

Data Wrangling 4000 Sensors

From Raw Data to Results

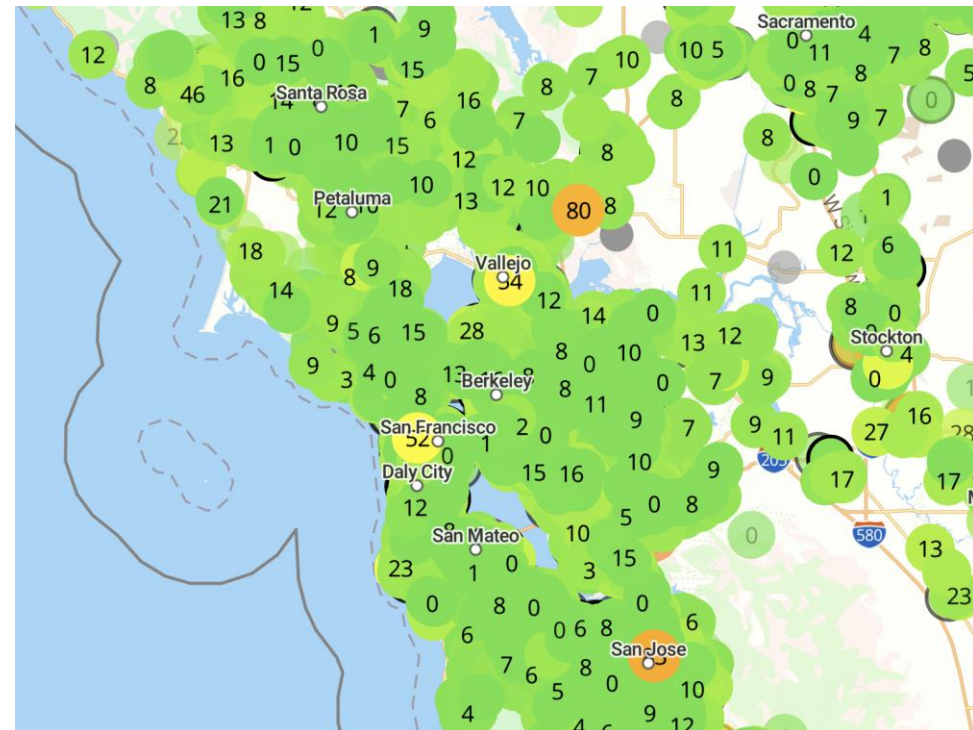
Helena Pliszka^a, Michael Flagg^b, Tim Dye^a, Katherine Hoag^b

- a. TD Environmental Services, LLC
- b. Bay Area Air Quality Management District



Bay Air Center

- Work with SLTs and Community groups
- Working with Bay Area communities through Bay Air Center
 - Bay Area Air Quality Management District
 - AQ Technical guidance, materials, and training resource
- Large network of air sensors in area already exists
 - Clarity, Aeroqual, PurpleAir
- Assemble data for variety of uses



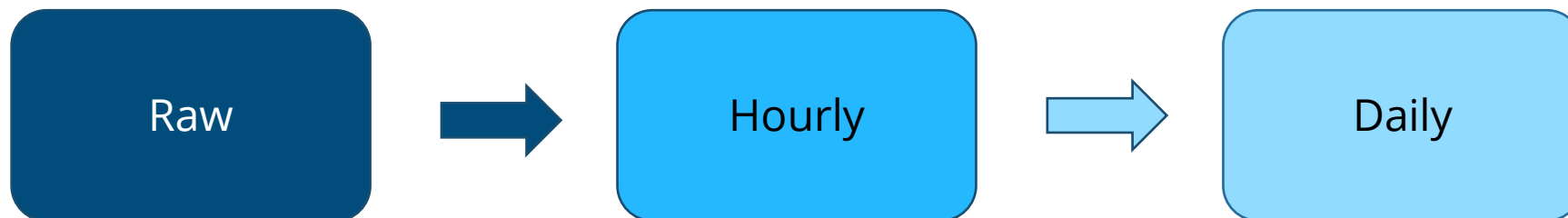
Purple Air map, July 11 2023, Bay Area

Potential Applications

- Inform both *ongoing community work, AD policy* and other *programmatic efforts* (like exceptional events and network assessments)
 - Answer questions about localized higher PM exposure
 - Identify information gaps in overburdened communities
 - Provide data insights to local community-led organizations for their air quality policy work
 - Support community-led advocacy and development of effective PM reduction strategies
 - Inform conceptual model of PM that highlights inequities in PM exposure
 - Visualize spatial variability of PM during different types of pollution episodes
 - Assess performance of EPA's Smoke & Fire Map correction factor
 - Add information for wildfire exceptional events demonstrations and PM monitoring network planning

