Measurement 230 (2024) 114529

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Field calibration of low-cost particulate matter sensors using artificial neural networks and affine response correction

Slawomir Koziel [a,b,*], Anna Pietrenko-Dabrowska [b], Marek Wojcikowski [b], Bogdan Pankiewicz [b]

[a] Engineering Optimization & Modeling Center, Reykjavik University, 102 Reykjavik, Iceland
[b] Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland

ARTICLE INFO

Keywords:
Air quality monitoring
Low-cost sensors
Neural networks
Particulate matter pollution
Sensor calibration
Surrogate modeling

ABSTRACT

Due to detrimental effects of atmospheric particulate matter (PM), its accurate monitoring is of paramount importance, especially in densely populated urban areas. However, precise measurement of PM levels requires expensive and sophisticated equipment. Although low-cost alternatives are gaining popularity, their reliability is questionable, attributed to sensitivity to environmental conditions, inherent instability, and manufacturing imperfections. The objectives of this paper include (i) introduction of an innovative approach to field calibration for low-cost PM sensors using artificial intelligence methods, (ii) implementation of the calibration procedure involving optimized artificial neural network (ANN) and combined multiplicative and additive correction of the low-cost sensor readings, (iii) demonstrating the efficacy of the presented technique using a custom-designed portable PM monitoring platform and reference data acquired from public measurement stations. The results obtained through comprehensive experiments conducted using the aforementioned low-cost sensor and reference data demonstrate remarkable accuracy for the calibrated sensor, with correlation coefficients of 0.86 for PM_1 and $PM_{2.5}$, and 0.76 PM_{10} (particles categorized as having diameter equal to or less than 1 μm, 2.5 μm, and 10 μm, respectively), along with low RMSE values of only 3.1, 4.1, and 4.9 μg/m^3.

1. Introduction

Researchers and healthcare experts widely agree on the detrimental impact of air pollution on human health and morbidity. Recent studies indicate that nearly nine million deaths occur globally each year due to ambient air pollution [1,2]. According to the European Environmental Agency (EEA) [3], atmospheric pollution stands out as the foremost environmental health risk factor in Europe. This exerts a substantial impact on the health of the European population, especially in urban areas, by causing chronic diseases and premature deaths [4]. Particulate matter (PM) is one of the common types of pollution characterized by the presence of tiny particles (or droplets) originating from various sources such as vehicle emissions, industrial processes, construction activities, and natural sources like wildfires and dust storms. Most of the existent studies on PM-related pollutants primarily focus on fine $PM_{2.5}$ particles with a diameter smaller than 2.5 μm. These are considered the most detrimental because, upon inhalation, they can penetrate deeply into the lungs [5]. Numerous studies consistently associate $PM_{2.5}$ with elevated occurrence of cardiovascular illnesses [6,7], various types of cancerous diseases [8,9], and preterm births [10]. It has been assessed that in 2020, 96 % of the urban population in the European Union (EU) was exposed to fine particulate matter ($PM_{2.5}$) concentrations above the World Health Organization (WHO) recommended level of 5 Âμg per cubic meter (μg/m^3) of air [3]. The latest estimates from EEA in 2020 indicate that the elevated $PM_{2.5}$ levels led to over 235,000 premature deaths in Europe [3]. Furthermore, it has been assessed that the exposure to $PM_{2.5}$ has been a cause of nearly 176,000 years of life with disability due to chronic obstructive pulmonary disease in 30 European countries [3]. The research studies indicate the two main sources of $PM_{2.5}$ [11]: residential combustion (contributing to approximately 45 % of $PM_{2.5}$ mass) and traffic-related emissions (around 30 % of $PM_{2.5}$ mass). The origins of the remaining quarter of $PM_{2.5}$ mass include mineral dust from construction works, and high-temperature processes, with steel processing as a main contributor [11].

Nowadays, accurate monitoring of $PM_{2.5}$ is carried out at fixed locations using government-authorized reference stations. Most acknowledged measurement methods are filter-based gravimetric techniques. Therein, particles are gathered on filters, stabilized, and

* Corresponding author.
E-mail addresses: koziel@ru.is (S. Koziel), anna.dabrowska%20@pg.edu.pl (A. Pietrenko-Dabrowska), marwojci1@pg.edu.pl (M. Wojcikowski), bogpanki%20@pg.edu.pl (B. Pankiewicz).

https://doi.org/10.1016/j.measurement.2024.114529
Received 9 February 2024; Received in revised form 17 March 2024; Accepted 18 March 2024
Available online 19 March 2024

subsequently their mass is measured in the accredited environmental laboratories by weighing the filters prior to and post-sampling. The filter-based approach offers accuracy and precision, but the measurements are costly, time-consuming, and offer limited spatiotemporal information, which makes it impossible to fully grasp ambient $PM_{2.5}$ concentrations [12,13]. Therefore, costly governmental stations typically act as a reference for calibrating data acquired by inexpensive mobile sensors.

Low-cost sensors (LCS) show great promise as tools for evaluating exposure to air pollution with increased spatial resolution [14,15,76]. The utilization of portable and more economical sensors offers significant potential benefits, which include the possibility of employing a greater number of sensors to enhance spatial coverage, simple usage and maintenance, decreased energy demand, and ease of relocation according to researcher's needs. LCSs are utilized standalone or supplement the existing governmental air quality monitoring facilities [16]. They may be incorporated in densely allocated stationary networks [17], mobile networks mounted on vehicles [18,19,79], or in the form of wearable devices [20–22]. Interesting studies on evaluating the quality of low-cost air quality sensors can be found in [73–75].

Most low-cost devices for PM monitoring rely on optical measurement techniques. Unfortunately, optical particle sensors exhibit numerous drawbacks, such as inferior accuracy, poor reliability and repeatability of measurements, as well as the need for calibration [23–26,77,78]. While the equipment of the reference facilities measures mass concentrations directly, the optical methods render merely an estimate of the mass. In general, the fundamental operating principle of the economical particulate matter sensors designed for commercial applications relies on light scattering which is fast but inaccurate [27]. Therefore, the researchers need to account for the observed considerable disparities between PM mass estimates performed using LCS and reference stations [28]. One of the reasons for these discrepancies is the fact that the elevated relative humidity can cause hygroscopic particle growth, thereby leading to an overestimation of a dry mass [29–31]. A possible mitigation involves accurate assessment of particle hydration and its incorporation in the calibration process, but also desiccation of particles while measuring. Yet another challenge is a consequence of inability of LCSs to detect particles featuring diameters below a certain threshold value, the reason being the utilized wavelength of the laser light, typically around 0.3 μm. Unfortunately, the peak in the size distribution of pollution particles is predominantly observed for particles below 0.3 μm. Furthermore, studies carried out in laboratory-controlled environments suggest significant variation in the accuracy of pollutant concentrations among optical sensors [32]. Consequently, it is indispensable to carry out a proper calibration prior to any responsible usage of data obtained from LCSs.

The boost in the sensor market instigated a surge in the development of calibration methods. Arguably the simplest techniques involve linear regression, where only the sensor data is employed as the input variable [33–35]. A somewhat more intricate approach constitutes multivariate linear regression, in which additional input variables like temperature, and humidity are incorporated for LCS calibration [36–39]. Another option offers gain-offset model [40–42], a subclass of linear regression models, which assesses the difference between the actual pollutant level and the one measured by LCS in terms of both additive and multiplicative bias. Nevertheless, such simple methods show significant limitations, as they are incapable of accounting for sensors' nonlinearities [43]. A considerable potential in addressing nonlinearities exhibit machine learning (ML) techniques [44–46]. Among most popular methods, random forests [47,48], gradient boosting algorithms [49,50] or support vector regression [51,52] should be mentioned. Nevertheless, recently, the employment of various types of neural networks (NNs) for calibration purposes of LCS-collected PM data has been steadily gaining in popularity. A multitude of calibration frameworks has been devised, which employ, e.g., Feedforward Neural Networks (FNN) [53], Recurrent Neural Networks (RNN) [54,55], Long Short-Term Memory Neural Networks (LTSM NNs) [56,57], convolutional neural networks (CNN) [58,59], as well as multi-layer perceptrons [72].

This study presents an inventive methodology for reliably correcting low-cost PM sensors. The proposed methodology combines multiplicative and additive corrections of the PM readings rendered by the low-cost sensor with the coefficients predicted using an artificial neural network (ANN) surrogate. Calibration inputs encompass environmental parameters such as temperature and humidity. Optimization of ANN hyper-parameters and calibration factors, specifically hidden layer size and affine scaling coefficients, is carried out to bolster the generalization capability of the surrogate. The proposed calibration technique has been comprehensively validated by applying it to a portable measurement platform developed at Gdansk University of Technology, Poland. The PM sensors on the platform are field calibrated using reference data acquired from several public monitoring stations in the city of Gdansk. The results demonstrate excellent accuracy in sensor correction, with achieved correlation coefficients of 0.86 for PM_1 and $PM_{2.5}$, and 0.76 for PM_{10}, respectively. Additionally, the RMSE values are only 3.1, 4.1, and 4.9 $\mu g/m^3$. This high level of precision establishes the calibrated sensor as a cost-effective solution for monitoring particulate matter pollution in urban areas.

The originality and the technical contributions of this work can be summarized as follows: (i) development of an innovative low-cost sensor technique which combines multiplicative and additive response correction with optimizable coefficient controlling the balance between the two, (ii) development of ANN-based calibration model with optimized architecture, and the complete calibration framework utilizing the aforementioned algorithmic tools, (iii) demonstrating the relevance of the presented calibration scheme for PM of three different types and high efficacy of the calibration process leading to correlation coefficients with the reference data close to 0.9 (for PM_1 and $PM_{2.5}$) and close to 0.8 for PM_{10}.

The remaining part of the manuscript is organized as follows. Section 2 introduces the portable low-cost particulate matter monitoring platform. Section 3 elaborates on acquisition of the reference data. The proposed calibration methodology is introduced in Section 4. Section 5 describes experimental setup, gathers the results and provides their discussion. Finally, Section 6 concludes the work.

2. Low-Cost particulate matter sensor. Design and properties

This section outlines the specifics of a portable system designed for monitoring particulate matter (PM) using low-cost PM sensors. It includes a description of the hardware design, embedded software, and the sensors integrated into the system. The calibration procedure for the PM sensor will be discussed in detail in Section 4.

2.1. Portable monitoring system. Hardware design

A low-cost SPS30 sensor from Sensirion [60] has been chosen to measure the concentration of particulate matter. It is a miniature (41 × 41 × 12 mm) optical sensor measuring PM particles in the range of 0–1000 $\mu g/m^3$, distinguishing several PM categories: PM_1, $PM_{2.5}$, PM_4, and PM_{10}. The accuracy of this sensor, according to the manufacturer's data, is ± 10 $\mu g/m^3$ for PM_1 and $PM_{2.5}$ and ± 25 $\mu g/m^3$ for PM_4 and PM_{10}. The sensor operates based on the principle of measuring scattered laser light on airborne pollutant particles within a fan-controlled airflow. While this concept appears straightforward, its practical implementation is intricate. Optical measurements effectively count and size particles, but accurately converting this data into concentration, expressed as weight per unit volume, necessitates complex processing algorithms. These algorithms consider factors such as particle shape and color, leading to limited accuracy. In outdoor settings, where environmental conditions like temperature, pressure, and humidity fluctuate, the sensor's accuracy may deviate from the manufacturer's declared specifications. The SPS30 sensor boasts a lifespan of ten years under

continuous operation, with a startup time of 30 s. Equipped with a built-in fan to facilitate air transportation, the sensor requires a weekly fan cleaning cycle for optimal performance. Fig. 1 illustrates both the external and internal views of the SPS30 sensor.

The PM sensor is powered by a 5 V power supply. The current consumption according to the catalog data is 45–65 mA for the measurement mode, 330 μA for the idle mode, and less than 50 μA for the sleep mode. Reading data from the sensor and controlling the sensor can be done via the Universal Asynchronous Receiver-Transmitter (UART) and Inter-Integrated Circuit (I2C) interfaces (user configurable); for the purposes of this project, the UART interface was used to communicate with the sensor.

The environmental conditions in which the PM sensor operates are monitored by a miniature, high-accuracy BME280 temperature, pressure, and humidity sensors from Bosch Sensortec [61]. One of the BME280 sensors was mounted on the edge of the housing to measure ambient conditions as closely as possible. The sensor is equipped with an I2C interface and is connected directly to the embedded computer board. Another BME280 sensor is mounted inside the enclosure, near the computer board and PM sensor and it is dedicated to measure the internal temperature and humidity, as they differ from the outside conditions due to heating generated by the processing hardware.

