International Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tres20

Assessing and monitoring air quality in cities and urban areas with a portable, modular and low-cost sensor station: calibration challenges

M. Tarazona Alvarado, J. L. Salamanca-Coy, K. Forero-Gutièrrez, L. A. Núñez, J. Pisco-Guabave, Fr. Escobar-Diaz & D. Sierra-Porta

INTERNATIONAL JOURNAL OF REMOTE SENSING
2024, VOL. 45, NO. 17, 5713–5736
https://doi.org/10.1080/01431161.2024.2373338

Assessing and monitoring air quality in cities and urban areas with a portable, modular and low-cost sensor station: calibration challenges

M. Tarazona Alvarado [a], J. L. Salamanca-Coy [b], K. Forero-Gutièrrez [b], L. A. Núñez [a], J. Pisco-Guabave [a], Fr. Escobar-Diaz [c] and D. Sierra-Porta [d]

[a]Escuela de Física, Universidad Industrial de Santander, Bucaramanga, Colombia; [b]IOT and Cloud Department, MakeSens, Bucaramanga, Santander, Colombia; [c]Departamento de Ingeniería Química y Ambiental, Universidad Nacional de Colombia, Bogotá, Colombia; [d]Facultad de Ciencias Básicas. Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia

ABSTRACT

Air pollution affects not only the air in cities but also extends to all indoor environments (homes, offices, schools, public places, transportation, etc.), where we spend between 80% and 90% of our time. Both indoor and outdoor air quality have emerged as significant health concerns and are integral to national strategies implemented by health and environmental institutes in each country. Recently, complaints regarding outdoor air quality have risen in cities, primarily due to automobile traffic and industrial activities in urban areas, and also indoors within homes, offices, and schools. The following paper presents a methodology for the calibration of low-cost monitoring stations based on measurements in a couple of cities in Colombia as part of the development of a project to reduce the environmental awareness gap in urban areas for the estimation of the air quality through low-cost, flexible, modular, and mobile air quality monitoring station design that could be used to assess air pollution in different indoor and outdoor environments. With the implementation of the low-cost stations, we have calibrated and evaluated the performance of the stations using usual linear regression methods, but we have also explored the use of unsupervised estimation with the help of machine learning algorithms, specifically with Random Forest estimators. We have found a significant improvement with using Random Forest for station calibration compared with those found using simple linear regressions for calibration effects. We have found that all the models offer a significant improvement in terms of RMSE. The regression model improves RMSE by up to 70%, while the multiple regression model does so by up to 73%. However, it is the Random Forest that shows the most remarkable improvement, with a reduction in RMSE of up to 86%.

ARTICLE HISTORY
Received 18 January 2024
Accepted 9 June 2024

KEYWORDS
Air quality; low-cost sensor; citizen science; calibration models

CONTACT M. Tarazona Alvarado jose.tarazona6@correo.uis.edu.co Escuela de Física, Universidad Industrial de Santander, Car 27 #9, Bucaramanga 680001, Colombia

1. Introduction

Atmospheric air quality assessment and detection of gaseous pollutants in urban areas and cities has recently been considered one of the important and growing research problems in recent decades (Bai et al. 2023; Mamun, Abdulla, and Rasit Yuce 2019; Mata et al. 2022; Singh et al. 2021; Sokhi et al. 2021). The need to have real measurement scenarios *in situ* using chemical analysers and sensors stations is a key task because it allows the generation of data that can be evaluated, quantified and analysed in real time (Baqer et al. 2023; Concas et al. 2021; Considine et al. 2023; Dai et al. 2023; Datta et al. 2020; Donnelly, Misstear, and Broderick 2015; H. Liu et al. 2021; Reis et al. 2023; Zhang et al. 2012) to establish scenarios and criteria so that, at another level of management, decision makers and policy-makers have the appropriate tools to apply mechanisms and create solutions based on accurate and updated information with technical and scientific arguments.

Furthermore, many cities in many countries take concrete steps in the consolidation and implementation of strategies, facilities and methodologies to evaluate spatially and temporally the air quality in their neighbourhoods, in order to estimate and solve problems related with human mobility, improvement of sanitation conditions in areas with high automobile traffic, or to limit this traffic in many areas in which high levels of environmental pollutants in the air have been reported. Such strategies contemplate static stations on main roads (Abu El-Magd et al. 2023; Ganji et al. 2023; Harleman et al. 2023; Nishitateno, Burke, and Arimura 2024), mobile stations in the city (Anjomshoaa et al. 2018; Apte et al. 2017; Blanco et al. 2022; Francesco et al. 2024; Ganji et al. 2023; Limon et al. 2023; Shaibal et al. 2015; Sonawani and Patil 2024; Van den Bossche et al. 2015), or even drones (Afshar-Mohajer and Wu 2023; Bagkis et al. 2021; Burgués and Marco 2023; Limon et al. 2023; Ranganathan et al. 2023; Renwick, Klein, and Hamann 2016; Wivou et al. 2016) that fly over critical areas equipped with rapid response sensors and with devices with wireless connections to visualize and analyse data in real time, which are used to give responses or establish immediate decision-making plans to alleviate critical conditions or in the medium term to strengthen policies and regulations appropriate to each situation. A really interesting case is from Google (Kerckhoffs, Khan, Hoek, Yuan, Hertel, et al. 2022; Apte et al. 2017), using the Google Street View vehicles equipped with a rapid response pollution measurement platform repeatedly sampling every street in a 30 km^2 area of Oakland, CA, collecting a large amount of urban air quality data with spatial precision 4 to 5 orders of magnitude greater than that possible with current environmental monitoring at the central site.

Some interesting applications also include the study and monitoring of air quality to promote and maintain healthy, safe, productive and comfortable student environments for students, teachers and school personnel (Annesi-Maesano et al. 2013; Becerra et al. 2020; de Gennaro et al. 2014; Ilie et al. 2022; Lee and Chang 2000; Smart et al. 2020; Wargocki et al. 2020); assessment of number concentrations, size distributions of ultrafine particles and time integrated black carbon, $PM_{2.5}$ mass, and chemical species on commercial flights (Baklanov and Zhang 2020; Hudda et al. 2020; Riley et al. 2021; Sher et al. 2021; Spengler and Wilson 2003; Waters et al. 2002); or also the indoor air pollution in hospitals associated with specific compounds emitted from various products used, health care activities and building materials (Amir et al. 2023; Baudet et al. 2022; Baurès et al. 2018; Jung et al. 2015; Mata et al. 2022).

The increase in low-cost detection to manage air pollution in cities has also been a constant in recent years in favour of commercial optical or chemical analysers. This is mainly due to several factors: the economic cost of professional stations is high (Marinov et al. 2016; Motlagh et al. 2020) and not accessible for some cities; the size of these stations is also large, it is necessary massive power consumption of sensor nodes and requires considerable space for installation (Castell et al. 2017; Jovašević-Stojanović et al. 2015), requiring adequate and appropriate spaces for their installation and they also consist of complicated devices that could not be easily accessed or maintained unless trained and skilled personnel are available.[1] Thus, several proposals for low-cost monitors and stations have emerged to overcome the above limitations. On the other hand, more accessible and low-cost stations allow greater investment for spatial monitoring covering more territories and for spatial and temporal analysis of air quality variability, especially in urban areas (Flores-Cortez, Adalberto Cortez, and Idalia Rosa 2019; Hernández-Gordillo et al. 2021; Marjovi, Arfire, and Martinoli 2017; Piersanti et al. 2015).

An underlying problem in the generation of low-cost monitoring stations is precisely the internal calibration process of the sensors. Although some global manufacturers and distributors ensure good performance conditions of individual specialized sensors, calibration is still an important part of the methodology to reduce production and development costs. Most of the time, low-cost stations measure adequately, but the sensitivity of the measurements may have adjustment problems, or malfunction in certain measurement ranges, or particularly exhibit a measurement lag mainly due to a deviation from the baseline measurement in the absence of contaminants (Feenstra et al. 2019; Karagulian et al. 2019).

Therefore, in the literature, it is abundant to find various ways to calibrate these sensors using statistical techniques (Cordero, Borge, and Narros 2018; Datta et al. 2020; Drajic and Gligoric 2020; Margaritis et al. 2021; Sahu et al. 2021) or neural networks (Esposito et al. 2016; Kow et al. 2022; Spinelle et al. 2014), or machine learning tools (Adong et al. 2022; Bagkis et al. 2021; De Vito et al. 2020).

The term 'low-cost monitoring station' can be defined in terms of several criteria that evaluate the affordability and effectiveness of these stations for measuring air quality. First, the initial cost of acquiring hardware and associated equipment is a critical factor (X. Liu et al. 2020). A low-cost station should be significantly less expensive than the more complex and specialized monitoring solutions used in professional applications. In addition to the initial cost, ongoing maintenance and operating expenses are also important. Low-cost solutions tend to have lower maintenance requirements and reduced operating costs compared to more advanced systems.

Simplicity of design is another key criterion. Low-cost stations typically use simple and inexpensive designs and components. This may include the use of commercially available sensors and basic data communication technologies (Badura et al. 2018; Jovašević-Stojanović et al. 2015; Spinelle et al. 2017). Although less expensive, the sensors used in these stations should provide accurate and reliable measurements of air quality parameters such as particulate matter (PM), gases (such as CO_2, NO_2, SO_2) and other relevant pollutants. Data quality and accuracy are important aspects to consider to ensure the usefulness and reliability of the measurements obtained.

The ability to scale and adapt these stations is also critical. Low-cost stations must be scalable, allowing multiple stations to be deployed in different locations without significant additional costs. This facilitates the creation of distributed monitoring networks that cover large and diversified areas, improving the coverage and representativeness of air quality measurements. In addition, these stations should be flexible and adaptable to different environmental conditions and specific monitoring requirements (Bai et al. 2023; Motlagh et al. 2020; Vitali, Arru, and Magnanini 2023).

A common approach for low-cost stations is their use in citizen or community science projects, where public participation in air quality monitoring is critical (Barros et al. 2023; Considine et al. 2023; Manshur et al. 2023). Therefore, ease of use and accessibility are key considerations. Easy access to the collected data and compatibility with standard analysis and visualization tools are also important aspects to maximize the usefulness and impact of these stations in citizen participation initiatives.

