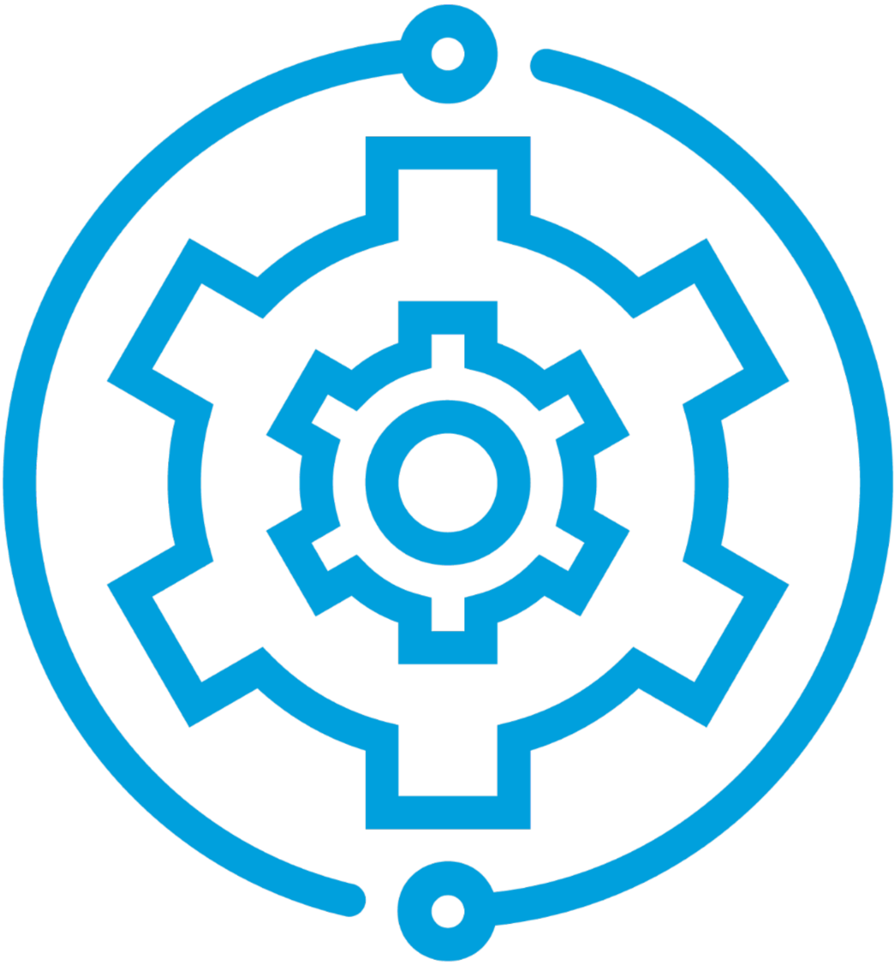
Operational Network of
Air quality Impact Resources

Best Practice Guide

BP301 | Implement and operate

Sensing device calibration



Introduction

Smart low-cost air quality sensing devices have enabled the collection of air quality data on a lower budget – and with higher spatial resolution – than has previously been possible using high grade air quality reference instruments alone, such as those used in air quality monitoring sites[1]. However, a challenge that can sometimes arise when using low-cost sensing devices is that the data generated by individual devices can be faulty or misleading, which negatively impacts the overall quality of the collected data.

This chapter provides information on how sensing device measurements can sometimes go wrong, and what approaches you can take to mitigate this. Depending on your exact application, it may be essential that you first perform co-location and calibration activities, well before collecting any data.

If the data gathered by your project will be used to inform critical decisions, then you need to have a very clear understanding of the accuracy of the data collected by your project's sensing device network. Important questions to ask include:

- How does your sensing device vendor quantify the measurement data accuracy of the devices?
- What is the deviation between your devices and the true measurement of pollutants (as measured by a certified reference grade instrument)?
- Does your device vendor have their own method or guidance for calibration of devices?

Who is this resource for?

This OPENAIR Best Practice Guide chapter is for local government staff responsible for managing and operating an air quality monitoring project. It is designed to be useful to project staff who may need to understand how to carry out co-location and calibration processes before an entire sensing network is deployed and activated (or who are tasked with hiring contractors to do this).

