Pre-Deployment Testing, Augmentation and Calibration of Cross-Sensitive Sensors

Balz Maag*, Olga Saukh*, David Hasenfratz†, Lothar Thiele*
*Computer Engineering and Networks Laboratory, ETH Zurich, Switzerland
†Sensirion AG, Staefa, Switzerland
{bmaag, saukh, thiele}@tik.ee.ethz.ch, david.hasenfratz@sensirion.com

Abstract

Over the past few years, many low-cost pollution sensors have been integrated into measurement platforms for air quality monitoring. However, using these sensors is challenging: concentrations of toxic gases in ambient air often lie at sensors' sensitivity boundaries, environmental conditions affect the sensor signal, and the sensors are cross-sensitive to multiple pollutants. Datasheet information on these effects is scarce or may not cover deployment conditions. Consequently the sensors need to undergo extensive pre-deployment testing to examine their feasibility for a given application and to find the optimal measurement setup that allows accurate data collection and calibration.

In this work, we propose a novel method to conduct in-field testing of low-cost sensors. The algorithm proposed is based on multiple least-squares and leverages the physical variation of urban air pollution to quantify the amount of explained and unexplained sensor signal. We verify (i) whether a sensor is feasible for air quality monitoring in a given environment, (ii) model sensor cross-sensitivities to interfering gases and environmental effects and (iii) compute the optimal sensor array and its calibration parameters for stable and accurate sensor measurements over long time periods. Finally, we apply our testing approach on five off-the-shelf low-cost sensors and twelve reference signals using over 9 million measurements collected in an urban area. We propose an optimized sensor array and show—compared to a state-of-the-art calibration technique—an up to 45% lower calibration error with better long-time stability of the calibration parameters.

Categories and Subject Descriptors

B.8 [**Performance and Reliability**]: Reliability, Testing, and Fault-Tolerance; G.3 [**Probability and Statistics**]: Correlation and regression analysis

General Terms

Algorithms, Verification, Performance

Keywords

sensors, multiple least-squares, in-field testing, sensor calibration, air quality

1 Introduction

Monitoring urban air quality is of great interest in modern society. Airborne pollutants can cause various diseases, such as respiratory illnesses, and are harmful to the environment. Usually, air pollutants are monitored by static measurement stations operated by governmental authorities, *e.g.*, the Swiss National Air Pollution Monitoring Network (NABEL) [29]. Numerous precise analytical instruments used in these stations cost several thousand dollars each [30] and require frequent calibration and maintenance [31, 26, 18]. Therefore, the number of deployed stations and, hence, the spatial resolution of the gathered data is limited.

Breakthroughs in sensor technology made new generation of small, cheap and portable air quality sensors available on the market. They are usually based on electrochemical cells, *e.g.*, [4], or semiconductor technologies, *e.g.*, [35], which allow a compact and inexpensive design. Researchers and start-ups integrate these sensors in their measurement platforms to monitor air pollution. Over the past years, numerous research projects, *e.g.*, CommonSense [7], MESSAGE [25], OpenSense [19], and public initiatives, such as, Air Quality Egg [1] and Data Canvas [2], were launched to explore opportunities of these new technologies and raise awareness in the society. However, results of laboratory tests and comparison against precise analytical instruments often report insufficient sensor accuracy, sensor drift and low correlation with reference measurements when trying to measure pollutant concentrations in ambient air [30, 17].

In this work, we present a novel method to quantify the real-world usability of a low-cost sensor for monitoring urban air pollutants by splitting the sensor's measurements into explained, unexplained and noise components. We show that many low-cost sensors can be used to monitor air quality, if deployed as part of a sensor array and jointly calibrated in combination with other low-cost sensors of the sensor array. Furthermore, we can compute more stable calibration parameters for a sensor array compared to a state-of-the-art calibration technique and, hence, need to calibrate the sensors less often.

Challenges. Using low-cost sensors to monitor pollutants in ambient air is challenging as (i) measured concentrations are low and often lie at sensors' sensitivity boundaries, because many sensors are primarily designed to sense higher pollution concentrations, *e.g.*, in the automotive industry [3]; (ii) environmental conditions—typically temperature and humidity—impact the speed of chemical reactions and, thus, the sensor output; and (iii) low-cost air quality sensors often suffer from low selectivity and their response is affected by a wide range of substances in the air, referred to as *cross-sensitivities*. Datasheet information on a sensor's cross-sensitivities is often scarce. Sensor producers may list measured cross-sensitivities per interfering pollutant based on laboratory experiments conducted in a certain fixed setting [4, 5] but more often provide no quantitative cross-sensitivity evaluation at all [15, 36, 6]. Consequently, considering only information from datasheets can limit the sensor performance during deployment. For instance, Eugster and Kling [16] use a methane (CH_4) sensor for rural air monitoring in Alaska. They derive a temperature and humidity correction model from information in the datasheet. Due to sparse testing at only three different humidity levels, the appliance of their model is limited to situations with a relative humidity larger than 40%.

Unless the sensors undergo extensive pre-deployment testing, using such sensors to monitor air quality is difficult. Outdoor air composition is complex and can exhibit notable variations over time, both daily and seasonally. Sensor testing and calibration in a laboratory requires exposing a sensor to a wide range of artificially created but feasible gas mixtures. This is labour intensive, requires complicated setup, *e.g.*, environmental chambers and gas mixtures, and assumes a very good knowledge of the air composition in the target environment and cross-sensitivities of the sensor. In contrast, testing and calibrating sensors in-field leverages existing high-quality measurement stations deployed outdoors and, thus, the results are more relevant for an outdoor deployment and no complex setup is needed.

Contributions and road-map. In this paper, we test and calibrate low-cost sensors in the field by conducting parallel measurements at a high-quality reference station. By analyzing obtained data, we can (i) verify whether air pollution monitoring with a given sensor in a given environment is feasible, (ii) characterize sensor cross-sensitivities and compute the sensor array, which can optimally monitor the specific pollutant, and (iii) compute calibration parameters for the sensors in the system. In this paper we make the following contributions:

- After introducing our assumptions and models in Sec. 2 and summarizing sensor calibration using multiple least squares [43] in Sec. 3, Sec. 4 describes our sensor testing methodology to uncover sensor cross-sensitivities. Given a limited number of references, we design an indicator that allows quantifying the amounts of captured and uncaptured cross-sensitivities and sensor noise. To the best of our knowledge, we are the first to provide detailed in-field characterization of low-cost sensors.

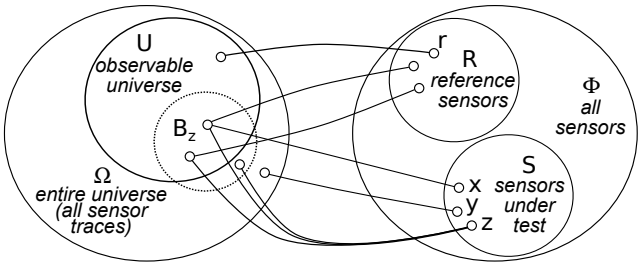


Figure 1. Relationship between different sets introduced in the paper.

- Sec. 5 reports the results of using the proposed testing methodology on real data. We use several low-cost sensors available on the market and previously used by other researchers [30, 17] to show that our approach allows to conclude on the feasibility of using a sensor in a given setting. We give positive and negative examples of sensors used for air quality monitoring in current and past projects. Furthermore, we leverage the approach to test and calibrate our urban air quality measurement system [23] equipped with low-cost cross-sensitive sensors.

2 Assumptions and Models

This section introduces basic terminology and discusses assumptions and models used throughout the paper.

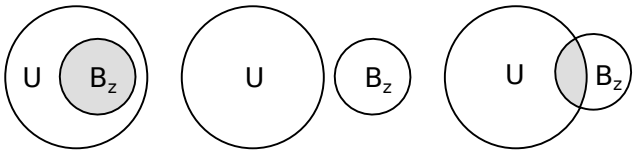
2.1 Observable Universe

Let Φ be a set of all sensors one can possibly build. Let $R \subset \Phi$ be a set of available sensors that accurately measure some phenomena of interest. That is, a sensor $r \in R$ accurately measures a single phenomenon, *e.g.*, ambient temperature. The sensors in R can be used as *reference sensors* to test the quality of other sensors, such as low-cost sensors, measuring the same phenomena.

