SENTINEL R Shiny Application User Guide

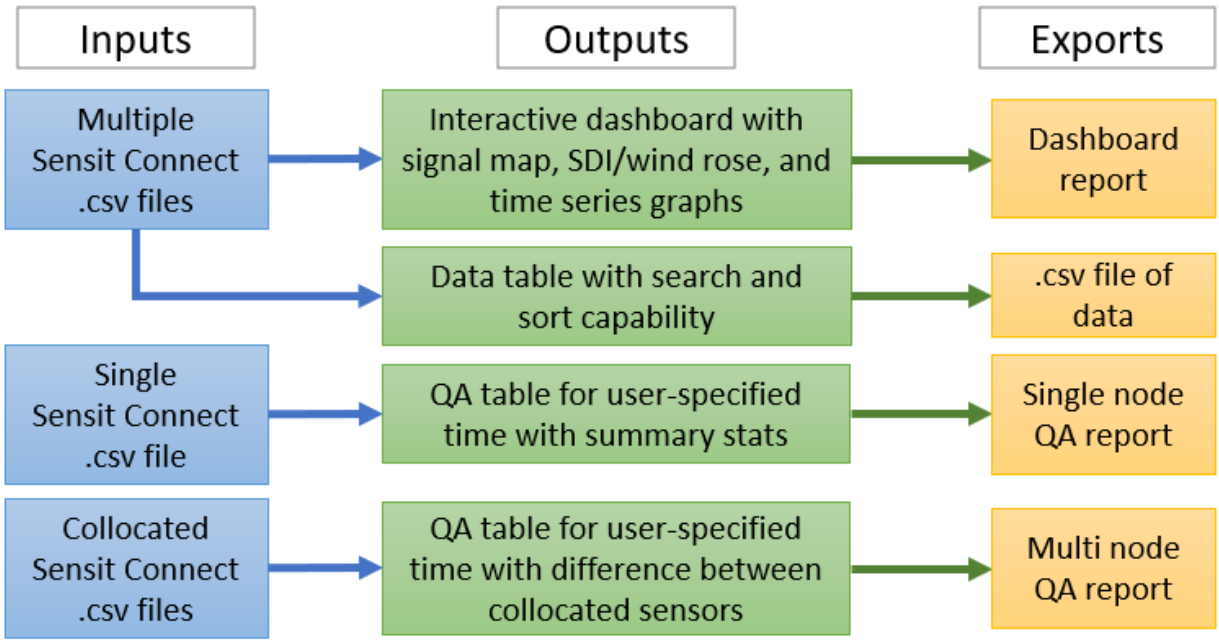
Version: 1.0 (January 2023)

Contents

App Overview

Sentinel (SENsor Intelligent Emissions Locator) is a R shiny application[1,2] that can be run with an access link on cloud.gov. It is currently in a development/prototype phase. This application is intended to provide non-coding users with visualization and graphical insights to Sensit low-cost fenceline sensor (SPod) data. These sensors can generate a large amount of data, which can be overwhelming for users to process manually. This application is also useful for generating Quality Assurance (QA) tables as required by the SOP for Sensit SPod Fenceline Sensor and Canister Grab Sample System Deployment and Operation (J-AMCD-SFSB-SOP-4380-2).

This app can be most easily run with the cloud.gov link but interested users can find the code on the Sentinel Bitbucket repository and run the code from R Studio. An overview diagram of the application inputs and outputs is shown below:



Downloading Data with Sensit Connect

Data can be downloaded from the Sensit Connect Website (sensitconnect.net) at a daily frequency for each individual sensor. Most SPod sensors are programmed automatically to report data at 30 seconds. This is acceptable by the app; however, the original code is programmed for 10 second frequency of data output. For more information on changing output settings on Sensit SPods, consult the Sensit Spod user manual.

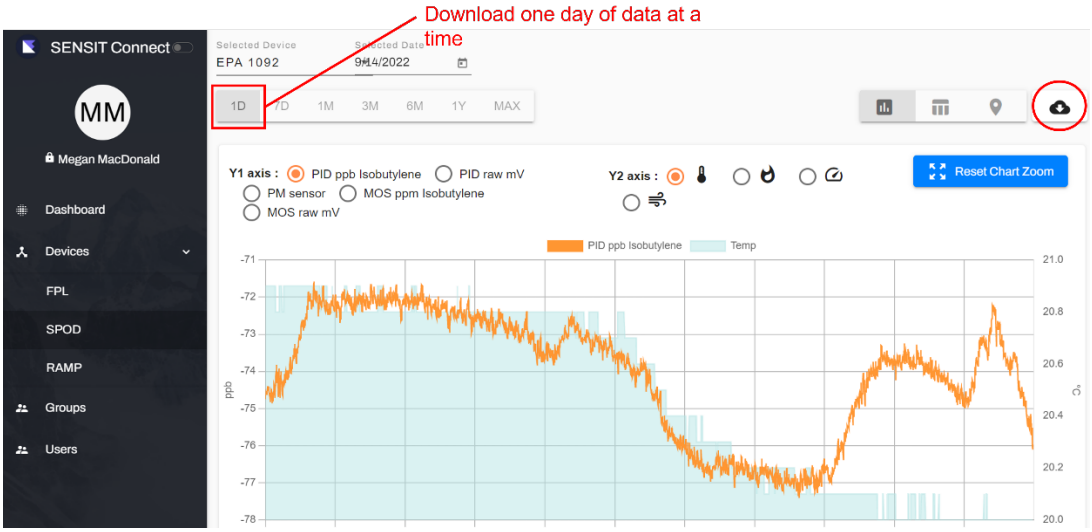


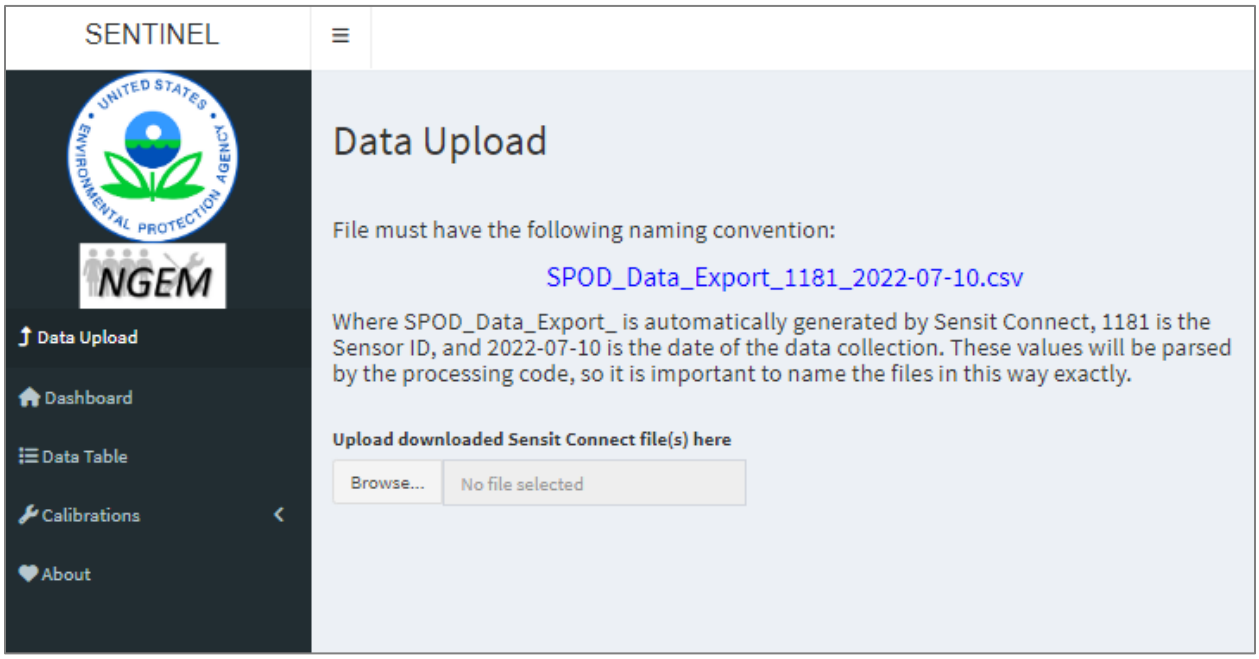
Figure 1: Dashboard of a specific SPod on the Sensit Connect website

