Documentation for Estimation of $PM_{2.5}$ in western US: Total and Attributed to Wildfires and Prescribed Fires

C.E. Reid[1], M.M. Maestas[1], E. Considine[1], G. Li[1],
N.H.F. French[2], M. Billmire[2], M. Jerrett[3]
[1]University of Colorado Boulder and [2]Michigan Technological University
and [3]University of California, Los Angeles

April 17, 2018

Abstract

The purpose of this document is to provide detailed information about the estimation of $PM_{2.5}$ (total and attributed to prescribed fires and wildfires) that our work could be reproduced. Figure 1 shows the study area of interest.

Contents

1 Data Sources for Machine Learning

For the creation of the spatiotemporal daily exposure surface via machine learning, a large number of data sets will be collected as discussed below. The dependent variable will be daily 24-hour $PM_{2.5}$ from monitoring data.

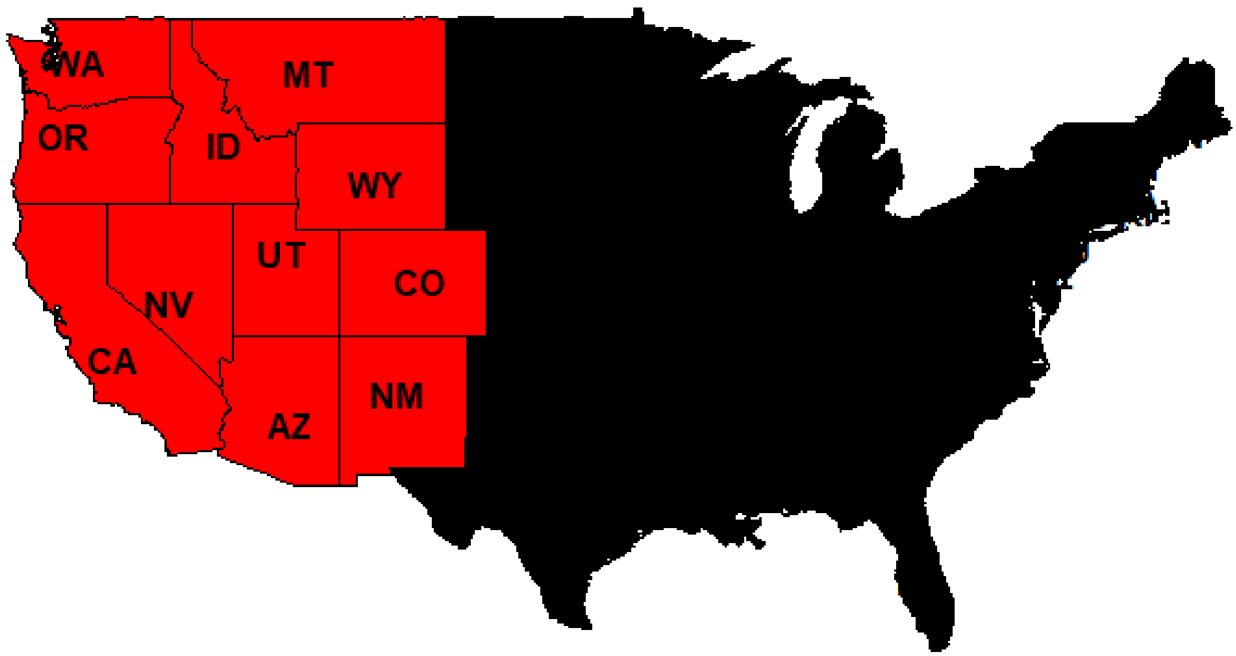

Figure 1: Map of 11-state study area.

1.1 All Monitor Locations

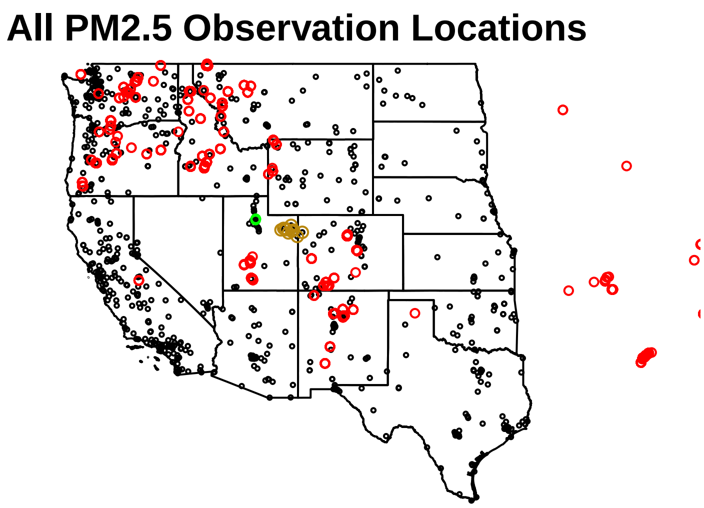

Figure 2: All Monitor Locations time series.

1.2 PM2.5 Monitor data from US EPA AQS Air Data Query Tool

Data Source

- **Contact**
  Can email the Air Quality Analysis Group (U.S. EPA Office of Air Quality Planning and Standards) on their website at https://www.epa.gov/outdoor-air-quality-data/forms/contact-us-about-outc
- **Citation/Link**
  United States Environmental Protection Agency. *Pre-Generated Data Files: Daily Summary Files, PM2.5 FRM/FEM Mass (88101) and PM2.5 non FRM/FEM Mass (88502), 2008-2014*. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent** 2008 through 2014
- **Acknowledgment**

Brief Description

We will download $PM_{2.5}$ data from both the US EPA AQS Air Data Query Tool (US EPA, 2017c) and the IMPROVE monitors that capture air quality information in more rural areas (US EPA, 2017e) for the 11-state region (Figure 1) including any of the following parameter codes: 88101, 88500, 88502, 81104 (US EPA, 2017a,d,f).

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1.

2.

Quality Control

Script Names

1.

Data File Names

1. daily_88101_2008.csv
2. daily_88101_2009.csv
3. daily_88101_2010.csv
4. daily_88101_2011.csv
5. daily_88101_2012.csv
6. daily_88101_2013.csv
7. daily_88101_2014.csv

8. daily_88502_2008.csv
9. daily_88502_2009.csv
10. daily_88502_2010.csv
11. daily_88502_2011.csv
12. daily_88502_2012.csv
13. daily_88502_2013.csv
14. daily_88502_2014.csv

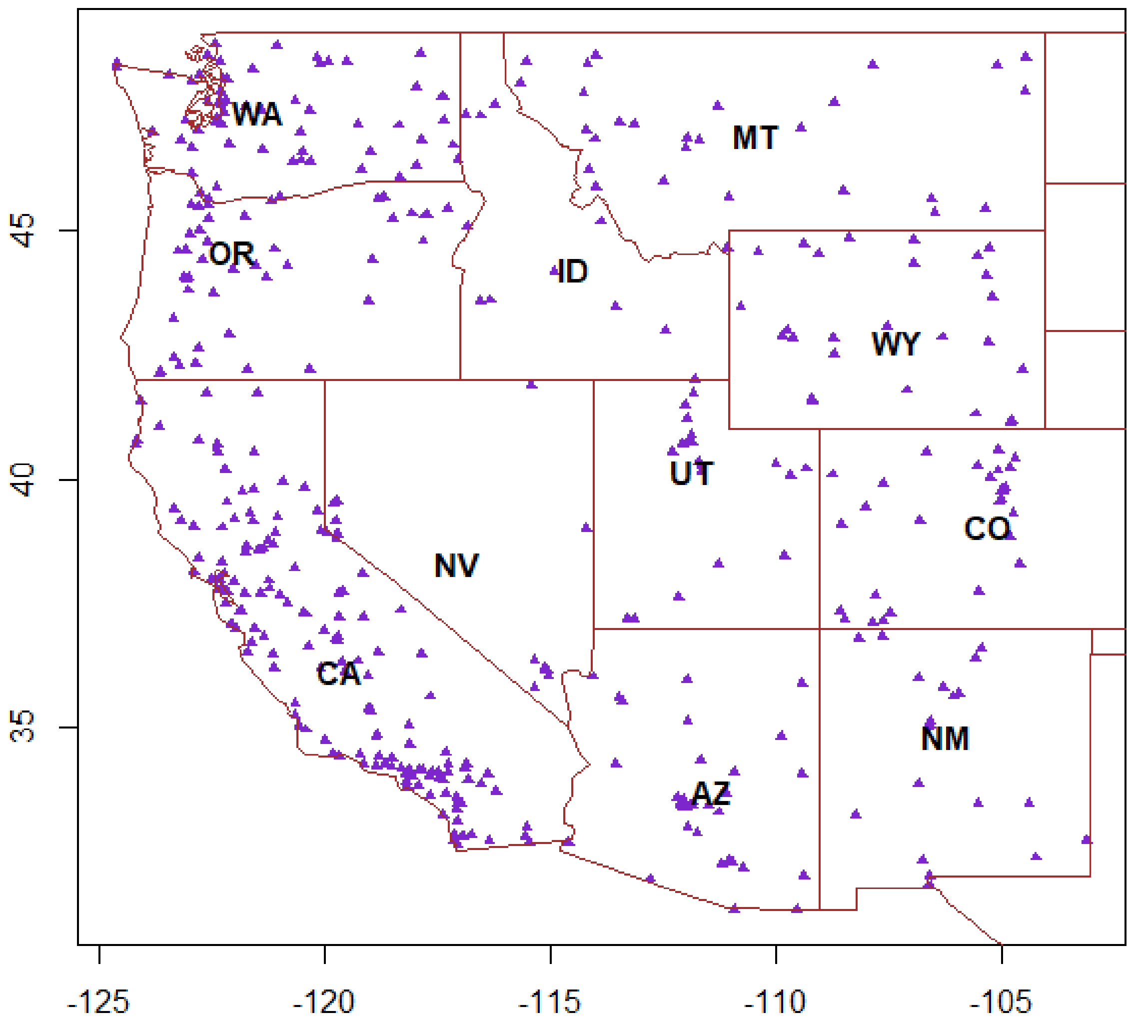

Figure 3: Map of 88101 and 88502 $PM_{2.5}$ Monitors.

