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],

June 14, 2019

Contents

1 Abstract

The objective of this project is to estimate $PM_{2.5}$ concentrations by day and ZIP code for the western US during 2008-2018 and categorize $PM_{2.5}$ concentrations as non-fire and by type of fire.

1.1 Key Words

2 Introduction

The increase in frequency and severity of wildfires occurring in the western US (??) has led to higher air pollution levels than would be expected without the fires (?). Prescribed fires (deliberate, controlled fires) are used as a management tool to reduce fuel loads and the risk of large uncontrolled wildfires while allowing ecological benefits of fire. Previous research indicates that prescribed fires impact air quality less than wildfires on a per-fire basis (?), however differences in chemical composition of the smoke could result in different health impacts from these two types of fires. To our knowledge, previous studies have not considered if air pollution from prescribed fires and wildfires pose differential risks to public health. Wildfire smoke is a complex mixture of gases and particles, but most studies of wildfire smoke and health focus on one air pollutant, particulate matter with aerodynamic diameter smaller than 2.5 μm ($PM_{2.5}$), which is the predominant pollutant of concern to health in wildfire smoke (?).

Recent extremely large fires such as the Camp Fire in Paradise, CA and the 2017 fires in Santa Rosa, CA have burned not only vegetation but also many buildings. We hypothesize that these urban-invasive fires could have different air pollution chemistry which could affect population health differently than wildfires that burn predominantly forests or grasses. Compared to our understanding of the health impacts of other sources of air pollution, we know much less about the health impacts of wildfire smoke exposure. Recent review articles document that wildfire smoke exacerbates asthma, although the evidence is less consistent for other respiratory and cardiovascular outcomes (???), despite clearer associations found with other sources of air pollution (?). It is possible that differences in findings across studies could be due to the type of fuels burned or other characteristics of each wildfire, or they could be due to the prevalence of chronic disease within the exposed populations (???). To date, however, no one has investigated what is driving heterogeneity of effects in wildfire-health studies. Given the increased risk of wildfires in the western US (Westerling, 2016b; ?)) and the increased population living within the wildland-urban interface (?), it is increasingly important to understand whether different kinds of wildfires pose differential health risks and if wildfires exert more health concerns in areas with higher pre-existing health burdens. As a pre-requisite to such a study, this paper presents the exposure data necessary.

Our interdisciplinary team has experience modeling fire emissions and trajectories of smoke plumes (?), estimating spatiotemporal air pollution exposures (??), and analyzing epidemiological relationships between air pollution and health (??). Our proposal addresses the following **Specific Aims**:

AIM 1: Determine whether there are differential associations between $PM_{2.5}$ and respiratory hospitalizations and emergency department (ED) visits during prescribed fires compared to wildfires.

Aim 1a – For a list of 200 wildfires and 150 prescribed fires that occurred in our study area (AZ, CA, CO, NM, OR, UT, WA) during 2008-2018, estimate the background and fire-affected $PM_{2.5}$ concentrations for each affected ZIP code-day.

Aim 1b – Using Poisson generalized estimating equations (GEE) models assess associations between prescribed fire $PM_{2.5}$ and wildfire $PM_{2.5}$, separately, with respiratory hospitalizations and ED visits, adjusted for pertinent confounding variables.

AIM 2: Assess whether associations between $PM_{2.5}$ during wildfires and respiratory and cardiovascular hospitalizations and ED visits differ by wildfire characteristics (e.g., predominant fuel type, size, duration, etc.).

Aim 2a – For each fire, estimate the association between daily wildfire-affected $PM_{2.5}$ and

daily counts of respiratory and cardiovascular hospitalizations and ED visits using Poisson GEE models, adjusted for pertinent confounding variables.

Aim 2b – Using meta-regression of the individual regressions by fire from Aim 2a, quantify the role of fire characteristics in describing the heterogeneity in relationships between $PM_{2.5}$ and respiratory and cardiovascular hospitalizations and ED visits across the study fires.

AIM 3: Assess whether associations between $PM_{2.5}$ during wildfires and respiratory and cardiovascular hospitalizations and ED visits differ by population health characteristics (e.g., prevalence of chronic disease or health behaviors).

Aim 3a – Using meta-regression of the individual regressions by fire from Aim 2a, quantify the role of population health characteristics in describing the heterogeneity in relationships between $PM_{2.5}$ and respiratory and cardiovascular hospitalizations and ED visits across the study fires.

3 Materials and Methods

3.1 Phase 1: Data Sources and Points of Interest 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.

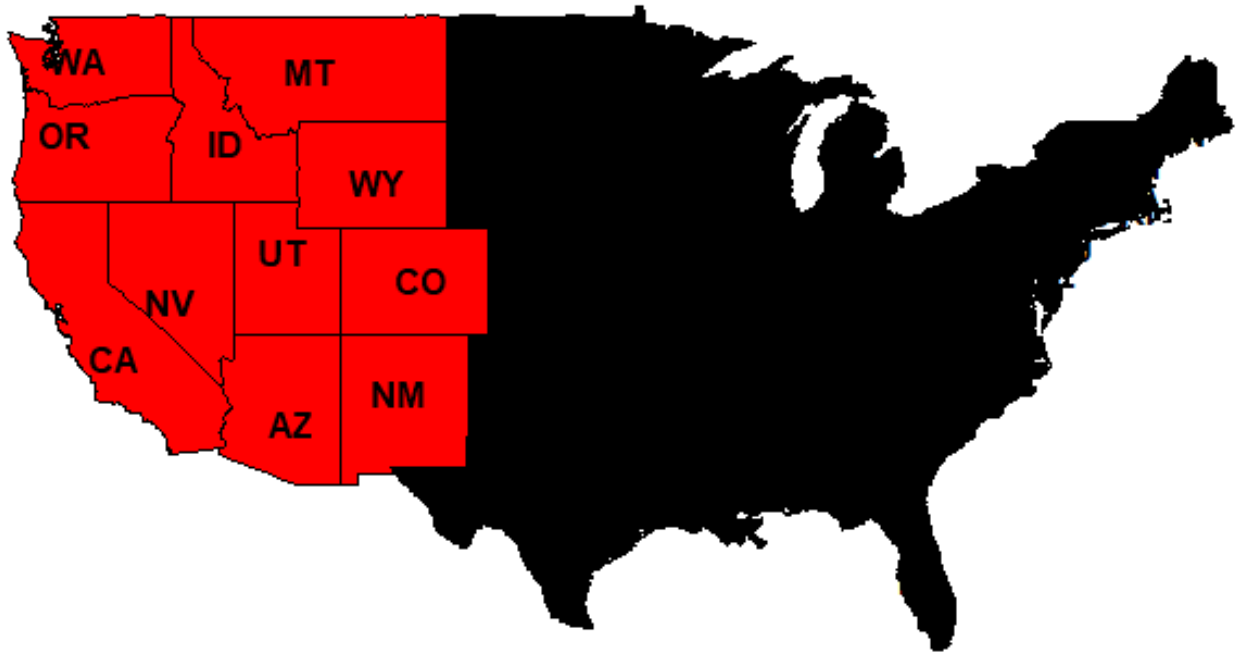

Figure 1: Map of 11-state study area.

3.1.1 $PM_{2.5}$ Monitor data from US EPA AQS Air Data Query Tool

To Do

1. Download the 2018 data again - the complete 2018 data should be available by July 2019.

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-outdoor-air-quality-data
- **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), 2008-2018*. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Daily. Download spreadsheet listing all AQS monitors with datums (https://aqs.epa.gov/aqsweb/airdata/aqs_monitors.zip) from "Monitor Listing" at https://aqs.epa.gov/aqsweb/airdata/download_files.html#Meta. The file name is aqs_monitors.csv in the AQS_Daily_Summaries folder in the S3 data.
- **Data (local)**
- **Geographic Extent**
- **Temporal Extent** 2008 through 2018. The 2018 file was downloaded June 14, 2019, and it is possible that more 2018 data will become available after this date.
- **Acknowledgment** EPA

Brief Description

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_88101_2015.csv
9. daily_88101_2016.csv
10. daily_88101_2017.csv
11. daily_88101_2018.csv * this is only part of 2018 - the rest will need to be downloaded once it becomes available
12. daily_88502_2008.csv
13. daily_88502_2009.csv
14. daily_88502_2010.csv
15. daily_88502_2011.csv
16. daily_88502_2012.csv
17. daily_88502_2013.csv
18. daily_88502_2014.csv
19. daily_88502_2015.csv
20. daily_88502_2016.csv
21. daily_88502_2017.csv

22. daily_88502_2018.csv * this is only part of 2018 - the rest will need to be downloaded once it becomes available

Script Names

1. process_PM25_EPA_data_source_function.R (called from Process_PM25_data_step1.R)

3.1.2 $PM_{2.5}$ data from the Fire Cache Smoke Monitor Archive

To Do

1. follow up about password protected monitor data (10 sites, see below)
2. also figure out if all of 2018 data is there and ask when that will be complete if it isn't

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.