Data must be downloaded as a .csv file with the following naming convention:

"SPOD_Data_Export_1181_2022-07-10.csv"

Where "SPOD_Data_Export_" is automatically generated by Sensit Connect, "1181" is the Sensor ID, and "2022-07-10" is the date of the data collection. These values will be parsed by the processing code, so it is important to name the files in this way exactly. These files can be saved anywhere, but it is recommended to keep them grouped in a folder for ease of uploading to the SENTINEL Shiny app.

Data Upload



All pages of SENTINEL contain the sidebar (which can be closed with the three horizontal lines button) and the main page. The sidebar options are Data Upload (the landing page), the interactive dashboard, the data table, calibration (single node or multi node) and an about page that contains resources. The user can click the NGEM logo to be taken to the NGEM webpage. The landing page is where the user can input their files that have been downloaded and named as stated above from the Sensit Connect site. These files should be in .csv form.

Once these files are uploaded, a status bar in the form of three blue vertical lines will appear as the files are processed. This processing involves applying automatic QA functions, that scan the data and look for values out range or repeating values and flag them accordingly in an appended QA column. This column is later visible in the Data Table viewer. Flag values in this column are as follows:

#	Description
0	No QA issues, passing values
1	Missing value or NA
2	Very negative PID (> -10 ppb)
3	Very high PID (> 1500 ppb)
4	Repeated concentration, wind speed or wind direction values > 10 times in a row
5	Check for illogical wind values (speed > 12, direction outside of 0-360)

The script will then conduct baseline correction on the 10 second data using the getBaseline function, which is stored in the app folder. This function calculates a baseline with the “df” input set to 4, which is a slowly varying fit. It then subtracts this out to minimize any environmental drift present in the data.

The u and v vector directions are then calculated based on the 2D sonic wind direction and wind speed. These will be vector averaged in the final step of summarizing the data to 5 minutes. If the wind

direction data were averaged without converting to u and v vectors, there would be inaccuracies when averaging 360 and 0 degrees. The formulas for these conversions are below:

$$u = Wind\ Speed \times \sin(2\pi \times \frac{Wind\ Direction}{360})$$

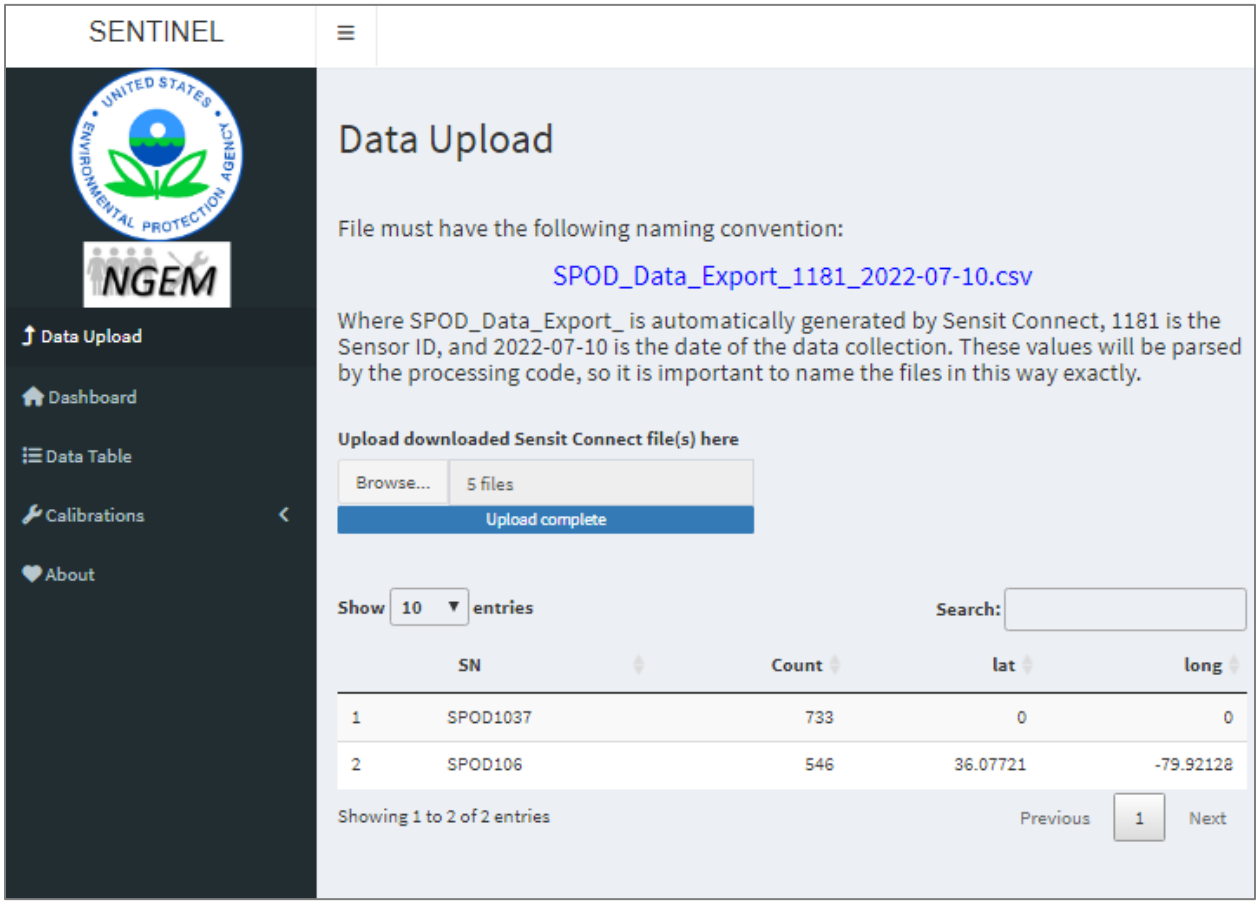
$$v = Wind\ Speed \times \cos(2\pi \times \frac{Wind\ Direction}{360})$$

The code then checks for latitude and longitude values. It is not required to have GPS configured on the sensor to use the Sentinel Shiny App. If latitude and longitude columns are not detected, the code will input 0 values for these columns and the mapping capabilities in the dashboard will be limited.