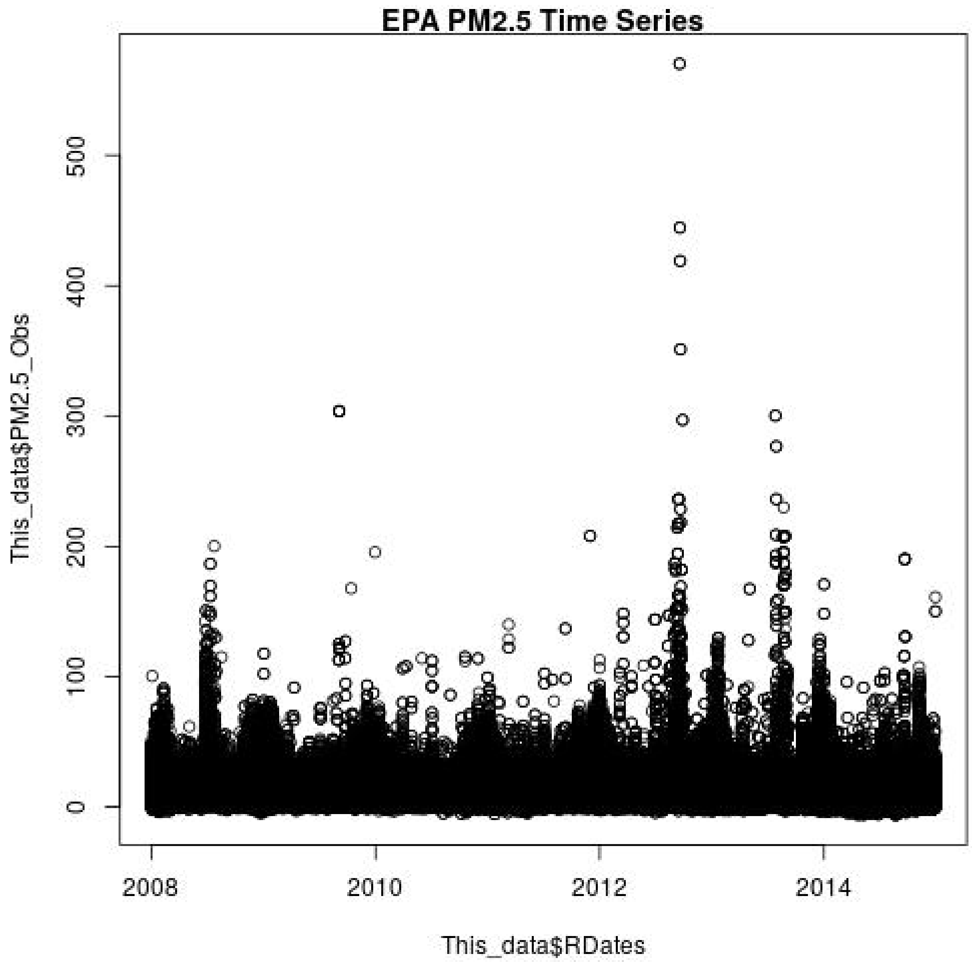

Figure 4: EPA PM2.5 time series. There are 49 data points (out of 1848393) with concentrations greater than 200 ug/m3

1.3 EPA PM2.5 Plots

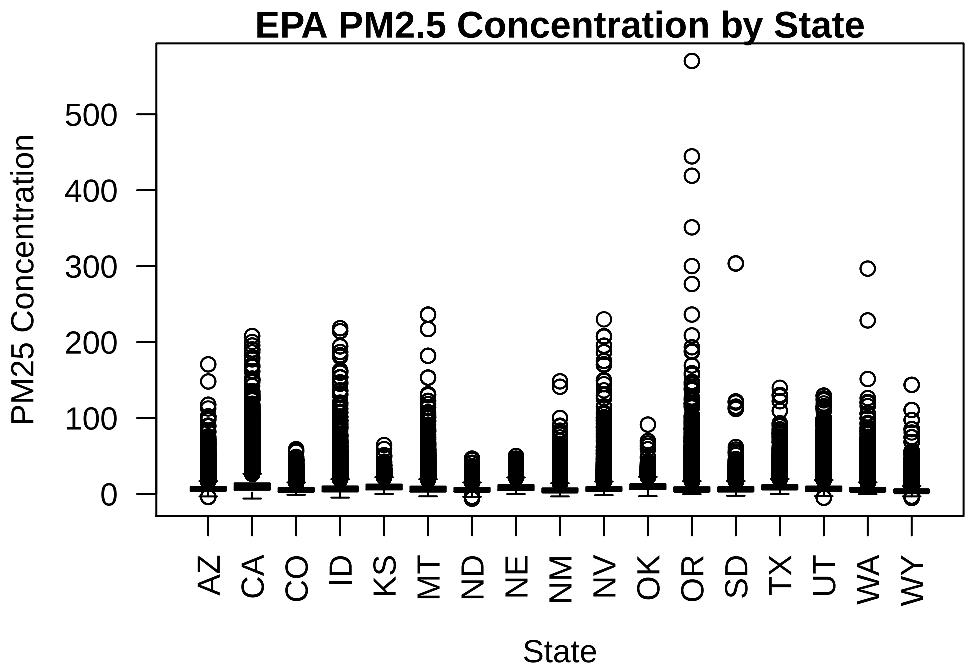

Figure 5: EPA PM2.5 box plots.

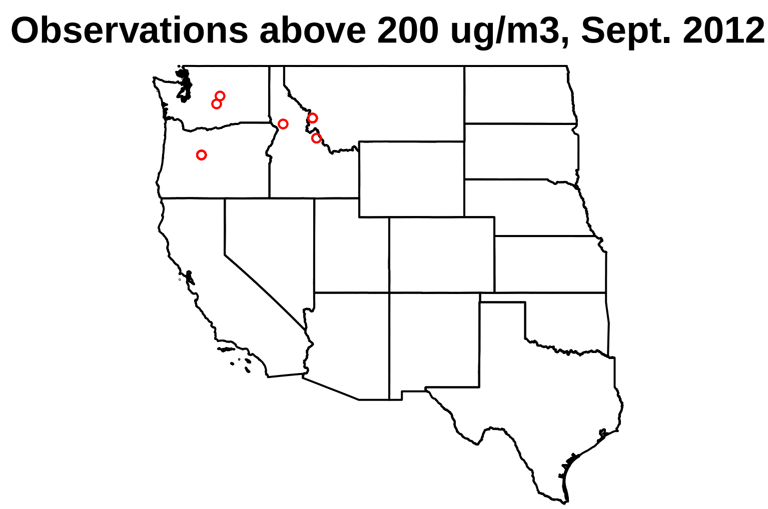

Figure 6: EPA PM2.5 map of locations with PM2.5 above 200 ug/m3 during Sept 2012.

1.4 PM2.5 Monitor data from IMPROVE network

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will download $PM_{2.5}$ data from both the US EPA AQS Air Data Query Tool (US EPA, 2017c) and the IMPROVE monitors that capture air quality information in more rural areas (US EPA, 2017e) for the 11-state region (Figure 1) including any of the following parameter codes: 88101, 88500, 88502, 81104 (US EPA, 2017a,d,f).

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1.
2.

Quality Control

Script Names

1.

Data File Names

1.

