

User Guide for the AirSensor DataViewer

How clean is
the air in my
neighborhood?

Which days had the
worst air quality last

Can I trust the
data from this
air sensor?

*You can answer these questions and more using
the AirSensor DataViewer!*



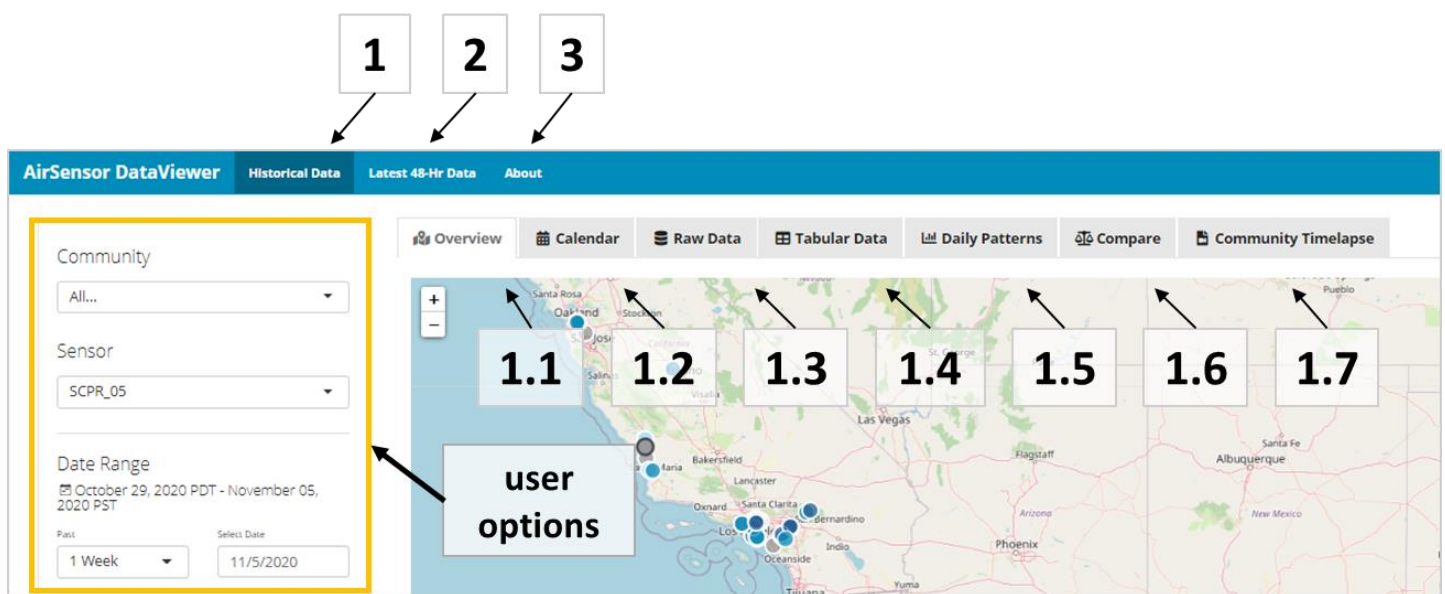
About the AirSensor DataViewer

Particulate matter sensors can answer questions about your local air quality. PM_{2.5} includes particles that are 2.5 µm (micrometers) in aerodynamic diameter or smaller. These particles are of concern to human health because they are able to penetrate deep into the lungs and in some cases enter the bloodstream. The US EPA funded South Coast AQMD to develop tools that you can use to learn about the levels of PM_{2.5} mass concentrations (µg/m³; micrograms per cubic meter of air) in your area.

The Science To Achieve Results (STAR) Grant provided funds to deploy networks of sensors in California communities. This guide will show you how to use the AirSensor DataViewer, which displays publicly accessible PM_{2.5} data from PurpleAir PA-II sensors deployed under the STAR Grant. You can focus on sensors in a specific community, search for PM_{2.5} data on specific dates, and discover local pollution patterns. The AirSensor DataViewer can generate visuals and graphs that help you interpret your data. If you have questions or comments about the AirSensor DataViewer, please contact the AQ-SPEC group at South Coast AQMD: <http://www.aqmd.gov/aq-spec/contact>.

Link to DataViewer Tool: <http://tools.mazamascience.com:6709/asdv/test/>

Overview of the interface:



At the top of the main/home page, there is a main menu bar that allows users to navigate between the “Historical Data”, “Latest 48-Hr Data”, and “About” pages. Below is a figure depicting the user options available on the “Overview tab” on the “Historical Data” page (on the left side of the tool). While available options will vary depending upon which tab is currently selected, the figure provides an idea of the user options available. Also depicted is the color scale used throughout the tool to illustrate pollutant levels.

User Options

Users can filter the sensors shown to focus on a specific community.

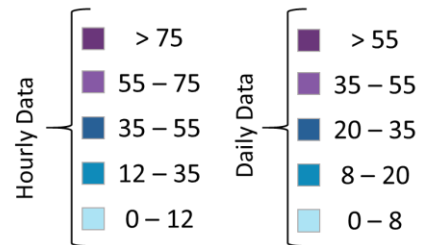
Here users can control the time frame by choosing (1) a date and (2) how many days previous to that date they would like to view.

Users can select a specific sensor for which to display the data.

“Share” provides a link directly to the visualization currently displayed.

“Download” provides users with data from the sensor and timeframe selected.

