

## LETTER

# Applying machine learning for large scale field calibration of low-cost PM<sub>2.5</sub> and PM<sub>10</sub> air pollution sensors

Priscilla Adong  | Engineer Bainomugisha | Deo Okure | Richard Sserunjogi

AirQo, Department of Computer Science, College of Computing and Information Sciences, Makerere University, Kampala, Uganda

**Correspondence**

Priscilla Adong, AirQo, Department of Computer Science, College of Computing and Information Sciences, Makerere University, Kampala, Uganda.  
Email: [priscilla@airqo.net](mailto:priscilla@airqo.net)

**Funding information**

Belgium through the Wehubit programme implemented by Enabel; Engineering and Physical Sciences Research Council / Global Challenges Research Fund; Google.org

**Abstract**

Low-cost air quality monitoring networks can potentially increase the availability of high-resolution monitoring to inform analytic and evidence-informed approaches to better manage air quality. This is particularly relevant in low and middle-income settings where access to traditional reference-grade monitoring networks remains a challenge. However, low-cost air quality sensors are impacted by ambient conditions which could lead to over- or underestimation of pollution concentrations and thus require field calibration to improve their accuracy and reliability. In this paper, we demonstrate the feasibility of using machine learning methods for large-scale calibration of AirQo sensors, low-cost PM sensors custom-designed for and deployed in Sub-Saharan urban settings. The performance of various machine learning methods is assessed by comparing model corrected PM using  $k$ -nearest neighbours, support vector regression, multivariate linear regression, ridge regression, lasso regression, elastic net regression, XGBoost, multilayer perceptron, random forest and gradient boosting with collocated reference PM concentrations from a Beta Attenuation Monitor (BAM). To this end, random forest and lasso regression models were superior for PM<sub>2.5</sub> and PM<sub>10</sub> calibration, respectively. Employing the random forest model decreased RMSE of raw data from 18.6  $\mu\text{g}/\text{m}^3$  to 7.2  $\mu\text{g}/\text{m}^3$  with an average BAM PM<sub>2.5</sub> concentration of 37.8  $\mu\text{g}/\text{m}^3$  while the lasso regression model decreased RMSE from 13.4  $\mu\text{g}/\text{m}^3$  to 7.9  $\mu\text{g}/\text{m}^3$  with an average BAM PM<sub>10</sub> concentration of 51.1  $\mu\text{g}/\text{m}^3$ . We validate our models through cross-unit and cross-site validation, allowing analysis of AirQo devices' consistency. The resulting calibration models were deployed to the entire large-scale air quality monitoring network consisting of over 120 AirQo devices, which demonstrates the use of machine learning systems to address practical challenges in a developing world setting.

**KEYWORDS**

air pollution, field calibration, low-cost sensors, machine learning applications, PM<sub>10</sub>, PM<sub>2.5</sub>

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Applied AI Letters* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Ambient air pollution is a major environmental health risk in cities all over the world with harmful effects on human health and the ecosystem.<sup>1,2</sup> According to the World Health Organisation (WHO),<sup>3</sup> ambient air pollution (AAP) causes 4.2 million deaths per year due to stroke, heart disease, lung cancer and chronic respiratory diseases. Populations in cities in low and middle-income countries (LMICs) are among those at high risk of exposures to levels that greatly exceed the WHO health guidelines.<sup>4-7</sup> While there is increasing evidence on the effects of ambient air pollution on human health, many cities in LMICs lack reliable high-resolution and long-term datasets to quantify the scale and magnitude of the problem and its impacts. Ambient air quality data collection is done using reference grade monitors (usually government-managed), for example, the Beta Attenuation Monitor (BAM) which measures Particulate Matter (PM). The BAM can be configured to measure either PM<sub>2.5</sub>, that is, particles smaller than 2.5  $\mu\text{m}$  in diameter or PM<sub>10</sub>, that is, particles smaller than 10  $\mu\text{m}$  in diameter.<sup>8</sup> The BAMs are highly accurate, but remain scarce in many cities in LMICs. For instance, in Sub-Saharan Africa (SSA) only a handful of cities have these monitors.<sup>9-11</sup> This can be attributed to a number of contextual challenges including high setup and maintenance costs, the need for highly skilled labour to operate them, and infrastructure bottlenecks such as unstable power supply<sup>12</sup> and intermittent internet connectivity. Consequently, many cities in SSA and other LMICs are characterised by sparse and limited spatial resolution of air pollution data, inadequate for modelling and analysis to inform decision making and mitigation actions. Recently, Internet of Things (IoT) platforms are being leveraged as a low-cost air quality monitoring approach.<sup>13-15</sup> Smart portable and low-cost air quality monitors can be deployed in diverse locations of the city allowing for measurement of pollution levels in near real-time. Low-cost air quality monitors (LCAQMs) are increasingly being adopted as a complementary approach to fill the air quality data gaps while increasing spatial resolution of air quality data.<sup>10,16-21</sup> Other than the possibility of massive reduction in monitoring costs; improved technical aspects including increased computational capabilities, wireless communication and high data relay frequency explain the noticeable gradual shift towards low-cost sensor platforms to complement traditional reference monitors.<sup>22,23</sup> Large scale deployment of LCAQMs and availability of high-resolution data is likely to spur innovative modelling and analytic approaches to contribute solutions to the growing ambient air pollution challenges, particularly in resource constrained settings where urban environments remain under monitored.

### 1.1 | The low-cost sensor calibration challenge

LCAQMs are more error prone than reference grade monitors, their accuracy degrades over time, they can be affected by external factors such as weather changes, that is, temperature and relative humidity (RH)<sup>24</sup> and suffer from cross-sensitivities between different ambient pollutants.<sup>9,25-27</sup> Previous research on low-cost PM sensors has also shown that low-cost PM sensors may overestimate PM<sub>2.5</sub> concentrations by a factor of approximately 1.5 to 5.0<sup>28-30</sup> and PM<sub>10</sub> concentrations by more than 1.5 for daily average concentrations.<sup>31</sup> Sensor calibration is a key requirement for LCAQMs to ensure data quality and reliability. This involves using appropriate statistical methods to correct measurements from low-cost sensors and validating against reference-grade monitors.<sup>17,26,32,33</sup> In this paper, we investigate machine learning (ML) approaches for sensor calibration on a large-scale air pollution network in urban environments with relatively higher levels of PM concentrations and variations than those studied before. We also investigate the issues involved in deploying such ML-based calibration models to a production system. Unique to this study is that we use AirQo<sup>34,35</sup> LCAQMs, custom designed in and for SSA urban environments. Models developed using our proposed approaches were deployed to calibrate AirQo devices on the AirQo network consisting of over 120 low-cost monitors deployed across urban towns in Uganda (East Africa). The deployment of ML-based calibration models to a production system provides a practical case study of ML for social good in a real-world setting.

### 1.2 | Previous work

Local calibration methods that require individual collocation of each low-cost device with a reference grade monitor are impractical for large networks as they require a great deal of human resource and time. Additionally, access to reference grade monitors is limited in low resource settings.<sup>19,27,30</sup> ML methods, therefore, are preferred for calibrating large sensor networks.

Previous studies<sup>11,19,30,36,37</sup> have achieved low Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) using Multivariate Linear Regression (MLR) and Multi-layer Perceptron (MLP), autoregressive model with extra input and Long-short Term Memory (LSTM), MLR, gaussian mixture regression and Random Forest regression (RF), respectively. However, we note that the pollutant levels considered in the studies are quite low, for example, the median PM<sub>2.5</sub> values were 3.2 µg/m<sup>3</sup> and 6.2 µg/m<sup>3</sup> for the two sites considered by Zaidan et al,<sup>37</sup> the mean and max PM levels were 23.10 µg/m<sup>3</sup> and 115 µg/m<sup>3</sup>, respectively in the study by Lee et al,<sup>36</sup> and mean PM<sub>2.5</sub> values of 9 µg/m<sup>3</sup> in the study by Barkjohn et al.<sup>19</sup> These values are relatively low when compared to the values in this study, that is, our mean PM<sub>2.5</sub> values were 37.9 µg/m<sup>3</sup> and 51.5 µg/m<sup>3</sup> from the BAM and low-cost devices, respectively with BAM PM<sub>2.5</sub> concentrations reaching 290 µg/m<sup>3</sup> while the mean PM<sub>10</sub> values were 51.1 µg/m<sup>3</sup> and 60.6 µg/m<sup>3</sup> from the BAM and low-cost devices, respectively. This partly explains the low RMSE and MAE achieved in these studies. Moreover, the performance of ML methods varies with pollutant levels and/or meteorological conditions for example MLR performance degrades when humidity is high.<sup>36</sup> The higher pollution concentrations and variations for urban areas considered in this paper present unique insights not obtained in previous studies, and therefore advance the case for testing the performance of existing methods, more so for large scale monitoring applications. This is an important step towards advancing low-cost sensing platforms and ML systems for sustained monitoring in resource-constrained settings like SSA.

