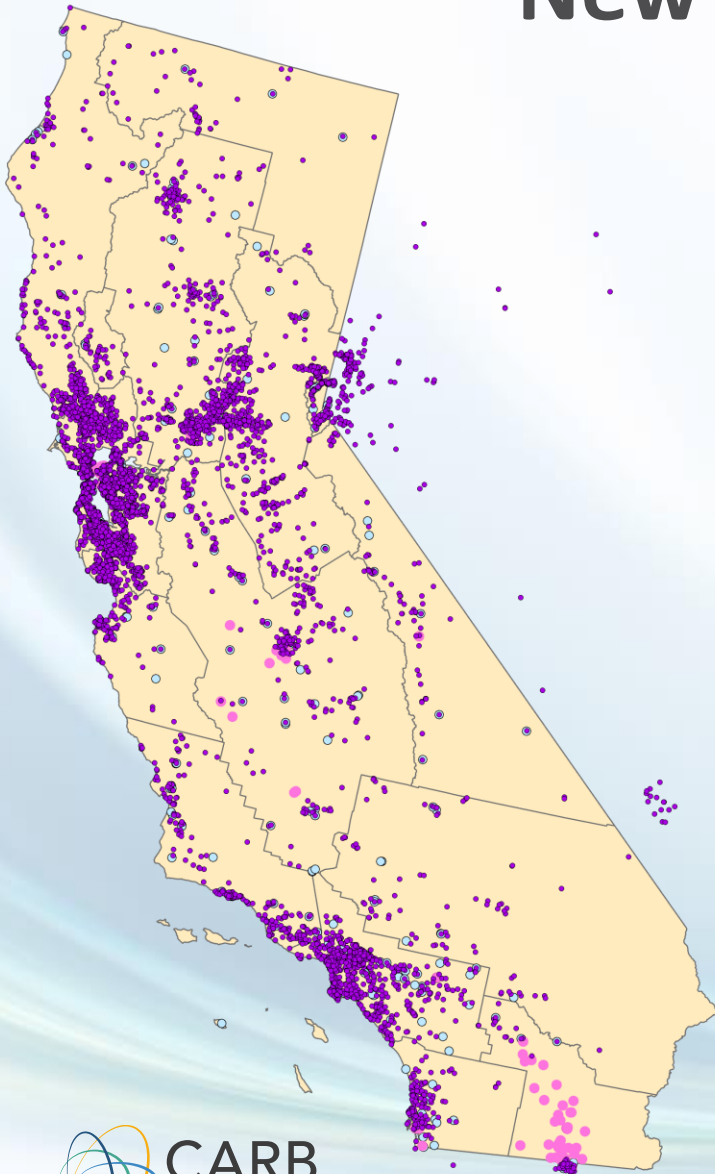# Community Air Quality Viewer (AQview)
Introduction and Data Quality Assessment

Raiford Hann, Ph. D.

Air Quality Planning and Science Division

CARB PQAO Training | February 2024

New Air Quality Data Challenges

Regulatory Monitors

Sites: ~250

Reporting Interval: Daily, Hourly

Community Monitors

Sites: 100+ and growing!

