Using Ground- and Space-based Observations to Quantify Within-day to Synoptic-scale Variance Budgets of Total Column Averaged Carbon Dioxide (XCO_2) Measurements

by

Anthony D. Torres

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Atmospheric, Oceanic, and Space Sciences)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Gretchen Keppel-Aleks, Chair
Professor Scott C. Doney, University of Virginia
Associate Professor J. Timothy Dvonch
Associate Professor Eric A. Kort

Anthony D. Torres
adtorres@umich.edu
ORCID iD: 0000-0002-0531-5883

DEDICATION

I dedicate this dissertation to everyone who is a first-generation student, of underrepresented and marginalized groups, disabled, and/or has faced serious hardships while seeking a degree. Thank you for all you have done to pave the pathway for me to reach this point in my life.

For those still in the midst of completing their degree, you are also taking part in paving the path for others. While the path still often seems impossibly difficult, know that no matter whatever you achieve, the difference your presence, experiences, and hard work contributes will enormously change the world for the better. Countless others will benefit from what you do. You've got this!

ACKNOWLEDGMENTS

I would like to first acknowledge my amazing support network. This is includes incredible family, friends, colleagues, University of Michigan faculty and staff, and most importantly – the love of my life – Jacqueline. Jackie, you have given me so much light in the time I've known you and your love and support has helped me through extremely difficult times.

I would also be remiss if I did not also dedicate this dissertation to my 15 year old Shih Tzu, Bella, who has provided a constant supply of love and joy throughout this entire process.

I also am thankful for the memories and amazing times I had with my cohort at CLaSP and my office mates in the "MAD Lab." I will miss all of the jokes, trips, and our regular hangouts. deeply appreciate the nonstop encouragement through my number of challenges, which has made the completion of this work possible. I seriously could not have done this without you.

I would have not reached the completion of my PhD without the phenomenal support from my advisor and Disseration Committee Chair, Gretchen Keppel-Aleks. You have gone above and beyond in accommodating me through my enormous challenges and your belief and faith in me carried me through the completion of my dissertation. I have learned so much from you and am a better scientist and person for the experience.

I would also like to acknowledge the support I had leading up to my Ph.D. As a young undergraduate student, I distinctly remember struggling through my Clouds and Precipitation Processes class. To that point, the thought of attending graduate school never crossed my mind. My then Professor, Derek Posselt, asked me why I hadn't considered graduate school – and told me why I would be a good candidate for a Ph.D. That encouragement marked a turning point in my life and marked the beginning of my love for research.

I would also like to thank all of the support I have gotten through the years from the SOARS (Significant Opportunities at Atmospheric Research) family at the National Center for Atmospheric Research. I learned so much from this program and have met so many incredible people and lifelong friends.

Last, but certainly not least, this research has been made possible by the financial support from NASA Award NNX15AH13G, the NASA Earth and Space Science Fellowship Award 80NSSC17K0382 as well as the University of Michigan's Rackham Merit Fellowship. I am extremely grateful and honored to have the privilege have the resources to do this research.

Table of Contents

List of Figures

Figure

List of Tables

Table

List of Appendicies

Abstract

Carbon dioxide (CO_2) is the single most important greenhouse gas contributor to increases in global temperatures. Since 1850, humans have emitted approximately 700 Gt of carbon. About half of that CO_2 has remained in the atmosphere with the rest being stored in the land and ocean. While global human emissions of CO_2 from fossil fuels are well-understood, regional-scale exchanges between the land and atmosphere and ocean and atmosphere remain highly uncertain and are a very active area of study. Inferences of carbon fluxes can be made by modeling carbon cycle processes, statistical upscaling of local fluxes, human emissions, or accurately partitioning drivers of spatiotemporal changes in atmospheric CO_2 observations.

The launch of space-based CO_2 monitoring systems in the 21st century has provided unprecedented global coverage of CO_2 observations. Instruments, such as NASA's Orbiting Carbon Observatory-2 (OCO-2), which launched in 2014, provide approximately 1 million total column-averaged CO_2 (XCO_2) soundings per day at a horizontal spatial resolution of 1.29×2.25 km^2, with a repeat cycle of every 16 days. OCO-3, which was installed on the International Space Station in 2019 provides the first space-based XCO_2 observations at different times of day, potentially allowing for the direct observation of the diurnal cycle of XCO_2 from space.

This thesis leverages the abundance of XCO_2 data available to quantify a subseasonal variance budget of XCO_2. At these timescales, we expect XCO_2 to vary as a function of local fluxes and advection via atmospheric transport. Here, we first quantify climatological variations in XCO_2 driven by fine mesoscale atmospheric transport, larger synoptic scale atmospheric transport, and local diurnal fluxes using ground-based XCO_2 observations from the Total Carbon Column Observing Network (TCCON). We then use the one dimensional tracer conservation equation to provide a statistical framework for estimating along-track mesoscale variations of XCO_2 detected by OCO-2

data. We also provide a framework to detect the diurnal cycle of XCO_2 from space-based missions, such as OCO-3.

This research contributes an early attempt at quantifying "XCO_2 meteorology," analogous to our understanding of subseasonal variations of atmospheric water vapor. These studies provide alternative methods to quantify transport-driven variability and random and spatially-coherent errors in XCO_2 retrievals, which can be applied to inverse models or evaluating the performance retrieval algorithms. Most importantly, these studies move us closer to intuitively quantifying the local-to-regional scale fluxes globally using XCO_2 observations.

Chapter 1

Introduction

1.1 Introduction

1.1.1 CO_2 contributions to climate change

The greenhouse effect plays a critical role in governing global surface temperatures and sustaining contemporary life on this planet. The greenhouse effect refers to the process in which trace gasses in the atmosphere – such as water vapor (H_2O), carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), nitrogen oxides (NO_x), among others – absorb outgoing longwave thermal radiation from the earth's surface and emits that radiation back to the surface. Without these greenhouse gasses, the earth's climate would be much colder with an average surface temperature of around -20ºC (0ºF), in contrast to the mean global surface temperature of 13.9ºC (57.0ºF) observed over the 20th century (NOAA National Centers for Environmental Information, 2023). At these temperatures, water can exist in liquid form, yielding a large system of lakes, aquifers, rivers, and oceans – and clouds that provide rainfall over land, transporting essential nutrients that support life.

Industrial emissions since around 1750 have continuously released greenhouse gasses into the atmosphere, increasing their global concentrations and the strength of the greenhouse effect. In this time period, atmospheric CO_2 concentrations have increased from approximately 275 ppm in 1750 to over 420 ppm today (Keeling et al, 2001). The United Nations Intergovernmental Panel on Climate Change's 2021 Physical Science Basis report (IPCC, 2021) includes a time series of both observed and simulated changes in global surface temperatures from 1850 to 2019 (Figure 1.1).

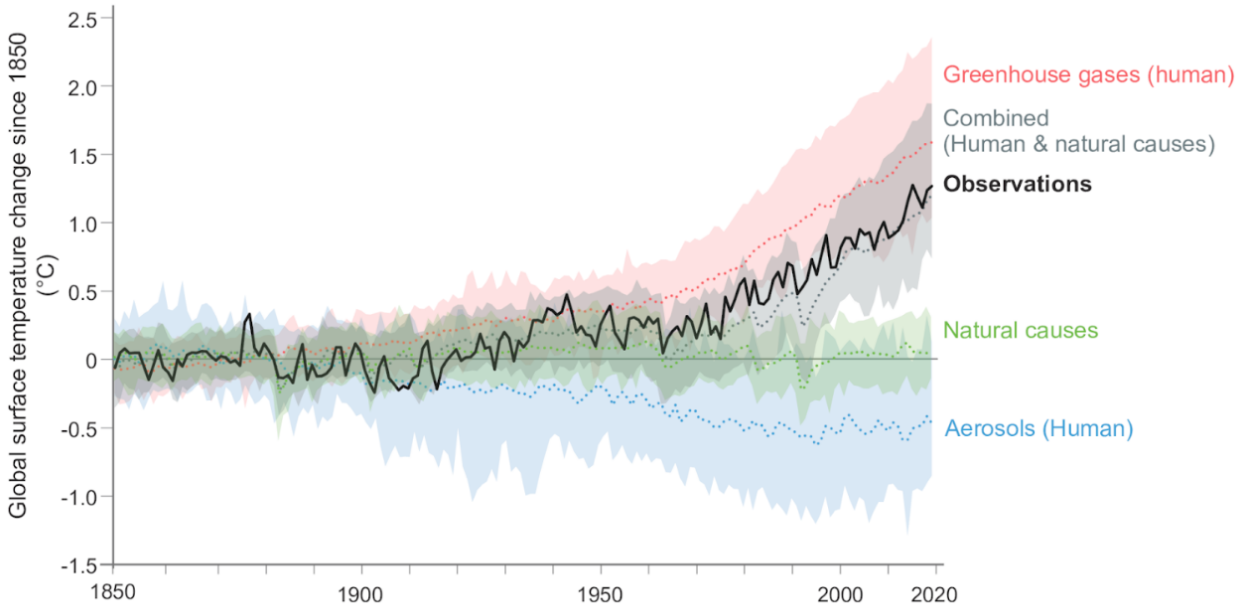

Figure 1.1: (From the IPCC, 2021) Observed global surface temperature changes from 1850-2019 (solid black line) compared to simulated temperatures based on human-induced forcing, such as greenhouse gasses (red shading) and aerosols (blue shading), natural causes (green shading), and the combined effects of human and natural causes (gray shading).

The simulated changes in temperature can be broken down and attributed to warming greenhouse gasses, net cooling from the release of aerosols, and natural causes. Only by including the net effects of human activities (warming from greenhouse gas emissions and cooling from aerosols) are simulations able to replicate observed warming in temperatures. In the absence of human activities, natural causes do not provide any appreciable long-term warming trends.

When partitioning warming among individual greenhouse gas species and aerosols, CO_2 is the largest contributor to observed temperature increases (Figure 1.2; IPCC, 2021). CO_2 is responsible for about 2 W m-2 approximately 1ºC (within a range of 0.5 to 1.5ºC) of warming from 1750 to 2019. The second largest contributor to warming, CH_4, only adds approximately 0.5ºC of warming in the same time period.

Annual CO_2 growth rates are overwhelmingly driven by fossil fuel emissions, with a secondary source from land use change (Figure 1.3; Friedlingstein et al, 2022). These anthropogenic sources of CO_2 are then distributed between the land, ocean, and atmospheric reservoirs. The key point

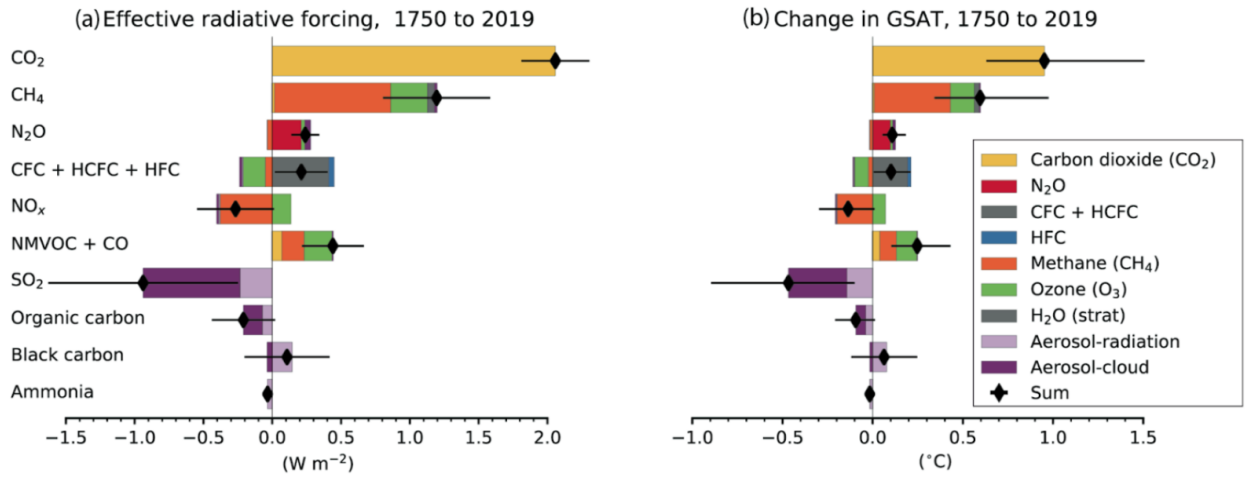

Figure 1.2: From IPCC (2021): (a) Effective radiative forcing and (b) changes in Global mean Surface Atmospheric Temperatures (GSAT) from 1750 to 2019 attributed to individual greenhouse gas species and aerosols.

here is that, despite interannual variability, the land and oceans each take up around 25 percent of emissions each year. The rest – approximately half – remains in the atmosphere contributing to global warming.

The longest continuous observations of CO_2 have been made at the Mauna Loa Observatory in Hawaii. Figure 1.4, widely known as the "Keeling curve," shows that atmospheric CO_2 concentrations have steadily increased from roughly 315 ppm in 1958 to approximately 420 ppm in 2023 (Keeling and Keeling, 2017). Prior to the industrial era in 1750, CO_2 concentrations were around 280 ppm (USGCRP, 2018). The last time CO_2 concentrations exceeded 400 ppm was approximately 10 to 20 million years ago; however, the average annual CO_2 growth rate of roughly 2 ppm per year is unprecedented in global records (IPCC, 2021). This underscores both the role human activity has contributed to increases, as well as CO_2's role in governing global surface temperatures.

1.1.2 The global carbon budget

Predicting how much CO_2 ends up in the atmosphere depends on both constraining carbon fluxes and an accurate representation of global carbon stocks. The global carbon budget contextualizes

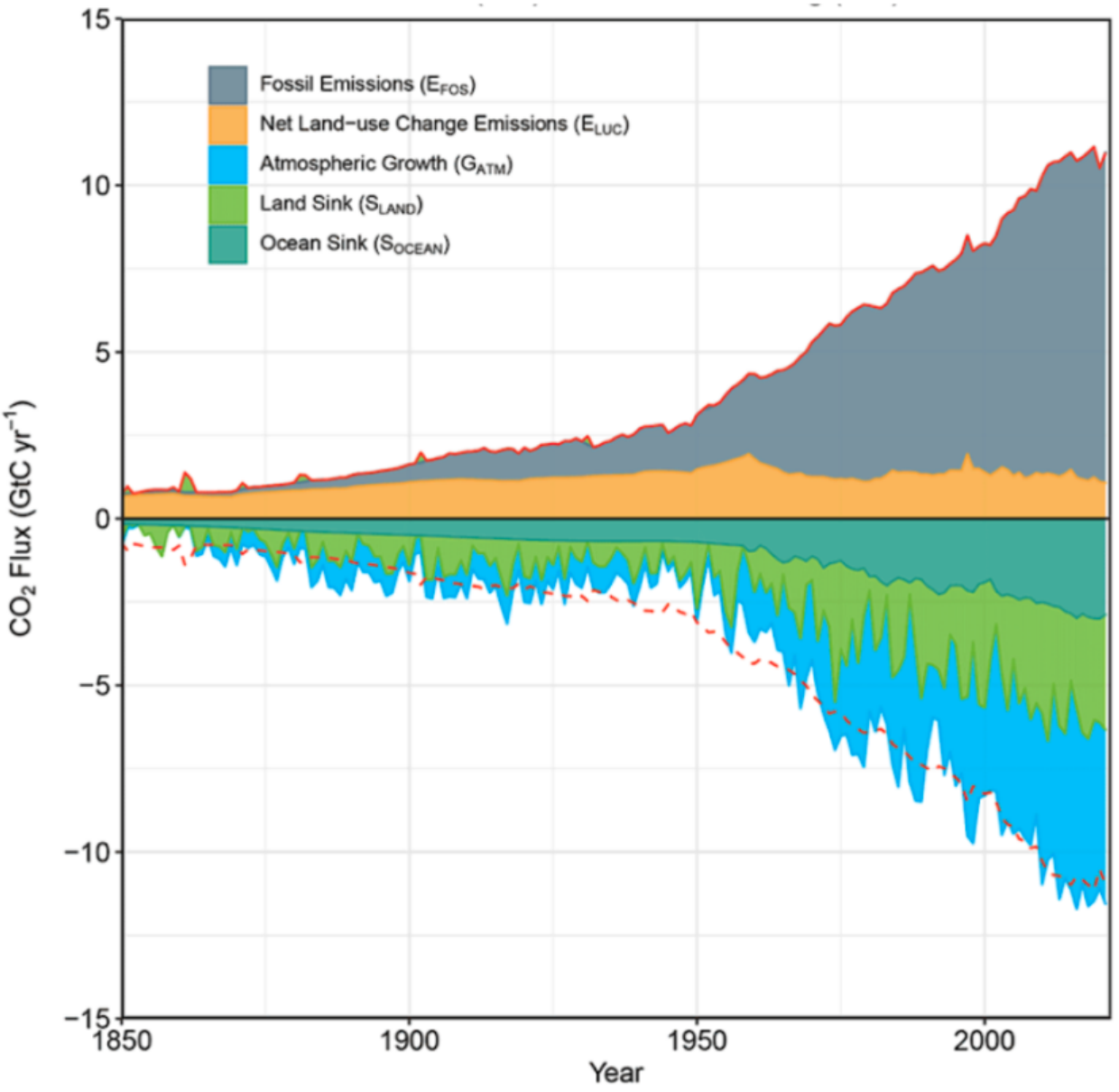

Figure 1.3: (Friedlingstein et al, 2022) Annual carbon emissions and their partitioning between the atmosphere, land, and ocean from 1850 to 2021.

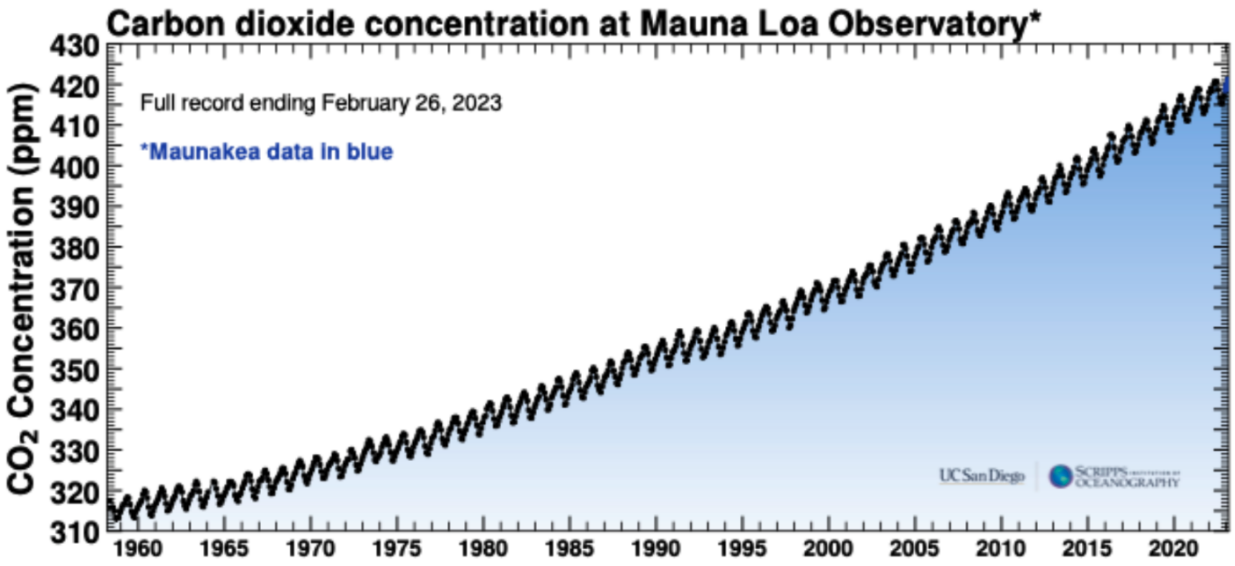

Figure 1.4: Carbon dioxide concentrations observed at the Mauna Loa Observatory from 1958 to 2023. The documentation for these measurements are described in Keeling and Keeling (2017).

global carbon stocks in the land, oceans, and atmosphere, as well as the natural exchanges between the three reservoirs and emissions of fossil fuels (Friedlingstein et al, 2022; Figure 1.5). The overwhelming majority of carbon on earth is stored in the lithosphere, the majority of carbon available for fast cycling in the modern carbon cycle is stored in the ocean and is on the order of 40,000 GtC. By comparison, the earth's atmosphere only stores a small fraction of what is in the land and oceans: 875 GtC. At annual timescales, approximately 25 percent of anthropogenic emissions have been historically stored in the oceans, with another 25 percent stored in the biosphere, with roughly just 50 percent remaining in the atmosphere (Figure 1.3; Friedlingstein et al, 2022). The primary source of anthropogenic CO_2 emissions to date, fossil fuels such as gas reserves (115 GtC), oil reserves (230 GtC), and coal reserves (230 GtC) store a cumulative 905 GtC (Figure 1.3; Friedlingstein et al, 2022). Still, atmospheric carbon sources, such as vegetation (450 GtC), soils (1700 GtC), and permafrost (1400 GtC) store 3550 GtC altogether, more than 3 times the cumulative carbon available from fossil fuels worldwide (Figure 1.3; Friedlingstein et al, 2022).

In this thesis, we focus on contemporary carbon exchanges that occur at much smaller time scales, such as within a day, as well as the processes that contribute to inferring these diurnal

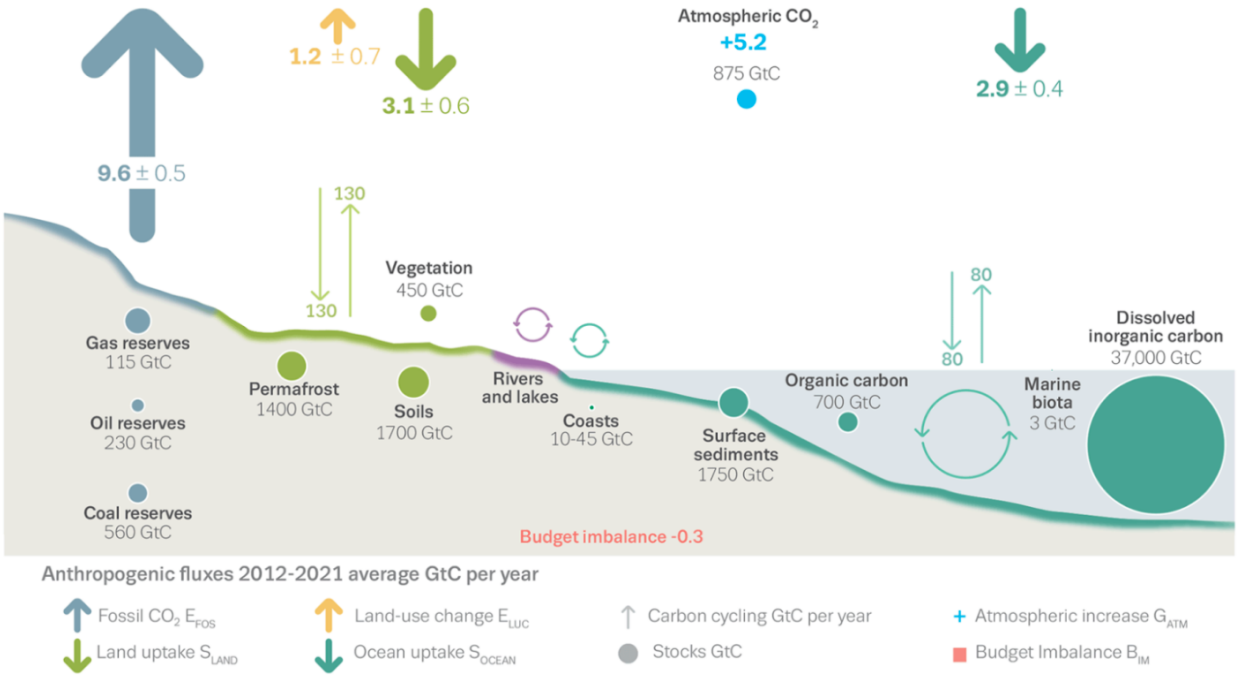

Figure 1.5: From Friedlingstein et al (2022): The Global Carbon Budget in 2021. Land-ocean-atmosphere fluxes are denoted by arrows. Upward (downward) pointing arrows represent the average carbon fluxes into the atmosphere (into the ocean or land reservoirs) each year from 2012 to 2021. Total carbon stocks by reservoir are denoted by circles. Note that 1 GtC is equal to 1 Pg of Carbon (PgC).

exchanges. These fluxes, and their response to environmental forcing, provide key insights to resolving uncertainties in global carbon uptake projections in earth system models (ESMs) (Bonan and Doney, 2018). Friedlingstein et al (2022) reports yearly carbon exchanges indicated by the arrows shown in Figure 1.5. Here, the net increase of carbon in the atmosphere is balanced by anthropogenic emissions and the biosphere and ocean sinks is approximately 5.2 GtC per year. Note that 1 GtC per year added to the atmosphere translates to an increase of CO_2 concentrations of roughly 0.5 ppm per year. Thus, the increase in atmospheric CO_2 translates to an approximate increase of around 2.6 ppm each year in this time period. However, even with these observations, there is still a budget imbalance of 0.3 GtC per year. This imbalance at annual timescales is more profoundly highlighted by the mismatch between sinks and sources in Figure 1.3, suggesting that the carbon budget remains to be closed.

1.1.3 Observing CO_2

1.1.3.1 In situ CO_2 observations

Sporadic samples of atmospheric CO_2 observations have been cited in academic literature dating back to Krogh (1904) through Glueckauf (1951). These first observations showed increases in CO_2 concentrations that could be attributed to the combustion of fossil fuels as early as Callendar (1938), Revelle and Suess (1957), Bolin and Erickson (1959). As described above, the first continuous measurements of in situ CO_2 began in 1958 at the Mauna Loa Observatory in Hawaii by Charles David Keeling. These measurements revealed the seasonal cycle of CO_2, in which enhanced drawdown was observed in the northern hemisphere during the summer, and a net release in the winter (Keeling, 1960). Keeling (1960) also included continuous measurements in Antarctica and California, in addition to Hawaii. These observations also showed for the first time in the magnitude of the seasonal cycle of CO_2 as a function of latitude, and the phasing of the seasonal cycle by hemisphere. Keeling's studies motivated broader observational networks to connect spatiotemporal changes of atmospheric CO_2 to our understanding of the contemporary global carbon cycle.

Most early studies of CO_2 primarily focused on exchanges between the ocean and atmosphere, as well anthropogenic emissions (Dingle, 1954; Revelle and Suess, 1957; Bolin, 1960; Bolin and Keeling, 1963; Broecker et al, 1979). Land fluxes to this point were considered negligible (Broeker at al, 1979). It was with a denser network of observations over land and oceans with the advent of NOAA's Geophysical Monitoring for Climatic Change (GMCC) program and observed partial pressures of CO_2 measurements over oceans that Tans et al (1990) revealed that oceanic CO_2 fluxes were too small to be the only sink of anthropogenic CO_2. This profound finding revealed how large of an imprint carbon fluxes within terrestrial ecosystems on CO_2 observations throughout the year and punctuated how we cannot ignore terrestrial fluxes if we wish to understand the contemporary global carbon cycle. The increase in spatial and temporal atmospheric CO_2 observations revealed the necessity for increasingly sophisticated methods for inferring global carbon fluxes, such as inverse models, to account for other drivers of variability of CO_2, such as atmospheric transport (Enting, 2002).

The network of in situ CO_2 observations has continued to expand into the present day (Crisp et al, 2022; Figure 1.6). In situ observations are routinely made over land, by sea routes and by aircrafts. However, the greatest density of CO_2 observations are still made in North America and Europe (Crisp et al, 2022; Figure 1.6). The gaps of regular in situ measurements over the tropics and boreal regions has contributed to the significant uncertainty in regional land fluxes (Gurney et al, 2002; Basu et al, 2018).

1.1.3.2 Space-based CO_2 observations

Satellites provide an unprecedented opportunity to fill in observational gaps with global, regular observations of CO_2. Several greenhouse gas satellites detect CO_2 throughout the entire atmospheric column, which is more commonly referred to as total column-averaged CO_2 (henceforth denoted as XCO_2) (Pan et al, 2022). XCO_2 can be interpreted as the vertical integration of CO_2 throughout the depth of the atmospheric column. Many of the early instruments retrieved CO_2 in the thermal infrared (TIR) spectra, which is most sensitive to CO_2 in the upper troposphere and

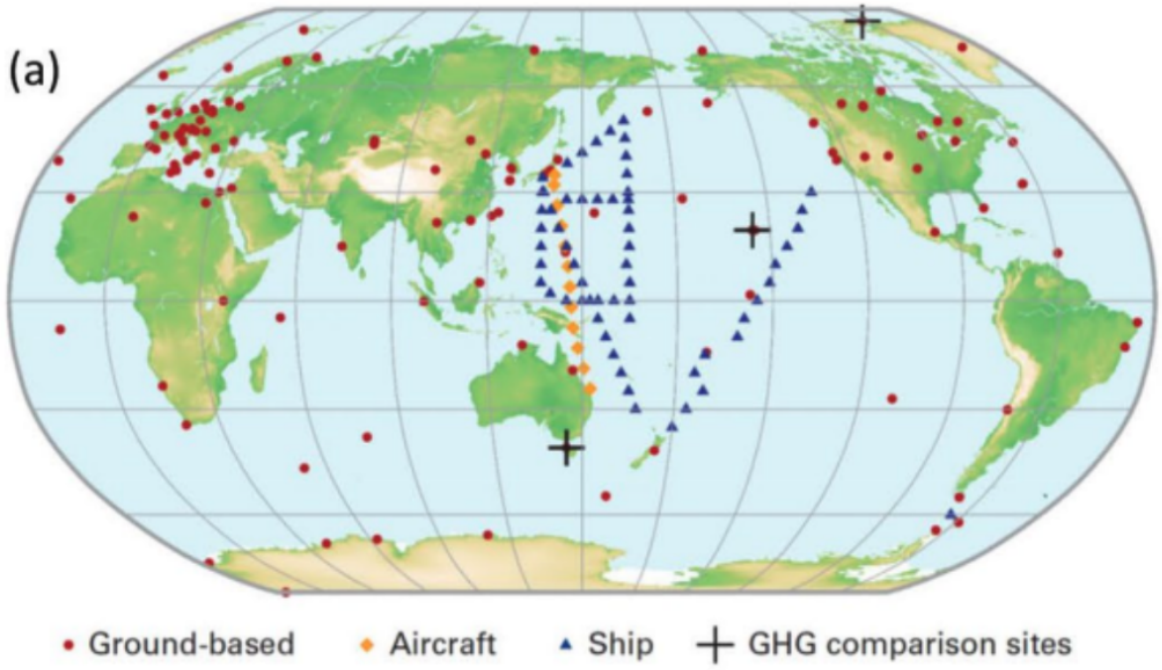

Figure 1.6: From Crisp et al, (2022) Global network of in situ CO_2 observations in 2022.

lower stratosphere and are not particularly sensitive to surface CO_2 (Pan et al, 2022). Unlike the TIR-observing instruments, observations in the near infrared spectra can detect XCO_2 through the entire column, including sensitivity to surface CO_2 (Chevalier et al, 2014). Characterizing subseasonal variations of XCO_2 that occur on timescales ranging from within-day to as much as 21 days is the focus of this dissertation. With this information, we expect that we can more directly infer the magnitude of local carbon fluxes globally using the increasing wealth and accuracy of available XCO_2 data.

The earliest satellites equipped with the capacity to monitor CO_2 were launched in 2002 (Figure 1.7; Pan et al, 2022). Several early satellites, such as those onboard NASA's Aqua, Aura, and European Space Agency's (ESA) Metop-A (Meteorological operational satellite - A) satellites, used CO_2 detection as ancillary products. From the order of satellites listed in the previous sentence, the CO_2-monitoring instruments included AIRS (Atmospheric Infrared Sounder), TES (Tropospheric Emission Spectrometer), and IASI (Infrared Atmospheric Sounding Interferometer), respectively (Pan et al, 2022). CO_2 retrievals from these instruments were primarily sensitive to CO_2 in the

mid-troposphere and did not directly detect the imprint of local surface fluxes (Figure 1.7; Pan et al, 2022). While the SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Cartography) instrument aboard the Envisat (Environmental Satellite) satellite could provide early retrievals of XCO_2, these retrievals are only accurate to 14 ppm, which is quite impractical for inferences of local fluxes (Figure 1.7; Pan et al, 2022).

The first satellites specifically designed to detect XCO_2 were Japanese Aerospace Exploration Agency's (JAXA) Greenhouse Gas Observing Satellite (GOSAT) (Yokota et al, 2009) and NASA's Orbiting Carbon Observatory (OCO; Crisp et al, 2004). These instruments leverage the NIR portion of the solar spectrum and therefore provide sensitivity close to the surface, where fluxes occur. GOSAT launched successfully in 2009 and began retrieving high-density XCO_2 and XCH_4 soundings at horizontal resolutions of 10 x 10 km^2 with a 3-day repeat cycle beginning in 2009 (Yokota et al, 2009). Unfortunately, the first OCO satellite failed to reach orbit due to a catastrophic malfunction in the launch vehicle during the same year(Eldering et al, 2017).

Most of our analysis is focused on the OCO-2 and OCO-3 datasets. OCO-2 launched in 2014 as a replacement to the original OCO satellite (Eldering et al, 2017). The OCO-2 satellite is equipped with three grating spectrometers that measure reflected sunlight in three narrow spectral bands centered around 0.76 μm for strong oxygen (O_2) absorption, 1.6 μm for weak CO_2 absorption, and 2.1 μm for strong CO_2 absorption with the highest spatial resolution to date of 1.29 × 2.25 km^2 (Crisp et al, 2008). This results in an average of approximately 1 million OCO-2 soundings per day that are accurate to approximately 1 ppm (Eldering et al, 2017). All of these instruments sit aboard satellites that have sun-synchronous polar-orbiting patterns that would allow for continuous observations at the same time of day globally. Other CO_2 monitoring satellites that have launched include China's TANSAT, which launched in 2016, and JAXA's GOSAT-2, which launched in 2018 (Figure 1.7; Pan et al, 2022).

The OCO-3 instrument is the first XCO_2-monitoring instrument that collects soundings through variable times of day from a space-based platform (Eldering et al, 2019). The OCO-3 payload, which has nearly identical design specifications as the OCO-2 instrument, was installed in the

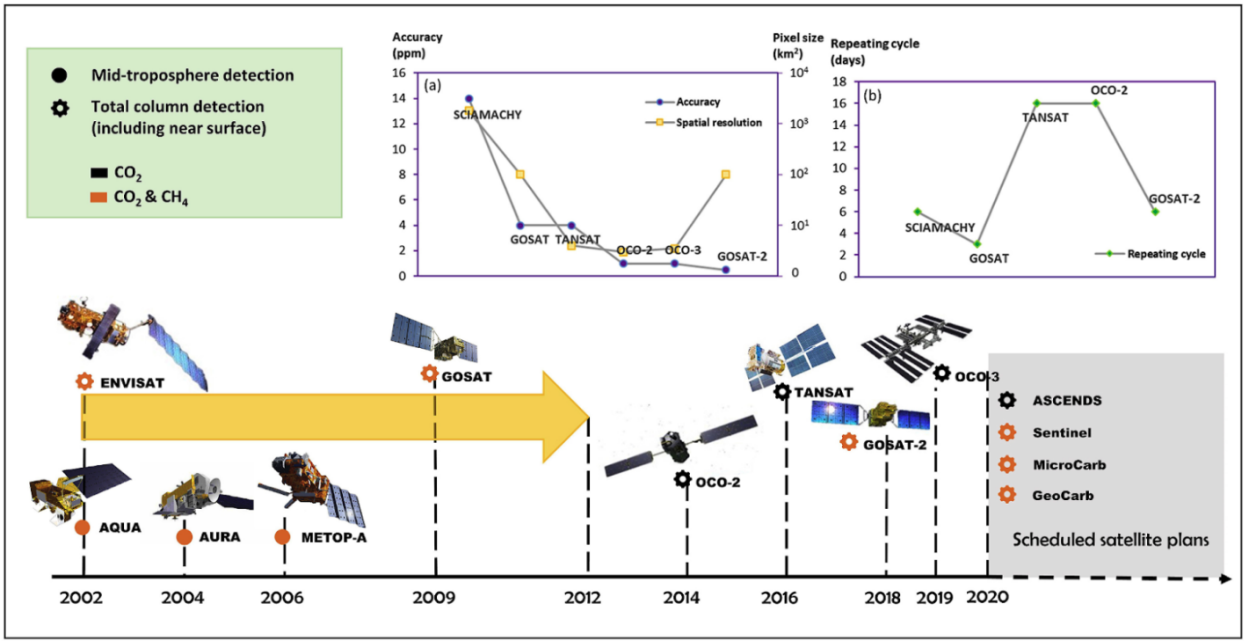

Figure 1.7: From Pan et al. (2020): "The launching timeline of satellites for CO_2 detection including scheduled satellite plans. (a)&(b) The accuracies, spatial resolutions and repeating cycles of current and previous CO_2 satellites (XCO_2 total column measurement). Among current satellites, OCO-2, OCO-3 and GOSAT-2 can provide relatively high detection accuracy. OCO-2, OCO-3 and TANSAT can detect kilometer level emission. GOSAT and GOSAT-2 can provide relatively frequent detection." Note: Future ASCENDS, MicroCarb, and GeoCarb missions have been canceled since the publishing of Pan et al. (2020).

International Space Station (ISS) in 2019 (Eldering et al, 2019). Since the ISS orbits in a precessing orbital pattern, overpasses occur at various times of day with no consistent repeat cycle (Eldering et al, 2019).

The studies in this dissertation only focus on OCO-2 and OCO-3 space-based and TCCON ground-based data. Additional details regarding specifications of these observing systems are described in greater depth in the subsequent chapters.

