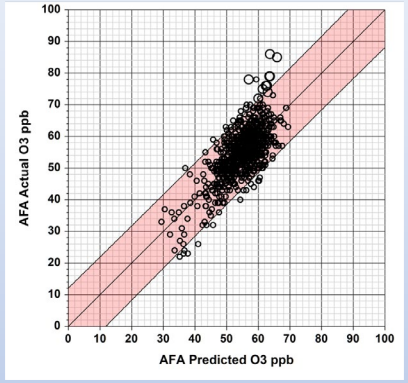
Using Cluster Analysis to Segregate Days Influenced by Smoke from Normal High Ozone Days and to Estimate Smoke Contributions to Ozone

National Exceptional Events Workshop * Hosted by the Western States Air Resources Council

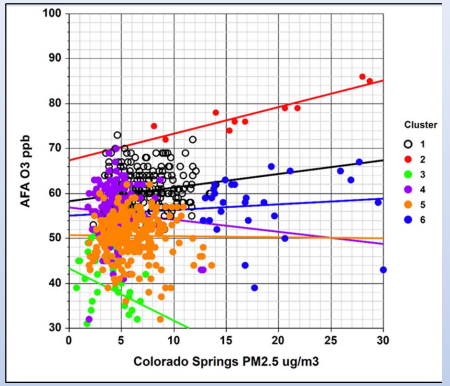
February 27-29, 2024

St. Louis, MO



Patrick Reddy

preddyresearch@gmail.com



PATRICK J. REDDY
ATMOSPHERIC SCIENTIST

Satellite, Air Quality, & Climate Data Analyst
Providing services in air quality, atmospheric, and environmental sciences

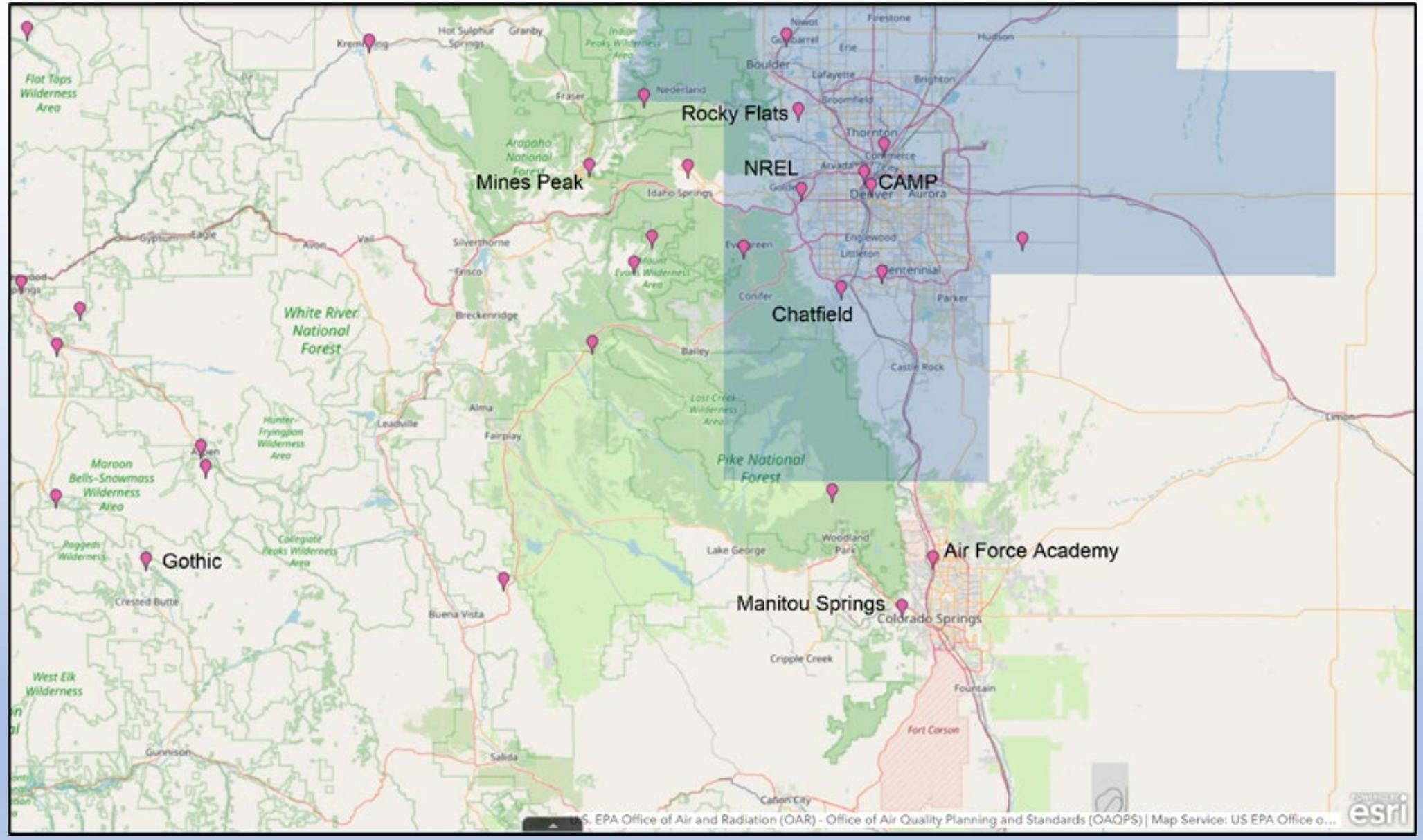
P.O. Box 451, Crestone, CO, 81131
preddyresearch@gmail.com

