Intelligent Air Pollution Sensors Calibration for Extreme Events and Drifts Monitoring

Martha Arbayani Zaidan, *Member, IEEE*, Naser Hossein Motlagh, Pak Lun Fung, Abedalaziz S. Khalaf, Yutaka Matsumi, Aijun Ding, Sasu Tarkoma, *Senior Member, IEEE*, Tuukka Petäjä, Markku Kulmala, and Tareq Hussein

***Abstract*—Air quality low-cost sensors (LCSs) are affordable and can be deployed in massive scale in order to enable high-resolution spatio-temporal air pollution information. However, they often suffer from sensing accuracy, in particular, when they are used for capturing extreme events. We propose an intelligent sensors calibration method that facilitates correcting LCSs measurements accurately and detecting the calibrators' drift. The proposed calibration method uses Bayesian framework to establish white-box and black-box calibrators. We evaluate the method in a controlled experiment under different types of smoking events. The calibration results show that the method accurately estimates the aerosol mass concentration during the smoking events. We show that black-box calibrators are more accurate than white-box calibrators. However, black-box calibrators may drift easily when a new smoking event occurs, while white-box calibrators remain robust. Therefore, we implement both of the calibrators in parallel to extract both calibrators' strengths and also enable drifting monitoring for calibration models. We also discuss that our method is implementable for other types of LCSs suffered from sensing accuracy.**

***Index Terms*—Air quality, Bayesian calibrator, drift monitoring, extreme event, indoor low-cost sensor (LCS).**

I. Introduction

INDOOR air quality has a direct impact on overall human health and significantly affects human work productivity. Based on the United States Environmental Protection Agency (EPA),[1] humans spend about 80%–90% of their time indoors. The levels of indoor air pollution are also often two to five times higher than outdoor levels. In some cases, the pollution levels might exceed 100 times than outdoor levels for the same pollutants. Indeed, excessive levels of indoor air pollutants would lead to immediate harmful effects. For example, incidental propane leaks in industrial plants [1] or excessive carbon monoxide (CO) in vehicles [2] would cause sudden death.

According to World Health Organization (WHO),[2] particulate matter (PM) is a common indicator for air pollution, which is more harmful in affecting human health than any other pollutants. PM indoors can be originated from outdoor origins or generated through human activities, such as cooking, burning candles, using kerosene heaters and smoking. Therefore, accurate indoor air quality measurement enables estimating health and safety risks in work and living environments.

Manuscript received 19 November 2021; revised 17 January 2022; accepted 4 February 2022. Date of publication 15 February 2022; date of current version 13 December 2022. This work was supported in part by the MegaSense and Nokia Center for Advanced Research (NCAR), in part by the Deanship of Scientific Research (DSR), University of Jordan, under Grant 2361, in part by the Scientific Research Support Fund (SRF), in part by the Jordanian Ministry of Higher Education under Grant WE-2-2-2017, in part by the Academy of Finland Center of Excellence under Grant 272041, in part by the Atmosphere and Climate Competence Center (ACCC) under Grant 337549, in part by Healthy Outdoor Premises for Everyone project under Grant UIA03-240, in part by the Academy of Finland Projects under Grant 324576, Grant 345008, and Grant 335934, and in part by the Technology Industries of Finland Centennial Foundation to Urban Air Quality 2.0 Project. The work of Tareq Hussein and Markku Kulmala was supported by the Eastern Mediterranean and Middle East–Climate and Atmosphere Research (EMME-CARE) Project, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant 856612 and the Government of Cyprus. Paper no. TII-21-5125. *(Corresponding authors: Martha Arbayani Zaidan; Tareq Hussein.)*

Martha Arbayani Zaidan, Tuukka Petäjä, and Markku Kulmala are with the Joint International Research Laboratory of Atmospheric and Earth System Sciences, Nanjing University, Nanjing 210093, China, and also with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, 00014 Helsinki, Finland (e-mail: martha.zaidan@helsinki.fi; tuukka.petaja@helsinki.fi; markku.kulmala@helsinki.fi).

Naser Hossein Motlagh and Sasu Tarkoma are with the Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland (e-mail: naser.motlagh@helsinki.fi; sasu.tarkoma@helsinki.fi).

Pak Lun Fung is with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, 00014 Helsinki, Finland (e-mail: pak.fung@helsinki.fi).

Abedalaziz S. Khalaf is with the Department of Physics, School of Science, University of Jordan, Amman 11946, Jordan (e-mail: abdulazizkhalaf@outlook.com).

Yutaka Matsumi is with the Institute for Space-Earth Environmental Research (ISEE), Nagoya University, Nagoya 464-8601, Japan (e-mail: matsumi@nagoya-u.jp).

Aijun Ding is with the Joint International Research Laboratory of Atmospheric and Earth System Sciences, Nanjing University, Nanjing 210093, China (e-mail: dingaj@nju.edu.cn).

Tareq Hussein is with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, 00014 Helsinki, Finland, and also with the Department of Physics, School of Science, University of Jordan, Amman 11946, Jordan (e-mail: tareq.hussein@helsinki.fi).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2022.3151782.

Digital Object Identifier 10.1109/TII.2022.3151782

[1][Online]. Available: https://www.epa.gov/report-environment/indoor-air-quality/

[2][Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

However, air quality in different rooms and spaces of a building varies from one to another. This may require installing multiple sensors indoors within different rooms. Fortunately, low-cost sensors (LCSs) can be utilized for such purposes [3]. LCS can then alert when excessive pollutants have reached a particular health threshold. Indeed, LCSs are affordable and relatively easy to install that can then be massively deployed in buildings [4].

Although LCSs are usually laboratory calibrated, they often suffer from low accuracy and low robustness when they are deployed in fields [5]. These issues usually occur due to sensor designs [6], sensor drifts, changes in environmental conditions, background changes, and fabrication variances [7]. For example, LCSs generally do not include a heater or dryer at their inlets, so the changes in temperature and relative humidity have a significant impact on the performance of low-cost PM sensors [8]. As a result, LCSs often are vulnerable to accurately measure air pollutants at very low and very high concentration levels [9], [10]. Fortunately, to overcome the challenges of LCSs measurement accuracy and robustness, many studies propose various solutions in terms of sensor deployment and sensor calibrations as presented in review studies in [5] and [11]. However, thanks to the advancement of computing technologies, data-driven, and machine-learning (ML) based approaches have recently emerged as a potential solution for these challenges [5], [12].

The state-of-the-art of indoor LCSs was reviewed comprehensively in [13] and [14]. Based on these studies, there is an immense need for performing research on indoor LCS-based measurements and calibrations. These studies highlight that most research activities for the indoor environments have been focused only on the sensors' data analytics (e.g., more than 60% from their reviewed papers), neglecting the evaluation of sensors' performance indoors through developing calibration methods. Another concern when deploying LCSs relates to the sensor drifts and calibrator drifts (also known as concept drift). While sensor drift indicates the aging of the sensors hardware overtime [15], that makes the reading of the sensor to deviate from the actual readings [16]. The calibrator drift refers to the situation where the performance of calibration models reduce due to the changes in environmental conditions [17]. To the best of our knowledge, none of the papers reviewed in the aforementioned articles propose a method that combines sensors calibration and drift detection together, especially for indoor environments, where reference instruments are usually not accessible nor remote sensing can penetrate indoors for sensors validation.

In this article, we contribute by proposing a novel sensor calibration method and a calibrators' drift detection method, which are evaluated in an indoor environment. The novelties of our study include: 1) performing controlled experiments to define scenarios for indoor extreme events (presented in Sections II and III), 2) deploying white-box and black-box calibrators in parallel for correcting LCSs measurements and detecting calibrators' drift (explained in Sections IV and V), and 3) discussing potential industrial applications extended from the proposed methods (discussed in Section VI-C).

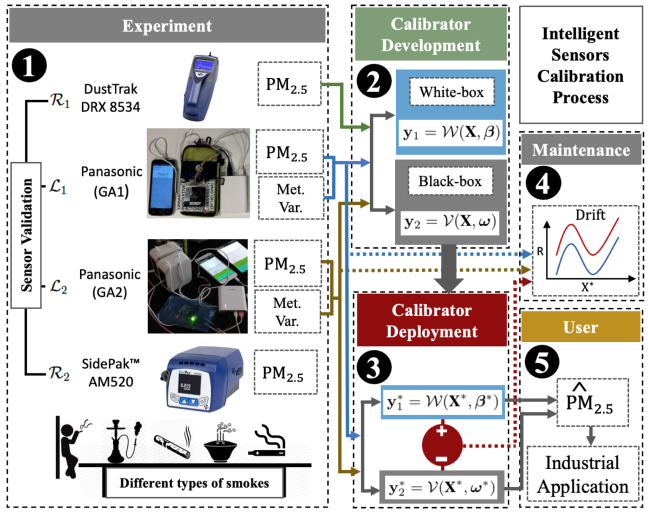

Fig. 1. Intelligent air pollution sensors calibration process.

II. Experiment: Indoor Pollutant Measurements

In the experiments, we use two types of reference instruments ($\mathcal{R}$) and two generations of LCSs also labeled as $\mathcal{L}$, where $\mathcal{R}$ refers to any high precision sensing instruments such as DustTrak and SidePak (as shown in Fig. 1). The measurements of $\mathcal{R}$ can be used as ground truth data for sensors calibration and validation purposes. In addition, LCSs are known to be affordable devices (i.e., the cost less than $2500 per unit [18]), which have evolved as efficient solutions for sensing indoor and outdoor air pollution monitoring [3]. In this study, the LCSs generation indicates the improvements on the LCSs' hardware and software (i.e., different LCS version). Both $\mathcal{R}$ and LCSs used in this study are shown in Fig. 1, part ❶, with labels $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{L}_1$, and $\mathcal{L}_2$.

A. Reference Instruments

DustTrak DRX 8534 (TSI Inc.), labeled as $\mathcal{R}_1$, is capable of simultaneously measuring size-segregated mass fraction aerosol concentrations in the range from 0.001 to 150 mg/m^3, corresponding to PM $_1$, $PM_{2.5}$, PM_4 (Respirable), PM_{10}, and total PM size fractions. Therefore, the instrument can measure contaminants such as dusts, smoke, fumes, and mists. The sensing technology of the instrument is based on light-scattering laser photometers. The instrument is battery operated, where data-logging can be done between $-20\,^{\circ}$C and 60 $^{\circ}$C with an operational humidity between 0% and 95%.

SidePak$^{\text{TM}}$ Personal Aerosol Monitor AM520 (TSI Inc.), labeled as $\mathcal{R}_2$, is capable of measuring aerosol mass concentrations in the range from 0.001 to 100 mg/m^3, corresponding to PM $_1$, $PM_{2.5}$, PM_4 (Respirable), PM_5 (China Respirable), PM_{10}, and 0.8 μm diesel particulate matter (DPM). Thus, the instrument provides real-time aerosol mass concentration readings of dusts, fumes, mists, smoke, and fog. The instrument is portable and battery operated. The sensing technology of the instrument is based on light-scattering laser photometers. It can also operate between $-20\,^{\circ}$C and 60 $^{\circ}$C with an operational humidity between 0 and 95%.

