On-the-fly Calibration of Low-cost Gas Sensors

David Hasenfratz, Olga Saukh, and Lothar Thiele

Computer Engineering and Networks Laboratory, ETH Zurich, Zurich, Switzerland
{hasenfratz,saukh,thiele}@tik.ee.ethz.ch

Abstract. Air quality monitoring is extremely important as air pollution has a direct impact on human health. Low-cost gas sensors are used to effectively perceive the environment by mounting them on top of mobile vehicles, for example, using a public transport network. Thus, these sensors are part of a mobile network and perform from time to time measurements in each others vicinity. In this paper, we study three calibration algorithms that exploit co-located sensor measurements to enhance sensor calibration and consequently the quality of the pollution measurements on-the-fly. Forward calibration, based on a traditional approach widely used in the literature, is used as performance benchmark for two novel algorithms: *backward* and *instant calibration*. We validate all three algorithms with real ozone pollution measurements carried out in an urban setting by comparing gas sensor output to high-quality measurements from analytical instruments. We find that both backward and instant calibration reduce the average measurement error by a factor of two compared to forward calibration. Furthermore, we unveil the arising difficulties if sensor calibration is not based on reliable reference measurements but on sensor readings of low-cost gas sensors which is inevitable in a mobile scenario with only a few reliable sensors. We propose a solution and evaluate its effect on the measurement accuracy in experiments and simulation.

1 Introduction

Urban air pollution is a major concern in modern cities and developing countries. Atmospheric pollutants considerably affect human health; they are responsible for a variety of respiratory illnesses (*e.g.*, asthma) and are known to cause cancer if humans are exposed to them for extended periods of time [18]. Additionally, air pollution is responsible for environmental problems, such as acid rain and the depletion of the ozone layer. Hence, air pollution monitoring is of utmost importance.

Nowadays, air pollution is monitored by networks of static measurement stations operated by official authorities (subsequently called fixed stations). Fixed stations used today are highly reliable and able to accurately measure a wide range of air pollutants using traditional analytical instruments, such as mass spectrometers and gas chromatographs. The drawbacks of these complex measurement systems are their large size, high price, and laborious maintenance. The extensive cost of acquiring and operating these stations severely limits the number of installations [11,27]. For example, the Swiss National Air Pollution Monitoring Network (NABEL) operates 16 measurement stations, distributed over an area of 41,285 km^2; the distance between two stations often exceeds 20 km, resulting in a *limited spatial resolution* of the published pollution maps, bearing in mind that NABEL's measurement density is still fairly high compared to

other international monitoring networks [23]. To assure high precision, the instruments are *manually calibrated* (adjusted in the parts-per-thousand range) every 14 days [4]. During calibration, a sensor is exposed to certain gas concentrations, and the sensor's calibration parameters are adjusted such that the difference between applied gas concentration and sensor output is minimized.

The concentration of air pollutants is highly location-dependent. Traffic junctions, urban canyons, industrial installations, and topological structure all have considerable impact on the local air pollution [26]. In recent years, several research groups started measuring the chemical pollutants in the atmosphere with low-cost solid-state gas sensors. Most of these sensors show an electrochemical reaction when exposed to a specific gas. The gas concentration is determined by measuring either the sensor's output current or the resistance of the sensor's tin dioxide layer. These solid-state gas sensors are inexpensive, small, and suitable for mobile measurements. Hence, researchers started to integrate them in mobile sensor nodes [12,15], for example, using public transport vehicles such as trams [3] and buses [7]. The nodes' mobility allows to obtain a higher spatial measurement resolution and to increase the covered area without the need of hundreds or thousands of sensors.

Challenges. The main drawbacks of low-cost gas sensors are their limited accuracy and resolution, low stability, and poor selectivity. Low-cost gas sensors are usually installed in industrial production factories to measure high-level concentrations (*e.g.*, in the automotive industry [24]). However, environmental gas monitoring requires to measure very low gas concentrations. Additionally, low-cost gas sensors are unstable [16]. Due to sensor aging [16] they must be re-calibrated every month [13,25] or even every week [17]. Another great challenge is the sensors' poor selectivity. The sensors' output is sensitive to ambient humidity and responsive to influence of interfering gases [17]. Hence, frequent sensor calibrations are required to improve or at least preserve the desired measurement accuracy. Manual sensor calibration, however, is an elaborate and time-consuming task.

Contribution and road-map. In this paper, we propose automatic calibration algorithms to improve the measurement accuracy of mobile low-cost gas sensors mounted on top of public transport vehicles. Throughout this paper we assume that *sensors are mobile* and equipped with GPS modules to determine their positions. Each gas sensor encounters fixed stations and other low-cost gas sensors. We let sensors exploit these rendezvous by making use of each others sensor readings to improve their calibration and thus to increase their measurement accuracy "on-the-fly." In Sec. 2 we introduce a calibration model that is used in Sec. 3 to describe our calibration algorithms. *Forward calibration* is based on a traditional approach that uses recent sensor readings to calculate new calibration parameters. We use forward calibration as a performance benchmark for our novel algorithms. *Backward calibration* re-evaluates measured ozone concentrations by integrating most up-to-date sensor readings into the calculation of the calibration parameters. However, these re-evaluated concentration measurements are not immediately available, introducing a significant information delay. *Instant calibration* is able to achieve almost the accuracy of backward calibration without any delay by constantly adjusting the calibration parameters.

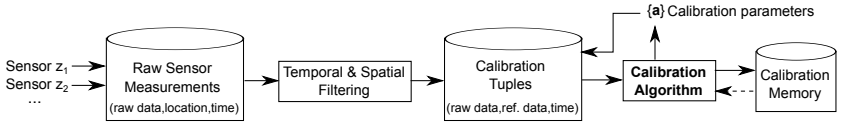


Fig. 1. Calibration pipeline. *Sensor readings are filtered based on their measurement time and location. Tuples are generated from the filtered data and used to calculate calibration parameters.*

In our scenario, sensor calibration is either based on reliable measurements from a fixed station (single-hop) or on unreliable sensor readings from low-cost sensors (multi-hop). We discuss the resulting difficulties if the calibration is based on inaccurate pollution concentration measurements and describe our solution approach in Sec. 4. We set up a six-week measurement campaign to analyze in Sec. 5 the accuracy of the described calibration algorithms using real pollution measurements. In order to know the ground truth pollution at all times, we use a static setting. The ground truth is delivered from a fixed station next to our installation. We simulate the mobility patterns of public transport vehicles in order to mimic that sensors are only from time to time in the vicinity of a fixed station. We show that instant calibration is able to calibrate the gas sensors such that they deliver measurements difference as low as ± 2 ppb from the ground truth of the fixed station, both when the calibration is based on reliable and unreliable sensor readings. This error is a factor of two smaller than achieved with forward calibration. Furthermore, we simulate sensor readings to investigate the measurement accuracy of up to 20 gas sensors that leverage each others measurements to improve their calibration. We find a linear dependency of the calibration accuracy on the number of calibration hops. We survey related work in Sec. 6, and conclude the paper in Sec. 7.

