United States Environmental Protection Agency

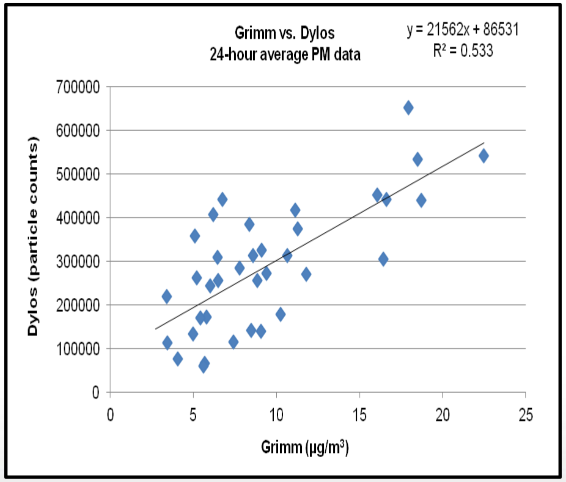
EPA Tools and Resources Webinar:
Low Cost Air Quality Sensors

Ron Williams
US EPA Office of Research and Development

April 17, 2019

Awaiting Low Cost Sensors (LCS)

- The value of emerging technologies to meet monitoring needs are unknown; key areas of uncertainty include:
  - Discovery - What sensors exist?
  - Evaluation - How well do they perform?
  - Application - How can they be used?



Anticipated Sensor Progression

Evaluations (Past)

Initial Performance Evaluations (in lab & field)

Short Term Studies/Applications

EPA Air Sensors Toolbox

https://www.epa.gov/air-sensor-toolbox

Air Quality (AQ) Spec

http://www.aqmd.gov/aq-spec

Networks (Present)

Smart Cities

Local Networks

Community Engagement

Near Source Monitoring

Long Term Performance Characterization

Sensor Evaluations

Integration (Future)

Data Quality

Data Interpretation

Data Management

Data Fusion

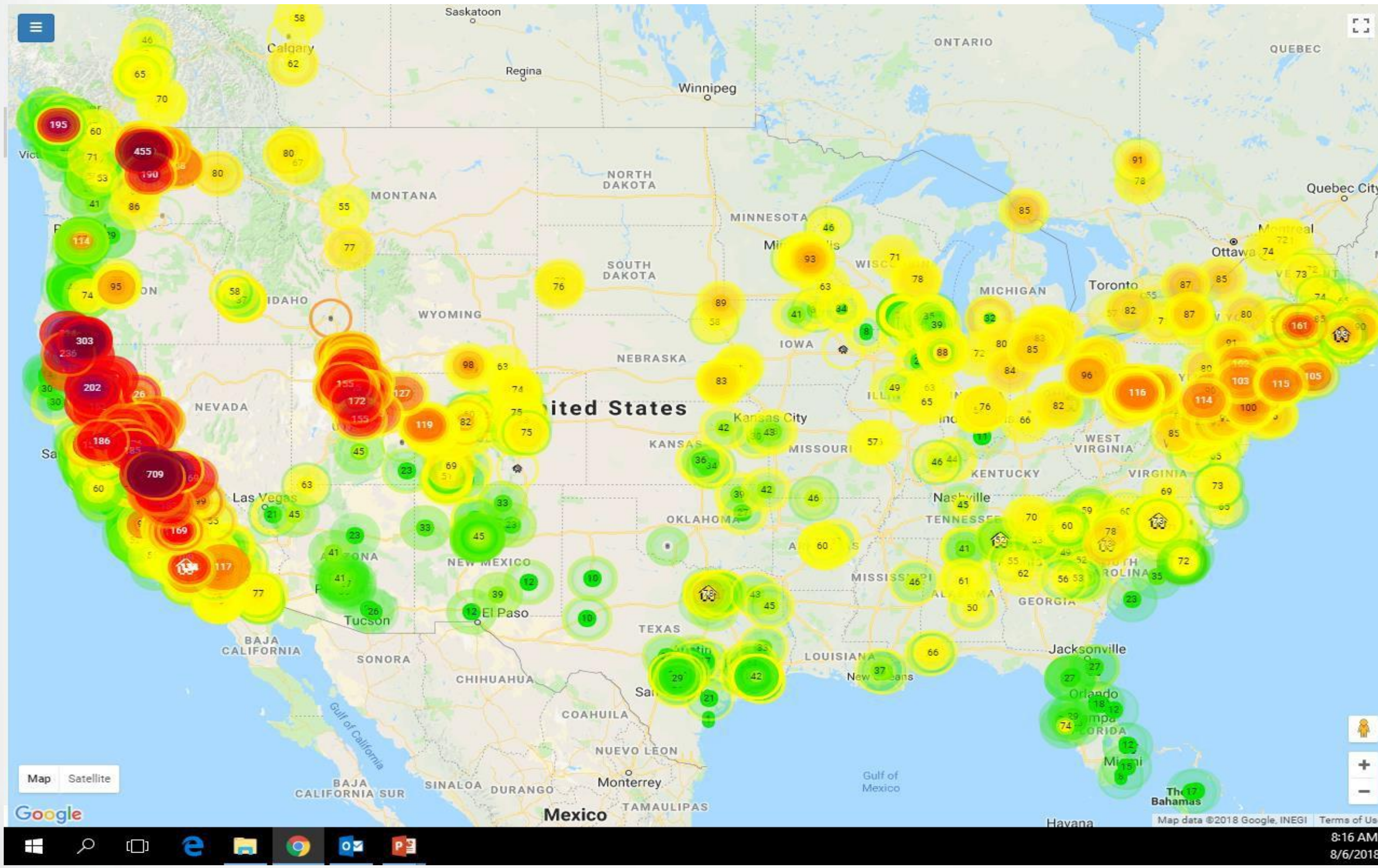
Certifications?

