

Near-Road Data Quality Dashboard

Contents

1	Summary	1
2	Tabs Overview	2
3	Network-wide Quality Indicator Tabs	2
3.1	Near-Road Network	2
3.2	Pollutant Summary	3
3.3	Near-Road Increment Calculator	5
3.4	Trends	8
3.5	Network Completeness	11
3.6	Completeness Calculation Methodology	12
4	Site Specific Quality Indicator Tabs	14
4.1	CBSA	14
4.2	Site and Method Summary	16
4.3	Time Series Investigation	17
4.4	Summary Statistics	19
4.5	Pollution Roses	20
4.6	Nonparametric Wind Regression	21
5	References	23

1 Summary

This application provides a Quality Assurance (QA) and data analysis dashboard for the Near-Road monitoring Network. Near-Road sites are required in all Core Based Statistical Areas (CBSAs) of 1 million or more persons. Those CBSAs with 2.5 million or more persons have two Near-Road sites. All Near-Road sites measure NO_2 , while at least one site in each CBSA measure CO and $PM_{2.5}$ in addition to NO_2 . Some sites have additional measurements that may include black carbon (BC), air toxics, and meteorology (e.g., wind speed, wind direction, ambient temperature, relative humidity). Assessments may be generated for each year when corresponding data are available from the AQS database; sample data from years 2016 through the present year are available for review in most tabs. The visual assessments are intended to assist agencies by providing quick summary access to information at the site or network level for the required pollutants measured at Near-Road stations. The assessments are expected to have three major advantages over existing assessments:

- Data are visualized to illustrate large amounts of information in a concise way.
- The site of interest is compared to all other sites, allowing agencies to understand whether their program's data quality is similar to other agencies or perhaps an outlier.
- The dashboard applications are expected to remain openly available so that users do not have to worry about user authentication IDs and passwords. In this way, we hope that users can efficiently use the data dashboard for their program's monitors.

The Near-Road Network Data Quality Dashboard provides multiple indicators and analysis tools for the key subset of parameters required to be measured at Near-Road monitoring stations.

- Site location and selection through a United States site map
- CBSA summaries that display Near-Road sites, paired ambient sites, and CBSA boundaries through a United States site map, NO_2 , CO , and $PM_{2.5}$ pollutant summaries and increments per CBSA (see Pollutant Summary and Increment Calculator tabs for details)
- Site and method summary that provides a tabular summary of key monitors, counts, and method types
- Generic time series investigation tool – combines a time series, scatter plot, and pollution rose visualization of any parameter or combination of two parameters
- Yearly summary statistics for parameters measured at the selected site
- Pollution roses for NO_2 and $PM_{2.5}$ with options for year and meteorological season subsets
- Source apportionment using nonparametric wind regression (NWR) algorithms

Additionally, the Near-Road Data Quality Dashboard provides a network wide indicators.

- Pollutant summary that provides graphical displays of annual and subannual NO_2 , CO , and $PM_{2.5}$ data
- Near-Road increment calculator for annual and subannual NO_2 , CO , and $PM_{2.5}$ data
- Trends for key required parameters at a network and site level
- Network completeness for NO_2 , CO , and $PM_{2.5}$ compared across all stations

2 Tabs Overview

A user can view two types of assessments within the data quality dashboard as shown in Figure 1. Tabs colored red indicate *network level assessments* that focus on network wide quality indicators. Tabs colored purple indicate *specific site-level assessments* focused on site-specific quality indicators and analysis tools. Red tabs will always be shown. With the exception of the CBSA and NWR tabs, purple tabs require that a site is selected on the Near-Road Network tab (see **Near-Road Network tab** section). The CBSA tab has a standalone map that does not interact with the rest of the tabs. Individual station trends in the Trends tab and the Pollution Roses tab are only displayed when a program station is (1) selected and (2) has the required data for display.



Figure 1: Dashboard tabs, shown for the Near-Road dashboard, categorized by color. Red are Network-level Assessments. Purple are Specific Site-level Assessments, and, with the exception of the CBSA tab, require that a site is selected on the Near-Road Network tab.

3 Network-wide Quality Indicator Tabs

Five network wide tabs are available: Near-Road Network, Pollutant Summary, Near-Road Increment Calculator, Trends, and Network Completeness.

3.1 Near-Road Network

The Near-Road Network tab is the default landing page for a data quality dashboard. Figure 2 shows the default map display for the Near-Road dashboard. A link to the Environmental Protection Agency’s (EPA) website NO_2 - Near-Road Monitoring webpage, where a user can find information about the Near-Road Network and download the most recent Near-Road site list, is beneath the tabs and above the map. Each circle on the map represents an individual monitoring site. By toggling between radio buttons above the map on the right, a user can update which sites are displayed. The categories of this filter may change based on monitoring program, but as seen in Figure 2, the Near-Road Dashboard allows a user to display either active (default selected) or inactive sites. Sites are color-coded to display whether a Near-Road site measures NO_2 only (i.e., does not measure CO or $PM_{2.5}$) (green) or NO_2 and CO and/or $PM_{2.5}$ (blue).

To **select a site**, a user should click a circle on the map. With the exception of the CBSA tab, a user must select a site on the map to view metrics on purple *Site Specific Quality Indicator* tabs. To **deselect a site**, a user may click the background of the map (i.e. any non-circle location on the map).

The table below the map provides site details, including location information, site code, site name, target road, distance to target road (m), distance to mainline (m), annual average daily traffic volume (2016) (AADT), fleet equivalent annual average daily traffic volume (2016) (FE-AADT), and whether the site measures multiple pollutants (i.e., NO_2 only or NO_2 and CO and/or $PM_{2.5}$). The table will update once a site is selected on the map to show information only for the selected site. Deselecting a site will re-populate the table with information for all sites.

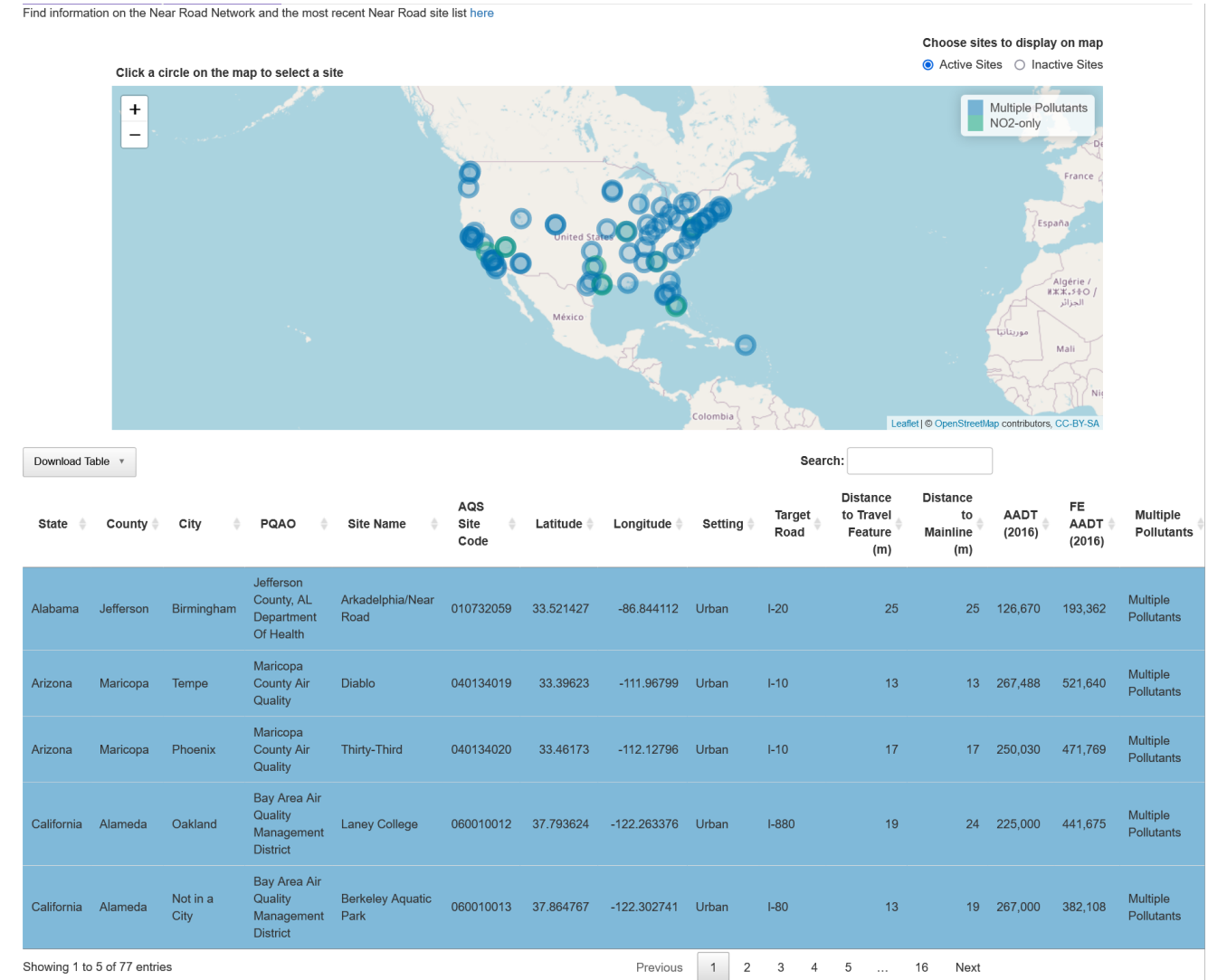


Figure 2: Near-Road Network tab (default landing page) for the Near-Road Dashboard. Near-Road sites shown in the map are also listed in the table below.

3.2 Pollutant Summary

The Pollutant Summary tab shows a customizable graphical display of annual and subannual NO_2 , CO , and $PM_{2.5}$ data recorded at sites in the Near-Road monitoring Network. The tab allows a user to select the pollutant, year, sample duration, and site(s) in drop-down menus, then render the boxplot (subannual)

or barplot (annual) by clicking the “Update Plot” button. The dropdown menus update depending on the availability of data (e.g., if a site does not measure CO , then the site will appear as a light shade of gray in the dropdown menu and will not be selectable). The site selection will only reset if a user chooses a different pollutant and/or different year. If a user toggles between different sample durations for the same pollutant and year, some sites may not be present in the plot and table upon clicking the “Update Plot” button.

The current pollutant-sample duration combinations include:

- NO_2 : hourly, 1-hour daily max (e.g., daily maximum of hourly data), and annual mean of 1-hour concentrations
- CO : hourly, 8-hour rolling average, daily max of the 8-hour average, annual second highest hourly, and annual second highest 8 hour rolling average concentrations
- $PM_{2.5}$: hourly, daily mean, and annual mean of hourly and daily mean concentrations

After the “Update Plot” button has been clicked and the desired graphical display appears below the display option panels, more options allow a user to enable or disable the appearance of outliers for the boxplot graphical display, to enable or disable the Air Quality Index (AQI) breakpoints for criteria air pollutants, and to arrange the sites by the median concentration (for boxplot display) or mean concentration (for barplot display) or by alphabetical order of the selected sites (Figure 3). Outliers are defined as 1.5 times the interquartile range above the upper quartile and below the lower quartile ($Q3 + 1.5 \times IQR$ and $Q1 - 1.5 \times IQR$). The outliers option, NAAQS threshold option, and sort option do not require a user to click the “Update Plot” button to activate.