A time-ordered sequence of discrete measurements $m = \{m(t_j)\}$ taken by a sensor at times t_j for $j \in \{1,2,...,n\}$ within a time interval $[t_1,t_n]$ is referred to as a *trace*. We consider a measurement as a point measurement, that is, it has no duration. The scenario a sensor is used in limits the number of possible traces of that sensor, *e.g.*, an air quality sensor reports different measurements in an automotive industry application than in monitoring outdoor air quality. Consider some fixed scenario of interest, let Ω and $U \subset \Omega$ be sets of corresponding traces produced by the sensors Φ and $R \subset \Phi$ respectively. Ω can be understood as the *entire universe* of all sensor traces and U as the *observable universe* determined by a set of references R, see Fig. 1.

2.2 Low-cost Sensors

Let $S \subset \Phi$ be a set of low-cost sensors under test. We assume no prior knowledge about the sensors. In order to explain a trace of a low-cost sensor y, we relate it to the traces of reference sensors in the observable universe U, *i.e.*, we represent $m_y = \{m_y(t_j)\}$ as a function of traces in U and an *unexplained*—or residual—trace in $\Omega \setminus U$. If m_y is solely representable as an unexplained trace in $\Omega \setminus U$, there is no possibility to explain measurements of sensor y with references in the given scenario. If a trace $m_x \in \Omega$ of a sensor



(a) Inclusion. (b) Exclusion. (c) Intersection.

Figure 2. Scenarios to quantify the amount of the observable universe U captured by measurements B_z.

$x \in S$ can be completely explained with a single reference trace m_r in U, one can compare the readings of sensor $x \in S$ to the response of a corresponding reference sensor $r \in R$ and calibrate it if needed.

2.3 Cross-sensitivity

Low-cost sensors show in general a close to *linear* or *polynomial response* to reference traces [4, 5, 35] but typically suffer from offsets and drifts that result in a substantial deviation of their trace from a reference trace. In order to minimize this deviation, the trace m_x of a sensor x needs to be *calibrated* to a reference trace m_r. Calibration is usually performed by representing the calibrated sensor trace $\hat{m}_x$ as a function *cal* of the raw trace, *i.e.*, $\hat{m}_x = cal(m_x)$. An optimal calibration function *cal* minimizes some norm of the difference between the calibrated sensor trace $\hat{m}_x$ and the reference sensor trace m_r.

An interesting sensor testing and calibration challenge arises if the trace of sensor $z \in S$ is a function of multiple traces $B_z \subset \Omega$, also known as sensor *cross-sensitivities*. In this case, calibrating a cross-sensitive sensor z to a reference r requires to describe the calibrated trace as a function of multiple traces, *i.e.*, $\hat{m}_z = cal(m_z, m_{s_1}, m_{s_2}, ...)$ given multiple sensors $s_i \in \Phi$ such that some norm of the difference between $\hat{m}_z$ and m_r is minimized. Given an observable universe U formed by some reference sensors R and a cross-sensitive sensor z, we distinguish three types of relationship between U and B_z as illustrated in Fig. 2:

Inclusion: $B_z \subseteq U$. The set of available reference sensors R can measure all cross-sensitivities of sensor z. Hence, the sensor response can be fully *explained* by the set of available references R, see also sensor x in Fig. 1. Moreover, R can be used to calibrate the trace of sensor z to any trace in B_z as is detailed later. In this case, we can unambiguously conclude on sensor quality and perform best-possible sensor calibration.

Exclusion: $B_z \cap U = \emptyset$. There is no relation between the sensor response and the observable universe U. In this case, the sensor response *can not be explained* by means of reference sensors $R \in \Omega$ and hence also not calibrated, see sensor y in Fig. 1.

Intersection: $B_z \cap U \neq B_z$. The most common case is that only a part of B_z can be explained by U, see sensor z in Fig. 1. We refer to $B_z \cap U$ as to *explained* part of z's response and to $B_z \setminus U$ as to its *unexplained* part. The usability of sensor z in a given scenario depends on whether the explained part of the trace dominates its unexplained part and sensor noise.

Low-cost gas sensors are often cross-sensitive, because their small sensing surface area and low power consumption requirements limit the selectivity [40]. Moreover, environmental parameters, such as ambient temperature and humidity, influence the speed of chemical reactions and, thus, often affect the sensor response [21]. We assume that a measurement of a cross-sensitive sensor is an *additive* combination of different, possibly non-linear effects describing the impact of different *phenomena*, *e.g.*, interfering gases or meteorological effects [11].

2.4 Sensor Array

Ignoring sensor cross-sensitivity or environmental parameters leads to poor sensor calibration. Cross-sensitive sensors are usually *augmented* with collocated sensors to a set of sensors $M \subseteq \Phi$, called *sensor array* throughout this paper. Sensor arrays are used to compensate for cross-sensitivities. A cross-sensitive sensor z can be calibrated to a reference sensor r using a *multiple regression* method, given low-cost sensors in M that cover *all* phenomena in B_z. These multiple regression methods can find the function of pre-processed and aggregated measurements from sensors in M that minimizes its deviation from a reference r. However, the knowledge of B_z is often incomplete or unknown due to scarce datasheet information obtained through basic tests conducted in the laboratory, or there is no information at all. Even if B_z is known (it might consist of multiple relevant phenomena [5]), the quality of sensor tests and calibration of sensor z is limited by the set of available reference sensors R. In this work, we give answers to the following questions: (i) Given a cross-sensitive sensor z and references R, how can we identify cross-sensitivities $B_z \cap U$ of sensor z? (ii) How should z be augmented to a sensor array M when using it in a measurement system to improve the measurement quality? (iii) If sensor z is sensitive to phenomena not covered by U *i.e.*, $B_z \cap U \neq B_z$, can z still be reasonably calibrated and used in a given scenario? Since the list of cross-sensitivities is typically long, $B_z \cap U \neq B_z$ presents a common case when dealing with gas sensors.

2.5 Test Deployment Conditions

Testing a cross-sensitive sensor z in a laboratory requires simulating common deployment conditions and varying concentrations of every substance in B_z. This is expensive, time and labour-intensive if the list of sensor cross-sensitivities is long. For instance, the datasheet of the Alphasense $NO_2 - B4$ sensor [5] lists 11 possible cross-sensitivities, whereof at least four can have a considerable impact on the sensor response depending on the scenario. In contrast, in-field sensor calibration with parallel measurements with reference sensors R is an alternative and gives the advantage that sensor packaging and deployment conditions are similar to those of the target deployment and environment. The latter is crucial when testing and calibrating low-cost sensors, which often measure at their sensitivity boundaries. In this work, we assume that the data collected for sensor testing and calibration is gathered under similar conditions as on the target deployment.

3 Sensor Calibration

Sensor calibration is necessary to maintain sensor data quality. Since the sensor response can be affected by multiple factors, a large body of work tackles the task to com-

pensate for these factors. This section recapitulates sensor calibration based on multiple least-squares [43]. Further, we discuss an application example for a cross-sensitive gas sensor.

3.1 Ordinary Least-Squares

Regression analysis is often used to calibrate sensor measurements according to a reference trace [8, 18]. The common approach is to calibrate a raw sensor measurement $m(t)$ at time t to a given reference sample $m_r(t)$ such that

$$m_r(t) = \beta_0 + \beta_1 \cdot m(t) + \varepsilon(t), \tag{1}$$

where β_0 and β_1 are calibration parameters describing intercept and slope of a *calibration line*, and ε is a regression error component with zero mean. *Ordinary Least Squares* (OLS) regression is typically used to compute estimates of the calibration parameters $\hat{\beta}_0$ and $\hat{\beta}_1$. A raw sensor measurement $m(t)$ can then be converted to its calibrated version $\hat{m}(t)$ as follows

$$\hat{m}(t) = \hat{\beta}_0 + \hat{\beta}_1 \cdot m(t). \tag{2}$$

3.2 Multiple Least-Squares

The measurements of cross-sensitive sensors are aggregated measurements of multiple phenomena. Consequently, the measurements correlate poorly to measurements of a single reference. Using the one-dimensional regression model described above to calibrate cross-sensitive sensors leads to poor calibration accuracy. The standard solution—also known as *multiple regression* [43]—is to include additional regressors $m_l \in M$, $l \in \{1,2,...,k\}$ into the model. The goal of multiple regression is to find coefficients β_i, $i \in \{0,1,...,k\}$ of the linear combination of different sensor measurements and compositions thereof $m_l(t)$ that best fits a reference measurement $m_r(t)$ as follows

