# Field Reconstruction-Based Non-Rendezvous Calibration for Low-Cost Mobile Sensors


Ji Luo
Tsinghua-Berkeley Shenzhen Institute, Tsinghua University
Shenzhen, China
luo-j21@mails.tsinghua.edu.cn

Yiling Hu
Shenzhen Environmental Thinking Science and Technology (ETST) Company Ltd.
Shenzhen, China
yiling510hu@gmail.com

Chengzhao Yu
Tsinghua-Berkeley Shenzhen Institute, Tsinghua University
Shenzhen, China
yucz21@mails.tsinghua.edu.cn

Chaopeng Hong
Shenzhen Key Laboratory of Ecological Remediation and Carbon Sequestration, Institute of Environment and Ecology, Shenzhen International Graduate School, Tsinghua University
Shenzhen, China
hongcp@sz.tsinghua.edu.cn

Xiao-Ping Zhang
Shenzhen International Graduate School, Tsinghua University
RISC-V International Open Source Laboratory
Shenzhen, China
Department of Electrical, Computer and Biomedical Engineering, Ryerson University
Toronto, Canada
xiaoping.zhang@ee.ryerson.ca

Xinlei Chen*
Shenzhen International Graduate School, Tsinghua University
Pengcheng Laboratory
RISC-V International Open Source Laboratory
Shenzhen, China
chen.xinlei@sz.tsinghua.edu.cn



## ABSTRACT

Low-cost air pollution sensors (LCS) deployed on urban vehicles (e.g., taxis, buses) have emerged as a cost-effective solution for fine-grained air pollution monitoring. However, these mobile LCSs suffer from measurement drifting in real-world scenarios, necessitating a post-deployment real-time calibration. Unfortunately, the limited availability of urban real reference stations (RRS) restricts the calibration opportunities for LCSs. This paper proposes a non-rendezvous method that addresses this challenge by establishing virtual reference stations (VRS), which offer additional calibration opportunities for LCSs. Through the air pollution field reconstruction, the readings of VRSs are inferred from RRSs' data. Furthermore, a confidence assessment mechanism is developed to quantify the uncertainty of established VRSs. Finally, a field experiment is conducted to demonstrate the effectiveness of the proposed method, showcasing a 25% improvement over the advanced baseline.


## CCS CONCEPTS

• **Computer systems organization** → **Sensor networks**; • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools.*


## KEYWORDS

Low-cost sensor, post-deployment calibration, field reconstruction

**ACM Reference Format:**
Ji Luo, Yiling Hu, Chengzhao Yu, Chaopeng Hong, Xiao-Ping Zhang, and Xinlei Chen. 2023. Field Reconstruction-Based Non-Rendezvous Calibration for Low-Cost Mobile Sensors. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), October 08–12, 2023, Cancun, Quintana Roo, Mexico.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3594739.3612908


## 1 INTRODUCTION

Urban air pollution has long posed a significant threat to public health, prompting the establishment of fine-grained city-scale air pollution monitoring systems in recent years [3]. The deployment of such systems serves multiple purposes: For city residents, they can reconstruct air pollution maps to improve the precision of travel planning [11, 12]. For environmental department, these systems play a crucial role in the precise implementation of pollution control measures, such as urban pollution source detection and water sprayers scheduling [13]. To achieve a higher spatial-temporal density in air quality sensing, a commonly applied method involves the deployment of a large number of low-cost sensors (LCS) on urban vehicles, such as taxis and buses [18, 29]. These mobile LCSs are specifically designed to be lightweight and cost-effective [15]. By utilizing these sensors, the existing network of static air pollution monitors can be augmented, leading to a more comprehensive and cost-efficient urban environmental monitoring [14].

However, despite their advantages in terms of reduced cost and weight compared to professional monitoring devices, LCSs often have issues with accuracy when deployed in real-world scenarios, such as measurement drifting [19, 25]. Consequently, post-deployment real-time calibration is crucial for the mobile LCS-based


