Quality Assurance Project Plan

Project 19-025
Apportioning the Sources of Ozone Production during the San Antonio Field Study

Prepared for
Texas Air Quality Research Program (AQRP)
The University of Texas at Austin

Prepared by

Tara I. Yacovitch (Principal Investigator)
Aerodyne Research, Inc.

October 5th, 2018
Version #2

Aerodyne Research, Inc. has prepared this QAPP following EPA guidelines for a Quality Assurance (QA) Category III Project: Secondary Data Project/Research Model Development and Applications Project. It is submitted to the Texas Air Quality Research Program (AQRP) as required in the Work Plan requirements.

QAPP Requirements: 0. Cover Page, 1. Project Description and Objectives, 2. Organization and Responsibilities, 3. Scientific Approach (including Model Selection), 4. Quality Metrics (Including Model Calibration and Model Verification). 5. Data Analysis, Interpretation and Management (including Model Documentation), 6. Reporting and 7. References.

QA Requirements: Technical Systems Audits - Not Required for the Project
Audits of Data Quality – 10% Required
Report of Findings – Required in Final Report

Approvals Sheet

This document is a Category III Quality Assurance Project Plan for the "Apportioning the Sources of Ozone Production during the San Antonio Field Study" project. The Principal Investigator for the project is Dr. Tara I. Yacovitch.

Electronic Approvals:

This QAPP was approved electronically on 10/15/2018 by Elena McDonald-Buller, The University of Texas at Austin.

Project Manager, Texas Air Quality Research Program

This QAPP was approved electronically on 9/24/2018 by Vincent M. Torres, The University of Texas at Austin.

Quality Assurance Project Plan Manager, Texas Air Quality Research Program

This QAPP was approved electronically on 09/21/2018 by Tara Yacovitch, Aerodyne Research, Inc.

Tara Yacovitch

Principal Investigator, Aerodyne Research Inc.

QAPP Distribution List

Texas Air Quality Research Program
David Allen, Director
Elena McDonald-Buller, Project Manager

Texas Commission on Environmental Quality
Bright Dornblaser, Project Liaison

Aerodyne Research, Inc.
Tara I. Yacovitch, Principal Investigator

Table of Contents

1 Project Description and Objectives

1.1 System to be studied

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.

This project will study the causes of ozone formation in and around San Antonio.

1.2 Project Purpose and Objectives

This project aims to discover which sources contribute to the formation of ground-level ozone in and around San Antonio, and in what quantities.

Raw data from the 2017 San Antonio Field Study (SAFS) will be examined closely and analyzed in full to identify characteristic sets of VOCs associated with different source types. Computer modeling of air transport will help identify the broad geographic areas where the measured air originated. An ozone formation computer model, in which individual source categories can be turned on, off, or varied, will be used to understand how each source type contributes to ozone formation in and around San Antonio.

The project has the following objectives:

- Perform high-resolution fitting of the mass-spectrometer-acquired data from the San Antonio Field Study (SAFS).
- Constrain the formation of ozone using photochemical box modeling and input data from SAFS.
- Identify the source regions of measured airmasses using HYSPLIT footprint analysis.
- Identify sets of chemical species that vary together using Positive Matrix Factorization (PMF).
- Interpret these covarying species into characteristic source categories (e.g. evaporative condensate emissions vs isoprene and other trees emissions).
- Perform a sector-based apportionment of OH reactivity in the SAFS area using the photochemical box model run with varying contributions of the identified source categories.

2 Organization and Responsibilities

2.1 Personnel

Project roles and responsibilities for scientists at ARI and subcontractor are described in this section. Technical roles and responsibilities are listed separately from quality assurance (QA) responsibilities.

Dr. Tara Yacovitch will serve as PI for this project. She will be supported by Drs. Herndon, Krechmer, Canagaratna, Roscioli, Lerner, Claflin and Mr. Daube. Dr. Berk Knighton, who also participated in the field study that collected this project's data, will serve again as a consultant here.

All project personnel will have the additional QA responsibility to peer-review analysis results presented by other team members during regular project meetings.

Person	Technical Responsibilities	QA Responsibilities
Dr. Tara I. Yacovitch	Principal Investigator Manage the overall project's scientific goals Assist with high-resolution fitting efforts and PMF analysis.	Ensure that QAPP is followed. Ensure timely submission of all required reports and deliverables. 10% Data Quality Audit
Dr. Scott C. Herndon	HYSPLIT footprint calculations. Lead the effort to perform photochemical box modeling.	QA for photochemical box model.
Dr. Manjula Canagaratna	Training and instruction for performing PMF analysis.	Technical oversight of the PMF analysis effort and any conclusions stemming from it. QA oversight of I-CIMS
Dr. Jordan Krechmer	Training and instruction to other team members in their high-resolution fit efforts.	Ensure consistent quality of high-resolution fits of I⁻ CIMS data.
Dr. J. Rob Roscioli	High-resolution fitting of mass-spectrometer data Assist with PMF analysis.	
Dr. Brian Lerner	Lead the efforts to do high-resolution fitting and PMF analysis of GC-TOF data.	Ensure consistent quality of high-resolution fits for GC-ToF data.

Dr. Megan Claflin	Analysis of GC-TOF data.	
Mr. Connor Daube	HYSPLIT footprint calculations Assist Dr. Herndon in these and other analysis efforts.	
Dr. W. Berk Knighton (consultant)	Assist with high-resolution analysis of PTR-ToF data. Identification of individual species and classes of species using the results of the high-resolution fits and PMF analysis.	Oversee efforts to assign uncertainties and response factors to PTR species. QA oversight on results of high-resolution fits for PTR-ToF and GC-ToF
Dr. Ed Fortner	High-resolution fitting of mass-spectrometer data.	

