



# Data Quality Assessment Methods for Community Air Quality Monitoring Data in AQview

Cheryl Winfield / Emily Gorrie

Air Sensors International Conference  
May 11-13, 2022

# AQview – CARB's Community Air Quality Data Portal

Initially created in response to California Assembly Bill 617 (AB 617)

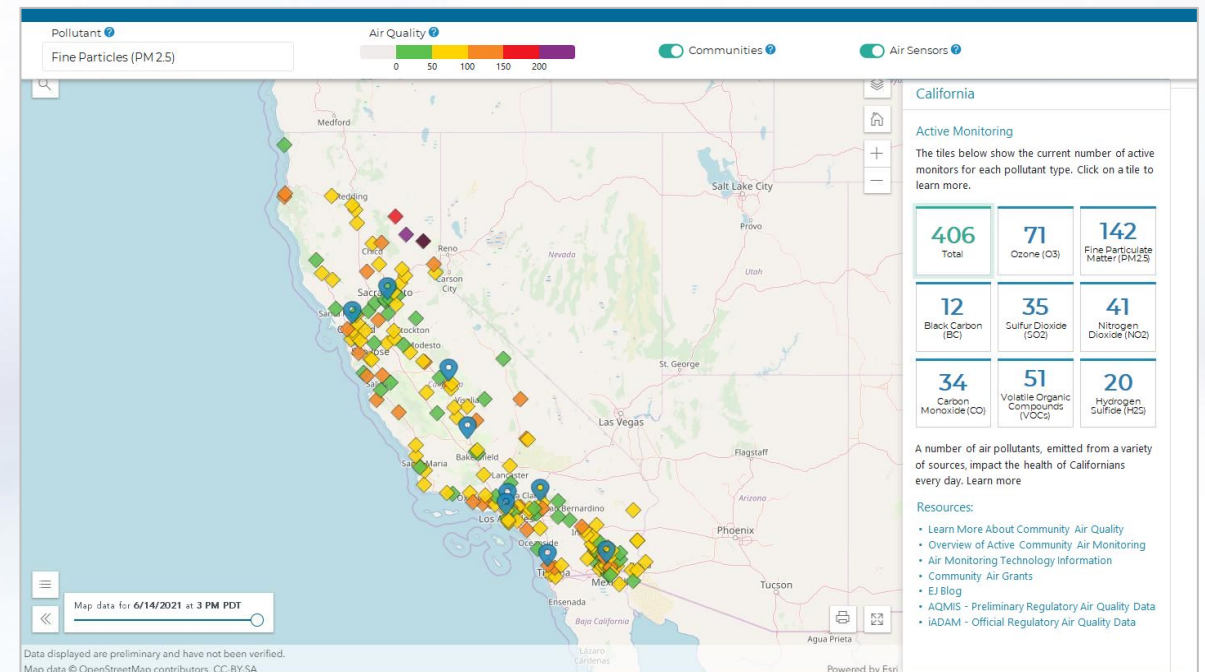
- Seeks to improve public health in communities experiencing high air pollution exposure burdens
- Community-focused and community-driven

Primary goal is to make community data easier to access and interpret through transparency of how data are collected and processed

Additional datasets will be added over time (e.g., regulatory network, research projects, special studies)

Key Features:

- Aggregates major sources of air quality data
- Provides straightforward data download
- Quality assurance and control (QA/QC) of data



# Challenges of Community Monitoring Data

AQview will house air quality data from low-cost sensors, regulatory monitors, and other monitors:

- How do we message differences in data quality?
- To what level do we assess data quality for different instrument types?
- Can we evaluate data from all instrument types on the same quality scale?
- How can we assess whether sensors are appropriately sited for ambient air quality measurements?
- How do we show all data together so that they are meaningful?