The information in this chapter is relevant to air quality sensing applications that are classified as Tier 2 or 3 (see the OPENAIR supplementary resource *A framework for categorising air quality sensing devices*, which defines air quality projects according to the quality of the data required). In Tier 2 or 3 applications, your project data will be used to make decisions as required by your business case, or will be used as supporting evidence to answer a research question.

How to use this resource

This chapter will provide an overview of why device co-location and calibration are required, and how to carry out these procedures. It will also help you to assign the necessary financial and human resources for your air quality monitoring project, by giving you a clear idea of the time and technical skills required.

[1] A list of measurement techniques used for air pollutants by the NSW Department of Planning and Environment are on their website here.

What is calibration?

Calibration is the process of configuring an instrument to provide a result within an acceptable range. Smart low-cost air quality sensing devices have great potential for local air quality monitoring projects, but without a clear understanding of their performance, the data they generate can end up being faulty or misleading.

A key method of addressing this is to compare the output of a low-cost sensing device to a ‘gold standard’ certified reference grade air quality instrument, which has a high degree of confidence in terms of precision and accuracy. This exercise is referred to as a **co-location measurement.** It is highly recommended that co-location is carried out, especially if the device manufacturer recommends it. This process provides solid evidence of sensing device performance, and the information it generates can be used to create correction factors to calibrate individual sensors.

What can go wrong without calibration?

Sensing devices can report incorrect results due to a range of issues, such as power outages, sensor blockage, and electronics failures. Even without those particular obstacles, an individual sensor (as a component within the sensing device) may still report measurements that differ from the true value.

At the highest level, this can appear as a **scale** and/or **offset** factor. This is demonstrated in Figure 1, which shows a hypothetical measurement (green) compared to the true (black) measurement over time. In all cases, there appears to be a general agreement in the reported data points, but in **Case A**, the relative heights of the measurements are different, and there is a general off-set. **Case B** shows no scaling issue, but there is a constant offset in the data. **Case C** shows the ideal case (often achieved after a successful calibration), where there is good agreement between the measurement and the true value.

There will always be a degree of error associated with the measurements of any sensor. These errors can be minimised by following the correct co-location and calibration processes. Figure 1 shows how an uncalibrated sensor can produce air pollution values that are multiple times higher than the true value, leading to false conclusions. This can also happen in the opposite direction, where readings are under-reported from the true value.

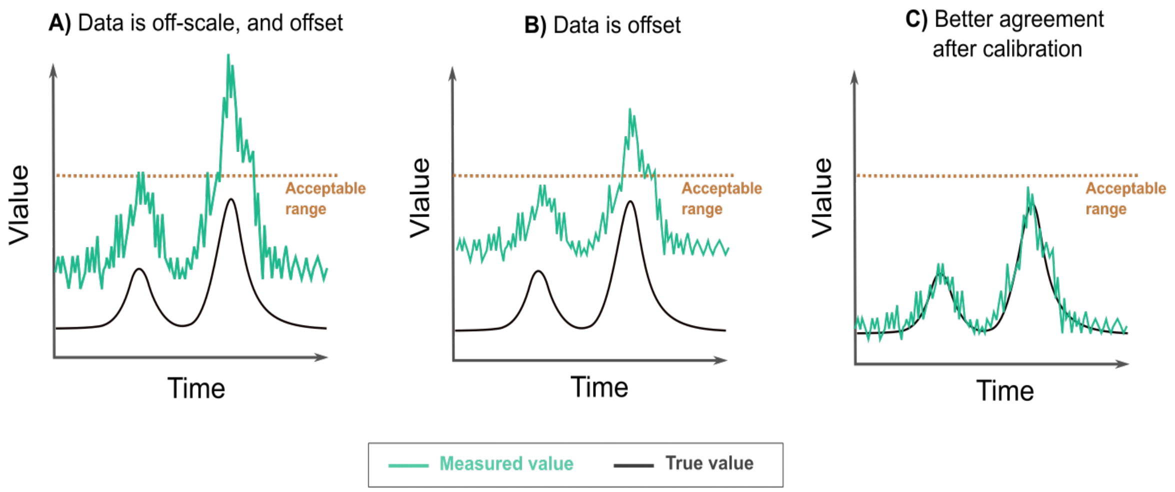


Figure 1. A simplified example of sensors that suffer calibration issues. In Case A, the collected data is off-scale and offset from the true values. The data in Case B has no scaling issue, but there is a constant offset from the true value. Under these conditions, a successful calibration results in the ideal Case C.

Why do readings from devices vary?

