

# AQRP Monthly Technical Report

<b>PROJECT TITLE</b>	<i>Evaluating the Ability of Statistical and Photochemical Models to Capture the Impacts of Biomass Burning Smoke on Urban Air Quality in Texas</i>	<b>PROJECT #</b>	22-003
<b>PROJECT PARTICIPANTS</b>	Matthew Alvarado, Jen Hegarty, Rick Pernak	<b>DATE SUBMITTED</b>	03-13-2022
<b>REPORTING PERIOD</b>	<b>From:</b> February 1, 2023 <b>To:</b> February 28, 2023	<b>REPORT #</b>	7

A Financial Status Report (FSR) and Invoice will be submitted separately from each of the Project Participants reflecting charges for this Reporting Period. I understand that the FSR and Invoice are due to the AQRP by the 15<sup>th</sup> of the month following the reporting period shown above.

## Detailed Accomplishments by Task for reporting period

During this month we merged comma-separated variable (CSV) files containing a flag to indicate if smoke occurred on a day at selected monitor locations in Houston and El Paso, TX with GAMs predictor dataset files used to predict O<sub>3</sub> and PM<sub>2.5</sub>. The smoke flag files were generated from NOAA Hazard Mapping System (HMS) smoke polygons for 2012 – 2021 that were converted to the GOES CONUS grid. The original GAMs dataset files included meteorological predictors as well as day of the year (doy), day of the week (dow) and year predictors. The production of the merged dataset files is a key milestone of *Task 1: Using GAMs to Quantify Urban Smoke Impacts* and *Deliverable 1 Q3 Compile meteorological and fire predictors into a single dataset*. During the 10-year analysis period there were days with missing HMS smoke polygon files and 1 day for which the smoke polygon file was unreadable. For these days the smoke flag was set to a “not a number” or “nan” value. On days for which the smoke polygon files were available but empty, we entered a no smoke value indicated by the number 0 into the CSV file. A smoke flag value of 1 in the CSV file indicates that smoke was observed over the monitoring location on that day.

After generating the merged datafiles with the binary smoke flag we performed initial GAM test. The binary smoke flag in these tests was included in the GAM as a factor predictor. Predictors can be included into the GAM as smooth functions only if they have at least 5 unique values. The smoke flag has only 2. The results are shown below and are compared with those generated without the smoke flag for TCEQ Work Order Number 582-22-31570.

## Preliminary Analysis

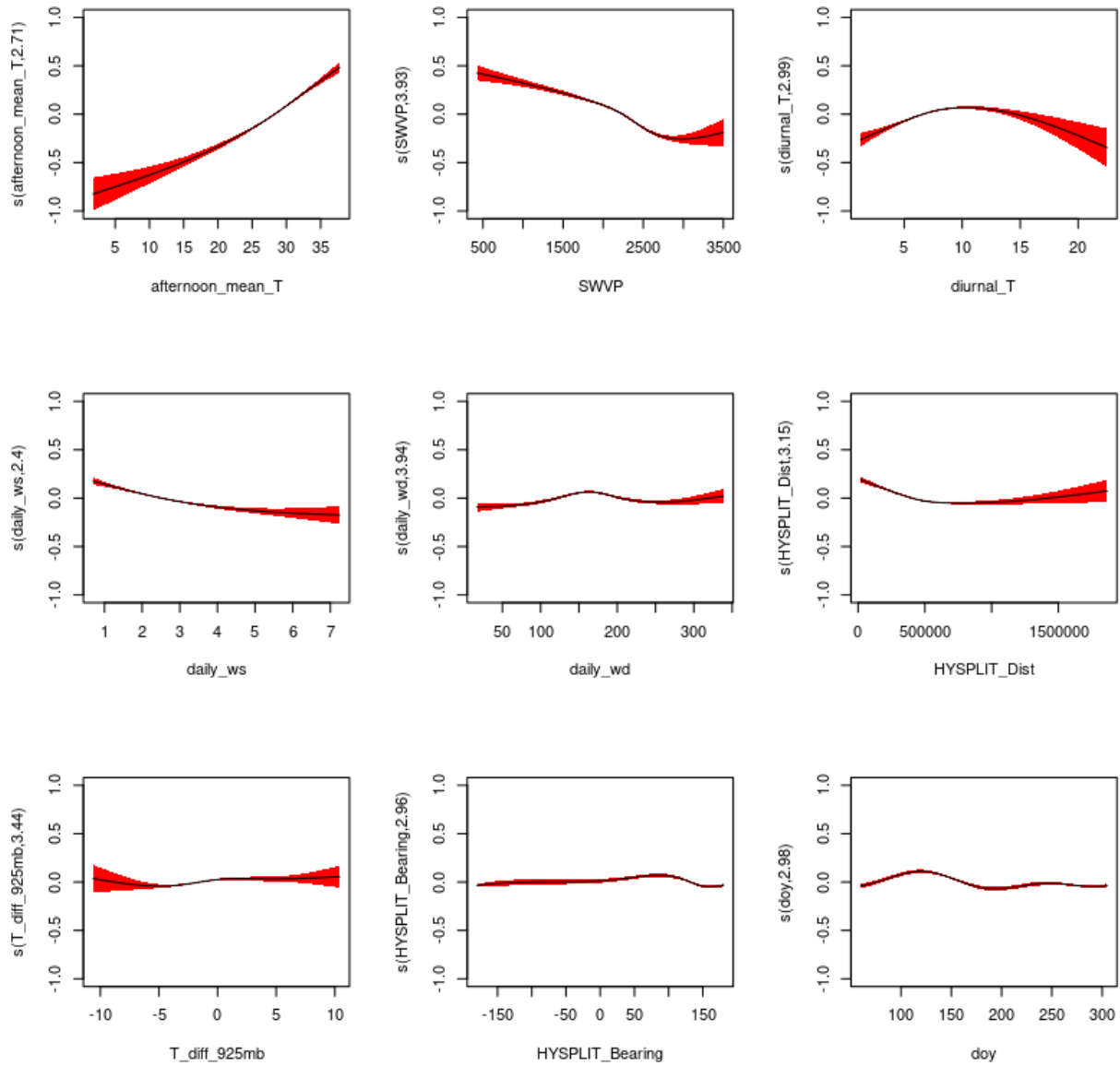
### *Total MDA8 O<sub>3</sub> GAM Results*

Table 1 organizes the total MDA8 O<sub>3</sub> GAM results for Houston and El Paso both without and with the binary smoke flag. This includes figure numbers for the results, the percentage deviance of the MDA8 O<sub>3</sub> values explained by the GAM model, and the level statistical significance of the meteorological predictors in Table 7 of the Final Report to TCEQ for Work

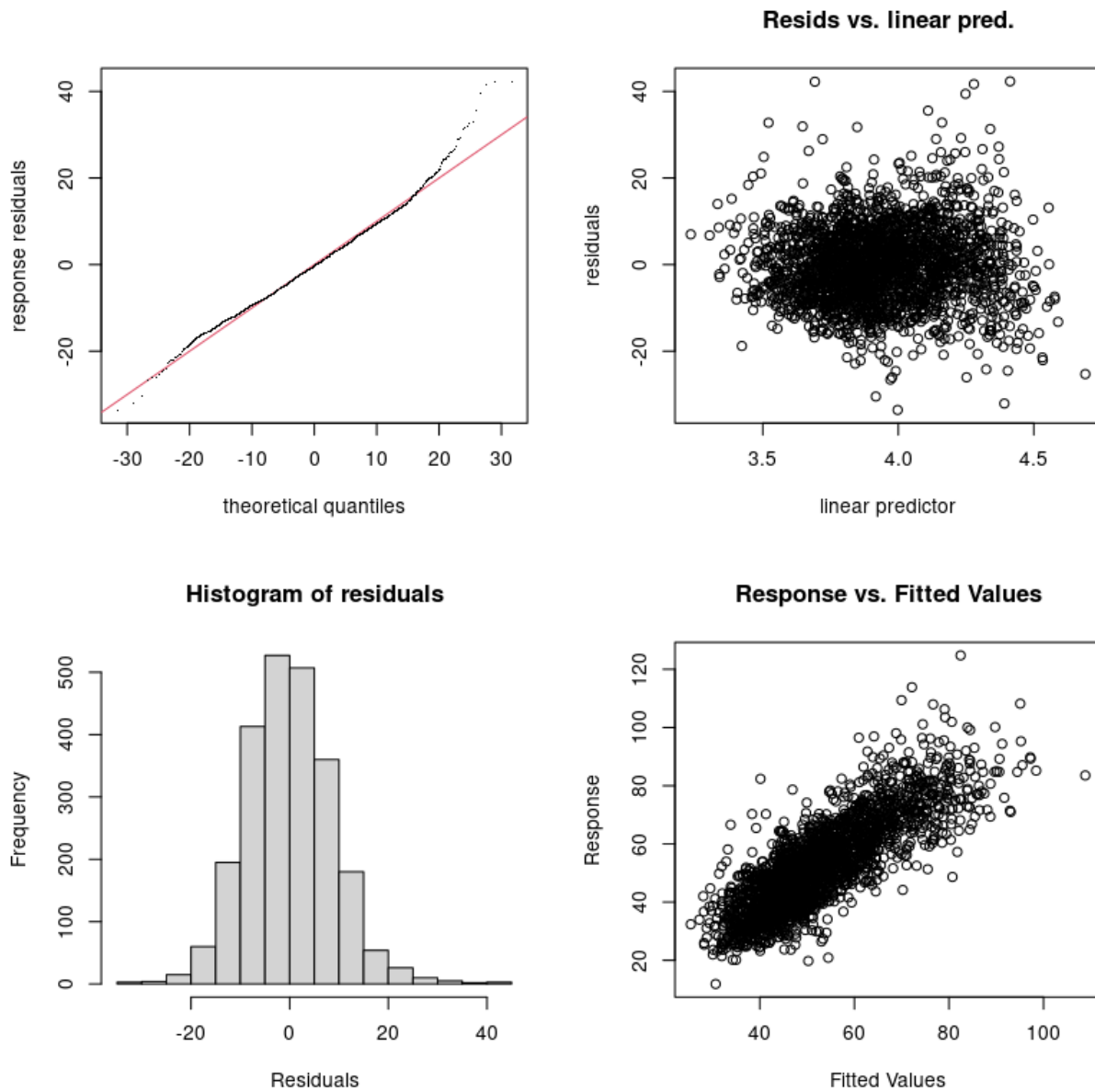
Order Number 582-22-31570, hereafter referred to simply as the TCEQ Report. Odd number figures are the smooth functions from the GAM fit of the natural logarithm of the total MDA8 O<sub>3</sub> values to the meteorological predictors with the 95% confidence intervals shown in red in the smooth functions. The day of year (*day*) function is also shown. There are no noticeable differences in these figures compared to corresponding ones in the TCEQ Report and there are no differences in the statistical significance levels of the meteorological predictors. The even number figures are the standard GAM evaluation plots (made with the *gam.check* function in the R *mgcv* package). As with the smooth function plots, no differences with those of the TCEQ Report are noticeable. The only notable difference between the GAMs is that the inclusion of the smoke flag increases the percentage of deviance explained by 0.2 % in Houston and 0.3 % in El Paso.

