Self-Calibration Methods for Uncontrolled Environments in Sensor Networks: A Reference Survey

Jose M. Barcelo-Ordinas[a,1,*], Messaud Doudou[b], Jorge Garcia-Vidal[a,1], Nadjib Badache[c]

[a] *Universitat Politecnica de Catalunya, 08034 Barcelona, Spain*
[b] *Sorbonne universits, Universit de technologie de Compigne, Heudiasyc Lab., Compigne, France*
[c] *CERIST Research Center, Algiers, Algeria*

Abstract

Growing progress in sensor technology has constantly expanded the number and range of low-cost, small, and portable sensors on the market, increasing the number and type of physical phenomena that can be measured with wirelessly connected sensors. Large-scale deployments of wireless sensor networks (WSN) involving hundreds or thousands of devices and limited budgets often constrain the choice of sensing hardware, which generally has reduced accuracy, precision, and reliability. Therefore, it is challenging to achieve good data quality and maintain error-free measurements during the whole system lifetime. Self-calibration or recalibration in ad hoc sensor networks to preserve data quality is essential, yet challenging, for several reasons, such as the existence of random noise and the absence of suitable general models. Calibration performed in the field, without accurate and controlled instrumentation, is said to be in an uncontrolled environment. This paper provides current and fundamental self-calibration approaches and models for wireless sensor networks in uncontrolled environments.

Keywords: Wireless Sensor Networks (WSN), Calibration, Low-cost sensors, Uncontrolled environments, Quality of Information (QoI)

1. Introduction

In recent years there has been a growing interest in the use of networks of monitoring nodes based on low-cost sensor technology. The monitored parameters measured by sensors include air quality parameters (e.g., concentration of gases such as NO_2, NO, CO, CO_2, and O_3, or concentration of particulate matter, such as PM2 or PM10), environmental parameters (e.g., temperature, humidity, radioactivity), noise, wind, tracking of objects or persons, presence of animals, and perimetric security. Many of these sensors are sold by manufacturers without individual calibration parameters except for some generic calibration values in the data-sheet description of sensor devices. In some situations, the sensing devices were individually calibrated by the manufacturer, usually under controlled conditions. In any case, calibration, either during the pre-installation phase or during the monitoring network operation, is necessary in order to achieve the data quality requirements.

Sensor calibration in the deployment of large-scale cyber-physical IoT systems and ad hoc sensor networks is mandatory due to the inherent process of device imperfection and noise in the massive data collected. In recent years there has been a growing interest in applying all the theoretical knowledge acquired in wireless sensor networks (WSN) to commercial or research deployments. Much of this research has focused on challenges related to communication protocols suited to WSN and the management of their energy consumption [1, 2, 3, 4]. From the very beginning, some works paid attention to the quality of the sensing data, [5, 6, 7, 8] but most of them focused on simple data applications, such as temperature or relative humidity, since these were the types of data measured by the first available sensors. Others [9, 10] used synthetic data generated with probability functions, such as Gaussian distribution functions, for validating calibration models. However, the concern for the quality of the data has increased with the growth of real wireless sensor deployments. For example, Buonadonna et al. [11] mentioned the disappointment when the values of light sensors (calibrated according to the manufacturer's specification) in a deployment were compared with the value from a calibrated sensor in an environmentally controlled chamber. Ramanathan et al. [12] described the difficulties in obtaining realistic readings of ion-selective electrode sensors used for monitoring water quality in a real deployment in Bangladesh. More recently, Snyder et al. [13] and Williams [14] mentioned the lack of knowledge on the performance and long-term reliability of low-cost air pollution sensors.

The classical calibration process mainly consists of calibrating the sensor device in a *controlled environment*, for example, in a laboratory with high-cost instrumentation, where the sensor response is measured under different controlled conditions. When it is not possible to use such a laboratory, or when sensor

*Corresponding author
Email addresses: joseb@ac.upc.edu (Jose M. Barcelo-Ordinas), messaoud.doudou@hds.utc.fr (Messaud Doudou), jorge@ac.upc.edu (Jorge Garcia-Vidal), badache@cerist.dz (Nadjib Badache)

[1] Declarations of interest: none

devices are under operation after deployment, the sensor parameters can be self-calibrated and adjusted in reference to another sensor of the network, whether calibrated with a ground-truth[2] [15, 16] reference node, calibrated with respect to already calibrated sensor nodes (e.g., distributed calibration), [17] or with respect to not-calibrated sensor nodes (e.g., blind calibration) [18].

When the sensor response cannot be measured in a controlled environment and the sensor parameters have to be adjusted according to other sensor nodes in the network, the calibration is said to be in an *uncontrolled environment*. In this case, poor or incomplete calibration can lead to significant errors in sensor measurements. Moreover, even if the sensors are calibrated before deployment, it is not possible to prevent sensor drift after deployment, especially when the lifetime of sensor systems can be as long as years. In some cases, even costly high-accuracy sensors produce faulty data when aging. Therefore, it is necessary to automatically detect drifts and miscalibration in order to correct sensors' measurement after deployment to ensure the trustworthiness of long-term WSNs.

While important, ensuring data quality in WSNs presents several challenges [19]. The quantity of data is large, but the quality is low, since the data can be subject to many types of faults. In a real sensor network deployment, Buonadonna et al. [11] observed that failures can occur in unexpected ways which provide inaccurate data. Koushanfar et al. [20] have identified several facets of faulty data and suggested five phases of testing sensor-based systems. Ni et al. [21] defined *data faults* to be data reported by a sensor that are inconsistent with the phenomenon of interest. Measurement faults can be produced by systematic and random errors. Each measurement has unpredictable random errors due to environmental noise, the precision of the equipment, or sensor manufacturing defects.

This survey provides insight into existing calibration approaches, models, and methods that have been proposed for an efficient calibration process when expensive equipment for controlled calibration is not available, or when the sensor nodes are already deployed in the field. Calibration is a technically challenging task mainly due to the existence of random noise and the absence of suitable general models. Calibration techniques and models depend on the physical space of the signal measured, on the availability of ground-truth data, and on the capabilities of the sensor nodes. Although there is a rich variety of papers related to calibration models and algorithms for different sensors, we have not found any survey that has collected the definitions, approaches, and models used for calibrating sensors in WSNs. The closest research area in which part of this knowledge is collected is sensor faults research, [21, 22] where calibration is considered one of the many causes of wireless sensor node faults.

Our main contributions are the following:

- We first define calibration in uncontrolled environments, stressing that an uncontrolled environment involves calibration in the field and not necessarily in the presence of accurate instrumentation. Moreover, we state the main challenges in calibrating low-cost sensors in uncontrolled environments in WSNs.
- We present calibration models and the difficulty in choosing a specific calibration model in the calibration process.
- We state the quality of information (QoI) metric generally used to evaluate the performance of a calibration process.
- We define calibration approaches and attributes in order to calibrate a sensor. These include specifying the conditions that the calibration process follows, specifically, what the measurement area is, how many sensors are involved in the calibration process, what the knowledge of the physical phenomenon is, what the position of the calibration sensor is with respect to reference data when reference data exist, how many times the calibration process is performed, how the information is processed with respect to the normal application operations, and where the calibration is performed.
- Finally, we present guidelines on how to calibrate a sensor network.

The aim of this survey, then, is to provide a systematically characterized taxonomy of approaches, attributes, and models for the calibration of sensors in WSN in uncontrolled environments. Sections 2, 3, and 4 present the basics of sensor calibration, calibration models, and different validation metrics used to evaluate the accuracy of a calibration model. Section 5 focuses on the different attributes and approaches of a calibration process. Section 6 gives guidelines in how to calibrate a low-cost sensor network. Section 7 discusses the existing research challenges in the field of calibration in uncontrolled environments. Finally, Section 8 summarizes and concludes the paper.

2. Background on Calibration

By definition, calibration refers to the process of correcting systematic errors (i.e., biases) in sensor readings often by comparing a known measure from a first device with an unknown measure of a second device to adjust the parameters that rule this second device, in order to provide an accurate measurement. The term has also often been used to refer to the process of adjusting the raw sensor readings to obtain corrected values by mapping them into standardized units. Traditional single-sensor calibration often relies on providing a specific stimulus with a known result, thus creating a direct mapping between sensor outputs and expected values. Consequently, such calibration for a sensor is often subject to specific ranges and operating-condition restrictions, which are reported in the manufacturer's specifications of the sensor. This type of calibration can be performed in the factory, during the production stage, manually in the field, or both. In addition to component-level calibrations, sensors usually must be calibrated at the device

[2]The *ground-truth value* is the reference value obtained from a perfectly calibrated sensor. In machine learning, the *ground-truth value* refers to the accuracy of the training set's classification for supervised learning techniques.

level when used as part of a measurement system. Moreover, recalibration is usually required in order to ensure proper operation of a measurement device, as aging and other factors affect sensors and other measurement hardware over time. In wireless sensor networks, when this process is performed in the field, in the absence of an environmentally controlled chamber, we call it self-calibration in an uncontrolled environment[3].

2.1. Calibration Challenges

Calibration in sensor networks is challenging because of several reasons [5, 6, 7, 11]. First, the sensor system network consists of a large number of devices typically with no calibration interface. Therefore, in-place sensor calibration schemes become impractical, time consuming, and difficult to achieve. Moreover, sensor nodes are exposed to environmental noise and hardware failures, and the mismatch between factory calibration conditions and in-field conditions makes the calibration of sensors challenging. In addition, different sensors require different calibration procedures, and the reference values might not be readily available. So, in general, even when the response function is known, calibration in uncontrolled environments is difficult. An example is an ozone (O_3) sensor whose output value depends not only on the O_3 concentrations but also on the environmental conditions – temperature and relative humidity – of the place in which the sensor is deployed. Conducting a multiple linear regression with respect to ground-truth O_3 concentrations in a place with different environmental conditions than those of the place in which the sensor node will be deployed can produce large errors in the predicted O_3 concentrations [16]. Yamamoto et al. [23] have recently observed a similar behavior in temperature sensors. Temperature sensors calibrated in a place behave differently when placed in another location because of the difference in environmental conditions (such as solar radiation, humidity, wind speed, rainfall, and azimuth) between the two locations.

In large-scale networks, two more challenges add to the inherent difficulties of calibrating a sensor: (i) the need to calibrate a massive number of sensors and (ii) the inconveniences of physically accessing the sensors, as they may be deployed in far and harsh, or even hostile, environments. Additionally, the sensors are presumed to stay active for long periods of time after deployment; therefore, they are expected to be checked regularly against standard instruments to ensure the measurements' quality and to allow periodic recalibration necessary because of the loss of accuracy caused by environmental conditions or internal defects. However, calibration is expected to minimize systematic and random errors; increase the accuracy of sensor readings with respect to the reference model; and manage the aforementioned constraints, or requirements, in addition to the limited capability of available sensors to provide accurate data at a low cost and without overhead mechanisms. Hence, a good calibration process should consider the following aspects [24]: i) time and monetary cost, ii) disruption to normal operation, iii) access to sensors in places difficult to access, and iv) calibration of a large number of sensors in the field. For example, in many large cities, sensors are installed in traffic lights or lamp-posts. Installing these sensors implies stopping vehicle traffic, and access to these sensors is difficult, costly, and time consuming.

2.2. Calibration Faults

Sensor error is defined as an unexpected value from the sensor device after its deployment. This type of error comprises precision degradation, reading bias, drift, noise, or sensor failure, which is generally due to miscalibration. The miscalibration error, also known as *measurement error*, can be defined as the difference between the values indicated by a sensor device and the true values provided by a reference model. Data faults due to miscalibration can manifest in different forms by lowering the accuracy of sensor measurement, its precision, or both. Usually, these errors are referred to as bias, gain, drift, or offset. The *offset, gain, and drift* are the well-known miscalibration errors that were first identified in [6, 12, 18] and discussed by [21, 22] as the main three different forms of sensor calibration faults in uncontrolled environments. *Offset* arises when the measured value Y differs from the true value X by a constant amount β_0, and it can be determined by measuring the sensed value when the ground-truth value is zero. When *offset* error is present, the measured values follow similar patterns to those of the expected phenomenon. *Gain* refers to the rate or the amount of change of the measured value with respect to the change in the underlying ground-truth value. In the specific case of a linear response function, the measured value is expressed by the following equation:

$$Y = f(\beta, X) = \beta_0 + \beta_1 X. \quad (1)$$

Here, the coefficients β_0 and β_1 represent the offset and gain parameters, respectively. The term *bias* refers to deviations that are systematic, not random, for example, because one can consistently over- or underestimate the measurements by X units. Bias, then, can be employed to express the gain, the offset, or both of them in sensor reading. *Drift* refers to the change over time of the gain and offset parameters associated with a sensor's original, or factory, calibration formulas, that causes the performance of the sensor to deviate from the real signal of the expected phenomenon. Honicky [25] observed that when a sensor is miscalibrated, its gain drift can often be bounded to a certain percentage range. For example, the gain of electrochemical sensors can drift no more than 5% per year, according to the data sheet [26]. Finally, *noise* refers to the random errors in sensor measurements that interfere with the calibration process and alter the calibration parameters. In calibration, the noise error ϵ is usually expressed as follows:

$$Y = f(\beta, X) + \epsilon, \quad (2)$$

[3] Henceforth, throughout the paper, whenever we speak of calibration, we refer to self-calibration in uncontrolled environments in WSNs. We stress that accurate instrumentation can be used in the calibration process in uncontrolled environments. The difference with respect to controlled environments is that in controlled environments, the calibration conditions can be manipulated manually (in a controlled chamber) to match the manufacturer's data-sheet conditions, while in non-controlled environments, there may be instrumentation that provides reference information but depends on conditions that cannot be manipulated.

where random noise often follows a probability distribution, for example, $\epsilon \sim N(0, \sigma^2)$ normal distribution with zero mean and variance σ^2, and most calibration approaches exploit such property to recover from random errors.

