

SCAN: Multi-Hop Calibration for Mobile Sensor Arrays

BALZ MAAG, ZIMU ZHOU, OLGA SAUKH, and LOTHAR THIELE, ETH Zurich

Urban air pollution monitoring with mobile, portable, low-cost sensors has attracted increasing research interest for their wide spatial coverage and affordable expenses to the general public. However, low-cost air quality sensors not only drift over time but also suffer from cross-sensitivities and dependency on meteorological effects. Therefore calibration of measurements from low-cost sensors is indispensable to guarantee data accuracy and consistency to be fit for quantitative studies on air pollution. In this work we propose *sensor array network calibration* (SCAN), a multi-hop calibration technique for *dependent* low-cost sensors. SCAN is applicable to sets of co-located, heterogeneous sensors, known as *sensor arrays*, to compensate for cross-sensitivities and dependencies on meteorological influences. SCAN minimizes error accumulation over multiple hops of sensor arrays, which is unattainable with existing multi-hop calibration techniques. We formulate SCAN as a novel constrained least-squares regression and provide a closed-form expression of its regression parameters. We theoretically prove that SCAN is free from regression dilution even in presence of measurement noise. In-depth simulations demonstrate that SCAN outperforms various calibration techniques. Evaluations on two real-world low-cost air pollution sensor datasets comprising 66 million samples collected over three years show that SCAN yields 16% to 60% lower error than state-of-the-art calibration techniques.

CCS Concepts: • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*; • **Hardware** → *Sensor applications and deployments*;

Additional Key Words and Phrases: Sensor Array; Calibration; Urban Sensing

ACM Reference format:

Balz Maag, Zimu Zhou, Olga Saukh, and Lothar Thiele. 2017. SCAN: Multi-Hop Calibration for Mobile Sensor Arrays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 19 (June 2017), 21 pages.

DOI: <http://doi.org/10.1145/3090084>

1 INTRODUCTION

The availability of portable and low-cost air quality sensors has made them promising not only for qualitative air pollution monitoring to raise public awareness, but also for quantitative analysis to facilitate public policies, infrastructure control and health studies. Installed on vehicles [3, 19, 23, 24] or in wearable devices [7, 8, 18, 31, 35], these sensor nodes travel citywide and their users, both professional and amateur, upload air quality measurements with time and location stamps. If collected in long-term, these air quality measurements can provide valuable insights on personal air pollution exposure and validation of high-resolution air pollution models.

To fully unlock the potential of the big urban air quality data collected by mobile, low-cost sensors, it is essential to calibrate the measurements to obtain a *consistent* dataset. However, the quality of measurements from

Author's address: B. Maag and Z. Zhou and O. Saukh and L. Thiele, Computer Engineering and Networks Laboratory, ETH Zurich; E-mail: {balz.maag, zimu.zhou, saukh, thiele}@tik.ee.ethz.ch.

Corresponding Author: Zimu Zhou.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2017 Association for Computing Machinery.

2474-9567/2017/6-ART19 \$15.00

DOI: <http://doi.org/10.1145/3090084>

low-cost air quality sensors are affected by multiple factors and can vary dramatically across sensors and over time. (i) Low-cost air quality sensors suffer from low selectivity, *i.e.*, they are *cross-sensitive* to various substances in the air. (ii) Changing environmental conditions, such as temperature and humidity, impact the sensor output. (iii) Many air quality sensors operate at their sensitivity boundaries when measuring pollution in ambient air, which leads to high noise in sensed data. (iv) Sensor sensitivity degrades with time due to sensor ageing effects.

To compensate for the dependencies of (i) and (ii), prior studies [20, 29] propose to augment a measurement system with additional sensors to form a *sensor array*. A sensor array consists of co-located sensors that measure, in addition to the target air pollutants, a set of correlated pollutants and environmental parameters *e.g.*, temperature. Previous research [20] has demonstrated the feasibility of resolving cross-sensitive dependencies of low-cost air quality sensors by jointly calibrating a set of measurements collected by sensor arrays. In fact, an increasing number of customized [29, 31, 35] and commercial [28] air quality sensing nodes are integrated with multiple correlated sensors and report measurements of pollutants and environmental parameters simultaneously.

Adopting sensor arrays alone is insufficient to ensure consistent data quality of the measurements. Due to noise effects, see (iii), and the ageing effect, see (iv), measurements of sensor arrays need to be calibrated to the accurate, static governmental-run air monitoring stations, which are sparsely deployed in cities. However, sensors may have infrequent or no access to the static stations, making it unreliable or infeasible to calibrate all sensors to the static references. To improve the opportunities for calibration, some studies [15, 22, 24] propose to calibrate noisy sensors during deployment by exploiting *rendezvous* [25]. During a rendezvous two sensors are in each others' spatial and temporal vicinity and thus sense the same phenomena and should have similar outputs. Rendezvous-based calibration allows newly calibrated sensors to recursively calibrate other sensors, known as *multi-hop calibration*. By constructing a calibration path starting from a static station and performing calibration for each pair of sensors that frequently meet, calibration can reliably propagate through all sensor arrays within a large-scale deployment. Compared to calibration solely based on rendezvous between mobile nodes and reference stations multi-hop calibration is proven to frequently calibrate a significantly larger number of nodes [21, 24, 34]. Thus, multi-hop calibration is a powerful tool for maintaining high data quality in large scale deployments over long time periods.

However, it is non-trivial to apply multi-hop calibration on sensor arrays. Multi-hop calibration is only effective and guarantees data consistency if the calibration error at each hop does not accumulate. Otherwise a newly calibrated sensor array after multiple hops will still be too noisy as a reference to calibrate its successor sensor array. State-of-the-art multi-hop calibration schemes [15] based on *ordinary least-squares regression* (OLS) suffer from error-accumulation over multiple hops due to regression dilution [33]. The results in [24] propose to use *geometric mean regression* (GMR) to minimise multi-hop error-accumulation. The scheme is efficient with single sensor calibration but inapplicable to sensor array calibration. Although there are several generalizations of GMR to higher dimensions [10, 32], they all perform poorly when calibrating sensor arrays over multiple hops, as shown in Sec. 4. Existing sensor array calibration schemes leverage multiple least squares (MLS) [20] or artificial neural networks (ANN) [29], but are primarily designed for one-hop calibration, *i.e.*, calibration with highly accurate references.

In this paper, we propose *sensor array network calibration* (SCAN), a low-error multi-hop calibration scheme for sensor arrays. We formulate multi-hop sensor array calibration as a novel *constrained least-squares regression*, and come up with a closed-form solution that minimizes error accumulation over multiple hops for multi-dimensional calibrations. We theoretically prove that SCAN minimizes error accumulation with practical array composition and measurement settings. Experiments with a one-year measurement comprising 10 million samples from a 11-hop metal oxide gas sensor array chain show that SCAN yields up to 38% lower error than multiple least-squares (MLS) and geometric mean regression (GMR). Evaluations on a public air pollution dataset [19] of 56 million samples collected over two years demonstrate the benefits of multi-hop calibration over one-hop calibration.

Additionally we show that SCAN outperforms MLS by 60%, which is a significant improvement in enabling low-cost sensors for quantitative analysis such as validation of air pollution models and personal exposure studies.

The contributions of this paper are summarized as follows.

- To the best of our knowledge, we are the first to theoretically tackle the error accumulation problem of multi-hop sensor array calibration. While we primarily target at low-cost gas sensors, SCAN is a generic multi-hop multi-dimensional calibration scheme and is widely applicable to various heterogeneous, correlated sensors working in an ad hoc manner.
- We formulate multi-hop sensor array calibration as a novel constrained least-squares regression and propose a closed-form solution that minimizes error accumulation. We prove that SCAN works with practical settings such as non-zero mean measurements and non-squared sensor array composition.
- We evaluate SCAN on two real-world low-cost air pollution sensor datasets consisting of 10 and 56 million samples respectively collected over three years. Experiments show that SCAN achieves 16% to 60% lower error than state-of-the-art calibration techniques, which significantly improves the quality of measurements for qualitative studies such as validation of high-resolution air pollution models and personal air pollution exposure.

In the rest of this paper, we formulate SCAN in Sec. 3.2 and come up with a closed-form solution in Sec. 3.3. We conduct extensive simulations and real-world experiments to evaluate the performance of SCAN in Sec. 4. Finally we summarise related work in Sec. 5 and conclude in Sec. 6.

2 ASSUMPTIONS AND MODELS

This section presents a primer on sensor arrays and defines the models for multi-hop sensor array calibration.

2.1 Sensor-Array Calibration

We refer to a *trace* as a time-ordered sequence of M discrete and instantaneous measurements $y_i = (y_{ij}) \in \mathbb{R}^M$ taken by sensor s_i at times t_j and spatial locations l_j for $j \in \{1, 2, \dots, M\}$ within a time interval $[t_1, t_M]$. A trace measured by a low-cost gas sensor is usually inaccurate due to (i) low selectivity, *i.e.*, low-cost sensors are cross-sensitive to multiple substances in the air, and (ii) the response of low-cost gas sensors is affected by meteorological conditions. To compensate for these dependencies, low-cost sensors are usually augmented by a set of heterogeneous sensors to form a *sensor array* (see Fig. 1a). Specifically, instead of measuring the target phenomenon using one low-cost sensor, a sensor array $A = [s_1, s_2, \dots, s_K]$ of sensors s_1, \dots, s_K is employed to simultaneously measure K phenomena. Prominent examples are gas sensors based on electrochemical cells designed to measure nitrogen dioxide (NO_2) concentrations, whose sensitivities are generally affected by ambient temperature and interfering gases, such as ozone (O_3) [29]. The accuracy of an electrochemical cell sensor is usually notably improved by concurrently acquiring ambient temperature and O_3 measurements to compensate for the sensors dependency. Calibrating a sensor array consisting of these multiple low-cost sensors to a NO_2 reference improves the overall measurement accuracy by a factor of 2 compared to simple calibration using only the NO_2 low-cost sensor (see Fig. 1b). Similar cross-sensitivities are in fact a well-known limitation for the majority of state-of-the-art low-cost environmental sensors [5, 11, 16, 20, 29].