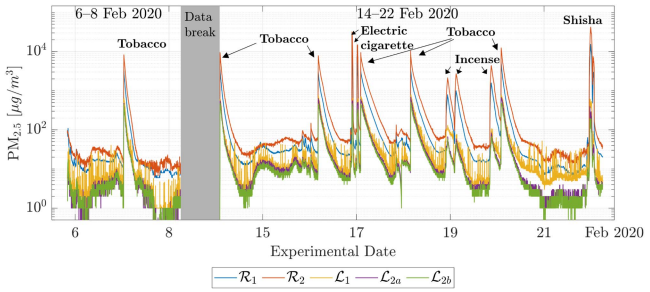

Fig. 2. Time-series data of $PM_{2.5}$ concentration obtained in the experiment.

B. Low-Cost Sensors (LCSs)

In Fig. 1, the LCS units refer to *sensor generation I* (labeled as $\mathcal{L}_1$) and *sensor generation II* (labeled as $\mathcal{L}_2$). The devices measure the mass concentration of PM with diameter smaller than 2.5 μm ($PM_{2.5}$). The thermal resistor in the sensor stimulates flow induced by temperature gradient. The sensor devices have an air inlet, a light sensor, and an infrared light source. They start measuring when air enters the sensor's air inlet, then the light source concentrates on sensing point. These sensor devices utilize light-scattering particle (LSP) sensing utilities for monitoring $PM_{2.5}$. LSP sensors are well-known low-cost solutions for particle concentrations measurements and monitoring. These portable sensor devices are utilized to perform real-time and spatial $PM_{2.5}$ measurements and monitoring [19]. In addition to the features of $\mathcal{L}_1$, *sensor generation II* (i.e., $\mathcal{L}_2$) is equipped with a case to reduce the effect of air turbulence in the inlet. Sensor $\mathcal{L}_2$ is also equipped with meteorological sensor utilities, including relative humidity (RH), temperature (Temp), and pressure (P). Moreover, an algorithm is embedded in $\mathcal{L}_2$ to filter the raw measured data such that it removes the spikes before data recording and monitoring.

C. Experiment

We carried out the experiments in two different time intervals. The first measurement was performed continuously between 6 and 8 Feb 2020, and the second measurement was performed between 14 and 22 Feb 2020. During the measurements, $\mathcal{R}_1$ and $\mathcal{R}_2$ were placed side by side with the LCSs, i.e., one unit of $\mathcal{L}_1$ and two units of $\mathcal{L}_2$ ($\mathcal{L}_{2a}$ and $\mathcal{L}_{2b}$), in a confined space, i.e., a room where the ventilation system was sealed off. The experimental setup is illustrated in Fig. 1, part ❶. The inlets of all instruments were placed exactly next to each other to ensure they extract the same amount of aerosol mass concentrations. Four types of smokes were generated using tobacco, electric cigarette, incense, and shisha, in which the measurements are depicted in Fig. 2. There were in total 12 experimental events for smoke measurements. Tobacco were smoked at events numbers 1, 2, 3, 6, 7, and 11; electric cigarette were smoked at events numbers 4 and 5; incense was lighted at events numbers 8, 9, and 10; and shisha was blown at event number 12. The experimental events were held by blowing the smoke next to the inlets of the experimental setup. During the experiment, we continuously recorded the measurements of $PM_{2.5}$ concentration and meteorological data, including Temp, RH, and P from all instruments.

TABLE I
LCS METEOROLOGICAL SENSORS: CONSISTENCY PERFORMANCE

Variables	Measurements		Validation metrics			
	Mean	Range	MAE	RMSE	MAPE	R
Temp ($^\circ C$)	25.22	20 - 30	0.80	0.83	3.24	0.995
RH (%)	29.25	20 - 40	0.77	0.85	2.61	0.990
P (mbar)	902.27	900 - 910	0.99	0.99	0.11	0.999

D. The Data

1) Data Preprocessing: The collected data from instruments and LCSs have different time resolution by default, thus, the data needs to be synchronized. The time resolution of $\mathcal{L}_1$ varies between 40 s and 1 min interval, whereas $\mathcal{L}_2$ has a fixed time-resolution at 1 min interval. Both $\mathcal{R}_1$ and $\mathcal{R}_2$ have a consistent measurement interval of 1 min. Hence, for our data analysis, we aggregate the data to be in 1 min resolution. Note that there is an experimental gap between 8 and 14 Feb 2020 (about a week) .

2) Smoking Events Characteristics: In this article, the whole experiment comprises the smoke and normal events. The median of $PM_{2.5}$ concentration for the whole experiment is 27.2 $\mu g/m^3$. The normal event is usually assumed if the $PM_{2.5}$ concentration is below this median level. However, as shown in Fig. 2, the experiment shows that the smoke does not dissipate quickly, since the ventilation system is OFF. In addition, before the $PM_{2.5}$ concentrations reach the median level, again another smoking event takes place. Therefore, we assume that the smoking event happens when the $PM_{2.5}$ concentrations crosses the 75% quantiles that is at 144.76 $\mu g/m^3$. Indeed, as shown in Fig. 2, the experiment highlights the gap between the measurements of $\mathcal{R}$ and LCSs, indicating that LCSs suffer from measurement accuracy that is the main concern in this article. Hence, to validate the measurements of LCSs, we use data collected from DustTrak ($\mathcal{R}_1$) as the ground truth data. The instrument performance has been approved in many scientific experiments [20].

3) Performance Metrics: We use performance metrics of Pearson correlation coefficient (R), mean absolute error (MAE), and mean absolute percentage error (MAPE), and root mean squared error (RMSE) for sensors and methods validation. The metrics are described in Appendix A.

III. SENSORS PERFORMANCE

In this section, we perform sensors validation using *consistency* and *accuracy* tests to evaluate the performance of sensors (as shown in Fig. 1, part ❶), whereas the term *consistency* refers to similarity in measurements of two LCS, the term *accuracy* indicates how similar are the measurement of LCS units with the measurement of a reference instrument.

A. Meteorological Variables: Consistency Test

The $\mathcal{L}_2$ is already equipped with meteorological sensors measuring variables Temp, RH, and P. To show the performance of

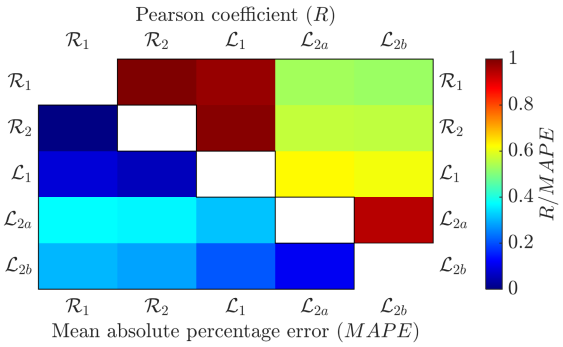

Fig. 3. Heatmap plot between reference instruments and LCSs.

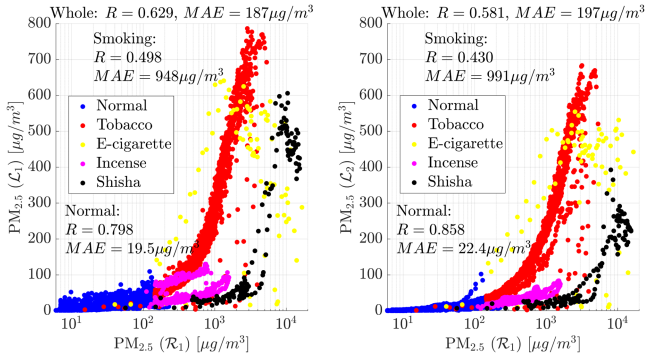

Fig. 4. $PM_{2.5}$ scatter plots between $\mathcal{R}_1$ and $\mathcal{L}_1$ (left) and $\mathcal{L}_{2a}$ (right).

LCSs and how consistent the measurements of meteorological variables are, we perform *consistency* test between $\mathcal{L}_{2a}$ and $\mathcal{L}_{2b}$ using the metrics of R, MAPE, MAE, and RMSE. The *consistency* test results are shown in Table I.

These results show that the meteorological measurements are almost identical and demonstrate consistent performance when they are compared between each other. In Table I, from the *measurements* column, the range and mean values indicate that the sensor readings are reliable. The Temp measurement ranges between 20 and 30 °C, with the mean value of 25.22 °C; RH ranges between 20% and 40%, with the mean value of 29.25%; and *P* ranges between 900 and 910 mbar with the mean value of around 902.27 mbar. These values show typical room conditions, where slight variations take place due to human activities. The performance of LCSs is clearly shown in *validation metrics* column with values of R, with all values approximately equal to 0.99 for all variables. Likewise, MAPE values are very low that is below 3.5% for all sensors. For example, in case of Temp, 3.24% MAPE for the mean Temp value of 25.22 °C can be considered to be relatively small. Similarly, all MAE and RMSE values for all meteorological variables are below 1, indicating that errors between two LCSs are so small that can be considered negligible.

B. Aerosol Sensors: Consistency and Accuracy Tests

We validate aerosol LCS measurements using the reference instruments. This validation is known as *accuracy test*, whereas the comparisons between the same type of devices are known as *consistency test*. Fig. 3 shows sensors validation heatmap matrix plot between reference instruments ($\mathcal{R}_1$ and $\mathcal{R}_2$) and LCSs ($\mathcal{L}_1$, $\mathcal{L}_{2a}$, and $\mathcal{L}_{2b}$). The figure consists of two performance metrics: the lower part illustrates MAPE, whereas the upper part shows Pearson correlation coefficient (R). The colors represent the level of R and MAPE values. When the color is closer to dark red, R between two devices is strong and MAPE is low. Inversely, when the color is closer to dark blue, R between two devices is low and MAPE is high.

The *consistency tests* between reference instruments show high correlation (i.e., high R value and small MAPE value). This explains that both reference instruments provide similar performance, and hence either of them can be used as ground truth. In addition, since the performance of $\mathcal{R}_1$ has been approved in many scientific experiments [20], thus, we select $\mathcal{R}_1$ as the ground truth sensing instrument for validating sensors and developing calibrators. Likewise, the *consistency tests* between both second generations of LCSs demonstrate high R correlation and very low MAPE value. This indicates that they are identical in terms of electronics and consistent in terms of performance. However, $\mathcal{L}_1$ and $\mathcal{L}_2$ have a minor performance difference (i.e., negligible) in terms of R and MAPE, allowing us to apply the same types of calibrators for the two generations of LCSs. The *accuracy test* between LCSs and the reference instruments shows that the correlation coefficients (R) are low at approximately about 0.6 (i.e., yellow color indicator), while their MAPE values are around 0.4 (light blue). These facts translate that LCSs do not meet the performance of reference instruments.