Once these initial processing steps are done, the code will use the dplyr[3] package to group the data in whatever frequency it is uploaded in to 5 minutes. The following columns are created in this aggregation for each 5-minute period, which will be used for analysis throughout the application. The method Detection Limit (MDL), Wind Speed (ws), and Wind Direction (wd) columns are appended after the 5-min aggregation. Canister columns will only be populated if the sensor is configured for canisters and canisters have been collected previously/at that time.

Column	Definition	Calculation
bc.pid.ppb	Background corrected data (ppb)	5-min mean
Pid.sd	The standard deviation of background corrected data (ppb)	5-min standard deviation
rawPID_ppb	Raw concentration data (ppb)	5-min mean
temp	Temperature data (C)	5-min mean
RH	Relative humidity data (%)	5-min mean
pressure	Pressure data (mbar)	5-min mean
u.wind	Calculated u values (see above formula)	5-min mean
v. wind	Calculated v values (see above formula)	5-min mean
s1temp	Sensor temperature (arbitrary units)	5-min mean
s1heat	Sensor heater output (0 = off, 255 = fully on)	5-min mean
set	Sensor setpoint (arbitrary units)	5-min mean
bat_volt	Battery voltage (volts)	5-min mean
chg.current	Charge current in milliamps	5-min mean
opp.current	Operating current in milliamps	5-min mean
trigportstat	Port status (NA indicates no collections)	List of unique values in 5-min period
trigactivestat	Trigger status (1 indicates active trigger)	List of unique values in 5-min period
trigactiveflag	Active Port (numbers indicate which port is currently active)	List of unique values in 5-min period
trigsampleflag	Event status (current sampling event or complete sampling event)	List of unique values in 5-min period
lat	Latitude (deg.)	Unique values in 5-min period
long	Longitude(deg.)	Unique values in 5-min period
QA	QA flags (See table on page 3)	List of unique values in 5-min period
MDL	Method Detection Limit	3 * median daily standard deviation
wd	Wind Direction (deg)	Atan2 function
ws	Wind Speed (mph)	Sqrt function
SN	Serial number of Unit	Parsed from file name

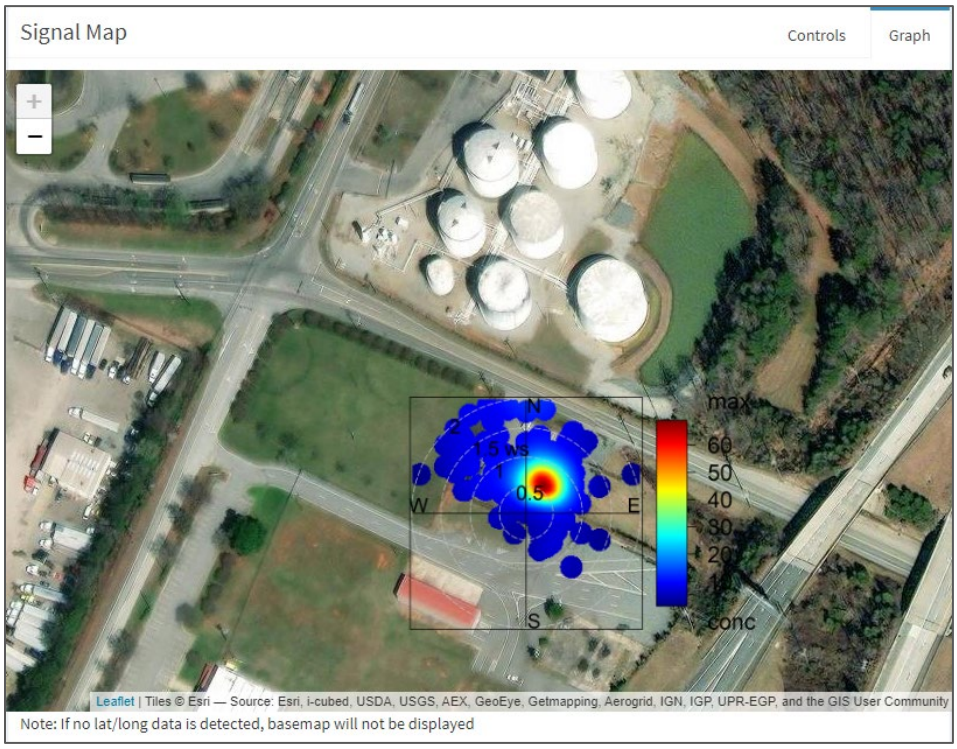
The application will display a simple summary table once all files are uploaded showing the unique serial numbers, latitude/longitude values, and a row count of how many entries are associated with that serial number. Any sensors with no latitude or longitude columns (not GPS-enabled) will show a 0 for latitude and longitude values.



Dashboard

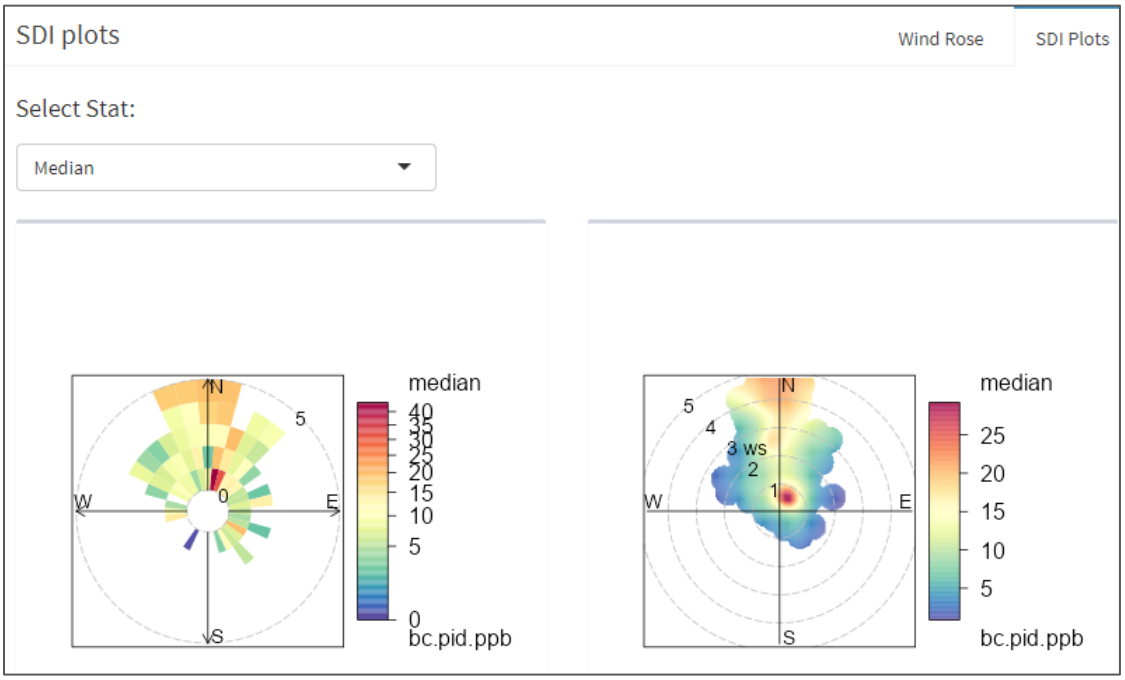
Once the user has uploaded their data through the data upload page, the dashboard components are loaded based on those files. The dashboard (developed using the shinydashboard package[4]) consists of a signal map, SDI plots, and Time series graphs. There is drop down menu that will be automatically populated with available units to display. There is also a report export button. More detail on each of the sections is given below.

