AQRP 19-025

Draft Final Report:
Apportioning the Sources of Ozone Production during the San Antonio Field Study

Prepared by

Tara Yacovitch, Rob Roscioli, Scott Herndon, Megan Claflin, Brian Lerner, Manjula Canagaratna and Jordan Krechmer

In Association with

W. Berk Knighton, Montana State University

QA Requirements: Audits of Data Quality: 10% Required

Report of Findings from Audits of Data Quality in Final Report

Abstract

Ozone high up in the stratosphere is protective against UV rays, but when it is present at ground-level, it is a pollutant that can cause shortness of breath and other respiratory health problems. With new federal ozone standards in effect, it is more important than ever to understand the causes of ozone in and around San Antonio.

Ozone is formed when volatile organic hydrocarbons (VOCs) react with the nitrogen oxides (NOx, the primary component in smog). A wide variety of VOCs are present in the air around cities such as San Antonio; they stem from sources as varied as vehicle exhaust, oil and gas extraction, and trees and vegetation. The OH radical, a short-lived reactive species, plays a crucial role in ozone photochemistry. In this project, we investigate which sources contribute to the OH reactivity in and around San Antonio, and in what quantities.

Raw data from the 2017 San Antonio Field Study (SAFS) is examined and analyzed in full to identify characteristic sets of VOCs associated with different source types. Positive Matrix Factorization, a mathematical technique, is used to identify and interpret co-correlating species among the hundreds of varying signals in the raw high-resolution datasets. Computer modeling of air transport identifies the broad land coverage types where the measured air originated. An ozone formation computer model, in which individual source categories can be turned on, off, or varied, is used to understand how each source type contributes to ozone formation in and around San Antonio.

Acknowledgements

The preparation of this report was funded by a grant from the Texas Air Quality Research Program (AQRP) at The University of Texas at Austin through the Texas Emission Reduction Program (TERP) and the Texas Commission on Environmental Quality (TCEQ). The findings, opinions and conclusions are the work of the author(s) and do not necessarily represent findings, opinions, or conclusions of the AQRP or the TCEQ.

Executive Summary

Ozone is formed when NOx and VOCs react in the presence of sunlight. In order to better understand ozone formation in the SAFS study area, this project looks at the VOC half of the equation. VOCs reacting with OH radical kick off a series of ozone-generating reactions. Thus, by tracing, quantifying and apportioning the formation of OH, we can better understand and quantify the causes of ozone formation in and around San Antonio.

We explored and analyzed the entirety of high-resolution datasets from the I-CIMS, PTR-ToF and GC-EI-ToF, three instruments that targeted different classes of atmospheric compounds. Previous efforts had only skimmed the surface, hand-picking species to report. Specifically, this analysis effort had three main priorities:

- Priority A: "How much do the different source categories contribute to the peak concentrations [of ozone] observed?" and "Can contributing source categories be identified and quantified from measurement data alone"?
- Priority B: "Collaborate and share data with other SAFS participants"
- Priority C: "How much do upwind source categories contribute [to ozone formation in San Antonio]"

All three priorities were addressed to varying degrees during this work. By coupling the chemical information into 0D box modelling, Priority A was addressed and partially answered, albeit not completely. During the course of the analysis Aerodyne communicated initial results in monthly updates, and has archived the data, fulfilling the goal of Priority B. Finally, by combining Positive Matrix Factorization (described below) with atmospheric transport modeling, footprint analysis, and chemical intuition Priority C was addressed and largely answered.

Positive Matrix Factorization, a mathematical technique, was crucial in identifying co-varying species among the hundreds of individual high-resolution signals (Figure 1). It was also run in a way that combined data from multiple instruments, and yielded interpretable factors, like diurnal trends due to isoprene, and location-specific signals due to oil and gas impact. This technique results in "factors" with distinctive time trends (Figure 1, left), and each factor has a chemical fingerprint (Figure 1, right) made up of individual chemical species. Individual chemical species can and do show up in more than one factor, which reflects the real complexity of sources in this campaign. For example, ethane (5th species from the left, Figure 1) shows up with significant intensity in the "oil and gas emissions" factor (blue), in "urban VOC oxidation products" (purple) and "anthropogenic (auto) emissions"; it has no contribution or much lower contributions to the other factors, particularly the biogenic factors (grey, red).

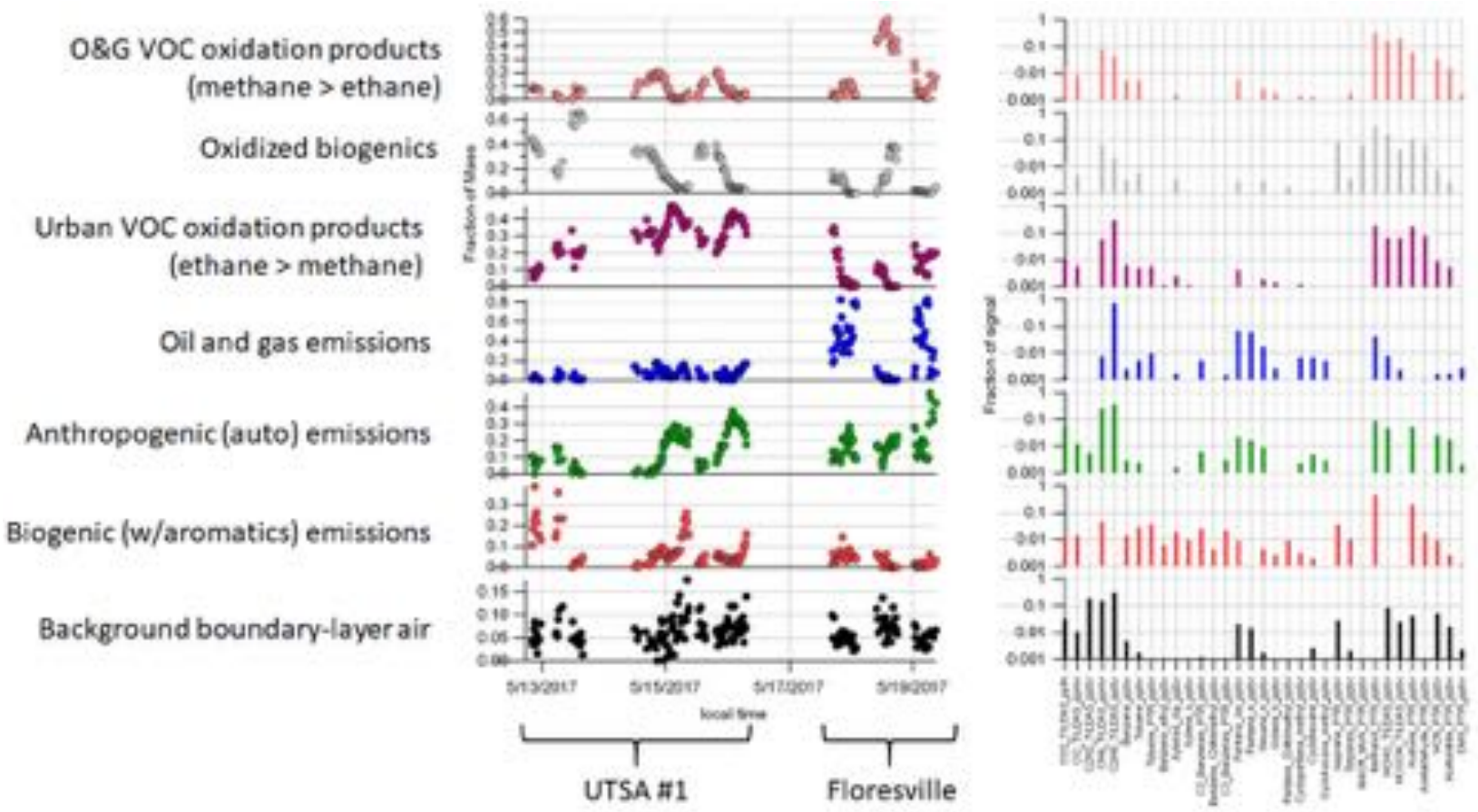

Figure 1. PMF results of combined datasets, with nominal description of each factor. Left: Individual factor contribution to total chemical burden. Right: chemical signature for each factor. The chemical signatures are shown as bar graphs, with each bar corresponding to a measured species. TILDAS refers to gas-phase species measured via Tunable Infrared Laser Direct Absorption Spectroscopy (with the exception of CO_2, which was measured via non-dispersive infrared spectroscopy (NDIR)); PTR refers to species measured via Proton Transfer Reaction Time of Flight Mass Spectrometry (PTR-ToF); all other species shown were measured via Gas Chromatography Electron Impact Time of Flight Mass Spectrometry (GC-EI-ToF or GC-ToF for short).

One important conclusion of this report is that biogenic sources of VOCs dominate OH reactivity in the SAFS study area. This is visible even in the simplest evaluation of measured data, shown in Figure 2. In this figure the campaign average concentrations of measured species, weighted by the number of carbons, are multiplied by their rate constants with OH, and normalized, yielding a ranking of OH reactivity per VOC.

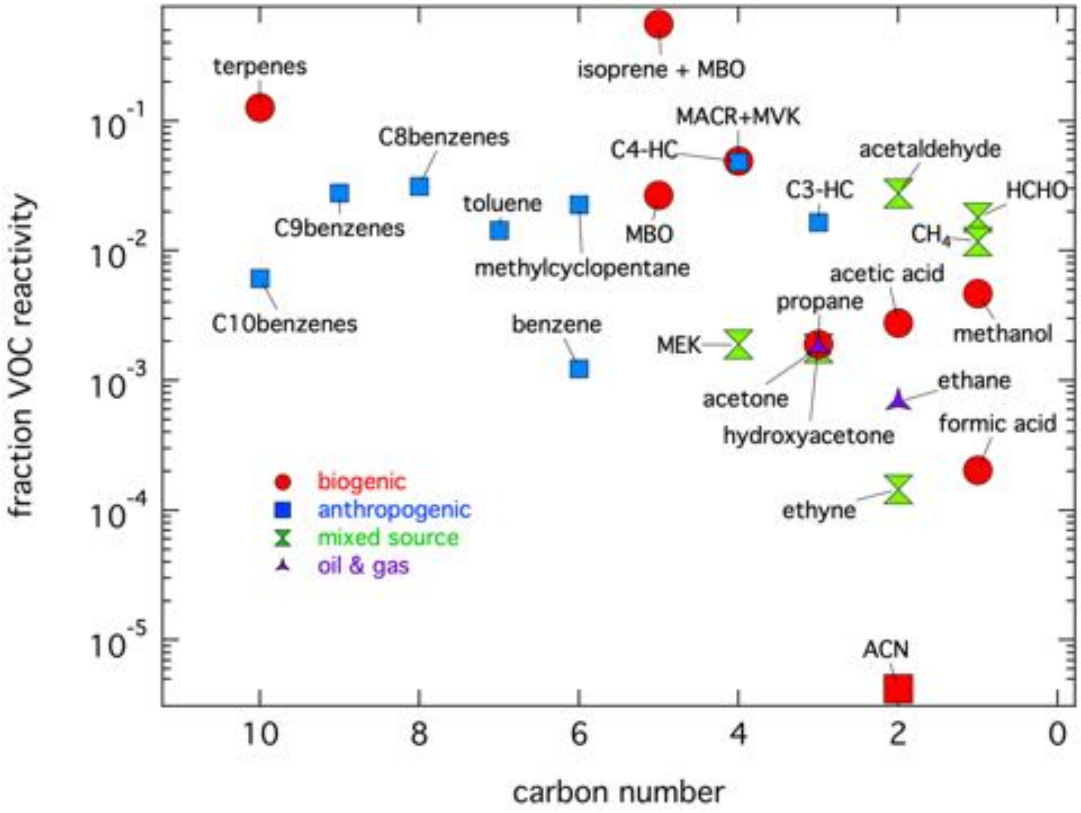

Figure 2. Relative VOC reactivity for selected VOCs.

An additional hypothesis of the original 2017 measurement campaign was that high-resolution analysis would allow identification of signature compounds linked to a specific source. The idea was that VOC intermediates can associate with NOx and leave signature compounds that still retain information about the source. Although parent VOC measurements are very useful, they do not directly indicate the history of ozone formation or its sensitivity to NOx and VOC emissions. Quantification of unique markers of oxidation, such as speciated oxygenated VOCs and alkyl nitrates could present an intriguing way to apportion ozone formation since they are created at the same time as the ozone is being produced.

The identification of such signature compounds was ultimately not successful, since the atmospheric processing results in fragmentation as well as functionalization [*Kroll et al.*, 2011]. The fragmentation process masks unique information about the parent's elemental formula. In principle, extensive additional laboratory experiments could guide

future field measurements and analysis to experimentally identify additional intermediates, or sets of intermediates, for a more direct attribution to sources.

To get around this lack of identifiable chemical intermediates linked to a specific source, a 0D chemical model was used to explicitly track photostationary state OH-formation and important intermediate radicals for given a set of measured constraints.

Table of Contents

Table of Figures

Table of Tables

Introduction

The Environmental Protection Agency (EPA) designates ozone as a criteria pollutant, with new 2015 standards set at 70 ppb (3-year average of the annual 4th maximum 8-hour daily maximum). The ozone monitor network in San Antonio indicates that the city is on the verge of being out of compliance for federal standards. The EPA monitor at Camp Bullis (C58), shows 14 maximum daily 8-hour averages exceeding 70 ppb since 2015. These exceed the 70 ppb EPA standard that came into effect in 2015, with enforcement of the standard expected to begin soon. The mitigation strategy is likely to involve curtailing emissions of ozone-producing volatile organic compounds (VOCs) emitted either in the city (for those VOCs that are oxidized rapidly) or further upwind in regions such as the Eagle Ford or in the refineries at Corpus Christi.

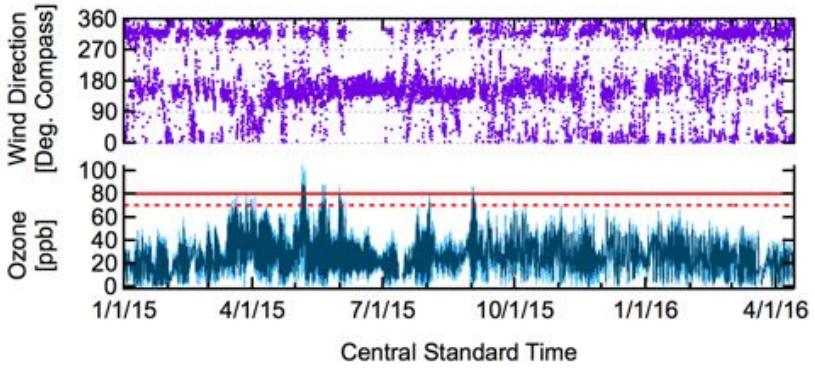

Figure 3. 1-hour ozone measurements from the Camp Bullis monitoring station are shown (pale blue) alongside 8-hour averages (dark blue) and wind direction (purple). EPA ozone limits (red) are shown at 80 ppb (solid, current standard) 70 ppb (dotted, 2015 standard, yet to be enforced).

Ozone is produced from sunlight, NOx and VOCs. VOCs are emitted from many different sources. They can be biogenic in origin, like isoprene and other terpenes; they can be oil & gas associated such as propane or various aromatic compounds. Numerous other anthropogenic compounds participate in ozone production chemistry.

During the SAFS field campaign, the Aerodyne Mobile Laboratory was operated at various sites between Corpus Christi and San Antonio. The strategy was to move the portable supersite to a location that is forecast (based on meteorology) to have needed characteristics to meet the science objectives for 1-3 days into the future. Although forecast conditions are always changing, the goal was to fully characterize the incoming Gulf air: through the refinery complex, as it passes through the Eagle Ford, into, and out of San Antonio. With the assistance of collaborators and the TCEQ, specific locations and logistics were chosen during the planning stages.

The Aerodyne Mobile Laboratory (AML) is a well-tested and extremely suitable measurement platform for the goals of the proposed study. Previous deployments have included measurements in urban polluted areas such as Mexico City during the 2006 MaxMEX/MILAGRO campaign [*Herndon et al.*, 2008; *Wood et al.*, 2009], the 2009 Queens, NY, study [*Massoli et al.*, 2012], or for more specific sources such an aircraft emissions [*Santoni et al.*, 2011] or oil and gas extraction [*Yacovitch et al.*, 2015].

Research and commercial instruments are installed into the AML to collect data while in motion for plume characterization, area mapping or portable deployment for photochemistry and transport experiments. Real-time monitoring of both gas-phase and particulate species is the key feature of the AML.

Most instruments proposed for this ozone study, including the TILDAS [*McManus et al.*, 2008; *Nelson et al.*, 2006] and the SP-AMS [*Onasch et al.*, 2012], have been successfully deployed by Aerodyne researchers and others in numerous field campaigns. Further descriptions of the most novel (I-CIMS-HRTOF) and new additions to the analytical payload are described below. The instrument manifest includes all of the combustion tracers (CO_2, CO, NO, NO_2, SO_2) the light alkanes (CH_4, C_2H_6, C_3H_8) and alkenes (C_2H_4, C_3H_6), and oxygenated and aromatic VOCs. The comprehensive table of the instrument payload can be found in the final report for AQRP project 17-053 at: https://herndon.homeunix.net/owncloud/index.php/s/DykTvaXkf8oeHl1

Reported here are the results of in-depth analysis of field data collected during the 2017 San Antonio Field Study (SAFS) in order to understand and apportion ozone formation in the studied area.

Reprocessing of the dataset with high-resolution methods allows us to identify previously unreported chemical species, while improving detection limits on existing results. With high-resolution data in hand, three core analysis pathways are followed that directly address the 2018-2019 priority research areas of the Texas Air Quality Research Program (AQRP).

- Priority A: “How much do the different source categories contribute to the peak concentrations [of ozone] observed?” and “Can contributing source categories be identified and quantified from measurement data alone”?
- Priority B: “Collaborate and share data with other SAFS participants”)
- Priority C: “How much do upwind source categories contribute [to ozone formation in San Antonio]”

Photochemical box modeling uses measured VOCs to constrain and understand the formation of ozone. Positive Matrix Factorization (PMF) is a mathematical technique that groups species with similar sources together, based on their time series. PMF is used to identify characteristic source categories, or “factors”, in the data. For example, PMF groups VOCs with the same diurnal profile, helping us to assign them as emissions from trees (Priority A). HYSPLIT footprint analysis draws borders on a map indicating where the sampled airmass likely originated (Priority C), and allows us to divide this footprint into categories corresponding to different land-use categories. Finally, experimental data is used to constrain and inform a 1-D box model that simulates ozone formation with full chemical specificity, allowing us to apportion OH reactivity in the SAFS study area (Priority A). **Together, this research will allow for a sector-based apportionment of OH reactivity in the SAFS area**

This report first outlines the methods used to analyze data from three core high-resolution (HR) mass-spectrometry-based instruments: Iodide Chemical Ionization Mass Spectrometer (I-CIMS), Proton-Transfer Reaction Time-of-Flight Mass Spectrometer (PTR-ToF) and Gas-Chromatography Time-of-Flight Mass Spectrometer (GC-ToF).

These methods include high-resolution analysis, and PMF. Descriptions of the methods used to calculate and analyze back-trajectory simulations are included, as is a description of the 0D model framework.

The Results and Discussion section first delves into an analysis of Species of Atmospheric Interest: the three instrumental dataset are querried, with the addition of some data from the CAMS station in Floresville, with the aim of identifying co-varying species and interpreting their identities or sources. Then, in the section Apportioning Ozone to Source Categories, model apportionment results are presented, as are results from a airmass source region analysis.

Deliverables for this project are listed below.

- Time series of all high-resolution fits (in order of preference: absolute calibrated data in ppb when available; estimated calibration factor based on chemical identification;[*Sekimoto et al.*, 2017] relative intensities.)
- Time series data descriptions including calibration factor or scaling factor applied.
- Time series of total OH reactivity (s^{-1})

Methods

This section describes relevant methodologies and advances used to produce the datasets and conclusions in this report.

Appendix A: High-Resolution Dataset Description lists and describes the data produced as part of this project.

Appendix B: Audits of Data Quality shows the data quality audit results (10% required).

High-Resolution fitting

Three of the instruments used to collect data during SAFS relied upon high-resolution mass-spectrometry for detection of species of atmospheric interest: PTR-ToF, CIMS and GC-ToF. Normally, a curated set of mass windows are fit in order to quantify select chemical species. The standard procedure divides the mass spectrum into 1 atomic mass unit (amu) sized windows, and integrates all signals within these bounds. As part of this analysis project, the full high-resolution mass spectra from each of these three instruments was leveraged and analyzed in order to reveal other signals of potential atmospheric interest.

Pseudo-HR strategy

Traditional approaches to analyzing mass spectral data rely on precise knowledge of the chemical formula of a species of interest, and its potential fragmentation products in

the instrument. This poses a problem when considering a full high-resolution dataset, since it is often possible to resolve several hundred individual signals, some of which occur as shoulders on top of larger neighboring peaks. Specifying and identifying every peak that needs fitting manually would be much too time consuming. Instead we leverage a peak-finding algorithm in the analysis software "TofWare".

The peak-finding algorithm was used to supplement the original peak lists manually developed during the SAFS campaign; the final peak lists thus consist of a number of named ions (i.e. with known chemical formula) plus a large number of unknown ions (i.e. with a mass precisely determined by the algorithm, but with no chemical identification). We term this peak list "pseudo-HR" since it includes both known and unknown ions.

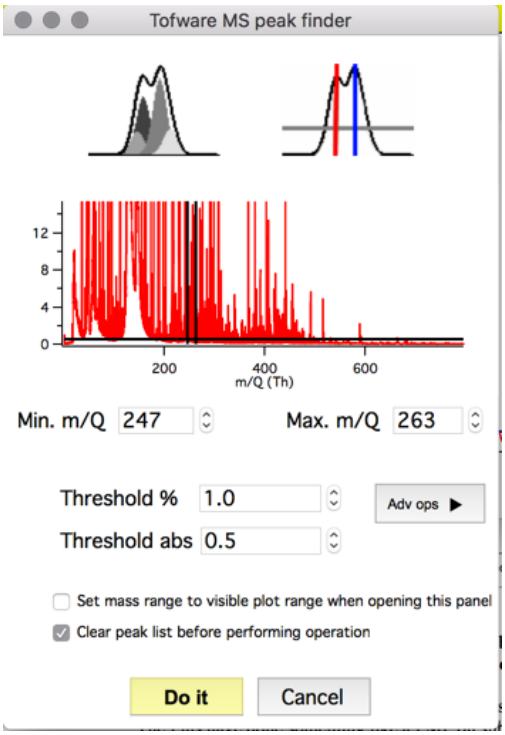

Figure 4. The Tofware MS peak finder window enables an automated fitting algorithm that will fit complex peaks with multiple individual masses.

Several details are important for ensuring a high-quality pseudo-HR peak list:

1. Mass calibrations must be properly executed.
2. Known ions (from previously developed peak lists) should be defined first.
3. A representative mass spectrum should be chosen
4. The peak-finding algorithm should be tuned to maximize the goodness-of-fit with the minimum number of peaks.

The reference spectrum chosen was determined by taking the month-long dataset and selecting or averaging spectra throughout the entire duration. This is especially important for SAFS, where three different locations were sampled, and a range of meteorological conditions were encountered.

As an example, the figure below shows the representative mass spectrum used to determine the pseudo-HR peak list for the PTR-ToF instrument. Every tenth mass spectrum was selected and averaged.

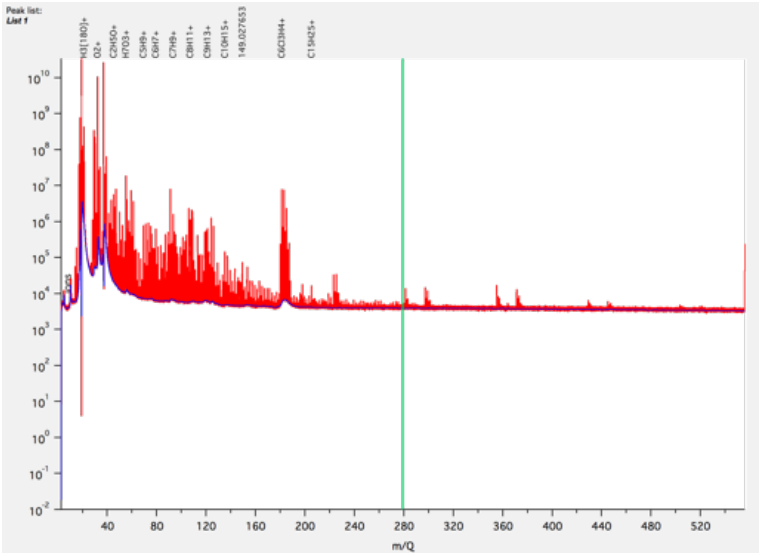

Figure 5. Representative mass spectrum used in determining PTR-ToF pseudo-HR peak lists. The red trace shows the signal plotting signal intensity versus mass-to-charge (m/Q).

PTR-ToF Ion Identification

Identification of compounds from the mass spectral data acquired by the Proton Transfr Reaction Time of Flight Mass Spectrometer (PTR-ToF-MS or PTR-ToF for short) starts with the determination of the elemental composition of the ion, which is made possible due the high resolution and mass accuracy of the time-of-flight mass spectrometer. The next step is to determine whether the ion is a protonated molecule (MH^+) or is a fragmentation product from a larger molecule. One generally starts with the premise that ions are protonated molecules and then evaluates how likely that compounds having that elemental composition would be present in the ambient atmosphere. Compilations such as Yuan et al. [2017] present comprehensive discussions on the assignment of specific compounds to selected ions and provide the core source for the identifications provided within. It must be recognized that any measurement technique based on a single metric, such as ion composition, can never be totally definitive. Over the past several decades the PTR-MS systems have gone through numerous intercomparison studies with other established techniques. These studies have provided a sense of validation for what compound/ion pairs can be reliably used for quantification. These validation studies, however, are often source dependent, and the end user must remain mindful that signal identities can change depending on the sampling environment.

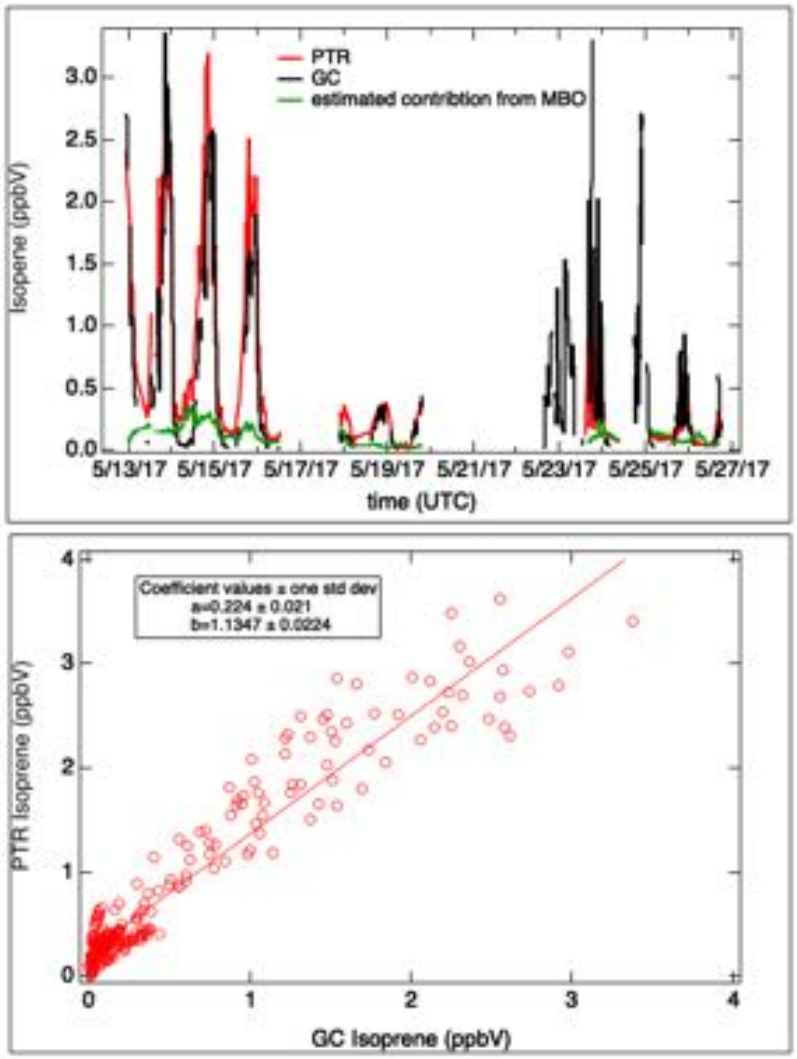

Figure 6. Comparison of the PTR-ToF-MS and GC-ToF-MS measurement of isoprene during the 2017 SAFS.

Biogenic emissions, particularly isoprene, are an important contributor to ozone formation in the San Antonio region. The ion, $C_5H_9^+$, has been assigned as isoprene + MBO (methylbutenol) even though the vast majority of the signal originates from isoprene. Both compounds are produced via the photosynthetic pathway, with oak trees being the primary source of isoprene and conifers for methylbutenol. Figure 6 provides a side-by-side comparison of the isoprene measurements reported by the PTR-ToF-MS and the GC-ToF-MS as well as an estimate of the MBO interference. The PTR-ToF-MS measurements have been averaged onto the GC trapping time base so the two measurements are directly comparable. Overall, the two data sets are in reasonable agreement with the PTR-ToF-MS measurements being slightly higher (13%). The MBO interference is estimated from a branching ratio determination made on a quadrupole PTR-MS instrument operated under the same drift tube reaction chamber conditions where it observed that MBO produces twice as much m/z 69 relative to the protonated molecule (m/z 87). Thus the MBO contribution illustrated represents the reported $C_5H_{11}O^+$ mixing ratio times two. While there is some contribution of MBO to the

isoprene signal it is not judged to be significant. We note that there are some time periods where the PTR-ToF-MS and GC-ToF-MS reported isoprene deviate and have no definitive explanation for this. Both instruments were challenged by elevated operating temperatures, and it is plausible that trapping efficiency of isoprene by the GC-ToF-MS may have been affected and contributed to the offset. Nighttime measurements are much more uncertain and in the absence of corroborating data (GC-MS) suggest that the $C_5H_9^+$ ion has an interference from another source. Several compounds such as the cyclohexanes [*Gueneron et al.*, 2015] and the sequiterpenes [*Kari et al.*, 2018] are reported to produce $C_5H_9^+$ fragment ions and may be responsible for the elevated signal reported as isoprene observed at night.

Ions that have been assigned names indicate that multiple criteria have been met including signal purity, known presence and temporal behavior related to other source metrics.

Appendix A contains a complete listing of the traces reported from the PTR-ToF-MS, and descriptions of the basis of their identification.

GC-ToF Data Processing

Gas chromatography (GC) is a technique used for the analysis and quantification of VOCs. The compounds of interest are injected onto an analytical column and separated based on their structure using a temperate gradient. The elution flow is then fed into a Time-of-Flight Mass Spectrometer (ToF-MS) for high resolution (HR) detection.

The HR mass calibration, conducted using typical methods for online techniques, was showing severe deviations when applied to the GC data set with the mass accuracy diverging to +/- 200 ppm. This is attributed to the combination of two effects 1) the temperature ramp that the GC eluent was exposed to before it enters the detector and 2) the GC separating common ambient ions used for mass calibration as resolved chromatographic peaks. The first effect appears to be an artifact of using ToF-MS as the detection method. To separate compounds using GC, the column is heated which causes the elution flow temperature to increase throughout the GC cycle. This change in eluent temperature correlates with changes in the mass calibration. Physically this can be ascribed to the known temperature dependence of the ToF-MS, where thermal expansion causes changes to the ion path length. This change in path length then results in mass calibration drift. The solution to both of these artifacts was to find ions that are being constantly fed into the detector, with sufficient signal, throughout the GC temperature ramp. It was found was that there are four classes of ions that can be appropriately utilized for HR mass calibration of GC data:

1) Carrier gas contaminants (e.g. Ar+, CO_2+)
2) GC column bleed (siloxane compounds from the column active phase)
3) Pulsed compounds from the EI-ToF calibration solution (applied at the beginning and end of each chromatogram)
4) Ions emitted from the EI filament assembly (ReO_3+, ReO_4+)

These four categories provide ions that are exposed to the entire GC path and some emitted directly inside of the detector. When this method is applied to the data, with

averaging every 1.5 minutes, mass accuracy of +/- 10 ppm can generally be achieved across the chromatogram (Figure 7).

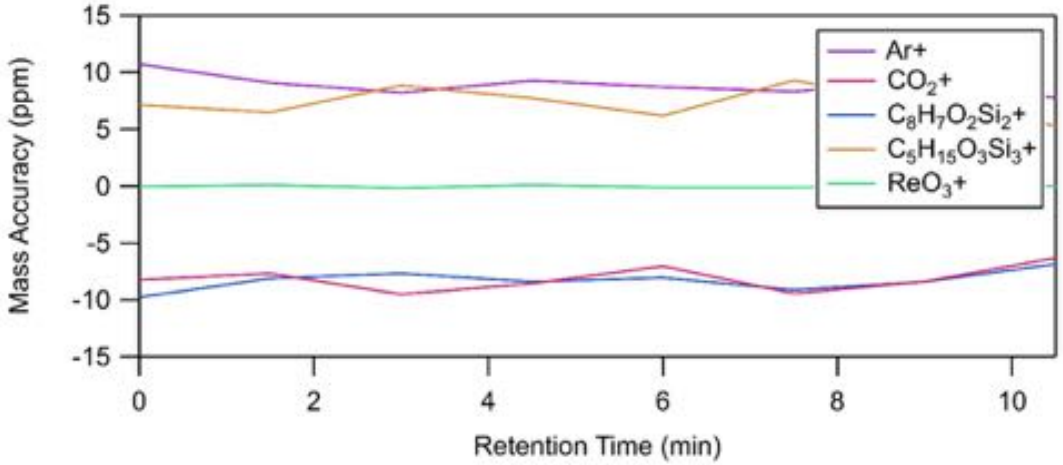

Figure 7. An example of mass accuracies achieved across a chromatogram using mass calibration ions found in the carrier gas and generated by column bleed and the detector filament.

Another instrument function that needed to be accounted for to obtain sufficient HR fitting was how turning the EI filament on/off was handled. Turning on/off the filament for EI-detection is a common practice used to extend the life of the filament, however, in the file this loss of signal (Figure 8) causes the mass calibration to diverge to unacceptable mass accuracies. To handle these signal changes, modifications to the HR fitting software, Tofware, had to be made. With this new version of the software, GC start and end times can be defined by the user, once these parameters are defined the mass calibration will only operate within those bounds.

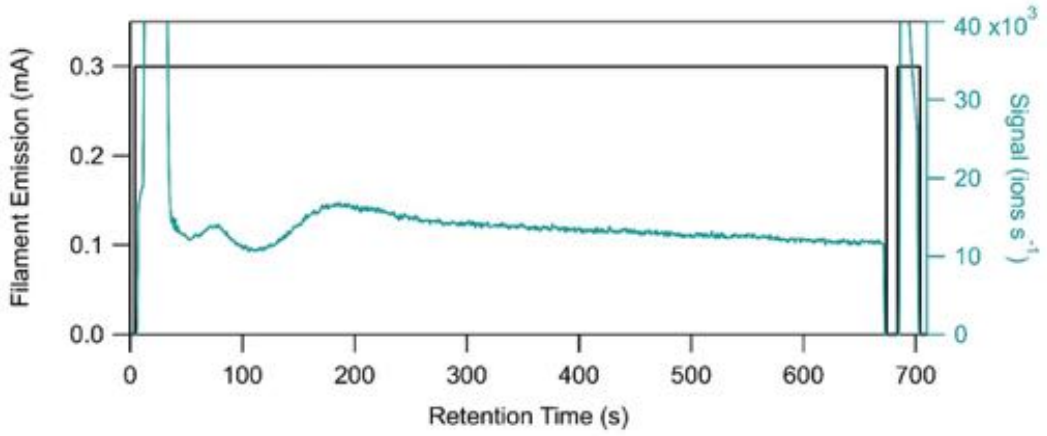

Figure 8. An example of the filament settings during each chromatographic run. The filament is turned on (0.3 mA) 5 and 685 s into the chromatogram and off (0 mA) at 675 and 705 s. The associated behavior of one of the mass calibration ions (Ar+) is shown, with the loss of signal seen when the filament is off.