2.2 Schedule and Key Milestones

High-resolution analysis will be undertaken first, Fall 2018. 0D box modeling (to understand ozone formation) and PMF Analysis (to understand source categories) will begin in Fall/Winter 2018, using preliminary results of the high-resolution analysis. First results will feed back into Task 1 to improve the chemical identification of species and to suggest improvements to the high-resolution analysis. Footprint analysis will begin in early 2019, in order to pinpoint emission source areas. Final report draft preparation will begin in the summer of 2019. Time has been allocated during this report preparation phase to re-run the 0D box model, PMF and footprint analyses as necessary with updated results. Throughout, collaboration with other SAFS study participants will be undertaken. The anticipated work schedule is charted below.

Task	Description	Year	18					19							
		Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Task 1	High Res. Analysis				x	x	x	x								
Task 2	0D Box Modeling						x	x	x	x					x	x
Task 3	PMF Analysis						x	x	x	x	x	x	x	x	x	
Task 4	Footprint Analysis									x	x	x	x			
Reporting	Final Report Preparation													x	x	x

Though the high-resolution fit results (Task 1) will ultimately be required before completing Tasks 2 and 3, progress on those tasks can begin even before Task 1 is complete.

Box modeling (Task 2), can begin using only previously-generated field data. This means that any data formatting requirements can be ironed out, and other potential challenges in getting the model to run and converge can be understood in preparation for the high-resolution data.

PMF analysis (Task 3) will be used in part to help identify high-resolution peaks that have been fit but not yet chemically assigned. This means that a preliminary high-resolution fit can be generated and PMF analysis begun even before Task 1 is finalized. Indeed, running PMF on such a set of preliminary results may help speed up the final steps of Task 1 by providing chemical intuition.

Footprint analysis, Task 4, relies only on archived meteorological data and so will be unaffected by delays or setbacks in other tasks.

Finally, final report preparation will require a major effort to synthesize results. However, some of this final report's contents will have been developed over the previous months: it will leverage the progress written up for monthly technical reports. Regular internal project meetings, with their "peer-review" format, will also help hone the major conclusions well in advance of the official start of this task.

3   Scientific Approach

3.1   Data Needed

A good deal of primary data (data which was collected for the purpose of this study) will be used in this project. This data was collected by the Aerodyne Mobile Laboratory (AML) and miniature Aerodyne Mobile Laboratory (minAML) during the 2017 SAFS campaign in the San Antonio region. This data was submitted as part of the AQRP Project 17-053. The quality-assurance document accompanying this pre-existing data is included as Appendices A and B to this document. A list of pre-existing primary data described in these Appendices is shown below:

Table 1. Pre-Existing SAFS Data from AQRP Project 17-053

Trace gas species	AMS data	Benzene and other VOC's	ICIMS data	VOC data from the GC-ToF	Meteorological, Position and other
CO_2	BC	Acetaldehyde	C3H4O3I	Benzene_124trimethyl	latitude
CO	Chl	Acetone	C4H6O4I	Benzene_ethyl	longitude
HCN	NH_4	Acetonitrile	C4H7NO5I	Benzene	TemperatureC
O_3	NO_3	Benzene	C5H8O3I	Cyclohexane_methyl	TrueWind
NO_2	Org	C_2Benzenes	C5H8O4I	Cyclohexane	TrueWind at 3m
NO	SO_4	C_3Benzenes	C5H9NO7I	Cyclopentane_methyl	jNO_3
N_2O		DMS	C5H9O4NI	Heptane_n	jO^1D
HCHO		Isoprene	C6H10O4I	Hexane_n	jHCHOa
HCOOH		MACR_MVK	C6H12O2I	Isoprene_norm_resp	jHCHOb
H_2O_2		Methanol	IC2H2O4	Octane_n	tower state
HONO		Terpenes	IC3H4O2	Pentane_224trimethyl	CO_background
CH_4		Toluene	IC3H4O4	Pentane_iso	site name
C_2H_6			IC5H10O3	Pentane_n	background mask
C_3H_8			IC6H6O	Toluene	
H_2O			IC6H10O5	Xylene_o	
SO_2			IC8H8O3	Xylenes_mp	
THC			IC10H16O3		
			ICH3Cl		
			ICH3NO2		
			ICl2		
			IClNO2		

Primary data will also be generated as part of this project. Task 1 in the workplan document describes the analysis effort that will be undertaken to produce time traces of additional chemical species through high-resolution fitting of raw mass spectrometer data. Three raw time-of-flight (ToF) datasets, all collected during SAFS, will be leveraged for this task:

- PTR-ToF
- I⁻CIMS
- GC-ToF

All of the above data needs to cover the time period of the San Antonio Field Study (SAFS): 5/8/17 00:00 – 5/31/17 18:51 (Times in UTC). The data also needs to cover the geographic areas visited, listed in Table 2.

Table 2. Times and Locations of the AML and minAML during the San Antonio Field Study (SAFS)

time (UTC)	Site Name
5/8/17 0:00	
5/10/17 20:04	UTSA
5/16/17 18:38	Transit
5/16/17 21:46	Floresville
5/21/17 16:38	Transit
5/21/17 21:24	Lake Corpus Christi
5/26/17 0:00	Transit
5/27/17 3:59	UTSA
5/31/17 18:51	

A specific format and time base (aka technological presentation) of the above data is required in order for this data to be used in subsequent PMF analysis.