SIGNAL MAP



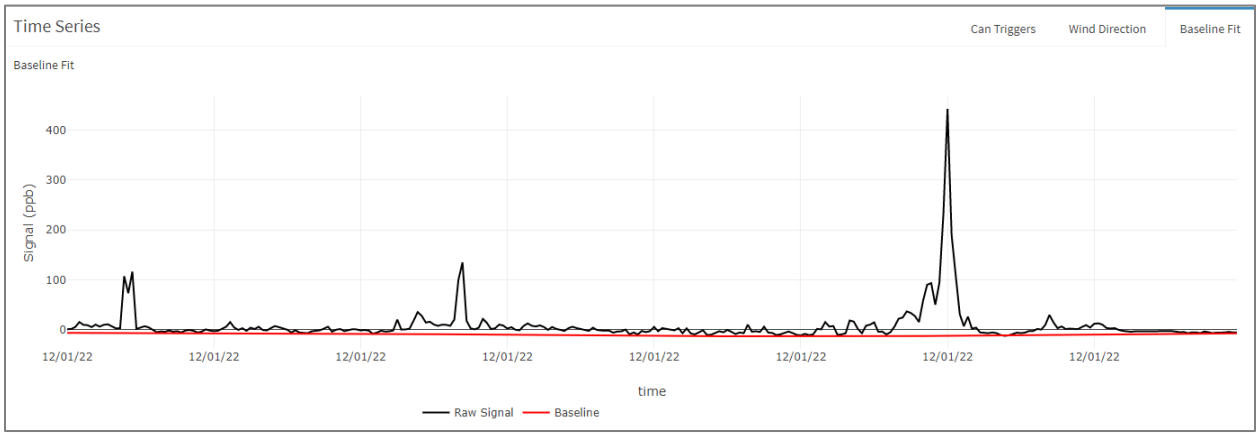
The signal map pane consists of 2 tabs. The first is a controls tab, which allows the user to select some inputs, and the second is the graph display. The graph display can be panned and zoomed with the mouse as well as using the +/- controls in the upper left-hand corner. These plots are built using the polarmap function in the openair package[5]. This is run with a leaflet basemap, from Esri World Imagery. We refer to this kind of plot as a Source Direction Indicator Plot. On the Controls tab, the user can use the slider to limit wind speed on the SDI plot to a certain range. The user can also choose the different statistic they would like to see applied to the graph, with the options of Median, weighted mean, or maximum. More information on these stats can be found in the SDI plot section. If the sensor did not include latitude and longitude data, the base map will not appear, and instead a grey screen will be displayed.

SDI PLOTS



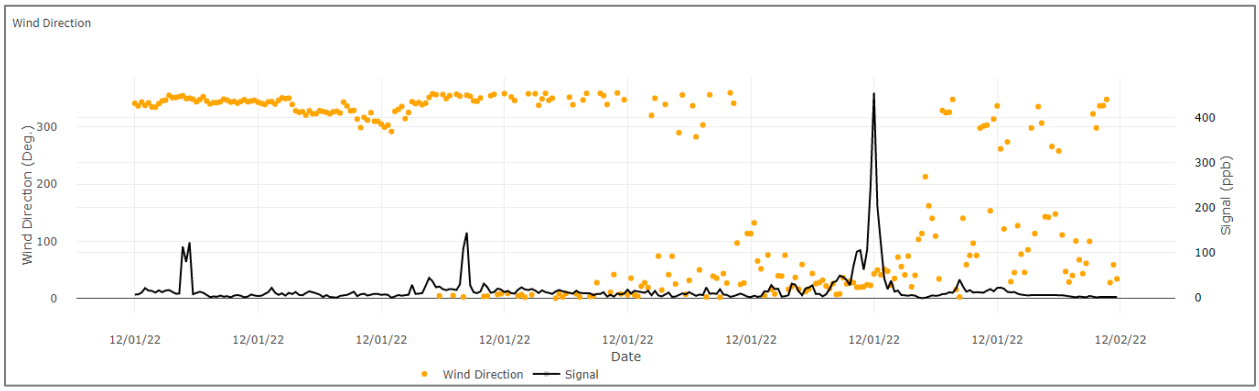
This panel contains two tabs, one containing a wind rose and the other containing two SDI plots. The SDI plot panel contains a polar frequency plot on the left, and an interpolated SDI plot (polarplot) on the right. More information about these functions can be found in the openair Book and the Openair user manual. The authors of these packages also provide some extra info in a companion R Journal article. The user has the option to select between several stats. These stats are each applied to the wind speed and wind direction bins shown in the polar frequency, and then interpolated with smoothing parameters in the SDI plot. The weighted mean stat is calculated as (concentration * frequency of occurrence). The wind rose plot shows overall wind conditions as a frequency of counts by wind direction. Wind speed is binned by color in the wind rose plots.

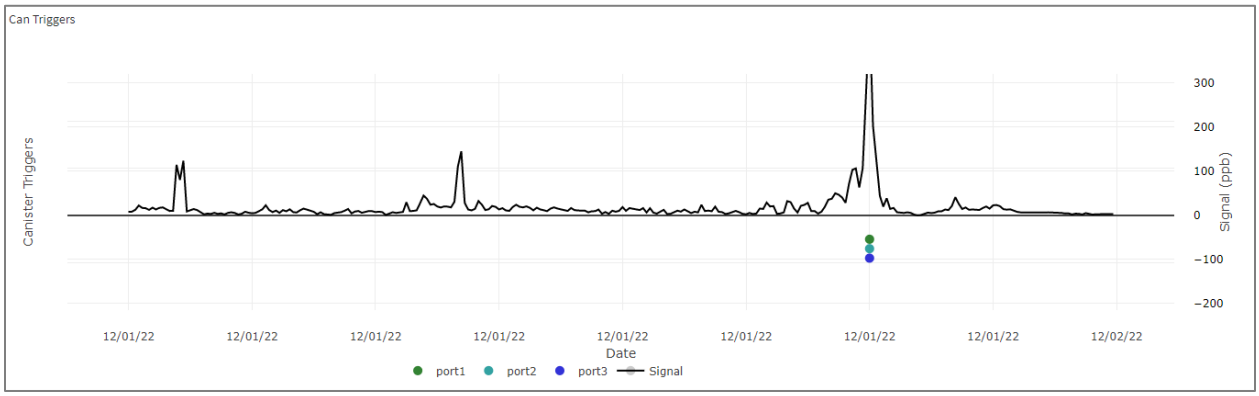
TIME SERIES GRAPHS



There are three tabs in the time series pane: a baseline fit tab, a wind direction and concentration tab, and a can trigger and concentration tab. All three of these plots are made using the plotly package[6], which allows for the user to hover over points and zoom/pan on the graph. If the user hovers the mouse