HR analysis of GC-TOF data

The gas chromatography (GC) detector used during the SAFS campaign is a high-resolution electron impact time of flight (EI-ToF) mass spectrometer, which allows for the reprocessing of collected data and conversion from unit-mass resolution (UMR) to high-resolution (HR) time series. These HR time series are constructed by fitting a set of peaks with a common peak shape to the overall high-resolution mass response.

An example is shown in Figure 9 for a chromatogram collected during lab testing before the SAFS study. In the right panel, a subsection of the total chromatogram is shown (565 – 605 sec of elution), with the UMR signal at m/z = 105 shown in red. This trace would indicate four significant chromatographic peaks, including 1,2,4-trimethylbenzene which elutes at retention time 584 sec. This section of data can be reprocessed using TofWare mass spectral analysis software, as shown in the left panel. The red trace of the total signal at m/z = 105 has a pronounced shoulder on the left side, indicating a second ion contribution to the UMR signal. We are able to fit the total signal with two ion peaks, $C_8H_9^+$ (aromatic fragment, exact mass = 105.070) and $C_7H_5O^+$ (oxidized aromatic peak, exact mass = 105.034). The time series for these ions are shown in the right panel, as the black ($C_8H_9^+$) and yellow traces ($C_7H_5O^+$). We can now see that for the elution time at 584 sec, there are actually two contributions to the m/z = 105 signal, from both 1,2,4-trimethylbenzene and also from benzaldehyde. Both of these compounds fragment in the EI ionizer to produce a significant fraction of m/z = 105 ion, which would appear as a single peak at the Aerodyne Research, Inc. GC resolution and separation method. But by separating these signals with the high-resolution re-analysis, we can distinguish these compounds. During the SAFS campaign, we fortuitously ran a modified elution method that allowed for the separation of 1,2,4-trimethylbenzene from benzaldehyde, as can be seen in Figure 10. Note the later elution time allows for better separation of compounds in this volatility range.

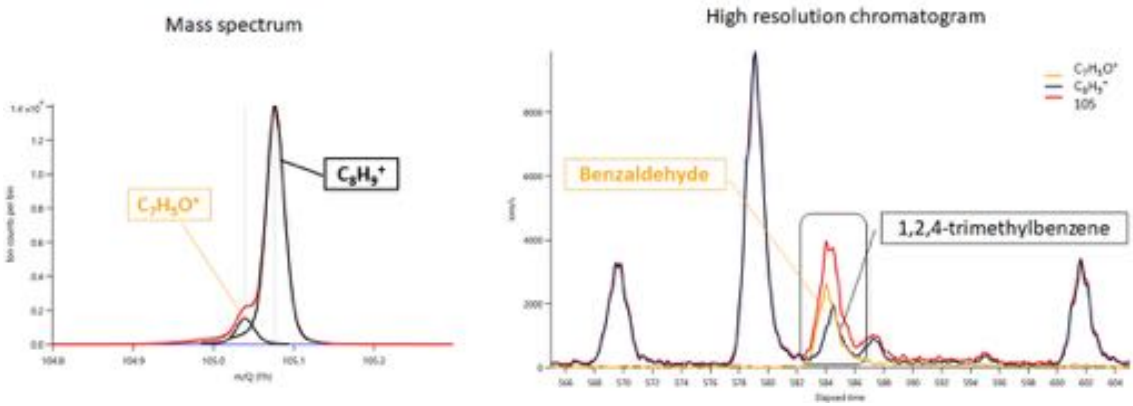

Figure 9. Left panel: contributions from individual ions to total UMR signal at m/z = 105. Right panel: Example chromatogram during pre-campaign testing showing UMR and individual ion chromatographic peaks at retention time = 584 sec.

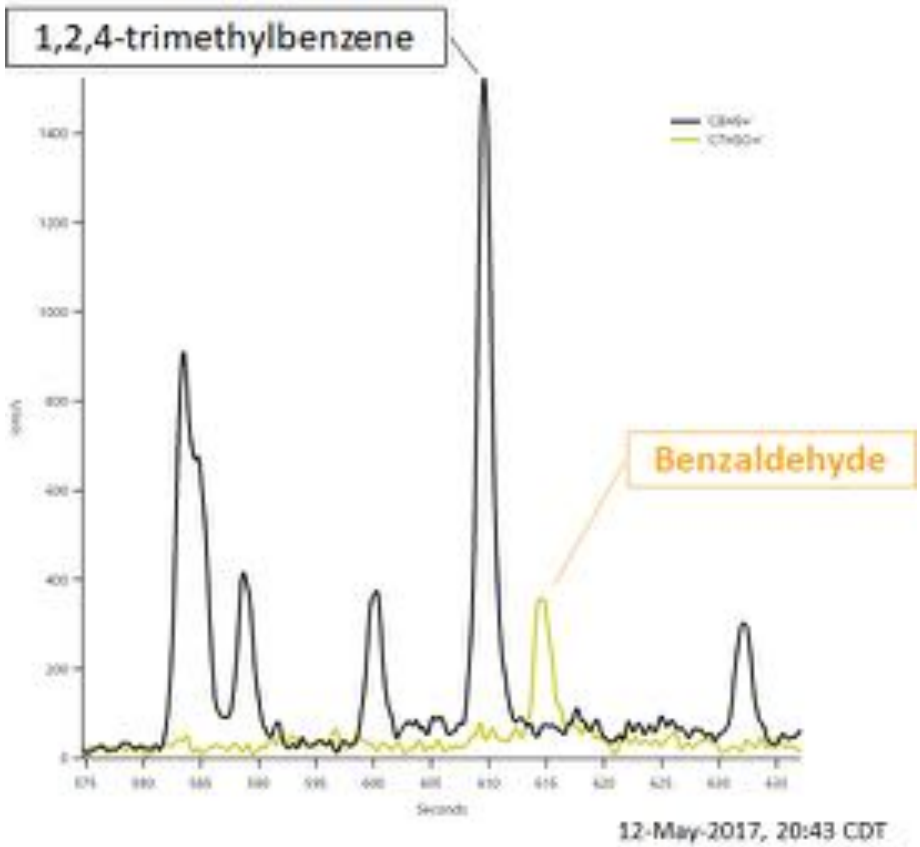

Figure 10. Chromatographic separation during SAFS for same range of compounds shown in Figure 9.

Examples of the re-analysis of the SAFS data with high-resolution mass spectral time series are shown in Figure 11, and a more extensive reprocessing is shown in Table 1. For the examples shown, we can see that there is good agreement between the high-resolution and unit-mass resolution data sets, including for 1,2,4-trimethylbenzene. In all cases except m/p-xylenes, the difference in slope between the high-resolution (HR) and unit mass resolution (UMR) data is less than the stated uncertainty for the measurement. But in all cases, the inter-comparison shows very good linearity ($R^2 > 0.98$) and therefore differences in peak area are accounted for in the calibration experiments.

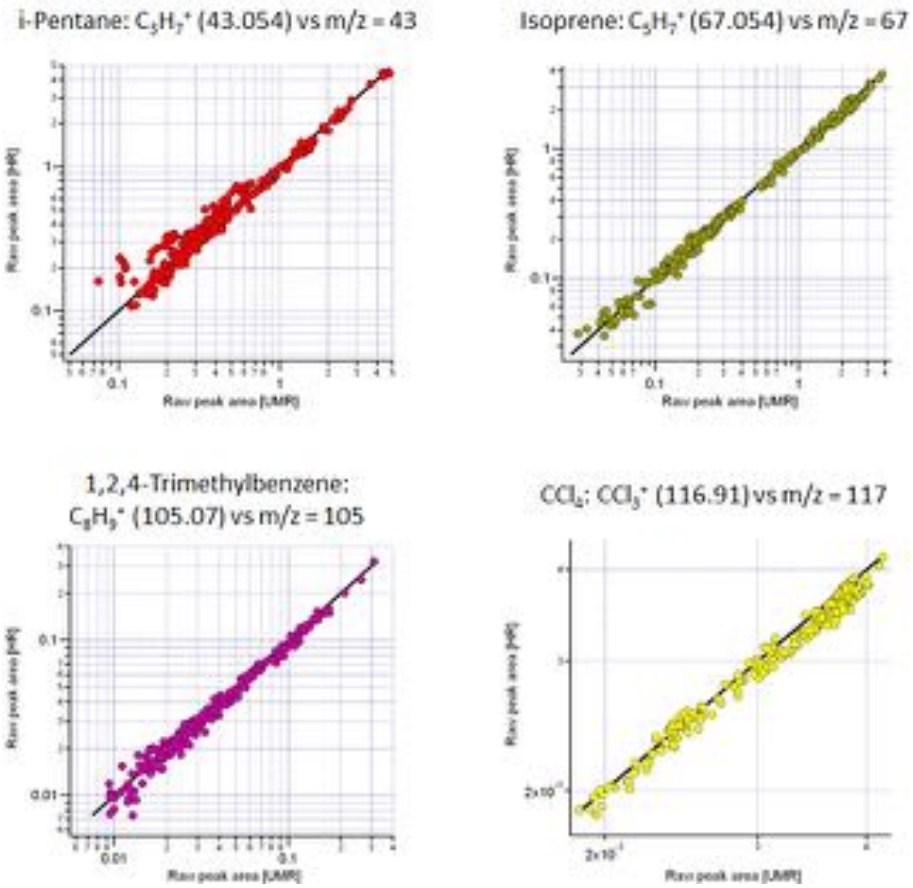

Figure 11. Inter-comparison of chromatographic peak areas derived from high-resolution (HR) and unit mass resolution (UMR) mass spectral data

Table 1. Comparison of chromatographic peak areas from HR vs UMR mass spectral peak fitting. Values in parentheses are 2σ uncertainties.

Species	Slope	Intercept	R^2
CCl4	0.94 (0.02)	0.012 (0.005)	0.983
i-Pentane	0.95 (0.01)	0.032 (0.012)	0.990
n-Pentane	1.01 (0.01)	0.081 (0.013)	0.995
Isoprene	0.95 (0.01)	0.012 (0.013)	0.994
Benzene	0.93 (0.01)	0.023 (0.008)	0.987
m/p-Xylenes	0.87 (0.01)	0.019 (0.003)	0.988
1,2,4-Trimethylbenzene	0.95 (0.01)	0.001 (0.001)	0.993

PMF overview and associated data quality metrics

PMF is a multivariate factor analysis technique developed by Paatero et al. ([*Paatero and Tapper*, 1994]; [*Paatero*, 1997]) to solve the bilinear factor model $x_{ij} = \Sigma_p g_{ip} f_{pj} + e_{ij}$ where x_{ij} are the measured values of j species in i samples and P are factors comprised of constant source profiles (f_j, mass spectra for CIMS data, or concentrations for combined dataset analysis) with varying contributions over the time period of the dataset (g_i, time

series). PMF does not require any a priori assumptions of either mass spectral or time profile ([*Lanz et al.*, 2007]; [*Ulbrich et al.*, 2009]). The bilinear PMF model is solved using the PMF2 algorithm, which computes the solution by minimizing the summed least squares error of the fit weighted with the error estimates of each data point. Solutions are also constrained to have non-negative values. The error weighting and non-negativity constraint result in more physically meaningful solutions that are easier to interpret compared to other receptor models. The PMF analysis yields factors which correspond to groups of ions that have similar time trends. The factor mass spectra provide the identities of the grouped ions that and the factor time trends provide the contributions of each group to the measured signal at any given time point. In this work the results of the PMF analysis are used to identify spectral signatures that can be used to follow the contributions of different sources and or photochemical processing regimes.

The PMF2 executable version 4.2 in the robust mode ([*Paatero*, 1997]) is used together with a custom software tool (PMF Evaluation Tool (PET) ([*Ulbrich et al.*, 2009]) in this analysis. Yan et al. (2016) have recently demonstrated application of PMF to unit mass resolution data obtained from a chemical ionization mass spectrometer (CIMS) while Massoli et al. (2018) have demonstrated PMF analysis of high-resolution CIMS data. This analysis utilizes the methods described in Massoli et al. (2018). The analysis is performed on a two-dimensional data matrix consisting of high resolution mass spectra (in units of Hz) measured as a function of time. Measured data (from a given instrument or from multiple instruments) from all sites are included together to yield a data matrix with dimensions of 297 time points and 2218 high resolution ions. The error estimates for each mass spectral data point (in Hz) are obtained from the calculated Poisson counting error for the measured ion counts. The errors of low S/N ions are downweighted according to Ulbrich et al. (2008). The PMF2 algorithm minimizes the quantity Q, which is expressed as

$$Q(E) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(E_{ij} / \sigma_{ij} \right)^2$$

where E_{ij} is an element of the residual matrix and σ_{ij} is the input error for each point. In the ideal case, when the fit is of good quality, E_{ij} and σ_{ij} are approximately equal, resulting in an overall ideal Q (Q_{exp}) that is approximately equal to the number of elements in the analyzed data matrix. Thus, the scaled residual (Q/Q_{exp}) that is obtained during the analysis is monitored in order to assess the quality of the fit. Ideal fits yield Q/Q_{exp} values of approximately 1. When the bilinear model is solved for different numbers of factors, different Q/Q_{exp} are obtained.

The "optimal" number of factors needed to solve the PMF problem is determined by monitoring the trend in Q/Qexp with the number of factors. Figure 12 shows an example of this for the PMF analysis that was performed for the I-CIMS dataset from this campaign, though this methodology is equally applicable to the combined dataset analysis. The I-CIMS PMF analysis was performed on a mass spectral data matrix that consisted of 297 time points and 2218 high resolution ions. As can be seen in this Figure, the Q/Qexp decreases sharply as the first few factors are added, but then plateaus around 5-7 factors. The fact that the Q/Qexp plot plateaus at these factor numbers indicates that addition of more factors does not improve the fit drastically.

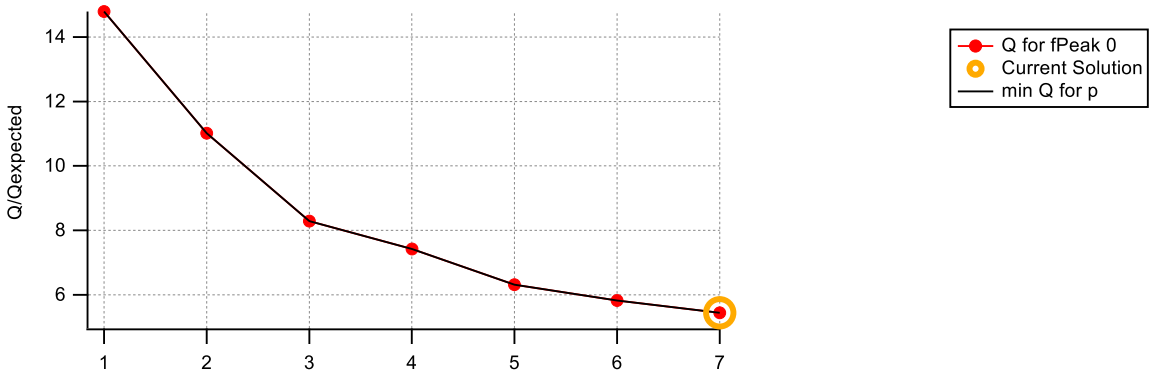

Figure 12. The Q/Qexp values observed as a function of number of factors in the PMF solution of the I-CIMS data.

In order to understand the detailed variation in Q/Qexp of the I-CIMS PMF analysis between 5-7 factors, the difference in Q/Qexp contributions from different high resolution ions and measurement time points is examined as a function of number of factors. Figure 12 shows that clear reductions in Q/Qexp are observed for many ions and most time points when PMF solutions with 1-3 factors are compared. Particularly large reductions in Q/Qexp are observed for the ions greater than m/z 500. The Q/Qexp does not drastically change for selected m/z or time points, however, once the 6th and 7th factors are added. Both the 6 and 7 factor solution yield similar total Q/Qexp values and have similar distributions of Q/Qexp contributions as a function of m/z and time. However, the 7 factor solution separates out two clear factors with daytime and night time diurnal cycles that are not separated out clearly in the 6 factor solution. As discussed later in the Results section of this report, the time trends of the two factors, their correlations with other measurements, and their signature ions are all interpretable in terms of previous laboratory and field studies. Therefore, based on the low Q/Qexp value and the interpretability of the solutions, the 7 factor solution is chosen as the "optimal" solution for this problem.

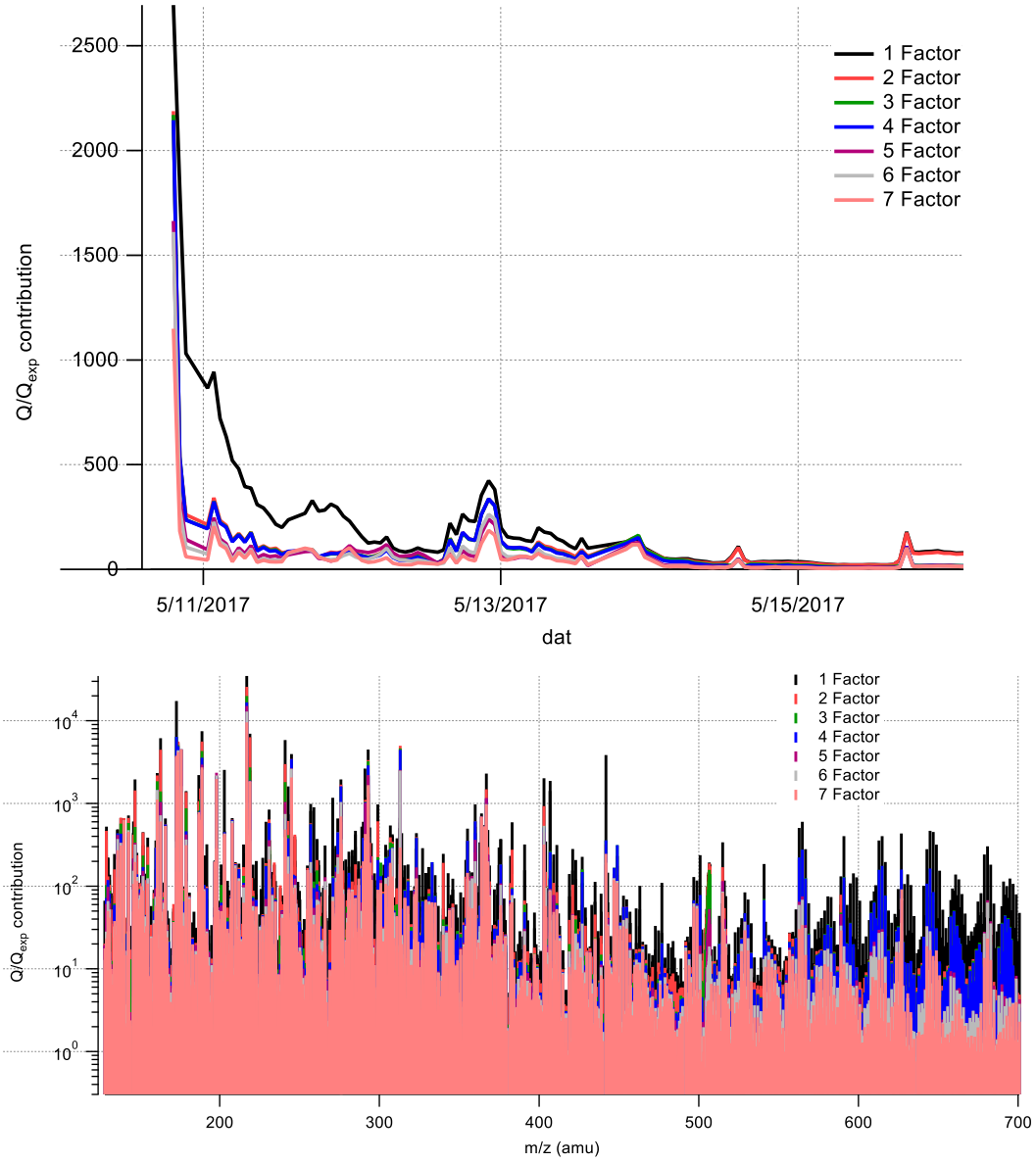

Figure 13. Q/Qexp calculated for the sum of all m/zs as a function of time (top panel); Q/Qexp calculated for all time points of the campaign as a function of m/z (bottom panel). The different colors correspond to PMF solutions obtained with different numbers of factors.

It is important to note that in Figure 12, the plateau in Q/Q_{exp} values is larger than the ideal value of 1. This could be due to prescribed errors that are too small. Spot checks for several high-resolution ions show that the prescribed errors appropriately capture the observed variability during time periods with constant composition, suggesting that the input error is reasonable.

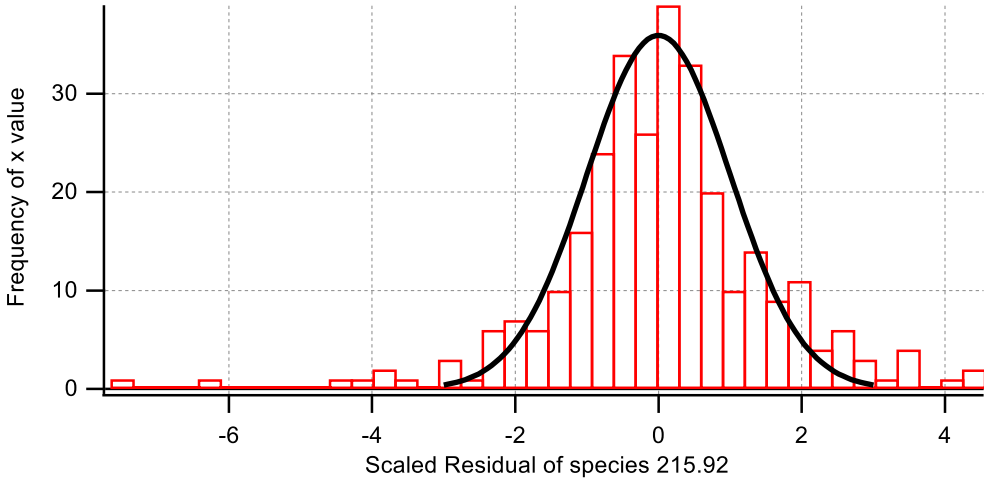

Figure 14. Scaled Residual of an individual ion. The black trace denotes a gaussian distribution width a half width that corresponds to the input error for this ion.

Another possible source of large Q/Qexp values is large residuals. Figure 14 shows the scaled residual, defined as the fit residual divided by the input error, of a single ion. The scaled residuals of this ion fall within a gaussian envelope (black line) with a half-width that corresponds to the input error for that ion, indicating that the ion is well fit. Figure 15 shows the distribution of the scaled residuals across several ions. This figure shows that the scaled residuals vary significantly between the different ions. Many ions are well fit, but several ions also have large scaled residuals. Since the ions that have higher scaled residuals do not correspond directly to any signature mass spectra measured in previous lab or field studies, it is not possible to unambiguously assign their residuals to a known source of variability. We also cannot discount the possibility that large residuals arise from a breakdown of the PMF assumption that observed I-CIMS spectral variability can be fully accounted for by a linear combination of non-varying, constant factor mass spectra.

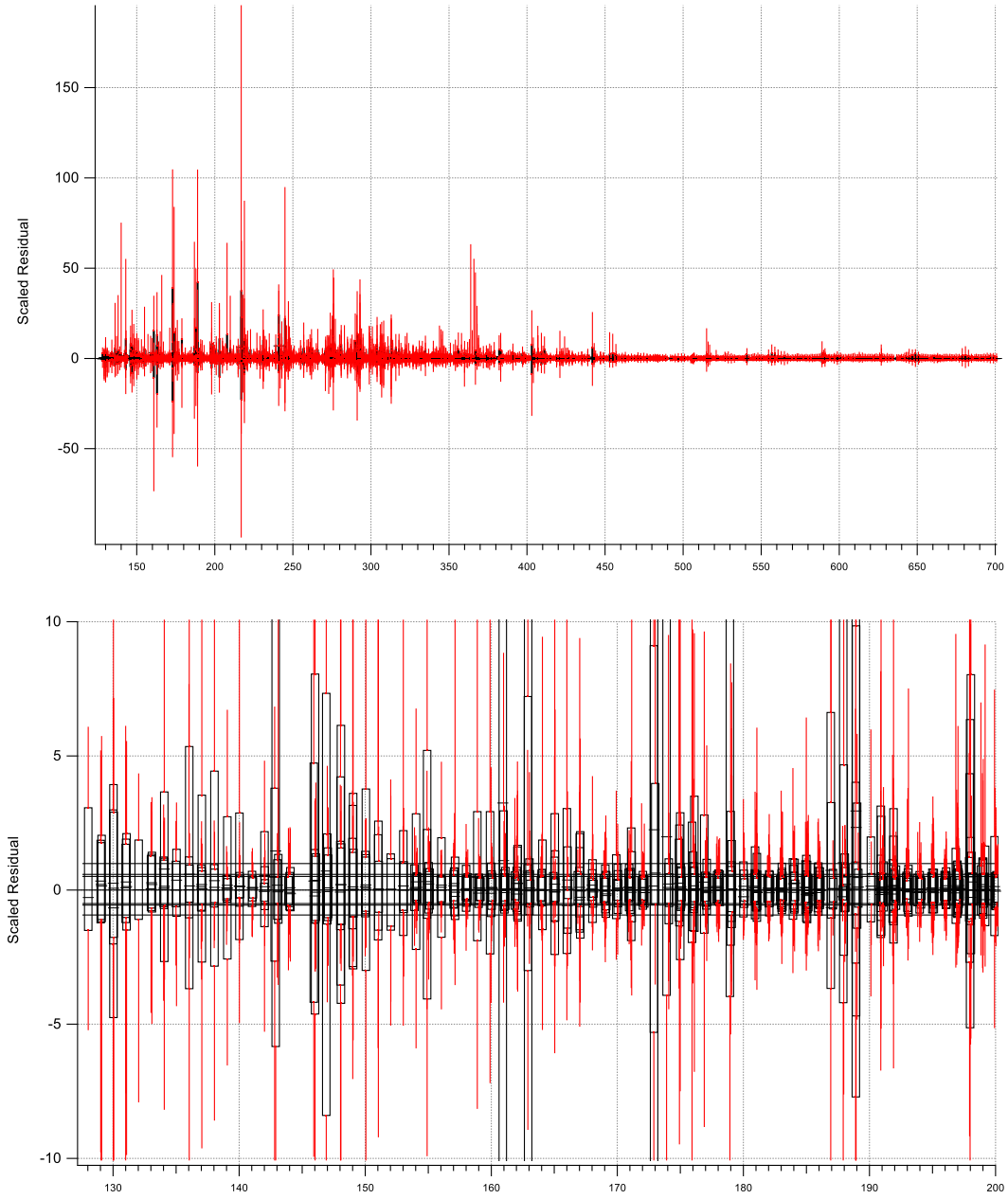

Figure 15. Distribution of scaled residuals across all ions in mass spectrum shows as a box and whisker plot. The bottom panel shows a blowup of a region of the mass spectrum shown in the top panel. The black boxes at each ion span the 25th and 75th percentiles and the red whiskers connect 2nd to 98th percentiles.

Methods for PMF analysis on GC-ToF datasets

Generally, PMF analysis is applied to measurement methods that use real-time sampling, where there are variations in signal with time. However, with GC-EI-ToF data there is an added time dimension so that within the data set there are 1) variations in mass spectral signal with GC retention time and 2) variation in chromatographic peak area with

sample time (Figure 16). This added time dimension complicates using PMF analysis for chromatography data where, if given the flexibility, PMF would associate a different factor for each chromatographic peak.

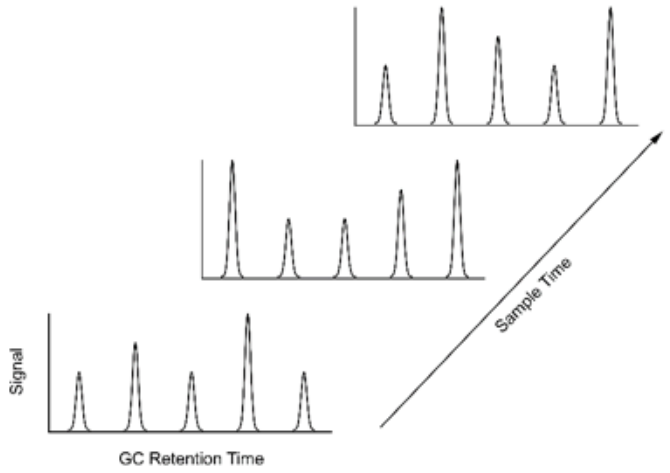

Figure 16. A visual depiction of the added time dimension, retention time, when applying PMF to GC data.

One way to avoid this complication is to put sections of the chromatogram into bins, and let PMF extract factors that describes changes in these bins over retention time (Figure 17). This binning technique retains some of the information gained through compound retention time, which is a function of compound volatility and functionality, while not allowing PMF to give one factor per chromatographic peak. While our goal was to apply this binning method to this data set, the intense temperature fluctuations in the environment surrounding the GC-EI-ToF caused significant retention time shifts in the chromatograms. Because the time shifts are non-linear across the chromatogram, where compounds that elute early experience more variability than those that elute later, this time shifting cannot be trivially corrected for samples across the campaign. And with retention times varying up to 60 s we could not trust that our prescribed bins would always be populated with the same chromatrographic peaks.

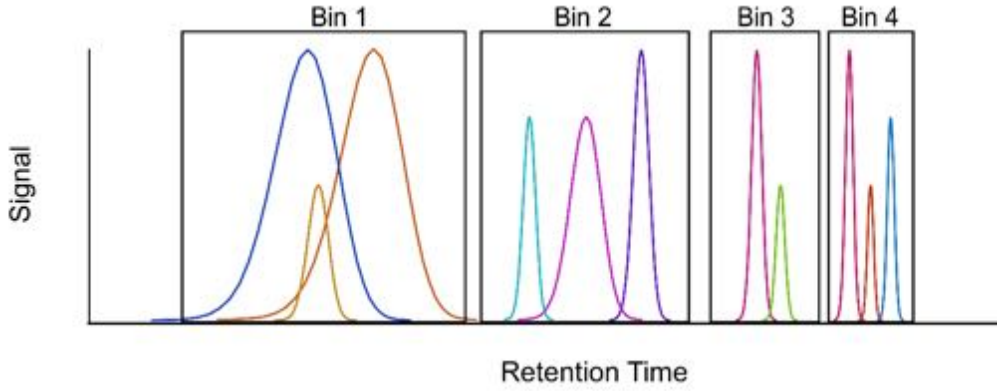

Figure 17. An example of how bins could be applied to divide chromatograms for PMF analysis.

Because of these discrepancies between chromatograms we have decided to remove the extra time dimension (GC retention time) and avoid issues associated with retention time shifting by stringing consecutive chromatograms together (Figure 18). We believe that this method will be able to distinguish between groups of compounds (different classes of hydrocarbons/oxygenates) and show correlations with time.

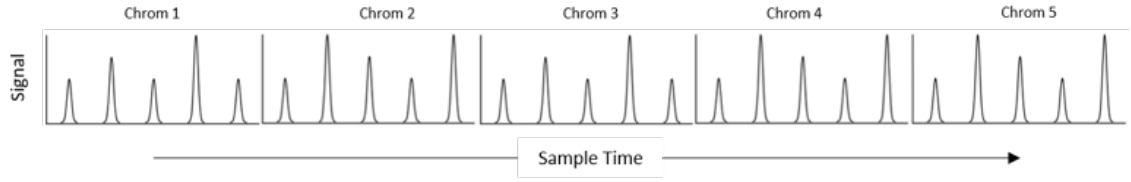

Figure 18. A schematic showing sequential chromatograms strung together to remove the added (retention) time dimension

However, another complication associated with GC data is that data collection speeds of ~ 5 Hz are required to obtain multiple data points across each chromatrographic peak (where peak widths can approach 1 – 2 s). This fast collection creates large data files and this file size limits how many chromatograms we can string together. At the moment we are able to string together 22 chromatograms (7.3 hours of sample time) while monitoring 215 masses before the matrix becomes too large for the PMF program to manage.

Because of these complications from both variations in the GC-ToF data due to changes in environmental conditions surrounding the instrument and limitations of the PMF program being applied to the large GC-ToF data files it was decided that time was better spend conducting traditional GC analysis. Looking ahead, we are seeking solutions to some of these issues that have been found by applying PMF to the GC-ToF data set so that this type of analysis can be used for future work. To address the file size issue, we are pursuing a 64-bit version of the PMF program, this upgrade would allow days of chromatograms to be strung together to look for diurnal changes. We have also recently become aware of a new auto-shift PMF method that is under development, this pre-treatment of GC data could potentially resolve the retention time shifting caused by changes in temperature surrounding the instrument.

0D Modeling overview and tests

The Dynamically Simple Model of Atmospheric Chemical Complexity (DSMACC) model was designed as a multipurpose chemical mechanism integrator with relevant submodules [*Emmerson and Evans*, 2009]. This model has been chosen for use in 0D photochemical simulations. The base GEOS-CHEM chemistry model was supplanted by the rebuilding the code to use the explicit Master Chemical Mechanism (MCM) version 3.3.

The model is operated in a mode where $j(NO_2)$ is constrained to the measured value. Quantified $j(NO_2)$ data from SAFS has been injected, supplanting the native clear sky photolysis rates. All other j values are computed via lookup using best practice to compare cross sections in the UV in the NCAR Tropospheric UltraViolet model (TUV).

One challenge we encountered was in the intermediate species. We found that as the model ran to steady state, the mixing ratios of unmeasured (or unconstrained) compounds increased in an unrealistic manner. We have adopted the methods described in Edwards et al. [2013], to mitigate the model artifact. Essentially, this works by introducing a depositional loss term that is a catch all value with $1/k$ time constant of 24 hrs for inert species and 10 hrs for compounds known to have additional true depositional losses. We are pursuing two paths for progress. The first involves running a simple set of constrained species to fine tune the model work-flow. The current method of operation involves stopping the model every half hour and reasserting the constrained species. The model stop/starts are introducing bottlenecks that are being addresses with more optimized code. Additionally, the large output files that contain the instantaneous rate computed for every reaction in the MCM are discarded until the final pass.

During model setup, we have identified additional data input needs (e.g. concentrations of certain alkanes). Depending on the identity of the desired input species, there may be data gaps during the campaign. One potential solution is to determine ratios of species of interest versus certain common denominator species that have complete data coverage during the campaign. Such denominator species include ethane and carbon monoxide. Identifying representative ratios and their typical ranges will allow us to give the model reasonable input concentrations, even during periods where data coverage was poor. The GC-ToF dataset has been mined to produce select ratios of chemical tracers. These ratios are used in the 0D model to ensure realistic defaults are used in the simulations.

Even with this ratio strategy, data coverage, particularly for VOC precursors measured by the then-prototype GC-ToF-MS instrument, is not sufficient. Thus, we have instead come up with a single synthetic "photochemical day". This photochemical day is based on the high ozone day which occurred on the 13th of May. To fill the gaps in data, relationships with other tracer compounds have been used to generate likely mixing ratios. Once the model workflow is improved, other days in the project will be run.

