Characterizing Air Quality in a Rapidly Changing World

A discussion on projects collecting large volumes of unique air quality data around the United States and how the collection of big data fits into the overall picture of air quality management and characterization.

The technology around lower cost air sensors is rapidly advancing, as are the deployments of these sensors throughout communities in the United States. As a result, disaggregated, big data sets characterizing local air quality are proliferating. These initiatives—often referred to as "Smart City" or "Internet of Things" (IoT)—are introducing entirely new sets of information. The U.S. Environmental Protection Agency (EPA), states, and tribes have been anticipating this development and published recommendations for advancing the understanding, collection, and use of these data just one year ago.[1] Recently, information technology (IT) companies—and users of their products—have emerged with technologies and tools that integrate air quality data into the larger IoT framework.

Collection of Big, Non-Traditional Datasets

Air quality data generated and collected by a variety of parties, including large IT companies, are growing at a rapid pace. IT companies are promoting the use of the cloud and advanced analytics (e.g., machine learning) to characterize air quality across the globe. They are often driven by the desire to have end users consume products (e.g., smartphone applications), provide data storage and access, or sell analytic services. Concurrently, a wide variety of private sector, academic, non-profit, and government entities are also developing new, integrated air quality sensor systems. These include wearables, urban sensor networks, mobile sensing systems, and application program interfaces. This presents a major shift in the United States, as government agencies have traditionally been the main resources for collecting, storing, sharing, and communicating air data. With a growing diversity of air monitoring participants (e.g., citizens, communities, researchers, businesses) and associated data sets (e.g., sensor, satellite remote-sensing, traditional regulatory networks) there comes a need to make sense of it all.

EPA has funded numerous internal projects, external grants, and challenges to explore the development and application of new sensor technologies.[2-8] Projects led by organizations outside of government are collecting data at an even larger scale. For example, Google street view vehicles repeatedly sampled street emissions in Oakland, California, creating the largest urban air quality data set of its type.[9] The Weather Company, which is owned by IBM, is collaborating with the instrument manufacturer Purple Air to augment its existing personal weather station network with particulate matter sensors.[10] The National Science Foundation is funding The Array of Things (AoT) in Chicago, a network of interactive, modular sensor boxes installed to collect real-time data on the city's environment.[11] The availability of very low-cost sensor equipment and supporting technical components is promoting the development of these systems. However, achieving sufficient quality data and a scientifically-grounded interpretation, especially as it relates to health, is a significant challenge.

Air Quality Management and Characterization Considerations

Collecting and managing data are only a part of characterizing air quality. Complexities are numerous and include understanding how sensor response may vary in environments that differ from the testing environment, ensuring measurements are accurate both initially and over time, having proper data that describes a measurement, relating measurements to health effects or emissions standards, and communicating information in a consistent manner.

To date, the majority of efforts characterizing the performance of air sensor technologies have been primarily conducted by collocation of sensors with reference grade equipment in ambient, stationary environments without nearby pollution sources.[12,13] Long-term (i.e., more than 12 months) performance of these sensors is generally poorly characterized, and some sensors have shown progressive drift with time.[14] These are important considerations when assessing the quality of measurements and use of data analytic approaches. The short-term (i.e., days to months) performance characterization or calibration developed at these locations may not hold when sensors are moved to near-source environments.

For example, particle counters using volume light scattering techniques can be influenced by particle size distribution, chemical composition, shape, and relative humidity. Any significant change from the initial calibration environment (e.g., a burst in large particles emitted by a nearby source) may result in measurement inaccuracies. Topography, varying air pollution mixtures, and near-road measurements can also introduce uncertainty in virtual, network-based corrections.[15,16] Additional research involving the collocation of sensors with reference-grade equipment near sources is needed, especially for localized pollutants such as particulate matter, nitrogen dioxide, carbon monoxide, and sulfur dioxide. Sensors mounted on mobile platforms add additional complexity, as one must consider the representativeness of instantaneous measurements, as well as potential measurement artifacts introduced by mobile monitoring (e.g., vibration or pressure effects). Metadata—data that describe the measurement and its data quality such as precision, bias, range, method detection limit, and calibration—are also critical in determining the end use of air quality information.

In general, the science on air pollution and health doesn't tell us what a few minutes of exposure to an elevated level of pollution will mean for an individual.[17] While air quality sensors currently have the capability to produce data on the order of seconds to minutes, health research has been primarily focused on longer-term exposures to air pollution and the resulting health endpoints. Thus, a shorter-term sensor measurement is not directly comparable to the National Ambient Air Quality Standards (NAAQS) or the related Air Quality Index (AQI) categories, which are based on a substantial body of research supporting health-based standards at longer averaging intervals (e.g., 24-hr $PM_{2.5}$ standard, 8-hr max ozone standard).