1.5 $PM_{2.5}$ data from the Federal Land Manager Environmental Database

Data Source

- **Contact** Tom Moore directed us to the website
- **Citation/Link** http://views.cira.colostate.edu/fed/DataWizard/Default.aspx
- **Download Date** March 15, 2018
- **Data (local)** $PM_{2.5}$ data from the Federal Land Manager Environmental Database
- **Geographic Extent** Nationwide
- **Temporal Extent** January 1, 2008 - December 31, 2014
- **Acknowledgment**

Brief Description

Downloading IMPROVE Aerosol, RHR II (New Equation) data (one parameter at at time):

1. Reports: Raw data
2. Datasets:“IMPROVE Aerosol, RHR II (New Equation)”
3. Sites: select all
4. Parameters:
   (a) Mass, PM2.5 (Fine): Code MF, Type PM2.5, Units ug/m$\hat{3}$ LC AQS ID 88101
5. Select Dates: By Years and Months: 2008-2014; select all months
6. Aggregations: Non-aggregated
7. Fields: Select All
8. Options: Text File; Generate one file containing all the data; Comma delimited, Standard (“wide” format); Data & Metadata, Display Column Headers, Don’t Display Section Titles, String Quotes: Double Quotes, Missing Values (blank); Date Format: 3/14/2002; Display Results: In a separate browser window; Show Report Log
9. Submit

Notes

Data sets in this database that don’t work for this project:

1. “SEARCH FRM" is only available 1998-2005
2. “SEARCH Best Estimate” is only available 1998-2005

File Formats

csv

Data Filtering and Processing

Final Variable(s)

Methods

1.

2.

Quality Control

Script Names

1.

Original Data File Names

1. Federal_Land_Manager_IMPROVE_RHR_II_2018315132109KL0L2K.csv

Processed/Cleaned Data File Names

1.

1.6 $PM_{2.5}$ data from the Fire Cache Smoke Monitor Archive

Data Source

- **Contact** Josh Walston at 775-673-7624; Amber Ortega directed us to the website and Scott Landis suggested that a good person to contact about the page would be Mike Broughton from the US Forest Service (michaelbroughton@fs.fed.us)
- **Citation/Link** https://wrcc.dri.edu/cgi-bin/smoke.pl
- **Data (local)** $PM_{2.5}$ data from the Fire Cache Smoke Monitor Archive
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

Notes

Several of the files were password protected, so we contacted Josh and they were able to unlock most of them. As of March 20, 2018, only "Smoke NCFS E-BAM # 3 is still password protected. (Need to try calling Josh again.) Here are some comments that the system administrator passed along to us (via Josh): the data does not get quality controlled, so we should do our own qa/qc. The monitors/data were designed for the fire community to see data in real time, not for research purposes. If we want to speak with the people who ran the monitors, we should contact Josh and the director can probably put us in contact.

These monitors were not included because the website indicated that it did not have data during our study period (January 1, 2008 - December 31, 2014):

1. Smoke E-BAM 418
2. Smoke E-BAM 591
3. Smoke E-BAM 592
4. Smoke E-BAM 882
5. Smoke E-BAM 969
6. Smoke USFS R2-922
7. Smoke USFS R2-923
8. Smoke USFS R2-924
9. Smoke USFS R8-34
10. Smoke USFS R8-35
11. Smoke USFS R8-55
12. Smoke USFS R8-56
13. Smoke USFS 3015
14. Smoke USFS 3016
15. Smoke USFS R9-3017
16. Smoke USFS R9-3018
17. RSF Smoke Monitor 1
18. Lolo NF Smoke Monitor #1
19. Lolo NF Smoke Monitor #2

File Formats

.csv

Data Filtering and Processing

Final Variable(s)

Methods

1.
2.

Quality Control

Script Names

1.

Original Data File Names

1. Fire_Cache_Smoke_DRI_FWS_Smoke_N1.csv
2. Fire_Cache_Smoke_DRI_Smoke_N11.csv
3. Fire_Cache_Smoke_DRI_Smoke_N13.csv
4. Fire_Cache_Smoke_DRI_Smoke_N15.csv
5. Fire_Cache_Smoke_DRI_Smoke_N16.csv
6. Fire_Cache_Smoke_DRI_Smoke_N17.csv
7. Fire_Cache_Smoke_DRI_Smoke_N19.csv
8. Fire_Cache_Smoke_DRI_Smoke_N20.csv
9. Fire_Cache_Smoke_DRI_Smoke_N21.csv
10. Fire_Cache_Smoke_DRI_Smoke_N22.csv
11. Fire_Cache_Smoke_DRI_Smoke_N23.csv
12. Fire_Cache_Smoke_DRI_Smoke_N24.csv
13. Fire_Cache_Smoke_DRI_Smoke_N25.csv
14. Fire_Cache_Smoke_DRI_Smoke_N65.csv
15. Fire_Cache_Smoke_DRI_Smoke_N66.csv
16. Fire_Cache_Smoke_DRI_Smoke_N67.csv
17. Fire_Cache_Smoke_DRI_Smoke_N68.csv
18. Fire_Cache_Smoke_DRI_Smoke_N69.csv
19. Fire_Cache_Smoke_DRI_Smoke_N84.csv
20. Fire_Cache_Smoke_DRI_Smoke_N86.csv
21. Fire_Cache_Smoke_DRI_Smoke_N215.csv
22. Fire_Cache_Smoke_DRI_Smoke_N216.csv
23. Fire_Cache_Smoke_DRI_Smoke_N217.csv
24. Fire_Cache_Smoke_DRI_Smoke_E_BAM_52.csv
25. Fire_Cache_Smoke_DRI_Smoke_E_BAM_65.csv
26. Fire_Cache_Smoke_DRI_Smoke_E-BAM_231.csv
27. Fire_Cache_Smoke_DRI_Smoke_E-BAM_840.csv
28. Fire_Cache_Smoke_DRI_Smoke_E-BAM_866.csv
29. Fire_Cache_Smoke_DRI_Smoke_E-BAM_925.csv
30. Fire_Cache_Smoke_DRI_Smoke_NCFS_E-BAM_N1.csv
31. Fire_Cache_Smoke_DRI_Smoke_NCFS_E-BAM_N2.csv
32. Fire_Cache_Smoke_DRI_Smoke_USFS_R1-39.csv
33. Fire_Cache_Smoke_DRI_Smoke_USFS_R1-52.csv
34. Fire_Cache_Smoke_DRI_Smoke_USFS_R1-53.csv

35. Fire_Cache_Smoke_DRI_Smoke_USFS_R1-306.csv
36. Fire_Cache_Smoke_DRI_Smoke_USFS_R1-307.csv
37. Fire_Cache_Smoke_DRI_Smoke_USFS_R2-69.csv
38. Fire_Cache_Smoke_DRI_Smoke_USFS_R2-78.csv
39. Fire_Cache_Smoke_DRI_Smoke_USFS_R2-264.csv
40. Fire_Cache_Smoke_DRI_Smoke_USFS_R2-265.csv
41. Fire_Cache_Smoke_DRI_Smoke_USFS_R3-28.csv
42. Fire_Cache_Smoke_DRI_Smoke_USFS_R3-86.csv
43. Fire_Cache_Smoke_DRI_Smoke_USFS_R5-39.csv
44. Fire_Cache_Smoke_DRI_Smoke_USFS_R5-49.csv
45. Fire_Cache_Smoke_DRI_Smoke_USFS_R8-33.csv
46. Fire_Cache_Smoke_DRI_Smoke_USFS_R9-15.csv
47. Fire_Cache_Smoke_DRI_Smoke_USFS_R9-16.csv
48. Fire_Cache_Smoke_DRI_Smoke_USFS_R9-17.csv
49. Fire_Cache_Smoke_DRI_Smoke_USFS_R9-60.csv
50. Fire_Cache_Smoke_DRI_Smoke_NPS_Yosemite_01_California.csv

Processed/Cleaned Data File Names

1.
2.

Download instructions

1. https://wrcc.dri.edu/cgi-bin/smoke.pl
2. Hover over the appropriate drop-down menu and click on the monitor you want to download e.g., "Cache Monitors" then "Smoke #11"
3. On the left-side menu, click on "Data Details"
4. Set the starting date: January 1, 2008 (or as far back as it goes if it doesn't go back to 2008)
5. Set the ending date: December 31, 2014 (or the last date possible if it ends before 2014)
6. Elements (ignore - default is to include all elements)
7. Options
8. Excel (.xls) (It had html code in the file if I chose other options.)
9. Data Source: Original
10. Represent missing data as: -9999.
11. Include data flags: Yes
12. Date format: MM/DD/YYYY hh:mm
13. Time format: LST 0-23
14. Table Header: Column header short descriptions
15. Field Delimiter: comma (,)
16. Select the Units: Metric
17. Leave Sub interval windows set to: January 01, December 31, Hours: 00 and 24
18. Submit Info
19. Open in excel
20. Save as: Fire_Cache_Smoke_DRI_*.csv Where * is the monitor name with spaces replaced with underscore and # symbols replaced with the letter N, e.g., the file name for monitor "Smoke #11" is "Fire_Cache_Smoke_DRI_Smoke_N11.csv"