We now formulate the basic model for sensor array calibration. We distinguish between *reference arrays*, which consist of precise and selective sensors, *i.e.*, they accurately measure their target phenomenon, and mobile *low-cost sensor arrays*, which suffer from mutual dependencies and drifts stated above. Let $Y = (y_{ij}) \in \mathbb{R}^{K \times M}$ be a matrix describing traces of K possibly correlated phenomena, such as pollutant concentrations and ambient temperature, where y_{ij} describes the j -th measurement of phenomenon y taken by a reference sensor s_i at the given time instance t_j and location l_j . Denote $\bar{X} = (\bar{x}_{ij}) \in \mathbb{R}^{K \times M}$ as the uncalibrated traces of a mobile low-cost sensor array A taken at the same time instances and locations. We assume that the uncalibrated trace $\bar{x}_i = (\bar{x}_{ij}) \in \mathbb{R}^M$

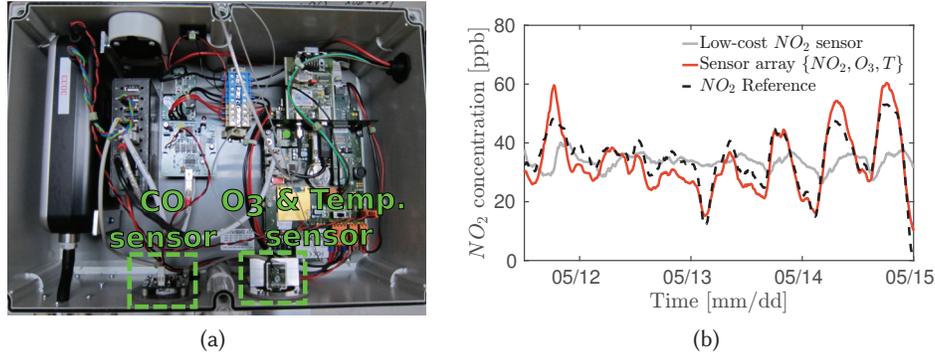


Fig. 1. (a) An air quality sensor box with multiple correlated sensors composing a sensor array. (b) Traces of a low-cost nitrogen dioxide (NO_2) electrochemical cell sensor before calibration and after calibration of temperature and ozone (O_3) dependency. Sensor array calibration improves the overall measurement accuracy by a factor of 2.

of a single cross-sensitive sensor is a linear combination of different effects describing the impact of different phenomena and sensor noise [6, 20]. Further, we assume that all sensors s_i for $i \in \{1, 2, \dots, K\}$ of an array A are sampled at the same time instance and at the same location, *i.e.*, a sensor array forms one physical unit. Then the uncalibrated traces of a low-cost sensor array can be defined as

$$\bar{X} = B^{-1} \cdot (Y + N_1), \quad (1)$$

where $N_1 \in \mathbb{R}^{K \times M}$ are K noise components (zero mean and uncorrelated to phenomena Y) and $B^{-1} \in \mathbb{R}^{K \times K}$ is an unknown matrix capturing the linear combination of phenomena and their magnitude for each individual sensor in the array, see Fig. 2.

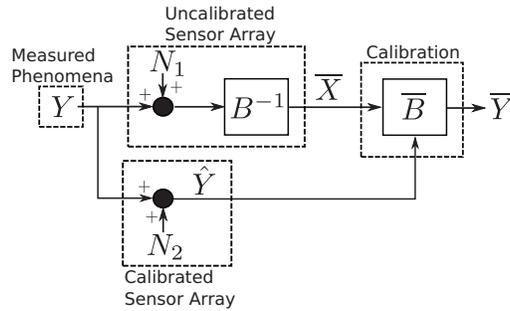


Fig. 2. Basic sensor array calibration model. The traces \hat{Y} of an already calibrated sensor array can be used as reference to calibrate the traces \bar{X} of an uncalibrated sensor array. If the calibrated sensor array is a reference array it measures the physical phenomena with perfect accuracy, *i.e.*, we assume the measurement noise $N_2 = \mathbf{0}_{K \times M}$.

The aim of sensor array calibration is to find a calibration matrix \bar{B} such that the calibrated measurements

$$\bar{Y} = \bar{B} \cdot \bar{X} = \bar{B} \cdot (B^{-1}(Y + N_1)) + C \quad (2)$$

minimize the calibration error, defined as a distance to the phenomena traces Y in some metric. Matrix $C \in \mathbb{R}^{K \times M}$ describes a constant additive offset. For simplicity of presentation we assume $C = \mathbf{0}_{K \times M}$ and all the rows in \bar{X} and

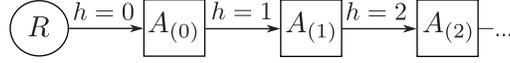


Fig. 3. In a first hop $h = 0$, an uncalibrated array $A_{(0)}$ is calibrated with traces from an accurate reference array R . In all consecutive hops $h > 0$ an uncalibrated array $A_{(q)}$ is calibrated with traces from an already calibrated array $A_{(q-1)}$.

Y have zero mean. We relax this assumption and present the calibration including a non-zero constant matrix C in Sec. 3.4. Further, we assume for a general low-cost sensor array that B is non-singular.

2.2 Multi-hop Sensor Array Calibration

Highly accurate reference arrays are often static and limited in a real-world deployment, and not all mobile sensor arrays have access to these reference arrays. To calibrate measurements of all sensor arrays, a practical strategy is to use calibrated sensor arrays as *virtual references* to calibrate another mobile sensor array. The rationale is that when two mobile sensor arrays are close in space and time, known as *rendezvous*, they are exposed to the same physical processes, and thus provide an opportunity for calibration. Such a rendezvous based calibration scheme can propagate along multiple hops and ultimately can cover all the deployed mobile sensor arrays.

One-hop calibration. Initially, we calibrate a mobile low-cost sensor array against a high quality reference array, which measures physical phenomena Y with perfect accuracy. Hence, we calibrate a sensor array $A_{(h=0)}$ using its traces $\bar{X}_{(0)} = B_{(0)}^{-1} \cdot (Y + N_1)$ by estimating the optimal $\bar{B}_{(0)}$. We refer to the calibration between reference and low-cost array as *one-hop calibration* with $h = 0$.

Multi-hop calibration. In the next hop, $h = 1$, given rendezvous between an already calibrated array $A_{(0)}$ and an uncalibrated array $A_{(1)}$, we take $A_{(0)}$ as a *virtual reference*. We use its measurement $\bar{X}_{(0)}$ at the current rendezvous to determine a set of calibrated measurements $\bar{Y}_{(0)} = \bar{B}_{(0)} \cdot \bar{X}_{(0)}$ with $\bar{B}_{(0)}$ from the one-hop calibration. The measured values $\bar{X}_{(1)}$ of array $A_{(1)}$ at the current rendezvous can now be calibrated with the calibrated measurements $\hat{Y}_{(1)} = \bar{Y}_{(0)}$ from array $A_{(0)}$, see Fig. 2. As illustrated in Fig. 3, this concept can be applied to consecutive hops $h > 1$ resulting in a *rendezvous path*. We refer to calibration of sensor arrays along rendezvous paths with $\max(h) > 0$ as *multi-hop calibration*. Specifically, at hop $h = q$ with $(q > 0)$, we calibrate the sensor array traces as $\bar{Y}_{(q)} = \bar{B}_{(q)} \cdot \bar{X}_{(q)}$, where $\bar{B}_{(q)}$ is obtained by the raw traces $\bar{X}_{(q)}$ and virtual reference traces $\hat{Y}_{(q)} = \bar{Y}_{(q-1)}$ measured by a calibrated sensor array $A_{(q-1)}$.

Thus, given uncalibrated measurements $\bar{X}_{(q)}$ from array $A_{(q)}$ and calibrated measurements $\hat{Y}_{(q)} = \bar{Y}_{(q-1)}$ from array $A_{(q-1)}$ we intend to calculate calibration parameters $\bar{B}_{(q)}$ such that

$$\left\| \bar{Y}_{(q-1)} - \bar{B}_{(q)} \cdot \bar{X}_{(q)} \right\| \quad (3)$$

with some norm $\| \cdot \|$ is minimized.

Discussions. (i) Our multi-hop sensor array calibration scheme is performed along a *static, pre-defined* rendezvous path. For example, the path can be determined by first selecting the mobile sensor array that most frequently meets the static reference array as the first hop, i.e. the mobile sensor array with the most rendezvous with a reference station, and recursively selecting the most frequently met array as the next hop. It is possible that a sensor array A meets a sensor array B frequently and the reference array R infrequently. In this case, we omit all the rendezvous between A and R . Determining the optimal rendezvous paths is out of the scope of this paper. We refer interested readers to [24] for calibration parent selection strategies and [13] on reference placement schemes to ensure network-wide calibrability. (ii) Our calibration scheme mainly serves as an important data cleaning technique to guarantee data accuracy and consistency in big air quality data collected by mobile sensor

arrays. We assume the dataset has been pre-processed to filter incomplete samples due to *e.g.*, GPS failures as in [25].

3 MULTI-HOP SENSOR ARRAY CALIBRATION

In this section we present in detail our novel multi-hop calibration method. In Sec. 3.1, we show that state-of-the-art *multiple least-squares* (MLS) is not suitable for multi-hop sensor array calibration due to its error accumulation over multiple hops. We tackle this problem by modifying MLS in order to make it suitable for multi-hop calibration. We formulate our *sensor array network calibration* (SCAN) solution as a new constrained least-squares regression in Sec. 3.2 and present a closed-form solution in Sec. 3.3.

