

A workshop on atmospheric boundary Layer processes

Howard University Beltsville Campus

Analyses of legacy air quality data sets

Instructors: Jose D. Fuentes (jdfuentes@psu.edu), Ariel Stein (ariel.stein@noaa.gov)

Goals

- A. Learn the most common methods of data visualization and presentations using Python.
- B. Apply techniques to detect trends and patterns in air quality data sets and develop Python code to ascertain whether air pollutant concentrations are changing with time.
- C. Perform analyzes of historical air quality data sets and develop tools to relate airborne pollutant levels with source regions using air parcel trajectories.

Expected outcomes

1. Participants will learn how to read data sets saved in Excel (csv, xls, xlsx), txt, or NetCDF format using Python.
2. Participants will gain skills on how to create publication-quality figures in Python.
3. Participants will learn the most common data visualization methods to report air quality information.
4. Participants will develop skills to perform temporal trend analyses of air quality data.
5. Participants will learn how to obtain air parcel trajectories using the HYSPLIT model.
6. Participants will devise methods to integrate air quality data and air parcel trajectories.
7. Participants will gain skills to interpret changes in air quality data trends.

Readings

Fuentes, J.D. and Dann, T.F., 1994. Ground-level ozone in eastern Canada: seasonal variations, trends, and occurrences of high concentrations. *Air & waste*, 44(8), pp.1019-1026.

Stein, A.F., Draxler, R.R., Rolph, G.D., Stunder, B.J., Cohen, M.D. and Ngan, F., 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society*, 96(12), pp.2059-2077.

Fujita, E.M., Stockwell, W.R., Campbell, D.E., Keislar, R.E. and Lawson, D.R., 2003. Evolution of the magnitude and spatial extent of the weekend ozone effect in California's South Coast Air Basin, 1981–2000. *Journal of the Air & Waste Management Association*, 53(7), pp.802-815.

Sources of air quality data

http://www.epa.gov/airdata/ad_data.html

https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw

Available data sets

Criteria gases and particulate matter

Pollutant	Chemical formula	Code
Ozone	O ₃	44201
Sulfur dioxide	SO ₂	42401
Carbon monoxide	CO	42101
Nitrogen dioxide	NO ₂	42602
Particle matter 2.5 mass [§]	PM _{2.5}	88101
Particulate matter 10 mass	PM ₁₀	81102

§: Filter-based Federal Reference Method (FRM)/ Filter-based Federal Equivalent Method (FRM)

Meteorological variables

Variable	Symbol	Code
Resultant wind speed	U	
Air temperature	T	62101
Barometric pressure	P	64101
Relative humidity	RH	42602
Dewpoint temperature	T _d	

The Beltsville, MD site

Latitude: 39.055277 °N, longitude: 76.878333 °W

Elevation: 49 m above mean sea level

Location setting: Suburban

State code: 24

County code: 33

Site number: 30

EPA site code: 024-033-0030

Air quality data analyses

Ozone data sets for Beltsville, MD during 2005 to 2019 are available to be analyzed as indicated below. There is a file for each year of ozone data. Hourly ozone values (in the mixing ratio units of parts per billion on a per volume basis, ppbv) are provided. The data analyses listed below are modeled after the study by Fuentes and Dann (1994), FD94 (Journal of the Air & Waste Management Association 44: 1019 - 1026). Write Python code to read the ozone data in to computer memory for all the years (need to combine the data), make the outlined calculations, and create figures. All figures need to have consistent line thickness, and font type and size for the labels.

1. Compute time series of daily (a) mean and (b) maximum ozone levels and plot the results (see FD94 Figure 2a).
2. For the months of June, July, and August (JJA), calculate the (a) average and (b) standard deviation of hourly ozone levels and create a plot, with average ozone \pm standard deviation on the y-axis and time of day on the x-axis. Figure axes need titles and appropriate units.
3. For JJA, estimate the (a) average and (b) standard deviation of hourly ozone levels for weekdays (Monday to Friday) and weekends (Saturday and Sunday). Create a one-page figure with two plots. The top plot is for weekdays and the bottom plot is for weekends. The y-axis is for average ozone \pm standard deviation and the x-axis is for time of day.