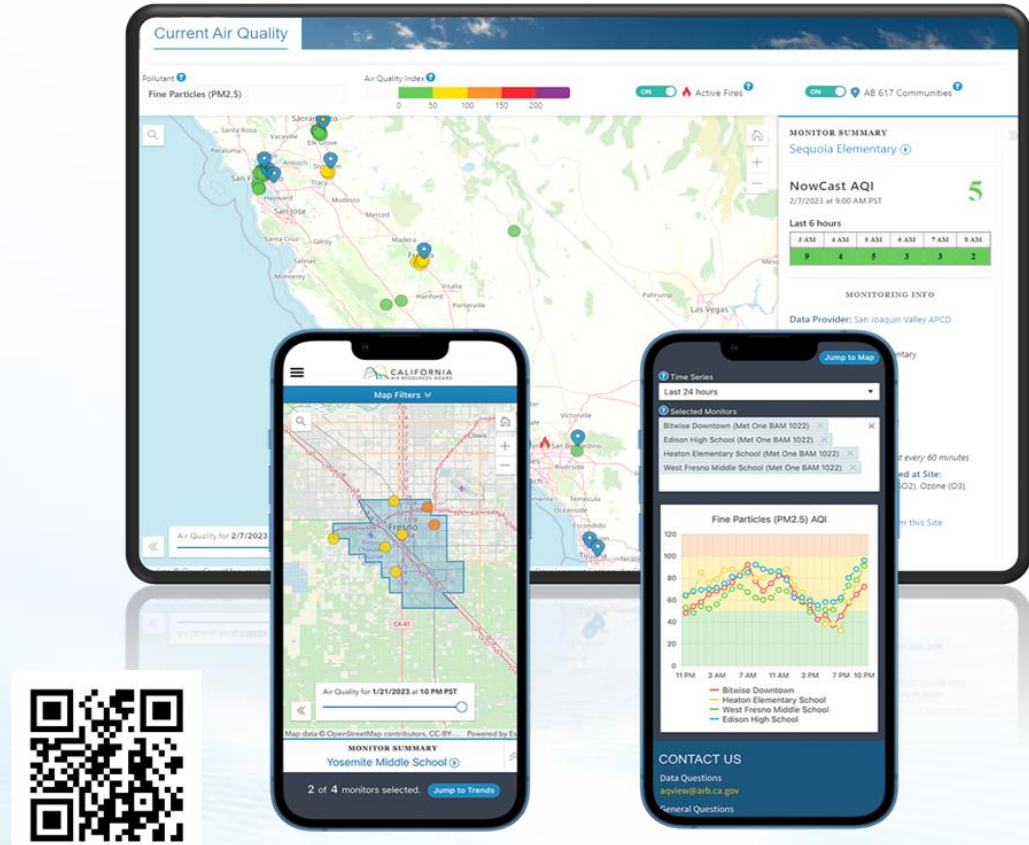Reporting Interval: Hourly, Subhourly

Purple Air Monitors

Sites: 7000+ and growing!

Reporting Interval: ~Minute

AQview is designed to support the volume, velocity, variety, and veracity of the data.

CARB

Introduction

- An innovative **cloud-based** air monitoring data system

- Provides visualization and easy access of air quality data from **diverse sources** in **near real-time**

- Designed with simple, intuitive, and **mobile-friendly** interfaces

- Provides **transparency** in how data are collected and processed

https://aqview.arb.ca.gov/

CARB

Main Features

Automated Quality Control
Upper and Lower Limits, Spikes, Repeating Values

Data Download
Measurement Data, Metadata

Reports Repository
Lab-Analyzed Data, Previous Studies, Reports

Data Visualization
Hourly PM$_{2.5}$ AQI Map, Time Series

Feedback Forum
Ask, Suggest, Report

Future Enhancements
Pesticides data, multilingual support, data summaries

Current Data by Region

Region	Communities/Networks	Sites	Pollutants*
Bay Area	3	71	$PM_{2.5}$, PM_{10}
Sacramento Valley	1	22	$PM_{2.5}$, PM_{10}, VOCs
San Joaquin Valley	5	25	$PM_{2.5}$, PM_{10}, Gases, VOCs
Great Basin	1	1	$PM_{2.5}$, PM_{10}
South Coast	7	18	$PM_{2.5}$, PM_{10}, Gases, VOCs
San Diego	2	7	Black Carbon
Imperial	2	46	$PM_{2.5}$, PM_{10}

Availability of pollutants varies by site.

See a complete data inventory for AQview here:
https://aqview.arb.ca.gov/resources/files/Pollutants_in_AQview.pdf

CARB

Challenges of Community Monitoring Data

AQview hosts air quality data from **low-cost sensors, regulatory-grade monitors, research-grade 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?