Model spin-up is required in order to come up with realistic concentrations of the unmeasured intermediate radicals. To do this the model is run with two separate integrators. The higher level integrator is run from Wavemetric's Igor Pro. This is where the measured data is injected into a run file for the DSMACC system, where the intial time is stepped backward slightly (e.g. 15 minutes). DSMACC simulates the intermediate radicals writing to a concentration file for ten minute steps. Igor captures the output file, interpolates all of the species that are not being constrained at the specific time of interest and produces a new run file at the next time step. Once the time period selected for modeling is covered, igor re-runs the process from the initial time but now uses the captured output of the intermediates as initial conditions. In this way, photostationary state is more closely approximated by minimizing the amount of time the model can produce unrealistic species in the channels where there are no constraints on the products. The number of passes that gives a high extent of convergence for OH (see Figure 19) and HO_2 is 5, however the model runs for 8.

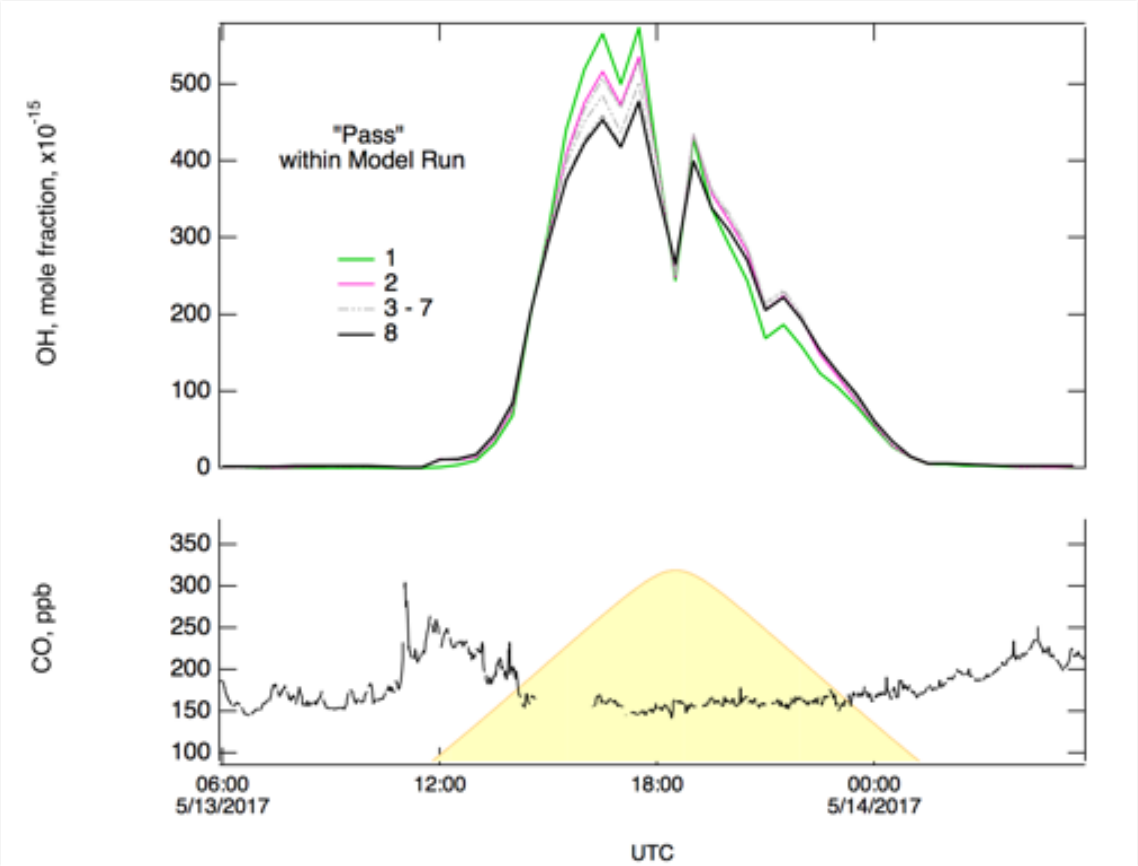

Figure 19. Convergence of the photochemical model to a photostationary state is observed via the OH radical.

At this point, concentrations have stabilized in a reproducible profile for the measurements, and the model photochemistry has reached a state where looking at intermediates and OH apportionment is meaningful. Figure 20 outlines a wrapper algorithm used to initialize and run the 0D model for this purpose.

```
Initialize:
        write first output file
        copy first file to input queue
Monitor:
        monitor the output queue
        copy out file to archive
        if( continue )
                write input file nudging measured values
                drop in input queue
```

Figure 20. Schematic of the wrapper algorithm used to execute the 0D chemical model for a photochemical day.

One blind spots in the MCM is RO2. RO2 only exists as a product species and internal reactant, and cannot be directly initialized. This means that it necessarily always must come up to speed at each run. This is thus the only intermediate that model tracks and generates that doesn't get passed forward.

Mining the model results for OH apportionment is a complicated venture. Thousands of individual chemical reactions produce (or consume) OH, and each reacting VOC is accounted for separately. For example, selecting methane as a primary VOC starts as two reactions, initial oxidation by atomic chlorine and hydroxyl radical, then quickly explodes as the implications of methyl-peroxy radical are filled in. Figure 21 shows the numerous resulting chemical equations required.

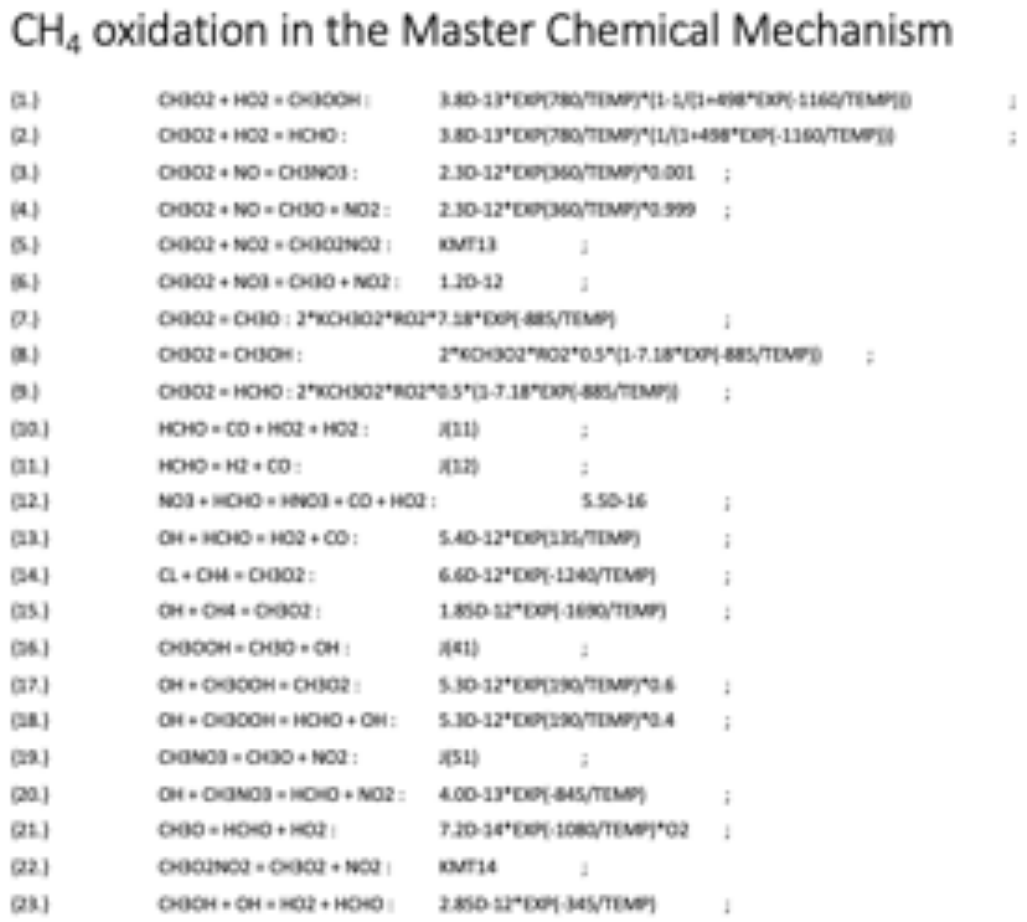

Figure 21. Methane oxidation in the MCM results in a cascade of photochemical reactions.

The DSMACC model automatically computes the instantaneous rate of all expressions in the model into the 'rates' file. As a result of this complexity, a parsing program has been written in IgorPro to sort through the model output and analyze the results.

Figure 22 illustrates the ranked contribution to OH loss, e.g. OH-reactivity based on the entire ensemble of constrained (measured) and unconstrained (modeled intermediates). Note that in the detailed oxidation of biogenic isoprene emissions, methyl

vinyl ketone (MVK) and methacrolein (MACR) are produced. These in turn produce formaldehyde. Formaldehyde is produced via the oxidation of many other VOC compounds. In subsequent sections, this model system is used to attribute the source of the HCHO using sensitivity analysis.

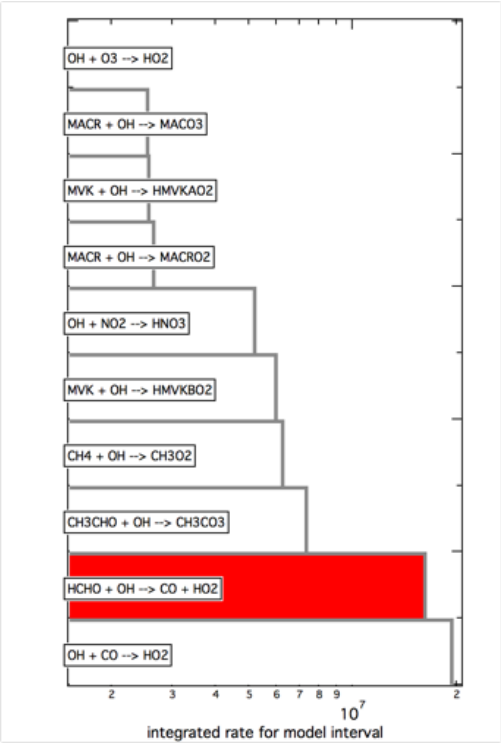

Figure 22. The top ten loss reactions that involve OH for the base model run 5/13/2017 at UTSA. HCHO is called out in red.

Results

Dataset Descriptions and Ratio Analyses

PTR-ToF dataset

In 2017, data from the Proton Transfer Reaction Time of Flight Mass Spectrometer (PTR-ToF-MS, or PTR-ToF for short) included ions present in the calibration tanks. By analyzing the full 2019 dataset in high resolution, we have produced a number of additional time traces.

High resolution peak extraction analysis of the PTR-ToF-MS mass spectra revealed 180 unique ions of which 64 are considered to be atmospherically relevant. Ions are deemed atmospherically relevant if their intensity was observed to decrease when the sample inlet was switched from ambient air to zero air. The response to zero criteria

provides a simple but robust method for distinguishing instrumental background ions generated from the internal mass calibration standard and the ion source to those formed from components in the ambient atmosphere. The chemical compositions were determined for all of the 64 atmospherically relevant ions. Thirteen (13) ions have been quantified as named species using calibrated sensitivity factors. Thirty-seven (37) ions have been given a tentative identification in the form of either a name or source. Fourteen (14) ions are referenced as unknown. Appendix A provides a detailed discussion on the ion identification.

The PTR instrument is selective and provides responses to those compounds whose proton affinity is greater than that of water, which includes most unsaturated hydrocarbons and VOCs containing a heteroatom such as O, N, and S. Alkanes are not normally detected with the PTR, but under conditions of high alkane loadings some are ionized through reaction with residual O_2^+ produced within the ion source. The O_2^+ reaction with the alkanes produces many of the same ions as the H_3O^+ reaction with alkenes, which means that there is always an uncertainty associated with the identification of the hydrocarbon ions observed.

Of the 64 atmospherically relevant ions reported 24 are hydrocarbons (CH), 35 contain C, H and O, 3 contain N, 1 sulfur and 1 chlorine. Figure xx provides an illustration of ion distribution as function of carbon number and the number of oxygen atoms present. Twenty (20) ions containing 1-O, twelve (12) contained 2-O and three (3) contained 3-O. No ions containing more than 3 oxygen atoms were identified.

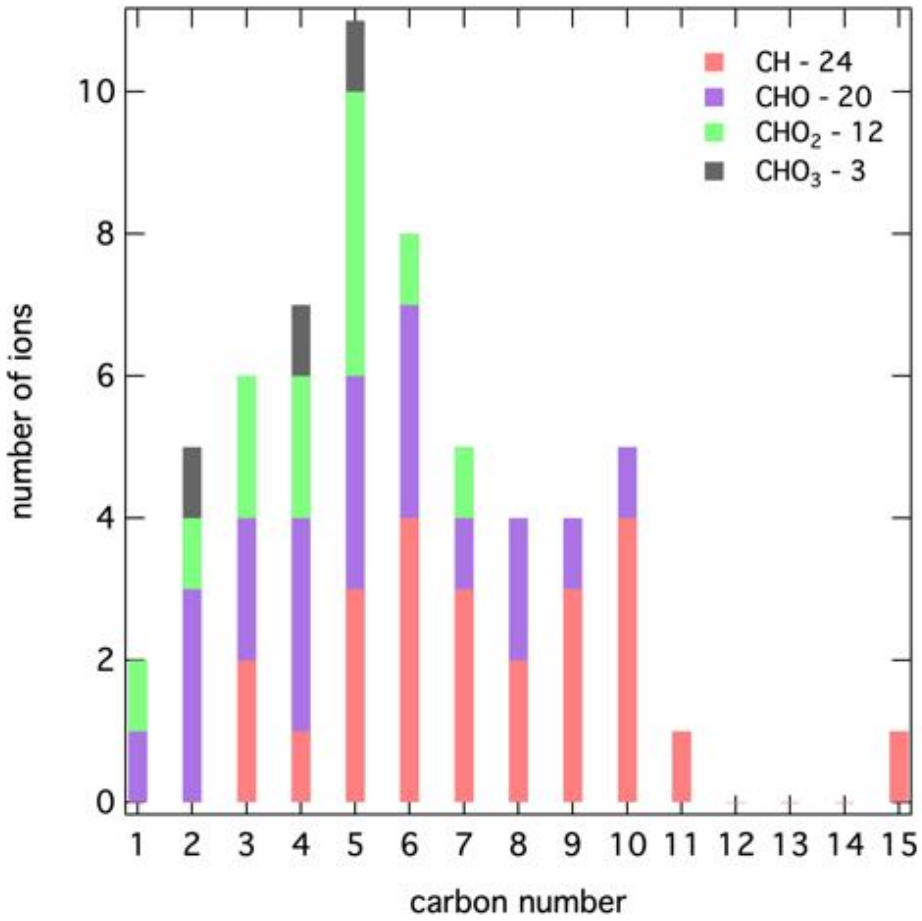

Figure 23. Distribution of identified ions in the PTR-ToF-MS as function of carbon number and oxygen content.

GC-EI-TOF dataset

Though PMF analysis is used in latter sections as a tool to discern multiple air masses observed during the study period, other simpler techniques are valuable to help validate the results of the PMF analysis. An example of a simple analysis type relies upon the ratio of iso- and n-pentane measured via GC-EI-ToF (gas chromatography electron impact ionization time of flight mass spectrometry) to distinguish between anthropogenic emissions typical of cities and from oil and gas activities. This technique has been described in the literature and relies upon the difference in ratio of straight-chain and branched alkanes found in refined gasoline versus unprocessed condensed liquids from oil and gas operations.

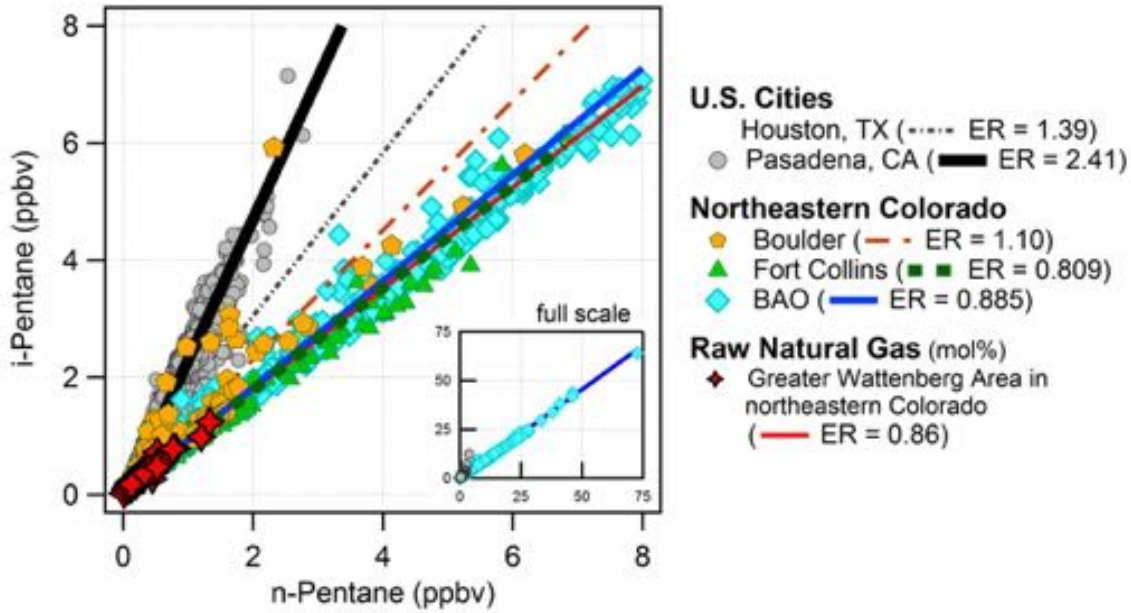

Figure 24. Observed iso- and n-pentane mixing ratios at various sites in Colorado and two cities (Houston, TX and Pasadena, CA).[*Gilman et al.*, 2013]

We apply this analysis to the gas-chromatograph measurements of iso- and n-pentane made during Texas AQRP 2017 at the UTSA field site May 13-16 and Floresville field site May 18-20. Figure 24(left) shows the time series of observed n-pentane mixing ratio for the two time periods, color-coded by iso-pentane mixing ratio, inside green and blue boxes for UTSA and Floresville, respectively. When we plot these mixing ratios against each other (Figure 24, right), we can see distinct and statistically significant differences in this ratio, in agreement with the observations by Gilman et al. [2013] shown in Figure 2. This indicates that the UTSA site shows iso- to n-pentane mixing ratios typical of U.S. cities, while the ratios observed at the Floresville site were typical of oil and gas operations.

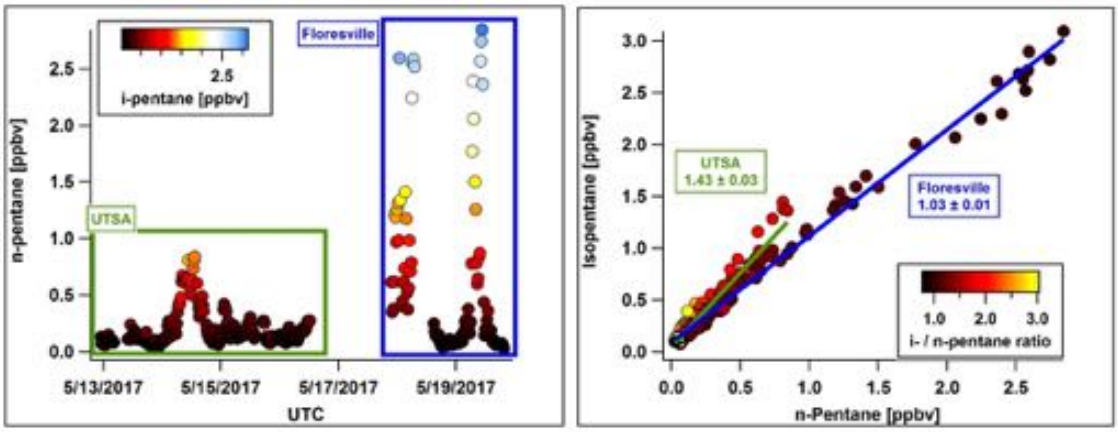

Figure 25. Left: Isopentane and n-pentane mixing ratios observed during Texas AQRP 2017. Right: Ratio of iso- to n-pentane mixing ratio for two sampling sites.

CAMS data in Floresville

The mobile lab was stationed at the Floresville AutoGC site from May 17 – May 22, 2017. Floresville due its proximity to the Eagle Ford oil and gas fields is heavily impacted by the emissions from this activity. This is readily apparent in the GC data reported by from the Floresville AutoGC facility, where there are numerous episodes where large enhancements in major alkane species are observed. Analysis of the AutoGC data set for the month of May 2017 shows that most of measured species are highly correlated to ethane when its concentration exceeds 20 ppb, as shown in Figure 26

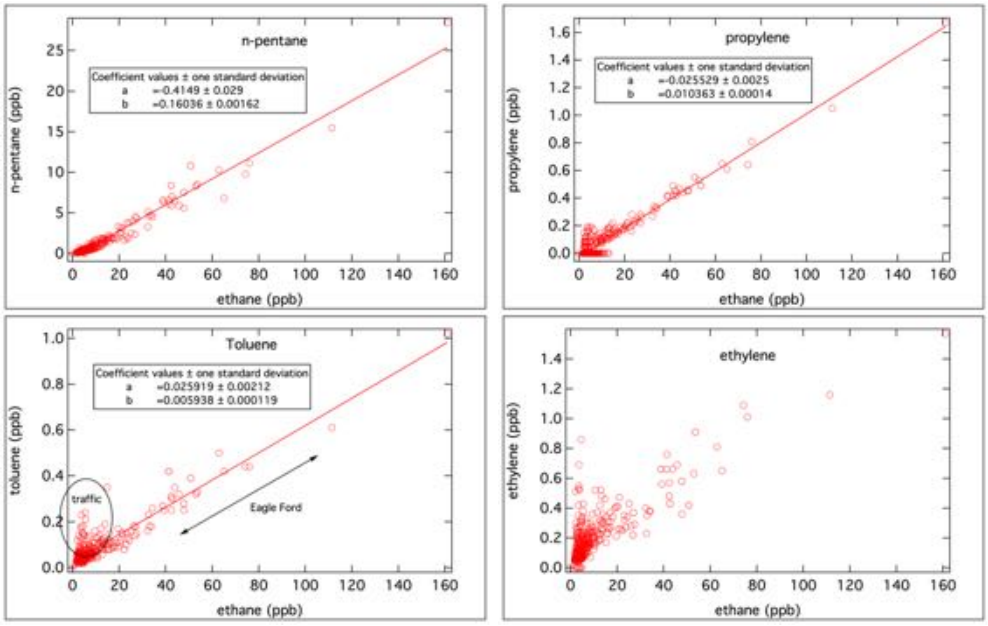

Figure 26. Correlation plots of selected hydrocarbons versus ethane. Data from the Floresville AutoGC.

The strong correlation between the different species under these conditions indicates that the episodic events share a common source. While it is intuitive to attribute the source of these episodic events to the Eagle Ford, it will be demonstrated that the composition of the air under these conditions closely resembles the composition of condensate tank emissions. Slopes from correlation plots, such as that shown in Figure 1, have been tabulated and were made for most of the reported species. These slopes yield a composition for each component scaled to ethane that can be examined to existing literature data sets for known sources. For the Floresville ambient air measurements we chose to compare them against those for condensate tank and gasoline exhaust emissions. This data is provided in the Table shown in Figure 29.

Compound_Floresville	Slope2Ethane	Intercept2Ethan	R_2Ethane	CondensateTank	Exhaust
propane	0.837	-1.419	0.995	1.135	0.2411
n-butane	0.45	-1.1	0.986	0.512	1.135
n-pentane	0.16	-0.415	0.98	0.211	0.9654
n-hexane	0.0432	-0.111	0.977	0.065	0.3998
n-heptane	0.0118	-0.0306	0.971	0.0311	0.3036
n-octane	0.00354	-0.007	0.963	0.0135	0.1059
isobutane	0.18	-0.372	0.989	0.189	0.1072
isopentane	0.15	-0.315	0.981	0.19	3.810
methylcyclopentane	0.0144	-0.062	0.946	0.0311	0.6097
methylcyclohexane	0.0102	-0.022	0.967	0.032	0.1791
2,2,4-trimethylpentar	0.0025	-0.0031	0.967	0.006	0.7060
benzene	0.0058	-0.0081	0.955	0.0244	1.219
toluene	0.0059	0.026	0.927	0.0177	1.78
m,p-xylene	0.0021	0.006	0.862	0.0058	

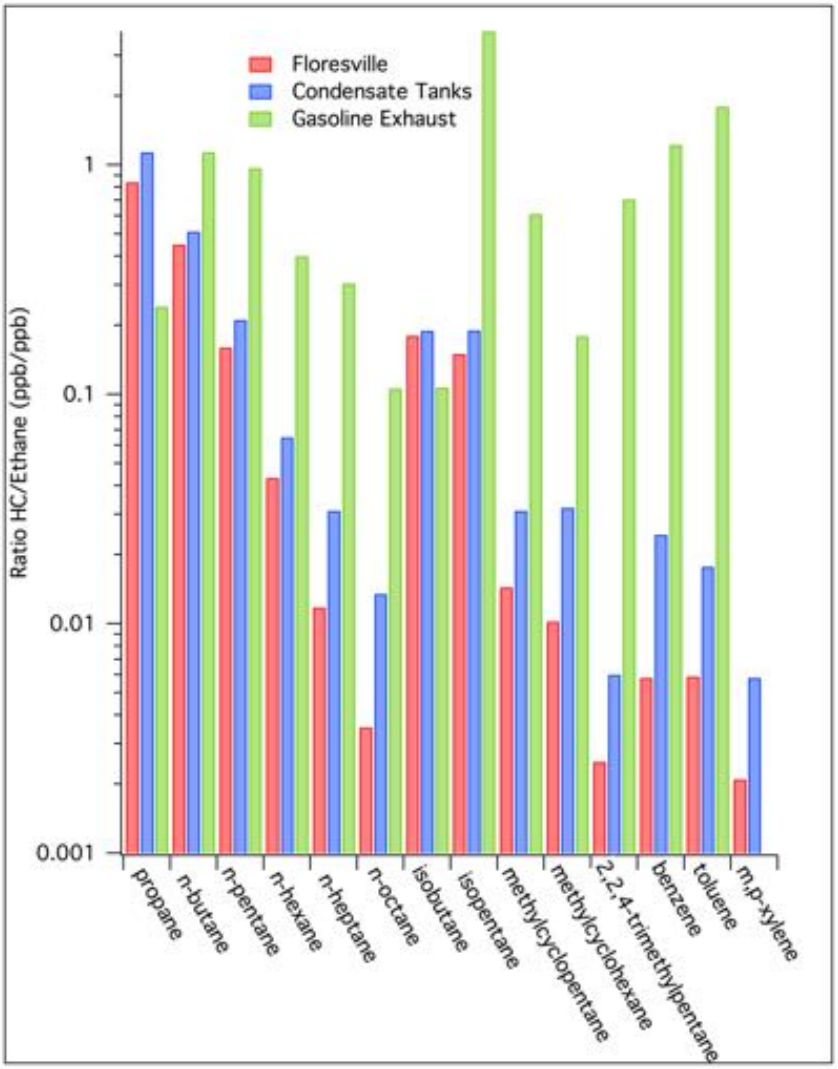

Figure 27. Floresville ambient air when experiencing high ethane (>20 ppb) conditions has a composition that closely resembles the emissions from condensate tanks.

It is instructive to examine whether other reactive alkene species, beyond those monitored by the AutoGC (C2-C4 alkenes) are associated with the Eagle Ford plume. While H_3O^+ reacts will all of the alkenes except ethylene, these species typically undergo a significant of fragmentation and produce ions predominately at m/z 43, 57, 71 and 85 depending on the number of carbons in the original molecule. The presence of additional alkenes would to lead to enhanced signals in these masses above what is anticipated from AutoGC data for the C3 and C4-alkenes. Figure 28 provides time series for the aforementioned species along with ethane and CO measurements recorded by Aerodyne. As might be expected, there are no significant enhancements for the alkene components. The signal containing the propylene response ($C_3H_7^+$) is general agreement (~ 1% of ethane).

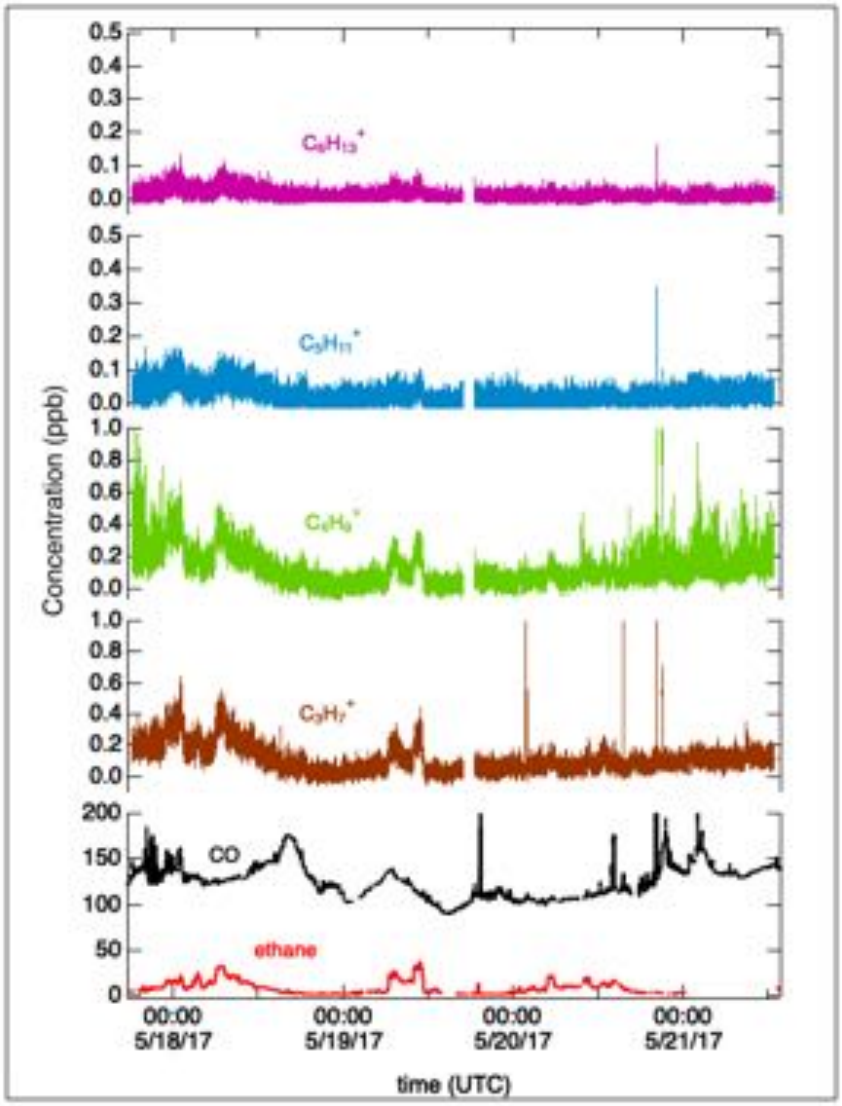

Figure 28. PTR-ToF-MS measurements from Floresville of ions produced from alkenes. Enhancements correlate with ethane and are in general agreement with what is anticipated from AutoGC measurements.

Positive Matrix Factorization

In this section, we report two sets of results that employ the positive matrix factorization (PMF) method. This method uses matrix factorization algebra to identify co-varying factors. A factor is made up of several chemical species or ion signals that all vary together. One advantage of this technique is that it is able to separate different time trends within the same compounds. This means that a single chemical species or ion can be present in multiple factors. This reflects the real complexity of the source environment in the SAFS study, where a single chemical tracer can have numerous sources, each with a signature variation with time or place.

PMF of the I-CIMS dataset

Since the I-CIMS detection scheme is based on the formation of detectable clusters between the iodide reagent ion and analyte of interest, it is most sensitive to multifunctional highly oxidized organic and inorganic species [*Lopez-Hilfiker et al.*, 2016]. Thus, the species observed by the I-CIMS are generally secondary in nature, though some primary species that are already oxidized, such as nitrophenol from biomass burning [*Lee et al.*, 2014] can also be observed. During this work the high-resolution I-CIMS mass spectra were fit to 2218 high resolution ions with varying degrees of oxidation. PMF was utilized to identify 7 groups of ions that covaried together in time. The covariations of the ions within each factor could be due to the fact that they originate from similar source regions and/or processing mechanisms. The mass spectra and time series of the 7 PMF factors chosen from this analysis are shown below.

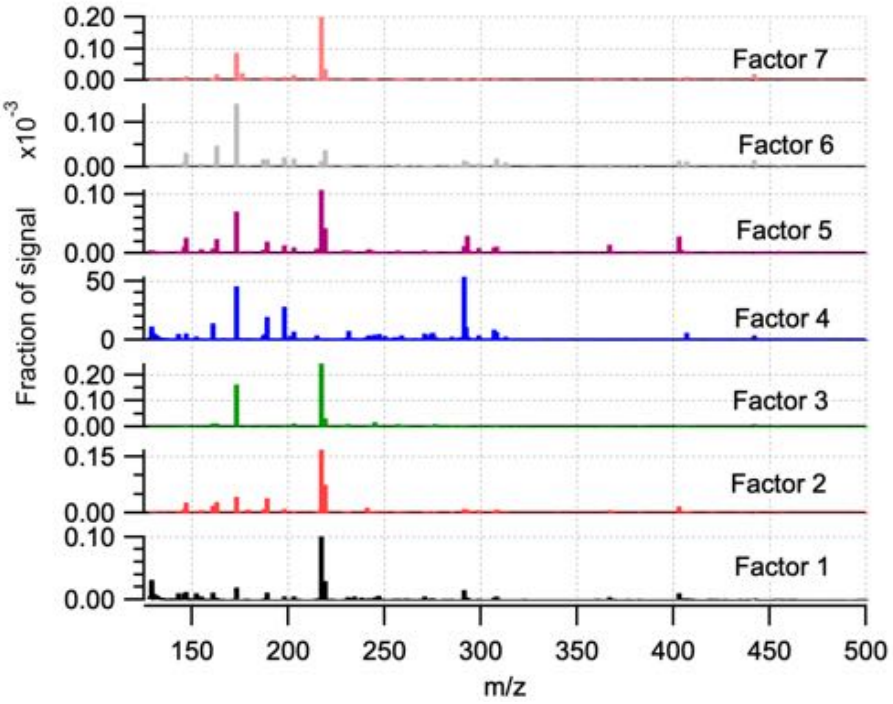

Figure 29. Mass spectra for the 7-factor PMF solution.

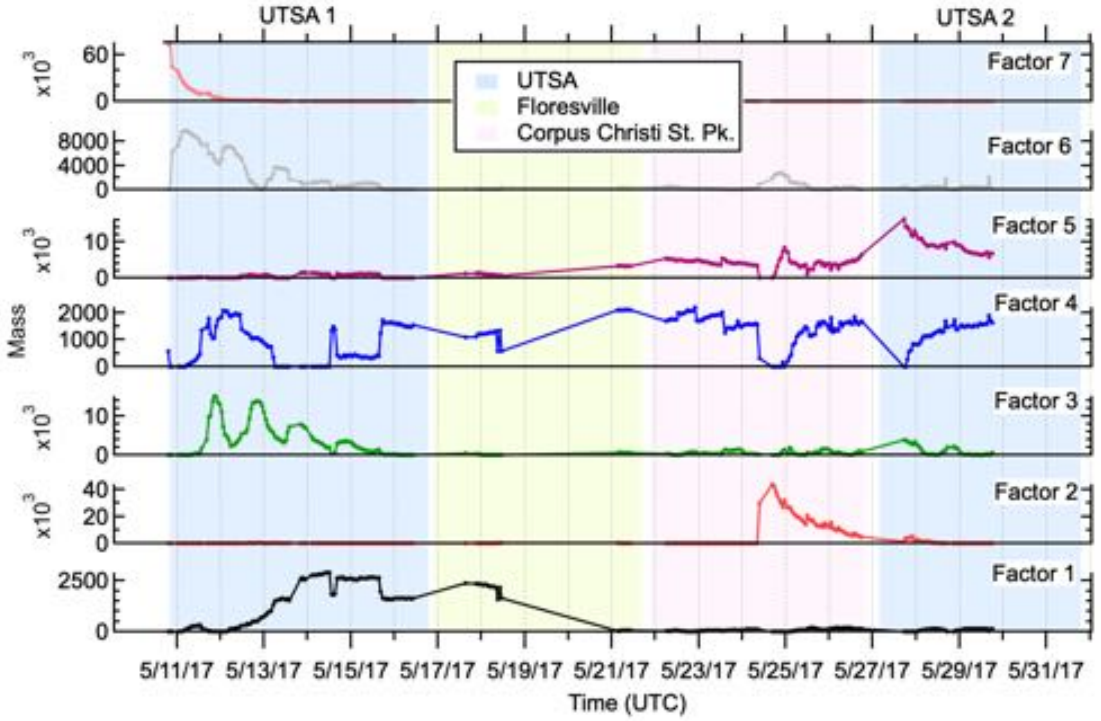

Figure 30. Time trends of the 7 factor PMF solution. Measurement locations are color-coded, and time periods of interest are labeled.