3. Calibration Model

Defining the calibration model is fundamental, yet its choice affects calibration accuracy. Sensors can be characterized by a specific *response function* relating to the measured parameters[4], defined by the set X, with the output parameter[5] defined as Y, that is $Y = f(\beta, X)$, where β is the calibration coefficient. The calibration curve is obtained by fitting an appropriate response function to a set of experimental data consisting of the measured signal relative to a known ground-truth signal. Depending on the type of phenomenon being measured, the response function associated with the calibration curve may be linear, logarithmic, exponential, or any other appropriate mathematical model. The following are the representative response functions found in sensor literature.

3.1. Linear Functions

Linear response is the most popular function in which the ground-truth value y is represented as a linear function of the measured value x, as follows [6, 11, 18]:

$$y \sim f(\beta, x) = \beta_0 + \beta_1 x, \tag{3}$$

where the calibration coefficients are offset β_0 and gain β_1. Temperature sensors[6] [6, 18] are examples of sensors in a WSN that follow linear function responses. A polynomial response function can also be used when the relationship between the response y and features x is curvilinear[7]:

$$y \sim f(\beta, x) = \beta_0 + \sum_{j=1}^{M} \beta_j x^j. \tag{4}$$

3.2. Multiple Linear Functions

In multiple linear response functions, the response variable y depends linearly on a set of parameters, as follows:

$$y \sim f(\beta, x) = \beta_0 + \sum_{j=1}^{M} \beta_j x_j. \tag{5}$$

Gas sensors, such as CO, O_3, and NO_2 sensors, are examples [15, 16, 27, 28, 29, 30, 31, 32, 33, 34] of sensors that follow multiple linear responses. Some gas sensors (see Table 1) are better fitted, with interactions of several features following a model such as [16, 27]:

$$y \sim f(\beta, x) = \beta_0 + \sum_{j=1}^{K} \beta_j x_1^{a_{1j}} \dots x_M^{a_{Mj}}. \tag{6}$$

An example [34] is an O_3 electrochemical sensor whose response depends on O_3 and temperature (T):

$$y \sim f(\beta, x) = \beta_0 + \beta_1 x_{O_3} + \beta_2 x_{O_3}^2 + \beta_3 x_T + \beta_4 x_T^2 + \beta_5 x_{O_3} x_T + \beta_5 x_{O_3}^2 x_T + \beta_6 x_{O_3} x_T^2. \tag{7}$$

3.3. Nonlinear Functions

Acoustic, seismic, and electromagnetic signals [7, 17, 35, 36] attenuate with the distance, d, from the source of the signal. Let us assume that sensor n is d_n meters from the source that transmits a signal with energy S. The attenuated signal, y_n, at the position of sensor n is [17]

$$y_n \sim f(\beta_n, x) = S\, f(\beta_n, d_n), \tag{8}$$

where $f(\beta_n, d_n)$ is a decreasing function of d_n, and β_n is a set of parameters of the signal decay function $f(\cdot)$ at sensor n. Some of the signal decay functions used are illustrated next.

Power law decay: The propagation of mechanical waves, such as acoustic and seismic signals, follows a power law decay [17]:

$$y_n \sim f(\beta_n, d_n) = S\, \frac{1}{(d_n / r_n)^{k_n}}, \tag{9}$$

where $\beta_n = \{k_n, r_n\}$; k_n is the decay parameter, and r_n is the reference distance determined by sensor shape.

Exponential decay: The intensity of light attenuates with the travel distance and follows an exponential decay [17]:

$$y_n \sim f(\beta_n, d_n) = S\, e^{\lambda_n d_n}, \tag{10}$$

where $\beta_n = \{\lambda_n\}$; λ_n is a decaying parameter, and it is referred to as the Lambert absorption coefficient.

Other phenomena are known to yield a linear response. However, when the samples are taken, due to imperfections of the sensor technology or other causes including the existence of drift or of a maximum dynamic range, the sensor response is nonlinear. For illustration of this fact, refer to Figure 1a, which shows the samples obtained from a metal-oxide ozone MIC2714 sensor of a WSN deployed in the H2020 CAPTOR project during summer 2017. The output of the uncalibrated sensor is in kiloohm. The ozone sensor is connected to a voltage divider circuit. The voltage divider has a load resistor and a variable resistor. The variable resistor varies according to the ozone concentration, which is what it is obtained at each measurement. The y-axis of Figure 1a shows the resistor values recorded at each sample time. The sampling rate is one sample per hour. The data set comprises approximately 900 samples,

[4]The input parameters are called *predictors, features, independent variables*, or *variables* in machine learning and statistical learning terminology.

[5]The output is also called *response* or *dependent variable* in machine learning and statistical learning terminology.

[6]Temperature sensors in other fields not considered WSN, such as a temperature sensor in a 3D printer, can have other function responses that are nonlinear.

[7]Polynomials on x belong to linear models since these ones are linear in coefficient β_j.

Table 1: Data models for arrays of gas sensors [16, 27].

Sensor model	Manufacturer	Gas	Multiple linear model
O3B4	αSense	O_3	$Y = \beta_0 + \beta_1 x_{O_3} + \beta_2 x_{NO_2} + \beta_3 x_{NO_2} \times x_{H_2O_2}$
O3_3E1F	Citytech	O_3	$Y = \beta_0 + \beta_1 x_{O_3} + \beta_2 x_{NO_2}$
NO2B4, NO2_3E50	αSense, Citytech	NO_2	$Y = \beta_0 + \beta_1 x_{NO_2} + \beta_2 x_{O_3} + \beta_3 x_T + \beta_4 x_{RH}$
MICS_2710	SGX-Sensotech	NO_2	$Y = \beta_0 + \beta_1 x_{NO_2} + \beta_2 x_{O_3} + \beta_3 x_T$
MICS_4514	SGX-Sensotech	NO_2	$Y = \beta_0 + \beta_1 x_{NO_2} + \beta_2 x_{O_3} + \beta_3 x_{NO} + \beta_4 x_T$
CairClip NO2	CairPol	NO_2	$Y = \beta_0 + \beta_1 x_{NO_2} + \beta_2 x_{O_3}$
NO_3E100	Citytech	NO	$Y = \beta_0 + \beta_1 x_{NO} + \beta_2 x_T + \beta_3 x_{RH}$
TGS-5042, MICS-4514	Figaro, e2V	CO	$Y = \beta_0 + \beta_1 x_{CO} + \beta_2 RH$
Gascard NG, S-100H	EdinburghSensors, ELT Sensors	CO_2	$Y = \beta_0 + \beta_1 x_{CO_2} + \beta_2 x_T + \beta_3 x_{RH}$

collected over 5.5 weeks. The WSN was collocated with a reference ground-truth node (see Section 5.4). Figure 1b shows the scatterplot of the sensor node data and reference node data. The values are normalized to their mean and standard deviation. In general, it is well-known [16, 27, 30] that ozone sensors can be calibrated using a multiple linear regression using an array of sensors (see Section 5.2 and Table 1). Figure 1c shows the calibrated sensor using a frequentist multiple linear regression (MLR). Although the error is low, $RMSE = 9.09$ μgr/m^3 with an $R^2 = 0.947$, we can observe from the scatterplot (Figure 1b) that a nonlinear model could improve the calibration error. The same sensor is calibrated using a nonlinear model called support-vector regression (SVR) having a $RMSE = 6.21$ μgr/m^3 with an $R^2 = 0.976$ in Figure 1d. This example shows that there is no golden rule for choosing a response function and that different models will calibrate the same sensor giving different QoI (quality of information), where QoI is measured in terms of the calibration error given by the model.

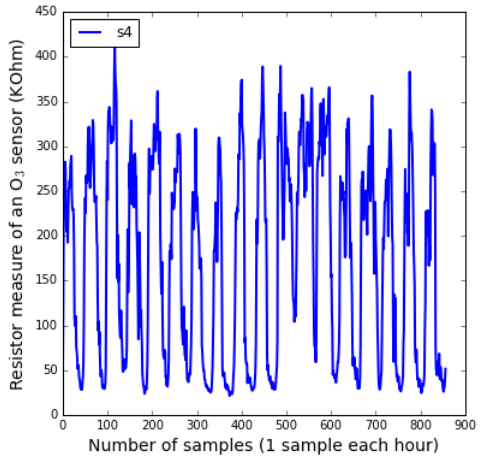
Nonlinear responses can be analyzed using a variety of techniques such as (i) maximum likelihood (ML) [17, 37] or maximum a posteriori (MAP) [38, 39, 40] in its Bayesian version, which finds the calibration parameters that better fit the data; (ii) splines [7, 35], which approximate nonlinear response functions by a continuous piecewise linear function with m segments (i.e., step functions); (iii) support vector regression (SVR) [31, 41], which classifies data by means of regression functions that depend on Lagrange multipliers and kernel functions; (iv) random forest [42], which constructs an ensemble of decision trees using a training data set; the mean value from that ensemble of decision trees is then used to predict the value for new input data; (v) artificial neural networks (ANN) [16, 27, 31], which consist of multiple layers formed by neural units. A single neuron has a number, I, of inputs, x_i, and one output, y. There are a weight, w_i, and a bias, w_0, associated with each input. The neuron response is called *activation rule* and captures the nonlinearity of the sensors.

3.4. Choosing a Model

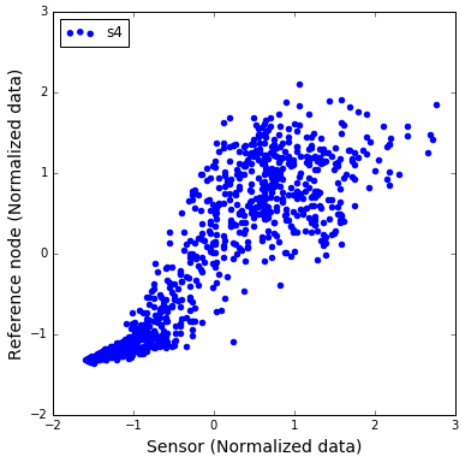
The choice of calibration model depends on the type of phenomena that the nodes measure, the resources of the wireless sensor deployment, the type of sensors, and the computation, storage, and communication capabilities of the sensor node. Moreover, the mathematical machinery that has been used in calibration in the past few years is more sophisticated when the sensor resources increase. More complex calibration approaches are being formulated with the increased number of wireless sensor network deployments with real applications. We have to mention that one of the most challenging tasks will be to directly calibrate on a low-cost node. Most low-cost wireless sensor nodes have too low capabilities to be able to be calibrated on-line. Light-computational models for low-capability nodes are one of the most difficult challenges that will have to be solved. For increasing wireless sensor capability nodes, higher computational techniques can be implemented, facilitating on-line calibration or recalibration mechanisms. For example, SVR and ANN require high computational resources that most wireless nodes do not include. In contrast, gradient descent models typically applied in optimization can be easily implemented in the node to perform on-line calibration in linear or multiple linear regression.

For instance, temperature sensors (Table 2) typically follow linear response functions and have been analyzed using linear regression [6, 11] for calibrating the sensors, Bayesian inference for modeling drift [43], and maximum likelihood [44], Gaussian processes [44], or kriging [45] for state-space modeling applications. Light point sensors have mainly been calibrated using splines [7, 35] or support-vector regression [41] due to the presence of nonlinearities, and they have been analyzed with distributed consensus [8] to show how light sensing is affected by sensor orientation. In the case of localization applications (Table 3), the most used calibration technique is maximum a posteriori [38, 39, 40] although distributed consensus [52] is also used when spatial redundancy is present. Target detection [17, 37] uses maximum likelihood to estimate detection probabilities. For synchronization applications, Stankovic et al. [10, 53] defined the distributed algorithms that converge fast to estimate the offset and gain of the calibration function. Ando et al. [54] used an iterative Bayesian algorithm for estimating the offset time of each sensor after the *k-th* event.

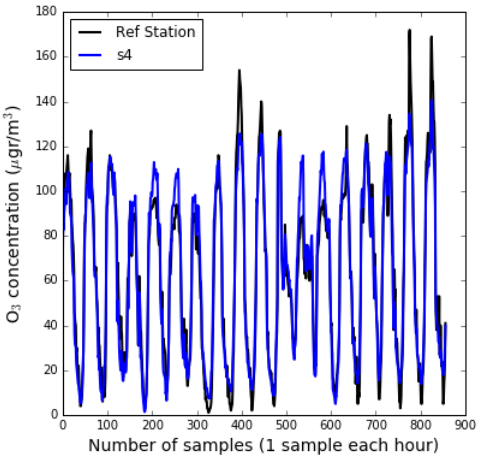
To finalize the set of examples, to monitor air pollution (Table 4), we mainly need arrays of sensors, because pollutants depend on several factors (Table 1). In general, air pollution sensors are analyzed using multiple linear regression (MLR) [12, 15, 28, 30, 31, 33, 55], although when nonlinearities due to the chemical composition of the sensor appear, techniques such as artificial neural networks (ANN), support-vector regression (SVR) or random forest are used [16, 27, 31, 42].