CARB

Assessing and Messaging Data Quality

Record-Level Quality Control Check		Overall Data Quality Assessment
Level 1 Preliminary QC	**Level 2** Enhanced QC	**Level 3** Overall Data Quality
Automated check for flagging obvious erroneous records	Automated QC checks with broader scope based on a longer time window	Cumulative statistical analysis. Are QC tests results remaining stable over time?
Results provided in graphical displays and data downloads	Spatial statistical tests applied to sensor clusters	Incorporates site and monitor QA factors
	Results provided in data downloads	Assess instrument reliability, performance, operation, and data provider reliability
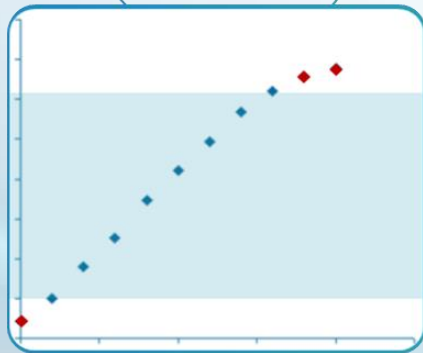
CARB

Level 1 – Preliminary QC
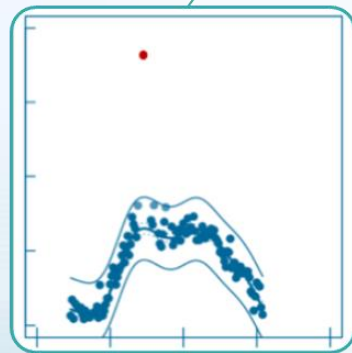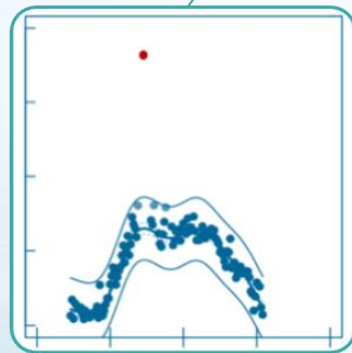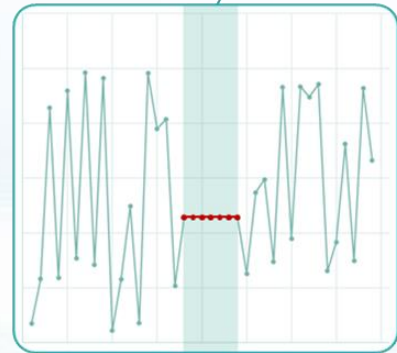
Example: Lower Detection Limit Check

PM$_{2.5}$ Lower Detection Limit Check Results

| Aug 2019 | Oct 2019 | Jan 2020 | Apr 2020 | Jul 2020 | Oct 2020 | Jan 2021 | Apr 2021 |

Site A · Site B · Site C · Site D · ✕ Failed QC Check
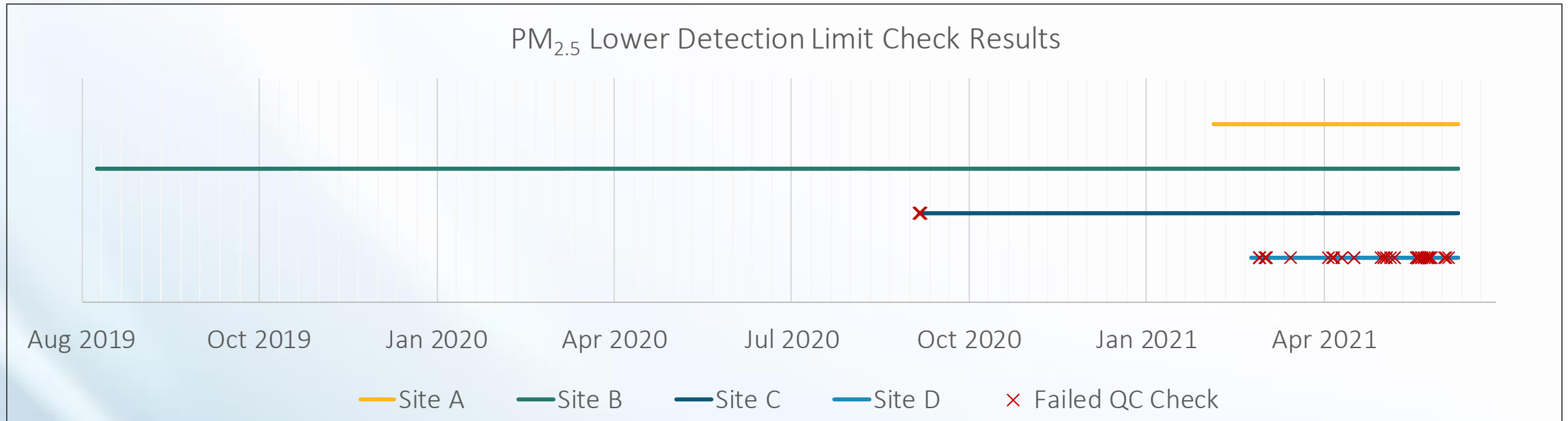
Site ID	Total Number of Records	% Pass	% Fail
Site A	2856	100	0
Site B	15822	100	0
Site C	6313	99.8	0.2
Site D	2380	97.7	2.3

