CRC Report No. A-135

UNDERSTANDING O_3 PRODUCTION AND $PM_{2.5}$ DISTRIBUTIONS DURING SMOKE EVENTS ACROSS THE ENTIRE U.S. USING SURFACE AND SATELLITE DATA FOR 2019-2023

Final Report

February 2025

COORDINATING RESEARCH COUNCIL, INC.
1 CONCOURSE PARKWAY • SUITE 800 • ATLANTA, GA 30328

The Coordinating Research Council, Inc. (CRC) is a non-profit corporation supported by the petroleum and automotive equipment industries with participation from other industries, companies, and governmental bodies on research programs of mutual interest. CRC operates through the committees made up of technical experts from industry and government who voluntarily participate. The five main areas of research within CRC are: air pollution (atmospheric and engineering studies); aviation fuels, lubricants, and equipment performance; heavy-duty vehicle fuels, lubricants, and equipment performance (e.g., diesel trucks); light-duty vehicle fuels, lubricants, and equipment performance (e.g., passenger cars); and sustainable mobility (e.g., decarbonization). CRC's function is to provide the mechanism for joint research conducted by industries that will help in determining the optimum combination of products. CRC's work is limited to research that is mutually beneficial to the industries involved. The final results of the research conducted by, or under the auspices of, CRC are available to the public.

Understanding O_3 production and $PM_{2.5}$ distributions across the U.S. using surface and satellite data for 2019-2023

Prepared by:

Daniel A. Jaffe
University of Washington
School of STEM and Department of Atmospheric Sciences
Bothell, WA 98011
djaffe@uw.edu

Prepared for:

Coordinating Research Council, Inc.
5755 North Point Parkway, Suite 265
Alpharetta, Georgia 30022

Table of Contents

Note on report organization and statistics

Statistical parameters were calculated using a combination of Excel (Microsfoft 365, Version 2410) and R software. Generalized Additive Modeling was completed using R software, version R-4.4.2. For this report, a statistically significant result implies a p value <0.05. For the correlation coefficient (R) we use the Pearson correlation coefficient. This provides a measure of association between two variables. Positive values of R represent a direct correlation, whereas negative values of R are inversely correlated. Note that the R^2 can be interpreted as to the degree to which the x variable explains the variance in the y variable. While there is no universal definition of “strong” or “weak” correlation, I consider correlations with an absolute value of R <0.5 to be “weak.” Finally, it should be noted that even weak correlations can still be statistically significant, if there are a sufficient number of observations.

The report is organized into sections by tasks. Within each section I include the methodology as well as the results for that specific task. In any case where results are separated for the Western, Central and Eastern U.S., I use these definitions for each continental U.S. region: Western US (> 102° W); Central US (82–102° W); Eastern US (< 82° W). Alaska and Hawaii are not included in these regions but are reported separately.

Executive Summary

The primary goals of this project were to develop a method to routinely identify smoke influence on $PM_{2.5}$ and O_3 concentrations at all regulatory air quality sites in the U.S. The analysis includes 802 O_3 monitors with 763,482 individual daily data records for the period April-October 2019-2023. To identify surface smoke, I used the NOAA Hazard Mapping System-Fire and Smoke Product (hereafter simply HMS) combined with surface $PM_{2.5}$. For each site, I define a "$PM_{2.5}$ criteria" that must be met to be considered a smoke influenced day. From this, I identified 17% of all days between April-October 2019-2023 that have some degree of smoke influence. This distribution of smoke days is highly variable in space and time, with sites in the Western U.S. most strongly impacted in 2020 and 2021 and sites in the Central and Eastern U.S. most strongly impacted in 2023.

From this dataset, I used the non-smoke data as "training data" for Generalized Additive Models (GAMs). GAMs were run individually for each site to predict the Maximum Daily Average 8-hour (MDA8) O_3 concentration from the observed meteorological parameters, satellite observations and other predictors. Each model was initiated with the same predictors and GAM equation, but a different number of predictors were statistically significant at each site. The model is very good at predicting the MDA8, with an overall R^2 of 0.84, for all data together, and a mean R^2 of 0.77 from the GAMs at all sites. I examined the model results using 10-fold cross validation to ensure good performance. The residual, defined as the observed MDA8 minus the model predicted MDA8, has a mean value of 0 and a standard deviation of 4.8 ppb for the training data. For each site, the same model is then used to predict the MDA8 for the smoke days. We can interpret the residual for the smoke days as the change in the MDA8 due to the presence of smoke. The overall mean and standard deviation of the residual on the smoke days at all sites is 3.8 ± 8.0 ppb, which indicates the average contribution of smoke to the MDA8. Overall, I found that out of 13,536 O_3 exceedance days, 6200 (45.8%) had smoke influence. Of these 6200 days, 4503 days (72.6%) had residuals that exceeded the 97.5^{th} percentile of non-smoke days. This metric is one that has been recommended by the U.S. EPA (U.S. EPA. 2016). Thus we find that smoke made a significant contribution to O_3 exceedance days in 2019-2023.

Finally, as part of this project, we developed an R-Shiny app that can be used to plot and display these results and run your own GAM/machine learning models. We believe this app will be useful to state and other agencies for both understanding smoke chemistry and in developing exceptional event demonstration packages.

Introduction and project goals

Surface ozone (O_3) is a criteria air pollutant that is formed from reactions of nitrogen oxides ($NO_x = NO+NO_2$) and volatile organic compounds (VOCs) in the presence of sunlight. O_3 has serious health impacts up to and including premature mortality. In the U.S., reductions in the precursor emissions, NO_x and VOCs, over the past several decades have reduced peak O_3 concentrations considerably (Simon et al., 2015), but at present, there are still more than 40 regions in the U.S. that exceed the current 8-hour O_3 standard, so more than 130 million Americans live in areas that do not meet the U.S. National Ambient Air Quality Standards (NAAQS). The current standard is met when the O_3 design value (ODV), defined as the annual fourth highest maximum daily 8-hour average (MDA8) averaged over 3 years, is 0.070 ppm or less. This standard has become stricter several times over the last few decades.

In addition to urban photochemistry, O_3 can also form due to wildfire emissions. Due to the increase in wildfires in California and other parts of the western U.S., many areas of the U.S. have experienced increases in $PM_{2.5}$ (McClure and Jaffe 2018a; Abatzoglou et al 2019; Williams et al 2019; Wilmot et al 2022) and O_3 (McClure and Jaffe 2018b; Jaffe et al 2020; Lee and Jaffe 2024). A number of studies have identified significant health impacts associated with this increasing smoke (e.g. O'Dell et al 2021; Childs et al 2022; Doubleday et al 2023; Burke et al 2023; Heft-Neal et al 2023; Wei et al 2023; Connolly et al 2024; Elser et al 2024).

In addition to the health impacts there are also regulatory challenges. Regions that would otherwise have met the NAAQS, except for the influence of wildfire smoke, may request exclusion of the influenced days under the CAA's Exceptional Event provision (U.S. EPA 2024). But the process is complex and some states have requested greater flexibility in the process (Governors Hobbs, Cox, Polis, & Gordon 2024).

There are two key challenges for both the health studies and regulatory requirements:

1) Lack of agreement on how to identify and define a "smoke influenced day" and
2) The slow timeline of current analyses that attempt to determine where and when there is smoke influence.

Particularly for regulatory consideration, states need rapid tools that can provide estimates of both the $PM_{2.5}$ and O_3 contributions due to smoke. In this work, our primary goal is to develop a set of tools that can routinely assess smoke impacts on $PM_{2.5}$ and O_3 for every air quality site in the U.S. Specific tasks are as follows:

Task 1: Develop Generalized Additive Models for the MDA8 for all US air quality sites measuring O_3 for April-October 2019-2023.

1a. Use the HMS satellite product with surface $PM_{2.5}$ to identify smoke influenced days at each site.

1b. Develop individual Generalized Additive Models for the MDA8 O_3 for each site.

Task 2: Use the GAM results to estimate the daily Smoke O_3 (SMO) contribution at each site.

Task 3: Examine the SMO values as a function of various factors to improve our understanding of what controls O_3 at air quality sites.

Task 4: Examine the hourly O_3 data to understand the SMO contributions for selected sites with heavy smoke influence.

Task 5: Develop an online GAM and visualization tool for use by state and other agencies.

This project builds on my team's past work to identify smoke influence on $PM_{2.5}$ and O_3 using the GAM approach (Gong et al 2017; 2018; McClure and Jaffe 2018b; Jaffe 2020; 2021; Jaffe et al 2020; Flynn et al 2021; Hu et al 2021; Lee and Jaffe 2024).

Data sources

I integrated data from a wide variety of sources. This includes surface pollution data from the U.S. Environmental Protection Agency's Air Quality System (AQS), backward airmass trajectories using the NOAA HYSPLIT model with NAM12 meteorology at 12 km resolution. I used the NOAA Hazard Mapping System Fire and Smoke Product (hereafter simply HMS; Rolph et al., 2009; Ruminski et al., 2011). I also used several products from NASA, including the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) global assimilation model. I also integrated NO_2 tropospheric column density and CH_2O total column density from the Tropospheric Ozone Monitoring Instrument (TropOMI) onboard the Sentinel-5 satellite.

For EPA data, these were mainly downloaded from the EPA's pre-generated data files (https://aqs.epa.gov/aqsweb/airdata/download_files.html) with subsequent updates and fill-ins using the AQS API (https://aqs.epa.gov/aqsweb/documents/data_api.html). It should be noted that data can change even after submission into AQS. Thus it is important to know that data are only current on the date when they were downloaded. For MDA8 values, we use the values calculated in accord with the 2015 O_3 standard, which differs slightly from the calculations for the 2008 O_3 standard. If a site had multiple valid observations (i.e. different POC values), we averaged these for each day. Table 1 shows a summary of all data sources used in this analysis.

Table 1. Data used in this analysis.

Source	Parameter	Unit	Description
EPA	MDA8 O_3	ppb	Daily maximum 8-hour average ozone concentrations
	$PM_{2.5}$	μg m^{-3}	Daily mean $PM_{2.5}$ concentrations
Calc.	DOY	-	Day of Year
TROPOMI (NASA)	NO2VCD	molec. cm^{-2}	Daily NO_2 tropospheric vertical column density (VCD) for a movin g average based on the center of a 15-day window
	HCHOVCD	molec. cm^{-2}	Daily HCHO tropospheric vertical column density (VCD) for a moving average based on the center of a 15-day window
MERRA-2 (NASA)	SRAD	W m^{-2}	Daily mean surface incoming shortwave flux
	T2MAX	K	Daily maximum surface temperature
	QV2M	kg kg^{-1}	Daily mean relative humidity
	U10M	m s^{-1}	Daily mean 10-meter eastward wind
	V10M	m s^{-1}	Daily mean 10-meter northward wind
	U500	m s^{-1}	Daily eastward wind at 500 hPa
	V500	m s^{-1}	Daily northward wind at 500 hPa
*HYSPLIT w/NAM12 (NOAA)	TrajH2O	g kg^{-1}	Daily water vapor mix ratio averaged over 12 hours along the back trajectory
	TrajDepth	M	Daily water mixing depth averaged over 12 hours along the back trajectory
	TrajDist	km	Daily point-to-point distance for a 12-hour backward trajectory
	TrajDir	deg	Daily point-to-point direction for a 12-hour backward trajectory
NOAA-HMS-FSP	HMS	Heavy/Med. /Light	Daily smoke polygons over each monitoring station
*Daily back-trajectories were initialized at 1 pm local standard time at 500 m above ground level for each site.			

.

Task 1: Develop Individual Generalized Additive Models for the MDA8 for all US air quality sites measuring O_3 April-October, 2019-2023.

a. Use the HMS satellite product with surface $PM_{2.5}$ to identify smoke influenced days at each site.

The NOAA HMS product is based on multiple satellite indicators to identify smoke influenced. But as it is based only on satellite data, it cannot indicate whether smoke is at the surface or not. For each site, we develop a surface “$PM_{2.5}$ criteria” to define smoke influenced days. We first use the HMS data to identify a set of days that are possibly smoke influenced. Any day with an overhead HMS smoke identification is termed HMS=1. For the $PM_{2.5}$ smoke criteria, previously we used the mean + 1 standard deviation of the non-HMS days (HMS=0) (Jaffe 2020; 2021). However, at some sites I found a strong log-normal distribution of $PM_{2.5}$ data, which means that the $PM_{2.5}$ criteria would be very high and we would miss many smoke influenced days. I have chosen to switch to a method based on the median plus the median absolute deviation (MAD) of the HMS=0 days to define the $PM_{2.5}$ smoke criteria. This method is preferred for non-normal distributions (Leys et al 2021). Because of the seasonality in the $PM_{2.5}$ concentrations, I chose to calculate the median and MAD for each month independently. So for our 5-year analysis, our $PM_{2.5}$ criteria is based on all days within each month with HMS=0. It’s important to use multiple years of data since in some months there was near continuous overhead smoke. For example, in June 2023, most sites in the Central and Eastern U.S. had overhead (HMS=1) smoke on more than 90% of days. Combing the HMS identification plus the surface $PM_{2.5}$ criteria, I identified 17% of all days in April-October 2019-2023 as having surface smoke influence, averaged across all sites.

Over the past five years there has been significant variability in surface smoke influence. Figure 1 shows a plot of the fractional smoke influence for the April-October time period by state for each year.