My Background:

- Forecasting, Meteorological Analyses, and Exceptional Event Analyses for the Colorado Department of Public Health & Environment, Air Pollution Control Division: 1990 - 2022.
- Independent/Freelance Research Scientist
- NASA Health and Air Quality Applied Sciences Team Ambassador, 2023 - present.
- www.linkedin.com/in/patrick-reddy-0981a8284

In This Presentation:

- I will present K-means cluster analysis and suggest how this could be used for comparisons between smoke days and non-smoke days with comparable meteorology.
- I will propose a way to estimate smoke contributions to ozone using cluster analysis results.
- I will demonstrate the use of K-means cluster analysis for screening and flagging wildfire smoke events affecting surface ozone concentrations in Colorado Springs, Colorado, focusing on the summer of 2020.



O_3 Sites in the Colorado Springs Region

Predictive Models and K-Means Cluster Analysis Can be Used for Event Screening and Weight-of-Evidence Analyses for an Exceptional Event Demonstration

- Predictive models can include traditional multiple linear regression or Generalized Additive Models (GAMs). (see Gong et al., 2017, https://pubs.acs.org/doi/10.1021/acs.est.7b03130, and Flynn et al., 2021, https://doi.org/10.1525/elementa.2020.00104).
- The predictive modeling process or a correlation analysis identifies meteorological variables suitable for cluster analyses.
- I have chosen multiple linear regression using Reddy and Pfister, 2016, as guidance for selecting meteorological variables (see: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015JD023840).
- K-means cluster analyses can be based on daily max 8-hour O_3, 24-hour PM2.5, satellite aerosol optical depth (AOD) and column carbon monoxide, and daily meteorological variables.
- The cluster analysis is an iterative process that involves testing the numbers of clusters needed to segregate normal high-O_3 days and days likely to be affected by smoke.
- See slide 15 for suggested data sources.

Simple Cluster Analysis for Colorado Springs Area Sites 2016-2020

Variables	Cluster 1 Hot, high heights, weak westerlies, high ozone	Cluster 2 Hot, high heights, weak westerlies, smoke, high Pm2.5, highest ozone	Cluster 3 Cool, weak westerlies, low heights, low ozone	Cluster 4 Moderately cool, low heights, strong westerlies aloft, moderate ozone	Cluster 5 Warm, low heights, strong westerlies, low ozone	Cluster 6 Warm, high heights, moderate westerlies, high Pm2.5, moderately high ozone
Colorado College PM25 (ug/m3)	6.95	17.83	3.89	4.45	5.90	19.52
Flagged for Smoke	No	**Yes**	No	No	No	No
AFA ozone ppb	60.5	78.0	39.2	55.8	50.6	57.8
Manitou Springs ozone ppb	59.2	77.0	38.2	55.3	49.0	55.6
Denver Grid 500 mb Heights (meters)	5915.9	5915.6	5705.0	5762.8	5874.0	5890.3
Next day's Four Corners Grid 500 mb Heights (meters)	5928.1	5924.1	5738.9	5780.6	5883.2	5906.7
Denver Grid 700 mb Temperature (deg F)	57	59	34	44	53	55
Denver Grid 700 mb Zonal Wind (m/s)	1.98	2.06	1.86	4.60	4.00	3.76
Colorado Springs max temperature (deg F)	88	91	57	75	84	82
Count	245	10	61	164	253	32