Rigorous QA/QC

- Starting with Purple Air
 - High-resolution (~2 min) data
 - 2018 – 2022 period
 - 4038 sites
 - 150 GB of data
- Needed ***streamlined, automated QA/QC***
 - Series of adaptable and flexible Python programs
- Combined ***established QC protocols and novel checks***
 - Leveraged methods from
 - AirNow Fire and Smoke Map
 - PA degradation paper (deSouza, Barkjohn, 2022)
 - AirNow QC
 - EPA CFR completeness criteria
 - Designed protocols to meet large dataset needs



Raw QC Checks

High quality data requires quality control at high-resolution

Existing QC

- Removed data outside of operational PA range ($>1000 \text{ ug/m}^3$)¹
 - Reduce bias in averaging
- Reduced dataset to entries where ALL data (PM_{2.5} A,B and RH) are valid¹
 - Ensures uniform data completeness across relevant parameters

Existing QC doesn't address potential inclusion of indoor PM_{2.5} data that would bias results

Designed QC

- Removed indoor sites based on metadata (24 sites)
 - Site name containing 'Indoor', 'Inside'
 - Hardware version PA-I, PA-I-SD

Hourly QC Checks

Existing QC

- Retained complete hourly data based on EPA 75% criteria²
 - Daily QC also based on EPA criteria
- Removed data entries with disagreeing channel A, B PM_{2.5}³
 - Abs diff > 5 ug/m³ or rel. diff > 0.7
- Applied PA wildfire smoke correction⁴
- Removed data below AirNow min threshold (<-5 µg/m³)⁵

Existing QC doesn't address potential for early operation data issues or for sensors flagged as 'outdoor' behaving like 'indoor'

Designed QC

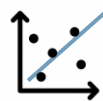
- Removed first 24 hours of data from each site
 - Early data showed anomalous PM_{2.5} spikes
- Removed high likelihood indoor sites based on PM_{2.5} and temp trends

Identifying Potential Indoor PAs

How can we use QC'ed data to start to gauge whether a sensor is indoors?