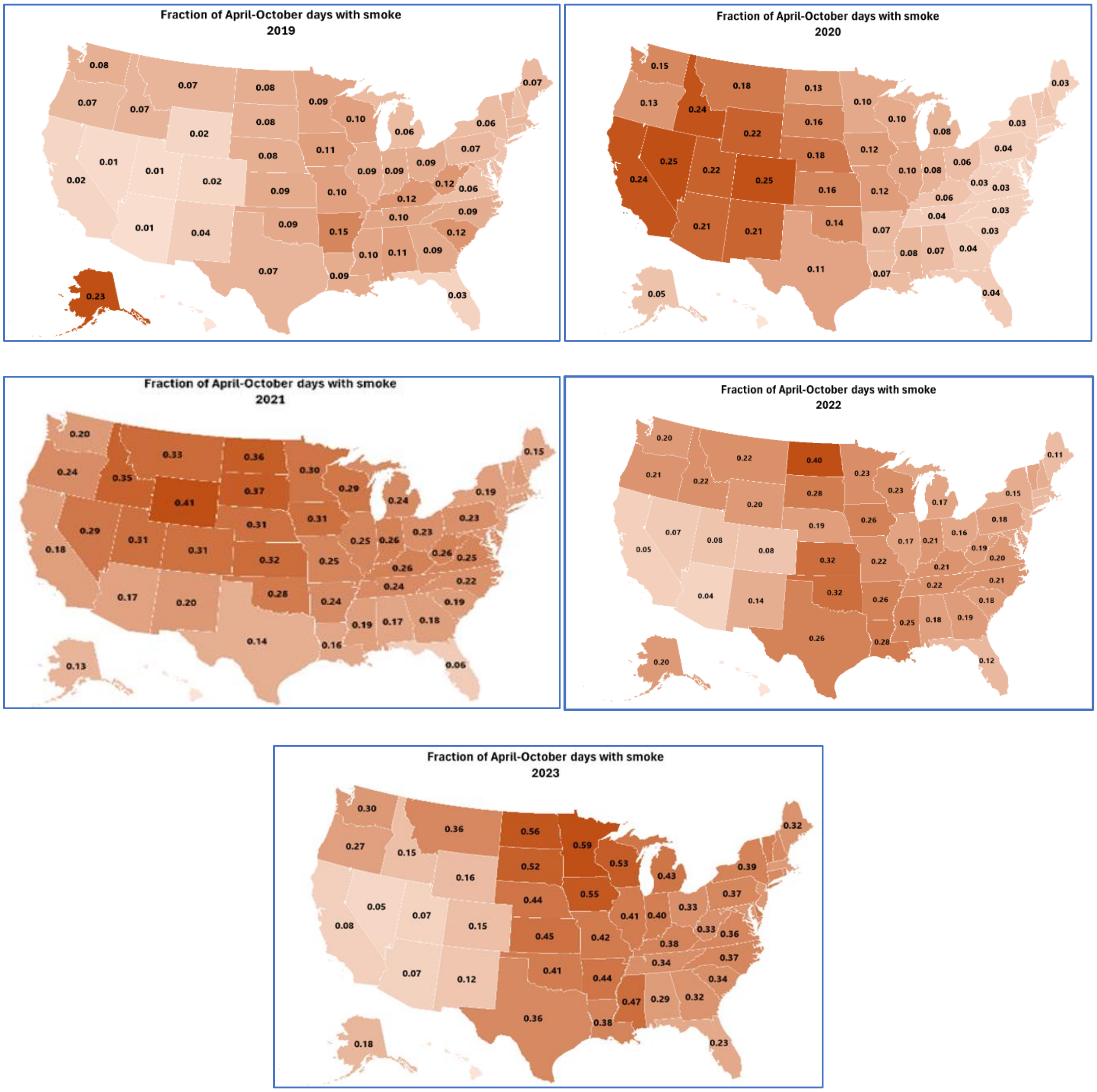

Figure 1. Fraction of smoke days (HMS+enhanced $PM_{2.5}$) for April-October 2019-2023.

b. Development of individual Generalized Additive Models for the MDA8 O_3 for each site.

GAMs are a type of model where the form of the relationship between the predictor and predictands are not predetermined (Wood 2017). Instead this relationship is determined by a training dataset. The general form of the GAM relationship is:

$$\boldsymbol{MDA8\ O_3} = \boldsymbol{f(predictor1) + f(predictor2) + \cdots + f(predictor\ x\ by\ predictor\ y) + \beta} \qquad \textbf{(Eq.1)}$$

Where predictor1, 2, etc are the individual MDA8 predictor variables and β is a constant. The form for each predictor is determined from the pattern of the training dataset. In some cases, we find that variables "interact", meaning that a variable will have a different relationship to the main predictand depending on the value of the second variable. Consistent with our previous work, we used a wide array of meteorological and other variables as predictors for the MDA8. In total, I considered 18 different variables plus interaction terms between variables. But a predictor was only retained in the model if it was deemed to be statistically significant. Table 1 shows a list of all predictor variables considered, along with the interaction terms.

I used the mcgv package in R software for these calculations. In the mcgv package, we can choose the smoothing parameter for each predictor variable. The "s" function is a spline fit. The smoother function "te" is used for interaction terms and includes the main (one dimensional) effect. The smoother function "ti" is used for an interaction term alone. Table 1 below shows a list of all predictors and interactions terms considered.

Table 2. List of main predictors and interactions used in this work.

	Main effects	Seasonality interactions	
$\boldsymbol{MDA8\ O_3} =$	$\boldsymbol{s(DOY) +}$ $\boldsymbol{s(SRAD) +}$ $\boldsymbol{s(T2MAX) +}$ $\boldsymbol{s(QV2M) +}$ $\boldsymbol{s(TrajH2O) +}$ $\boldsymbol{s(TrajDepth) +}$ $\boldsymbol{s(TrajDist) +}$ $\boldsymbol{s(TrajDir) +}$ $\boldsymbol{s(NO2VCD) +}$ $\boldsymbol{s(HCHOVCD) +}$	$\boldsymbol{ti(DOY, SRAD) +}$ $\boldsymbol{ti(DOY, T2MAX) +}$ $\boldsymbol{ti(DOY, QV2M) +}$ $\boldsymbol{ti(DOY, TrajH2O) +}$ $\boldsymbol{ti(DOY, TrajDepth) +}$ $\boldsymbol{ti(DOY, TrajDist) +}$ $\boldsymbol{ti(DOY, TrajDir) +}$ $\boldsymbol{ti(DOY, NO2VCD) +}$ $\boldsymbol{ti(DOY, HCHOVCD) +}$	$\boldsymbol{ti(TrajDist, TrajDir) +}$ $\boldsymbol{te(U10M, V10M) +}$ $\boldsymbol{te(U500, V500) +}$
			$\boldsymbol{ti(T2MAX, SRAD) +}$ $\boldsymbol{ti(T2MAX, QV2M) +}$ $\boldsymbol{ti(T2MAX, TrajDepth) +}$ $\boldsymbol{ti(QV2M, SRAD) +}$ $\boldsymbol{ti(QV2M, TrajH2O) +}$
			$\boldsymbol{ti(T2MAX, NO2VCD) +}$ $\boldsymbol{ti(T2MAX, HCHOVCD) +}$ $\boldsymbol{ti(NO2VCD, HCHOVCD)}$

The model is trained and optimized at each air quality site independently using the non-smoke day dataset and the predictors for that site (met, satellite observations, etc). For each individual site/GAM, I use a stepwise backward process to remove non-significant predictors based on their F-statistic, which is an indicator of variable importance. This process continues until all remaining predictors are statistically significant. This results in a different number of optimized predictors for the GAM at each individual site. From this process we can calculate a number of key metrics including the overall R^2 and standard deviation of the residual (observed MDA8 – predicted MDA8). The mean residual must be zero, or else this indicates a problem with the calculation.

Once the best model is identified, I used 10-fold cross validation (CV) as a quality control check on each GAM. In this step, the model is recomputed with 90% of the training dataset and then evaluated with the remaining 10% of the dataset that was not used to train the model. This is repeated 10 times so that all of the training dataset is eventually used as CV data. The mean performance (R^2) of the 10 CV runs is then a measure of the models ability to predict the MDA8 for data that was never used to train the model. The difference between R^2 values for the full training dataset and the CV results can indicate if the model is "over-fitting"; i.e. has too many predictor variables.

Once the optimum model for each site is developed, the same model is applied to the smoke dataset. In this way, we can compute the expected MDA8 given the meteorological conditions, seasonality, etc. for each smoke day. Calculating the same metrics as above (R^2, mean and standard deviation of the residual), provide different information than with the training dataset. In general, we find that the R^2 using the smoke dataset is lower and the standard deviations are higher, compared to the training dataset. This reflects the greater variability that smoke introduces. For nearly all sites, the mean residual for the smoke days is positive due to additional O_3 from wildfire emissions. Figure 2 shows a schematic diagram of the overall process used to develop the GAMs.

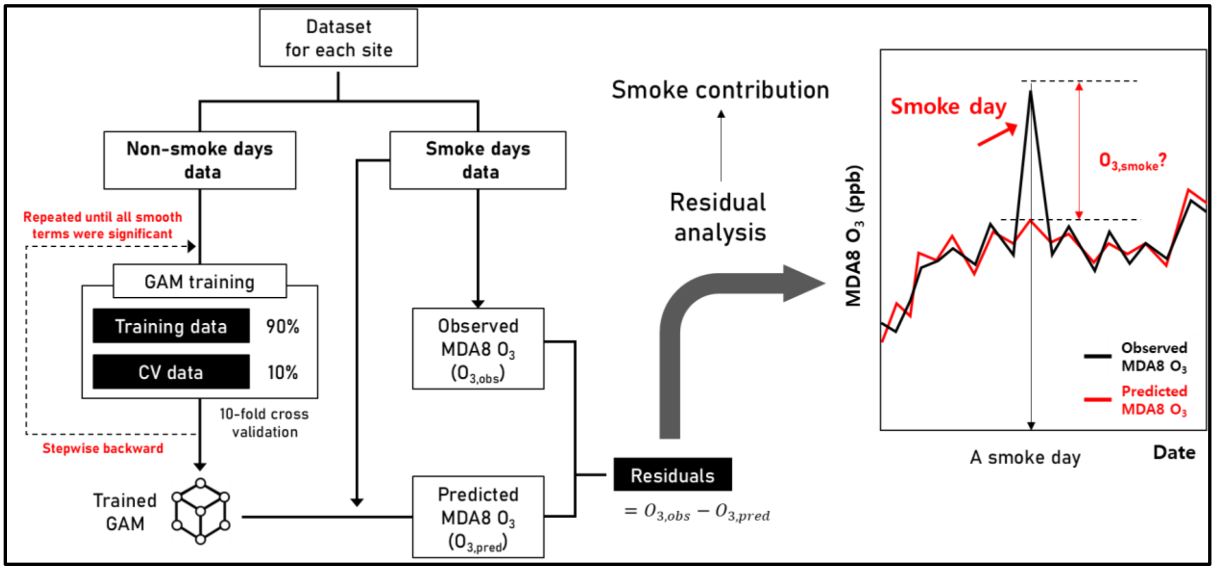

Figure 2. Overall process used to create the GAMs at each site.

c. GAM results

A summary of the GAM results is shown in Table 3. Figure 3 shows a comparison of all GAM results for the training dataset. The overall R^2 is 0.84 using the training dataset and 0.74 using the CV results. The GAMs at individual sites show a mean R^2 value of 0.77 and a mean of 0.62 using the CV results.

Table 3. Summary of GAM results.

	Training	CV	Smoke
N	633760	633760	129722
R^2 (alldat)	0.84	0.74	0.59
R^2 (mean all sites)	0.77 ± 0.05	0.62 ± 0.07	0.45 ± 0.14
Residual Mean+SD	0.00 ± 4.78	-0.01 ± 6.16	3.80 ± 7.99
Residual RMSE	4.78	6.16	8.85

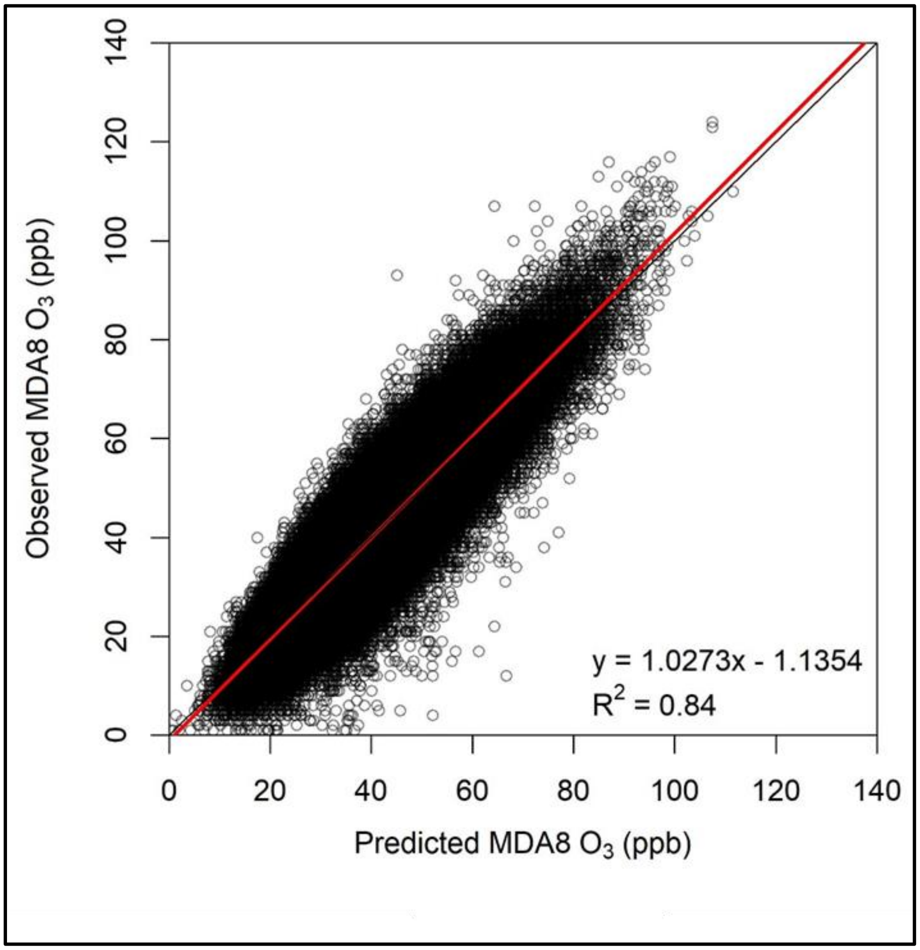

Figure 3. Observed vs predicted MDA8 for all training data.

Task 2: Estimates of the daily Smoke O_3 (SMO) contribution at each site.

In using the GAM results to estimate the smoke contribution to the MDA8 it is essential to consider the distribution of the residuals. Figure 4 shows the GAM residuals for both training and smoke day datasets. As noted previously, the mean residual for training data is 0, whereas the mean residual for the smoke day data is 3.8 ppb. While the mean residual from smoke provides information on the overall mean smoke contribution to the MDA8, for individual days, we must consider variations in both distributions. Based on published guidance from the U.S. EPA (U.S. EPA. 2016), I apply a statistical metric on the residual to determine if it sufficiently different from the training dataset. In this case, the residual must be greater than the 97.5th percentile of the non-smoke residuals. If it is, then we can consider that MDA8 to have been enhanced by smoke. The 97.5th percentile of the non-smoke residuals is calculated individually at each site. The overall mean 97.5th percentile at all sites is 9.6 ppb, with a range of 5.2-16.0

ppb. In general, sites that have a better model performance (higher R^2) have a lower 97.5th percentile, although the relationship is not strong. So in summary, when the GAM residual exceeds the 97.5th percentile, this is a strong indicator that smoke significantly enhanced the MDA8.