## 2 | MATERIALS AND METHODS

### 2.1 | Study locations

We consider a real-world air quality monitoring network with over 120 nodes deployed in cities within a developing world setting in SSA. Urban environments in developing cities are characterised by unique pollution profiles due to the sources, weather, demography and economic dynamics as well as technology infrastructure not usually observed elsewhere. The experimental setup for the calibration data includes two monitoring sites that are part of a large sensor network located in Kampala city in Uganda where low-cost air quality monitors and reference grade monitors were co-located for an extensive period. The first co-location site (Reference site 1) and second co-location site (Reference site 2) both have AirQo low cost air quality monitors and reference grade monitors (BAM 1022). The two sites are about 4.62 km apart (see Appendix A). Reference site 1 is located at 0.33347 N latitude, 32.56854 W longitude, is 1237 m above sea level and is about 0.6 km from a major road. The site is an institutional setting (University) with a resident population (parish-level<sup>24</sup>)\* of over 5000, has distinct surface covers built, car parking, tarmacked/paved roads, vegetation areas and no immediate pollution generating activities beyond routine traffic. Reference site 2 is located at 0.33260 N latitude, 32.61004 W longitude, is 1188 m above sea level, is about 0.3 km from a major road. The site is an urban location and resident population (parish-level) of over 26 000, and potentially much higher transient population. In contrast, Site 2 is situated within the buffer zone of major pollution generating activities comprising multiple diffuse sources including traffic, domestic and commercial outdoor burning activities, industries and construction.

### 2.2 | Sensor technology

We consider calibration for PM<sub>2.5</sub> and PM<sub>10</sub>, which are among the priority/public health pollutants. PM concentrations are measured using Met One Beta Attenuation Monitors (BAM) 1022 as reference grade monitors† and custom low-cost AirQo devices deployed at fixed site locations (Reference site 1 and Reference site 2). AirQo devices are built with dual plantower sensors (PMS 5003)‡ and utilise light-scattering technology to measure PM<sub>1.0</sub>, PM<sub>2.5</sub> and PM<sub>10</sub> with an effective measurement range of 0-500 µg/m<sup>3</sup>. AirQo devices collect global positioning system (GPS) coordinates and internal temperature and humidity. All embedded sensors interface with a custom printed circuit board (PCB) with a micro-controller and their measurements are transmitted every 90 seconds to a cloud-based platform via a local cellular/Global system for mobile communication (GSM) network. AirQo devices are designed to withstand the challenges that are common in SSA settings for example they are weather proof, can be powered by either mains or solar thus can be used in areas with no electricity. Additionally, the use of cellular data transmission is ideal for LMICs setting where WiFi coverage is often limited to a few indoor environments.<sup>35</sup>

## 2.3 | Data collection and pre-processing

AirQo devices are tested for consistency in the laboratory environment for at least 7 days prior to field deployment. The two PM sensors in each device are compared to ensure that their measurements are consistent, that is, data offset should be  $\pm 5 \mu\text{g}/\text{m}^3$  for both hourly and daily averages else the monitor is recalled for diagnosis and fault-fixing. Acceptable device to device correlation and device intra-sensor correlations are set to correlation  $\geq 0.99$  and  $R^2 \geq 0.98$ .

PM data was collected using a total of eight AirQo devices and two BAMs collocated at reference site 1 between 15th July 2020 and 17th July 2021 and reference site 2 from 30th September to 26th October 2021. Due to different time-resolutions between the measurements of AirQo devices and BAMs, we averaged AirQo devices' data on an hourly basis so that it can be directly compared with the data from the BAMs. Before averaging, we check to ensure that R values between the two PM sensors in each AirQo device is greater than 0.99 else the data samples for that hour are discarded. This is done to check for any anomalies in the data. We take the PM reading of an AirQo device to be the average of the two PM sensor readings. There were some periods of missing data from AirQo devices as well as the BAMs in the database. The missing data was due to technical problems associated with faults in data acquisition, sensor measurements and communication loss (e.g, due to power outages and national internet shutdown). We excluded all low-cost sensor data for which no reference data was available. After this data pre-processing procedure, the average data completeness for all devices used in this study was approximately 87.6%. Mean hourly  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  concentrations as measured by the two BAMs were  $37.8 \mu\text{g}/\text{m}^3$  and  $45.1 \mu\text{g}/\text{m}^3$  for  $\text{PM}_{2.5}$  and  $51.1 \mu\text{g}/\text{m}^3$  and  $52.6 \mu\text{g}/\text{m}^3$  for  $\text{PM}_{10}$ . A wide range of PM concentration is observed across the datasets with hourly minimum and maximum  $\text{PM}_{2.5}$  of  $1.3 \mu\text{g}/\text{m}^3$  and  $290 \mu\text{g}/\text{m}^3$ , respectively, and minimum and maximum  $\text{PM}_{10}$  concentration of  $2.2 \mu\text{g}/\text{m}^3$  and  $202.3 \mu\text{g}/\text{m}^3$ , respectively.

The BAMs also monitor meteorological variables including ambient temperature and humidity so for our reference sites we use temperature and humidity readings from the BAM. For other sites with no reference monitor, we use temperature and humidity readings from a Trans-African Hydro-Meteorological Observatory (TAHMO) station nearest to the monitoring site. These readings are accessed via an API. Sites considered in this study lie within the tropical rain forest climate zone<sup>38</sup> characterised by somewhat high temperatures and RH. Temperature values across the entire datasets were between 16 and  $33.5^\circ\text{C}$  while RH values ranged from 31 to 99%. Overall temperature and RH means were approximately  $22.7^\circ\text{C}$  and 78.5%, respectively.

## 2.4 | Algorithm selection and validation

In order to find an optimal calibration model, we evaluate the performance of various ML algorithms for low-cost  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  calibration using datasets from the collocation of AirQo devices and BAM at Reference site 1. The algorithms evaluated include  $k$ -nearest neighbours (KNN),<sup>39</sup> support vector machine (SVM),<sup>40</sup> MLR,<sup>41</sup> ridge regression (ridge),<sup>42</sup> lasso regression (lasso),<sup>42</sup> elastic net regression,<sup>42</sup> XGBoost,<sup>43</sup> MLP,<sup>44</sup> RF<sup>43</sup> and gradient boosting (GB).<sup>43</sup> All algorithms were implemented using the scikit-learn Python library, an open source module distributed under a BSD license.  $\text{PM}_{2.5}$  data was collected between 15 July 2020 and 23 March 2021 while  $\text{PM}_{10}$  data was collected between 30 June and 17 July 2021. After data cleaning and pre-processing, we had 4734 (79%) paired hourly averaged BAM-AirQo device  $\text{PM}_{2.5}$  data points and 408 (100%)  $\text{PM}_{10}$  data points to develop the calibration models. 80% of the dataset was randomly selected to construct a training dataset and the remaining 20% was used as a test dataset. The performance of different algorithms was evaluated using the same training and validation datasets. Performance evaluation of each algorithm was done using the RMSE, MAE,  $R^2$  and Pearson's correlation coefficient.

### 2.4.1 | Input variable selection

We performed sensitivity analysis to select the best variable combinations, using a set of input variables including hourly  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  from the low-cost sensor, atmospheric temperature (AT), RH, features derived from timestamp (month and hour [hr]), features from PM including  $\text{errorPM}_{2.5}$  which is the difference between low-cost sensor 1 and sensor 2  $\text{PM}_{2.5}$ ,  $\text{errorPM}_{10}$  which is the difference between low-cost sensor 1 and sensor 2  $\text{PM}_{10}$ ,  $\text{PM}_{2.5} - \text{PM}_{10}$  which is the difference between low-cost average  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$ , with BAM  $\text{PM}_{2.5}$  or  $\text{PM}_{10}$  as the dependent variable.