1.1.3.3 Total Carbon Column Observing Network (TCCON)

While space-based observations provide a high density of XCO_2 observations, these observing systems combined still provide sparse temporal observations. These temporal gaps are partially filled by ground-based networks that collect XCO_2. One such network that is heavily used in this

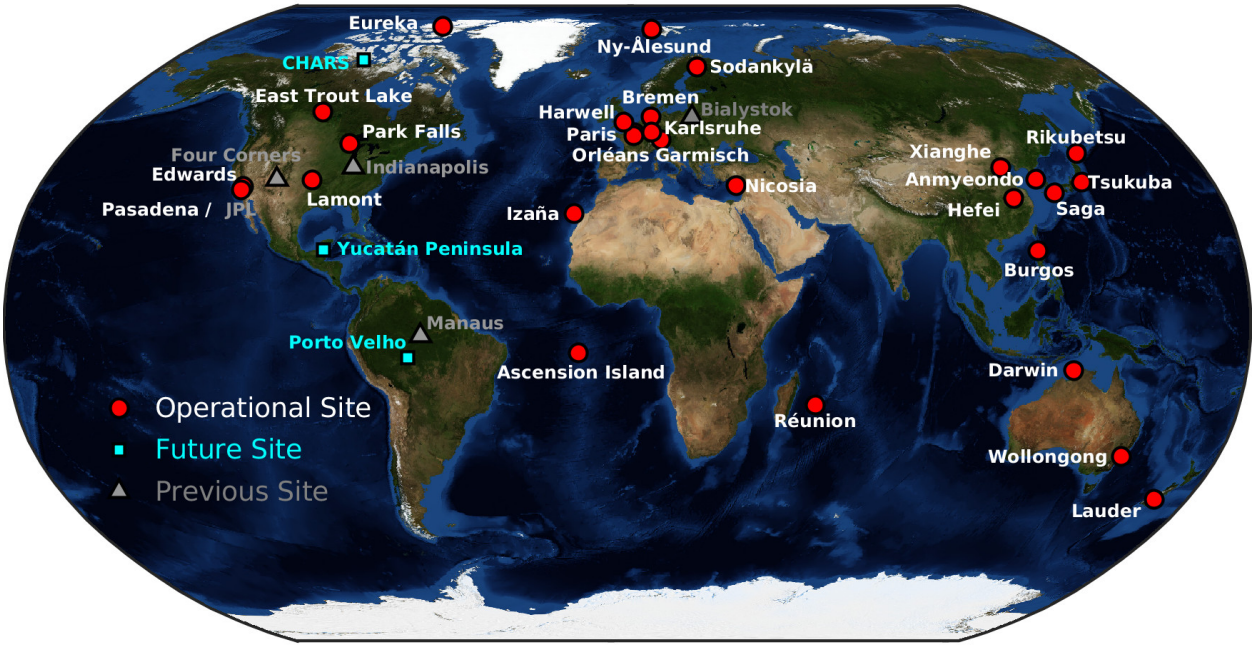

Figure 1.8: A map of operational, future, and previous Total Carbon Column Observing Network (TCCON) sites. https://tccondata.org/

dissertation is the Total Carbon Column Observation Network (TCCON; Wunch et al, 2011). TCCON contains a network of ground-based Fourier Transform Spectrometers measure direct solar absorption spectra in the near-infrared to measure total column-averaged dry air mole fractions of atmospheric trace gasses, including CO_2, CH_4, N_2O, CO, H_2O, and HDO (Wunch et al. 2011a). There are TCCON sites spread across the planet with the greatest density of sites in North America, Europe, and eastern Asia (Figure 1.8). The oldest TCCON site is located in Park Falls, Wisconsin, which has been in nearly continuous operation since 2004 (Washenfelder et al, 2006).

TCCON provides ground-based XCO_2 soundings at a repeat cycle of approximately 2 minutes throughout the daylight hours in sunny to mostly sunny conditions across several sites globally and is also used for validation for OCO (Wunch et al, 2011; 2017) and GOSAT (Parker et al, 2011;

Cogan et al, 2012; Yoshida et al, 2013; Kulawik et al, 2016) soundings. The longevity, fidelity, and high temporal resolution of XCO_2 data collected in the TCCON network has also provided key insights into the behavior of XCO_2, such as sensitivity to atmospheric transport (Keppel-Aleks et al, 2012; Torres et al, 2019), diurnal fluxes (Yang et al, 2007; Torres et al, 2019), and seasonal exchanges (Basu et al, 2011; Messerschmidt et al, 2012). In this thesis, we use TCCON data to partition the sources and quantify the magnitude of variability of XCO_2 across diurnal to synoptic-scale time scales.

1.1.4 Inferring carbon fluxes

Understanding the biogeochemical processes that govern global carbon uptake is critical for reducing outstanding uncertainties in climate model projections through 2100 (IPCC, 2021). In Bonan and Doney (2018), model uncertainty in Earth System Models (ESMs) is largely driven by uncertainty in representing biogeochemical processes. Through 2100, model uncertainty represents about 10 percent of uncertainty in simulated carbon uptake over the oceans, but roughly 80 percent of the uncertainty over land (Figure 1.9; Bonan and Doney, 2018). While uncertainty in anthropogenic emission scenarios dominate ocean uptake uncertainty, uncertainty in biogeochemical processes become increasingly important at regional and biome scales (Bonan and Doney, 2018). In order to reduce model uncertainty in carbon uptake, we have to improve our understanding of the biogeochemical processes that govern contemporary carbon fluxes and how these processes respond to changes in climate forcing through 2100 (Bonan and Doney, 2018). Providing constraints to global carbon fluxes at increasing spatiotemporal resolution is a key way to increasing our understanding of the natural processes that govern model uncertainty (Crisp et al, 2021).

1.1.4.1 "Bottom-up" approaches

Carbon fluxes are traditionally inferred using "bottom-up" or "top-down" approaches. While this dissertation primarily focuses on top-down approaches, we briefly describe bottom-up approaches here. Bottom-up data products often infer carbon fluxes by scaling up processes or statistics col-

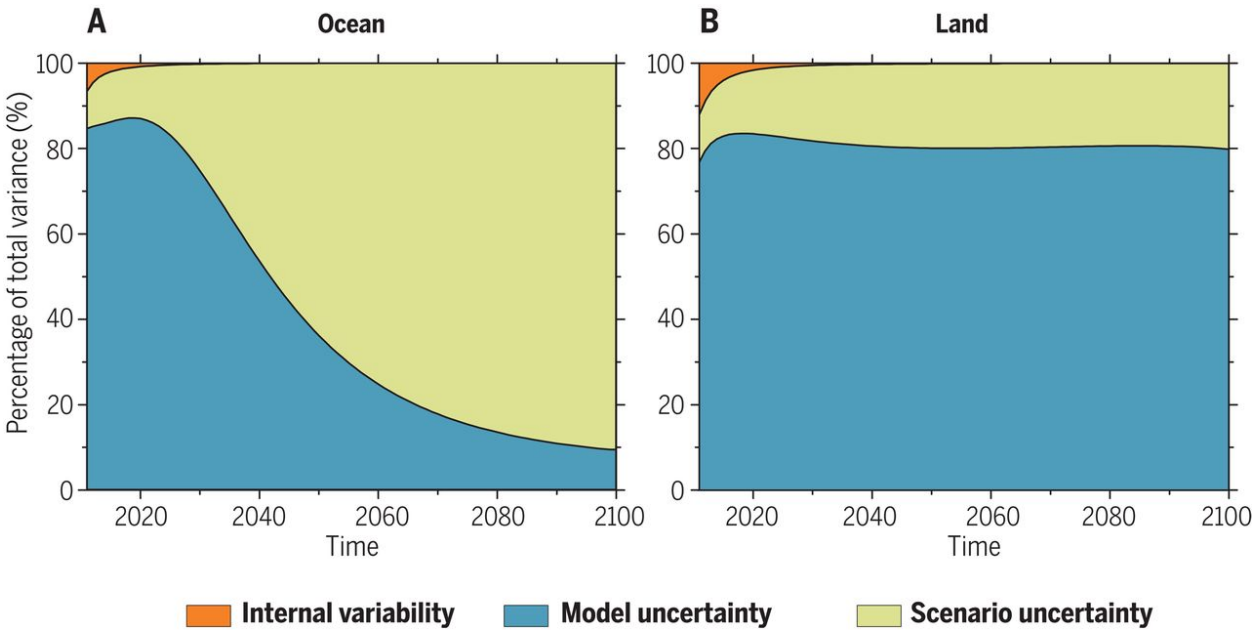

Figure 1.9: From Bonan and Doney (2018): "The percentage of total variance attributed to internal variability, model uncertainty, and scenario uncertainty in projections of cumulative global carbon uptake from 2006 to 2100 differs widely between (A) ocean and (B) land. The ocean carbon cycle is dominated by scenario uncertainty by the middle of the century, but uncertainty in the land carbon cycle is mostly from model structure. Data are from 12 ESMs using four different scenarios."

lected at local, regional, and global levels. Statistically-based products include fossil fuel inventories, such as the Emissions Database for Global Atmospheric Research (EDGAR; Muntean et al, 2018), the scaling-up of eddy covariance flux tower measurements, such as FLUXNET (Baldocchi et al, 2001), or inferring ocean-atmosphere fluxes by scaling up measurements of partial pressure of CO_2 in the oceans, such as the Surface Ocean pCO_2 Mapping intercomparison (SOCOM) initiative (Rödenbeck et al, 2015).

Process-based bottom-up products include terrestrial biosphere models (TBMs) and oceanic biogeochemical models. The accuracy of bottom-up estimates of carbon fluxes largely depend on the accurate representation of all relevant processes and the fidelity of the statistics collected. Intercomparison projects, such as the North American Carbon Program (NACP) regional interim synthesis, have benchmarked an ensemble of TBMs ability to estimate carbon fluxes over North America (Huntzinger et al, 2012). Within NACP regional interim synthesis, the discrepancies among TBMs were so large that the models could not decipher whether North America was a net sink or source of CO_2 (Huntzinger et al, 2012). This has motivated longer-term intercomparison projects, such as the Multi-Scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) to systematically identify structural differences within member TBMs to better infer global carbon fluxes (Figure 1.10; Hutzinger et al, 2013). Still, differences in the representation of processes driving CO_2 fluxes among TBMs are large enough that the inter-model spread of the cumulative land sink from 1959 to 2010 varies from a net source of CO_2 to a net sink of over 200 PgC per year (Figure 1.10; Huntzinger et al, 2017).

1.1.4.2 "Top-down" approaches

An alternative to bottom-up methods are top-down approaches, which are heavily referenced throughout the remainder of this dissertation. Top-down methods are generally more agnostic to processes that govern carbon fluxes and are more focused on spatiotemporal quantification of mass exchange between the atmosphere and the surface. Top-down methods infer carbon fluxes from spatiotemporal changes in CO_2 observations. The earliest top-down model inferred CO_2 fluxes at

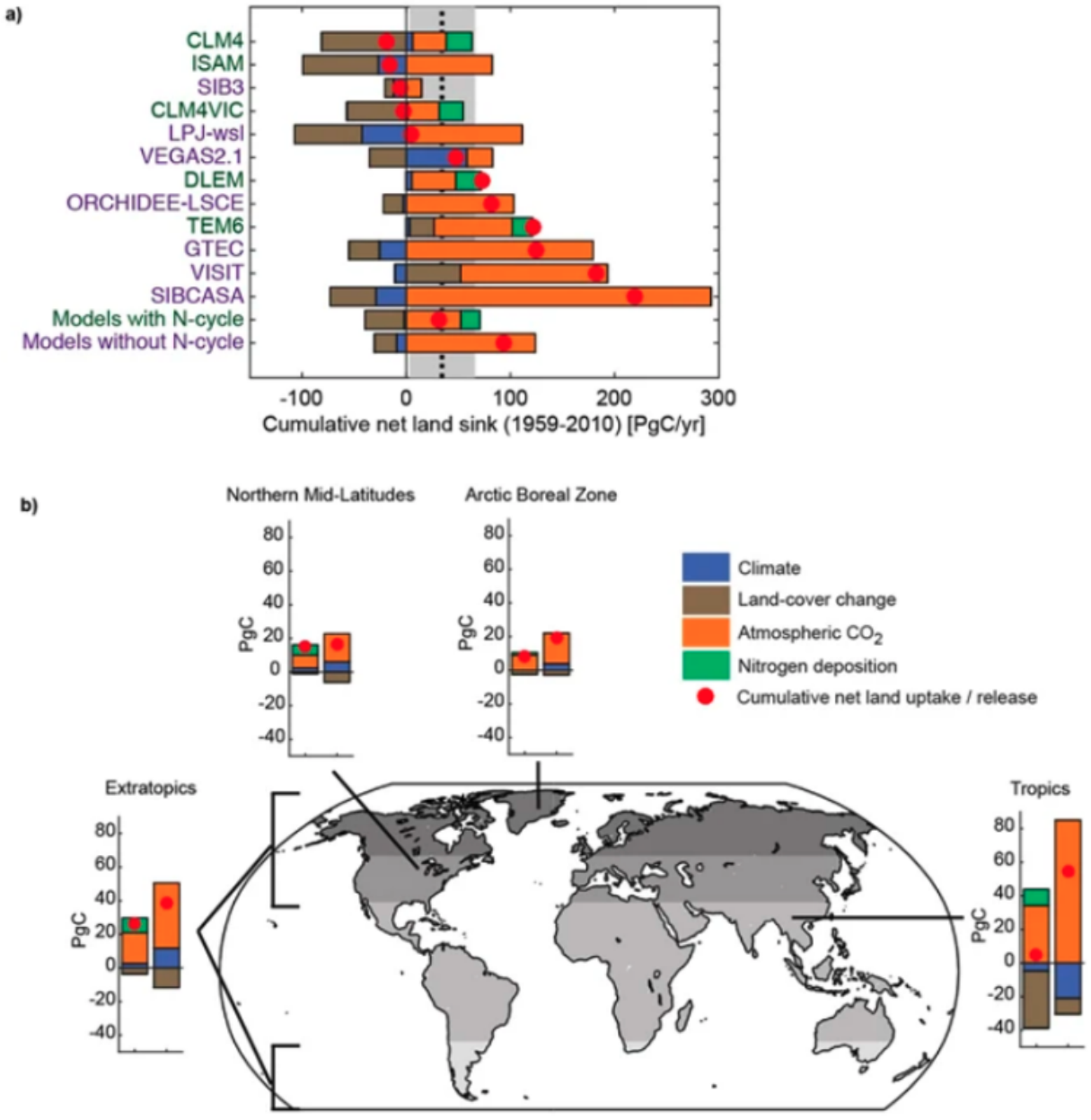

Figure 1.10: The cumulative net land carbon partitioned by climate forcing, land-cover change, atmospheric CO_2, nitrogen, and cumulative net land uptake/release among (a.) 12 TBMs and (b.) by region (Huntzinger et al, 2017). ”

the hemispheric scale, which crudely concluded the tropical oceans as a net source of CO_2 (Bolin and Keeling, 1963). While the actual magnitude of the CO_2 fluxes may have been inaccurate, this paper provided a robust roadmap for using atmospheric CO_2 observations to infer carbon fluxes (Broeker et al, 1979; Keeling et al, 1989; Tans et al, 1990).

The most common top-down method involves using inverse models, which are incrementally more robust to errors in observations, atmospheric transport, and other proxies (Enting et al, 1995). As an illustrative example of inverse models, a relatively common way to estimate CO_2 fluxes uses Bayesian least squares as a mode of posterior distribution (Enting, 2002). Here, the cost function (J) which is minimized to estimate the value of the "true carbon flux" ($\mathbf{x}$) (Equation 1.1). The difference between CO_2 observations ($\mathbf{c}$) and $\mathbf{x}$, which is scaled by a gain matrix ($\mathbf{G}$), is weighted against the difference between $\mathbf{x}$ and a priori assumptions of the carbon fluxes ($\mathbf{z}$). Each term in the cost function is weighted with inverse covariance matrices $\mathbf{X}$ and $\mathbf{W}$, which represents uncertainty in the measurements and a priori fluxes. $\mathbf{G}$ represents external factors that connect $\mathbf{c}$ to $\mathbf{x}$, which in the case of CO_2 observations and fluxes, is often atmospheric transport. Thus, the outcome of these models depends on which value for $\mathbf{x}$ minimizes J, which then provides the best estimate of "true" carbon flux, $\mathbf{x}$. The accuracy of posterior estimates depend heavily on the quality of the observations, a priori assumptions, and accurate constraints of errors in both the observations and assumed fluxes.

$$J(\mathbf{x}) = (\mathbf{c} - \mathbf{G}\mathbf{x})^T \mathbf{X} (\mathbf{c} - \mathbf{G}\mathbf{x}) + (\mathbf{x} - \mathbf{z})^T \mathbf{W} (\mathbf{x} - \mathbf{z}) \tag{1.1}$$

Inverse models have also long been subject to significant uncertainties in regional carbon fluxes (Figure 1.11; Gurney et al, 2002). As shown in Equation 1.1, discrepancies among flux inversions are a function of differing prior assumptions, sparse observations along with poorly characterized measurement errors in observations, and errors in quantifying atmospheric transport. Misrepresentations in atmospheric transport can contribute up to systemic biases of up to 1.7 GtC per year in inverse model flux inferences (Schuh et al, 2019). Again, to compare to the Global Carbon Budget, the atmosphere gains approximately 5.2 GtC per year (Friedlingstein et al, 2022). Biases in

transport can lead to the misattribution of carbon sinks and sources over key carbon rich biomes, such as the tropics, temperate, boreal regions, and/or oceanic sinks (Gurney et al, 2002; Stephens et al, 2007; Schuh et al, 2019). Basu et al (2018) demonstrated that the dependence of transport on posterior fluxes can be somewhat mitigated simply by incorporating the abundance of global XCO_2 observations. Given the importance and intractability of accurately representing fine-scale atmospheric transport of CO_2, an overview of what is known about atmospheric transport is described in the next section 1.1.5.

1.1.5 Atmospheric Transport

1.1.5.1 Vertical mixing and the "Rectifier Effect"

As demonstrated in section 1.1.4.2, top-down estimates of carbon fluxes heavily depend on spatiotemporal variations of CO_2 observations. The drivers of these spatiotemporal variations are a combination of fluxes, human emissions, and atmospheric transport. In this section, I will discuss the mechanisms in which atmospheric transport can drive errors in estimates of carbon fluxes, as well as how large the imprint of atmospheric transport can be on any given observation of XCO_2.

Since CO_2 has an average residence time of approximately 5 to 200 years in the atmosphere (IPCC, 2021), the vertical and horizontal atmospheric transport of CO_2 happens across the large spectrum of spatiotemporal scales (Figure 1.12). Many studies quantify transport errors by simply running an ensemble of transport models (Baker et al, 2006; Gloor 2007; Miller et al, 2015; Schuh et al, 2019). Overall, transport models are able to accurately resolve larger-scale atmospheric processes, such as synoptic-scale variability and beyond, but struggle with resolving finer-scale transport, such as mesoscale processes (Lauvaux and Davis, 2014) and vertical mixing (Denning et al, 1994; Peters et al, 2004; Stephens et al, 2007; Schuh et al, 2019).

The reason vertical mixing between the planetary boundary layer (PBL) and free troposphere (FT) is such a significant source of errors in flux inversions is described aptly by the "Rectifier Effect" (Figure 1; Denning et al, 1995). The PBL describes the approximately lowest 2 km of the atmosphere where air flows are often turbulent due to interactions with the surface. PBL variations

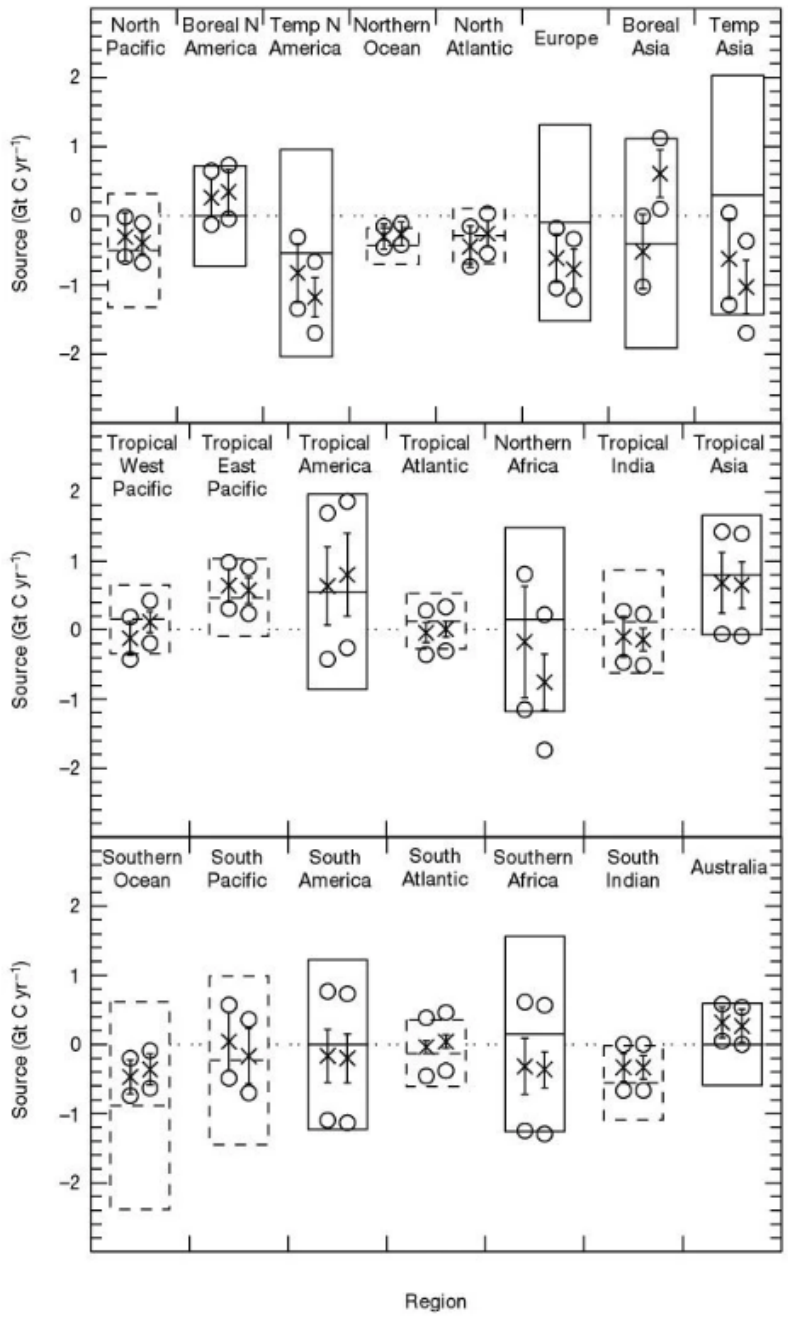

Figure 1.11: "Left-hand symbols in each box are for the control inversion, right-hand symbols are for an inversion without the background seasonal biosphere flux. Mean estimated fluxes are shown by crosses, and include all background fluxes except fossil fuel. Positive values indicate a source to the atmosphere. The prior flux estimates and their uncertainties are indicated by the boxes (solid for land, dashed for ocean); the central horizontal bar indicates the prior flux estimate, and the top and bottom of the box give the prior flux uncertainty range. The mean estimated uncertainty across all models (the 'within-model' uncertainty) is indicated by the circles. The standard deviation of the models' estimated fluxes (the 'between-model' uncertainty) is indicated by the 'error bars'. Regions are shown in their approximate north–south and east–west relationship." (Gurney et al, 2002)

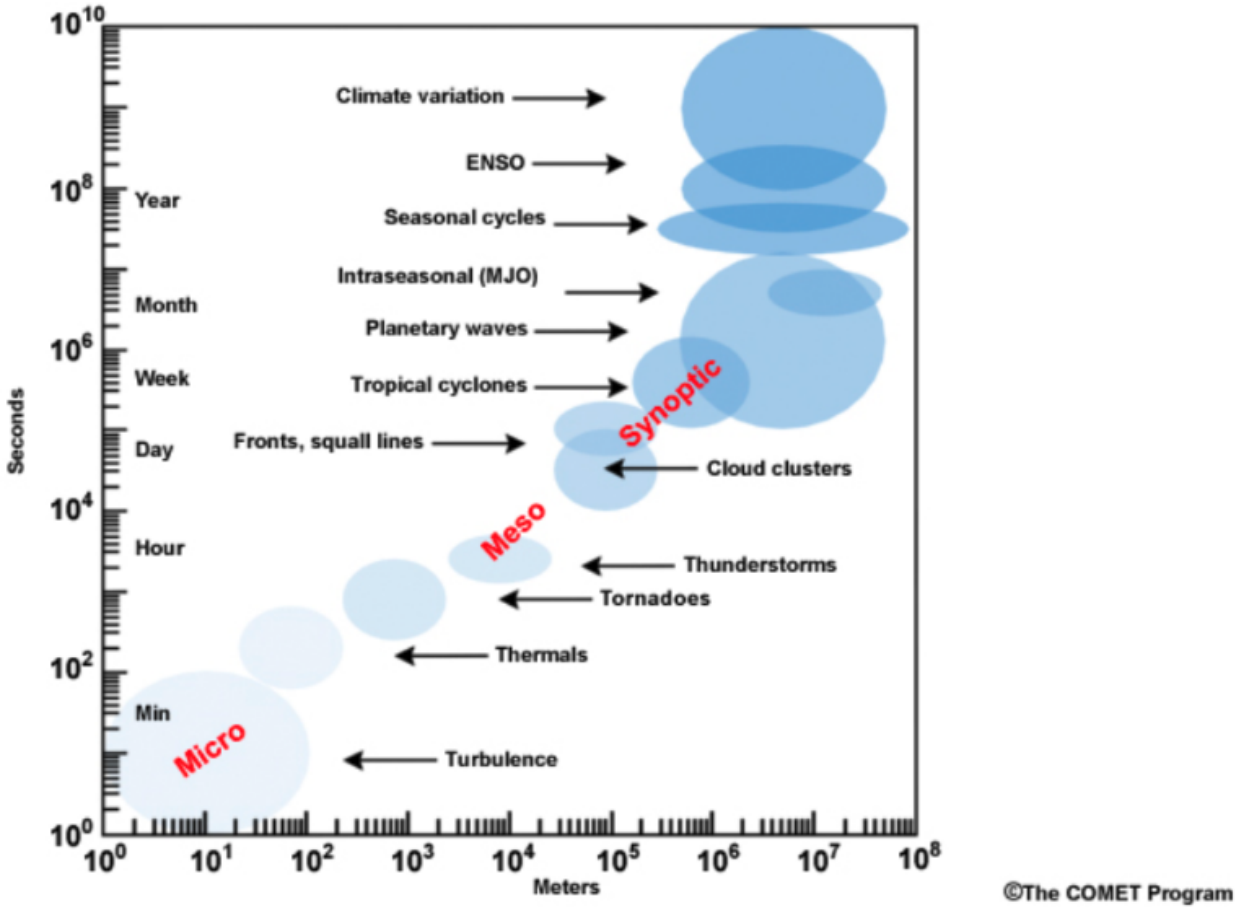

Figure 1.12: Space and time-scales of dynamical atmospheric processes. The source of this material is the COMET Website at http://meted.ucar.edu/ of the University Corporation for Atmospheric Research (UCAR), sponsored in part through cooperative agreement(s) with the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce (DOC). ©1997-2023 University Corporation for Atmospheric Research. All Rights Reserved.

in CO_2 are driven by surface fluxes as well as vertical and horizontal advection. PBL height and convective mixing vary across all spatiotemporal scales outlined in Figure 1.12. Denning et al (1995) demonstrate that seasonal variations in PBL height and convective mixing are strongly correlated with fluxes over the northern mid and high-latitude regions (Denning et al, 1995; Figure 1.13). Since CO_2 fluxes have minimal dependence on PBL height and mixing, this introduces that poor characterization of the rectifier effect results in significant errors in estimates of surface fluxes. In the warm season in the midlatitudes, advection in the PBL accounts for 60 to 70 percent of day-to-day variability (Parazoo et al, 2008). This venting of CO_2 from the PBL is commonly known as the "Rectifier Effect."

Poor characterization of vertical mixing is also evident when using tracers with well-constrained fluxes (or in this case – emissions), such as sulfur hexafluoride (SF6). Peters (2004) flagged the impact atmospheric transport errors can have on posterior flux estimates from inverse models by modeling well-known global concentrations of SF6. Ultimately, errors in vertical mixing between the boundary layer and free troposphere from transport models led to a 19 percent error in the global meridional gradient of SF6 (Peters, 2004). TransCom 3, which was an intercomparison of several inverse models, estimated carbon fluxes across 22 distinct regions using 78 global CO_2 measurement sites and quantified transport errors using an ensemble of 13 transport models (Baker et al, 2006).

Any top-down flux inference using atmospheric transport models must properly account for vertical mixing. By improperly accounting for vertical gradients in CO_2, Stephens et al (2007) found that flux inversions can overestimate the magnitude of northern hemisphere terrestrial uptake by over 1 PgC per year, and underestimate tropical uptake by almost 2 PgC per year (Figure 1.11), which is quite significant compared to average global annual CO_2 growth rates from 2012-2021 of 5.2 PgC per year (Friedlingstein et al, 2022). For the tropics, this flux discrepancy is large enough to switch the region from what was considered a net source to a potentially a net sink of carbon (Stephens et al, 2007).

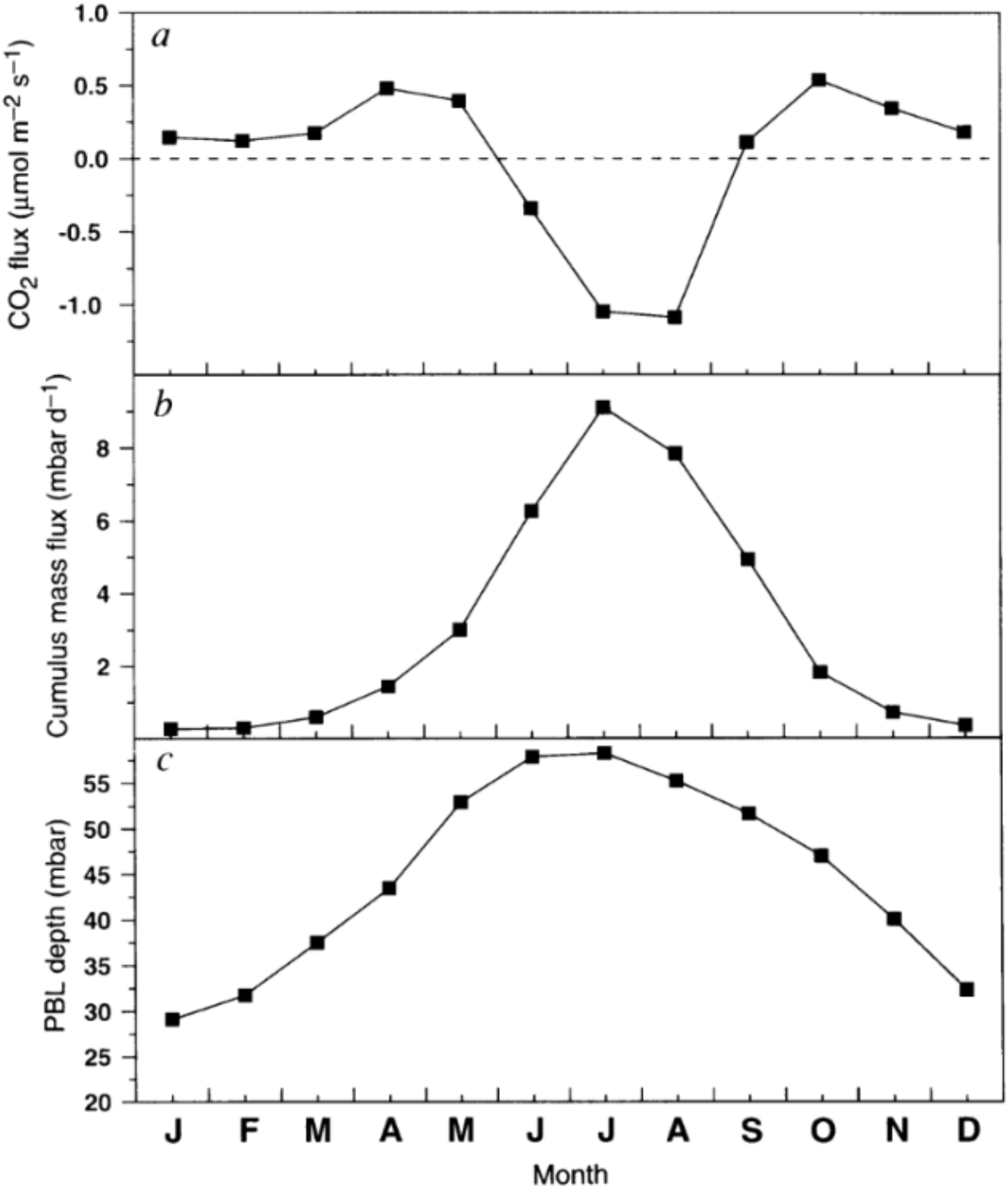

Figure 1.13: "Seasonal variations: a, of prescribed surface CO_2 flux from the terrestrial biosphere to the atmosphere, b, of simulated atmospheric mass flux due to cumulus convection at the top of the top of the atmospheric planetary boundary layer (PBL) and c, of the simulated depth of the atmospheric PBL. Values are area-weighted means for land points north of 28ºN, averaged for each calendar month." (Denning et al, 1995)

1.1.5.2 Horizontal atmospheric transport

CO_2, which is introduced into the FT from the PBL via vertical mixing, ultimately gets transported horizontally by mesoscale, synoptic-scale, and planetary scale eddies across the mid-to-high latitudes (Parazoo et al, 2011). Parazoo et al (2011) shows that the net effect of these eddies is the meridional transport of CO_2 into the high latitudes (Figure 1.14). While the bulk of CO_2 variations by month are driven by surface fluxes in the mid-latitudes, monthly variations of CO_2 in the Arctic are largely attributed to meridional eddy transport from synoptic- and planetary-scale advection. These larger-eddy-driven variations are both quite pronounced in XCO_2 observations and are maximized by the magnitude of the meridional gradient of XCO_2, which is maximized during the growing season and minimized during the winter (Keppel-Aleks et al, 2011; 2012).

A clear example of horizontal advection of CO_2 in the FT is shown by Keppel-Aleks et al (2012; Figure 1.15). Here, two vertical profiles of CO_2 were sampled 3 hours apart over Park Falls, Wisconsin. CO_2 in the PBL remained relatively constant. However, in the FT, CO_2 varies by about 5 ppm. Assuming the FT comprises about 70 percent of the atmospheric column, this would result in a 3 hour change of XCO_2 of approximately 3.5 ppm. For reference, a typical diurnal change of XCO_2 over Park Falls, Wisconsin during the same time of year is 1.2 ppm (Torres et al, 2019).

Large-scale horizontal advections of XCO_2 has been shown that it can be approximately modeled by using proxies for atmospheric transport, such as 700 mb potential temperature (Keppel-Aleks et al, 2011; 2012). A case study during the Atmospheric Carbon and Transport - America (ACT-America) field campaign illustrates the near perfect alignment between potential temperature and vertical variations in CO_2 in the FT during the passage of a warm and cold front (Figure 1.16; Samaddar et al, 2021). Note that at spatial scales that are 100 km or 1 degree, the correlation between potential temperature and CO_2 weakens yielding the need for alternative representations of mesoscale and finer scale atmospheric transport (Samaddar et al, 2021). Spatially, mesoscale transport can be represented using high pass filters and semivariograms (Torres et al, 2019; Mitchell et al, 2022).

