Safety and Health at Work 10 (2019) 452–460

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Original Article

Calibration of Portable Particulate Matter–Monitoring Device using Web Query and Machine Learning

Byoung Gook Loh [1], Gi Heung Choi [2,*]

[1] Department of Applied IT Engineering, Hansung University, Seoul, Republic of Korea
[2] Department of Mechanical Systems Engineering, Hansung University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 8 May 2019
Received in revised form
28 July 2019
Accepted 9 August 2019
Available online 19 August 2019

Keywords:
Calibration
Machine learning
Monitoring and control
Particulate matter
Web query

ABSTRACT

Background: Monitoring and control of $PM_{2.5}$ are being recognized as key to address health issues attributed to $PM_{2.5}$. Availability of low-cost $PM_{2.5}$ sensors made it possible to introduce a number of portable $PM_{2.5}$ monitors based on light scattering to the consumer market at an affordable price. Accuracy of light scattering–based $PM_{2.5}$ monitors significantly depends on the method of calibration. Static calibration curve is used as the most popular calibration method for low-cost $PM_{2.5}$ sensors particularly because of ease of application. Drawback in this approach is, however, the lack of accuracy.
Methods: This study discussed the calibration of a low-cost $PM_{2.5}$-monitoring device (PMD) to improve the accuracy and reliability for practical use. The proposed method is based on construction of the $PM_{2.5}$ sensor network using Message Queuing Telemetry Transport (MQTT) protocol and web query of reference measurement data available at government-authorized PM monitoring station (GAMS) in the republic of Korea. Four machine learning (ML) algorithms such as support vector machine, k-nearest neighbors, random forest, and extreme gradient boosting were used as regression models to calibrate the PMD measurements of $PM_{2.5}$. Performance of each ML algorithm was evaluated using stratified K-fold cross-validation, and a linear regression model was used as a reference.
Results: Based on the performance of ML algorithms used, regression of the output of the PMD to $PM_{2.5}$ concentrations data available from the GAMS through web query was effective. The extreme gradient boosting algorithm showed the best performance with a mean coefficient of determination (R^2) of 0.78 and standard error of 5.0 $\mu g/m^3$, corresponding to 8% increase in R^2 and 12% decrease in root mean square error in comparison with the linear regression model. Minimum 100 hours of calibration period was found required to calibrate the PMD to its full capacity. Calibration method proposed poses a limitation on the location of the PMD being in the vicinity of the GAMS. As the number of the PMD participating in the sensor network increases, however, calibrated PMDs can be used as reference devices to nearby PMDs that require calibration, forming a calibration chain through MQTT protocol.
Conclusions: Calibration of a low-cost PMD, which is based on construction of $PM_{2.5}$ sensor network using MQTT protocol and web query of reference measurement data available at a GAMS, significantly improves the accuracy and reliability of a PMD, thereby making practical use of the low-cost PMD possible.

1. Introduction

There has been much research activity concerning potential adverse health risks and mortality rates associated with environmental and occupational exposures to particulate matter (PM) with a diameter of 2.5 microns or less ($PM_{2.5}$) [1–5]. To improve public health by securing protection from $PM_{2.5}$ exposure, the US Environmental Protection Agency [6], World Health Organization [7], and European Union [8] set guidelines for daily and annual exposure standard of $PM_{2.5}$. A first step to investigate $PM_{2.5}$ exposure would be to measure either indoor or outdoor environmental $PM_{2.5}$ concentration through a network of the $PM_{2.5}$ monitors.

In practice, $PM_{2.5}$ can be monitored at fixed locations such as government-authorized PM monitoring stations (GAMSs) or at

* Corresponding author. Department of Mechanical Systems Engineering, Hansung University, 116 Samseongyoro-16gil Sungbuk-gu, Seoul, 136-792, Republic of Korea.
E-mail address: gihchoi@hansung.ac.kr (G.H. Choi).

2093-7911/$ – see front matter
https://doi.org/10.1016/j.shaw.2019.08.002

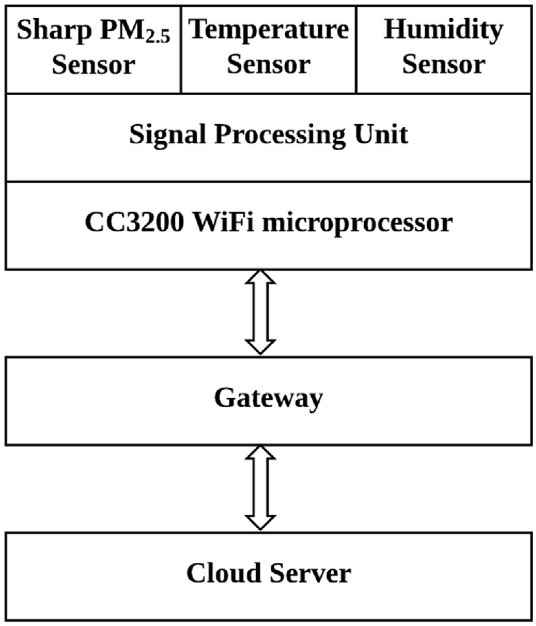

Fig. 1. Functional block diagram of PMD proposed in [11]. PMD, $PM_{2.5}$-monitoring device.

multiple locations using portable light scattering–based PM-monitoring devices [9]. Equipment used at the GAMS provides accurate measurements. These devices are, however, large and prohibitively expensive (50K–100K USD) to construct a grid of sensing network. Sparse $PM_{2.5}$ monitoring data make it difficult to have a full grasp of ambient $PM_{2.5}$ concentration resulting from urban emissions, atmospheric transformations, and transport mechanism [10].

Small inexpensive $PM_{2.5}$-monitoring devices (PMDs), relying on newly available off-the-shelf sensors are cheap and portable and thus suitable for deployment at large scale. Loh and Choi [11] developed a portable PMD that incorporates the concept of internet of things (IoT). Yang Q et al. [12] and Yang X [13] proposed an open platform for PMD with Wi-Fi and Message Queuing Telemetry Transport (MQTT) protocol for interdevice communication. Zhuang et al. [14] designed and implemented a context-sensing portable device for personal air quality monitoring. Saad et al. [15] used a server to collect sensor data from a PMD and make collected data available through Web.

