MOMA

A Virtual Calibration Method for Air Quality Sensors

Geoff Henshaw
& Lena Weissert

March 2022

aeroqual™

Abstract

Aeroqual has developed a virtual, continuous calibration technique for air quality sensor networks. It is applicable to any air quality network that contains multiple tiers of instruments such as regulatory instruments, mid-tier or near-reference monitors and sensors.

The technique derives slope and offset estimates by matching the mean and variance of the time-averaged distribution of recent sensor data to values derived from specific reference monitors ("proxies").

Proximity and land use similarity form the basis for selecting these proxies. The method is sensor agnostic and is reliable for urban, rural, and industrial air quality sensor networks from a variety of vendors.

Introduction

Traditional regulatory air quality measurements rely on expensive certified instruments and frequent, rigorous site calibration. Along with the capital cost of establishing a regulatory site, there is also a significant and ongoing operational cost in the maintenance and calibration of site instruments.

Over the past few years, air quality measurement using sensors in distributed networks are being used to provide information to professionals and citizens at a local scale. Studies demonstrate the success of such networks to identify local pollutant patterns not evident in the more sparsely distributed regulatory networks[1-6].

With the rapid uptake of this technology, international standards organizations and Government agencies have developed, or are developing, performance evaluation protocols to build confidence in the “out of the box” data quality from such devices[7-9].

Ongoing operation of sensor networks can be challenging, especially with respect to calibration and data validation. As with all instrumentation, there is the potential for response drift, adversely impacting measurement accuracy.

Introduction

Continued

However, it is impractical and cost-prohibitive to undertake an individual site calibration of sensors in a distributed network. Less intensive, virtual calibration methods are needed that are well-defined and transparent so that users of the data understand its quality and make reliable decisions based upon it.

Calibration by co-location prior to and post-deployment is a common method for small sensor networks[10]. This process is also resource intensive, resulting in data loss and raising questions about whether the calibration is transferable from the co-location site.

Calibration using a mobile transfer standard is also possible but is limited by the vagaries of meteorological and pollutant conditions. This can make it difficult to ensure adequate data is obtained to establish reliable slope and bias estimates.

It is important to distinguish between “correction” and “calibration”. The term “calibration” means an adjustment of a sensor’s pollutant reading via changes in the slope and bias of the reading. The term “correction” applies to the use of algorithms that combine multiple inputs to derive a pollutant concentration.

Introduction

Continued

Such algorithms, which include statistical methods such as multiple linear regression[11] and machine learning methods such as neural networks[12], are optimized (trained) by comparing their outputs to regulatory station data. If the algorithm is regularly updated against regulatory station data, it may account for changes in sensor drift and performing the dual functions of correction and calibration. However, these correction models[13] can often require large training sets, high computation costs, and still only be applicable over limited geographical and seasonal domains.

Studies have also demonstrated that variability between sensors can be high, requiring device-specific corrections[14].

There is also the question of what variables are appropriate to include in an algorithm. Air quality agency researchers[15] have warned against the use of “questionable” parameters such as wind speed, proximity to roads, or time of day in sensor correction algorithms.

Sensor vendors are starting to offer some version of a field calibration. One vendor[16] proposes a co-location of all sensors for at least two weeks to derive the sensors’ humidity and temperature response prior to deployment.

