Scope of Work

Project #22-003
Evaluating the Ability of Statistical and Photochemical Models to Capture the Impacts of Biomass Burning Smoke on Urban Air Quality in Texas

Prepared for

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

By

Matthew J. Alvarado
Atmospheric and Environmental Research, Inc.

September 9, 2022
Version #2

QA Requirements: Audits of Data Quality: 10% Required
Report of QA Findings: Required in Final Report

NOTE: The Workplan package consists of three independent documents: Scope of Work, Quality Assurance Project Plan (QAPP), and budget and justification. Please deliver each document (as well as all subsequent documents submitted to AQRP) in Microsoft Word format.

Approvals

This Scope of Work was approved electronically on **<u>08/24/2022</u>** by Elena McDonald-Buller, The University of Texas at Austin

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

This Scope of Work was approved electronically on **<u>09/12/2022</u>** by Chola Regmi, Texas Commission on Environmental Quality

Chola Regmi
Project Liaison, Texas Commission on Environmental Quality

Scope of Work

1 Introduction

Understanding the impact of domestic fire smoke on urban air quality (AQ) requires understanding (i) the chemistry of the smoke before it reaches the city and (ii) the changes in the urban production rate of O_3 and $PM_{2.5}$ caused by the smoke. The relative importance of these two pathways on the air quality impacts of domestic fire smoke is not well understood and it is unclear which processes should be targeted to reduce the overall uncertainty.

In addition, three-dimensional (3D) photochemical models like CAMx can have trouble representing the near-source chemistry of the smoke plume and the impact of smoke mixing with urban pollution due to a combination of low spatial resolution near fires and incorrect representation of the chemistry of smoke-specific VOCs (e.g., Baker et al., 2016). These limitations in physical approaches have led to the development of a variety of statistical approaches to estimate the impact of biomass burning on urban AQ (e.g., Gong et al., 2017; de Foy et al., 2021). However, little work has been done to compare the statistical and 3D photochemical approaches or to identify priorities for further development of both approaches. Thus, the US EPA and US Forest Service organized assessment of smoke research needs noted this was a key priority for future smoke chemistry research (Alvarado et al., 2022). ***A statistical analysis of the impacts of domestic fire emission on urban air quality in Texas and a statistical evaluation of the ability of the CAMx model to simulate these impacts would greatly help TCEQ air quality managers understand the impacts of domestic fires on Texas air quality and human health.***

Thus, the objectives of this project are to:

(1) Use generalized additive models (GAMs) driven with satellite and surface observations to examine the impact of fires on background and total O_3 and $PM_{2.5}$ in Texas urban areas.

(2) Examine the ability of CAMx photochemical model to simulate these fire impacts by applying similar statistical methods to the CAMx results

(3) Use any statistically significant differences found to prioritize different approaches to improve the ability of CAMx to simulate the impacts of domestic fires on air quality

In this project we will examine the impact of fires on urban AQ in Texas using statistical modeling (Task 1, Section 3.1). Two urban areas will be examined: Houston-Galveston-Brazoria (HGB) and El Paso. Background O_3 and $PM_{2.5}$ concentrations will be estimated using the lowest value observed at sites near the border of the area of interest, as TCEQ has done in the past (e.g., Berlin et al., 2013). Analyzing the impacts of fires on background and urban sites separately will allow us to examine the change in O_3 and $PM_{2.5}$ due to the mixing of smoke with urban pollution separately from the impact of smoke before it mixes with urban pollution. We will then apply the same statistical methods to both the real-world surface observations and CAMx-simulated surface observations to determine if the impact of fires on urban air quality as simulated in CAMx is statistically equivalent to the impacts seen in the real-world data (Task 2, Section 3.2). We will examine any statistically significant differences to determine avenues for improving the handling of smoke and urban air chemistry in the photochemical models. (Task 3, Section 3.3).

2 Models and Data Sources

2.1 Generalized Additive Models

GAMs are a form of linear modeling which allows non-linear functions of individual predictors within a regression framework (Wood, 2017). This is like standard linear regression techniques, which optimize scalar coefficients (a_k) for each predictor (x_k) for k = 1 to p:

$$\hat{y} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_p x_p$$

except that the coefficients are replaced with potentially non-linear smooth functions:

$$\hat{y} = s_0 + s_1(x_1) + s_2(x_2) + \ldots + s_p(x_p)$$

The GAM approach optimizes these smooth functions. An advantage of GAMs over neural networks and similar machine learning techniques is that it is easy to isolate the effects of the individual variables in GAMs. This will allow us to separate the impact of fire-related variables from the impact of the rest of the predictors. A potential disadvantage relative to standard linear regression is that the smooth functions can overfit the data, and so care needs to be taken to ensure that the derived smooth functions are realistic and robust to changes in the training data set. To address this, the R package mgcv includes routines to fit GAMs, examine the models graphically, and test their robustness via k-fold cross-validation and other techniques.

