Atmos. Meas. Tech., 16, 4723–4740, 2023
https://doi.org/10.5194/amt-16-4723-2023

Atmospheric Measurement Techniques
Open Access
EGU

Development of low-cost air quality stations for next-generation monitoring networks: calibration and validation of NO_2 and O_3 sensors

Alice Cavaliere[1], Lorenzo Brilli[1], Bianca Patrizia Andreini[2,★], Federico Carotenuto[1], Beniamino Gioli[1], Tommaso Giordano[1], Marco Stefanelli[2,★], Carolina Vagnoli[1], Alessandro Zaldei[1], and Giovanni Gualtieri[1]

[1]National Research Council – Institute of BioEconomy (CNR–IBE), Via Caproni 8, 50145 Florence, Italy
[2]ARPAT, Tuscany Region Environmental Protection Agency, Via Porpora, 22, 50144 Florence, Italy
★These authors contributed equally to this work.

Correspondence: Alice Cavaliere (alice.cavaliere@ibe.cnr.it)

Received: 14 April 2023 – Discussion started: 2 May 2023
Revised: 12 August 2023 – Accepted: 18 September 2023 – Published: 20 October 2023

Abstract. A pre-deployment calibration and a field validation of two low-cost (LC) stations equipped with O_3 and NO_2 metal oxide sensors were addressed. Pre-deployment calibration was performed after developing and implementing a comprehensive calibration framework including several supervised learning models, such as univariate linear and non-linear algorithms, and multiple linear and non-linear algorithms. Univariate linear models included linear and robust regression, while univariate non-linear models included a support vector machine, random forest, and gradient boosting. Multiple models consisted of both parametric and non-parametric algorithms. Internal temperature, relative humidity, and gaseous interference compounds proved to be the most suitable predictors for multiple models, as they helped effectively mitigate the impact of environmental conditions and pollutant cross-sensitivity on sensor accuracy. A feature analysis, implementing dominance analysis, feature permutations, and the SHapley Additive exPlanations method, was also performed to provide further insight into the role played by each individual predictor and its impact on sensor performances. This study demonstrated that while multiple random forest (MRF) returned a higher accuracy than multiple linear regression (MLR), it did not accurately represent physical models beyond the pre-deployment calibration dataset, so a linear approach may overall be a more suitable solution. Furthermore, as well as being less computationally demanding and generally more suitable for non-experts, parametric models such as MLR have a defined equation that also includes a few parameters, which allows easy adjustments for possible changes over time. Thus, drift correction or periodic automatable recalibration operations can be easily scheduled, which is particularly relevant for NO_2 and O_3 metal oxide sensors. As demonstrated in this study, they performed well with the same linear model form but required unique parameter values due to intersensor variability.

1 Introduction

Low-cost (LC) air quality sensors are gaining more and more interest as they can provide near-real-time observations, with high spatial and temporal resolution. Their observations can be integrated into the current official regulatory networks, usually monitoring air quality at lower space and time resolution, and thus providing useful information to support policymakers and stakeholders in understanding air pollution dynamics (Brilli et al., 2021; Morawska et al., 2018). Dramatic advances in LC sensor technology have been made since their very first applications for monitoring CO, NO_2, and NO_x (De Vito et al., 2009), O_3 (Williams et al., 2013), and particulate matter (Holstius et al., 2014). Among gaseous species, NO_2 and O_3 are the most commonly investigated, since both short- and long-term exposure to these pollutants are associated with a higher risk to human health (Lin et al., 2018; Nuvolone et al., 2018; Meng et al., 2021; World Health Organization, 2021).

Published by Copernicus Publications on behalf of the European Geosciences Union.

Typically, LC NO_2 and O_3 monitors use electrochemical (EC) or metal oxide sensors (MOSs) (Narayana et al., 2022; Concas et al., 2021; Idrees and Zheng, 2020), which produce an analogue signal proportional to pollutant concentration.

In their simplest configuration, EC sensors are based on a redox reaction within an electrochemical cell in which the target analyte oxidizes the anode or the cathode (Gäbel et al., 2022). As for MOSs, they have an exposed metal oxide surface film that changes its electrical properties when exposed to the target gas (Masson et al., 2015; Fine et al., 2010).

MOSs have a longer lifetime, can operate at higher temperatures, and have a shorter response time and a wider operating range than EC sensors. In contrast, EC sensors have a lower power consumption, as they do not require powering an electric heater and are less impacted by high humidity levels (Narayana et al., 2022; Concas et al., 2021).

Overall, choosing between MOS and EC sensors depends on the goals of the deployment. EC sensors should be preferred in areas with steady temperatures and weather conditions (Concas et al., 2021), while MOSs are more suited for long-term monitoring (Concas et al., 2021; Narayana et al., 2022; Burgués and Marco, 2018). LC sensors are affected by environmental factors, such as air temperature and relative humidity (Barcelo-Ordinas et al., 2019; Mueller et al., 2017; Mead et al., 2013), and suffer from cross-sensitivity with other air pollutants (Rai et al., 2017; Bart et al., 2014), thus complicating robust measurement recovery. These issues depend on sensor characteristics such as the type of electrolyte, electrode, or semiconductor material used (Spinelle et al., 2015). Unfortunately, the lack of information or inconsistency in data sheets from sensor manufacturers makes it challenging to accurately interpret the readings (Narayana et al., 2022). As a result, these issues must be addressed in the calibration process to ensure the accuracy and reliability of LC field measurements.

Two main approaches to calibrating LC sensors exist (Spinelle et al., 2013), namely pre-deployment calibration and field calibration.

Pre-deployment calibration is typically performed in a controlled environment, where LC sensors are exposed to a gas of known concentration in order to properly tune a calibration model (e.g. Claveau et al., 2022; Wei et al., 2018). Field calibration, on the other hand, consists of co-locating LC sensors near reference (official) stations that provide measured concentrations so as to develop a calibration model in real-world conditions (e.g. Spinelle et al., 2015). However, this approach may lead to potential inaccuracies when the calibrated LC sensors are deployed on locations with varying air compositions and weather conditions (e.g. Spinelle et al., 2017; Aleixandre et al., 2013).

Both pre-deployment and field calibration models are developed using a variety of mathematical methods ranging from simple univariate regression models to more advanced machine learning techniques (Aula et al., 2022). The latter include various supervised learning techniques such as artificial neural networks (ANNs), random forest (RF), and support vector regression (SVR; e.g. Karagulian, 2023; Karagulian et al., 2019; Cordero et al., 2018). In addition, the use of covariates such as temperature, relative humidity, and interfering gases such as NO_2, NO, and O_3 can increase accuracy in the calibration process (Concas et al., 2021; Peterson et al., 2017; Piedrahita et al., 2014). To date, while the accuracy of LC calibration algorithms has been widely investigated, there is a lack of studies addressing crucial issues associated with these techniques, such as the (i) transferability of field calibration beyond the training range (as highlighted in Nowack et al., 2021; Zauli-Sajani et al., 2021; De Vito et al., 2020; Esposito et al., 2018); (ii) pre-deployment calibration complemented by a later field validation for EC and MOSs (as mentioned in Maag et al., 2018); and (iii) the weight or importance of each feature included in multiple calibration models, particularly for black box techniques that cannot rely on statistical inference techniques (as mentioned in Sahu et al., 2021).

This study aims at addressing these issues by (i) implementing a pre-deployment calibration procedure for two LC stations measuring NO_2 and O_3 concentrations; (ii) identifying the optimal calibration that results in the highest accuracy; (iii) performing a long-term (more than 1 year) field validation against a regulatory station located in a different site; and (iv) critically discussing the transferability and scalability of the selected calibration model for multiple devices. These goals have been pursued by using 10 models among parametric, non-parametric univariate, and multiple algorithms. Additionally, the investigation focused on delving deeper into the influence of internal temperature on LC sensors. To ensure comprehensive analysis, the covariate set for the multiple models was expanded to incorporate other essential factors, such as humidity and gaseous interference compounds. Furthermore, the study utilized model-agnostic techniques, including SHapley Additive exPlanations (SHAP; Lundberg and Lee, 2017), to assess the model's generalization ability in a field environment. While SHAP has been employed in previous pollution-related studies (e.g. Wang et al., 2023; Chakraborty et al., 2022; Vega García and Aznarte, 2020), this research provides an original contribution by applying SHAP specifically to MOSs. This application aims to provide both local and global interpretations, resulting in a deeper understanding of the sensor's behaviour on individual data points and gaining insights into its overall performance.

2 Materials and methods

2.1 AirQino low-cost stations

The study focuses on two AirQino LC air quality monitoring stations (hereinafter AQ) developed by the National Research Council – Institute of BioEconomy (CNR–IBE) in

Florence (Italy), namely AQ1 and AQ2, which are equipped with MOSs to measure O_3 and NO_2 concentrations (Zaldei et al., 2017; Di Lonardo et al., 2014). AQ consists of an Arduino-Shield-compatible electronic board that integrates LC and high temporal resolution sensors (2–3 min data acquisition frequency) to monitor environmental parameters and atmospheric pollutants such as relative humidity, internal and external temperature, CO, CO_2, O_3, NO_2, volatile organic compounds (VOCs), $PM_{2.5}$, and PM_{10}. As for the atmospheric pollutants examined in this study (NO_2 and O_3), their concentrations are collected by SGX Sensortec MOSs, with the models MiCS-2714 for NO_2 (Sensortech, a) and MiCS-2614 for O_3 (Sensortech, b). These sensors consist of a micro-metal-oxide semiconductor diaphragm, with an integrated heating resistor (temperature ranges from 350 to 550 °C). The resistor-produced heat catalyses the reaction, which in turn affects the electrical resistance of the oxide layer itself. After the initial pre-heating period, the sensor detects gas changes in time intervals below 2 s. The output signal from the sensor is passed through an analogue-to-digital converter (ADC) circuit with a 10 bit output. The ADC converts the analogue signal to a digital value between 0 and 1023 counts. This signal in counts is the primary output provided by the sensors (raw data). External air temperature (extT) and relative humidity (RH) are measured by an AM2305 (Asair, 2021) sensor protruding from the device enclosure. The internal temperature (intT) of the enclosure is monitored by a DS18B20 sensor (Maxim Integrated, 2021) that is mounted directly on the electronic board. Sensor readings are collected by the onboard microprocessor and sent to a PostgreSQL database via a general packet radio service (GPRS) connection.

2.2 Reference instruments

During pre-deployment calibration, reference pollutant concentrations were measured using two HORIBA instruments (HORIBA, Ltd.; Ambient Air Pollution AP Series analysers). The HORIBA APNA-370 model is an ambient nitrogen oxide monitor based on the chemiluminescence principle, allowing a continuous measurement of NO and NO_2 concentrations. The HORIBA APOA-370 model was used to collect O_3 concentrations, based on a cross-flow-modulated ultraviolet absorption method (Fig. 1).