**Table 1.** Performance of GAMs with and without smoke for Total MDA3 O<sub>3</sub>.

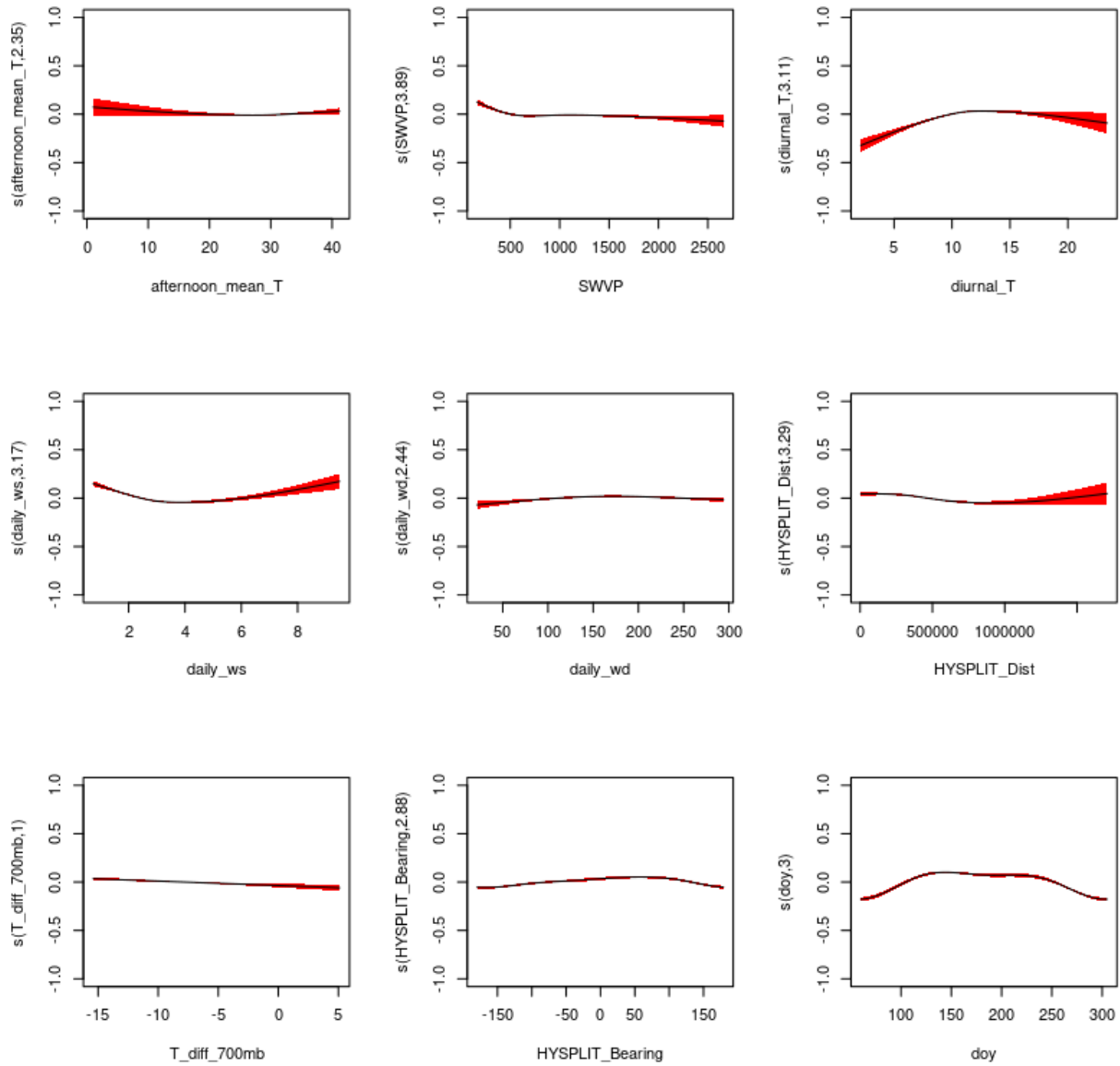
<b>Urban Area / Experiment</b>	<b>Figure for Smooth Functions</b>	<b>Figure for GAM Evaluation</b>	<b>% Deviance Explained</b>	<b>Significance of Meteorological Predictors</b>
<b>Houston no smoke</b>	TCEQ Report	TCEQ Report	67.0	8 terms at $\alpha=0.001$
<b>Houston smoke flag</b>	1	2	67.2	8 terms at $\alpha=0.001$
<b>El Paso no smoke</b>	TCEQ Report	TCEQ Report	53.9	7 terms at $\alpha=0.001$ 1 term at $\alpha=0.05$
<b>El Paso smoke flag</b>	3	4	54.2	7 terms at $\alpha=0.001$ 1 term at $\alpha=0.05$



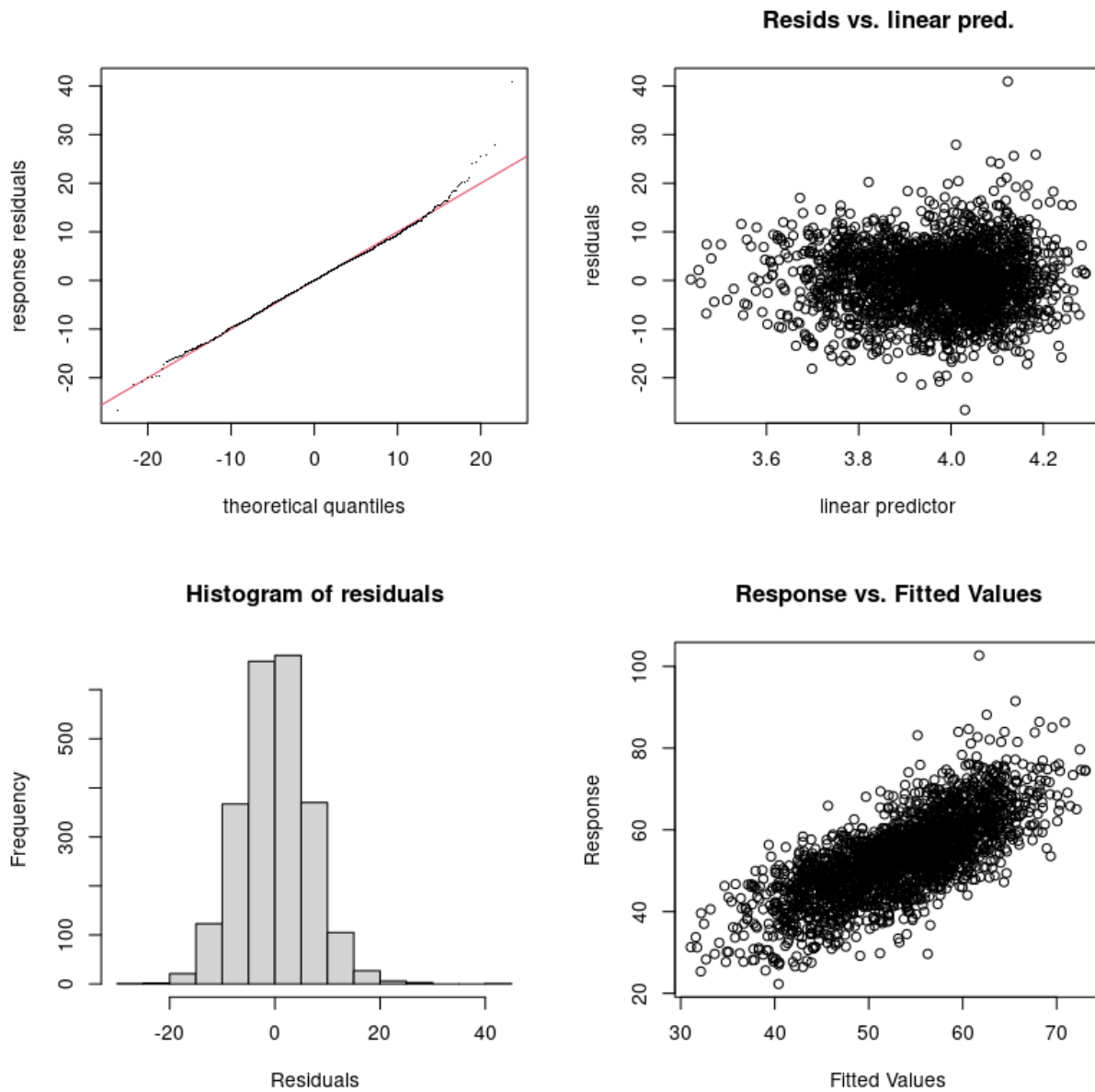
**Figure 1.** Smooth functions for the total MDA8 O<sub>3</sub> GAM fit with the binary smoke flag as a factor predictor in Houston, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the MDA8 O<sub>3</sub> in ppbv from its mean value.