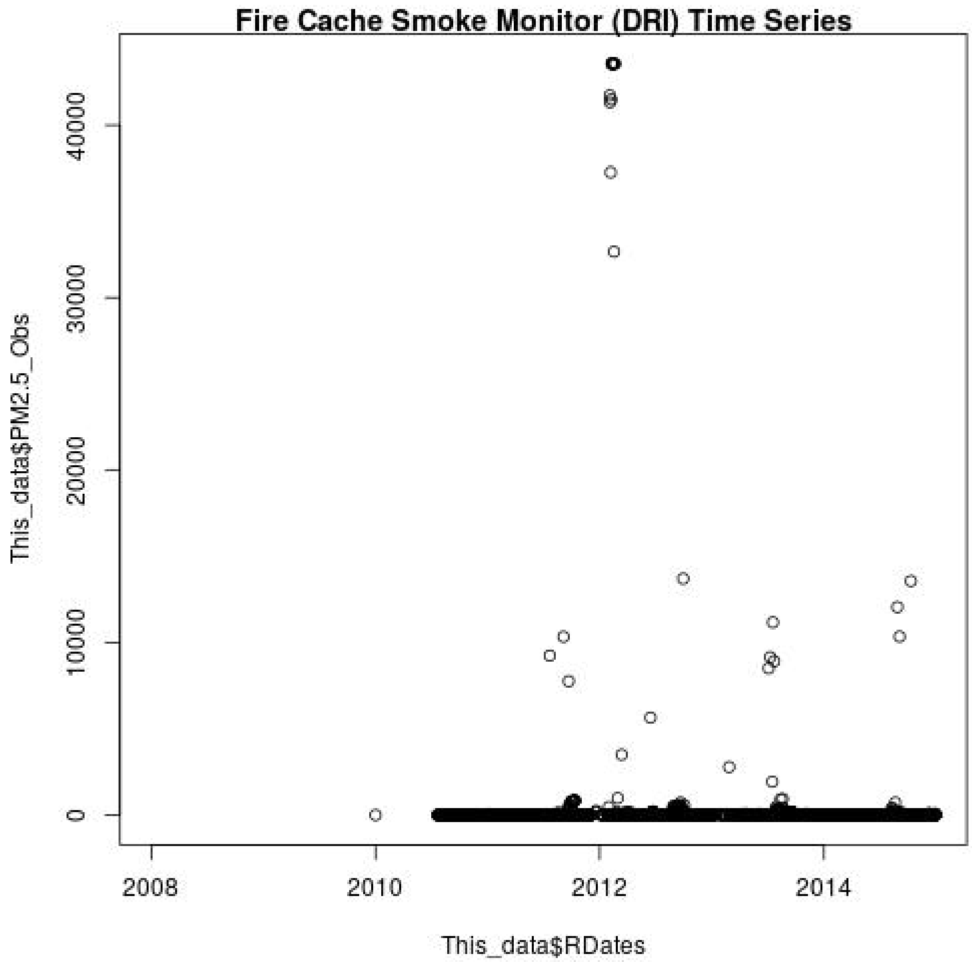

Figure 7: Fire Cache Smoke Monitor (DRI) time series. There are 104 data points (out of 4796) with concentrations greater than 200 ug/m3

21. Upload file to S3 bucket: https://732215511434.signin.aws.amazon.com/console
22. Click on S3
23. Earthlab-reid-group
24. Fire_Cache_Smoke_DRI (folder)
25. Check the following:
26. The name of the monitor is in cell A1
27. The header is spread across rows 2-4
28. There are 34 columns of data (goes through columns "AH" in excel)
29. Concentration in the 11th ("K") columns
30. List the file names in the overleaf documentation (PM25_Fire_Cache_Smoke_Monitor_Archive.tex)

1.7 Fire Cache Smoke Monitor (DRI) Plots

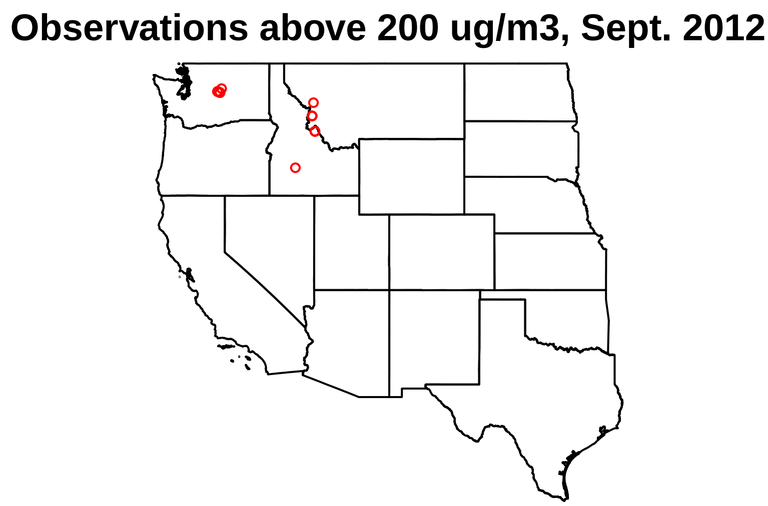

Figure 8: Fire Cache Smoke Monitor (DRI) map of locations with PM2.5 above 200 ug/m3 during Sept 2012.

1.8 $PM_{2.5}$ Monitor data from Uintah Basin

Data Source

- **Contact** Seth Lyman
- **Citation/Link** seth.lyman@usu.edu
- **Data (local)** $PM_{2.5}$ measurements from 10 sites in Uintah Basin, Utah
- **Geographic Extent** Uintah Basin, Utah
- **Temporal Extent** October 2009 - March 2017
- **Acknowledgment** $PM_{2.5}$ data from the Uintah Basin were provided by Seth Lyman at Utah State University.

Brief Description

$PM_{2.5}$ data were provided by Seth Lyman at Utah State University via email on January 16, 2018. The .xlsx file has $PM_{2.5}$ data from 10 stations during 2009-2017. The .png file has the longitude and latitude of each site.

Notes

Additional information from Seth's email:
"I've attached most of the PM2.5 observations that have ever been collected in the Uintah Basin. What are in the Excel file are 24-hr average data. Data from Roosevelt, Vernal, Ouray, Red Wash, Myton, and Rangely are from the EPA AQS database.
Data from Horsepool are from a BAM 1020 monitor that we operate every winter. Data in Ft. Duchesne and Randlett are 24-hr filter samples that were analyzed gravimetrically. Data from Rabbit Mountain are from a BAM 1020, and data through mid-2013 are in the AQS database.

I have hourly data from Horsepool and Rabbit Mountain if you'd rather have that.

Site locations are given in the list of monitoring stations for the Basin below."

The .png file is easier to read in some programs than others, e.g., it looks fine in "Paint," but not "Photos."

File Formats

Excel and png

Data Filtering and Processing

FinalPM2.5_multiyear_thruwint2017_sheet1.csv is the first sheet of FinalPM2.5_multiyear_thruwint2017.xlsx converted to .csv, and the second row of the header was merged into the first (24hr avg PM2.5).

FinalPM2.5_multiyear_thruwint2017_GISsheet.csv is the third sheet of FinalPM2.5_multiyear_thruwint2017. converted to .csv and gives the latitude and longitude of each site. This sheet originally did not have location information from the Rangely site, so this was filled in by hand with the numbers form UintahBasinSiteLocations.png.

Final Variable(s)

Methods

1.

2.

Quality Control

Script Names

1.

Original Data File Names

1. FinalPM2.5_multiyear_thruwint2017.xlsx
2. UintahBasinSiteLocations.png

Processed/Cleaned Data File Names

1. FinalPM2.5_multiyear_thruwint2017_sheet1.csv
2. UintahBasinSiteLocations.png

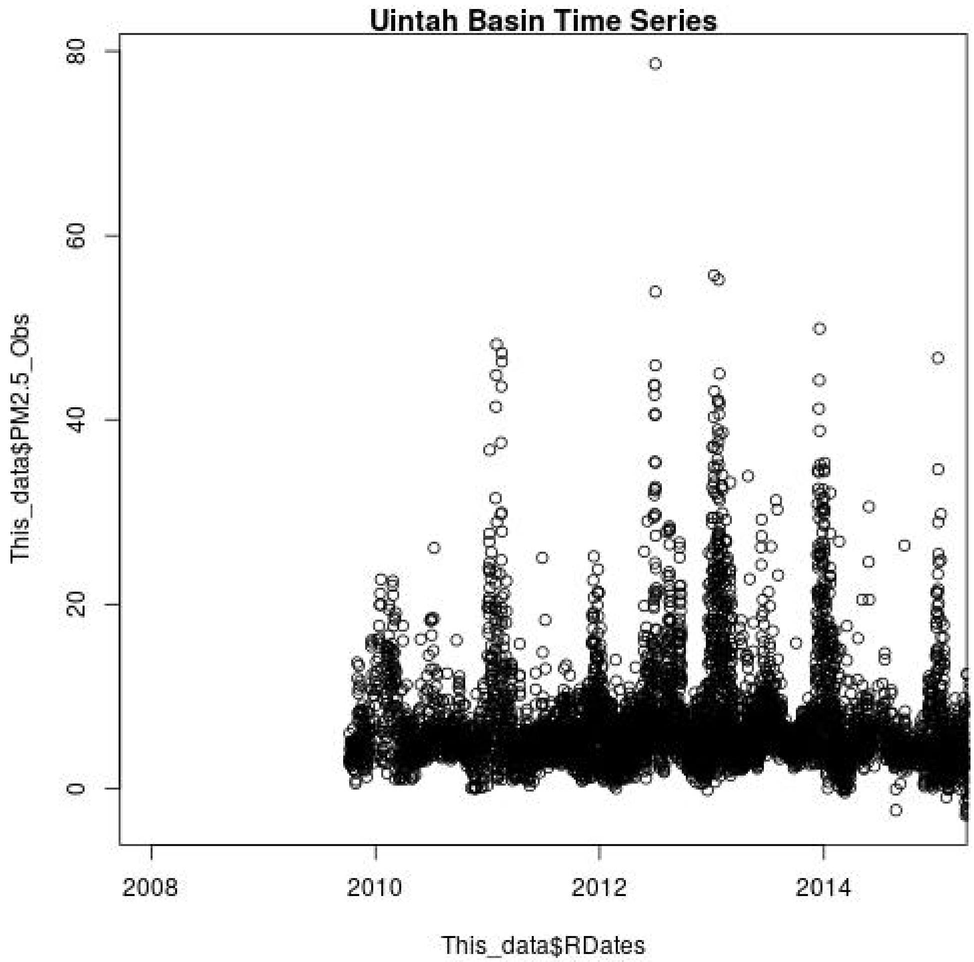

Figure 9: Uintah Basin time series. There are 0 data points (out of 27330) with concentrations greater than 200 ug/m3