Sensing device vendors will advertise that their devices operate within a certain performance range, but there are many factors that can cause device readings to differ (some of which may not be the fault of the vendor). Despite efforts in making sensors identical, in practice it is often difficult for them to report the exact same values. For instance, particle sensors rely on a precise alignment of light sources and detectors to measure particles that are smaller than the width of a human hair. Uncertainty in manufacturing tolerances, light source power fluctuations, and detector alignment can all contribute to errors in the sensor readings.

Figure 2 shows readings from two 'identical' particle sensors, placed at the same location across the same time period (Device 2's data has been artificially shifted up for clarity). At first glance, the readings are very similar. However, on closer inspection, there are minor differences in the relative shape and height of the peaks associated with air pollution events. These two sensors agree with an R^2 value[2] of 91%, which is relatively high, but there are still discrepancies.

[2] R^2 is the represents the Pearson correlation coefficient. It is the most common parameter to measure the strength of linear association between two variables. It will be further explained later in this chapter.

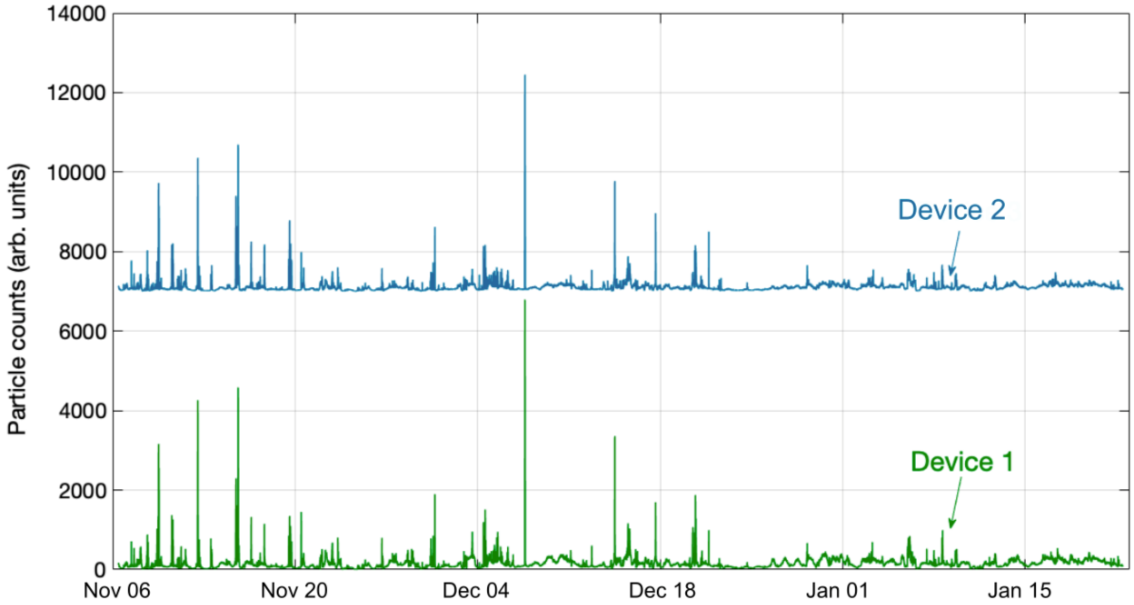


Figure 2. Comparison of data sets from two particle counters located in the same position (Device 2's values have been artificially shifted for clarity). There is an overall agreement between the counts over time, but closer inspection shows slight differences caused from errors in the devices themselves.

There are also environmental factors that will affect the readings, in ways that might be quite different to those stated in a sensing device vendor's data sheet. Some factors to note include:

- **Temperature and humidity**

  These factors can affect the electronics of the sensing device (such as the photodetector's sensitivity, or an electrochemical sensor's efficiency). High levels of humidity can cause certain pollutants to 'swell' and effect sensor readings. Advanced sensors may have a heated inlet to remove excess humidity, but this feature is not common in low-cost sensing devices.

- **Local climate conditions**

  The immediate environment of the site where your sensing devices are located can affect their readings (e.g. if they are inside a wind tunnel, or in a low-circulation zone). This means that individual sensor readings may not be representative of the wider region.

