Atmospheric Environment 286 (2022) 119251

Contents lists available at ScienceDirect

Atmospheric Environment

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

Can data reliability of low-cost sensor devices for indoor air particulate matter monitoring be improved? – An approach using machine learning

H. Chojer[a,b], P.T.B.S. Branco[a,b], F.G. Martins[a,b], M.C.M. Alvim-Ferraz[a,b], S.I.V. Sousa[a,b,*]

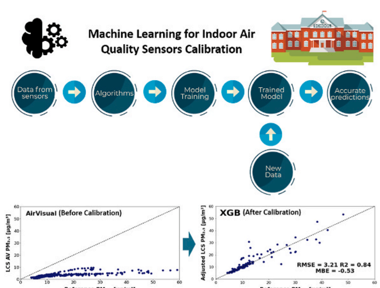
[a] *LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal*

[b] *ALiCE – Associate Laboratory in Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal*

HIGHLIGHTS

- Low-cost sensors (LCS) showed low accuracy in schools' measurements.
- LCS devices showed improved accuracy with field calibration using machine learning.
- Boosting regression models performed best and were most robust.
- The LCS responded differently in classrooms in comparison with the lunchroom.

GRAPHICAL ABSTRACT



ARTICLE INFO

Index Terms:
Indoor air quality
Low-cost sensors
Particulate matter
Nursery and primary school
Machine learning
Calibration
Data analysis

ABSTRACT

Poor indoor air quality has adverse health impacts. Children are considered a risk group, and they spend a significant time indoors at home and in schools. Air quality monitoring has traditionally been limited due to the cost and size of the monitoring stations. Recent advancements in low-cost sensors technology allow for economical, scalable and real-time monitoring, which is especially helpful in monitoring air quality in indoor environments, as they are prone to sudden peaks in pollutant concentrations. However, data reliability is still a considerable challenge to overcome in low-cost sensors technology. Thus, following a monitoring campaign in a nursery and primary school in Porto urban area, the present study analyzed the performance of three commercially available low-cost IoT devices for indoor air quality monitoring in real-world against a research-grade device used as a reference and developed regression models to improve their reliability. This paper also presents the developed on-field calibration models via machine learning technique using multiple linear regression, support vector regression, and gradient boosting regression algorithms and focuses on particulate matter (PM_1, $PM_{2.5}$, PM_{10}) data collected by the devices. The performance evaluation results showed poor detection of particulates in classrooms by the low-cost devices compared to the reference. The on-field calibration algorithms showed a considerable improvement in all three devices' accuracy (reaching up to $R^2 > 0.9$) for the light scattering technology based particulate matter sensors. The results also show the different performance of low-cost devices in the lunchroom compared to the classrooms of the same school building, indicating the need for calibration in different microenvironments.

* Corresponding author. LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal.

E-mail address: sofia.sousa@fe.up.pt (S.I.V. Sousa).

https://doi.org/10.1016/j.atmosenv.2022.119251
Received 15 March 2022; Received in revised form 23 June 2022; Accepted 24 June 2022
Available online 28 June 2022

1. Introduction

Air pollution is responsible for over 7 million premature deaths every year globally (WHO, 2022). A significant number of fatalities arises as more than 90% of the world's population live in places where air pollution exceeds its set guidelines (WHO, 2022). Pollution in indoor environments has been garnering much attention recently, where the pollutant concentration levels can be much higher than outdoors (Jones, 1999) and their numerous adverse health impacts have been documented in the literature (Rosário Filho et al., 2021; Schraufnagel et al., 2019; Sousa et al., 2012; Weschler et al., 1989). As humans spend an increasing amount of time indoors, more than 90%, (Klepeis et al., 2001), monitoring indoor air quality (IAQ) becomes essential. Moreover, people can modify indoor environmental exposures, making its monitoring even more important.

Children's exposure to air pollutants is especially alarming as they are considered a risk group due to the development of respiratory and the immune systems during childhood and higher inhalation rate (per kilogram of their body weight) than adults, which leave them susceptible to health risks (Branco et al., 2020a; Mudway et al., 2019; Nieuwenhuijsen et al., 2006; Schwartz, 2004; Sousa et al., 2012). The resulting ramifications of this early exposure may have long-term effects on their health as well (Schwartz, 2004). Past research associates children's exposure to poor air quality with adverse health impacts, including reduced lung function, asthma and allergies (Gehring et al., 2013; Mendell, 2007; Schwartz, 2004; Sousa et al., 2012). Hence, monitoring children's exposure to air pollutants is of great significance. In the context of indoor environments, nursery and primary schools are a unique case study for two main reasons: i) children spend more time there than in any other environment besides home, being the first place of social activity in life (Branco et al., 2014b); and ii) previous studies have evidenced that the poor indoor air quality often found in nursery and primary schools impairs children's health (Branco et al., 2020b; Nunes et al., 2016; Sá et al., 2017).

Recently, low-cost sensors (LCS) for air quality monitoring have witnessed remarkable advancements (Sá et al., 2022; Snyder et al., 2013). Generally, the term LCS implies inexpensive sensor nodes costing less than 100 US dollars (Chojer et al., 2020; Morawska et al., 2018; Rai et al., 2017). LCS are a portable, user-friendly and economical solution that can provide near real-time air quality analysis while offering scalability and widespread availability (Castell et al., 2013; Thompson, 2016; White et al., 2012). They spearhead the paradigm shift in air quality monitoring Snyder et al. (2013) by being a potential supplement to the enormous, expensive traditional monitoring methods by implementing relatively cheaper technologies such as electrochemical cell (EC), metal oxide semiconductor (MOS), nondispersive infrared (NDIR), nephelometry, and optical particle counters (OPC) among others (Hagan and Kroll, 2020; White et al., 2012). Such LCS based Internet of Things (IoT) devices for indoor air quality monitoring are especially promising as portable devices that can easily be deployed where bulky monitoring devices are not feasible.

Many such devices are currently commercially available for any common citizen. They do not require qualified technicians to operate them, thus they can be used ubiquitously. There have been several studies using LCS devices developed or deployed to monitor various indoor environments like homes/residences (Singer and Delp, 2018; Zamora et al., 2020), schools (Wang et al., 2017), and offices (Parkinson et al., 2019). However, with the design flaws accompanying the lower cost, data reliability is still a concern associated with this technology (Peterson et al., 2017; White et al., 2012). The poor reliability stems from the weak reproducibility, high cross-sensitivity, frequent recalibration requirement, variability in measurements with changing ambient conditions and the short life spans, to name a few (Morawska et al., 2018; Peterson et al., 2017; White et al., 2012; Zhang et al., 2014).

LCS for particulate matter (PM) monitoring usually employ nephelometry. The sensors detect the aerosols by the detection of light scattered or reflected. The scattered light detected by the detector is dependent upon properties like particle size distribution, shape and refractive index (Hinds, 1999). The particle properties vary in real-world and hence, the PM mass concentrations with which these sensors are calibrated in the lab (with a reference mass measurement) will be different when deployed for uncontrolled monitoring (Hagan and Kroll, 2020). Moreover, it also adds the limitation of varying environmental parameters like humidity to the performance of these sensors. The hygroscopic influences can modify particle properties like shape, density and refractive index as the particles take up water (Hagan and Kroll, 2020). Therefore, while these nephelometers are small and inexpensive, they have associated concerns as they are not a direct mass measurement (Snyder et al., 2013). The devices housing nephelometers usually also lack a heater/dryer at the inlet, which is useful to remove the moisture that would influence the performance of these sensors (Barkjohn et al., 2021; Giordano et al., 2021; Levy Zamora et al., 2019). These concerns for LCS and, specifically, LCS for PM monitoring pose some challenges and measurement limitations with these sensors.

