Evaluation of OCO-2 $X_{\text{CO}_2}$ Variability at Local and Synoptic Scales using Lidar and In Situ Observations from the ACT-America Campaigns

Emily Bell, Christopher W. O’Dell, Kenneth J. Davis, Joel Campbell, Edward Browell, A. Scott Denning, Jeremy Dobler, Wayne Erxleben, Tai-Fang Fan, Susan Kooi, Bing Lin, Sandip Pal, and Brad Weir

Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA; Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA; NASA Langley Research Center, Hampton, VA, USA; STARSS-III Affiliate, NASA Langley Research Center, Hampton, VA, USA; Spectral Sensor Solutions LLC, Fort Wayne, IN, USA; L3Harris Technologies, Fort Wayne, IN, USA; Atmospheric Science Division, Department of Geosciences, Texas Tech University, Lubbock, TX, USA; Universities Space Research Association, Columbia, MD, USA; NASA Goddard Space Flight Center, Greenbelt, MD, USA

Abstract

With nearly 1 million observations of column-mean carbon dioxide concentration ($X_{\text{CO}_2}$) per day, the Orbiting Carbon Observatory 2 (OCO-2) presents exciting possibilities for monitoring the global carbon cycle, including the detection of subcontinental column $CO_2$ variations. While the OCO-2 data set has been shown to achieve target precision and accuracy on a single-sounding level, the validation of $X_{\text{CO}_2}$ spatial gradients on subcontinental scales remains challenging. In this work, we investigate the use of an integrated path differential absorption (IPDA) lidar for evaluation of OCO-2 observations via NASA’s Atmospheric Carbon and Transport (ACT)-America project. The project has completed eight clear-sky underflights of OCO-2 with the Multifunctional Fiber Laser Lidar (MFLL)—along with a suite of in situ instruments—giving a precisely colocated, high-resolution validation data set spanning nearly 3,800 km across four seasons. We explore the challenges and opportunities involved in comparing the MFLL and OCO-2 $X_{\text{CO}_2}$ data sets and evaluate their agreement on synoptic and local scales. We find that OCO-2 synoptic-scale gradients generally agree with those derived from the lidar, typically to ±0.1 ppm per degree latitude for gradients ranging in strength from 0 to 1 ppm per degree latitude. $CO_2$ reanalysis products also typically agree to ±0.25 ppm per degree when compared with an in situ-informed $CO_2$ “curtain.” Real $X_{\text{CO}_2}$ features at local scales, however, remain challenging to observe and validate from space, with correlation coefficients typically below 0.35 between OCO-2 and the MFLL. Even so, ACT-America data have helped investigate interesting local $X_{\text{CO}_2}$ patterns and identify systematic spurious cloud-related features in the OCO-2 data set.

1. Introduction

Earth’s climate is warming significantly, and as it continues within the next few centuries, it will bring about extreme economic, social, and ecological change. Greenhouse gases are the primary driver of this warming. Earth’s most influential greenhouse gas, carbon dioxide ($CO_2$), has risen from 280 to approximately 400 parts per million (ppm) in atmospheric mole fraction since the start of the Industrial Revolution, when the burning of fossil fuels began. However, while it seems straightforward to assume a correlation between anthropogenic emissions and atmospheric concentrations, the relationship between the two is extremely complex. In fact, an estimated 50% of $CO_2$ from anthropogenic sources is removed from the atmosphere via carbon sinks, although the location of these sinks remains uncertain (Le Quéré et al., 2018). Atmospheric $CO_2$ measurements help to shed light on the complex nature of the carbon cycle—historically, primarily through ground-based observation networks. The relatively sparse data density of such networks, however, limits their ability to resolve continental and ocean basin fluxes. Within the past decade, remote sensing of $CO_2$ has developed the potential to fill in the spatial gaps of ground-based data sets. The Orbiting Carbon Observatory 2 (OCO-2, launched in 2014) produces a $CO_2$ column-averaged dry air mole fraction ($X_{\text{CO}_2}$) data set with sufficient global coverage and sampling to complement ground-based in situ networks.
Typical column-integrated concentrations of $X_{CO_2}$ vary globally from around 380 to 420 ppm, with an interhemispheric gradient on the order of 2 ppm when averaged annually (Tans et al., 1990). It is with such precision in mind that the OCO-2 instrument was designed. Nearly 5 years of OCO-2 data have now successfully been shown to achieve this level of precision, 0.1–0.5%, or about 0.4–2 ppm error per individual sounding (O’Dell et al., 2018; Wunch et al., 2017).

However, spatial differences in CO$_2$ within continents and weather systems are considerably larger than 2 ppm and can be used to constrain fluxes with subcontinental and subseasonal spatial and temporal resolution. Large differences in tropospheric CO$_2$ associated with the passage of midlatitude cyclones have been identified using tower-based observations (Hurwitz et al., 2004), and consequently simulated numerically (Chan et al., 2004; Wang et al., 2007), but the spatial density of ground-based data was not sufficient to draw direct, synoptic-scale comparisons between models and observations. Large subcontinental gradients in atmospheric boundary layer (ABL) CO$_2$ have also been identified in areas of strong net uptake of CO$_2$ by agriculture. Miles et al. (2012) documented monthly mean daytime CO$_2$ differences of 10 ppm between tower sites in the corn belt of the upper Midwestern United States and tower sites in surrounding forest or grassland ecosystems. Ogle et al. (2015), Schuh et al. (2013), and Lauvaux et al. (2012) successfully used these strong gradients as the basis of inferences of net CO$_2$ uptake by these ecosystems.

Variations in the CO$_2$ field on these scales produce corresponding variations in the column-averaged value. On regional to local scales, spatial variations in $X_{CO_2}$, driven primarily by synoptic-scale transport, can be as large as several ppm (Keppel-Aleks et al., 2011; Pal et al., 2019). These subcontinental gradients should be detectable by OCO-2 and would serve as powerful constraints on regional CO$_2$ fluxes, similar to the results achieved with tower networks (e.g., Lauvaux et al., 2012; Schuh et al., 2013), but with data available around the globe.

Some work has already been done to evaluate OCO-2 local-scale $X_{CO_2}$ observations and subcontinental variability. On local scales, a study by Nassar et al. (2017) diagnosed estimates from individual power plants in the United States, India, and South Africa using OCO-2 overpass data. Validation on scales this small is difficult, though, as demonstrated by the Nassar et al. (2017) data set: in 2 years of data, only six clear-sky scenes were colocated with strong point sources suitable for case studies. On synoptic scales, a comparison study by Wunch et al. (2017) between OCO-2 and the Total Column Carbon Observing Network (TCCON) shows that after bias correction (against TCCON) and filtering, the OCO-2 $X_{CO_2}$ version 7 target-mode data, typically spanning approximately 0.2° longitude by 0.2° latitude, as well as nadir and glint data in boxes of 5° latitude by 10° longitude, can still have residual biases of up to 1.5 ppm in root mean square (RMS) difference from their colocated TCCON sites. Worden et al. (2017) evaluated sources of variability on scales of 100×10.5 km$^2$ by using comparisons of observed and predicted noise, and concluded both that some uncertainty in local-scale gradients remains unaccounted for, and that some spurious soundings may be due to variations in surface properties or solar zenith angle. The OCO-2 versions 7 and 8 data did reveal some sensitivity to surface features such as topography and albedo, though in large part, the adjustments made in version 9 have eliminated the topographical sensitivity (Kiel et al., 2019). There is also potential for spurious soundings due to clouds and aerosols, as OCO-2 relies on reflected sunlight (it is a “passive” instrument) and thus cannot distinguish between photons from intermediate scatterers and photons reflected from Earth’s surface. This can lead to modified optical path lengths, effectively changing the extent of the measured column and producing anomalous $X_{CO_2}$ values (O’Dell et al., 2018). Such anomalies in the OCO-2 data set have the ability to affect the observed subcontinental spatial features, but validation of those spatial features has, up to this point, been lacking.

NASA has funded the Atmospheric Carbon and Transport (ACT)-America mission in part to help evaluate OCO-2 $X_{CO_2}$ observations at subcontinental scales. The ACT-America mission goals are threefold: to quantify and reduce atmospheric transport uncertainties; to improve regional-scale, seasonal prior estimates of CO$_2$ and CH$_4$ fluxes; and to evaluate the sensitivity of OCO-2 column measurements to regional variability in tropospheric CO$_2$ (Davis et al., 2017).

The first two goals can be achieved with in situ sensors, with which both mission aircrafts (NASA C-130 and B-200) are equipped. These are utilized in fair weather conditions and near frontal boundaries, to help assess surface fluxes and transport of CO$_2$ due to mesoscale and synoptic transport mechanisms. Related to the third goal and subcontinental spatial gradients, Davis et al. (2018) have analyzed ACT-America flight data and found CO$_2$ differences throughout the depth of the troposphere, ranging from a few ppm in the
upper free troposphere to tens of ppm in the ABL. Large spatial gradients have also been identified in fair weather conditions, though these are most often limited to the ABL (Davis et al., 2017).

In pursuit of ACT-America’s third goal, the Multifunctional Fiber Laser Lidar (MFLL)—jointly developed, evaluated, and demonstrated by ITT Exelis (now part of the Harris Corp.) and NASA Langley Research Center (LaRC) —was flown on the C-130 along OCO-2 orbital tracks, at (nearly) concurrent times and at an altitude typically between 8 and 9 km. The MFLL measures CO₂ column optical depths between the C-130 and the surface, and this measurement is converted to XCO₂ via a simple retrieval algorithm (Dobbs et al., 2008; Dobler et al., 2013). We use the MFLL as a benchmark against which to compare the OCO-2 observations, as it is an active instrument and does not suffer from the sensitivities to scattering from even thin layers of cloud and aerosol, from which the passive satellite measurements can suffer. MFLL data are taken at a frequency of 10 Hz along track, or 7- to 9-km spatial resolution when averaged into 60-second bins. A diagram of a typical OCO-2 underflight plan is shown in Figure 1b, along with its geographic location in Figure 1c. A total of 12 OCO-2 underflights were completed during four campaigns, three in each season, and the eight clear-sky flights used in this study are shown in Figure 1a. This work seeks to assess the agreement of spatial gradients and local-scale variability between the OCO-2 and MFLL XCO₂ data sets for eight relatively clear-sky cases, with assistance from in situ data and models.