2 Calibration System and Model

In this section, we describe our calibration system and introduce the model that is used to describe our calibration algorithms. For the sake of convenience, we only introduce those parts of our calibration model required to calibrate a low-cost sensor based on reference measurements from a perfect sensor, *i.e.*, *single-hop* calibration. In Sec. 4 we discuss the necessary model extensions to support *multi-hop* calibration, where sensor calibrations are based on unreliable sensor readings.

Calibration System. We start by giving a high-level overview of the calibration system as illustrated in Fig. 1. We use data streams from different sensors containing raw data, measurement time, and measurement location to improve sensor calibration. Two sensors are exposed to very similar gas concentrations if they appear to be at similar locations at similar times. Hence, potentially the sensors can use each others sensor readings to improve their calibration quality, *i.e.*, reduce the difference between the calibrated measurement and the actual phenomenon. We stream all measurements through a data filter in order to select those sensor readings that were taken by two spatially and temporally close sensors. Whether two sensors are considered to be in each others vicinity, depends on the type of gas sensor used and the chemical properties of the monitored pollutant. We use the measurements that passed the filter to construct calibration tuples containing the measurements of two close-by sensors. The calibration algorithm

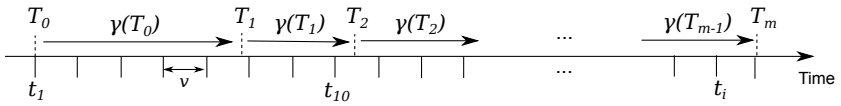


Fig. 2. Calibration model. *A sensor takes measurements with a constant sampling interval v. Calibration parameters evaluated at time T_m are denoted with $\gamma(T_m)$.*

calculates new calibration parameters as soon as either a certain number of calibration tuples is collected or if a predefined period of time has elapsed since the most recent calibration. The accumulated calibration tuples and possibly data from the calibration memory with older tuples used in previous calibrations serve as input to the calibration algorithms.

Model. We continue with the underlying calibration model. Let a phenomenon of interest H exhibit a continuous measurable signal $h : T \to D$ with time $T \subseteq \mathbb{R}^+$ and domain of measured values $D \subseteq \mathbb{R}$. A sensor z takes measurements of H with a constant sampling interval $v \in \mathbb{R}^+$. This leads to a sequence of discrete measurements $\{p(t_i)\}$ at times $t_i \in T$, $i \in \mathbb{N} \setminus \{0\}$ with $t_{i+1} - t_i = v$ as shown in Fig. 2. We consider a measurement as point measurement, that is, it has no duration. If sensor z is perfect at any point in time then $p(t_i) = h(t_i)$, where $h(t_i)$ is the actual phenomenon. However, low-cost gas sensors typically suffer from the effects of aging [16,19] and are also highly sensitive to ambient humidity and interfering gases [17]. Both have a significant influence on the sensor measurements and result in a deviation of $p(t_i)$ from $h(t_i)$. Sensors try to minimize this deviation by adjusting their *calibration curve*. Let $c : \mathbb{R}^{k+1} \times D \to D$ denote the sensor's calibration curve abstracted with a polynomial of order k with a vector of *calibration parameters* $\boldsymbol{a} = (a_0, a_1, \ldots, a_k) \in \mathbb{R}^{k+1}$. Thus,

$$c(\boldsymbol{a}, x) = \sum_{n=0}^{k} a_n \cdot x^n \tag{1}$$

where x is a raw sensor reading. Let $\gamma : T \to \mathbb{R}^{k+1}$ be a function that returns the calibration parameters of the most recent sensor calibration, *i.e.*, $\boldsymbol{a} = \gamma(t_i)$ (see Fig. 2). The calibrated measurement $\tilde{p}(t_i)$ of sensor z at time t_i is then:

$$\tilde{p}(t_i) = c(\gamma(t_i), p(t_i)) \tag{2}$$

A *perfect sensor* has $\gamma(\cdot) = (0, 1, 0, 0, \ldots, 0) \in \mathbb{R}^{k+1}$, *i.e.*, $\tilde{p}(t_i) = p(t_i) = h(t_i)$.

Consider two sensors z_1 and z_2 taking measurement sequences $\{p(t_i)\}$ and $\{q(t_j)\}$ with constant sampling intervals.We assume that z_2 is a perfect sensor providing accurate measurements. Sensor z_1 improves its calibration parameters by comparing its measurements to those from z_2. In this case sensor z_2 is the calibration parent of z_1. Additionally, sensor z_1 can be mobile (*e.g.*, placed on top of a public transport vehicle) and thus, in general, only from time to time be located in the vicinity of z_2. We use a filter to consider only sensor readings of z_1 and z_2 as useful calibration input if they were measured in each others temporal and spatial vicinity. Consider that set C contains the sensor readings that passed the filter. These are combined into tuples

$$(p(t_i), \tilde{q}(t_j), t_l) \tag{3}$$

with sensed raw data $p(t_i)$ from z_1, calibrated reference measurement $\tilde{q}(t_j)$ from z_2, and the *joint measurement time* of this tuple $t_l = (t_i + t_j)/2$.

Due to aging, correlation of raw data and reference measurement of a tuple loose significance as time advances. T_m denotes the point in time when a calibration was performed, where T_0 indicates the initial sensor calibration (see Fig. 2). We evaluate the remaining expressiveness of a data tuple by means of a function $w : T \rightarrow [0, 1]$, *i.e.*, at time of calibration T_m the remaining expressiveness of a tuple with joint measurement time t_l is $w(T_m - t_l)$. Because of aging, w monotonically decreases with increasing time difference $T_m - t_l$, $T_m \geq t_l$ since only past tuples are used to calculate calibration parameters. Note that if no sensor aging is considered then $w(t) \equiv 1$ for any $t \in T$. Assume that at time T_m sensor z_1 is calibrated using sensor z_2 as reference. Then the new calibration parameters $\gamma_1(T_m)$ are chosen such that the sum of differences between $\tilde{p}(t_i) = c(\gamma_1(T_m), p(t_i))$ and $\tilde{q}(t_j) = c(\gamma_2(T_m), q(t_j))$ weighted with $w(T_m - t_l)$ are minimized. We apply linear regression by using the method of least squares [6] to calculate the calibration parameters $\gamma_1(T_m)$. The age-dependent weight determines the influence of each tuple on the final calibration curve. Thus up-to-date tuples have higher impact than outdated measurements. All tuples in C and possibly former tuples from the calibration memory contribute to the above sum.

