Journal of Aerosol Science 157 (2021) 105809

Contents lists available at ScienceDirect

Journal of Aerosol Science

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

Evaluation of nine machine learning regression algorithms for calibration of low-cost $PM_{2.5}$ sensor

Vikas Kumar [a], Manoranjan Sahu [a,b,c,*]

[a] Aerosol and Nanoparticle Technology Laboratory, Environmental Science and Engineering Department, Indian Institute of Technology Bombay, Mumbai 400076, India
[b] Inter-Disciplinary Program in Climate Studies, Indian Institute of Technology Bombay, Mumbai 400076, India
[c] Centre for Machine Intelligence and Data Science, Indian Institute of Technology Bombay, Mumbai 400076, India

ARTICLE INFO

Keywords:
$PM_{2.5}$
Low-cost sensor
Machine learning
Calibration
Regression algorithims
Gradient boosting

ABSTRACT

Low-cost sensors (LCS) can construct a high spatial and temporal resolution $PM_{2.5}$ network but are affected by environmental parameters such as relative humidity and temperature. The data generated by LCS are inaccurate and require calibration against a reference instrument. This study has applied nine machine learning (ML) regression algorithms for Plantower PMS 5003 LCS calibration and compared their performance. The nine ML algorithms applied in this study are: (a) Multiple Linear Regression (MLR); (b) Lasso regression (L1); (c) Ridge regression (L2); (d) Support Vector Regression (SVR); (e) k- Nearest Neighbour (kNN); (f) Multilayer Perceptron (MLP); (g) Regression Tree (RT); (h) Random Forest (RF); (i) Gradient Boosting (GB). The comparison exhibits that kNN, RF and GB have the best performance out of all the algorithms with train scores of 0.99 and test scores of 0.97, 0.96 and 0.95 respectively. This study validates the capability of ML algorithms for the calibration of LCS.

1. Introduction

$PM_{2.5}$ exposure affects more humans globally than any other air pollutant and caused 4.2 million deaths, i.e., 7.6% of global deaths in 2015 and is the 5th leading risk factor for death (Forouzanfar et al., 2016; Cohen et al., 2017; WHO, 2018). $PM_{2.5}$ measurement is significant for the formulation of air pollution control measures, policies and frameworks to counter the potentially damaging effects on humans and climate change. For example, the National Clean Air Program (NCAP) launched by GoI to monitor and control PM emissions in Indian cities (Ganguly et al., 2020). There are two primary methods to measure $PM_{2.5}$ viz. Federal Reference Methods (FRM) and Federal Equivalent Methods (FEM) (Noble et al., 2001). FRM requires a gravimetric method where particle mass concentration is determined by weighing the filters before and after the sampling period as a measurement technique. FRM is the most accurate method and is widely used by regulatory bodies. However, there are certain disadvantages of the FRM, such as it is not a real-time form of measurement and only provides a 24-h average. FRM also has a high operation cost, has manual process involved and lacks portability (Ayers et al., 1999; Le et al., 2020; Noble et al., 2001). Beta attenuation monitor (BAM) and tapered element oscillating microbalance (TEOM) are two standard FEM. FEM have a higher temporal resolution, i.e., provide 1-h average and have relatively low operational cost compared to FRM but have high installation cost (Ayers et al., 1999; Chung et al., 2001; Le et al., 2020).

* Corresponding author. 507, Aerosol and Nanoparticle Technology Laboratory, Environmental Science and Engineering Department, Indian Institute of Technology, Bombay Powai, Mumbai, 400076, India.
E-mail address: mrsahu@iitb.ac.in (M. Sahu).

https://doi.org/10.1016/j.jaerosci.2021.105809
Received 30 January 2021; Received in revised form 1 May 2021; Accepted 4 May 2021
Available online 11 May 2021

$PM_{2.5}$ concentration is mostly organized spatiotemporally and thereof; the concentration and exposure can vary notably even in the nearby areas. The FRM and FEM cannot be installed in large numbers to create a dense network of high spatial resolution due to the lack of portability and high installation as well as operational cost. Low-cost sensors (LCS) can be a possible alternative and can be installed in large numbers to construct a high spatial and temporal resolution $PM_{2.5}$ network (Di Antonio et al., 2018; Bulot et al., 2020). LCS are cost-effective compared to FRM/FEM and can collect real-time data, i.e., minute resolution. LCS are compact, light-weight, portable and require low maintenance. However, there are certain shortcomings associated with using LCS which need to be addressed before being used extensively. LCS are based on laser light scattering (LLS) technology and uses a light beam to estimate the concentration based on the light scattered by particles passing through an air stream (Badura et al., 2018) instead of measuring their mass concentration or counting particles directly (He et al., 2020). Aerosol composition and size distributions, meteorological conditions such as relative humidity (RH) and temperature (T) can significantly influence the performance and accuracy of LCS (Rai et al., 2017; Badura et al., 2018; Zheng et al., 2018; Loh & Choi, 2019; Bai et al., 2020; Chu et al., 2020; Qin et al., 2020; Zusman et al., 2020). In high RH conditions, the aerosol particles absorb water and change the size, morphology, refractive index etc., known as hygroscopic growth (Lee et al., 2008), which regulates the amount of light scattered and so the particle concentration. The critical reason for RH and T's inclusion while calibrating the LCS with an FRM/FEM is the difference in operational RH and T of these instruments during measurement. FRM/FEM instruments generally operate at normal conditions (T: 20–23 °C, RH: 30–40%), while the LCS measures data in ambient conditions, which can cause disagreement in reported concentrations (Malings et al., 2020). The $PM_{2.5}$ concentration from the LCS alone is not adequate to explain the disparity in the LCS measurements (Zheng et al., 2018) and so RH and T are the crucial parameters for the LCS calibration (Lee et al., 2020). Also, the introduction of RH and T in the calibration model significantly improves the accuracy and can account for up to 17% and 7% variation in the $PM_{2.5}$ measurement reported by the low-cost sensor respectively which aids in achieving the highest accuracy possible (Chen et al., 2018; Gao et al., 2015; Jiang et al., 2021; Lee et al., 2020; Lin et al., 2018; Rai et al., 2017; Zheng et al., 2018).

Therefore, even though the cost of LCS is an advantage compared to FRM/FEM, the accuracy of concentration reported by the LCS is questionable (Li et al., 2020) and can vary concerning site/season due to change in environmental conditions. So, the LCS requires rigorous calibration against a reference instrument to account for RH and T's effect to meet regulatory instruments standards. Although LCS manufacturers develop calibration factors for the sensors but in laboratory conditions, which is not the same as the site where the sensor is installed and so the calibration factors become incompetent. So, it is vital to re-calibrate the sensors at the installation site to report accurate concentrations (Badura et al., 2018; Di Antonio et al., 2018; Bulot et al., 2020; Si et al., 2020). Besides, LCS may also require developing multiple calibration models specific to site or season and regular calibration to ensure accuracy of the particle concentration reported (Miskell et al., 2018; Zheng et al., 2018). Due to this, calibration of LCS manually can be hectic and is not a feasible option and some sophisticated models need to be developed. The calibration models have to be highly accurate and efficient in terms of time and computational power with the ability to handle large datasets (Wang et al., 2020). Machine learning can play a crucial role in this due to the wide range of algorithms for various kinds of datasets and applications.