- All data shall be output in Igor Pro Format (.ibw or .itx file specifications)
- All times will be in universal coordinated time (UTC).
- All data (except for GC-ToF) will be averaged onto the same time base. This time base is a data series containing one data point every minute, and spanning the length of the SAFS study, and is described in detail in Appendix A. Pre-existing data is already on this time base.
- GC-ToF data will be output on its own time-base to account for the need for both a start and stop time for each measurement.

Only a limited amount of secondary data – data not specifically collected for the purpose of this study – will be used:

- Weather data for back-trajectory calculations (HYSPLIT). This data covers a larger geographic bound than the primary data in order to simulate the origin of air parcels. It also needs to be in a format appropriate for use by the back-trajectory model: gridded data in ARL format. The National Centers for Environmental Prediction (NCEP) produces both national and global weather analyses products. The group is part of the National Weather Service. The "High Resolution Rapid Refresh" (HRRR) data is the highest resolution product relevant to the HYSPLIT work proposed here unless a special product is generated for the study domain. The process of collecting the .hrrra meteorological files is complete.
- TCEQ monitoring site data. TCEQ monitoring site data for the duration of the SAFS may also be leveraged in early parts of the project in order to help with data QA. The data in question will need to cover the SAFS time period, and should be geographically as close to the mobile lab as possible. For the Floresville site, for example, the SAFS measurements were co-located with a TCEQ sensor. Any data format capable of being read into Igor Pro is acceptable (e.g. delimited text format).

3.2 Data Sources

National Oceanic and Atmospheric Administration (NOAA) meteorological data archives (Air Resources Laboratory, 2017) will be used to provide gridded weather data for HYSPLIT. This data has been collected and formatted by the Air Resources Laboratory, a division of NOAA, a federal agency with particular expertise in air monitoring, meteorological data collection and weather forecasting. The meteorological datasets from NOAA will thus be preferred over alternative options which would be incomplete and inadequate, such as using measured wind on the mobile laboratories during the SAFS to simulate a weather field over broader geographic span. Several practical considerations also support the use of NOAA weather fields, notably the availability of data at a variety of geographic grid resolutions, and data directly available in a format used by the HYSPLIT model.

3.3 Model Selection

Two project goals require the use of models. The first and most important goal relates to the apportionment of ozone formation to different source categories. Ozone formation is related to OH reactivity, k[OH], which must be modeled. By varying the model inputs to reflect real measured VOC signatures, k[OH] can be attributed to specific sets of VOCs, and, ultimately, to their specific source categories. Models with varying degrees of complexity and of chemical specificity could be used to model k[OH]. However, since this project aims to apportion ozone formation to different source categories, it is important for the selected model to include chemical pathways specific to different individual VOCs.

Limitations in the applicability of the model are expected. For example, some of the species of interest may not be explicitly included in the model's available chemical mechanisms and reaction rates. Errors, inaccuracies or assumptions in the model may also limit its ability to reproduce measured atmospheric concentrations of key tracers. However, this project's use of this model does not aim to further validate the model itself or to use the model to predict future scenarios. Rather, its purpose is to help explain the intertwined chemical mechanisms leading to ozone formation in the San Antonio area. In this context, many valuable insights can still be gleaned despite an imperfect model.

The second modeling goal relates to identifying source regions of measured air. Instead of limiting data interpretation to "upwind" and "downwind" as has been done previously (Schade and Roest, 2016), a specific "footprint", or source region, will be calculated. This source footprint will be used as evidence to help identify common source categories of measured chemical tracers.

This model use will be narrowly focused on the region and time period of the SAFS and will not attempt to forecast future atmospheric transport scenarios. The model will identify a source region for a given measurement time, but the size of this region will have "fuzzy" edges (contours at different levels of probability). Thus, this model's applicability will largely serve as an aid to data interpretation.

3.3.1 Required Capabilities

Modeling Ozone Formation: This model must have been developed for modeling photochemical reactions. The model should rely on chemically explicit equations. It must be applicable to the chemistry and reactions typical of the San Antonio study area (including both biogenic and anthropogenic VOC reaction pathways). It must also allow for independently varying source concentrations and for determining intermediate species like k[OH]. The model need only be applicable to the time period of the 2017 SAFS and does not need to cover multiple locations at the same time (0D model, location independent except for inputs like temperature and solar elevation angle). The model needs a minimum temporal resolution of 1-hour to account for variations in air masses, but better resolutions are also acceptable. It should be compatible with the Igor Pro analysis software, which will be used for the majority of the data analysis for this project.

It must also be well-documented and relatively simple to spin-up and use since this modeling task needs to be balanced with the numerous other analysis efforts necessary for project completion. There is synergy between this simplified modeling and AQRP Project 18-040 led by Dr. Ezra Wood, which will include more sophisticated photochemical box modeling. We will share data and collaborate closely to enhance the impact of this modeling effort.

Modeling Source Footprints: This model should be a well-established back-trajectory model for use in determining the source location of measured air. It should leverage real historical meteorological data in performing the back-trajectories. It must cover a spatial extent much broader than the SAFS study area, including, for example, possible source regions like the Yucatan peninsula (Mexico). It must have a minimum time resolution of 1 hour or better to include changes in weather fronts and air masses. It must allow for determining not just source locations or trajectories, but "footprints": outlines of source areas of interest. Finally, it should be compatible with the Igor Pro analysis software, which is used for the majority of the data analysis for this project.