This comparison is nuanced due to certain systematic differences between the MFLL and OCO-2 measurements. First, they sample different portions of the atmospheric column. OCO-2 is in a polar orbit in the NASA A-Train (or Afternoon Train), a series of satellites which cross the equator along nearly the same path around 1:30 pm local time in their ascending orbits. OCO-2 effectively measures CO₂ absorption from the full atmospheric column. The MFLL, on the other hand, flies on the NASA C-130 at a nominal altitude of...
9 km and measures the column between this altitude and the surface. The inclusion or exclusion of CO₂ concentrations between 9 km and top of atmosphere (TOA) has a nonzero effect on the column average, which must be accounted for in these comparisons. In addition, OCO-2 and MFLL measurements weight the CO₂ absorption in the vertical layers of the atmosphere differently—pushing the column average toward those concentrations with higher weights in each case. This, too, has a nonzero effect on column average values and must be accounted for.

We seek to validate two aspects of the OCO-2 data set in this work: spatial gradients at synoptic scales, on the order of hundreds of kilometers, and variability on local scales, which we define as a few tens of kilometers. These validation efforts will additionally shed light on the capabilities of MFLL, in situ data, and CO₂ reanalysis products as validation tools, and will help identify remaining sources of spurious features in the OCO-2 XCO₂ data set.

The contents of this paper are thus structured as follows: section 2 gives an overview of the OCO-2 and MFLL instruments and data sets, including some details of the retrieval process and MFLL retrieval sensitivities, as well as an overview of the additional validation data sets used. Section 3 details the methods we used in this project and contains two subsections. Section 3.1 describes our model-derived “adjustments,” which enable an “apples-to-apples” comparison between the various data sets, and section 3.2 describes the statistics we use in the along-track comparisons between data sets. Section 4 shows results and provides discussion in the context of the goals outlined above. We conclude with a summary of our findings thus far, their implications, and a discussion of future work, in section 5.

2. Instruments and Data sets

2.1. OCO-2

Launched in July 2014, OCO-2 travels a sun-synchronous polar orbit in NASA’s A-Train and records nearly 1 million soundings per day on eight footprints across its approximately 10-km swath, with an average single-footprint area of about 1.25 by 3 km. Around 100,000 of these are sufficiently cloud free and suitable for full-column XCO₂ analysis; the XCO₂ data typically achieve a precision of 0.3% (1 ppm) or better, in line with OCO-2’s mission goals and instrument design (Eldering et al., 2017). OCO-2 completes a global cycle (between about 60N and 60S) every 16 days.

The OCO-2 instrument is a passive sensor, using a high-resolution grating spectrometer to retrieve, from reflected sunlight, the amount of atmospheric absorption due to CO₂ in two spectral bands: the weak band at 1.6 μm, and the strong band at 2.0 μm. It also measures absorption in the O₂ A-band at 0.76 μm. The instrument measures at high spectral resolution, with 1,016 spectral channels in each of the three bands (Eldering et al., 2017).

Cloud-free soundings from the two CO₂ bands are processed through the Atmospheric Carbon Observations from Space (ACOS) retrieval algorithm to produce the operational column-averaged XCO₂ product, which is used in this study (Boesch et al., 2019). The ACOS retrieval algorithm employs an optimal estimation scheme, which, in using an a priori estimate and observed values, a cost function is minimized to determine the most probable posterior state vector, as described in O’Dell et al. (2012) and O’Dell et al. (2018).

The data used in this study are version 9 (OCO-2 Science Team/Michael Gunson, 2018) and have been pre-filtered using rough cloud and aerosol screens described in Taylor et al. (2015). We also apply the standard quality flag filter, which depends on a variety of factors including observed surface albedo, surface roughness, aerosol optical depth (AOD), and offset from prior model estimates. The data analyzed in this study employ the standard bias correction (Kiel et al., 2019) and have quality flag 0 (good).

The OCO-2 XCO₂ quantities required by this study are partial column values corresponding to the partial column sampled by the MFLL. As the OCO-2 operational product provides no validated or bias-corrected partial column values, we calculate the partial column average using some simple assumptions and a model CO₂ profile; section 3.1 discusses the method of this calculation, and our estimates of its uncertainty, in more detail. We will refer to this calculation as one of our model-derived “adjustments.”

2.2. Multifunctional Fiber Laser Lidar

2.2.1. MFLL Instrument and Retrievals

The MFLL, developed by Harris Corp., with support and evaluation by NASA’s LaRC (Dobbs et al., 2008; 2013), was designed specifically to address the goals set by the Active Sensing of CO₂ over Nights, Days, and
Seasons (ASCENDS) mission and is the first intensity-modulated continuous-wave (IM-CW) laser absorption spectrometer (LAS). This section provides only a brief overview of the MFLL and its operation during ACT-America; further details of the MFLL instrument for ACT-America can be found in Campbell et al. (2020).

Figure 2 shows the standard setup of sampling wavelengths on the MFLL absorption line. Operational since 2005, the MFLL emits at three wavelengths on a single CO$_2$ absorption line centered at 1571.112 nm, which was chosen for MFLL operation due to its relative insensitivity to both relative humidity (RH) and temperature. One emission wavelength is at the center of the line (referred to as the “ON” line), and the other two are placed further out on the wings (“OFF” lines). For all but one flight (5 August 2016) in this study, the configuration had the two OFF wavelengths at 50 picometers to either side of the ON line. We refer to these as the short (S) and long (L) lines, at 1571.062 and 1571.162 nm respectively. The MFLL CO$_2$ optical depth ($\tau$) column measurement uses the integrated path differential absorption (IPDA) method (Browell et al., 1979), in which the differential absorption ($\Delta \tau$) between two of the three sampling wavelengths—either ON and OFF$_S$ or ON and OFF$_L$—is calculated by taking ratios of the transmitted and received powers for those two wavelengths. The two retrieval results from the two wavelength pairs can differ by up to about 0.5% (or about 2 ppm) along a given flight track. This is presumably due in part to the higher water vapor sensitivity of the OFF$_L$ wavelength, and for partly this reason, the ON/OFF$_S$ retrieval alone has been selected for Level 1 and Level 2 processing (Lin et al., 2018).

The MFLL column CO$_2$ retrieval algorithm relies on a relatively simple manipulation of Beer’s law given the ratio of the signal extinction due to CO$_2$ at the two wavelengths, ON and OFF. Equation (1) shows the key relation of the calculation, which serves as the basis of all differential absorption lidar (DIAL) measurements (Korb & Weng, 1983).

\[
\ln \left( \frac{P_{R,OFF}}{P_{R,ON}} \cdot \frac{P_{T,ON}}{P_{T,OFF}} \right) = \frac{2}{\mu} (\Delta \tau_{CO_2} + \Delta \tau_{H_2O})
\]  

(1)

In this equation, the received and transmitted powers, $P_r$ and $P_t$, respectively, are measured; $\Delta \tau_{H_2O}$ is assumed to be known, either from a model (in the case of the MFLL) or observations—or both, in the case of OCO-2; and the target variable is $\Delta \tau_{CO_2}$. The variable $\mu$ is the cosine of the viewing angle $\theta$, and $\Delta \tau = \tau_{ON} - \tau_{OFF}$ is known as the differential optical depth ($\Delta \tau$). The optical depth of any gas between two arbitrary pressures $P_{lower}$ and $P_{upper}$ at a given wavelength can be calculated as

\[
\tau_{gas}(\lambda) = \int_{P_{lower}}^{P_{upper}} \sigma_{gas}(\lambda, P, T, q) \, dN_{gas}.
\]  

(2)
where $\sigma_{\text{gas}}$ is the gas absorption cross section and $N_{\text{gas}}$ is the number of molecules of that gas per unit area. The absorption cross section $\sigma_{\text{gas}}$ depends on wavelength ($\lambda$), pressure ($P$), temperature ($T$), and humidity ($q$). Instruments such as OCO-2 and the MFLL, built particularly to gather data on CO$_2$, rely on model data for estimates of all but wavelength. $N_{\text{gas}}$ also depends on both pressure and humidity and thus also relies on model data. Spectroscopic tables are used to look up estimates of the corresponding cross sections once estimates of $T$, $P$, and $q$ can be obtained from either in situ spiral measurements or, as in this study, a meteorological model. In this study, we primarily use a HITRAN 2016 precursor table (later referred to as pre-HITRAN16) for $\sigma_{\text{CO}_2}$, generated with an open-source code available online at the HITRAN website (Gordon et al., 2017). The CO$_2$ spectroscopy uses first-order line mixing with a speed-dependent Voigt lineshape. We use HITRAN2012 tables for $\sigma_{\text{H}_2\text{O}}$ (Rothman et al., 2013) and MERRA2 reanalysis data for meteorological inputs (Gelaro et al., 2017). For the full details of our spectroscopy models and associated assumptions, please see Campbell et al. (2020).

To calculate $X_{\text{CO}_2}$ from the $\Delta r_{\text{CO}_2}$ in equation (1), we could in principle use optimal estimation, as is done in many retrievals, including OCO-2. However, because we can write down one equation and one unknown, we use a simple analytical estimate of $X_{\text{CO}_2}$ as follows:

$$X_{\text{CO}_2, \text{obs}} = \frac{\Delta r_{\text{CO}_2, \text{obs}} - \Delta r_{\text{H}_2\text{O}, \text{ap}}}{\Delta r_{\text{CO}_2, \text{ap}}} \times X_{\text{CO}_2, \text{ap}}.$$  \hspace{1cm} (3)

This equation assumes that the true column of CO$_2$ can be calculated by scaling a prior estimation ($ap$). We report retrieval results which use a uniform prior column value of 400 ppm. The resultant $X_{\text{CO}_2, \text{obs}}$ depends explicitly on the water vapor profile used to calculate $\Delta r_{\text{H}_2\text{O}}$. It also depends upon the assumed temperature, surface pressure, and spectroscopic models used for the calculation of $\sigma$ in equation (2). We discuss these retrieval sensitivities in further detail in section 2.2.2.

Early tests of the MFLL instrument demonstrated successful $X_{\text{CO}_2}$ measurements in a variety of conditions and over a variety of surface types (Browell et al., 2008; Dobler et al., 2013). The first flight campaigns found signal-to-noise ratios (SNRs) higher than 250 for 1-s averaged data over land and yielded CO$_2$ concentrations as precise as 0.6 ppm, in line with the ≤1 ppm precision of OCO-2 measurements (Browell et al., 2008; 2010). A series of 2010 flights measured SNRs better than 600 over desert, 500 over vegetation, and 150 over ocean for 1-s data; 2011 flights showed $X_{\text{CO}_2}$ column agreement to within 0.65 ppm between MFLL and in situ measurements (Dobler et al., 2013). While the MFLL data in this study are averaged to 60 s, we found similar noise levels in the data as compared with the literature.

