

Analysis and Improvement of Two Low-Cost Air Quality Sensor Measurements' Uncertainty



Marios Panourgias and Kostas Karatzas

Abstract Measurements resulting from the operation of two different low-cost air quality monitoring devices (LCAQMD) are used as a basis for a data analytics and modelling procedure towards the improvement of the uncertainty of sensor readings. A data processing method for missing value and outliers handling, followed by the implementation of computational intelligence-oriented algorithms aimed to the PM₁₀ modelling. Descriptive statistics and correlation coefficients are used for a primary evaluation of data analytics results, while modelling outcomes are compared with the aid of the relative expanded uncertainty, as well as via the model performance evaluation metrics, to determine the most efficient model. Results suggest that the advanced artificial neural network oriented computational intelligence algorithms, may lead to significant improvement of the performance of the two LCAQMD, this being applicable for a certain concentration range (18–65 $\mu\text{g}/\text{m}^3$), indicating that additional future work and more advanced computational techniques are required for further improvement of their performance.

Keywords Low-cost air quality monitoring devices · Measurement uncertainty · Data quality · Computational intelligence

1 Introduction

Air pollution has been proven to have various adverse effects on health, climate, and sustainable development [1]. The increase in the number of air quality (AQ) monitoring devices improves our knowledge on air pollution and allows for better regulatory as well as information provision actions towards better quality of life. Such devices are usually of high cost, this being a limiting factor concerning their

M. Panourgias · K. Karatzas (✉)
Environmental Informatics Research Group, School of Mechanical Engineering, Aristotle University, Thessaloniki, Greece
e-mail: kkara@auth.gr

M. Panourgias
e-mail: m.panourgias@student.maastrichtuniversity.nl

wide-spread application. An alternative that has gained ground in recent years, especially in relation to citizen science initiatives, is that of low-cost air quality monitoring devices (LCAQMD) [2]. Unfortunately, the quality of these measurements has been strongly doubted as they often differ significantly from those of high quality or reference instruments that operate in accordance with the standards of European legislation [3]. Nevertheless, improving the performance of LCAQMD will help in their operational use as supporting methods for the assessment of air pollution. For this purpose, we apply computational intelligence methods to model and improve the performance of two different LCAQMD tested in Thessaloniki, Greece. In Chap. 2 we describe the materials and methods employed, Chap. 3 presents and discusses the results, and Chap. 4 draws the conclusions of this study.

2 Materials and Methods

2.1 Study Area and Experimental Setup

The study was conducted in Thessaloniki, the financial and educational centre of the Macedonian region in Northern Greece, with approx. one million inhabitants. The city has the sea to its south and southwest (Thermaikos gulf), with the Chortiatis mountains framing its southeast border and the Seich-Sou forest lying to the northeast outskirts of the urban web. The local climate is Mediterranean with hot, dry summers and mild, wet winters. The annual mean temperature is 15.6 °C, while mean relative humidity ranges from 53.2% (July) to 78% (December), with an annual mean rate of 67.3%.

The Aristotle University experimental location

The devices participating in the experiment were placed on the roof of Building E14 of the Faculty of Engineering at the Aristotle University of Thessaloniki main campus. The location is next to a busy street (named “3rd of September”) and close to the city centre (Fig. 1). The installation is less than 50 m away from a road junction, so measurements are expected to display sharp gradients of PM₁₀ concentrations [4]. A metro construction site is located at a distance less than 200 m. Two devices have been used:

- (a) The Purpleair PA-II LCAQMD (manufacturer: Purpleair llc, USA), which employs the Plantower PMS5003 optical sensor for particulate matter (PM). It has a built-in microscopic fan to ensure the necessary air flow and can detect a wide range of particle sizes up to 10 μm, therefore being used for PM_{2,5} and PM₁₀ concentration estimations. The generated signals of the sensors are processed by special company algorithms for their conversion into concentration units of the corresponding mass. The PMS5003 sensors are calibrated in a certified laboratory by the manufacturer. Moreover, this specific device uses two

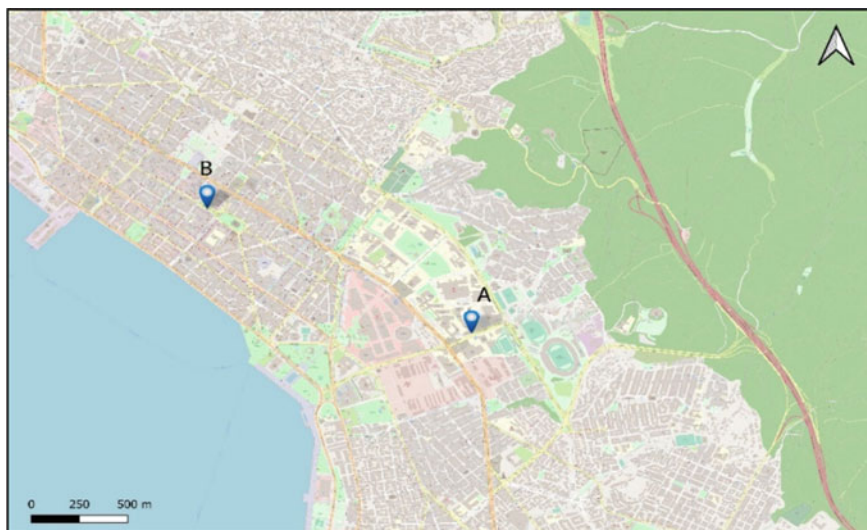


Fig. 1 Location of the air quality device installations in Thessaloniki, Greece: **a** Aristotle University campus installation and **b** Thessaloniki city centre installation

optical sensor units connected to each other under the same protective frame, while it also includes an air temperature and relative humidity sensor.