**Figure 2.** Evaluation plots of the GAM with the binary smoke flag for total MDA8 O<sub>3</sub> in Houston, TX.



**Figure 3.** Smooth functions for the total MDA8 O3 GAM fit with the binary smoke flag as a factor predictor in El Paso, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the MDA8 O3 in ppbv from its mean value.



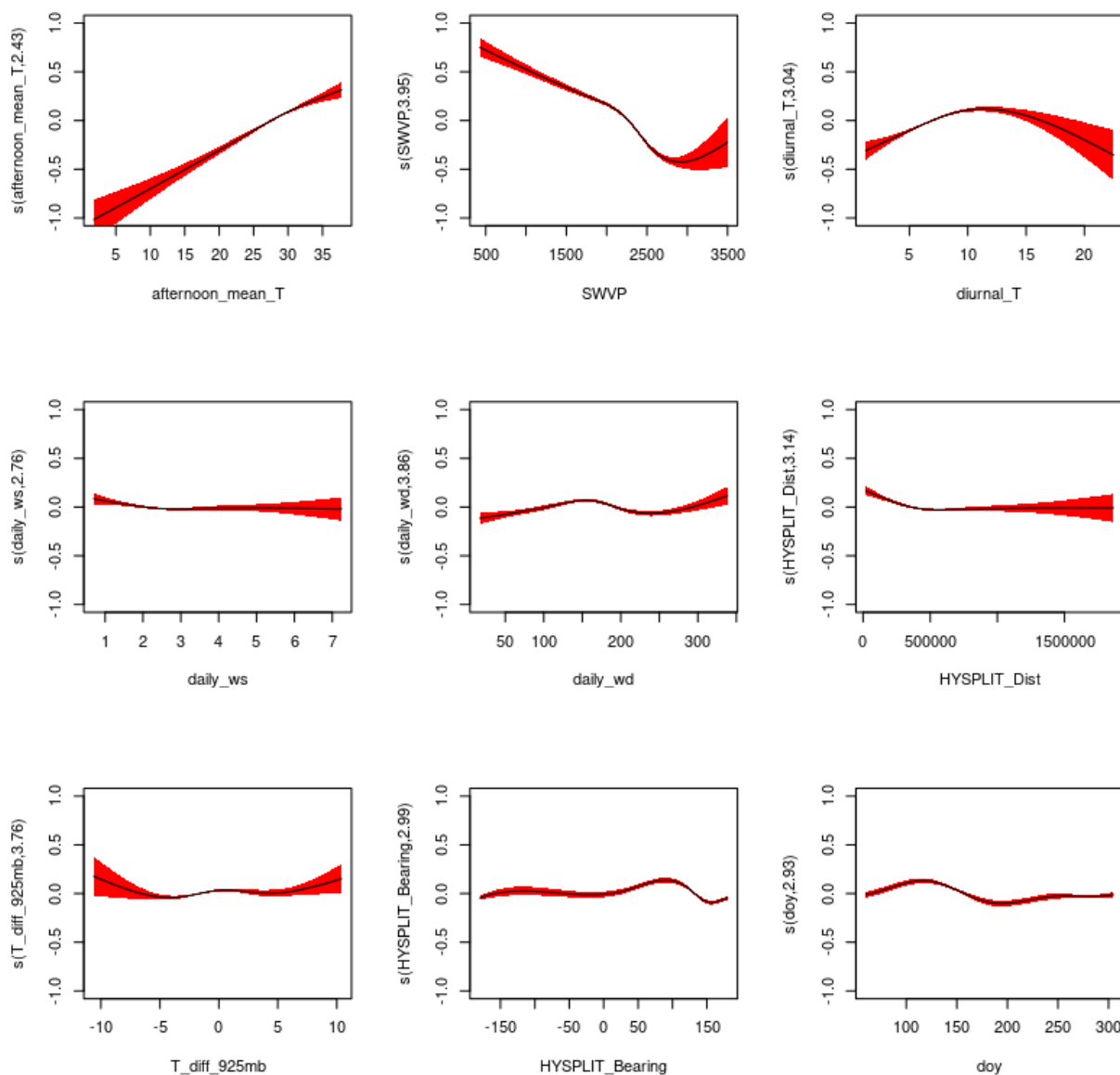
**Figure 4.** Evaluation plots of the GAM with the binary smoke flag for total MDA8 O<sub>3</sub> in El Paso, TX.

### ***Background MDA6 O3 GAM Results***

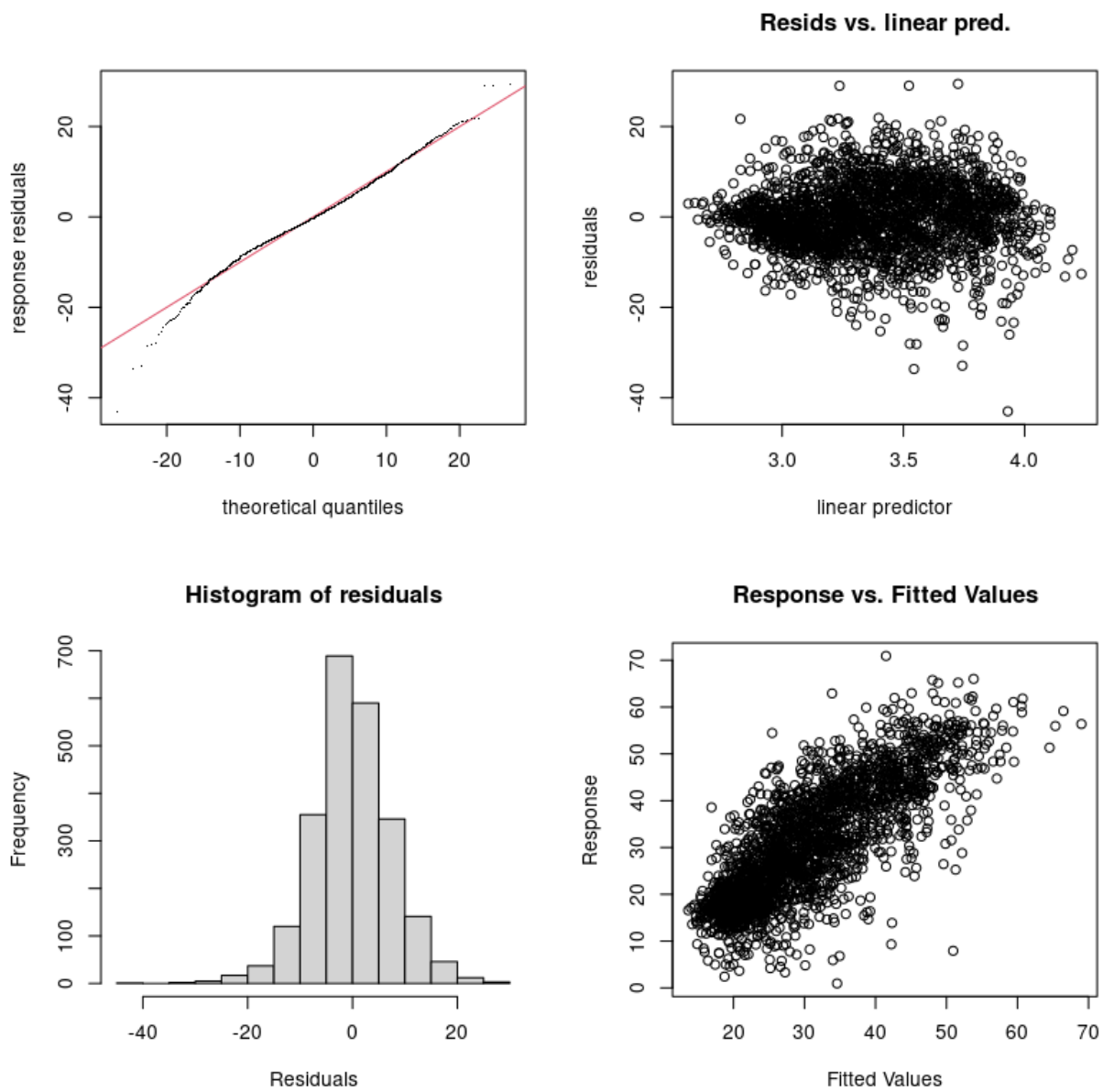
Table 2 summarizes the background MDA8 O<sub>3</sub> GAM results for Houston and El Paso both without and with the binary smoke flag. Table 2 is organized as in Table 1. As with the total MDA8 O<sub>3</sub>, differences in the figures with those of the TCEQ Report are not evident. However, the smoke flag improved the significance level of one of the meteorological predictors at Houston. In addition, the percentage of deviance explained improved by 0.3 % in Houston and 0.5 % in El Paso.

