



[www.agilairecorp.com](http://www.agilairecorp.com)

# Agency Air Quality Data Management Systems And Real-Time Small Sensor Networks

Steve Drevik  
Agilaire

[info@agilaire.com](mailto:info@agilaire.com)

# Small Sensors - Real-Time Challenges

Small air sensors (SAS) have been with us for several years now.

For most of the projects, data corrections from correlations / calibrations and QC validations have been applied 'after the fact' when the entire data set is known.

In the future, we'd like to be able to do this in real time (e.g., hourly, 10-minute, etc) to make the data truly useful.

Reference Monitor



15.1 ug/m<sup>3</sup>

Small Sensor Network



12.1 ug/m<sup>3</sup>



16.9 ug/m<sup>3</sup>



24.5 ug/m<sup>3</sup>

*Small sensor data needs correction and validation to be useful in real-time*

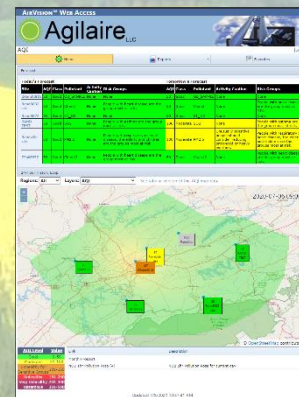
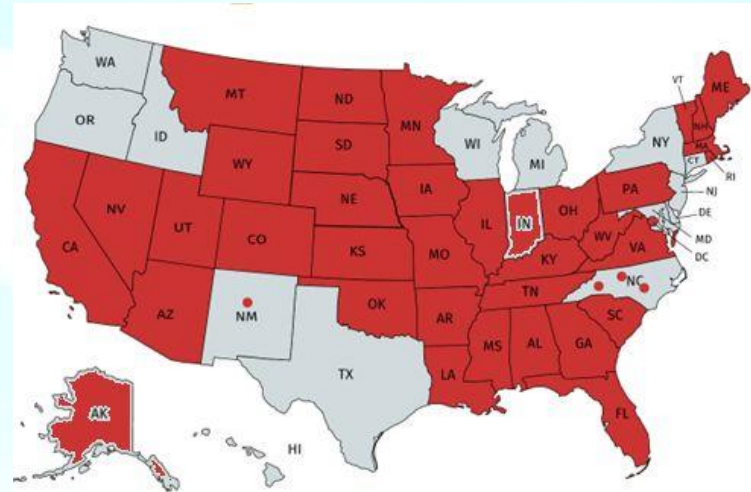
# Background: Agilaire

Leading provider for data management systems for state, local, and tribal (SLT) agencies, covering ~ 75% of US.

Deep expertise in dealing with data interfaces, real-time data collection, automated, and manual validation / QC tools.

Deployed the first interfaces for automatic, continuous collection and correction of small sensor (Purple Air) data in real-time.

Focused on continuously improvement, including building library of off the shelf drivers for automated collection of small sensors significant to the SLT user base and industrial sources.



# Small Air Sensors

Primary challenges from a data handling perspective, especially trying to build a real-time network, have been:

1. Acquiring the data (protocols, gateways, etc)
2. Quality assuring the data (is it good or bad?)
3. Correcting the data (more 'NAAQS\*-like')
4. Visualization of the Data

Most critical? Step 2 and 3. Garbage in = Garbage out.

*\* National Ambient Air Quality Standards, e.g., regulatory quality monitor*

# Acquiring The Data

Most SASs push data to a web gateway, and most now offer good API interfaces to collect the data.