Recently, there have been several studies trying to address this issue of data accuracy and precision by evaluating the LCS performance for particulate matter monitoring (Barkjohn et al., 2021; Levy Zamora et al., 2019; McFarlane et al., 2021; Singer and Delp, 2018; Zamora et al., 2020). Levy Zamora et al. (2019) evaluated the three Plantower PMSA003 sensors in field and laboratory settings. and found the accuracy ranged from 13 to more than 90% compared to reference values. They also found the accuracy of the sensors decrease with increasing humidity. They conducted their indoor field experiments in a residential apartment of a coastal city. Zamora et al. (2020) evaluated the performance of three LCS devices (AirVisual Pro, Speck, and AirThinx) for over a year. They found that AirVisual Pro exhibited the best accuracy (about 86%) compared to the filter. They also concluded that high accuracies could be observed for AirVisual Pro and AirThinx with one or two calibrations during one year, although, monthly calibration was needed for achieving highest accuracies. Singer and Delp (2018) simulated particles from typical residential sources in a laboratory to test LCS devices like AirBeam, AirVisual, Foobot and PurpleAir. They found that the LCS devices under-reported the concentration peaks and even missed events for particles emitted below 0.3 μm in diameter. Tryner et al. (2021) designed a sampling platform of LCS for IAQ monitoring and tested 9 units with reference monitors in an occupied home. They used Plantower PMS5003 (the sensor used in PurpleAir) for aerosol monitoring and found that the sensor overestimated the $PM_{2.5}$ concentrations compared to the reference. Hence, there is a lack of agreement in literature about the behaviour of these LCS for aerosol monitoring, especially for indoor monitoring, as in some studies they are reported to be understating the PM concentrations while overstating in others. A potential reason of this lack of consensus can be the variation of properties and size distribution of the particulates from one environment to the other as the nephelometers are heavily influenced by these parameters.

In general, the current research for improving the data reliability of LCS is focused on developing correction or calibration models. Giordano et al. (2021) best described it as the process of measuring true aerosol concentrations using LCS side-by-side with a trusted reference device and finding a calibration algorithm that best describes their relationship. Barkjohn et al. (2021) used more than 10,000 PurpleAir sensors in ambient air and developed a United States wide correction for $PM_{2.5}$ measurements using a simple linear regression method. Magi et al. (2020) evaluated PurpleAir devices in near-road urban ambient settings and used multiple linear regression (MLR) models to improve the accuracy (27–57% improvement) of the $PM_{2.5}$ data. Recently, McFarlane et al. (2021) showed in their monitoring campaign in Accra, Ghana that MLR and random forest regression models that have been previously shown to improve accuracy between PurpleAir and reference data, did not result in significant improvement in Accra. They used gaussian mixture regressions to achieve high correlation and accuracy ($R^2 = 0.88$

and MAE = 2.2 μg/m^3). Low-cost sensors are known to exhibit different behaviour with different chemical composition of aerosols (Giordano et al., 2021; Singer and Delp, 2018) Thus, the extensive literature survey solidifies the hypothesis that the environment has a major role to play in the accuracy and calibration of LCS devices for PM monitoring.

Apart from the effect of environment on LCS, the behaviour of pollutants indoors is also known to be significantly different than outdoors, especially for particulate matter whose concentrations can be up to five times higher indoors (Branco et al., 2014a). The particulate matter (PM) concentrations indoors are also highly prone to peaks in comparison to smoother patterns of concentration outdoors (Branco et al., 2014a; Nunes et al., 2015). Even within the same indoor environment, significant variations can occur in PM chemical compositions and sizes between different microenvironments (Amato et al., 2014; Moreno et al., 2014). Due to these reasons, it is of great significance to study the performance of low-cost sensors in different indoor environments. Moreover, and as far as the authors' knowledge goes, no similar studies were performed using machine learning (ML) models for LCS calibration in nurseries and primary schools. Schools are a specific microenvironment that can potentially have different composition, size, and concentration levels of PM pollution.

Thus, the current work intended to evaluate the performance and improve data reliability of commercially available LCS devices for indoor air particulate matter monitoring in nurseries and primary schools by using advanced ML algorithms to build on-field calibration models. It also intended to evaluate and compare several models for calibration, namely multiple linear regression (MLR), support vector regression (SVR) and boosting regression models.

2. Methodology

2.1. Deployed devices

The IAQ monitoring was conducted in a nursery and primary school in the urban area of Porto city, Portugal. Three commercially available IAQ monitoring devices were deployed in four different rooms of the same building, including three classrooms and one lunchroom for infants (<3 years old), pre-schoolers (3–5 years old) and primary school children (6–10 years old). The research campaign was carried out from June 3, 2019 to July 8, 2019. The deployed low-cost devices were AirVisual Pro, PurpleAir PAII SD, uRAD Monitor Model A3 (IQAir, 2021; PurpleAir, 2021; uRAD, 2021), alongside the research-grade device (TSI DUSTTRAK DRX Aerosol Monitor) used as reference, as can be observed in Table 1. PurpleAir houses two Plantower PMS5003 sensors and the an average of the two sensors was taken for our analysis. The calibration of TSI DUSTTRAK was performed prior to the monitoring campaign by the manufacturer (TSI) as per standard ISO 12103–1, A1 test dust (Arizona dust). All the devices were deployed side-by-side on a table or a shelf, usually near a wall or in the centre of the microenvironment, at about the height of the children's breathing. Natural ventilation was adopted in all the studied microenvironments by opening windows (to the playgrounds, the residential area or the street, depending on the ME location) and/or doors (inside the building to the hallways).

Regression models were implemented for all three particulate matter fractions (PM_1, $PM_{2.5}$, and PM_{10}). All three LCS devices collected temperature and relative humidity data and used light scattering technology for PM monitoring. They all came with prior factory calibration settings, and the ML models were implemented for the on-field calibration phase.

2.2. Data analysis

The methodology flowchart of the work from raw data to the final best models is shown in Fig. 1.

2.2.1. Data preprocessing

The pre-processing and visualization involved merging datasets and removing the null data points from the merged datasets and data visualization through distribution plots to observe skews and scatter plot matrices. Arithmetic mean was taken to harmonize the sampling rate, and the datasets were merged with the basis of timestamps to further eliminate any missing values between the devices to establish a common ground for comparison of the devices and the creation of calibration models. Literature (Giordano et al., 2021), and references therein, show that due to the lack of a heater or dryer to remove the moisture at the sensor inlet, RH influences the low-cost PM sensors. Prior studies have also demonstrated that temperature can be a significant predictor in low-cost PM sensor response (McFarlane et al., 2021). Hence, it was concluded that T and RH, along with low-cost device measurements, should be used for developing the models.

Data sorting involved randomly splitting the dataset into training and testing subsets. For training, 80% of the data were selected, and the rest were used for final model testing. Regression algorithms were trained individually for each pollutant of each device for all the studied rooms. The datasets were split using the train_test_split function from scikit-learn. A fixed random state was used to ensure that the same split was used to draw a consistent comparison between different models.

2.2.2. Regression training algorithms

The regression algorithms used were MLR, SVR and boosting regressions (GBR - gradient boosting regression and XGB- extreme gradient boosting). The input variables were temperature (T), relative humidity (RH), and LCS concentrations trained to match all models' reference concentrations. The output variable was the adjusted LCS concentration.

MLR model is an extension of simple linear regression, and it was used here due to multiple explanatory variables and is defined in Equation (1).

$$y_{i,pred} = p_0 + p_1 x_1 + \ldots + p_n x_n \tag{1}$$

where $y_{i,pred}$ is the ith predicted value of the model, p_i are the regression coefficients and x_i are the explanatory variables. The *statsmodels* Python module was used to ascertain the regression summary and significance test results (Seabold and Statsmodels, 2010). The level of statistical significance was set at a p-value of 0.05, except when stated otherwise.

Support vector machine was a training algorithm initially developed as a classifier (Boser et al., 1992) and quickly found its application in regression analysis called SVR (Drucker et al., 1996). The fundamental concept of training an SVR is solving a convex optimization problem:

Table 1
The deployed devices in the research campaign.

Device	Type	Pollutants Monitored	PM Sensor	Monitoring Interval	IoT
AirVisual Pro	Low-cost	PM_1, $PM_{2.5}$, PM_{10}, CO_2	AVPM25b	10 s	Enabled
PurpleAir PAII SD	Low-cost	PM_1, $PM_{2.5}$, PM_{10}	Plantower PMS5003 (2)	2 min	Enabled
uRADMonitor Model A3	Low-cost	PM_1, $PM_{2.5}$, PM_{10}, CO_2, O_3, CH_2O, VOCs	Winsen ZH03A	1 min	Enabled
Reference PM: TSI DUSTTRAK DRX Aerosol Monitor	Research-grade	PM_1, $PM_{2.5}$, PM_4, PM_{10}, Total Suspended Particles (TSP)	–	10 s	No

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i + C \sum_{i=1}^{l} \xi_i^* \tag{2}$$

$$\text{subject to } w^T \Phi(x_i) + b - z_i \leq \varepsilon + \xi_i, \rightarrow \left| z_i - w^T \Phi(x_i) - b \right| \leq \varepsilon + \xi_i^*, \rightarrow \xi_i, \xi_i^* \geq 0, i = 1, \ldots, l, \tag{3}$$

where $\varphi(x_i)$ maps x_i into a higher-dimensional space, w is the vector variable, C is the regularisation parameter and ξ denotes the deviation from the margin beyond the maximum error ε (giving an additional hyperparameter) (Chang and Lin, 2011; Smola and Schölkopf, 2004). The present study analyses both linear and non-linear (RBF - radial basis function) kernels for SVR training.