3.1 Limitations of Multiple Least-Squares

Multiple least-squares (MLS) minimizes the squared calibration error in each individual hop $h = q$, that is the deviation of virtual references $\hat{Y}_{(q)} = \bar{Y}_{(q-1)}$ taken by array $A_{(q-1)}$ and calibrated signals $\bar{Y}_{(q)} = \bar{B}_{(q)} \cdot \bar{X}_{(q)}$ taken by array $A_{(q)}$. For simplicity of presentation we set $\bar{X} = \bar{X}_{(q)}$, $\bar{B} = \bar{B}_{(q)}$ and $\hat{Y} = \hat{Y}_{(q)} = \bar{Y}_{(q-1)}$ in all the following sections. Hence, the minimization problem looks as follows [4]

$$\min_{\bar{B}} \text{tr} \left((\hat{Y} - \bar{B}\bar{X}) (\hat{Y} - \bar{B}\bar{X})^T \right) \quad (4)$$

where *tr* is the *trace* operator [4]. The solution to (4) is given by

$$\bar{B} = \hat{Y}\bar{X}^T \left(\bar{X}\bar{X}^T \right)^{-1}. \quad (5)$$

Since the traces of a virtual reference array are usually imperfect, *i.e.*,

$$\hat{Y} = Y + N_2, \quad (6)$$

where N_2 are K noise components related to the sensor readings of the virtual reference sensor, *e.g.*, due to calibration error, a major challenge in multi-hop calibration is *error accumulation* over the rendezvous path [24]. We apply MLS to multi-hop sensor array calibration at hop $h = q$ with $q > 0$, where N_1 denotes the noise when reading the sensors of the array $A_{(q)}$ that needs to be calibrated, *i.e.*, $\bar{X} = B^{-1}(Y + N_1)$. The calibration matrix \bar{B} calculated by MLS becomes

$$\begin{aligned} \bar{B} &= (Y + N_2)(Y + N_1)^T B^{-T} \cdot B^T \left[(Y + N_1)(Y + N_1)^T \right]^{-1} B \\ &= \left[YY^T + YN_1^T + N_2Y^T + N_2N_1^T \right] \cdot \left[YY^T + YN_1^T + N_1Y^T + N_1N_1^T \right]^{-1} B. \end{aligned} \quad (7)$$

Because we assume independent noise and no correlation between noise and phenomena, it holds that $YN_1^T = N_1Y^T = N_2Y^T = N_2N_1^T = \mathbf{0}_{K \times K}$ and finally

$$\bar{B} = \left[YY^T \right] \left[YY^T + N_1N_1^T \right]^{-1} B. \quad (8)$$

From (8) we see that MLS underestimates \bar{B} under the presence of noise N_1 with $N_1N_1^T \neq \mathbf{0}_{K \times K}$, better known as *regression dilution* or *bias towards zero* [12]. This effect grows with increasing sensor noise variance $N_1N_1^T$. Thus, calibration parameters \bar{B} estimated by MLS depend on sensor noise. The bias towards zero of the parameters of a calibrated array affects the uncalibrated array of the consecutive hop. As a result calibration error is accumulated over the whole rendezvous path, as we will show in the simulations and real-world experiments in Sec. 4.

3.2 Sensor Array Network Calibration

To reduce error accumulation in multi-hop sensor array calibration, we propose a solution formulated as a constrained least-squares regression referred to as *sensor array network calibration* (SCAN). As far as we are aware of, this is the first work that formulates the multi-hop sensor array calibration problem into a novel constrained least-squares regression formulation with closed-form solutions.

We first define the symmetric matrices $\hat{Y}\hat{Y}^T = YY^T + N_2N_2^T$ and $\bar{X}\bar{X}^T = B^{-1}(YY^T + N_1N_1^T)B^{-T}$. For the following results, suppose $\bar{X}\bar{X}^T$ and $\hat{Y}\hat{Y}^T$ are non-singular. The SCAN regression problem can be formulated as follows:

$$\underset{\bar{B}}{\text{minimize}} \quad \text{tr} \left((\hat{Y} - \bar{B}\bar{X})(\hat{Y} - \bar{B}\bar{X})^T \right) \quad (9)$$

$$\text{subject to} \quad \bar{B}\bar{X}\bar{X}^T\bar{B}^T = \hat{Y}\hat{Y}^T \quad (10)$$

Thus, SCAN minimizes the least-squares error (9) with the constraint on the regression parameters stated in (10). As we will show later on, (10) reduces the bias towards zero and eliminates it under certain realistic assumptions. To the best of our knowledge, we are the first to formulate this constraint regression problem.

3.3 Closed-form Solution to SCAN

Degrees of freedom. Let $\bar{X} = U_X D_X V_X^T$ be the singular value composition of the matrix $\bar{X} \in \mathbb{R}^{K \times M}$, where $U_X \in \mathbb{R}^{K \times K}$ and $V_X \in \mathbb{R}^{M \times K}$ are orthogonal matrices and $D_X \in \mathbb{R}^{K \times K}$ is a diagonal matrix holding the singular values of \bar{X} on the diagonal. Further let $\hat{Y} = U_Y D_Y V_Y^T$ be the singular value decomposition of $\hat{Y} \in \mathbb{R}^{K \times M}$. The constraint in (10) can be formulated as the following equivalent equation

$$\begin{aligned} \bar{B}U_X D_X V_X^T V_X D_X^T U_X^T \bar{B}^T &= U_Y D_Y V_Y^T V_Y D_Y^T U_Y^T \\ D_Y^{-1} U_Y^T \bar{B} U_X D_X D_X U_X^T \bar{B}^T U_Y D_Y^{-1} &= I \\ \Leftrightarrow F = D_Y^{-1} U_Y^T \bar{B} U_X D_X &\wedge FF^T = I, \end{aligned} \quad (11)$$

where $F \in \mathbb{R}^{K \times K}$ is an orthogonal matrix. We can re-formulate the constraint (10) to

$$\bar{B} = U_Y D_Y F D_X^{-1} U_X^T, \quad (12)$$

where the orthogonal matrix F defines the *degrees of freedom* of the calibration matrix \bar{B} .

Minimizing least-squares error. We now present a method to determine F that minimizes the least-squares error (9). Recall the minimization problem (9)

$$\begin{aligned} \min_{\bar{B}} \text{tr} \left((\hat{Y} - \bar{B}\bar{X})(\hat{Y} - \bar{B}\bar{X})^T \right) &= \min_{\bar{B}} \text{tr}(\hat{Y}\hat{Y}^T + \bar{B}\bar{X}\bar{X}^T\bar{B}^T - \bar{B}\bar{X}\hat{Y}^T - (\bar{B}\bar{X}\hat{Y}^T)^T) \\ &= \max_{\bar{B}} \text{tr}(\bar{B}\bar{X}\hat{Y}^T). \end{aligned} \quad (13)$$

Thus, we simplify the original minimization problem to a maximization problem stated in (13). Recall constraint (12) on \bar{B} , it follows

$$\begin{aligned} \max_{\bar{B}} \text{tr}(\bar{B}\bar{X}\hat{Y}^T) &= \max_F \text{tr}(U_Y D_Y F D_X^{-1} U_X^T U_X D_X V_X^T V_Y D_Y U_Y^T) \\ &= \max_F \text{tr}(F V_X^T V_Y D_Y^2). \end{aligned} \quad (14)$$

We define $Q = V_X^T V_Y D_Y^2 \in \mathbb{R}^{K \times K}$ and its singular value decomposition $Q = U_Q D_Q V_Q^T$ and followingly

$$\max_F \operatorname{tr} \left(F V_X^T V_Y D_Y^2 \right) = \max_F (FQ) = \max_F \left(F U_Q D_Q V_Q^T \right). \quad (15)$$

The solution for F that maximizes (15) and (14) respectively is given by

$$F = V_Q U_Q^T. \quad (16)$$

From (12) and (16) it follows that

$$\bar{B} = U_Y D_Y V_Q U_Q^T D_X^{-1} U_X^T \quad (17)$$

minimizes the least-squares error (9) under the constraint (10).

PROOF. The proof for the solution to (14) is similar to the *orthogonal Procrustes problem* [26] as well as a consequence of the Courant-Fischer theorem [4]. It holds

$$\operatorname{tr}(FQ) = \operatorname{tr} \left(F U_Q D_Q V_Q^T \right) = \operatorname{tr} \left(\left[F U_Q D_Q^{\frac{1}{2}} \right] \left[V_Q D_Q^{\frac{1}{2}} \right]^T \right) = F U_Q D_Q^{\frac{1}{2}}, V_Q D_Q^{\frac{1}{2}}. \quad (18)$$

According to the *Cauchy-Schwarz* theorem [4] it follows

$$F U_Q D_Q^{\frac{1}{2}}, V_Q D_Q^{\frac{1}{2}} \leq \left\| F U_Q D_Q^{\frac{1}{2}} \right\|_2 \left\| V_Q D_Q^{\frac{1}{2}} \right\|_2,$$

where the equality holds if $F = V_Q U_Q^T$ and thus maximizes $\operatorname{tr}(FQ)$, *i.e.*,

$$\operatorname{tr}(FQ) = F U_Q D_Q^{\frac{1}{2}}, V_Q D_Q^{\frac{1}{2}} = V_Q D_Q^{\frac{1}{2}}, V_Q D_Q^{\frac{1}{2}} = \left\| D_Q^{\frac{1}{2}} \right\|_2 \left\| D_Q^{\frac{1}{2}} \right\|_2 = \operatorname{tr}(D_Q). \quad (19)$$

Thus, $F = V_Q U_Q^T$ is an orthonormal transformation that maximizes $\operatorname{tr}(\bar{B} \bar{X} \hat{Y}^T)$ and consequently minimizes the least-squares error (9) under constraint (10). \square

3.4 Discussions

Existence. There always exists a solution to the SCAN regression. The closed-form expression of the regression parameters \bar{B} stated in (17) involves the singular value decomposition of $\bar{X} \in \mathbb{R}^{K \times M}$, $\hat{Y} \in \mathbb{R}^{K \times M}$ and $Q \in \mathbb{R}^{K \times K}$. Because there always exists a singular value decomposition for any general real matrix [4], there also exists a regression parameter matrix $\bar{B} \in \mathbb{R}^{K \times K}$ according to (17).

Relationship to GMR. For the two-dimensional calibration problem where we calibrate a single sensor on a single reference sensor, *i.e.*, $K = 1$, the constraint (10) of SCAN reduces to $\bar{B} = \pm \frac{\overline{Y Y^T}}{\overline{X X^T}}$. This is in fact the solution for the regression slope according to GMR [33].