*Xinlei Chen is the corresponding author





*UbiComp/ISWC '23 Adjunct, October 08–12, 2023, Cancun, Quintana Roo, Mexico*



ACM ISBN 979-8-4007-0200-6/23/10.
https://doi.org/10.1145/3594739.3612908

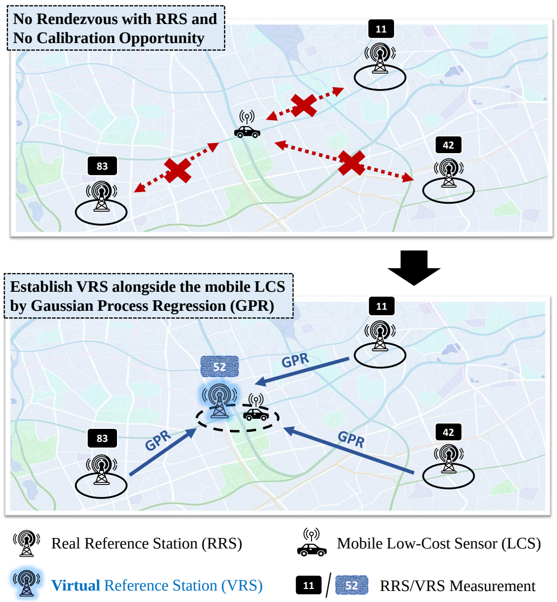


**Figure 1: Illustration of the proposed method.**

monitoring systems. Unlike laboratory calibration, post-deployment calibration operates "on-the-fly" [17]. Its primary aim is to establish and sustain a real-time calibration function for each mobile LCS, transforming raw readings into more accurate ones.

Several related works have attempted to address such post-deployment calibration. Initially, the approach involved calibrating LCSs when they rendezvous with static monitoring stations in the city [26]. These static monitoring stations, commonly referred as "reference stations" in calibration systems, are typically equipped with high-cost professional devices that consistently provide accurate air quality measurements. Unfortunately, due to cost constraints in city management, these reference stations are often sparsely deployed across a city, making rendezvous between reference stations and vehicles carrying LCSs unlikely [22]. Therefore, subsequent studies have focused on increasing calibration opportunities, leading to the emergence of multi-hop calibration scheme [16, 21]. In addition to rendezvous with static stations, multi-hop calibration assist the mobile LCS in leveraging information obtained from rendezvous with other mobile LCSs, thereby increasing the calibration opportunities. However, the reliability of the results obtained through multi-hop calibration may be compromised due to the potential accumulation of errors [21].

**This paper proposes a non-rendezvous method to calibrate mobile air pollution sensors**. Aiming to increase the calibration opportunities, the proposed method tries to explore additional reference information beyond situations where direct rendezvous is possible. Specifically, through the air pollution field reconstruction, virtual reference stations (VRS) are established adjacent to the mobile LCSs, which are located at a distance from the real reference stations (RRSs). The readings of these VRSs are inferred from other RRSs in the city by the Gaussian Process Regression (GPR) method. This design enables the mobile LCSs to collect reference data without physically meeting the RRSs. Additionally, a confidence assessment mechanism is developed based on the results of GPR to quantify the uncertainty of these established VRSs. Once a mobile LCS has collected a sufficient amount of reference data with varying levels of confidence from both the VRSs and RRSs, neural networks are employed to incorporate the confidence information into the training of calibration function. To evaluate the effectiveness of the proposed method, 22 mobile sensing devices are deployed in one city for an 8-month field experiment. The result demonstrates a 25% improvement compared to the best baseline method. The contributions of this paper are summarized as follows:

- Proposing a non-rendezvous method to calibrate mobile air pollution LCSs, which utilizes GPR to establish VRS, providing more calibration opportunities for mobile LCSs.
- Developing a confidence assessment mechanism to quantify the uncertainty of established VRS and designing a neural network to integrate such confidence information into mobile LCSs' calibration.
- Deploying 22 mobile sensing devices in one city and conducting an 8-month field experiment to validate the effectiveness of the proposed method.

The remainder of the paper is organized as follows: Section 2 introduces the problem formulation. Section 3 details the methodology of the proposed method. Section 4 presents the evaluation process. Finally, Section 5 concludes the paper.

## 2 PROBLEM DEFINITION

In this paper, three object concepts are specified as follows:

- **Low-cost Sensor (LCS):** Owing to their cost-effectiveness and lightweight attributes, LCSs have found wide-ranging applications in mobile platforms, particularly swarms [5, 6]. Notwithstanding, they frequently exhibit issues concerning accuracy. The focus of this paper is on the calibration of LCSs. The variable $x$ represents raw LCS readings.
- **Real Reference Station (RRS):** These sensors are deemed to be highly accurate and are employed as reference in the context of mobile air pollution sensing. Typically, RRSs are stationary monitoring stations equipped with high-cost devices. They serve the purpose of calibrating the LCSs or assessing their performance. The readings obtained from the RRSs are denoted by the variable $y$.
- **Virtual Reference Station (VRS):** In situations where a vehicle carrying a mobile LCS is not located in close proximity to an RRS, a VRS is established alongside the mobile LCS to facilitate its calibration, as illustrated in Figure 1. The readings of the VRS are inferred from the neighboring RRSs, which will be elaborated upon in Section 3.1. The reliability assessment of the VRS's data will be explored in Section 3.2.