Fig. 4 shows scatter plots of $PM_{2.5}$ between $\mathcal{R}_1$ and $\mathcal{L}_1$ and $\mathcal{L}_{2a}$. The scatter plot of $\mathcal{L}_{2b}$ is not shown in the figure, as it would demonstrate similar pattern. In the figure, the normal event is illustrated by blue color, whereas the smoking events are shown by other colors. Each color shows a different deviation path that interestingly forms a cluster for each type of smoke. It can be seen that the relationship between $\mathcal{R}_1$ and LCSs is correlated nonlinearly for the concentration distribution within each smoke type (cluster). The figure also presents the values of R and MAE for *whole*, *normal,* and *smoking* event scenarios. During *normal event*, R values for both LCSs are still high (≈ 0.8) and MAE values are low (< 23 $\mu g/m^3$). These results explain that the performance of LCSs is similar to the reference instrument in normal conditions. However, during *smoking events*, the measurement error between the reference instrument and LCSs become larger as $PM_{2.5}$ concentration increases (such that $R <$ 0.5 and MAE > 900 $\mu g/m^3$).

In practice, since LCSs are incapable of measuring high levels of $PM_{2.5}$ concentrations and extreme events; thus, relying on their measurements for these smoking events would be harmful. As a result, to improve LCSs' $PM_{2.5}$ measurement, they need to be calibrated. In next section, we explain our proposed sensors calibration method.

IV. Sensors Calibration

A. Calibration Process

In Fig. 1 (part ❷), we illustrate the development of sensor calibrators, where it consists of two calibrator models, called

white-box ($\mathbb{W}$) and black-box ($\mathbb{V}$) calibrators. In general, there are two approaches for developing $\mathbb{W}$. The first approach relies on physics-based models, and the second approach uses statistical models, where the relationship between the inputs and the outputs are visible and transparent [20]. Therefore, white-box calibrator ($\mathbb{W}$) is usually suitable for modeling a calibrator if the measurements of LCSs and reference instruments exhibit regular patterns. For example, in our case as illustrated in Fig. 4, the relationship between reference instrument and LCSs presents exponential shapes. The black-box calibrator $\mathbb{V}$ provides little explanatory insight into the relative influence of the independent variables (e.g., inputs variables) in the prediction process (e.g., outputs), but they are often effective in dealing with air quality and environmental data, which are nonlinear [21]. For example, neural-networks are known as a general approximator that can relatively well deal with most nonlinear problems, such as sensors calibration and virtual sensors [9].

Both calibrators ($\mathbb{W}$ and $\mathbb{V}$) are then trained independently using the datasets obtained from the experiments (Fig. 1, part ❶). Even though, our sensors calibration process (see Fig. 1) allows flexibility in terms of models choice for $\mathbb{V}$ and $\mathbb{W}$. In our study, we select a Bayesian linear model (BLM) as $\mathbb{W}_2$ and a Bayesian neural-network (BNN) as $\mathbb{V}_2$. We select Bayesian framework, because, first, Bayesian models are robust from overfitting due to the presence of regularization. Second, Bayesian inference leads to probability distributions in their model coefficients and predictive distribution, which enables analyzing them statistically [20]. For comparison of $\mathbb{W}_2$ and $\mathbb{V}_2$, we also redevelop the most popular calibration methods as mentioned in [5]. These calibration methods include multivariate linear regression (MLR) and artificial neural-network (ANN) representing white-box ($\mathbb{W}_1$) and black-box ($\mathbb{V}_1$) models, respectively.

Next, we deploy both trained calibrators ($\mathbb{W}_2$ and $\mathbb{V}_2$) in parallel to ensure that they complement the strengths and weaknesses of each others (see Fig. 1, part ❸). We further compute the residual [see (3): $\mathbb{R}$] between $\mathbb{V}_2$ and $\mathbb{W}_2$ to monitor the calibrators drift (see Fig. 1, part ❹). Finally, the outputs from the calibrators provide accurate $PM_{2.5}$ concentration information for users (see Fig. 1, part ❺). In addition, as described in Section VI-C various industrial applications can benefit from the calibrated $PM_{2.5}$ measurements.

B. Calibration Models

In the calibrator development phase (see Fig. 1, part ❷), the $\mathbb{W}$ and $\mathbb{V}$ calibrators can be expressed mathematically as

$$\mathbf{y}_1 = \mathcal{W}(\mathbf{X}, \boldsymbol{\beta}) + \varepsilon_1 \tag{1}$$

$$\mathbf{y}_2 = \mathcal{V}(\mathbf{X}, \boldsymbol{\omega}) + \varepsilon_2 \tag{2}$$

where $\mathcal{W}$ and $\mathcal{V}$ are white-box and black-box calibration functions, respectively; and $\mathbf{y}_1$ and $\mathbf{y}_2$ are the outputs of calibrators $\mathbb{W}$ and $\mathbb{V}$, respectively. It is worth noting that $\mathbf{y}_1$ and $\mathbf{y}_2$ are the calibration outputs during the training process. The symbol $\boldsymbol{\beta}$ represents the model coefficients of $\mathcal{W}$ and the symbol $\boldsymbol{\omega}$ embodies the weights of $\mathcal{V}$. In both calibrators, ε refers to errors that follow a Gaussian distribution with zero mean and σ^2 noise variance, given by $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. The inputs $\mathbf{X}$ for both calibrators are obtained from the LCS measurements, including $PM_{2.5}$ concentration and meteorological variables. As described in Section IV-A, the calibrator functions of $\mathbb{W}_2$ and $\mathbb{V}_2$ are selected to be a BLM and a BNN, respectively. Therefore, the optimization of models' coefficients is then performed using Bayesian inference. In the calibrator deployment phase (see Fig. 1, part ❸), $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ are the calibrators' outputs during the testing process, which are in the form of Gaussian predictive distribution symbolized by $p(\mathbf{y}_1^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}_1)$ and $p(\mathbf{y}_2^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}_2)$, for $\mathbb{W}_2$ and $\mathbb{V}_2$, respectively. In both calibrators, symbols $\mathbf{X}^*$ are the test data obtained from LCS measurements. The derivation of both calibrators is described in [20] and also briefly presented in Appendices B and C.

C. Drift Monitoring Methods

In real deployment, due to various hardware and environmental reasons, the calibration models would become less effective throughout the time. In this article, we call this phenomenon as *calibrator drift* and we propose two methods for monitoring the calibrators drift (see Fig. 1, part ❹) including: 1) monitoring the outputs of calibrators' residual between $\mathbb{W}$ and $\mathbb{V}$, and 2) monitoring one of the key variables, which may affect the calibrators' effectiveness.

The first method computes the predictive distribution of calibrator residual ($\mathbb{R}$) between two deployed calibrators ($\mathbb{W}$ and $\mathbb{V}$), shown as the red dashed lines in Fig. 1. In our case, since the predictive distributions for $\mathbb{W}_2$ and $\mathbb{V}_2$ are in the form of a Gaussian distribution (as explained in the Section IV-B), thus, the drift monitoring residual ($\mathbb{R}$) results in Gaussian distribution as

$$\mathbb{R} \sim \mathcal{N}(\mu^*_{\mathbf{y}_2} - \mu^*_{\mathbf{y}_1}, \Sigma^*_{\mathbf{y}_2} + \Sigma^*_{\mathbf{y}_1}) \tag{3}$$

where the notations of $\mu^*_{\mathbf{y}_2}$ and $\mu^*_{\mathbf{y}_1}$ represent the mean of predictive Gaussian distributions for $\mathbb{V}_2$ and $\mathbb{W}_2$, respectively, whereas the notations of $\Sigma^*_{\mathbf{y}_2}$ and $\Sigma^*_{\mathbf{y}_1}$ denote the variance of predictive Gaussian distributions for $\mathbb{V}_2$ and $\mathbb{W}_2$, respectively. The derivation is described in Appendix D.

The second method enables drift detection by complementing the first method through monitoring the updates of one of the key variables measured by LCSs. This is shown as the blue and brown dashed lines in Fig. 1. Due to the simplicity and transparency of the model, the key variables affecting calibrators can be identified by analyzing the model coefficients of the calibrator $\mathbb{W}$. For example, in our case $PM_{2.5}$ is the key variable affecting the calibration. If the calibrators were trained with *normal event*, then the calibrators may drift when LCSs are deployed on smoking events. Let us recall that *normal event* refers to scenarios, where there is no smoking and generally the $PM_{2.5}$ concentration is considered to be low, while smoking event indicates to the scenarios, where LCSs measuring $PM_{2.5}$ concentration is high.

To enable the drift detection, an outlier limit (L) can be computed by calculating the upper limit of quantile (q) from the training data. For example, the outlier limit (L) can be set by computing the qth quantile of the training data of $PM_{2.5}$

Group	Scenario	Training	Testing
$\mathbf{G}_1$	$\mathcal{S}_1$	Random 70%	Random 30%
$\mathbf{G}_2$	$\mathcal{S}_2$	$\mathcal{L}_{2a}$	$\mathcal{L}_{2b}$
	$\mathcal{S}_3$	$\mathcal{L}_{2b}$	$\mathcal{L}_{2a}$
$\mathbf{G}_3$	$\mathcal{S}_4$	$\mathcal{L}_{2a}$	$\mathcal{L}_1$
	$\mathcal{S}_5$	$\mathcal{L}_{2b}$	$\mathcal{L}_1$
	$\mathcal{S}_6$	$\mathcal{L}_1$	$\mathcal{L}_{2a}$
	$\mathcal{S}_7$	$\mathcal{L}_1$	$\mathcal{L}_{2b}$
$\mathbf{G}_4$	$\mathcal{S}_8$	$\mathcal{L}_{2a}$ except Tobacco	$\mathcal{L}_{2b}$
	$\mathcal{S}_9$	$\mathcal{L}_{2a}$ except E-cigarrete	$\mathcal{L}_{2b}$
	$\mathcal{S}_{10}$	$\mathcal{L}_{2a}$ except Incense	$\mathcal{L}_{2b}$
	$\mathcal{S}_{11}$	$\mathcal{L}_{2a}$ except Shisha	$\mathcal{L}_{2b}$
$\mathbf{G}_5$	$\mathcal{S}_{12}$	Normal events	Smoking events

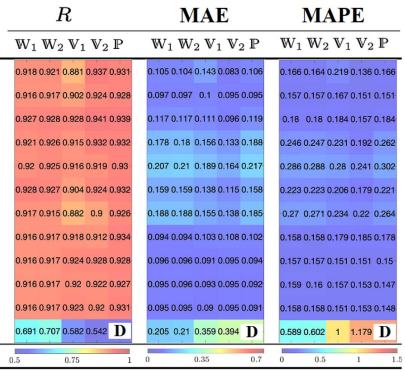

Fig. 5. Calibration results for different scenarios. Note: one alternative solution is that we put the numbers as tables and put the alternative colors behind the numbers. The letter **D** in $\mathbf{G}_5$ indicates to the detection of drift.

concentrations obtained from LCSs ($\mathbf{X}_{\mathrm{PM}_{2.5}}$). Whenever new $\mathrm{PM}_{2.5}$ measurement ($\mathbf{X}^*_{\mathrm{PM}_{2.5}}$) is bigger than L, this is considered as outlier in the test data. Indeed, the outlier in test data is one of the indicators for drift occurrence. To this end, the number of outliers in test data needs to be counted. We show this counting with $\mathcal{C}$. Finally, the accepted percentage of outlier $\mathbf{X}^*_{\mathrm{PM}_{2.5}}$ (denoted as $\mathcal{P}$) is computed by $\frac{\mathcal{C}}{l} \times 100\%$, where l is the number of $\mathbf{X}_{\mathrm{PM}_{2.5}}$ data points.