No bias towards zero property. We now show that the calibration matrix obtained from our SCAN regression is free from regression dilution even in presence of noise in measurements. Let \bar{B} a solution that minimizes the least-squares error (9) and satisfies constraint (10), then it follows by (1) and (6)

$$\bar{B} \bar{B}^{-1} \left(Y Y^T + N_1 N_1^T \right) \bar{B}^{-T} \bar{B}^T = Y Y^T + N_2 N_2^T. \quad (20)$$

In the case where the variance of the noise of the two sensor arrays are equal and N_1 and N_2 are uncorrelated, we assume

$$N_1 N_1^T = N_2 N_2^T = N N^T = \lambda \cdot I, \quad (21)$$

where $\frac{\lambda}{M} \in \mathbb{R}$ describes the variance of noise components in N and $I \in \mathbb{R}^{K \times K}$ is the unit matrix. Further, let $YY^T + NN^T = UD^2U^T$ be the eigenvalue decomposition of the symmetric matrix $YY^T + NN^T = YY^T + \lambda \cdot I$ and accordingly it holds $YY^T = U(D^2 - \lambda \cdot I)U^T$. Hence, (20) simplifies to

$$\begin{aligned} \bar{B}\bar{B}^{-1} (YY^T + NN^T) B^{-T}\bar{B}^T &= YY^T + NN^T \\ \bar{B}\bar{B}^{-1} (UD^2U^T) B^{-T}\bar{B}^T &= UD^2U^T \\ (D^{-1}U^T\bar{B}\bar{B}^{-1}UD)(D^{-1}U^T\bar{B}\bar{B}^{-1}UD)^T &= I \\ \Leftrightarrow G = D^{-1}U^T\bar{B}\bar{B}^{-1}UD \wedge GG^T &= I, \end{aligned} \quad (22)$$

where G is an orthogonal matrix. Recall the maximization problem stated in (13), it follows

$$\max_{\bar{B}} \text{tr}(\bar{B}\bar{X}\hat{Y}^T) = \max_G \text{tr}(G(D^2 - \lambda \cdot I)), \quad (23)$$

where due to the diagonal form of $(D^2 - \lambda \cdot I)$ the maximum is achieved if $G = I$ [26]. In this case the calculated regression parameters $\bar{B} = B$. That means, even if $NN^T \neq \mathbf{0}_{K \times K}$, SCAN estimates the true underlying calibration parameters B . Hence, \bar{B} calculated by SCAN are not affected by regression dilution, whereas MLS underestimates the calibration parameters due to its bias towards zero.

In a real-world deployment we cannot assume that (21) ideally holds. In this case \bar{B} does depend on a relation between N_1N_1 and N_2N_2 as stated in (20). In Sec. 4 we discuss this assumption in detail and experimentally show that SCAN reduces error accumulation over multiple hops and outperforms various state-of-the-art calibration techniques.

Relaxing the zero-mean variable assumption. Both \hat{Y} and \bar{X} contain in general rows with a non-zero mean. In this case, the calibration also needs to compensate for an offset term C , see (2). This is done similar to MLS [4] and basic GMR [33]. Let $\tilde{Y} = \hat{Y} + \text{mean}(\hat{Y})$ and $\tilde{X} = \bar{X} + \text{mean}(\bar{X})$ be non-zero mean representations of \hat{Y} and \bar{X} , where $\text{mean}(\hat{Y}) \in \mathbb{R}^{K \times M}$ and $\text{mean}(\bar{X}) \in \mathbb{R}^{K \times M}$ are mean values of each row in \hat{Y} and \bar{X} respectively. It follows

$$\begin{aligned} \tilde{Y} &= \bar{B}\tilde{X} + C \\ \hat{Y} + \text{mean}(\hat{Y}) &= \bar{B}(\bar{X} + \text{mean}(\bar{X})) + C. \end{aligned} \quad (24)$$

Let $C = \text{mean}(\hat{Y}) - \bar{B} \cdot \text{mean}(\bar{X})$, then (24) equals to the calibrated measurements by SCAN with zero-mean variables. Following the calibrated measurements are calculated as $\bar{Y} = \bar{B}\bar{X} + \text{mean}(\hat{Y}) - \bar{B} \cdot \text{mean}(\bar{X})$. Accordingly we can reformulate the constraint (10) using non zero-mean variables and the calculated offset to

$$\begin{aligned} (\hat{Y} + \text{mean}(\hat{Y})) (\hat{Y} + \text{mean}(\hat{Y}))^T &= (\bar{B}(\bar{X} + \text{mean}(\bar{X})) + C) (\bar{B}(\bar{X} + \text{mean}(\bar{X})) + C)^T \\ \Leftrightarrow \hat{Y}\hat{Y}^T + \text{mean}(\hat{Y}) \cdot \text{mean}(\hat{Y})^T &= \bar{B}\bar{X}\bar{X}^T\bar{B}^T + \text{mean}(\hat{Y}) \cdot \text{mean}(\hat{Y})^T \end{aligned} \quad (25)$$

which is equal to constraint (10) of our SCAN regression with zero-mean variables. Thus, the offset calculation does not affect the *no bias towards zero* property, *i.e.*, error accumulation.

Relaxing the squared calibration matrix assumption. So far we assumed that a sensor array consisting of K different sensors is always calibrated to an array with K sensors as well. In some situations this setup is not suitable. For instance, although a sensor of an array is cross-sensitive to a certain phenomenon it might not be possible to calibrate the array to said phenomenon due to the lack of highly accurate reference measurements. Therefore, a calibration of K sensors to L references with $1 \leq L < K$ is in certain cases required.

Let $\tilde{Y} = S(Y + N_2)$ be L phenomena measured by a calibrated sensor array. $S \in \mathbb{R}^{L \times K}$ is a matrix with $S_{ij} = 1$ if phenomenon $j \in \{1, \dots, K\}$ equals the phenomenon $i \in \{1, \dots, L\}$ measured by the calibrated array and all other

elements equal 0. Further, it holds $SS^T = I$. Let $\tilde{B} \in \mathbb{R}^{L \times K}$ be the calibration matrix of an uncalibrated sensor array to the L phenomena. Accordingly the constraint (10) on \tilde{B} looks as follows

$$\tilde{B}B^{-1}(YY^T + N_1N_1^T)B^{-1}\tilde{B}^T = SY Y^T S^T + SN_2N_2^T S^T.$$

If we again assume equal noise variances, *i.e.*, $N_1N_1^T = N_2N_2^T = NN^T = \lambda \cdot I$, then we find the true calibration parameters $\tilde{B} = SB$. This can be proofed similar to (20) - (23) and, thus, we do not present a detailed proof. Hence, the *no bias towards zero* property is preserved.

In the case of non-equal noise variances, the constraints of the calibration with L and K phenomena differ. Unlike MLS, SCAN calculates different calibration parameters depending on the number of phenomena in \hat{Y} , which is elaborated in more detail in Sec. 4.

4 EXPERIMENTAL EVALUATION

This section presents the evaluations of SCAN on both simulated and real-world datasets.

4.1 Simulation

We first show through simulations the advantages of SCAN over multiple least-squares (MLS, see Sec. 3.1), which motivates our multi-hop sensor array calibration scheme, as well as performance comparison with the state-of-the-art calibration schemes. We then investigate the robustness of SCAN with various impacting factors.

4.1.1 Setup. We artificially generate 20 different reference sensor arrays $Y_{(h)} \in \mathbb{R}^{K \times M} \sim \mathcal{N}(\ln Y_{(h)}; 0, \gamma)$ representing M ground-truth measurements of K different phenomena with standard deviation $\gamma \in [0.2, 0.45]$. The log-normal distribution models typical air pollution and meteorological measurements in an urban area [17, 30]. These reference arrays serve as basis for a rendezvous path consisting of 20 noisy sensor arrays. In all experiments the number of physical phenomena K is set to 4. The measurements of each array at hop h with $h \in \{0, 1, \dots, 19\}$ are defined by $\bar{X}_{(h)} = B_{(h)}^{-1} \cdot (Y_{(h)} + N_1)$, where $B_{(h)}$ is a general matrix with randomly chosen entries $B_{ij} \in [0.2, 2]$ and $N_1 \in \mathbb{R}^{K \times M} \sim \mathcal{N}(0, \sigma^2)$ describes sensor noise. These sensor arrays are in line with our findings in Sec. 4.2 and Sec. 4.3, where we use real-world sensors that exhibit similar cross-sensitivity intensities, as well as with existing literature [16, 20, 24, 29]. In the first hop $h = 0$ the low cost sensor array is calibrated with the reference array. In all consecutive hops the measurement of the previously calibrated array $\hat{Y}_{(h)} = \bar{B}_{(h-1)} \cdot (B_{(h-1)}^{-1}(Y_{(h)} + N_1))$ are used as reference array. We used 500 samples for training the calibration parameters and 500 samples for evaluation. The 500 samples for evaluating the performance of an array are also used to train the calibration parameters of the consecutive sensor array. Each experiment is run 100 times with re-sampled reference measurements and ground-truth calibration parameters $B_{(h)}$.

Evaluation metrics. We calculate the difference between the calculated calibration parameters \bar{B} and the ground-truth B using the Froebinus norm, defined as

$$\|B - \bar{B}\|_F = \left(\text{tr} \left((B - \bar{B})(B - \bar{B})^T \right) \right)^{\frac{1}{2}}. \quad (26)$$

Additionally we investigate the root-mean-square value of the calibration error (RMSE) of the calibrated sensor array to one phenomena $y_4 \in Y$, *i.e.*,

$$RMSE = \left(\frac{1}{M} \sum_{j=1}^M (y_{4j} - \bar{y}_{4j})^2 \right)^{\frac{1}{2}}. \quad (27)$$

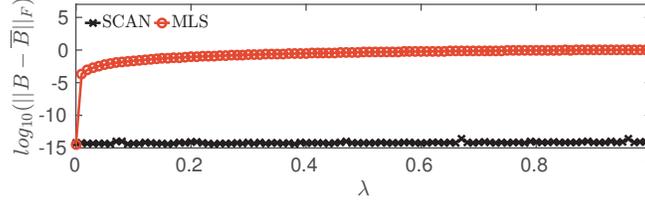


Fig. 4. Difference between estimated \bar{B} and ground-truth calibration parameters B for different sensor noise variance λ .

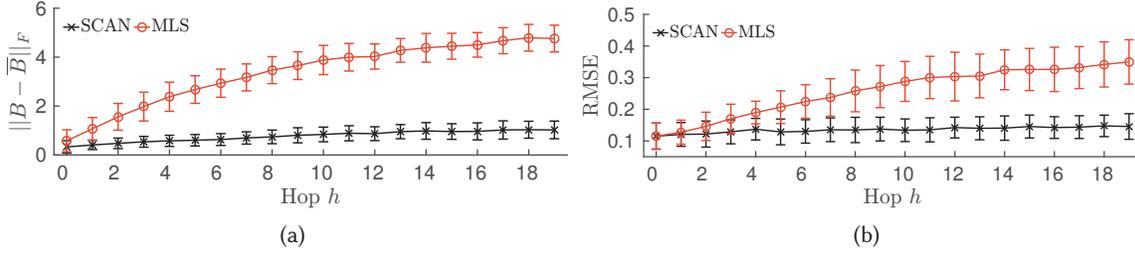


Fig. 5. SCAN calculates calibration parameters \bar{B} that deviate over all hops to a less extent from the groundtruth B compared to MLS, depicted in Fig. 5a. Hence, SCAN calibrates in Fig. 5b all sensor arrays of a 20-hop calibration path with clearly reduced error accumulation.