We use ambient temperature and RH as inputs because they have a direct impact on light scattering particles (PM<sub>2.5</sub>).<sup>25</sup> Previous studies have found that temperature and humidity significantly improved PM<sub>2.5</sub> predictions, furthermore, RH improved consistency in prediction biases between different locations.<sup>19,37</sup>

## 2.4.2 | Algorithm validation methods

We use two different validation methods. (1) Cross unit validation, where we conduct performance evaluation for the proposed models using data from other AirQo devices within the same site, that is, Reference site 1. We use AQ\_G502, AQ\_G505 and AQ\_G506 for PM<sub>2.5</sub> calibration and AQ\_88 for PM<sub>10</sub> calibration. Through this validation, we ensure that the calibration model developed using one AirQo device is accurate when used on other AirQo devices in the same location. (2) Cross site validation where we conduct a performance evaluation for the proposed models using other AirQo devices (ie, AQ\_91 and AQ\_98) collocated with the BAM at Reference site 2. Through this validation, we ensure that calibration models developed at one site are accurate when used to calibrate AirQo devices at different sites.

## 3 | PRESENTATION OF RESULTS

### 3.1 | Preliminary data analysis

We compared hourly measurements from AirQo devices located at both co-location sites, Reference sites 1 and 2. This was done to establish the relationship between AirQo devices in the field. Our results show high device to device correlation with a mean correlation for both PM<sub>2.5</sub> and PM<sub>10</sub> concentration of 0.99 and 0.96 at Reference site 1 and Reference site 2, respectively and a mean  $R^2$  of 0.98 for both PM<sub>2.5</sub> and PM<sub>10</sub> concentrations and both sites.

Additionally, we assessed the relationship between hourly PM<sub>2.5</sub> and PM<sub>10</sub> values of BAM and AirQo devices (AQ\_88 and AQ\_G501) which were used to develop the calibration models. Figure 1 presents scatter plots comparing AirQo device and BAM data before calibration as well as the influence of temperature and humidity on PM. While the measurements from the two devices follow a similar trend, AirQo devices consistently overestimated PM<sub>2.5</sub> and PM<sub>10</sub> values when compared with BAM values. Mean PM<sub>2.5</sub> values were 51.3  $\mu\text{g}/\text{m}^3$  for the AirQo device compared to 37.8  $\mu\text{g}/\text{m}^3$  for BAM while the mean PM<sub>10</sub> value was 60.6  $\mu\text{g}/\text{m}^3$  for the AirQo device and 51.1  $\mu\text{g}/\text{m}^3$  for BAM. The performance of AirQo devices before calibration is summarised in the first row of Tables 1 and 2.

Furthermore, we evaluated the relationship between the two reference monitors (BAMs) used in this study. The two BAMs were collocated at Reference site 1 and were both configured to measure PM<sub>10</sub> from 30th June to 17th July 2021 and then configured to measure PM<sub>2.5</sub> from 09 to 19 August 2021. The RMSE, MAE,  $R^2$  and correlation values between measurements of the two BAMs were 5.1  $\mu\text{g}/\text{m}^3$ , 4.0  $\mu\text{g}/\text{m}^3$ , 0.96 and 0.98, respectively for PM<sub>10</sub> and 6.2  $\mu\text{g}/\text{m}^3$ , 4.1  $\mu\text{g}/\text{m}^3$ , 0.92 and 0.96, respectively for PM<sub>2.5</sub>. These results were used as a benchmark for the calibration models.

### 3.2 | Algorithm selection

For all the algorithms considered in this study, we tested several input variable combinations and the best performance for most algorithms was achieved using input variable combinations in Equations (1) and (2) for PM<sub>2.5</sub> and PM<sub>10</sub> calibration, respectively.

$$\text{TargetPM}_{2.5} = \text{RF}(\text{PM}_{2.5}, \text{AT}, \text{RH}, \text{PM}_{10}, \text{errorPM}_{2.5}, \text{errorPM}_{10}, \text{PM}_{2.5} - \text{PM}_{10}, \text{month}, \text{hr}) \quad (1)$$

$$\text{TargetPM}_{10} = \text{Lasso}(\text{PM}_{10}, \text{AT}, \text{RH}, \text{PM}_{2.5}, \text{hr}) \quad (2)$$

While all algorithms showed good performance for PM<sub>2.5</sub> calibration, RF had the best performance. Performance of MLP and GBM models was comparable to the performance of RF but MLP and GBM were harder to tune than RF, that is, it was difficult to maximise the model's performance without overfitting. On the other hand, lasso regression had the best performance for low-cost PM<sub>10</sub> calibration. Therefore, RF and lasso regression models were used for further

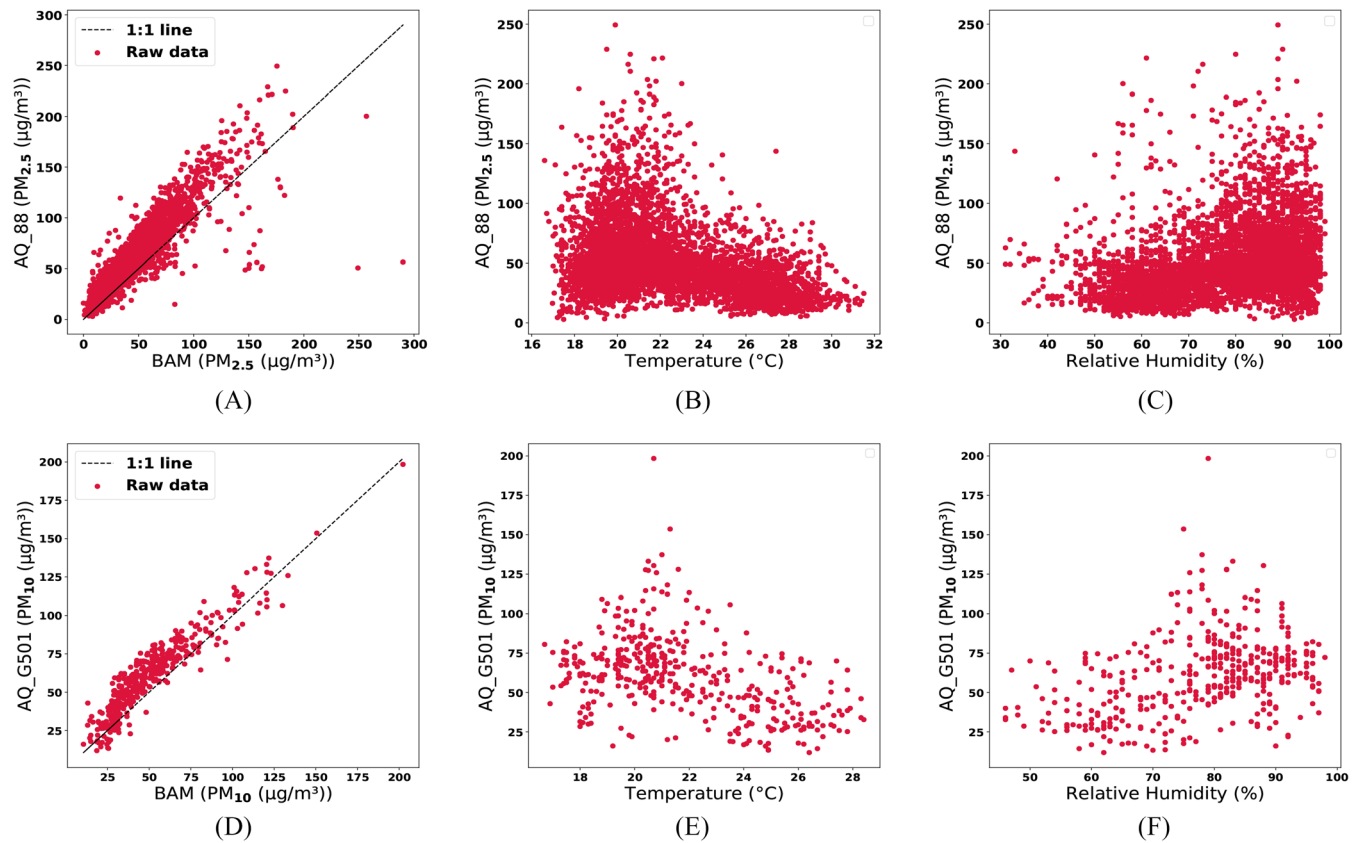


FIGURE 1 Part A,D,  $PM_{2.5}$  and  $PM_{10}$  comparison between BAM and AirQo devices (AQ\_88 and AQ\_G501) collocated at Reference site 1 before calibration. Part B,C,E,F, relationship between low-cost PM and temperature and humidity