CARB

9

Level 2 - Enhanced QC

Spatial-based anomaly identification
(e.g., cluster models, machine learning)

Temporal-based parametric statistical anomaly identification
(i.e., based on expected statistical distribution)

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

CARB
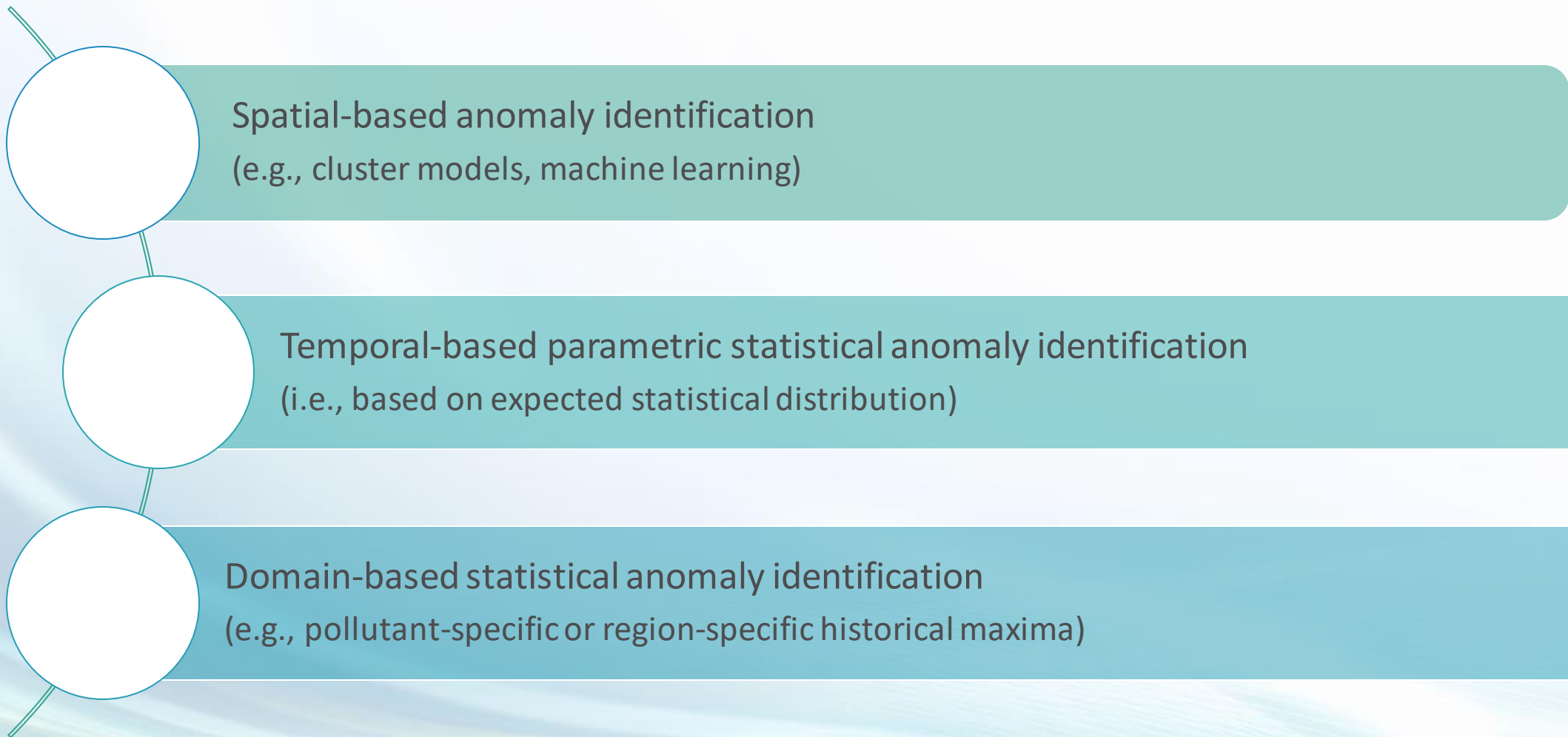
Example: PurpleAir Data Quality Control Checks