The autonomous embedded computer system has been provided to support the measurement procedure of the low-cost PM sensor used in the system. The main part of the computer system is a Linux-based all-in-one robotics computer Bealgebone® Blue [62] integrated on a small (87 x 55 mm) printed circuit board. The Beaglebone® computer contains an ARM®-based microprocessor along with many peripheral devices such as: Wi-Fi, Bluetooth modules, Li-Po battery charging monitor with a 2-cell balancer, multiple UART, Serial Peripheral Interface (SPI), I2C and General-Purpose Input/Output (GPIO) ports. The microprocessor mounted on Beaglebone board is OSD3358 System-in-Package integrated circuit from Octavo Systems [63] build around AM3358 processor from Texas Instruments [64]. It consists of an ARM Cortex-A8 processor clocked at 1 GHz, 32 KB L1 cache, 256 KB L2 cache, 64 KB RAM and 176 KB ROM, PRU-ICSS coprocessor (programmable real-time units), 3D graphics accelerator, encryption accelerator and many on-chip peripheral elements (serial and parallel interfaces, memory interfaces, real-time clock, watchdog, etc.). The block diagram of the Beaglebone computer has been shown in Fig. 2.

A Li-Po battery consisting of two cells with a total output voltage of 7.4 V and a capacity of 4400mAh is connected to the computer board, ensuring autonomous operation of the system for at least 24 h. However, an external power supply should be used anyway. By default, the BeagleBone® Blue computer is equipped with a Wi-Fi wireless network module, but to ensure long-distance communication, a cellular network modem u-GSM shield from itbrainpower.net [65] with an installed BG96 module from Quectel [66] on a small PCB of approximate dimensions of 45 × 27 mm was added to the system. The modem supports communication compatible with Long-Term Evolution (LTE), Frequency-Division Duplexing (FDD), Time-Division Duplexing (TDD) and Enhanced General Packet Radio Service (EGPRS) 800/900/1800/1900 MHz wireless

Fig. 1. Low-cost PM sensor SPS30 from Sensirion [60]: (a) top view, (b) sensor's internals.

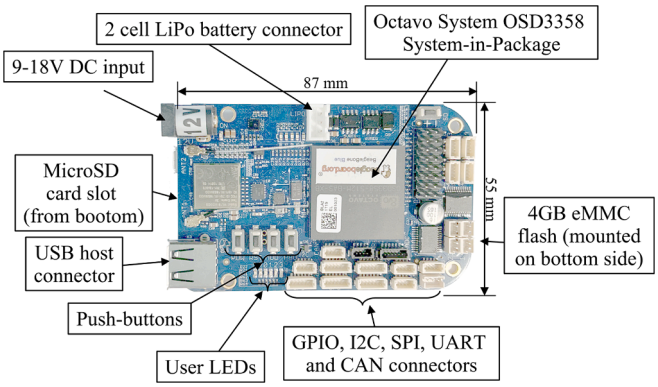


Fig. 2. Block diagram of the Beaglebone Blue embedded computer board. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

broadband communication standards for mobile devices. In addition to communication, the BG96 modem also supports the following satellite navigation systems: Galileo, Global Positioning System (GPS), Global Navigation Satellite System (GLONASS) and BeiDou/Compass. These are used to provide the geographical coordinates of the location during each measurement. The *1nce* operator [67] was chosen to provide communication, which provides low-bandwidth communication dedicated to IoT devices in 165 countries in Europe, Asia, North America, South America and Oceania for a one-time initial fee, enabling operation in 2G, 3G, 4G/LTE-M and NB-IoT networks (depending on the country) for 10 years. The internals of the measurement platform have been shown in Fig. 3.

The computer, modem, environmental parameter sensor, PM sensor module, GSM antenna, battery, and other components were affixed to a base mounting plate constructed using a 3D printer. This base plate was then enclosed within a housing (see Fig. 4), also created through 3D printing, featuring dedicated mounting brackets. The housing design facilitates easy removal of the base plate for potential repairs or diagnostics. Engineered to withstand various weather conditions, the casing was crafted from PET-G material, ensuring durability and resilience.

2.2. *Portable monitoring system. Software*

The Beaglebone computer was equipped with the Ubuntu Linux version 18.04 LTS operating system. To facilitate hardware interfacing with the PM sensor, environmental sensor, and modem, dedicated drivers were developed in C and Python. The primary PM measurement software was coded in Python, utilizing multi-threaded functions with

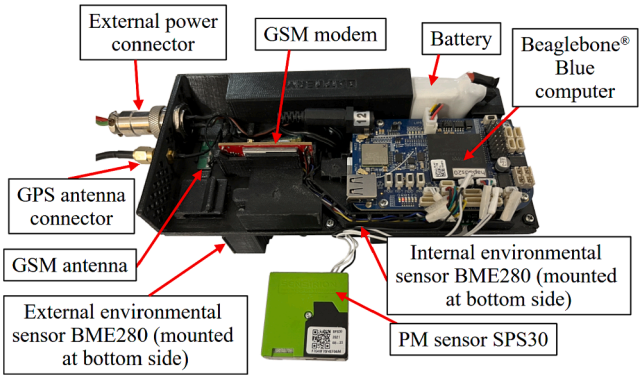


Fig. 3. Internals of the system with Beaglebone Blue embedded computer and PM sensor SPS30. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

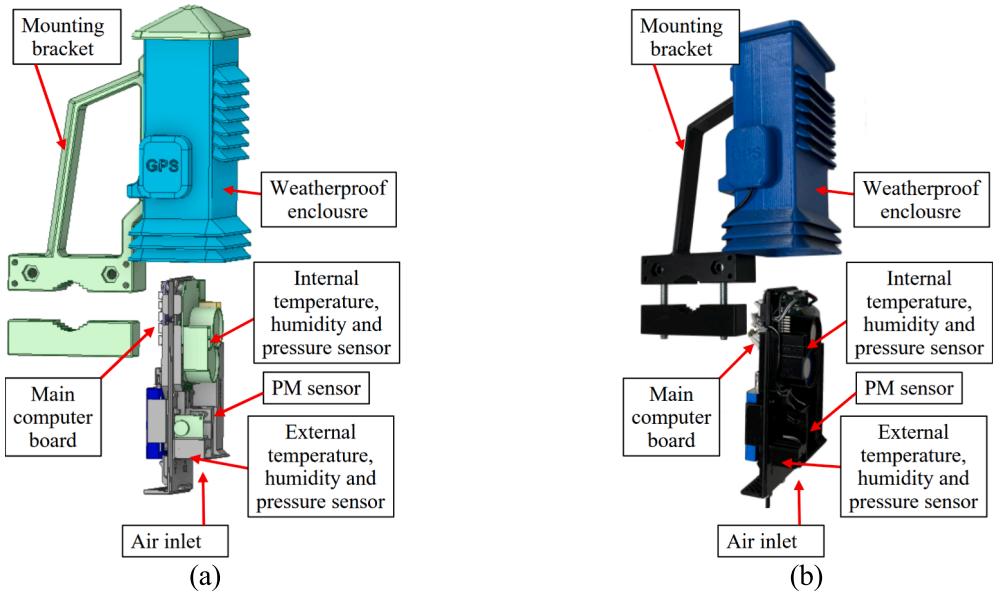


Fig. 4. Assembled system with weatherproof enclosure: (a) - project, (b) - real picture.

mutexes serving as an inter-process synchronization mechanism. Within the main software module, Python class instances representing software objects were instantiated and linked to the hardware modules of the PM sensor, temperature sensor, and GPS unit, as illustrated in Fig. 5.

These objects were endowed with methods enabling their utilization, with mutexes employed to safeguard against concurrent access attempts to hardware resources and their corresponding drivers. The primary software routine orchestrates the following tasks on a regular basis: (i) preparing the PM sensor for measurement, (ii) conducting PM, temperature, pressure, and humidity measurements every hour while retrieving geographical coordinates from the GPS system, and (iii) transmitting the measurement results to a cloud data collection platform via a GSM modem. Simultaneously, the PM sensor undergoes a cleaning procedure mandated by the device manufacturer every 24 h. A simplified depiction of these activities is provided in Fig. 6.

3. Reference data acquisition. Public monitoring stations

The low-cost PM sensor-based particulate matter measurement devices, as detailed in Section 2, were deployed directly at the monitoring

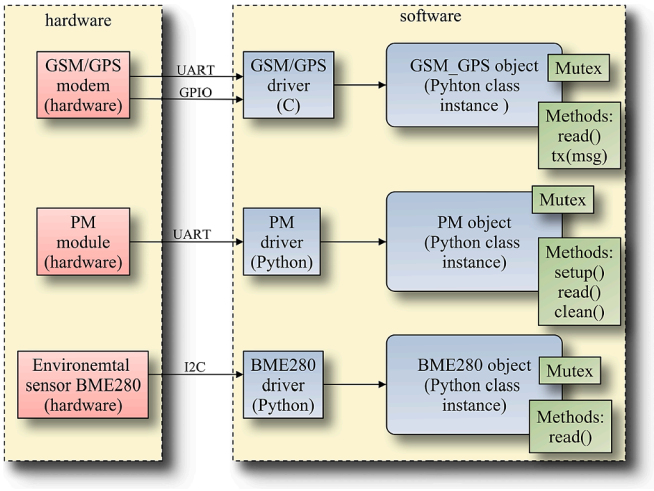


Fig. 5. Simplified structure of hardware-software interaction and software objects with relevant classes and methods.

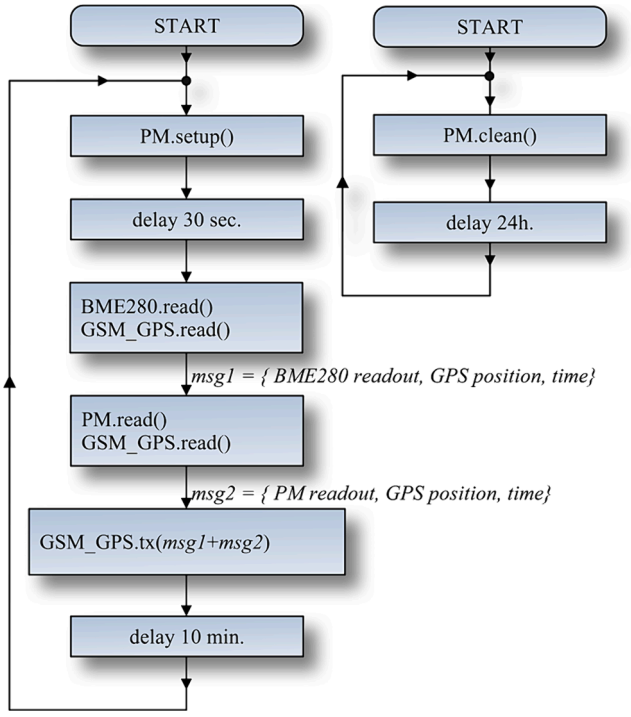


Fig. 6. The flow diagram of software operation to perform measurements and sensor maintenance. Note that the operations are performed in loops therefore there are no termination conditions indicated.

stations operated by the ARMAG foundation [68], responsible for air quality surveillance in Gdansk, Poland. Gdansk, with a population of approximately 0.5 million people, ranks as the sixth-largest city in Poland. The ARMAG Foundation manages a network of air-conditioned containers strategically positioned across the city, each equipped with professionally calibrated and serviced measuring instruments. At these stations, the specific sensors employed are GRIMM 180 analysers, as depicted in Fig. 7. These analysers serve as the reference data source for the field calibration of the low-cost sensors.

Air pollution measurements, along with various environmental parameters such as temperature, humidity, wind speed, and sunlight, are

Fig. 7. Equipment for reference PM measurement (Gimm 180 analyser) installed in the reference Station 3 of ARMAG foundation.

conducted hourly at each reference station. The results of these measurements are published daily on the foundation's website, accessible to the public free of charge.

To gather reference data for this publication, specialized software was created. This software runs on a server hosted at the Gdansk University of Technology. It regularly reads the website and saves the collected reference data into a CSV file.

The tested measurement devices, as described in Section 2, were installed near the reference stations. They were placed either on the roof or adjacent to the containers, depending on the configuration of the local station. This setup allowed for direct comparison between the measurement results of the tested devices and those of the reference devices operated by the ARMAG foundation. An illustration of the installation of one of the tested devices at a reference station is provided in Fig. 8. The measurements conducted by the tested devices followed the same hourly intervals as the reference stations, and the results were transmitted via GSM modems to a cloud-based data collection platform. Subsequently, the data from the cloud was downloaded into a CSV file for further analysis.

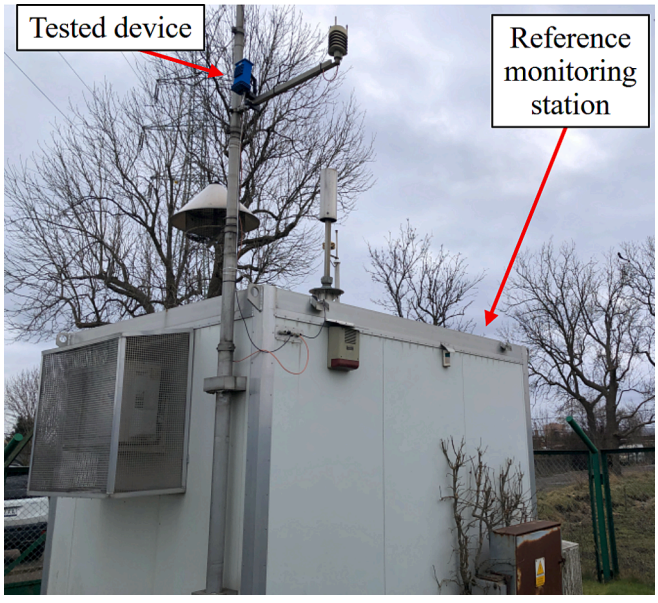


Fig. 8. One of the reference stations of ARMAG foundation (Station 2) with tested device installed on the mast on top of the container.