in the upper right-hand corner of the graph, there are a suite of options available. One of these, the camera icon, can export a .png image of the graph. The user can also select an axis and drag it up or down to pan down only one axis at a time. The baseline fit tab shows the raw signal as a black trace and the baseline fit as a red trace. The wind direction tab shows the wind direction points as yellow dots and the signal as a black trace. This is useful for matching the periods of elevated signal with wind direction. Finally, the Canister trigger graph shows the signal as a black line trace and any canister acquisitions during that period color coded by port. Note that the y axis value of the canister trigger is not used – there are separated by distance in the event of multiple triggers in one 5-minute period (shown below). Note that no triggers will be present if the sensors are not configured to collect canister samples or if no triggers were recorded during this time frame.

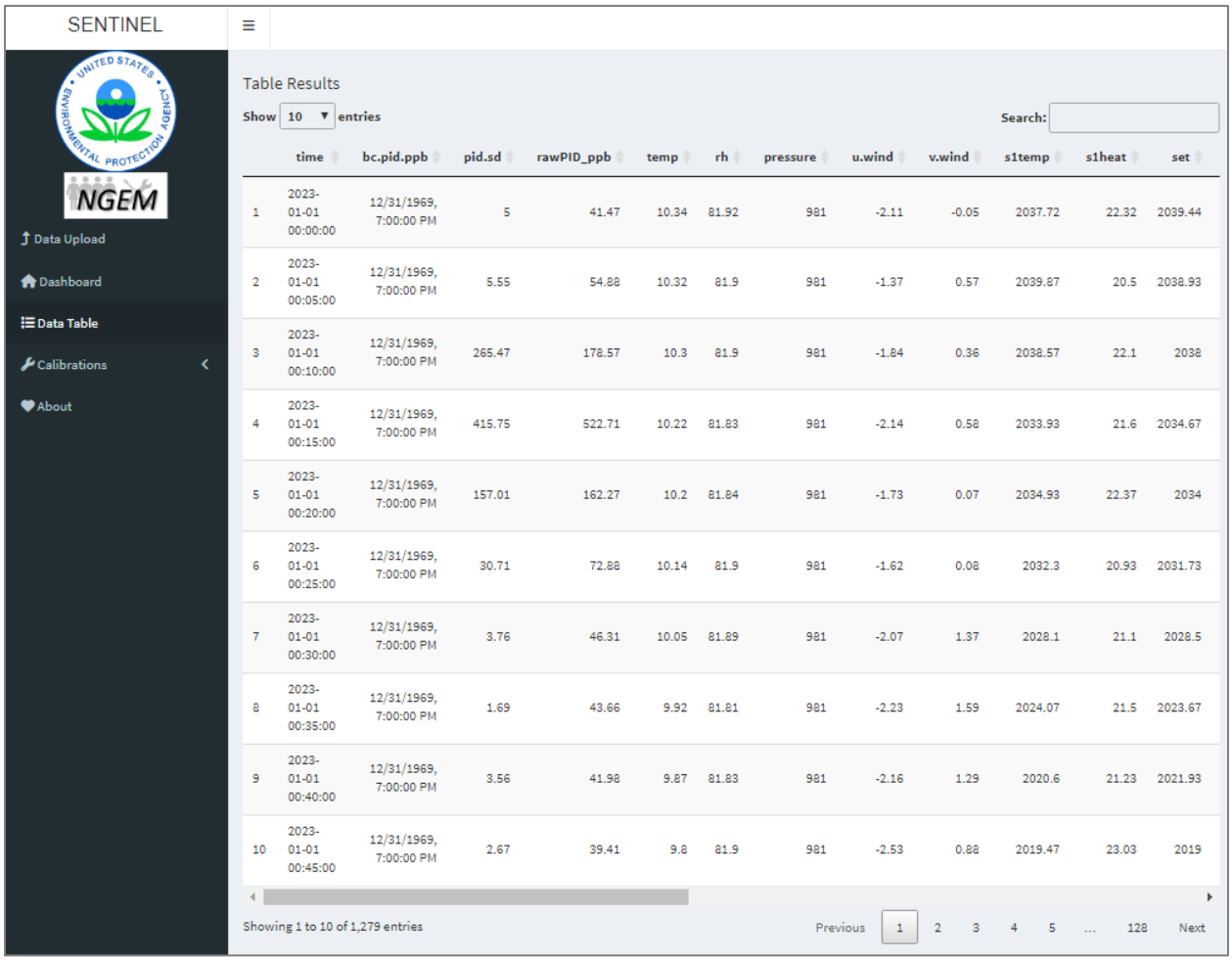




EXPORT TO PDF

The top of the dashboard provides an option for a user to generate a report of the data displayed on the dashboard. This report will be output as a pdf and will contain the SDI plot, the Wind Rose, and the three time series graphs (no zoom options can be saved in that export). The signal map cannot be exported since it is not a static image. The report can be saved to the user's device, edited in adobe acrobat, and printed as a document.

Data Table



	time	bc.pid.ppb	pid.sd	rawPID_ppb	temp	rh	pressure	u.wind	v.wind	s1temp	s1heat	set
1	2023-01-01 00:00:00	12/31/1969, 7:00:00 PM	5	41.47	10.34	81.92	981	-2.11	-0.05	2037.72	22.32	2039.44
2	2023-01-01 00:05:00	12/31/1969, 7:00:00 PM	5.55	54.88	10.32	81.9	981	-1.37	0.57	2039.87	20.5	2038.93
3	2023-01-01 00:10:00	12/31/1969, 7:00:00 PM	265.47	178.57	10.3	81.9	981	-1.84	0.36	2038.57	22.1	2038
4	2023-01-01 00:15:00	12/31/1969, 7:00:00 PM	415.75	522.71	10.22	81.83	981	-2.14	0.58	2033.93	21.6	2034.67
5	2023-01-01 00:20:00	12/31/1969, 7:00:00 PM	157.01	162.27	10.2	81.84	981	-1.73	0.07	2034.93	22.37	2034
6	2023-01-01 00:25:00	12/31/1969, 7:00:00 PM	30.71	72.88	10.14	81.9	981	-1.62	0.08	2032.3	20.93	2031.73
7	2023-01-01 00:30:00	12/31/1969, 7:00:00 PM	3.76	46.31	10.05	81.89	981	-2.07	1.37	2028.1	21.1	2028.5
8	2023-01-01 00:35:00	12/31/1969, 7:00:00 PM	1.69	43.66	9.92	81.81	981	-2.23	1.59	2024.07	21.5	2023.67
9	2023-01-01 00:40:00	12/31/1969, 7:00:00 PM	3.56	41.98	9.87	81.83	981	-2.16	1.29	2020.6	21.23	2021.93
10	2023-01-01 00:45:00	12/31/1969, 7:00:00 PM	2.67	39.41	9.8	81.9	981	-2.53	0.88	2019.47	23.03	2019