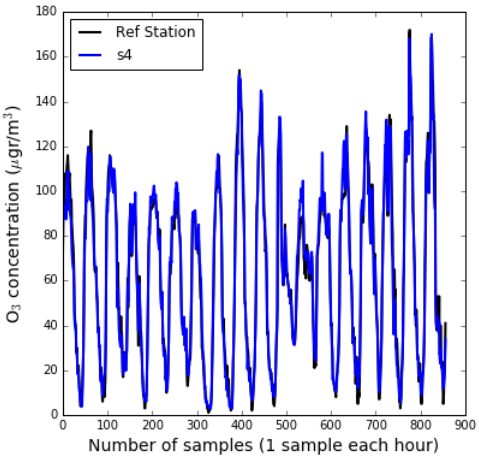
(a) Resistor values recorded by the sensor node.



(b) Scatterplot: x-axis with normalized values recorded by the sensor; y-axis with normalized values given by the reference node.



(c) Calibrated using multiple linear regression (MLR).



(d) Calibrated using support-vector regression (SVR).

Figure 1: H2020 CAPTOR wireless sensor node measuring ozone concentrations.

3.5. Open Calibration Data Sets

There is no centralized repository of open data sets for the calibration of low-cost sensors. Most of the data sets found on the Internet are data collected by sensors operating on real networks. For example, CRAWDAD[8] contains data from sensors such as temperature, accelerometers, location, or RF signal strength sensors. The major drawback of these repositories is that the data they contain are from applications and are not data collected in the calibration phase. This makes it difficult in most cases to use these data to investigate algorithms or calibration mechanisms.

In the field of calibration, some authors have published data obtained from experiments or deployments designed for the calibration of sensors at WSNs. For example, the UCI machine learning repository[9] contains some data sets related to the machine learning algorithms used in the calibration of low-cost sensors. Examples are the data set published by Fonollosa et al. [57] used in their investigation of calibration of chemical sensors [58, 59] and the data set published by De Vito et al. [60, 61] and used in the calibration of air pollution sensors.

4. Accuracy of the Model

Given the twofold goal of a calibration process – to recover the true values of a phenomenon and to detect sensory faults

[8] https://crawdad.org/

[9] https://archive.ics.uci.edu/ml/

Table 2: Simple applications include light, temperature, relative humidity, vibration, and accelerometer sensors.

References	Calibration model	Evaluation metric	Data set	Application
[6]	Linear regression	Mean error	Real (thermocouple)	Temperature
[7]	Splines/optimization	Confidence interval	Real (photovoltaic)	Point-lights
[8]	Distributed consensus	Mean error	real	Light
[11]	Linear regression	Mean error	Real	Light, temp., humidity
[18]	Linear regression	Mean & median error	Real (thermistor)	Temp., humidity
[35]	Nonlinear/splines	Mean squared error	Real	Light
[36]	Hidden Markov model	Recognition accuracy	Real	Motion (accelerometer)
[41]	Support vector regression (SVR)	Mean squared error	Real	Light
[43]	Bayesian	Root mean squared error	Synthetic	Temperature
[44]	Maximum likelihood	Absolute error	Synthetic	Temperature
[45]	Kriging	Root mean squared error	Real	Temp., humidity, light
[46, 47, 48]	Gaussian process	K-L divergence	Real	Temperature
[49]	Linear/nonlinear optimization	Mean absolute error	Real	Vibration (water flow)
[50]	Distributed consensus	Mean squared error	Synthetic	Temp., humidity, sound
[51]	PCA + compressive sensing	Mean squared error	Real	Temperature

Table 3: Localization, synchronization, and target location applications.

References	Calibration model	Evaluation metric	Data set	Application
[5]	Least squares	Mean error	Real (acoustic)	Localization
[17]	Nonlinear (maximum likelihood)	Detection probability	Synthetic/real	Target detection
[38]	Bayesian (maximum a posteriori)	Cramèr-Rao bound	Real (acoustic)	Localization
[39, 40]	Bayesian (maximum a posteriori)	K-L divergence	Synthetic	Localization
[37]	Nonlinear (maximum likelihood)	Bayes risk	Synthetic	Target detection
[52]	Distributed consensus	Mean, avg. residual	Real (RSSI)	Localization
[53, 10]	Distributed macro-calibration	Mean squared error	Synthetic	Synchronization
[54]	Bayesian (iterative)	Error variance	Synthetic	Synchronization

– a good estimation of the calibration parameters and the accuracy of the process determine the effectiveness of the model. The way to quantify the accuracy or QoI of a fitting model is by minimizing an error function [62] that measures the misfit between the output Y and the response function f(β,x) for any given value of β and the data set X. If the size of the data set Y is K, then a choice of an error function is the square of the errors between the predictions $f(\beta, x_k)$ for each data point x_k and the target value y_k:

$$E(\beta) = RSS = \sum_{n=1}^{K} (f(\beta, x_k) - y_k)^2, \tag{11}$$

where RSS stands for residual sum of squares. Then, the goal is to minimize the error function

$$\begin{array}{ll} \min & E(\beta) \\ \text{var.} & \beta \end{array} . \tag{12}$$

The calibration parameters, β^*, are the solution of the minimization problem. The error is positive except whenever the function $f(\beta^*, x_k)$ passes exactly through each target point y_k, in which case the error is zero. The following QoI metrics can be defined.

(i) *The mean squared error (MSE)* measures the average of the squares of the errors. It is the second moment (about the origin) of the error and thus incorporates the variance of the calibration curve.

(ii) *The root mean squared error (RMSE)* allows comparing different sizes of data sets, because it is measured on the same scale as the target value y_k.

(iii) *The coefficient of determination (R^2)* measures the proportion of variability in Y that can be explained using X, and it is bounded between 0 and 1. A value of R^2 close to 1 indicates that a large proportion of the variability in the response has been explained by the regression.

(iv) *The normalized mean bias error (NMBE)* represents effectively a total percent error. The use of percent differences rather than absolute differences normalizes the size of errors in the calibrated measurements.

(v) *The mean absolute error (MAE)* is a quantity used to measure how close the calibrated measurements are to the ground-truth data.

Other times, the calibration parameters are represented by a probability distribution. An example is a response function that follows a normal distribution of mean y_k=$f(\beta, x_k)$ and variance σ^2. Now, the objective is to obtain a distribution with mean $\hat{\beta}$ and variance $\hat{\sigma}^2_{\beta}$. In this case, other QoI metrics can also be obtained.

(vi) *The Cramèr-Rao bound (CRB), or information inequality*, expresses a lower bound on the variance of a parameter's estimators.

There are metrics that allow measuring the accuracy or precision of the estimation and depend on the algorithm or application defined by the authors. One example is the Kullback-Leibler (KL) divergence, which measures how one probability distribution diverges from a second expected probability distribution [40, 48]. Other examples are the average value of the relative bias error (RBE), the average value of the relative deviation error (RDE) [56], and the Bayes risk in event detection applications [37].

Table 4: Air pollution and water chemistry applications.

References	Calibration model	Evaluation metric	Data set	Application
[12]	Multiple linear regression, splines	Coeff. of determination	Real	Water chemistry
[15]	Multiple linear regression	Mean absolute error	Real	O_3
[16, 27]	Multiple linear regression, neural networks	Mean bias error	Real	O_3, CO, CO_2,NO, NO_2
[28]	Bayesian multiple linear regression	Residual sum of squares	Real	CO, CH_4, C_3H_8, CeO_2, NiO
[29]	Linear and geometric regression	Root mean squared error	Real	O_3, CO
[30]	Multiple linear regression	Root mean squared error	Real	O_3, CO, NO_2
[31]	Multiple linear regression, Gaussian processes, neural networks, support vector regression	Mean absolute error	Real	O_3, NO, NO_2
[32]	Optimization	Mean squared error	Real	O_3, CO, NO_2
[33]	Multiple linear regression	Root mean squared error	Real	O_3
[55]	Multiple linear regression	Mean error	Real	Photosynthetically active radiation (PAR)
[56]	Maximum likelihood	Relative bias error	Synthetic	Air pollution

Tables 2, 3, and 4 show how different sensor applications use different QoI metrics to assess the performance of the sensor network. The most used metrics are MSE, $RMSE$, and R^2. Other authors use the target diagram [63] that provides information in one plot about the mean bias (MB), the standard deviation, the RMSE, the centered root mean square error (CRMSE), and the correlation coefficient, R. However, the QoI metrics in general are linked to the calibration model selected and to the application. For example, R^2 and $RMSE$, MAE, and MSE are good metrics to assess whether the assumptions of the model are correct and estimate whether the error is large.

5. Calibration Approaches and Attributes

Calibration in sensor networks has been applied to many fields including environmental and air quality monitoring [16, 29, 56], weather monitoring [6, 18, 24, 44, 46, 50, 55], localization [5, 38, 39, 52], synchronization [10, 54], target discovery [17, 35], robotic, electronic, and radio sensing [64, 65, 66], and water flow monitoring [49]. These calibration attempts have taken different *forms, or approaches,* and have employed a number of calibration *models* (section 3). A calibration *approach* is the description of the set of attributes that will characterize the calibration of the sensor. For example, some authors assume that the node has ground-truth data available, while others assume that no such data are available. Some calibration models are applied at a centralized server, while others are distributed. In general, there is a set of *attributes* that capture a specific calibration approach and that have been used and defined in the literature. We next describe the attributes of the general characteristics that a calibration approach has. In order to calibrate a wireless sensor network in uncontrolled environments, the following questions have to be answered:

- How is the measurement area? Calibration can be *micro* (i.e., performed at given points) or *macro* (i.e., performed in given areas).
- What is the number of sensors involved? Calibration can be *single* (i.e., using only one sensor) or *sensor fusion*, also called *multi-sensor data fusion*. Sensor fusion includes the case of an *array of sensors* (i.e., using multiple sensors).
- What is the knowledge of the physical phenomenon? Calibration can be *non-blind* (i.e., with full information), *semi-blind* (i.e., with partial information), or *blind* (i.e., with no information).
- What is the position of the uncalibrated nodes with respect to the ground-truth node? Calibration can be performed by nodes that are *collocated*, *multi-hop* (i.e., iterative), or *model-based*.
- How many times and when is the calibration performed? Calibration can be carried out during *pre-deployment* (i.e., before the deployment of nodes), *post-deployment* (i.e., after the deployment of nodes), *opportunistically* (i.e., whenever possible), or *periodically* (i.e., at given intervals of time).
- How is the information processed with respect to the normal operation? Calibration can be *off-line* (i.e., network not operative, normally related to calibrating with a set of data samples) or *on-line* (i.e., network operative, normally related to calibrating at each sample arrival).
- Where is the calibration performed? The process can be *centralized* (i.e., localized at a central station) or *distributed* (i.e., among nodes).

Answering these questions is fundamental and facilitates the understanding of different sensor calibration approaches and of the calibration models that have been employed. These calibration attributes are described in the following subsections; Figure 2 shows a map of the calibration attributes.

5.1. Calibration Area

Sensor nodes are usually deployed in a specific area. Depending on the area of interest, calibration can be performed at a given point, in which case it is known as micro, or device-level, calibration. If the objective is a whole area involving a set of sensors or all of them, the calibration is called macro, or system-level, calibration.

Micro-calibration refers to the method that tunes each individual sensor to output accurate readings at a specific given point (location) of the area, as illustrated in Figure 3. Micro-calibration algorithms calibrate every sensor according to a reference node in order to have accurate sensor measurements of the phenomenon monitored. Examples of micro-calibration are the calibration of a sensor light at a given location [38] and the

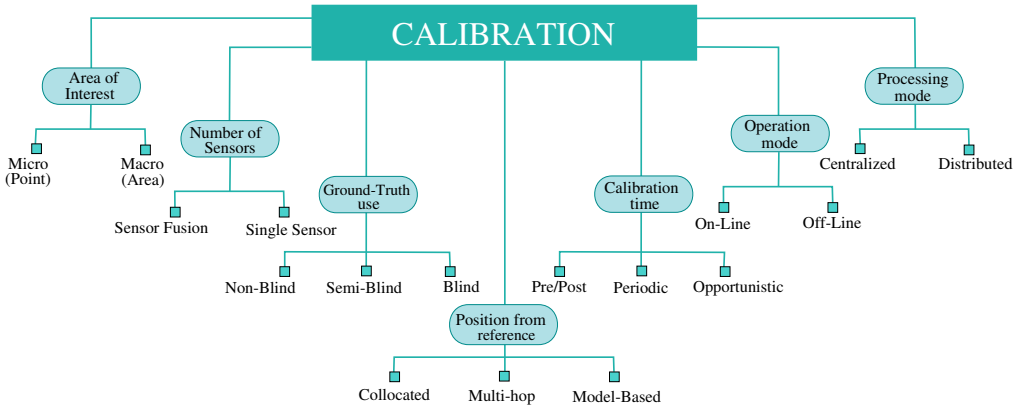


Figure 2: Illustration of the calibration attributes that make up a calibration approach.

calibration of a NOx sensor [16] with respect to a ground-truth reference station.