- **Chemical make-up and cross-contamination**

  Certain pollutants, mainly gases such NO_2, may cross-contaminate low-cost sensors designed to detect another pollutant - leading to false readings. This is known to be the case for electrochemical sensors that are not highly selective to the pollutants they were designed for. The chemical make-up of pollutants locally may also be different to those present in the location where the sensing devices were initially tested. For instance, an inland region may have high levels of dust, whereas coastal areas may have high levels of sea salt particles.

Some of these factors are described in more detail in the OPENAIR supplementary resource *Sensing device performance evaluation methodology*.

How much variation should I expect?

Given that all sensors have inherent errors in their readings, it is important to consider how these values are reported, and how to interpret them. Not all sensing device vendors report their errors in the same way, which can make this process difficult. Most vendors will describe their sensing device performance in general terms, including the following factors:

- **Range or limit of detection –** the operating limits of the sensor, or the smallest value able to be detected by the device
- **Precision –** how reliable/repeatable multiple measurements are over time (often characterised by an R^2 value with a reference instrument)
- **Accuracy –** how close the measured value is to the true value (quoted as an absolute value, such as ± 10 µg/m^3).

Be careful when interpreting these values. It is useful to consider when and where the test was done to derive these numbers, whether it was a lab test or outside test, which environmental conditions were present during the test, and whether the metrics were derived before or after a correction factor was applied.

As an example, the vendors of the *Clarity Node-S* sensing device have published the margins of error for their particulate matter (PM [3]) sensor, as shown in Figure 3 (Clarity, 2023). Below 100 µg/m^3, the accuracy is quoted at ± 10 µg/m^3, whereas above this value, it is quoted as ± 10%. The precision is characterised in terms of an 'R^2 > 0.8', when tested against a US EPA FEM instrument[4]. These values can be used to guide your device procurement decisions, but should not be the only deciding factor.

PARAMETER	TECHNOLOGY	RANGE	PERFORMANCE AFTER CALIBRATION
Particulate Matter $PM_{2.5}$ [µg/m^3]	Laser Light Scattering with Remote Calibration	0-1000 µg/m^3 1 µg/m^3 resolution	Accuracy: < 100 µg/m^3: ± 10 µg/m^3; ≥ 100 µg/m^3: within ± 10% of measured value Correlation (R^2) with USEPA FEM instrument > 0.8

Figure 3. A comprehensive list of sensor specifications provided by (Clarity, 2023) for their particulate matter readings. This provides details on the range, resolution, and accuracy for specific pollutant levels. It also provides an R^2 correlation value compared to a US EPA FEM instrument.

Importantly, these reported values are "performance after calibration". This means a mathematical correction has been done on the raw data to correct for the range of error-producing effects mentioned above. Depending on the vendor, this correction may be proprietary intellectual property, and it may not be possible to figure out exactly which calculations were performed. There is, however, a wealth of established literature on this topic, which will be explored in the next section of this chapter.

[3] PM (particulate matter) refers to airborne solids or liquids. Its size is measured in micrometres and is indicated by the subscript. E.g. $PM_{2.5}$ has a diameter of 2.5 micrometres or less. (NSW Health, 2020)
[4] FEM (Federal Equivalent Method) instruments meet a strict measurement performance criteria to ensure data quality (Clements, 2019).

What to expect when conducting calibrations

Now that you are aware of *why* calibration is important, the next step is to evaluate how much time, effort, and resources are needed to conduct calibration procedures at varying levels of complexity. The exact level of calibration your sensing devices will require depends on your project goals. For example, if you are investigating pollution hotspots, or running a long-term supplementary network to a nearby monitoring station, it is highly recommended that you carry out a co-location study before deploying a full sensing device network.

An example of a co-location study on low-cost sensing devices. The devices were mounted on a regulatory air quality monitoring station to gather data for several months. The data sets from both devices were later used for comparison and calibration purposes. Image source: University of Sydney

To conduct a co-location test, follow these steps:

1. **Determine a relevant co-location site**
   Find a location that has access to high-quality air quality data, able to be traced back to a certified air quality reference instrument. You should ideally locate two sites to get a good geographic distribution.
2. **Establish access to the site, and build relationship with the site managers**
   Running a co-location test will require physical access to the site, access to reference instrument data, and regular communication with on-site staff. Local governments can organise co-location testing by contacting their relevant state authorities that manages their air quality network. A minimum data collection period of one month is generally necessary. Staff from your own organisation will need to be designated to oversee the co-location testing. Data sharing and access protocols will also need to be established prior to the co-location test.