- (b) The Dust Sentry PM_{10} (DS- PM_{10} , manufacturer: Aeroqual Ltd, New Zealand), which was used as the high-quality measuring instrument at site. Its operation is based on a light scattering nephelometer with a sharp cut cyclone. It also contains a built-in sensor for detecting and balancing temperature changes, air purification filter to keep the surface of optical sensors clean, automatic baseline drift correction and optical fibres to control the procedure of optical components.

The corresponding data set consists of 8296 h values received from 01/01/2019 to 12/12/2019.

The city centre experimental location

This is the second location of the experiment, where an official air quality reference station is installed (at Ermou Street, near Agias Sofias square, Fig. 1). This is a busy city centre area, therefore expected to depict high gradients of traffic-related pollutant concentrations. The area is surrounded by high-density buildings and hosts a metro construction site at a distance approx. 250 m. Two devices have been installed here also:

- (a) The AQY, used as the LCAQMD (manufacturer: Aeroqual Ltd, New Zealand). It uses the Nova Fitness SDS011 particle optical sensor, which detects $PM_{2.5}$ and PM_{10} using an optical laser particle counter. The same device also detects NO_2 and O_3 using sensitive electrochemical sensors (Gas Sensitive Electrochemical-GSE) and Gas Sensitive Semiconductor (GSS) sensors, respectively.

Table 1 Parameters monitored by the devices participating in the study

Instrument	Parameters monitored
Purpleair	PM ₁ , PM _{2.5} , PM ₁₀ , Temperature (ambient), Rel. Humidity
DS-PM ₁₀	PM ₁₀ , O ₃ , O _x , NO ₂ , SO ₂ , Temperature (ambient), Rel. Humidity
AQY	PM _{2.5} , PM ₁₀ , O ₃ , NO ₂ , Temperature (ambient), Rel. Humidity
Reference	PM _{2.5} , PM ₁₀ , O ₃ , NO ₂ , Temperature (ambient), Rel. Humidity

- (b) A reference instrument, that measures concentration levels of PM₁₀ and PM_{2.5} using an Eberline FH 62 IR analyser together with a b-type damper. Both the low-cost and the reference instruments were accompanied by temperature and relative humidity measurements. The resulting dataset is composed of 3960 hourly instances for the period 27/03/2019–08/09/2019.

Both installation locations display similar characteristics concerning the proximity to main roads of the city, while the metro construction sites add an extra factor that draws the attention to compare these two occasions. Monitored parameters are reported in Table 1.

2.2 Data Pre-processing

It was made evident that technical issues led to considerable number of missing values, thus it was decided to remove all records corresponding to those of the DS-PM₁₀ and of the reference instruments. For the Purpleair and AQY devices, the kNN data imputation method was implemented: For each time series vector the algorithm identifies the k-nearest observations closer to a missing value based on their Euclidean distance and then calculates their weighted average. This new value is assigned to the missing point. The procedure is repeated until every missing point is substituted by the average of its k-nearest neighbour of the corresponding variable. The number of k-nearest neighbours is a parameter that the user needs to select based on computational experiments [5]. In this study five nearest neighbours were selected, for each variable in all the available datasets, as the parameter more suitable to achieve the best results.

Outliers were also identified in the studied datasets using the criteria described below and then they were assigned as missing values following the procedure described previously:

- (i) *The standard deviation criterion:* Values greater than the sum of $3\mu + \sigma$, where μ is the average (mean) and σ the standard deviation of each variable, were flagged as outliers.

- (ii) *The Iglewicz-Hoaglin criterion* [6]: Considering the $x_1, x_2, x_3, \dots, x_N$ values of a vector X , their mean value x_m and M_x the median value of the vector, for each single value the Mean Absolute Difference (MAD) is calculated as:

$$MAD_i = |x_i - M_x| \quad (1)$$

Estimating the median value M_{MAD} of the newly formed vector of MAD_i , the variable M_i for each single value can be evaluated as following:

$$M_i = \frac{0.675(x_i - x_m)}{M_{MAD}} \quad (2)$$

where M_{MAD} is the median value of the vector formed by the Eq. 1. Values with $M_i > 3.5$ are labelled as outliers.

2.3 Intercomparison of Measurements

To analyse the measurements taken via the LCAQMDs, several descriptive statistics were employed, while correlation coefficients were also calculated. Moreover, the relative expanded uncertainty of the measurements was estimated, as a key parameter for assessing and improving the performance of the low-cost devices. Table 3 presents the mean value, the standard deviation, the skewness, and kurtosis coefficients concerning PM_{10} concentration. While the mean and the standard deviation describe the location and spreading of the measurements, skewness and kurtosis are a measure of symmetry and heavy tailing of the values.