Below we describe the data sources that will be used to derive the O_3 and $PM_{2.5}$ dependent variables (the predictands or "y" variables, Section 1.2.1.1) and for the meteorological (Section 1.2.1.2) and fire/smoke (Section 1.2.1.3) predictor (independent, "x") variables for a ten-year period (2012-2021).

2.2 Ambient Air Quality Data

Maximum daily 8-hour average (MDA8) O_3 mixing ratios and daily average $PM_{2.5}$ concentrations will be calculated for each HGB and El Paso site from the surface air quality data in TAMIS. Sites in each urban area will be separated into background (for sites on the outskirts of the city) and urban (for sites near the city core). For each urban area, the minimum value of MDA8 O_3 and daily average $PM_{2.5}$ from background sites upwind of the urban core will be selected as the background estimate for that day. As the concentrations of O_3 and $PM_{2.5}$ can vary within the urban core, all sites in the urban core will be included in the urban GAM model. The impacts of the fire and meteorological predictors will be assumed to be the same for all sites, with an additional site-specific random intercept used to account for the mean differences between sites (i.e., a Generalized Additive Mixed Effects Model, or GAMM, where the only random effect is in the intercept, so that the intercept varies between sites, but all other effects are the same between sites).

2.2.1 Meteorological Predictors

The meteorological predictors to be used in this study (Table 1) are based on our previous GAM studies of the ability of meteorological predictors to estimate the concentrations of O_3 and $PM_{2.5}$ at urban and background monitoring sites in Texas. Our team has shown that these meteorological predictors, plus the previous day's MDA8 O_3 or daily average $PM_{2.5}$, can explain approximately 70% of the variability in background and urban O_3, and about 30-40% of the variability in background and urban $PM_{2.5}$ (e.g., Alvarado et al., 2015b; McVey et al., 2018; Pernak et al., 2019; Brown-Steiner et al., 2021).

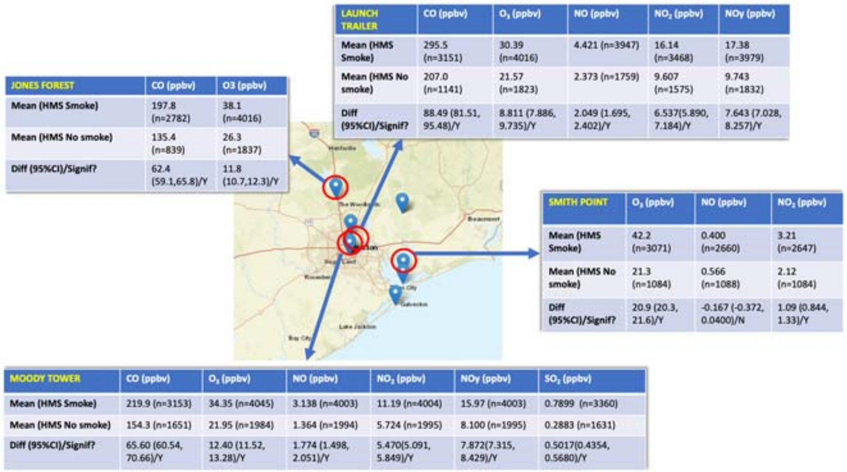
Table 1. Meteorological parameters used in the GAMs. The column name is given in italics.

1)	Afternoon mean temperature (^{o}C, *afternoon_mean_T, 1-4 PM CST*)
2)	Diurnal temperature change (^{o}C, *diurnal_T*)
3)	Daily average wind speed (m/s, *daily_ws*)
4)	Daily average wind direction (degrees clockwise from North, *daily_wd*)
5)	Daily average water vapor density (g/m^3, *SWVP*)
6)	Morning surface temperature difference (1200 UTC) (temperature at 925 or 700 mb–temperature at surface at 1200 UTC) (^{o}C, *T_dif_925mb* or *T_dif_700mb*)
7)	Transport direction (degrees clockwise from North, *HYSPLIT_Bearing*)
8)	Transport distance (m, *HYSPLIT_dist*)