(Spatial-based Anomaly Identification)

- Identify questionable sensors by assessing the data from a group based on their spatial locations

Step 1: Parametric statistical tests against sensors within a local cluster

| Bias Relative to Cluster Median Test | Bias Relative to Standard Deviation Test | Mean Deviation Test | Standard Deviation Test |

If more than two sensors fail on a test cycle

Step 2: k-means clustering for anomalous sensor reporting as a member of a local cluster

CARB

# Example: Real-Time Outlier Detection

(Temporal-based Parametric Statistical Anomaly Identification)

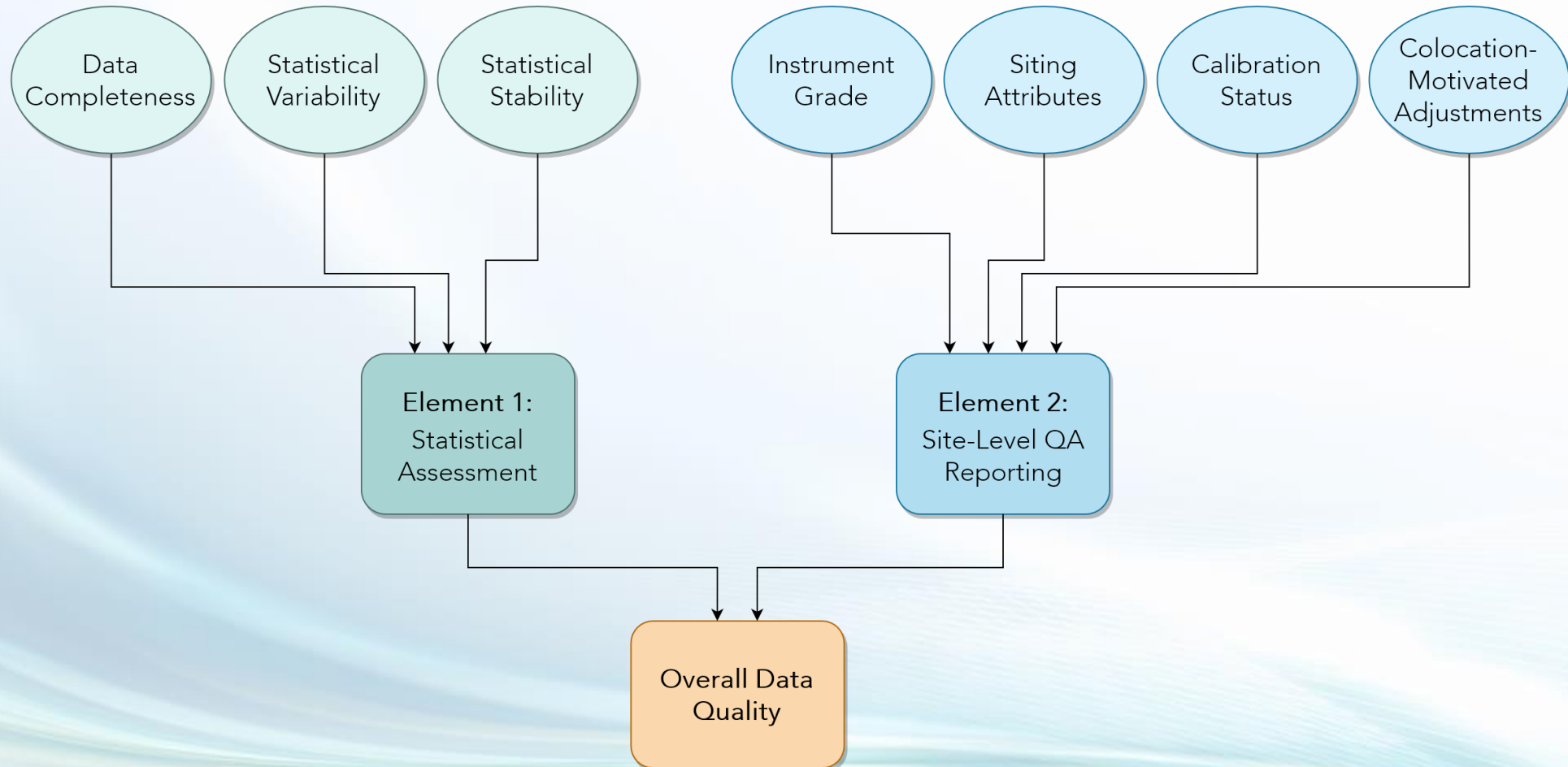Hourly Ozone Concentrations with Threshold Confidence Limits
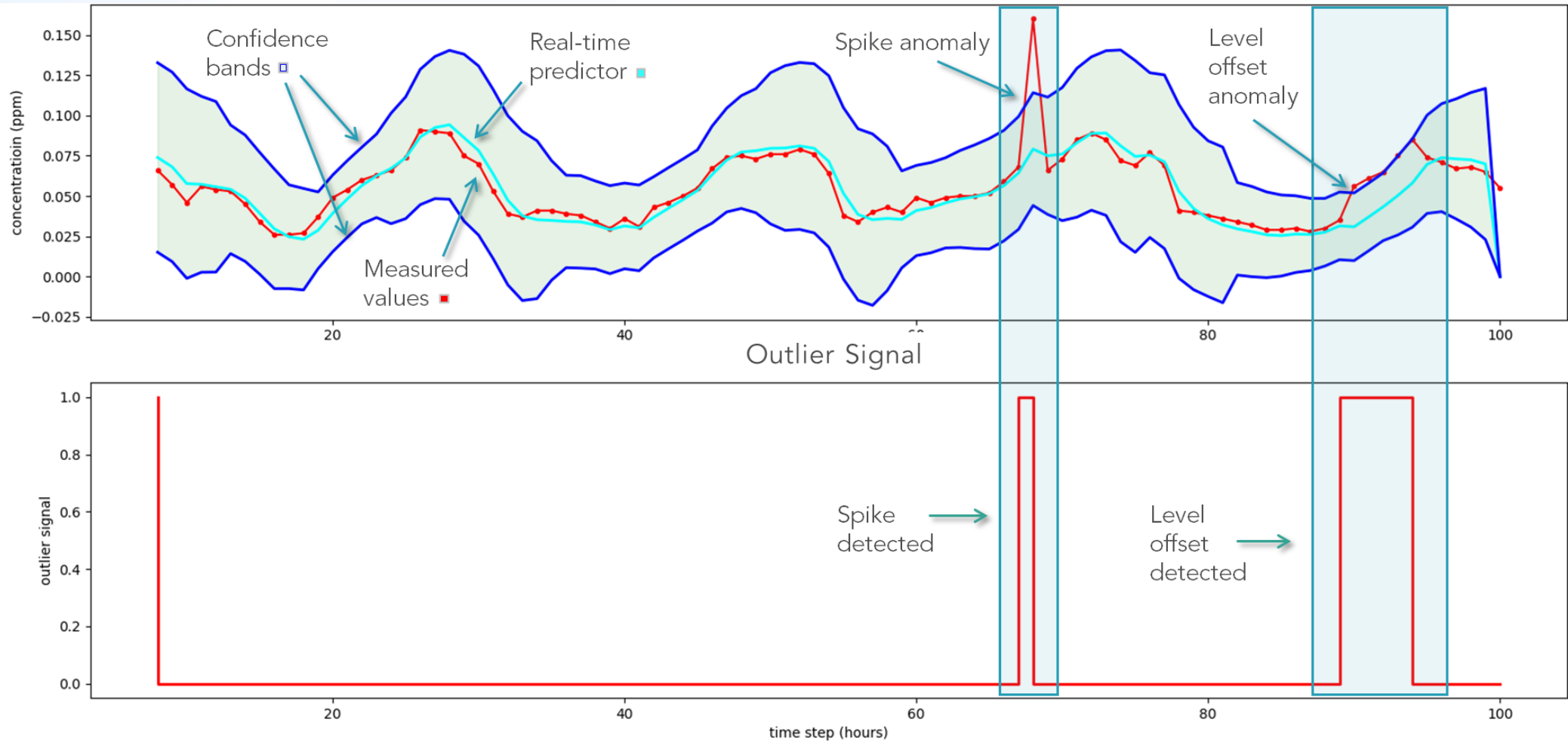
Level 3 - Overall Data Quality