**Table 2.** Performance of GAMs with and without smoke for Background MDA3 O<sub>3</sub>.

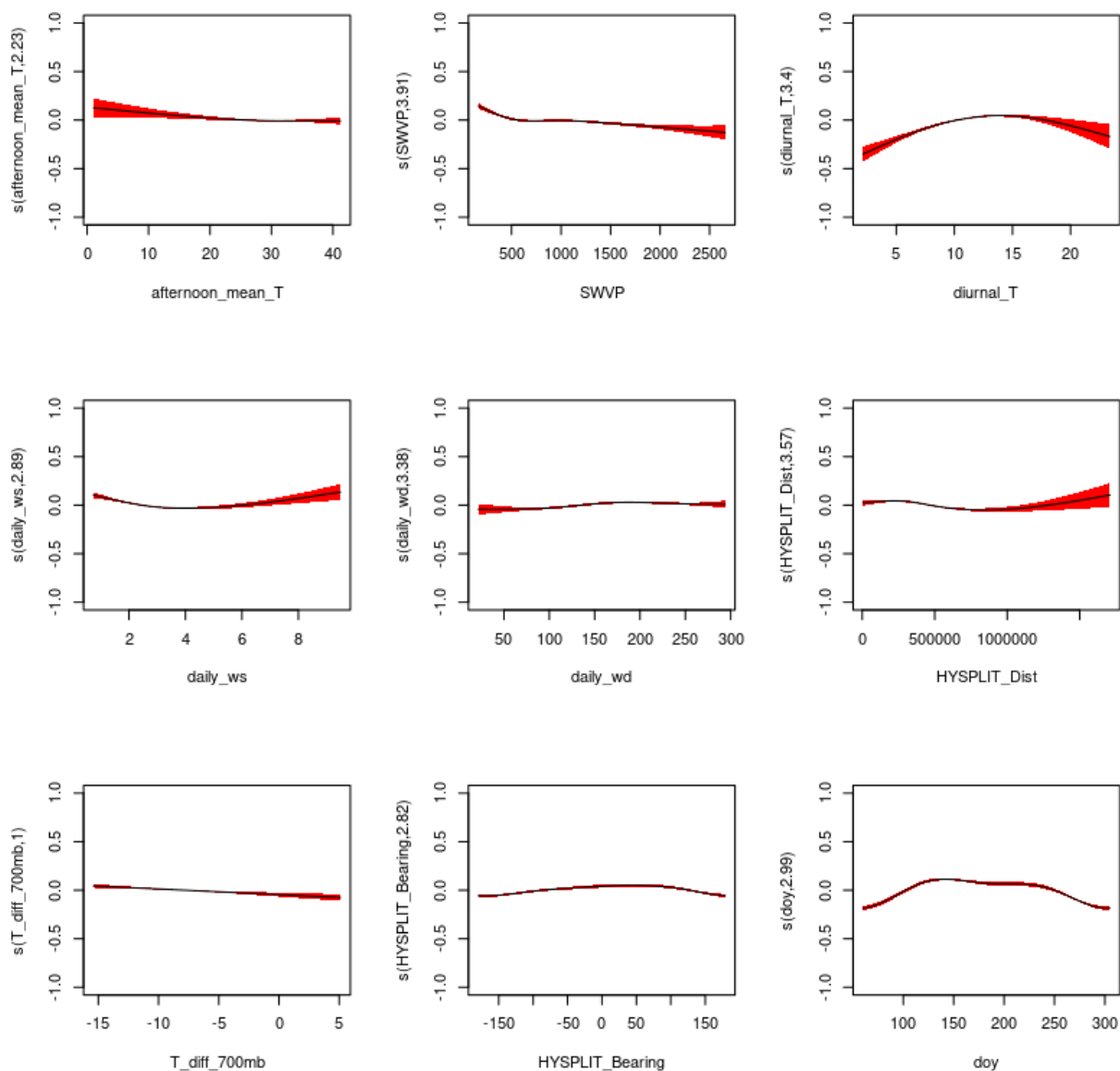
<b>Urban Area / Experiment</b>	<b>Figure for Smooth Functions</b>	<b>Figure for GAM Evaluation</b>	<b>% Deviance Explained</b>	<b>Significance of Meteorological Predictors</b>
<b>Houston no smoke</b>	TCEQ Report	TCEQ Report	63.4	7 terms at $\alpha=0.001$ 1 term at $\alpha=0.01$
<b>Houston smoke flag</b>	5	6	63.7	8 terms at $\alpha=0.001$
<b>El Paso no smoke</b>	TCEQ Report	TCEQ Report	51.0	7 terms at $\alpha=0.001$ 1 term at $\alpha=0.05$
<b>El Paso smoke flag</b>	7	8	51.5	7 terms at $\alpha=0.001$ 1 term at $\alpha=0.05$



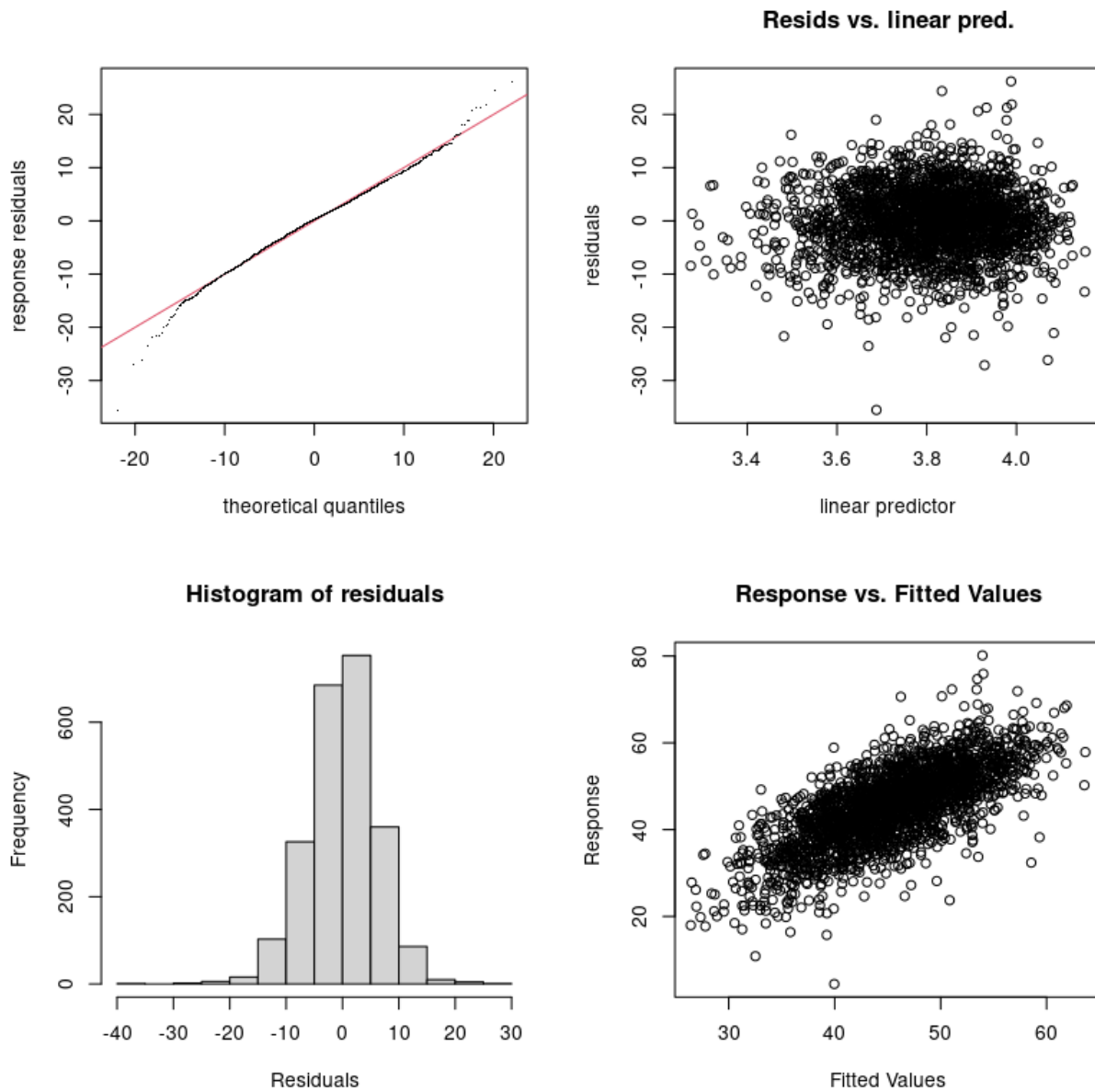
**Figure 5.** Smooth functions for the background MDA8 O<sub>3</sub> GAM fit with the binary smoke flag as a factor predictor in Houston, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the MDA8 O<sub>3</sub> in ppbv from its mean value.



**Figure 6.** Evaluation plots of the GAM with the binary smoke flag for background MDA8 O<sub>3</sub> in Houston, TX.



**Figure 7.** Smooth functions for the background MDA8 O3 GAM fit with the binary smoke flag as a factor predictor in El Paso, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the MDA8 O3 in ppbv from its mean value.



