Green and Smart Mining Engineering 2 (2025) 301–312

Contents lists available at ScienceDirect

Green and Smart Mining Engineering

journal homepage: www.keaipublishing.com/en/journals/gsme

Smart sensors, smart calibration: Applications in machine learning for coal dust monitoring

Nana A. Amoah[a], Mirza Muhammad Zaid[a], Xiaosong Du[b], Yang Wang[c], Guang Xu[a,*]

[a] Department of Mining Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA
[b] Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA
[c] Department of Chemical, Environmental and Materials Engineering, University of Miami, Coral Gables, FL 33124, USA

ARTICLE INFO

Keywords:
Smart sensors
Smart calibration
Machine learning
Coal dust monitoring
Artificial neural network

ABSTRACT

The recent resurgence of pneumoconiosis among coal miners in the United States has been linked to their exposure to excessive levels of coal dust. PDM3700 monitors are used in the mining industry to measure each miner's coal dust exposure levels and control overexposure. However, the high cost of the PDM3700 hinders its use in measuring the exposure levels of all miners. Plantower PMS5003 low-cost particulate matter (PM) sensors can measure coal dust concentrations with high spatial resolution in real-time owing to their low cost and small size. However, these sensors require extensive calibration to ensure a high accuracy over long deployment periods. Because they have only been calibrated for mining-induced PM monitoring using linear regression models, the objective of this study was to leverage machine learning algorithms for calibration of coal-dust-monitoring sensors. Laboratory collocation tests were performed using the PDM3700 and aerodynamic particle sizer as reference monitors in a wind tunnel at a wide range of concentrations (0–3 mg/m^3), temperatures (20–32°C), and relative humidities (23%–43%). The results revealed that nonlinear machine learning techniques significantly outperformed traditional linear regression models for low-cost sensor calibration. With the artificial neural network (ANN) being the strongest calibration model, Pearson's correlation of the PMS5003 sensors reached 0.98 and 0.97, those of the Airtrek sensors reached of 0.89 and 0.91, and those of the GasLab sensors reached 0.93 and 0.92. This shows a 2%–11% improvement in model performance over the linear regression model using ANN calibration. The success of the machine learning algorithms used in this study demonstrates the feasibility of deploying low-cost PM sensors for coal dust monitoring in mines.

1. Introduction

Overexposure to respirable coal mine dust has been associated with coal workers' pneumoconiosis (CWP), also known as "black lung" [1,2]. Miners that work in mines where respirable coal mine dust contains at least 5% silica dust have also been diagnosed with silicosis [3]. An increase in the prevalence of CWP has been observed among US coal miners since 2000, and this recent resurgence is considered the most severe in history, with a recent prevalence of 11% among long-term coal miners [4–6]. Coal miners are currently at an increased risk of contracting CWP and silicosis, which can cause permanent long-term disability and premature death. Contrary to the common belief that CWP and silicosis are associated only with long-term exposure to respirable coal dust, recent data indicate that they have also been diagnosed in younger coal miners [7,8].

Effective personal dust monitoring is essential to control overexposure and prevent CWP. Personal real-time-monitoring units can detect overexposure in a timely manner and provide alerts for early control measures. The National Ambient Air Quality Standards and the Mine Safety and Health Administration (MSHA), which regulate particulate matter (PM) concentrations in ambient and mining environments, respectively, have implemented various PM monitoring guidelines [7–9]. To protect miners from respirable coal mine dust (RCMD) overexposure, the MSHA has promulgated regulations requiring underground mines to maintain RCMD end-of-shift concentrations at a time-weighted average of 1.5 mg/m^3 in underground mines [7]. However, these compliance agencies only approve the use of federal equivalent methods (FEMs) and federal reference methods (FRMs) for PM monitoring. The personal dust monitor (PDM) model 3700 is the only MSHA-certified coal-dust-monitoring unit. However, PDMs are

Peer review under the responsibility of Editorial Board of Green and Smart Mining Engineering.

* Corresponding author.
E-mail address: guang.xu@mst.edu (G. Xu).

https://doi.org/10.1016/j.gsme.2025.09.010
Received 5 November 2024; Received in revised form 8 April 2025; Accepted 11 April 2025
Available online 19 September 2025

expensive and bulky [9–11]. A unit of PDM costs approximately \$19,000 and weighs 2.0 kg [11,12]. Because of these drawbacks, only a few miners can use a PDM primarily for regulatory compliance monitoring. These are mainly miners who are exposed to the highest coal dust concentrations at their work locations and those who have already been diagnosed with pneumoconiosis [7]. Therefore, the exposure levels for many miners are not measured.

The limitations of PDMs warrant the need for more affordable coal dust monitoring in mines. Recently, the use of low-cost PM sensors for real-time PM exposure monitoring has gained considerable attention. These sensors can characterize PM with high spatial and temporal resolutions owing to their fast response, low cost, small size, and light weight [13–15]. Several studies have been conducted on these sensors, and these sensors have achieved impressive performance in comparison with FEMs [16,17], FRMs [18,19], and research-grade PM monitors [19–21]. Because of the potential of low-cost PM sensors, regulatory agencies such as the Environmental Protection Agency (EPA) have developed calibration guidelines to improve their accuracy and have designed a roadmap for their adaptation [9]. Laboratory studies have shown that while low-cost sensors provide highly linear responses to reference instruments, their accuracy can be strongly influenced by particle composition, size distribution, and environmental conditions such as relative humidity and temperature [22]. Similarly, field evaluations demonstrated that these sensors could capture fine-scale variations in ambient air pollution, though careful calibration remains essential for reliable deployment in diverse environments [23]. The South Coast Air Quality Management District has examined ways of incorporating low-cost PM sensors into local communities [24,25]. Although these sensors have been widely investigated for ambient PM monitoring, they represent a new type of technology for mining-induced PM monitoring and exhibit questionable accuracy that necessitates the need for more powerful calibration methods. The simple linear regression (SLR) model is the only calibration model used to calibrate low-cost PM sensors for coal dust monitoring [15]. The long-term stability of calibration models for low-cost sensors is significantly influenced by environmental factors, such as temperature and humidity. Continuous recalibration and integration of adaptive calibration models using machine learning techniques are recommended to address sensor drift and ensure accurate air quality monitoring over time. Studies have shown that low-cost sensors exhibit significant measurement biases owing to variations in the temperature and relative humidity (RH). For example, high humidity levels can lead to the hygroscopic growth of particulates, which affects the measurement accuracy of low-cost sensors, particularly when the RH exceeds 75% [26]. Low-cost sensors exhibit variable performance across a range of humidity and temperature conditions, necessitating the development of correction algorithms to mitigate these effects [26]. Moreover, the drift of sensor performance over time is a significant concern. Periodic recalibration is essential to maintain the accuracy of low-cost sensors because they are prone to drift and other errors that may degrade their performance. For example, a single calibration cycle can improve accuracy for only a limited duration before drift occurs, underscoring the need for ongoing calibration efforts [27]. The dynamic adaptability of calibration models is also crucial in addressing the challenges posed by environmental variability. Machine learning approaches have been proposed to enhance calibration models by continuously adjusting for drift and cross-sensitivities to other pollutants [28]. For instance, one researcher found that machine learning could effectively develop corrections for low-cost sensors based on real-time data, thereby improving measurement reliability under fluctuating environmental conditions [29]. This adaptability is essential for maintaining the integrity of air-quality-monitoring networks that rely on low-cost sensors, as they are increasingly deployed in diverse and dynamic urban environments [30]. Recent advancements in mine dust monitoring have significantly evolved, driven by the urgent need to address the health impacts of coal dust exposure on miners. The integration of innovative technologies, including low-cost sensors and machine learning techniques, has transformed the landscape of occupational health monitoring in mining environments [31,32]. One of the most significant achievements is the development of low-cost PM sensors, which have attracted attention in both surface and underground coal mines [32].One of the recent research comprehensively evaluated these sensors and observed their potential for real-time monitoring of inhalable dust [33]. Despite these advantages, commercially available PM sensors often lack inherent calibration, leading to concerns regarding their measurement accuracy. This highlights the need for the development of robust calibration methods to improve the reliability of these sensors in mining applications. One of recent study reported the importance of continuous sampling throughout a miner's shift, as mandated by the MSHA [6,8,33]. They found that prolonged exposure to respirable coal dust, even at concentrations below the permissible exposure limit (PEL) of 1.5 mg/m^3, could contribute to severe health issues, including CWP. The integration of low-cost sensors into routine monitoring practices can facilitate timely interventions in mitigating dust exposure and improving miners' health [34].

SLR calibration methods cannot account for other environmental factors, such as temperature and RH. This affects their transferability to environments that differ from their calibration environments, which alters their accuracy in ambient deployment [23,35]. A more advanced MLR model has been developed to consider the effects of temperature and RH on low-cost PM sensors. However, the MLR algorithm is more dependent on random variations in the sensor outputs, making it more sensitive to the test conditions [36,37]. Therefore, the limitations of these calibration models necessitate the use of more advanced calibration models based on machine learning algorithms for evaluating and calibrating low-cost PM sensors [26–29]. A study reported the use of a feedforward neural network (FNN) model to calibrate Plantower PMS7003 low-cost sensors, which improved the accuracy from 0.618 to 0.905 using *R*-squared as an indicator against the beta attenuation monitor (BAM-1020) [38]. Building on this, random forest (RF) and support vector machine (SVM) algorithms have also been shown to deliver superior performance when compared with other machine learning approaches [39,40]. Furthermore, another investigation evaluated and calibrated a Plantower PMS5003 low-cost sensor using the Synchronized Hybrid Ambient Real-time Particulate (SHARP) monitor [41]. In that work, simple regression models (SLR and MLR) were compared with two machine-learning calibration algorithms, XGBoost and the FNN, and it was found that the FNN provided the most accurate results, with a root-mean-squared error of 3.91 [42]. Despite the potential of these methods, machine learning algorithms have not yet been applied to the calibration of low-cost PM sensors for coal dust monitoring under underground mining conditions.