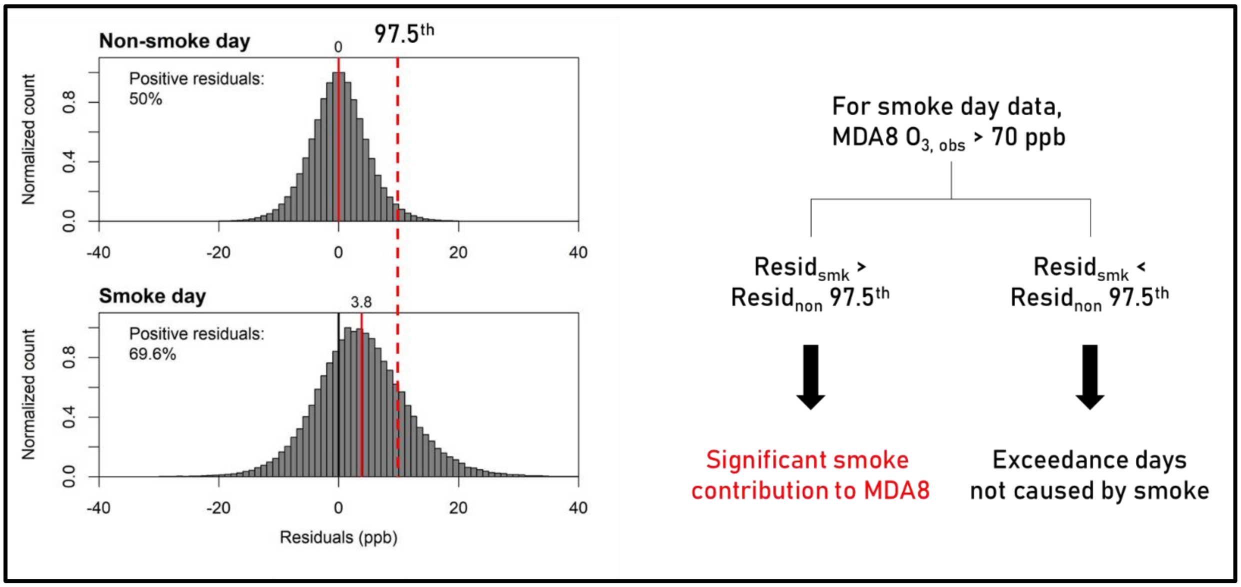

Figure 4. Distribution of residuals from the training and smoke day datasets.

Using this metric, I find that 2.5% of the non-smoke residuals are greater than the 97.5th percentile (by definition), but 21.3% of the smoke residuals are greater than this metric. This demonstrates a signficant enhancement in the MDA8 due to smoke on more than 27,000 days over this 5-year time period. Of the O_3 exceedance days that we determined to be smoke influenced, 73% had residuals that exceeded the 97.5th percentile. I define the Smoke O_3 (SMO) contribution to be equal to the residual on days with smoke.

Figure 5 shows examples of especially strong smoke influence on the MDA8 for 8/21/2020 and 6/2/2023 and Figure 6 shows the SMO for these same dates. On 8/21/20, large fires in California led to significant $PM_{2.5}$ and O_3 enhancements in a large region of the western U.S. from California to Colorado. On this day, 83 monitors exceeded an 8-hour average of 70 ppb. Of these 83 monitors, 77 had smoke and 67 had residuals that exceeded the 97.5th percentile. The average smoke contribution to the MDA8 on this day was 25.9 ppb. On 6/2/23, large fires in Canada led to significant $PM_{2.5}$ and O_3 enhancements in a large region of the Central and Eastern U.S. On this day, there were 115 monitors that exceeded an 8-hour average of 70 ppb, 107 of which had smoke. Of these 107 monitors, 65 had residuals that exceeded the 97.5th percentile with an average smoke contribution to the MDA8 of 18.5 ppb. While smoke from the 2021 California fires resulted in greater $PM_{2.5}$ and O_3 enhancements (at least in the U.S.) compared to the Canadian smoke events in 2023, the 2023 smoke events impacted a larger fraction of the U.S. population.

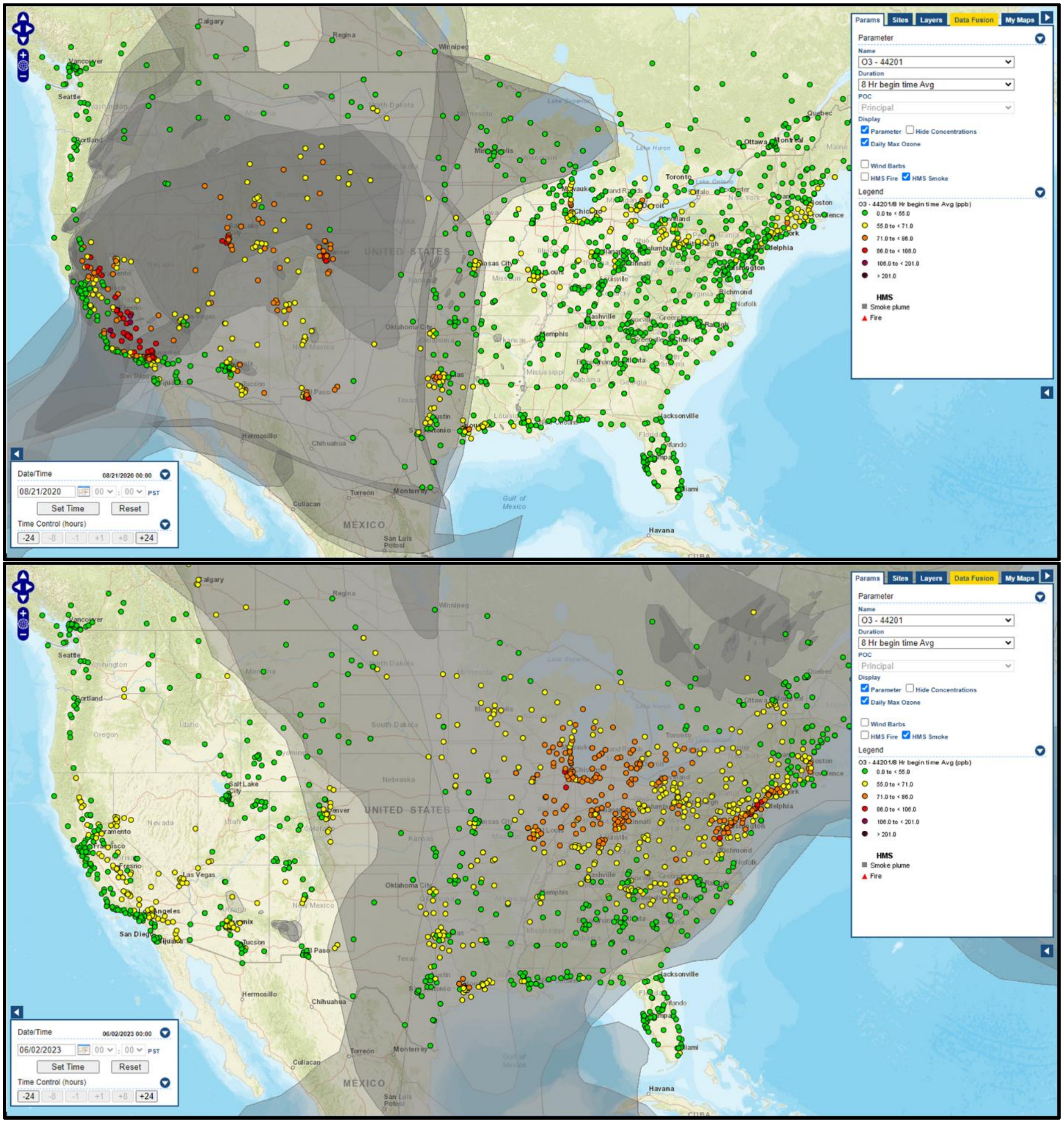

Figure 5. Examples of strong smoke influence on the MDA8 for 8/21/2020 (top) and 6/2/2023 (bottom). Maps generated using AirNowTech.org

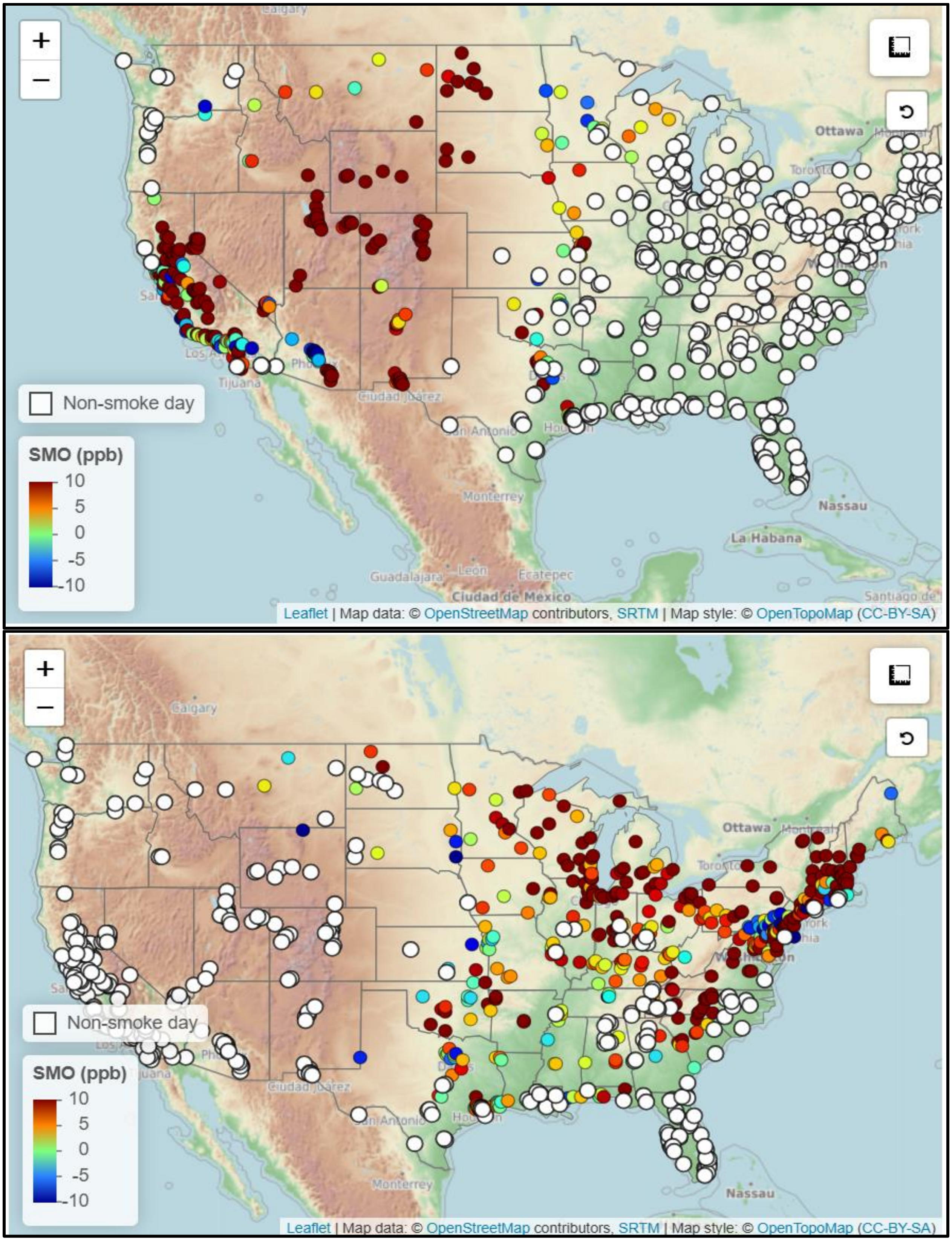

Figure 6. Calculated SMO contributions to the MDA8 for 8/21/2020 (top) and 6/2/2023 (bottom). Maps generated using authors' R-Shiny app (more details presented under Task 5).

Table 4 shows the results by state. This includes number of exceedance days, number of exceedance days with smoke and number of exceedance days with smoke with residuals greater than the 97.5th percentile summed for all 5 years (2019-2023). In the appendix (Tables A1-A5), I show the same information by year. Overall, I find that 33% of exceedance days in the U.S. likely have a significant smoke contribution (smoke days with residuals greater than 97.5th percentile). States in the Western U.S., such as California, have their highest number of exceedance days and exceedance days with smoke in 2020 and 2021. In the East and Central U.S. the highest numbers are in 2023. Dates that have smoke and residuals greater than the 97.5th percentile are, from a scientific and statistical perspective, demonstrate the strongest cases for a smoke influence on the MDA8. That said, the U.S. EPA uses a "weight of evidence" approach to designate exception events and no one piece of evidence can be considered conclusive.

Table 4. Total of exceedance days, exceedance days with smoke and exceedance days with smoke that exceed the 97.5th percentile by state for 2019-2023. Tables with the same information for each year are shown in the appendix.