Introduction

Continued

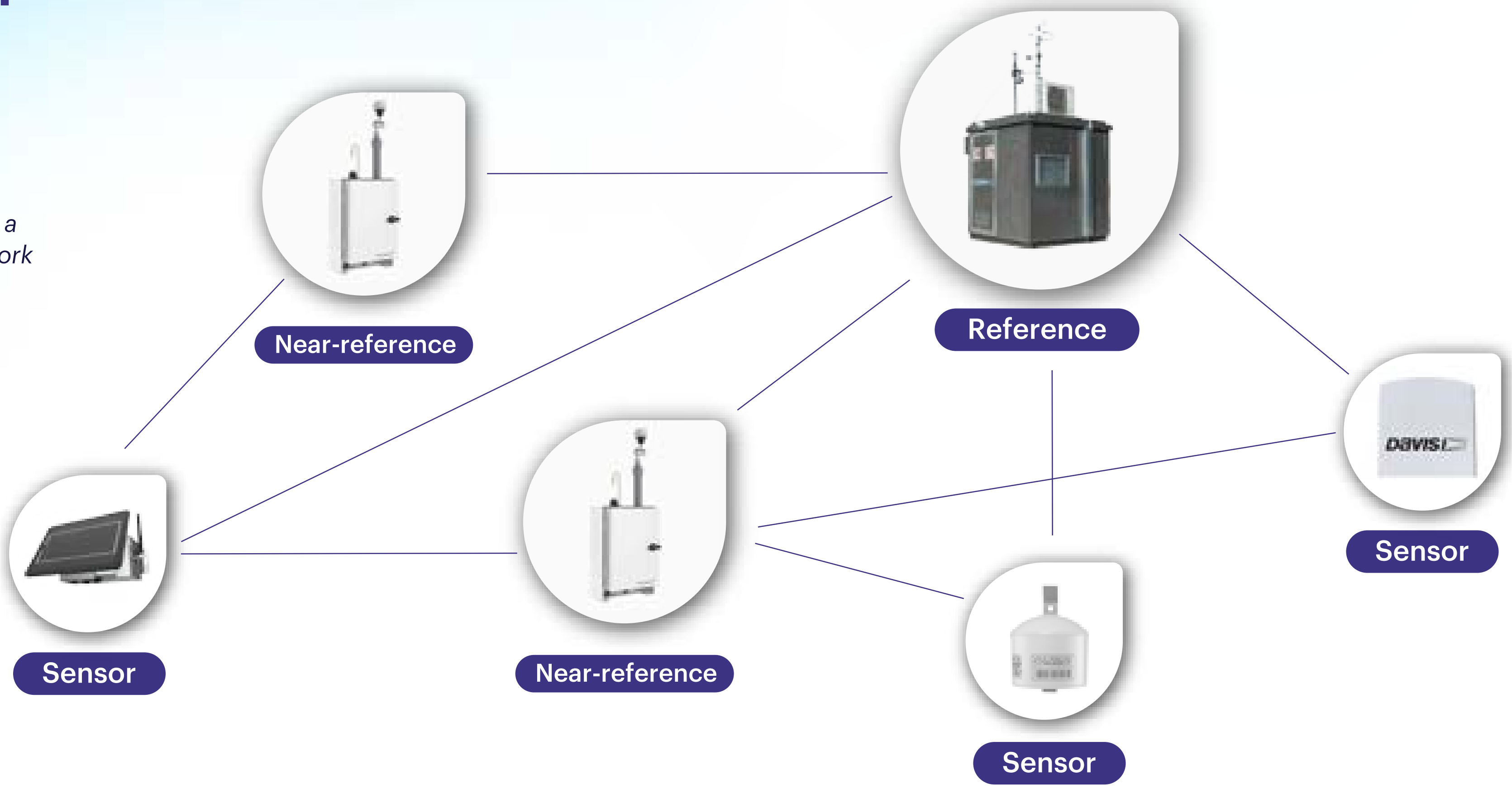
This is uploaded and the network is deployed, bar one “canary” air quality sensor which stays behind at the colocation site. Periodic virtual adjustment of the network then occurs, accounting for seasonal and environmental changes (though there is typically no description of how this occurs). The lack of transparency makes it difficult to evaluate the validity of the calibration process.

To address these issues, Aeroqual, in conjunction with the University of Auckland, has developed a continuous method for virtual calibration of air quality sensors that is transparent and does not require a co-location step. This method uses the idea of a mixed-asset network, in which a small number of well-calibrated reference monitors act as ‘proxies’ to which other air quality sensors are referred.

Introduction

Continued

Figure 1: Network architecture for a MOMA calibrated air quality network



Theory

The MOMA calibration method assumes linearity of the sensor response. This is not a significant constraint since sensor linearity is widely demonstrated for air quality sensors, at least over ambient pollutant concentration ranges. If $X_i(t)$ denotes the true data value at location *i* and time *t*, then the sensor data, $Y_i(t)$ will satisfy:

$$X_i(t)=a_0+a_1Y_i(t)$$

We assume the frequency distribution of sensor and proxy data have a similar functional form over a given time interval. For example, distributions might be lognormal or be similar around the mean but differ in the extremes.