2.3 Sensor calibration

As detailed in Table S1 in the Supplement, the pre-deployment calibration of the AQ1 and AQ2 stations against HORIBA analysers was performed at CNR–IBE headquarters in Florence, Italy (43°47′52″ N, 11°11′ E; Fig. 1). The AQ stations were mounted on a dedicated outdoor rack, while the HORIBA instruments were placed indoors in a laboratory setting. For outdoor air pollution sampling, approximately 2 m long sampling probes were employed to collect outside air and channel it directly to each of the reference instruments. HORIBA returned measurements at 3 min resolution collected across a 73 d period (19 July 2017–30 September 2017). To ensure data validity, measurements associated with RH > 99 %, following Wang et al. (2010), or classified as outliers by an interquartile range (IQR) method (Dekking et al., 2005) were removed from the dataset, eventually resulting in 58 949 valid records for NO_2 and 59 261 valid records for O_3 concentrations. The workflow of the pre-deployment calibration process is shown in Fig. 2.

Prior to implementing the calibration techniques, an exploratory data analysis (EDA) was conducted using the correlation matrix to identify important insights. The study further explored the potential generalizable aspect of the relationship between O_3 MOSs and temperature, as highlighted in Spinelle et al. (2016), leveraging observations from the correlation matrix. The core of the calibration framework consisted of a set of supervised learning algorithms, previously evaluated in the literature, that fall into the following two categories: univariate and multiple models. The former are based on a single predictor (pollutant raw data), while the latter include additional predictors. Both categories included linear and non-linear algorithms. During the training phase, the datasets containing both LC and reference measurements were divided into a training subset consisting of 67 % of the data and a testing subset consisting of the remaining 33 %.

The suite of algorithms for univariate calibration linear methods included linear regression (Mijling et al., 2018; Maag et al., 2016) and robust regressions (Cavaliere et al., 2018), while the non-linear approaches comprised support vector machine (Bigi et al., 2018; Gu et al., 2018), random forest (Han et al., 2021; Zimmerman et al., 2018), and gradient boosting (Lin et al., 2018; Johnson et al., 2018). Multiple models, which considered temperature, humidity, and cross-sensitivity parameters for prediction, consisted of both parametric models and non-parametric models (Gäbel et al., 2022; Sayahi et al., 2020; Spinelle et al., 2017).

2.3.1 Univariate models

The suite of univariate algorithms included a total of 10 models that fall into the following three main categories: (i) simple linear regression (SLR), (ii) non-linear regression (SNLR), and (iii) support vector machine (SVM). Five regression models are included in SLR, namely linear regression (LR); polynomial regression of the second (PLR2) and third (PLR3) degree; Huber regression (HBLR), which is a robust regression technique to outliers that uses a different loss function rather than the traditional least squares; and Cook's distance regression (CDLR; Cook, 1977), which summarizes how much all values in the regression model change when the ith observation is removed. The non-linear regression (SNLR) included parametric and non-parametric models. The former included power non-linear regression (PNLR) and logarithm regression (LNLR), which con-

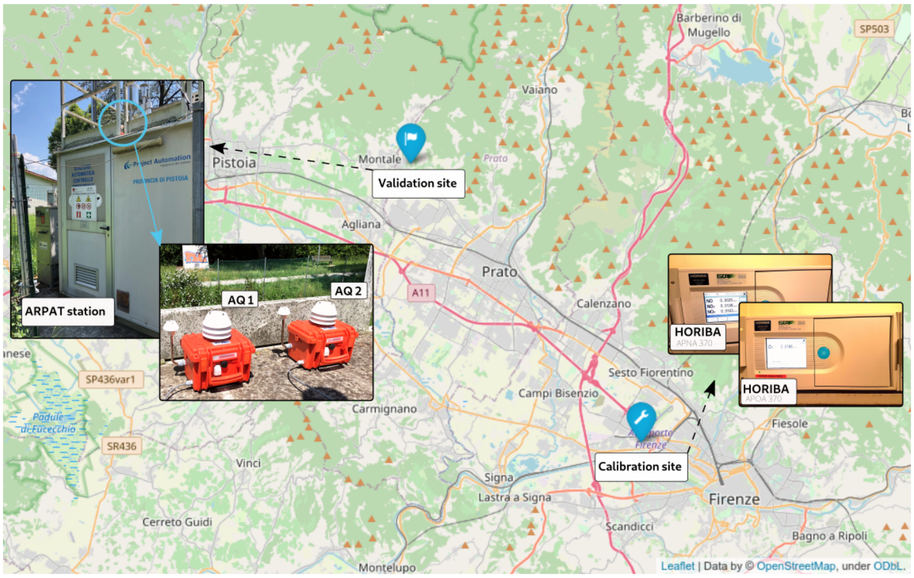

Figure 1. Map highlighting the calibration and validation locations for AQ1 and AQ2 LC air quality monitoring stations in Tuscany, Italy. At the calibration site (Florence), the HORIBA instruments used for calibrating the LC stations are shown, while at the validation site (Montale), the LC stations are pictured as installed on the roof of the reference ARPAT station (air quality station EoI code IT1553A).

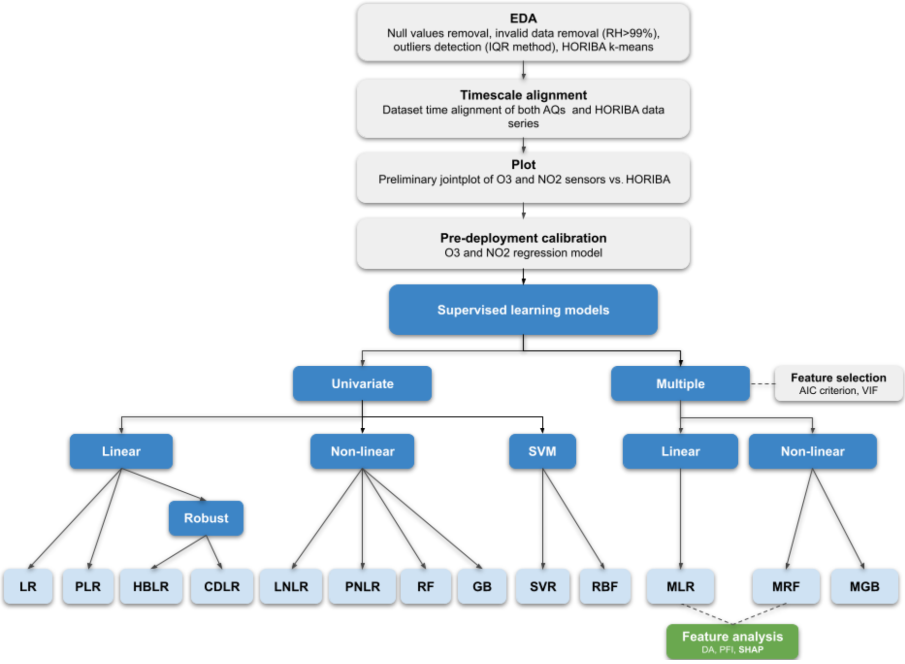
EDA
Null values removal, invalid data removal (RH>99%), outliers detection (IQR method), HORIBA k-means
Timescale alignment
Dataset time alignment of both AQs and HORIBA data series
Plot
Preliminary jointplot of O3 and NO2 sensors vs. HORIBA
Pre-deployment calibration
O3 and NO2 regression model
Supervised learning models
Univariate
Multiple
Feature selection
AIC criterion, VIF
Linear
Non-linear
SVM
Linear
Non-linear
Robust
LR
PLR
HBLR
CDLR
LNLR
PNLR
RF
GB
SVR
RBF
MLR
MRF
MGB
Feature analysis
DA, PFI, SHAP

Figure 2. Workflow of the pre-deployment calibration process performed in the work. The model abbreviations are also listed in the Appendix.

sider the estimation of coefficients through the Levenberg–Marquardt algorithm. The latter included random forest (RF) and gradient boosting (GB). RF conducts the optimal splitting of data samples into smaller sample sets, which then are fitted, respectively, along the tree paths, while GB built an additive model based on gradient boosting decision trees and in each stage in which a regression tree was fit. In the present calibration, the RF model used the mean square error as a fitting function in order to evaluate each decision split. Finally, SVM included a support vector regression using linear kernel (SVR) and radial basis function (RBF). In SVM, the kernel allows us to identify a hyperplane with maximum margin such that the maximum number of data points is within that margin. For each non-parametric model, the grid search method was used to optimize the default hyper parameter values (Pedregosa et al., 2011; Smets et al., 2007).

2.3.2 Multiple models

Multiple models included both linear (MLR) and non-linear models, with the latter consisting of multiple random forest (MRF) and multiple gradient boosting (MGB). While implementing an MLR model, a linear stepwise multi-regression analysis was carried out by automatically generating all possible models, starting from a list of explanatory variables. In the case of NO_2 and O_3 sensors, the latter included internal temperature (intT), external temperature (extT), and relative humidity (RH). In order to solely include statistically significant variables, thus excluding possible collinearity between them, the variance inflation factors (VIFs) were examined for each generated model. To refine the choice between internal and external temperature, a multiple linear model was used that alternatively incorporated both temperatures, followed by a cross-validation. Once a subset of significant explanatory variables was identified during the multiple linear regression (MLR) implementation, the multiple random forest (MRF) and multiple gradient boosting (MGB) models were also applied. MGB was selected, as GB is the univariate model that improves the results obtained by the supervised machine learning model, while MRF was selected as being a model that is widely used in the literature (e.g. Bisignano et al., 2022; Bigi et al., 2018; Zimmerman et al., 2018). To compare the performance between models, specified metrics were evaluated, such as the adjusted R^2 (AdjR^2; Draper and Smith, 1998). In Table S2, a concise summary of the initialization hyperparameters applied to the models is provided.

2.3.3 Multiple-model interpretation

To gain a better understanding of the impact due to different predictors and an insightful interpretation of the multiple-model results, several analysis techniques have been applied, such as permutation feature importance (PFI; Breiman, 2001), dominance analysis (DA; Azen and Budescu, 2003), and the SHapley Additive exPlanation (SHAP; Lundberg and Lee, 2017) analysis.

PFI is a model inspection technique that measures the global variable importance by observing the effect of randomly shuffling each explanatory variable. DA is a common procedure for identifying the relative importance of predictors in a linear model. In this work, the following five different DA statistics were evaluated: (i) interactional dominance (IntD); (ii) individual dominance (ID); (iii) average partial dominance (APD); (iv) total dominance (TD); and (v) percentage relative importance (PRI).

SHAP analysis is a model-agnostic approach based on the game theory that can be applied to any machine learning model as a post hoc interpretation technique. According to the SHAP analysis, each machine learning model's prediction, $f(x)$, can be represented as the sum of its computed SHAP values, plus a fixed base value, as shown in Eq. (1):

$$f(x) = \Phi_0 + \sum_{i=1}^{p} \Phi_i, \tag{1}$$

where Φ_0 is the base value of the model, which represents the average prediction across all inputs, and Φ_i is the SHAP value for feature i for the input x. Each Φ_i is computed as in Eq. (2):

$$\Phi_i = \sum_{S \subseteq 1,2,\ldots,p \setminus i} \frac{(p-|S|-1)! \cdot |S|!}{p!} \cdot [f(x_{S\cup i}) - f(x_S)], \tag{2}$$

where p is the total number of features, S is a subset of all features except for feature i, $|S|$ is the number of features in subset S, $f(x_S)$ is the model's prediction for input x with features in subset S, and $f(x_{S\cup i})$ is the model's prediction for input x with features in subset S and feature i included.