3 Calibration Algorithms

Low-cost gas sensors are either delivered uncalibrated or the factory calibration is not intended for low concentration measurements as found when monitoring urban air pollutants. Additionally, factory calibration parameters are often only based on measurements of two to three different gas concentrations under one specific temperature and humidity setting. Possibly, the chosen setting is not suitable for the intended measurement campaign. Hence, it is essential for the customers to adjust the calibration parameters of these low-cost gas sensors to their needs.

There are two common approaches for the calibration of gas sensors that are used to monitor urban air pollutants. The calibration can take place either in the laboratory using artificial gas mixtures [12] or in the field with real pollution measurements by placing the sensor close to a fixed station providing reliable measurements [7,10,17]. The disadvantage of calibrating the sensors with real pollution measurements is the dependency on the weather conditions and the local pollution dispersion that both cannot be controlled. For a precise calibration, reference measurements under a wide range of environmental conditions are desirable, *e.g.*, low, middle, and high concentrations of the target pollutant under distinct humidity and temperature settings [25]. Nevertheless, we calibrate our gas sensors in the field, since for us the advantages prevail: *(i)* the sensor is calibrated under very similar conditions as in the later deployment (*e.g.*, same hardware and software components, same water and dust cover), *(ii)* reliable measurements from fixed stations are freely available, and *(iii)* with a large number of reference measurements it is possible to calculate precise calibration parameters.

In this paper, we introduce three calibration algorithms: forward calibration, backward calibration and instant calibration. We begin with *forward calibration*, which is based on a traditional approach to calibrate gas sensors widely used in the litera-

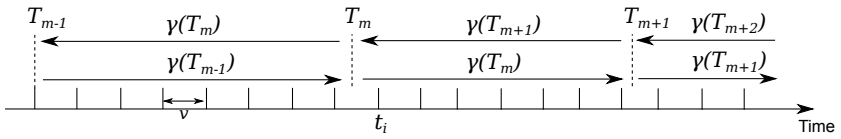


Fig. 3. Backward calibration. *Calibration parameters $\gamma(T_m)$ are used to get temporary measurements for sensor readings in $[T_m, T_{m+1})$ and to re-evaluate sensor readings in $[T_{m-1}, T_m)$.*

ture [10,12,17]. We use it in our evaluation as performance benchmark for two novel calibration algorithms, *backward* and *instant calibration*, which show different trade-offs between measurement accuracy and calibration delay.

Forward calibration. As done by many traditional approaches, forward calibration calculates new calibration parameters based on recent sensor readings. Calibration parameters can be calculated in regular time intervals or as soon as certain number of calibration tuples is available in C. At time T_m the calibration parameters $\gamma(T_m)$ are chosen to minimize the weighted sum of squared differences between $c(\gamma(T_m), p(t_i))$ and $\tilde{q}(t_j), \forall\, (p(t_i), \tilde{q}(t_j), t_l) \in C$:

$$\arg\min_{\gamma(T_m)} \sum_{(p(t_i),\tilde{q}(t_j),t_l)\in C} w(T_m - t_l) \cdot \Big(c(\gamma(T_m), p(t_i)) - \tilde{q}(t_j)\Big)^2 \qquad (4)$$

Note that all calibration tuples in C have a joint measurement time t_l in the interval $[T_{m-1}, T_m)$. Only these tuples are considered by most traditional calibration approaches in order to accurately capture the momentary sensor characteristics. Forward calibration uses the calibration parameters $\gamma(T_m)$ for all $p(t_i)$ where $t_i \in [T_m, T_{m+1})$.

The main drawback of using forward calibration is that the data used to determine the calibration parameters are always based on slightly outdated sensor readings. The currentness of the calibration parameters depends on the time passed between two calibration instances, which can be arbitrarily long.

Backward calibration. In contrast to forward calibration, backward calibration additionally re-evaluates already calibrated sensor readings offline. At calibration time T_m new calibration parameters are calculated as defined in (4). The calibration parameters $\gamma(T_m)$ are used to *(i)* calibrate future sensor readings $p(t_i)$ where $t_i \in [T_m, T_{m+1})$ and *(ii)* to *recalculate* former sensor readings $p(t_i)$ where $t_i \in [T_{m-1}, T_m)$, as depicted in Fig. 3. The recalculation improves the measurement accuracy if the sensor characteristics significantly differ during time periods $[T_{m-2}, T_{m-1})$ and $[T_{m-1}, T_m)$.

The main disadvantage of backward calibration is that the recalculated sensor measurements are only available after a delay, which depends on the time passed between two calibration instances, that is $T_m - T_{m-1}$. Additionally, the sensor readings $p(t_i)$ have to be recorded in order to enable the recalculation of these measurements with new calibration parameters. We continue with instant calibration that trades off these disadvantages with a slight loss of calibration accuracy.

Instant calibration. In contrast to forward and backward calibration, instant calibration does not only use calibration tuples in C to calculate new calibration parameters, but also former calibration tuples from the *calibration memory*. Therefore, instant calibration can calculate new calibration parameters more frequently; new calibration param-

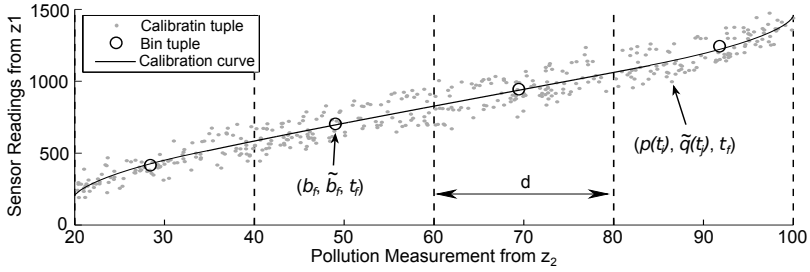


Fig. 4. Instant calibration. *The measurement range is partitioned into equally spaced bins. Each bin is represented by a single averaged value pair* $(b_f, \tilde{b}_f, t_f)$.

eters are calculated each time a new tuple is available in C. The sensor's measurement range is partitioned into several bins in order to increase the significance of the new calibration tuple. For that reason, all calibration tuples of the calibration memory that belong to the same bin, are age-based weighted and represented by one averaged tuple per bin, as detailed below.