Color Scale for PM_{2.5} (µg/m³)

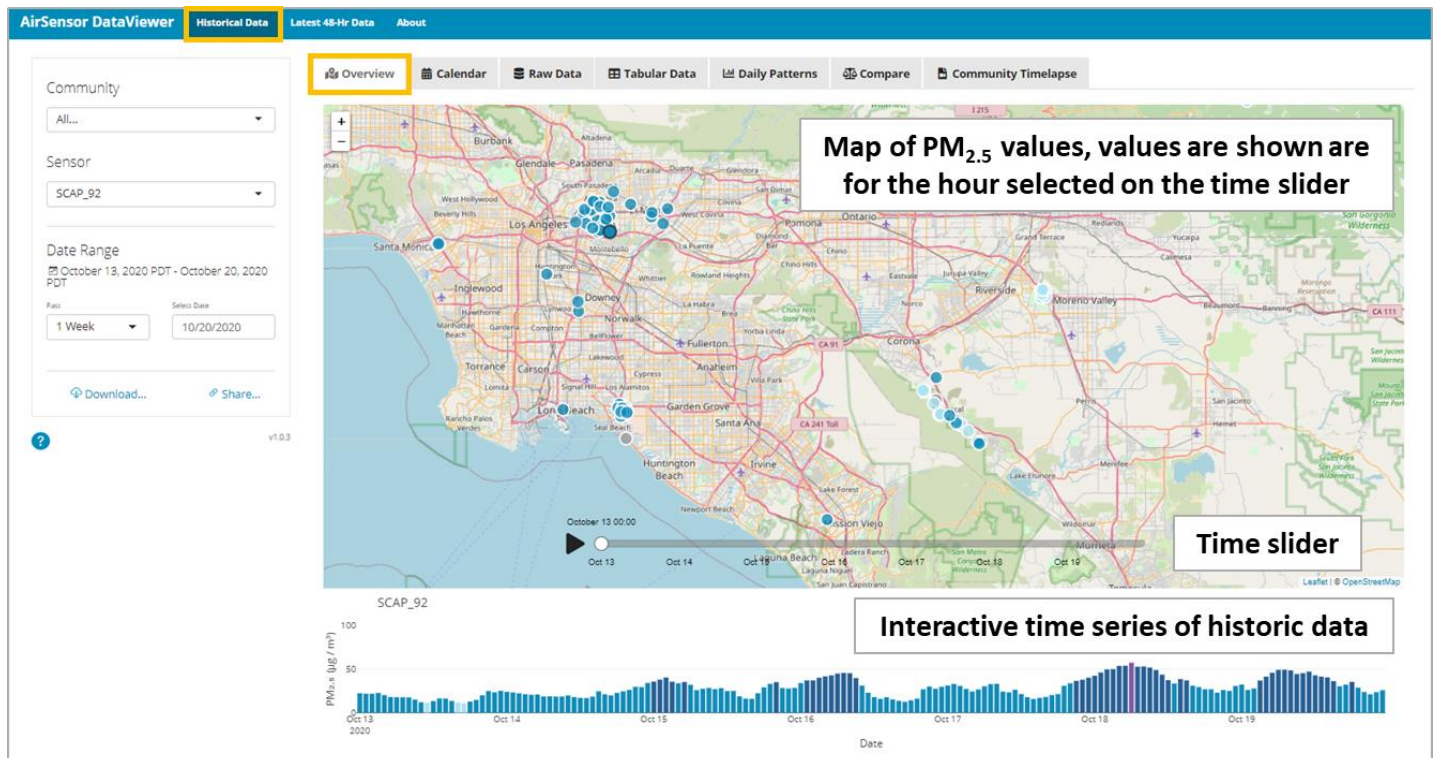


Throughout the DataViewer the above color scale is used to indicate different PM_{2.5} levels.

1. The “**Historical Data**” page is where users can analyze sensor data using a variety of different data visualizations, each tab leads to a different visualization.