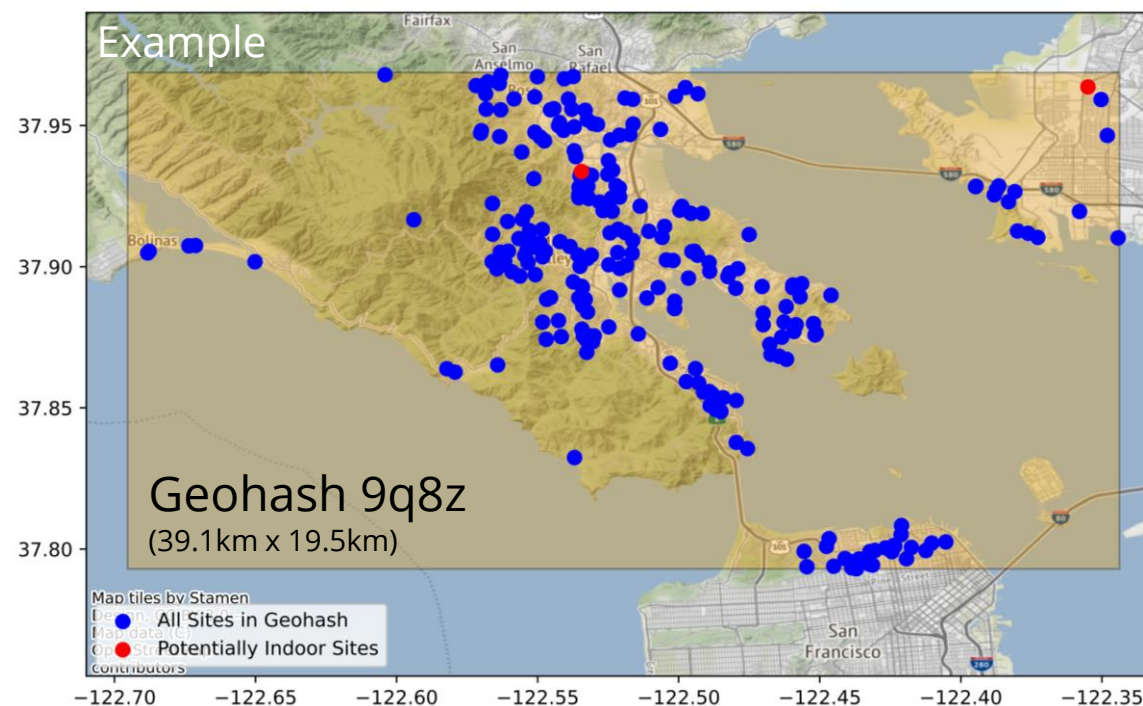


Warmer winter temperature + Uncorrelated PM_{2.5}



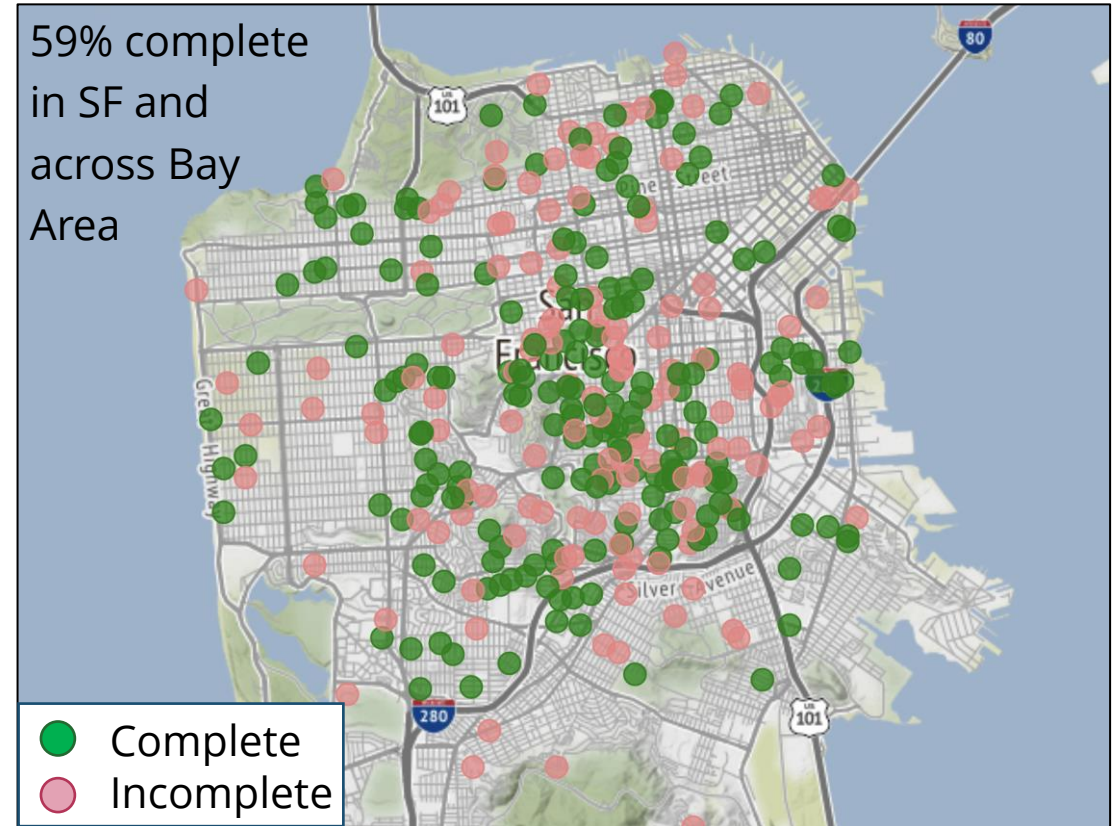
- Grouped sites by geohash
- Sites *automatically* flagged when both conditions satisfied
 - Site winter month avg (Feb) – area median $\geq 10^{\circ}\text{F}$
 - Site average hourly PM_{2.5} correlation ≤ 0.5
- Identified 17 potential indoor sensors
- Followed by *manual* review to double check
 - 10 highest likelihood indoor sites

Shows potential for large dataset QC design (e.g. buddy check, 'clean air' levels)