As more high time resolution data become available, appropriate context will be needed. AirNow is the current tool used by EPA to promote real-time data exchange and protect public health. It uses quality assured data from long-term ambient monitoring networks to display and send air quality alerts. AirNow presents data in a way that is consistent with the NAAQS and published health evidence using official AQI colors (i.e., green, yellow, orange, red, purple, and maroon) to communicate information to the general public.

Meanwhile, a number of private sector groups have developed a variety of new air quality communication platforms, which are using a myriad of approaches to communicate air quality conditions. While there is a tremendous opportunity to learn from new and more localized datasets collected by these groups, creation of unique air quality communication platforms, visualizations, interpretations, use of the AQI or AQI colors in different ways, and alerts have the potential to confuse the public. Care needs to be taken to ensure air quality information is communicated with a scientifically-grounded approach. Further, understanding how people react to high-resolution temporal data is a subject of needed research.

E-Enterprise Advanced Monitoring Team Updates

As EPA monitors the rapid explosion of sensor networks, EPA, states, and tribes continue to make progress on advanced monitoring priorities through the collaborative E-Enterprise initiative (see Table 1).[18] Specifically discussed in this article are updates on priority initiatives 1, 3, and 4 shown in Table 1.

Since one of the biggest concerns about sensors is the quality of the data generated, EPA, states, and tribes are actively exploring the feasibility of an independent third-party certification program (Priority 1). The certification program concept is unique in that it would be voluntary and based upon defensible data quality objectives for common non-regulatory applications, as well as other justifiable technology requirements

Table 1. E-Enterprise advanced monitoring priorities.

1	Options and Feasibility Analysis for an Independent Third-Party Certification Program
2	Technology Scan, Screen, and User Support Network
3	Data Interpretation
4	Data Exchange Standards
5	Lean the Current Technology Approval Process

supporting good data quality. The team has reached out to standards organizations to determine the level of interest and potential structure in supporting a long-term strategy and program. Recognizing that it may take several years to establish such a program, EPA is planning a workshop in spring 2018 to discuss potential performance targets for air sensors used in non-regulatory applications. The purpose of the workshop will be a comprehensive review and analysis of information related to determining performance targets and test methods that would support an eventual third-party certification program for sensor technologies.

The Priority 3 data interpretation team has identified candidate pilot cities to test new and existing communication and visualization strategies for disparate datasets (e.g., air sensor and regulatory monitoring network data). Strategies for communicating data from pilot cities will explore providing short-term measurements alongside the AQI; applying weighted averages to hourly sensor measurements (i.e., the NowCast algorithm); presenting hourly values for one or more pollutants; giving data additional context, including geographical and meteorological information; and providing appropriate data caveats such as marking data as "raw," "provisional," or "final."

Consistent with the charge of evaluating existing data standards, the data exchange standards team (Priority 4) has completed an analysis and comparison of standards for continuous monitoring results. The team plans to recommend a set of data standards by the end of 2017. Concurrently, the team continues to explore the development of metadata standards, as well as proposed data architecture.

Conclusion

With this rapid onset of new sensor technology and increased availability of new air quality data, it will be important for air quality professionals and data management companies to work together moving forward. Stakeholders in domestic and international governments, academia, commercial interests, and communities will need to collaborate to understand the unique characteristics of air quality measurements, how the data relate to health effects and other factors, and the different data handling and analytic approaches that could be used to "improve" sensor measurements.

To promote engagement, EPA, the California Air Resources Board (CARB), and the University of California, Davis Air Quality Research Center are planning an Air Sensor International Conference (https://sehall4.wixsite.com/asic/home-landing) in September 2018, which will focus on advancing the science and engaging communities. In order to harness the opportunity provided by new data, it will be important to first understand and address some of the challenges. This is an ever changing landscape and will require long-term strategic thinking from all parties involved. **em**

Kristen J. Benedict and **Richard A. Wayland** are both with the Office of Air and Radiation; and **Gayle S.W. Hagler** is with the Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC.