Macro-calibration assumes a system-wide calibration in the area of interest and focuses on optimizing the measurements of that area as a whole. The main goal of macro-calibration is not to adjust sensor calibration according to a reference signal. Instead, calibration algorithms try to maximize the similarity among the measurements of all sensor nodes in the area. Hence, most macro-calibration algorithms do not require access to reference measurements [10, 17, 44]. For example, Stankovic et al. [10, 53] proposed a distributed macro-calibration approach based on the generalized consensus problem, in which all the equivalent sensor gains and offsets in the monitoring area covered by the sensor network should converge asymptotically to equal values, reducing the overall network error. Another example is that of [17], in which a detection mechanism was leveraged for surveillance applications. Sensors give their readings to a centralized node (a cluster-head), which computes the calibration coefficients for each sensor node that participates in the process, with the goal of maximizing the detection probability subject to an upper bound of the false alarm rate specified by the application.

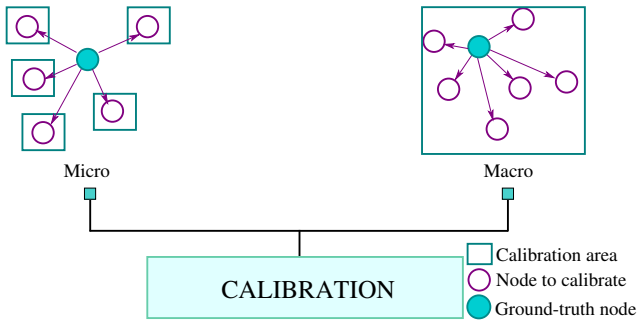


Figure 3: Illustration of micro- and macro-calibration in sensor networks.

5.2. Number of Sensors

Many calibration techniques use the measurements from a *single sensor*. This approach is very common [6, 7, 8, 10, 11, 35, 40, 43, 46, 50, 51, 52, 54] in sensor network calibration, and these measurements are used to establish the calibration model for the sensor and estimate the corresponding parameters.

In contrast, the goal of *sensor fusion*, or *multi-sensor data fusion*, is to combine information from two or more data sources into a single one that provides a more accurate description than that of any of the individual data sources, as illustrated in Figure 4. Several approaches can be considered for sensor fusion. The most common approach is to reduce calibration errors by jointly considering the measurements of multiple sensors. This technique is called *array of sensors* and aims to reduce the uncertainty of calibration parameters in the data model. Arrays of different classes of gas sensors [16, 27, 28, 30, 32, 33] have proven useful to qualitatively identify gas species using pattern recognition approaches and quantitatively determine gas composition based on regression models. An example is a NO_2 sensor that is regressed using NO_2, O_3, temperature, and relative humidity sensors. Moreover, arrays or networks with the same class of sensors [33] can be represented by a *virtual* value that stands for the set of sensors. In this case, the set of nodes sensing the same physical phenomenon can use a multivariate model, a hierarchical Bayes model, or a consensus algorithm to obtain the virtual value representing the calibrated sensor. Other techniques used for sensor fusion are the following: Tan

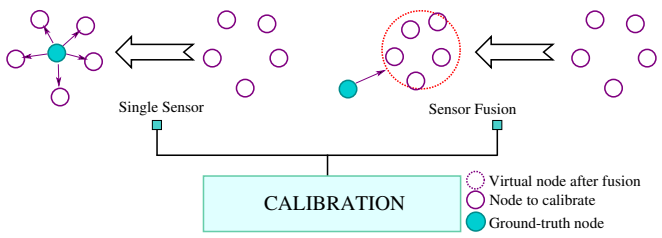


Figure 4: Calibration with single sensor and with sensor fusion.

et al. [17] proposed a two-tier system-level calibration of a sensor network. As a first step, each sensor learns and transmits its local sensing model to a head node, also called fusion-head. Then, the received sensors' measurements are fused in the second tier, where a common model is established, and sensors are

globally calibrated to optimize the system-wide performance. Similarly, Fabeck et al. [37] considered a network of nodes that send measured data to a centralized node according to a binary hypothesis testing problem for detecting the presence of a target. The received decisions are combined to yield a final decision. The decision fusion problem can be viewed as a hypothesis testing problem, with local detection results being the observations and a Bayes optimal fusion rule taking the form of a likelihood ratio test. The minimum probability of error associated with the optimal fusion rule is given by a Bayes risk probability function. Gao et al. [36] used multi-sensor fusion of four sensors attached to the waist, chest, thigh, and side of the body for activity recognition. They relied on Bayesian techniques to increase the system estimation and achieved between 70.88% and 97.66% accuracy.

5.3. Knowledge of Ground-Truth Data

Calibration can be blind, semi-blind, or non-blind depending on whether or not sensory data are processed in the presence of controlled stimuli, a reference model, or high-fidelity ground-truth values (Figure 5).

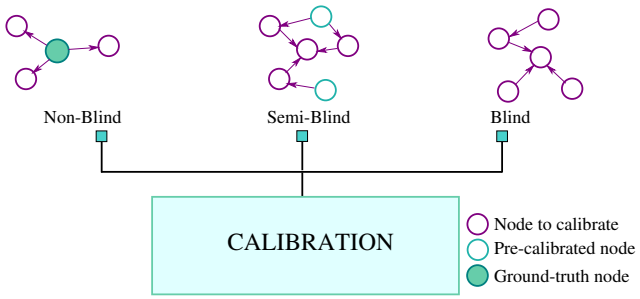


Figure 5: Illustration of blind, semi-blind, and non-blind calibration.

In *non-blind calibration*, also known as *reference-based*, sensors are calibrated leveraging some reference information. These approaches are based on a priori knowledge of *ground-truth* data, from high-quality sensors, which can be useful for comparing the response with the expected values. Accordingly, the calibration parameters are adjusted using these known data inputs. A good example in which non-blind calibration techniques are easy to apply is temperature and gas sensors. In the case of gas sensors, in many countries there are organizations[10] that deploy high-accuracy reference stations to measure air pollution. These reference stations are not densely deployed due to their high costs (approximately 100 thousands euros)[11], but they can be used to calibrate low-cost sensors located behind them (pair-wise or collocated calibration; see Section 5.4). Non-blind calibration was applied for sensor network calibration by [12, 15, 27, 30, 37, 49, 55]. Moreover, some researchers use commercial instruments to provide ground-truth measurements that are later used to evaluate calibration models [16, 27]. These instruments are not calibration chambers in which the conditions can be manually adjusted; instead, they are accurate devices that allow comparing the low-cost readings against reference data.

Blind calibration is the approach that calibrates sensor networks without relying on any controlled stimuli or high-fidelity ground-truth data [18]. Instead, blind calibration schemes leverage signal processing theory, a priori knowledge, such as physical models and constraints, of the sensed phenomenon, and exploit the advantages of dense sensor deployments, such as redundant measurements and temporal or spatial correlation among groups of sensors, to determine the fitting function and estimate the calibration parameters. Many blind calibration works have been proposed for sensor networks employing different techniques [9, 10, 18, 35, 50, 51, 67, 68, 69, 70]. Balzano et al. [18] used least squares to solve the linear blind calibration problem; later, Lipor and Balzano [67] also proposed blind calibration with model mismatch in which the subspace measured is not perfectly known, solving the problem using total least squares. Ramakrishnan et al. [9] estimated the signal signature and simultaneously calibrated the sensor nodes using gossip-based algorithms. Stankovic et al. [10] proposed a distributed algorithm for blind macro-calibration of large sensor networks. Wang et al. [35] proposed a density guided blind calibration (DGC) scheme for nonlinear mobile sensors by approximating the nonlinear calibration functions using piecewise linear functions. DGC relies on the observation that different sensors that move in the same region record similar field statistics. Other authors [51] proposed a learning phase in which principal component analysis (PCA) is used on a subspace of drift-free samples, followed by a calibration phase in which compressed sensing is used to estimate the sensor drift. Compressed sensing (CS) exploits the fact that the signal is M-sparse, meaning that only M<K out of the K components are non-zero. The authors of [68, 69, 70] also applied compressed sensing to blind calibration frameworks with unknown gains.

Semi-blind calibration is an intermediate case between blind and non-blind calibration. In semi-blind calibration, a sensor requires partial ground-truth information sent from a subset of nodes to calibrate itself [17, 32, 38, 41, 43, 45, 54, 56] or when interacting with a calibrated sensor [8, 71], as illustrated in Figure 5. For instance, in the semi-blind calibration proposed by [8], uncalibrated nodes interact with calibrated nodes and run an average consensus algorithm using their measurements as partial ground-truth for calibration. Maag et al. [32] proposed a multi-hop calibration scheme in which sensors are calibrated using already calibrated sensors.

5.4. Position from Ground-Truth Nodes

Depending on the relative position of the uncalibrated node with respect to the reference model, calibration can be performed between collocated nodes, iteratively hop-by-hop (multi-hop) from the ground-truth, or with the local presence of a reference model (model-based), as illustrated in Figure 6.

[10] The European Environment Agency (EEA, http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-7) and the United States Environmental Protection Agency (EPA, https://www.epa.gov/outdoor-air-quality-data) publish air pollution ground-truth data from reference stations of many countries in Europe and in the US, respectively.

[11] Similar cost in US dollars.

Following the definition of Saukh et al. [29], a process of interest P is a continuous measurable signal $\eta{:}T \times L \rightarrow D$ with domain $T \subseteq \mathbb{R}^+$ and $L \subseteq \mathbb{R} \times \mathbb{R}$, where T represents time and L location. Let us assume two sensors s_1 and s_2 that take samples at two times and locations such as (t,l) and (t',l'); then, the process P is bounded in time and space:

$$| \eta(t,l) - \eta(t',l') | \leq \gamma(| t - t' |, | l - l' |), \tag{13}$$

with γ being a monotonically non-decreasing function. As the authors note, the slower the function γ grows, the more correlated will be the data of the two sensors.

Saukh et al. [29] defined a *rendezvous* as the set of spatially and temporally close pairs of measurements $\Phi^{s1,s2}$ between sensors s_1 and s_2 within a time interval T as

$$\Phi^{s_1,s_2} = \{(x_{s1}(t_i,l_i), x_{s2}(t_j,l_j)) \mid (t_i,t_j \in T) \wedge (| t_i - t_j | \leq \triangle t) \wedge (| l_i - l_j | \leq \triangle d)\}, \tag{14}$$

where $\{x_{s_1}\} \in D_{s_1}$, $\{x_{s_2}\} \in D_{s_2}$, and $\triangle t$ and $\triangle d$ are temporal and spatial constraints on the required closeness of measurements, respectively. As it can be observed, the rendezvous defines how close two sensors have to be to calibrate the uncalibrated sensor with respect to the reference sensor and the temporal synchronization level of the samples from both sensors.

Collocated calibration refers to the situation where a node is calibrated using interaction with a close neighbor node (collocated) that can sense the same phenomenon and provide reference measurements. In terms of the process P, collocated calibration involves a sensor at place (t,l) that gives ground-truth values, a sensor at place (t',l') that is going to be calibrated, and equation (14) to be fulfilled with closed temporal and spatial constraints. This approach, then, is associated with non-blind calibration where a node is exposed to the true values available from a reference ground-truth node [11, 16, 24, 30, 27, 49].

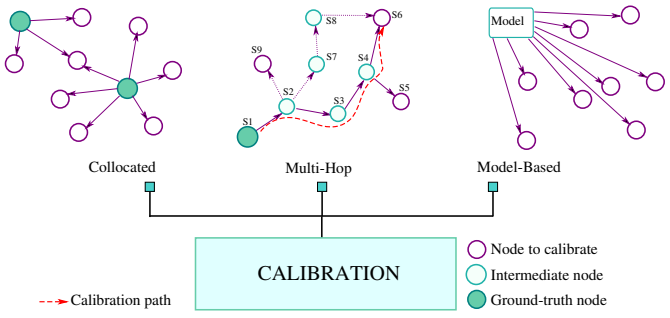


Figure 6: Illustration of collocated calibration, iterative, or multi-hop, calibration, and model-based.

In *iterative, or multi-hop, calibration*, a node is calibrated using a set of nodes already calibrated, which are not necessarily ground-truth nodes. The name of multi-hop comes from the case in which the uncalibrated sensor is h-hops from the ground-truth node, where each hop has iteratively been calibrated. For illustrating the process, let us define the *calibration graph* (Figure 6) as a directed graph $G{=}(V,E)$, where $V{=}\{1,2,...,N\}$ is the set of nodes, and $E{\subseteq}V{\times}V$ is the set of edges. The edge e_{ij} between sensor node s_i and sensor node s_j means that s_j is calibrated using node s_i as reference node. In this case, an uncalibrated sensor can be calibrated using an h-hop path. An example (Figure 6) is a path $s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4 \rightarrow s_6$, with s_1 being a ground-truth node. Then, s_2 would be calibrated using node s_1, node s_3 using node s_2, s_4 using node s_3, and s_6 using node s_4.

Collocated calibration would be the case of having a 0-hop calibration. Another special case is when h=1, called *single-hop*. In this situation, a node is calibrated using already existing neighboring calibrated sensor nodes[12]. In a multi-hop calibration approach, there can be many paths to calibrate an uncalibrated sensor. Thus, the objective is to minimize the errors committed as a result of the calibration process [6, 29, 32].