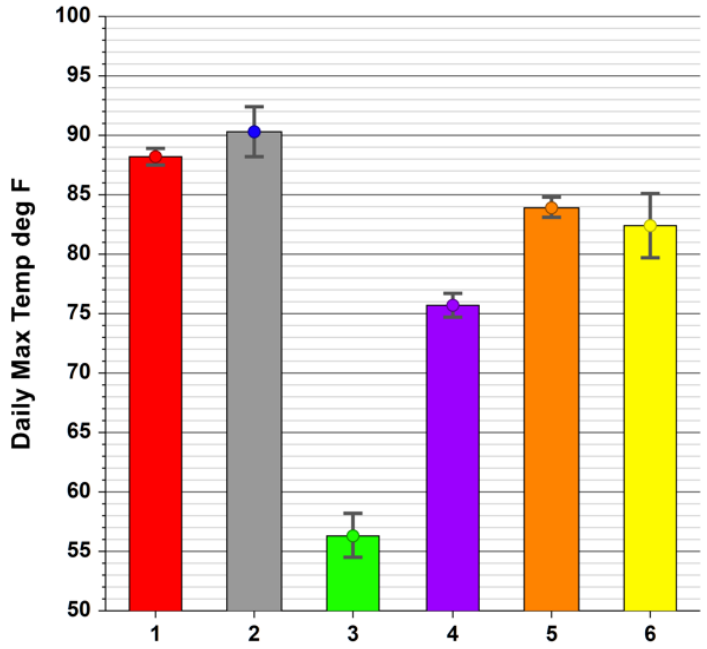
Cluster analysis of 8 meteorological and air quality variables and CDPHE smoke event flags for May-September 2016-2020.

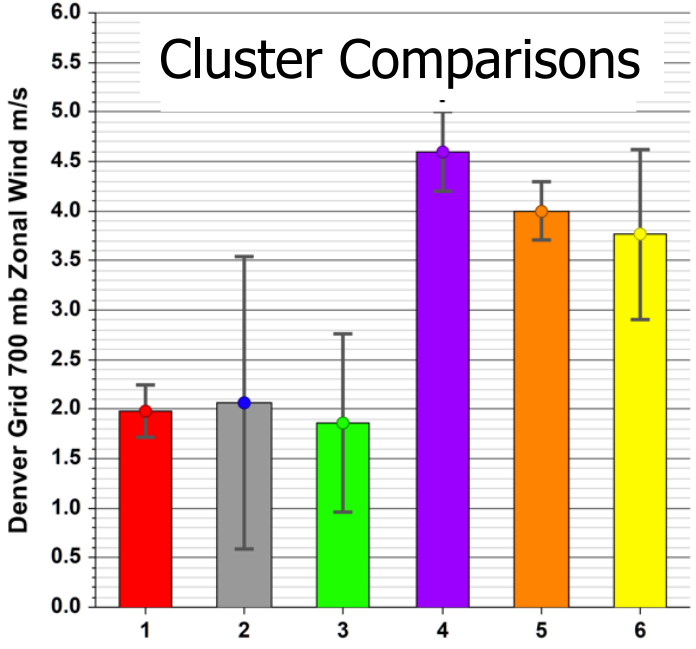
Cluster 1 represents normal high-O_3 weather conditions. Cluster 2 represents nearly identical conditions, but all 10 members were flagged and PM2.5 was high. ***Mean O_3 differences ~ 18 ppb.***

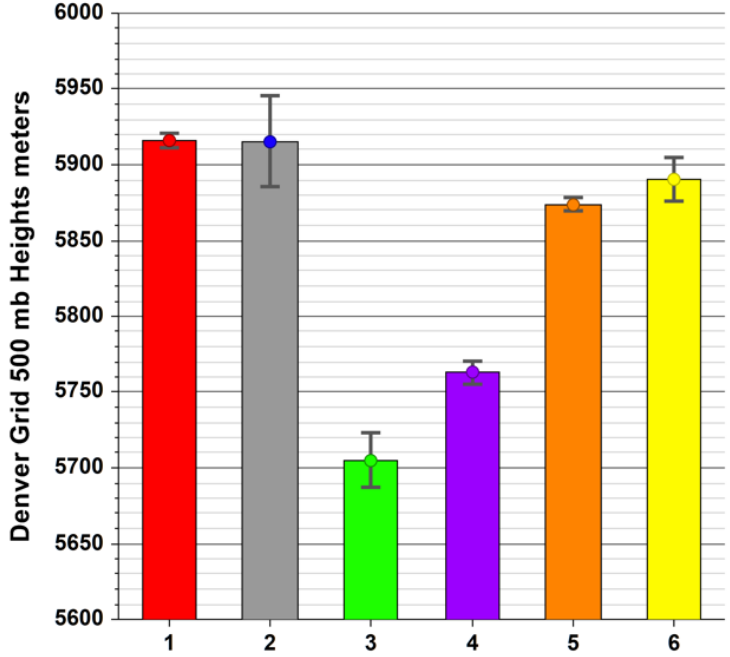
Cluster 6 represents cooler, windier conditions and includes September 2020 smoke events.