Let $Z_i(t)$ denote the reference (proxy) data and $Y_i(t)$ denote the sensor data for site i over the time interval *(t)*. Then estimates for the slope a_1 and offset a_0 for corrected sensor data are derived as:

$$\hat{a}_1=\sqrt{\frac{var[Z(t)]}{var[Y(t)]}}$$

$$\hat{a}_0=E[Z(t)]-\hat{a}_1E[Y(t)]$$

Theory

Continued

Where E< > denotes the mean and var< > denotes the variance. Then, the estimate of the true data at site *i* at time *t*, $X_{i,t}$ from the sensor data is given by:

$$X_{i,t}=a_0+a_1Y_{i,t}$$

This approach differs from linear regression in that it doesn't rely on a pairwise correlation. Instead, it matches the probability distributions of the sensor and proxy data. The time interval is chosen to obtain a strong estimate of the proxy and sensor distributions and to emphasize the longer-term, regular component of the concentration variation within the distribution of values.

A key step in the method is the mapping of each of the air quality sensors to a suitable proxy reference monitor.

We studied a variety of proxy selection methods and have shown that the optimum selection method depends on the pollutant and the nature of the emission sources in the airshed. If the pollutant is largely regional in nature (for example, ozone), the most suitable proxy has proven to be the nearest. This is unless there is significant titration from nitrogen monoxide (NO) emissions at the sensor site.

Theory

Continued

For pollutants with a dispersed set of localized sources within an airshed, such as nitrogen dioxide (NO_2) the proxy is chosen based on land-use similarities between the sensor and proxy. Land use has proven to be capable of explaining a significant fraction of the total variance of concentration across a region[17] and therefore is a reliable method for finding sites with similar pollutant patterns. Land use variables are selected based on a systematic literature review on land-use regression (LUR) models.

MOMA Performance

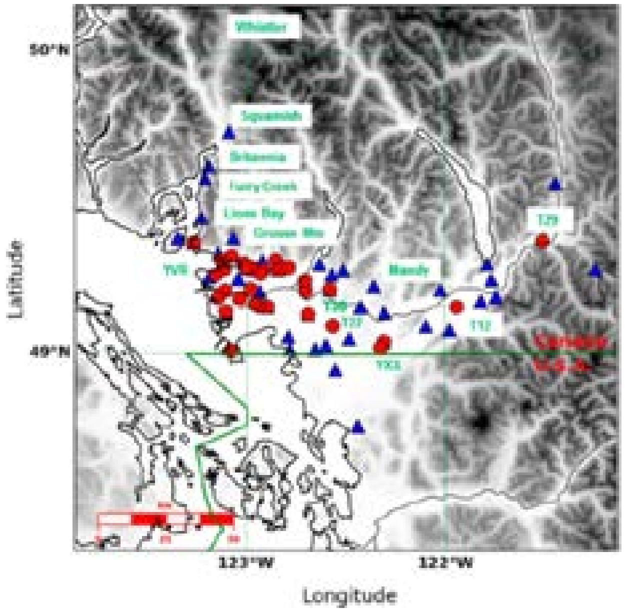
We have seen the performance of the MOMA method evaluated in several published field studies[18-24]. The test approach first involved applying the MOMA method to a co-located sensor using a distant proxy. Corrected sensor data was then compared to that of the co-located regulatory station. The examples that follow demonstrate the effectiveness of the method for O_3, NO_2 and $PM_{2.5}$. The method can also readily extend to other pollutants and vendor devices, provided suitable proxies exist in the measurement region.

MOMA Performance

Example 1

Ozone measurement around the Lower Frazer Valley of British Columbia

Fifty solar-powered ozone sensors were deployed throughout the Lower Frazer Valley of British Columbia over summer, May to September 2012, to study ozone formation, transport and distribution (20).