- 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

Overall Data Quality

Are these data usable for analysis purposes?

How optimal is instrument operation?

CARB
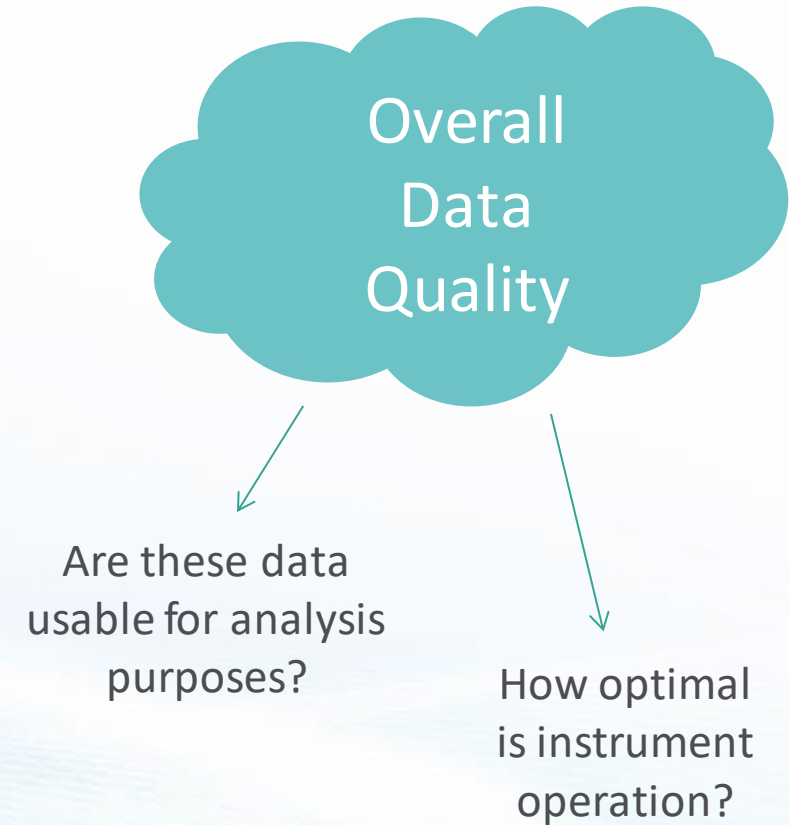
Level 3 - Overall Data Quality Elements *
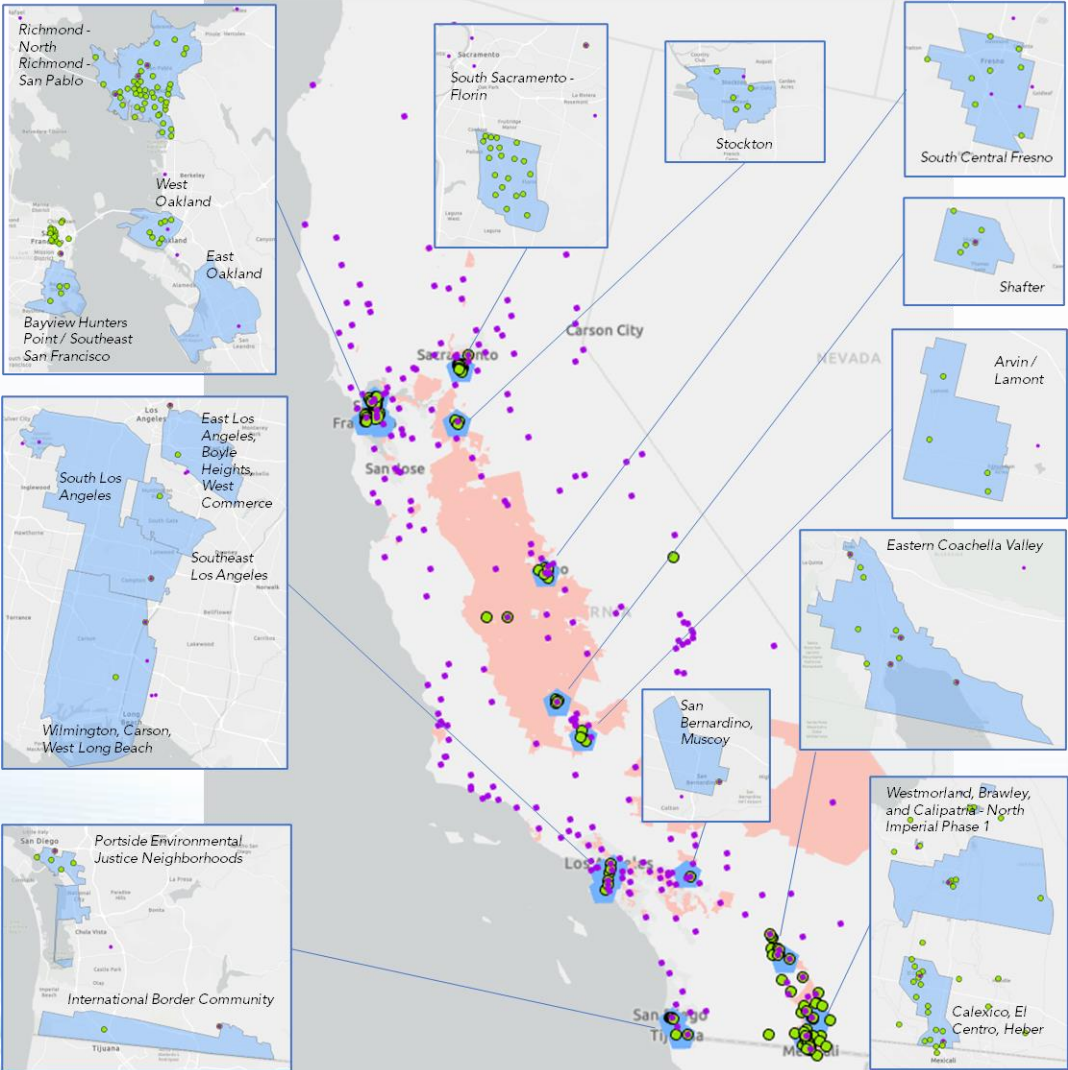
* Conceptual Framework

14

Data Available in AQview

More to come!

- Additional Data in AQview
  - Public sensor networks
  - Regulatory monitoring
  - More community monitoring

- Analysis and Interpretation of air monitoring data

○ Community-scale Monitoring Site

● Regulatory Monitoring Site (active)

⬠ AB 617 Community

▮ SB 535 Disadvantaged Community

CARB

AQview Team

Contact Us: aqview@arb.ca.gov

- Raiford Hann
- Emily Gorrie
- Annie Flores
- Eugene Kim
- Allison Cook
- Yanju Chen

Visit: https://aqview.arb.ca.gov

CARB