SHAP values are calculated for each feature and value present in the dataset, and they approximate the contribution towards the output given by that data point. To compute SHAP values for different types of machine learning models, various SHAP implementations are available. In this study, the shap.LinearExplainer function was used for MLR predictors, while the FastTreeSHAP explainer (Yang, 2021) was used for other models. Compared to the widely used TreeSHAP algorithm, FastTreeSHAP provides faster computation of feature importance values for tree-based models.

2.4 Field validation

To test pre-deployment calibration models, the AQ stations were subject to a field validation based on hourly measurements collected during 429 consecutive days (19 June 2018–22 August 2019) by a reference air quality station operated by the Regional Agency of Tuscany for the Environmental Protection (ARPAT). High-resolution NO_2 and O_3 concentrations measured by the AQ stations over the same period were averaged hourly in order to be aligned to the reference data. Overall, datasets of valid hourly records, ranging 7383–9340 for NO_2 and 7344–9303 for O_3 concentrations, were used (Table S1). The reference air quality station (EoI code IT1553A) was located at Montale, a small town in Tuscany located between the cities of Prato and Pistoia (43°54′57″ N, 11°00′26″ E) and classified as a suburban background station (Fig. 1). The ARPAT reference station and the HORIBA APNA-370 analyser used the same method for measuring NO_2, while a different method (ultraviolet photometry) was used by ARPAT to measure O_3.

2.5 Statistics and libraries

The performances of each AQ station during both pre-deployment calibration and field validation were computed using various statistical measures, including the Pearson correlation coefficient (r); coefficient of determination (R^2); AdjR^2; root mean squared error (RMSE); normalized RMSE (nRMSE), which takes into account the range of values by dividing the RMSE by the difference between the maximum and minimum values; mean absolute error (MAE); and mean bias error (MBE). The variance impact factor (VIF) and Akaike information criterion (AIC) were also applied to discriminate between MLR models. All calculations related to calibration procedure and the analysis of the performance of the calibrated units are implemented using Python scikit-learn library (Pedregosa et al., 2011) and Python statsmodels module (Seabold and Perktold, 2010). Finally, a feature evaluation of MLR and MRF models was performed using the Python Dominance-Analysis library (Shekhar et al., 2019), SHAP library (Lundberg and Lee, 2022), FastTreeSHAP library (Yang, 2022), and ELI5 Permutation Importance library (TeamHG-Memex, 2022).

3 Results

3.1 Exploratory data analysis

After applying the humidity threshold and IQR procedure, 2 % and 12 % of the records were withdrawn from the initial datasets of the AQ1 and AQ2 stations, respectively. The comparison between the resulting O_3 and NO_2 data and the HORIBA reference concentrations is shown in Fig. 3. Based on the analysis of Pearson's correlation (Fig. S3), three patterns for both AQ stations emerged as conforming to the existing literature. HORIBA NO_2 and O_3 had a negative Pearson's r ($r_{\mathrm{AQ1}} = -0.77$; $r_{\mathrm{AQ2}} = -0.75$), which is compatible with the chemical coupling of O_3 and $NO_x = NO + NO_2$ (Han et al., 2011). AQ intT had a high positive correlation with HORIBA O_3 ($r_{\mathrm{AQ1}} = 0.79$; $r_{\mathrm{AQ2}} = 0.80$), which is compatible with the fact that high temperatures can increase the rate of O_3 formation through photochemical reactions (Han et al., 2011). AQ RH had a high negative correlation with HORIBA O_3 ($r_{\mathrm{AQ1}} = -0.75$; $r_{\mathrm{AQ2}} = -0.74$), which is compatible with the fact that high relative humidity is generally associated with lower O_3 levels (Camalier et al., 2007). Moreover, as a result of the convective heat transfer equation, a strong positive correlation was observed between intT and extT for each AQ ($r_{\mathrm{AQ1,AQ2}} = 1$). On average, the temperature difference between intT and extT remains relatively constant at around 8 °C. A visual representation of the difference between the two temperatures, plotted against their mean, can be found in the Bland–Altman plots in Fig. S4. No significant correlation was observed between NO_2 raw and either temperature or RH. Moderate positive associations were instead found between O_3 raw and both intT ($r_{\mathrm{AQ1}} = 0.55$; $r_{\mathrm{AQ2}} = 0.55$) and extT ($r_{\mathrm{AQ1}} = 0.52$; $r_{\mathrm{AQ2}} = 0.53$).

3.2 Univariate models

The results of the supervised linear (SLR), supervised non-linear (SNLR), and support vector machine (SVM) models applied for both AQ stations and pollutants are reported in Table 1. For both AQ stations, the best performances were found using the GB model, with O_3 concentrations generally fitting better than NO_2 concentrations.

3.3 Multiple regression

EDA suggested that the inclusion in multiple regression models of both intT and extT may result in unstable results due to their strong collinearity ($r_{\mathrm{AQ1,AQ2}} = 1$). This was confirmed by the variance inflation factor (VIF) for the MLR model, which was higher than 5 when both variables were used (Table S5). To ensure consistent selection of the optimal temperature variable in the model, a cross-validation procedure was conducted on the calibration dataset for the MLR model, alternately including intT and extT in the covariate set. The results of the five-split cross-validation (Table 2) showed no significant differences when using intT or extT, while the use of intT provided a slightly higher mean accuracy and a lower mean RMSE.

Following the previous result, the final subset of predictors used for all models consisted of intT, RH, and raw signal from both sensors (Tables S6 and S7). Accordingly, for both stations, Eqs. (3) and (4) were the best model formulas for O_3 and NO_2 sensors, respectively:

$$\begin{aligned} O_3 = {} & \beta_0 + \beta_1 \cdot NO_2\mathrm{raw} + \beta_2 \cdot O_3\mathrm{raw} \\ & + \beta_3 \cdot \mathrm{RH} + \beta_4 \cdot \mathrm{intT} \end{aligned} \tag{3}$$

$$\begin{aligned} NO_2 = {} & \beta_0 + \beta_1 \cdot NO_2\mathrm{raw} + \beta_2 \cdot O_3\mathrm{raw} \\ & + \beta_3 \cdot \mathrm{RH} + \beta_4 \cdot \mathrm{intT}. \end{aligned} \tag{4}$$

The calibration coefficients achieved for the MLR model are reported in Table 3, while the scores of the MLR, MGB, and MRF model application are reported in Table 4.

Overall, O_3 concentrations were better fitted than NO_2 concentrations, while MRF proved to be the finest model, generally outperforming the MGB and particularly the MLR model.

3.4 Multiple-model interpretation

In terms of traditional statistical inference techniques such as DA and PFI during MLR and MRF O_3 calibration, the results confirmed first O_3 raw and second intT as being the most significant predictors (Table 5), which were consistent with those reported by Masson et al. (2015). Overall, the DA analysis showed that O_3 raw and intT were the most important features for both stations and pollutants. In particular,

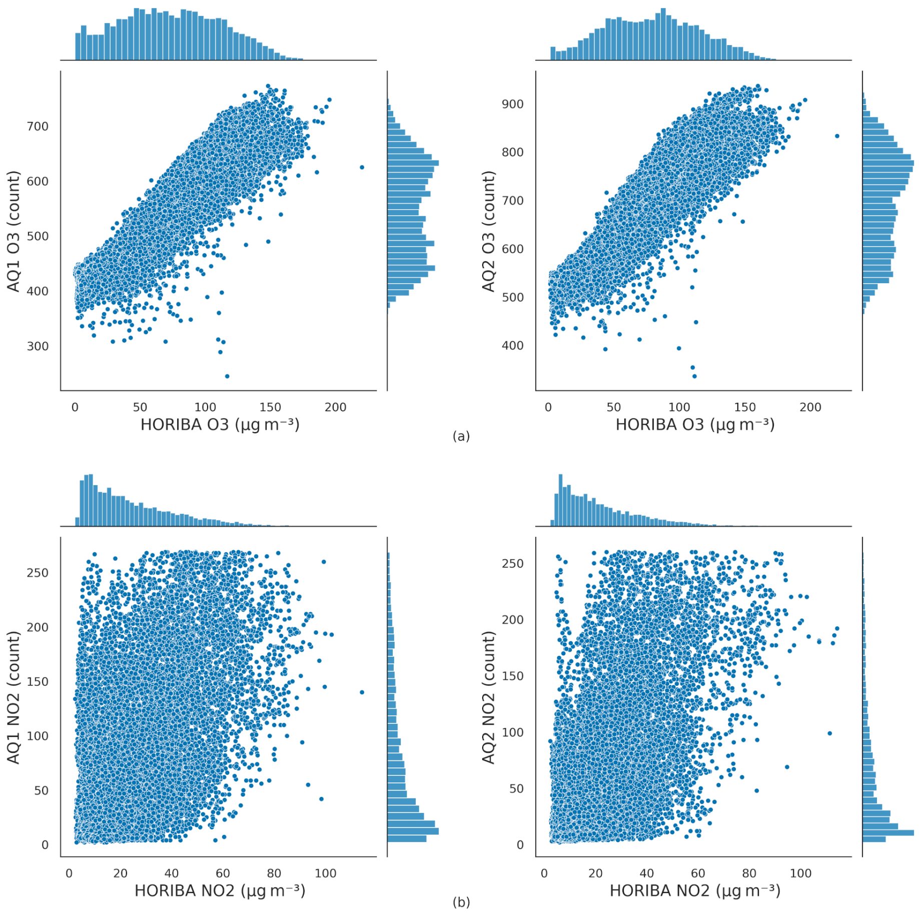

Figure 3. Scatterplots of 3 min sampled AQ1 and AQ2 signals vs. HORIBA reference concentrations observed during the pre-deployment calibration for O_3 **(a)** and NO_2 **(b)**.

for O_3 concentrations, O_3 raw data resulted in the highest PRI value, explaining 38.96 % and 34.95 % of the R^2 of the MLR model for AQ1 and AQ2, respectively, followed by intT (28.64 % and 31.51 %, respectively). Also for NO_2 concentrations, O_3 raw data had the highest PRI value, explaining 55.18 %–51.13 % of the R^2 of the MLR model, followed by NO_2 raw data (23.78 %–26.79 %). In the O_3 MRF regression, O_3 raw was the most important feature for AQ1, while it was intT for AQ2. Conversely, in the NO_2 MRF regression, O_3 raw was the most important feature for both AQ stations, followed by RH for AQ1 and by NO_2 for AQ2. Notably, for both MLR and MRF models, O_3 raw proved to be a more important feature in NO_2 calibration than in O_3 calibration.