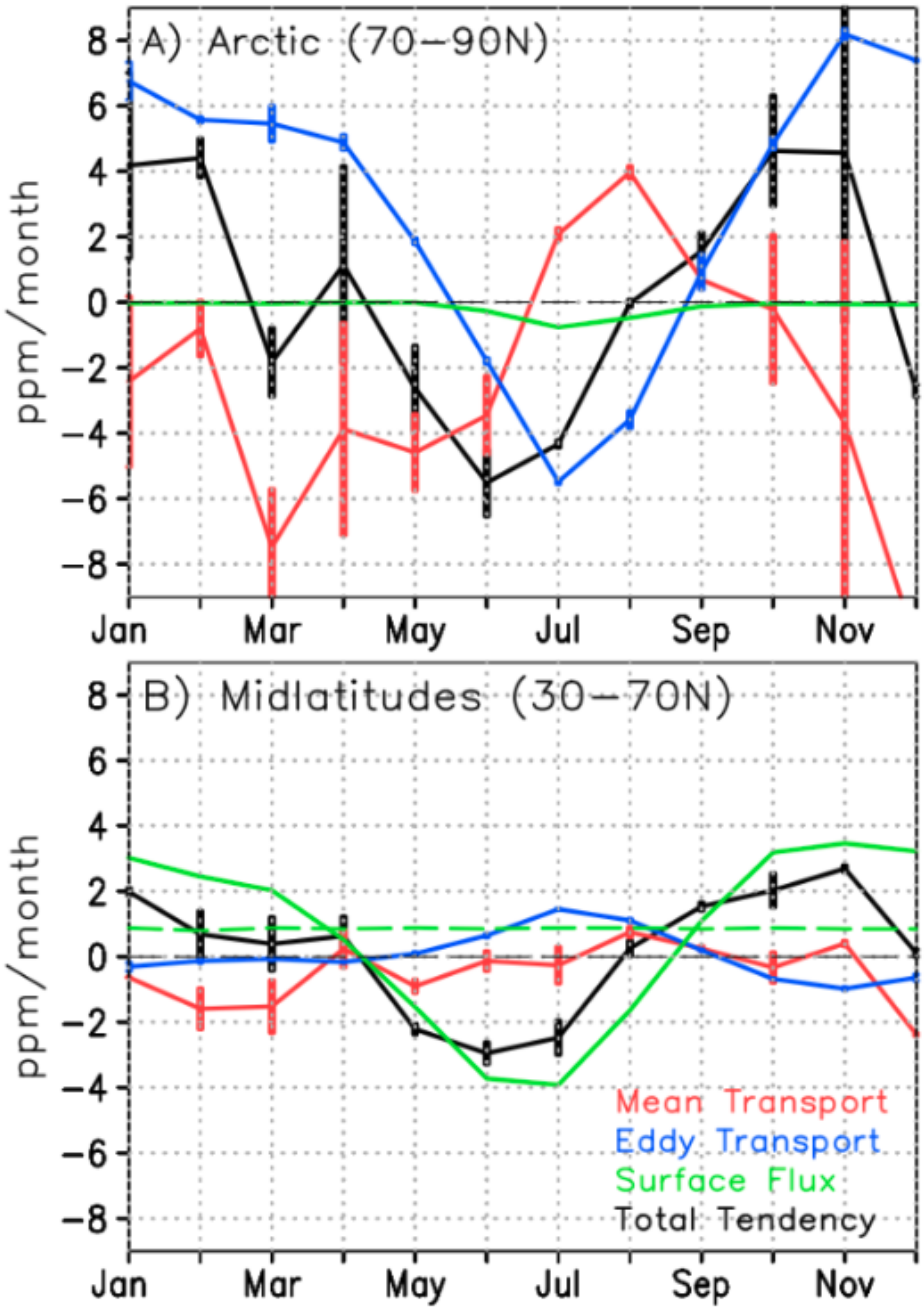

Figure 1.14: "Column integrated seasonal CO_2 budget tendencies (ppm month^{-1}) on θ_e binned into (a) high latitudes and (b) mid-latitudes. Transport by transient and stationary eddies is plotted in blue, mean meridional transport in red, the total CO_2 tendency in black, and the total surface flux of carbon from land, ocean, and fossil fuels in green. The fossil fuel tendency is plotted as a dashed green line. The sum of individual tendencies (red, blue and green lines) is equal to the total tendency (black line). Error bars represent the root mean squared error." (Parazoo et al, 2011)

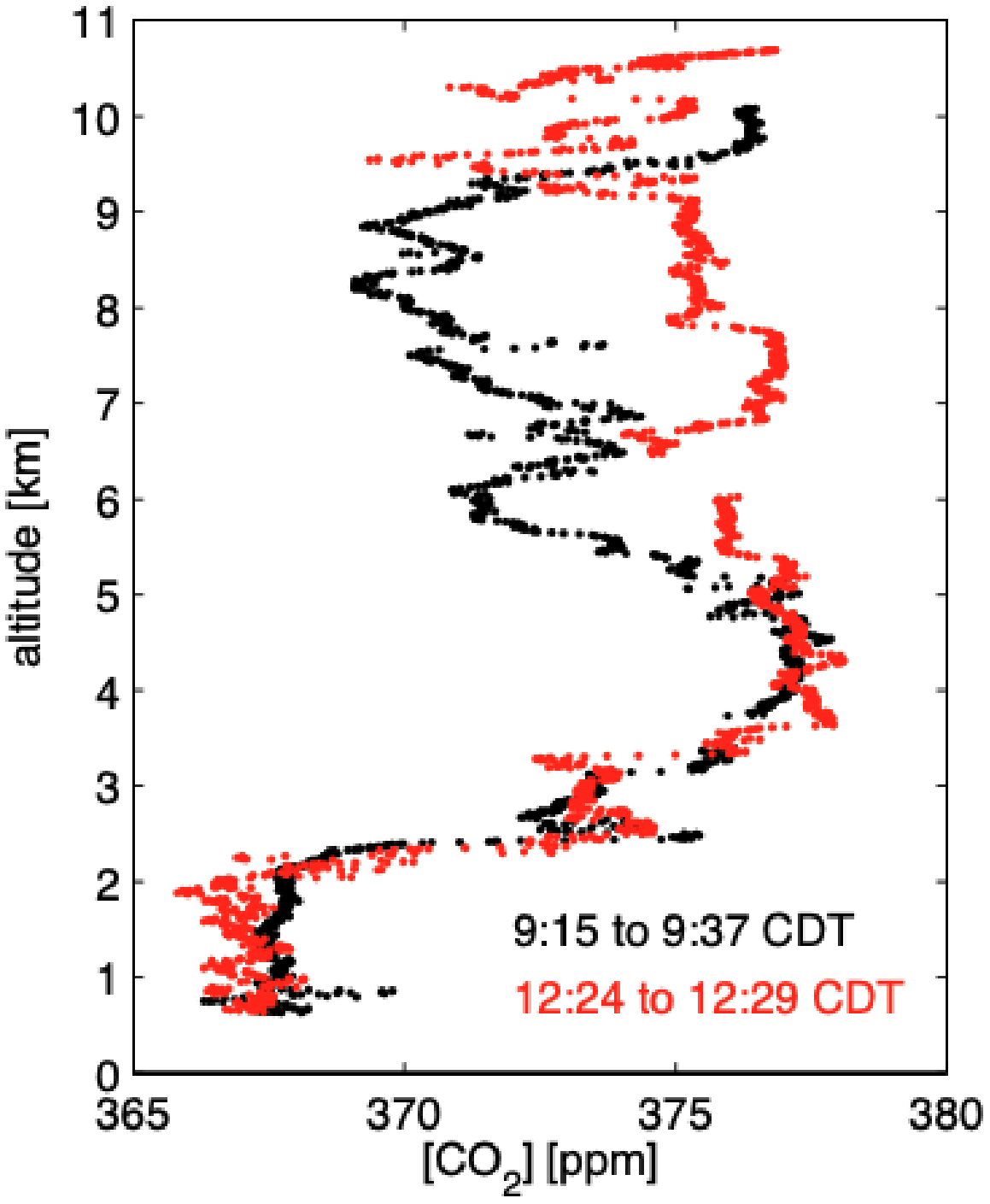

Figure 1.15: From Keppel-Aleks et al (2012): "Two profiles obtained above Park Falls on 15 July 2004 within three hours during the INTEX-NA campaign. A frontal system moved through the region between the profiles, leading to [an approximately] 5 ppm increase in free tropospheric CO_2 between 5-9 km. The decrease in boundary layer CO_2 is notably smaller."

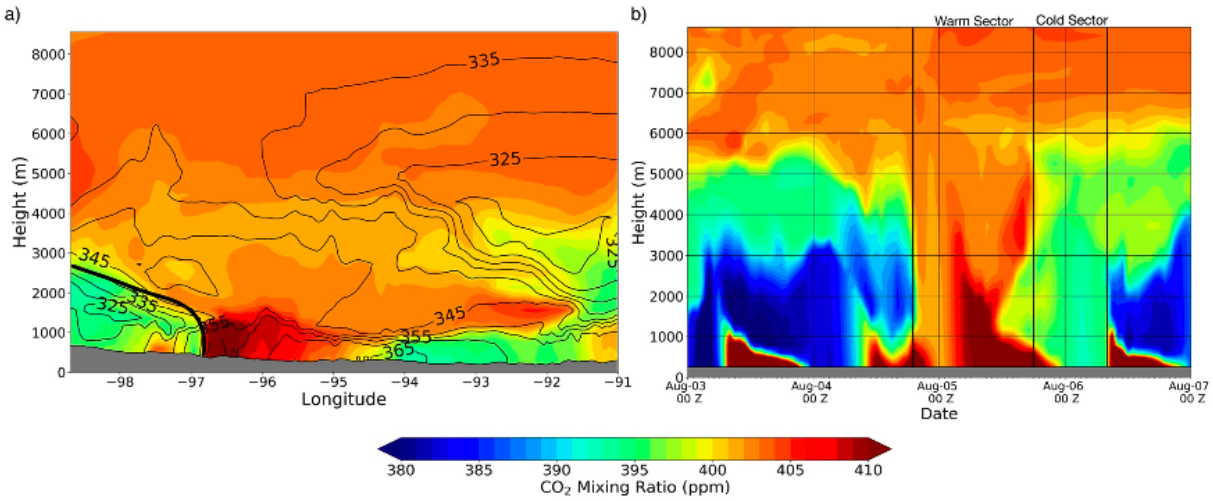

Figure 1.16: From Samaddar et al (2021): "Vertical distribution of CO_2 during a cold front passage. (a) Vertical distribution (MSL) of CO_2 along the transect (white line in Figure 6b) shown in Figure 6 highlighting the warm and cold sector of the front on August 4th at 18Z. The bold black line shows the slanted structure of the front in the cold sector with lower CO_2 mole fractions. (b) Time evolution of CO_2 mole fractions over the reference location (white star in Figure 6b at 40.9 N and 96.9 W) from August 3rd to August 7th 00Z. The gray regions show the terrain. The vertical black lines in panel (b) show the period of frontal influence from August 4th 18Z to August 6th 09Z over the reference location. The black vertical lines highlight the period of warm and cold sector passage over the location."

1.1.6 Thesis Overview

In this dissertation, we attempt to partition the sources of variability of XCO_2 at subseasonal timescales. Throughout the thesis, variations in XCO_2 are broadly explained in terms shown in Equation 1.2:

$$XCO_2 = XCO_{2,trend} + XCO_{2,transport} + XCO_{2,fluxes} + \epsilon \tag{1.2}$$

where variations in XCO_2 are driven by anthropogenic emissions ($XCO_{2,trend}$), multi-scale atmospheric transport ($XCO_{2,transport}$), multi-scale fluxes ($XCO_{2,fluxes}$), and measurement errors (ϵ). Note that Equation 1.2 neglects to mention interannual and seasonal variations in XCO_2. For simplicity in this introduction, these terms are implicitly folded into the flux and transport terms. In the subsequent chapters, the XCO_2 variance budget is partitioned into smaller subcategories.

The overarching science question addressed throughout this thesis is:

How can we use statistical analyses of XCO_2 observations to partition variations of XCO_2 into component parts of atmospheric transport and local diurnal fluxes (as shown in Equation 1.1.)?

Chapter 2 includes my published work from 2019 in the Journal of Geophysical Research: Atmospheres, entitled "A geostatistical framework for quantifying the imprint of mesoscale atmospheric transport on satellite trace gas retrievals" (Torres et al. 2019). In this paper, we use semivariograms, which is a measure of along-track spatial variability, to estimate the magnitude of fine scale variability ($<$250 km) of XCO_2 retrievals from the OCO-2 satellite, which we attribute primarily to mesoscale atmospheric transport. We also include preliminary work to independently quantify sources of errors within the elements of the state vector of the XCO_2 retrieval algorithm, such as albedo and aerosol optical depth.

Chapter 3 focuses on our ability to leverage OCO-3's precessing orbital pattern aboard the International Space Station to make global estimates of the diurnal cycle of XCO_2. Specifically, we use TCCON observations (Wunch et al, 2011), as well as synthetic XCO_2 from the 2019 version of the CarbonTracker model (Jacobson et al., 2020), to estimate the minimum number of overpasses

required to reasonably quantify the magnitude of the diurnal cycle of XCO_2. From Torres et al. (2019), the magnitude of the climatological diurnal cycle of XCO_2 directly scales with the magnitude of local diurnal fluxes. Thus, this research provides a potential framework for directly inferring the magnitude of local fluxes globally using space-based observations. The manuscript for this project is in prep with an expected submission in May 2023.

Finally, this thesis closes with a brief conclusion in Chapter 4 tying all of the research together along with ideas for future work.

Chapter 2

A Geostatistical Framework for Quantifying the Imprint of Mesoscale Atmospheric Transport On Satellite Trace Gas Retrievals

This chapter has been published in the Journal for Geophysical Research: Atmospheres (see Torres et al, 2019).

2.1 Key Points

1. We developed a framework to relate high-frequency spatial variations to transport-induced temporal fluctuations in atmospheric tracers
2. We use geostatistical analysis to quantify the variance budget for XCO_2 and XH_2O retrieved from NASA's OCO-2 satellite
3. Accounting for random errors, systematic errors, and real geophysical coherence in remotely sensed trace gas observations may yield improved flux constraints

2.1.1 Abstract

NASA's Orbiting Carbon Observatory-2 (OCO-2) satellite provides observations of total column-averaged CO_2 mole fractions (XCO_2) at high spatial resolution that may enable novel constraints on surface-atmosphere carbon fluxes. Atmospheric inverse modeling provides an approach to optimize surface fluxes at regional scales, but the accuracy of the fluxes from inversion frameworks depends on key inputs, including spatially and temporally dense CO_2 observations and reliable

representations of atmospheric transport. Since XCO_2 observations are sensitive to both synoptic and mesoscale variations within the free troposphere, horizontal atmospheric transport imparts substantial variations in these data, and must be either resolved explicitly by the atmospheric transport model or accounted for within the error covariance budget provided to inverse frameworks. Here, we used geostatistical techniques to quantify the imprint of atmospheric transport in along-track OCO-2 soundings. We compare high-pass filtered (<250 km, spatial scales that primarily isolate mesoscale or finer scale variations) along-track spatial variability in XCO_2 and XH_2O from OCO-2 tracks to temporal synoptic and mesoscale variability from ground-based XCO_2 and XH_2O observed by nearby Total Carbon Column Observing Network (TCCON) sites. Mesoscale atmospheric transport is found to be the primary driver of along-track, high frequency variability for OCO-2 XH_2O. For XCO_2, both mesoscale transport variability and spatially coherent bias associated with other elements of the OCO-2 retrieval state vector are important drivers of the along-track variance budget.

2.1.2 Plain Language Summary

Numerous efforts have been made to quantify sources and sinks of atmospheric CO_2 at regional spatial scales. A common approach to infer these sources and sinks requires accurate representation of variability of CO_2 observations attributed to transport by weather systems. While numerical weather prediction models have a fairly reasonable representation of larger-scale weather systems, such as frontal systems, representation of smaller-scale features (<250 km), is less reliable. In this study, we find that the variability of total-column averaged CO_2 observations attributed to these fine-scale weather systems accounts for up to half of the variability attributed to local sources and sinks. Here, we provide a framework for quantifying the drivers of spatial variability of atmospheric trace gases rather than simply relying on numerical weather prediction models. We use this framework to quantify potential sources of errors in measurements of total-column averaged CO_2 and water vapor from NASA's Orbiting Carbon Observatory-2 (OCO-2) satellite.

2.2 Introduction

Knowledge of regional surface-atmosphere carbon dioxide (CO_2) fluxes are required to understand anthropogenic and climatic influences on the global carbon cycle. Despite longstanding research efforts to develop a robust budget for surface fluxes of CO_2 (Bolin and Keeling, 1963; Enting and Mansbridge, 1989; Tans et al., 1990; Keeling et al., 1996; Peters et al., 2005; Chevallier et al., 2010; Peylin et al., 2013; Basu et al., 2018), these studies diverge in their estimates of the geographic distribution of sources and sinks of CO_2 (Gurney et al., 2002; Baker et al., 2006; Stephens et al., 2007). For example, Gurney et al. (2002) found uncertainties in regional scale carbon fluxes were greater than 0.5 Gt C yr-1 across various inversion frameworks. In these studies, carbon fluxes are inferred from spatial and temporal variations in atmospheric CO_2 observations via atmospheric inverse methods. Atmospheric inversions typically apply Bayesian optimal estimation methods to optimize assumed (a priori) fluxes that have been used as boundary conditions to simulate spatiotemporal CO_2 variations in an atmospheric transport model. Mismatches between the simulated and observed atmospheric CO_2 provide the basis for scaling the assumed fluxes. The optimization requires rigorous attention to errors associated with the assumed flux structure, the observations, and the fidelity of atmospheric transport modeled by the framework (Rogers, 2000). One limitation to inverse modeling studies has been the density and geographic distribution of atmospheric observations available to constrain surface fluxes (Gurney et al., 2002). Traditionally, observations of atmospheric CO_2 have been measured in situ or via flask sampling within the atmospheric boundary layer. These observatories are concentrated within Northern Hemisphere temperate latitudes, and there is a scarcity of observations in key regions for the global carbon cycle, including the tropics (Stephens et al., 2007) and the Southern Ocean (Landschützer et al, 2015). The sparse in situ network for atmospheric CO_2 observations was the impetus for the launch of several satellites, including Japan Aerospace Exploration Agency's (JAXA) Greenhouse gases Observing SATellite (GOSAT) (Yokota et al, 2009; Ross et al., 2013), NASA's Orbiting Carbon Observatory-2 satellite (OCO-2) (Crisp et al., 2004; Eldering et al., 2017), and Chinese National Space Administration's (CNSA) TanSat (Yang et al., 2018). These low Earth, polar-orbiting satellites measure the total

column averaged dry air mole fraction of atmospheric CO_2 (denoted as XCO_2) at high spatial density. For example, OCO-2 acquires approximately 1 million soundings every day, each with a footprint on the order of 1 km2 (Crisp et al., 2004; Eldering et al., 2017). Theoretical studies have hypothesized that the high spatiotemporal density of XCO_2 observations may allow for a reduction of errors in flux inferences from inversion models (Rayner and O'Brien, 2001; Baker et al, 2006; Miller et al., 2007).

A second limitation to the fidelity of inverse modeling studies is the inverse modeling framework itself, via either incorrect parameterization of atmospheric transport (Masarie et al., 2011; Williams et al., 2014; Chevallier et al., 2010; Houwelling et al., 2015; Basu et al., 2018) or inappropriate representation of error covariance structures (Chevallier and O'Dell, 2013). The goal of this paper is to discuss the unique requirements for atmospheric transport fidelity and the description of variance budgets for $XCO_2$from the OCO-2 satellite. One advantage of measuring the column averaged mole fraction is that its variations can be used more effectively to constrain surface fluxes via mass balance. Measurements made within the planetary boundary layer are sensitive not only to fluxes at the surface, but also to the rate at which the surface flux signal is entrained into the free troposphere. The column, however, is unaffected by the vertical entrainment rate, so in theory it is more directly related to surface fluxes via mass balance (Rayner and O'Brien, 2001; Olsen and Randerson, 2004). For example, Basu et al. (2018) concluded that fluxes inferred from perfect, or error-free, satellite observations of XCO_2 are less sensitive to uncertainty in atmospheric transport than perfect in situ observations of CO_2 in the planetary boundary layer by applying flux inversion techniques to the output from different atmospheric transport models forced with the same CO_2 initial and boundary conditions.

A potential complication of using XCO_2, however, is that it is sensitive to CO_2 within the free troposphere, where most weather occurs. The variance budget is therefore strongly affected by horizontal advection (Geels et al., 2004; Keppel-Aleks et al., 2011), not just surface fluxes, which are the real target of obtaining and inverting atmospheric observations. In fact, at subseasonal timescales, horizontal advection dominates the variance budget in XCO_2 (Keppel-Aleks et al.,

2011). Keppel-Aleks et al. (2012) used ground-based observations of XCO_2 at several midlatitude sites in the Total Carbon Column Observing Network (TCCON) to show that synoptic-scale variations, which occur on spatial scales on the order of 1000 km and temporal scales of about one to two weeks, could be up to half the peak-to-trough seasonal cycle in XCO_2. Likewise, horizontal advection drives up to 60 to 70 percent of diurnal variations of boundary layer CO_2 in the midlatitudes (Parazoo et al., 2008), and these horizontal eddy-induced variations are roughly half the size of the seasonal CO_2 variations driven by regional net ecosystem exchange of XCO_2(Parazoo et al., 2011). At finer spatial scales on the order of 100 km, mesoscale variability in XCO_2, which occurs on timescales of around one day, can be larger than diurnal variations in XCO_2 resulting from CO_2 exchange with the local terrestrial ecosystem (Keppel-Aleks et al., 2012). Mesoscale transport imposes especially large errors on flux inversions over cities at time scales smaller than a month (Lauvaux et al, 2016). However, even with global scale inversions, the variations on XCO_2 imparted by fine scale transport may ultimately degrade the inverted fluxes. This is largely due to the fact that signal imposed by fine and large-scale atmospheric transport is spatially correlated, thus these variations cannot simply be addressed simply by averaging multiple observations (Miller et al, 2015).

Thus, efforts to use XCO_2 from OCO-2 for flux inference must reliably account for transport-induced time/space variations, either through explicit simulation within the atmospheric transport model or by representation of transport-induced errors within the error covariance matrix. We note that larger-scale synoptic weather systems are more likely to be simulated explicitly by atmospheric inverse modeling frameworks, which generally have horizontal resolutions between 0.5º to 5º (Corbin et al., 2008), whereas mesoscale systems occur at spatial scales smaller than the grid-cell resolution for all but the highest resolution atmospheric transport models. These smaller mesoscale systems, therefore, may not be represented explicitly by atmospheric transport models despite affecting the distribution of XCO_2. Because mesoscale or frontal systems may also have clouds, which obscure space-based XCO_2 measurements, it is important to quantify the variance and spatial coherence of XCO_2 that will be averaged from satellite measurements before for com-

parison with a single model grid-cell value (Corbin et al., 2008).

In this study, we use complementary information from space-based (OCO-2) and ground-based (TCCON) remote sensing of XCO_2 to quantify the imprint of mesoscale atmospheric transport and to refine the variance budget of OCO-2 XCO_2. While the current suite of carbon observing satellites provide spatially dense observations, the time in between satellite overpasses at a specific location is too long (16 days for OCO-2) to sample temporal variations of XCO_2 driven by mesoscale (i.e., the duration of a thunderstorm) or synoptic-scale systems (i.e., the time in between frontal systems, which is typically on the order of a week in mid latitude regions). In contrast, ground-based networks, such as TCCON, provide temporally dense, but spatially sparse observations. To use these observations together, we must develop a framework that relates the spatial variations in OCO-2 data to the temporal variations in TCCON data.

Throughout our analysis, we are cognizant of the fact that observing system error may also introduce variance to satellite data (Baker et al., 2010; Chevallier et al., 2014). For example, Worden et al. (2017) showed that natural variability (i.e., variations due to natural surface fluxes, anthropogenic emissions, and atmospheric transport) of XCO_2 along simulated representative OCO-2 tracks was negligibly small (approximately 0.08 ppm over 100 km neighborhoods) in comparison to variations of 1.28 ppm per 100 km attributed to instrument noise and slow varying biases, such as those caused by surface pressure or albedo variations, observed in OCO-2 data. Therefore, while the primary purpose of this study is to understand synoptic and mesoscale variations of XCO_2, we also leverage retrievals of total column-averaged mole fractions of water vapor (XH_2O) from OCO-2 (Nelson et al., 2016) and TCCON observations to validate our approach to estimating terms in Equation 2.5. Retrievals of water vapor from OCO-2 have a high signal-to-noise ratio (from several hundred to greater than 1000) (Nelson et al., 2016). Therefore, we expect that synoptic and mesoscale variations of atmospheric transport XH_2O are more readily quantifiable from space-based observations.

This manuscript is organized around answering the following science questions.

1.) What is the imprint of synoptic and mesoscale systems on XCO_2 (and XH_2O)?

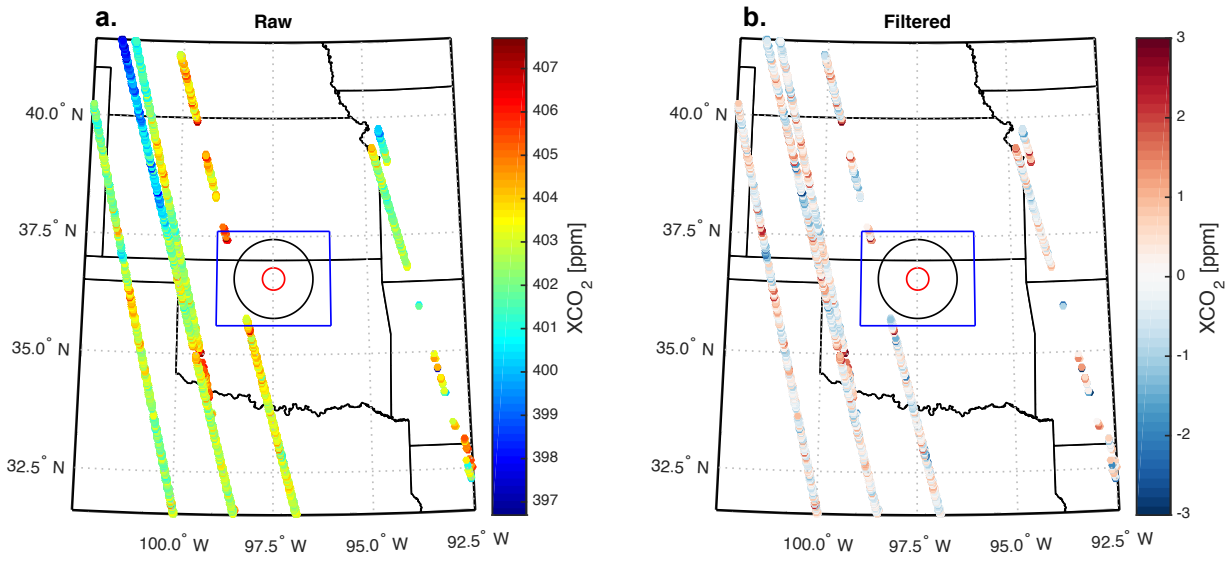

Figure 2.1: XCO_2 tracks over a 10°×10° domain centered on Lamont, Oklahoma for one 16-day repeat cycle in early July 2016. (a.) Raw XCO_2 soundings and (b.) High-pass filtered XCO_2. The radius of the red and black circle represents the monthly mean range (denoted as a_{space}) of explained variability of XCO_2 and XH_2O, respectively in July. The blue box represents a typical 3°×2° grid cell used in atmospheric inversion models, such as those used in Basu et al. (2018).

2.) How predictive are large-scale spatial gradients in XCO_2 (or XH_2O) of the imprint of synoptic and mesoscale atmospheric transport on OCO-2 observations?

3.) How large are other sources of fine-scale variation in XCO_2 (and XH_2O) in the OCO-2 variance budget?

In Section 2, we describe the methods and framework we used to quantify variability attributed to synoptic and mesoscale atmospheric transport from both TCCON and along-track OCO-2 observations of XCO_2 and XH_2O. In Section 3, we describe the variance budgets for OCO-2 XCO_2 and XH_2O in the context of validation data from TCCON. In Section 4, we provide discussion and recommendations for future work toward robust flux influence from the satellite data.

2.3 Methods

2.3.1 Framework to compare temporal and spatial variability of trace gases

This framework provides the basis to which we compare temporal mesoscale variability of along-track XCO_2 and XH_2O observed at TCCON ground sites to along-track spatial mesoscale variability from OCO-2 data. We define along-track mesoscale spatial variability of XCO_2 and XH_2O for tracks that occur within a 10º by 10º box of TCCON sites (Figure 2.1). The domains of analysis chosen were large enough to encompass several representative atmospheric transport model grid cells, such as the 3º by 2º grid cells used by Basu et al. (2018) to infer carbon fluxes.

We start from the tracer conservation equation in one dimension:

$$\frac{\partial c}{\partial t} = -u\frac{\partial c}{\partial x} + S_c \tag{2.1}$$

where c represents the tracer concentration, in this case XCO_2, u represents the column-weighted wind velocity in one direction (assumed along the OCO-2 track), and S_c represents the column-average surface sources and sinks of CO_2 (with appropriate scaling to convert from flux to XCO_2 column-average variation). In this equation, we have neglected molecular diffusion of XCO_2, which is small relative to the other terms, and any variations in u and CO_2 in the vertical profile by simply using the total column averages. We decompose c into its mean and variable components (Equation 2.2)

$$c = \overline{c} + c' \tag{2.2}$$

and Reynolds average Equation 2.1 to yield an equation for the time rate of change of $\overline{c}$ (Equation 2.3). For our analysis, we assume that the filter used to determine the average concentration, $\overline{c}$, results in a c' that reflects mesoscale variations in the tracer concentration while synoptic and slower- and larger-scale variations remain in $\overline{c}$.

$$\frac{\partial \bar{c}}{\partial t} = -\overline{u}\frac{\partial \bar{c}}{\partial x} - \frac{\partial \overline{u'c'}}{\partial x} + \overline{S_c} \tag{2.3}$$

The first term on the right hand side represents the advection of the mean gradient in c by the mean wind, while the second term represents turbulent flux divergence. Equation 2.3 underscores that spatial gradients in the mean tracer concentration give rise to temporal variations through the action of atmospheric transport. We can subtract Equation 2.3 from Equation 2.1, expanded by replacing u and c with the corresponding mean and anomaly terms from Equation 2.2 (and equivalent equation for u), to yield an equation for the time rate of change for the fluctuating component, c':

$$\frac{\partial c'}{\partial t} = -\overline{u}\frac{\partial c'}{\partial x} - u'\frac{\partial \bar{c}}{\partial x} + \frac{\partial \overline{u'c'}}{\partial x} + S'_c \tag{2.4}$$

In Equation 2.4, the first term on the right hand side represents the advection of mesoscale gradients by the mean wind, the second term and third terms represent the production of mesoscale anomalies in c by eddies acting on the mean gradient and mesoscale gradient, respectively. The fourth term represents the turbulent flux convergence. We can simplify Equation 2.4 by assuming that the production term from eddies acting on mesoscale gradients and the turbulent flux convergence are both small. We also neglect variations in sources, Sc', since our framework accounts for only climatological mean surface fluxes (described in detail in 3.2.1 below). We can then use scaling arguments to approximate these terms:

$$\frac{\langle c' \rangle_{time}}{\tau_{time}} = \overline{u}\frac{\langle c' \rangle_{space}}{a_{space}} + \langle u' \rangle \langle \frac{\partial \bar{c}}{\partial x} \rangle \tag{2.5}$$

In Equation 2.5, $\langle c' \rangle_{time}$ represents the characteristic magnitude of temporal variations at a TCCON site over a relevant mesoscale timescale time. The variable $\langle c' \rangle_{space}$ represents the characteristic magnitude of along-track spatial variations from OCO-2 over a relevant mesoscale length scale, a_{space}. The last term on the right-hand side (RHS) represents the advection of the mean gradient $\langle \frac{\partial \bar{c}}{\partial x} \rangle$ by mesoscale transport $\langle u' \rangle$.

The crux of our analysis is to compare $\langle c' \rangle_{time}$ and time inferred from empirical analysis of TCCON observations with $\langle c' \rangle_{space}$ and a_{space} inferred from geostatistical analysis of high-pass filtered OCO-2 tracks. This analysis is conducted with an eye toward using the OCO-2 derived estimates of $\langle c' \rangle_{space}$ and a_{space} to improve the representation of fine scale transport errors within the error covariance budget provided to inverse models used for flux inference.

2.3.2 TCCON

We quantified temporal synoptic and mesoscale variations in XCO_2 and XH_2O using ground-based remote sensing data from sites in the TCCON network (Table 2.1). TCCON sites are instrumented with ground-based Fourier Transform Spectrometers that acquire direct solar absorption spectra approximately every two minutes during sunny conditions (Wunch et al., 2015). TCCON instruments obtain near infrared spectra in the same spectral region as OCO-2 (0.65 – 2.63 μm), and total column CO_2 is retrieved in the 1.58 and 1.60 μm absorption bands and total column H2O is retrieved in the 1.54 – 1.65 μm absorption bands using the GFIT algorithm (Wunch et al., 2011). Because TCCON measures direct solar absorption spectra, the signal to noise ratio is higher compared to that of OCO-2, and the uncertainties on TCCON XCO_2 have a calibration accuracy of 0.4 ppm (Wunch et al., 2010). TCCON data are calibrated to the World Meteorological Organization (WMO) standard ensuring absolute accuracy of measurements better than 0.25 percent (Washenfelder et al., 2006; Wunch et al., 2011). We analyzed data from TCCON sites that have data records longer than 5 years and that observe across a full annual cycle to minimize biases introduced by seasonal and interannual variations.

2.3.2.1 Removing diurnal cycle climatology of XH_2O and XCO_2 to quantify temporal synoptic and mesoscale variability

We separated the imprint of synoptic and mesoscale systems on variations in TCCON XCO_2 and XH_2O by assuming that the only sources of variations were surface fluxes or atmospheric transport. For both XCO_2 and XH_2O, we assumed that flux-driven diurnal variations could be accounted

TCCON Site	Location	Observational Periods	Citation
Bialystok, Poland	53.33ºN, 23.03ºE	March 13, 2009 – April 14, 2017	Deustcher et al. (2014)
Karlsruhe, Germany	49.10ºN, 8.44ºE	April 19, 2010 – January 24, 2018	Hase et al. (2014)
Orleans, France	47.97ºN, 2.11ºE	August 29, 2009 – April 29, 2017	Warneke et al. (2014)
Garmisch, Germany	47.48ºN, 11.06ºE	July 16, 2007 – March 16, 2018	Sussman and Rettinger (2014)
Park Falls, Wisconsin	45.95ºN, 90.27ºE	June 2, 2004 – December 31, 2017	Wennberg et al. (2014a), Washenfelder et al. (2006)
Lamont, Oklahoma	36.60ºN, 97.49ºW	July 6, 2008 – December 31, 2017	Wennberg et al. (2014b)
Darwin, Australia	12.42ºS, 130.89ºE	August 28, 2005 – March 28, 2017	Griffith et al. (2014), Deutscher et al., (2010)
Reunion Island, France	20.90ºS, 55.49ºE	September 16, 2011 – January 30, 2018	De Mazière et al. (2014)
Lauder, New Zealand	45.04ºS, 169.69ºE	February 2, 2010 – November 1, 2017	Sherlock et al. (2014)

Table 2.1: Locations of TCCON sites and observational periods analyzed in this study with associated references.

for by calculating a monthly climatological daily cycle of XCO_2 variations for each site, since atmospheric transport patterns may be random but surface fluxes are phase-locked to the diurnal cycle. We note there are changes in surface fluxes in response to physical climate changes, such as thunderstorms/rain, cloud coverage, or boundary layer temperature, induced by mesoscale and synoptic-scale systems (Baldocchi et al., 2001). There is, however, no easy way to attribute the changes in XCO_2 and XH_2O to either changing fluxes or synoptic/mesoscale transport without running a coupled atmosphere/land model. We choose instead to use an empirical, data-driven approach that necessitates neglecting weather-driven changes in surface fluxes.

For each calendar month, we binned all available observations (after removing the long-term trend) from the multi-year time series into half-hour increments to reveal the characteristic diurnal cycle (Figure **??**-2). For any given month, we limited our analysis to daytime observations obtained at solar zenith angle less than 75º to reduce the influence of spectroscopic errors at high air masses. We then removed the climatological daily cycle from each calendar day with observations, and assumed that the residual was the component of variability driven by transport. We note that this approach is a simplification, and expect that at least some of the residual were due to synoptic, intraseasonal, and interannual variability of surface fluxes. Our approach does, however, allow us to approximate the influence of local fluxes on the observations without relying on an ecosystem model or sparse flux tower data with limited spatial footprints.

Given our assumption that temporal variability of XCO_2 and XH_2O is derived from either local fluxes or atmospheric transport, we can then estimate the influence of atmospheric transport-driven variations from the time series of residuals. We calculated the standard deviation from the half-hourly bin averaged residuals at bi-weekly time intervals to approximate variability at synoptic and smaller timescales. We likewise calculated the standard deviation of the residuals within each day to approximate mesoscale variability. These time periods were sufficient to sample variability attributed to multiple synoptic scale weather systems, such as high and low pressure systems and frontal passages, or mesoscale systems, such as individual thunderstorms.

We evaluate our approach for calculating the influence of climatological fluxes on the diurnal

cycle of XCO_2 at the Park Falls TCCON site since it is co-located with an Ameriflux eddy covariance (EC) tower that provides observations of diurnally varying NEE (Desai et al 2015). We estimate the influence that the observed eddy covariance fluxes have on the daily cycle of XCO_2 (and denote this quantity as $XCO_{2,EC}$ using equations 6 and 7:

$$\langle \frac{\mathrm{dXCO}_{2,EC}}{\mathrm{d}t} \rangle = \frac{\mathrm{NEE}_{EC} \cdot \mathrm{g} \cdot \mathrm{MW}_{\text{dry air}}}{P_s} \tag{2.6}$$

$$\mathrm{XCO}_{2,EC} = \int_{\tau_{AM}}^{\tau_{PM}} \langle \frac{\mathrm{dXCO}_{2,EC}}{\mathrm{d}t} \rangle \, \mathrm{d}\tau \tag{2.7}$$

where NEE_{EC} represents the observed net ecosystem exchange, g represents the gravitational constant of 9.81 ms-2, $\mathrm{MW}_{\text{dry air}}$ represents the molecular weight of dry air, Ps represents the surface pressure, and τ represents time. We calculated $XCO_{2,EC}$ at hourly time steps over a period from when the local solar zenith angle crosses 70 degrees in the morning and afternoon. The seasonal cycle of the within-day variation in XCO_2 observed by the TCCON instrument agrees well with the seasonal cycle of the expected within-day variation in XCO_2 from NEE observations (R2 of 0.8; Figure 2.2a). The magnitude of the error bars derived from NEE, which represent the standard deviation among individual days, are substantially smaller than the magnitude of the error bars derived from the TCCON XCO_2 drawdown (Figure 2.2b). During winter, the average standard deviation for $XCO_{2,EC}$ is less than 0.1 ppm while the average standard deviation from XCO_2,FTS is about 0.4 ppm. In contrast, the average summer standard deviation is about 0.3 ppm for $XCO_{2,EC}$ and 1.2 ppm for XCO_2,FTS. Across seasons, the uncertainty from assuming a climatological within-day drawdown therefore reflects at most 30 percent of the total variability across the days on which observations are obtained. This suggests that most of the within-in day variation for XCO_2 results from processes other than local fluxes, confirming the motivation of the present study.

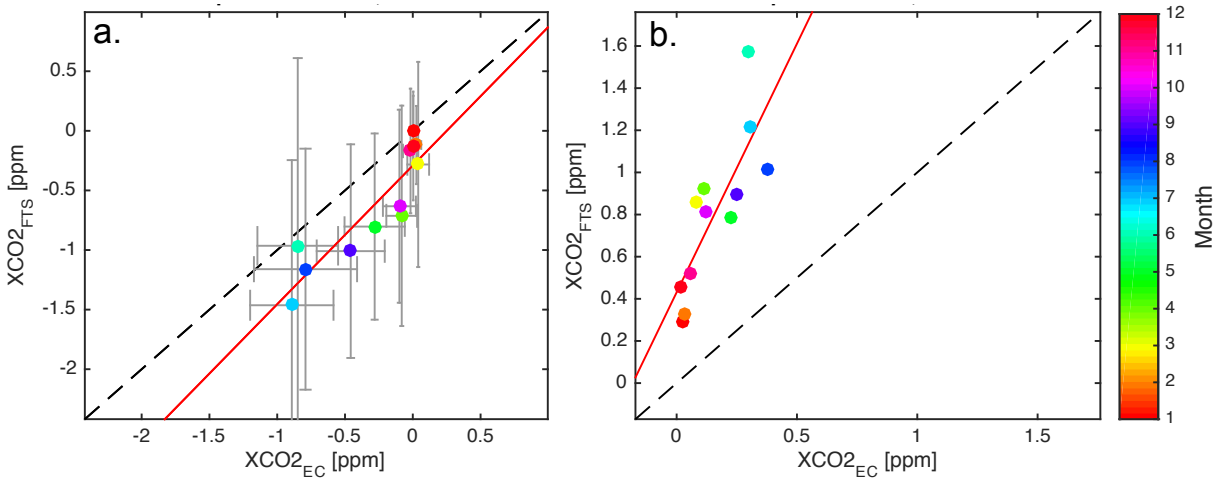

Figure 2.2: (a.) A comparison of the climatological monthly mean diurnal mean amplitude of XCO_2 observed at the TCCON site (XCO_2,FTS) compared to the estimated imprint of drawdown based on the observed net ecosystem exchange at the adjacent FluxNet eddy covariance tower in Park Falls, Wisconsin ($XCO_{2,EC}$). (b.) The standard deviation of monthly mean XCO_2,FTSS compared to the standard deviation of $XCO_{2,EC}$ plotted in panel a. Note that the axes in panel b. are different.