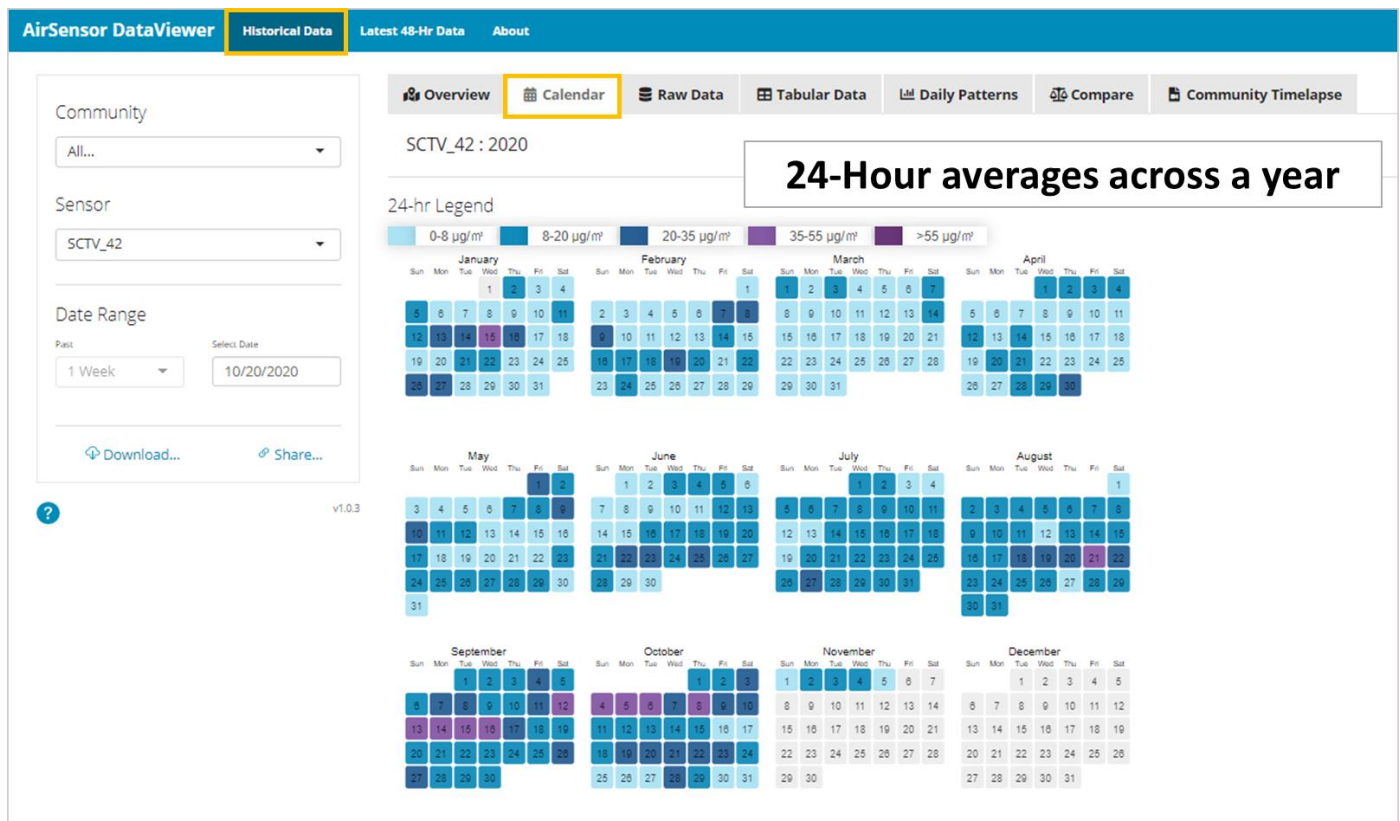
1.1 There are different ways you can use this tool to analyze your data.

- Beginning with the “**Overview**” tab on the “**Historical Data**” page, a map illustrates where sensors are located and provides a customizable and interactive time series. Sensors are located in West Los Angeles, Alhambra/Monterey Park, El Monte, South Gate, Seal Beach, Redlands, Riverside, Sycamore Canyon, Temescal Valley, Imperial Valley, Nipomo, Paso Robles, Oakland and Richmond.
- Looking at the time series, consider how PM_{2.5} levels have varied in the previous week or over the previous 30 days.
- You can also examine the effects of specific events on local PM_{2.5}, for example, how did 4th of July fireworks emissions impact your neighborhood?



1.2 The “Calendar” tab on the “Historical Data” page, provides an overview of the data from an entire year.

- Use this plot to spot interesting trends or events.
- How do $PM_{2.5}$ levels vary from season to season?
- Do the days where elevations in $PM_{2.5}$ occur correspond to days when you might expect elevated $PM_{2.5}$ levels?