The challenge with traditional feature selection methods like DA and PFI is that they may produce misleading results when features are highly correlated or the data are noisy. These methods in fact do not consider interactions or correlations between predictors, and DA is only applicable to linear models. To overcome these limitations, the study utilized SHAP analysis. The SHAP analysis was performed in order to gain insight into both the global and local contribution of each feature at both individual instance level and across the population, resulting in the SHAP bee swarm plots for MLR and MRF, as shown in Figs. 4 and 5, respectively. The bee swarm plot ranks the input features from the highest to the lowest mean absolute SHAP values for the entire dataset. For each variable, every instance of the dataset appears as its own point. The points are distributed horizontally along the x axis according to their SHAP value. In places where there is a high density of SHAP values, the points are stacked vertically. The colour bar corresponds to the raw values of each feature for each instance, providing a visual representation of the feature's contribution to the outcome prediction.

As for the MLR model for both AQ stations, high levels of both O_3 raw and intT data had a strong and positive impact on O_3 output, as indicated by high and positive SHAP values (Fig. 4a), while high levels of NO_2 raw data had a strong and positive impact on the NO_2 output (Fig. 4b). Herein, however, high levels of O_3 raw and intT data had a greater impact on decreasing the predicted values of NO_2 than the raw data of NO_2.

Also, for the MRF model, O_3 raw data had a high influence on O_3-predicted values (Fig. 5a) because higher values of O_3 raw data increased the O_3 prediction, while lower

Table 1. Statistics of the univariate regression models applied to the AQ1 and AQ2 stations. Note that for non-linear models (LNLR and PNLR), R^2 is not a useful metric, while it is useful for linear models that use polynomials to model the curvature in the data (Spiess and Neumeyer, 2010).

			SLR					SNLR				SVM	
Pollutant	AQ ID	Stat.	LR	PLR2	PLR3	CDLR	HBLR	LNLR	PNLR	RF	GB	SVR	RBF
O_3	AQ1	R^2	0.81	0.81	0.82	0.81	0.81	0.00	–	0.82	0.82	0.81	0.70
		RMSE	16.92	16.92	16.75	16.99	17.00	17.15	18.85	16.78	16.58	17.18	21.31
		MAE	13.42	13.41	13.40	13.19	13.18	13.47	14.94	13.42	13.27	13.12	16.06
		MBE	−0.28	−0.28	−0.30	1.56	1.69	−0.33	−0.96	−0.18	−0.20	2.89	3.53
	AQ2	R^2	0.77	0.77	0.77	0.76	0.77	–	–	0.77	0.78	0.76	0.60
		RMSE	17.58	17.55	17.41	17.86	17.75	17.97	18.46	17.58	17.36	18.11	23.06
		MAE	14.11	14.10	14.10	13.81	13.85	14.30	14.79	14.18	14.05	13.79	17.38
		MBE	0.14	0.14	0.15	3.03	2.39	0.19	-0.48	0.24	0.14	4.17	3.55
NO_2	AQ1	R^2	0.34	0.34	0.34	0.32	0.33	–	–	0.33	0.35	0.33	0.31
		RMSE	14.22	14.18	14.16	14.40	14.28	14.51	14.46	14.35	14.14	14.35	14.50
		MAE	10.91	10.84	10.82	10.77	10.79	11.20	11.22	10.84	10.75	10.76	10.81
		MBE	0.09	0.11	0.11	2.26	1.21	0.16	−0.11	0.06	0.09	1.73	2.08
	AQ2	R^2	0.38	0.38	0.38	0.35	0.37	–	–	0.36	0.38	0.36	0.32
		RMSE	12.86	12.85	12.85	13.12	12.95	13.45	13.06	13.04	12.85	13.02	13.39
		MAE	9.83	9.83	9.84	9.69	9.69	10.37	10.06	9.94	9.81	9.67	9.83
		MBE	−0.09	−0.09	−0.09	2.56	1.53	−0.05	−0.24	−0.05	−0.10	2.05	2.60

Table 2. R^2 and RMSE ($\mu g\,m^{-3}$) values by covariate set, including intT or extT variables of the cross-validation procedure applied to the MLR model.

			Covariate set (mean ± SD)	
AQ ID	Pollutant	Stat.	O_3, NO_2, intT, RH	O_3, NO_2, extT, RH
AQ1	O_3	R^2	0.93 ± 0.03	0.93 ± 0.02
		RMSE	9.52 ± 2.51	9.55 ± 2.10
AQ2	O_3	R^2	0.91 ± 0.04	0.91 ± 0.05
		RMSE	9.52 ± 2.86	9.72 ± 2.85
AQ1	NO_2	R^2	0.57 ± 0.21	0.56 ± 0.24
		RMSE	10.75 ± 1.31	10.83 ± 1.43
AQ2	NO_2	R^2	0.61 ± 0.07	0.61 ± 0.07
		RMSE	9.87 ± 2.07	9.89 ± 2.02

values had a negative effect. This also applies to the NO_2 output values (Fig. 5b), as higher values of O_3 raw data decreased the NO_2 prediction and lower values had a positive effect. Herein, however, high or low levels of NO_2 raw had no significant influence on the prediction. The mean absolute SHAP values for all features of both MLR and MRF models are reported in Fig. S8.

In order to provide a local interpretability, a heatmap for the SHAP values of the NO_2 MLR model was also elaborated (Fig. 6). The heatmap showed that lower model predictions $f(x)$, computed using Eq. (1), were linked to a dark colour for O_3 and a light colour for NO_2 for both AQ stations. This suggested that O_3 raw data had a more significant impact, mostly on the lower NO_2 concentrations than the NO_2 raw data had, while the impact of the NO_2 raw data became significant at higher concentrations.

3.5 Field validation

The scores of field validation involving the MLR- and MRF-calibrated models are summarized in Table 6. Model accuracy in predicting the O_3 concentrations is confirmed to be higher than in predicting the NO_2 concentrations. In terms of Pearson's r values, the MLR model outperforms the MRF model, exhibiting r values (0.92–0.93 for O_3 and 0.75–0.78 for NO_2) higher than MRF (0.81–0.76 for O_3 and 0.76–0.65 for NO_2), while the opposite applies in terms of the standard deviation, as MRF returns lower values than MLR. Notably,

Table 3. Statistics of the MLR model applied to the AQ1 and AQ2 stations. β_0 values are the intercepts, and β_i values are the calibration coefficients.

		Coefficient					Stat.
Pollutant	AQ ID	β_0	β_1	β_2	β_3	β_4	AdjR^2
O_3	AQ1	−180.76	−0.11	0.23	0.15	3.79	0.95
	AQ2	−133.43	−0.16	0.14	0.03	3.58	0.95
NO_2	AQ1	144.78	0.05	−0.14	−0.32	−0.93	0.69
	AQ2	126.78	0.08	−0.10	−0.23	−0.87	0.69

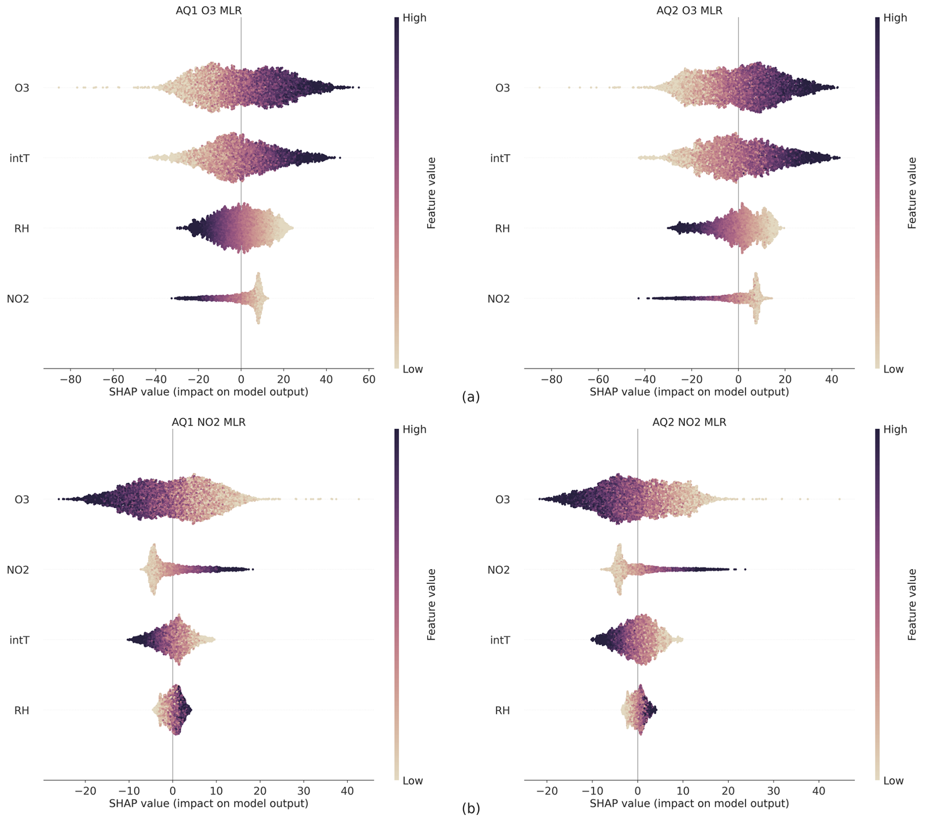

Figure 4. Bee swarm plot showing the SHAP values calculated for each feature and instance using the linear explainer the MLR model for O_3 **(a)** and NO_2 **(b)**.

a significant difference by AQ station may be observed in MRF scores, while it is not the case for MLR.

Taylor diagrams of the pre-deployment MLR- and MRF-calibrated models assessed against the ARPAT reference station for O_3 (Fig. S10a) and NO_2 (Fig. S10b) concentrations may be found in Fig. S10, while weekly concentrations predicted by the models against the reference station are given in Fig. 7.

A seasonal analysis was also performed for MLR field validation (Table 7). O_3 concentrations were well predicted across all seasons; for both stations, the lowest nRMSE values were registered in the summers of 2018 and 2019 and the highest in winter from 2018–2019. Notably, all statistical scores during the summer in 2019 proved to be worse compared to the summer in 2018, suggesting a likely drift in sensor accuracy after 1 year of deployment. As for NO_2, the highest (and thus more meaningful) concentrations were measured in winter from 2018–2019. The NO_2 scores during this period, however, are confirmed to be worse than those affecting O_3 during the period of highest O_3 concentrations (i.e. summers of 2018 and 2019). Furthermore, the scores of seasonal analysis addressed for MRF field validation may be found in Table S12.

