Calibration of Atmospheric Measurements in Low-cost Sensors

Pak Lun Fung
pak.fung@helsinki.fi

KEYWORDS

calibration, atmospheric measurements, linear regression, multiple linear regression, neural networks, low-cost sensors

1 INTRODUCTION

Urban air pollution is a major concern in modern cities and developing countries. Ambient pollutants considerably affect human health, especially causing a variety of respiratory diseases, such as asthma and lung cancer. Additionally, air pollution is responsible for environmental problem, for instance acid rain and the depletion of the ozone layer. As a result, air pollution monitoring is of paramount importance.

Nowadays, urban air pollution is monitored by networks of static measurement stations (hereinafter called reference stations). Reference stations used today has been demonstrated to meet the data quality and trace-ability requirements of international programmes such as World Meteorological Organization (WMO)/ Global Atmosphere Watch (GAW), and therefore are highly reliable. They accurately measure a wide range of air pollutants using traditional analytical instruments, such as mass spectrometers and gas chromatographs. One example in Finland is Stations Measuring Earth Surfaces and Atmosphere Relations (SMEAR).

The disadvantages of these complex measurement systems are their large size, high price, and complex maintenance. The extensive cost of acquiring and operating these stations severely limits the number of installations. Therefore, it is necessary to deploy the massive use of low-cost sensors, where they are usually defined as an initial capital cost reduction of at least one order of magnitude over reference instruments, to increase the measurement coverage, and hence to better understand the current situation. The data quality of low-cost sensors is highly variable and real-word performance can vary very much due to different data correction and calibration approach [Lewis et al. 2018].

2 LOW-COST SENSORS AND CURRENT CHALLENGES

The gaseous air pollutants that are typically measured using sensors include nitrogen monoxide (NO), nitrogen dioxide (NO2), ozone (O3), sulphur dioxide (SO2), carbon monoxide (CO), and total volatile organic compounds (VOCs). Aerosols such as particulate matter (PM) are also measured together with other gaseous air pollutants. They are important due to their direct and indirect adverse health and ecosystem effects and their role as precursors to form secondary pollutants.

Measurements of the gaseous pollutants are typically reported either as a mixing ratio (e.g. ppm or ppb), or in mass concentration units (e.g. ug/m3). It is also relevant to note that sensor performance, e.g. sensitivity and measurement error might be different not only between sensors but also between pollutants measured by the same sensor. Below shows a list of types of sensors and Figure 1 shows a typical low-cost sensor[Spinelle et al. 2013].

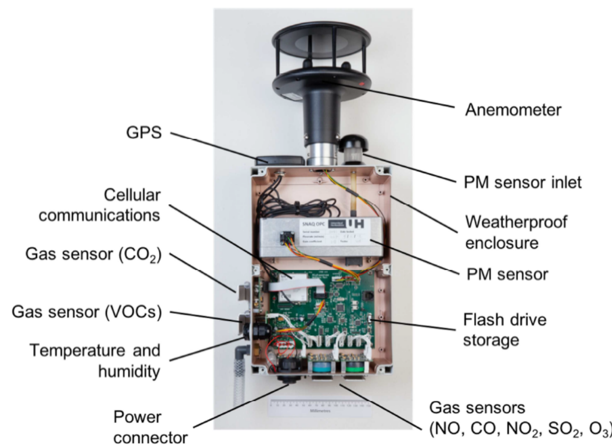

Figure 1: A typical low-cost sensor gadget measuring different parameters at one time [Lewis et al. 2018].

- `Electrochemical (EC) sensors`: They have interferences with relative humidity and temperature, requiring additional measurements to be made in order to obtain reliable results.
- `Metal-oxide semiconductor (MOS) sensors`: This relationship is in general non-linear in nature and these sensors have some sensitivity to changing environmental conditions, and interferences from other gases that may be present.
- `Miniature photo-ionization detectors (PIDs)`: Only some compounds are efficiently ionized (and detected) while other compounds are less efficiently ionized (and less efficiently detected).

Apart from the sensor-specific problems listed above, problems like short-term and long-term drifting, cross-sensitivity between NO2 and O3, size-detection limit for ultrafine particles, too low detection limit for SO2, particle hygroscopicity effect for PM, and so, are also current challenges for low-cost sensors.

DSNS '19, Spring 2019, Helsinki, Finland

ACM ISBN .
https://doi.org/

3 LABORATORY AND IN-FIELD CALIBRATION

Calibration of a low-cost sensor is needed to establish a relationship between the output of a low-cost sensor and a measurement standard, where the measurement standard in this context can be either a calibrated reference instrument or a gas/particle reference material.

Based on the literature from over the past ten years, laboratory-based sensor calibrations performed under controlled laboratory conditions tend to produce better analytical agreements between low-cost sensors and reference instruments than that in the field of naturally varying atmospheric composition.

Laboratory calibrations are useful for testing the functionality of the sensors. The laboratory experiments are usually designed that a dynamic system is needed for generating, pre-mixing and delivering a known concentration of a test gas or vapour in air. An exposure chamber is used for the test of sensors, under independently controlled airflow, concentration, temperature and relative humidity of the test gas and of any added interference [Spinelle et al. 2013].

In-field calibrations of gas phase sensors are widely considered as the more direct and appropriate method for comparing different measurement approaches in the real world, although sensor performance can differ when used in a different locations and exposed under different concentration levels. Due to different environmental and meteorological factors including ambient temperature, humidity and also other common atmospheric compounds, there is a need to perform temperature and humidity corrections and also cross-sensitivities among other atmospheric variables [Spinelle et al. 2013].