A key feature of the MFLL as an active remote sensing instrument is the ability to identify intermediate scatterers in the observed column. This is accomplished via a range-encoded intensity-modulation technique which, using the magnitude and timing of the returned signal, differentiates surface-reflected signals from those reflected off of intermediate scatterers. The timing of the returned signal can be used to calculate the range to the surface, which has been shown to be better than 3 m in both precision and accuracy (Dobler et al., 2013; Lin et al., 2013; 2015). This level of accuracy was also found in range calculations to cloud and aerosol levels.

Native MFLL measurements are made at 10 Hz, and for OCO-2 underflights, the C-130, carrying the MFLL, flies nominally at an altitude of 9 km above sea level. The winter 2017 campaign was an exception, as a coating degradation on the MFLL viewing window significantly reduced measurement SNR, so the C-130 flew at a maximum of around 6 km along OCO-2 underflight tracks. MFLL optical depths throughout the mission demonstrated some range-dependent bias, the cause of which is currently unknown. As discussed more thoroughly in Campbell et al. (2020), the value of this bias varied between individual flights, but by comparing MFLL optical depth measurements to in situ spiral data from the aircraft, a line was fit to the data to determine the difference in lidar optical depth versus in situ optical depth. An average bias correction for each campaign was empirically developed by the LaRC data processing team and applied to all flights in each campaign, effectively removing this range dependence. We note that the bias correction development for the summer 2016 campaign differs from that of the proceeding campaigns; updates to the summer 2016 data are ongoing, but were unfinished at the time of publication, and will be available in the next version of the data release.

The data used in this study come from the Level 1 (L1) files, which are publicly available online (ACT-America Data Archive, 2019). A variety of filters are involved in the L1 data processing, descriptions
of which can be found in the archive as well. The L1 cloud_ground_flag is used to filter out optically thick cloudy soundings, and the data_quality_flag filters out soundings with poor signal strength from the surface, as well as pitch and roll angles larger than 5°. The 10-Hz L1 data are then averaged into 1-s bins to reduce retrieval run time, and retrievals are run on the resulting 1-s ODs, providing us with 1-s XCO₂ data.

Due to the nature of the retrieval, the MFLL XCO₂ calculation employs a pressure weighting function that is quite different from that of OCO-2. We apply a model-derived “adjustment” to the 1-s average values, which accounts for this effect and makes the OCO-2 and MFLL retrieved values more directly comparable. The details of this adjustment will be explored further in section 3.1. After the adjustment is applied, the MFLL XCO₂ data are further averaged to 60 s for our comparisons to OCO-2.

2.2.2. MFLL Retrieval Sensitivities
As previously stated, the simplicity of the MFLL retrieval leads to some significant sensitivity to spectroscopic and meteorological inputs. There is also a sensitivity to the assumed wavelength of the ON and OFF channels, but we found that these wavelengths were determined accurately in flight and were stable enough to not result in a substantial retrieval error (Campbell et al., 2020). Sensitivities to CO₂ spectroscopy, H₂O spectroscopy, and meteorology have been tested as follows. We use a baseline setup for retrieval input variables: pre-HITRAN2016 CO₂ spectroscopy (Gordon et al., 2017; Campbell et al., 2020), HITRAN2012 H₂O spectroscopy, and meteorology...
spectroscopy (Rothman et al., 2013), and MERRA2 meteorology (Gelaro et al., 2017). We then test two alternatives for each input variable. To test CO₂ spectroscopy, for example, we hold the H₂O spectroscopy and meteorology constant by using HITRAN2012 and MERRA2, respectively, for all retrievals. We then calculate X_{CO₂} using first the pre-HITRAN2016 CO₂ lookup tables, then HITRAN2012, and then ABSCO4.2 (Payne & Thompson, 2013). The results are three different sets of retrieved X_{CO₂}, with CO₂ spectroscopy the only difference between them. We assume that the differences between these three spectroscopic models are of a similar magnitude to the difference between our default spectroscopy model and the true spectroscopy. The results of this test are shown in Figure 3a. We can use these results to evaluate the effects of CO₂ spectroscopy on the spatial variability of the data. On synoptic scales, we use a linear least-squares fit for the three retrieved data sets in Figure 3a to determine their linear south-to-north (S–N) gradients in X_{CO₂}. The range of these three gradients is referred to as “Δgrad” and represents the uncertainty in the MFLL gradient associated with CO₂ spectroscopy. Figure 3b differences the two alternate retrievals (denoted HT12 and ABSCO in the figure) from the baseline retrieval and makes the difference in the gradients more clearly visible.

To evaluate the effects of CO₂ spectroscopy on smaller-scale variability, we use the retrieval differences in Figure 3b. We remove their large-scale S–N gradients and are left with Figure 3c. We use the standard deviation of these values as a measure of the overall change in small-scale variability along track, which we call Δδ. From Figure 3c, we obtain two values of Δδ, and we can difference the HT12 and ABSCO results for a third. We report the mean of these three values as a representation of the expected uncertainty in small-scale variability due to uncertainty in CO₂ spectroscopy.

We repeat these calculations in a similar fashion to test the effects of H₂O spectroscopy and meteorology on the S–N gradient Δgrad and small-scale variability Δδ. Figures 3d–3f and 3g–3i show these results. We find generally among the eight flights that the MFLL retrieval is particularly sensitive to water vapor via both the meteorology and spectroscopy, as evidenced by the large differences in retrieved X_{CO₂}, and relatively large small-scale along-track variability when testing H₂O spectroscopy (Figures 3e and 3f). The similar spatial patterns between Figures 3f and 3i support this—while changing meteorology inputs changes many variables, the water vapor profile is one of them, and changes to meteorology produce similar spatial patterns to changes in the H₂O spectroscopy. We surmise that this sensitivity to water vapor is due to the nontrivial amount of water vapor absorption in the MFLL wavelength region around 1571.11 nm.

In addition to the tests above, we also examine the impact of the prior profile shape on the retrieved X_{CO₂}, in a similar manner. Figure 4 shows the results of these tests for two different dates. For each case, we run the retrieval with a flat profile of 400 ppm (“Flat_400”) and then using the two resampled models which compare best to the in situ spiral on that day. We find that in winter/fall cases (Figures 4d–4f) where the model and in situ profiles are relatively flat, the retrieved X_{CO₂} shifts by a few tenths of a ppm at most. In summer cases where there is strong CO₂ drawdown near the surface (Figures 4a–4c), the profile has a sharp vertical gradient, and the retrieved X_{CO₂} can shift by 1 ppm or more. However, via Δgrad and Δδ, we find that this has an insignificant effect on the synoptic and local spatial variability in most cases.

Table 1 summarizes Δgrad and Δδ numerically for all eight flights. These sensitivity tests help establish our uncertainties, particularly in the gradients. We can define an MFLL gradient uncertainty due to retrieval variability as

$$
\epsilon_{ret} = \sqrt{\Delta\text{grad}^2_C + \Delta\text{grad}^2_H + \Delta\text{grad}^2_M + \Delta\text{grad}^2_P},
$$

where subscripts C, H, M, and P stand for CO₂ spectroscopy, H₂O spectroscopy, meteorology, and prior, respectively. \( \epsilon_{ret} \) represents only one piece of the MFLL gradient uncertainty, particularly that associated with input assumptions to the retrieval; we discuss additional sources of MFLL uncertainty in section 3.

2.3. Global Models

In this study, we use three global models to serve as both a reference against which to compare our observations, as well as a tool in making our model-derived “adjustments,” which are more fully described in section 3.1.

The Copernicus Atmosphere Modeling Service (CAMS) provides a forecast model of atmospheric components as a part of the European Centre for Medium-Range Weather Forecasts (ECMWF) (Agusti-Panareda et al., 2014). The CO₂ forecasts assimilate the GOSAT BESD X_{CO₂} product (Heymann et al., 2015), using a
4D-Var system to optimize fluxes. In this study, we use CAMS data at 6-hourly time steps on a 0.25° latitude by 0.25° longitude grid, with 137 vertical levels from surface to top of atmosphere (TOA). We calculate partial-column $X_{CO2}$ values from CAMS as a means of first-order validation and include CAMS adjustments to OCO-2 and the MFLL in our estimate of model-derived error.

CarbonTracker, a product based on posterior flux products from NOAA ESRL in Boulder, Colorado, provides near-real-time analyses which include ground-based flask measurements (Peters et al., 2007, with updates

| Table 1 |

| MFLL Retrieval Sensitivities Summary |

| Date       | $CO_2$ spec | $H_2O$ spec | Met | Prior $CO_2$ spec | $H_2O$ spec | Met | Prior |
|------------|-------------|-------------|-----|-------------------|-------------|-----|-------|
| 27 July 2016 | 0.096       | 0.244       | 0.191 | 0.107             | 0.085       | 0.372 | 0.319 | 0.093 |
| 5 August 2016 | 0.076       | 0.116       | 0.081 | 0.038             | 0.044       | 0.119 | 0.150 | 0.040 |
| 15 February 2017 | 0.015       | 0.019       | 0.042 | 0.018             | 0.018       | 0.012 | 0.039 | 0.014 |
| 8 March 2017 | 0.272       | 0.017       | 0.238 | 0.023             | 0.044       | 0.018 | 0.199 | 0.021 |
| 22 October 2017 | 0.067       | 0.095       | 0.154 | 0.008             | 0.011       | 0.020 | 0.069 | 0.005 |
| 27 October 2017 | 0.058       | 0.018       | 0.080 | 0.044             | 0.019       | 0.004 | 0.064 | 0.012 |
| 27 April 2019 | 0.031       | 0.114       | 0.199 | 0.003             | 0.018       | 0.035 | 0.078 | 0.019 |
| 29 April 2018 | 0.020       | 0.069       | 0.167 | 0.016             | 0.020       | 0.012 | 0.066 | 0.017 |
Figure 5. (a) The PICARRO CO$_2$ data in latitude from both the B-200 and C-130 flight tracks on 27 July 2016. (b) The resulting “curtain” after assimilation of the PICARRO data into the GMAO GEOS-5 model.
Figure 6. Operational retrieval weighting schemes compared with a straight pressure weighting scheme. Note that while the red line is a true weighting function, the yellow is the OCO-2 averaging kernel: retrieved X\textsubscript{CO}_2 from OCO-2 is actually a weighted average of the “observed” profile and an a priori guess. The yellow values shown here represent the weights used on the “observed” profile—in this example, a sample CO\textsubscript{2} profile taken from the CAMS model.

curtain can be sampled in time and space as desired to make X\textsubscript{CO}_2 estimates along the full OCO-2 under-flight track. We sample the curtain column up to the C-130 height to produce a partial-column value which uses a straight pressure weighting function. This data set will be listed as “GMAO\_ACT” in plot legends. As described in the next section, we attempt to put both the MFLL and OCO-2 on similar footing with the curtain and model X\textsubscript{CO}_2 values—which are all straight pressure-weighted—by accounting for their respective weighting function/averaging kernel effects.