- APIs should be well documented, and vendors should (and have) worked with data acquisition system providers to provide test environments.
- JSON data payloads vary widely in structure, time formats, etc.
- Lessons learned? The data system must have a *very flexible system* for parsing the JSON payloads. There is a wide variation in the way data is portrayed, parameters are identified in the list, etc.

```
{ "status": "OK", "desc": "Operation success", "values":  
  { "record": { "measureid": 1794, "measuredate": "2021-04-19  
00:09:01", "chromfile": "pyxis_method_Chrom_2021_04_19_  
_00_09_01", "gasname": "Benzene", "idgas": "71-43-  
2", "conc": 0.163354999999983 }, { "measureid": 1794, "measur  
edate": "2021-04-19
```

Column Number	Data Field	Parse Format	Flag Map
4	Date/Time	yyyy-MM-dd HH:mm:ss	
6	ParmIdentifier from meta tag		
8	Value		

# Validating the Data

The data in the sensors often has the information needed to self-validate, but most gateways don't offer this.

- DAS vendors- again, a *very flexible system* that allows users to create multi-condition rules that look at all available data can help validate data in real-time.
- Sensor vendors should take on some of this work(e.g. compare A/B sensors, look at RH levels, etc) and consider including quality codes with their data stream.

But right now, this work falls on the agency / end user data management systems.



# Automatic Data Validation Processor (ADVP)

## Example design:

- User can create any number of rules
- Rules can have multiple conditions (e.g., A and (B or C)).
- Each condition can look at properties of various measurements (RH>x, PM increasing or decreasing, difference between two sensors > 10%, day or nighttime conditions, etc).
- Offers flexibility in what to do if rule met (adjust, mark suspect, invalidate, add text notation, send email)

The screenshot displays the ADVP Editor software interface. The main window is titled 'ADVP Editor' and contains several panes:

- Rule List:** A list of rules including '633Stat', 'AUDIT annotation', 'Bad PMA', 'BAM > 300', 'BAM 985', 'BAM\_GT\_984', 'BenzeneToluene', 'BenzeneToluene\_xflag', 'Benzene\_NotZero', 'Burning Xmas Trees', 'Carbon Tetrachloride', 'Compare 2 Sites', 'CrossSiteExample', 'CrossSiteTest', 'CrossSiteTest (2)', 'DiffTest', 'EthyleneEthane', 'EthyleneEthane\_xflag', 'Ethylbenzene\_NotZero', 'Formaldehyde\_Low', 'GRIC Test', 'HighOzoneTest', 'Historical', 'Isoprene\_NotZero', 'Lookback', 'N-Butane\_NotZero', 'N-Hexane\_NotZero', 'Night Isoprene' (selected), 'O3RollupCompare', 'OzoneOut', 'PAMS > TNMOC', 'PM\_Wind\_KNOX', 'Propane\_NotZero', 'PropylenePropane', 'PropylenePropane\_xflag', 'Shelter Temp OZONE', 'Shelter Temp SO2', 'UndecaneDecane', and 'UndecaneDecane\_xflag'.
- Rule Information:** Shows details for the selected rule 'Night Isoprene', including 'Rule Name', 'Description', 'Average Interval' (001h), and 'Max Lookback Intervals'.
- Actions:** Configuration for actions triggered by the rule, such as 'Site', 'Parameter Template' (Isoprene\_HR), 'Apply Null Code' (DA - Aberrant Data), 'Assign Value', 'Assign Data Grade', 'Apply Flag' (I - Invalidated By Edit), and 'Clear Flag'.
- Conditions Triggering Rule:** A table showing the logical structure of conditions:

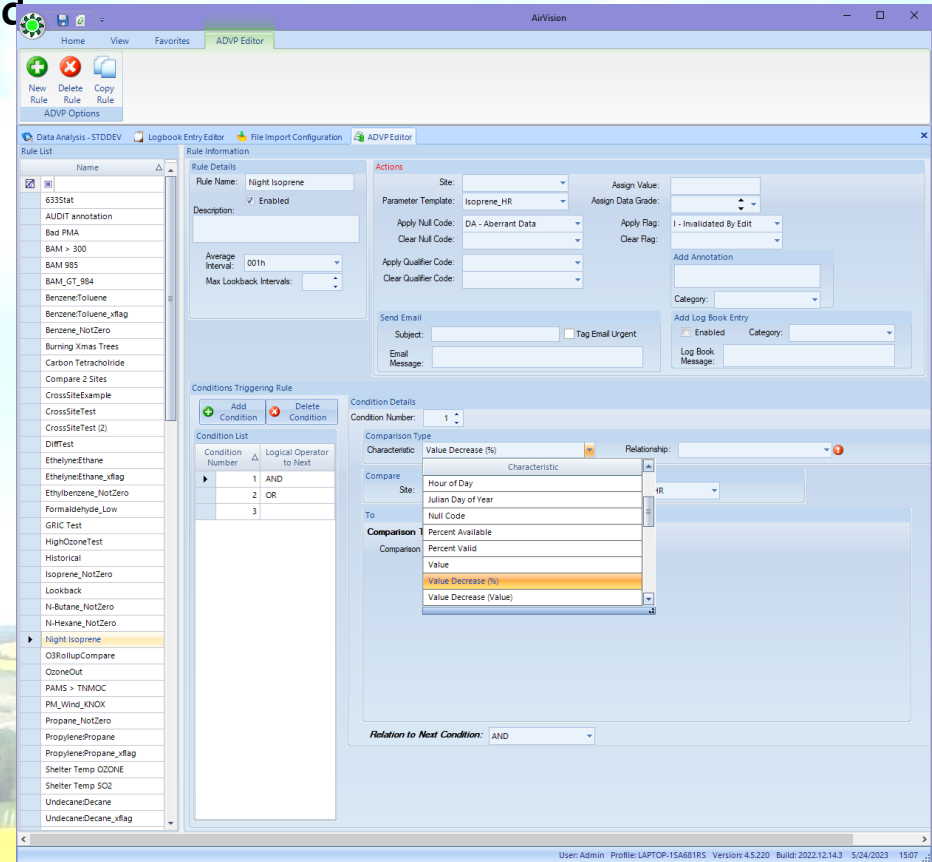
Condition Number	Logical Operator	to Next
1	AND	
2	OR	
3		
- Condition Details:** Configuration for a specific condition (Condition Number: 1), including 'Comparison Type' (Value Decrease (%)), 'Characteristic' (Hour of Day), 'Relationship' (IR), and 'Comparison' (Percent Available, Percent Valid, Value, Value Decrease (%), Value Decrease (Value)).

The status bar at the bottom indicates the user is 'Admin', the profile is 'LAPTOP-15A681RS', the version is '4.5.220', the build is '2022.12.14.3', and the date is '5/24/2023 15:07'.

# Automatic Data Validation Processor (ADVP)

Example rules of such a design (rules derived from manufacturer or from user experience/testing):

- RH > 90%
- PM2.5 > PM10
- Value varies > 30% from last hour
- Value differs > 30% from a nearby NAAQS sensor site
- Value differs > 30% from nearby small sensor site (“buddy site”)
- A / B sensors differ > 10% from each other
- Validation vs. internal diagnostic parameters, if available.