3.4 Model Identity

Two models have been identified for use in this project:

Modeling Ozone Formation: The model selected is a 0D Photochemical Box Model: the Master Chemical Mechanism (MCM) (Jenkin et al., 2003; Saunders et al., 2003) to model

complex photochemistry. MCM version 3.3.1 (MCMv331) will be used (Jenkin et al., 2015). The MCM is a large, community-developed, near-explicit sequence of chemical reactions that govern the tropospheric oxidation of emitted VOCs (anthropogenic and biogenic). It tracks primary emissions and partial oxidation products that are not commonly quantified. The compounds measured analytically during the field campaign with known calibration factors will be constrained with the reactive intermediate species allowed to run to photo-stationary state. The ensemble of species will be used to calculate OH reactivity. An example use case of such a model is described in Mao et al. (2010).

Several front-ends are available to run the MCM: the Dynamically Simple Model for Atmospheric Chemical Complexity (DSMACC); the Framework for 0-D Atmospheric Modeling (F0AM), or the Kinetic PreProcessor (KPP). Selection of the front-end will be done in consultation with Dr. Wood and his post-doc (Project 18-040). Consideration will be given to primarily to ease-of-use, spin-up, and the ability to achieve the requirements described in Section 3.5.

Modeling Source Footprints: The NOAA HYSPLIT model has been selected for this purpose (Draxler and Hess, 1997; Draxler and Hess, 1998; Draxler, 1999; Stein et al., 2015).

3.5   Model Requirements

PHOTOCHEMICAL MODELING: The model will be run in a mode constrained by observations taken during SAFS. This means that the model will need measurements of temperature, pressure, water vapor, O_3, NO_2, CO, CH_4, HCHO, methanol, acetone, acetaldehyde, isoprene, propane, ethane, ethyne, monoterpenes, toluene, *n*-pentane, *n*-hexane, *n*-heptane, *n*-octane, xylenes, ethyl benzene, 1,2,4-trimethylbenzene, benzene, and cyclohexane in order to anchor the photostationary state. No speciated observations of monoterpenes or xylenes are available, and so the composition will need to be assumed (e.g. all monoterpenes assigned to α-pinene, and the measured total *m*- and *p*-xylene assumed to occur in a 50% split).

HYSPLIT: The HYSPLIT model will be used to calculate footprints using the concentration mode, tracking 50,000-120,000 particles backwards in time from the measurement latitude, longitude and altitude. In order to achieve good resolution for these footprints, a grid of resolution 0.03/0.05-degree or better will be targeted for the study region. Since ground-level sources are of interest, only those particles within 100 m of ground level should be considered. Other model parameters were set to the default values.

The GDAS1 and GFS0p25 meteorological data were used in the in-field simulations. However, in this work, the high-resolution rapid refresh (hrrr) meteorological files will be employed and the calculations rerun. The hrrr files contain much higher resolution

data (~6x) and will require more computer time than was used during the project. The time duration for the backward concentration mode will be extended to 71 hours.

3.6   Quality Requirements

There is no specific data quality requirement for the source data (primary and secondary) for this project.

Accuracy/Precision: No specific accuracy or precision requirements are necessary for this project. However, data precision shall be reported at the 1-sigma level for all primary data generated in this project. If comparable data exists for this newly generated primary data, data accuracy shall also be evaluated by comparison.

Representativeness: All data must be representative of the 2017 SAFS (see Table 2).

Completeness: No specific completeness metric is required for this project. Missing time periods in the primary data due to calibrations, zeroes, instrumentation failure or other causes shall be excised from the dataset.

Comparability: No specific comparability metric is required for this project. For some chemical species (e.g. certain hydrocarbons), comparable data may exist from TCEQ monitoring sites. However, in most cases, the primary data generated for this project will consist of unique chemical tracers with no comparable measurements.

Audit of data quality 10%: Though no specific accuracy/precision metrics are specified for this project, it is important that the primary data generated (results of high-resolution fits) be quality-assured to remove spurious data (e.g. in-field calibrations or zeroing, instrument issues). A 10% data quality audit will be performed by the project PI, Tara Yacovitch. 10% of the generated data series will be selected. The procedure is described in Section 4.2.

3.7   Quality Assessment Procedures

PHOTOCHEMICAL MODELING: The 0-D photochemical model includes a number of components. Each component will be treated separately to assess the model performance. The primary goal of the modeling effort is to understand detailed species contributions to total OH reactivity.

- Mechanism – The MCM has been chosen for this work. No assessment of this well-documented mechanism will be attempted as part of this project. However, a complimentary project by Dr. Wood (Project 18-040) will touch on this component.

- Integrator – the front-end of the model (aka integrator) is used to drive the underlying mechanism. A sensitivity matrix, which can be generated by the integrator for each model run, can highlight which model parameters most impact the results. This result will be used to guide interpretation of errors and bias.
- Modules (TUV) – Certain model parameters were not measured, and require generation through use of an internal model. One such parameter is accurate photolysis rates. The Tropospheric Ultraviolet and Visible radiation (TUV) module will be used to determine solar irradiation, which can be converted to photolysis rate. This model will be assessed and calibrated through use of a measured proxy, the photolysis rate of nitrogen dioxide, or $j(NO_2)$.
- Measurements – Finally, the model will be used with numerous measured inputs (chemical species) and can be run in either a constrained or semi-constrained way. Since OH reactivity is the primary focus, but was not directly measured, other indicators of OH reactivity are needed for quality assessment. Formaldehyde (HCHO) fits the bill: it is produced as part of photolysis reactions of most carbon-containing VOCs, and was measured with high precision during the SAFS by a laser based instrument. The model can thus be run with HCHO unconstrained, and the HCHO results compared with measurements. Reasons for disagreement can then be investigated using the sensitivity matrix, for example.

