# Description of data and code files for “Wildfires increase concentrations of hazardous air pollutants in downwind communities.”

In the primary analysis, we analyzed pre-generated data files of daily air pollutant measurements from the US EPA Air Quality System (AQS) database (<https://aqs.epa.gov/aqsweb/airdata/download_files.html>) and smoke plume extents from the NOAA Hazard Mapping System (HMS) 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”, “daily\_HAPS\_year.zip”, and “daily\_SPEC\_year.zip” files for the years 2006-2020. From the HMS website, we manually downloaded the HMS SMOKE shapefiles from 01/01/2006 to 12/31/2020. Burned area extents for select fires in Figure 2 were obtained from the Monitoring Trends in Burn Severity website (<https://www.mtbs.gov/direct-download>). Descriptions of the Python and R scripts used for the analysis in this study follow below. They can be run in the following order to reproduce the results of this study.

# Python and R Code:

## 1. AQS monitor and HMS smoke spatial link.py

This script combines each of the input PM2.5 speciation and HAPs AQS files together and labels smoke impact using the HMS shapefiles.

## 2. join\_PM2.5\_with\_haps-nutrients-combined\_dataset.ipynb

This script adds total PM2.5 to the project dataset. The final project dataset is produced, named “2006-2020\_combined\_linkedtoplumes\_data\_2\_15\_pm2.5\_added\_6-14-2022.csv”.

## 3. HAPs unit\_conversion.ipynb

This script converts gaseous species’ units from ppbC to μg/m3. The output of this script is analyzed in the following scripts. A copy of the dataset (with a different shape) is also generated as an input to the permutation test.

## 4. Maps.ipynb

This script generates the map figures, Figure 1 and Figure S1.

## 5. HAPs station-specific mean differences.ipynb

This script calculates the year- and station-specific concentration differences between smoke and non-smoke days. Figure 3 and data for Table 1 (excluding the permutation test) are generated.

## 6. permutation test main.r

The permutation tests included in this study are performed in this script.

## 7. permutation test summary.ipynb

Permutation test results are aggregated in this file. Figure 4 and Table S4 are generated.

## 8. HAPs PM2.5 correlation.ipynb

Correlation between HAPS and total PM2.5 on smoke and non-smoke days is calculated in this script. Figure S3 is generated.

## 9. san jose jackson time-series.ipynb

Measurements at the San Jose site are analyzed in this file. Figures 5, S4a, and S4b are generated.

## 10. health reference values.ipynb

Comparisons of HAPs concentrations to reference values for human health effects are performed in this script. Table 2, Figure 6, Figure 7, and Figure S5 are generated in this script.

## 11. table s1 - summary statistics.ipynb

Table S1, including summary statistics of each of the included HAPs, is created using this script.

## 12. table 1 - western us analysis results.ipynb

Table 1 is created using outputs from the permutation test and calculations of concentration differences in this script.

## 13. table s5 - max values on smoke vs non-smoke days.ipynb

Table S5 is generated using this script.

## 14. table s2 - observation count by year.ipynb

Table S2 is generated using this script.

## data.py

This file contains convenience functions for loading the project dataset. This file is not run directly, but the code in this file is imported into the other scripts.

## naaqs\_fires\_analysis.py

This script contains the code performing the GIS analysis required to label AQS monitoring days as smoke impacted or not. The functions in this script are imported into the script named “1. AQS monitor and HMS smoke spatial link.py”. This file is not run directly.