Figure 30 shows some clear temporal differences between the factors that are indicative of changes in the gas phase species measured by the I- CIMS. For example, during the first UTSA time period (UTSA 1), Factor 3 (green traces) shows diurnal peaks in the afternoons while Factor 6 (gray traces) shows diurnal peaks at night. Factor 7 (pink traces) and Factor 2 (Red) show large episodic peaks during the first UTSA measurement time period and during Corpus Christi State Park measurements, respectively. Factor 1 (black traces) becomes more important at the end of UTSA 1 while Factor 4 (blue traces) and Factor 5 (purple traces) are observed in Corpus Christi State Park and the second UTSA period (UTSA 2). Temporal correlations of the groups of species linked to each factor with external tracers can help in the interpretation of the sources that influence the factor. For example, higher temporal correlations of Factor 3 with measurements of precursor species such as isoprene and secondary isoprene oxidation species such as methacrolein and methyl vinyl ketone suggest that this factor represents species that are enhanced in airmasses that are influenced by isoprene sources. On the other hand, factor 6, which peaks at night in San Antonio, correlates well temporally with measurements of C2 and C3 benzenes and terpenes.

Detailed information on the interpretation of the I-CIMS factors is enabled by the identification of source-specific ions that have been observed in previous laboratory and field measurements. For this analysis, signature ions in I- CIMS spectra from the photooxidation of isoprene, a-pinene, and biomass burning are used [*Ehn et al.*, 2012; *Gaston et al.*, 2016; *Lee et al.*, 2014; *Lee et al.*, 2016; *Mohr et al.*, 2013; *Murschell et al.*, 2017; *Schobesberger et al.*, 2016; *Veres et al.*, 2015]. Results by Bin Yuan et al. during

the FIREX campaign are also leveraged, as are recent unpublished I- CIMS measurements at Aerodyne of trimethyl benzene (TMB) laboratory photooxidation under high-NOx conditions.

Before matching source signature ions with ions observed in factor mass spectra, a pre-selection was performed to identify the most useful and unique signature ions for any given factor. This pre-selection was necessary since the signals of most individual ions were distributed across multiple factor mass spectra. The ions likely to be most useful as signatures for a given factor were identified by examining the correlations between time trends of all measured ions every factor. A total of 53, 673, 34, 5, 14, 162, and 937 high resolution ions were selected as factor-specific signatures for factors 1 to 7 respectively since they showed higher temporal correlation factors ($R^2 > 0.4$). Of these factor-specific high-resolution ions, 13, 196, 52, 14, 12, 41, and 477 were found to match the source-specific signature ions that have been observed in previous work. Figure 31 shows the distribution of the signature ions from each factor that match the signature ions for isoprene, alpha pinene, biomass burning and photooxidation products of trimethyl benzene (TMB) under high-NOx conditions. The panels in the figure are also labeled by their temporal behavior as shown in Figure 30. While no individual factor is found to uniquely indicate a given source, the distributions vary with the relative importance of various source signatures. Factors 1 and 5 contain signature ions from all sources and thus appear to reflect mixed source influence. Factors 2 and 4 contain signatures that also correspond to all the source-specific spectra. The fact that their time series decay sharply at the beginning of each UTSA measurement time period, however, suggest that these factors are caused by instrumental artifacts. The only clear unique source-specific features are found in Factors 3 and 6,which contain unique marker ions for isoprene and α-pinene precursors respectively. The likely source influence for each factor is summarized in Table 2. The specific ions corresponding to the source influences that are observed during this study are available as a table in Appendix C.

Table 2. Likely source influences for I-CIMS PMF factors

	Source Influences
Factor 1	Mixed Source
Factor 2	Instrument Artifact
Factor 3	Isoprene signature
Factor 4	Mixed Source
Factor 5	Mixed Source
Factor 6	a-pinene signatures
Factor 7	Instrument Artifact

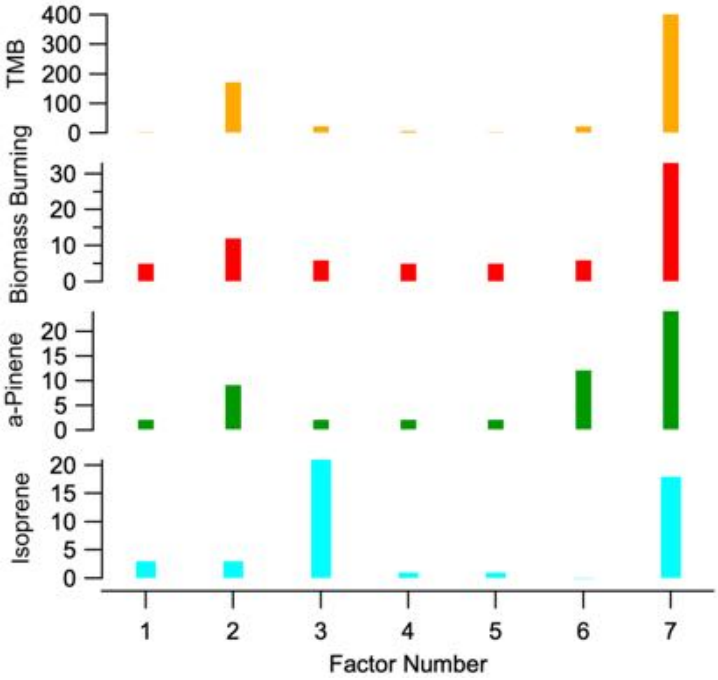

Figure 31. Signature ions of each factor categorized according to their match with known source-specific ions.

PMF of combined datasets

PMF was used on the GC-EI-ToF-derived VOC mixing ratio time series for the campaign. To expand the number of chemical species input into the analysis, additional mixing ratio data from PTR-ToF and gas-phase species (Tunable Infrared Direct Laser Absorption Spectroscopy, TILDAS or Non-dispersive infrared, NDIR). All species were modeled using their mixing ratio in ppb except CO and CO_2, which were used in units of ppm and ppth, respectively, to ensure that the absolute numerical variance scaled roughly in the same dynamic range as the VOC species. These measurements were averaged to the GC data collection start and stop times (5 minutes every 20 minutes). The final list of chemical species used, along with their source instrument is shown in Table 3.

Table 3. Chemical species used for PMF analysis of time series for SAFS, grouped by measurement type. Mixing ratios for all compounds in ppb unless denoted in brackets. Species with (*) indicate summed isomers.

TILDAS Species	GC-EI-TOFMS Species	PTR-TOFMS Species
CO [ppm]	Benzene	Isoprene
Methane	Toluene	Monoterpenes*
Ethane	Ethylbenzene	Methyl vinyl ketone & methacrolein*
Ethyne	m&p-Xylenes*	Methanol
Formaldehyde	o-Xylene	Acetone
Formic acid	1,2,4-trimethylbenzene	Acetaldehyde
	i-Pentane	Hydrogen cyanide
NDIR Species	n-Pentane	Acetonitrile
CO_2 [ppth]	n-Hexane	Dimethyl sulfide
	n-Octane	
	2,2,4-trimethylpentane	
	Methylcyclopentane	
	Cyclohexane	
	Methylcyclohexane	

PMF analysis requires that all species have complete time series with no missing data points, so for any GC sample period where any compound had unreported data, the sample was removed from the data set for all species. As a result of this, the original time series of 541 data points was winnowed down to 194 rows of 33 gas species. PMF analysis is highly-sensitive to the uncertainties for all reported species. Here, the data uncertainties were defined based simply on the measurement technique rather than on a compound-by-compound basis. For GC data, uncertainties were defined as 12% + 0.005 ppb; for PTR data, 20% + 0.01 ppb and for other measurements, 10% + 0.005 maxing ratio unit.

PMF analysis requires the user to determine the number of factors needed to describe the variability in the data, and the modeling software provides diagnostic tools to evaluate the results to minimize factors and avoid over-fitting of the data. Figure 32 shows two plots of these diagnostics for the VOC dataset described above. The Q-plot (left side) is an attempt to estimate the number of degrees of freedom that exist in the dataset, where additional factors significantly below $Q/Q_{expected} = 1$ likely contribute little. In this case, we have selected seven factors, or degrees of freedom. R-R plot (right) compares the chemical co-variance of the seven factors with their temporal variance. While we see significant chemical co-variance for several factors, they have poor temporal co-variance. This is likely due to air masses having similar emission sources but different levels of photooxidative processing (see below).

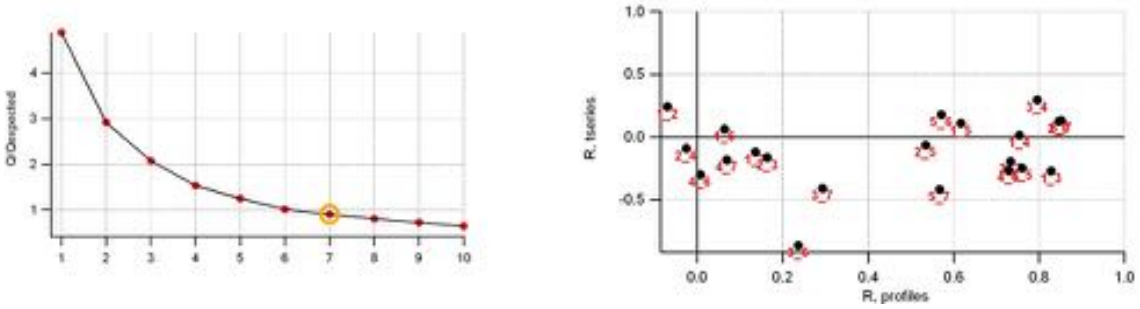

Figure 32. PMF diagnostics. Left, Q-plot, showing the enhancement factor (Q/Q Expected) for estimate of number of degrees of freedom; Right, R-R plot, comparing temporal co-variance vs. chemical co-variance.

As noted, we found that seven factors provided adequate description of the observed VOC atmospheric burden. These factors are shown in Figure 33, with the chemical signature of each factor on the right and the individual temporal response of each factor in the center. The factors have been given nominal assignments based upon their chemical signature, temporal variability and location when observed [*Abeleira et al.*, 2017]. A brief descriptor of each:

- Background boundary-layer air: long-lived chemical species (e.g. benzene), a small contribution (5-10%) throughout the time period modeled. This represents a small fraction of the VOC background consisting of long-lived species that does not have a significant diurnal or wind-direction variance. This is typical of continental boundary layer air, e.g. Abeleira et al., 2017
- Biogenic emissions: a mixture of isoprene / monoterpenes and some aromatic species, with small contribution except for short time periods (overnight) at UTSA.
- Anthropogenic (auto): pentanes / hexane with some combustion markers (e.g. formaldehyde), showed large enhancements during the day at UTSA, typically 10-30% of VOC burden
- Oil and gas emissions: alkane / cycloalkane / aromatic signature with little oxidized VOC, negligible at UTSA but dominant (up to 80%) at times in Floresville.
- Oxidized urban VOCs (ethane > methane): OVOC dominant signature, with ethane contribution larger than scaled methane. This a typical urban oxidation signature and contributed 20-40% to the VOC burden at UTSA, and typically <20% at Floresville.
- Oxidized biogenics: OVOC signature with biogenic oxidation markers (MVK, MACR, methanol). This was an important contributor to the UTSA VOC burden in the afternoons (up to 40%), but had a smaller contribution in Floresville.
- Oxidized oil and gas VOCs (methane > ethane): OVOC signature similar to urban, but with methane signal larger than ethane. This signal had one large event at Floresville, contributing about half of the VOC burden. Interestingly, it was a non-negligible signal at UTSA, twice observed contributing about 20% of the

VOC burden. (First event was 14-May, 18:00-24:00 local time; second event was 15-May, 19:00-21:00 local time)

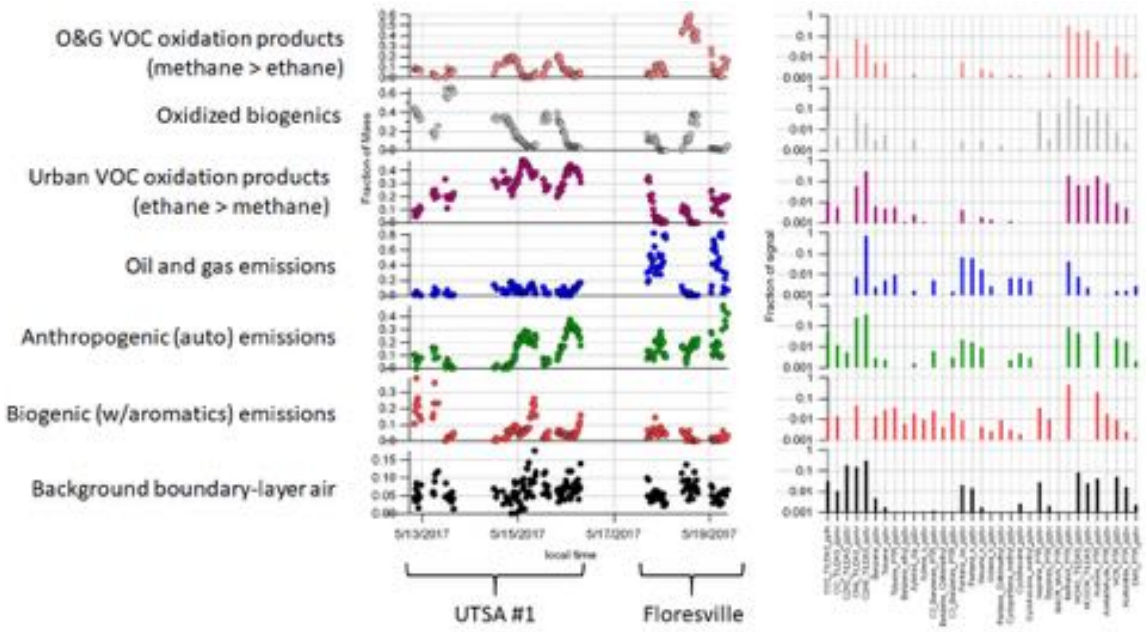

Figure 33. PMF results of combined datasets, with nominal description of each factor. Left: Individual factor contribution to total chemical burden. Right: chemical signature for each factor. The chemical signatures are shown as bar graphs, with each bar corresponding to a measured species. TILDAS refers to gas-phase species measured via Tunable Infrared Laser Direct Absorption Spectroscopy (with the exception of CO_2, which was measured via non-dispersive infrared spectroscopy (NDIR)); PTR refers to species measured via Proton Transfer Reaction Time of Flight Mass Spectrometry (PTR-ToF); all other species shown were measured via Gas Chromatography Electron Impact Time of Flight Mass Spectrometry (GC-EI-ToF or GC-ToF for short).

To summarize, we can describe the observed VOC variability as a combination of three major emission sources (urban / auto; biogenic; oil and gas) and the oxidation products of those sources. This analysis is in good agreement with the PTR-ToF and CAMS site Auto-GC analysis presented in previous sections.

One benefit of PMF analysis that is present in the above results is the ability of this methodology to separate contributions of different sources with different time trends to the same chemical species. For example, these results aspect

Back-Trajectory Analysis and Landcover Mapping

HYSPLIT Back-Trajectory Footprints

A complete dataset of HYSPLIT back-trajectory footprints has been run on a high-powered computational computer. These trajectories are run in particle mode to generate a probabilistic footprint of the airmass source. This is in contrast to the simpler back-trajectory mode, which follows only the cloud centroid and results in a trajectory "string"

on a map. The particle mode of the model is releasing 100,000+ particles and tracking them and allows particles to randomly spill into other grid cells. This analysis leverages the best-available meteorological datasets and takes about 30 minutes to complete 1 hour of simulated time. The results are output in a format that is accessible in the analysis software of choice, Igor Pro, and summary .PNG figures were output for ease of more qualitative analysis.

The full set of hourly back trajectory PNGs are available in the dataset deliverables for this project (see also Appendix D). The calculations have also been done for the UTSA location for the entire campaign, in order to support other research groups doing stationary measurements at this location. A movie of these UTSA back-trajectories, which gives an overview of the general wind directions and transport during the campaign, can be viewed here:

https://herndon.homeunix.net/owncloud/index.php/s/peG6fXDR9FQtx5r

An example of the Hysplit back-trajectory footprint figures is shown below. These two simulations show the origin of the airmass sampled during the campaign during an hour-long period. The first, on 05/17/2017 shows clear influence from the ocean; the second, only a few days later on 05/23/2017, shows more influence from on-land surfaces. These two time periods are compared with measured tracers (like ethane and pentanes) as well as PMF factor results (which represent groupings of species), to better understand potential sources of observed species.

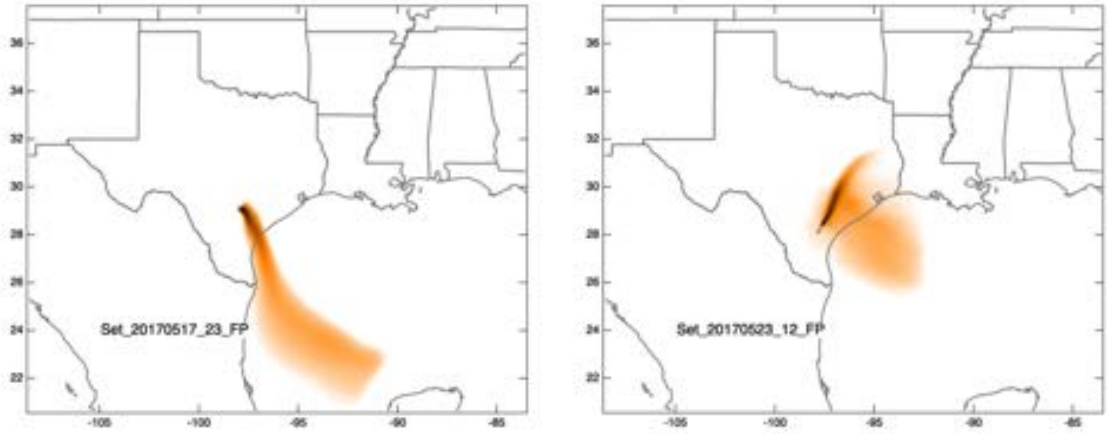

Figure 34. Example of HYSPLIT back-trajectory footprints.

The results of a test of the release height are depicted in Figure 35. The figure suggests that the long-range transport contours are not strongly coupled to the choice of release height. In the case of the measurements, the air was sampled at ~10m. The convention within the back-trajectory calculation community, 10 meters is considered too low to the ground for the model to accurately calculate a footprint. This figure suggests that for long-range interpretation of airmass sources, the choice of release height (within the bounds explored here) may not matter. Note that the next steps will be to compare the actual magnitude of the footprint. In this analysis, owing to the use of a log scale, some numerical differences due to the simulated release height will not be discerned. This release height insensitivity lends confidence, however, to the planned use of these results in this research project: figures such as those below will be used in a qualitative way to

look at the geographic area sampled for better understanding of the potential emission sources.

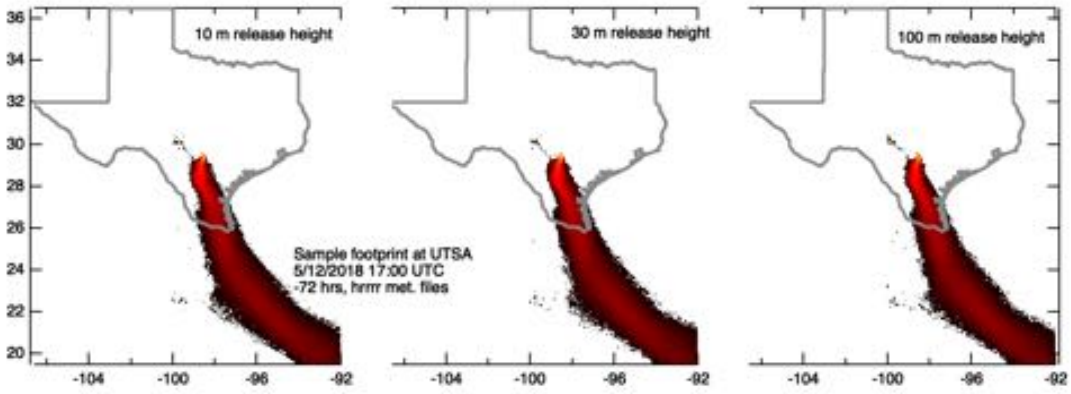

Figure 35. Sensitivity of the model release height. The three panels (from left the right) are the result of the HYSPLIT model calculation of airmass 'footprint' at 10, 30, 100 meters respectively. The color scale is a logarithmic representation of likely surface sources during the prior three days.

Combining Hysplit Footprints with Land Use Data

Ecological mapping information was obtained from the Ecological Mapping Systems of Texas database (EMST) published by the Texas Parks & Wildlife Department [*Elliott et al.*, 2014]. The data were obtained in raster format at 10m resolution. GIS tools were used to resample the dataset onto a 1000 m grid size for loading into the analysis software, Igor Pro.

398 individual vegetation or ground cover classes were coarsely sorted into broad categories: urban (0), oak (4), pine (3) and other vegetation (2), using keyword matching of the common name (e.g. Bastrop Lost Pines: Loblolly Pine Forest, or Rolling Plains: Mixedgrass Prairie, Table 4 below). Land cover types that did not fall into these categories were marked as “sparse” (1).

Table 4. Broad categorization of the EMST dataset [*Elliott et al.*, 2014] classes.

Keyword	Category ID	Broad Category
urban	0	Urban
barren, sparse, water, playa, dune, etc.	1	Sparse
herb, grass, swamp, prairie, marsh, woodland, shrubland, wood, forest	2	Other Veg.
pine	3	Pine
oak	4	Oak

The ecological data was resampled onto a 0.025 degree grid to match the resolution of the Hysplit back-trajectory calculations, and bounded to the area spanning 26 to 32 latitude by -100 to -94 longitude. The resulting 4-category map is shown in the figure

below. The city limits of San Antonio [*City of San Antonio*, 2019] are outlined in black. The urban areas of Austin and Houston are also evident (dark blue spots).

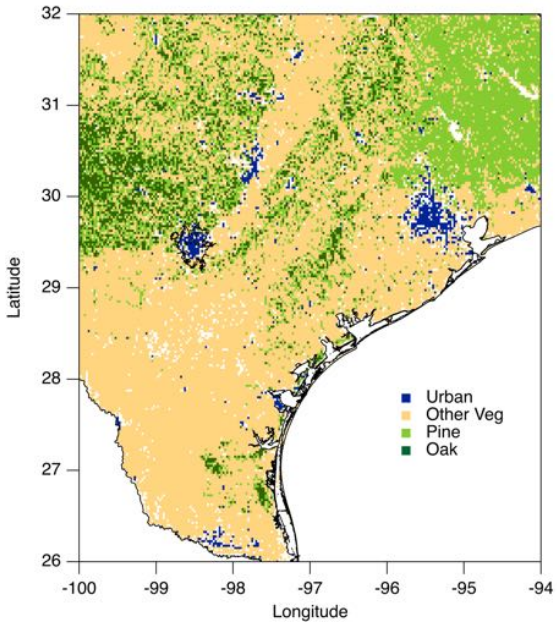

Figure 36. Ecological and oil and gas mapping information is shown gridded onto a 0.025 x 0.025 decimal degree grid. The city bounds of San Antonio (center left) are outlined in black.

The extent of the Eagle Ford, Barnett and Haynesville Plays were obtained from the EIA [*U.S. Energy Information Administration*, 2016] and gridded onto the same grid. The oil and gas play classification was allowed to overlap with other land cover types.

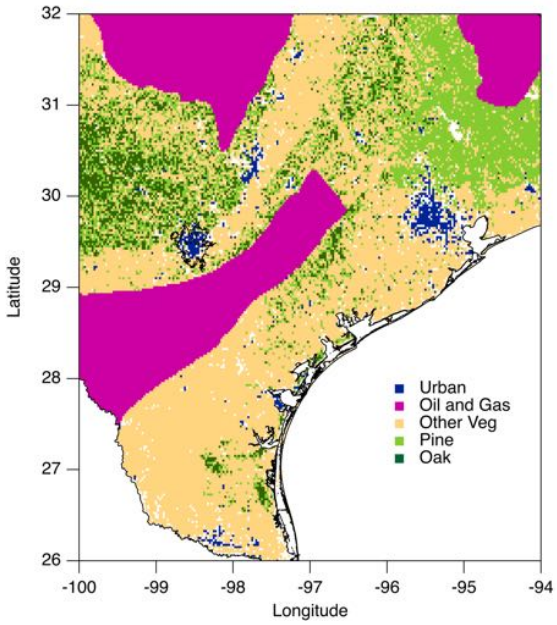

Figure 37. Oil and gas mapping information (purple) is shown gridded onto a 0.025 x 0.025 decimal degree grid overlaying the ecological and land use data. The Eagle Ford Shale Play (center left) is shown along with some of the Barnett (top left) and Haynesville (top right) plays. The city bounds of San Antonio are outlined in black.

An overlap analysis was done with results from Hysplit footprint simulations. The Hysplit footprint was put onto a log scale and clipped to within bounds to eliminate artifacts. The footprint overlap as a function of time is calculated only within the domain shown in the figure above. The calculated overlap does not further normalize results, in order to faithfully represent the calculated footprints and their log scale intensities, which sometimes extend beyond the bounds of the calculated domain. An example overlap figure is reproduced below.

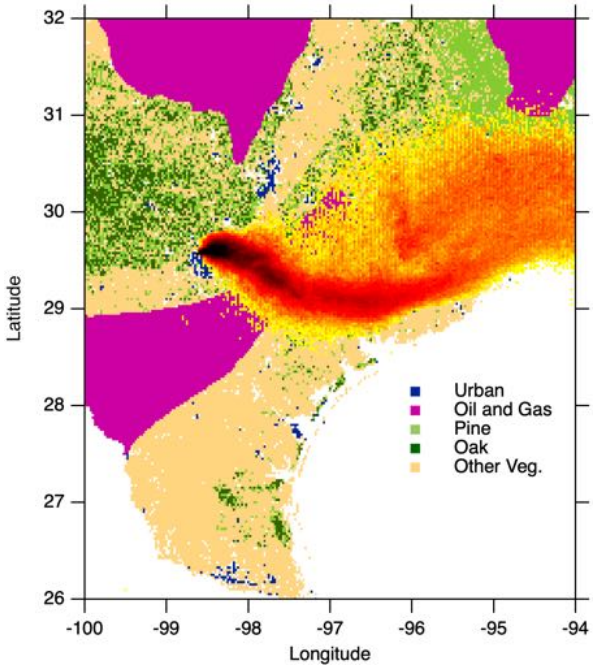

Figure 38. Example Hysplit overlap figure. The black/red/gold cloud shows the Hysplit footprint overlaid upon a map of Texas.

The overlap as a function of time is shown in the figure below along with a few indicative tracers.

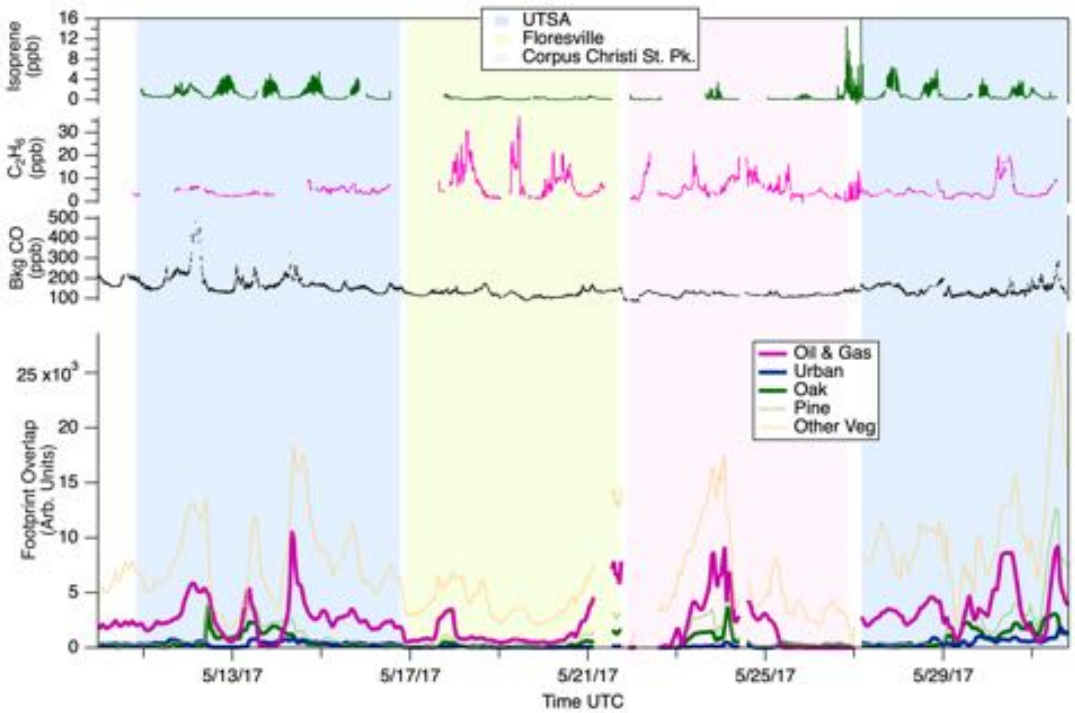

Figure 39. Hysplit footprint overlap with 5 different land cover types. Measured isoprene, ethane and background carbon monoxide time traces are also shown. Shaded areas represent time spent at UTSA (blue), Floresville (green) and Corpus Christi State Park (pink).

Additional footprint calculations have been done to try to separate the impacts of different urban centers on the SAFS dataset. The urban land classification shown above was categorized based on rough bounds for a select few cities:

Table 5. Rough city bounds, in decimal degrees, applied to the Urban land cover type shown in Figure 37

City	Lat S	Lat N	Lon W	Lon E
Corpus Christi	27.60	27.90	-97.75	-97.10
San Antonio	29.25	29.70	-98.80	-98.30
Houston	29.25	30.00	-96.00	-94.75
Austin	29.85	30.70	-98.00	-97.40

These data series have been output onto the standard SAFS 1-minute time series.

- MM_footprint_Urban
- MM_footprint_Pine
- MM_footprint_OtherVeg
- MM_footprint_OilGasPlays
- MM_footprint_Oak

Source Locations and Measured Airmasses

A challenge in interpreting results from the SAFS study has been understanding the source of the measured airmass. Measurements of wind direction give a simplistic understanding of airmass source (e.g. coastal air from the south vs continental air from the north). Hysplit simulations, which trace air particles back in time, give a much more detailed and quantitative view. Coupling these two-dimensional Hysplit footprint maps with maps of Texas land cover allow us to see the different source categories that may be influencing measured results. In this analysis, we produced time traces of this overlap that can be directly compared to, and correlated with, time traces of various measured species.

An overlap analysis was done with results from Hysplit footprint simulations (see Methods section).

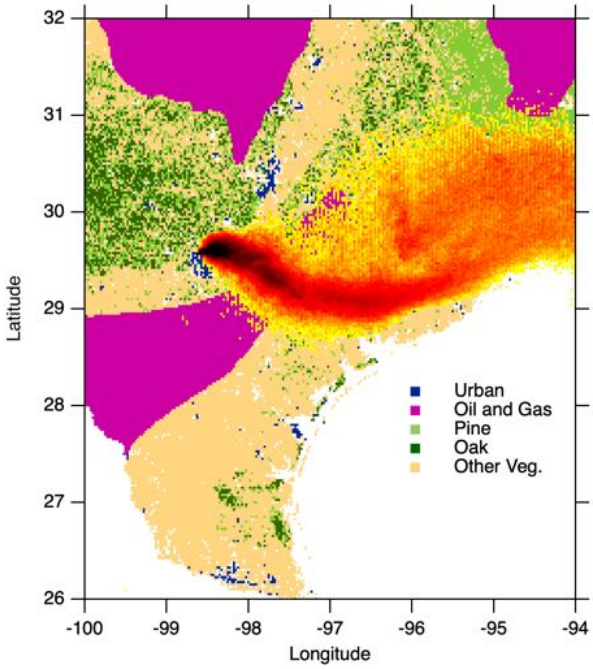

Figure 40. Oil and gas mapping information (purple) is shown gridded onto a 0.025 x 0.025 decimal degree grid overlaying the ecological and land use data. The Eagle Ford Shale Play (center left) is shown along with some of the Barnett (top left) and Haynesville (top right) plays. The city bounds of San Antonio are outlined in black. The black/red/gold cloud shows a single Hysplit footprint overlaid upon the map of Texas.

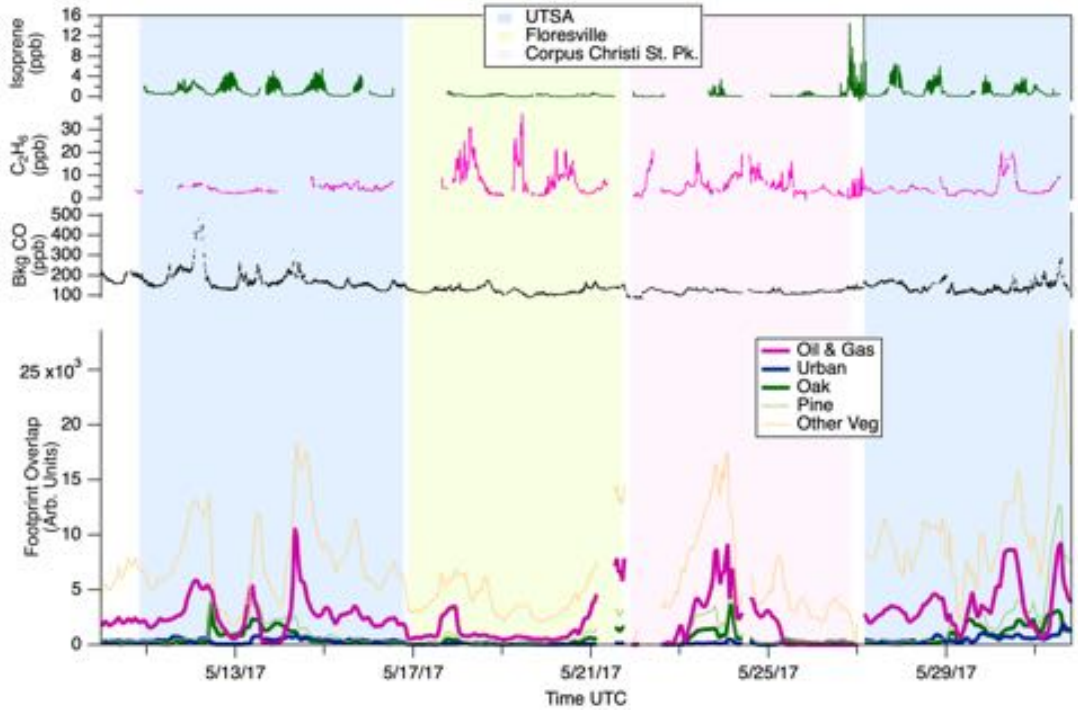

Figure 41. Hysplit footprint overlap with 5 different land cover types. Measured isoprene, ethane and background carbon monoxide time traces are also shown. Shaded areas represent time spent at UTSA (blue), Floresville (green) and Corpus Christi State Park (pink).