4.1.2 Overall Performance. In this set of experiments, we first verify the correctness of our theoretical findings by comparing with MLS, and then compare SCAN with various state-of-the-art multi-hop calibration schemes.

True calibration matrix estimation. As theoretically shown in Sec. 3.4, SCAN is able to calculate the true underlying calibration parameters $\bar{B} = B$ if the virtual reference and uncalibrated array suffer from noise with equal variance, *i.e.*, $N_1 N_1^T = N_2 N_2^T = \lambda \cdot I$. We first investigate this finding and simulate the assumption for an arbitrary hop in the rendezvous path by enforcing $\hat{Y}_{(h)} \cdot \hat{Y}_{(h)}^T = Y_{(h)} \cdot Y_{(h)}^T + \lambda \cdot I$ and $\bar{X}_{(h)} \cdot \bar{X}_{(h)}^T = B_{(h)}^{-1} (Y_{(h)} \cdot Y_{(h)}^T + \lambda \cdot I) B_{(h)}^{-T}$. Fig. 4 reflects our theoretical findings in Sec. 3.4. We observe that SCAN is able to calculate the true B up to machine precision independent of λ . Further, we observe the estimation by MLS is dependent on $\lambda \neq 0$.

Error accumulation.

Fig. 5 shows the results of the calibration at each hop when the standard deviation σ of the noise for all sensor arrays is randomly chosen from $[0.05, 0.2]$. Due to the small number of samples, the noise components are in general correlated, *i.e.*, $N_1 N_1^T$ has a general form with dominating diagonal elements, which relates to real-world sensor arrays as we will show later in Sec. 4.2. In Fig. 5a we observe that our SCAN approach clearly outperforms MLS with respect how accurate the underlying $B_{(h)}$ is estimated in each hop. This finding is reflected in the calibration error over multiple hops in Fig. 5b, where SCAN achieves a reduced error accumulation over 20 hops compared to MLS. In fact, already after 5 hops MLS shows a 38% higher average error than SCAN. Although SCAN outperforms MLS, we also observe an error accumulation for SCAN over 20 hops. Relative to the first hop the error increased by less than 26% at the last hop. The reasons are the different sensor noise variances σ^2 of each individual sensor, *i.e.*, assumption (21) is not satisfied, and the different measurement ranges the sensor samples of each array lie in. The impact of these two reasons is elaborated in detail in the following section.

Comparison with other techniques. Table 1 summarizes the calibration error for seven different techniques,

Method	$h = 0$	$h = 5$	$h = 19$
SCAN	0.12 ± 0.04	0.12 ± 0.04	0.14 ± 0.04
MLS	0.11 ± 0.04	0.21 ± 0.05	0.34 ± 0.07
GMR	0.4 ± 0.06	0.54 ± 0.18	0.54 ± 0.15
TLS	0.14 ± 0.07	0.62 ± 0.3	4.2 ± 2.7
ANN	0.19 ± 0.3	> 100	> 100
Draper	0.13 ± 0.05	0.24 ± 0.05	0.34 ± 0.07
Tofallis	0.14 ± 0.06	0.25 ± 0.06	0.35 ± 0.07

Table 1. Calibration error of different techniques for one- and multi-hop calibration.

namely our sensor array network calibration regression (SCAN), 2-dimensional geometric mean regression (GMR [24]), multiple least-squares (MLS [20]), total least-squares (TLS [14]), artificial neural networks (ANN [29]) and two different generalizations of the basic geometric mean regression to multiple dimension by Draper *et al.* [10] and Tofallis [32]. The comparison in Table 1 is based on 100 runs with a 20-hop rendezvous path and $\sigma \in [0.05, 0.2]$.

As previously shown SCAN clearly outperforms MLS over multiple hops due to the reduction of error accumulation. For one-hop calibration MLS is more accurate. The 2-dimensional GMR calibration is not able to compensate for cross-sensitivities. Thus, the calibration error on all hops is severely larger compared to the other multi-dimensional regression techniques. TLS is a multi-dimensional error-in-variables regression, *i.e.*, it assumes that both X and Y are affected by errors, which relates to our multi-hop calibration problem. However as experimentally shown, TLS is also suffering from an error accumulation that is even more severe compared to MLS over multiple hops. ANNs have been widely used for sensor array calibration due to their ability of modelling complex and possibly non-linear relationships between sensor and reference measurements. In our setup we use a simple network with 1 hidden layer and 10 neurons [9]. We show that already after 5 hops the accuracy of the calibration exceeds all other techniques. This can be traced back to the fact that neural networks tend to overfit and are usually a prominent choice for compensating for non-linear cross-sensitivities. Finally, we compare SCAN to two different generalizations of the geometric mean regression to multiple dimensions introduced by Draper *et al.* [10] and Tofallis [32]. From our results we observe that both generalizations suffer from considerable error accumulation over multiple hops and, thus, are not suited for multi-hop sensor array calibration. Overall, we conclude that our SCAN approach outperforms all other tested techniques for multi-hop calibration and MLS is the best choice for one-hop calibration.

4.1.3 Robustness. In this set of experiments, we investigate the impact of various factors that may violate the assumptions of our theoretical analysis to assess the robustness of SCAN.

Impact of noise variance. The amount of sensor noise has an important effect on the multi-hop calibration performance. Fig. 6a shows the comparison between different intervals for σ , *i.e.*, $\sigma \in [0.05, 0.2]$ (bold lines) and $\sigma \in [0.2, 0.3]$ (dashed lines). For both SCAN and MLS the average error over all hops is increased with increasing variance in sensor noise. As shown before, SCAN reduces the error accumulation compared to MLS. In fact, the accumulated error over 7 hops is below 13% with SCAN as shown in Fig. 6b. When calibrating with MLS the amount of noise affects the error accumulation per hop. After $h = 5$ the accumulated error per hop decreases and the calibration error stabilizes with MLS under high sensor noise. This is because the calculated calibration parameters \bar{B} already strongly deviate from the ground-truth due to the bias towards zero of \bar{B} with MLS, as theoretically proved in (8). With lower noise MLS achieves a lower average error compared to SCAN and MLS under higher noise conditions, but accumulated over 95% error over 7 hops.

Impact of dependent noise. So far we affected all sensors with noise whose variance lied in a narrow interval.

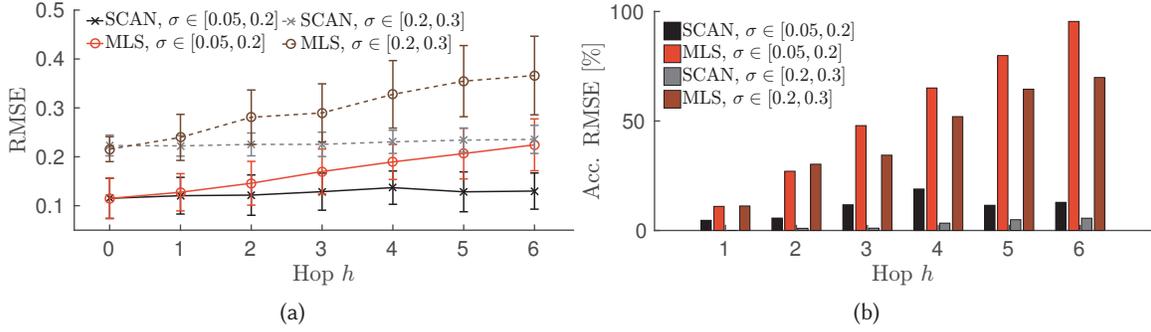


Fig. 6. Calibration error over seven hops (Fig. 6a) and total error accumulation (Fig. 6b) after each hop relative to $h = 0$ when the sensor signals are affected by low ($\sigma \in [0.05, 0.2]$) and high noise ($\sigma \in [0.2, 0.3]$).

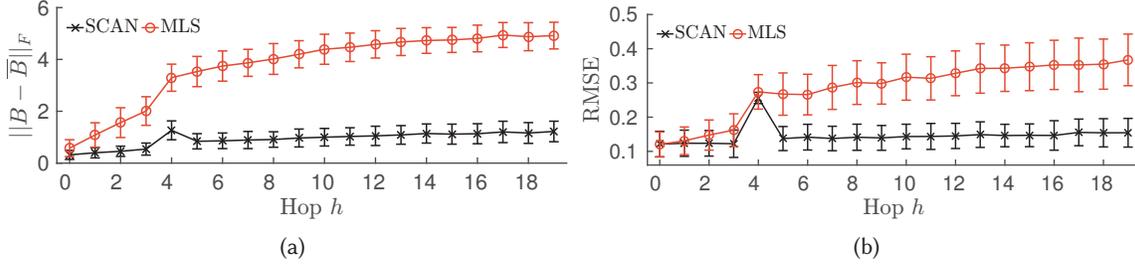


Fig. 7. Impact on calibration parameter estimation (Fig. 7a) and error (Fig. 7b) of increased sensor noise on a single sensor array at hop $h = 4$. SCAN is able to accurately calculate calibration parameters independent of the sensor noise of its parent.

In a real-world deployment this assumption may not hold. Noise of individual arrays might be significantly higher compared to others, especially in networks with heterogeneous nodes. In Fig. 7 the standard deviation of the noise of all sensors in the uncalibrated array at hop $h = 4$ is set to $\sigma = 0.3$ and the one of its neighbours $h = 3$ and $h = 5$ to $\sigma = 0.12$. Despite the increased calibration error of the array at hop $h = 4$, arrays at hops $h = [5, 19]$ are not notably affected when calibrating with SCAN. In contrast the increased noise at $h = 4$ has a strong impact on the parameter estimation by MLS and, hence, sensor arrays at hops $h = [5, 19]$ suffer from a high calibration error. This result shows that the calibration parameter estimation of a child sensor array by SCAN depends on the child sensor array noise but barely on the noise of its parent.

Impact of variable measurement range. In the previous experiments we always assumed that the measurements of all sensor arrays along the rendezvous path include samples over the whole range of the underlying phenomena, *i.e.*, the interval defined by the smallest and highest absolute value of the phenomena measurements. For instance, the ambient temperature in a deployment ranges from 5°C to 30°C , then so far we assumed that all arrays along a rendezvous path contain temperature samples within the whole range. This assumption is difficult to enforce in a real-world deployment. Physical phenomena typically show large variance depending on time and location. Therefore, capturing samples over the whole range of all target phenomena is a difficult task without enforcing rendezvous between sensor arrays.