The low-cost PMDs frequently use a static calibration curve or equation because of its ease of use. However, device–device variations and lack of standard calibration method typically lead to substantial measurement errors. If a consistent and practical method to calibrate the low-cost PMD can be implemented, wide use of it would be, therefore, possible. Holstius et al. [10] performed field calibration of a PMD at a regulatory monitoring site in the USA and found a good correlation between the measurement of the PMD and that of the official stationary $PM_{2.5}$ monitor. Cheng et al. developed a cloud-based air quality monitoring system using cloud air quality analytic engine for calibration [16]. Manikonda et al. [4] compared well-characterized reference instruments with low-cost PMDs and concluded that if properly calibrated and deployed at scale, the PMDs could be applied for high-spatial temporal monitoring of $PM_{2.5}$ over extended time intervals and offer useful exposure estimate for health effect research. Sousan et al. [17] also evaluated low-cost PMDs and demonstrated that once calibrated, low-cost $PM_{2.5}$ sensors can be used in the occupational settings.

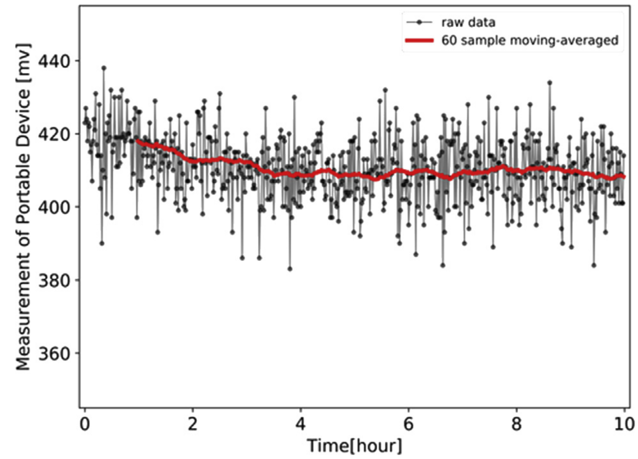

Fig. 2. Raw measurement of PMD. PMD, $PM_{2.5}$-monitoring device.

In this study, a method to calibrate a low-cost PMD using hourly measurements at a GAMS in the Republic of Korea available in the form of RESTful web service was proposed. The calibration method is based on machine learning (ML). Four ML algorithms were tested for Web-based calibration of the PMD, and individual performance of each ML was evaluated in comparison with the linear regression model.

2. Materials and methods

2.1. Construction and working principle of low-cost PMD

The PMD developed by Loh and Choi [11] was used for investigating the calibration analysis. It is comprised of the $PM_{2.5}$ sensor module (DN7C3CA006 [18]; Sharp, Japan), a signal processing unit, temperature and humidity sensors, and a Wi-Fi microprocessor (CC3200 LaunchPad, Texas Instrument, USA). The PMD measures 80 mm in width, 80 mm in height, and 200 mm in length, and the measurements of the PMD can be sent to a cloud server through a gateway for data collection and analysis as illustrated in Fig. 1.

To measure $PM_{2.5}$ concentration, air is drawn into an air conduit by a motorized fan in the sensor module. With input voltage of the sensor module set to 5 V, the motorized fan rotates at a fixed rotational speed. Flow rate is specific to the sensor module, and no change in the flow rate is allowed. In the particle separation chamber called a virtual impactor, only particles smaller than 2.5 μm are introduced from the air into the light scattering unit, while particles greater than $PM_{2.5}$ pass through the main air flow cavity. As $PM_{2.5}$ in the air enters the light scattering unit consisting of an infrared light source and a photodiode, it scatters infrared light rays. Intensity of scattered light is measured by the photodiode placed at an angle of 120 degrees from the light source, resulting in voltage proportional to $PM_{2.5}$ concentration via nephelometric detection [19].

2.2. Preprocessing of PMD-measured signal

The measured voltage from the sensor module was sampled at a frequency of 100 Hz and then moving-averaged with a 60-sample moving window to remove high-frequency noises as shown in Fig. 2. The preprocessed measurement requires calibration with respect to the reference measurements to convert the measurement of the PMD in voltage into μg/m^3. A flowchart of preprocessing of the measured signal and calibration of the PMD is shown in Fig. 3. The PMD is equipped with Wi-Fi network capability

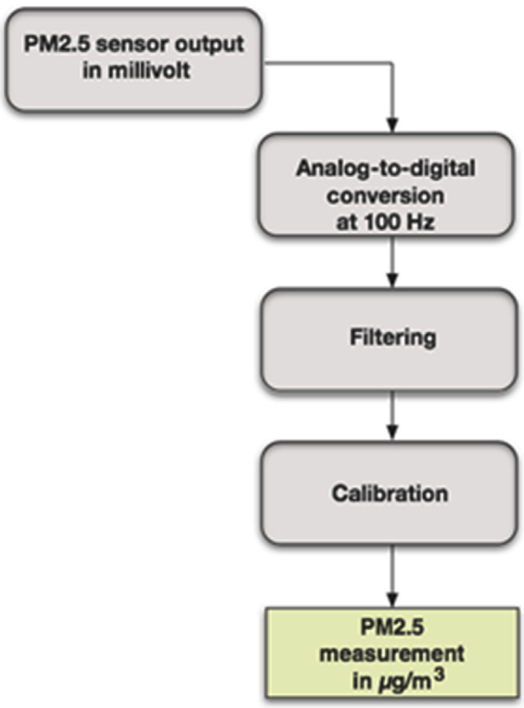

Fig. 3. Flowchart of $PM_{2.5}$ signal processing and calibration.


```xml
<so2Value>0.004</so2Value>
<coValue>0.5</coValue>
<o3Value>0.024</o3Value>
<no2Value>0.023</no2Value>
<pm10Value>40</pm10Value>
<pm10Value24>44</pm10Value24>
<pm25Value>19</pm25Value>
<pm25Value24>21</pm25Value24>
<khaiValue>64</khaiValue>
<khaiGrade>2</khaiGrade>
<so2Grade>1</so2Grade>
<coGrade>1</coGrade>
<o3Grade>1</o3Grade>
<no2Grade>1</no2Grade>
<pm10Grade>2</pm10Grade>
<pm25Grade>2</pm25Grade>
<pm10Grade1h>2</pm10Grade1h>
<pm25Grade1h>2</pm25Grade1h>
```


Fig. 5. An example of $PM_{2.5}$ XML data measured at a GAMS and obtained from the open data portal using HTTP protocol. GAMS, government-authorized PM monitoring station; HTTP, Hypertext Transfer Protocol; PM, particulate matter; XML, Hypertext Markup Language.

by which the preprocessed voltage is transmitted to the cloud server via the local Wi-Fi network at every minute.