Another useful aspect of the analysis focused on the calculation of the Pearson and Spearman correlation coefficients among the time series of each parameter. Their main difference is that Pearson does not reflect non-linear relationships, whereas the existence of extreme values strongly affects its magnitude. Spearman on the other hand, reflects the monotonic relationship that may exist between the variables. Positive values of these coefficients indicate that two factors increase in parallel, while negative ones demonstrate opposite trends.

In addition to the above metrics, relative expanded uncertainty (REU) is a key factor for estimating the quality of air pollutants measurements. In the field of air quality monitoring, the concept of uncertainty has been introduced through the European Air Quality Framework Directive 2008/50/EC. It constitutes a robust criterion of the established Data Quality Objectives (DQO) which an observation must meet to be considered equivalent to the one derived from the use of the reference methods, or to be defined as an indicative one as a function of its relative expanded uncertainty that characterizes it. Values of the REU used for classifying measurements to different “quality” groups are defined in Annex I of 2008/50/EC. According to the technical specifications being developed by the Sensor Working Group [7] of

the Technical Committee on Air Quality Standardization CEN/TC 264/WG 42 [7], low-cost sensor measurements can be categorized as follows [8]:

- Class 1 sensor systems: A monitoring device whose measurements comply with the DQO criteria of the indicative methods set out in Directive 2008/50/EC. In the case of PM_{10} , REU in this case should be below 50%.
- Class 2 sensor systems: A monitoring device whose measurements comply with the DQO criteria of the specified objective estimations in Directive 2008/50/EC. In the case of PM_{10} , REU in this case should be between 50 and 100%.
- Class 3 sensor systems: A monitoring device whose measurements do not appear to any kind of officially established uncertainty limits. This is the case for PM_{10} if the REU is greater than 100%.

The REU is used in accordance with the procedure described in the report published by the EC Working Group on the Guidance for the Demonstration of Equivalence [9] and the technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service [10]. On this basis, the initial performance of the Purpleair and AQY LCAQMDs, compared to the high-precision DS- PM_{10} and reference instrument correspondingly, is evaluated, as well as the efficiency of the results obtained from the modelling procedure for the improvement of their performance.

2.4 Data-Driven Modelling

In a further analysis, an effort was made to develop data-driven mathematical models that receive as an input the variables monitored by the LCAQMD and have as a target-variable the desired parameter of the high precision instrument, this being PM_{10} in the frame of the current study. For the two intercomparisons performed, the inputs were all the parameters mentioned in Table 1 for the low-cost devices (Purpleair and AQY) apart from PM_{10} , while the outputs produced were the PM_{10} which they were later compared with the actual values of PM_{10} vectors of the high-precision instruments. For this reason, computational intelligence algorithms were employed.

Polynomial Regression (PR)

This method examines the relationship between two variables through a polynomial curve [11], which is expressed as the sum of the products of the constant coefficients $\beta_1, \beta_2, \dots, \beta_n$ with the independent variable, having a power exponent greater than 1. The general form of the equation describing the relationship between the independent and dependent variables is:

$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \dots + \beta_n \cdot x^n \quad (3)$$

The most common technique of adjusting the curve to the points is the least squares method that calculates the best fitted curve trying to minimize the sum of the

squares of the distances that the plane points have from this curved line. However, one of its major disadvantages is its sensitivity and the incapability to find a suitable adjusted curve under the presence of many outliers.

Support Vector Regression (SVR)

SVR algorithm belongs to the category of non-linear regression models [12]. In this case the model tries to find the most appropriate insensitive region in the N-dimensional space, which is defined by a central hyperplane (maximum margin hyperplane) and two others, at a distance ε on each side of it. Ultimate target is to include as many points of the data space as possible, minimizing therefore the loss function which is translated as the distance of the excluded points from the limits of this insensitive region. The loss function is described by the following equation:

$$\min_w \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (4)$$

where w is the normalized vector at the external surfaces of the hyperplane, ξ_i and ξ_i^* are the excluded points at each side of the plane respectively and C is a factor that reinforces the minimization.

Random Forest (RF)

The RF algorithm belongs to the family of Decision Trees and is used for classification and regression purposes [13]. The algorithm builds regression models in the form of a tree structure through a sequence of logical decisions. These logical decisions aim to the division of the initial dataset into smaller subsets depending on the information entropy criterion, to obtain the greatest possible gain of useful information. RF is an ensemble method combining the behaviour of several trees, each trained in a subset of the initial dataset, resulting from a bootstrapping (sampling with replacement) procedure. For the specific analysis 10 trees were selected as the parameter, which deduces the greatest results. The final model result is based on weighted voting. The points showing highest gain of useful information are selected as nodes in the development of the tree.

Artificial Neural Networks (ANNs)

The Multilayer Perceptron feedforward backpropagation ANN was utilized in this study [14]. Many network architectures were tested to improve model results; two hidden layers of 10 neurons correspondingly rendered the optimum outcome, avoiding at the same time the adverse phenomenon of overfitting with the parallel examination of train and test sets values of loss function. The structure of the algorithm contained additional early stop settings to avoid overfitting.