The primary calibration approach involves utilizing RRSs to create VRSs for calibrating the LCSs. The ultimate objective of real-time post-deployment calibration is to determine the optimal calibration function $\widehat{f}$ for each of the $N_{LCS}$ mobile LCSs, which is:

$$\underset{\widehat{f}}{argmin} \sum_{i=1}^{N_{LCS}} \sum_{t=t_0}^{t_n} d(\widehat{f_i}(x_{i,t}), y_{i,t}), \tag{1}$$

where $d(\cdot)$ represents the error evaluation function, and $t_0$ and $t_n$ indicate the sensing start and end time, respectively. The optimal fitting method for the $\widehat{f}$ function will be discussed in Section 3.3.

# 3 METHODOLOGY

## 3.1 Data Inference for VRS

To ensure precise estimation of VRS readings, we employ GPR, a prevalent method for field reconstruction [10], to infer data for the VRS's spatial-temporal region.

*3.1.1 Establishment of the Gaussian Process Model.* For each meta-area characterized by the spatial-temporal coordinate $\mathbf{r}$, the pollution data $g(\mathbf{r})$ is regarded as a function that maps $\mathbf{r}$. $g(\mathbf{r})$ is treated as a function drawn from a Gaussian Process (GP), which is:

$$g(\mathbf{r}) \sim \mathcal{GP}(m(\mathbf{r}), k(\mathbf{r}, \mathbf{r}^{'})). \tag{2}$$

$m(r)$ represents the mean function and $k(\mathbf{r}, \mathbf{r}^{'})$ represents the covariance function. Since we normalize all training data using the z-score normalization method [23], we employ a mean function with a constant value of 0. For the covariance function, we use the exponential kernel function to measure the covariance of different spatial-temporal points:

$$k(\mathbf{r}, \mathbf{r}^{'}) = \sigma_l * exp(-(\mathbf{r} - \mathbf{r}^{'})^2/(2\lambda_l^2)), \tag{3}$$

where $\lambda_l$ and $\sigma_l$ are the scale parameters, which need to be optimized using the RRS's data.

Eq. 2 implies that for any meta-area with the spatial-temporal coordinate $\mathbf{r}$, their pollution concentrations $y$ satisfy a multivariate joint normal distribution. In our case, the meta-areas are classified into two types: $\mathrm{R} = \dot{\mathrm{R}} \cup \mathrm{R}^*$, where $\dot{\mathrm{R}}$ represents the RRS's areas and $\mathrm{R}^*$ represents VRS's areas. We denote $\dot{Y}$ and $Y^*$ as the pollution concentration data for these two kinds of areas, respectively. Thus, we have $\dot{R} = (\dot{\mathbf{r}}_1, \dot{\mathbf{r}}_2, ..., \dot{\mathbf{r}}_{\dot{N}})^T, R^* = (\mathbf{r}_1^*, \mathbf{r}_2^*, ..., \mathbf{r}_{N^*}^*)^T, \dot{Y} = (\dot{y}_1, \dot{y}_2, ..., \dot{y}_{\dot{N}})^T, Y^* = (y_1^*, y_2^*, ..., y_{N^*}^*)^T$, where $\dot{N}$ and $N^*$ represent the number of these two kinds of areas, respectively. They satisfy the following joint normal distribution:

$$\begin{pmatrix} \dot{Y} \\ Y^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} M(\dot{R}) \\ M(R^*) \end{pmatrix} \begin{pmatrix} K(\dot{R}, \dot{R}) & K(\dot{R}, R^*) \\ K(R^*, \dot{R}) & K(R^*, R^*) \end{pmatrix}\right). \tag{4}$$

Here, $K(\cdot)$ is the matrix form of the $k(\cdot)$, which means $K(\cdot)$ outputs a matrix containing every pairwise relationship for the two input vectors. We can now use the Bayesian conditional probability formula to obtain $p(Y^*|\dot{R}, \dot{Y}, R^*)$. However, before that, we will first use RRS's data to determine the hyperparameters $\lambda_l$ and $\sigma_l$ in the covariance function, optimizing the entire GP model.

*3.1.2 Optimization of the Gaussian Process Model.* To optimize the GP model, specifically, to determine the parameters $\lambda_l$ and $\sigma_l$ of the kernel function in Eq. 3, the probability of occurrence of $\dot{Y}$ in the current GP model must be maximized. In this paper, the Marginal Log-likelihood is employed to find the optimal values for $\lambda_l$ and $\sigma_l$:

$$\begin{aligned} &logp(\dot{Y}|\lambda_l, \sigma_l) = log\mathcal{N}(0, K(\dot{R}, \dot{R}|\lambda_l, \sigma_l)) \\ &= -\frac{1}{2}\dot{Y}^T K(\dot{R}, \dot{R}|\lambda_l, \sigma_l)^{-1}\dot{Y} - \frac{1}{2}log|K(\dot{R}, \dot{R}|\lambda_l, \sigma_l)| - \frac{\dot{N}}{2}log(2\pi). \end{aligned} \tag{5}$$

Here, "L-BFGS-B" optimization method [31] is utilized to find the most suitable values for $\lambda_l$ and $\sigma_l$.