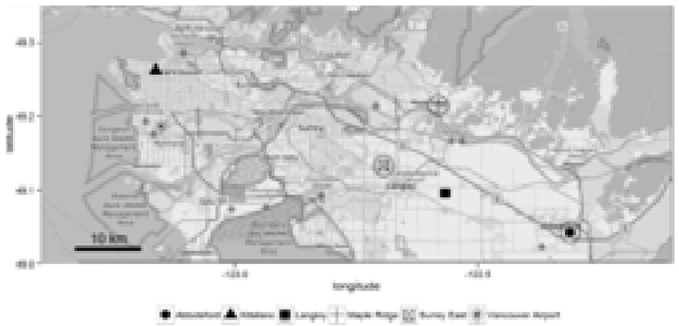
MOMA Performance

Example 1

To derive scientific insights from the data, the MOMA calibration technique was applied post-deployment. Ozone sensor drift was evident at several sites but was corrected by MOMA to yield useful data.

The validity of the MOMA process was established by comparing the collocated raw sensor and MOMA-corrected data (based on a distant proxy) to the regulatory site data. An example of the performance is shown in Figure 2 for the sensor located at the Langley regulatory site, using the Maple Ridge regulatory station as a proxy.

The uncorrected data shows significant drift over time, while the MOMA calibrated data set maintains a high correlation coefficient. The mean absolute error (MAE) for the raw data was 6.4 ppb, whereas the MAE dropped to 3 ppb for the MOMA-corrected data set over the study period.



MOMA Performance

Example 1

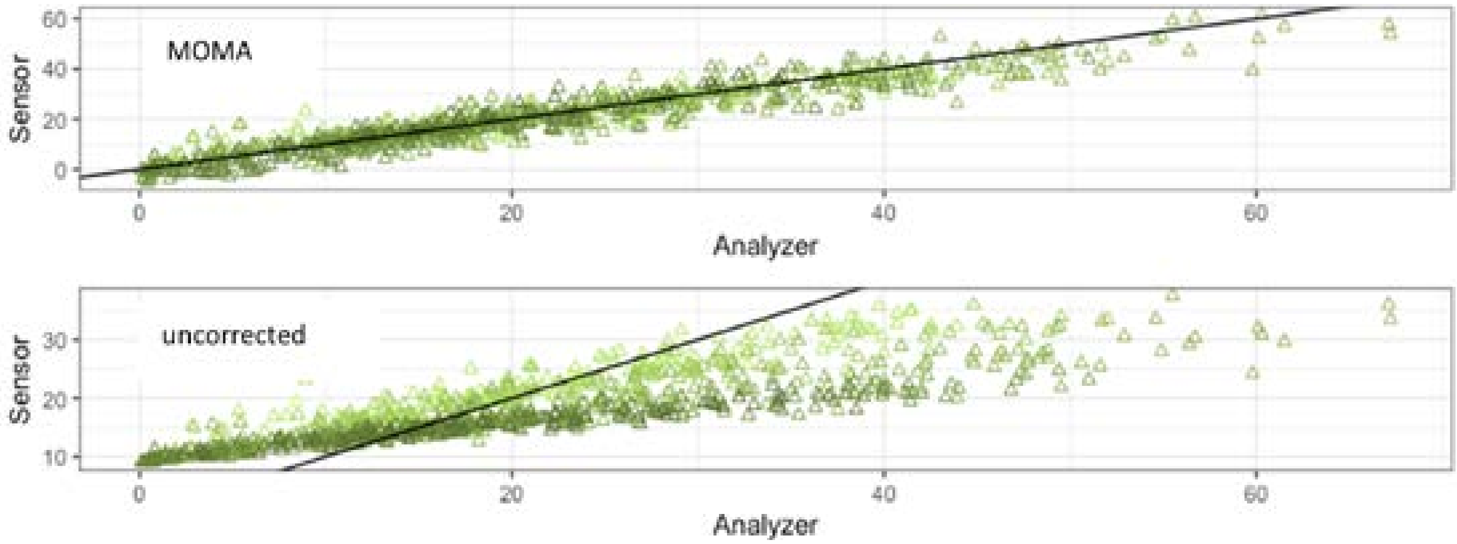


Figure 2: Two-month scatter plots of hourly averaged sensor versus regulatory analyzer data at the Langley site, both uncorrected and MOMA-corrected.