The figure shows a user interface with three main sections:

- 1. Choose primary pollutant summary settings.** This section contains four dropdown menus:
 - a. Select a Pollutant:** Nitrogen dioxide (NO₂)
 - b. Select a year:** 2018
 - c. Select a sample duration:** Daily 1 hour max
 - d. Choose site(s) to display:** Nothing selected
- 2. Update plot based on primary settings.** This section contains a single blue button labeled "Update Plot".
- 3. Choose secondary display settings.** This section contains three groups of radio button options:
 - Display outliers (boxplot only):** "Show" (selected) and "Hide".
 - Display NAAQS threshold (criteria air pollutants only):** "Show" (selected) and "Hide".
 - Arrange graph by:** "Median (boxplot) or mean (barplot)" (selected) and "Alphabetical Order".

Figure 3: Example dropdown menu selection for 2018 NO_2 1-hour daily max.

Each monitoring site is shown on the y-axis and is labeled by the city (or county), state, and AQS site number. CO or $PM_{2.5}$ data are labeled as Federal Reference Methods (FRM), Federal Equivalent Methods (FEM), or mixed FRM/FEM methods, followed by the POC. NO_2 data are labeled as “True NO_2 (direct or photolytic)”, “Traditional (chemiluminescence)”, or mixed True NO_2 /Traditional methods, followed by the POC. Traditional (chemiluminescence) is the most common NO_2 measuring technique in the Near-Road Network. Mixed method types designate POCs where the method type was switched throughout the year. For example, an individual POC measuring $PM_{2.5}$ throughout the year may have switched from an FEM method to an FRM method, thereby designating the pollutant concentration as having been measured by “mixed FRM/FEM methods”.

If a site has multiple POCs for a selected year, the boxplots (subannual) or barplots (annual) are displayed next to each other aligned with the same location. Figure 4 shows an example output of 1-hour daily max NO_2 for 2018.

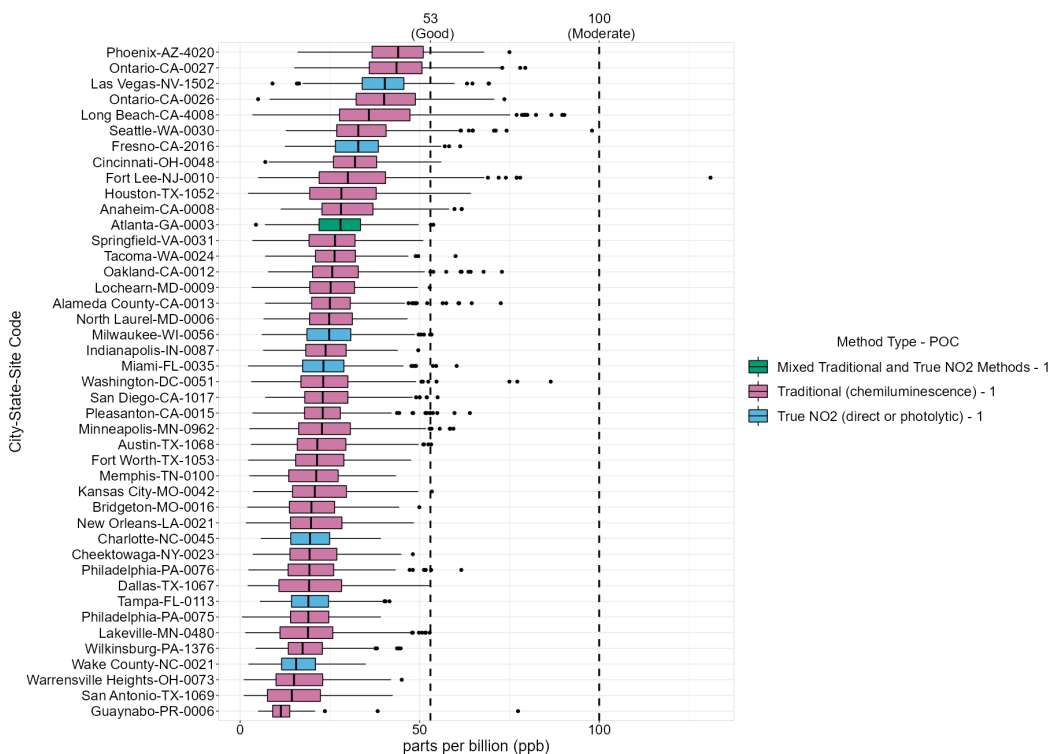


Figure 4: 1-hour daily max NO_2 for 2018 with AQI breakpoints.

A table below the plot shows metadata of a user’s selection from the settings, including the Near-Road location, AQS sitecode, POC, method, method code, and method type(s). The metadata table updates after a user click the “Update Plot” button to reflect the selected pollutant, year, sample duration, and Near-Road site(s).

3.3 Near-Road Increment Calculator

The Near-Road Increment Calculator tab shows the difference (increment) between annual and subannual NO_2 , CO , and $PM_{2.5}$ concentrations measured at Near-Road sites and paired-non-Near-Road sites (hereafter “ambient” sites). Increments quantify the contribution of the nearby roadway to the ambient pollutant concentrations. The format of the Near-Road Increment Calculator tab is similar to the Pollutant Summary tab, with an identical top panel that allows a user to select a pollutant, year, increment duration, and site(s). Further, the dropdown menus update depending on the availability of data. Near-Road monitoring sites are used to calculate increments only if they meet the following criteria:

1. Have $\geq 75\%$ annual completeness for the pollutant
2. Have at least one ambient site within 40 km that also has $\geq 75\%$ annual completeness

The completeness criteria does not differentiate between multiple POCs at a single site, so an individual POC at a site may not meet the completeness threshold of $\geq 75\%$. Two different methods are available to calculate the increment: nearest-neighbor and inverse distance weighted (IDW) average of multiple monitors. Nearest-neighbor increments are defined as the difference between the concentrations measured at the Near-Road site and concentrations measured at the closest ambient site. IDW average increments consider all ambient sites within 40 km with $\geq 75\%$ annual completeness of a Near-Road site; each ambient site is weighted based on the inverse distance squared (d^{-2}), so that ambient sites closer to the Near-Road sites more strongly influence the average, or ambient concentration. Some Near-Road sites are paired only

with a single ambient site, as no other sites in the 40 km radius meet the completeness threshold; therefore, the increment graph for Near-Road sites paired with a single ambient site will not change when a user toggles between the nearest-neighbor and IDW method.

After choosing to calculate the increment using the nearest-neighbor or IDW method, click the “Update Plot” button to render the boxplot (subannual) or barplot (annual). The user has the option of sorting the Near-Road sites from top to bottom:

- Median Near-Road pollutant concentration (subannual only)
- Median Increment (subannual only)
- Mean Near-Road pollutant concentration (annual only)
- Increment (annual only)
- Distance from target road (m)
- 2016-estimated Fleet Equivalent Annual Average Daily Traffic volume (FE-AADT)
- Alphabetical order of U.S. state Near-Road locations

The user has the option to show matching, non-matching, or both matching and non-matching methods of the Near-Road and ambient sites. The non-matching/matching methods toggle only works with subannual nearest neighbor increments; IDW averaged background concentrations are aggregates of multiple ambient monitors that do not necessarily have the same method for measuring a pollutant, thereby making the IDW averaged increments impractical to compute in real-time. The sort options and matching/non-matching method options do not require a user to click the “Update Plot” button to activate. If the increment duration is set to annual (e.g., annual mean or percentile) and sort options are set to either “Distance from target road (m)” or “FE-AADT”, the increment is aggregated by Near-Road site, and the display matching/non-matching method option becomes unavailable.

Like the Pollutant Summary tab, CO or $PM_{2.5}$ data are labeled as FRM, FEM, or mixed FRM/FEM methods, followed by the POC. NO_2 data are labeled as “True NO_2 (direct or photolytic)”, “Traditional (chemiluminescence)”, or mixed True NO_2 /Traditional methods, followed by the POC. Figure 5 shows an example output of 1-hour daily max NO_2 nearest neighbor increments for 2018.

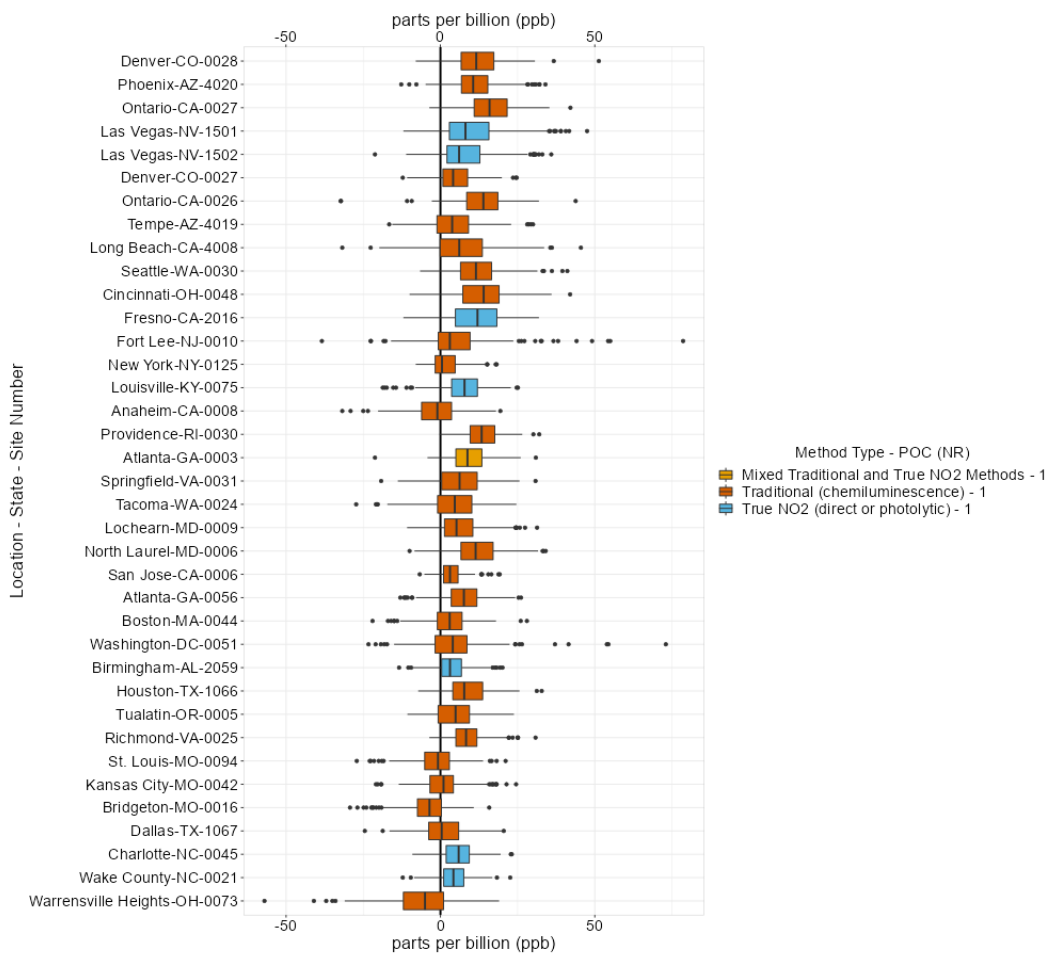


Figure 5: 1-hour daily max NO_2 nearest-neighbor increments for 2018 sorted by the median increment.