So far, many statistical techniques such as Gaussian Process regression (GPR) and simple/multiple linear regression (Badura et al., 2019; Zheng et al., 2019; Si et al., 2020; Patra et al., 2021) have been implemented for calibrating the LCS with the reference instruments. A limited number of studies have also applied machine learning (ML) algorithms such as k-nearest neighbours (Loh & Choi, 2019), support vector machines (Loh & Choi, 2019; Wang et al., 2020), artificial neural network (Badura et al., 2019; Si et al., 2020), decision tree (Wijeratne et al., 2019), random forest (Loh & Choi, 2019; Wang et al., 2019, 2020) and gradient boosting (Johnson et al., 2018; Loh & Choi, 2019; Si et al., 2020). The problem with ML models is that they can suffer from overfitting. Overfitting is when the model learns the training data too well and provides a high train score but lower test score. Overfitted models are unable to perform accurately on new datasets and cannot be generalized. One way to overcome overfitting is to train multiple ML algorithms of a wide range and working principles and evaluate their performance to find the most appropriate model for the problem. No attempts have been made previously to apply a range of ML algorithms for LCS calibration.

The following objectives are addressed in this paper: (a) to investigate the performance of Plantower PMS 5003 LCS compared to Thermal Fisher Scientific SHARP model 5030 as a reference method for measurement of $PM_{2.5}$ at Alberta, Canada (Si, 2019) and (b) evaluate and compare the performance of nine ML algorithms for calibration of LCS. The nine ML algorithms applied in this study are: (a) multiple linear regression (MLR); (b) Lasso regression (L1); (c) Ridge regression (L2); (d) support vector regression (SVR); (e) k -nearest neighbour (kNN); (f) multilayer perceptron (MLP); (g) regression tree (RT); (h) random forest (RF); (i) gradient boosting (GB). The predicted values from all the algorithms are compared with the LCS and reference instrument measurements and each other.

2. Methodology

2.1. Data

The Plantower PMS 5003 LCS is evaluated in this study and measures $PM_{2.5}$, RH and T. The Thermal Fisher Scientific SHARP model 5030 is used as a reference instrument to evaluate the LCS. The data used in this study (Si, 2019) was collected at Calgary Region Airshed Zone (CRAZ) in Calgary, Alberta, Canada from December 7, 2018 to April 26, 2019 at an interval of 6s. The data was aggregated to derive the hourly concentrations. The exact sampling method, details of instruments, and the data pre-processing techniques applied can be found in Si et al. (2020).

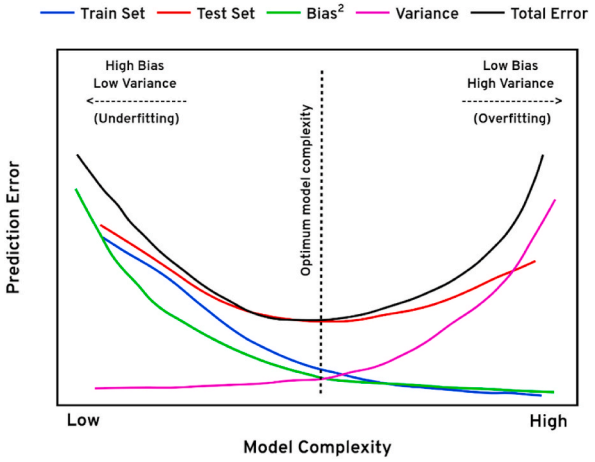

Fig. 1. Bias-variance dilemma (Fortmann-Roe, 2012; Hastie et al., 2009).

2.2. Algorithms

Regression is a powerful and fundamental statistical technique in machine learning and finds its application in economics, psychology, geography, and so forth (Sammut & Webb, 2011; Mendenhall & Sincich, 2014). Regression analysis is the study of dependence or relation between random variables of interest and infers mathematical functions to explain their behaviour, known as regression models (Rawlings et al., 1998; Mendenhall & Sincich, 2014). Regression analysis requires two types of real-valued variables, target/dependent and independent represented by y and x respectively. The objective of regression is to map a function such that $y = f(x) + \varepsilon$, where ε is the error (Kroese et al., 2019; Zaki & Meira, 2020). The cases in which there is more than one independent variable are known as multi-regression, where the equation transforms to $y = f(x_1, x_2, x_3, \ldots, x_n) + \varepsilon$, where $(x_1, x_2, x_3, \ldots, x_n)\ \varepsilon\ x$. In this study, we have three independent variables viz. $PM_{2.5}$, temperature, and relative humidity from the LCS represented by $PM_{2.5\_LCS}$, T, and RH respectively and one target variable, $PM_{2.5}$ from the reference instrument represented by $PM_{2.5\_REF}$. The regression algorithms applied in this study attempt to estimate functions to explain the effect of temperature and relative humidity on the measurement of $PM_{2.5}$ from the LCS to calibrate it with the reference instrument measurements. Mathematically, the objective function is $PM_{2.5\_REF} = f(PM_{2.5\_LCS}, Temp, RH) + \varepsilon$. Nine different ML regression algorithms were applied in this study which is discussed briefly below.

2.2.1. Linear models

Linear regression (LR) is the simplest regression model as it involves only one independent variable and assumes a linear function (straight line) between x and y (Bishop, 2006; Mendenhall & Sincich, 2014). The objective of LR is to fit a line ($y = w.\ x + b$, where b is the bias/intercept and w is the slope/weight) to a set of data points (x) to predict future values of y for given values of x, known as best fit line (Mendenhall & Sincich, 2014; Rawlings et al., 1998). The slope and intercept can be calculated using equations (1) and (2) respectively.

$$w = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2 - n(\bar{x})^2} \tag{1}$$

$$b = \bar{y} - w\,.\,\bar{x} \tag{2}$$

where $\bar{x}$ and $\bar{y}$ is the mean of x and y respectively and n is the sample size. The deviation between the actual (y_{act}) and predicted (y_{pred}) value of y is known as error. LR applies the least square error criterion to find out the best fit line. The best fit line is the one where the sum of the errors (SE) is zero and the sum of the squares of the errors (SSE) is minimum (Mendenhall & Sincich, 2014). SSE is the cost function (CF) of LR. SE and SSE can be calculated using equations (3) and (4) respectively.

$$SE = \sum_{i=1}^{n}\left(y_{pred_i} - y_{act_i}\right) = \sum_{i=1}^{n}\left(w.\,x_i + b - y_{act_i}\right) \tag{3}$$

$$SSE = \sum_{i=1}^{n}\left(y_{pred_i} - y_{act_i}\right)^2 = \sum_{i=1}^{n}\left(w.x_i + b - y_{act_i}\right)^2 \tag{4}$$

Multiple linear regression (MLR) is an extended version of LR and involves more than one independent variable. In MLR, the objective is to find the best-fit plane instead of line and the equation broadens to the form of $y = b + w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots + w_n x_n$, where $x_1, x_2, x_3, \ldots, x_n$ are the independent variables and (w_1, w_2, w_3, …w_n) ε w are the respective weights and w is the weight vector (Kroese et al., 2019; Mohri et al., 2018; Zaki & Meira, 2020). The method of fitting for MLR is identical to LR.

For regression, the trained model's error can be fragmented as the sum of error due to variance and error due to bias, known as bias-variance decomposition. Generally, mean square error (MSE) is used to quantify the prediction error which is equivalent to the sum of variance and squared bias, as seen in equation (5). The bias component describes how the average predicted value differs from the mean of actual values. In contrast, the variance component explains predicted values' inconsistency for a given data point from different models (Bishop, 2006; Fortmann-Roe, 2012; Hastie et al., 2009; Sammut & Webb, 2011).

$$MSE(x) = E\left[(f(x) - y)^2\right] = (E[f(x)] - y)^2 + E\left(f(x) - E\left((f(x) - E(f(x)))^2\right)\right) = bias^2 + variance \tag{5}$$

A high bias suggests that the model does not fit the data well and incorporates more assumptions about the target variable whereas a low bias means the model made fewer assumptions. A high variance value reflects random noise in the training data while a low value suggests predicted values close to each other. A high bias value causes low variance in order to reduce the model error according to equation (5), which leads to a phenomenon known as underfitting. In underfitting, the model cannot capture the underlying trend between dependent and independent variables and neither performs well on the training nor the test set, producing high prediction error for both. Increasing the model complexity reduces training error but too much training increases variance on the cost of bias causing overfitting in which the model adapts to the data too well. Overfitted models do not generalize well on other datasets and produce low training but comparatively larger test errors. There is a constant trade-off between bias and variance to find the optimum value in order to get the least model error possible, known as bias-variance dilemma as illustrated in Fig. 1 (Bishop, 2006; Fortmann-Roe, 2012; Hastie et al., 2009; Sammut & Webb, 2011).