Different paths contribute differently to the post-calibration *skew* of the sensors, where skew is defined as the difference between the calibrated value of sensor j and the actual value [71]. The goal in multi-hop, or iterative, calibration is to study how the errors propagate along the network path and find calibration paths with minimum skews. The skew depends on the length of the paths and on the individual calibration errors of the sensors that the path traverses. As an example, if sensor s_1 is calibrated against a reference sensor and produces measured data in the range $[x_1\text{-}\epsilon_1, x_1\text{+}\epsilon_1]$ and then a sensor s_2 is calibrated against s_1, reporting measured data in the range $[x_2\text{-}\epsilon_2, x_2\text{+}\epsilon_2]$, the post-calibration skew of sensor s_2 will be at most $[\epsilon_1\text{+}\epsilon_2]$. Akan [71] proved that the skew is at most the sum of the absolute errors along the calibrating path.

In a more recent work, Saukh et al. [29] investigated whether linear fitting models suppress multi-hop error propagation in a large-scale mobile sensor network. Saukh et al. [29] showed that ordinary least squares regression suffers from *regression attenuation* or *regression dilution* [72]. Regression dilution is the biasing of the regression slope towards zero caused by errors in the independent variable. This effect increases as the number of hops increases. The reason is that at every hop, the independent variable x of the linear fitting adds an error to the model. For example, let us have a path $s_1 \rightarrow s_2 \rightarrow s_3$, where s_1 is a ground-truth node; then, assuming a simple linear model,

$$\begin{cases} x_1 \sim y_2 = & \beta_{02} + \beta_{12} x_2 + \epsilon_2 \\ y_2 \sim y_3 = & \beta_{03} + \beta_{13} x_3 + \epsilon_3 \end{cases}, \tag{15}$$

with ϵ_2 and ϵ_3 normal distributed errors, x_1 is the ground-truth data, and x_2 and x_3 are respectively the measured sensor data and are defined according to equation (14) in subsection 5.4. In this case, at each hop, the calibration coefficient is decreased proportionally by the variance of the error given by that hop (regression dilution) [72]. The solution is compensating the bias

[12] A *pair-wise* calibration is also used in this context. Pair-wise implies a rendezvous between two sensor nodes in the calibration. If one of them is a ground-truth node, we call the pair-wise a 0-hop, or collocated, calibration. If one of them is an already calibrated node, we call the pair-wise a 1-hop, or single-hop, calibration. It was common in the early literature on calibration in which multi-hop was not well defined to call pair-wise any rendezvous, regardless of whether it was an already calibrated node or a ground-truth node.

in the slope estimation if the variance of the error of the sensor node is known. The authors of [29] mentioned the difficulty of this estimation due to drift in real sensors and changing environmental conditions. As an alternative solution, [29] proposed to use other fitting techniques that do not suffer from regression dilution, such as geometric mean regression (GMR). They tested the multi-hop calibration in a real network of O_3 and CO sensors, showing that ordinary least regression (OLR) produces regression dilution, that GMR has very low hop-by-hop error propagation, and that the calibration quality is independent of the calibration error of its calibration parent. Maag et al. [32] proposed sensor array network calibration (SCAN), a multi-hop micro-calibration scheme for mobile sensor arrays. SCAN minimizes the accumulated error over multiple hops of calibrated sensors. SCAN formulates a constrained least squares regression optimization problem reducing, and even eliminating under certain conditions, the regression dilution effect.

In *model-based calibration*, a node is calibrated using a set of ground-truth nodes that are not in the vicinity of the non-calibrated node. In this case, the ground-truth values allow by means of a mathematical model to produce a reference value at the position of the non-calibrated node. The values produced by this model now act as ground-truth reference data for the estimation of the calibration parameters [7, 36, 38, 44, 45]. The accuracy of model-based calibration is less than that of the collocated calibration, since the values estimated in the non-calibrated node are produced by a model that introduces an error, even if this model uses ground-truth data. An example is a source of light. The intensity of the light at source location d can be obtained by an exponential decaying function [7] that depends on the intensity of the light source and distance (see equation (10). In another example, Moses et al. [38] considered the problem of location and orientation of sensors by measuring the time of arrival (TOA) and direction of arrival (DOA) of the signal emitted by that source. A third example is given by Whitehouse et al. [5], who used received signal strength information (RSSI) in fusion with acoustic time of flight (TOF) as a method of auto-calibration for localization services. Some distances between sensors are known a priori, and this allows taking advantage of anchor nodes or pre-calibrated nodes.

5.5. Calibration Time and Frequency

In most cases, the calibration process is performed before the nodes are deployed, *pre-deployment calibration* [7, 16, 35, 45] among others. However, there are situations in which the sensors are replaced or recalibrated with certain periodicity. Examples are ozone (O_3) campaigns that occur in summer in Europe. After a summer campaign, a recalibration of the sensors can give information on how the sensors have aged and how good the results are. This kind of calibration is called *pre-post-deployment* and has two phases: a first calibration before the sensors are deployed and a second phase after the deployment. If the sensors have aged, the data from the first half of the campaign are predicted with the pre-calibrated parameters, while the data from the second part of the campaign are post-calibrated.

Periodic calibration occurs when the nodes are recalibrated at given time intervals (see Figure 7). The authors of [29, 30, 32] used two stations of the local governmental measurement network (NABEL and OstLuft stations) located in the city center of Zurich as their high-quality measurements to calibrate low-cost sensors deployed on top of public streetcars. Given the known itinerary of public transportation, the mobile sensors on top of the streetcars are periodically calibrated whenever they enter in contact with reference stations that provide ground-truth reference data. Characterized by a meeting point with a distance of 50 meters and a time of 5 minutes, the rendezvous calibration graph of the ozone measurements comprises 500 meeting points between the reference stations and mobile sensors within an interval of 10 days.

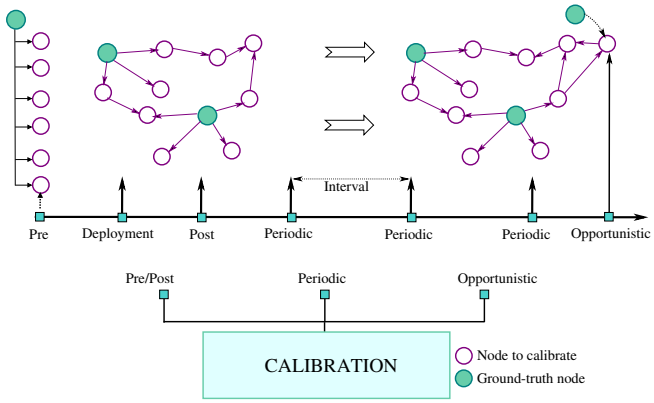


Figure 7: Illustration of pre- and post-deployment calibration, periodic calibration, and opportunistic calibration.

Opportunistic calibration is used whenever the mobility patterns of sensors are unknown and the calibration of sensor measurements relies on the possible rendezvous with a ground-truth or freshly calibrated node. As an example, Miluzzo et al. [8] exploited the opportunistic encounter with ground-truth, or calibrated, sensors to calibrate uncalibrated mobile sensors. The convergence time of the approach depends on the mobility patterns of mobile sensors and on the density of the ground-truth nodes.

5.6. Mode of Operation

Off-line calibration involves performing sensor calibration when the device is not under operation, as illustrated in Figure 8. A sensor is under operation when it is taking data that an application is using. Ramanathan et al. [12] discussed ways to increase the quantity and quality of data on-line during a deployment. In off-line calibration[13], a node takes probe data to calibrate the sensor. In general, off-line calibration requires that

[13] We define off-line and on-line based on ideas presented in the paper [12]. In this paper, Ramanathan et al. call on-line calibration the case when in-field users can detect and compensate for problems as they occur. We extend the definition to the general case of sensors that are operative. Some previous papers, such as Feng et al. [7], named on-line calibration what was later called blind-calibration in the literature.

the sensor send the data to a node with enough storage and computation capabilities to perform the calibration process. Calibration, then, can be carried out using a set of measurements that are recorded in a repository and later used along with some reference signal and a known model to calibrate the device. The goal is to obtain a function that maps the raw data recorded by the low-cost sensor to the real value with a minimum prediction error. Examples of off-line calibration can be found in [12, 16, 27, 28, 33, 36, 38, 49].

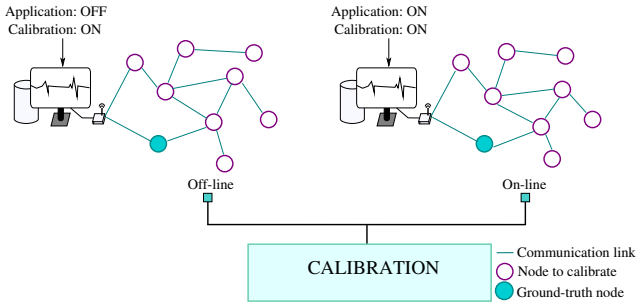


Figure 8: Illustration of on-line and off-line calibration in sensor networks.

On-line calibration enables performing the calibration process during the normal operation of sensors [8, 15, 29, 40, 41, 43, 44, 50]. For example, in the on-line temperature sensor calibration approach in [44], sensors cooperate using a gossip-based algorithm and run the expectation-maximization algorithm that enables them to converge at runtime on local aggregate means in order to compute a set of calibration parameters. In general, single-sensor on-line calibration requires communication with the ground-truth node or among the nodes that participate in the calibration process.

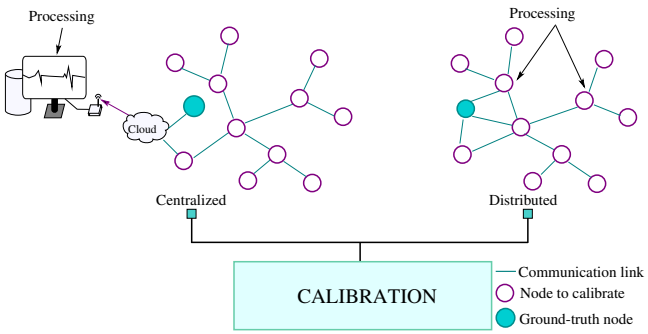


Figure 9: Illustration of centralized and distributed calibration in sensor networks.

5.7. Mode of Calibration Processing

In many situations, the sensor nodes store the sample data or send the measurements taken in an interval of time to a centralized node in the sensor network or to a centralized server on the Internet. After having a set of samples, the centralized node or server calculates the calibration parameters (see Figure 9). In this case, the calibration is called *centralized calibration*. The calibration parameters obtained in the centralized server are then used to produce calibrated data [11, 18, 27, 31, 33, 49].

In *distributed calibration*, a sensor node and its neighboring, or multi-hop, sensor nodes collaborate to calculate the calibration parameters [8, 10, 35, 41, 52]. There are several possibilities for building distributed system architectures. The following are some of the approaches.

5.7.1. Consensus Algorithms

The consensus problem goal in a distributed system with multiple agents is the agreement among the agents for a single data value. In the case of calibration, a network of uncalibrated nodes interacts with calibrated nodes to solve a consensus problem. The average consensus algorithm measures the sensor samples' disagreement between the uncalibrated node and a set of calibrated neighbors. The algorithm converges to a consensus value among the set of nodes and leads to the discovery of the actual disagreement between the uncalibrated node's sensor and calibrated nodes' sensors. In a consensus algorithm, [73], the state of agent n in graph G=(V, E) is denoted as x_n. A consensus algorithm to reach an agreement regarding the state of n agents with dynamics $\dot{x}_i$=u_i, where $\dot{x}$ denotes derivative, can be written as an *nth*-order linear system:

$$\dot{x}_i(t) = \sum_{j \in N_i} (x_j(t) - x_i(t)) + b_i(t), \tag{16}$$

with $x_i(0)$=$z_i \in R$ and $b_i(0)$=0. It can be shown [73] that the iterative discrete consensus algorithm can be expressed as

$$x_i(k+1) = (1-\epsilon)x_i(k) + \epsilon \sum_{j \in N_i} \frac{(x_j(k) - x_i(k))}{|N_i|}. \tag{17}$$

CaliBree [8] is a model-based calibration network in which ground-truth calibrated nodes inform non-calibrated nodes, using a beacon protocol, that they will participate in the calibration process. These nodes, then, use a discrete consensus distributed algorithm to calibrate the uncalibrated node by computing the degree of disagreement between the reference nodes and the uncalibrated node. The nodes are separated in two sets: nodes that are calibrated and nodes that are not calibrated. Let N_i be the neighboring set of calibrated nodes of an uncalibrated node i. The state, denoted as $s_i(k)$, is the data sample taken at time k, and $\bar{d}_i(k)$ is the average disagreement measured by node i up to round k. The disagreement is defined as the difference between the state and the consensus value, that is, $d_i(k)$=$s_i(k)$-α. The disagreement at step k=0, d_i^{uncal}, is the difference between the uncalibrated node i and a calibrated node. The formulation of the disagreement consensus problem for calibration of data is then [8]

$$\bar{d}_i(k+1) = \begin{cases} (1-\epsilon)\bar{d}_i(k) + \epsilon \sum_{j \in N_i} \frac{(s_j(k) - s_i(k))}{|N_i|} & k > 0 \\ d^{uncal} & k = 0 \end{cases}. \tag{18}$$

CaliBree [8] considered that neighboring calibrated nodes periodically send beacons with their ground-truth or calibrated sensed data. Uncalibrated nodes run the consensus algorithm

Table 5: Attributes used in calibration approaches classified by applications.

