

UNIVERSITY OF OKLAHOMA  
GRADUATE COLLEGE

CARBON-BASED POLLUTANT ANALYSIS AND REMOTE SENSOR  
VALIDATION USING COLUMN-OBSERVING FOURIER TRANSFORM  
INFRARED (FTIR) SPECTROMETERS DURING THE TRACER CAMPAIGN

A THESIS  
SUBMITTED TO THE GRADUATE FACULTY  
in partial fulfillment of the requirements for the  
Degree of  
MASTER OF SCIENCE

By

ELIZABETH SPICER  
Norman, Oklahoma  
2023

CARBON-BASED POLLUTANT ANALYSIS AND REMOTE SENSOR  
VALIDATION USING COLUMN-OBSERVING FOURIER TRANSFORM  
INFRARED (FTIR) SPECTROMETERS DURING THE TRACER CAMPAIGN

A THESIS APPROVED FOR THE  
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Feng Xu (Chair)

Dr. Sean Crowell

Dr. Petra Klein

© Copyright by ELIZABETH SPICER 2023  
All Rights Reserved.

## Acknowledgments

I would like to express my gratitude to each of the members in my committee for their consistent support of my research and professional growth. Dr. Sean Crowell, my advisor, has provided useful guidance since day one. Acclimating to the world of meteorology and carbon cycle research would have been an insurmountable task without his direction. My committee chair, Dr. Feng Xu, and third committee member, Dr. Petra Klein, have both shown immense support for my research. My graduate school experience would not be the same without each of my committee members. I am eternally grateful for their guidance, kindness, and understanding.

I would also like to thank the organizations that facilitated this work. I acknowledge funding from National Aeronautics and Space Administration (NASA) through the GeoCarb Mission under award 80LARC17C0001. Additionally, some of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSKER) at the University of Oklahoma (OU). This research was supported by and data were obtained from the Atmospheric Radiation Measurement (ARM) User Facility, a U.S. Department of Energy (DOE) Office of Science User Facility managed by the Biological and Environmental Research Program. Additionally, thanks to the National Oceanic and Atmospheric Administration (NOAA) National Severe Storms Laboratory (NSSL), who provided data from the Collaborative Lower Atmospheric Mobile Profiling Systems (CLAMPS) collected during TRACER for analysis in this work.

I would also like to thank those who have worked with me on EM27/SUN deployments and data analysis. First, thank you to Dr. Wesley T Honeycutt, for being a fantastic mentor and constructive critic. Thank you to Dr. Nicole Jacobs, whose experience and data processing assistance were instrumental to my research progress. Thank you to Dr. Debra Wunch of the University of Toronto and the other researchers within the EM27/SUN and TCCON communities that have provided invaluable insight and assistance. Huge thanks to all of the incredible volunteers for the GeoCarb-TRACER Campaign, Dr. Gregory Osterman and others from the



NASA Jet Propulsion Laboratory for their OCO-related assistance and for loaning their EM27/SUN instruments, Dr. Frank Hase and colleagues of Karlsruhe Institute of Technology for loaning an EM27/SUN and troubleshooting help, the folks at the ARM Southern Great Plains site, the organizers and operators of ARM TRACER, as well as Dr. James Flynn from the University of Houston and his team involved in ARM TRACER. Without the support of ARM and James, my field research would not have been possible.

Finally, thank you to all of those who have supported me personally and professionally through the ups and downs of life as a graduate student. My parents, friends, and kitty Nara have all provided immense support that I am incredibly grateful for. Thank you to Andrew Shearer in particular, my partner, motivator, and late night coding buddy, for all of the above and more.

# Table of Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>List Of Tables</b>	<b>viii</b>
<b>List Of Figures</b>	<b>ix</b>
<b>Abstract</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Chemicals of Interest and Their Impacts on Climate Change . . . . .	1
1.1.1 Carbon Dioxide . . . . .	2
1.1.2 Methane . . . . .	3
1.1.3 Carbon Monoxide . . . . .	4
1.2 The Need for Continued Carbon-Based Pollutant Monitoring . . . . .	5
1.3 Pollutant Monitoring During the TRACER Campaign . . . . .	7
<b>2 Methods</b>	<b>10</b>
2.1 Instrumentation for Campaigns . . . . .	10
2.1.1 EM27/SUN Spectrometers . . . . .	10
2.1.1.1 EM27/SUN Deployment Methods . . . . .	13
2.1.2 Secondary Instrumentation . . . . .	15
2.2 Experimental Design of Field Campaigns . . . . .	15
2.2.1 TCCON Calibration Campaign . . . . .	16
2.2.2 GeoCarb-TRACER Campaign . . . . .	18
2.2.2.1 OCO-2/3 Collocation . . . . .	24
2.3 EM27/SUN Retrieval Algorithms . . . . .	25
2.4 Data Analysis Techniques . . . . .	29
2.4.1 Choice of Machine Learning Algorithms . . . . .	29
2.4.2 Data Standardization for Clustering Applications . . . . .	31
2.4.3 Hyperparameter Tuning for DBSCAN and UMAP . . . . .	33
2.4.4 Application of Machine Learning Techniques . . . . .	35
<b>3 EM27/SUN Data Quality Assurance</b>	<b>39</b>
3.1 Quality Control Filtering . . . . .	39
3.1.1 Combating Solar Tracking Issues . . . . .	41

3.1.2	Filtering Artificially Heightened Concentrations . . . . .	43
3.1.3	An Exception to the Rule . . . . .	46
3.2	Retrieval Algorithm Comparison . . . . .	49
3.2.1	Evaluation of Solar Zenith Angle Dependence . . . . .	51
3.2.2	EM27/SUN Comparison to TCCON . . . . .	52
3.2.3	EM27/SUN Intercomparison . . . . .	54
3.2.4	Correction Factors for Instrument Bias . . . . .	55
<b>4</b>	<b>TRACER Campaign Results</b>	<b>58</b>
4.1	Seasonal Trends during TRACER . . . . .	58
4.2	Sub-Seasonal Pollutant Analysis with UMAP and DBSCAN . . . . .	64
4.2.1	The Process of Clustering: High Variability $XCO_2$ as a Representative Sample . . . . .	64
4.2.2	$XCO_2$ Cluster Analysis Findings . . . . .	70
4.3	Satellite Validation . . . . .	74
<b>5</b>	<b>Conclusions</b>	<b>79</b>
5.1	Summary of Work and Findings . . . . .	79
5.2	Future Work . . . . .	82
	<b>References</b>	<b>84</b>
	<b>Appendix</b>	<b>90</b>
A	Lessons Learned from the GeoCarb-TRACER Campaign . . . . .	90
B	Procedures for EM27/SUN Operation . . . . .	96
B.1	Description of the Secondary Instrumentation Suite for Remote Deployments . . . . .	96
B.2	Combined Procedures for Tabletop and Enclosure-Based Deployments . . . . .	97
B.3	Troubleshooting Guide . . . . .	122
C	Procedures for Retrieval Algorithm Processing . . . . .	125
C.1	PROFFAST Version 2.2 Procedure . . . . .	125
C.2	GGG2020 Procedure . . . . .	129

# List Of Tables

2.1	Sources and temporal resolution of near-surface meteorological data at EM27/SUN deployment sites during the TRACER Campaign . . . .	22
2.2	State vector space for UMAP reduction of the TRACER dataset when the feature of interest is $XCO_2$ . . . . .	31
3.1	Bias of guest EM27/SUN instruments from Blanche estimated during the TCCON calibration campaign such that $Xgas_{Blanche-corrected\ guest} = Xgas_{guest} - bias$ . . . . .	56
3.2	Difference between Blanche and TCCON $Xgas$ during the TCCON calibration campaign . . . . .	57
4.1	Description of the high and low variability daytime datasets collected between 6/23/22 and 6/25/22 . . . . .	62
4.2	High variability $XCO_2$ clusters generated using recursive DBSCAN .	67
4.3	OCO-2 and OCO-3 collocations during the TRACER Campaign . . .	75
4.4	Bias of OCO-2 and OCO-3 compared to EM27/SUN instruments for TRACER collocations . . . . .	76
A.1	Summarized feedback from the lessons learned session for the GeoCarb-TRACER Campaign . . . . .	91

# List Of Figures

1.1	The unofficial GeoCarb-TRACER Campaign logo, designed and drawn by myself . . . . .	7
1.2	Blanche the EM27/SUN spectrometer, actively observing trace gases during a teaching demonstration . . . . .	8
2.1	A cartoon of an EM27/SUN deployment. The EM27/SUN has been deployed during clear-sky conditions so that direct sunlight (yellow beam) can be observed by the EM27/SUN. The primary and secondary mirrors of the external optics (yellow discs on top of the EM27/SUN) reflect the direct sunlight to the tertiary external mirror (not pictured) and then into the instrument cavity following the red dotted arrows. The internal optics and sensors of the EM27/SUN record information about the direct radiation they observe. Any trace gases present will alter the radiation that reaches the instrument. Emission sources, generalized by the industrial plant on the right, may produce large amounts of these trace gases, represented by the simplified gray plume from the smoke stack. . . . .	11
2.2	EM27/SUN enclosure near instrument trailers during the TRACER Campaign . . . . .	13
2.3	CAD drawing of the original EM27/SUN enclosure design created by Dr. Bianca Baier, Tim Newberger, and Philip Handley of the NOAA Global Monitoring Laboratory. . . . .	14
2.4	Deployment of Blanche at the Guest Instrument Facility during the TCCON calibration campaign. The EM27/SUN can be seen on top of the platform in the picture on the left. The picture on the right demonstrates a typical "tabletop" EM27/SUN deployment on the raised platform. . . . .	16
2.5	Deployment of Blanche and guest EM27/SUN instruments next to the ARM SGP TCCON station (white building) during the TCCON calibration campaign during October 17-19, 2022. . . . .	17
2.6	TRACER Campaign site locations in the Houston, TX area . . . . .	21
2.7	Central Houston wind roses generated using wind speed data at approximately 10 m elevation collected at the Houston Dunn ASOS station (identifier MCJ) from 2009 through 2021 (Iowa Environmental Mesonet 2022) . . . . .	23

2.8	One EM27/SUN interferogram captured in OPUS, where the bottom x-axis represents wavenumber ( $cm^{-1}$ ), the upper x-axis represents wavelength ( $\mu m$ ), and the y-axis is the intensity of the solar radiation . . .	27
2.9	The sorted $k$ -distance graph for the high variability $XCO_2$ dataset, where $k = 20$ . The optimal value for $eps$ should be approximately 0.26.	34
2.10	The distribution the of the number of data points in the daytime datasets collected during the TRACER Campaign. Only daytime datasets with greater than 30 one-minute averaged data points are included. . . . .	35
2.11	Flowchart of the data preparation and grouping utilized during cluster analysis. Specifics about the UMAP and DBSCAN implementations can be found in Figure 2.12. . . . .	38
2.12	Flowchart of the UMAP and DBSCAN applications utilized for this study	38
3.1	Preliminary data quality and solar tracking failure filters applied to EM27/SUN data collected on May 29, 2022 by Blanche the GeoCarb EM27/SUN during the TRACER Campaign at the UH Coastal Center site . . . . .	42
3.2	Application of the percent change filter on $XCO$ and $Xluft$ data collected on May 29, 2022 by the GeoCarb EM27/SUN during the TRACER Campaign at the UH Coastal Center site. These data were previously filtered with the preliminary data quality and solar tracking failure filters as show in Figure 3.1. . . . .	44
3.3	Comparison of May 29, 2022 EM27/SUN data with preliminary and entire collection of data quality filters applied . . . . .	45
3.4	Blanche and the two guest EM27/SUN instruments at SGP on 10/17/22, where a plume of smoke blew into the viewing path of the instruments.	46
3.5	$XCO$ data collected with Blanche from October 17, 2022 with all filters except for the $XCO$ percent change filter applied to the points in blue, resulting in the filtered (green) points. . . . .	47
3.6	Blanche $XCO$ data from October 17, 2022 with all filters, including the $XCO$ percent change filter, applied. There is a lack of data preserved (green points) while observing the smoke plume. . . . .	47
3.7	Bias in $XCO$ between Blanche and TCCON from October 17-19, 2022. The data from October 17 were removed from the box plot on the right, while data from all three days are included in the box plot on the left.	48
3.8	Preliminarily filtered $Xluft$ retrieved with GGG. These data were collected with Blanche on October 17, 2022. . . . .	50

3.9	Preliminarily filtered <i>Xluft</i> retrieved with a modified version of GGG with an adjusted $\delta\nu$ parameter to remove the sawtooth pattern from the data as well as air-mass independent and dependent correction factors from (Mostafavi Pak et al. 2023). These data were collected with Blanche on October 17, 2022. . . . .	51
3.10	Solar zenith angle dependence during the TCCON calibration campaign using Blanche’s data retrieved using PROFFAST version 2.2 (left), unmodified GGG2020 (center), and GGG2020 with $\delta\nu$ , AICF, and ADCF modifications (right) . . . . .	52
3.11	Bias of Blanche compared to the SGP TCCON during the calibration campaign for $XCO_2$ (left), $XCO$ (center), and $XCH_4$ (right). The green shading represents the estimated uncertainty of TCCON as reported by Mostafavi Pak et al. (2023). The sample sizes ( $n$ ) reported for $XCO_2$ are consistent for all <i>Xgas</i> . . . . .	53
3.12	Bias of Blanche compared to the guest EM27/SUN instruments using one-minute averaged data for $XCO_2$ . The green band represents the bias between the SGP TCCON and Blanche, reported in Table 3.2. . . . .	54
3.13	Bias of Blanche compared to the guest EM27/SUN instruments using one-minute averaged data for $XCO$ (left) and $XCH_4$ (right). The number of samples per comparison is consistent with Figure 3.12. The green bands represents the biases between the SGP TCCON and Blanche for each <i>Xgas</i> , reported in Table 3.2. . . . .	55
3.14	Bias of Blanche compared to the SGP TCCON during the TCCON calibration campaign. Instruments were compared using ten-minute averaged $XCO_2$ data . . . . .	57
4.1	All quality filtered EM27/SUN $XCO_2$ data collected during the TRACER Campaign . . . . .	59
4.2	All EM27/SUN $XCO_2$ (top), $XCO$ (middle), and $XCH_4$ (bottom) data collected between 10:00 and 16:00 local time during the TRACER Campaign. Each box represents data collected on a single day at one deployment site. Data have been separated into high and low variability daytime datasets as defined in Section 2.4. 6/23/22 through 6/25/22 have multiple box plots, as multiple EM27/SUN instruments were deployed at different sites during these days. . . . .	61
4.3	All meteorological data collected between 10:00 and 16:00 local time during the TRACER Campaign. Each box represents data collected on an individual day. . . . .	63
4.4	Histograms of one minute averaged $XCO_2$ , $XCO$ , and $XCH_4$ standard deviations where data have been limited to a local time window from 10:00 to 16:00 . . . . .	65

4.5	UMAP three-dimensional representation of the high variability $XCO_2$ group, colored by the feature of interest, or standardized $XCO_2$ ( $z_{XCO_2}$ ).	65
4.6	Optimized DBSCAN clustering results of the high variability $XCO_2$ dataset. Each cluster is represented by a different color, where the -1 cluster contain data the algorithm considers noise.	66
4.7	Recursive DBSCAN clustering results of the high variability $XCO_2$ dataset	67
4.8	DBSCAN clusters of the high variability $XCO_2$ dataset. The dashed line represents $\overline{XCO_2}$ for the dataset and the shaded area represents one standard deviation above and below this mean.	68
4.9	Density plots of various state vector space variables for the high variability $XCO_2$ dataset. The only clusters shown are the high and low concentration clusters.	69
4.10	Correlation between all data in cluster 4 of the high variability $XCO_2$ dataset	71
4.11	$XCO_2$ plotted against $XCO$ (left column), $XH_2O$ (middle column), and $XCH_4$ (right column). The plots in the top row are of data from the high variability, high concentration $XCO_2$ cluster 4, while the bottom row plots were created using the high variability, low concentration $XCO_2$ cluster 0.	72
4.12	$XCO_2$ and wind direction plotted from cluster 4 (top) and cluster 0 (bottom). The time period shown for cluster 4 plot begins before the EM27/SUN data collection began.	74
4.13	Box plots demonstrate the distribution of quality filtered $XCO_2$ from OCO-2, OCO-3, and the collocated EM27/SUN instruments during the TRACER Campaign for each collocation period. EM27/SUN observations have been separated based on the satellite they were collocated with. The number of samples that met the collocation criteria and passed quality filtering are given as $n$	75
4.14	Comparison of EM27/SUN and OCO-2/3 observed $XCO_2$ using the preliminarily filtered PROFFAST version 2.2 dataset (left) and the quality-filtered modified GGG2020 EM27/SUN dataset (right). Note that the left plot contains data from 8/4/22, which was later removed from analysis due to low OCO-3 sample size.	77



## Abstract

Greenhouse gases methane ( $\text{CH}_4$ ), and carbon dioxide ( $\text{CO}_2$ ), along with carbon monoxide ( $\text{CO}$ ), while produced by both anthropogenic and natural sources, all contribute to atmospheric warming. Additionally,  $\text{CO}$  poses health risks to individuals. If the atmospheric dynamics in a region are understood, it should be possible to use regional-scale sensors to evaluate emissions from upwind sources with respect to atmospheric variability.

The GeoCarb-TRACER Campaign was designed to observe these trace gases and their dynamics as a part of the TRacking Aerosol Convection interactions ExpeRiment (TRACER), organized by the U.S. Department of Energy’s (DOE) Atmospheric Radiation Measurement (ARM) user facility in Houston, TX. During this campaign, portable Bruker EM27/SUN Fourier transform spectrometers were deployed at various urban and background sites in the summer of 2022. Each EM27/SUN captures high-resolution ( $0.5 \text{ cm}^{-1}$ ) spectra in the near- and shortwave-infrared wavelength range. Multiple EM27/SUN spectrometers were deployed simultaneously alongside instruments gathering boundary layer, aerosol, and near-surface meteorological information. Spectra were analyzed to retrieve column-averaged concentrations of  $\text{CO}_2$ ,  $\text{CO}$ , and  $\text{CH}_4$ , in reference to water vapor in the atmosphere. Researchers used unsupervised machine learning techniques to identify relationships between heightened EM27/SUN concentration measurements and local meteorological and anthropogenic source information. Targeting certain conditions for in-depth case studies identified by the machine learning analysis of local emission sources and co-emitted pollutants will inform further study. This cluster analysis approach highlights potential relationships between heightened EM27/SUN concentrations, surface meteorological conditions, and local industrial sources that may have been overlooked with a daily case study analysis.

Each EM27/SUN instrument was validated by intra- and inter-device comparison using a higher-resolution spectrometer from the TCCON (Total Carbon Column Observing Network) corrected to World Meteorological Organization (WMO) standards

to ensure data accuracy, demonstrating minimal bias between instruments using the GGG2020 retrieval algorithm to process raw data. Empirical modifications to the retrieval algorithm were implemented to further correct the EM27/SUN data for solar zenith angle dependence and bring the retrieved concentration data up to the WMO standard, ultimately providing the most accurate representation of the data when compared to TCCON. Additionally, a series of automated data quality filters were developed to remove erroneous data during loss of tracking episodes. EM27/SUN TRACER data also validated the Orbiting Carbon Observatories, OCO-2 and OCO-3, indicating the bias between the EM27/SUN instruments and satellites were small, supporting the assertion that the OCO instruments provide an accurate representation of CO<sub>2</sub> concentrations in the atmosphere even in proximity to urban and industrial pollutant sources.

# Chapter 1

## Introduction

Carbon-based trace gases play key roles in the important issues of climate change and air quality. Though many trace gases exist naturally, anthropogenic sources can enhance concentrations of these atmospheric constituents, which have been proven to contribute to the warming of Earth's atmosphere (IPCC 2021). However, many questions remain surrounding the variability of these constituents, how we can accurately measure them, and how they impact both the climate and human health. Monitoring these pollutants using multiple data sources can build understanding about data precision and accuracy, source/sink relationships, climate forcing, and potential impacts on atmospheric warming. Additionally, validated instrument data can inform and monitor the effect of mitigation efforts to limit the impacts of increasing pollutant concentrations.

### 1.1 Chemicals of Interest and Their Impacts on Climate Change

There are many airborne pollutants whose concentrations have grown alongside mean atmospheric temperatures since the rise of industrialized society around the late 1700s (Seinfeld and Pandis 2016). When emitted into the atmosphere in excessive amounts, many of these pollutants, called climate forcers, contribute to an increased

”greenhouse effect.” This effect occurs when pollutant molecules absorb and re-emit longwave radiation (also referred to as heat energy) emitted from Earth’s surface, causing the temperature of the atmosphere to rise (Ramanathan et al. 2019). Though all climate forcers contribute to atmospheric warming, four species are responsible for the majority of the greenhouse effect: carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), and fluorinated gases, in descending order of contribution (Seinfeld and Pandis 2016; US EPA 2023b). While many atmospheric pollutants play important roles in Earth’s rising temperatures and changing climate, three carbon-based species take the focus of this work due to their impacts on climate change and excessive anthropogenic emissions: carbon dioxide, methane, and carbon monoxide (CO).

### **1.1.1 Carbon Dioxide**

In addition to being closely related to natural processes, such as plant growth and decay, CO<sub>2</sub> concentrations are highly influenced by anthropogenic processes, including both urban and industrial emissions. In the Northern Hemisphere, CO<sub>2</sub> is known to oscillate from low concentrations during the growing season, due to plant photosynthesis, and high concentrations during the remainder of the year due to plant decay and respiration. However, this annual cycle occurs with higher average concentrations as years progress. This global trend of increasing CO<sub>2</sub> has been repeatedly tied to increases in anthropogenic activity. In the majority of cases, transportation, power generation, and various industrial processes (e.g., cement and steel production (Ramanathan et al. 2019), as well as organic chemical and petrochemical processes (Saygin and Gielen 2021)) are primarily responsible for increases in CO<sub>2</sub>, only overshadowed by less-common natural occurrences (Arya 1999). No matter the source, greater

concentrations of CO<sub>2</sub> lead to a greater global warming potential (GWP). As the most abundant greenhouse gas, CO<sub>2</sub> reduction is necessary to limit the effects of climate change. Anthropogenic CO<sub>2</sub> is especially important in this effort, as it is specifically considered the largest contributor to the greenhouse gas-induced imbalance of the Earth's energy budget (Hu et al. 2018).

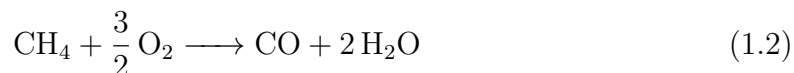
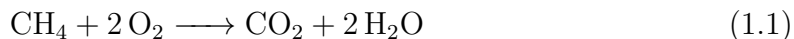
### **1.1.2 Methane**

While less abundant than CO<sub>2</sub>, greenhouse gas CH<sub>4</sub> is almost 30 times as potent at trapping radiation in the atmosphere over a 100-year period (IPCC 2021). Like CO<sub>2</sub>, CH<sub>4</sub> comes from a variety of natural and anthropogenic sources, but the non-natural sources hold special importance in the climate discussion due to "super-emitter" sources (a small percentage of sources that contribute a majority of emissions) and uncertainties in emission contributions (Duren et al. 2019; Zavala-Araiza et al. 2015). Since the early 1800s, atmospheric CH<sub>4</sub> has more than doubled accompanying the advent and global growth of many industrial processes (US EPA 2022b). Recent trends in the United States suggest CH<sub>4</sub> has been subtly declining since 1990 due to limited coal mining, reformed natural gas and petroleum systems, as well as the presence of persistent El Niño conditions (US EPA 2023b; IPCC 2021). However, these trends are not shared globally (IPCC 2021). Generally speaking, the potency and relatively short atmospheric life span (approximately 12 years) make CH<sub>4</sub> a significant concentration of interest, as any concentration reduction would have a more profound impact on climate change than a comparable reduction of CO<sub>2</sub> (US EPA 2022b, 2023b). Reducing emissions from the oil and gas industry in particular, host

to many methane super-emitters, could dramatically mitigate climate impacts (Duren et al. 2019; Zavala-Araiza et al. 2015).

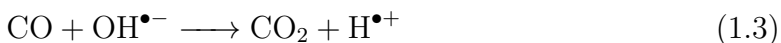
### 1.1.3 Carbon Monoxide

Carbon monoxide shares many sources with CO<sub>2</sub>, but has notably different impacts when emitted into the atmosphere. Like CO<sub>2</sub>, CO is a product of a combustion reaction, or reaction with oxygen. While complete combustion of carbon-based material generates CO<sub>2</sub> (Equation 1.1), combustion with insufficient oxygen, also known as incomplete combustion, generates CO (Equation 1.2). When a material is combusted, both complete and incomplete combustion reactions occur, co-emitting CO<sub>2</sub>, CO, and H<sub>2</sub>O.



Unlike CO<sub>2</sub>, CO is not a greenhouse gas because its ability to directly trap and re-emit longwave radiation is minimal. However, CO is unique in its ability to indirectly contribute to atmospheric warming via chemical reactions that alter the concentrations of CH<sub>4</sub>, CO<sub>2</sub>, and O<sub>3</sub> (Daniel and Solomon 1998). In the atmosphere, CO is the most common reactant with the hydroxyl radical (OH<sup>•</sup>), called the “detergent of the atmosphere” for its proclivity to react out trace gases in the atmosphere (Riedel and Lassey 2008). This reaction produces CO<sub>2</sub> (Equation 1.3), whose climate-related

importance was discussed previously. When not reacting with CO,  $\text{OH}^{\bullet-}$  prefers to react with  $\text{CH}_4$ , another important climate forcer (Riedel and Lassey 2008).



Aside from impacting the climate, CO can also have profound impacts on human health. Elevated concentrations of CO near the Earth’s surface can become a health concern, especially for people with certain kinds of heart disease, and fatal if exposed to high doses or in confined spaces (US EPA 2022a). The US Environmental Protection Agency (EPA) categorizes CO as a “criteria air pollutant,” and thus has specific standards for indoor and outdoor exposure to avoid health concerns (US EPA 2023a).

## 1.2 The Need for Continued Carbon-Based Pollutant Monitoring

As the scientific and political landscapes evolve their understanding of climate change and more regulations are set in place to curtail pollution, continued monitoring of emissions is critical. This is particularly true at regional and city scales, where source attribution uncertainty surges due to coarse resolution emission inventories and limited datasets. Source attribution of  $\text{CO}_2$  becomes increasingly difficult as spatial resolution increases from global scales, where uncertainty is often around 10-40%, to city-scale, where uncertainty becomes over 150% (Hu et al. 2018). Spatial data gaps in  $\text{CH}_4$  data have also proven problematic when trying to assess climate change at regional scales (IPCC 2021). Emission uncertainties for  $\text{CH}_4$  and CO are especially complicated due to their shared connections with  $\text{OH}^{\bullet-}$  (IPCC 2021; Daniel and Solomon 1998).

Continued data collection would ensure scientific understanding of these pollutants continues to evolve, and that emission inventories and climate models are held to high accuracy standards.

Continued ground-based monitoring is especially important to assure that global continuous monitoring via space-based sensors provides an accurate representation of the state of the atmosphere. Satellites are crucial for providing a holistic view of the Earth's atmosphere, as lower atmospheric measurements are comparatively sparse (Pan et al. 2021). However, space-based sensors often deal with uncertainty sources that ground-based sensors are less receptive to, like albedo changes (Zhou et al. 2022). Therefore, ground-based monitoring is imperative to validate satellite data. These comparisons are done frequently and have been shown to benefit understanding and establish proper data correction methods for satellite data like those from the NASA Orbiting Carbon Observatories, OCO-2 and OCO-3 (e.g., Jacobs et al. (2020); Wunch et al. (2017); Zhou et al. (2022)).

Additionally, continued emission monitoring is necessary to properly inform stakeholders of urban and industrial environments. Accurate monitoring is crucial for those living in areas with increased emissions, as they should remain well informed of any hazards they face, such as those associated with exposure to CO (US EPA 2022a; Voiland 2015). Additionally, accurate emission data are critical for creation and proper enforcement of climate policy, especially compliance laws (Shiga et al. 2014; Pan et al. 2021).



# 1.3 Pollutant Monitoring During the TRACER Campaign

The GeoCarb-TRACER Campaign, generally referred to as the TRACER Campaign throughout this report, was designed to contribute to the continued pollutant monitoring effort, focusing on CO<sub>2</sub>, CH<sub>4</sub>, CO, and water vapor measurements in and around the coastal city of Houston, TX from late May through mid-September of 2022. As a part of the greater US Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) user facility

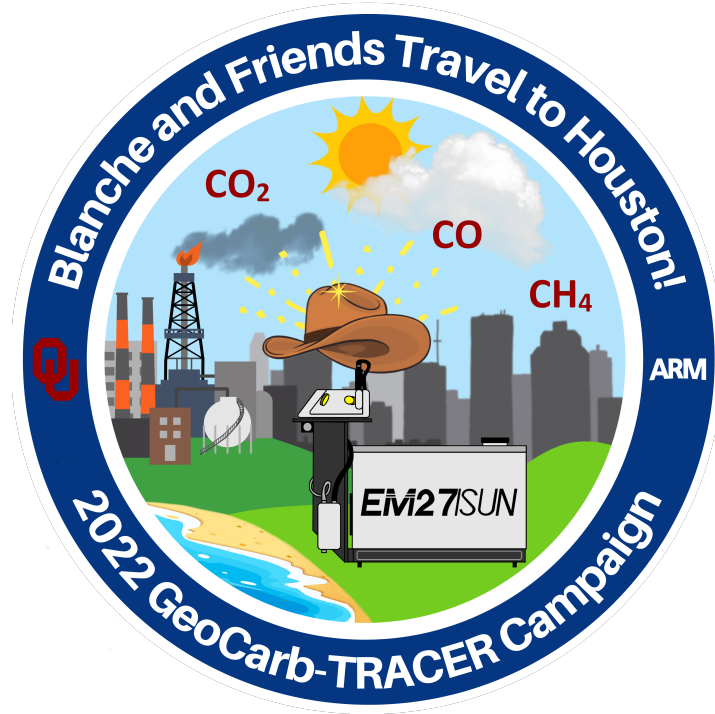


Figure 1.1: The unofficial GeoCarb-TRACER Campaign logo, designed and drawn by myself

Tracking Aerosol Convection interactions Experiment (TRACER), GeoCarb-TRACER was focused on observing urban and industrial carbon-based pollutants, as well as satellite validation.

Houston was an ideal location for the GeoCarb-TRACER Campaign, as the city is the fourth most populated in the United States and home to many expansive industrial processes (Wallace et al. 2018). The city is as a global leader in both petrochemical and

plastics manufacturing (Medlock III 2021; Phillips 2019), where  $\text{CO}_2$ ,  $\text{CO}$ , and  $\text{CH}_4$ , among other pollutants, such as aerosols, are released into the atmosphere as fugitive emissions (IPCC 2019). Many of these manufacturing facilities are concentrated along an area known as the Houston Ship Channel, located to the east and southeast of the city center (Wallace et al. 2018). Limited zoning laws in the area allow for residential and industrial areas to exist in alarmingly close proximity of one another (Miller et al. 2020), emphasizing the need for accurate pollutant monitoring.



Figure 1.2: Blanche the EM27/SUN spectrometer, actively observing trace gases during a teaching demonstration

During the TRACER Campaign,  $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{CO}$ , and water vapor were observed using Bruker (Billerica, Massachusetts, USA) EM27/SUN spectrometers, one of which can be seen in Figure 1.2. These portable Fourier transform spectrometers capture high resolution spectra in the near- and shortwave-infrared wavelength range (Gisi et al. 2012). These spectra are commonly analyzed to retrieve column-averaged abundances of species present

in the atmosphere, primarily  $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{CO}$ , and water vapor, at regional and sub-regional scales. During the campaign, I coordinated and participated in the deployment of up to three EM27/SUN instruments simultaneously

alongside ARM TRACER instruments gathering boundary layer, aerosol, and near-surface meteorological information. Deployments occurred at sites expected to be influenced by local urban and industrial pollutant sources (e.g., traffic, petrochemical facilities, etc.) as well as background sites, which experienced fewer urban/industrial influences.

The EM27/SUN instruments were also used to collect trace gas data at the ARM Southern Great Plains (SGP) site in Lamont, OK before and after the TRACER Campaign. These measurements validate the EM27/SUN spectrometers against a Bruker IFS 125HR spectrometer belonging to TCCON, the Total Carbon Column Observing Network. The TCCON instruments record higher resolution spectra ( $0.02\text{ cm}^{-1}$ ) than the EM27/SUN instruments and are calibrated to the World Meteorological Organization (WMO) trace gas standard using repeated aircraft profile measurements at each site (Wunch et al. 2011). For these reasons, the TCCON instruments are often considered the standard of comparison for any total column trace gas measurements.

This thesis will discuss the methodology of EM27/SUN deployments, summarize the data collected, and discuss important findings from data analysis. The approach to analysis is three-pronged, focusing first on ground-based instrument intercomparisons including variations introduced by retrieval algorithm choice and modification. Then, seasonal to sub-daily trends within the TRACER dataset are explored, generating hypotheses for the presence of significantly high and low observed concentrations. Unsupervised machine learning enhances the process, paired with statistical analysis. Finally, ground-up EM27/SUN data are used to contextualize OCO-2 and OCO-3 top-down observations during the TRACER Campaign.