Variables 1-5 in Table 1 were calculated from the surface meteorological data in the Texas Air Monitoring Information System (TAMIS). Variable 6, which reflects the vertical stability of the atmosphere each day, was calculated from upper atmosphere data in the Integrated Global Radiosonde Archive (IGRA Version 2). Given El Paso's higher elevation, an upper atmosphere level of 700 mbar was used for this city as opposed to the 925 mbar value used for all other urban areas. Variables 7 and 8 were calculated from 24-hour NOAA Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT) back-trajectories driven with 12 km horizontal resolution NAM data. As in Camalier et al. (2007), these back-trajectories are calculated assuming an initial height of 300 m above ground level (AGL) and are started at noon local solar time. The endpoints of the back-trajectories were used to calculate the 24-hour transport direction and distance for each urban area for the 2007-2016 period.

2.2.2 Fire and Smoke Predictors

The NOAA Hazard Mapping System (HMS) Fire and Smoke product will be our primary source of fire predictors, as we have shown that days where the HMS product indicates smoke over Houston tend to be associated with enhancements in CO, O_3, and NO_x (Figure 1). **This gives us confidence in the ability of our statistical modeling to be able to identify a statistically significant impact of fires on air quality in HGB and El Paso.** To make the HMS, satellite analysts compare automated fire detections to the infrared satellite images used to produce them to ensure each fire exists (Ruminski et al., 2006; Schroeder et al., 2008; Brey et al., 2018). Small fires are more difficult to detect and are underreported (*e.g.*, Hu et al., 2016). False fire detections are removed, and fires that were not automatically detected are added manually.



LAUNCH TRAILER	CO (ppbv)	O$_3$ (ppbv)	NO (ppbv)	NO$_2$ (ppbv)	NOy (ppbv)
Mean (HMS Smoke)	295.5 (n=3151)	30.39 (n=4016)	4.421 (n=3947)	16.14 (n=3468)	17.38 (n=3979)
Mean (HMS No smoke)	207.0 (n=1141)	21.57 (n=1823)	2.373 (n=1759)	9.607 (n=1575)	9.743 (n=1832)
Diff (95%CI)/Signif?	88.49 (81.51, 95.48)/Y	8.811 (7.886, 9.735)/Y	2.049 (1.695, 2.402)/Y	6.537(5.890, 7.184)/Y	7.643 (7.028, 8.257)/Y

JONES FOREST	CO (ppbv)	O3 (ppbv)
Mean (HMS Smoke)	197.8 (n=2782)	38.1 (n=4016)
Mean (HMS No smoke)	135.4 (n=839)	26.3 (n=1837)
Diff (95%CI)/Signif?	62.4 (59.1,65.8)/Y	11.8 (10.7,12.3)/Y

SMITH POINT	O$_3$ (ppbv)	NO (ppbv)	NO$_2$ (ppbv)
Mean (HMS Smoke)	42.2 (n=3071)	0.400 (n=2660)	3.21 (n=2647)
Mean (HMS No smoke)	21.3 (n=1084)	0.566 (n=1088)	2.12 (n=1084)
Diff (95%CI)/Signif?	20.9 (20.3, 21.6)/Y	-0.167 (-0.372, 0.0400)/N	1.09 (0.844, 1.33)/Y

MOODY TOWER	CO (ppbv)	O$_3$ (ppbv)	NO (ppbv)	NO$_2$ (ppbv)	NOy (ppbv)	SO$_2$ (ppbv)
Mean (HMS Smoke)	219.9 (n=3153)	34.35 (n=4045)	3.138 (n=4003)	11.19 (n=4004)	15.97 (n=4003)	0.7899 (n=3360)
Mean (HMS No smoke)	154.3 (n=1651)	21.95 (n=1984)	1.364 (n=1994)	5.724 (n=1995)	8.100 (n=1995)	0.2883 (n=1631)
Diff (95%CI)/Signif?	65.60 (60.54, 70.66)/Y	12.40 (11.52, 13.28)/Y	1.774 (1.498, 2.051)/Y	5.470(5.091, 5.849)/Y	7.872(7.315, 8.429)/Y	0.5017(0.4354, 0.5680)/Y

Figure 1. Map of selected surface air quality monitoring sites in HGB. Tables show concentrations of species measured at each circled site, divided into HMS-smoke and HMS-no smoke categories.

