Description of data and code files for “*Beyond particulate matter mass: heightened levels of lead and other pollutants associated with destructive fire events in California”*

For our analysis, we utilized data from the EPA AQS database (<https://aqs.epa.gov/aqsweb/airdata/download_files.html>) and the NOAA Hazard Mapping System database (<https://www.ospo.noaa.gov/Products/land/hms.html#data>). From the AQS website, we manually downloaded the “aqs\_monitors.csv”, “daily\_88101\_year.zip”, and “daily\_SPEC\_year.zip” files for the years 2006-2018. From the HMS website, we manually downloaded the HMS SMOKE shapefiles from 01/01/2006 to 12/31/2018. The following are code files in Python and R to prepare and analyze the data.

**Python and R Code:**

*naaqs\_fires\_analysis\_byron.py*

Defines data preparation functions for AQS data, smoke plume data, and fire location data as well as GIS functions used in analysis. This file does not need to be run; it is imported to define functions in subsequent code files (files 1-4; note: files are numbered and organized in the general order performed).

*1. Run smoke analysis on 2006-2018 AQS datasets california speciation only 4-9-2021.py*

Performs spatial link of AQS monitor locations to HMS satellite-detected smoke plumes. Imports the raw data for AQS monitor locations and the raw HMS smoke polygon data. Exports AQS monitor locations with a Boolean indicator of whether they fall within a smoke plume (intermediate file: california AQS speciation monitors linked to HMS 2006-2018.csv).

*2. Run fire analysis on 2006-2018 AQS datasets california speciation only 5-17-2021.py*

Calculates the distance to the nearest fire on each day for each unique AQS monitor location. Exports AQS monitor locations and the distance to the nearest satellite-detected fire on each day (intermediate file: california AQS speciation monitors distance to closest fire 2006-2018.csv).

*3. Join HMS analysis data to PM Spec CA data 2006-2018 BR 5-18-2021.py*

Adds outputs from 1. (boolean indicator of overhead smoke plume) and 2. (distance to closest fire location) to AQS concentration data. Exports AQS data with joined fields (intermediate file: 2006-2018 HMS smoke and fire analysis PM2.5 SPEC CA 11-3-2021.csv).

*4. Quantile Regression of PM SPEC CA 2006-2018 distance to closest HMS-detected fire.py*

Perform quantile regression analysis of AQS monitor concentrations with respect to distance to the nearest fire for all monitoring days with an overhead smoke plume. Export quantitative results and descriptive figure.

*pm\_linking\_final.py*

This code imports the ‘2006-2018 HMS smoke and fire analysis PM2.5 SPEC CA 11-3-2021.csv’ file, which is output from *3. Join HMS analysis data to PM Spec CA data 2006-2018 BR 5-18-2021.py*, and links the data to the concentration of PM2.5 at each station on the same day. Exports the file: 2006-2018\_pm\_speciation\_labeled\_joined\_pm\_11\_3\_21.csv.

*functions.py*

This code contains two functions: **count** and **stations**. The first function, count, is used to count the number of measurements of zero on smoke and non-smoke days as well as the percentage of smoke days and percentage of non-zero smoke days. The second function, stations, calculates the following: a) the mean across all stations, for each element, each year, b) the mean at each station, for each element, each year on smoke and non-smoke days, c) the percentage difference between smoke and non-smoke averages across each element, station, and year, d) the deviation of smoke days from non-smoke days across each element, station, and year, e) the concentration difference between smoke and non-smoke averages across each element, station, and year, and f) the average values of c) d) and e) across all stations for each year. This file does not need to be run; it is imported to define functions in subsequent code files (5-7).

*5. Code for table analysis.py*

This code creates the following files: '2006-2018\_pm\_speciation\_labeled\_new\_11\_3\_21.csv', 'non\_zero\_values\_11\_3.csv', and 'differences\_for\_table\_1.csv'. The first file, '2006-2018\_pm\_speciation\_labeled\_new\_11\_3\_21.csv', is the combined smoke and non-smoke list, with labels, for permutation analysis in R. The second file, 'non\_zero\_values\_11\_3.csv', contains the data found in **Supplementary Table 3** with number of measurements, including measurements of zero, number of total days, and % of smoke days. The third file, 'differences\_for\_table\_1.csv', contains the data found in **Supplementary** **Table 1** with median and maximum percentage change and maximum concentration difference. The code also finds the potassium:lead ratio found in **Supplementary Figure 5**.

*6. Code for figures 1, 2, 3, and 5.py*

This code creates **Figures 1b** (map of measurement stations), **2** (percent differences for all stations, all years, all chemicals), **3** (absolute and percentage concentration differences for all stations, year-by-year), and **Supplementary Figures 2** (same as Figure 3) and **4** (linear regression of bromine and sodium and PM2.5).

*7. Code for figures 6 and 7.py*

This code creates **Figures 4 and 5** (chemical profiles for different chemical breakdowns during multiple fire events).

*permutation\_test\_shuffling.R*

This code shuffles 10,000 permutations for each element. It shuffles the indices for each element, each year. It creates a file for each element called ‘Element\_updated\_Data\_2006-2018\_Permute\_Data\_11\_15.Rdata’.

*permutation\_test\_run.R*

This code uses the permuted indices to shuffle the imported data at each site. It calculates the p-value for the year by counting how many of the permuted samples had a difference value that was greater than or equal to the difference value in the original dataset. It then creates a boxplot of the concentrations, with p-value at the bottom for each year. This information is used in **Supplementary Tables 1 and 2** (permutation test results and summary of results). The code needs to be run separately for each element, switching out the element name and information for placement of p-value on the boxplot.