# Chapter 2

## Methods

This chapter summarizes the technical and logistical details pertaining to EM27/SUN deployments and data processing. First, Section 2.1 discusses EM27/SUN spectrometers, deployment methods, and related measurements useful for analysis. Section 2.2 covers the TCCON calibration campaign and the TRACER Campaign, including the motivation for both. Finally, EM27/SUN data processing and machine learning analysis techniques are discussed in Sections 2.3 and 2.4, respectively.

### 2.1 Instrumentation for Campaigns

Both field campaigns discussed in this work involved deploying Bruker EM27/SUN spectrometers. These instruments and their deployment methods are discussed in detail in Section 2.1.1 before a brief discussion of secondary instrumentation recommended for co-deployment in Section 2.1.2.

#### 2.1.1 EM27/SUN Spectrometers

Both the TRACER and TCCON calibration campaigns involved deploying one EM27/SUN at minimum to collect trace gas concentration data. These instruments directly measure infrared solar radiation in a wavelength range from 4000 to 11000  $cm^{-1}$  at a spectral resolution of 0.5  $cm^{-1}$  (Hase et al. 2016; Gisi et al. 2012) and varied

temporal resolution of approximately 15 seconds. The solar radiation observed by the spectrometer contains information about the concentrations of atmospheric constituents present between the instrument and the top of the atmosphere, as every gas and particle in the atmosphere absorbs solar radiation at different wavelengths depending on their physical and chemical makeup.

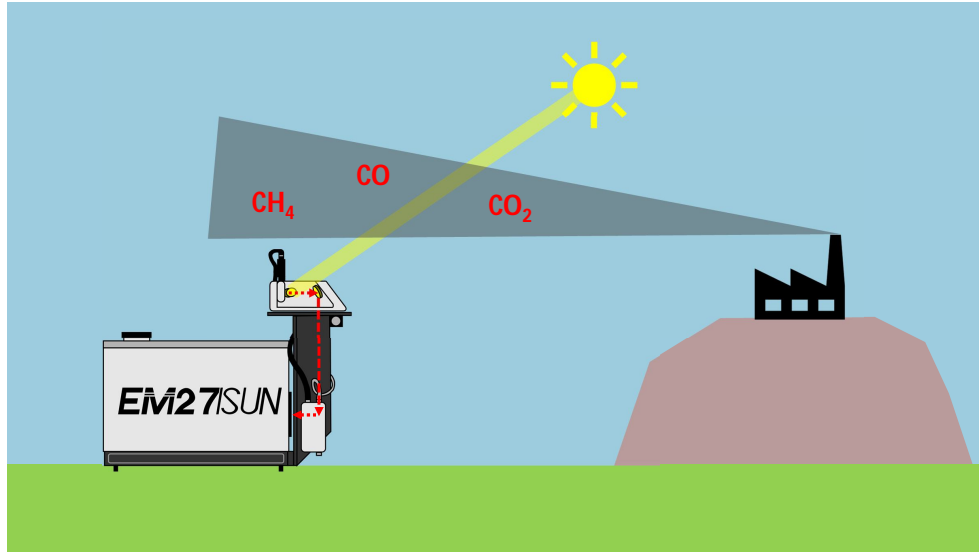


Figure 2.1: A cartoon of an EM27/SUN deployment. The EM27/SUN has been deployed during clear-sky conditions so that direct sunlight (yellow beam) can be observed by the EM27/SUN. The primary and secondary mirrors of the external optics (yellow discs on top of the EM27/SUN) reflect the direct sunlight to the tertiary external mirror (not pictured) and then into the instrument cavity following the red dotted arrows. The internal optics and sensors of the EM27/SUN record information about the direct radiation they observe. Any trace gases present will alter the radiation that reaches the instrument. Emission sources, generalized by the industrial plant on the right, may produce large amounts of these trace gases, represented by the simplified gray plume from the smoke stack.

In the field, solar radiation data are collected by aligning the external optics on the instrument so that direct sunlight can reach the internal optics and sensors. The external optical system is comprised of three 3" gold mirrors. The first two mirrors,

represented by the yellow discs in the cartoon in Figure 2.1, rotate to match the elevation and solar zenith angle of the sun. The CamTracker software (Gisi et al. 2011) operates the motors that rotate these mirrors and allow for continuous automatic solar tracking. Sunlight reflected by the rotating mirrors is reflected in turn by the third external gold mirror through a specialty longpass filter glass window on the side of the EM27/SUN onto the internal optics and dual indium gallium arsenide (InGaAs) detectors (Hase et al. 2016; Gisi et al. 2012).

Through retrieval algorithm processing, the solar radiation data collected by the EM27/SUN can be used to infer total column dry-air mole fractions, or total column concentrations, of different trace gases present in the atmosphere. These total column concentrations represent the concentration, in ppb or ppm, of a trace gas in the column of atmosphere between the instrument and the top of the atmosphere. In the retrieval algorithm, these concentrations are calculated using Equation 2.1,

$$X_{gas} = \frac{VC_{gas}}{VC_{O_2}} * 0.2095 \quad (2.1)$$

where  $X_{gas}$  represents the total column concentration,  $VC_{gas}$  is the retrieved total (vertical) column abundance of a measured gas,  $VC_{O_2}$  is the retrieved total column abundance of oxygen, and the multiplicative factor represents the assumed mole fraction of oxygen in dry air (Gisi et al. 2012). The total column concentrations frequently retrieved from raw EM27/SUN data using this method include  $X_{CO_2}$ ,  $X_{CO}$ ,  $X_{CH_4}$ , and  $X_{H_2O}$ , which can also be found in literature represented as  $X_{CO_2}$ ,  $X_{CO}$ ,  $X_{CH_4}$ , and  $X_{H_2O}$ , depending on the publication. More information about data processing with retrieval algorithms is provided in Section 2.3.

### 2.1.1.1 EM27/SUN Deployment Methods

These spectrometers were designed to be small and lightweight, allowing for easy tabletop deployments. However, enclosure-based deployments are recommended to protect the EM27/SUN from the environment and minimize instrument maintenance.



Figure 2.2: EM27/SUN enclosure near instrument trailers during the TRACER Campaign

The enclosure used while deploying the GeoCarb EM27/SUN, shown in Figure 2.2, was able to cool the EM27/SUN and protect instrumentation from the elements with some modifications. In addition to the EM27/SUN, the enclosure could host the field laptop, along with any other sensors or items necessary for deployment. The original enclosure design, shown in the computer aided design (CAD) drawing in Figure 2.3, was created and kindly shared by Dr. Bianca Baier, Tim Newberger, and Philip Handley of the National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Laboratory.

The latest version of the enclosure used in the field was equipped with two ThermoTEC Series 500 BTU thermoelectric air conditioners (EIC Solutions, Inc.,

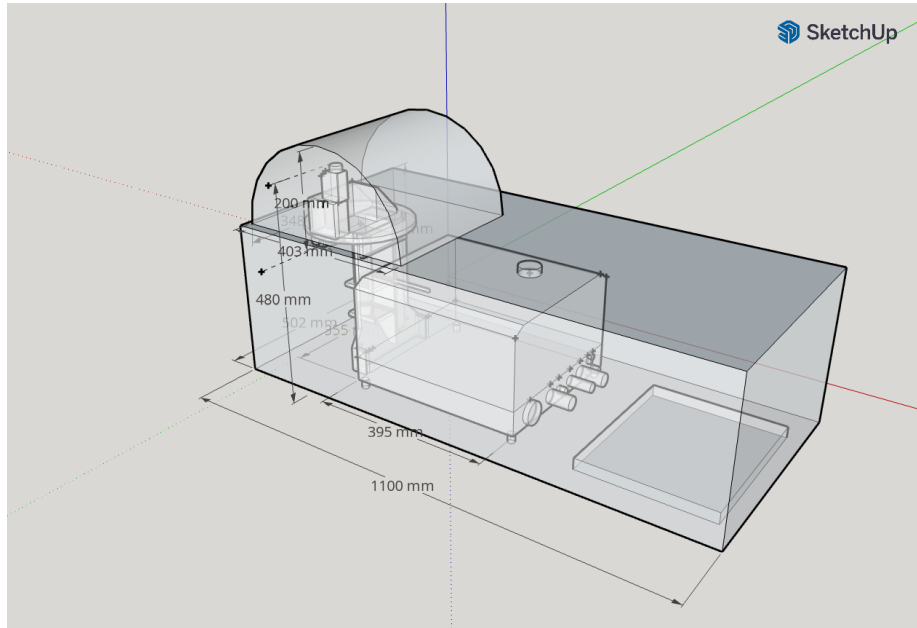


Figure 2.3: CAD drawing of the original EM27/SUN enclosure design created by Dr. Bianca Baier, Tim Newberger, and Philip Handley of the NOAA Global Monitoring Laboratory.

Warminster, Pennsylvania, USA) for sufficient cooling in extreme heat environments after a single unit proved to be insufficient during Houston deployments. These air conditioners utilize Peltier coolers and fans, which limit vibration compared to traditional compressor-based units. Minimal vibration is necessary to avoid impacting the EM27/SUN optics. These air conditioners also operate with low power requirements, ensuring that their operation will not interrupt the instrumentation suite during data collection. Suggested improvements for future enclosure designs are included in the lessons learned discussion in Appendix A. Additionally, all procedures and troubleshooting advice for both tabletop- and enclosure-based deployment options are detailed in Appendix B.



### **2.1.2 Secondary Instrumentation**

Apart from the tabletop- or enclosure-based EM27/SUN setup, deployments could include use of a pressure sensor, weather station, pyranometer, and rain sensor for more information about the meteorological conditions present while deploying. A secondary instrumentation suite was in the process of being developed during the course of this work. More details about this instrumentation suite and related deployment procedures are detailed in Appendix B.

While the instrumentation suite was being developed and made ready for field deployments, the necessary meteorological measurements were made using pre-existing instrumentation at the deployment sites. These measurements included near-surface temperature, pressure, relative humidity, wind speed and wind direction. Information about the instrumentation used at each deployment site is provided below in Section 2.2.

## **2.2 Experimental Design of Field Campaigns**

Blanche the EM27/SUN has been involved in two field campaigns to date: the TCCON calibration campaign and the GeoCarb-TRACER Campaign. Section 2.2.1 includes the logistics and experimental details for the TCCON calibration campaign, which centered around EM27/SUN validation. The TRACER Campaign, which used the validated EM27/SUN instruments to gather data in the industrial city of Houston, TX, is discussed in detail in Section 2.2.2.

## 2.2.1 TCCON Calibration Campaign

The TCCON calibration campaign was the first field campaign conducted involving the GeoCarb EM27/SUN, “Blanche.” The campaign involved a series of single- and multiple-day deployments at the ARM SGP atmospheric observatory near Lamont, OK. The purpose of these deployments was two-fold. The primary objective was to assess the precision and accuracy of Blanche the EM27/SUN via comparison with the on-site TCCON station, host to a high resolution Bruker IFS 125HR spectrometer corrected to WMO standards and frequently used as a comparison standard for lower-resolution spectrometers. However, the extended timeline of the campaign generated a second objective: to test essential components of the remote deployment setup, discussed in detail in Appendix B.



Figure 2.4: Deployment of Blanche at the Guest Instrument Facility during the TCCON calibration campaign. The EM27/SUN can be seen on top of the platform in the picture on the left. The picture on the right demonstrates a typical “tabletop” EM27/SUN deployment on the raised platform.

The calibration campaign consisted of 13 single-day deployments from October 29, 2021 through November 21, 2022. With the exception of the final three, all deployments involved collecting data with Blanche without the use of an enclosure. During these deployments, Blanche collected data from a raised platform attached to the SGP

Guest Instrument Facility (GIF), pictured in Figure 2.4. This facility was located approximately one-tenth of a mile away from the Lamont TCCON station and the near-surface Automatic Weather Station, MAWS. Data utilized from MAWS included one-minute near-surface temperature, pressure, relative humidity, wind speed, and wind direction measurements (Keeler and Kyrouac n.d.).

The final three deployments during the calibration campaign involved deploying Blanche along with the two guest EM27/SUN instruments featured in the GeoCarb-TRACER Campaign, kindly loaned by Dr. Gregory Osterman from NASA Jet Propulsion Laboratory (JPL) and Dr. Frank Hase at the Karlsruhe Institute of Technology (KIT). The deployments took place beside the SGP TCCON station, pictured in Figure 2.5. This location was chosen to easily co-deploy all EM27/SUN instruments with TCCON, as the raised GIF platform did not support multiple simultaneous EM27/SUN deployments. Additionally, the TCCON station conveniently provided power to the EM27/SUN instruments and Blanche's enclosure.



Figure 2.5: Deployment of Blanche and guest EM27/SUN instruments next to the ARM SGP TCCON station (white building) during the TCCON calibration campaign during October 17-19, 2022.

These deployments served as an important calibration experiment in addition to the ongoing single-EM27/SUN comparison with TCCON. The purpose of this experiment was to evaluate the agreement between the co-deployed EM27/SUN instruments. Evaluating the bias between the EM27/SUN instruments was crucial for the TRACER Campaign, where each EM27/SUN was expected to collect accurate data while deployed at different sites.

### **2.2.2 GeoCarb-TRACER Campaign**

The GeoCarb-TRACER Campaign was conducted during the summer and early fall of 2022, from May 26 through September 17, in Houston, Texas. GeoCarb-TRACER was a child campaign of ARM TRACER, whose intensive operational period (IOP) lasted from June 1 through September 30, 2022 in the greater Houston/Galveston area of Texas.

While the GeoCarb-TRACER Campaign and ARM TRACER occurred almost simultaneously, each campaign had very unique goals. ARM TRACER focused primarily on analyzing aerosols, cloud life cycles, and the impacts of aerosols on cloud evolution. Houston was chosen for this analysis as the proximity to the Gulf of Mexico generates complex and frequently convective meteorological conditions in the presence of urban- and industrially-driven aerosol pollution (Jensen 2019). The GeoCarb-TRACER Campaign, however, focused on carbon-based pollution monitoring with EM27/SUN spectrometers. The primary goals of the GeoCarb-TRACER Campaign were as follows:

- To monitor the carbon-based pollutants present in the highly urbanized and industrialized Gulf environment and attempt to attribute anomalous concentrations,
- To contextualize emissions in the Houston atmosphere by comparing EM27/SUN data with aerosol, boundary layer, and other datasets collected by other research groups during ARM TRACER,
- To validate space-based remote sensors including OCO-2 and OCO-3.

The first two goals of the GeoCarb-TRACER Campaign were based on the pivotal hypothesis that EM27/SUN spectrometers were capable of detecting *Xgas* enhancements from local urban and industrial pollutant sources. Houston provided the ideal testing ground for this hypothesis. As mentioned in Section 1.3, petrochemical and other industrial processes are very common in the Houston area. Consequently, various urban/industrial pollutant signals could be observed in the area, included co-emitted CO<sub>2</sub>, CO, and H<sub>2</sub>O from combustion processes (e.g., vehicle emissions, building heating, etc.) as well as fugitive emissions of CO<sub>2</sub>, CO, and CH<sub>4</sub> from petrochemical and organic chemical processes (IPCC 2019). Co-emission of all *Xgas* concentrations, including water vapor, could occur at petrochemical and related industrial facilities via flares, common safety systems that operate constantly to combust gaseous emissions instead of releasing them directly to the atmosphere. The amount of flaring that occurs a facility is a function of how well a process is running. While flares are operated constantly, emergency flaring can cause more than one million pounds of gas per hour to pass through the flare (Peterson et al. 2007). Fugitive emissions from flares can contain a variety of chemicals that go uncombusted, as these systems are notoriously difficult to operate at peak efficiency, especially during non-routine process operation

(Plant et al. 2022; Peterson et al. 2007). Poorly controlled flares alone are estimated to emit enough  $\text{CH}_4$  to quintuple the expected emissions from oil and gas production (Plant et al. 2022).

During the GeoCarb-TRACER Campaign, EM27/SUN data were collected during a total of 58 daily observation periods across 45 days, made possible with the inclusion of three loaned EM27/SUN instruments. One of these instruments was loaned by the Karlsruhe Institute of Technology (EM27/SUN serial number 83) while the other two were loaned from the NASA JPL (serial numbers 42 and 175). Unfortunately, technical issues in the field with JPL 42 excluded this instrument from the campaign. The remaining guest EM27/SUN instruments were used multiple times during the campaign, and will be referred to simply as KIT and JPL.

Multiple EM27/SUN deployment sites were necessary to accomplish the goals set by the GeoCarb-TRACER, or simply, TRACER Campaign. While ARM TRACER featured three primary sites for instrument co-deployment, a series of sites around Houston were considered. Each deployment site was chosen according to the following criteria:

1. Ease of access. Sites already planned for use during ARM TRACER were prioritized to avoid siting difficulties while ensuring instrument collocation.
2. Status as either an urban or background site in anticipation of measuring concentration gradients.
3. Location with respect to seasonal wind patterns. Monthly historic mean wind directions were evaluated to ensure that regions upwind of sites were either urban/industrial or suburban. In each month of the campaign, at least one deployment site needed to fall into each category.

Using the above site location criteria, the ARM Mobile Facility 1 (AMF1) site (later referenced as the ARM Main site as well as site 5), was adopted by GeoCarb-TRACER along with the University of Houston’s Coastal Center (UHCC, site 6), main campus (UH Main, site 7), and Aldine, Texas (Aldine, site 8) sites, all of which can be viewed in Figure 2.6. Each site hosted a variety of instrumentation used in ARM TRACER, including various near-surface meteorological stations, whose data were utilized for EM27/SUN retrievals and analysis. All meteorological data sources utilized during the campaign are summarized in Table 2.1.

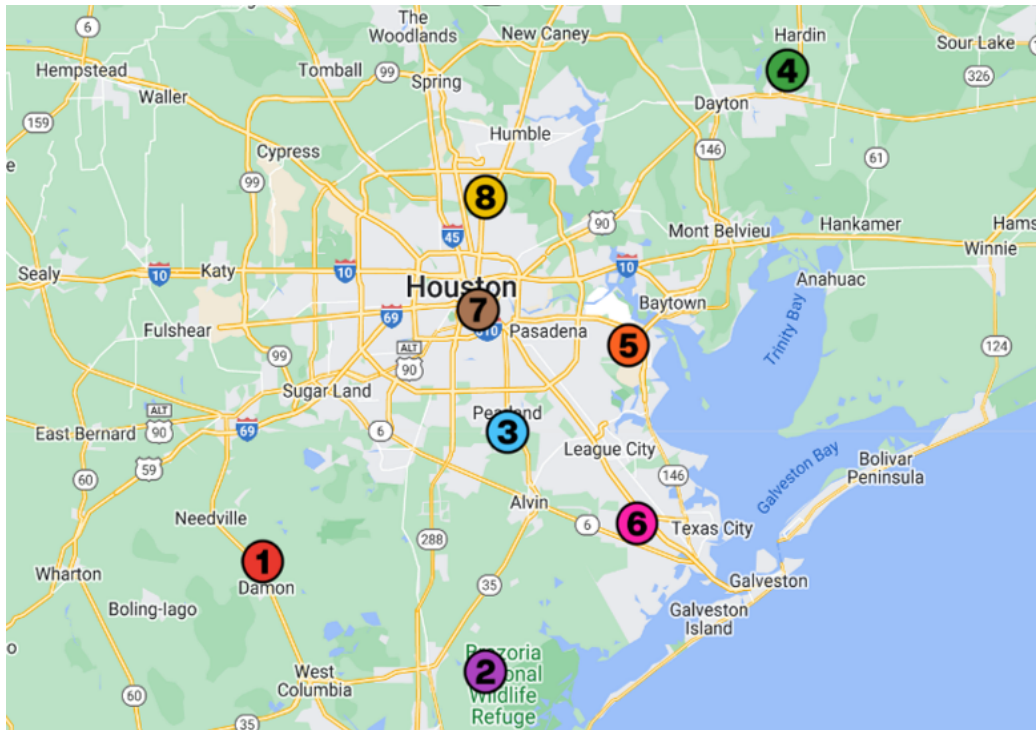


Figure 2.6: TRACER Campaign site locations in the Houston, TX area

Table 2.1: Sources and temporal resolution of near-surface meteorological data at EM27/SUN deployment sites during the TRACER Campaign

Deployment Site	Data Source	Temporal Resolution
ARM Main (5)	ARM-hosted MAWS (Keeler and Kyrouac n.d.)	One minute
UHCC (6)	NOAA National Severe Storms Laboratory (NSSL) Collaborative Lower Atmospheric Mobile Profiling Systems, CLAMPS-2 (Wagner et al. 2019)	One second
UH Main (7)	UH-hosted sensors operated by James H. Flynn <sup>1</sup>	Ten second
UH Aldine (8)	CLAMPS-1 (Wagner et al. 2019)	One second

Sites 1-4 represent the ARM ancillary site, CopterSonde (Segales et al. 2020) launch site near Brazoria National Wildlife Refuge, ARM CSAPR site (host to CSAPR2, the second-generation C-band Scanning ARM Precipitation Radar), and UH Liberty, TX site, respectively. These sites were considered for EM27/SUN deployments but were not used for a variety of reasons including their locations with respect to local trace gas sources, relative difficulty to access, and, for some, lack of instruments to co-deploy alongside.

While daily wind patterns near the Gulf Coast are quite complex due to the bay and sea breeze circulations (Caicedo et al. 2019), the monthly climatological 10 m winds are relatively consistent through the summer months. Surface winds in the Houston area from May through August largely agree with the wind rose on the

---

<sup>1</sup>Pressure data were collected with a Vaisala (Vanta, Finland) PTB110, temperature and relative humidity were collected with a Vaisala HMP-45c, and both wind speed and direction were measured with an R. M. Young Company (Traverse City, Michigan, USA) YOUNG Model 05305 AQ Wind Monitor.



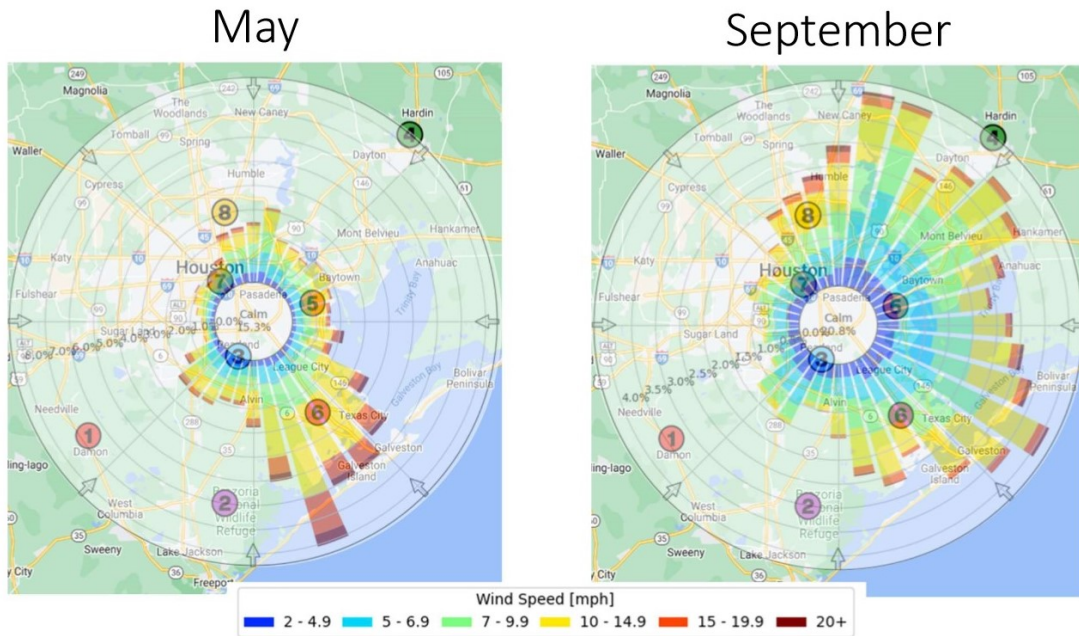


Figure 2.7: Central Houston wind roses generated using wind speed data at approximately 10 m elevation collected at the Houston Dunn ASOS station (identifier MCJ) from 2009 through 2021 (Iowa Environmental Mesonet 2022)

left of Figure 2.7, generated from an Automated Surface Observing System (ASOS) station near the heart of Houston (Iowa Environmental Mesonet 2022). The wind resource changes significantly in September (right side of Figure 2.7), which suggested during campaign planning that the sites chosen for GeoCarb-TRACER needed to be multi-purpose. For the majority of the campaign, the UHCC and ARM main sites provided a background signal, while UH Main and Aldine most often served as urban sites. However, due to the large amount of urbanization in the area, the presence of many industrial plants, as well as the populated island of Galveston, all deployment sites were thought to likely see some degree of an urban or industrial signal.

The majority of the TRACER Campaign featured tabletop deployments of EM27/SUN instruments while the air-conditioned enclosure introduced in Section 2.1.1.1 was being

manufactured. These tabletop deployments proved difficult for both the EM27/SUN instruments and their operators due to the harsh summer heat and humidity. Consequently, once the enclosure was ready for deployment, it was promptly transported to Houston for use in the final 21 deployments to house Blanche. During the enclosure deployments, guest instruments were kept in storage while waiting on the manufacturing of other enclosures to prevent damage from harsh conditions.

### **2.2.2.1 OCO-2/3 Collocation**

During the TRACER Campaign, EM27/SUN instruments were deployed near the observation paths of the OCO-2 and OCO-3 satellites to validate  $XCO_2$  measured by the space-based remote sensors. Both OCO-2 and OCO-3 can collect  $XCO_2$  data in nadir mode (measurements taken over land facing the earth's surface and oriented orthogonally) as well as glint mode (measurements taken over water pointing towards glint spots on the water). The OCO-2 satellite also has a target mode for TCCON validations. These satellites collect thin swaths of  $XCO_2$  data in these observation modes globally (Eldering et al. 2017, 2019). However, OCO-3 is unique in that it can collect large swaths of data (around 80 km<sup>2</sup>) at specific targets in snapshot area mode (SAM) (Eldering et al. 2019).

The latest bias-corrected datasets for OCO-2 and OCO-3, Version 11r and Version 10.4r, respectively, were used for comparison. These data can be found online from OCO-2/OCO-3 Science Team (2020, 2022). All satellite data were quality filtered using built-in flags before comparison such that only good quality data were analyzed and lower quality data were omitted, designated by the *xco2\_quality\_flag* provided with the data.

The EM27/SUN instruments were considered collocated with OCO-2 or OCO-3 if the satellite data were collected within 50 km of the EM27/SUN deployment site. Additionally, EM27/SUN data were considered collocated with OCO-2/3 if collected within two hours before and after the satellite overpass. These collocation criteria were established by Zhou et al. (2022) for monitoring OCO-2 and OCO-3 bias in urban areas. To simplify the spatial criterion for practical purposes, OCO data were considered collocated within a 100 km box around the deployment site as opposed to a 50 km radius circle centered on the site. The coordinates of these 100 km spatial collocation boxes were found using the great circle method, which can be used to provide the radial distance from a location (Weintrit and Kopacz 2011; Williams n.d.).

## 2.3 EM27/SUN Retrieval Algorithms

Prior to analysis, raw EM27/SUN data, also known as interferograms, must be pre-processed using a retrieval algorithm to extract  $X_{gas}$  from the absorption spectra. Interferograms are primarily processed using one of two open-source retrieval algorithms: GGG (Alberti et al. 2022) and PROFFAST (Wunch et al. 2011). The GGG algorithm is the official retrieval algorithm of TCCON, but has been adapted for processing EM27/SUN data as well. The PROFFAST algorithm was designed for use with the EM27/SUN network called COCCON, the Collaborative Carbon Column Observing Network. The latest PROFFAST algorithm and its documentation can be found on GitLab (Feld et al. 2023). The GGG retrieval algorithm code can be downloaded by following the instructions at <https://tcccon-wiki.caltech.edu/Main/WebHome> after creating a user account. Procedures for running each of the retrieval algorithms are included in Appendix C.

Each of the retrieval algorithms produces  $X_{gas}$  concentrations for  $\text{CO}_2$ ,  $\text{CO}$ ,  $\text{CH}_4$ , and  $\text{H}_2\text{O}$ , denoted as  $X_{\text{CO}_2}$ ,  $X_{\text{CO}}$ ,  $X_{\text{CH}_4}$ , and  $X_{\text{H}_2\text{O}}$ , respectively. Additionally, both algorithms calculate a total column concentration of air, which is commonly used for quality assurance of the retrieved trace gas concentrations. However, each algorithm defines this quantity differently. In GGG, this parameter is called  $X_{luft}$ , defined in Equation 2.2, where  $VC$  indicates a column density (molecules per  $\text{cm}^2$ ). The column density of dry air,  $VC_{dryair}$ , is calculated from surface pressure measured during EM27/SUN deployment. As in Equation 2.1,  $VC_{\text{O}_2}$  represents the total vertical column abundance of oxygen. The multiplicative factor represents the mole fraction of oxygen in dry air (Mostafavi Pak et al. 2023). The PROFFAST alternative to this parameter,  $X_{air}$ , is simply the reciprocal of  $X_{luft}$  (Frey et al. 2019). This parameter should ideally always equal one, making it a useful indicator for EM27/SUN performance quality.

$$X_{luft} = \frac{VC_{dryair}}{VC_{\text{O}_2}} * 0.2095 \quad (2.2)$$

In order to retrieve the  $X_{gas}$  data, each algorithm follows the same basic process flow for EM27/SUN data (Wunch et al. 2011; Mostafavi Pak et al. 2023; Alberti et al. 2022):

1. Conversion of interferograms into spectra, where the data are a function of wavenumber, via Fourier transform. Figure 2.8 shows a single spectra recorded with the EM27/SUN in OPUS, the data recording software for the EM27/SUN. This spectral view in OPUS is purely for visualization purposes. OPUS saves the data as interferograms and the retrieval algorithm performs a Fourier transform during processing.

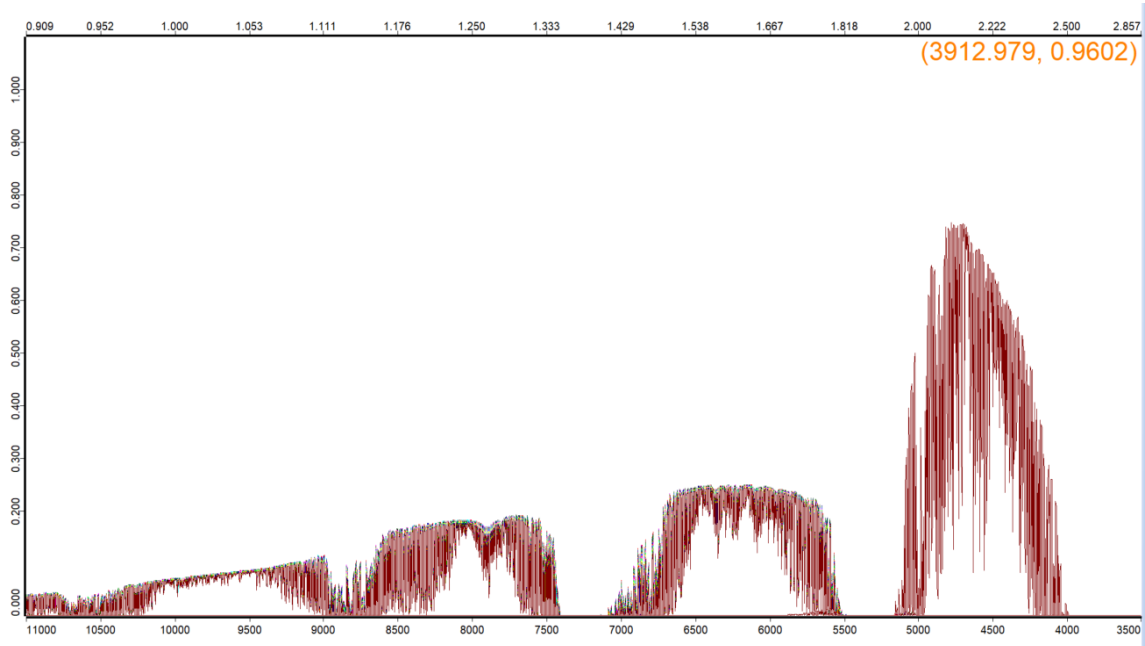


Figure 2.8: One EM27/SUN interferogram captured in OPUS, where the bottom x-axis represents wavenumber ( $cm^{-1}$ ), the upper x-axis represents wavelength ( $\mu m$ ), and the y-axis is the intensity of the solar radiation

2. Nonlinear least squares fitting of the EM27/SUN spectra with modeled spectra generated with *a priori* profiles and iteratively scaled to improve fit.
3. Integration of the modeled spectra to calculate column abundances of the chemical species of interest.
4. Calculation of total column concentrations of  $CO_2$ ,  $CO$ ,  $CH_4$ , and  $H_2O$ .

Both retrieval algorithms utilize the same *a priori* data. Each *a priori* profile is generated by the Goddard Earth Observing System Forward Processing for Instrument Teams (GEOS-FP-IT) atmospheric data assimilation system. This system produces *a priori* data files, or priors, every three hours for each day (Lucchesi 2015). Information including temperature, pressure, and gas concentrations are generated on 51 vertical

levels in the atmosphere. These levels are structured to provide a higher density of data closer to the surface, with level spacing increasing with altitude from 0.4 km spacing at sea level to 2.4 km spacing at an altitude of 70 km (Laughner et al. 2022). Additionally, both algorithms require a detailed record of atmospheric pressure at the same altitude of the EM27/SUN to accurately contextualize the depth of the column of air for the retrieval (Wunch et al. 2011).

