Journal of Environmental Management 328 (2023) 116910

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Two step calibration method for ozone low-cost sensor: Field experiences with the *UrbanSense* DCUs

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

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

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

ARTICLE INFO

Keywords:
Low-cost sensor
Air quality
Air pollution
Ozone
Calibration
Statistical model
Machine learning

ABSTRACT

Urban air pollution is a global concern impairing citizens' health, thus monitoring is a pressing need for city managers. City-wide networks for air pollution monitoring based on low-cost sensors are promising to provide real-time data with detail and scale never before possible. However, they still present limitations preventing their ubiquitous use. Thus, this study aimed to perform a post-deployment validation and calibration based on two step methods for ozone low-cost sensor of a city-wide network for air pollution and meteorology monitoring using low-cost sensors focusing on the main challenges. Four of the 23 data collection units (DCUs) of the *UrbanSense* network installed in Porto city (Portugal) with low-cost sensors for particulate matter (PM), carbon monoxide (CO), ozone (O_3), and meteorological variables (temperature, relative humidity, luminosity, precipitation, and wind speed and direction) were evaluated. This study identified post-deployment challenges related to their validation and calibration. The preliminary validation showed that PM, CO and precipitation sensors recorded only unreliable data, and other sensors (wind speed and direction) very few data. A multi-step calibration strategy was implemented: inter-DCU calibration (1st step, for O_3, temperature and relative humidity) and calibration with a reference-grade instrument (2nd step, for O_3). In the 1st step, multivariate linear regression (MLR) resulted in models with better performance than non-linear models such as artificial neural networks (errors almost zero and $R^2 > 0.80$). In the 2nd step, the calibration models using non-linear machine learning boosting algorithms, namely Stochastic Gradient Boosting Regressor (both with the default and post-tuning hyper-parameters), performed better than artificial neural networks and linear regression approaches. The calibrated O_3 data resulted in a marginal improvement from the raw data, with error values close to zero, with low predictability ($R^2 \sim 0.32$). The lessons learned with the present study evidenced the need to redesign the calibration strategy. Thus, a novel multi-step calibration strategy is proposed, based on two steps (pre and post-deployment calibration). When performed cyclically and continuously, this strategy reduces the need for reference instruments, while probably minimising data drifts over time. More experimental campaigns are needed to collect more data and further improve calibration models.

1. Introduction

Urban air pollution is a global concern representing a significant health burden to citizens. Only in 2018, about 492,600 premature deaths were attributed to long-term exposure to particulate matter with aerodynamic diameter of less than 2.5 μm ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3) in 41 European cities (the vast majority of them in the EU-27) (EEA, 2020). As such, to prevent harm to the population exposed to high levels of air pollutants, monitoring and managing air pollution are pressing needs for urban planners and city managers (Penza et al., 2014), for criteria pollutants such as carbon monoxide (CO), O_3, NO_2, sulphur dioxide (SO_2), and particulate matter (PM, highlighting PM_{10} and $PM_{2.5}$) (Guerreiro et al., 2014; Kumar et al., 2015). Conventionally, urban air pollutant concentrations are obtained

* Corresponding author. LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal.
E-mail address: sofia.sousa@fe.up.pt (S.I.V. Sousa).

https://doi.org/10.1016/j.jenvman.2022.116910
Received 10 May 2021; Received in revised form 26 September 2022; Accepted 26 November 2022
Available online 7 December 2022

from: i) static monitoring stations; ii) mobile monitoring stations; iii) passive samplers; and/or iv) models. Static and mobile monitoring stations are equipped with certified reference instruments, usually to provide data over long and short time periods, respectively. However, the installation and maintenance including quality assurance and quality control has high costs leading to sparse coverages, which while sufficient for regulatory purposes is insufficient to provide complete information on human exposure, e.g., about the spatial distribution of pollutants and/or to identify pollution hotspots. This contributes to the lack of information in many areas (Mead et al., 2013; Rai et al., 2017; Schneider et al., 2017). Although mobile monitoring stations possibly increase the spatial sampling density, their temporal coverage is incomplete. Similarly, passive samplers cannot identify short-term pollutant episodes or even track common temporal patterns, because they only allow the quantification of cumulative air pollutants' levels and they are not as accurate as reference instrumentation (Castell et al., 2016; Küçükaçıl Artun et al., 2017; Rosario et al., 2016; Schneider et al., 2017). Modelling can be an effective tool to supplement air quality monitoring, mainly for forecasting (e.g. spatial distribution of pollutants), but it requires a highly specialised knowledge and its accuracy is relatively limited with recurring systematic errors (Pannullo et al., 2016; Rai et al., 2017; Vardoulakis et al., 2003).

Recently, some technological advances have been developed increasing acquisition of air quality data beyond the traditional monitoring methods, being able to provide less expensive and high-resolution spatiotemporal solutions to real-time air quality monitoring (Castell et al., 2013; Heimann et al., 2015; Rai et al., 2017; Velasco et al., 2016). Corroborating it, recent literature, including review studies, have been suggesting low-cost sensors for air quality monitoring as a promising alternative (Bhanarkar et al., 2016; Chojer et al., 2020; Gozzi et al., 2016; Jovašević-Stojanović et al., 2015; Kumar et al., 2015; McKercher et al., 2017; Morawska et al., 2018; Rai et al., 2017; Snyder et al., 2013; Thompson, 2016; Williams, 2020; Zhou et al., 2015). This new generation of air sensors are presented as low-cost, small size and weight, and low power consumption (Sun et al., 2016), and the installation and maintenance costs could be reduced, offering the possibility of getting larger spatial and temporal coverage, especially in situations where traditional monitors are impractical (Miskell et al., 2017, 2019). As such, low-cost sensor city-wide networks may improve urban air quality monitoring by: i) supplementing conventional air pollution monitoring (Rai et al., 2017); ii) supporting decision-making and informing the public (e.g. designing new strategies for air pollution control and mitigation) (Castell et al., 2016); iii) detecting pollution hotspots (Kumar et al., 2015); (iv) assessing real-time exposure (Piedrahita et al., 2014; Rai et al., 2017; Viana et al., 2015); v) validating atmospheric dispersion models (Borrego et al., 2016); and vi) performing spatially detailed mapping of air pollution (Schneider et al., 2017).

However, low-cost sensors still present limitations worthy of research (Ródenas García et al., 2022; WMO, 2021). The lack of selectivity and stability of sensors are generally found problematic (Borrego et al., 2016; Spinelle et al., 2015), as well as the cross-sensitivity and the influence of local conditions (e.g. temperature, relative humidity) (Afzal et al., 2012; Mead et al., 2013). The ageing and wear suffered over time (reducing the sensor lifetime) represent other disadvantages (Kim et al., 2022; Tancev, 2021). Another main challenge is the lower accuracy than reference instruments (Fishbain et al., 2017; Mead et al., 2013; Piedrahita et al., 2014). These limitations are still preventing the total adoption of low-cost sensors for air quality monitoring, whereby these sensing technologies are currently still being tested against reference monitoring methods (Borrego et al., 2016; Castell et al., 2016; Fishbain et al., 2017; Miskell et al., 2017; Schneider et al., 2017).

Calibration methods have been developed by modelling the association between measured concentrations with reference instruments and corresponding low-cost sensor responses in order to address these limitations (Morawska et al., 2018; Spinelle et al., 2017). Miskell et al. (2018) tested a simple, remote, continuous calibration technique for O_3 (semi-blind calibration), suitable for application in a hierarchical network, featuring a few well-maintained validated instruments and a more significant number of low-cost devices. Kim et al. (2018) explored an automated, in situ strategy for the calibration of sensors embedded in an air quality sensor network, consisting of a model that includes cross-sensitivities, chemical conservation equations, global and/or regional background of pollutants and temperature dependence. Spinelle et al. (2015, 2017) evaluated the performance of different calibration methods between low-cost sensors and reference instruments for O_3, NO_2, nitrogen oxide (NO), CO and carbon dioxide (CO_2), in a semi-rural area of Po valley (Italy). These authors used both linear and multivariate linear regression (MLR) and Artificial Neural Networks (ANN) with raw data, with standardised data (normally rescaled values), and with calibrated (MLR) data. Cordero et al. (2018) applied MLR and more complex calibration models, namely Random Forests, Support Vector Machines and ANN in a preliminary study for determining how well low-cost sensors can accurately assess NO_2 concentrations.

However, there are still limited research regarding the validation and calibration of low-cost sensors in a city-wide monitoring network. Moreover, as far as known, no studies performed a multi-step calibration that included both the calibration between low-cost sensors to ensure that all are measuring in the same range and the calibration against a reference-grade instrument. Additionally, although some supervised machine learning algorithms were used before to build calibration models, boosting algorithms were never used to improve weak calibration models.

Thus, this work's main objective was to perform a post-deployment validation and calibration based on two step methods for ozone low-cost sensor of a city-wide network for air quality and meteorology monitoring using low-cost sensors, highlighting the major challenges encountered. The specific objectives were: i) to validate and calibrate ozone low-cost sensors using a multi-step strategy; ii) to compare the performance of different models for calibration, including both linear regression (multivariate) and advanced machine learning algorithms (ANN and adaptive and gradient boosting algorithms); and iii) to propose a novel calibration strategy, redesigned based on the lessons learned.