In the following note, we address the problem of calibrating low-cost sensors for particulate matter in a couple of cities in Colombia, where an awareness strategy has been promoted through the construction and distribution of low-cost stations in schools and official state institutions to narrow the gap in the estimation of urban air quality and advance in the construction of clean and smart cities. The main objective is to evaluate the effectiveness of three calibration models for PM_{10} and $PM_{2.5}$ measurements conducted by low-cost stations, emphasizing the significance of environmental conditions in the calibration process, as well as on-site meteorological variables.

2. Air quality monitoring stations

2.1. Low-cost stations descriptions

For the results below, we present the calibration procedures in two major cities in Colombia in which two different versions of low-cost stations have been installed.

The low-cost monitoring stations have been shipped and distributed in several locations in both localities and for calibration purposes, all stations have been placed during the first months in locations close to professional, robust and expensive monitoring stations, acquired by the Ministry of Environment, government and mayoral offices and other institutions related to the environmental quality of the cities through official contracts. The low-cost monitoring stations are located within 50 metres of the professional reference stations.

The first of the cities is Bucaramanga located in the northwest of the Department of Santander in the Colombian Andes. Bucaramanga has the fifth largest economy in Colombia and is the ninth most populated city in the country, with a population of 613,400 inhabitants (projection for 2023).

In this location, we consider the low-cost station called *RACIMO-AIRE*, which is a citizen science project born in 2018 from a project funded by the Ministry of Science and Technology of Colombia. It is installed at the Normal Superior school in Bucaramanga, Santander (7°07′38″N, 73°07′02″W, see Table 1). The station is one of the first versions using low-cost sensors for PM_{10} and $PM_{2.5}$. It is an open-source and open-hardware device, and more information about it can be found at https://github.com/JoseSalamancaCoy/RACIMO_AIRE.

Table 1. Main characteristics of both measurement locations and low-cost air quality monitoring stations.

Item	Station 1	Station 2
Model name	*RACIMO-AIRE* (S_1)	*EVA* (S_2)
Located in City	Bucaramanga	Bogotá
Coordinates (Lat (N), Lon (W))	(7.12, −73.11)	(4.70, −74.07)
Altitude (m.a.s.l)	990	2650
Average temperature (°C)	26	13
Installation building	School	Road way
Station kind	Urban-Background	Urban-Traffic

On the other hand, the second low-cost station is located in Bogotá, which is the capital of the Republic of Colombia and the department of Cundinamarca. It is made up of 20 localities or districts and is the political, economic, administrative, industrial, artistic, cultural, sports and tourist epicentre of the country. It is the third highest capital city in the world, at an average of 2625 metres above sea level. The low-cost station considered en Bogotá is named *EVA* (4°42′00″N, 74°04′12″W, see Table 1).

Both projects also include a proposal for work in citizen science for the empowerment of resources, methodologies, community participation, data analysis and generation of human resources in data analysis. The main characteristics of both measurement locations and the stations in operation are presented in Table 1.

2.1.1. *Sensor specifications for air quality monitoring*

To monitor air quality at low-cost stations in Bucaramanga and Bogotá, Colombia, several specialized sensors are used. Below are the technical characteristics and specific uses of each sensor in the context of the monitoring stations.

The Sensirion SHT35 sensor is utilized for measuring humidity and temperature. This sensor operates in a humidity range of 0% to 100% RH and a temperature range of −40°C to 90°C, with an accuracy of ±1.5% RH and ±0.1°C, respectively. To allow the assessment of redundancy in measurements in the stations, we include another sensor BMW280, essential for validating data quality and accuracy.

The LTR-390 UV-01 sensor is used for measuring illuminance. This sensor can detect both UV and visible light, making it particularly useful for outdoor environmental studies where these measurements are crucial. The LTR-390 UV-01 is sensitive to both ultraviolet and visible light, offering a fast and accurate response, enabling assessment of solar light exposure, a significant factor in air quality and environmental comfort.

The PMS7003 is a particulate matter sensor that measures concentrations of PM_1, $PM_{2.5}$, and PM_{10}, which are particles with aerodynamic diameters of ≤1, 2.5, and 10 micrometres, respectively. This sensor is critical for evaluating air quality, as fine particles can penetrate the lungs and cause respiratory issues. The stations are equipped with two units of this sensor, providing a way to verify measurement consistency and accuracy by comparing data collected from each sensor.

The BME280 is a small sensor that measures atmospheric pressure, humidity, and temperature. Used in conjunction with the SHT35, it provides a secondary data source for humidity and temperature, enhancing measurement reliability. The BME280 operates with an accuracy of ±1 hPa for pressure, ±3% RH for humidity, and ± 1°C for temperature. This sensor is invaluable for correlating meteorological changes with variations in air quality.

The use of these sensors in the monitoring stations of Bucaramanga and Bogotá enables the collection of data on various environmental parameters, which is crucial for assessing air quality and designing appropriate strategies for improvement. Redundancy in particulate matter, humidity, and temperature sensors ensures measurement validity, providing a solid foundation for environmental decision-making.

2.2. Reference stations

The reference stations belong to the air quality monitoring networks in their respective cities. In Bucaramanga, we considered a station from the Metropolitan Aqueduct of Bucaramanga, ABM, located at Colegio Normal Superior Sede C, at 27th Street #29–69. This station is part of the air quality surveillance system – Type III (SVCA) that the city is equipped with, and it became operational in October 2018. For more details, please consult this https://www.amb.gov.co/calidad-del-aire/. Meanwhile, in the city of Bogotá, we considered the 'Las Ferias' station, located at 69Q–50, 80 Avenue (4°41'26.52'N, 74°4'56.94'W). This station belongs to the Bogotá Air Quality Monitoring Network (RMCAB), a system composed of 19 regulatory grade stations around the city which has been in operation since 1997. For more details, please consult this http://rmcab.ambientebogota.gov.co/home/map.

The reference monitoring stations consist of monitors, analysers, and automatic sensors that collect data every hour on the state of air quality in several locations. This information is stored and sent via the internet to the central server of the *Secretaría de Ambiente*. It is prevalidated there and subsequently published in real-time on the entity's website. The monitors and analysers in the network operate under specific measurement methods established in Title 40 of the CFR (Code of Federal Regulations) (US 2023), which are approved by the United States Environmental Protection Agency (EPA). For each pollutant, a specific reference method is defined, in accordance with the equivalent technique for the operation of each monitor, also established in the appendices of Part 50 of Title 40 of the CFR, in case of particulate matter, they are measured through a MetOne Betta Attenuation Monitor (BAM 1020), an U.S. EPA equivalent method for automatic $PM_{2.5}$ or PM_{10} monitoring. The BAM 1020 has a measurement cycle consisting of beta irradiation for 8 min at the beginning (zero reading) and 8 min at the end of each hour with an air sampling period of 42 min between measurements using a glass fibre filter tape. The attenuation of beta ray due to trapped particles is used to determine the mass of them on the fibre for calculating the volumetric concentration in micrograms per cubic metre.

Since 2018, RMCAB is governed by what is established in Resolution 2254 of 2017 from the Ministry of Environment and Sustainable Development, which sets the air quality standard or immission level and adopts provisions for the management of this resource in the national territory. The goal is to ensure a healthy environment and minimize the risk to human health caused by pollutants in the atmosphere. The document establishes the maximum permissible levels for criteria pollutants at different times and exposure scales for the declaration of environmental alerts, ranges and conditions of the air quality index, definition of source areas, and mechanisms for result dissemination.

The composition of RMCAB is based on the guidelines defined in the system design manual for air quality monitoring, the Protocol for Air Quality Monitoring and

Follow-up, adopted through Resolution 650 of 2010 by the then Ministry of Environment (Ministry of Environment and Sustainable Development 2017), Housing, and Territorial Development. It includes aspects such as station locations, monitor and analyser types, equipment maintenance and verification activities, and data analysis processes, among others. Additionally, the operational and design guidelines established in the Quality Assurance Handbook for Air Pollution Measurement Systems by the Environmental Protection Agency (EPA) of the United States are considered.

3. Description of data used

In this study, data collected during specific periods in Bogota and Bucaramanga, Colombia, were analysed, considering the drought and rainfall seasons characteristic of each region.

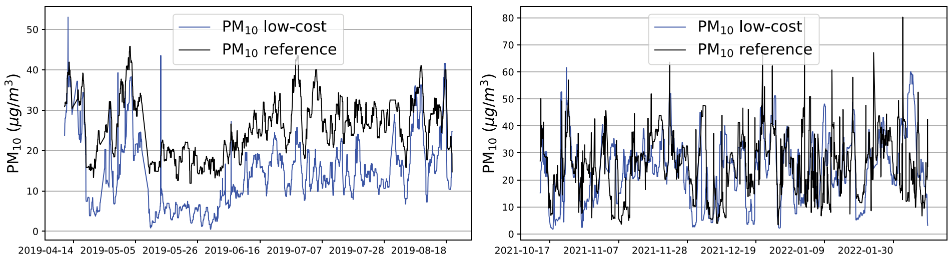
In Bogota, the data used covers the period from 14 October 2021 to 9 February 2022, which mainly covers the dry season in the region. The main dry season in Bogota extends from December to April, with less than 3 days of rain per month during these months. In July, there is again a reduction in rainfall, although less pronounced. May and June, as well as August to November, are the rainiest months in Bogota. The rainy season in this city lasts approximately 8 months, from March to December, with April and November being the rainiest months. Throughout the year, rainfall is distributed in two dry seasons and two rainy seasons, with the months of January, February, July and August being predominantly dry. The rainy seasons extend from late March to early June and from late September to early December.

For Bucaramanga, the data corresponds to the period from 11 April 2019 to 20 August 2019. This time it also covers the main dry season in Bucaramanga, which is recorded in the months of December, January and February, with less than 5 days of rain per month. During the months of May to November, rains are more frequent, with an average of 10–15 days of rain per month. The main rainy season in Bucaramanga runs from June to November, with a rainfall frequency of 20–24 days per month during this period. May and December act as transition months with moderate rainfall.

