Removal of Smoke Events from Particulate Background Concentration Data

Smoke events cause an increase in background concentrations used in dispersion modeling analyses that are not representative of normal conditions. In their April 2019 memo "Additional Methods, Determinations, and Analyses to Modify Air Quality Data Beyond Exceptional Events" EPA suggests that it may be appropriate to modify monitoring data for use in modeling analyses where the data are not representative to characterize background concentrations[1]. This document describes the method used by the Iowa DNR to identify these extreme smoke events and remove their influence from the data used to calculate background concentrations.

Note: This method does not apply to the calculation of AQI or monitor design values

Identification of Smoke Events

The DNR uses two criteria to identify days that will be excluded from background data used in dispersion modeling.

1. Outliers in the ambient monitoring data
2. Smoke observations

Identifying Outliers

Traditionally, outliers in a dataset are determined using a certain number of standard deviations from the mean. However, both the mean and the standard deviation are heavily influenced by extreme outliers, like those from smoke events. An alternate approach is to use the Median Absolute Deviation (MAD). This technique is less influenced by outliers, making it a more ideal robust metric for identifying smoke events in the ambient monitoring data.

Ambient pollutant concentrations naturally follow a lognormal distribution[2,3]. To account for right-skewness without inflating baseline variance during extreme smoke events, daily concentration data (x_i) can be log-transformed ($y_i = \ln(x_i)$). After which, the data approximate a normal distribution. An upper-MAD approach is applied to strictly focus on variations above the median, ensuring that low concentration noise does not influence the outlier threshold[4]. On the log-transformed scale the median is determined, and the upper-MAD is calculated using only those values greater than or equal to the median.

Once the upper-MAD has been determined for a data set it is multiplied by a consistency constant. Most commonly, 1.4826 is used, which makes MAD a consistent estimator of the standard deviation in normal distributions[5]. Because the log-transformed data approximate a normal distribution we can use 1.4826 directly, eliminating the need to derive distribution-specific consistency constants for each monitor/year.

The log-transformed upper outlier threshold is defined as: $\text{median}_y + k\ (1.4826 \times \text{upper MAD}_y)$, where k is the acceptable number of deviations from the median. The selection of k is somewhat subjective and commonly ranges between 2-3. The DNR selected k=2, which corresponds to an approximate 95.4% upper confidence bound on the log scale (compared to 99.7% for k=3). A threshold of k=2 was selected because it allows for the identification of less extreme smoke events, and because the threshold itself is not the only criterion that will be used to determine if each data point will actually be removed.

---

[1] US EPA (2019). *Additional Methods, Determinations, and Analyses to Modify Air Quality Data Beyond Exceptional Events*

[2] Ott, W. R. (1990). *A Physical Explanation of the Lognormality of Pollutant Concentrations*. Journal of the Air & Waste Management Association, 40(10), 1378–1383

[3] US EPA (2002). *Calculators and Software for Environmental Applications: The Lognormal Distribution in Environmental Applications*. EPA/600/R-97/006

[4] https://aakinshin.net/posts/harrell-davis-double-mad-outlier-detector/

[5] Leys, C., et al. (2013). *Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median*, Journal of Experimental Social Psychology

Finally, the log-transformed outlier threshold is converted back into the original concentration scale via exponentiation: Threshold = exp ($Threshold_{log}$). Any day exceeding the resulting threshold is flagged as an outlier.

Outlier thresholds are computed independently for each monitor using a trailing 3-year period, which prevents a single severe smoke year from artificially inflating the threshold. Once a threshold is calculated for a given year, it is frozen and retained as the window rolls forward to subsequent years. While this approach provides consistency, evaluating each year with its own trailing 3-year window requires that a monitor have five years of data before a background concentration can be calculated.

Smoke Observations

Smoke data is retrieved from NOAA's Hazard Mapping System (HMS)[6] for every day of the period being evaluated. An initial cursory review of each day is conducted to determine if a smoke plume was present over any part of the state on each day. For days where smoke was present somewhere in the state a more detailed review is conducted for each ambient monitor location. Each day with a smoke plume present over a specific monitor is flagged. In addition, the 4th of July is flagged to account for the possible presence of firework smoke that would not be depicted in the HMS maps.

Removal of Smoke Events

Specific daily concentrations at each monitor that are flagged as both a statistical outlier and a smoke event are removed from the background dataset. Using these criteria in tandem is important because high concentrations can occur in the absence of a smoke event, and because the smoke data only indicates that a smoke plume was observed in the column of air above the monitor, not that it was actually impacting the monitor at the surface. Requiring both criteria to be met ensures that only data points attributable to smoke events are removed.

This method appears to provide a more representative estimate of what background concentrations should be in the absence of extreme events like wildfires. The DNR will continue to track the frequency and impact of regional smoke events and will re-evaluate this background adjustment procedure as atmospheric conditions and regulatory guidance evolve.

[6] https://www.ospo.noaa.gov/Products/land/hms.html