2.3. *Framework of PMD calibration with web query*

The hourly $PM_{2.5}$ measurements are available from the open data portal www.data.go.kr which is connected to the GAMSs across Korea in a form of RESTful web service as shown in Fig. 4. Using Hypertext Transfer Protocol (HTTP) Get protocol allows for obtaining the $PM_{2.5}$ measurement along with other atmospheric environment measurements in Hypertext Markup Language (XML) format as shown in Fig. 5. The measured $PM_{2.5}$ can be retrieved using XML parser (see xml tag <pm25Value> in Fig. 5) and used as reference data to calibrate the PMD.

2.4. *Construction of PMD sensor network using MQTT*

There exist only 191 GAMSs in the Republic of Korea with the capability of measuring $PM_{2.5}$. The number is far smaller than that required to provide sufficient amount of measured data across the country. Properly calibrated PMDs can be used to form a $PM_{2.5}$ sensing network in conjunction with the GAMS using the Message Queuing Telemetry Transport (MQTT) protocol [16]. The MQTT is publish–subscribe–based message protocol that works on transmission control protocol/Internet protocol (TCP/IP), suitable for connections with remote locations and limited bandwidth. To create a sensor network, a MQTT $PM_{2.5}$ server retrieves hourly $PM_{2.5}$ data via RESTful HTTP from the open data portal and periodically publishes fetched data to the MQTT broker through which the PMD receives $PM_{2.5}$ data using the MQTT subscribe protocol as shown in Fig. 6. It is noted that under the MQTT model, the PMDs are not allowed to directly communicate with each other and inter-PMDs communication is possible solely through the MQTT broker.

The publisher in Fig. 6 publishes data to the broker under the self-descriptive topics, and the subscriber can subscribe to the topics of interest to receive data when new data for the subscribed topic are available. MQTT topics are hierarchically constructed in a similar way to a file system with the forward slash (/) as a delimiter. In building topic hierarchy, a preferred approach would be the semantic method in which things are named for where they are first and what they measure subsequently [20]. Following the semantic topic hierarchy, the PMD can publish its $PM_{2.5}$ measurement under the topic “Country code/Postal code/Latitude Longitude/PM2.5.” A publish topic example of the PMD located at Hansung University in Seoul, Korea, is illustrated in Fig. 7. The first topic in the hierarchical topic tree is the country code, and the second is the postal code of the PMD. Because the postal codes are usually assigned to geographical areas, they can be conveniently used for identifying the areas in the country. The third topic represents the latitude and longitude of the PMD location which not only provides the exact location but also serves as the unique device identifier. The last topic is the type of measurement, $PM_{2.5}$ in $\mu g/m^3$. Any PMDs

Fig. 4. $PM_{2.5}$ data fetch using HTTP protocol from the open data portal www.data.go.kr connected to GAMSs in Korea. GAMS, government-authorized PM monitoring station; HTTP, Hypertext Transfer Protocol; PM, particulate matter.

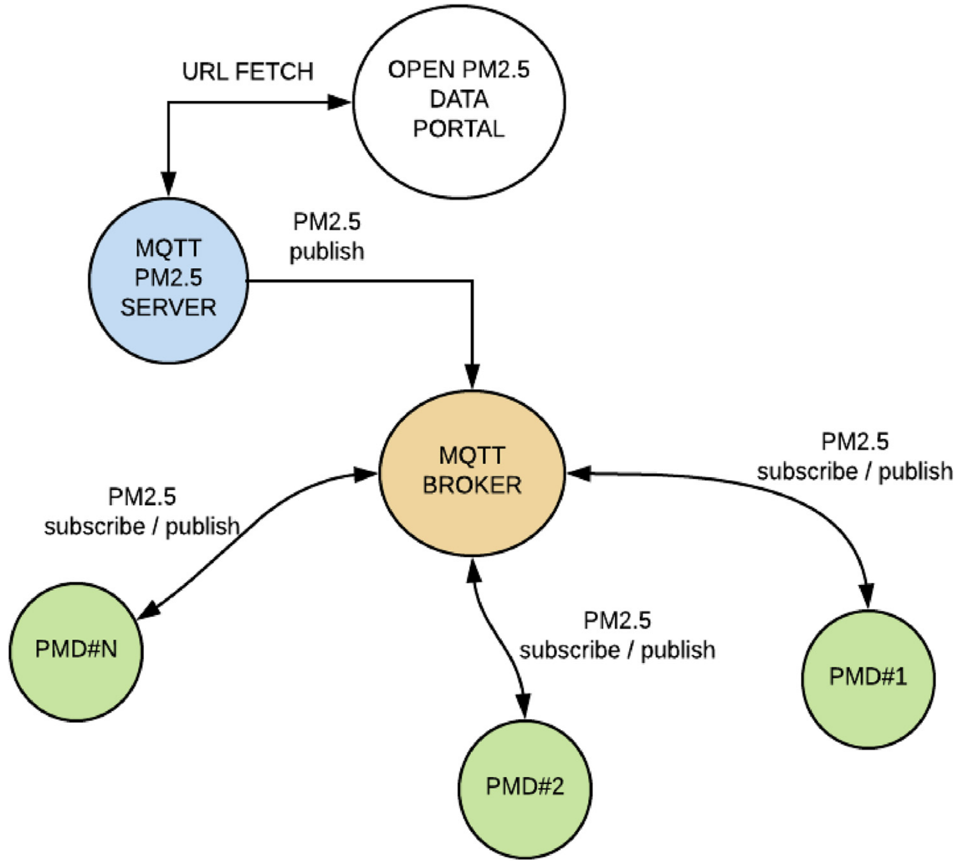

Fig. 6. Illustration of $PM_{2.5}$ data sharing among PMDs using MQTT protocol. MQTT, Message Queuing Telemetry Transport; PMD, $PM_{2.5}$-monitoring device.

subscribing the topic receive the $PM_{2.5}$ data when new data are published under the topic.