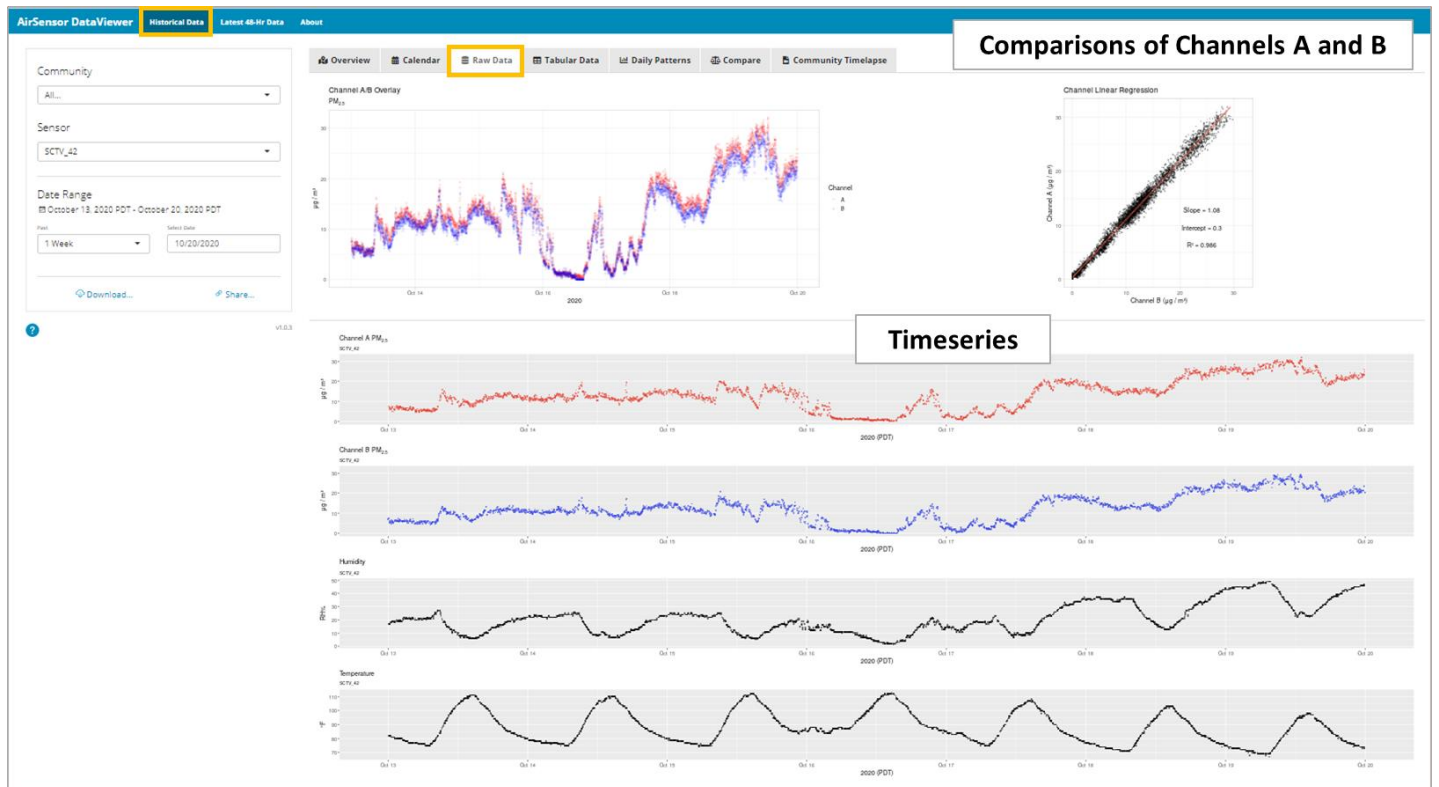


Usage:

- Select a sensor using the drop-down menus.
- The calendar shows the average daily $PM_{2.5}$ mass concentrations. The greater the $PM_{2.5}$ concentration, the darker the day is shaded. If the daily cell is white, a value was not recorded for the day. Hover over a specific day to view the exact 24-hour average $PM_{2.5}$ mass concentration.

1.3 The “Raw Data” tab on the “Historical Data” page, displays the data available from each sensor prior to processing and QA/QC.

- You’ll notice that there are two graphs of PM_{2.5} data labeled Channel A and Channel B, this is because there are two sensors inside each individual PurpleAir unit.
- For more information on Channels A and B data and the data processing please refer to the final page of this guide.
- At the bottom, this tab provides plots directly comparing Channels A and B, which may be used to better understand how well a sensor is functioning.
- Additionally, you can see temperature and humidity data in this tab.



Usage:

- Select a sensor using the drop-down menus and select a timeframe using “Select Date” and “Past”.
- The plots in this tab show the “raw data” (or data that comes directly from the sensor, before any time averaging or processing has been applied).
- Below the time series plots are two more plots that illustrate how well the two PM_{2.5} signals from Channels A and B agree with each other – significant differences between the channels may be an indication that either one or both of the duplicate sensors is malfunctioning (agreement is signified by a ‘Slope’ and an R² close to 1.0).

1.4 The “*Tabular Data*” tab, on the “*Historical Data*” page, displays raw numeric data from the selected sensor in a table format. On this page, users can view, search, and download data from the specified period and from the selected sensor. You can also use this tool to view the precise values that occurred at specific times, for example to find the peak value on a given day.

The screenshot shows the AirSensor DataViewer interface. The top navigation bar includes 'AirSensor DataViewer', 'Historical Data' (highlighted), 'Latest 48-Hr Data', and 'About'. Below this is a secondary navigation bar with tabs: 'Overview', 'Calendar', 'Raw Data', 'Tabular Data' (highlighted), 'Daily Patterns', 'Compare', and 'Community Timelapse'. On the left, a sidebar contains filters for 'Community' (set to 'All...'), 'Sensor' (set to 'SCAP_14'), and 'Date Range' (set to 'October 13, 2020 PDT - October 20, 2020 PDT'). The 'Past' filter is set to '1 Week' and 'Select Date' is '10/20/2020'. The main area displays a table with the following data:

Sensor Name	Community	Sensor Model	Longitude	Latitude	State	Country
SCAP_14	Alhambra/Monterey Park	PA-II (PurpleAir)	-118.11	34.08	CA	US