Algorithm 1 presents our proposed parallel calibration deployment and drift detection ($\mathbb{P}$). The Algorithm operates such that from lines 1 to 3, it uses three determined thresholds including the maximum accepted residual ($\mathcal{T}_1$), maximum accepted percentage of $\mathbf{X}^*_{\mathrm{PM}_{2.5}}$ ($\mathcal{T}_2$), and the quantile outlier (qth). The Algorithm performs computations for the two methods (explained earlier) from lines 4 to 23 (while LCSs are deployed and perform measurements). The first method (lines 6 to 8) computes both calibrators ($\mathbb{W}$ and $\mathbb{V}$) and the residual $\mathbb{R}$. In line 9, using the available training data ($\mathbf{X}_{\mathrm{PM}_{2.5}}$), the second method computes the outlier limit (L), where in our case, we select $q = 0.99$. The lines 10 and 11 compute the outlier test data ($\mathcal{C}$) and the accepted percentage of outlier $\mathbf{X}^*_{\mathrm{PM}_{2.5}}$ ($\mathcal{P}$), respectively.

In lines 12 and 13, if the mean($\mathbb{R} < \mathcal{T}_1$), then our proposed calibration is executed using $\mathbb{V}$, which is known to be more accurate. In our study, since 100 μg/m^3 residual between two calibrators already indicate the drift in the calibrator $\mathbb{V}$, thus, we select $\mathcal{T}_1 = 100$. From lines 14 to 22, if $\mathbb{R}$ value crosses the defined threshold ($\mathcal{T}_1$), this indicates that $\mathbb{V}$, which is known to be less robust, begins to drift. Hence, our proposed calibration switches to execute calibrator $\mathbb{W}$ (line 16). In the lines 17–18, when $\mathcal{P}$ crosses the threshold $\mathcal{T}_2$ (e.g., in our case, we select it to be 25%), then calibrator drifts are declared. This means that both calibrators $\mathbb{V}$ and $\mathbb{W}$ do not function properly (line 19). Therefore, a mitigation such as recalibration is required (line 20), as explained in Section VI-B.

Algorithm 1: Deployment of Parallel Calibrators and Drift Detector ($\mathbb{P}$).

1: Determine maximum accepted residual: $\mathcal{T}_1$
2: Determine maximum accepted percentage of $\mathbf{X}^*_{\mathrm{PM}_{2.5}}$: $\mathcal{T}_2$
3: Determine q^{th} quantile outlier threshold
4: **while** LCSs measurements are being performed **do**
5: From LCS measurements, obtained {$\mathrm{PM}_{2.5}$, Temp, RH, P} to form matrix input $\mathbf{X}^*$
6: Compute $\mathbb{W}$: $\mathbf{y}^*_1 = \mathcal{W}(\mathbf{X}^*, \boldsymbol{\beta}^*)$
7: Compute $\mathbb{V}$: $\mathbf{y}^*_2 = \mathcal{V}(\mathbf{X}^*, \boldsymbol{\beta}^*)$
8: Compute $\mathbb{R}$
9: Compute the outlier limit: $L =$ quantile ($\mathbf{X}_{\mathrm{PM}_{2.5}}$, q)
10: Count $\mathcal{C}$: the occurrence number of $\mathbf{X}^*_{\mathrm{PM}_{2.5}} > L$
11: Compute the accepted percentage of $\mathbf{X}^*_{\mathrm{PM}_{2.5}}$ outliers : $\mathcal{P} = \frac{\mathcal{C}}{l} \times 100\%$ (l is the number of $\mathbf{X}_{\mathrm{PM}_{2.5}}$ data points)
12: **if** mean($\mathbb{R}$) $< \mathcal{T}_1$ **then**
13: Calibrate LCS using $\mathbb{V}$
14: **else if** mean($\mathbb{R}$) $> \mathcal{T}_1$ **then**
15: $\mathbb{V}$ does not function well:
16: Calibrate LCS using $\mathbb{W}$
17: **if** $\mathcal{P} > \mathcal{T}_2$ **then**
18: Calibrator drift is declared!
19: $\mathbb{V}$ and $\mathbb{W}$ do not function well
20: Mitigation (Section VI-B)
21: **end if**
22: **end if**
23: **end while**

V. Results

A. Calibration Performance

In order to evaluate the performance of calibrators $\mathbb{W}$ and $\mathbb{V}$, we design 12 different scenarios within five groups. As shown in Fig. 5, the groups are labeled by $\mathbf{G}_1 - \mathbf{G}_5$ and the scenarios are labeled by $\mathcal{S}_1 - \mathcal{S}_{12}$. These grouped scenarios are planned to evaluate the calibrators' performance within four approaches including *cross-units validation*, *cross-different-units validation*, *benchmark validation,* and *calibrators drift validation*.

The *cross-units validation* refers to calibrators' performance evaluation when we train the calibrators on one unit and then test them on another unit of the same type. This approach enables evaluating the calibrators' sensitivity and accuracy. In addition, this validation is beneficial for evaluating calibrators' resilience

against sensor fabrication variance. The *cross-different-units validation* aims to investigate the calibrators' performance when we train the calibrators on one unit and then test the them on another unit of different type. We use this approach to evaluate the calibrators' accuracy. The *calibrators drift validation* aims to investigate the calibrators drift due to the lack of information in the training data (for example, when calibrators have never experienced smoking events). Finally, *benchmark validation* is planned to evaluate the calibrators performance using a standard modeling process, which typically uses 70% random data for training and the remaining 30% of the data for testing. In our study, we use *benchmark validation* to compare its performance with the other validation approaches.

The first group (G_1), which includes scenario $\mathcal{S}_1$ aims to evaluate the calibrators using the *benchmark validation* approach. The second group (G_2), which includes scenarios $\mathcal{S}_2$ and $\mathcal{S}_3$ is designed to observe the accuracy of calibrators utilizing the *cross-units validation* approach. The third group ($\mathbf{G}_3$) that includes scenarios $\mathcal{S}_4 - \mathcal{S}_7$ uses the *cross-different-units validation* approach to investigate the calibrators' accuracy across different types of LCSs. The fourth group (G_4), which includes scenarios $\mathcal{S}_8 - \mathcal{S}_{11}$, is designed to perform *cross-units validation* approach in order to observe the sensitivity of the developed calibrators. In the scenarios in G_4, we use all data except one particular smoke from the sensor $\mathcal{L}_{2a}$ for training the calibrators. Then, we test the calibrator on sensor $\mathcal{L}_{2b}$. For instance, in scenario $\mathcal{S}_8$, we train the calibrators using all dataset from $\mathcal{L}_{2a}$ except for tobacco and test it on sensor $\mathcal{L}_{2b}$. The fifth group ($\mathbf{G}_5$) that consists of only the scenario $\mathcal{S}_{12}$ is planned to perform *calibrators drift validation*. In this scenario, we use $\mathcal{L}_{2a}$ to train calibrators using the whole normal events data, and test the trained calibrators with all of the smoking events.

Fig. 5 shows the performance results of different calibrators, including BLM ($\mathbb{W}_2$), BNN ($\mathbb{V}_2$), and our proposed calibrator ($\mathbb{P}$) for different scenarios. In the figure, we also include the most popular white-box ($\mathbb{W}_1$) and black-box ($\mathbb{V}_1$) calibration methods in order to compare the performance results of the calibrators $\mathbb{V}_2$, $\mathbb{W}_2$, and $\mathbb{P}$. As presented in figure, we use the performance metrics R, MAE, and MAPE.

Using *benchmark validation* approach, which is the case of $\mathbf{G}_1$, $\mathbb{W}_2$ and $\mathbb{V}_2$ calibrators demonstrate to have a better performance than $\mathbb{W}_1$ and $\mathbb{V}_1$ using all performance metrics. The existence of regularization factor in Bayesian inference makes $\mathbb{W}_2$ and $\mathbb{V}_2$ calibrators more generalized than $\mathbb{W}_1$ and $\mathbb{V}_1$. In addition, the performance of $\mathbb{V}_2$ is better than $\mathbb{W}_2$, shown by all performance metrics. Through this approach, our proposed method ($\mathbb{P}$) shows better performance than the rest of the calibrators, except in case of $\mathbb{V}_2$ that has just minor performance difference with $\mathbb{P}$. The reason for this minor difference might be that the training data already contain the outliers, while the test data do not contain the outliers.

The *cross-units validation* approach that is evaluated within G_2 consists of the scenarios $\mathcal{S}_2$ and $\mathcal{S}_3$. The values of R for $\mathbb{V}_2$ are consistently higher than $\mathbb{W}_2$ for both scenarios. This implies $\mathbb{V}_2$ generates better calibrators' accuracy. Likewise, the values of metrics MAE and MAPE for $\mathbb{V}_2$ is lower than $\mathbb{W}_2$, indicating that $\mathbb{V}_2$ is more accurate than $\mathbb{W}_2$ in these scenarios. Both $\mathbb{W}_2$

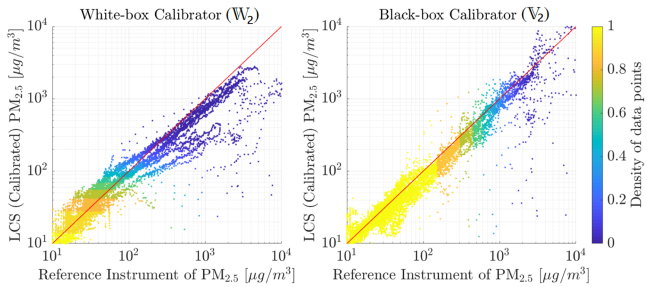

Fig. 6. Scatter plots between the reference instrument and the calibrated LCS using $\mathbb{W}_2$ (left) and $\mathbb{V}_2$ (right) for scenario $\mathcal{S}_2$.

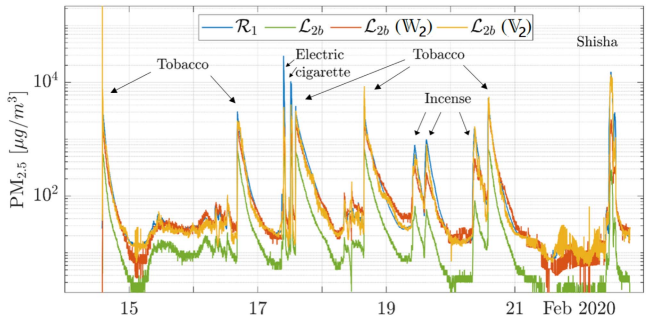

Fig. 7. Time-series plot representing the ground truth ($\mathcal{R}_1$), uncalibrated LCS ($\mathcal{L}_{2b}$), calibrated LCS ($\mathcal{L}_{2b}$) using $\mathbb{W}_2$ and $\mathbb{V}_2$, tested on scenario $\mathcal{S}_2$.

and $\mathbb{V}_2$ calibrators outperform $\mathbb{W}_1$ and $\mathbb{V}_1$ due to the same reasons explained previously for scenario $\mathcal{S}_1$. In scenario $\mathcal{S}_2$, $\mathbb{P}$ demonstrates to have better performance than all calibrators. In scenario $\mathcal{S}_3$, $\mathbb{P}$ also outperforms all of the calibrators, except the $\mathbb{V}_2$ with a very minor difference. The reason for the minor difference is explained in $\mathcal{S}_1$. To conclude, the performance metrics evaluations confirm that the calibrators function well across units of the same type.