Simultaneous subscription to multiple topics is possible with wild card #, enabling to receive $PM_{2.5}$ measurements from the multiple PMDs. For example, subscribing to the topic “Korea/02786/#” allows $PM_{2.5}$ data to be received from all PMDs located in the area with Korean postal code of 02786, as shown in Fig. 8. Data aggregation can be conveniently achieved for analysis in the IoT clouds, personal computers, and mobile phones depending on the needs. To validate the concept of calibration based on the sensor network using MQTT, two PMDs were built. The MQTT $PM_{2.5}$ server was implemented using Amazon Web Service [21], and Eclipse Mosquitto open-source MQTT broker [22] was used. It was found that MQTT message transmission latency was less than 500 ms, and the PM2.5 server latency based on HTTP protocol was on the order of seconds. As the $PM_{2.5}$ data from the GAMS are updated every hour, the effect of transmission latency on the performance of calibration using the sensor network is negligible. Calibration was performed off-line for the study because of low computing capability of the CC3200 processor of the PMD. However, real-time built-in calibration can be readily implemented in the gateway.

2.5. *Characteristics of raw data and linear regression*

Fig. 9 shows the measurements of the temperature, humidity, and volumetric $PM_{2.5}$ from the PMD along with the measurements of

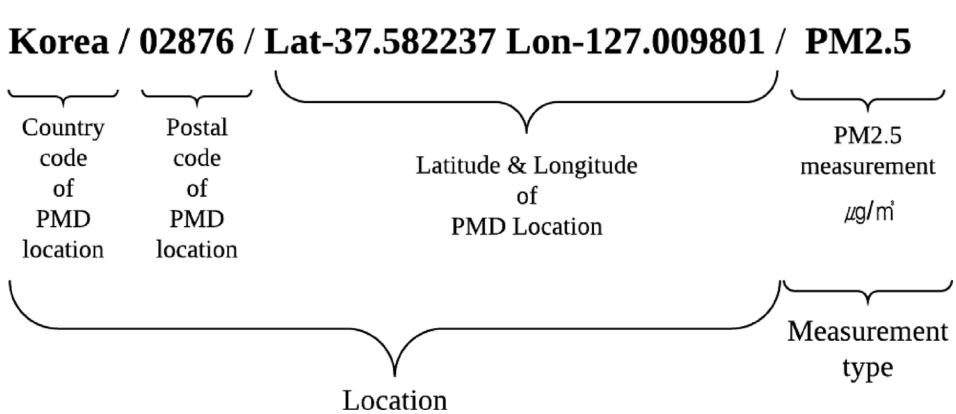

Fig. 7. MQTT hierarchical topic tree example for publishing PM2.5 data from PMD. MQTT, Message Queuing Telemetry Transport; PMD, $PM_{2.5}$-monitoring device.

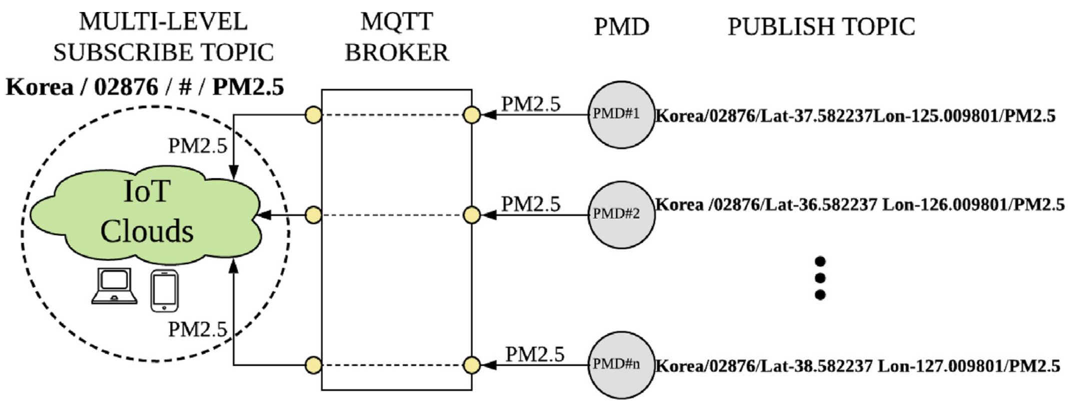

Fig. 8. Multiple subscriptions to MQTT topics to retrieve $PM_{2.5}$ data from multiple PMDs. IoT, Internet of things; MQTT, Message Queuing Telemetry Transport; PMD, $PM_{2.5}$-monitoring device.

$PM_{2.5}$ from the GAMS. Unit of $PM_{2.5}$ is millivolt in PMD measurements and μg/m^3 in GAMS measurements. Temperature and humidity measurements of the PMD were calibrated based on the sensor manufacturer's conversion formula specified in the sensor's datasheet [23]. The temperature and humidity measurements shown in Fig. 9(B) and (C) serve the purpose of showing correlation between the PM2.5 sensor measurement of the PMD and the temperature and humidity measurements. The measurements were performed for 319 hours consecutively from the single PMD, resulting in 319 data samples for regression analysis. Temperature was measured in degree Celsius, and humidity, in %. The geographical distance between the locations of the PMD and GAMS is 3.2 km, as shown in Fig. 10, and no significant air pollution sources exist between them.