The retrieval algorithms differ in their application of data quality filters, correction factors, and, arguably most notably, instrument-specific parameters. Minor optical irregularities, detector sensitivity, and other differences or changes in the optical alignment of an EM27/SUN can modify the raw data collected by the instrument and lead to biased retrieval products (Alberti et al. 2022; Mostafavi Pak et al. 2023). These changes are quantified by the instrument line shape (ILS). PROFFAST encourages the use of instrument-specific ILS parameters, which are periodically measured for each EM27/SUN in COCCON (Alberti et al. 2022). In contrast, the current version of GGG, GGG2020, does not implement instrument-specific ILS, instead using the ideal ILS of a perfectly aligned spectrometer (Mostafavi Pak et al. 2023).

Data collected during the TCCON calibration campaign have been processed with the latest versions of PROFFAST and GGG upon the time of analysis, PROFFAST version 2.2 and GGG2020, respectively. Section 3.2 compares retrieved data from these algorithms, along with the retrieval products from a modified version of GGG2020 utilizing EM27/SUN-specific *post hoc* correction factors developed by Mostafavi Pak et al. (2023). The results of this analysis dictated the choice of retrieval algorithm for the TRACER dataset.

## 2.4 Data Analysis Techniques

Due to the nature of the TRACER Campaign and its goals, robust data analysis techniques were essential. The campaign dataset was large and the data were expected to be influenced by the complex local meteorological conditions and pollution sources. A manual approach to analyze how the EM27/SUN data were influenced by these factors on different time scales would have been an arduous, if not impossible task. Therefore, unsupervised machine learning techniques were utilized to provide a more holistic data analysis approach.

Unsupervised machine learning techniques exist to uncover different ways that a dataset may be organized, revealing relationships in the data. One unsupervised technique, cluster analysis, is especially useful for visualizing data in ways that allow the relationships between different data to stand out. Cluster analysis is known for being generally robust to noise and outliers present in the data. However, these methods are also inherently subjective. Different approaches to cluster analysis will group the data in different ways, but it is ultimately up to the user to decide whether a clustered representation of the data is acceptable or not (James et al. 2021).

### 2.4.1 Choice of Machine Learning Algorithms

DBSCAN, or Density-Based Clustering of Applications with Noise, was chosen as the clustering algorithm for the TRACER dataset. This method is especially useful for recognizing clusters in data that are not organized in circular or spherical point clouds, which other methods can be heavily drawn towards (Ester et al. 1996). The TRACER dataset is not easily grouped into circular or spherical clusters due to the time-dependence of the data, among other qualities.

The DBSCAN algorithm works by identifying ideal neighborhood clusters, called epsilon neighborhoods, which are determined by the presence of core and boarder points in a higher density space than the space surrounding said points. The algorithm optimally shapes these clusters using two hyperparameters, the radius of the epsilon neighborhood and the density threshold for the neighborhood (Ester et al. 1996).

Uniform Manifold Approximation and Projection (UMAP), a dimensional reduction technique, was implemented to help DBSCAN identify clusters with the high-dimensional TRACER dataset. UMAP is based in topology and represents complex datasets as lower-dimensional topological manifolds that preserve as much information about the original dataset as possible. This technique is stochastic, yet relatively stable, meaning that each time the dimension reduction is performed, a slightly different reduced dataset will be created. While the variations between solutions are small, the algorithm can be forced to provide a specific solution for repeatability by specifying a seed number in the code.

In this study, UMAP is used to gain information about each of the carbon-based *Xgas*, which are called the features of interest of the algorithm. To learn about a feature of interest in UMAP, all data related to or expected to influence the feature of interest are provided to the algorithm. The combination of these data are known as the state vector space (McInnes et al. 2018). The feature of interest is withheld from the state vector space to avoid information leakage. In other words, if UMAP were to receive direct information about the variable it is expected to estimate, the algorithm will focus on it, clouding the algorithm's ability to highlight relationships within the state vector space that impact the feature of interest (Kaufman et al. 2012). For example, if trying to use UMAP to identify patterns in  $XCO_2$  with the TRACER



dataset, only the variables in Table 2.2 would be included in the state vector space and  $XCO_2$  itself would be omitted.

Table 2.2: State vector space for UMAP reduction of the TRACER dataset when the feature of interest is  $XCO_2$

General Information	EM27/SUN Data	Meteorological Data
<ul style="list-style-type: none"> <li>• Time of averaged measurements</li> <li>• Site code, stored as the single digit number from the code (e.g., 5 for site t5)</li> </ul>	<ul style="list-style-type: none"> <li>• One-minute averaged <math>XCO</math>, <math>XCH_4</math>, and <math>XH_2O</math></li> <li>• Two-minute rate of change of <math>XCO_2</math>, <math>XCO</math>, <math>XCH_4</math>, and <math>XH_2O</math></li> <li>• One-minute averaged SZA</li> </ul>	<ul style="list-style-type: none"> <li>• One-minute averaged temperature, pressure, relative humidity, wind speed, and wind direction</li> <li>• Two-minute rate of change of temperature, pressure, relative humidity, and wind speed <sup>2</sup></li> </ul>

## 2.4.2 Data Standardization for Clustering Applications

Unsupervised machine learning methods can be influenced by the scale of the data they analyze. To standardize the data, each data type (e.g.,  $XCO_2$ , ambient temperature, relative humidity, etc.) is mean-centered and scaled by standard deviation as seen in Equation 2.3 after being one-minute averaged to standardize the temporal resolution.

$$X_{Standardized} = \frac{X - \mu}{\sigma_X} \quad (2.3)$$

<sup>2</sup>Two-minute rate of change for wind direction was not included in the UMAP state vector space due to concerns about measurement accuracy at low wind speeds when using data from cup anemometers and similar instruments. Ideally, this measurement would also be included in the analysis.

Typically, standardization occurs over the entire dataset, such that a single mean and standard deviation are used per data type for standardization. For this analysis, standardization was performed for each unique combination of date and deployment site, meaning that data collected during each EM27/SUN deployment were standardized independently of the rest of the dataset. This approach was chosen to limit the influence of seasonal trends on analysis, as  $XCO_2$  and other variables were expected to exhibit seasonal as well as sub-seasonal trends.

While z-score standardization is straightforward for most data types, wind speed and wind direction require special attention. Wind speed and direction must be transformed into  $u$  and  $v$  wind components before accurately calculating the means and standard deviations for standardization. Equations 2.4 and 2.5 are used to calculate the mean wind speed and direction, respectively, from the  $u$  and  $v$  wind components. Note that the subtraction from  $270^\circ$  in Equation 2.5 is necessary to define the wind direction meteorologically, where North is represented by both  $0^\circ$  and  $360^\circ$  (NCAR n.d.). Equations 2.6 and 2.7 are used to calculate standard deviation for wind speed and direction, respectively, formulated via propagation of uncertainty assuming  $u$  and  $v$  are independent. These equations must be carefully applied to the data when coding, as each language may perform the operations differently. The Python package MetPy (May et al. 2022) was utilized in this study to directly calculate mean wind speed and wind direction from  $u$  and  $v$ .

$$\overline{WSPD} = \sqrt{\overline{u^2} + \overline{v^2}} \quad (2.4)$$

$$\overline{WDIR} = 270 - \tan^{-1} \left( \frac{\overline{v}}{\overline{u}} \right) * \frac{180}{\pi} \quad (2.5)$$

$$s_{\overline{WSPD}} = \sqrt{\left(\frac{\bar{u}}{\overline{WSPD}}\right)^2 s_u^2 + \left(\frac{\bar{v}}{\overline{WSPD}}\right)^2 s_v^2} \quad (2.6)$$

$$s_{\overline{WDIR}} = \frac{180}{\pi} \frac{1}{1 + \frac{\bar{v}^2}{\bar{u}}} \left| \frac{\bar{v}}{\bar{u}} \right| \sqrt{\left(\frac{s_v^2}{\bar{v}}\right)^2 + \left(\frac{s_u^2}{\bar{u}}\right)^2} \quad (2.7)$$

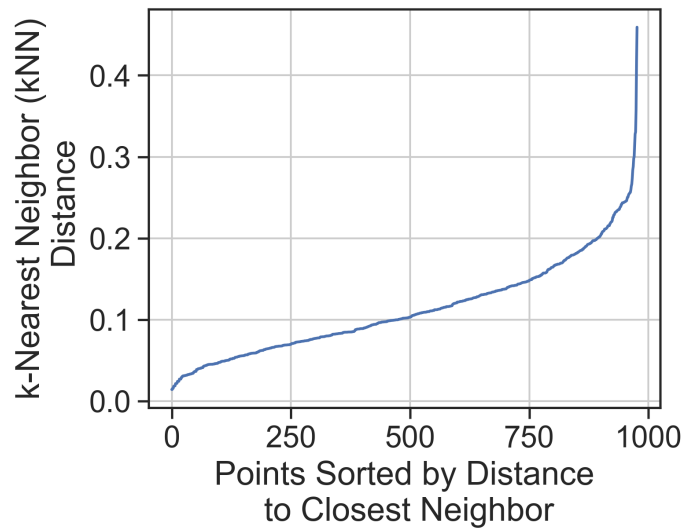
### 2.4.3 Hyperparameter Tuning for DBSCAN and UMAP

Once the data are standardized, the iterative process of tuning hyperparameters and processing the data through the UMAP and DBSCAN algorithms can begin. As previously mentioned, unsupervised machine learning techniques like the ones used here are subjective. It is up to the user to appropriately tune hyperparameters and evaluate if the products from an algorithm appear to reasonably depict the data. Both UMAP and DBSCAN have a variety of hyperparameters that the user can manipulate to produce the subjectively best results.

The UMAP algorithm contains a variety of hyperparameters. The two hyperparameters changed from default values in this study are *n\_components* and *n\_neighbors*. The *n\_components* parameter simply determines the number of dimensions that the reduced dimension dataset should have (McInnes et al. 2018). An *n\_components* value of 3 was chosen here as a compromise between a less drastic dimensional reduction (higher *n\_components* would be ideal) and an easily visualized dataset (lower *n\_components* results in fewer plots to represent the reduced dataset). The *n\_neighbors* parameter, however, determines if UMAP will focus on small-scale, local structures within the data (lower values of *n\_neighbors*) or more global structures (higher values of *n\_neighbors*). This parameter should be kept between 2 and 200, but should otherwise be tuned

repeatedly by the user to find the best representation of the data (McInnes et al. 2018).

As mentioned previously, the two necessary hyperparameters for DBSCAN are the radius of the epsilon neighborhood and the density threshold of the neighborhood, represented by *eps* and *min\_samples*, respectively in the scikit-learn Python library (Pedregosa et al. 2011). The *min\_samples* parameter should be at least one greater



than the original dimension of the data when the dataset has at least three dimensions. The

Figure 2.9: The sorted *k*-distance graph for the high variability  $XCO_2$  dataset, where  $k = 20$ . The optimal value for *eps* should be approximately 0.26.

*eps* parameter can be determined using a sorted *k*-distance graph (Figure 2.9). The distance between each point and its *k*-th nearest neighbor are plotted in order of distance, where *k* should be one less than the dimension of the original dataset. The distance where the slope of the curve changes most dramatically (the elbow of the plot) should be used for *eps* (Ester et al. 1996).

## 2.4.4 Application of Machine Learning Techniques

Before applying cluster analysis techniques to the TRACER dataset, the data underwent minor pre-processing. First, near-surface meteorological data from each deployment site were combined with corresponding EM27/SUN data to provide context on atmospheric conditions during the deployments. Both meteorological and EM27/SUN data were averaged over one-minute periods to standardize the temporal resolution of the dataset. A single day of data

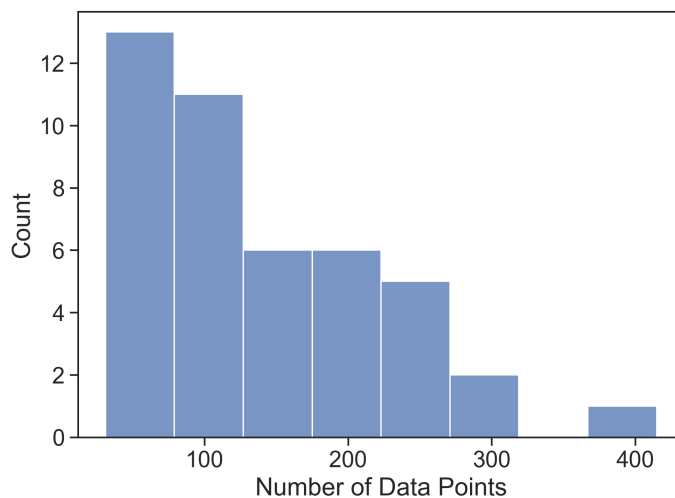


Figure 2.10: The distribution the of the number of data points in the daytime datasets collected during the TRACER Campaign. Only daytime datasets with greater than 30 one-minute averaged data points are included.

would only be considered for further analysis if more than 30 data points existed within the 10:00 to 16:00 local time window after averaging, in accordance with the Central Limit Theorem. Almost 90% of the daily TRACER datasets contained more than 30 data points during the daytime window. The distribution of daytime (within the 10:00 to 16:00 local time window) dataset length is provided in Figure 2.10.

For each daytime dataset, the standard deviation of  $X_{CO_2}$ ,  $X_{CO}$ , and  $X_{CH_4}$  were calculated within the 10:00 to 16:00 time window. These distributions were used to separate daily datasets into high and low variability categories for each  $X_{gas}$

using the 75<sup>th</sup> percentile. High variability groups were expected to contain data collected during uncommon circumstances compared to the low variability groups, such as unique changes to local emission sources or infrequently observed changes in meteorological conditions. Low variability groups, however, were expected to isolate common conditions associated with heightened or lowered pollutant concentrations, while high variability groups were thought to contain data collected during more anomalous circumstances. Data partitioning proved necessary for the combined UMAP and DBSCAN approach, as continued analysis with the full TRACER dataset did not yield significant insight.

Once the daily datasets had been assigned to the high or low variability groups, each of the groups were z-score standardized using Equation 2.3 to prepare them for dimensionality reduction via UMAP and later, cluster analysis with DBSCAN. Seed 20220718 was set in the UMAP code for repeatability, as discussed above. This seed was chosen pseudo-randomly, as the results it generated were not considered while choosing the seed.

Each three-dimensional dataset was passed through DBSCAN for clustering twice, where the second DBSCAN iteration was run on the previous iteration's unclustered data. During this process, the data standardization techniques discussed in Section 2.4.2 were applied as stated to the first DBSCAN iteration. In order for the second iteration of DBSCAN to identify any new clusters, the hyperparameter tuning approach was modified to enhance *eps* by around 5%.

Finally, "high concentration" and "low concentration" clusters were isolated for further analysis according to the feature of interest mean of the cluster. For example,  $X_{CO_2}$  high and low concentration clusters were defined according to Equations 2.8 and 2.9, respectively, where  $s_{X_{CO_2}}$  represents the standard deviation of  $X_{CO_2}$  in the

dataset. Equation 2.10 represents the same relationship for determining a high concentration cluster as Equation 2.8 using the standardized, z-score notation. Similarly, Equation 2.11 shows the z-score notation for determining a low concentration cluster. While useful for dividing up data, this method does have one important caveat: since dataset-wide statistics are used to compare to, seasonal trends inadvertently impact the magnitude of  $\overline{XCO_2}_{dataset}$ . An alternative approach to defining high and low concentration clusters should be explored in future work.

$$\overline{XCO_2}_{high\ concentration\ cluster} = \overline{XCO_2}_{dataset} + s_{XCO_2, dataset} \quad (2.8)$$

$$\overline{XCO_2}_{low\ concentration\ cluster} = \overline{XCO_2}_{dataset} - s_{XCO_2, dataset} \quad (2.9)$$

$$\overline{z_{XCO_2}}_{high\ concentration\ cluster} > 1 \quad (2.10)$$

$$\overline{z_{XCO_2}}_{low\ concentration\ cluster} < -1 \quad (2.11)$$

This data analysis approach, including data preparation and partitioning for analysis, is summarized in the flowcharts contained in Figures 2.11 and 2.12.

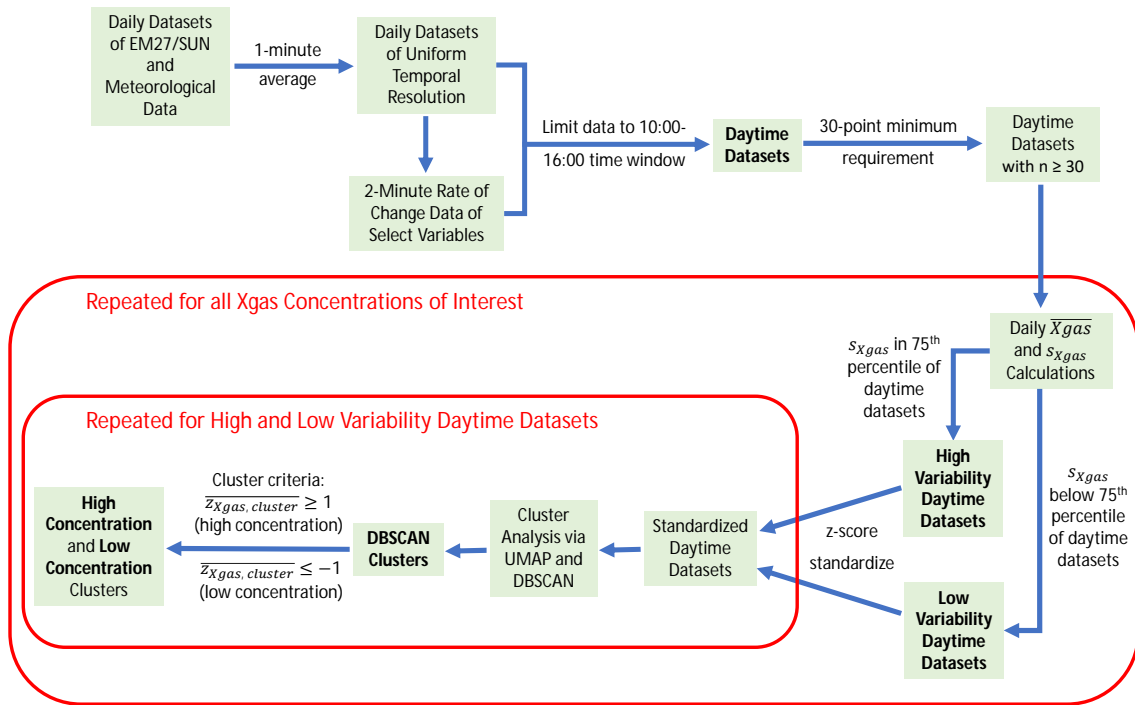


Figure 2.11: Flowchart of the data preparation and grouping utilized during cluster analysis. Specifics about the UMAP and DBSCAN implementations can be found in Figure 2.12.

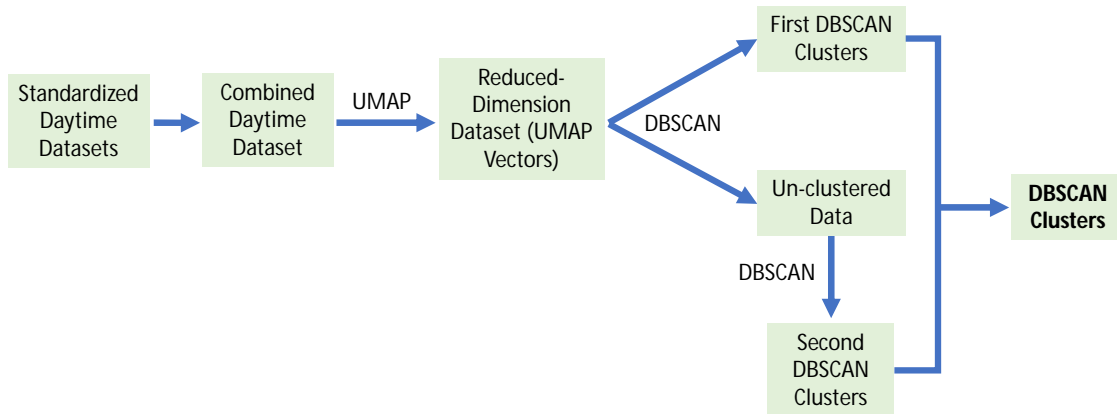


Figure 2.12: Flowchart of the UMAP and DBSCAN applications utilized for this study



## Chapter 3

### EM27/SUN Data Quality Assurance

EM27/SUN data quality, accuracy with respect to other spectrometers, and retrieval algorithm performance are explored throughout this chapter. Section 3.1 discusses the data quality control filters developed with both the TCCON calibration campaign and TRACER Campaign datasets to limit erroneous EM27/SUN data that could hinder analysis. Then, Section 3.2 evaluates the accuracy of EM27/SUN data with respect to TCCON when EM27/SUN data are processed with GGG2020 and PROFFAST version 2.2 retrieval algorithms.

#### 3.1 Quality Control Filtering

Before analyzing the EM27/SUN data, the retrieved  $X_{gas}$  values must undergo a series of data quality checks to ensure that the final dataset is representative of the conditions in the atmosphere and is void of artificially heightened concentrations. During the retrieval process, bad quality data are identified for a variety of reasons (low solar intensity, low/high internal temperature, etc.) by the algorithm. By default, PROFFAST removes any data that fails quality checks before presenting the data for the end user instead of providing all processed data with any associated flags. The GGG algorithm, however, assigns bad quality flags to certain data points for the end user to remove. For retrieved data from both algorithms, the first step to acquiring a

quality-filtered dataset is to remove any data that have been assigned quality flags. Once this has been done, all EM27/SUN data pass through three preliminary data quality filters:

1. Any data collected with an associated  $X_{air}$  (generated by PROFFAST) or  $X_{luft}$  (generated by GGG) value outside of the mole fraction range of  $0.97 \leq (X_{air} \text{ or } X_{luft}) \leq 1.03$  are considered unreasonable and must be removed. As previously mentioned,  $X_{air}$  and  $X_{luft}$  should ideally always equal 1. The acceptable parameter bounds defined above provide a stricter filter than the example quality control filters for a TCCON station (the Park Falls TCCON station, specifically) which are provided with GGG2020, discussed in detail in Section 2.3. The  $X_{air}$  or  $X_{luft}$  values outside of this range occur rarely, but can be more common for EM27/SUN than TCCON instruments due to the higher precision, temperature-stabilized instrumentation used in TCCON. Whenever  $X_{air}$  or  $X_{luft}$  is outside of this range, an EM27/SUN is considered to be operating improperly and cannot record raw data accurately.
2. Any data collected at a solar zenith angle (SZA) greater than  $75^\circ$  are removed from the dataset, as these values are expected to have high airmass, or SZA, dependence, which can artificially heighten retrieved concentrations (Sha et al. 2020). Data taken with large SZA only occur in the early morning or late evening when the sun is near the horizon, making the column of air that the instrument observes very long and more difficult to accurately observe. These data have been removed in lieu of more advanced data processing.
3. Concentrations of 0 ppm or 0 ppb have been removed, as these concentrations of the trace gases of interest do not exist in the real atmosphere.

4. GGG occasionally returns excessively large concentrations when the algorithm is not able to properly process raw data. Values of 99999.9 ppm for  $X_{H_2O}$  and  $X_{CO_2}$ , 99999.9 ppb for  $X_{CO}$ , and 999.9999 ppm for  $X_{CH_4}$  are reported in these cases. All of these values are removed during the filtering process.

### 3.1.1 Combating Solar Tracking Issues

Whenever an EM27/SUN has difficulty following the path of the sun, it may continue to collect data if part of the sun can still be seen by the optics. This happened multiple times during the TRACER Campaign and TCCON calibration campaign due to transient cloud cover and solar tracking failures. While the specific source of these solar tracking failures is still being explored, the impact on the data is understood.

If an EM27/SUN experiences a solar tracking failure, the external optics and solar tracking software will fail to keep the EM27/SUN optics aligned with the sun. Eventually, the sun will move far enough away that no direct sunlight will be reflected to the internal optics of the instrument. At this point, the recorded interferograms will record very low solar intensities, though notably higher than the solar intensity thresholds set in the retrieval algorithm quality control filtering, resulting in a lack of retrievable *Xgas* data. Before the EM27/SUN fully loses sight of the sun, however, it continues to record data using the limited direct solar radiation it receives. While these interferograms can still be processed through a retrieval algorithm, the resulting *Xgas* are often unrealistic due to incorrect assumptions about the path length. If an EM27/SUN is in the process of a tracking failure, it will continue to collect these lower intensity interferograms for around 15 minutes if the optics are not re-aligned by the operator before the instrument fully loses sight of the sun.

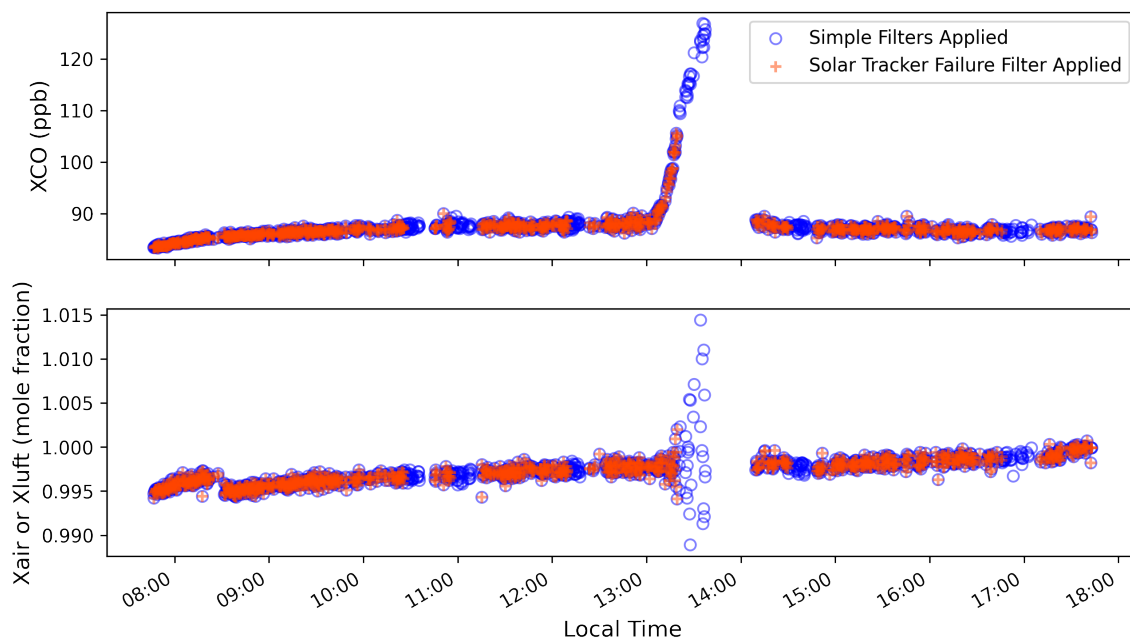


Figure 3.1: Preliminary data quality and solar tracking failure filters applied to EM27/SUN data collected on May 29, 2022 by Blanche the GeoCarb EM27/SUN during the TRACER Campaign at the UH Coastal Center site

The first filter created to address loss of tracking episodes was a simple bulk data removal. The filter looks for periods of time during the day when  $X_{gas}$  are missing for five minutes or longer. This length of time without data is unlikely to occur unless the EM27/SUN has completely lost solar tracking. If the code encounters a period of missing data, it removes the prior 30 data points, approximately 15 minutes of raw EM27/SUN data. This time estimate holds some uncertainty, as the EM27/SUN data are collected in uneven time intervals (approximately two data points per 30 seconds). Figure 3.1 demonstrates the effect of this filter on data collected on May 29, 2022 by the GeoCarb EM27/SUN during the TRACER Campaign at the UH Coastal Center. While some suspect data around 13:30 local time have been removed, the filtered  $XCO$  signal still demonstrates a sharp increase. Looking closely at  $X_{luft}$ , the scatter is large

at the same time as the *XCO* signal spikes. This indicates that the *XCO* concentration increase is not due to an increase in CO in the total column, as the *Xluft* signal is uncharacteristically disturbed at the same time. Therefore, additional filtering is necessary to further address unreasonable data. Without additional filtering, sharp concentration increases like the *XCO* spike seen here could be incorrectly interpreted as real signal increases by an untrained data user.

### 3.1.2 Filtering Artificially Heightened Concentrations

While the previously discussed filters are useful for catching egregiously erroneous data, they fail to assess more subtle impacts on data quality. For example, an EM27/SUN may begin to lose solar tracking, only to be quickly corrected by an operator within a couple of minutes. Though this would still constitute a loss of solar tracking, the solar tracking failure filter would not properly identify the issue, as the EM27/SUN would not have lost direct sight of the sun for a sufficient amount of time. However, in this time, *Xgas* may be artificially magnified as the column of air the EM27/SUN observed was obscured by the operator's presence.

Many methods were explored for omitting artificially heightened concentrations while preserving good quality data. Ultimately, a filter was designed utilizing the percent change of both *Xluft* (or *Xair*) and *XCO* for quality assurance. As mentioned previously, *Xair* and *Xluft* are data quality metrics, making them ideal candidates for filtering. However, *XCO* can be useful for filtering as well, since it responds as much as three orders of magnitude greater than the *Xair* or *Xluft* response.

The data quality filter designed for removing artificially heightened concentrations utilizes the percent change of a moving 20-point window of both *Xluft* (or *Xair*) and

$XCO$ . If  $XCO$  changes more than 2% or if  $X_{lufft}$  (or  $X_{air}$ ) changes more than 0.1% from the twentieth point before it, the data from that timestamp are removed. The size of the moving window, along with the percent change thresholds, were determined empirically with both the TCCON calibration and TRACER campaign datasets. Figure 3.2 demonstrates the removal of points by percent change using the May 29, 2022 data previously filtered with the preliminary and solar tracking failure filters.

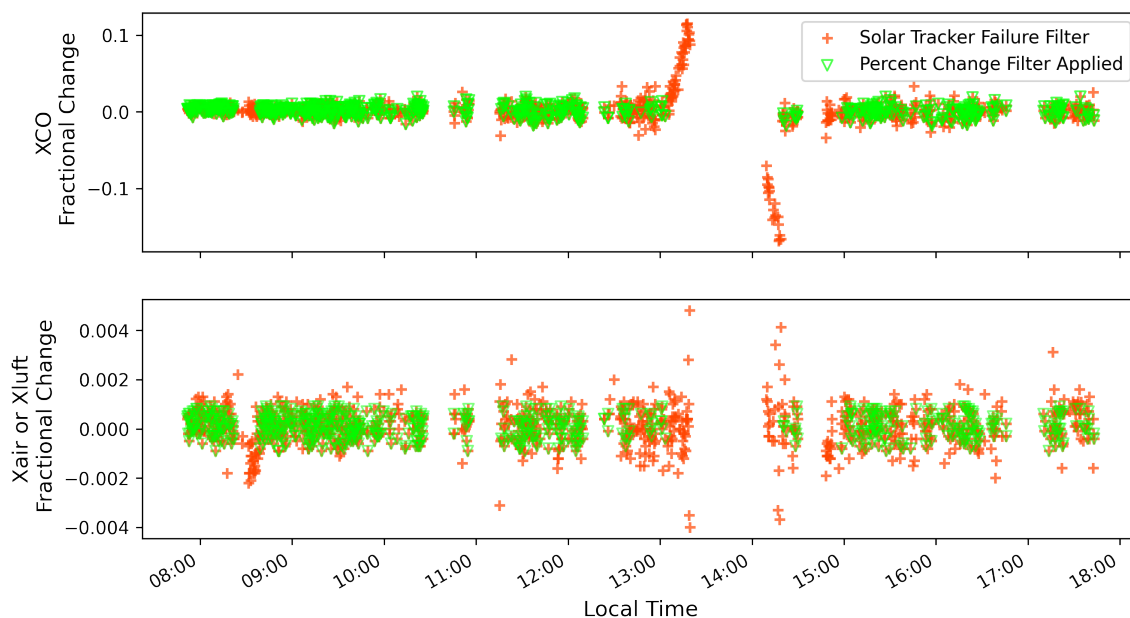


Figure 3.2: Application of the percent change filter on  $XCO$  and  $X_{lufft}$  data collected on May 29, 2022 by the GeoCarb EM27/SUN during the TRACER Campaign at the UH Coastal Center site. These data were previously filtered with the preliminary data quality and solar tracking failure filters as show in Figure 3.1.

Rigorous tuning of the percent change filter criteria effectively removed the majority of data impacted by loss of tracking episodes, but occasionally left artificially heightened points behind. More stringent criteria tended to remove data from periods of time without tracking issues or other quality concerns. To remove the remaining artificially heightened concentrations, the data point before and after a removed point by this filter

were also removed. Additionally, the point on either end of a 10-point span of filtered out data were removed as well. These additional checks appeared to appropriately cull the datasets, leaving the most representative data of the observed atmosphere behind. The final filtered data from May 29, 2022 are plotted in Figure 3.3.