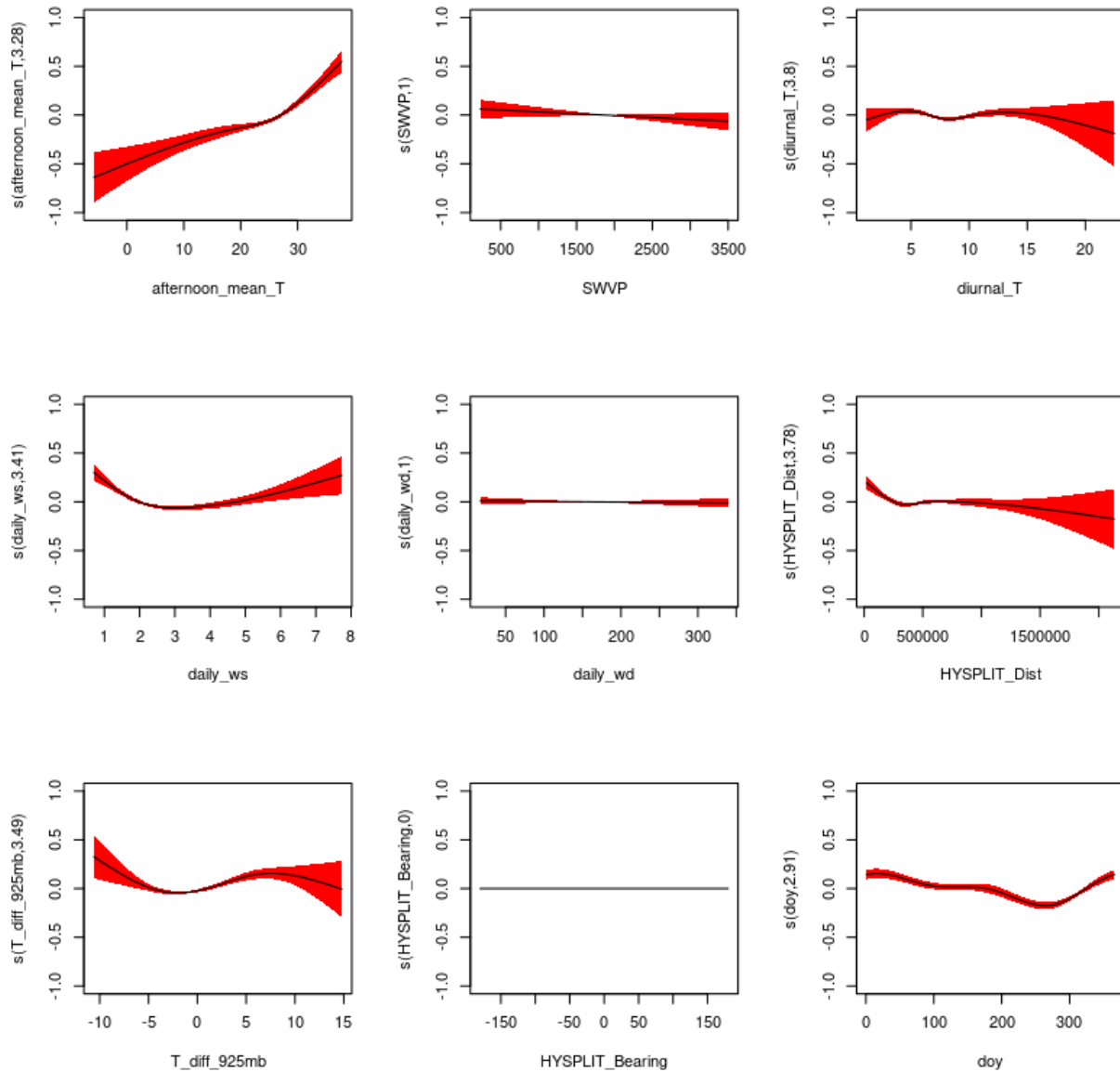
**Figure 8.** Evaluation plots of the GAM with the binary smoke flag for background MDA8 O<sub>3</sub> in El Paso TX.

**Total PM<sub>2.5</sub> GAM Results**

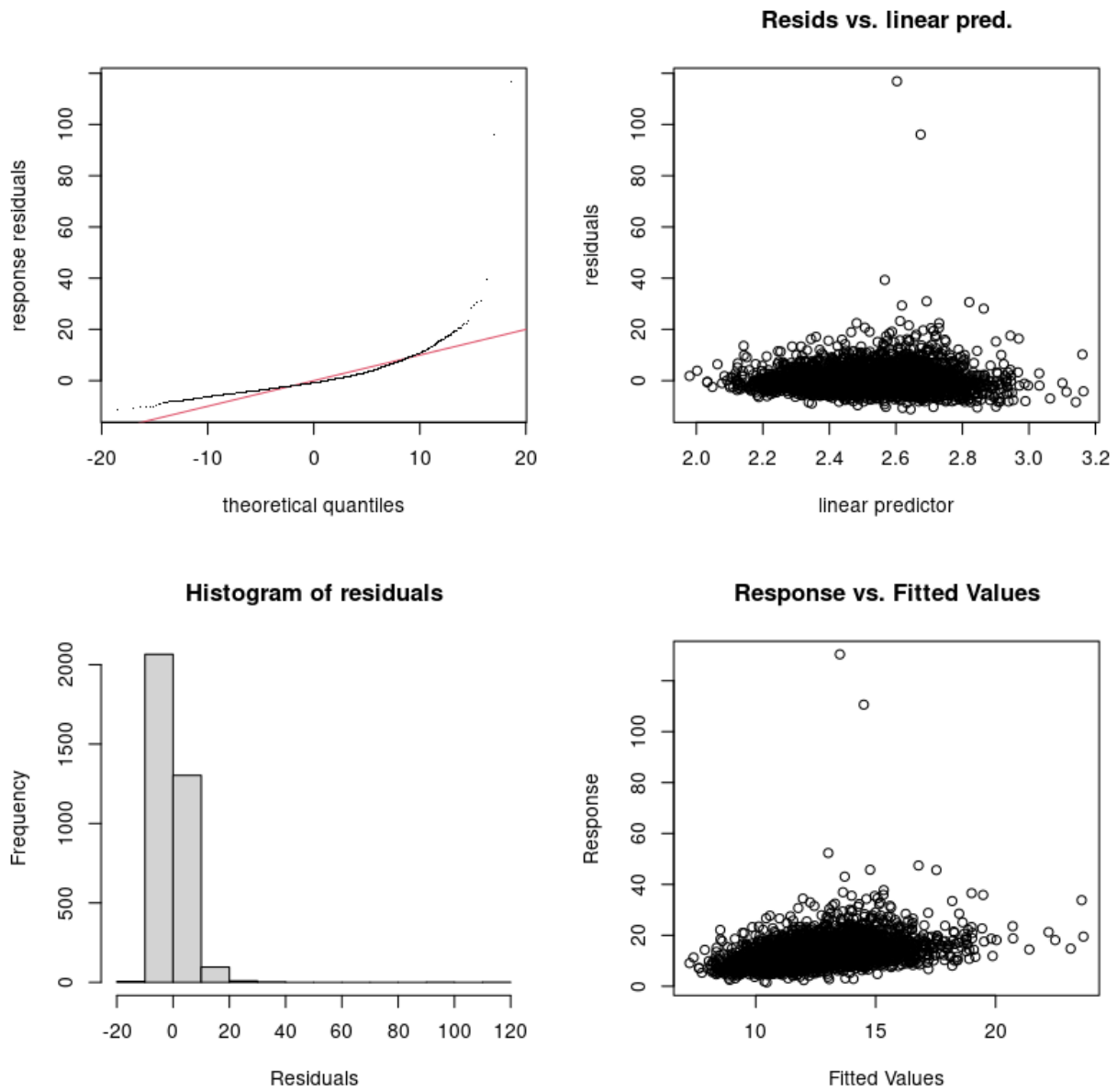
Table 3 organizes the total daily average PM<sub>2.5</sub> GAM results for Houston and El Paso as in Tables 1 and 2. For this case there are minor differences noticeable in the smooth function and GAM evaluation plots. Most notable are the narrowing of the 95 % confidence interval in the doy plots and some minor changes to the scatter in the Response vs. Fitted interval plots compared to corresponding plots in the TCEQ report. In addition, Table 3 indicates some shifting of the significance levels of the meteorological predictors. The percentage of explained deviance increases by 0.4 % in Houston and 0.9 % in El Paso.