*3.1.3 Inference by the Gaussian Process Model.* Once the optimal GP model with the most suitable parameters $\lambda_l$ and $\sigma_l$ is obtained, the inference of data in VRS's areas can commence. From Eq. 4, for any VRS's area $\mathbf{r}^* \in R^*$, the Bayesian method provides:

$$\begin{aligned} &p(y^*|\dot{R}, \dot{Y}, R^*) \sim \mathcal{N}(y^*|\mu^*, \sigma^*), \\ &\mu^* = K(\mathbf{r}^*, \dot{R})K(\dot{R}, \dot{R})^{-1}\dot{Y}, \\ &\sigma^* = K(\mathbf{r}^*, \mathbf{r}^*) - K(\mathbf{r}^*, \dot{R})K(\dot{R}, \dot{R})^{-1}K(\dot{R}, \mathbf{r}^*). \end{aligned} \tag{6}$$

The distribution of pollution concentration in the VRS's areas is obtained, and the mean value $\mu^*$ is used as the inference result for the pollution concentration value of the VRS's reading. The variance value $\sigma^*$ is employed to measure the confidence of the predicted value of VRS's data. Notably, a variance of 0 is assigned to RRS' data.

## 3.2 Confidence Assessment for VRS

Generally, a distribution with higher variance is considered to have lower confidence. Additionally, to perform timely calibration, earlier sensed data is assigned lower confidence. Due to these factors, the following conversion formula is designed to obtain the confidence $c$ from the predicted distribution's variance:

$$c = \gamma^{|t_0 - t_s|}/(\sigma + \epsilon), \tag{7}$$

where $t_0$ represents the current time and $t_s$ represents the time when the data is sensed (or virtually sensed). $\gamma$ is a hyperparameter used to adjust the importance of data's immediacy. $\epsilon$ is another scale hyperparameter that adjusts whether the calibration focuses more on RRS's data or VRS's data, as all RRS's data will have $\sigma^*$ set to 0. The smaller the value of $\epsilon$, the larger the RRS's data's $c$ relative to the inferred data with non-zero variance.

## 3.3 Neural Network for Confidence Weighting

Upon completion of the data inference and confidence assessment process, the VRSs can be established with confidence-based readings at any given time and location. In other words, as for each mobile LCS within the time interval $t_0$ to $t_n$, a set of calibration data $(x_1, \mathbf{m_1}, c_1, y_1), (x_2, \mathbf{m_2}, c_2, y_2), ..., (x_N, \mathbf{m_N}, c_N, y_N)$ can be acquired. This set comprises sensors' raw readings $x_i$, meteorology data $\mathbf{m_i}$ (temperature and humidity), corresponding confidence values $c_i$, and reference data $y_i$ obtained from both VRSs and RRSs. Subsequently, using this dataset, each LCS is required to train a calibration function $\widehat{f}$ that transforms its raw readings into more precise and accurate values.

A crucial challenge here lies in handling the confidence values $c$. To address this, a neural network with a specific loss function is employed. The neural network, denoted as $\widehat{f}(\cdot; \theta)$ with parameter $\theta$, consists of an input layer that incorporates LCS raw readings $x$ and meteorological data $m$, and an output layer with a single neuron to output the post-calibrated data $y$. The loss function employed incorporates confidence weighting, as expressed by the following equation:

$$L(\theta) = \frac{\sum_{i=1}^{N} c_i|\widehat{f}(x_i, \mathbf{m}_i; \theta) - y_i|^2}{\sum_{i=1}^{N} c_i}. \tag{8}$$

By considering confidence values $c_i$ in the loss function, the neural network training process assigns more significance to data points with higher confidence, resulting in a calibration function that accurately fits reliable data. This weighted approach improves system calibration by accommodating varying degrees of confidence in different inference data.

Here, the influence of temperature and humidity on LCS calibration has been established in previous studies [2]. Given the higher accuracy of temperature and humidity sensors compared to air pollution LCSs [25], the propose method utilizes temperature and humidity measurements as meteorology training data $\mathbf{m}_i$, while disregarding the associated reading errors.