Regularization is a technique to control overfitting which discourages the regression weights from reaching large values by decreasing the variance at the expense of increasing the bias slightly. Regularization adds a penalty term to the cost function depending on the weights' norm (Bishop, 2006; Flach, 2012; Hastie et al., 2009; Kroese et al., 2019). Two different regularization techniques are applied in this study Lasso (L1) and Ridge (L2) have subtle but significant differences. Lasso applies the L1 (Manhattan/Taxicab) norm while Ridge applies the L2 (Euclidean) norm. Due to this, L2 regression weights may remain small but still non-zero, while L1 favors sparse model and drives many weights to zero acting as a feature extraction method, especially in the case of multi-regression (Bishop, 2006; Flach, 2012; Hastie et al., 2009; Zaki & Meira, 2020). The regularized cost function's for L1 and L2 are in equations (6) and (7):

$$CF_{L1} = \sum_{i=1}^{n} \left(w.x_i + b - y_{act_i}\right)^2 + \lambda |w| \tag{6}$$

$$CF_{L2} = \sum_{i=1}^{n} \left(w.x_i + b - y_{act_i}\right)^2 + \lambda w^2 \tag{7}$$

where λ ($\geq$0) is the regularization constant. λ controls the trade-off between the regularization and SSE components of the regularized cost function. When $\lambda = 0$, there is no regularization and the model will have low bias and possibly high variance. However, if $\lambda \to \infty$, all the weights would tend to zero producing a low variance and high bias overfitted model. Varying the λ estimates the best balance between bias and variance for an optimal predictive model. A small positive value of λ always guarantees a solution (Bishop, 2006; Flach, 2012; Hastie et al., 2009; Mohri et al., 2018; Sammut & Webb, 2011). L1 and L2 are simply LR with a normed cost function and so the method of fitting remains the same. Further details about these techniques can be found in Rawlings et al. (1998), Bishop (2006), Hastie et al. (2009), Mendenhall and Sincich (2014).

Support Vector Regression (SVR) is a powerful and robust modeling technique in ML. SVR follows the same paradigm of finding a function that fits the training data well to minimize the prediction error as discussed in LR but with specific crucial differences. SVR attempts to construct a tube of width $\varepsilon > 0$ (user-specified) around the regression function and treats deviations outside the margin as noise. A small value of ε may lead to a tube that does not enclose the entire data, while a very high value would mean that outliers are the only points that define the regression equation and produce an insignificant prediction model. SVR tries to minimize the absolute error instead of SSE as in LR (Bishop, 2006; Gunn, 1998; Hastie et al., 2009; Mohri et al., 2018; Yang, 2019). SVR implements a loss function known as ε-insensitive loss defined as:

$$L_\varepsilon(y) = \begin{cases} 0 & for\ |w\ .\ x + b - y\ | < \varepsilon \\ |w\ .\ x + b - y| - \varepsilon & otherwise \end{cases} \tag{8}$$

Due to the ε-insensitive loss function, only the deviation of points outside the tube contributes to the final error. The deviation of points inside the tube is ignored in the optimization. The points on the border of the tube and outside are known as support vectors as they support the regression line. The ε -insensitive loss function reduces the risk of overfitting as large outliers have a restricted effect on the regression equation. Mathematically, the optimization problem for SVR can be written as:

$$\min_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} |y_i - (w.\ x_i + b)| \tag{9}$$

subject to the constraints:

$$\begin{cases} y_i - w.\ x_i - b \leq \varepsilon \\ w.\ x_i + b - y_i \leq \varepsilon \end{cases} \tag{10}$$

where $C > 0$ is the penalty parameter that controls the trade-off between amount up to which deviation greater than ε are tolerated and flatness of the regression model (tube) (Clarke et al., 2009; Gunn, 1998; Kuhn & Johnson, 2013; Witten & Frank, 2005; Yang, 2019). More details on the derivation and implementation of SVR can be found in Gunn (1998), Clarke et al. (2009), Kuhn and Johnson (2013) and Yang (2019).

2.2.2. k-Nearest Neighbour regression

k- Nearest Neighbour (kNN) is a simple and easy to implement learning-by-memorizing-based ML algorithm. kNN is developed on the assumption that similar things exist near each other and can be applied for regression and classification. kNN does not create any stringent abstractions about the data's underlying structure. Instead, it merely memorizes the training data and predicts the value for new instances based on the closest samples. kNN algorithm is a five-step process: (a) select distance metric; (b) select number of nearest neighbours ($k < n$); (c) compute distance from other data points to desired point; (d) sort the points in increasing order of distance; (e) compute the average of k nearest neighbours' responses. Euclidean distance is the most commonly used metric for regression, although some other metrics are available (Hastie et al., 2009; Kuhn & Johnson, 2013; Sammut & Webb, 2011; Yang, 2019). For a p dimensional and n sample size data, the Euclidean distance between two points with attribute values $a_1, a_2, a_3, \ldots, a_p$ and $b_1, b_2, b_3, \ldots, b_p$ can be calculated using equation (11) and the value of response variable for new instance can be calculated using equation (12):

$$Euclidean\ Distance = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2 + \ldots + (a_p - b_p)^2} \tag{11}$$

$$y_{new}(x_{new}) = \frac{1}{k} \sum_{i \in k\ (x_{new})} y_i \tag{12}$$

where $k\ (x_{new})$ is the set of k nearest neighbours of x_{new}. k decides the accuracy of the model and so choosing the optimal value of k is of critical significance. A minimal value of k pulls few data points for accurate estimation and becomes sensitive to noise in data accepting high variance in predictions that may lead to overfitting. While a very high value of k may include irrelevant data points that are not actual neighbours and increases the bias at the cost of variance, reducing the model's accuracy. kNN is also known as lazy algorithm as it is computationally expensive and significantly slower for large datasets since it needs to compute and sort the distances for the entire training dataset every time to find the nearest neighbours. This makes kNN inappropriate for scenarios where rapid predictions are required (Hastie et al., 2009; Kuhn & Johnson, 2013; Sammut & Webb, 2011; Yang, 2019).

2.2.3. Multilayer perceptron

Multilayer Perceptron (MLP) is a feed-forward (connections between nodes do no form loop) artificial neural network (ANN). The primary computing device of MLP is perceptron, a mathematical model of a neuron. MLP is a network of connected perceptrons organized in a layered formation. Each perceptron possesses an activation function and weights for each input. The objective of MLP is to learn these weights from the data such that the prediction accuracy is optimized. MLP consists of one input and output layer and can have more than one hidden layer.

The working mechanism of MLP is straightforward and can be summarized as a five-step process: (a) initialize the input data and random but small weights; (b) train the network in the forward direction and predict the output; (c) calculate the cost/error (MSE is used mainly for regression); (d) backpropagate the error and update the weights accordingly; (e) repeat the steps (b) and (c) until the error is marginal (Kroese et al., 2019; Kubat, 2017; Shalev-Shwartz & Ben-David, 2014; Yang, 2019; Zaki & Meira, 2020). MLP's are effective modeling techniques and can deal with several kinds of data. MLP falls into deep learning, a set of ML which deals with neural networks and is extensive. The mathematical derivations associated with training of the network and backpropagation of error are beyond the scope of this paper and can be found in Shalev-Shwartz and Ben-David (2014), Kubat (2017) and Yang (2019).