Update 2018-05-2018: sent email to Josh requesting the "Smoke NCFS E-BAM # 3" data with flags and with the other formatting settings we used on the other files. Also asked how to determine which datum is assoiciated with the latitude/longitude data.

File Formats

.csv

Data Filtering and Processing

Final Variable(s)

Methods

1.
2.

Quality Control

Script Names

1.

Processed/Cleaned Data File Names

1.
2.

Download instructions

1. For each file:
   (a) https://wrcc.dri.edu/cgi-bin/smoke.pl
   (b) Hover over the appropriate drop-down menu and click on the monitor you want to download e.g., "Cache Monitors" then "Smoke #11"
   (c) On the left-side menu, click on "Data Details"
   (d) Set the starting date: January 1, 2008 (or as far back as it goes if it doesn't go back to 2008)
   (e) Set the ending date: Last date possible
   (f) Elements (ignore - default is to include all elements)
   (g) Options
      i. Excel (.xls) (It had html code in the file if I chose other options.)
      ii. Data Source: Original
      iii. Represent missing data as: -9999.
      iv. Include data flags: Yes
      v. Date format: MM/DD/YYYY hh:mm
      vi. Time format: LST 0-23
      vii. Table Header: Column header short descriptions
      viii. Field Delimiter: comma (,)
      ix. Select the Units: Metric
      x. Leave Sub interval windows set to: January 01, December 31, Hours: 00 and 24
      xi. Submit Info
   (h) Open in excel (this takes several seconds and sometimes there is an alert dialog box, answer yes)
   (i) Check the following:
      i. The name of the monitor is in cell A1
      ii. The header is spread across rows 2-4
      iii. There are 34 columns of data (goes through columns "AH" in excel)
      iv. Concentration in the 11th ("K") columns
   (j) 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". If a window pops up asking whether to keep the csv format, answer yes.
   (k) List the file names in the latex documentation (PM25_Fire_Cache_Smoke_Monitor_Archive.tex)

2. Once all files are downloaded locally
   (a) Upload file to S3 bucket in the Earthlab-reid-group folder within the Fire_Cache_Smoke_DRI subfolder https://732215511434.signin.aws.amazon.com/console

Original Data File Names and notes about monitors (downloaded February 26-27, 2019)

1. Cache Monitors
   1 Smoke #11: August 2010 – December 2018; Fire_Cache_Smoke_DRI_Smoke_N11.csv
   2 Smoke #13: January 2010 – February 2019; Fire_Cache_Smoke_DRI_Smoke_N11.csv
   3 Smoke #15: August 2010 – December 2018; Fire_Cache_Smoke_DRI_Smoke_N15.csv

4 Smoke #16: June 2011 – June 2018; Fire_Cache_Smoke_DRI_Smoke_N16.csv
5 Smoke #17: September 2010 – July 2018; Fire_Cache_Smoke_DRI_Smoke_N17.csv
6 Smoke #19: August 2010 – October 2018; Fire_Cache_Smoke_DRI_Smoke_N19.csv
7 Smoke #20: June 2011 – October 2018; Fire_Cache_Smoke_DRI_Smoke_N20.csv
8 Smoke #21: April 2011 – December 2018; Fire_Cache_Smoke_DRI_Smoke_N21.csv
9 Smoke #22: February 2011 – November 2018; Fire_Cache_Smoke_DRI_Smoke_N22.csv
10 Smoke #23: March 2011 – February 2019; Fire_Cache_Smoke_DRI_Smoke_N23.csv * This file seems to have a few rows of data shifted by two columns. Those rows look like they could be problematic, so I'll leave them as-is and let the processing scripts remove them.
11 Smoke #24: March 2011 – October 2018; Fire_Cache_Smoke_DRI_Smoke_N24.csv * This file seems to have a few rows of data shifted by two columns. Those rows look like they could be problematic, so I'll leave them as-is and let the processing scripts remove them.
12 Smoke #65: March 2013 – August 2018; Fire_Cache_Smoke_DRI_Smoke_N65.csv
13 Smoke #66: March 2012 – October 2018; Fire_Cache_Smoke_DRI_Smoke_N66.csv
14 Smoke #67: January 2012 – June 2018; Fire_Cache_Smoke_DRI_Smoke_N67.csv
15 Smoke #68: March 2012 – November 2018; Fire_Cache_Smoke_DRI_Smoke_N68.csv
16 Smoke #69: January 2012 – October 2018; Fire_Cache_Smoke_DRI_Smoke_N69.csv
17 Smoke #84: March 2013 – July 2018; Fire_Cache_Smoke_DRI_Smoke_N84.csv * This file seems to have a few rows of data shifted by a few columns. Assuming those rows could be problematic, I'll leave them as-is and let the processing scripts remove them.
18 Smoke #215: January 2014 – October 2018; Fire_Cache_Smoke_DRI_Smoke_N215.csv
19 Smoke #216: January 2014 – February 2019; Fire_Cache_Smoke_DRI_Smoke_N216.csv
20 Smoke # 217: January 2014 – October 2018; Fire_Cache_Smoke_DRI_Smoke_N216.csv * This file seems to have a few rows of data shifted by a few columns. Assuming those rows could be problematic, I'll leave them as-is and let the processing scripts remove them.
21 Smoke E-BAM 231: November 2014 – December 2018; Fire_Cache_Smoke_DRI_Smoke_E-BAM_231.csv
22 Smoke E-BAM 418: July 2017 – November 2018; **PASSWORD PROTECTED, CAN'T DOWNLOAD**
23 Smoke E-BAM 591: July 2017 – September 2017 **PASSWORD PROTECTED, CAN'T DOWNLOAD**
24 Smoke E-BAM 592: July 2017 – August 2018 **PASSWORD PROTECTED, CAN'T DOWNLOAD**
25 Smoke E-BAM 840: October 2014 – February 2019; Fire_Cache_Smoke_DRI_Smoke_E-BAM_840.csv
26 Smoke E-BAM 866: November 2014 – November 2018; Fire_Cache_Smoke_DRI_Smoke_E-BAM_866.csv
27 Smoke E-BAM 882: July 2017 – September 2018 **PASSWORD PROTECTED, CAN'T DOWNLOAD**
28 Smoke E-BAM 925: November 2014 – February 2019; Fire_Cache_Smoke_DRI_Smoke_E-BAM_925.csv * This file seems to have a few rows of data shifted by a few columns.

Assuming those rows could be problematic, I'll leave them as-is and let the processing scripts remove them.