2. Materials and methods

2.1. Study location and data collection

Under the European Union (EU) funded project Future Cities, the deployment of the *UrbanSense* platform was the last step from three urban scale testbeds in Porto (Portugal), one of the oldest European cities, with around 235,000 inhabitants and the centre of a metropolitan area with around 1.7 million inhabitants (Luis et al., 2016). *UrbanSense* was an IoT based city-wide platform for ubiquitous environmental monitoring. It was composed of monitoring units called Data Collection Units (DCUs), containing low-cost sensors to monitor: i) air pollution, including carbon monoxide (CO), ozone (O_3) and particulate matter (PM); ii) meteorological parameters, including temperature (T), relative humidity (RH), luminosity, precipitation, and wind speed and direction; and iii) noise. Data on noise was not considered, as it was beyond this study's scope. Immediately after an inter-DCUs comparison process carried out at the Faculty of Engineering of the University of Porto (FEUP) campus (described in section 2.4), DCUs were deployed in the city between July and October 2015, logging data of the various air pollutants and meteorological parameters from that moment until August 2017 (when the last DCUs stopped communicating data) continuously in 1-min intervals. The principle of operation and data collection of the DCUs (and their sensors) has already been described elsewhere (Luis et al., 2016; Santos et al., 2018). The 23 DCUs were deployed in Porto city in strategic locations identified as representative of typical urban environments: parks/gardens, residential, industrial, educational, historical/touristic, heavy traffic, and beachfront areas

(Fig. S1 in Supplementary Material).

2.2. Preliminary validation

Although all DCUs included CO and PM low-cost sensors, there were not enough reliable data available from them. The CO concentrations were consistently below the lower detection limit of the sensors in the DCUs. Data on PM was not possible to compare with that from a reference-grade instrument as it was not clear the size fraction that the low-cost sensors were measuring, and there was no PM data from the reference DCUs. Similarly, the precipitation sensor, encompassed only in some of the DCUs (Table S3), presented incomplete (communication failures reaching 90.5%) and inconsistent data (values not comparable with the meteorological station) over the monitoring period, precluding to proceed with further analysis. In the pre-processing phase, evident extreme values, negative values, and continuing zeros were removed to improve the results' credibility and quality, whereby some DCUs had to be completely excluded from further analyses. In addition, sensors from DCUs located in beachfront areas (DCU_11 and DCU_12) suffered irreversible damages from sea spray aerosol shortly after their deployment (approximately 2 months after their deployment, with DCU_12 presenting 31% of O_3 drift). Thus, only four DCUs based on the highest data availability (DCU_9, DCU_16, DCU_21 and DCU_23) were included in subsequent analyses. In turn, two additional identical DCUs were used as reference DCUs (DCU_Ref1 and DCU_Ref2), being DCU_Ref2 placed in the same location of DCU_23. The location zone types and the parameters analysed in each studied DCU are summarised in Table S1, while the main characteristics of each sensor assessed in this study are summarised in Table S2 in Supplementary Material. In addition, Table S3 shows a summary of the communication failures on the low-cost sensors, O_3 drift and errors, as well as the relative humidity errors of each DCU.

The time-series profiles were evaluated to validate the low-cost sensors data (O_3 and meteorological). Hourly mean values were calculated, considering those with at least 50% of raw data. Daily, weekly, monthly and seasonal mean profiles were plotted for O_3 and meteorological variables from each DCU whenever data was available. These profiles were compared with the analogous patterns reported in previous publications using reference-grade instruments in Porto. Wind-speed and direction data obtained from the DCUs were compared with those from a meteorological reference-grade station located in a representative place at FEUP campus. A wind rose profile was created to visualise wind data (aggregating wind speed and direction), by using the openair package (Carslaw and Ropkins, 2012) from R software, version 3.4.3 (R Core Team, 2018).

2.3. Multi-step calibration strategy

A multi-step calibration strategy was planned to mitigate the low accuracy associated with the data from the low-cost sensors (Fig. 1). Two steps were applied: i) the inter-DCU calibration in the 1st step, in which all studied DCUs were calibrated by a similar DCU designated as the reference DCU (DCU_Ref1); and ii) the calibration with a reference instrument in the 2nd step, in which the reference DCU (DCU_Ref2) was calibrated with a reference-grade instrument.

According to the Portuguese legislation (Decreto-Lei nº 102/2010), any equipment using the reference method or equivalent can be considered a reference-grade instrument for air quality monitoring. Thus, *Aeroqual s500* was used as a reference instrument in this study. The 2nd step of the calibration was only performed for O_3, as this was the sensor available in both reference DCU and reference-grade instrument (*Aeroqual s500*).

The use of this multi-step calibration strategy is expected to significantly reduce the number of reference instruments required for the calibration of the low-cost sensors in a city-wide air quality monitoring network like *UrbanSense*, as it allows the possibility of extrapolating the model obtained in the 2nd step of the calibration (model 2) to the other DCUs producing calibrated data (DCU_i_cal).

2.4. Inter-DCU calibration

To allow inter-comparability between data from different DCUs, the inter-DCU calibration (1st step of the calibration) aimed to adjust the measurement ranges and correction of signal changes between the studied DCUs (DCU_9, DCU_16, DCU_21 and DCU_23) using another DCU as the reference DCU (DCU_Ref1). For each studied DCU, a separate reference DCU dataset was used according to the corresponding measurement period. Consequently, a model for each 'i' DCU was developed (Model1_DCU_i) and applied, resulting in the adjusted dataset, from now on defined as DCU_i_adj. The reference DCU and each of the analysed DCUs measured concomitantly with 1-min logged data, for at least 4 days at FEUP campus outdoors (Table S4 in Supplementary Material), simulating the conditions of the city's measurement locations. After

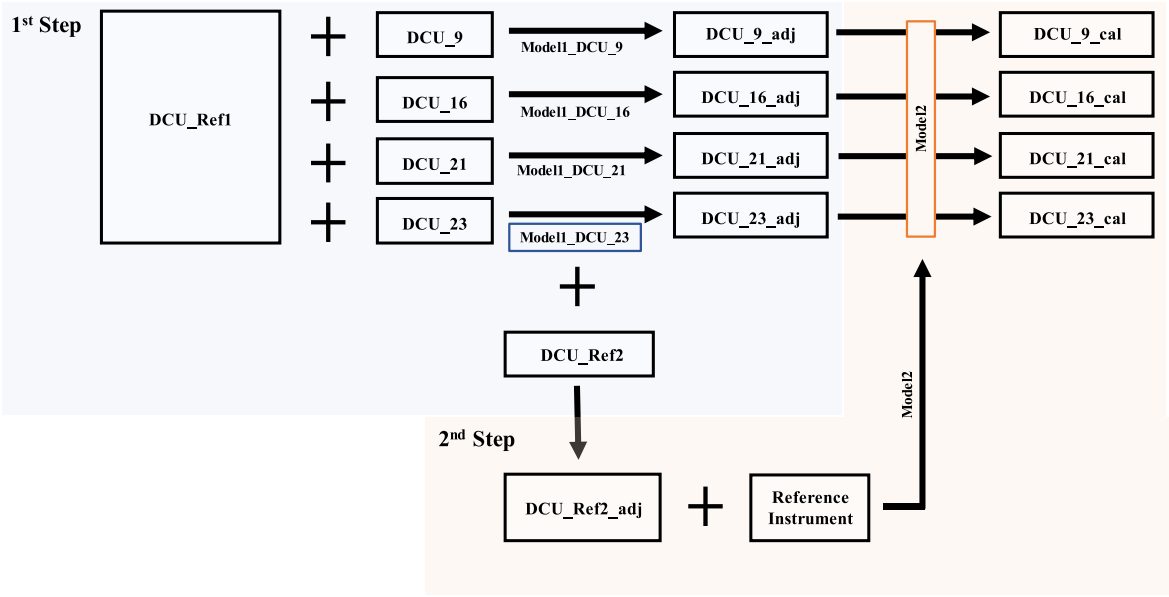

Fig. 1. Multi-step calibration strategy.

that, they were installed in their respective locations in Porto city (Fig. S1 in Supplementary Material).

In this 1st step of the calibration, data collected from the low-cost sensors in the studied DCUs were calibrated against the reference DCU (DCU_Ref1). This calibration was done using multivariate linear regression (MLR) models, thus assuming the linearity between each low-cost sensor data and the data from the reference DCU (DCU_i vs DCU_Ref1). An MLR model is an extension of a simple linear regression model, considering more than one explanatory variable in a prediction equation, for a response variable (Tranmer and Elliot, 2008). MLR general equation was represented by Equation (1).

$$\widehat{Y} = P_0 + P_1X_1 + \ldots + P_nX_n \quad (1)$$

where $\widehat{Y}$ represented the DCU sensor response and X_i (i = 1, …, n) corresponded to the explanatory variables (predictors) and P_i (i = 0, 1, …, n) represented the regression coefficients, usually estimated by least square. A full MLR model was initially built with all the potential explanatory variables in each case (Supplementary Material, Table S5). Then, a stepwise model selection was run to select the MLR model that best fitted the data. Data were tested for normality with Shapiro-Wilk and Kruskal-Wallis tests, and when normal distribution was not followed, data were normalised with a z-score normalisation (mean of 0 and standard deviation of 1).

This 1st step of the calibration was also performed using artificial neural networks (ANN), in an attempt to improve the calibration results obtained with linear models. Sometimes, linear models cannot incorporate the nonlinearities of the results of some air pollutants or meteorological variables, whereby in some cases, the ANN may suit better for this type of analyses (Sousa et al., 2007). ANN consists of pattern recognition without any target attribute. The most widely used ANN models are perceptron-based neural networks, which were inspired by the human brain's events. Typically, an ANN consists of three layers: the input layer, the hidden layer, and the output layer. The input layer receives the data and begins its transmission in the ANN structure by the subsequent layers until the output layer produces the structure's end output (Park et al., 2018). The weighted sum of the inputs is formed to compose the activation of the unit. The activation signal is passed through an activation function (usually sigmoid, hyperbolic tangent or linear) to produce the unit's output (Spinelle et al., 2015). It has been referred that an ANN with a single hidden layer, including enough neurons, can approximate any function with the desired accuracy (Mba et al., 2016). ANN can be represented by the general Equation (2).

$$y = h\left(w_0 + \sum_{i=0}^{n} w_i u_i\right) \quad (2)$$

where h is the activation function, w_i are the weights and u_i are the inputs.