Annual increments are displayed as a bar graph. If a method at a Near-Road site is paired with multiple methods from a single ambient site, the median of the annual increments is displayed, with the minimum and maximum increment represented as left and right whiskers, respectively, extending from the median. Figure 6 shows an example output of the annual mean of 24-hour $PM_{2.5}$ nearest neighbor increments for 2020. In the example, the minimum, maximum, and median annual mean increment is approximately $0.75 \mu g/m^3$, $2 \mu g/m^3$, and $1.5 \mu g/m^3$, respectively. The minimum and maximum whiskers are not present if the IDW average increment is used because each increment is averaged based on distance.

Although the Near-Road and paired ambient site separately have $\geq 75\%$ annual completeness per pollutant, the overlap period between both sites, or number of daily increments, may not meet that threshold. Daily increments of NO_2 , CO , and $PM_{2.5}$ are calculated using daily 1-hour max, daily max of 8-hour averages, and 24-hour concentrations, respectively. A button to the left of the increment display allows a user to only display the increments that have $\geq 75\%$ annual completeness.

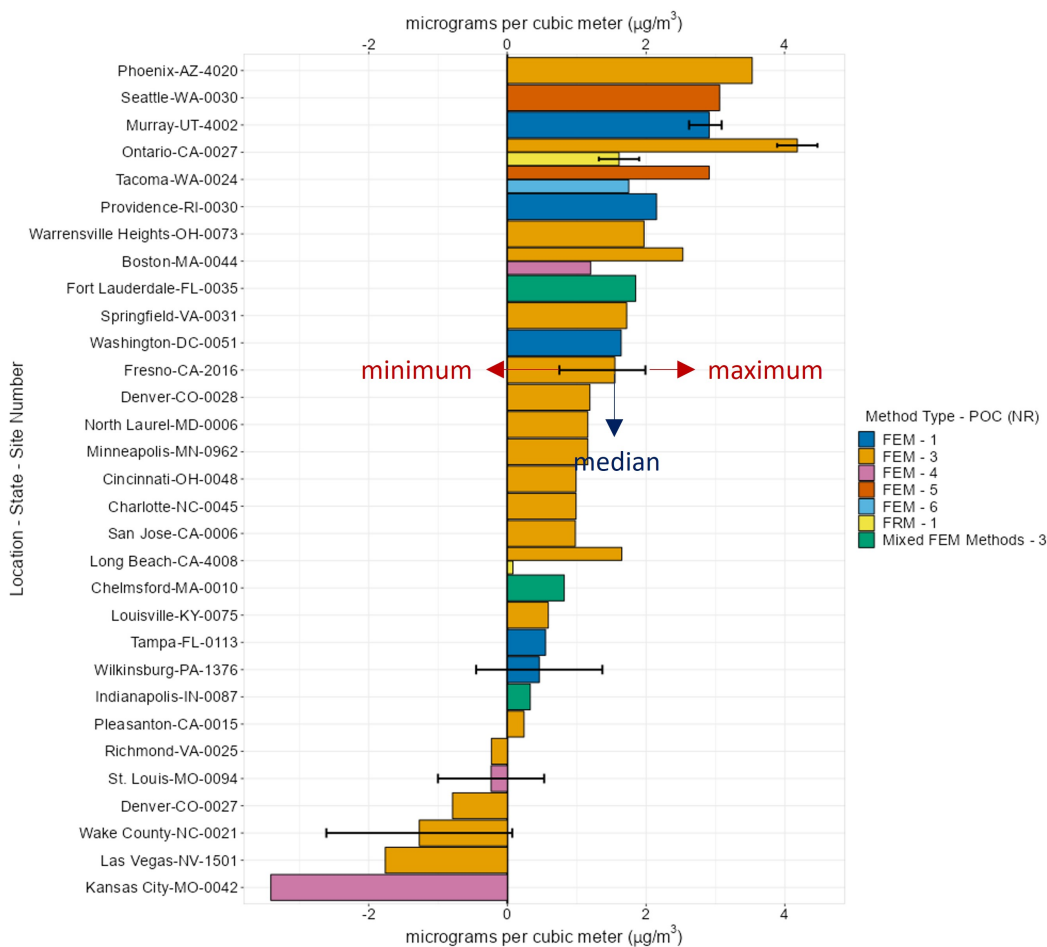


Figure 6: Annual mean of 24-hour $PM_{2.5}$ nearest-neighbor increments for 2020 sorted by the mean increment.

Near-Road and ambient site daily 1-hour max NO_2 and daily max of 8-hour average CO and annual CO (second highest 1-hour average CO and the second highest 8-hour average CO) are quantified independently before increments are calculated. Therefore, the concentrations calculated at the Near-Road and its paired ambient site(s) may not have occurred during the same hour. In other words, the daily and annual ambient concentrations are not matched based on the hour that was used to calculate the daily and annual Near-Road concentrations.

A table below the plot shows metadata of a user's selection from the settings, including the Near-Road sitecode, location, Near-Road site name, POC, paired ambient sitecode, site name of the ambient site, whether the ambient site is the nearest-neighbor, the distance from the Near-Road site to the ambient site in kilometers, annual means of the Near-Road and paired ambient site, annual increment, units, the number of POCs from the ambient site that are paired with the Near-Road site, number of daily increments for the selected year and pollutant, and the annual increment completeness (%). The metadata table updates after a user clicks the "Update Plot" button to reflect the selected pollutant, year, and Near-Road site(s).

3.4 Trends

The Trends tab provides annual trends figures for required monitoring program pollutants. Figure 7 shows the top half of the standard trends analysis tab. The boxplot shown in Figure 7 shows the distribution of slopes across the network over the past 5 years for each required pollutant. The annual percent change in

concentration is plotted on the x-axis, and each required pollutant is plotted on the y-axis. A reference line at a slope change of 0% is plotted as a dotted gray line.

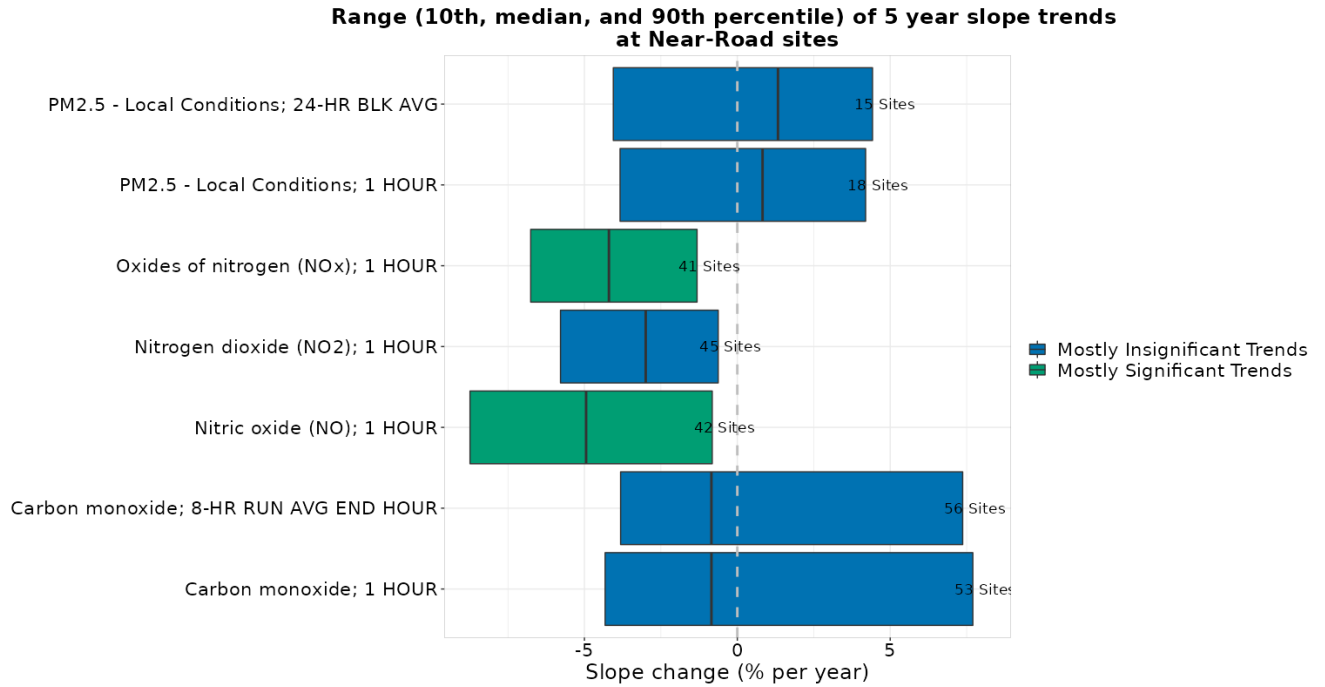


Figure 7: Trends tab with Slope Trends figure displayed.

Each box plot represents the collection of slopes across all sites that measure the pollutant. The number of sites represented by each box is noted to the right of the box. The left end of the box is drawn at the 10th percentile slope (% change per year), the right end of the box is drawn at the 90th percentile slope, and the median slope is drawn as a black line within the box. Boxes are color coded to indicate whether f-test values of individual trends are statistically significant at the 95% confidence level at more than 50% of the available trend stations; boxes that are blue indicate most site trends are statistically significant and red boxes indicate that trends are mostly insignificant (i.e., not significantly different from a flat trend). See Figure 8 for reference. Slope trends are normalized by site, so the network distributions show the three distributional points (10th, 50th, 90th percentile) percentage change per year per pollutant.