3. **Conduct a site visit**
Additional work and resources will be required to set up the sensing devices appropriately on-site. A site visit will ensure all set-up details are understood (such as mechanical mounting, power requirements, and data communication and storage).

4. **Make any required modifications**
Make appropriate plans to ensure ease of device set-up (such as providing mechanical mounting panels, or electrical wiring). It is best to assume that a staff member (or a contractor you hire) will be needed to manage this.

5. **Install devices on-site; run co-location tests**
Commence the co-location process, which may run for 1-3 months. During this time, site visits may be needed to check on progress.

6. **Collect and clean data; compare results**
The data collected by your devices will need to be cleaned, so that it can be usefully compared to reference instrument data. This involves checking for consistency in the measurement units, dealing with any data dropouts, and synchronising the timestamps of each measurement. After this, data analysis can be conducted, including the calculation of the R^2 correlation value. For a staff member with basic data analysis experience, this can take several days. The on-site staff may be able to assist, but this service should not be assumed. More detailed studies of the data will require specialists or researchers to be involved (e.g. PhD students or academics at a local university).

7. **Document the co-location test in detail**
Documenting the entire co-location process is critical to the future success of your air quality project. If there is ever any doubt about the performance of the sensing devices in your network, this co-location data will be used to resolve any problems or inaccuracies. Make sure you report in detail on all procedures (e.g. when, where, and how sensing devices were mounted; or any notable environmental disruptions during the co-location, such as bushfires).

8. **Apply calibration; commence air quality project**
Once the calibration factors are determined, these can be applied as a mathematical correction factor to future measurements. You are now ready to deploy your air quality sensing device network.

TIP: Do not underestimate the work required for co-location testing

It is highly recommended that you allocate a staff member to oversee this time-consuming process of co-location. If this is not possible, ensure you have good relationships and arrangements with the staff at your closest air quality regulatory monitoring station. Early stakeholder engagement of this nature will be critical, especially if you have limited resources.

Common questions about co-location

There is no 'one-size-fits-all' approach to device calibration using co-location. However, here are some general answers to commonly asked questions about co-location tests:

How long should co-location take?

Typically, co-location studies are done for a period of 1-3 months, but there is no set rule about what kind of time period is best. It has been suggested that 30-60 days of data is a minimum to get a statistically significant data sample (Liang, 2021). However, it is also the case that – in certain circumstances – doing a co-location study for *longer than* 60 days can potentially make it harder to characterise the sensors. This is because some low-cost sensors have shown degrading performance beyond this time period, and seasonal variations in the environment start to take effect. For this reason, it is best practice to choose a time period during which the environment does not significantly change.

Do we need to recalibrate?

Performing regular calibrations on the same sensing device is recommended, because calibration is only effective under the conditions in which the co-location study was conducted. Seasonal shifts and changes to the immediate environment (e.g. new highways or construction sites) may affect the performance of a prior calibration. In addition, sensor performance will degrade over time. In some cases, for example, gas sensors showed significant drift even within just one month (Miskell et al., 2016). There are niche methods to measure this degradation 'on the fly', so that you can decide when a recalibration is necessary, but these methods are not yet widely used (Hasenfratz et al., 2012).

Do we need to calibrate all sensors?

It is good practice to calibrate every sensor that you will be testing. However, there are circumstances where this is not practical. Although every sensor differs slightly in its performance, similar brands/units will face the same types of errors, as their fundamental sensing mechanisms are the same. Therefore, it may be possible to calibrate one unit, and then apply that same correction factor to other units of the same make/model. For example, the Breathe London project used a select number of devices for periodic co-location with reference instruments, and transferred those calibration factors to nearby units. For more details, see the *Breathe London Technical Report* (Breathe London, 2021).

Are there any additional costs?

A co-location test will require key personnel to regularly check the data feeds. To simplify this process, it is best to arrange for the sensor data to be remotely accessible (if it is not already). This saves on travel time to the site to inspect the device. Bear in mind that data communication will come at a fixed cost (for instance, a LoRaWAN gateway service can cost approximately AU$100-200 per month, depending on the provider package).

How to apply correction factors and further analysis

After performing a co-location study, you may need to apply correction factors yourself, depending on the sensing device vendor you have selected. Certain vendors include this as a service, where you provide the co-location data to them, and they derive the calibration factors and apply them to the devices in real time. Other vendors may have co-location testing included in their product offering, but the exact correction factors used may be proprietary information.