In all of the aforementioned modelling approaches, the validation procedure of both the random train-test split and the 10-fold cross validation were tested in order to create the training and the test set, nevertheless the first one rendered slightly better results in most of the occasions, so it was selected as the common validation technique

for all models. In particular, 70% of the data were used as the training sample, while the rest 30% formed the test set, having been split in a random mode. These percentages were selected after conducting computational tests, which indicated these two as the most efficient ones.

Convolutional Neural Networks (CNN 1D)

Convolutional Neural Networks are algorithms derived from the field of Deep Learning [15]. Although they initially targeted at image recognition problems, their basic logic can also be used for time series modelling problems. The main goal of the algorithm is to reduce the dimensions of the input dataset making use of a transformation process based on Artificial Neural Network principles. The transition from a 2D matrix dataset to a 3D structure, to comply with the CNN characteristics, is accomplished by the manual addition of a 3rd dimension to the array, which is the number of the time step used to divide the initial dataset into smaller subsets. Meanwhile, the output is defined to consist of one variable which is the outcome of the network. Contrary to the previous methods of training test data random split, in the CNN case a sequential division is preferred to avoid repetition of the information provided by the multiple sub-sets. A time step equal to 3 was selected as the time-step parameter, meaning that each input sub-set is composed of 3 instances, covering thus sequentially the whole dataset. In each of these smaller sets there was the corresponding single value of the variable-target vector; thereby it is secured the preservation of the initiatory relations between the variables. The Rectifier function was selected as the activation parameter and the best resulting architecture was a network composed of one hidden layer consisted of 64 neurons. A sequential split of 70% of the data as training set and 30% as test set was implemented. Contrary to the previous models' random split, the sequential split was chosen for the CNN model, due to the additional time-step slicing of dataset that has been performed. Validation of the performance of this algorithm was achieved by monitoring the curves of training and test sub-sets loss function minimization. Early stop settings were also used to avoid overfitting. The similarity in the behaviour of the curves is considered as the criterion of validation procedure.

Long Short-Term Memory Recurrent Neural Networks (LSTM)

Long Short-Term Memory (LSTM) networks is a special type of Recurrent Neural Networks (RNN). RNNs use as input not just the present records but also the data coming from a previous time step [16]. There are multiple networks for each time instance, connected with inner loops as the extracted values of a specific moment is utilized for the input of the next period, allowing thus information to persist over a time frame: the decision that a network of this kind will take at time step $t-1$ affects the decision it will make the exact next moment at time step t . For this reason, RNNs are considered to receive two sources of input, the present and the recent past. Hence, memory is the main characteristic of RNNs. It should be mentioned that RNNs function suffers from two problems: vanishing gradient and exploding gradient causing numerous hurdles. A way around this issue was provided by the invention of Long Short-Term Memory units, which preserves the backpropagated

error. Maintenance of a constant error can permit to recurrent nets learning over many precedent time steps. The main logic behind LSTM algorithms is that they produce outputs with a forecasting approach, while the previous work on a nowcasting mode. The preferable architecture selected in this study consists of 4 LSTM layers with 50 neurons respectively. Furthermore, a memory of 24 h instances (or 1 day) was attributed to the dataset, following concurrently the sequential split of 70% training and 30% test of CNN networks for the same reason of time-step slicing. LSTM algorithm uses by default Hyperbolic Tangent and Sigmoid functions for the main and recurrent activation correspondingly.

3 Results and Discussion

3.1 Data Pre-processing

First step at the data pre-processing procedure was the determination of the percentage of missing values in the datasets, to decide which periods were the most appropriate and representative for the computational experiments. This is useful in order to decide whether a whole instance will be subtracted, or the corresponding missing values will be substituted as described in Chap. 2. Most missing values for Purpleair and DS-PM₁₀ was detected during the period July–August of 2019 on account of technical problems caused by the high temperatures and high levels of relative humidity, contrary to the pair of AQY-Reference instruments. For this cause, it was decided not to use the exact same period for both experiments. Based on missing values as well as outlier identification and handling, the pre-processed data for both experimental locations are visualised in Fig. 2. It is evident that the high accuracy instrument (DS-PM₁₀) provided data that were of lower concentration value on average, in comparison to the Purpleair device; in the case of the AQY device, its values were on average lower than those provided by the high accuracy reference instrument.

3.2 Intercomparison of Devices

Basic descriptive statistics (Table 2) show that the Standard Deviation of PM₁₀ time series for the DS-PM₁₀ Aeroqual measurements is greater than its mean value. A potential explanation of this phenomenon is the influence stemming from the sharp gradients of PM₁₀ concentrations due to the nearby traffic load. This can be observed also from the relatively high value of kurtosis and skewness [17].