$$m_r(t) = \beta_0 + \beta_1 \cdot m_1(t) + ... + \beta_k \cdot m_k(t) + \varepsilon(t) \tag{3}$$

and accordingly in matrix form

$$\mathbf{m}_r = \mathbf{M} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4}$$

where $\boldsymbol{m}_r \in \mathbb{R}^{n\times 1}$, $\mathbf{M} \in \mathbb{R}^{n\times(k+1)}$, $\boldsymbol{\beta} \in \mathbb{R}^{(k+1)\times 1}$, $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times 1}$ and n is the number of samples at times t_j, $j \in \{1,2,...,n\}$. The estimates of the regression parameters $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{(k+1)\times 1}$ are calculated by *multiple least-squares* (MLS) [43] and the set of raw sensor traces $\mathbf{M}$ is calibrated by applying

$$\hat{\mathbf{m}} = \mathbf{M} \cdot \hat{\boldsymbol{\beta}}. \tag{5}$$

Calibration quality. The goal of least-squares based regressions is to minimize some norm of the regression error $\boldsymbol{\varepsilon}$ with

$$\boldsymbol{\varepsilon} = \hat{\mathbf{m}} - \mathbf{m}_r. \tag{6}$$

We use the *root-mean-square error* (RMSE) between $\mathbf{m}_r$ and $\hat{\mathbf{m}}$ to evaluate calibration accuracy. RMSE is a standard metric [34, 8, 13, 14] to quantify calibration quality and is computed as follows

$$\mathrm{RMSE}(\hat{\mathbf{m}}, \mathbf{m}_r) = \left(\frac{1}{n} \sum_{j=1}^{n} (\hat{m}(t_j) - m_r(t_j))^2 \right)^{\frac{1}{2}}. \tag{7}$$

(a) NABEL air quality measurement station.

(b) Air-inlets of the NABEL measurement station and our sensor testing box.

Figure 3. Sensor box deployment at the NABEL measurement station.

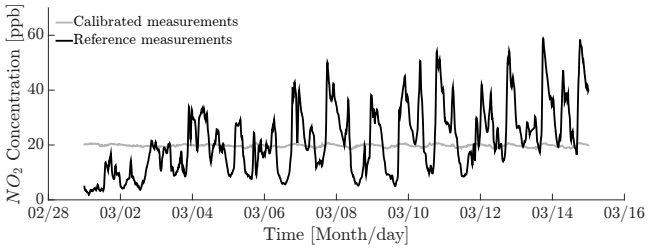


Figure 4. Ordinary least-squares calibrated $m = NO_2^*$ measurements and $r = NO_2$ reference measurements over time. The calibrated measurements do not correlate to the reference.

3.3 Application Example: NO_2 Sensor

When calibrating a cross-sensitive sensor, OLS is usually not suited. For instance, when calibrating a nitrogen dioxide (NO_2) sensor from AlphaSense [5], we use OLS to calculate the calibration parameters based on a training dataset of two weeks gathered in February 2014. As reference we use NO_2 measurements from a static, high-quality NABEL reference station (see Fig. 3) in an urban area, where our sensors are installed on the roof of the station to ensure collocated measurements. The outcome of the calibration for a test dataset of two weeks during March 2014 is presented in Fig. 4, which shows the calibrated measurements and the corresponding NO_2 reference over time. The calibrated measurements remain nearly constant over the whole calibration period. Due to sensor cross-sensitivities there is no correlation between uncalibrated sensor measurements and the reference. As a result the slope of the calibration line has a strong bias towards zero and the overall RMSE of the calibration is 12.4 ppb.

In order to improve the calibration quality, we apply MLS on measurements from multiple sensors M, measuring phenomena to which our sensor to be calibrated is cross-sensitive to. However, the calibration quality heavily depends on the choice of the sensors in M. We choose a sensor array consisting of $M = \{NO_2^*, H^*, T^*\}$, where NO_2^*, H^* and T^* are low-cost nitrogen dioxide, humidity and temperature sensors. In the reminder of this paper, we denote all low-cost sensors that need calibration with an asterisk (*). The grey line in Fig. 5 shows the calibration outcome of MLS during the same two weeks in March. The calibrated mea-

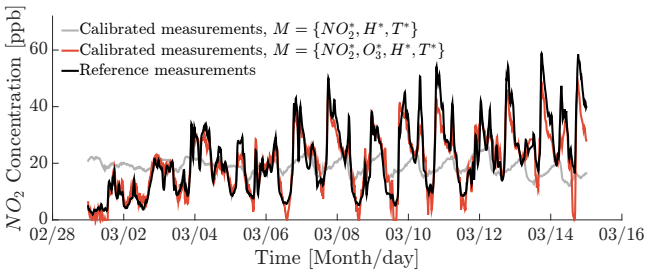


Figure 5. Multiple least-squares calibration with different sensor arrays M and $r = NO_2$ reference measurements over time. The calibration quality heavily depends on the choice of traces in M.

surements still correlate poorly to the reference and have a notable RMSE of 12.8 ppb. The reason for the poor performance is the cross-sensitivity of the NO_2 sensor to ozone (O_3).

The calibrated measurements when a collocated O_3 sensor is added to M, *i.e.*, $M = \{NO_2^*, O_3^*, H^*, T^*\}$, is plotted in Fig. 5. The calibration quality is improved significantly. There is a clear correlation between calibrated measurements and reference and an almost 3 times smaller RMSE of 4.6 ppb. We conclude that MLS is able to calibrate a cross-sensitive sensor when *augmented* with appropriate sensors to a sensor array M. However, due to scarce or no information about cross-sensitivities, the set of sensors in M and their respective impact on the measurements of the cross-sensitive sensor to be calibrated is often unknown.

3.4 Discussion

The example of the low-cost NO_2 sensor in Sec. 3.3 emphasizes the need of a pre-deployment testing methodology. The calibration accuracy of a cross-sensitive sensor z is limited without the thorough knowledge of the phenomena B_z to which the sensor is sensitive. If a sensor trace is an additive combination of multiple phenomena, it correlates poorly to a single reference trace. In order to calibrate the sensor measurements to a reference, the phenomenon of interest needs to be *segregated* from the measurements. This is only possible, if all remaining phenomena in the measurements can be compensated for. A sensor may not be sensitive to the same extent on the individual phenomena and it is possible—depending on the scenario—that some cross-sensitivities have minor impact on the sensor behaviour. It is therefore important to identify, which phenomena need to be measured and their individual impact on the sensor trace under application-related circumstances.

The sensitivity list of a cross-sensitive sensor allows to augment it with additional sensors to a sensor array M. Having measurements from sensors in array M, which measures all phenomena in B_z, it is possible to accurately calibrate the measurements using multiple least-squares to the corresponding reference trace, as described in Sec. 3.2. The fit of the linear combination in Eq. (3) can then be seen as the segregation of the part in a cross-sensitive sensor trace $m_z \in M$ that is induced by the phenomenon of interest by compensating for the other measured phenomena and fitting it to a reference.

4 Testing Methodology

Datasheet information on sensors' cross-sensitivities and their dependency on meteorological parameters is often scarce [16]. Even though all sensors undergo laboratory testing and calibration, test settings only cover a few points in the sensing range. Given the usually long list of sensor cross-sensitivities, extensive tests and sensor calibration is highly time-consuming and is, therefore, hardly possible. To solve the problem, we propose a novel method that uncovers sensor dependencies under deployment-related conditions. We ignore any prior knowledge about the sensor, i.e., do not rely on any information given in the datasheet, and treat the sensor as a black-box. We choose the observable universe U as a set of all relevant phenomena for a given application.

Our testing methodology consists of three steps depicted in Fig. 6:

Standardization. All input traces are converted to a standardized representation with zero mean and unit variance in order to get scale-invariant sensor traces and thereby scale-invariant cross-sensitivity factors.

Inverse calibration. Multiple least-squares is used to regress the standardized measurements from the phenomena in the observable universe U on the measurements of low-cost sensor z. The resulting regression parameters give insights about the composition of the sensor measurements, *i.e.*, identifies cross-sensitivities of the sensor.