Fig. 4 shows an example with calibrated pollution measurements in the range from 20 ppm to 100 ppm which are portioned into equally sized bins of size $d = 20$ ppm. Each bin f is represented by a single bin tuple $(b_f, \tilde{b}_f, t_f)$ denoting the age-based weighted moving average of the sensor readings from sensor z_1, the calibrated measurements from sensor z_2, and the time of the most recent bin adjustment t_f, respectively. The calibration process starts by selecting an appropriate bin $f = \lfloor p(t_i)/d \rfloor$ based on the calibration tuple $(p(t_i), \tilde{q}(t_j), t_l) \in C$. Then, the bin tuple $(b_f, \tilde{b}_f, t_f)$ is at calibration time T_m adjusted as follows:

$$b_f^{(m)} = \frac{\sum_{s<m} w(T_m - t_f^{(s)}) \cdot b_f^{(s)} + w(0) \cdot p(t_i)}{\sum_{s<m} w(T_m - t_f^{(s)}) + w(0)} \tag{5}$$

$$\tilde{b}_f^{(m)} = \frac{\sum_{s<m} w(T_m - t_f^{(s)}) \cdot \tilde{b}_f^{(s)} + w(0) \cdot \tilde{q}(t_j)}{\sum_{s<m} w(T_m - t_f^{(s)}) + w(0)} \tag{6}$$

$$t_f^{(m)} = t_l \tag{7}$$

Previously calculated bin tuples only contribute to the above sums if $w(T_m - t_f^{(s)}) > 0$. The adjusted bin tuple is used to calculate the calibration parameters $\gamma(T_m)$ that minimize the weighted sum of squared differences between $c(\gamma(T_m), b_f)$ and $\tilde{b}_f$ for all bins:

$$\arg\min_{\gamma(T_m)} \sum_{\forall(b_f, \tilde{b}_f, t_f)} w(T_m - t_f) \cdot (c(\gamma(T_m), b_f) - \tilde{b}_f)^2 \tag{8}$$

The bin size defines the adjustment speed of the calibration parameters. The change in value of a single bin tuple $(b_f, \tilde{b}_f, t_f)$ has almost no effect on the calibration parameters if there are many bins, but if there are only a few the parameters change quickly.

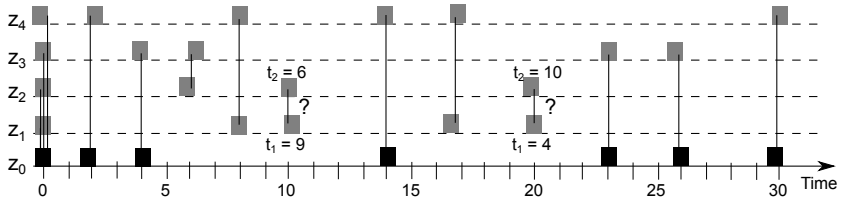


Fig. 5. Multi-hop calibration. *Connected sensors are in each others temporal and spatial vicinity.*

4 Multi-hop Calibration

In the previous section, we introduced three algorithms to calibrate sensors based on reliable measurements of a perfect sensor (*i.e.*, single-hop calibration). In this section, we describe the extensions to support multi-hop calibration in order to facilitate the calibration of sensors which are rarely or never in the vicinity of a perfect sensor. Hence, their calibration is mostly based on unreliable measurements.

Consider five mobile sensors $z_0, \ldots, z_4$ that happen to be in each others vicinity from time to time. We denote with z_0 a perfect sensor while z_1 to z_4 are low-cost gas sensors. The movement patterns of the sensors determine whether and how regularly two sensors are at similar locations, as illustrated in Fig. 5. The lines connect sensors which are in each others temporal and spatial vicinity. The ground truth concentration is unknown if both instruments are low-cost sensors delivering unreliable measurement. We approximate the actual concentration by weighting the measurements with the *calibration ages* of the two sensors that depend on: the time passed since the latest calibration and the quality of the used reference measurements for that calibration.

Hence, we define the age of a calibration as the sum of *(i)* the time passed since a sensor's most recent calibration T_m and *(ii)* the average calibration age of the sensor's calibration parents. A perfect sensor has calibration age 0 at all times. To compute the calibration age locally, we additionally store with each calibration tuple the calibration ages of the sensors building the calibration tuple. Let the calibration ages of sensors z_1 and z_2 be t_1 and t_2, respectively. Assume that sensors z_1 and z_2 are at time step t_l in each others vicinity. None of the two sensors is able to accurately measure the ground truth concentration. We combine the measurements of the two sensors and assume for this reason that all low cost sensors are identically constructed and thus have similar precision and sensitivity characteristics, much like [12,15]. We solely examine the calibration ages of z_1 and z_2 to decide on their measurement quality. Therefore, given the calibration ages in Fig. 5, we assume that at time step 10 the output of sensor z_2 is closer to the ground truth concentration than the output of z_1, whereas at time step 20 the output of z_1 is preferred. We extend the tuple in (3) to

$$\left(p(t_i), \frac{t_1 \cdot \tilde{q}(t_j) + t_2 \cdot \tilde{p}(t_i)}{t_1 + t_2}, t_l, t_2\right) \tag{9}$$

$$\left(q(t_j), \frac{t_1 \cdot \tilde{q}(t_j) + t_2 \cdot \tilde{p}(t_i)}{t_1 + t_2}, t_l, t_1\right) \tag{10}$$

which weights the sensor measurements inversely proportional to the sensor calibration ages. The above transformation has the following two properties: *(i)* the approximated

ground truth concentration is exactly the averaged measurement of both sensors if both sensor calibration ages are equal; *(ii)* if one of the sensors is perfect, its calibration age 0 and the ground truth is given by (3).

5 Evaluation

In this section, we evaluate experimentally the accuracy of our calibration algorithms, and analyze multi-hop calibration and sensor aging in simulation.