References	Micro	Macro	Single Sensor	Sensors Fusion	Non-Blind	Semi-Blind	Blind	Collocated	Multi-Hop	Model-based	Pre/Post	Periodic	Opportunistic	Off-Line	On-Line	Centralized	Distributed	Application
Attributes used in light, temperature, relative humidity, vibration and accelerometer applications																		
[6]		X	X				X		X		X				X		X	Temperature
[7]		X	X		X					X	X				X	X		Point-lights
[8]	X	X	X		X	X			X				X		X		X	Light
[11]	X		X		X			X			X				X		X	Light, Temp., Hum.
[18]	X		X				X		X		X			X		X	X	Temp., Hum.
[35]	X		X				X		X		X				X	X	X	Light
[36]	X			X	X					X	X			X		X		Motion (accelerometer)
[41]	X		X			X			X		X				X		X	Light
[43]		X	X			X			X		X				X		X	Temperature
[44]		X	X			X				X	X				X		X	Temperature
[45]	X		X			X				X	X			X	X		X	Temp. Hum., Light
[46, 47, 48]	X		X			X			X		X				X		X	Temperature
[49]		X		X	X			X			X			X		X		Vibration (Water Flow)
[50]	X		X				X		X				X		X		X	Temp., Hum., Sound
[51]		X	X				X		X		X				X		X	Temperature
Attributes used in localization, synchronization and target detection applications																		
[5]		X		X	X					X	X				X		X	Localization
[17]		X		X		X			X	X	X				X	X	X	Target Detection
[37]	X			X	X			X			X				X		X	Target Detection
[38]	X			X		X				X	X			X		X		Localization
[39, 40]	X		X			X				X	X				X		X	Localization
[52]	X		X			X			X		X				X		X	Localization
[53, 10]		X	X				X		X			X			X		X	Synchronization
[54]	X		X			X				X		X			X		X	Synchronization
Attributes used in air pollution and water chemistry applications																		
[12]	X			X	X			X			X			X		X		Water Chemistry
[15]	X			X	X			X	X			X	X		X		X	O_3
[16, 27]	X			X	X			X			X			X		X		O_3,CO,CO_2,NO,NO_2
[28]	X			X	X			X			X			X		X		CO,CH_4,C_3H_8,CeO_2,NiO
[29]	X			X	X				X				X		X		X	O_3,CO
[30]	X			X	X			X			X			X		X		O_3,CO,NO_2
[31]	X			X	X			X			X			X		X		O_3, NO, NO_2
[32]	X			X	X	X			X		X	X			X		X	O_3, CO, NO_2
[33]	X			X	X			X			X			X		X		O_3
[55]	X			X	X			X			X			X		X		Photosynthetically Active Radiation (PAR)
[56]		X		X		X				X	X			X			X	Air pollution

to achieve a consensus value, showing that the algorithm converges to a minimum disagreement $\bar{d}_i^{min}$. The CaliBree protocol performance is tested using an uncalibrated mobile light sensor node that comes in the range of calibrated sensor nodes.

In a similar work, Bolognani et al. [52] proposed a distributed algorithm to micro-calibrate the radio strength signal indicator (RSSI) that is applied to localization and tracking for WSNs. The algorithm uses mean squares to estimate wireless channel parameters, whereas a consensus algorithms is introduced to estimate the sensor offset that affects the measured received strength of the receiving node due to fabrication mismatches in the radio chip and that can produce ±6dB in RSSI measurements.

Stankovic et al. [10] proposed a distributed non-blind or blind macro-calibration mechanism to calibrate the coefficients of a network of N sensors measuring a common discrete signal $\{x(t)\}$ that is considered a random process. The algorithm can be considered a non-trivial extension of consensus algorithms. Each sensor n produces an output signal $y_n(t) = \beta_{0n} + \beta_{1n}x(t)$. The goal is to achieve a calibration function $z_n(t) = a_n y_n + b_n = a_n\beta_{1n}x(t) + a_n\beta_{0n} + b_n = g_n x(t) + f_n$ so that when there is no ground-truth reference signal (distributed blind macro-calibration), the sensor produces asymptotically equal outputs $z_n(t)$. The mechanism may also be adapted to the case in which a sensor gives ground-truth reference data (distributed non-blind macro-calibration), producing a gain g_n=1 and offset f_n=0 for all n=1, ..., N. If θ_n=$[a_n, b_n]$. The distributed blind calibration algorithm minimizes the instantaneous difference between the signal measured by a sensor and its neighborhood:

$$J_i = \sum_{j \in N_i} \gamma_{ij}(z_i(t) - z_j(t))^2, \tag{19}$$

where γ_{ij} are nonnegative scalar weights reflecting the relative importance of the neighboring nodes. If we apply a *stochastic gradient descent algorithm*, the parameter θ_n is expressed as

$$\hat{\theta}_n(t+1) = \hat{\theta}_n(t) + \delta(t)\nabla_\theta J_n = \hat{\theta}_n(t) + \delta(t) \sum_{j \in N_i} \gamma_{ij}[z_i(t) - z_j(t)] \begin{bmatrix} y_n(t) \\ 1 \end{bmatrix}, \tag{20}$$

where $\delta(t)$ is the step length at each step t of the algorithm. The algorithm has to begin with an initial point $\theta(0)$=[1, 0]. At each step, node n has to receive the updates of its neighboring nodes. Stankovic et al. [53] gave conditions for the convergence of the algorithm and showed how to modify the algorithm in case a sensor acts as a ground-truth sensor, giving reference data to its neighbors.

5.7.2. Gossip Algorithms

Average consensus gossiping [74, 75] is an asynchronous protocol where a node chosen uniformly at random wakes up, contacts a neighbor randomly within its connectivity radius, and exchanges a state variable to produce a computation update. The problem is having a network of n nodes that sample an area with initial values $x(0)=[x_1(0), ..., x_n(0)]^T$ and calculating the average of the entries of $x(0)$, $\bar{x} = \sum_i x_i(0)$ in a distributed manner. Although average gossip algorithms perform data fusion (e.g., achieving an average of the samples of the network), gossip algorithms allow distributing the computational burden and thus also apply to distributed calibration.

Ramakrishnan et al. [9] presented a distributed signature learning and node calibration (D-SLANC) gossip algorithm for sensor calibration in WSNs. The algorithm relays on distributed measurements of sensors to estimate source signal's signature and to estimate calibration parameters using gossip-based distributed consensus. This is a special case of blind calibration in which there is no a priori knowledge of the signal subspace [18]. The mechanism can also be classified as micro-calibration, because although there is a network of sensors that collaborate to estimate the signal, each one of the sensors will have its own estimated calibration coefficients. In D-SLANC, N nodes measure a common signal of interest with different gains that are time-invariant. Each node n=1, ...N, has the following model:

$$y_n = \begin{bmatrix} y_{n1} \\ \vdots \\ y_{nK} \end{bmatrix} = \begin{bmatrix} \beta^n_{11} & \cdots & \beta^n_{1M} \\ \vdots & \ddots & \dots \\ \beta^n_{K1} & \cdots & \beta^n_{KM} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_K \end{bmatrix} = \mathbf{A_n}(\beta_n)\mathbf{x}+\epsilon, \quad (21)$$

where $y_i \in R^K$ is the noisy observation, $\epsilon_i \in R^m$ is additive Gaussian noise with zero mean and covariance $\sigma^2\mathbf{I}$, and $x \in R^M$ is the common signal to be estimated. The whole network can be expressed as follows:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \vdots \\ \mathbf{y_N} \end{bmatrix} = \begin{bmatrix} \mathbf{A_1}(\beta_1) \\ \vdots \\ \mathbf{A_N}(\beta_N) \end{bmatrix} \mathbf{x} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix} = \mathbf{A}(\beta)\mathbf{x} + \epsilon. \quad (22)$$

The signature x and the calibration parameter β_n are the unknown parameters to be estimated. Here, there are no ground-truth data, the sensor measurements act as y_n, and the unknowns are both the calibration coefficients and the true sensing signal (signature estimation). Since the joint distribution of the unknowns is non-convex, D-SLANC splits the parameter space into two parts, the signature space and the calibration coefficients space, and then applies *alternating minimization* that consists of performing maximum likelihood (ML) estimation of one set while keeping the other one fixed. D-SLANC converges to a local maximum of the likelihood. To perform the signature estimation, a projected-gradient-based coordinate descent algorithm can be developed in which a distributed consensus algorithm forms the basis of each step of the coordinate algorithm [9]. Using the gradient descent algorithm, each node computes each one of the M components of the signature signal x. At each step of the algorithm it is necessary to calculate the gradient of the signature. This gradient is obtained using a distributed consensus mechanism. Ramakrishnan et al. [9] provided sufficient conditions to guarantee convergence of the coordinate descent for the exact consensus and with high probability for the approximate consensus algorithm. Finally, the local calibration factors are calculated given the current estimate of the signature.

Buadhachain et al. [44] combined noise compensation using the EM (expectation-maximization) algorithm and a gossip-based protocol for sharing calibration information in a network of temperature sensors. The authors presented a state-space representation for the system and a control methodology that calibrates sensor nodes by modifying their output values given local sensors' outputs and the models.

5.7.3. Gaussian Process Regression

Fujino et al. [46, 48] proposed to calibrate a network of N thermal sensors using Gaussian processes. All the sensors measure the same phenomenon, and then they are neighbors from the point of view of seeing similar correlated data, being the calibration form of type collocated or single-hop. The authors considered that a set of sensors acts as ground-truth nodes and proposed an on-line version of the architecture, instead of the classical off-line. The network can be represented by $p(y_i \mid \theta) \sim N(a_i x_i + b_i, \sigma^2)$, where θ consists of all sensors' parameters [a_1,...,a_N, b_1,...,b_N]. The likelihood can be obtained from $p(y_i \mid \theta)$, and the solution of the problem is found using the expectation-maximization (EM) algorithm.

Gaussian process regression, also known as kriging, is a technique used in geostatistical models to predict the value of a function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point. Kumar et al. [45] used simple kriging to predict the humidity value of a sensor from neighboring nodes, and the measured drift was then Kalman filtered to get the correct drift estimates. The method was tested in a wireless network of 54 real sensors, improving RMS error with respect to averaging-based prediction models [76]. Although [45] used kriging to estimate the value of neighboring sensors and then correct the drift, kriging can also be used as a model-based calibration. Given a set of reference calibrated nodes, the value of interest can be obtained by interpolation in a given location using kriging. This value can then be used as the ground-truth value for calibrating a sensor located in the same place. The accumulated errors now come from the kriging process and from the calibration model used.

5.8. Combining Calibration Attributes

The attributes described in section 5 are not disjoint, and it is crucial to define the system approach in which the calibration will be performed. Multiple approaches can result from combining the calibration attributes. In fact, while describing the calibration attributes, examples of combinations have just appeared. Some examples are (i) centralized, collocated, non-blind, micro, off-line, pre-calibration, [12]; (ii) centralized, model-based, non-blind, macro, on-line, pre-calibration, [7]; (iii) centralized, multi-hop, non-blind, micro-calibration, array of sensors, [32]; and (iv) distributed, semi-blind, on-line, micro-calibration, [41]. There is no golden rule for selecting a calibration approach for a specific application. In general, depending on the resources available for the deployment, different combinations of attributes can yield similar results, being the difference and the quality of information parameters obtained in terms of calibration errors. As an example, calibrating a sensor against a collocated ground-truth sensor node will always produce better performance than calibrating the same sensor against a 1-hop already-calibrated sensor node or against the value predicted by a kriging process. Choosing one option or another will depend on the availability of the collocated ground-truth node or on the availability of other already calibrated nodes. Table 5 describes how the examples in Tables 2, 3, and 4 combine the attributes, forming complex calibration approaches for several applications.

6. Calibration Guidelines

In general, there is no defined methodology that indicates how to calibrate the sensors of a network. However, in figure 10, we show a flowchart that specifies the stages we recommend to follow in the calibration of a wireless sensor network. These steps are described below:

1. **Application requirements:** This stage defines the intrinsic requirements of the application, the ranges of measurement and the QoI metrics.

2. **Definition of network calibration approach:** In this step, the attributes of the calibration are defined. The set of attributes chosen for calibration determines the calibration approach.

3. **Sensor modeling:** during this stage the mathematical models suitable for the calibration of an individual sensor are chosen, such as a linear model or a non-linear model. It also analyzes aspects that may influence the calibration such as the time response of the sensor, the dynamic range of the sensor or the cross sensitivity.

4. **Sensor calibration process:** at this stage, the size of the training set, the size of the test set, or aspects such as the calibration window in case of an opportunistic calibration are defined. It is also determined how the environmental conditions or the place of calibration may affect on the place where the calibrated data will finally be taken. In addition, it is analyzed if it is necessary to recalibrate the sensor due to aspects such as drift or aging. Finally, each sensor is calibrated and its QoI is assessed.

5. **Sensor calibration validation:** in this phase, the individual calibration of the sensors is validated. In case of low performance, we go back to the "sensor modeling" phase where we evaluate if it is necessary to use other models or if there are other aspects that have not been considered to influence the calibration.

6. **Network modeling:** during this stage, the appropriate mathematical models are chosen for the calibration of the sensor network, such as a consensus or a gossip model.