For *cross-different-units validation* approach, we consider the scenarios $\mathcal{S}_4 - \mathcal{S}_7$ in group G_3. Similar to group G_2, the performance metrics in the scenarios in group G_3 show that generally $\mathbb{W}_2$ and $\mathbb{V}_2$ have better performance than $\mathbb{W}_1$ and $\mathbb{V}_1$, respectively. However, in these scenarios, $\mathbb{V}_2$ does not outperform $\mathbb{W}_2$, indicating that white-box calibrators perform slightly better than black-box calibrators when they are tested on different unit type. Nevertheless, $\mathbb{P}$ still outperforms all other calibrators, indicating that $\mathbb{P}$ shows promising results when it is tested on different unit type. As outcome of the *cross-different-units validation*, the performance results demonstrate that all of the calibrators still function well across different units.

Similar to group G_2, the group G_4 that includes the scenarios $\mathcal{S}_8$–$\mathcal{S}_{11}$ also evaluates the *cross-units validation* approach. In scenarios of G_4, calibrators $\mathbb{W}_2$ and $\mathbb{V}_2$ still outperform $\mathbb{W}_1$ and $\mathbb{V}_1$. However, in some cases (e.g., $\mathcal{S}_8$ and $\mathcal{S}_{11}$), the results of performance metrics show that $\mathbb{W}_2$ slightly perform better than $\mathbb{V}_2$. Therefore, as an outcome of *cross-units validation*, $\mathbb{V}_2$ seems to be more sensitive when facing a new smoking event. For example, in $\mathcal{S}_8$, the calibrator $\mathbb{W}_2$ works better than $\mathbb{V}_2$, because $\mathbb{W}_2$ is more robust to outliers than $\mathbb{V}_2$. As a result, $\mathbb{V}_2$ does not accurately calibrate the LCS on the tobacco smoking event. Nevertheless, $\mathbb{P}$ outperforms all of the calibrators, although the

TABLE II
SUMMARY OF R VALUES FOR DIFFERENT GROUPS OF SCENARIOS

Group	Scenario	$\mathbb{W}_1$	$\mathbb{W}_2$	$\mathbb{V}_1$	$\mathbb{V}_2$	$\mathbb{P}$
$\mathbf{G}_1$	$\mathcal{S}_1$	0.918	0.921	0.881	0.937	0.931
$\mathbf{G}_2$	$\mathcal{S}_2$ - $\mathcal{S}_3$	0.921	0.923	0.915	0.932	0.933
$\mathbf{G}_3$	$\mathcal{S}_4$ - $\mathcal{S}_7$	0.922	0.923	0.904	0.919	0.930
$\mathbf{G}_4$	$\mathcal{S}_8$ - $\mathcal{S}_{11}$	0.916	0.917	0.921	0.921	0.930
$\mathbf{G}_5$	$\mathcal{S}_{12}$	0.690	0.707	0.582	0.542	Drift

test data contains outliers. This is due to the fact that the parallel implementation in $\mathbb{P}$ enables switching from $\mathbb{V}_2$ to $\mathbb{W}_2$ when the residual $\mathbb{R}$ increases due to outliers.

To investigate the *calibrators drift validation* approach, we consider the group $\mathbf{G}_5$ that includes the scenario $\mathcal{S}_{12}$. Let us recall that in this scenario, all smoking events data are excluded in calibrators' training. The results show that both calibrators $\mathbb{V}_1$ and $\mathbb{V}_2$ clearly drift by presenting small values for the metric R and values higher than 1 for MAPE. While $\mathbb{W}_1$ and $\mathbb{W}_2$ maintain the performance to an acceptable level by showing a value about 0.7 for the metric R. Indeed, calibrators $\mathbb{W}_1$ and $\mathbb{W}_2$ are more robust than calibrators $\mathbb{V}_1$ and $\mathbb{V}_2$, that is, because white-box calibrators have less modeling complexity. In this scenario, our proposed method $\mathbb{P}$ alerts the calibrators drift as described in Algorithm 1. This is highlighted by D, i.e., calibrators drift for scenario $\mathcal{S}_{12}$ in Fig. 5. The calibrators drift analysis will be explained in Section IV-C.

Next, we generate scatter plots (see Fig. 6) and time-series plots (depicted in Fig. 7) to provide further insights about the results presented in Fig. 5. Since most results indicate that $\mathbb{V}_2$ is more accurate than $\mathbb{W}_2$, in this case, as an example, we consider further analyzing scenario $\mathcal{S}_2$, which is also a simpler scenario to understand. Fig. 6 depicts scatter plots between the reference instrument ($\mathcal{R}_1$) and calibrated LCS ($\mathcal{L}_{2b}$), for calibrators $\mathbb{W}_2$ (left subfigure) and $\mathbb{V}_2$ (right subfigure). In this figure, the colors indicate the density of data points for $PM_{2.5}$ measurement. The plot shows that the data points of $PM_{2.5}$ concentrations scatter around the red reference lines for both calibrators. The results of scatter plot indicate that both calibrators perform well by correcting the measurements of $\mathcal{L}_{2b}$ and making them similar to the measurements of $\mathcal{R}_1$. In addition, $\mathbb{V}_2$ calibrate $PM_{2.5}$ more accurately than $\mathbb{W}_2$, especially at high $PM_{2.5}$ concentrations. Nevertheless, both $\mathbb{W}_2$ and $\mathbb{V}_2$ calibrate $PM_{2.5}$ to an acceptable level. This is confirmed by Fig. 7, where both calibrators $\mathbb{W}_2$ and $\mathbb{V}_2$ are tracking very well the reading of $\mathcal{R}_1$. Fig. 7 also illustrates that the calibrators are able to capture the extreme smoking events effectively. As a result, implementing both of the calibrators enables detecting and avoiding false negative situations, which may be harmful for human.

The results of different scenarios presented in Fig. 5 show that both calibrators have strengths and weaknesses. Indeed, $\mathbb{V}_2$ tends to drift drastically when a completely new situation emerges (as the case in scenario $\mathcal{S}_{12}$), however, $\mathbb{W}_2$ performs adequately with acceptable performance degradation. Indeed, these facts had motivated us to deploy both calibrators in parallel ($\mathbb{P}$) as they have two different characteristics. In order to highlight the performance results of all of the calibrators and $\mathbb{P}$, in Table II, we summarize the mean of R values for the scenarios in each group. Indeed, this table concludes the results presented in Fig. 5 by presenting that 1) $\mathbb{W}_2$ and $\mathbb{V}_2$ are generally better than the most popular calibration methods $\mathbb{W}_1$ and $\mathbb{V}_1$, 2) $\mathbb{V}_2$ is better than $\mathbb{W}_2$ for most scenarios, 3) our proposed approach $\mathbb{P}$ outperforms the other calibrators, and 4) $\mathbb{P}$ enables calibrator drift detection as shown in scenario $\mathcal{S}_{12}$.

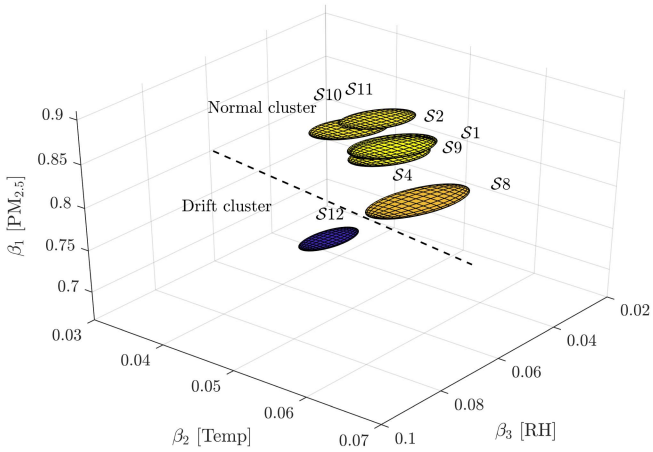

Fig. 8. Model coefficients of calibrator $\mathbb{W}$ in the forms of ellipsoids for scenarios $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_4$, and $\mathcal{S}_8$–$\mathcal{S}_{12}$.

It is worth noting that the drift detection is important because LCSs and reference instruments usually are not installed or placed near each other. Consequently, it is challenging to detect calibrator drifts in the absence of a reference instrument, which provides ground truth data. As described in Section IV-C, deploying two types of calibrators allows cross-checking them. This process which is called *drift monitoring* aims to ensure both calibrators perform effectively by enabling detecting the calibrators drifts. The next section provides further analysis about the calibrators drifts.

B. Drift Analysis

As explained in Section IV-C, analyzing the model coefficients of calibrator $\mathbb{W}$ provides insights about the variables impacting the LCSs measurements. Fig. 8 depicts the model coefficients of calibrators $\mathbb{W}_2$ (obtained using the data from $\mathcal{L}_{2a}$) for scenarios $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_4$, and $\mathcal{S}_8$–$\mathcal{S}_{12}$. Since the calibrators $\mathbb{W}_2$ in these scenarios are based on BLMs, their model coefficients ($\boldsymbol{\beta}$) are in the form of Gaussian distribution, following $p(\mu_\beta, V_\beta)$, with mean μ_β and variance V_β. These model coefficients ($\boldsymbol{\beta}$) are depicted in Fig. 8 with the ellipsoids, where the core and radius represent the mean and standard deviation of multivariate Gaussian distribution, respectively.

In the figure, the largest magnitude of coefficient $\boldsymbol{\beta}$ indicates the most dominant variable in LCSs measurements. The variables include $PM_{2.5}$, Temp, and RH, which are associated with β_1, β_2, and β_3, respectively. It can be seen that while $PM_{2.5}$ that is associated with β_1 plays a major role in calibration as their values range between 0.7 and 0.9, which are one magnitude bigger than the values in β_2 and β_3. The variations of Temp and RH measurements have less influence in calibrators performance. In addition, the role of pressure (P) is trivial with the mean of β_4 for all scenario is closed to -0.003 (not including in the figure). Moreover, as illustrated in Fig. 8, the ellipsoids position that

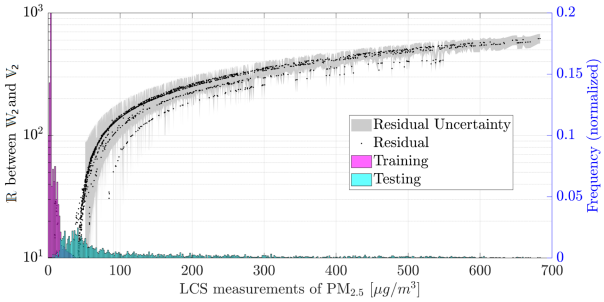

Fig. 9. Calibrators drift monitoring for scenario $\mathcal{S}_{12}$.

divide between normal (yellow) and drift (dark blue) clusters are dominated by the magnitude of β_1. This means that (as described in Algorithm 1) monitoring the changes on the test data $PM_{2.5}$ ($\mathbf{X}^*_{PM_{2.5}}$) provides an indication about the calibrator drifts.