3. Methods

This section consists of two subsections. Section 3.1 details the model-derived “adjustments” which we make to OCO-2 and the MFLL, and section 3.2 details the statistics we use to draw our along-track comparisons between the two instruments and both synoptic and local scales.

3.1. Column Extent and Weighting Function Adjustments

While OCO-2 has a relatively mature X\textsubscript{CO}_2 retrieval, the MFLL retrieval is quite new, and its X\textsubscript{CO}_2 observations from the approach described in section 2.2.1 have not been thoroughly tested. Because the MFLL and OCO-2 observe different atmospheric columns, their measurements are different in systematic ways. In this subsection, we describe the main differences between MFLL and OCO-2 column CO\textsubscript{2} observations, and analyze their effects on the retrieved X\textsubscript{CO}_2 values. The subsections detail our methods of accounting for those effects, which we call model-derived “adjustments.”

Figure 6 illustrates the differing principles of the OCO-2 and MFLL X\textsubscript{CO}_2 calculations, using model CO\textsubscript{2} profiles taken from CAMS.

The first of two key differences between the MFLL and OCO-2 measurements is the vertical extent of the sampled column. OCO-2, flying well above TOA, samples CO\textsubscript{2} absorption throughout the full atmospheric
Figure 7. Top panels show CAMS XCO₂ values for both full-column average and lower-column average beneath the mean height of the C-130 aircraft, for one summer and one winter example. The bottom plots show the difference between these averages.

column. The MFLL, on the other hand, pointing at the surface from the altitude of the C-130, only samples the atmospheric column up to the aircraft height—typically around 9 km, or between 300 and 400 hPa. Since the portion of the column above this level (which we will refer to as the “upper column,” or A, the column above the aircraft) typically accounts for approximately 30% of the total column mass—and includes the typically lower CO₂ concentrations of the upper troposphere and stratosphere, as shown in Figure 6—its inclusion in the column average can have a large effect on the retrieved XCO₂ value. Figure 7 shows that, especially in the winter months when there is no surface drawdown of CO₂, the difference between a full-column average and a partial-column average can be multiple ppm in magnitude and can display spatial variability both on track-length (synoptic) scales and on smaller (local) scales. We conclude that full-column and partial-column values are not directly comparable. In order to highlight signals from surface sources and sinks, we choose to focus on partial columns, so our first “adjustment” will be to calculate a partial-column value for OCO-2. This is the topic of section 3.1.1. We will refer to this partial column below the aircraft as B, or the “lower column.”

The second key difference between the MFLL and OCO-2 data sets is the shape of the weighting functions used in the retrievals to calculate XCO₂. The ACOS retrieval for OCO-2 produces an a posteriori XCO₂ value

Figure 8. Top panels show 2 days of CAMS XCO₂ results when applying the weighing schemes shown in Figure 6 to the CO₂ profiles in the lower column. Blue is differenced from red and yellow to create the bottom panel.
which is essentially a weighted average of an a priori guess and the observation, as given in equation (5) below. The weights used on the observation are called the averaging kernel and are shown in yellow in Figure 6 for a sample case. These weights taper off in the upper column, but are very nearly flat in the lower column. The MFLL retrieval, on the other hand, calculates \(X_{\text{CO2}}\) using a pressure weighting function which, compared with OCO-2, overweights \(\text{CO}_2\) concentrations near the aircraft and underweights \(\text{CO}_2\) concentrations near the surface. We discuss the source of this weighting function shape in section 3.1.2. In the example shown in Figure 6, this would result in a higher column average: \(\text{CO}_2\) values above 700 hPa, which are consistently greater than 401 ppm, get weighted more strongly than the \(\text{CO}_2\) values below 700 hPa, which are consistently less than 400 ppm. Figure 8 shows that for lower-column \(X_{\text{CO2}}\) calculations, the OCO-2 averaging kernel yields a similar \(X_{\text{CO2}}\) value to a straight-pressure-weighted calculation, but the MFLL weighting function can lead to differences of nearly 2 ppm in summer, with notable changes in retrieved spatial variability as well. Given the comparability of the OCO-2 retrieval and the straight-pressure-weighted retrieval, the simplest way to establish comparability between all data sets is to account for the influence of the MFLL weighting function shape on the retrieved MFLL \(X_{\text{CO2}}\). This will be our second adjustment, detailed in section 3.1.2.

3.1.1. OCO-2 Adjustments: Full Column to Partial Column

The first adjustment we use to ensure comparability among the data sets is the calculation of a partial-column \(X_{\text{CO2}}\) for OCO-2. We start with the basic principles of the ACOS retrieval: for a given sounding, we use the following formula to calculate \(X_{\text{CO2}}\):

\[
X_O = \sum_{i=1}^{20} h_i (a_i u_i + (1 - a_i) u_{\text{ap},i}).
\]

where \(X_O\) is the full-column average as retrieved by OCO-2. It is calculated as the weighted average of the “true” or “observed” vertical \(\text{CO}_2\) profiles, \(u\), and an a priori guess, \(u_{\text{ap}}\). The weights of \(u\) and \(u_{\text{ap}}\) are dictated by the normalized averaging kernel \(a\). The weights of each vertical level, of which there are 20 in ACOS, are determined by the pressure weighting function \(h\). For more details on the ACOS retrieval, refer to O’Dell et al. (2012).

In theory, we could calculate \(X_O\) for any partial column we like by using a chosen subset of \(u, u_{\text{ap}}, h,\) and \(a\), renormalizing \(h\) and \(a\) as appropriate. However, the only operationally bias-corrected and validated quantity from the ACOS retrieval is the full-column \(X_{\text{CO2}}\). The vertical profile is neither validated nor bias corrected in any way, so the fidelity of the retrieved \(\text{CO}_2\) profile is fairly nebulous. Thus, instead of using the retrieved \(u\) in our partial-column calculations, we use a model profile, \(u_m\).

We use \(u_m\) to calculate the upper column average \(A_O\) in the same way as \(X_O\), using subsets of \(h\) and \(a\) corresponding to this portion of the column. We denote these subsets with a subscript \(A\): \(\hat{h}_A\) is the appropriate subset of the weighting function and is renormalized according to equation (6):

\[
\hat{h}_A, i = \frac{h_i}{\sum_A h_i},
\]

where the summation is understood to only include levels above the aircraft. We thus have

\[
A_O = \sum_A h_{A,i} (a_i u_{m,i} + (1 - a_i) u_{\text{ap},i})
\]

for the partial column average above the aircraft. Then, given the ACOS-retrieved value of the full column \((X_O)\), we assume a simple weighted average to estimate the \(X_{\text{CO2}}\) of the lower portion, as given by

\[
X_O = f_A A_O + (1 - f_A) B_O,
\]

where \(B_O\) is the partial-column average we seek. Weights \(f_A\) and \(1 - f_A\) represent the fraction of the column above and below the plane, respectively, calculated using the native OCO-2 pressure weighting function. \(f_A\) is typically about 30% of the column.

We can solve equation (8) for \(B_O\) to calculate the partial-column \(X_{\text{CO2}}\) below the aircraft, as in equation (9). This is our “adjusted” OCO-2 measurement quantity:

\[
B_O = \frac{1}{1 - f_A} (X_O - f_A A_O).
\]
Figure 9. (a) Visualization of the separate pieces involved in the OCO-2 adjustment calculation. A is derived from a model. (b, c) OCO-2 adjustments ($\Delta_{\text{OCO2}}$ according to equation 10) as derived from two different models, for one summer and one winter flight. The gray line shows the absolute value of their difference.

An example of the values and physical extent of $X_O$, $A_O$, and $B_O$ is shown in Figure 9a. We define the “OCO-2 adjustment” itself, or $\Delta_{\text{OCO2}}$, as the change induced in $X_{CO2}$ such that

$$\Delta_{\text{OCO2}} = B_O - X_O.$$  \hspace{1cm} (10)

The values of $\Delta_{\text{OCO2}}$ are shown in Figures 9b and 9c when derived from two different model profiles. The value of this adjustment can be as large as several ppm, can vary significantly in space—especially in summer (left)—and can differ by up to 1 ppm between models. The difference between model adjustments gives an indication of the uncertainty in this adjustment, and we will discuss the incorporation of this uncertainty into our comparisons in Section 3.2.1.

### 3.1.2. MFLL Adjustments: Pressure Weighting Function Effects

As discussed in the beginning of section 3.1, the second key difference between the OCO-2 and MFLL partial-column retrievals is the shape of their weighting schemes. The OCO-2 averaging kernel is comparably flat in the column below the aircraft, but the MFLL weighting function is curved in such a way that can affect both the magnitude and variability of the retrieved $X_{CO2}$.

As done for OCO-2, a straight pressure weighting function can be defined as

$$h_i = \frac{N_{dry} dz_i}{\sum N_{dry} dz_i},$$  \hspace{1cm} (11)

where $N_{dry} dz_i$ is the number of molecules per square meter of dry air in atmospheric layer $i$. The MFLL pressure weighting function, on the other hand, is defined as

$$w_i = \frac{\Delta \tau_{CO2,i}}{\sum \Delta \tau_{CO2,i}},$$  \hspace{1cm} (12)

per the MFLL retrieval as described in section 2.2.1. Via equation (2) in that section,

$$\Delta \tau_{CO2,i} = \Delta \sigma_{CO2,i} N_{dry} dz_i.$$  \hspace{1cm} (13)

Thus, the shape of the MFLL pressure weighting function is dependent on the shape of $\Delta \sigma$, which is the difference between $\sigma$ values at the ON and OFF_SHORT wavelengths. Pressure broadening dictates that $\Delta \sigma$ will be smaller near the surface and larger near the aircraft. This is the shape seen in red in Figure 6.
Figure 10. Visualization for the calculation and application of the MFLL adjustment. This example is taken from the location of the 27 July 2016 spiral profile, for observation-versus-model context.

We can now define the partial-column $X_{\text{CO}_2}$ from the MFLL ($B_M$) as well as the partial-column $X_{\text{CO}_2}$ using a straight pressure weighting function ($B$):

$$B = \sum_B h_i u_i,$$

$$B_M = \sum_B w_i u_i,$$

where $u_i$ is the CO$_2$ concentration on each vertical level $i$. To evaluate the impact of the MFLL weighting function on the retrieved $X_{\text{CO}_2}$, we can simply difference equations (14) and 15:

$$\Delta_{\text{MFLL}} = B - B_M = \sum_B (h_i - w_i) u_i.$$

Note here that because both $h$ and $w$ are normalized, each of their sums is equal to unity, and thus,

$$\sum_B (h_i - w_i) = 0.$$

Figure 11. MFLL adjustment ($\Delta_{\text{MFLL}}$ from equation 16) as calculated for (a) one summer date and (b) one winter date, using two different sets of model profiles. The difference between the two model-derived adjustments is shown in gray and represents an estimate of the uncertainty in the value of $\Delta_{\text{MFLL}}$. 
Figure 12. Comparisons between in situ and model profiles. At the bottom of each plot is the mean absolute difference between the in situ and model concentrations.