2.3.3 OCO-2

We analyzed spatial variations in XCO_2 and XH_2O retrieved from OCO-2 satellite observations. OCO-2 is a sun-synchronous, polar-orbiting satellite with a spatial footprint for individual soundings of 2.4 km along-track and 1.25 km cross-track; the instrument measures 8 cross-track bins at each time step resulting in a narrow (approximately 10 km wide) sampling swath (Eldering et al., 2017). The satellite acquires a repeat cycle of approximately every 16 days using three scanning modes, described below. The instrument comprises three grating spectrometers that measure radiances from reflected near-infrared sunlight in two CO_2 bands, the 1.61 μm weak absorption band (WCO_2) and the 2.06 μm strong CO_2 (SCO_2) absorption band, and in the 0.72 μm oxygen (O_2A) absorption band. These radiances are used in a full physics retrieval algorithm (version 8r, O'Dell et al., 2012; 2018), which uses optimal estimation to infer the vertical column of both CO_2 and O_2 while simultaneously adjusting other elements of the retrieval state vector, including the surface albedo for each band, aerosol optical depth (AOD), and other parameters that affect measured ra-

diances (O'Dell et al., 2012; 2018). The reported error for each XCO_2 sounding is estimated using instrument noise, and then post-processed to account for errors associated with the forward model used in the retrieval algorithm (O'Dell et al., 2012; Wunch et al., 2017). The measurements are bias-corrected by accounting for biases in individual cross-track observations using multivariate linear regression to identify physically unrealistic correlations between XCO_2 and other elements of the retrieval state vector (such as surface pressure, aerosols, or unphysical variations of the retrieved vertical profile of CO_2) and systematic offsets of OCO-2 XCO_2 target mode retrievals in comparison to TCCON observations (O'Dell et al., 2018). The (lower bound) bias-corrected single sounding errors for retrieved XCO_2 are generally less than 1 ppm (compared to a mean global value of approximately 410 ppm; Tans and Keeling, 2018), with the largest errors over land and higher latitudes (generally above 45º N or S) and smallest errors over the ocean (Eldering et al., 2017). Similarly, XH_2O was retrieved from OCO-2 using the 1.61 and 2.06 μm weak and strongly absorbing H2O spectral bands with mean biases of approximately 70 ppm, compared to typical XH_2O concentrations that varies from roughly 700 to 9000 ppm globally (Nelson et al., 2016).

OCO-2 uses three scanning modes to optimize retrievals over land and ocean surfaces, including nadir (land only), glint (over ocean and land), where the instrument is pointed at the glint angle to maximize reflected light over water surfaces, and target where the instrument angle is adjusted to point towards a targeted location (typically a ground-based validation site). In this analysis, we investigate nadir and glint observations separately, and only used soundings without a quality warning flag (Osterman et al., 2018). Note that many tracks exhibit significant missing data because of cloud cover.

2.3.3.1 Geostatistical analysis

We used geostatistical analysis to quantify the variance budget for OCO-2 data. We removed low frequency variations using a 250 km Hamming high-pass filter. To apply the filter, the data were pre-processed by averaging up to 8 cross-track soundings into 1.1 km bins in nadir mode or 1.3 km bins in glint mode to create a 1-dimensional track. We gap-filled empty bins with a distance-

weighted mean of the nearest filled bin. For each 10º by 10º box, we filtered tracks containing valid observations in at least 96 bins in glint mode or 113 bins in nadir mode (i.e. one-half of the rolling window filter size). To minimize edge effects on the high-pass filter, we attached a 250-point buffer made up of the average of the first 250 bins (i.e. the length of variability passed through the high pass filter) to the beginning and end of each satellite track. After running the filter, we repopulated each sounding with the filtered bin-averaged and gap-filled values and began our semivariogram analysis described below.

We separated variance of XCO_2 and XH_2O along OCO-2 tracks into random errors ("unexplained variance") and the component that is spatially coherent, or systematic, ("explained variance") by calculating semivariograms for the high frequency component of XH_2O and XCO_2. We calculated the semivariance (γ^*) for lag d at position xk for sounding values Z using Equation 2.8,

$$\gamma^*(d) = \frac{1}{2N(d)} \sum_{k=1}^{N} [Z(\vec{x}_k) - Z(\vec{x}_k + d)]^2 \tag{2.8}$$

where N is the number of soundings separated by lag d (Cressie and Hawkins, 1980). We fit a spherical model (Equation 2.9) to estimate the total variance, c_∞, and the spatial range of total variance, denoted as a_{space} (as in Equation 2.5), for each semivariogram (Figure **??**). For the spherical model fits, we fixed the unexplained variance, c_0, to the semivariance calculated from the observations at the smallest observed lag (1.1 km in nadir or 1.3 km in glint mode).

$$\gamma(d) = \begin{cases} c_0 + (c_\infty - c_0)\left[\frac{3d}{2a} - \frac{1d^3}{2a^3}\right] & \text{for } d \leq a \\ c_\infty & \text{for } d > a \end{cases} \tag{2.9}$$

We calculated the explained variance, denoted as $\langle c'_{space} \rangle$, by subtracting the unexplained variance from the total variance, $c_\infty - c_0$. In this framework, the explained variance relates to spatially coherent patterns, which could be due to real atmospheric gradients owing to fine scale transport or errors arising from spatially coherent correlations between XH_2O and XCO_2 and other elements of the state vector. We compared the square root of unexplained and explained variances, denoted as unexplained and explained variability, to temporal variations observed at adjacent TCCON sites,

described in more detail in Section 2.2.

2.3.3.2 North-south gradient calculation

To investigate the mesoscale tracer transport term on the RHS of Equation 2.5, we quantified the relationship between fine-scale spatial variations and the large-scale gradient in XH_2O and XCO_2. We calculated the North-South (N-S) gradient from three different datasets. For OCO-2, we aggregated data within a 10º by 10º box centered at the TCCON sites listed in Table 2.1. We calculated the gradient for each track within the targeted domain by fitting OCO-2 soundings to a simple least squares linear regression model weighted by self-reported errors from the version 8 level 2 retrieval algorithm. Because OCO-2 tracks may have data gaps associated with seasonal variations or cloud cover, we filtered the north-south gradients by quantifying the uncertainty (σ^2_{NS}) of the estimated N-S gradient using Equation 2.8 (Glover et al., 2011), where x_i represents the latitude and σ_i is the OCO-2 reported retrieval error at point i for N total soundings. We then discarded regression fits that had an uncertainty larger than 0.01 ppm/degree.

$$\sigma^2_{NS} = \frac{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \cdot \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2})^2} \tag{2.10}$$

We compared monthly mean observed N-S gradients from OCO-2 to two additional datasets: the monthly mean N-S gradients derived from assimilated 2017 CarbonTracker (CT2017) output from the OCO-2 period (from 2014-2017, with observations ongoing) and the N-S gradients inferred from the High-Performance Instrumented Airborne Platform for Environmental Research (HIAPER) Pole-to-Pole Observations (HIPPO) flight transects over the Pacific Ocean that took place between 2009 and 2011. CarbonTracker is a data assimilation system that provides three-dimensional atmospheric CO_2 fields based on assimilating surface CO_2 observations from NOAA's cooperative sampling network (Peters et al., 2007; with updates documented at http://carbontracker.noaa.gov). XCO_2 was computed in their 2017 (CT2017) dataset with simple pressure-weighted vertical integration of CO_2. During the HIPPO campaign, partial columns of

CO_2 were measured from roughly 300 to 8500 m altitude from aircraft transects spanning from 67ºS to 85ºN across the Pacific Ocean during all seasons between 2009 and 2011 (Wofsy et al., 2011). XCO_2 was then inferred by integrating a pressure-weighted mean concentration using reference static pressure from the GV Paroscientific Model 1000 sensor (Wofsy et al., 2017). We did not apply averaging kernels to either the CT2017 or HIPPO data since we were not attempting to directly compare individual columns to their OCO-2 or TCCON counterparts, but rather to approximate the large-scale features in XCO_2.

With sufficient data density, the N-S gradients derived from OCO-2 overpasses were broadly consistent with CT2017 output and HIPPO transects. However, when satellite data were characterized by gaps or low coverage during the winter season, the satellite estimate of the N-S gradient was inconsistent with HIPPO and CarbonTracker. Given this pattern of agreement and the need for year-round N-S gradient information, we used the CarbonTracker gradient to quantify monthly-mean N-S gradients and to evaluate the impact of the gradient on temporal synoptic-scale and mesoscale variability and along-track high frequency explained variability.

2.4 Results

2.4.1 Temporal Variations at TCCON

2.4.1.1 Flux-driven diurnal variations

Local diurnal fluxes account for up to 1 to 2.0 ppm of within-day temporal variability of XCO_2 during the growing season, with the largest diurnal signal observed during boreal summer (Figure 2.3a). For example, in Lamont, Oklahoma, local ecosystem drawdown contributed a decrease of XCO_2 of 1.1 ppm between 7:00 am to 5:30 pm LST (local standard time) in July, whereas it showed almost no change throughout the day (10 am to 2 pm) during winter months (Figure **??**), as expected given the relatively dormant winter biosphere. At most midlatitude TCCON sites, local diurnal fluxes of XCO_2 accounted for less than 0.3 ppm of within-day variability during the winter

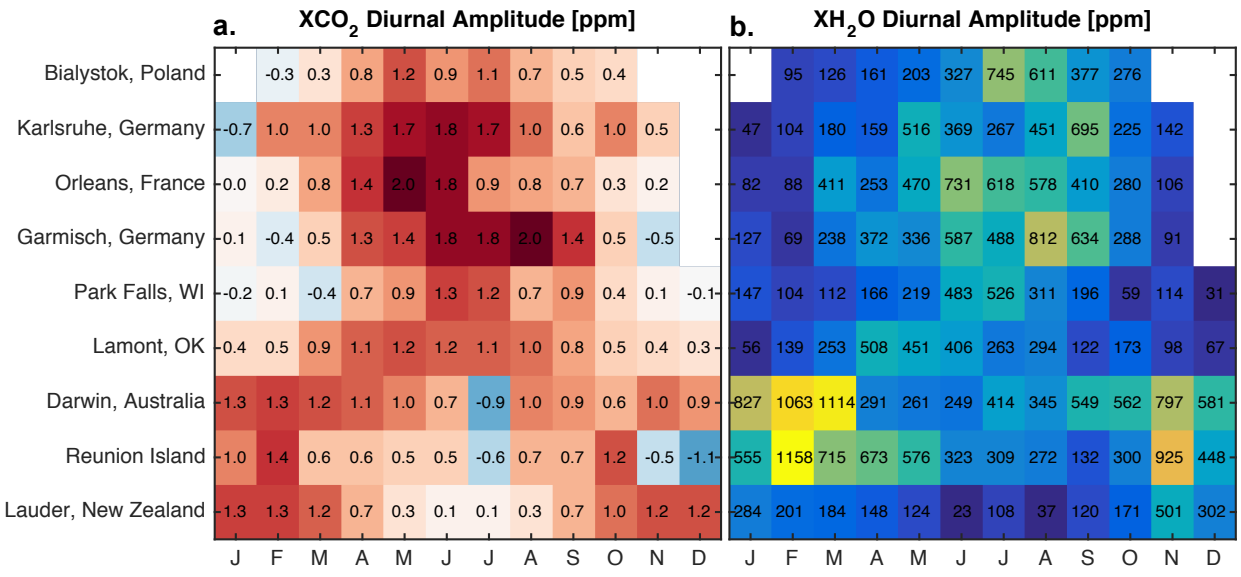

Figure 2.3: Climatological daytime diurnal range of a. $XCO_2$and b. XH_2O . We calculate the range between morning and evening, with a limit of solar zenith angle less than 75º.

(Figure 2.3a). The seasonal cycle of XCO_2 variability driven by diurnal fluxes at tropical TCCON sites, such as Darwin, Australia, was tied to the onset of the wet and dry seasons and varied from 0.1 ppm and 1.1 ppm (Figure 2.3a). We note that these are typical within-day variations of XCO_2 attributed to diurnal fluxes, and that the actual diurnal fluxes depend on weather, anthropogenic, and other natural interactions.

The climatological diurnal cycle of XH_2O had a strong seasonal cycle across all TCCON sites, with defined wet and dry seasons in the tropics, and lower winter and higher summer peak-to-trough diurnal cycle amplitudes in the midlatitudes (Figure 2.3b). In the Northern Hemisphere midlatitudes, the within-day local imprint was maximum in the summer (around 200-600 ppm) and smallest during boreal winter (around 5-100 ppm). Within-day, flux-driven variations were largest at the two tropical TCCON sites, which are both located in the Southern Hemisphere tropics. Within-day variations in these regions could exceed 500 ppm during austral summer but were generally 300-500 ppm during austral winter (Figure 2.3b). We note that while many atmospheric processes are analogous for XCO_2 and XH_2O, condensation and precipitation drive additional spatial and temporal variability in XH_2O (Dai and Wang, 2002). The values we report in Figure

2.3b are the peak-to-trough difference in within-day XH_2O climatology. During summer, most TCCON sites showed a maximum value of XH_2O in mid-to-late afternoon (1400 to 1700h, Figure **??**), consistent with the diurnal phasing of precipitable water reported by Dai and Wang (2002).

2.4.1.2 Synoptic-scale (bi-weekly) variability

Mean temporal synoptic-scale (bi-weekly) variations in both XH_2O and XCO_2 from TCCON were larger in magnitude to the typical daily cycle (Figure 2.3 and 4). For XH_2O, synoptic-scale variations were, on average, 4 times larger than variability attributed to diurnal fluxes (Figure 2.3 and 4). For some months, synoptic-scale variability of XH_2O was over 10 times larger than the magnitude of the imprint of local diurnal fluxes. For XCO_2 mean synoptic-scale variations were approximately twice as large local flux-driven variability, although for both gases, the differences varied seasonally. These variations in both XH_2O and XCO_2 were also tied to the seasonal cycle in the large-scale N-S gradient (Table **??**). As described in Section 2.2.1, we quantified temporal synoptic-scale variability by taking the standard deviation of the bi-weekly residual in XH_2O and XCO_2 after accounting for the climatological peak-to-trough within-day signal at each TCCON site.

Temporal synoptic-scale variations of XH_2O across all TCCON sites were of order 100 to 1000 ppm with strong regional and seasonal dependence (Figure 2.4a). For example, we observe synoptic-scale variations in XH_2O of 200-400 ppm in Lamont and Park Falls during boreal winter and peak synoptic-scale variability of over 1000 ppm during the Northern Hemisphere summer (Figure 2.4a). At similar latitudinal regions in Europe, synoptic scale variability of XH_2O only varies from 150 to 800 ppm throughout the year. We acknowledge that on multi-week timescales, many processes other than atmospheric transport can alter the atmospheric water vapor mole fraction, including diabatic processes in the atmosphere. This complexity is evident in the different seasonal patterns in and magnitudes of bi-weekly variability at TCCON sites, which varies even within a given latitude band. We therefore present this analysis to parallel the XCO_2 analysis described below.

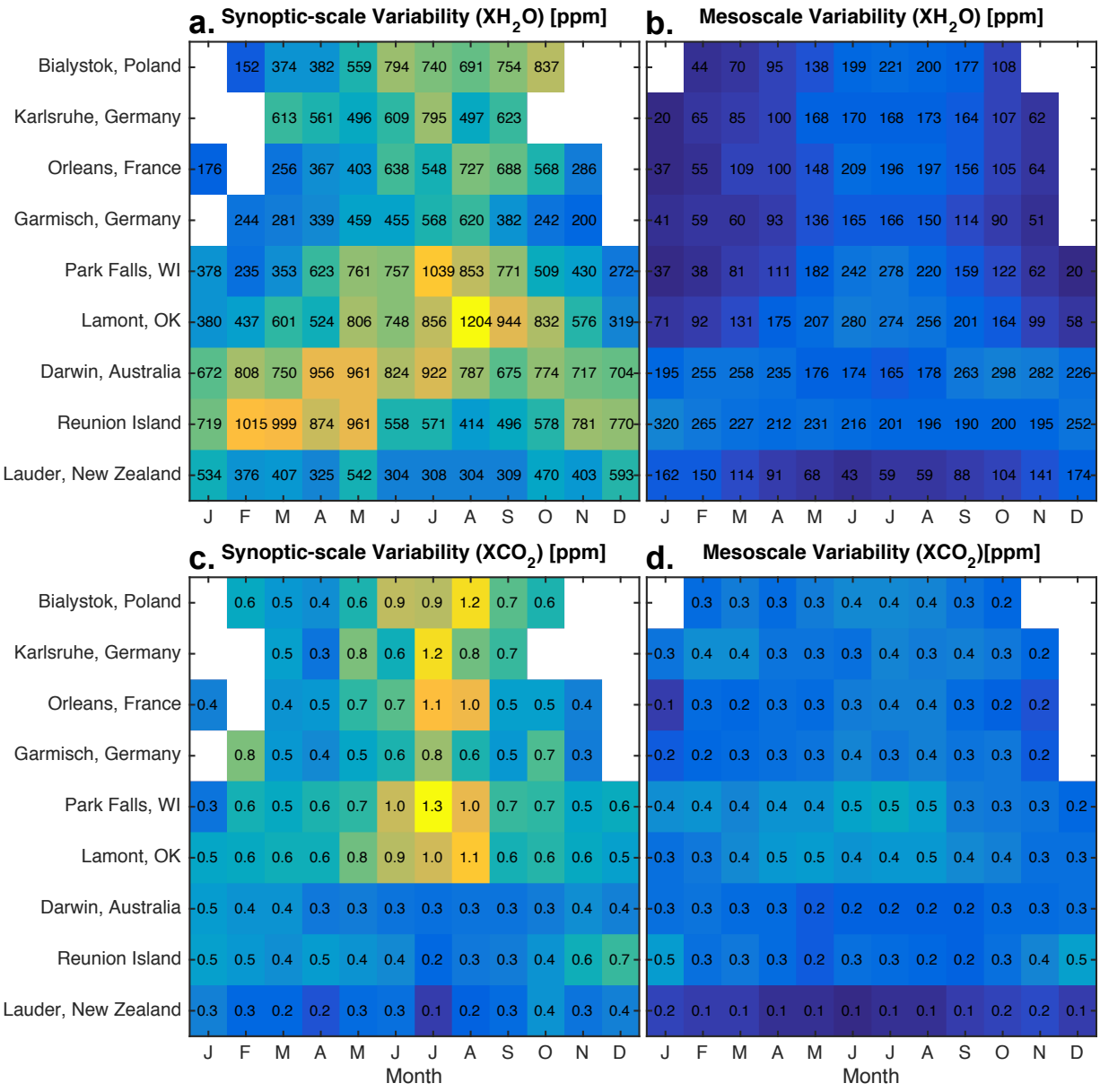

Figure 2.4: Comparison of synoptic-scale and variability in TCCON observations. (a.) XH_2O monthly mean synoptic-scale (bi-weekly) variability, (b.) XH_2O mesoscale (within-day) variability, (c.) XCO_2 monthly mean synoptic-scale variability, and (d.) XCO_2 monthly mean mesoscale variability. Note that the color scales for XH_2O and XCO_2 are different.

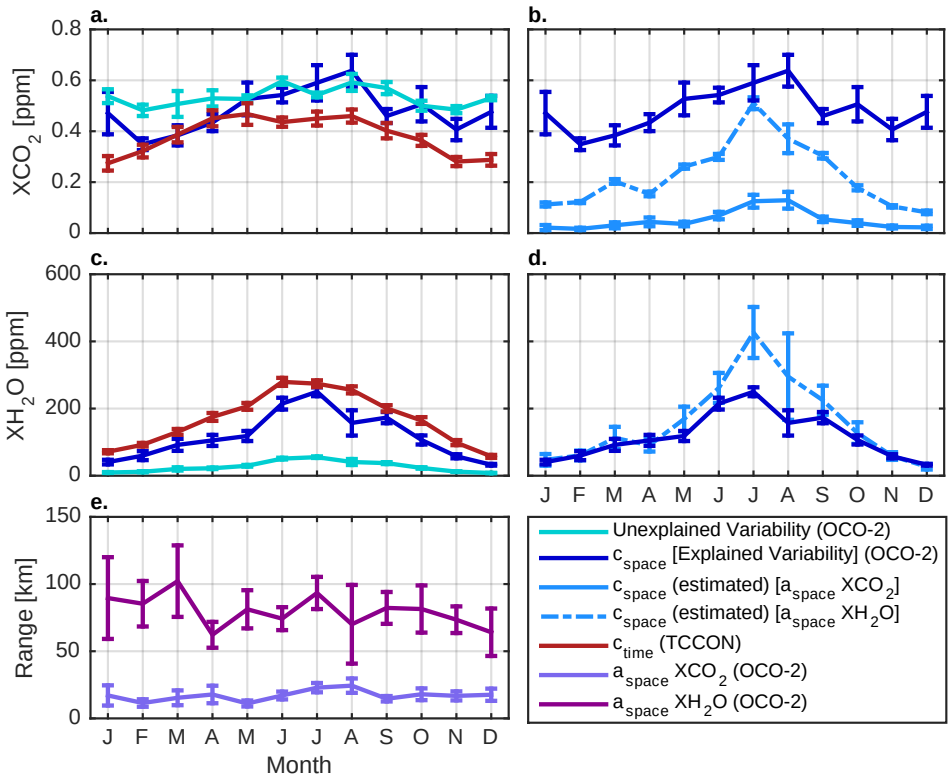

Figure 2.5: A comparison of monthly mean TCCON temporal mesoscale $\langle c' \rangle_{time}$ (within-day) variability (red), OCO-2 observed along-track high frequency unexplained (teal) and explained ($\langle c'_{space} \rangle$) (dark blue) spatial variability, and explained variability ($\langle c'_{space} \rangle$) estimated from Equation 2.12 (light blue) using ranges (a_{space}) observed by OCO-2 for XH_2O (dashed line) and XCO_2 (solid line). Panels (a.-b.) shows variability for $XCO_2$and panels (c.-d.) XH_2O in Lamont, Oklahoma. The bottom panel, e., represents the range (a_{space}) of along-track high frequency explained spatial variability of XH_2O (dark purple) and XCO_2 (light purple). The error bars for observed parameters represent the standard error. For estimated $\langle c'_{space} \rangle$, the error bars represent the error propagation using climatological monthly mean standard error of each observed parameter.

For XCO_2, typical synoptic-scale variations ranged from 0.1 to 1.3 ppm across all TCCON sites, with the largest variations (in excess of 1 ppm) observed over Northern Hemisphere TCCON sites during July and August (Figure 2.4c). These locations also had the largest seasonal cycles of synoptic-scale variability (Figure 2.4c). There was less pronounced synoptic-scale variability of XCO_2 in the tropics and the southern hemisphere, where synoptic-scale variability ranged from 0.1 in the boreal summer to 0.7 ppm in boreal winter (Figure 2.4c). The magnitude and seasonality of these variations are mostly tied to the meridional (N-S) gradient in XCO_2, as we discuss below.

Synoptic-scale variations of XCO_2 were correlated with the magnitude of the N-S gradient at northern hemisphere midlatitude TCCON sites (Table **??**). As described in Section 2.3.2, we fit a regression slope to estimate the absolute value of N-S gradients of XCO_2 derived from OCO-2 overpasses to those derived from CT2017 output and HIPPO transects (Figure **??**). In the Northern Hemisphere midlatitudes, the correlation between monthly mean temporal synoptic-scale variations observed from TCCON (Fig 4c.) and the monthly mean N-S gradient (Figure **??**-S6) was generally around 0.6 to 0.9 (Table **??**). This relationship is consistent with the argument that synoptic variations arise from transient eddies acting on the large-scale gradient, as indicated by the second RHS term in Equation 2.5. The weaker correlation (R=0.41) observed at Garmisch, Germany was an outlier among Northern Hemisphere midlatitude TCCON sites, which may be due to limited observations. In the tropics and Southern Hemisphere, the relative absence of N-S gradients of XCO_2 resulted in weak relationship with synoptic-scale variations. We note that the length-scales estimated from many of the slopes of our best-fit linear regressions between synoptic scale variability and the N-S gradient of XCO_2 (2.1º to 6.5º; or roughly 200 to 600 km at Northern Hemisphere midlatitude TCCON sites) are on the smaller end of those of typical synoptic-scale systems.

The temporal synoptic variability of XH_2O correlates with the N-S gradient of XH_2O across most TCCON sites (Table **??**). The correlation coefficients were between 0.4 to 0.9 across TCCON sites. The highest correlation coefficients were observed at mid-latitude TCCON sites (Bialystok, Orleans, Park Falls, and Lamont), and at these sites, the slope of the relationship was consistently

between 7 and 9 ppm/(ppm/degree) (which can be written as a length scale, degree). This is approximately consistent with typical length scales of synoptic-scale variability of the order of magnitude of 1000 km and larger than that estimated for XCO_2.

2.4.1.3 Mesoscale (within-day) variability

Mesoscale variations (Figure 2.4b, 5c) of XH_2O are, on average, a factor of 1.4 times larger in magnitude to variability attributed to local diurnal fluxes (Figure 2.3b). For XH_2O, mesoscale variations were generally a factor of 5 smaller than synoptic-scale variability across all TCCON sites (Figure 2.4b). As expected based on synoptic scale variations in XH_2O, patterns of mesoscale variation showed strong regional variations. For example, TCCON sites in North America (Park Falls and Lamont), mesoscale variations of XH_2O were almost 300 ppm during the boreal summer, but less than 100 ppm during the boreal winter (Figure 2.4b, 5c). In contrast, at TCCON sites at similar latitudes in Europe, mesoscale variations of XH_2O were generally less than 200 ppm all year. At tropical TCCON sites, such as Darwin and Reunion Island, mesoscale variations were between 160 and 320 ppm all year round, with lower ($<$200 ppm) mesoscale variations occurring during the dry season (Figure 2.4b).

Mesoscale variations of XH_2O, like synoptic variations, were generally correlated with the N-S gradient of XH_2O across all midlatitude TCCON sites (R values of 0.4 to 0.9 at TCCON sites with statistically significant slopes; Table **??**). The slopes, however, were much lower, typically around 2º. We expect that these correlations do not necessarily suggest that the large-scale N-S gradient drives mesoscale variations, but rather to the fact that both quantities change seasonally and have strong temperature dependence via the Clausius-Clapeyron relationship.

Temporal mesoscale variability in XCO_2, which we assume is primarily driven by advection from small-scale weather features, was less than 0.5 ppm across all TCCON sites and all months (Figure 2.4, 5). This represents about half the magnitude of variability attributed to local diurnal fluxes during the growing season at northern hemisphere TCCON sites (Figure 2.4). In the winter, mesoscale variations of XCO_2 become larger than the imprint of variability attributed to local

diurnal fluxes. In Park Falls, Wisconsin, the combined imprint of mesoscale and synoptic scale transport was 1 to 2 ppm during summer (Figure 2.4), substantially larger than the potential bias from assuming climatological fluxes of about 0.3 ppm (Figure 2.2). These mesoscale variations were approximately 30 to 50 percent magnitude of synoptic-scale variability (Figure 2.4c and 4d). Mesoscale variations in XCO_2were only moderately correlated with N-S gradients (R-values less than 0.52) in the mid-to-high latitudes in both the northern and southern hemispheres (Table **??**), consistent with our expectation that the length scale of mesoscale systems is small in comparison to the length scale of the N-S gradient. In the tropics, there was likewise no correlation between mesoscale variability and the N-S gradient of XCO_2 (Table **??**).

We calculated a typical timescale for mesoscale (within-day) variations based on the autocorrelation of within-day residuals with climatological local fluxes removed (Figure S7). We found that the autocorrelation of the residuals typically decayed to values between e-1 and 0 over about 3 hours. This timescale was consistent for both XH_2O and XCO_2 across all TCCON sites. We therefore used this mean lag time as time in Equation 2.5 to compare temporal variations to spatial variations of XH_2O and XCO_2 (Section 3.2.1)

Location	XH_2O						XCO_2					
	Synoptic (TCCON)		**Mesoscale (TCCON)**		**Explained (OCO-2)**		**Synoptic (TCCON)**		**Mesoscale (TCCON)**		**Explained (OCO-2)**	
	Slope [deg]	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**
Bialystok, Poland	**7.4 ± 3.1**	**0.75**	**2.2 ± 0.5**	**0.92**	**1.3 ± 0.6**	**0.80**	**6.5 ± 5.2**	**0.44**	0.6 ± 1.6	0.06	-0.7 ± 4.0	0.01
Karlsruhe, Germany	1.5 ± 3.7	0.12	**1.8 ± 0.9**	**0.58**	**1.9 ± 1.3**	**0.64**	**2.9 ± 2.6**	**0.51**	0.4 ± 0.5	0.08	-0.0 ± 0.7	0.00
Orleans, France	**7.6 ± 3.4**	**0.69**	**1.9 ± 1.2**	**0.50**	0.7 ± 1.6	0.11	**3.1 ± 1.2**	**0.76**	0.7 ± 0.4	0.52	-0.4 ± 0.9	0.13
Garmisch, Germany	3.5 ± 3.7	0.28	**1.4 ± 1.0**	**0.42**	**2.1 ± 1.0**	**0.72**	0.8 ± 1.2	0.17	**0.4 ± 0.3**	**0.29**	-0.3 ± 1.4	0.02
Park Falls, WI	**7.5 ± 2.2**	**0.79**	**2.7 ± 0.6**	**0.87**	**1.5 ± 1.2**	**0.41**	**3.6 ± 1.6**	**0.62**	**0.9 ± 0.8**	**0.30**	0.3 ± 1.2	0.04
Lamont, OK	**8.7 ± 6.9**	**0.35**	1.5 ± 2.4	0.12	1.1 ± 2.2	0.08	**2.1 ± 0.9**	**0.67**	0.5 ± 0.5	0.32	**0.8 ± 0.4**	**0.49**
Darwin, Australia	**0.8 ± 0.5**	**0.46**	**0.3 ± 0.2**	**0.39**	**-0.4 ± 0.3**	**0.36**	**1.1 ± 0.7**	**0.41**	0.4 ± 0.4	0.27	0.3 ± 1.0	0.04
Reunion Island	4.6 ± 6.2	0.15	0.8 ± 1.2	0.12	1.1 ± 1.2	0.23	-1.0 ± 3.7	0.02	-0.7 ± 3.0	0.02	0.4 ± 1.1	0.04
Lauder, New Zealand	**4.5 ± 3.7**	**0.33**	**2.8 ± 1.1**	**0.69**	**0.7 ± 0.6**	**0.58**	4.1 ± 9.5	0.06	2.5 ± 3.6	0.14	0.2 ± 9.9	0.00

Table 2.2: Regression statistics for magnitude of variability in trace gases versus their N-S gradient. We report the slope (± 95 percent confidence interval) in units of degrees and the correlation coefficient, R, of a best-fit linear regression line for the N-S gradient of XH_2O and XCO_2 output from CT2017 compared to bi-weekly (synoptic-scale) and within-day (mesoscale) temporal variability of XH_2O observed by TCCON, and explained along-track high frequency spatial variability observed by OCO-2. Bolded slopes and regressions indicate statistically significant fits at the 95 percent confidence interval.

2.4.2 OCO-2 along-track spatial variability

This section compares the relationship between the high-frequency along-track spatial variability observed by OCO-2 to mesoscale temporal variability of XH_2O and XCO_2 from TCCON using the theoretical framework outlined in Section 2.1. Variations in both XH_2O and XCO_2 evolve in response to local surface fluxes and atmospheric transport, and for XH_2O, condensation, evaporation, and precipitation within the atmosphere. While the details of the surface fluxes and in situ atmospheric processes differ for the two tracers, they experience the same atmospheric advection and mixing fields. The advantage of a joint analysis of these two gases is that both XH_2O and XCO_2 are observed simultaneously by TCCON and OCO-2, and the precision of XH_2O is substantially larger, providing a framework for assessing the XCO_2 results. Specifically, if the calculated and observed explained variability of either species, $\langle c'_{space} \rangle$, are in agreement, then we can assume that mesoscale atmospheric transport is the dominant source of high-frequency variability of that gas.

2.4.2.1 Explained Variability ($\langle c'_{space} \rangle$)

Along-track, high frequency (<250 km) explained spatial variations of XH_2O from OCO-2 spanned between 20 and 300 ppm across all TCCON sites (Figure 2.6b). The smallest explained variations (20 to 60 ppm) were observed at Northern Hemisphere midlatitude sites during the boreal winter. The largest explained variations (> 200 ppm) occurred over most Northern Hemisphere midlatitude sites during the boreal summer and over the tropical sites (Darwin and Reunion Island) during the local wet season. Across all months and TCCON sites, the spatial range (a_{space}) of explained variability generally spanned between 40 to 140 km (Figure 2.7b). The explained high-frequency spatial variability of XCO_2 was generally between 0.2 and 1.0 ppm across all TCCON sites (Figure 2.6e). The highest explained variations (>0.5 ppm) were observed over Northern Hemisphere TCCON sites. Smaller explained variations (0.2 to 0.5 ppm) were observed at southern hemisphere TCCON sites. In contrast to XH_2O, the a_{space} for XCO_2 occurred at much smaller spatial scales from 10 to 40 km, with mean a_{space} values for explained variability around 20 km (Figure 2.7a).

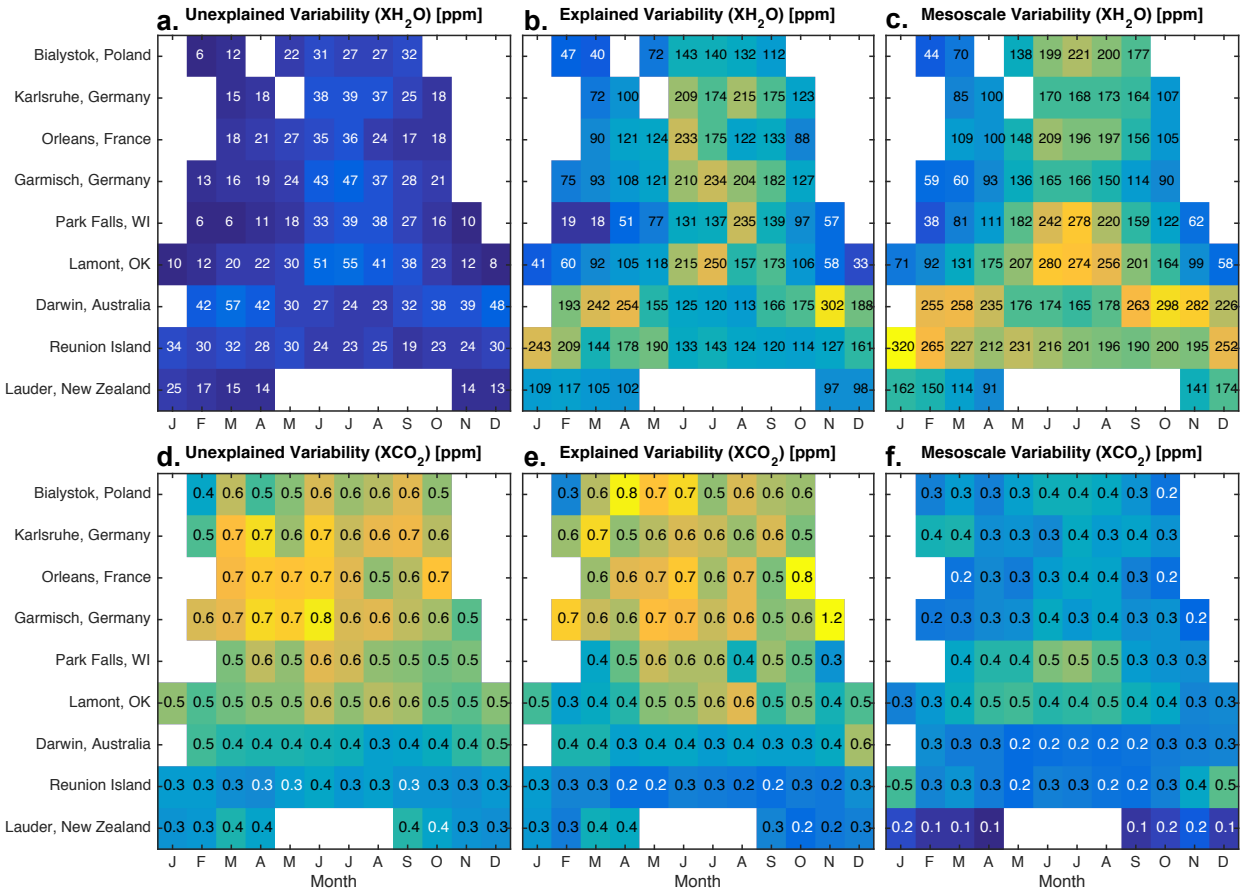

Figure 2.6: Variability metrics for remotely sensed XH_2O and XCO_2. Monthly mean (a.) unexplained variability for XH_2O and (b.) explained variability for XH_2O derived from high-pass filtered OCO-2 observations. c. Monthly mean temporal mesoscale (within-day) variability in XH_2O derived from TCCON observations. Panels d-f are similar, except we show values for XCO_2 from OCO-2 (d-e) and TCCON (f) observations.