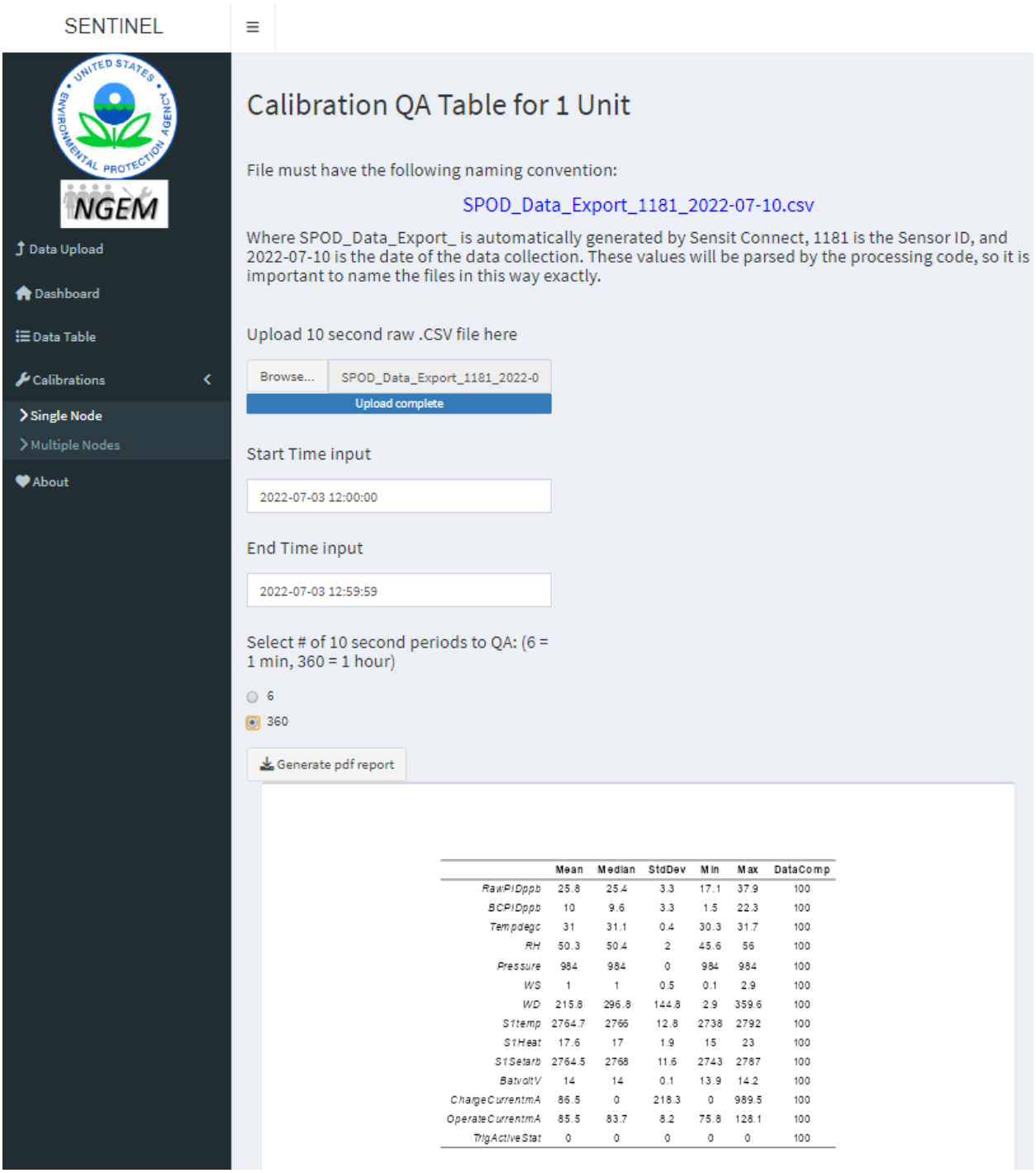
The Data Table page allows the user to see the uploaded data in a 5-minute aggregated tabular form. Selecting the “Download .CSV data” button will export a .csv file of the compiled data and calculations. More entries can be displayed if the user changes the drop-down menu in the top left-hand corner. In the top right, there is a search bar where the user can search for values. The data is automatically organized by date, but the user can sort by other columns using the arrows at the top of each column. More rows can be seen by using the previous and next buttons in the bottom left. Scrolling to the right will show more columns. Definitions for these columns can be found on page 4 of this guide. Note that values in the canister trigger columns are encoded. Explanations of these values are shown in the table below:

Port Status (canister column 1)	
15	Port 1-4 installed; none collected
31	Port 1-4 installed; port 1 collected
63	Port 1-4 installed; port 1-2 collected
127	Port 1-4 installed; port 1-3 collected
255	Port 1-4 installed; port 1-4 collected
Trigger Status (canister column 2)	
1	Trigger occurring in this time frame
Active Port (canister column 3)	
1	Port 1 collecting
2	Port 2 collecting
4	Port 3 collecting
8	Port 4 collecting
Event Status (canister column 4)	
1	Port 1 started
17	Port 1 started; Port 1 complete
19	Port 2 started; Port 1 complete
51	Port 2 complete; Port 2 complete; Port 1 complete
55	Port 3 started; Port 2 complete; Port 1 complete
119	Port 3 complete; Port 2 complete; Port 1 complete
127	Port 4 started; Port 3 complete; Port 2 complete; Port 1 complete
255	Port 4 complete; Port 3 complete; Port 2 complete; Port 1 complete

Calibrations

These options (single node and multi node) allow the user to create a QA table based on a time frame for a single sensor or two collocated sensors. These are useful for ensuring the sensor is running nominally as well as determining how similar two sensors located side-by side are reporting. These tables are required to be collected during calibrations or cal-checks as defined in the SOP.

