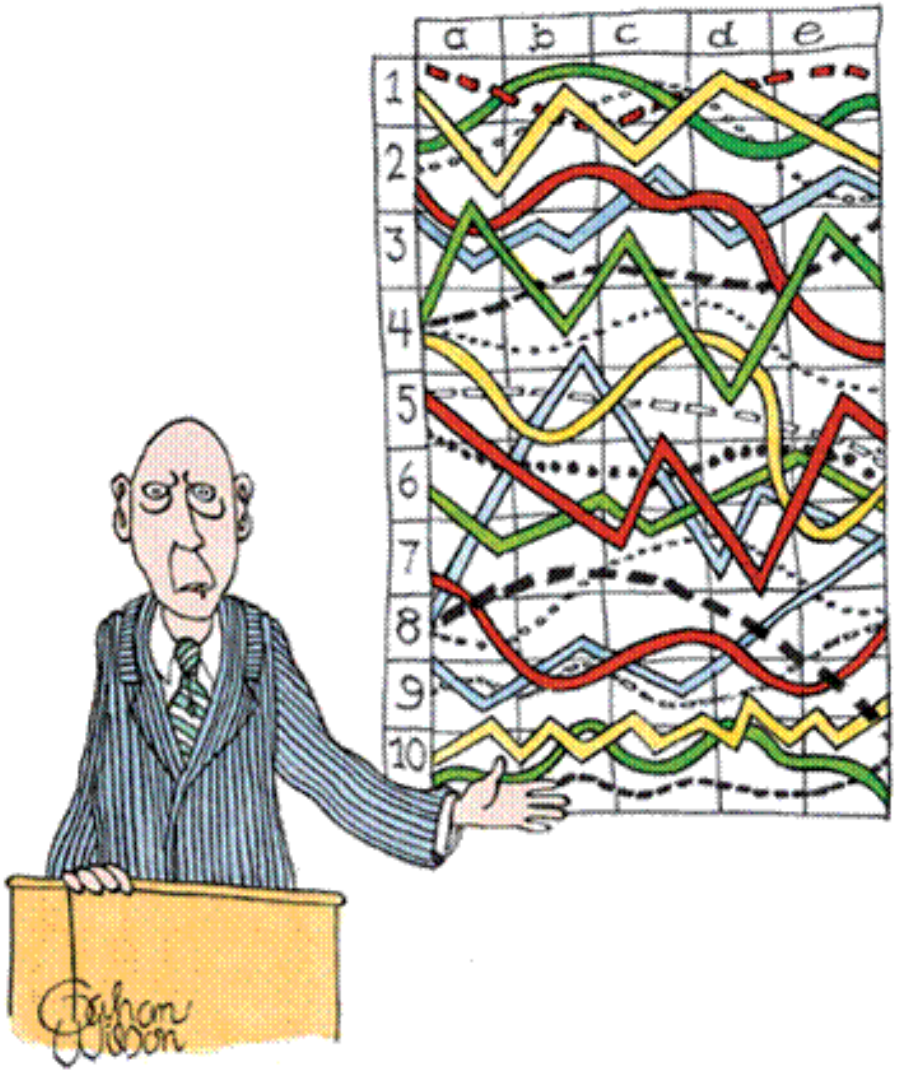
Sensor Data Limitations: Interpretation, Messaging, and Uses

Dena Vallano, PhD, ORISE Fellow, Air Division
U.S. Environmental Protection Agency, Region 9



"I'll pause for a moment so you can let this information sink in."

Data Interpretation: What does it mean?

- Me: How does air pollution affect my health? What is my least polluted commute route?
- Communities: Is my neighborhood air quality ok? Are our kids playing in a safe environment?
- Local governments and planning agencies: How well are we balancing growth, development, and public health?
- Governmental air agencies: How to effectively address community concerns and apply sensor results?

Data Interpretation: Challenges

- Good data interpretation starts with identifying specific objectives, careful study design, QA, and measurement uncertainty
  - Guidance is needed for users on choosing which sensors/projects best meet their needs and understanding results to make better use of measurements
- Sensors presents several unique challenges related to analysis and interpretation:
  - Availability of sensors (affordability)
  - Mobility of the sensors
  - Results in large data sets (“Big Data”) with high temporal and spatial resolution (sampling intervals of seconds to minutes)
  - Local influences
- Real-time air pollution monitor measurements should be validated prior to their analysis and interpretation

Making Sense of Big Data

- Personal sensors do not equate to regulatory data
  - NAAQS are set with long-term datasets
  - Regulatory monitors have very rigorous quality requirements and oversight
- Interpretation of high resolution data in the context of regulatory standards
  - Consideration of spatial and temporal representativeness
- Example: Sensor Ozone Measurements
  - 8-hr ozone standard is 75 parts per billion (ppb), but how should the public interpret the health implications of shorter-term averages if they exceed the standard?
  - Is it safe for ozone levels to be at 100 ppb for only one hour or one minute?
- EPA recognizes that accurate messaging is needed for short-term personal air quality measurements that guide exposure mitigation and behavior change