The aim of this study was to calibrate low-cost PM sensors for coal dust monitoring using machine learning algorithms. Laboratory tests were performed under various coal dust concentration, temperature, and RH conditions. The MSHA FEM Personal Dust Monitor (PDM) model 3700 and the research grade aerodynamic particle sizer (APS) model 3321 were used as reference monitors. Based on the data from the experiment, the sensors were calibrated using five machine learning algorithms and compared to determine the most accurate calibration method. Widely used SLR and MLR models, together with three other machine learning algorithms—artificial neural network (ANN), SVM, and RF regressor (RFR)—were selected in this study.

2. Experimental

The experimental setup used for the calibration experiment is described in Ref. [15]. The experimental setup consisted of a 0.5 m × 0.5 m U-shaped wind tunnel designed to simulate underground mine airways. It was equipped with a dust injection system to control the coal dust concentrations within the wind tunnel and monitoring systems to regulate the temperature and RH.

2.1. Reference monitors

The EPA FRM PDM model 3700 developed by the National Institute of Occupational Safety and Health (NIOSH) and the research-grade APS were the reference monitors used in this study. The PDM is a real-time personal coal dust monitor that operates based on the principle of a tapered element oscillation microbalance (TEOM) [43]. Its accuracy, precision, and comfortability are validated by NIOSH and approved for regulatory monitoring by the MSHA. Laboratory tests by NIOSH have demonstrated with 95% confidence that PDM measurements are within ± 25% of true concentrations. Field precision tests also showed that the PDM had a standard deviation of 0.078% [44]. The PDM is equipped with a respirable size coal dust cyclone with a cutoff size modeled to closely mimic the respirable curve of the human respiratory system. Therefore, only respirable-sized particles pass through the mass transducer, and the mass concentration is determined using the TEOM method.

The APS measures the PM mass and number concentration by particle aerodynamic sizes of 0.5–20.0 μm using time-of-flight principles. A PM concentration of particle size smaller than 4.37 μm from the APS was adopted in this study to make it comparable with the PDM because it compares with the 4.37 μm D_{50}, where D_{50} refers to the aerodynamic diameter at which a cyclone or impactor collects 50% of particles and allows the other 50% to pass through. The APS received ambient PM-laden air into the monitor through a nozzle at an accelerated flow rate of 5.0 L/min. The accelerated airflow passed through the sensing zone where the PM concentrations were measured. The particle size distribution was then reported every 15 s based on the settings used in this study [45].

2.2. Low-cost sensors

Three low-cost PM sensors were calibrated in this study: Plantower PMS5003 (PMS) low-cost PM, Airtrek PM, and GasLab CM-505 multigas sensors. Two units of each monitor were evaluated to compare their intramodel precision. The GasLab sensor (Fig. 1(a)) measured $PM_{2.5}$ and PM_{10} particle concentrations, together with oxygen, carbon dioxide, and carbon monoxide concentrations. GasLab used a combination of non-dispersive infrared (NDIR), electrochemical, and fluorescent sensors to measure PM and other gas concentrations. In contrast, the Airtrek sensors (Fig. 1(b)) used light-scattering principles similar to those of the Plantower PMS5003 sensor. The Airtrek sensors measure PM in four size bins: 1.0, 2.5, 4.0, and 10.0 μm.

PMS sensors are inexpensive PM detection and measurement sensors that operate based on light-scattering principles. These sensors are commercially available at a cost of approximately $30 per unit. Each sensor draws the ambient airflow through an inlet into the sensing area using a fan. As illustrated in Fig. 1(c), the light-emitting diode radiated laser-induced light into the sensing area to target the particles within the measuring cavity. The light scattered by the particles was detected by a photodetector, which transmitted pulses of the electrical signal to the inbuilt microprocessor. The number and intensity of the electrical signals measured by the microprocessor were then converted into number and mass concentrations, respectively, based on Mie scattering (MIE) theory.

These PMS sensors characterize PM by size as PM_1, $PM_{2.5}$, and PM_{10}. The manufacturers of the PMS sensors state that PM_1 is measured for particle sizes of 0.3–1.0 μm, $PM_{2.5}$ for particle sizes of 1.0–2.5 μm, and PM_{10} for particles sizes of 2.5–10.0 μm. For each size category, the PMS reports 2 PM outputs: one without any form of correction factor (CF), called the standard PM concentration (or CF = 1), and the other with an atmospheric calibration factor, called the environmental PM concentration. The manufacturers have not published any details regarding the calibration factor or its development. Therefore, because of considerable uncertainties regarding manufacturer calibration, standard PM concentrations were used in this study. The manufacturer's specifications indicate that the performance of these sensors is not affected by non-PM aerosols, owing to the inbuilt shielding technology.

Together with a DHT-22 temperature and RH sensor, NodeMCU ESP8266, and a four-line by 20-character liquid crystal display (LCD) screen, the PMS was integrated into a prototype monitor (low-cost PM monitor), as shown in Fig. 1(e). This monitor was programmed with an Arduino IDE to display real-time concentrations of $PM_{2.5}$, PM_{10}, temperature, and RH readings every 1.0 s on the LCD screen. The sensor and its components were housed in a 10.0 cm × 4.0 cm × 1.5 cm acrylic

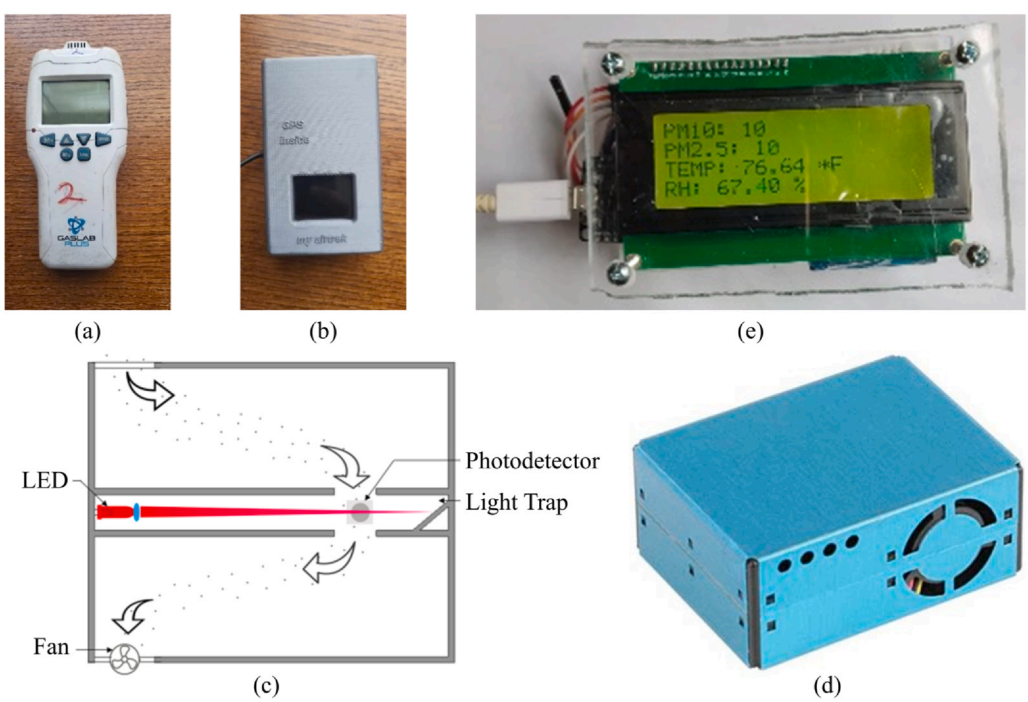

Fig. 1. Sensors and reference monitors used in the experiments: (a) GasLab CM-505 multigas sensor; (b) Airtrek PM sensor; (c) schematic of PMS5003 sensor; (d) Plantower PMS5003 sensor; (e) PM monitor.

plastic box to protect the low-cost PM sensor and its critical components from rainfall, sunlight, and physical damage. This also enabled the team to observe the screen readings in real time through the transparent acrylic case. The sides of the acrylic case were strongly fastened but left uncovered to allow for normal operation of the PMS sensor without interference. The low-cost PM monitor was continuously powered using a 5.0 V USB cable. For data analysis, the PM monitor was interfaced with a ThingSpeak MATLAB-based online Internet of things (IOT) platform that served as a cloud in which all data were transmitted through Wi-Fi.

2.3. Data collection

A custom-built wind tunnel was constructed using acrylic glass for the experiments. The wind tunnel was equipped with a particle generator that was used to control the coal dust concentrations within the tunnel. A heater and humidifier were installed to control temperature and RH, respectively. A constant air velocity of 1.0 m/s was maintained in the wind tunnel throughout the experiment, whereas the coal dust concentrations were varied between 0 and 3 mg/m^3. Temperatures ranging from 20 to 32°C were maintained within the wind tunnel using its heating system and the heating, ventilation, and air-conditioning system of the laboratory building. RH of 23%–43% was generated within the wind tunnel using a humidifier. The sensors and reference monitors were exposed to a combination of these conditions and the coal dust concentrations. The selection of a coal dust concentration range of 0–3 mg/m^3 for the experiments is justified by the regulatory framework established by MSHA, which stipulates a PEL of 1.5 mg/m^3 for an 8-h time-weighted average. The rationale behind this regulatory standard is based on extensive epidemiological studies demonstrating a clear correlation between cumulative dust exposure and the onset of respiratory diseases in coal miners [45,46]. Moreover, the upper limit of 4 mg/m^3 is consistent with findings of various studies that indicate that although the PEL is set to 1.5 mg/m^3, actual exposure levels in certain mining environments may exceed this threshold. For instance, coal dust levels have been observed to reach 4.81 mg/m^3 which is close to the experimental values obtained in our study [47].

The PDM recorded concentrations every minute; however, other sensors and monitors were programmed to record data every 15 s. The experiment lasted for approximately 6 h per d for 8 d. The clock times on all sensors and monitors were reset to synchronize their time stamps.