7. **Network calibration process:** at this point, the parameters participating in the network calibration are determined and the information to be exchanged between the nodes of the network is specified. Finally, the network is calibrated and its QoI is assessed.

8. **Network calibration validation:** here, the network calibration is validated. In case of low performance, we go back to the "network modeling" phase and evaluate other parameters that may influence the calibration by repeating the process.

7. Open Issues

Although the survey covers two decades of calibration literature in uncontrolled environments, there are still issues and challenges that have to be solved. We group the open issues into three categories: (i) open issues in calibration models deal with aspects related to the calibration models and algorithms and their effect on sensor calibration, (ii) open issues in calibrating specific sensor technologies deal with specific aspects of sensor technologies, and (iii) open issues in calibration approaches deal with aspects related to how combining attributes can produce more complex calibration approaches.

7.1. Open Issues in Calibration Models

In the case of calibration using linear and nonlinear models in centralized off-line approaches, there is not a great knowledge on the size of the data set needed to calibrate the sensors. In general, the data set is divided into a training and a validation data set. There are few works that analyze what the minimum size of the data set should be in order to correctly calibrate a sensor in a real deployment. The minimum size of the data set depends on the kind of sensors and sometimes on the environmental conditions. The size of the training set is also important in mobile networks in which the sensors are in contact with reference nodes during a window of time [15]. The time window refers to the number of samples obtained during an interval of time. What should be the quality of the calibration in opportunistic or periodic calibration in mobile networks is another challenge that has to be addressed.

In the past years, there has been an increasing number of papers researching more complex calibration techniques based on linear and nonlinear models for centralized, micro, non-blind,

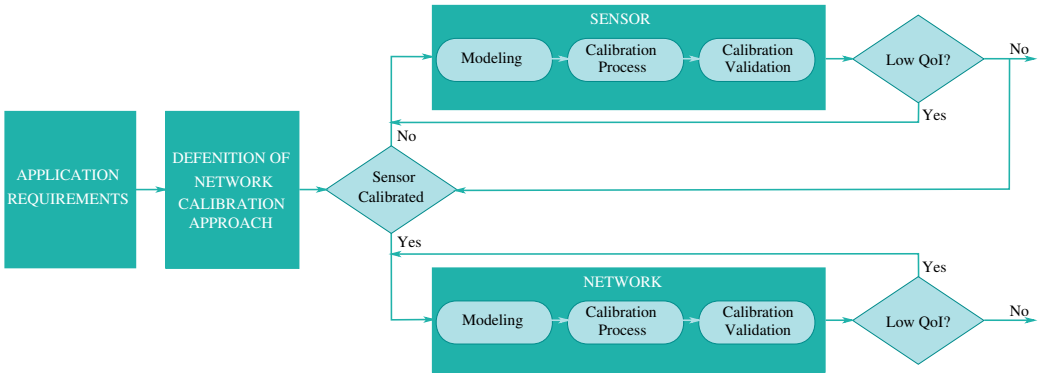


Figure 10: Calibration process flowchart.

collocated, and off-line approaches. As previously mentioned, the initial models were based on linear models when the sensors were for temperature, humidity, or vibration. Light-point sensors were modeled with splines, and air pollution with multiple linear regression models. Some sensors present nonlinearities, because their response function is nonlinear or because of other causes such as drift, aging, or achieving their maximum dynamic range. Some recent works including [16, 27, 31, 41] proposed using deep learning models such as artificial neural networks or support-vector regression. These models are said to characterize better the nonlinearities than splines but at the cost of greater computational complexity. Comparative studies on the complexity of these techniques and models, which ones are appropriate, and how they perform on different families of sensors are needed.

Most calibration techniques have high computational complexity that makes an on-line distributed approach difficult. High-performance light calibration techniques that can be implemented on-line in low-computational or in low-capability nodes are needed. Moreover, if the sensors have nonlinear responses, it is necessary to research how nonlinear models can be applied to on-line distributed approaches.

Many times there is a lack of ground-truth nodes in the network. One of the possibilities in calibration in uncontrolled environment networks is to calibrate the sensors using an already calibrated node or interpolating (e.g., by kriging or other interpolation schemes) the ground-truth values in a point or location and calibrate using these values as reference values. A better knowledge of how the error produced by these techniques against a collocated ground-truth is needed.

Finally, there is no specific repository dedicated to the calibration of low-cost sensors where one can find data sets or code used by the calibration research community to allow the comparison of algorithms, approaches, and calibration methodologies.

7.2. Open Issues in Calibrating Specific Sensor Technologies

There is a lot of research in the communications area on wireless sensor lifetime. However, this does not take into account the lifetime of the sensor subsystem. Depending on the sensor application, sensors have different lifetimes. For example, air pollution low-cost sensors for WSNs based on metal-oxide technology are said to have an average lifetime no longer than a year and a half, and electro-chemical sensors no longer than a year. Moreover, the drift of these sensors is not well known after months of operation [77]. How these sensors drift, what their real lifetime is, after how long the sensors should be recalibrated are questions to be researched. Takruri et al. [43, 78] have also investigated the drift in environmental sensors such as temperature sensors, explaining that detecting drifting sensors and correcting their measurements would increase the effective life of the network. Wang et al. [51] also stated the difficulty in separating the detection of drift and calibration process, and mentioned that more information such as temporal correlation of sensory data, statistical features, or some other prior knowledge can be employed to achieve a higher detection rate.

Sensors from the same manufacturer in applications such as air pollution sometimes present high variability [33, 77]. For example, Peterson et al. [77] analyzed metal-oxide sensors and the effect of warm-up, time, and variability between the same sensors under the same environmental conditions in the calibration process, showing that sensors have to be calibrated regularly and highlighting the importance of calibration at a specific location. Learning which sensor technologies are more stable, what differences exist between sensors from different manufacturers for the same application, and under what conditions the sensors have to be calibrated will trigger the deployment of real applications. Studying temperature and relative humidity sensors, Yamamoto et al. [23] reported a mismatch of the calibration error produced in a calibration place and the error obtained in the prediction in another place with different environmental conditions. The same problem has been stated in air pollution low-cost sensors. Castell et al. [79] mentioned the effect of different environmental conditions when calibrating NO_2 and O_3 sensors in a real WSN deployment. Understanding the different cross-correlations between different phenomena and how these affect the calibration of a specific sensor still needs to be investigated.

7.3. Open Issues in Calibration Approaches

Few research studies on multi-hop approaches exist [6, 29, 32, 71]. Multi-hop approaches have the problem of regression

dilution. Recently, Maag et al. [32] have proposed how to minimize or eliminate the regression dilution effect in multi-hop approaches by using geometric mean regression. Relative to the accumulated error in multi-hop approaches, other calibration techniques robust against regression dilution and use of a ground-truth node or an already calibrated node as reference in the multi-hop approaches are other issues to be investigated.

Most of the works that analyzed blind-calibration used temperature data sets [18, 50, 51] or synthetic data sets [10, 12, 67, 68]. We believe that although blind calibration is a promising calibration technique, it has to be tested with wider sets of sensors and in real scenarios or testbeds. Moreover, knowing the complexity of blind calibration used in low-computational sensor networks and in distributed approaches is another question to be answered.

Most of the calibration approaches found in the literature are distributed-based. However, they are analyzed as centralized-based systems. This is probably because of the small number of real testbeds deployed and the difficulty of implementing real-scenario distributed approaches. We believe that distributed approaches, in general, are an open area that has to be developed and most of the approaches that are distributed have to be proven that they can work in a distributed manner, for example, by evaluating their complexity and feasibility in a real testbed.

8. Conclusions

A broad range of calibration models was developed to correct measurement errors in uncontrolled environments of sensor networks. However, most of the existing calibration approaches are built upon a number of assumptions, such as the availability of high-quality reference measurements; prior knowledge on the true signal, the error model, or both; redundant measurements; spatial correlation, temporal correlation, or both; mobility; and interaction or cooperation of nodes. This helps provide some basic information required for establishing the calibration process.

Throughout the paper, we have defined the main challenges in the calibration process in WSNs in uncontrolled environments. These include defining calibration approaches, calibration models, calibration errors, and the accuracy of the calibration process. We have first stated that calibration in uncontrolled environments implies calibration in the field instead of in environmentally controlled chambers with accurate instrumentation. Nevertheless, this does not mean that accurate instrumentation cannot participate in the calibration process. What it means is that the sensor node can interact in the field with nodes that are more accurate or that were previously calibrated, but sensors are not subjected to the processes of a chamber whose conditions can be manually controlled.

Moreover, we have described the main models used in calibrating sensors in uncontrolled environments. These can be divided into linear and nonlinear models. Choosing a specific model is difficult. In general, this choice depends on the behavior of each sensor. Even sensors of the same family can behave differently, with some being quite linear and others presenting nonlinearities, making the choice of a calibrating model challenging. We have described the typical metrics used to evaluate QoI in the calibration of sensor networks, emphasizing those most used: $RMSE$, R^2, and MAE.

There are a number of attributes that define a specific calibration approach. Describing these attributes allows us to classify how the calibration is going to be performed. These attributes help identify in which area the calibration is performed, how many sensors participate in the calibration, whether there is information available for performing the calibration, what the position of the calibrated nodes with respect to the uncalibrated node is, at which time the calibration is done, whether the network is operative at the time of the calibration, and finally if the calibration is done at a specialized node or among all the sensor nodes.

Knowing these attributes is the foundation necessary to build complex calibration approaches that provide estimation parameters with increasingly sophisticated mathematical models. Throughout the paper, we have defined and explained how other authors have used these attributes to calibrate low-cost sensors in WSNs for different applications including temperature, light-point, vibration, humidity, location, synchronization, and air pollution (e.g., O_3, CO, CO_2, NO, and NO_2) sensors. In summary, we have shown how authors mix calibration attributes to construct calibration approaches and apply mathematical models to solve calibration problems that appear in WSNs.

We think that our paper provides researchers with a comprehensive review of a broad set of calibration approaches and models applied to WSNs in uncontrolled environments, while also showing researchers and engineers how to solve real calibration application problems.

Acknowledgements

This work is supported by the National Spanish funding TIN2016-78473-C3-1-R, regional project 2017-SGR990, and H2020 CAPTOR project.

9. Bibliography

[1] N. A. Pantazis, D. D. Vergados, A survey on power control issues in wireless sensor networks., IEEE Communications Surveys and Tutorials 9 (1-4) (2007) 86–107.

[2] P. Huang, L. Xiao, S. Soltani, M. W. Mutka, N. Xi, The evolution of mac protocols in wireless sensor networks: A survey, IEEE communications surveys & tutorials 15 (1) (2013) 101–120.

[3] N. A. Pantazis, S. A. Nikolidakis, D. D. Vergados, Energy-efficient routing protocols in wireless sensor networks: A survey, IEEE Communications surveys & tutorials 15 (2) (2013) 551–591.

[4] J. M. Barcelo-Ordinas, J.-P. Chanet, K.-M. Hou, J. García-Vidal, A survey of wireless sensor technologies applied to precision agriculture, in: Precision agriculture13, Springer, 2013, pp. 801–808.

[5] K. Whitehouse, D. Culler, Calibration as parameter estimation in sensor networks, in: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, WSNA '02, 2002, pp. 59–67.

[6] V. Bychkovskiy, S. Megerian, D. Estrin, M. Potkonjak, A collaborative approach to in-place sensor calibration, in: Proceedings of the 2Nd International Conference on Information Processing in Sensor Networks, IPSN'03, Springer-Verlag, Berlin, Heidelberg, 2003, pp. 301–316.

[7] J. Feng, S. Megerian, M. Potkonjak, Model-based calibration for sensor networks, in: Proceedings of IEEE Sensors 2003 (IEEE Cat. No.03CH37498), Vol. 2, 2003, pp. 737–742 Vol.2.

[8] E. Miluzzo, N. D. Lane, A. T. Campbell, R. Olfati-Saber, Calibree: A self-calibration system for mobile sensor networks, in: Proceedings of the 4th IEEE International Conference on Distributed Computing in Sensor Systems, DCOSS '08, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 314–331.

[9] N. Ramakrishnan, E. Ertin, R. L. Moses, Gossip-based algorithm for joint signature estimation and node calibration in sensor networks, IEEE Journal of Selected Topics in Signal Processing 5 (4) (2011) 665–673.

[10] M. S. Stankovic, S. S. Stankovic, K. H. Johansson, Distributed blind calibration in lossy sensor networks via output synchronization, IEEE Transactions on Automatic Control 60 (12) (2015) 3257–3262.

[11] P. Buonadonna, D. Gay, J. M. Hellerstein, W. Hong, S. Madden, Task: sensor network in a box, in: Wireless Sensor Networks, Second European Workshop, EWSN, Istanbul, Turkey, January 31 - February 2, 2005, Proceedings, 2005, pp. 133–144.

[12] N. Ramanathan, L. Balzano, M. Burt, D. Estrin, T. Harmon, C. Harvey, J. Jay, E. Kohler, S. Rothenberg, M. Srivastava, Rapid deployment with confidence: Calibration and fault detection in environmental sensor networks, in: Tech. Rep. Center for Embedded Network Sensing (UCLA), 2006.