Goals for Low Cost Sensors

- More spatial data
- Higher temporal frequency
- Reduction in purchase and operation costs
- Reduced technical training and labor to operate
- Ease of data collection/recovery/transmission
- Replace (or at least supplement) regulatory monitoring
- Democratize air quality monitoring
- Provide developing countries the ability to define their air quality situation
- Provide enhanced risk assessment/epidemiological data

What is the Reality?

Particular Matter (PM) - More spatial data are a reality

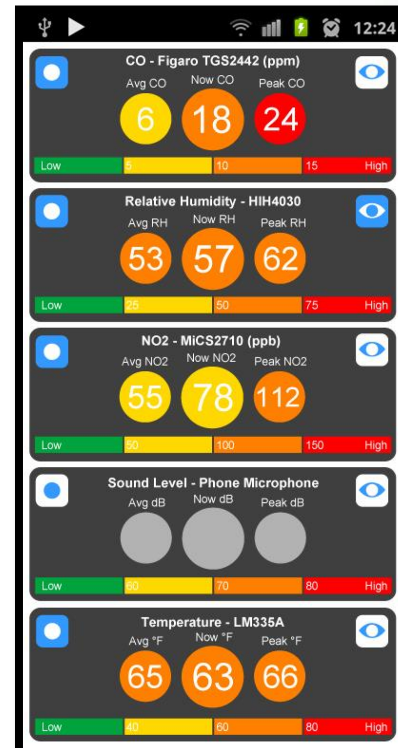


Numerous portals are now available reporting air quality sensor data

Purple Air network is one of many vendor-based data sources

What is the Reality?

Extensive spatial data coverage is often not a reality for NO_2, SO_2, CO, and VOCs



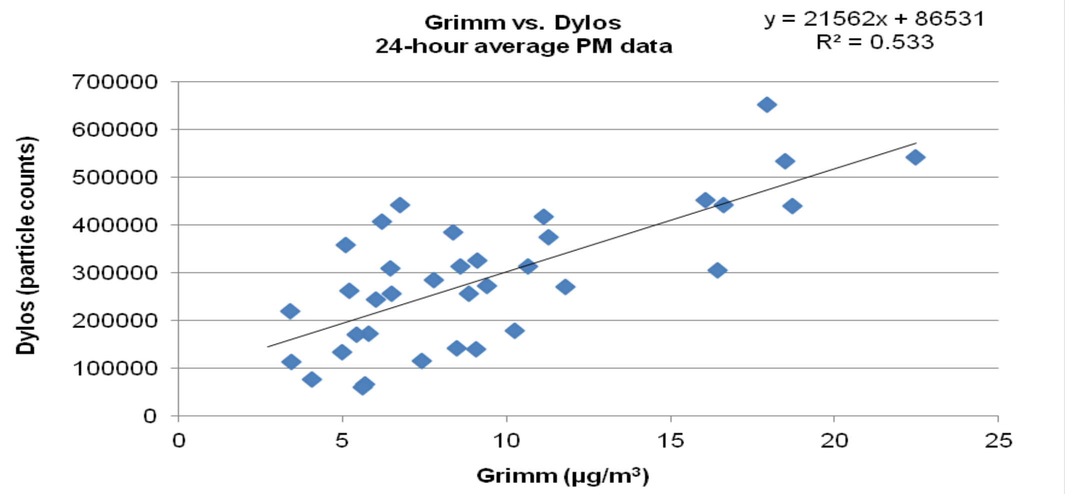
Courtesy of Michael Heimbinder, Habitat Map, Brooklyn NY