Error decomposition. The regression error $\boldsymbol{\varepsilon}$ is used as an indicator for missing phenomena in U and substantial sensor noise of z. We distinguish the latter on the frequency characteristics of physical phenomena and, therefore, can decompose the error by applying a low-pass filter.

We explain the three steps in more detail below.

4.1 Standardization

The measurements we use for testing can be any pollutant concentration, temperature, and relative humidity and, thus, all these measurements have different scales and units. In order to get scale-invariant results, all variables need to be standardized [10], *i.e.*, they need to be centred and have unit variance. We denote in the reminder of this paper the standardized form of a vector $\mathbf{x} \in \mathbb{R}^{n\times 1}$ as $\overline{\mathbf{x}} = \frac{\mathbf{x}-\mu_x}{\sigma_x}$, where μ_x and σ_x are mean and standard deviation of $\mathbf{x}$, respectively.

4.2 Inverse Calibration

The primary goal of the testing procedure is to uncover the explained part $B_z \cap U$ of a low-cost sensor $z \in S$, *i.e.*, expose the physical phenomena B_z the sensor is sensitive to (see Fig. 1). This is achieved by decomposing the sensor measurements into single phenomena of the observable universe U. In contrast to calibration, where usually multiple sensors are regressed on a reference, we reverse the process. Hence, given collocated measurements $\mathbf{z} \in \mathbb{R}^{n\times 1}$ of sensor z and references traces $\mathbf{U} \in \mathbb{R}^{n\times|U|}$ of multiple reference sensors in U, the standardized regression equation is

$$\overline{\mathbf{z}} = \overline{\mathbf{U}} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{8}$$

The estimated regression parameters $\hat{\boldsymbol{\beta}}$ will give insights about the extent of any cross-sensitivities or dependency on meteorological effects of sensor z. With that knowledge it is possible to determine, whether additional sensors need to

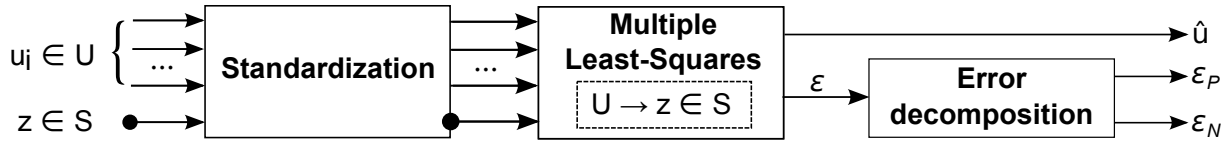


Figure 6. Sensor testing methodology.

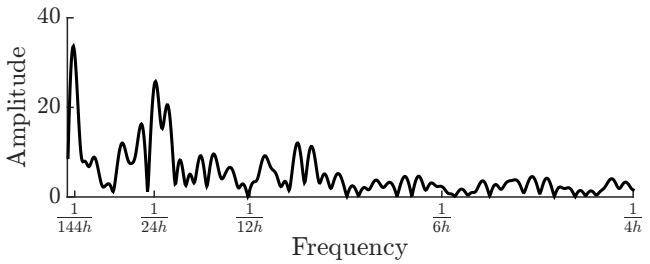


Figure 7. Frequency spectrum of O_3 reference measurements with a peak at frequency $\frac{1}{24h}$.

augment z forming a sensor array to accurately measure the target phenomenon. For instance, whether we need temperature and humidity values to compensate for meteorological dependencies.

4.3 Error Decomposition

It is possible that a sensor measures a phenomenon not captured by the observable universe U. In this case, we are not able to explain a certain part of the sensor trace with U, limiting the benefit of using the sensor in the given application. Hence, it is important to determine the fraction of the explained and unexplained parts of z given U.

The unexplained part of z depends on the performance of the inverse calibration, defined as regression error $\boldsymbol{\varepsilon}$, *i.e.*, the root-mean-square error

$$RMSE(\overline{\mathbf{z}}, \hat{\mathbf{u}}), \tag{9}$$

where $\hat{\mathbf{u}} = \overline{\mathbf{U}} \cdot \hat{\boldsymbol{\beta}}$ is the regression estimation solved by MLS. The larger the RMSE, the larger is the unexplained part of the sensor measurements.

The contribution of the unexplained part is often two-fold, *(i)* the sensor is impacted by physical phenomena, such as interfering gases, which can not be explained with the current universe U, and *(ii)* the measurements suffer from sensor noise.

In order to distinguish between these two cases, we exploit that the underlying phenomena and noise differ in their frequency representations. Sensor noise is often a high-frequent signal, whereas physical phenomena show distinct low-frequent variation patterns. For instance, the concentrations of primary air pollutants usually reach their maxima during the day and drop in the night [33], which is based on the increased activity of pollutant sources such as traffic or industrial plants during daytime. We exploit this daily periodicity by using it as indicator for any possible missing phenomena in U, indicated by a substantial low-frequent part (*e.g.*, with frequency $\geq \frac{1}{24h}$) in the error $\boldsymbol{\varepsilon}$ induced by the missing phenomenon. For instance, Fig. 7 shows the frequency spectrum of O_3 reference measurements recorded during April 2014. We observe substantial frequency components at frequencies larger or equal than $\frac{1}{24h}$. Consequently, the unexplained part of any sensor that is sensitive to O_3 will contain a notable low-frequent part, if O_3 is not included in universe U during the inverse calibration step. We use a low-pass filter to decompose $\boldsymbol{\varepsilon}$ in a low-frequency part $\boldsymbol{\varepsilon}_P$, which represents *uncaptured* periodic phenomena, and in a high-frequent part $\boldsymbol{\varepsilon}_N$, which we treat as the *noise* component of the sensor. For simplicity, we use $\frac{1}{24h}$ as cut-off frequency.

To quantify the impact of each error component, we compute the root-mean-square of a time-series signal. The $\text{RMS}(\boldsymbol{\varepsilon}_P) \in [0,1]$ serves as a measure for uncaptured phenomena in our model. High $\text{RMS}(\boldsymbol{\varepsilon}_P)$ indicates that it is likely that the sensor is cross-sensitive to a phenomenon not included in U but still related to a relevant phenomenon. By contrast, high $\text{RMS}(\boldsymbol{\varepsilon}_N) \in [0,1]$ is attributed to high sensor noise. Finally, the amount that can be explained with U is measured with $\text{RMS}(\hat{\mathbf{u}}) \in [0,1]$.

Depending on the decomposed errors, we can draw conclusions about the feasibility of deploying the sensor under test in a given environment. Assuming a sensor z can be fully explained with reference variables in U and is not affected by noise, *i.e.*, $B_z \cap U = B_z$, the regression estimation equals to the sensor measurements, *i.e.*, $\hat{\mathbf{u}} = \overline{\mathbf{z}}$. Consequently, both $\text{RMS}(\boldsymbol{\varepsilon}_P)$ and $\text{RMS}(\boldsymbol{\varepsilon}_N)$ are zero and $\text{RMS}(\hat{\mathbf{u}})$ corresponds to the standard deviation of $\overline{\mathbf{z}}$, *i.e.*, equals one. We can expect values close to zero and one respectively for any good low-cost sensor given an adequate observable universe.

4.4 Sensor Signature

The determined error components $\text{RMS}(\boldsymbol{\varepsilon}_P)$ and $\text{RMS}(\boldsymbol{\varepsilon}_N)$ and the explained part $\text{RMS}(\hat{\mathbf{u}})$ of a sensor z with a given observable universe U describe a *sensor signature*. Based on the sensor signature it is possible to determine the sensor array that can be used for compensating cross-sensitivities. If the testing methodology for sensor z is conducted multiple times with different compositions of universe U, the universe U that optimizes the sensor signature, *i.e.*, minimizes $\text{RMS}(\boldsymbol{\varepsilon}_P)$ and $\text{RMS}(\boldsymbol{\varepsilon}_N)$ and maximizes $\text{RMS}(\hat{\mathbf{u}})$, best describes the set of phenomena B_z the sensor under test is sensitive to. If $h \in B_z$ is a phenomenon of interest, then the necessary sensor array M to measure h is created by augmenting sensor z with low-cost sensors that measure phenomena $B_z \setminus h$.