In this calibration step, ANN models were trained using one to three hidden layers, although the results were in general better with only one. A cross-validation method was applied to both linear models and ANN, namely a holdout method, by dividing the dataset into two sets (training and testing) of each air pollutant or meteorological variable analysed. In all cases, 80% of the data was considered for training data and 20% for testing, randomly selected from the complete calibration dataset. The training dataset was used to determine the model topology, while the testing dataset was used to compute the model performance.

The performance indexes computed were mean bias error (MBE) and root mean square error (RMSE) for both MLR and ANN. In addition, the coefficient of determination (R^2) was calculated for MLR model. MBE indicates how observed results are over or underpredicted, while RMSE measure residual errors, which summarise the difference between the observed and predicted values (Gardner and Dorling, 2000; Humphrey et al., 2017). In turn, coefficient of determination indicates the proportion of the variance in the dependent variable (that is predicted by the model) and the predictor variable (Nagelkerke, 1991). Thus, models were selected based on the lowest MBE and RMSE and the highest R^2 (whenever available).

R software version 3.4.3 (R Core Team, 2018) was used for this analysis, and ANN models were created using the *neuralnet* package (Fritsch et al., 2019). The level of statistical significance was set at 0.05, except when stated otherwise.

2.5. Calibration with a reference instrument

In the calibration with the reference-grade instrument (the 2nd step of the calibration) model 2 was created by calibrating adjusted data from a reference DCU (DCU_Ref2_adj, dataset adjusted using the selected best model in the 1st step of the calibration) with data from the reference-grade instrument (*Aeroqual s500*). For that, model 1 resulting from DCU_23 calibration in the 1st step was used as the best model because DCU_Ref2 was placed side-by-side with the reference-grade instrument in the sampling site of DCU_23. O_3, T and RH were monitoring continuously by DCU_Ref2 with 1-min log intervals, from 11th May to July 2, 2016 (Table S4 in Supplementary Material).

In this 2nd step of the calibration, non-linear models using machine learning (ML) were used for the regression analysis. As the first step, data acquired from DCU and reference-grade instrument were merged after removing null or missing data points from the datasets and the calibration was implemented on the 10-min means. Resampling methods were used to randomly divide the merged dataset into training and test sets in an 80–20 ratio.

ANN and several boosting models were implemented after the data resampling. The ANN is a particulate case of ML algorithms used in this 2nd step of the calibration. In this step, a multi-layer perceptron ANN was used, and several structures were tested (number of hidden layers and neurons). Rectified linear unit (ReLU) function was used as the activation function. The data were standardised to compensate for the skews in distributions using the StandardScalar function of scikit-learn library before the training step. Adam algorithm was the optimisation function for training the model, and it based the regression on the minimisation of mean square error loss.

Besides ANN, boosting ensemble algorithms were also implemented. These consist of a family of algorithms that convert weak learners to strong ones. They create a sequence of models that attempt to correct the models' mistakes before them in the sequence. Three different boosting ensemble algorithms were used: Adaptive Boosting (AdaBoost) regressor, Stochastic Gradient Boosting Regressor (GBR), and Extreme Gradient Boosting (XGBoost) regressor.

All the models underwent the optimisation known as "hyper-parameter tuning". Hyper-parameters are the parameters that control the learning process of the models. For ANN, the hyper-parameters can be network structure parameters like number of hidden layers, number of neurons for each layer, and they can also be the algorithm training parameters like learning rate, batch size, among others. For boosting models, the hyper-parameters most commonly optimised are learning rate and number of boosting iterations. K-fold cross-validation was used for hyper-parameter tuning of the boosting models, which estimates the performance of a model with less variance than a single train-test split. It splits the dataset into *k* subsets, where the model training occurs in *k-1* subsets, and the testing occurs on the final subset (Rodríguez et al., 2010). Grid search method with cross-validation from scikit-learn library was used in to tune the hyper-parameters in this study. Table 1 shows the hyper-parameters tuned in each model.

Statistical performance indexes computed in this calibration step were R^2, RMSE and MBE. Before and after hyperparameter tuning, all the models were checked for convergence, over/under-fitting, and the statistical metrics.

Python 3.7 was used with Jupyter Notebook interface for data analysis in this 2nd step of the calibration. The *pytorch* library was used for ANN models, and the scikit-learn library was used for the other ML

Table 1
Hyper-parameters tuned in each model.

Model	ANN	GBR	AdaBoost	XGB
Hyper-parameters	number of hidden layers, number of neurons, learning rate	number of boosting stages, learning rate, maximum depth	number of boosting stages, learning rate	number of boosting stages, learning rate, maximum depth, subsample, gamma

ANN – Artificial Neural Network; GBR – Gradient Boosting Regression; AdaBoost – Adaptive Boosting; XGB – Extreme Gradient Boosting.

models.

3. Results and discussion

3.1. Preliminary validation

Although DCUs included low-cost sensors for various air pollutants, there were no PM, CO, and precipitation data due to the reasons pointed out in the materials and methods section. Two relevant lessons must be learned from this, and taken into account by those responsible for developing and installing the DCUs or the network. First of all, they must ensure that the chosen low-cost sensors can detect within the typical range of concentrations observed in the city or monitoring site, i.e., above the lower and below the upper limit of detection. Secondly, they should ensure that the reference DCU have all the low-cost sensors. Otherwise, the proposed multi-step approach cannot be completed. In the case of PM, they must also ensure a sensor's choice with a known and adequate PM size fraction (at least PM_{10} and/or $PM_{2.5}$).

The datasets obtained from the DCUs had a relevant number of missing values, mainly due to the low-cost sensors' failure readings and data communication failures (Table S3). Some O_3 and RH raw data suffered from error and inconsistent values (evident extreme values, negative values and continuing zeros) that were removed (reaching 100% of the data in some cases for both O_3 and RH), thus contributing to missing data. Precipitation data were not possible to analyse due to the reasons pointed out in the materials and methods section: DCUs presented incomplete and inconsistent data over the monitoring period (with communication failures varying between 4.7% and 90.5%). In addition, a wide range of communication failures was registered for the RH sensor (4.3%–93.9%). However, error values were registered among the available data (reaching 100% in some cases). Temperature, luminosity, and wind speed and direction sensors suffered from communication failures (up to 79.8%, 88.7%, 89.0%, 91.0%, respectively), but all the data recorded was used, while O_3 data presented a maximum of 10.1% of data drift (among of the four studied DCUs). Most of that missing data corresponded to the last monitoring periods of the sensors. O_3 sensors seemed to start drifting in different periods depending on the DCU: from the 1st month on DCU_23, 3rd month on DCU_16 and 8th month on DCU_9 and DCU_21 (Fig. S2 in Supplementary Material). Such behaviour seems to evidence a rapid degradation of the low-cost sensors along the time, indicating they may have a reduced lifetime. Aleixandre and Gerboles (2012) have already stated that low-cost sensors' responses change over time, thus needing recalibration and, in some cases, replacement. With the cleaned data obtained from the DCUs, mean values (daily, weekly, monthly, and seasonal) were calculated, and time-series profiles were drawn to represent mean scenarios for O_3, T, luminosity and wind speed and direction. RH was not included, as none of the four studied DCUs had more than 50% of RH raw data available in

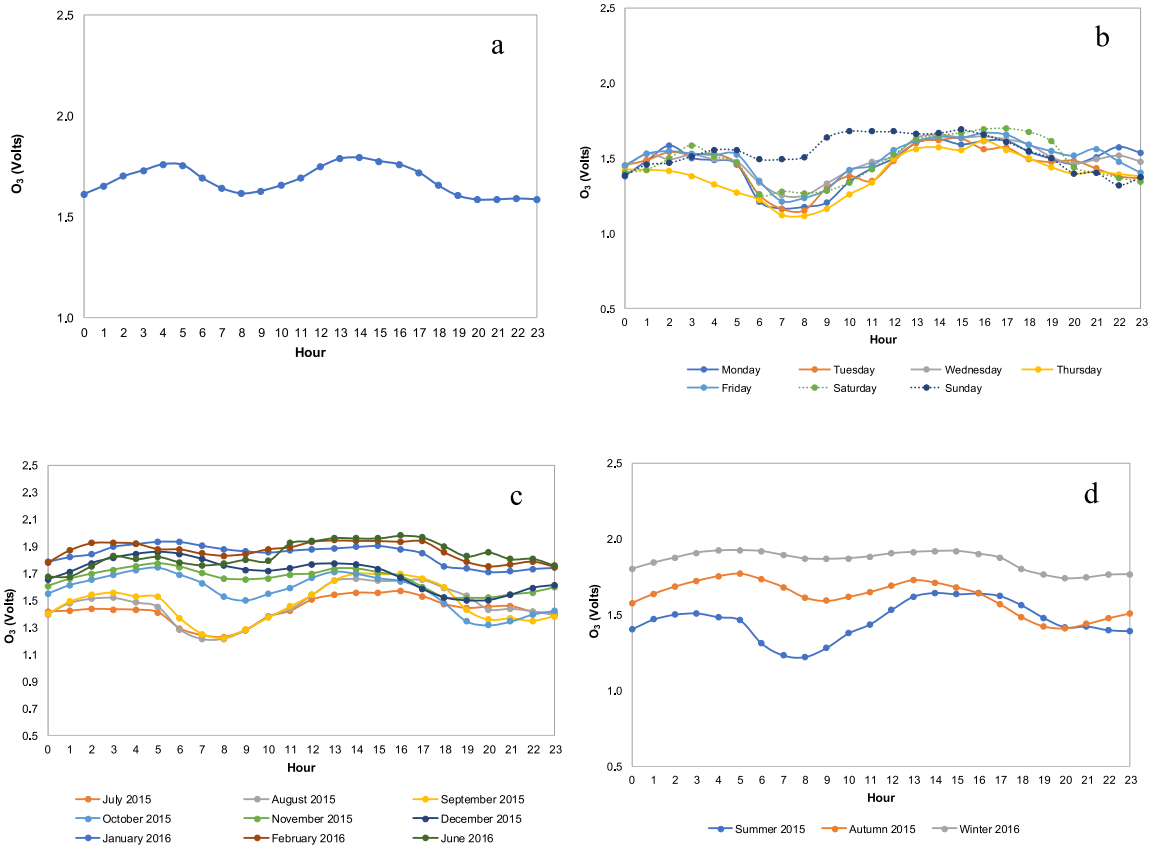

Fig. 2. Daily mean profiles of O_3 readings in the DCU_9 considering data: (a) complete; (b) by weekday; (c) by month; and (d) by season.

this preliminary validation period.