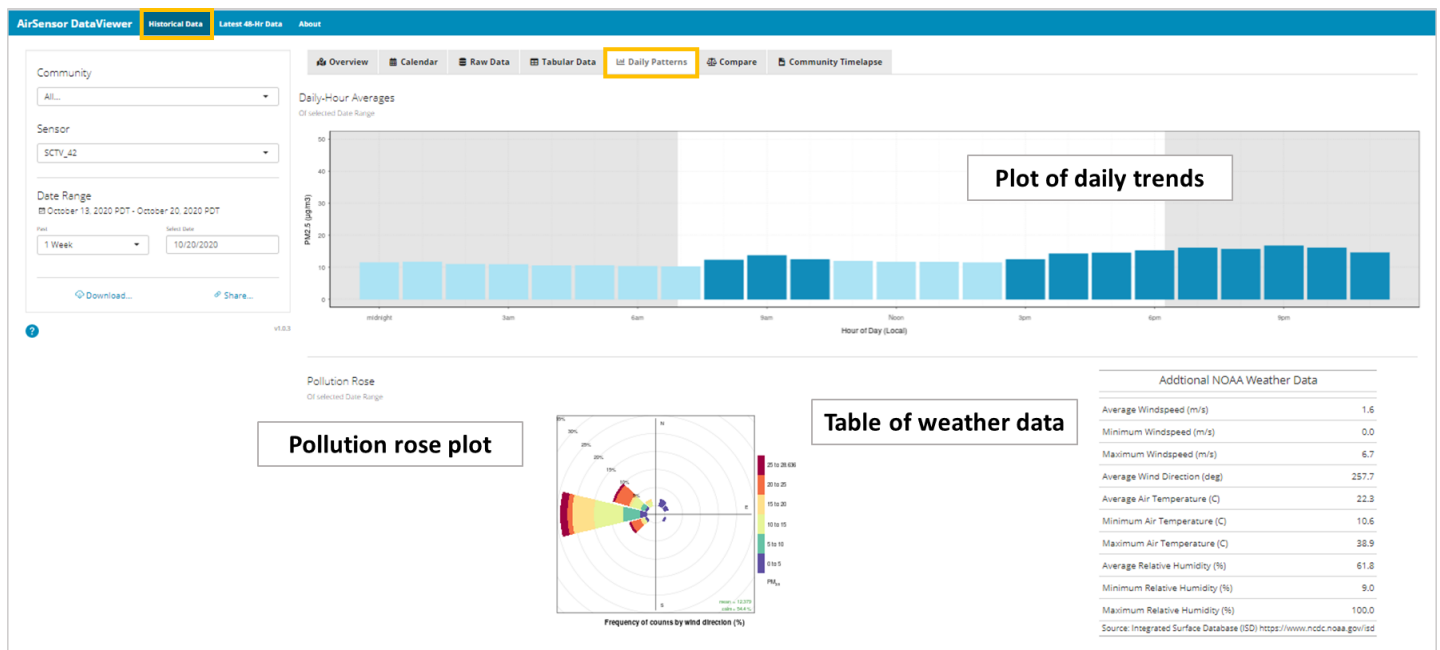
	Datetime (UTC)	PM2.5 Ch. A ($\mu\text{g} / \text{m}^3$)	PM2.5 Ch. B ($\mu\text{g} / \text{m}^3$)	Temperature (F)	Relative Humidity (%)
1	10/13/2020, 12:00:00 AM	15.53	16.39	79	40
2	10/13/2020, 12:02:00 AM	14.5	16.22	79	40
3	10/13/2020, 12:04:00 AM	15.29	16.28	79	41
4	10/13/2020, 12:06:00 AM	15.36	16.49	79	41
5	10/13/2020, 12:08:00 AM	16.4	15.66	79	42
6	10/13/2020, 12:10:00 AM	16.14	15.53	79	42
7	10/13/2020, 12:12:00 AM	15.97	16.69	79	41
8	10/13/2020, 12:14:00 AM	15.34	15.09	79	41
9	10/13/2020, 12:16:00 AM	16.24	16	79	40
10	10/13/2020, 12:18:00 AM	14.93	14.78	78	41
11	10/13/2020, 12:20:00 AM	16.14	16.31	78	41
12	10/13/2020, 12:22:00 AM	14.68	15.39	78	41
13	10/13/2020, 12:24:00 AM	15.63	15.88	78	41
14	10/13/2020, 12:26:00 AM	15.21	15.76	78	41
15	10/13/2020, 12:28:00 AM	15.41	15.67	78	41
16	10/13/2020, 12:30:00 AM	15.43	15.02	78	40
17	10/13/2020, 12:32:00 AM	14.81	15.54	78	40
18	10/13/2020, 12:34:00 AM	14.48	15.46	78	40
19	10/13/2020, 12:36:00 AM	15.4	14.7	78	40

Usage:

- Select a sensor using the drop-down menus and select a timeframe using “Select Date” and “Past”.
- You can use the search bar in the top right to help you explore the data.
- You can also download the data as a .csv file.

1.5 In the “Daily Patterns” tab on the “Historical Data” page, there is another plot to help reveal interesting trends as well as tables and plots providing supplemental weather data.

- Looking at the Daily Trends plot, is there a particular time of day when PM_{2.5} is highest? When is PM_{2.5} lowest?
- What can the pollution rose tell you about potential sources of PM_{2.5}?
- In a pollution rose, the shape summarizes where the wind was coming from during the specified period and the colors indicate where the wind was coming from when various pollutant concentrations were observed.
- Additionally, the table provides some statistics related to weather conditions, these can help you check whether rain may have occurred during the selected timeframe, or learn what the temperature highs and lows were.