2.2.4. Tree-based models

Tree-based models are among the most popular ML models due to their iterative divide-and-conquer nature and have its root in the data structure. They are easy to implement and efficient but are computationally intensive. The tree-based model constructs a set of highly interpretable logical (if-then) conditions by recursively partitioning the decision space into smaller subspaces using training data and presents the decision process in the form of a tree graphically. They implicitly perform feature selection and can be applied for both regression and classification on datasets with large numbers of cases and/or variables (Flach, 2012; Hastie et al., 2009; Sammut & Webb, 2011; Yang, 2019).

There are numerous techniques for constructing a tree. However, here we follow the procedure of constructing the regression tree (RT) from one of the most utilized frameworks called classification and regression trees (CART) of Breiman et al. (1984). For a regression problem, the entire training dataset (D) is initially at the tree's root node and specific logical tests are conducted. The test partitions the data into two group (D_1 and D_2), one with the values satisfying the test and remaining to the other such that the overall SSE is minimized:

$$SSE = \sum_{i \in D_1} \left(y_i - \bar{y}_{D_1}\right)^2 + \sum_{j \in D_2} \left(y_j - \bar{y}_{D_2}\right)^2 \tag{13}$$

where $\bar{y}_{D_1}$ and $\bar{y}_{D_2}$ are the mean of the training set predictions for D_1 and D_2 respectively. The process is repeated until convergence

Table 1
Statistical performance parameters between $PM_{2.5\_REF}$ and $PM_{2.5\_LCS}$.

Parameter	$PM_{2.5\_REF}$ vs $PM_{2.5\_LCS}$
R^2	0.74
MSE	72.24
MAE	4.95
Slope	1.86 ± 0.02
Intercept	−2.38 ± 0.17

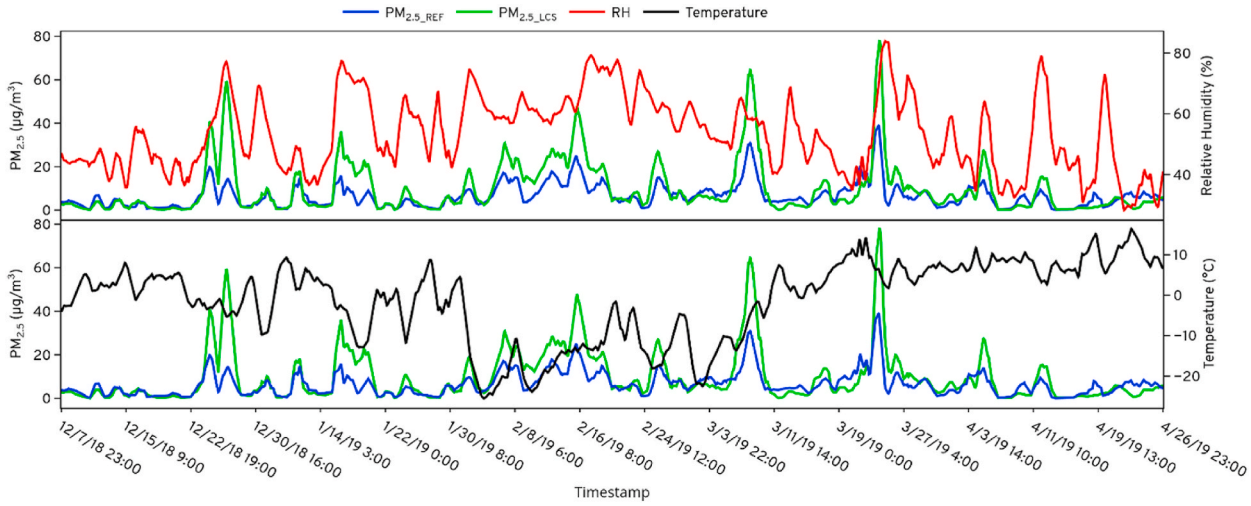

Fig. 2. Effect of relative humidity and temperature on LCS $PM_{2.5}$ measurements.

criteria are met and no more splits are possible. The last nodes of the tree are called decision nodes (or leaves) and contain the predicted value for the target variable (Bishop, 2006; Hastie et al., 2009; Kuhn & Johnson, 2013; Sammut & Webb, 2011; Yang, 2019).

Bagging (short for bootstrap aggregating) and boosting are two popular ensemble learning techniques (models that combine the output of multiple models) in ML. The main idea behind bagging and boosting is constructing multiple decision trees and combining their predictions by averaging (in regression) or voting (in classification) to reduce variance and increase prediction accuracy. The difference is that bagging creates individual trees and assigns equal weight to all trees. Contrary to that, in boosting the new trees are influenced by the previous ones' performance and assigns weight based on the trees' performance (Bishop, 2006; Hastie et al., 2009; Sutton, 2005; Yang, 2019). Two ensemble techniques used in this study viz. Random Forest (RF) and Gradient Boosting (GB) are extensions of bagging and boosting respectively. The detailed discussion of these algorithms is beyond the scope of this paper but can be seen in Sutton (2005), Hastie et al. (2009) and Yang (2019).

3. Results and discussion

3.1. Evaluation of LCS measurements

LCS measurements under ambient conditions are affected by high relative humidity (RH) and sensitive to temperature (T). The comparative analysis of $PM_{2.5}$ measurements of LCS ($PM_{2.5\_LCS}$) and reference instrument ($PM_{2.5\_REF}$) indicates a high variation in measured values, sometimes almost double when the value is above 10 µg/m^3. For values less than that, the values are closer comparatively. The mean(±sd), minimum and maximum values measured by reference instrument are 6.57(±5.57), 0.00 and 38.87 respectively, while for LCS the values are 9.89(±12.09), 0.03 and 77.8. The statistical performance parameters between $PM_{2.5\_REF}$ and $PM_{2.5\_LCS}$ are calculated and presented in Table 1.

This drift produced in the LCS measurements is due to the variation in RH or T and sometimes a combination of both, as shown in Fig. 2. The reference instrument's measurements are not affected by RH and T due to a drying system. Some other environmental parameters might also be affecting the LCS measurements but we are ignoring them in this study due to lack of data. The objective is to calibrate the $PM_{2.5\_LCS}$ with the $PM_{2.5\_REF}$ that incorporates RH and T's effect and reduce the measurements' variance. In this study, we have applied nine ML regression algorithms to calibrate the LCS.

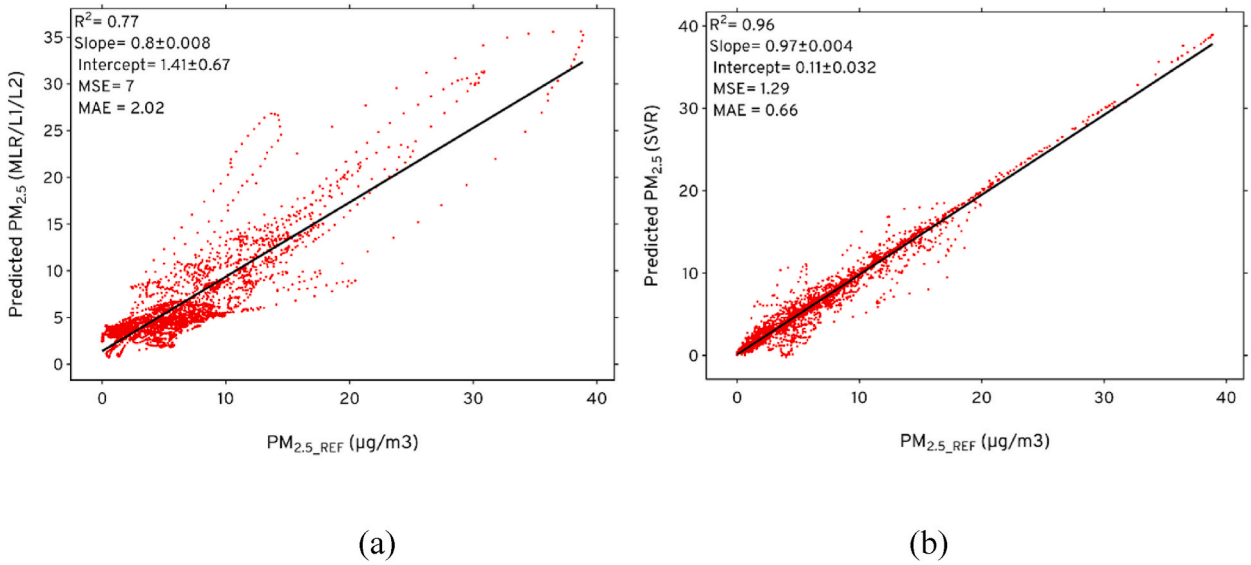

Fig. 3. Scatter plot of $PM_{2.5\_REF}$ and values predicted from (a) MLR/L1/L2, (b) SVR.

