Description of data and code files for “*Wildfires in the western United States are mobilizing PM2.5-associated nutrients and may be contributing to downwind cyanobacteria blooms”*

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 2006-2020. From the HMS website, we manually downloaded the HMS SMOKE shapefiles from 01/01/2006 to 12/31/2020. Satellite cyanobacteria imagery and time series csv files were downloaded from the California Harmful Algal Blooms (HABs) Portal (<https://fhab.sfei.org/>) for each lake of interest during the dates listed in the manuscript. The following are code files in Python and R to prepare and analyze the data.

**Python and R Code:**

*naaqs\_fires\_analysis.py*

Defines data preparation functions for AQS data and smoke plume 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

*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.

*AQS monitor and HMS smoke spatial link.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: 2006-2020\_combined\_linkedtoplumes\_data\_2\_15.csv).

*monitor location map.py*

This code creates **Figures 1** (map of measurement stations).

*Nutrients.py*

This code creates **Figures 2 and S1** (percent and concentration differences of all species on smoke-impacted vs. non-smoke days), **Table S4** (average percent change on smoke days compared to non-smoke days for each species separated by year), **Figures 3 and S4** (chemical profiles for case study fires). This code contains data found in **Supplementary Table 1** (number of measurements, including measurements of zero, number of total days, and % of smoke days) and **Supplementary** **Table 3** (median and maximum percentage change and maximum concentration difference).

*permutation test main-nutrients.r*

 A permutation test of the null hypothesis that smoke and non-smoke observations originate from the same population is performed on the speciated PM2.5 measurements labeled by smoke impact status, both for all years combined and for individual years. This script outputs figures and tables, and the resulting information is used in **Supplementary Table 2** (permutation test results and summary of results).

*permutation\_test\_combo\_sampling\_combine\_years.r*

The function defining the permutation test for all years combined. This file does not need to be run; it is imported to define functions in subsequent code files

permutation\_test\_combo\_sampling\_year\_specific.r

The function defining the permutation test for individual years. This file does not need to be run; it is imported to define functions in subsequent code files

*Plot lake CI.R*

This code was used to plot monthly time series cyanobacteria index plots for **Figures 5, 6, and S6**. The code needs to be run separately for each lake or group of lakes plotted together, switching out the lake name when applicable.

*timeseries.py*

This code creates the annual cyanobacteria index plots for **Figures 5, 6, and 7.**