	Sum of all exceedance days	Sum of all exceedance days with smoke	Sum of all exceedance days with smoke that exceed 97.5th percentile	Fraction of all exceedance days that have smoke and exceed the 97.5th percentile
Alabama	39	29	23	0.59
Alaska	0	0	0	NA
Arizona	720	195	119	0.17
Arkansas	21	12	9	0.43
California	6837	2123	1369	0.20
Colorado	494	314	230	0.47
Connecticut	260	153	120	0.46
Delaware	26	17	12	0.46
D.C.	26	18	14	0.54
Florida	27	14	11	0.41
Georgia	80	60	51	0.64
Hawaii	0	0	0	NA
Idaho	29	27	24	0.83
Illinois	471	326	201	0.43
Indiana	201	157	123	0.61
Iowa	65	62	53	0.82
Kansas	49	37	24	0.49

Kentucky	89	64	53	0.60
Louisiana	76	54	45	0.59
Maine	13	7	7	0.54
Maryland	133	78	60	0.45
Massachusetts	44	33	31	0.70
Michigan	170	111	80	0.47
Minnesota	76	66	64	0.84
Mississippi	22	18	18	0.82
Missouri	123	88	66	0.54
Montana	9	8	8	0.89
Nebraska	15	14	9	0.60
Nevada	361	194	180	0.50
New Hampshire	9	8	8	0.89
New Jersey	97	64	50	0.52
New Mexico	171	67	57	0.33
New York	128	86	77	0.60
North Carolina	37	28	22	0.59
North Dakota	41	41	41	1.00
Ohio	184	113	83	0.45
Oklahoma	136	102	85	0.63
Oregon	24	21	19	0.79
Pennsylvania	174	120	93	0.53
Rhode Island	36	22	22	0.61
South Carolina	13	8	8	0.62
South Dakota	58	44	36	0.62
Tennessee	57	46	38	0.67
Texas	1005	510	318	0.32
Utah	466	289	258	0.55
Vermont	2	2	2	1.00
Virginia	14	13	13	0.93
Washington	12	9	9	0.75
West Virginia	0	0	0	NA
Wisconsin	364	296	229	0.63
Wyoming	32	32	31	0.97
Grand Total	13536	6200	4503	0.33

Figure 7 shows the impacts of smoke on exceedance days by year and region. Days in red were identified as smoke influenced and have residuals greater than the 97.5th percentile. In the Western U.S. the largest impacts from smoke were seen in 2020 and 2021, whereas in the Eastern and Central U.S. the largest impacts were seen in 2023.

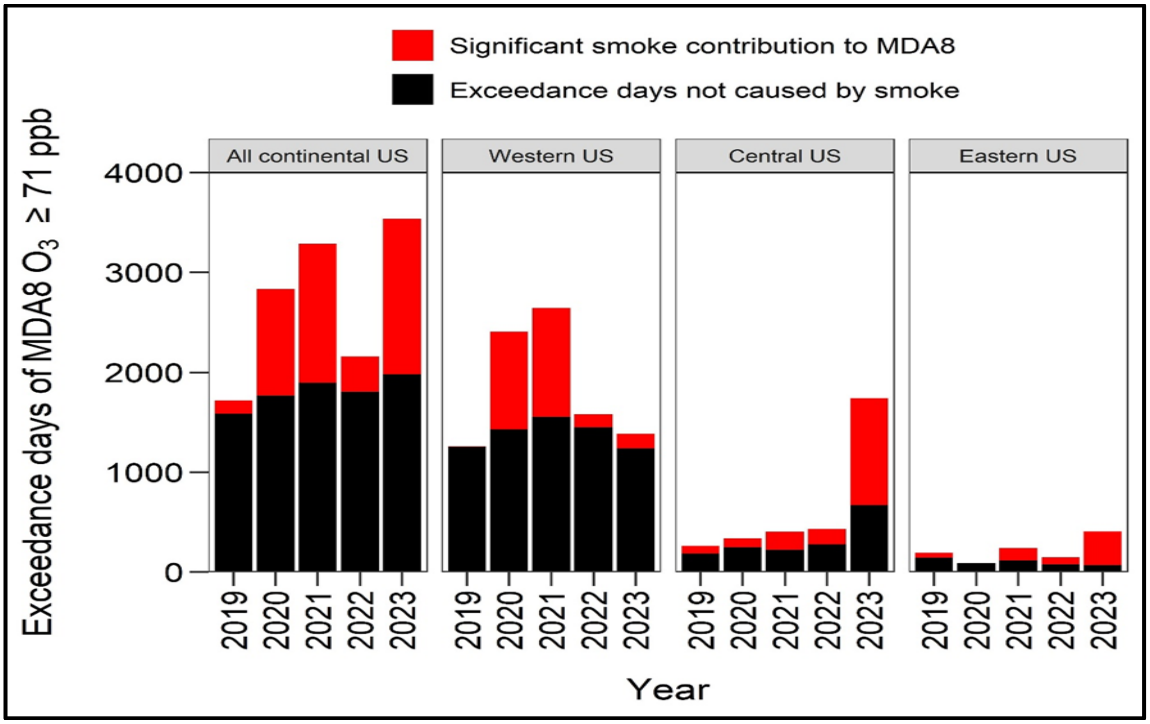

Figure 7. Number of exceedance days by year and region. The bars in red count days that were identified as smoke influenced and have residuals greater than the 97.5th percentile.

Task 3: Examine the SMO values as a function of various factors to improve our understanding of what controls O_3 at air quality sites.

As noted in the sections above, the GAM residuals for smoke days show a significant positive mean bias compared to the non-smoke days. While the difference between smoke and non-smokes is statistically significant, there remains a large degree of variability. So here, I examine how the smoke day residuals vary as a function of surface $PM_{2.5}$, satellite NO_2 and HMS smoke patterns. Figure 8 shows the GAM residual vs daily mean $PM_{2.5}$. This plot includes all smoke days, so despite a very low R^2, this is a statistically significant relationship. Nonetheless, it is clear that $PM_{2.5}$ can explain only a very small fraction of the total variance in the GAM residuals (ca 3%). There are a number of complicating factors with this approach. First, the amount of $PM_{2.5}$ does not tell us anything about the quantify of O_3 precursors (especially NOx and VOCs) which can vary considerably from fire to fire and depending not only on the amount of biomass burned, but also combustion conditions and fuel. Second, $PM_{2.5}$ may act to reduce photolysis rates, and consequently O_3 production.

Figure 9 looks at how NOx may impact smoke O_3. It shows a plot of the satellite NO_2 vertical column density (log scale). We might expect this relationship to show a significant correlation if local NOx were a driver of smoke induced O_3 production (see discussion in Jaffe et al 2020). In this case we see an even worse relationship. Again, there are a number of complicating factors for this analysis. First, the satellite data is from the OMI satellite and is fairly low resolution (1^o x 1^o), with frequent missing data. To be useful in this analysis, I had to do 15 day averaging. A second complication is that while NOx is needed for O_3 production, at higher NOx levels, we know that it will suppress O_3 formation.

Finally, I look at whether the HMS satellite product gives any useful information on the smoke O_3 production. Table 1 shows $PM_{2.5}$ and the GAM residuals sorted by HMS classification (light, medium or heavy smoke). There is a statistically significant increase in the $PM_{2.5}$ concentrations and the GAM residuals with heavier smoke. But the variability remains large.

I would not say this part of the analysis has led to any new insights. The variability is too large and none of the variables that I have used provide strong correlations. The only conclusion is that $PM_{2.5}$ and the HMS smoke category do tend to predict higher O_3 GAM residuals, but there remains a large amount of variability. Future work on this analysis could include:

i. Use of higher resolution satellite retrievals (such as the recently launched TEMPO instrument).

ii. More information on the smoke source (e.g. fire characteristics) and transport (for example time of transport, meteorology during transport, etc).

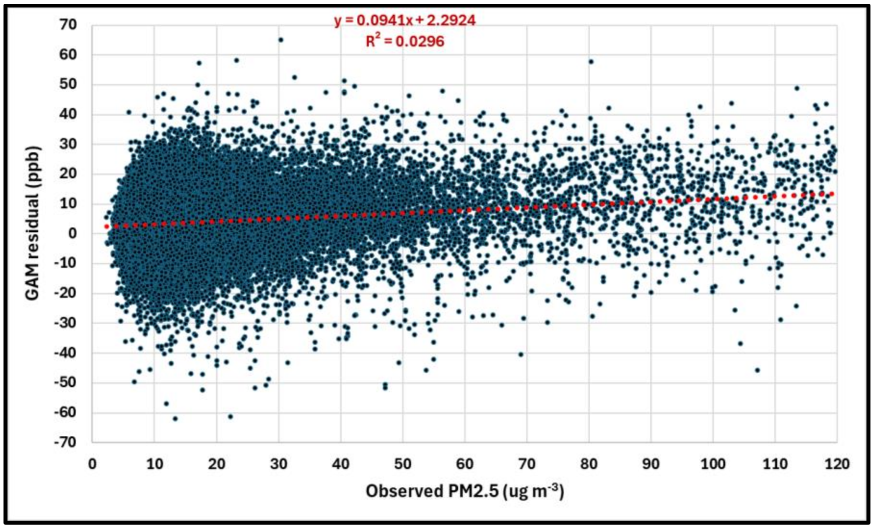

Figure 8. Plot of GAM residuals vs observed daily mean $PM_{2.5}$. This plot contains 120,788 smoke data points.

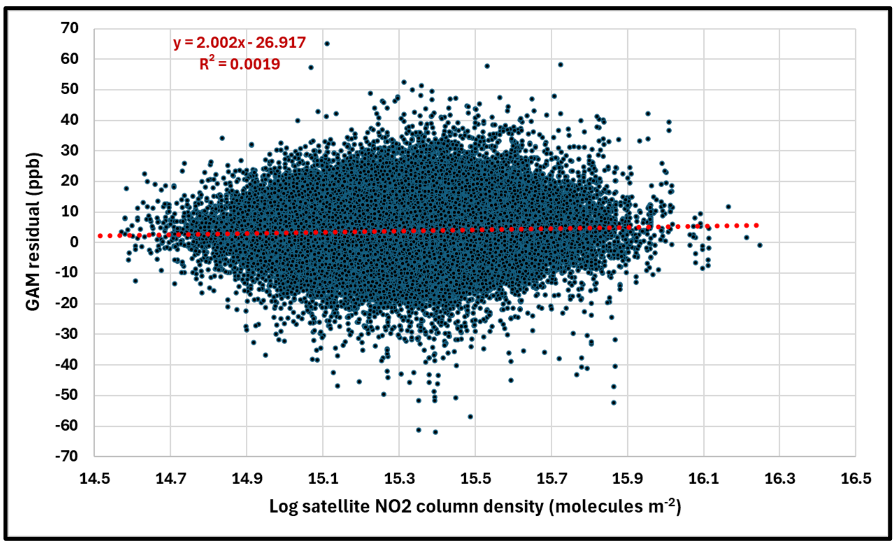

Figure 9. Plot of GAM residuals vs the log of OMI satellite NO_2 column density. This plot contains 120,788 smoke data points.

Table 5. Statistics on PM2.5 and GAM O3 residuals by HMS smoke classification. This table includes 129,722 smoke day data points.

HMS Smoke category:	Light	Medium	Heavy
Count	79047	34979	15696
Average PM2.5 (ug m^{-3})	12.21	15.65	32.05
S.D. of PM2.5 (ug m^{-3})	5.40	10.66	33.95
Average of GAM residuals (ppb)	3.08	4.53	5.78
S.D. of GAM residuals (ppb)	7.38	8.00	10.14

Task 4: Examine the hourly O_3 data to understand the SMO contributions for selected sites with heavy smoke influence.

We can gain insight into whether smoke O_3 at each air quality site is locally produced or produced in the smoke plume prior to arrival at the monitoring location by looking at the hourly O_3 data. In the first instance, we would expect to see an enhanced rate of O_3 production (dO_3/dt) on days with smoke and in the second case we would expect to see more uniform O_3 enhancements across the day. Both types of O_3 enhancements are possible and both have been demonstrated in the literature (Buysse et al 2010, Rickly 2023). To start the analysis, I want to focus on states and years with high fire influence on O_3 exceedance days. Tables A1-A5 shows the number of all exceedances days by state for the years 2019-2023 and also the number of exceedance days with smoke. From this, I find that California in 2020 had the highest number of smoke influenced exceedance days for all states over this time period (1012). Of these, 312 smoke influenced exceedance days occurred in August, 392 in September and 296 in October. In 2023, sites in the Midwest had a large number of exceedance days with smoke, including Texas (291) and Illinois (251). For Texas, the highest month was September and for Illinois, it was June. So I will examine the hourly data for these states and times.

For this task, I downloaded hourly data files for each calendar year from the EPA's Air Data website. These files are very large, containing more than 9 million rows of data. To make the files more manageable, I split these by state and reconfigure the data format to a more useful form. Finally, I calculate the MDA8 from the hourly data and ensure my calculated MDA8 matches the regulatory value. Next, I add a smoke or no smoke day classification for each site for each day.

Figures 10 shows the mean diurnal cycle of O_3 across all California sites for August, September and October 2020, respectively for smoke and non-smoke days. This period had extensive fires in California and a huge number of $PM_{2.5}$ and O_3 exceedance days. A similar procedure was used to evaluate impacts from the large Canadian fires in 2023 on MDA8 O_3 in Illinois and Texas. Figures 11 shows the mean diurnal cycle of O_3 across the Illinois and Texas sites in June and September 2023, respectively for smoke and non-smoke days. The diurnal O_3 profiles (Figures 10-11) strongly suggest that most of the enhanced O_3 on these smoke days was due to enhanced local photochemical production. This conclusion comes from the fact that:

i. morning O_3 was nearly identical on the smoke and non-smoke days;

ii. the rate of increase (dO_3/dt) was much higher on smoke influenced days.

Table 6 and 7 quantifies these results for each month by calculating mean statistics for a number of key parameters by smoke and non-smoke day. Table 6 and 7 also gives some additional information on the number of exceedance days, which demonstrates the extent of smoke impacts in 2020 (in California) and 2023 (Illinois and Texas).

To estimate the contribution to the higher afternoon O_3 due to baseline enhancement, I use:

$$\%\ \boldsymbol{contribution\ from\ baseline} = \frac{\boldsymbol{mean\ morning\ difference}}{\boldsymbol{mean\ afternoon\ difference}}\ \boldsymbol{x\ 100}$$

Note that this is usually positive, but if the average baseline O_3 is lower on smoke days, then this can yield a negative value. To estimate the contribution from photochemistry, we can do this two ways, first using:

$$\%\ \boldsymbol{contribution\ from\ photochemistry} = \frac{\boldsymbol{mean\ slope\ difference\ (per\ hr)x\ 8\ hrs}}{\boldsymbol{mean\ afternoon\ difference}}\ \boldsymbol{x\ 100}$$

This can also be calculated by simple subtraction:

$$\%\ \boldsymbol{contribution\ from\ photochemistry} = \mathbf{100} - \%\ \boldsymbol{contribution\ from\ baseline}$$

In all of these periods across these three states, the % contribution from photochemistry far outweighs the contributions from baseline enhancements. So these calculations support the idea that **local** photochemical O_3 production from enhanced smoke precursors was mainly responsible for the enhanced MDA8 values seen for these months. This is supported by the lack of significant enhancement in morning O_3 for each month. It is likely that in each case, the wildfire emissions contributed, mainly, reactive VOCs, while the local anthropogenic emissions were the source or NOx. Once these precursors mixed and in the presence of solar radiation, O_3 production was strongly enhanced on a smoke day, compared to a non-smoke day.