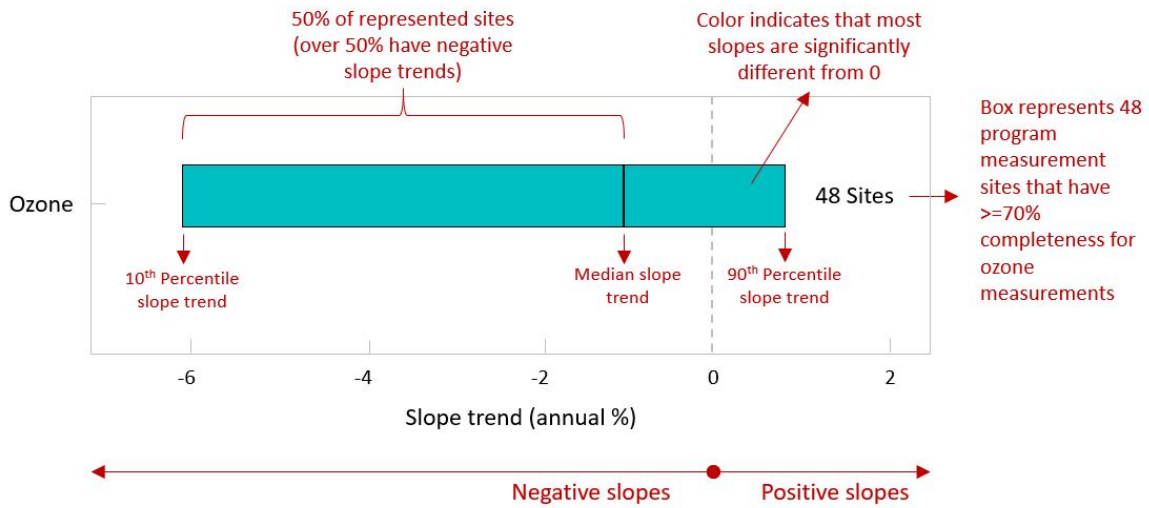


Figure 8: Reference for Slope Trends figure box.

The bottom image on the Trends tab (Figures 9 and 10) shows Individual Station Trends for a selected pollutant and sample duration. When a site is selected on the map that has a valid and complete trend, it is highlighted as a colored line on this figure. This figure allows a user to identify how annual trends at any measurement site compares to nationwide trends at other stations for the same time period. If the site selected does not have a valid or complete trend, the colored line will not appear in this figure.

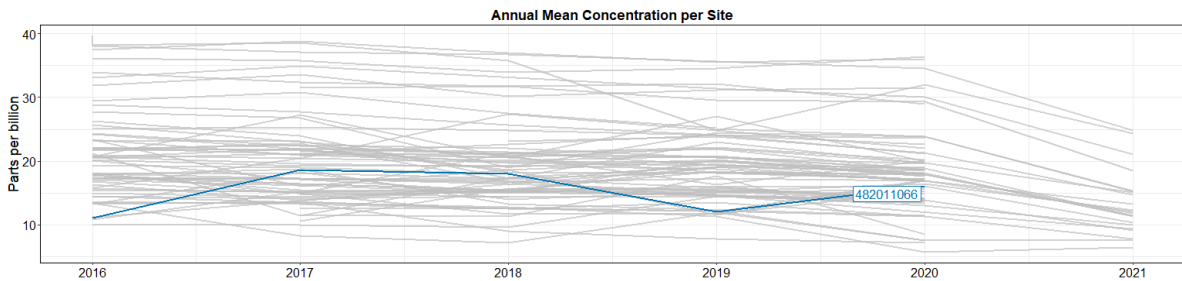


Figure 9: Individual Station Trends figure with site selected.

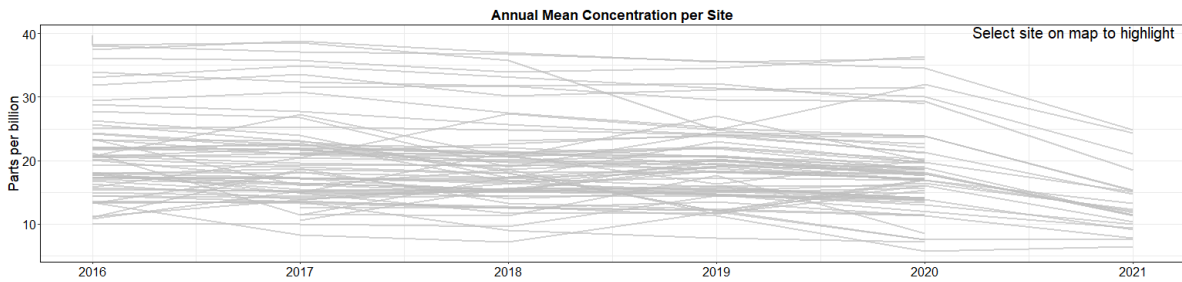


Figure 10: Individual Station Trends figure with no site selected.

3.5 Network Completeness

Figure 11 shows the Network Completeness tab display for NO_2 , CO , and $PM_{2.5}$ at Near-Road stations. The figure shows the percent completeness of pollutant parameters for the program network by site and year. The main table shows parameter and duration percent completeness for the site selected on the Near-Road Network tab (if one has been selected). If a site has multiple POCs, the lowest POC or the highest percent completeness is used in the calculation of the site completeness. The site completeness of the non-lowest POC or the POC(s) with the lower percent completeness is blank because the percent completeness did not factor into the site completeness. Further, only the highest POC or highest percent completeness POC is displayed in the figure.

Year may be changed using the drop-down menu above the plot. The menu only shows complete years of monitoring data. The dropdown menu does not allow a user to select partial years (before reporting deadlines to AQS). This means that the most recent complete year is updated in April of the following year (e.g. 2020 will become available in April, 2021). A “Select Parameter(s)” box allows a user to select or deselect NO_2 , CO , and $PM_{2.5}$, which will automatically update the completeness figure. Completeness statistics, including rank order, and the figure update when the year and selected parameters are updated. Sites appear in the figure only if they measure the selected parameter(s).

Each monitoring site is represented by a gray line segment according to rank-ordered completeness, plotted on the x-axis. Completeness (%) per parameter is plotted on the y-axis. Each parameter measured at a site is represented by a colored shape and plotted on the gray line segment representing the site.

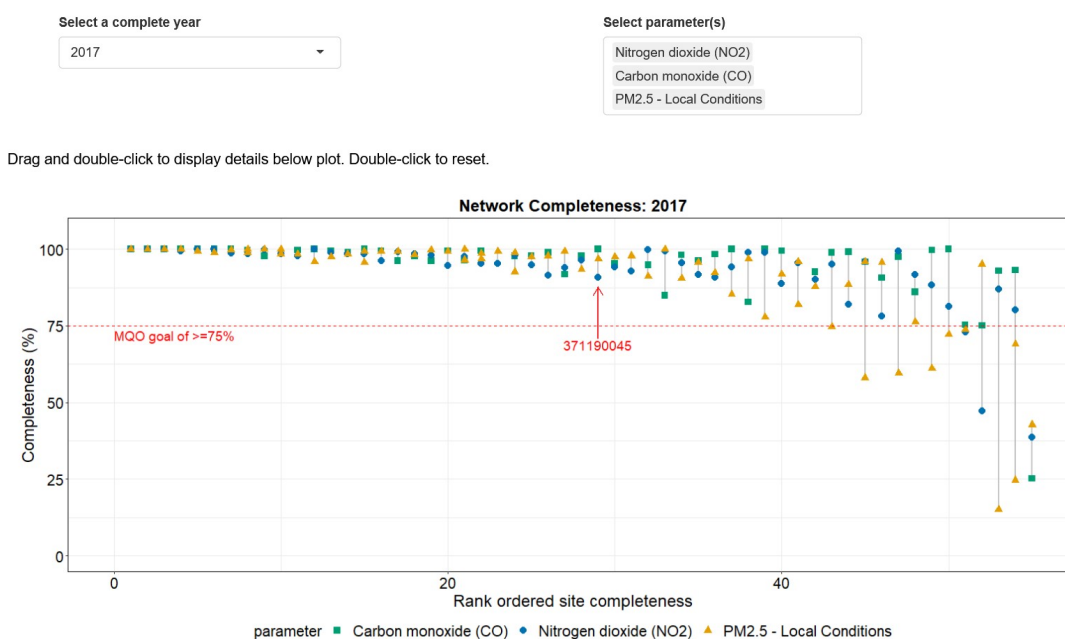
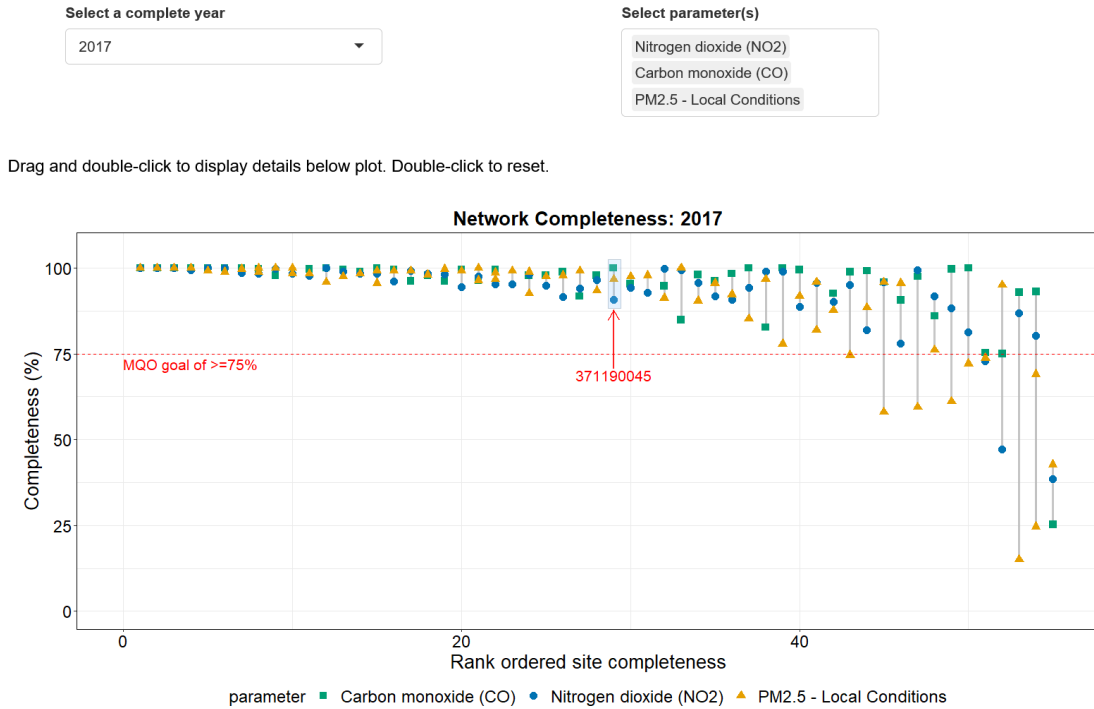


Figure 11: Network Completeness tab for 2017 with site 371190045 selected.

If a site is selected on the Near-Road Network tab, the segment representing that site is highlighted on the plot. A user can drag and double-click across a subset of the plot to view detailed information about the data point enclosed within the drag-box, which will appear in a temporary table between the plot and the main table (see Figure 12).