2.4. Data processing

The time-weighted average (TWA) concentration data reported by the PDM were converted to minute-by-minute real-time concentrations using Eq. (1) for consistency with other sensor concentration units to clean up the data from these sensors for analysis.

$$C = (\mathrm{TWA}_n \times \frac{T_n}{T}) - (C_1 + C_2 + C_3 + \cdots + C_{n-1}) \tag{1}$$

where TWA_n is the time-weighted average at each time step, C_n is the measured real-time concentration, T is the time interval between successive measurements, and T_n is the total number of minutes at time n.

Further processing required the four 15-s concentrations reported each minute by the sensors to be averaged to obtain minute-by-minute concentrations and make them comparable with the PDM data. A boxplot method of outlier detection was used to detect outliers. The PMS sensors were the only sensors that recorded outlier concentrations. Because the number of detected outliers was significantly small, these data points were deleted without any effect on the analysis.

3. Data analysis

3.1. Simple linear regression

The performance of these low-cost PM sensors was evaluated using the SLR method to establish calibration curves. For light-scattering low-cost PM sensors, establishing a linear relationship with PM concentration is desirable. In this study, this relationship was determined by establishing correlations between the sensor output and true concentrations determined using the PDM. The concentrations obtained by the reference monitors were used as independent variables, whereas those measured by the sensors were used as dependent variables. Each sensor was evaluated and monitored to obtain a calibration curve. The least-squares method was used to fit the calibration curve through the data points to determine the calibration equation, as expressed by Eq. (2), where y is the reference monitor concentration, α is the y intercept, β is the slope of the calibration curve, and x is the sensor outputs. This method uses the sum of squares method to establish the curve with the least sum of squares as the curve of best fit to calibrate the sensors. By using the ordinary least-squares linear regression method, R-squared values were calculated to quantify the relationship between sensor outputs and true concentrations and to determine the percentage of true concentration that could confidently be predicted using these sensors. Pearson's correlations were used to determine the strength of the linear relationships.

The mean squared-error (MSE) was determined using the test dataset to evaluate the performance of the model to compare the results of the evaluation with those of other methods used in this study. However, a major disadvantage of this calibration method is that it can only estimate the relationship between two variables. Low-cost PM sensors are influenced by other factors, such as temperature and RH, which cannot be considered in SLR models.

$$y = \alpha + \beta x \tag{2}$$

3.2. Multiple linear regression

Because SLR models are limited to single dependent and independent variables, MLR models were used to account for the effects of temperature and RH. These factors, together with the sensor outputs, were then considered multiple independent variables, whereas the true concentration determined by the reference monitor remained the dependent variable. Similar to SLR, the least-squares approach was used to estimate the values of the model coefficients required to establish the MLR calibration model. The coefficients were then used to determine the minimum sum of the squared deviations between the data points and the plane that established the plane of best fit. The nature of the equation for the plane, considered the calibration model, is expressed by Eq. (3). Here, y is the true concentration, β_0 is the constant term, β_1 to β_3 is the slope coefficient for each independent variable, and C is the model error term. The correlation values for each sensor were calculated to assess the strength of the correlation between the sensor outputs and the independent variables.

On the basis of a 70:30 train–test split, the model was trained using a training dataset. The R-squared and Pearson correlation values were calculated to determine the true concentration that could be correctly determined using low-cost sensors. P-values were calculated using statistical tests to determine the statistical significance of the regression model. A test dataset was used to evaluate the performance of the model. Performance was evaluated using the MSE.

$$y = \beta_0 + \beta_1 \times \mathrm{Concentration} + \beta_2 \times \mathrm{Temperature} + \beta_3 \times \mathrm{RH} + C \tag{3}$$

3.3. Artificial neural network

Low-cost PM sensors respond to PM concentration, and the effects of temperature and RH may not always be linear. A more comprehensive algorithm, such as the ANN, is crucial for modeling nonlinear relationships within a dataset. The ANN algorithm is inspired by biological neural networks within the human brain. The neural network within an ANN trains the data by matching the inputs of the samples with their corresponding outputs. This matchmaking method is used to

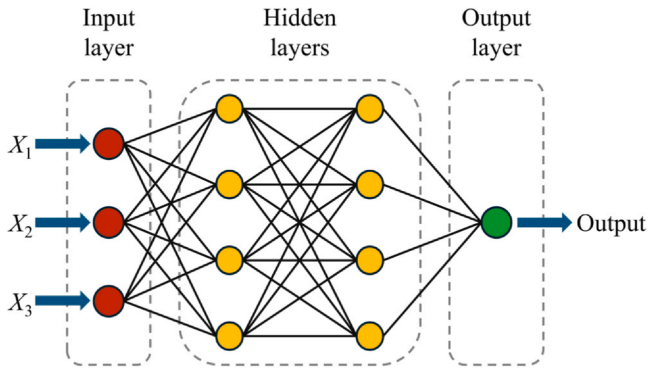

Fig. 2. Structure of the ANN architecture.

generate and optimize the probability-weighted relationship between the inputs and outputs. A schematic of this algorithm and the structure of the ANN architecture are shown in Fig. 2. As shown in Fig. 2, the input layer introduces the input variable data into the model for further processing. In this application, the input layer consists of the sensor outputs, temperature, and RH. Two hidden layers were used in the algorithm. These layers are responsible for assigning and optimizing weights to each connection and adding a bias term to its sum, resulting in Eq. (4). In the case of a two-layer hidden layer in this model, this process is carried over to the second hidden layer, where a new set of weights is assigned, and a new set of biases is added. An activation function is then applied to the resulting linear equation before moving to the next hidden layer. A ReLU activation function was used in this model and programmed in the hidden layers. The results of these computations are sent to the output layer of the model.

The true concentrations determined using the PDM were used as nodes in the output layer. The dataset was randomized and split into training and testing datasets. The 70:30 train–test split method was used. Thus, 70% of the dataset was used as the training set to train the algorithm for prediction, and 30% was used as the testing dataset to evaluate the performance of the algorithm. The MSE was used to analyze model performance.

$$Z = f(W_1X_1 + W_2X_2 + W_3X_3 + \dots + W_iX_n + B) \tag{4}$$

where W_i represents the assigned weights, X_n represents the input variables, B is the added bias, and $f(x)$ is the rectifier linear unit (ReLU) activation function.

3.4. Random forest

The RFR algorithm is a supervised machine learning model that uses an ensemble of decision trees to solve regression problems. RF models combine a multitude of decision trees, each of which trains inputs to their respective outputs to predict the outputs. Each decision tree begins with a root node that repeatedly splits into internal nodes through a set of branches until a leaf node is obtained and a decision is made.

A sample decision tree regressor (Fig. 3) illustrates the sequence of the algorithm. The 70:30 train–test split was applied in this algorithm, and a bootstrapped dataset was generated from the training dataset. This method involves the random sampling of a specific number of variables from the dataset of each decision tree to obtain a multitude of trees. As shown in the sample calibration tree in Fig. 3(a), the tree starts with the assigned root node, with the coal dust concentration being the strongest predictor. This node is split into branches with concentrations exceeding 2.0 mg/m^3. If this condition is satisfied, the algorithm follows the path of determining the concentration to be 2.5 mg/m^3. However, the algorithm follows the path to the next node if the concentration is lower than 2.0 mg/m^3. Although this explains the sequence of the decision tree, these steps are repeated in several interior nodes until the algorithm reaches the leaf node when a decision is made on concentration prediction, considering the temperature and RH variables.

An ensemble of decision trees was developed based on this algorithm and trained using the same data. The results of all models were combined via an ensemble-learning process. The results of all the trees used in the model were combined to determine the final RF prediction, as shown in Fig. 3(b). This model makes concentration predictions based on the outputs of all decision trees involved in the training, and ultimately determines the predicted concentration as the mean concentration from all trees. The MSE criterion was calculated to evaluate model performance.

3.5. Support vector machines

This algorithm is a supervised method that predicts discrete values that are similar in principle to categorical SVM. Similar to linear regression principles, the support vector regressor (SVR) determines the line of best fit through the data points, and in the case of higher-dimensional data, obtains the hyperplane of best fit. However, unlike linear regression, which minimizes the sum of the squared errors between data points and the plane of best fit, the objective of SVR is to establish the line of best fit within a threshold error value.

The SVR algorithm achieves this by minimizing the objective function expressed by Eq. (5), which minimizes the error between the line of best fit and the margins. The corresponding error term in this case is then considered the constraint given in Eq. (6), where the absolute error is set to be less than or equal to the acceptable error of the proposed model. However, some data points are likely to fall outside the constraint margins. A slack variable (the second term in Eq. (5)) is added to minimize the error margin to account for these points in the optimum hyperplane generation.

The difference between the hyperplane of best fit and the data points outside the error margins introduces errors in the algorithm. Consequently, the deviations are added to the new objective function and minimized to account for these data points and minimize the error margins. This algorithm maximizes the prediction of coal dust concentration within an acceptable margin of error of 5%. The 70:30 train–test split was applied to train and test the model, and MSE was used to evaluate the performance of the model.

$$\mathrm{MIN} = \frac{1}{2}\,||\boldsymbol{w}||^2 + C\sum_{i=1}^{n}|\boldsymbol{\xi_i}| \tag{5}$$

$$|\boldsymbol{y}_i - \boldsymbol{w}_i\boldsymbol{x}_i| \le \varepsilon + |\boldsymbol{\xi}_i| \tag{6}$$

where, $\boldsymbol{w}$ denotes the weight vector of the model, $\boldsymbol{w}_i$ is i-th component of the weight vector $\boldsymbol{w}$ while $\boldsymbol{x}_i$ is the input feature vector for the i-th data sample. The scalar $\boldsymbol{y_i}$ represents the corresponding true target value. The slack variable ξ_i accounts for deviations beyond the ε-insensitive margin, with ε specifying the tolerance band within which no penalty is applied for prediction errors. The regularization parameter $C > 0$ balances the trade-off between minimizing model complexity, represented by the squared Euclidean norm $||\boldsymbol{w}||^2$, and penalizing the magnitude of deviations captured by ξ_i.