It is important to note that, although the data come from a single time period at each location, climatic conditions in Colombia tend to be relatively stable throughout the year due to its location in the tropics. This implies that abrupt climatic variations are less common compared to mid- or high latitude areas. These climatic factors provide an important context for the interpretation of air quality data and other environmental parameters collected during the study.

During the periods analysed, there is a consistent and detailed dataset for Bogota and Bucaramanga. In the case of Bogota, a total of 2846 data records are available, while for Bucaramanga there are 3143 records. These data were collected for the low-cost stations with a measurement frequency of 2 min, but because the reference stations only provide a sampling frequency of 1 h, then the data have been resampled to the maximum temporal resolution of 1 h, which provides an accurate temporal resolution in both data sets.

Regarding the completeness of the data, it is observed that in Bogotá, approximately 13.70% of the data have missing values, while in Bucaramanga, this figure drops to 7.57%.



(a) Bucaramanga (b) Bogotá

Figure 1. Sequence of measurements for PM_{10} and $PM_{2.5}$ particulate matter sensors for a low-cost station and a reference station located in Bucaramanga between April and September 2019 (Panel (a), left). Similarly, measurements from a low-cost station and a reference station located in Bogotá between October 2021 and January 2022 are presented (Panel (b), right). Both sets of stations are located within a few meters of each other.

Most of these omissions are attributed to occasional events of malfunctioning of the monitoring stations. However, it is important to note that this amount of missing data represents a relatively small proportion compared to the regularity and high temporal resolution of the data available for the study. These conditions and nearly continuous data provide a solid basis for the analysis and evaluation of air quality and meteorological conditions in Bogota and Bucaramanga during the selected periods, despite the presence of some missing data due to eventual failures of the monitoring stations. For the subsequent analysis, these missing data have been removed from the dataset to make complete and accuracy models.

In addition to PM_{10} and $PM_{2.5}$ particulate measurements, the low-cost monitoring stations also recorded various meteorological variables, including temperature, humidity, and atmospheric pressure, irradiance and illuminance. These additional environmental data are essential to better understand the context in which air quality measurements take place.

Figures 1 and 2 present the historical measurements for PM_{10} and $PM_{2.5}$ sensors for the low-cost and reference stations in the corresponding time span between April and September 2019.

4. Methods and techniques

For the purpose of evaluating the performance of the low-cost monitoring station under field operation conditions, we have used and applied three calibration methods: simple linear models (LR), multivariate linear estimation (MLR) and finally use of Machine Learning technique specifically with Random Forest estimators with decision trees.

4.1. Linear regression (LR)

The first method we use to calibrate stations consists in the calibration functions that were established by ordinary linear regression, using the minimization of the residuals of the low-cost sensor responses with respect to the reference measurement. In this case, calibration is done as usual using the simple model

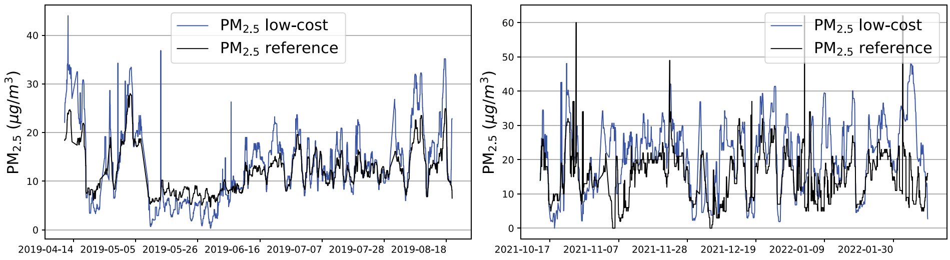


Figure 2. Sequence of measurements for PM_{10} particulate matter sensors for a low-cost station and a reference station located in Bucaramanga between April and September 2019 (Panel (a), left). Similarly, measurements from a low-cost station and a reference station located in Bogotá between October 2021 and January 2022 are presented (Panel (b), right). Both sets of stations are located within a few meters of each other.

$$\hat{y}_i = \alpha_0 + \alpha_1 y_i, \tag{1}$$

where $\hat{y}_i$ represents the low-cost sensor responses calibrated, α_0 and α_1 represent the intercept and slope of the least-squares fit, respectively, and y_i the corresponding reference measurements.

4.2. PCA

Principal component analysis (PCA) (Abdi and Williams 2010; Bro and Smilde 2014) is a statistical technique used to reduce the dimensionality of a data set by identifying the most important variables or components that explain the variability in the data. This technique is widely used in data science and can be applied to different areas, including air quality.

To understand and analyse air quality, meteorological data such as temperature, humidity, wind speed, wind direction, solar radiation, among others, are collected as these factors can influence the dispersion and concentration of air pollutants.

In this section, we apply PCA to establish the importance of meteorological variables for the generation of low-cost sensor calibration models. PCA can be used to analyse meteorological data in terms of air quality. For example, a meteorological data set that includes multiple variables such as temperature, humidity, wind speed and direction can be complex and difficult to analyse individually. PCA can help simplify these data by identifying the most important variables that explain the variability in the data.

In other words, PCA can also be useful in air quality modelling and prediction. For example, variables identified using PCA can be used as inputs to weather forecasting models or pollutant dispersion models, which can improve the accuracy of air quality predictions.

PCA is a useful statistical technique for analysing meteorological data in terms of air quality. It allows us to identify the most important variables that explain variability in the data, simplify data interpretation, and find relationships or patterns in meteorological data that may influence air quality. ACP can be a valuable tool in air quality management, helping to understand and predict the factors that influence air pollution and make informed decisions to mitigate its effects.

Figure 3 shows the process of extracting the main characteristics for the importance of the meteorological variables that can be used as good estimators in the generation of robust (linear or nonlinear) calibration models of low-cost stations.

Additionally, in Figure 4 we can observe that the most important variables that can help to develop forecast models as well as calibration models for low-cost stations are essentially air temperature, solar radiation (negatively correlated with component 1 and 2), as well as precipitation and relative humidity (positively correlated with component 1 and negatively correlated with component 2). These last four variables account for most of the total variance explained by the variation of the data from the monitoring stations and represent the most important predictors for model generation.

4.3. Multivariate linear regression (MLR)

Calibration was performed using the least squares method taking into account more than one predictor variable. The models were established by adding ambient temperature, pressure and relative humidity information also collected with the sensors arranged in the low-cost station. As in the case of the LR, the calibration functions consisted of equations of multiple linear type

$$\hat{y}_i = \alpha_0 + \alpha_1 y_i + \sum_{j=2}^{4} \alpha_j y_{j,i}, \tag{2}$$

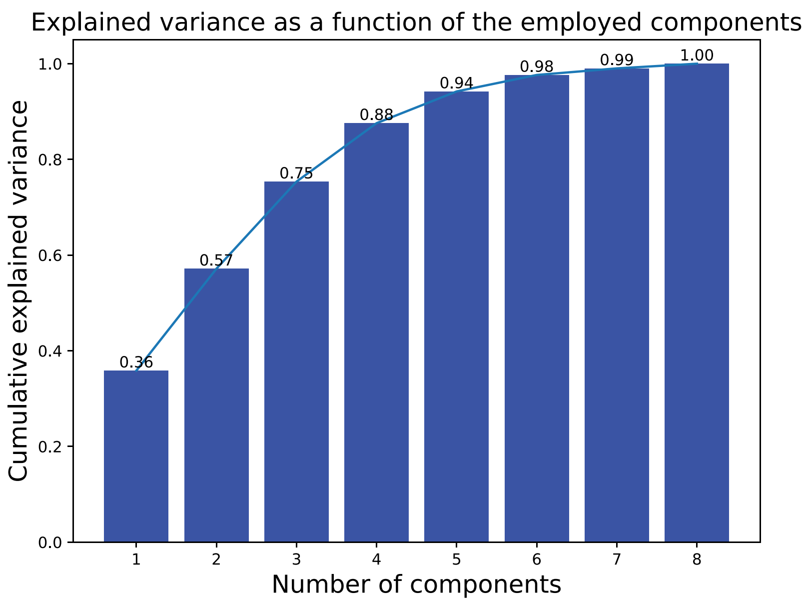


Figure 3. Percentage of cumulative significance for the components analyzed using meteorological data from air quality monitoring stations.

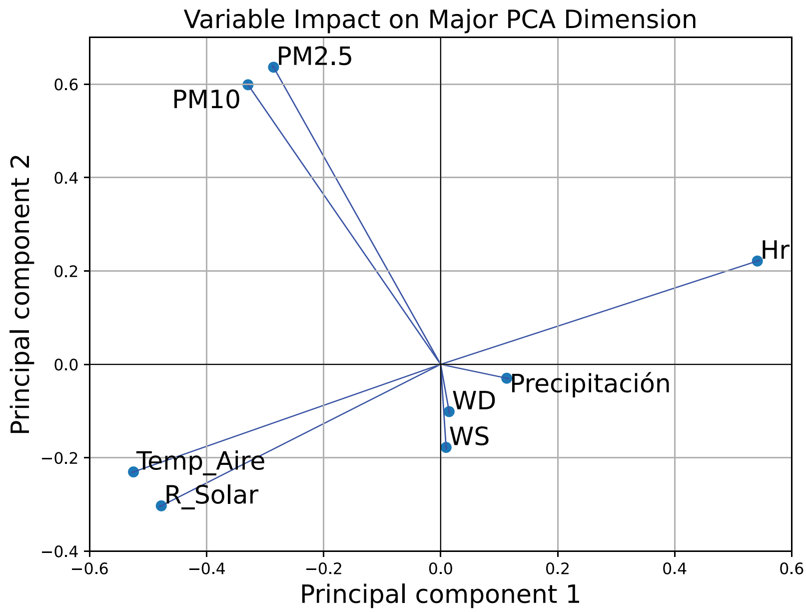


Figure 4. Total variance explained for variables related to particulate matter pollutants along two principal components.

where $\hat{y}_i$ represents the low-cost sensor responses, α_0 and α_1 represent the intercept and slope of the least-squares fit, respectively, for the corresponding reference measurements (y_i) and α_2, α_3 and α_4 represent the corresponding weights for the temperature (y_2), pressure (y_3) and relative humidity (y_4) variables, respectively, according to more important variables in results in Figures 3 and 4 in above section.