TABLE 1 Random forest using optimal parameters and various input variable combinations

Input variables	RMSE ( $\mu\text{g}/\text{m}^3$ )	MAE ( $\mu\text{g}/\text{m}^3$ )	$R^2$	Correlation
Factory calibrated (Raw $PM_{2.5}$ )	18.6	14.6	0.52	0.9
$PM_{2.5}$ , AT, RH	10.4	6.02	0.85	0.92
$PM_{2.5}$ , AT, RH, $PM_{10}$	9.3	5.6	0.88	0.94
$PM_{2.5}$ , AT, RH, $PM_{10}$ , error $PM_{2.5}$	9.1	5.3	0.88	0.94
$PM_{2.5}$ , AT, RH, $PM_{10}$ , error $PM_{2.5}$ , error $PM_{10}$	8.5	5.1	0.90	0.95
$PM_{2.5}$ , AT, RH, $PM_{10}$ , error $PM_{2.5}$ , error $PM_{10}$ , $PM_{2.5} - PM_{10}$	7.6	4.8	0.92	0.96
$PM_{2.5}$ , AT, RH, $PM_{10}$ , error $PM_{2.5}$ , error $PM_{10}$ , $PM_{2.5} - PM_{10}$ , month	7.4	4.7	0.92	0.96
$PM_{2.5}$ , AT, RH, $PM_{10}$ , error $PM_{2.5}$ , error $PM_{10}$ , $PM_{2.5} - PM_{10}$ , month, hr	7.2	4.6	0.92	0.96
Collocated BAMs (Benchmark)	6.2	4.1	0.92	0.96

analyses for low-cost  $PM_{2.5}$  and  $PM_{10}$  calibration, respectively. Detailed results for  $PM_{2.5}$  and  $PM_{10}$  calibration using the various algorithms are summarised in Tables B1 and B2 in Appendix B.

### 3.3 | $PM_{2.5}$ calibration

The proposed RF model was trained using Equation (1) and optimal parameters max\_depth of 50, 1000 n\_estimators, max\_features equal to the square root of the total number of features. The parameters were selected using a grid search.

TABLE 2 Lasso regression using optimal parameters and various input variable combinations

Input combinations	RMSE ( $\mu\text{g}/\text{m}^3$ )	MAE ( $\mu\text{g}/\text{m}^3$ )	$R^2$	Correlation
Factory calibrated ( $\text{PM}_{10}$ )	13.4	11.3	0.72	0.93
$\text{PM}_{10}$ , AT, RH	9.0	6.9	0.91	0.96
$\text{PM}_{10}$ , AT, RH, $\text{PM}_{2.5}$	8.2	6.3	0.93	0.96
$\text{PM}_{10}$ , AT, RH, $\text{PM}_{2.5}$ , error $\text{PM}_{10}$	8.2	6.3	0.93	0.96
$\text{PM}_{10}$ , AT, RH, $\text{PM}_{2.5}$ , error $\text{PM}_{10}$ , error $\text{PM}_{2.5}$	8.2	6.3	0.93	0.96
$\text{PM}_{10}$ , AT, RH, $\text{PM}_{2.5}$ , error $\text{PM}_{10}$ , error $\text{PM}_{2.5}$ , $\text{PM}_{2.5} - \text{PM}_{10}$	8.2	6.3	0.93	0.96
$\text{PM}_{10}$ , AT, RH, $\text{PM}_{2.5}$ , error $\text{PM}_{10}$ , error $\text{PM}_{2.5}$ , $\text{PM}_{2.5} - \text{PM}_{10}$ , month	8.2	6.3	0.93	0.96
$\text{PM}_{10}$ , AT, RH, $\text{PM}_{2.5}$ , error $\text{PM}_{10}$ , error $\text{PM}_{2.5}$ , $\text{PM}_{2.5} - \text{PM}_{10}$ , month, hr	7.9	6.0	0.93	0.97
$\text{PM}_{10}$ , AT, RH, $\text{PM}_{2.5}$ , hr	7.9	6.0	0.93	0.97
Collocated BAMs (Benchmark)	5.1	4.0	0.96	0.98

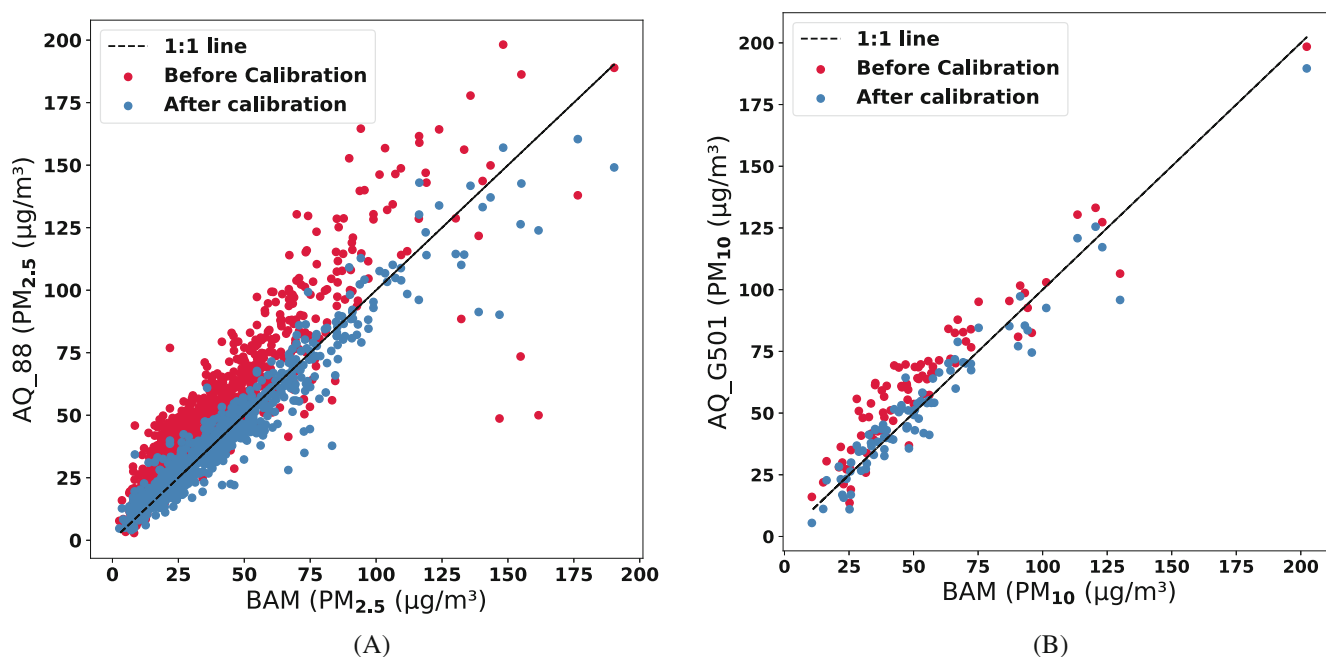
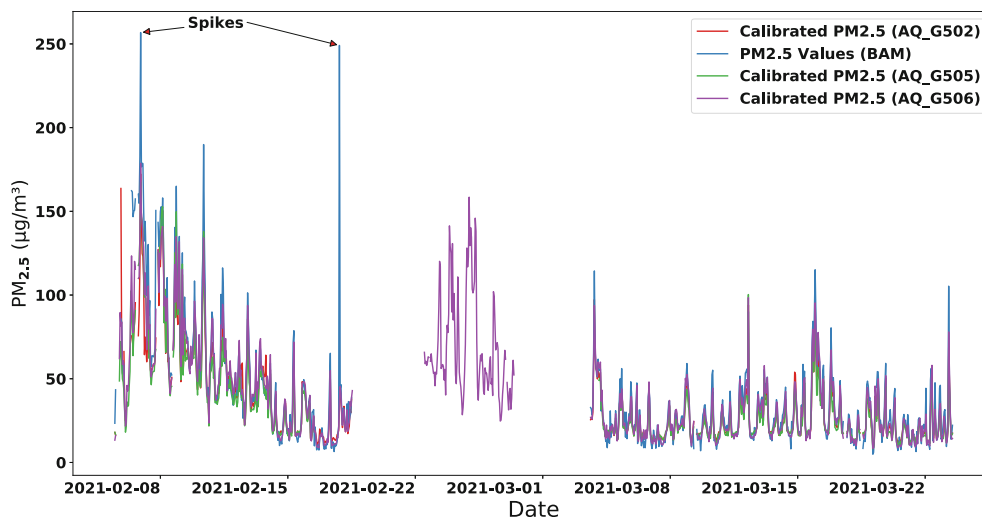