This analysis results in time traces representing an airmass footprint overlap with each of the selected land use types as a function of time (see Figure 41).

This data shows periods of time with more or less influence from oil and gas regions. It also shows relatively little variation in the biogenic signatures (Oak, Pine, Other Vegetation) sources with the exception of the time spent in Corpus Christi State Park. Measured isoprene shows the strongest enhancements at San Antonio; we attribute this enhanced isoprene to the hyper-local sampling environment (nearby oak trees on campus).

Oaks are by far the biggest expected isoprene source. However, isoprene is only emitted during the day, with an emission rate sensitive to light and temperature. This footprint experiment does not include the impact of the expected diurnal and temperature-dependent emissions of such biogenic tracers. However, when the SAFS dataset contains traces for both solar elevation angle (MM_SEA) and ambient temperature (MM_temperatureC).

Various correlations can be explored using these traces. For example, isoprene is plotted against the Oak footprint below, showing poor correlation, likely due to the factors described above. On the other hand, ethane can be plotted against the oil and gas footprint and shows a much better correlation. Particularly notable is the correlation of the urban footprint with background CO concentrations (i.e. with spikes due to local traffic and other sources removed). This indicates that background CO is a good tracer for impact of urban land cover types on the SAFS measurements.

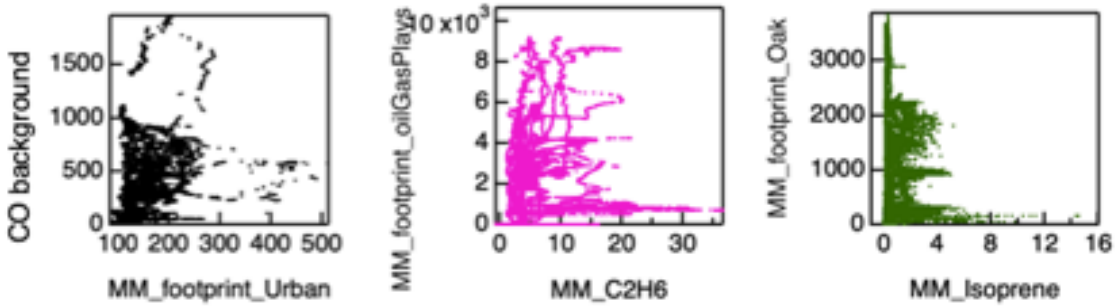

Figure 42. Overlap footprint plotted versus selected chemical tracer. The Urban footprint (black) is plotted against background carbon monoxide (background_CO) showing good correlation. The oak footprint (green) is plotted against isoprene, showing poor correlation. The oil and gas play footprint (pink) is plotted against ethane (C_2H_6) showing better correlation.

Given that winds in Texas often originate from the South East, one might ask what the potential impact would be of other urban areas on San Antonio's air. The urban footprint overlap was divided into the impacts of four cities, as shown in Figure 43 (see Methods for details). This figure shows that measurements in UTSA were strongly impacted by the City of San Antonio, as expected, but there were also times when Austin (5/13/2017) or even Houston (5/31/2018) had an impact. Corpus Christi has a small influence on measurements at all three sites. Winds were only rarely from the SE when sampling in Corpus Christi State Park, accounting for the very minor impact of Corpus Christi urban areas to the sampled air there. This analysis was done overlapping Hysplit

footprints that were clipped and logarithmically scaled in order to eliminate artifacts and visually show the full extent of back-trajectories on a map; this same analysis using unscaled Hysplit footprints washes out the contributions of other cities and increases the contribution of San Antonio itself.

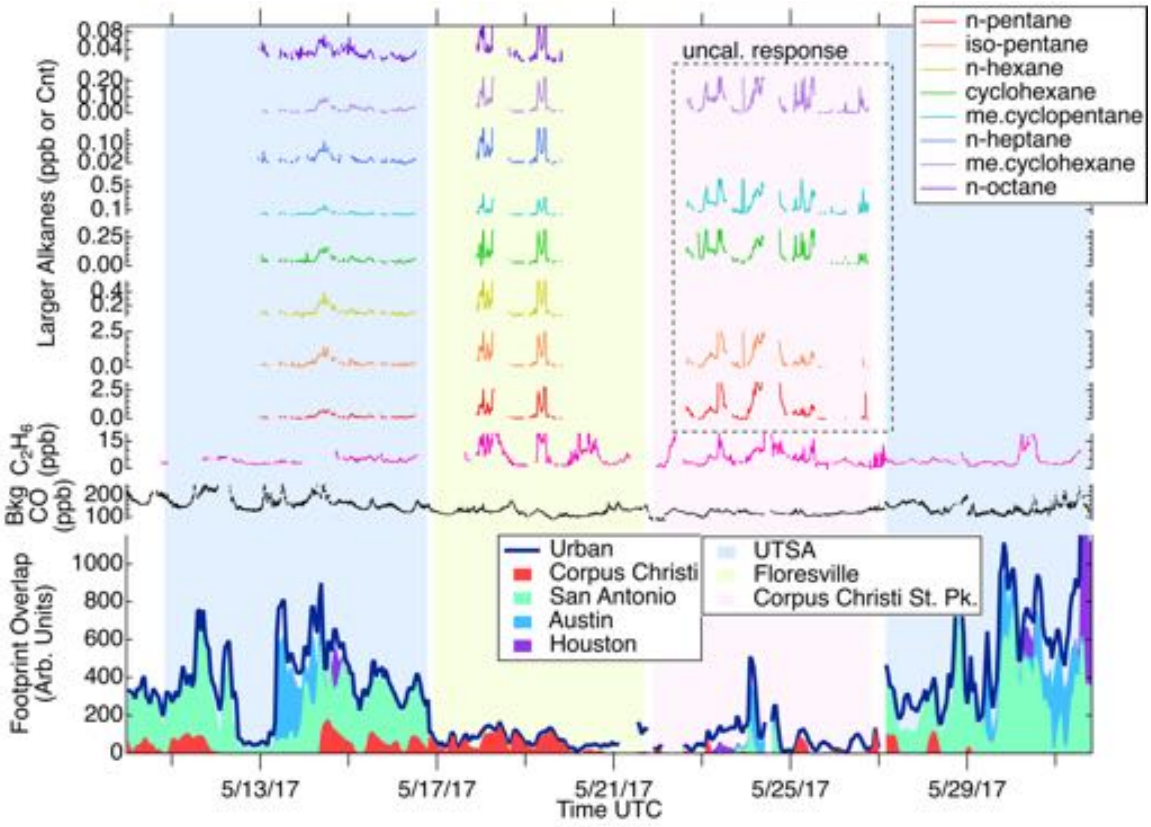

Figure 43. Hysplit footprint overlap with the Urban land cover type, divided into four urban areas: Corpus Christi, San Antonio, Austin and Houston. Measured tracers including background carbon monoxide, ethane selected larger alkanes are shown. Shaded areas represent time spent at UTSA (blue), Floresville (green) and Corpus Christi State Park (pink). Larger alkane measurements in Corpus Christi State Park are uncalibrated in their response.

An investigation of chemical tracers is inconclusive as to the impact Corpus Christi or the nearby refineries on measurements during SAFS. Figure 43. Shows a selection of alkanes plotted along with the urban footprint overlaps. The Corpus Christi footprint overlap has approximately equal impact during the first San Antonio period and during Floresville measurements (both times when the same column was used and good calibrations were available). In contrast, all of the hydrocarbons plotted show increases in Floresville. This implies that a unique hydrocarbon characteristic of Corpus Christi industrial activities has not been identified in this dataset.

Apportioning Ozone to Source Categories

Concentrations and OH Reactivity

A simple way of examining the relative importance of VOCs during the SAFS campaign is shown in Figure 44 and Figure 45. First, average concentrations of a measured species is computed, in parts-per-billion carbon (ppbC). This lends extra weight to larger molecules and aims to show VOCs according to their ozone forming potential. The normalized fraction of OH reactivity of individual VOCs is computed using their weighted concentration and their reaction rate constant with OH: kOH*ppbC/sum of kOH*ppbC, where ppbC is campaign averaged concentration multiplied by its carbon number. The top 20 species measured by PTR-ToF are shown alongside some key trace gases like formaldehyde (HCHO), methane (CH_4), ethane and propane. Species that show up high in both graphs will be most important towards ozone photochemistry, most notably isoprene and MBO. This simple investigation agrees with conclusions from PMF analysis of I-CIMS data and from 0D modeling results. This is a simplest way of scaling all of the components based on their reactivity to OH without needing to know the actual OH concentration. Adding up all of the fractions yields 1 so missing carbon is not included. This analysis includes all of the data and so doesn't reflect changes with sampling location or differences between daytime/nighttime photochemistry.

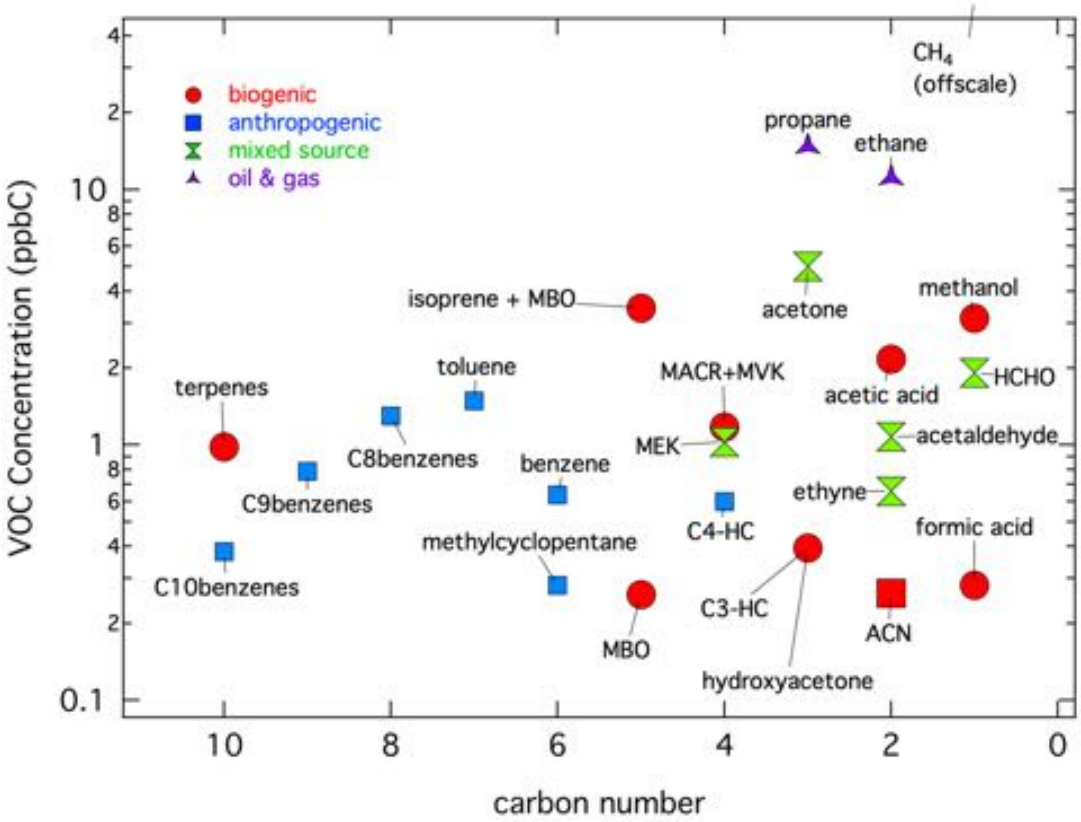

Figure 44. VOC burden for selected VOCs.

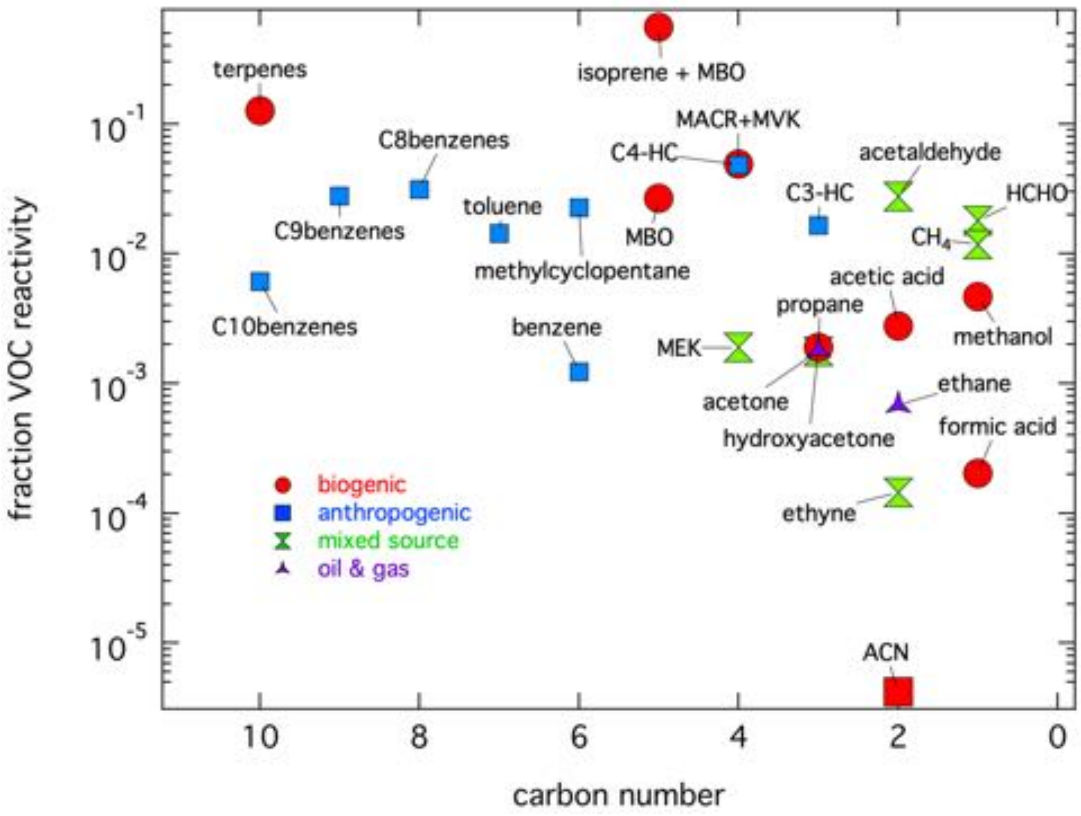

Figure 45. Relative VOC reactivity for selected VOCs.

Photochemical Modeling to Apportion OH reactivity

There are two main types of results that come from the 0D photochemical model. First, mixing ratios of various species can be determined. They can either be constrained by measurements (as in the CO result in Figure 46) or left unconstrained (for unmeasured species, or for model benchmarking). The concentration of OH radical is the primary time trace of interest from the 0D model.

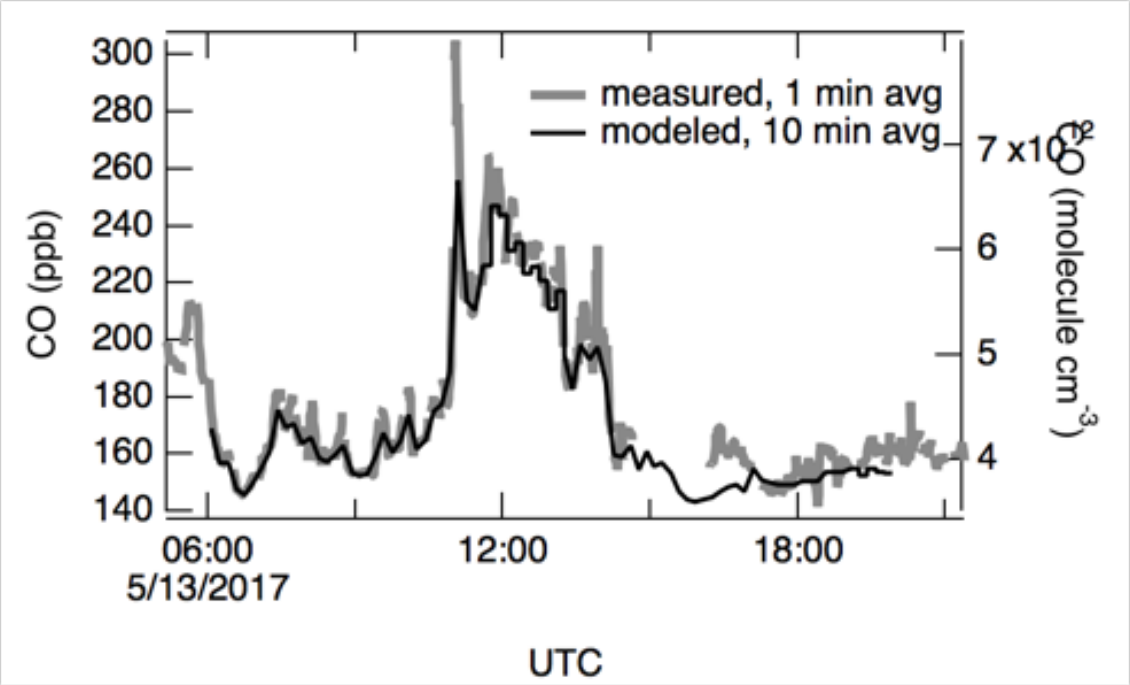

Figure 46. 0D model CO results versus measurements for a representative photochemical day in at UTSA.

The second type of result enabled by photochemical modeling is a breakout of the various species reacting with OH radical and contributing to the formation of ozone. The model, using the thousands of individual reactions included in the framework, can keep track of each OH radical formed, and the reacting parent VOC. The VOCs are grouped into source categories to produce pie charts of reactivity.

Figure 47 shows one such pie chart, apportioning OH reactivity at UTSA to categories like carbonyls (formaldehyde - HCHO and acetaldehyde - CH3CHO), intermediates from isoprene oxidation (methylvinylketone - MVK, and methacrolein - MACR). This figure shows that biogenics dominate OH reactivity at UTSA on the modeled days (5/12 – 5/14 2017), since a significant portion of the carbonyls are expected to be due to isoprene. An alternate view of this graph, with quantified OH reactivities, is shown below.

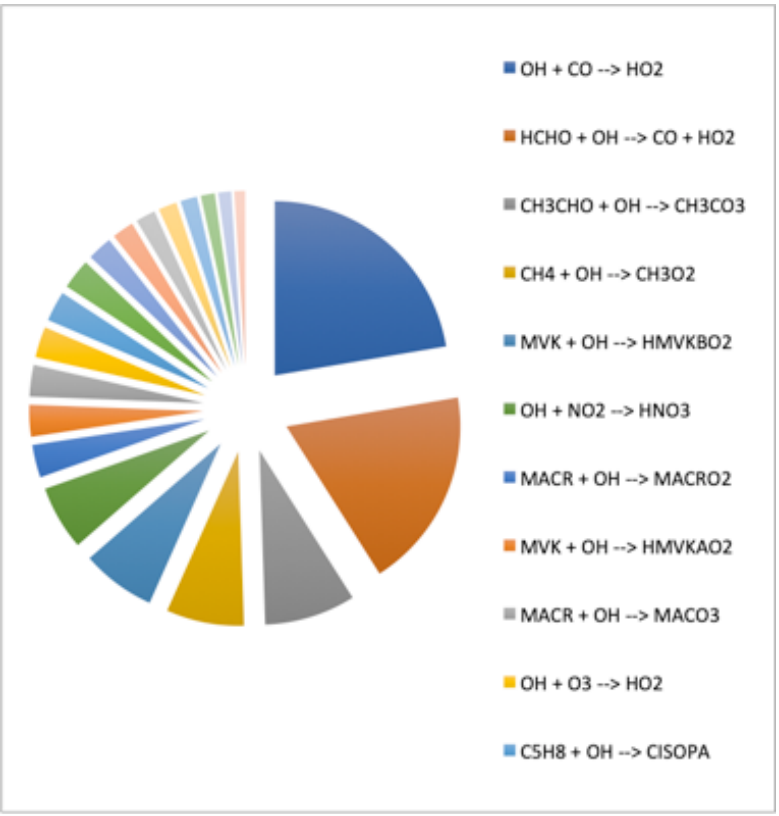

Figure 47. Apportionment of OH reactivity for a representative photochemical day at UTSA on 5/12-5/14 2017 (wind from the North). The pie chart shows contributions to OH reactivity from select VOC source categories. Categories include carbonyls (formaldehyde, HCHO and acetaldehyde, CH3CHO), intermediates from isoprene oxidation (methylvinylketone, MVK, and methacrolein (MACR)).

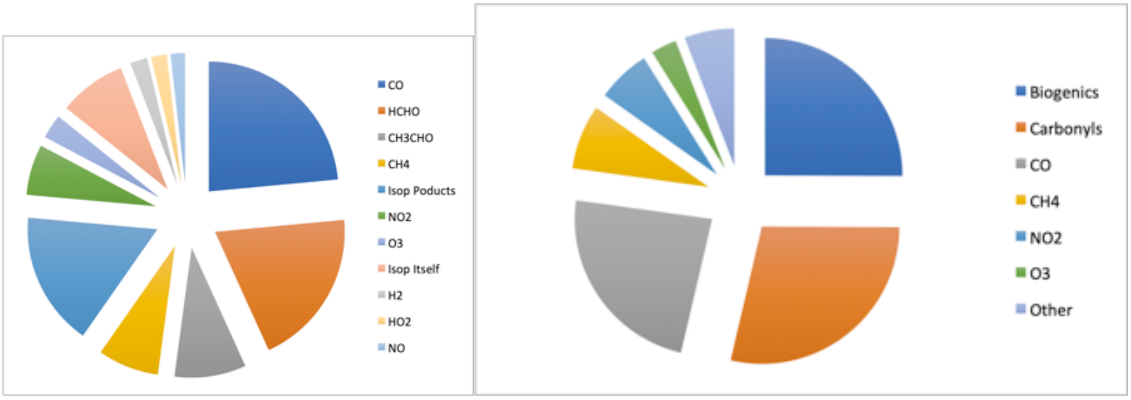

Figure 48. These two pie charts group the OH reactivity shown in Figure 47 into coarser and coarser source categories. The coarsest categories are chosen to match those listed by Anderson et al. [2019].

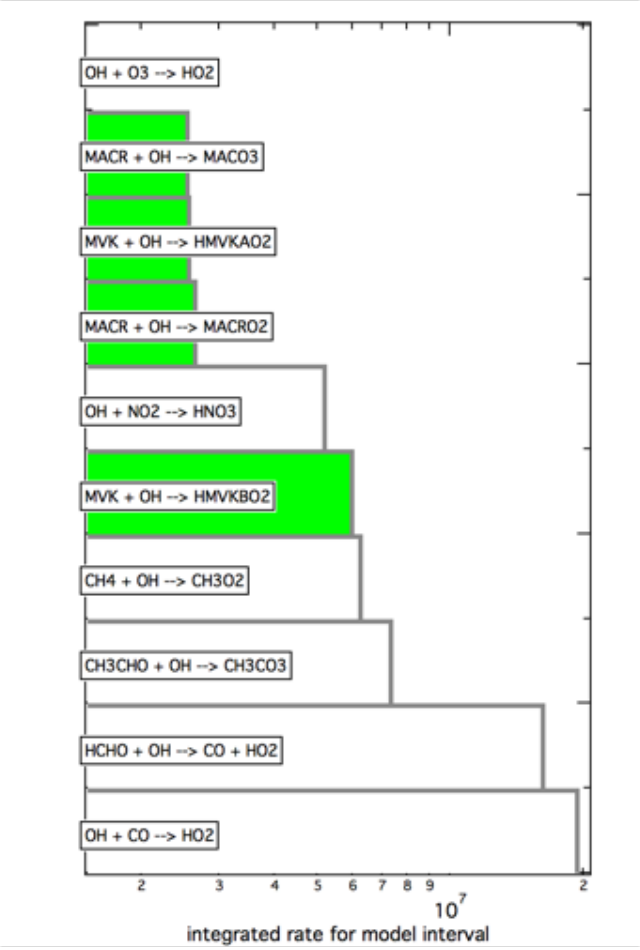

Figure 49. Initial Igor/DSMACC model result. Photostationary local noon on 5/13

Figure 49 above shows how important certain carbonyls like formaldehyde are to the total OH reactivity. Unlike MVK and MACR, which are predominantly biogenic, formaldehyde is formed via numerous different pathways. Thus, additional modeling work is required to further apportion formaldehyde to source categories. One such technique would involve nudging the modeled input of isoprene (biogenic) or toluene (urban/oil and gas) and observing the change in formaldehyde. This type of work is essentially a sensitivity analysis and allows additional apportionment beyond what the chemically explicit model can provide.

The result shown above is known to underestimate the total reactivity for a few reasons. First, the modeled mixing ratios of MVK + MACR have not been constrained to their measured values, which are approximately double the modeled values. Furthermore, longer chain alkanes (C4 and longer) have not been input. Both these additions would increase the total size of the “pie”.

Despite the shortfalls, this apportionment result agrees favorably with results calculated by Anderson et al. [2019] using entirely different model framework. The model mechanics developed as part of this project will be able to contribute significantly to the understanding of OH reactivity, particularly as it relates to carbonyls.

A second set of 0D calculations were done at the Floresville measurement location using the same methodology. A comparison of these two locations is shown in Figure 50 below. The two bars are the contribution to OH reactivity by species or compound class for the measurements at Floresville and the UTSA sites on the photochemical days 5/18/2017 and 5/13/2017, respectively. The full model state at photo-stationary state at 1700 UTC for each location was parsed according to the top 22 contributions. In the case of the first-generation isoprene oxidation compounds, their contribution was combined for clarity.

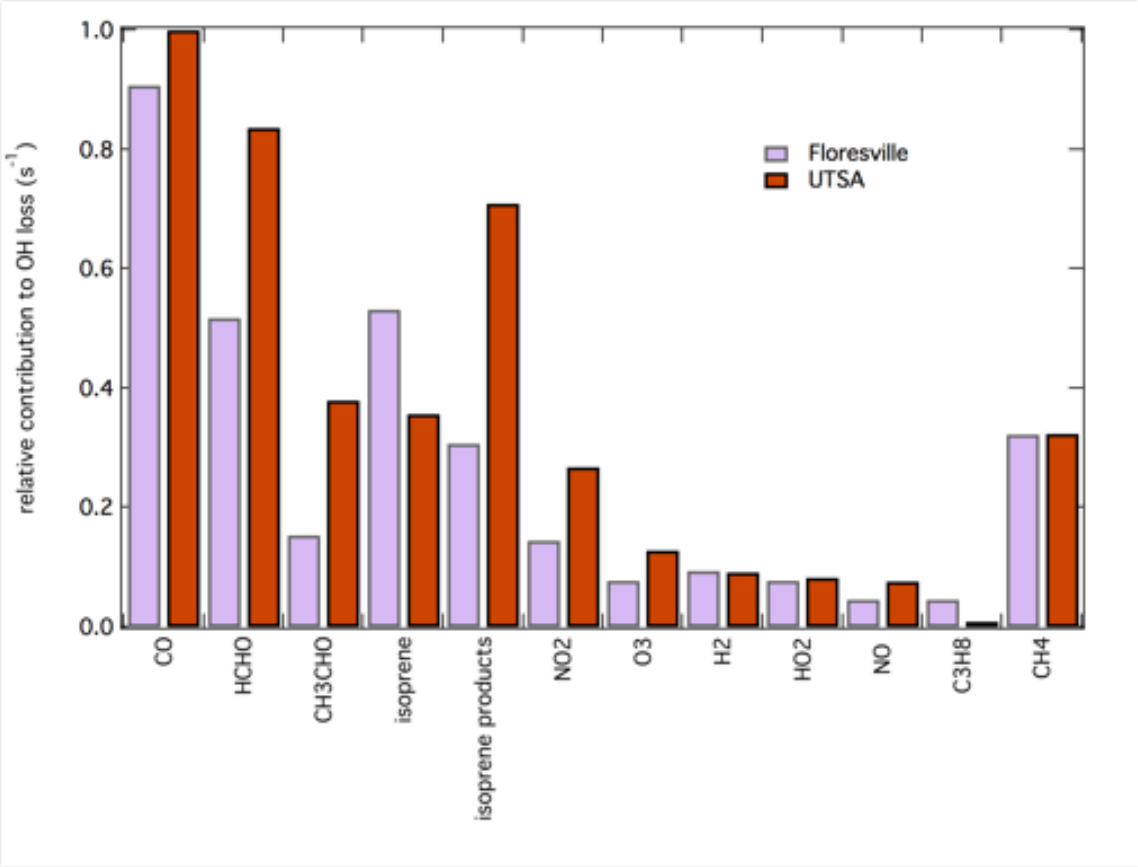

Figure 50. Relative contribution to OH loss by compound class. Classes include carbon monoxide (CO), formaldehyde (HCHO), acetaldehyde (CH_3CHO), isoprene and its products, propane (C_3H_8) and methane (CH_4), in addition to several ozone (O_3), NOx and HOx-related classes.

This graph shows several important differences between the two sites. First, the rate of OH formation (d[OH]/dt) at UTSA is ~ 32% greater than in Floresville, with the difference attributed primarily to formaldehyde (HCHO) and isoprene products. Second, propane, an oil and gas product, is detectable at Floresville, but is still minor relative to isoprene. Finally, though isoprene itself was greater in Floresville, the oxidation products and some contribution from formaldehyde suggest that the biogenic influence on OH-reactivity is strong at both locations.

Formaldehyde Sensitivity Analysis

One challenge with interpreting the model results is the dominance of formaldehyde (HCHO) in the apportionment results; HCHO is a product of numerous reaction channels from source VOCs from different source classes.

Here, we present preliminary results probing the sensitivity of the photochemical model's HCHO during peak photochemical hours at UTSA. In this analysis, the model is allowed to converged. Then, the converged atmospheric composition is used to re-initialize the model for a run at a peak photochemical hour (after noon), taking species *x* (e.g. C_5H_8, isoprene) as-is (sensitivity of 1.00), then altering species *x* concentration slightly (e.g. 95% of that value for a sensitivity of 0.95), etc. HCHO concentrations are not constrained during this process. The resulting HCHO is tabulated, leading to a sensitivity graph as shown below:

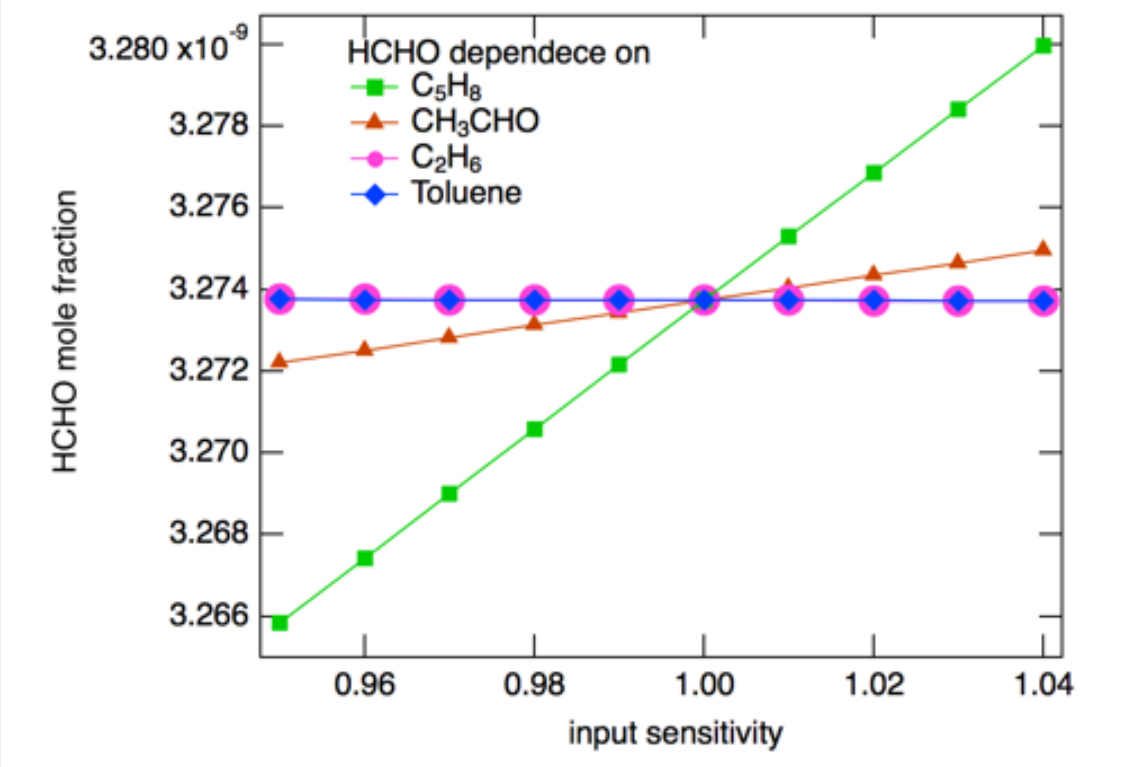

Figure 51. Formaldehyde (HCHO) sensitivity analysis for the UTSA 5/13 modeled photochemical day at peak photochemical times. The dependency on isoprene (C_5H_8), acetaldehyde (CH_3CHO), ethane (C_2H_6) and toluene are shown.

Note that above, the ethane and toluene sensitivities are nearly flat. In fact, their response is slightly negative. The reason is that this analysis was conducted for peak photochemical times. At these times, the radical intermediates generated from ethane and toluene aren't leading to formaldehyde quickly. Isoprene, on the other hand, generates HCHO quickly and this is apparent in the sensitivity analysis.

Discussion

A challenge in interpreting any atmospheric data is that it provides an incomplete window into the total chemical content of the atmosphere. Recent advances in mass spectrometry (i.e. HR TOF coupled to chemical ionization) have afforded more complete pictures of this chemical profile. Nonetheless, in this case the analysis was challenged by the lack of observed atmospherically processed chemical intermediates that could be linked to a single source category. Notably, no clear nitrogen-containing intermediate was found in the I-CIMS dataset that could be uniquely identified to a single source category such as oil and gas.

One thing that has become apparent in the high-resolution analysis of the I-CIMS dataset is the dominance of oxygenated products. While the I-CIMS detection scheme provides elemental formulas for individual chemical species, identification of the precise source that produced the detected species can be complicated because the same elemental formula can correspond to multiple structural isomers. Identification of the precise precursors that produced the oxidized products detected with the I-CIMS can also be complicated by the fact that oxidation processes result in fragmentation as well as functionalization [*Kroll et al.*, 2011]. The fragmentation process masks unique information about the parent's elemental formula since it produces molecular species that have smaller carbon skeletons than the original precursor. In recent laboratory experiments at Aerodyne, for example, we have studied the I-CIMS spectra obtained from the photooxidation of several aromatic precursors including 1,3,5-trimethyl benzene, 1,2,4-trimethyl benzene, 1-methyl naphthalene, propyl benzene and isopropyl benzene. Preliminary analysis indicates that many of the products formed from these different precursors result from fragmentation. These fragmentation products from the different precursors are observed to have identical elemental formulas. Fragmentation of the aromatic ring in these precursors also results in several small non-aromatic products that are also common to the oxidation of isoprene and terpenes as well. In this way, fragmentation can lead to confounding signals that are difficult to interpret without an understanding of oxidation and fragmentation pathways.

To get around these difficulties, the PMF analysis of the I-CIMS dataset leveraged laboratory and field experiments under known and controlled conditions to identify what we should expect to see in the mass spectrum when we are sampling aromatic oxidation or biomass burning, for example. This analysis was able to identify two clear biogenic signatures (isoprene and a-pinene oxidation), but the anthropogenic or oil and gas sources could not be conclusively identified, and they are identified as "mixed sources".