Fig. 9 illustrates the locations of the measurement sites where the ARMAG network reference stations are situated, alongside the portable monitoring platforms installed as described in Section 2. The measurement campaign spanned two months, between March and May 2023. The reference and measurement data collected during this period were utilized to develop the calibration model detailed in subsequent sections (Section 4 and 5).

4. Field calibration of Low-Cost PM sensor by Machine learning and affine response scaling

This section introduces the proposed machine-learning-based calibration procedure for correcting low-cost PM sensor. The operation of our algorithm will be illustrated using the portable monitoring platform outlined in Section 2, using the reference data acquired as detailed in Section 3. A description of the calibration strategy is preceded by a rigorous formulation of the sensor correction problem in Section 4.1, followed by a definition of a mixed multiplicative and additive response correction in Section 4.2. The artificial neural network (ANN) calibration model is elucidated in Section 4.3 along with a description of the hyper-parameter optimization process. The complete operating flow of PM measurement using calibrated low-cost sensor is covered in Section 4.4.

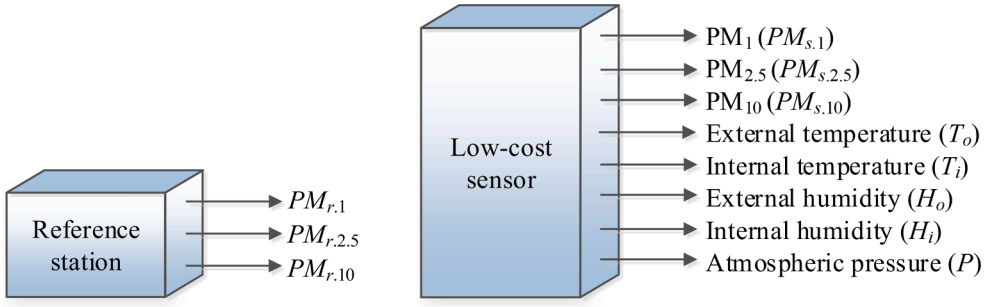
4.1. Low-Cost PM sensor calibration. Notation and terminology. Problem statement

Fig. 10 illustrates data produced by the reference stations and the low-cost sensor of Section 2. As for the reference, we are only interested in the measured level of particulate matter (PM), here denoted as $PM_{r.x}$, where, later, x will be substituted by a specific PM type (1, 2.5, or 10). On the other hand, the low-cost sensor provides the PM readings $PM_{s.x}$, and a set of environmental parameters, namely, internal and external temperature and humidity, as well as atmospheric pressure. As elucidated in Section 2, internal and external conditions are different due to the heating effects associated with the electronic circuitry within the measurement platform. As the sensor operation is affected by the temperature and humidity, utilizing both sets of parameters (i.e., internal and external) as calibration variables may enhance the calibration process reliability. Fig. 10(c) gathers notation utilized to denote the reference and low-cost sensor PM as well as the environmental variables.

The total number of data samples acquired for the purpose of sensor calibration is N, and the dataset is split into the training part (N_b samples), and the testing part (N_t samples). The testing part encapsulated about twenty percent of the total number of samples. The specific number of samples are as follows (cf. Section 5.1 for more details): N = 3756, 5846, and 6410 for PM_1, $PM_{2.5}$, and PM_{10}, respectively; $N_t = 672$ for PM_1, and 1008 for $PM_{2.5}$ and PM_{10}; finally, $N_b = N - N_t$. These numbers are obtained after removing corrupted data from the low-cost sensor. The details concerning data division will be discussed in the verification section (Section 5). The training set was exclusively utilized to identify the calibration model, whereas the testing set was exclusively used for validation purposes. For the purpose of formulating the calibration problem, we denote the reference training samples as $PM_{r.x}^{(b.j)}$, $j = 1, \ldots, N_b$. The low-cost sensor training samples are denoted as $PM_{s.x}^{(b.j)}$, $j = 1, \ldots, N_b$, and $\boldsymbol{v}^{(b.j)}$, $j = 1, \ldots, N_b$ (environmental data). The testing data will be denoted as $PM_{r.x}^{(t.j)}$, $j = 1, \ldots, N_t$, $PM_{s.x}^{(t.j)}$, $j = 1, \ldots, N_t$, and $\boldsymbol{v}^{(t.j)}$, $j = 1, \ldots, N_t$, respectively.

In the following, the low-cost sensor calibration model will be denoted as $C(PM_{s.x},\boldsymbol{v};\boldsymbol{p})$, where $\boldsymbol{p}$ is an aggregated vector of model parameters (e.g., hyper-parameters of the neural network, cf. Section 4.3). The calibration model output represents the corrected low-cost sensor measurement. Parameter vector $\boldsymbol{p}$ is identified to fit the calibration model to the reference data over the entire training set. The fitting process is carried out to minimize the MSE error, i.e., we have

Fig. 9. Location of the reference stations in the city of Gdansk. Stations 1 through 5 will be referred to as AM1, AM2, AM3, AM7, and AM8 later in the paper, according to the ARMAG's nomenclature.



(a) (b)

Data source	Data type	Symbol	Comment
Reference station	PM_1	$PM_{r.1}$	Denoted as $PM_{r.x}$ when no specific PM type is referred to
	$PM_{2.5}$	$PM_{r.2.5}$	
	PM_{10}	$PM_{r.10}$	
Low-cost sensor (portable measurement platform of Section 2)	PM_1	$PM_{s.1}$	Denoted as $PM_{s.x}$ when no specific PM type is referred to
	$PM_{2.5}$	$PM_{s.2.5}$	
	PM_{10}	$PM_{s.10}$	
	External temperature	T_o	Collectively denoted as a vector of environmental conditions $\boldsymbol{v} = [T_o\ T_i\ H_o\ H_i\ P]^T$
	Internal temperature	T_i	
	External humidity	H_o	
	Internal humidity	H_i	
	Atmospheric pressure	P	

(c)

Fig. 10. Description of data produced by the reference stations and the low-cost portable monitoring platform of Section 2: (a) PM_x reading from the reference station; (b) PM_x reading from the low-cost sensor under calibration. The sensor also produces auxiliary outputs: external and internal temperature (T_o and T_i, respectively), external and internal humidity (H_o and H_i respectively), and atmospheric pressure (P); (c) symbols of data produced by the reference station and low-cost sensor.

$$\mathbf{p}^* = \underset{\mathbf{p}}{\operatorname{argmin}} \frac{1}{N_b} \sum_{j=1}^{N_b} \left(PM_{r.x}^{(b.j)} - C\left(PM_{s.x}^{(b.j)}, \mathbf{v}^{(b.j)}, \mathbf{p}\right) \right)^2 \tag{1}$$

where $\boldsymbol{p}^*$ is the optimum set of calibration model parameters to be found.

The corrected low-cost sensor output will be denoted as

$$PM_{corr.x} = C(PM_{s.x}, \mathbf{v}, \mathbf{p}^*) \tag{2}$$

Reliability of the calibration process will be validated using the correlation coefficient between the reference and corrected low-cost sensor, and the RMSE error, both computed for the testing data.

In particular, we have

$$r = \frac{\sum_{j=1}^{N_t} \left(PM_{corr.x}^{(t.j)} - \overline{PM}_{corr.x}\right)\left(PM_{r.x}^{(t.j)} - \overline{PM}_{r.x}\right)}{\sqrt{\sum_{j=1}^{N_t} \left(PM_{corr.x}^{(t.j)} - \overline{PM}_{corr.x}\right)^2 \sum_{j=1}^{N_t} \left(PM_{r.x}^{(t.j)} - \overline{PM}_{r.x}\right)^2}} \tag{3}$$

where

$$\overline{PM}_{r.x} = \frac{1}{N_t}\sum_{j=1}^{N_t} PM_{r.x}^{(t.j)} \overline{PM}_{corr.x} = \frac{1}{N_t}\sum_{j=1}^{N_t} PM_{corr.x}^{(t.j)} \tag{4}$$

and

$$RMSE = \sqrt{\sum_{j=1}^{N_t} \frac{\left(PM_{corr.x}^{(t.j)} - PM_{r.x}^{(t.j)}\right)^2}{N_t}} \tag{5}$$

4.2. Mixed multiplicative and additive response correction of Low-Cost sensor

The calibration approach adopted in this study for the low-cost sensor involves a combination of additive and multiplicative response corrections, akin to affine scaling. This scheme introduces additional degrees of freedom, which could enhance the reliability of the calibration process.

The basic low-cost sensor correction scheme is arranged as follows:

$$PM_{corr.x} = A_m(PM_{s.x} + A_a) \tag{6}$$

where $PM_{corr.x}$ is the corrected sensor reading. Using (6), the calibration function $C(PM_{s.x},\boldsymbol{v},\boldsymbol{p})$ takes a more specific form of

$$C(PM_{s.x}, \mathbf{v}, \mathbf{p}) = A_m(PM_{s.x}, \mathbf{v}, \mathbf{p})(PM_{s.x} + A_a(PM_{s.x}, \mathbf{v}, \mathbf{p})) \tag{7}$$

Using the above notation, the regression problem (1) takes the form of

$$\mathbf{p}^* = \underset{\mathbf{p}}{\operatorname{argmin}} \frac{1}{N_b} \|\mathbf{PM}_r - \mathbf{C}(\mathbf{p})\|^2 \tag{8}$$

where

$$\mathbf{PM}_r = \begin{bmatrix} PM_{r.x}^{(b.1)} \\ \vdots \\ PM_{r.x}^{(b.N_b)} \end{bmatrix} \tag{9}$$

and

$$\mathbf{C}(\mathbf{p}) = \begin{bmatrix} C(PM_{s.x}^{(b.1)}, \mathbf{v}^{(1)}, \mathbf{p}) \\ \vdots \\ C(PM_{s.x}^{(b.N_b)}, \mathbf{v}^{(N_b)}, \mathbf{p}) \end{bmatrix} = \begin{bmatrix} A_m(PM_{s.x}^{(b.1)}, \mathbf{v}^{(1)}, \mathbf{p})\left[PM_{s.x}^{(b.1)} + A_a(PM_{s.x}^{(b.1)}, \mathbf{v}^{(1)}, \mathbf{p})\right] \\ \vdots \\ A_m(PM_{s.x}^{(b.N_b)}, \mathbf{v}^{(N_b)}, \mathbf{p})\left[PM_{s.x}^{(b.N_b)} + A_a(PM_{s.x}^{(b.N_b)}, \mathbf{v}^{(N_b)}, \mathbf{p})\right] \end{bmatrix} \tag{10}$$

It should be noted that solution to (8) is not unique because various combinations of multiplicative and additive correction coefficients may lead to the same outcome. Here, a uniqueness is enforced by introducing a calibration hyper-parameter $0 \le \alpha \le 1$, which controls the distribution of weights between the additive and multiplicative coefficients. The correction coefficients computed over the training data are then given as

$$A_a(PM_{s.x}^{(b.j)}, \mathbf{v}^{(b.j)}, \mathbf{p}) = \alpha\left[PM_{r.x}^{(b.j)} - PM_{s.x}^{(b.j)}\right] \tag{11}$$

and

$$A_m(PM_{s.x}^{(b.j)}, \mathbf{v}^{(b.j)}, \mathbf{p}) = \frac{PM_{r.x}^{(b.j)}}{PM_{s.x}^{(b.j)} + \alpha(PM_{r.x}^{(b.j)} - PM_{s.x}^{(b.j)})} \tag{12}$$

The parameter α will be optimized together with the calibration model hyper-parameters to ensure the best possible correction quality. Also, because the typical variations of the low-cost sensor readings are lower than those of the reference, the recommended value of α is lower than unity, which would translate into scaling coefficient A_m being higher than one. To the best knowledge of the authors, a combination of multiplicative and additive correction in the form similar to that described in this section was not employed so far in any studies available in the literature.

4.3. Calibration model implementation using artificial neural networks

The calibration model of choice in this study is artificial neural network (ANN). As we need a relatively simple regression model, a feedforward ANN, specifically, a multi-layer perceptron (MLP) [69,70], will be employed. Initial experiments indicate that utilization of three fully connected hidden layers ensure sufficient flexibility, while being sufficiently immune against overtraining.

However, the number of neurons in those layers are treated as hyper-parameters, adjusted during model training, as elucidated below. The coefficient α discussed in Section 4.2 is an additional hyper-parameter.