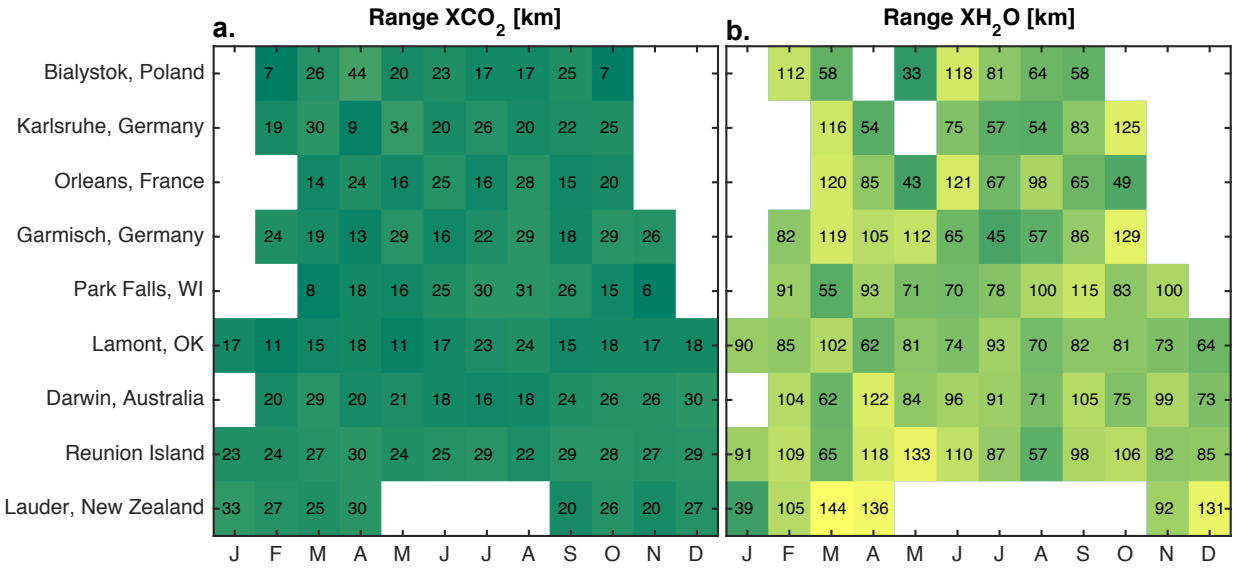

Figure 2.7: Monthly mean spatial range (a_{space}) of a. XH_2O and b. XCO_2derived from OCO-2 data near each TCCON site. Note both panels use the same color scale.

We can put the explained spatial variability of both XH_2O and XCO_2 along OCO-2 tracks ($\langle c'_{space} \rangle$) by comparing variations in each gas with the corresponding mesoscale variations at TCCON sites (Section 2.3.1) using Equation 2.5 to relate temporal variations at a given location to tracer anomaly advection and turbulent production. We expect the time tendency (Term A in Equation 2.9) to be balanced by the advection of tracer variations by the mean wind (Term B) or by small-scale production of variation by eddies acting on the mean gradient (Term C). Take, for example, a typical temporal mesoscale variation of XH_2O, $\langle c'_{time} \rangle$, of about 280 ppm (as is approximately the case in Lamont, Oklahoma in June; Figs. 4-6). We computed the following scales:

$$
\begin{aligned}
&\text{(A)}\ \frac{c'_{time}}{\tau_{time}} = \frac{280\ \text{ppm}}{3\ \text{hours}} = 90 \frac{\text{ppm}}{\text{hour}} \\
&\text{(B)}\ \overline{u} \frac{c'_{space}}{a_{space}} = 8 \left(\frac{\text{m}}{\text{s}} \right) \frac{280\ \text{ppm}}{70\ \text{km}} = 90 \frac{\text{ppm}}{\text{hour}} \\
&\text{(C)}\ u' \frac{\partial \overline{c}}{\partial x} = 1 \left(\frac{\text{m}}{\text{s}} \right) \frac{280\ \text{ppm}}{140\ \text{degrees}} = 5 \frac{\text{ppm}}{\text{hour}}
\end{aligned}
\tag{2.11}
$$

We used a within-day period (time) of 3 hours (Section 3.1.3; Figure S7). For Lamont in June, we computed total column pressure and H2O vertical profile weighted mean horizontal wind speed ($\overline{u}$) of 8 m/s and mesoscale variability of horizontal winds (u') of 1 m/s by vertically integrating wind output from CT2017 (Figure S8). Since CarbonTracker is run at relatively coarse resolution, we compare these values against calculations derived from the North American Regional Reanalysis (NARR, Mesinger et al., 2006), which provides wind fields at 0.3 degrees resolution. At the two North American TCCON sites, u was the same as that estimated from CarbonTracker fields, while u' was larger by a factor of two (about 2 m/s), an expected result given the higher horizontal resolution of NARR. Nevertheless, when we apply this larger u' estimate into Equation 2.11, our scale analysis remains unchanged. We also used CT2017 output to calculate a typical summertime Northern Hemisphere midlatitude N-S gradient of around 140 ppm/degree (Figure **??**). The scaling exercise suggests that the time tendency is mostly balanced by the mesoscale anomaly advection term (Term B) rather than the turbulent production term (Term C), at least for these time-scales. The small contribution of the production term is consistent with the fact that mesoscale variations at TCCON sites were not highly correlated with the mean N-S gradient. This analysis can also be applied to relate spatial and temporal mesoscale variations in XCO_2or another atmospheric tracer.

Based on the scaling arguments above, we rearrange Equation 2.5 to solve for the expected $\langle c' \rangle_{space}$ along OCO-2 tracks and neglect the turbulent production term (Equation 2.12):

$$\langle c' \rangle_{space} = \frac{a_{space}}{\overline{u}} \left(\frac{\langle c' \rangle_{time}}{\tau_{time}} - \langle u' \rangle \langle \frac{\partial \overline{c}}{\partial x} \rangle \right) \approx \frac{a_{space}}{\overline{u}} \left(\frac{\langle c' \rangle_{time}}{\tau_{time}} \right) \tag{2.12}$$

We then use Equation 2.12 to estimate the expected magnitude of $\langle c' \rangle_{space}$ for XH_2O in Lamont in June as approximately 226 ppm which agrees within 10 percent with the observed explained variability of 214 ppm from OCO-2 (Figure 2.6b and 8b). We found that if we applied Equation 2.12 to compute $\langle c' \rangle_{space}$ at each TCCON site and each month, our estimated $\langle c' \rangle_{space}$ values match observed explained variations from OCO-2 to within 30 percent (Figure 2.8a and b). The agreement between the observed and estimated $\langle c' \rangle_{space}$ of XH_2O suggests that the observed explained variations of XH_2O from OCO-2 are primarily driven by mesoscale atmospheric (Figure 2.5 and

2.8a,b). These results also validate our choice of a 250 km high-pass filter to isolate mesoscale transport and exclude larger-scale synoptic systems.

We likewise calculated the $\langle c' \rangle_{space}$ of XCO_2 in Lamont in June using the estimated total column pressure and CO_2 vertical profile weighted mean horizontal wind speed (u) m/s and a spatial range, a_{space}, of 20 km fit from semivariogram analysis (Figure 2.7a). Using this framework, the calculated spatial mesoscale variability of 0.1 ppm was much smaller than the observed $\langle c' \rangle_{space}$ (0.6 ppm). The empirical a_{space} of 20 km for XCO_2 is suspect, since mesoscale variations in XH_2O from OCO-2 showed an a_{space} of 70 km in the same month and are consistent with the expected length-scale for mesoscale systems. When we instead used a_{space} of 70 km based on the analysis of XH_2O, estimated mesoscale variability increases to 0.3 ppm (Equation 2.13).

$$\langle c' \rangle_{space,XCO_2} = \frac{70 \text{ km}}{10 \text{ m/s}} \left(\frac{0.4 \text{ ppm}}{3 \text{ hours}} \right) \approx 0.3 \text{ ppm} \tag{2.13}$$

We note that an estimate of 0.3 ppm, while more reasonable in magnitude, is still about 40 percent smaller than the observed value of 0.5 ppm for the OCO-2 explained XCO_2 variability.

Together, these relationships suggest first that the temporal and spatial scaling within our framework is consistent with mesoscale variations of XH_2O quantified using TCCON and OCO-2 data. Second, the spatial range (a_{space}) for XCO_2 variability derived from the geostatistical analysis of OCO-2 data is too small to be driven by mesoscale systems. Third, the results suggest that the XCO_2 $\langle c' \rangle_{space}$ value is larger than what is calculated assuming mesoscale systems are the only driver of high frequency spatial variability along OCO-2 tracks. As shown for Lamont, Oklahoma, there is no overlap between the estimated and observed $\langle c' \rangle_{space}$ even when accounting for uncertainty in both terms (Figure 2.5b). We estimated the uncertainty $\langle c' \rangle_{space}$ using error propagation of the standard error of each of the measured terms in Equation 2.12. Although we only plot the uncertainty on the $\langle c' \rangle_{space}$ calculation for Lamont (Figure 2.5), we quantified uncertainty across all TCCON sites presented in the paper and this result is robust, meaning that the differences between the calculated $\langle c' \rangle_{space}$ (Figure 2.8c,d) and the observed $\langle c'_{space} \rangle$(Figure 2.8e) represent real

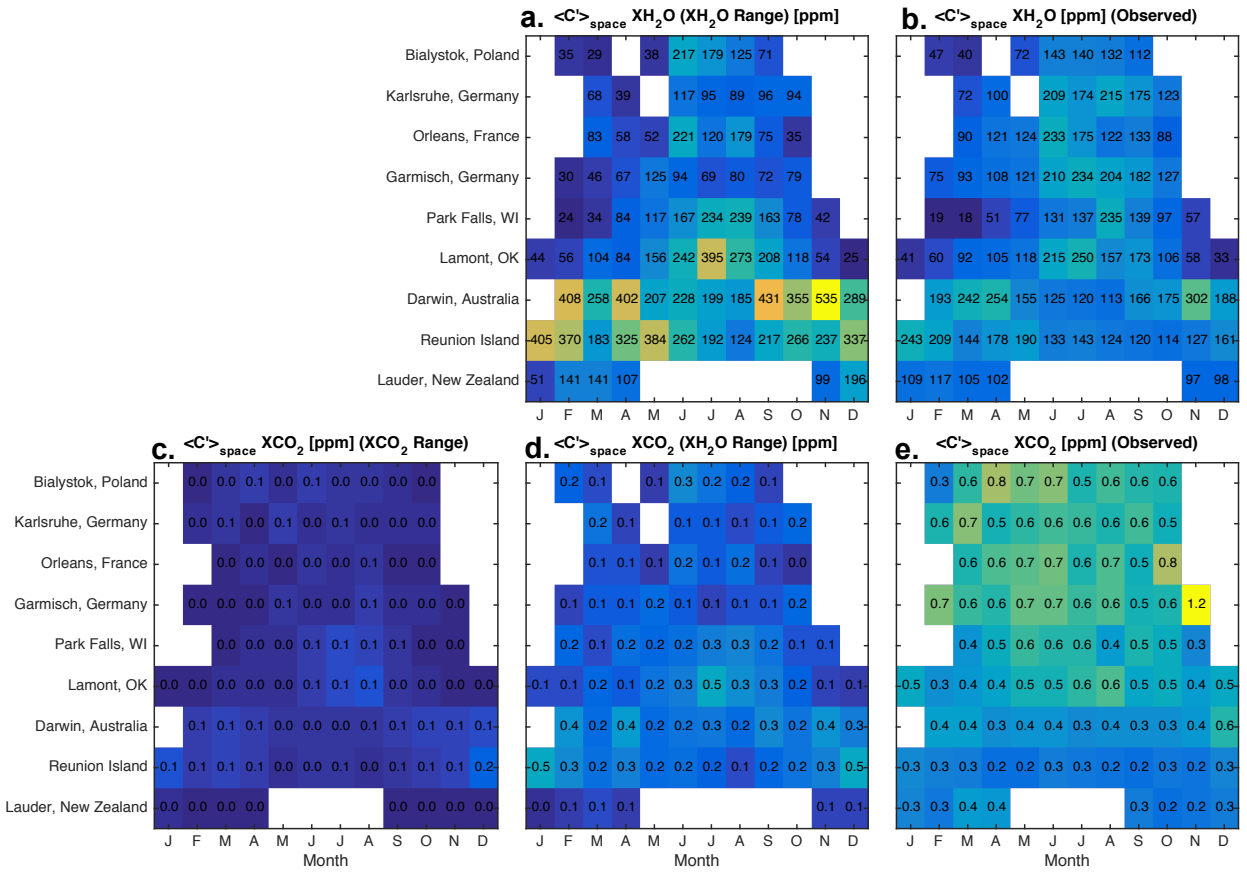

Figure 2.8: Comparison of calculated and observed $\langle c'_{space} \rangle$. The top row shows (a.) Calculated annual cycle in $\langle c' \rangle_{space}$ for XH_2O using Equation 2.12, assuming a_{space} values derived from XH_2O (Figure 2.7b) and $\langle c' \rangle_{time}$ values from TCCON mesoscale variations (Figure 2.4b). (b.) Observed annual cycle in $\langle c' \rangle_{space}$ for XH_2O. Note that this quantity is identical to the explained variability of XH_2O (Figure 2.6b). The bottom row shows the same quantities, except for XCO_2. (c.) Calculated annual cycle in $\langle c' \rangle_{space}$ for XCO_2 using Equation 2.12, assuming a_{space} values for XCO_2 (Figure 2.7a) and $\langle c' \rangle_{time}$ values from TCCON (Figure 2.4d). (d.) Calculated annual cycle in $\langle c'_{space} \rangle$, except we use a_{space} values from XH_2O. (e.) Observed annual cycle in $\langle c' \rangle_{space}$ for XCO_2 (identical to Figure 2.6e).

and significant disagreement. Thus, we conclude that some other factor imparts spatially coherent variability on OCO-2 XCO_2 that depresses the a_{space} and augments the $\langle c'_{space} \rangle$. One possibility is coherent biases or errors in the OCO-2 XCO_2 retrieval, as discussed more in section 3.3. These relationships are true across all TCCON sites (Figure 2.8c and 8e), where the observed XCO_2 $\langle c' \rangle_{space}$ was generally larger than what would be expected based on $\langle c' \rangle_{time}$ and the observed XCO_2 a_{space} was small relative to the value calculated from XH_2O observations (Figure 2.8a and 8b). We found that XCO_2 $\langle c' \rangle_{space}$ was more comparable to observed explained variations from OCO-2 when we used observed a_{space} values for XH_2O (Figure 2.8d and 8e).

We do not expect that the inflated XCO_2 $\langle c' \rangle_{space}$ values result from aliasing of synoptic scale variability. First, the water vapor analysis confirmed that our 250 km filter properly isolates mesoscale atmospheric transport. Second, the explained variability from OCO-2 was not correlated with the N-S gradients (Table **??**). If our $\langle c' \rangle_{space}$ values reflected synoptic scale variability, these quantities should be correlated.

2.4.2.2 Unexplained Variability

The unexplained high frequency spatial variability of XH_2O was up to 50 ppm (Figure 2.5 and 6), about 50 percent larger than the random errors reported by the v8 OCO-2 retrieval algorithm data product. These results were consistent with arguments from Connor et al. (2008) that reported random errors from the OCO-2 retrieval algorithm represent a lower bound on actual error. However, our estimate for unexplained variability along OCO-2 tracks was still less than 20 percent of the temporal and spatial mesoscale XH_2O variability (Figure 2.5 and 6), suggesting that the signal-to-noise ratio of XH_2O retrievals are large enough to observe mesoscale variations.

The unexplained spatial variability in XCO_2 was 0.3 to 0.8 ppm, which is the same order of magnitude as the spatial and temporal variations that may reflect mesoscale variations and is also generally consistent with random errors reported by the v8 OCO-2 retrieval algorithm data product (Figure 2.6d). These unexplained variations were consistent with the mean standard deviation of the cross-track soundings we averaged into each along-track bin (Section 2.3.2; Figure S9).

OCO-2 tracks adjacent to Southern Hemisphere TCCON sites typically had smaller unexplained variations (0.3-0.4 ppm), while tracks adjacent to Northern Hemisphere TCCON sites had slightly larger unexplained variations (0.5-0.7 ppm). This difference appears to arise due to the fact that the Southern Hemisphere and tropical OCO-2 tracks we analyzed contained ocean observations, whereas the Northern Hemisphere tracks contained only land observations. When we conducted semivariogram analysis around the latitude band between 40-50 degrees North, the unexplained variations over the ocean were approximately half the magnitude observed over land (Figure 2.9). Errors associated with retrieving XCO_2 over land, where topography and albedo can influence the XCO_2 retrieval, likely increase the unexplained variability. The unexplained variations over land did not show dependence on nadir versus glint observing mode. We note that the estimates for unexplained variability were not sensitive to the cutoff for the high-pass filter and were also robust when we explicitly fitted, rather than fixed, the unexplained variance in the spherical semivariogram model.

2.4.3 Spatially correlated variance from the state vector

Correlations between high frequency along-track spatial variations in XCO_2 and other elements of the OCO-2 retrieval state vector likely contributed to the larger than expected spatially coherent (explained) variability in OCO-2 XCO_2, compared to mesoscale variations at corresponding TCCON sites (Table **??**), and the smaller geostatistical spatial range (a_{space}) for XCO_2 than XH_2O. We therefore tested whether the total and high frequency along-track spatial variability of XCO_2 were correlated with other elements of the OCO-2 retrieval state vector. We selected aerosol optical depth (AOD) and albedo in the O_2 and weak and strong CO_2 absorption bands as variables that were likely to have spatial structures that, if correlated with XCO_2, could obscure spatially coherent transport patterns. The correlations between elements of the state vector and both the total variations and the high frequency along-track spatial variations of XCO_2 were small (R 0.2; Table **??**), but statistically significant. This analysis suggests that correlations between XCO_2 and these state vector elements may have depressed the apparent a_{space} and increased the explained variabil-

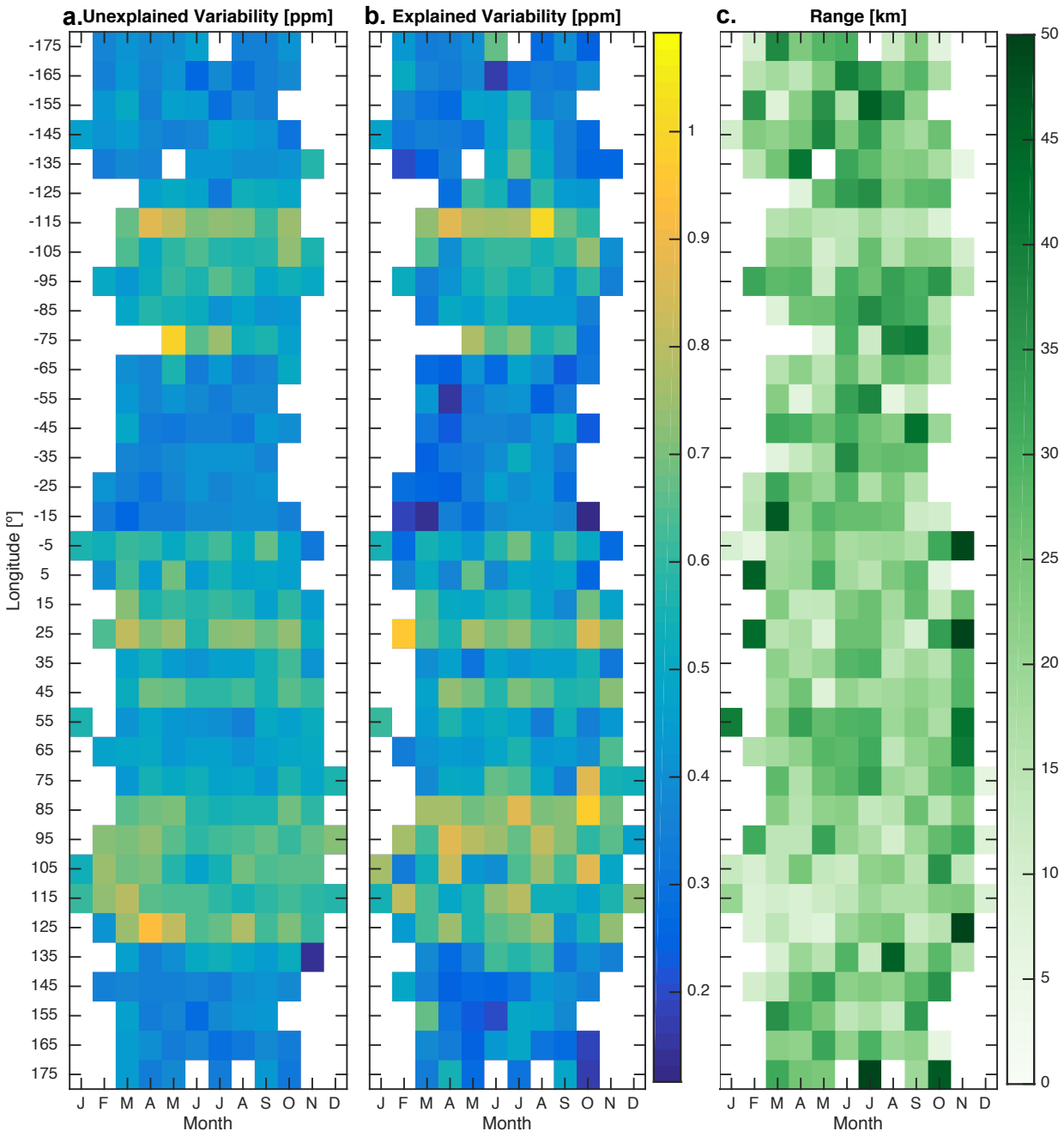

Figure 2.9: Global analysis of along-track high-frequency spatial (a.) unexplained variability, (b.) explained variability, and (c.) range of explained variability (a_{space}) of XCO_2 within 10°×10° grid cells across a latitudinal band centered at 45ºN. In general, grid cells over ocean show lower values of explained and unexplained error than those over land.

ity along OCO-2 tracks. We note that the analysis of albedo and AOD presented here was in no way exhaustive, but rather, our results show that these correlations were important contributors to the overall variance budget.

In contrast, high frequency variations in XH_2O were generally independent of other state vector elements. There were some weak to moderate, statistically significant correlations between the unfiltered XH_2O data and albedo or AOD (Table **??**). These correlations could represent real geophysical relationships among XH_2O, albedo, and AOD, but could also be attributed large-scale, coherent systematic biases in the retrieval of XH_2O that can be attributed to errors in AOD and albedo. When we high-pass filtered these variables, however, the correlations between the high frequency variations of XH_2O and albedo were not statistically significant, with the exception of AOD. This result suggests that at smaller spatial scales (less than 250 km), the variations in XH_2O were independent of the state vector elements we tested here and therefore the explained variability did not contain the imprint of spatially coherent biases

Location	XH_2O						XCO_2					
	Synoptic (TCCON)		**Mesoscale (TCCON)**		**Explained (OCO-2)**		**Synoptic (TCCON)**		**Mesoscale (TCCON)**		**Explained (OCO-2)**	
	Slope [deg]	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**	**Slope [deg]**	**R**
Bialystok, Poland	**7.4 ± 3.1**	**0.75**	**2.2 ± 0.5**	**0.92**	**1.3 ± 0.6**	**0.80**	**6.5 ± 5.2**	**0.44**	0.6 ± 1.6	0.06	-0.7 ± 4.0	0.01
Karlsruhe, Germany	1.5 ± 3.7	0.12	**1.8 ± 0.9**	**0.58**	**1.9 ± 1.3**	**0.64**	**2.9 ± 2.6**	**0.51**	0.4 ± 0.5	0.08	-0.0 ± 0.7	0.00
Orleans, France	**7.6 ± 3.4**	**0.69**	**1.9 ± 1.2**	**0.50**	0.7 ± 1.6	0.11	**3.1 ± 1.2**	**0.76**	0.7 ± 0.4	0.52	-0.4 ± 0.9	0.13
Garmisch, Germany	3.5 ± 3.7	0.28	**1.4 ± 1.0**	**0.42**	**2.1 ± 1.0**	**0.72**	0.8 ± 1.2	0.17	**0.4 ± 0.3**	**0.29**	-0.3 ± 1.4	0.02
Park Falls, WI	**7.5 ± 2.2**	**0.79**	**2.7 ± 0.6**	**0.87**	**1.5 ± 1.2**	**0.41**	**3.6 ± 1.6**	**0.62**	**0.9 ± 0.8**	**0.30**	0.3 ± 1.2	0.04
Lamont, OK	**8.7 ± 6.9**	**0.35**	1.5 ± 2.4	0.12	1.1 ± 2.2	0.08	**2.1 ± 0.9**	**0.67**	0.5 ± 0.5	0.32	**0.8 ± 0.4**	**0.49**
Darwin, Australia	**0.8 ± 0.5**	**0.46**	**0.3 ± 0.2**	**0.39**	**-0.4 ± 0.3**	**0.36**	**1.1 ± 0.7**	**0.41**	0.4 ± 0.4	0.27	0.3 ± 1.0	0.04
Reunion Island	4.6 ± 6.2	0.15	0.8 ± 1.2	0.12	1.1 ± 1.2	0.23	-1.0 ± 3.7	0.02	-0.7 ± 3.0	0.02	0.4 ± 1.1	0.04
Lauder, New Zealand	**4.5 ± 3.7**	**0.33**	**2.8 ± 1.1**	**0.69**	**0.7 ± 0.6**	**0.58**	4.1 ± 9.5	0.06	2.5 ± 3.6	0.14	0.2 ± 9.9	0.00

Table 2.3: Mean correlation coefficients between OCO-2 trace gases and other elements of the state vector, including aerosol optical depth (AOD), and surface albedo in the oxygen-absorbing band (O_2A), weakly absorbing CO_2 band (WCO_2), and strongly absorbing CO_2 band (SCO_2) retrieved from OCO-2 observations. Bolded values denote statistically significant non-zero mean correlation coefficients at a 95 percent confidence interval.

Table 2.4: Mean correlation coefficients between OCO-2 trace gases and other elements of the state vector, including aerosol optical depth (AOD), and surface albedo in the oxygen-absorbing band (O2A), weakly absorbing CO2 band (WCO2), and strongly absorbing

Location	XCO2								XH2O							
	Total Variability				**High Frequency Variability**				**Total Variability**				**High Frequency Variability**			
	AOD	**O2A**	**WCO2**	**SCO2**	**AOD**	**O2A**	**WCO2**	**SCO2**	**AOD**	**O2A**	**WCO2**	**SCO2**	**AOD**	**O2A**	**WCO2**	**SCO2**
Bialystok, Poland	**0.24**	0.11	**0.14**	**0.13**	**0.12**	0.02	0.05	0.05	**0.09**	**0.15**	**0.12**	0.06	**0.08**	**0.13**	**0.12**	**0.05**
Karlsruhe, Germany	**0.18**	0.07	0.09	0.08	**0.12**	-0.01	0.01	0.02	**0.16**	**0.12**	0.07	0.04	**0.11**	**0.12**	**0.08**	0.03
Orleans, France	**0.13**	0.09	**0.14**	**0.16**	**0.07**	0.01	0.06	**0.09**	**0.20**	**0.17**	**0.13**	**0.11**	**0.17**	**0.16**	**0.14**	**0.09**
Garmisch, Germany	**0.17**	-0.01	**0.10**	**0.10**	**0.11**	0.01	**0.07**	**0.08**	**0.13**	**0.09**	0.05	0.02	**0.10**	**0.09**	**0.06**	0.02
Park Falls, WI	**0.14**	0.09	**0.16**	**0.13**	**0.07**	0.04	**0.10**	**0.09**	**0.10**	**0.11**	**0.10**	0.07	**0.08**	**0.12**	**0.14**	**0.09**
Lamont, OK	0.06	0.00	**-0.14**	**-0.14**	**0.04**	0.03	-0.01	0.00	**0.14**	**0.07**	**0.06**	0.04	**0.06**	**0.13**	**0.16**	**0.12**
Darwin, Australia	0.03	**-0.32**	**-0.40**	**-0.41**	0.01	-0.01	0.00	-0.01	**0.07**	0.01	-0.02	-0.03	0.03	**0.06**	0.04	0.01
Reunion Island	**0.18**	-0.03	**-0.13**	**0.12**	**0.08**	0.03	0.00	**0.05**	**-0.07**	**0.07**	0.04	**0.14**	**-0.09**	**0.09**	**0.07**	**0.18**
Lauder, New Zealand	**0.14**	-0.05	**-0.13**	0.03	**0.08**	0.02	-0.03	0.02	**0.16**	0.02	0.06	**0.12**	0.02	-0.03	**-0.08**	-0.04

2.5 Discussion and Conclusions

We developed a framework that allows us to leverage spatially dense soundings from the OCO-2 satellite and temporally dense soundings from TCCON to quantify the variance budget for XH_2O and XCO_2, with a focus on estimating the imprint of mesoscale transport on OCO-2 observations. We first developed a method to separate variations from local diurnal surface-atmosphere fluxes, synoptic-scale atmospheric transport, and mesoscale atmospheric transport from the overall variability in TCCON observations. We found that variability from synoptic-scale transport was, on average, 3 times larger than that attributed to diurnal fluxes for XH_2O and of 2 times larger magnitude for XCO_2. On average, and mesoscale variations in XH_2O and XCO_2 were similar in magnitude to the variability from local diurnal fluxes. The large contributions of mesoscale and synoptic scale transport in driving tracer variability underscores the importance of accounting for uncertainties in atmospheric transport and its subgrid-scale impact when using XCO_2 in a carbon flux inference system.

Second, we used geostatistical analysis to quantify explained (spatially coherent) and unexplained (random) variations in XCO_2 and XH_2O along OCO-2 tracks. We applied a 250-km high pass filter that, for XH_2O, isolated mesoscale variations that were the primary driver of along-track high frequency variability of XH_2O from OCO-2. We confirmed that the explained variations in XH_2O were primarily related to mesoscale transport using the tracer transport framework to compare temporal variability to spatial variations. For XH_2O, observed explained spatial variations of XH_2O were consistent with estimated explained variations within this framework (Figure 2.8a and 8b).

In contrast, we were not able to fully characterize the variance budget for XCO_2. Within our physical framework, the explained variations of XCO_2 observed by OCO-2 (Figure 2.8c and e) were too large to be explained solely by mesoscale atmospheric transport. We note that this mismatch was particularly acute when using the geostatistically-estimated a_{space} from OCO-2 (around 20 km) for XCO_2, but was also true when we substituted the observed a_{space} for XH_2O (Figure 2.8d). Together, these suggest another source of spatially coherent variance in OCO-2 XCO_2 that

both shortens the length scale of coherence and contributes additional spatially coherent variability. Based on correlation analysis with other elements of the state vector, we conclude that high-pass filtered XCO_2 fields from v8 of the OCO-2 Level 2 XCO_2 retrieval impart such structure and may explain part, though perhaps not all, of the mismatch in mesoscale time-space XCO_2 variability from TCCON and OCO-2. Correlations among elements of the retrieval state vector have also manifested in biases in XCO_2 retrieval output from GOSAT data, resulting in uncertainty of the magnitude and sign of posterior regional-scale flux estimates (Chevallier et al., 2014). Our estimates of explained and unexplained variability can be compared against other studies that leverage the OCO-2 data set. For instance, Worden et al., (2017) analyzed the contribution of natural variations simulated by NASA's high resolution GEOS-5 simulation to the OCO-2 error budget. They found that the natural variability of XCO_2 within 100 km neighborhoods was only about 0.08 ppm, and that the observed variability from OCO-2 exceeded this value due to bias from other elements of the OCO-2 v7 retrieval vector. We used our framework to estimate the natural (or mesoscale) contribution to variations in XCO_2 with our space-for-time framework. Based on mesoscale variations at northern hemisphere TCCON stations, we expect that the imprint of mesoscale systems along the satellite track should be about 0.4 ppm (over 250 km). That our value is larger than the natural variability in model output reported by Worden et al. (2017) underscores the utility of our geostatistical approach for quantification of the variance budget based on observations themselves, rather than model output that may contain its own set of biases. Further, our methodology enables us to quantify the random error in the observations (unexplained variability) as well as the spatially coherent error (the portion of the explained variability caused by systematic bias in the retrieval).

Our results suggest that our analysis framework yields robust quantification of the influence of mesoscale transport for XH_2O, despite the fact that local surface processes are likely also to impart local variations on this quantity. We acknowledge that our method to subtract local influence based on a climatological diurnal signal may result in biases of up to 30 percent of the fraction of variability in TCCON data being attributed to transport (Figure 2.2). Ultimately, the methodology to account for local signals at TCCON sites would benefit from additional information on local

fluxes across all sites, either from observations, such as eddy covariance fluxes, or from mechanistic coupled atmosphere-land models. Despite this caveat, the agreement between TCCON and XH_2O variations suggests that the 250 km filter used and the assumptions made in this manuscript are sufficient to account for the structure imparted by fine spatial scale transport.

Although there were complications with XCO_2, our results are promising for the use of geostatistical methods for parameterizing errors in inverse modeling frameworks. While we had hypothesized that our approach would allow us to separate the influence of mesoscale transport on XCO_2, we note that for robust flux inference, proper accounting for spatially coherent non-transport structures within the data is also necessary. The approach presented in this study can be applied to each subsequent data release, and we anticipate that as the retrieval algorithm becomes more mature, the importance of correlated errors will decrease and the role of mesoscale transport will be revealed more clearly. Our results highlight the importance of continued development of the OCO-2 retrieval algorithm, since correlations between XCO_2 and other elements of the state vector may induce bias and spurious spatial correlation in XCO_2 that mask the influence of atmospheric transport.

Our results also highlight the importance of constraining variations of XCO_2 attributed to atmospheric transport for improved inferences of carbon fluxes from inversion models. While the influence of random errors can be minimized by averaging multiple soundings, transport-driven processes introduce variability on XCO_2 observations that are both spatially and temporally correlated. For example, taking a 10 s average of XCO_2 observations along the OCO-2 track (about 70 km), similar to the method presented in Crowell et al. (2019) will reduce the unexplained error (Figure 2.6d) by 1N where N is the number of soundings in the averaging bin, because the unexplained errors for each sounding are assumed to be independent. In contrast the mean error for a bin decreases only by 1Neffective for the spatially correlated transport variability (Figure 2.6e), where Neffective can be approximated as the bin length divided by the geostatistical range (or autocorrelation length), which for XCO_2 is roughly 2-3 and may actually be closer to 1 if the XH_2O ranges are used as more appropriate. Since the length-scales at which the resolved variability is correlated

are comparable to the spatial distance encompassed by a 10-s along track average (Figure 2.7), the mesoscale variance is a key element of the signal that will be used in the inversions even when observations are aggregated. Given that the imprint of mesoscale and synoptic scale variance is not spatially or temporally uniform, ignoring it in an inverse modeling framework will lead to over confidence in some observational aggregates and underconfidence in others, ultimately shifting the distribution of fluxes. This is analogous to conducting a simple linear regression in which uniform error bars are assumed instead of assigning realistic errors to individual points; the resulting slope and intercept will differ depending on the method of assigning errors. Inversion techniques typically require uncertainty estimates and their correlations on the inversion grid-scale, which spans from roughly mesoscale (100km) to several times mesoscale depending on the transport model resolution. The appropriate uncertainty estimate will incorporate both instrument and algorithm error and spatially/temporally coherent subgrid-scale variability induced by transport and surface flux processes. Based on our analysis of both TCCON and OCO-2, the coherent mesoscale variability signal for XCO_2 is substantial relative to sounding errors alone and may be larger than transport variability estimates produced by most carbon cycle models, which may inadequately resolve mesoscale dynamics. Since mesoscale systems may have also been associated with frontal cloud coverage, lack of ability to constrain mesoscale variations may have resulted in large representation errors in inverse modeling (Corbin et al. 2008). Our results suggested that as a first step, we could use the explained variability derived from geostatistical analysis of OCO-2 data to inflate error estimates within inverse modeling systems. As a next step, we recommend the development of coupled high-resolution CO_2-weather models that fully capture fine-to-large scale spatial and temporal variations in carbon fluxes, as an alternative to constrain the imprint of atmospheric transport from XCO_2observations provided by OCO-2, OCO-3, and other emerging CO_2 monitoring satellites."

2.6 Acknowledgments

The authors thank the leadership and participants of the NASA OCO-2 mission and acknowledge financial support from NASA Award NNX15AH13G. A.D. Torres also acknowledges support from the NASA Earth and Space Science Fellowship Award 80NSSC17K0382. We thank TCCON for providing observations. We thank A. Jacobson and the National Oceanographic and Atmospheric Administration Earth System Research Laboratory in Boulder, CO for providing CarbonTracker CT2017 data, available online at http://carbontracker.noaa.gov. We thank S. Wofsy for providing HIPPO data, funded by the National Science Foundation and NOAA and available online at https://www.eol.ucar.edu/field_projects/hippo. The TCCON Principal Investigators acknowledge funding from their national funding organizations. TCCON data were obtained from the archive at https://tccondata.org. NARR data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/.

Chapter 3

Can We Infer the Imprint of Local Biospheric Fluxes in XCO_2 Observations from Space?

Note: This study remains in prep with a plan to submit for publication during the summer in 2023. In this chapter, we provide an overview of the current status of this study, as well outline the remaining work before submission.

3.1 Abstract

NASA's Orbiting Carbon Observatory-3 (OCO-3) instrument aboard the International Space Station is the first CO2-monitoring mission that collects total column-averaged CO_2 (XCO_2) data at various times of day. OCO-3 provides the first opportunity to observe the diurnal cycle of XCO_2 from space. We analyze OCO-3 observations in an attempt to calculate the diurnal cycle within a given calendar month by compositing data obtained across different days, focusing on three locations within the Total Carbon Column Observing Network (TCCON), which provides validation for space-based XCO_2 observations. At these three sites (Park Falls, Wisconsin, Lamont, Oklahoma, and Darwin, Australia), we find that OCO-3 observations are not sufficiently dense to infer the diurnal cycle in light of uncertainty in the observations and geophysical variation due to atmospheric transport. Instead, we leverage long-term records of XCO_2 provided by TCCON to quantify the minimum number of satellite overpasses as a function of time of day each month necessary to infer the climatological diurnal cycle of XCO_2. TCCON-based bootstrapping estimates suggest a minimum of 10 to 20 observations, which are made under sunny to mostly sunny

sky conditions, to reach a minimum of 100 percent relative error in the diurnal cycle amplitude estimates. This result agrees well with a similar approach applied to the and the 2019B version of the CarbonTracker posterior CO_2 fields, (CT2019B). Our study provides a roadmap to measure diurnal carbon fluxes globally for OCO-3, or similar remote sensing platforms that sample across the diurnal cycle, as more data are collected during the extended mission, as well as for future CO_2-monitoring satellite missions.