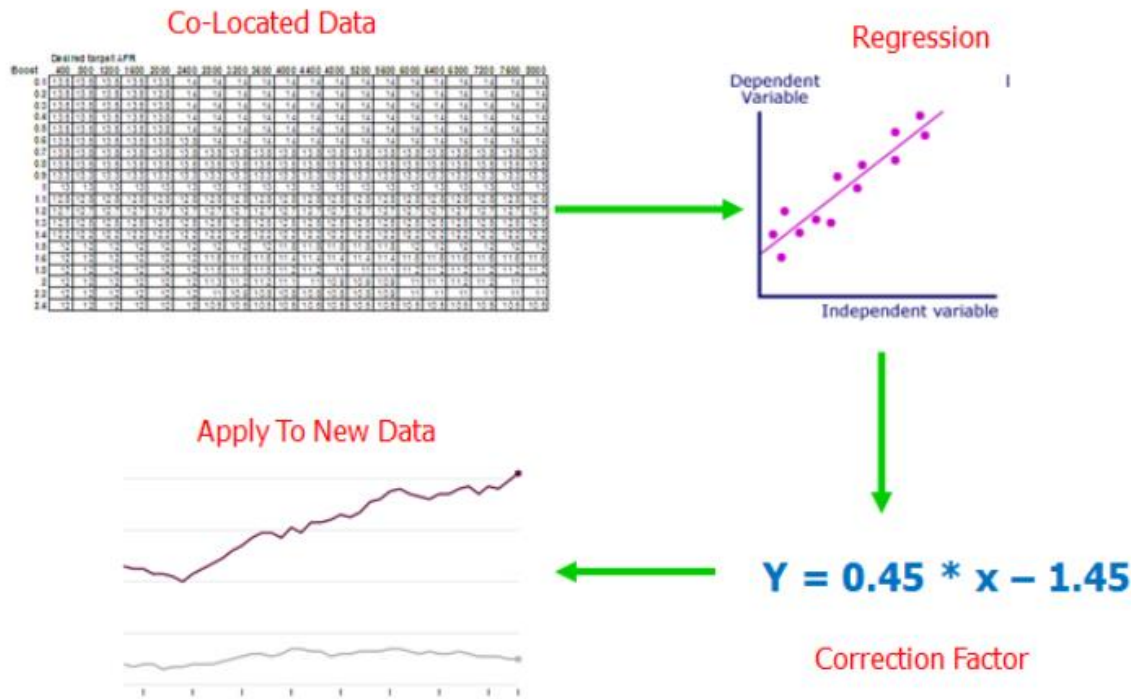
*As you learn more about a particular sensor's quirks, can add additional qualifiers like daytime / nighttime, etc.*



# Data Correction

So far, a lot of this work has been “one-shot” correction of data after the fact, with time to look at various regression and machine learning methods to make the data look great.

Implementing a real-time network is a different animal. You have to pick an algorithm, initial correction factors, and decide if you’re going to update them based on new data, age of instrument, etc