4 CALIBRATION METHOD

At present, many users rely on factory calibration settings, conversion from light absorption, voltage, or conductivity, as the main method of calibration. However, there is limited evidence to suggest, at least for the current generation of low-cost sensors, this is sufficient to provide long-term accurate data across the possible environments. To determine whether or not a factory calibration is sufficient, a validation of the data should be performed in an environment similar to the one in which the low-cost sensors will be used. Linear regression and multiple linear regression are more commonly used, while modern machine learning methods, for example Artificial Neural Network (ANN) and random forest, are on the rise.

4.1 Linear regression and Multiple linear regression

For each sensor a calibration function was established by assuming the linearity of the sensor responses with reference measurement for each pollutant. Ordinary linear regression was used with the minimization of square residuals of the sensor responses versus reference measurements. The calibration functions were of the type

$$Rs = a \cdot X + b, \tag{1}$$

where Rs represents the sensor responses and X is the corresponding reference measurements of air pollutant. Finally, the measuring function, the converse equation

$$X = (Rs - b)/a, \tag{2}$$

was applied to all sensor responses in order to predict air pollutant levels.

Multiple linear regression has similar principles but involves in more parameters than the linear regression. The additional parameters are usually meteorological data, such as relative humidity, temperature and wind speed, and gas compounds with which chemically interact [Duvall et al. 2016; Johnson et al. 2018; Spinelle et al. 2015].

4.2 Artificial Neural Network and Random Forest

The advantage of using machine learning is to save plenty of time in selecting features. Meanwhile, it can easily overfit the model by including everything. ANN receives a number of inputs and also hidden layers with hidden units. The weighted sum of the inputs is formed to compose the activation of the unit. The activation signal is passed through an activation function to produce the output of the unit. Random Forest works by constructing an ensemble of decision trees using a training data set and the users have to specifies the maximum number of trees that make up the forest. In both methods, performance can be improved by adjusting the number of hidden layers and number of trees, respectively. However, the black boxes in the models are usually discreet and not open to public.

The use of ANN for calibration purpose appears to be the most efficient in terms of uncertainties with a smaller drift over time on O3 and NO2. Humidity/temperature dependence was also corrected, without the needs of such measurements [Spinelle et al. 2015].

The use of random forest was proven to work better than traditional regressions, especially on CO2, NO2 and O3. It also seems to eliminate the cross-sensitive effect if we include additional measurement of air pollutants [Zimmerman et al. 2018].

4.3 Calibration evaluation

Performance comparison/calibration is often defined by the correlation statistics between the reference and sensor time series, the linearity of the sensors to the compound concentrations and the variability of the sensors compared to reference. Many studies currently use a combination of the correlation coefficient (R2), root mean squared error (RMSE), and mean absolute error (MAE) to describe their model performance. They are the most direct statistical description for the calibration performance. However, inter-sensor statistics, low-cost sensors and reference comparative pattern analysis and performance analysis on seasonal trend, which are less common in the current literature database, can be carried out to see their long-term performance [Hasenfratz et al. 2012].

4.4 Quality control of sensors

Currently, there is a set of guidelines to regulate the data quality control in the framework of the Indicative Measurements of the EU Directive on Ambient Air (2008/50/EC). Air pollution low-cost sensors should be treated as any other analytical instrument and they require regular calibration and show long-term changes in drift and sensitivity.

Several parameters from low-cost sensors that should be monitored over time including baseline drift (change in intercept) and changes in sensitivity (e.g. changes in slope). They should be corrected over time with data post-processing. Data quality has to be maintained by periodical comparison at regular intervals of 6 months of the values obtained with a low-cost sensor to a co-located reference monitor [Lewis et al. 2018].

5 CONCLUSION

This article illustrates the current situation of atmospheric low-cost sensors, especially on its performance, calibration and quality control. Currently, calibrations are divided into laboratory and field calibration, which is the real-world, posing different meteorological challenges. Simple regression is still the most common way for calibration. However, machine learning methods, like ANN, can be useful in merging sensitivity analysis as to their hidden layers. Calibration can be evaluated by statistical parameters like R2, RMSE, and MAE. Calibration should be carried out at a regular interval due to a possible long-term drift and different meteorological conditions.

The purpose of using low-cost sensors are not for perfect accuracy, but for a wider spatial resolution and an more extensive measurement network. The improvement of sensors by calibration would provide more insights from places other than the existing reference stations.

REFERENCES

Rachelle Duvall, Russell Long, Melinda Beaver, Keith Kronmiller, Michael Wheeler, and James Szykman. 2016. Performance evaluation and community application of low-cost sensors for ozone and nitrogen dioxide. *Sensors* 16, 10 (2016), 1698.

David Hasenfratz, Olga Saukh, and Lothar Thiele. 2012. On-the-fly calibration of low-cost gas sensors. In *European Conference on Wireless Sensor Networks*. Springer, 228–244.

Karoline K Johnson, Michael H Bergin, Armistead G Russell, and Gayle SW Hagler. 2018. Field test of several low-cost particulate matter sensors in high and low concentration urban environments. *Aerosol Air Qual. Res* 18 (2018), 565–578.

Alastair Lewis, W Richard Peltier, and Erika von Schneidemesser. 2018. Low-cost sensors for the measurement of atmospheric composition: Overview of topic and future applications. (2018).

L Spinelle, M Aleixandre, and M Gerboles. 2013. Protocol of evaluation and calibration of low-cost gas sensors for the monitoring of air pollution. *Publication Office of the European Union, Luxembourg* (2013).

Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto Bonavitacola. 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical* 215 (2015), 249–257.

Naomi Zimmerman, Albert A Presto, Sriniwasa PN Kumar, Jason Gu, Aliaksei Hauryliuk, Ellis S Robinson, Allen L Robinson, and R Subramanian. 2018. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques* 11, 1 (2018).