5.1 Experimental Results

Two ozone gas sensors are integrated into a *CoreStation* [8] and installed next to a fixed station which delivers reliable ozone concentration measurements. We assume that the fixed station has a perfect sensor and take its measurements as the ground truth. The measurements from the fixed station are used *(i)* as calibration reference and *(ii)* to evaluate the accuracy of the three calibration algorithms.

Setup. The deployed gas sensors are two identical MiCS-OZ-47 ozone sensing heads from e2v [1]. They are delivered with a transmitter PCB and are equipped with temperature and humidity sensors providing data over a RS232-TTL interface. The reported ozone concentration is based on the resistance of the sensor's tin dioxide layer. As the resistance is heavily temperature dependent, we use the on-board temperature sensor to calculate a temperature compensation. The gas sensors are factory calibrated and reported to work in the range from 20 ppb to 200 ppb with an accuracy of $\pm$20 ppb. The supported temperature range is between 10 °C and 40 °C, the humidity range between 20 % and 90 %. Due to aging effects the lifetime is limited to three years. We model the sensor's sensitivity loss w using an exponential function that maps $t \mapsto \alpha^t$ where $\alpha = 0.9$.

We assume that the fixed station has a perfect sensor. It measures the ozone concentration once per minute, but we only have access to 10-minute averages. The two ozone sensors are installed on the outer wall of the station. They measure the ozone concentration once per minute and transmit the computed ozone concentration (based on the factory calibration) as well as raw data (resistance of the tin dioxide layer, temperature, and humidity) to our server.

We analyze data from a six-week measurement campaign, which results in the evaluation of over 8,000 measurements from the fixed station and more than 160,000 readings from the two gas sensors. First, we study the accuracy of the factory calibration. Next, we use the reference measurements from the fixed station to initially calibrate the sensors according to our needs. Then, we apply either forward, backward, or instant calibration to periodically recalibrate the sensors and compare the accuracy of the different algorithms. Finally, we show the impact of multi-hop calibration on the accuracy of the sensor measurements.

Performance Measures. We evaluate the efficacy of a calibration algorithm by looking at the differences of calibrated sensor readings $\tilde{p}(t_i)$ and measurements of the fixed station $\tilde{q}(t_j)$ during time interval $[T_S, T_E)$ with $T_S, T_E \in T$. The set $B(T_S, T_E) = \{t_l \mid T_S < t_l \leq T_E\}$ contains all joint measurement times t_l of calibration tuples

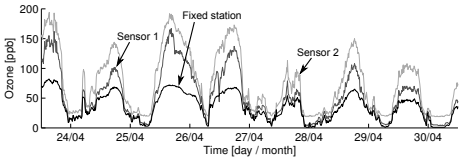


Fig. 6. Factory calibrated gas sensor measurements compared to data from the fixed station.

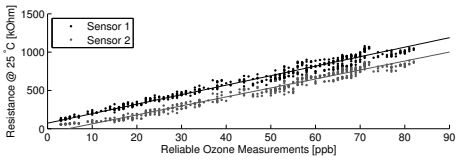


Fig. 7. Initial gas sensor calibration based on measurements from the fixed station.

generated between T_S and T_E. For all these tuples we calculate $d(t_l) = |\tilde{p}(t_i) - \tilde{q}(t_j)|$ and evaluate the *mean absolute error* M and the *standard deviation* S:

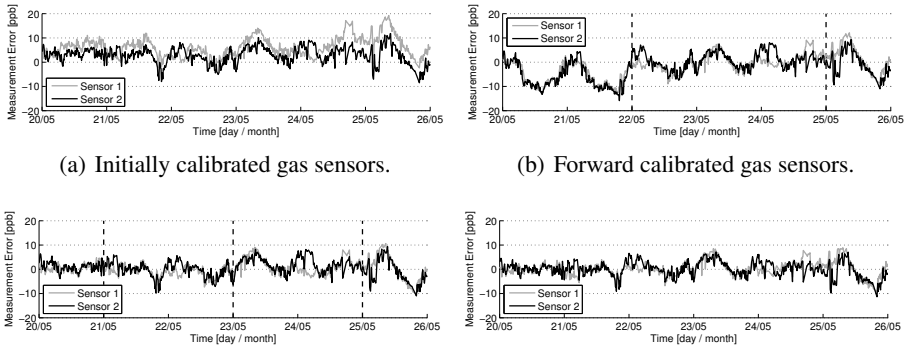
$$M = \frac{1}{|B(T_S, T_E)|} \cdot \sum_{t_l \in B(T_S, T_E)} d(t_l) \tag{11}$$

$$S = \Big(\frac{1}{|B(T_S, T_E)|} \cdot \sum_{t_l \in B(T_S, T_E)} (d(t_l) - M)^2 \Big)^{\frac{1}{2}} \tag{12}$$

Initial Calibration. The gas sensors are delivered with a factory calibration that does not fulfill our accuracy demands. Fig. 6 shows the measured ozone concentration from the two gas sensors and the fixed station over a period of one week. The measurements are off the mark by up to a factor of two. We use sensor readings from the first installation days to calculate custom initial calibration parameters $\gamma(T_0)$ for both gas sensors. Having the tin dioxide layer resistance R at ambient temperature $\mathcal{T}$, we calculate the resistance $\tilde{R}$ at the reference temperature $\mathcal{T}_0 = 25\,^{\circ}\mathrm{C}$ with $\tilde{R} = R \cdot e^{K \cdot (\mathcal{T} - \mathcal{T}_0)}$ where $K = 0.019$ is the temperature coefficient given in the sensor data sheet [1]. The calibration parameters $\gamma(T_0)$ are chosen to minimize the sum of squared differences between the sensor readings $\tilde{R}$ and the reliable measurements from the fixed station. Fig. 7 depicts sensor readings, reference measurements, and computed calibration curves for the two gas sensors. The output of the two sensors are already from the very beginning fairly different, which underlines the need for individual sensor calibrations.

Fig. 8(a) depicts the measurement errors of a one-week extract from our measurement campaign. We calculate the errors by comparing the gas sensor readings to the measurements from the fixed station. It can be seen that the measurement error varies significantly over time. Performing an initial calibration leads to considerably more precise sensor measurements than using the factory calibration, reducing the average measurement error by a factor of four.

Periodic Calibration. After the initial calibration, sensors are periodically calibrated to keep the calibration parameters up to date using either forward, backward, or instant calibration. We consider a setup where the two gas sensors are part of a mobile air quality monitoring system deployed on top of public transport vehicles. We assume the following: *(i)* public transport vehicles are moving on a known and fixed track, *(ii)* vehicles pass by a fixed station every 40 minutes, and *(iii)* vehicles continuously operate between 5.00 AM and 1.00 AM. We use measurements from our static setting to simulate the above described mobility pattern. Hence, the gas sensors only have access



(a) Initially calibrated gas sensors.