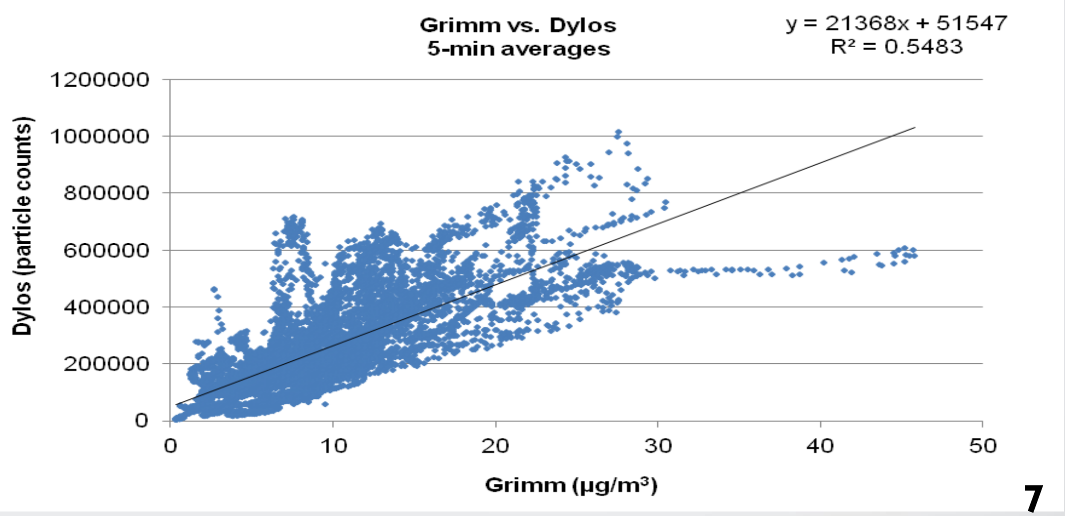
Limitations are probably a result of unknown performance of sensors (consumer confidence)

Reality-Higher Temporal Coverage

- Sensors often have the ability to detect/report data at 1 second intervals

- Is this valuable? Averaging intervals on data quality must be considered





Key Negative Considerations

- The amount of data being produced can become staggering. As an example:
  - A single monitor operating 24 hrs/day at 1 second time resolution for 1 week would produce >600K one second data points!
- Need more sophisticated data recovery and manipulation software; often earth mapping software is required to make sense of data (visual representation)
- Monitors are not without bias and noise – some pre-determined plan should exist for reducing this effect (either during or following data collections); basic bias and noise features of the monitor should be known before sampling is initiated