After identifying fire locations, HMS analysts use imagery from multiple NOAA and NASA satellites to identify the geographic extent of smoke plumes (Rolph et al., 2009; Ruminski et al., 2006). Due to the frequent interference by cloud cover, the number and extent of smoke plumes reported in the HMS represents a conservative estimate.

In addition to the HMS, we will use the Fire Inventory from NCAR (FINN, Wiedinmyer et al., 2011) to obtain high resolution (~1 km^2) estimates of total NO$_x$, volatile organic compounds (VOC), and PM$_{2.5}$ emissions from fires. These emissions will be convolved with daily estimates of upwind influences on HGB and El Paso measurements (i.e., upwind surface "footprints" calculated from meteorological models; see Section 1.2.2) to determine the daily transport of O$_3$ precursors and primary PM$_{2.5}$ to each city.

We will test several fire and smoke predictors based on the HMS and FINN data, including:

a. A binary predictor for the presence of smoke according to the NOAA HMS (Figure 1)
b. The total HMS fire counts in surrounding regions binned by distance from the city (e.g., every 500 km out to 2000 km from the city) and direction (e.g., NE, SE, NW, or SW quadrant). The exact criteria for each distance and wind direction bin will be determined during model training.
c. The FINN NO$_x$, VOC, and PM$_{2.5}$ emissions multiplied by the daily surface "footprints" (Section 1.2.2). The footprints will be separated into domestic and international surface influences to isolate the impacts of domestic fire emissions, and into fresh and aged smoke to examine how smoke age affects the impacts.

2.3 Surface footprints: STILT

The Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al., 2003; Nehrkorn et al., 2010) is an enhanced version of the HYSPLIT model (Draxler and Hess, 1998) aimed at mass conservation, a critical consideration for inversion work. STILT computes the "footprint" (adjoint of the transport field, Figure 2) by following an ensemble of tracer particles backwards in time from the location of each measurement ("receptor"). The footprint (units: ppm/μmol m^{-2} s^{-1}) quantifies the concentration enhancement at the receptor at each point in time due to unit surface flux at each upwind location. Here we will generate daily STILT footprints for each urban area driven with 12 km NAM meteorological data. As with the HYSPLIT predictors in Section 1.2.1.2, these footprints will be initialized at local noon at a height of 300 m and will go at least 72 hours back in time.

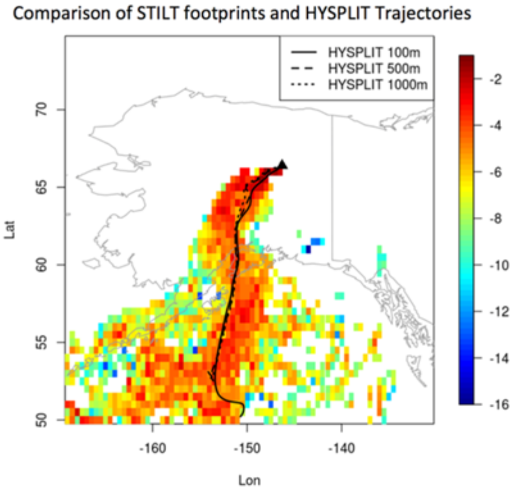


Figure 2. STILT footprints (shaded, units of ppmv (μmol m^{-2} s^{-1})$^{-1}$) and HYSPLIT back-trajectories from receptors located at 100 (solid), 500 (dashed), and 1000 m (dotted black lines) at 19:35 UTC 20 August 2012. From Henderson et al. (2015).

Note that, as STILT does not account for the chemistry along the transport path, it cannot directly account for the chemical formation of O_3 and $PM_{2.5}$ as the smoke is transported from the fire. Instead, we will use the STILT footprints multiplied by the FINN fire emissions as an indicator of smoke transport and see how this smoke indicator correlates with O_3 and $PM_{2.5}$. We will also examine how the predictions depend on smoke age by segregating the footprint predictors into fresh smoke (< 24 hours transport time from fire to city) and aged smoke (≥ 24 hours).

2.4 CAMx

CAMx is a widely used 3D Eulerian chemical transport model that simulates regional air quality over spatial scales ranging from neighborhoods to continents. Meteorological inputs are supplied to CAMx from separate weather prediction models (e.g., Weather Research and Forecasting Model, WRF) while emission inputs are supplied from external pre-processing systems (e.g., Sparse Matrix Operator Kernel Emissions, SMOKE). Here we will use TCEQ CAMx air quality simulations from 2019 (at 4 km resolution over Houston and El Paso, mapped to TAMIS monitoring sites) as predictands for extended versions of our GAM models to assess the ability of CAMx to simulate the air quality impacts of fires (Task 2, Section 1.3.2). Using this pre-existing TCEQ generated output will allow us to evaluate the CAMx model while keeping the cost and complexity of the project modest.