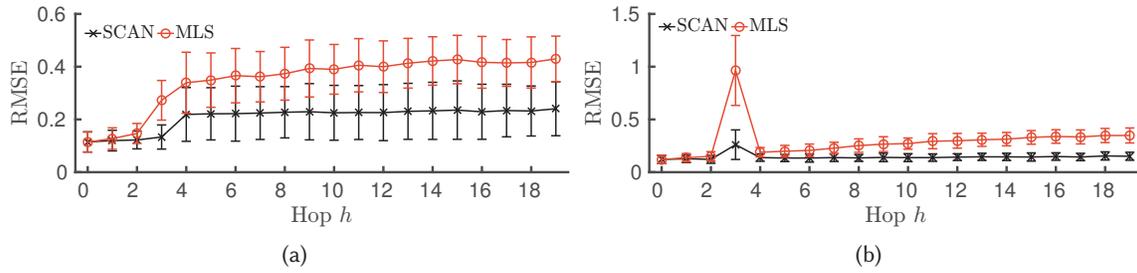


Fig. 8. A lower measurement range between arrays at $h = 4$ introduces an increased calibration error in Fig. 8a for both SCAN and MLS in all hops $h > 3$. In Fig. 8b an increased measurement range only affects sensor array at hop $h = 3$ without effect on the subsequent sensor arrays.

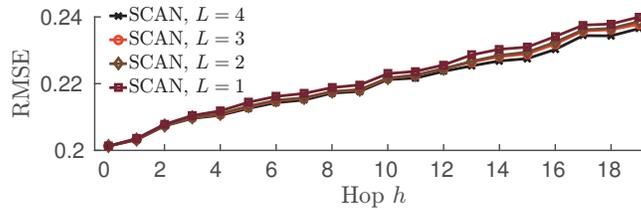


Fig. 9. Calibration error of our SCAN method with different numbers L of reference signals. The effects on the error accumulation are negligible.

In Fig. 8 we show the effect of a rendezvous between array at hop $h = 3$ and $h = 4$ with a different measurement range compared to all other sensor arrays in the path. We decrease the range by a factor of 3 in Fig. 8a and observe that both SCAN and MLS are affected. The lower measurement range introduces an increase of the calibration error for sensor array at $h = 4$ and all its subsequent arrays. The reason for this effect is that the calibration parameter estimation for all sensor arrays with hop $h > 4$ is only valid for the smaller range of array at $h = 4$ and, thus, the array can only be used to calibrate subsequent arrays with equal measurement range. In Fig. 8b we increase the measurement range of the rendezvous between arrays at $h = 3$ and $h = 4$ by a factor of 3. This only affects the sensor array at $h = 3$ but not the ones at hops $h = [4, 19]$. The sensor array at $h = 3$ is not able to estimate parameters for its whole range, only for the smaller range of its parent array. This does not affect sensor arrays at hops $h = [4, 19]$ because their measurement range is equal to the one at hop $h = 2$.

We conclude that it is important to use calibration data including a large measurement range of equal size for all sensor arrays in a rendezvous path. In a real-world deployment the density of the network and, hence the number of rendezvous between nodes, can therefore have an impact on the accuracy of the calibration.

Impact of number of reference signals. Unlike MLS, SCAN calculates different calibration parameters of an array to a certain reference depending on the number of other references in the regression setup. As shown in Sec. 3.3 this does however not violate the *no bias towards zero* property under the assumption of equal sensor noise of parent and child array. We now show that even if this assumption does not hold, the impact on the calibration error with different number of reference signals is negligible. We investigate the calibration error to phenomena $y_4 \in Y$ with different number of reference signals $L \in \{1, 2, 3, 4\}$. That means, if $L = 1$ each sensor array is only calibrated to y_4 . Accordingly, if $L = 2$ the array is calibrated to two phenomena $[y_3; y_4]$, and so forth. The standard deviation of the noise is set to $\sigma \in [0.2, 0.3]$ for all sensor arrays. Fig. 9 shows the calibration error

$y_4 \in Y$ over multiple hops for different L . We observe, that the number of references does not have a notable impact on the error. The differences only become clear after multiple hops, where SCAN with $L = 4$ achieves the lowest average error. In fact, at $h = 19$ the calibration errors for all L values differ by less than 3% .

Summary. Through extensive simulations, we demonstrate that our SCAN scheme outperforms the state-of-the-arts in multi-hop calibration for sensor arrays. Our SCAN scheme still accumulate mild errors in presence of numerous practical factors including noise variance, noise dependency, variable measurement range, etc., but still significantly outperforms MLS, which becomes unusable in these situations.

In the next two subsections, we further evaluate the performance of SCAN on two real-world air quality datasets, showing both the advantages of our novel SCAN scheme and the lessons learned in calibrating large-scale air quality measurements collected by low-cost mobile sensor arrays. We mainly compare our SCAN scheme with MLS, which we show is the best choice for one-hop sensor array calibration, and GMR, the state-of-the-art for multi-hop sensor calibration.

4.2 Metal Oxide Sensor Array

In this set of experiments, we evaluate the performance of SCAN on measurements collected by low-cost metal oxide based gas sensors. The sensor is a prototype featuring an array of sensing layers, whereof each individual layer in the array exhibits a different sensitivity to certain environmental gases. The sensitivity of these layers can be controlled by setting the temperature of the sensing layer to a specific value, which is a common technique for metal oxide based sensors [5]. Due to its small size and low power consumption, the sensor is suitable for a large variety of Internet-of-Things, wearable devices and crowd-sensing applications.

4.2.1 Setup. We deploy the low-cost metal oxide based gas sensors next to a static and highly accurate air monitoring station in Duebendorf, Switzerland to monitor the ambient ozone (O_3) concentration. We heat the sensing layers of the gas sensors to different temperatures to simulate heterogeneous sensor nodes with different noise levels. Further, we deploy a temperature sensor [27] for ambient temperature measurements. The sensors are sampled in an interval of 30 sec. As sensor array we use measurements from two of the gas sensor layers and from the temperature sensor, *i.e.*, $K = 3$. We collect measurements from July 2015 to July 2016.

Rendezvous path. We construct a rendezvous path of 11 hops because there were at most eleven sensor arrays measuring at the same time. Each calibration is trained on 200 samples within a time frame of at most two weeks. The calibration is tested on 200 different samples within the consecutive two weeks. The whole evaluation over one year of data is executed 500 times with a randomly re-sampled calibration path for every execution. For the calibration in the first hop to measurements of the reference station we use MLS instead of our SCAN method. As shown in Table 1 and also confirmed by evaluating the dataset, MLS performs better for one-hop calibration and, thus, the overall multi-hop calibration achieves better accuracy.

Ground-truth. For training the calibration parameters of the array at the initial hop and evaluating the calibration of all arrays in the path we use highly accurate ozone and temperature measurements from the station as reference.

4.2.2 Performance. We compare the multi-hop sensor array calibration performance of our SCAN approach to MLS. 2-Dimensional GMR is used to compare sensor array and simple sensor calibration by solely calibrating the measurements from a single sensing layer in the array that exhibits the highest sensitivity to ozone.

Error accumulation. The average RMSE over the whole time period for each calibrated array per hop is depicted in Fig. 10a. The results highlight the two major advantages of our SCAN sensor array calibration method; *(i)* SCAN reduces error accumulation over multiple hops compared to MLS and *(ii)* improves the overall calibration accuracy compared to the simple sensor calibration based on GMR. Over eleven hops MLS increases the average calibration error from 3.5 ppb to 5.85 ppb which is equal to a relative increase of over 60% error. SCAN considerably reduces this error accumulation to a relative increase of 30%. Compared to the simple sensor calibration with GMR, sensor

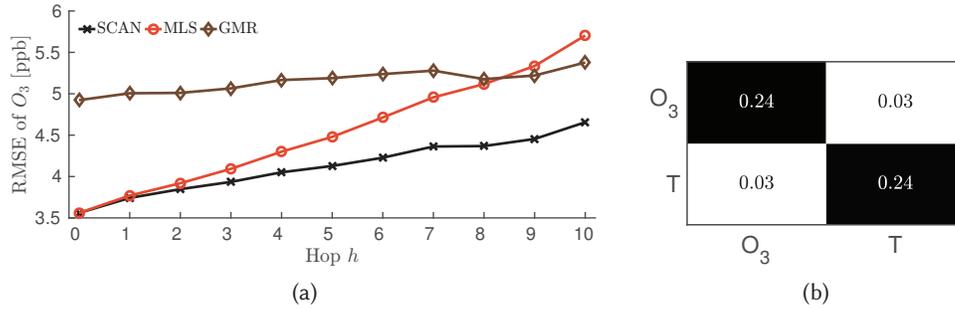


Fig. 10. (a) Sensor array calibration error over multiple hops when calibrated to ambient ozone (O_3) concentration. SCAN clearly outperforms MLS and GMR. (b) Covariance matrix of sensor noise after sensor calibration to O_3 and temperature.

array calibration based on SCAN improves the calibration error by up to 1.36 ppb O_3 concentration. Over all hops SCAN achieves a 16% to 38% smaller error than GMR and an up to 23% smaller error than MLS.

Investigation on noise characteristics. As shown in Sec. 3.4, our SCAN approach is able to completely remove the bias towards zero of its regression parameters if the noise components of the sensor array are uncorrelated, see assumption (21). We therefore investigate if this assumption holds for our sensor arrays. Because N_1 is not directly measurable without the knowledge of the true calibration matrix \bar{B} , we calculate the covariance matrix $N_2 N_2^T$ of the calibration error N_2 from the one-hop sensor array calibration using MLS, *i.e.*, the best possible calibration for each array. The reference traces for ozone and temperature have different units and ranges. Thus, we scale them to assure that the noise components of the low-cost sensor arrays have equal impact, *i.e.*, equal variance. Fig. 10b shows the average covariance matrix for all eleven sensor arrays. We observe that the two noise components, *i.e.*, the calibration error of ozone O_3 and temperature T , are not completely uncorrelated. However, the diagonal elements dominate the off-diagonal elements, *i.e.*, the $N_2 N_2^T$ matrix resembles a diagonal matrix. This result relates to our assumptions on the noise components in Sec. 3 and Sec. 4.1. In conclusion, our SCAN approach considerably improves the calibration accuracy even if the noise components are correlated, *i.e.*, (21) is not satisfied.

4.3 Mobile Air Pollution Sensor Network

In this set of experiments, we evaluate the performance of SCAN on a large dataset from a real-world mobile air pollution sensor network deployment [19]. We first demonstrate the benefits of using multi-hop calibration over one-hop calibration in a real-world deployment. Further, we show that our sensor array network calibration outperforms MLS and GMR for arrays of sensors with low selectivity.