A strong linear correlation between the PMD and GAMS measurements of $PM_{2.5}$ is observed in Fig. 9(A). Temperature and relative humidity also show moderate correlation with the PMD measurements, as shown in Fig. 9(B) and (C), respectively. $PM_{2.5}$

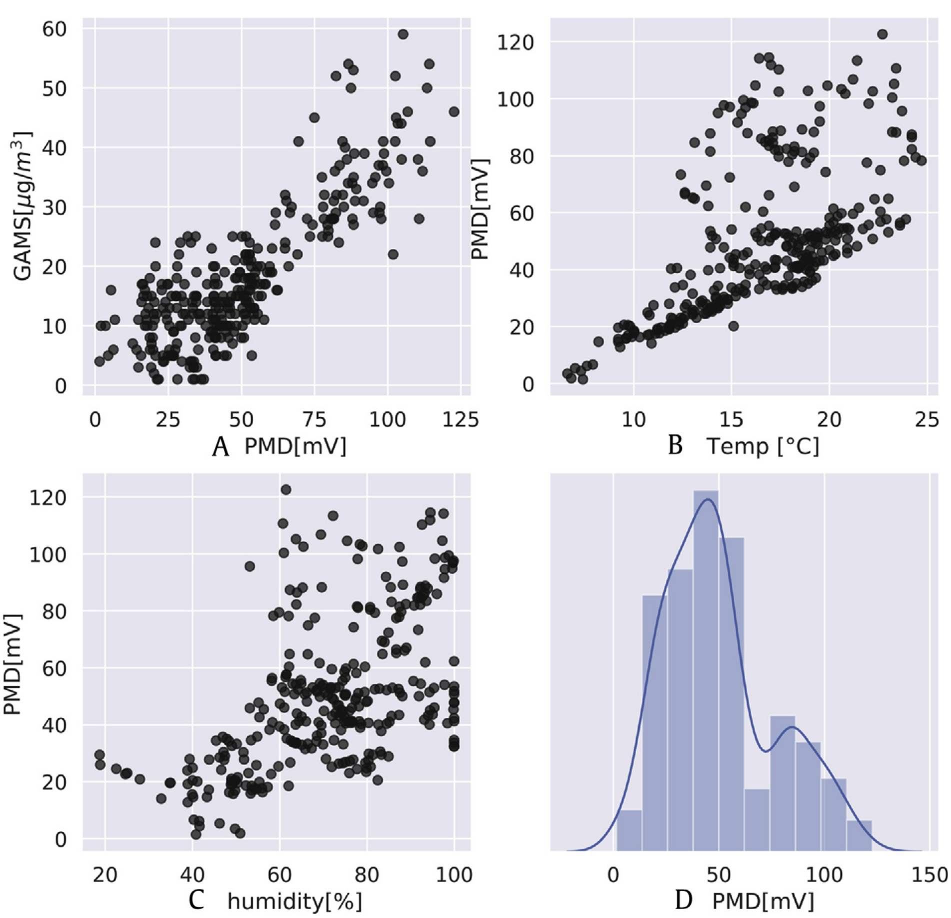

Fig. 9. Characteristics of measurements from PMD and GAMS. (A) Comparison of PMD measurements of $PM_{2.5}$ in millivolt with GAMS measurements of $PM_{2.5}$ in μg/m^3. (B) Variation of PMD measurements of $PM_{2.5}$ in millivolt with temperature in degree Celsius. (C) Variation of PMD measurements of $PM_{2.5}$ with humidity in %. (D) Histogram of PMD measurements of $PM_{2.5}$. GAMS, government-authorized PM monitoring station; PMD, $PM_{2.5}$-monitoring device.

Fig. 10. Locations of PMD and GAMS in Seoul, Korea, for data measurements. GAMS, government-authorized PM monitoring station; PMD, $PM_{2.5}$-monitoring device.

concentrations appear to increase as both temperature and relative humidity increase. More specifically, at low temperature and humidity, a firm linear relationship is seen in Fig. 9(B) and (C), while more pronounced scatter is observed at an elevated temperature and humidity. In general, humidity of the air can affect the performance of optical PM sensors in several ways: failure of the electronic circuits/sensors, overestimated outputs in sensor signals due to hygroscopic growth of some particles, detection of water droplets in fog or mist as particles [24–27], and deterioration of sensor performance due to absorption of infrared radiation by water [28]. Histogram of PMD $PM_{2.5}$ in Fig. 9(D) shows the normal distribution with a mean value of PMD $PM_{2.5}$ at 50 mV.

To find a correlation between the measurements of the PMD and GAMS measurements, the linear regression method was used as a reference model and subsequently compared with the other regression models based on four widely used ML algorithms such as support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), and extreme gradient boosting (XGB).

3. Results and discussion

3.1. Performance of linear regression as the reference model

The PMD measures $PM_{2.5}$ volumetric concentration which is affected by temperature and humidity. In constructing a linear regression model, two cases were considered with a regression sample size of 319. For Case I, only the PMD measurement is taken into account as the predictor, whereas for Case II, the temperature and humidity along with the PMD measurement were used as the predictors of the linear regression model. Increase in the number of the predictor always leads to increase in R^2. Therefore, adjusted R^2, which is a modified version of R^2 that has been adjusted for the number of predictors in the model, was instead used to investigate the effect of the number of the predictor on regression performance. As shown in Table 1, inclusion of temperature and humidity measurements increases the adjusted R^2 by 10%. Fig. 11 compares the predictions of $PM_{2.5}$ concentration based on the linear regression model for Cases I and II with the observations made at the GAMS.