4.3.1. Network architecture

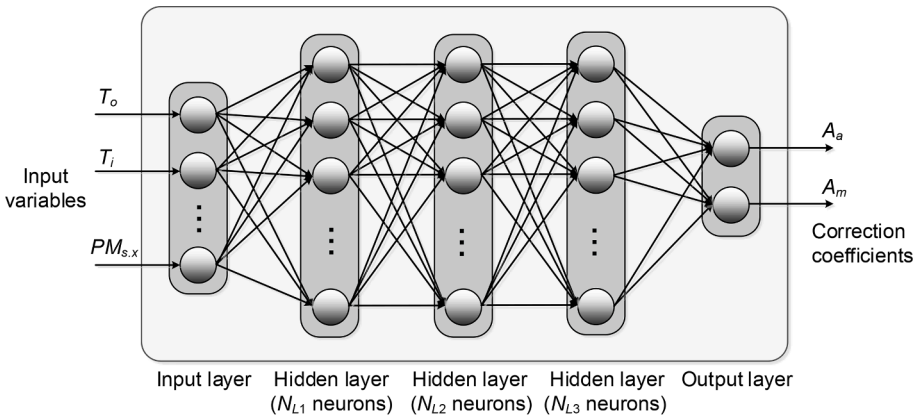
Fig. 11 shows the overall architecture of the MLP model employed in this study along with its training setup. The model inputs are environmental parameters (temperature, humidity, atmospheric pressure) gathered in vector $\boldsymbol{v}$, as well as the PM reading from the low-cost sensor $PM_{s.x}$. Using this data, the calibration model predicts the affine scaling coefficients A_a and A_m. Simplicity of the network architecture is intentional and allows us to smoothen the noise inherent to PM measurement data (pertaining to both the reference and low-cost sensor readings). Furthermore, the model training time is short (a few seconds), which enables evaluation of many different architectural variations, specifically, adjustment of hyper-parameters mentioned in the previous paragraph.

4.3.2. Model identification

Fig. 12 shows the flow diagram of calibration model identification including adjustment of the hyper-parameters. For simplicity, we have three parameters, namely, N_L, M_L, and α, gathered in a vector $\boldsymbol{H} = [N_L\ M_L\ \alpha]^T$. The network architecture, i.e., the number of neurons in the ith layer N_{Li}, $i = 1, 2, 3$, is determined by N_L and M_L as follows:

$$[N_{L1}\ \ N_{L2}\ \ N_{L3}] = \left[I_N(M_L N_L)\ \ N_L\ \ I_N\left(\frac{N_L}{M_L}\right)\right] \tag{13}$$

where $I_N(.)$ is the rounding function (i.e., $I_N(x)$ is the nearest integer to x). Due to fast training of the MLP model, hyper-parameters $\boldsymbol{H}$ are found through exhaustive search within a discrete parameters space defined as all combinations of the sets $N_L \in \{6, 8, 10, 12, 15, 20\}$, $M_L \in \{2, 1, 0.5\}$, and $\alpha \in \{0.6, 0.7, 0.8, 0.9, 1.0\}$. This space corresponds to eighteen different ANN architectures and, independently, five combinations of the affine correction coefficients. For each value of the vector $\boldsymbol{H}$, the neural network is trained twenty times, and the best-performing model



(a)

Calibration model training setup	
Algorithm	Backpropagation Levenberg-Marquardt routine [71]
Activation function	Sigmoid
Maximum number of epochs	1000
Performance evaluation	Mean-square error (MSE)
Training/testing data division	Random

(b)

Fig. 11. Artificial neural network surrogate used as the calibration model: (a) basic model architecture, here, a multi-layer perceptron (MLP) with three fully-connected hidden layers; the number of neurons N_{L1}, N_{L2}, and N_{L3} are treated as hyper-parameters; detailed description of the network inputs (i.e., environmental parameters and low-cost sensor PM readings) can be found in Fig. 10(c); (b) MLP training setup.

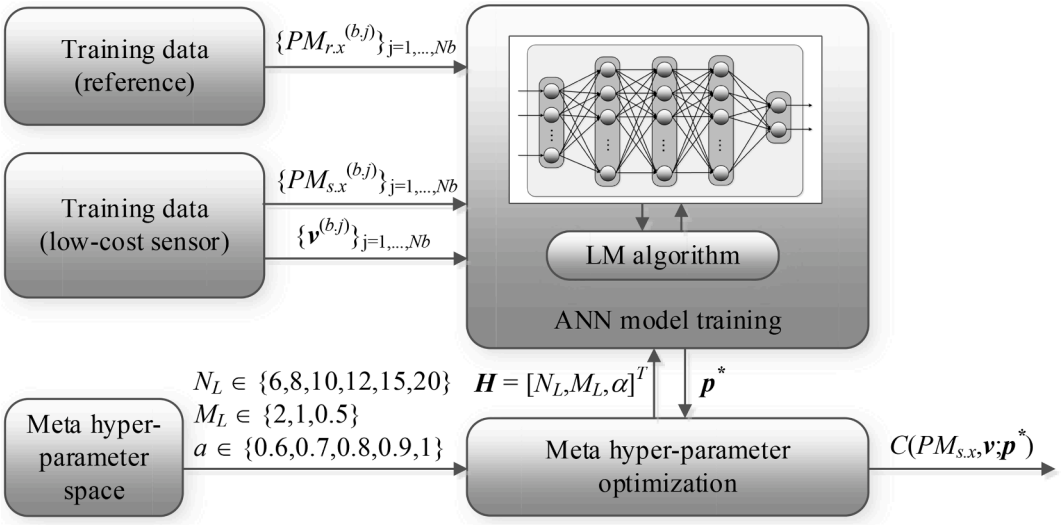


Fig. 12. Identification of the calibration model with concurrent adjustment of the hyper-parameters $\boldsymbol{H} = [N_L\ M_L\ \alpha]^T$. The ANN model is trained using the Levenberg-Marquardt algorithm for all combinations of hyper-parameters to identify the best MLP configuration and the scaling coefficient α. For each value of $\boldsymbol{H}$, the MLP is trained twenty times, and the best-performing model is selected.

is selected. This is necessary because of random data division (into training and testing samples) carried out within the training cycle, which leads to different results for each run.

4.4. *Operating flow of particulate matter detection by means of calibrated sensor*

The operating flow of the sensor calibration process has been shown in Fig. 13. The first step of the process is to obtain predicted correction coefficients A_m and A_a from the MLP calibration model, using the raw PM_x reading from the low-cost sensor and the values of environmental parameters (vector $\boldsymbol{v}$). These coefficients are then applied to compute the calibrated sensor output using affine transformation (6).

5. Results and discussion

In this section, we demonstrate the operation and performance of the proposed calibration scheme, which is applied to the portable monitoring platform outlined in Section 2. The reference data is acquired from the public stations briefly discussed in Section 3. The content of this section is structured as follows. In Section 5.1, we overview the reference and low-cost data, as well as its division into training and

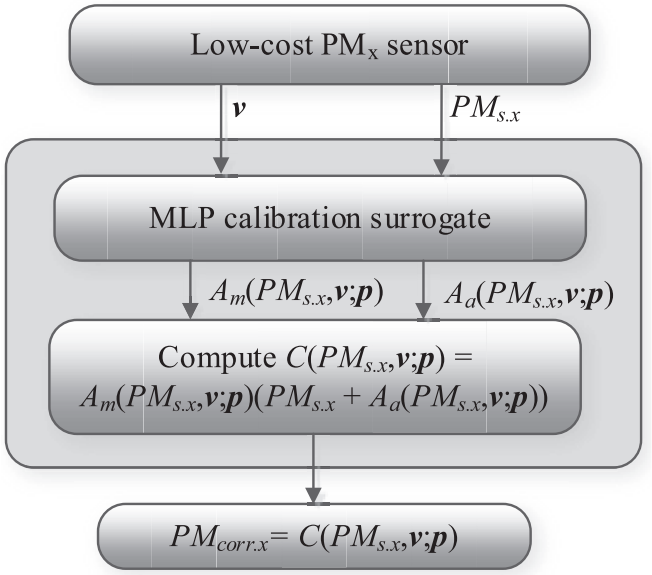


Fig. 13. Operating flow of the low-cost sensor calibration procedure proposed in this article. The PM_x and environmental parameter readings (vector $\boldsymbol{v}$) are used as the input of the calibration MLP model. The latter produces multiplicative and additive correction coefficients, which are applied to obtain the calibrated sensor output $PM_{corr.x}$.

testing sets. Section 5.2 outlines the experimental setup. Among others, it enlists the various calibration scenarios considered for the purpose of comparative analysis. We also investigate the effects of incorporating mixed multiplicative and additive correction versus additive-only scheme, as well as quantify the impact of optimizing the MLP calibration model architecture. The same section gathers the numerical results and provides their visualization. Section 5.3 offers a summary of findings and discusses the performance of the presented calibration strategy.

6. Reference and Low-Cost sensor data

The proposed calibration procedure is validated using the reference data acquired from the public measurement stations allocation in the city of Gdansk, Poland, and outlined in Section 3. The data has been gathered hourly over the period of almost two months, from the beginning of March 2023 through beginning of May 2023. The portable platforms outlined in Section 2 have been allocated in the proximity of the respective stations, and the measurements were takes at the same times. Most of this data (about 80 %) has been used for training, whereas about 20 % served for testing purposes. The testing data was allocated in one-week-long time intervals (four weeks for PM_1, and six weeks in total for $PM_{2.5}$, and PM_{10}). The details concerning data acquisition are as follows:

- Reference data was acquired in the period from March 9 to May 9, 2023, from five reference stations allocated in the city of Gdansk, Poland (cf. Section 3);
- PM_1 data was only provided by stations AM1, AM3, and AM7, whereas $PM_{2.5}$ and PM_{10} measurements were carried out by all five stations;
- Low-cost sensor data was gathered from cost-efficient platforms outlined in Section 2, allocated in close proximity of the respective reference stations;
- Data recorded every hour;
- Overall number of samples almost 4,000 for PM_1, and around 6,600 for $PM_{2.5}$ and PM_{10} (corrupted samples were removed).

The training data constituted approximately eighty percent of the dataset and was utilized to establish the calibration models. Training data was taken from results concatenated for all monitoring stations (and respective portable measurement platforms). The testing data was assigned as follows:

- Testing data corresponds to several one-week periods (four for PM_1, and six for $PM_{2.5}$ and PM_{10}), as specified below;
- The number of testing data points is $N_t = 4 \times 168 = 672$ (for PM_1), and $N_t = 6 \times 168 = 1008$ (for $PM_{2.5}$ and PM_{10});
- Specific allocation of testing samples
  - PM_1: March 26 - April 2 (AM1), March 17–24 (AM3), March16-23 (AM7), April 6–13 (AM7);
  - $PM_{2.5}$: April 17–23 (AM1), April 11–18 (AM2), March 11–18 (AM3) April 14–21 (AM3), March 19–26 (AM7), March 9–16 (AM8).
  - PM_{10}: March 26 – April 2 (AM1), April 27 – May 3 (AM1), March 29 – April 5 (AM2), March 9–16 (AM3), April 10–17 (AM3), April 9–16 (AM7).

Fig. 14 shows reference and low-cost sensor readings concatenated for all measurement locations with testing periods marked grey.

6.1. Results

This part of the paper provides the results of field-calibration of the low-cost PM_x sensor integrated into the portable monitoring platform outline in Section 2. The three types of particulate matter pollution (PM_1, $PM_{2.5}$, and PM_{10}) are treated separately.

For each of these, several calibration setups are considered, different by which of the hyper-parameters contained in the vector $\boldsymbol{H} = [N_L\ M_L\ \alpha]^T$ are optimized. For the first three scenarios, we keep $\alpha = 1$, which corresponds to sensor correction being purely additive (multiplicative coefficient $A_m = 0$). At the same time, either none of the MLP size parameters (N_L, M_L), only N_L, or both are optimized. The fourth and the fifth scenario assume variable α, and either restricted or full MLP architecture optimization. The details of hyper-parameter adjustment can be found in Section 4.3.2. The details of the considered calibration process configurations can be found in Table 1.

The setups pertinent to PM_1, $PM_{2.5}$, and PM_{10} are marked as A, B, and C, respectively. For each setup, the MLP model was trained fifty times, and the best run in terms of the loss function value was utilized as the ultimate calibration model. The multiple training runs are necessary due to the presence of stochastic components (random internal training/testing data division).

The numerical results from all scenarios have been gathered in Table 2, presenting the correlation coefficient and modeling error (RMSE) for both training and testing data. Clearly, the figures for the testing data are of primary interest. The definitions of the correlation coefficient and RMSE have been provided in Section 4.1 (equations (3) and (5), respectively). Figs. 15, 16, and 17 show the reference and calibrated low-cost sensor PM_1, $PM_{2.5}$, and PM_{10} readings for the selected subsets of the training and testing data, along with the scatter plots, respectively. In each case, we visualize the last setup for each PM type (A.5, B.5, and C.5), which incorporates optimization of all three hyper-parameters N_L, M_L, and α.