4.3.1 Setup. The dataset was collected by air quality measurement boxes mounted on top of ten streetcars of the public transport network in the city of Zurich, Switzerland, depicted in Fig. 11a and Fig. 1a. Each measurement box includes an ozone (O_3) [28], a carbon monoxide (CO) [1] and a temperature sensor [27]. The sampling interval of the sensor array is set to 30 sec. Each box is equipped with a GPS receiver to record location and time of each measurement. All the sampled data is transferred to a global database via GSM. We filter incomplete samples due to network connection or sensor failures and only focus on calibrating errors induced by sensor dependencies as mentioned in Sec. 1. Previous studies [20, 29] have uncovered a substantial dependency on ambient temperature of the deployed low-cost O_3 and CO sensor. Therefore we augment the gas sensors with the temperature sensor to an array and calibrate the concurrent measurements to the corresponding high-quality reference signals.

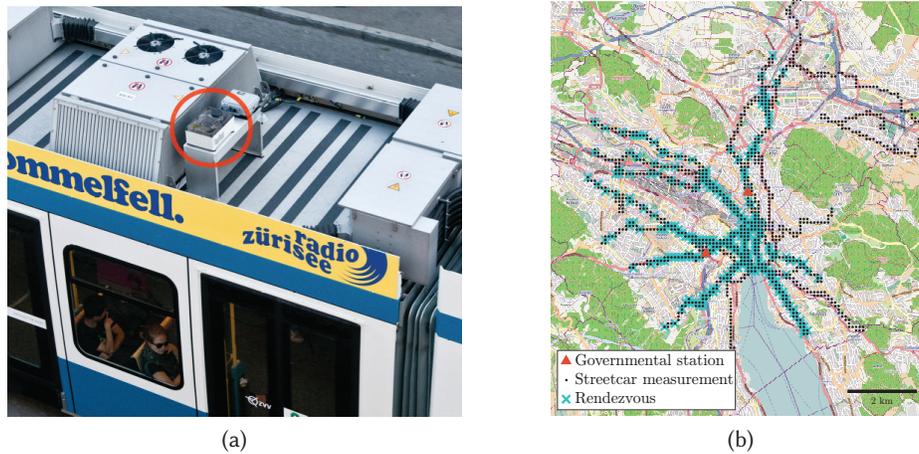


Fig. 11. (a) Measurement box mounted on top of a streetcar. (b) Locations of sensor measurements and rendezvous between two or more sensor arrays.

Rendezvous path. As shown in Fig. 11b, the streetcars meet occasionally at different times and locations. We exploit these rendezvous between streetcars to construct rendezvous paths. In the initial hop of a rendezvous path one sensor array is calibrated with measurements from the governmental station. In all following hops an already calibrated sensor array calibrates an uncalibrated one using measurements at rendezvous between two streetcars. We define a rendezvous between two sensor arrays as within a time interval of 5 *min* and spatial closeness of 50 *m*, which has been validated through extensive testing in [24]. Each rendezvous path consists of at least 200 samples per sensor array pair within a time window of at most four weeks. The average length of all rendezvous paths is 3 hops.

Ground-truth. As reference signals we use O_3 , CO and temperature measurements from two static monitoring stations within the deployment area. We use data from both stations to train the calibration of any initial hop. This assures that we sufficient measurements, *i.e.*, at least 200 samples, for one-hop calibration and removes potential error sources, such as variable measurement range as described in Sec. 4.1.3 Each calibration is calculated on a training dataset within four weeks. The performance of each sensor array calibration is evaluated on a dataset from the consecutive four weeks. These two datasets do not overlap in time. The location of these two monitoring stations is depicted in Fig. 11b. The streetcars operate daily on various routes in the city and, thus, achieve a high spatial and temporal measurement coverage, illustrated in Fig. 11b.

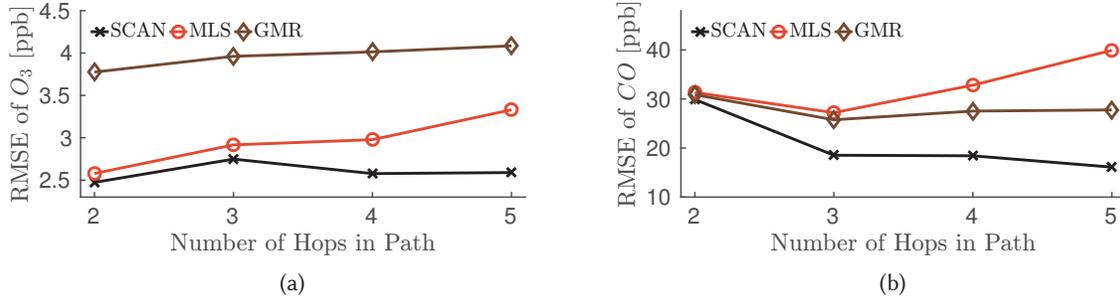
4.3.2 Performance. To benchmark the multi-hop calibration performance we compare a particular array at the last hop of a rendezvous path and its baseline calibration. The baseline calibration is obtained by calibrating the array to reference measurements from the governmental stations using MLS, *i.e.*, the best possible calibration for the array in question. This approach removes effects on the overall measurements accuracy of each array that cannot be compensated by calibration, such as mobility influences [2, 3].

Overall, we use roughly $2.2 \cdot 10^5$ rendezvous between seven of the ten streetcars and over 550 different rendezvous paths out of 56 million sensor samples recorded from March 2014 to March 2016 for evaluation.

One-hop versus multi-hop calibration. This evaluation aims to show the necessity of multi-hop calibration when calibrating datasets collected by a mobile sensor network. Depending on the availability of reference

	Max. calibrated arrays	Avg. calibrated arrays
One-hop	3	32.7%
Multi-hop	7	94.3%

Table 2. Number of calibrated arrays with one-hop and multi-hop calibration

Fig. 12. Calibration error of sensor arrays compared to baseline depending on the number of hops in the rendezvous path. SCAN achieves best results for both target pollutants, ozone (O_3 : Fig. 12a) and carbon monoxide (CO : Fig. 12b).

stations one-hop calibration is able to calibrate only a fraction of all the sensor arrays in the network. In the streetcar deployment there are two highly accurate reference stations that can be used to calibrate sensor arrays on streetcars that pass by these two stations. All the remaining sensors can be calibrated using multi-hop calibration. Table 2 shows that at most three out of seven arrays can be calibrated with one-hop calibration whereas it is possible to calibrate all arrays with multi-hop calibration. Over the two years the multi-hop approach calibrates in average over 94% of all arrays every month which is an improvement by a factor of 2.88 compared to one-hop calibration. This result clearly shows the benefit of using multi-hop calibration over one-hop calibration. Reasons for not calibrating 100% of all sensors are the irregular schedules and routes of the streetcars over two years or missing sensor data. As a consequence certain streetcars did not meet other streetcars often enough for a successful calibration on a monthly basis. Note that dynamically selecting rendezvous paths to ensure network-wide calibration is out of the scope of this paper. We expect that in a more dense network the rate of calibrated sensors is even higher.

Sensor array versus simple sensor calibration. This evaluation aims to validate the effectiveness of our SCAN scheme on large-scale real-world mobile sensor networks. Fig. 12 shows the calibration error depending on the number of hops in the rendezvous path for ozone (O_3 , Fig. 12a) and carbon monoxide (CO , Fig. 12b). We show again the two important contributions of our proposed SCAN approach. SCAN achieves the lowest error accumulation over multiple hops and sensor array calibration improves the overall accuracy compared to the simple sensor calibration based on GMR for both target pollutants. Overall SCAN achieves an up to 42% lower calibration error than GMR and up to 60% compared to MLS over 5 hops, which can dramatically benefit further quantitative analysis such as validation on air pollution models and health studies.

In conclusion, multi-hop calibration considerably increases the number of calibrated sensor arrays in the deployment. Additionally, our proposed SCAN method improves the calibration accuracy, especially when compared to the simple sensor calibration based on GMR.

Discussions. While evaluations on the mobile sensor network deployment show that our SCAN scheme outperforms the state-of-the-arts, we also observe that the benefits are not as distinct as in the simulations and the metal oxide sensor arrays. Here we discuss multiple issues to further clarify the applicability of SCAN.

- **SCAN works in practice even if the assumptions on noise may not hold.** As shown in section Sec. 4.2, noise from real-world sensors can correlate but exhibit typically low cross-correlation, therefore SCAN still works and notably outperforms MLS.
- **SCAN shows more notable benefits for long rendezvous paths.** The current deployment consists only of seven nodes with 5 hops at most. SCAN reduces calibration errors by 42% to 60% after 5 hops, and we expect more notable gain in minimizing error accumulation over multiple hops for larger networks with many low-cost sensor arrays.
- **SCAN is mainly designed to tackle cross-sensitivities in multi-hop calibration.** There are possibly additional error sources in a mobile sensor deployment, such as network faults and mobility. For instance, recent studies [3] tackle the impact of slow sensor dynamics in a mobile deployment. SCAN can be combined with these schemes to further reduce errors in sensor measurements.

5 RELATED WORK

This section summarizes existing work related to the two main topics addressed in this paper, sensor arrays and multi-hop calibration.

Sensor Array Calibration. Calibrating a set of multiple different sensors to target phenomena is a popular approach to tackle the problems of low-cost sensors such as cross-sensitivity, low stability and dependency on environmental conditions. A large body of work [11, 20, 29] shows the benefits of sensor array calibration.

Eugster and Kling [11] compensate for the dependency to environmental conditions of a low-cost methane (CH_4) sensor. They augment the gas sensor with a temperature and relative humidity sensor and successfully calibrate the array to ambient methane concentrations using a multi-dimensional linear regression.

Spinelle *et al.* [29] present the calibration of 13 different gas sensors. They improve the measurement accuracy by compensating for cross-sensitivities to ozone, temperature and humidity of those sensors with the appliance of multiple least-squares and neural networks.

In this work we adopt the idea of sensor array calibration and present a regression technique that can also be applied to multi-hop calibration.

Multi-hop Calibration. Exploiting rendezvous between sensors in a mobile network is a popular approach for sensor calibration [15, 21, 24, 25, 34].

Markert *et al.* [21] present a privacy-protecting multi-hop calibration which enables its appliance for participatory environmental sensing. They demonstrate the benefits of multi-hop over one-hop calibration while protecting the location privacy of participating users.