4. Results and discussion

The results of the sensor evaluation and the performance of the calibration models were analyzed to compare the performances of the five calibration models used in this study. The performances of the sensors were also evaluated to determine the feasibility of adopting these low-cost PM sensors for coal dust monitoring using the investigated calibration models.

4.1. Preliminary data analysis

The raw data obtained from the experiments are presented in Fig. 4(a) and (b). Fig. 4(a) and (b) shows the time-series data of the raw

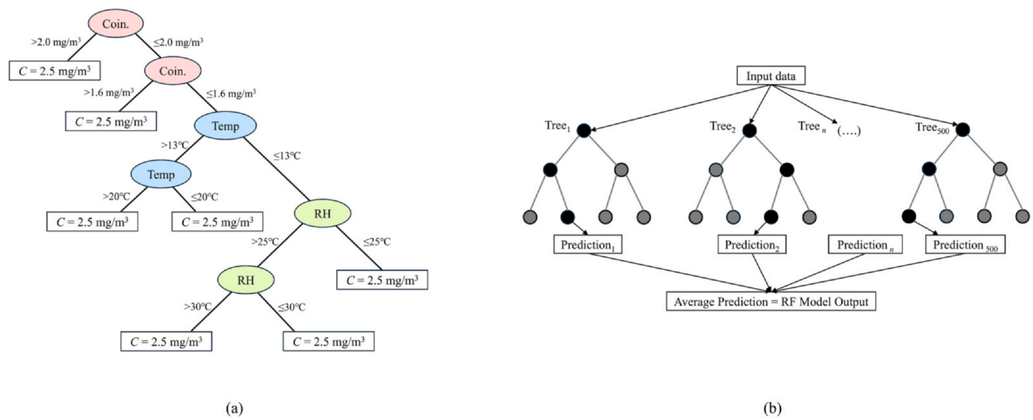

Fig. 3. Representation of RFR algorithm: (a) RF decision tree structure; (b)RF architecture.

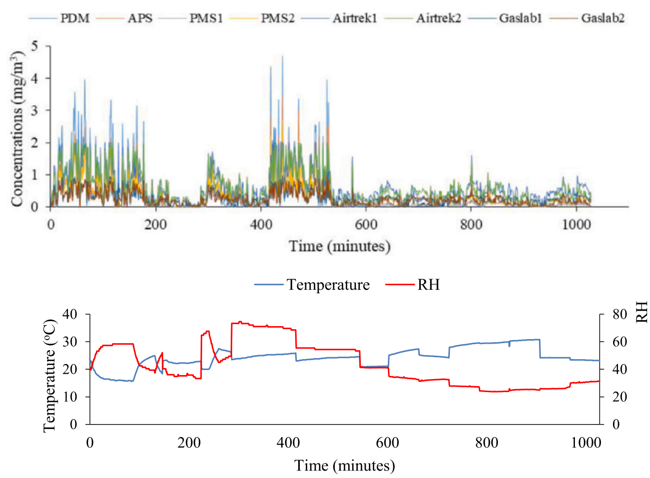

Fig. 4. (a) Time-series data comparing the responses of the reference monitors and low-cost PM sensors. This figure shows a total of 1028 min (17.2 h) of monitoring data. (b) Time-series data of RH and temperature conditions throughout the testing period.

sensor outputs recorded over the testing period for all six low-cost sensors and two reference monitors. The corresponding time series for RH and temperature conditions are presented in separate plots in Fig. 4(b) to visualize the trends under these conditions during the testing period. Sensor outputs were corrected using the calibration models developed in this study. The data presented here represent the combined results from several days of monitoring. Variations in the outputs and the presence of spikes in these data were the result of changes in the concentration within the wind tunnel throughout the testing period. The performance of the calibration models on these data is discussed in subsequent sections.

4.2. Simple linear regression

The results of the SLR calibration models are presented in Figs. 5 and 6. Fig. 5 shows a pairwise correlation between all sensors and monitors used in this study presented as scatter plots, with the correlation coefficients for each scatter plot displayed. These plots were generated using uncalibrated data from the sensors to clarify the behavior of the outputs generated by the sensors relative to the reference monitors. All the sensors exhibited a positive response to changes in concentration, with correlation values ranging from 0.83 to 0.98. In general, the response of the sensors were in better agreement with the reference monitors than with the results of a previous study [48]. This strong performance is linked to the considerably higher number of data points used in this study. A higher number of data points for regression has been demonstrated to statistically improve model performance. However, although the two PMS sensors yielded an entirely linear response with correlation coefficients of 0.95, the Airtrek and GasLab sensors had significantly lower correlations with PDM. The lower correlation of the Airtrek and GasLab sensors was caused by their limited upper concentration limits of 2.0 and 1.0 mg/m^3, respectively, creating an exponential trend for their data. These results are consistent with those of a previous study that highlighted the measurement limitations of the GasLab and Airtrek sensors [48].

A SLR calibration model, developed to correct the sensor response, was trained using 70% of the dataset and tested using the remaining 30%. The model was trained using the ordinary least-squares regression method to generate an SLR model. Each model built for each sensor was in the format discussed in Section 3.1, using the coefficients of the sensor outputs, slope of the calibration curve, and intercept. Fig. 6 shows a comparison of the actual and predicted values generated from this model using the testing dataset, with the correlation coefficients for each model shown in the plots. These results were compared with the uncalibrated responses of the sensors. The actual values are represented by the PDM concentrations on the x-axis, whereas the predicted values are indicated by the sensor-corrected concentrations on the y-axis. The Pearson correlation coefficient was used to evaluate these models over the coefficient of determination because, whereas the latter is used to determine the correlation of the model, the Pearson correlation determines the strength and direction of the linear relationship. These results indicate slight improvement in the sensor performance, with minimal to no improvement in the correlation values, compared to the uncalibrated performance.

A significant drawback observed in these models is that for models whose y-intercept was negative, as shown in Fig. 6(b), (c), (e), and (f), the model predicted negative concentrations for PDM concentrations lower than the x-intercept. Another critical limitation of using an SLR model is its inability to account for multiple independent variables.

4.3. Multiple linear regression

The effects of RH and temperature were considered and analyzed in the proposed MLR model, which was developed to calibrate low-cost PM sensors. This was expected to yield a better performance than the SLR method because temperature and RH influence the outputs of optical PM sensors. The results of these models are presented in Fig. 7, which compares the performance of the MLR calibration model with that of the SLR model. An MLR calibration model was generated for each sensor in the format described in Eq. (3), Section 3.2. This model was trained using 70% of the sensor dataset and tested using the remaining 30%.

After accounting for RH and temperature, the MLR models calibrated the low-cost PM sensors slightly better than the SLR models.

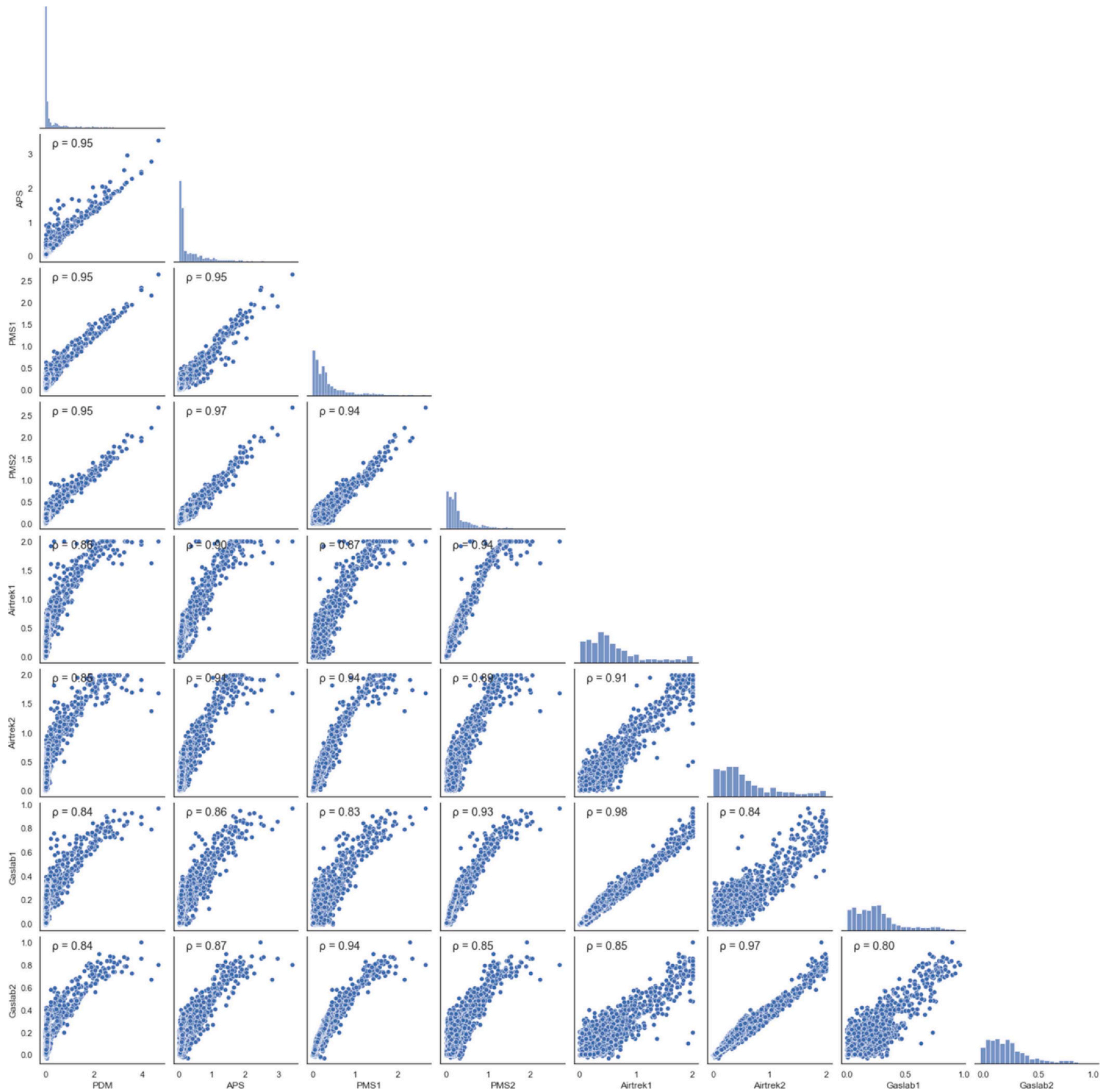

Fig. 5. Pairwise correlation among the six sensors and the PDM and APS using the 70% training dataset. The *x*- and *y*-axes represent the concentration outputs of the sensors and monitors (in mg/m^3), respectively.