6.2. Discussion

The primary objective of the numerical experiments carried out in this section was to assess the efficiency of the proposed calibration strategy. In particular, we are interested in evaluating the reliability of the PM_x readings provided by the corrected low-cost sensor in comparison to the reference data. Furthermore, we investigate the relevance of the affine scaling (i.e., a combination of the multiplicative and additive correction), as well as the effects of adjusting the calibration model architecture (i.e., optimization of the hyper-parameters N_L and M_L). One should also recall that the calibration process is challenging as

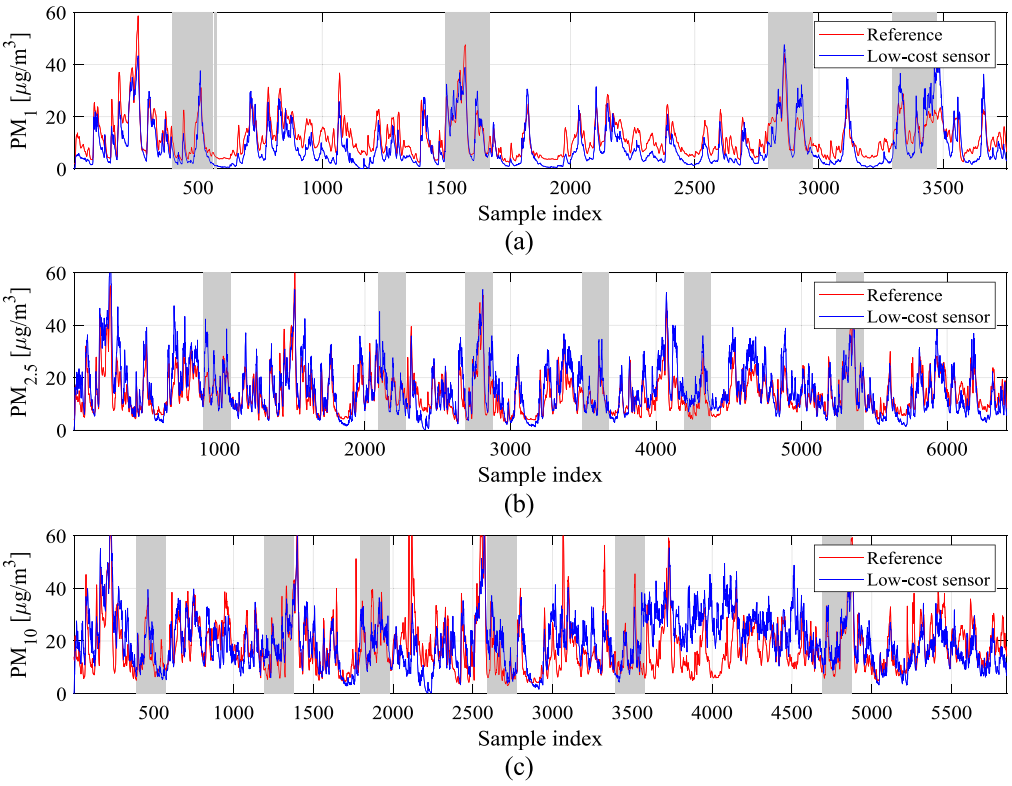


Fig. 14. Reference and low-cost sensor data utilized to calibrate the sensor of Section 2. The plots show concatenated data from all reference stations and the corresponding portable platforms. Grey areas correspond to the testing data. The remaining samples are used for calibration model training: (a) PM_1, (b) $PM_{2.5}$, (c) PM_{10}.

Table 1
Calibration model input configurations considered in verification experiments.

PM type	Calibration setup	Calibration hyper-parameters		
		Affine scaling coefficient $\alpha^{\$}$	$N_L{}^{\#}$	$M_L{}^{\&}$
PM_1	A.1	1 (fixed)	10 (fixed)	1 (fixed)
	A.2	1 (fixed)	Optimized	1 (fixed)
	A.3	1 (fixed)	Optimized	Optimized
	A.4	Optimized	Optimized	1 (fixed)
	A.5	Optimized	Optimized	Optimized
$PM_{2.5}$	B.1	1 (fixed)	10 (fixed)	1 (fixed)
	B.2	1 (fixed)	Optimized	1 (fixed)
	B.3	1 (fixed)	Optimized	Optimized
	B.4	Optimized	Optimized	1 (fixed)
	B.5	Optimized	Optimized	Optimized
PM_{10}	C.1	1 (fixed)	10 (fixed)	1 (fixed)
	C.2	1 (fixed)	Optimized	1 (fixed)
	C.3	1 (fixed)	Optimized	Optimized
	C.4	Optimized	Optimized	1 (fixed)
	C.5	Optimized	Optimized	Optimized

$^{\$}$ Fixed $\alpha = 1$ corresponds to additive sensor correction (for $\alpha = 1$, multiplicative coefficient $A_m = 1$).

$^{\#}$ Fixed $N_L = 10$ corresponds to MLP model architecture of the form [10 10 10] (all hidden layers having ten neurons each).

$^{\&}$ Fixed $M_L = 1$ corresponds to MLP model architecture having the same number of neurons N_L in each hidden layer.

the disparities between the reference and uncorrected low-cost sensor data is significant, whereas the dynamic range of PM_x readings is high, from close to zero to almost sixty μg/m^3. Finally, PM_x levels often fluctuate substantially within short timeframes (cf. Fig. 14).

The results gathered in Table 2 demonstrate that the proposed calibration approach exhibits remarkable performance both in terms of the correlation coefficient and the RMSE values. For PM_1 and $PM_{2.5}$, the correlation coefficient reached the values of about 0.86, whereas for PM_{10}, it is about 0.76. These numbers are impressive, especially having in mind considerable misalignment between the uncorrected low-cost sensor and reference measurements, for which the correlation coefficients are 0.40, 0.44, and 0.17, respectively.

It should be emphasized that the reliability of the low-cost sensor is particularly poor for PM_{10}, yet our calibration technique allows elevating it to usable levels. These improvements can be observed in Figs. 15, 16, and 17. In particular, the scatter plots for the calibrated sensor are concentrated much closer to the identify function than those for corresponding to the raw sensor. In terms of RMSE, the values for the calibrated sensor are as low as about 3.1, 4.1, and 4.9 μg/m^3 for PM_1, $PM_{2.5}$, and PM_{10}, respectively. The values for uncorrected sensor are 5.3, 5.6, and 9.6 μg/m^3. Again, noticeable improvement can be observed due to the proposed calibration. Finally, the average relative error of the calibrated low-cost sensor is about 20, 18, and 26 percent for PM_1, $PM_{2.5}$, and PM_{10}, respectively.

As can be observed, the matching between the reference and corrected low-cost sensor data is better for the training set. This is expected as the calibration model is exclusively trained using the training data, and the training process attempts to improve the alignment between the reference and low-cost sensor readings over this set. The testing set is not used in this process whatsoever. It is therefore expected that the discrepancies for the testing set will be higher. Although the calibration process accounts for the overall behavior and properties of the low-cost sensor, it is obviously not capable of accounting for local fluctuations present in the sensor data for testing locations. Here, a critical part is to ensure that the calibration model is relatively simple so that its approximation capability is limited, otherwise (e.g., in the case of interpolating models), approximation capability would by far surpass generalization capability, thereby leading to inferior results over the

Table 2
Sensor calibration performance: correlation coefficients and RMSE.

PM type	Calibration setup	Training data			Testing data			Calibration model architecture	
		Correlation coefficient r	RMSE [μg/m^3]	Relative error [%]	Correlation coefficient r	RMSE [μg/m^3]	Relative error [%]	α	MLP architecture
PM_1	A.1	0.86	2.57	19.3	0.817	3.64	23.9	1	[10 10 10]
	A.2	0.91	2.06	14.9	0.820	3.76	21.5	1	[12 12 12]
	A.3	0.90	2.18	16.1	0.828	3.67	21.2	1	[16 8 4]
	A.4	0.90	2.13	15.1	0.839	3.56	20.9	0.6	[10 10 10]
	A.5	0.91	2.11	15.1	0.859	3.13	20.1	0.8	[40 20 10]
$PM_{2.5}$	B.1	0.88	2.61	16.6	0.814	4.69	23.9	1	[10 10 10]
	B.2	0.83	3.04	18.5	0.843	4.31	22.5	1	[8 8 8]
	B.3	0.84	2.95	18.2	0.850	4.21	22.1	1	[12 6 3]
	B.4	0.82	3.19	18.8	0.856	4.13	19.6	0.8	[8 8 8]
	B.5	0.92	2.10	14.9	0.860	4.08	18.2	0.7	[6 12 24]
PM_{10}	C.1	0.70	5.73	22.9	0.721	5.37	31.5	1	[10 10 10]
	C.2	0.79	4.79	23.0	0.743	4.98	28.4	1	[8 8 8]
	C.3	0.87	3.80	17.8	0.755	5.02	28.1	1	[30 15 8]
	C.4	0.88	3.64	17.6	0.756	5.01	26.5	0.6	[15 15 15]
	C.5	0.87	3.83	20.8	0.762	4.92	26.2	0.8	[30 15 8]

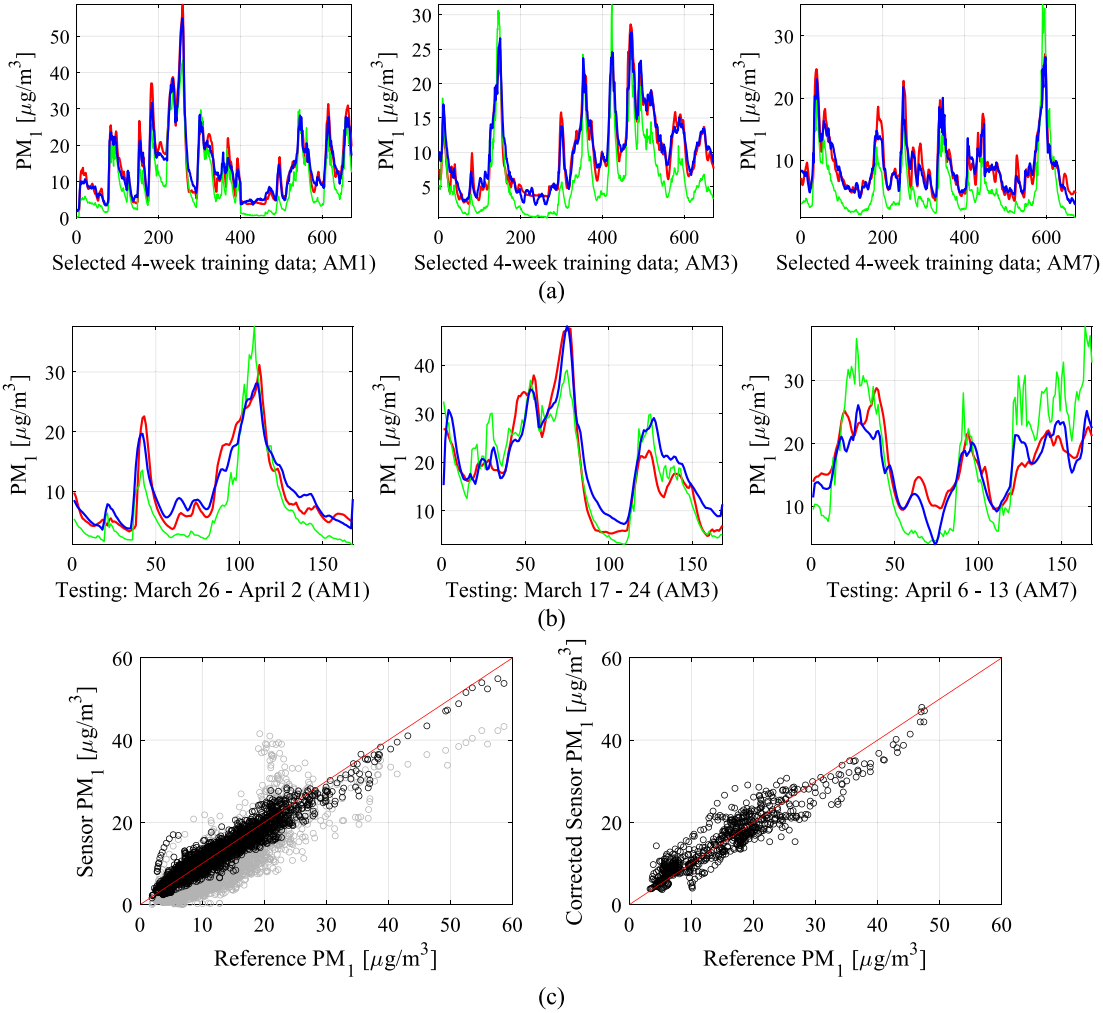


Fig. 15. Sensor calibration performance for PM_1 using setup A.5 (cf. Table 1): (a) selected subsets of the training data (reference – red, uncorrected sensor – green, corrected sensor – blue); (b) selected subsets of the testing data (reference – red, uncorrected sensor – green, corrected sensor – blue); (c) scatter plots for the training data (left: uncorrected – gray, corrected – black) and the testing data (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

testing set. In this work it is prevented by using a simple ANN architecture as elucidated in Section 4.