Boosting is a type of ensemble learning which converts a sequence of weak learners into complex models to predict by combining all learners in the end (Friedman, 2002). GBR and XGB were used in the present study as they apply a similar idea for regressions (Scikitlearn, 2021). An extensive grid search for several hyperparameters was done for boosting algorithms, with a considerable emphasis on optimizing the number of iterations and learning rate.

Hyperparameters optimization was performed for SVR, XGB, and GBR models via an exhaustive grid search performed with 3-fold cross-validation. The different hyperparameters that were optimized are shown in Table 2. The hyperparameters that exhibit a strong influence on the model were generally chosen to be optimized.

Subsequently, models underwent final testing with the single split holdout testing dataset, and the best performing models were obtained for all the low-cost devices in every room and each pollutant individually.

2.2.3. Performance indexes

The performance indexes considered for the evaluation of the models were the coefficient of determination (R^2), the root mean square error (RMSE) and mean bias error (MBE), given by Equations (4)–(6), respectively.

$$R^2 = 1 - \frac{SS_r}{SS_t} = 1 - \frac{\sum_i \left(y_i - y_{i,pred}\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_i \left(y_i - y_{i,pred}\right)^2} \tag{5}$$

$$MBE = \frac{1}{n} \sum_i \left(y_i - y_{i,pred}\right) \tag{6}$$

where SS_r is residual sum of squares, SS_t is total sum of squares of deviations in relation to the global reference mean ($\bar{y}$), y_i is the ith true value (reference), and $y_{i,pred}$ is the ith predicted value of the model.

The data treatment was done using the open-source Python 3.7 with Jupyer Notebook interface (Beg et al., 2021; Pilgrim and Willison, 2009). The regression modelling and analysis was performed using the *scikit-learn* library (Pedregosa et al., 2011).

3. Results & discussion

3.1. Data preprocessing and visualization

Table 3 shows the availability of devices for each classroom during the experimental campaign.

Only one classroom (S01_A) and the lunchroom (P01_LR) had all the devices available for the entire period of the experimental campaign. Data acquired were not uniform for all devices due to differences in time resolutions and data loss incurred by the devices. The data loss were especially evident in the case of uRAD monitor for several classrooms. Other low-cost devices lost almost no data points during the monitoring campaign. As mentioned in Table 1, PurpleAir had a minimum reporting interval of 2 min. Hence, taking 10-min means after removing the null/blank values resolved the issue and the datasets were then merged based on timestamps, which gave a common ground for comparison between all the devices. The discussion of results specific to the classrooms is detailed in section 3.2.1 and for the lunchroom in section 3.2.2., respectively. The concentrations of all two PM fractions for the monitoring period in the school can be seen in Fig. S1 (supplementary material).

Fig. 2 shows the scatter plot of $PM_{2.5}$ as measured by the three LCS devices for classroom S01_A (as an example, with n = 712). A distinction between the occupancy and non-occupancy periods shows that most non-occupancy periods exhibited low $PM_{2.5}$ concentrations, whereas there were higher concentrations for occupancy periods where re-suspension of particles might be more prevalent. The necessity of the field calibration is apparent as none of the three LCS concentrations even broke the 12 μg/m^3 mark compared to the reference, which showed up to five times higher average concentrations. PurpleAir exhibited almost perfect linearity at low concentrations of $PM_{2.5}$ but failed to corroborate with the reference at higher concentrations. uRAD monitor also showed some linearity at lower concentrations but showed poorer performance at higher concentrations. Similar results were obtained for other classrooms. The results imply that all three LCS devices performed well for concentrations lower than 15 μg/m^3, but their performance worsened at elevated $PM_{2.5}$ concentrations inside the classroom. This corresponds also for some parts with the period of non-occupancy. Paradoxically, the concept of employing the LCS is rendered useless if they are not able to track higher pollutant concentrations, especially when the indoor spaces are occupied and the air is being inhaled by the people. It is precisely at those periods of elevated concentrations that these devices, with their real-time monitoring, are supposed to aid the end-user. In our specific environment of classrooms in a school building, the 3 LCS devices struggled to detect the particulates. Even if these devices were to be used for indicative purposes, they did not indicate a significant rise in PM concentrations in the classroom. Our results differ from that of some prior studies that have shown the LCS devices tested here (AirVisual and PurpleAir) to be able to detect PM concentrations of up to 350 μg/m^3 (Singer and Delp, 2018; Tryner et al., 2021; Zamora et al., 2020), although those studies were conducted in different microenvironments and were not monitoring inside school classrooms. The findings from the experiments of Singer and Delp (2018) are especially of interest in interpreting our results as they found that AirVisual and PurpleAir devices both did a poor job in detecting particles from sources like dust from a mop (coarser particles) and from sources of ultrafine particles. Hence, an explanation of the discrepancy found here might be that the PM in the classroom is composed of particles in a size range that these sensors struggle to detect.

For all three LCS devices, significant overlaps can be observed between occupancy and non-occupancy period for $PM_{2.5}$ concentrations. A probable explanation might be that the transition from occupancy period to non-occupancy period is taken by fixed boundary conditions as provided by the timetable of the school authorities as per their schedules. In reality, there might have been some activity even after the

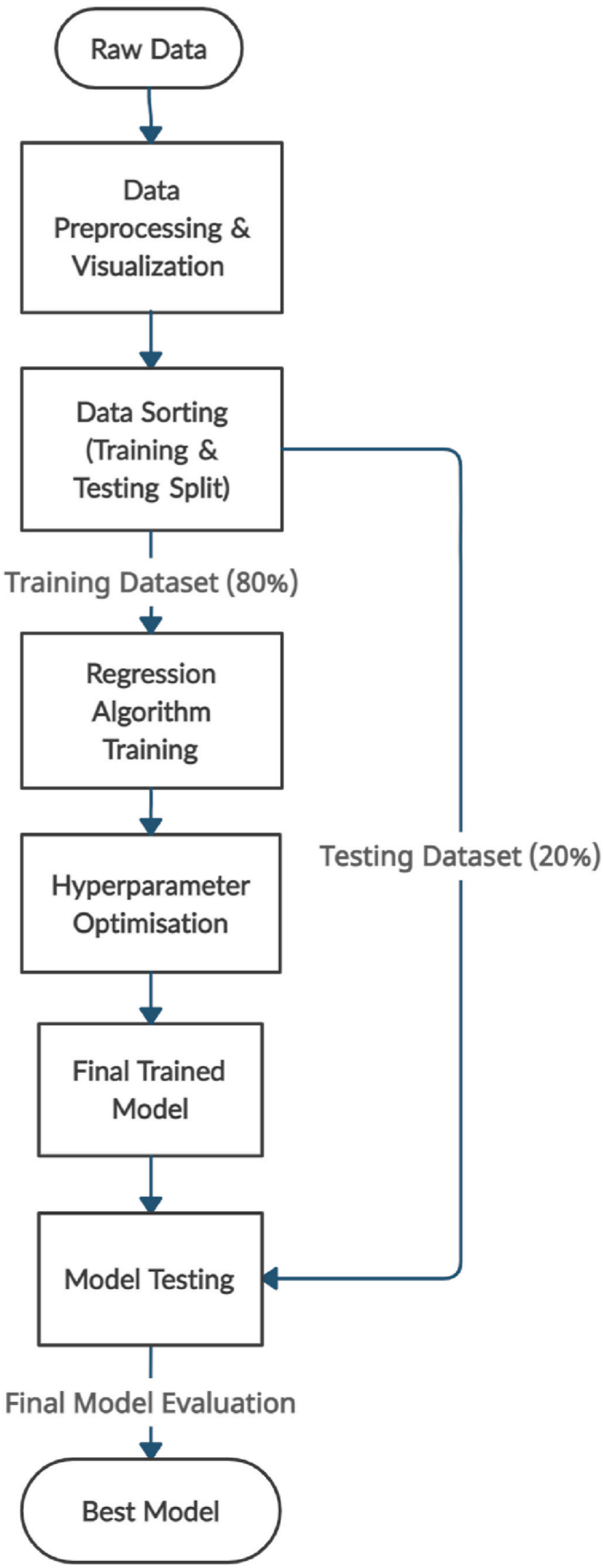


Fig. 1. Methodology flowchart of the data analysis.