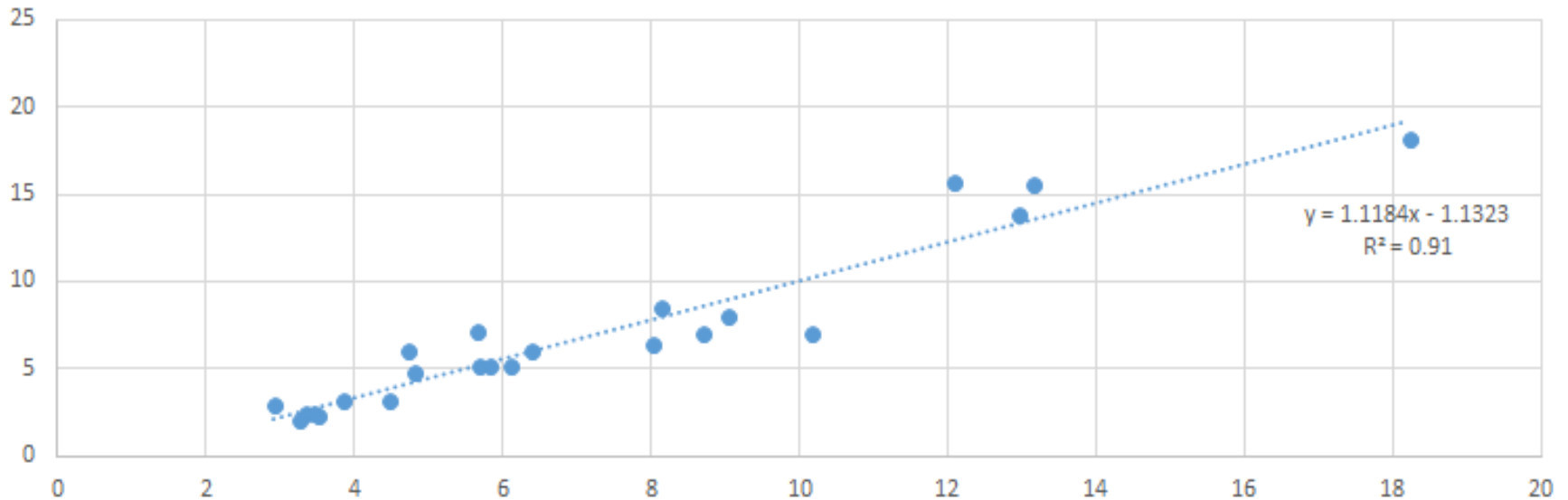


# Data Correction

Again, flexibility is key here.

- EPA started with a simple, correction for PA2 (mX+b, RH factor)
- Now we have a 5-part 'piecewise' algorithm based on PM levels. Unclear why Purple Air hasn't implemented in their web portal / API?
- Correlation curves could be specific to each instrument

Corrected Purple Air vs BAM: 9/28/17-12/2/17



# Data Correction: AirVision

The original simple correction for Purple Air 2 has been supported by AirVision's Central Math functions since the beginning.

Same system can be used to support other instruments, if not available in the instrument or manufacturer's gateway.

AirVision will have a new Central Math function to easily implement the new 5-part piecewise algorithms for Purple Air 2 in an updated release (July 2023) as a function "EPAPA2021", with arguments for PM (cfatm) and relative humidity.

Raw values and final values can be stored, along with meta data.

The screenshot displays the AirVision software interface. The top window is the "Math Equation Editor" with a table of equations:

Equation Name	Prevent Recalculation	Equation Description	Math Equation
PA	<input type="checkbox"/>		
PA2_1101	<input type="checkbox"/>	PM correction for purple air 2 (old)	$0.534 * PA\_cf1 - 0.0844 * RELHUM + 5.604$
PA2 Correct 5 piece	<input type="checkbox"/>	New EPA 5 piece algorithm	$EPAPA2021(PM\_cfatm, RELHUM)$

The bottom window shows the "System" tree on the left and the "Parameter Template" configuration for "PM25\_Corr" on the right. The configuration includes:

- Site: ANT\_QLD
- Parameter: PM25\_Corr
- Parent Parameter: [dropdown]
- Parameter Group: [dropdown]
- Website Display Name: [text field]
- Enabled:  Enable AIRNow Reporting:  Filter From Web Site:
- Parameter Data Type:  Average / Continuous,  Continuous Sample,  Sample / Non-Continuous
- Description: Purple Air Corrected
- Math Equation (if Calculated): PA2 Correct 5 piece
- EPA POC: 1
- EPA Method: 111
- EPA Units: 105 - Micrograms/cubic meter (LC)
- EPA Parameter: [dropdown]
- Reported Digits: 4
- Precision: 1
- Calibration Precision: [dropdown]
- Truncate Round Rule:  Round,  Truncate
- Reported Units: UG/M3
- Analyzer Units (if different): [dropdown]
- Graph Minimum: 0.00
- Graph Maximum: 1000.0
- Calibration Span: [text field]
- Instrument Detection Limit: [text field]
- Limit Of Quantization: [text field]
- Minimum Detectable Limit: [text field]
- Practical Quantitation Limit: [text field]
- Parameter Report Order: [dropdown]
- Totalize in Reports  Minimum in Reports

# End Benefit- Higher Quality Real-Time Data

We have networks of high accuracy regulatory monitors, but just not enough to cover the spatial gaps between them as much as we would like.

So, being able to validate and adjust such data in real-time provides benefits to agencies, in particular during wildfire or emergency air quality events.