Another point of interest are the effects of incorporating mixed multiplicative and additive correction (cf. Section 4.2), as well as optimization of the MLP calibration model hyper-parameters. As indicated in Table 2, affine sensor correction (variable scaling coefficient α) brings

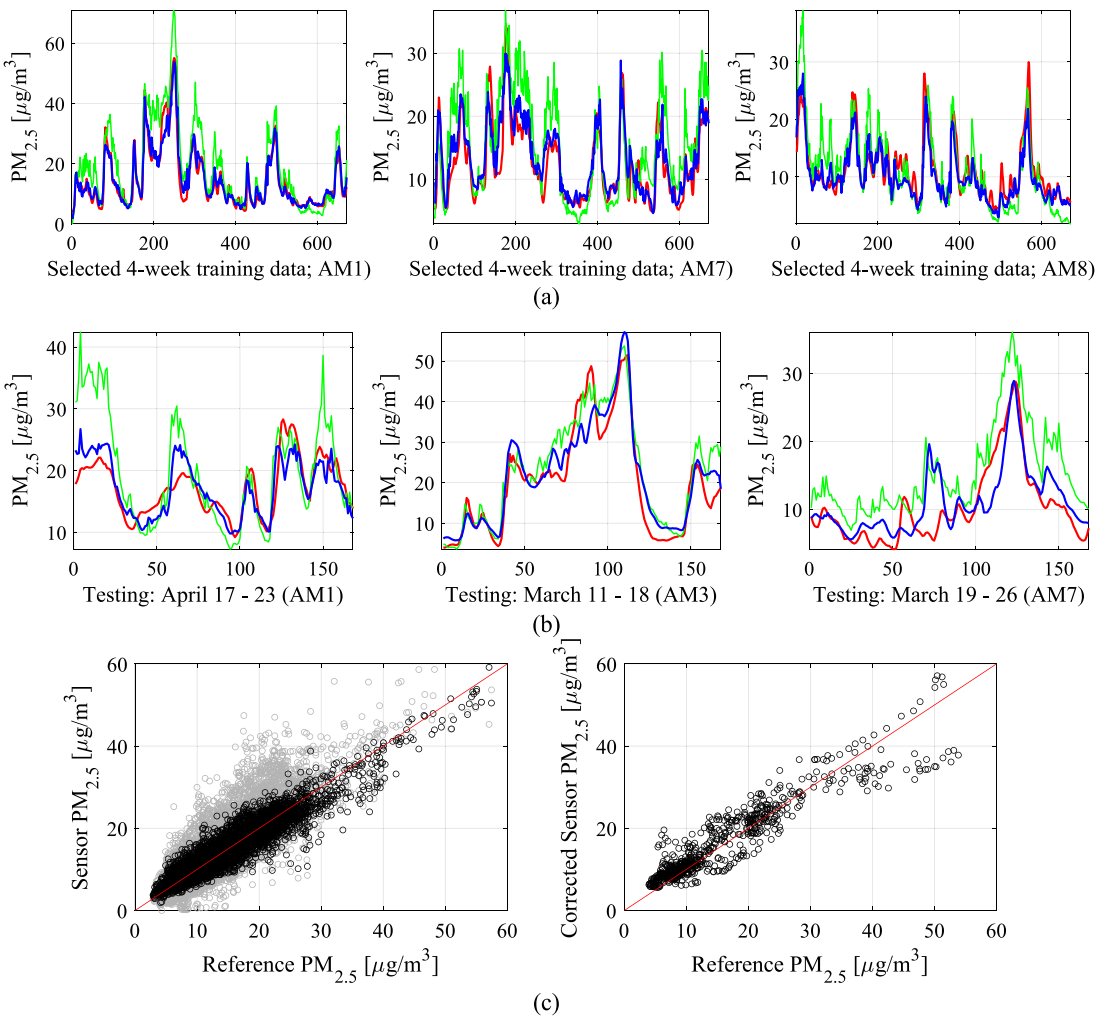


Fig. 16. Sensor calibration performance for $PM_{2.5}$ using setup B.5 (cf. Table 1): (a) selected subsets of the training data (reference – red, uncorrected sensor – green, corrected sensor – blue); (b) selected subsets of the testing data (reference – red, uncorrected sensor – green, corrected sensor – blue); (c) scatter plots for the training data (left: uncorrected – gray, corrected – black) and the testing data (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

certain advantages as compared to traditional additive correction (α fixed at the value of unity). On the average, it leads to about 0.04 improvement of the correlation coefficient for PM_1 (setup A.5 versus A.3), and up to 0.02 improvement for $PM_{2.5}$ and PM_{10}. At the same time, similar benefits are observed due to optimization of the MLP model hyper-parameters N_L and M_L, which contribute to about 0.03 to 0.04 improvement of the correlation coefficient for all PM types. Both factors also contribute to a reduction of the RMSE values with their collective effect being as high as about 0.5 μg/m^3.

For the sake of supplementary validation, the calibration process has been repeated with the white Gaussian noise added to the low-cost sensor readings. The calibration process was repeated for PM_1, $PM_{2.5}$, and PM_{10} using the calibration setups A.5, B.5, and C.5. In each case, four noise levels were applied (different by the maximum noise amplitude): 1 μm/m^3, 2 μm/m^3, 5 μm/m^3, and 10 μm/m^3. The results have been gathered in Table 3. It should be emphasized that the noise levels of 5 μm/m^3 and 10 μm/m^3 are huge and by far exceeding the typical RMSE values obtained for the respective PM_x types and calibration setups for noise-free data. As a matter of fact, even the noise level of 2 μm/m^3 is significant, i.e., comparable with the average error values. The analysis of data included in Table 3 indicates that incorporating noise obviously has detrimental effects on both approximation and generalization capability of the calibration model. However, the effects are more pronounced for the training data, which is due to the fact that the presence of noise disturbs (systematic) relationships between the reference and the low-cost sensor readings. At the same time, worsening of the correlation coefficient and RMSE for the testing data, while still noticeable, is less pronounced than for the training samples. This effect is a result of poorer alignment between the reference and the calibrated sensor across the testing set (as compared to the training data), which means that the presence of noise is not as harmful here. On the other hand, this corroborates the robustness of the presented calibration strategy and its relative immunity to noise.

Overall, the calibration technique proposed in this study has been demonstrated to exhibit remarkable efficiency. The reliability of the corrected low-cost sensor is excellent, especially when compared to its raw version. The proposed mixed multiplicative and additive correction allows us to elevate the correlation coefficient with the reference data to practically acceptable level even for the most challenging case of PM_{10}, where the initial correlation coefficient was as low as 0.17 (increased to 0.76 upon calibration). On the practical side, the calibration process can be executed offline, i.e., applied to the PM_x measurements acquired from the portable platform before presenting the results to the user. Another approach would be to implement the calibration scheme within the platform using its built-in computational resources, as outlined in Section 2.

For the purpose of supplementary validation, the calibration approach proposed in this study has been compared to several bench-

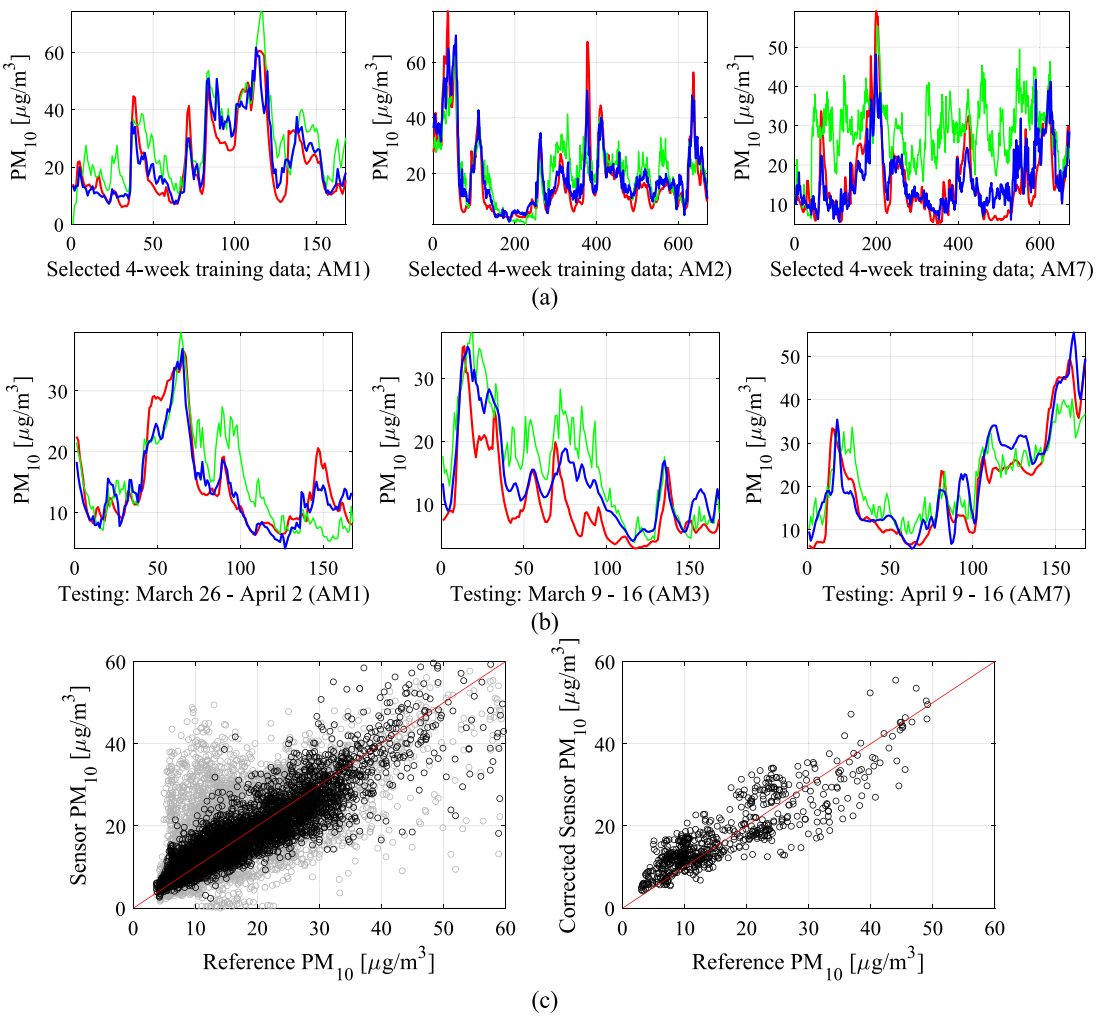


Fig. 17. Sensor calibration performance for PM_{10} using setup C.5 (cf. Table 1): (a) selected subsets of the training data (reference – red, uncorrected sensor – green, corrected sensor – blue); (b) selected subsets of the testing data (reference – red, uncorrected sensor – green, corrected sensor – blue); (c) scatter plots for the training data (left: uncorrected – gray, corrected – black) and the testing data (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Sensor calibration including white Gaussian noise added to low-cost sensor readings.

PM type	Calibration setup	Noise amplitude [μg/m^3]	Training data		Testing data	
			Correlation coefficient *r*	RMSE [μg/m^3]	Correlation coefficient *r*	RMSE [μg/m^3]
PM_1	A.5	0 (noise-free)	0.91	2.11	0.86	3.13
		1	0.90	2.15	0.83	3.69
		2	0.88	2.20	0.81	3.86
		5	0.85	2.45	0.80	3.95
		10	0.68	3.87	0.78	4.03
$PM_{2.5}$	B.5	0 (noise-free)	0.92	2.10	0.86	4.08
		1	0.91	2.19	0.84	4.61
		2	0.87	2.37	0.82	4.79
		5	0.82	2.71	0.80	5.04
		10	0.73	3.86	0.78	5.20
PM_{10}	C.5	0 (noise-free)	0.87	3.83	0.76	4.92
		1	0.84	4.20	0.73	5.31
		2	0.83	4.32	0.71	5.49
		5	0.81	4.59	0.69	5.81
		10	0.75	5.17	0.66	5.87

mark methods, including linear regression, support vector regression (SVR), Random Forest Regression, ANN-based calibration, and calibration implemented using a convolutional neural network (CNN) [71]. In all cases, the respective calibration models were set up to directly predict the model output (as opposed to predicting correction coefficients as in the method considered in Section 3). The linear regression is a model of the form

$$S(\mathbf{v}, PM_{s.x}) = \alpha_0 + \alpha_1 T_o + \alpha_2 T_i + \alpha_3 H_o + \alpha_4 H_i + \alpha_5 P + \alpha_6 PM_{s.x} \quad (14)$$

i.e., it uses all environmental parameters measured by the portable monitoring platform as well as the $PM_{s.x}$ readings from the low-cost sensor. The model coefficients are established using conventional least-square regression based on the training data. The SVR model employs Gaussian kernels, and its hyper-parameters are optimized using expected improvement as the acquisition function [80]. Random forest regression utilizes least-square boosting as the ensemble aggregation methods, as well as the hyper-parameter optimization. The ANN uses the same architecture as described in Section 4 (three fully-connected hidden layers). We consider three variations different in terms of the number of neurons (10, 15, and 20 for versions ANN_1, ANN_2, and ANN_3, respectively). The ANN model is trained using the back-propagation Levenberg-Marquardt algorithm with random data division. The CNN architecture uses filters of the size 4 × 1 × 1, three convolution layers of spatial sizes 32, 16, and 8, followed by a fully connected layer of the size 64 neurons (CNN_1), layers of sizes 64, 32, 16 (CNN_2), and 128, 64, and 32 (CNN_3). Furthermore, batch normalization and ReLU layers are included between the convolution layers. CNN is trained using the ADAM's algorithm with a mini batch size of 1000 [71].