Concerning correlation coefficients (Table 3), results provide a first look at the divergence of the examined instruments. AQY indexes for PM₁₀ are higher (Pearson = 0.52, Spearman = 0.54), implying a stronger relationship between the low- and

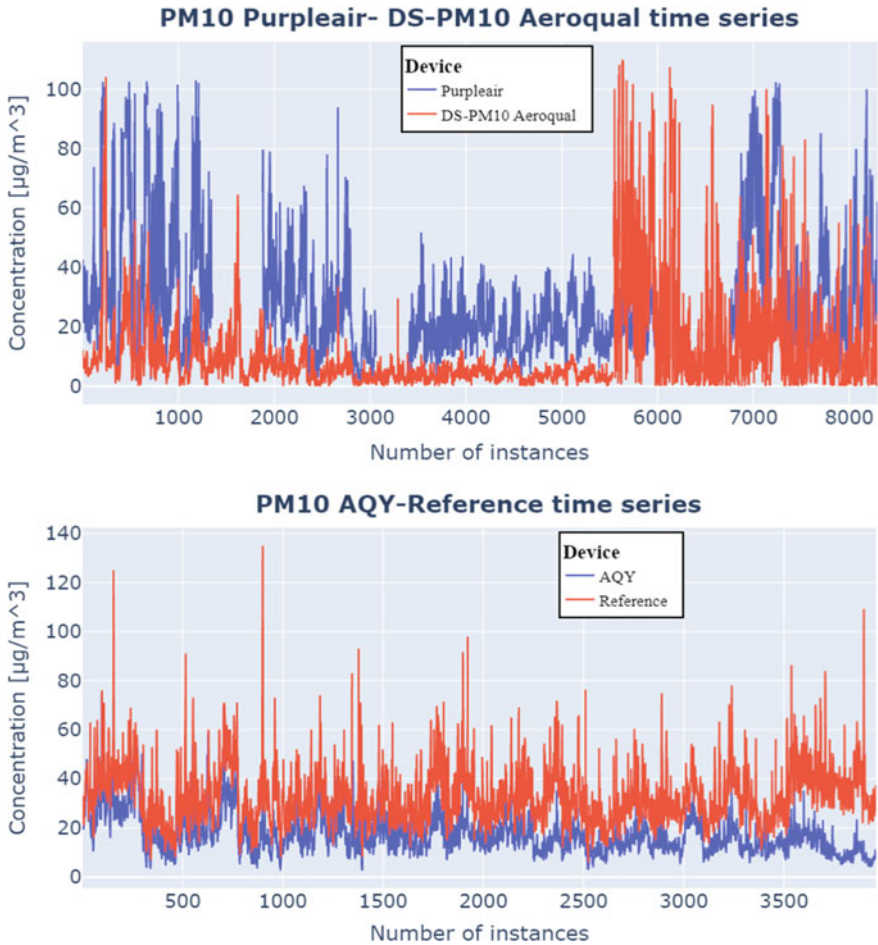


Fig. 2 Time series intercomparison for PM₁₀ concentration levels concerning Purpleair and DS-PM₁₀ Aeroqual (up), and AQY-Reference instrument (down)

Table 2 Basic statistical quantities for the low-cost and high-precision instrument measurements concerning PM₁₀ (in µg/m³)

Device	Mean	Std	Skewness	Kurtosis
Purpleair	26.85	18.62	1.58	5.61
DS-PM ₁₀	10.54	14.67	2.96	13.21
AQY	17.33	8.42	1.57	6.81
Reference	32.65	11.79	1.26	7.38

Table 3 Correlation coefficients for PM₁₀ between the low-cost and high-precision instruments

LCAQMD	High precision	Pearson	Spearman
Purpleair	DS-PM ₁₀	0.16	0.48
AQY	Reference	0.52	0.54

high-quality measurements for this case, contrary to the pair of Purpleair-DS-PM₁₀ (Pearson = 0.16, Spearman = 0.48).

3.3 Uncertainty of PM₁₀ Measurements

The REU of the initial PM₁₀ measurements of the LCAQMDs as well as of the values resulting from their computational improvement are presented in Fig. 3 and in Fig. 4. In the case of the Purpleair device the initial PM₁₀ observations significantly deviate in comparison to the measurements of the high precision instrument (DS-PM₁₀ Aeroqual). More specifically, low concentration levels in the range up to 15 µg/m³ account for uncertainty levels which start at the level of 800% and then register a sharp drop reaching the level of approximately 250% at 15 µg/m³ reference concentration. From that point and after, uncertainty levels represent a gradual upward trend, levelling off at a rate marginally below 1800% (not presented in the relevant graph), as reference values keep rising at the same time, indicating thus a positive correlation above 15 µg/m³.

Referring to AQY, the REU of the initial PM₁₀ values exhibits a better behaviour in comparison to the desired performance of the reference instrument. In detail, the curve begins from a level slightly lower than 300%, then decreases achieving almost

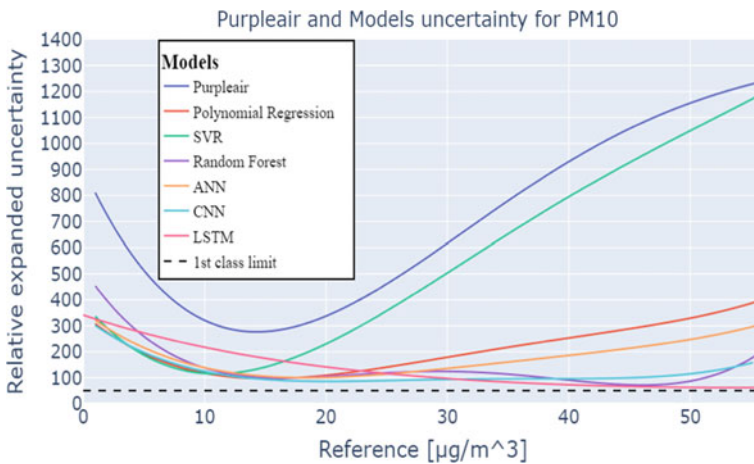


Fig. 3 Purpleair REU (initial and after the computational improvement) for PM₁₀. Dashed line indicates the limit of 1st class measurements of 50%