Table 2
Hyperparameters optimized for the models.

Model	Hyperparameters
SVR	Kernel, Regularisation parameter C
GBR	number of boosting stages, learning rate, maximum depth
XGB	number of boosting stages, learning rate, maximum depth, subsample, gamma

Table 3
Device availability for each classroom.

Classroom	AirVisual	PurpleAir	URAD	Reference PM
S01_B	✔	✔	×	✔
S01_A	✔	✔	✔	✔
P01_A	×	✔	✔	✔
P01_LR	✔	✔	✔	✔

Reference PM: TSI DUSTTRAK DRX Aerosol Monitor.

designated times as the rooms would have been slowly vacated, causing further re-suspension, and, consequently, higher PM concentrations. Similarly, it is hard to quantify how long before the designated time were the rooms slowly occupied. Although the authors note the differentiating behavioural trends between occupancy and non-occupancy periods with a visual inspection, the calibration results and discussion were conducted for the entire period.

Fig. 3 shows the strength of linear correlation for the monitoring period in the representative classroom S01_A between all the variables involved via a Pearson correlation matrix plot. AirVisual showed the strongest linear relationship with reference, followed by uRAD monitor while PurpleAir showed the weakest overall linearity. The results observed here differ from previous studies (Barkjohn et al., 2021; Magi et al., 2020; Malings et al., 2020) that showed a better correlation ($r > 0.7$ for all three studies) for PurpleAir. It should be noted that these studies were conducted for ambient air monitoring. Hence, as discussed above, this difference in findings might be due to the difference in the monitoring environments and the types and sizes of pollutants therein.

Temperature and RH, expectedly, showed a negative correlation towards each other. Temperature and PM concentrations showed weak correlations, which is in line with the findings in the literature (Zamora et al., 2020).

For two classrooms (S01A and S01B) the RH values varied from 30% to 70%, whereas for the classroom P01A and lunchroom P01LR the variation was between 50% and 70%. Levy Zamora et al. (2019) showed in their year long study in a coastal city (in field and lab settings) that their Plantower PMSA003 sensors (similar to the one used in PurpleAir) were affected by RH levels higher than 50%. Other studies have also shown the influence of high RH on Plantower sensors (Jayaratne et al., 2018; Malings et al., 2020). The findings from the classrooms in the present study differ from those observed previously as PurpleAir and uRAD showed weak correlations with RH (maximum pearson correlation coefficient of 0.25). Neither of these devices has a dryer at the inlet, implying that the unmodified air should influence the PM sensor due to the moisture. Perhaps, the two devices already have a correction factor applied for RH by the manufacturer (no specific information from the manufacturer regarding this issue). Another possible explanation might be that the mean RH values in the present study were around 55% for the entire monitoring period.

3.2. Model training and testing results

3.2.1. Classroom results

The results obtained from all the classrooms were similar. The set consisting of 80% of the total dataset in S01_A classroom yielded around 570 data points used to train the models. The initial regression models were trained with the default hyperparameters provided by *scikit-learn*.

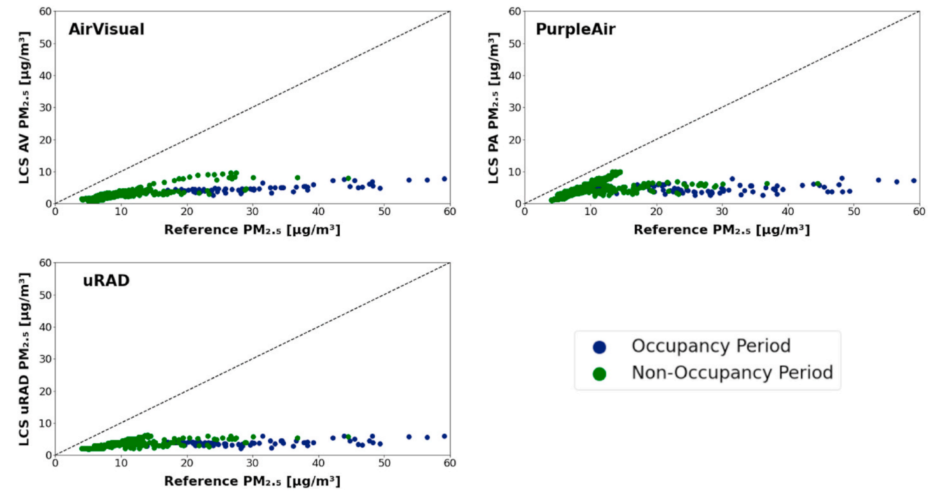


Fig. 2. Scatter plot of the AirVisual Pro, PurpleAir, and uRAD Monitor $PM_{2.5}$ 10-min mean measurements with reference using raw data for classroom S01_A; LCS: Low-cost sensor.

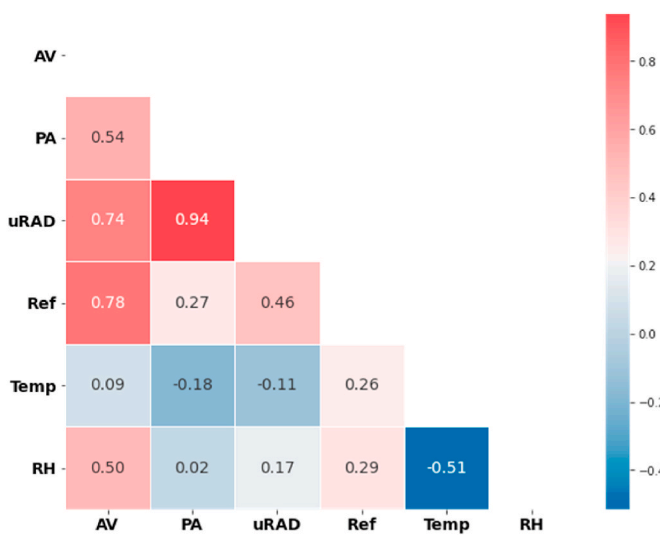


Fig. 3. Pearson correlation matrix plot of all the variables involved in $PM_{2.5}$ measurements for classroom S01_A; AV: AirVisual Pro, PA: PurpleAir, uRAD: uRAD Monitor, Ref: Reference, Temp: Temperature, RH: Relative Humidity.

It was done to have some sort of baseline performance of the models for the given data. This performance was evaluated via the three statistical metrics described in the methodology: R^2, RMSE, MBE. For S01_A $PM_{2.5}$ monitoring with AirVisual Pro (taken as an example case for brevity), all the MLR coefficients obtained were statistically significant (p-value $<$ 0.01) and hence, considered to be valid.

The models then underwent hyperparameter optimization on the parameters mentioned in Table 2. It should be noted that the optimization methodology was followed, but manual tuning was done whenever necessary. For example, in cases where overfitting was suspected due to low cross-validation scores evaluated for the hyperparameter optimization, the overfitting suspected was also verified by making learning and deviance curves (learning curves of all four models in Supplementary Material Fig. S4). In these cases, manual optimization was done by a hit and trial manner to rectify the overfitting. The results presented here include the test set results of the final trained models. The learning curves also show that most models would have continued to improve if a larger dataset was available, with only the GBR model showing a slight decline in R^2 towards the end.

Fig. 4 shows the scatter plot of LCS device AirVisual Pro and reference $PM_{2.5}$ values of final test set results arising from implementing the four models. The subfigures show that the machine learning strategy improved the results with all four models showing significant improvement compared to the scatter observed in Fig. 2. It can be noted that after the application of the models, the LCS device AirVisual Pro did not seem to be understating the pollutant concentrations.

Moving the discussion further to the comparison between the four models, MLR model showed the lowest R^2 and the highest error and bias values for our example case. SVR showed the most significant improvement, followed very closely by GBR and XGB models. The hyperparameter optimization results showed that RBF kernel with a regularisation parameter fixed at 100 was the best configuration for SVR. Following the trend of the uncorrected scatter plots, at lower pollutant concentrations, the LCS showed a very strong resemblance to the reference values, but the deviations increased as the pollutant concentrations increased and the 95% confidence interval widens (Fig. S 2. In the supplementary material). Further, the negative MBE values for the two boosting models (and for all three LCS devices) imply that even after implementing these models, the predicted values mostly underestimated the PM2.5 concentrations compared to the reference values. Although, the extent of the underestimation is much lower in comparison to the raw data. SVR also showed the lowest MBE values compared to the other three models for AirVisual. In general, taking all three metrics (R^2, RMSE, and MBE) into consideration, the three devices were significantly improved by SVR, GBR and XGB, and they only marginally outperformed each other in most instances.