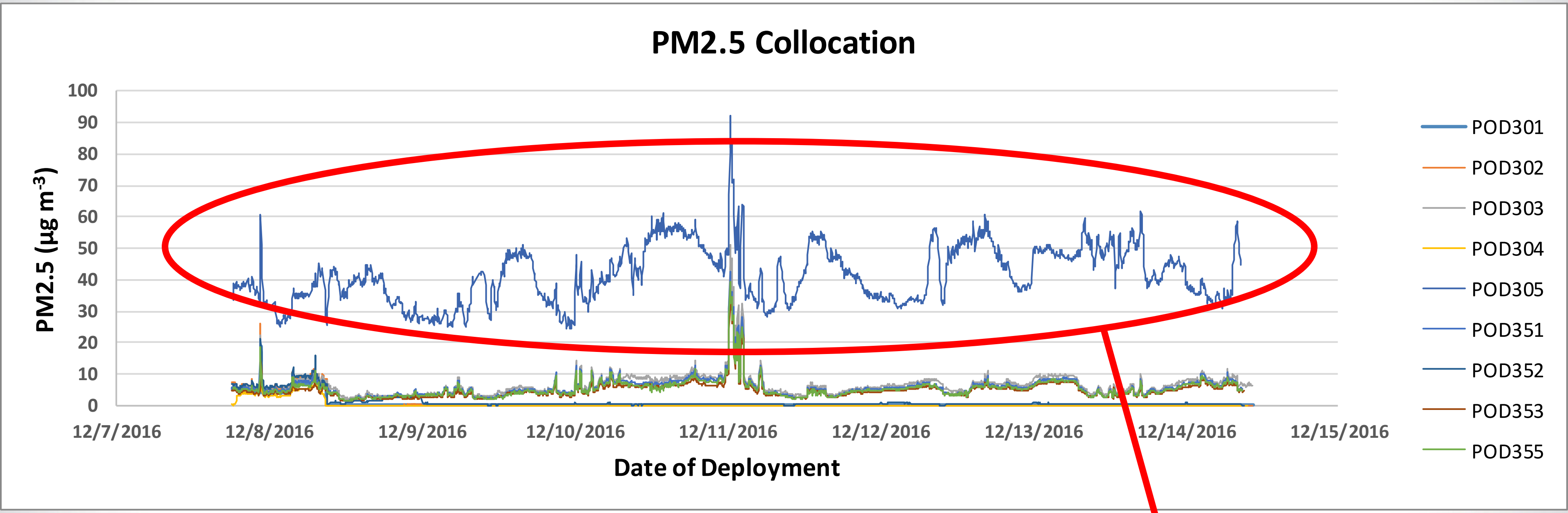
Goal: Lower Costs

- Most air quality sensors retail for \$100-\$2500
- Minimal or limited technical support often encountered
- Gas phase sensors have limited life span (~ 6 months to 1 year)
- PM sensors have longer lifespans (~1-3 years)
- Unforeseen costs (WiFi, cellular SIMs, vendor server costs) can exceed \$200/year
- Data collection often result in millions of data points
- Data analyses can result in significant expenditures or overwhelm end users

Reality: Skill Level of Operators

- Experience has revealed that many LCSs require ability to program script or other data handling activities
- Sensors may produce an output, but it takes an experienced eye to ferret out malfunctions or non-sensical data
- Data validation and tabulation becomes a major activity; this often is not an Excel type of data handling: SAS, Python, MATLAB, R or other tools needed to manage these extremely large datasets
- Automated quality assurance routines are needed to detect outliers and invalid output