29 Smoke E-BAM 969: July 2017 – September 2017; **PASSWORD PROTECTED, CAN'T DOWNLOAD**

2. USFS Regional Monitors

1 Smoke USFS R-39: February 2013 – October 2018; Fire_Cache_Smoke_USFS_R1-39.csv
2 Smoke USFS R1-52: January 2013 – February 2019; Fire_Cache_Smoke_USFS_R1-52.csv
3 Smoke USFS R1-53: February 2013 – February 2019; Fire_Cache_Smoke_USFS_R1-53.csv
4 Smoke USFS R1-306: November 2014 – February 2019; Fire_Cache_Smoke_USFS_R1-306.csv
5 Smoke USFS R1-307: November 2014 – September 2017; Fire_Cache_Smoke_USFS_R1-307.csv
6 Smoke USFS R2-69: May 2013 – February 2019; Fire_Cache_Smoke_USFS_R2-69.csv
7 Smoke USFS R2-78: April 2013 – October 2017; Fire_Cache_Smoke_USFS_R2-78.csv
8 Smoke USFS R2-264: November 2014 – October 2017; Fire_Cache_Smoke_USFS_R2-264.csv
9 Smoke USFS R2-265: December 2014 – October 2018; Fire_Cache_Smoke_USFS_R2-264.csv
10 Smoke USFS R2-922: June 2015 – June 2018; Fire_Cache_Smoke_USFS_R2-922.csv
11 Smoke USFS R2-923: June 2015 – October 2016; Fire_Cache_Smoke_USFS_R2-923.csv
12 Smoke USFS R2-924: June 2015 – February 2019; Fire_Cache_Smoke_USFS_R2-924.csv (website claims August of 1941 has data, but this is highly dubious.)
13 Smoke USFS R3-28: March 2013 – October 2018; Fire_Cache_Smoke_USFS_R3-28.csv * This file seems to have a few rows of data shifted by a few columns. Assuming those rows could be problematic, I'll leave them as-is and let the processing scripts remove them.
14 Smoke USFS R3-86: March 2018 – February 2019; Fire_Cache_Smoke_USFS_R3-86.csv
15 Smoke USFS R5-39: June 2013 – July 2014; Fire_Cache_Smoke_USFS_R5-39.csv
16 Smoke USFS R5-49: June 2013 – July 2016; Fire_Cache_Smoke_USFS_R5-49.csv * concentrations are all 0 or -9999 ug/m3. Air Flow is correspondingly 0 or -9999 L/min, so all of these data will be removed by the processing scripts.
17 Smoke USFS R8-33: December 2014 – January 2017; Fire_Cache_Smoke_USFS_R8-33.csv
18 Smoke USFS R8-34: September 2015 – August 2016; Fire_Cache_Smoke_USFS_R8-34.csv (website claims August of 1941 has data, but this is highly dubious.)
19 Smoke USFS R8-35: August 2016 – July 2017; Fire_Cache_Smoke_USFS_R8-35.csv
20 Smoke USFS R8-55: November 2016 – July 2017; **PASSWORD PROTECTED,**

CAN'T DOWNLOAD

21 Smoke USFS R8-56: November 2016 – July 2017; **PASSWORD PROTECTED, CAN'T DOWNLOAD**

22 Smoke USFS R9-15: February 2013 – April 2018; Fire_Cache_Smoke_USFS_R9-15.csv

23 Smoke USFS R9-16: February 2013 – May 2018; Fire_Cache_Smoke_USFS_R9-16.csv

24 Smoke USFS R9-17: February 2013 – May 2018; Fire_Cache_Smoke_USFS_R9-17.csv

25 Smoke USFS R9-60: October 2012 – July 2018; Fire_Cache_Smoke_USFS_R9-60.csv * This file seems to have a few rows of data shifted by a few columns. Assuming those rows could be problematic, I'll leave them as-is and let the processing scripts remove them.

26 Smoke USFS 3015: December 2015 – February 2019; Fire_Cache_Smoke_USFS_3015.csv

27 Smoke USFS 3016: May 2016 – February 2019; Fire_Cache_Smoke_USFS_3016.csv (website claims 1941 has data, but this is highly dubious.)

28 Smoke USFS R9-3017: December 2015 – February 2017; Fire_Cache_Smoke_USFS_R9-3017.csv

29 Smoke USFS R9-3018: January 2016 – February 2017; Fire_Cache_Smoke_USFS_R9-3018.csv

3. Miscellaneous Monitors

1 Smoke #25: January 2011 – September 2015; Fire_Cache_Smoke_DRI_Smoke_N25.csv * This file seems to have a few rows of data shifted by a few columns. Assuming those rows could be problematic, I'll leave them as-is and let the processing scripts remove them.

2 Smoke #86: October 2012 – February 2019 Fire_Cache_Smoke_DRI_Smoke_N86.csv

3 Smoke E_BAM 52: April 2012 – April 2015; Fire_Cache_Smoke_DRI_Smoke_E-BAM_52.csv

4 Smoke E_BAM 65: June 2013 – September 2015; Fire_Cache_Smoke_DRI_Smoke_E-BAM_65.csv * This file seems to have a few rows of data shifted by a few columns. Assuming those rows could be problematic, I'll leave them as-is and let the processing scripts remove them.

5 FWS Smoke #1: June 2012 – November 2018; Fire_Cache_Smoke_DRI_FWS_Smoke_N1.csv

6 Smoke NCFS E-BAM #1: June 2014 – December 2016; * This file has 10 fewer columns of data that most of the files, which have 34, and the ordering of the columns that are there is different. Not sure if this will cause a problem when compiling the data. Fire_Cache_Smoke_DRI_Smoke_NCFS_E_BAM_N1.csv

7 Smoke NCFS E-BAM #2: June 2014 – February 2015; Fire_Cache_Smoke_DRI_Smoke_NCFS_E-BAM_N2.csv * This file has 10 fewer columns of data that most of the files, which have 34, and the ordering of the columns that are there is different. Not sure if this will cause a problem when compiling the data. It is a short file and doesn't appear to have any days of data that will pass the quality checks (negative concentrations, too few hourly observations within a day).

8 Smoke NCFS E-BAM #3: June 2014 – February 2015; **PASSWORD PROTECTED,**

CAN'T DOWNLOAD

9 Smoke NPS Yosemite 01 California: September 2014 – Sepember 2018; Fire_Cache_Smoke_DRI_Sm

10 RSF Smoke Monitor 1: November 2015 – July 2016; Fire_Cache_Smoke_DRI_RSF_Smoke_Monitor_

11 Lolo NF Smoke Monitor #1: February 2016 – October 2018; **PASSWORD PROTECTED, CAN'T DOWNLOAD**

12 Lolo NF Smoke Monitor #2: February 2016 – April 2016; **PASSWORD PROTECTED, CAN'T DOWNLOAD**

3.1.3 $PM_{2.5}$ data from the Federal Land Manager Environmental Database

To Do

1. Check again to see if the 2018 data is available (as of February 2019, it was not available)

Data Source

- **Contact** Bret Schichtel
- **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 - Decemer 31, 2017
- **Acknowledgment** - need to fill in

We downloaded IMPROVE $PM_{2.5}$ data from the Federal Land Manager Environmental Database maintained by CIRA and Colorado State University. The IMPROVE monitors capture air quality information in more rural areas (US EPA, 2017d). We are including any of the following parameter codes: 88101, 88500, 88502, 81104 (US EPA, 2017a,c,e).

The data does not come with datum information. When processing the data, the datum is input as WGS84 per an email from Bret Schichtel on October 22, 2018.

A few sites only had only two decimal places for latitude and/or longitude. I contacted Bret Schichtel (bret.schichtel@colostate.edu), who put me in contact with Scott Copeland (scott.copeland@colostate.ed and Anthony Prenni (anthony_prenni@nps.gov). Scott Copeland sent Site_Meta_Master_10_2018.csv, a master file of the IMPROVE sites and Anthony Prenni referred me to the "Current Site List" at http://vista.cira.colostate.edu/Improve/improve-data/. The Current Site List seems to have the most decimal places for location information, but does not have all sites. This file is used first and then locations are filled in from Site_Meta_Master_10_2018.csv. Latitude and Longitude data from the main data files are ignored. See process_PM25_IMPROVE_data_source_functions.R.

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:
    1 Mass, PM2.5 (Fine): Code MF, Type PM2.5, Units ug/m$\hat{3}$ LC AQS ID 88101
    2 Mass, PM2.5 Reconstructed (Fine): Code RCFM, Type PM2.5 Units ug/m$\hat{3}$ LC, AQS ID 88401
5. Select Dates: By Years and Months: 2008-2014; select all months
6. Aggregations: Non-aggregated
7. Fields: Select All

8. When 2008-2014 data was downloaded: 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. When 2008-2017 data was downloaded: Options: Text File; Generate one file containing all the data; Comma delimited, Standard ("wide" format); Data & Metadata, String Quotes: Double Quotes, Missing Values (blank); Date Format: 3/14/2002; Display Results: In a separate browser window;

10. Submit

Repeat the downloading steps above, except replace step #2 with these Datasets and parameters:

1. IMPROVE Aerosol, RHR III (DRAFT - Preliminary Most Impaired Days dataset)

    1 Mass, PM2.5 (Fine) is listed twice

After downloading data, save each file as *_top_removed.csv and remove all rows above the main section of data (approximately 270 rows).