FIGURE 2 Comparison between BAM and low-cost PM from the test set. In part A, we present a scatter plot that shows the relationship between BAM and lowcost (AQ\_88)  $\text{PM}_{2.5}$  before and after calibration using the proposed RF model while in part B, we present a scatter plot that shows the relationship between BAM and lowcost (AQ\_G501)  $\text{PM}_{10}$  before and after calibration using the proposed lasso regression model

We observed a strong relationship between  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  concentrations. The correlation between these values was 0.99. Including  $\text{PM}_{10}$  and features from  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  significantly reduced RMSE and MAE and increased  $R^2$  and correlation as shown in Table 1. Performance of the RF model with the different variable combinations is summarised in Table 1. Figure 2A is a scatterplot comparing AirQo device (AQ\_88) and BAM  $\text{PM}_{2.5}$  using the test set before and after calibration. From the figure, we observe a very close fit between calibrated low-cost and BAM  $\text{PM}_{2.5}$  values, before calibration, AirQo device values were higher than BAM values.

### 3.4 | $\text{PM}_{10}$ calibration

The lasso regression model had consistently better performance for  $\text{PM}_{10}$  with all variable combinations compared to other models considered in this study, however we observed variables error $\text{PM}_{2.5}$ , error $\text{PM}_{10}$ ,  $\text{PM}_{2.5} - \text{PM}_{10}$  and month did not have any effect on the performance of the lasso regression model as shown in Table 2 therefore these were omitted in the final implementation. We used lassoCV with iterative fitting to automatically select optimal parameters for the lasso regression model using cross-validation.



**FIGURE 3** Cross-unit validation results for  $PM_{2.5}$  calibration using the RF model. We present hourly comparison between BAM and calibrated low-cost  $PM_{2.5}$  for AirQo devices (AQ\_G502), (AQ\_G505) and (AQ\_G506)

**TABLE 3** Cross-unit validation performance comparison by device

Device	Method	RMSE ( $\mu\text{g}/\text{m}^3$ )	MAE ( $\mu\text{g}/\text{m}^3$ )	$R^2$	Correlation
AQ_G502	Factory calibrated ( $PM_{2.5}$ )	19	11.5	0.61	0.85
	RF	15.9	7.5	0.79	0.92
	RF (spikes excluded)	13.5	7.1	0.83	0.94
AQ_G505	Factory calibrated ( $PM_{2.5}$ )	16.5	8.8	0.60	0.89
	RF	13.7	8.1	0.84	0.96
	RF (spikes excluded)	13.3	8.2	0.84	0.96
AQ_G506	Factory calibrated ( $PM_{2.5}$ )	20.3	12.6	0.68	0.87
	RF	14.8	6.3	0.85	0.94
	RF (spikes excluded)	10.2	5.8	0.92	0.97
AQ_88	Factory calibrated ( $PM_{10}$ )	19.6	17.1	0.50	0.91
	lasso	11.9	8.6	0.78	0.92

Results from performance evaluation of the lasso regression model for  $PM_{10}$  calibration are summarised in Table 2. Figure 2B presents a scatter plot showing the relationship between AirQo device and BAM  $PM_{10}$  before and after calibration using the test set. From the figure, it is clear that before calibration AirQo devices over estimate BAM PM and after calibration these values are lowered and thus closer to the BAM values.

### 3.5 | Algorithm validation

#### 3.5.1 | Cross unit validation

Figure 3 compares calibrated  $PM_{2.5}$  of each AirQo device with collocated BAM  $PM_{2.5}$ . The devices were collocated with BAM from 5th February to 23rd March 2021. Table 3 summarises the performance of the RF model for each AirQo device. From Figure 3 and Table 3, we observe significant improvement in data accuracy for all devices with a drop in RMSE and MAE by up to 10.20 and 6.8  $\mu\text{g}/\text{m}^3$ , respectively when spikes are excluded and an increase in  $R^2$  and correlation by up to 0.24 and 0.1, respectively. Spikes are momentarily high pollution levels, from our contextual knowledge these could be due to a sudden pollution event in the neighbourhood.

Likewise, the lasso regression model used for  $PM_{10}$  calibration also showed good performance with cross unit validation using AirQo device (AQ\_88) collocated with BAM from 30th June to 17th July 2021. The performance is

TABLE 4 Cross-site validation performance comparison by device

Device	Method	RMSE ( $\mu\text{g}/\text{m}^3$ )	MAE ( $\mu\text{g}/\text{m}^3$ )	$R^2$	Correlation
AQ_91	Factory calibrated ( $\text{PM}_{2.5}$ )	13.5	11.2	0.57	0.91
	RF	5.7	4.2	0.90	0.95
	Factory calibrated ( $\text{PM}_{10}$ )	16.5	12.3	0.67	0.85
	Lasso	13.7	8.5	0.78	0.9
AQ_98	Factory calibrated ( $\text{PM}_{2.5}$ )	11.3	9.4	0.66	0.92
	RF	5.1	3.8	0.92	0.96
	Factory calibrated ( $\text{PM}_{10}$ )	15.1	10.9	0.69	0.86
	Lasso	14.4	9.1	0.76	0.91

summarised in Table 3. A visual comparison between calibrated low-cost  $\text{PM}_{10}$  and collocated BAM  $\text{PM}_{10}$  is given in Figure C2 in Appendix C.

### 3.5.2 | Cross site validation

Table 4 summarises the performance of the RF and lasso regression models with each AirQo device AQ\_91 and AQ\_98 used for cross site validation. The BAM at Reference site 2 was configured to measure  $\text{PM}_{2.5}$  from 15th to 26th October and  $\text{PM}_{10}$  from 30th September to 15th October. The visual comparisons between calibrated low-cost  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  for each AirQo device collocated with the BAM PM are given in Figures C3 and C4, respectively in Appendix C.

## 4 | DISCUSSION

While the measurements from LCAQMs and reference monitors follow a similar trend, LCAQMs consistently overestimate  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  concentrations when compared with reference PM concentrations even after laboratory calibration (see Figure 1). This demonstrates the need for field calibration of low-cost devices. Previous studies have shown that multivariate calibration can reduce sensor cross sensitivities and makes it possible to generalise models for different units and monitoring sites with varying environmental conditions. ML methods provide a strong framework for multivariate regression algorithms that can support sensor calibration.<sup>19,45</sup> In order to choose the most optimal calibration model for AirQo devices, we compared the performance of various multivariate ML methods using the same input variables and data. Based on our results, all algorithms evaluated in this study showed relatively good performance for  $\text{PM}_{2.5}$  but the performance of non-linear algorithms was generally poor for  $\text{PM}_{10}$  calibration. This can be attributed to the smaller dataset that was available for training the models (408 data points). RF and lasso regression methods were superior for  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  calibration, respectively and their performance is comparable to the results (benchmark) obtained by collocating the two BAMs used in this study with the maximum difference in RMSE and MAE of  $1 \mu\text{g}/\text{m}^3$  for  $\text{PM}_{2.5}$  and  $2.8 \mu\text{g}/\text{m}^3$  for  $\text{PM}_{10}$ . The best performing algorithms were further evaluated with different units (cross-unit validation) and sites (cross site validation). From our cross-unit validation results (see Figure 3 and Table 3), we draw the following conclusions. Cross unit validation shows promising results, with calibrated  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$  values of AirQo devices being very close to BAM values. However, the RF model tends to under-predict spikes, this raises the RMSE and MAE and also contributed to the lower  $R^2$  and correlation values obtained (see Table 3). This is because RF does not extrapolate to values beyond its training range but may constantly predict the maximum values encountered in the training set.<sup>46</sup> When these spikes are excluded from the data, the RMSE and MAE values reduced by up to  $4.6 \mu\text{g}/\text{m}^3$  and  $0.6 \mu\text{g}/\text{m}^3$  for AQ\_G506. RF and lasso regression models also performed well when tested with devices from a different site other than the training site. We observe a significant improvement in performance for both devices (ie, AQ\_91 and AQ\_98). We were able to achieve reasonable accuracy with cross-unit and cross-site validation because AirQo devices have strong precision. The correlation between two collocated AirQo devices is typically above 0.96 and  $R^2$  above 0.98. This shows that AirQo devices do not have to be calibrated individually.