**Table 3.** Performance of GAMs with and without smoke for total PM<sub>2.5</sub>.

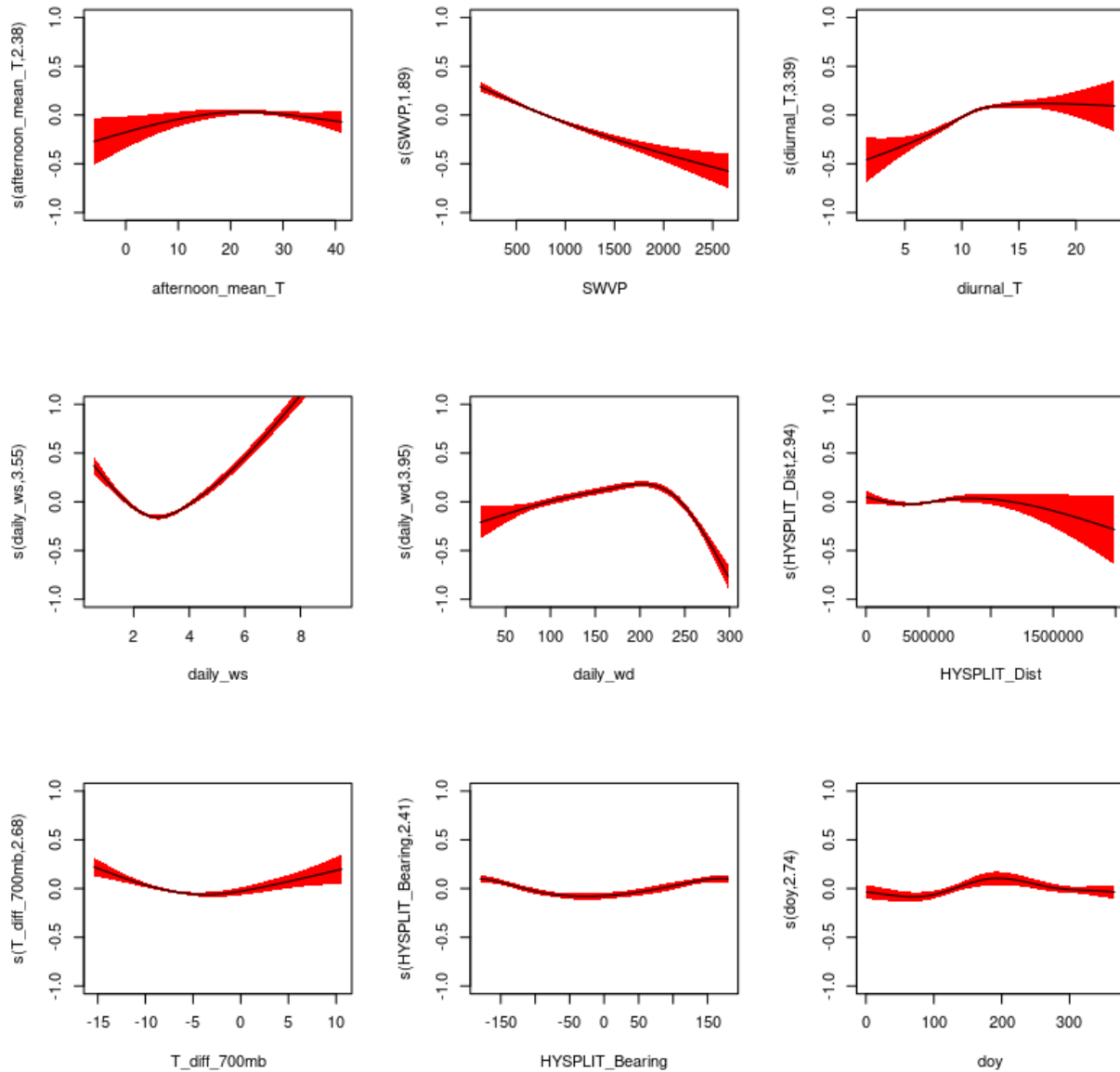
<b>Urban Area / Experiment</b>	<b>Figure for Smooth Functions</b>	<b>Figure for GAM Evaluation</b>	<b>% Deviance Explained</b>	<b>Significance of Meteorological Predictors</b>
<b>Houston no smoke</b>	TCEQ Report	TCEQ Report	14.7	4 terms at $\alpha=0.001$ 1 term at $\alpha=0.01$ 1 term at $\alpha=0.1$ 2 terms at $\alpha=1$
<b>Houston smoke flag</b>	9	10	15.1	4 terms at $\alpha=0.001$ 1 term at $\alpha=0.01$ 3 terms at $\alpha=1$
<b>El Paso no smoke</b>	TCEQ Report	TCEQ Report	32.0	6 terms at $\alpha=0.001$ 1 term at $\alpha=0.05$ 1 term at $\alpha=0.1$
<b>El Paso smoke flag</b>	11	12	32.9	6 terms at $\alpha=0.001$ 2 terms at $\alpha=0.05$



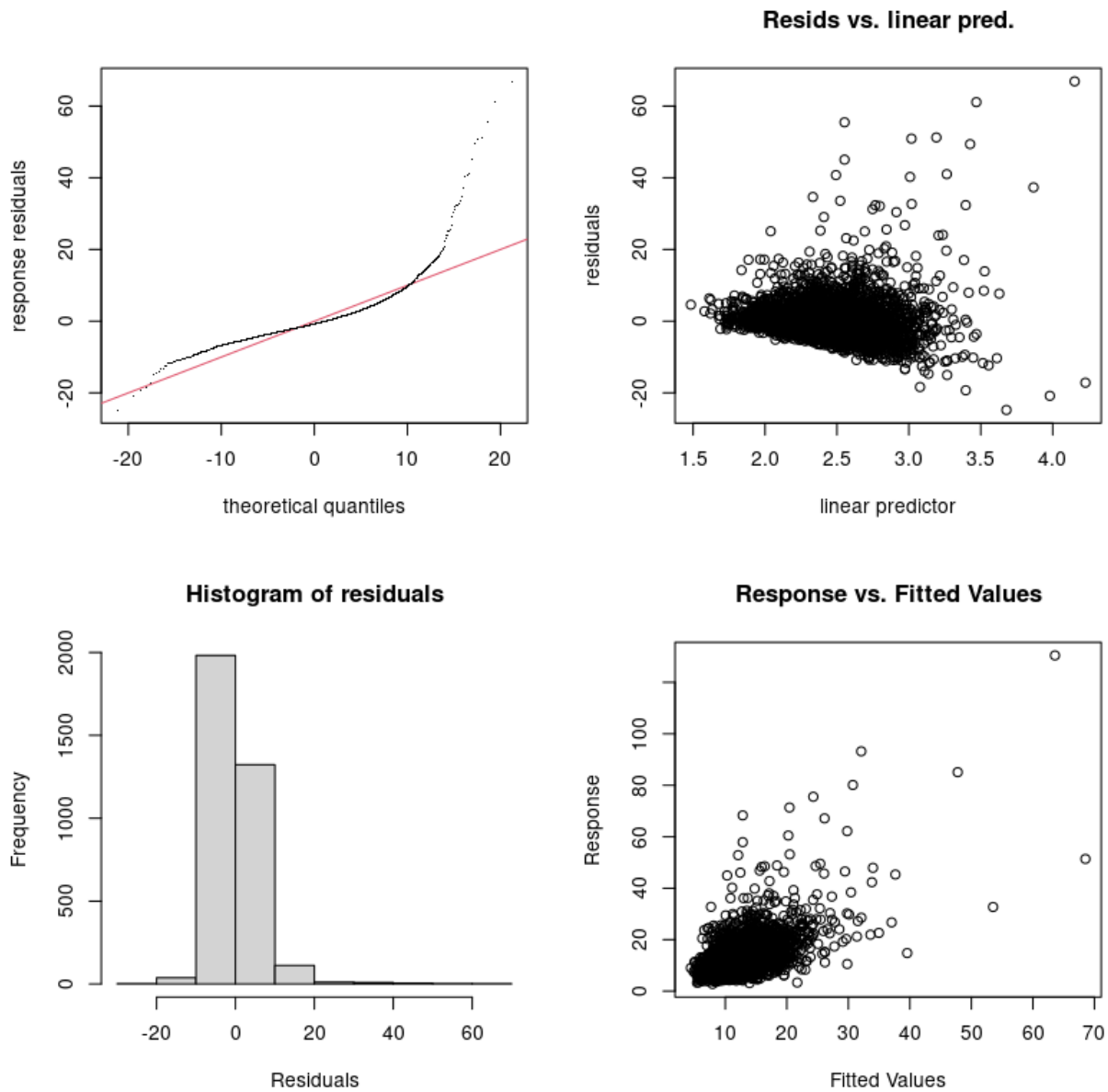
**Figure 9.** Smooth functions for the total daily average  $PM_{2.5}$  GAM fit with the binary smoke flag as a factor predictor for Houston, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the total daily average  $PM_{2.5}$  in  $\mu g/m^3$  from its mean value.



**Figure 10.** Evaluation plots of the GAM with the binary smoke flag for total daily average PM<sub>2.5</sub> in Houston, TX.



**Figure 11.** Smooth functions for the total daily average PM<sub>2.5</sub> GAM fit with the binary smoke flag as a factor predictor for El Paso, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the total daily average PM<sub>2.5</sub> in  $\mu\text{g}/\text{m}^3$  from its mean value.