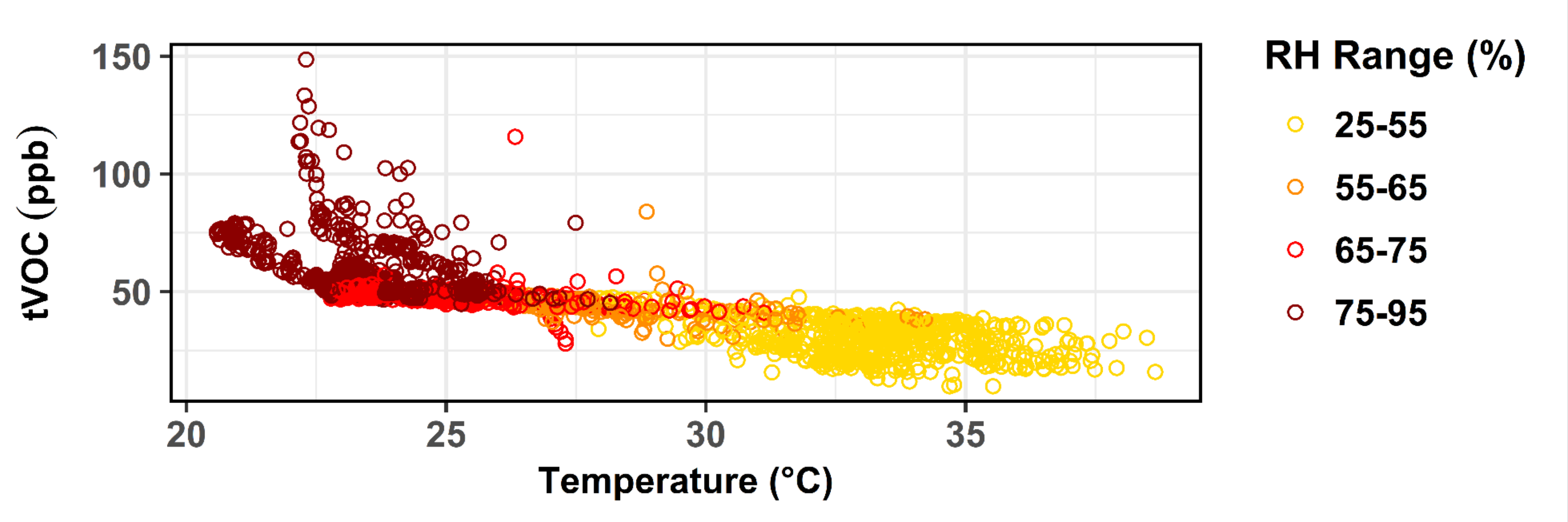
Unreasonable PM Response-Example



Response statistically different than other collocated replicates

Expect the Unexpected

An example of multiple response scenarios for a single total volatile organic compound (tVOC) sensor for relative humidity (RH) and temperature



Data need to be carefully examined for quality assurance features

Sensor Failure-PM Example

A-H Log 1	RTC_date	RTC_time	Shinyei 1 (	CC 1 O3/N	AQ 1 O3 (p	Dylos 1 Sm	Dylos 1 Lg	Dylos 2 Sm	Dylos 2 Lg (pt/0.01 cf
A-H Log 1	10/27/2015	0:00:02	2.376	255	0.001	146	11	883	83
A-H Log 1	10/27/2015	0:01:00	2.664	255	0	141	9	891	65
A-H Log 1	10/27/2015	0:02:00	2.25	255	0.002	110	6	816	56
A-H Log 1	10/27/2015	0:03:00	2.07	255	0.003	118	5	773	45
A-H Log 1	10/27/2015	0:04:01	2.214	255	0.003	105	7	777	43
A-H Log 1	10/27/2015	0:05:01	2.106	255	0.002	95	5	753	42
A-H Log 1	10/27/2015	0:06:01	2.052	255	0.002	112	6	749	40
A-H Log 1	10/27/2015	0:07:01	1.602	255	0.002	98	5	761	39
A-H Log 1	10/27/2015	0:08:01	1.656	255	0.001	97	5	751	43
A-H Log 1	10/27/2015	0:09:02	1.422	255	0.003	96	6	754	40
A-H Log 1	10/27/2015	0:10:02	1.8	255	0.002	92	3	746	37
A-H Log 1	10/27/2015	0:11:00	1.476	255	0.003	94	5	723	38
A-H Log 1	10/27/2015	0:12:02	1.44	255	0.001	92	4	706	35
A-H Log 1	10/27/2015	0:13:00	2.142	255	0.003	81	2	722	36
A-H Log 1	10/27/2015	0:14:00	1.512	255	0.003	99	5	708	33

Note, the repetitive 255 value from the Cairpol Cairclip sensor. Just because there is a data value output does not mean the value is useful. Represents a non-defined manufacturer fail state.