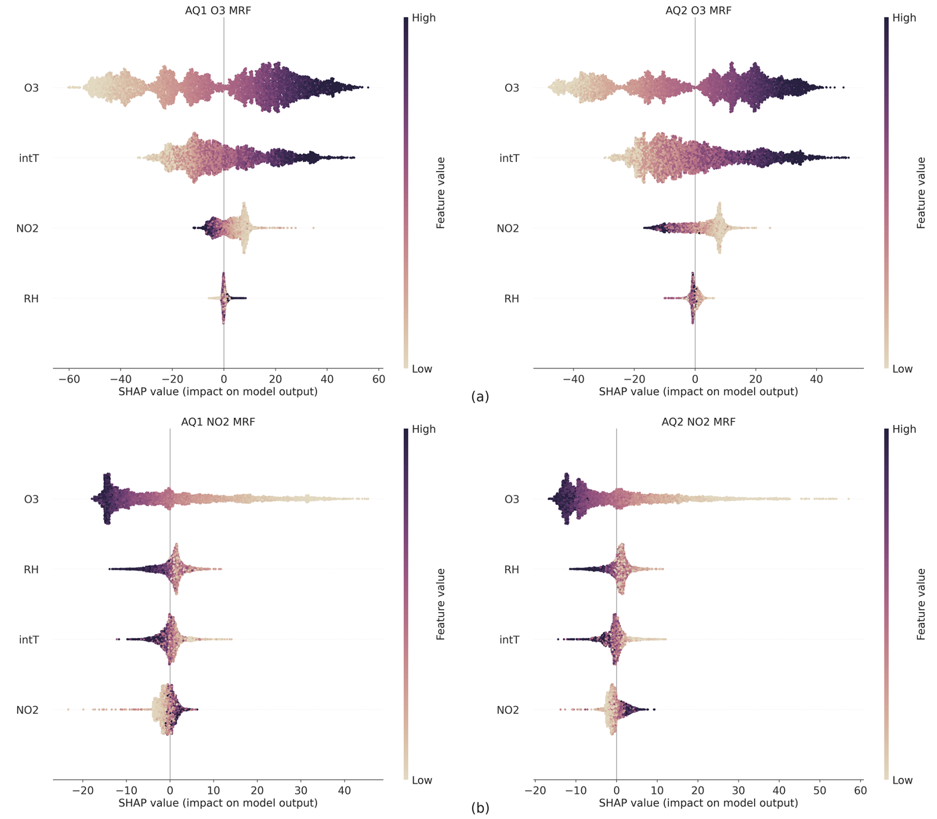

Figure 5. Bee swarm plot showing the SHAP values calculated for each feature and instance using the FastTreeSHAP of the MRF model for O_3 **(a)** and NO_2 **(b)**.

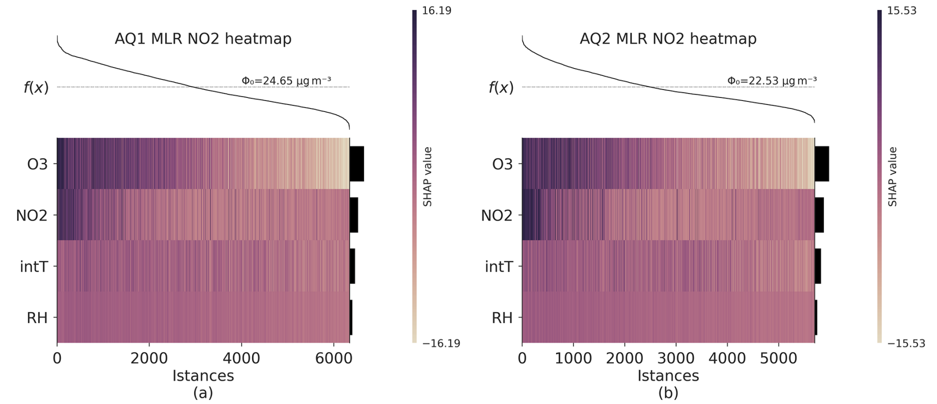

Figure 6. Heatmap of SHAP values of the NO_2 MLR model for AQ1 **(a)** and AQ2 **(b)**. The heatmap displays the contribution of each feature to the model's predictions, with positive contributions represented by dark-coloured cells and negative contributions by light-coloured cells. The colour intensity denotes the magnitude of the contribution. The output of the model, $f(x)$, is shown above the heatmap matrix, which is centred around the explanation's base value (ϕ_0), and the global importance of each model input is shown in the bar plot on the right-hand side of the plot. Observations have been ordered by the sum of the SHAP values over all features.

4 Discussion

Current outcomes achieved for the MOS O_3 and NO_2 pre-deployment calibration were generally consistent with those found in the literature. MOS NO_2 calibration exhibited low accuracy in linear univariate models, as demonstrated by Nowack et al. (2021), and the MiCS-2710 NO_2 sensor achieves a poor R^2 (0.21) compared to the O3_3E1F EC sensor value (0.845), as reported by Spinelle et al. (2015). In contrast, the O_3 sensor calibration returns high R^2 values,

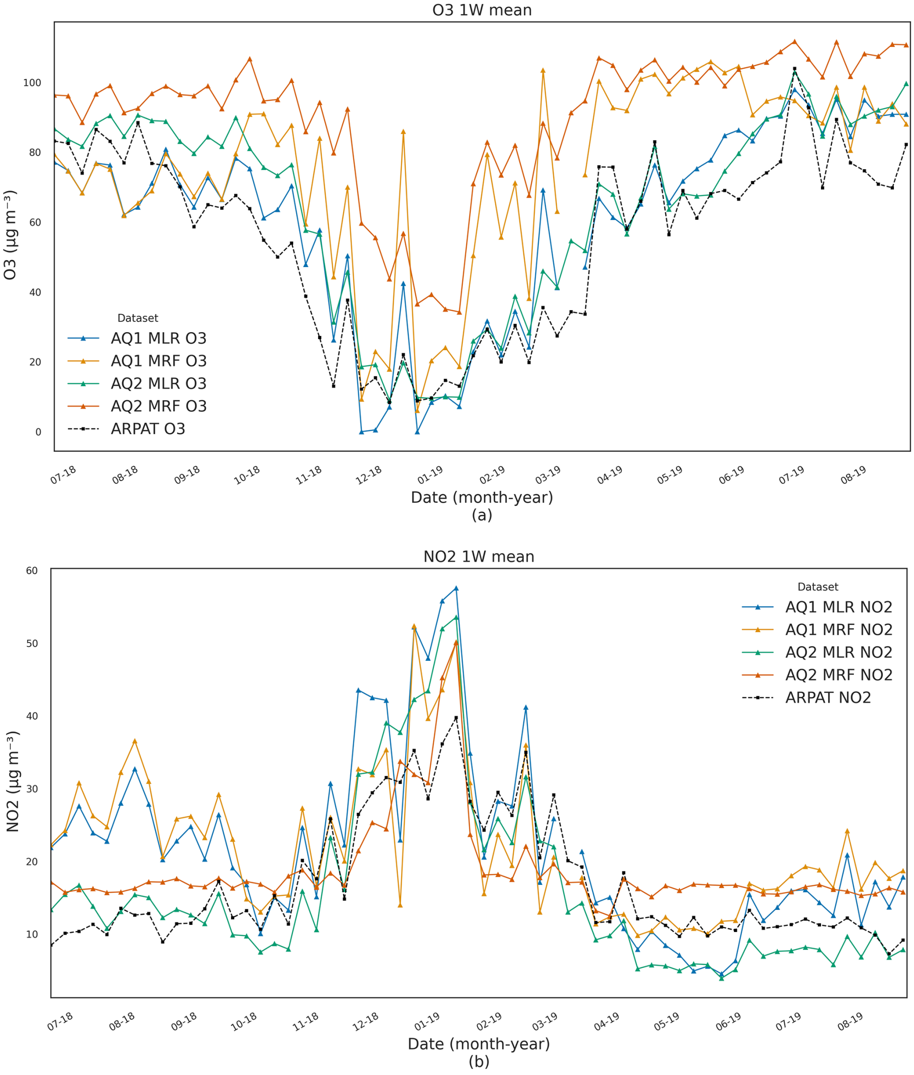

Figure 7. Trend analysis of 7 d average O_3 and NO_2 concentrations measured at the validation site by the calibrated AQ1 and AQ2 stations compared to the ARPAT reference station (17 June–22 August 2019).

suggesting limited potential for improvement using more complex univariate techniques like SVR, RF, or GB, as noted in Sales-Lérida et al. (2021). For the NO_2 calibration, incorporating multiple covariates like temperature, humidity, and gaseous interference compounds was instead essential for better performance, as emphasized in studies cited in Karagulian et al. (2019).

This study confirmed that both linear and non-linear multiple models resulted in a slight improvement in the O_3 calibration and a significant one in the NO_2 prediction when compared to univariate models (Table 4). In particular, the MLR model improved the accuracy of the simple LR by more than 14 %–18 % for O_3 and 31 %–35 % for NO_2. Notably, both MOSs performed well when using the same model form, but due to intersensor variability, each sensor necessitated a distinct set of coefficients to achieve optimal performance. Moreover, taking into account the observed multicollinearity issue between temperatures and the slightly higher mean accuracy, as well as the lower mean RMSE observed when using the internal one (Table 2), the study drew upon insights from existing literature to identify the most suitable set of covariates (e.g. Miech et al., 2021; Schmitz et al., 2021). As a result, the inclusion of internal temperature as a significant factor was given priority, as it offers a more accurate representation of the operating conditions of the MOSs within the system. This approach was also adopted to tackle potential challenges in the board's analogue-to-digital converter circuit.

However, among the multiple models, MRF proved to be the most effective in the pre-deployment calibration (e.g. Bisignano et al., 2022; Johnson et al., 2018). The SHAP methodology proved to be particularly insightful in gaining a

Table 4. Statistics of the multiple regression models applied to the AQ1 and AQ2 stations.

			Multiple models		
Pollutant	AQ ID	Stat.	MLR	MGB	MRF
O_3	AQ1	AdjR^2	0.95	0.97	0.98
		RMSE	8.62	7.30	6.04
		MAE	6.30	5.40	4.31
		MBE	−0.10	−0.01	−0.01
	AQ2	AdjR^2	0.95	0.96	0.98
		RMSE	8.58	6.86	5.51
		MAE	6.50	5.17	4.05
		MBE	−0.03	−0.03	0.09
NO_2	AQ1	AdjR^2	0.69	0.80	0.86
		RMSE	9.68	7.84	6.63
		MAE	7.36	5.76	4.72
		MBE	−0.03	0.06	0.06
	AQ2	AdjR^2	0.69	0.80	0.85
		RMSE	9.07	7.28	6.30
		MAE	6.83	5.35	4.46
		MBE	−0.08	0.03	0.05

comprehensive understanding of the behaviour of both MLR and MRF models in the pre-deployment calibration dataset. It enabled the identification of the relationships between input features (O_3, NO_2, internal temperature, and relative humidity) and the predicted outcomes. Additionally, the use of SHAP allowed for the diagnosis of potential issues, such as the non-parametric models' ability to extrapolate and predict pollution levels beyond the scope of the training calibration dataset (e.g. Nowack et al., 2021; Malings et al., 2019). These issues were confirmed through the validation process against ARPAT official reference.