O_3 profiles were similar for all locations (in Fig. 2 from DCU_9 and from the remaining DCUs in Supplementary Material Figs. S3–S5), characterised by an increase from the end of the morning until the end of the afternoon, where the concentration peak varied between 13:00 and 16:00. In some locations, a slight increase from the early dawn until the early morning was also observed, also known as nocturnal maxima. These are the usual O_3 profiles at Porto city (Alvim-Ferraz et al., 2010; Sousa et al., 2011), more accentuated in some locations than in others as expected. Another study, carried out in Auckland city (New Zealand), used a network of low-cost gas sensors deployed in 12 sites and reached similar daily O_3 profiles in most studied locations (Weissert et al., 2017). However, contrary to what was expected, spring and summer concentrations were lower than autumn and winter, which may indicate that the O_3 sensor suffered some wear along the monitoring period, leading to the registered drifts (Fig. S2 in Supplementary Material). Consequently, at least a monthly calibration is expected to be necessary.

Temperature and luminosity data from the DCUs (in Figs. 3–4 from DCU_9 and DCU_21, respectively, as well as from the remaining DCUs in Supplementary Material Figs. S6–S11) had very similar patterns, characterised by a typical increase at the beginning of the morning and a decrease at the end of the afternoon showing the influence of the solar radiation intensity. Similar profiles were found in studies carried out in European cities, including Porto, that evaluated urban air temperature using reference-grade monitoring stations (Sousa et al., 2011), low-cost sensors/networks (Meier et al., 2017) or modelling (Borrego et al., 2014; Steeneveld et al., 2016). Temperature and luminosity differences between seasons were consistent with the climate in Porto city (Sousa et al., 2008), with the highest values in summer and the lowest in winter. Temperature and luminosity low-cost sensors did not show data issues like error, inconsistent data or drift along the time, as the O_3 sensors. The study of these variables, especially temperature, in urban areas are essential to evaluate episodes of air pollution, because temperature is often related to situations of stagnant atmospheric circulation and formation of photochemical pollutants (Borrego et al., 2014). In this study, there was no apparent correlation between T/luminosity and O_3 (correlation matrices in Supplementary Material Tables S5–S8), which may also be related to O_3 low-cost sensors' drift problems already mentioned.

The wind rose from DCU_21 and from the reference-grade meteorological station in July 2016 was plotted in Fig. 5 as an example, while the other wind data collected were plotted in Supplementary Material, Figs. S12 and S13. The wind roses obtained from DCU data showed dominant winds from north (N) and varying from the Northwest (NW) to Northeast (NE), with speeds reaching 32.6 km h^{-1}. Dominant directions were similar to those registered by the meteorological reference-grade station. In this study, the monitoring period was mostly during summer months, so the dominant wind direction was coincident with what was found in the literature – main circulations in Porto city are from West-Northwest (WNW) in summer (Monteiro, 1993; Santos et al., 2002). In general, the low-cost anemometer from the DCU registered higher wind speed values than the meteorological reference-grade station. Moreover, there were discrepancies in the wind profiles when comparing data from the DCU with that from the meteorological reference-grade station. The interpretation of these results was difficult as this wind speed and direction analysis was merely based on a single DCU. Moreover, other factors could have also contributed to those differences, namely a poor accuracy of the low-cost sensors used in the DCU, as well as local conditions which may have influenced the results (although located in a representative site, the meteorological reference-grade station was not installed in the exact location as the DCU_21). Porto's city is located in the Atlantic Ocean coast, from where most of the moisture affecting western Iberia Peninsula arrives, with the prevailing winds occurring mostly in winter months (Russo et al., 2014; Trigo et al., 2002). Contrary to the expected, the strongest winds were

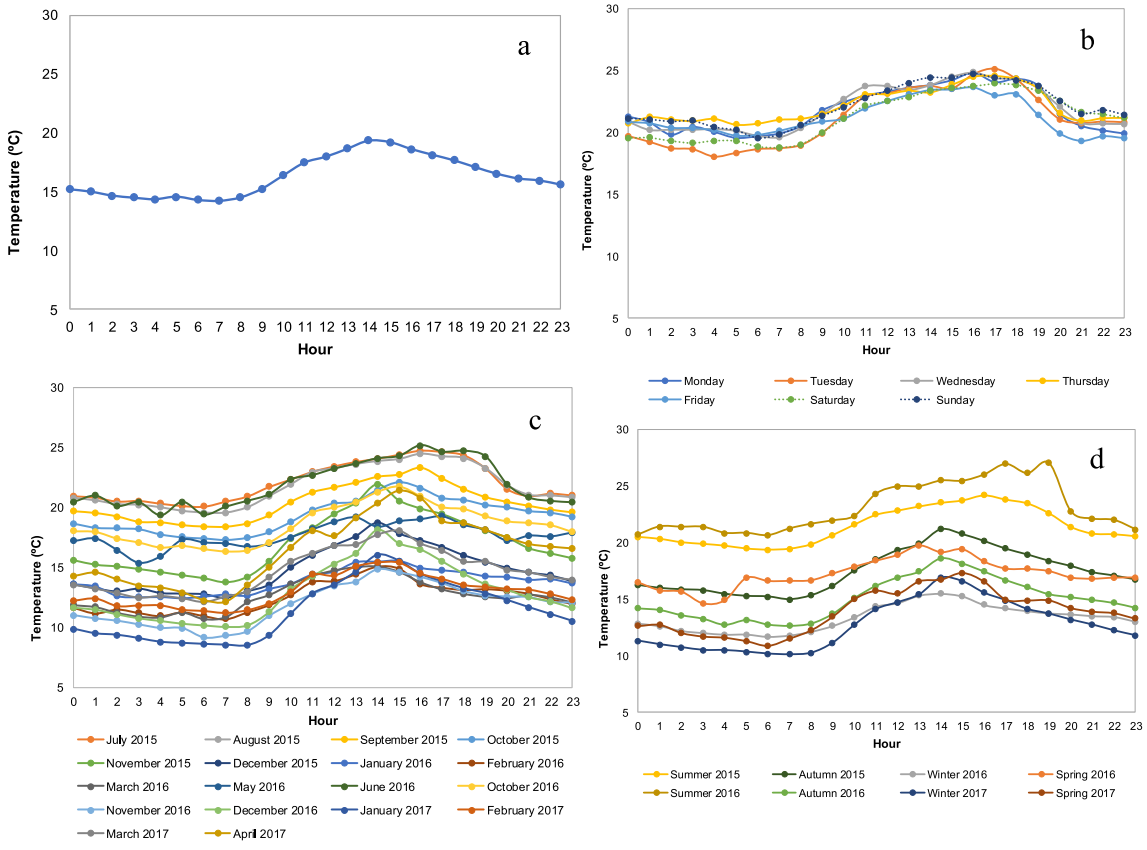

Fig. 3. Daily mean profiles of temperature readings in the DCU_9 considering data: (a) complete; (b) by weekday; (c) by month; and (d) by season.

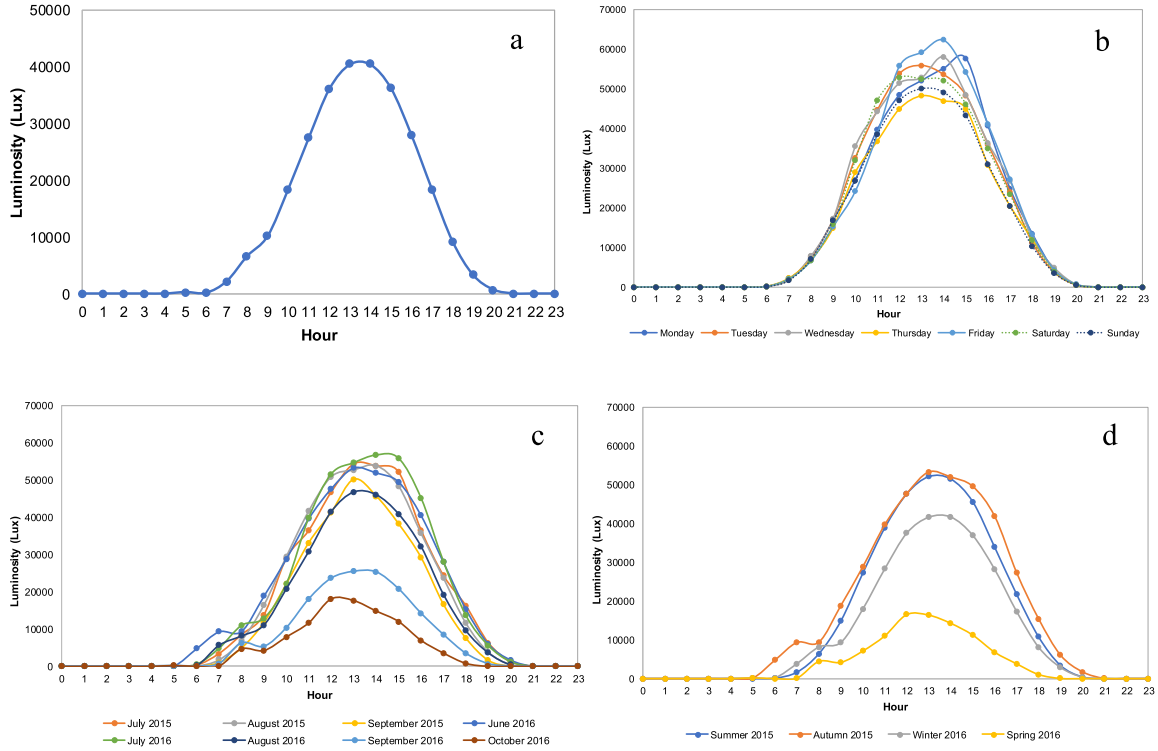

Fig. 4. Daily mean profiles of luminosity readings in the DCU_21 considering data: (a) complete; (b) by weekday; (c) by month and (d) by season.