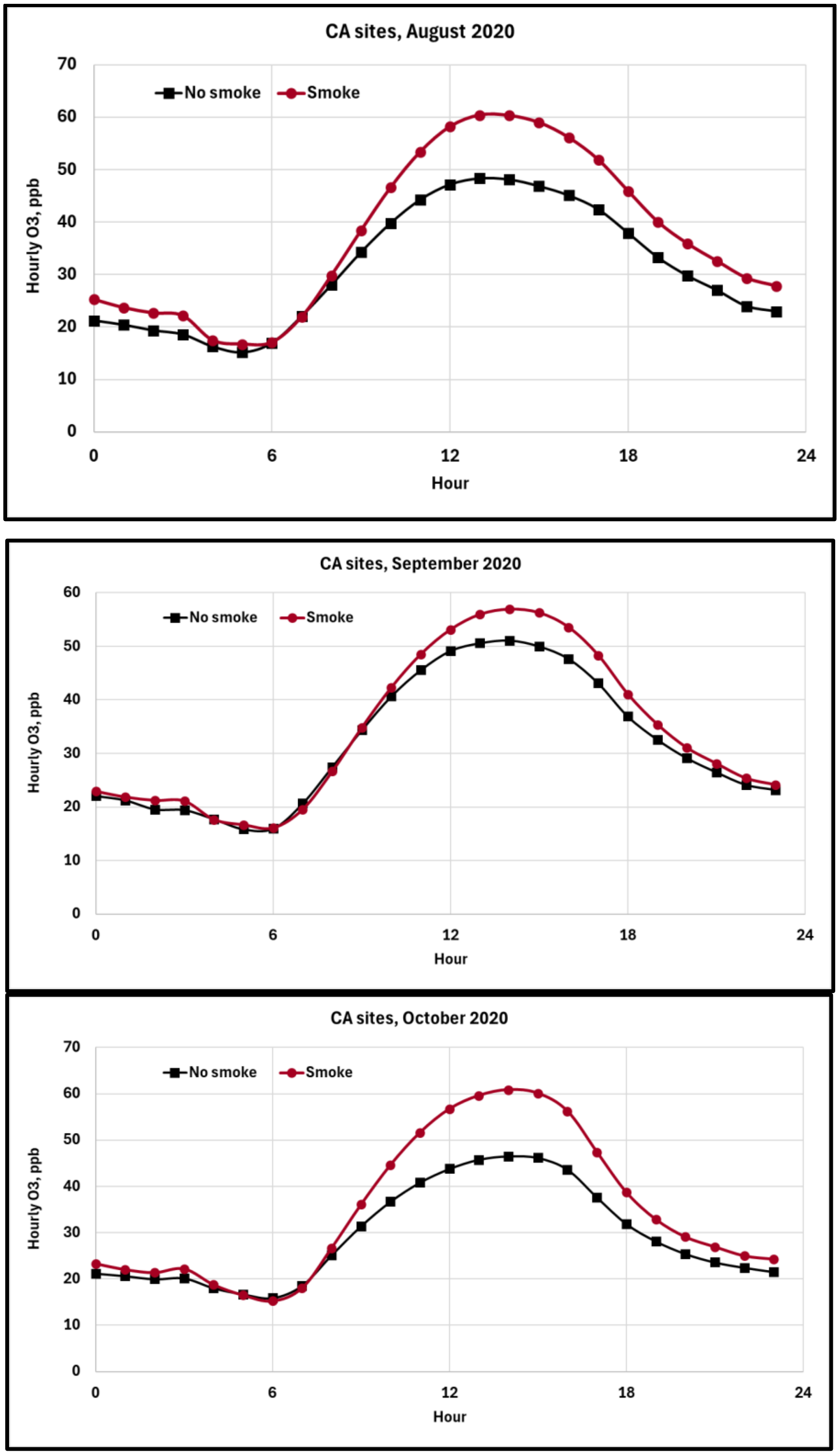

Figure 10. Hourly O_3 for all California sites in August (top), September (middle) and October 2020 with and without smoke.

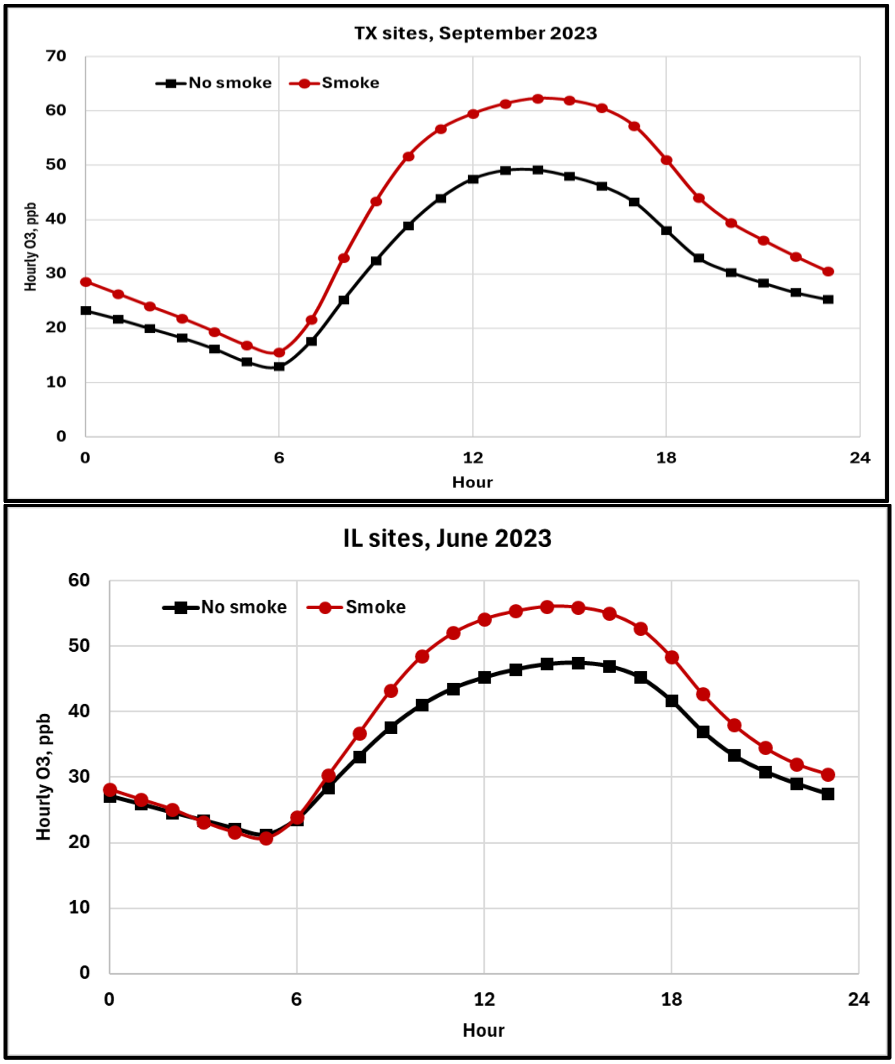

Figure 11. Hourly O3 for all Texas sites in September 2023 (top) and Illinois sites in June 2023 (bottom) with and without smoke.

Table 6. Estimated contribution to enhanced afternoon (2pm) O3 from local photochemistry vs enhanced baseline for California sites in August-October, 2020.

California August 2020		
Mean morning difference (6 am smoke day - non smoke day)	0.1	ppb
Mean afternoon difference (2 pm smoke day - non smoke day)	12.2	ppb
Mean slope difference (6 am to 2pm slope smoke day - non smoke day)	1.5	ppb/hr
Contribution from morning (baseline)	0.8	%
Contribution from enhanced photochemistry (from slope)	99.4	%
Contribution from enhanced photochemistry (by subtraction)	99.2	%
Percent of all days that are smoke days	39.2	%
Percent exceedance days	12.9	%
Percent exceedance days that are smoke days	69.8	%
California September 2020		
Mean morning difference (6 am smoke day - non smoke day)	0.1	ppb
Mean afternoon difference (2 pm smoke day - non smoke day)	5.8	ppb
Mean slope difference (6 am to 2pm slope smoke day - non smoke day)	0.7	ppb/hr
Contribution from morning (baseline)	2.2	%
Contribution from enhanced photochemistry (from slope)	91.6	%
Contribution from enhanced photochemistry (by subtraction)	97.8	%
Percent of all days that are smoke days	72.4	%
Percent exceedance days	14.6	%
Percent exceedance days that are smoke days	79.2	%
California October 2020		
Mean morning difference (6 am smoke day - non smoke day)	-0.6	ppb
Mean afternoon difference (2 pm smoke day - non smoke day)	14.4	ppb
Mean slope difference (6 am to 2pm slope smoke day - non smoke day)	1.9	ppb/hr
Contribution from morning (baseline)	-4.5	%
Contribution from enhanced photochemistry (from slope)	104.4	%
Contribution from enhanced photochemistry (by subtraction)	104.5	%
Percent of all days that are smoke days	53.1	%
Percent exceedance days	9.8	%
Percent exceedance days that are smoke days	88.5	%

Table 7. Estimated contribution to enhanced afternoon (2pm) O3 from local photochemistry vs enhanced baseline for Illinois (June) and Texas (September) in 2023.

Illinois June 2023		
Mean morning difference (6 am smoke day - non smoke day)	4.1	ppb
Mean afternoon difference (2 pm smoke day - non smoke day)	17.8	ppb
Mean slope difference (6 am to 2pm slope smoke day - non smoke day)	1.7	ppb/hr
Contribution from morning (baseline)	23.2	%
Contribution from enhanced photochemistry (from slope)	74.4	%
Contribution from enhanced photochemistry (by subtraction)	76.8	%
Percent of all days that are smoke days	88.4	%
Percent exceedance days	32.7	%
Percent exceedance days that are smoke days	96.8	%
Texas September 2023		
Mean morning difference (6 am smoke day - non smoke day)	2.7	ppb
Mean afternoon difference (2 pm smoke day - non smoke day)	13.1	ppb
Mean slope difference (6 am to 2pm slope smoke day - non smoke day)	1.3	ppb/hr
Contribution from morning (baseline)	20.2	%
Contribution from enhanced photochemistry (from slope)	79.8	%
Contribution from enhanced photochemistry (by subtraction)	79.8	%
Percent of all days that are smoke days	35.1	%
Percent exceedance days	9.6	%
Percent exceedance days that are smoke days	76.3	%

Task 5: Develop an online GAM and visualization tool for use by state and other agencies.

As part of this project, we developed an R-Shiny web app to allow state and other agencies the ability to run their own GAM and to visualize our existing results. The app can be found at:

https://smoke.shinyapps.io/rsGAM/

At the initial start page, we first request an email from all users and, optionally, their name and affiliation. This is so we can track users and let them know about updates to the app. Once the user enters the app main page, there is an introduction page, similar to the page below.

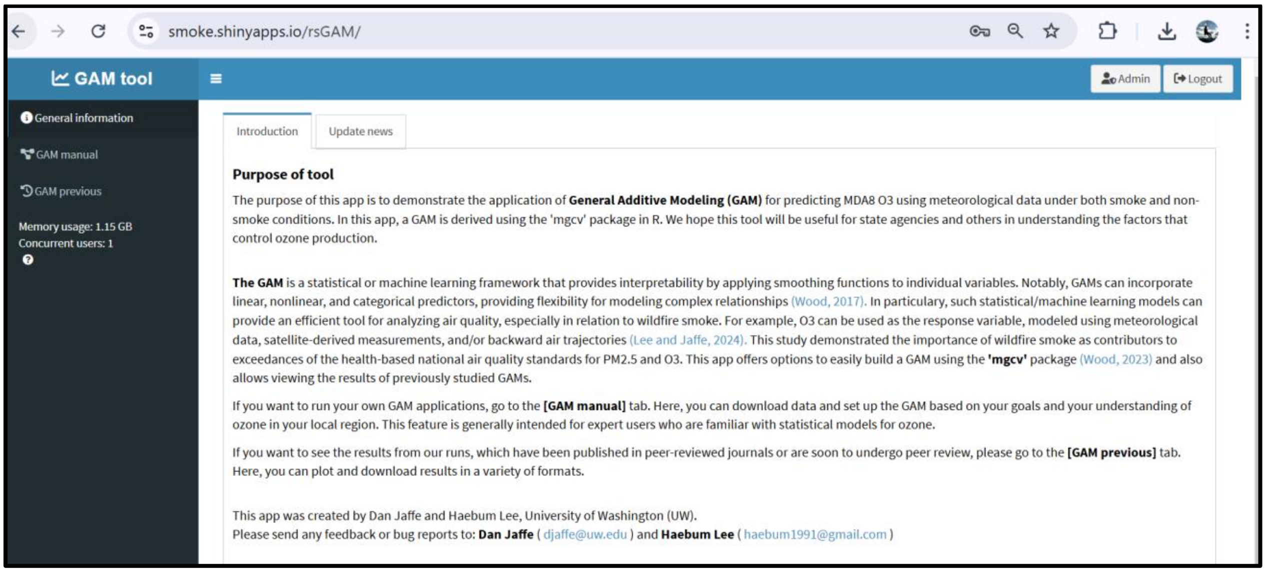

Figure 12. Opening screen of R-Shiny app.

Here the user makes a choice of downloading air quality and meteorological data and running their own GAM ("GAM manual") or plotting and downloading results from one of our existing GAM runs ("GAM previous").

If the user chooses "GAM manual" from the main menu, they will get to a page that looks like this. Here the user can identify the O_3, $PM_{2.5}$ and meteorological sites that they want to use.