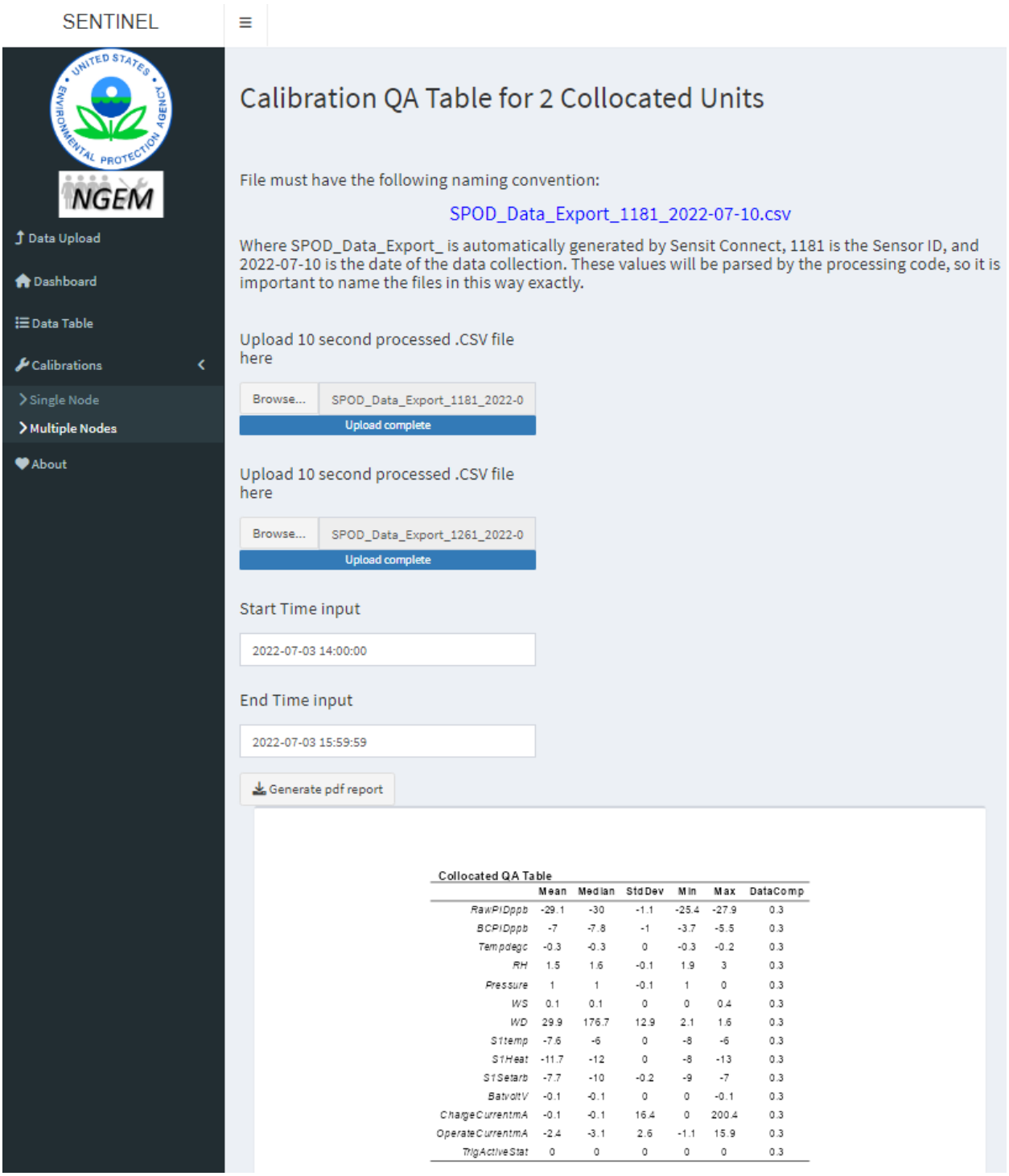
SINGLE NODE



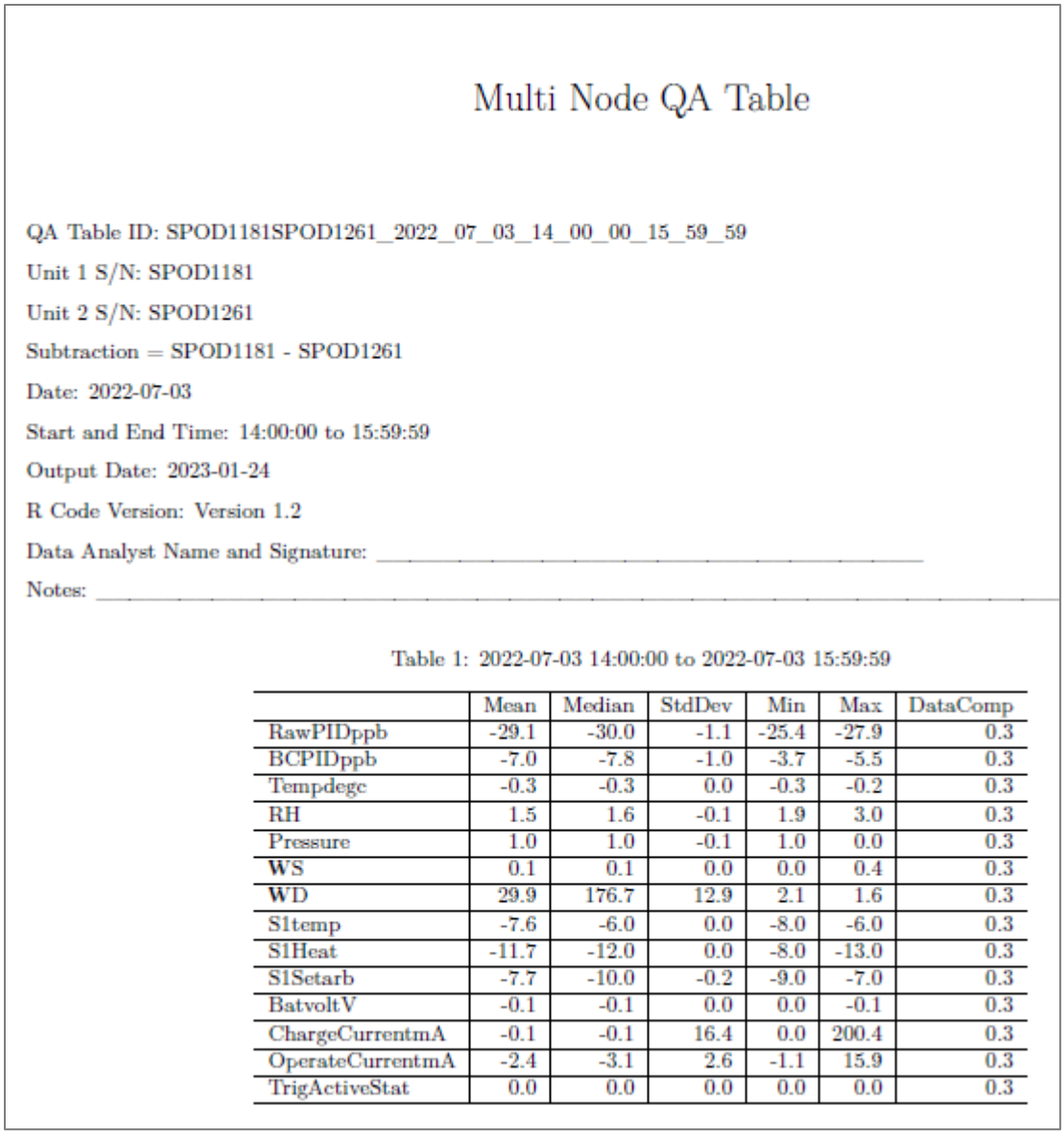
	Mean	Median	StdDev	Min	Max	DataComp
RawPIDppb	25.8	25.4	3.3	17.1	37.9	100
BCPIDppb	10	9.6	3.3	1.5	22.3	100
Tempdegc	31	31.1	0.4	30.3	31.7	100
RH	50.3	50.4	2	45.6	56	100
Pressure	984	984	0	984	984	100
WS	1	1	0.5	0.1	2.9	100
WD	215.8	296.8	144.8	2.9	359.6	100
S1temp	2764.7	2766	12.8	2738	2792	100
S1Heat	17.6	17	1.9	15	23	100
S1Setarb	2764.5	2768	11.6	2743	2787	100
BatvoltV	14	14	0.1	13.9	14.2	100
ChargeCurrentmA	86.5	0	218.3	0	989.5	100
OperateCurrentmA	85.5	83.7	8.2	75.8	128.1	100
TrigActiveStat	0	0	0	0	0	100

The Single node calibration requires the user to enter their Sensit connect raw file (named in the same way as general files are named) and the start and end time that they would like to QA. For a calibration on 10 second data, this will likely be only 1 minute. For other QA purposes, this could be 1 hour. The user should select either 6 seconds (1 minute) or 360 seconds (1 hour), so the table shows the correct data completeness value. The table that is generated will show summary stats for data categories which can used to judge if the sensor is calibrating or performing as expected during that time frame. The user can then select the “Generate PDF report” button to get a pdf output with the QA table. This is an excellent option for record keeping.