Therefore, the value of $\Delta_{\text{MFLL}}$ depends on the shape of the assumed profile $u$, and in the case of a flat CO$_2$ profile, would be identically zero. Since we do not explicitly measure the shape of the CO$_2$ profile, we rely on models to calculate $\Delta_{\text{MFLL}}$, which lends it an inherent uncertainty.

Our second model-derived adjustment is thus to apply a model-derived $\Delta_{\text{MFLL}}$ to the retrieved MFLL $X_{\text{CO}_2}$, making it more directly comparable with OCO-2, so that our final adjusted value is

$$B_{\text{adj}} = B_{\text{obs}} + \Delta_{\text{MFLL}}.$$  (18)

Figure 10 provides a visual representation of this adjustment. Figure 11 shows that the value of $\Delta_{\text{MFLL}}$ can be on the order of multiple ppm. It can differ by as much as 0.5 ppm between models, but varying in space.

Because both the MFLL and OCO-2 adjustments can vary between models, it is essential to use the model which most accurately represents the true atmospheric state on each date. In order to do this, we select the model that most closely matches the in situ spiral profile(s). For each case in which the GMAO curtain is available, it replicates the in situ values most closely. Figure 12 shows an example of two such comparisons.

Despite the good agreement between the curtain and in situ at the location of the spiral, there is one date where we use CT-NRT rather than the curtain for the adjustments. On 27 April 2018, the B-200 diverted from the OCO-2 underflight track for an urgent landing and refueling in an urban area. The PICARRO data from this diversion shows strong CO$_2$ enhancements in the boundary layer of the metropolitan St. Louis area; we find that the assimilation of this data increases boundary layer CO$_2$ significantly along much of the northern portion of the underflight track, while all other data sets indicate no such increase. Such an anomaly is unique to this date and renders the curtain unsuitable for MFLL weighting function adjustments.
Figure 13. 27 July 2016 colocation of MFLL and OCO-2 soundings. Each colored section of the C-130 track represents 60 s of MFLL data, and the OCO-2 footprints highlighted in the same color have been colocated within the same bin.

Note that we always use the same model for both the MFLL adjustments below the aircraft and the OCO-2 adjustments above, though it is possible that one model may not accurately replicate the profile in both sections of the column. However, with no validation data above the aircraft, we use the comparison below as first-order estimate of agreement above. To estimate the uncertainty associated with this choice, we can use the range of \( X_{\text{CO}_2} \) results produced from a variety of model-derived adjustments. This will be discussed in further detail in the next section.

### 3.2. Along-Track Comparison Statistics

This section details, first, the collocation and averaging of data and then the calculation of comparison statistics. Section 3.2.1 covers the statistics used for synoptic-scale comparisons (hundreds of kilometers), and section 3.2.2 covers the statistics used for local-scale comparisons (tens of kilometers).

The first step to these comparisons is proper colocation of the data. OCO-2 and MFLL data, as well as curtain data, are colocated in the following manner. First, MFLL data are averaged into 60-s bins in order to reduce noise. Based on the average speed of the C-130 over the course of 60 s, this is equivalent to a spatial resolution of between 7 and 9 km. We calculate the geographical center of each bin, and the distance to its edge (3.5 to 4.5 km) is used as the radius of a circle in which OCO-2 data are considered physically close enough for comparison. Note that this allows for some small difference in both latitude and longitude. A minimum number of soundings is required in each bin for valid comparison. Typically this number is 20 for the MFLL and 3 for OCO-2, due to the difference in the instruments’ data densities. Bins containing fewer soundings than this threshold are discarded. Figure 13 provides a visual example of this colocation method from the 27 July 2016 flight, in which the colored sections of the MFLL track are colocated with the OCO-2 (quality flag “good”) footprints which are highlighted in the same color. The full data sets are thus narrowed down to only the sufficiently dense data points lying along the underflight track.

We do not filter for time differences. For all but one flight, the typical maximum \( \Delta t \) between MFLL and OCO-2 measurements is less than 1 h. To test the effect that these time differences might have on the \( \text{CO}_2 \) field, and thus the \( X_{\text{CO}_2} \) gradient, we have sampled the GMAO curtain at overpass times of both instruments. Figure 14 shows two examples—27 July 2016, which has the largest time difference due to a very long C-130 spiral, and 15 February 2017, which is more representative of the rest of the flights. We find that the difference in the observed south-to-north gradient at the two overpass times is usually small compared with the uncertainties derived from measurement noise or model adjustments, 27 July being the exception due to the excessively long spiral on that day.
Figure 14. Example study, using the GMAO curtain, of the differences in south-to-north gradient when sampling at OCO-2 versus MFLL overpass times. (a) and (e) show the curtain CO₂ field at OCO-2 overpass times; (b) and (f) at MFLL overpass times. (c) and (g) show the time difference between the two in hours, and (d) and (h) show the XCO₂ at the two overpass times.

3.2.1. Synoptic Scales

We average the MFLL, OCO-2, and curtain data into these 60-s bins and calculate the slope of the best-fit line through these points in latitude. This slope represents the strength of the S–N gradient in units of ppm per degree latitude and will be referred to as the S–N gradient (or simply “gradient”) moving forward. The parameter we use to evaluate MFLL and OCO-2 synoptic-scale agreement is the difference between their S–N gradients.

For both OCO-2 and the MFLL, the standard deviation of the observations within each averaging bin is used as a rough estimate of the retrieval error. These values are then scaled to achieve a reduced chi-squared of 1 from the linear fit, which accounts for additional variability in the data set, due to factors such as clouds or aerosols or altitude-related effects. The resulting values are denoted here as σ₀ and σₘ for OCO-2 and the MFLL, respectively. This approach assumes any deviation from linearity to be the retrieval error, which is then propagated into the gradient error. The results of this calculation give errors for MFLL 60-second errors which range from approximately 0.15 ppm on flat, clear-sky days, to 0.5 ppm on more variable, cloudy days.

We also include an additional model-based uncertainty estimate in our calculations, based on previous discussion. In section 3.1, we show that using different models produces different values for the MFLL and OCO-2 adjustments and thus produce different XCO₂ results in both magnitude and spatial variability—this includes the value of the S–N gradient. To address this uncertainty, for each underflight, we calculate the XCO₂ using adjustments from up to four models: CAMS, CarbonTracker, GMAO, and the GMAO curtain. Each model gives a slightly different end result for both MFLL and OCO-2 adjustments, resulting in a range of meridional gradients for each data set. We calculate the range of these gradients and divide by 2 and designate this as our model-derived uncertainty for each gradient. We call this εₘₐₐₜ,O for OCO-2 and εₘₐₐₜ,M for the MFLL. We then add εₘₐₐₜ in quadrature to each instrument’s σ value to obtain an estimate of the total uncertainty on the gradient. See equations (19) and 20. Note that for the MFLL, we have an additional
retrieval-derived uncertainty $\epsilon_{\text{ret}}$, which is explained in section 2.2.1.

$$\sigma_{\text{tot,O}} = \sqrt{\sigma_{O}^2 + \epsilon_{\text{mod,O}}^2}$$  \hspace{1cm} (19)$$

$$\sigma_{\text{tot,M}} = \sqrt{\sigma_{M}^2 + \epsilon_{\text{mod,M}}^2 + \epsilon_{\text{ret}}^2}$$  \hspace{1cm} (20)$$

We then calculate the uncertainty on the OCO-2/MFLL gradient difference by adding the individual gradient uncertainties in quadrature, as in equation (21).

$$\sigma_{\text{tot,OM}} = \sqrt{\sigma_{\text{tot,O}}^2 + \sigma_{\text{tot,M}}^2}$$  \hspace{1cm} (21)$$

We will refer to $\sigma_{\text{tot,OM}}$ as the total combined error. With an estimate of their combined error, we can determine whether their gradients differ within the uncertainty of that difference.

### 3.2.2. Local Scales

Once the data are averaged into their colocated bins and the values of their gradients have been calculated, we remove the synoptic S–N trend using the best-fit line. We then calculate correlation coefficients ($r$) between the residuals, as well as the standard deviation of their differences (note that we do not compare OCO-2 to MFLL data in scenes where the OCO-2 data are filtered out due to clouds). In the case of models (CarbonTracker, GMAO, CAMS), we do not average the data into colocated bins for this calculation; instead, we interpolate the model $X_{CO_2}$ field directly to the latitudes desired. Correlations should be high if the $X_{CO_2}$ patterns behave similarly on a local point-to-point scale, on the order of tens of kilometers.
4. Results and Discussion

This study summarizes the results for eight relatively clear-sky OCO-2 underflights, two from each season. Table 2 provides a brief summary of each flight. The synoptic setups for these flights were characterized by high-pressure anticyclonic conditions, with no convective activity along the OCO-2 tracks. Flight plans were constructed such that at least one vertical profile of CO₂ was obtained along each OCO-2 track. Section 4.1 focuses on validation of synoptic-scale S–N gradients, and section 4.2 focuses on validation of local-scale features.

4.1. S–N Gradients

Figure 15 shows the XCO₂ results from the MFLL, OCO-2, models, curtain, and in situ spiral for three of eight flights. Take for example the top panel, from the first OCO-2 underflight of the summer campaign, on 27 July 2016. All data sets exhibit a decreasing trend in XCO₂ from south to north, which gives a negative S–N gradient across the approximately 400 km of flight track. The gradient of the curtain, MFLL, and OCO-2, along with the CAMS model, qualitatively look quite similar, with the CarbonTracker 2017 reanalysis product and the GMAO model slight outliers. The in situ spiral (green star) agrees well with both the MFLL and OCO-2 at that location.

Gaps in the OCO-2 data in these plots—for example, between 41° and 42° latitude on 27 July 2016—are primarily due to clouds. We often still have MFLL data in such cloudy regions, however, because the MFLL data are taken at a much finer temporal and spatial resolution and can often see through the gaps between clouds or fly below the clouds altogether. In addition, sometimes the OCO-2 measurement is contaminated by clouds which lie above the C-130 aircraft and which therefore do not interfere with the downward-looking MFLL instrument.