Fig. 9 illustrates the relationship between residual ($\mathbb{R}$) and $PM_{2.5}$ measurements data gathered during the testing process ($\mathbf{X}^*_{PM_{2.5}}$), for $\mathcal{S}_{12}$. While the blue histogram shows the $\mathbf{X}^*_{PM_{2.5}}$, the pink histogram is $PM_{2.5}$ measurements data collected during the training process ($\mathbf{X}_{PM_{2.5}}$). In the figure, x-axis represents $PM_{2.5}$ measurements from LCS prior to calibration, the left y-axis shows the residual ($\mathbb{R}$) between $\mathbb{V}_2$ and $\mathbb{W}_2$, and the right y-axis presents the frequency of histograms.

As described in Algorithm 1, drifting detection can be performed by monitoring $\mathbb{R}$ between $\mathbb{V}_2$ and $\mathbb{W}_2$. In the figure, $\mathbb{R}$ shows incremental pattern (with uncertainty) when the LCS PM $_{2.5}$ measurement concentration increases. In this case, $\mathbb{W}_2$ maintains the calibration performance to an acceptable level, but both calibrators fail when LCS $PM_{2.5}$ measurements ($\mathbf{X}^*_{PM_{2.5}}$) are too large (i.e., mean($\mathbb{R}$) $> \mathcal{T}_1$). In the figure, this is shown when the $\mathbb{R}$ reaches 100 μ g/m^3 in the left y-axis.

Furthermore, while the outlier limit (L) lies on the edge of the pink histogram's right tail (about 50 μg/m^3 on x-axis at $q = 0.99$). It is obvious that the blue histogram has deviated (expanded) largely from the pink histogram, indicating that the accepted percentage of $\mathbf{X}^*_{PM_{2.5}}$ already crosses the threshold (i.e., $\mathcal{P} > \mathcal{T}_2$). This indicates that the calibrator drift is declared (according to Algorithm 1) and both calibrators are unable to calibrate the readings of LCSs.

Obtaining a reliable drifting monitoring also enables detecting the wear in sensors hardware when they are in real use. As the wear of hardware usually provides inconsistent reading, therefore, residual evaluation would assist in identifying the sources of errors. The drifting monitoring allows ensuring the sensors calibrators and hardware function accurately in the field deployment. If they do not function accurately, then the maintenance can be performed based on the information provided by drifting monitoring.

VI. Discussion

A. Comparison With the State-of-the-Art

LCSs increasingly use ML-based calibration methods to improve the accuracy of sensor measurements [5]. The studies in the state-of-the-art, present specific ML-based calibration methods, however, in contrast, we propose a generic strategy in applying parallel ML-based calibration models ($\mathbb{P}$). Indeed, most of the studies in literature implement either white-box ($\mathbb{W}$) or black-box ($\mathbb{V}$) models to perform calibration. Our proposed method offers flexibility in choosing any ML model to represent $\mathbb{W}$ and $\mathbb{V}$ models. Thus, we selected BLM and BNN in our proposed method $\mathbb{P}$.

These studies in literature use different datasets generated in different environments, seasons, and locations, while each dataset has different characteristics. Hence, comparing the performance results of the calibration models seems to be inappropriate. Nevertheless, to show the performance of our proposed method ($\mathbb{P}$), we redeveloped the most popular calibration methods [5], i.e., MLR and ANN, and then we, respectively, compared them with our selected calibration methods, which are BLM and BNN. Indeed, as presented in Section V-A, our proposed method (which implements parallel ML models) outperforms individual selected methods (i.e., BLM and BNN) as well as the most popular methods (i.e., MLR and ANN). Our proposed method indeed promotes the use of Bayesian models and parallel deployment for LCSs calibration methods.

Furthermore, as the deployment of sensor networks in smart cities has recently increased, the drifts of calibration models have become challenging during their in-field operation time. The drifts result from various reasons including clean air policies, e.g., traffic, changes in humans consumption patterns such as fuel and gas [22], or temporal effects such as forest fires and volcano eruptions [23]. To detect the drifts, the methods in the state-of-the-art use statistical difference in distributions of air pollution measurements [12]. However, in contradiction our proposed method uses two layers of detection methods, first by computing residual between $\mathbb{W}$ and $\mathbb{V}$, and then by monitoring the changes of one of the key variables measured by LCSs, e.g., $PM_{2.5}$. The two layers implementation would reduce the probability of receiving *false positive* alarms if the proposed method was applied in the earlier mentioned scenarios causing the drifts (such as policies and temporal effects).

Moreover, to the best of our knowledge, it is also the first time, the drift detection method is tested in indoor environments. As we have performed comprehensive experiments by testing and evaluating our proposed method in an indoor environment (by various smoking events), while according to recent literature survey study focused on the use of LCSs indoors [13], the majority of the works in literature do not calibrate nor validate the LCSs used in their studies. For example, based on this survey study there are approximately 77.5% of works did not include details about the calibration of their LCSs [14]. We also demonstrate how extreme events such as smoking activities can alter significantly the LCSs reading, leading to *false negative*. It is worth noting that *false negative* situation in sensors reading can be harmful to human exposure as there are no alarm alerting people when the pollution concentration is very high in indoor environments.

In summary, in our paper, we propose a generic parallel ML-based calibration method, which (as mentioned earlier) provides many advantageous compared which the works in literature. Calibration and drift detection methods might perform differently in

various environments, e.g., meteorological conditions. However, the dataset we used in our study is limited to only to one type of indoor environment having a specific characteristics such as room size, ventilation, and other influencing factors. Hence, our proposed method requires more evaluations using different and comprehensive datasets obtained from various environmental characteristics. Therefore, the use of comprehensive datasets can assist investigating different LCSs calibration and drift detection methods.

B. Suggested Solutions for Drifting

Besides sensors recalibration, investigating the causes of drifts helps understanding the sources of problems and therefore enables improving the calibration models and the LCS hardware design. We envision three methods to minimize the drift in calibrators:

Method 1: Extensive laboratory experiments can be performed for testing different scenarios on new design LCSs. Different kinds of aerosol particles with varying meteorological variables are inserted to an experimental chamber, where the LCSs are placed. The idea aims to mimic as many scenarios in which the LCSs may encounter in the field deployment as possible. For example, if LCSs are designed to be deployed indoors, they should be tested on different indoor scenarios, e.g., smoking and fire sensing. Thus, based on these experiments, effective calibrators can be developed.

Method 2: Adaptive calibration model can be used. The adaptive model can be developed if the ground truth data available, e.g., from a nearby reference instrument or other calibrated LCS, which can communicate via Internet. For example, adaptive calibrators can be developed using federated learning techniques [24].

Method 3: Robust calibrators can be developed such as $\mathbb{W}$, where the calibrators do not drift easily under unexpected circumstances. For example, in our approach, we coupled $\mathbb{W}$ and $\mathbb{V}$. Hence, if a drift is detected then $\mathbb{W}$ still function to an acceptable limit compared to $\mathbb{V}$ in some new cases before retraining the calibrators. The best robust calibrators are physics-based models, where the underlying physical relationship between LCS and reference instrument can be derived.

C. Industrial Applications

Our proposed method can be potentially extended on various industrial applications using the calibrated $PM_{2.5}$ concentration (as shown in Fig. 1, part ❺). Following are examples of few potential industrial applications:

1) Personalized Health Device: Accurate measurements of $PM_{2.5}$ concentration enables deriving personalized health information from LCS devices [25]. This provides information of individual deposited dosage [26], which can be integrated via wearable devices [27].

2) Smoking Detector: When smoking indoors, the smoke lingers in the air, because the smoke particles sizes are too small such that 85% of them are invisible and odorless.[3] Our experiment shows that high $PM_{2.5}$ concentrations remain in the room for hours, which can cause longer breathing issue for humans. Recent development in automatic image and video analytics has enabled smoke detection with a high accuracy [28]. However, adopting this method is expensive since cameras need to be installed in all rooms. Using our proposed methodology for smoking detection is economically beneficial.

3) Fire Detector: Current indoor fire detectors are based on ionization and photoelectric technologies [29]. However, these technologies might not always be effective in detecting very small increase of PM concentration triggered by fires in early stages. Thus, to complement, applying our proposed method of calibrated $PM_{2.5}$ LCSs contributes to early fire detection.

4) Poisonous Gases Detector and Monitoring: LCSs can also be used for detecting poisonous gases indoors such as CO [30]. Indeed, CO is a colorless, tasteless, and odorless gas produced by incomplete combustion of carbon-containing materials. Similar to LCSs of $PM_{2.5}$, other LCSs capable of measuring CO require calibration. Extending and embedding our proposed method to low-cost gas sensors such as CO enables detecting accurately the poisonous gas concentrations.

5) Engineering Assets Monitoring: Accurate LCSs deployment can help monitor engineering assets. For example, more affordable accurate sensors can be deployed massively to monitor atmospheric corrosion. Different gases such as CO_2, SO_2, and dust can accelerate corrosion in various types of metals [31]. Accurate monitoring of such pollutants enables engineers to perform preventive maintenance.

6) Electronic Nose (e-nose): E-nose is known as an electronic sensing device intended to detect odors. E-nose devices are widely used in research and development, quality control, process and production, health, and security purposes. Although e-nose devices currently are used in many application areas, they are still considered as unreliable solutions [7] due to low accuracy as air quality LCSs. Indeed, our proposed method can be adopted to improve e-nose sensing performance.

VII. Conclusion

Air quality LCSs suffer from sensing accuracy when they were used for measuring extreme events. In this article, we proposed an intelligent sensor calibration process that enables effectively correcting LCS readings as well as identifying the calibrators' drift. Therefore, we performed controlled experiments in an indoor environment for defining scenarios for extreme events. These scenarios included 12 different indoor smoking activities. We used the data collected from these controlled experiments for obtaining insight about smoking events and we also utilize dthe data for developing calibrators and investigating their performance. We further used Bayesian framework for developing white-box ($\mathbb{W}$) and black-box ($\mathbb{V}$) calibrators. Then, we deployed these calibrators in parallel ($\mathbb{P}$) in order to correct LCSs measurements and enable detecting calibrators drift. Then, we

[3][Online]. Available: https://www.nhsinform.scot/campaigns/take-it-right-outside

TABLE III
PERFORMANCE METRICS FOR SENSORS AND CALIBRATORS VALIDATIONS

Performance Metrics	Formulation
Pearson Correlation Coefficient	$R = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})\,(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$
Mean Absolute Error	$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert\hat{y}_i - y_i\rvert$
Root Mean Squared Error	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$
Mean Absolute Percentage Error	$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i-\hat{y}_i\rvert}{y_i}$

evaluated the calibrators in a controlled experiment under different types of smoking events within 12 scenarios. For instance, in scenario 2 (i.e., $\mathcal{S}_2$), we trained the calibrator on one LCS and test it on another LCS of the same type. Another example was scenario 12 (i.e., $\mathcal{S}_{12}$), which is designed to mimic the calibrators' drift. We then evaluated the developed calibrators on all designed scenarios using different performance metrics. The performance results showed that our proposed method accurately estimates the aerosol mass concentration in different scenarios, except for $\mathcal{S}_{12}$. Because the calibrators in $\mathcal{S}_{12}$ were established only using normal data (not for extreme events). Nevertheless, we demonstrated that our proposed drift monitoring was able to detect the calibrators' drift for $\mathcal{S}_{12}$. Finally, we discussed how our proposed method was extendable to various industrial applications, such as smoking, fire and poisonous gas detectors, engineering assets monitoring, and health informatics.