1.9 Uintah Basin Plots

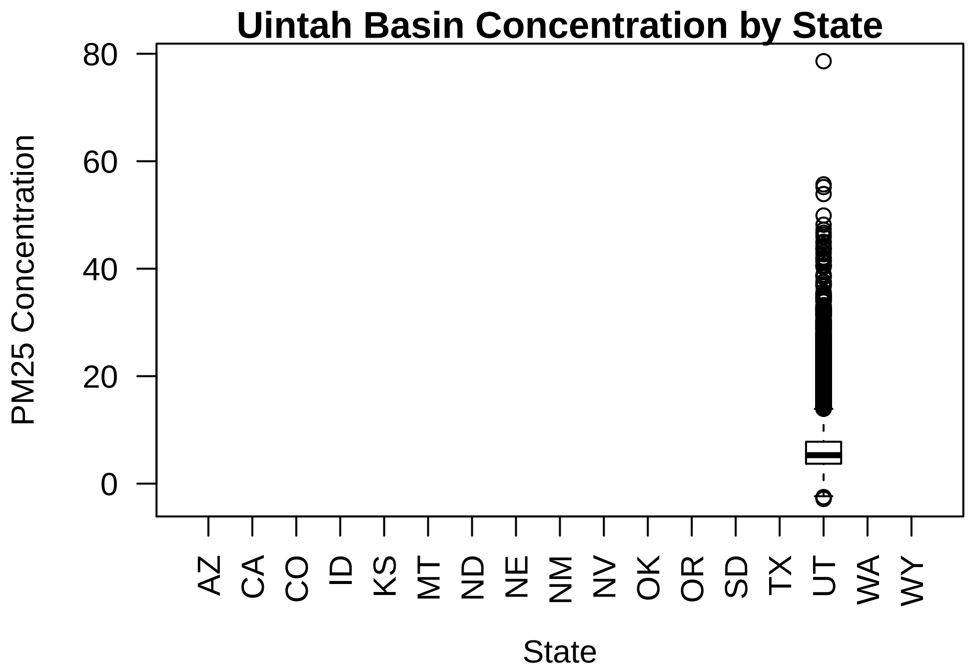

Figure 10: Uintah Basin box plots.

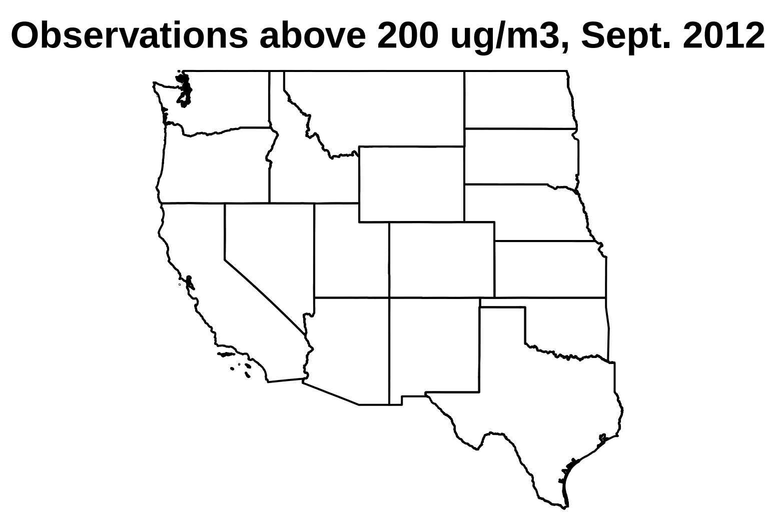

Figure 11: Uintah Basin map of locations with PM2.5 above 200 ug/m3 during Sept 2012.

1.10 $PM_{2.5}$ data from PCAPS in the Salt Lake Valley

Data Source

- **Contact** Dr. Geoff Silcox in Chemical Engineering at the University of Utah (geoff@chemeng.utah.edu)
- **Citation/Link** Publication: https://www.sciencedirect.com/science/article/pii/S1352231011011204 (Data was received from Dr. Silcox via email on February 6, 2018.)
- **Data (local)** $PM_{2.5}$ data from the Persistent Cold Air Pool Study (PCAPS)
- **Geographic Extent** Salt Lake Valley
- **Temporal Extent** January - February, 2011
- **Acknowledgment** Dr. Geoff Silcox

Brief Description

Notes

File Formats

.xlsx

Data Filtering and Processing

PCAPS_Site_Locations.csv is the same data as Table 1 of final_publication.pdf, and has the site locations and elevation.

Final Variable(s)

Methods

1.

2.

Quality Control

Script Names

1.

Original Data File Names

1. final_publication.pdf (Publication of paper)

2. MiniVol_data.xlsx

Processed/Cleaned Data File Names

1. MiniVol_data.csv

2. PCAPS_Site_Locations.csv

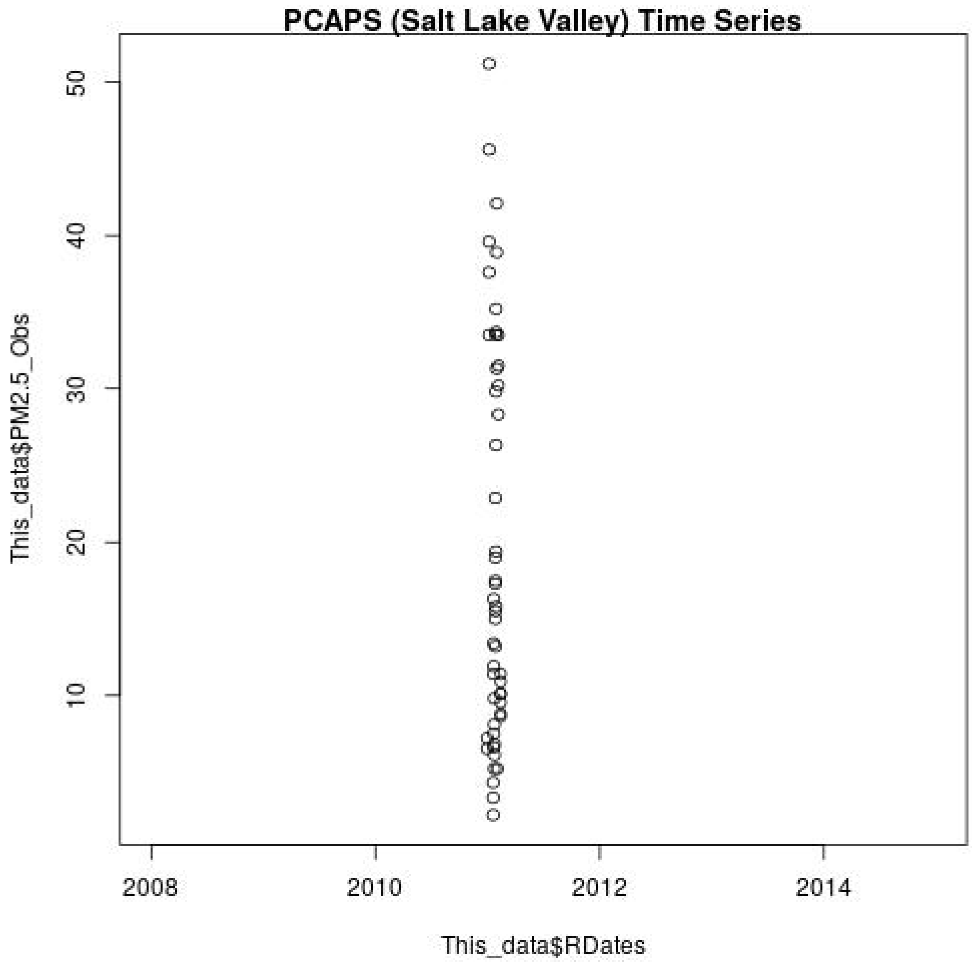

Figure 12: PCAPS (Salt Lake Valley) time series. There are 0 data points (out of 59) with concentrations greater than 200 ug/m3