## 5 | DEPLOYMENT OF CALIBRATION MODELS IN PRODUCTION

The resulting calibration model is deployed as part of an urban air quality sensing system that is accessible to users via different channels including an open-air quality API, an analytics dashboard,\*\* and a mobile app. As such, the deployment serves as a demonstration of the use of a ML system in addressing societal challenges, in this case ambient urban air pollution. A key requirement for the deployment is scalability and availability. We adopt the microservice architecture powered by the Kubernetes infrastructure.<sup>47</sup> The calibration models, that is, RF and lasso regression models are encapsulated as a microservice that are exposed as REST APIs. The CI/CD pipeline is powered by GitHub Actions and the deployed models run on a multi node kubernetes cluster to ensure high availability and scalability. Raw sensor measurements from all AirQo devices on the network are streamed to a cloud-based IoT platform followed by data streaming pipeline using kafka connectors running on the same Kubernetes orchestrated environment. We leverage kafka streams to perform data cleaning and other ETL functions. The cleaned raw sensor measurements are then stored in a sharded MongoDB which runs on a separate multi-node Kubernetes cluster. Data is serialized and transmitted using bytes to reduce the size of messages. For security reasons, all topics (data stores) within the pipeline require authentication before data enters or leaves the data store. Authentication credentials are stored in secret files within the cluster and additional configurations are stored in configuration maps. The simplified pipeline of the system is shown in Figure D5 in Appendix D.

Raw hourly PM concentrations are re-sampled and fed into the calibration model together with corresponding hourly temperature and humidity readings from the nearest meteorological station to generate corresponding calibrated PM concentrations. The calibrated PM concentrations are then stored alongside the raw hourly PM concentrations. We store our data in mongoDB from where it can be retrieved and shared or used to conduct further analysis.

## 6 | IMPLICATIONS AND FUTURE WORK

In this work, we developed general calibration approaches for low-cost (AirQo device) PM<sub>2.5</sub> and PM<sub>10</sub> based on collocation with reference monitors (BAMs). In order to improve the performance of various ML algorithms, we explored the performance of selected ML algorithms with different variable combinations (features) and chose the combination that provides the best performance in terms of RMSE, MAE,  $R^2$  and correlation. We showed that our proposed models can be used with similar devices and in different environments for which they have not been previously calibrated against through cross-unit and cross-site validation. We also give details of how our proposed models are deployed to production and used to calibrate over 120 low-cost devices.

Ideally, our calibration models developed by collocation with a reference monitor would be trained using data covering the entire range of conditions such as PM levels, temperature and humidity that can be experienced in a given location, however, practically, this may not be possible especially when sensors are deployed in new geographical regions previously not monitored like Uganda. This and the fact that we had access to only two static reference monitors made it difficult to get a larger training dataset. This therefore means that, in order to obtain reasonable calibration accuracy in environments other than that where the training data was collected, the feature space also should include meteorological factors such as temperature and humidity and concentrations of relevant cross-sensitive species, such as PM<sub>10</sub> or PM<sub>2.5</sub> when calibrating PM<sub>2.5</sub> or PM<sub>10</sub>, respectively.<sup>19</sup> It is also important to retrain these models periodically with new data in order to cater for seasonal and condition-specific dependency of calibration factors.<sup>11,29,30</sup> We shall consider seasonal re-calibration of models as part of our future work. These efforts will improve the accuracy of ML methods in predicting wider range of PM concentration levels in different environments and drive the applicability of low-cost air quality sensors.

Both calibration models (RF and lasso regression) proposed in this work are able predict PM levels with reasonable accuracy; however, they are still unable to predict spikes. These spikes are usually a result of local pollution sources that vanish rapidly resulting in significant spatial variations. Prediction of spikes is an important area of future research.

### AUTHOR CONTRIBUTIONS

Engineer Bainomugisha: conceptualisation, supervision, review, writing and editing. Deo Okure: writing, review and editing. Priscilla Adong: data curation, calibration, visualisation, writing-original draft and editing. Richard Sserunjogi: code review, writing and editing.

## ACKNOWLEDGEMENTS

The authors would like to appreciate the feedback and discussions from the AirQo team and partners from government, academia, private sector and civil society. This work was supported by [Google.org](https://www.google.org), the Global Challenges Research Fund (GCRF), and Belgium through the Wehubit programme implemented by Enabel.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Priscilla Adong  <https://orcid.org/0000-0002-4884-3387>

## ENDNOTES

\* A parish is the second smallest administrative unit. Hierarchy of the administrative units in Uganda are of the order: Country > district > county > sub-county > parish and village (smallest unit).

† <https://metone.com/wp-content/uploads/2022/02/BAM-1022-9800-Rev-G.pdf>.

‡ [https://www.plantower.com/en/products\\_33/74.html](https://www.plantower.com/en/products_33/74.html).

§ <https://tahmo.org/>.

¶ [https://tahmo.org/docs/TAHMO\\_API\\_documentation\\_latest.pdf](https://tahmo.org/docs/TAHMO_API_documentation_latest.pdf).

\*\* <https://platform.airqo.net/>.

**How to cite this article:** Adong P, Bainomugisha E, Okure D, Sserunjogi R. Applying machine learning for large scale field calibration of low-cost PM<sub>2.5</sub> and PM<sub>10</sub> air pollution sensors. *Applied AI Letters*. 2022;3(3):e76. doi:10.1002/ail2.76

## REFERENCES

1. He C, Clifton O, Felker-Quinn E, et al. Interactions between air pollution and terrestrial ecosystems: perspectives on challenges and future directions. *Bull Am Meteorol Soc*. 2021;102(3):E525-E538.
2. Anenberg SC, Belova A, Brandt J, et al. Survey of ambient air pollution health risk assessment tools. *Risk Anal*. 2016;36(9):1718-1736. doi:10.1111/risa.12540
3. World Health Organization. *Air Pollution*. WHO Health Topics; 2021. [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health#:~:text=Ambient%20\(outdoor\)%20air%20pollution%20in,and%20respiratory%20disease%2C%20and%20cancers](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health#:~:text=Ambient%20(outdoor)%20air%20pollution%20in,and%20respiratory%20disease%2C%20and%20cancers). Accessed November 01, 2021
4. Brunekreef B, Holgate ST. Air pollution and health. *Lancet*. 2002;360(9341):1233-1242.
5. Cohen AJ, Brauer M, Burnett R, et al. 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. *Lancet*. 2017;389(10082):1907-1918.
6. Katoto PD, Byamungu L, Brand AS, et al. Ambient air pollution and health in Sub-Saharan Africa: current evidence, perspectives and a call to action. *Environ Res*. 2019;173:174-188.
7. Green P, Okure D, Adong P, Sserunjogi R, Bainomugisha E. Exploring PM<sub>2.5</sub> variations from a low-cost sensor network in Greater Kampala, during COVID-19 imposed lockdown restrictions: lessons for policy. *Clean Air J*. 2022;32(1):37-50.
8. United State Environmental Protection Agency. *Particulate Matter (PM) Basics*. WHO Health Topics; 2022. <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>. Accessed April 28, 2022.
9. McFarlane C, Isevulambire PK, Lumbuenamo RS, et al. First measurements of ambient PM<sub>2.5</sub> in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo using field-calibrated low-cost sensors. *Aerosol Air Qual Res*. 2021;21:200619.
10. Ngo NS, Asseko SVJ, Ebanega MO, Allo'o SMA, Hystad P. The relationship among PM<sub>2.5</sub>, traffic emissions, and socioeconomic status: evidence from Gabon using low-cost, portable air quality monitors. *Transport Res D: Transp Environ*. 2019;68:2-9.
11. McFarlane C, Raheja G, Malings C, Appoh EK, Hughes AF, Westervelt DM. Application of Gaussian mixture regression for the correction of low cost PM<sub>2.5</sub> monitoring data in Accra, Ghana. *ACS Earth Space Chem*. 2021;5(9):2268-2279.
12. Pinder RW, Klopp JM, Kleiman G, Hagler GS, Awe Y, Terry S. Opportunities and challenges for filling the air quality data gap in low- and middle-income countries. *Atmos Environ*. 2019;215:116794.