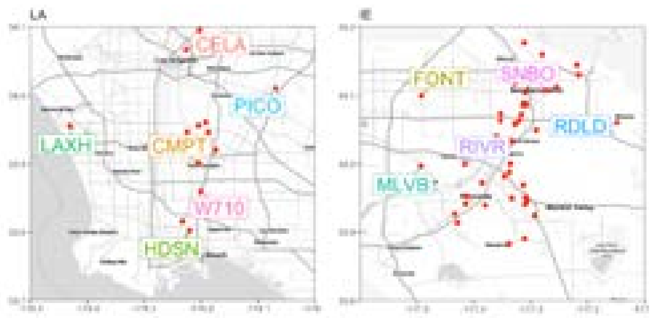
MOMA Performance

Example 2

NO_2 measurements across the Los Angeles region

Two Aeroqual AQY sensor networks were established in the Los Angeles region(23). One consisted of 20 units around the city of Los Angeles, with the other consisting of 45 units in the Inland Empire region. In each local network, five AQY NO2 sensors were collocated with one of the regulatory analyzer stations (designated by a 4-letter code) used for validation of the framework. A distant regulatory station served as the proxy for the AQY sensor.

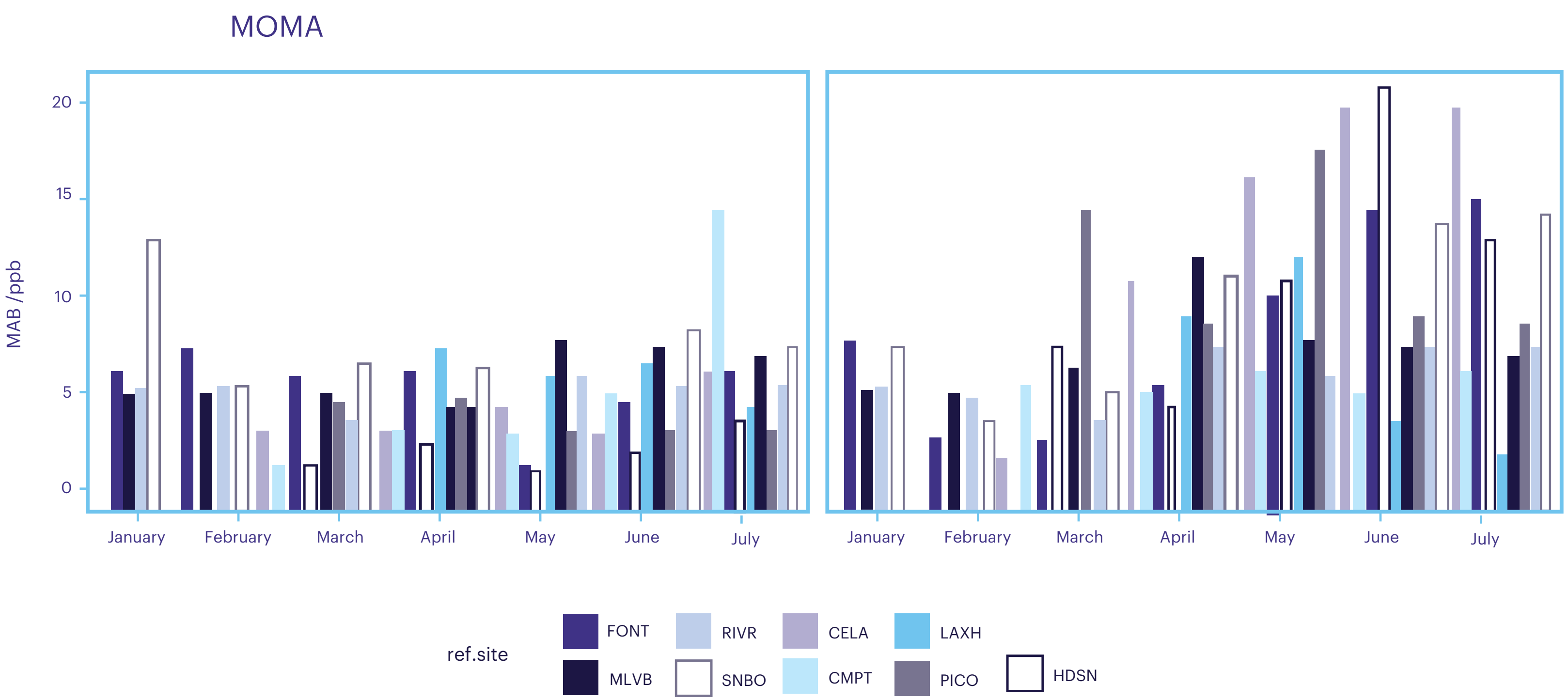
The effect of MOMA (Figure 3) was to reduce long-term sensor drift and improve accuracy, particularly after 4 months of deployment. Analysis of spatiotemporal O3/NO2 patterns further demonstrated that MOMA can be extended to a larger sensor network without losing information about local effects24.