Figure 15. XCO₂ results from (a) 27 July 2016, (b) 15 February 2017, and (c) 8 March 2017.
However, the OCO-2 quality flag does not eliminate all soundings affected by clouds. There are $X_{\text{CO}_2}$ features in numerous flights which appear to be spurious due to scattering and path-lengthening effects from clouds. For example, near 42° north in Figure 15a, MODIS visible imagery (in Figures 16a and 16b) indicates that the few available soundings escape through narrow gaps in a popcorn cumulus cloud field. Figures 16d and 16e show a similar case from 8 March 2017. In both cases, OCO-2 is the only data set that indicates any enhancement in $X_{\text{CO}_2}$ at these locations.

In Figure 15b, there is a similar $X_{\text{CO}_2}$ enhancement on 15 February 2017. At about 39.2° N, the OCO-2 $X_{\text{CO}_2}$ exhibits a coherent peak exceeding 1 ppm in magnitude which neither the MFLL nor the GMAO curtain observe. Again, using MODIS visible imagery, we find in Figure 16c that this peak is actually just offset from the location of a high cloud. We expect that this is another case of three-dimensional cloud scattering effects which result in path lengthening. These soundings don’t get filtered out by the OCO-2 quality flags because they’re not actually colocated with the cloud; we can see that the quality flags do actually catch a few bad

### Table 3

| Date            | OCO-2 gradient | MFLL gradient | Total combined uncertainty ($2\sigma_{\text{tot,OM}}$) | Gradient difference |
|-----------------|----------------|---------------|------------------------------------------------------|---------------------|
| 27 July 2016    | -0.96 (-1.00)  | -1.11 (-1.11) | 0.99 (0.98)                                          | 0.15 (0.11)         |
| 5 August 2016   | -0.42          | -0.48         | 0.38                                                 | 0.06                |
| 15 February 2017| 0.05 (-0.06)   | -0.07 (-0.07) | 0.20 (0.19)                                          | 0.02 (0.01)         |
| 8 March 2017    | 1.36 (0.65)    | -0.23 (-0.44) | 1.00 (0.91)                                          | 1.59 (1.09)         |
| 22 October 2017 | 0.08           | 0.03          | 0.38                                                 | 0.05                |
| 27 October 2017 | -0.16          | -0.08         | 0.18                                                 | 0.08                |
| 27 April 2018   | -0.22          | -0.12         | 0.50                                                 | 0.11                |
| 29 April 2018   | -0.01          | -0.08         | 0.38                                                 | 0.07                |

**Note.** (Bolded) values have had spurious cloudy soundings subjectively removed.
Figure 17. Along-track gradients and associated uncertainties for the MFLL, OCO-2, models, and in situ curtain. Both the MFLL and OCO-2 data have our model-derived adjustments applied, as described in section 3.1. Each vertical panel represents one OCO-2 underflight. Error bars correspond to the estimated 2σ uncertainty in the OCO-2 and MFLL gradients, where the 1σ uncertainty is given by equations (19) and 20.

soundings at the physical cloud location. Figure 16f shows a similar case from the 8 March 2017 overpass, around 38.2° N.

The inclusion of cloud-contaminated data in local flux models could lead to misattribution of XCO₂ features; the recurring presence of such features has highlighted the need for either enhanced cloud screening, or perhaps some kind of three-dimensional scattering “bias correction.” Such developments are currently underway by members of the OCO-2 science team. Preliminary tests indicate that more rigorous filtering of such features only changes the global bias by approximately 0.1 ppm, so these types of features appear to have little effect on the large-scale biases.

Given the observation of these spurious XCO₂ features in multiple cases, we calculate our comparison statistics, listed in Table 3, in two ways for each cloud-contaminated flight. We first use the data set which has been filtered using the standard OCO-2 quality flag (unbolded), which includes the spurious feature, and then, as proof of concept, we subjectively remove the soundings which we visually believe to be spurious (bolded, in parentheses). In the 15 February 2017 case, in Figure 15, for example, this would cut out the XCO₂ peak and leave a flatter trend line. We do the same for the local-scale statistics in the next section.

For each date in Table 3, the first two columns show the XCO₂ S–N gradients, in units of ppm per degree latitude. The third column in Table 3 shows the total combined uncertainty of the gradient difference, calculated from equations (19), 20, and 21. The final column shows the difference between the gradients. We consider the OCO-2 and MFLL gradients to be in agreement if they differ to within 2σ of their total combined uncertainty. This 2σ value is meant to represent the 95% confidence interval for the gradient, but we recognize that within such a statistical framework, this interval is dependent upon several somewhat arbitrary assumptions, and may in fact overestimate the true uncertainties. We will thus simply refer to the 2σ value as our estimate of uncertainty going forward.
Table 4
Correlation Coefficients with OCO-2 Partial-Column \(X_{\text{CO}_2}\)

| Date                  | MFLL     | CAMS     | CT       | GMAO     | GMAO Curtain |
|-----------------------|----------|----------|----------|----------|--------------|
| 27 July 2016          | 0.10 (0.16) | 0.04 (0.17) | -0.18 (-0.08) | 0.39 (0.24) | 0.08 (0.07)  |
| 5 August 2016         | 0.43     | -0.43    | 0.41     | -0.35    | -0.11        |
| 15 February 2017      | 0.16 (0.21) | -0.27 (0.02) | -0.07 (0.20) | -0.004 (0.08) | 0.16 (-0.34) |
| 8 March 2017          | 0.22 (0.17) | 0.09 (0.32) | -0.07 (-0.06) | 0.63 (0.02) | 0.43 (-0.31) |
| 22 October 2017       | 0.12     | 0.49     | -0.19    | -0.38    | 0.59         |
| 27 October 2017       | 0.35     | 0.36     | 0.21     | -0.05    | -0.05        |
| 27 April 2018         | 0.21     | N/A      | 0.44     | N/A      | 0.30         |
| 29 April 2018         | 0.33     | N/A      | -0.07    | N/A      | 0.64         |

Note. (Bolded) values have had spurious cloudy soundings subjectively removed.

The OCO-2 and MFLL meridional gradients agree within this estimated uncertainty in seven of the eight cases, though this agreement may be overstated if our uncertainties are overestimated. An equally relevant metric is the difference in the estimates relative to the mean signal. It is also worth noting that even in cases where the gradient is small, the MFLL and OCO-2 agree on the sign. Figure 17 provides a visual overview of the findings reported in Table 3. Both show that we typically achieve an agreement within 0.1 ppm per degree latitude between OCO-2 and the MFLL. Figure 17 also shows that the gradients from three models tend to agree with our closest “truth” estimate, the GMAO curtain, within 0.25 ppm per degree latitude. Exceptions are on the two cloudiest days, 27 July 2016 and 8 March 2017, and on 27 April 2018, when the B-200 PICARRO sampled a metropolitan area during diversion for an emergency refuel, driving up the S–N gradient estimated by the curtain.

The singular case in which the OCO-2 and MFLL disagree is 8 March 2017. As previously discussed, this case includes a myriad of cloud-contaminated OCO-2 soundings, most notably at the northern end of the track as shown in Figures 15c and 16e. These high \(X_{\text{CO}_2}\) values at the end of the track drive the value of the OCO-2 gradient upward, resulting in the high value shown in the top panel of Figure 17. Manually removing these soundings from the data set results in a shorter track for MFLL/OCO-2 colocation, and changes the gradients of all data sets as a result, as shown in the bottom plot of Figure 17. Both the MFLL and OCO-2 gradients change when the northern portion of the track is eliminated, so their difference remains large. The OCO-2 gradient remains larger than those of all other data sets, and closer examination reveals that this is in part due to the fact that the C-130 flew at 6-km altitude on this date rather than 9 km. As a result, all three (four, including the curtain) models show a relatively strong S–N gradient in the unmeasured column above 6 km and impose a strong gradient on the partial column as a result. The OCO-2 full-column values before the model adjustment show an S–N gradient of 0.125 ppm per degree, which is much more similar to the curtain gradient of 0.020 ppm per degree. The reason for the steep negative slope in the MFLL data is unclear; it exists with or without the model adjustments and seems to follow the trend of the measured optical depth.

4.2. Local Scale Variability

Having established good agreement between OCO-2 and the MFLL on synoptic scales, we now turn to local scales. Table 4 shows the correlation coefficients between OCO-2 and each of the validation data sets after linear regression. As stated in the section 3.2, this is the correlation between approximately 7 to 9 km wide MFLL bins and colocated OCO-2 and curtain bins. The correlation coefficient between OCO-2 and the MFLL reaches a maximum value of 0.43 on 5 August 2016, but is otherwise consistently below 0.35. We conclude from these low correlations that for an average scene with no strong variability in the \(X_{\text{CO}_2}\) field, OCO-2 and the MFLL do not typically “see” the same small-scale features. Table 5 tells a similar story: on days with more \(X_{\text{CO}_2}\) variability, such as 27 July and 8 March, the difference between the OCO-2 and MFLL residuals is especially large, indicating that even on days with more significant small-scale features, the two do not see the same variability. On days with cloud contamination effects, this standard deviation metric does improve when we manually remove spurious soundings—most notably on 15 February, where the \(X_{\text{CO}_2}\) field really appears quite flat in both the MFLL and OCO-2 when we eliminate the high cloud shadow from the OCO-2 data.
Table 5

| Date           | MFLL   | CAMS   | CT     | GMAO   | GMAO Curtain |
|----------------|--------|--------|--------|--------|--------------|
| 27 July 2016   | 1.31 (1.19) | 1.32 (1.17) | 1.38 (1.24) | 1.16 (1.09) | 1.24 (1.14) |
| 5 August 2016  | 0.69   | 1.02   | 0.70   | 0.79   | 0.80         |
| 15 February 2017 | 0.71 (0.58) | 0.50 (0.19) | 0.49 (0.18) | 0.49 (0.20) | 0.46 (0.20) |
| 8 March 2017   | 1.49 (0.76) | 1.51 (0.75) | 1.51 (0.60) | 1.56 (0.61) | 1.41 (0.65) |
| 22 October 2017| 0.46   | 0.32   | 0.37   | 0.39   | 0.31         |
| 27 October 2017| 0.37   | 0.32   | 0.39   | 0.39   | 0.40         |
| 27 April 2018  | 0.40   | N/A    | 0.35   | N/A    | 0.35         |
| 29 April 2018  | 0.62   | N/A    | 0.65   | N/A    | 0.59         |

Note. (Bolded) values have had spurious cloudy soundings subjectively removed.