3 Methodology

3.1 Task 1: Using GAMs to Quantify Urban Smoke Impacts

In this task, we will train GAMs to determine the impact of smoke on background and urban core MDA8 O_3 and daily average $PM_{2.5}$. Examining background and total urban O_3 and $PM_{2.5}$ separately will allow us to examine the impact of long-range transport of fire smoke separately from the impacts that result when the smoke mixes with urban pollution.

We will assemble a dataset with a variety of meteorological predictors (including but not limited to those in Table 1) and the fire/smoke predictors described in Section 1.2.1. We will perform a correlation analysis of the fire and meteorological predictors to identify if any meteorological predictors are highly correlated ($R^2 > 0.6$) with the fire/smoke predictors. The final set of predictors will be chosen to remove such correlated variables to minimize attributing pollutant variability due to fires with that due to meteorological variability. We will also examine which of the proposed fire predictors is most correlated with background and urban O_3 and $PM_{2.5}$ to determine which predictors will be used in the GAM training.

A variety of GAMs will be trained to determine the optimal set of predictors. All meteorological and fire predictors will be simulated as smooth functions using cubic spline basis set, with periodic splines used to account for the effects of the day of year and hour of day and a random intercept used to account for variability between sites. The models will predict the natural logarithm of O_3 and $PM_{2.5}$ concentrations as these are usually log-normally distributed. We will optimize the GAM by varying hyperparameters (i.e., the smoothing constant and the number of knots allowed for each smooth function). After the training of the initial GAM, variables that are not significant at an $\alpha = 0.1$ level will be removed. The remaining variables will be tested using a leave-one-out process, where the least significant variable is removed, and the resulting GAM is compared with the original via ANOVA. If the two models are not significantly different according to the F-ratio test ($p < 0.05$), then the variable is removed and the next least significant variable is tested. This process is repeated until removing the least significant variable results in a significant degradation of the GAM's ability to predict the pollutant concentrations. Five-fold cross-validation will be used to estimate the prediction error of the GAM for data it was not trained on.

We will examine the GAM residuals and functional forms to determine the quality of the fit and to ensure that the statistical relationships are qualitatively consistent with our knowledge about how meteorology and fires can impact background and urban O_3 and $PM_{2.5}$ (e.g., an increase in smoke transported to the city should increase O_3 and $PM_{2.5}$). We will ensure the residuals are normally distributed and that the residuals are not significantly different between fire and non-fire days. We also expect that adding the fire predictors should not significantly change the functional forms of the meteorological predictors but should increase the skill of the model. Several graphical tests will be used to assess the underlying GAM assumptions of homogeneity, normality, and independence, including plotting the residuals against each explanatory variable that was used in the model and examining the autocorrelation of the original O_3 and/or $PM_{2.5}$ data and residuals as in Gong et al. (2017).

Once the final GAMs are produced, we will apply the fitted smooth functions for each fire predictor to create a derived fire predictor variable. For example, if the GAM fitting found that the square of the fire counts was the best fit function, we would create an additional variable in our dataset that is the square of the fire counts. The final set of observationally derived fire predictor variables will be used in Task 2 (Section 1.3.2) which will allow us to quantitatively determine if the dependence of O_3 and $PM_{2.5}$ on fires in CAMx is consistent with the real-world data.

The functional dependence of O_3 and $PM_{2.5}$ on the fire predictors will allow us to quantify the impact of wildfires on O_3 and $PM_{2.5}$ in HGB and El Paso, thereby addressing two of the TCEQ priority questions: *"What are concentrations of PM and ozone, and their precursors, transported into Texas, from domestic wildfires and wildland-urban fires?"* and *"What role do domestic and international smoke emissions have in exceptional events?"*

3.2 Task 2: Evaluating CAMx Urban Smoke Simulations with GAMs

To assess if CAMx can reproduce the observed relationships between fires and air quality, we will add the MDA8 O_3 and daily average $PM_{2.5}$ values at each monitoring site calculated from the 2019 CAMx output to our dataset. The CAMx output and real-world data will be labeled with a dummy variable (0 for real world data, 1 for CAMx) to allow us to quantitatively test if the differences between the real-world and CAMx GAMs are statistically significant.