HYSPLIT: Secondary data includes historical weather data used in HYSPLIT simulations, and (possibly) TCEQ monitoring station data used for intercomparison with measured results. No quality assessment will be performed on the raw weather data. Instead the quality of the resulting HYSPLIT footprints will be evaluated qualitatively. In order to test the model output for consistency, best practice will be followed (personal communication Stein, A.) These will include:

- For a subset of the model runs, adjustments to the lowest level height will be performed and compared to the base run of 100m. Qualitative agreement should be observed to gain confidence that the choice of layer height is not excessively steering the result.
- For a subset of the model runs, the hrrr meteorological database will be substituted with the NAM hybrid sigma pressure coordinate data. If the results are relatively similar with different met data sets, we will be more confident that meteorological uncertainties are not playing a significant role.

TCEQ Data: The quality of the TCEQ data will also not be directly assessed. Indeed, should it be relevant, the TCEQ data will be used to evaluate the quality of the primary data used or generated as part of this project.

Since some primary data will be generated as part of this project, those quality assessment procedures are discussed in the Section 5 dealing with data analysis.

3.7.1   Quality Disclaimer

No project-specific data quality requirement exists for this project. Furthermore, the source data will not be evaluated by the EPA. The following data quality disclaimer will be included with all project deliverables:

"The quality of the source data for this deliverable has not been evaluated by the EPA"

3.8   Model Calibration

PHOTOCHEMICAL MODELING: The photochemically driven master chemical mechanism model itself does not intrinsically rely on parameters that require calibration or tuning. The explicit MCM encompasses a vetted kinetics model with atmospheric pressure and temperature dependencies built-in. One important model component, the TUV module used to determine photolysis rate, does require calibration. Measurements of the photolysis rate of NO_2, $j(NO_2)$, are available, but do not directly account for all the other light-reactive species present in the atmosphere. TUV module calibration will involve translating $j(NO_2)$ into a representative proxy for all other photolysis rates.

HYSPLIT: No model calibration will be attempted for the HYSPLIT model used to determine footprints of measured airmasses. Expert judgment will be used to evaluate whether the model results are reasonable; Dr. Herndon and Mr. Daube have both taken the week-long NOAA training workshop on HYSPLIT use. They maintain relationships with and receive counsel from the HYSPLIT user community through participation on the bulletin board.

3.8.1   Procedure and Data used for Model Calibration

PHOTOCHEMICAL MODELING: The data that will be used to convert the measured NO_2 photolysis rate, $j(NO_2)$ into other photolysis rates will rely on the clear-sky calculations using the National Center for Atmospheric Research Tropospheric Ultraviolet and Visible Radiation Model (TUV). The first order approach is to scale the ratio of clear-sky photolysis rates by the empirical $j(NO_2)$. This requires other calibrations based of the likely influences of cloud and aerosol loading on the relative differences in the wavelength-resolved cross section (Stark et al., 2007; Bohn et al., 2008; Vaida, 2009).

HYSPLIT: The high-resolution rapid refresh meteorological assimilations that will be used for the footprint calculations is hourly data, 3 km horizontal resolution with 37 layer in the vertical. The data covers the continental U.S. and portions of Mexico and Canada. Meteorological files at coarse resolution will be added to better represent air flow from the Gulf of Mexico.

3.8.2 Acceptance Criteria for Calibration

PHOTOCHEMEICAL MODEL: Since the primary goal of the photochemical model is to perform a species-based allocation of OH reactivity, the photolysis rate (as assessed via HCHO, see Section 3.7) will be the primary quality acceptance metric. The target acceptance criteria for this metric will be a < 20% difference from measured HCHO.

HYSPLIT: In many of the HYSPLIT best practice procedures recommended to end-users, a qualitative agreement is the target acceptance criteria (NOAA, 2018). For this project though, because we are taking the additional step of unwinding the binary results file into the Igor Pro analysis software, we will compute and document a spatially resolved difference between calculations for 5% of the footprints in this project.

1. Runs setting the lowest layer to 75 and 125 m will be compared to the base run of 100m.
2. For three different transport conditions (a. from NW "continental"; b. from SW "Mexico" and c. from SE "gulf flow"), the hrrr meteorological database will be substituted with the NAM hybrid sigma pressure coordinate data. Each condition will be compared to the base run.

The results of the numerical differences in footprint will need evaluation prior to establishing a numerical acceptance criteria. We will consult with other experienced users of HYSPLIT for advice.

3.9 Model Verification

DSMACC/MCM: Model verification will occur in the initial phases of model use. The numerical accuracy of the built model will be verified by running install-test-case. This will only verify that the model and components were installed properly. The actual usage challenges described below will be performed.

HYSPLIT: No model verification will be attempted for the HYSPLIT model since this model is well-established, developed and maintained by a respected federal research body, and will be used to support data interpretations at a qualitative level.