Potential explanations for this lack of agreement on local scales are numerous. For example, in section 2.2.2, we found that the MFLL retrieval is particularly sensitive to water vapor. Atmospheric water vapor varies far more than CO₂, so it is conceivable that between the times of the MFLL and OCO-2 overpasses, the water vapor field along track changes enough to make the MFLL-observed small-scale variability significantly different from OCO-2. In addition, the OCO-2 retrieval is sensitive to things like surface pressure and aerosols, among others, which can vary not only on 10-km scales but also between individual footprints. It is extremely difficult to pin down the source of every variation in retrieved XCO₂ on these scales, especially given two instruments and retrievals which are so inherently different; averaging the data in space to reduce this kind of noise is the most effective way to account for such variability and explains why we see disagreement on 10 km features but agreement on 100+ km features.

An important implication of these results concerns the appropriate scale of along-track averaging of XCO₂ retrievals from OCO-2 for use in flux estimation using tracer transport inversions (Crowell et al., 2018; 2019). Flux estimation is based on minimizing the difference between simulated and observed XCO₂ subject to Bayesian constraints, but the drastic mismatch in spatial scales between models and retrievals inevitably introduces aggregation error into flux estimates (Kaminski et al., 2001; Engelen et al., 2002). At the same time, individual retrievals of XCO₂ are subject to both random and systematic errors associated with spectroscopy and nuisance variables such as unresolved clouds and aerosol. The grid cells of tracer transport models used for flux estimation from concentration data typically span several degrees of latitude and longitude; they are thus four to five orders of magnitude larger than the footprint of an individual OCO-2 retrieval. One analytical practice has been the preprocessing of OCO-2 along-track retrievals into 10-s along-track (roughly 70 km) granules for comparison with simulations made with trial fluxes in global models (Crowell et al., 2019; Kulawik et al., 2019). Our results show that such averages are likely to substantially reduce noise introduced by spurious variations at smaller scales, while retaining information about physical gradients in XCO₂ in the real atmosphere as revealed by both the in situ curtains and the MFLL retrievals. A handful of successive 10-s granules would constitute each of our underflights and correctly represent XCO₂ gradients to nearly ±0.1 ppm per degree of latitude. Smaller along-track granules of order 1 s would be overly influenced by retrieval error, and larger granules are unnecessary. There may be some intermediate spatial averaging length between the local and synoptic scales of this study, a threshold at which the point-by-point correlations are higher and more spatial detail would be retained, rather than just one value of an along-track gradient; this type of detailed resolution study has not yet been attempted with the ACT-America data.

4.3. Local-Scale Feature Case Studies

Beyond these bulk analyses, intriguing results arise when we look closely at individual OCO-2 overpass features. While we generally find low correlations between the MFLL and OCO-2 on small scales across the full length of the underflight track, we have still observed a number of interesting features on local scales. The cloud contamination effects discussed earlier are examples of such features, but they are not the only ones. The remainder of this section will discuss particular OCO-2 XCO₂ observations which ACT-America data have helped to investigate more deeply.

One such feature comes from the 8 March 2017 case. Figure 18a shows that between 38.4° and 39.6° latitude, we observe a sinusoidal-like wave pattern in the full column OCO-2 XCO₂ field. This portion of the track...
is located in eastern Virginia, just west of the Washington, DC, area. The observed wave pattern has an approximate peak-to-trough magnitude of 1 ppm in $X_{CO_2}$ and between 10 and 20 km in space from peak to trough. Figure 18c shows the OCO-2 and MFLL partial-column $X_{CO_2}$ data at this location averaged into 0.02° latitude bins. While the correlation coefficient between the two is relatively low ($r = 0.32$) due to MFLL noise, the MFLL data do indeed display a qualitatively similar sinusoidal pattern to OCO-2.

We investigate potential sources of this variability using the Cloud Physics Lidar (CPL) aboard the C-130, which provides aerosol backscatter data below the aircraft along the flight track (McGill et al., 2002). Unfortunately, due to instrument malfunctions at this time, there is no CPL data at the exact location of the observed OCO-2 pattern, but just south of the observed pattern, the CPL reveals visible wave structures in both the boundary layer and an elevated aerosol layer between 4- and 6-km altitude, as shown in Figure 18d. PICARRO data at this location contain some sinusoidal variation at altitudes below 1,500 m, but while the approximate spatial resolution of these waves in both the elevated aerosol layer and the PICARRO data corresponds nicely to the spatial resolution of the observed OCO-2 $X_{CO_2}$ waves, we cannot conclude a statistical correlation without exact colocation. In addition, while the CPL confirms vertical motion within various atmospheric layers, it is unclear whether this type of vertical motion could produce 1 ppm changes in the partial column average. It is possible that this $X_{CO_2}$ feature could be real; however, it is equally possible that the OCO-2 retrieval has trouble dealing with the undulating height of the aerosol layer, creating a sinusoidal bias in $X_{CO_2}$. We thus have no conclusive results on the nature of these observed features. However, such an example serves as proof that comparisons on such small scales can be incredibly nuanced and difficult to parse.

Investigation into other retrieved variables from the OCO-2 retrieval produces the strongest (inverse) correlation with the retrieved ice AOD ($r = -0.408$), and there is a similar wave pattern over a portion of the 27 July 2016 track over Pennsylvania, with a similarly high correlation ($r = -0.399$ with quality filtering) to the retrieved ice AOD. However, while the magnitude of the variability on 27 July—shown in Figure 19—is similar to the 8 March case in both ppm and spatial resolution, the CPL backscatter image on this date shows nothing comparable with the structure seen in Figure 18d. In addition, though not shown here, the MFLL does not observe any similar $X_{CO_2}$ variability to OCO-2 on 27 July, which indicates that perhaps the
Figure 19. (a) Quality-filtered OCO-2 $X_{CO2}$ on 27 July 2016. (b) The same $X_{CO2}$ field without quality filters applied, for illustrative purposes. (c) The retrieved ice AOD without quality filters applied. The correlation between the filtered $X_{CO2}$ and ice AOD fields is -0.399 on this date, close to that of 8 March 2017.

The cause of the 27 July feature is not the same as on 8 March. The OCO-2 variability was initially thought to be correlated with local topography, but analysis ($r = 0.256$ between $X_{CO2}$ and surface elevation) indicates a stronger correlation to the ice AOD. We also note that this $X_{CO2}$ feature was present in version 8 of the OCO-2 data, but when a pointing correction—which correlated with topographical changes—was applied in version 9, the $X_{CO2}$ feature remained and, in fact, was strengthened slightly due to slightly looser quality flag restrictions and thus higher data throughput.

There is one additional case which, pending further study, may shed some light on OCO-2 retrieval biases. On 22 October 2017, there is a small signal in the retrieved sulfate AOD which is colocated with an approximately 0.85 ppm $X_{CO2}$ enhancement in the full column. Figure 20 illustrates this feature. Using visible imagery available online at NASA Worldview, we can actually track the development in time of what we believe to be a smoke plume from a controlled burn at a farm in rural Kansas. It is unclear whether an enhancement of this magnitude is reasonable given the size and type of the burned field (shown in Figure 20d from LandSat), or whether some of this 0.85 ppm signal may in fact be due to aerosol-related biases. Given that MFLL does not see any enhancement at this location, we suspect the latter, but further investigation is required. It seems unlikely that the MFLL would miss such a signal if it were real—perhaps the aircraft was flying too low, or the plume had dissipated or shifted trajectory by the time the MFLL passed this same location 11 min later.

While none of these cases are particularly conclusive with regards to OCO-2 or MFLL biases, they do serve as proof that comparison between the two instruments is easily complicated by a variety of atmospheric factors. Variables like aerosols or ice, as well as water vapor and other trace gases, make the local comparisons we attempt here in section 4.2 challenging.

5. Summary and Conclusions

The satellite era has brought major developments in Earth monitoring systems. OCO-2 in particular has provided a high-resolution data set, both spectrally and spatially, with which to study the global carbon cycle. The global data set has been shown to compare well with ground-based measurements of high precision and accuracy, but while $CO_2$ concentrations at singular locations show promise, little study has been done on the $CO_2$ gradients across regions ranging from local to synoptic scales. In this work, we use aircraft data from
the NASA ACT-America mission, spread over four campaigns and using both in situ and remote sensing data, to show that OCO-2 does indeed have a capability to detect trends on synoptic scales.

We have thus made the first comparisons of OCO-2 results to airborne lidar measurements, for multiple cases during all four seasons. We primarily compare to the MFLL, from whose measurements we can derive a similar quantity as OCO-2: column-averaged dry air mole fraction of CO₂ \( (X_{CO₂}) \). We make adjustments to both the MFLL and OCO-2 measurements, however, to assure their comparability: using model data, we remove the upper portion of the column from the OCO-2 measurement, and we adjust the MFLL pressure weighting function to an equivalent straight pressure weighting function. The values of these adjustments rely heavily upon the choice of model, and we show that both the OCO-2 and MFLL \( X_{CO₂} \) values can change by up to a few ppm when different models are used. To account for this, we choose the most suitable model by comparing to in situ data.

In our comparisons between the MFLL and OCO-2, we seek to evaluate their agreement in two respects: first, in the S–N gradient across the length of the flight track, a length of a few hundred kilometers; second, in the correlation of local-scale features, at a scale of tens of kilometers. We also seek to identify spurious features in the OCO-2 data which may be attributable to other atmospheric signals.
On synoptic scales, across the length of each underflight track, we find that in seven of eight clear-sky cases, the values of the MFLL and OCO-2 S-N gradients agree to within 0.1 ppm per degree latitude, well below their combined gradient uncertainty estimate. This includes cases in which the gradient itself is only on the order of a few tenths of a ppm per degree latitude—even in such cases, the MFLL and OCO-2 agree remarkably well on both the sign and the magnitude of the gradient.

In the eighth case, the OCO-2 gradient appears too strongly positive, which is due to the fact that the OCO-2 full-column XCO₂ gradient is nearly flat, while the model-derived upper column gradient is strongly negative. This, combined with the fact that the C-130 flew at 6 km rather than 9 km on this date, means that the model-derived adjustment leads to a lower-column gradient with a strong positive trend. This is perhaps an example of a case in which the upper troposphere and stratosphere of the models do not adequately represent reality.

To determine agreement on local scales, we perform a linear regression and calculate correlation coefficients between the colocated, 7- to 9-km residuals for the MFLL and OCO-2. We see little to no correlation between the two in any of the eight flights, with a maximum $r$ value of 0.43, but more consistently below 0.35. This could be due to a variety of challenges with the measurements on these scales. For example, we show that the MFLL retrieval is particularly sensitive to water vapor, and given the approximately 1-h maximum time difference between the two overpasses, small changes in the water vapor field along the track might conceivably change the MFLL small-scale variability significantly. Variables such as aerosols and surface pressure can also affect OCO-2 retrievals on these scales. We conclude overall that our ability to replicate subtle local features in the OCO-2 data set is limited on these local scales.