Evaluating the performance of the model on 30% of the data unseen by the model revealed a negligible increase in Pearson's correlation coefficient by 0–3% among the sensors compared with those of the SLR models and yielded moderate MSE values (0.20–0.33 mg/m^3). This evaluation compared the actual concentrations recorded by the PDM with the concentrations predicted based on the raw sensor output, which were corrected using the MLR model. These results suggest that accounting for temperature and RH correction in linear regression models to calibrate low-cost sensors can improve their performance, depending on the degree of influence of temperature and RH. This demonstrates the influence of temperature and RH on the performance of low-cost PM sensors, making it necessary to consider them in sensor calibration models, although the performance increase is marginal using MLR models. However, in this case, temperature and RH minimally impacted the performance of the PMS sensors, which explains why the performance increased slightly, accounting for temperature and RH. The results showed that the performance increase was relatively higher for the Airtrek and GasLab sensors, which exhibited exponential responses, suggesting that these sensors were influenced more by temperature and RH. The results also indicated that the exponential response in the Airtrek and GasLab data could not be corrected using the MLR model. Moreover, the model predicted negative values for concentrations lower than the *x*-intercept of the plane of best fit.

4.4. Random forest regressor

The RFR calibration model significantly outperformed both the SLR and MLR models for all six sensors. The results of the RFR prediction

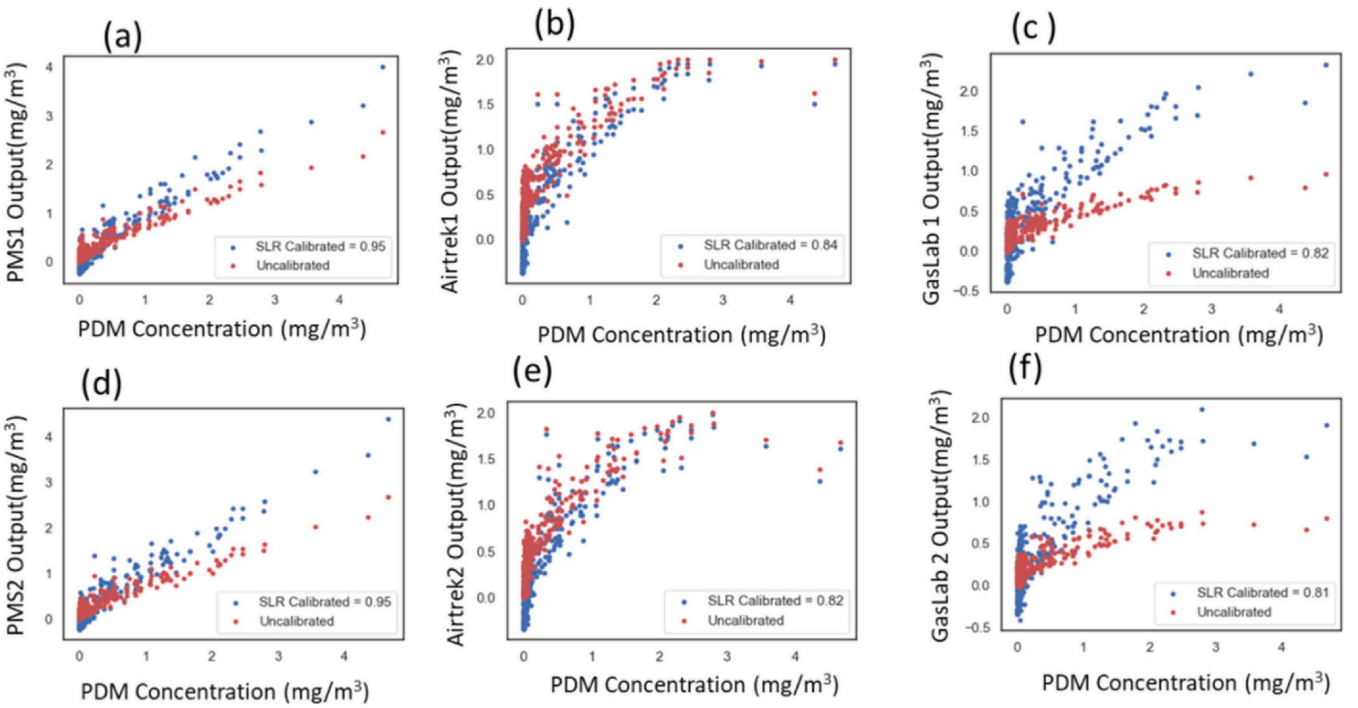

Fig. 6. SLR calibration model showing predicted calibrated sensor outputs vs. PDM concentration. The sensors are (a) PMS1, (b)Airtrek1, (c) GasLab1, (d) PMS2, (e)Airtrek2, and (f) GasLab2.

model are presented in Fig. 8, which compares its prediction performance with that of the MLR model. This model emerged as one of the best-performing calibration models investigated in this study, showing an improved Pearson's correlation (0.97 for both PMS1 and PMS2), with an MSE of 0.02 mg/m^3. The Airtrek sensors yielded improved correlations of 0.89 and 0.91 for Airtrek1 and Airtrek2, respectively, with an MSE of 0.07 mg/m^3 for both sensors. The GasLab sensors also showed improved correlations of 0.93 and 0.92 for GasLab1 and GasLab2, respectively, with corresponding MSE values of 0.04 and 0.06 mg/m^3. As shown in Fig. 8, the poor performance of the regression models at lower concentrations, leading to the prediction of negative concentrations, was corrected by the RFR model, showing a significant improvement in correction performance for all six sensors, as indicated by the Pearson correlations shown in the plots. Moreover, a near-perfect 1:1 linear relationship was established, even for sensors that initially yielded exponential curves. The results of this model showed that the RFR calibration values were more condensed around the calibration curve, indicating a better fit than regression models. This is more

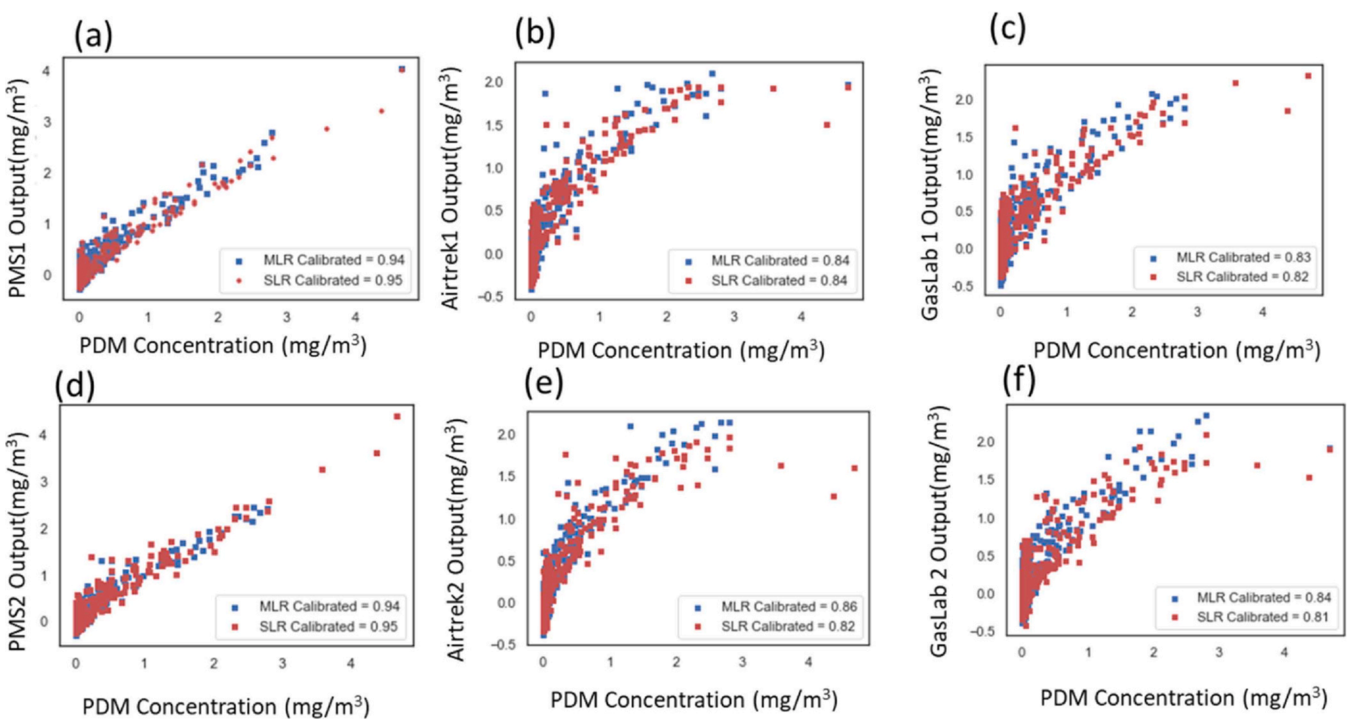

Fig. 7. MLR results showing its comparison with the SLR model. The sensors are (a) PMS1, (b)Airtrek1, (c) GasLab1, (d) PMS2, (e) Airtrek2, and (f) GasLab2.