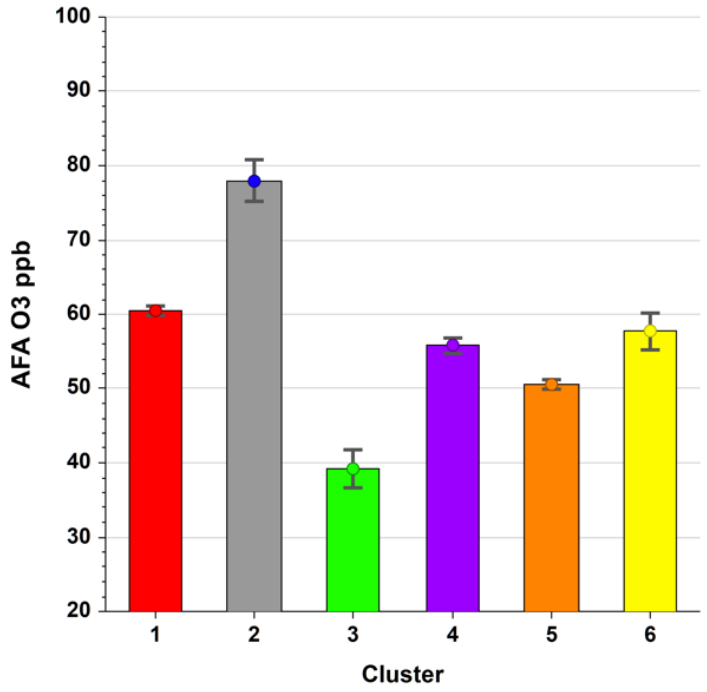
Comparison of clusters 1 and 2 takes the place of comparing event days with a limited number of similar days, and it is a more robust way to look at the differences.