MOMA Performance

Example 2

Figure 3: Mean absolute bias (MAB) values for the MOMA and uncorrected NO2 sensor data over a 6-month period



MOMA Performance

Example 3

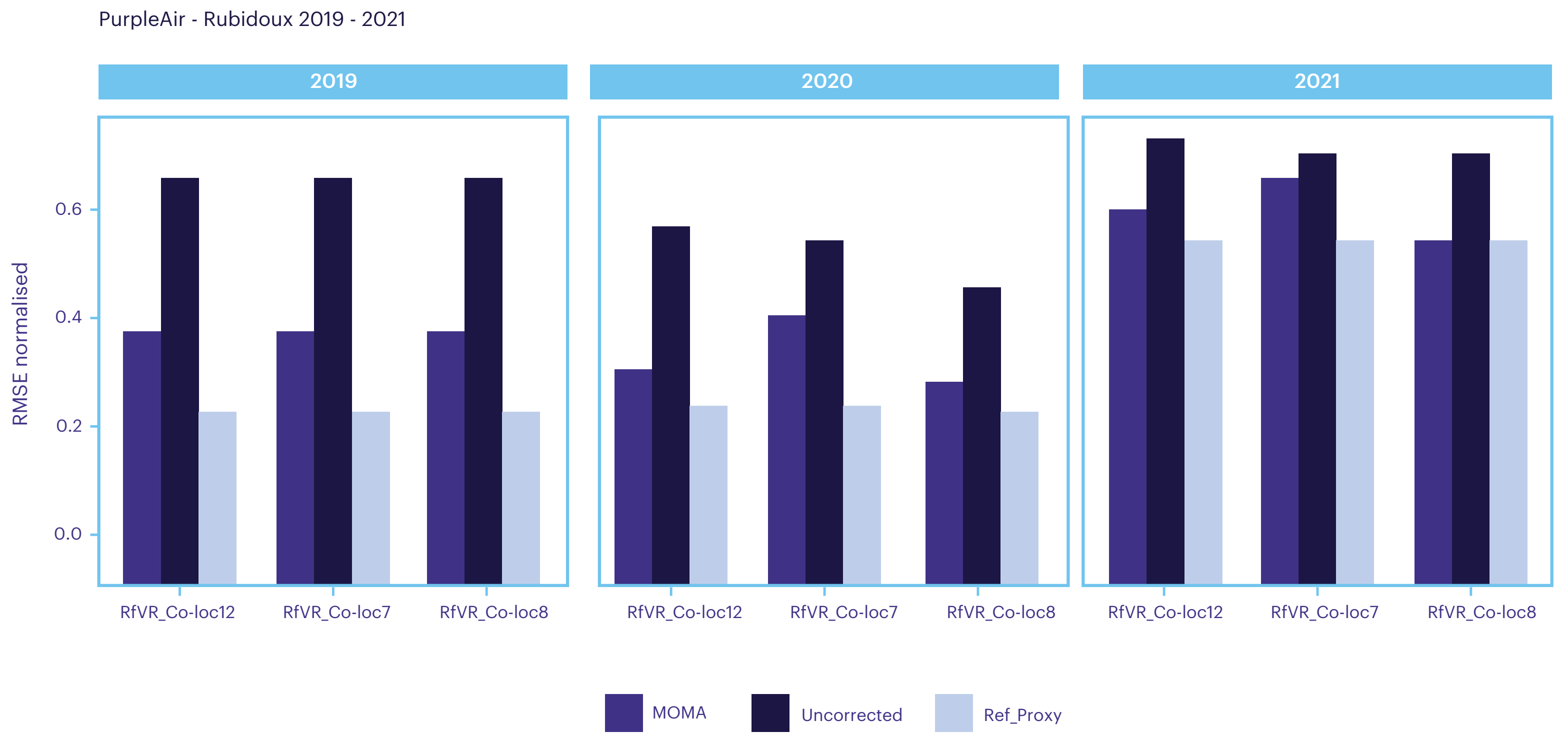
PurpleAir $PM_{2.5}$ measurement

PurpleAir (PA) sensors have been studied by various researchers and shown to exhibit significant humidity and accuracy issues without field calibration. Several PA sensors were collocated at the Rubidoux regulatory station managed by South Coast Air Quality Management District (SC AQMD) for a multi-year period. The impact of undertaking an automated MOMA calibration process over a 3-year period is found in Figure 4. This figure shows the yearly averaged, daily normalized RMSE error (nRMSE) for the uncorrected PA data, MOMA-calibrated data (using a distant proxy station) and the proxy against the Rubidoux Ref for the period 2019-2021. In short, the MOMA calibration considerably improved the PurpleAir data quality over each of the three years. In most cases the nRMSE between sensor and reference is similar to the nRMSE between proxy and reference.

MOMA Performance

Example 3

Figure 4: The yearly averaged, daily normalized RMSE error for the uncorrected PA data, MOMA calibrated and the proxy versus the Rubidoux Station data.



Conclusions

The MOMA calibration method is a transparent and resource-light virtual calibration technique for a network of air quality sensors. The technique functions by leveraging a network of regulatory and/or managed, near-reference level monitors. The regulatory instruments act as proxy references for the denser network of lower-cost sensors. A network of managed monitors provide a means of extending MOMA calibration coverage beyond the regulatory network.

Such a network can be operated with an optimum tradeoff between data accuracy and operating cost. The underlying theory indicates that the method can be successfully applied to a wide range of sensor types and vendors.

References:

1. Bart, M., Williams, D., Ainslie, B., McKendry, I., Salmond, J., Grange, S., ... Henshaw, G. (2014). High density ozone monitoring using gas sensitive semi-conductor sensors in the Lower Fraser Valley, British Columbia. Environmental Science & Technology, 48(7), 3970-3977. doi:10.1021/es404610t

2. Brienza, S., Galli, A., Anastasi, G., & Bruschi, P. (2015). A low-cost sensing system for cooperative air quality monitoring in urban areas. Sensors (Switzerland), 15(6), 12242-12259. doi:10.3390/s150612242

3. Capezzuto, I., Abbamonte, L., De Vito, S., Fattoruso, G., Massera, E., Buonanno, A., ... Di Francia, G. (2015). An app based air quality social sensing system built on open source Hw/Sw tools doi:10.1007/978-3-319-09617-9_54

4. Cheadle, L., Deanes, L., Sadighi, K., Casey, J., Collier-Oxandale, A., & Hannigan, M. (2017). Quantifying neighborhood-scale spatial variations of ozone at open space and urban sites in Boulder, Colorado using low-cost sensor technology. Sensors (Switzerland), 17(9) doi:10.3390/s17092072

5. Cross, E., Williams, L., Lewis, D., Magoon, G., Onasch, T., Kaminsky, M., ... Jayne, J. (2017). Use of electrochemical sensors for measurement of air pollution: Correcting interference response and validating measurements. Atmospheric Measurement Techniques, 10(9), 3575-3588. doi:10.5194/amt-10-3575-2017

6. Weissert, L., Salmond, J., Miskell, G., Alavi-Shoshtari, M., Grange, S., Henshaw, G., & Williams, D. (2017). Use of a dense monitoring network of low-cost instruments to observe local changes in the diurnal ozone cycles as marine air passes over a geographically isolated urban centre. Science of the Total Environment, 575, 67-78. doi:10.1016/j.scitotenv.2016.09.229

7. AQ-SPEC program operated by South Coast AQMD (http://www.aqmd.gov/aq-spec) .

8. US Environmental Protection Agency, Air Sensor Performance Targets and Testing Protocols, (https://www.epa.gov/air-sensor-toolbox/air-sensor-performance-targets-and-testing-protocols) retrieved 14 January 2022.