We will then fit a modified version of the GAMs from Task 1 to this larger dataset. We will use the derived fire predictors variables discussed in Task 1 so that the effect of fires can be included as linear terms. For example, if the GAM fitting found that the square of the fire counts was the best fit function, we would use the square of the fire counts as a linear predictor rather than fitting a new smooth function of the fire counts. This will allow us to include interaction terms between the coefficients of the derived fire variables and the dummy variable. By using this approach, we will be able to test if the CAMx dependence on fires is significantly different than that in the real data by examining the significance of the coefficients for the interaction terms between the dummy variable and the derived fire variables. This will answer the TCEQ priority question *"Is the atmospheric chemistry of fire plume interaction with urban air accurately captured in photochemical models?"* As in Task 1, this process will be performed for both background and total urban O_3 and $PM_{2.5}$ to separately examine how well CAMx simulates the impacts of long-range smoke transport and the mixing of the smoke with urban pollution.

3.3 Task 3: Identify Potential Improvements to CAMx

If statistically significant differences are found between CAMx and the real-world observations in Task 2, we will perform several sensitivity tests to identify the likely cause of the differences. For example, we will investigate if errors in the WRF meteorology used to drive CAMx are a potential source of the difference by replacing the observed meteorological predictors with their values in the CAMx input met files. If this swap removes the differences between the CAMx and real-world GAMs, then errors in the WRF meteorology are a likely cause of the differences. We will also test replacing the FINN emissions used in the Task 1 and 2 GAMs with the fire emissions used in CAMx. If neither of those changes reduces the difference between the GAMs for CAMx and the real-world observations, it would suggest that the problems are related to how the smoke chemistry is handled in the model. Examining the background and urban GAMs separately will help to determine if the CAMx errors are due to errors in the chemistry taking place during transport of the smoke to the city, due to errors in the chemistry after the smoke mixes with urban emissions, or both. These sensitivity tests will suggest potential avenues for improving the ability of CAMx and other photochemical models to represent the impacts of fires on urban air quality.

4 Work Schedule and Deliverables

The proposed schedule and deliverables for this project are summarized in Table 2. The required monthly and quarterly progress reports are not explicitly listed but will be included in the project Work Plan, along with the draft and final reports. We plan to hold a training session on GAMs for TCEQ staff near the time of the AQRP workshop.

Table 2. Project work schedule and deliverables.

2022	
Q3	**Deliverable 1:** Work Plan and Quality Assurance Project Plan. **Due:** 10 business days after notification of funding.
	Compile meteorological and fire predictors into a single dataset (Task 1).
	Gather and evaluate 2019 CAMx output from TCEQ (Task 2).
Q4	Train and evaluate initial GAMs to assess the impact of smoke on background and urban air quality monitors in Texas (Task 1).
	Determine functional forms of fire predictors to use in Task 2 (Task 1).
2023	
Q1	Apply GAMs to CAMx model output and evaluate differences from Task 1 (Task 2).
	Investigate causes of the differences (Task 2).
	Deliverable 2: Software and input/output files for the GAM models from Task 1, along with a brief technical memo on the models and evaluation. **Due:** March 31, 2023
Q2	Identify conditions under which CAMx mis-estimates fire impacts (Task 3).
	Deliverable 3: Software and input/output files for the GAM models from Task 2, along with a brief technical memo on the models and evaluation. **Due:** June 30, 2023
Q3	Write final report and draft presentation to AQRP workshop.
	Deliverable 4: Presentation at AQRP Workshop **Due:** Approximately 1 month before end of project.
	Deliverable 5: Training for TCEQ staff on GAMs and the software from this study. **Due:** Approximately 1 month before end of project.
	Deliverable 6: Draft Final Report **Due:** August 15, 2023
	Deliverable 7: Final Report **Due:** August 31, 2023

5 Key Personnel

Dr. Matthew Alvarado of AER will be the Principal Investigator for this project and will be responsible for directing this project's day-to-day activities. He will also ensure that project quality standards are met on all deliverables. He is well qualified to carry out this work due to his past research on the chemistry and impacts of biomass burning smoke (e.g., Alvarado et al., 2010, 2015a; Lonsdale et al., 2017a, 2020), as well as his work on using statistical models to predict urban O_3 and $PM_{2.5}$ extremes in Texas (e.g., Pernak et al., 2019; Brown-Steiner et al., 2021).