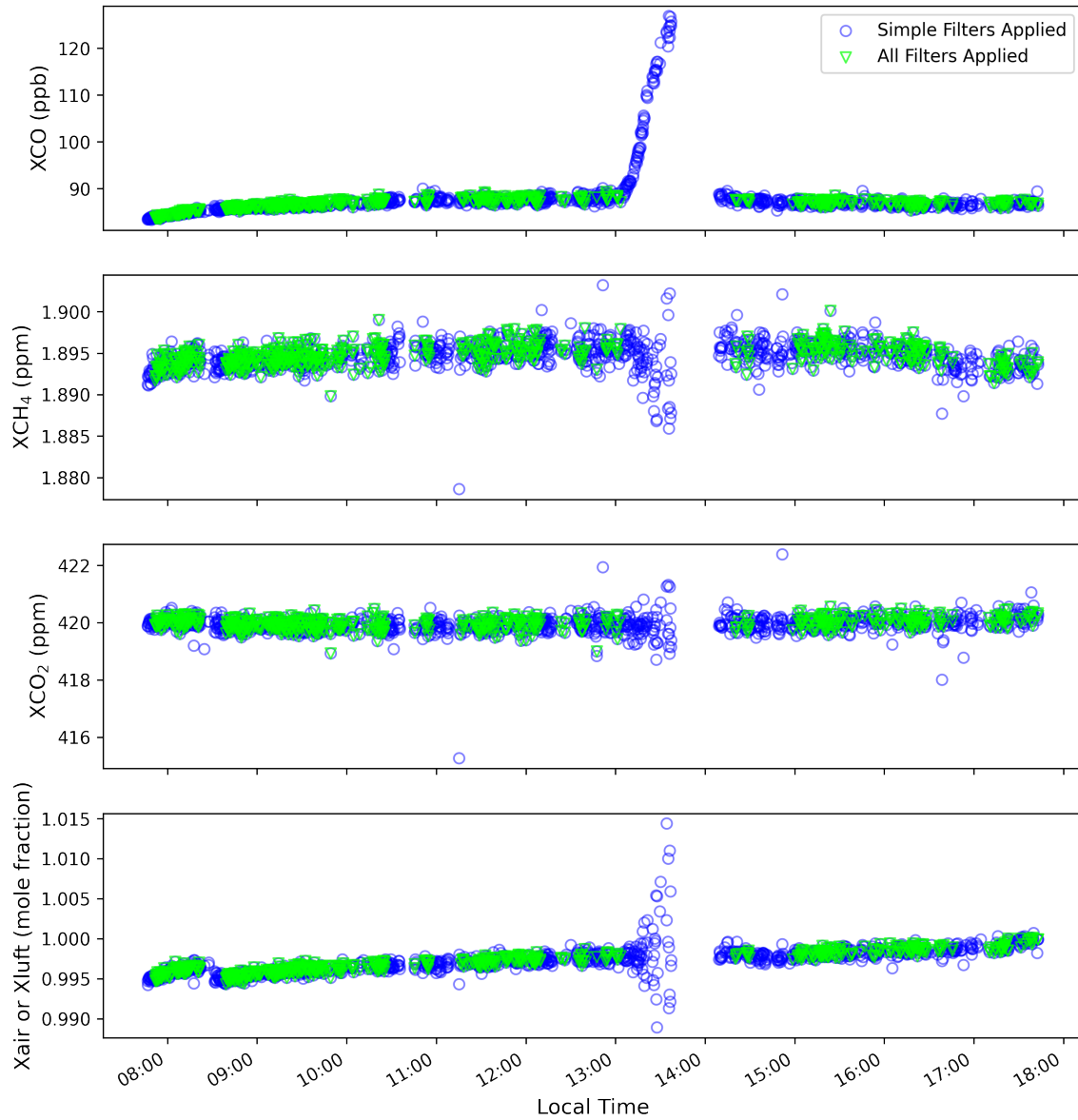


Figure 3.3: Comparison of May 29, 2022 EM27/SUN data with preliminary and entire collection of data quality filters applied

While this percent change filter presents an effective approach to data omission via parameter tuning, a more extensible method may be available. Section 5.2 discusses alternative methods for quality filtering based on empirical and statistical approaches. These methods should be explored during future data analysis.

### 3.1.3 An Exception to the Rule

While the filters described in previous sections appeared effective for most of the EM27/SUN data collected during the campaigns, one exception is notable. On October 17, 2022, the first day of the TCCON calibration campaign, the EM27/SUN and TC-



CON instruments all collected data from atmospheric columns that contained smoke from a large nearby fire, seen in Figure 3.4. As a consequence, these

Figure 3.4: Blanche and the two guest EM27/SUN instruments at SGP on 10/17/22, where a plume of smoke blew into the viewing path of the instruments.

columns were rich in CO in the lower atmosphere. The transient behavior of the dispersed smoke plume led to frequent large changes in  $XCO$ , unlike any other data collection period in either the TCCON calibration or TRACER campaigns. Due to the presence of the smoke plume, all filters except for the  $XCO$  percent change filter described above were applied to the EM27/SUN data collected on this day.



Due to the nature of the *XCO* percent change filter, most of the data sampled when the smoke plume intersected the sampled column of air would have been removed if this filter were applied, demonstrated using data from Blanche in Figure 3.5. The green points in Figure 3.6 demonstrate the effect on the data when all filters are applied except for the *XCO* percent change. These are the data points saved for Blanche on this date.

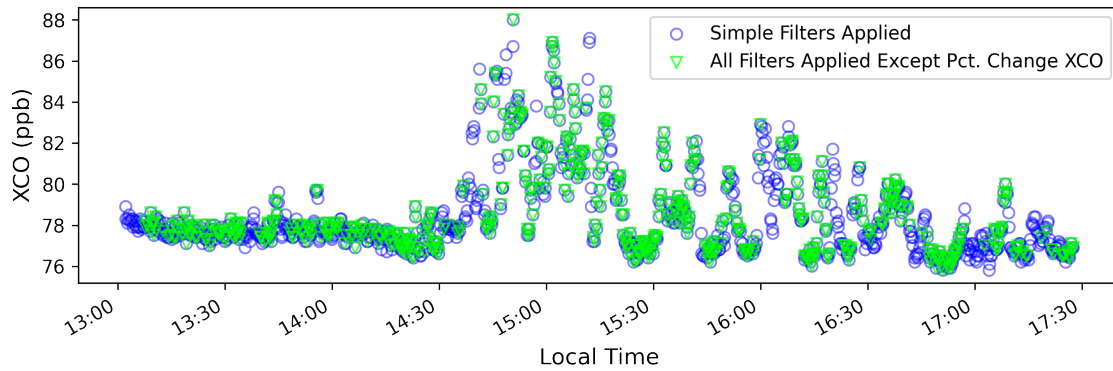


Figure 3.5: *XCO* data collected with Blanche from October 17, 2022 with all filters except for the *XCO* percent change filter applied to the points in blue, resulting in the filtered (green) points.

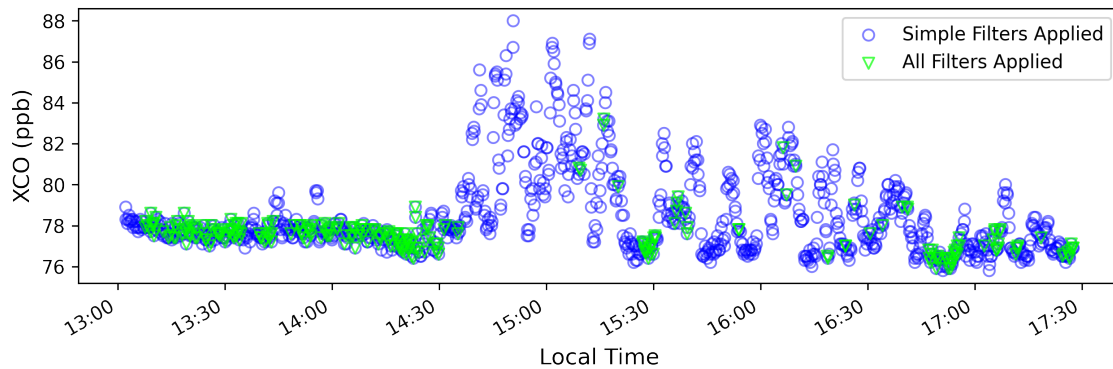


Figure 3.6: Blanche *XCO* data from October 17, 2022 with all filters, including the *XCO* percent change filter, applied. There is a lack of data preserved (green points) while observing the smoke plume.

The  $XCO$  signals from all EM27/SUN instruments and the TCCON station were in close agreement while sampling the smoke plume, so the data are not considered to be bad quality. Figure 3.7 demonstrates the agreement between Blanche and the TCCON station during the period of the TCCON calibration campaign where the three EM27/SUN instruments were collocated with the TCCON station. More about this comparison and the associated bias estimates are discussed in Section 3.2.2. The box plot on the left of Figure 3.7 demonstrates the bias of Blanche from TCCON using all data during the three-day collocation period, including the data collected while sampling the smoke plume.

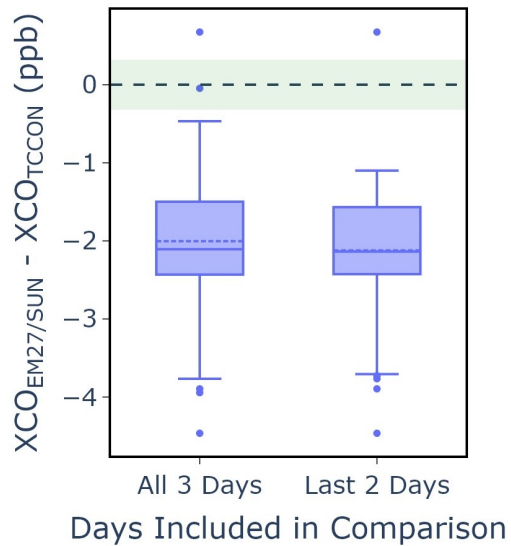


Figure 3.7: Bias in  $XCO$  between Blanche and TCCON from October 17-19, 2022. The data from October 17 were removed from the box plot on the right, while data from all three days are included in the box plot on the left.

The box plot on the right leaves out all data from October 17. The median bias does not appreciably change between these comparisons, indicating that the presence of erratic  $XCO$  from the smoke plume did not impact the agreement between the instruments. Since Blanche and the TCCON station effectively agree during the smoke observation period, the  $XCO$  percent change filter can soundly be neglected for circumstances like these (i.e., directly sampling a smoke plume) in order to keep these unique observations as good quality data.

## 3.2 Retrieval Algorithm Comparison

Once the EM27/SUN data were processed with both retrieval algorithms and quality filtered, collocation with TCCON during the TCCON calibration campaign provided the opportunity to evaluate the two retrieval algorithms. Due to the differences discussed in Section 2.3, each retrieval algorithm provides slightly different  $X_{gas}$  from the same raw EM27/SUN dataset. Evaluating these differences using data from a comparison standard is a useful approach to decide which algorithm to use, as this comparison also ensures that the EM27/SUN data are in line with the standard dataset. As mentioned previously, TCCON stations provide the ideal standard of comparison for the EM27/SUN instruments.

After the TRACER Campaign, from October 17 through October 19, 2022, the main three EM27/SUN instruments used in Houston were transferred to the SGP ARM site for side-by-side measurements with the Lamont, OK TCCON station. While Blanche data collected at SGP prior to TRACER allowed for the agreement between Blanche and TCCON to be evaluated, the co-deployment of three EM27/SUN instruments allowed for the analysis of two important additional metrics:

1. Lack of SZA-dependence of the data.
2. Agreement between Blanche and the guest EM27/SUN instruments.

While preparing data collected by Blanche for the SZA-dependence and inter-instrument evaluation, an initial comparison of the data processed by the GGG and PROFFAST revealed an unexpected difference between the resulting  $X_{gas}$  data. Time series of Blanche's data retrieved with GGG exhibited a sawtooth-like pattern for all  $X_{gas}$ , shown in Figure 3.8. The TRACER dataset exhibited this behavior as well,

which was theorized to be caused by an unstable laser in the spectrometer. While the TCCON instruments are built with temperature-stabilized lasers, the EM27/SUN instruments are not, which minimizes their cost and size. Therefore, the location of spectral peaks collected by a TCCON instrument are less likely to change than the spectral peaks of an EM27/SUN with changing atmospheric temperatures. Retrieval algorithms specify a frequency shift, which adjusts where the code expects to find spectral information about the atmospheric species of interest. Since the primary purpose of GGG is to process TCCON spectra, this frequency shift is small. Within GGG, the  $\delta\nu$  parameter can be modified to allow for a larger frequency shift, or wavelength shift, depending on your perspective, to properly handle data from an instrument with an unstable laser.

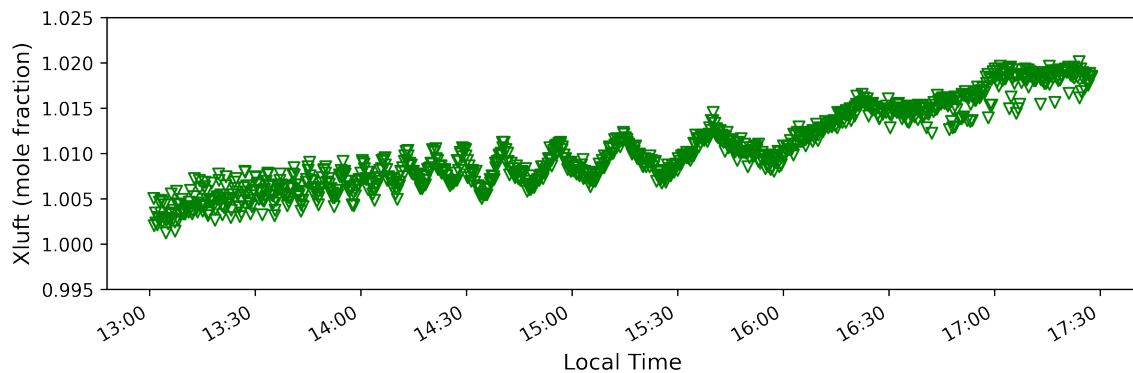


Figure 3.8: Preliminarily filtered *Xluft* retrieved with GGG. These data were collected with Blanche on October 17, 2022.

The  $\delta\nu$  parameter was adjusted for Blanche’s data in accordance with the procedure in Appendix B, which successfully removed the sawtooth pattern from the retrieved data (Figure 3.9). Additionally, air-mass independent and dependent correction factors (AICF, ADCF) from Mostafavi Pak et al. (2023) were applied *post hoc* to all

GGG-retrieved EM27/SUN data to minimize the SZA-dependence and correct the data to WMO standards, respectively, resulting in a new, modified GGG dataset.

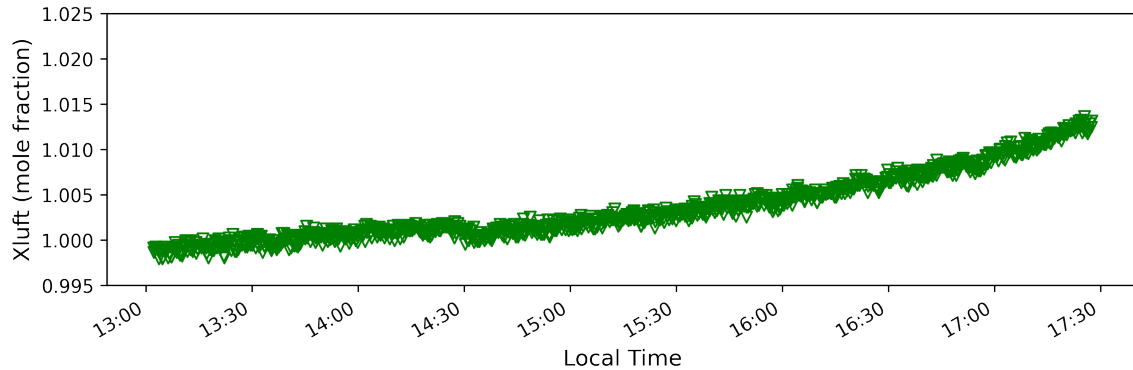


Figure 3.9: Preliminarily filtered  $X_{luft}$  retrieved with a modified version of GGG with an adjusted  $\delta\nu$  parameter to remove the sawtooth pattern from the data as well as air-mass independent and dependent correction factors from (Mostafavi Pak et al. 2023). These data were collected with Blanche on October 17, 2022.

### 3.2.1 Evaluation of Solar Zenith Angle Dependence

The SZA-dependence of the data was analyzed for each of the three datasets, including the modified GGG dataset. For this analysis, the daily mean for  $X_{CO_2}$ , denoted as  $\overline{X_{CO_2}}$  in the figure below, was subtracted from the one-minute-averaged daily dataset and plotted as a function of SZA. This bias from the daily mean should ideally be constant and equal to zero across a wide SZA range (Mostafavi Pak et al. 2023). Figure 3.10, created for Blanche’s data during the TCCON calibration campaign, demonstrates that the modified GGG dataset provided the least dependent data on SZA.

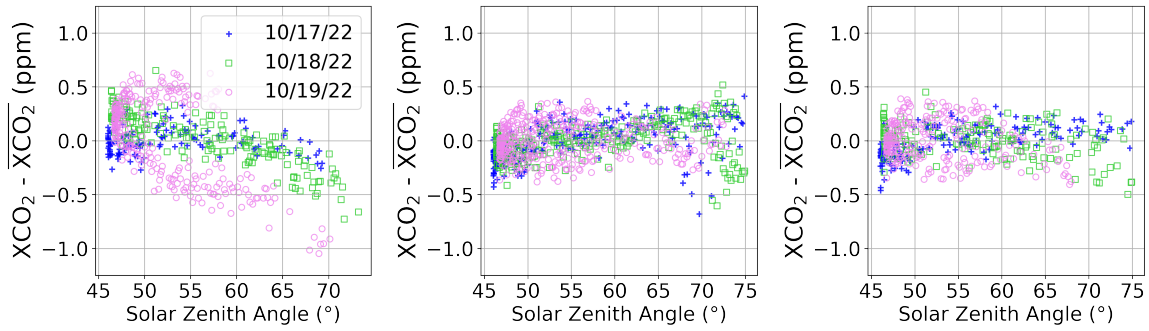


Figure 3.10: Solar zenith angle dependence during the TCCON calibration campaign using Blanche’s data retrieved using PROFFAST version 2.2 (left), unmodified GGG2020 (center), and GGG2020 with  $\delta\nu$ , AICF, and ADCF modifications (right)

### 3.2.2 EM27/SUN Comparison to TCCON

Comparing Blanche to TCCON during the October 17-19, 2022 EM27/SUN comparison period affirmed the use of modified GGG to process EM27/SUN data. For this comparison, EM27/SUN and TCCON data (Wennberg et al. 2022) were averaged over ten-minute periods to minimize the difference of temporal resolution between the instruments. Then, the difference in  $X_{gas}$  between Blanche and the TCCON station was calculated to evaluate the agreement between the instruments (Figure 3.11).

Modified GGG provided the smallest bias between the instruments for both  $XCO_2$  and  $XCH_4$ . The modified GGG-retrieved bias in  $XCO_2$  was reduced to nearly zero, where zero represents perfect agreement between the instruments. Additionally, the median  $XCO_2$  and  $XCH_4$  biases presented are well within the TCCON uncertainties of 0.64 ppm for  $CO_2$  and 6.1 ppb for  $CH_4$ , respectively, estimated by Mostafavi Pak et al. (2023) using GGG2020-retrieved TCCON data. These biases are also congruent with the experimental findings of Mostafavi Pak et al. (2023), where multiple EM27/SUN instruments were collocated with the SGP TCCON station to determine the SGP-specific biases of the EM27/SUN instruments.

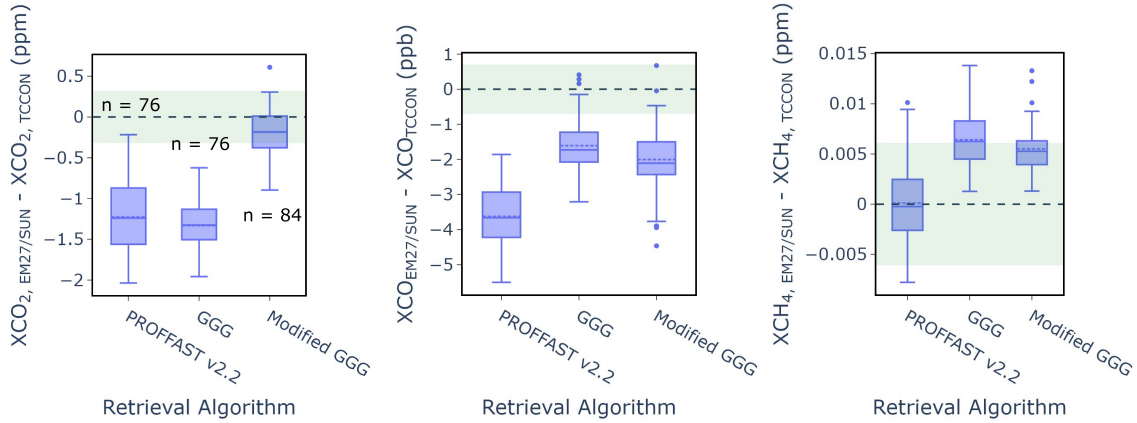


Figure 3.11: Bias of Blanche compared to the SGP TCCON during the calibration campaign for  $XCO_2$  (left),  $XCO$  (center), and  $XCH_4$  (right). The green shading represents the estimated uncertainty of TCCON as reported by Mostafavi Pak et al. (2023). The sample sizes ( $n$ ) reported for  $XCO_2$  are consistent for all  $Xgas$ .

Median biases in  $XCO$  were larger than anticipated when compared to both the experimental SGP-specific bias, near 0 ppb, and the estimated TCCON bias of 1.4 ppb (Mostafavi Pak et al. 2023). Omitting data from October 17 during the smoke observation period, discussed in Section 3.1.3, did not decrease the median bias. Despite the non-ideal bias estimated with this comparison, the improvement between PROFFAST-retrieved data to GGG-retrieved data is notable. Nevertheless, these results suggest that an offset is necessary for correcting Blanche’s  $XCO$  data.

Generally speaking, the comparison between Blanche and the SGP TCCON demonstrated minimal median biases during the TCCON calibration campaign. These biases were largely minimized by the modified GGG retrieval algorithm. Therefore, this modified retrieval approach appears to provide the most accurate interpretation of the EM27/SUN data evaluated against the comparison standard of TCCON.

### 3.2.3 EM27/SUN Intercomparison

To further assess the accuracy of modified GGG-retrieved data from each EM27/SUN used during the TRACER Campaign,  $X_{gas}$  from each guest EM27/SUN was compared directly to data from Blanche. Similar to the comparison with TCCON, the bias between Blanche and the other EM27/SUN instruments was minimized with modified GGG for  $X_{CO_2}$  (Figure 3.12). Bias between Blanche and the guest EM27/SUN instruments was also minimized for  $X_{CH_4}$  and  $X_{CO}$ , though retrieval algorithm choice appeared to have minimal impact on  $X_{CO}$ , as seen in the left panel of Figure 3.13. The increase in  $X_{CH_4}$  bias magnitude for unmodified GGG-retrieved data was due to the sawtooth pattern induced by the limited frequency shift by the retrieval algorithm. Again, the presence of the ADCF and AICF modifications in GGG appeared to provide the best representation of the dataset out of all the retrieval algorithms investigated.

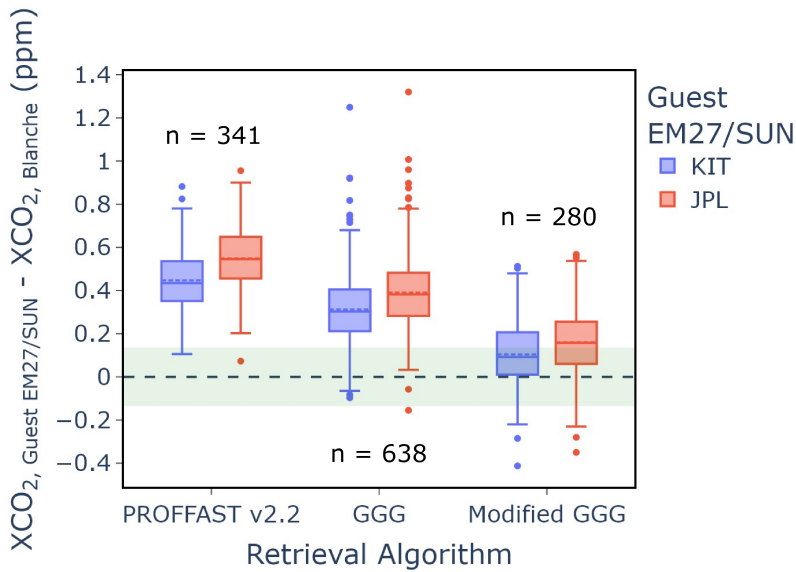


Figure 3.12: Bias of Blanche compared to the guest EM27/SUN instruments using one-minute averaged data for  $X_{CO_2}$ . The green band represents the bias between the SGP TCCON and Blanche, reported in Table 3.2.



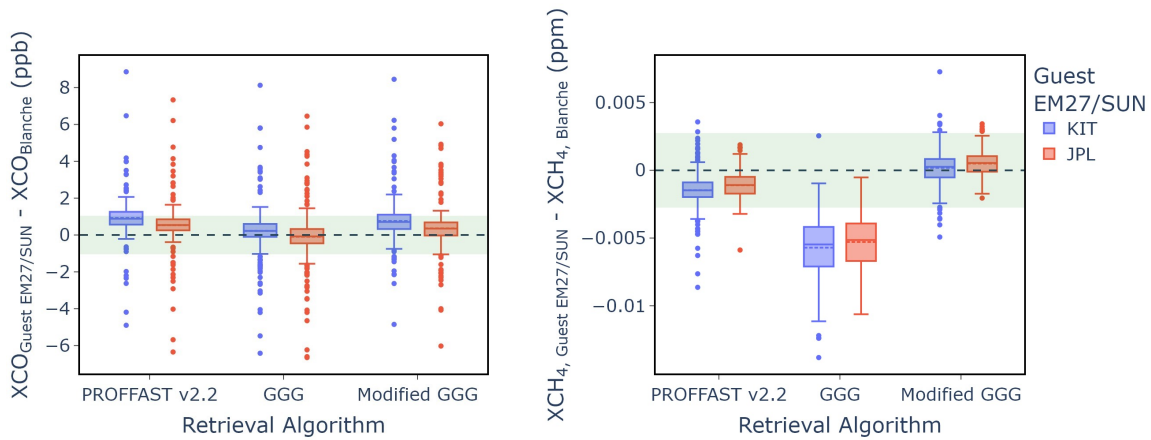


Figure 3.13: Bias of Blanche compared to the guest EM27/SUN instruments using one-minute averaged data for  $XCO$  (left) and  $XCH_4$  (right). The number of samples per comparison is consistent with Figure 3.12. The green bands represents the biases between the SGP TCCON and Blanche for each  $Xgas$ , reported in Table 3.2.

The GGG2020 dataset with air-mass dependent and independent correction factors from Mostafavi Pak et al. (2023), as well as and Blanche-specific  $\delta\nu$  modifications, presented the best representation of the EM27/SUN data from the TCCON calibration campaign with respect to SZA-dependence and bias between instruments. Therefore, the remainder of the data analysis will utilize datasets processed with this modified GGG approach.

### 3.2.4 Correction Factors for Instrument Bias

To account for biases between instruments, correction factors were created using the TCCON comparison dataset. In accordance with Mostafavi Pak et al. (2023), supplemental EM27/SUN data should be corrected with respect to a standard EM27/SUN. For this study, Blanche is the standard EM27/SUN. The median biases between the guest EM27/SUN instruments and Blanche using one-minute averaged modified GGG are provided in Table 3.1.

Table 3.1: Bias of guest EM27/SUN instruments from Blanche estimated during the TCCON calibration campaign such that  $Xgas_{Blanche-corrected\ guest} = Xgas_{guest} - bias$

<b>Guest EM27/SUN</b>	<b>Xgas</b>	<b>Median Bias From Blanche</b>	<b>RMSE</b>
KIT	XCO <sub>2</sub>	0.0938 ppm	0.1802 ppm
	XCO	0.7542 ppb	1.3765 ppb
	XCH <sub>4</sub>	0.0 ppm	0.0013 ppm
	XH <sub>2</sub> O	-1.2029 ppm	13.5108 ppm
JPL	XCO <sub>2</sub>	0.1687 ppm	0.2259 ppm
	XCO	0.3625 ppb	1.2163 ppb
	XCH <sub>4</sub>	0.0005 ppm	0.0011 ppm
	XH <sub>2</sub> O	-1.3713 ppm	11.6777 ppm

Additionally, comprehensive median biases between Blanche and the SGP TCCON were determined using the entire TCCON comparison campaign dataset. Figure 3.14 demonstrates the spread of XCO<sub>2</sub> measured during the campaign. The biases for each Xgas with respect to TCCON are provided in Table 3.2.

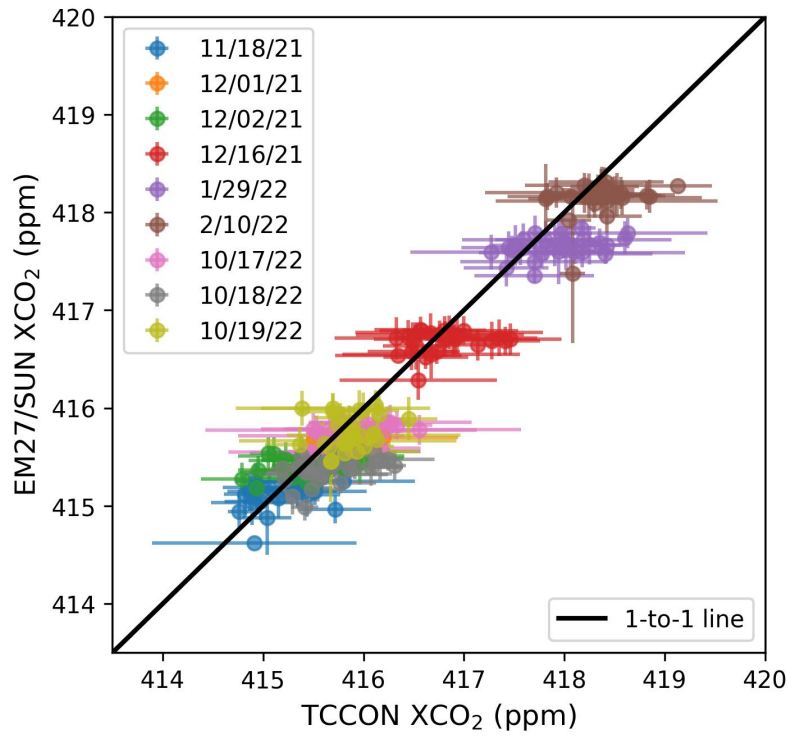


Figure 3.14: Bias of Blanche compared to the SGP TCCON during the TCCON calibration campaign. Instruments were compared using ten-minute averaged  $XCO_2$  data

Table 3.2: Difference between Blanche and TCCON  $X_{gas}$  during the TCCON calibration campaign

$X_{gas}$	Median Bias From TCCON	RMSE
$XCO_2$	-0.2711 ppm	0.3452 ppm
$XCO$	-2.0857 ppb	2.2435 ppb
$XCH_4$	0.0055 ppm	0.0070 ppm

## Chapter 4

### TRACER Campaign Results

Analysis of EM27/SUN data collected during the TRACER Campaign involves three main components: time series analysis on seasonal scales (Section 4.1), sub-seasonal investigation of *Xgas* trends using machine learning (Section 4.2), and remote sensor validation of OCO-2 and OCO-3  $XCO_2$  (Section 4.3)

#### 4.1 Seasonal Trends during TRACER

The EM27/SUN data collected during the TRACER Campaign spanned nearly five months, allowing for observation of diurnal and seasonal trends in *Xgas*. The TRACER dataset contained highly variable data, which were hypothesized to have been impacted by changes in boundary layer development, meteorological conditions, local emission sources, and, in the case of  $XCO_2$ , Northern Hemisphere seasonal changes. All quality filtered  $XCO_2$  data collected during the campaign are included in Figure 4.1 to demonstrate the level of variability present in the dataset.

A clear seasonal trend was observed with  $XCO_2$ , where the highest daily mean concentrations were predominantly observed during the beginning of the campaign. These mean concentrations generally decreased as the campaign progressed, with the lowest daily mean concentrations present in September, seen in both Figure 4.1 and the top panel of Figure 4.2. As discussed in Section 2.4.4, the time period 10:00 to

16:00 local time was chosen to analyze each daily dataset as most of the datasets contained observations during this time period. Additionally, this time window aligned with boundary layer conditions dominated by daytime processes.

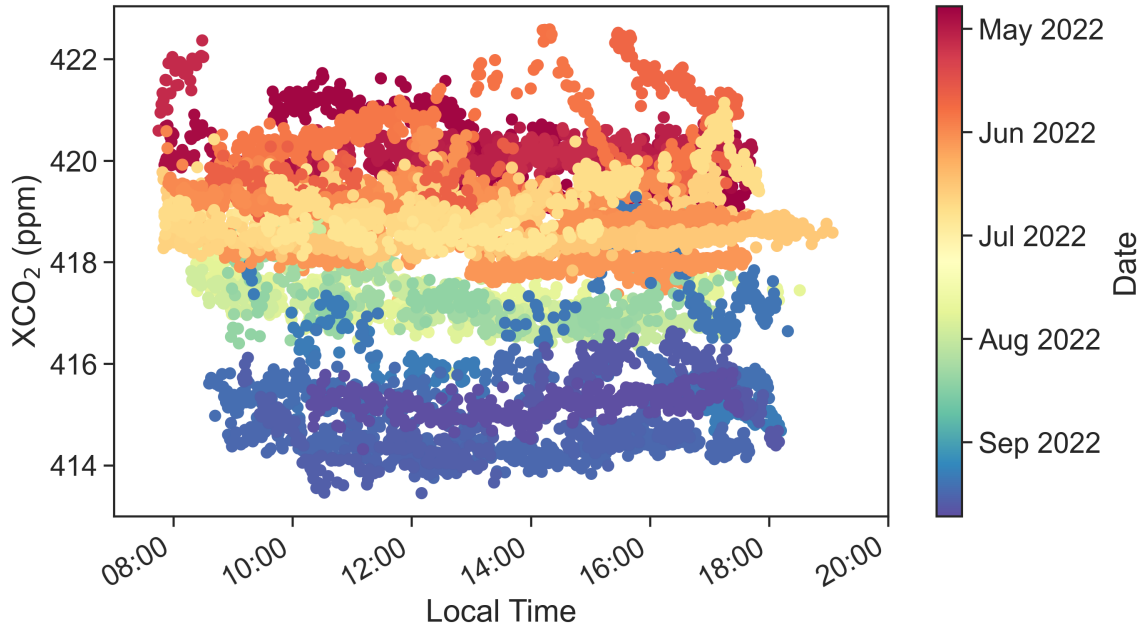


Figure 4.1: All quality filtered EM27/SUN  $XCO_2$  data collected during the TRACER Campaign

The decreasing trend of  $XCO_2$  suggests that on seasonal scales,  $CO_2$  concentrations in Houston follow the Northern Hemisphere  $CO_2$  cycle, controlled by natural processes. Therefore, the declining trend in daily mean concentrations from the beginning to the end of the campaign reflects the dominant role of background  $CO_2$  concentrations, caused by increased plant growth and photosynthesis during the Northern Hemisphere growing season. While the general trend of  $XCO_2$  is decreasing for the duration of the campaign, some days stand out from the trend with higher or lower mean concentrations. Additionally,  $XCO_2$  variability within each daily dataset changes

from day to day during the campaign. This motivated data analysis on smaller time scales to investigate the sources of variability and higher mean concentrations.