File Formats

csv

Original Data File Names

1. Federal_Land_Manager_IMPROVE_RHR_II_88101_201922513514530P22rMs.csv
2. Federal_Land_Manager_IMPROVE_RHR_II_88401_2019225135623221212 1t.csv
3. Federal_Land_Manager_RHR_III_88101_first_param_2019225135946946xJ0L22.csv

3.1.4 $PM_{2.5}$ data from the California State Air Quality and Meteorological Information System (AQMIS)

To Do

1. Check for availability of 2015-2018 data

Data Source

- **Contact** Denise Odenwalder, Denise.Odenwalder@arb.ca.gov
- **Citation/Link** To AQMIS: https://www.arb.ca.gov/aqmis2/aqmis2.php
- **Data (local)**
- **Geographic Extent** Whole state of California, wherever there are monitors
- **Temporal Extent** 2008-2014, daily averages
- **Acknowledgment** California Air Resources Board was very helpful in gathering and sending us this data.

Brief Description

- PM2.5 measurements at all monitoring stations in CA
- Some entries are 24-hour measurements while others are the average of hourly measurements
- One entry per 3 days

Notes

Reached out to aqmis@arb.ca.gov after determining that there was data being collected in CA that is not published on the EPA AQS website. They emailed us within a week, with a file of the data we requested.

File Formats

xlsx spreadsheet

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.

3.1.5 $PM_{2.5}$ Monitor data from Uintah Basin

To Do

1. Check to see if there is any more recent data available

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

3.1.6 $PM_{2.5}$ 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

3.1.7 $PM_{2.5}$ data from the Utah Department of Environmental Quality

To Do

1. check for data availability between 2014-2018

Data Source

- **Contact**
- **Citation/Link** http://www.airmonitoring.utah.gov/dataarchive/archpm25.htm
- **Data (local)**
- **Geographic Extent** Varies...
- **Temporal Extent** Hourly Value CSVs
- **Acknowledgment**

Brief Description

PM2.5 data from all monitoring stations in Utah

Notes

There was a lot of overlap with the EPA AQS data, so we took data only from the PM2.5 stations not reported by the EPA. This ended up being one or more of three stations (NP, HC, and RS) for 2009, 2010, 2012, and 2013.

Information about the monitoring stations: http://www.airmonitoring.utah.gov/network/Counties.htm

Meta information about monitors obtained from http://www.airmonitoring.utah.gov/dataarchive/2016DailyMaxPM25.pdf

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.

3.1.8 Processing $PM_{2.5}$ data

Below are the scripts that process and compile the $PM_{2.5}$ data. The "*" in each of the file names refers to the current "processed_data_version" (set in general_project_functions.R) since compiling the data is an ongoing process.

1. Script1_Install_Pkgs.R » install packages
2. Process_PM25_data_step1.R » compiles the various $PM_{2.5}$ data sources into a single data frame. The only eliminations of data are geographic, to remove states that are not in our study area. Update time frame of study if necessary. The output from this script is a csv file and sink .txt for each $PM_{2.5}$ data source as well as a file with all of the $PM_{2.5}$ data sources merged together ("PM25_Step1_part_*.csv"). This script takes about 6 minutes to run on a laptop.
    1 PM25_Step1_part_*.csv (the main data file)
    2 PM25_CARB_Step1_part_*.csv
    3 PM25_CARB_Step1_part_*_combining_sink.txt
    4 PM25_EPA_Step1_part_*.csv
    5 PM25_EPA_Step1_part_*_combining_sink.txt
    6 PM25_FireCacheDRI_Step1_part_*.csv
    7 PM25_FireCacheDRI_Step1_part_*_combining_sink.txt
    8 PM25_IMPRHR2MF88101_10010_Step1_part_*.csv
    9 PM25_IMPRHR2MF88101_10010_Step1_part_*_combining_sink.txt
    10 PM25_IMPRHR2RCFM88401_10010_Step1_part_*.csv
    11 PM25_IMPRHR2RCFM88401_10010_Step1_part_*_combining_sink.txt
    12 PM25_IMPRHR3MF88101_10006_Step1_part_*.csv
    13 PM25_IMPRHR3MF88101_10006_Step1_part_*_combining_sink.txt
    14 PM25_PCAPS_Step1_part_*.csv
    15 PM25_PCAPS_Step1_part_*_combining_sink.txt
    16 PM25_UintahBasin_Step1_part_*.csv
    17 PM25_UintahBasin_Step1_part_*_combining_sink.txt
    18 PM25_UtahDEQ_Step1_part_*.csv
    19 PM25_UtahDEQ_Step1_part_*_combining_sink.txt

    Notes that are useful when incorporating new data:

    1 Process_PM25_data_step1.R: change n_data_sets to higher number if adding new data source

    2 For the Federal Land Manager Database (IMPROVE) data: Download data as described in Section 3.1.3. Edit "skip_n_files" in process_PM25_parallel_wrapper_function.R so that FMLEdata_Parameter_MetaData will load the row with the header 'DatasetID, Parameter, Code, AQSCode, Units, Description' for each IMPROVE file.

    Note about flag added to data:

    1 For DRI data, put in flags for voltage data outside the range 11-17 V. (These thresholds are somewhat arbitrary, but it was noticed that when the voltage was outside this range, the $PM_{2.5}$ concentrations were often absurdly high, e.g., greater than 24,000 ug/m3.

    Data to include in next batch (f) of PM2.5 data:

1. 37 CA mobile air monitors - Charles Pearson indicated he'll email files after they are internally downloaded and approved for release
2. Fire Cache Monitors DRI (USFS) 10 sites - password protection now removed
3. 2015-2018 CARB data
4. Uintah Basin - need to check with Utah contact if there is more data now
5. UDEQ - need to check for new data
6. Any additional state data we can acquire (e.g., NV, MT) - check with Ellen

3. Process_PM25_data_step2.R » cleans the data. This script takes about 5 minutes on a laptop. This script outputs the following files:
   1. PM25_Step2_part_*.csv (main cleaned data file)
   2. PM25_Step2_part_*_sink.txt (description and summaries of the data at each step of the quality cutting)
   3. PM25_Step2_part_*_Locations.csv (list of unique locations from main cleaned data file)
   4. PM25_Step2_part_*_Locations_Dates.csv (list of unique locations/dates from main cleaned data file)
   5. Data_Removed_in_PM25_Step2_part_*.csv (list of unique locations from main cleaned data file)

   The following is a list of the quality cuts and changes made to the data:
   1. Replace "UNKNOWN" datum in EPA data with "NAD27" per Colleen's advice.
   2. Remove negative and NA $PM_{2.5}$ concentrations. This includes removing all data for a monitor on a given day if any of the hourly observations were negative.
   3. For the hourly data, remove monitor-days that do not have at least 18/24 observations.
   4. For DRI data, remove data with voltage flags (which includes flags that came with the data and flags that were put in because the battery voltage was outside the range 11-17 V.
   5. For DRI data, remove data at or below 0 L/min for flow. Think about whether a minimum value of flow should be set (higher than zero).
   6. June 6, 2014 24-hr average $PM_{2.5}$ concentration from monitor "Smoke NCFS E-BAM #1" (Fire_Cache_Smoke_DRI_Smoke_NCFS_E_BAM_N1.csv) is 24,203 ug/m3. There's nothing apparent wrong with the hourly data, however, this is the only day of data that made it through the other quality checks from this data file. This suggests that this monitor is suspect, and will be removed.
   7. Remove data points with lat/lon outside this box: (50,-126) to (25,-101). These values are defined in general_project_functions.R.
   8. Remove data outside the study period (defined in general_project_functions.R).
   9. Remove data with "Event_Type" = "Excluded"
   10. Remove data with more than 0.001 degrees variation in Lat/lon within a day
   11. Remove data from monitor "USFS R2-265" (Fire_Cache_Smoke_USFS_R2-265.csv) between October 2016 - May 2017. The concentrations (some higher than 65,000,000 ug/m3) and the behavior of the concentrations (frequently changing by exactly 1000 ug/m3 from one hour to the next) are unrealistic. The data outside this time frame look more reasonable.