[13] E. G. Snyder, T. H. Watkins, P. A. Solomon, E. D. Thoma, R. W. Williams, G. S. Hagler, D. Shelow, D. A. Hindin, V. J. Kilaru, P. W. Preuss, The changing paradigm of air pollution monitoring, Environmental science & technology 47 (20) (2013) 11369–11377.

[14] D. E. Williams, G. S. Henshaw, M. Bart, G. Laing, J. Wagner, S. Naisbitt, J. A. Salmond, Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network, Measurement Science and Technology 24 (6).

[15] D. Hasenfratz, O. Saukh, L. Thiele, On-the-fly calibration of low-cost gas sensors, in: European Conference on Wireless Sensor Networks, Springer, 2012, pp. 228–244.

[16] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, 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 (2015) 249–257.

[17] R. Tan, G. Xing, Z. Yuan, X. Liu, J. Yao, System-level calibration for data fusion in wireless sensor networks, ACM Trans. Sen. Netw. 9 (3) (2013) 28:1–28:27.

[18] L. Balzano, R. Nowak, Blind calibration of sensor networks, in: Proceedings of the 6th International Conference on Information Processing in Sensor Networks, IPSN '07, ACM, New York, NY, USA, 2007, pp. 79–88.

[19] A. Kotsev, S. Schade, M. Craglia, M. Gerboles, L. Spinelle, M. Signorini, Next generation air quality platform: Openness and interoperability for the internet of things, Sensors 16 (3).

[20] F. Koushanfar, M. Potkonjak, A. Sangiovanni-Vincentelli, On-line fault detection of sensor measurements, in: Proceedings of IEEE Sensors 2003 (IEEE Cat. No.03CH37498), Vol. 2, 2003, pp. 974–979 Vol.2.

[21] K. Ni, N. Ramanathan, M. N. H. Chehade, L. Balzano, S. Nair, S. Zahedi, E. Kohler, G. Pottie, M. Hansen, M. Srivastava, Sensor network data fault types, ACM Trans. Sen. Netw. 5 (3) (2009) 25:1–25:29.

[22] A. B. Sharma, L. Golubchik, R. Govindan, Sensor faults: Detection methods and prevalence in real-world datasets, ACM Transactions on Sensor Networks (TOSN) 6 (3) (2010) 23.

[23] K. Yamamoto, T. Togami, N. Yamaguchi, S. Ninomiya, Machine learning-based calibration of low-cost air temperature sensors using environmental data, Sensors 17 (6) (2017) 1290.

[24] Y. Yu, H. Li, Virtual in-situ calibration method in building systems, Automation in Construction 59 (2015) 59 – 67.

[25] R. J. Honicky, Understanding and using rendezvous to enhance mobile crowdsourcing applications, Computer 44 (6) (2011) 22–28.

[26] C. MICROceL, Carbon Monoxide (CO) Gas Sensor, City Technology, www.citytech.com, 7th Edition (Nov. 2015).

[27] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, F. Bonavitacola, Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. part b: NO, CO and CO2, Sensors and Actuators B: Chemical 238 (2017) 706–715.

[28] Y. Liu, K. Zhou, Y. Lei, Using bayesian inference framework towards identifying gas species and concentration from high temperature resistive sensor array data, Journal of Sensors 2015.

[29] O. Saukh, D. Hasenfratz, L. Thiele, Reducing multi-hop calibration errors in large-scale mobile sensor networks, in: Proceedings of the 14th International Conference on Information Processing in Sensor Networks, IPSN '15, ACM, New York, NY, USA, 2015, pp. 274–285.

[30] B. Maag, O. Saukh, D. Hasenfratz, L. Thiele, Pre-deployment testing, augmentation and calibration of cross-sensitive sensors, in: Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks, EWSN '16, Junction Publishing, USA, 2016, pp. 169–180.

[31] E. Esposito, S. De Vito, M. Salvato, G. Fattoruso, G. Di Francia, Computational intelligence for smart air quality monitors calibration, in: International Conference on Computational Science and Its Applications, Springer, 2017, pp. 443–454.

[32] B. Maag, Z. Zhou, O. Saukh, L. Thiele, Scan: Multi-hop calibration for mobile sensor arrays, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (2) (2017) 19.

[33] J. M. Barcelo-Ordinas, J. Garcia-Vidal, M. Doudou, S. Rodrigo-Muñoz, A. Cerezo-Llavero, Calibrating low-cost air quality sensors using multiple arrays of sensors, in: Wireless Communications and Networking Conference (WCNC), IEEE, 2018, pp. 1–6.

[34] M. Mueller, J. Meyer, C. Hueglin, Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of zurich, Atmospheric Measurement Techniques 10 (10) (2017) 3783–3799.

[35] C. Wang, P. Ramanathan, K. K. Saluja, Blindly calibrating mobile sensors using piecewise linear functions, in: 6th Annual IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2009, pp. 1–9.

[36] L. Gao, A. K. Bourke, J. Nelson, A system for activity recognition using multi-sensor fusion, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 7869–7872.

[37] G. Fabeck, R. Mathar, In-situ calibration of sensor networks for distributed detection applications, in: International Conference on Intelligent Sensors, Sensor Networks and Information, 2007, pp. 161–166.

[38] R. L. Moses, R. Patterson, Self-calibration of sensor networks, in: AeroSense 2002, International Society for Optics and Photonics, 2002, pp. 108–119.

[39] A. T. Ihler, J. W. Fisher, III, R. L. Moses, A. S. Willsky, Nonparametric belief propagation for self-calibration in sensor networks, in: Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks, IPSN '04, ACM, New York, NY, USA, 2004, pp. 225–233.

[40] A. T. Ihler, J. W. Fisher, R. L. Moses, A. S. Willsky, Nonparametric belief propagation for self-localization of sensor networks, IEEE Journal on Selected Areas in Communications 23 (4) (2005) 809–819.

[41] R. Rossini, E. Ferrera, D. Conzon, C. Pastrone, Wsns self-calibration approach for smart city applications leveraging incremental machine learning techniques, in: 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2016, pp. 1–7.

[42] N. Zimmerman, A. A. Presto, S. P. Kumar, J. Gu, A. Hauryliuk, E. S. Robinson, A. l. L. Robinson, R. Subramanian, A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring., Atmospheric Measurement Techniques 11 (1).

[43] M. Takruri, S. Challa, R. Chakravorty, Recursive bayesian approaches for auto calibration in drift aware wireless sensor networks, Journal of Networks.

[44] S. O. Buadhachain, G. Provan, A model-based control method for decentralized calibration of wireless sensor networks, in: 2013 American Control Conference, 2013, pp. 6571–6576.

[45] D. Kumar, S. Rajasegarar, M. Palaniswami, Automatic sensor drift detection and correction using spatial kriging and kalman filtering, in: IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), 2013, pp. 183–190.

[46] T. Fujino, S. Honda, Automatic calibration method for networked thermal sensors, in: SICE Annual Conference 2011, 2011, pp. 1528–1532.

[47] T. Fujino, S. Honda, An extension of regression-based automatic calibration method for sensor networks, in: 2012 Ninth International Conference on Networked Sensing (INSS), 2012, pp. 1–2.

[48] T. Fujino, S. Honda, Automatic calibration of sensing systems for distributed physical fields, SICE Journal of Control, Measurement, and System Integration 6 (3) (2013) 221–229.

[49] Y. Kim, H. Park, M. B. Srivastava, A longitudinal study of vibration-based

water flow sensing, ACM Transactions on Sensor Networks (TOSN) 9 (1) (2012) 8.
[50] B.-T. Lee, S.-C. Son, K. Kang, A blind calibration scheme exploiting mutual calibration relationships for a dense mobile sensor network, IEEE Sensors Journal 14 (5) (2014) 1518–1526.
[51] Y. Wang, A. Yang, Z. Li, X. Chen, P. Wang, H. Yang, Blind drift calibration of sensor networks using sparse bayesian learning, IEEE Sensors Journal 16 (16) (2016) 6249–6260.
[52] S. Bolognani, S. Del Favero, L. Schenato, D. Varagnolo, Consensus-based distributed sensor calibration and least-square parameter identification in wsns, International Journal of Robust and Nonlinear Control 20 (2) (2010) 176–193.
[53] M. S. Stankovic, S. S. Stankovic, K. H. Johansson, Distributed macro calibration in sensor networks, in: 2012 20th Mediterranean Conference on Control Automation (MED), 2012, pp. 1049–1054.
[54] S. Ando, N. Ono, A bayesian theory of cooperative calibration and synchronization in sensor networks, Trans. Soc. Instru. Control Engineers (SICE) E-S-1 (2006) 21–26.
[55] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, W. Hong, A macroscope in the redwoods, in: Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems, SenSys '05, ACM, New York, NY, USA, 2005, pp. 51–63.
[56] C. Xiang, P. Yang, C. Tian, H. Cai, Y. Liu, Calibrate without calibrating: An iterative approach in participatory sensing network, IEEE Transactions on Parallel and Distributed Systems 26 (2) (2015) 351–361.
[57] J. Fonollosa, I. Rodríguez-Luján, R. Huerta, Chemical gas sensor array dataset, Data in brief 3 (2015) 85–89.
[58] A. Vergara, S. Vembu, T. Ayhan, M. A. Ryan, M. L. Homer, R. Huerta, Chemical gas sensor drift compensation using classifier ensembles, Sensors and Actuators B: Chemical 166 (2012) 320–329.
[59] I. Rodriguez-Lujan, J. Fonollosa, A. Vergara, M. Homer, R. Huerta, On the calibration of sensor arrays for pattern recognition using the minimal number of experiments, Chemometrics and Intelligent Laboratory Systems 130 (2014) 123–134.
[60] S. De Vito, M. Piga, L. Martinotto, G. Di Francia, Co, no2 and nox urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization, Sensors and Actuators B: Chemical 143 (1) (2009) 182–191.
[61] S. De Vito, E. Esposito, M. Salvato, O. Popoola, F. Formisano, R. Jones, G. Di Francia, Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative machine learning approaches, Sensors and Actuators B: Chemical 255 (2018) 1191–1210.
[62] C. M. Bishop, Pattern recognition, Machine Learning 128 (2006) 1–58.
[63] A. Pederzoli, P. Thunis, E. Georgieva, R. Borge, D. Carruthers, D. Pernigotti, Performance criteria for the benchmarking of air quality model regulatory applications: The 'target' approach, Int. J. of Environment and Pollution 50 (2012) 175 – 189.
[64] A. LaMarca, W. Brunette, D. Koizumi, M. Lease, S. B. Sigurdsson, K. Sikorski, D. Fox, G. Borriello, Making Sensor Networks Practical with Robots, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 152–166.
[65] V. Divi, G. W. Wornell, Blind calibration of timing skew in time-interleaved analog-to-digital converters, IEEE Journal of Selected Topics in Signal Processing 3 (3) (2009) 509–522.
[66] S. Ling, T. Strohmer, Self-calibration and biconvex compressive sensing, Inverse Problems 31 (11) (2015) 115002.
[67] J. Lipor, L. Balzano, Robust blind calibration via total least squares, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 4244–4248.
[68] C. Bilen, G. Puy, R. Gribonval, L. Daudet, Convex optimization approaches for blind sensor calibration using sparsity, IEEE Transactions on Signal Processing 62 (18) (2014) 4847–4856.
[69] R. Gribonval, G. Chardon, L. Daudet, Blind calibration for compressed sensing by convex optimization, in: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, IEEE, 2012, pp. 2713–2716.
[70] C. Schulke, F. Caltagirone, F. Krzakala, L. Zdeborová, Blind calibration in compressed sensing using message passing algorithms, in: Advances in Neural Information Processing Systems, 2013, pp. 566–574.
[71] H. Akcan, On the complexity of energy efficient pairwise calibration in embedded sensors, Applied Soft Computing 13 (4) (2013) 1766 – 1773.
[72] C. Frost, S. G. Thompson, Correcting for regression dilution bias: comparison of methods for a single predictor variable, Journal of the Royal Statistical Society: Series A (Statistics in Society) 163 (2) (2000) 173–189.
[73] R. Olfati-Saber, J. A. Fax, R. M. Murray, Consensus and cooperation in networked multi-agent systems, Proceedings of the IEEE 95 (1) (2007) 215–233.
[74] S. Boyd, A. Ghosh, B. Prabhakar, D. Shah, Randomized gossip algorithms, IEEE/ACM Transactions on Networking (TON) 14 (SI) (2006) 2508–2530.
[75] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, A. Scaglione, Broadcast gossip algorithms for consensus, IEEE Transactions on Signal processing 57 (7) (2009) 2748–2761.
[76] B. Krishnamachari, S. Iyengar, Distributed bayesian algorithms for fault-tolerant event region detection in wireless sensor networks, IEEE Transactions on Computers 53 (3) (2004) 241–250.
[77] P. J. Peterson, A. Aujla, K. H. Grant, A. G. Brundle, M. R. Thompson, J. Vande Hey, R. J. Leigh, Practical use of metal oxide semiconductor gas sensors for measuring nitrogen dioxide and ozone in urban environments, Sensors 17 (7) (2017) 1653.
[78] M. Takruri, S. Rajasegarar, S. Challa, C. Leckie, M. Palaniswami, Online drift correction in wireless sensor networks using spatio-temporal modeling, in: Information Fusion, 2008 11th International Conference on, IEEE, 2008, pp. 1–8.
[79] N. Castell, F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, A. Bartonova, Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, Environment international 99 (2017) 293–302.