However, for many smart low-cost sensing devices, you will have to carry out this process yourself, and make sure the data output is corrected. This section of the chapter will guide you on how to do this.

Once you have completed a co-location study, you will have two sets of data: one for your sensing device's measurements; and one for the reference-grade instrument. If your sensing device records more than one parameter, you may have multiple data sets to analyse.

The next steps depend on what level of accuracy your application demands, and the degree to which you are willing to apply correction factors.

Figure 4 describes the methods used broadly in the literature to calculate and apply correction factors, explored in more detail in (Liang, 2021). It is important to understand that more sophisticated methods may improve the overall agreement between the sensors in a specific co-location test, but there is no guarantee this will work in all cases. For instance, when the same calibration is applied in a different context (e.g. in a new location), it may not be effective, since the calibration was highly tuned to the previous location.

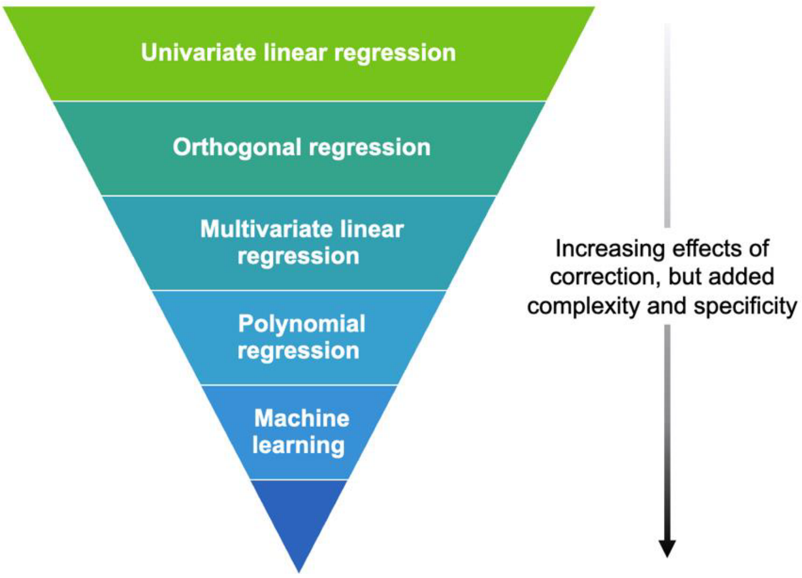


Figure 4. The levels of 'corrections' that can be applied to sensing device outputs after a co-location study is done. Deeper layers of the diagram point to more sophisticated methods that may have increasing correction effects, but come at the cost of increased complexity.

Often, the simplest method is also the most versatile. Univariate linear regression can be performed to establish a linear relationship between the two data sets. In a co-location study, the independent variable is the low-cost sensing device's output, and the reference instrument is the dependent variable. This establishes a linear relationship as:

$$\mathbf{S_{LCS} = \alpha_0 + \alpha_1 S_{ref} + \varepsilon}$$

Equation 1. A simple linear relationship between the low-cost sensing device data (S_{LCS}) and the reference instrument data (S_{ref})

S_{LCS} are the low-cost sensing device measurements; S_{ref} are the reference instrument readings; α_0 represents the constant offset (intercept); α_1 is the regression coefficient (gradient); and ε is the measure of error. Ideally, if there are no offsets or scaling factors, the coefficient α_0 is zero, and α_1 is simply 1.

Conducting a regression analysis will provide numerical values for the above relationship. Figure 5 shows an example where the sensor data is plotted against the reference instrument, and the linear relationship is derived (and subsequently used to correct the readings). The raw results are seen on the left, with the data being slightly skewed from 'line of agreement'. The linear regression has identified the relationship through the coefficients in Equation 1. Once this has been done, you can simply invert the offset and bias to create a 'calibrated' data set.