# 4 EVALUATION

## 4.1 Experimental Set-up

*4.1.1 Low-cost sensor.* The field experiment was conducted in Nanjing, China, from April 2022 to December 2022. In this experiment, the ETAR7002 Mobile Air Quality Monitor, developed by ETST Company Ltd, was chosen as the experimental LCS for calibration purpose [20]. The ETAR7002 integrates multiple components, including a particle sensor unit (PMS5003T model), a gas sensing module (Alphasense OX-B431 model for Ozone measurement), a communication module, a control module, and a power module. A total of 22 such LCSs were mounted on mobile vehicles (buses and city management vehicles) to collect data. Initially, these devices were positioned in close proximity to the Ruijin State-Control Station for rigorous performance testing from 26th April 2022 to 13th May 2022. Subsequently, these LCSs were deployed in the field to commence the experimental phase.

*4.1.2 Real reference station.* The experimental area comprises 12 city-controlled air monitoring stations, each equipped with high-cost devices. In this experiment, these stations serve as the reference stations. More specifically, three of these stations were utilized for calibrating the mobile LCSs, whereas the remaining nine stations were dedicated to testing the proposed method.

*4.1.3 Data preprocessing and test method.* The data are gridded into a spatial resolution of 0.2 km x 0.2 km with a frequency of 1-minute blocks. The values within each block are averaged, and the Three Sigma Criterion is applied to remove outliers [24]. Here, if two vehicles or stations occupy the same spatial-temporal block, they are considered to have a "rendezvous". In the experiment, a test is conducted when a mobile LCS rendezvouses with a test reference station. We ultimately obtained 6,364 pieces of data for testing.

*4.1.4 Performance metric.* In the experiments, the Root Mean Square Error (RMSE) is used to measure the calibration result's error [4]. RMSE is highly sensitive to outliers, making it well-suited for evaluating the performance of the calibration system.

*4.1.5 Implementation Details and Reproducibility.* The implementation is based on TensorFlow 2.4.0 framework [1] using Python 3.8. A neural network with 4 layers, 30 neurons per layer is used. The calibration function's updating time span is set to 2 hours, and a learning rate of 0.001 is used in Adam for neural networks' training. The values of $\gamma$ and $\epsilon$ are set to 0.995 and 0.01, respectively, in Eq. 7.

## 4.2 Overall Performance

This section presents an evaluation of the entire method, focusing on the assessment of the most common air pollutants: PM2.5, PM10, and O3. We consider the following approaches as baselines:

**Table 1: Overall Performance**

| | RMSE of Calibration Result ($\mu g/m^3$) | | |
|---|---|---|---|
| | PM2.5 | PM10 | O3 |
| NC | 11.49 | 16.62 | 17.26 |
| RC | 8.88 | 13.85 | 13.51 |
| MC | 7.20 | 11.31 | 10.58 |
| VRS-TS | 7.03 | 11.44 | 10.91 |
| **VRS-NN** | **5.93** | **10.25** | **9.46** |

- **No Calibration (NC)**: This approach involves utilizing the raw sensor readings directly as the post-calibrated results.
- **Rendezvous Calibration (RC)**: Only reference data collected when mobile LCS meet with the reference station are used. Neural networks are used for calibration function's fitting.
- **Multi-hop Calibration (MC)**: Building upon the RC approach, if a mobile LCS meets with another mobile LCS, the most recently calibrated LCS's reading is used as the reference to provide reference data to the earlier calibrated LCS.
- **Threshold Strategy (VRS-TS)**: In this approach, VRSs are created to assist in calibration. However, only data from the VRS that exceeds a fixed confidence threshold are used. The calibration function is trained using these selected data points with equal weighting. To differentiate this baseline from the proposed method, we refer to it as "-TS" and use the suffix "-NN" to represent the proposed method, which employs a confidence-weighted strategy to use neural networks fitting the calibration function.

Table 1 presents the average results obtained from each testing opportunity for the three types of air pollution concentrations. The following observations can be made: (i) "RC" shows only marginal improvement in data accuracy, primarily due to the limited number of rendezvous opportunities. (ii) "MC" leverages rendezvous opportunities more effectively, resulting in increased reference data. As a result, the final performance of the MC method exhibits improvement compared to RC, albeit still limited. (iii) The VRS-TS approach establishes VRSs but relies on a simplistic approach to process data confidence. Consequently, the performance achieved by VRS-TS is suboptimal. (iv) In contrast, the proposed VRS-NN method demonstrates superior performance across all three types of air pollution concentrations. This indicates the generalizability of the VRS-NN method for calibrating various air pollution concentration types.

## 4.3 Evaluation of Different VRS's Establishment Methods

The main focus of this paper is to construct VRSs for assisting in the collection of reference data. In this subsection, various establishment methods for the VRS will be compared.