5 Experimental Evaluation

In this section, we apply our testing methodology to different types of low-cost sensors. In Sec. 5.1 we test three sensors, which have different sensitivity profiles, *i.e.*, have different cross-sensitivities to different extents. We analyze their cross-sensitivities and meteorological dependen-

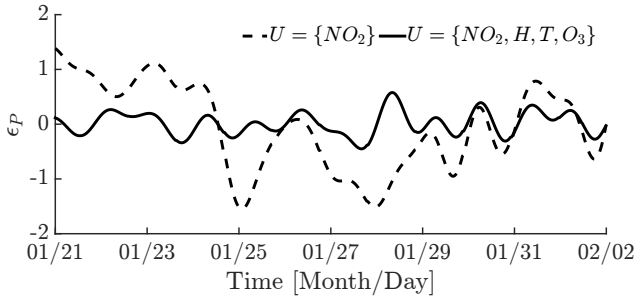


Figure 8. Low-frequent error component ε_P for different observable universes U and $z = \{NO_2^*\}$ over time. Including O_3 in U lowers the amplitude of ε_P, which highlights the sensor's cross-sensitivity to O_3.

cies and show how collocated measurements of multiple sensors in a sensor array can be used to accurately calibrate the measurements to references gases. Further, we investigate the stability of the calibration parameters over time. Finally, in Sec. 5.2, we present results from two sensors, which are not suitable for air quality monitoring in our setting. For these sensors we were not able to explain the sensor measurements with reference variables to an adequate extent.

5.1 Sensor Array Testing and Calibration

Our goal is to use low-cost sensors for monitoring major pollutants, namely O_3, CO and NO_2, in an urban environment. To achieve this goal, we build a measurement system consisting of multiple low-cost sensors. In order to find the optimal sensor array that ensures accurate measurements, we apply the presented testing methodology. We deploy three sensors at a station of the Swiss National Air Pollution Monitoring Network (NABEL) in Duebendorf, Switzerland, depicted in Fig. 3. The sensors are an electrochemical-based NO_2 sensor[1], a metaloxide-based O_3 sensor[2] and an electrochemical-based CO sensor[3]. The sensors are placed inside a box, which is mounted on the roof of the station next to the air inlets of the highly accurate devices. Therewith we ensure collocated measurements of low-cost sensors and reference devices.

The following evaluations are based on over 9 million measurement samples gathered during 15 months from January 2014 to March 2015. We use all sensing modalities measured by the official air quality measurement station as reference variables to build our observable universe. Further, we use temperature and humidity reference measurements to gain insights about the meteorological dependencies of the low-cost sensors. Because low-cost sensors usually do not show a completely linear dependency to phenomena, the samples of all references in U have additionally been included in quadratic and cubic form in the regression Eq. (8).

Sensor testing. In order to show the feasibility of our testing methodology, we start with an observable universe U, which consists only of the reference corresponding to each sensor. For example, the universe $U = \{NO_2\}$ is used to initially perform the testing methodology for the low-cost sensor $z = \{NO_2^*\}$. The universe is then gradually extended with further references and the testing methodology repeated to highlight the impact of adding references to universe U.

Fig. 9(a) shows the evolution of RMS($\boldsymbol{\varepsilon}_P$), RMS($\boldsymbol{\varepsilon}_N$) and RMS($\hat{\mathbf{u}}$) for the NO_2^* low-cost sensor, where each set of bars corresponds to a different U. The values are calculated in steps of two weeks over the whole measurement period and the height of the bars and the whiskers indicate the average and standard deviation respectively over all tests.

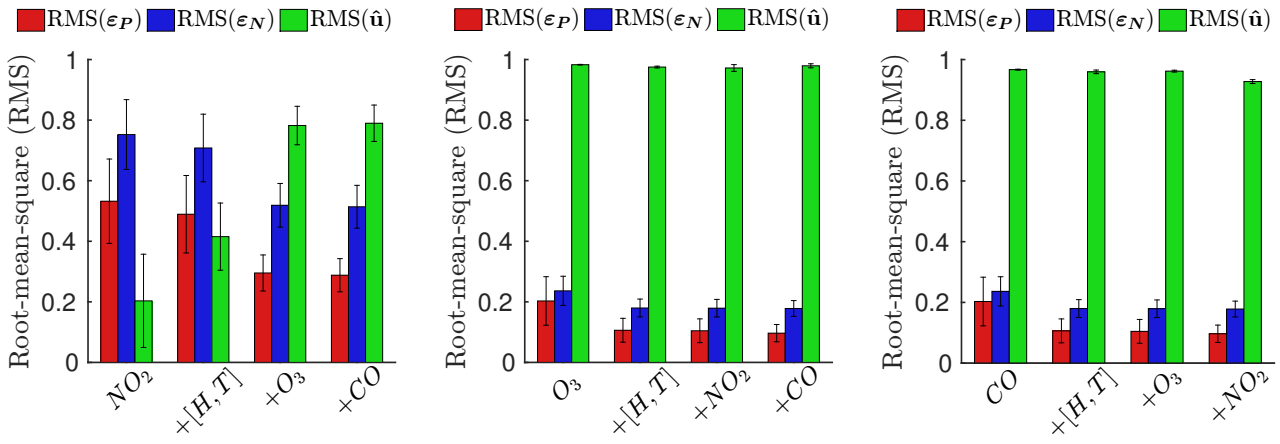
We make the following observations:

- The results for the initial $U = \{NO_2\}$ show a remarkably larger periodic error RMS($\boldsymbol{\varepsilon}_P$) and noise component RMS($\boldsymbol{\varepsilon}_N$) than the explained part RMS($\hat{\mathbf{u}}$). This finding points out that the sensor is to a large extent cross-sensitive to some phenomena beyond NO_2 and is affected by noise.
- The second set of bars shows the results when humidity and temperature measurements including their squared and cubed versions are added to $U = \{NO_2, H, T\}$. Although we observe an increase of the explained part of the sensor measurements, the unexplained low-frequency estimate RMS($\boldsymbol{\varepsilon}_P$) remains dominant. Though the sensor is influenced by meteorological effects, it still is sensitive to other phenomena, which are not included in U yet.
- Extending the observable universe $U = \{NO_2, H, T, O_3\}$ with O_3 modality decreases the unexplained part of the sensor measurements by roughly 35% and confirms the sensor's strong cross-sensitivity to O_3. This result is reflected in Fig. 8, where we observe a smaller amplitude of ε_P compared to the initial universe $U = \{NO_2\}$.
- We fail to improve the sensor performance with the addition of CO modality to the universe $U = \{NO_2, H, T, O_3, CO\}$ and, thus, conclude that our NO_2^* sensor is not impacted by a change of CO concentration in the ambient air.

In contrast to the cross-sensitive NO_2^* sensor, the results of the O_3^* (Fig. 9(b)) and CO^* sensor (Fig. 9(c)) are different. Both sensors are highly sensitive to their target gases and, hence, the initial U suffices already to explain sensor measurements to a great extent. In fact, only including temperature and humidity references in the regression lessens the unexplained part of the error. Introducing additional references does not improve the outcome of the testing procedure, because both sensors are not cross-sensitive to any of the tested interfering gases.

The above findings show that the NO_2^* sensor can be used to monitor air quality in our setting only if the measurement system additionally acquires collocated O_3 measurements to compensate for the sensor's cross-sensitivity to O_3. Moreover, all gas sensors (O_3^*, NO_2^* and CO^*) depend on meteorological conditions and, thus, should be augmented with temperature and relative humidity sensors to achieve accurate calibration.

[1] AlphaSense NO_2-B4 [5]
[2] SGX Sensortech MiCS-OZ-47 O_3 [35]
[3] AlphaSense CO-B4 [4]



(a) NO_2 sensor [5] (b) O_3 sensor [35] (c) CO sensor [4]

Figure 9. Average root-mean-square value of the periodic (low-frequent) part ε_P, noise (high-frequent) part ε_N of the regression error and the regression estimation $\hat{u}$ for three low-cost sensors. The x-axis shows the evaluation for different observable universes U.

Table 1. Comparison of calibration accuracy (mean RMSE $\pm$ standard deviation) between OLS and MLS and the $75th$ percentile of the actual pollutant concentrations. MLS outperforms OLS for all three target pollutants.

