Sensing our air: the quest for big data about our air quality

Gayle Hagler, PhD
EPA Office of Research and Development
Purpose of this talk

- Share emerging technology trends for measuring air quality and migration from “small, high quality data” towards “big, variable-quality data”
- Tee up for discussion some significant gaps
Measuring the air

Air quality is in the public consciousness

New York Times, April 1, 2013

Air Pollution Linked to 1.2 Million Premature Deaths in China

Chicago Tribune, June 27, 2014

EPA finds rail yards transfer pollutants as well as freight

USA Today, March 20, 2015

Air quality can affect students’ decisions to study abroad

France 24, April 10, 2015

AIR POLLUTION LINKED WITH BEHAVIORAL ISSUES IN CHILDREN, STUDY SAYS

Paris mayor races to improve air quality ahead of marathon
Measuring the air

In the United States, traditional paradigm for air quality monitoring:

Example: Measuring fine particles in New York

Rigorous protocols and methods for regulatory applications

- Expensive instruments (>$10K)
- Specialized training required
- Large physical footprint
- Large power draw
Recent explosion of lower cost sensor technology for measuring air quality:

**The driver:** What is my exposure? What is my child’s exposure? What is the air quality like in my neighborhood?
Measurement challenges

Researchers using advanced data processing strategies to get meaningful information from low-cost (~$20) sensors:

“We’re compensating for a bad sensor with machine learning…”
- Illah Nourbakhsh, sensor developer at Carnegie Mellon University, during a recent interview

“…a separate model was selected for each sensor....Fifth order polynomial models that included relative humidity (RH %) and temperature (C) was found to best convert PUWP signals into PM$_{2.5}$…”
- Gao et al., 2015, A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM$_{2.5}$ in Xi’an China. Environmental Pollution
Currently a disordered evolution

Buy sensor device

Evaluate sensor performance

Use sensor to discover air quality trends

Data analytics, model/data fusion

Database (sometimes data analytics)

Publish findings in peer-reviewed literature

6 months to 1 year after measurement

Publish in real-time to web or app

Evolving business models:
1. Sell the device, uncontrolled processing of data
2. Sell the device, manufacturer controls data processing
3. Manufacturer controls device, sells the data

Now!
Our case study: Village Green

- Data pushed to web server every minute
- Data quality reviewed by algorithms
- Quality-checked data made available to public immediately

**Project website:** epa.gov/villagegreen

**Locations:** Durham, NC; Washington, DC; Philadelphia, PA; Kansas City, KS; Oklahoma City, OK*; Hartford, CT*; Chicago, IL* (*to be installed in summer 2015)
Where this is tracking

Exponential increase in new measurements

Existing data sources
- Monitoring networks

Data analytics, data fusion, model-data fusion

Satellite remote sensing data

Timely and detailed information to manage air quality
- Individual action
- Community action
- Industry action
- Government action
Challenges:

• Who has access to the data?

• What are the best practices – sensor use, data analytics, data communication?

• Who provides the database, data schema, data rules?

• What can be said about data collected at unprecedented rates (1 second, 1 minute) in advance of relevant health research?

• Should there be a certification process for sensors?

• How to minimize erroneous information to the public but not stifle the evolution of IoT and IoE for air?