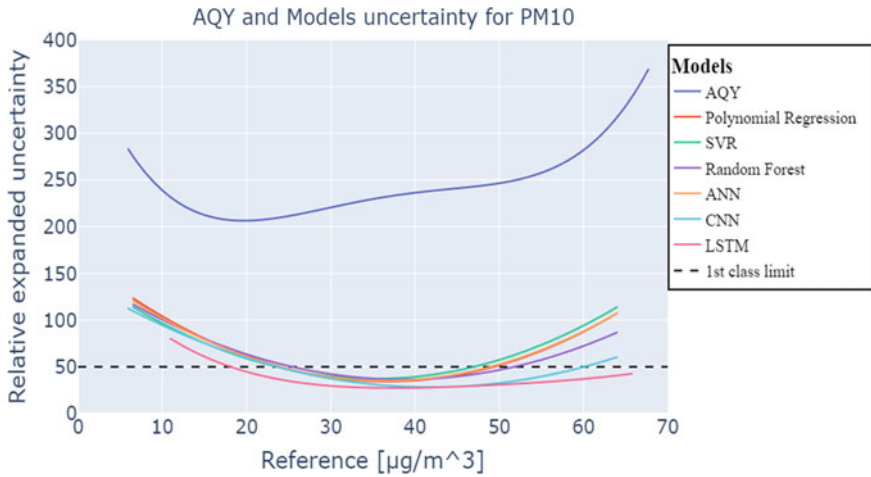


Fig. 4 AQY REU (initial and after the computational improvement) for PM_{10} . Dashed line indicates the limit of 1st class measurements of 50%

200% for reference concentrations of about $20 \mu\text{g}/\text{m}^3$, ending up at 370% for the highest amount of PM_{10} concentrations of $68 \mu\text{g}/\text{m}^3$. However, for both low-cost instruments, the level of 50% of 1st class category of indicative measurements has not been approached. On the contrary, both instruments demonstrate a REU greater than 100%, therefore rendering them as class 3 sensor systems.

Based on the computationally improved PM_{10} REU curves for Purpleair, it is evident that LSTM and CNN demonstrate the most significant upgrade of measurements followed by RF and the simple ANN. More specifically, LSTM started from the level of 320% and progressively approached the limit of 1st class measurements of 50% without falling under it for a PM_{10} concentration between 40 and $50 \mu\text{g}/\text{m}^3$. On the other hand, CNN illustrated a different trend: for a PM_{10} concentration range of approximately $0\text{--}25 \mu\text{g}/\text{m}^3$, the REU drops nearer to the 1st class limit in comparison to LSTM, and then flattens slightly above it. The two other algorithms appear to have an almost aligned behaviour. However, the goal of 50% uncertainty rate is not achieved by any of the algorithms, while the 100% class 2 sensor limit is achieved for a certain concentration value range by LSTM and CNN. Model performance metrics [18], for the evaluation of the models' performance (reported in Table 4), confirm the findings based on the REU graphs as LSTMs obtain the best indices, achieving the lowest RMSE and MAE and the highest R^2 .

Coming to the AQY computational improvement of PM_{10} measurements, it is evident that the LSTM model led to the lowest uncertainty rates. Its curve starts at the level of approximately 75% for low concentration values and then falls under the limit of 50% for the concentration range $18\text{--}65 \mu\text{g}/\text{m}^3$. The 1st class goal is achieved from all other algorithms and especially from CNN which follows LSTM in performance ranking, as it drops below 50% for a concentration range $22\text{--}60 \mu\text{g}/\text{m}^3$.

Table 4 Indices for the performance of the models for Purpleair and AQY for improving PM₁₀ measurements in comparison to PM₁₀ measurements of high-quality instruments

Models	Purpleair			AQY		
	RMSE	MAE	R ²	RMSE	MAE	R ²
Pol. Regr	12.39	7.11	0.32	8.42	6.24	0.37
SVR	13.81	6.26	0.15	8.55	6.23	0.35
RF	10.09	5.55	0.55	8.71	6.45	0.33
ANN	12.3	7.19	0.32	8.55	6.37	0.35
CNN	10.87	6.21	0.46	7.19	5.5	0.55
LSTM	11.76	7.5	0.57	6.18	4.5	0.67

Model performance metrics also verify the best performance achieved by LSTM and CNN, reaching an R² equal to 0.67 for LSTM.