As evident from Table 7, the MLR calibration model outperformed the MRF approach, thus showcasing a better transferability across diverse spatial and temporal settings. Besides, even though the pre-deployment dataset mainly represented a summer period, the physical patterns identified in the MLR model remained valid across seasons. Additionally, the SHAP heatmap (Fig. 6) provided insightful evidence of the O_3 sensor's ability to handle the lower reading limit of the NO_2 sensor for both AQ stations. This observation is important, especially in conditions with low-NO_2 concentrations, where the NO_2 sensor's accuracy in providing readings might be compromised.

On the contrary, MRF did not align perfectly with the expected underlying physical model. Instead, it appeared to be "true to the data", due to its ability to memorize specific patterns from the pre-deployment dataset, based on what emerged by the SHAP analysis (Fig. 6b). However, this characteristic posed challenges when trying to apply the model to unseen data, leading to unsatisfactory performance in the field. This lack of generalization capability hindered the MRF model's effectiveness when faced with differing concentration regimes.

The seasonal analyses presented in Table 7 provided an overview of these seasonal changes in the stability and biases of the AQ1 and AQ2 O_3 and NO_2 sensors for the application of the MLR calibration model after deployment. The O_3 pre-deployment calibration showed good performance in all seasons, and for both stations, the lowest nRMSE value was registered in summer 2018 and in summer 2019, and the highest value of nRMSE was recorded in winter 2018–2019. The decline in performance during the winter period was minimal, despite the fact that the pre-deployment calibration was mainly performed in the summer. Furthermore, a comparison of the summer period of 2018 and 2019 showed a decrease of 2 % in nRMSE for both AQ stations and pollutants (O_3 and NO_2). The decrease in nRMSE for O_3 was accompanied by an increase in the magnitude of MAE and MBE, pointing towards a possible linear drift in the O_3 sensor readings after a year of use. Conversely, for pre-deployment calibration of NO_2, a decrease in MBE was observed. The decrease in MBE for NO_2 and the prominent role of O_3 raw readings and its negative impact on prediction, as identified through feature importance analysis of the pre-deployment MLR model, further reinforced the idea of a linear drift in O_3 sensor readings. Similarly, the pattern of the lowest and highest nRMSE values for O_3 validation also remained consistent for the MRF model, with the values being the lowest in the summer of 2018 and 2019 and the highest in the winter of 2018–2019 (Table S12). Notably, AQ1 outperforms AQ2 in both models; however, as mentioned earlier, the differences in nRMSE values between the MLR and MRF models were quite significant.

5 Conclusions and perspective

In this study, the pre-deployment calibration and field validation of two low-cost (LC) stations named AirQino, developed by CNR–IBE in Florence (Italy), were addressed. The stations were equipped with O_3 and NO_2 MOSs and meteorological sensors. Pre-deployment calibration was performed after developing and implementing a comprehensive calibration framework, consisting of several elements, including parametric, non-parametric univariate, and multiple algorithms, that allowed us to identify the optimal calibration pathway. Ultimately, this resulted in robust LC performances outside the training conditions and the ability for easy adjustments to cope with changes in sensor performance over time. While selecting the most suitable LC calibration models, necessarily going beyond mere accuracy, this study primarily recommends (i) including multiple covariates, such as internal (rather than external) temperature, relative humidity, and gaseous interference compounds, into the multiple regression models and (ii) analysing the importance of the fea-

Table 5. DA statistics and PFI weights achieved for MLR and MRF models and applied to the AQ1 and AQ2 stations.

			DA					PFI
Pollutant	AQ ID	Variable	IntD	ID	APD	TD	PRI	Weight
O_3	AQ1	O_3	0.09	0.82	0.29	0.37	38.96	0.66 ± 0.01
		intT	0.07	0.62	0.20	0.27	28.64	0.48 ± 0.01
		RH	0.00	0.57	0.10	0.19	19.92	0.01 ± 0.00
		NO_2	0.02	0.26	0.10	0.12	12.47	0.16 ± 0.01
	AQ2	O_3	0.06	0.77	0.25	0.33	34.95	0.47 ± 0.01
		intT	0.11	0.64	0.22	0.30	31.51	0.55 ± 0.01
		RH	0.00	0.55	0.09	0.18	19.10	0.01 ± 0.00
		NO_2	0.03	0.31	0.10	0.14	14.44	0.20 ± 0.01
NO_2	AQ1	O_3	0.16	0.66	0.36	0.39	55.18	1.12 ± 0.02
		NO_2	0.02	0.34	0.15	0.17	23.78	0.21 ± 0.01
		intT	0.02	0.20	0.06	0.08	11.93	0.18 ± 0.01
		RH	0.03	0.18	0.02	0.06	9.11	0.22 ± 0.01
	AQ2	O_3	0.12	0.64	0.33	0.35	51.13	1.01 ± 0.04
		NO_2	0.04	0.38	0.16	0.19	26.79	0.19 ± 0.01
		intT	0.03	0.21	0.06	0.09	13.34	0.18 ± 0.01
		RH	0.03	0.18	0.02	0.06	8.74	0.16 ± 0.01

Table 6. Statistics of the MLR- and MRF-calibrated models assessed during the field validation procedure. CRMSD is the centred root mean squared difference.

Pollutant	AQ ID	Stat.	MLR	MRF
O_3	AQ1	*r*	0.92	0.81
		CRMSD	15.98	24.13
		RMSE	16.52	29.08
		MAE	12.96	22.22
		MBE	4.20	16.23
	AQ2	*r*	0.93	0.76
		CRMSD	15.35	25.44
		RMSE	17.88	42.98
		MAE	13.99	36.24
		MBE	9.17	34.64
NO_2	AQ1	*r*	0.75	0.65
		CRMSD	11.25	11.83
		RMSE	12.64	13.48
		MAE	9.40	9.97
		MBE	5.75	6.46
	AQ2	*r*	0.78	0.58
		CRMSD	10.63	11.77
		RMSE	10.65	11.94
		MAE	7.72	9.12
		MBE	−0.69	2.03

tures used in the multiple models to disclose their role when the calibrated LC stations are operated under field conditions rather than in a controlled environment.

As a novelty applied to LC MOS calibration, the SHapley Additive exPlanations (SHAP) method was used to provide further insight into the role played by individual model predictors and their global and local impact on the overall LC sensor performances. This method was also used to hypothesize the capability of the model to accurately describe conditions beyond the pre-deployment calibration period.

This study confirmed that machine learning models, such as MRF, can effectively calibrate LC sensors and mitigate the impact of environmental conditions and pollutant cross-sensitivity. However, while the MRF model demonstrated higher accuracy than MLR during pre-deployment calibration, it faced challenges in accurately representing physical models and struggled to generalize on the field validation dataset. Furthermore, as well as being less computationally demanding and generally more suitable for non-experts, parametric models such as MLR have a defined equation that also includes a few parameters, which allows – when needed – easy adjustments for possible changes over time. Thus, drift correction or periodic automatable recalibration operations can be readily scheduled, making parametric models advantageous. This aspect is particularly relevant for NO_2 and O_3 MOSs, as demonstrated in this study. Both sensors performed well with the same linear model form, requiring unique parameter values due to intersensor variability.

A limitation of the present work is that the LC stations have been calibrated during a period that is not particularly long (73 d) and a typically summer period, thus when pollution levels are generally meaningful for O_3, but they are not meaningful for NO_2 concentrations. Indeed, conducting a pre-deployment calibration during a winter period, when

Table 7. Seasonal analysis of MLR validation. Minimum and maximum values ($\mu g\,m^{-3}$) represent the minimum and maximum concentrations measured by the reference station, while intT (°C) is the average internal temperature measured by the AQ stations.

						Stat.			
Year	Season	Pollutant	AQ ID	Min–max	intT	r	nRMSE	MAE	MBE
2018	Summer	O_3	AQ1	6–166	34.65	0.94	9.17	11.69	−5.17
			AQ2	6–166	34.20	0.94	8.80	11.07	7.57
		NO_2	AQ1	1–47	34.62	0.69	35.74	14.04	13.83
			AQ2	1–47	34.16	0.69	16.97	5.94	2.90
	Autumn	O_3	AQ1	2–146	28.06	0.93	13.24	15.08	11.35
			AQ2	2–146	25.53	0.94	15.87	19.32	18.37
		NO_2	AQ1	1–62	28.07	0.73	19.66	8.62	5.57
			AQ2	1–62	25.54	0.71	16.57	7.60	−2.40
	Winter	O_3	AQ1	2–65	16.60	0.93	17.94	8.07	0.11
			AQ2	2–72	14.53	0.93	18.97	8.97	2.50
		NO_2	AQ1	3–88	16.59	0.71	21.01	13.97	10.44
			AQ2	2–88	14.53	0.70	18.34	12.16	4.62
2019	Spring	O_3	AQ1	2–132	22.41	0.86	13.32	13.98	3.36
			AQ2	2–132	21.14	0.84	13.91	14.40	4.63
		NO_2	AQ1	2–63	22.32	0.74	13.18	6.10	−2.09
			AQ2	2–63	21.09	0.67	16.19	7.31	−5.36
	Summer	O_3	AQ1	7–185	34.98	0.92	10.09	14.62	10.21
			AQ2	7–185	32.29	0.92	10.68	15.98	12.71
		NO_2	AQ1	0–47	34.94	0.63	17.82	6.36	3.58
			AQ2	0–47	32.25	0.68	13.68	5.01	−3.05

NO_2 concentrations are typically higher, would be a valuable addition to the study. This step would provide a more comprehensive understanding of the AQ station validation performance under varying pollution conditions and help address the limitation of the current calibration period biased towards summer data. Moreover, conducting a similar validation outside of Italy, in regions with differing pollution and meteorological conditions, would be of great interest. For this purpose, in the ongoing activity, the AirQino LC stations are planned to be deployed outside Italy, such as in Nice and Aix-en-Provence (France), Barcelona (Spain), Budapest (Hungary), Tirana (Albania), and Niamey (Niger).

Furthermore, in the future, a new sensor for monitoring NO could hopefully be integrated into the LC stations and validated. As such, the combined monitoring of NO, NO_2, and O_3 concentrations and their daily and seasonal variability would allow a comprehensive pattern of the oxidant capacity of atmosphere, which is particularly effective in southern Mediterranean countries such as Italy (Pancholi et al., 2018). In addition, once the AQ VOC sensor is validated, it will enable the monitoring of all O_3 precursors (VOC and NO_x). This comprehensive monitoring, combined with the application of SHapley Additive exPlanations (SHAP) method, will lead to a full characterization of photochemical pollution in various areas of interest, including urban, sub-urban, or rural regions. Moreover, the portability of LC sensors makes them ideal devices for filling knowledge gaps in regions that are difficult to access, such as the open sea. When mounted on buoys or ships, for example, LC sensors could collect the high-O_3 levels that typically occur over these areas in summer due to high solar activity and rather low mixing height combined with a lack of O_3-consuming NO emissions.