12 Remove data from monitor "USFS R2-264" (Fire_Cache_Smoke_USFS_R2-264.csv") between October 2016 - October 2017. The behavior of the concentrations (frequently changing by exactly 1000 ug/m3 from one hour to the next) are unrealistic. The data outside this time frame look more reasonable.
13 Remove data from monitor "FWS Smoke #1" (Fire_Cache_Smoke_DRI_FWS_Smoke_N1.csv) between February 11-14, 2017. The behavior of the concentrations (frequently changing by exactly 1000 ug/m3 from one hour to the next) are unrealistic. The data outside this time frame look more reasonable.
14 Remove data from monitor "Smoke #22" (Fire_Cache_Smoke_DRI_Smoke_22.csv) on June 15, 2012. The behavior of the concentrations (frequently changing by exactly 1000 ug/m3 from one hour to the next) are unrealistic. The data outside this time frame look more reasonable.
15 Removing data from monitor "USFS-R2-69" during August 31, 2016 - September 5, 2016 becuase the concentrations look unrealistic and can shift by more than 10,000 ug/m3 from one hour to the next. The data outside this time frame look more reasonable.
16 Removing data from monitor "USFS-R1-307" during May 4-19, 2015 because the concentrations look unrealistic and can shift by more than 10,000 ug/m3 from one hour to the next. The data outside this time frame look more reasonable.
17 Removing data from monitor "Smoke # 216" during May 16- June 17, 2018 because the concentrations look unrealistic and are constant at 851 ug/m3 for extended hours at at time. The data outside this time frame look more reasonable.
18 Removing data from monitor " Smoke USFS R2-922" during July 26-28, 2016 because the concentrations look unrealistic and can shift by thousands of ug/m3 from one hour to the next. The data outside this time frame look more reasonable.

4. Process_PM25_data_step3.R » convert all PM2.5 data to the same datum (NAD83). Take the converted location info and put it into the data frame with the daily $PM_{2.5}$ data. This script also rounds all lat/lon info to 5 digits. This script takes about 3 minutes on a laptop. This script outputs these files:
   1 PM25_Step3_part_*_NAD83.csv (main data file)
   2 PM25_Step3_part_*_Locations_NAD83_include_old_projection.csv
   3 PM25_Step3_part_*_Locations_NAD83.csv
   4 PM25_Step3_part_*_Locations_Dates_NAD83_include_old_projection.csv
   5 PM25_Step3_part_*_Locations_Dates_NAD83.csv
5. Process_PM25_data_step4_parallel.R » composite replicate data and data where there are co-located monitors. This script produces two different versions of the data:
   1 PM25_Step4_part_*_de_duplicated_aves_ML_input.csv takes the averages of all available co-located data
   2 PM25_Step4_part_e_de_duplicated_aves_prioritize_24hr_obs_ML_input.csv prioritizes data that was originally a 24-hour observation (typically FRM/filter-based measurements) over the data that was an average of hourly observations.

   Calls these functions:
   1 PM25_station_deduplicate_aves_parallel.fn

2 PM25_station_deduplicate_aves_parallel.fn

3 prioritize_daily_obs_over_hourly.fn (only called if de_duplication_method is set to "prioritize_24Hour_Obs"

4 fill_input_mat_aves.fn

i. concatinate_within_column_function.R

ii. concatinate_vector_of_strings.fn

6. Process_PM25_data_create_report.R » map locations of monitors by data source/year

3.1.9 Determine which dates/locations are new to most recent "processed_data_version"

Separate_Locations_Dates_by_processed_data_version.R » Take difference between parts d and b to find what locations/dates are only in part d. This script takes a few minutes on a laptop.

1. part a: early version created while writing code. Disregard.
2. part b: first batch of PM2.5 data that was used to extract predictor data, years 2008-2014
3. part c: county centroids, 2008-2014. This work flow has now been moved to the "Locations_of_interest".
4. part d: second batch of PM2.5 data, adds AQS data for 2015-2018.
5. part e: updates IMPROVE and Fire Cache data to include 2008-2018 (whatever portion of that was available for download)

3.1.10 Notes about very high data points

All files with daily average concentrations above 1000 ug/m3 were individually inspected. The following Fire Cache monitors have concentrations above 1000 ug/m3 and were kept because the file looked ok (at least there was nothing obviously wrong). (Some files shifted hourly concentrations in increments of 1000 ug/m3 and those were removed as described in Step 2 above.)

1. RSF Smoke Monitor 1
2. Smoke 215

3.2 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. Step 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. Step 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.

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

   Run script ‘modis_aod_create_csv_file.py‘. This script will take all the .hdf files that you have downloaded and store the lat, long and aod value for non-null pixels from the ‘Deep_Blue_Aerosol_Opt SDS. A .csv file will be created for each corresponding .hdf file.

4. Step 4: Create .shp file for each .csv file

   Run ‘modis_aod_convert_csv_to_shapefile.py‘. This script will read in the .csv files and convert them to .shp files using multiprocessing, which speeds up the process.

5. Step 5: Project .shp files to US Albers Equal Area Conic

   Run ‘modis_aod_project_to_albers.py‘. This script will reproject the .shp files to be US Albers Equal Area Conic (ESRI:102003).

6. Step 6: Combine .shp files for same date and convert to raster with 10km resolution

   Run ‘modis_aod_create_daily_averages.py‘. This will combine all .shp files from the same date and then produce a raster for each with a 10km resolution. Then, the interpolated grids are clipped to the 11 western states (our study area) with a 100km buffer.

7. Step 7: Extract MODIS AOD value at EPA monitor locations

   Using ExtractValuesToPoints tool in ArcGIS.

Quality Control

Script Names

1. modis_aod_create_csv_file.py
2. modis_aod_convert_csv_to_shapefile.py
3. modis_aod_project_to_albers.py
4. modis_aod_create_daily_averages.py

Data File Names

1. western_states_merge.shp

3.3 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

websites: https://www.ncdc.noaa.gov/data-access/satellite-data/satellite-data-access-datasets
https://www.ncdc.noaa.gov/data-access/satellite-data/satellite-data-access-datasets
Order form for data: https://www.ncdc.noaa.gov/has/has.dsselect
https://www.ncdc.noaa.gov/doclib/index.php?choice=dsi&searchstring=3635&submitted=1&submitted=Search

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1. Download Data

   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'. Select GOES-West for satellite ID, GASP-AOD-GZ for data type, and appropriate start and end date. Select "Yes" for Submit Batch. Enter email address and submit order. You

will get emails later on with FTP links to your data. Run ‘Generic_FTP_download_to_S3.py‘ on an EC2 instance passing in the FTP url as the argument. This will download the data and upload it to S3 (and then delete it off the EC2 instance).

Quality Control

Script Names

1. Generic_FTP_download_to_S3.py

Data File Names

1.

3.4 Hazard Mapping System (HMS)

Data Source

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

Brief Description

See **?** and references therein (especially Rolph et al 2009) to see descriptions of HMS data.

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.

3.5 MERRA-2

Data Source

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

3.6 MAIAC

Data Source

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

Brief Description

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1.

Quality Control

Script Names

1. Contacted NASA DeepBlue team via email and was given the FTP site for their research data output. Public data set not yet available. But should be in several months under the name ‘MCD19‘.

