Environmental Science & Technology

Viewpoint

Cite This: Environ. Sci. Technol. XXXX, XXX, XXX−XXX

pubs.acs.org/est

Air Quality Sensors and Data Adjustment Algorithms: When Is It No Longer a Measurement?

Gayle S. W. Hagler,*[,†] Ronald Williams,[†] Vasileios Papapostolou,[‡] and Andrea Polidori[‡]

[†]United States Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Research Triangle Park, North Carolina 27709, United States

[‡]South Coast Air Quality Management District, Diamond Bar, California 91765, United States

Sensor technology to measure outdoor air pollution is becoming ubiquitous. Sensors are currently developed and deployed by a wide number of start-up technology companies, academic institutions, government organizations, community groups, traditional air quality instrument manufacturers, and other commercial entities.[1] Developers seek to maximize the quality and quantity of information from sensor technologies, while minimizing the cost to build and maintain. The original equipment manufacturer (OEM) sensor components used for detection of atmospheric gases and particles generally trade off measurement selectivity, sensitivity, and reproducibility for miniaturization, power, and price. Additionally, performance targets for OEM sensors or integrated sensor devices are not currently established. Air quality sensors, therefore, have a variety of known measurement artifacts that those developing and applying the technology seek to overcome.

A growing trend in air sensor applications is to improve the data quality from sensors through applying multiple linear regression,[2,3] machine learning,[2] or other complex mathematical algorithms.[4] To develop a data adjustment method, the sensor device is usually collocated with a reference-grade monitor in an environment that is representative of the sampling conditions. This collocation time frame serves as the training period for which a correction algorithm is developed that incorporates the sensor raw data and adjusts the data to most closely match the reference-grade data. Thereafter, the sensor device is relocated to another environment for ongoing use and the correction algorithm is applied, based upon the presumption that the ongoing sampling conditions are within range of the calibration period. In some approaches, sensor data at one location are adjusted based upon measurements in other places, assuming there is homogeneity in air pollution concentrations over a specific geographic area and time frame;[5] for example, this approach appears to be supported via commercially available software (e.g., Advanced Normalization Tool for AirVision; http://agilaire.com/pdfs/ANT.pdf). These emerging strategies raise a number of questions for debate, such as: How confident are we in the approach of calibrating sensors at one location for a short period of time, then deploying at other locations under potentially differing conditions and for longer time spans? What are the appropriate parameters to include in sensor data postprocessing? At what point do sensor data depart the identity of an independent measurement, but are now considered a model output to some degree, and does this distinction matter?

A measurement purist would argue that the only parameters that are appropriate for inclusion into a sensor data adjustment algorithm are those that are definitively proven to cause measurement response error or bias. For example, optical particle sensors often display artifacts under increasing humidity. This effect is due to the condensation of water to the particles, altering their light-scattering properties and introducing inaccuracy in the estimated particulate matter mass concentrations. Optical particle sensors also have lower particle size limits for their detection capability (e.g., 300 nm). Numerous gas-phase sensors have known cross-sensitivities, whereby an electrochemical or metal oxide sensor that is identified as sensing a specific gas may also have some degree of responsiveness to another gas type. Complicating this further, gas sensors may also have measurement artifacts related to temperature and humidity. Finally, some low-cost sensors drift in their measurement response over time.[3] These complex factors collectively create a multidimensional problem, for which a variety of groups attempt to solve through sophisticated data postprocessing.