Site Rank	Site Code	Year	Parameter	POC	Method	Duration	% Annual Complete (Parameter)	% Annual Complete (Site)
29	371190045	2017	Carbon monoxide (CO)	1	INSTRUMENTAL - Gas Filter Correlation Thermo Electron 48i-TLE	1 HOUR	100.00	95.80
29	371190045	2017	Nitrogen dioxide (NO2)	1	Teledyne-API Model 200EUP or T200UP - Photolytic-Chemiluminescence	1 HOUR	90.68	95.80
29	371190045	2017	PM2.5 - Local Conditions	1	R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC - Gravimetric	24 HOUR	96.72	95.80

Figure 12: Network Completeness tab with dragbox over the representing site 371190045 (rank order site completeness = 29). The temporary table is shown below the plot.

3.6 Completeness Calculation Methodology

C statistics are reported in annual data to AQS. For complete years, the completeness statistics displayed are pulled directly from annual data.

Incomplete years are years for which data is not yet required to be submitted to AQS through the end of the year. Annual data from incomplete years may be unavailable or inaccurate. The completeness of a datastream for incomplete years is estimated according to the methodology below.

3.6.1 Completeness estimations for incomplete years

3.6.1.1 Step 1: Identify the “required” date The deadline for submitting data to AQS is two quarters before the most recent complete quarter.

E.g. If today is June 6, 2022, the most recent complete quarter is Q1 2022 which ended on March 31, 2022. 2 quarters before this date is the end of Q3 2021, **September 30, 2021**.

3.6.1.2 Step 2: Identify incomplete years Incomplete years include all years greater than or equal to the year of the required date.

Examples:

- On June 6, 2022, the required date is September 30, 2021. **Incomplete years include 2021 and 2022.**
- On July 1, 2022, the required date is December 31, 2021. **Incomplete years include only 2022** because data is required through the end of 2021.

3.6.1.3 Step 3: Pull datastream completeness statistics for complete years Completeness statistics for complete years are pulled directly from the **Observation Percent** column in annual data submitted to AQS. This completeness is delivered per unique datastream at a site (per site, poc, parameter, method).

3.6.1.4 Step 4: Estimate datastream completeness statistics for incomplete years Completeness per datastream (unique site, poc, parameter, method) for incomplete years is calculated against EITHER the required date, or the maximum date of available data if data extends beyond the required date.

The expected number of observations per data stream is calculated using the sample frequency recorded in sample data from AQS. If more than one sample frequency is available for a datastream across available years, the most recent recorded sample frequency is used.

A complete list of sample frequencies can be found here. The expected number of observations per year are calculated per sample frequency as shown in the list below (e.g. Every 4th day = $365 / 4 = 91.25$ expected days). Seasonal designations are ignored in this estimation (e.g. PAMS). Sample frequencies without an obvious cadence (seasonal, random, episodic, etc.) are assigned 365. **Decimal days are rounded in subsequent steps.**

- 'EVERY DAY' = 365 expected days
- 'EVERY 24TH DAY' = 15.208 expected days
- 'EVERY 30 DAYS' = 12.167 expected days
- 'EVERY OTHER DAY' = 182.5 expected days
- 'EVERY 3RD DAY' = 121.6677 expected days
- 'EVERY 4TH DAY' = 91.25 expected days
- 'EVERY 5TH DAY' = 73 expected days
- 'EVERY 6TH DAY' = 60.833 expected days
- 'EVERY 12TH DAY' = 30.417 expected days
- 'STRATIFIED RANDOM' = 365 expected days
- 'RANDOM' = 365 expected days
- 'EVERY 90 DAYS' = 4.05555556 expected days
- 'DAILY: 24 - 1 HR SAMPLES -PAMS' = 365 expected days
- 'DAILY: 8 - 3 HR SAMPLES - PAMS' = 365 expected days
- 'DAILY: 1 - 3 HR SAMPLE - PAMS' = 365 expected days
- 'DAILY: 1 - 24 HR SAMPLE - PAMS' = 365 expected days
- 'DAILY: 4 - 6 HR SAMPLES - PAMS' = 365 expected days
- 'DAILY: 4 - 3 HR SAMPLES - PAMS' = 365 expected days
- 'EVERY 3RD DAY:24-1 HR - PAMS' = 121.667 expected days
- 'EVERY 3RD DAY:8-3 HOUR - PAMS' = 121.667 expected days
- 'EVERY 3RD DAY:1-3 HR - PAMS' = 121.667 expected days
- 'EVERY 3RD DAY:1-24 HR - PAMS' = 121.667 expected days
- 'EVERY 3RD DAY:4-6 HR - PAMS' = 121.667 expected days
- 'EVERY 3RD DAY:4-3 HR - PAMS' = 121.667 expected days
- 'EVERY 6TH DAY:24-1 HR - PAMS' = 60.833 expected days
- 'EVERY 6TH DAY:8-3 HR - PAMS' = 60.833 expected days
- 'EVERY 6TH DAY:1-3 HR - PAMS' = 60.833 expected days
- 'EVERY 6TH DAY:1-24 HR - PAMS' = 60.833 expected days
- 'EVERY 6TH DAY:4-3 HR - PAMS' = 60.833 expected days
- 'EPISODIC SAMPLING' = 365 expected days
- 'SEASONAL' = 365 expected days

- ‘5 OUT OF 7 DAYS (FOR PSI)’ = 260.7142857 expected days
- ‘EVERY 6TH DAY: 3-3 HR - PAMS’ = 60.83333333 expected days

Completeness per incomplete year is estimated as follows:

$$Completeness(\%) = 100 * \frac{Obs_{recorded}}{Obs_{possible}}$$

Where:

- *IncompleteMaxDate* : The maximum of the required date, and the most recent observation available in AQS (within the year in question)
- *IncompleteMinDate* : The first day of the incomplete year (e.g. if the required date is 2021-09-30, Incomplete Min Date = 2021-01-01)
- *FracRequiredDays* : Expected yearly observations based on sample frequency (see list above) / 365
- *Obs_{recorded}* : Number of complete daily observations ($\geq 75\%$ completeness in sample data) between *IncompleteMinDate* and *IncompleteMaxDate*
- *Obs_{possible}* : (Number of days between *IncompleteMinDate* and *IncompleteMaximumDate*) * *FracRequiredDays*

Notes and Assumptions

- Completeness estimates $> 100\%$ are capped at 100% . Estimates greater than 100% could result from inaccurately recorded sample frequencies.
- Completeness for seasonal datastreams may be underestimated.
- Leap years are not accounted for in this estimation
- Decimal days are rounded to the nearest day

3.6.1.5 Step 5: Calculate site completeness statistics Average site completeness is calculated as the mean completeness per year across all datastreams at a site. Minimum and maximum site completeness is calculated as the minimum and maximum completeness values per year across all datastreams at a site.

4 Site Specific Quality Indicator Tabs

4.1 CBSA

Core-based statistical areas (CBSA) are areas delineated by a central metropolitan and socioeconomically connected, adjacent counties. CBSAs may span multiple states. The CBSA tab combines elements of the Pollutant Summary tab and Increment Calculator tab to display NO_2 , CO , and $PM_{2.5}$ data. Instead of displaying data per Near-Road site, the CBSA tab summarizes data by CBSA. The CBSA tab has a standalone map of Near-Road sites (blue) and paired ambient sites (green) and CBSA boundaries (Figure 13). Ambient sites are those sites that are paired with the Near-Road sites from the Increment Calculator tab. Some Near-Road sites do not have a paired ambient site; in this case, the Near-Road site will not have a visible ambient site near it or in the CBSA boundary. The standalone map does not interact with other tabs.

Near Road and Paired Ambient Sites

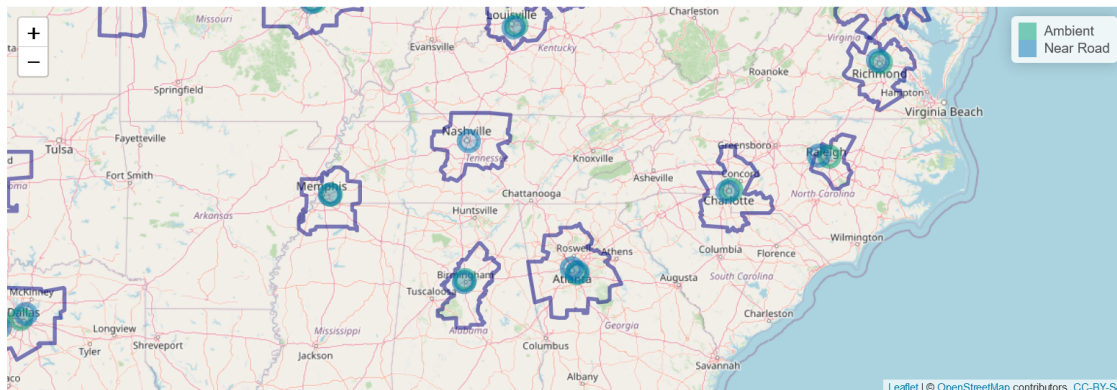


Figure 13: A zoomed-in view of Near-Road Network sites, paired ambient sites, and CBSA boundaries that contain Near-Road sites.

To **select a site**, a user should click a circle on the map. Below the map is a table with the selected Near-Road sitecode, Site Designation (i.e., Near-Road or Ambient), Site Name, City, State, and CBSA. Below the table are display and filter options for the tables and figure. A user can filter by year, pollutant, and EPA region.

Below the display and filter options is a metadata table that shows the CBSA, the number of Near-Road sites in the CBSA, the EPA region, the state(s), and the number of Near-Road sites measuring NO_2 , $PM_{2.5}$ (1-hour and 24-hour), and CO . This metadata table may change from year to year as pollutant monitors are added to or removed from Near-Road sites.

The boxplot below the metadata table displays pollutant summary data grouped by CBSA. The data displayed here is the same as the data displayed in the Pollutant Summary tab, except the boxplots are grouped by CBSA rather than individual Near-Road sites, and color-coded by EPA region. CO data are 1-hour, NO_2 data are 1-hour daily max, and $PM_{2.5}$ data are 24-hour. Figure 14 shows 1-hour daily max NO_2 for all CBSA's in EPA regions 3, 4, and 6. The CBSA selected on the map will be highlighted in the figure, assuming the EPA region that encompasses the CBSA is selected to display.