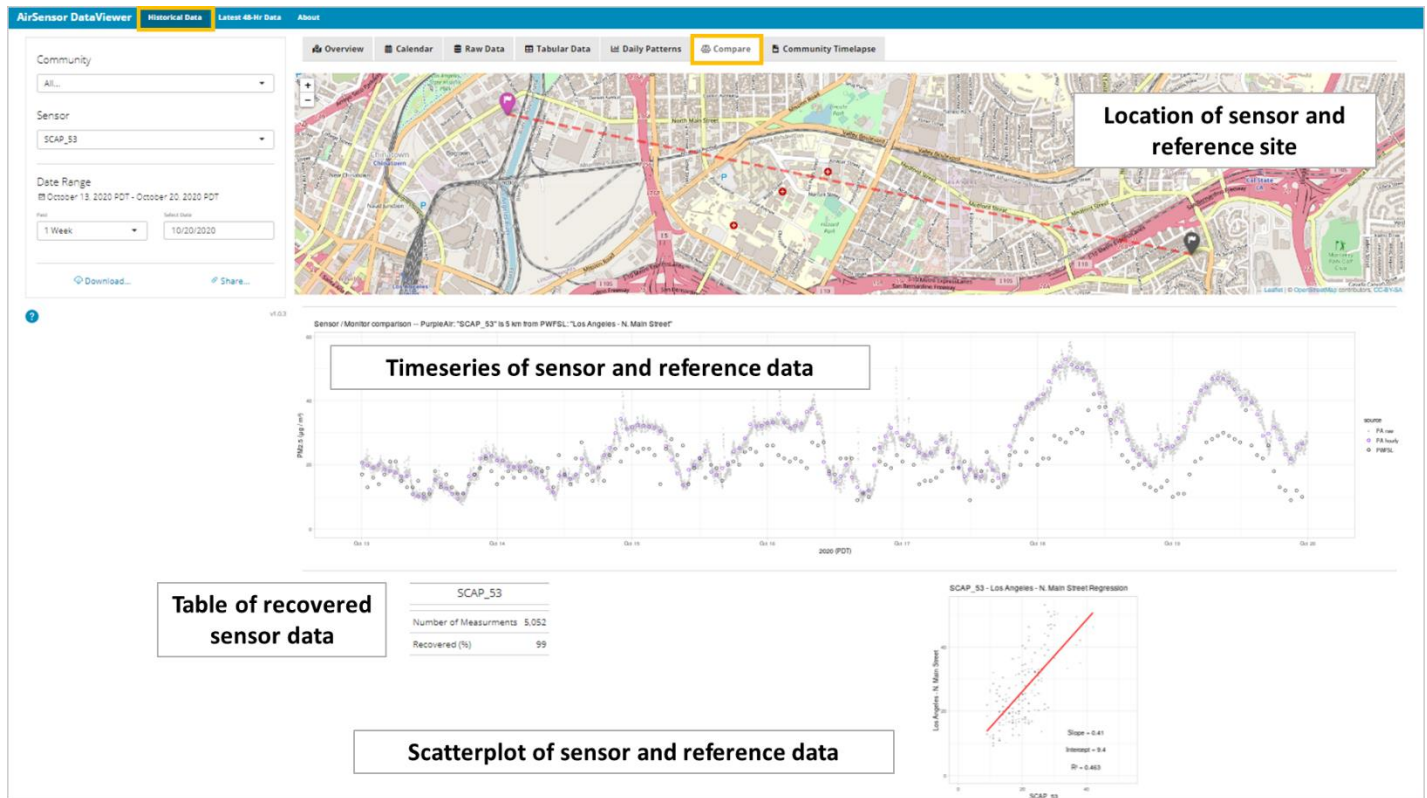


Usage:

- Select a sensor using the drop-down menus and select a timeframe using “Select Date” and “Past”.
- The first plot shows the average for each hour of the day across the timeframe selected.
- Below the plot, the table of weather data and the pollution rose can provide additional useful information. For example, a high humidity value may be an indication that a precipitation event occurred in that timeframe while the pollution rose can illustrate the direction from which the wind was blowing when high or low PM_{2.5} levels were observed – possibly providing some insight into potential sources.
- The weather data is accessed from the National Oceanic and Atmospheric Administration’s (NOAA) Integrated Surface Database (ISD) using the worldmet R-package.

1.6 The “Compare” tab on the “Historical Data” page, allows users to compare the data from a sensor to the data from the nearest reference site (or monitoring site operated and maintained by a government agency).

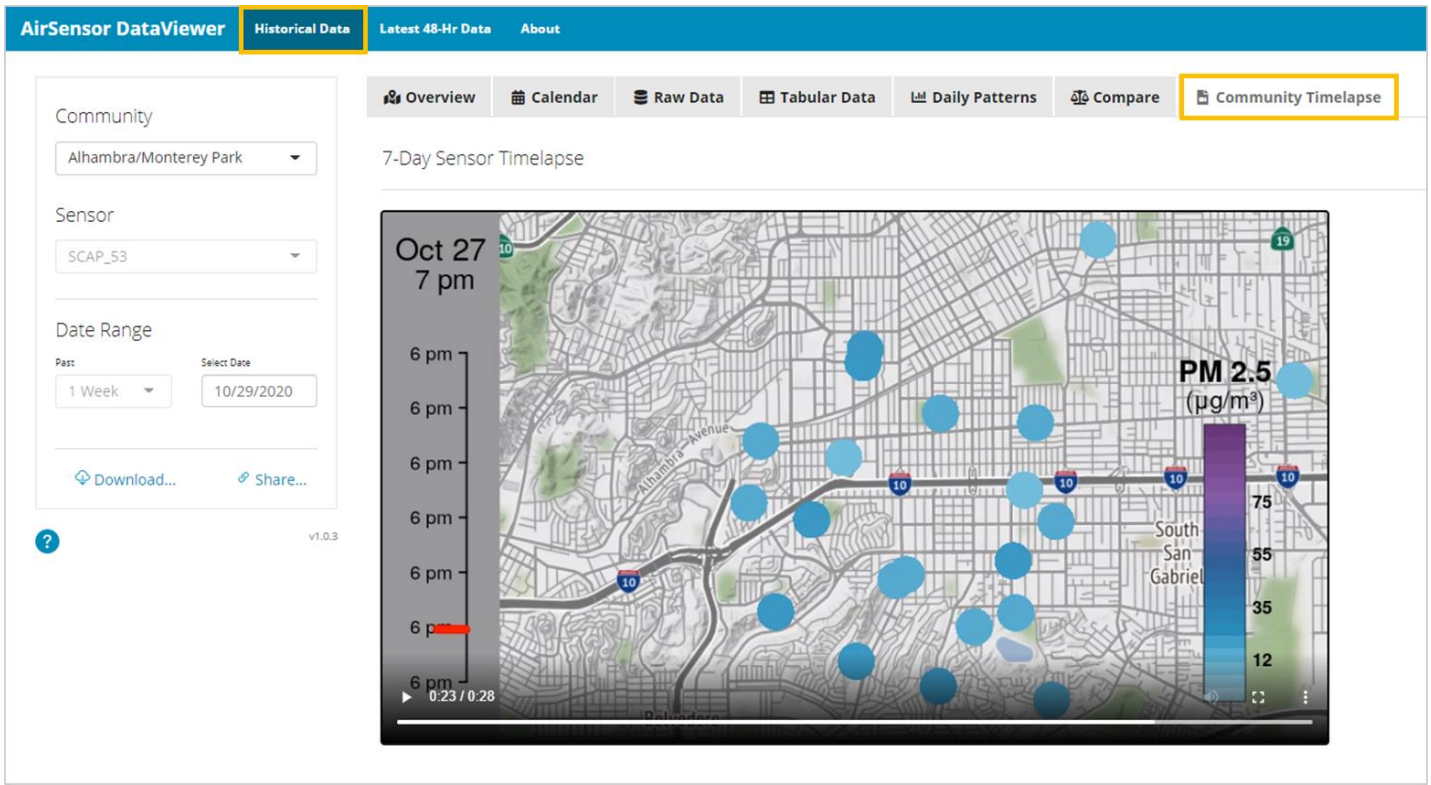
- Consider, does the sensor data agree with the reference data in terms of trends and levels?
- If not, why might this be the case? Consider the distance to the reference site, the geography of the region, and potential sources nearer to one site or the other.