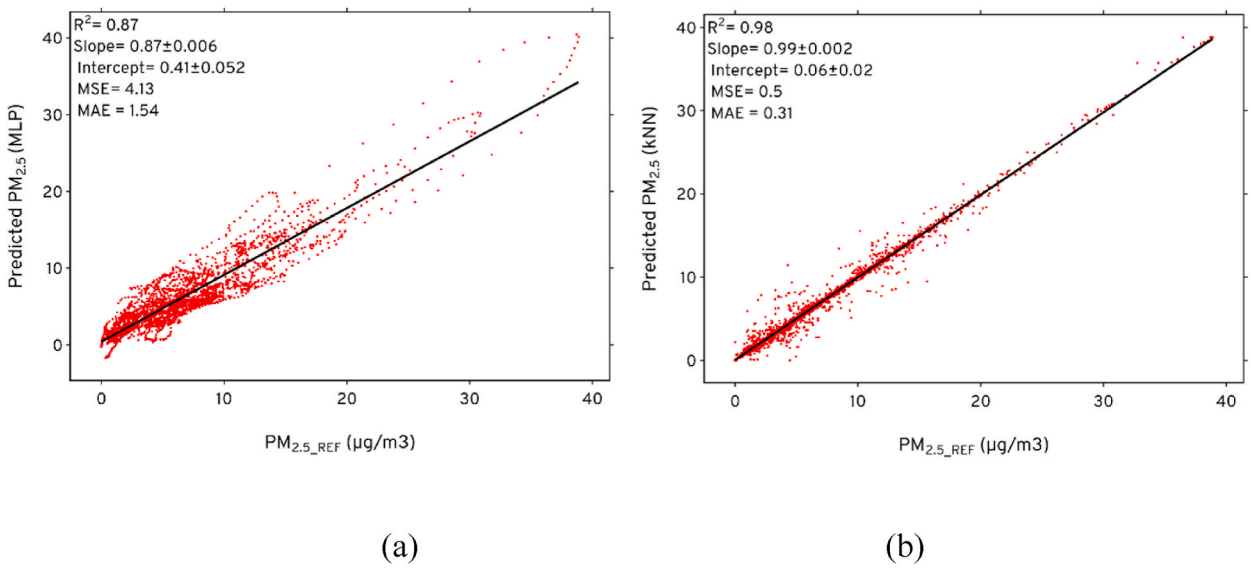

Fig. 4. Scatter plot of $PM_{2.5\_REF}$ and values predicted from (a) MLP, (b) kNN.

3.2. Calibration results

To understand the regression algorithms' capabilities applied in this study, the trained models need to be tested on new data (unavailable to the model) to check the model's performance. The entire dataset was split into two parts: train and test (70/30). The entire pipeline is divided into two steps: parameter optimization and training/prediction. In parameter optimization, a grid search was performed to infer appropriate parameters for the data with ten cross-fold validation for all the models simultaneously. The second step comprises applying the parameters derived from parameter optimization and comparing the results obtained by the models, which is discussed below.

3.2.1. Linear models

L1 and L2 models are producing the same result as MLR. A regularization constant (λ) of 0.01 and 0.1 was applied in L1 and L2 models respectively inferred through grid search. SVR model was trained on $\varepsilon = 0.2$ and penalty parameter (C) = 1. Out of four linear models, MLR, L1 and L2 have a low train (0.78) and test score (0.75) and only SVR is producing a high train (0.97) and test score (0.94). The linear models except for SVR, i.e., MLR/L1/L2, have reduced the mean (6.63 ± 5.03) and maximum (35.62) values of LCS

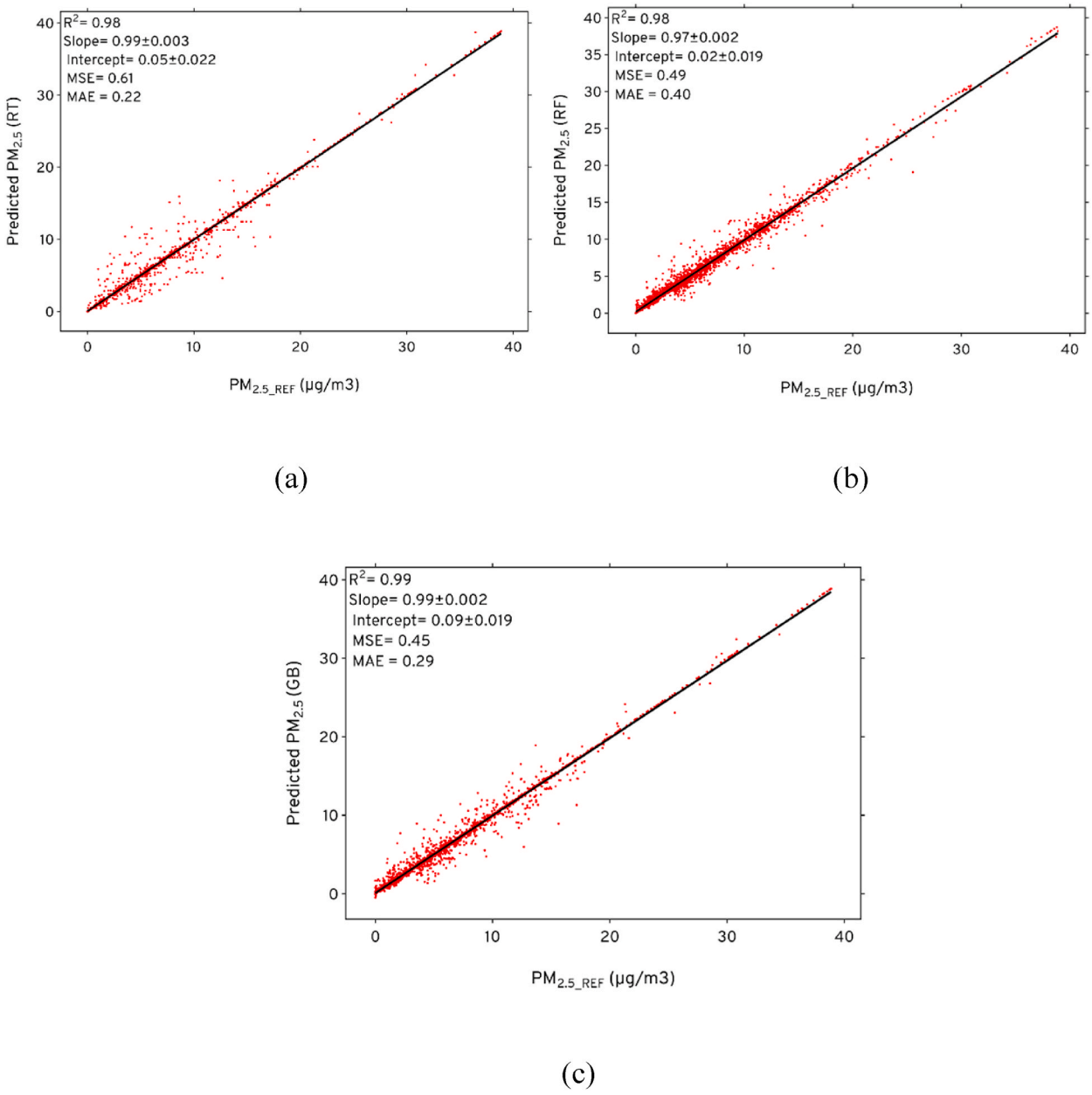

Fig. 5. Scatter plot of $PM_{2.5\_REF}$ and values predicted from (a) RT, (b) RF, (c) GB.

measurements close to the range of reference instrument measurements but are unable to provide an accurate prediction as can be seen in Fig. 3(a). Along with the train and test scores being low for these models, the MSE (7.00) and MAE (2.02) is low compared to actual values but is high compared to other models and vice-versa for R^2 (0.77). The weights of $PM_{2.5\_LCS}$, RH and T are 0.46, −0.11 and −0.03 respectively, and the bias is 7.94.