(b) Forward calibrated gas sensors.

(c) Backward calibrated gas sensors.

(d) Instant calibrated gas sensors.

Fig. 8. Measurement errors depicted for a time period of one week. *The vertical dashed lines denote the times of calibration when using forward and backward calibration. Instant calibration adjusts the calibration parameters every 40 minutes.*

to reference measurements once every 40 minutes between 5.00 AM and 1.00 AM, which correspond to around 30 reliable measurements per day.

Figs. 8(b)–(d) show measurement errors of the two gas sensors when compared to data from the fixed station. The calibration period is denoted with dashed lines for forward and backward calibration, instant calibration is adjusting the calibration parameters after every available calibration tuple, *i.e.*, every 40 minutes. In Fig. 8(b) forward calibration shows a distinct offset compared to the reference measurements, which is fixed with the calibration on the 22nd May. The offset probably originates from the calibration performed on the 19th May based on outdated data. With backward and instant calibration (see Fig. 8(c) and Fig. 8(d)) the measurement errors are considerably reduced compared to forward calibration. The absolute mean errors M and the standard deviations S of both gas sensors when compared to the measurements of the fixed station during the six-week deployment period are shown in Fig. 9. The achieved measurement accuracy of instant calibration is very similar to the one of backward calibration with the advantage that accurate measurements are immediately available. The measurement error of ± 2 ppb is acceptable for ozone concentration monitoring, given that ordinary daily measurements range between 0 ppb and 70 ppb, and that the European Union sets the information and alert thresholds to 90 ppb (1 h average) and 120 ppb (1 h average), respectively [2]. We show in Table 1 that with the improved calibration accuracy it is possible to precisely measure whether the ozone concentration is above a certain threshold, *e.g.*, 60 ppb. Using backward calibration the relative error of the calculated total hours sensed above 60 ppb are below 1 % for both sensors when compared to the hours evaluated by the fixed station. With instant calibration, still both sensor's relative errors in measuring this time duration are below 5 %.

Furthermore, we evaluate the measurement errors M and S of instant calibration for different bin sizes d. The errors are minimized if the measurement range is segmented in 3 to 4 bins ($d = 25$ ppb), hence, the calibration process is able to quickly adapt the calibration parameters to varying sensor characteristics. The calibration period has a large impact on the measurement errors when using forward and backward calibration. Fre-

Table 1. The total amount of time while the calibrated sensors measured more than 60 ppb (left) and their relative errors when compared to the 66.2 h measured by the fixed station (right).

Calibration	Initial	Forward	Backward	Instant	Initial	Forward	Backward	Instant
Sensor 1	90.5 h	54.3 h	65.8 h	64.6 h	36.7 %	18.4 %	0.6 %	2.4 %
Sensor 2	231.3 h	69.5 h	65.6 h	69.3 h	249.4 %	5.0 %	0.8 %	4.7 %

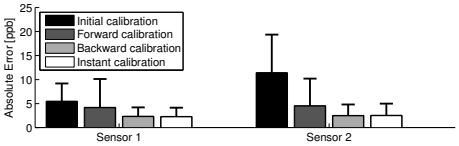


Fig. 9. Mean and standard deviation of the absolute measurement errors of the two gas sensors.

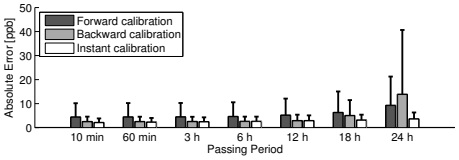


Fig. 10. Calibration accuracy for passing periods between 10 minutes and 1 day.

quent calibrations help to quickly adjust to temporal changing of sensor characteristics. However, a too short calibration period involves the danger of calculating calibration parameters that do not optimally represent the sensor characteristics as the parameters are only based on a few reference measurements. Extensive evaluations with different calibration periods show that it is best to calibrate the sensors every three and two days for forward and backward calibration, respectively. The period for forward calibration is longer, because it relies on the knowledge of the average sensor characteristics to calculate upcoming sensor readings as effectively as possible. Backward calibration tries to capture short-term temporal differences to recalculate sensor measurements, hence its calibration period is shorter.

Until now, we assumed that the gas sensors are in the vicinity of a fixed station every 40 minutes (passing period). Since we assume that the sensors are deployed on top of pubic transport vehicles, our influence on this passing period is very limited. Hence, we analyze the dependency of the calibration accuracy on the passing period, that has a direct effect on the number of available reference measurements. We evaluate the calibration accuracy for various passing periods between 10 minutes and 24 hours. The decreasing number of available reference measurements has the least impact on the measurement accuracy of instant calibration, as shown in Fig. 10. Forward and backward calibration encounter problems as soon as new calibration parameters are calculated based on only a few reference measurements.

Multi-hop calibration. We continue using our static setting to evaluate the loss in measurement accuracy if the gas sensors are calibrated over multiple hops. To this end, we assume that only one of the two sensors passes by a fixed station. This sensor is single-hop calibrated as its calibration is based on reliable measurements. The other sensor uses the sensor readings of the calibrated gas sensor as reference. Just as before, we assume that the two gas sensors are in each others vicinity every 40 minutes. Fig. 11 shows the slight increase of the mean measurement error M and its standard deviation S if the gas sensors are calibrated over two hops. With two sensors we can at most compare the calibration accuracy of one and two hops. We circumvent this restriction by extending our multi-hop analysis with simulated sensor readings.

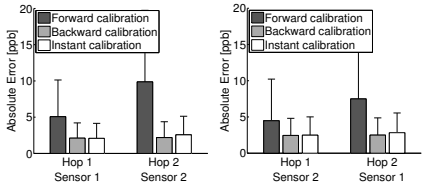


Fig. 11. Absolute measurement errors when sensors are calibrated over one and two hops.

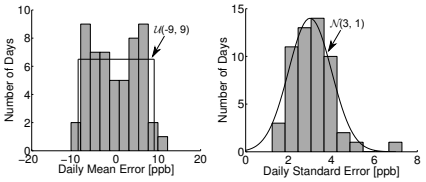


Fig. 12. Mean and standard deviation of the daily average measurement error from sensor 1.