The same methodology was applied for all PM fractions and the rest of the devices. Table 4 shows the performance indexes for both training and testing stage results using the four models for S01_A classroom as an example (similar results for the other two classrooms can be seen in the supplementary material - Tables S1 and S2). It is prudent to base the performance discussion on testing set results, while the training set results signify that the models were not underfitting.

Throughout the results for all PM fractions and classrooms, before applying the supervised learning models, all devices showed very low concentrations for all PM concentrations, similar to the ones observed in Fig. 2. AirVisual Pro consistently performed better than the other two devices. Table 4 shows low R^2 values of PurpleAir and uRAD monitor even after applying the MLR model, which implies that PurpleAir and uRAD monitor did a poor job detecting periods of elevated PM pollution in real-time, specifically in the classroom setting. AirVisual Pro showed moderate to good R^2 scores and lower error values in comparison. It shows that AirVisual is better capable of tracking the aerosol concentration fluctuations in near real-time but the displayed concentrations

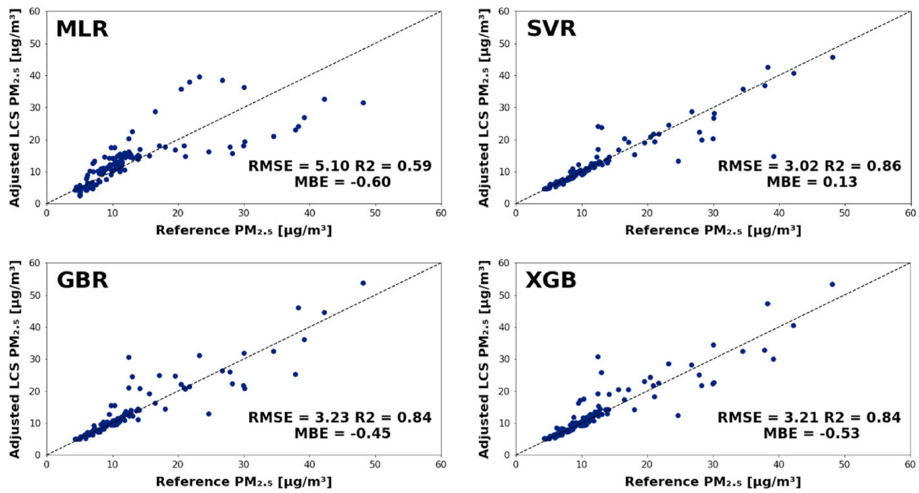


Fig. 4. Model test set scatter plot for all four models for $PM_{2.5}$ 10-min mean measurements of AirVisual Pro for classroom S01_A; MLR: Multiple Linear Regression, SVR: Support Vector Regression, GBR: Gradient Boosting Regression, XGB: Extreme Gradient Boosting.

Table 4
Performance indexes of each model applied to the three low-cost devices for PM for the classroom S01_A.

PM_1	Model	AirVisual			PurpleAir			URAD		
		R^2	RMSE	MBE	R^2	RMSE	MBE	R^2	RMSE	MBE
Training	MLR	0.629	4.567	−0.000	0.114	7.054	−0.000	0.105	7.090	1.357
	SVR	0.956	1.563	0.207	0.818	3.196	0.464	0.785	3.476	0.597
	GBR	0.985	0.913	−5.604	0.919	2.129	−0.000	0.905	0.231	0.000
	XGB	0.988	0.812	0.988	0.999	0.172	−0.001	0.992	0.649	−0.000
Testing	MLR	0.598	4.786	−0.642	0.182	6.827	−0.094	0.159	6.921	1.224
	SVR	0.841	3.015	0.105	0.736	3.876	0.410	0.706	4.092	0.443
	GBR	0.780	3.539	−0.679	0.744	3.823	0.092	**0.712**	**4.049**	−0.495
	XGB	**0.857**	**2.852**	−0.436	**0.802**	**3.361**	−0.162	0.696	4.164	−0.234
$PM_{2.5}$										
Training	MLR	0.645	4.728	−0.000	0.136	7.377	−0.000	0.134	7.387	1.344
	SVR	0.955	1.681	0.149	0.831	3.257	0.483	0.804	3.517	0.529
	GBR	0.993	0.670	3.225	0.863	2.940	−0.000	0.957	1.638	1.514
	XGB	0.975	1.243	0.975	0.915	2.309	0.004	0.908	2.404	0.037
Testing	MLR	0.594	5.100	−0.597	0.215	7.093	−0.062	0.195	7.179	1.261
	SVR	**0.857**	**3.024**	0.130	0.742	4.062	0.536	0.739	4.091	0.371
	GBR	0.838	3.225	−0.447	**0.743**	**4.061**	−0.429	**0.765**	**3.882**	−0.096
	XGB	0.839	3.209	−0.528	0.650	4.733	−0.260	0.724	4.203	−0.329
PM_{10}										
Training	MLR	0.833	6.872	0.000	0.111	15.849	−0.000	0.090	16.034	3.038
	SVR	0.992	1.491	0.074	0.838	6.776	0.986	0.752	8.37	1.153
	GBR	0.999	0.564	−0.000	0.909	5.060	−0.000	0.937	4.232	−0.000
	XGB	0.992	1.480	0.021	1.000	0.226	0.001	0.898	5.373	0.136
Testing	MLR	0.801	7.326	−0.985	0.193	14.747	−0.404	0.161	15.037	2.709
	SVR	0.902	5.145	−0.062	0.722	8.650	1.125	**0.742**	**8.334**	0.121
	GBR	0.831	6.751	−1.078	**0.727**	**8.583**	−0.840	0.647	9.758	−1.506
	XGB	**0.923**	**4.560**	−0.662	0.626	10.043	−0.254	0.679	9.301	−1.215

Bold numbers represent best performing models; MLR: Multiple Linear Regression; SVR: Support Vector Regression; GBR: Gradient Boosting Regression; XGB: Extreme Gradient Boosting; The units for error values are μg/m.[3].

remain lower relative to the reference concentrations. On the other hand, the weak linear relationship between PurpleAir, uRAD monitor and the reference for all PM fractions was evident with the weaker performance results even after using the MLR model. In general, the results show that using SVR and boosting models yielded better calibration results than MLR model for the LCS devices.

Boosting and SVR models after hyperparameter optimization showed a significant improvement in the performance of all three devices for all PM fractions. For SVR, the optimization always favoured the rbf kernel while the regularisation parameter C varied. Both the boosting models were steady in their performance throughout and outperformed other models in most cases. The SVR, GBR and XGB models adjusted LCS values showed strong association with the reference, but some caution is needed here because of the relatively small size of the dataset, which can easily lead to overfitting in complex models. However, in the current study, the extensive model evaluation upon hyperparameter optimization with learning curves and deviance curves increases confidence in the developed models.

The results for the devices were similar throughout for all PM fractions and the classrooms (Supplementary Material – Tables S1–S3), with PM_{10} showing slightly higher RMSE values compared to $PM_{2.5}$ and PM_1 even after implementing the models. As the errors are squared before being averaged, RMSE accentuates large errors. It implies that errors were slightly larger for PM_{10} predicted values compared to reference than for the other two PM fractions. AirVisual for PM_{10} monitoring (both pre and post ML corrections) resembled the reference better in terms of R^2 score. After applying the MLR model, it achieved a high R^2 score of more than 0.8. It might imply that AirVisual is more suited to monitor coarse PM fractions than finer ones. These results somewhat deviate from those of Wang et al. (2020), who showed in their "accuracy for quantifying event integrated PM_{10}" results that PM_{10} data from LCS devices might have lower consistency than the $PM_{2.5}$ data. However, they also noted that one LCS device reported PM_{10} similar to the reference for several sources.

3.2.2. Lunchroom results

Fig. 5 shows that all LCS devices showed strong linearity (r > 0.95) even for higher concentrations. This behaviour is different than that observed in the classrooms, which showed linearity only for lower concentrations. In the classrooms, the LCS devices weren't even able to show an increase in PM concentrations, as is evident from Fig. 2. For the lunchroom, all three devices were able to correctly indicate increasing PM concentrations, albeit still understating the concentrations. The calibration benefited from this response, and MLR models, in this case, showed very high R^2 scores (>0.9) and very low error values, which was different from the results from the classrooms (Tables S1–S3 in the supplementary material).