Appendix A: Nomenclature

APD	Average partial dominance	LC	Low cost	RBF	Radial basis function
AQ	AirQino	LNLR	Logarithm regression	RF	Random forest
CDLR	Cook's distance regression	LR	Linear regression	RH	Relative humidity
DA	Dominance analysis	MGB	Multiple gradient boosting	SHAP	SHapley Additive exPlanations
EC	Electrochemical	MLR	Multiple linear regression	SLR	Supervised linear regression
EDA	Exploratory data analysis	MOS	Metal oxide sensors	SNLR	Supervised non-linear regression
extT	External temperature	MRF	Multiple random forest	SVM	Support vector machine
GB	Gradient boosting	PFI	Permutation feature importance	SVR	Support vector regression
HBLR	Huber regression	PLR2	Polynomial regression of second degree	TD	Total dominance
ID	Individual dominance	PLR3	Polynomial regression of third degree	VIF	Variance impact factor
IntD	Interactional dominance	PNLR	Power non-linear regression		
intT	Internal temperature	PRI	Percentage relative importance		

Code and data availability. All data (HORIBA reference data, AirQino raw signal data, and ARPAT validation data) and codes (Jupyter Notebook) to recreate the results discussed here are provided at https://doi.org/10.5281/zenodo.7826791 (Cavaliere, 2023).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/amt-16-4723-2023-supplement.

Author contributions. Conceptualization: AC, BG, FC, and AZ. Methodology: AC, BG, FC, and GG. Software: AC. Formal analysis: AC, GG, and BG. Investigation: LB, FC, BG, and AZ. Data curation: AC, FC, and TG. Original draft preparation: AC and LB. Review and editing: AC, LB, GG, and BG. Visualization: BPA, MS, and CV. Project administration: AZ and CV. All authors have read and agreed to the published version of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Review statement. This paper was edited by Albert Presto and reviewed by Mark Joseph Campmier and one anonymous referee.

References

Aleixandre, M., Gerboles, M., and Spinelle, L.: Report of the laboratory and in-situ validation of micro-sensors and evaluation of suitability of model equations NO9: CairClipNO2 of CAIRPOL (F), Publications Office of the European Union, Luxembourg, oCLC: 1111194588, 2013.

Asair: Datasheet AM2305C, https://asairsensors.com/wp-content/uploads/2021/09/Data-Sheet-AM2315C-Humidity-and-Temperature-Module-ASAIR-V1.0.02.pdf, 29 September 2021.

Aula, K., Lagerspetz, E., Nurmi, P., and Tarkoma, S.: Evaluation of Low-Cost Air Quality Sensor Calibration Models, ACM Transactions on Sensor Networks, 3512889, https://doi.org/10.1145/3512889, 2022.

Azen, R. and Budescu, D. V.: The Dominance Analysis Approach for Comparing Predictors in Multiple Regression, Psychol. Meth., 8, 129–148, https://doi.org/10.1037/1082-989X.8.2.129, 2003.

Barcelo-Ordinas, J. M., Ferrer-Cid, P., Garcia-Vidal, J., Ripoll, A., and Viana, M.: Distributed Multi-Scale Calibration of Low-Cost Ozone Sensors in Wireless Sensor Networks, Sensors, 19, 2503, https://doi.org/10.3390/s19112503, 2019.

Bart, M., Williams, D. E., Ainslie, B., McKendry, I., Salmond, J., Grange, S. K., Alavi-Shoshtari, M., Steyn, D., and Henshaw, G. S.: High Density Ozone Monitoring Using Gas Sensitive Semi-Conductor Sensors in the Lower Fraser Valley, British Columbia, Environ. Sci. Technol., 48, 3970–3977, https://doi.org/10.1021/es404610t, 2014.

Bigi, A., Mueller, M., Grange, S. K., Ghermandi, G., and Hueglin, C.: Performance of NO, NO_2 low cost sensors and three calibration approaches within a real world application, Atmos. Meas. Tech., 11, 3717–3735, https://doi.org/10.5194/amt-11-3717-2018, 2018.

Bisignano, A., Carotenuto, F., Zaldei, A., and Giovannini, L.: Field calibration of a low-cost sensors network to assess traffic-related air pollution along the Brenner highway, Atmos. Environ., 275, 119008, https://doi.org/10.1016/j.atmosenv.2022.119008, 2022.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brilli, L., Carotenuto, F., Andreini, B. P., Cavaliere, A., Esposito, A., Gioli, B., Martelli, F., Stefanelli, M., Vagnoli, C., Venturi, S., Zaldei, A., and Gualtieri, G.: Low-Cost Air Quality Stations' Capability to Integrate Reference Stations in Particulate Matter Dynamics Assessment, Atmosphere, 12, 1065, https://doi.org/10.3390/atmos12081065, 2021.

Burgués, J. and Marco, S.: Low Power Operation of Temperature-Modulated Metal Oxide Semiconductor Gas Sensors, Sensors, 18, 339, https://doi.org/10.3390/s18020339, 2018.

Camalier, L., Cox, W., and Dolwick, P.: The effects of meteorology on ozone in urban areas and their use in assessing ozone trends, Atmos. Environ., 41, 7127–7137, https://doi.org/10.1016/j.atmosenv.2007.04.061, 2007.

Cavaliere, A.: Code and dataset for: Development of Low-Cost Air Quality Stations for Next Generation Monitoring Networks: Calibration and Validation of NO_2 and O_3 Sensors, Zenodo [code and data set], https://doi.org/10.5281/zenodo.7826791, 2023.

Cavaliere, A., Carotenuto, F., Di Gennaro, F., Gioli, B., Gualtieri, G., Martelli, F., Matese, A., Toscano, P., Vagnoli, C., and Zaldei, A.: Development of Low-Cost Air Quality Stations for Next Generation Monitoring Networks: Calibration and Validation of $PM_{2.5}$ and PM_{10} Sensors, Sensors, 18, 2843, https://doi.org/10.3390/s18092843, 2018.

Chakraborty, S., Mittermaier, S., Carbonelli, C., and Servadei, L.: Explainable AI for Gas Sensors, in: 2022 IEEE Sensors, 1–4, IEEE, https://doi.org/10.1109/SENSORS52175.2022.9967180, 2022.

Claveau, C., Giraudon, M., Coville, B., Saussac, A., Eymard, L., Turcati, L., and Payan, S.: Performance comparison between electrochemical and semiconductors sensors for the monitoring of O_3, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2022-75, 2022.

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

Cook, R. D.: Detection of Influential Observation in Linear Regression, Technometrics, 19, 15–18, https://doi.org/10.1080/00401706.1977.10489493, 1977.

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

De Vito, S., Piga, M., Martinotto, L., and Di Francia, G.: CO, NO_2 and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization, Sensor. Actuator. B, 143, 182–191, https://doi.org/10.1016/j.snb.2009.08.041, 2009.

De Vito, S., Di Francia, G., Esposito, E., Ferlito, S., Formisano, F., and Massera, E.: Adaptive machine learning strategies for network calibration of IoT smart air quality monitoring devices, Pattern Recogn. Lett., 136, 264–271, https://doi.org/10.1016/j.patrec.2020.04.032, 2020.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., and Meester, L. E.: A Modern Introduction to Probability and Statistics, Springer Texts in Statistics, Springer London, London, https://doi.org/10.1007/1-84628-168-7, 2005.

Di Lonardo, S., Zaldei, A., Toscano, P., Matese, A., Gioli, B., Rocchi, L., Vagnoli, C., De Filippis, T., Gualtieri, G., and Martelli, F.: The SensorWebBike for air quality monitoring in a smart city, in: IET Conference on Future Intelligent Cities, 2, Institution of Engineering and Technology, London, UK, https://doi.org/10.1049/ic.2014.0043, 2014.

Draper, N. R. and Smith, H.: Applied regression analysis, Wiley series in probability and statistics, Wiley, New York, 3rd edn., ISBN 978-0-471-17082-2, 1998.

Esposito, E., Salvato, M., Vito, S. D., Fattoruso, G., Castell, N., Karatzas, K., and Francia, G. D.: Assessing the Relocation Robustness of on Field Calibrations for Air Quality Monitoring Devices, in: Sensors and Microsystems, Lecture Notes in Electrical Engineering, edited by: Leone, A., Forleo, A., Francioso, L., Capone, S., Siciliano, P., and Di Natale, C., Springer International Publishing, Cham, 457, 303–312, https://doi.org/10.1007/978-3-319-66802-4_38, 2018.

Fine, G. F., Cavanagh, L. M., Afonja, A., and Binions, R.: Metal Oxide Semi-Conductor Gas Sensors in Environmental Monitoring, Sensors, 10, 5469–5502, https://doi.org/10.3390/s100605469, 2010.

Gäbel, P., Koller, C., and Hertig, E.: Development of Air Quality Boxes Based on Low-Cost Sensor Technology for Ambient Air Quality Monitoring, Sensors, 22, 3830, https://doi.org/10.3390/s22103830, 2022.

Gu, K., Qiao, J., and Lin, W.: Recurrent Air Quality Predictor Based on Meteorology- and Pollution-Related Factors, IEEE T. Ind. Inform., 14, 3946–3955, https://doi.org/10.1109/TII.2018.2793950, 2018.

Han, P., Mei, H., Liu, D., Zeng, N., Tang, X., Wang, Y., and Pan, Y.: Calibrations of Low-Cost Air Pollution Monitoring Sensors for CO, NO_2, O_3, and SO_2, Sensors, 21, 256, https://doi.org/10.3390/s21010256, 2021.

Han, S., Bian, H., Feng, Y., Liu, A., Li, X., Zeng, F., and Zhang, X.: Analysis of the Relationship between O_3, NO and NO_2 in Tianjin, China, Aerosol Air Qual. Res., 11, 128–139, https://doi.org/10.4209/aaqr.2010.07.0055, 2011.

Holstius, D. M., Pillarisetti, A., Smith, K. R., and Seto, E.: Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California, Atmos. Meas. Tech., 7, 1121–1131, https://doi.org/10.5194/amt-7-1121-2014, 2014.

Idrees, Z. and Zheng, L.: Low cost air pollution monitoring systems: A review of protocols and enabling technologies, Journal of Industrial Information Integration, 17, 100123, https://doi.org/10.1016/j.jii.2019.100123, 2020.

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

Karagulian, F.: New Challenges in Air Quality Measurements, in: Air Quality Networks, Environmental Informatics and Modeling, edited by: De Vito, S., Karatzas, K., Bartonova, A., and Fattoruso, G., Springer International Publishing, Cham, https://doi.org/10.1007/978-3-031-08476-8_1, 1–18, 2023.

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

Lin, Y., Dong, W., and Chen, Y.: Calibrating Low-Cost Sensors by a Two-Phase Learning Approach for Urban Air

Quality Measurement, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2, 1–18, https://doi.org/10.1145/3191750, 2018.

Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, Advances in neural information processing systems, arXiv, 2, https://doi.org/10.48550/ARXIV.1705.07874, 2017.

Lundberg, S. and Lee, S.-I.: SHAP, https://shap.readthedocs.io/en/latest/ (last access: 16 June 2022), 2022.