Sub-seasonal analysis is also beneficial for  $X_{CO}$  and  $X_{CH_4}$ , as these pollutants do not share the same vegetation cycle dependence as  $X_{CO_2}$ . The bottom two panels of Figure 4.2 demonstrates a lack of obvious seasonality for these gases. However, these gases exhibit similar day-to-day trends as  $X_{CO_2}$  with changes in both the daily mean and diurnal variability of the concentration values. Therefore, analysis on smaller temporal scales is necessary to understand why some days have higher concentrations and different levels of variability than others.

The variability of all  $X_{gas}$  as discussed here motivated the choice of data classification based on standard deviation for cluster analysis as described in Section 2.4. By isolating high variability days, sub-seasonal analysis attempts to identify unique events that impacted  $X_{gas}$ . These high variability days were targeted specifically to determine if the EM27/SUN data were able to resolve small-scale changes in pollutant concentrations and if the causes of these concentration changes could be identified. High and low variability days are reflected in Figure 4.2 as red and blue box plots, respectively. During the period between 6/23/22 and 6/25/22, the EM27/SUN instruments were co-deployed at different sites, which is why for these days Figure 4.2 shows more than one box plot. Table 4.1 details the high and low variability data collected during these three days.

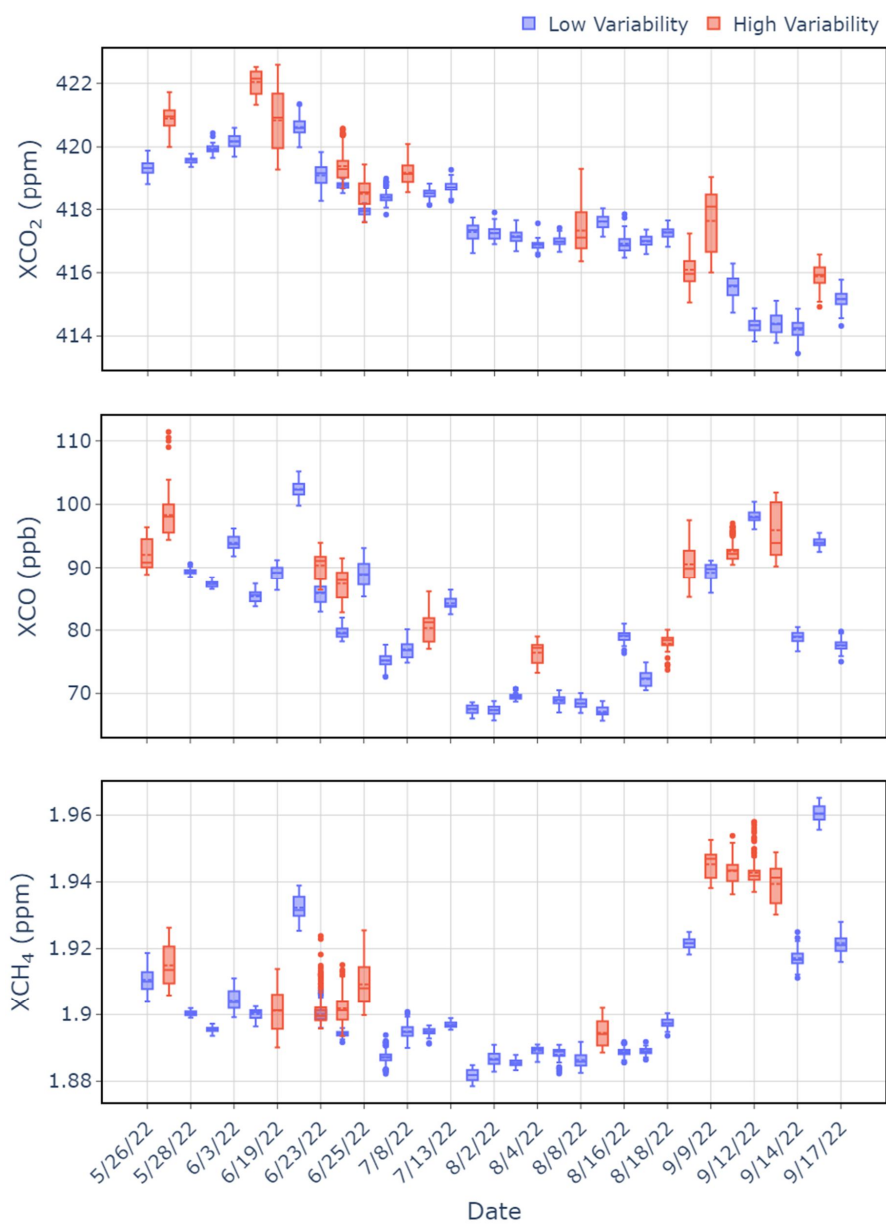


Figure 4.2: All EM27/SUN  $XCO_2$  (top),  $XCO$  (middle), and  $XCH_4$  (bottom) data collected between 10:00 and 16:00 local time during the TRACER Campaign. Each box represents data collected on a single day at one deployment site. Data have been separated into high and low variability daytime datasets as defined in Section 2.4. 6/23/22 through 6/25/22 have multiple box plots, as multiple EM27/SUN instruments were deployed at different sites during these days.

Table 4.1: Description of the high and low variability daytime datasets collected between 6/23/22 and 6/25/22

Day	Site	XCO <sub>2</sub>	XCO	XCH <sub>4</sub>
6/23/22	Site 6	Low	Low	High
	Site 7	Low	High	Low
6/24/22	Site 6	Low	Low	Low
	Site 7	High	High	High
6/25/22	Site 6	Low	Low	High
	Site 7	High	Low	High

In addition to the *Xgas* variations mentioned previously, the meteorological conditions in Houston were also highly variable. Figure 4.3 demonstrates the lack of obvious seasonality during the campaign aide from the decreasing temperatures and wind direction in mid-August and September. The proximity to the Gulf of Mexico and various bays were expected to contribute to this variability, due to the presence of sea, bay, and land breezes which commonly propagate through the area (Caicedo et al. 2019). The presence of these complex boundary layer processes and overall data variability further highlights the importance of cluster analysis for this study.

Since the purpose of cluster analysis is to identify trends within the state vector space that influenced a feature of interest, only days with every data type in the state vector space were able to be analyzed. Due to unknown circumstances, relative humidity data were not collected by the MAWS system at the ARM main site (site 5) on a total of seven EM27/SUN deployment days: 5/31, 6/1, 6/14, 6/15, 7/7, 7/9, and 7/10. Therefore, these days were omitted from analysis and do not appear in Figures 4.2 and 4.3, which represent the entire TRACER dataset available for cluster analysis.



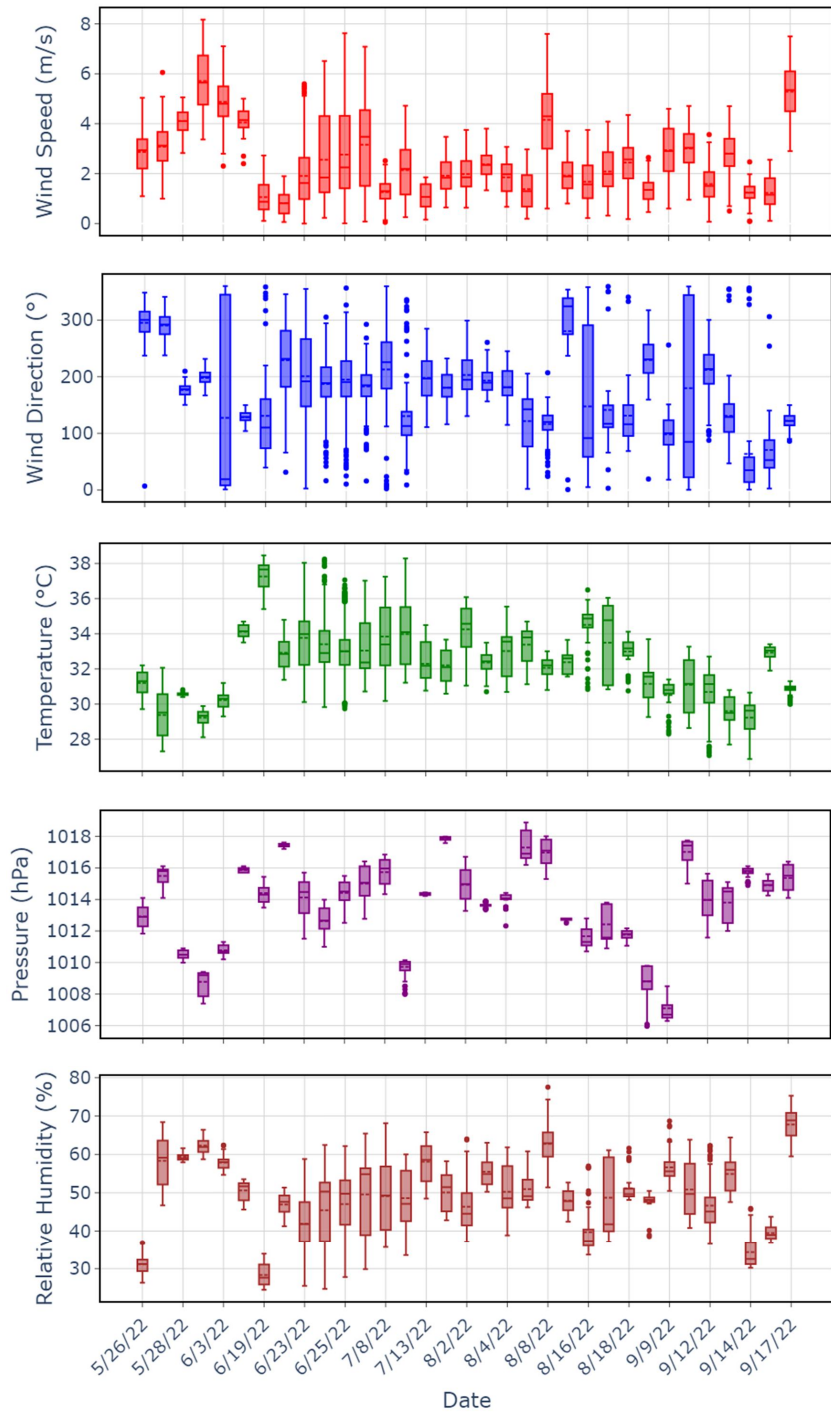


Figure 4.3: All meteorological data collected between 10:00 and 16:00 local time during the TRACER Campaign. Each box represents data collected on an individual day.

## 4.2 Sub-Seasonal Pollutant Analysis with UMAP and DBSCAN

Cluster analysis was chosen to provide a holistic view of the pollutant variations during the TRACER Campaign. Following Section 2.4.4, EM27/SUN and surface meteorological data from 10:00 to 16:00 local time were analyzed using UMAP and DBSCAN to emphasize groups of data with unique characteristics. Clusters marked by high and low daily mean concentrations were analyzed to hypothesize relationships between *Xgas*, pollutant sources in the Houston area, and meteorological processes.

### 4.2.1 The Process of Clustering: High Variability $XCO_2$ as a Representative Sample

Before standardizing the data for use with UMAP and DBSCAN, the daytime datasets were parsed into high and low variability groups. The distribution of standard deviations for each *Xgas* during the TRACER Campaign are provided in Figure 4.4. As described in Section 2.4, daytime datasets with *Xgas* standard deviations in the 75<sup>th</sup> percentile were considered high variability datasets for a particular *Xgas*, while all other daytime datasets were classified as low variability datasets.

Each of the standardized high and low variability groups for  $XCO_2$ ,  $XCO$ , and  $XCH_4$  ( $z_{XCO_2}$ ,  $z_{XCO}$ , and  $z_{XCH_4}$ , respectively) were analyzed individually with the clustering algorithm. Prior to clustering, each dataset first passed through UMAP to project the data onto a lower-dimensional space. The remainder of this analysis will discuss the data clustering process with a focus on the high variability  $XCO_2$  dataset.

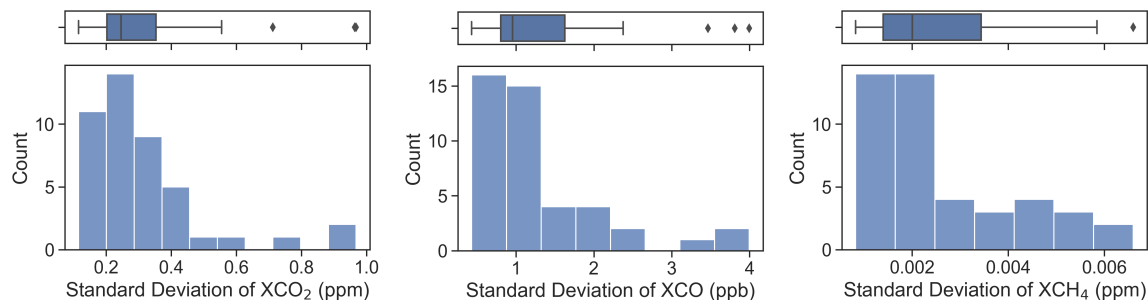


Figure 4.4: Histograms of one minute averaged  $XCO_2$ ,  $XCO$ , and  $XCH_4$  standard deviations where data have been limited to a local time window from 10:00 to 16:00

The three-dimensional UMAP projection of high variability  $XCO_2$  is presented in Figure 4.5. By coloring the UMAP projection by the feature of interest, the hidden structures of the data within the state vector space are revealed with respect to the feature of interest. As discussed in Section 2.4.3, repeated hyperparameter tuning of both UMAP and DBSCAN occurred before settling on this projection as the subjectively best representation of the high variability  $XCO_2$  dataset.

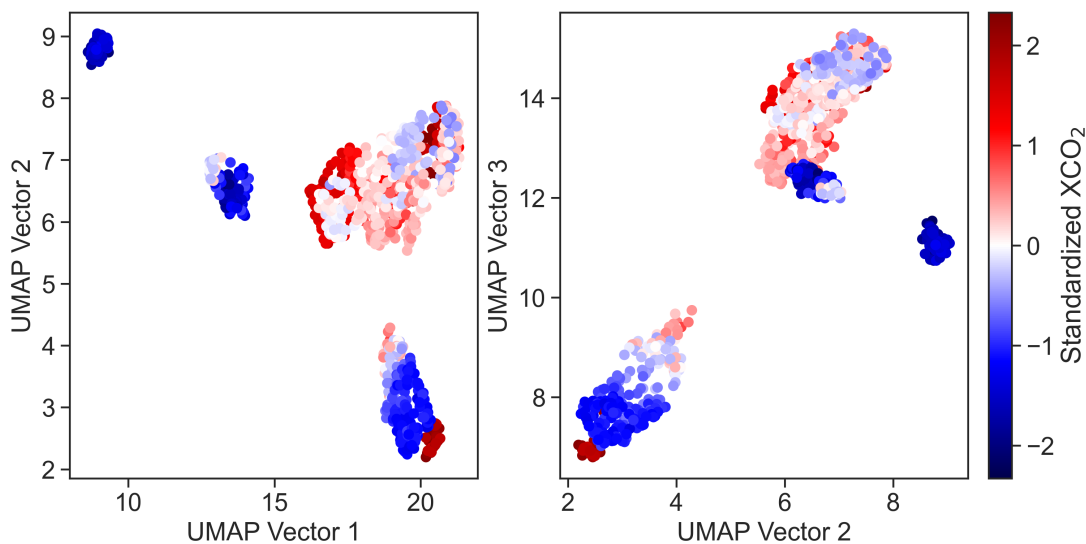


Figure 4.5: UMAP three-dimensional representation of the high variability  $XCO_2$  group, colored by the feature of interest, or standardized  $XCO_2$  ( $z_{XCO_2}$ ).

After dimensional reduction via UMAP, the lower-dimensional representation of the dataset was analyzed with DBSCAN to isolate clusters of related data for further analysis. Figure 4.6 illustrates the clusters chosen by the algorithm after hyperparameter tuning. Comparing these clusters to the UMAP vectors colored by standardized  $XCO_2$  ( $z_{XCO_2}$ ), it is evident that the DBSCAN representation does not contain all of the visually apparent clusters in Figure 4.5.

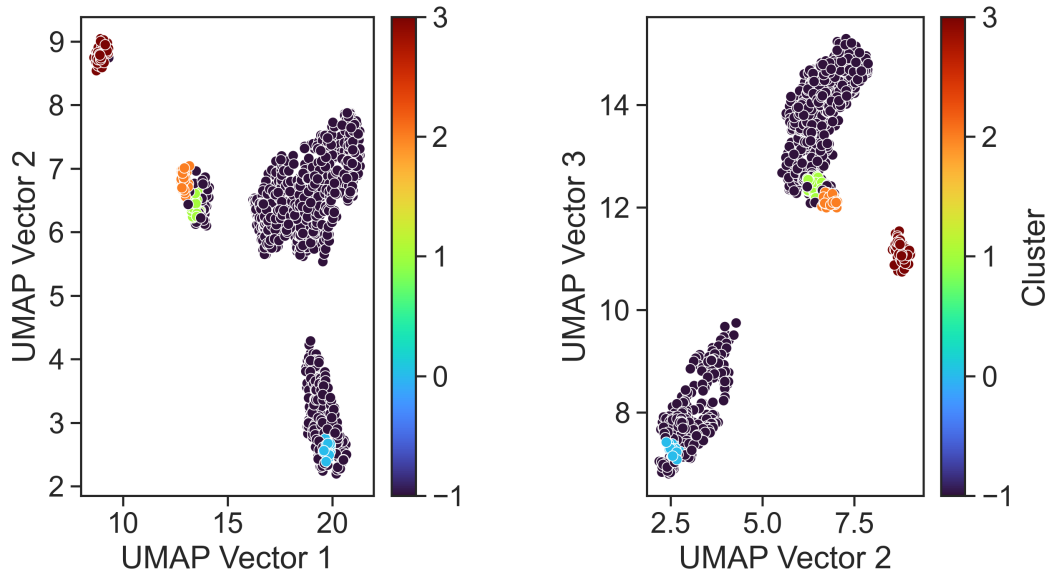


Figure 4.6: Optimized DBSCAN clustering results of the high variability  $XCO_2$  dataset. Each cluster is represented by a different color, where the -1 cluster contain data the algorithm considers noise.

To improve the clustering, the DBSCAN algorithm was run recursively with modified hyperparameters on the data considered to be noise by the algorithm during the first run. Figure 4.7 represents the clusters from both DBSCAN runs. This recursive approach to DBSCAN largely improves the clustering, successfully isolating data that appear unique using the UMAP projection as a guide. Each of the clusters that DBSCAN isolates represents a group of data with shared characteristics that

are unique compared to the rest of the dataset. Table 4.2 includes basic information about each of the high variability  $XCO_2$  clusters generated via recursive DBSCAN.

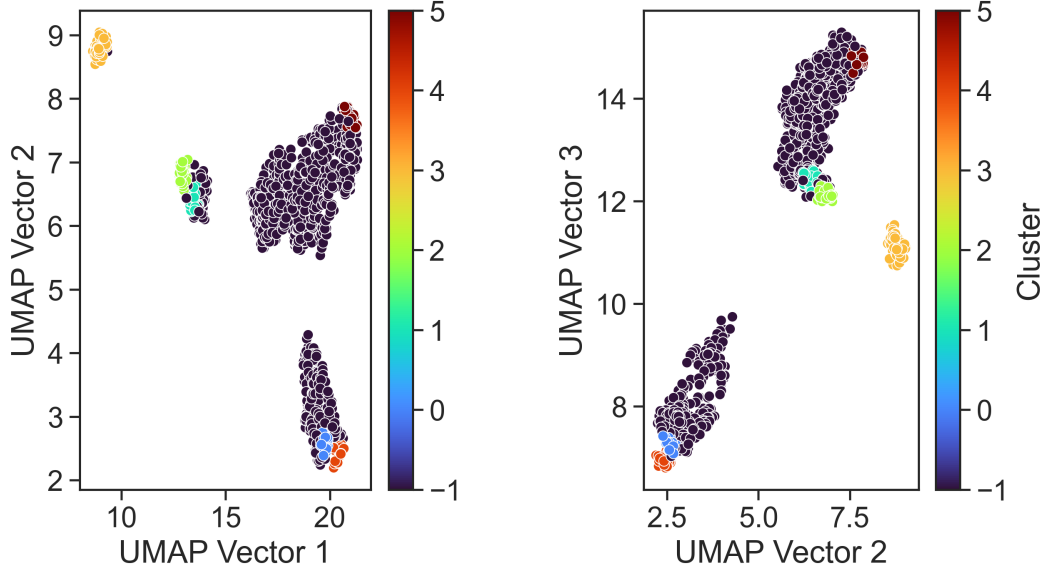


Figure 4.7: Recursive DBSCAN clustering results of the high variability  $XCO_2$  dataset

Table 4.2: High variability  $XCO_2$  clusters generated using recursive DBSCAN

Cluster	Mean $XCO_2$ (ppm)	Site Code	Date	Local Time
0	416.784	5	8/8/22	14:47 to 15:59
1	415.753	8	9/8/22	11:23 to 14:54
		5	9/9/22	10:22, 11:11
2	417.931	5	9/9/22	13:39 to 15:54
3	415.900	7	9/15/22	13:58 to 15:43
4	422.061	5	6/16/22	15:26 to 16:00
5	419.073	7	6/19/22	14:35 to 15:59

While each cluster assigned by DBSCAN contains information about related  $XCO_2$  observations, not every cluster represents data that are significantly different than the

mean  $XCO_2$  for the dataset. Figure 4.8 presents the  $XCO_2$  data contained in each cluster. To better understand the relationships between the state vector space and high or low values of  $XCO_2$ , the mean  $z_{XCO_2}$ , or  $\overline{z_{XCO_2}}$ , for each cluster was evaluated. Clusters containing  $\overline{z_{XCO_2}}$  greater than 1 (deemed high concentration clusters) or less than -1 (low concentration clusters) were kept for further analysis. In other words, only clusters with a mean  $XCO_2$  abiding by Equation 4.1 were kept,

$$\overline{XCO_2}_{cluster} = \overline{XCO_2}_{dataset} \pm s_{XCO_2} \quad (4.1)$$

where  $\overline{XCO_2}_{cluster}$  is the mean  $XCO_2$  of the cluster,  $\overline{XCO_2}_{dataset}$  represents the mean  $XCO_2$  of the high variability dataset, and  $s_{XCO_2}$  represents the standard deviation of  $XCO_2$  in the high variability dataset. To simplify further discussions, clusters that meet these criteria will be referred to as high and low concentration clusters.

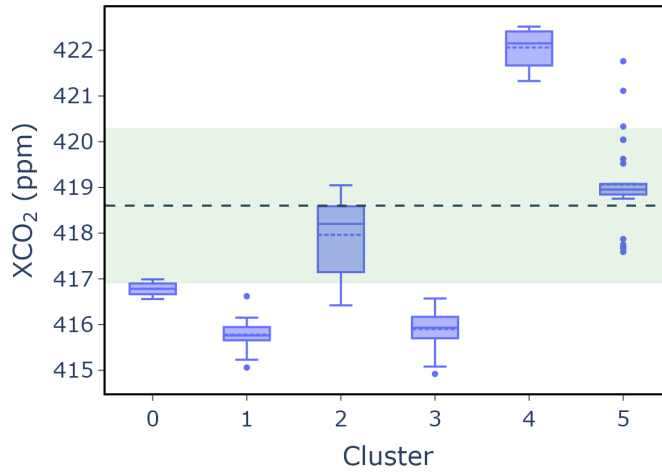


Figure 4.8: DBSCAN clusters of the high variability  $XCO_2$  dataset. The dashed line represents  $\overline{XCO_2}$  for the dataset and the shaded area represents one standard deviation above and below this mean.

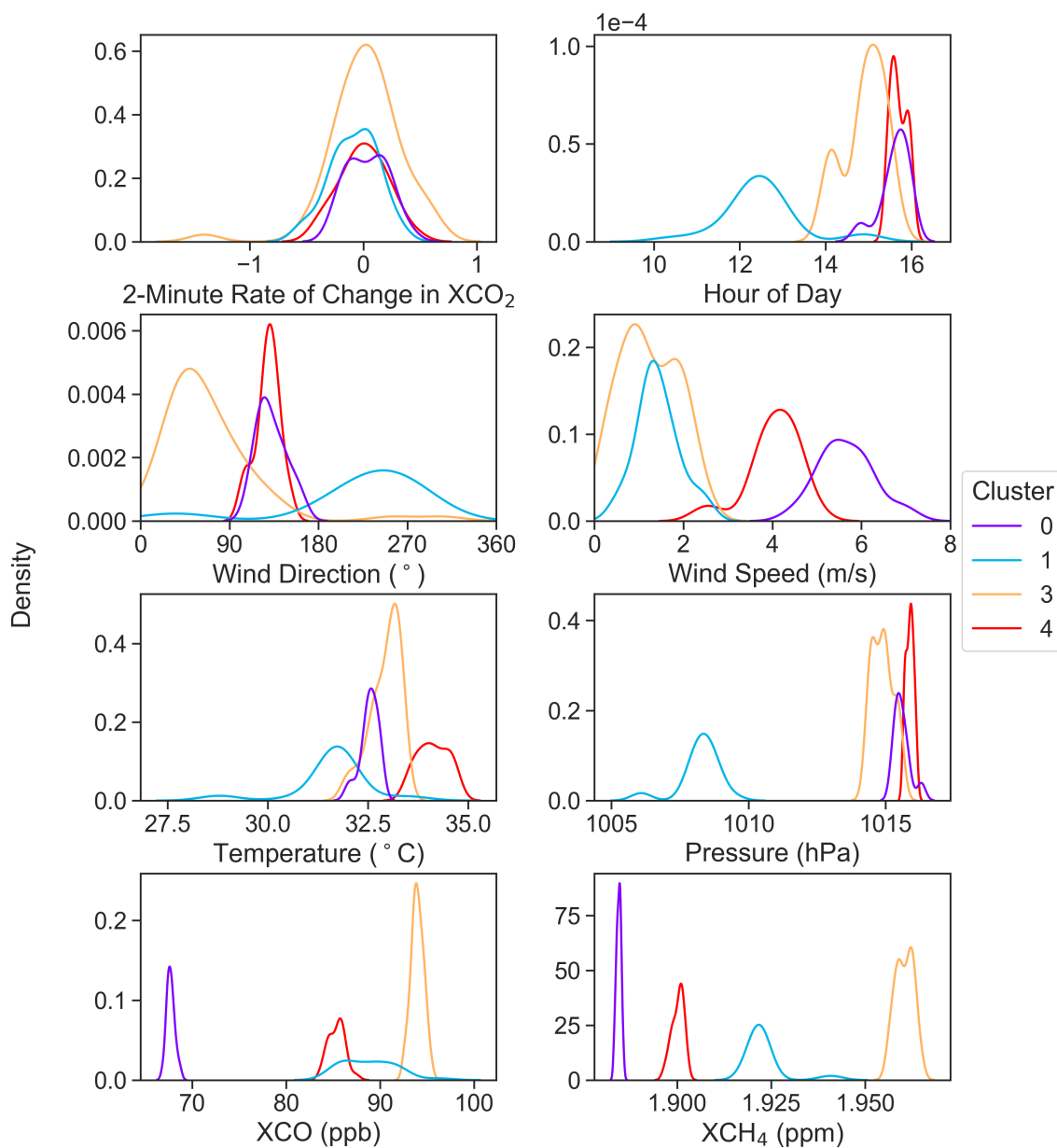


Figure 4.9: Density plots of various state vector space variables for the high variability  $XCO_2$  dataset. The only clusters shown are the high and low concentration clusters.

Of the six clusters generated by recursive runs of DBSCAN for the high variability  $XCO_2$  data, only four clusters represented  $XCO_2$  data at least one standard deviation above or below the mean of the dataset. Cluster 4 contained information about

higher concentrations, while clusters 0, 1, and 3 contained information about lower concentrations. Isolating these clusters provides a succinct view of the data so that inferences can be drawn about how the state vector space influenced the concentrations of  $XCO_2$  in Houston. Figure 4.9 highlights some of the relationships between  $XCO_2$  and the state vector space. These results, along with findings from other cluster analyses will be discussed in the following section.

### 4.2.2 $XCO_2$ Cluster Analysis Findings

The high variability  $XCO_2$  dataset contained four clusters of high and low concentrations, as defined in the previous section. The single high concentration cluster, cluster 4, contained data collected in the afternoon from the ARM main site on June 16. The outstanding feature highlighted by the density plot array in Figure 4.9 is the mean wind direction, representative of southeasterly winds. The Houston Ship Channel surrounds much of the site to the east with multiple industrial facilities to the north and south of the site as well. This suggests that this above average  $XCO_2$  cluster may have been influenced by local industrial sources. Figure 4.9 also includes density spikes for  $XCO$  and  $XCH_4$ , which simply indicates that  $XCO$  and  $XCH_4$  data were collected at the same time as the  $XCO_2$  data. However, correlation analysis of the state vector space and  $XCO_2$ , represented by the heatmap in Figure 4.10, shows that  $XCO_2$ ,  $XCO$ ,  $XH_2O$ , and  $XCH_4$  were all significantly correlated, with Pearson coefficients almost equal to one for each species. To further explore these relationships, these concentration data were plotted against each other to reveal the trends in the data. These results can be seen in the top row of Figure 4.11, where each  $Xgas$  is clearly positively correlated with  $XCO_2$  for cluster 4.



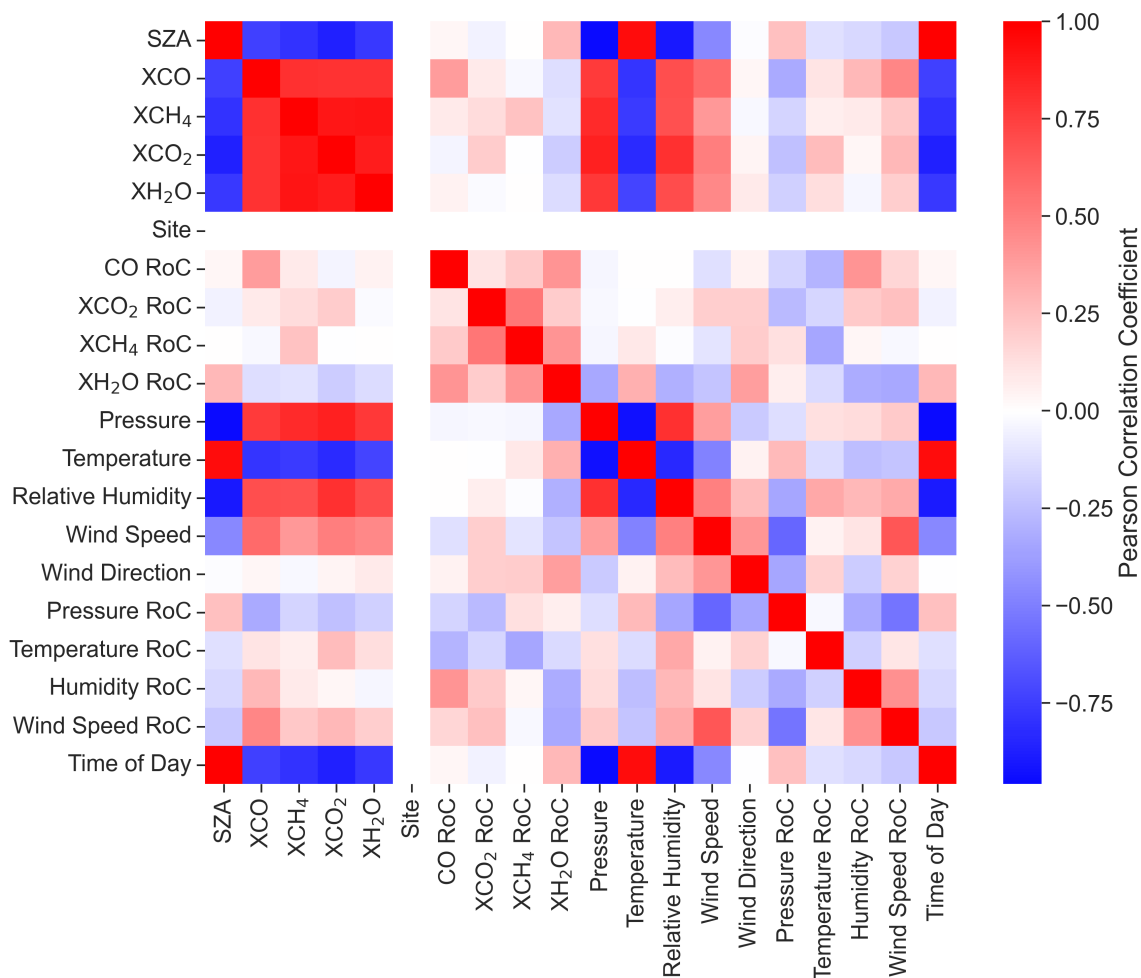


Figure 4.10: Correlation between all data in cluster 4 of the high variability  $XCO_2$  dataset

The three high variability, low concentration  $XCO_2$  clusters (clusters 0, 1, and 3) all included data from August and September. The data in cluster 0, like cluster 4, were collected on 8/8 at the ARM Main site (site 5). Cluster 1 included data from 9/8 at UH Aldine (site 8) and from 9/9 at the ARM Main site. Cluster 3 contained data from 9/15 at the UH Main site (site 7). From a cursory glance at the density curves associated with these clusters in Figure 4.9, these three clusters do not hold many similarities. However, cluster 0 and cluster 4 appear to be quite similar. Both of these

clusters contained data from the ARM Main site that were collected in the afternoon. Additionally, the mode of the wind direction for each cluster is representative of southeasterly winds. However, as stated before, cluster 4 contained high concentration  $XCO_2$  data, while cluster 0 contained low concentration  $XCO_2$ .