APPENDIX

A. Performance Metrics

Performance metrics used to validate the sensors and calibrators are presented in Table III. The notations $\mathbf{y}$ and $\bar{\mathbf{y}}$ are the measurements through reference instrument and its mean value, respectively. Whereas the notations $\hat{\mathbf{y}}$ and $\bar{\hat{\mathbf{y}}}$ represent LCS measurements and the mean of LCS measurements, respectively, before calibration (for sensors validation) or after calibration (for calibrators validation).

B. BLM: White-Box Calibrator ($\mathbb{W}$)

A Bayesian linear calibrator, $\mathbf{y}_1$, can be modeled as

$$\mathbf{y}_1 = \mathcal{W}(\mathbf{X}, \boldsymbol{\beta}) + \varepsilon_1 \tag{4}$$

where ε_1 is a random error term, which follows a Gaussian distribution, with zero mean, and σ^2 noise variance, $\varepsilon_1 \sim \mathcal{N}(0, \sigma^2 I)$. The symbol $\mathcal{W}$ is a function of white-box BLM. $\mathcal{W}(\mathbf{X}, \boldsymbol{\beta}) = \Phi(\mathbf{X})\boldsymbol{\beta}$, where $\Phi(\mathbf{X})$ is an $N \times D$ design matrix for the inputs. In this case, the design matrix is choice to be: $\Phi(\mathbf{X}) = [X_1, X_2, \ldots, X_D]$. The calibrator $\mathcal{W}$, expressed in the (4) can be called a white-box because the relationship between the inputs and output are visible and transparent. In order to optimize variable $\boldsymbol{\beta}$, Bayesian inference use Bayes' rule: posterior $\propto$ likelihood $\times$ prior.

1) Prior Distributions: of the calibrator coefficients, $\boldsymbol{\beta}$ is modeled as a Gaussian distribution: $p(\boldsymbol{\beta}) \sim \mathcal{N}(\mu_o, \sigma_0^2)$. Informative prior can be determined by applying linear regression on the data. Therefore, variable μ_0 can be estimated and the variable σ_0^2 are chosen three times larger than its mean value.

2) Likelihood Function: for this model is the conditional probability of observing the measurement data ($\mathbf{X}$) and the model parameters ($\boldsymbol{\beta}, \sigma^2$). The likelihood also follows a Gaussian distribution and it can be written as

$$p(\mathbf{y}_1|\mathbf{X}, \boldsymbol{\beta}) = \mathcal{N}(\mathbf{y}_1|\Phi(\mathbf{X})\boldsymbol{\beta}, \sigma^2 I). \tag{5}$$

3) Posterior Distributions: can be computed using the likelihood function and the prior distribution, based on Bayes' theorem to give

$$\underbrace{p(\boldsymbol{\beta}|\mathbf{y}_1, \mathbf{X})}_{\text{posterior dist.}} \propto \underbrace{p(\mathbf{y}_1|\mathbf{X}, \boldsymbol{\beta})}_{\text{likelihood func.}}\ \underbrace{p(\boldsymbol{\beta})}_{\text{prior dist.}}. \tag{6}$$

The probabilistic model above is linear, and therefore the posterior distribution can be computed analytically, resulting in another Gaussian distribution to give

$$p(\boldsymbol{\beta}) = p(\boldsymbol{\beta}|\boldsymbol{\beta}^*, V^*) \tag{7}$$

where $\boldsymbol{\beta}^*$ and V^* are the mean and variance, respectively. They can be computed by

$$\boldsymbol{\beta}^* = V^*(V^{-1}\bar{\boldsymbol{\beta}} + \Phi(\mathbf{X})\mathbf{y}_1/\sigma^2) \tag{8}$$

$$V^* = (V^{-1} + \Phi(\mathbf{X})^T\Phi(\mathbf{X})/\sigma^2)^{-1}. \tag{9}$$

4) Predictive Distribution: is also in the form of Gaussian distribution, symbolized by $p(\mathbf{y}_1^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}_1)$. This can also be computed analytically by using posterior distribution, to give

$$p(\mathbf{y}_1^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}_1) = (\mathbf{y}_1^*|\Phi(\mathbf{X}^*)\boldsymbol{\beta}^*, \Phi(\mathbf{X}^*)V^*\Phi(\mathbf{X})^{*T} + \sigma^2 I). \tag{10}$$

C. BNN: Black-Box Calibrator ($\mathbb{V}$)

Neural networks are usually considered as a black-box model ($\mathbb{V}$), since they provide little explanatory insight into the relative influence of the independent variables in the prediction process. The black-box calibrator can then be modeled as

$$\mathbf{y}_2 = \mathcal{V}(\mathbf{X}, \boldsymbol{\omega}) + \varepsilon \tag{11}$$

where ε_2 is a random error term following a Gaussian distribution with zero mean and γ precision. The symbol $\mathcal{V}$ is a black-box BNN function and $\mathbf{X}$ is input data measurements.

A neural network, $\mathcal{V}(\mathbf{X}, \boldsymbol{\omega})$, can be viewed as a probabilistic model, that follows a Gaussian distribution, given by

$$p(\mathbf{y}_2|\mathbf{X}, \boldsymbol{\omega}, \gamma) = \mathcal{N}(\mathbf{y}_2|\mathcal{V}(\mathbf{X}, \boldsymbol{\omega}), \gamma^{-1}) \tag{12}$$

where the notations of $\mathbf{X}, \boldsymbol{\omega}$, and γ are the inputs, the neural network weights, and the precision of the Gaussian distribution, respectively. Equation (12) is also known as a likelihood function.

In a Bayesian framework, a prior distribution needs to be assigned, where in this case, the prior follows a Gaussian distribution with mean zero and the precision of α, given by

$$p(\boldsymbol{\omega}|\alpha) = \mathcal{N}(\boldsymbol{\omega}|\mathbf{0}, \alpha^{-1}\mathbf{I}). \tag{13}$$

Using the prior distribution and likelihood function, the posterior distribution for the BNN can be computed based on Bayes

theorem to give

$$\underbrace{p(\boldsymbol{\omega}|\mathbf{y}_2,\mathbf{X},\alpha,\gamma)}_{\text{posterior dist.}} \propto \underbrace{p(\mathbf{y}_2|\mathbf{X},\boldsymbol{\omega},\gamma)}_{\text{likelihood func.}}\ \underbrace{p(\boldsymbol{\omega}|\alpha)}_{\text{prior dist.}}. \tag{14}$$

The inclusion of the prior distribution leads to a regularization, which then counters overfitting. Furthermore, BNN provides a degree of belief on the estimated output, which can be used to assess the quality of the predictions. In our case, the confidence interval will be used to monitor drifting detection, which will be described in Appendix D.

Due to the nonlinear dependence of $\mathcal{V}(\mathbf{X},\boldsymbol{\beta})$ on $\boldsymbol{\omega}$, the posterior distribution calculation is intractable. Therefore, the posterior distributions as well as predictive distribution, $p(\mathbf{y}_2^*|\mathbf{X}^*,\mathbf{X},\mathbf{y}_2)$, can be approximated using Laplace approximation or variational inference as described in [20].

D. Calibrators Residual: Drifting Monitoring

The calibrators' drifting is monitored through the residual of two predictive distributions, that is between $\mathbb{W}$ and $\mathbb{V}$. Let $\mathbb{W}$ and $\mathbb{V}$ be independent random variables that are normally distributed, then their residual is also normally distributed

$$\mathbb{W} \sim \mathcal{N}(\mathbf{y}_1^*|\mu_{\mathbf{y}_1}^*,\Sigma_{\mathbf{y}_1}^*) \tag{15}$$

$$\mathbb{V} \sim \mathcal{N}(\mathbf{y}_2^*|\mu_{\mathbf{y}_2}^*,\Sigma_{\mathbf{y}_2}^*) \tag{16}$$

$$\mathbb{R} = \mathbb{V} - \mathbb{W} \tag{17}$$

where $\mathbb{R}$ is a residual function for monitoring calibrators. Then, this results in another Gaussian predictive distribution, given by

$$\mathbb{R} \sim \mathcal{N}(\mathbf{r}|\mu_{\mathbf{y}_2}^* - \mu_{\mathbf{y}_1}^*, \Sigma_{\mathbf{y}_2}^* + \Sigma_{\mathbf{y}_1}^*). \tag{18}$$

The sensors inspection and maintenance decision can then be determined using information from the residual distribution, by extracting the mean, $\mu_{\mathbb{R}} = \mu_{\mathbf{y}_2}^* - \mu_{\mathbf{y}_1}^*$ and the standard deviation, $\sigma_{\mathbb{R}} = \sqrt{\Sigma_{\mathbf{y}_2}^* + \Sigma_{\mathbf{y}_1}^*}$.

Proof: For independent random variables $\mathbb{V}$ and $\mathbb{W}$, the distribution $f_{\mathbb{R}}$ of $\mathbb{R} = \mathbb{V} - \mathbb{W}$ equals to the convolution of $f_{\mathbb{V}}$ and $f_{\mathbb{W}}$

$$f_{\mathbb{R}} = \int_{-\infty}^{\infty} f_{\mathbb{V}}(\mathbf{y}_1^* - \mathbf{r}) f_{\mathbb{W}}(\mathbf{y}_1^*) d\mathbf{y}_1^*. \tag{19}$$

Given $f_{\mathbb{W}}$ and $f_{\mathbb{V}}$ are Gaussian distribution following:

$$f_{\mathbb{W}} = \mathcal{N}(\mathbf{y}_1^*|\mu_{y_1}^*,\Sigma_{y_1}^*) \tag{20}$$

$$f_{\mathbb{V}} = \mathcal{N}(\mathbf{y}_2^*|\mu_{y_2}^*,\Sigma_{y_2}^*). \tag{21}$$

Therefore, by substituting (20) and (21) into (19), we obtain

$$f_{\mathbb{R}} = \frac{1}{\sqrt{2\pi(\Sigma_{\mathbf{y}_2}^* + \Sigma_{\mathbf{y}_1}^*)}} \exp\left[-\frac{(\mathbf{r} - (\mu_{\mathbf{y}_2}^* - \mu_{\mathbf{y}_1}^*)^2}{2(\Sigma_{\mathbf{y}_2}^* + \Sigma_{\mathbf{y}_1}^*)}\right]. \tag{22}$$

ACKNOWLEDGMENT

This research was part of a close collaboration between Joint International Research Laboratory of Atmospheric and Earth System Sciences (JirLATEST) at Nanjing University, Institute for Atmospheric and Earth System Research (INAR/Physics) and Department of Computer Science at Helsinki University, Environmental and Atmospheric Research Laboratory (EARL) at University of Jordan, and Institute for Space-Earth Environmental Research (ISEE) at Nagoya University. The sole responsibility of this publication lies with the authors. The European Union is not responsible for any use that may be made of the information contained therein.