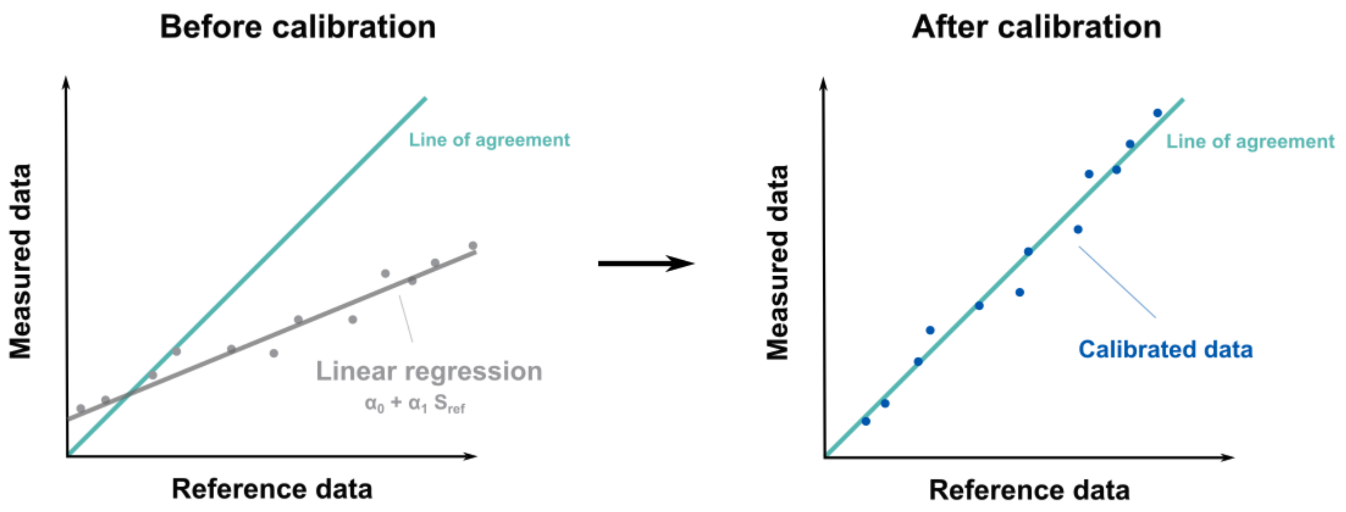


Figure 5. Application of a univariate regression analysis to a hypothetical data set. (Left) Applying the linear regression to the data set establishes a linear relationship between the two data sets. (Right) Using the coefficients, one can then 'correct' for the measured data, so that it best aligns with the line of agreement for the reference data.

This correction will help align any future data to a more relevant range that is indicative of the reference instrument with which it was co-located. How well the linear regression has performed can then be quantified, by measuring the 'R^2 correlation coefficient'.

The R^2 correlation coefficient

The Pearson correlation coefficient is the most common parameter to measure the strength of linear association between two variables – denoted by *R.* The coefficient can take values ranging from +1 (strong positive correlation) to -1 (strong negative correlation). By taking the squared value (*R^2),* which ranges from 0 to 1, we consider the proportion of variance in the measurement that can be explained by reference data in the regression model. Higher values indicate a stronger correlation between the data sets.

There are no clear, set values of R^2 that enable one to define sensing devices as being in 'good agreement' with one another; it all depends on the data use case requirements. Typical R^2 values, and the circumstances in which they were derived, are outlined in (Rai et al., 2017) for a wide range of air quality sensing devices (as a general guide, values of $R^2 > 0.8$ are considered 'excellent', and tests achieving $R^2 < 0.1$ are considered to be 'extremely poor performance'). The exact thresholds of a 'good agreement' between devices is highly dependent on the chosen application, but you should record the R^2 correlation coefficient of your co-location test for future reference.

The limitations of R^2

Although the R^2 test is a common method to assess the performance of a co-location study, it has some severe limitations when used in isolation. One of the most critical factors is that it does not take into account **the range of pollutant concentrations**.

For example, if a co-location test is conducted for NO_2 levels, but the true values do not rise by any significant amount, the comparison becomes very difficult. The R^2 value expresses how one variable responds as the other changes, but if this change is small to begin with, it will naturally be low.

It is therefore necessary to make sure that the natural variation of the pollutant is significant, or to also report the 'mean absolute error' values as well. This topic is explored in more depth in (Micalizzi, 2020).

Employing more sophisticated methods

In addition to simple univariate linear regression, there are other methods that can be applied to improve agreement between sensors. It is beyond the scope of this chapter to go into detail about each method, but there are well-established and documented steps in the technical literature to conduct them, should they be necessary to your project. For example, using multivariate linear regression expands on Equation 1 by including more parameters that may affect accuracy (such as humidity, temperature, pressure, wind speed, dew point, and precipitation).