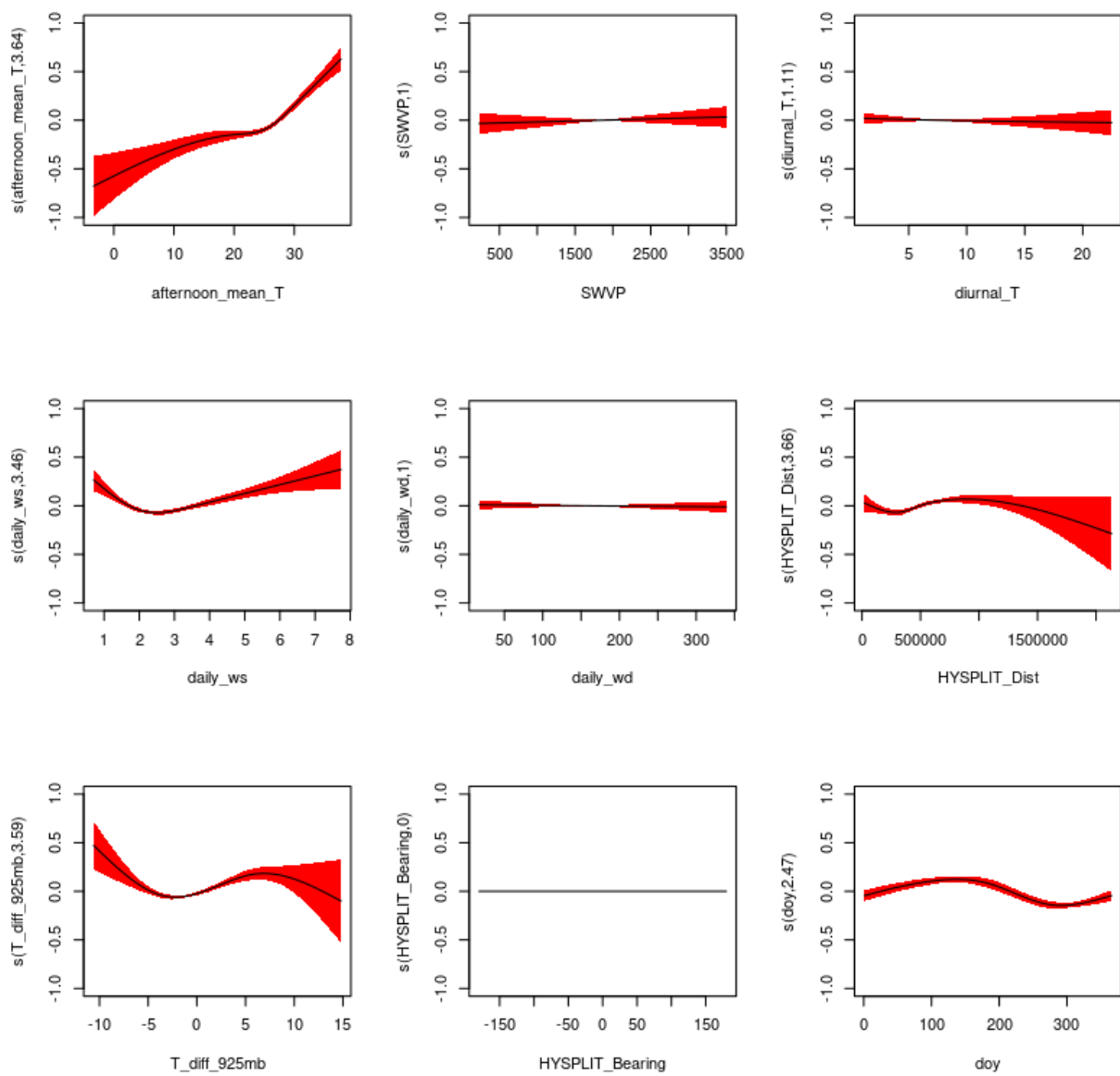
**Figure 12.** Evaluation plots of the GAM with the binary smoke flag for total daily average PM<sub>2.5</sub> in El Paso, TX.

**Background PM<sub>2.5</sub> GAM Results**

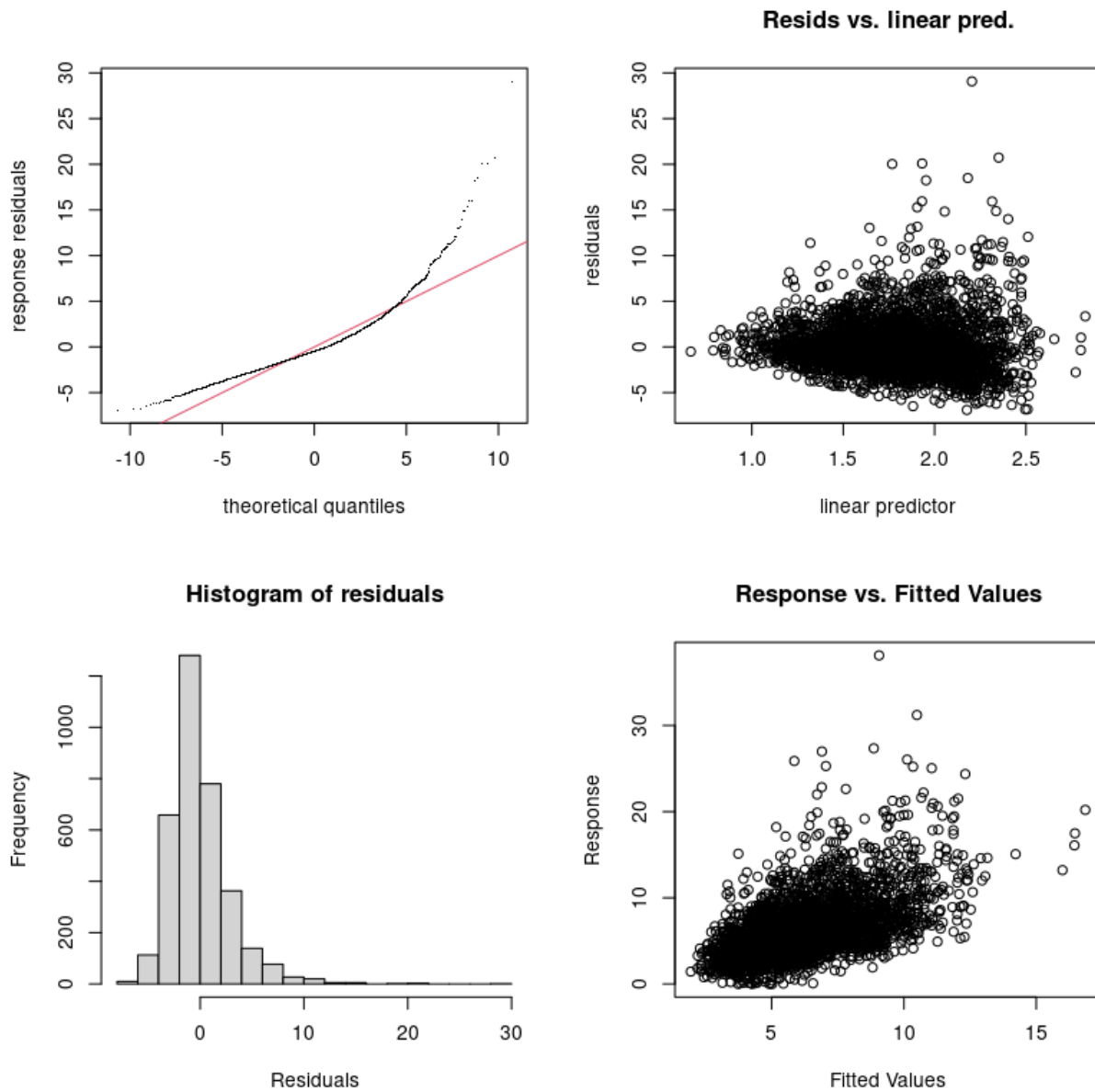
Table 4 organizes the background PM<sub>2.5</sub> GAM results for Houston and El Paso as in Tables 1, 2 and 3. Similar to the total daily average PM<sub>2.5</sub> case there are minor differences noticeable in the smooth function and GAM evaluation plots. Most notable are the narrowing of the 95 % confidence interval in the doy plots and some minor changes to the scatter in the Response vs. Fitted interval plots compared to corresponding plots in the TCEQ report. In addition, Table 4 indicates some shifting of the significance levels of the meteorological predictors. The percentage of explained deviance increases by 0.4 % in Houston and 0.9 % in El Paso.