MULTI NODE



The multi node tab operates similarly to the single node tab expect the user is prompted to enter a secondary node to be compared to the first node. For this comparison to be effective, the two selected nodes should be collocated, so sensor agreement can be evaluated. The user once again enters the start and end time, and a table will appear. This table is the simple difference between the two nodes (the first node – the second node). The user has the option to export this QA table to a pdf for record keeping. These outputs are built in R Markdown (example shown below).



Multi Node QA Table

QA Table ID: SPOD1181SPOD1261_2022_07_03_14_00_00_15_59_59

Unit 1 S/N: SPOD1181

Unit 2 S/N: SPOD1261

Subtraction = SPOD1181 - SPOD1261

Date: 2022-07-03

Start and End Time: 14:00:00 to 15:59:59

Output Date: 2023-01-24

R Code Version: Version 1.2

Data Analyst Name and Signature: _____

Notes: _____

Table 1: 2022-07-03 14:00:00 to 2022-07-03 15:59:59

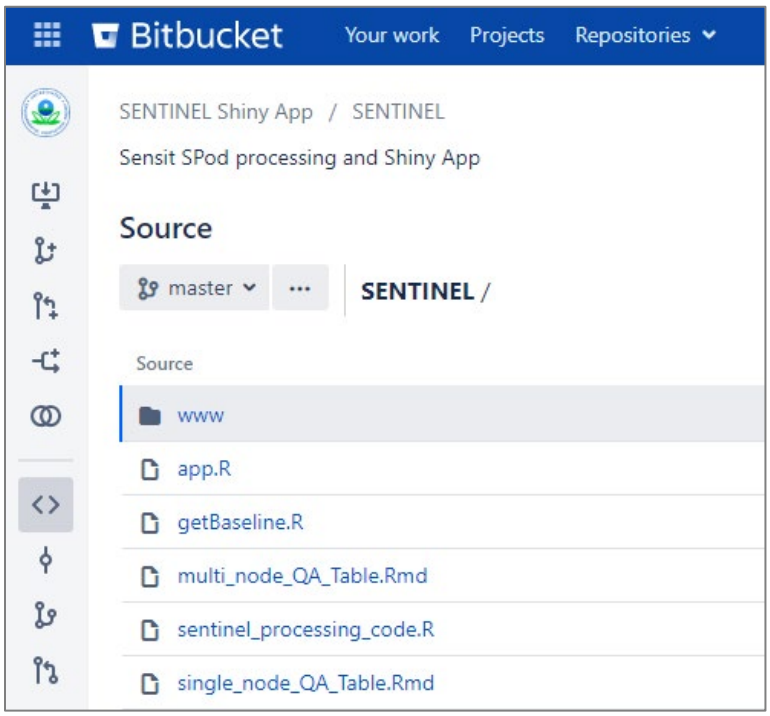
	Mean	Median	StdDev	Min	Max	DataComp
RawPIDppb	-29.1	-30.0	-1.1	-25.4	-27.9	0.3
BCPIDppb	-7.0	-7.8	-1.0	-3.7	-5.5	0.3
Tempdegc	-0.3	-0.3	0.0	-0.3	-0.2	0.3
RH	1.5	1.6	-0.1	1.9	3.0	0.3
Pressure	1.0	1.0	-0.1	1.0	0.0	0.3
WS	0.1	0.1	0.0	0.0	0.4	0.3
WD	29.9	176.7	12.9	2.1	1.6	0.3
S1temp	-7.6	-6.0	0.0	-8.0	-6.0	0.3
S1Heat	-11.7	-12.0	0.0	-8.0	-13.0	0.3
S1Setarb	-7.7	-10.0	-0.2	-9.0	-7.0	0.3
BatvoltV	-0.1	-0.1	0.0	0.0	-0.1	0.3
ChargeCurrentmA	-0.1	-0.1	16.4	0.0	200.4	0.3
OperateCurrentmA	-2.4	-3.1	2.6	-1.1	15.9	0.3
TrigActiveStat	0.0	0.0	0.0	0.0	0.0	0.3

About Page

The about page contains the version number of the code, and the contact information if any user finds questions or bugs (macodnald.megan@epa.gov). There are also some links on the about page that might be helpful for new users. These include some information about SDI plots, the Sensit user manual, and some EPA presentations/articles about fenceline sensors. The associated SOP and User manual are also included on this page, as well as acknowledgements to contributors to this project.

Accessing Code

For users who are curious about the actual Shiny R Code used to build the app, the code folder is available on the SENTINEL EPA Bitbucket Repository. This requires a user to be given access to see the code repository; please email macdonald.megan@epa.gov for access requests). Once the user has access, they can see the files in the code folder on Bitbucket. This includes the www folder (where images on the app are stored), the getBaseline function, the R Markdown template documents for pdf outputs, and the "app.R" code. Bitbucket is updated with new code commits and version control records as the app is further developed.



References

1. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
2. Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert and Barbara Borges (2022). shiny: Web Application Framework for R. R package version 1.7.2. https://CRAN.R-project.org/package=shiny
3. Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.9. https://CRAN.R-project.org/package=dplyr
4. Winston Chang and Barbara Borges Ribeiro (2021). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.2. https://CRAN.R-project.org/package=shinydashboard
5. Carslaw, D. C. and K. Ropkins (2012). openair --- an R package for air quality data analysis. Environmental Modelling & Software. Volume 27-28, 52-61.
6. Carson Sievert (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida.

Resources

More information on R Shiny Applications:

- Mastering Shiny Bookdown Site

More information on SPod deployments:

- Rubbertown Next Generation Emissions Measurement Demonstration Project (Journal article)
- Demonstration of VOC Fenceline Sensors and Canister Grab Sampling near Chemical Facilities in Louisville, Kentucky (Journal article)
- Sensor Pod (SPod): An Approach for VOC Fenceline Monitoring and Data Analysis (EPA Tools and Resources Webinar)
- Next Generation Emission Measurements (NGEM) Advancements (EPA ORISE Meets the World Seminar)
- Fenceline and Community Sensor Applications and Comparisons (Air Sensors International Conference video)