Goal: Ease of Data Transmission

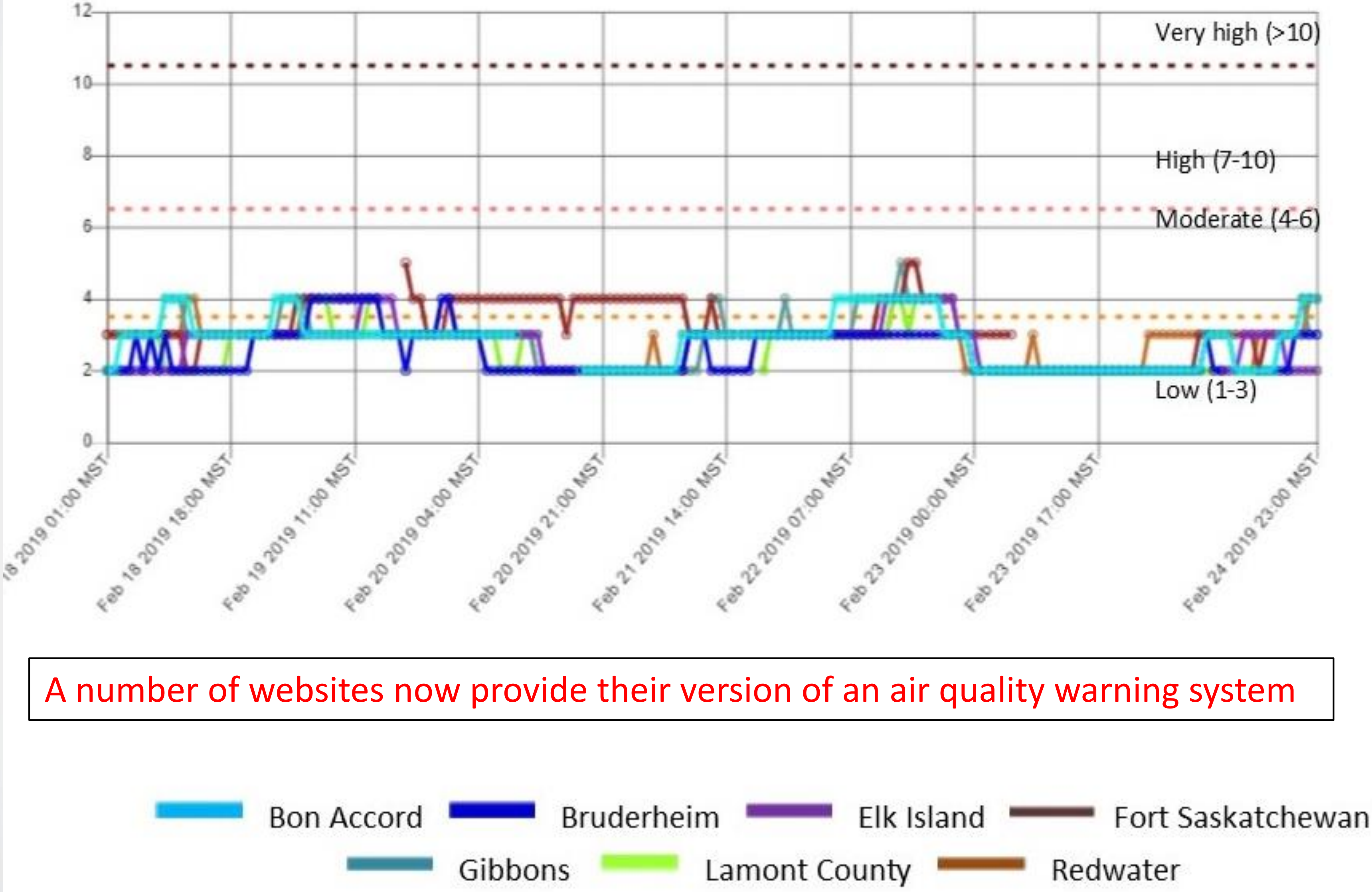
- Many LCS promise ease of use features relative to data transmission events
- WiFi and cellular often defined as turn-key features
- Hardships occur when users have to deviate from vendor-defined specifications
- Many vendors are unable to provide fast technical support to overcome data transmission troubles
  - Vendor provided script "buggy"
  - Vendor script produces data outputs resulting in a host of issues (microprocessor failure, reboots, etc.)
- End user data handling often requires a high level of coding and engineering skills

Goal: Replacement/Supplement of Regulatory Data

Regulatory officials and those governed by regulatory requirements (e.g., industry) are often hesitant to accept LCS data relative to being actionable; these situations may be associated with:

- Unknown data quality of the LCS and how it was operated
- Undefined features of the LCS with respect to interferences, range of applicability
- Lack of a QAPP (hypothesis driven data collection)
- Lack of sufficient data analyses needed to validate raw data
- Non-data defined conclusions (unsupported by data collections/analyses)
- Inappropriate data conclusions (e.g., use of 5 min value to reflect health risk for a 24-hr based NAAQS)
- Vendor based health indices (pseudo AQIs) using real-time LCS data often undefined with respect to their underlying science or statistical basis