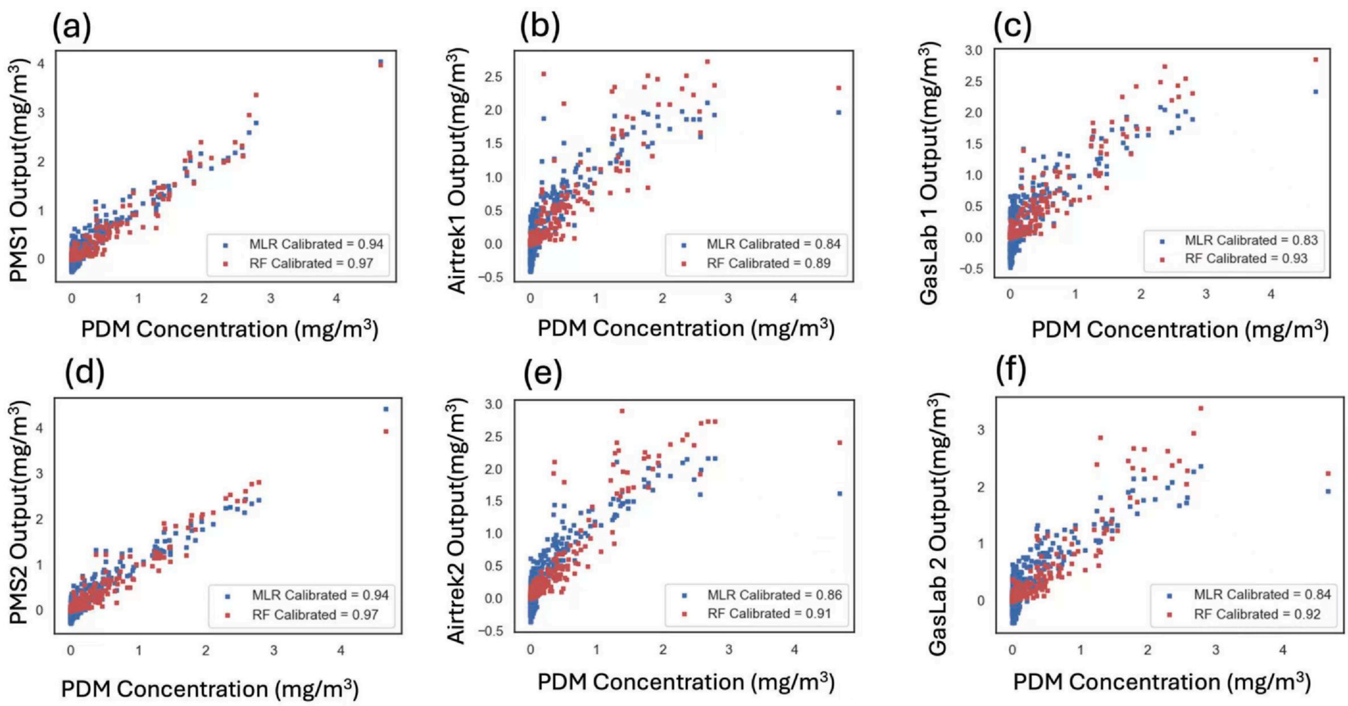

Fig. 8. RFR model performance for low-cost PM sensors using the testing dataset. The correlation plots show actual data values (PDM concentration) against predicted values, which is considered sensor corrected outputs (sensor outputs) for all six sensors. Each plot shows the model performance of each model using Pearson's correlation coefficient. The sensors are (a) PMS1, (b)Airtrek1, (c) GasLab1, (d) PMS2, (e)Airtrek2, and (f) GasLab2.

significant for the GasLab and Airtrek sensors, which initially showed exponential trends. These findings are consistent with those of other studies that have demonstrated the potential of RF models to outperform regression models. For example, a study improved the correlation between the HK-B3 PM sensor and the MicroPEM monitor from 0.87 using linear regression to 0.98 using the RFR model [49]. This is not surprising because of the ability of the RF model to balance datasets with numerous variables and variations, making it suitable for complex models.

As demonstrated by these results, RF can potentially be used as the calibration model of choice for low-cost PM sensors in coal dust monitoring. However, the model can be complicated by multiple parameters that must be considered when developing the model. If incorrect values are used, this will directly impact the model performance. In this study, the hyperparameters were tuned using the RandomizedSearchCV method of Scikit-Learn by performing a five-fold cross validation with each iteration. In each iteration, a different combination of model settings is used to search for the optimum hyperparameters. The optimum combination of hyperparameters obtained from this search was used to train the training dataset to obtain the results. In this study, the optimum hyperparameters included 500 trees, "auto" number of features in consideration at every split, a maximum number of levels allowed at each decision tree of 100, a minimum sample number of 3 to split at a node, a minimum sample leaf of 3, and a "true" bootstrapped method for sampling data points.

The RF calibration model structure highlights the relative importance of the explanatory variables evaluated in this study. For these low-cost PM sensors, the coal dust concentration response from the sensors emerged as the most crucial predictor variable in the calibration, followed by RH. The temperature had the least effect on the decision tree. This was determined by examining the rate of change in the MSE when each variable was permuted in a decision tree. The decision trees in this model were grown based on the importance of each explanatory variable for predicting the true concentration. The MSE was calculated each time, and an explanatory variable was included in the bootstrapped dataset. The importance of each parameter was determined by calculating the rate of increase in the MSE during the bootstrapping process. For a variable that significantly influenced the prediction of the true concentration, permuting the variables in the model significantly increased the MSE.

4.5. Artificial neural network

The ANN model used in this study to calibrate the sensors yielded similar results with the RF model. Similar to the RF model, the ANN model significantly outperformed the SLR and MLR models. The model corrected the negative concentrations predicted by the regression models while optimizing the 1:1 relationship between the PDM and sensors. The prediction performance of the ANN model was compared with that of the RFR model side-by-side (Fig. 9). Together, both models exhibited the best performance among all the models used in this study, with a similar Pearson's correlation. Although the ANN model slightly outperformed the RF model by 1% and 2% for PMS1 and Airtrek1, respectively, the Pearson's correlations for the remaining four sensors remained the same for both models. Therefore, the ANN is considered the best-performing model for low-cost PM sensor calibration in coal dust monitoring, although it can be interchanged with the RF model because of the closeness of their performance.

The parameters used in the ANN model determined the results of the model. Tuning the hyperparameters of the layers in the ANN model revealed that the optimum learning rate was 0.1 and the batch size was 30 with 60 epochs using the MSE as the loss function. With three input values in the input layer, the values were passed to an optimum two-layer hidden layer, each with 10 neurons. Finally, the output layer contained one neuron, which contained the predicted coal dust concentration output. These parameters must be appropriately deployed to ensure model accuracy and prevent suboptimal model performance. For example, running the ANN model used in this calibration application for more than 60 epochs is computationally expensive and offers insignificant improvement in model performance.

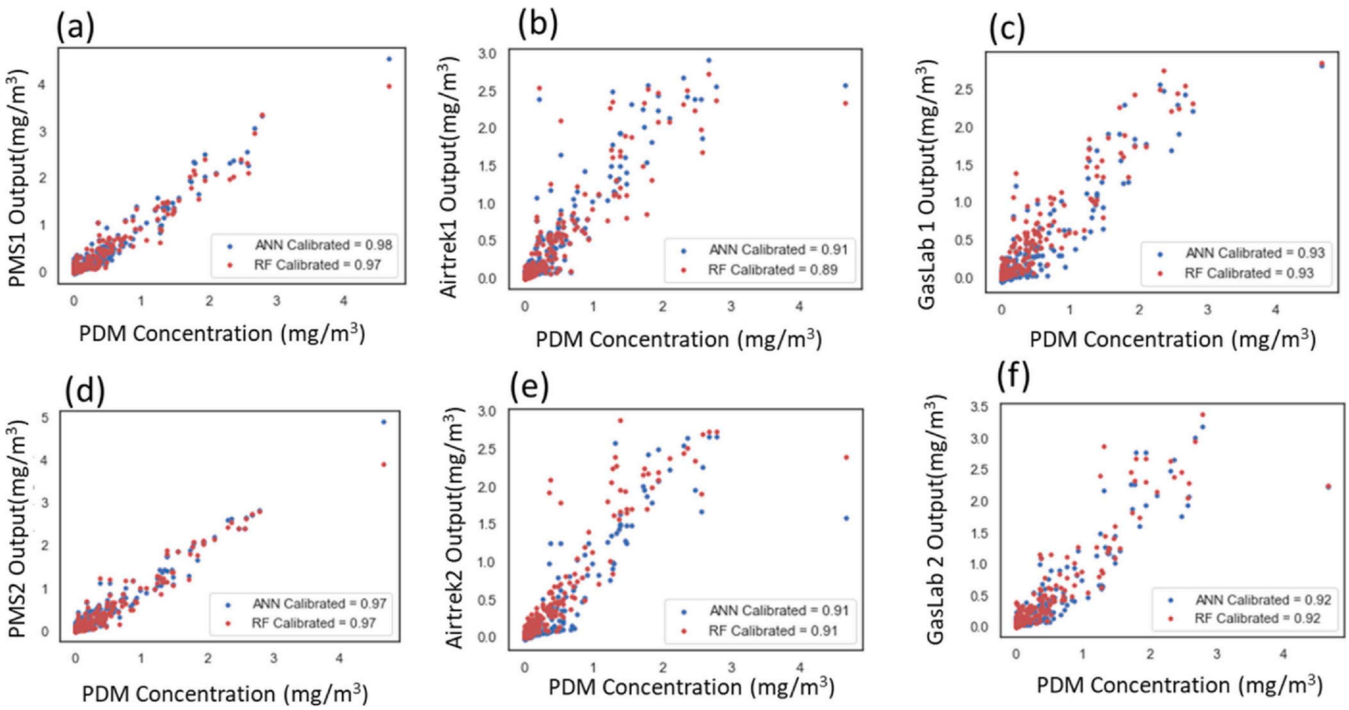

Fig. 9. Results for ANN models showing calibrated sensor outputs vs. PDM concentration, showing comparisons between the ANN and RFR models. The sensors are (a) PMS1, (b)Airtrek1, (c) GasLab1, (d) PMS2, (e)Airtrek2, and (f) GasLab2.