Conversely, PMF analysis leveraging directly emitted species (as measured by the gas-phase instrumentation and the GC-ToF) better addresses these non-biogenic source categories. In this combined instrument analysis, the ratios of chemical species in the PMF-derived factors was used to identify factors associated with anthropogenic and oil and gas activities. This analysis succeeds in identifying the impact of these sources on the measured airmass, but does not directly address ozone formation. Notably, the insights obtained here are therefore the consequence of using multiple sampling techniques – chemical ionization, gas chromatography, and precision laser-based methods – to provide a more complete picture of the relevant sources.

Ultimately, ozone formation in the San Antonio study area is best understood through the photochemical box modeling approach. The OH radical is a key intermediate in the set of equations that dictate ozone formation. Since the OH radical is not a measured species, it is modeled. In doing this modeling, we take care to track each individual OH radical formed, and its source VOC. This modeling supports the conclusion that biogenic emissions are crucial in explaining ozone formation in San Antonio.

This modeling has also revealed the importance of carbonyls such as formaldehyde (HCHO) in ozone formation chemistry. HCHO presents a challenge in interpreting OH reactivity pie charts because it is produced via numerous chemical pathways with different types of source VOCs. The HCHO sensitivity analysis that we conducted shows, again, the dominance of biogenic species such as isoprene at the UTSA site at peak photochemical times (mid-day). In this case, small changes in isoprene have a strong impact on the resulting HCHO. Other species like toluene or ethane, which can also produce HCHO, have nearly no impact.

Importantly, the box modeling approach used here is a 0D chemically explicit approach – this means that we are examining the results of a steady-state atmosphere with full chemical specificity. However, this approach does not couple to atmospheric transport or emissions models. A separate investigation on airmass sources was done for this campaign using the results from Hysplit back-trajectory analyses. Land-cover was incorporated into this investigation in order to estimate the time periods in the campaign that were most likely to be influenced by air originating from nearby oil and gas plays, from biogenic sources, or from upwind cities like Corpus Christi. These efforts revealed instances in which biogenic emissions contributed to the measured airmass, but did not fully account for biogenic impact at UTSA, highlighting the impact of the hyper-local sampling environment (oak trees on campus) at the UTSA site.

Conclusions

Ozone is formed when NOx and VOCs react in the presence of sunlight. In order to better understand ozone formation in the SAFS study area, this project looks at the VOC half of the equation. VOCs reacting with OH radical kick off the canonical series of ozone-generating reactions. Thus, by tracing, quantifying and apportioning the formation of OH, we can better understand and quantify the causes of ozone formation in and around San Antonio.

We explored and analyzed the entirety of high-resolution datasets from the I-CIMS, PTR-ToF and GC-EI-ToF, three instruments that targeted different classes of atmospheric compounds. Previous efforts had only skimmed the surface, hand-picking species to report.

Positive Matrix Factorization, a mathematical technique, was crucial in identifying co-varying species among the hundreds of individual high-resolution signals. It was also run in a way that combined data from multiple instruments, and yielded interpretable factors, like diurnal trends due to isoprene, and location-specific signals due to oil and gas impact.

One challenge in interpreting the data was the lack of atmospherically processed chemical intermediates that could be linked to a single source category. To get around this fact, a 0D chemical model was used to track OH formation given a set of measured constraints.

The main conclusion of this report is that biogenic sources of VOCs dominate OH reactivity in the SAFS study area.

Recommendations for Future Work

Results presented within this report motivate a number of additional research avenues.

First, additional investigation is warranted on emissions in the SAFS study area of biogenic ozone precursors like isoprene. Such research might include work on emissions inventories, which could be coupled to the spatial distribution of sources and their impact on various downwind areas (perhaps similar to the coupled Hysplit and land use work presented here).

Second, the use of multiple high-resolution mass-spectrometer based instrumentation provides a superior level of detail on the composition of the atmosphere. However, this amount of data comes at a cost, since manual identification and calibration of individual time traces is time consuming and borderline impossible for some of the larger datasets. Instead, use of methodologies that examine correlations amid the entire mass spectrum (PMF, R^2 plots) should be used early in the process to identify which high-resolution traces are of greatest interest.

On a practical side, SAFS measurements at three distinct locations, each conducted for enough time to experience variations in meteorology, was crucial in many of the conclusions presented here. Colocation with the AutoGC in Floresville presented an excellent opportunity for data validation and intercomparison. Future measurement campaigns should consider this strategy.

Gaps in measured results due to instrumental problems presented challenges when using SAFS data in modeling. Efforts to maximize data coverage for all instruments are warranted in future campaigns, as are efforts to determine reasonable default concentrations for the study area for each reported species.

Finally, an analysis of potential ozone mitigation strategies for the SAFS study area is warranted, underpinned by the scientific results presented here, but including policy and economic considerations unique to the San Antonio area.

References

Abeleira, A., I. B. Pollack, B. Sive, Y. Zhou, E. V. Fischer, and D. K. Farmer (2017), Source characterization of volatile organic compounds in the Colorado Northern Front Range Metropolitan Area during spring and summer 2015, *J. Geophys. Res., [Atmos.]*, *122*(6), 3595-3613, doi:10.1002/2016jd026227.

Cappellin, L., T. Karl, M. Probst, O. Ismailova, P. M. Winkler, C. Soukoulis, E. Aprea, T. D. Märk, F. Gasperi, and F. Biasioli (2012), On Quantitative Determination of Volatile Organic Compound Concentrations Using Proton Transfer Reaction Time-of-Flight Mass Spectrometry, *Environ. Sci. Technol.*, *46*(4), 2283-2290, doi:10.1021/es203985t.

City of San Antonio, GIS (Geographic Information Services): Downloadable Maps. https://gis.sanantonio.gov/Download/CoSABoundary.zip (Accessed 04/01/2019).

Edwards, P. M., et al. (2013), OH reactivity in a South East Asian tropical rainforest during the Oxidant and Particle Photochemical Processes (OP3) project, *Atmos. Chem. Phys.*, *13*(18), 9497-9514, doi:10.5194/acp-13-9497-2013.

Ehn, M., et al. (2012), Gas phase formation of extremely oxidized pinene reaction products in chamber and ambient air, *Atmos. Chem. Phys.*, *12*(11), 5113-5127, doi:10.5194/acp-12-5113-2012.

Elliott, L. F., A. Treuer-Kuehn, C. F. Blodgett, C. D. True, D. German, and D. D. Diamond, Ecological Systems of Texas: 391 Mapped Types. Texas Parks & Wildlife Department and Texas Water Development Board: Austin, Texas, 2014. http://www.tpwd.state.tx.us/gis/data/downloads#EMS-T (Accessed 04/01/2019).

Emmerson, K. M., and M. J. Evans (2009), Comparison of tropospheric gas-phase chemistry schemes for use within global models, *Atmos. Chem. Phys.*, *9*(5), 1831-1845, doi:10.5194/acp-9-1831-2009.

Gaston, C. J., F. D. Lopez-Hilfiker, L. E. Whybrew, O. Hadley, F. McNair, H. Gao, D. A. Jaffe, and J. A. Thornton (2016), Online molecular characterization of fine particulate matter in Port Angeles, WA: Evidence for a major impact from residential wood smoke, *Atmos. Environ.*, *138*, 99-107, doi:https://doi.org/10.1016/j.atmosenv.2016.05.013.

Gilman, J. B., B. M. Lerner, W. C. Kuster, and J. A. de Gouw (2013), Source Signature of Volatile Organic Compounds from Oil and Natural Gas Operations in Northeastern Colorado, *Environ. Sci. Technol.*, *47*(3), 1297-1305, doi:10.1021/es304119a.

Gueneron, M., M. H. Erickson, G. S. VanderSchelden, and B. T. Jobson (2015), PTR-MS fragmentation patterns of gasoline hydrocarbons, *Int. J. Mass Spectrom.*, *379*, 97-109, doi:https://doi.org/10.1016/j.ijms.2015.01.001.

Herndon, S. C., et al. (2008), Correlation of secondary organic aerosol with odd oxygen in Mexico City, *Geophysical Research Letters*, *35*(15), doi 10.1029/2008GL034058, doi:10.1029/2008GL034058.

Kari, E., P. Miettinen, P. Yli-Pirilä, A. Virtanen, and C. L. Faiola (2018), PTR-ToF-MS product ion distributions and humidity-dependence of biogenic volatile organic compounds, *Int. J. Mass Spectrom.*, *430*, 87-97, doi:https://doi.org/10.1016/j.ijms.2018.05.003.

Kroll, J. H., et al. (2011), Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, *Nature Chem.*, *3*, 133, doi:10.1038/nchem.948.

Lanz, V. A., M. R. Alfarra, U. Baltensperger, B. Buchmann, C. Hueglin, and A. S. H. Prévôt (2007), Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, *Atmos. Chem. Phys.*, *7*(6), 1503-1522.

Lee, B. H., F. D. Lopez-Hilfiker, C. Mohr, T. Kurtén, D. R. Worsnop, and J. A. Thornton (2014), An Iodide-Adduct High-Resolution Time-of-Flight Chemical-Ionization Mass Spectrometer: Application to Atmospheric Inorganic and Organic Compounds, *Environ. Sci. Technol.*, *48*(11), 6309-6317, doi:10.1021/es500362a.

Lee, B. H., et al. (2016), Highly functionalized organic nitrates in the southeast United States: Contribution to secondary organic aerosol and reactive nitrogen budgets, *Proc. Nat. Acad. Sci.*, *113*(6), 1516-1521, doi:10.1073/pnas.1508108113.

Lopez-Hilfiker, F. D., S. Iyer, C. Mohr, B. H. Lee, E. L. D'Ambro, T. Kurtén, and J. A. Thornton (2016), Constraining the sensitivity of iodide adduct chemical ionization mass spectrometry to multifunctional organic molecules using the collision limit and thermodynamic stability of iodide ion adducts, *Atmos. Meas. Tech.*, *9*(4), 1505-1512, doi:10.5194/amt-9-1505-2016.

Massoli, P., et al. (2012), Pollution Gradients and Chemical Characterization of Particulate Matter from Vehicular Traffic near Major Roadways: Results from the 2009 Queens College Air Quality Study in NYC, *Aerosol Sci. and Technol.*, *46*(11), 1201-1218, doi:DOI: 10.1080/02786826.2012.701784.

McManus, J. B., J. H. Shorter, D. D. Nelson, M. S. Zahniser, D. E. Glenn, and R. M. McGovern (2008), Pulsed quantum cascade laser instrument with compact design for rapid, high sensitivity measurements of trace gases in air, *Appl. Phys. B*, *92*(3), 387, doi:10.1007/s00340-008-3129-9.

Mohr, C., et al. (2013), Contribution of Nitrated Phenols to Wood Burning Brown Carbon Light Absorption in Detling, United Kingdom during Winter Time, *Environ. Sci. Technol.*, *47*(12), 6316-6324, doi:10.1021/es400683v.

Murschell, T., S. R. Fulgham, and D. K. Farmer (2017), Gas-phase pesticide measurement using iodide ionization time-of-flight mass spectrometry, *Atmos. Meas. Tech.*, *10*(6), 2117-2127, doi:10.5194/amt-10-2117-2017.

Nelson, D. D., Jr., J. B. McManus, S. C. Herndon, J. H. Shorter, M. S. Zahniser, S. Blaser, L. Hvozdara, A. Muller, M. Giovannini, and J. Faist (2006), Characterization of a

near-room-temperature, continuous-wave quantum cascade laser for long-term, unattended monitoring of nitric oxide in the atmosphere, *Opt. Lett.*, *31*, 2012-2014.

Onasch, T. B., A. Trimborn, E. C. Fortner, J. T. Jayne, G. L. Kok, L. R. Williams, P. Davidovits, and D. Worsnop (2012), Soot Particle Aerosol Mass Spectrometer: Development, Validation, and Initial Application, *Aerosol Sci. and Technol.*, *46*(7), 804-817, doi:DOI:10.1080/02786826.2012.663948.

Paatero, P. (1997), A weighted non-negative least squares algorithm for three-way 'PARAFAC' factor analysis, *Chemometrics Intell. Lab. Syst.*, *38*(2), 223-242.

Paatero, P., and U. Tapper (1994), Positive Matrix Factorization - a Nonnegative Factor Model with Optimal Utilization of Error-Estimates of Data Values, *Environmetrics*, *5*(2), 111-126.

Santoni, G. W., B. H. Lee, E. C. Wood, S. C. Herndon, R. C. Miake-Lye, S. C. Wofsy, J. B. McManus, D. D. Nelson, and M. S. Zahniser (2011), Aircraft Emissions of Methane and Nitrous Oxide during the Alternative Aviation Fuel Experiment, *Environmental Science & Technology*, *45*(16), 7075-7082, doi:10.1021/es200897h.

Schobesberger, S., et al. (2016), High upward fluxes of formic acid from a boreal forest canopy, *Geophys. Res. Lett.*, *43*(17), 9342-9351, doi:10.1002/2016gl069599.

Sekimoto, K., S.-M. Li, B. Yuan, A. Koss, M. Coggon, C. Warneke, and J. de Gouw (2017), Calculation of the sensitivity of proton-transfer-reaction mass spectrometry (PTR-MS) for organic trace gases using molecular properties, *Int. J. Mass Spectrom.*, *421*, 71-94, doi:10.1016/j.ijms.2017.04.006.

U.S. Energy Information Administration, Maps: Low permeability oil and gas play boundaries in Lower 48 States. 2016. https://www.eia.gov/maps/maps.htm (Accessed 04/16/2019).

Ulbrich, I. M., M. R. Canagaratna, Q. Zhang, D. R. Worsnop, and J. L. Jimenez (2009), Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, *Atmos. Chem. Phys.*, *9*(9), 2891-2918.

Veres, P. R., J. M. Roberts, R. J. Wild, P. M. Edwards, S. S. Brown, T. S. Bates, P. K. Quinn, J. E. Johnson, R. J. Zamora, and J. de Gouw (2015), Peroxynitric acid (HO₂NO₂) measurements during the UBWOS 2013 and 2014 studies using iodide ion chemical ionization mass spectrometry, *Atmos. Chem. Phys.*, *15*(14), 8101-8114, doi:10.5194/acp-15-8101-2015.

Wood, E. C., et al. (2009), A case study of ozone production, nitrogen oxides, and the radical budget in Mexico City, *Atmospheric Chemistry and Physics*, *9*(7), 2499-2516.

Yacovitch, T. I., S. C. Herndon, G. Pétron, J. Kofler, D. Lyon, M. S. Zahniser, and C. E. Kolb (2015), Mobile Laboratory Observations of Methane Emissions in the Barnett Shale Region, *Environ. Sci. Technol.*, *49*(13), 7889-7895, doi:10.1021/es506352j.

Yuan, B., A. R. Koss, C. Warneke, M. Coggon, K. Sekimoto, and J. A. de Gouw (2017), Proton-Transfer-Reaction Mass Spectrometry: Applications in Atmospheric Sciences, *Chemical Reviews*, *117*(21), 13187-13229, doi:10.1021/acs.chemrev.7b00325.

Appendix A: High-Resolution Dataset Description

High-Resolution Fits: New primary data (time-series from high-resolution fits) that will be produced as part of this project are described in this appendix, to accompany the final data deliverable.

The data description includes, if relevant, the following items: instrument that collected the data, units, reason for any missing data, how the data was analyzed (high-resolution fit,), calibration factor or response factor applied, and source of the factor (measured cal tank, literature reaction rate, etc.), interferences identified, eliminated, or suspected, and uncertainty (see descriptive statistics below).

Uncertainty (precision) will be reported at the one sigma level for each of the various time series generated as a result of the high-resolution fits. Such statistics will be preferentially determined using segments of data when concentrations are low and stable, or, when the instrument was sampling from a stable source such as a zero-air tank or calibration cylinder. When comparable data is available (e.g. from TCEQ monitoring sites) it will further be used to assess accuracy and correct any bias of the data.

The total systematic uncertainty at 95% confidence limits will be the combination of the uncertainty and the uncertainty of the calibration standard used in-field, pre- and post-campaign. All errors will be accounted for and estimated.

Calculated sensitivities and uncertainties for the GC-TOF-MS instrument during 2017 Texas AQRP

Introduction

Instrument sensitivity for the Aerodyne Research, Inc. GC-TOF-MS is determined in a two-step process. First, instrument response is normalized on a per-chromatogram basis, using the measurement of CCl4, a halocarbon with invariant tropospheric mixing ratio on the time-scale of the SAFS campaign. Instrument sensitivities are then determined as normalized counts, based upon instrument response to calibration standards routinely measured throughout the sample period.

A calibration scheme using three methods was designed for Texas AQRP:

1. Frequent (twice daily) in-field measurement of a custom-blended field standard (Cylinder CC358900), a diluted natural gas standard (Restek natural gas standard #2, p/n 34439). This serves to provide a single-point measure of instrument response, as well as allow for retention-time characterization. Uncertanties of compound mixing ratios for this tank were 7-12%.

2. Bi-weekly In-field multi-point calibration of Instrument response to PTR-TOF calibration standard (Cylinder CC428331, Apel-Riemer Environmental, Inc.), uncertainties 5%, diluted in zero air via two mass flow controllers and added to the sample inlet in excess. Total uncertainty for this method is geometric sum of 5% (tank) + 2% (MFC #10 + 2% (MFC #2) = 6%.
3. Post-campaign measurement of the transfer standard (CC358900) used in (1), at various lengths of time flowing the standard through the sample trap (1-5 minutes) to generate a multi-point calibration.

Additionally, the instrument was co-located with a TCEQ AutoGC instrument during the active sampling period in Floresville, TX. This allows for inter-comparison of observed mixing ratios for a subset of reported species, albeit with disparate sample periods.

Field Results

The instrument was operated with two (2) sample traps during the campaign, therefore calibrations are required for both traps. The first trap was pulled from the instrument at the mid-point of the campaign (19-May) to preserve it for post-campaign calibration. The second trap was installed on 19-May, showed very poor sensitivity to lower volatility compounds and then failed – lost adsorbent packing – on 26-May. The GC sample path was determined to be contaminated with the adsorbent, at which point it was pulled from operation for the remainder of the campaign. Post-campaign work was completed with the first sample trap (operated 12-May to 19-May) only.

The instrument showed a strong overall response variance that was dependent upon trailer temperature (Figure 1, left panel). We used CCl_4 response (since CCl_4 is effective constant in the air – sampled during SAFS) as a scale factor to correct for this sensitivity change. To scale instrument response, we used an arbitrary raw response of 0.39 mV-sec. The scaling factor for any chromatogram is simply calculated as the ratio of CCl_4 raw response to 0.39 (Figure 1, right panel). All other integrated peak areas are divided by the scaling factor to correct instrument sensitivity. When CCl_4 is not present in a sample at ambient mixing ratio, i.e. a calibration point, we interpolate the scale factor from ambient measurements before and after the calibration sample. The scaling factor for the GC was determined for both sample traps across the entire field campaign, through 26-May.

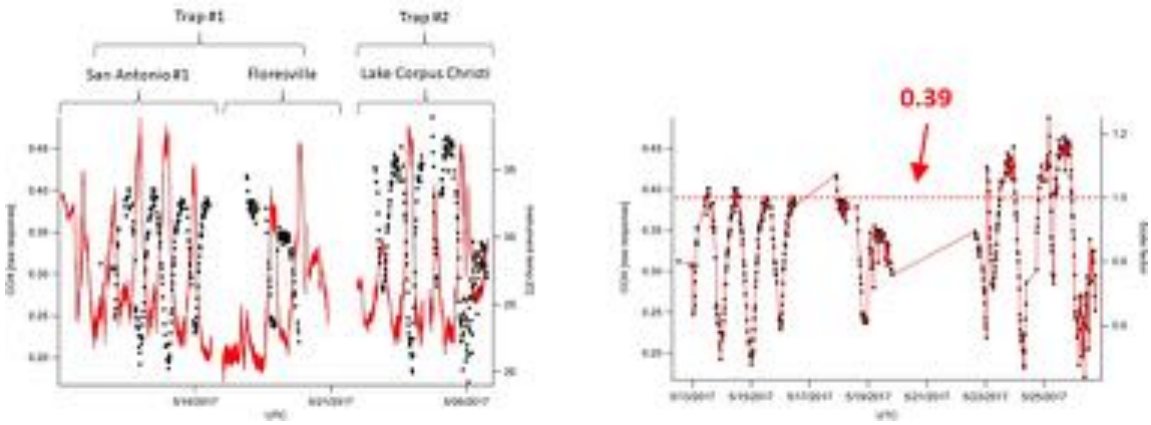

Figure S1. GC Scaling Factor.

Single-point field calibration

These measurements did not serve as valid calibration tests, due to differences in sample pressure between the calibration samples and ambient samples. The sample trap was exposed to variable super-ambient pressure from the calibration gas flow.

The data from these calibration points were valuable as retention time markers, especially with the severe time-shifting experienced due to the wide range of trailer temperatures during Texas AQRP.

The instrument flow-path was revised to allow for inlet overflow (at the instrument inlet) to allow for these calibrations to become quantitative (hopefully).

Multi-point field calibration

Calibrations were performed on the VOC sample inlet (shared by the PTR-TOF and GC-TOF) at irregular intervals throughout the campaign. The calibrant (CC428331) consists of the following components (all at ±5% uncertainty):

Table S1. VOC calibration tank CC428331 composition.

Species	Mixing Ratio (ppbv)
Acetaldehyde	53
Methanol	495
Acetone	504
Acetonitrile	498
Benzene	502
Toluene	494
p-Xylene	491
1,2,4-Trimethylbenzene	470

Note that the last 4 compounds on the list – in bold – are the only species that were actually identified for reporting by the GC-TOF. The GC sampling system was designed for hydrocarbon analysis only for this campaign.

Calibrations were performed at the following times, typically via standard addition of cal gas into 4500 sccm zero air. Total inlet intake flow was 4000 sccm for the campaign. All flows were controlled via MFC (calibrated at the start of the campaign by B. Knighton, U. Montana).

Calibrations occurred on the following dates / times:

13-May: 1450 – 1610 UTC
15-May: 1420 – 1540
15-May: 2220 – 2240
19-May: 1700 – 1820
22-May: 1950 – 2050 (First multi-point cal on second trap)
23-May: 1510 – 1550
25-May: 1400 – 1440

These calibrations provide four measures of the instrument response for the first sample trap, and 3 measures for the second trap. The results of these calibrations are shown in Figure 2. Also shows in the panels for benzene, toluene and p-xylene from the post-campaign calibrations, discussed below.

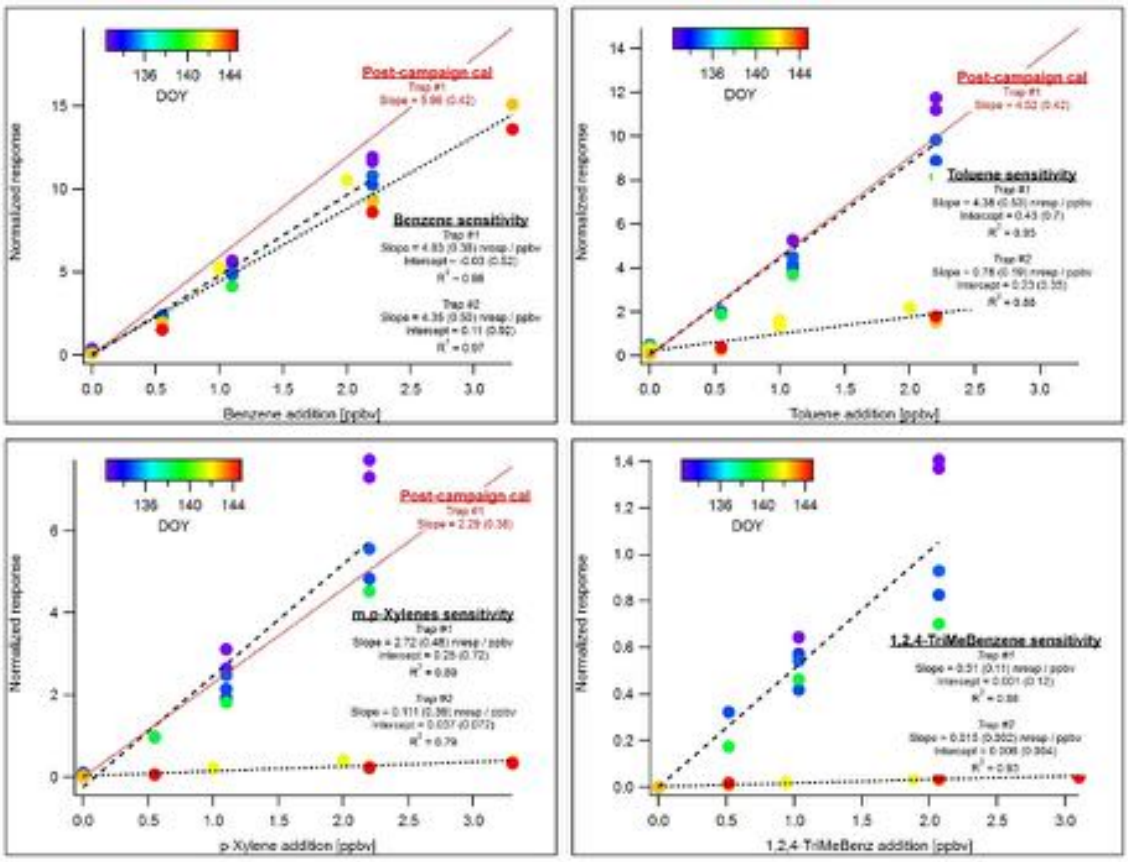

Figure S2. GC Column calibrations

Post-campaign calibrations

Post-campaign calibrations were performed on the first sample trap via the dilute natural gas standard (cylinder CC358900) (see Table 1 below). An example of the results of these calibrations is in Figure 3, which shows calibrations over two days at multiple standard additions.

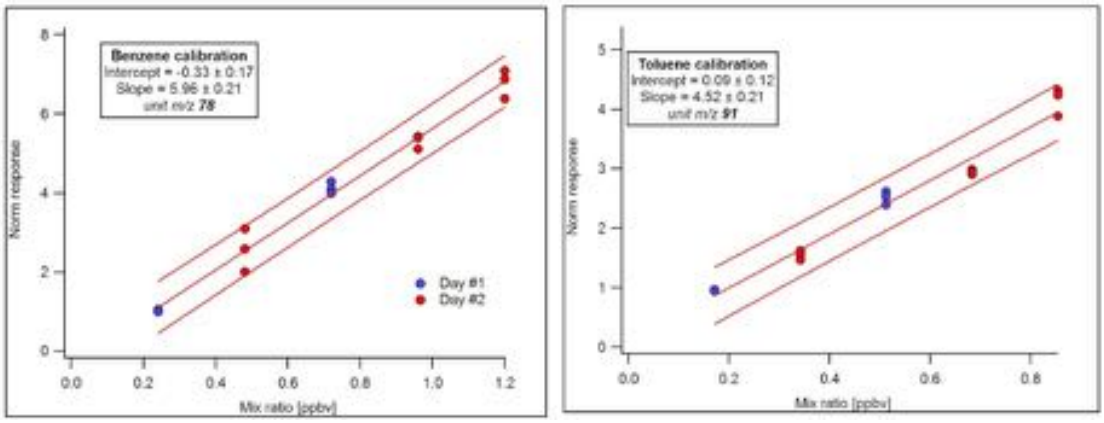

Figure S3. Benzene and toluene calibrations post-campaign

Results from the set of calibrations are shown in Table S2 below. In this case, the uncertainty refers to error of the linear fit of the normalized calibration points versus mixing ratio. We estimate LOD by using the nominal minimum detectable signal for the TOF multiplied by three = 0.005 normalized counts. To estimate the total uncertainty for the campaign, we include the uncertainties of the calibration tank mixtures (described above) and the difference in lab and field sensitivity when available.

Table S2. Species mixing ratio for calibration cylinder CC358900. Post-campaign measured instrument sensitivity and uncertainty for each species, in units of ppb / norm counts. LOD is determined using 0.005 norm counts (see text). Total uncertainty includes additional estimates of error.

Compound	Mixing Ratio	Sens	Unc	LOD	Total Unc
Pentane, iso-	102 ppb	1.68	0.07	0.003	10%
Pentane, n-	102	2.96	0.09	0.002	11%
Hexane, n-	21.4	2.06	0.05	0.002	9%
Hexane, 2-methyl-	1.57	6.3	0.18	0.001	19%
Heptane, n-	4.99	1.56	0.04	0.010	8%
Octane, n-	0.59	0.82	0.04	0.006	8%
Pentane, 2,2,4-trimethyl	0.22	4.54	0.15	0.001	17%
Cyclopentane, methyl-	3.56	2.59	0.09	0.002	11%
Cyclohexane	2.95	1.13	0.03	0.004	8%
Cyclohexane, methyl-	0.60	1.97	0.06	0.003	9%
Benzene	1.19	5.96	0.21	0.001	22%
Toluene	0.85	4.52	0.21	0.001	22%
Ethylbenzene	0.14	2.61	0.19	0.002	20%
Xylenes, m- and p-	0.25	2.29	0.19	0.002	20%
Xylene, o-	0.08	0.82	0.06	0.006	9%

Isoprene sensitivity

No isoprene calibration standard was available during the Texas AQRP campaign, or immediately before or after. We were able to obtain a multi-component gas standard in December-2017 from Apel-Riemer Environmental, with 15 components at nominal 1 ppm mixing ratio and actual concentrations certified to 5% uncertainty. This standard includes 5 compounds also quantified in the GC natural gas calibration standard, allowing for overlap comparison between the standards.

In Jan to March, 2018, the Aerodyne Research, Inc. GC-EI-TOFMS was deployed to the MIT environmental chamber as a part of the Total-C oxidation experiment. We took this opportunity to run a set of calibrations with both gas standards on the GC system. It is worth noting here that GC sensitivity can vary due to differences in the detector sensitivity, which is dependent upon filament excitation current and MCP voltage (and other variables). Since filament and MCP setting were changed during the MIT experiment, we typically ran a cal at the start and end of each day, using the NG standard for a full 5-minute sample, and a cal daily during Texas.

First, the natural gas standard was used to demonstrate linearity and establish relative sensitivities. The following day, we ran a second set of calibrations with the Apel-Riemer standard. The detector MCP voltage was increased after the first day, and the instrument sensitivity was found to have increase by 45% (on average). The signal from the first day was therefore multiplied by a factor of 1.45 to allow inter-comparison between the tanks. Note that calibration via the NG standard is performed by varying the sample collection time (1-5 min) while the A-R standard mixing ratio was changed via dilution (cal MFC setting) and collected for the full 5 min. These results indicate that these methods agree well, and we could extend dilution dynamic range by combining these methods.

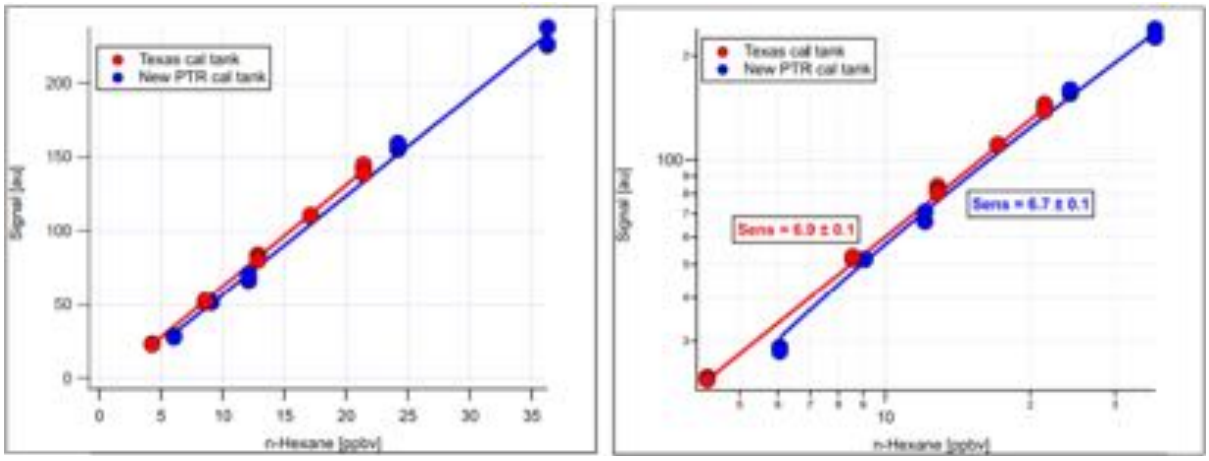

Figure S4. Example (n-hexane)

These plots show the comparison between the NG standard (labeled “Texas” here) calibration, using a 1.45x multiplication factor, and the Apel-Riemer tank (labeled “New PTR”) without a multiplication factor. Both experiments show very good linearity over the decadal range, as expected. The sensitivities agree within 3%, well within the overall uncertainty for the tanks.

We performed this inter-comparison with other species common to both tanks; unfortunately, only n-hexane allows for overlapping measured mixing ratios.

Examples:

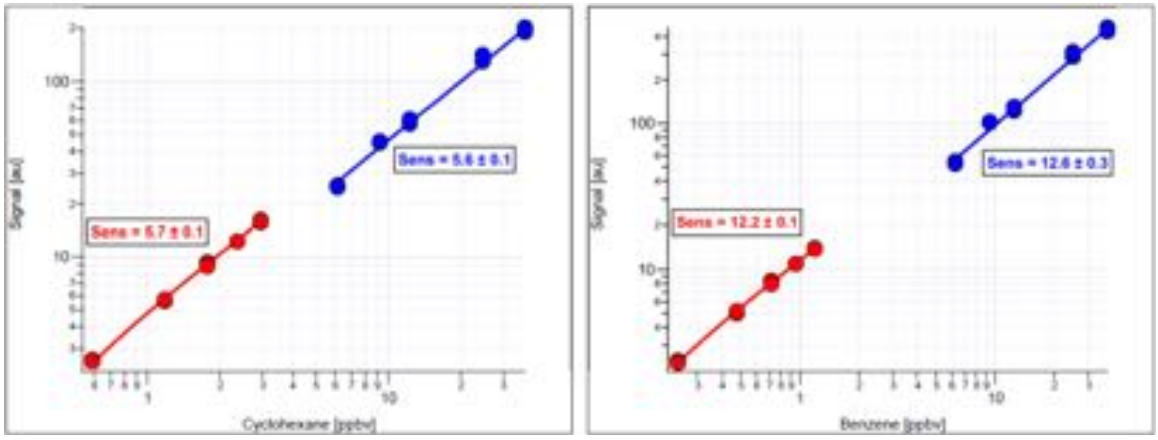

Figure S5. Example (cyclohexane and benzene)

Here, the color scheme is the same as above (red = NG; blue = A-R). Again, the signal is linear, now extended multiple decades of mixing ratio. Again, the sensitivities agree within 3% or better.