*(But let's shy away from the term "AQI" if we can...)*

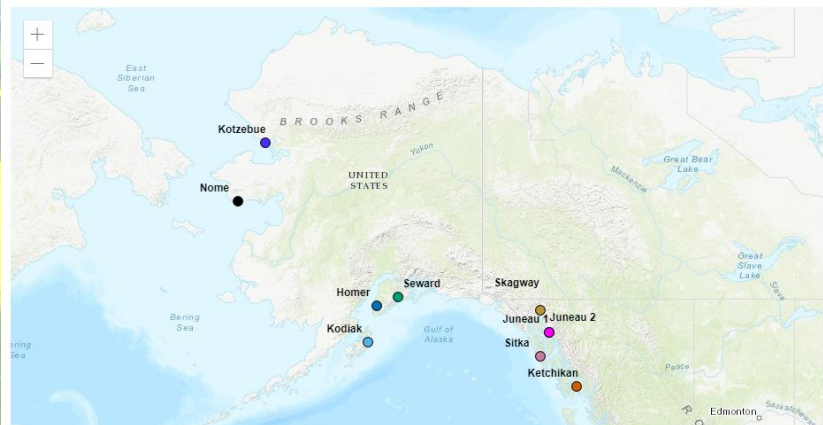
## COMMUNITY-BASED AIR MONITORING PILOT PROJECT

The site is undergoing maintenance. The map feature is currently under repair, but please scroll down to the Data Visualization section to see real-time data plots. Be sure to click on the parameter of interest to see the data display (ex. PM2.5).

DEC is in the process of deploying AQMesh sensor pods to hub communities throughout Alaska as part of a Community-Based Air Monitoring Pilot Project. This webpage will display data from the pods as they are deployed and begin reporting data. Data from the pods are not available until approximately 15 minutes after the top of the hour.

### AQMesh Deployment Map

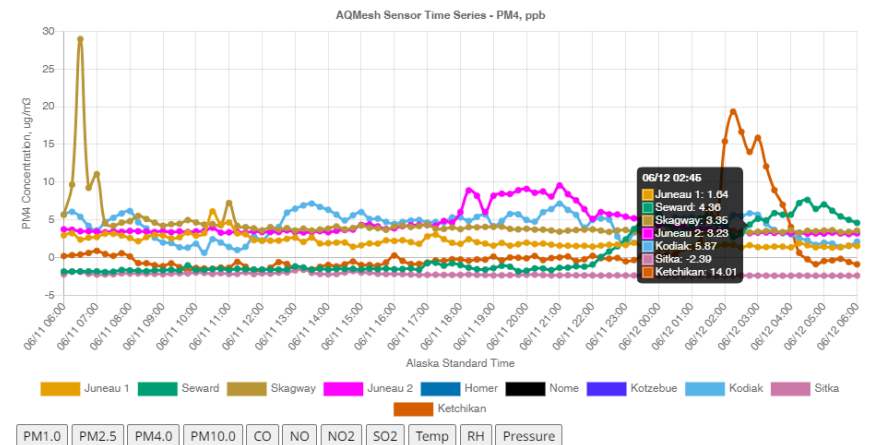
The map displays the locations of each AQMesh pod.



Alaska DEC AQMesh project w/AirVision

## Data Visualization

Data from all of the operational sensor units are shown as 15-minute average values for the preceding 24-hour period. Please note the time series is presented in Alaska Standard Time. Click buttons below the plot to view different parameters. To turn stations on or off, click on the site name in the legend.



[Public Notices](#) • [Regulations](#) • [Statutes](#)  
[Press Releases](#) • [Contact](#) • [Accessibility](#)

Department of Environmental Conservation

Mailing Address: P.O. Box 111800

Juneau, Alaska 99811

Location: 410 Willoughby Avenue, Juneau



# Summary of QA procedures

QA type	✓, X, N/A	Description (required)
Check of individual components	X	
Other hardware QA	X	
Sensor software QA	X	
Lab testing of sensors	X	
Field testing of sensors	X	
<b>Mathematical QA of individual data points</b>	✓	Via Automatic Data Validation Processor (ADVP)
<b>Mathematical QA of individual sensors</b>	✓	ADVP: based on available internal diagnostic / values
<b>Mathematical network-wide QA</b>	✓	Via Automatic Data Validation Processor (ADVP)
<b>Data fusion QA</b>	✓	ADVP: “buddy site” comparisons
<b>Mathematical QA using information other than pollutant of interest</b>	✓	Via Automatic Data Validation Processor (ADVP)
<b>Aggregator acceptance criteria</b>	✓	ADVP can use any measurement in the network for validation rules for any other parameter, with multi-condition combinational logic (if X and Y or Z).
<b>Other?</b>	✓	Data value corrections or applying calibration factors / curves