# Assessing and Messaging Data Quality

## Record-Level Quality Control/Check

### Level 1 - Preliminary QC:

- Automated check for flagging obvious erroneous records
- Provided in graphical displays and data downloads

### Level 2 - Enhanced QC:

- Automated QC checks with broader scope based on a longer time window
- Semi-automated assessments
- Provided in data downloads

## Overall Data Quality Assessment

### Level 3 - Overall Data Quality:

- Cumulative statistical analysis
- Incorporates site and monitor QA factors
- Used to assess instrument reliability, performance, and operation, as well as data provider reliability for data analyses

→ Methods must be transparent and defensible



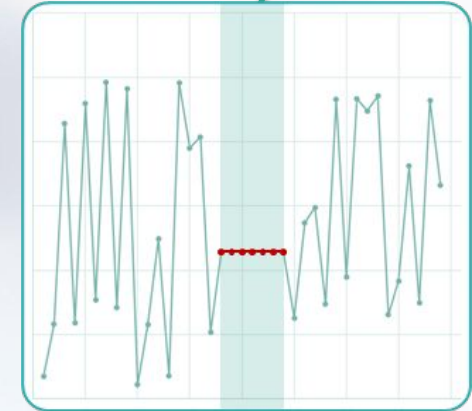
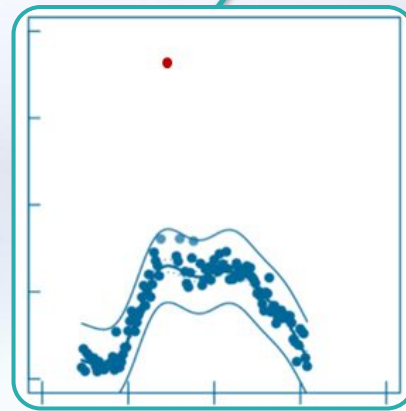
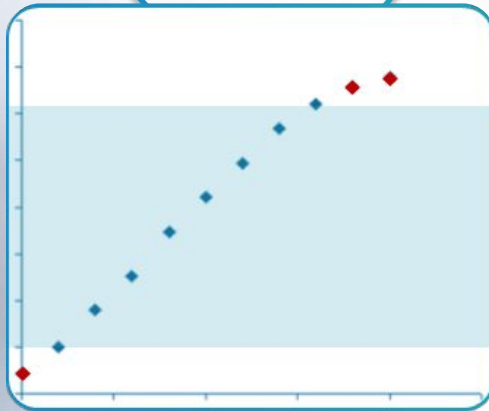
# Level 1 – Preliminary QC