1-hour daily max NO₂ Pollutant Summaries of Near Road sites

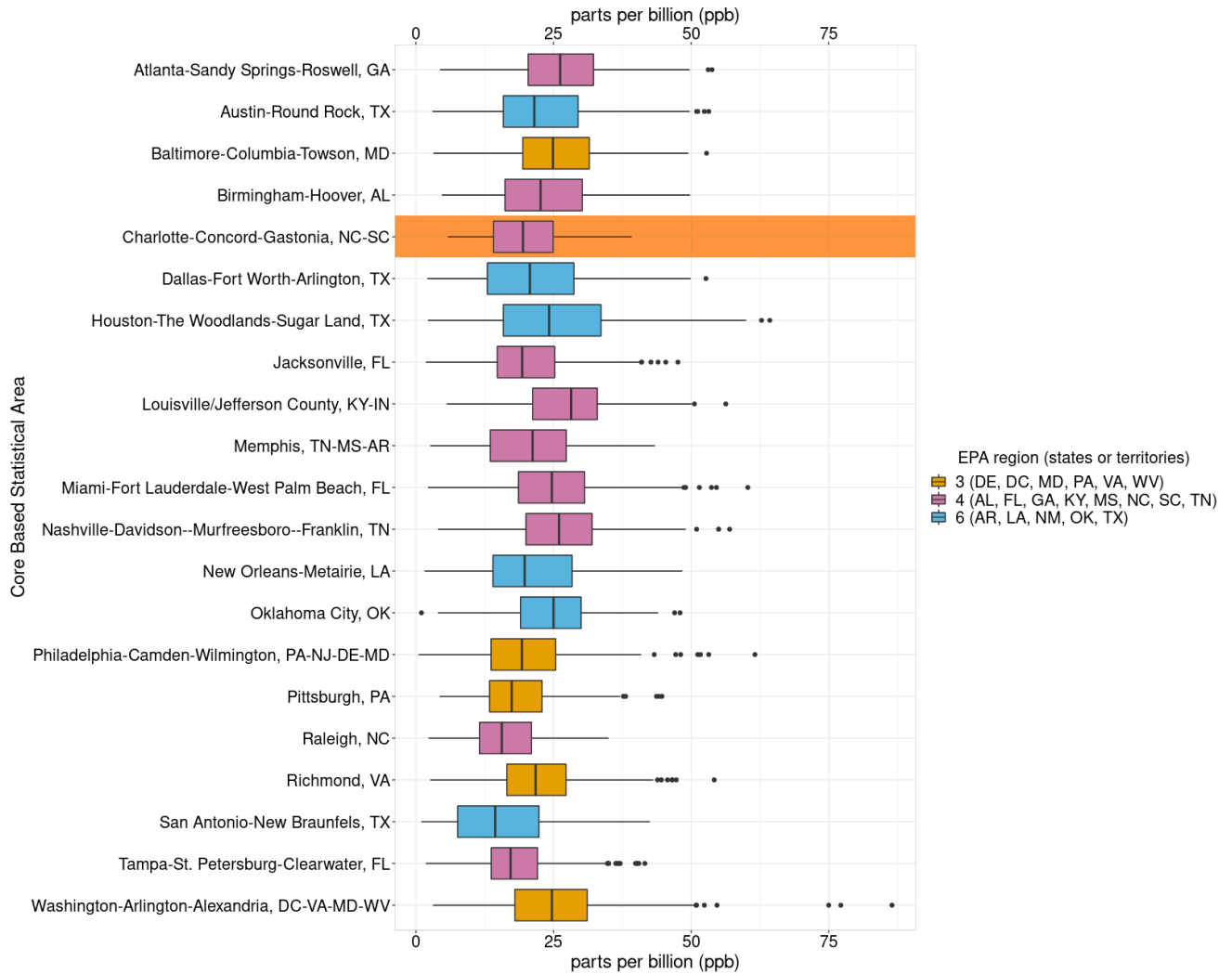


Figure 14: 2018 1-hour daily max NO₂ pollutant summary boxplots with CBSA 'Charlotte-Concord-Gastonia, NC-SC' selected on the map.

Below the pollutant summary boxplot is a table that shows Near-Road IDW increments from the Increment Calculator tab. This table displays all increments of all Near-Road sites within the user-selected CBSA. The boxplots and metadata tables will update first by clicking on a different site within a different CBSA; the boxplots and metadata tables will update further based on the display and filter options described above.

4.2 Site and Method Summary

The Site and Method Summary tab provides an annually selectable set of Quality indicators for the required parameters and a subset of optional parameters. Figure 15 shows an example for the San Antonio Interstate 35 station data in 2019. Directly below the tabs is a description of the selected site and its metadata. A selectable drop-down menu allows the user to select a year of interest and available groups – data is currently available for 2016-2021. Finally, the main table shows individual required parameters and monitor specific information including method, POC, sample duration, and units. The color-coding of the table indicates the type of parameter being measured. Types include:

- **Required Gases** : required program parameters that are gas-phase measurements
- **Required PM** : required program parameters that are particle-phase measurements
- **Meteorology** : required program parameters that are meteorology measurements
- **Additional Supporting** : optional program parameters
- **Other Meteorology** : optional meteorological parameters

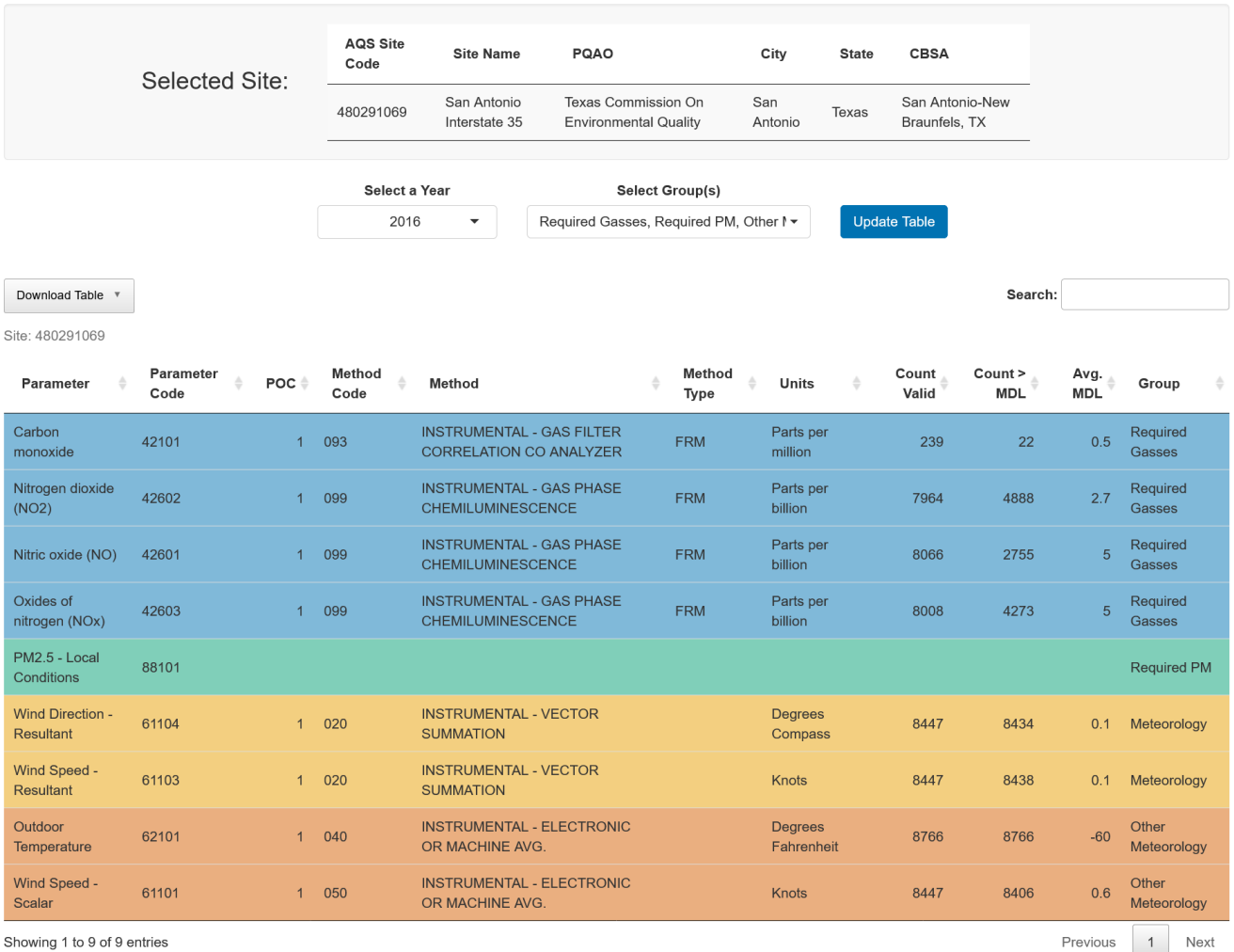


Figure 15: Example of the Site and Method Summary tab for the San Antonio Interstate 35 Near-Road station.

4.3 Time Series Investigation

The Time Series investigation tab allows a user to compare two parameters, one hourly, and the other either hourly or daily, in a time series, pollution rose, and a scatter plot. **Scatter plot comparisons are only available if both parameters are hourly.**

Options at the top of the tab allow a user to select a primary hourly parameter to display, and optionally, a secondary hourly or daily parameter to compare. Once a user selects a radio button other than “None” to display a secondary parameter, a drop-down menu will appear with the available options based on chosen duration (either 1-hour parameters or 24-hour parameters).

A map at the bottom of the page displays an urban-scale (~25 km) map showing the site location within its local area.

If one hourly parameter is chosen:

- The timeseries plot will display the primary parameter concentration as a solid line.
- A single pollution rose will display the primary parameter concentration.
- Scatter plot comparison is unavailable.

If two hourly parameters are chosen:

- The timeseries plot will display the primary parameter concentration as a solid line and the secondary parameter as a dashed line. A secondary y-axis (right-hand side) will be included to reference the secondary parameter values. (Figure 16)
- Two pollution roses will display the primary and secondary parameter concentrations, respectively. (Figure 17)
- A scatter plot comparison will display, with the option to overlay a one-to-one line and/or a linear regression line using check-box selections. (Figure 18)

If one hourly parameter and one daily parameter is chosen:

- The timeseries plot will display the primary parameter concentration as a solid line and the secondary parameter as points. A secondary y-axis (right-hand side) will be included to reference the secondary parameter values.
- A single pollution rose will display the primary parameter concentration, and a second pollution rose will display the secondary parameter concentrations (24-hour wind directions are vector averages of 1-hour resultant winds).
- Scatter plot comparison is unavailable.

4.3.1 Plot Reactivity

A user may zoom in on a particular time range in the timeseries plot by clicking-and-dragging a box across the desired subset of data on the plot, and double-clicking within the drawn box to zoom in. If displayed, pollution roses and the scatterplot will update to reflect this zoomed time period. To reset the zoom, double-click the timeseries plot.

Two reference lines are available to overlay on a scatter plot comparison: a one-to-one line, and a linear regression line. These may be drawn by selecting one or both checkboxes to the right of the scatter plot. A one-to-one line may not appear within the plot boundaries if the scales of the primary and secondary parameters do not overlap (e.g. ozone and barometric pressure).