4.4. Random Forest (RF)

In the analysis procedure, explanatory methods should be used in accordance with the existing data. Thus, the best model is the one that identifies the correct model complexity. Therefore, in addition to the types of procedures that were reviewed in the development of the present project (e.g. linear regression, and multiple linear regressions); finally, the use of the family of decision tree algorithms has also been considered. Specifically, the Random Forests algorithm (Breiman 2001) was selected because it is fast, computationally effective, robust in the presence of noisy data, offers possibilities for the explanation of the data, and error estimates can be made (Tsymbal, Pechenizkiy, and Cunningham 2006). Moreover, it is a comparable method and, according to the tests carried out, in most cases it has a better predictive capacity than classical regression methods. Finally, it is a non-parametric procedure that is very useful when there are many correlated variables and few data, or when there are complex interactions between predictor variables or when there is a lot of missing data. Random Forests is a classifier-regressor that consists of an independent, identically distributed, random collection of classifiers organized into trees, with each tree contributing a single vote. Basically, for clustering or regression problems, Random Forests randomly selects a subset of the attributes and then re-selects the best cut among them. Subsequently, the process is repeated for each of the trees (many trees grow in the same way) to build a forest. Finally, all the trees are used in the final result from the average of the results of each of the trees.

Once the model has been made and calculated a file is stored on the hard disk and subsequently included in the software structure of the low-cost station to translate the sensor's original measurements to the calibrated measurement with the model. This ensures consistency in future measurements.

4.5. Performance evaluation methods

To establish comparisons between calibration models for the station sensors, we have used a set of error estimators that could evaluate the performance of the calibrations made at the low-cost stations. The usual error estimators such as mean absolute error, R^2 score, accuracy, mean square error have been used in this performance evaluation. Additionally, we will use as an estimator for the robustness evaluation of Goodness-of-Fit of Regressions the metric defined by the root mean squared error (RMSE), described by $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$.

Additionally, to measure the behaviour of the models tested and compare results, we use the R^2-Score metric that measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It is usually also known as the Coefficient of Determination.

Finally, we have also used the correlation coefficient calculated between the time series for the particulate matter measurements before and after calibration, that is, the measure of linear association between the original measurements (from the reference station) and the measurements resulting from the application of each of the models to the original measurements from the low-cost station.

5. Results and discussion

Considering the data described in the previous section, we first show some descriptive statistics for the data collected prior to the performance and application of the calibration methods.

The descriptive statistics of air quality measurements from low-cost stations compared to reference stations in Bogotá reveal several insights. In terms of particulate matter (PM_{10} and $PM_{2.5}$), the average concentrations are slightly lower in low-cost stations compared to reference stations, with PM_{10} averaging 24.60 $\mu g/m^3$ and $PM_{2.5}$ averaging 21.81 $\mu g/m^3$ in low-cost stations, versus 25.63 $\mu g/m^3$ and 15.39 $\mu g/m^3$ respectively in reference stations. The variability indicated by the standard deviation, is also comparable between the two types of stations. The range of PM_{10} values is between 1.75 $\mu g/m^3$ and 61.58 $\mu g/m^3$ in low-cost stations, similar to the range in reference stations (from 3.4 $\mu g/m^3$ to 80.2 $\mu g/m^3$), for PM_{10}. Meanwhile, the range of $PM_{2.5}$ values is from 0 $\mu g/m^3$ to 48.09 $\mu g/m^3$ in low-cost stations, while in reference stations, it ranges from 0 $\mu g/m^3$ to 62 $\mu g/m^3$.

The average temperature recorded in low-cost stations (15.29 °C) is similar to that in reference stations (15.37 °C), with a relatively low standard deviation (approximately 0.92 °C for both low-cost and reference). The temperature range is from 10.98 °C to

22.48 °C similar both in low-cost stations and reference stations. The mean humidity is slightly higher in low-cost stations (66.05%) compared to reference stations (63.34%), with a standard deviation of approximately 7.17%. The humidity ranges from 36.42% to 78.40% in both low-cost stations and reference stations.

In other words, the average irradiance measured in low-cost stations (22.52 W/m^2) is consistent with that in reference stations, with a standard deviation of approximately 38.16 W/m^2. The range of irradiance values varies from 0 W/m^2 to 798.56 W/m^2.

Results from basic descriptive statistics for Bucaramanga city are similar that the above. In terms of particulate matter (PM_{10} and $PM_{2.5}$), the average concentrations are slightly lower in low-cost stations compared to reference stations. Specifically, PM10 averages around 14.28 µg/m^3 and PM2.5 averages approximately 13.26 µg/m^3 in low-cost stations, while reference stations show slightly higher values with PM10 averaging 25.69 µg/m^3 and PM2.5 averaging 12.12 µg/m^3. The variability, as indicated by the standard deviation, varies across these measurements, in case of PM_{10} and $PM_{2.5}$ with 8.37 µg/m^3 and 7.67 µg/m^3 for low-cost stations, and 7.14 µg/m^3 and 4.57 µg/m^3 for reference stations, respectively.

Regarding temperature and humidity, the average air temperature recorded in low-cost stations (30.87 °C) is higher than in reference stations (23.79 °C), with a relatively low standard deviation (approximately 1.26 °C for both low-cost and reference). Humidity levels are also comparable, with slightly higher mean relative humidity in reference stations (74.17%) compared to low-cost stations (74.17%), showing moderate variability within the data.

No significant variation is exhibited for solar radiation measurements with the relation of reference stations recording a wider range of values (from 0.4 W/m^2 to 233.1 W/m^2) compared to low-cost stations.

In the first instance, a simple linear regression using least squares methods has been used to adjust the values of the interesting variables of the low-cost stations, PM_{10} and $PM_{2.5}$ in both cities in terms only of the corresponding variables measured by the reference stations. The results of such a calibration procedure for each variable of the four stations available in the testing phase are shown in Table 2 as well in Figures 9 and 10. The coefficient of determination (R^2) as well as for each calibration is given in Table 2.

To generalize the results obtained from linear regression calibration, we proceed with multiple linear regression using additional predictor variables beyond PM_{10} and $PM_{2.5}$ from reference stations. Specifically, we incorporate temperature, humidity, pressure, and solar irradiance as predictor variables. These auxiliary variables are selected based on the outcomes established through dimensionality reduction using PCA (see Figure 3 and 4). By including these variables in our regression model, we aim to enhance the robustness

Table 2. Summary of calibration results of low-cost station with linear regression method in field using Equation 1.

Parameter/Species	PM_{10} BGA	$PM_{2.5}$ BGA	PM_{10} BOG	$PM_{2.5}$ BOG
a_0 (intercept)	15.6514	4.6797	17.9986	8.1577
a_1 (slope)	0.7056	0.5600	0.3174	0.3377
RMSE	3.7808	1.5595	10.0832	6.1013
R^2 score	0.7011	0.8725	0.1410	0.2310
Correlation (r)	0.8374	0.9340	0.3774	0.4812

and accuracy of our air quality predictions. This approach leverages comprehensive environmental data to develop a more holistic understanding of the relationships between pollutant concentrations and key meteorological factors, facilitating more nuanced and reliable air quality assessments across the monitoring network. The results and coefficient of determination (R^2) for each calibration are given in Table 3.

Tables 2 and 3 show the calibration evaluation using linear regressions and multiple linear regression, respectively, for entire time intervals of data. In the case of PM_{10}, the regression coefficients for linear regression are $\alpha_0 = 15.6514$ (intercept), $\alpha_1 = 0.7056$ (slope), and for $PM_{2.5}$ the coefficients are $\alpha_0 = 4.6797$ (intercept), $\alpha_1 = 0.5600$ (slope); while for the multiple linear regression we have for PM_{10} the coefficients 526.3294, 0.7537, −0.9931, −0.4952, −0.6671, 0.0016 and for $PM_{2.5}$ the coefficients 500.2963, 0.5820, −0.4402, −0.5208, −0.2482, −0.0007, respectively, for intercepts and weight for specie variable, temperature, pressure, relative humidity and solar radiation.

For the case of using Random Forest we have used $N_f = 4$ features, $N_e = 800$ estimators, and in addition, we have used 50% of the data to construct the random data set for training and 50% of the data for testing. The results of this calibration with these parameters have been summarized in Table 4.

The summary of the calibration results obtained from the Random Forest model for low-cost stations provides valuable information on the predictive performance and significance of features across different parameters and pollutant species in the field.

First, the feature importance analysis reveals that the most significant predictors for PM_{10} and $PM_{2.5}$ measurements at the Bucaramanga station are sensor-specific parameters, indicating a strong dependence on direct sensor readings for accurate predictions. In contrast, for the Bogota station, sensor-specific parameters are still important, but show relatively less importance compared to temperature, pressure, relative humidity and

Table 3. Summary of calibration results of low-cost station with linear multiple regression method in field.

Parameter/Species	PM_{10} BGA	$PM_{2.5}$ BGA	PM_{10} BOG	$PM_{2.5}$ BOG
α_0 (intercept)	526.3294	500.2963	1121.0498	263.8161
α_1 (specie slope)	0.7537	0.5820	0.2818	0.2999
α_2 (Temperature)	−0.9931	−0.4402	−0.2563	0.1239
α_3 (Pressure)	−0.4952	−0.5208	−1.4726	−0.3526
α_4 (Rel. Humidity)	−0.6671	−0.2482	0.1355	0.1282
α_5 (R. Solar)	0.0016	−0.0007	0.0152	0.0015
RMSE	3.3674	1.3267	9.9632	6.0091
R^2 score	0.7787	0.9127	0.1571	0.2474
Correlation (r)	0.8824	0.9554	0.3984	0.4982

Table 4. Summary of calibration results of low-cost station with Random Forest estimator in field.