4. For all months of the year, compute the monthly (a) average maximum and standard deviation and (b) mean and standard deviation ozone levels. Create a one-page figure with two plots (similar to FD94 Figure 3). The top plot is for average maximum \pm standard deviation and the bottom plot is for mean \pm standard deviation. The y-axis is for average ozone \pm standard deviation and the x-axis is for time of day.
5. For JJA, compute the composite yearly average 95, 90, 75, 50, 25, 10, and 5 percentiles. Also, for each year, compute the mean ozone values. Plot the results (as in FD94 Figure 4a) using the Python **boxplot** command (https://matplotlib.org/3.1.0/gallery/statistics/boxplot_color.html#sphx-glr-gallery-statistics-boxplot-color-py).
6. For the results obtained in 5, estimate temporal trends in ozone levels (all percentiles) and determine whether temporal changes in ozone levels are statistically significant (use your own choice of probability (p) values).

Air parcel trajectories

Air parcel trajectories for the Beltsville, MD site will be calculated from 2005 to 2019 and for JJA months. Once air parcel trajectories are executed, they need to be integrated with ozone data to create the results outlined below. Groups of two students need to be established. Each group will do 1 or 2 years of air parcel trajectories. A general description of the methodology used by the HYSPLIT model to calculate trajectories can be found in the following link: https://www.ready.noaa.gov/documents/Tutorial/html/traj_eqns.html

1. Download the NARR (North American Regional Reanalysis) or WRF (Weather Research Forest) 27-km meteorological datasets (<ftp://arlftp.arlhq.noaa.gov/archives/>) for 1 or 2 years within the period of 2005 to 2019. Calculate backward air parcel trajectories starting from Beltsville, MD, starting between 7 UTC (Universal Coordinated Time) and 10 UTC for a duration of 48 hours. Save all the trajectories in files stored on your personal computer. To establish the influence of source regions on ozone levels, select the period during 12:00 to 15:00 hours (Eastern Time) to produce averages for ozone levels. A way to automate the calculation of the trajectory can be found in the following link: https://www.ready.noaa.gov/documents/Tutorial/html/traj_freq.html
2. Group the ozone data based on the four quadrants representing 0-90 degrees (North East = NE), 90-180 degrees (South East = SE), 180-270 degrees (South West = SW) and 270-360 degrees (North West = NW). Compute the ensemble ozone averages and associated standard deviation corresponding the four quadrants. Plot the average and standard deviation values for ozone on a histogram-type plot, with the ozone levels on the y-axis and source regions (i.e., NE,) on the x-axis. See sample plot below.
3. Group the air parcel trajectories based on the four quadrants representing 0-90 degrees (North East = NE), 90-180 degrees (South East = SE), 180-270 degrees (South West = SW) and 270-360 degrees (North West = NW). For each day, compute and plot the total trajectory point count. Plot the total trajectory point count on the y-axis and day of year on the axis. Below is a sample figure.
4. For the time period of 12:00 to 15:00 hours, plot the individual air parcel trajectories on a map to identify the likely regions that provide the highest ozone levels. Short line segments will appear the map, away from the Beltsville site (see sample figure below).

5. Perform a cluster analysis following the guidance provided in this link: https://www.ready.noaa.gov/documents/Tutorial/html/traj_clus.html. Compute and plot the average backward trajectories. The sample figure below provides the averaged trajectories for three clusters. A description of the theoretical basis for the cluster analysis can be found in the following link: https://www.ready.noaa.gov/documents/Tutorial/html/traj_cluseqn.html
6. Plot all the 48-hour air parcel trajectories on a map. This type of plot is called the 'spaghetti' air parcel trajectories. The thick lines represent averaged air parcel trajectories.
7. For your chosen time period, compute and plot on a map the probability density distribution of air parcel trajectories as shown in the figure below.