1.11 PCAPS (Salt Lake Valley) Plots

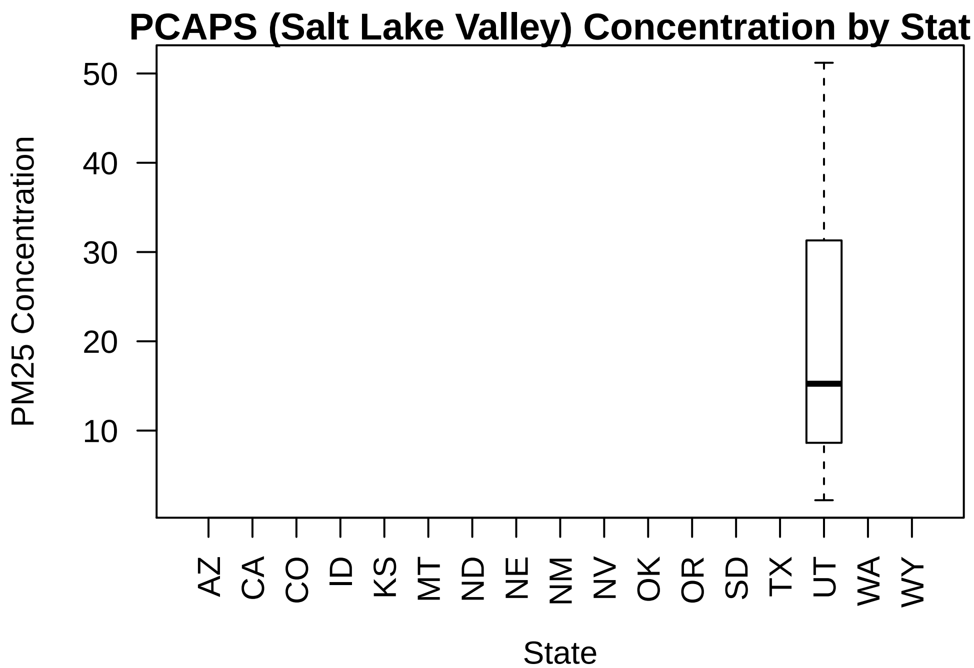

Figure 13: PCAPS (Salt Lake Valley) box plots.

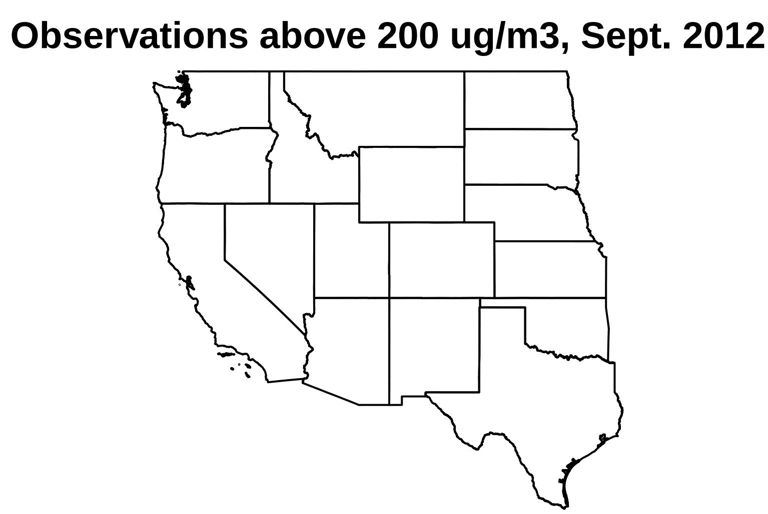

Figure 14: PCAPS (Salt Lake Valley) map of locations with PM2.5 above 200 ug/m3 during Sept 2012.

1.12 Arizona Department of Environmental Quality (ADEQ)

Data Source

- **Contact** Phone: 602-771-7676; also email option on website
- **Citation/Link** http://azdeq.gov/node/2204
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent** Only has data archived 2016 - 2017
- **Acknowledgment**

Brief Description

Notes

File Formats

Data Filtering and Processing

Final Variable(s)

Methods

1.

2.

Quality Control

Script Names

1.

Original Data File Names

1.

2.

Processed/Cleaned Data File Names

1.

2.

1.13 MODIS AOD

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will use AOD estimates from the Deep Blue retrieval algorithm for AOD from the MODIS instrument on the NASA Terra and Aqua satellites (MOD04_L2 and MYD04_L2) (Sayer et al., 2013). The MODIS product is available twice daily at a 10 km spatial resolution for cloud-free scenes and is available longer than our 2008-2014 study period (NASA LAADS DAAC, 2017a,b).

AOD products use cloud filtering algorithms that often remove pixels in the center of the smoke plumes because they are assumed to be clouds due to high reflectivity (Kondragunta and Seybold, 2009). Given that these can be in the middle of smoke plumes, often the locations most heavily impacted by smoke have missing data for a key variable, AOD. In our previous work in summer in California when rain clouds are incredibly rare, we could be confident that missing values not along the coast were not clouds. However, for this larger study region and time period, this will be a bigger challenge. We will attempt to isolate smoke plumes from true clouds using satellite imagery and smoke plume polygons from NOAA's Hazard Mapping System Fire Smoke Product (NOAA OSPO, 2017). We will then estimate missing values within validated smoke plumes, but not within clouds, using radial basis functions as was done in our previous work (Reid et al., 2015). Radial basis functions are exact interpolation functions that will return observed AOD values where they exist but can interpolate higher values than nearby observations in missing locations, which is needed since the missing values were removed due to their high reflectivity (Reid et al., 2015).

Notes

File Format

.hdf

Data Filtering and Processing

Final Variable(s)

Methods

1. Download the MODIS AOD data sets from both Terra and Aqua sensors:

   Using the NASA EarthData online search tool, search for the 'MOD04' (Terra) data set. Set temporal extent by drawing polygon and set spatial extent by adjusting the appropriate filter on the web interface. Select the collection and proceed to download data. For data download options, specify "Stage for Delivery" through the "FTPPull" distribution option. Specify the email address for orders to be sent to. Orders will be sent to your email with instructions on how to connect to the FTP server and pull the ordered data into your local workspace through the command line. Because the amount of data being requested is large, the orders will come through several separate emails. Repeat this step for the 'MYD04'

(Aqua) data set. All of the raw downloaded data from this step will be in .hdf file format.

2. Set up file system for data processing:

   Create a directory locally named ‘collected_data‘. In this directory, make two child directories named "MOD04_terra" and "MYD04_aqua". Follow instructions in email to download data through FTP into the appropriate MODIS directory (‘MOD04_terra‘ or ‘MYD04_aqua‘) depending on whether the order is from the Terra or Aqua sensor. Create another directory locally named "processed_data" at the same level as "collected_data" (this is where the processed data will eventually go). This file naming convention is important because the python scripts that are run later depend on this hierarchy.

3. Extract lat, long, and aod values from .hdf files and save into .csv files

   Run script ‘modis_aod_create_csv_file.py‘ (will need to change input and output filepath at top of script in order to match those of your local setup). This script will take all the .hdf files that you have downloaded and store the lat/long and aod value for non-null pixels. A .csv file will be created for each corresponding .hdf file and stored in your ‘processed_data’ folder.

4. Add UTC date and time information as columns to each .csv file

   Run script ‘modis_aod_add_utc_to_csv.py‘. This will create columns "year", "month", "day", "hour_min" in each .csv file and populate the values. These values are taken from the file name and correspond to UTC.

5. Merge csv files into one

   Run script ‘modis_aod_merge_csv_files.py‘ to combine multiple .csv files into one with script that uses multiprocessing (25 processes). Run script again to combine 25 files into 2 (2 process, header=None). At this step, do not merge the last two .csv files because it will take too long in Python using this script. So, used Postgres to combine final dataset (load the last two .csv files into two separate Postgres tables then combine those two tables into one). This script also excludes the ‘point‘ column, which is unnecessary.

6. Adjust UTC datetime to local datetime

   Create another column in Postgres table with the correct TZ value from the tz database. Create additional column that adjusts UTC timestamp value to local timestamp value using the TZ value.

7. Create 24 hour averages for EPA station points

Quality Control

Script Names

1. modis_aod_create_csv_file.py

Data File Names

1. n/a

1.14 GASP-West AOD

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will use AOD estimates from the Geostationary Operational Environmental Satellite West (GOES-West) Aerosol Smoke Product (GASP-West AOD). The GASP product is available at a 4 km resolution at nadir with retrievals every 30 minutes during daylight hours and is available from 2006 onward (NOAA NCEI, 2017).

AOD products use cloud filtering algorithms that often remove pixels in the center of the smoke plumes because they are assumed to be clouds due to high reflectivity (Kondragunta and Seybold, 2009). Given that these can be in the middle of smoke plumes, often the locations most heavily impacted by smoke have missing data for a key variable, AOD. In our previous work in summer in California when rain clouds are incredibly rare, we could be confident that missing values not along the coast were not clouds. However, for this larger study region and time period, this will be a bigger challenge. We will attempt to isolate smoke plumes from true clouds using satellite imagery and smoke plume polygons from NOAA's Hazard Mapping System Fire Smoke Product (NOAA OSPO, 2017). We will then estimate missing values within validated smoke plumes, but not within clouds, using radial basis functions as was done in our previous work (Reid et al., 2015). Radial basis functions are exact interpolation functions that will return observed AOD values where they exist but can interpolate higher values than nearby observations in missing locations, which is needed since the missing values were removed due to their high reflectivity (Reid et al., 2015).