Example of a Community-based Air Quality Index

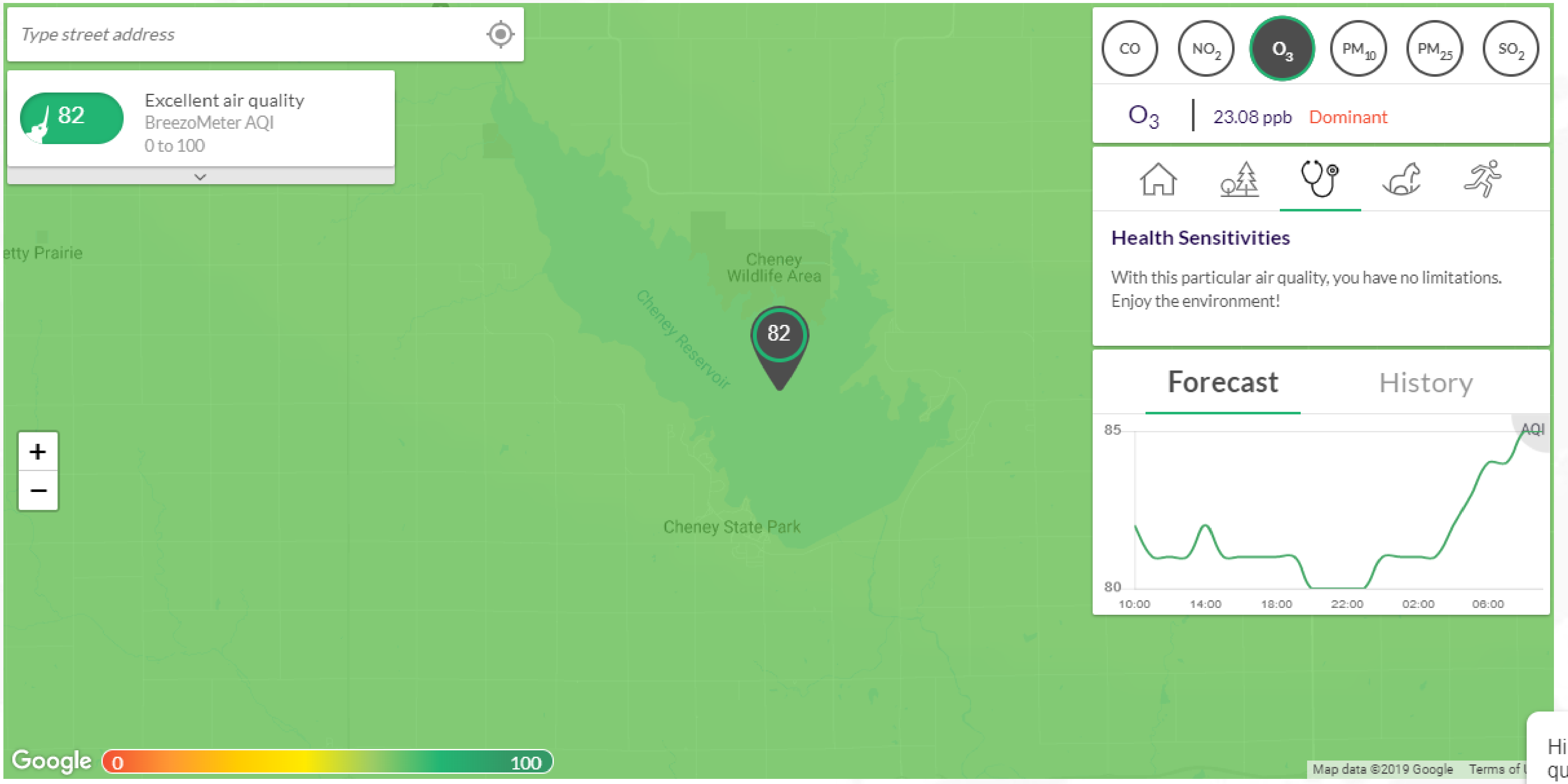


A number of websites now provide their version of an air quality warning system

Example of a Vendor-based Air Quality Index

Air Quality at queried location

Another example of a non-Air Quality Index reporting network



Vendor-based Continuous Health Risk Warnings

What are the Pitfalls of Such Health Indices?

- Includes assumptions about duration and impact of a short-term value representing a long-term health risk
- Health risk not associated with statistically-defined epidemiological findings
- Monitoring device (LCS) often have accuracy errors of 50-100% and typically biased high-potentially false warnings
- Risk associated with only one or a series of pollutant species
- Indoor/mobile/occupational monitoring locations but sensor uses ambient-based health indices

The New Air Quality Paradigm

- It is vital that an objective perspective be used in establishing the value of data from LCS
- Data should not be discarded by regulatory/industry officials just because it was obtained by LCS
- Data should not be considered accurate just because the LCS yielded a value
- Key is defining data quality and the fit for purpose attributes of the measurement/data set
- Monetization of LCS data by a host of parties is of potential concern
  - Data quality/integrity
  - How it is being used
  - How it is being viewed