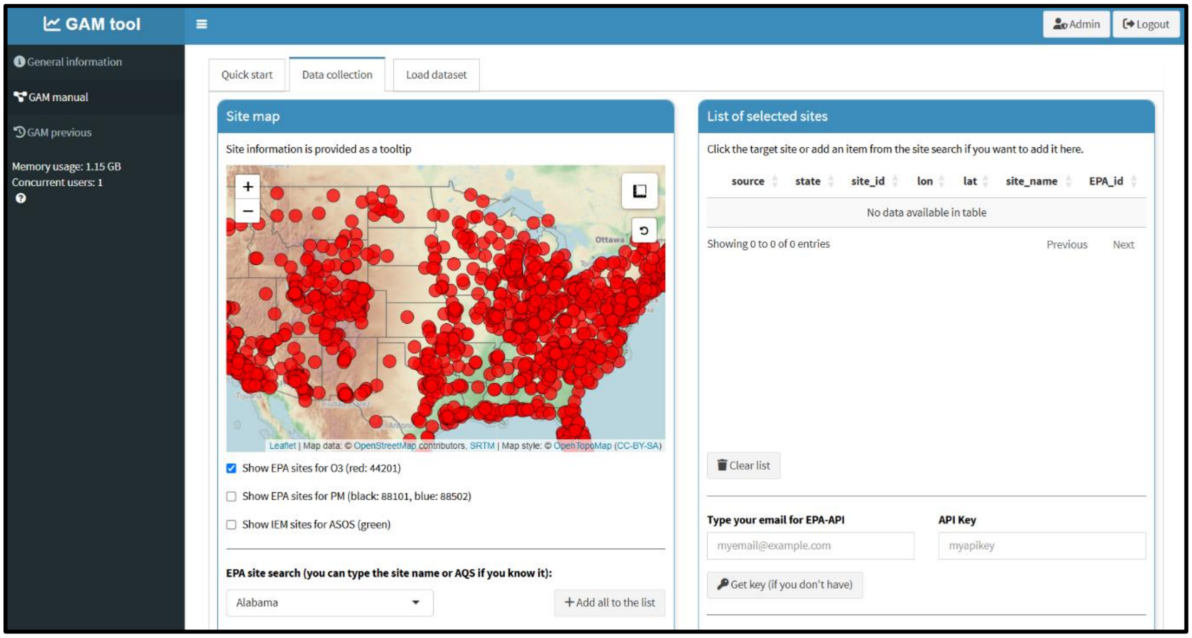

Figure 13. Screenshot showing "GAM manual" page in R-Shiny app.

There are some things to consider in running your own GAM:

1. The application can identify, download and merge large amounts of O_3, $PM_{2.5}$ and meteorological data in real time fairly efficiently. There is an interface that allows a user to easily select the time frame and sites.
2. The application can then identify smoke days using one of several statistical approaches that we recommend.
3. From the merged dataset, the user can then setup a GAM and run it, including doing CV runs on their dataset.
4. The current configuration has a limitation on the type of met data that is automatically downloaded. As the trajectory data require a more complex calculation for incorporation into the GAMs, at present trajectory data can not be included into the GAMs automatically. This means for most sites, our previous GAM runs, which include trajectory data, will probably have a better R^2 than those a user can do on this app.   It is possible for a user to upload any type of dataset, including analyzed trajectory data, and these could then be incorporated into the GAM.

5. Running the GAMs requires some advanced statistical knowledge. Thus, most users will want to stick with results from one of our previous GAM runs.

If the user chooses "GAM previous" from the main menu, they will get to a page that looks like this. The points shown in green are AQS sites that were included in our previous GAM run.

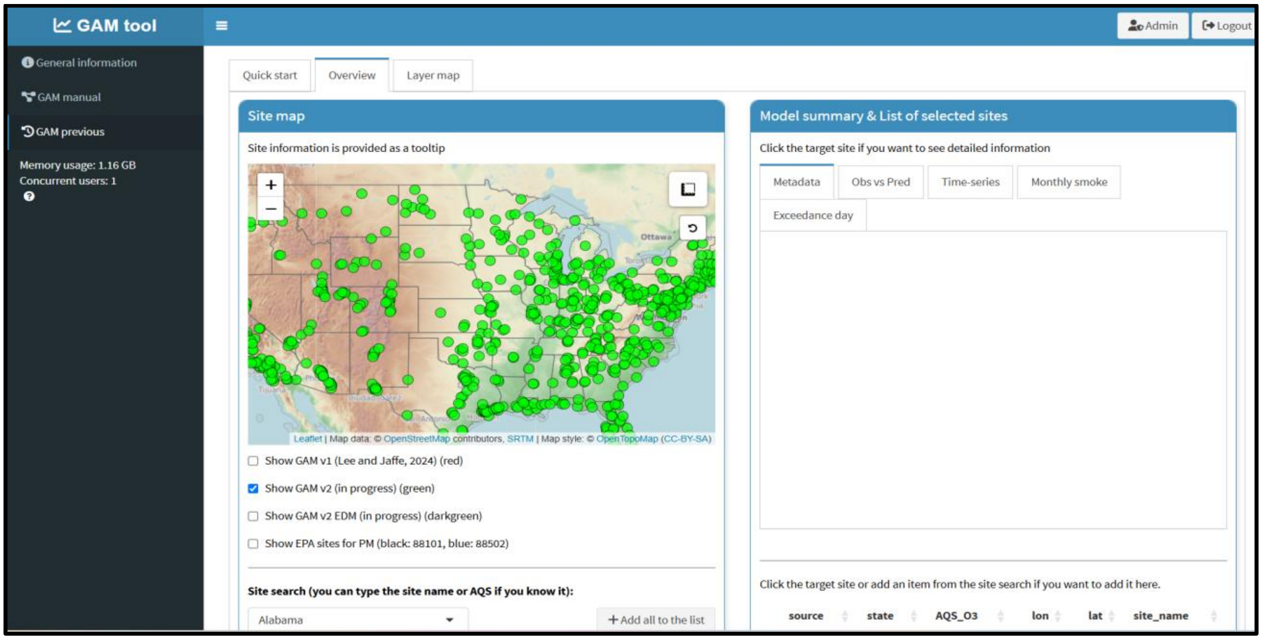

Figure 14. Screenshot showing "GAM previous" page in R-Shiny app.

Clicking on one of these sites gives information on the GAM results for that site. Below, I show results from the South DeKalb site in Georgia. On the right side of the screen are the summary GAM results from this site.

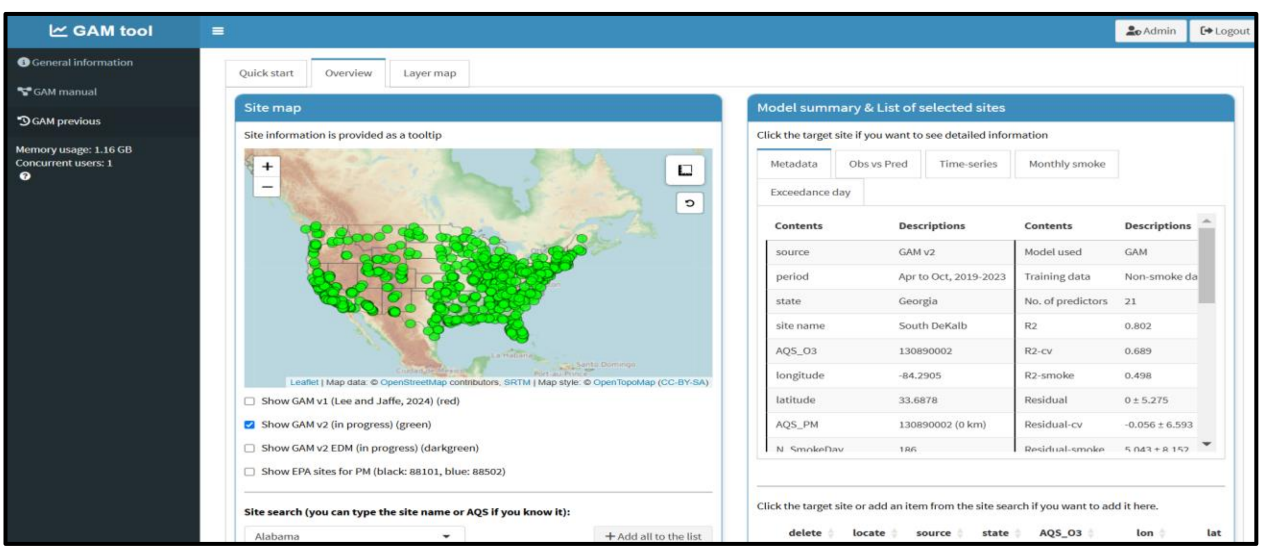

Figure 15. Choosing specific sites on "GAM previous" page in R-Shiny app.

On the top right side of this screen, there are various choices to display the South DeKalb GAM results. For example, there is a time series of the results. This shows the daily MDA8 values, whether we identified this day as a smoke day or not, the GAM predicted MDA8 values and the smoke contribution to the MDA8.

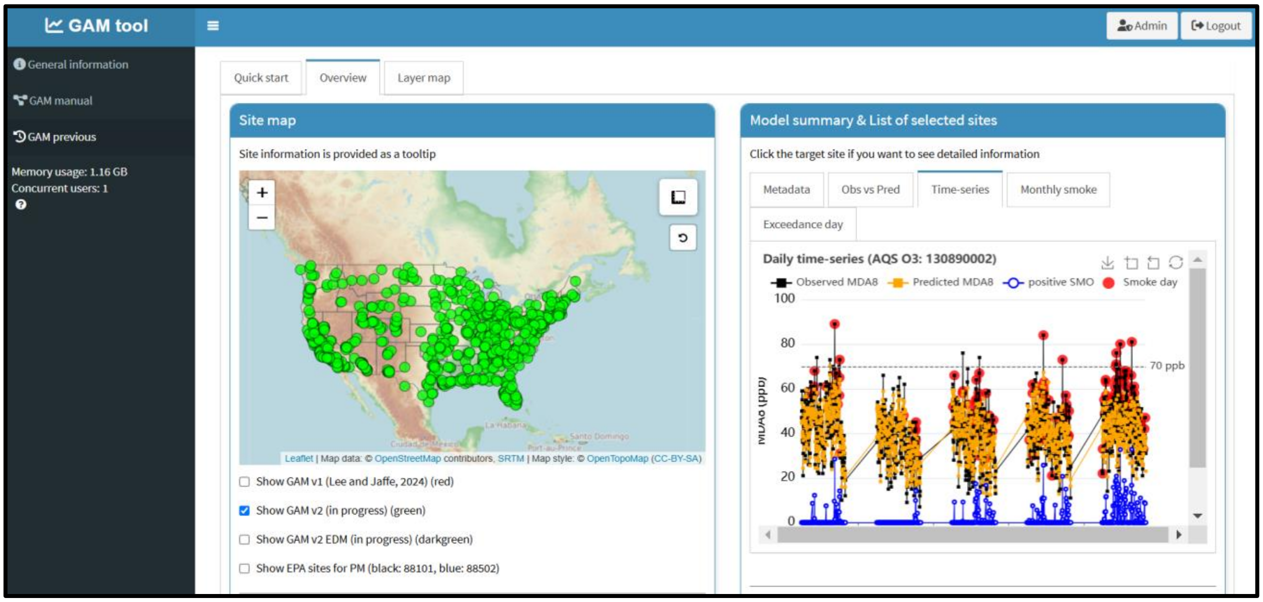

Figure 16. Example plots from "GAM previous" page in R-Shiny app.

Another useful set of plots is found in the "Layer map" tab. Here one can plot results for individual days, or summarize results by state, etc. An example of this was shown earlier in this report in Figure 6.

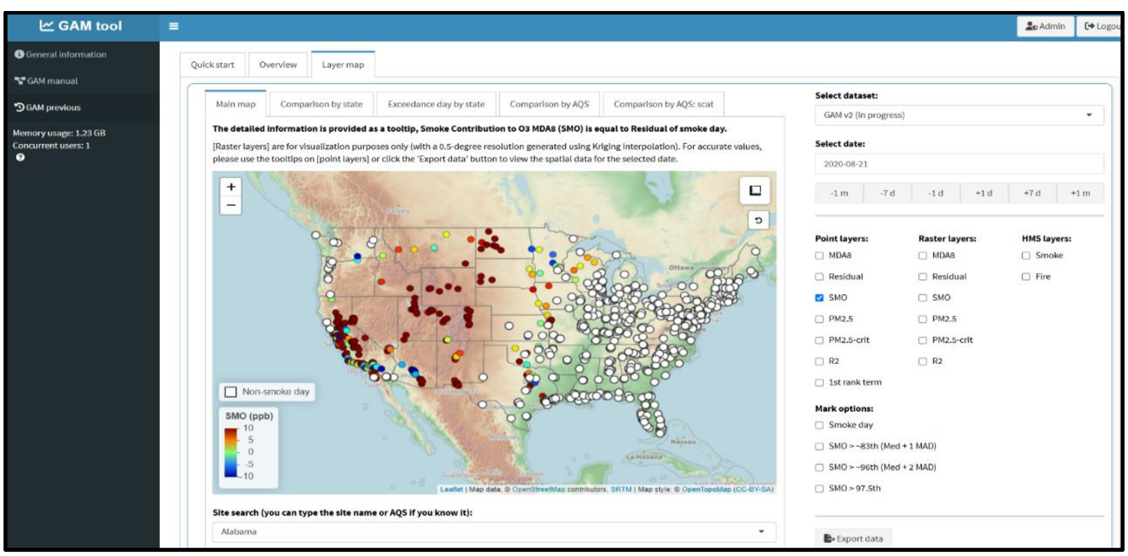

Figure 17. Map of SMO values from "GAM previous" page in R-Shiny app.

But a particularly useful feature is to toggle on/off the check box for SMO>97.5th (circled in red). This allows a user to identify those SMO values that exceeded the 97.5th percentile.