The results discussed in this section are for the application of the optimum hyperparameters in training the training dataset and testing using the testing dataset, which constituted 30% of the entire dataset. The model was evaluated using Pearson's correlation coefficient, which yielded values of 0.98 and 0.97 for PMS1 and PMS2, respectively. These results are consistent with those of other studies that applied ANN to calibrate the Plantower PMS7003 sensor, increasing its R^2 value from 0.86 to 0.97 [49]

4.6. Support vector machines

In addition to the SLR and MLR machine model, the SVM exhibited the worst calibration performance among the evaluated models. Because linear regression models have the drawback of not fitting no-so-linear dataset with high accuracy, the SVM model was expected to significantly correct the outputs of the sensors. However, inaccuracy tendencies remained in the predictive evaluation. The results for the

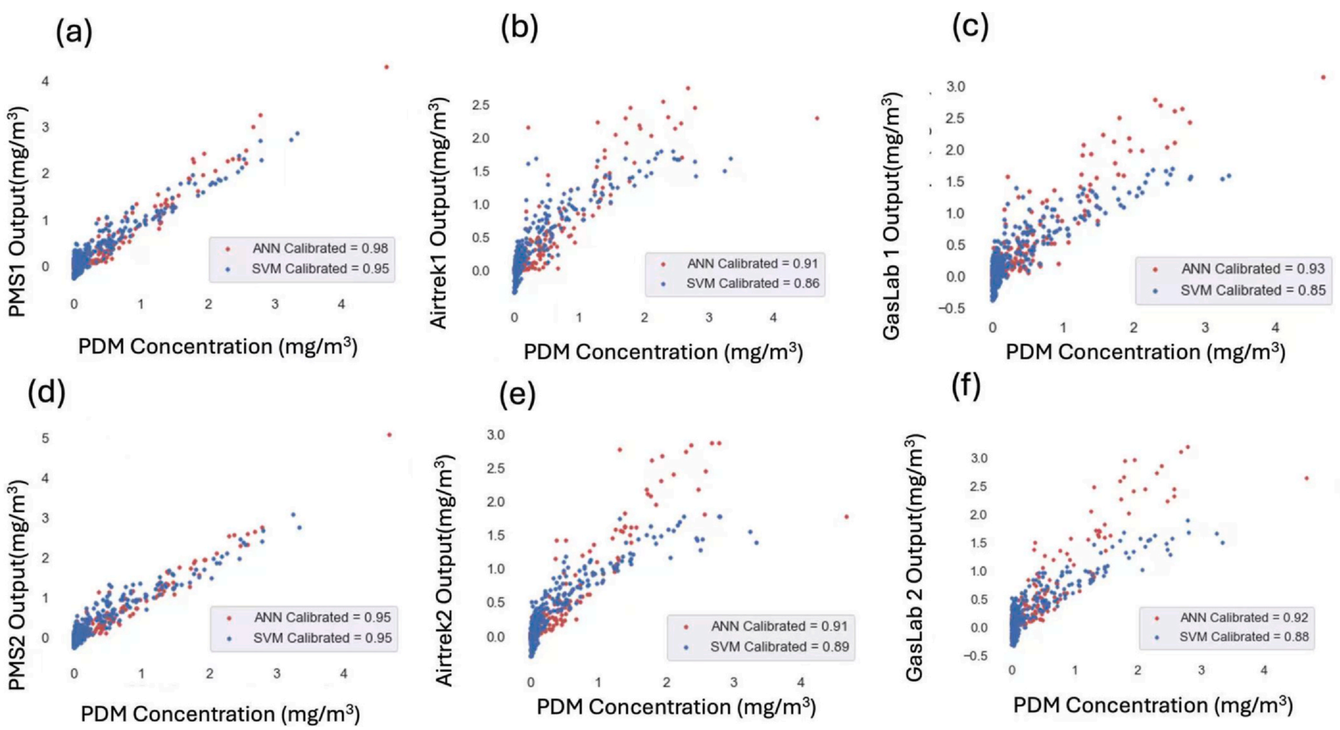

Fig. 10. Model performance of SVM and ANN model testing datasets for variations in PDM concentration with predicted sensor calibration output. The sensors are (a) PMS1, (b)Airtrek1, (c) GasLab1, (d) PMS2, (e)Airtrek2, and (f) GasLab2.

SVM model are presented in Fig. 10 as scatter plots of the PDM concentrations against the predicted calibrated sensor outputs compared with the ANN model results. The results highlight the significant difference in the performance between these two models. Although the SVM model results showed minimal improvement in sensor performance for the Airtrek and GasLab sensors, it did not perform better than other models tested for the PMS sensors after calibration. The correlations of Airtrek1 and Airtrek2 increased from 0.84 and 0.82 to 0.86 and 0.89, respectively, corresponding to MSEs of 0.30 and 0.29 mg/m^3, respectively. The GasLab sensors recorded an increase in correlations from 0.82 and 0.81 to 0.85 and 0.88, respectively, with corresponding MSEs of 0.32 and 0.30 mg/m^3. However, Pearson's correlations for PMS1 and PMS2 showed no improvement, with Pearson's correlation of 0.95 and an MSE value of 0.20 mg/m^3. In addition to these correlations, the MSE values generated by this model were significantly higher than those generated by the RFR and ANN models.

A detailed analysis of the results revealed the drawbacks of linear regression models. First, even after tuning the model hyperparameters and applying the optimum parameters, the SVM model predicted a significant number of negative concentrations for most sensors. This is because the plane of best fit of the SVM model has a negative intercept on the PDM concentration. When this occurs, lower concentrations are predicted to be negative, which can be misleading. In our study, this phenomenon was primarily caused by the features of the raw dataset, which had negative y-intercept values. Second, an exponential relationship was observed over the desired linear relationship between the predicted sensor outputs and PDM concentrations for the Airtrek and GasLab sensors. These results suggest the inability of the SVM model to correct the portions of the sensor dataset that were oddly correlated, although better performances have been observed in a previous study.

5. Conclusions

Overexposure to respirable coal mine dust in underground coal mines has been linked to the recent resurgence of pneumoconiosis among coal workers. Effective real-time personal coal dust monitoring in underground coal mines provides timely detection of overexposure, enabling the implementation of control measures. However, the currently used PDM3700 personal coal dust monitor is very expensive for all miners, which hinders many miners from knowing their exposure levels. Low-cost PM sensors are used to measure personal exposure levels with high spatial and temporal resolutions owing to their low cost, small size, and light weight. However, these sensors require powerful calibration to ensure accuracy during long-term deployment. Although these sensors have been previously calibrated using SLR models, a more accurate calibration model is required for improved accuracy. Therefore, this study compared the performance of five machine-learning algorithms for calibrating low-cost PM sensors for coal dust monitoring.

The main conclusions of this study are as follows. Superiority of machine learning models (ANN and RFR) over traditional regression techniques for sensor calibration was demonstrated. Significant improvements in the correlation coefficients and accuracy were achieved using low-cost PM sensors calibrated using ANN and RFR. These calibrated sensors can be implemented for coal dust monitoring in mining environments. This streamlined approach will emphasize the contributions and practical implications of this study.

CRediT authorship contribution statement

Nana A. Amoah: Conceptualization, Methodology, Investigation, Writing – original draft. **Mirza Muhammad Zaid:** Investigation, Writing – review & editing. **Xiaosong Du:** Methodology, Validation. **Yang Wang:** Investigation, Validation, Resources. **Guang Xu:** Supervision, Resources, Project administration.

Declaration of Competing Interest

Guang Xu is an editorial board member for this journal and was not involved in the editorial review or the decision to publish this article. All other 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.

Acknowledgements

This research is partially funded by the National Science Foundation (No. 2034198), and the National Institute for Occupational Safety and Health (NIOSH) (No. 75D30123C17714). The findings and conclusions in this report are those of the authors and do not represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention; nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