This is a case of under-fitting where the model cannot completely identify the underlying relation between independent and dependent variables distribution. The regularization techniques also cannot handle the complexity of the problem and do not show any improvement compared to MLR. On the other hand, SVR exhibits high train and test scores and a higher R^2 between actual and predicted values than its siblings. But higher MSE (1.29) and MAE (0.66) sometimes almost double compared to the values of other algorithms such as kNN (0.31) and GB (0.29). SVR can also fit the line to the data without getting sensitive to outliers, as illustrated in Fig. 3(b).

3.2.2. Multilayer perceptron

MLP model's train (0.87) and test (0.85) scores are higher than MLR, L1 and L2 but lower than SVR. The MLP was trained with three

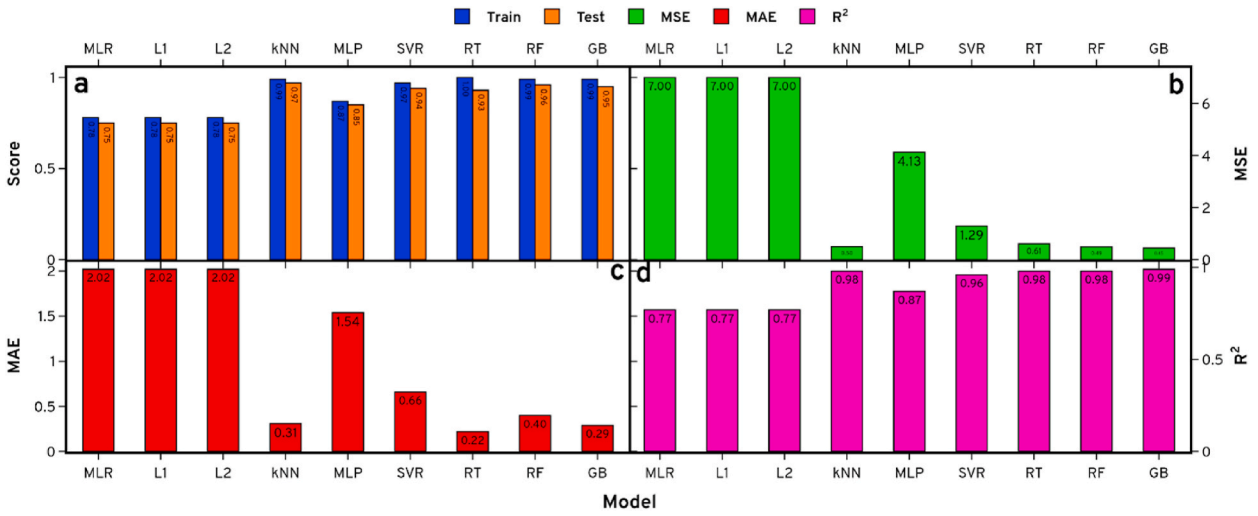

Fig. 6. Comparison of (a) train and test scores, (b) MSE, (c) MAE, (d) R^2 for the nine ML regression algorithms.

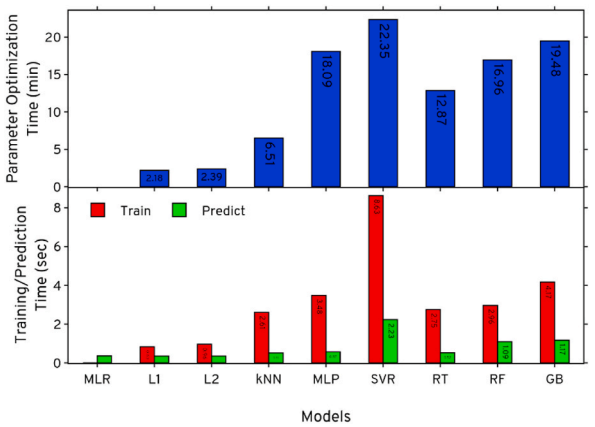

Fig. 7. Comparison of computation time for the algorithms.

hidden layers of size five each and tanh as activation function and some other peripheral parameters. The MSE (4.13) and MAE (1.54) between the model's predicted and actual values are relatively high compared to some models. The R^2 (0.87) value is significantly low compared to other algorithms. The predicted data is still susceptible to outliers, as shown in the scatter plot in Fig. 4(a). MLP can be extremely powerful if trained regressively and more accurate models could be produced. We acknowledge that the model could not be trained with more hidden and deep layers due to computational competency.

3.2.3. k- Nearest Neighbour

kNN is building a very accurate model with train and test scores as high as 0.99 and 0.97. The kNN model was trained for k = 2. The model's accuracy decreases on increasing the value of k. The MSE, MAE and R^2 values for kNN are 0.5, 031 and 0.98, respectively. The model's prediction and actual values can be fitted by a line that explains their relationship without getting affected by outliers/noise in data, as illustrated in Fig. 4(b).

3.2.4. Tree models

All three tree-based models produce quality results with train scores of 1.00, 0.99 and 0.99 and test scores of 0.93, 0.96, and 0.95 for RT, RF and GB, respectively. RT, RF and GB are trained with trees of maximum depth 28, 15 and 5 respectively. RT is producing the best train score. However, the difference between train and test scores for RT (0.07) is higher compared to RF (0.03) and GB (0.04). RT is capable of overfitting as they apply greedy algorithms due to which sometimes the optimal tree cannot be found. This is a case of overfitting. The RT model is learning the training data too well and cannot be generalized. RF and GB are capable of fixing the overfitting issue of RT, as seen in this case. The scatter plots for RT, RF and GB can be seen in Fig. 5(a), (b) and (c) respectively. The data points for RF and GB are closer to the fit line compared to RT. The MSE (0.61) for RT is higher compared to RF (0.49) and GB (0.45) but MAE (0.22) is lower than RF (0.4) and GB (0.29). The R^2 value is in the same range. RF and GB are providing the best results on all

Table 2
Comparison with previous studies.

Model	References	R^2	MSE	MAE
MLR	Johnson et al. (2018)	0.44	10.98	–
	Badura et al. (2019)	0.85	19.68	–
	Si et al. (2020)	0.6	21.61	3.09
	Wang et al. (2020)	0.76	47.33	–
	This study	0.77	7	2.02
L2	Johnson et al. (2018)	0.45	10.78	–
	This study	0.77	7	2.02
SVR	Loh and Choi (2019)	0.76	31.36	–
	Wang et al. (2020)	0.94	27.3	–
	This study	0.96	1.29	0.66
kNN	Loh and Choi (2019)	0.78	29.16	–
	This study	0.98	0.5	0.31
MLP/ANN	Badura et al. (2019)	0.79	26.62	–
	Si et al. (2020)	0.67	17.61	2.63
	This study	0.87	4.13	1.54
RT	Wijeratne et al. (2019)	0.99	–	–
	This study	0.98	0.61	0.22
RF	Loh and Choi (2019)	0.8	27.04	–
	Wang et al. (2019)	0.98	–	–
	Wang et al. (2020)	0.94	10.34	–
	This study	0.98	0.49	0.4
GB	Johnson et al. (2018)	0.72	5.51	–
	Loh and Choi (2019)	0.82	24.5	–
	Si et al. (2020)	0.72	15.26	2.38
	This study	0.99	0.45	0.29

parameters.