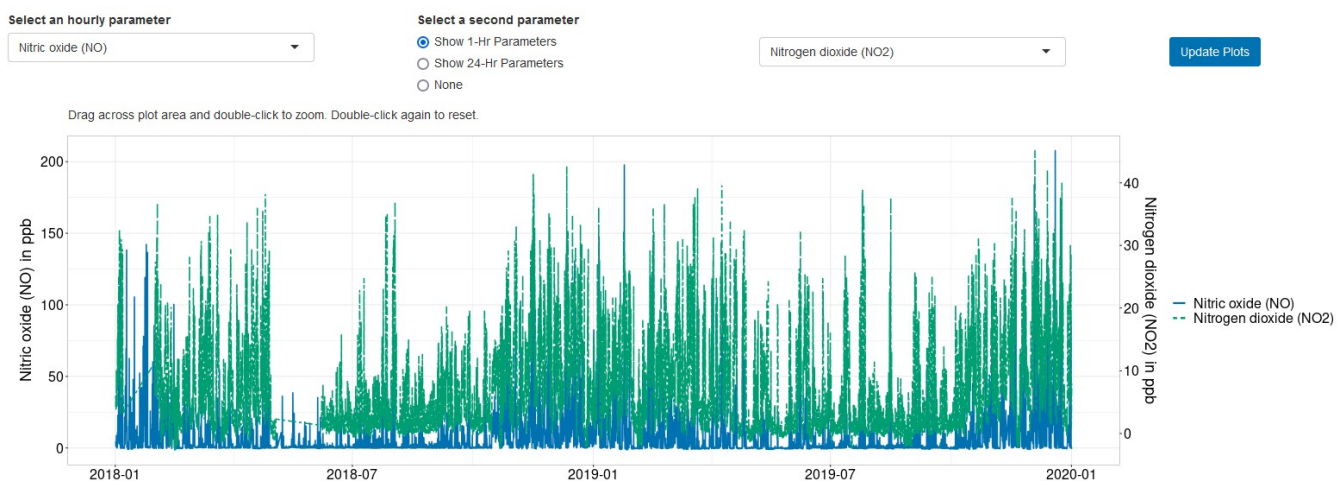
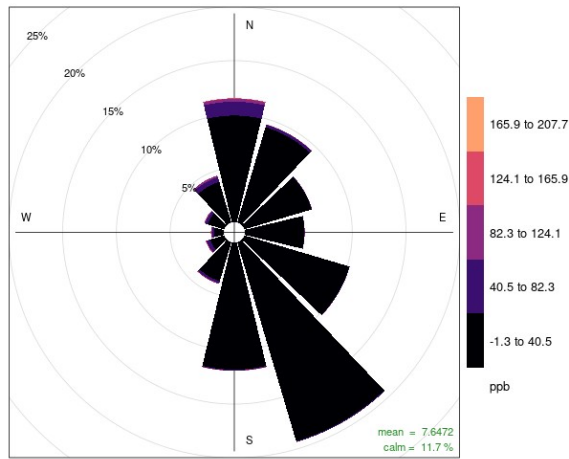


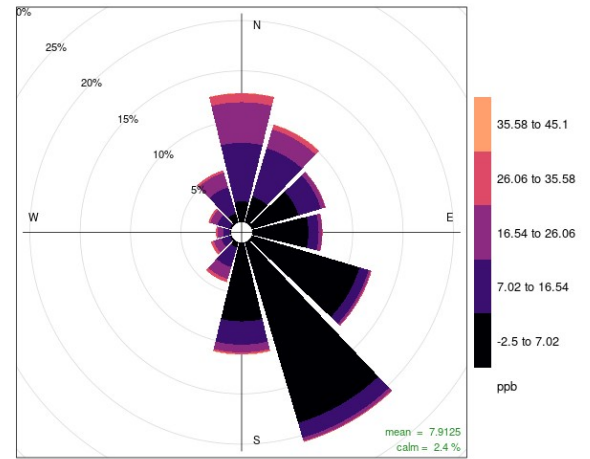
Figure 16: Time series plot a primary (hourly) and secondary parameter (hourly). The right-hand y-axis represents the secondary parameter.

Nitric oxide (NO)



Frequency of counts by wind direction (%)

Nitrogen dioxide (NO2)



Frequency of counts by wind direction (%)

Figure 17: Pollution roses for primary (hourly) and secondary parameters (hourly).

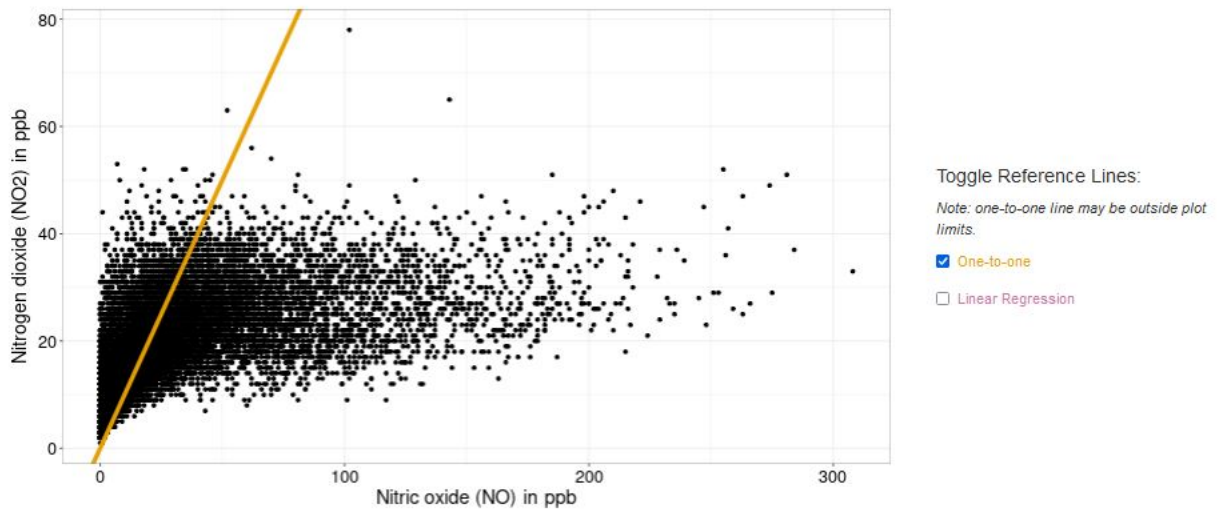


Figure 18: Scatter plot comparing primary (hourly) and secondary parameter (hourly), and one-to-one and linear regression lines shown. A scatter plot comparison is only available when two hourly parameters are chosen.

4.4 Summary Statistics

The Summary Statistics tab (Figure 19) is intended to provide a quick glance at summary statistics for each of the key program parameters. The format is very similar to the Site and Method Summary tab, with an identical site metadata table, followed by the year drop-down menu, the available group(s) drop-down menu, and then the table with summary statistics by monitor.

The color-coding of the primary data table is the same as the Site and Method Summary tab. Parameters are sorted first by required groups, and then the optional groups. The primary difference between this tab and the Site and Method Summary tab is the yearly data table provides the summary values (maximum, average, median, minimum) for each parameter. Additional quality indicators columns indicate the number of valid reported samples, the number of samples above the method detection limit (MDL), and the percentage of

samples above the MDL.

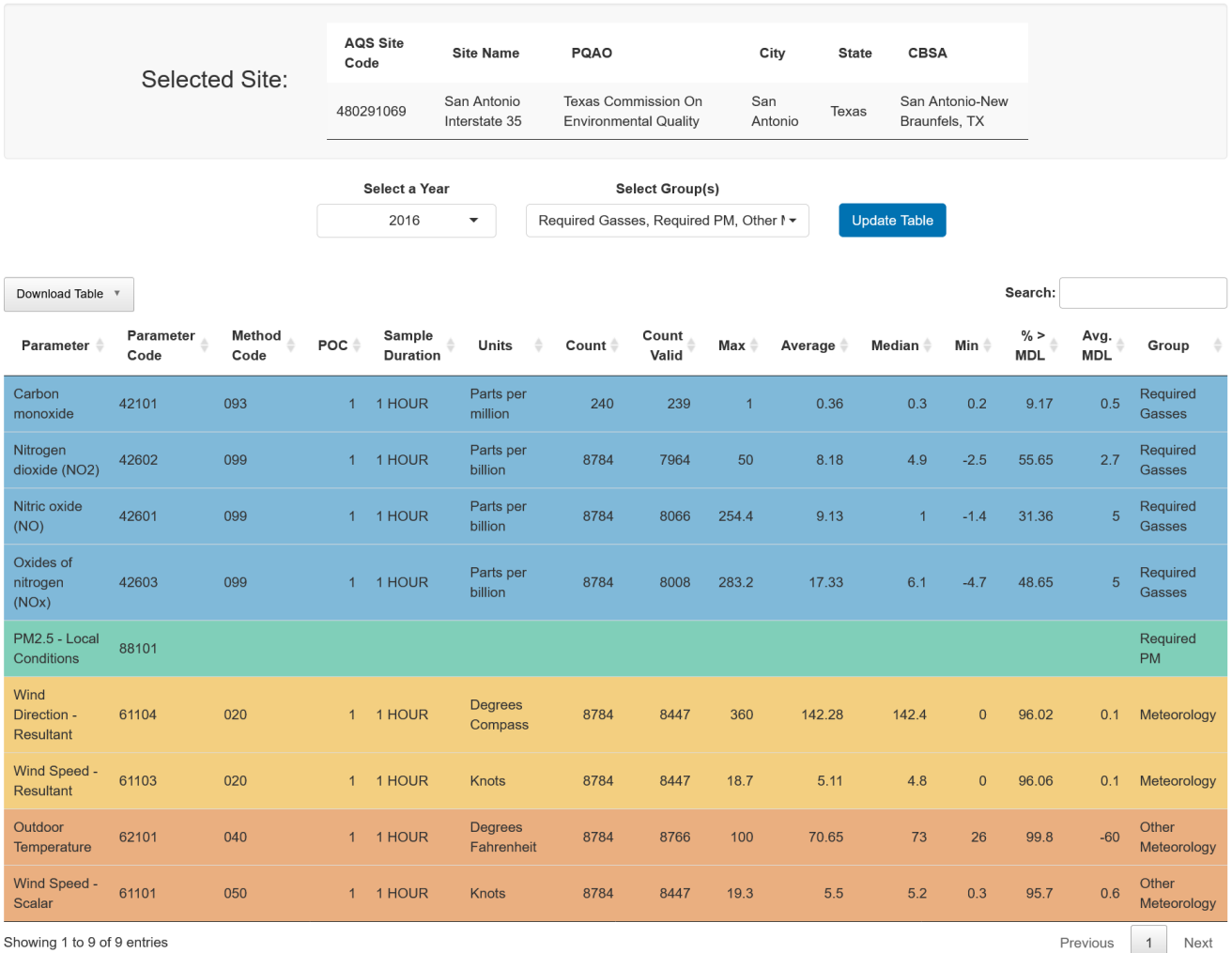


Figure 19: Example of the Summary Statistics tab, shown for the Near-Road Dashboard.