Better evaluation of a linear regression model is achieved when the linear regression model is verified against data sets unseen to the model. Therefore, the data set was split into the training and validation sets with a split ratio of 80% to 20%, respectively. Linear regression was performed on the training set (255 samples, 80% of 319 data samples) and evaluated against the unseen validation set (64 samples, 20% of 319 data samples). The value of R^2 obtained from the validation set was 0.75, with root mean square error of 5.6

Table 1
R^2 and adjusted R^2 of Cases I and II for linear regression analysis

Predictors	R^2	Adjusted R^2
PMD measurement (Case I)	0.682	0.681
PMD measurement, temperature, humidity (Case II)	0.769	0.767

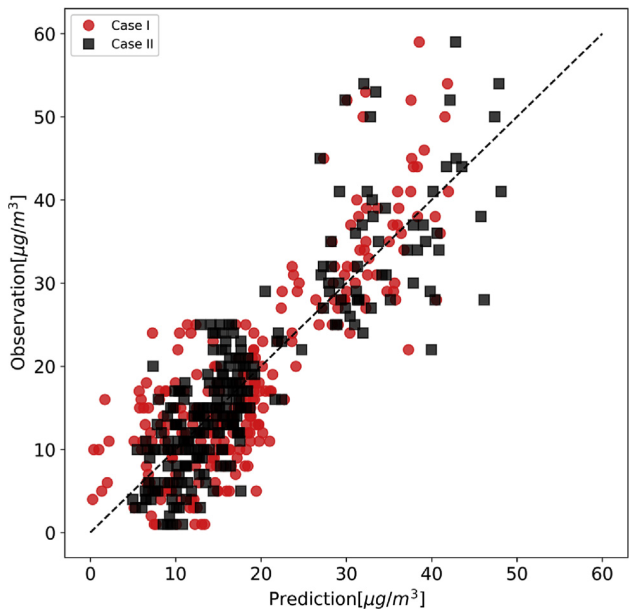

Fig. 11. Predictions of $PM_{2.5}$ concentration by linear regression for Cases I and II. The values of R^2 of Cases I and II are 0.68 and 0.77, respectively.

μg/m^3. To graphically indicate the degree to which the predictions agree with observation, the predictions from the verification set are arranged in an ascending order and shown along with observation in Fig. 12.

3.2. Regression models using ML algorithms with stratified K-fold cross-validation

With the linear regression model as the reference model for calibration of the PMD, four ML algorithms such as the SVM, KNN, RF, and XGB were studied as the calibration models. The input parameters of the ML-based calibration model were temperature, humidity, and $PM_{2.5}$ sensor measurements. To gauge the performance of each ML algorithm, stratified K-fold cross-validation is used [29]. In stratified K-fold cross-validation, the given sample is partitioned into K equal size subsamples in a way that each subsample can be considered as a good representative of the whole data by maintaining the mean response value of all folds

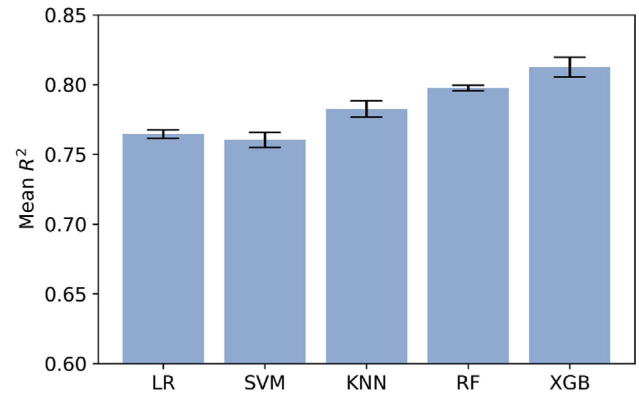

Fig. 13. Mean R^2 of linear regression and four different machine learning algorithms. Machine learning configurations are as follows: number of neighbors = 10 for KNN, linear kernel and C = 1 for SVM, minimum_samples_leaf = 2 for RF. Error bar represents the standard error. KNN, k-nearest neighbors; LR, linear regression; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting.

approximately equal. The K-1 subsamples out of the K subsamples are used as training data for regression, and the remaining subsample is retained as validation data for regression model trained with the training data. This process continues K times until each of K subsamples is used exactly once as the test data for the regression model. Once K results from the processes are obtained, they are averaged to create a single metric to compare effectiveness of each ML. In this study, data are evenly split into 5 subsamples and the number of the total data used for analysis was 319.

The mean coefficient of determination, R^2, and mean root mean squared error (RMSE) were evaluated using fivefold cross-validation. Among the four ML algorithms investigated, XGB showed the best result with a mean R^2 of 0.78 and standard error of 5.0 μg/m^3, which is 8% increase in R^2 and 12% decrease in RMSE in comparison with the linear regression model, as shown in Fig. 13, Fig. 14, and Fig. 15.

3.3. Calibration of PMD with web query

The number of samples included in regression analysis has a significant effect on the performance of regression. To investigate the minimum number of sample data required for calibrating the PMD using web query, the effect of the number of sample data on the coefficient of determination was examined as the number of

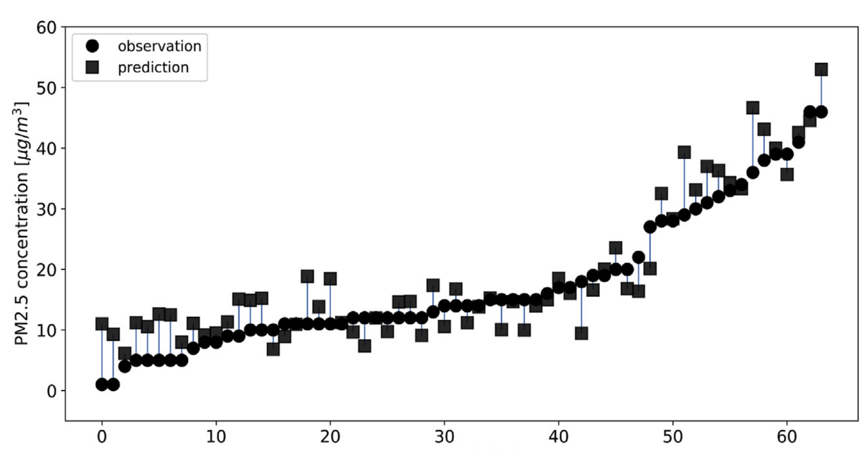

Fig. 12. Comparison of prediction of PMD from 20% validation set with observations at the GAMS. Vertical line represents difference between prediction and observation. GAMS, government-authorized PM monitoring station; PMD, $PM_{2.5}$-monitoring device.