Table 4 gathers the numerical results for all considered benchmark techniques and all three PM types. As it can be noticed, the proposed calibration approach performs significantly better than the benchmark, especially the best configurations (A.5, B.5, and C.5), both in terms of the correlation coefficient and RMSE. Only CNN in some of its variants has been shown to be close to our method, yet slightly inferior. Also, Random Forest calibration works well on the testing data, but its generalization capability is not as good as that of the proposed framework. This indicates that utilization of affine correction with optimized balance factor α is superior to direct prediction of the calibrated sensor.

7. Conclusion

This article introduces a streamlined field calibration scheme designed for low-cost particulate matter (PM) sensors. The proposed methodology relies on a combination of mixed multiplicative and additive correction applied to the readings of the low-cost sensor. The balance between these correction mechanisms is governed by an optimizable hyper-parameter within the procedure. Correction coefficients are predicted through a feedforward neural network, specifically a multilayer perceptron (MLP), utilizing environmental parameters (temperature, humidity, atmospheric pressure) measured by the sensor. The architecture of the MLP is also optimized to enhance the model's generalization capability. To validate and showcase the effectiveness of our calibration strategy, it was implemented on a portable monitoring platform developed at Gdansk University of Technology, Poland. This platform integrates PM_x and environmental sensors, along with electronic circuitry for implementing measurement protocols and wireless data transmission. Numerical experiments were conducted using reference data obtained from public monitoring stations situated in the city of Gdansk.

The results obtained for three types of particulate matter, namely PM_1, $PM_{2.5}$, and PM_{10}, underscore the exceptional performance of our calibration methodology. On one hand, the achieved correlation coefficients with the reference data are notably high, reaching 0.86 for

Table 4
Results of comparative studies.

PM type	Calibration model	Training data			Testing data		
		Correlation coefficient r	RMSE [μg/m^3]	Relative error [%]	Correlation coefficient r	RMSE [μg/m^3]	Relative error [%]
PM_1	Linear regression	0.78	3.3	24.5	0.74	4.5	28.2
	SVR	0.83	1.8	11.9	0.70	4.9	24.2
	Random Forest	0.98	1.0	5.9	0.78	4.1	23.4
	ANN_1[#]	0.93	1.8	12.6	0.50	6.3	33.9
	ANN_2[#]	0.83	2.8	20.6	0.62	5.5	28.5
	ANN_3[#]	0.84	2.8	17.8	0.58	5.8	25.3
	CNN_1[$]	0.88	2.4	19.0	0.78	4.2	23.9
	CNN_2[$]	0.89	2.3	18.1	0.78	4.1	24.1
	CNN_3[$]	0.90	2.2	17.2	0.82	3.7	22.1
$PM_{2.5}$	Linear regression	0.76	3.6	21.6	0.78	5.1	22.6
	SVR	0.88	2.6	13.1	0.82	4.7	28.5
	Random Forest	0.97	1.2	6.5	0.79	5.0	29.3
	ANN_1[#]	0.89	2.5	15.2	0.74	5.6	29.2
	ANN_2[#]	0.89	2.5	15.1	0.65	6.5	33.0
	ANN_3[#]	0.95	1.7	10.3	0.47	7.9	34.0
	CNN_1[$]	0.81	3.2	18.8	0.85	4.1	25.2
	CNN_2[$]	0.83	3.0	17.9	0.86	4.1	24.7
	CNN_3[$]	0.85	2.8	16.8	0.84	4.2	24.9
PM_{10}	Linear regression	0.54	7.1	34.6	0.64	6.1	39.1
	SVR	0.78	4.9	18.1	0.67	5.9	37.7
	Random Forest	0.93	2.6	12.2	0.64	6.1	40.13
	ANN_1[#]	0.74	5.3	25.1	0.62	6.3	40.2
	ANN_2[#]	0.85	4.0	18.6	0.60	6.5	40.7
	ANN_3[#]	0.90	3.2	13.9	0.61	6.4	39.5
	CNN_1[$]	0.70	5.7	27.7	0.68	5.8	39.4
	CNN_2[$]	0.73	5.4	25.2	0.68	5.7	39.6
	CNN_3[$]	0.77	5.0	23.7	0.67	5.9	39.9

[#] Three ANN architectures utilized, all with three fully connected hidden layers, and the following number of neurons per layer: 10 (ANN_1), 15 (ANN_2), 20 (ANN_3).

[$] Three CNN architectures utilized, all with filters of the size 4 × 1 × 1, three convolution layers of spatial sizes 32, 16, and 8, followed by a fully-connected layer of the size 64 neurons (CNN_1), layers of sizes 64, 32, 16 (CNN_2), and 128, 64, and 32 (CNN_3), as well as batch normalization and ReLU layers in between the convolution layers. CNN is trained using the ADAM's algorithm with a mini batch size of 1000 [71].

PM_1 and $PM_{2.5}$, and 0.76 for PM_{10}. This marks an improvement over the uncorrected sensor, where the corresponding correlation coefficients were 0.40, 0.44, and 0.17. On the other hand, the attained error levels (RMSE) are merely about 3.1, 4.1, and 4.9 $\mu g/m^3$, emphasizing the practical utility and reliability of the calibrated low-cost sensor. Extensive comparative experiments further validate the importance of incorporating affine scaling, showing an improvement of up to 0.04 in the correlation coefficient compared to an additive-correction-only scheme. Simultaneously, the optimization of the multilayer perceptron (MLP) calibration model architecture has demonstrated a similar contribution to enhancing reliability, with an improvement of up to 0.03 in terms of the correlation coefficient. Although the proposed calibration scheme is generic, it seems that the fundamental limitation is related to the size of the training dataset. If the available number of training samples becomes extremely large, identification of the calibration model may become problematic (and, definitely more time consuming). Also, the MLP architecture might need to become more complex, which translates into more intricate optimization process of its hyperparameters.

Future efforts will concentrate on enhancing the efficiency of the calibration process. One avenue for exploration involves incorporating additional inputs into the calibration model, such as estimated time derivatives of environmental parameters and time series data of prior PM_x readings from the low-cost sensor. Another option is the development of supplementary correction mechanisms designed to globally improve the correlation between reference and sensor data. Furthermore, the consideration of more advanced machine learning tools, including convolutional neural networks, will be explored. Another objective of the future work will be investigation of the effects of changing external conditions due to different seasons (winter, summer, etc.) on the calibration process. This will require acquisition of more extensive reference and low-cost sensor data, which is one of the goals of the future measurement campaign. Yet another research direction is investigating potential cross-sensitivity of the low-cost sensor to different PM types, which may be explored as a potential way of improving calibration process efficacy by incorporating the readings of other PM types as the calibration inputs. Finally, the incorporation of statistical models will be considered as well, such as support vector regression, Gaussian process regression, etc.

CRediT authorship contribution statement

Slawomir Koziel: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Anna Pietrenko-Dabrowska:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis. **Marek Wojcikowski:** Writing – review & editing, Supervision, Software, Resources, Project administration, Funding acquisition. **Bogdan Pankiewicz:** Writing – review & editing, Validation, Software, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The research leading to these results has received funding from the Norway Grants 2014-2021 via the National Centre for Research and Development, Poland, grant NOR/POLNOR/HAPADS/0049/2019-00. This work was also supported in part by the Icelandic Research Fund Grant 217771. The authors would also like to thank the Agency of Regional Atmospheric Monitoring Gdansk-Gdynia-Sopot (ARMAG) for providing free of charge data from the reference measurement stations.

References

[1] J. Lelieveld, K. Klingmüller, A. Klaus, P. Pozzer, M. Ulrich, A.D. Fnais, T. Münzel, "Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions, Eur. Heart J. 20 (2019) 1590–1596.
[2] S. Khomenko, M. Cirach, E. Pereira-Barboza, N. Mueller, J. Barrera-Gomez, D. Rojas-Rueda, Premature mortality due to air pollution in european cities: a health impact assessment, Lancet 5 (3) (2021) 121–134.
[3] "Air quality in Europe 2022," Report no. 05/2022, European Environment Agency, doi: 10.2800/488115, 2022.
[4] C.D. Koolen, G. Rothenberg, Air pollution in Europe, Chem.-Susta.-Energy-Mater. 12 (1) (2018) 164–172.
[5] M. Santibanez-Andrade, Y.I. Chirino, I. Gonzalez-Ramirez, Y. Sanchez-Perez, C. M. Garcia-Cuellar, Deciphering the code between air pollution and disease: the effect of particulate matter on cancer hallmarks, Mol. Sci. 21 (136) (2020).
[6] "GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019," *Lancet*, vol. 396, pp. 1204-1222, 2019.
[7] C. Krittanawong, Y. Kamran Qadeer, R.B. Hayes, Z. Wang, S. Virani, G.D. Thurston, C.J. Lavie, "$PM_{2.5}$ and cardiovascular health risks, current problems in cardiology, Int. J. Cardiol. Cardiovasc. Risk Prev. 48 (6) (2023) 101670.
[8] T. Zhang, W. Mao, J. Gao, et al., The effects of $PM_{2.5}$ on lung cancer-related mortality in different regions and races: a systematic review and meta-analysis of cohort studies, Air Qual. Atmos. Health. 15 (2022) 1523–1532.
[9] Z. Zhang, D. Zhu, B. Cui, R. Ding, X. Shi, P. He, Association between particular matter air pollution and lung cancer, Thorax 75 (1) (2020) pp.
[10] B.L. Alman, J.A. Stingone, M. Yazdy, L.D. Botto, T.A. Desrosiers, S. Pruitt, et al., Associations between $PM_{2.5}$ and risk of preterm birth among liveborn infants, Annals of Epidemiol. 39 (2019) 46–53.
[11] K. Juda-Rezler, M. Reizer, K. Maciejewska, B. Błaszczak, K. Klejnowski, Characterization of atmospheric $PM_{2.5}$ sources at a central European urban background site, Sci. Total Environ. 713 (2020) 136729.
[12] A. Mehadi, H. Moosmüller, D.E. Campbell, W. Ham, D. Schweizer, L. Tarnay, et al., Laboratory and field evaluation of real-time and near real-time $PM_{2.5}$ smoke monitors, J. Air Waste Manag. Assoc. 70 (2) (2020) 158–179.
[13] E. Bagkis, T. Kassandros, K. Karatzas, Learning calibration functions on the fly: hybrid batch online stacking ensembles for the calibration of low-cost air quality sensor networks in the presence of concept drift, Atmos. 13 (2022) 416.
[14] L. Morawska, P.K. Thai, X. Liu, A. Asumadu-Sakyi, G. Ayoko, A. Bartonova, et al., Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone? Environ. Int. 116 (2018) 286–299.
[15] T. Zheng, M.H. Bergin, K.K. Johnson, S.N. Tripathi, S. Shirodkar, M.S. Landis, et al., Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments, Atmos. Meas. Tech. 11 (8) (2018) 4823–4846.
[16] A. Datta, A. Saha, M. Levy Zamora, C. Buehler, L. Hao, F. Xiong, Statistical field calibration of a low-cost $PM_{2.5}$ monitoring network in baltimore, Atmos. Environ. 242 (2020) 117761.
[17] E. Bainomugisha, J. Ssematimba, D. Okure, Design considerations for a distributed low-cost air quality sensing system for urban environments in low-resource settings, Atmos. 14 (2) (2023) 354.
[18] J.S. Apte, K.P. Messier, S. Gani, M. Brauer, T.W. Kirchstetter, M.M. Lunden, et al., High-resolution air pollution mapping with google street view cars: exploiting big data, Environ. Sci. Technol. 51 (12) (2017) 6999–7008.
[19] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, et al., "Deriving high-resolution urban air pollution maps using mobile sensor nodes, Pervasive Mob. Comput. 16 B (2015) 268–285.
[20] F. Kane, J. Abbate, E.C. Landahl, M.J. Potosnak, Monitoring particulate matter with wearable sensors and the influence on student environmental attitudes, Sensors 22 (2022) 1295.
[21] S. Palomeque-Mangut, F. Meléndez, J. Gómez-Suárez, S. Frutos-Puerto, P. Arroyo, et al., Wearable system for outdoor air quality monitoring in a WSN with cloud computing: design, validation and deployment, Chemosphere 307 (3) (2022) 135948.
[22] L. Zamora, M. Pulczinski, J.C. Johnson, N. Garcia-Hernandez, R. Rule, A. Carrillo, et al., Maternal exposure to $PM_{2.5}$ in South Texas, a pilot study, Sci. Total Environ. 628 (2018) 1497–1507.
[23] N. Castell, F.R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, et al., Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environ. Int. 99 (2017) 293–302.
[24] C. Malings, R. Tanzer, A. Hauryliuk, P.K. Saha, A.A. Robinson, R. Subramanian, Fine particle mass monitoring with low-cost sensors: corrections and long-term performance evaluation, Aerosol Sci. Technol. 54 (2020) 160–174.
[25] M.R. Giordano, C. Malings, S.N. Pandis, A.A. Presto, V.F. McNeill, 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. 158 (2021) 105833.