To extrapolate between the experiments at MIT and the field data collected during 2017 Texas AQRP, we need to compare the relative sensitivities measured. This was performed for 6 species (iso-pentane, n-pentane, n-hexane, methylcyclopentane, cyclohexane, benzene):

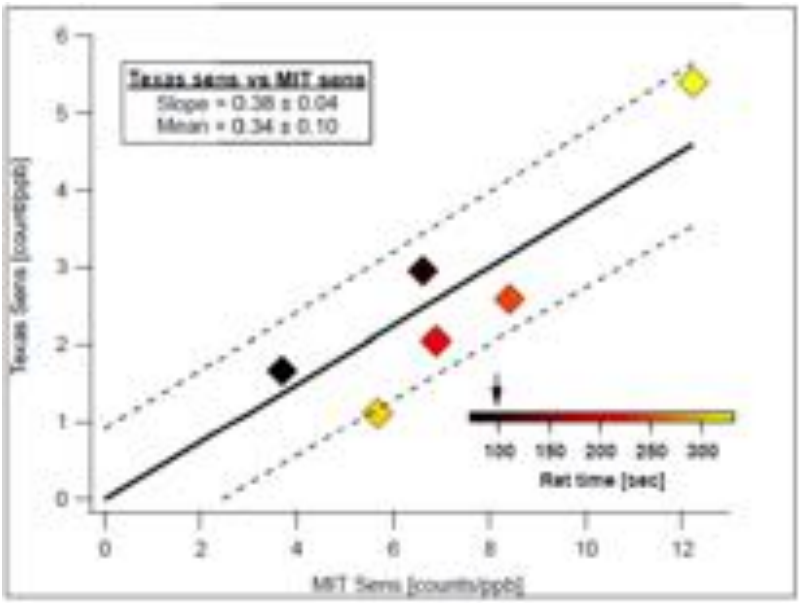

Figure S6. Comparison between experiments at MIT and field data during SAFS

This inter-comparison is not very satisfactory, as we see significant variance between species. Markers are color-coded by retention time (compounds were ordered by RT above), and this doesn't seem to be a significant cause of the variance. It is worth noting

that the sample trap was operated at 30°C during Texas AQRP but at 15°C during the MIT study. It is not known how this effects the various sensitivities but could account for the differences observed here.

The GC sensitivity to isoprene was measured as per the other species discussed above via the A-R cal standard:

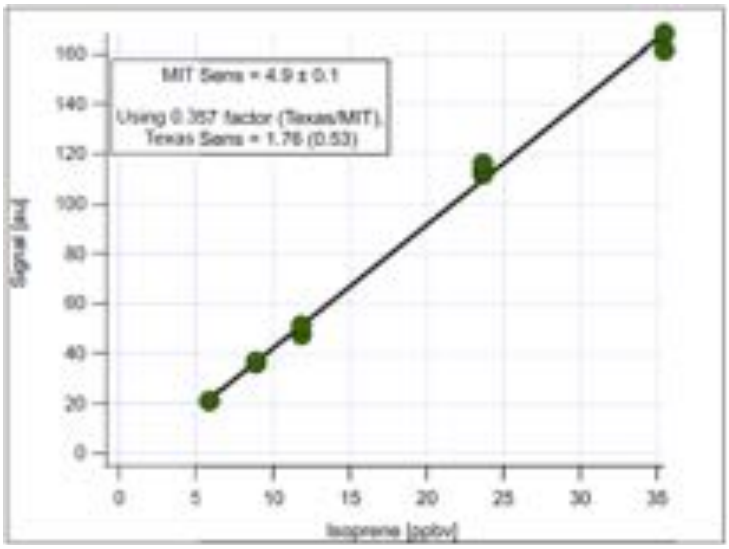

Figure S7. GC sensitivity to isoprene

Since we have no *a priori* sensitivity estimate for isoprene, the total uncertainty for extrapolating from MIT to Texas will be estimated from the average difference shown in the figure above (0.10 / 0.34 = 30%), and the factor used will be 0.36 (between the fitted line and the average value). Uncertainty is the geometric sum of uncertainties of the cal mixing ratio, tank uncertainty, linear fit and conversion factor (5%, 5%, 2%, 29%), which is 31%. This yields a sensitivity of 1.76 (0.55) counts / ppbv. This is the sensitivity used for the 2017 Texas AQRP isoprene mixing ratios reported. Using the standard resolvable peak height of 0.005 counts, this yields a LOD of 3 pptv.

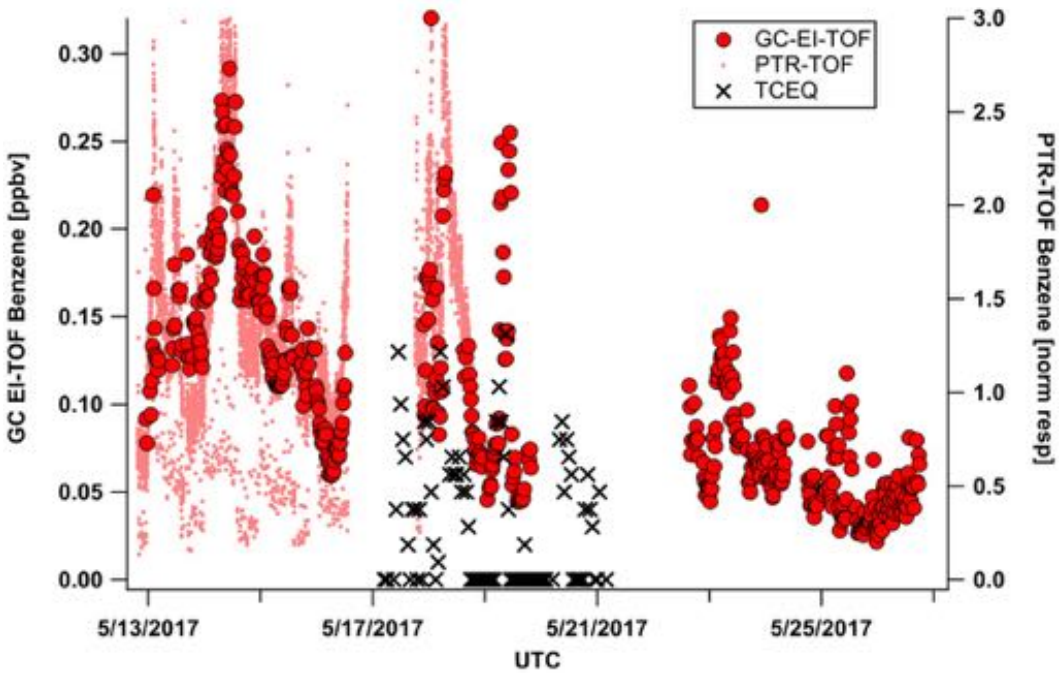

Figure S8. Comparison of GC-ToF, PTR-ToF and TCEQ AutoGC measurements of Benzene.

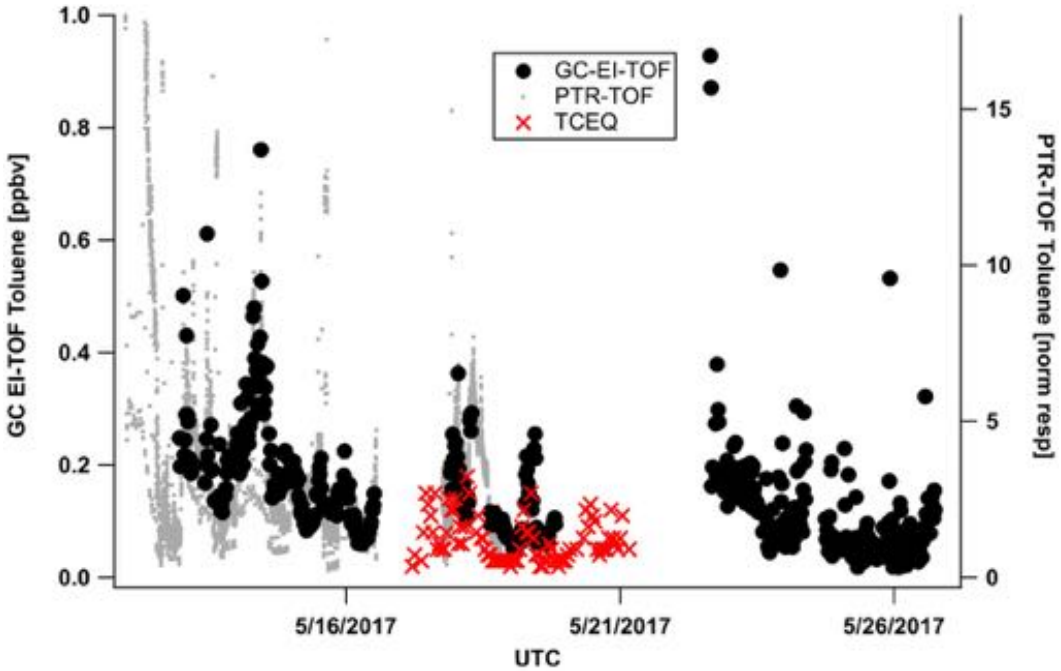

Figure S9. Comparison of GC-ToF, PTR-ToF and TCEQ AutoGC measurements of Toluene.

PTR-ToF-MS Operation and calibration

The PTR-ToF-MS instrument utilized in the SAFS study was an instrument on loan from Environment Canada and was operated using the recommended parameters provided in Table S3

Table S3. PTR-ToF-MS Instrument Parameters

Parameter	Value
Drift Tube Pressure	2.15 mbar
Drift Tube Temperature	60 ^{o}C
E/N	141 Td
Inlet flow into PTR	145 sccm

The PTR-ToF-MS subsampled 2 slpm of air from the main inlet whose total flow was 10 slpm. A 3-way solenoid valve was controlled to allow the instrument to sample ambient air for 55 minutes and switch to VOC free air processed by passing ambient air through a heated platinum catalyst for 5 minutes. Concentrations derived from the PTR-ToF-MS are deduced from the difference between the ion intensities measured from the ambient samples and those obtained from VOC free processed air. On one occasion from 20:00 5/17/17 – 14:00 5/19/17 the 3-way solenoid valve program was not activated and no VOC free background measurements were performed. This event coupled with the discovery that there was a small vacuum leak in the PTR-ToF-MS sample inlet interface at the end of project has significant consequences on some of the measured species during aforementioned time period. At night when the Aerodyne Mini-Mobile vehicle doors were closed fumes from the on-board generator fuel tank would build up leading to enhanced signals associated with ions associated with gasoline fuel components. For most of the campaign the instrument zeros adequately account for the contributions from this small leak. Although, there are artifacts evident in the data resulting from the interpolation routine used to define the background signal.

Mass scale calibrations were performed using TOFWERK 2.5 using standard peak fitting procedure. These calibrations used $H_3{}^{18}O^+$ (m/z 21.022086), NO^+ (m/z 29.997440), $H_3{}^{18}O(H_2O)^+$ (m/z 39.032651), $C_3H_7O^+$ (m/z 59.049141), $C_6H_4Cl_3^+$ (m/z 280.937310) and $C_6H_4{}^{37}ClCl_2^+$ (m/z 182.934909). Accuracy of the peak fitting for the calibrated ions was generally better than 40 ppm and resulted in a resolution (m/dm) of 3000 for the high mass ions. Elemental ion compositions were deduced with the aid of the TOFWARE program, which performed the peak extractions for the high resolution time series.

Ion intensities produced for the high resolution peak extraction process were converted to normalized counts per second (ncps) by dividing the measured ion intensity by the sum of the H_3O^+ and $H_3O^+(H_2O)$ reagent ions as follows:

$$I_{MH+}\ (ncps) = 1e6 * I_{MH+}/(I_{H318O+} * 500 + I_{H318O+(H2O)} * 250)$$

The resulting normalized signals have their background signals (zeros) evaluated and averaged for each zero period. These backgrounds are then interpolated back onto the original time base and subtracted from the measured ambient signals. These background corrected signals can be converted to mixing ratios by application of the appropriate

sensitivity factor. By convention PTR-MS sensitivity factors are expressed as ncps/ppb so the conversion to ppb is accomplished by dividing the background corrected signals (ncps) by the sensitivity factor (ncps/ppb). Sensitivity factors can be determined either directly through calibration using standards or estimated from calculated proton transfer reaction constants.

Sensitivity factors for a selected set of compounds were evaluated from calibrations that were performed periodically throughout the campaign by overblowing the inlet with a dynamically diluted calibration standard. All but one of calibrations was performed using a multicomponent calibration standard containing: methanol, acetonitrile, acetaldehyde, acetone, benzene, toluene, p-xylene and 1,2,4-trimethylbenzene. A single calibration was performed using a standard borrowed from the University of Houston that contained: methanol, acetonitrile, acetaldehyde, acetone, dimethylsulfide, isoprene, methacrolein, methylvinylketone, methylethylketone, benzene, toluene and camphene. For all of the components except methanol that were common to both standards the calibration factors differed by less than 10%. For methanol the sensitivity factors differed by 18%. A total of 11 calibrations were performed during the campaign and the results of those measurements are reported in Table S4.

Table S4. PTR-ToF-MS Calibration tank species and sensitivity factors.

Compound	Measured Sensitivity (ncps/ppb)	[a]Measured Sensitivity (std dev)	Branching Fraction	[b]k (10^{-9} cm^3/s)	Calculated Sensitivity (ncps/ppb)
Methanol	6.38	1.23	1	2.14	9.78
Acetonitrile	13.29	1.73	1	3.82	17.46
Acetaldehyde	10.91	1.62	1	3.02	13.81
Acetone	14.35	2.13	1	3.21	14.68
Dimethylsulfide	7.77		1	2.10	9.60
Isoprene	4.08		0.475	1.95	8.92
[c]MACR+MVK	9.76		1	[c]3.15	14.40
Methylethylketone	12.65		1	3.14	14.36
Benzene	9.06	1.39	1	1.93	8.82
Toluene	10.49	1.67	1	2.08	9.51
p-xylene	11.12	1.79	1	2.26	10.33
[d]1,2,4-TMB	11.41	2.10	1	2.40	10.97
Camphene	5.32		0.51	2.46	11.25

a – standard deviations are not reported for compounds with only a single determination
b – values from Table S1 at E/N 140 Td from Cappellin et al. Table S1 [*Cappellin et al.*, 2012]
c – weighted average for methacrolein(MACR) and methylvinylketone(MVK)
d – trimethylbenzene (TMB)

The reported concentrations for named species that have measured sensitivity factors are determined using the measured sensitivity factors reported in Table S4. It is recognized (see discussion below) that several of the measured sensitivity factors appear to be in error, but without additional information to corroborate that there were problems

with the calibration or that the concentrations within the calibration standard are incorrect. All other concentrations reported were computed using 10 ncps/ppb as the sensitivity factor. This value is reasonable approximation for most compounds that do not extensively fragment upon ionization with this instrument. For compounds that do not fragment, one can using the information presented below, to estimate more appropriate sensitivity factors from computed proton reaction rate constants. Fragmentation results in lower apparent sensitivity due to the loss of signal for the protonated molecule. It is difficult to estimate the extent of fragmentation and no attempt has been made to account for this directly.

It is well known that the extent of ionization of molecules within the drift tube reaction chamber is controlled by the amount of reagent ion, the concentration of the substrate, the proton transfer reaction rate constant and drift time of reagent ion. It is straight forward to calculate the sensitivity (ncps/ppb) expected:

$$\text{Sensitivity (ncps/ppb)} = k*t*\text{number\_density}*1e\text{-}9*1e6$$

Where k = reaction rate constant, t = drift time (100 microseconds), the number density (molecules/ml) is computed via the ideal gas law using the drift tube pressure and temperature, 1e-9 converts to ppb and 1e6 represents 1 million H_3O^+ reagent ions. Table S5 lists the calculated sensitivity for each of components that were calibrated. What is unknown and instrument specific is the extraction and detection efficiency of the ions, but is typically dependent on the ions mass. It is typical to plot the ratio of the measured sensitivity to the calculated sensitivity versus ion mass and then fit the data to an expression. Effectively this provides a transmission curve, which accounts for all of mass dependent ion loss processes. Figure S10 shows a plot of this data for all of the calibrated species. It is important to recognize that the calculated sensitivity assumes all of response is detected as a single ion and the measured sensitivity must be corrected for fragmentation and the natural ^{13}C isotopic abundance. The transmission curve allows one to compute a "measured" (instrument specific) sensitivity factor for any compound for which the proton transfer reaction rate constant is known or can be calculated. These instrument specific sensitivity factors then are corrected to account for fragmentation and natural ^{13}C isotopic abundance.

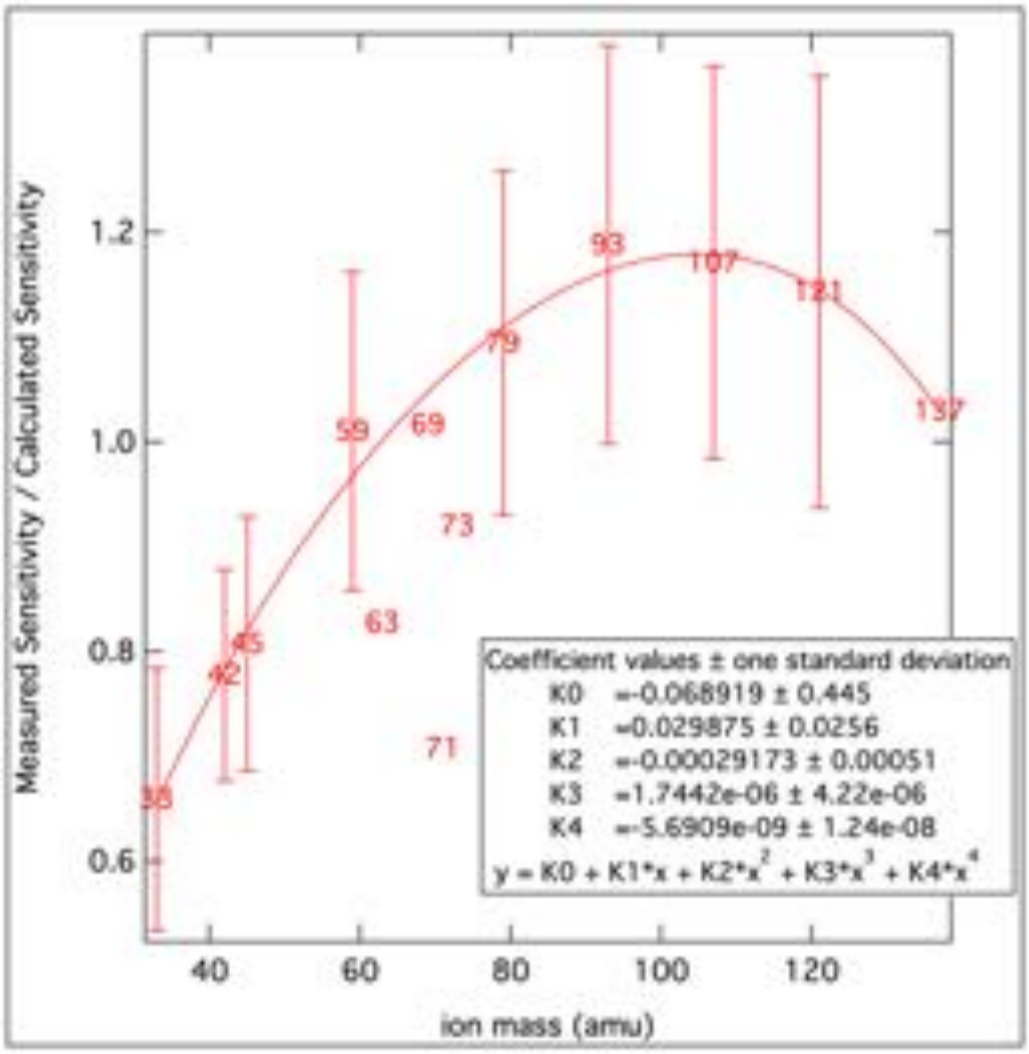

Figure S10. Transmission curve for the PTR-ToF-MS used in SAFS.

PTR-ToF Ion Identifications

For a discussion of the challenges of ion identification and interpretation, particularly as it relates to isoprene, see the Methods section.

The majority of the ion identifications are taken from Yuan et al. [2017], which provides a comprehensive review of both the identification and potential chemical interferences that can impact quantification. All of that detail is appropriate for the current study and is not repeated here. Additional details specific to the SAFS interpretation are included and referenced.

- m/z 33 – CH_5O^+ – methanol
- m/z 41 – $C_3H_5^+$ There are multiple sources contributing to this ion. A significant fraction of the isoprene signal is detected at this mass. During high isoprene loadings the majority of the $C_3H_5^+$ signal is attributable to isoprene and is accounted for in the isoprene quantification. The residual $C_3H_5^+$ most likely originates from the fragmentation of larger unidentified hydrocarbons.

- m/z 42 – $C_2H_4N^+$ – acetonitrile
- m/z 43 – $C_2H_3O^+$ This ion appears to be almost exclusively derived from the fragmentation of acetic acid. A correlation plot of $C_2H_3O^+$ versus $C_2H_5O_2^+$ yields a slope of 0.79 and correlation coefficient of 0.9.
- m/z 43 – $C_3H_7^+$ This ion represents the sum of propene and fragmentation products from larger hydrocarbons. While it most likely that bulk of the signal originates from fragmentation of larger hydrocarbons, one can use this signal to place an upper limit on propene.
- m/z 45 – $C_2H_5O^+$ – acetaldehyde
- m/z 46 – NO_2^+ This signal is attributed to a collection of organic nitrates. It is important to recognize that the NO_2^+ signal reported is from an atmospherically generated product and not from residual background signal generated from the hollow cathode ion source. Organic nitrates are known to produce a NO_2^+ fragment in the PTR-MS (Duncianu et al. 2017, Hansel et al. 2000).
- m/z 47 – $CH_3O_2^+$ – formic Acid
- m/z 47 – $C_2H_7O^+$ This ion is assumed to be ethanol. The reported concentration should viewed as a lower limit as ethanol undergoes a significant amount of fragmentation within the PTR-MS.
- m/z 54 – $C_3H_4N^+$ – acrylonitrile
- m/z 57 – $C_3H_5O^+$ – acrolein
- m/z 57 – $C_4H_9^+$ This ion represents the sum of butenes and fragmentation products from larger hydrocarbons. While it most likely that bulk of the signal originates from fragmentation of larger hydrocarbons, one can use this signal to place an upper limit on butenes.
- m/z 59 – $C_3H_7O^+$ – acetone
- m/z 61 – $C_2H_5O_2^+$ – This ion represents the sum of acetic acid and glycolaldehyde.
- m/z 63 – $C_2H_7S^+$ – dimethylsulfide
- m/z 67 – $C_5H_7^+$ – This ion signal is predominately from the fragmentation of isoprene.
- m/z 69 – $C_4H_5O^+$ – furan

  m/z 69 – $C_5H_9^+$ – Isoprene is the dominate contributor during the daytime, but other species such as methylbutenol (MBO) can contribute. Comparison with the GC-ToF-MS indicated reasonable agreement with the PTR-ToF-MS measurements being 13% higher.
- m/z 71 – $C_4H_7O^+$ – Sum of methylacrolein and methylvinylketone produced from the photooxidation of isoprene.
- m/z 71 – $C_5H_{11}^+$ This ion represents the sum of pentenes and fragmentation products from larger hydrocarbons. While it most likely that bulk of the signal originates from

fragmentation of larger hydrocarbons, one can use this signal to place an upper limit on pentenes.

- m/z 73 – $C_3H_5O_2^+$ – Methylgyloxal is considered to be the most logical assignment for this ion. Bregonzio-Rozier et al. 2016 identified methylglyoxal as significant component of isoprene oxidation in a chamber experiment.
- m/z 73 – $C_4H_9O^+$ – methylethylketone
- m/z 75 – $C_3H_5O_2^+$ – hydroxyacetone
- m/z 77 – $C_2H_5O_3^+$ – fragment ion of peroxyacetyl nitrate (PAN).
- m/z 79 – $C_6H_7^+$ – benzene
- m/z 81 – $C_6H_9^+$ – This ion is primarily a fragment ion of the monoterpenes. In most cases the carbon presence in this ion signal is accounted for in the quantification of the monoterpenes. There are times when this ion is significantly enhanced in the absence of any monoterpene signal. This source has not been identified.
- m/z 83 $C_5H_7O^+$ This ion is thought to be methylfuran, which is a known oxidation product of isoprene.
- m/z 83 $C_6H_{11}^+$ Most PTR-MS literature identifies this ion as methylcyclopentane. Corresponding measurements of methylcyclopentane made by a GC-MS indicate no correlation between the C6H11+ ion signal and cyclopentane while the mobile lab was at Floresville. Based on this result we conclude that C6H11+ is not methylcyclopentane.
- m/z 85 $C_5H_9O^+$ No identification has been made.
- m/z 85 $C_6H_{13}^+$ This ion represents the parent ion of the C6-alkenes and is also a fragment ion of larger hydrocarbons.
- m/z 87 $C_4H_7O_2^+$ Bregonzio-Rozier et al. 2016 have identified this ion as a component of isoprene oxidation in a chamber experiment.
- m/z 87 $C_5H_{11}O^+$ This ion is tentatively identified as methylbutenol which is emitted from conifers in an analogous pathway to that of isoprene. It is only emitted during the daytime.
- m/z 89 $C_4H_9O_2^+$ No identification has been made.
- m/z 93 $C_7H_9^+$ – Toluene
- m/z 95 $C_7H_{11}^+$ This ion appears to be predominately as a fragment from the monoterpenes, but there are other sources that contribute.
- m/z 97 $C_5H_5O_2^+$ No identification has been made.
- m/z 97 $C_6H_9O^+$ No identification has been made.
- m/z 97 $C_7H_{13}^+$ No identification has been made.
- m/z 99 $C_4H_3O_3^+$ No identification has been made.

- m/z 99 $C_5H_7O_2^+$ Bregonzio-Rozier et al. 2016 identified this ion as a component of isoprene oxidation in a chamber experiment.
- m/z 99 $C_6H_{11}O^+$ No identification has been made.
- m/z 101 $C_5H_9O_2^+$ No identification has been made.
- m/z 101 $C_6H_{13}O^+$ No identification has been made.
- m/z 103 $C_5H_{11}O_2^+$ No identification has been made.
- m/z 104 $C_8H_9^+$ – Styrene
- m/z 107 $C_7H_7O^+$ – Benzaldehyde
- m/z 107 $C_8H_{11}^+$ – Sum of the C8Benzenes which includes ethylbenzene plus the three xylene isomers.
- m/z 113 $C_5H_5O_3^+$ Bregonzio-Rozier et al. 2016 identified this ion as a component of isoprene oxidation in a chamber experiment.
- m/z 115 $C_6H_{11}O_2^+$ No identification has been made.
- m/z 117 CCl_3^+ This is the fragment ion of carbon tetrachloride and most likely produced via O_2^+ as the proton affinity of CCl_4 is less than that of water.
- m/z 119 $C_9H_{11}^+$ No identification has been made.
- m/z 121 $C_8H_9O^+$ This could be a tolualdehyde but the overall signal is very weak and no identification has been made.
- m/z 121 $C_9H_{13}^+$ Sum of the C9Benzenes which includes a total of 8 isomers representing 3-trimethylbenzenes, 3-ethyltoluenes and 2-propylbenzenes.
- m/z 123 $C_9H_{15}^+$ No identification has been made.
- m/z 127 $C_8H_{15}O^+$ No identification has been made.
- m/z 129 $C_{10}H_9^+$ – naphthalene
- m/z 129 $C_7H_{13}O_2^+$ No identification has been made.
- m/z 131 $C_{10}H_{11}^+$ No identification has been made.
- m/z 135 $C_{10}H_{15}^+$ Sum of the C10Benzenes.
- m/z 137 $C_{10}H_{17}^+$ Sum of the monoterpenes.
- m/z 139 $C_9H_{15}O^+$ This ion has been observed as an emission product from Ponderosa pines. It temporal pattern is consistent with it being a biogenic emission from conifers.
- m/z 149 $C_{11}H_{17}^+$ Sum of the C11Benzenes.
- m/z 153 $C_{10}H_{17}O^+$ This is most likely a terpenoid compound coming from conifers.
- m/z 205 $C_{15}H_{25}^+$ This is parent ion for the sesquiterpenes. These compounds undergo a significant amount of fragmentation so the concentration represents a lower limit.

References:

Bregonzio-Rozier et al., Secondary organic aerosol formation from isoprene photooxidation during cloud condensation-evaporation cycles. Atmos. Chem. Phys., 16, 1747-1760, doi:10.5194/acp-16-1747-2016, 2016.

Duncianu et al., Measurement of alkyl and multifunctional organic nitrates by proton-transfer-reaction mass spectrometry. Atmos. Meas. Tech. 10, 1445-1463, doi:10.5194/amt-10-1445-2017, 2017.

Gueneron, M.; Erickson, M.H.; VanderSchelden, G.S.; Jobson, B.T. PTR-MS fragmentation patterns of gasoline hydrocarbons, Int. J. Mass. Spectrom. 379, 97-109, 2015.

Hansel, A., Wisthaler, A., A Method for Real-Time Detection of PAN, PPN and MPAN in Ambient Air. Geophys. Res. Lett., 27(6) 895-898, 2000.

Kari, E.; Miettinen, P.; Yli-Pirila, P.; Virtanen, A.; Faiola, C.L. PTR-ToF-MS product ion distributions and humidity-dependence of biogenic volatile organic compounds. Int. J. Mass Spectrom. 430, 87-97, 2018.

Yuan et al., Proton-Transfer-Reaction Mass Spectrometry: Application in Atmospheric Sciences, Chem. Rev. 117, 13187-13229, doi: 10.1021/acs.chemrev.7b00325, 2017.

I-CIMS Operation and Data Description

The iodide Chemical Ionization Mass Spectrometer (I-CIMS) relies on VOC ionization by either creating an I^-VOC adduct, or by charge exchange and/or fragmentation in reaction with an I^- or $I^-(H_2O)_n$ cluster. Once ionized the ions are injected into a high-resolution time-of-flight (TOF) mass spectrometer, which provides mass-to-charge resolution in excess of 4000:1. The mass spectrometer during the SAFS field campaign recorded one-second mass spectra over the course of the field campaign. To speed up analysis, the fast data was averaged to 20-seconds before high-resolution fitting of the mass spectra. Mass spectral fitting consists of calibrating the mass spectral m/z axis based upon a set of common, intense species easily identified in the mass spectrum over the course of the campaign. The stability of that mass calibration is a reflection of instrument stability and drift. The resulting uncertainty in mass calibrated peak positions from the SAFS campaign was <5 parts per million (i.e. the peak position is established to better than 5 parts in 1 million).

After mass calibration, peaks can be identified and fit, allowing for conversion to mass signal intensity. As shown below, there can be many overlapping peaks, requiring fitting of the peak shapes to multiple contributions. In this case, we identified 2446 potential mass peaks in the range of 30-700 amu. We refined this peak list by only focusing on peaks that have responded to the application of zero air to the instrument inlet. The resulting list is used in the PMF analysis of the data, and chemical

interpretation. The peak intensities are corrected for zero offset, and calibrated to known concentrations where appropriate.

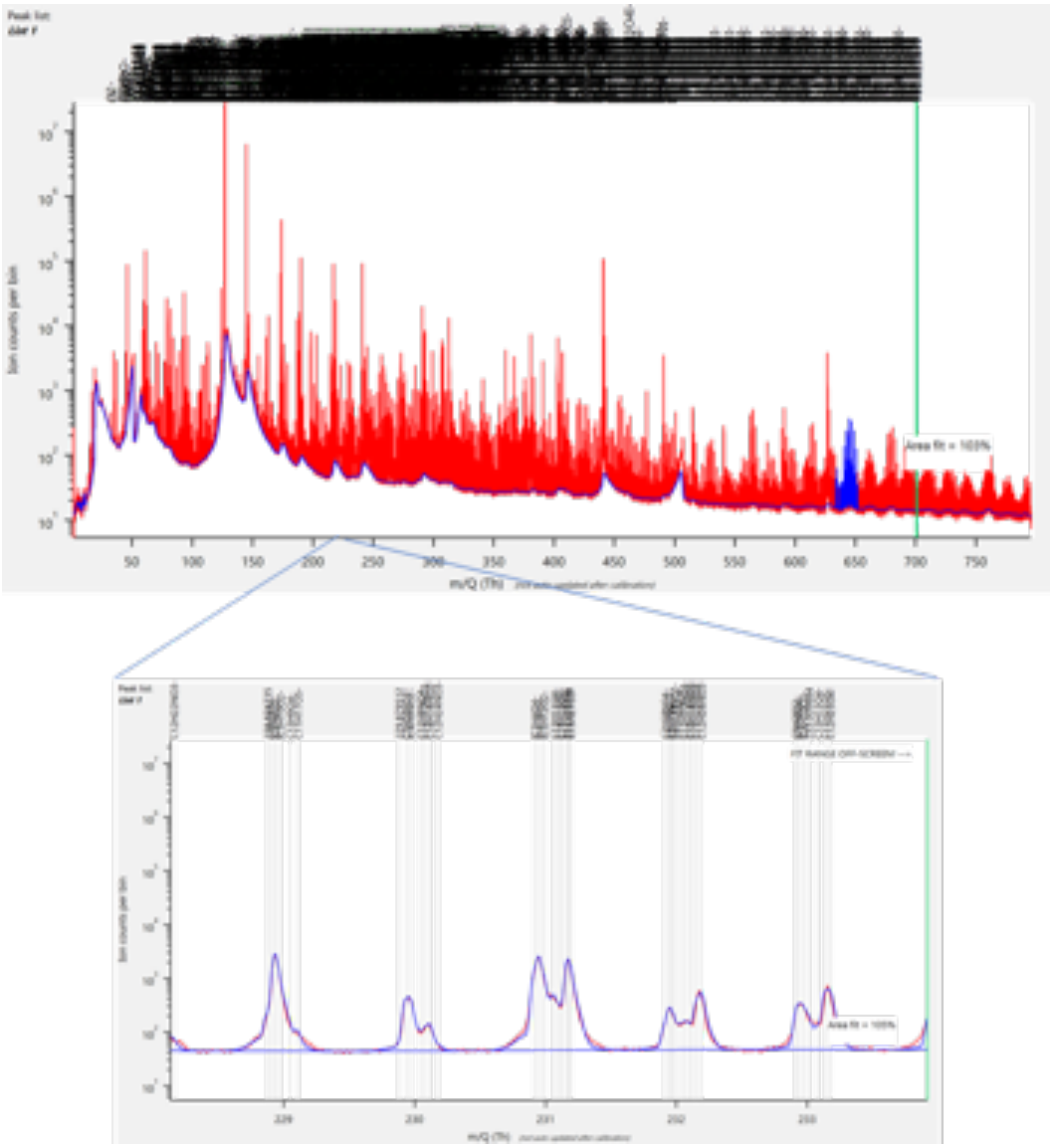

Figure S11. I-CIMS massspectrum and high-resolution fit

Hysplit and Land Use Overlap Integrals

Overlap integrals of a Hysplit back-trajectory with Texans land cover types, in arbitrary units.

An overlap analysis was done using land cover data from the Texas Parks & Wildlife Department [*Elliott et al.*, 2014] and results from Hysplit footprint simulations performed in this study. A map of the spatial domain of these overlaps including 5 different

categories is shown below. The oil and gas play type was allowed to overlap with the other 4 land cover types. A detailed explanation of this procedure is produced in the Methods section of this final report.

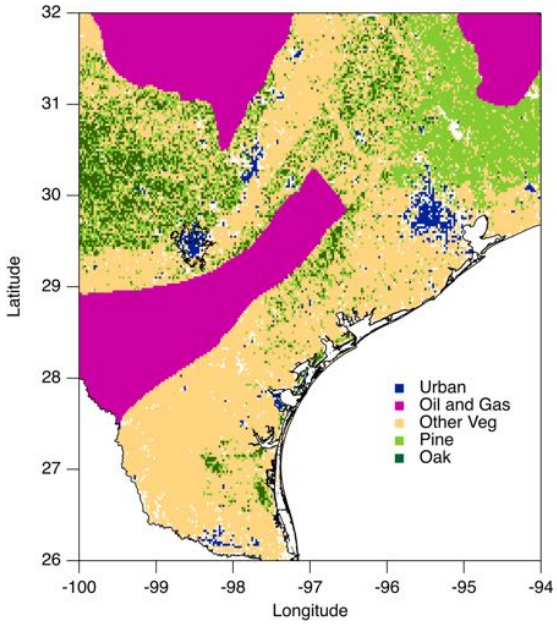

Figure S12. Oil and gas mapping information (purple) is shown gridded onto a 0.025 x 0.025 decimal degree grid overlaying the ecological and land use data. The Eagle Ford Shale Play (center left) is shown along with some of the Barnett (top left) and Haynesville (top right) plays. The city bounds of San Antonio are outlined in black.