The establishment of the VRS can be divided into two parts: the data inference module and the confidence assessment module. In this section, an additional experiment is conducted using only PM2.5 data from RRSs. The dataset is randomly divided into a training set comprising 70% of the data and a test set comprising 30% of the data. Subsequently, the training set's data is used to

**Table 2: Performance of VRS's data inference methods**

| | RMSE of Inference Result ($\mu g/m^3$) |
|---|---|
| PP | 7.91 |
| BI | 4.78 |
| **GPR-M** | **3.66** |

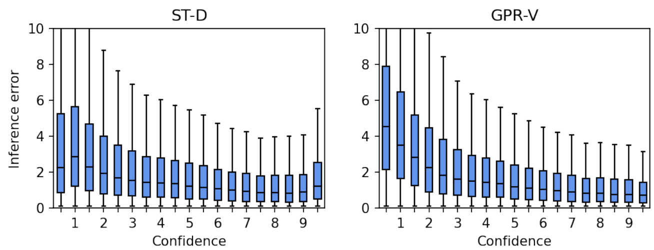


**Figure 2: Confidence and inference error's correlation compare of two confidence assessment methods for VRS.**

establish the VRS at the spatial-temporal locations of the test set's data. The VRS's inference results are evaluated using the ground truth from the test set, and the assessment of confidence is evaluated based on the test error.

*4.3.1 Different data inference methods for VRS.* In order to assess the effectiveness of the proposed GPR-M method in data inference, which utilizes the mean in the GPR result as the virtual monitoring value, two baselines are considered in this evaluation.

- **Proximity Principle (PP)**: Directly using the monitoring value of the closest reference station as the VRS's virtual monitoring value.
- **Bilinear Interpolation (BI)**: Reconstructing the field using the bilinear interpolation method and using its result as the VRS's virtual monitoring value.

Table 2 presents the final results, demonstrating that the GPR model exhibits the best performance with the lowest inference RMSE.

*4.3.2 Different confidence assessment methods for VRS.* The proposed GPR-V method leverages the inverse ratio of variance in the GPR result as the measure of confidence for the VRS. In this evaluation of confidence assessment methods, we use a commonly used method as the baseline:

- **Spatial-temporal Distance (ST-D)**: Using the inverse ratio of spatial-temporal distance between VRS and RRS as the confidence. This is a commonly used quality assessment method applied in many localization [27] and sensing systems [30]. Here, to ensure a clear comparison between these two methods, we introduce a multiplication factor of 10 for the ST-D baseline, aiming to align the scales.

In the comparison evaluation of confidence levels yielded by the two methods, the inferred data associated with confidence, ranging from 0 to 10, are divided into 19 bins. Each bin spans 0.5 units based on successive confidence. To ensure robust analysis, bins with fewer than 20 data points are excluded.

Figure 2 presents the box plots depicting the averaged inference errors and their respective variances for each confidence bin. It is important to note that, when compared to the baseline ST-D, the confidence measurement of GPR-V exhibits a stronger negative correlation with both the mean and variance of the test error within each bin. This observation suggests that higher confidence values obtained from GPR-V correspond to lower errors, indicating a more reliable inference. Conversely, lower confidence values in GPR-V may imply potential uncertainty, resulting in either high or low error values. In summary, these findings underscore the effectiveness of the proposed GPR-V method in delivering more accurate and reliable inference compared to the baseline ST-D approach.

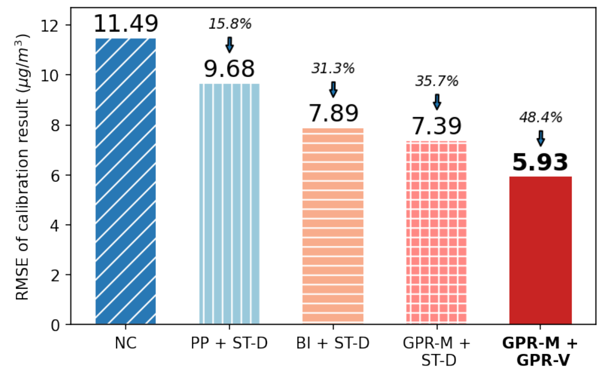


**Figure 3: Performance of different combinations of data inference and confidence assessment methods for VRS.**

*4.3.3 Different combinations.* At last, we revisit the initial mobile LCS calibration experiment to evaluate various combinations of data inference and confidence assessment methods. The results are depicted in Figure 3. It is evident that the proposed method combinations outperforms other combinations, yielding the most accurate calibration result.

## 5 CONCLUSION

This paper addresses the critical issue of post-deployment calibration for mobile LCSs used in urban air pollution monitoring. A novel approach is proposed, leveraging GPR and inferring data from established reference stations to construct VRS. The VRSs enhance the calibration dataset, enabling non-rendezvous calibration and improving LCS reading accuracy. A confidence assessment mechanism is introduced to evaluate the reliability of VRS data. Experimental studies in Nanjing, China validate the approach, showing a significant 25% improvement over other baselines.