A comparison of the train and test scores (R^2) for the models is presented in Fig. 6(a). The MSE, MAE and R^2 between actual ($PM_{2.5\_REF}$) and values predicted from the models were compared and are presented in Fig. 6(b), (c), and (d) respectively. The MSE is a measurement of goodness of fit. Lower MSE means a lower error or better fit. The overall evaluation and comparison of all the nine regression algorithms exhibit that kNN, RF and GB are the best-suited LCS calibration models.

3.2.5. Comparison of computation time

Another predominant parameter that dictates the real-time application of a calibration methodology is the time taken to train the calibration model. Since the calibration methods applied in this study are in two stages, the computation time is also segregated accordingly and is presented in Fig. 7. The less complex computational models have their parameter optimization time of less than 10 min. For more complex high-end ML algorithms, the parameter optimization time is around 20 min, with SVR (22.35 min) taking the most time. The training time is closer to 5 s for almost all the algorithms except SVR taking 8.63 s. The prediction time is less than 2.5 s for all the algorithms. The evaluation of computation time displays that none of the high-end ML algorithms that produce accurate results have an advantage concerning computational time for this dataset. However, the computational time can depend based on the size of the dataset used and the computational power available to the modeler.

3.2.6. Comparison with previous studies

The results obtained on applying the algorithms are compared with previous studies, which have applied the respective algorithms and are presented in Table 2. It is clear from the comparison that the less complex computation models provide better results than Si et al. (2020) but are still underfitting. On the contrary, the high-end ML algorithms produce highly accurate results in this study, similar to the previous studies. Differences in hyperparameter optimization technique and train test split ratio could be the reason for variation in the performance of the algorithms compared to Si et al. (2020). The primary criteria that affect the algorithms' performance are correct parameters for the respective algorithms and the selection of essential variables to achieve the best results. We acknowledge that this study applied the algorithms on a particular dataset measured at a specific location. It would be interesting to investigate these algorithms' performance for multi-location and seasonal datasets and compare their performance, which will be conducted in future studies.

4. Conclusion

Nine machine learning (ML) regression algorithms have been examined to calibrate low-cost sensors (LCS). This was demonstrated on the dataset collected from a Plantower PMS 5003 LCS and Thermal Fisher Scientific SHARP model 5030 as a reference instrument. Multiple Linear (MLR), Lasso (L1) and Ridge (L2) regression models giving the same result and are underfitting the data while the Regression Tree (RT) is overfitting. Support Vector Regression (SVR) and Multilayer Perceptron (MLP) are giving good results compared to MLR, L1 and L2 but can be improved with increased computational capability. K- Nearest Neighbour (kNN), Random Forest (RF) and Gradient Boosting (GB) provide highly accurate results. kNN, RF and GB have the best performance out of all the

algorithms with train scores of 0.99 and test scores of 0.97, 0.96 and 0.95 respectively. The overall evaluation and comparison demonstrate that kNN, RF and GB are the best-suited models for the calibration of LCS.

Data availability

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.3473833 (Si, 2019).

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.

Acknowledgments

This work has been supported by the Central Pollution Control Board as a part of the study "Pilot Study for Assessment of Reducing Particulate Air Pollution in Urban Areas by Using Air Cleaning System (sometimes called as Smog Tower)" (Grant no: RD/0120-CPCB000- 001). Partial support from the study "Application of Nanoparticles in ESP for Inactivation of Microorganisms and Degradation of VOCs for Air Purification" (Grant no: RD/0119- DST0000-048) is acknowledged.

References

Ayers, G. P., Keywood, M. D., & Gras, J. L. (1999). TEOM vs. manual gravimetric methods for determination of PM2.5 aerosol mass concentrations. *Atmospheric Environment, 33*(22), 3717–3721. https://doi.org/10.1016/S1352-2310(99)00125-9

Badura, M., Batog, P., Drzeniecka-Osiadacz, A., & Modzel, P. (2018). Evaluation of low-cost sensors for ambient PM2.5 monitoring. *Journal of Sensors*, 1–16. https://doi.org/10.1155/2018/5096540

Badura, M., Batog, P., Drzeniecka-Osiadacz, A., & Modzel, P. (2019). Regression methods in the calibration of low-cost sensors for ambient particulate matter measurements. *SN Applied Sciences, 1*(6). https://doi.org/10.1007/s42452-019-0630-1

Bai, L., Huang, L., Wang, Z., Ying, Q., Zheng, J., Shi, X., & Hu, J. (2020). Long-term field evaluation of low-cost particulate matter sensors in nanjing. *Aerosol and Air Quality Research, 20*(2), 242–253. https://doi.org/10.4209/aaqr.2018.11.0424

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. CRC press.

Bulot, F. M. J., Russell, H. S., Rezaei, M., Johnson, M. S., Ossont, S. J. J., Morris, A. K. R., Basford, P. J., Easton, N. H. C., Foster, G. L., Loxham, M., & Cox, S. J. (2020). Laboratory comparison of low-cost particulate matter sensors to measure transient events of pollution. *Sensors, 20*(8). https://doi.org/10.3390/s20082219

Chen, C.-C., Kuo, C.-T., Chen, S.-Y., Lin, C.-H., Chue, J.-J., Hsieh, Y.-J., Cheng, C.-W., Wu, C.-M., & Huang, C.-M. (2018). *Calibration of low-cost particle sensors by using machine-learning method. 2018 IEEE asia pacific Conference on Circuits and systems (APCCAS)* (pp. 111–114). https://doi.org/10.1109/apccas.2018.8605619

Chu, H.-J., Ali, M. Z., & He, Y.-C. (2020). Spatial calibration and PM2.5 mapping of low-cost air quality sensors. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-79064-w

Chung, A., Chang, D. P. Y., Kleeman, M. J., Perry, K. D., Cahill, T. A., Dutcher, D., McDougall, E. M., & Stroud, K. (2001). Comparison of real-time instruments used to monitor airborne particulate matter. *Journal of the Air & Waste Management Association, 51*(1), 109–120. https://doi.org/10.1080/10473289.2001.10464254

Clarke, B., Fokoue, E., & Zhang, H. H. (2009). *Principles and theory for data mining and machine learning*. New York: Ny Springer. New York.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., & Pope, C. A. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the global burden of diseases study 2015. *The Lancet, 389*(10082), 1907–1918. https://doi.org/10.1016/s0140-6736(17)30505-6

Di Antonio, A., Popoola, O., Ouyang, B., Saffell, J., & Jones, R. (2018). Developing a relative humidity correction for low-cost sensors measuring ambient particulate matter. *Sensors, 18*(9), 2790. https://doi.org/10.3390/s18092790

Flach, P. (2012). *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press.

Forouzanfar, M. H., Afshin, A., Alexander, L. T., Anderson, H. R., Bhutta, Z. A., Biryukov, S., Brauer, M., Burnett, R., Cercy, K., Charlson, F. J., Cohen, A. J., Dandona, L., Estep, K., Ferrari, A. J., Frostad, J. J., Fullman, N., Gething, P. W., Godwin, W. W., Griswold, M., & Hay, S. I. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the global burden of disease study 2015. *The Lancet, 388*(10053), 1659–1724. https://doi.org/10.1016/s0140-6736(16)31679-8

Fortmann-Roe, S. (2012). *Understanding the bias-variance tradeoff*. Retrieved from https://scott.fortmann-roe.com/docs/BiasVariance.html/. (Accessed 3 January 2021).