Instrument-Based  
Upper Limit  
Check

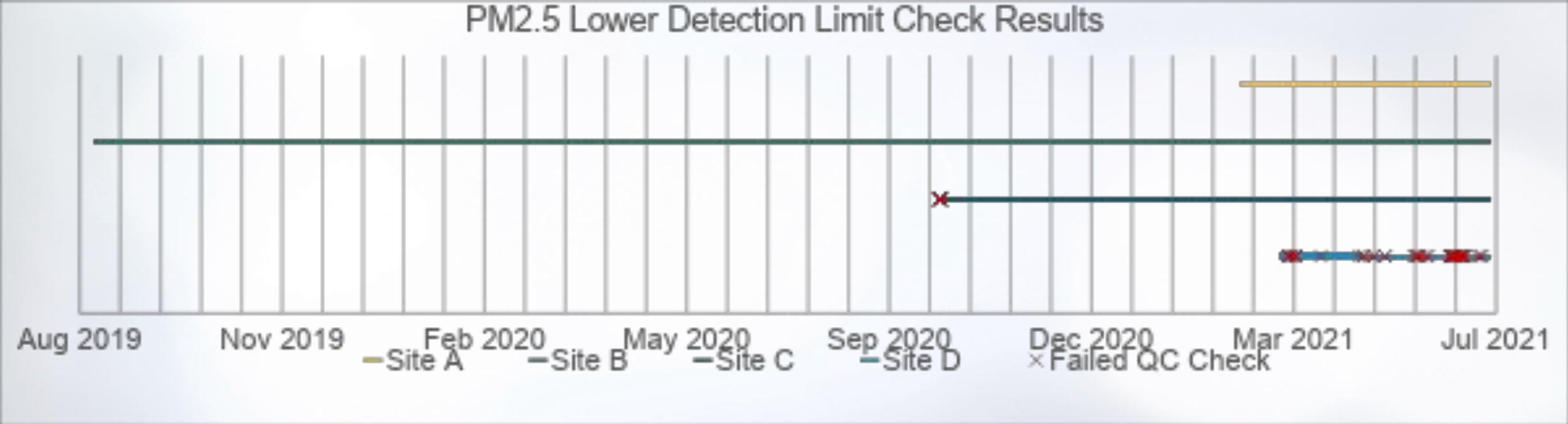
Instrument-Based  
Lower Limit  
Check

Statistical  
Outlier  
Check

Repeating  
Values  
Check

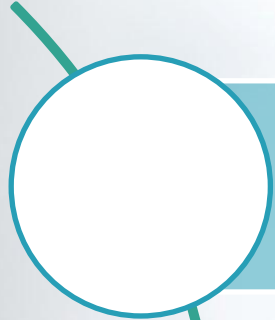


# Example: Level 1 QC - Lower Detection Limit Check

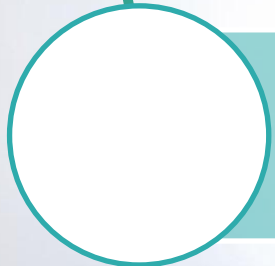


| Site ID | Total Number of Records | % Pass | % Fail |
|---------|-------------------------|--------|--------|
| Site A  | 2856                    | 100    | 0      |
| Site B  | 15822                   | 100    | 0      |
| Site C  | 6313                    | 99.8   | 0.2    |
| Site D  | 2380                    | 97.7   | 2.3    |

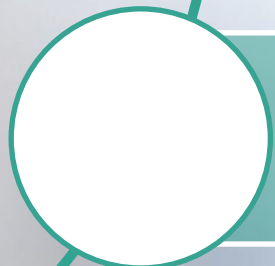
# Level 2 - Enhanced QC



Domain-based statistical anomaly identification  
(e.g., pollutant-specific or region-specific historical maxima)



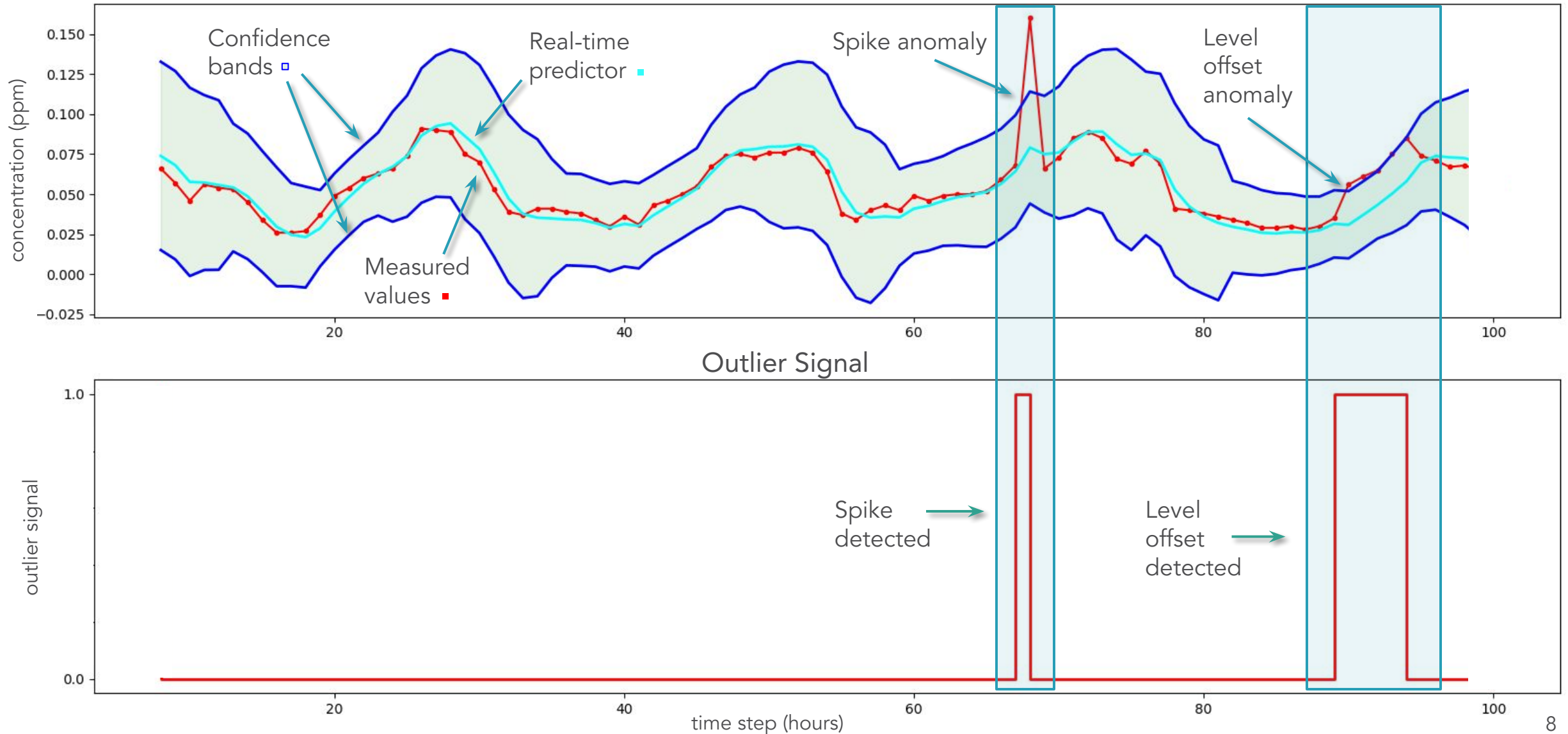
Parametric statistical anomaly identification  
(i.e., based on expected statistical distribution)