However, given the typically high SNR of the MFLL at these scales, we have been able to invalidate some spurious local features in the OCO-2 data set. These XCO₂ enhancements appear coherent in the OCO-2 data with magnitudes often upwards of 1 ppm, but are absent at that location in both MFLL and curtain data. MODIS visible imagery shows clouds adjacent to these features, indicating a cloud shadowing effect, in which the photon path length is increased due to scattering, increasing the absorption signal. These are features which, without careful consideration, could be interpreted as real sources in flux estimation studies. This has provided further context for the improvement of the OCO-2 filtering process, particularly relating to cloud three-dimensional effects. Such work is currently underway, and planned future work will evaluate the effects of these features on regional flux inversions.

There have also been several individual cases of interesting small-scale features in the OCO-2 data, which ACT-America data have helped to parse. In two cases, we observe wave structures in the XCO₂ field on the order of 10–20 km, which we believe to be unrelated to local topography. In the first, the XCO₂ feature is qualitatively replicated in the MFLL data and may be due to either gravity wave effects or a retrieval bias from a lofted aerosol layer observed nearby on the same track. In the second, the MFLL sees no such feature, so the underlying cause may be different. The observation of a strong XCO₂ enhancement in a smoke plume on another date may be another example of aerosol interference in OCO-2, as the MFLL does not see any similar enhancement. While we have been unable to pin down the true underlying causes of any of these examples, they all serve to show how difficult local-scale comparisons of the two data sets can be: retrievals are sensitive to the variability of many factors, in addition to time mismatches and instrument differences.

These results can inform the use of OCO-2 data in the flux inversion community. We have shown that on the single-sounding level, and up to resolutions of a few tens of kilometers, patterns observed by OCO-2 appear to be driven primarily by either noise or spurious signals from clouds and aerosols. Assimilation of data on this scale is thus not particularly useful to inversion modelers and should be accounted for in error estimates if assimilated at all. However, on synoptic scales, OCO-2 successfully reproduces meridional gradients observed by our validation data sets. There is some level of aggregation between the 10- and 400-km scales tested here, at which data users can minimize aggregation error and maximize SNRs: we estimate something between 50 and 100 km, given our results. Finding the precise value of the optimal aggregation length is not the subject of this study—and in fact that length would vary from case to case—but testing a broader range of aggregation scales may be a useful topic of further research.

This work demonstrates that even with directly colocated remote sensing data at high spatial resolution, instrument differences and retrieval sensitivities, as well as SNR limitations, make the validation of small-scale XCO₂ patterns challenging. However, though real variations at tens of kilometers remain difficult
to both observe and validate, we are generally able to reproduce patterns on synoptic scales and attribute several local OCO-2 features to cloud-related effects. We have also taken steps toward a deeper understanding of OCO-2 retrieval sensitivity, both to real changes in the CO₂ field and to the presence of atmospheric aerosols. Further development of the OCO-2 data product, MFLL retrieval algorithm, and in situ curtain construction methods may yet shed light on smaller-scale patterns: the state of greenhouse gas monitoring science continues to advance, and both OCO-2 and the ACT-America mission are strong evidence of its progress.

References

ACT-America Data Archive (2019). Accessed: January 21, 2020 https://www-air.larc.nasa.gov/missions/ACT-America/

Agusti-Panareda, A., Massart, S., Chevallier, F., Boussetta, S., Balsamo, G., Beljaars, A., et al. (2014). Forecasting global atmospheric CO₂. Atmospheric Chemistry and Physics, 14(21), 11,959–11,983. https://doi.org/10.5194/acp-14-11959-2014

Boesch, H., Brown, L., Castano, R., Christi, M., Connor, B., Crisp, D., et al. (2019). Orbiting Carbon Observers (OCO) - 2 Level 2 Full Physics Algorithm Theoretical Basis Document: ASA Jet Propulsion Laboratory and California Institute of Technology. Version 2.0, Rev 3.

Browell, E. V., Dobbs, M. E., Dobler, J., Kooi, S., Choi, Y., Harrison, F. W., et al. (2008). Airborne demonstration of 1.57-micron laser absorption spectrum for atmospheric CO₂ measurements. Presented at the 24th International Laser Radar Conference in Boulder, Colorado.

Browell, E. V., Dobler, J., Kooi, S. A., Choi, Y., Harrison, F. W., Moore, B., & Zaccheo, T. S. (2010). Airborne validation of laser remote measurements of atmospheric carbon dioxide. In EGU general assembly conference abstracts (Vol. 12, pp. 3721). Vienna, Austria.

Browell, E. V., Wilkerson, T. D., & McLain, T. J. (1979). Water vapor differential absorption lidar development and evaluation. Applied Optics, 18, 3474–3483.

Campbell, J. F., Lin, B., Oblad, M. D., Dobler, J., Erziehen, W., McGregor, D., et al. (2020). Field evaluation of column CO₂ retrievals from intensity-modulated continuous-wave differential absorption lidar measurements during ACT-America. arXiv e-prints, arXiv:2003.10971.

Chan, D., Yuen, C. W., Higuchi, K., Shashkov, A., Liu, J., Chen, J., & Worthy, D. (2004). On the CO₂ exchange between the atmosphere and the biosphere: The role of synoptic and mesoscale processes. Tellus Series B Chemical and Physical Meteorology, 56, 194. https://doi.org/10.1034/j.1600-0889.2004.00040.x

Chen, Hans, M., Zhang, L. N., Zhang, F., Davis, K. J., Lauvaux, T., Pal, S., et al. (2019). Evaluation of regional CO₂ mole fractions in the ECFMF 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, 112, D02S07. https://doi.org/10.1029/2007JD009492

Chevallier, F., Engelen, R. J., Carouge, C., Conway, T. J., Peilin, P., Pickett-Heaps, C., et al. (2009). AIRS-based versus flask-based estimation of carbon surface fluxes. Journal of Geophysical Research, 114, D20S03. https://doi.org/10.1029/2009JD011031

Crowell, S., Baker, D., Schuhi, A., Basu, S., Jacobson, A. R., Chevallier, F., et al. (2019). The 2015-2016 carbon cycle as seen from OCO-2 and the global in situ network. Atmospheric Chemistry and Physics, 19(15), 9797–9831. https://doi.org/10.5194/acp-19-9797-2019

Crowell, S. M. R., Randolph Kawa, S., Browell, E. V., Hammerling, D. M., Moore, B., Schaefer, K., & Doney, S. C. (2018). On the ability of space-based passive and active remote sensing observations of CO₂ to detect flux perturbations to the carbon cycle. Journal of Geophysical Research: Atmospheres, 123, 1460–1477. https://doi.org/10.1002/2017JD027836

Darmenov, A. S., & da Silva, A. M. (2015). The Quick Fire Emissions Dataset (QFED): Documentation of Versions 2.1, 2.2 and 2.4. Volume 38: Technical Report Series on Global Modeling and Data Assimilation (Tech. Rep.): NASA Goddard Space Flight Center. Davis, K. J., Baier, B., Baker, D., Barkley, Z., Bell, E., Bowman, K. W., et al. (2017). Progress toward improving regional atmospheric CO₂ measurements using airborne measurements: Results from ACT-America, ACO Full Meeting Abstracts (pp. AS1M-03). New Orleans, Louisiana.

Davis, K. J., Pal, S., Lauvaux, T., Browell, E. V., DiGangi, J. P., Gaudet, B. J., et al. (2018). Synoptic and meso scale variability in greenhouse gases across fronts in four seasons. In AGU Fall Meeting Abstracts (Vol. 81, pp. A33N-3136). Washington, D.C.

Engelen, R. J., Denning, A. S., & Gurney, K. R. (2002). On error estimation in atmospheric CO₂ inversions. Geophysical Research Letters, 29, 2031. https://doi.org/10.1029/2001GL014565

Engelen, R. J., Stavrakou, T., De Arellano, M., Gobbi, G. P., Hensen, B., et al. (2016). The MACC-II data assimilation system: its development and initial evaluation. Geoscientific Model Development, 9, 191–213. https://doi.org/10.5194/gmd-9-191-2016

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. K., Viatte, C., et al. (2017). The Orbiting Carbon Observatory-2: First 18 months of science data products. Atmospheric Measurement Techniques, 10, 549–563. https://doi.org/10.5194/amt-10-549-2017

Eldering, A., Wengenna, P. O., Crisp, D., Schimel, D. S., Gunson, M. K., Chatterjee, A., et al. (2017). The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. Science, 358(6360), eaam5747. https://doi.org/10.1126/science.aam5747

Engelen, R. J., Denning, A. S., & Gurney, K. R. (2002). On error estimation in atmospheric CO₂ inversions. Journal of Geophysical Research, 107, 4635. https://doi.org/10.1029/2002JD002195

Gerbig, C., Kort, E. A., Wofsy, S., Daube, B., Brink, J., Wiggins, S., et al. (2007). The North American regional trace gas network: origins, development and early results. Atmospheric Measurement Techniques, 1, 607–636. https://doi.org/10.5194/amt-1-607-2008

Gordon, I. E., Rothman, L. S., Hill, C., Kochanov, R. V., Bernath, P. F., et al. (2007). The HITRAN2004 molecular spectroscopic database. Journal of Quantitative Spectroscopy and Radiative Transfer, 96, 3–69. https://doi.org/10.1016/j.jqsrt.2007.07.008

Heymann, J., Reuter, M., Hilker, M., Buchwitz, M., Schneising, O., Bovensmann, H., et al. (2015). Consistent satellite XCO₂ retrievals from SCIAMACHY and using the BESD algorithm. Atmospheric Measurement Techniques, 8, 2961–2980. https://doi.org/10.5194/amt-8-2961-2015

Hurwitz, M. D., Ricciuto, D. M., Bakwin, P. S., Davis, K. J., Wang, W., Yi, C., & Butler, M. P. (2004). Transport of carbon dioxide in the presence of storm systems over a Northern Wisconsin Forest. Journal of Atmospheric Sciences, 61, 607–618. https://doi.org/10.1175/1520-0469(2004)061<0607:TCDOT3>2.0.CO;2

Kaminski, T., Rayner, P. J., Heimann, M., & Enting, I. G. (2001). On aggregation errors in atmospheric transport inversion. Journal of Geophysical Research, 106, 4703–4715. https://doi.org/10.1029/2000JD005851

BELL ET AL. 29 of 31

Journal of Geophysical Research: Atmospheres 10.1029/2019JD031400
Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., et al. (2017). Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO₂ measurements with TCCON. *Atmospheric Measurement Techniques, 10*(6), 2209–2238. https://doi.org/10.5194/amt-10-2209-2017