3.9.1 Model Verification Approach

PHOTOCHEMICAL MODEL: The various outputs of the box model that are relevant to the science objectives (e.g. OH reactivity and reactivity classes, photostationary radical levels) will eventually be generated with the model run in a mode constrained by measurements. In the constrained mode, when the integrator computes an instantaneous change for the time interval, d[X]/dt where X is a measured species, it is

actually held at the measured level. In order to validate the model, however, we will let test species, (formaldehyde; acetone) float to a photostationary state based on balancing production and loss rates. The model predicted number density of the test species will be compared to the measured level and a numeric error or model bias established. We will examine the dependence of the bias on other factors, such as overall VOC loading, sunlight, ambient temperature to ascertain whether the bias is explained by an input. By this mechanism, we will identify the sources of bias.

3.9.2   Model Uncertainty and Sensitivity

PHOTOCHEMICAL MODEL: The integrator will run in a mode to calculate the model sensitivities. These can be coupled to the empirical uncertainties in the analytical measurements. An overall uncertainty in the modeled vectors of interest, such as OH reactivity will be computed. This approach is likely to result in some measurement uncertainties being more important than others. This will be a useful result for future measurement project design.

3.9.3   Data requirements

PHOTOCHEMICAL MODEL: We anticipate that the model validation and uncertainty estimate will be most robust if data from the three stationary measurement locations are combined. In this way, the model will be challenged to reproduce the test species (formaldehyde) during periods of a) relatively clean gulf air; b) air with oil & gas emissions; c) mixed urban air. The data will need to be free of time periods where the assumption of photostationary state does not hold, such as the recent passage of a recreational water craft upwind. Depending on how long the model takes to run to photostationary state, we anticipate working on two time bases: a) when the GC-TOF collected samples and b) 1 minute time base for all other measurements.

3.9.4   Criteria for Verification

PHOTOCHEMICAL MODEL: The independent data in the model verification challenge is the measurements of formaldehyde and acetone. Each of these measurements is certain to better than 15%. A model predicted concentration vector that differs by no more than 20% will be considered acceptable for the work proposed here. It is important to note though, that the model/measurement bias will be probed for cause in other independent variables. It will be part of the research endeavor to employ appropriate analysis techniques based on the results of that investigation.

4   Data Analysis, Interpretation and Management

4.1   Data Reporting Requirements

Results of the high-resolution fits will be reported on one of two acceptable time bases (a one-minute time base for most species; or a start-stop time base for GC-ToF data). Internally-developed software has been written to aid in this interpolation onto unified time bases. A selection of this source code is included as Appendix C. Data will be accompanied by a QA document similar to the one reproduced in Appendix A.

Results of the HYPSLIT footprint analyses will be output as .png files and as Igor Pro-readable .itx files for ease of making contour plots and other graphs.

All other analysis tasks outlined in the workplan will report results in monthly technical reports and in the final report.

4.2   Data Validation Procedures

High-Resolution Fitting: In order to ensure consistent high-quality fitting among all project participants, an internal analysis workshop has been planned. This workshop is organized by Dr. Jordan Krechmer, and will be given by Dr. Harald Stark, the scientist who has written the analysis software. The course will include the following topics:

- data loading
- mass calibration
- instrument function definition (peak width, peak shape)
- high-resolution peak identification
- high-resolution peak assignment
- high-resolution peak fitting or integration
- high-resolution time series plotting
- elemental analysis
- data export for external analysis / archiving

A training video covering these topics is also available (requires permission to access): https://sites.google.com/site/citofms/analysis-software/user-manual#TOC-Training-videos

No single data quality metric is appropriate for determining the validity of high-resolution fits. However, a recent peer-reviewed publication by Cubison and Jimenez (2015) outlines the many considerations involved in multivariate fitting of high-resolution data and quantifies the subsequent effects on the precision of the data. Cubison and Jimenez further state:

Imprecision in such a constrained fitting procedure may arise from (i) noise in the measurement distribution, particularly from counting statistics of the ions of interest; (ii) the m/Q calibration, which is limited itself by statistical noise on the reference ions used to determine it; (iii) the discrete nature of the measurements, i.e. the spacing of the discrete measurement points; and (iv) the instrument transfer function ("peak shape"), errors in the determination of which propagate directly into retrieval of ion signals from the fits yet must still be empirically determined and thus will contain errors.

Such causes of imprecision will be examined as part of the high-resolution fitting procedure.

Calibrating fit results: The raw units of the high-resolution fits will be in ion counts. A response factor is necessary to convert these into real units of concentration like parts-per-billion (ppb).

Several species have had their response factors directly measured with data from in-field calibrations. This is the best-case scenario, and these direct calibration factors will be applied to each species possible.

However, the bulk of the analysis for this project will rely on the numerous other species without direct calibration experiments available. One strategy to determine response factors for such species is to first correlate instrument sensitivity in ion counts per ppb with known reaction rate for the chemical ionization reaction in question. For the PTR instrument, the relevant reaction rate is k^{H3O+}, or the rate of proton transfer; for the I-CIMS instrument, it will be k^{I-}, the rate of iodide ion transfer, in units of 10^{-9} cubic centimeters per molecule second (10^{-9} cm^3/molec*s). This correlation plot is constructed for those chemical species with known calibration factors (e.g. acetone, toluene, xylene, MEK, etc. for the PTR). Then, an estimated response factor for other species without calibrations can be determined if a literature reaction rate (either measured or calculated) is found. Seikimoto et al. (2017) have recently published a list of such rate constants calculated for numerous PTR-MS species.