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1. Navigate to NCEI's Archive Information Request System (AIRS). Scroll down and click on 'Satellite' to expand menu. Click on 'Goes Products' to expand menu. Click on 'Order Data'.

2. Select appropriate Satellite ID for time frame of interest (we selected GOES-11 for 01/01/2008-02/13/2012 and GOES-15 for 02/14/2012-12/31/2014 to encompass our study time period of 2008-2014). Select appropriate data type (GASP-AOD-GZ-).

3. Select "Yes" for Submit Batch

4. Enter email address and submit order

Quality Control

Script Names

1.

Data File Names

1.

1.15 MODIS Thermal Anomalies/Fire Daily L3 Global 1km (MOD14 and MYD14)

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will collect data about fire detection locations, size, and fire radiative power from the MODIS Thermal Anomalies/Fire Daily L3 Global 1km (MOD14 and MYD14) (Giglio et al., 2006; Hawbaker et al., 2017). Using GIS techniques, we will create daily clusters of fire points and use these to calculate: (1) the distance to the nearest fire cluster by day and (2) the sum of Fire Radiative Power (FRP) of the nearest clusters of fires by day as it is likely that smoke levels are higher closer to fires. The MODIS product spans longer than our study period (2008-2014) at daily temporal resolution and has a spatial resolution of 1 km.

Notes

File Format

.hdf

Data Filtering and Processing

Final Variable(s)

Methods

1. Run script and pass two arguments: the first is the data set name and the second is the local directory path to save files to (i.e. "MOD14" "C:/Users/User/MOD14_Downloads")

Quality Control

Script Names

1. MODIS_FTP_Download.py

Data File Names

1. n/a

1.16 Landsat-derived burned area essential climate variable (BAECV) fire activity data

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will collect data about fire detection locations, size, and fire radiative power from the Landsat-derived burned area essential climate variable (BAECV) fire activity data, (LP DAAC, 2017). Using GIS techniques, we will create daily clusters of fire points and use these to calculate: (1) the distance to the nearest fire cluster by day and (2) the sum of Fire Radiative Power (FRP) of the nearest clusters of fires by day as it is likely that smoke levels are higher closer to fires. The BAECV can detect fires larger than 4 km^2 and provides an estimate of the date of the fire and is available from 1984-2015.

Notes

File Format

.shp

Data Filtering and Processing

Final Variable(s)

Methods

1. Navigate to USGS BAECV download page and click years to download.

Quality Control

Script Names

1. n/a

Data File Names

1. n/a

1.17 MODIS/Terra and Aqua Burned Area Monthly L3 Global 500 m SIN Grid V006 (MCD64A1)

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will collect data about fire detection locations, size, and fire radiative power from MODIS/Terra and Aqua Burned Area Monthly L3 Global 500 m SIN Grid V006 (MCD64A1) (Schroeder et al., 2014). Using GIS techniques, we will create daily clusters of fire points and use these to calculate: (1) the distance to the nearest fire cluster by day and (2) the sum of Fire Radiative Power (FRP) of the nearest clusters of fires by day as it is likely that smoke levels are higher closer to fires.

Notes

File Format

.hdf

Data Filtering and Processing

Final Variable(s)

Methods

1. Run script and pass two arguments: the first is the data set name and the second is the local directory path to save files to (i.e. "MCD64A1" "C:/Users/User/MCD64A1_Downloads")

Quality Control

Script Names

1. MODIS_FTP_Download.py

Data File Names

1.

1.18 Visible Infrared Imaging Radiometer Suite (VIIRS) (VNP14IMGTDL_NRT)

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will collect data about fire detection locations, size, and fire radiative power from the Visible Infrared Imaging Radiometer Suite (VIIRS) (VNP14IMGTDL_NRT) (Schroeder et al., 2014). Using GIS techniques, we will create daily clusters of fire points and use these to calculate: (1) the distance to the nearest fire cluster by day and (2) the sum of Fire Radiative Power (FRP) of the nearest clusters of fires by day as it is likely that smoke levels are higher closer to fires. The MODIS product spans longer than our study period (2008-2014) at daily temporal resolution and has a spatial resolution of 1 km. VIIRS was launched in 2011 and has 12 h temporal resolution with 750 m resolution. The BAECV can detect fires larger than 4 km^2 and provides an estimate of the date of the fire and is available from 1984-2015.

Notes

File Format

.csv

Data Filtering and Processing

Final Variable(s)

Methods

1. Go to the NASA EarthData Fire Information for Resource Management System (FIRMS) online tool and navigate to the Archive section. Click 'Create New Request' and specify spatial and temporal resolution. Also choose 'VIIRS' from Fire Data Source. Choose 'csv' as file type and enter email address. Wait for email, which will contain a .zip file with the data.

Quality Control

Script Names

1. n/a

Data File Names

1. fire_archive_V1_2770.csv

1.19 Classified land cover information from the Landsat-derived NLCD 2011

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

Classified land cover information from the Landsat-derived NLCD 2011 (Homer et al., 2017) will be used to calculate estimates of the percentage of urban development (codes 22, 23, and 24), agriculture (codes 81 and 82), and vegetated area other than agricultural land (codes 21, 41, 42, 43, 52, and 71) within buffer radii of 100 m, 250 m, 500 m, and 1000 m around each monitor. The buffer distance that is most highly correlated with $PM_{2.5}$ will be entered into each model. NLCD 2011 has a spatial resolution of 30 m and uses circa 2011 Landsat satellite data.

Notes

File Format

.shp

Data Filtering and Processing

Final Variable(s)

Methods

1. Go to the NLCD Download Page and click ' NLCD 2006 Land Cover (2011 Edition)'. This will begin the download process. Once finished, save and unzip the file.

Quality Control

Script Names

1. n/a

Data File Names

1.

1.20 MODIS Snow Cover Daily L3 Global 500m Grid, Version 6 (MOD10A1 and MYD10A1)

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will use snow cover data from the MODIS Snow Cover Daily L3 Global 500m Grid, Version 6 (MOD10A1 and MYD10A1) (Hall and Riggs, 2016) because snow coverage is a known contributor to wintertime $PM_{2.5}$ concentrations in mountain valleys (Whiteman et al., 2014). Daily MOD10A1 and MYD10A1 data are available since 2002 and have 500 m spatial resolution.

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1.

2.

Quality Control

Script Names

1.

Data File Names

1.

1.21 Elevation

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

Elevation can influence $PM_{2.5}$ concentrations; for example, $PM_{2.5}$ can accumulate in mountain valleys during persistent cold air pools (commonly referred to as inversions) during winter (Whiteman et al., 2014). We will get elevation data from the 3D Elevation Program, which has resolution of 1/3 arc-second. This resolution is approximately 10 m north/south and varies east/west with latitude (USGS, 2017).

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1. Navigate to the https://viewer.nationalmap.gov/basic/?basemap=b1&category=ned,nedsrc&title=3DEP%20 Elevation Program (3DEP) The National Map Viewer. Only the "1/3 arc-second DEM" filter should be selected. The file format should be ArcGrid. Click "Find Products"
2. Click "results" for the Elevation Products (3DEP).
3. Add all results to cart
4. Click "View Cart"
5. Click "Export Items to CSV"
6. Save downloaded CSV
7. Run script "NED_bulk_download.py". Pass the location of the CSV file as the first argument and the desired save path as the second argument.

Quality Control

Script Names

1. NED_bulk_download.py

Data File Names

1. CSV file with list of download URLs

1.22 Meteorological Data

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

We will obtain meteorological data from the National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) (Mesinger et al., 2006; NCEP, 2005) because it includes all of the standard meteorological variables but also has planetary boundary layer height, which has proved to be an important variable for converting AOD to $PM_{2.5}$ (Liu et al., 2005). We will calculate 24-hour averages from 3-hourly data for temperature, relative humidity, sea level pressure, surface pressure, planetary boundary layer height, dew point temperature, precipitation, and the U and V components of wind speed. NARR has 32 km resolution and is available from 1979 onward.

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1.
2.

Quality Control

Script Names

1.

Data File Names

1.

1.23 Dust Storms

Data Source

- **Contact**
- **Citation/Link**
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent**
- **Acknowledgment**

Brief Description

Dust storm records will be included in the machine learning algorithm because they can be a significant indicator of airborne particulate matter from sources other than fires. Dust storm records are available from 1993-2017. The spatial resolution varies, but includes either forecast zone or county (US National Weather Service, 2017b,c,a).

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1.

2.

Quality Control

Script Names

1.

Data File Names

1.