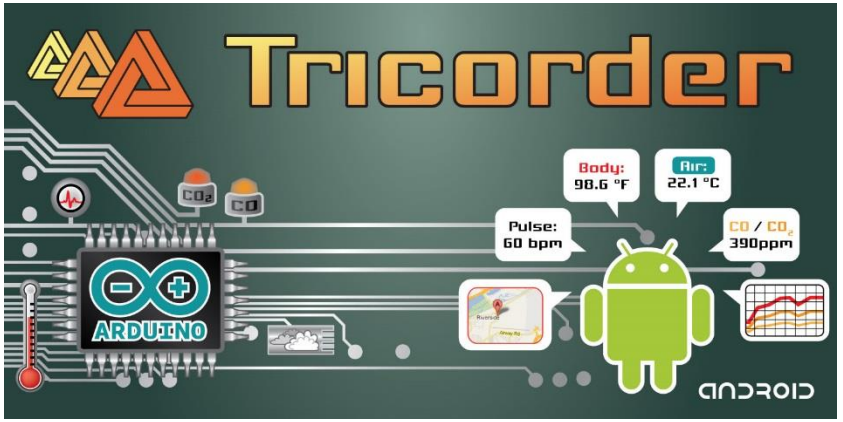
Data Reporting

- Privacy issues, including a general apprehension of users to share sensitive data

- Training users to understand technical information and gain confidence in their data-collecting skills is critical for active engagement
  - identification of objective
  - data-collection and methods
  - tracking and sharing of metadata
  - handling data quality issues post-collection (averaging, quality assurance)
  - data interpretation
  - data fusion with model and regulatory observations
  - data visualization and presentation (i.e. conveying uncertainty)

Data Uses: Education

- Using sensors in educational settings for STEM (science, technology, engineering, and math) curricula and promotion



Wright Brothers Institute Student Project 2012

- Example: Sensors are provided to students to monitor and understand air quality issues – and they have a blast doing it

Data Uses: Information/Awareness

- Using sensors for informal air quality awareness

- Example: A sensor is used to compare air quality at people's home, work, in their car, local park, or at their child's school.

Data Uses: Personal Exposure Monitoring

- Monitoring the air quality that a single individual is exposed to while doing normal activities

- Example: An individual having a clinical condition increasing sensitivity to air pollution wears a sensor to identify when and where he or she is exposed to pollutants potentially impacting their health

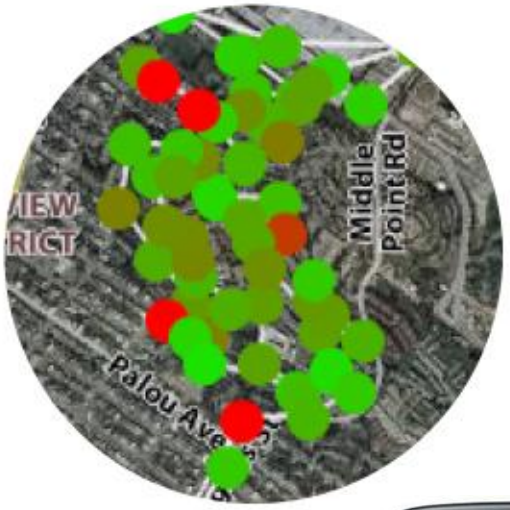
Data Uses: Research

- Scientific studies aimed at discovering new information about air pollution

- Example: A network of air sensors is used to measure particulate matter variation across a city, a neighborhood, a few blocks, etc.

Data Uses: Supplemental Monitoring

- Placing sensors within an existing state/local regulatory monitoring area to fill in coverage and assess network adequacy



- Example: A sensor is place in an area between regulatory monitors to better characterize the concentration gradient between the different locations

Data Uses: Source Identification and Characterization

- Investigate possible emission sources by monitoring near the suspected source.

- Example: A sensor is placed downwind of an industrial facility or near a busy intersection to monitor variations in air pollutant concentrations over time.

Data Uses: Policy Implications …an EPA perspective

- Sensor data is currently not used to determine whether an area is in compliance with the NAAQS
- Non-regulatory (i.e. secondary) data has informed boundaries for nonattainment areas and to support additional monitoring in areas of concern
- EPA does not expect personal sensors to be used for regulatory decisions
  - Guidance would help clarify appropriate uses of secondary data from sensors

"Huge volumes of data may be compelling at first glance, but without an interpretive structure they are meaningless."

— Tom Boellstorff, Ethnography and Virtual Worlds: A Handbook of Method

Thanks!

Contact information: Dena Vallano
(vallano.dena@epa.gov)

Disclaimer: Mention of commercial products does not constitute endorsement or recommendation for use and are provided here solely for informational purposes