The final fallback option, when no direct calibration or indirect sensitivity factor is available, will be to produce a time trace using a simple relative intensity versus a known species like acetone. This final option, though it does not produce data in ppb, is still extremely valuable, particularly in the context of the positive matrix factorization (PMF) analysis task. PMF looks at correlations between species, and is inherently insensitive to the input units. In fact, performing PMF on uncalibrated ion count data, or even on uncalibrated unit mass resolution data (instead of high-resolution data), can provide valuable insights into ions of interest, and can help identify those species deserving additional attention in the high-resolution fitting procedure.

10 % Data Audit: Tara Yacovitch will perform 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.

PMF: Positive Matrix Factorization (PMF) is a matrix-based technique used to identify covarying chemical tracers. In order to ensure consistent high-quality analysis procedures among all project participants, an internal analysis course has been planned. Dr. Manula Canagaratna will give this course and serve as technical oversight for all PMF analysis in this project. The course will follow the contents of the "PMF-AMS Analysis Guide" produced by CIRES and the Jimenez Research group ("Jimenez Research Group Wiki", 2018b).

Positive matrix factorization (PMF) is an advanced analysis technique with no specific validation procedure. However, Dr. Manjula Canagaratna is an expert in this method. She will serve as technical oversight for this type of analysis and will evaluate the results. The validation procedures for this analysis will be similar to those involved typical peer-review: data will be presented and explained in written or presentation format. Other project scientists, led by Dr. Canagaratna, will then identify inconsistencies or problems with the analysis results or interpretation. These problems will then be addressed, and the peer-review performed again until all parties agree on the conclusions.

4.3 Data Description Procedures

High-Resolution Fits: New primary data (time-series from high-resolution fits) that will be produced as part of this project will be described in a data QA document to accompany the final data deliverable. This document will resemble Appendix A.

The data description will include, 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 in section below).

Analysis results: All other results from this study (0D Box model output, HYSPLIT simulations, PMF analysis) will be described in a format similar to a paper draft: figures will be produced and discussed, conclusions will be deduced. Intermediate results will be reported as part of the monthly technical reports, and will be summarized in the final report.

4.3.1   Descriptive Statistics

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.

4.4   Model Documentation

No model development will occur during this project. Instead, existing well-established models (MCM, HYSPLIT) will be used to better understand the causes of ozone during the SAFS.

The following information will be included in deliverables that include model results:

- Model Description
    - references to model documentation: ("The Master Chemical Mechanism", 2018a) for MCM and Draxler et al. (1999) for HYSPLIT
    - underlying assumptions

- Description of input data
    - data base description, including actual input data (type and format) used
    - parameter values and sources
    - boundary conditions used in the model (if different from model default)

- Description of model results
    - overview of the immediate (non-manipulated or -post processed) results of the model runs (if post-processing is done)

      o   output of model runs and interpretation

4.5   Data Storage Requirements

The storage requirements for the results of the experimental SAFS campaign (AQRP Project 17-053) were 618 MB for the final analyzed data; for the total project, the storage required was 694 GB, due largely to the raw ToF data from numerous instruments. For the current project, we expect an additional ~1 GB of data from the additional high-resolution mixing ratio traces.

The PI will retain all data, analysis products and reports, whether in electronic or hard copy format, for a minimum of five years.

5   Reporting

5.1   Expected Deliverables

AQRP requires certain reports to be submitted on a timely basis and at regular intervals. A description of the specific reports to be submitted and their due dates are outlined below. One report per project will be submitted (this project has no offical collaborators). The lead PI will submit the reports, unless that responsibility is otherwise delegated with the approval of the Project Manager. All reports will be written in third person and will follow the State of Texas accessibility requirements as set forth by the Texas State Department of Information Resources. Report templates and accessibility guidelines found on the AQRP website at http://aqrp.ceer.utexas.edu/ will be followed.

Abstract: At the beginning of the project, an Abstract will be submitted to the Project Manager for use on the AQRP website. The Abstract will provide a brief description of the planned project activities, and will be written for a non-technical audience.

Abstract Due Date: Friday, August 31, 2018

Quarterly Reports: Each Quarterly Report will provide a summary of the project status for each reporting period. It will be submitted to the Project Manager as a Microsoft Word file. It will not exceed 2 pages and will be text only. No cover page is required. This document will be inserted into an AQRP compiled report to the TCEQ.

Monthly Technical Reports (MTRs): Technical Reports will be submitted monthly to the Project Manager and TCEQ Liaison in Microsoft Word format using the AQRP FY16-17 MTR Template found on the AQRP website.

Financial Status Reports (FSRs): Financial Status Reports will be submitted monthly to the AQRP Grant Manager (Maria Stanzione) by each institution on the project using the AQRP FY16-17 FSR Template found on the AQRP website.

Draft Final Report: A Draft Final Report will be submitted to the Project Manager and the TCEQ Liaison. It will include an Executive Summary. It will be written in third person and will follow the State of Texas accessibility requirements as set forth by the Texas State Department of Information Resources. It will also include a report of the QA findings.

Final Report: A Final Report incorporating comments from the AQRP and TCEQ review of the Draft Final Report will be submitted to the Project Manager and the TCEQ Liaison. It will be written in third person and will follow the State of Texas accessibility requirements as set forth by the Texas State Department of Information Resources.

Project Data: All project data is listed below and will be submitted to the AQRP Project Manager within 30 days of project completion (September 30, 2019). The data will be submitted in a format that will allow AQRP or TCEQ or other outside parties to utilize the information. It will also include a report of the QA findings.