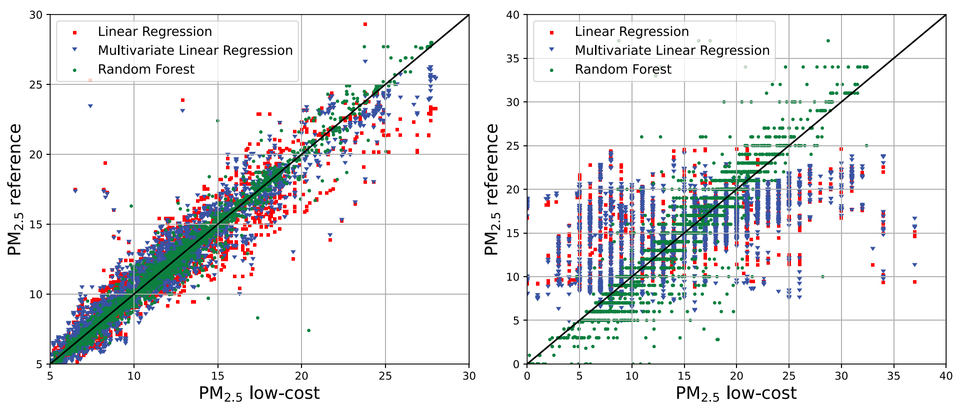
Parameter/Species	PM_{10} BGA	$PM_{2.5}$ BGA	PM_{10} BOG	$PM_{2.5}$ BOG
Import. Param. (sensor specie)	0.8171	0.9023	0.3541	0.4523
Import. Param. (Temperature)	0.0339	0.0278	0.1816	0.1592
Import. Param. (Pressure)	0.0369	0.0257	0.1791	0.1404
Import. Param. (Rel. Humidity)	0.0924	0.0338	0.1973	0.1769
Import. Param. (R. Solar)	0.0193	0.0101	0.0878	0.0710
RMSE	1.7905	0.9617	6.1498	5.0392
Accuracy (%)	95.4243	95.3937	78.2196	79.4427
R^2 score	0.9375	0.9556	0.8293	0.8430
Correlation (r)	0.9682	0.9776	0.8251	0.8267

solar irradiance. This suggests that environmental factors play a more substantial role in the prediction of air pollutant concentrations in Bogotá.

However, having very low relative importance in RF model, it takes the data from the reference station of the specific contamination sensor and is able to predict the correct value with which the calibration works. In other words, what this means is that the environmental factors are important but the model is able to calibrate itself basically with the measurement of the corresponding pollution sensor. This is interesting since this means that if the low-cost station is moved to another location with different conditions the calibration could continue to work for much longer.

The RMSE values provide information on the accuracy of the model in predicting pollutant concentrations. The lower RMSE values for $PM_{2.5}$ (see Figures 5–7), in both Bucaramanga and Bogotá, indicate higher accuracy compared to PM_{10} (see Figures 8–10), with Bucaramanga achieving slightly better accuracy overall.

Model performance is further evaluated by accuracy percentages and R^2 scores, which indicate high levels of predictive accuracy for both PM_{10} and $PM_{2.5}$ measurements in Bucaramanga and Bogotá. Strong correlation coefficients (r) further support these



(a) Bucaramanga (b) Bogotá

Figure 5. Dispersion of measurements of the low-cost station with respect to the reference station for $PM_{2.5}$ species (after calibration) for all calibration methods explored.

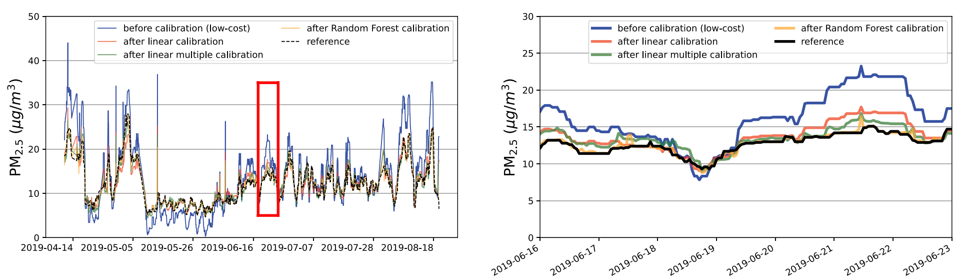


Figure 6. Evaluation of the calibration models in PM_{10} measurements for the low-cost station in terms of the reference station, after calibration and before calibration.

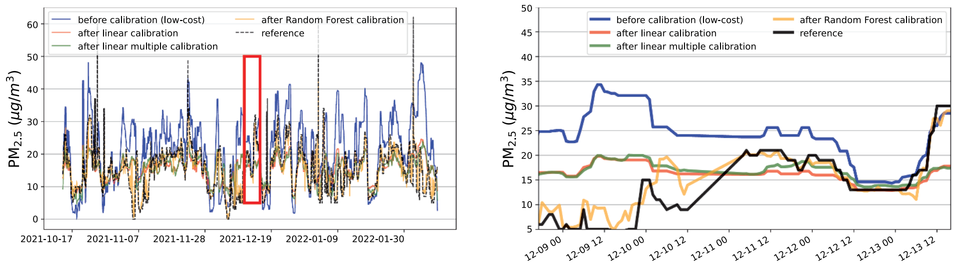
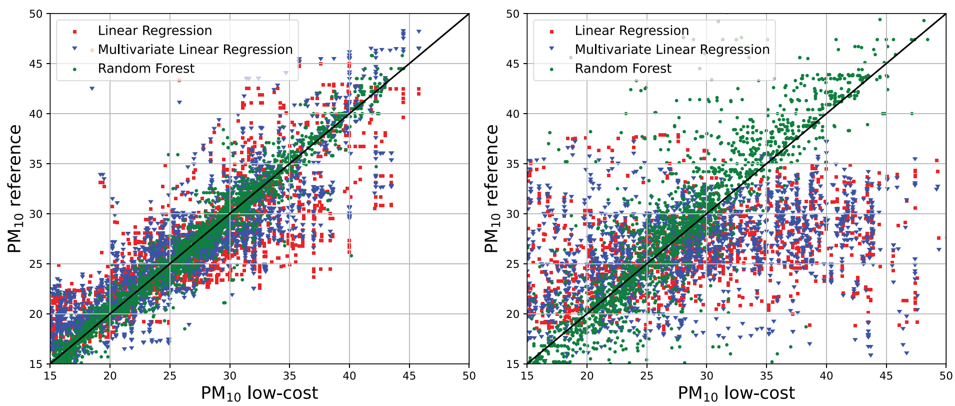


Figure 7. Evaluation of the calibration models in PM25 measurements for the low-cost station in terms of the reference station, after calibration and before calibration.



(a) Bucaramanga (b) Bogotá

Figure 8. Dispersion of measurements of the low-cost station with respect to the reference station for PM_{10} species (after calibration) for all calibration methods explored.

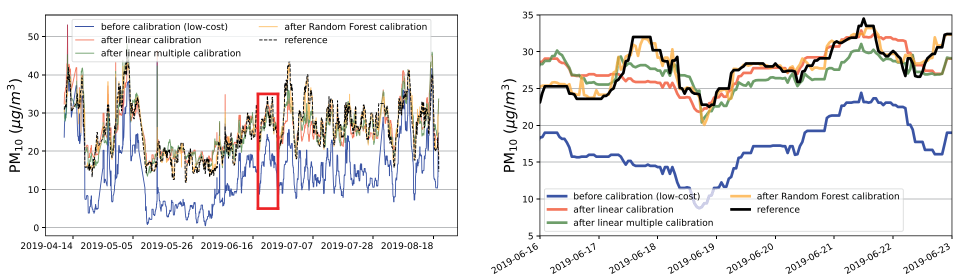


Figure 9. Evaluation of the calibration models in PM_{10} measurements for the low-cost station in terms of the reference station, after calibration and before calibration.

findings, highlighting robust relationships between predicted and observed values for air pollutants.

The Random Forest model demonstrates promising results in predicting air pollutant concentrations at low-cost stations, with different levels of importance assigned to specific sensor parameters and environmental factors at different locations. The high accuracy and strong correlations suggest that the incorporation of multiple predictors

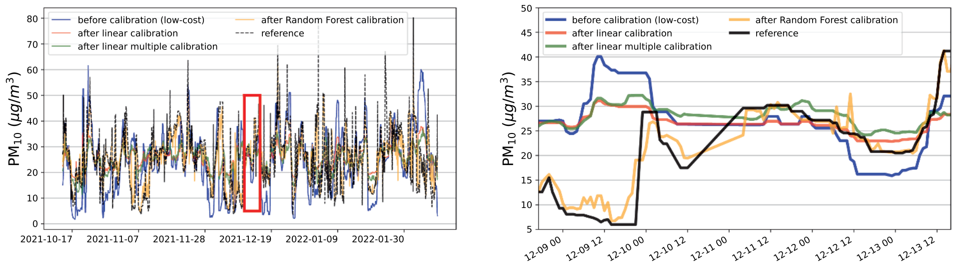


Figure 10. Evaluation of the calibration models in PM$_{10}$ measurements for the low-cost station in terms of the reference station, after calibration and before calibration.

improves the predictive capabilities of the model, underscoring the potential of machine learning techniques for reliable air quality monitoring and assessment in urban environments.

One of the big challenges of using low-cost sensors is that they are very susceptible to significant variations and changes with respect to temperature and relative humidity (Aleixandre and Gerboles 2012; Mead et al. 2013; Schneider et al. 2017) in the near environment. The manufacturers of the sensors, we use in our low-cost stations bring with them generic data describing the relationship between sensor current response, temperature and relative humidity, which are supplied by the manufacturer. Although we have considered these corrections in our stations, we can see that there are conditions that cannot be understood from linear (multiple) regression algorithms, although certainly with much more optimization with machine learning algorithms, which are able to understand nonlinear models of data behaviour from many descent trees.

Our results demonstrate that each sensor exhibits a unique response pattern, necessitating individualized and customized calibration for accurate field deployment, but unique, because environmental factors, although they help the model, are relatively small in importance. Frequently, manufacturer-provided information is incomplete for effectively compensating sensor behaviour under real-world measurement conditions, characterized by significant fluctuations in temperature and relative humidity. This knowledge gap underscores the importance of recent studies utilizing post-processing techniques such as multiple regression, neural networks, and machine learning to mitigate the impact of environmental variables on the performance of low-cost sensors (Morawska et al. 2018; Spinelle et al. 2015, 2017).