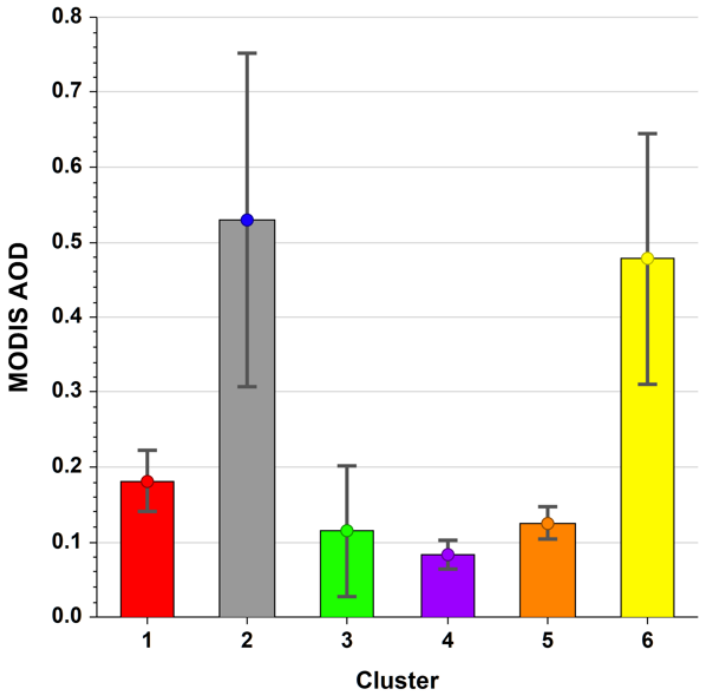
Cluster Comparisons

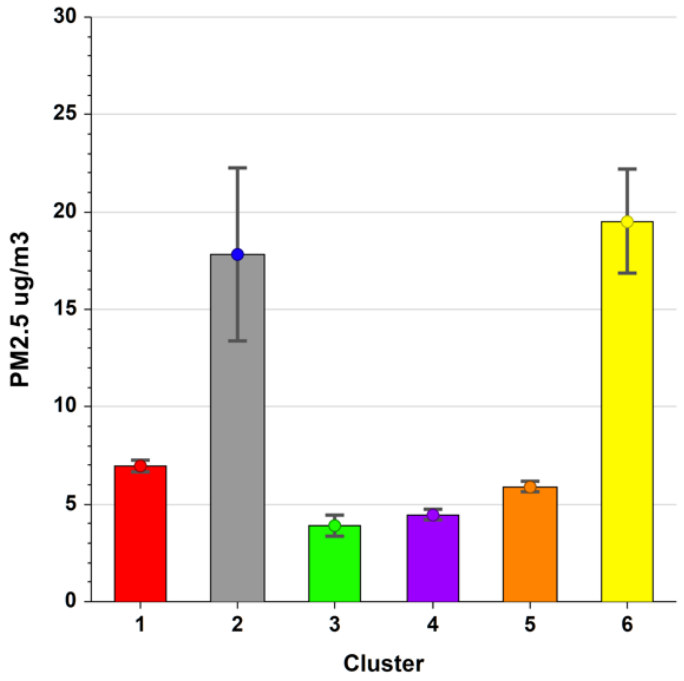




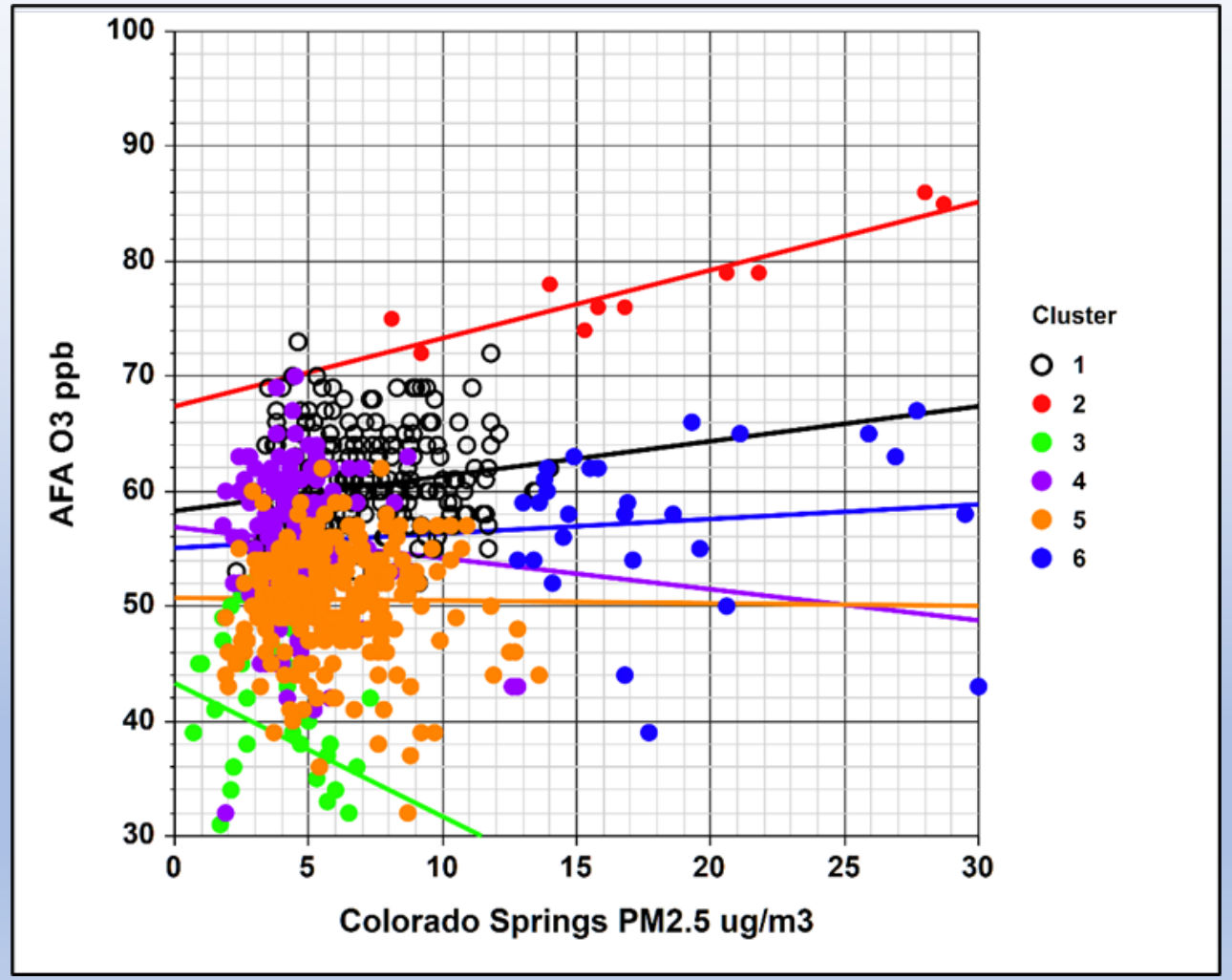








AFA daily max 8-hour ozone versus PM2.5 by cluster with linear regressions - for May-September 2016-2020. Smoke events are categorized as cluster 2. R-squared for regression between O_3 and PM2.5 for cluster 2 days = 0.84.



EPA Method: Estimates of Smoke Contributions Based on Subtraction of the 95th Percentiles for Positive Model Residuals from the Day-Specific Predictive Model Residuals.