- 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.)
- Data description including calibration factor or scaling factor applied.
- Time series of total OH reactivity
- All other analysis results to be submitted in monthly technical and final reports.

AQRP Workshop: A representative from the project will present at the AQRP Workshop in the first half of August 2019.

Presentations and Publications/Posters: All data and other information developed under this project which is included in published papers, symposia, presentations, press releases, websites and/or other publications shall be submitted to the AQRP Project Manager and the TCEQ Liaison per the Publication/Publicity Guidelines included in Attachment G of the Subaward.

5.2 Expected Final Products

All of the following data products will use primary data (pre-existing and generated as part of this project). Depending on the specific product, secondary data (HYSPLIT weather; TCEQ monitoring station results) may also be included.

Data Products: Final Data, Technical reports, Draft Final and Final Report, Scientific Journal Articles, Scientific conference presentations and posters.

6 References

2018a. The master chemical mechanism, version 3.3.1. Available at http://mcm.leeds.ac.uk/MCM/. Accessed 09/17/2018.

2018b. Jimenez research group wiki: PMF-AMS analysis guide. Available at http://cires1.colorado.edu/jimenez-group/wiki/index.php/PMF-AMS_Analysis_Guide. Accessed 09/01/2018.

Air Resources Laboratory. 2017. Gridded meteorological data archives. Accessed Sep 1, 2018.

Bohn B, Corlett GK, Gillmann M, Sanghavi S, Stange G, Tensing E, Vrekoussis M, Bloss WJ, Clapp LJ, Kortner M, Dorn HP, Monks PS, Platt U, Plass-Dülmer C, Mihalopoulos N, Heard DE, Clemitshaw KC, Meixner FX, Prevot ASH, Schmitt R. 2008. Photolysis frequency measurement techniques: Results of a comparison within the accent project. *Atmos Chem Phys* **8**(17): 5373-5391. DOI:10.5194/acp-8-5373-2008.

Cubison MJ, Jimenez JL. 2015. Statistical precision of the intensities retrieved from constrained fitting of overlapping peaks in high-resolution mass spectra. *Atmos Meas Tech* **8**(6): 2333-2345. DOI:10.5194/amt-8-2333-2015.

Draxler RR, Hess GD. 1997. Description of the HYSPLIT_4 modeling system. Silver Spring, MD: 24. ERL ARL-224.

Draxler RR, Hess GD. 1998. An overview of the HYSPLIT_4 modeling system of trajectories, dispersion, and deposition. *Aust Meteor Mag* **47**: 295-308.

Draxler RR. 1999. HYSPLIT_4 user's guide Silver Spring, MD: NOAA Air Resources Laboratory. ERL ARL-230. Available at https://www.arl.noaa.gov/documents/reports/arl-230.pdf. Accessed 2017-10-06.

Jenkin ME, Saunders SM, Wagner V, Pilling MJ. 2003. Protocol for the development of the master chemical mechanism, mcm v3 (part b): Tropospheric degradation of aromatic volatile organic compounds. *Atmospheric Chemistry and Physics* **3**: 181-193.

Jenkin ME, Young JC, Rickard AR. 2015. The mcm v3.3.1 degradation scheme for isoprene. *Atmospheric Chemistry and Physics* **15**(20): 11433-11459. DOI:10.5194/acp-15-11433-2015.

Mao J, Ren X, Chen S, Brune WH, Chen Z, Martinez M, Harder H, Lefer B, Rappenglück B, Flynn J, Leuchner M. 2010. Atmospheric oxidation capacity in the summer of houston 2006: Comparison with summer measurements in other metropolitan studies. *Atmos Environ* **44**(33): 4107-4115. DOI:10.1016/j.atmosenv.2009.01.013.

NOAA. 2018. HYSPLIT cheat sheet. Available at https://ready.arl.noaa.gov/documents/ppts/Cheat_Sheet_2018.pdf. Accessed 09/10/2018.

Saunders SM, Jenkin ME, Derwent RG, Pilling MJ. 2003. Protocol for the development of the master chemical mechanism, mcm v3 (part a): Tropospheric degradation of

non-aromatic volatile organic compounds. *Atmospheric Chemistry and Physics* **3**: 161-180.

Schade GW, Roest G. 2016. Analysis of non-methane hydrocarbon data from a monitoring station affected by oil and gas development in the eagle ford shale, Texas. *Elem Sci Anth* **4**(000096): 17. DOI:10.12952/journal.elementa.000096.

Sekimoto K, Li S-M, Yuan B, Koss A, Coggon M, Warneke C, de Gouw J. 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.

Stark H, Lerner BM, Schmitt R, Jakoubek R, Williams EJ, Ryerson TB, Sueper DT, Parrish DD, Fehsenfeld FC. 2007. Atmospheric in situ measurement of nitrate radical (NO3) and other photolysis rates using spectroradiometry and filter radiometry. *J Geophys Res, [Atmos]* **112**(D10). DOI:doi:10.1029/2006JD007578.

Stein AF, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, Ngan F. 2015. Noaa's HYSPLIT atmospheric transport and dispersion modeling system. *Bull Am Meteorol Soc* **96**(12): 2059-2077. DOI:10.1175/bams-d-14-00110.1.

Vaida V. 2009. Spectroscopy of photoreactive systems: Implications for atmospheric chemistry. *The Journal of Physical Chemistry A* **113**(1): 5-18. DOI:10.1021/jp806365r.