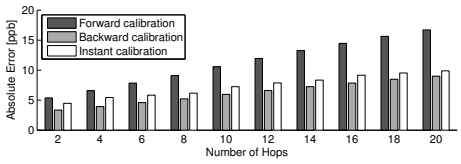


Fig. 13. Simulative analysis of the measurement errors when calibrating over multiple hops.

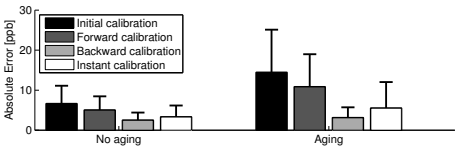


Fig. 14. Simulative analysis of the influence of sensor aging on the calibraiton accuracy.

5.2 Simulations

Analyzing the measurement accuracy of the initially calibrated sensors from our six-week measurement campaign results in the following observations. The measurement errors *(i)* vary significantly over time as depicted in Fig. 8(a) and *(ii)* if only initially calibrated are normally distributed $\mathcal{N}(\mu, \sigma^2)$ with $\mu \sim \mathcal{U}(-9, 9)$ ppb and $\sigma \sim \mathcal{N}(3, 1)$ ppb over the period of one day as shown in Fig. 12. Accordingly, we generate artificial sequences of occurring ozone concentrations, $h(t_i)$, and corresponding gas sensor measurements, $p(t_i)$, for $t_i \in [0, 365]$ days with a sampling interval v of 10 minutes. For each day a new μ and σ is chosen, *i.e.*, for 144 consecutive sensor readings. We only consider positive values for σ.

Multi-hop calibration. We simulate the measurements of 20 gas sensors $z_1, \ldots, z_{20}$. To obtain the maximum hop distance, we assume that the sensors are calibrated in line, *i.e.*, calibration of sensor z_i is based on measurements from z_{i-1}. Sensor z_0 is the perfect sensor of a fixed station not requiring any calibration. Fig. 13 shows the absolute mean calibration error for sensors calibrated over 1 to 20 hops. The measurement errors are averages of three simulation runs over a simulated time of 365 days. As expected the error linearly increases with the number of calibration hops. In general, public transport networks are dense, hence, the maximum calibration hop distance is rather limited. As long as a sensor using instant calibration is calibrated over less than six hops, its measurement error is smaller compared to calibrating over a single hop using forward calibration.

Sensor aging. Solid-state gas sensors suffer from loss of sensitivity over time, which limits their lifetime to about three years. The sensitivity loss reduces the measurement accuracy. We simulate these aging effects by adjusting the uniform distribution of μ over time using $\mu \sim \mathcal{U}(-9 - \frac{d}{5}, 9 + \frac{d}{5})$, where d denotes the number of days the sensor is in use. Thus, the maximum possible measurement error increases each day by

0.2 ppm [16]. Again, we simulate sensor readings over a period of 365 days and show in Fig. 14 the average measurement errors and standard deviations of three simulation runs. Sensor aging has the least effect on the accuracy of backward calibration since the measured concentrations are postprocessed. The measurement errors of instant calibration is considerably lower than those of forward calibration. Comparing these results to previous simulations with multi-hop calibration reveals that calibrating sensors over multiple hops has a higher influence on the measurement error than sensor aging as long as backward or instant calibration is used.

6 Related Work

A wide range of research projects have recently emerged dealing with air pollution monitoring using low-cost gas sensors [22,25]. There is significant amount of work about calibrating sensors in general, but only little attention has been given to the fact that these sensors are often installed in sensor nodes and are part of a mobile sensor network [12,15], *e.g.*, by using public transport networks [3,7]. We exploit co-located sensor measurements to improve sensor calibration.

Miluzzo *et al.* propose CaliBree [21], a self-calibration system for mobile sensor nodes. A sensor is considered to be at the same location as a ground-truth node if both their readings are virtually identical. Additionally packets are broadcast by the ground-truth nodes to allow sensor nodes to detect whether they are approaching a ground-truth node by evaluating the received signal strength indicator (RSSI). Our sensor nodes have a built-in GPS module. As a result, it is possible to accurately check the position of a node in order to evaluate whether it is in the vicinity of another node.

Both CaliBree [21] and the in-place sensor calibration from Bychkovskiy *et al.* [9] assume to have enough nodes with perfect sensors delivering reliable measurements to provide a single-hop calibration to all low-cost sensors in the network. Both attempts do not consider that a calibration of a low-cost sensor can also be based on sensor readings of another (calibrated) low-cost sensor. We distinguish single-hop calibration (between a low-cost sensor and a perfect sensor), where reliable reference measurements are available, from multi-hop calibration (between two low-cost sensors), where the sensor measurements are not reliable. Our algorithm deals with the limited accuracy of the reference measurements of low-cost sensors appearing when calibration is performed over multiple hops.

Balzano and Nowak introduce blind calibration [5] whereby no perfect sensor is required for calibration. By oversampling the signal of interest, they recover the unknown sensor gain. Just the same as a large group of calibration techniques [9,15,21,17,25] these all strongly depend on the assumption that only aging effects invalidate calibration parameters, *i.e.*, these calibration approaches only adjust gain and offset which are the calibration curve's slope and the sensor's output in clear air. Against this, our field experiments reveal that low-cost gas sensor characteristics are more complex. Besides the aging effect also changing environmental conditions have a serious effect on the sensor readings. Our algorithms do not have the above mentioned constraints regarding the nature of the captured signal.

Already some previous work [10,17] proposed to initially calibrate gas sensors based on real pollution measurements. We pursue this road and also perform (besides the initial calibration) the re-calibrations based on real pollution data, such as [7].

In contrast to previous work presenting automatic calibration algorithms with rather simple sensors measuring the temperature [21], humidity [5], or light intensity [9], we evaluate our calibration approaches with ozone gas sensors in an urban setting and compare the sensor measurements to highly accurate reference data from a fixed station.

Gas sensor calibration reveals similarities with time synchronization. Time synchronization protocols such as FTSP [20] try to estimate the clock drift over time. The drift is calculated by analyzing previously measured clock deviations and using linear regression for its compensation. Forward and instant calibration show similarities with FTSP which uses linear regression to estimate clock drifts based on prior clock drift measurements. Backward calibration bears resemblance to Sundial [14] that reconstructs global timestamps offline. Compared to these time synchronization protocols we introduce an age-dependent weight in the linear regression to give up-to-date measurements a higher impact on the calculated calibration parameters.