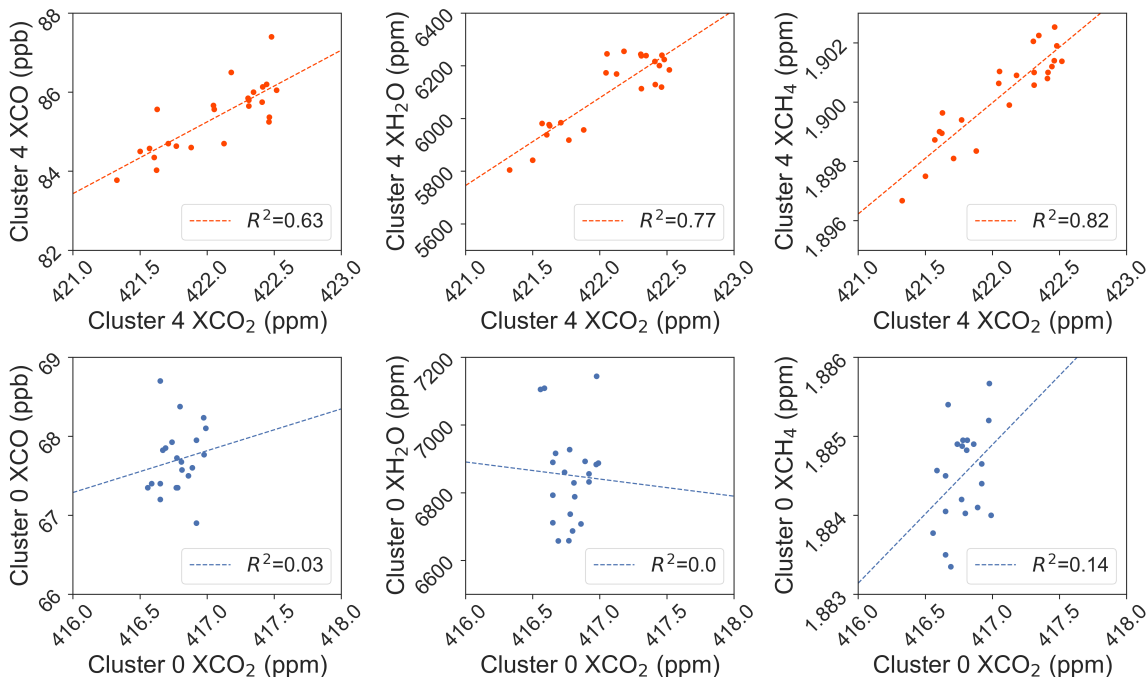


Figure 4.11:  $XCO_2$  plotted against  $XCO$  (left column),  $XH_2O$  (middle column), and  $XCH_4$  (right column). The plots in the top row are of data from the high variability, high concentration  $XCO_2$  cluster 4, while the bottom row plots were created using the high variability, low concentration  $XCO_2$  cluster 0.

Revisiting Figure 4.11, cluster 4 clearly shows positive, linear relationships between the  $Xgas$ , while cluster 0 shows no correlation. The strong correlation between all  $Xgas$  in cluster 4 suggests that these data were influenced by a local industrial signal. Correlation between  $XCO_2$ ,  $XCO$ , and  $XH_2O$  suggests that these are co-emitted combustion products, but the included correlation with  $XCH_4$  is unique. Since the ARM Main site is in close proximity to the Houston Ship Channel and other industrial

facilities, including petrochemical, chemical, and plastic production facilities, each of these *Xgas* could have easily been co-emitted as fugitive emissions by these local sources.

Since the data from clusters 0 and 4 were collected at the same deployment site, the same local sources would have existed, given some variation, during each data collection period. One possible reason for the difference in  $XCO_2$  between these clusters is an unexpected variation in the local industrial sources. As discussed in Section 2.2.2, flares at petrochemical and related facilities are difficult to control, as the composition of the gases sent to the flare can change frequently. Co-emitted  $CO_2$ ,  $CO$ ,  $H_2O$ , and  $CH_4$  as seen in cluster 4 may have occurred due to a unique flaring event, correctly isolated by the high variation data binning prior to cluster analysis, where the composition of gas sent to a flare was different than usual. Following this theory, heightened  $XCO_2$  supports that the industrial facility may have undergone a process disruption during this time, causing more gas to be sent to the flare.

To further explore the concentration magnitude difference between clusters 4 and 0, the near-surface wind resource present during each data collection period was analyzed. Figure 4.12 shows the relationship between  $XCO_2$  and wind direction for each cluster. Unfortunately, no EM27/SUN data existed before approximately 15:15 local time during the cluster 4 data collection period. The wind direction data from before this time were plotted to provide context about the meteorological conditions before the high concentrations exhibited at the beginning of this cluster. While not conclusive, the shift to approximately  $100^\circ$  and decrease in spread of wind direction seen in the cluster 4 data appear to coincide with the highest  $XCO_2$  concentrations for the cluster. As the wind direction shifts back to a range between  $110$ - $130^\circ$  and the variability increases again,  $XCO_2$  decreases. Meanwhile, no similar shifts in wind

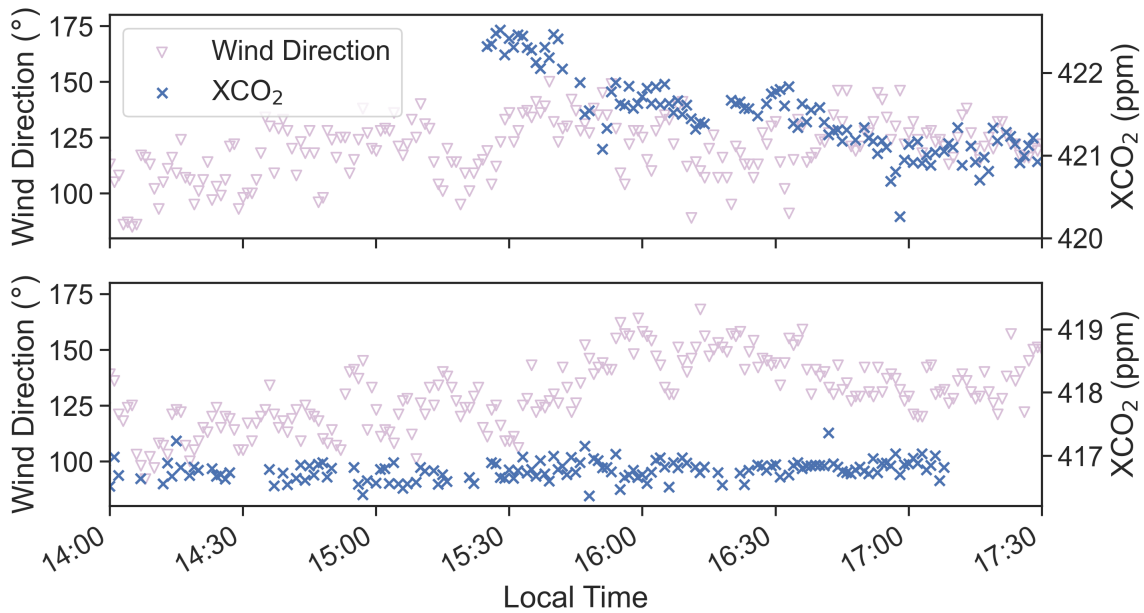


Figure 4.12:  $XCO_2$  and wind direction plotted from cluster 4 (top) and cluster 0 (bottom). The time period shown for cluster 4 plot begins before the EM27/SUN data collection began.

direction or  $XCO_2$  concentration can be seen in the cluster 0 data. Therefore, the shift in wind direction and decrease in wind direction variability shown in cluster 4 may have contributed to the high concentrations that define this cluster and separate it from cluster 0.

### 4.3 Satellite Validation

Using the collocation criteria from Zhou et al. (2022) described in Section 2.2.2.1, an EM27/SUN was only meaningfully collocated with either OCO-2 or OCO-3 a total of five times during the TRACER Campaign (Table 4.3). The distributions of  $XCO_2$  from each spectrometer in this comparison study, along with the number of data points that passed quality filtering, are represented in Figure 4.13.

Table 4.3: OCO-2 and OCO-3 collocations during the TRACER Campaign

Satellite	Date and Time (CDT)	Viewing Mode	EM27/SUN Site
OCO-2	7/7/22 14:37	Nadir	ARM Main
	8/8/22 14:37	Transition	ARM Main
	8/17/22 14:31	Glint	Aldine
OCO-3	7/5/22 9:05	SAM (Descending)	UHCC
	8/8/22 11:09	SAM (Ascending)	ARM Main

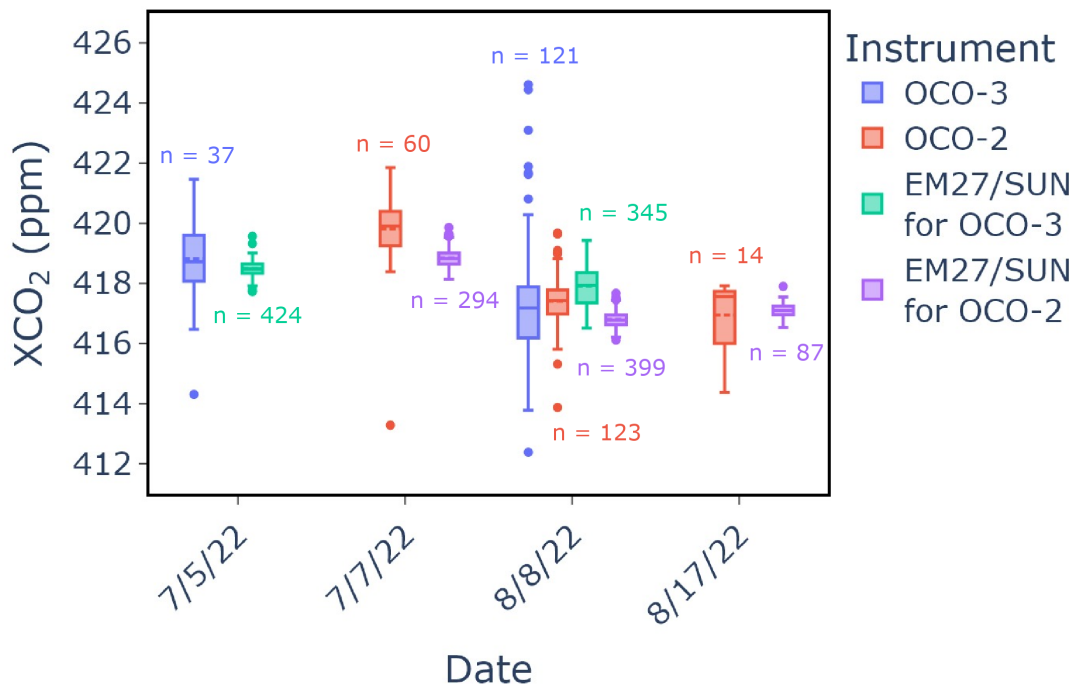


Figure 4.13: Box plots demonstrate the distribution of quality filtered  $XCO_2$  from OCO-2, OCO-3, and the collocated EM27/SUN instruments during the TRACER Campaign for each collocation period. EM27/SUN observations have been separated based on the satellite they were collocated with. The number of samples that met the collocation criteria and passed quality filtering are given as  $n$

Collocations were likely minimized due to sky conditions in the Houston area during the summer months. Due to the proximity to the coast, small non-convective cumulus clouds were present many days during the campaign, limiting both EM27/SUN and OCO-2/3 data.

Despite the limited data available for comparison, the ground- and space-based instruments still appeared to agree. The right panel of Figure 4.14 demonstrates the similarity of  $XCO_2$  reported by the EM27/SUN instruments and the satellites. The OCO data here have been quality filtered as previously discussed in Section 2.2.2.1. All EM27/SUN data used for this comparison are from the modified GGG dataset. The EM27/SUN data from the OCO-3 collocation on 7/5 were collected with KIT, and thus the KIT  $XCO_2$  correction factor from Table 3.1 was applied to reduce the bias compared to Blanche, the travel standard EM27/SUN. An averaging kernel correction, common for comparisons between different kinds of spectrometers, was not applied to the data, and would likely further improve the agreement between the instruments. The median biases between the OCO and modified GGG-retrieved EM27/SUN data are presented in Table 4.4

Table 4.4: Bias of OCO-2 and OCO-3 compared to EM27/SUN instruments for TRACER collocations

Satellite	Median Bias in $XCO_2$ From EM27/SUN (ppm)
OCO-2	0.5945
OCO-3	-0.7804

The choice of EM27/SUN retrieval algorithm was critical for evaluating the bias between these ground- and space-based measurements. The left panel of Figure 4.14 demonstrates the agreement between OCO-2/3 and the EM27/SUN instruments when

EM27/SUN data are processed with PROFFAST version 2.2. The large bias between collocated observations as seen with this comparison likely misrepresents the level of agreement between the instruments, considering that the agreement between Blanche and TCCON discussed previously was partially algorithm-dependent.

While the size of the comparison dataset does is not conducive to a thorough statistical analysis, this comparison suggests the CO<sub>2</sub> concentrations reported by OCO-2 and OCO-3 are representative of the concentrations present in the greater Houston area despite the presence of aerosols and cloud-contamination in the industrial, coastal environment. This comparison upholds the idea that EM27/SUN instruments can be used as transfer standards from the TCCON network. This is especially important for challenging urban environments like Houston that are far removed from TCCON stations, as increased confidence in space-based XCO<sub>2</sub> measurements fosters further scientific discovery with these datasets.

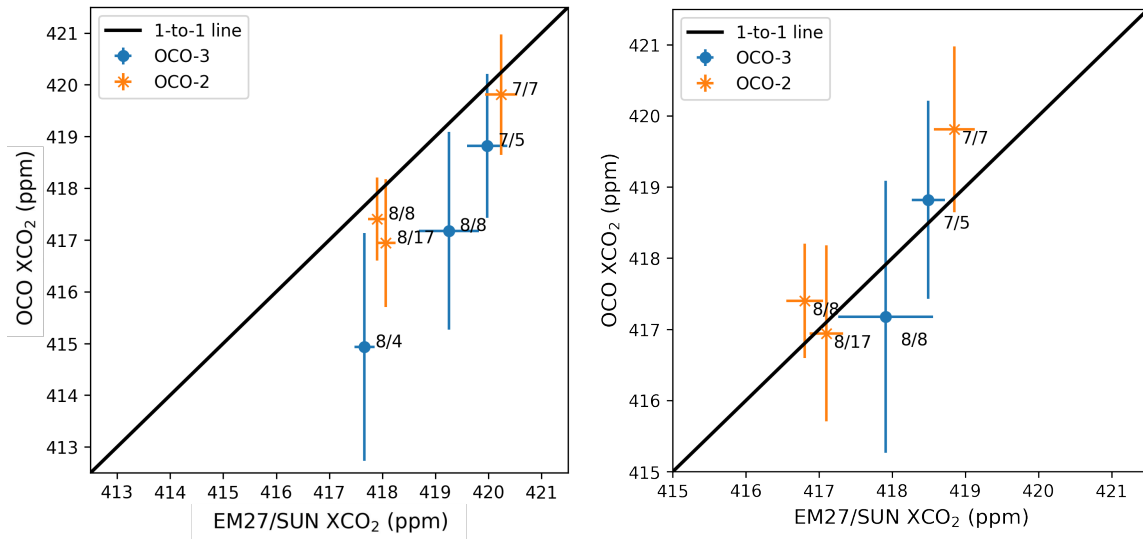


Figure 4.14: Comparison of EM27/SUN and OCO-2/3 observed XCO<sub>2</sub> using the preliminarily filtered PROFFAST version 2.2 dataset (left) and the quality-filtered modified GGG2020 EM27/SUN dataset (right). Note that the left plot contains data from 8/4/22, which was later removed from analysis due to low OCO-3 sample size.

This comparison also appears to agree with some of the findings from Zhou et al. (2022), the source of the collocation criteria used in this study, where OCO-2 measurements demonstrated minimal bias from EM27/SUN data collected in the highly populated city of Beijing, and the nearby suburb of Xianghe, China. However, the agreement between EM27/SUN and OCO-3 SAM observations in the Zhou et al. (2022) study demonstrated a notable bias. While SAM measurements appeared to clearly capture gradients of  $XCO_2$  between the cities, these measurements had a bias between 0.9 and 1.4 ppm from the EM27/SUN collocated measurements in both locations. Zhou et al. (2022) concludes more data are necessary to confirm that this bias is uniformly experienced in these cities, just as more data are necessary to conclusively determine the biases between instruments in this study. Future data collection should feature extended EM27/SUN deployments during different times of the year to more fully describe the differences between the ground- and space-based sensors. More details for future work are proposed in Section 5.2.



## Chapter 5

### Conclusions

EM27/SUN observations of  $XCO_2$ ,  $XCO$ , and  $XCH_4$  from two field campaigns were analyzed to evaluate biases between carbon cycle observing instruments and to investigate localized influences on observations in urban and industrial environments. Each of the trace gases of interest are important climate forcers, which require accurate monitoring to understand the state of the climate and enforce emission regulations. Monitoring is especially useful for urban and industrial environments where anthropogenic emission sources are concentrated and certain pollutants (e.g., CO) can impact human health.

#### 5.1 Summary of Work and Findings

Prior to trace gas analysis, a stringent automated data quality filtering regime was developed to omit artificially heightened concentrations. Quality filters were created with loss of solar tracking episodes in mind, which can be caused by significant cloud interference and issues with instrumentation. These episodes can produce wildly erroneous data that do not trigger quality flags within the retrieval algorithms, presenting the data user with artificially heightened concentration data that could be interpreted as legitimate signal enhancements. The quality filters developed removed heightened concentrations from these loss of tracking episodes while retaining highly

variable data that did not appear correlated with instrumental issues, and likely related to changes in environmental conditions.

To ensure measurement accuracy, three EM27/SUN instruments were collocated with a higher resolution, WMO-standardized TCCON instrument. EM27/SUN data from this three-day collocation period were processed using two retrieval algorithms, PROFFAST Version 2.2 and GGG2020. Algorithm performance was evaluated by investigating the SZA-dependence of the data, agreement between the standard EM27/SUN instrument (Blanche) and TCCON, as well as the agreement between the EM27/SUN instruments. A modified version of GGG2020, which included empirical *ad hoc* airmass-dependent and airmass-independent correction factors developed by Mostafavi Pak et al. (2023) performed the best during the evaluation, limiting SZA-dependence as designed and definitively improving the agreement between all instruments for  $XCO_2$ .

During the TRACER Campaign, the three EM27/SUN instruments — Blanche, KIT, and JPL — were deployed in Houston at urban and background sites from late May through mid-September of 2022. The underlying trend of  $XCO_2$  throughout the campaign suggested that the greater Houston area followed the Northern Hemisphere seasonal cycle of  $XCO_2$ , driven by background concentrations. Variability in  $X_{gas}$  on smaller time scales were hypothesized to be heavily influenced by local pollutant sources and meteorological conditions.

Cluster analysis utilizing UMAP and DBSCAN algorithms was explored to test the local influence hypothesis. By focusing the analysis on days with high variability  $XCO_2$  data (where daily standard deviation of  $XCO_2$  was in the 75<sup>th</sup> percentile of all daily  $XCO_2$  standard deviations), unique patterns in the data that occurred rarely during the campaign were highlighted. Two clusters in particular stood out from this

subset of data due to similarities in deployment site, time of day, and mean wind direction. However, one cluster was characterized by high  $\overline{XCO_2}$  while the other, low  $\overline{XCO_2}$ . Further investigation suggested that the high concentration cluster contained data from a probable period of pollutant co-emission from local industrial sources, as  $XCO_2$ ,  $XCO$ ,  $XH_2O$ , and  $XCH_4$  were all strongly correlated. A sharp shift in wind direction appeared to align temporally with the highest  $XCO_2$  concentrations in the cluster. Additionally, understanding of flare operation suggests that a rare flaring event may have occurred to not only co-emit pollutants, but co-emit high concentrations of pollutants. These theories further suggest that the EM27/SUN was able to briefly observe co-emitted fugitive emissions from a nearby industrial source.

Finally, OCO-2 and OCO-3 data were compared to EM27/SUN data collected during TRACER. These satellites, and many others like them, provide frequent global views of the carbon cycle and climate-forcing trace gases. While the satellites are validated at TCCON sites regularly, the validation is spatially limited. During the TRACER Campaign, the EM27/SUN instruments served as travel standards for TCCON, extending satellite validation to the challenging urban and coastal environment of Houston. This satellite validation effort, though limited in data, supported the consensus that the data these satellites collect accurately represent the pollutants in the atmosphere, shown with minimal bias between space- and ground-based instruments. The satellite comparison also highlighted the importance of retrieval algorithm choice when analyzing spectrometer data. When processed with PROFFAST version 2.2, the EM27/SUN data appeared to be biased approximately 2 ppm higher than the OCO instruments. Processing the data with GGG2020, however, revealed a relatively small bias between the instruments. Since this analysis, PROFFAST version 2.3 has been released with modified parameters to minimize

SZA-dependence and correct instruments to TCCON standards. While these changes are expected to improve the comparison between EM27/SUN instruments and other spectrometers, this analysis highlights the importance of understanding how a chosen retrieval algorithm processes data before using it.

## 5.2 Future Work

Continued EM27/SUN analysis should first consider alternative approaches to quality control filtering. While the tuned percent change method for bad quality data omission appeared to work well for the data analyzed in this thesis, a statistically- or empirically-based method may exist that performs similarly to the filter used here, but ensures a more universal method. Options to explore include filtering data using the standard deviation of *Xgas* signals and using data generated by CamTracker (Gisi et al. 2011) during deployments to develop criteria for loss of tracking episodes. A time-window approach, as opposed to point-window approach, would also be useful for improving filter applicability.

More EM27/SUN data are necessary to strengthen the conclusions from initial TRACER analysis. The satellite validation effort would benefit from this especially, as there were only five collocation periods during the TRACER Campaign. EM27/SUN data from non-summer months would be especially helpful to avoid persistent non-convective cumulus clouds. Additionally, future satellite validation efforts should include averaging kernel corrections while processing EM27/SUN data to ensure accurate bias estimates.

Including different data types in future analysis would further contextualize the EM27/SUN data collected during the TRACER Campaign. More meteorological data

(i.e., vertical profiles of wind direction and speed, boundary layer height, sea/bay/land breeze passages) would provide invaluable information about the state of the atmosphere where the *Xgas* data were collected and where local contributions to *Xgas* may have originated. Joint analysis of trace gas and aerosol data could further link *Xgas* enhancements to local pollutant sources, as  $XCO_2$  is known to be co-emitted with aerosols by industrial processes (e.g., Vogt et al. (2011); Majeed and Svendsen (2018)).

In addition to the data sources mentioned above, future cluster analysis approaches should consider the inclusion of land use data. These datasets already exist for Houston, allowing for easy implementation. If included, these data would provide the models with information that could be used to infer local emission sources.

Though the cluster analysis was able to draw attention to interesting collections of data throughout the TRACER Campaign, better methods may exist for this analysis. While DBSCAN is a useful algorithm for complex datasets, high-density non-homogeneous point clouds, like those produced by UMAP in this study, can be handled poorly. This is one reason why the majority of data per cluster analysis was considered noise by the DBSCAN algorithm. Using an algorithm that better understands how to handle these point clouds could massively improve clustering of this data, allowing for more thorough analysis of high and low concentration *Xgas* data during the campaign. There have been multiple attempts to build off of the DBSCAN algorithm to address these issues. These methods (e.g., Nagaraju et al. (2016); Cai et al. (2020)) should be explored in future work.

## References

- Alberti, C., and Coauthors, 2022: Improved calibration procedures for the EM27/SUN spectrometers of the COllaborative Carbon Column Observing Network (COCCON). **15 (8)**, 2433–2463, <https://doi.org/10.5194/amt-15-2433-2022>, publisher: Copernicus GmbH.
- Arya, S. P., 1999: *Air Pollution Meteorology and Dispersion*. Oxford University Press.
- Cai, Z., J. Wang, and K. He, 2020: Adaptive Density-Based Spatial Clustering for Massive Data Analysis. **8**, 23 346–23 358, <https://doi.org/10.1109/ACCESS.2020.2969440>, conference Name: IEEE Access.
- Caicedo, V., and Coauthors, 2019: Bay breeze and sea breeze circulation impacts on the planetary boundary layer and air quality from an observed and modeled discover-aq texas case study. *Journal of Geophysical Research: Atmospheres*, **124 (13)**, 7359–7378, <https://doi.org/https://doi.org/10.1029/2019JD030523>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD030523>.
- Daniel, J. S., and S. Solomon, 1998: On the climate forcing of carbon monoxide. **103**, 13 249–13 260, <https://doi.org/10.1029/98JD00822>.
- Duren, R. M., and Coauthors, 2019: California’s methane super-emitters. **575 (7781)**, 180–184, <https://doi.org/10.1038/s41586-019-1720-3>.
- Eldering, A., T. E. Taylor, C. W. O’Dell, and R. Pavlick, 2019: The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data. **12 (4)**, 2341–2370, <https://doi.org/10.5194/amt-12-2341-2019>, publisher: Copernicus GmbH.
- Eldering, A., and Coauthors, 2017: The Orbiting Carbon Observatory-2: first 18 months of science data products. **10 (2)**, 549–563, <https://doi.org/10.5194/amt-10-549-2017>, publisher: Copernicus GmbH.
- Ester, M., H.-P. Kriegel, J. Sander, and X. Xu, 1996: A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, AAAI Press, 226–231, KDD’96, <https://doi.org/10.5555/3001460.3001507>.
- Feld, L., D. Dubravica, and B. Herkommer, 2023: PROFFASTpylot. URL <https://gitlab.eudat.eu/coccon-kit/proffastpylot>.

- Frey, M., and Coauthors, 2019: Building the COllaborative Carbon Column Observing Network (COCCON): long-term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer. **12 (3)**, 1513–1530, <https://doi.org/10.5194/amt-12-1513-2019>.
- Gisi, M., F. Hase, S. Dohe, and T. Blumenstock, 2011: Camtracker: a new camera controlled high precision solar tracker system for FTIR-spectrometers. **4 (1)**, 47–54, <https://doi.org/10.5194/amt-4-47-2011>, publisher: Copernicus GmbH.
- Gisi, M., F. Hase, S. Dohe, T. Blumenstock, A. Simon, and A. Keens, 2012: XCO<sub>2</sub>-measurements with a tabletop FTS using solar absorption spectroscopy. **5 (11)**, 2969–2980, <https://doi.org/10.5194/amt-5-2969-2012>, publisher: Copernicus GmbH.
- Hase, F., M. Frey, M. Kiel, T. Blumenstock, R. Harig, A. Keens, and J. Orphal, 2016: Addition of a channel for XCO observations to a portable FTIR spectrometer for greenhouse gas measurements. **9 (5)**, 2303–2313, <https://doi.org/10.5194/amt-9-2303-2016>, publisher: Copernicus GmbH.
- Hu, C., T. J. Griffis, X. Lee, D. B. Millet, Z. Chen, J. M. Baker, and K. Xiao, 2018: Top-Down Constraints on Anthropogenic CO<sub>2</sub> Emissions Within an Agricultural-Urban Landscape. **123 (9)**, 4674–4694, <https://doi.org/10.1029/2017JD027881>.
- Iowa Environmental Mesonet, 2022: Iowa Environmental Mesonet: Wind Roses. URL [https://mesonet.agron.iastate.edu/sites/windrose.phtml?station=MCJ&network=TX\\_ASOS](https://mesonet.agron.iastate.edu/sites/windrose.phtml?station=MCJ&network=TX_ASOS).
- IPCC, 2019: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories — IPCC. URL <https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/>.
- 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*. Cambridge University Press, <https://doi.org/10.1017/9781009157896>.
- Jacobs, N., and Coauthors, 2020: Quality controls, bias, and seasonality of CO<sub>2</sub> columns in the boreal forest with Orbiting Carbon Observatory-2, Total Carbon Column Observing Network, and EM27/SUN measurements. **13 (9)**, 5033–5063, <https://doi.org/10.5194/amt-13-5033-2020>.
- James, G., D. Witten, T. Hastie, and R. Tibshirani, 2021: *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics, Springer US, <https://doi.org/10.1007/978-1-0716-1418-1>.

- Jensen, M. P., 2019: BNL-212068-2019-INRE, 1561242. Tracking Aerosol Convection Interactions Experiment (TRACER) Science Plan. BNL-212068-2019-INRE, 1561242 pp., <https://doi.org/10.2172/1561242>.
- Kaufman, S., S. Rosset, C. Perlich, and O. Stitelman, 2012: Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data*, **6** (4), <https://doi.org/10.1145/2382577.2382579>.
- Keeler, E., and J. Kyrouac, n.d.: Automatic Weather Station (MAWS). <https://doi.org/10.5439/1162061>.
- Laughner, J. L., and Coauthors, 2022: A new algorithm to generate a priori trace gas profiles for the GGG2020 retrieval algorithm. <https://doi.org/10.5194/amt-2022-267>.
- Lucchesi, R., 2015: File Specification for GEOS-5 FP-IT. *GMAO Office Note No. 2 (Version 1.3)*, 60, URL [http://gmao.gsfc.nasa.gov/pubs/office\\_notes](http://gmao.gsfc.nasa.gov/pubs/office_notes).
- Majeed, H., and H. F. Svendsen, 2018: Characterization of aerosol emissions from CO<sub>2</sub> capture plants treating various power plant and industrial flue gases. **74**, 282–295, <https://doi.org/10.1016/j.ijggc.2018.04.016>.
- May, R. M., and Coauthors, 2022: MetPy: A Meteorological Python Library for Data Analysis and Visualization. *Bulletin of the American Meteorological Society*, **103** (10), E2273 – E2284, <https://doi.org/10.1175/BAMS-D-21-0125.1>.
- McInnes, L., J. Healy, N. Saul, and L. Großberger, 2018: UMAP: Uniform Manifold Approximation and Projection. *JOSS*, **3** (29), 861, <https://doi.org/10.21105/joss.00861>.
- Medlock III, K. B., 2021: The future of houston as energy transitions. *Baker Institute Report*, **5**.
- Miller, D. J., and Coauthors, 2020: Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas. **54** (4), 2133–2142, <https://doi.org/10.1021/acs.est.9b05523>.
- Mostafavi Pak, N., and Coauthors, 2023: Using portable low-resolution spectrometers to evaluate Total Carbon Column Observing Network (TCCON) biases in North America. *Atmos. Meas. Tech.*, **16** (5), 1239–1261, <https://doi.org/10.5194/amt-16-1239-2023>.
- Nagaraju, S., M. Kashyap, and M. Bhattacharya, 2016: An Effective Density Based Approach to Detect Complex Data Clusters Using Notion of Neighborhood Difference. **14** (1), 57–67, <https://doi.org/10.1007/s11633-016-1038-7>, publisher: Machine Intelligence Research.



- NCAR, n.d.: Wind Direction Quick Reference. URL <https://www.eol.ucar.edu/content/wind-direction-quick-reference>.
- OCO-2/OCO-3 Science Team, V. P., Abhishek Chatterjee, 2020: OCO-2 Level 2 bias-corrected XCO<sub>2</sub> and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing V11r. Goddard Earth Sciences Data and Information Services Center (GES DISC), <https://doi.org/10.5067/5Q8JLZL1VD4A>.
- OCO-2/OCO-3 Science Team, V. P., Abhishek Chatterjee, 2022: OCO-3 Level 2 bias-corrected XCO<sub>2</sub> and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing v10.4r. Goddard Earth Sciences Data and Information Services Center (GES DISC), <https://doi.org/10.5067/970BCC4DHH24>.
- Pan, G., Y. Xu, and J. Ma, 2021: The potential of CO<sub>2</sub> satellite monitoring for climate governance: A review. **277**, 111–423, <https://doi.org/10.1016/j.jenvman.2020.111423>.
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.
- Peterson, J., N. Tuttle, H. Cooper, and C. Baukal, 2007: Minimize facility flaring. *Hydrocarbon Processing*, **86**, 111–115.
- Phillips, A., 2019: Plastics Pollution on the Rise. URL <https://environmentalintegrity.org/reports/plastics-pollution-on-the-rise/>, 45 pp.
- Plant, G., and Coauthors, 2022: Inefficient and unlit natural gas flares both emit large quantities of methane. **377 (6614)**, 1566–1571, <https://doi.org/10.1126/science.abq0385>, publisher: American Association for the Advancement of Science.
- Ramanathan, V., and Coauthors, 2019: *Bending the Curve: Climate Change Solutions*.
- Riedel, K., and K. Lassey, 2008: "detergent of the atmosphere". **16 (1)**, 22–23, URL <https://niwa.co.nz/publications/water-and-atmosphere/vol16-no1-march-2008/detergent-of-the-atmosphere>.
- Saygin, D., and D. Gielen, 2021: Zero-Emission Pathway for the Global Chemical and Petrochemical Sector. **14 (13)**, 3772, <https://doi.org/10.3390/en14133772>, number: 13 Publisher: Multidisciplinary Digital Publishing Institute.
- Segales, A. R., B. R. Greene, T. M. Bell, W. Doyle, J. J. Martin, E. A. Pillar-Little, and P. B. Chilson, 2020: The CopterSonde: an insight into the development of a smart unmanned aircraft system for atmospheric boundary layer research. **13 (5)**, 2833–2848, <https://doi.org/10.5194/amt-13-2833-2020>.