	$NO_2\,[ppb]$	$O_3\,[ppb]$	$CO\,[ppm]$
OLS	9.45 ± 2.74	4.9 ± 1.68	0.051 ± 0.018
MLS	5.13 ± 1.05	2.8 ± 0.96	0.048 ± 0.019
75^{th} perc.	22.3	30.8	0.338

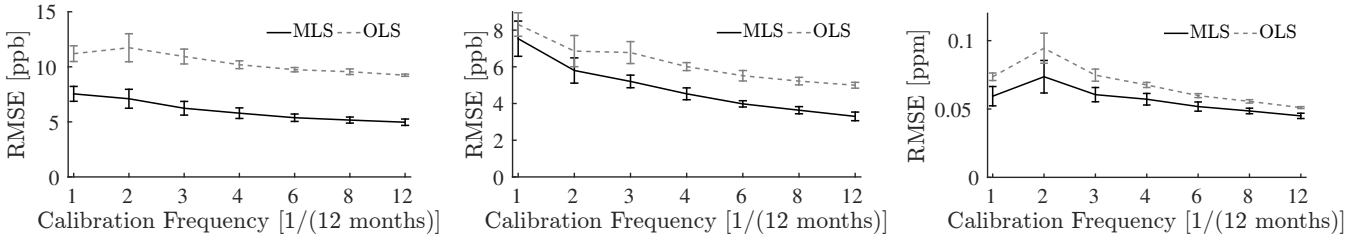
Sensor array calibration. Our novel testing methodology allows revealing cross-sensitivities and their extent for every sensor under test. These results immediately suggest the necessary measurement system augmentation to segregate mutual dependencies between cross-sensitive sensors. In the next step, the measurement system can be calibrated using MLS to measure the desired physical phenomena. Therefore, we construct a sensor array, which features the tested NO_2^*, O_3^* and CO^* low-cost sensors as well as a temperature and a relative humidity sensor. In this section, we investigate the calibration accuracy of this sensor array and compare the performance of the multiple least-squares (MLS) and ordinary least-squares (OLS) approaches. Measurements from all five sensors as well as the quadratic form of the samples are used to build M and calibrate the sensor array to the O_3, CO and NO_2 references using MLS[4]. For the OLS approach we use the sensor measurements corresponding to the reference, *i.e.*, the O_3^* sensor measurements are calibrated to the O_3 reference using OLS. The calibration parameters have been repeatedly trained with data over four weeks and used to calibrate the consecutive four weeks. The average RMSE for the OLS and MLS calibration over the whole deployment period of 15 months is summarized in Table 1. As already seen in Sec. 3.3, MLS manages to achieve accurate calibration of the cross-sensitive NO_2^* sensor and outperforms the OLS approach by up to 45%. Further, the sensor array calibration to O_3 and CO references is beneficial as well. Although both sensors have no significant cross-sensitivities to interfering gases, MLS compensates for the meteorological influences resulting in a lower calibration error compared to OLS.

These results emphasize the necessity of uncovering the sensitivity profiles of low-cost sensors by our pre-testing methodology and then augmenting the measurement system with appropriate sensors.

Calibration stability. Various works, which use an univariate calibration approach such as OLS [42, 31, 8] or calibration techniques based on artificial neural networks (ANN) [20] emphasize that low-cost sensors need frequent re-calibration, *e.g.*, in the range of one to every four weeks. We show that if our measurement system is augmented with appropriate sensors and then calibrated with MLS, it needs less-frequent re-calibration than reported above.

We compare the calibration error of MLS and OLS with different re-calibration frequencies over a period of 12 months. We calculate the regression parameters for both techniques using a training dataset of four weeks. The resulting parameters are then used to calibrate and evaluate a testing dataset between the end of the current training dataset and the end of the consecutive one. The individual training intervals are uniformly spread over a period of 12 months, *i.e.*, a calibration frequency of four re-calculates the calibration parameters every three months. The procedure is performed 12 times for each re-calibration frequency setting, where each time the start of the initial training dataset is increased by one week starting at January 10^{th} 2014. Fig. 10 shows the average RMSE for the MLS and OLS calibration to NO_2, O_3 and CO references. We observe a decreasing error with an increasing calibration frequency for all three references and both techniques. As shown before, OLS is not suited to calibrate the cross-sensitive NO_2^* sensor and conse-

[4] In contrast to the sensor testing, we do not standardize the variables for calibration.



(a) Calibration to NO_2 reference measurements (b) Calibration to O_3 reference measurements (c) Calibration to CO reference measurements

Figure 10. Average RMSE over 12 months with different calibration frequencies of multiple least-squares and ordinary least-squares. Calibrating a sensor array with multiple least-squares needs less frequent re-calibration than ordinary least-squares.

quently MLS clearly outperforms OLS for all calibration frequency settings in Fig. 10(a). Of more interest are the results of the sensor array calibration to O_3 and CO references. In Fig. 10(b) we reveal a similar error for both MLS and OLS when the calibration to the O_3 reference is only performed once at the beginning of the year. However, already calibrating twice during 12 months using MLS achieves a lower average RMSE compared to the case when OLS is used to calibrate four times in the same period. Further, calibrating four times during 12 months using MLS always outperforms OLS. We reveal a different behaviour when calibrating to the CO reference. For both techniques the average error increases when calibrating twice during 12 months. The reason for this behaviour are unstable regression parameters calculated during summer. However, OLS is never able to outperform MLS. It is necessary to re-calibrate more than four times using OLS in order to achieve the same performance when MLS is used to calibrate three times.

In conclusion, we can see that calibrating our sensor array with multiple least-squares needs less frequent re-calibration to achieve the same performance when compared to the state-of-the art ordinary least-squares based calibration.

5.2 Unqualified Sensors

Low-cost sensors available on the market may fail under certain application conditions. Fig. 11 shows test results for two sensors that can not be used to monitor air quality in our setting. The *Air quality* sensor [15] from Shenzhen Dovelet Sensors Technology is a metaloxide sensor, which is—according to its datasheet—sensitive to a long list of pollutants even at low concentrations. Unfortunately, no information about the extent of each sensitivity is provided. The *Particle* sensor [36] from Shinyei is a sensor that measures the concentration of particle matter, *i.e.*, dust particles with diameter larger than $1\,\mu m$ (*PM1*).

In Fig. 11, we observe for both sensors and each observable universe that the error components dominate and the explained part $\mathrm{RMS}(\hat{\boldsymbol{u}})$ does not exceed values above 0.6. Besides temperature and humidity references we used all pollutants measured by the static measurement station to construct the universe for testing the *Air quality* sensor. We observe that the sensor is sensitive to meteorological influences but not to any pollutants, which are helpful to explain sensor measurements. We suspect that reasons for the negative result might be the sensors' low accuracy and its design for higher pollution concentration.

On the contrary, even though the measurements of the *Particle* sensor can be partly explained with *PM1*, temperature and humidity, it has still a remarkable periodic error component with a $\mathrm{RMS}(\boldsymbol{\varepsilon}_P) \approx 0.65$. The result indicates that the sensor is impacted by a phenomenon that we failed to identify. The sensor may be useful to measure PM1 if one can find and compensate for that phenomenon in question.

6 Related Work

This section summarizes existing work on sensor selectivity, testing and calibration. We list techniques that are often used on cross-sensitive sensors to split and filter measurements, and discuss how this work advances state-of-the-art.