QA Overview

- Bias- generally undefined by vendor and most researchers
- Precision- can be quite good (<10% error) but anomalies are often observed
- Calibration- chamber calibrations are often high (>95% agreement) but ambient conditions are so/so
- Detection limit- often quite acceptable (SO_2 being an exception)
- Response time- very acceptable for all situations except mobile applications
- Linearity of sensor response- high in chambers but interferences impact ambient response
- Measurement frequency- longevity of LCS lifetimes vary widely
- Data aggregation- higher time averaging improves agreement with reference measures
- Specificity- PM sensors respond to all light scattering materials; EC and MOS sensors respond to a host of gases
- Interferences- RH, temperature often found to influence response
- Sensor poisoning and expiration- chamber studies have shown poisoning to be a real concern
- Dynamic range- usually well within the ability of most LCS (PM, gases, tVOCs)
- Drift- established for some light scattering devices, undefined for most gas phase LCS
- Accuracy of timestamp- inconsistent nature of timestamps often a reality
- Data completeness- sudden or unknown failures often observed

Reported Literature Application Categories

- Air quality forecasting
- Air quality index (AQI) reporting
- Community near-source monitoring
- Control strategy effectiveness
- Data fusion
- Emergency response
- Epidemiological studies
- Exposure reduction (personal)
- Hot-spot detection
- Model input
- Model verification
- Process study research
- Public education
- Public outreach
- Source identification
- Supplemental monitoring

Frequency of DQOs/DQIs Reported

Performance Characteristic/DQI	$PM_{2.5}$	PM_{10}	Carbon Monoxide (CO)	Nitrogen Dioxide (NO_2)	Sulfur Dioxide (SO_2)	Ozone (O_3)
Accuracy/Uncertainty	84% (16)	77% (10)	65% (11)	68% (15)	80% (4)	76% (19)
Bias	5% (1)	8% (1)	18% (3)	9% (2)	40% (2)	16% (4)
Completeness	26% (5)	31% (4)	12% (2)	14% (3)	40% (2)	16% (4)
Detection Limit	26% (5)	8% (1)	47% (8)	32% (7)	80% (4)	24% (6)
Measurement Duration	26% (5)	8% (1)	18% (3)	14% (3)	0% (0)	20% (5)
Measurement Frequency	26% (5)	15% (2)	35% (6)	23% (5)	0% (0)	32% (8)
Measurement Range	47% (9)	46% (6)	35% (6)	32% (7)	80% (4)	40% (10)
Precision	42% (8)	31% (4)	29% (5)	36% (8)	80% (4)	32% (8)
Response Time	0% (0)	0% (0)	29% (5)	32% (7)	80% (4)	20% (5)
Selectivity	11% (2)	8% (1)	24% (4)	23% (5)	80% (4)	16% (4)
Other	5% (1)	8% (1)	0% (0)	0% (0)	0% (0)	8% (2)
% All Information Sources	40% (19)	27% (13)	35% (17)	46% (22)	10% (5)	52% (25)

( ) represents the number of references used in the statistic

Take Home Messages

- Low cost air quality sensors are being developed and used world-wide
- Much work remains in understanding sensor performance
- EPA is sharing tools and knowledge with all of its stakeholders
- There is a common goal in understanding how these sensors can be used purposefully
- The use of networked sensors, new analysis and visualization tools are bringing insight to the questions

Upcoming Events or Activities

- Publication of the 2018 Sensor Performance Targets Workshop Summary (Atmospheric Environment) – spring 2019
- EPA's Performance Targets discussions on PM_{10}, NO_2, SO_2, CO (RTP, NC) – tentatively summer 2019 - save the date notice released soon (https://www.epa.gov/air-sensor-toolbox); states and other partners welcomed
- EPA's Sensor Loan Program (ORD & Regions) – ongoing
- EPA's Long Term Performance Evaluations: 6 locations across US with common group of LCS – summer 2019
- 2019 Air Sensor International Conference – summer 2019

Resources and Contact Information

Future EPA points of contact:

Vasu Kilaru
Kilaru.Vasu@epa.gov

Gayle Hagler
Hagler.Gayle@epa.gov

Andrea Clements
Clements.Andrea@epa.gov

https://www.epa.gov/air-sensor-toolbox

Disclaimer: Name or inclusion of any sensor here is not endorsement or recommendation for use by US EPA