The calibration process involved the application of various techniques, including simple linear models (LR), multivariate linear estimation (MLR), and the utilization of machine learning, specifically Random Forest estimators. The results demonstrated a marked improvement in calibration accuracy when employing Random Forest models. Notably, all models exhibited significant enhancements in terms of RMSE, with Random Forest showing the most remarkable reduction, up to 86%. This underscores the potential of machine learning algorithms in addressing calibration challenges associated with low-cost sensors.

The results demonstrate that Random Forest allows for calibration where meteorological variables have relatively little importance. While this may pose a challenge for environmental conditions with significant differences, calibration can still be valid in other scenarios with similar environmental conditions. In citizen science projects employing these devices, it is common to have a pair of reference equipment (very expensive) for cities with similar environmental conditions, which do not cover the entire area of interest, and supplement them with low-cost stations. Thus, calibration is carried out with the nearest reference equipment before distributing the low-cost stations in the city.

This study underscores the importance of addressing air pollution, not only in outdoor environments but also within indoor spaces, where people spend a significant proportion of their time. The development of low-cost, flexible air quality monitoring stations represents a promising approach to assess air pollution in diverse settings. Calibration plays a crucial role in ensuring measurement accuracy, and machine learning algorithms, particularly Random Forest, offer substantial improvements in calibration results. However, the unique responses of sensors require customized calibration, emphasizing the complexity of mitigating environmental influences on low-cost sensor data. Despite these challenges, this research demonstrates the feasibility of maintaining stable and accurate calibration over an extended period, providing valuable insights for ongoing air quality monitoring efforts.

Note

1. In the case of Colombia, we have detailed a complete and concise protocol to management of monitor stations, analysis, operation, etc, for example: https://www.minambiente.gov.co/wp-content/uploads/2021/06/Protocolo_Calidad_del_Aire_-_Manual_Diseno.pdf.

Acknowledgements

The authors would like to express our sincere gratitude to the invaluable contributors who made this research possible. We extend our deepest thanks to MakeSens, the company behind the development of low-cost monitoring stations. Their dedication to providing innovative tools for air quality assessment has been instrumental in the success of this project. We appreciate their support and the critical role their technology plays in expanding our understanding of air pollution in both indoor and outdoor environments.

We also extend our heartfelt thanks to the Departmental Air Quality Monitoring Systems in Bogotá and Bucaramanga for providing the essential data from the reference stations. Their commitment to maintaining comprehensive air quality data allowed us to develop and refine our calibration models, significantly enhancing the accuracy and reliability of our measurements. Without their data, this research would not have been possible. We are deeply appreciative of their collaboration and their contribution to advancing our knowledge of air quality and its impact on public health and the environment.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

M. Tarazona Alvarado http://orcid.org/0000-0001-9935-1094
L. A. Núñez http://orcid.org/0000-0003-4575-5899

J. Pisco-Guabave http://orcid.org/0000-0002-3714-2032
D. Sierra-Porta http://orcid.org/0000-0003-3461-1347

References

Abdi, H., and L. J. Williams. 2010. "Principal Component Analysis." *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4): 433–459. https://doi.org/10.1002/wics.101.

Abu El-Magd, S., G. Soliman, M. Morsy, and S. Kharbish. 2023. "Environmental Hazard Assessment and Monitoring for Air Pollution Using Machine Learning and Remote Sensing." *International Journal of Environmental Science and Technology* 20 (6): 6103–6116. https://doi.org/10.1007/s13762-022-04367-6.

Adong, P., E. Bainomugisha, D. Okure, R. Sserunjogi, and P. Nachev. 2022. "Generative Model-Enhanced Human Motion Prediction." *Applied AI Letters* 3 (2): e76. https://doi.org/10.1002/ail2.63.

Afshar-Mohajer, N., and C.-Y. Wu. 2023. "Use of a Drone-Based Sensor As a Field-Ready Technique for Short-Term Concentration Mapping of Air Pollutants: A Modeling Study." *Atmospheric Environment* 294:119476. https://doi.org/10.1016/j.atmosenv.2022.119476.

Aleixandre, M., and M. Gerboles. 2012. "Review of Small Commercial Sensors for Indicative Monitoring of Ambient Gas." *Chemical Engineering Transactions* 30 (10.3303): 169–174.

Amir, F., H. Muhammad, N. A. Abdullah, S. Rizal, R. Thaib, and H. Umar. 2023. "Performance Analysis of Heat Recovery in Heat Pipe Heat Exchanger on Room Air Conditioning Systems." *AIMS Energy* 11 (4): 612–627. https://doi.org/10.3934/energy.2023031.

Anjomshoaa, A., F. Duarte, D. Rennings, T. J. Matarazzo, P. deSouza, and C. Ratti. 2018. "City Scanner: Building and Scheduling a Mobile Sensing Platform for Smart City Services." *IEEE Internet of Things Journal* 5 (6): 4567–4579. https://doi.org/10.1109/JIOT.2018.2839058.

Annesi-Maesano, I., N. Baiz, S. Banerjee, P. Rudnai, S. G. Solenne Rive. 2013. "Indoor Air Quality and Sources in Schools and Related Health Effects." *Journal of Toxicology and Environmental Health Part B* 16 (8): 491–550. https://doi.org/10.1080/10937404.2013.853609.

Apte, J. S., K. P. Messier, S. Gani, M. Brauer, T. W. Kirchstetter, M. M. Lunden, J. D. Marshall, C. J. Portier, R. C. Vermeulen, and S. P. Hamburg. 2017. "High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data." *Environmental Science & Technology* 51 (12): 6999–7008. https://doi.org/10.1021/acs.est.7b00891.

Badura, M., P. Batog, A. Drzeniecka-Osiadacz, P. Modzel. 2018. "Evaluation of Low-Cost Sensors for Ambient PM 2.5 Monitoring." *Journal of Sensors* 2018:1–16. https://doi.org/10.1155/2018/5096540.

Bagkis, E., T. Kassandros, M. Karteris, A. Karteris, and K. Karatzas. 2021. "Analyzing and Improving the Performance of a Particulate Matter Low Cost Air Quality Monitoring Device." *Atmosphere* 12 (2): 251. https://doi.org/10.3390/atmos12020251.

Bai, X., H. Tian, C. Zhu, L. Luo, Y. Hao, S. Liu, Z. Guo, et al. 2023. "Present Knowledge and Future Perspectives of Atmospheric Emission Inventories of Toxic Trace Elements: A Critical Review." *Environmental Science & Technology* 57 (4): 1551–1567. https://doi.org/10.1021/acs.est.2c07147.

Baklanov, A., and Y. Zhang. 2020. "Advances in Air Quality Modeling and Forecasting." *Global Transitions* 2:261–270. https://doi.org/10.1016/j.glt.2020.11.001.

Baqer, N. S., H. Mohammed, A. S. Albahri. 2023. "Development of a Real-Time Monitoring and Detection Indoor Air Quality System for Intensive Care Unit and Emergency Department." *Signa Vitae* 19 (1). 77–92

Barros, N., P. Sobral, R. S. Moreira, J. Vargas, A. Fonseca, I. Abreu, and M. Simas Guerreiro. 2023. "SchoolAIR: A Citizen Science IoT Framework Using Low-Cost Sensing for Indoor Air Quality Management." *Sensors* 24 (1): 148. https://doi.org/10.3390/s24010148.

Baudet, A., E. Baurès, O. Blanchard, P. Le Cann, J.-P. Gangneux, and A. Florentin. 2022. "Indoor Carbon Dioxide, Fine Particulate Matter and Total Volatile Organic Compounds in Private Healthcare and Elderly Care Facilities." *Toxics* 10 (3): 136. https://doi.org/10.3390/toxics10030136.

Baurès, E., O. Blanchard, F. Mercier, E. Surget, P. Le Cann, A. Rivier, J.-P. Gangneux, and A. Florentin. 2018. "Indoor Air Quality in Two French Hospitals: Measurement of Chemical and Microbiological Contaminants." *Science of the Total Environment* 642:168–179. https://doi.org/10.1016/j.scitotenv.2018.06.047.

Becerra, J. A., J. Lizana, M. Gil, A. Barrios-Padura, P. Blondeau, and R. Chacartegui. 2020. "Identification of Potential Indoor Air Pollutants in Schools." *Journal of Cleaner Production* 242:118420. https://doi.org/10.1016/j.jclepro.2019.118420.

Blanco, M. N., A. Gassett, T. Gould, A. Doubleday, D. L. Slager, E. Austin, E. Seto, T. V. Larson, J. D. Marshall, and L. Sheppard. 2022. "Characterization of Annual Average Traffic-Related Air Pollution Concentrations in the Greater Seattle Area from a Year-Long Mobile Monitoring Campaign." *Environmental Science & Technology* 56 (16): 11460–11472. https://doi.org/10.1021/acs.est.2c01077.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Bro, R., and A. K. Smilde. 2014. "Principal Component Analysis." *Analytical Methods* 6 (9): 2812–2831. https://doi.org/10.1039/C3AY41907J.

Burgués, J., and S. Marco. 2023. "Drone-Based Monitoring of Environmental Gases." In *Air Quality Networks. Environmental Informatics and Modeling*, edited by S. De Vito, K. Karatzas, A. Bartonova, G. Fattoruso, 115–137. Cham: Springer. https://doi.org/10.1007/978-3-031-08476-8_6.

Castell, N., F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova. 2017. "Can Commercial Low-Cost Sensor Platforms Contribute to Air Quality Monitoring and Exposure Estimates?" *Environment International* 99:293–302. https://doi.org/10.1016/j.envint.2016.12.007.

Concas, F., J. Mineraud, E. Lagerspetz, S. Varjonen, X. Liu, K. Puolamäki, P. Nurmi, and S. Tarkoma. 2021. "Low-Cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey and Critical Analysis." *ACM Transactions on Sensor Networks* 17 (2): 1–44. https://doi.org/10.1145/3446005.

Considine, E. M., D. Braun, L. Kamareddine, R. C. Nethery, and P. deSouza. 2023. "Investigating Use of Low-Cost Sensors to Increase Accuracy and Equity of Real-Time Air Quality Information." *Environmental Science & Technology* 57 (3): 1391–1402. https://doi.org/10.1021/acs.est.2c06626.