In the future, we will extend the existing mobile device scheduling algorithm [9, 28] to the non-rendezvous calibration scenario. Efforts will also be directed towards the potential application of the calibration system to novel mobile platforms [7, 8].

## ACKNOWLEDGMENTS

This paper was supported by the National Key R&D program of China (2022YFC3300703), Guangdong Innovative and Entrepreneurial Research Team Program (2021ZT09L197), Shenzhen 2022 Stabilization Support Program (WDZC20220811103500001), and Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan (JC20220011).

The authors would like to thank Shenzhen Environmental Thinking Science and Technology (ETST) Company Ltd. for their assistance in system deployment and data collection.

# REFERENCES


[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: a system for large-scale machine learning.. In *Osdi*, Vol. 16. Savannah, GA, USA, 265–283.

[2] Lu Bai, Lin Huang, Zhenglu Wang, Qi Ying, Jun Zheng, Xiaowen Shi, Jianlin Hu, et al. 2020. Long-term field evaluation of low-cost particulate matter sensors in Nanjing. *Aerosol and Air Quality Research* 20, 2 (2020), 242–253.

[3] Thomas Becnel, Kyle Tingey, Jonathan Whitaker, Tofigh Sayahi, Katrina Lê, Pascal Goffin, Anthony Butterfield, Kerry Kelly, and Pierre-Emmanuel Gaillardon. 2019. A distributed low-cost pollution monitoring platform. *IEEE Internet of Things Journal* 6, 6 (2019), 10738–10748.

[4] Tianfeng Chai and Roland R Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development* 7, 3 (2014), 1247–1250.

[5] Xinlei Chen, Aveek Purohit, Carlos Ruiz Dominguez, Stefano Carpin, and Pei Zhang. 2015. DrunkWalk: Collaborative and Adaptive Planning for Navigation of Micro-Aerial Sensor Swarms. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems* (Seoul, South Korea) *(SenSys '15)*. Association for Computing Machinery, New York, NY, USA, 295–308. https://doi.org/10.1145/2809695.2809724

[6] Xinlei Chen, Aveek Purohit, Shijia Pan, Carlos Ruiz, Jun Han, Zheng Sun, Frank Mokaya, Patric Tague, and Pei Zhang. 2017. Design Experiences in Minimalistic Flying Sensor Node Platform through SensorFly. *ACM Trans. Sen. Netw.* 13, 4, Article 33 (nov 2017), 37 pages. https://doi.org/10.1145/3131779

[7] Xinlei Chen, Carlos Ruiz, Sihan Zeng, Liyao Gao, Aveek Purohit, Stefano Carpin, and Pei Zhang. 2020. H-DrunkWalk: Collaborative and Adaptive Navigation for Heterogeneous MAV Swarm. *ACM Trans. Sen. Netw.* 16, 2, Article 20 (apr 2020), 27 pages. https://doi.org/10.1145/3382094

[8] Xuecheng Chen, Haoyang Wang, Zuxin Li, Wenbo Ding, Fan Dang, Chengye Wu, and Xinlei Chen. 2023. DeliverSense: Efficient Delivery Drone Scheduling for Crowdsensing with Deep Reinforcement Learning. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(UbiComp/ISWC '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 403–408. https://doi.org/10.1145/3544793.3560412

[9] Xinlei Chen, Susu Xu, Jun Han, Haohao Fu, Xidong Pi, Carlee Joe-Wong, Yong Li, Lin Zhang, Hae Young Noh, and Pei Zhang. 2020. PAS: Prediction-Based Actuation System for City-Scale Ridesharing Vehicular Mobile Crowdsensing. *IEEE Internet of Things Journal* 7, 5 (2020), 3719–3734. https://doi.org/10.1109/JIOT.2020.2968375

[10] Xinlei Chen, Susu Xu, Xinyu Liu, Xiangxiang Xu, Hae Young Noh, Lin Zhang, and Pei Zhang. 2020. Adaptive hybrid model-enabled sensing system (HMSS) for mobile fine-grained air pollution estimation. *IEEE Transactions on Mobile Computing* (2020).

[11] Xinlei Chen, Xiangxiang Xu, Xinyu Liu, Hae Young Noh, Lin Zhang, and Pei Zhang. 2016. Hap: Fine-grained dynamic air pollution map reconstruction by hybrid adaptive particle filter. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 336–337.

[12] Yun Cheng, Xiucheng Li, Zhijun Li, Shouxu Jiang, Yilong Li, Ji Jia, and Xiaofan Jiang. 2014. AirCloud: A cloud-based air-quality monitoring system for everyone. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. 251–265.