Usage:

- Select a sensor to compare to the nearest reference site by clicking on the map or by using the drop-down menus, then select a timeframe using “Select Date” and “Past”.
- The plots below the map illustrate (1) how well the sensor and nearest reference site agree with each other (see the ‘Sensor-Monitor Correlation’ plot) and (2) how the trends between the sensor and the reference site compare (see the ‘Sensor-Monitor Comparison’ plot). Remember, that many factors influence how similar or different the data may be from the two sources. This tab is intended for informational purposes and not to quantitatively assess the performance of an individual sensor.
- The distance between the reference site and the sensor is available on the map and it is stated in the title of the ‘Sensor-Monitor Comparison’ plot.
- In terms of the statistics, a slope close to 1.0 indicates that the low-cost sensor and the reference site are reflecting similar levels. A slope greater than 1.0 indicates that higher values are seen at the regulatory monitoring site and vice versa for a slope of less than 1.0. The intercept can be an indicator of bias (e.g., whether the sensor may be consistently under or over-predicting pollutant concentrations). The coefficient of determination R^2 tells us how well the trends agree between the two sites; an R^2 closer to 1.0 indicates more agreement and an R^2 closer to 0.0 indicates less agreement.
- There is also a sensor status table, which provides insight into how well the sensor is functioning. For example, if the percent of data recovered is small, then this may be an indication of a performance issue with the sensor.

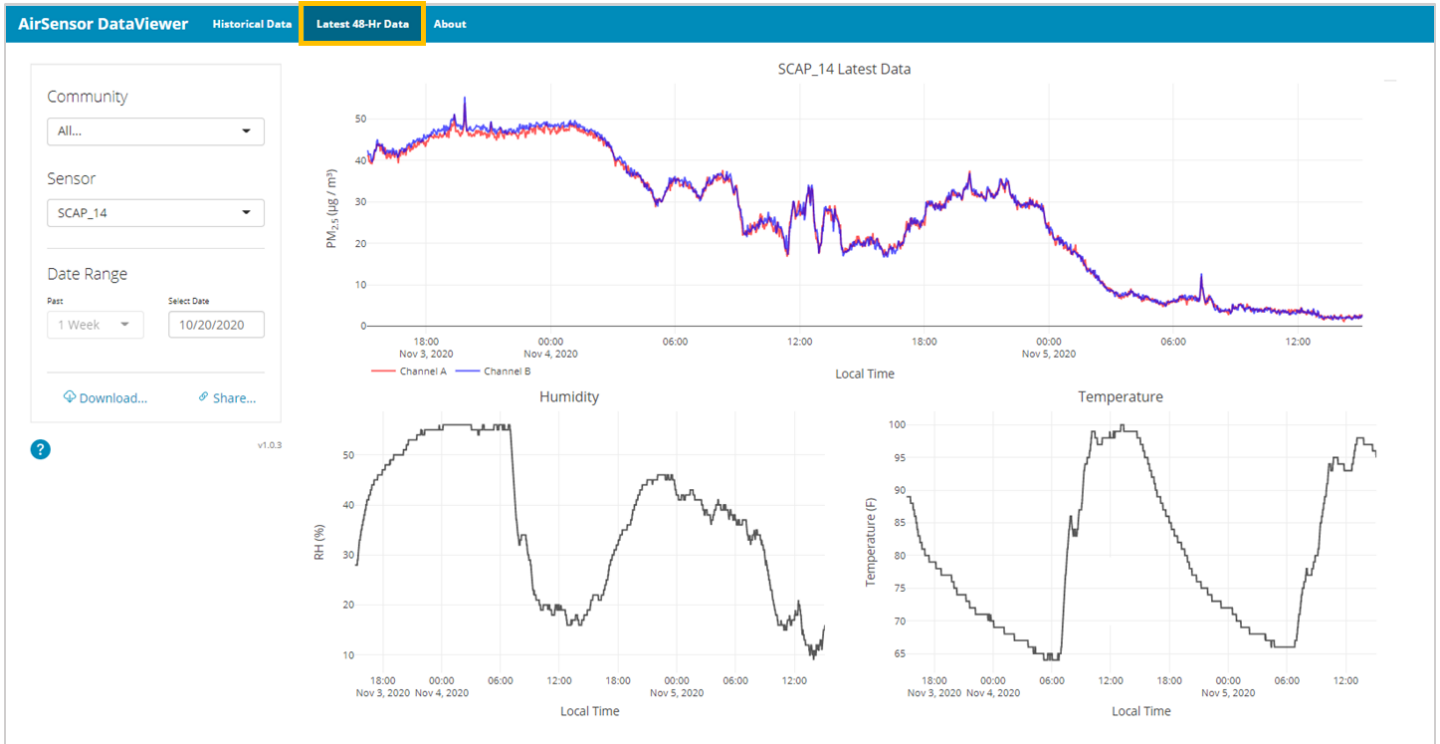
1.7 The “*Timelapse*” tab on the “*Historical Data*” page, provides users with animations of $PM_{2.5}$ and how they change over time in each STAR Grant community. Are the changes consistent throughout the community? Do any sensors stand out?