Appendix B: Audits of Data Quality – 10% Required

A data quality audit on 10 % of the generated high-resolution time-series. The chosen data will include traces from each of the three instruments undergoing high-resolution analysis, and will include the entire time period of interest. This choice (complete time coverage, 10% data series coverage) is preferred over the alternate (10% time coverage, all data series) because any problems found in the data at a given time are likely to be present at the same time in other data from that instrument. This choice results in a better chance of finding errors.

The data audit will be done by plotting the selected time series along with pre-existing data for diagnostic tracers like carbon monoxide, and, if available, TCEQ data for the species of interest. The following problems will be identified, flagged and communicated to the scientist responsible for the data for correction: zeroes or calibration periods present; glitches in the data showing unphysical mixing ratios; bias in the data versus TCEQ reference (if available); and others.

PTR-ToF

A total of 68 PTR-ToF data series were produced. They are listed in Appendix A. They include both chemical species which were originally produced in low-resolution format for the 2017 dataset, and additional species identified during this 2019 project. An overview of these traces is shown in the figure below. This figure is used to qualitatively assess data coverage and the presence of glitches or other major issues. Coverage is excellent, with only a few gaps.

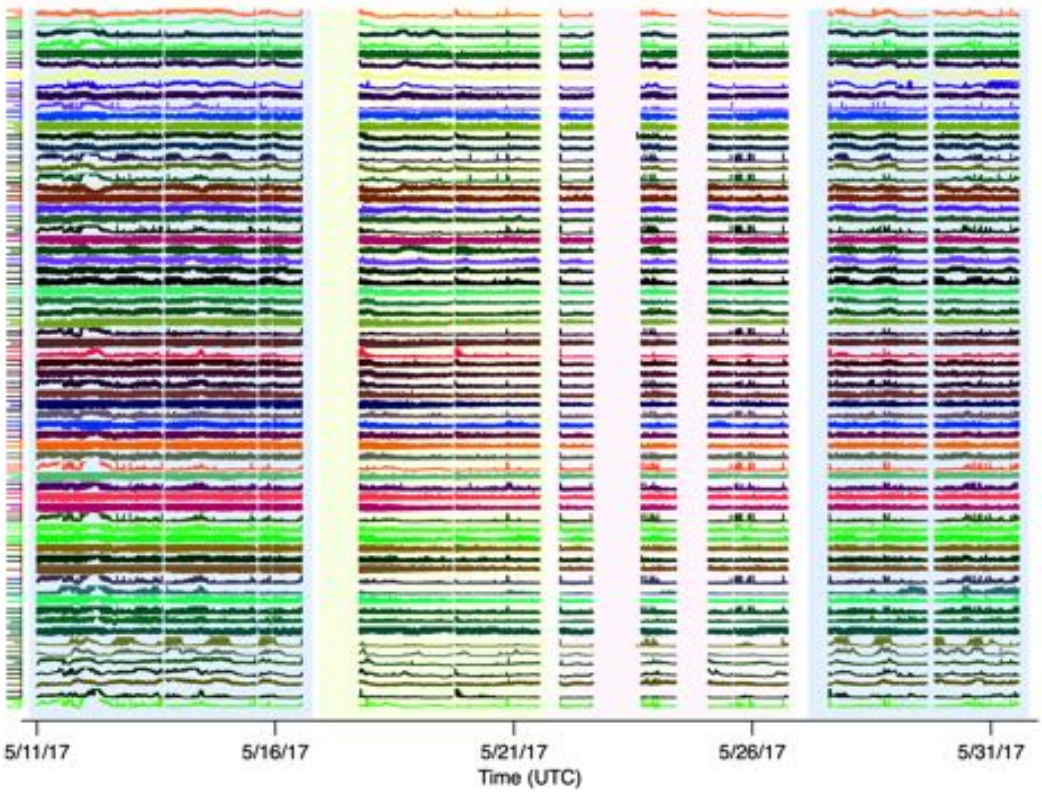

Figure S13. Overview of PTR-ToF dataset showing data coverage. The vertical axes are scaled to show baseline variations. The graph is shaded according to the three locations visited during SAFS.

The following data quality indicators were assessed

- ***Zeroes or calibration periods present:*** ***None present***
- ***Glitches in the data showing unphysical mixing ratios:*** ***None present***
- ***Biases in data versus reference (if available):*** ***See below***
- ***Other quality notes:***
    - ***Baseline wander for small periods of time for some tracers. See below.***
    - ***Ensured large data gaps in fast data were set to NaN in final 1-minute average data.***

A selection of the produced PTR-ToF dataset was compared with low-resolution results produced from the original 2017 project. Only those species which were resolved in the 2017 project were available for comparison; this cannot be done with the newly identified tracers. Differences in calibration factors were noted, with toluene as an example below. These differences were flagged and set off a review of the full PTR-ToF dataset. We identified major changes in the calibration factors after high-resolution fitting, which had not been incorporated into the final dataset. Updated data using calibration factors appropriate to the high-resolution fits was produced as a result of this QA procedure.

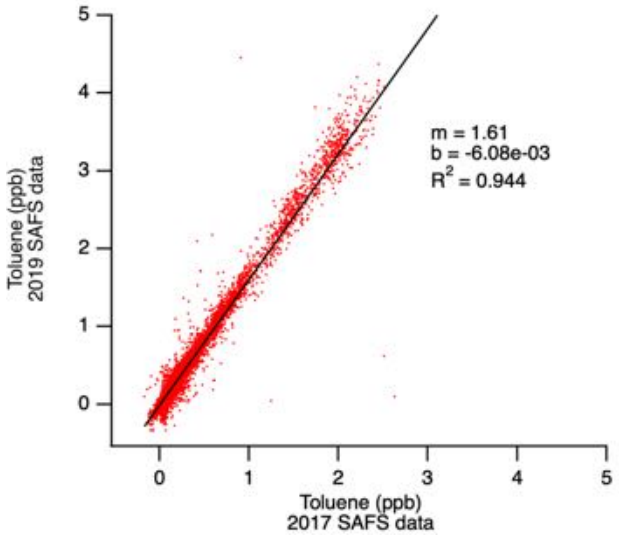

Figure S14. Difference in calibration of toluene data in 2019 versus 2017 dataset.

Occasionally, the PTR-ToF high-resolution data shows baseline deviations that are unphysical (i.e. fall below zero). An example is shown below.

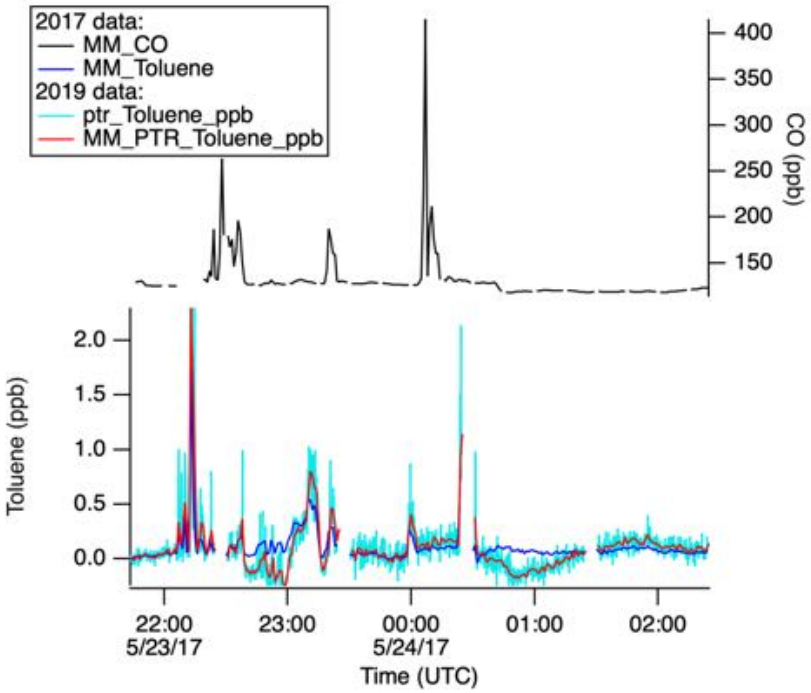

Figure S15. Example of baseline wander in high-resolution data (red) versus 2017 low-resolution dataset.

I-CIMS

The I-CIMS data produced are listed in Appendix A. They include both chemical species which were produced in low-resolution format for the 2017 dataset, and additional species identified during this 2019 project.

The ***following data quality indicators were assessed***

- ***Zeroes or calibration periods present:*** ***None present***
- ***Glitches in the data showing unphysical mixing ratios:*** ***None present***
- ***Biases in data versus reference (if available):*** ***See below***
- ***Other quality notes:***
  - ***Updated results fill in missing periods compared to 2017 results***
  - ***Integration problem initially yielded zero mixing ratios in Floresville (see below).***
  - ***Ensured large data gaps in fast data were set to NaN in final 1-minute average data.***

In general the high-resolution I-CIMS results show greater coverage compared to 2017 versions of the data. An example is shown below for $IClNO_2$, where the original data ended on 5/29, but updated data covered up until the end of the campaign on 5/31.

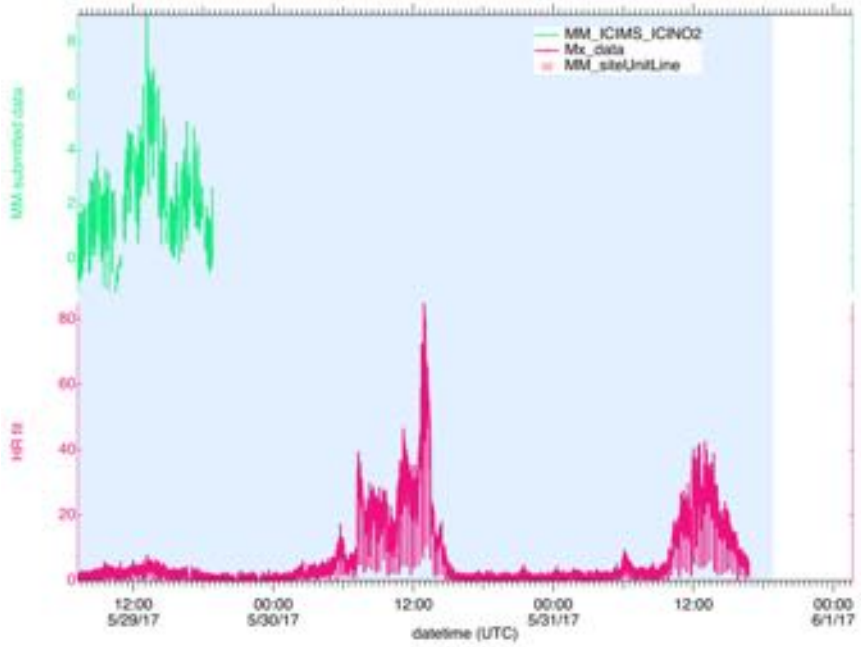

Figure S16. Previous data (green) versus high-resolution data (red) produced as part of this campaign. The $IClNO_2$ data trace is shown here.

Furthermore, the high-resolution fit data is better able to capture certain species during periods which yielded unphysical in the previous data.

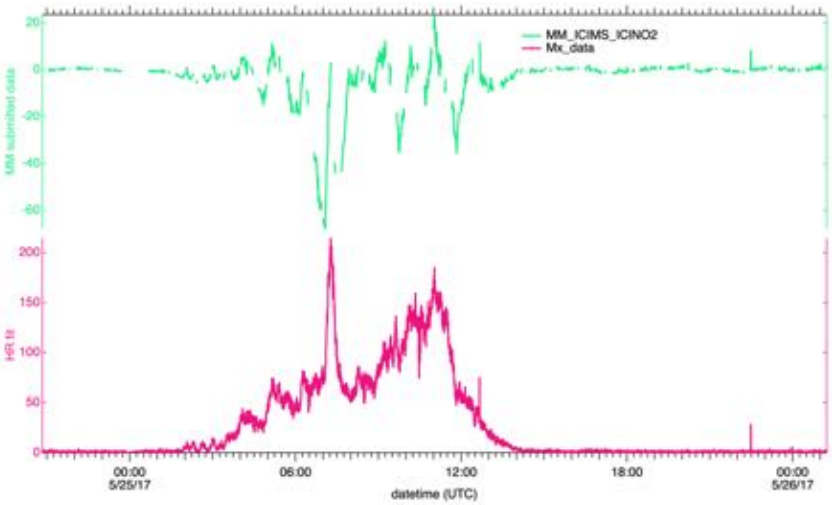

Figure S17. Previous data (green) versus high-resolution data (red) produced as part of this campaign. The $ICINO_2$ data trace is shown here. Previous data showed unphysical results at some times, which are not present in the current dataset.

However, the high-resolution I-CIMS dataset showed a major problem which was identified as part of this data quality audit: There is a period in time in Floresville, where the fit is yielding exactly zero. This is shown in Figure S18 for $ICINO_2$ but is true for all I-CIMS species as shown in Figure S19. This issue has been corrected in the final data.

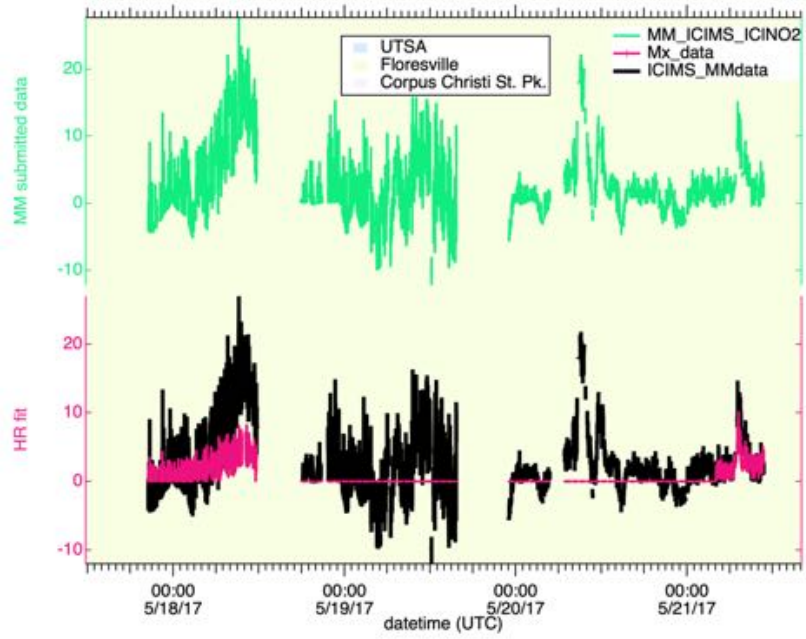

Figure S18. Highlight showing high-resolution fit problem between 5/18 and 5/21. The black and green traces correspond to $ICINO_2$ data produced as part of the 2017 project. The pink trace shows the initial attempt at a high-resolution fit.

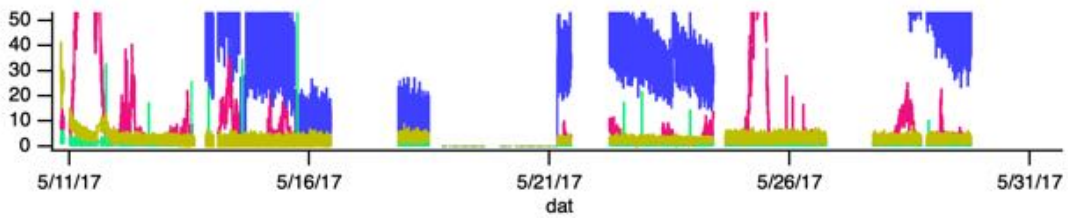

Figure S19. The high-resolution fit problem is apparent for all I-CIMS species in the original high-resolution fit.

GC-ToF

A total of 18 GC-ToF data series were produced. They are listed in Appendix A. They include chemical species which were originally produced for the 2017 dataset, but include additional time periods analyzed as part of this 2019 project. An overview of these traces is shown in the figure below. This figure is used to qualitatively assess data coverage and the presence of glitches or other major issues. Coverage for this data series has some gaps; the instrument in question was a first-of-its-kind prototype instrument at the time of deployment. Data reported in the latter half of the campaign (dotted lines) are in estimated response only.

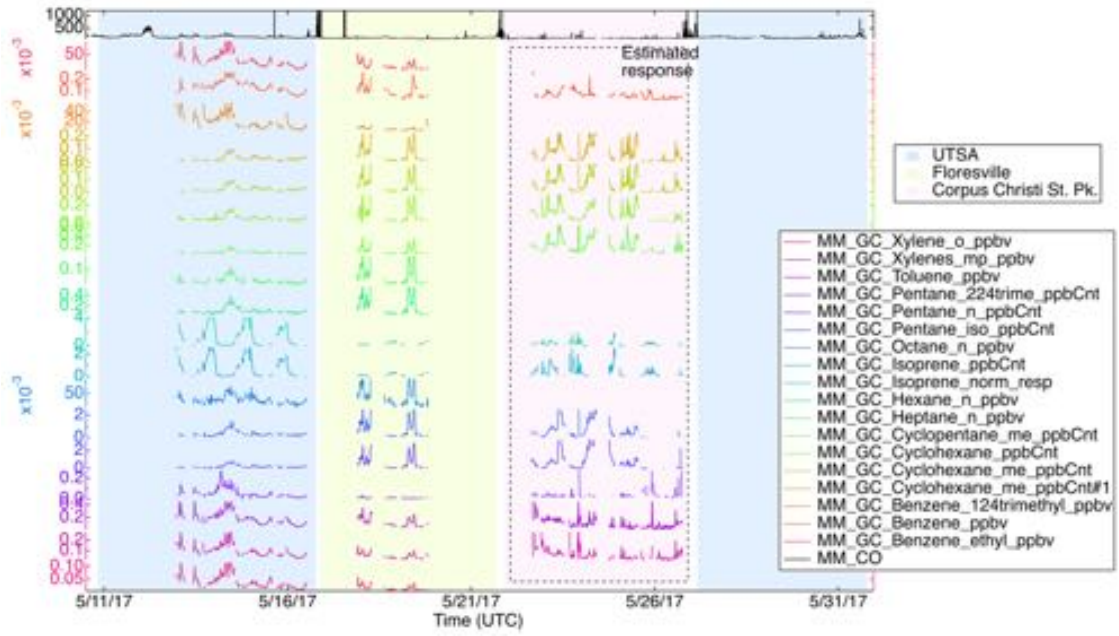

Figure S20. Overview of GC-ToF dataset showing data coverage. The graph is shaded according to the three locations visited during SAFS. GC-ToF data reported in Corpus Christi State Park have been reported in counts; the rest of the time series is in calibrated parts-per-billion.

The ***following data quality indicators were assessed***

- ***Zeroes or calibration periods present:*** ***None present***
- ***Glitches in the data showing unphysical mixing ratios:*** ***None present***
- ***Biases in data versus reference (if available):*** ***See below***
- ***Other quality notes:***

- ***Updated results fill in missing periods compared to 2017 results***
- ***Ensured large data gaps in fast data were set to NaN in final 1-minute average data.***

Appendix C: Ion Lists

PTR-ToF-MS Ion Identification

The table below provides a listing of the ions that were identified and quantified using the PTR-ToF-MS instrument during the 2017 SAFS. Concentrations were derived for each species (ion) using the sensitivity factors listed. Sensitivity factors that were derived from calibration experiments (values not equal to 10) and account for both the inherent sensitivity (proton transfer reaction rate) and any loss in the MH^+ signal intensity due to fragmentation. Mathematically $S_{measured} = S_{theory}$*Branching Fraction. Uncalibrated species have been assigned a $S_{measured}$ = 10 ncps/ppb. Concentrations for uncalibrated species can be improved by applying the appropriate sensitivity factor by multiplying the reported concentrations by the ratio of $(10/S_{calculated})$. Mixing ratio averages reflect the average of over the entire campaign. The note section provides tentative identifications, which are discussed in greater detail in the following section.

Table S5. PTR-ToF-MS HR Ion Identification List

Nominal Mass	Carbon Number	ion	measured m/z	exact m/z	Sensitivity (ncps/ppb)	Branching fraction	mixing ratio average (ppb)	mixing ratio std dev (ppb)	mixing ratio average (ppbC)	mixing ratio std dev (ppbC)	note
21		H318O+	21.021872	21.0181							reagent ion
33	1	CH5O+	33.033491	33.03404	6.38	1	5	2.5	5	2.5	methanol
39		H318O+ (H2O)	39.032622	39.0262							reagent ion
41	3	C3H5+	41.038577	41.039125	10	1	0.46	0.41	1.38	1.23	isoprene fragment + hydrocarbon fragment
42	2	C2H4N+	42.033826	42.034374	13.29	1	0.21	0.09	0.42	0.18	acetonitrile
43	2	C2H3O+	43.017841	43.0184	10	1	0.86	0.56	1.72	1.12	acetic acid fragment
43	3	C3H7+	43.054227	43.0548	10	1	0.21	0.26	0.63	0.78	C3-alkene + hydrocarbon fragment
45	2	C2H5O+	45.033491	45.034	10.91	1	0.86	0.51	1.72	1.02	acetaldehyde
46		NO2+	45.992355	45.9929	10	1	0.3	0.27			organonitrate fragment
47	1	CH3O2+	47.012756	47.01433	10	1	0.45	0.36	0.45	0.36	formic acid
47	2	C2H7O+	47.049141	47.0497	10	1	0.13	0.3	0.26	0.6	ethanol
54	3	C3H4N+	54.033826	54.0344	10	1	0.006	0.12	0.018	0.36	acrylonitrile
57	3	C3H5O+	57.033491	57.034	10	1	0.086	0.087	0.258	0.261	acrolein
57	4	C4H9+	57.069877	57.0704	10	1	0.24	0.34	0.96	1.36	C4-alkene + hydrocarbon fragment
59	3	C3H7O+	59.049141	59.0497	14.35	1	2.67	1.18	8.01	3.54	acetone
61	2	C2H5O2+	61.028406	61.0189	10	1	0.87	0.65	1.74	1.3	acetic acid + glycoaldehyde
63	2	C2H7S+	63.026298	63.0164	7.77	1	0.026	0.042	0.052	0.084	dimethylsulfide

Nominal Mass	Carbon Number	ion	measured m/z	exact m/z	Sensitivity (ncps/ppb)	Branching fraction	mixing ratio average (ppb)	mixing ratio std dev (ppb)	mixing ratio average (ppbC)	mixing ratio std dev (ppbC)	note
67	5	C5H7+	67.054227	67.0437	10	1	0.031	0.04	0.155	0.2	isoprene fragment +
69	4	C4H5O+	69.033491	69.0227	10	1	0.015	0.02	0.06	0.08	furan
69	5	C5H9+	69.069877	69.0704	4.08	0.475	1.1	1.4	5.5	7	isoprene + MBO
71	4	C4H7O+	71.049141	71.0497	9.76	1	0.47	0.5	1.88	2	methacrolein + methylvinylketone
71	5	C5H11+	71.085527	71.0861	10	1	0.053	0.11	0.265	0.55	C5-alkene + hydrocarbon fragment
73	3	C3H5O2+	73.028406	73.02896	10	1	0.081	0.076	0.243	0.228	isoprene oxidation product maybe methyglyoxal
73	4	C4H9O+	73.064791	73.0653	12.65	1	0.41	0.24	1.64	0.96	methylethylketone
75	3	C3H7O2+	75.044056	75.0446	10	1	0.21	0.17	0.63	0.51	isoprene oxidation product maybe hydroxyacetone
77	2	C2H5O3+	77.02332	77.0239	10	1	0.034	0.043	0.068	0.086	PAN fragment
79	6	C6H7+	79.054227	79.0548	9.06	1	0.17	0.25	1.02	1.5	benzene
81	6	C6H9+	81.069877	81.0704	10	1	0.1	0.2	0.6	1.2	terpene fragment +
83	5	C5H7O+	83.049141	83.0497	10	1	0.038	0.043	0.19	0.215	isoprene oxidation product maybe methylfuran
83	6	C6H11+	83.085527	83.0861	10	1	0.075	0.081	0.45	0.486	
85	5	C5H9O+	85.064791	85.0653	10	1	0.032	0.035	0.16	0.175	

Nominal Mass	Carbon Number	ion	measured m/z	exact m/z	Sensitivity (ncps/ppb)	Branching fraction	mixing ratio average (ppb)	mixing ratio std dev (ppb)	mixing ratio average (ppbC)	mixing ratio std dev (ppbC)	note
85	6	C6H13+	85.101177	85.1017	10	1	0.02	0.036	0.12	0.216	C6-alkene + hydrocarbon fragment
87	4	C4H7O2+	87.044056	87.0446	10	1	0.07	0.072	0.28	0.288	isoprene oxidation product
87	5	C5H11O+	87.080441	87.081	10	1	0.083	0.059	0.415	0.295	maybe methylbutenol
89	4	C4H9O2+	89.059706	89.0603	10	1	0.036	0.039	0.144	0.156	
93	7	C7H9+	93.069877	93.0704	10.49	1	0.34	0.62	2.38	4.34	toluene
95	7	C7H11+	95.085527	95.0861	10	1	0.02	0.027	0.14	0.189	terpene fragment +
97	5	C5H5O2+	97.028406	97.02896	10	1	0.011	0.034	0.055	0.17	
97	6	C6H9O+	97.064791	97.0653	10	1	0.006	0.022	0.036	0.132	
97	7	C7H13+	97.101177	97.1017	10	1	0.031	0.036	0.217	0.252	
99	4	C4H3O3+	99.00767	99.0082	10	1	0	0.045	0		
99	5	C5H7O2+	99.044056	99.0446	10	1	0.044	0.053	0.22	0.265	isoprene oxidation product
99	6	C6H11O+	99.080441	99.081	10	1	0.032	0.037	0.192	0.222	
101	5	C5H9O2+	101.059706	101.0603	10	1	0.049	0.075	0.245	0.375	
101	6	C6H13O+	101.096091	101.0966	10	1	0.03	0.03	0.18	0.18	
103	5	C5H11O2+	103.075356	103.0759	10	1	0.0004	0.025	0.002	0.125	
105	8	C8H9+	105.069877	105.0704	10	1	0.019	0.036	0.152	0.288	styrene
107	7	C7H7O+	107.049141	107.0497	10	1	0.024	0.033	0.168	0.231	benzaldehyde

Nominal Mass	Carbon Number	ion	measured m/z	exact m/z	Sensitivity (ncps/ppb)	Branching fraction	mixing ratio average (ppb)	mixing ratio std dev (ppb)	mixing ratio average (ppbC)	mixing ratio std dev (ppbC)	note
107	8	C8H11+	107.085527	107.0861	11.12	1	0.26	0.5	2.08	4	C8benzenes
113	5	C5H5O3+	113.02332	113.0239	10	1	0.049	0.089	0.245	0.445	isoprene oxidation product
115	6	C6H11O2+	115.075356	115.0759	10	1	0.026	0.039	0.156	0.234	
117	1	CCl3+	116.906009	116.9066	10	1	0.07	0.03	0.07	0.03	carbon tetrachloride
119	9	C9H11+	119.085527	119.0861	10	1	0.01	0.023	0.09	0.207	
121	8	C8H9O+	121.064791	121.0653	10	1	0.002	0.023	0.016	0.184	
121	9	C9H13+	121.101177	121.1017	11.41	1	0.14	0.28	1.26	2.52	C9-benzenes
123	9	C9H15+	123.1174	123.1174	10	1	0.012	0.17	0.108	1.53	
127	8	C8H15O+	127.111742	127.1123	10	1	0.008	0.016	0.064	0.128	
129	10	C10H9+	129.069877	129.0704	10	1	0.022	0.039	0.22	0.39	naphthalene
129	7	C7H13O2+	129.091006	129.0916	10	1	0.014	0.027	0.098	0.189	
131	10	C10H11+	131.101	131.0861	10	1	0.002	0.013	0.02	0.13	
135	10	C10H15+	135.116827	135.1174	10	1	0.061	0.107	0.61	1.07	C10-benzenes
137	10	C10H17+	137.132477	137.133	5.32	0.51	0.157	0.2	1.57	2	terpenes
139	9	C9H15O+	139.111742	139.1123	10	1	0.019	0.024	0.171	0.216	conifer biogenic
149	11	C11H17+	149.132477	149.133	10	1	0.014	0.026	0.154	0.286	C11-benzenes
153	10	C10H17O+	153.127392	153.1279	10	1	0.038	0.083	0.38	0.83	terpenoid
205	15	C15H25+	205.195077	205.1956	10	1	0.009	0.022	0.135	0.33	sesquiterpene

Factor-Specific Signature Ions from I- CIMS

SAFS_ionName	SAFS_mz
Isoprene	
C5H7NO4I-	271.9425354
C4H7NO5I-	275.937439
C5H7NO5I-	287.937439
C4H7NO6I-	291.932373
C5H11NO5I-	291.96875
C5H7NO6I-	303.932373
C5H9NO6I-	305.947998
C5H7NO7I-	319.9272766
C5H9NO7I-	321.9429321
C5H11NO7I-	323.9585876
C5H7NO8I-	335.9221802
C5H9NO8I-	337.9378357
C6H11NO8I-	351.9534912
C5H10N2O8I-	352.9487305
C5H9NO9I-	353.9327393

SAFS_ionName	SAFS_mz
Alpha Pinene	
C10H15NO5I-	356
C8H11NO7I-	359.9586
C8H13NO7I-	361.9742
C10H15NO6I-	371.995
C9H13NO7I-	373.9742
C10H17NO6I-	374.0106
C8H11NO8I-	375.9535
C9H15NO7I-	375.9899
C10H19NO6I-	376.0262
C10H15NO7I-	387.9899
C9H13NO8I-	389.9691
C10H17NO7I-	390.0055
C8H11NO9I-	391.9484
C10H15NO8I-	403.9848
C10H15NO9I-	419.9797

SAFS_ionName	SAFS_mz
Biomass Burning	
IC6H6O2-	236.9418
IC6H5NO3-	265.932
IC9H10O2-	276.9731
IC8H8O3-	278.9524
IC9H12O2-	278.9887
IC7H7NO3-	279.9476
IC6H5NO4-	281.9269
IC6H10O5-	288.9579
IC10H12O2-	290.9887
IC9H10O3-	292.968
IC10H10O3-	304.968
IC10H12O3-	306.9837
IC10H14O3-	308.9993
IC10H14O3-	308.9993
IC6H4N2O5-	310.9171
IC17H20O4-	415.0412
IC18H22O4-	429.0568
IC20H24O5-	471.0674
IC20H22O6-	485.0467

SAFS_ionName	SAFS_mz
1,3,5 Trimethyl Benzene	
IC2H2O3-	200.9054
C2H4O3I-	202.9211
203.914337	203.9143
203.973007	203.973
C9H18NO4-	204.1241
C10H22NO3-	204.1605
C3H7INO2-	215.9527
C2H2O4I-	216.9003
IC2H4O4-	218.916
C3H2IO4-	228.9003
IC4H6O3-	228.9367
IC3H4O4-	230.916
C2H3INO4-	231.9112
IC3H6O4-	232.9316
IC4H10O3-	232.968
IC6H8O2-	238.9575
IC5H6O3-	240.9367
241.899384	241.8994
IC4H4O4-	242.916
C5H8O3I-	242.9524
C6H12O2I-	242.9888
C4H6O4I-	244.9316
IC4H8O4-	246.9473
C5H2IO4-	252.9003
IC6H6O3-	252.9367
C6H9INO2-	253.9684
IC5H4O4-	254.916
C16H34NO-	256.2646
C4H2IO5-	256.8952
C5H8O4I-	258.9473
C14H14NO4-	260.0928
IC5H10O4-	260.9629
IC6H2O4-	264.9003
IC6H6O4-	268.9316
IC6H8O4-	270.9473
C4H2IO6-	272.8902
IC5H6O5-	272.9265
C4H4IO6-	274.9058
C5H8IO5-	274.9422

C5H8IO5-	274.9422
IC9H10O2-	276.9731
IC8H8O3-	278.9524
IC9H12O2-	278.9887
C13H13O7-	281.0667
IC6H5NO4-	281.9269
IC6H4O5-	282.9109
IC7H8O4-	282.9473
IC8H12O3-	282.9837
IC6H6O5-	284.9265
C7H10O4I-	284.9629
IC8H14O3-	284.9993
C5H4IO6-	286.9058
IC6H8O5-	286.9422
C8H3INO3-	287.9163
C5H6O6I-	288.9214
IC6H10O5-	288.9579
C5H8IO6-	290.9371
IC10H12O2-	290.9887
IC9H11NO2-	291.984
C5H10IO6-	292.9528
IC9H10O3-	292.968
C7H5INO4-	293.9269
C19H20NO2-	294.15
C7H4IO5-	294.9109
IC9H12O3-	294.9837
IC10H16O2-	295.0201
C6H4INO5-	296.914
IC8H10O4-	296.9629
C6H4IO6-	298.9058
C8H12IO4-	298.9786
C9H16IO3-	299.015
C6H6IO6-	300.9214
IC7H10O5-	300.9579
C7H12IO5-	302.9735
C5H6O7I-	304.9164
IC10H10O3-	304.968
C8H5INO4-	305.9269
C9H9INO3-	305.9633
IC9H8O4-	306.9473
C9H11INO3-	307.9789

C13H26NO7-	308.1715
IC6H4N2O5-	310.9171
IC8H8O5-	310.9422
IC9H12O4-	310.9786
C10H16O3I-	311.015
IC7H7NO5-	311.9374
IC9H15NO3-	312.0102
IC8H10O5-	312.9579
IC9H14O4-	312.9942
IC10H18O3-	313.0306
C7H8IO6-	314.9371
C7H8IO6-	314.9371
IC8H12O5-	314.9735
C8H15INO4-	316.0051
C6H6IO7-	316.9164
IC7H10O6-	316.9528
C9H4IO5-	318.9109
320.88736	320.8874
C8H18IO5-	321.0204
C8H6IO6-	324.9214
C10H14IO4-	324.9942
325.882904	325.8829
C8H9INO5-	325.9531
C9H13INO4-	325.9895
C11H21INO2-	326.0623
C11H4IO4-	326.916
C9H12IO5-	326.9735
328.168549	328.1685
C7H6IO7-	328.9164
IC8H10O6-	328.9528
C9H14IO5-	328.9891
C9H17INO4-	330.0208
IC8H12O6-	330.9684
C9H16IO5-	331.0048
C6H7INO7-	331.9273
C7H11INO6-	331.9637
C6H6IO8-	332.9113
IC7H10O7-	332.9477
IC7H10O7-	332.9477
C9H4IO6-	334.9058
IC11H12O4-	334.9786

IC12H22O3-	341.0619
C11H4IO5-	342.9109
C9H12IO6-	342.9684
C13H28IO2-	343.114
C9H14IO6-	344.9841
C7H8O8I-	346.9269
C7H8O8I-	346.9269
C11H8IO5-	346.9422
C8H12O7I-	346.9633
C6H7NO8I-	347.9222
C12H15INO3-	348.0102
C10H6IO6-	348.9214
C9H13NO6I-	357.9793
C8H11NO7I-	359.9586
C9H15NO6I-	359.995
370.964508	370.9645
C9H13NO7I-	373.9742
C9H12O8I-	374.9583
C9H12O8I-	374.9583
C8H11NO8I-	375.9535
376.97467	376.9747
388.882935	388.8829
C9H13NO8I-	389.9691
C10H17NO7I-	390.0055
390.092987	390.093
C8H11NO9I-	391.9484
400.911438	400.9114
404.078888	404.0789

Appendix D: Hysplit Trajectories

The full set of hourly back trajectory figures is available in the dataset deliverables for this project. They may also be viewed for a limited time here:

https://herndon.homeunix.net/owncloud/index.php/s/5DeukFZ4CcbEHUM

The calculations have also been done for the UTSA location for the entire campaign, in order to support other research groups doing stationary measurements at this location. A movie of these UTSA back-trajectories can be viewed here:

https://herndon.homeunix.net/owncloud/index.php/s/peG6fXDR9FQtx5r