**Table 4.** Performance of GAMs with and without smoke for background PM<sub>2.5</sub>.

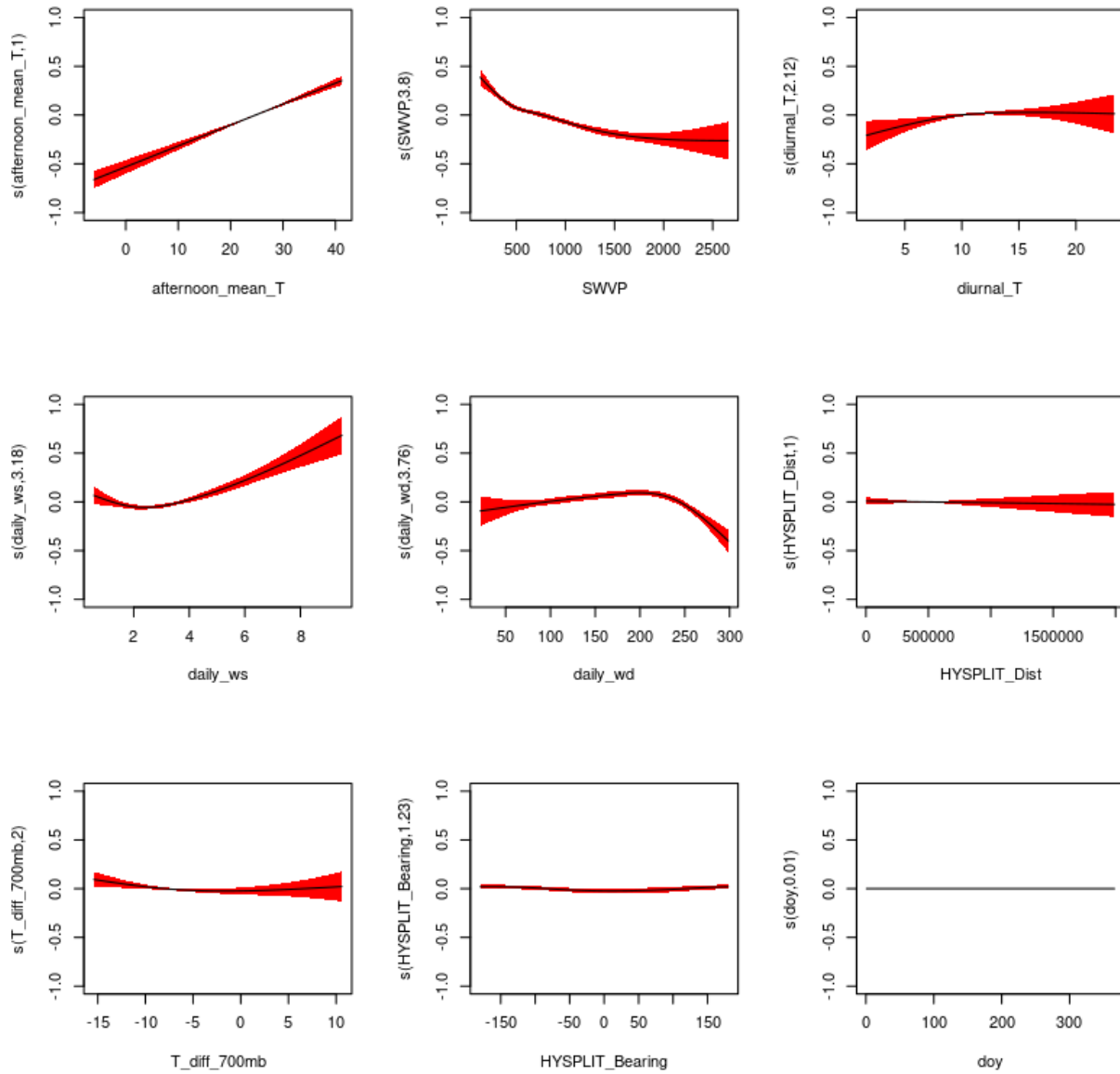
<b>Urban Area / Experiment</b>	<b>Figure for Smooth Functions</b>	<b>Figure for GAM Evaluation</b>	<b>% Deviance Explained</b>	<b>Significance of Meteorological Predictors</b>
<b>Houston no smoke</b>	TCEQ Report	TCEQ Report	31.6	4 terms at $\alpha=0.001$ 4 terms at $\alpha=1$
<b>Houston smoke flag</b>	13	14	32.0	4 terms at $\alpha=0.001$ 4 terms at $\alpha=1$
<b>El Paso no smoke</b>	TCEQ Report	TCEQ Report	21.9	4 terms at $\alpha=0.001$ 2 terms at $\alpha=0.05$ 2 terms at $\alpha=1$
<b>El Paso smoke flag</b>	15	16	22.8	4 terms at $\alpha=0.001$ 2 terms at $\alpha=0.05$ 1 term at $\alpha=0.1$ 1 terms at $\alpha=1$



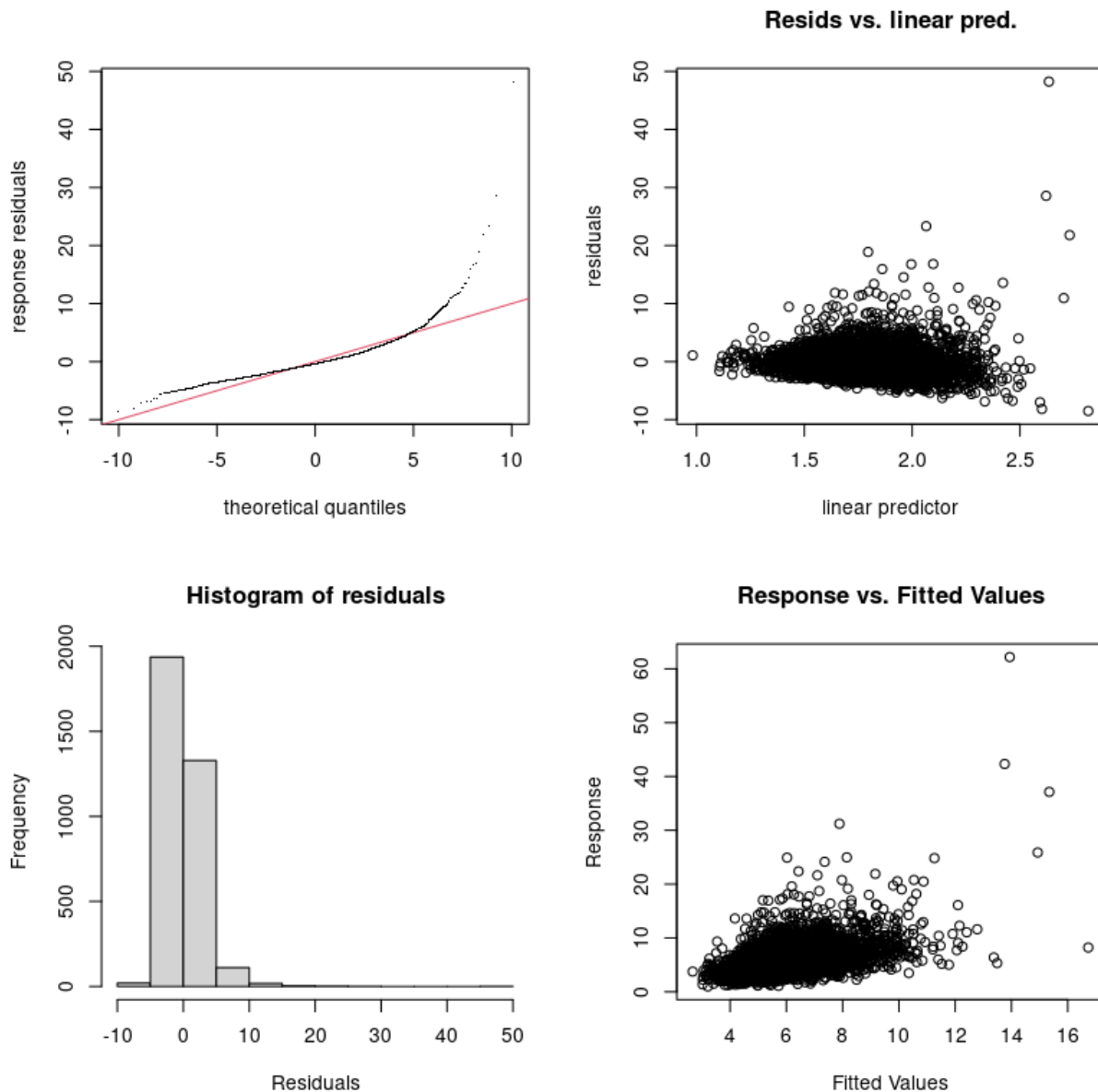
**Figure 13.** Smooth functions for background PM<sub>2.5</sub> GAM fit with the binary smoke flag as a factor predictor for Houston, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the total daily average PM<sub>2.5</sub> in µg/m<sup>3</sup> from its mean value.



**Figure 14.** Evaluation plots of the GAM with the binary smoke flag for background PM<sub>2.5</sub> in Houston, TX.



**Figure 15.** Smooth functions for background PM<sub>2.5</sub> GAM fit with the binary smoke flag as a factor predictor for El Paso, TX. The y-axis scale is the scale of the “linear predictor”, i.e., the deviation of the natural logarithm of the total daily average PM<sub>2.5</sub> in  $\mu\text{g}/\text{m}^3$  from its mean value.



**Figure 16.** Evaluation plots of the GAM with the binary smoke flag for background PM<sub>2.5</sub> in El Paso, TX

### Data Collected

None to date.

### Identify Any Problems or Issues Encountered and Proposed Solutions or Adjustments

The HMS to GOES scripts have many python package dependencies. We discovered that it is critical to start with a “clean” conda environment and install all the necessary packages at once for the installation to work properly. It is also recommended to install all of the packages from the same repository to avoid incompatibilities. We have documented this process and will include this information with the code delivered to TCEQ upon completion of this project.

## **Goals and Anticipated Issues for the Succeeding Reporting Period**

During this next reporting period we add the HMS fire counts in surrounding regions binned by distance from the city to the GAM input datasets. We will also work to include FINN NO<sub>x</sub>, VOC, and PM<sub>2.5</sub> emissions multiplied by the daily surface “footprints” to the GAM input datasets.

## **Detailed Analysis of the Progress of the Task Order to Date**

Completed initial drafts of Work Plan and QAPP. Held project kickoff. Started work on STILT analyses and gathering data for GAMM studies. Wrote and tested code to determine whether smoke occurred at a monitor location on a given day from the HMS smoke polygons. Performed initial GAM analysis.

**Do you have any publications related to this project currently under development? If so, please provide a working title, and the journals you plan to submit to.**

Yes       No

**Do you have any publications related to this project currently under review by a journal? If so, what is the working title and the journal name? Have you sent a copy of the article to your AQRP Project Manager and your TCEQ Liaison?**

Yes       No

**Do you have any bibliographic publications (ie: publications that cite the project) related to this project that have been published? If so, please list the reference information. List all items for the lifetime of the project.**

Yes       No

**Do you have any presentations related to this project currently under development? If so, please provide working title, and the conference you plan to present it (this does not include presentations for the AQRP Workshop).**

Yes       No

**Do you have any presentations related to this project that have been published? If so, please list reference information. List all items for the lifetime of the project.**

Yes       No

**Have any personnel changes occurred that were not listed in the original proposal? If so, please include a detailed description of the personnel change(s) below.**

Yes       No

**Are any delays expected in the progress of the research? If so, please include a detailed description of the potential delay below.**

Yes       No

**Describe any possible concerns/issues (technical or non-technical) that AQRP should be made aware of.**

None

**Are you anticipating using all the available funds allocated to this project by the end date? If not, why and approximately what is the amount to be returned?**

Yes       No

---

Submitted to AQRP by  
Matthew J. Alvarado