3.2 Introduction

Accurate representation of carbon cycle processes is a critical step in constraining future estimates of carbon dioxide (CO_2) concentrations under various greenhouse gas emission scenarios (Peters et al., 2017). Historically, lands and oceans have together taken up roughly half of all anthropogenic CO_2 emissions, with approximately a quarter of total emitted CO2 stored within each reservoir (Friedlingstein et al, 2022). According to the Global Carbon Project (Friedlingstein et al, 2022), uncertainty in global fluxes remains high at nearly 1 PgC Yr^{-1} between the land and atmosphere, and 0.5 PgC Yr^{-1} between the ocean and atmosphere. This uncertainty, which increases at smaller local and regional scales, is due to unresolved processes in terrestrial ecosystem models (O'Sullivan et al, 2022), poorly resolved constraints in assumptions used in atmospheric inversion models (Crowell et al, 2019; Chandra et al, 2022), as well as uncertainties in ocean biogeochemical process-based and observationally-based models (Fay and McKinley, 2022).

There are several approaches used within the broader scientific community to estimate carbon fluxes. These approaches are often broken down into two distinct categories: bottom-up and top-down. Briefly, bottom-up approaches make inferences of carbon fluxes by estimating CO_2 exchange at the process-level – such as tracking fossil fuel emissions, modeling plant behavior, or scaling up fluxes measured at eddy covariance flux towers. Our study focuses on top-down approaches, which infer carbon fluxes by tracking spatiotemporal changes in atmospheric CO_2 concentrations.

Examples of top down approaches include from simple box models (Bolin and Keeling, 1963) that make basic assumptions of hemispheric exchanges of CO_2 to more complex inverse models that are largely based on the framework used by Enting (2002). Regardless of approach, the robustness of top-down inferences have a major dependence on the spatiotemporal density of high quality CO_2 observations (Gurney et al, 2002; Baker et al, 2006; Masarie et al, 2011). This dependency is due to multiple sources and spatiotemporal scales of variability in CO_2 observations that each top-down method must constrain in order to infer fluxes. For example, in situ CO_2 observations are heavily influenced by vertical mixing between the free troposphere and boundary layer (Denning et al, 1995; Stephens et al, 2007) in addition to horizontal advection (Parazoo et al, 2008; Parazoo et al, 2011; Williams et al, 2014; Kerr et al, 2020). Inadequate representation of vertical and horizontal atmospheric transport processes yields large errors and uncertainties in estimated posterior fluxes in inversions (Gurney et al, 2002; Baker et al 2006, Houwelling et al, 2015; Miller and Michalak, 2020).

To fill critical global observational gaps in atmospheric CO_2, a series of CO_2-monitoring satellites launched. The first satellites equipped with CO_2-monitoring instruments launched with Aqua and Envisat in 2002, Aura in 2004, and Metop-A in 2006; however their spatial footprints and accuracy limited their ability to infer regional-scale carbon fluxes (Pan et al, 2022). By 2009, Japan Aerospace Exploration Agency's Greenhouse gases Observing SATellite (GOSAT-1; Ross et al., 2013; Yokota et al., 2009). GOSAT-1 provides high density measurements of total vertical column integrated CO_2 and methane (henceforth denoted as XCO_2 and XCH4). GOSAT-2, a second satellite launched in 2018 (Kataoka et al., 2017). National Aeronautics and Space Administration (NASA) launched the Orbiting Carbon Observatory-2 (OCO-2) in 2014 (Crisp et al, 2004; Eldering et al, 2017), and launched OCO-3 in 2019 and sits aboard the International Space Station (ISS) (Eldering et al, 2019). Additionally, China launched Tansat in 2016 (Yang et al, 2018). Together, these satellite observations, combined with existing networks of in situ observations, increase the robustness of posterior estimates of fluxes from inverse models (Basu et al, 2018).

3.2.1 Budgeting the diurnal variability of XCO_2

Our research is predicated upon the decomposition of the XCO_2 time series over a range of temporal scales (Equation 3.1). We build on the work of Guan et al (2023), which examines the interannual variability of XCO_2, and Mitchell et al (2023), which focuses on mesoscale variations of XCO_2. In Section 3, we discuss the methodology used to account for each term of Equation 3.1 in detail.

$$XCO_2 = XCO_{2,trend} + XCO_{2,IAV} + XCO_{2,season} + XCO_{2,synoptic} + XCO_{2,mesoscale} + XCO_{2,local} + \epsilon \tag{3.1}$$

At the longest timescales, XCO_2 increases year upon year due to anthropogenic emissions ($XCO_{2,trend}$). Over the period of years to decades, we also fold in some variations of XCO_2 driven by interannual variability ($XCO_{2,IAV}$), which primarily lies beyond the scope of this particular study. Within a given year, XCO_2 oscillates seasonally due to hemispheric scale seasonal fluxes ($XCO_{2,season}$), synoptic-scale atmospheric transport ($XCO_{2,synoptic}$), mesoscale atmospheric transport ($XCO_{2,mesoscale}$), and local fluxes ($XCO_{2,local}$) (as noted in Mitchell et al, 2023). We also include a bulk error term (ϵ), which represents variations attributable to measurement errors or other unaccounted features.

Our study focuses on the the ability to infer typical magnitudes of changes in column CO_2 due to local diurnal fluxes ($XCO_{2,local}$) using space-based instruments, such as OCO-3. Here, we define $XCO_{2,local}$ as diurnal variations of XCO_2 that are the result of net mass exchanges of CO_2 between the atmosphere and land (or possibly the atmosphere and oceans). At diurnal timescales, the magnitude $XCO_{2,local}$ varies both by season and latitude (Torres et al, 2019). For example, in northern mid-latitude regions $XCO_{2,local}$ can contribute as much as 2 ppm of daytime XCO_2 variability per day during the growing season, but as little as 0.1 ppm per day in the winter (Torres et al, 2019). At lower latitudes in the tropics, $XCO_{2,local}$ driven daytime variability is around 1 ppm per day nearly year-round (Torres et al, 2019).

Atmospheric transport drives significant variations in XCO_2 that often drowns out variations driven by $XCO_{2,local}$ (Keppel-Aleks et al, 2011; Keppel-Aleks et al, 2012; Torres et al, 2022). Atmospheric transport advects CO_2 gradients across all ranges of spatiotemporal scales ranging from the microscale eddies, to mesoscale circulations like thunderstorms, land-sea breeze interactions, to synoptic-scale frontal systems, to planetary scale Rossby waves and inter-hemispheric exchange. In the northern hemisphere midlatitude regions during the warm season, when the north-south gradient is maximized, advection of XCO_2 via large-scale eddies in the form of synoptic- to planetary-scale weather systems ($XCO_{2,synoptic}$) can contribute to approximately the same amount or greater of within-day variability of XCO_2 as $XCO_{2,local}$ (Torres et al., 2019). On a typical day, variability driven by $XCO_{2,mesoscale}$ can contribute more than half of the diurnal variability from $XCO_{2,local}$ (Torres et al, 2019; Mitchell et al, 2023).

The partitioning of XCO_2, $XCO_{2,local}$, $XCO_{2,mesoscale}$ and $XCO_{2,synoptic}$ depends on the length of averaging across spatiotemporal scales (Keppel Aleks, 2012; Torres et al, 2019). Torres et al (2019) demonstrated using 13 years of data that mean total XCO_2 variability scales directly with local net ecosystem exchange observations. This implies that with a sufficient timespan of observations, long-term contributions driven by $XCO_{2,mesoscale}$ and $XCO_{2,synoptic}$ on the diurnal XCO_2 diurnal variance budgetcancels out. As a result, a climatological representation of the diurnal variance budget of XCO_2 is primarily driven by $XCO_{2,local}$ in the absence of systemic biases embedded within the ϵ term in Equation 3.1 (Torres et al, 2019).

3.2.2 Study Overview

With the exception of the OCO-3 instrument, most of the other CO_2-monitoring satellites follow a polar orbit and only make measurements at approximately 13:00 LT (Crisp et al 2004; Yokota et al, 2009; Ross et al, 2013; Eldering et al, 2017; Eldering et al, 2019). As a result of the ISS's precessing orbit discussed in Section 2.1, OCO-3 collects data throughout the entire day (Eldering et al, 2019). This unique orbital pattern provides an unprecedented opportunity to observe the diurnal cycle of XCO_2 globally from space. For this study, we first provide a contemporary assessment of

OCO-3's capacity to directly infer the climatological diurnal cycle of XCO_2. We then leverage the longer and more complete TCCON and CarbonTracker records to define a minimum threshold of observations to reasonably estimate the diurnal cycle of XCO_2.

This chapter addresses the following science questions:

1. How well does OCO-3 currently estimate the climatological diurnal cycle of XCO_2?

2. What would be the minimum number of required space-based observations spread over the day to reliably reproduce the diurnal cycle of XCO_2?

Section 3.3 of this chapter discusses the datasets we use in this analysis in detail. Section 3.4 describes how we estimate each term of the XCO_2 diurnal cycle budget and our methods for identifying the minimum number of required observations to estimate the diurnal cycle of XCO_2. We show our results in Section 3.5 and discuss the implications of this study in Section 3.6.

3.3 Data

3.3.1 OCO-3

NASA began the installation the OCO-3 instrument on May 4, 2019 (Eldering et al, 2019). OCO-3 observations began in August 2019 with the extended mission expected to continue through 2029 (https://ocov3.jpl.nasa.gov/). The precessing orbit allows for the sampling of XCO_2 at different times of day within 52 degrees of the equator (Eldering et al, 2019). In a typical month, there are roughly 140,000 globally XCO_2 soundings over land and water that are quality flagged (QF) as "good" (Eldering et al, 2019). Unlike OCO-2, which has a consistent 16-day repeat cycle globally (Crisp et al, 2004), the OCO-3 repeat cycle is much more complex and geographically dependent (Eldering et al, 2019). Globally, the average repeat cycle of OCO-3 is approximately 70 days with very few soundings below 30 degrees south and as much as over 300 soundings at 55 degrees North (Eldering et al, 2019). While overpasses can occur at any time per day, only observations made with a solar zenith angle of 63 degrees pass the QF as "good" (Eldering et al, 2019).

Observations are made with OCO-3 in three scanning modes: land nadir, ocean glint, and Snapshot Area Mapping (SAMs), all of which were used in this study (Eldering et al, 2019). SAMs represents a new scanning mode that is not available with OCO-2. Here, a pointing mirror assembly, takes 100 km by 100 km samples of various land (and potentially oceanic) regions to observe potential CO_2 emission hotspots, volcanos, and other areas of interest (Eldering et al, 2019). The closest analog to SAMs with OCO-2 is "target mode," in which the satellite rotates such that grating spectrometers on board are pointed in the direction of the target (Crisp et al, 2004).

The OCO-3 instrument has a nearly identical design (described below) to that of OCO-2, which has been collecting data in low earth orbit since September 2014. To infer XCO_2, the OCO-3 instrument consists of three grating spectrometers, measuring reflected sunlight in strong CO_2 absorption band centered around 2.0 μm, a weak CO_2 absorption band near 1.6 μm, and the oxygen A band near 0.76 μm, (Eldering et al, 2019). Each measurement consists of eight cross-track footprints of 1.6 by 2.2 km^2 every 0.33 seconds (Eldering et al, 2019). Unlike OCO-2, the spectrometers on OCO-3 are guided by a pointing mirror assembly (PMA), which allows for measurements of irradiances at various solar zenith angles making SAMs possible (Eldering et al, 2019).

In this study, we use data from the OCO-3 B10 Level 2 bias-corrected XCO_2 from the full-physics retrieval product starting on August 8, 2019 and ending November 30, 2022 (O'Dell et al, 2018). These are derived from the measured irradiances, combined with ancillary meteorological and physics-based data, using the Atmospheric Carbon Observations from Space (ACOS) retrieval algorithm (O'Dell et al, 2018). For OCO-3, XCO_2 retrievals are accurate to within 1 ppm after calibration with the TCCON network (Eldering et al, 2019). This data is available for download online at https://disc.gsfc.nasa.gov/ labeled with the prefix "OCO3_Lite_FP_10" for each file. We use output from the quality control parameter in the OCO3_Lite_FP_10 output to retain observations that pass the quality flags as "good.".

3.3.2 TCCON

We calculate the climatological diurnal cycle in XCO_2 using observations from three test sites in the Total Carbon Column Observing Network (TCCON). TCCON is a global, ground-based network of Fourier transform spectrometers (FTS) that measure direct solar absorption spectra in the near-infrared, from which the total column-averaged mole fraction of atmospheric trace gasses like CO_2, CH4, N2O, CO, H2O, and HDO (Wunch et al. 2011a) are inferred. TCCON is regularly used for validating OCO-2 and -3 XCO_2 retrievals (Wunch et al, 2011b; Liang et al, 2017; Wu et al, 2018) and provides a long-term record with dense temporal coverage. In the absence of all cloud cover, a spectrum can be obtained in an approximately 2 minutes repeat-cycle.

For our analysis, we focus on three TCCON sites. We selected two mid-latitude sites in North America and one tropical TCCON site in Australia (Table 1). We choose these different latitude zones because we expect that patterns of atmospheric transport, which vary by latitude, may impact our ability to estimate the diurnal cycle (Torres et al, 2019). In addition to atmospheric transport, the seasonality of local diurnal fluxes as observed by TCCON also varies greatly by latitude (Torres et al, 2019). To compare TCCON with OCO-3, we closely follow the coincident criteria described by Wunch et al (2017) in which any retrieval within a 10 degree longitude and 5 degree latitude centered over a TCCON location is considered. We partition all of the available data within half-hourly bins for each OCO-3 overpass. Given the observational length from these sites, this aligns with anywhere between 0 and 7 (and most often just 1) coinciding observations from OCO-3 collected at a given hour of the day for a given month of the year (Figure 1).

Retrievals of XCO_2 are derived from absorption in the 1.58 and 1.60-μm absorption band (Wunch et al., 2011). We use TCCON data that has been processed using the GGG2014 retrieval algorithm (De Mazière et al, 2014). TCCON data typically have precision of 0.4 ppm and are tied to the World Meteorological Organization (WMO) CO_2 scale by periodic aircraft and AirCore overpasses (Wunch et al., 2010). We filter out any data where the solar zenith angle is greater than 55º.

TCCON Site	Location	Citation	Dates of TCCON measurements
Park Falls, WI, US	45.93ºN, 90.44ºW	Wennberg et al (2022a)	May 26, 2004 - August 31, 2022
Lamont, OK, US	36.69ºN, 97.56ºW	Wennberg et al (2022b)	July 6, 2008 - February 27, 2022
Darwin, Australia	12.46°S, 130.84° E	Deustcher et al (2010)	August 28, 2005 - April 30, 2020

Table 3.1: The list of TCCON sites and dates of measurements used in this analysis.

3.3.3 CarbonTracker

We use CO_2 output from the CarbonTracker to corroborate inferences about the detectibility of the diurnal cycle made from TCCON. CarbonTracker creates a 4-dimensional estimate (x, y, z, and t) of atmospheric CO_2 fields by assimilating multiple networks of in situ CO_2 observations from the surface, towers, aircrafts, and ships. CarbonTracker uses the TM5 atmospheric transport model bounded by meteorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-interim reanalysis product (Jacobson et al, 2020). In this study, we use the CO_2 and temperature products provided by the 2019B version of CarbonTracker (CT2019B). The temporal resolution of CT2019B output is every 3 hours, spanning from January 2000 to the end of December 2018. The horizontal resolution over our study region in North America is on 1 by 1 degree grid cells and 3 by 2 degree grid cells in Australia (Jacobson et al, 2020). We integrate CO_2 vertically using pressure-weighting and a uniform vertical averaging kernel for the nearest horizontal grid cell to each TCCON site to compute XCO_2 in CT2019B (Supplemental Figure B.1).

3.4 Methods

3.4.1 Estimating the climatological diurnal cycle of XCO_2

Torres et al (2019) demonstrated that the climatological diurnal drawdown XCO_2 scales directly with local net ecosystem exchange. To isolate the local (diurnal) imprint on XCO_2, we first account for the long term trend ($XCO_{2,trend}$) by fitting a linear polynomial to all available observations. After subtracting this trend, we infer $XCO_{2,season}$ by taking the mean all thedetrended

data within a given calendar month. By subtracting $XCO_{2,trend}$ and $XCO_{2,season}$, we are left with variations of XCO_2 driven by local fluxes, sub-seasonal transport, and in the case of OCO-3 and TCCON, measurement errors (Equation 3.1). We denote the detrended and deseasonalized XCO_2 as $XCO_{2,detrended}$. For TCCON and CT2019B, where there are sufficiently long records of data, we detrend and deseasonalize using XCO_2 data from their respective datasets. Given the limited data available with OCO-3, we deseasonalize and detrend OCO-3 XCO_2 data using values derived from the coincident TCCON site.

Note, we neglect the interannual variability (IAV) term from Equation 3.1 at timescales between $XCO_{2,trend}$ and $XCO_{2,season}$. This is in part because, as we discuss below, we average across multiple seasonal cycles, averaging away IAV, and in part because IAV on the order of 0.5 to 1 ppm is fairly small in comparison to long-term trends driven by anthropogenic emissions and seasonal cycle amplitudes (Guan et al., 2023).

We calculate the climatological diurnal cycle at each site using $XCO_{2,detrended}$ data from TCCON. To obtain representative samples of $XCO_{2,detrended}$ for each hour of observations, we first partition each observation into hourly local time bins at each TCCON site. We take the mean of all the data in each bin provided there are a minimum of 10 observations in that bin. Otherwise that bin is filtered out from our calculations. Once we obtain representative hourly $XCO_{2,detrended}$ data, we are then able to directly compute the climatological diurnal cycle of XCO_2. To remove potential biases folded in by large air masses, we filter out all TCCON and OCO-3 data obtained with a solar zenith angle of greater than 55 degrees. We then further partition all of the available representative $XCO_{2,detrended}$ data into their respective 12 month by 24 hour bins and take the mean of all available data.

We treat the climatological diurnal cycle derived from TCCON data as the "true" diurnal cycle of XCO_2. This provides us the basis to evaluate the performance of estimates of the climatological diurnal cycle of XCO_2 inferred with OCO-3 data.

3.4.2 Bootstrapping the diurnal cycle of XCO_2

We use a bootstrapping approach to estimate the uncertainty on the diurnal cycle drawdown given the sample size of the observations. For TCCON observations, we again partition all of the $XCO_{2,detrended}$ data into hourly bins as described above in Section 3.4.1, filtering out bins with fewer than 10 observations. We then bootstrap diurnal cycle computations using 1000 random samples of length N. For TCCON data, N increases from 1 to 100, while N increases from 1 to 400 with CT2019B data. We keep track of any discrepancy by computing 95 percent confidence intervals of the bootstrapped diurnal cycle for each sample N, and compare the magnitude of the confidence interval to the magnitude of the amplitude of the climatological diurnal cycle of XCO_2. We follow the same procedure for CT2019B data, although model output is only available at 3 hour local time intervals.

3.5 Results

3.5.1 Climatological diurnal cycle of XCO_2

OCO-3 shows no discernible trend in the diurnal cycle of XCO_2 for any month across the three locations we analyze (Figure 3.1). For example, in August over Park Falls, OCO-3 shows that XCO_2 falls from -0.4 ppm at 10:00 to -1.7 ppm by 11:00 LT, followed by an increase to 0.2 ppm between 11:00 and 12:00 LT; resulting in a net decrease of 1.3 ppm in one hour followed by a sudden increase of 1.9 ppm in the subsequent hour (Figure 3.1a). In the same time period, TCCON observes a steady drop from 0.2 ppm to 0.0 pm from 11:00 to 12:00 LT (Figure 3.1b). While the downward trend of XCO_2 continues into the late afternoon with TCCON with a drop down to -0.5 ppm by 15:00 LT, XCO_2 anomaliees from OCO-3 randomly oscillates between positive and negative values (Figure 3.1). Similar are observed in Lamont, Oklahoma in August (Figure 3.1c and d). In Darwin, Australia, the "warm" season with the greatest data coverage with TCCON and OCO-3 is November (Figure 3.1 e and f). TCCON consistently shows hour-to-hour drops in

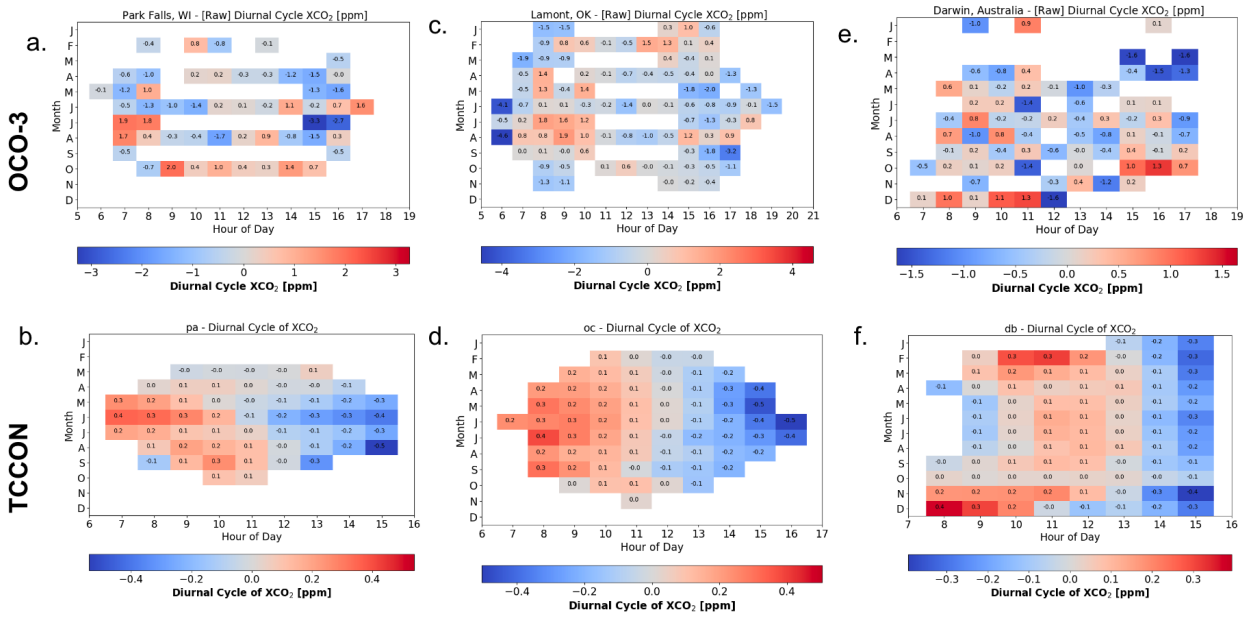

Figure 3.1: Diurnal cycle of XCO_2 as a function of month of year and hour of day over (a. and b.) Park Falls, Wisconsin, US, (c. and d.) Lamont, Oklahoma, US, and (e. and f.) Darwin, Australia. Climatological diurnal cycles derived from OCO-3 data are in panels (a.), (c.), and (e.); and TCCON in (b.), (d.), and (f.).

XCO_2 from 0.2 ppm at 08:00 LT to -0.4 ppm by 15:00 LT (Figure 3.1f). With the limited OCO-3 data available in November, we again note considerable hour-to-hour variability in XCO_2 with no discernable trend (Figure 3.1e).

The lack of a diurnal cycle almost certainly spurs from the sparsity of observations made by OCO-3 within each hourly bin (Figure 3.2). Using August again as an example, we note a maximum of 5 OCO-3 overpasses over an hourly bin in Park Falls and 7 in Lamont (Figure 3.3). In Darwin, there are only 1 to 3 observations within each hourly bin in November. In general 3 or fewer, with frequently only 0 or 1 observations with the hourly bins each month across each site (Figure 3.3). In contrast, there is an excess of 100 or TCCON observations within each bin when the solar zenith angle is less than 55 degrees at each site (Figure 3.4). During the late morning and afternoon hours in the peak growing season, TCCON observes over 300 to 400 hourly-averaged observations within each bin in Park Falls and Lamont (Figure 3.4a and b). In Darwin, the peak number of TCCON observations (300 to nearly 500) occur in the late afternoon during the dry sea-

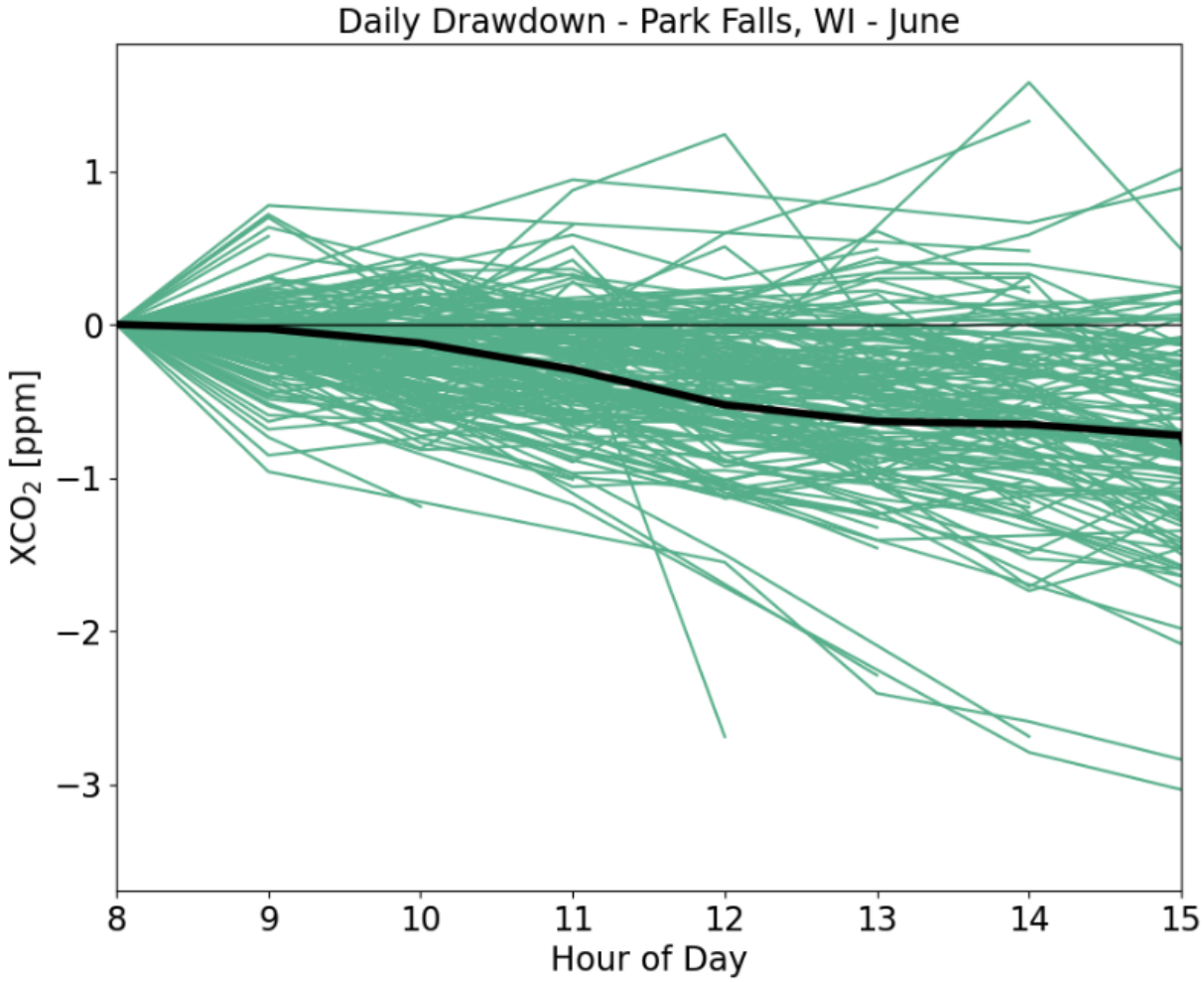

Figure 3.2: The diurnal drawdown of XCO_2 observed at the TCCON site in Park Falls, Wisconsin in June. The blue lines represent the drawdown on each individual day. The thick black line presents the climatological diurnal cycle of XCO_2.

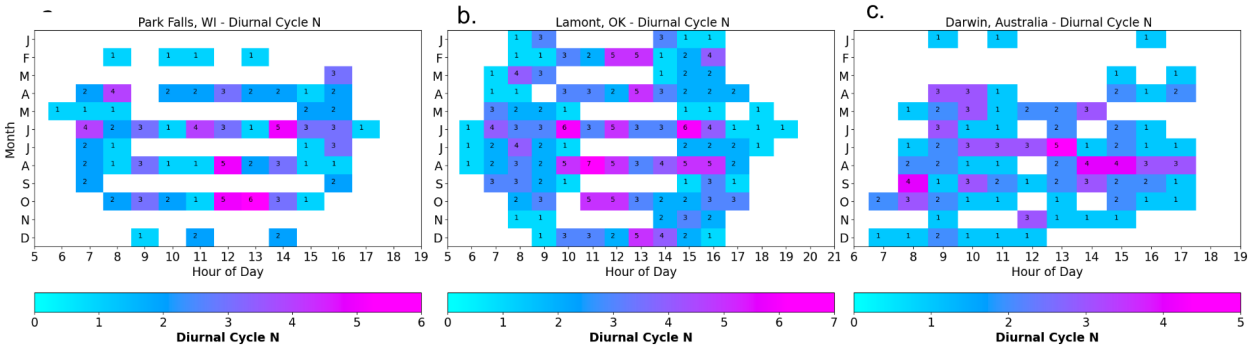

Figure 3.3: OCO-3 overpasses as a function of month of year and hour of day over (a.) Park Falls, Wisconsin, US, (b.) Lamont, Oklahoma, US, and (c.) Darwin, Australia, from August 8, 2019 through November 30, 2022.

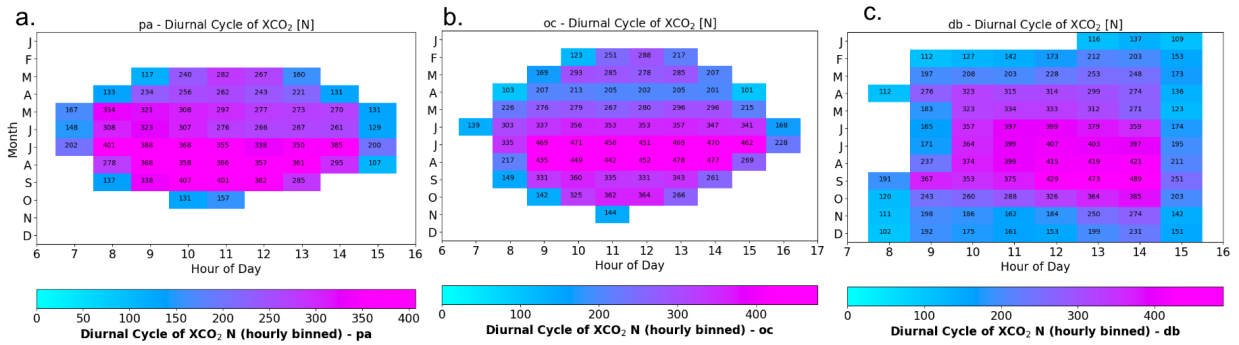

Figure 3.4: The same as Figure 3.3, but using all available TCCON data.

son from June to October (Figure 3.4c). Unfortunately, this time period is challenging to compare against OCO-3 since the dry season coincides with the weakest diurnal drawdown observed by TCCON in Darwin.

We note that the overall patterns in the diurnal cycle of XCO_2 described with TCCON also match closely what we observe with CT2019B.

3.5.2 Bootstrapping XCO_2 as a function of sample size

The trajectory of XCO_2 varies significantly from day-to-day (Figure 3.2). We attribute much of this day-to-day variability to random atmospheric transport patterns and random and systematic retrieval errors (Equation 3.1). Given the randomness of the spread of the evolution of the XCO_2

time series on a given day, as the number of observations increase, the more representative does the mean $XCO_{2,detrended}$ value becomes to the "true" climatological diurnal mean, as shown with TCCON. Thus, increasing the coverage of satellite observations over time, as well as the quality of retrievals, would provide adequate direct estimates of the diurnal cycle of XCO_2 globally.

Across each site, the uncertainty in estimates of the climatological diurnal cycle each hour decreases as the number of observations increases (Figure 3.5). At the northernmost site of our analysis in Park Falls, Wisconsin in July, we note that the range of the 95 percent confidence interval decreases from 4 ppm with 5 observations to 0.5 ppm with 100 observations with CT2019B data (Figure 3.5a). We note approximately identical numbers with TCCON in Park Falls (Figure 3.5b). At Lamont, Oklahoma, we observe both slightly lower estimates of error with fewer samples (approximately 1.5 ppm with 5 samples), but here we also have a smaller "true" diurnal cycle amplitude (Figure 3.5c and d). We see the smallest error over Darwin, Australia in February at only approximately 1 ppm with 5 samples observed by TCCON and just under 1 ppm observed by CT2019 (Figure 3.5e and f). In Darwin, we note the magnitude of the diurnal amplitude of XCO_2 is observed by CT2019B remarkably low at only 0.1 ppm, compared to the approximately 0.7 ppm observed by TCCON (Figure 3.5e and f).

To achieve 100 percent relative error on the true diurnal cycle, we require 10 to 20 observations within an hourly bin. Across each site, the uncertainty asymptotes to under 0.5 ppm (about 50 percent relative error) with sample sizes of roughly 50 or greater, which we roughly consistent with the magnitude day-to-day variability of the imprint of local diurnal fluxes estimated at Park Falls in Torres et al (2019). In Park Falls (Figure 3.6a) and Lamont (Figure 3.6b), we observe remarkable consistency between TCCON and CT2019B – which increases our confidence in results. However, in February in Darwin there are significant unexplained differences between TCCON and CT2019B data, again owing to the differences in the diurnal cycle amplitude. It is worth noting that despite differences in the magnitude of the diurnal drawdown, the trajectory of the range of the 95th percentile estimated error of the climatological cycle of XCO_2 in CT2019B remains remarkably similar to the range observed by TCCON. Overall, the estimated error in CT2019B is

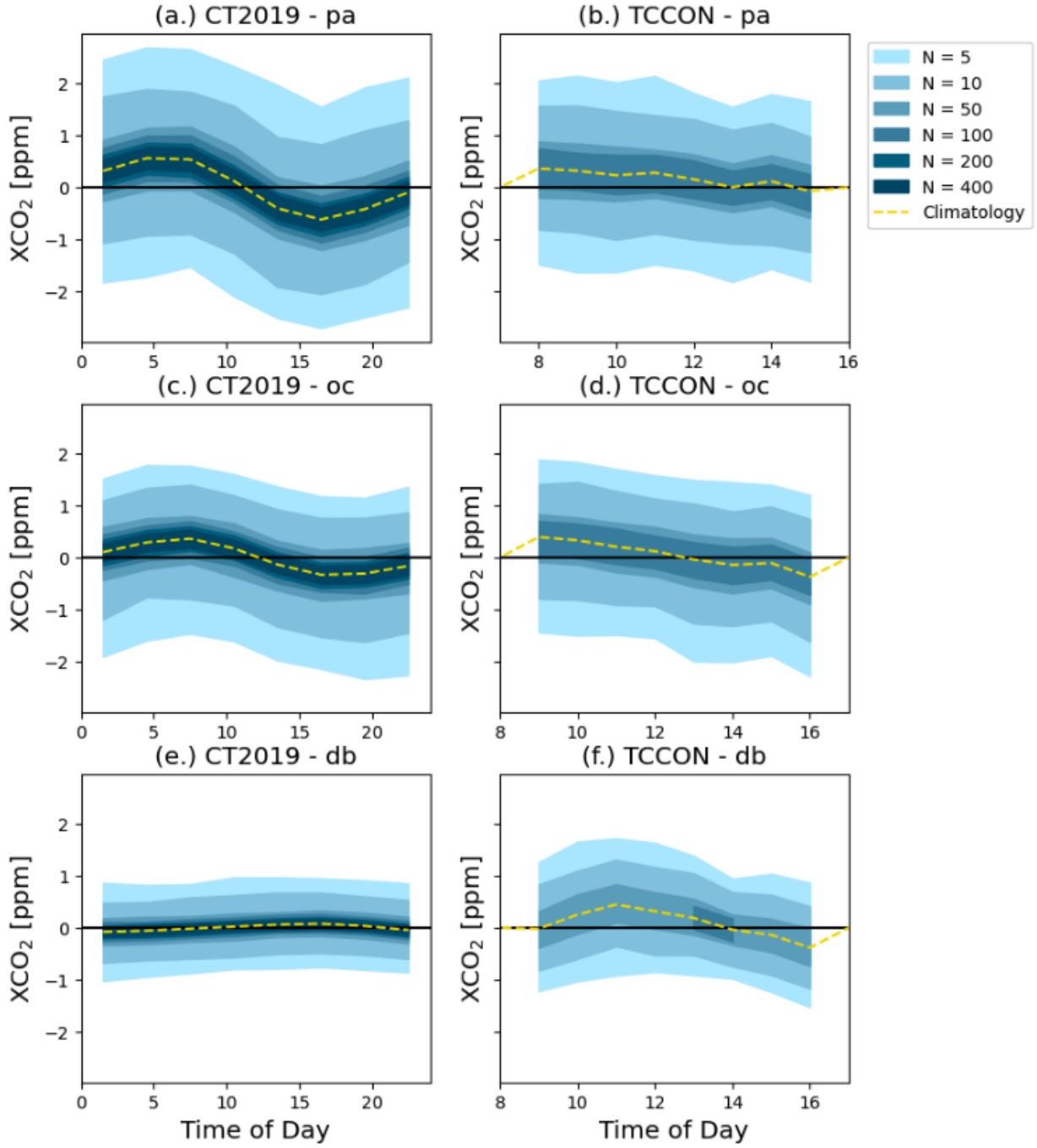

Figure 3.5: The range of the bootstrapped diurnal cycles of XCO_2, shaded by the sample size of N. The shaded region represents the range of the middle 95th-percentile of the diurnal cycle of XCO_2 each hour. The yellow dashed line represents the "true" diurnal cycle at each respective site. The right (left) column represents bootstrapped data for (a. and b.) Park Falls, Wisconsin, US, (c. and d.) Lamont, Oklahoma, US, and (e. and f.) Darwin, Australia. Panels (a.), (c.), and (e.) represent TCCON data and panels (b.), (d.), and (f.) are from CT2019B.