4.5 Pollution Roses

The Pollution Roses tab shows pollution roses for NO_2 and $PM_{2.5}$ at the monitoring site selected on the Near-Road Network tab. There are drop-down menus for year (or year range) and meteorological season (or seasons). The seasons are defined as:

- Spring: March, April, and May
- Summer: June, July, and August
- Fall: September, October, and November
- Winter: December, January, and February

Two pollution roses are displayed side-by-side for NO_2 and $PM_{2.5}$ (Figure 20). Below the two pollution roses is an urban-scale (~25 km) map showing the site location within its local area.

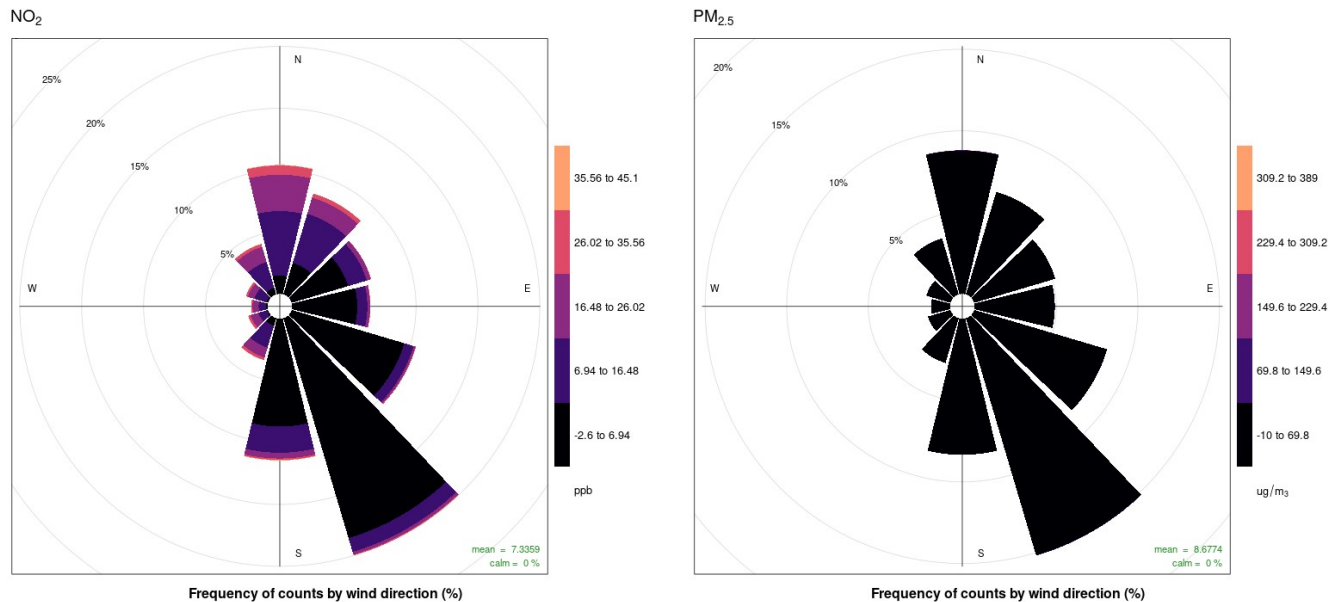


Figure 20: Pollution Roses tab showing concentrations of $PM_{2.5}$ ($\mu g/m^3$) and NO_2 (ppb) by wind arc. A table of observational completeness and a local map are also displayed.

4.6 Nonparametric Wind Regression

Nonparametric wind regression (NWR) is a pollutant source apportionment model that can be used to identify and quantify the possible source regions of pollutants as defined by wind direction sectors and wind speed bins. NWR quantifies the expected average concentration of a pollutant as a function of wind direction and wind speed using nonparametric regression and kernel smoothing methods, or a type of weighted average. NWR algorithms and output are explained in detail in “Source Region Identification Using Kernel Smoothing” (Henry et al. 2009).

The NWR tab allows a user to select a site on the map on the Near-Road Network tab, filter the input data, and run the NWR algorithms to generate 1) **polar plots** of the average concentration of a pollutant displayed on a continuous surface in polar coordinates, 2) **Cartesian plots** of the same data used in the polar plots but on a Cartesian (x and y) surface, 3) **sector apportionment density plots** normalized to the maximum average concentration, and 4) **sector mean and uncertainty plots**. A user can hover over the sector apportionment curve in the Sector apportionment density plot to display the data below the figure. Below the four plots is a map that displays a transparent version of the polar plot imposed over the Near-Road site location, and a wind rose with wind degree binning options. A table at the bottom of the page displays basic annual statistics of the input data.

To **select a site**, a user should click on a site on the Near-Road Network tab. A message will appear upon a site selection that describes the availability and source of wind data at the selected site. Near-Road sites that do not measure wind parameters but are located within 40 km of a site that measures wind parameters can still be selected; a message will appear at the top of the tab page that lists the source of the wind data. selected Near-Road sitecode, Site Name, City, State, and CBSA. Below the table is a drop down panel explaining the plots in detail and a drop down panel explaining how to use the user input section. Below the drop down panels are display and filter options for the tables and figure. A user can choose any parameter that is collected at a Near-Road site with wind data, filter the POC(s), year(s), month(s), day(s) of week, hour(s), and exclude entire days. Further, a user can set a maximum wind speed limit, maximum value for the selected parameter, keep or remove calm winds (< 1 mph), and define smoothing constants for wind speed and wind direction. User inputs and input and output metadata are displayed to the right of the input section, and will update upon new inputs. Finally, a user can select one of eight color schemes to apply to the polar and Cartesian plots. Figure 21 shows an example of user input.

Primary Data Filtering

Parameter <input type="text" value="Nitrogen dioxide (NO2)"/>	POC(s) <input type="text" value="1"/>	Year(s) <input type="text" value="2020, 2021, 2022"/>	Parameter: Nitrogen dioxide (NO2) POC(s): 1 Year(s): 2020, 2021, 2022 Month(s): Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec Day(s): Mon, Tue, Wed, Thu, Fri, Sat, Sun Hour(s): 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 Excluded Date(s): None
Month(s) <input type="text" value="Jan, Feb, Mar, Apr, M..."/>	Day(s) of Week <input type="text" value="Mon, Tue, Wed, Thu, I..."/>	Hour(s) <input type="text" value="00:00, 01:00, 02:00, 0..."/>	

Optional Inputs

Remove Calm Winds < 1 mph (0.45 m/s)? <input checked="" type="radio"/> Keep <input type="radio"/> Remove	Exclude Date(s) <input type="text"/>	Max. WS (m/s) <input type="text"/>	Maximum WS in input data (m/s): 8.23 Maximum pollutant concentration in input data: 51.7 Parts per billion Maximum NWR-estimated pollutant concentration: 16.4 Parts per billion WS smoothing constant (m/s): 1 WD smoothing constant (°): 10
Max. concentration <input type="text"/>	WS smoothing (m/s) <input type="text" value="1"/>	WD smoothing (°) <input type="text" value="10"/>	

Create Plots
Reset Inputs and Plots
Download Options

Select Color Bar

Figure 21: Filtering and display options for the NWR Tab.

Buttons allow a user to download 1) metadata file, 2) polar and Cartesian plots, 3) sector apportionment plots, 4) input data, 5) NWR results, 6) sector apportionment results, 7) the wind rose, and 8) the polar plot overlaid onto the map. The downloaded data reflect the data that are displayed in the figures. Figure 22 shows an example of a downloaded polar and Cartesian plot showing NWR-estimated average NO_2 .

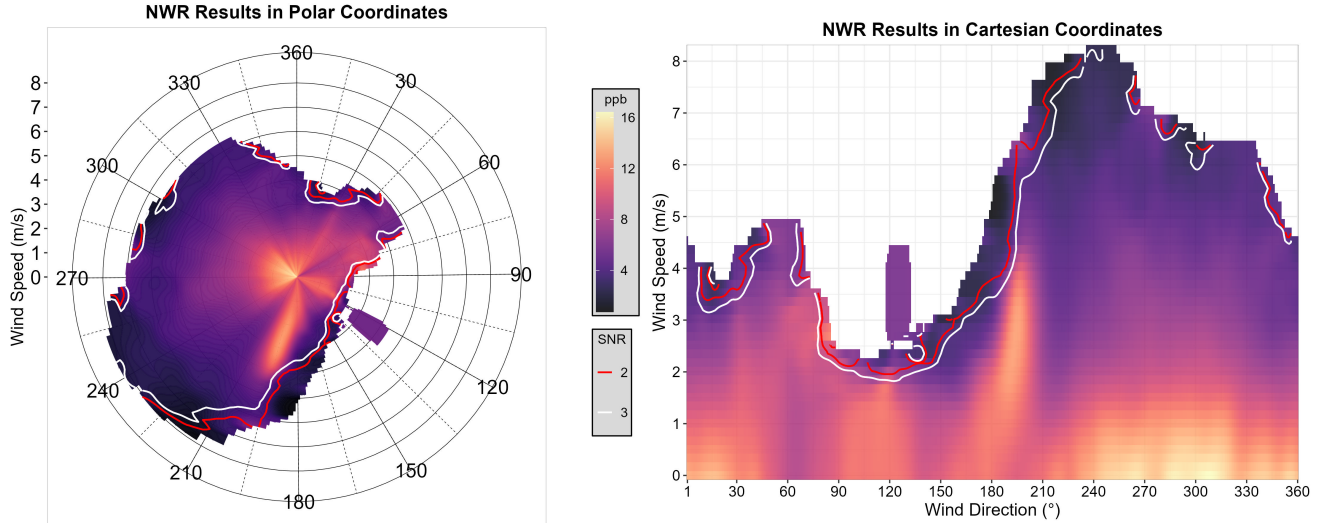


Figure 22: An example of downloaded NWR polar and Cartesian plots using data specified in the previous figure.

The map near the bottom of the page displays a transparent version of the polar plot overlaid onto the Near-Road site location. Point sources within 25 km of the Near-Road site from the 2020 National Emissions Inventory (NEI) data set are shown as icons on the map. A user can hover over the icons to see the name and type of point source (e.g., mining, airport) and click on an icon to display the average annual emissions of relevant parameters.

4.6.1 NEI Point Source Selection Methodology

4.6.1.1 Parameter Selection Point sources that measure *CO*, *Elemental Carbon*, *Organic Carbon*, *NO_x*, *PM_{2.5}* (filterable and condensable), and *Diesel – PM_{2.5}* (*PM_{2.5}* from certain diesel engines) were selected from the entire NEI Point Source data set. These parameters reflect parameters that may be measured at a Near-Road site.

4.6.1.2 Distance Filtering Point sources within 25 km of any given Near-Road site were selected from the point sources remaining after parameter filtering.

4.6.1.3 Emissions Because some point sources have multiple measurements for the same parameter, the average annual emissions were calculated per site and parameter. Sources with emissions less than the 10th percentile of average annual emissions were filtered from the data set.

5 References

Henry, Ronald, Gary A. Norris, Ram Vedantham, and Jay R. Turner. 2009. “Source Region Identification Using Kernel Smoothing.” *Environmental Science & Technology* 43 (11): 4090–97. <https://doi.org/10.1021/es8011723>.