Ganguly, T., Selvaraj, K. L., & Guttikunda, S. K. (2020). National Clean Air Programme (NCAP) for Indian cities: Review and outlook of clean air action plans. *Atmospheric Environment X, 8*, 100096. https://doi.org/10.1016/j.aeaoa.2020.100096

Gao, M., Cao, J., & Seto, E. (2015). A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *Environmental Pollution, 199*, 56–65. https://doi.org/10.1016/j.envpol.2015.01.013

Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS Technical Report, 14*(1), 5–16.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. In *Data mining, inference, and prediction* (2nd ed.). Springer.

He, M., Kuerbanjiang, N., & Dhaniyala, S. (2020). Performance characteristics of the low-cost Plantower PMS optical sensor. *Aerosol Science and Technology, 54*(2), 232–241. https://doi.org/10.1080/02786826.2019.1696015

Jiang, Y., Zhu, X., Chen, C., Ge, Y., Wang, W., Zhao, Z., Cai, J., & Kan, H. (2021). On-field test and data calibration of a low-cost sensor for fine particles exposure assessment. *Ecotoxicology and Environmental Safety, 211*, 111958. https://doi.org/10.1016/j.ecoenv.2021.111958

Johnson, N. E., Bonczak, B., & Kontokosta, C. E. (2018). Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment. *Atmospheric Environment, 184*, 9–16. https://doi.org/10.1016/j.atmosenv.2018.04.019

Kroese, D. P., Botev, Z. I., Taimre, T., & Vaisman, R. (2019). *Data science and machine learning: Mathematical and statistical methods*. CRC Press, Taylor & Francis Group.

Kubat, M. (2017). *An introduction to machine learning*. Springer International Publishing.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Ny Springer.

Lee, H., Kang, J., Kim, S., Im, Y., Yoo, S., & Lee, D. (2020). Long-term evaluation and calibration of low-cost particulate matter (PM) sensor. *Sensors, 20*(13), 3617. https://doi.org/10.3390/s20133617

Lee, A. K. Y., Ling, T. Y., & Chan, C. K. (2008). Understanding hygroscopic growth and phase transformation of aerosols using single particle Raman spectroscopy in an electrodynamic balance. *Faraday Discussions, 137*, 245–263. https://doi.org/10.1039/b704580h

Le, T. C., Shukla, K. K., Chen, Y. T., Chang, S. C., Lin, T. Y., Li, Z., Pui, D. Y. H., & Tsai, C. J. (2020). On the concentration differences between PM2.5 FEM monitors and FRM samplers. *Atmospheric Environment, 222*, 117138. https://doi.org/10.1016/j.atmosenv.2019.117138

Li, J., Mattewal, S. K., Patel, S., & Biswas, P. (2020). Evaluation of nine low-cost-sensor-based particulate matter monitors. *Aerosol and Air Quality Research, 20*(2), 254–270. https://doi.org/10.4209/aaqr.2018.12.0485

Lin, Y., Dong, W., & Chen, Y. (2018). Calibrating low-cost sensors by a two-phase learning approach for urban air quality measurement. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(1), 1–18. https://doi.org/10.1145/3191750

Loh, B. G., & Choi, G. H. (2019). Calibration of portable particulate matter–monitoring device using web query and machine learning. *Safety and Health at Work, 10*(4), 452–460. https://doi.org/10.1016/j.shaw.2019.08.002

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., & Subramanian, R. (2020). Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation. *Aerosol Science and Technology, 54*(2), 1–15. https://doi.org/10.1080/02786826.2019.1623863

Mendenhall, W., & Sincich, T. (2014). *A second course in statistics: Regression analysis*. Pearson.

Miskell, G., Salmond, J. A., & Williams, D. E. (2018). Solution to the problem of calibration of low-cost air quality measurement sensors in networks. *ACS Sensors, 3*(4), 832–843. https://doi.org/10.1021/acssensors.8b00074

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT Press.

Noble, C. A., Vanderpool, R. W., Peters, T. M., McElroy, F. F., Gemmill, D. B., & Wiener, R. W. (2001). Federal reference and equivalent methods for measuring fine particulate matter. *Aerosol Science and Technology, 34*(5), 457–464. https://doi.org/10.1080/02786820121582

Patra, S. S., Ramsisaria, R., Du, R., Wu, T., & Boor, B. E. (2021). A machine learning field calibration method for improving the performance of low-cost particle sensors. *Building and Environment, 190*, 107457. https://doi.org/10.1016/j.buildenv.2020.107457

Qin, X., Hou, L., Gao, J., & Si, S. (2020). The evaluation and optimization of calibration methods for low-cost particulate matter sensors: Inter-comparison between fixed and mobile methods. *The Science of the Total Environment, 715*, 136791. https://doi.org/10.1016/j.scitotenv.2020.136791

Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., Yasar, A., & Rickerby, D. (2017). End-user perspective of low-cost sensors for outdoor air pollution monitoring. *The Science of the Total Environment, 607–608*, 691–705. https://doi.org/10.1016/j.scitotenv.2017.06.266

Rawlings, J. O., Dickey, D. A., & Pantula, S. G. (1998). *Applied regression analysis: A research tool*. New York: Ny Springer.

Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.

Si, M. (2019). *Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine learning methods (version v0) [data set]. Zenodo*. http://doi.org/10.5281/zenodo.3473833.

Si, M., Xiong, Y., Du, S., & Du, K. (2020). Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods. *Atmospheric Measurement Techniques, 13*(4), 1693–1707. https://doi.org/10.5194/amt-13-1693-2020

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of Statistics, 24*, 303–329. https://doi.org/10.1016/s0169-7161(04)24011-1

Wang, Y., Du, Y., Wang, J., & Li, T. (2019). Calibration of a low-cost PM2.5 monitor using a random forest model. *Environment International, 133*, 105161. https://doi.org/10.1016/j.envint.2019.105161

Wang, W. C. V., Lung, S. C. C., & Liu, C. H. (2020). Application of machine learning for the in-field correction of a PM2.5 low-cost sensor network. *Sensors, 20*(17), 5002. https://doi.org/10.3390/s20175002

Wijeratne, L. O. H., Kiv, D. R., Aker, A. R., Talebi, S., & Lary, D. J. (2019). Using machine learning for the calibration of airborne particulate sensors. *Sensors, 20*(1), 99. https://doi.org/10.3390/s20010099

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufman.

World Health Organization (WHO). (2018). *Ambient (outdoor) air quality and health*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health/. (Accessed 15 January 2021).

Yang, X. S. (2019). *Introduction to algorithms for data mining and machine learning*. Elsevier.

Zaki, M. J., & Meira, W. (2020). *Data mining and machine learning: Fundamental concepts and algorithms*. Cambridge University Press.

Zheng, T., Bergin, M. H., Johnson, K. K., Tripathi, S. N., Shirodkar, S., Landis, M. S., Sutaria, R., & Carlson, D. E. (2018). Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments. *Atmospheric Measurement Techniques, 11*(8), 4823–4846. https://doi.org/10.5194/amt-11-4823-2018

Zheng, T., Bergin, M. H., Sutaria, R., Tripathi, S. N., Caldow, R., & Carlson, D. E. (2019). Gaussian process regression model for dynamically calibrating and surveilling a wireless low-cost particulate matter sensor network in Delhi. *Atmospheric Measurement Techniques, 12*(9), 5161–5181. https://doi.org/10.5194/amt-12-5161-2019

Zusman, M., Schumacher, C. S., Gassett, A. J., Spalt, E. W., Austin, E., Larson, T. V., Carvlin, G., Seto, E., Kaufman, J. D., & Sheppard, L. (2020). Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environment International, 134*, 105329. https://doi.org/10.1016/j.envint.2019.105329