Data File Names

1. n/a

3.7 MODIS Thermal Anomalies/Fire Daily L3 Global 1km (MCD14DL)

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

.shp

Data Filtering and Processing

Final Variable(s)

Methods

1. Navigate to the NASA EarthData FIRMS Archive Download site
2. Select "Create new Request"
3. In the dropdown for region, select "Custom Region" and draw a bounding box around study area
4. In the dropdown for fire data source, select "MODIS C6"
5. Select dates for study time period
6. In the dropdown for file type, select "Shapefile (.shp)"
7. Enter your email address
8. You will get an email with a download link containing a zipfile with the data
9. Run active_fire.py with the required arguments. This script will spatially join the fire data with the timezone data. This is a necessary step for adjusting the timestamp from UTC to local in the next step.

10. Create fields "ACQ_TIME" and "ACQ_DATE" in ArcMap, QGIS, or any other method of your choosing in the output shp file from the previous step. Extract the time and date from the "adj_time" column to populate these new fields.

11. Run buffers.py with the required arguments. This script will take in a csv file with lat, lon, and dates (the PM2.5 stations/points of interest), a shp file with buffers around each of these observation (specified as arguments), the fire shp file that was edited in the previous step, and an output csv filename. The output is a csv file similar to the input csv, but with an additional columns that includes the number of active fires in each buffer. Note: the buffer shp file used in this study was created with geodesic buffering due to the large buffer sizes used (25km, 50km, 100km, 500km radiuses). We decided that it would be better to use geodesic buffering for large areas (as opposed to planar buffering) because the they will not be affected by distortion introduced from the projection, which is a bigger issue with larger areas.

12. Run merge_with_zeros.py with the required arguments. This merges the dataframe created in the previous step with the original dataframe of station locations and dates with over 1 million rows. It matches the rows with fire counts and give a value of 0 for all else.

Quality Control

Script Names

1. active_fire.py
2. buffers.py
3. merge_with_zeros.py

Data File Names

1. timezones_western_us.json

3.8 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. BAECV data set already downloaded by EarthLab fire group. Navigate to the 'earthlab-ls-fire' S3 bucket, then the v1.1 subdirectory. Here you will find yearly .tar.gz files. Have not spent time decompressing files and exploring data yet but my guess is that within each yearly file, we will find more detailed, daily burn data.

Quality Control

Script Names

1. n/a

Data File Names

1. n/a

3.9 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 'MODIS_FTP_download.py' 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_ Update: 'MODIS_FTP_download.py' is obsolete because NASA LAADS decomissioned their FTP site in favor of HTTPS. So, a new all-purpose script will need to be written to do this download that does HTTPS retrievals instead.

Quality Control

Script Names

1. MODIS_FTP_Download.py

Data File Names

1.

3.10 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. Navigate to the NASA EarthData FIRMS Archive Download site
2. Select "Create new Request"
3. In the dropdown for region, select "Custom Region" and draw a bounding box around study area
4. In the dropdown for fire data source, select "VIIRS"
5. Select dates for study time period
6. In the dropdown for file type, select "Shapefile (.shp)"
7. Enter your email address
8. You will get an email with a download link containing a zipfile with the data
9. Progress stopped here, as we chose to proceed with the MODIS Thermal Anomalies dataset for the active fire input for the project as of Fall 2018. But, follow along in the steps for

the MODIS Thermal Anomalies workflow to continue. The steps are the same, as the data comes from the same source.

Quality Control

Script Names

1. n/a

Data File Names

1. fire_archive_V1_2770.csv

3.11 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. Navigate to the National Map Viewer and find products for "National Land Cover Database (NLCD)" at the National extent. From the search results, download "NLCD 2011 Land Cover (2011 Edition, amended 2014)". This will download a zipfile with the data.

2. Run nlcd_process.py with the required arguments. This script computes zonal statistics between a buffer shp file and an classified raster tif (in our use case, a reclassified NLCD raster). The computed value is percent area of developed high density land cover in each buffer. The output is another csv, which is the input csv with an an extra column denoting the data. Note: the buffer shp file used in this study consisted of 1km, 5km, and 10km radius buffers using planar buffering.

Quality Control

Script Names

1. nlcd_process.py

Data File Names

1. n/a

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

   Using the NASA EarthData online search tool, search for the 'MOD10A1' (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 'MYD10A1' (Aqua) data set. All of the raw downloaded data from this step will be in .hdf file format.

Quality Control

Script Names

1.

Data File Names

1.

3.12.1 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 arc-second. This resolution is approximately 30 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=3DEPViewNational Map Viewer site and find products for Elevation Products (3DEP), 1 arc-second DEM, IMG file format. Once results are returned, select "Save as Text", which will download a text file containing server links to each NED tile.

2. Download the data using the download_tiles.py script, which will access the text file that you just downloaded.

3. Extract the elevation values using the extract_values_to_points.py script. Step-by-step instructions:

```
1  Start t2.medium EC2 instance on AWS with 250 BG storage
2  docker pull earthlab/estimate-pm25
3  docker run -d earthlab/estimate-pm25
4  docker ps
5  docker exec -it [docker_name] /bin/bash
6  aws configure
7  mkdir NED
8  cd NED
9  mkdir data
10 aws s3 cp s3://earthlab-reid-group/NED/data/ /home/jovyan/NED/data/ –recursive
```

i. This step takes a few hours

11 aws s3 cp s3://earthlab-reid-group/Processed_Data/PM25_Locations_Dates/ /home/jovyan/NED/ –recursive

12 git clone https://github.com/earthlab/estimate-pm25.git

13 cd estimate-pm25

14 cd download-earth-observations/NED/

15 python extract_values_to_points.py –NED_directory "/home/jovyan/NED/data/" –input_csv_file "/home/jovyan/NED/Part_e_not_in_bd_Locations.csv" –output_csv_file "/home/jovyan/NED/ned_pa

16 See also: https://docs.google.com/document/d/1IysKoAS8l1WH6nN4IksWe9MWqNZEOG5DQD9DY edit?usp=sharing

Quality Control

Script Names

1. download_tiles.py
2. extract_values_to_points.py

Data File Names

1. n/a

3.13 MODIS Normalized Difference Vegetation Index (MOD13A3)

Data Source

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

Brief Description

MODIS NDVI description

Notes

File Format

Data Filtering and Processing

Final Variable(s)

Methods

1. Download data using the MODIS_HTTPS_download.py script with required arguments.
2. Translate, mosaic, and reproject data using the translate_mosaic_reproject.py script with required arguments.
3. Extract to points using the extract_to_values.py script with required arguments.

Quality Control

Script Names

1. MODIS_HTTPS_download.py
2. translate_mosaic_reproject.py
3. extract_to_values.py

Data File Names

1. n/a

3.14 Meteorological Data

Data Source

North American Mesoscale, Analysis (NAM)

- **Contact**
- **Citation/Link** https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesc https://nomads.ncdc.noaa.gov/data/namanl/
- **Geographic Extent** North America
- **Temporal Extent** Available March, 2004 - present with slight delay
- **Acknowledgment**

Brief Description

We will obtain meteorological data from the North American Mesoscale, Analysis (NAM) because it includes all of the standard meteorological variables, including 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 6-hourly data for temperature, relative humidity, sea level pressure, surface pressure, planetary boundary layer height, dew point temperature, precipitation, snow coverage, and the U and V components of wind speed. NAM has 12 km resolution and is available 2004 onward.

Notes

File Format

Prior to 2018, the files are in *.grb ("grib1") format, while 2018 data is in *.grb2 ("grib2") format.