A critical issue for debate in the scientific community is the appropriate design of sensor postprocessing algorithms. Of chief concern are the inclusion of parameters for which there is no demonstrated measurement artifact or rely upon untested assumptions about the state of the atmosphere. In the era of big data, it is tempting to maximize the ability of sensors to

Received: April 6, 2018

DOI: 10.1021/acs.est.8b01826
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Table 1. Sensor Data Adjustment Parameters: Defendable and Questionable

defendable parameters	questionable parameters
• relative humidity, for which measurement artifact has been established	• wind speed or direction
• temperature, for which measurement artifact has been established	• gases for which no cross-sensitivity is indicated
• other gases for which cross-sensitivity has been established	• data from neighboring monitors (reference-grade or sensor) that are not proven as suitable reference point[a]
• elapsed time since manufactured or deployed, if aging has been demonstrated to cause change in sensor response.	• local emission information or surrogates for emissions (e.g., traffic patterns, population density)
• accessory measurements indicating aerosol refractive index, for pm sensors	• temporal factors other than elapsed time of use (e.g., time of day, day of week)
• autozero data, if equipped to self-zero	• atmospheric mixing height
• monitors in close proximity, if established to have comparable data under specific conditions[a]	• location relative to sources (e.g., proximity to a road)

[a]This is a subject of needed research and likely location-specific, as well as pollutant-specific.

reproduce reference monitor data or to produce trends meeting predetermined expectations, and meet this goal through introducing questionable parameters into data processing approaches (Table 1). These parameters build assumptions into the processed air sensor data that can introduce error and lose the integrity of the data as "ground truth". For example, a machine learning algorithm for one air pollutant, incorporating another measured pollutant's values for which there is no established cross-sensitivity, has now arguably created an empirically modeled value. As another example, network-based approaches that incorporate reported values of neighboring reference or sensor monitors may also introduce errors to the data, particularly for pollutants with high spatiotemporal variability.

The important question is, does it matter? High quality air measurement data are commonly used to ground-truth predictive air quality models, serve as comparison data for satellite remote sensing data, determine impacts of source emissions, communicate air quality conditions to the public and as inputs for epidemiological studies. If sensors are to be used for similar purposes, the incorporation of questionable parameters leads to a significant data integrity issue and undermines the usability of the data. For these and other uses of air quality data, it is essential that adjustments to raw sensor data avoid becoming a predictive model. Transparency is essential to build trust in air sensor data, which is a challenging issue for many sensor developers where algorithms applied are valuable intellectual property. This limitation may be overcome through the provision of unprocessed, original sensor data output, allowing scientists to develop and openly document independent algorithms. Secondarily, trust in the processed data would be increased if developers share which parameters are incorporated in postprocessing, communicate when algorithms are updated, and show the comparison of unadjusted and adjusted data.

As air sensor technology expands globally, research informing best practices in air sensor application and data processing is critical. While secondary data products, such as estimated air pollution exposure surfaces, are highly valuable and may assimilate a wide variety of information, it is essential to maintain original observational data that represents actual conditions. The envisioned bright future of widely available air sensor technology hinges on the integrity of the data.

AUTHOR INFORMATION

Corresponding Author

*E-mail: hagler.gayle@epa.gov.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Snyder, E. G.; Watkins, T. H.; Solomon, P. A.; Thoma, E. D.; Williams, R. W.; Hagler, G. S. W.; Shelow, D.; Hindin, D. A.; Kilaru, V. J.; Preuss, P. W. The changing paradigm of air pollution monitoring,. *Environ. Sci. Technol.* **2013**, *47*, 11369−77.

(2) Zimmerman, N.; Presto, A.; Kumar, S.; Gu, J.; Hauryliuk, A.; Robinson, E.; Robinson, A.; Subramanian, R. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech.* **2018**, *11*, 291−313.

(3) Jiao, W.; Hagler, G.; Williams, R.; 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.* **2016**, *9*, 5281−5292.

(4) Cross, E. S.; Williams, L. R.; Lewis, D. K.; Magoon, G. R.; Onasch, T. B.; Kaminsky, M. L.; Worsnop, D. R.; Jayne, J. T. Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. *Atmos. Meas. Tech.* **2017**, *10*, 3575−3588.

(5) Moltchanov, S.; Levy, I.; Etzion, Y.; Lerner, U.; Broday, D. M.; Fishbain, B. On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Sci. Total Environ.* **2015**, *502*, 537−547.

DOI: 10.1021/acs.est.8b01826
Environ. Sci. Technol. XXXX, XXX, XXX−XXX