Date	AFA Ozone ppb	Manitou Springs Ozone ppb
6/17/2020	9.1	3.7
8/21/2020	3.6	3.6
8/22/2020	7.2	5.1
8/23/2020	10.4	11.7
8/24/2020	3.4	4.8
8/25/2020	1.7	3.3

Estimated Smoke Contributions Derived by Subtraction of 95th Percentiles for High-O_3 Cluster 1 Days from Concentrations on Exceptional Event Days

Date	AFA O_3 ppb	Manitou Springs O_3 ppb	95 Percentile O_3 in ppb for *Cluster 1* at Both Sites	AFA Estimated Smoke Impact on O_3	Manitou Springs Estimated Smoke Impact on O_3
17-Jun-20	78	73	69	9	4
21-Aug-20	79	77	69	10	8
22-Aug-20	85	81	69	16	12
23-Aug-20	86	85	69	17	16
24-Aug-20	79	79	69	10	10
25-Aug-20	76	76	69	7	7

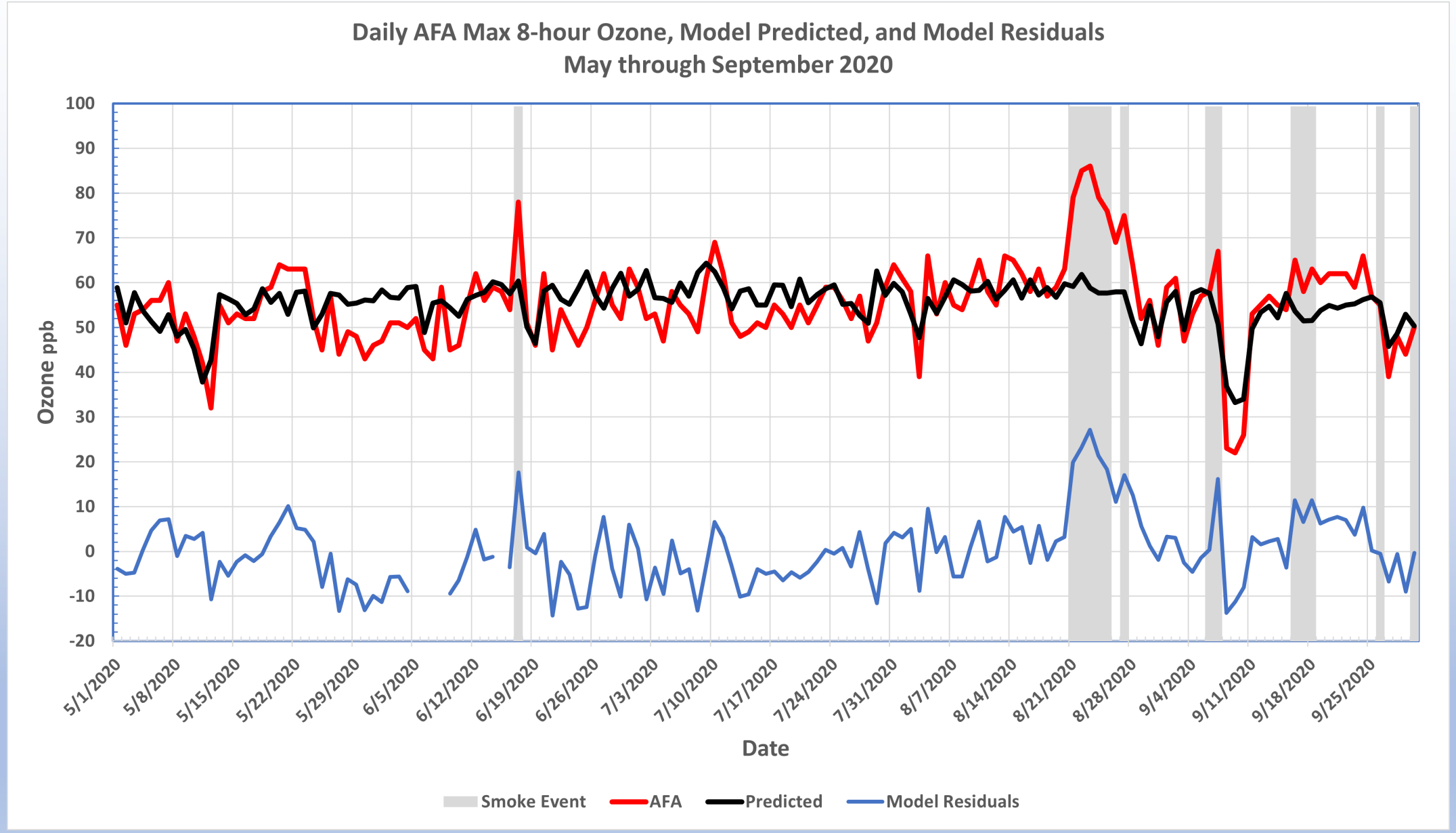
(95th percentiles for Cluster 1 were identical at both sites.)