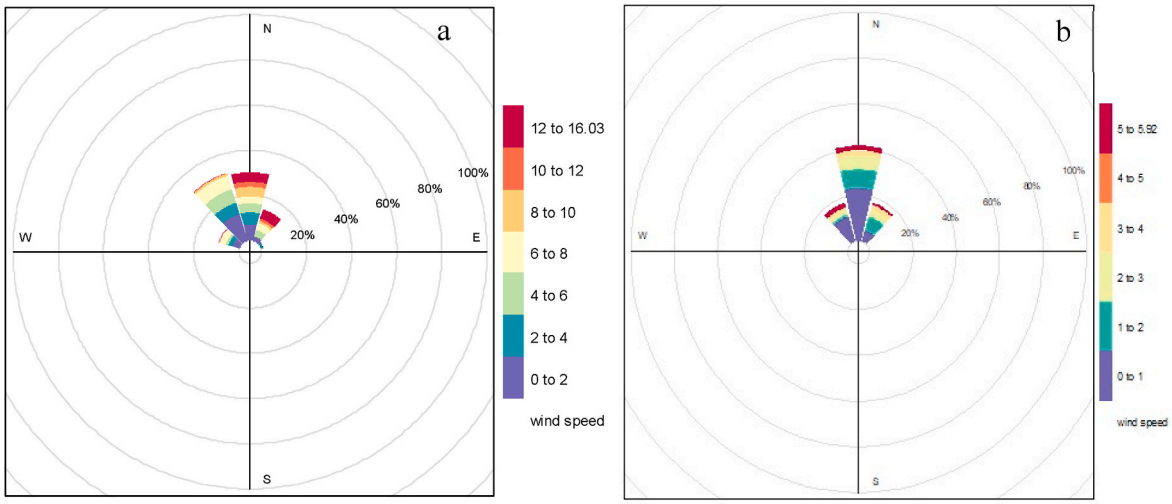

Fig. 5. Wind roses for July 2016 from (a) DCU_21 and from (b) reference-grade meteorological station located at FEUP campus.

recorded in August (maximum of 32.6 km h^{-1}) and the weakest in January (maximum of 9 km h^{-1}). Nevertheless, it should be noted that there was only one month of recorded data in Winter (January 2016), which did not represent the whole winter season, whereby extrapolations should not be made.

Summarising, the results from this initial validation of the low-cost sensors in the studied DCUs showed three distinct situations: i) PM and CO low-cost sensors only recorded unreliable data, and RH and precipitation low-cost sensors did not collect enough data, making it impossible to conduct any analysis; ii) wind speed and direction low-cost sensors provided limited data (in quantity), making it impossible to conduct further analyses, although in some cases that data seemed to show patterns close to the reference or as expected; and iii) O_3, T and luminosity low-cost sensors provided promising data, with patterns similar to those from the reference or as expected.

3.2. Inter-DCU calibration

The inter-DCU calibration was considered the 1st step of the proposed calibration strategy. In the present study, this step could only be performed for O_3, T and RH as these were the low-cost sensors available in the reference DCU (DCU_Ref1). Although there were limited RH data from the DCUs for the preliminary validation, RH data were available from the 1st step of calibration (inter-DCU calibration). On the other hand, luminosity data from the DCUs were used as a potential explanatory variable in the models at this step, although the reference DCU did

not have that sensor. The available data for this step varied from 5 consecutive days (DCU_16), to 12 consecutive days (DCU_21), 13 consecutive days (DCU_9) and 14 consecutive days (DCU_23). Data collection periods were detailed in Supplementary Material (Table S4). To understand the behaviour of O_3 concentrations along the time in each DCU, a time-series plot was drawn between the reference DCU (DCU_Ref1) and each studied DCU_i (Fig. S14 in Supplementary Material). Similar profiles were observed for each comparison. Moreover, the results showed a strong correlation of each variable (O_3, T and RH) between each DCU and the reference DCU, with $r > 0.89$ (Supplementary Material Tables S6–S9). In each DCU, O_3 was positively correlated with T, and negatively with RH. There was a strongly negative correlation between T and RH ($r < -0.88$), except in DCU_23 ($r = -0.35$).

Table 2 shows the selected best MLR models' performance indexes in both training and testing periods at this inter-DCU calibration (1st step of the calibration, Reference DCU vs DCU_i) for O_3, T and RH. The performance indexes were generally good for both training and testing periods. The R^2 were consistently higher than 0.80, confirming the selected models' good fit. Regarding O_3, it was possible to verify very good results in the four studied DCUs with R^2 varying between 0.822 and 0.958 and MBE and RMSE close to zero, for both training and testing periods. T and RH sensors also presented very good and similar performance indexes. All the best models included T and RH as explanatory variables, evidencing some sensor sensitivity to these variables, which may indicate their relevance in the calibration process, particularly in O_3 low-cost sensors.

The linear regression models (MLR) showed similar performance compared with the non-linear ANN models (Supplementary Material Tables S10–S13); however, MBE and RMSE were slightly lower for MLR models. Thus, for this 1st step of the calibration proposed strategy (inter-DCU calibration), MLR models should be used instead of non-linear ANN models due to the higher complexity of the latter. Still, one cannot exclude that a more significant amount of data may benefit the performance of ANN.

3.3. Calibration with a reference instrument

The calibration of the O_3 sensor in the reference DCU (DCU_Ref2) against a reference-grade instrument (*Aeroqual s500*) was considered the 2nd step in the proposed strategy. During the initial pre-processing phase, 1-min values logged by the devices were acquired. The DCU showed many errors or inconsistent values, namely evident extreme values due to sensor malfunctioning and missing data points. These issues diminished by taking 10-min means. Hence, this 2nd step of the calibration was performed considering 10-min mean values. The mutually exclusive 10-min mean data points available for all the sensors were merged together (n = 4599). To visualise the data, a scatter plot matrix for all the sensors presents in the reference DCU and the reference-grade instrument was initially drawn (Supplementary Material, Fig. S16). Pearson correlation coefficient (Supplementary Material Table S14) elucidated that both T and RH correlated with O_3. Due to the very strong correlation of RH and T, only T was used as the explanatory variable for subsequent steps. T was chosen over RH because T data from DCUs' low-cost sensors showed higher quality in this study's previous stages (preliminary validation and 1st step). Moreover, the low-cost RH sensor from the DCU_Ref2 recorded more invalid and missing data than the T sensor.

Before undergoing ANN regression modelling, the data were checked for skewness in distribution. The O_3 values from the reference-grade instrument showed a minimal skewness (0.151), but the adjusted O_3 and T data from the DCU_Ref2 showed considerable skewing (−0.545 and 0.832 respectively). Hence, the datasets underwent standardisation before modelling. Both time-series plot and scatter plot of O_3 concentrations were drawn to better visualise a linear relation between the O_3 values from the reference DCU and from the reference-grade instrument (Supplementary Material Figs. S15 and S17). The time-series plot showed that there were several missing data, which made a more detailed analysis difficult. However, both reference-grade instrument and low-cost sensor from DCU did not always seem to follow the same trend, so it is foreseen the application of more complex calibration models. In addition, a weak correlation was observed (r = 0.375) evidencing the low data reliability of these sensors. Other authors (Spinelle et al., 2015, 2017) have also reported poor correlations and weak data accuracy of low-cost sensors, and concluded that the use of ANN increased the strength of association between estimated and reference data.

From the application of ANN in this 2nd step of the calibration, an initial model was obtained. Table 3 shows the training and test phases performance index results of all the models (base model refers to default hyper-parameters).

For ANN, a minimal improvement was found by adding more hidden layers and changing the number of neurons of the initial/base model (two hidden layers with 16 neurons and a learning rate of 0.01 with ReLU activation function). Although final ANN model showed a good convergence, poor accuracy could be observed even after implementation of hyper-parameter tuning. Moreover, the results of the application of Adaboost regressor, XGB regressor, and GBR were better than those with ANN (Table 3).

Model performance enhanced upon hyper-parameter tuning of boosting models by varying the parameters mentioned in Table 1. The automated tuning using exhaustive grid search and 5-fold cross-validation yielded the favorable parameters. The calculated hyper-

Table 2
Performance indexes of the selected linear models in the inter-DCU calibration (1st step of the calibration, Reference DCU vs DCU_i), for O_3, temperature and relative humidity.

Sensor	DCU_i	Model [Explanatory Variables]	R^2		MBE		RMSE	
			Training	Test	Training	Test	Training	Test
O_3	DCU_9	MLR [O_3, RH, T]	0.947	0.952	4.22E-18	3.08E-03	3.74E-02	3.50E-02
	DCU_16		0.956	0.958	−4.11E-17	−2.62E-04	3.49E-02	3.45E-02
	DCU_21		0.822	0.824	4.73E-18	−3.25E-04	8.58E-02	8.57E-02
	DCU_23		0.909	0.909	−9.83E-18	−5.07E-04	6.12E-02	6.07E-02
T	DCU_9	MLR [T, RH]	0.887	0.887	−9.49E-17	7.39E-03	0.675	0.668
	DCU_16		0.964	0.961	−7.97E-16	−1.09E-02	0.410	0.430
	DCU_21		0.855	0.862	3.23E-16	1.41E-04	0.650	0.653
	DCU_23		0.943	0.945	−2.2E-16	1.06E-02	0.402	0.398
RH	DCU_9	MLR [RH, T]	0.828	0.813	4.66E-17	−4.42E-02	2.00	1.96
	DCU_16		0.947	0.947	1.90E-15	2.11E-02	1.16	1.32
	DCU_21		0.864	0.875	−1.51E-16	1.21E-02	2.57	2.63
	DCU_23		0.968	0.970	−5.01E-16	−2.99E-02	1.27	1.24

O_3 – ozone; T-temperature; RH – relative humidity; R^2 – Coefficient of Determination; MBE – mean bias error; RMSE - root mean squared error; MLR - multivariate linear regression.