13. Dhingra S, Madda RB, Gandomi AH, Patan R, Daneshmand M. Internet of things mobile–air pollution monitoring system (IoT-Mobair). *IEEE Internet Things J.* 2019;6(3):5577–5584. doi:10.1109/JIOT.2019.2903821
14. Korunoski M, Stojkoska BR, Trivodaliev K. Internet of things solution for intelligent air pollution prediction and visualization. *IEEE EUROCON 2019 -18th International Conference on Smart Technologies.* Novi Sad, Serbia: IEEE; 2019:1–6.
15. Fuertes W, Carrera D, Villacás C, et al. Distributed system as internet of things for a new low-cost, air pollution wireless monitoring on real time. *2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT).* Chengdu, China: IEEE; 2015:58–67.
16. Alhasa KM, Mohd Nadzir MS, Olalekan P, et al. Calibration model of a low-cost air quality sensor using an adaptive neuro-fuzzy inference system. *Sensors.* 2018;18(12):4380.
17. Maag B, Zhou Z, Thiele L. A survey on sensor calibration in air pollution monitoring deployments. *IEEE Internet Things J.* 2018;5(6):4857–4870.
18. Mead M, Popoola O, Stewart G, et al. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmos Environ.* 2013;70:186–203.
19. Barkjohn KK, Gantt B, Clements AL. Development and application of a United States-wide correction for PM 2.5 data collected with the purple air sensor. *Atmos Meas Techn.* 2021;14(6):4617–4637.
20. Amegah AK. Proliferation of low-cost sensors. What prospects for air pollution epidemiologic research in Sub-Saharan Africa? *Environ Pollut.* 2018;241:1132–1137.
21. Pope FD, Gatari M, Ng'ang'a D, Poynter A, Blake R. Airborne particulate matter monitoring in Kenya using calibrated low-cost sensors. *Atmos Chem Phys.* 2018;18(20):15403–15418.
22. Snyder EG, Watkins TH, Solomon PA, et al. The changing paradigm of air pollution monitoring. *Environ Sci Technol.* 2013;47(20):11369–11377.
23. Morawska L, Thai PK, Liu X, et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone? *Environ Int.* 2018;116:286–299.
24. Okure D, Ssematimba J, Sserunjogi R, Gracia NL, Soppelsa ME, Bainomugisha E. Characterization of ambient air quality in selected urban areas in Uganda using low-cost sensing and measurement technologies. *Environ Sci Technol.* 2022;56(6):3324–3339.
25. Concas F, Mineraud J, Lagerspetz E, et al. Low-cost outdoor air quality monitoring and sensor calibration: a survey and critical analysis. *ACM Trans Sens Netw (TOSN).* 2021;17(2):1–44.
26. Giordano MR, Malings C, Pandis SN, et al. From low-cost sensors to high-quality data: a summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors. *J Aerosol Sci.* 2021;158:105833.
27. Malings C, Tanzer R, Hauryliuk A, et al. Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. *Atmos Meas Tech.* 2019;12(2):903–920.
28. Badura M, Batog P, Drzeniecka-Osiadacz A, Modzel P. Evaluation of low-cost sensors for ambient PM<sub>2.5</sub> monitoring. *J Sens.* 2018;2018:1–16.
29. Sayahi T, Butterfield A, Kelly K. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ Pollut.* 2019;245:932–940.
30. Wang WCV, Lung SCC, Liu CH. Application of machine learning for the in-field correction of a PM<sub>2.5</sub> low-cost sensor network. *Sensors.* 2020;20(17):5002.
31. Carratu M, Ferro M, Paciello V, Sommella P, Lundgren J, O'Nils M. Wireless sensor network calibration for PM<sub>10</sub> measurement. *2020 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA).* Tunis, Tunisia: IEEE; 2020:1–6.
32. Spinelle L, Gerboles M, Villani MG, Aleixandre M, Bonavitacola F. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sens Actuators B.* 2015;215:249–257.
33. Hasenfratz, D., Saukh, O., Thiele, L. (2012). On-the-Fly Calibration of Low-Cost Gas Sensors. In: Picco, G.P., Heinzelman, W. (eds) *Wireless Sensor Networks. EWSN 2012. Lecture Notes in Computer Science*, vol 7158. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-28169-3\\_15](https://doi.org/10.1007/978-3-642-28169-3_15)
34. AirQO. *AirQo Monitoring System.* Kampala, Uganda: AirQO; 2022.
35. Coker ES, Amegah AK, Mwebaze E, Ssematimba J, Bainomugisha E. A land use regression model using machine learning and locally developed low cost particulate matter sensors in Uganda. *Environ Res.* 2021;199:111352.
36. Lee H, Kang J, Kim S, Im Y, Yoo S, Lee D. Long-term evaluation and calibration of low-cost particulate matter (PM) sensor. *Sensors.* 2020;20(13):3617.
37. Zaidan MA, Motlagh NH, Fung PL, et al. Intelligent calibration and virtual sensing for integrated low-cost air quality sensors. *IEEE Sens J.* 2020;20(22):13638–13652.
38. Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci Data.* 2018;5(1):1–12.
39. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat.* 1992;46(3):175–185.
40. Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat Comput.* 2004;14(3):199–222.
41. Rencher AC, Christensen WF. Chapter 10, multivariate regression—Section 10.1, introduction. *Methods of Multivariate Analysis, Wiley Series in Probability and Statistics.* Vol 709. New Jersey, USA: John Wiley & Sons; 2012:19.

42. Ogutu JO, Schulz-Streeck T, Piepho HP. *Genomic Selection Using Regularized Linear Regression Models: Ridge Regression, Lasso, Elastic Net and their Extensions*. Rennes, France: Springer; 2012:1-6.
43. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artif Intell Rev*. 2021;54(3):1937-1967.
44. Zhang Z. *Multivariate Time Series Analysis in Climate and Environmental Research*. Cham, Switzerland: Springer; 2018:1-35.
45. Esposito E, De Vito S, Salvato M, Fattoruso G, Di Francia G. *Computational Intelligence for Smart Air Quality Monitors Calibration*. Cham, Switzerland: Springer; 2017:443-454.
46. Nowack P, Konstantinovskiy L, Gardiner H, Cant J. Machine learning calibration of low-cost NO<sub>2</sub> and PM<sub>10</sub> sensors: non-linear algorithms and their impact on site transferability. *Atmos Meas Tech*. 2021;14(8):5637-5655.
47. Bernstein D. Containers and cloud: from LXC to Docker to Kubernetes. *IEEE Cloud Comput*. 2014;1(3):81-84. doi:10.1109/MCC.2014.51

## APPENDIX A: STUDY LOCATION

We provide images of the the two study sites considered in this study (Figure A1).



(A) Reference site 1



(B) Reference site 2



(C) Location of the reference sites

**FIGURE A1** Monitoring sites used in this study. Part A, shows AirQo devices and BAMS installed at Makerere University (Reference site 1), part B, shows AirQo devices and BAM installed at Nakawa (Reference site 2) and part C, is a map of Kampala city showing the locations of the two monitoring sites considered in this study

## APPENDIX B: ALGORITHM SELECTION

Here, we present in more detail the results from performance evaluation of the various ML algorithm evaluated in this study (Tables B1 and B2).

$$\text{Target}PM_{10} = \text{Lasso}(PM_{10}, AT, RH, PM_{2.5}, \text{error}PM_{2.5}, \text{error}PM_{10}, PM_{2.5} - PM_{10}, \text{month}, \text{hr}) \quad (\text{B1})$$

TABLE B1 Results from performance evaluation of various ML algorithms for PM<sub>2.5</sub> calibration using Equation (1)