Resources about this file type:

- rNOMADS is an R package for accessing grb* files. It is mostly geared for grib2 files. https://cran.r-project.org/web/packages/rNOMADS/rNOMADS.pdf
- Explanation of what grib files are: http://www.cpc.ncep.noaa.gov/products/wesley/reading_grib.html,
- wgrib program information: http://www.cpc.ncep.noaa.gov/products/wesley/wgrib.html

Data Filtering and Processing

1. Use the earthlab/r-reidgroup docker image, which has wgrib and wgrib http://www.cpc.ncep.noaa.gov/products/wesley/wgrib.html and wgrib2 http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/ installed on it.
    1 Start EC2-session on AWS (using m5a.xlarge session with 30 GB attached memory)
    2 Install and start the earthlab/r-reidgroup docker image, see https://hub.docker.com/r/earthlab/r-reidgroup/. See also https://docs.google.com/document/d/1hQdZgbnwwBMACzVvTOqdjA27JgyL4EtNyCyVa4/edit?usp=sharing
2. Process_NAM_data_step1.R reads in locations file and outputs the '*_NextDay' csv file, which includes the next day for each location/day listed in the first file. The purpose of this is so all of the necessary NAM files can be processed. UTC dates can go into the next day

for western US time zones. This script runs in serial and takes approximately 6 minutes to run. This step uses these input files and R packages and functions:

1 Input files:
   i. PM25_Step3_part_*_Locations_Dates_NAD83.csv
   ii. CountyCentroid_Locations_Dates_2008-01-01to2018-12-31.csv
2 Packages:
   i. (none)
3 Files with custom functions:
   i. general_project_functions.R
   ii. NAM_processing_functions.R
4 Output files:
   i. NAM_Step1_part_*_Locations_Dates_PM25_Locations_Dates_wNextDay.csv
   ii. NAM_Step1_part_e_Locations_Dates_CountyCentroid_wNextDay.csv

3. Process_NAM_data_step2_parallel.R downloads each NAM file, extracts relevant data, and deletes the original NAM data. (All of the NAM files together would be about 1.6 Tb.) This file operates in parallel, and will use n-1 cores, where n is the number of cores on the computer. The output is 1 csv with all locations of interest for a given date and time step. The time steps for the NAM are 0Z, 6Z, 12Z, and 18Z. The output files have the format Locations_Dates_of_PM25_Obs_DeDuplicate_YYYY_MM_DD_XXUTC.csv where XX refers the to the timestep. Change the study start and stop dates for the dates to be processed. This step uses these input files and R packages and functions:

1 Input files:
   i. NAM_Step1_part_*_Locations_Dates_PM25_Locations_Dates_wNextDay.csv
   ii. NAM_Step1_part_e_Locations_Dates_CountyCentroid_wNextDay.csv
2 Packages:
   i. rNOMADS
   ii. parallel
3 Files with custom functions:
   i. general_project_functions.R
   ii. NAM_processing_functions.R
   iii. extract_NAM_data_parallel_function.R
   iv. define_project_bounds_function.R
   v. loop_NAM_run_times.parallel_function.R
   vi. merging_data_functions.R
4 Output files:
   i.

ii.

1 Locations_Dates_of_PM25_Obs_DeDuplicate.csv - Data file with dates (local) and locations where you want the NAM data

2 MeteoVariablesNAM.csv - listing of meteorological variables to be extracted from NAM data

3 rNOMADS R package (which calls wgrib and wgrib2) https://cran.r-project.org/web/packages/rNOMADS/rNOMADS.pdf

4 parallel R package

5 grb1to2_conversion_prep_function.R - This script downloads the files that will be necessary to run grb1to2.pl, created by the Climate Prediction Center http://www.cpc.ncep.noaa.gov/products/wesley/grb1to2.html

6 loop_NAM_run_times.parallel_function.R - this function loops through the time steps on a given day and calls function (listed below) to extract meteo data at locations of interest

7 define_project_bounds_function.R - the bounding box for the study area is defined in this function. The scripts can run faster if the entire NAM domain does not need to be loaded into memory.

8 extract_NAM_data_parallel_function.R - this function extracts the NAM data at points of interest

9 which_type_of_grib_file_function.R - this function determines whether the data for a given time step are grib1 or grib2 format

10 convert_grib1to2_function.R - convert file type from grib1 to grib2, unless it's already a grib2 file. This is essentially a wrapper for grb1to2.pl created by the Climate Prediction Center http://www.cpc.ncep.noaa.gov/products/wesley/grb1to2.html

4. Process_NAM_data_step3.R merges all of the files from step 2 into a single file and adds a column at the beginning giving the UTC time stamp. This script will take approximately 20 minutes (runs in serial).

5. Process_NAM_data_step4.R Add time zones and local times.

6. Process_NAM_data_step5.R merges the 4 time steps to give a 24-hr summary. Min, max, mean, etc. is set in MeteoVariabesNAM.csv. This script is expected to take a little over an hour on 16-cores.

Final Variable(s)

See MeteoVariablesNAM.csv

Quality Control

3.15 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.

3.16 Locations of Interest

This section describes the code for identifying and compiling the predictor variables for locations of interest, such as county centroids.

3.16.1 County Centroids

1. CountyCentroid_CreateLatLonDateFiles.R » Create two csv files in the /home/Processed_Data/CountyCentroid/ folder. This script takes approximately 6 minutes to run on a laptop.

    1 Locations of county centroids for the study area: CountyCentroid_Locations.csv ($\sim$ 30 KB)

    2 Locations of county centroids for study area expanded across all dates in study period: CountyCentroid_Locations_Dates_[Study Start Date]to[Study End Date].csv ($\sim$ 140 MB)

2. CountyCentroid_PlotLocations.R » Plot centroid locations and create summaries of the locations-only and dates-locations centroids files in CountyCentroid_Locations_File_Summary.txt, which is stored in the same folder as the data. This script takes a few seconds to run on a laptop.

3.16.2 Population-weighted county centroids

1. Extract_county_pop_centroids.R **To Do:** update this code to have a similar process as the geographic county centroid codes listed above.

3.17 Phase 2: Extraction to Observation Locations and Points of Interest

3.18 Phase 3: Merge extracted data

This phase both the input file for the ML training and the input file for the predictions to points of interest.

3.18.1 Predictor input files for points of interest

Once the predictor variables have been extracted to the points of interest, these need to merged into a single input file for each type of point of interest (e.g., geometric county centroid, population-weighted county centroid, etc.).

1. ML_PM25_estimation_merge_predictors.R » Merge the various predictor variables together with the monitor data or dates/locations of interest.

2. ML_PM25_estimation_plot_predictors.R » Plot the training input file

    1 predictor variables vs date

    2 predictor variables vs $PM_{2.5}$

1. Merge_predictors_to_points_of_interest.R » Merge the predictor variables to the locations of interest for each set of points of interest. The file names for the source files will need to be updated as more predictor data is processed. This script takes about 2 minutes on a laptop. This script calls this function:

    1 Merge_predictors_to_points_of_interest_parallel_wrapper_function.R

2. Plot_Predictor_Inputs.R » Plot prediction input files that were created with above script. This script takes several minutes to run on laptop.

3.19 Phase 4: Machine Learning Methods

3.20 ML Techniques and Calculations

Need to describe how R^2 is calculated.

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.

3.21 ML Scripts

1. ML_PM25_estimation_step1.R » ML training algorithms
2. ML_PM25_estimation_step1.R » create data frame of the dates/locations for which we want to predict PM2.5

3.22 Phase 5: Predictions to Points of Interest

4 Results

5 Discussion

Discuss a comparison of temporal trends between our work and O'Dell et al ?.

6 Ideas, To Do, Resources, etc

code up fires by type of land coverage

Consider using the work of Westerling et al for a comprehensive fire history (up through 2012) http://science.sciencemag.org/content/313/5789/940, http://www.pnas.org/content/108/32/13165, http://rstb.royalsocietypublishing.org/content/371/1696/20150178 Westerling (2016b,a) Also look into the fire histories referenced in Westerling Westerling (2016b,a): http://fam.nwcg.gov/fam-web/weatherfirecd/fire_files.htm and http://fam.nwcg.gov/fam-web/kcfast/mnmenu.htm See also http://www.nifc.gov

look into the Fire and Smoke Model Evaluation Experiment (FASMEE) http://www.fasmee.net

Compare our results with EPA Downscaler https://www.epa.gov/air-research/downscaler-model-predicting-

Look at Kollanus et al. (2016) again for references for PM2.5 paper, especially the introduction. Consider using NAAPS in our study.

read ?

read ?

see also https://www.5280.com/2018/09/can-colorado-burn-its-way-out-of-a-wildfire-crisis/

Could we use inciweb to distinguish prescribed fires?

look up Global Fire Emissions Database (GFED3) - maybe it would be useful for our study as an input to the machine learning? see Liu et al. (2016)

see ? for potential data sources for ML project

emissions vary by temperature https://cires.colorado.edu/news/wildfire-temperatures-key-better-understand and https://www.atmos-chem-phys.net/18/9263/2018/

read Monitoring Trends in Burn Severity MTBS, 2014. Data Access: Fire Level Geospatial Data. US Department of Agriculture, Forest Service and US Department of Interior, Geological Survey. http://mtbs.gov/data/individualfiredata.html/.