4 Conclusions

The use of LCAQMD as alternative sources of information concerning air pollution concentration estimation, is limited by the low quality of their measurements, in comparison to the one obtained by high quality and especially reference instruments. The REU is being used for categorizing low-cost instruments into three classes according to the Technical Committee on Air Quality Standardization CEN/TC264/WG42 [7, 8], while also being a key criterion for a recently proposed certification protocol for the evaluation of sensor systems dedicated to the ambient air quality monitoring [19]. The performance of such devices can be improved with the aid of computational methods which make use of monitored data as inputs, targeting the values of the parameter of interest as recorded by high quality monitoring instruments. This approach has been applied for two LCAQMD installed and tested in Thessaloniki, Greece. In both cases, parameters monitored via the low-cost devices were used as model inputs, and a total of six computational intelligence algorithms were employed for improving their performance, using PM₁₀ as the parameter of interest. The initial REU for Purpleair was higher in comparison to the REU of AQY, and in both cases the levels of uncertainty were well above the 50 µg/m³ criterion for class 1 measurements. The result of the modelling approach demonstrated that CNN and LSTM performed best in terms of REU improvement: In the case of Purpleair achieving an uncertainty level close to the 50% criterion and below the 100% criterion (LSTM being the best performing algorithm concerning the latter), while in the case of AQY achieving an uncertainty lower than the 50% criterion, therefore resulting in measurements that can be considered to be as indicative (i.e. device class 1), for a concentration range 18–65 µg/m³, concerning their overall REU. Both algorithms achieved the best results in terms of the RMSE, ME and R², with the latter

reaching 0.67 for the calibrated PM_{10} measurements of AQY in comparison to the reference instrument measurements, and 0.57 for the Purpleair vs DS- PM_{10} . These are results directly comparable with the ones that have been reported for the AQY instrument at the same location [20], where a more advanced feature selection and modelling procedure was followed, aiming at maximizing the improvement of the LCAQMD performance in terms of the REU. Overall, LSTMS could be proposed as a method that renders the most desirable outcomes, however the selection of the most appropriate algorithm depends highly on the examined datasets.

Further research should aim at testing more advanced, ensemble-oriented methods like the ones reported in a number of different LCAQMD and at different locations [21]. The ultimate goal would be the development of generalizable and transferrable data-driven models for sensor improvement at an operational level. In this direction, results reported in [22] are very supportive and provide with a way towards achieving this goal.

Acknowledgements The authors greatly acknowledge kartECO S.A. for providing access to the AQY and the Dust Sentry PM_{10} Aeroqual measurements.

Appendix A

Pearson and Spearman correlation coefficients [23]

Pearson coefficient:

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left[\frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y} \right] \quad (5)$$

Spearman coefficient:

$$= 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)} \quad (6)$$

where \bar{y} and \bar{x} are the mean values of the measurements coming from the low cost and of the high-quality devices respectively, s_y and s_x the standard deviations of the two vectors, D is the difference between the simultaneous values of the two variables.

Relative expanded uncertainty:

For this purpose, the measurements of the examined instrument Y_i are linearly related to those of the reference instrument X_i according to the following equation:

$$Y_i = a + bX_i \quad (7)$$

The orthogonal regression calculating procedure is involved for calculating parameters a(slope) and b(intersection):

$$b = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4(S_{xy})^2}}{2S_{xy}} \tag{8}$$

where:

$$S_{xx} = \sum_{i=1}^N (x_i - \bar{x})^2 \tag{9}$$

$$S_{yy} = \sum_{i=1}^N (y_i - \bar{y})^2 \tag{10}$$

$$S_{xy} = \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \tag{11}$$

$$\alpha = \bar{y} - b\bar{x} \tag{12}$$

Here \bar{y} and \bar{x} are the mean values of the measurements coming from the low cost and of the high-quality devices, respectively.

The uncertainty of a single value Y_i (symbolized as u) can then be evaluated as:

$$u(Y_i) = \sqrt{\frac{RSS}{n - 2} - u(x_i)^2 + [a + (b - 1)x_i]^2} \tag{13}$$

where RSS is the sum of relative residuals, and it is calculated as follows:

$$RSS = \sum_{i=1}^N (y_i - a - bx_i)^2 \tag{14}$$

and $u(x_i)$ is the uncertainty of the reference instrument. Especially for the particulate matter case, $u(x_i)$ can be assigned with the constant value $u(x_i)^2 = 0,67[\mu\text{gm}^{-3}]^2$ in accordance with the indications of European Directive 2008/50/EC. Finally, the relative expanded uncertainty is calculated inserting the coverage factor $k = 2$ [9] as follows:

$$u(y_i)_r = \frac{2 \cdot u(Y_i)}{Y_i} \tag{15}$$

References

1. Rajagopalan, S., Park, B., Palanivel, R., Vinayachandran, V., Deiuliis, J.A., Gangwar, R.S., Das, L.M., Yin, J., Choi, Y., Al-Kindi, S., Jain, M.K., Hansen, K.D., Biswal, S.: Metabolic effects of air pollution exposure and reversibility. *J. Clin. Investig.* **130**, 6034–6040 (2020)
2. Kassandra, T.H., Bakousi, A., Gavros, A., Karatzas, K.: Citizens in the loop for air quality monitoring in Thessaloniki, Greece. In Kamilaris, A., Wohlgemuth, V., Karatzas, K., Athanasiadis, I. (eds.) *Advances and New Trends in Environmental Informatics, Digital Twins for Sustainability*, pp. 121–130. Springer AG, Switzerland (2020)
3. Pinho, P., Lopes, S., Panourgias, M., Reis, J., Karatzas, K.: Intercomparison between IoT air quality monitoring devices for PM₁₀ concentration estimations. In Kamilaris, A., Wohlgemuth, V., Karatzas, K., Athanasiadis, I. (eds.) *Environmental Informatics-New perspectives in Environmental Information Systems: transport, Sensors, Recycling, Adjunct Enviroinfo 2020 proceedings*, pp. 139–144, Shaker Verlag, Kassel, Germany (2020)
4. Ganguly, R., Batterman, S., Isakov, V., et al.: Effect of geocoding errors on traffic-related air pollutant exposure and concentration estimates. *J. Expo. Sci. Environ. Epidemiol.* **25**, 490–498 (2015)
5. Aishwarya, S.: A practical introduction to K-nearest neighbors algorithm for regression (with Python code). <https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/>. Accessed 28 Dec. 2021
6. Iglewicz, B., Hoaglin, D.: The ASQC basic references in quality control: statistical techniques. In: Mykytka, E.F. (ed.) *How to Detect and Handle Outliers*, vol. 16. ASQC Quality Press, Milwaukee, USA (1993)
7. European Committee Standardization.: CEN/TC 264/WG 42: ambient air—Air quality sensors (2019)
8. Gerboles, M.: Performance evaluation of sensors for gaseous pollutants and particulate matter monitoring in ambient air. In: *Online Presentations of the UC Davis Air Quality Research Center, Oakland Convention Center, California USA* (2018). https://asic2018.aqrc.ucdavis.edu/sites/g/files/dgvnsk3466/files/inline-files/Michel%20Gerboles%20-%20TC_264_Sep2018.pdf
9. EC Working Group on Guidance for the Demonstration of Equivalence.: *Guide to the demonstration of equivalence of ambient air monitoring methods* (2010). <https://ec.europa.eu/environment/air/quality/legislation/pdf/equivalence.pdf>
10. Spinelle, L., Alexandre, M., Gerboles, M.: Protocol of evaluation and calibration of low-cost gas sensors for the monitoring of air pollution. European Commission JRC Technical Reports (2013). <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/protocol-evaluation-and-calibration-low-cost-gas-sensors-monitoring-air-pollution>
11. Pant, A.: Introduction to linear and polynomial regression. <https://towardsdatascience.com/introduction-to-linear-regression-and-polynomial-regression-f8adc96f31cb>. Accessed 30 Dec. 2021
12. Awad, M., Khanna, R.: Support vector regression. In: *Efficient Learning Machines: Theories, Concepts and Applications for Engineers and System Designers*, p. 67–80. Open Access. <https://link.springer.com/book/10.1007%2F978-1-4302-5990-9>
13. Breiman, L.: Random forests. *Mach. Learn.* **45**, 5–32 (2001)
14. Tan, P.N., Steinbach, M., Karpatne, A., Kumar, V.: *Introduction to Data Mining*, 2nd edn. Michigan State University, University of Minnesota (2019)
15. Wu, J.: *Introduction to Convolutional Neural Networks*. National Key Lab for Novel Software Technology Nanjing University, China (2017). <https://cs.nju.edu.cn/wujx/paper/CNN.pdf>
16. Nicholson, C.: A beginner's guide to LSTMs and recurrent neural networks. <https://wiki.patmind.com/lstm/>. Accessed 30 Dec. 2021
17. Brown, S.: Measures of shape: skewness and kurtosis (2020). <https://brownmath.com/stat/shape.htm>. Accessed 26 Oct. 2020

18. Mishra, D.: Regression: an explanation of regression metrics and what can go wrong. <https://towardsdatascience.com/regression-an-explanation-of-regression-metrics-and-what-can-go-wrong-a39a9793d914>. Accessed 6 Dec. 2021
19. Spinelle, L., Mace, T.: A first certification protocol for the evaluation of sensor systems dedicated to the ambient air quality monitoring, presentation at the CIM2021—Green deal challenges for chemistry (2021). <https://www.ineris.fr/sites/ineris.fr/files/2021-09/Pre0164-SPINELLE.pdf>
20. Bagkis, E, Kassandra, T., Karteris, M, Karteris, A, Karatzas, K: Analyzing and improving the performance of a particulate matter low cost air quality monitoring device. *Atmosphere* **12**(2), 251 (2021). <https://doi.org/10.3390/atmos12020251>
21. Bagkis, E., Kassandra, T.H., Karatzas, K.: Learning calibration functions on the fly: hybrid batch-online stacking ensembles for the calibration of low-cost air quality sensor networks in the presence of concept drifts. *Atmosphere* **13**(3), 416 (2022). <https://doi.org/10.3390/atmos13030416>
22. Bush, T., Papaioannou, N., Leach, F., Pope, F.D., Singh, A., Thomas, G.N., Stacey, B., Bartington, S.: Machine learning techniques to improve the field performance of low-cost air quality sensors. *Atmos. Meas. Tech. Discuss.* <https://doi.org/10.5194/amt-2021-282> (2021)
23. de Winter, J.C.F., Gosling, S.D., Potter, J.: Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychol. Methods* **21**, 273–290 (2016)