Ye *et al.* [34] present a multi-hop calibration approach for a large network of mobile barometric sensors embedded into commercial smartphones. The work shows two major challenges; *(i)* common low-cost barometric sensors show a considerable cross-sensitivity to temperature and wind regimes and *(ii)* errors accumulate along rendezvous paths. In order to reduce error accumulation over multiple hops they propose an approach that finds paths with minimal length and calibrates all sensors in the network. They show that this is a NP-complete problem and propose a heuristic solution. We are convinced that our SCAN approach can *(i)* compensate for the barometric sensors cross-sensitivities when augmented to an appropriate sensor array and *(ii)* does not require to find short rendezvous paths for improved measurement quality.

Saukh *et al.* [25] introduce the concept of sensor rendezvous and its potential for applications such as sensor fault detection and calibration. In a further work [24] they present multi-hop calibration based on a rendezvous path. They show that ordinary least-squares (OLS) suffers from error accumulation when applied to multi-hop

calibration and propose the use of the geometric mean regression (GMR). GMR does not suffer from error accumulation and, thus, improves network-wide data accuracy. However, GMR can only be applied to single sensors and, hence, can not compensate for limiting effects of common low-cost sensors, such as low selectivity. Our proposed method can be applied to sensor arrays while also minimizing error accumulation and, thus, improves data accuracy of low-cost sensor deployments even more.

6 CONCLUSION

Monitoring air pollution with mobile wireless sensor networks has received increasing research interest in recent years. Low-cost, portable air quality sensors on the market introduced the opportunity for large scale deployments with high spatial coverage. However, maintaining high-quality measurements from low-cost air pollution sensors is challenging. Low-cost air pollution sensors not only drift over time, but are also cross-sensitive to interfering gases and depend on meteorological conditions. Therefore calibrating the air quality measurements is vital if the dataset is for quantitative analysis such as air pollution modeling and health studies. Pioneer research has explored constructing sensor arrays to compensate cross-sensitivities and meteorological dependencies, yet existing multi-hop calibration techniques lead to dramatic error accumulation when applied to sensor arrays, making multi-hop sensor array calibration an open question.

We propose *sensor array network calibration* (SCAN), a novel constrained multi-dimensional linear regression technique, that (i) calibrates sensor arrays and (ii) reduces error accumulation over multiple hops. We theoretically prove that SCAN is free from regression dilution, the root cause of error accumulation, even in presence of measurement noise. Extensive evaluations on two datasets of 56 million samples collected over three years demonstrate the benefits of SCAN over the state-of-the-art calibration techniques. SCAN compensates for all major limiting factors to maintain high-quality measurements from low-cost air pollution sensors, thus improving reliability of large-scale air quality datasets. We envision SCAN as a general calibration technique for not only air pollution monitoring, but also a range of mobile sensor network applications with dependent sensors, especially in participatory and crowdsourcing sensing.

REFERENCES

- [1] Alphasense. 2014. CO-B4 4-Electrode carbon monoxide sensor (datasheet). <http://goo.gl/egp6Sm>. (2014).
- [2] Adrian Arfire, Ali Marjovi, and Alcherio Martinoli. 2016. Enhancing Measurement Quality through Active Sampling in Mobile Air Quality Monitoring Sensor Networks. In *Proc. of AIM*. IEEE, 1022–1027.
- [3] Adrian Arfire, Ali Marjovi, and Alcherio Martinoli. 2016. Mitigating Slow Dynamics of Low-Cost Chemical Sensors for Mobile Air Quality Monitoring Sensor Networks. In *Proc. of EWSN*. ACM, 159–167.
- [4] Sudipto Banerjee and Anindya Roy. 2014. *Linear Algebra and Matrix Analysis for Statistics*. Taylor & Francis Group.
- [5] N. Barsan, D. Koziej, and U. Weimar. 2007. Metal oxide-based gas sensor research: How to? *Sensors and Actuators B: Chemical* 121, 1 (2007), 18 – 35.
- [6] R. Bro. 2003. Multivariate calibration What is in chemometrics for the analytical chemist? *Analytica Chimica Acta* 500 (2003), 185–194.
- [7] Matthias Budde, Rayan El Masri, Till Riedel, and Michael Beigl. 2013. Enabling Low-cost Particulate Matter Measurement for Participatory Sensing Scenarios. In *Proc. of MUM*. ACM, 19:1–19:10.
- [8] Matthias Budde, Marcel Köpke, and Michael Beigl. 2015. Robust In-situ Data Reconstruction from Poisson Noise for Low-cost, Mobile, Non-expert Environmental Sensing. In *Proc. of ISWC*. ACM, 179–182.
- [9] Yaron Danon and Mark Embrechts. 1992. Least Squares Fitting Using Artificial Neural Networks. *Intelligent Engineering Systems through Artificial Neural Networks 2* (1992).
- [10] Norman Draper and Yonghong Yang. 1997. Generalization of the Geometric Mean Functional Relationship. *Computational Statistics and Data Analysis* 23, 3 (1997), 355–372.
- [11] W. Eugster and G. W. Kling. 2012. Performance of a low-cost methane sensor for ambient concentration measurements in preliminary studies. *Atmospheric Measurement Techniques* 5, 8 (2012), 1925–1934.
- [12] Chris Frost and Simon G. Thompson. 2000. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society Series A* 163, 2 (2000), 173–189.

- [13] Kaibo Fu, Wei Ren, and Wei Dong. 2017. Multihop Calibration for Mobile Sensing: k-hop Calibratability and Reference Sensor Deployment. In *Proc. of INFOCOM*. IEEE.
- [14] Gene H. Golub and Charles F. Van Loan. 1980. An Analysis of the Total Least Squares Problem. *SIAM J. Numer. Anal.* 17, 6 (1980), 883–893. <http://www.jstor.org/stable/2156807>
- [15] David Hasenfratz, Olga Saukh, and Lothar Thiele. 2012. On-the-fly calibration of low-cost gas sensors. In *Proc. of EWSN*. ACM, 228–244.
- [16] W. Jiao, G. Hagler, R. Williams, R. Sharpe, R. Brown, D. Garver, R. Judge, M. Caudill, J. Rickard, M. Davis, L. Weinstock, S. Zimmer-Dauphinee, and K. Buckley. 2016. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmospheric Measurement Techniques* 9, 11 (2016), 5281–5292.
- [17] Henry D. Kahn. 1973. Note On The Distribution of Air Pollutants. *Journal of the Air Pollution Control Association* 23, 11 (1973), 973–973.
- [18] Sunyoung Kim, Eric Paulos, and Mark D. Gross. 2010. WearAir: Expressive T-shirts for Air Quality Sensing. In *Proc. of TEL*. ACM, 295–296.
- [19] Jason Jingshi Li, Boi Faltings, Olga Saukh, David Hasenfratz, and Jan Beutel. 2012. Sensing the Air We Breathe – The OpenSense Zurich Dataset. In *Proc. of AAAI*. AAAI, 323–325.
- [20] Balz Maag, Olga Saukh, David Hasenfratz, and Lothar Thiele. 2016. Pre-Deployment Testing, Augmentation and Calibration of Cross-Sensitive Sensors. In *Proc. of EWSN*. ACM, 169–180.
- [21] Jan F. Markert, Matthias Budde, Gregor Schindler, Markus Klug, and Michael Beigl. 2016. Private Rendezvous-based Calibration of Low-Cost Sensors for Participatory Environmental Sensing. In *Proc. of Urb-IoT*. ACM, 82–85.
- [22] Emiliano Miluzzo, Nicholas D. Lane, Andrew T. Campbell, and Reza Olfati-Saber. 2008. CaliBree: A Self-calibration System for Mobile Sensor Networks. In *Proc. of DCOSS*. IEEE, 314–331.
- [23] Luis Sánchez, Verónica Gutiérrez, Jose Antonio Galache, Pablo Sotres, Juan Ramón Santana, Javier Casanueva, and Luis Muñoz. 2013. SmartSantander: Experimentation and service provision in the smart city. In *Proc. of WPMC*. IEEE, 1–6.
- [24] Olga Saukh, David Hasenfratz, and Lothar Thiele. 2015. Reducing Multi-Hop Calibration Errors in Large-Scale Mobile Sensor Networks. In *Proc. of IPSN*. ACM/IEEE, 274–285.
- [25] Olga Saukh, David Hasenfratz, Christoph Walser, and Lothar Thiele. 2013. On Rendezvous in Mobile Sensing Networks. In *Proc. of RealWSN*. Springer, 29–42.
- [26] Peter H. Schönemann. 1966. A generalized solution of the orthogonal procrustes problem. *Psychometrika* 31, 1 (1966), 1–10.
- [27] Sensirion AG. 2016. SHTC1 Humidity and Temperature Sensor IC (datasheet). <https://goo.gl/boJh2T>. (2016).
- [28] SGX Sensortech. 2014. MiCS-OZ-47 ozone sensor (datasheet). <http://goo.gl/C49tcw>. (2014).
- [29] Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto Bonavitacola. 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical* 215 (2015), 249 – 257.
- [30] Yele Sun, Guoshun Zhuang, Ying Wang, Lihui Han, Jinghua Guo, Mo Dan, Wenjie Zhang, Zifa Wang, and Zhengping Hao. 2004. The air-borne particulate pollution in Beijing – concentration, composition, distribution and sources. *Atmospheric Environment* 38, 35 (2004), 5991 – 6004.
- [31] Rundong Tian, Christine Dierk, Christopher Myers, and Eric Paulos. 2016. MyPart: Personal, Portable, Accurate, Airborne Particle Counting. In *Proc. of CHI*. ACM, 1338–1348.
- [32] Chris Tofallis. 2002. Model Fitting for Multiple Variables by Minimising the Geometric Mean Deviation. In *Total Least Squares and Errors-In-Variables Modeling: Algorithms, Analysis And Applications*.
- [33] E. B. Woolley. 1941. The method of minimized areas as a basis for correlation analysis. *Econometrica* 9(1) (1941), 38–62.
- [34] Haibo Ye, Tao Gu, Xianping Tao, and Jian Lu. 2014. SBC: Scalable Smartphone Barometer Calibration Through Crowdsourcing. In *Proc. of MobiQuitous*. Springer, 60–69.
- [35] Yan Zhuang, Feng Lin, Eun-Hye Yoo, and Wen Yao Xu. 2015. AirSense: A Portable Context-sensing Device for Personal Air Quality Monitoring. In *Proc. of MobileHealth*. ACM, 17–22.

Received February 2017; revised March 2017; accepted May 2017