REFERENCES

[1] F. Chraim, Y. B. Erol, and K. Pister, "Wireless gas leak detection and localization," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 768–779, Apr. 2016.
[2] C. Nandi, R. Debnath, and P. Debroy, "Intelligent control systems for carbon monoxide detection in IoT environments," in *Guide to Ambient Intelligence in the IoT Environment*. Berlin, Germany: Springer, 2019, pp. 153–176.
[3] N. H. Motlagh *et al.*, "Low-cost air quality sensing process: Validation by indoor-outdoor measurements," in *Proc. 15th IEEE Conf. Ind. Electron. Appl.*, 2020, pp. 223–228.
[4] N. H. Motlagh *et al.*, "Toward massive scale air quality monitoring," *IEEE Commun. Mag.*, vol. 58, no. 2, pp. 54–59, Feb. 2020.
[5] F. Concas *et al.*, "Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis," *ACM Trans. Sensor Netw.*, vol. 17, no. 2, pp. 1–44, 2021.
[6] B. Alfano *et al.*, "A review of low-cost particulate matter sensors from the developers' perspectives," *Sensors*, vol. 20, no. 23, 2020, Art. no. 6819.
[7] S. Marco and A. Gutierrez-Galvez, "Signal and data processing for machine olfaction and chemical sensing: A review," *IEEE Sensors J.*, vol. 12, no. 11, pp. 3189–3214, Nov. 2012.
[8] M. R. Giordano *et al.*, "From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors," *J. Aerosol Sci.*, vol. 158, 2021, Art. no. 105833.
[9] M. A. Zaidan *et al.*, "Intelligent calibration and virtual sensing for integrated low-cost air quality sensors," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13638–13652, Nov. 2020.
[10] T. Zheng *et al.*, "Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments," *Atmos. Meas. Techn.*, vol. 11, no. 8, pp. 4823–4846, 2018.
[11] B. Maag, Z. Zhou, and L. Thiele, "A survey on sensor calibration in air pollution monitoring deployments," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4857–4870, Dec. 2018.
[12] S. De Vito, E. Esposito, N. Castell, P. Schneider, and A. Bartonova, "On the robustness of field calibration for smart air quality monitors," *Sensors Actuators B, Chem.*, vol. 310, 2020, Art. no. 127869.
[13] H. Chojer, P. Branco, F. Martins, M. Alvim-Ferraz, and S. Sousa, "Development of low-cost indoor air quality monitoring devices: Recent advancements," *Sci. Total Environ.*, vol. 727, 2020, Art. no. 138385.
[14] J. Saini, M. Dutta, and G. Marques, "Sensors for indoor air quality monitoring and assessment through Internet of Things: A systematic review," *Environ. Monit. Assessment*, vol. 193, no. 2, pp. 1–32, Jan. 2021.
[15] Z. Wang *et al.*, "Analysis and compensation of benchmark drift of micromachined thermal wind sensor caused by packaging asymmetry," *IEEE Trans. Ind. Electron.*, vol. 69, no. 1, pp. 950–959, Jan. 2022.
[16] K. Huang, X. Zhang, and N. Karimi, "Real-time prediction for IC aging based on machine learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 12, pp. 4756–4764, Dec. 2019.
[17] T. Chen, B. Lin, H. Geng, S. Hu, and B. Yu, "Leveraging spatial correlation for sensor drift calibration in smart building," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 40, no. 7, pp. 1273–1286, Jul. 2021.
[18] N. H. Motlagh *et al.*, "Feasibility of low-cost sensors for pollution hot-spot detection," in *Proc. IEEE 17th Int. Conf. Ind. Informat.*2019, pp. 1083–1090.
[19] T. Nakayama, Y. Matsumi, K. Kawahito, and Y. Watabe, "Development and evaluation of a palm-sized optical PM2.5 sensor," *Aerosol Sci. Technol.*, vol. 52, no. 1, pp. 2–12, 2018.
[20] M. A. Zaidan, D. Wraith, B. E. Boor, and T. Hussein, "Bayesian proxy modelling for estimating black carbon concentrations using white-box and black-box models," *Appl. Sci.*, vol. 9, no. 22, 2019, Art. no. 4976.
[21] K. Gu, J. Qiao, and W. Lin, "Recurrent air quality predictor based on meteorology-and pollution-related factors," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3946–3955, Sep. 2018.

[22] B. C. McDonald *et al.*, "Volatile chemical products emerging as largest petrochemical source of urban organic emissions," *Science*, vol. 359, no. 6377, pp. 760–764, 2018.
[23] R. Valente de Almeida and P. Vieira, "Forest fire finder-DOAS application to long-range forest fire detection," *Atmos. Meas. Techn.*, vol. 10, no. 6, pp. 2299–2311, 2017.
[24] W. Sun, S. Lei, L. Wang, Z. Liu, and Y. Zhang, "Adaptive federated learning and digital twin for industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5605–5614, Aug. 2021.
[25] C.-T. Chiang, "Design of a high-sensitivity ambient particulate matter 2.5 particle detector for personal exposure monitoring devices," *IEEE Sensors J.*, vol. 18, no. 1, pp. 165–169, Oct. 2017.
[26] J. Madureira *et al.*, "Assessment of indoor air exposure at residential homes: Inhalation dose and lung deposition of PM10, PM2. 5 and ultrafine particles among newborn children and their mothers," *Sci. Total Environ.*, vol. 717, 2020, Art. no. 137293.
[27] N. H. Motlagh *et al.*, "Transit pollution exposure monitoring using low-cost wearable sensors," *Transp. Res. Part D, Transp. Environ.*, vol. 98, 2021, Art. no. 102981.
[28] K. Muhammad, S. Khan, V. Palade, I. Mehmood, and V. H. C. De Albuquerque, "Edge intelligence-assisted smoke detection in foggy surveillance environments," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1067–1075, Feb. 2020.
[29] A. Gaur *et al.*, "Fire sensing technologies: A review," *IEEE Sensors J.*, vol. 19, no. 9, pp. 3191–3202, May 2019.
[30] L. Dong *et al.*, "The gas leak detection based on a wireless monitoring system," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6240–6251, Dec. 2019.
[31] E. Kusmierek and E. Chrzescijanska, "Atmospheric corrosion of metals in industrial city environment," *Data Brief*, vol. 3, pp. 149–154, 2015.

Martha Arbayani Zaidan (Member, IEEE) received the Ph.D. degree in automatic control and systems engineering from Sheffield University, Sheffield, U.K., in 2014.

He was a Postdoctoral Research Associate with Maryland University, College Park, MD, USA, and a Fellow with Aalto University, Espoo, Finland. Currently, he acts as a Research Associate Professor and a Researcher with Nanjing University and Institute for Atmospheric and Earth System Research (INAR), Helsinki University, respectively. His research interests include artificial intelligence and machine learning for intelligent control systems, health monitoring technologies, applied physics, atmospheric, and environmental sciences.

Naser Hossein Motlagh received the D.Sc. degree in networking technology from the School of Electrical Engineering, Aalto University, Espoo, Finland, in 2018.

He is currently a Postdoctoral Researcher with the Department of Computer Science, the University of Helsinki within the Nokia Center for Advanced Research (NCAR). Previously, he was a Postdoctoral Fellow within Helsinki Institute for Information Technology (HIIT)—Helsinki Center for Data Science (HiDATA) program. His research interests include the Internet of Things, wireless sensor networks, environmental sensing, smart buildings, and unmanned aerial and underwater vehicles.

Pak Lun Fung received the Ph.D. degree with dissertation on proxy derivation of air-quality parameters in urban regions from the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, Helsinki, Finland, in January 2022.

He is currently conducting air-quality research in INAR, specialized in low-cost sensor measurements and statistical modeling using machine learning techniques. He is also interested in data visualization and science communication.

Abedalaziz S. Khalaf received the master's degree in physics from Environmental and Atmospheric Research Laboratory (EARL), University of Jordan, Amman, Jordan, in 2021.

He works on low-cost sensors to improve their accuracy by comparing with standard instruments and calibrating them by regression methods. His research interests include health exposure, aerosol measurements and artificial intelligence techniques.

Yutaka Matsumi is currently a Professor Emeritus with the Institute for Space–Earth Environmental Research, Nagoya University, Nagoya, Japan. His works include the development of measuring devices, based on new principles, for aerosols and trace gas elements in the atmosphere. These devices are then used for atmospheric measurement, enabling clarification for the reactive processes and fluctuations of trace materials in the atmosphere, and studying the impact of these on the Earth's environment.

Aijun Ding received the Ph.D. degree in meteorology from Nanjing University, Nanjing, China, in 2004.

He is currently a Professor of atmospheric environment and atmospheric physics and the Dean with the School of Atmospheric Sciences, Nanjing University, China. He had about 8-year research experiences in the Hong Kong Polytechnic University during 2011–2009, as Research Assistant, Post-Doc, and Research Fellow, and has become a Full Professor with Nanjing University, since 2009. His research interests include air pollution-weather/climate interactions, tropospheric ozone, chemical transport modeling and Lagrangian dispersion modeling.

Sasu Tarkoma (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Helsinki, Helsinki, Finland, in 2006.

He is currently a Professor of computer science with the University of Helsinki. He is a Visiting Professor with the 6 G Flagship at the University of Oulu. He has authored four textbooks and has authored or coauthored more than 250 scientific articles. He holds 10 granted U.S. patents. His research interests include Internet technology, distributed systems, data analytics, and mobile and ubiquitous computing.

Tuukka Petäjä received the Ph.D. degree and Docent title in physics from the University of Helsinki, Helsinki, Finland, in 2006 and 2011, respectively.

He was a Postdoctoral Researcher with the U.S. National Center for Atmospheric Research (NCAR), Boulder, CO, USA. He is currently a Professor of experimental atmospheric sciences with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. He and his team are currently in charge of the development of aerosol particles measuring equipment for continuous measurements. His research interest includes atmospheric aerosol particles and their role in climate change and air quality.

Markku Kulmala received the Ph.D. degree in theoretical physics from the University of Helsinki, Helsinki, Finland, in 1988.

He is currently an Academy Professor and the Head of Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. He is the Founder of the International (Station for Measuring Ecosystem-Atmosphere Relations) SMEAR observation networks. He has authored or coauthored more than 800 SCI articles (including more than 40 articles in Science and Nature). His research interests include atmospheric aerosol nucleation and growth mechanisms, kinetics of atmospheric aerosols and clusters, and biosphere-aerosol-cloud-climate interactions.

Prof. Kulmala is currently a member of the Academy of Europe and a Foreign Academician of the Chinese Academy of Sciences.

Tareq Hussein received the Ph.D. degree from the Division of Atmospheric Sciences, University of Helsinki, Helsinki, Finland, in 2005, and the Docent title from the University of Helsinki, in 2008, both in physics.

He is currently a Professor of Physics with the University of Jordan, Amman, Jordan, and a Visiting Professor with the Institute for Atmospheric and Earth System Research (INAR), University of Helsinki. His research interests include urban and indoor air quality and exposure.