Dr. Alvarado will be assisted with training and evaluating the statistical modeling in Tasks 1 and 2 by Mr. Rick Pernak, a Senior Staff Scientist at AER with experience in statistical modeling and machine learning (e.g., Pernak et al., 2019). In addition, Dr. Jennifer Hegarty, an AER Staff Scientist with expertise in pollutant transport modeling (e.g., Hegarty et al., 2013) and air quality modeling (e.g., Lonsdale et al., 2017b), will assist in developing the meteorological predictors for the GAMs in Tasks 1 and 2, as well as with the interpretation of the results in Task 3.

6 References

Alvarado, M. J., Logan, J. A., Mao, J., *et al.* (2010), Nitrogen oxides and PAN in plumes from boreal fires during ARCTAS-B and their impact on ozone: an integrated analysis of aircraft and satellite observations, *Atmos. Chem. Phys.,* **10***,* 9739–9760, doi:10.5194/acp-10-9739-2010.

Alvarado, M. J., Lonsdale, C. R., Yokelson, R. J., Akagi, S. K., Coe, H., Craven, J. S., Fischer, E. V., McMeeking, G. R., Seinfeld, J. H., Soni, T., Taylor, J. W., Weise, D. R., and Wold, C. E. (2015a), Investigating the links between ozone and organic aerosol chemistry in a biomass burning plume from a prescribed fire in California chaparral, *Atmos. Chem. Phys.*, **15**, 6667–6688, doi:10.5194/acp-15-6667-2015.

Alvarado, M. J., C. R. Lonsdale, M. E. Mountain, & **J. D. Hegarty** (2015b). *Investigating the Impact of Meteorology on* O_3 *and* $PM_{2.5}$ *Trends, Background Levels, and NAAQS Exceedances*, Final Report to Texas Commission on Environmental Quality (TCEQ) for Work Order No. 582-15-54118-01 under TCEQ Contract No. 582-15-50415, August 31.

Alvarado, M. J., K. Barsanti, S. Chung, D. Jaffe, & T. Moore (2022). “Smoke Chemistry”, in *Wildland fire smoke in the United States: a scientific assessment*, Peterson, D.L.; McCaffrey, S.M.; Patel-Weynand, T., eds. Dordrecht: The Netherlands, ISBN-13: 978-3030870447*.*

Baker, K. R., Woody, M. C., Tonnesen, G. S., Hutzell, W., Pye, H. O. T., Beaver, M. R., Pouliot, G., & Pierce, T. (2016). Contribution of regional-scale fire events to ozone and $PM_{2.5}$ air quality estimated by photochemical modeling approaches. *Atmos. Environ., 140*, 539–554. doi:10.1016/j.atmosenv.2016.06.032

Brey, S. J., Ruminski, M., Atwood, S. A., and Fischer, E. V. (2018), Connecting smoke plumes to sources using Hazard Mapping System (HMS) smoke and fire location data over North America, *Atmos. Chem. Phys*., **18**, 1745–1761, https://doi.org/10.5194/acp-18-1745-2018.

Brown-Steiner, B., Zhou, X., **Alvarado, M. J.,** & Russell, B. T. (2021). Prediction of High-ozone Events Using GAM, SMOTE, and Tail Dependence Approaches in Texas (2005-2019). *Aerosol and Air Quality Research*, *21*, 210077.

Camalier, L., Cox, W., & Dolwick, P. (2007). The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmos. Environ.*, *41*(33), 7127-7137.

Dayalu, A., Alvarado, M. J., & Sun, Q. (2021). *Using Satellite Observations to Quantify Surface $PM_{2.5}$ Impacts from Biomass Burning Smoke*, Final Report to Texas Air Quality Research Program (AQRP) Project 20-005, August 31.

de Foy, B., Saroar, M. G., Salam, A., & Schauer, J. J. (2021). Distinguishing Air Pollution Due to Stagnation, Local Emissions, and Long-Range Transport Using a Generalized Additive Model to Analyze Hourly Monitoring Data. *ACS Earth and Space Chemistry*, *5*(9), 2329-2340.

Draxler, R. R., & Hess, G. D. (1998). An overview of the HYSPLIT_4 modelling system for trajectories. *Australian meteorological magazine*, *47*(4), 295-308.

Gong, X., Kaulfus, A., Nair, U., & Jaffe, D. A. (2017). Quantifying O_3 impacts in urban areas due to wildfires using a generalized additive model. *Environ. Sci. & Technol.*, *51*(22), 13216-13223.

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., ... & Andrews, A. (2013). Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases. *Journal of Applied Meteorology and Climatology*, *52*(12), 2623-2637.