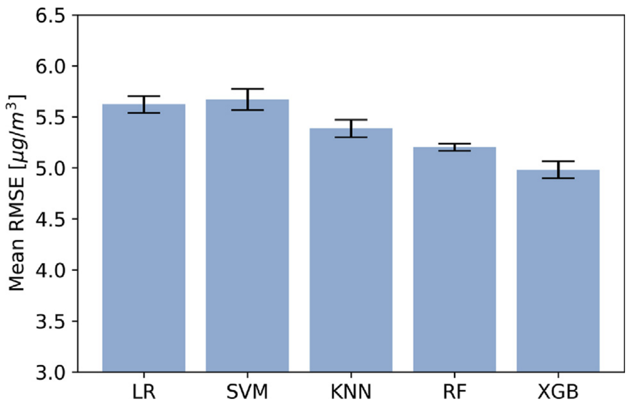

Fig. 14. Mean RMSE of linear regression and four different machine learning algorithms. Machine learning configurations are the same as in Fig. 13. Error bar represents standard error. KNN, k-nearest neighbors; LR, linear regression; RF, random forest; RMSE, root mean squared error; SVM, support vector machine; XGB, extreme gradient boosting.

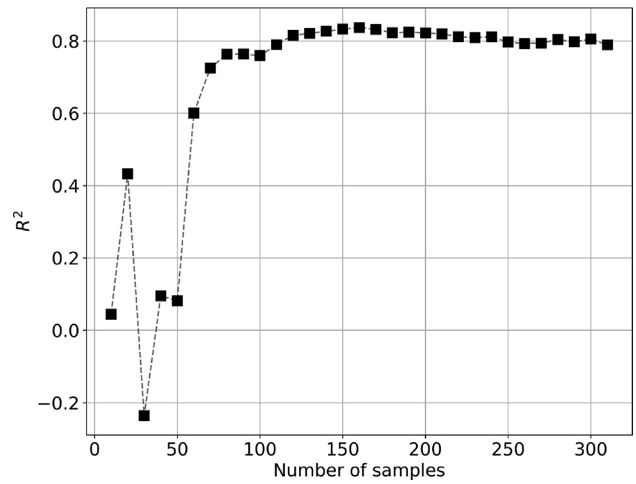

Fig. 16. Coefficient of determination (R^2) versus the number of regression data samples.

sample data increased from 10 to 310 with an increment of 10, as shown in Fig. 16. The coefficient of determination, R^2, oscillates between 0.1 and 0.6 as the number of the data included in regression approaches 60, beyond which the R^2 progressively converges to 0.78 with an increase in the number of samples and shows little variation as the number of the data grows greater than 100. Since the coefficient of determination of 0.78 indicates a strong correlation, the minimum number of samples of 100 can be safely used for calibration of the PMD considered in this study [11]. For best calibration of the PMD with web query, initial 100 data samples are required, which corresponds to 100 hours of lead time before the PMD is calibrated to its full capacity. For other types of PMDs, however, the minimum number of data samples required for calibration can vary and, therefore, should be identified by observing a change in R^2 with the number sample data in regression.

The proposed calibration method has a limitation in that the location of the PMD needs to be in the vicinity of the GAMS. With only little over one hundred GAMSs in the Republic of Korea, it is reasonable to assume that PMD readings at a particular location are most correlated with readings in the GAMS at the nearest location. To determine the range that a PMD needs to be in relation to the GAMS, more substantial experiments with a large number of the participating PMDs at different locations are required. In this study, however, the primary focus was on establishing a calibration method based on web query and ML and then evaluating the proposed method. As the number of the PMD participating in the sensor network increases, however, calibrated PMDs can be used as a reference device to nearby PMDs that require calibration, forming a calibration chain through the MQTT protocol. This will significantly improve the accuracy of PMDs in the network, thereby making practical use of these low-cost devices possible.

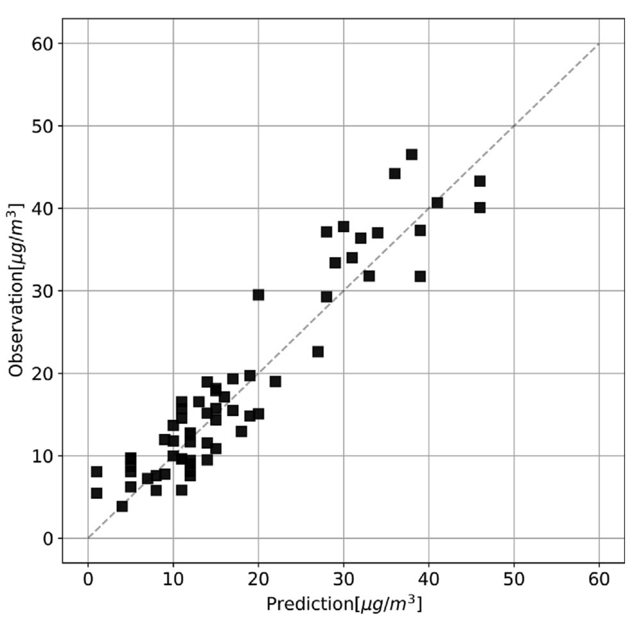

Fig. 15. Comparison of predictions made by XGB algorithm with observations at the GAMS. R^2 is 0.78, and RMSE is 5.0 μg/m^3. Predictions were made with a validation set with the size of 64. GAMS, government-authorized PM monitoring station; RMSE, root mean squared error; XGB, extreme gradient boosting.

4. Conclusions

Calibration of a low-cost PMD was studied to improve the accuracy and reliability of the PMD for practical use. The proposed method is based on construction of $PM_{2.5}$ sensor network using MQTT protocol and web query of reference measurement data available at the GAMS in the Republic of Korea. Four ML algorithms such as SVM, KNN, RF, and XGB were used as regression models to calibrate the PMD measurements of $PM_{2.5}$. Measurements made at the GAMS provide accurate data and were used as reference data for calibration. With a linear regression model as a reference, stratified K-fold cross-validation method was used to evaluate the performance of these ML algorithms.