Maag, B., Saukh, O., Hasenfratz, D., and Thiele, L.: Pre-Deployment Testing, Augmentation and Calibration of Cross-Sensitive Sensors., in: EWSN, 169–180, 2016.

Maag, B., Zhou, Z., and Thiele, L.: A Survey on Sensor Calibration in Air Pollution Monitoring Deployments, IEEE Internet Things, 5, 4857–4870, https://doi.org/10.1109/JIOT.2018.2853660, 2018.

Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S. P. N., Zimmerman, N., Kara, L. B., Presto, A. A., and R. Subramanian: Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring, Atmos. Meas. Tech., 12, 903–920, https://doi.org/10.5194/amt-12-903-2019, 2019.

Masson, N., Piedrahita, R., and Hannigan, M.: Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring, Sensor. Actuator. B, 208, 339–345, https://doi.org/10.1016/j.snb.2014.11.032, 2015.

Maxim Integrated: Datasheet DS18B20, https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf, last access: 29 September 2021.

Mead, M., Popoola, O., Stewart, G., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J., McLeod, M., Hodgson, T., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J., and Jones, R.: 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, 2013.

Meng, X., Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A. M., Milojevic, A., Guo, Y., Tong, S., Coelho, M. D. S. Z. S., and Saldiva, P. H. N.: Short Term Associations of Ambient Nitrogen Dioxide with Daily Total, Cardiovascular, and Respiratory Mortality: Multilocation Analysis in 398 Cities, BMJ, n534, https://doi.org/10.1136/bmj.n534, 2021.

Miech, J., Stanton, L., Gao, M., Micalizzi, P., Uebelherr, J., Herckes, P., and Fraser, M.: Calibration of Low-Cost NO_2 Sensors through Environmental Factor Correction, Toxics, 9, 281, https://doi.org/10.3390/toxics9110281, 2021.

Mijling, B., Jiang, Q., de Jonge, D., and Bocconi, S.: Field calibration of electrochemical NO_2 sensors in a citizen science context, Atmos. Meas. Tech., 11, 1297–1312, https://doi.org/10.5194/amt-11-1297-2018, 2018.

Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai, F., Christensen, B., Dunbabin, M., Gao, J., Hagler, G. S., Jayaratne, R., Kumar, P., Lau, A. K., Louie, P. K., Mazaheri, M., Ning, Z., Motta, N., Mullins, B., Rahman, M. M., Ristovski, Z., Shafiei, M., Tjondronegoro, D., Westerdahl, D., and Williams, R.: 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, 2018.

Mueller, M., Meyer, J., and Hueglin, C.: Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich, Atmos. Meas. Tech., 10, 3783–3799, https://doi.org/10.5194/amt-10-3783-2017, 2017.

Narayana, M. V., Jalihal, D., and Nagendra, S. M. S.: Establishing A Sustainable Low-Cost Air Quality Monitoring Setup: A Survey of the State-of-the-Art, Sensors, 22, 394, https://doi.org/10.3390/s22010394, 2022.

Nowack, P., Konstantinovskiy, L., Gardiner, H., and Cant, J.: Machine learning calibration of low-cost NO_2 and PM_{10} sensors: non-linear algorithms and their impact on site transferability, Atmos. Meas. Tech., 14, 5637–5655, https://doi.org/10.5194/amt-14-5637-2021, 2021.

Nuvolone, D., Petri, D., and Voller, F.: The Effects of Ozone on Human Health, Environ. Sci. Pollut. Res., 25, 8074–8088, https://doi.org/10.1007/s11356-017-9239-3, 2018.

Pancholi, P., Kumar, A., Bikundia, D. S., and Chourasiya, S.: An observation of seasonal and diurnal behavior of O_3–NO_x relationships and local/regional oxidant ($OX = O_3 + NO_2$) levels at a semi-arid urban site of western India, Sustain. Environ. Res., 28, 79–89, https://doi.org/10.1016/j.serj.2017.11.001, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., and Cournapeau, D.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Peterson, P., Aujla, A., Grant, K., Brundle, A., Thompson, M., Vande Hey, J., and Leigh, R.: Practical Use of Metal Oxide Semiconductor Gas Sensors for Measuring Nitrogen Dioxide and Ozone in Urban Environments, Sensors, 17, 1653, https://doi.org/10.3390/s17071653, 2017.

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., Li, K., Dick, R. P., Lv, Q., Hannigan, M., and Shang, L.: 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, 2014.

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

Sahu, R., Nagal, A., Dixit, K. K., Unnibhavi, H., Mantravadi, S., Nair, S., Simmhan, Y., Mishra, B., Zele, R., Sutaria, R., Motghare, V. M., Kar, P., and Tripathi, S. N.: Robust statistical calibration and characterization of portable low-cost air quality monitoring sensors to quantify real-time O_3 and NO_2 concentrations in diverse environments, Atmos. Meas. Tech., 14, 37–52, https://doi.org/10.5194/amt-14-37-2021, 2021.

Sales-Lérida, D., Bello, A. J., Sánchez-Alzola, A., and Martínez-Jiménez, P. M.: An Approximation for Metal-Oxide Sensor Calibration for Air Quality Monitoring Using Multivariable Statistical Analysis, Sensors, 21, 4781, https://doi.org/10.3390/s21144781, 2021.

Sayahi, T., Garff, A., Quah, T., Lê, K., Becnel, T., Powell, K. M., Gaillardon, P.-E., Butterfield, A. E., and Kelly, K. E.: Long-term calibration models to estimate ozone concentrations with a metal oxide sensor, Environ. Pollut., 267, 115363, https://doi.org/10.1016/j.envpol.2020.115363, 2020.

Schmitz, S., Towers, S., Villena, G., Caseiro, A., Wegener, R., Klemp, D., Langer, I., Meier, F., and von Schneidemesser, E.: Unravelling a black box: an open-source methodology for the field calibration of small air quality sensors, Atmos. Meas. Tech., 14, 7221–7241, https://doi.org/10.5194/amt-14-7221-2021, 2021.

Seabold, S. and Perktold, J.: statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference, Austin, Texas, 28–30 June 2021, https://doi.org/10.25080/Majora-92bf1922-011, 2010.

Sensortech, S. G. X.: Datasheet MiCS-2714, https://www.sgxsensortech.com/content/uploads/2014/08/1107_Datasheet-MiCS-2714.pdf (last access: 29 September 2021), a.

Sensortech, S. G. X.: Datasheet MiCS-2614, https://sensorsandpower.angst-pfister.com/fileadmin/products/datasheets/188/MOS-Ozone-MiCS-2614_1620-21530-0006-E-0714.pdf (last access: 29 September 2021), b.

Shekhar, S., Bhagat, S., Kunjithapatham, S., and Kolluri, B. K.: Dominance-Analysis, https://github.com/dominance-analysis/dominance-analysis, (last access: 29 September 2022), 2019.

Smets, K., Verdonk, B., and Jordaan, E. M.: Evaluation of performance measures for SVR hyperparameter selection, in: 2007 International Joint Conference on Neural Networks, IEEE, 637–642, 2007.

Spiess, A.-N. and Neumeyer, N.: An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach, BMC Pharmacology, 10, 6, https://doi.org/10.1186/1471-2210-10-6, 2010.

Spinelle, L., Aleixandre, M., and Gerboles, M.: Protocol of evaluation and calibration of low-cost gas sensors for the monitoring of air pollution, Publications Office of the European Union, Luxembourg, https://doi.org/10.2788/9916, 2013.

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., and Bonavitacola, F.: Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide, Sensor. Actuator. B, 215, 249–257, https://doi.org/10.1016/j.snb.2015.03.031, 2015.

Spinelle, L., Gerboles, M., Aleixandre, M., and Bonavitacola, F.: Evaluation of metal oxides sensors for the monitoring of O_3 in ambient air at ppb level, Chem. Engineer. Trans., 54, 319–324, https://doi.org/10.3303/CET1654054, 2016.

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., and Bonavitacola, F.: Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2, Sensor. Actuator. B, 238, 706–715, https://doi.org/10.1016/j.snb.2016.07.036, 2017.

TeamHG-Memex: Eli5, https://eli5.readthedocs.io/en/latest/ (last access: 25 October 2022), 2022.

Vega García, M. and Aznarte, J. L.: Shapley additive explanations for NO_2 forecasting, Ecol. Inform., 56, 101039, https://doi.org/10.1016/j.ecoinf.2019.101039, 2020.

Wang, A., Machida, Y., deSouza, P., Mora, S., Duhl, T., Hudda, N., Durant, J. L., Duarte, F., and Ratti, C.: Leveraging Machine Learning Algorithms to Advance Low-Cost Air Sensor Calibration in Stationary and Mobile Settings, Atmos. Environ., 301, 119692, https://doi.org/10.1016/j.atmosenv.2023.119692, 2023.

Wang, C., Yin, L., Zhang, L., Xiang, D., and Gao, R.: Metal Oxide Gas Sensors: Sensitivity and Influencing Factors, Sensors, 10, 2088–2106, https://doi.org/10.3390/s100302088, 2010.

Wei, P., Ning, Z., Ye, S., Sun, L., Yang, F., Wong, K., Westerdahl, D., and Louie, P.: Impact Analysis of Temperature and Humidity Conditions on Electrochemical Sensor Response in Ambient Air Quality Monitoring, Sensors, 18, 59, https://doi.org/10.3390/s18020059, 2018.

Williams, D. E., Henshaw, G. S., Bart, M., Laing, G., Wagner, J., Naisbitt, S., and Salmond, J. A.: Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network, Meas. Sci. Technol., 24, 065803, https://doi.org/10.1088/0957-0233/24/6/065803, 2013.

World Health Organization: WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide, World Health Organization, Geneva, https://apps.who.int/iris/handle/10665/345329 (last access: 26 September 2022), section: xxi, 273 pp., 2021.

Yang, J.: Fast TreeSHAP: Accelerating SHAP Value Computation for Trees, arXiv, 3, https://doi.org/10.48550/ARXIV.2109.09847, 2021.

Yang, J.: FastTreeSHAP, https://github.com/linkedin/FastTreeSHAP (last access: 29 November 2022), 2022.

Zaldei, A., Camilli, F., De Filippis, T., Di Gennaro, F., Di Lonardo, S., Dini, F., Gioli, B., Gualtieri, G., Matese, A., Nunziati, W., Rocchi, L., Toscano, P., and Vagnoli, C.: An integrated low-cost road traffic and air pollution monitoring platform for next citizen observatories, Transport. Res. Proced., 24, 531–538, https://doi.org/10.1016/j.trpro.2017.06.002, 2017.

Zauli-Sajani, S., Marchesi, S., Pironi, C., Barbieri, C., Poluzzi, V., and Colacci, A.: Assessment of air quality sensor system performance after relocation, Atmos. Pollut. Res., 12, 282–291, https://doi.org/10.1016/j.apr.2020.11.010, 2021.

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., and Subramanian, R.: A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, Atmos. Meas. Tech., 11, 291–313, https://doi.org/10.5194/amt-11-291-2018, 2018.