Table 3
Performance indexes of all the models used in the 2nd step of calibration.

	Model	R^2		MBE		RMSE	
		Training	Test	Training	Test	Training	Test
Base Models	GBR	0.387	0.307	8.91	1.57	29.1	29.5
	Ada	0.241	0.218	1.50	3.26	32.4	31.3
	XGB	0.754	0.243	−0.001	1.21	18.5	30.8
	ANN	0.184	0.144	−0.302	−6.45	33.4	30.1
Post-Tuning	GBR	0.423	0.319	0.000	1.59	28.2	29.2
	Ada	0.252	0.247	0.287	1.80	32.2	30.7
	XGB	0.373	0.324	1.04	2.64	29.4	29.1
	ANN	0.177	0.177	1.64	−3.65	33.5	29.5

ANN – Artificial Neural Network; GBR – Gradient Boosting Regression; AdaBoost – Adaptive Boosting; XGB – Extreme Gradient Boosting.

parameters were then used to train the models and the resulting performance indexes can be observed in Table 3. Fig. 6 shows the plots of the all the tuned models versus the reference (with respective performance indexes).

The results show that hyper-parameter tuning only marginally improved the models, which could not be sufficient to get close to the reference concentrations. Models GBR and XGB performed the best out of all four models trained reaching very low MBE values and improved R^2 values of 0.32 (Fig. 6 (c) and (d)).

3.4. Limitations of the study

Despite this study's novelty, particularly by testing a novel multi-step calibration strategy in devices using low-cost sensors to monitor air pollution and meteorological variables in a city-wide network with 23 DCUs, this study was not free from limitations.

Errors and failures associated with DCUs and their low-cost sensors, partially explained by their reduced lifetime, became a relevant limitation of this study. That significantly limited the amount of available DCUs (limiting the analyses to four DCUs) and data in both preliminary validation and calibration steps, affecting the implementation of the multi-step calibration strategy proposed initially.

Although the calibration models improved O_3 data quality in the 1st step and marginally in the 2nd step of the calibration strategy proposed, the limited sampling time, did not allow to quantify the sensors' drift along the time. While it was not certain that a more extended sampling period could have improved raw data quality, it would certainly improve the ML calibration models' performance and, consequently, the accuracy of the calibrated data. However, to get better results and correlations between devices, it might be interesting to test the performance of other low-cost ozone sensors based on different technology, such as electrochemical cells. Other study limitations included: i) the absence of meteorological sensors in the reference DCUs, which did not allow the calibration of the low-cost luminosity sensors; and ii) using two different reference DCUs (DCU_Ref1 and DCU_Ref2) could have introduced some issues related to reproducibility, repeatability, and accuracy. The inherent variability in the manufacturing process results in differences in the reactivity of the metal oxide substrate of individual sensors, which influences the repeatability and hampers the robustness (Zhang et al., 2014).

3.5. Proposal of a new calibration approach

The lessons learned with this study and the results obtained evidence that the initially proposed multi-step calibration strategy needs to be redesigned. It is not possible, however, to define one universal calibration model that fits all situations. This is confirmed by the variability of the results obtained in this study and supported by the literature.

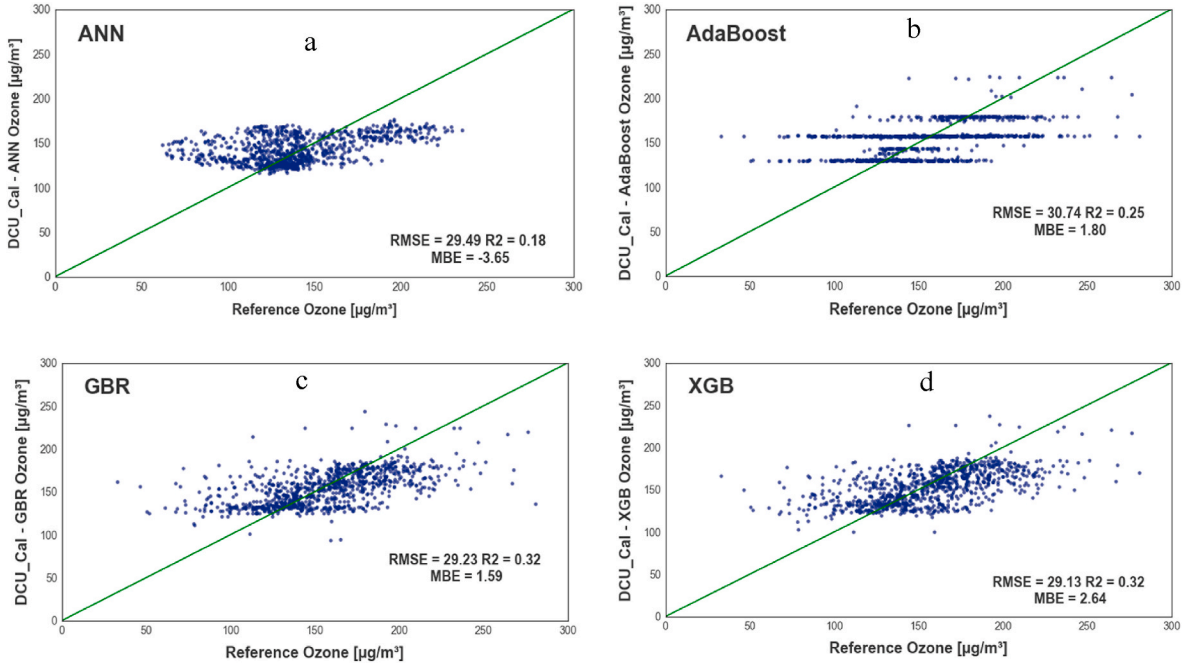

Fig. 6. Scatter plots of the test phase results of the four regression models implemented in the 2nd step of the calibration post hyper-parameter tuning: (a) ANN: Artificial Neural Network; (b) AdaBoost: Adaptive Boosting; (c) GBR: Gradient Boosting Regressor; and (d) XGB: Extreme Gradient Boosting.

Cordero et al. (2018) concluded that it was not possible to propose a specific model to be generally used for future applications, due to the variability of results, after applying MLR to the data, and testing more complex calibration models (random forests and support vector machines).

Thus, a new multi-step calibration strategy is now suggested (Fig. 7). Before the actual calibration, the low-cost sensors to be installed in the DCUs should be appropriately validated, meaning that it should be sensible to what is going to measure (either by self-testing or by access to manufacturers' tests results), confirming that they can monitor the typical trends and ranges in the site where they will be deployed. Afterwards, the multi-step calibration strategy should be applied in 2 steps. In the 1st step (pre-deployment calibration), each DCU (low-cost sensors) to be deployed is calibrated with multiple reference DCUs together (model 1), resulting in the adjusted data. Simultaneously, those reference DCUs are calibrated with a reference-grade instrument. All this 1st step occurs in the same calibration site and before deployment. In the 2nd step (post-deployment calibration), each DCU is deployed on its final monitoring site, and its adjusted data is calibrated using one reference DCU previously calibrated with the reference-grade instrument (model 2), resulting in the calibrated data of each DCU. This procedure can be done cyclically and continuously in time, as the reference DCUs can be mobile going through every deployed DCU to correct them for data deviations. The calibration procedure should be applied monthly to try to minimise data drifts. Thus, the ideal scenario would be to have a 1 to 4 ratio of reference DCUs to deployed DCUs. It is crucial to ensure that: i) the reference DCU is similar to the other DCUs (although designated as reference); ii) the reference DCUs have all the same low-cost sensors as the most complete deployed DCU; and iii) there are reference-grade instruments (high-performance sensors) to monitor all the variables measured by the low-cost sensors in the DCUs. This new strategy may reduce the number of reference-grade instruments needed, bringing more reliability and robustness to the low-cost sensors' data.

Although it is not guaranteed that the proposed novel approach will result in high data accuracy, it is expected to work as most of the challenges encountered along the post-deployment validation and calibration are met.

4. Conclusions

Although the city-wide networks for air pollution monitoring based on low-cost sensors are promising to analyse the city's environmental conditions in real-time with detail and scale never before possible, this study identified several post-deployment challenges related to their validation and calibration. Those challenges were related to data acquisition or accuracy of the data, and they must be considered by those responsible for developing and installing these networks. The preliminary validation of the low-cost sensors installed in the DCUs showed three situations: i) PM, CO and precipitation only recorded unrealiable data, and RH did not collect enough data, making it impossible to conduct any analysis; ii) wind speed and direction provided limited data (in quantity), making it impossible to compare with reference-grade instruments; and iii) O_3, T and luminosity resulted in promising data, with patterns similar to those from the reference or as expected.

In the 1st step of the calibration, MLR led to more accurate calibration models than non-linear ANN, with slightly better performances for training and testing data (low MBE and RMSE, with a R^2 higher than 0.80). Thus, MLR models were suggested to be used for inter-DCU calibration instead of ANN models due to the latter's higher complexity. In the 2nd step, the calibrated O_3 data resulted in a marginal improvement from the raw data. Results from the ANN models almost reduced RMSE values to zero, yet they showed no real predictability, based on R^2 results. Results from other supervised ML techniques (boosting algorithms), namely results from GBR (both with default and post tuning hyper-parameters) were able to increase the predictability by more than double while also decreasing the error values close to zero. Still, the calibrated O_3 data could not achieve sufficient predictability ($R^2 \sim 0.32$).