[13] Yun Cheng, Zimu Zhou, and Lothar Thiele. 2022. iSpray: Reducing Urban Air Pollution with Intelligent Water Spraying. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–29.

[14] H Chojer, PTBS Branco, FG Martins, MCM Alvim-Ferraz, and SIV Sousa. 2020. Development of low-cost indoor air quality monitoring devices: Recent advancements. *Science of The Total Environment* 727 (2020), 138385.

[15] Francesco Concas, Julien Mineraud, Eemil Lagerspetz, Samu Varjonen, Xiaoli Liu, Kai Puolamäki, Petteri Nurmi, and Sasu Tarkoma. 2021. Low-Cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey and Critical Analysis. *ACM Trans. Sen. Netw.* 17, 2, Article 20 (may 2021), 44 pages. https://doi.org/10.1145/3446005

[16] Kaibo Fu, Wei Ren, and Wei Dong. 2017. Multihop calibration for mobile sensing: K-hop calibratability and reference sensor deployment. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 1–9.

[17] David Hasenfratz, Olga Saukh, and Lothar Thiele. 2012. On-the-fly calibration of low-cost gas sensors. In *Wireless Sensor Networks: 9th European Conference, EWSN 2012, Trento, Italy, February 15-17, 2012. Proceedings 9*. Springer, 228–244.

[18] Yidan Hu, Guojun Dai, Jin Fan, Yifan Wu, and Hua Zhang. 2016. BlueAer: A fine-grained urban PM2. 5 3D monitoring system using mobile sensing. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.

[19] Xinyu Liu, Xinlei Chen, Xiangxiang Xu, Enhan Mai, Hae Young Noh, Pei Zhang, and Lin Zhang. 2017. Delay effect in mobile sensing system for urban air pollution monitoring. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. 1–2.

[20] Xinyu Liu, Xiangxiang Xu, Xinlei Chen, Enhan Mai, Hae Young Noh, Pei Zhang, and Lin Zhang. 2017. Individualized calibration of industrial-grade gas sensors in air quality sensing system. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. 1–2.

[21] Balz Maag, Zimu Zhou, Olga Saukh, and Lothar Thiele. 2017. SCAN: Multi-hop calibration for mobile sensor arrays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–21.

[22] Mannam Veera Narayana, Devendra Jalihal, and S. M. Shiva Nagendra. 2022. Establishing A Sustainable Low-Cost Air Quality Monitoring Setup: A Survey of the State-of-the-Art. *Sensors* 22, 1 (2022). https://doi.org/10.3390/s22010394

[23] SGOPAL Patro and Kishore Kumar Sahu. 2015. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462* (2015).

[24] Friedrich Pukelsheim. 1994. The three sigma rule. *The American Statistician* 48, 2 (1994), 88–91.

[25] Abdul Samad, Daniel Ricardo Obando Nuñez, Grecia Carolina Solis Castillo, Bernd Laquai, and Ulrich Vogt. 2020. Effect of relative humidity and air temperature on the results obtained from low-cost gas sensors for ambient air quality measurements. *Sensors* 20, 18 (2020), 5175.

[26] Olga Saukh, David Hasenfratz, Christoph Walser, and Lothar Thiele. 2014. On rendezvous in mobile sensing networks. In *Real-World Wireless Sensor Networks: Proceedings of the 5th International Workshop, REALWSN 2013, Como (Italy), September 19-20, 2013*. Springer, 29–42.

[27] Haoyang Wang, Xuecheng Chen, Yuhan Cheng, Chenye Wu, Fan Dang, and Xinlei Chen. 2022. H-SwarmLoc: Efficient Scheduling for Localization of Heterogeneous MAV Swarm with Deep Reinforcement Learning. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 1148–1154.

[28] Susu Xu, Xinlei Chen, Xidong Pi, Carlee Joe-Wong, Pei Zhang, and Hae Young Noh. 2020. iLOCuS: Incentivizing Vehicle Mobility to Optimize Sensing Distribution in Crowd Sensing. *IEEE Transactions on Mobile Computing* 19, 8 (2020), 1831–1847. https://doi.org/10.1109/TMC.2019.2915838

[29] Xiangxiang Xu, Xinlei Chen, Xinyu Liu, Hae Young Noh, Pei Zhang, and Lin Zhang. 2016. Gotcha ii: Deployment of a vehicle-based environmental sensing system. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 376–377.

[30] Chengzhao Yu, Ji Luo, Rongye Shi, Xinyu Liu, Fan Dang, and Xinlei Chen. 2022. ST-ICM: spatial-temporal inference calibration model for low cost fine-grained mobile sensing. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 910–912.

[31] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)* 23, 4 (1997), 550–560.