7 Conclusions

We have addressed the challenge of on-the-fly calibration of low-cost gas sensors. Our calibration algorithms exploit the sensors' mobility to repeatedly improve the calibration parameters of the sensors. This is required as the sensor characteristics are highly temporally dependent due to the influence of sensor aging and the cross-sensitivity to gases in the atmosphere. We show with real pollution measurements and simulations that with instant calibration the measurement error is reduced by a factor of two compared to the traditional approach widely used in the literature. Especially, using instant calibration our gas sensors are able to measure the ozone concentration with an error of ± 2 ppb only when compared to measurements of high-quality instruments. This is remarkable as the accuracy given in the datasheet is ± 20 ppb [1]. We also find a linear dependency of the calibration accuracy on the number of calibration hops. The accuracy loss is tolerable as long as the number of calibration hops is rather limited, as found in public transport networks. Furthermore, we reveal that among the investigated calibration approaches, instant calibration's measurement accuracy is least dependent on the available number of reference measurements. This is important because both the location of fixed stations and the schedule of the mobile sensor nodes can usually not be changed.

Acknowledgements. We want to thank Christoph Hueglin and Beat Schwarzenbach for sharing their expertise in the field of air pollution monitoring and for enabling us to carry out the measurements at the NABEL station in Duebendorf, Zurich. Further, we thank Federico Ferrari, Marco Zimmerling, and the anonymous reviewers for their valuable feedback that helped us to improve this paper. Finally, the experimental evaluation would not have been possible without the PermaSense platform and the technical support from Jan Beutel, Bernhard Buchli, Tonio Gsell, Matthias Keller, Roman Lim, Christoph Walser, and Mustafa Yuecel. This work was funded by NanoTera.ch with Swiss Confederation financing.

References

1. MiCS-OZ-47. http://www.e2v.com/assets/media/files/sensors_datasheets/Metal_Oxide/mics-oz-47.pdf.
2. EU Directive 2008/50/EC. In *Official Journal of the European Union No. L 152*.
3. K. Aberer, S. Sathe, D. Chakraborty, A. Martinoli, G. Barrenetxea, B. Faltings, and L. Thiele. OpenSense: Open community driven sensing of environment. In *ACM IWGS*, 2010.
4. BAFU and EMPA. Messresultate NABEL. 2009.
5. L. Balzano and R. Nowak. Blind calibration of sensor networks. In *ACM/IEEE IPSN*, 2007.
6. A. Björck. Numerical methods for least squares problems. In *SIAM*, 1996.
7. A. Boscolo and C. Mangiavacchi. Pollution dynamic monitoring system for city air quality control. In *IEEE IMTC*, 1998.
8. B. Buchli, M. Yuecel, R. Lim, T. Gsell, and J. Beutel. Demo abstract: Feature-rich platform for WSN design space exploration. In *ACM/IEEE IPSN*, 2011.
9. V. Bychkovskiy, S. Megerian, D. Estrin, and M. Potkonjak. A collaborative approach to in-place sensor calibration. In *ACM/IEEE IPSN*, 2003.
10. M. C. Carotta, G. Martinelli, L. Crema, C. Malagù, M. Merli, G. Ghiotti, and E. Traversa. Nanostructured thick-film gas sensors for atmospheric pollutant monitoring: quantitative analysis on field tests. In *Elsevier Sensors and Actuators B: Chemical*, 2001.
11. A. Carullo, S. Corbellini, and S. Grassini. A remotely controlled calibrator for chemical pollutant measuring-units. In *IEEE TIM*, 2007.
12. S. Choi, N. Kim, H. Cha, and R. Ha. Micro sensor node for air pollutant monitoring: Hardware and software issues. In *Sensors MEMS*, 2009.
13. S. De Vito, E. Massera, M. Piga, L. Martinotto, and G. Di Francia. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. In *Elsevier Sensors and Actuators B: Chemical*, 2007.
14. J. Gupchup, R. Musaloiu-E., and A. Szalay A. Terzis. Sundial: Using sunlight to reconstruct global timestamps. In *Springer EWSN*, 2009.
15. R. J. Honicky, E. A. Brewer, E. Paulos, and R. M. White. N-smarts: Networked suite of mobile atmospheric real-time sensors. In *ACM NSDR*, 2008.
16. C. Huan, L. Zhiyu, and F. Gang. Analysis of the aging characteristics of SnO2 gas sensors. In *Elsevier Sensors and Actuators B: Chemical*, 2011.
17. M. Kamionka, P. Breuil, and C. Pijolat. Calibration of a multivariate gas sensing device for atmospheric pollution measurement. In *Elsevier Sensors and Actuators B: Chemical*, 2006.
18. D. Mage, G. Ozolins, P. Peterson, A. Webster, R. Orthofer, V. Vandeweerd, and M. Gwynne. Urban air pollution in megacities of the world. In *Elsevier Atmospheric Environment*, 1996.
19. L. Makadmini and M. Horn. Self-calibrating electrochemical gas sensor. In *IEEE TRANSDUCERS*, 1997.
20. M. Maróti, B. Kusy, G. Simon, and Á. Lédeczi. The flooding time synchronization protocol. In *ACM SenSys*, 2004.
21. E. Miluzzo, N. D. Lane, A. T. Campbell, and R. Olfati-Saber. CaliBree: A self-calibration system for mobile sensor networks. In *IEEE DCOSS*, 2008.
22. R. N. Murty, G. Mainland, I. Rose, A. R. Chowdhury, A. Gosaint, J. Berst, and M. Welsh. Citysense: An urban-scale wireless sensor network and testbed. In *IEEE HST*, 2008.
23. U. Nyffeler. Das Nationale Beobachtungsnetz für Luftfremdstoffe. In *BUWAL*, 2001.
24. C. Pijolat, C. Pupier, M. Sauvan, G. Tournier, and R. Lalauze. Gas detection for automotive pollution control. In *Elsevier Sensors and Actuators B: Chemical*, 1999.
25. W. Tsujita, A. Yoshino, H. Ishida, and T. Moriizumi. Gas sensor network for air-pollution monitoring. In *Elsevier Sensors and Actuators B: Chemical*, 2005.
26. S. Vardoulakis, B.E.A. Fisher, K. Pericleous, and N. Gonzalez-Flesca. Modelling air quality in street canyons: a review. In *Elsevier Atmospheric Environment*, 2003.
27. N. Yamazoe and N. Miura. Development of gas sensors for environmental protection. In *IEEE CPMT*, 1995.