[26] K.K. Barkjohn, B. Gantt, A.L. Clements, Development and application of a United States-wide correction for PM2.5 data collected with the PurpleAir sensor, Atmos. Meas. Tech. 14 (2021) 4617–4637.
[27] H. Khreis, J. Johnson, K. Jack, B. Dadashova, E.S. Park, Evaluating the performance of low-cost air quality monitors in dallas, Texas. Int. J. Environ. Res. Public Health 19 (2022) 1647.
[28] P. deSouza, R. Kahn, T. Stockman, W. Obermann, B. Crawford, A. Wang, et al., Calibrating networks of low-cost air quality sensors, Atmos. Meas. Tech. 15 (2022) 6309–6328.
[29] K.E. Kelly, J. Whitaker, A. Petty, C. Widmer, A. Dybwad, D. Sleeth, R. Martin, A. Butterfield, Ambient and laboratory evaluation of a low-cost particulate matter sensor, Environ. Pollut. 221 (2017) 491–500.
[30] M. Badura, P. Batog, A. Drzeniecka-Osiadacz, P. Modzel, Evaluation of low-cost sensors for ambient $PM_{2.5}$ monitoring, J. Sens. (2018) 5096540.
[31] R. Jayaratne, X. Liu, P. Thai, M. Dunbabin, L. Morawska, The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog, Atmos. Meas. Tech. 11 (2018) 4883–4890.
[32] D. Kim, D. Shin, J. Hwang, Calibration of low-cost sensors for measurement of indoor particulate matter concentrations via laboratory/field evaluation, Aerosol Air Qual. Res. 23 (2023) 230097.
[33] K.K. Barkjohn, M.H. Bergin, C. Norris, J.J. Schauer, Y. Zhang, M. Black, M. Hu, J. Zhang, Using low-cost sensors to quantify the effects of air filtration on indoor and personal exposure relevant $PM_{2.5}$ concentrations in Beijing, China, Aerosol and Air Qual. Res. 20 (2) (2020).
[34] G. Tancev, C. Pascale, The relocation problem of field calibrated low-cost sensor systems in air quality monitoring: a sampling bias, Sensors 20 (2020) 6198.
[35] H.Y. Liu, P. Schneider, R. Haugen, M., Performance assessment of a low-cost $PM_{2.5}$ sensor for a near four-month period in Oslo, Norway, Atmos. 10 (2) (2019) 41.
[36] H. Lee, J. Kang, S. Kim, Y. Im, S. Yoo, D. Lee, Long-term evaluation and calibration of low-cost particulate matter (PM) sensor, Sensors 20 (2020) 3617.
[37] B.I. Magi, C. Cupini, J. Francis, M. Green, C. Hauser, Evaluation of $PM_{2.5}$ measured in an urban setting using a low-cost optical particle counter and a federal equivalent method beta attenuation monitor, Aerosol Sci. Technol. 54 (2019) 147–159.
[38] N. Zimmerman, A.A. Presto, S.P.N. Kumar, J. Gu, A. Hauryliuk, E. Robinson, et al., A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, Atmos. Meas. Tech. 11 (2018) 291–313.
[39] M. Masiol, S. Squizzato, D. Chalupa, D.Q. Rich, P.K. Hopke, Evaluation and field calibration of a low-cost ozone monitor at a regulatory urban monitoring station, Aerosol. Air Qual. Res. 18 (2018) 2029–2037.
[40] G.H. Hong, T.C. Le, J.W. Tu, C. Wang, S.C. Chang, J.Y. Yu, G.Y. Lin, et al., Long-term evaluation and calibration of three types of low-cost $PM_{2.5}$ sensors at different air quality monitoring stations, J. Aerosol Science 157 (2021) 105829.
[41] A. Cavaliere, F. Carotenuto, F. Di Gennaro, B. Gioli, G. Gualtieri, F. Martelli, A. Matese, P. Toscano, C. Vagnoli, A. Zaldei, Development of low-cost air quality stations for next generation monitoring networks: calibration and validation of $PM_{2.5}$ and PM_{10} sensors, Sensors 18 (2018) 2843.
[42] J. Hofman, et al., Mapping air quality in IoT cities: cloud calibration and air quality inference of sensor data, Rotterdam, Netherlands, Conf. IEEE Sensors, 2020, pp. 1–4.
[43] M.V. Narayana, D. Jalihal, S. Nagendra, "Establishing a sustainable low-cost air quality monitoring setup: a survey of the state-of-the-art, Sensors 22 (2022) 394.
[44] M. Zusman, C.S. Schumacher, A.J. Gassett, E.W. Spalt, E. Austin, T. Larson, et al., Calibration of low-cost particulate matter sensors: model development for a multi-city epidemiological study, Environ. Int. 134 (2020) 105329.
[45] E.M. Considine, C.E. Reid, M.R. Ogletree, T. Dye, Improving accuracy of air pollution exposure measurements: statistical correction of a municipal low-cost airborne particulate matter sensor network, Environ. Pollut. 268 (2021) 115833.
[46] L. Liang, Calibrating low-cost sensors for ambient air monitoring: techniques, trends, and challenges, Environ. Res. 197 (2021) 111163.
[47] J. Venkatraman Jagatha, A. Klausnitzer, M. Chacón-Mateos, B. Laquai, E. Nieuwkoop, et al., "Calibration method for particulate matter low-cost sensors used in ambient air quality monitoring and research, Sensors 21 (2021) 3960.
[48] Y. Wang, Y. Du, J. Yanjun, T. Li, "Calibration of a low-cost $PM_{2.5}$ monitor using a random forest model, Environ. Int. 133 (2019) 105161.
[49] M. Si, K. Du, Development of a predictive emissions model using a gradient boosting machine learning method, Environ. Technol. Innov. 20 (2020) 101028.
[50] B.G. Loh, G.H. Choi, Calibration of portable particulate matter–monitoring device using web query and machine learning, Safety Health at Work 10 (4) (2019) 452–460.
[51] S. De Vito, E. Esposito, M. Salvato, O. Popoola, F. Formisano, R. Jones, G. Di Francia, Calibrating chemical multisensory devices for real world applications: an indepth comparison of quantitative machine learning approaches, Sensor. Actuator. B Chem. 255 (2018) 1191–1210.
[52] S. Mahajan, P. Kumar, Evaluation of low-cost sensors for quantitative personal exposure monitoring, Sustain. Cities Soc. 57 (2020).
[53] C. -C. Chen et al., "Calibration of low-cost particle sensors by using machine-learning method," *IEEE Asia Pacific Conf. Circuits and Systems (APCCAS)*, Chengdu, China, 2018, pp. 111-114.
[54] V. Athira, P. Geetha, R. Vinayakumar, K.P. Soman, DeepAirNet: applying recurrent networks for air quality prediction, Procedia Comput. Sci. 132 (2018) 1394–1403.
[55] X. Dai, J. Liu, Y. Li, A recurrent neural network using historical data to predict time series indoor $PM_{2.5}$ concentrations for residential buildings, Indoor Air 31 (2021) 1228–1237.
[56] H. Jeon, J. Ryu, K.M. Kim, J. An, The development of a low-cost particulate matter 2.5 sensor calibration model in daycare centers using long short-term memory algorithms, Atmos. 14 (2023) 1228.
[57] S. Ali, F. Alam, K.M. Arif, J. Potgieter, Low-cost CO sensor calibration using one dimensional convolutional neural network, Sensors 23 (2023) 854.
[58] H. Yu, et al., A deep calibration method for low-cost air monitoring sensors with multilevel sequence modeling, IEEE Trans. Instrument. Meas. 69 (9) (2020) 7167–7179.
[59] R.R. Kureshi, B.K. Mishra, D. Thakker, R. John, A. Walker, S. Simpson, et al., Data-driven techniques for low-cost sensor selection and calibration for the use case of air quality monitoring, Sensors 22 (2022) 1093.
[60] Datasheet SPS30, Particulate matter sensor for air quality monitoring and control, Sensirion (2024). https://sensirion.com/media/documents/8600FF88/616542B5/Sensirion_PM_Sensors_Datasheet_SPS30.pdf.
[61] Humidity Sensor BME280, Bosch Sensortec.. Available: https://www.bosch-sensortec.com/products/environmental-sensors/humidity-sensors-bme280/.
[62] 'BeagleBone® Blue', BeagleBoard. Accessed: Feb. 07, 2024. [Online]. Available: https://www.beagleboard.org/boards/beaglebone-blue.
[63] TI AM335x System-In-Package - Octavo - OSD335x ARM A8, 1GB DDR3', Octavo Systems, 2024. https://octavosystems.com/octavo_products/osd335x/.
[64] Arm-based processors | TI.com, 2024: https://www.ti.com/microcontrollers-mcus-processors/arm-based-processors/overview.html.
[65] Arduino and RaspberryPI modems - u-GSM shield LTE CAT M1, NB IoT, LTE CAT NB1, LTE CAT NB2, LTE CAT4, LTE CAT1, UMTS, and GSM: presentation, 2024. https://itbrainpower.net/u-GSM/features.php LPWA BG96 Cat M1/NB1/EGPRS, Quectel, 2024: https://www.quectel.com/product/lpwa-bg96-cat-m1-nb1-egprs.
[66] LPWA BG96 Cat M1/NB1/EGPRS, Quectel, 2024: https://www.quectel.com/product/lpwa-bg96-cat-m1-nb1-egprs.
[67] The World's First IoT Lifetime Flat, 1NCE, 2024: https://1nce.com/en-eu/.
[68] ARMAG Foundation, 2024: https://armaag.gda.pl/en/index.htm.
[69] R. Vang-Mata (Ed.), *Multilayer perceptrons*, Nova Science Pub. Inc., 2020.
[70] S. Dlugosz, Multi-layer perceptron networks for ordinal data analysis, Logos Verlag (2008).
[71] C.C. Aggarwal, Neural networks and deep Learning, A Textbook, Springer, New York, 2018.
[72] M. Casari, L. Po, L. Zini, AirMLP: a multilayer perceptron neural network for temporal correction of $PM_{2.5}$ values in Turin, Sensors 23 (2023) 9446.
[73] Y. Kang, L. Aye, T.C. Ngo, J. Zhou, Performance evaluation of low-cost air quality sensors: a review, Sci. Total Environ. 818 (2021).
[74] S. Vikram, A. Collier-Oxandale, M.H. Ostertag, M. Menarini, C. Chermak, S. Dasgupta, T. Rosing, M. Hannigan, W.G. Griswold, Evaluating and improving the reliability of gas-phase sensor system calibrations across new locations for ambient measurements and personal exposure monitoring, Atm. Meas. Techn. 12 (8) (2019) 4211–4239.
[75] M.V. Narayana, D. Jalihal, S.M. Shiva Nagendra, Establishing a sustainable low-cost air quality monitoring setup: a survey of the state-of-the-art, Sensors 22 (1) (2022) 394.
[76] J. Khan, M. Ketzel, K. Kakosimos, M. Sorensen, S.S. Jensen, Road traffic air and noise pollution exposure assessment – a review of tools and techniques, Sc. Tot. Environment 634 (2018) 661–676.
[77] F. Maccantelli, G. Peruzzi, A. Pozzebon, "Traffic level monitoring in urban scenarios with virtual sensing techniques enabled by embedded machine learning", IEEE sensors applications symposium (SAS), Ottawa, ON, Canada, 2023, pp. 1–6.
[78] H. Forehead, J. Barthelemy, B. Arshad, N. Verstaevel, O. Price, P. Perez, Traffic exhaust to wildfires: $PM_{2.5}$ measurements with fixed and portable, low-cost LoRaWAN-connected sensors, Plos One 15 (4) (2020) e0231778.
[79] L. Rothkrantz, "Measurement of air pollution by measurement of traffic density", Smart City symposium Prague (SCSP), Prague, Czech Republic, 2022, pp. 1–8.
[80] V. Kecman, T.M. Huang, M. Vogt, Iterative single data algorithm for training kernel machines from huge data sets: theory and performance, in: L. Wang (Ed.), Support Vector Machines: Theory and Applications, Springer-Verlag, Berlin, 2005, pp. 255–274.