Annual Completeness

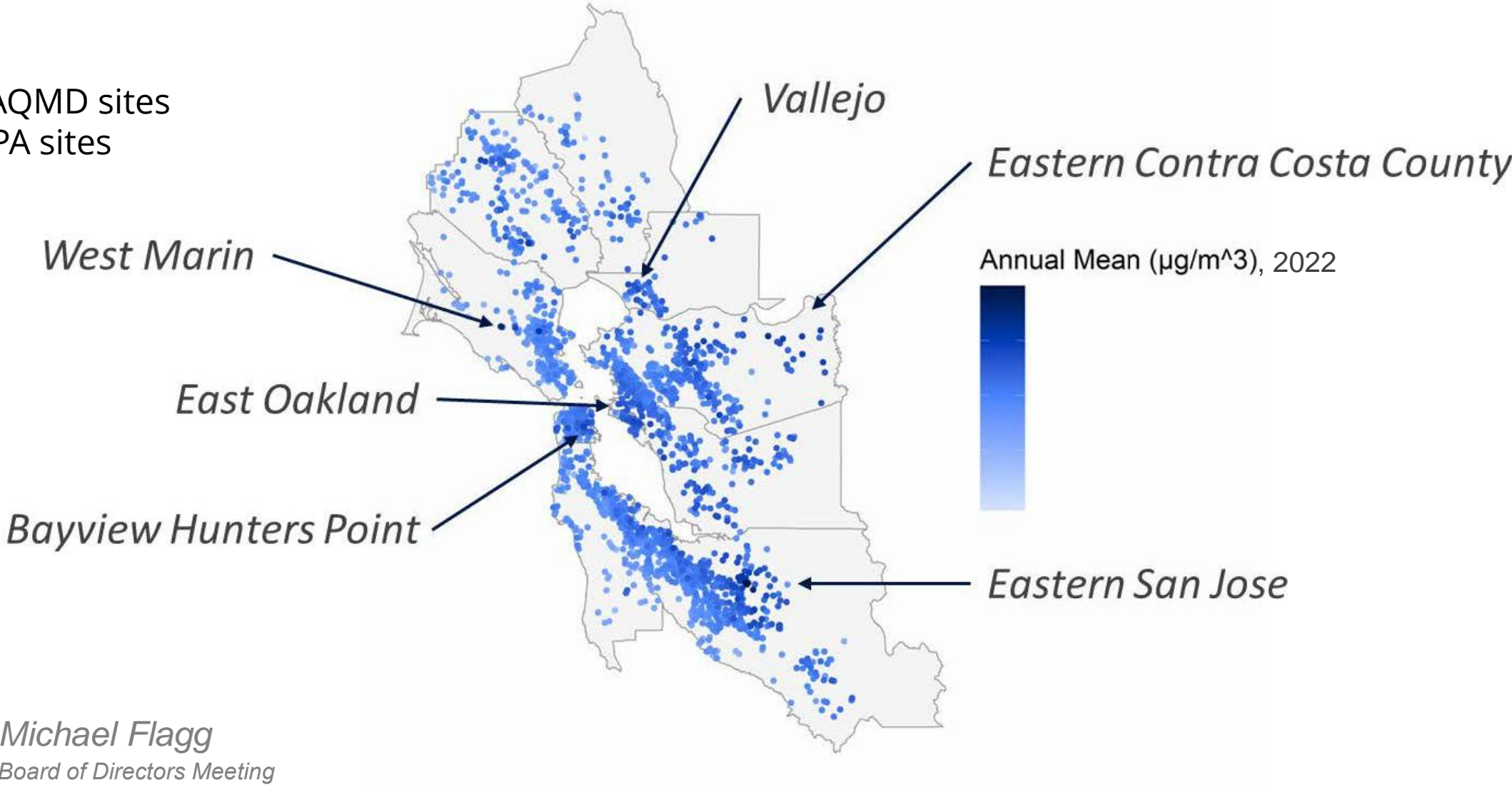
- High standard for screening data
 - EPA requires completeness for all four calendar quarters
- High incompleteness will limit dataset application
 - 2338 out of 3972 complete
 - Can't assess EPA annual metrics for many sites
 - Still allows for case study application



San Francisco Annual Completeness 2022

Data Preview

- 17 BAAQMD sites
- 2,338 PA sites



Michael Flagg
Board of Directors Meeting
July 19, 2023

Contact

Helena Pliszka

Helena.Pliszka@tdenviro.com

404-786-3556

Contact us to help with any aspect of your air monitoring programs:

- ⑩ Study design
- ⑩ Measurements
- ⑩ Data management
- ⑩ Analytics
- ⑩ Community engagement
- ⑩ Training & mentoring
- ⑩ Capacity building

References

1. DeSouza Priyanka, K. Barkjohn, A. Clements, J. Lee, R. Kahn, B. Crawford, AND P. Kinney. An analysis of degradation in low-cost particulate matter sensors. 2022
2. 40 CFR Part 50, Appendix N, Section 4.0
3. AirNow. AirNow Fire and Smoke Map Questions and Answers. 2022.
4. Johnson Barkjohn, K., A. Holder, A. Clements, S. Frederick, AND R. Evans. Sensor data cleaning and correction: Application on the AirNow Fire and Smoke Map. To be Presented at American Association for Aerosol Research Conference, Albuquerque, NM, October 18 - 22, 2021
5. AirNow QC, 2022