At a perfunctory glance, the results might seem contradictory. Hence, concentration ranges, median concentrations, occupancy and the temperature and RH conditions in all the rooms were analyzed. Table 5 shows the $PM_{2.5}$ concentration range (as an example, similar results were obtained for other PM fractions) in all the classrooms and the lunchroom (P01_LR). The idea was to check if the majority of PM concentration in the lunchroom lies in the low concentration regime, which would imply similar behaviour for the devices in all the monitored rooms inside the school. However, both the minimum and maximum concentrations observed in the lunchroom were higher than in the classrooms. It also showed higher median concentrations than in the classrooms. The daily average duration of occupancy in all the rooms was also similar. Moreover, the temperature and RH conditions in all the rooms were also similar. Hence, these were excluded as a possible explanation for this behaviour.

The reason for this discrepancy could lie in the difference of PM compositions present in the classrooms and the lunchroom. Past studies showed that different sources of pollutants exist for different rooms in a school (Branco et al., 2014a, 2019). For the present study, the school's kitchen was also adjacent to the lunchroom, and they shared a window and a door. Therefore, the PM composition in the lunchroom could indeed be different from the ones encountered in the classrooms. Past studies (in controlled/lab settings) for indoor and outdoor environments (Giordano et al., 2021; Liu et al., 2017; Salimifard et al., 2020) also showed that the chemical composition plays an important role in the linearity and sensitivity of the sensor response for low-cost PM sensors. The PM compositions can differ not only due to the chemical compositions, but also due to different particle size distributions found in these two microenvironments, which has an even larger impact on the sensor response (Hagan and Kroll, 2020; Ouimette et al., 2021). In the present field analysis of the PM concentrations in rooms within a school having similar temperature and RH conditions and with a PM concentration range that was not exorbitantly different, it can be hypothesized that the difference in the particle size distributions and chemical compositions showed a huge impact in the linearity exhibited by the 3 LCS devices.

Table 5
$PM_{2.5}$ concentration ranges and median values as monitored by reference device in all the rooms monitored.

Room	Concentration Range Observed (μg/m^3)	Median (μg/m^3)
S01_B	4.0–35.2	8.4
S01_A	4.03–59.1	10.1
P01_A	4.7–83.3	14.1
P01_LR	16.5–137.3	28.1

To emphasize the difference between these two microenvironments on the behaviour of LCS devices and the potential ramifications to the calibration models developed, the model developed in one classroom (S01A) was implemented on the raw uncorrected data from another classroom and on the data from the lunchroom. Fig. 6 shows the scatter plot along with the performance metrics for both the cases. The error values found when the classroom model was deployed for the lunchroom were very high. The adjusted values highly overestimated the PM concentration. The ability of the LCS devices to better determine the concentration of finer particles that might have been prevalent in the lunchroom implies that the classroom corrections were not applicable there. The higher coefficient of determination for the lunchroom results implies higher linearity of the lunchroom data that was already observed earlier. The classroom calibration model when applied to another classroom showed better results with lower error values. It also implies that the calibration model developed can be used in another similar microenvironment and improve the data reliability and accuracy of the LCS devices for PM monitoring.

These results have a high significance on the usage of LCS devices.

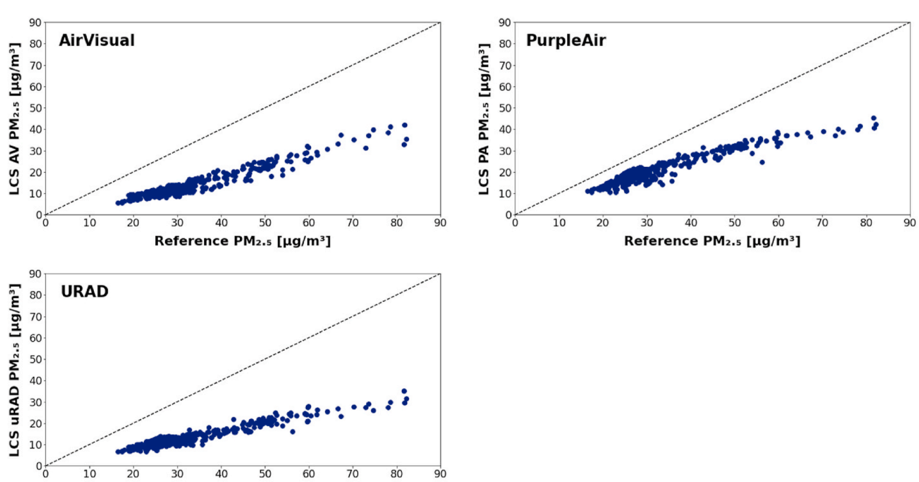


Fig. 5. Scatter plot of the AirVisual Pro, PurpleAir, and uRAD Monitor $PM_{2.5}$ 10-min mean measurements with reference using raw data for the lunchroom; LCS: Low-cost sensor.

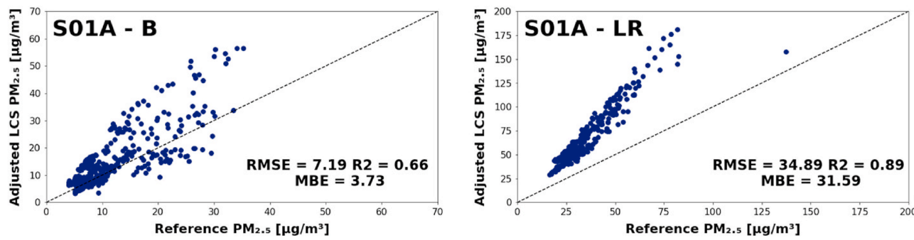


Fig. 6. Plot of the AirVisual Pro $PM_{2.5}$ 10-min mean measurements with implementation of model developed in S01A classroom to S01B (left), and to lunchroom (right); LCS: Low-cost sensor.

These LCS devices are generally manufactured and sold to monitor IAQ in a broad range of indoor environments like homes, offices, schools, hospitals, clinics, etc. This study shows that the LCS devices might behave differently even in different rooms within the same building. While the LCS might register similar PM concentrations for two rooms (say, kitchen and bedroom), the actual concentrations might be poles apart. Much attention needs to be paid to this aspect of low-cost aerosol monitors, and it might be crucial to calibrate them for specific micro-environments. This study's calibration results also prove that making generic calibration models for different microenvironments might not provide reliable concentrations for PM monitoring using LCS devices.

In some aspects, our results differed from prior studies. The classroom monitoring results especially showed unique results that showed poor detection of elevated concentrations of LCS devices. The lunchroom results exhibited strong linearity by all three LCS devices compared to the reference. The difference in sources of pollutants might explain the difference in results shown here and in prior studies. Further studies monitoring IAQ in various school rooms are needed to corroborate the current findings. As these devices begin to gradually become ubiquitous, the duty to disseminate the information on their performance in various environments lies in the scientific community. Deploying these devices in huge numbers might give us the information to work on mitigating pollution. But it might not be the case here as many of the measurements recorded by the devices in the classrooms vastly differed from the reference measurements. Thorough calibration, validation and verification of the data quality of these sensors and devices would ensure that false information is either curbed or rectified. From the overall results of this study, it is also inferred that field calibration using boosting models proved to be very robust for improving the data accuracy of the LCS devices in different indoor microenvironments. These models can be used reliably for similar purposes in future scenarios.

3.3. Limitations

Models made using a small dataset may show strong prediction results because it is easier to over-fit the models than with extensive data modelling. Hence, the single-split holdout dataset results should be taken with a grain of salt and further studies with more extended monitoring periods and larger datasets are needed to confirm this study's results. Generally, it is reasonable to use simpler models for small datasets as they reduce the chance of overfitting. Moreover, longer monitoring periods with multiple identical devices might also show drift in LCS calibration and inter-device variations, which was not studied here.

4. Conclusions

The foremost conclusion from the present study is that the commercially available low-cost devices showed unreliable results by massively understating the pollutant concentrations in real-world settings (an urban school). The data suggested that using these commercially available devices in their current plug and play form, as advertised to be used, understated the pollutant concentrations in the specific environment of classrooms. Even for indicative purposes, the devices did not show enough sensitivity to the PM peaks, and they were not able to provide the warnings necessary in classrooms. The advanced models developed in the present study could improve the data reliability of commercially available LCS devices for IAQ monitoring. Four ML models: MLR, SVR, GBR and XGB were used with hyperparameter optimization to make the corrections. Generally, the in-field calibration approach achieved high accuracy with boosting and SVR models for PM (PM_1, $PM_{2.5}$ and PM_{10}) sensors. For AirVisual Pro, linear models worked well to improve PM sensors' data reliability and can be relied upon due to their simplicity and robust nature, which is especially important to mention for the present work where the datasets are smaller, and overfitting is easier. Although the other two devices, PurpleAir and uRAD monitor, showed a weak linear relationship with the reference and MLR models, they did not show sufficient improvement in their classroom performance.