A Simple K-Means Cluster Analysis That Uses Satellite Data and Does not Use Predetermined Informational Flags. This Analysis Could be Used to Determine Which Days to Flag.

A Cluster Analysis Without Predetermined Flagged Days for May-Sep 2020 That Includes Three Satellite Variables: MODIS Tera AOD, MODIS Tera Cloud Fraction, and TROPOMI CO

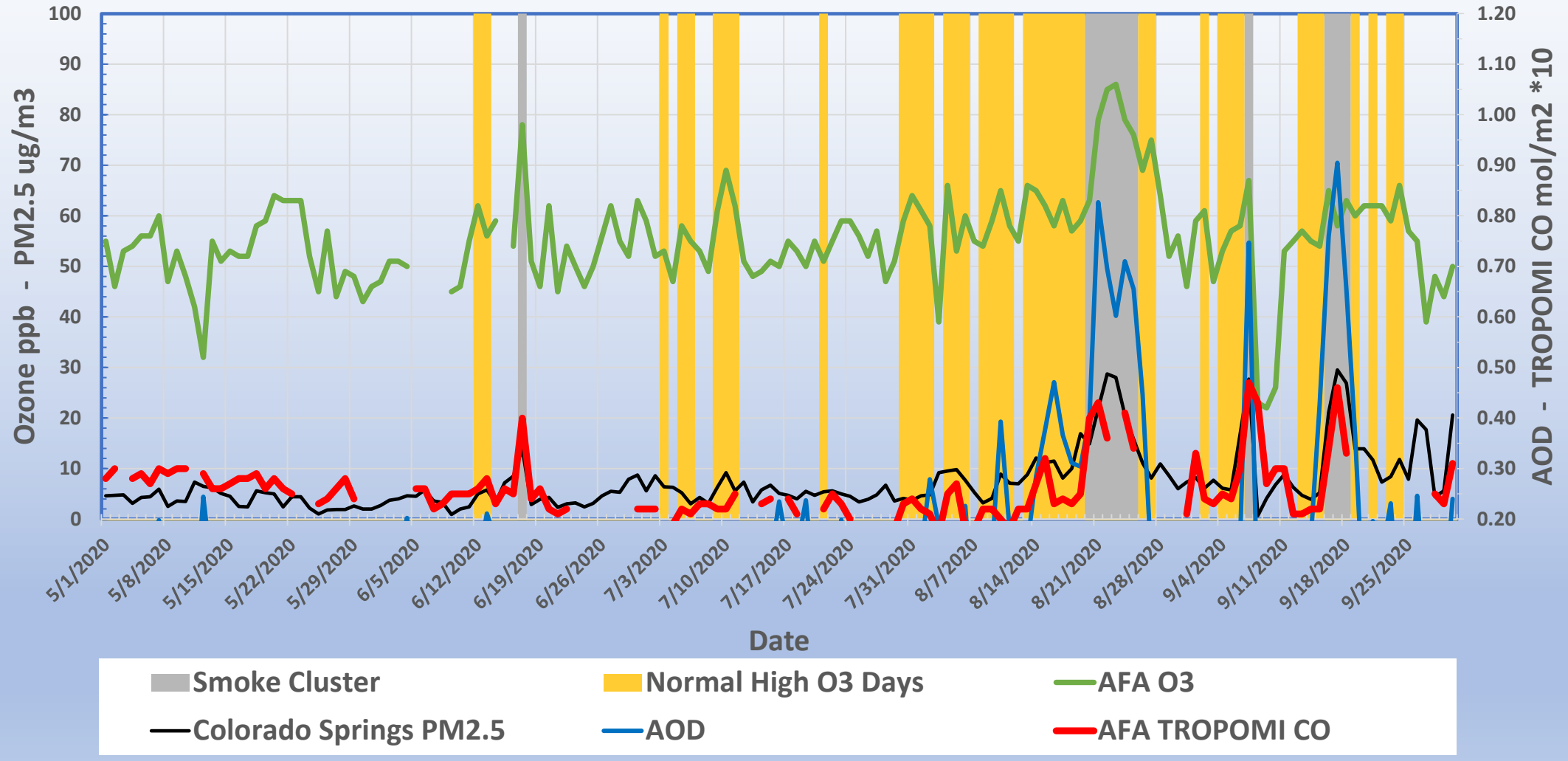
Variables	Cluster 1: Windy Aloft	Cluster 2: Cool & Cloudy	Cluster 3: Cloudy & Lower Heights	Cluster 4: Normal High O_3 Days	Cluster 5: Smoke Days	Cluster 6: Cool, Windy, & Lower Heights
Colorado Springs 24-hr PM2.5 ug/m3	6.6	3.3	4.8	7.8	22.6	4.9
TROPOMI CO mol/m2	0.024	0.027	0.022	0.024	0.039	0.027
MODIS Tera AOD	0.09	0.10	0.15	0.19	0.64	0.10
AFA Daily Max O_3 ppb	54	39	52	60	73	53
Denver 500 mb heights m	5849	5733	5891	5916	5908	5747
Denver 700 mb zonal winds m/s	5.6	-0.3	2.6	2.3	1.9	4.9
Colorado Springs Max Temperature Deg K	304	287	303	305	305	296
MODIS Tera Cloud Fraction	0.20	0.81	0.70	0.13	0.08	0.17
Count	33	9	35	40	11	25