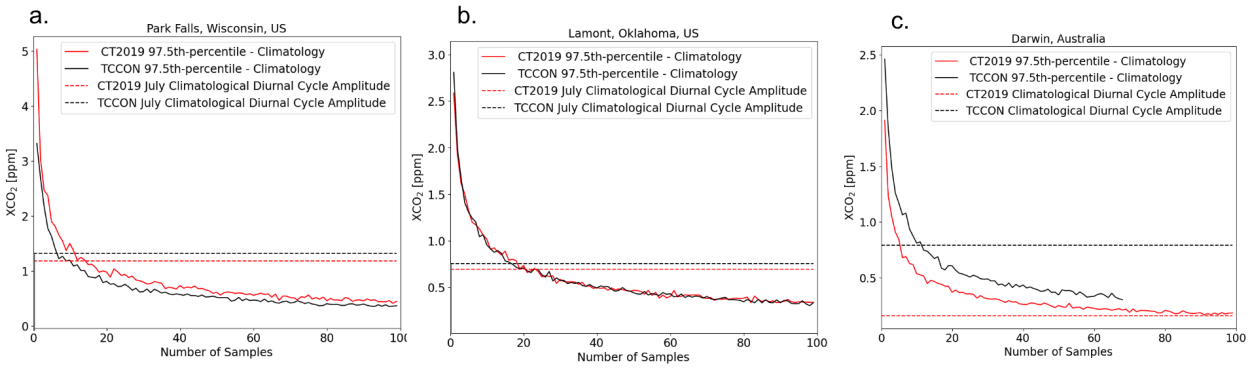

Figure 3.6: The difference between the 97.5th percentile of the diurnal cycle of the bootstrapped XCO_2 (red lines) compared to climatology (black lines) as a function of the total number of samples for (a.) Park Falls, Wisconsin, US, (b.) Lamont, Oklahoma, US, and (c.) Darwin, Australia. We use July climatology for Park Falls and Lamont, but February climatology at Darwin. The solid lines use TCCON data and the dashed lines represent CT2019 data.

approximately 0.2 ppm smaller than that observed by TCCON, which we speculate may be due to the coarser 3º x 2º resolution of the global CT2019B output (compared to the 1º x 1º horizontal resolution output available at the North American sites) (Figure 3.6c). Still, TCCON shows that 10 to 20 observations are sufficient to reduce the relative error in climatological diurnal cycle of XCO_2 estimates in Darwin, which is consistent with our results in Park Falls, and Lamont (Figure 3.6).

3.6 Discussion and Conclusions

Our results show that that – to date – OCO-3 has not collected enough data in any of the three sites we analyzed (Park Falls, Wisconsin, Lamont, Oklahoma, and Darwin, Australia) to make direct inferences of the climatological diurnal cycle of XCO_2 (Figures 1 and 3). However, we are able to leverage the longevity of coincident temporally dense TCCON observations at each site to obtain the "true" climatological diurnal cycle of XCO_2 (Figure 3.1).

Across each site, we estimate that there must be a minimum of 10 to 20 observations to get 100 percent relative error on the estimate of the climatological value XCO_2 each hour. The uncertainty of these estimates when using 10 to 20 samples is approximately equal to the magnitude

of the peak-to-trough diurnal drawdown of XCO_2 (Figure 5). It is noteworthy that this result was consistent across each site using TCCON, and Park Falls and Lamont with CT2019B data. The minimum threshold of 10 to 20 samples is consistent with the representative timescales of larger synoptic-scale weather systems in previous studies if observations were made over the course of sequential days (Torres et al 2019; Keppel-Aleks et al, 2011; and Keppel-Aleks et al, 2012). We note a divergence in the results between TCCON and CT2019B in Darwin which are driven by a large difference in the magnitude of the diurnal cycle amplitude in February between the two datasets (Figure 3.6). Using TCCON data exclusively, the results still suggest 10 to 20 observations are sufficient to reach 100 percent relative error in estimates of the diurnal cycle. However, the amplitude of the diurnal cycle in CT2019 in Darwin of 0.1 ppm is smaller than the variability of XCO_2 with even 50 to 100 observations (Figure 3.6). Overall, these results underscore the need to increase the number of global observations of XCO_2 or the need to understand the processes that govern the spatiotemporal variability of XCO_2 with fewer observations.

Another caveat is that this statistical analysis and bootstrapping methods used in this chapter of this thesis are preliminary. Prior to publication, we will need to include more robust analysis. The approach we use will likely continue bootstrapping each hour of binned data. However, we will instead estimate the magnitude of diurnal drawdown of XCO_2 using linear regression instead of peak-to-trough analysis used in this thesis. We will also determine if we can reduce the total number of XCO_2 samples each hour to infer reasonable estimates of the climatological diurnal cycle accounting for the imprint of larger-scale synoptic-scale atmospheric transport as discussed in section 3.3. We hypothesize that we can account for synoptic-scale atmospheric transport using 700 mb potential temperature data as a proxy for transport (Equation 2). From Torres et al (2019), the imprint of synoptic-scale transport imposes a signal that is approximately an equal order of magnitude of the diurnal drawdown of XCO_2 in the midlatitudes. While we will unlikely include finer scale atmospheric transport in this analysis, which can be estimated using semivariograms, it can account for variations in XCO_2 that are about half the magnitude of the climatological diurnal drawdown across most latitudinal bands (Torres et al, 2019; Mitchell et al, 2023).

Even when accounting for atmospheric transport, there still remain several uncategorized variations in XCO_2 that are all embedded within the error term, ϵ, in Equation 3.1. We anticipate that these variations likely include retrieval errors in TCCON or OCO-3, spurious fossil fuel emissions, or other potential sources. Accurate representations of each of these sources of variability are needed in order to back out the "true" imprint of local carbon fluxes. Even in the absence of retrieval errors with CT2019B, we note a remarkable consistency in results when compared to TCCON data – including the minimum number of observations to have 100 percent relative error or less in estimates of the diurnal cycle of XCO_2. This increases our confidence in our results and the number of observations are required to properly estimate typical contributions to XCO_2 variability by local diurnal fluxes.

We expect that improvements in the characterization of the diurnal variability budget of XCO_2 in Equation 3.1 will increase our ability to more directly quantify the climatological imprint of local carbon fluxes with space-based observations. Additionally, with an additional three years worth of data from OCO-3's extended mission and a consequence of the orbital pattern of XCO2, we anticipate that there could be increasing utility in making early estimates of the climatological diurnal cycle of XCO_2 using OCO-3 in regions with greater than 10 to 20 overpasses at a particular time of day and month. This threshold could potentially be reduced with better constraints on other terms in Equation 3.1. These estimates could be compared against estimates of fluxes derived from inverse models, terrestrial biosphere models, and other methods of inferring local carbon fluxes. This study not only examines the ability for OCO-3 to estimate the diurnal cycle of XCO_2, but could also inform the design of future CO_2-monitoring satellite missions.

3.7 Future Work

We anticipate that we will submit a paper based on the work done in this chapter in June. Prior to publication, our efforts are concentrated in two areas:

1. Increasing the robustness of the bootstrapping method. Instead of simply taking the mean of

each hourly bin to estimate the diurnal cycle of XCO_2, we are bootstrapping each hourly bin from 5 to 100 samples and bootstrapping the best fit linear regression slope of the diurnal drawdown of XCO_2. By multiplying the slope by the number of daylight hours observations available, this provides an alternative estimate of the diurnal drawdown of XCO_2 for each month. Then we will be able to evaluate the minimum number of samples each hour are required to estimate the climatological drawdown of XCO_2.

2. We will attempt to account for the imprint of synoptic-scale transport of XCO_2 for each observation. We will then repeat the bootstrapping analysis using the synoptic-scale atmospheric transport corrections. We will be able to determine if these corrections are able to improve estimates of the diurnal drawdown of XCO_2 with fewer observations. We note that we are still neglecting to account for mesoscale atmospheric transport – as well as accounting for potential contributions to the error term in Equation 3.1.

3.8 Acknowledgments

The authors thank the leadership and participants of the NASA OCO-2 mission and acknowledge financial support from NASA Award NNX15AH13G. A.D. Torres also acknowledges support from the University of Michigan's Rackham Merit Fellowship. We thank TCCON for providing observations. We thank A. Jacobson and the National Oceanographic and Atmospheric Administration Earth System Research Laboratory in Boulder, CO for providing CarbonTracker CT2019B data, available online at http://carbontracker.noaa.gov. The TCCON Principal Investigators acknowledge funding from their national funding organizations. TCCON data were obtained from the archive at https://tccondata.org.

Chapter 4

Conclusions

CO_2 is one of the largest contributors to the warming of the planet and understanding how much CO_2 will remain in the atmosphere is critical for understanding future climate scenarios through 2100. Future projections of CO_2 require that we accurately model the processes that govern the mass exchange of carbon between the land and atmosphere and ocean and atmosphere (Bonan and Doney, 2018). By constraining contemporary carbon fluxes at increasing spatiotemporal resolution, we move closer to understanding the processes that govern the contemporary carbon cycle – and can potentially learn how these processes will respond to different environmental forcing. In this dissertation, we examine the spatiotemporal behavior of XCO_2 at timescale that ranges from within a day through up to 21 days. Spatially, this represents mesoscale to synoptic-scale processes.

In Chapter 2, we used simultaneous measurements of XCO_2 and XH_2O to evaluate and introduce a framework for quantifying along-track mesoscale variability of XCO_2 from the OCO-2 retrievals using semivariograms. We also quantified the diurnal amplitude of XCO_2 across several TCCON sites, as well as estimated the climatological imprint of mesoscale and synoptic scale variability of XCO_2 using TCCON. In doing so, we demonstrated that climatological drawdown of XCO_2 scales directly with local carbon fluxes (net ecosystem exchange). As an ancillary outcome, we demonstrated the utility semivariograms provide as an alternative independent measure of retrieval error. The framework we use in this study can be applied to future satellite missions and other relatively inert trace gasses.

In Chapter 3, we evaluate the potential to directly infer the climatological diurnal cycle of

XCO_2 using OCO-3 data. While, to date, there is not enough OCO-3 data available to make reliable inferences of the diurnal cycle of XCO_2, we use TCCON and CarbonTrack 2019B data to estimate the minimum number of observations required to reliably estimate the diurnal drawdown of XCO_2 from space. At all three sites studied (Park Falls, Wisconsin, Lamont, Oklahoma, and Darwin, Australia), we determined that a minimum of 10 to 20 observations were needed to reach within 100 percent relative error in any estimate of the climatological diurnal drawdown of XCO_2. Future work in Chapter 3 will involve evaluating how many observations would be needed if we properly account for synoptic-scale atmospheric transport. From there, we can iterate by removing mesoscale and finer scale atmospheric transport, as well as accounting for biases in retrievals. The work done in chapter 3 provides a roadmap for directly observing the diurnal cycle – and by extension – the mass exchange of CO_2 from space.

By understanding the drivers of variability of XCO_2, we can provide independent estimates of regional scale carbon fluxes that complement research from inverse models, process-based carbon cycle models, and the scaling up of local flux data.

Appendix A

Supplemental Figures for Chapter 2

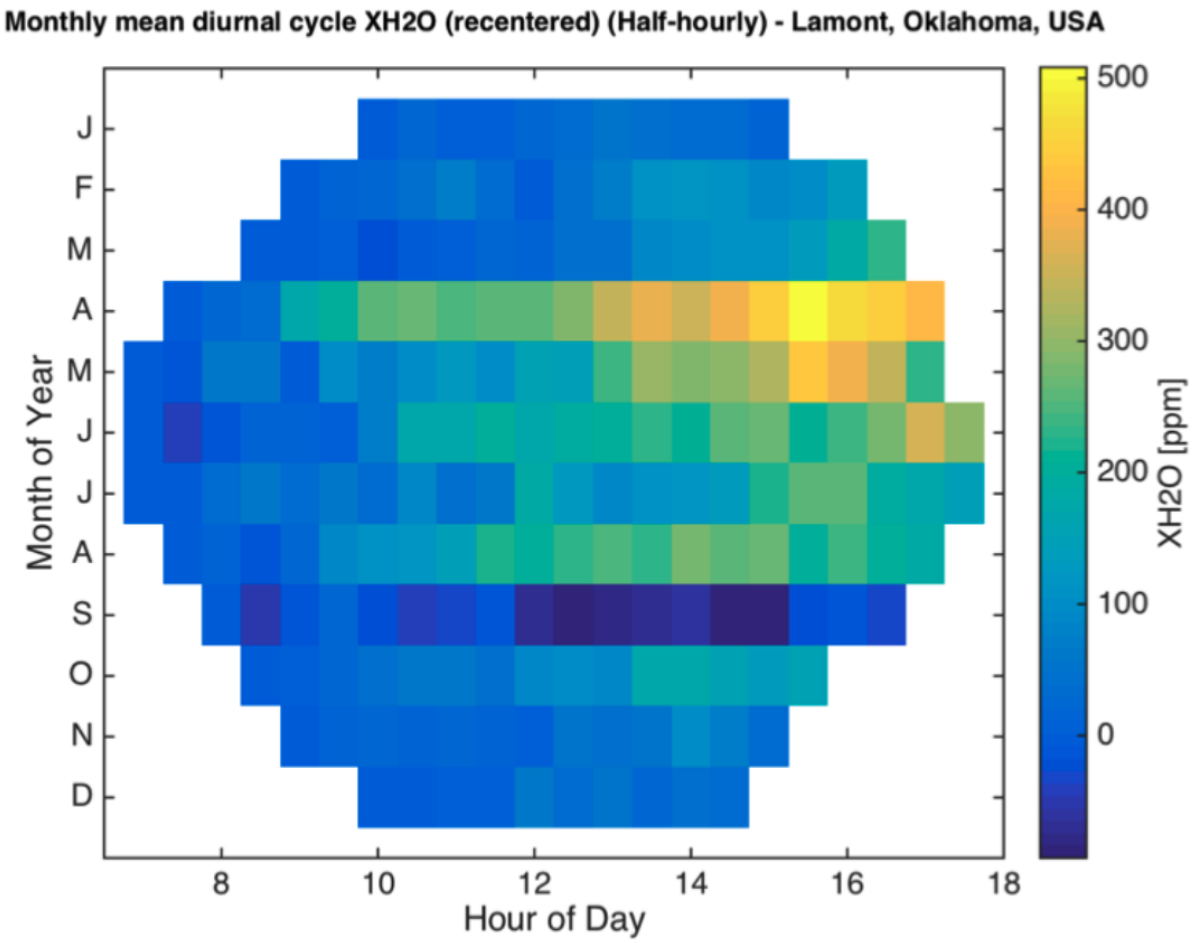

Figure A.1: Climatological diurnal drawdown of XH_2O observed at the TCCON site in Lamont, Oklahoma partitioned into half-hourly bins.

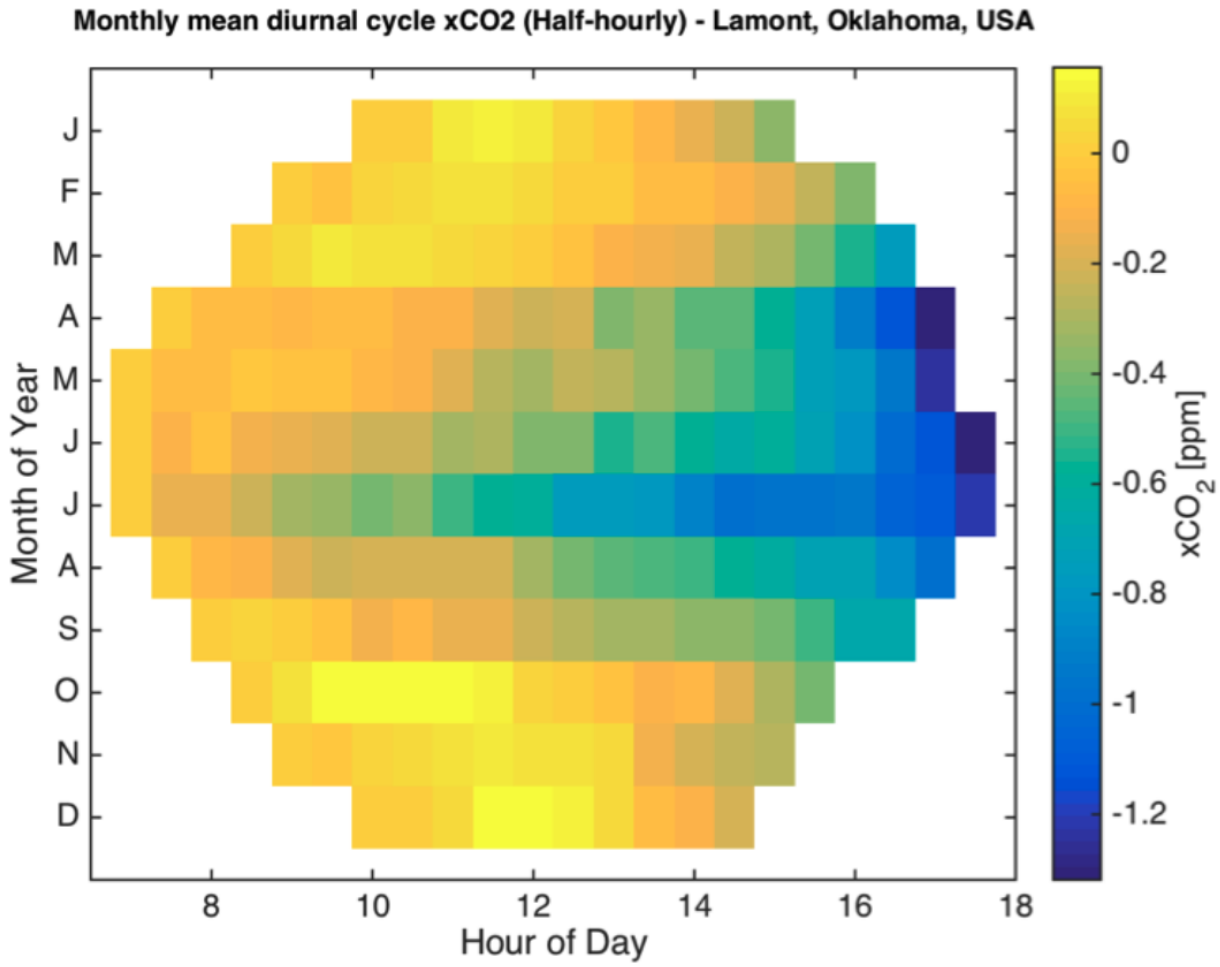

Figure A.2: Similar to Supplemental Figure 1, but for XCO_2.

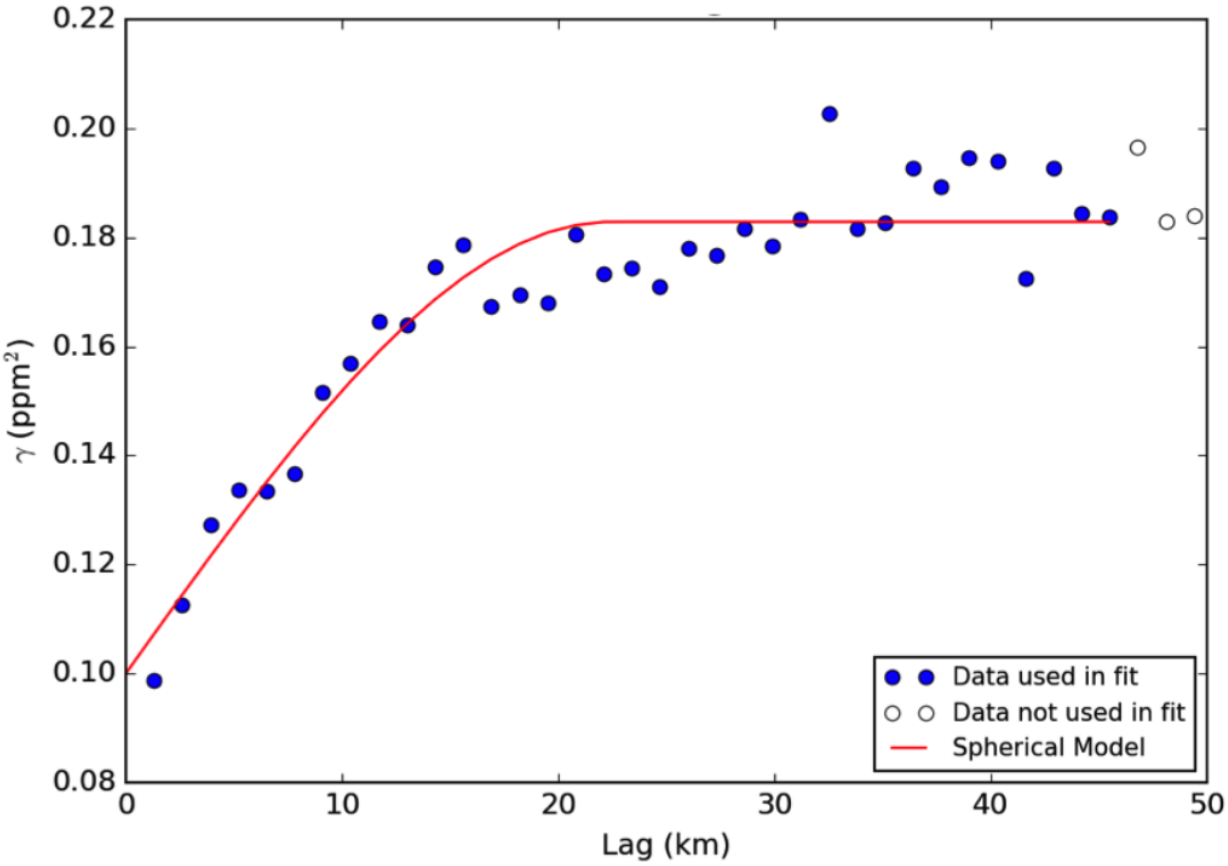

Figure A.3: A sample semivariogram of XCO_2 derived from an OCO-2 overpass fitted to a spherical model (red line).

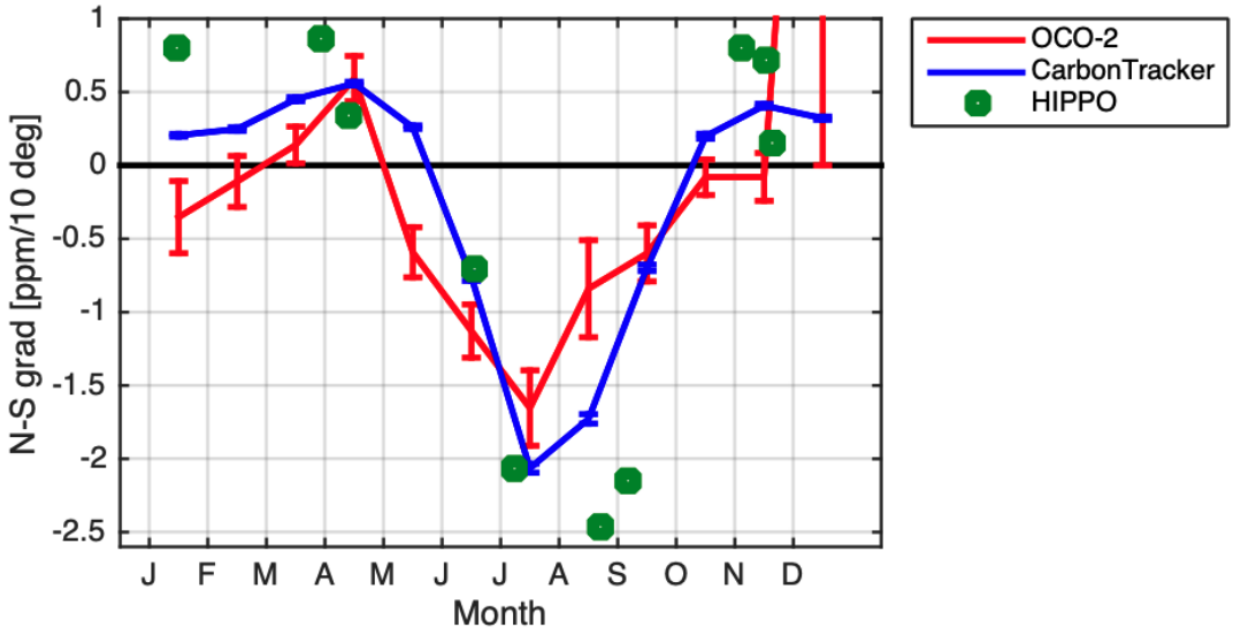

Figure A.4: Monthly mean N-S gradient of XCO_2 calculated from OCO-2 data (with at least over 30 coincident overpasses) (red) and CT2017 output (blue) within a 10ºx10º grid centered on Lamont, Oklahoma, and HIPPO transects over the Pacific Ocean between 25º and 45ºN (green). The error bars represent the standard error of monthly mean N-S gradients.

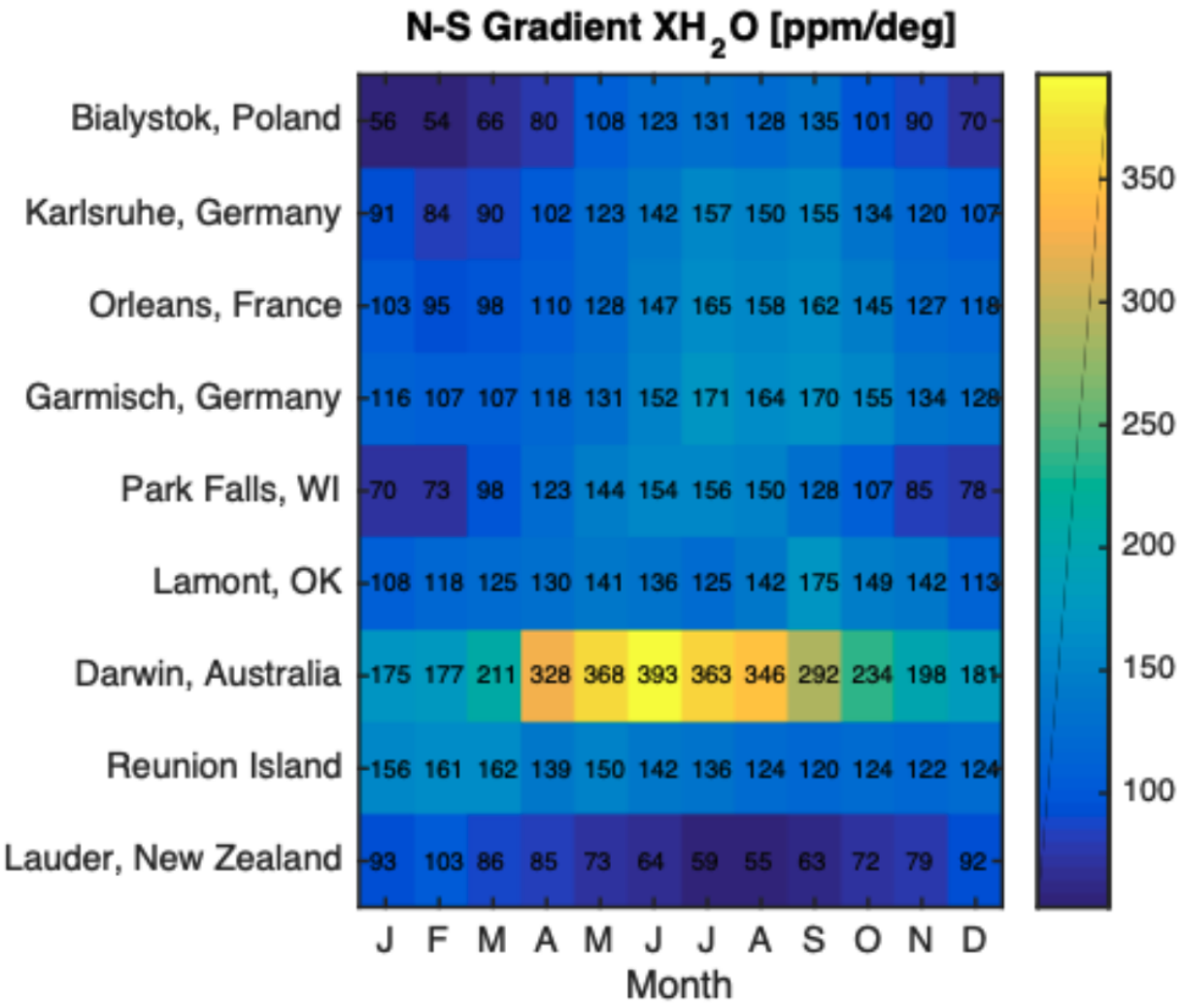

Figure A.5: Monthly mean N-S gradients of XH_2O at each TCCON site derived from CT2017 data.

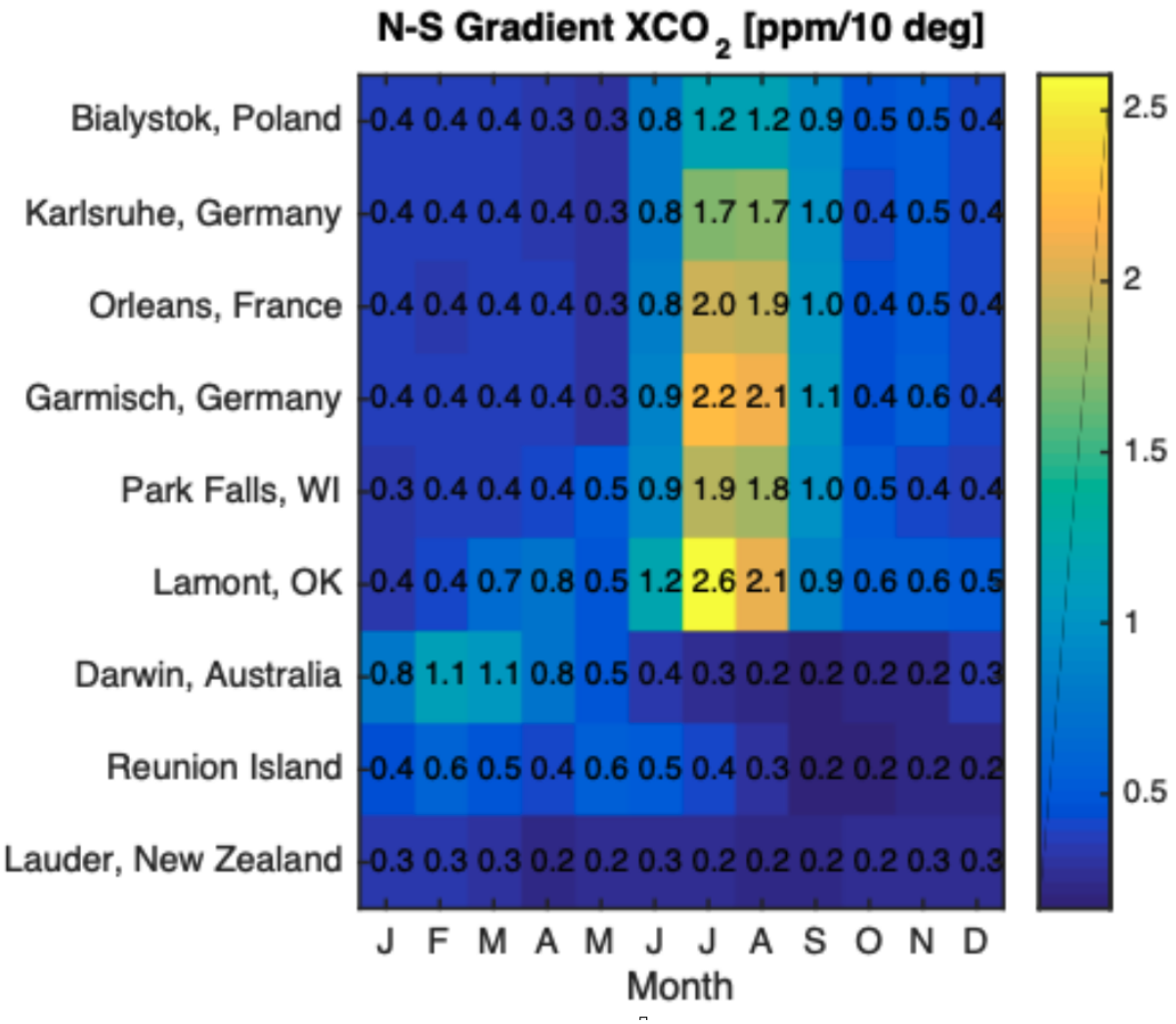

Figure A.6: Similar to Supplemental Figure 6, but for XCO_2.

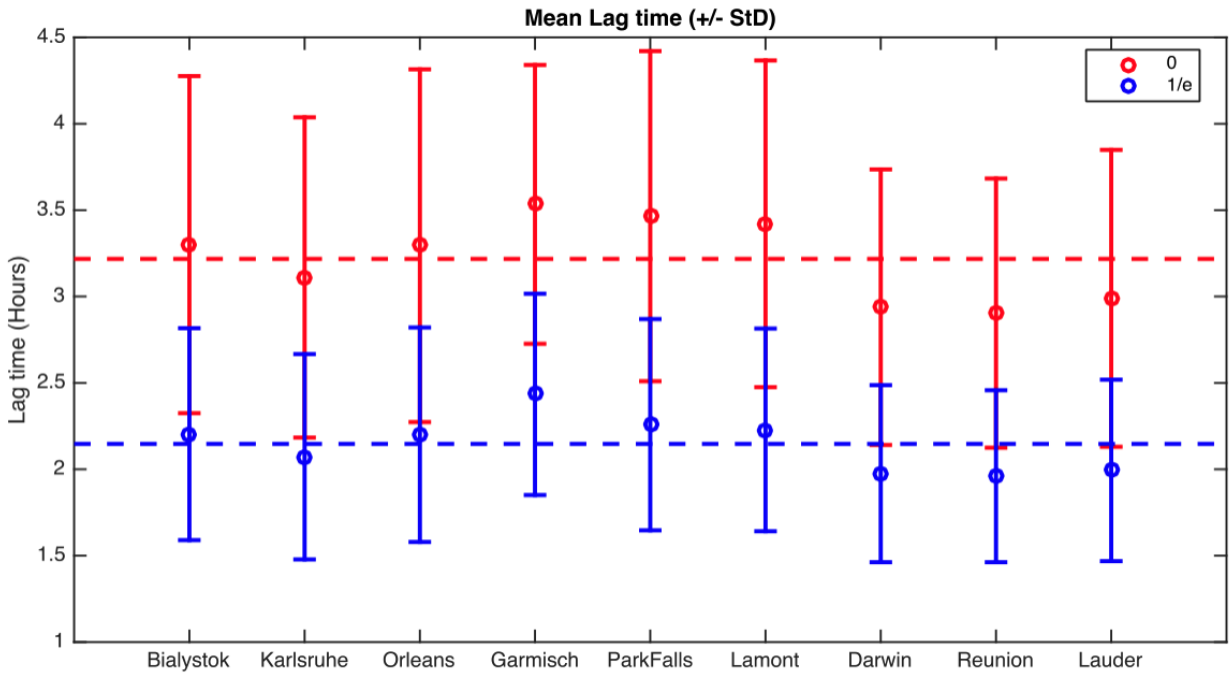

Figure A.7: Mean lag time of within-day variability of residual XCO_2 across each TCCON site from 800 to 1800 LST every day (dashed lines) for when autocorrelation values drop to less than 0 (red) and e^{-1}(blue). The circles represent the mean lag times at each TCCON site. The error bars represent the standard deviation of the mean lag times at each TCCON site.