The lunchroom results were different from the classrooms. All three LCS devices showed strong linearity and were able to indicate increasing aerosol concentrations correctly, although they were still understating the PM concentrations. But it could be concluded that the LCS devices behaved differently in the lunchroom and classrooms. MLR models were able to improve the results for all three devices, which wasn't the case for classroom LCS devices data. The difference in the results found in classrooms and the lunchroom was hypothesized to be due to differences in PM composition: chemical composition and particle size distribution. Further analysis and monitoring of IAQ in school classrooms are required to corroborate the results observed here.

While this study showed improvement in R^2 scores and low error values with ML models, it lacked the long-term analysis of sensors performance (using the developed models). One of the significant issues related to low-cost sensor devices is the calibration drift over time. Thus, in the future, the research campaigns for such a study will require long-term IAQ monitoring with the devices along with reference instruments and add sensor age as one of the calibration variables. Moreover, future work should also consider investigating how the developed models perform versus the constant need to recalibrate the sensors.

CRediT authorship contribution statement

H. Chojer: Software, Formal analysis, Data curation, Writing – original draft, Visualization. **P.T.B.S. Branco:** Methodology, Investigation, Writing – review & editing. **F.G. Martins:** Conceptualization, Writing – review & editing. **M.C.M. Alvim-Ferraz:** Conceptualization, Writing – review & editing. **S.I.V. Sousa:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors express immense gratitude to the school for their cooperation. This work was financially supported by LA/P/0045/2020 (ALiCE), UIDB/00511/2020 and UIDP/00511/2020 (LEPABE), funded by national funds through FCT/MCTES (PIDDAC); Project PTDC/EAM-AMB/32391/2017 - POCI-01-0145-FEDER-032391, funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds (PIDDAC) through FCT/MCTES; and Project 2SMART - engineered Smart materials for Smart citizens, with reference NORTE-01-0145-FEDER-000054, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). H. Chojer thanks the Portuguese Foundation for Science and Technology (FCT) for the individual research grant SFRH/BD/05092/2021.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2022.119251.

References

Amato, F., Rivas, I., Viana, M., Moreno, T., Bouso, L., Reche, C., et al., 2014. Sources of indoor and outdoor PM2.5 concentrations in primary schools. Sci. Total Environ. 490, 757–765. https://doi.org/10.1016/j.scitotenv.2014.05.051.

Barkjohn, K.K., Gantt, B., Clements, A.L., 2021. Development and application of a United States-wide correction for PM2.5 data collected with the PurpleAir sensor. Atmos. Meas. Tech. 14, 4617–4637. https://doi.org/10.5194/amt-14-4617-2021.

Beg, M., Taka, J., Kluyver, T., Konovalov, A., Ragan-Kelley, M., Thiery, N.M., Fangohr, H., 2021. Using jupyter for reproducible scientific workflows. Comput. Sci. Eng. 23, 36–46. https://doi.org/10.1109/2Fmcse.2021.3052101.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. In: A Training Algorithm for Optimal Margin Classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory. Association for Computing Machinery, Pittsburgh, Pennsylvania, USA, pp. 144–152.

Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2014a. Indoor air quality in urban nurseries at Porto city: particulate matter assessment. Atmos. Environ. 84, 133–143. https://doi.org/10.1016/j.atmosenv.2013.11.035.

Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2014b. The microenvironmental modelling approach to assess children's exposure to air pollution – a review. Environ. Res. 135, 317–332. https://doi.org/10.1016/j.envres.2014.10.002.

Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2019. Quantifying indoor air quality determinants in urban and rural nursery and primary schools. Environ. Res. 176, 108534 https://doi.org/10.1016/j.envres.2019.108534.

Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Ferraz, C., Vaz, L.G., Sousa, S.I.V., 2020a. Asthma in urban and rural pre- and primary schoolchildren according to the latest GINA definition. Allergy 75, 1771–1776. https://doi.org/10.1111/all.14201.

Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Ferraz, C., Vaz, L.G., Sousa, S.I.V., 2020b. Impact of indoor air pollution in nursery and primary schools on childhood asthma. Sci. Total Environ. 745, 140982 https://doi.org/10.1016/j.scitotenv.2020.140982.

Castell, N., Viana, M., MiC, Minguillón, Guerreiro, C., Querol, X., 2013. Real-world Application of New Sensor Technologies for Air Quality Monitoring. ETC/ACM Technical Paper. 16.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 https://doi.org/10.1145/1961189.1961199. Article 27.

Chojer, H., Branco, P.T.B.S., Martins, F.G., Alvim-Ferraz, M.C.M., Sousa, S.I.V., 2020. Development of low-cost indoor air quality monitoring devices: recent advancements. Sci. Total Environ. 727, 138385 https://doi.org/10.1016/j.scitotenv.2020.138385.

Drucker, H., Burges, C., Kaufman, L., Smola, A., Vapnik, V., 1996. In: Jordan, M.I., Petsche, T. (Eds.), Support Vector Regression Machines, vol. 96. NIPS, pp. 155–161.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.

Gehring, U., Gruzieva, O., Agius, R.M., Beelen, R., Custovic, A., Cyrys, J., et al., 2013. Air pollution exposure and lung function in children: the ESCAPE Project. Environ. Health Perspect. 121, 1357–1364. https://doi.org/10.1289/ehp.1306770.

Giordano, M.R., Malings, C., Pandis, S.N., Presto, A.A., McNeill, V.F., Westervelt, D.M., et al., 2021. From low-cost sensors to high-quality data: a summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors. J. Aerosol Sci., 105833 https://doi.org/10.1016/j.jaerosci.2021.105833.

Hagan, D.H., Kroll, J.H., 2020. Assessing the accuracy of low-cost optical particle sensors using a physics-based approach. Atmos. Meas. Tech. 13, 6343–6355. https://doi.org/10.5194/amt-13-6343-2020.

Hinds, W.C., 1999. Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles. John Wiley & Sons.

IQAir, 2021. Air Visual monitor. https://www.iqair.com/air-quality-monitors/airvisual-series. (Accessed 22 September 2021).

Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., Morawska, L., 2018. The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. Atmos. Meas. Tech. 11, 4883–4890. https://doi.org/10.5194/amt-11-4883-2018.

Jones, A.P., 1999. Indoor air quality and health. Atmos. Environ. 33, 4535–4564. https://doi.org/10.1016/S1352-2310(99)00272-1.

Klepeis, N.E., Nelson, W.C., Ow, R., Robinson, J.P., Tsang, A.M., Switzer, P., et al., 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J. Expo. Sci. Environ. Epidemiol. 11, 231–252. https://doi.org/10.1038/sj.jea.7500165.

Levy Zamora, M., Xiong, F., Gentner, D., Kerkez, B., Kohrman-Glaser, J., Koehler, K., 2019. Field and laboratory evaluations of the low-cost plantower particulate matter sensor. Environ. Sci. Technol. 53, 838–849. https://doi.org/10.1021/acs.est.8b05174.

Liu, D., Zhang, Q., Jiang, J., Chen, D.-R., 2017. Performance calibration of low-cost and portable particular matter (PM) sensors. J. Aerosol Sci. 112, 1–10. https://doi.org/10.1016/j.jaerosci.2017.05.011.

Magi, B.I., Cupini, C., Francis, J., Green, M., Hauser, C., 2020. Evaluation of PM2.5 measured in an urban setting using a low-cost optical particle counter and a Federal Equivalent Method Beta Attenuation Monitor. Aerosol. Sci. Technol. 54, 147–159. https://doi.org/10.1080/02786826.2019.1619915.

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P.K., Robinson, A.L., Presto, A.A., et al., 2020. Fine particle mass monitoring with low-cost sensors: corrections and long-term performance evaluation. Aerosol. Sci. Technol. 54, 160–174. https://doi.org/10.1080/02786826.2019.1623863.