The lessons learned with the present study evidenced that the multi-step calibration strategy initially proposed needed to be redesigned. Thus, a multi-step calibration strategy was proposed, based on two steps: i) pre-deployment calibration (1st step), when each DCU to be deployed is calibrated with multiple reference DCUs together resulting in the adjusted data; and ii) post-deployment calibration (2nd step) when each DCU is deployed, and its adjusted data is calibrated using one reference

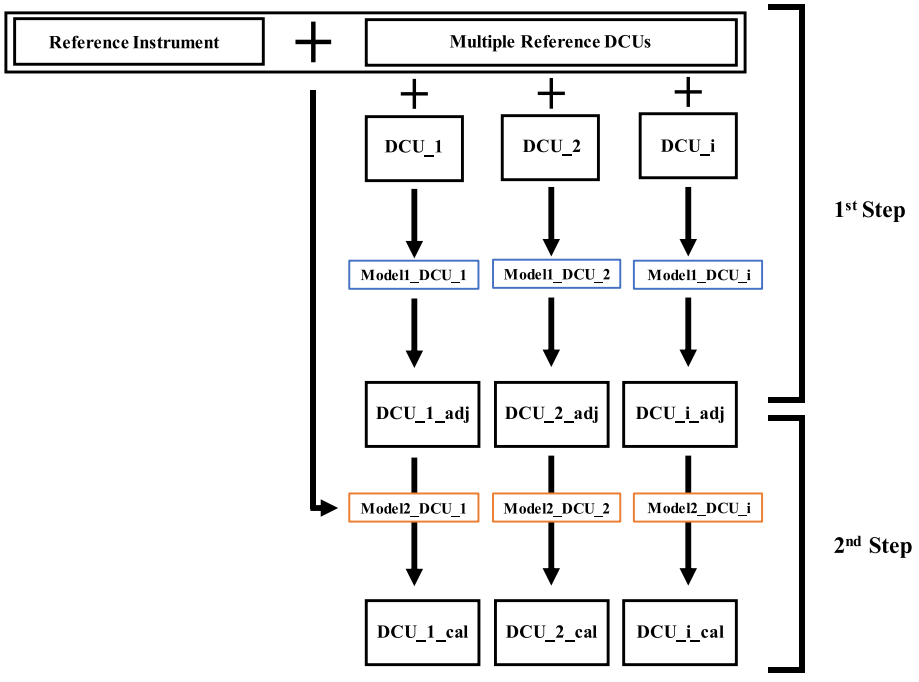

Fig. 7. New multi-step calibration strategy suggested.

DCU previously calibrated with the reference-grade instrument, resulting in the calibrated data. When performed cyclically and continuously, this strategy can be advantageous as it may reduce the number of reference-grade instruments needed, while probably minimising data drifts over time, thus bringing more reliability and robustness to the data collected with low-cost sensors.

More experimental campaigns using the same and other low-cost sensors are needed, which would lead to more data collection and further improvement of the calibration models. In fact, the use of other low-cost sensors (more suited for the study) is highly encouraged. Further studies should also evaluate the use of the proposed multi-step calibration strategy.

CRediT author statement

Juliana Sá: Methodology, Formal Analysis, Visualization, Writing-Original draft preparation**. Hiten Chojer**: Data curation, Formal Analysis, Software. **Pedro Branco**: Visualization, Investigation. **Maria Alvim-Ferraz**: Supervision, Writing- Reviewing and Editing. **Fernando Martins**: Writing- Reviewing and Editing, Software. **Sofia Sousa**: Conceptualization, Methodology, Supervision, Writing- Reviewing and Editing.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

Authors are grateful to the Municipality of Porto for all the support in deploying, operating and maintaining the DCUs, and to Ana Aguiar and her team at *Instituto de Telecomunicações* for the management of the data platform. This work was financially supported by: LA/P/0045/2020 (ALiCE) and UIDB/00511/2020 - UIDP/00511/2020 (LEPABE) funded by national funds through FCT/MCTES (PIDDAC); project 2SMART - engineered Smart materials for Smart citizens, with reference NORTE-01-0145-FEDER-000054, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2022.116910.

References

Afzal, A., Cioffi, N., Sabbatini, L., et al., 2012. NOx sensors based on semiconducting metal oxide nanostructures: progress and perspectives. Sensor. Actuator. B Chem. 171, 25–42. https://doi.org/10.1016/j.snb.2012.05.026.

Aleixandre, M., Gerboles, M., 2012. Review of small commercial sensors for indicative monitoring of ambient gas. Chem. Eng. Trans. 30, 169–174. https://doi.org/10.3303/CET1230029.

Alvim-Ferraz, M., Mesquita, M., Isabel Ferreira, M., et al., 2010. Influence of land-sea breezes on nocturnal ozone maxima observed in urban sites. Int. J. Environ. Waste Manag. 6, 293–308. https://doi.org/10.1504/ijewm.2010.035064.

Bhanarkar, M.K., Godase, S.M., Korake, P.M., et al., 2016. Review on WSN based outdoor air pollution monitoring system. Int. J. Ser. Eng. Sci. 1–13. https://doi.org/10.1000/IJSES.V0I0.81.

Borrego, C., Martins, H., Amorim, J., et al., 2014. Air quality, climate change and resilience in the Porto urban area. WIT Trans. Ecol. Environ. 183, 3–13. https://doi.org/10.2495/AIR140011.

Borrego, C., Costa, A.M., Ginja, J., et al., 2016. Assessment of air quality microsensors versus reference methods: the EuNetAir joint exercise. Atmos. Environ. 147, 246–263. https://doi.org/10.1016/j.atmosenv.2016.09.050.

Carslaw, D.C., Ropkins, K., 2012. Openair — an R package for air quality data analysis. Environ. Model. Software 27–28, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008.

Castell, N., Viana, M., Minguillón, M.C., et al., 2013. Real-world Application of New Sensor Technologies for Air Quality Monitoring, vol. 16. ETC/ACM Technical Paper., 2014.

Castell, N., Dauge, F.R., Schneider, P., et al., 2016. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environ. Int. https://doi.org/10.1016/j.envint.2016.12.007.

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

Cordero, J.M., Borge, R., Narros, A., 2018. Using statistical methods to carry out in field calibrations of low cost air quality sensors. Sensor. Actuator. B Chem. 267, 245–254. https://doi.org/10.1016/j.snb.2018.04.021.

EEA, 2020. Air Quality in Europe - 2020 Report. European Environment Agency, Copenhagen. https://doi.org/10.2800/786656. EEA Report No 09/2020. ISSN 1977-8449.

Fishbain, B., Lerner, U., Castell, N., et al., 2017. An evaluation tool kit of air quality micro-sensing units. Sci. Total Environ. 575, 639–648. https://doi.org/10.1016/j.scitotenv.2016.09.061.

Fritsch, S., Guenther, F., Wright, M.N., et al., 2019. Package 'neuralnet' - training of neural network. URL: https://github.com/bips-hb/neuralnet.

Gardner, M.W., Dorling, S.R., 2000. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmos. Environ. 34, 21–34. https://doi.org/10.1016/S1352-2310(99)00359-3.

Gozzi, F., Della Ventura, G., Marcelli, A., 2016. Mobile monitoring of particulate matter: state of art and perspectives. Atmos. Pollut. Res. 7, 228–234. https://doi.org/10.1016/j.apr.2015.09.007.

Guerreiro, C.B.B., Foltescu, V., de Leeuw, F., 2014. Air quality status and trends in Europe. Atmos. Environ. 98, 376–384. https://doi.org/10.1016/j.atmosenv.2014.09.017.

Heimann, I., Bright, V.B., McLeod, M.W., et al., 2015. Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. Atmos. Environ. 113, 10–19. https://doi.org/10.1016/j.atmosenv.2015.04.057.

Humphrey, G.B., Maier, H.R., Wu, W., et al., 2017. Improved validation framework and R-package for artificial neural network models. Environ. Model. Software 92, 82–106. https://doi.org/10.1016/j.envsoft.2017.01.023.

Jovašević-Stojanović, M., Bartonova, A., Topalović, D., et al., 2015. On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter. Environ. Pollut. 206, 696–704. https://doi.org/10.1016/j.envpol.2015.08.035.

Kim, J., Shusterman, A.A., Lieschke, K.J., et al., 2018. The Berkeley Atmospheric CO2 Observation Network: field calibration and evaluation of low-cost air quality sensors. Atmos. Meas. Tech. 11, 1937–1946. https://doi.org/10.5194/amt-11-1937-2018.

Kim, H., Müller, M., Henne, S., et al., 2022. Long-term behavior and stability of calibration models for NO and NO2 low-cost sensors. Atmos. Meas. Tech. 15, 2979–2992. https://doi.org/10.5194/amt-15-2979-2022.

Küçükaçıl Artun, G., Polat, N., Yay, O.D., et al., 2017. An integrative approach for determination of air pollution and its health effects in a coal fired power plant area by passive sampling. Atmos. Environ. 150, 331–345. https://doi.org/10.1016/j.atmosenv.2016.11.025.

Kumar, P., Morawska, L., Martani, C., et al., 2015. The rise of low-cost sensing for managing air pollution in cities. Environ. Int. 75, 199–205. https://doi.org/10.1016/j.envint.2014.11.019.

Luis, Y., Santos, P.M., Lourenco, T., et al., 2016. UrbanSense: an Urban-Scale Sensing Platform for the Internet of Things. IEEE International Smart Cities Conference (ISC2), Trento, p. 6. https://doi.org/10.1109/ISC2.2016.7580869. Italy 1.