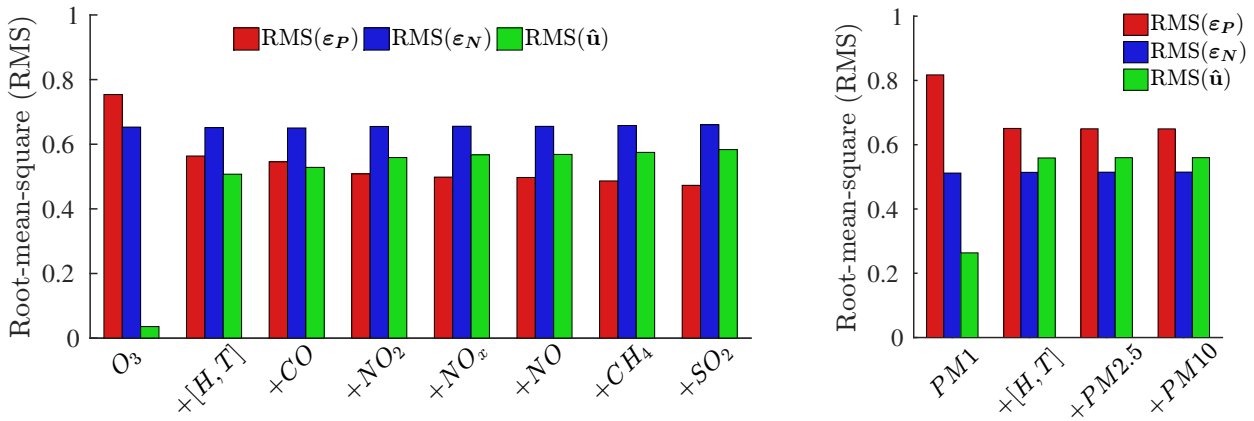
6.1 Testing Cross-Sensitive Sensors

Sensors are traditionally tested in a fully controlled environment in a laboratory [27, 24, 44]. This type of testing allows for (i) fast sensor characterization (ii) for a given interval of interest and (iii) under controlled value changes of a few parameters of interest (*e.g.*, temperature and humidity), but fixed values for all other parameters (*e.g.*, interfering pollutants) [22, 39].

Martin *et al.* [24] and Vaughn *et al.* [44] characterize the response of various air quality sensors by exposing them to different target gas concentrations under a limited number of ambient temperature and humidity settings. Although both works are able to uncover cross-sensitivities of their sensors under test, they are not able to guarantee that the list of cross-sensitivities is complete and, hence, that the sensor performs well under deployment conditions.

Morsi [27] conducts testing with the same CH_4 sensor used by Eugster and Kling [16] under laboratory conditions. Morsi uncovers various sensor cross-sensitivities—in particular to CO_2—which are not mentioned in the datasheet. These sensor cross-sensitivities are however not affecting the sensor behaviour under the specific deployment conditions in [16].

The examples above motivate the need of testing a specific sensor on its feasibility in a given scenario. We need to uncover the sensor's sensitivity, environmental dependency and operating range and test whether these match application requirements. Consequently, latest works on air quality monitoring favour in-field sensor testing against reference



(a) *Air quality* sensor [15]

(b) *Particle* sensor [36]

Figure 11. Average root-mean-square value of the periodic (low-frequent) part ε_P, noise (high-frequent) part ε_N of the regression error and the regression estimation $\hat{u}$ for two unqualified low-cost sensors. The x-axis shows the evaluation for different observable universes U.

sensors over laboratory tests [37, 41, 30]. In-field sensor testing allows sensor evaluation at a wide range of varying environmental conditions and pollutant concentrations under deployment conditions.

Spinelle *et al.* [37] test different low-cost sensors with collocated measurements at an official reference station by applying MLS. They use sensor and additional reference measurements as regressors to regress the targeted reference trace. They infer sensor feasibility from the coefficient of determination (R^2) of the regression. However, they do not investigate the reasons for low R^2 values. Our testing methodology finds all sensor cross-sensitivities given a set of reference measurements and clearly indicates if the sensor is cross-sensitive to phenomena, which are not included in the reference traces.

Traversa *et al.* [41] build their own metaloxide gas sensor and use in-field testing to show preliminary feasibility of their sensor for air quality monitoring.

Piedrahita *et al.* [30] test various low-cost gas sensors using measurements collected in a laboratory and collocated to an official air monitoring system. They were able to detect different sensor responses between tests during winter and summer and point out the advantages of in-field testing over laboratory testing. Although they use a multiple regression approach, they only consider temperature and humidity and neglect any possible cross-sensitivities to pollutants.

This paper advocates in-field sensor evaluation at reference stations located at the deployment site or featuring close to deployment environmental conditions and air composition. We use multiple least-squares (MLS)[5] to decide on sensor qualification for the environment of interest. Further, we draw conclusions about the unexplained part of the sensor measurements by leveraging frequency characteristics of physical phenomena.

[5] *Multiple least-squares* is sometimes referred to as *multivariable least-squares* or *multivariate least-squares* in the related literature.

6.2 Calibrating Cross-Sensitive Sensors

Low-cost sensors usually suffer from substantial deviation when compared to highly accurate references. In order to minimize this deviation sensors need to be calibrated. Under the assumption that there is a linear relationship between sensor and reference measurements a widely used approach is to apply univariate linear regression techniques, such as OLS [18]. These techniques, however, only perform well if the low-cost sensor is highly selective to the target gas. If the sensor is affected by cross-sensitivities to interfering gases or depends on meteorological conditions a multiple model should be used instead [11]. These models, however, require collocated measurements of additional sensors—often referred to as sensor array—to compensate for sensor cross-sensitivities. Popular approaches to calibrate a sensor array to reference measurements can be classified into multiple linear regression techniques, such as MLS [20, 28, 47, 9, 43], and artificial neural networks (ANN) based techniques [46, 45, 38, 12, 32]. Calibration can either be conducted in a laboratory setup or in-field under deployment related conditions.

Multiple regression. Kamionka *et al.* [20] simulate typical urban air compositions, *e.g.*, automotive traffic pollution, in a laboratory setup and apply a multiple regression approach to calibrate a metaloxide sensor. Although the calibration performs good, the results are not representative for outdoor applications. This is a common problem with laboratory calibration approaches due to a complex urban air composition. Consequently, a large body of work tackles the problem of calibrating sensors under real application conditions during deployment.

Morsi [28] uses MLS to calibrate gas sensor readings for meteorological effects with collocated measurements of temperature, humidity and wind speed. Due to a possible non-linear response to the meteorological effects, Morsi also includes quadratic values of the samples in the regression equation.

Zhang and Wang [47] use a sensor array consisting of

eight different gas sensors to derive the quality of peaches. The sensor array is monitoring the air in a storage. They apply MLS to calibrate the sensor array on a set of quality characteristics including sugar content and *pH*.

Bayer *et al.* [9] derive a large feature set of spectral characteristics of soil using spectroscopy and apply a multiple regression approach to calibrate the feature set on typical soil constituents such as organic carbon or iron oxides. The presented problem is tightly linked to sensor cross-sensitivities, because the derived spectral features usually overlap and are influenced by a variety of physical soil properties.

Artificial Neural Networks. The increasing popularity of ANNs in the last few years is responsible for a large body of work that tackles the challenge to calibrate cross-sensitive sensors using different types of ANNs. Calibration based on back-propagation ANNs [46, 45] and multilayer perceptrons [38, 32] successfully solve the cross-sensitivity issue and accurately calibrate sensor arrays. Hence, it is possible to use ANNs instead of MLS. However, the link between sensor calibration and the inverse calibration used for our testing methodology is easier to interpret with MLS.

7 Conclusions

Nowadays, low-cost air quality sensors are integrated in an increasing number of measurement platforms for air quality monitoring. Calibrating these sensors to reference measurements is however challenging. These sensors typically suffer from cross-sensitivities, poor stability and sensor noise. Information about these limiting effects is often not provided by the manufacturers. Even if the information is given in a datasheet, it is often scarce and reflects sensor performance under laboratory conditions. Neglecting sensor cross-sensitivities and deployment settings usually results in poor sensor performance, frequent calibration necessity and calibration failures. This arises the need for pre-deployment sensor testing under application conditions.

In this work, we present an in-field sensor testing and augmentation methodology and calibration for low-cost possibly cross-sensitive sensors. Our novel algorithm is based on multiple least-squares and uses collocated measurements of low-cost sensors and various reference sensors to quantify the amount of captured and uncaptured cross-sensitivities, and substantial sensor noise. With the obtained testing results we are able (i) to conclude about the usability of a given sensor under test in the given setting, (ii) identify fundamental cross-sensitivities and compute the sensor array, which can optimally measure a specific pollutant and (iii) compute calibration parameters that provide accurate measurements with long-term stability. We extensively evaluate our algorithm with various low-cost sensors using a dataset of 9 million sensor measurements and show the improved accuracy and long-term parameter stability when calibrating an augmented sensor array to reference measurements. We believe the proposed algorithm can be an essential step in the design of measurement platforms with low-cost cross-sensitive sensors.

Acknowledgment

We thank our shepherd Oliver Amft for his valuable feedback. This work was funded by NanoTera.ch with Swiss Confederation financing.

8 References

[1] Air Quality Egg: community-led sensing network. http://airqualityegg.com.