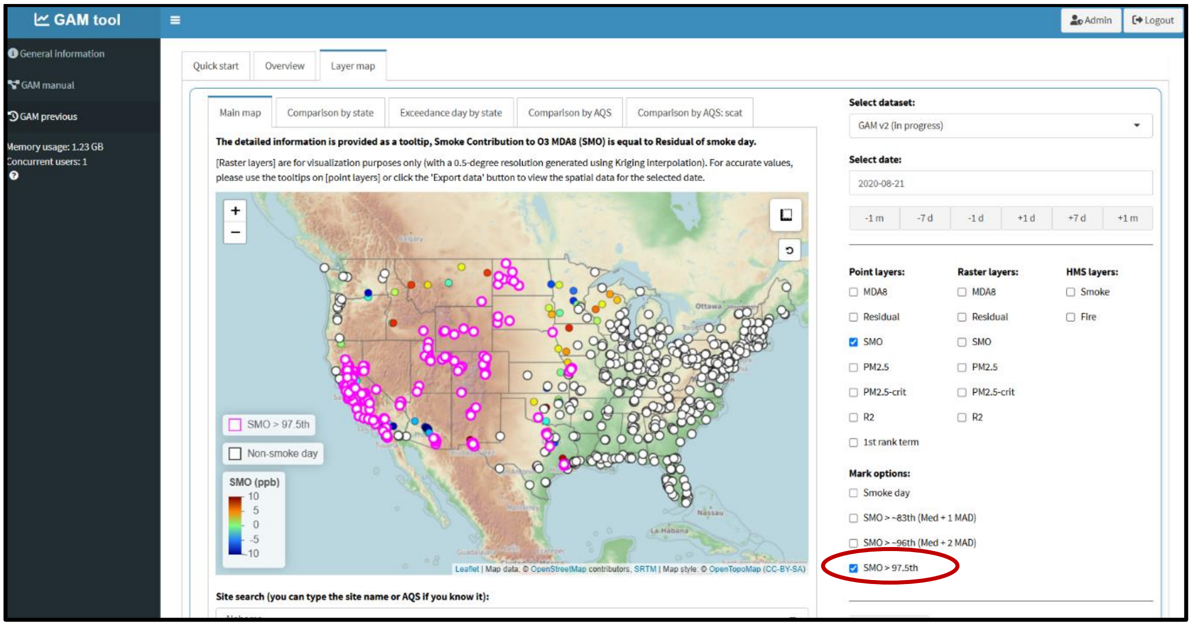

Figure 18. Map of SMO values with check box to indicate SMO values greater than 97.5th percentile from "GAM previous" page in R-Shiny app.

We believe that showing the spatial pattern in the smoke influence, the GAM residual and the residuals greater than the 97.5th percentile will significant add to our confidence that smoke did indeed influence a specific monitoring location.

There are a number of other features in the app. We encourage any readers to explore the app. As it is a new application and a work in progress, please let me know about any issues and/or suggested improvements.

Summary and suggestions for future research.

The primary goals of this project were to develop a method to routinely identify smoke influence on $PM_{2.5}$ and O_3 concentrations at all regulatory air quality sites in the U.S. To start, I generated a dataset of daily O_3, $PM_{2.5}$, satellite and meteorological parameters for 802 Air Quality Sites in all 50 states, plus the District of Columbia for every day in 2019-2023, resulting in 763,482 individual daily data records. To identify days with a significant contribution from smoke I used the NOAA Hazard Mapping System-Fire and Smoke Product (hereafter simply HMS) combined with surface $PM_{2.5}$.

The HMS product is based on multiple satellite indicators to identify smoke influence. But as it is based only on satellite data, it cannot indicate whether smoke is at the surface or not. For each site, we develop a surface "$PM_{2.5}$ criteria" to define smoke influenced days. We first use the HMS data to define a set of days that are possibly smoke influenced. Any day with an overhead HMS smoke identification is termed HMS =0. For the $PM_{2.5}$ smoke criteria, previously we used the mean + 1 standard deviation of the non-HMS days (HMS=0) (Jaffe 2020; 2021). However, at some sites I found a strong log-normal distribution of $PM_{2.5}$ data, which means that the $PM_{2.5}$ criteria would be very high and we would miss many smoke influenced days. I chose to switch to a method based on the median plus the median absolute deviation (MAD) of the HMS=0 days to define the $PM_{2.5}$ smoke criteria. This method is preferred for non-normal distributions (Leys et al 2021). Using this approach, I identified 17% of all days in April-October 2019-2023 as having surface smoke influence. This compares with 14% of all days for May-September 2019-2023, using the mean + 1 standard deviation as the $PM_{2.5}$ criteria.

The distribution of smoke days is highly variable in space and time. Sites in the Central U.S. had the highest frequency of smoke impacted days, 20.3% of days for April-October 2019-2023. But this increased to 39.6% of April-October days in 2023 due to the massive Canadian wildfires that burned in that year. This also more than doubled the number of O_3 exceedance days in the Central U.S. In general, we find a strong relationship between the number of O_3 exceedance days and smoke frequency in all areas of the country.

I used the non-smoke data as "training data" for the GAMs. For each site, GAMs were run individually to predict the Maximum Daily Average 8-hour (MDA8) O_3 concentration from the observed meteorological parameters, satellite observations and other predictors. Each model was initiated with the same predictors and GAM equation, but a different number of predictors were statistically significant at each site. The model is very good at predicting the MDA8, with an overall R^2 of 0.84, for all data together, and a mean R^2 of 0.77 using the GAMs from all sites. I examined the model results using 10-fold cross validation and other metrics. The residual for the training dataset, defined as the observed MDA8 minus the model predicted MDA8, has a mean value of 0 and a standard deviation of 4.8 ppb. For each site, the same model is used to predict the MDA8 for the smoke days. Since smoke, nor $PM_{2.5}$, are included in the model we can interpret the residual for the smoke days as the change in the MDA8 due to the presence of smoke. The overall mean and standard deviation of the residual on the smoke days at all sites is 3.8 ± 8.0 ppb. Overall, I found that out of 13,536 exceedance days, 6200 (45.8%) had smoke influence.

While the residual on individual smoke days is a measure of the enhancement of the MDA8 due to the smoke, we must consider the large variability in these residuals. Based on published guidance from the U.S. EPA (U.S. EPA. 2016), I apply a statistical metric on the residual to determine if it sufficiently different from the training dataset. In this case, the residual must be greater than the 97.5th percentile of the non-smoke residuals. If it is, then we can consider that MDA8 to have been enhanced by smoke. Using this metric, I find that 2.5% of the non-smoke residuals are greater than the 97.5th percentile (by definition), but 21% of the smoke residuals are greater than this metric. This demonstrates a signficant enhancement in the MDA8 due to smoke on more than 27,000 days over this 5-year time period. Of the O_3 exceedance days that we determined to be smoke influenced, 73% had residuals that exceeded the 97.5th percentile.

Knowing that smoke enhances the MDA8, but with a large degree of variability, we would like to better understand the factors that control this variability. I examined several factors that might be expected to have some influence on the urban O_3 production in the presence of smoke, including: $PM_{2.5,}$ NO_2 tropospheric column density, and HMS smoke type (heavy, medium, light). Unfortunately none of these parameters showed a strong relationship with the GAM residual. Only the HMS smoke classification (heavy, medium or light) showed any relationship with the residual, but it was rather weak. This tells us that smoke chemistry and O_3 production in urban areas can not be explained by a simple one-factor prediction approach. It seems likely that multiple factors are needed to explain the urban O_3 enhancements we see in smoke.

Next we used hourly O_3 data on smoke days to examine whether the enhanced O_3 we see is most likely associated with enhanced local O_3 production or transport of O_3 formed elsewhere into the region. Enhanced O_3 production would be indicated by an enhanced rate of O_3 production (dO_3/dT) during the daytime hours, whereas transport would not show any particular diurnal pattern. Examining data from several time periods with strong smoke influence shows that local photochemical production is the most important mechanism. This is likely from the combination of smoke VOCs and urban/industrial NO_x. This finding has implications for O_3 control, since local production of O_3 still depends on local NO_x emissions.

Finally, as part of this project, we developed an R-Shiny app that can be used to plot and display these results and run your own GAM/machine learning models. The app includes the ability to visualize and extract results from our existing GAM runs. In addition, expert users can use this app to download and merge large amounts of air quality and meteorological data and run a GAM for smoke conditions in future years. We believe this app will be useful to state and other agencies for both understanding smoke chemistry and in developing exceptional event demonstration packages.

Several aspects of smoke chemistry remain rather murky. In particular, understanding what controls O_3 production in urban areas during smoke events remains unclear. Given the results in Task 4 using the hourly O_3 data, it is reasonable to conclude that most O_3 production occurs locally from a combination of smoke VOCs and anthropogenic NOx. Future work could combine and/or compare statistical modeling approaches with detailed photochemical modeling. Another aspect that is ripe for future work is comparing these results to different approaches. At present, there are few examples of national analyses to calculate the contribution of smoke to surface $PM_{2.5}$ and O_3, but this is likely to change in the future. When other approaches become available, it will be very useful to compare the results from different methodologies.

There are a number of uncertainties to consider in this analysis:

1) Smoke days determined by HMS and $PM_{2.5}$-criteria may be incorrectly identified.
   For all sites, the values for the $PM_{2.5}$ smoke criteria ranged from 4.8 to 22.5 µg m^{-3}, with the highest value observed for the Pomona site in California. At this site, although 20% of days had an overhead HMS signal (i.e., HMS = 1), only 9% of the days experienced daily $PM_{2.5}$ exceeding the $PM_{2.5}$-criteria, resulting in a smoke day frequency of only 4%. In other words, the higher $PM_{2.5}$-criteria, results in a reduced ability to detect moderate smoke levels
2) Uncertainty arises from the data merging process.
   In instances where O_3 data were available but $PM_{2.5}$ data were not, we combined the O_3 data with $PM_{2.5}$ data from the closest station within a 25 km radius. In this study, only 55% have identical coordinates (i.e. co-location) for both O_3 and $PM_{2.5}$ monitoring stations. Increasing the co-location of $PM_{2.5}$ and O_3 monitoring sites in the future would reduce this uncertainty

References

Abatzoglou, J. T.; Williams, A. P.; Barbero, R. 2019. Global Emergence of Anthropogenic Climate Change in Fire Weather Indices. *Geophys. Res. Lett.* 46, 326–336. DOI: 10.1029/2018GL080959

Buysse C.E. Kaulfus A. Nair U. and Jaffe D.A. 2019. Relationships between particulate matter, ozone, and nitrogen oxides during urban smoke events in the western US. Environ. Sci. Technol. DOI: 10.1021/acs.est.9b05241.

Burke, M., Childs, M. L., de la Cuesta, B., Qiu, M., Li, J., Gould, C. F., Heft-Neal, S., & Wara, M. 2023. The contribution of wildfire to PM2.5 trends in the USA. *Nature*, *622*(7984), 761–766. https://doi.org/10.1038/s41586-023-06522-6

Childs, M. L., Li, J., Wen, J., Heft-Neal, S., Driscoll, A., Wang, S., Gould, C. F., Qiu, M., Burney, J., & Burke, M. 2023. Daily Local-Level Estimates of Ambient Wildfire Smoke $PM_{2.5}$ for the Contiguous US. *Environmental Science and Technology*, *56*(19), 13607–13621. https://doi.org/10.1021/ACS.EST.2C02934.

Connolly, R., Marlier, M. E., Garcia-Gonzales, D. A., Wilkins, J., Su, J., Bekker, C., Jung, J., Bonilla, E., Burnett, R. T., Zhu, Y., & Jerrett, M. 2024. Mortality attributable to PM2.5 from wildland fires in California from 2008 to 2018. *Science Advances*, *10*(23). https://doi.org/10.1126/sciadv.adl1252

Doubleday, A., Schulte, J., Sheppard, L., Kadlec, M., Dhammapala, R., Fox, J., & Busch Isaksen, T. 2020. A case-crossover study. *Environ Health* . https://doi.org/10.1186/s12940-020-0559-2

Elser H, Frankland TB, Chen C, et al. Wildfire Smoke Exposure and Incident Dementia. *JAMA Neurol.* Published online November 25, 2024. doi:10.1001/jamaneurol.2024.4058

Flynn M.T., Mattson E.J., Jaffe D.A., and Gratz L.E. 2021. Spatial patterns in summertime surface ozone in the southern Front Range of the Rocky Mountains, USA. Elementa: Science of the Anthropocene. 9 (1): 00104. https://doi.org/10.1525/elementa.2020.00104.

Gong X., Kaulfus A., Nair U. and Jaffe D.A. 2017. Quantifying O_3 impacts in urban areas due to wildfires using a Generalized Additive Model. Envir. Sci. Tech. DOI: 10.1021/acs.est.7b03130.

Gong X., Hong S. Jaffe D.A. 2018. Ozone in China: Spatial distribution and leading meteorological factors controlling O_3 in 16 Chinese cities. Aer. Air Qual. Res, DOI: 10.4209/aaqr.2017.10.0368.

Governors Hobbs, Cox, Polis, & Gordon. 2024. *Arizona, Colorado, Utah, and Wyoming send letter to President Biden on Ozone – Westar*. Accessed 11/27/24. http://westar.cira.colostate.edu/wp-content/uploads/2024/05/letter-to-Biden_ozone.pdf

S. Heft-Neal, C.F. Gould, M.L. Childs, M.V. Kiang, K.C. Nadeau, M. Duggan, E. Bendavid, M. Burke, 2023. Emergency department visits respond nonlinearly to wildfire smoke, Proc. Natl. Acad. Sci. U.S.A. 120 (39) e2302409120, https://doi.org/10.1073/pnas.2302409120.

Hu C., Kang P., Jaffe D.A. Li C., Zhang X., Wu K., and Zhou M. 2021. Understanding the impact of meteorology on ozone in 334 cities of China. Atmos. Envir., 248, 118221, https://doi.org/10.1016/j.atmosenv.2021.118221,

Jaffe D. Role of Meteorology, Emissions and Smoke on Ozone in the South Coast Air Basin. Final project report for CRC Project A-118, Coordinating Research Council, Alpharetta, GA, January 2020. Available at: http://crcao.org/wp-content/uploads/2020/01/CRC-Project-A-118-Final-Report_Jan2020.pdf, 2020.

Jaffe D. Evaluation of Ozone Patterns and Trends in 8 Major Metropolitan Areas in the U.S. Final project report for CRC Project A-124, Coordinating Research Council, Alpharetta, GA, March 2021. Available at: http://crcao.org/wp-content/uploads/2021/04/CRC-Project-A-124-Final-Report_Mar2021.pdf

Jaffe D.A., O'Neill S.M., Larkin N.K., Holder A.L, Peterson D.L., Halofsky J.E. and Rappold A.G. 2020. Wildfire and prescribed burning impacts on air quality in the United States, J. Air and Waste Mgt. Assn., DOI: 10.1080/10962247.2020.1749731.

Lee, H., & Jaffe, D. A. 2024. Impact of wildfire smoke on ozone concentrations using a Generalized Additive model in Salt Lake City, Utah, USA, 2006–2022. *Journal of the Air and Waste Management Association*, *74*(2), 116–130. https://doi.org/10.1080/10962247.2023.2291197

Leys, Christophe, et al. 2013. Detecting Outliers: Do Not Use Standard Deviation around the Mean, Use Absolute Deviation around the Median. J.Expt. Soc. Psych., vol. 49, no. 4, pp. 764–66, doi:10.1016/j.jesp.2013.03.013.