- Seinfeld, J. H., and S. N. Pandis, 2016: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. 3rd ed., John Wiley and Sons, Inc.
- Sha, M. K., and Coauthors, 2020: Intercomparison of low- and high-resolution infrared spectrometers for ground-based solar remote sensing measurements of total column concentrations of CO<sub>2</sub>, CH<sub>4</sub>, and CO. **13 (9)**, 4791–4839, <https://doi.org/10.5194/amt-13-4791-2020>, publisher: Copernicus GmbH.
- Shiga, Y. P., A. M. Michalak, S. M. Gourdj, K. L. Mueller, and V. Yadav, 2014: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO<sub>2</sub> measurements. **41 (12)**, 4381–4388, <https://doi.org/10.1002/2014GL059684>.
- US EPA, 2022a: Basic Information about Carbon Monoxide (CO) Outdoor Air Pollution. URL <https://www.epa.gov/co-pollution/basic-information-about-carbon-monoxide-co-outdoor-air-pollution>.
- US EPA, 2022b: Importance of Methane. URL <https://www.epa.gov/gmi/importance-methane>.
- US EPA, 2023a: NAAQS Table. URL <https://www.epa.gov/criteria-air-pollutants/naaqs-table>.
- US EPA, 2023b: Overview of Greenhouse Gases. URL <https://www.epa.gov/ghgemissions/overview-greenhouse-gases>.
- Vogt, M., E. D. Nilsson, L. Ahlm, E. M. Mårtensson, and C. Johansson, 2011: The relationship between 0.25–2.5 μm aerosol and CO<sub>2</sub> emissions over a city. **11 (10)**, 4851–4859, <https://doi.org/10.5194/acp-11-4851-2011>, publisher: Copernicus GmbH.
- Voiland, A., 2015: Fourteen years of carbon monoxide from MOPITT. URL <https://climate.nasa.gov/news/2291/fourteen-years-of-carbon-monoxide-from-mopitt>.
- Wagner, T. J., P. M. Klein, and D. D. Turner, 2019: A new generation of ground-based mobile platforms for active and passive profiling of the boundary layer. **100 (1)**, 137–153, <https://doi.org/10.1175/BAMS-D-17-0165.1>, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.
- Wallace, H. W., N. P. Sanchez, J. H. Flynn, M. H. Erickson, B. L. Lefer, and R. J. Griffin, 2018: Source apportionment of particulate matter and trace gases near a major refinery near the houston ship channel. **173**, 16–29, <https://doi.org/10.1016/j.atmosenv.2017.10.049>.

- Weintrit, A., and P. Kopacz, 2011: A novel approach to loxodrome (rhumb line), orthodrome (great circle) and geodesic line in ECDIS and navigation in general. *Methods and Algorithms in Navigation, Marine Navigation and Safety of Sea Transportation*, 123–132.
- Wennberg, P. O., and Coauthors, 2022: TCCON data from Lamont (US), Release GGG2020.R0 (Version R0). CaltechDATA, <https://doi.org/10.14291/tccon.ggg2020.lamont01.R0>.
- Williams, E., n.d.: Aviation Formulary V1.47. URL <http://edwilliams.org/avform147.htm#GCF>.
- Wunch, D., and Coauthors, 2011: The Total Carbon Column Observing Network. **369 (1943)**, 2087–2112, <https://doi.org/10.1098/rsta.2010.0240>.
- Wunch, D., and Coauthors, 2017: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO<sub>2</sub> measurements with TCCON. **10 (6)**, 2209–2238, <https://doi.org/10.5194/amt-10-2209-2017>.
- Zavala-Araiza, D., D. Lyon, R. A. Alvarez, V. Palacios, R. Harriss, X. Lan, R. Talbot, and S. P. Hamburg, 2015: Toward a functional definition of methane super-emitters: Application to natural gas production sites. **49 (13)**, 8167–8174, <https://doi.org/10.1021/acs.est.5b00133>, publisher: American Chemical Society.
- Zhou, M., and Coauthors, 2022: CO<sub>2</sub> in Beijing and Xianghe Observed by Ground-Based FTIR Column Measurements and Validation to OCO-2/3 Satellite Observations. **14 (15)**, 3769, <https://doi.org/10.3390/rs14153769>.

## A Lessons Learned from the GeoCarb-TRACER Campaign

During the GeoCarb-TRACER Campaign, there were multiple unexpected complications, but few issues that could not be resolved by the EM27/SUN group and others involved in ARM TRACER. There was one near miss near the middle of the campaign where volunteers did not monitor the deployed EM27/SUN instruments when there was a slight chance or higher of rain. Consequently, the two EM27/SUNs in the field were lightly rained on. Neither instrument was harmed by this incident, as concluded by a thorough investigation and conversation with the instrument owners. However, the incident dramatically changed the communication and deployment approach of the campaign. These were the most notable changes:

1. The wording of the weather conditions disclaimer in the operating procedure (highlighted in yellow in the first step of the procedure) was clarified to point volunteers directly to the National Weather Service's forecast for rain potentials.
2. All future deployments were limited to one EM27/SUN in an air conditioned enclosure to limit travel between sites and weather impacts. The enclosure eventually used for deployments was not completely built at the time of the incident, so all GeoCarb-TRACER deployments were canceled until the enclosure was ready for the field.
3. Formal training sessions were held with all future volunteers to clearly introduce all human- and instrument-related safety considerations needed while on site. While informal training sessions were held prior to the incident, these formal sessions ensured information was uniformly shared between volunteers, who met with me in a group setting to facilitate conversation and address any questions that arose such that everyone was aware of the appropriate answers.
4. For all future deployments where I was not on site, I would have morning check-ins every day during the deployment discussing weather conditions and protocol for the day.
5. Finally, the two EM27/SUN instruments that were rained on were not deployed again until they could be thoroughly investigated to ensure that no harm was done to the instruments during the incident. Both instrument owners were contacted and given an exhaustive report of the scenario, the investigation after, and the conclusion by the PI of the project and myself that the instruments were not harmed during the near miss.

After the end of the GeoCarb-TRACER Campaign, all volunteers were asked to share thoughts about the campaign on a survey. All volunteers were asked the following questions, which they could answer anonymously if they desired:

- During your deployment, what went well? What did not go so well?
- What would have made your deployment experience better?
- Do you have any recommendations for future EM27/SUN-based field campaigns?
- Do you have any other comments?

All responses were summarized, made anonymous, and shared with all volunteers in a meeting to generate more conversation and solutions for issues identified during the campaign. The following table summarizes the lessons learned from the survey and discussion.

Table A.1: Summarized feedback from the lessons learned session for the GeoCarb-TRACER Campaign

General Topic	Comments	Recommendations
EM27/SUN Transportation	<ul style="list-style-type: none"> <li>• Securing items in the cargo van for travel between deployments was not trivial.</li> <li>• Equipment could bounce around when driving over rough roads if not secured properly. This is concerning for the EM27/SUN instruments.</li> </ul>	<ul style="list-style-type: none"> <li>• Minimize the amount of travel that the EM27/SUN instruments do.</li> <li>• Cover proper tie-down procedures during volunteer training.</li> <li>• Consider investing in shock-mount pallets or corner mounted vibration control systems for in-case EM27/SUN stabilization.</li> </ul>

- 
- |                               |  |   |
|-------------------------------|--|---|
| Documentation of Daily Events | <ul style="list-style-type: none"><li>• Daily events were not recorded during the entirety of the campaign.</li><li>• The Google Docs lab notebook used during later IOPs was fine for documentation.</li><li>• Different volunteers documented observations differently and provided different levels of details.</li></ul> | <ul style="list-style-type: none"><li>• Finalize documentation method prior to the start of the campaign.</li><li>• Create a lab notebook template for volunteers to fill out to ensure that the right information is captured.</li><li>• If possible, automate collection of values not already recorded.</li><li>• Consider using notion.so instead of Google Docs to help organize everything.</li></ul> |
|-------------------------------|--|---|

- 
- |                      |  |   |
|----------------------|--|---|
| Instrument Operation | <ul style="list-style-type: none"><li>• Deployments that aimed to deploy instruments at multiple sites were challenging for a single pair of field scientists/volunteers.</li><li>• The number of active deployment sites should match the number of operator teams present.</li><li>• The buddy system is a safe and useful system, but sometimes it seems like deployments would be more efficient without it.</li></ul> | <ul style="list-style-type: none"><li>• Make sure someone is on-site when an instrument is running, or at least within a short distance away from the location if remote control is not established and thoroughly tested.</li><li>• Keep the buddy system but increase the number of people in the field to cover multiple deployment aspects at once.</li></ul> |
|----------------------|--|---|
-

- 
- Communication Surrounding Operation
- The procedures were nicely detailed.
  - The procedures were excessively long.
  - There were situations where volunteers used some information from the documentation and assumed the rest without asking questions.
  - Assigning management and decision tasks to volunteers should be avoided.
  - Volunteer training generally seemed effective at conveying organizational, safety, and basic operational information. However, some incidents and deviations from the procedures still occurred in the field.
  - Consider creating a 1–2-page checklist of the essentials in addition to the procedures. This document could have links to the other procedures.
  - Train volunteers more extensively to instill safe, reasonable behavior in the field. Make the training as interactive as possible.
  - No amount of documentation can solve some issues; we are always going to be partly reliant on the hope that everyone in the field shares a similar safety and operational culture to avoid incidents.
  - The graduate student should be in the field during every deployment. Techs should be hired to support the student.
  - Have a meteorologist on call for forecasting advice.
-

---

Campaign  
Design

- The deployment expectations were too high for this campaign.
- A full multi-IOP campaign this underprepared should not have occurred. We planned for deployments to be close to fully automated and were not able to meet this goal.
- Approaching a campaign with high expectations is not uncommon.
- If a campaign does not end up having the infrastructure anticipated, the goals need to be adjusted like they were during this campaign.
- After we had an incident during a deployment, the approach to deployments, documentation, and communication were altered. Deployments after did not have any issues.
- It is not necessary to call off a campaign unless you find a better use of resources, safety becomes an issue, or you cannot meet any of the science goals.



---

Enclosure  
Design

- The enclosure was very large, heavy, and difficult to maneuver.
  - Enclosure worked efficiently when equipped with two coolers. The original design only had one cooler, so the final version was even heavier and bulkier to move.
  - The EM27/SUN hatch was not large enough to safely close around the EM27/SUN when the mirrors were in any given position.
  - The EM27/SUN hatch was a potential pinch point. The automatic lid control would not stop if something was in its way.
  - The wheels on the enclosure had “self-leveling” capabilities which worked poorly. One of the wheels broke during deployment.
  - The build quality was poor and consequentially had some sharp edges.
  - The inside of the enclosure was difficult to organize.
  - An enclosure re-design should feature the following:
    - Larger wheels, better for difficult terrain.
    - One cooler with sufficient cooling power.
    - More insulation and a reflective coating to aid cooler.
    - Larger EM27/SUN hatch with at minimum, caution stickers to indicate pinch point risk.
    - A different leveling system not attached to the wheels.
    - Internal storage.
  - Consider a more modular enclosure design to reduce weight.
  - Investigate repurposing a garage door opener for the hatch control to create an inherently safer design.
-

---

Technology Implementation	<ul style="list-style-type: none"> <li>• The top priority moving forward should be acquiring and testing remote capabilities.</li> <li>• Precipitation sensors for automatic lid control would be extremely helpful.</li> <li>• Places like Houston are non-ideal for instruments that need clear sky conditions and are sensitive to high temperatures and humidity.</li> </ul>	<ul style="list-style-type: none"> <li>• Both the enclosure and automated systems should be rigorously designed and tested before another large-scale campaign like TRACER.</li> <li>• Parameters for EM27/SUN operation should be developed to ensure good quality data are taken even in difficult deployment conditions (e.g., the Houston summer heat).</li> <li>• Advanced lid control using precipitation sensors would be useful, but no matter the level of complexity introduced, there should always be a manual override available.</li> </ul>
---------------------------	--	---

---

## B Procedures for EM27/SUN Operation

### B.1 Description of the Secondary Instrumentation Suite for Remote Deployments

The following procedures include details necessary for deploying the secondary instrumentation suite which has been in development during the course of this work. Ideally, these instruments can be deployed along with an EM27/SUN to provide near-surface meteorological data and important metrics for automated enclosure lid control to protect the EM27/SUN from precipitation.

As discussed in Section 2.3, near-surface pressure data from the EM27/SUN deployment site are necessary for processing raw EM27/SUN data. The secondary instrumentation suite includes a Vaisala PTB330 pressure sensor and Campbell Scientific (Logan, Utah, USA) CR6 datalogger with Wifi in order to collect and store this data. When used in conjunction with a NETGEAR (San Jose, California, USA) Nighthawk

hotspot router and cellular data plan, instrument operation, troubleshooting, and data retrievals could take place remotely.

Other instrumentation included Hydreon (Eden Prairie, Minnesota, USA) RG-11 rain gauges are included in the instrumentation suite to further allow for remote instrument deployment. These rain gauges operate by detecting droplets on the sensor. This information could ideally be used to automatically open and close the enclosure lid to shield the EM27/SUN from rain. Additionally, a Lufft (Fellbach, Germany) WS500 smart weather sensor and a Campbell Scientific Apogee SP-230-SS heated pyranometer are included to collect data on the solar irradiance, temperature, wet bulb temperature, relative humidity, wind speed, and wind direction in the vicinity of the EM27/SUN. These secondary instruments were seldom used during the deployments discussed in this thesis other than for testing purposes, since most sites had similar sensors deployed in the vicinity by ARM or the University of Houston, depending on the deployment site. Continued testing of this instrumentation suite is underway to support future EM27/SUN deployments.

## **B.2 Combined Procedures for Tabletop and Enclosure-Based Deployments**


This procedure was designed to be used by all skill levels. The DETAILS section provides steps that can be easily followed by someone experienced with these tasks. The BACKGROUND provides information especially helpful for newer users. Once a user is comfortable with all the information by performing this procedure multiple times, the STEPS section can simply be used as a checklist.

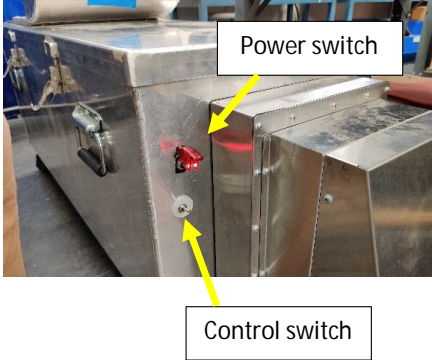

Any text in blue corresponds to instructions for EM27/SUN deployments with an enclosure only. Similarly, any text relating to deploying the secondary instrumentation suite is in green. Ignore any text in blue and green if only wanting to conduct a tabletop EM27/SUN deployment with no secondary instrumentation.

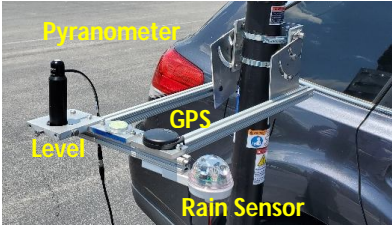

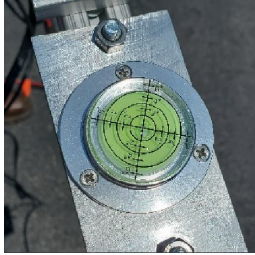


STEPS	DETAILS	BACKGROUND
Organize before deploying the EM27	<ol style="list-style-type: none"> <li>1. Check the weather and make sure the conditions are favorable for testing. Ideal conditions are minimal to no cloud cover and minimal chance of precipitation. The EM27s can handle some cloud cover and should be able to keep tracking the sun after a couple minutes of cloud disturbance. More disturbance may call for realignment (discussed later).</li> </ol>	<p>Here are some potentially useful sites for monitoring weather conditions:</p> <ul style="list-style-type: none"> <li>• National Weather Center (NWS) forecasts:  <a href="https://www.weather.gov/">https://www.weather.gov/</a> <ul style="list-style-type: none"> <li>○ Useful for hourly predictions.</li> </ul> </li> <li>• GOES satellite view:  <a href="https://weather.cod.edu/satrad/">https://weather.cod.edu/satrad/</a> <ul style="list-style-type: none"> <li>○ Useful for viewing current cloud cover.</li> </ul> </li> <li>• Radar:  <a href="https://weather.cod.edu/satrad/nextrad/">https://weather.cod.edu/satrad/nextrad/</a> </li> </ul>

The EM27s **CANNOT be rained on**. If deploying on a day with a **slight chance of precipitation or higher**, stay nearby to pack up the instrument quickly before the rain comes. Determine slight chance of precipitation or higher by using the hourly NWS weather forecast (see link in right column).


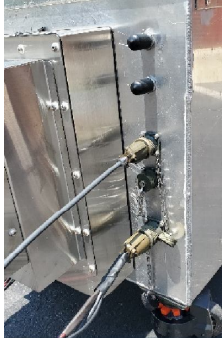


STEPS	DETAILS	BACKGROUND
Organize before deploying the EM27 (continued)	<ol style="list-style-type: none"> <li>2. Record the expected weather conditions in the digital lab notebook.</li> <li>3. Get equipment organized. Most items needed for deployment will stay with the EM27s, but some may not (i.e., laptops).</li> <li>4. Gather everything in the checklist and travel to the deployment site, in whatever order is required.</li> </ol>	<p>Checklist of items for deployment of a single EM27:</p> <ul style="list-style-type: none"> <li>• The EM27 in its travel case</li> <li>• EM27 enclosure + power cord (optional)</li> <li>• EM27 power cables</li> <li>• Ethernet cable</li> <li>• USB cable</li> <li>• Extension cord</li> <li>• The EM27 laptop</li> <li>• Laptop charger</li> <li>• Pressure sensor + datalogger</li> <li>• Hotspot + power cable, Ethernet cable</li> </ul>
<b>Optional:</b> Set up the enclosure	<ol style="list-style-type: none"> <li>5. If using an enclosure, place it on relatively flat, level ground such that the enclosure is oriented north/south with the rounded hatch pointing south. <b>If not using an enclosure, skip to step 16.</b></li> <li>6. Locate the green power cord with one female and one male end.  If you are setting up the enclosure after it has been stored in place at its location, the power cord is likely not plugged in and instead stored inside the enclosure itself. Open the enclosure by unlocking any locks hinging the top of the enclosure open.</li> <li>7. Close the enclosure lid if you had to open it. <ol style="list-style-type: none"> <li>a. Partially closing and opening the lid will disengage the locking mechanism that holds the lid in place so you can fully close the lid.</li> </ol> </li> </ol>	

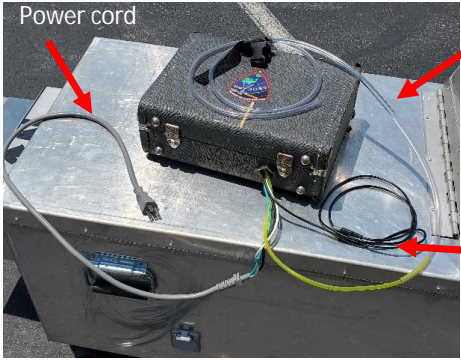

STEPS	DETAILS	BACKGROUND
<p><b>Optional:</b> Set up the enclosure (continued)</p>	<ol style="list-style-type: none"> <li>8. Connect the enclosure to power using the green cord and the electrical port on the outside of the enclosure.</li> <li>9. Make sure the enclosure has power by checking the coolers. If the fans are turning, you have power.</li> <li>10. Power the rounded EM27 hatch control using the red covered switch. The picture on the right shows the linear actuator power switch in the ON position.</li> <li>11. Raise the hatch by toggling the bottom switch up.</li> <li>12. Once the linear actuator has opened the EM27 hatch, flip the power switch to OFF and put the control switch back in neutral.</li> <li>13. View the current conditions provided on the NWS page for your deployment site (<a href="https://www.weather.gov/">https://www.weather.gov/</a>).</li> <li>14. Record current conditions in the lab notebook (temperature, humidity, dew point, wind speed/direction, cloud cover, etc.)</li> </ol>	<p>We want to limit the amount of time the enclosure is open in case the outside air is more humid than the air inside the box. This will help keep water vapor from condensing inside.</p> 

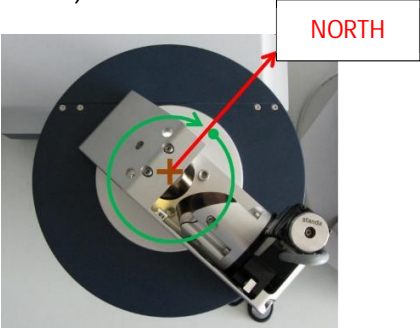
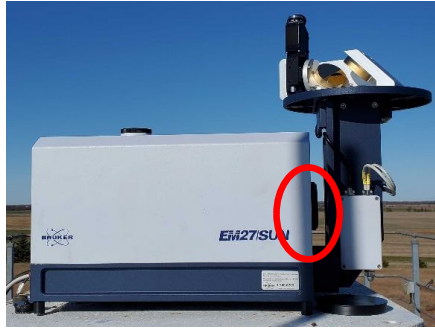
STEPS	DETAILS	BACKGROUND
<p><b>Optional:</b> Set up the enclosure (continued)</p>	<p>15. Set the cooler setpoint to three degrees greater than the dewpoint (in Fahrenheit) on the thermocouple readout (taped to the roof of the enclosure).</p> <p>a. To adjust the setpoint, hold the asterisk button and use the arrow keys to toggle to the correct temperature.</p>	<p>Linear actuator switches:</p>  <p>You should be able to access the thermocouple readout by reaching into the EM27 hatch when the enclosure lid is down. If this isn't possible, lift the lid of the enclosure to adjust the setpoint and close it once you are done.</p>
<p><b>Optional:</b> Set up secondary instrumentation</p>	<p>16. If setting up the secondary instrumentation suite, first place the Hurry-Up mast inside the drive-on mounting plate at a location at least 5 m from any tree or building.</p> <p><b>If not deploying secondary instrumentation, skip to step 25.</b></p> <p>17. Secure the mast plate with a car.</p> 	<p>The mast used here is a manual push-up telescoping mast created by Will-Burt (Orrville, Ohio, USA).</p> <p>Deploying away from tall objects is necessary to collect accurate data with the weather station.</p>

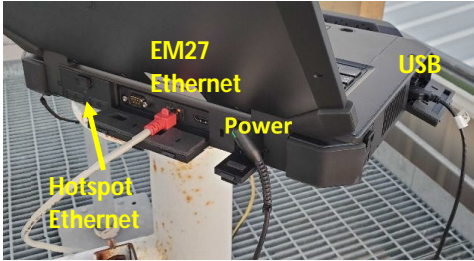

STEPS	DETAILS	BACKGROUND
<p><b>Optional:</b> Set up secondary instrumentation (continued.)</p>	<p>18. Attach the pyranometer, level, GPS, and rain sensor housing to the mast.</p>  <p>Secure the housing by tightening the two hose clamps to the mast with a flathead screwdriver.</p>  <p>19. Adjust the angle of the housing until the instruments are level with the ground.</p> <p>A level level:</p>  <p>20. Remove the green cap from the pyranometer so that it can collect data.</p> 	<p>This housing was custom-built and designed such that the angle of the housing could be adjusted manually using adjustable bolts in curved slots. After loosening the bolts, the housing could be tilted to the appropriate angle then secured by tightening the bolt.</p> 






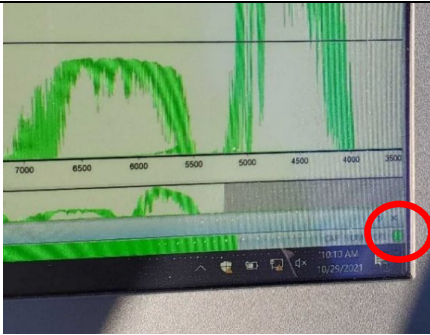
STEPS	DETAILS	BACKGROUND
<p><b>Optional:</b> Set up secondary instrumentation (continued.)</p>	<p>21. Anchor the weather station to the top of the mast. Ensure that the North arrow marked on the weather station is actually pointing North before securing the sensor.</p>  <p>22. Connect the weather station, pyranometer, GPS, and rain sensor to power and the datalogger.</p> <p>If using an enclosure, unique connectors are built into the side of the enclosure that are used for both power and data transfer to the datalogger.</p>   	<p>North marker</p> <p>Unique connectors:</p>


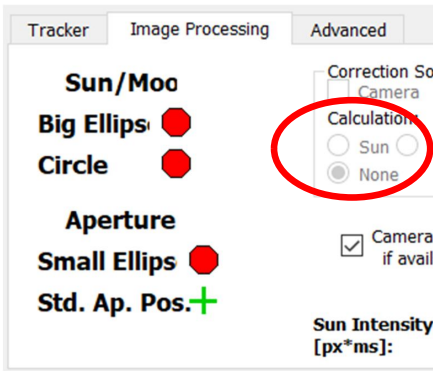
STEPS	DETAILS	BACKGROUND
<p><b>Optional:</b> Set up secondary instrumentation (continued.)</p>	<p>23. Locate the datalogger/pressure sensor case and ensure that the pressure sensor sample tube is unobstructed.</p>  <p>24. Plug in the datalogger/pressure sensor case to building power.</p>	<p>Sample tube</p> <p>USB connection for data retrieval from the logger</p>
<p>Set up EM27</p>	<p>25. Remove the EM27 from the travel case.</p> <p>26. If using an enclosure, open the enclosure lid. Set the EM27 on the raised platform in the enclosure such that the mirrors can “look” out of the enclosure through the EM27 hatch.</p> <p><b>If not using an enclosure, skip to step 27.</b></p>	<p>There may be a plastic cover over the mirror housing on the EM27. Set this aside before moving the EM27.</p> 

STEPS	DETAILS	BACKGROUND
Set up EM27 (continued)	<p>27. <b>If not using an enclosure</b>, place the EM27 on a flat testing surface (e.g., folding table) with the metal plane at the back of the mirror housing perpendicular to the North/South direction, with the metal plane farther South than the mirror.</p> <p>28. Check to see if the EM27 is level by checking the bubble level located on the blue platform below Mirrors 1 and 2.</p> <ol style="list-style-type: none"> <li>If the EM27 is level, skip to step 29.</li> <li>If the EM27 is not level, adjust the platform/enclosure that the EM27 is on until the instrument is level.</li> </ol> <p>29. Remove the lens cover from the body of the EM27.</p> <p>To remove the lens cover, pull the cover away from the body of the EM27 until it is detached, <b><u>making sure to not touch the third mirror or lens in the process.</u></b></p> <p>30. If using an enclosure, place the fabric EM27 collar around the mirror platform like the picture in step 26. Connect the Velcro to the enclosure as best as possible. Make sure to not obstruct the thermocouple readout in the process.</p>	<p>Mirror housing alignment (courtesy of Bruker):</p>  <p>Small terry towels are useful for folding and placing under table legs or travel case wheels to make the system level.</p> <p>Lens location:</p> 

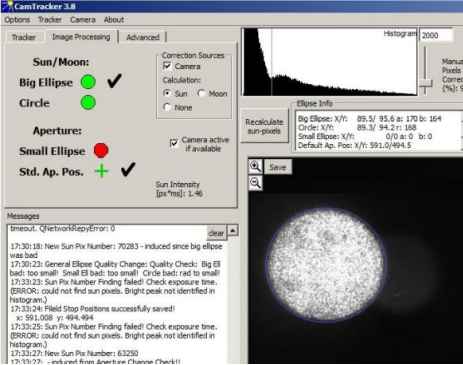

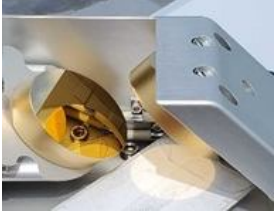

STEPS	DETAILS	BACKGROUND
<p>Connect the EM27 to power and laptop</p>	<p>31. If using an enclosure, connect the EM27, laptop, hotspot, and datalogger case power cords to the power strip located inside the enclosure. Also connect these cords to their respective devices.</p> <p><b>If not using an enclosure</b>, connect the all power cords (e.g., EM27, laptop, hotspot, and datalogger case) directly to the building or mobile power supply. Also connect these cords to their respective devices.</p> <p>Cable configuration on field laptop:</p>  <p>The EM27 Ethernet cable <b>must</b> be plugged into the correct port for everything to work properly.</p> <p>32. Connect the USB and Ethernet cables to the EM27 and the laptop.</p> <p>33. If using an enclosure, make sure that the open end of the clear tube connected to the datalogger case is positioned outside of the enclosure. <b>If not using an enclosure, skip to step 34.</b></p>	<p>If using an enclosure, the laptop and datalogger case can be kept inside the enclosure to keep them cool. If not using an enclosure, try to keep the equipment in the shade to avoid overheating effects.</p> <p>Cable configuration on the back of the EM27:</p>  <p>The metal enclosure acts as a Faraday cage, so the hotspot will not be able to work properly if inside the enclosure without antennae on the outside. Keep the hotspot outside of the enclosure unless you have supplementary antennae (future editions of the procedure will include this).</p>

STEPS	DETAILS	BACKGROUND
<p>Connect the EM27 to power and laptop (continued)</p>	<p>34. Connect the hotspot Ethernet cable (white cable with clear connections) to the hotspot and the field laptop.</p> 	<p>The pressure sensor needs to be able to sample ambient air outside of the enclosure to accurately measure pressure.</p> <p>Hotspot Ethernet connection</p> <p>Hotspot power connection</p>
<p>EM27 Startup</p>	<p>35. Turn the laptop on and log in.</p> <p>36. Press the ON/OFF button on the back of the EM27 and make sure the light above it eventually stays green.</p>  <p>37. Open OPUS using the icon at the bottom of the screen. The password for OPUS is OPUS</p> 	<p>Note: the touch pad on the laptop is difficult to use. Use the attached stylus on the touch screen for more effective interfacing.</p> <p>The light above the ON/OFF button should flash red and green several times before staying green.</p> <p>Avoid opening OPUS versions not pinned to the taskbar at the bottom of the screen. These versions may not be compatible with the EM27 anymore.</p>