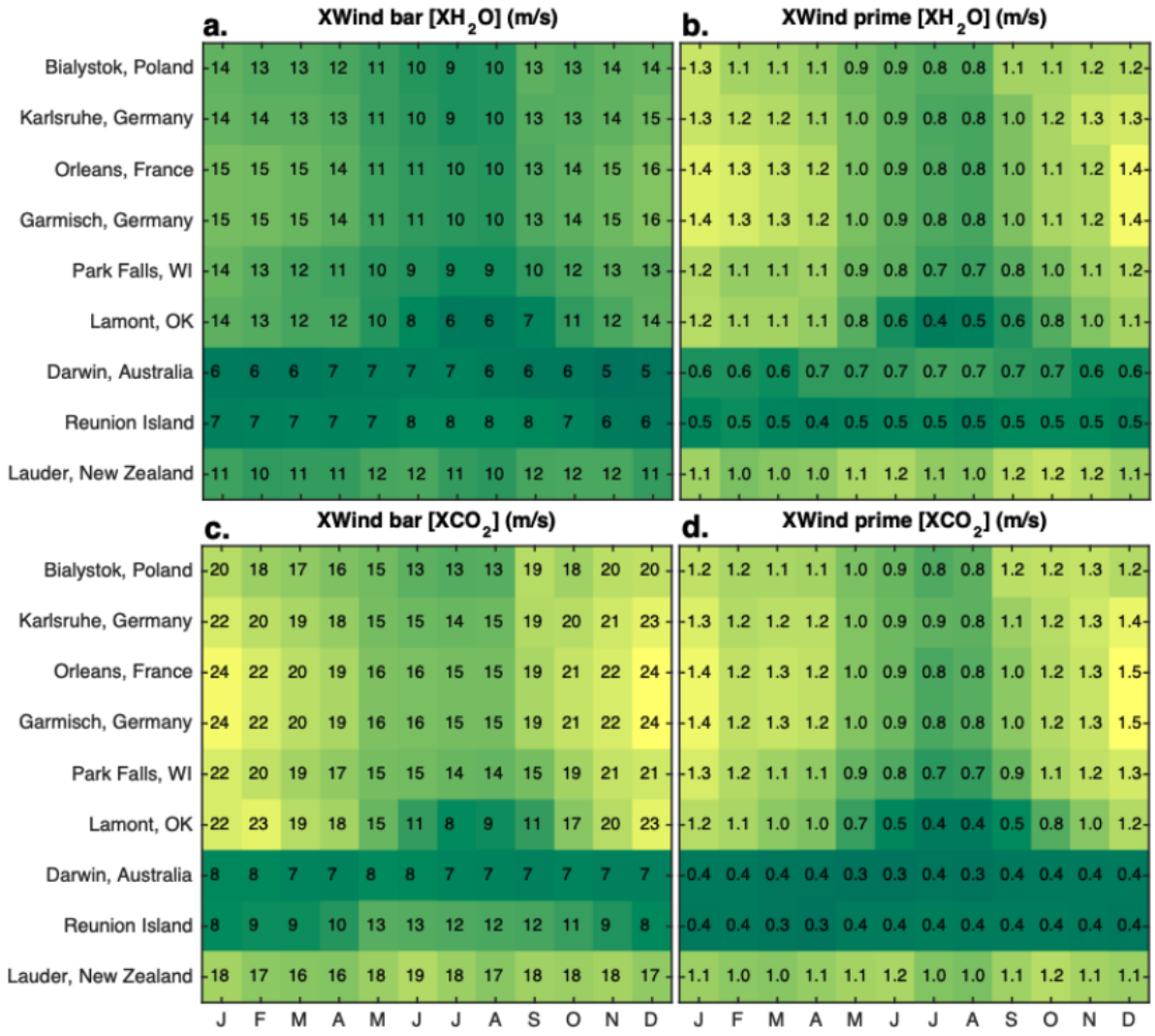

Figure A.8: Monthly within-day (1000 to 1800 LST) (a. and c.) mean and (b. and d.) standard deviation of total column averaged horizontal wind speeds weighted by vertical profiles of (a. and b.) H_2O and (c. and d.) CO_2 at each TCCON site derived from CT2017 output.

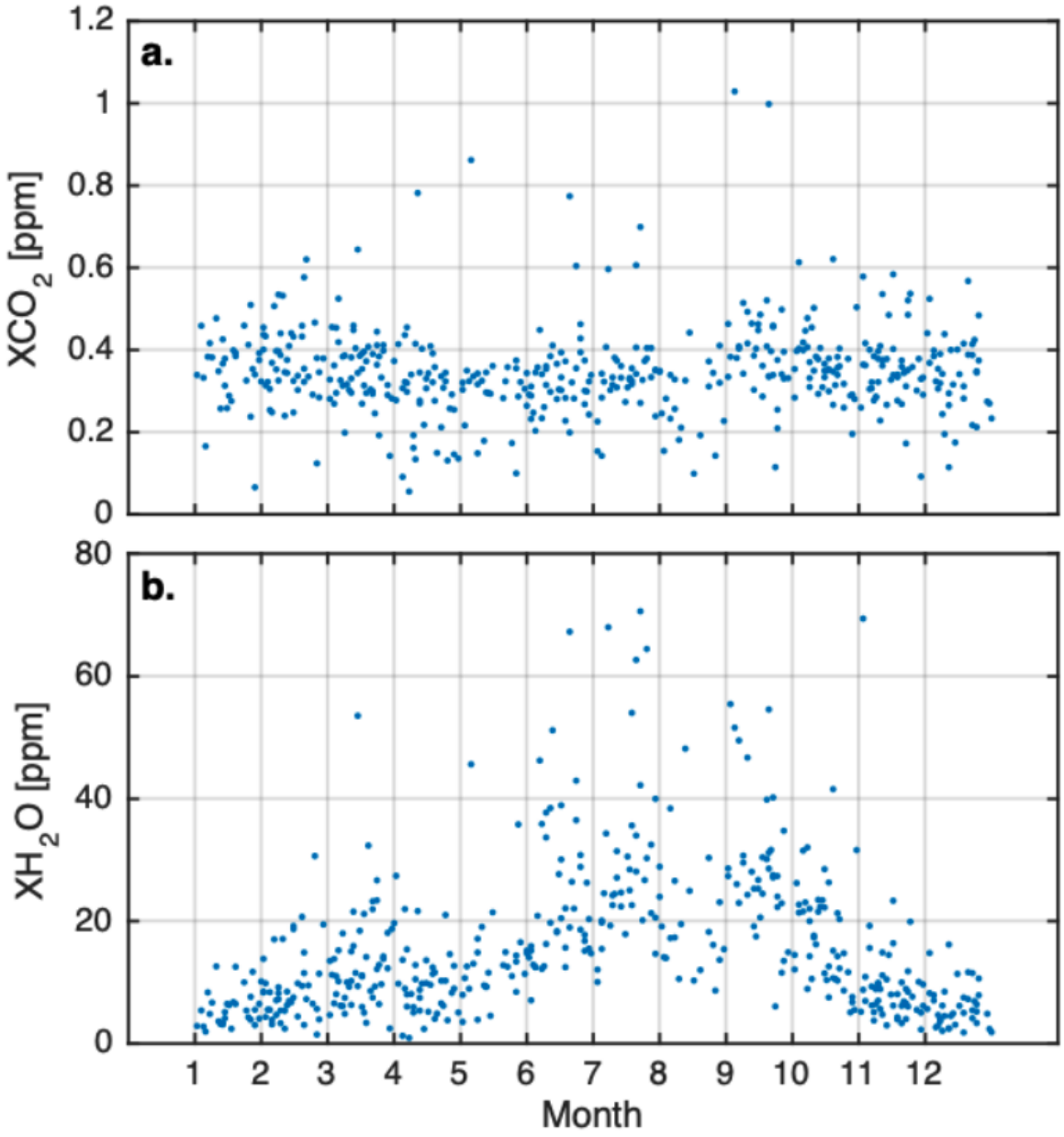

Figure A.9: Mean standard deviation of a. XCO_2 and b. XH_2O for every cross-track retrieval averaged into each bin described in Section 2.2.1 for each OCO-2 track over Lamont, Oklahoma.

Appendix B

Supplemental Figures for Chapter 2

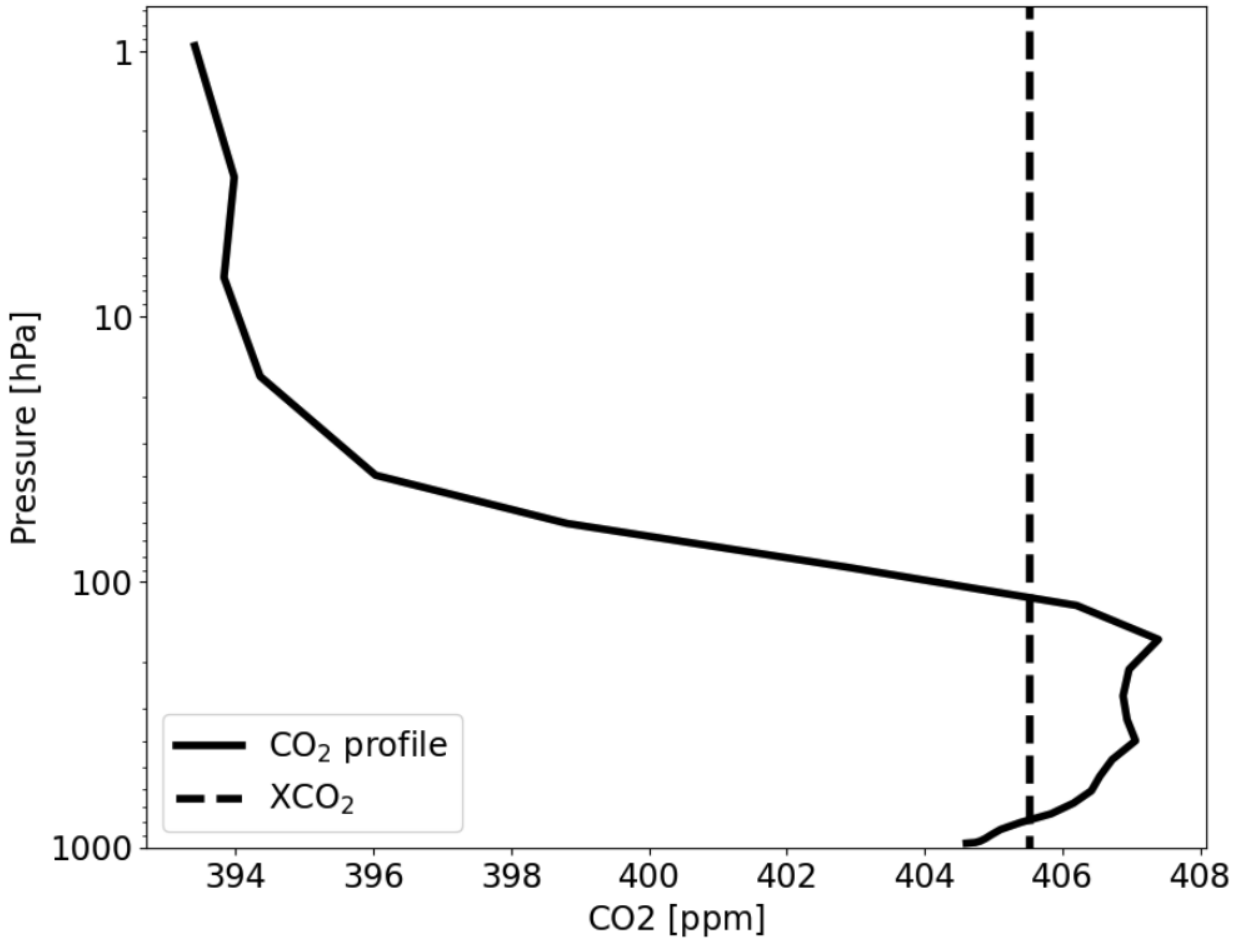

Figure B.1: The range of the bootstrapped diurnal cycles of XCO_2, shaded by the sample size of N. The shaded region represents the range of the middle 95th-percentile of the diurnal cycle of XCO_2 each hour. The yellow dashed line represents the "true" diurnal cycle at each respective site. The right (left) column represents bootstrapped data for (a. and b.) Park Falls, Wisconsin, US, (c. and d.) Lamont, Oklahoma, US, and (e. and f.) Darwin, Australia. Panels (a.), (c.), and (e.) represent TCCON data and panels (b.), (d.), and (f.) are from CT2019B.

References

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., et al. (2006, January 7). TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988-2003. Global Biogeochemical Cycles. American Geophysical Union (AGU). https://doi.org/10.1029/2004gb002439

Baker, D.F., H. Bösch, S.C. Doney, D. O'Brien, and D.S. Schimel, 2010: Carbon source/sink information provided by column CO2 measurements from the Orbiting Carbon Observatory, Atmosp. Chem. Physics, 10, 4145-4165. https://doi.org/10.5194/acp-10-4145-2010

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001, November). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. Bulletin of the American Meteorological Society. American Meteorological Society. https://doi.org/10.1175/1520-0477(2001)082¡2415:fantts¿2.3.co;2

Basu, S., Houweling, S., Peters, W., Sweeney, C., Machida, T., Maksyutov, S., et al. (2011, December 15). The seasonal cycle amplitude of total column CO2: Factors behind the model-observation mismatch. Journal of Geophysical Research: Atmospheres. American Geophysical Union (AGU). https://doi.org/10.1029/2011jd016124

Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., and Miller, J. B. (2018, May 24). The impact of transport model differences on CO2 surface flux estimates from OCO-2 retrievals of column average CO2 Atmospheric Chemistry and Physics. Copernicus GmbH. https://doi.org/10.5194/acp-18-7189-2018

Bolin, B. (1960, August). On the Exchange of Carbon Dioxide between the Atmosphere and the Sea. Tellus. Stockholm University Press. https://doi.org/10.1111/j.2153-3490.1960.tb01311.x

Bolin, B., and Keeling, C. D. (1963, July 1). Large-scale atmospheric mixing as deduced from the seasonal and meridional variations of carbon dioxide. Journal of Geophysical Research. American Geophysical Union (AGU). https://doi.org/10.1029/jz068i013p03899

Bonan, G. B., and Doney, S. C. (2018, February 2). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. Science. American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.aam8328

Broecker, W. S., Takahashi, T., Simpson, H. J., and Peng, T.-H. (1979, October 26). Fate of Fossil Fuel Carbon Dioxide and the Global Carbon Budget. Science. American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.206.4417.409

Callendar, G. S. (1938, April). The artificial production of carbon dioxide and its influence on temperature. Quarterly Journal of the Royal Meteorological Society. Wiley. https://doi.org/10.1002/qj.49706427503

Chandra, N., Patra, P. K., Niwa, Y., Ito, A., Iida, Y., Goto, D., et al. (2022, July 18). Estimated regional CO2 flux and uncertainty based on an ensemble of atmospheric CO2 inversions. Atmospheric Chemistry and Physics. Copernicus GmbH. https://doi.org/10.5194/acp-22-9215-2022

Chevallier, F., Engelen, R. J., and Peylin, P. (2005). The contribution of AIRS data to the estimation of CO2 sources and sinks. Geophysical Research Letters. American Geophysical Union (AGU). https://doi.org/10.1029/2005gl024229

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., et al. (2010). CO2 surface fluxes at grid point scale estimated from a global 21 year re-

analysis of atmospheric measurements. Journal of Geophysical Research, 115(D21), D21307. https://doi.org/10.1029/2010JD013887

Chevallier, F., and O'Dell, C. W. (2013). Error statistics of Bayesian CO2 flux inversion schemes as seen from GOSAT. Geophysical Research Letters, 40(6), 1252–1256. https://doi.org/10.1002/grl.50228

Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P. (2014). Toward robust and consistent regional CO2 flux estimates from in situ and spaceborne measurements of atmospheric CO2. Geophysical Research Letters, 41(3), 1065–1070. https://doi.org/10.1002/2013gl058772

Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J.-F. L., et al. (2012, November 1). Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations. Journal of Geophysical Research: Atmospheres. American Geophysical Union (AGU). https://doi.org/10.1029/2012jd018087

Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D. (2008). Orbiting Carbon Observatory: Inverse method and prospective error analysis. Journal of Geophysical Research: Atmospheres, 113(D5), D05305. https://doi.org/10.1029/2006jd008336

Corbin, K. D., Denning, A. S., Lu, L., Wang, J.-W., and Baker, I. T. (2008, January 16). Possible representation errors in inversions of satellite CO2 retrievals. Journal of Geophysical Research. American Geophysical Union (AGU). https://doi.org/10.1029/2007jd008716

Cressie, N., and Hawkins, D. M. (1980). Robust estimation of the variogram: I. Journal of the International Association for Mathematical Geology, 12(2), 115–125. https://doi.org/10.1007/BF01035243

Crisp, D., Atlas, R. M., Breon, F.-M., Brown, L. R., Burrows, J. P., Ciais, P., et al. (2004, January). The Orbiting Carbon Observatory (OCO) mission. Advances in Space Research. Elsevier BV. https://doi.org/10.1016/j.asr.2003.08.062

Crisp, D., Dolman, H., Tanhua, T., McKinley, G. A., Hauck, J., Bastos, A., et al. (2022, May 6). How Well Do We Understand the Land-Ocean-Atmosphere Carbon Cycle? Reviews of Geophysics. American Geophysical Union (AGU). https://doi.org/10.1029/2021rg000736

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., et al. (2019, in review). The 2015-2016 Carbon Cycle As Seen from OCO-2 and the Global In Situ Network. Atmospheric Chemistry and Physics Discussions, 1–79. https://doi.org/10.5194/acp-2019-87

Dai, A., Wang, J., Ware, R. H., and Van Hove, T. (2002). Diurnal variation in water vapor over North America and its implications for sampling errors in radiosonde humidity. Journal of Geophysical Research: Atmospheres, 107(D10), ACL 11-1-ACL 11-14. https://doi.org/10.1029/2001JD000642

Davis, K. J., Browell, E. V., Feng, S., Lauvaux, T., Obland, M. D., Pal, S., et al. (2021, September). The Atmospheric Carbon and Transport (ACT)-America Mission. Bulletin of the American Meteorological Society. American Meteorological Society. https://doi.org/10.1175/bams-d-20-0300.1

Denning, A. S., Fung, I. Y., and Randall, D. (1995, July). Latitudinal gradient of atmospheric CO2 due to seasonal exchange with land biota. Nature. Springer Science and Business Media LLC. https://doi.org/10.1038/376240a0

Desai, A. (2016). AmeriFlux US-PFa Park Falls/WLEF [Data set]. AmeriFlux; University of Wisconsin. https://doi.org/10.17190/amf/1246090

Dingle, A. N. (1954, November). The Carbon Dioxide Exchange Between the North Atlantic Ocean and the Atmosphere. Tellus. Stockholm University Press. https://doi.org/10.1111/j.2153-3490.1954.tb01129.x

Dickinson, R. E., and Cicerone, R. J. (1986, January). Future global warming from atmospheric trace gases. Nature. Springer Science and Business Media LLC. https://doi.org/10.1038/319109a0

Deutscher, N. M., Griffith, D. W. T., Bryant, G. W., Wennberg, P. O., Toon, G. C., Washenfelder, R. A., et al. (2010). Total column CO2 measurements at Darwin, Australia – site description and calibration against in situ aircraft profiles. Atmospheric Measurement Techniques, 3(4), 947–958. https://doi.org/10.5194/amt-3-947-2010

Deutscher, N., Notholt, J., Messerschmidt, J., Weinzierl, C., Warneke, T., Petri, C., et al. (2014). TCCON data from Bialystok, Poland, Release GGG2014R1. https://doi.org/10.14291/tccon.ggg2014.bialystok01.R1/1183984

Enting, I. G., and Mansbridge, J. V. (1989). Seasonal sources and sinks of atmospheric CO2: Direct inversion of filtered data. Tellus B: Chemical and Physical Meteorology, 41(2), 111–126. https://doi.org/10.3402/tellusb.v41i2.15056

Enting, I. G., Trudinge, C. M., and Francey, R. J. (1995, February). A synthesis inversion of the concentration and delta13 C of atmospheric CO2. Tellus B. Stockholm University Press. https://doi.org/10.1034/j.1600-0889.47.issue1.5.x

Enting, I. G. (2002, September). An empirical characterisation of signal versus noise in CO2 data. Tellus B. Stockholm University Press. https://doi.org/10.1034/j.1600-0889.2002.01303.x

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., et al. (2017, February 15). The Orbiting Carbon Observatory-2: first 18 months of science data products. Atmospheric Measurement Techniques. Copernicus GmbH. https://doi.org/10.5194/amt-10-549-2017

Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R. (2019, April 15). The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data. Atmospheric Measurement Techniques. Copernicus GmbH. https://doi.org/10.5194/amt-12-2341-2019

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022, November 11). Global Carbon Budget 2022. Earth System Science Data. Copernicus GmbH. https://doi.org/10.5194/essd-14-4811-2022

Geels, C., Doney, S. C., Dargaville, R., Brandt, J., and Christensen, J. H. (2004). Investigating the sources of synoptic variability in atmospheric CO2 measurements over the Northern Hemisphere continents: a regional model study. Tellus B: Chemical and Physical Meteorology, 56(1), 35–50. https://doi.org/10.3402/tellusb.v56i1.16399

Glover, D. M., Jenkins, W. J., and Doney, S. C. (2011). Modeling Methods for Marine Science. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511975721

Gloor, M., Dlugokencky, E., Brenninkmeijer, C., Horowitz, L., Hurst, D. F., Dutton, G., et al. (2007, August 15). Three-dimensional SF6data and tropospheric transport simulations: Signals, modeling accuracy, and implications for inverse modeling. Journal of Geophysical Research. American Geophysical Union (AGU). https://doi.org/10.1029/2006jd007973

Glueckauf, E. (1951). The Composition of Atmospheric Air. Compendium of Meteorology. American Meteorological Society. https://doi.org/10.1007/978-1-940033-70-9_1

Griffith, D. W. T., Deutscher, N., Velazco, V. A., Wennberg, P. O., Yavin, Y., Aleks, G. K., et al. (2014). TCCON data from Darwin, Australia, Release GGG2014R0. https://doi.org/10.14291/tccon.ggg2014.darwin01.R0/1149290

Guan, Y., Keppel-Aleks, G., Doney, S. C., Petri, C., Pollard, D., Wunch, D., et al. (2023, May 12). Characteristics of interannual variability in space-based XCO2 global observations. Atmospheric Chemistry and Physics. Copernicus GmbH. https://doi.org/10.5194/acp-23-5355-2023

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., et al. (2002, February). Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models. Nature. Springer Science and Business Media LLC. https://doi.org/10.1038/415626a

Hase, F., Blumenstock, T., Dohe, S., Gross, J., and Kiel, M. (2014). TCCON data from Karlsruhe, Germany, Release GGG2014R1. https://doi.org/10.14291/tccon.ggg2014.karlsruhe01.R1/1182416

Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., et al. (2015). An intercomparison of inverse models for estimating sources and sinks of CO2 using GOSAT measurements. Journal of Geophysical Research: Atmospheres, 120(10), 5253–5266. https://doi.org/10.1002/2014JD022962

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., et al. (2012, May). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. Ecological Modelling. Elsevier BV. https://doi.org/10.1016/j.ecolmodel.2012.02.004

Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., et al. (2013, December 17). The North American Carbon Program Multi-Scale Synthesis and Terrestrial Model Intercomparison Project – Part 1: Overview and experimental design. Geoscientific Model Development. Copernicus GmbH. https://doi.org/10.5194/gmd-6-2121-2013

Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., et al. (2017, July 6). Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. Scientific Reports. Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-017-03818-2

IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change[Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, In press, doi:10.1017/9781009157896.

Jacobson, A. R., Schuldt, K. N., Miller, J. B., Oda, T., Tans, P., Arlyn Andrews, et al. (2020). CarbonTracker CT2019B. NOAA Global Monitoring Laboratory. https://doi.org/10.25925/20201008

Kataoka, F., Crisp, D., Taylor, T., O'Dell, C., Kuze, A., Shiomi, K., et al. (2017, November 11). The Cross-Calibration of Spectral Radiances and Cross-Validation of CO2 Estimates from GOSAT and OCO-2. Remote Sensing. MDPI AG. https://doi.org/10.3390/rs9111158

Keeling, C. D. (1960, January). The Concentration and Isotopic Abundances of Carbon Dioxide in the Atmosphere. Tellus. Stockholm University Press. https://doi.org/10.3402/tellusa.v12i2.9366

Keeling, R. F., Piper, S. C., and Heimann, M. (1996). Global and hemispheric CO2 sinks deduced from changes in atmospheric O2 concentration. Nature, 381(6579), 218–221. https://doi.org/10.1038/381218a0

Keeling, C. D., Bacastow, R. B., Carter, A. F., Piper, S. C., Whorf, T. P., Heimann, M., et al. (1989). A three-dimensional model of atmospheric CO2 transport based on observed winds: 1. Analysis of observational data. Aspects of Climate Variability in the Pacific and the Western

Americas. American Geophysical Union. https://doi.org/10.1029/gm055p0165

Keeling, R. F., and Keeling, C. D. (2017). Atmospheric Monthly In Situ CO2 Data - Mauna Loa Observatory, Hawaii. In Scripps CO2 Program Data [Data set]. UC San Diego Library Digital Collections. https://doi.org/10.6075/J08W3BHW

Keppel-Aleks, G., Wennberg, P. O., and Schneider, T. (2011, April 18). Sources of variations in total column carbon dioxide. Atmospheric Chemistry and Physics. Copernicus GmbH. https://doi.org/10.5194/acp-11-3581-2011

Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., et al. (2012, March 1). The imprint of surface fluxes and transport on variations in total column carbon dioxide. Biogeosciences. Copernicus GmbH. https://doi.org/10.5194/bg-9-875-2012

Kerr, G. H., Waugh, D. W., Miller, S. M. (2021, January 9). Jet Stream-Surface Tracer Relationships: Mechanism and Sensitivity to Source Region. Geophysical Research Letters. American Geophysical Union (AGU). https://doi.org/10.1029/2020gl090714

Kulawik, S., Wunch, D., O'Dell, C., Frankenberg, C., Reuter, M., Oda, T., et al. (2016, February 29). Consistent evaluation of ACOS-GOSAT, BESD-SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON. Atmospheric Measurement Techniques. Copernicus GmbH. https://doi.org/10.5194/amt-9-683-2016

Lauvaux, T., and Davis, K. J. (2014, January 27). Planetary boundary layer errors in mesoscale inversions of column-integrated CO2 measurements. Journal of Geophysical Research: Atmospheres. American Geophysical Union (AGU). https://doi.org/10.1002/2013jd020175

Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., van Heuven, S., et al. (2015). The reinvigoration of the Southern Ocean carbon sink. Science, 349(6253), 1221–1224. https://doi.org/10.1126/science.aab2620

Liang, A., Gong, W., Han, G., and Xiang, C. (2017, October 10). Comparison of Satellite-Observed XCO2 from GOSAT, OCO-2, and Ground-Based TCCON. Remote Sensing. MDPI AG. https://doi.org/10.3390/rs9101033

Masarie, K. A., Pétron, G., Andrews, A., Bruhwiler, L., Conway, T. J., Jacobson, A. R., et al. (2011). Impact of CO2 measurement bias on CarbonTracker surface flux estimates. Journal of Geophysical Research, 116(D17), D17305. https://doi.org/10.1029/2011JD016270

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., et al. (2006). North American Regional Reanalysis. Bulletin of the American Meteorological Society, 87(3), 343–360. https://doi.org/10.1175/bams-87-3-343

Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., et al. (2007). Precision requirements for space-based data. Journal of Geophysical Research: Atmospheres, 112(D10), D10314. https://doi.org/10.1029/2006JD007659

Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I., and Liu, J. (2015). Biases in atmospheric CO2; estimates from correlated meteorology modeling errors. Atmospheric Chemistry and Physics, 15(5), 2903–2914. https://doi.org/10.5194/acp-15-2903-2015

Miller, S. M., Michalak, A. M., Yadav, V., and Tadić, J. M. (2018). Characterizing biospheric carbon balance using CO2; observations from the OCO-2 satellite. Atmospheric Chemistry and Physics, 18(9), 6785–6799. https://doi.org/10.5194/acp-18-6785-2018

Miller, S. M., and Michalak, A. M. (2020, January 8). The impact of improved satellite retrievals on estimates of biospheric carbon balance. Atmospheric Chemistry and Physics. Copernicus GmbH. https://doi.org/10.5194/acp-20-323-2020

Muntean, M., Janssens-Maenhout, G., Song, S., Giang, A., Selin, N. E., Zhong, H., et al. (2018, July). Evaluating EDGARv4.tox2 speciated mercury emissions ex-post scenarios and their impacts on modelled global and regional wet deposition patterns. Atmospheric Environment. Elsevier BV. https://doi.org/10.1016/j.atmosenv.2018.04.017

Mitchell, K. A., Doney, S. C., and Keppel-Aleks, G. (2023, January 27). Characterizing Average Seasonal, Synoptic, and Finer Variability in Orbiting Carbon Observatory-2 XCO2 Across North America and Adjacent Ocean Basins. Journal of Geophysical Research: Atmospheres. American Geophysical Union (AGU). https://doi.org/10.1029/2022jd036696

Nelson, R. R., Crisp, D., Ott, L. E., and O'Dell, C. W. (2016). High-accuracy measurements of total column water vapor from the Orbiting Carbon Observatory-2. Geophysical Research Letters, 43(23), 12,212-261,269. https://doi.org/10.1002/2016gl071200

NOAA National Centers for Environmental Information (2023). State of the Climate: Global Climate Report for 2022. Accessed March 1, 2023, from https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202213 .

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., et al. (2012). The ACOS CO2 retrieval algorithm-Part 1: Description and validation against synthetic observations. Atmospheric Measurement Techniques, 5(1), 99–121. https://doi.org/10.5194/amt-5-99-2012

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., et al. (2018). Improved Retrievals of Carbon Dioxide from the Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. Atmospheric Measurement Techniques Discussions, 1–57. https://doi.org/10.5194/amt-2018-257

Olsen, S. C., and J. T. Randerson (2004), Differences between surface and column atmospheric CO2 and implications for carbon cycle research, J. Geophys. Res., 109, D02301. https://doi.org/10.1029/2003JD003968

Parazoo, N. C., Denning, A. S., Kawa, S. R., Corbin, K. D., Lokupitiya, R. S., and Baker, I. T. (2008). Mechanisms for synoptic variations of atmospheric CO2 in North America, South America and Europe. Atmospheric Chemistry and Physics, 8(23), 7239–7254. https://doi.org/10.5194/acp-8-7239-2008

Parazoo, N. C., Denning, A. S., Berry, J. A., Wolf, A., Randall, D. A., Kawa, S. R., et al. (2011). Moist synoptic transport of CO2 along the mid-latitude storm track. Geophysical Research Letters, 38(9), L09804. https://doi.org/10.1029/2011gl047238

Pan, G., Xu, Y., and Ma, J. (2021, January). The potential of CO2 satellite monitoring for climate governance: A review. Journal of Environmental Management. Elsevier BV. https://doi.org/10.1016/j.jenvman.2020.111423

Parker, R., Boesch, H., Cogan, A., Fraser, A., Feng, L., Palmer, P. I., et al. (2011, August). Methane observations from the Greenhouse Gases Observing SATellite: Comparison to ground-based TCCON data and model calculations. Geophysical Research Letters. American Geophysical Union (AGU). https://doi.org/10.1029/2011gl047871

Peters, W. (2004). Toward regional-scale modeling using the two-way nested global model TM5: Characterization of transport using SF6. Journal of Geophysical Research. American Geophysical Union (AGU). https://doi.org/10.1029/2004jd005020

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., et al. (2005). An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations. Journal of Geophysical Research, 110(D24), D24304.

https://doi.org/10.1029/2005jd006157

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., et al. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. Proceedings of the National Academy of Sciences, 104(48), 18925–18930. https://doi.org/10.1073/pnas.0708986104

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., et al. (2013). Global atmospheric carbon budget: Results from an ensemble of atmospheric CO2 inversions. Biogeosciences, 10(10), 6699–6720. https://doi.org/10.5194/bg-10-6699-2013

Rayner, P. J., and O'Brien, D. M. (2001). The utility of remotely sensed CO2 concentration data in surface source inversions. Geophysical Research Letters, 28(1), 175–178. https://doi.org/10.1029/2000gl011912

Revelle, R. and Suess, H. E. (1957, February). Carbon Dioxide Exchange Between Atmosphere and Ocean and the Question of an Increase of Atmospheric CO2 during the Past Decades. Tellus. Stockholm University Press. https://doi.org/10.1111/j.2153-3490.1957.tb01849.x

Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., et al. (2015, December 11). Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean CO2 Mapping intercomparison (SOCOM). Biogeosciences. Copernicus GmbH. https://doi.org/10.5194/bg-12-7251-2015

Rodgers, C. D. (2000). Inverse Methods for Atmospheric Sounding: Theory and Practice. Singapore: World Scientific. https://doi.org/10.1142/9789812813718_0001

Ross, A. N., Wooster, M. J., Boesch, H., and Parker, R. (2013). First satellite measurements of carbon dioxide and methane emission ratios in wildfire plumes. Geophysical Research Letters, 40(15), 4098–4102. https://doi.org/10.1002/grl.50733

Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., et al. (2019, April). Quantifying the Impact of Atmospheric Transport Uncertainty on CO 2 Surface Flux Estimates. Global Biogeochemical Cycles. American Geophysical Union (AGU). https://doi.org/10.1029/2018gb006086

Sherlock, V., Connor, B., Robinson, J., Shiona, H., Smale, D., and Pollard, D. (2014). TCCON data from Lauder, New Zealand, 120HR, Release GGG2014R0. https://doi.org/10.14291/tccon.ggg2014.lauder01.R0/1149293

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., et al. (2007, June 22). Weak Northern and Strong Tropical Land Carbon Uptake from Vertical Profiles of Atmospheric CO 2. Science. American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.1137004

Sussmann, R., and Rettinger, M. (2014). TCCON data from Garmisch, Germany, Release GGG2014R0. https://doi.org/10.14291/tccon.ggg2014.garmisch01.R0/1149299

Tans, P. P., Fung, I. Y., and Takahashi, T. (1990). Observational Contrains on the Global Atmospheric CO2 Budget. Science, 247(4949), 1431–1438. https://doi.org/10.1126/science.247.4949.1431

Tans, P. P., and Keeling, C. D. (2017). Atmospheric Monthly In Situ CO2 Data - Mauna Loa Observatory, Hawaii. In Scripps CO2 Program Data [Data set]. UC San Diego Library Digital Collections. https://doi.org/10.6075/j08w3bhw

Torres, A. D., Keppel-Aleks, G., Doney, S. C., Fendrock, M., Luis, K., De Mazière, M., et al. (2019, September). A Geostatistical Framework for Quantifying the Imprint of Mesoscale

Atmospheric Transport on Satellite Trace Gas Retrievals. Journal of Geophysical Research: Atmospheres. American Geophysical Union (AGU). https://doi.org/10.1029/2018jd029933

USGCRP, 2018: Second State of the Carbon Cycle Report (SOCCR2): A Sustained Assessment Report. [Cavallaro, N., G. Shrestha, R. Birdsey, M. A. Mayes, R. G. Najjar, S. C. Reed, P. Romero-Lankao, and Z. Zhu (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 878 pp., doi: 10.7930/SOCCR2.2018

Warneke, T., Messerschmidt, J., Notholt, J., Weinzierl, C., Deutscher, N., Petri, C., et al. (2014). TCCON data from Orleans, France, Release GGG2014R0. https://doi.org/10.14291/tccon.ggg2014.orleans01.R0/1149276

Washenfelder, R. A., Toon, G. C., Blavier, J.-F., Yang, Z., Allen, N. T., Wennberg, P. O., Vay, S. A., Matross, D. M., and Daube, B. C. (2006). Carbon dioxide column abundances at the Wisconsin Tall Tower site. In Journal of Geophysical Research (Vol. 111, Issue D22). American Geophysical Union (AGU). https://doi.org/10.1029/2006jd007154

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R., et al. (2014). TCCON data from Park Falls, Wisconsin, USA, Release GGG2014R0. https://doi.org/10.14291/tccon.ggg2014.parkfalls01.R0/1149161

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J.-F., Toon, G. C., Allen, N., et al. (2014). TCCON data from Lamont, Oklahoma, USA, Release GGG2014R0. https://doi.org/10.14291/tccon.ggg2014.lamont01.R0/1149159

Williams, I. N., Riley, W. J., Torn, M. S., Biraud, S. C., and Fischer, M. L. (2014). Biases in regional carbon budgets from covariation of surface fluxes and weather in transport model inversions. Atmospheric Chemistry and Physics, 14(3), 1571–1585. https://doi.org/10.5194/acp-14-1571-2014

Wofsy, S. C. (2011). HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369(1943), 2073–2086. https://doi.org/10.1098/rsta.2010.0313

Wofsy, S., Daube, B., Jimenez, R., Kort, E., Pittman, J., Park, S., et al. (2017). HIPPO Pressure-Weighted Mean Total, 10-km, and 100-m Interval Column Concentrations. Version 1.0 (Data set). UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.3334/cdiac/hippo_011. Accessed 03 Mar 2018.

Worden, R. J., Doran, G., Kulawik, S., Eldering, A., Crisp, D., Frankenberg, C., et al. (2017). Evaluation and attribution of OCO-2 XCO2 uncertainties. Atmospheric Measurement Techniques, 10(7), 2759–2771. https://doi.org/10.5194/amt-10-2759-2017

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., et al. (2010). Calibration of the Total Carbon Column Observing Network using aircraft profile data. Atmospheric Measurement Techniques, 3(5), 1351–1362. https://doi.org/10.5194/amt-3-1351-2010

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., et al. (2011, May 28). The Total Carbon Column Observing Network. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. The Royal Society. https://doi.org/10.1098/rsta.2010.0240

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G., and Wennberg, P. O. (2015). The Total Carbon Column Observing Network's GGG2014 Data Version. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee,

U.S.A. https://doi.org/10.14291/tccon.ggg2014.documentation.R0/1221662

Wunch, D., P. O. Wennberg, G. B. Osterman, B. Fisher, B. Naylor, C. M. Roehl, C. O'Dell, L. Mandrake, C. Viatte, M. Kiel, D. W. T. Griffith, N. M. Deutscher, V. A. Velazco, J. Notholt, T. Warneke, C. Petri, M. De Maziere, M. K. Sha, R. Sussmann, M. Rettinger, D. Pollard, J. Robinson, I. Morino, O. Uchino, F. Hase, T. Blumenstock, D. G. Feist, S. G. Arnold, K. Strong, J. Mendonca, R. Kivi, P. Heikkinen, L. Iraci, J. Podolske, P. W. Hillyard, S. Kawakami, M. K. Dubey, H. A. Parker, E. Sepulveda, O. E. Garc?a, Y. Te, P. Jeseck, M. R. Gunson, D. Crisp, and A. Eldering (2017), Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON, Atmos. Meas. Tech., 10(6), 2209–2238, doi:10.5194/amt-10-2209-2017.

Yang, D., Liu, Y., Cai, Z., Chen, X., Yao, L., and Lu, D. (2018). First Global Carbon Dioxide Maps Produced from TanSat Measurements. Advances in Atmospheric Sciences, 35(6), 621–623. https://doi.org/10.1007/s00376-018-7312-6

Yang, Z., Washenfelder, R. A., Keppel-Aleks, G., Krakauer, N. Y., Randerson, J. T., Tans, P. P., et al. (2007, June 23). New constraints on Northern Hemisphere growing season net flux. Geophysical Research Letters. American Geophysical Union (AGU). https://doi.org/10.1029/2007gl029742

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S. (2009). Global Concentrations of CO2 and CH4 Retrieved from GOSAT: First Preliminary Results. SOLA. Meteorological Society of Japan. https://doi.org/10.2151/sola.2009-041

Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., et al. (2013, June 7). Improvement of the retrieval algorithm for GOSAT SWIR XCO2 and XCH4; and their validation using TCCON data. Atmospheric Measurement Techniques. Copernicus GmbH. https://doi.org/10.5194/amt-6-1533-2013