Based on the performance of ML algorithms used, regression of the output of the PMD to $PM_{2.5}$ concentration data available from the GAMS through web query appears to be effective, resulting in the best performance of R^2 of 0.78 and standard error of 5.0 μg/m^3 with the XGB algorithm. These numbers correspond to 8% increase in R^2 and 12% decrease in RMSE in comparison with the linear regression model. The result indicates that after calibration, the PMD can be used to measure $PM_{2.5}$ with reasonable accuracy in environmental settings as either an individual device or a participatory device in the sensor network using the MQTT protocol. A minimum calibration period of 100 hours was required to calibrate the PMD to its full capacity. The calibration method proposed in this study has a limitation in that the location of the PMD needs to be in the vicinity of the GAMS. As the number of the PMD participating in the sensor network increases, however, calibrated PMDs can serve as a reference to adjacent PMDs needing calibration, forming a calibration chain through the MQTT protocol. This will significantly enhance the accuracy of a PMD in the network, enabling practical use of the low-cost PMD.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

This research was financially supported by Hansung University, Republic of Korea.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.shaw.2019.08.002.

References

[1] Boldo E, Medina S, Le Tertre A, Hurley F, Mücke H-G, Ballester F, et al. Apheis: health impact assessment of long-term exposure to PM 2.5 in 23 European cities. Eur J Epidemiol 2006;21(6):449–58.
[2] Bonn G. Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. World Health Organization; 2003. p. 7–9.
[3] Grant WB. Air pollution in relation to US cancer mortality rates: an ecological study; likely role of carbonaceous aerosols and polycyclic aromatic hydrocarbons. Anticanc Res 2009;29(9):3537–45.
[4] Manikonda A, Zíková N, Hopke PK, Ferro AR. Laboratory assessment of low-cost PM monitors. J Aerosol Sci 2016;102:29–40.
[5] Sørensen M, Daneshvar B, Hansen M, Dragsted LO, Hertel O, Knudsen L, et al. Personal PM2. 5 exposure and markers of oxidative stress in blood. Environ Health Perspect 2003;111(2):161–6.
[6] Agency USEP. NAAQS Table; 2019. Available from: https://www.epa.gov/criteria-air-pollutants/naaqs-table.html.
[7] Organization WH. Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. World Health Organization; 2006.
[8] Agency EE. Air quality in Europe — 2018 report. EEA Rep 2018;(12):60.
[9] Proceedings of the 12th ACM conference on embedded network sensor systems. In: Cheng Y, Li X, Li Z, Jiang S, Li Y, Jia J, et al., editors. AirCloud: a cloud-based air-quality monitoring system for everyone. ACM; 2014.
[10] Holstius DM, Pillarisetti A, Smith K, Seto E. Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. Atmos Meas Tech 2014;7(4):1121–31.
[11] Loh BG, Choi GH. Development of IoT-based PM2. 5 measuring device. J Korean Soc Saf 2017;32(1):21–6.
[12] 2015 IEEE international wireless symposium (IWS 2015). In: Yang Q, Zhou G, Qin W, Zhang B, Chiang PY, editors. Air-kare: a Wi-Fi based, multi-sensor, real-time indoor air quality monitor. IEEE; 2015.
[13] 2017 13th IEEE International Conference on Control & Automation (ICCA). In: Yang X, Yang L, Zhang J, editors. A WiFi-enabled indoor air quality monitoring and control system: the design and control experiments. IEEE; 2017.
[14] Proceedings of the 2015 workshop on pervasive wireless healthcare. In: Zhuang Y, Lin F, Yoo E-H, Xu W, editors. Airsense: a portable context-sensing device for personal air quality monitoring. ACM; 2015.
[15] International conference on electrical, electronics and system engineering (ICEESE); 2013. In: Saad SM, Saad ARM, Kamarudin AMY, Zakaria A, Shakaff AYM, editors. Indoor air quality monitoring system using wireless sensor network (WSN) with web interface. IEEE; 2013.
[16] Chen L-J, Ho Y-H, Lee H-C, Wu H-C, Liu H-M, Hsieh H-H, et al. An open framework for participatory PM2. 5 monitoring in smart cities. IEEE Access 2017;5:14441–54.
[17] Sousan S, Koehler K, Thomas G, Park JH, Hillman M, Halterman A, et al. Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols. Aerosol Sci Technol 2016;50(5):462–73.
[18] Sharp. Device specification for PM2.5 sensor module model No. DN7C3CA006. Electronic components and devices division; 2014.
[19] Cao TT, Jonathan E. Portable, ambient PM2.5 sensor for human and/or animal exposure studies. Anal Lett 2017;50(4):712–23.
[20] Perez X. MQTT topic naming convention; 2017. Available from: https://tinkerman.cat/mqtt-topic-naming-convention.
[21] AmazonWebServices. Amazon EC2 - secure and resizable compute capacity in the cloud; 2019. Available from: https://aws.amazon.com/ec2/.
[22] Eclipse-Foundation. IoT standards; 2019. Available from: https://iot.eclipse.org/standards/.
[23] Texas-Instrument. HDC1000 low power, high accuracy digital humidity sensor with temperature sensor; 2016. Available from: http://www.ti.com/lit/ds/symlink/hdc1000.pdf.
[24] Allen G, Sioutas C, Koutrakis P, Reiss R, Lurmann FW, Roberts PT. Evaluation of the TEOM® method for measurement of ambient particulate mass in urban areas. J Air Waste Manag Assoc 1997;47(6):682–9.
[25] Badura M, Batog P, Drzeniecka-Osiadacz A, Modzel P. Evaluation of low-cost sensors for ambient PM2. 5 monitoring. J Sens 2018;2018.
[26] Crilley LR, Shaw M, Pound R, Kramer LJ, Price R, Young S, et al. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. Atmos Meas Tech 2018:709–20.
[27] Jayaratne R, Liu X, Thai P, Dunbabin M, Morawska L. The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. Atmos Meas Tech 2018;11(8):4883–90.
[28] Wanjura JD, Shaw BW, Parnell C, Lacey RE, Capareda SC. Comparison of continuous monitor (TEOM) and gravimetric sampler particulate matter concentrations. Trans ASABE 2008;51(1):251–7.
[29] Scikit-learn. Cross-validation: evaluating estimator performance; 2018. Available from: https://scikit-learn.org/stable/modules/cross_validation.html.