Idea: look at ambulance calls and PM2.5, similar to what Salimi et al. (2016) did in Australia.

read ?

Database of planned/proposed prescribed burns: WRAP's Fire Emissions Tracking System: http://wrapfets.org/index.cfm

See Di et al., 2016 and Johnston et al., 2012, Rappold et al., 2014 in ? - combine modelled and monitored/satelite data to estimate PM2.5

See page 11 of ? for discussion of discrepancies related to burned area estimates

http://www.ptep-online.com/ctan/symbols-a4.pdf

US National Atlas http://nationalmap.gov/small_scale/atlasftp.html

Thought: Using DigitalGlobe for fire data compared to NASA: would have higher spatial resolution, but not consistently viewing all areas (no cost to CU people)

Look up Openair R package

Papers/resources to look into: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1293

https://www.fs.fed.us/psw/publications/4451/psw_2009_4451-001.pdf

https://labcit.ligo.caltech.edu/~ethrane/Resources/UNIX/

https://community.tableau.com/thread/141548

According to ?, GEOS-Chem "can be classified according to emission source", that implies that we could tag the emissions as wildfire vs prescribed fire vs urban. Would there be any advantages of this model over CAMx?

could analyze data with NAAQS and WHO PM2.5 standards

projection/datum info: https://gis.stackexchange.com/questions/664/whats-the-difference-between-a-projec http://resources.esri.com/help/9.3/arcgisengine/dotnet/89b720a5-7339-44b0-8b58-0f5bf2843393.htm http://grindgis.com/blog/wgs84-vs-nad83

Monitoring Trends in Burn Severity (MTBS) MTBS, 2016: Data Access: Fire Level Geospatial Data. USDA Forest Service/U.S. Geological Survey, accessed 8 October 2016, https://mtbs.gov/direct-download. Eidenshink, J., B. Schwind, K. Brewer, Z.-L. Zhu, B. Quayle, and S. Howard, 2007: A project for monitoring trends in burn severity. Fire Ecol., 3, 3–21, https://doi.org/10.4996/fireecology.0301003.

Idea: Maybe instead of just distance to closest fire, we should follow the example of [Baek2016] and do distributed lags with concentric circles with information about fires in each concentric circle... also, instead of just distance to fire, maybe we could come up with a variable that is something like [distance*size of fire] since both are important.

Fire stats/records: https://www.nifc.gov/fireInfo/fireInfo_statistics.html

See **?** for description of fire perimeter data that perhaps we could use (CA only)

See **?** for info about MTBS and Active Fire Mapping Program and NWS smoke products. See also Lassman et al **?** cited therein.

Read these papers cited in **?**: Yao and Henderson, 2014; Henderson et al 2011; Liu et al 2015; Gan et al 2017; and look at their sources of PM2.5 data to see if we could add any of those to our project.

7 PM2.5 Surface Paper Notes

Discussion of trends in anthro PM2.5: **?**

7.1 Papers published in Atmospheric Environment - use as style example

Need to go through these papers

- Brokamp et al. (2017) (partially done, done through intro)
- Sampson et al. (2013)
- Anyenda et al. (2016)
- Torvela et al. (2014)
- Whiteman et al. (2014)

Put in Brokamp et al. (2017); Larsen et al. (2017)

8 Papers to cite/discuss in Introduction and/or Discussion

Westerling (2016b,a)

try to find English version http://80.24.165.149/webproduccion/PDFs/15CAP03.PDF

For fire identification, consider using NOAA's Hazard Mapping System and BlueSky

8.1 Notes on Papers

See J. et al. (2016) for statistics about wildfires in western US, e.g., % started by humans, number of fires, etc.

9 Fire attribution paper

revisit **?**

include **?** - does a good job of summarizing the debate about more vs less prescribed burns

sources of fire data **?**, **?**

will need to compare our work to **?**

include Westerling (2016b,a) and Abatzoglou and Williams (2016)

See **?** for an alternative method of attributing PM2.5 to wildfire smoke (instead of CAMx)

See Le et al 2014 **?**

See Huff et al **?**

See https://eos.org/articles/new-eyes-on-wildfires?utm_source=eos&utm_medium=email&utm_campaign=EosBuzz050319

9.1 text written for the COPD paper - variation of this may be useful

Larsen et al., 2017 Larsen et al. (2017) found that, on average, ground-level $PM_{2.5}$ concentrations increased by 2.9 $\mu g \cdot m^{-3}$ (2.8, 3.0) when there was a visible wildfire smoke plume overhead (from satellite imagery), as well as a 2.6 ppb (2.5-2.7) increase in O_3. Satellite data provides a wealth of data and can provide information about air quality where monitors are not present. However, satellite imagery inherently comes with a substantial uncertainty in that satellite data describes the entire atmospheric column and not specifically just air pollution at the ground level, where people are breathing.

References

Abatzoglou, J. T. and Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42):11770–11775.

Anyenda, E. O., Higashi, T., Kambayashi, Y., Thao, N. T. T., Michigami, Y., Fujimura, M., Hara, J., Tsujiguchi, H., Kitaoka, M., Asakura, H., Hori, D., Yamada, Y., Hayashi, K., Hayakawa, K., and Nakamura, H. (2016). Exposure to daily ambient particulate polycyclic aromatic hydrocarbons and cough occurrence in adult chronic cough patients: A longitudinal study. *Atmospheric Environment*, 140(Supplement C):34 – 41.

Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., and Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environment*, 151:1–11.

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.

J., F. E., T., A. J., K., B. J., T., F. J., and A., B. B. (2016). Quantifying the human influence on fire ignition across the western usa. *Ecological Applications*, 26(8):2390–2401.

Kollanus, V., Tiittanen, P., Niemi, J. V., and Lanki, T. (2016). Effects of long-range transported air pollution from vegetation fires on daily mortality and hospital admissions in the Helsinki metropolitan area, Finland. *Environ Res*, 151:351–358.

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.

Larsen, A. E., Reich, B. J., Ruminski, M., and Rappold, A. G. (2017). Impacts of fire smoke plumes on regional air quality, 2006-2013. *Journal of Exposure Science & Environmental Epidemiology*.

Liu, J. C., Wilson, A., Mickley, L. J., Dominici, F., Ebisu, K., Wang, Y., Sulprizio, M. P., Peng, R. D., Yue, X., Anderson, G. B., and Bell, M. L. (2016). Wildfire-specific Fine Particulate Matter and Risk of Hospital Admissions in Urban and Rural Counties. *Epidemiology*, 28:77–85.

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.

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.

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.

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.

Salimi, F., Henderson, S. B., Morgan, G. G., Jalaludin, B., and Johnston, F. H. (2016). Ambient particulate matter, landscape fire smoke, and emergency ambulance dispatches in Sydney, Australia. *Environ Int*.

Sampson, P. D., Richards, M., Szpiro, A. A., Bergen, S., Sheppard, L., Larson, T. V., and Kaufman, J. D. (2013). A regionalized national universal kriging model using partial least squares regression for estimating annual pm2.5 concentrations in epidemiology. *Atmospheric Environment*, 75:383 – 392.

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.

Torvela, T., Tissari, J., Sippula, O., Kaivosoja, T., Leskinen, J., Virén, A., Lähde, A., and Jokiniemi, J. (2014). Effect of wood combustion conditions on the morphology of freshly emitted fine particles. *Atmospheric Environment*, 87(Supplement C):65 – 76.

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, 2017b). *Outdoor Air Quality Data Download Daily Data*. https://www.epa.gov/outdoor-air-quality-data/download-daily-data.

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

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

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

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.

Westerling, A. L. (2016a). Correction to 'increasing western us forest wildfire activity: sensitivity to changes in the timing of spring'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371(1707).

Westerling, A. L. (2016b). Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring. *Philos Trans R Soc Lond B Biol Sci*, 371(1696). bibtex: westerling_increasing_2016.

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.