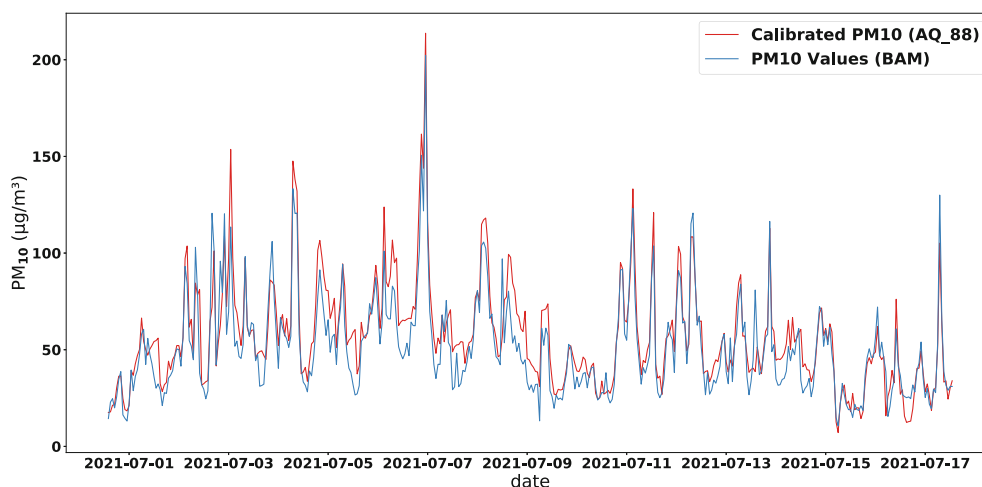
Method	RMSE ( $\mu\text{g}/\text{m}^3$ )	MAE ( $\mu\text{g}/\text{m}^3$ )	$R^2$	Correlation
Factory calibrated	18.6	14.6	0.52	0.9
SVM	10.4	5.5	0.84	0.93
Lasso	9.7	5.8	0.86	0.93
Elastic net	9.7	5.8	0.86	0.93
Ridge	9.5	5.9	0.87	0.93
MLR	9.5	5.9	0.87	0.93
KNN	8.9	5.3	0.89	0.94
XGBoost	7.6	4.9	0.92	0.96
GB	7.2	4.7	0.92	0.96
MLP	7.2	4.8	0.92	0.96
RF	7.2	4.6	0.92	0.96

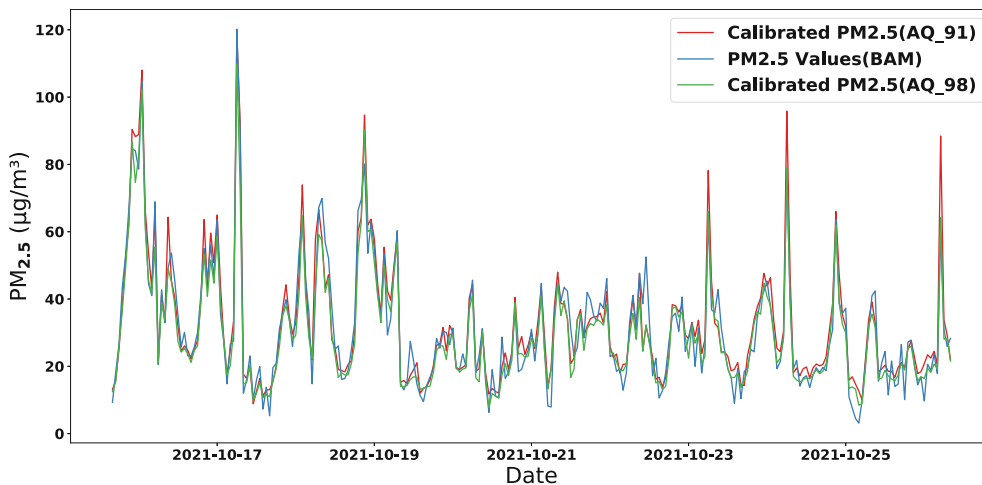
TABLE B2 Results from performance evaluation of various ML algorithms for PM<sub>10</sub> calibration using Equation (B1)

Method	RMSE ( $\mu\text{g}/\text{m}^3$ )	MAE ( $\mu\text{g}/\text{m}^3$ )	$R^2$	Correlation
Factory calibrated (PM <sub>10</sub> )	13.4	11.3	0.72	0.93
SVM	20.6	9.8	0.54	0.79
KNN	11.8	7.0	0.85	0.93
RF	11.4	7.1	0.86	0.93
XGBoost	11.2	6.8	0.86	0.93
GB	10.9	6.7	0.87	0.94
MLP	8.0	6.0	0.93	0.97
Elastic net	8.0	6.0	0.93	0.97
Ridge	7.9	6.1	0.93	0.97
MLR	7.9	6.1	0.93	0.97
Lasso	7.9	6.0	0.93	0.97

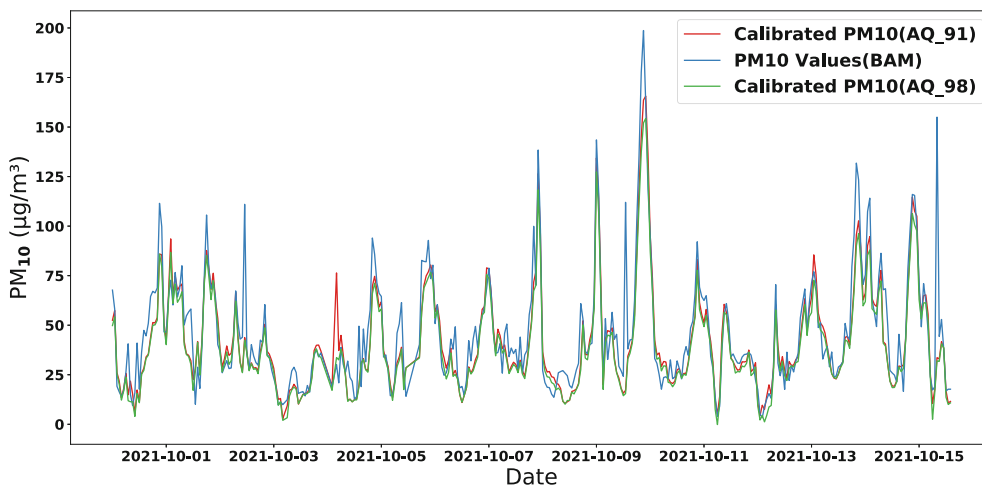
## APPENDIX C: ALGORITHM VALIDATION

We provide additional graphs presenting results from cross unit and cross site validations. In Figure C2, we present cross unit validation results by comparing calibrated low cost PM<sub>10</sub> and collocated BAM PM<sub>10</sub> while in Figures C3

FIGURE C2 Hourly comparison between BAM and calibrated low-cost PM<sub>10</sub> for AirQo devices collocated with BAM at Reference site 1



**FIGURE C3** Hourly comparison between BAM and calibrated low-cost PM<sub>2.5</sub> for AirQo devices AQ\_91 and AQ\_98 collocated with BAM at Reference site 2

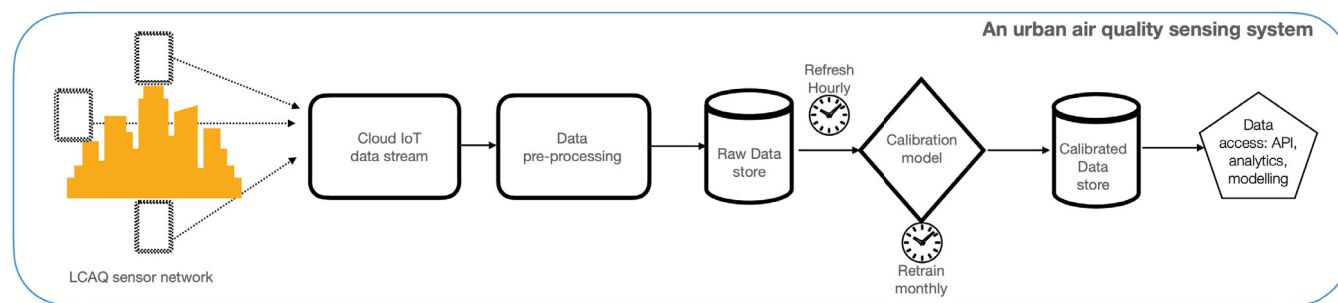


**FIGURE C4** Hourly comparison between BAM and calibrated low-cost PM<sub>10</sub> for AirQo devices AQ\_91 and AQ\_98 collocated with BAM at Reference site 2

and C4, we present cross site validation results for PM<sub>2.5</sub> and PM<sub>10</sub>, respectively. We observe a close fit between calibrated PM and BAM PM with both cross unit and cross site validation.

#### APPENDIX D: DEPLOYMENT PIPELINE

We provide a diagram of the calibration model pipeline in production (Figure D5).



**FIGURE D5** Deployment pipeline of the calibration model in an urban air quality sensing system