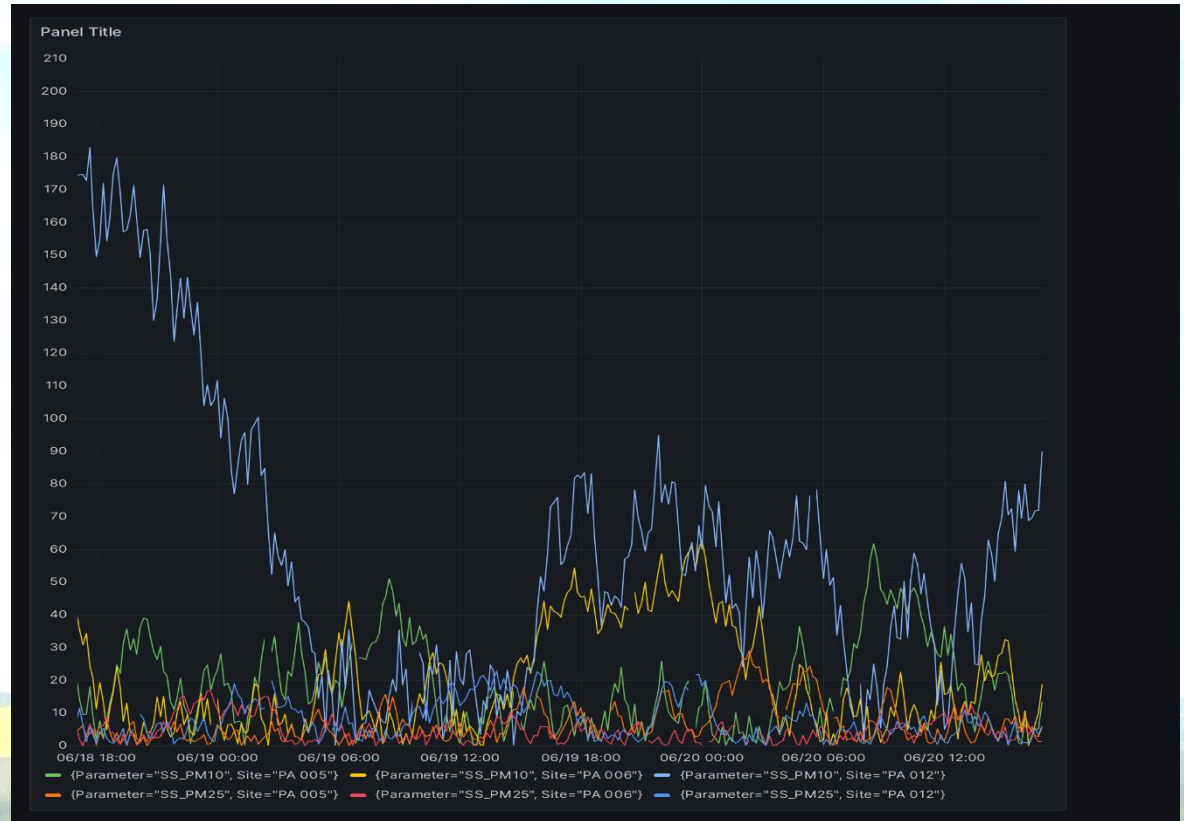
www.agilairecorp.com

# Questions?

Steve Drevik

info@agilaire.com

Agilaire



*Purple Air 2 data depicted in Grafana*