McFarlane, C., Raheja, G., Malings, C., Appoh, E.K.E., Hughes, A.F., Westervelt, D.M., 2021. Application of Gaussian mixture regression for the correction of low cost PM2.5 monitoring data in Accra, Ghana. ACS Earth and Space Chemistry 5, 2268–2279. https://doi.org/10.1021/acsearthspacechem.1c00217.

Mendell, M.J., 2007. Indoor residential chemical emissions as risk factors for respiratory and allergic effects in children: a review. Indoor Air 17, 259–277. https://doi.org/10.1111/j.1600-0668.2007.00478.x.

Morawska, L., Thai, P.K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., et al., 2018. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone? Environ. Int. 116, 286–299. https://doi.org/10.1016/j.envint.2018.04.018.

Moreno, T., Rivas, I., Bouso, L., Viana, M., Jones, T., Àlvarez-Pedrerol, M., et al., 2014. Variations in school playground and classroom atmospheric particulate chemistry. Atmos. Environ. 91, 162–171. https://doi.org/10.1016/j.atmosenv.2014.03.066.

Mudway, I.S., Dundas, I., Wood, H.E., Marlin, N., Jamaludin, J.B., Bremner, S.A., et al., 2019. Impact of London's low emission zone on air quality and children's respiratory health: a sequential annual cross-sectional study. Lancet Public Health 4, e28–e40. https://doi.org/10.1016/S2468-2667(18)30202-0.

Nieuwenhuijsen, M., Paustenbach, D., Duarte-Davidson, R., 2006. New developments in exposure assessment: the impact on the practice of health risk assessment and epidemiological studies. Environ. Int. 32, 996–1009. https://doi.org/10.1016/j.envint.2006.06.015.

Nunes, R.A.O., Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2015. Particulate matter in rural and urban nursery schools in Portugal. Environ. Pollut. 202, 7–16. https://doi.org/10.1016/j.envpol.2015.03.009.

Nunes, R.A.O., Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2016. Gaseous pollutants on rural and urban nursery schools in Northern Portugal. Environ. Pollut. 208, 2–15. https://doi.org/10.1016/j.envpol.2015.07.018.

Ouimette, J.R., Malm, W.C., Schichtel, B.A., Sheridan, P.J., Andrews, E., Ogren, J.A., et al., 2021. Evaluating the PurpleAir monitor as an aerosol light scattering instrument. Atmos. Meas. Tech. Discuss. 1–35. https://doi.org/10.5194/amt-2021-170, 2021.

Parkinson, T., Parkinson, A., Dear, Rd, 2019. Continuous IEQ monitoring system: performance specifications and thermal comfort classification. Build. Environ. 149, 241–252. https://doi.org/10.1016/j.buildenv.2018.12.016.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peterson, P.J.D., Aujla, A., Grant, K.H., Brundle, A.G., Thompson, M.R., Vande Hey, J., et al., 2017. Practical use of metal oxide semiconductor gas sensors for measuring nitrogen dioxide and ozone in urban environments. Sensors 17. https://doi.org/10.3390/s17071653. Basel, Switzerland.

Pilgrim, M., Willison, S., 2009. Dive into. In: Python, 3. Springer.

PurpleAir, 2021. PurpleAir Monitor. https://www2.purpleair.com/products/purpleair-pa-ii?variant=40067691774049. (Accessed 22 September 2021).

Rai, A.C., Kumar, P., Pilla, F., Skouloudis, A.N., Di Sabatino, S., Ratti, C., et al., 2017. End-user perspective of low-cost sensors for outdoor air pollution monitoring. Sci. Total Environ. 607–608, 691–705. https://doi.org/10.1016/j.scitotenv.2017.06.266.

Rosário Filho, N.A., Urrutia-Pereira, M., D'Amato, G., Cecchi, L., Ansotegui, I.J., Galán, C., et al., 2021. Air pollution and indoor settings. World Allergy Organization Journal 14, 100499. https://doi.org/10.1016/j.waojou.2020.100499.

Sá, J.P., Branco, P.T.B.S., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2017. Evaluation of low-cost mitigation measures implemented to improve air quality in nursery and primary schools. Int. J. Environ. Res. Publ. Health 14. https://doi.org/10.3390/ijerph14060585.

Sá, J.P., Alvim-Ferraz, M.C.M., Martins, F.G., Sousa, S.I.V., 2022. Application of the low-cost sensing technology for indoor air quality monitoring: a review. Environ. Technol. Innovat. 28, 102551 https://doi.org/10.1016/j.eti.2022.102551.

Salimifard, P., Rim, D., Freihaut, J.D., 2020. Evaluation of low-cost optical particle counters for monitoring individual indoor aerosol sources. Aerosol. Sci. Technol. 54, 217–231. https://doi.org/10.1080/02786826.2019.1697423.

Schraufnagel, D.E., Balmes, J.R., Cowl, C.T., De Matteis, S., Jung, S.-H., Mortimer, K., et al., 2019. Air pollution and noncommunicable diseases: a review by the forum of international respiratory societies' environmental committee, Part 1: the damaging effects of air pollution. Chest 155, 409–416. https://doi.org/10.1016/j.chest.2018.10.042.

Schwartz, J., 2004. Air pollution and children's health. Pediatrics 113, 1037–1043.

Scikitlearn, 2021. Gradient Boosting Regressor. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html. (Accessed 24 March 2021).

Seabold, S., Statsmodels, Perktold J., 2010. Econometric and Statistical Modeling with python.

Singer, B.C., Delp, W.W., 2018. Response of consumer and research grade indoor air quality monitors to residential sources of fine particles. Indoor Air 28, 624–639. https://doi.org/10.1111/ina.12463.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S.W., et al., 2013. The changing paradigm of air pollution monitoring. Environ. Sci. Technol. 47, 11369–11377. https://doi.org/10.1021/es4022602.

Sousa, S.I.V., Ferraz, C., Alvim-Ferraz, M.C.M., Vaz, L.G., Marques, A.J., Martins, F.G., 2012. Indoor air pollution on nurseries and primary schools: impact on childhood asthma – study protocol. BMC Publ. Health 12, 435. https://doi.org/10.1186/1471-2458-12-435.

Thompson, J.E., 2016. Crowd-sourced air quality studies: a review of the literature & portable sensors. Trends in Environmental Analytical Chemistry 11, 23–34. https://doi.org/10.1016/j.teac.2016.06.001.

Tryner, J., Phillips, M., Quinn, C., Neymark, G., Wilson, A., Jathar, S.H., et al., 2021. Design and testing of a low-cost sensor and sampling platform for indoor air quality. Build. Environ. 206, 108398 https://doi.org/10.1016/j.buildenv.2021.108398.

uRAD, 2021. uRAD Monitor Model A3. https://www.uradmonitor.com/uradmonitor-model-a3/. (Accessed 22 September 2021).

Wang, Y., Boulic, M., Phipps, R., Chitty, C., Moses, A., Weyers, R., et al., 2017. Integrating Open-Source Technologies to Build a School Indoor Air Quality Monitoring Box (SKOMOBO). 4th Asia-Pacific World Congress on Computer Science and Engineering. APWC on CSE).

Wang, Z., Delp, W.W., Singer, B.C., 2020. Performance of low-cost indoor air quality monitors for PM2.5 and PM10 from residential sources. Build. Environ. 171, 106654 https://doi.org/10.1016/j.buildenv.2020.106654.

Weschler, C.J., Shields, H.C., Naik, D.V., 1989. Indoor ozone exposures. JAPCA 39, 1562–1568. https://doi.org/10.1080/08940630.1989.10466650.

White, R.M., Paprotny, I., Doering, F., Cascio, W.E., Solomon, P.A., Sensors, L.A.G., 2012. Sensors and apps for community-based atmospheric monitoring. Equip. Manag.: AIR AND WASTE MANAGEMENT ASSOCIATION'S MAGAZINE FOR ENVIRONMENTAL MANAGERS 36–40.

WHO, 2022. Air Pollution. World Health Organization. https://www.who.int/health-topics/air-pollution#tab=tab_1. (Accessed 21 February 2022).

Zamora, M.L., Rice, J., Koehler, K., 2020. One year evaluation of three low-cost PM2.5 monitors. Atmos. Environ. 235, 117615 https://doi.org/10.1016/j.atmosenv.2020.117615.

Zhang, L., Tian, F.-C., Peng, X.-W., Yin, X., 2014. A rapid discreteness correction scheme for reproducibility enhancement among a batch of MOS gas sensors. Sensor Actuator Phys. 205, 170–176. https://doi.org/10.1016/j.sna.2013.11.015.