Cordero, J. M., R. Borge, and A. Narros. 2018. "Using Statistical Methods to Carry Out in Field Calibrations of Low Cost Air Quality Sensors." *Sensors and Actuators: B, Chemical* 267:245–254. https://doi.org/10.1016/j.snb.2018.04.021.

Dai, X., W. Shang, J. Liu, M. Xue, and C. Wang. 2023. "Achieving Better Indoor Air Quality with IoT Systems for Future Buildings: Opportunities and Challenges." *Science of the Total Environment* 895:164858. https://doi.org/10.1016/j.scitotenv.2023.164858.

Datta, A., A. Saha, M. Levy Zamora, C. Buehler, L. Hao, F. Xiong, D. R. Gentner, and K. Koehler. 2020. "Statistical Field Calibration of a Low-Cost PM2. 5 Monitoring Network in Baltimore." *Atmospheric environment* 242:117761. https://doi.org/10.1016/j.atmosenv.2020.117761.

de Gennaro, G., P. R. Dambruoso, A. Demarinis Loiotile, A. Di Gilio, P. Giungato, M. Tutino, A. Marzocca, A. Mazzone, J. Palmisani, and F. Porcelli. 2014. "Indoor Air Quality in Schools." *Environmental Chemistry Letters* 12 (4): 467–482. https://doi.org/10.1007/s10311-014-0470-6.

De Vito, S., G. D. Francia, E. Esposito, S. Ferlito, F. Formisano, and E. Massera. 2020. "Adaptive Machine Learning Strategies for Network Calibration of IoT Smart Air Quality Monitoring Devices." *Pattern Recognition Letters* 136:264–271. https://doi.org/10.1016/j.patrec.2020.04.032.

Donnelly, A., B. Misstear, and B. Broderick. 2015. "Real Time Air Quality Forecasting Using Integrated Parametric and Non-Parametric Regression Techniques." *Atmospheric Environment* 103:53–65. https://doi.org/10.1016/j.atmosenv.2014.12.011.

Drajic, D. D., and N. R. Gligoric. 2020. "Reliable Low-Cost Air Quality Monitoring Using Off-The-Shelf Sensors and Statistical Calibration." *Elektronika Ir Elektrotechnika* 26 (2): 32–41. https://doi.org/10.5755/j01.eie.26.2.25734.

EPA, U.S. Environmental Protection Agency | US. 2023. *Title 40 of the CFR (Code of Federal Regulations)*. U.S. Government Publishing Office. https://www.ecfr.gov/current/title-40.

Esposito, E., S. De Vito, V. B. Maria Salvato, R. Lewis Jones, O. Popoola, and O. Popoola. 2016. "Dynamic Neural Network Architectures for on Field Stochastic Calibration of Indicative Low

Cost Air Quality Sensing Systems." *Sensors and Actuators: B, Chemical* 231:701–713. https://doi.org/10.1016/j.snb.2016.03.038.

Feenstra, B., V. Papapostolou, S. Hasheminassab, H. Zhang, B. Der Boghossian, D. Cocker, and A. Polidori. 2019. "Performance Evaluation of Twelve Low-Cost PM2. 5 Sensors at an Ambient Air Monitoring Site." *Atmospheric Environment* 216:116946. https://doi.org/10.1016/j.atmosenv.2019.116946.

Flores-Cortez, O. O., R. Adalberto Cortez, and V. Idalia Rosa. 2019. "A Low-Cost IoT System for Environmental Pollution Monitoring in Developing Countries." In *2019 MIXDES-26th International Conference"Mixed Design of Integrated Circuits and Systems"*, Rzeszow, Poland, 386–389. IEEE.

Francesco, B., E. Brattich, C. Cintolesi, A. Ghafoor Nizamani, S. Di Sabatino, M. Milelli, E. E. Peerlings, S. Polder, G.-J. Steeneveld, and A. Parodi. 2024. "Performance Evaluation of MeteoTracker Mobile Sensor for Outdoor Applications." *Atmospheric Measurement Techniques* 17 (10): 3255–3278.

Ganji, A., O. Youssefi, J. Xu, K. Mallinen, M. Lloyd, A. Wang, A. Bakhtari, S. Weichenthal, and M. Hatzopoulou. 2023. "Design, Calibration, and Testing of a Mobile Sensor System for Air Pollution and Built Environment Data Collection: The Urban Scanner Platform." *Environmental Pollution* 317:120720. https://doi.org/10.1016/j.envpol.2022.120720.

Harleman, M., L. Harris, M. D. Willis, B. Ritz, P. Hystad, and E. L. Hill. 2023. "Changes in Traffic Congestion and Air Pollution Due to Major Roadway Infrastructure Improvements in Texas." *Science of the Total Environment* 898:165463. https://doi.org/10.1016/j.scitotenv.2023.165463.

Hernández-Gordillo, A., S. Ruiz-Correa, V. Robledo-Valero, C. Hernández-Rosales, and S. Arriaga. 2021. "Recent Advancements in Low-Cost Portable Sensors for Urban and Indoor Air Quality Monitoring." *Air Quality, Atmosphere & Health* 14 (12): 1931–1951. https://doi.org/10.1007/s11869-021-01067-x.

Hudda, N., L. W. Durant, S. A. Fruin, and J. L. Durant. 2020. "Impacts of Aviation Emissions on Near-Airport Residential Air Quality." *Environmental Science & Technology* 54 (14): 8580–8588. https://doi.org/10.1021/acs.est.0c01859.

Ilie, A. M. C., N. McCarthy, L. Velasquez, M. Moitra, and H. Michael Eisl. 2022. "Air Pollution Exposure Assessment at Schools and Playgrounds in Williamsburg Brooklyn NYC, with a View to Developing a Set of Policy Solutions." *Journal of Environmental Studies and Sciences* 12 (4): 838–852. https://doi.org/10.1007/s13412-022-00777-7.

Jovašević-Stojanović, M., A. Bartonova, D. Topalović, I. Lazović, B. Pokrić, and Z. Ristovski. 2015. "On the Use of Small and Cheaper Sensors and Devices for Indicative Citizen-Based Monitoring of Respirable Particulate Matter." *Environmental Pollution* 206:696–704. https://doi.org/10.1016/j.envpol.2015.08.035.

Jung, C.-C., P.-C. Wu, C.-H. Tseng, and H.-J. Su. 2015. "Indoor Air Quality Varies with Ventilation Types and Working Areas in Hospitals." *Building & Environment* 85:190–195. https://doi.org/10.1016/j.buildenv.2014.11.026.

Karagulian, F., M. Barbiere, A. Kotsev, L. Spinelle, M. Gerboles, F. Lagler, N. Redon, S. Crunaire, and A. Borowiak. 2019. "Review of the Performance of Low-Cost Sensors for Air Quality Monitoring." *Atmosphere* 10 (9): 506. https://doi.org/10.3390/atmos10090506.

Kerckhoffs, J., J. Khan, G. Hoek, Z. Yuan, T. Ellermann, O. Hertel, M. Ketzel, S. Solvang Jensen, K. Meliefste, and R. Vermeulen. 2022. "Mixed-Effects Modeling Framework for Amsterdam and Copenhagen for Outdoor NO2 Concentrations Using Measurements Sampled with Google Street View Cars." *Environmental Science & Technology* 56 (11): 7174–7184. https://doi.org/10.1021/acs.est.1c05806.

Kerckhoffs, J., J. Khan, G. Hoek, Z. Yuan, O. Hertel, M. Ketzel, S. Solvang Jensen, F. Al Hasan, K. Meliefste, and R. Vermeulen. 2022. "Hyperlocal Variation of Nitrogen Dioxide, Black Carbon, and Ultrafine Particles Measured with Google Street View Cars in Amsterdam and Copenhagen." *Environment international* 170:107575. https://doi.org/10.1016/j.envint.2022.107575.

Kow, P.-Y., I.-W. Hsia, L.-C. Chang, and F.-J. Chang. 2022. "Real-Time Image-Based Air Quality Estimation by Deep Learning Neural Networks." *Journal of Environmental Management* 307:114560. https://doi.org/10.1016/j.jenvman.2022.114560.

Lee, S. C., and M. Chang. 2000. “Indoor and Outdoor Air Quality Investigation at Schools in Hong Kong.” *Chemosphere* 41 (1–2): 109–113. https://doi.org/10.1016/S0045-6535(99)00396-3.

Limon, I. A., A. Daud Hossain, K. Farhan Ibne Faruque, M. Rejwan Uddin, and M. Hasan. 2023. “Drone-Based Real-Time Air Pollution Monitoring for Low-Access Areas by Developing Mobile-Smart Sensing Technology.” In *2023 3rd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, Dhaka, Bangladesh, 90–94. IEEE.

Liu, H., G. Yan, Z. Duan, and C. Chen. 2021. “Intelligent Modeling Strategies for Forecasting Air Quality Time Series: A Review.” *Applied Soft Computing* 102:106957. https://doi.org/10.1016/j.asoc.2020.106957.

Liu, X., R. Jayaratne, P. Thai, T. Kuhn, I. Zing, B. Christensen, R. Lamont, et al. 2020. “Low-Cost Sensors As an Alternative for Long-Term Air Quality Monitoring.” *Environmental Research* 185:109438. https://doi.org/10.1016/j.envres.2020.109438.

Mamun, A., M. Abdulla, and M. Rasit Yuce. 2019. “Sensors and Systems for Wearable Environmental Monitoring Toward IoT-Enabled Applications: A Review.” *IEEE Sensors Journal* 19 (18): 7771–7788. https://doi.org/10.1109/JSEN.2019.2919352.

Manshur, T., C. Luiu, W. R. Avis, V. Bukachi, M. Gatari, J. Mulligan, J. Radcliffe, et al. 2023. “A Citizen Science Approach for Air Quality Monitoring in a Kenyan Informal Development.” *City and Environment Interactions* 19:100105. https://doi.org/10.1016/j.cacint.2023.100105.