References

[1] A.S. Laney, M.D. Attfield, Coal workers' pneumoconiosis and progressive massive fibrosis are increasingly more prevalent among workers in small underground coal mines in the United States, Occup. Environ. Med. 67 (6) (2010) 428–431.
[2] H.B. Liu, Z.F. Tang, Y.L. Yang, D. Weng, G. Sun, Z.W. Duan, J. Chen, Identification and classification of high risk groups for coal Workers' pneumoconiosis using an artificial neural network based on occupational histories: a retrospective cohort study, BMC Public Health 9 (2009) 366.
[3] R.A. Cohen, E.L. Petsonk, C. Rose, B. Young, M. Regier, A. Najmuddin, J.L. Abraham, A. Churg, F.H. Green, Lung pathology in U.S. Coal workers with rapidly progressive pneumoconiosis implicates silica and silicates, Am. J. Respir. Crit. Care Med. 193 (6) (2016) 673–680.
[4] D.J. Blackley, C.N. Halldin, A.S. Laney, Continued increase in prevalence of coal workers' pneumoconiosis in the United States, 1970–2017, Am. J. Public Health 108 (9) (2018) 1220–1222.
[5] B.C. Doney, D. Blackley, J.M. Hale, C. Halldin, L. Kurth, G. Syamlal, A.S. Laney, Respirable coal mine dust in underground mines, United States, 1982–2017, Am. J. Ind. Med. 62 (6) (2019) 478–485.
[6] M.M. Zaid, N. Amoah, A. Kakoria, Y. Wang, G. Xu, Advancing occupational health in mining: investigating low-cost sensors suitability for improved coal dust exposure monitoring, Meas. Sci. Technol. 35 (2) (2023) 025128.
[7] Mine Safety and Health Administration, Major provisions and effective dates MSHA's final rule to lower miners' exposure to respirable coal mine dust, MSHA, 2014.
[8] M.M. Zaid, G. Xu, N.A. Amoah, Accuracy of low-cost particulate matter sensor in measuring coal mine dust—A wind tunnel evaluation, in: Underground Ventilation, CRC Press, 2023, pp. 274–284.
[9] U.S. Environmental Protection Agency, Roadmap for Next-Generation Air Monitoring, EPA, 2013. 〈https://www.epa.gov/sites/default/files/2014-09/documents/roadmap-20130308.pdf〉 (Accessed September 11, 2023).
[10] R. Williams, A. Kaufman, T. Hanley, S. Rice, J. Garvey, EPA Sensor Evaluation Report, U.S. Environmental Protection Agency, 2014. 〈https://www.aqmd.gov/docs/default-source/aq-spec/resources-page/us-epa—sensor-evaluation-report.pdf〉 (Accessed September 11, 2023).
[11] M. Badura, P. Batog, A. Drzeniecka-Osiadacz, P. Modzel, Evaluation of low-cost sensors for ambient $PM_{2.5}$ monitoring, J. Sens. 2018 (1) (2018) 5096540.
[12] Z.Al Barakeh, P. Breuil, N. Redon, C. Pijolat, N. Locoge, J.P. Viricelle, Development of a normalized multi-sensors system for low cost on-line atmospheric pollution detection, Sens. Actuat. B Chem. 241 (2017) 1235–1243.
[13] J.C. Volkwein, R.P. Vinson, L.J. McWilliams, D.P. Tuchman, S.E. Mischler, Performance of a new personal respirable dust monitor for mine use, National Institute for Occupational Safety and Health, Cincinnati, OH, 2004.
[14] M.U. Khan, A.D.S. Gillies, Real-time monitoring of DPM, airborne dust and correlating elemental carbon measured by two methods in underground mines in USA, in: E. Sarver, S. Schafrik, E. Jong, K. Luxbacher (Eds.), 15th North American Mine Ventilation Symposium, 2015, pp. 1–7.
[15] N.A. Amoah, G. Xu, Y. Wang, J.Y. Li, Y.M. Zou, B.S. Nie, Application of low-cost particulate matter sensors for air quality monitoring and exposure assessment in underground mines: a review, Int. J. Miner. Metall. Mater. 29 (8) (2022) 1475–1490.
[16] 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.
[17] L. Spinelle, M. Aleixandre, M. Gerboles, Protocol of evaluation and calibration of Low-cost gas sensors for the monitoring of air pollution, Publications Office of the European Union, Luxembourg, 2013.
[18] D.M. Holstius, A. Pillarisetti, K.R. Smith, E. Seto, Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in california, Atmos. Meas. Tech. 7 (4) (2014) 1121–1131.
[19] S. Kelleher, C. Quinn, D. Miller-Lionberg, J. Volckens, A low-cost particulate matter ($PM_{2.5}$) monitor for wildland fire smoke, Atmos. Meas. Tech. 11 (2) (2018) 1087–1097.

[20] A. Polidori, V. Papapostolou, H. Zhang, Laboratory evaluation of low-cost air quality sensors: laboratory setup and testing protocol, South Coast Air Quality Management District, 2016. 〈https://www.aqmd.gov/docs/default-source/aq-spec/protocols/sensors-lab-testing-protocol6087afefc2b66f27bf6fff00004a91a9.pdf〉 (Accessed September 9, 2023).

[21] T. Sayahi, A. Butterfield, K.E. Kelly, Long-term field evaluation of the plantower PMS low-cost particulate matter sensors, Environ. Pollut. 245 (2019) 932–940.

[22] Y. Wang, J. Li, H. Jing, Q. Zhang, J. Jiang, P. Biswas, Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement, Aerosol Sci. Technol. 49 (11) (2015) 1063–1077.

[23] T.S. Zheng, M.H. Bergin, K.K. Johnson, S.N. Tripathi, S. Shirodkar, M.S. Landis, R. Sutaria, D.E. Carlson, Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments, Atmos. Meas. Tech. 11 (8) (2018) 4823–4846.

[24] B. Feenstra, V. Papapostolou, S. Hasheminassab, H. Zhang, B. Der Boghossian, D. Cocker, A. Polidori, Performance evaluation of twelve low-cost $PM_{2.5}$ sensors at an ambient air monitoring site, Atmos. Environ. 216 (2019) 116946.

[25] A. Polidori, V. Papapostolou, B. Feenstra, H. Zhang, Field evaluation of low-cost air quality sensors: field setup and testing protocol, South Coast AQMD, 2017. 〈http://www.aqmd.gov/aq-spec/evaluations/field〉 (Accessed September 9, 2023).

[26] M. Tagle, F. Rojas, F. Reyes, Y. Vásquez, F. Hallgren, J. Lindén, D. Kolev, Å.K. Watne, P. Oyola, Field performance of a low-cost sensor in the monitoring of particulate matter in Santiago, Chile, Environ. Monit. Assess. 192 (3) (2020) 171.

[27] A. Samad, D.R. Obando Nuñez, G.C. Solis Castillo, B. Laquai, U. Vogt, Effect of relative humidity and air temperature on the results obtained from low-cost gas sensors for ambient air quality measurements, Sensors 20 (18) (2020) 5175.

[28] F. Concas, J. Mineraud, E. Lagerspetz, S. Varjonen, X.L. Liu, K. Puolamäki, P. Nurmi, S. Tarkoma, Low-cost outdoor air quality monitoring and sensor calibration, ACM Trans. Sen. Netw. 17 (2) (2021) 1–44.

[29] A. Gonzalez, A. Boies, J. Swanson, D. Kittelson, Measuring the air quality using low-cost air sensors in a parking garage at University of Minnesota, USA, Int. J. Environ. Res. Public Health 19 (22) (2022) 15223.

[30] Z. Farooqui, J. Biswas, J. Saha, Long-term assessment of PurpleAir low-cost sensor for $PM_{2.5}$ in California, USA, Pollutants 3 (4) (2023) 477–493.

[31] M. Ghamari, C. Soltanpur, P. Rangel, W.A. Groves, V. Kecojevic, Laboratory and field evaluation of three low-cost particulate matter sensors, IET Wirel. Sens. Syst. 12 (1) (2022) 21–32.

[32] M. Ghamari, H. Kamangir, K. Arezoo, K. Alipour, Evaluation and calibration of low-cost off-the-shelf particulate matter sensors using machine learning techniques, IET Wirel. Sens. Syst. 12 (5–6) (2022) 134–148.

[33] N.A. Amoah, M.M. Zaid, A.R. Kumar, P. Chang, G. Xu, Optimized canopy air curtain dust protection using a two-level manifold and computational fluid dynamics, Min. Metall. Explor 41 (4) (2024) 1807–1818.

[34] S. Munir, M. Mayfield, D. Coca, S.A. Jubb, O. Osammor, Analysing the performance of low-cost air quality sensors, their drivers, relative benefits and calibration in cities—A case study in sheffield, Environ. Monit. Assess. 191 (2) (2019) 94.

[35] 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 (9) (2018) 2843.

[36] P. Nowack, L. Konstantinovskiy, H. Gardiner, J. Cant, Machine learning calibration of low-cost NO_2 and PM_{10} sensors: Non-linear algorithms and their impact on site transferability, Atmos. Meas. Tech. 14 (8) (2021) 5637–5655.

[37] C.C. Chen, C.T. Kuo, S.Y. Chen, C.H. Lin, J.J. Chue, Y.J. Hsieh, C.W. Cheng, C.M. Wu, C.M. Huang, Calibration of low-cost particle sensors by using machine-learning method, 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), IEEE, Chengdu, China, 2018, pp. 111–114.

[38] Y.W. Wang, Y.J. Du, J.N. Wang, T.T. Li, Calibration of a low-cost $PM_{2.5}$ monitor using a random forest model, Environ. Int. 133 (2019) 105161.

[39] L.O.H. Wijeratne, D.R. Kiv, A.R. Aker, S. Talebi, D.J. Lary, Using machine learning for the calibration of airborne particulate sensors, Sensors 20 (1) (2020) 99.

[40] N. Zimmerman, A.A. Presto, S.P.N. Kumar, J. Gu, A. Hauryliuk, E.S. Robinson, A.L. Robinson, R. Subramanian, A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, Atmos. Meas. Tech. 11 (1) (2018) 291–313.

[41] M.X. Si, Y. Xiong, S. Du, K. Du, Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods, Atmos. Meas. Tech. 13 (4) (2020) 1693–1707.

[42] W.V. Wang, S.C. Lung, C.H. Liu, Application of machine learning for the in-field correction of a $PM_{2.5}$ low-cost sensor network, Sensors 20 (17) (2020) 5002.

[43] J.C. Volkwein, R.P. Vinson, S.J. Page, L.J. McWilliams, G.J. Joy, S.E. Mischler, D.P. Tuchman, Laboratory and Field Performance of a Continuously Measuring Personal Respirable Dust Monitor, National Institute for Occupational Safety and Health, Cincinnati, OH, 2006.

[44] TSI Incorporated, 3321 aerodynamic particle sizer spectrometer—specification sheet, TSI, 2016. 〈https://www.kenelec.com.au/wp-content/uploads/2016/06/TSI-3321-Aerodynamic-Particle-Sizer-SpecSheet.pdf〉 (Accessed September 11, 2023).

[45] N.B. Hall, D.J. Blackley, C.N. Halldin, A.S. Laney, Current review of pneumoconiosis among US coal miners, Curr. Environ. Health Rep. 6 (3) (2019) 137–147.

[46] A.I. Batool, N.H. Naveed, M. Aslam, J. da Silva, M.F.U. Rehman, Coal dust-induced systematic hypoxia and redox imbalance among coal mine workers, ACS Omega 5 (43) (2020) 28204–28211.

[47] N.A. Amoah, G. Xu, A.R. Kumar, Y. Wang, Calibration of low-cost particulate matter sensors for coal dust monitoring, Sci. Total Environ. 859 (2023) 160336.

[48] L. Wang, F.F. Li, Y.Y. Guo, Q.Z. Li, T.M. Chen, J.J. Wu, Numerical simulation study of dust transport of comprehensive mining working surface, in: A.J. Tallón-Ballesteros (Ed.), Modern Management based on Big Data III, IOS Press, 2022, pp. 440–446.

[49] M.L. Zamora, F. Xiong, D. Gentner, B. Kerkez, J. Kohrman-Glaser, K. Koehler, Field and laboratory evaluations of the low-cost plantower particulate matter sensor, Environ. Sci. Technol. 53 (2) (2019) 838–849.