However, these complex models have only shown marginal improvements in some cases, and run the risk of overcomplicating the relationship between sensors. Machine learning is a collective term to include various data-driven approaches used to establish relationships between the sensors. Certain studies have indicated that a 10% improvement in R^2 correlation results were observed using this approach, when compared against linear regression results (Mahajan & Kumar, 2020).

Specific methods also exist to correct for the influence of humidity on particle counters readings. Aerosols are known to undergo hygroscopic growth at high relative humidity (>60%), and low-cost sensors need to be corrected for using κ-Köhler theory, as described in (Crilley et al., 2020). This process involves additional calculations, and assumes humidity data is available.

In summary, there are multiple correction factors that can be applied to overcome specific issues related to sensor readings and calibration. However, every intervention comes with a risk of biasing the data towards what is intended as a 'good result'. The assumptions underlying these corrections should always be justified and disclosed. Often, this requires deep technical knowledge, and these methods are only recommended if you have access to experienced personnel (or contractors) as part of your project team.

References

Breathe London. (2021). Breathe London Technical Report Pilot Phase (2018-2020). https://www.aqmesh.com/media/ygfdyyxm/bl-pilot-final-technical-report.pdf

Clarity. (2023). Node-S Technical Specifications. https://www.clarity.io/products/clarity-node-s

Clements, A. V., Robert (2019). *FRMs/FEMs and Sensors: Complementary Approaches for Determining Ambient Air Quality*. US EPA. https://www.epa.gov/sites/default/files/2019-12/documents/frm-fem_and_air_sensors_dec_2019_webinar_slides_508_compliant.pdf

Crilley, L. R., Singh, A., Kramer, L. J., Shaw, M. D., Alam, M. S., Apte, J. S., Bloss, W. J., Hildebrandt Ruiz, L., Fu, P., Fu, W., Gani, S., Gatari, M., Ilyinskaya, E., Lewis, A. C., Ng'ang'a, D., Sun, Y., Whitty, R. C., Siyao, Y., Young, S., & Pope, F. D. (2020). Effect of aerosol composition on the performance of low-cost optical particle counter correction factors. *Atmospheric Measurement Techniques, 13*(3), 1181-1193.

Hasenfratz, D., Saukh, O., & Thiele, L. (2012). On-the-Fly Calibration of Low-Cost Gas Sensors Wireless Sensor Networks, Berlin, Heidelberg.

Liang, L. (2021). Calibrating low-cost sensors for ambient air monitoring: Techniques, trends, and challenges. *Environmental Research*(111163).

Mahajan, S., & Kumar, P. (2020). Evaluation of low-cost sensors for quantitative personal exposure monitoring. *Sustainable Cities and Society, 57*(102076).

Micalizzi, P. (2020). How to assess air quality sensor accuracy: R^2. *Clarity*. https://www.clarity.io/blog/how-to-assess-sensor-accuracy-r2

Miskell, G., Salmond, J., Alavi-Shoshtari, M., Bart, M., Ainslie, B., Grange, S., McKendry, I. G., Henshaw, G. S., & Williams, D. E. (2016). Data Verification Tools for Minimizing Management Costs of Dense Air-Quality Monitoring Networks. *Environmental science & technology, 50*(2), 835-846.

NSW Health. (2020). *Particulate matter (PM10 and PM2.5)*. NSW Government. https://www.health.nsw.gov.au/environment/air/Pages/particulate-matter.aspx

Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., Yasar, A., & Rickerby, D. (2017). End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Science of the Total Environment* 691-705.

Associated OPENAIR resources

Supplementary resources

A framework for categorising air quality sensing devices

This resource presents a new framework for categorising air quality sensing devices in an Australian context. It identifies four tiers of device types, separated in terms of functionality, and the quality and usability of their data output. It is designed to assist with the selection of devices that are appropriate to meeting the needs of a project and an intended data use case.

Sensing device performance evaluation methodology

This resource presents a methodology for evaluating the performance (and corresponding data quality) of smart low-cost air quality sensing devices.

Further information

For more information about this project, please contact:

Peter Runcie

Project Lead, NSW Smart Sensing Network (NSSN)
Email: peter@natirar.com.au

This Best Practice Guide chapter is part of a suite of resources designed to support local government action on air quality through the use of smart low-cost sensing technologies. It is the first Australian project of its kind. Visit www.openair.org.au for more information. OPENAIR is made possible by the NSW Government's Smart Places Acceleration Program.

Document No: 20231208 BP301 Sensing device calibration Version 1 Final