Mba, L., Meukam, P., Kemajou, A., 2016. Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region. Energy Build. 121, 32–42. https://doi.org/10.1016/j.enbuild.2016.03.046.

McKercher, G.R., Salmond, J.A., Vanos, J.K., 2017. Characteristics and applications of small, portable gaseous air pollution monitors. Environ. Pollut. 223, 102–110. https://doi.org/10.1016/j.envpol.2016.12.045.

Mead, M.I., Popoola, O.A.M., Stewart, G.B., et al., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. 70, 186–203. https://doi.org/10.1016/j.atmosenv.2012.11.060.

Meier, F., Fenner, D., Grassmann, T., et al., 2017. Crowdsourcing air temperature from citizen weather stations for urban climate research. Urban Clim. 19, 170–191. https://doi.org/10.1016/j.uclim.2017.01.006.

Miskell, G., Salmond, J., Williams, D.E., 2017. Low-cost sensors and crowd-sourced data: observations of siting impacts on a network of air-quality instruments. Sci. Total Environ. 575, 1119–1129. https://doi.org/10.1016/j.scitotenv.2016.09.177.

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

Miskell, G., Pattinson, W., Weissert, L., et al., 2019. Forecasting short-term peak concentrations from a network of air quality instruments measuring PM2.5 using boosted gradient machine models. J. Environ. Manag. 242, 56–64. https://doi.org/10.1016/j.jenvman.2019.04.010.

Monteiro, A., 1993. O clima urbano do Porto: contribuição para a definição das estratégias de planeamento e ordenamento do território. Thesis. Porto, Portugal.
Morawska, L., Thai, P.K., Liu, X., et al., 2018. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone? Environ. Int. 116, 286–299. https://doi.org/10.1016/j.envint.2018.04.018.
Nagelkerke, N.J., 1991. A note on a general definition of the coefficient of determination. Biometrika 78, 691–692.
Decreto-Lei nº 102, 2010. Legislation from Ministério do Ambiente e do Ordenamento de Território. Diário da República - I Série, 186, pp. 4189–4204.
Pannullo, F., Lee, D., Waclawski, E., et al., 2016. How robust are the estimated effects of air pollution on health? Accounting for model uncertainty using Bayesian model averaging. Spat. Spatio-temp. Epidemiol. 18, 53–62. https://doi.org/10.1016/j.sste.2016.04.001.
Park, S., Kim, M., Kim, M., et al., 2018. Predicting PM10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN). J. Hazard Mater. 341, 75–82. https://doi.org/10.1016/j.jhazmat.2017.07.050.
Penza, M., Suriano, D., Villani, M.G., et al., 2014. Towards air quality indices in smart cities by calibrated low-cost sensors applied to networks. Journal 2012–2017. https://doi.org/10.1109/ICSENS.2014.6985429, 2014-December.
Piedrahita, R., Xiang, Y., Masson, N., et al., 2014. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. Atmos. Meas. Tech. 7, 3325–3336. https://doi.org/10.5194/amt-7-3325-2014.
R Core Team, 2018. R: a language and environment for statistical computing, 2018 R Found. Statis. Comput. Vienna, Austria. URL: http://www.Rproject.org/. (Accessed 20 May 2018) Accessed on.
Rai, A.C., Kumar, P., Pilla, F., et al., 2017. End-user perspective of low-cost sensors for outdoor air pollution monitoring. Sci. Total Environ. 607, 691–705. https://doi.org/10.1016/j.scitotenv.2017.06.266, 608.
Ródenas García, M., Spinazzé, A., Branco, P.T.B.S., et al., 2022. Review of low-cost sensors for indoor air quality: features and applications. Appl. Spectrosc. Rev. 1–33. https://doi.org/10.1080/05704928.2022.2085734.
Rodríguez, J.D., Pérez, A., Lozano, J.A., 2010. Sensitivity analysis of kappa-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32, 569–575. https://doi.org/10.1109/tpami.2009.187.
Rosario, L., Pietro, M., Francesco, S.P., 2016. Comparative analyses of urban air quality monitoring systems: passive sampling and continuous monitoring stations. Energy Proc. 101, 321–328. https://doi.org/10.1016/j.egypro.2016.11.041.
Russo, A., Trigo, R.M., Martins, H., et al., 2014. NO2, PM10 and O3 urban concentrations and its association with circulation weather types in Portugal. Atmos. Environ. 89, 768–785. https://doi.org/10.1016/j.atmosenv.2014.02.010.
Santos, F.D., Forbes, K., Moita, R., 2002. Climate Change in Portugal: Scenarios, Impacts and Adaptation Measures: SIAM Project. SIAM Project, first ed. Gradiva. Depósito legal, pp. 181–782. 2002.
Santos, P.M., Rodrigues, J.G., Cruz, S.B., et al., 2018. PortoLivingLab: an IoT-based sensing platform for smart cities. IEEE Internet Things J. 5, 523–532. https://doi.org/10.1109/JIOT.2018.2791522.
Schneider, P., Castell, N., Vogt, M., et al., 2017. Mapping Urban Air Quality in Near Real-Time Using Observations from Low-Cost Sensors and Model Information. Environment International. https://doi.org/10.1016/j.envint.2017.05.005.
Snyder, E.G., Watkins, T.H., Solomon, P.A., et al., 2013. The changing paradigm of air pollution monitoring. Environ. Sci. Technol. 47, 11369–11377. https://doi.org/10.1021/es4022602.
Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., et al., 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environ. Model. Software 22, 97–103. https://doi.org/10.1016/j.envsoft.2005.12.002.
Sousa, S.I.V., Martins, F.G., Pereira, M.C., et al., 2008. Influence of atmospheric ozone, PM10 and meteorological factors on the concentration of airborne pollen and fungal spores. Atmos. Environ. 42, 7452–7464. https://doi.org/10.1016/j.atmosenv.2008.06.004.
Sousa, S.I.V., Alvim-Ferraz, M.C.M., Martins, F.G., 2011. Identification and origin of nocturnal ozone maxima at urban and rural areas of Northern Portugal – influence of horizontal transport. Atmos. Environ. 45, 942–956. https://doi.org/10.1016/j.atmosenv.2010.11.008.
Spinelle, L., Gerboles, M., Villani, M.G., et al., 2015. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: ozone and nitrogen dioxide. Sensor. Actuator. B Chem. 215, 249–257. https://doi.org/10.1016/j.snb.2015.03.031.
Spinelle, L., Gerboles, M., Villani, M.G., et al., 2017. Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2. Sensor. Actuator. B Chem. 238, 706–715. https://doi.org/10.1016/j.snb.2016.07.036.
Steeneveld, G.-J., Klompmaker, J.O., Groen, R.J.A., et al., 2016. An Urban Climate Assessment and Management Tool for Combined Heat and Air Quality Judgements at Neighbourhood Scales. Resources, Conservation and Recycling. https://doi.org/10.1016/j.resconrec.2016.12.002.
Sun, L., Wong, K.C., Wei, P., et al., 2016. Development and application of a next generation air sensor network for the Hong Kong marathon 2015 air quality monitoring. Sensors 16. https://doi.org/10.3390/s16020211.
Tancev, G., 2021. Relevance of drift components and unit-to-unit variability in the predictive maintenance of low-cost electrochemical sensor systems in air quality monitoring. Sensors 21, 3298. https://doi.org/10.3390/s21093298.
Thompson, J.E., 2016. Crowd-sourced air quality studies: a review of the literature & portable sensors. Trends Environ. Anal. Chem. 11, 23–34. https://doi.org/10.1016/j.teac.2016.06.001.
Tranmer, M., Elliot, M., 2008. Multiple linear regression. Cathie Marsh Cent. Census Surv. Res. (CCSR) 5, 30–35.
Trigo, R.M., Osborn, T.J., Corte-Real, J.M., 2002. The North Atlantic Oscillation influence on Europe: climate impacts and associated physical mechanisms. Clim. Res. 20, 9–17. https://doi.org/10.3354/cr020009.
Vardoulakis, S., Fisher, B.E.A., Pericleous, K., et al., 2003. Modelling air quality in street canyons: a review. Atmos. Environ. 37, 155–182. https://doi.org/10.1016/S1352-2310(02)00857-9.
Velasco, A., Ferrero, R., Gandino, F., et al., 2016. A mobile and low-cost system for environmental monitoring: a case study. Sensors 16. https://doi.org/10.3390/s16050710.
Viana, M., Rivas, I., Reche, C., et al., 2015. Field comparison of portable and stationary instruments for outdoor urban air exposure assessments. Atmos. Environ. 123, 220–228. https://doi.org/10.1016/j.atmosenv.2015.10.076.
Weissert, L.F., Salmond, J.A., Miskell, G., et al., 2017. Use of a dense monitoring network of low-cost instruments to observe local changes in the diurnal ozone cycles as marine air passes over a geographically isolated urban centre. Sci. Total Environ. 575, 67–78. https://doi.org/10.1016/j.scitotenv.2016.09.229.
Williams, D.E., 2020. Electrochemical sensors for environmental gas analysis. Curr. Opin. Electrochem. 22, 145–153. https://doi.org/10.1016/j.coelec.2020.06.006.
WMO, 2021. An Update on Low-Cost Sensors for the Measurement of Atmospheric Composition. World Meteorological Organisation, Geneva, Switzerland. WMO-No. 1215. https://library.wmo.int/doc_num.php?explnum_id=10620.
Zhang, L., Tian, F.-C., Peng, X.-W., et al., 2014. A rapid discreteness correction scheme for reproducibility enhancement among a batch of MOS gas sensors. Sens. Actuators A: Phys. 205, 170–176. https://doi.org/10.1016/j.sna.2013.11.015.
Zhou, X., Lee, S., Xu, Z., et al., 2015. Recent progress on the development of chemosensors for gases. Chem. Rev. 115, 7944–8000. https://doi.org/10.1021/cr500567r.