References

1. Hindin, D.; Grumbles, B.; Wyeth, G.; Benedict, K.; Watkins, T.; Aburn Jr., G.; Ulrich, M.; Lang, S.; Poole, K.; Dunn, A.D. Advanced Monitoring Technology: Opportunities and Challenges. A Path Forward for EPA, States, and Tribes; *EM*, November 2016; http://pubs.awma.org/flip/EM-Nov-2016/hindin.pdf.
2. Jiao, W.; Hagler, G.S.W.; Williams, R.W.; Sharpe, R.N.; Weinstock, L.; Rice, J. Field Assessment of the Village Green Project: An Autonomous Community Air Quality Monitoring System; *Environ. Sci. Technol.* **2015**; DOI: 10.1021/acs.est.5b01245.
3. Jiao, W.; Hagler, G.S.W.; Williams, R.W.; Sharpe, R.; Brown, R.; Garver, D.; Judge, R.; Caudill, M.; Rickard, J.; Davis, M.; Weinstock, L.; Zimmer-Dauphinee, S.; Buckley, K. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States; *Atmos. Meas. Tech.* **April 2016**; https://doi.org/10.5194/amt-9-5281-2016.
4. Duvall, R.M.; Long, R.W.; Beaver, M.R.; Kronmiller, K.G.; Wheeler, M.L.; Szykman, J.J. Performance Evaluation and Community Application of Low-Cost Sensors for Ozone and Nitrogen Dioxide; *Sensors* **2016**; DOI:10.3390/s16101698.
5. EPA Air Pollution Monitoring for Communities Grants. See https://www.epa.gov/air-research/air-pollution-monitoring-communities-grants.
6. EPA's Smart City Air Challenge. See https://developer.epa.gov/smart-city-air-challenge-resource-pages/.
7. EPA Air Sensors Toolbox. See https://www.epa.gov/air-sensor-toolbox/air-sensor-toolbox-what-epa-doing#pane-1.
8. Wildland Fire Sensors Challenge. See https://www.challenge.gov/challenge/wildland-fire-sensors-challenge/.
9. Apte, J.S.; Messier, K.P.; Gani, S.; Brauer, M.; Kirchstetter, T.W.; Lunden, M.M.; Marshall, J.D.; Portier, C.J.; Vermeulen, R.C.H.; Hamburg, S.P. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data; *Environ. Sci. Technol.* **May 2017**; DOI: 10.1021/acs.est.7b0089.
10. The Weather Company Collaborates with PurpleAir to Provide Community Air Quality Data Across Its Consumer Properties. See http://www.theweather-company.com/newsroom/2017/07/19/weather-company-collaborates-purpleair-provide-community-air-quality-data-across.
11. Array of Things. See https://arrayofthings.github.io.
12. Air Quality Sensor Performance Evaluation Center (AQ-SPEC). See http://www.aqmd.gov/aq-spec.
13. EPA Office of Research and Development Evaluation of Emerging Air Pollution Sensor Performance. See https://www.epa.gov/air-sensor-toolbox/evaluation-emerging-air-pollution-sensor-performance.
14. Jiao, W.; Hagler, G.S.W.; Williams, R.W.; Sharpe, R.; Brown, R.; Garver, D.; Judge, R.; Caudill, M.; Rickard, J.; Davis, M.; Weinstock, L.; Zimmer-Dauphinee, S.; Buckley, K. Community Air Sensor Network (CAIRSENSE) project: Evaluation of low-cost sensor performance in a suburban environment in the southeastern United States; *Atmos. Meas. Tech.* **April 2016**; https://doi.org/10.5194/amt-9-5281-2016.
15. Karner, A.A.; Eisinger, D.S.; Niemeier, D.A. and references therein, Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data; *Environ. Sci. Technol.* **June 2010**; DOI: 10.1021/es100008x.
16. Baldauf, R.W.; Heist, D.; Isakov, V.; Perry, S.; Hagler, G.W.S; Kimbrough, S.; Shores, R.; Black, R.; Brixey, L. Air Quality Variability Near a Highway in Complex Urban Environment; *ScienceDirect* **January 2013**; https://doi.org/10.1016/j.atmosenv.2012.09.054.
17. Keating, M.; Benedict, K.; Evans, R.; Jenkins, S,; Mannshardt, E.; Stone, S.L. Interpreting and Communicating Short-Term Air Sensor Data; *EM* November 2016; http://pubs.awma.org/flip/EM-Nov-2016/keating.pdf.
18. *Note:* The progress on E-Enterprise advanced monitoring priorities described in this article is part of EPA's broader effort to meet the opportunities and challenges posed by advanced monitoring technologies as described in an article in the November 2016 issue of *EM*. See "Advanced Monitoring Technology: Opportunities and Challenges" by Hindin et al.; http://pubs.awma.org/flip/EM-Nov-2016/hindin.pdf.