Daily max temperature is a key variable universally. Other key meteorological variables will need to be determined on a subregion-specific basis. TROPOMI Satellite Column CO Acquired from Google Earth Engine; see sample code on last slide.



Only smoke days have AFA residuals greater than 10 ppb.

Daily Max 8-hour Ozone, 24-hour PM2.5, and TROPOMI CO for Colorado Springs with Denver Grid MODIS Satellite AOD - May - September 2020



TROPOMI Satellite Column CO Acquired from Google Earth Engine

Data and Websites Useful for Screening Days for Smoke-Ozone Event Flags (All sites allow access to archived data.)

- MODIS Combined Dark Target and Deep Blue AOD at 0.55 micron for land and ocean and Tera cloud fraction are available at: https://giovanni.gsfc.nasa.gov/giovanni/.
- Daily NCEP Reanalysis meteorological variables are available at: https://psl.noaa.gov/data/timeseries/daily/.
- Sentinel 5P satellite TROPOMI column carbon monoxide (CO) is available at: Sentinel-5P OFFL CO: Offline Carbon Monoxide | Earth Engine Data Catalog | Google for Developers. See sample javascript code on last slide.
- AerosolWatch site: GOES and VIIRS visible imagery, satellite aerosol optical depth (AOD), fire detections, smoke masks, and TROPOMI satellite CO beginning in 2022: https://www.star.nesdis.noaa.gov/smcd/spb/aq/AerosolWatch/
- NCAR ACOM Worldview: Visible satellite imagery; fires; a wealth of satellite measurements; TROPOMI satellite CO; AOD; modeled O_3, PM2.5, and CO; and more: https://worldview.acom.ucar.edu/
- AirNow Tech Navigator: Air quality data, HMS smoke coverage, trajectories, O_3 wind roses, and meteorology: https://www.airnowtech.org/navigator/index.cfm#
- InciWeb fire information: https://inciweb.nwcg.gov/
- HYSPLIT forward and back trajectories: Air Resources Laboratory - HYSPLIT - Hybrid Single Particle Lagrangian Integrated Trajectory model (noaa.gov)
- Weather maps: https://psl.noaa.gov/cgi-bin/data/narr/plotday.pl

```
// Script to plot and extract TROPOMI total column carbon monoxide (CO).
// Lat Long coordinates for monitoring site
var geometry = ee.Geometry.Point([-104.81638372975587, 38.96096752365792]);

// Select date range. The Google Earth Engine image limit is 5000 or ~ 5 months.

var startdate = "2020-05-01"
var enddate = "2020-09-30"

// Import sentinel-5P OFFL CO

var CO_collection = ee.ImageCollection('COPERNICUS/S5P/OFFL/L3_CO')
.filterBounds(geometry)
.filterDate(startdate,enddate)
.select('CO_column_number_density');

print(CO_collection)

// Make Time Series Chart
// Substitute a project specific title, e.g., 'Air Force Academy CO Column Number Density'
var chart = ui.Chart.image.series(CO_collection, geometry)
.setOptions({
title: 'CO Column Number Density',
vAxis: {title: 'Concentration (mol/m²)'},
hAxis: {title: 'Date'}
})
print(chart)

// Click on symbol in upper right of the chart to expand it.
// In the upper right of the expanded chart there are options
// for saving the chart image or downloading a csv file.
```

Google Earth Engine Javascript Code to Extract TROPOMI Column Carbon Monoxide for a Point.

Sign up for a Google Earth Engine account and read Google's policies for a noncommercial account.

Go to: *Sentinel-5P OFFL CO: Offline Carbon Monoxide | Earth Engine Data Catalog | Google for Developers*

At the bottom of the page open the Code Editor and replace the code with the text (left).

Run the code. Results will show up in the console. Follow additional directions in code comments.