Hu, X., Yu, C., Tian, D., Ruminski, M., Robertson, K., Waller, L. A., and Liu, Y. (2016), Comparison of the Hazard Mapping System (HMS) fire product to ground-based fire records in Georgia, USA, *J. Geophys. Res.-Atmos.,* **121**, 2015JD024448, https://doi.org/10.1002/2015JD024448.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., & Grainger, C. A. (2003). A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model. *Journal of Geophysical Research: Atmospheres*, *108*(D16).

Lonsdale, C. R., Brodowski, C. M., & **Alvarado, M. J.** (2017a). *Improving the Modeling of Wildfire Impacts on Ozone and Particulate Matter for Texas Air Quality Planning*, Final Report to Texas Air Quality Research Program (AQRP) Project 16–024, August 31.

Lonsdale, C. R., **Alvarado, M. J**., Yokelson, R. J., Travis, K. R., and Fischer, E. V. (2014), A sub- grid scale parameterization of biomass burning plume chemistry for global and regional air quality models, presented at the 2014 Community Modeling and Analysis System (CMAS) Conference, Chapel Hill, NC, 27-29 Oct.

Lonsdale, C. R., **Hegarty, J. D.,** Cady-Pereira, K. E., **Alvarado, M. J**., Henze, D. K., Turner, M. D., ... & Scarino, A. J. (2017b). Modeling the diurnal variability of agricultural ammonia in Bakersfield, California, during the CalNex campaign. *Atmospheric Chemistry and Physics*, *17*(4), 2721-2739.

Lonsdale, C. R., C. M. Brodowski, and **M. J. Alvarado** (2017c), *Improving the Modeling of Wildfire Impacts on Ozone and Particulate Matter for Texas Air Quality Planning*, Final Report to Texas Air Quality Research Program (AQRP) Project 16 – 024, August 31.

Lonsdale, C. R., **Alvarado, M. J.**, Hodshire, A. L., Ramnarine, E., & Pierce, J. R. (2020). Simulating the forest fire plume dispersion, chemistry, and aerosol formation using SAM-ASP version 1.0. *Geoscientific Model Development*, *13*(9), 4579-4593.

McDonald-Buller, E., Y. Kimura, C. Wiedinmyer, C. Emery, Z. Liu, and G. Yarwood (2015), *Targeted Improvements in the Fire Inventory form NCAR (FINN) Model for Texas Air Quality Planning,* Final Report to Texas Air Quality Research Program (AQRP) for Project 14-011, December 2015.

McVey, A., R. Pernak, J. Hegarty, & **M. Alvarado** (2018). *El Paso Ozone and $PM_{2.5}$ Background and Totals Trend Analysis*, Final Report to Texas Commission on Environmental Quality (TCEQ) for Work Order No. 582-18-81763-07 under TCEQ Contract No. 582-15-50414, June 29.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., & Freitas, S. (2010). Coupled weather research and forecasting–stochastic time-inverted lagrangian transport (WRF–STILT) model. *Meteorology and Atmospheric Physics*, *107*(1), 51-64.

Pernak, R., **Alvarado, M.**, Lonsdale, C., Mountain, M., **Hegarty, J.,** & Nehrkorn, T. (2019). Forecasting surface O_3 in Texas urban areas using random forest and generalized additive models. *Aerosol and Air Quality Research*, *19*(12), 2815-2826.

Rolph, G. D., Draxler, R. R., Stein, A. F., Taylor, A., Ruminski, M. G., Kondragunta, S., Zeng, J., Huang, H.-C., Manikin, G., McQueen, J. T., and Davidson, P. M. (2009), Description and verification of the NOAA smoke forecasting system: The 2007 fire season, *Weather Forecast.*, **24**, 361–378, https://doi.org/10.1175/2008WAF2222165.1.

Ruminski, M., Kondragunta, S., Draxler, R., and Zeng, J. (2006), Recent changes to the Hazard Mapping System, 15th Int. Emiss. Inventory Conf, (Reinventing Inventories), available at: https://www.researchgate.net/publication/228625934_Recent_changes_to_the_Hazard_Mapping_System (last access: 11 January 2020).

Schroeder, W., Ruminski, M., Csiszar, I., Giglio, L., Giglio, E., Schmidt, C., and Morisette, J. (2008), Validation analyses of an operational fire monitoring product: The Hazard Mapping System, *Int. J. Remote Sens.,* **29**, 6059–6066.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011). The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, *4*(3), 625-641.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.