Margaritis, D., C. Keramydas, I. Papachristos, D. Lambropoulou. 2021. “Calibration of Low-Cost Gas Sensors for Air Quality Monitoring.” *Aerosol & Air Quality Research* 21 (11): 210073. https://doi.org/10.4209/aaqr.210073.

Marinov, M. B., I. Topalov, E. Gieva, and G. Nikolov. 2016. “Air Quality Monitoring in Urban Environments.” In *2016 39th International Spring Seminar on Electronics Technology (ISSE)*, Pilsen, Czech Republic, 443–448. IEEE.

Marjovi, A., A. Arfire, and A. Martinoli. 2017. “Extending Urban Air Quality Maps Beyond the Coverage of a Mobile Sensor Network: Data Sources, Methods, and Performance Evaluation.” In *Proceedings of the International Conference on Embedded Wireless Systems and Networks*, Uppsala, Sweden, February 20-22, 2017, 12–23.

Mata, T. M., A. A. Martins, C. S. Calheiros, F. Villanueva, N. P. Alonso-Cuevilla, M. Fonseca Gabriel, and G. Ventura Silva. 2022. “Indoor Air Quality: A Review of Cleaning Technologies.” *Environments* 9 (9): 118. https://doi.org/10.3390/environments9090118.

Mead, M. I., O. A. M. Popoola, G. B. Stewart, M. C. Peter Landshoff, M. Hayes, J. J. Baldovi, J. J. Baldovi, et al. 2013. “The Use of Electrochemical Sensors for Monitoring Urban Air Quality in Low-Cost, High-Density Networks.” *Atmospheric Environment* 70:186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060.

Ministry of Environment and Sustainable Development. 2017. “Resolution 2254 of 2017 from the Ministry of Environment and Sustainable Development.” https://www.ccacoalition.org/partners/colombia .

Morawska, L., P. K. Thai, X. Liu, A. Asumadu-Sakyi, G. Ayoko, A. Bartonova, A. Bedini, et al. 2018. “Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone?” *Environment international* 116:286–299. https://doi.org/10.1016/j.envint.2018.04.018.

Motlagh, N. H., E. Lagerspetz, P. Nurmi, X. Li, S. Varjonen, J. Mineraud, M. Siekkinen, et al. 2020. “Toward Massive Scale Air Quality Monitoring.” *IEEE Communications Magazine* 58 (2): 54–59. https://doi.org/10.1109/MCOM.001.1900515.

Nishitateno, S., P. J. Burke, and T. H. Arimura. 2024. “Road Traffic Flow and Air Pollution Concentrations: Evidence from Japan.” *International Journal of Economic Policy Studies* 1–29. https://doi.org/10.1007/s42495-024-00132-4.

Piersanti, A., L. Vitali, G. Righini, G. Cremona, and L. Ciancarella. 2015. “Spatial Representativeness of Air Quality Monitoring Stations: A Grid Model Based Approach.” *Atmospheric Pollution Research* 6 (6): 953–960. https://doi.org/10.1016/j.apr.2015.04.005.

Ranganathan, R. H., S. Balusamy, P. Partheeban, C. Mani, M. Sridhar, and V. Rajasekaran. 2023. “Air Quality Monitoring and Analysis for Sustainable Development of Solid Waste Dump Yards Using

Smart Drones and Geospatial Technology." *Sustainability* 15 (18): 13347. https://doi.org/10.3390/su151813347.

Reis, J., D. Lopes, D. Graça, A. Patrcia Fernandes, A. Isabel Miranda, and M. Lopes. 2023. "Using Low-Cost Sensors to Assess Real-Time Comfort and Air Quality Patterns in Indoor Households." *Environmental Science and Pollution Research* 30 (3): 7736–7751. https://doi.org/10.1007/s11356-022-22771-w.

Renwick, J. D., L. J. Klein, and H. F. Hamann. 2016. "Drone-Based Reconstruction for 3D Geospatial Data Processing." In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, Reston, VA, USA, 729–734. IEEE.

Riley, K., R. Cook, E. Carr, and B. Manning. 2021. "A Systematic Review of the Impact of Commercial Aircraft Activity on Air Quality Near Airports." *City and Environment Interactions* 11:100066. https://doi.org/10.1016/j.cacint.2021.100066.

Sahu, R., A. Nagal, K. Kumar Dixit, H. Unnibhavi, S. Mantravadi, S. Nair, Y. Simmhan, et al. 2021. "Robust Statistical Calibration and Characterization of Portable Low-Cost Air Quality Monitoring Sensors to Quantify Real-Time O_3 and NO_2 Concentrations in Diverse Environments." *Atmospheric Measurement Techniques* 14 (1): 37–52. https://doi.org/10.5194/amt-14-37-2021.

Schneider, P., N. Castell, M. Vogt, F. R. Dauge, W. A. Lahoz, and A. Bartonova. 2017. "Mapping Urban Air Quality in Near Real-Time Using Observations from Low-Cost Sensors and Model Information." *Environment International* 106:234–247. https://doi.org/10.1016/j.envint.2017.05.005.

Shaibal, M., S. Luther, B. Halley, S. Casson, N. Lucas, K. Sue, and W. Ronald. 2015. "Comparison of Modeled Traffic Exposure Zones Using On–Road Air Pollution Measurements." *Atmospheric Pollution Research* 6 (1): 82–87. https://doi.org/10.5094/APR.2015.010.

Sher, F., D. Raore, J. Jaromr Klemeš, P. Muhammad Rafi-Ul Shan, M. Khzouz, K. Marintseva, and O. Razmkhah. 2021. "Unprecedented Impacts of Aviation Emissions on Global Environmental and Climate Change Scenario." *Current Pollution Reports* 7 (4): 549–564. https://doi.org/10.1007/s40726-021-00206-3.

Singh, D., M. Dahiya, R. Kumar, and C. Nanda. 2021. "Sensors and Systems for Air Quality Assessment Monitoring and Management: A Review." *Journal of Environmental Management* 289:112510. https://doi.org/10.1016/j.jenvman.2021.112510.

Smart, N. R., C. J. Horwell, T. S. Smart, and K. S. Galea. 2020. "Assessment of the Wearability of Facemasks Against Air Pollution in Primary School-Aged Children in London." *International Journal of Environmental Research and Public Health* 17 (11): 3935. https://doi.org/10.3390/ijerph17113935.

Sokhi, R. S., N. Moussiopoulos, A. Baklanov, J. Bartzis, I. Coll, S. Finardi, R. Friedrich, et al. 2021. "Advances in Air Quality Research–Current and Emerging Challenges." *Atmospheric Chemistry and Physics Discussions* 22 (7): 1–133.

Sonawani, S., and K. Patil. 2024. "Air Quality Measurement, Prediction and Warning Using Transfer Learning Based IOT System for Ambient Assisted Living." *International Journal of Pervasive Computing and Communications* 20 (1): 38–55. https://doi.org/10.1108/IJPCC-07-2022-0271.

Spengler, J. D., and D. G. Wilson. 2003. "Air Quality in Aircraft." *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering* 217 (4): 323–335. https://doi.org/10.1243/095440803322611688.

Spinelle, L., M. Gerboles, M. Gabriella Villani, M. Aleixandre, and F. Bonavitacola. 2014. "Calibration of a Cluster of Low-Cost Sensors for the Measurement of Air Pollution in Ambient Air." *SENSORS, 2014 IEEE*, Valencia, Spain, 21–24. IEEE.

Spinelle, L., M. Gerboles, M. Gabriella Villani, M. Aleixandre, and F. Bonavitacola. 2015. "Field Calibration of a Cluster of Low-Cost Available Sensors for Air Quality Monitoring. Part A: Ozone and Nitrogen Dioxide." *Sensors and Actuators: B, Chemical* 215:249–257. https://doi.org/10.1016/j.snb.2015.03.031.

Spinelle, L., M. Gerboles, M. Gabriella Villani, M. Aleixandre, and F. Bonavitacola. 2017. "Field Calibration of a Cluster of Low-Cost Commercially Available Sensors for Air Quality Monitoring. Part B: NO, CO and CO_2." *Sensors and Actuators: B, Chemical* 238:706–715. https://doi.org/10.1016/j.snb.2016.07.036.

Tsymbal, A., M. Pechenizkiy, and P. Cunningham. 2006. "Dynamic Integration with Random Forests." In *Machine Learning: ECML 2006. ECML 2006. Lecture Notes in Computer Science*, edited by J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, Vol. 4212, 801–808. Berlin, Heidelberg: Springer. https://doi.org/10.1007/11871842_82.

Van den Bossche, J., J. Peters, J. Verwaeren, D. Botteldooren, J. Theunis, and B. De Baets. 2015. "Mobile Monitoring for Mapping Spatial Variation in Urban Air Quality: Development and Validation of a Methodology Based on an Extensive Dataset." *Atmospheric Environment* 105:148–161. https://doi.org/10.1016/j.atmosenv.2015.01.017.

Vitali, G., M. Arru, and E. Magnanini. 2023. "A Scalable Device for Undisturbed Measurement of Water and CO 2 Fluxes Through Natural Surfaces." *Sensors* 23 (5): 2647. https://doi.org/10.3390/s23052647.

Wargocki, P., J. Ali Porras-Salazar, S. Contreras-Espinoza, and W. Bahnfleth. 2020. "The Relationships Between Classroom Air Quality and children's Performance in School." *Building & Environment* 173:106749. https://doi.org/10.1016/j.buildenv.2020.106749.

Waters, M. A., T. F. Bloom, B. Grajewski, and J. Deddens. 2002. "Measurements of Indoor Air Quality on Commercial Transport Aircraft." In *Proceedings of indoor air*, OH, USA, 782–787.

Wivou, J., L. Udawatta, A. Alshehhi, E. Alzaabi, A. Albeloshi, and S. Alfalasi. 2016. "Air Quality Monitoring for Sustainable Systems via Drone Based Technology." In *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, Galle, Sri Lanka, 1–5. IEEE.

Zhang, Y., M. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov. 2012. "Real-Time Air Quality Forecasting, Part I: History, Techniques, and Current Status." *Atmospheric Environment* 60:632–655. https://doi.org/10.1016/j.atmosenv.2012.06.031.