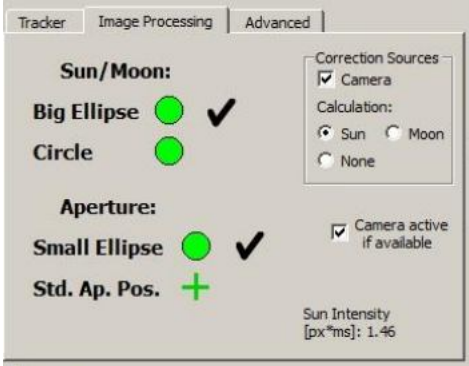
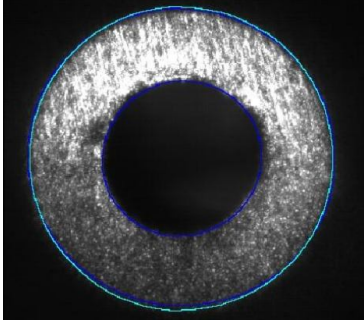
STEPS	DETAILS	BACKGROUND																		
EM27 Startup (continued)	<p>38. Ensure that everything is connected properly by checking the circle in the lower right-hand corner of the screen.</p> <ol style="list-style-type: none"> <li>If it is green, continue.</li> <li>If it is gray, wait a minute to see if it turns green. If not, ensure all connections are secure, restart OPUS, and check again.</li> </ol> <p>39. Check to see if the time on the laptop is correct. If it is different, even by a minute, change the time manually.</p>																			
CamTracker Startup	<p>40. Open "CamTrackerConfig.txt" located in C:\Users\Public\Documents\Bruker\camtracker_3_9_1_0</p> <p>41. Make sure the information on lines 9-11 match the information for the site you are testing at by using the table on the right.</p> <ol style="list-style-type: none"> <li>If testing at a site not listed in the table, use Google Maps to find the coordinates and <a href="https://apps.nationalmap.gov/elevation/">https://apps.nationalmap.gov/elevation/</a> to find the ground-level elevation, which should be modified if setting up on a platform, building, etc.</li> </ol>	<table border="1"> <thead> <tr> <th data-bbox="915 1010 1045 1163">Location</th> <th data-bbox="1045 1010 1127 1163">Site Code</th> <th data-bbox="1127 1010 1344 1163">Details (Latitude, Longitude in degrees; elevation in km)</th> </tr> </thead> <tbody> <tr> <td data-bbox="915 1163 1045 1255">Lamont, OK next to TCCON</td> <td data-bbox="1045 1163 1127 1255">LT</td> <td data-bbox="1127 1163 1344 1255">Lat: 36.6038 Lon: -97.4856 Elev: 0.3181 +</td> </tr> <tr> <td data-bbox="915 1255 1045 1348">TRACER Main ARM Site</td> <td data-bbox="1045 1255 1127 1348">T5</td> <td data-bbox="1127 1255 1344 1348">Lat: 29.6692 Lon: -95.0595 Elev: 0.0079 +</td> </tr> <tr> <td data-bbox="915 1348 1045 1503">TRACER UH Coastal Center Site</td> <td data-bbox="1045 1348 1127 1503">T6</td> <td data-bbox="1127 1348 1344 1503">Lat: 29.3884 Lon: -95.0425 Elev: 0.0055 +</td> </tr> <tr> <td data-bbox="915 1503 1045 1625">TRACER UH Main Campus Site</td> <td data-bbox="1045 1503 1127 1625">T7</td> <td data-bbox="1127 1503 1344 1625">Lat: 29.7240 Lon: -95.3392 Elev: 0.0121 +</td> </tr> <tr> <td data-bbox="915 1625 1045 1717">TRACER UH Aldine Site</td> <td data-bbox="1045 1625 1127 1717">T8</td> <td data-bbox="1127 1625 1344 1717">Lat: 29.9012 Lon: -95.3262 Elev: 0.0214 +</td> </tr> </tbody> </table> <p data-bbox="915 1717 1344 1776">+: elevation includes 3 ft for deployment on a table or instrument travel case</p>	Location	Site Code	Details (Latitude, Longitude in degrees; elevation in km)	Lamont, OK next to TCCON	LT	Lat: 36.6038 Lon: -97.4856 Elev: 0.3181 +	TRACER Main ARM Site	T5	Lat: 29.6692 Lon: -95.0595 Elev: 0.0079 +	TRACER UH Coastal Center Site	T6	Lat: 29.3884 Lon: -95.0425 Elev: 0.0055 +	TRACER UH Main Campus Site	T7	Lat: 29.7240 Lon: -95.3392 Elev: 0.0121 +	TRACER UH Aldine Site	T8	Lat: 29.9012 Lon: -95.3262 Elev: 0.0214 +
Location	Site Code	Details (Latitude, Longitude in degrees; elevation in km)																		
Lamont, OK next to TCCON	LT	Lat: 36.6038 Lon: -97.4856 Elev: 0.3181 +																		
TRACER Main ARM Site	T5	Lat: 29.6692 Lon: -95.0595 Elev: 0.0079 +																		
TRACER UH Coastal Center Site	T6	Lat: 29.3884 Lon: -95.0425 Elev: 0.0055 +																		
TRACER UH Main Campus Site	T7	Lat: 29.7240 Lon: -95.3392 Elev: 0.0121 +																		
TRACER UH Aldine Site	T8	Lat: 29.9012 Lon: -95.3262 Elev: 0.0214 +																		

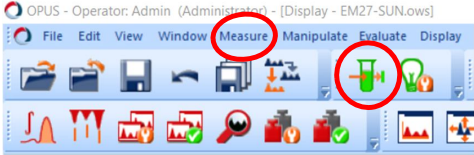
STEPS	DETAILS	BACKGROUND
	<p>42. Save the file and close it.</p> <p>43. Open CamTracker. </p> <p>44. If you modified the CamTracker configuration file, select "Read Configuration File" under the "Options" drop down menu.</p> <p>45. In the "Tracker" tab, click "Initialize Tracker."</p> <p>46. Navigate to the "Imaging Processing" tab and click on the bubble next to the "Sun" option under "Calculation:"</p>	 <p>The screenshot shows the 'Advanced' tab of the CamTracker software. Under the 'Calculation:' section, the 'Sun' radio button is selected and circled in red. Other options include 'None' and 'Camera'. There are also checkboxes for 'Correction So' and 'Camera if avail', and a label for 'Sun Intensity [px*ms:]'.</p>



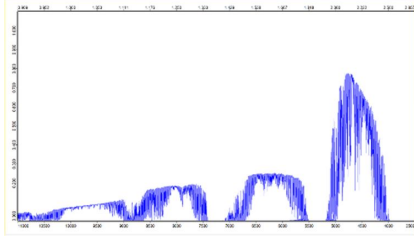
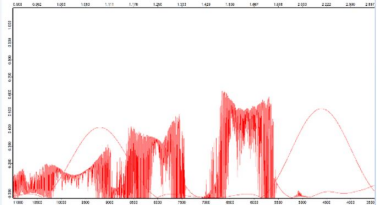

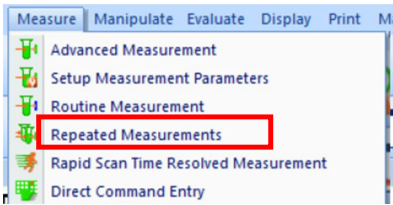
STEPS	DETAILS	BACKGROUND
<p>CamTracker Startup (continued)</p>	<p>47. Navigate back to the "Tracker" tab and align the mirrors so they direct the sunlight into the EM27. Use the "+" and "-" buttons for Motor 1 and Motor 2 to rotate the elevation and azimuthal mirrors, respectively.</p> <p>c. Motor adjustment size can be changed by typing new values in the "size" box before clicking the "+" and "-" buttons.</p> <p><b>First</b>, make sure all incident light is being reflected from Mirror 2 into the tube to Mirror 3.</p> <p><b>Then</b>, make minor (~0.5 or less) motor adjustments while monitoring the histogram and camera preview in CamTracker. The histogram will grow when more light is present. <u>The mirrors will be aligned when there is a circle of light in the center of the camera preview and there is a thin blue circle around the edge of the circle of light</u> (shown below, minus the centered image requirement. Picture courtesy of Bruker)</p> 	<p>Pieces of paper are useful for checking the alignment after Mirror 2 (the azimuthal mirror) and Mirror 3 (by the lens). Blocking incident light not from Mirror 1 (the elevation mirror) and blocking the light path from mirror to mirror and help identify alignment issues:</p> <p>Blocking incident light:</p>  <p>Blocking light path:</p>  <p><b>Proper alignment should result in the following observations:</b></p> <ul style="list-style-type: none"> <li>• No sunlight reflected from mirror housing when looking at the sunlight reflected from Mirror 2.</li> <li>• Perfect circles of light reflected from Mirrors 2, 3.</li> <li>• Sharp edges of the circle reflected from mirror 3. Very minor adjustments can be made to make this happen (motor adjustments of around 0.02).</li> </ul> <p>A good example of the light reflected from Mirror 3 when Mirrors 1 and 2 are aligned:</p> 

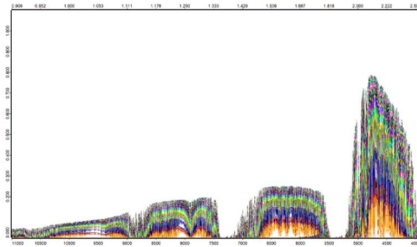


STEPS	DETAILS	BACKGROUND
CamTracker Startup (continued)	<p>48. Once the mirrors appear to be aligned, remove any paper used for adjustments and click on the "Image Processing" tab.</p> <ul style="list-style-type: none"> <li>b. If there are green circles next to "Big Ellipse" and "Circle," along with check marks by these same indicators, continue to step 49.</li> <li>c. If one of the above conditions is not met, adjust the mirrors to achieve better alignment and check again.</li> </ul> <p>49. Once the good quality conditions are met in step 48, click the box next to "Camera" under the header that reads "Correction Sources."</p> <p>50. Wait for the EM27 to adjust. Watch the condition indicators and the sunlight reflection image in the "Image Processing" tab.</p> <p>51. Once all indicators are green (image below courtesy of Bruker), CamTracker is working properly and should successfully track the Sun unless interrupted by clouds, night, etc.</p> 	<p>The sunlight reflection should look as follows, with blue circles surrounding the inner and outer ellipses (courtesy of Bruker):</p> 

STEPS	DETAILS	BACKGROUND
CamTracker Startup (continued)	<p>52. Record the time, elevation and azimuth angles, as well as the motor offsets in the lab notebook. These can be useful for simplifying the alignment process during later deployments at the same site.</p>	
Data Collection in OPUS	<p>53. Keeping CamTracker open, navigate back to OPUS.</p> <p>54. Take a single measurement by either clicking on the Measurement button (green test tube) in the main ribbon of the GUI or selecting "Advanced Measurement" from the "Measure drop-down menu.</p>  <p>The screenshot shows the OPUS software interface. The title bar reads 'OPUS - Operator: Admin (Administrator) - [Display - EM27-SUN.ows]'. The main ribbon includes 'File', 'Edit', 'View', 'Window', 'Measure', 'Manipulate', 'Evaluate', and 'Display'. The 'Measure' ribbon is active, and a green test tube icon is circled in red. Below the ribbon, there are several tool icons, including a magnifying glass and a green checkmark.</p>	

STEPS	DETAILS	BACKGROUND
Data Collection in OPUS (continued)	<p>55. Ensure that all the settings are appropriate. The most important are listed here:</p> <ul style="list-style-type: none"> <li>d. In the "Advanced" tab: <ul style="list-style-type: none"> <li>i. File name: bl&lt;Y&gt;&lt;m&gt;&lt;d&gt;.ifg.0000</li> <li>ii. Path: either an existing or new folder of your chosen name under C:\Data</li> <li>iii. Resolution: <b>0.5 cm<sup>-1</sup></b></li> <li>iv. Sample scan time: <b>2 scans</b></li> <li>v. Background scan time: <b>1 scans</b></li> <li>vi. Save data from: <b>100 to 15797.9 cm<sup>-1</sup></b></li> <li>vii. Data blocks to be saved: Single Channel, Sample Interferogram</li> </ul> </li> <li>e. In the "Optic" tab: <ul style="list-style-type: none"> <li>i. Sample preamp gain: <b>B</b></li> <li>ii. Background preamp gain: <b>A</b></li> </ul> </li> <li>f. In the "Acquisition" tab: <ul style="list-style-type: none"> <li>i. Wanted high/low frequency limit: set to match the range given in the "Advanced" tab</li> </ul> </li> <li>g. In the "FT" tab: <ul style="list-style-type: none"> <li>i. Phase resolution: <b>12</b></li> <li>ii. Phase correction mode: <b>Mertz</b></li> <li>iii. Apodization function: <b>Boxcar</b></li> <li>iv. Zerofilling factor: <b>2</b></li> </ul> </li> </ul> <p>56. Back in the "Basic" tab, click the "Sample Single Channel" button.</p>	<p>Add folders as needed or desired, just make sure you note where OPUS is saving data.</p> <p>The upper limit of the frequency range may not be accepted by OPUS. Just use the highest number allowed that is below 15798 cm<sup>-1</sup>.</p>

STEPS	DETAILS	BACKGROUND
<p>Data Collection in OPUS (continued)</p>	<p>57. Look at the interferogram in OPUS. If it looks reasonable, skip to step 59.</p> <p>The interferogram should look something like this when looking at the interferogram generated by both detector channels:</p>  <p>58. If the interferogram does not look reasonable, go back to step 55, and repeat until this point, using different preamp gains in the "Optic" tab until a reasonable interferogram is generated.</p> <p>59. Under the "Measure" drop-down menu, select the "Repeated Measurements" option.</p>	 <p>Unreasonable interferograms may be warped or have pseudo sinusoidal curves present like this interferogram:</p> <p>To view the interferogram generated by both detector channels, click on the "2 CHN" button (highlighted in red below) under the interferogram name in the OPUS Browser that you want to view.</p>  <p>Other preamp gain combinations that may work better depending on the conditions:</p> <ul style="list-style-type: none"> <li>• Sample gain of B, background gain of C</li> <li>• Sample gain of B, background gain of B</li> <li>• Sample gain of A, background gains of A/B/C (only use if the sun intensity is very low)</li> </ul> 



STEPS	DETAILS	BACKGROUND
Data Collection in OPUS (continued)	<p>60. Adjust the settings according to Step 55 with the following exceptions:</p> <ul style="list-style-type: none"> <li>h. In the “Basic” tab: <ul style="list-style-type: none"> <li>i. Repeat the measurement: <b>3000 times</b></li> <li>ii. Delay between measurements: <b>0 sec</b></li> </ul> </li> </ul> <p>61. Navigate to the “Check Signal” tab and wait for the signal to stabilize.</p> <p>62. Record the amplitude and position in the lab notebook.</p> <p>63. Click the “Save Peak Position” button.</p> <p>64. Back in the “Basic” tab, click the “Repeated Sample Single Channel” button.</p> <p>65. If using an enclosure, place the laptop inside the enclosure with the screen still open. Shut the lid to the enclosure so ensure the inside will stay conditioned.</p>	<p>This will ensure that the spectrometer collects data all day.</p>
EM27 monitoring	<p>66. If the EM27 appears to be operating as expected (CamTracker’s quality indicators are green, OPUS is recording reasonable looking interferograms), you can sit back and enter babysitter mode. Make sure to place the laptop in the shade if possible to avoid overheating.</p> <p>Recall that if there <i>is a slight chance or higher of rain/thunderstorms</i>, you must stay at the site so you can cover the EM27 or <a href="#">close the EM27 hatch on the enclosure</a> before rain begins.</p>	<p>If set up properly, OPUS should gather spectra that look like the one shown in step 57 above. Throughout the day, interferogram peak magnitude will fluctuate according to the solar zenith angle (SZA). Here is an example of a full day of data:</p> 



STEPS	DETAILS	BACKGROUND
EM27 monitoring (continued)	<p>67. Throughout the day, especially if you have reason to believe that something (clouds, people, etc.) interrupted the column of air that the EM27 is analyzing, check CamTracker and OPUS to make sure everything is working as it was when you left it during step 66.</p> <p>Record observations in the lab notebook (i.e., sky cover, temperature, time window of deployment, purpose of deployment, any issues encountered, etc.).</p>	<p>If the EM27 stops tracking the sun due to a disturbance, realign the mirrors using the same process you used earlier.</p>
Halt data collection and shut down EM27	<p>68. If at the end of the day, OPUS is still trying to collect data, right click on the status bar at the bottom of the screen.</p> <p>69. Halt the data collection by selecting the "Abort Task Group" option in OPUS. Move on to the next step only once the status bar says "No Active Task" like it did when you first opened OPUS.</p> <p>70. Record what time you ended the data collection in the lab notebook.</p> <p>71. Close OPUS.</p> <p>72. With CamTracker open, record the elevation and azimuth angles, as well as the motor offsets in the lab notebook.</p> <p>73. In the "Tracker" tab of CamTracker, click the button that reads "Move Tracker to park pos."</p> <p>74. Once both mirrors on the EM27 have stopped moving, close CamTracker.</p>	<p>CamTracker may ask if you really want to quit. Simply click 'yes.'</p>


STEPS	DETAILS	BACKGROUND
<p><b>Optional:</b> Prepare EM27 for next day deployment with enclosure</p>	<p>75. <b>If not using an enclosure</b>, skip to step 84.</p> <p>If using an enclosure, determine if you are going to be deploying at the same site the next day.</p> <p>76. If you are deploying at the same site the next day, leave the EM27 powered on and connected to the computer.</p> <p>77. Unplug the datalogger case from power to keep it from taking data overnight.</p>	<p>Leaving the EM27 on overnight will decrease your chances of dew forming on the external optics since the EM27 itself produces heat.</p>
<p><b>Optional:</b> Prepare EM27 for next day deployment with enclosure (continued)</p>	<p>78. Check the NWS' predicted low temperature for the night.</p> <p>79. Change the enclosure temperature set point to at least two degrees above the low temperature.</p> <p>80. Lower the EM27 hatch using the toggle switches described in steps 10-12.</p> <p>81. Close the enclosure lid for the night.</p> <p>82. Secure the enclosure by locking the lid and chaining it to nearby permanent objects if able.</p> <p>83. When deploying the next day, repeat the steps above as needed. Continue repeating these steps until you are ready to leave the site and then follow the remaining steps to properly store everything.</p>	<p>This is another safeguard to prevent dew from forming on the optics overnight.</p>
<p>Shut down and store the instruments</p>	<p>84. Power off the EM27.</p> <p>85. Put the EM27 lens cap over the lens <b><u>without touching any of the mirrors or the lens itself.</u></b></p>	

STEPS	DETAILS	BACKGROUND
Shut down and store the instruments (continued)	<ol style="list-style-type: none"> <li data-bbox="483 342 834 405">86. Disconnect all cables from the EM27, laptop, and hotspot.</li> <li data-bbox="483 436 862 527">87. Gather the cords neatly using the provided twist-ties and/or cord holders.</li> <li data-bbox="483 558 889 680">88. Locate the large plastic bag used for cord storage. Place all cords in the bag except for the extension cord and enclosure power cord, if used.</li> <li data-bbox="483 711 870 802">89. Unplug the datalogger case from power but leave the power cables attached to the case.</li> <li data-bbox="483 833 886 924">90. Screw the port connectors back on the USB, power, and Ethernet ports on the EM27.</li> <li data-bbox="483 955 867 1136">91. Place the EM27 back in the travel case as you found it earlier. <b>The EM27's orientation within the travel case is very important to properly store it.</b> See the pictures below for guidance.</li> </ol>	

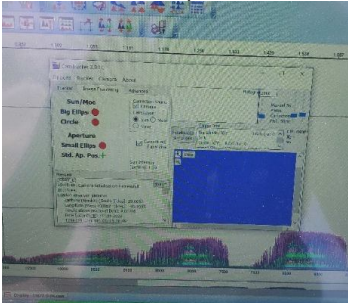
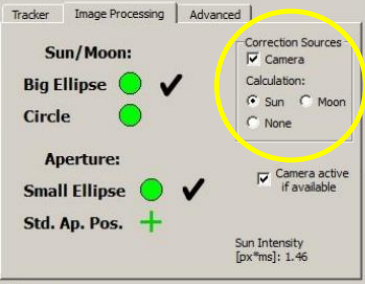



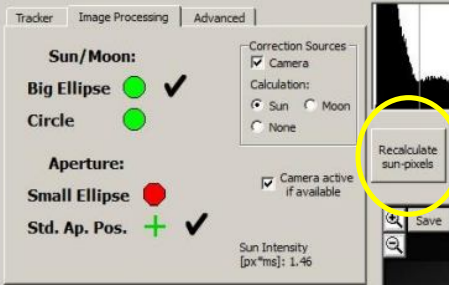
STEPS	DETAILS	BACKGROUND
<p>Shut down and store the instruments (continued)</p>	<p>92. Place the laptop, cords (minus the extension cord and/or enclosure power cord), and ratchet strap in the case. For our EM27/travel case combination, use the following layout:</p>  <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div data-bbox="448 884 574 947" style="border: 1px solid black; padding: 2px;">Laptop</div> <div data-bbox="626 884 776 947" style="border: 1px solid black; padding: 2px;">Cord Bags</div> <div data-bbox="824 884 976 961" style="border: 1px solid black; padding: 2px;">Ratchet Strap</div> </div> <p>93. Once the case contents are properly stored, carefully put the lid back on the travel case.</p> <p>94. If using an enclosure, unplug the enclosure power cord and store it inside the enclosure.</p> <p>95. If using secondary instrumentation, unplug all instruments from power and store as necessary for future deployments.</p> <p>96. Remove towels or other objects from beneath the testing surface (or enclosure) that may have been used to level the EM27.</p> <p>97. Store everything <i>except for the laptop, hotspot, and datalogger</i> as needed.</p>	

STEPS	DETAILS	BACKGROUND
Shut down and store the instruments (continued)	98. Set the laptop and hotspot up in a secure location such that they are both connected to power and to each other. <u>Sign out of Administrator, but do not power down the laptop.</u>	
Transfer data from field equipment	<p>99. Within your EM27 data folder on the field laptop, create folders for every date you deployed.</p> <p>100. Sort all the raw EM27 data files into the appropriate date folders.</p> <p>101. Leave the laptop connected to the hotspot and building power. Every day at 19:00 CST, new directories and their contents within C:\Data will be automatically copied to Schooner via SCP.</p> <p><b>The following steps are for downloading data from the secondary instrumentation.</b></p> <p>102. Plug the USB cord attached to the datalogger case to the field laptop.</p> <p>103. On the field laptop, open LoggerNet using the desktop icon.</p>  <p>104. On the LoggerNet GUI, select "Connect" under the Main tab.</p>  <p>105. Make sure that "CR6Series" is listed under the "Stations" menu.</p>	<p>There is no mandatory naming convention for the date folders, but YYYYMMDD is customary.</p>

STEPS	DETAILS	BACKGROUND
<p>Transfer data from field equipment (continued)</p>	<p>106. Click on "CR6Series" to select it if it is not already selected.</p> <p>107. Click the "Connect" button.</p> <p>Once the logger is connected, the "Connect" button will change to a "Disconnect" button.</p>  <p>108. Click on the "Custom" button in the same ribbon as the "Connect"/"Disconnect" button.</p> <p>109. In the Custom Collection screen, adjust the following settings:</p> <ol style="list-style-type: none"> <li>Collect Mode: Data From Selected Date and Time</li> <li>File Mode: Append to End of File</li> <li>File Format: ASCII Table Data, Long Header (TOA5)</li> <li>In the "Starting Date/Time" section, specify the date and time options as desired to save the appropriate data</li> <li>In the "Format Options" section, only select Include Timestamp, Include Record Number</li> <li>In the "Table Collection" section, select the necessary files to save</li> </ol> <p>110. Once all settings have been adjusted, click "Start Collection."</p> <p>A new popup window entitled "Data Collection Results" should appear when the collection is complete.</p> <p>111. Click "Ok."</p> <p>112. The data will be saved in *.dat files located in the directory/directories specified in the "Table Collection" section mentioned above.</p>	<p>The "Append to End of File" option ensures that the new data you save is appended to the existing file instead of creating a new file or replacing an existing file. This is important to note when processing data for analysis.</p>

## B.3 Troubleshooting Guide

ISSUE	RECOMMENDATIONS
<p><b><u>CamTracker camera preview is completely white or blue.</u></b></p> <p>Blue screen example:</p> 	<p>Suggestion 1:</p> <ol style="list-style-type: none"> <li>1. Laptops can easily overheat due to high ambient temperatures and excessive incident light. If the laptop is hot to the touch, move it into the shade if possible.               <ol style="list-style-type: none"> <li>a. If there is no shade available, consider shutting the laptop down and leaving it in an air-conditioned building until it has cooled off before continuing.</li> </ol> </li> </ol> <p>Suggestion 2:</p> <ol style="list-style-type: none"> <li>2. Check the “Camera” box in the “Correction Sources” box under the “Image Processing” tab in CamTracker.            </li> <li>3. Wait for the motors to stop moving.</li> <li>4. If the camera preview changes back to black, check the “Camera” box again, make any necessary motor adjustments, and continue taking data.</li> </ol> <p>Suggestion 3:</p> <ol style="list-style-type: none"> <li>5. Uncheck the box labeled “Camera active if available”</li> <li>6. Wait up to a minute to see if the camera preview changes.</li> <li>7. Recheck the box.</li> </ol> <p>Suggestion 4:</p> <ol style="list-style-type: none"> <li>8. If this does not help, put the tracker back in park position.</li> <li>9. Restart CamTracker.</li> <li>10. Make the tracker track the sun (see main Manual Operation procedure for more details).</li> </ol> <p>Suggestion 5:</p> <ol style="list-style-type: none"> <li>11. If the camera preview is still white or blue, halt data collection.</li> <li>12. Put the tracker in park.</li> <li>13. Restart the computer.</li> <li>14. Follow the steps for operating OPUS and CamTracker to begin taking data again (see the main Manual Operation procedure).</li> </ol>

ISSUE	RECOMMENDATIONS
<p><b><u>CamTracker quality checks are not all green and the sun intensity is reading low values or 'nan.'</u></b></p>	<p>Suggestion 1:</p> <ol style="list-style-type: none"> <li>The sky may not be clear enough for data collection. Even cirrus clouds (thin clouds in the picture on the right) or relatively slow-moving cumulus clouds (smaller clouds in the foreground of the picture) can keep the EM27 from being able to track properly. Wait for more favorable sky conditions before re-aligning the mirrors and continuing to take data.</li> </ol>  <p>Suggestion 2:</p> <ol style="list-style-type: none"> <li>If the sky is clear, the mirrors might not be aligned perfectly. Make sure all sun hitting the first mirror is being transferred to the second mirror.</li> <li>Adjust the motor positions if necessary.</li> <li>Make sure all sun hitting the third mirror is going into the lens in a crisp, full circle of light by looking at the reflection on the lens. See the main Manual Operation procedure for more details and example pictures.</li> <li>Adjust the motor positions if necessary.</li> </ol> <p>Suggestion 3:</p> <ol style="list-style-type: none"> <li>If the mirrors appear to be aligned properly and the issue persists, click the "Recalculate sun pixels" button in CamTracker.</li> </ol>  <p>Suggestion 4:</p> <ol style="list-style-type: none"> <li>Restart the alignment process by putting the mirrors back in park position and realigning.</li> </ol>

ISSUE	RECOMMENDATIONS
<p><b><u>Continued: CamTracker quality checks are not all green and the sun intensity is reading low values or 'nan.'</u></b></p>	<p>Suggestion 5:</p> <ol style="list-style-type: none"> <li>8. If the issue persists, document the issue and how you tried to fix it. There may be an issue with the EM27 that needs to be addressed by a Bruker technician.</li> </ol>
<p><b><u>CamTracker quality checks are not all green and the sun intensity is displaying values much larger than previously.</u></b></p>	<p>Suggestion 1:</p> <ol style="list-style-type: none"> <li>1. Check the "Camera" box in the "Correction Sources" box under the "Image Processing" tab in CamTracker.</li> <li>2. If the motors move, continue to Step 3. If they do not move, skip to Suggestion 2.</li> <li>3. Wait for the motors to completely stop moving. This may take around a minute or longer as the motors will be moving very slowly.</li> <li>4. Click on the "Camera" box again.</li> <li>5. Make any necessary motor adjustments to realign the mirrors.</li> </ol> <p>Suggestion 2:</p> <ol style="list-style-type: none"> <li>6. Put the tracker into park position.</li> <li>7. Restart CamTracker and realign mirrors.</li> </ol>

# C Procedures for Retrieval Algorithm Processing

## C.1 PROFFAST Version 2.2 Procedure

STEPS	DETAILS	BACKGROUND
<p>Prepare the data files to be processed</p>	<ol style="list-style-type: none"> <li>1. Locate your raw EM27 data files.</li>   <li>2. If not already done, organize files by day and instrument. Main folders should correspond to the instrument serial numbers in SN### format (e.g., SN155). Nested folders should correspond to the deployment date in YYMMDD format (e.g., 221019 for a deployment on October 19, 2022).</li>   <li>3. Ensure that all raw EM27 data files are named according to the following file naming convention: YYMMDDSN.###</li>   <li>4. Locate your surface pressure data. The location as well as file name, structure, and type can be customized to your liking, but should be repeatable between files. Make note of the path to your data.</li>   <li>5. If not already done, collect the a priori profile (*.map) data files. It is recommended that you use the sub-daily files created using TCCON data, which can be made on request for any site and date range.</li>   <li>6. Make note of the path to your *.map files.</li> </ol>	<p>The location of the files does not matter so long as the code in your <i>proffastpylot</i> directory is able to access the file location without the need for extra permissions (i.e., do not store raw EM27 data on a remote cluster and try to run the PROFFAST code locally).</p> <p>The four digits making up the file extension are assigned by the OPUS software for the EM27.</p> <p>See the TCCON wiki for more information on gathering *.map files: <a href="https://tcon-wiki.caltech.edu/Main/WebHome">https://tcon-wiki.caltech.edu/Main/WebHome</a></p>





STEPS	DETAILS	BACKGROUND
<p>Set up the retrieval directory in <i>proffastpylot</i> (continued)</p>	<p>and altitude in the input file, make sure no path is listed on <a href="#">line 37</a>.</p> <p>e. <a href="#">Line 41</a>: offset in hours from UTC.</p> <p>f. <a href="#">Line 62</a> (optional): the smallest file size you wish to process in megabytes</p> <p>g. <a href="#">Lines 78-79</a> (optional): date range you wish to process. Not necessary if wanting to process all dates</p> <p>h. <a href="#">Line 125</a>: full path to EM27 data location</p> <p>i. <a href="#">Line 126</a>: full path to location of *.map files</p> <p>j. <a href="#">Line 127</a>: full path to pressure data location</p> <p>15. Save and close your renamed input file.</p> <p>16. Open <i>log_type_pressure.yml</i> and edit the following lines:</p> <p>a. Lines 6-8: filename parameters for identifying your pressure data files. Each line is defined in the comments of the *.yml file.</p> <p>b. Lines 16-24: file structure parameters for parsing the data inside your pressure data files. Each line is defined in the comments of the *.yml file.</p> <p>c. Line 38: offset in hours from UTC. Used for converting the times in your pressure file to UTC.</p> <p>17. Save and close your pressure *.yml file.</p>	<p>Example: use -6.0 if you collected data in Central Standard Time.</p> <p>1.0 was used for processing TRACER data.</p> <p>Note that data collected at different sites should be processed in a separate retrieval run.</p> <p>The separator (<i>sep</i>, line 24) specified in the example file ("<i>\t</i>") indicates that the pressure file is tab delimited.</p> <p>This offset may be different than the one used in step 15 depending on what time zone the pressure sensor was set up to collect data in</p>

STEPS	DETAILS	BACKGROUND
Set up the retrieval directory in <i>proffastpylot</i> (continued)	<p>18. Open <i>run.py</i> and edit the following lines:</p> <ul style="list-style-type: none"> <li>a. <u>Line 24</u>: the name of your renamed input file</li> <li>b. <u>Lines 10, 18-21</u>: delete the contents. These lines are only meant to be used with the example data provided with the code.</li> </ul> <p>19. Save and close <i>run.py</i>.</p>	
Run the retrieval	<p>20. Navigate back to the <i>proffastpylot</i> folder.</p> <p>21. Activate the Python virtual environment using the following command: <code>source prf_venv/bin/activate</code></p> <p>22. Navigate to your retrieval directory.</p> <p>23. Run the retrieval algorithm using the following command: <code>python run.py</code></p> <p>24. After the retrieval has finished running, find your retrieved data in the results sub-directory of your retrieval directory. <i>*invparms*</i> files contain ASCII tables with retrieved trace gas results. <i>*.spc</i> files contain information about the different spectral fits generated by the retrieval algorithm.</p>	

## C.2 GGG2020 Procedure

Directions on how to install and run the GGG2020 retrieval algorithm code and EM27/SUN GGG interferogram (EGI) processing suite, necessary for running GGG with EM27/SUN data, can be found at <https://tcon-wiki.caltech.edu/Main/WebHome>. You must create a user profile to view this information. Conveniently, test data and example files are provided with the code so a new user can easily follow along with the instructions on the website.

A brief summary of how to set up and run a retrieval post-installation is detailed here, including information on how to adjust the  $\delta\nu$  parameter and apply EM27/SUN-specific AICF and ADCF, like those from Mostafavi Pak et al. (2023) used during this study.

- Just as with PROFFAST, GGG requires EM27/SUN interferograms, surface meteorological data, and *a priori* profile data (referred to as \*.map files in the PROFFAST procedure above). As discussed in the PROFFAST procedure, the *a priori* data should be downloaded for each EM27/SUN deployment date and site combination using the instructions given on the TCCON Wiki (see link above).
  - GGG expects to be given data with specific filenames. Raw EM27/SUN data should be named as follows: *xxYYYYMMDD.ifg.####*. Here, *xx* represents the two-letter instrument code, *YYYYMMDD* is the date, and *####* represents interferogram number as assigned by OPUS. Surface meteorological data should be named by date in *YYYYMMDD.txt* format.
  - Additionally, surface meteorological data should be formatted precisely as detailed on the TCCON wiki (see link above).
  - Data file locations do not matter so long as the paths to the files are provided to the code. For GGG, these directory locations must be specified in the EGI file location list (*file\_loc\*.lst*). A unique version of this file should be created for each EM27/SUN. An example file is included along with GGG upon downloading the software. This example file is located in the following directory: `.../egi/setup_examples/file_loc_ex2020.lst`.
- If modifying the  $\delta\nu$  parameter (necessary for Blanche in this study), an override file needs to be created for the instrument. The example file included in GGG is located in `.../egi/setup_examples/setup_test/xa_override.dat`, where *xa* should be the same two-letter instrument code used in the interferogram file naming convention and as provided in the file location list (*file\_loc\*.lst*).
  - Copy the example file and replace *xa* with the appropriate instrument code.

- Modify line 21 as desired to adjust the  $\delta\nu$  parameter. The minimum  $\delta\nu$  can be calculated using  $\delta\nu = \delta\nu * (1 + frequency\ shift * 1e - 06)$ . For Blanche,  $\delta\nu$  was set to 0.2410564839.
- If wanting to apply air-mass independent and dependent correction factors to the retrieved data, multiple files must be modified.
  - If wanting to implement air-mass **dependent** correction factors, modify the existing values for each gas in the following file as desired:  
 .../ggg2020/tccon/corrections\_airmass\_postavg.dat
    - \* In order for GGG to properly implement the air-mass dependent correction factors, the latest version of EGI should be installed. Software versions predating 2022 may require code modifications to properly apply these correction factors to EM27/SUN data.
  - If wanting to implement air-mass **independent** correction factors, modify the values provided for each gas in the following file as desired:  
 .../ggg2020/tccon/corrections\_insitu\_postavg.dat
- Once these corrections have been made, the retrieval algorithm can be run using the EGI wrapper following the instructions on the TCCON website, provided above.