[2] Data Canvas: Sense Your City sensing network. http://datacanvas.org.

[3] Gas detection for automotive pollution control. *Sensors and Actuators B: Chemical*, 59(23):195 – 202, 1999.

[4] Alphasense. CO-B4 4-Electrode carbon monoxide sensor (datasheet). http://goo.gl/egp6Sm, 2014.

[5] Alphasense. NO2-B4 4-Electrode nitrogen dioxide sensor (datasheet). http://goo.gl/IpmhPb, 2014.

[6] Amphenol Advanced Sensors. Telaire 6703 series CO2 module (datasheet). http://goo.gl/C4FePf, 2015.

[7] P. Aoki, R. J. Honicky, A. Mainwaring, C. Myers, E. Paulos, S. Subramanian, and A. Woodruff. A vehicle for research: Using street sweepers to explore the landscape of environmental community action. In *SIGCHI*, pages 375–384, 2009.

[8] L. Balzano and R. Nowak. Blind calibration of sensor networks. In *IPSN*, pages 79–88, 2007.

[9] A. Bayer, M. Bachmann, A. Mueller, and H. Kaufmann. A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded south african ecosystem. *Applied and Environmental Soil Science*, 2012:20, 2012.

[10] J. Bring. How to standardize regression coefficients. *The American Statistician*, 48(3):209–213, 1994.

[11] R. Bro. Multivariate calibration what is in chemometrics for the analytical chemist? *Analytica Chimica Acta*, 500:185–194, 2003.

[12] K. Brudzewski and S. Osowski. Gas analysis system composed of a solid-state sensor array and hybrid neural network structure. *Sensors and Actuators B: Chemical*, 55(1):38 – 46, 1999.

[13] S. Chen, J. S. Brantley, T. Kim, and J. Lach. Characterizing and minimizing synchronization and calibration errors in inertial body sensor networks. In *BodyNets*, 2010.

[14] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang. Aircloud: A cloud-based air-quality monitoring system for everyone. In *SenSys*, 2014.

[15] Dovelet. TP-401A indoor air quality gas sensor (datasheet). http://goo.gl/AVAQJ6.

[16] W. Eugster and G. W. Kling. Performance of a low-cost methane sensor for ambient concentration measurements in preliminary studies. *Atmospheric Measurement Techniques*, 5(8):1925–1934, 2012.

[17] EveryAware Project Consortium. Report on: sensor selection, calibration and testing; EveryAware platform; smartphone applications. http://goo.gl/ckqp36, 2012.

[18] D. Hasenfratz, O. Saukh, and L. Thiele. On-the-fly calibration of low-cost gas sensors. In *EWSN*, pages 228–244, 2012.

[19] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, and L. Thiele. Pushing the spatio-temporal resolution limit of urban air pollution maps. In *PerCom*, pages 69–77, 2014.

[20] M. Kamionka, P. Breuil, and C. Pijolat. Calibration of a multivariate gas sensing device for atmospheric pollution measurement. *Sensors and Actuators B: Chemical*, 118(12):323–327, 2006.

[21] G. Korotcenkov, I. Blinov, V. Brinzari, and J. Stetter. Effect of air humidity on gas response of SnO2 thin film ozone sensors. *Sensors and Actuators B: Chemical*, 122(2):519 – 526, 2007.

[22] S. Laurent, G. Michel, and A. Manuel. Laboratory and in-situ validation of ozone microsensors, α-sense, model B4 O3 sensors. Technical report, Publications Office of the European Union, 2014.

[23] J. J. Li, B. Faltings, O. Saukh, D. Hasenfratz, and J. Beutel. Sensing the air we breathe – the OpenSense Zurich dataset. In *AAAI Conference on Artificial Intelligence*, 2012.

[24] M. A. Martin, J. Santos, H. Vsquez, and J. Agapito. Study of the interferences of NO2 and CO in solid state commercial sensors. *Sensors and Actuators B: Chemical*, 58(13):469 – 473, 1999.

[25] M. Mead, O. Popoola, G. Stewart, P. Landshoff, M. Calleja, M. Hayes, J. Baldovi, M. McLeod, T. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J. Saffell, and R. Jones. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203, 2013.

[26] E. Miluzzo, N. D. Lane, A. T. Campbell, and R. Olfati-Saber. CaliBree: A self-calibration system for mobile sensor networks. In *DCOSS*, pages 314–331, 2008.

[27] I. Morsi. A microcontroller based on multi sensors data fusion and artificial intelligent technique for gas identification. In *IECON*, 2007.

[28] I. Morsi. Electronic noses for monitoring environmental pollution and building regression model. In *IECON*, 2008.

[29] U. Nyffeler. Das Nationale Beobachtungsnetz für Luftfremdstoffe. In *BUWAL*, 2001.

[30] R. Piedrahita, Y. Xiang, N. Masson, J. Ortega, A. Collier, Y. Jiang, K. Li, R. P. Dick, Q. Lv, M. Hannigan, and L. Shang. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques*, 7(10):3325–3336, 2014.

[31] N. Ramanathan, L. Balzano, M. Burt, D. Estrin, T. Harmon, C. Harvey, J. Jay, E. Kohler, S. Rothenberg, and M. Srivastava. Rapid deployment with confidence: Calibration and fault detection in environmental sensor networks. *TR 62 CENS*, 2006.

[32] D. Reza Nadafi, S. Nejad, M. Kabganian, and F. Barazandeh. Neural network calibration of a semiconductor metal oxide micro smell sensor. In *DTIP*, pages 154–157, 2010.

[33] A. Ribas and J. Peuelas. Temporal patterns of surface ozone levels in different habitats of the north western mediterranean basin. *Atmospheric Environment*, 38(7):985 – 992, 2004.

[34] O. Saukh, D. Hasenfratz, and L. Thiele. Reducing multi-hop calibration errors in large-scale mobile sensor networks. In *IPSN*, pages 274–285, 2015.

[35] SGX Sensortech. MiCS-OZ-47 ozone sensor (datasheet). `http://goo.gl/C49tcw`, 2014.

[36] Shinyei. PPD42NS Particle Sensor (datasheet). `http://goo.gl/2btsP3`, 2010.

[37] L. Spinelle, M. Gerboles, M. Villani, M. Aleixandre, and F. Bonavitacola. Calibration of a cluster of low-cost sensors for the measurement of air pollution in ambient air. In *SENSORS, 2014 IEEE*, pages 21–24, Nov 2014.

[38] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola. 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*, 215:249 – 257, 2015.

[39] A. Szpakowski, C. Tyszkiewicz, and T. Pustelny. Multivariate analysis in gas sensing applications. *Acta Physica Polonica A*, 114:239–242, 2008.

[40] A. A. Tomchenko, G. P. Harmer, B. T. Marquis, and J. W. Allen. Semiconducting metal oxide sensor array for the selective detection of combustion gases. *Sensors and Actuators B: Chemical*, 93(13):126 – 134, 2003.

[41] E. Traversa, Y. Sadaoka, M. C. Carotta, and G. Martinelli. Environmental monitoring field tests using screen-printed thick-film sensors based on semiconducting oxides. *Sensors and Actuators B: Chemical*, 65(13):181 – 185, 2000.

[42] W. Tsujita, A. Yoshino, H. Ishida, and T. Moriizumi. Gas sensor network for air-pollution monitoring. *Sensors and Actuators B: Chemical*, 110(2):304 – 311, 2005.

[43] S. Vaihinger and W. Goepel. *Multi-Component Analysis in Chemical Sensing*, pages 191–237. 2008.

[44] D. L. Vaughn, T. S. Dye, P. T. Roberts, A. E. Ray, and J. L. DeWinter. Characterization of low-cost NO2 sensors. Technical report, Sonoma Technology, Inc., 2010.

[45] S. D. Vito, E. Massera, M. Piga, L. Martinotto, and G. D. Francia. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129:750 – 757, 2008.

[46] S. D. Vito, M. Piga, L. Martinotto, and G. D. Francia. CO, NO2 and NOx urban pollution monitoring with on- field calibrated electronic nose by automatic bayesian regularization. *Sensors and Actuators*, 143:182–191, 2009.

[47] H. Zhang and J. Wang. Evaluation of peach quality attribute using an electronic nose. In *Sensors and Materials, Vol. 21*, 2009.