References:

Continued

9. Ineris and LNE certification protocol for air quality sensors based on the CEN /TC264/WG42 normative group's work (https://prestations.ineris.fr/en/certification/certification-sensors-system-air-quality-monitoring) retrieved January 14 2022

10. Sadighi, K., Coffey, E., Polidori, A., Feenstra, B., Lv, Q., Henze, D.K., Hannigan, M., 2018.

11. Intra-urban spatial variability of surface ozone in Riverside, CA: Viability and validation of low-cost sensors. Atmospheric Measurement Techniques 11, 1777-1792.

12. Malings, C., et al., 2019. Fine particle mass monitoring with low-cost sensors: corrections and long-term performance evaluation. Aerosol. Sci. Technol. 54 (2), 160-174

13. Si, M., Xiong, Y., Du, S., Du, K., 2020. Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods. Atmos. Meas. Tech. 13 (4), 1693-1707.

14. Considine, E., Reid, C., Olgetree, M., Dye, T. 2021 Improving accuracy of air pollution exposure measurements: Statistical correction of a municipal low-cost airbourne particulate matter sensor network., Environmental Pollution 268, 115833

15. M. A. Zaidan et al., Intelligent Calibration and Virtual Sensing for Integrated Low-Cost Air Quality Sensors, (2020), IEEE Sensors, 20 (22), pp. 13638-13652

16. Hagler et al., 2018, Air quality sensors and data adjustment algorithms: when is it No longer a measurement? Environ. Sci. Technol., 52 (10) 5530–31

17. https://www.clarity.io/landing-pages/download-guide-to-accurate-particulate-matter-measurements-with-usepa-performance-targets#download-guide-form accessed 23 December 2021.

18. Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regresssion models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42, 7561–7578.

References:

Continued

19. Miskell, G.; Salmond, J.; Alavi-Shoshtari, M.; Bart, M.; Ainslie, B.; Grange, S.; McKendry, I. G.; Henshaw, G. S.; Williams, D. E. Data Verification Tools for Minimizing Management Costs of Dense Air-Quality Monitoring Networks. Environmental Science & Technology 2016, 50, 835-846.

20. Alavi-Shoshtari, M.; Salmond, J. A.; Giurcaneanu, C. D.; Miskell, G.; Weissert, L.; Williams, D. E. Automated data scanning for dense networks of low-cost air quality instruments: Detection and differentiation of instrumental error and local to regional scale environmental abnormalities. Environmental Modelling & Software 2018, 101, 34-50.

21. Miskell, G.; Salmond, J. A.; Williams, D. E. Solution to the Problem of Calibration of Low-Cost Air Quality Measurement Sensors in Networks. ACS Sensors 2018, 3, 832-843.

22. Weissert, L.F., Alberti, K., Miskell, G., Pattinson, W., Salmond, J.A., Henshaw, G., Williams, D.E., 2019. Low-cost sensors and microscale land use regression: Data fusion to resolve air quality variations with high spatial and temporal resolution. Atmospheric Environment 213, 285-295. 10.1016/j.atmosenv.2019.06.019

23. Miskell, G., Alberti, K., Feenstra, B., Henshaw, G., Papapostolou, V., Patel, H., Polidori, A., Salmond, J.A., Weissert, L.F., Williams, D.E., 2019. Reliable data from low-cost ozone sensors in a hierarchical network, Atmospheric Environment 214, 116870.

24. Lena Weissert, Elaine Miles ,Georgia Miskell, Kyle Alberti, Brandon Feenstra, Geoff S Henshaw, Vasileios Papapostolou, Hamesh Patel, Andrea Polidori, Jennifer A Salmond, David E Williams, Hierarchical network design for nitrogen dioxide measurement in urban environments, Atmospheric Environment 228 (2020) 117428.

25. Weissert, LF, Alberti, K., Miles, E., Miskell, G., Feenstra, B., Henshaw, G. Papapostolou, V., Patel, H., Polidori, A., Salmond, J.A., Williams, D.E., 2020. Low-cost sensor networks and land-use regression: Interpolating nitrogen dioxide concentration at high temporal and spatial resolution in Southern California, Atmospheric Environment 223, 2020, 117287.