Usage:

- Select a community and a date in the panel on the left.
- Note that the date selected will be the end date of the time-lapse animation.
- The corresponding 7-day time-lapse illustrates the changes in $PM_{2.5}$ pollutant concentration over time.

2. On the “**Latest 48-Hr Data**” page, you can view the most recent data from the selected sensor. This page is a particularly useful reference during air quality events, for example when a sensor is detecting wildfire emissions. You can use this page to compare sensor data to your own observations about the air quality, such as what you see or smell. Consider whether your observations and the sensor data agree or disagree?



Usage:

- Select a sensor using the drop-down menus and select a timeframe using “Select Date” and the most recent up-to-date data will automatically be shown.

3. On the “About” page, you can find background information on the DataViewer and its intended purpose. You can also find a description of the QA/QC procedures applied and important disclaimer information.

Information Regarding the Data Processing and QA/QC Procedures:

Each PurpleAir PA-II sensor unit includes two raw, duplicate sensors referred to as Channel A and Channel B. Processed data is data for which Channels A and B have been compared and averaged together. Unless otherwise stated, all the visuals in the DataViewer display processed data (averages of both Channels). Wherever unprocessed or “raw” data is shown (e.g., the “Raw Data” tab on the “Historical Data” page or on the “Latest 48-Hr Data” page), both Channel A and Channel B are shown separately. Processed data has undergone the following QA/QC procedures: (1) removal of values outside of the manufacturer defined specifications for the sensors, (2) pollutant values for Channel A and Channel B are averaged on an hourly basis, (3) if the proportion of points contributing to the hourly average meets the minimum requirement and the hourly averages are judged to be not statistically different, according to a student’s t-test, then the hourly averages for Channels A and B are averaged together – producing a single value for each hour. More detail on the procedures and functions used is available in the AirSensor R-Package documentation (<https://mazamascience.github.io/AirSensor/index.html>).



Guide last updated: November 10, 2020

About the DataViewer

The DataViewer tool was built using the code and functionalities available in an open-source R package, called the AirSensor package. The DataViewer tool as well as the AirSensor R-Package were developed through a collaboration between the AQ-SPEC group at the South Coast Air Quality Management District (South Coast AQMD), a regional governmental agency in Southern California, and Mazama Science, a software company in Seattle, WA. This tool is intended to support data exploration and analysis by community members participating in the US EPA-funded Science To Achieve Results (STAR) Grant at the South Coast AQMD, entitled “Engage, Educate and Empower California Communities on the Use and Applications of Low-cost Air Monitoring Sensors.” Funding for the development of this tool was provided through the abovementioned US EPA STAR Grant.

The tool accesses publicly available data from low-cost sensors in the Particulate Matter (PM) sensor (model PA-II, PurpleAir, Draper, UT) network associated with the STAR Grant project. Unless otherwise stated, the DataViewer displays processed, hourly-averaged PM_{2.5} data. Processed data is data to which quality assurance and quality control (QA/QC) procedures have been applied in order to exclude data from sensors that may be malfunctioning.

A journal article is available introducing the AirSensor package and DataViewer web application: Feenstra, B., Collier-Oxandale, A., Papapostolou, V., Cocker, D., & Polidori, A. (2020). “The AirSensor open-source R-package and DataViewer web application for interpreting community data collected by low-cost sensor networks.” *Environmental Modelling & Software*, 134, 104832. <https://doi.org/10.1016/j.envsoft.2020.104832>

Disclaimer

This tool is intended to be used for educational and informational purposes only. Furthermore, the code used to build this tool, the QA/QC procedures, and the different features of this tool may be subject to revision at any time depending on the needs of the project. This work was developed under Assistance Agreement No. RD83618401 awarded by the U.S. Environmental Protection Agency to the South Coast Air Quality Management District. It has not been formally reviewed by EPA. The views expressed in this work are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this work.