Prediction-based anomaly identification  
(e.g., machine learning)

# Example: Level 2 Real-Time Outlier Detection

Hourly Ozone Concentrations with Threshold Confidence Limits



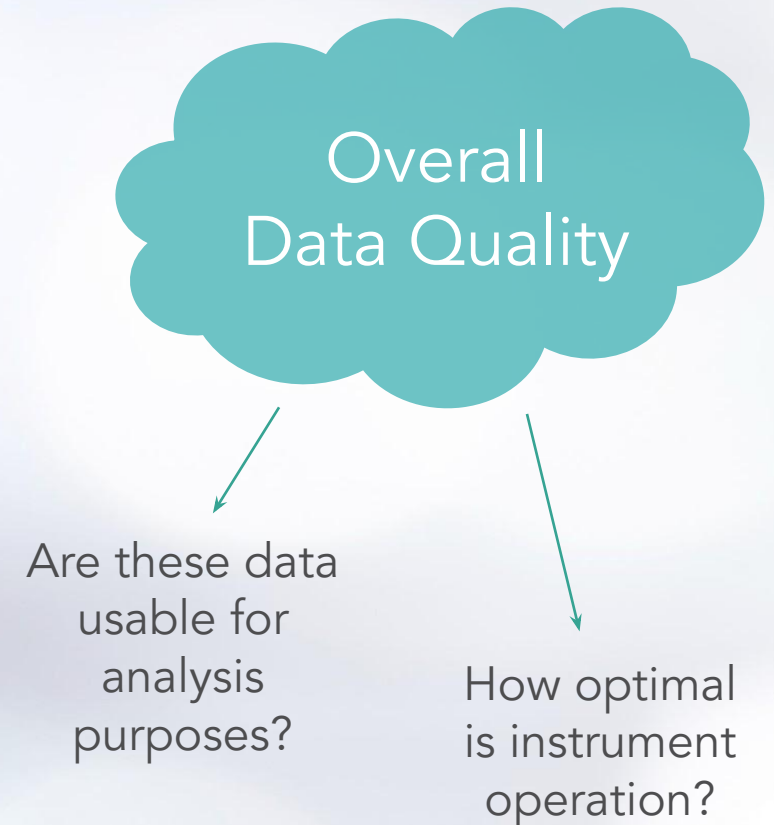


# Level 3 - Overall Data Quality

Data quality will be periodically assessed over a longer timeframe to assign an Overall Data Quality to a dataset.

Overall Data Quality will be assigned on an instrument level via scores in two fields:

- Statistical assessment of automated long-term QC processing results
- Data providers' site-level QA information



# Level 3 - Overall Data Quality Elements \*

\* Conceptual Framework