McClure, C. D., & Jaffe, D. A. 2018a. US particulate matter air quality improves except in wildfire-prone areas. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), 7901–7906. https://doi.org/10.1073/pnas.1804353115

McClure C.D. and Jaffe D.A. 2018b. Investigation of High Ozone Events due to Wildfire Smoke in an Urban Area. Atmos. Envir. https://doi.org/10.1016/j.atmosenv.2018.09.021.

O'Dell, K., Bilsback, K., Ford, B., Martenies, S. E., Magzamen, S., Fischer, E. V., & Pierce, J. R. 2021. Estimated Mortality and Morbidity Attributable to Smoke Plumes in the United States: Not Just a Western US Problem. *GeoHealth*, *5*(9), e2021GH000457. https://doi.org/10.1029/2021GH000457

Rickly, P. S. et al. 2023. Influence of Wildfire on Urban Ozone: An Observationally Constrained Box Modeling Study at a Site in the Colorado Front Range. *Environ. Sci. Technol. 57*, 1257−1267, doi: 10.1021/acs.est.2c06157.

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., & Davidson, P. M. 2009. Description and verification of the NOAA smoke forecasting system: The 2007 fire season. In *Weather and Forecasting* (Vol. 24, Issue 2, pp. 361–378). https://doi.org/10.1175/2008WAF2222165.1

Ruminski, M., Kondragunta, S., Draxler, R., & Zeng, J. 2011. *Recent Changes to the Hazard Mapping System*. https://www3.epa.gov/ttnchie1/conference/ei15/session10/ruminiski.pdf

Simon, H, Reff, A, Wells, B, Xing, J and Frank, N. 2015. Ozone trends across the United States over a period of decreasing NOx and VOC emissions. *Environ Sci Technol* 49(1): 186–195. DOI: https://doi. org/10.1021/es504514z, 2015.

U.S. EPA. 2016. Guidance on the preparation of exceptional events demonstrations for wildfire events that May influence ozone concentrations. Accessed 9/8/2024. https:// www.epa.gov/sites/default/files/2016-09/documents/excep tional_events_guidance_9-16-16_final.pdf

U.S. EPA. 2024. Treatment of Air Quality Monitoring Data Influenced by Exceptional Events. Accessed 11/27/24. https://www.epa.gov/air-quality-analysis/treatment-air-quality-monitoring-data-influenced-exceptional-events.

Wei, J.; Wang, J.; Li, Z.; Kondragunta, S.; Anenberg, S.; Wang, Y.; Zhang, H.; Diner, D.; Hand, J.; Lyapustin, A.; Kahn, R.; Colarco, P.; da Silva, A.; Ichoku, C. 2023. Long-term Mortality Burden Trends Attributed to Black Carbon and $PM_{2.5}$ from Wildfire Emissions Across the

Continental USA from 2000 to 2020: A Deep Learning Modeling Study. *The Lancet Planetary Health* 2023, 7 (12), E963–E975. DOI: 10.1016/S2542-5196(23)00235-8.

Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., & Lettenmaier, D. P. 2019. Observed Impacts of Anthropogenic Climate Change on Wildfire in California. *Earth's Future*, *7*(8), 892–910. https://doi.org/10.1029/2019EF001210

Wilmot, T. Y., Mallia, D. V., Hallar, A. G., & Lin, J. C. 2022. Wildfire activity is driving summertime air quality degradation across the western US: a model-based attribution to smoke source regions. *Environmental Research Letters*, *17*(11). https://doi.org/10.1088/1748-9326/ac9a5d

Wood, S.N. 2017. Generalized Additive Models, An Introduction with R, Second Edition. Chapman and Hall/CRC, Boca Raton, FL.

Appendix

Tables A1-A5. Total of exceedance days, exceedance days with smoke and exceedance days with smoke that exceed the 97.5th percentile by state for each year (2019-2023).

2019	Sum all exceedance days	Sum all exceedance days with smoke	Sum all exceedance days with smoke that exceed 97.5th percentile	Fraction of all exceedance days that have smoke and exceed the 97.5th percentile
Alabama	20	10	6	0.30
Alaska	0	0	0	NA
Arizona	57	1	0	0.00
Arkansas	1	1	1	1.00
California	1086	11	4	0.00
Colorado	27	2	1	0.04
Connecticut	58	25	18	0.31
Delaware	4	0	0	0.00
D.C.	6	1	1	0.17
Florida	7	1	1	0.14
Georgia	28	18	14	0.50
Hawaii	0	0	0	NA
Idaho	0	0	0	NA
Illinois	22	12	12	0.55
Indiana	11	4	4	0.36
Iowa	1	0	0	0.00
Kansas	1	1	1	1.00
Kentucky	6	0	0	0.00
Louisiana	7	1	1	0.14
Maine	1	1	1	1.00
Maryland	38	6	6	0.16
Massachusetts	1	1	1	1.00
Michigan	5	2	1	0.20
Minnesota	3	2	2	0.67
Mississippi	2	2	2	1.00
Missouri	9	3	3	0.33
Montana	0	0	0	NA
Nebraska	0	0	0	NA
Nevada	7	0	0	0.00
New Hampshire	0	0	0	NA
New Jersey	16	2	2	0.13
New Mexico	33	0	0	0.00
New York	19	6	5	0.26
North Carolina	13	5	4	0.31

North Dakota	0	0	0	NA
Ohio	28	11	11	0.39
Oklahoma	4	2	2	0.50
Oregon	2	1	1	0.50
Pennsylvania	20	2	2	0.10
Rhode Island	2	1	1	0.50
South Carolina	4	0	0	0.00
South Dakota	4	1	1	0.25
Tennessee	4	1	1	0.25
Texas	116	15	11	0.09
Utah	28	0	0	0.00
Vermont	0	0	0	NA
Virginia	2	1	1	0.50
Washington	0	0	0	NA
West Virginia	0	0	0	NA
Wisconsin	15	11	11	0.73
Wyoming	0	0	0	NA

2020	Sum all exceedance days	Sum all exceedance days with smoke	Sum all exceedance days with smoke that exceed 97.5th percentile	Fraction of all exceedance days that have smoke and exceed the 97.5th percentile
Alabama	1	1	1	1.00
Alaska	0	0	0	NA
Arizona	127	84	53	0.42
Arkansas	3	0	0	0.00
California	1959	1012	729	0.37
Colorado	93	83	65	0.70
Connecticut	28	0	0	0.00
Delaware	2	0	0	0.00
D.C.	0	0	0	NA
Florida	3	0	0	0.00
Georgia	1	0	0	0.00
Hawaii	0	0	0	NA
Idaho	3	3	3	1.00
Illinois	75	25	13	0.17
Indiana	25	17	12	0.48
Iowa	0	0	0	NA
Kansas	3	1	1	0.33
Kentucky	9	0	0	0.00
Louisiana	10	2	1	0.10
Maine	1	0	0	0.00
Maryland	5	0	0	0.00
Massachusetts	4	0	0	0.00
Michigan	39	21	14	0.36
Minnesota	2	1	1	0.50
Mississippi	1	0	0	0.00
Missouri	6	1	0	0.00
Montana	0	0	0	NA
Nebraska	0	0	0	NA
Nevada	75	47	46	0.61
New Hampshire	0	0	0	NA
New Jersey	6	0	0	0.00
New Mexico	27	19	18	0.67
New York	11	0	0	0.00
North Carolina	0	0	0	NA
North Dakota	0	0	0	NA
Ohio	32	5	1	0.03
Oklahoma	3	0	0	0.00

Oregon	5	4	3	0.60
Pennsylvania	14	0	0	0.00
Rhode Island	7	0	0	0.00
South Carolina	1	0	0	0.00
South Dakota	2	1	1	0.50
Tennessee	0	0	0	NA
Texas	110	43	33	0.30
Utah	74	45	41	0.55
Vermont	0	0	0	NA
Virginia	0	0	0	NA
Washington	0	0	0	NA
West Virginia	0	0	0	NA
Wisconsin	56	25	20	0.36
Wyoming	9	9	9	1.00

2021	Sum all exceedance days	Sum all exceedance days with smoke	Sum all exceedance days with smoke that exceed 97.5th percentile	Fraction of all exceedance days that have smoke and exceed the 97.5th percentile
Alabama	2	2	2	1.00
Alaska	0	0	0	NA
Arizona	201	62	35	0.17
Arkansas	7	6	3	0.43
California	1611	752	494	0.31
Colorado	244	201	144	0.59
Connecticut	52	36	29	0.56
Delaware	5	2	1	0.20
D.C.	9	6	6	0.67
Florida	3	0	0	0.00
Georgia	7	2	2	0.29
Hawaii	0	0	0	NA
Idaho	21	21	18	0.86
Illinois	48	27	21	0.44
Indiana	12	7	7	0.58
Iowa	0	0	0	NA
Kansas	4	2	2	0.50
Kentucky	10	9	9	0.90
Louisiana	12	10	10	0.83
Maine	7	5	5	0.71
Maryland	31	20	18	0.58
Massachusetts	10	3	3	0.30
Michigan	31	10	10	0.32
Minnesota	4	3	2	0.50
Mississippi	1	1	1	1.00
Missouri	14	10	10	0.71
Montana	9	8	8	0.89
Nebraska	1	1	1	1.00
Nevada	190	141	129	0.68
New Hampshire	2	1	1	0.50
New Jersey	24	16	14	0.58
New Mexico	50	34	30	0.60
New York	38	25	22	0.58
North Carolina	2	2	2	1.00
North Dakota	5	5	5	1.00
Ohio	17	8	8	0.47
Oklahoma	15	10	9	0.60

Oregon	6	6	6	1.00
Pennsylvania	40	24	14	0.35
Rhode Island	7	2	2	0.29
South Carolina	1	1	1	1.00
South Dakota	12	10	9	0.75
Tennessee	8	4	4	0.50
Texas	181	79	55	0.30
Utah	257	208	186	0.72
Vermont	0	0	0	NA
Virginia	3	3	3	1.00
Washington	8	7	7	0.88
West Virginia	0	0	0	NA
Wisconsin	43	30	24	0.56
Wyoming	23	23	22	0.96

2022	Sum all exceedance days	Sum all exceedance days with smoke	Sum all exceedance days with smoke that exceed 97.5th percentile	Fraction of all exceedance days that have smoke and exceed the 97.5th percentile
Alabama	1	1	0	0.00
Alaska	0	0	0	NA
Arizona	149	18	11	0.07
Arkansas	5	2	2	0.40
California	1165	166	68	0.06
Colorado	87	16	11	0.13
Connecticut	59	32	17	0.29
Delaware	0	0	0	NA
D.C.	3	3	1	0.33
Florida	3	2	2	0.67
Georgia	9	7	5	0.56
Hawaii	0	0	0	NA
Idaho	4	2	2	0.50
Illinois	45	11	2	0.04
Indiana	33	13	9	0.27
Iowa	0	0	0	NA
Kansas	7	4	3	0.43
Kentucky	14	5	3	0.21
Louisiana	6	4	3	0.50
Maine	4	1	1	0.25
Maryland	4	2	1	0.25
Massachusetts	8	8	7	0.88
Michigan	25	10	7	0.28
Minnesota	0	0	0	NA
Mississippi	5	5	5	1.00
Missouri	10	5	3	0.30
Montana	0	0	0	NA
Nebraska	1	1	1	1.00
Nevada	41	4	4	0.10
New Hampshire	3	3	3	1.00
New Jersey	9	4	4	0.44
New Mexico	49	11	7	0.14
New York	14	11	7	0.50
North Carolina	6	6	2	0.33
North Dakota	0	0	0	NA
Ohio	34	21	20	0.59
Oklahoma	41	33	27	0.66

Oregon	7	6	5	0.71
Pennsylvania	17	11	8	0.47
Rhode Island	6	5	5	0.83
South Carolina	1	1	1	1.00
South Dakota	2	1	1	0.50
Tennessee	14	11	9	0.64
Texas	172	82	52	0.30
Utah	59	23	18	0.31
Vermont	0	0	0	NA
Virginia	1	1	1	1.00
Washington	1	0	0	0.00
West Virginia	0	0	0	NA
Wisconsin	36	21	16	0.44
Wyoming	0	0	0	NA

2023	Sum all exceedance days	Sum all exceedance days with smoke	Sum all exceedance days with smoke that exceed 97.5th percentile	Fraction of all exceedance days that have smoke and exceed the 97.5th percentile
Alabama	15	15	14	0.93
Alaska	0	0	0	NA
Arizona	186	30	20	0.11
Arkansas	5	3	3	0.60
California	1016	182	74	0.07
Colorado	43	12	9	0.21
Connecticut	63	60	56	0.89
Delaware	15	15	11	0.73
D.C.	8	8	6	0.75
Florida	11	11	8	0.73
Georgia	35	33	30	0.86
Hawaii	0	0	0	NA
Idaho	1	1	1	1.00
Illinois	281	251	153	0.54
Indiana	120	116	91	0.76
Iowa	64	62	53	0.83
Kansas	34	29	17	0.50
Kentucky	50	50	41	0.82
Louisiana	41	37	30	0.73
Maine	0	0	0	NA
Maryland	55	50	35	0.64
Massachusetts	21	21	20	0.95
Michigan	70	68	48	0.69
Minnesota	67	60	59	0.88
Mississippi	13	10	10	0.77
Missouri	84	69	50	0.60
Montana	0	0	0	NA
Nebraska	13	12	7	0.54
Nevada	48	2	1	0.02
New Hampshire	4	4	4	1.00
New Jersey	42	42	30	0.71
New Mexico	12	3	2	0.17
New York	46	44	43	0.93
North Carolina	16	15	14	0.88
North Dakota	36	36	36	1.00
Ohio	73	68	43	0.59
Oklahoma	73	57	47	0.64

Oregon	4	4	4	1.00
Pennsylvania	83	83	69	0.83
Rhode Island	14	14	14	1.00
South Carolina	6	6	6	1.00
South Dakota	38	31	24	0.63
Tennessee	31	30	24	0.77
Texas	426	291	167	0.39
Utah	48	13	13	0.27
Vermont	2	2	2	1.00
Virginia	8	8	8	1.00
Washington	3	2	2	0.67
West Virginia	0	0	0	NA
Wisconsin	214	209	158	0.74
Wyoming	0	0	0	NA