2 Data Sources for CAMx Modeling of Source-Attributed Air Quality Modeling

For meteorological inputs, the CAMx modeling will use archived daily 27-km Advanced Research Weather Research and Forecasting (WRF-ARW) grids available via NOAA Real-time Environmental Applications and Display sYstem (READY) servers for the entire study area and time period (Wang et al., 2007; Rolph et al., 2017). For the study years 2008-2012 and 2014, we will use fire emissions datasets prepared by the Western Regional Air Partnership (WRAP) and the National Emissions Inventory (NEI) (US EPA, 2017b) based on aggregated source-tagged fire occurrence data sources, the FCCS (Ottmar et al., 2007), and Consume (Prichard et al., 2009) modeling. For the study year 2013, we will prepare a fire emissions dataset using the same aggregated source-tagged fire occurrence data sources and FCCS/Consume modeling framework in the NASA-funded Wildland Fire Emissions Information System (WFEIS) (MTRI, 2017) developed by Co-I's French and Billmire (French et al., 2014). Fire occurrence datasets include MODIS (MOD14/MYD14 and MCD64A1) and VIIRS (VNP14IMGTDL_NRT) fire data products (Giglio et al., 2006; LP DAAC, 2017; Schroeder et al., 2014). For non-fire emissions during the entire study period, we will use the dataset prepared by WRAP for year 2008.

Look into using spot forecasts to help distinguish between wild and prescribed fires: http://www.weather.gov/spot/monitor/

3 CAMx Modeling

4 Compiling Data

4.1 Processing PM2.5 data

These are the scripts that process and compile the PM2.5 data:

1. Script1_Install_Pkgs.R » install packages
2. Create_ML_Input_File.R » compiles the various PM2.5 data sources into a single data frame. The only eliminations of data are geographic, to remove states that are neither in our study area nor adjacent to it.
3. Clean_ML_Input_File.R » clean data, e.g., negative concentrations

to be written » composite replicate data

4. Plot_ML_Input_File.R » create plots, maps, and statistical summary

to be written » merge with satelite and other data

5 Machine Learning Methods

setting aside a portion of the PM2.5 data set and then doing 10-fold cross validation on the rest of the data

see http://www.cvent.com/events/nasa-aist-machine-learning-workshop/event-summary-1f5144a5d1734ca aspx and particularly the very end of https://global.gotomeeting.com/public/recording-player.html?id=owZDmUustOjaW9sJGQ5u9cUG2pBa4D for list of resources and papers to read.

6 Machine Learning Results

[Currently, results below are derived from the example data/code from Colleen.]

References

French, N. H. F., McKenzie, D., Erickson, T., Koziol, B., Billmire, M., Endsley, K. A., Scheinerman, N. K. Y., Jenkins, L., Miller, M. E., Ottmar, R., and Prichard, S. (2014). Modeling Regional-Scale Wildland Fire Emissions with the Wildland Fire Emissions Information System. *Earth Interactions*, 18(16):1–26.

Giglio, L., Csiszar, I., and Justice, C. O. (2006). Global distribution and seasonality of active fires as observed with the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. *Journal of Geophysical Research: Biogeosciences*, 111(G2). G02016; https://modis.gsfc.nasa.gov/data/dataprod/mod14.php.

Hall, D. K. and Riggs, G. A. (2016). MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 6. *NASA National Snow and Ice Data Center Distributed Active Archive Center*. http://dx.doi.org/10.5067/MODIS/MYD10A1.006.

Hawbaker, T. J., Vanderhoof, M. K., Beal, Y.-J., Takacs, J. D., Schmidt, G. L., Falgout, J. T., Williams, B., Fairaux, N. M., Caldwell, M. K., Picotte, J. J., Howard, S. M., Stitt, S., and Dwyer, J. L. (2017). Mapping burned areas using dense time-series of Landsat data. *Remote Sensing of Environment*, 198(Supplement C):504 – 522.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., and Megown, K. (2017). Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information. *Photogrammetric Engineering & Remote Sensing*, 81(5):345 – 354. https://www.mrlc.gov/nlcd2011.php.

Kondragunta, S. and Seybold, M. (2009). Revisions to GOES Aerosol and Smoke Product (GASP) Algorithm. http://www.ssd.noaa.gov/PS/FIRE/GASP/gasp.html.

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J., and Koutrakis, P. (2005). Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing. *Environ Sci Technol*, 39(9):3269–78.

LP DAAC (2017, accessed November 12, 2017). MCD64A1: MODIS/Terra and Aqua Burned Area Monthly L3 Global 500 m SIN Grid V006. https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd64a1_v006.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, 87(3):343–360.

MTRI (2017 accessed November 7, 2017). *Wildland Fire Emissions Information System*. http://wfeis.mtri.org/.

NASA LAADS DAAC (2017, accessed November 2, 2017a). MOD04_L2 - MODIS/Terra Aerosol 5-Min L2 Swath 10km. https://ladsweb.modaps.eosdis.nasa.gov/api/v1/productPage/product=MOD04_L2.

NASA LAADS DAAC (2017, accessed November 2, 2017b). MYD04_L2 - MODIS/Aqua Aerosol 5-Min L2 Swath 10km. https://ladsweb.modaps.eosdis.nasa.gov/api/v1/productPage/product=MYD04_L2.

NCEP (2005). NCEP North American Regional Reanalysis (NARR). http://rda.ucar.edu/datasets /ds608.0/.

NOAA NCEI (2017, accessed November 2, 2017). *Satellite Data Access by Datasets*. https://www.ncdc.noaa.gov/data-access/satellite-data/satellite-data-access-datasets.

NOAA OSPO (2017, accessed November 3, 2017). *Hazard Mapping System Fire and Smoke Product*. http://www.ospo.noaa.gov/Products/land/hms.html.

Ottmar, R. D., Sandberg, D. V., Riccardi, C. L., and Prichard, S. J. (2007). An overview of the Fuel Characteristic Classification System — Quantifying, classifying, and creating fuelbeds for resource planning. *Canadian Journal of Forest Research*, 37(12):2383–2393.

Prichard, S. J., Ottmar, R. D., and Anderson, G. A. (2009). Consume 3.0 user's guide. *USDA Forest Service Pacific Wildland Fire Sciences Laboratory Rep.*, page 239.

Reid, C. E., Jerrett, M., Petersen, M. L., Pfister, G. G., Morefield, P. E., Tager, I. B., Raffuse, S. M., and Balmes, J. R. (2015). Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning. *Environ Sci Technol*, 49(6):3887–96.

Rolph, G., Stein, A., and Stunder, B. (2017). Real-time Environmental Applications and Display sYstem: READY. *Environmental Modelling & Software*, 95(Supplement C):210 – 228.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., and Jeong, M.-J. (2013). Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data. *Journal of Geophysical Research: Atmospheres*, 118(14):7864–7872.

Schroeder, W., Oliva, P., Giglio, L., and Csiszar, I. A. (2014). The New VIIRS 375m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of Environment*, 143(Supplement C):85 – 96.

US EPA (2017, accessed November 2, 2017a). *AQS Memos - Technical Note on Reporting PM2.5 Continuous Monitoring and Speciation Data to the Air Quality System (AQS)*. https://www.epa.gov /aqs/aqs-memos-technical-note-reporting-pm25-continuous-monitoring- and-speciation-data-air-quality.

US EPA (2017, accessed November 2, 2017c). *Outdoor Air Quality Data Download Daily Data*. https://www.epa.gov/outdoor-air-quality-data/download-daily-data.

US EPA (2017, accessed November 2, 2017d). *Parameters*. https://aqs.epa.gov/aqsweb/documents/codetables/parameters.html.

US EPA (2017, accessed November 2, 2017e). *PM 2.5 - Visibility (IMPROVE)*. https://www3.epa.gov/ttnamti1/visdata.html.

US EPA (2017, accessed November 2, 2017f). *Sampling Methods for All Parameters*. https://aqs.epa.gov/aqsweb/documents/codetables/methods_all.html.

US EPA (2017, accessed October 23, 2017b). *National Emissions Inventory (NEI)*. https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei.

US National Weather Service (2016, accessed November 2, 2017a). *National Weather Service Instruction 10-1605*. https://www.ncdc.noaa.gov/stormevents/pd01016005curr.pdf.

US National Weather Service (2017, accessed November 2, 2017b). *Storm Events Database*. https://www.ncdc.noaa.gov/stormevents/.

US National Weather Service (2017, accessed November 2, 2017c). *Storm Events Database: Database Details*. https://www.ncdc.noaa.gov/stormevents/details.jsp.

USGS (2017, accessed November 6, 2017). *About 3DEP Products and Services*. https://nationalmap.gov/3DEP/3dep_prodserv.html.

Wang, W., Barker, D., Bray, J., Bruyere, C., Duda, M., Dudhia, J., Gill, D., and Michalakes, J. (2007). User's Guide for Advanced Research WRF (ARW) Modeling System Version 3. *Mesoscale and Microscale Meteorology Division–National Center for Atmospheric Research (MMM-NCAR)*.

Whiteman, C. D., Hoch, S. W., Horel, J. D., and Charland, A. (2014). Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. *Atmospheric Environment*, 94(Supplement C):742 – 753.