Improving Raw Readings from Low-Cost Ozone Sensors Using Artificial Intelligence for Air Quality Monitoring

Guillem Montalban-Faet, Eric Meneses-Albala, Santiago Felici-Castell, Juan J. Perez-Solano, and Jaume Segura-Garcia

Departament de Informàtica, ETSE, Universitat de València, Avd. de la Universidad S/N, 46100 Burjassot, (Valencia), Spain

Correspondence: Santiago Felici-Castell (felici@uv.es)

Abstract. Ground-level ozone (O3) is a highly oxidizing gas with very reactive properties, harmful at high levels, generated by complex photochemical reactions when "primary" pollutants from combustion of fossil materials react with sunlight. Thus, its concentration indicates the activity of other air pollutants and plays a crucial role in air quality monitoring systems in smart cities. To increase its spatial sampling resolution over the city map, low-cost ozone sensors are an interesting alternative, but they have a lack of accuracy. In this context, artificial intelligence techniques, particularly ensemble machine learning methods, can improve the raw readings from these sensors taking into account additional environmental information. In this paper, we analyze, propose and compare different techniques, reducing the estimation error by approximately 94.05%, with a mean relative error of 7.21% using the Gradient Boosting algorithm and outperforming the related work using a sensor approximately 10 times less expensive.

1 Introduction

Air Quality (AQ) is a fundamental aspect of environmental health, addressing the composition and purity of atmospheric gases, in terms of fine Particulate Matter (PM), Nitrogen Oxides (such as NO, NO2 and total NOx), Sulfur Dioxide (SO2), Total Volatile Organic Compounds (TVOC) and ground-level Ozone (O3), now on named simply as O3.

AQ has a direct impact on both human health and the environment (Manisalidis et al. (2020)). According to World Health Organization (WHO) (H. Adair-Rohani (2024)), 99% of the world's population breathes air that exceeds the limit values of the recommended safety Air Quality Guideline (AQG) (Organization et al. (2021)). This guideline specifies recommended levels for these pollutants for both short-term and long-term exposure. It is regularly reviewed and updated to incorporate the latest scientific evidence on the health effects of air pollution. This helps governments and authorities establish and implement policies to protect human health from the adverse effects of air pollution.

Among these pollutants, we focus on O3, a highly oxidizing gaseous pollutant, that has very reactive properties and is harmful at high levels. Notice that in this AQG with regard to O3, the target is to achieve a concentration of 100 $\mu g/m^3$ measured on average of daily maximum 8 hours. Continued exposure to levels above those recommended by this AQG may lead to respiratory irritation, lung inflammation, aggravation of respiratory diseases such as asthma or bronchitis, cell damage

and may have associated effects on the cardiovascular system. Those at the highest risk include children, older adults, people with respiratory or heart conditions, and individuals who spend significant time outdoors (Garcia et al. (2021)).

This gas is very important to monitor, because it is called a secondary pollutant, which is generated in cities by complex photochemical reactions when "primary" pollutants from combustion of fossil materials (such as NO, NO2 and SO2) react with sunlight (Seinfeld and Pandis (2016)). Thus, its concentration indicates the activity of other air pollutants and plays a crucial role in AQ monitoring systems in smart cities to help their citizens improve their quality of life. It is worth mentioning that it is being recommended to increase the spatial sampling resolution of this pollutant, ideally at least one sample per 100 m^2, according to Annex III-B of the European (Directive 2008/50/EC (2008)). And Low-Cost Sensor (LCS) are becoming increasingly important, an interesting alternative, but they do not have good accuracy (Borrego et al. (2016)) in comparison with the official equipment, due to limitations in their sensing technology, lack of frequent calibration, sensitivity to environmental factors, cross-sensitive issues, use of less durable materials and the absence of rigorous certification processes. While official equipment uses advanced technologies and is subject to strict standards of accuracy and reliability, LCS are designed to offer basic monitoring at a low price, which involves sacrifices in accuracy and durability. So, in this context it is a challenge to estimate the official measurements from these LCS with a reduced error (García et al. (2022); Borrego et al. (2016)).

Artificial Intelligence (AI) techniques are valuable for environmental research due to their capacity to process large datasets and identify patterns that enhance system explainability and clarify the behavior of these AQ parameters (Zhu et al. (2023)).

In this paper, we show that Machine Learning (ML) models, in particular ensemble models, can correct the raw readings from LCS by taking into account additional environmental information, such as Temperature (Temp), Relative Humidity (RH), as well as other pollutants, being able to use these sensors to extend the resolution of air quality monitoring networks at low cost, but assuming a small error. This is our main objective. We propose and compare different techniques, reducing the estimation error up to 94.05% based on Mean Absolute Error (MAE) measurements, with a Mean Relative Error (MRE) of 7.21%, achieving the best results with the Gradient Boosting (GB) algorithm and outperforming the related work, using sensors approximately 10 times less expensive. We also carry out the calibration process using Random Forest (RF), Adaptive Boosting (ADA) and Decision Tree (DT) models.

The rest of the paper is structured as follows. Section 2 introduces the related work. Section 3 explains the experimental work carried out for the deployment of LCS and shows the data processing, as well as the use of ML algorithms for the O3 calibration of these LCS. The results are shown in Section 4 and finally, the conclusions and future work are presented in Section 5.

2 Related work

Regarding AQ LCS, due to the increasing market demand, a wide variety of them are available to measure different pollutants, gases and particles. These sensors are available in different price ranges and are more affordable compared to regulated measuring station.

Module	Sensors	Price range
SDSO11 (Nova Fitness Co., Ltd. (2024))	Temp, RH, PM, PA	Low
DL-LP8P (DecentLab, Ltd. (2024))	Temp, RH, CO2, PA	Low
MiCS-6814 (SGX, SensorTech (2024))	CO, NO2, C2H5OH, NH3, CH4	Low
ZPHS01B (Zhengzhou Winsen Electronics Technology Co. (2024))	Temp, RH, PM1-10, CO, CO2, O3, NO2, TVOC	Mid-Low
Sensit RAMP (Sensit (2024))	PM2.5, CO, CO2, NO, NO2, O3	High
AirSensEUR (Van Poppel et al. (2023))	NO, NO2, O3, CO, PM2.5, PM10, PM1, CO2	Mid-High

Table 1. AQ Sensor modules with cost estimate: Low (less than 10$), Mid-Low (100-200$), Mid-High (600-1000$) and High (≈<4000$).

Since in AQ different pollutants are considered and each sensor measures only one, we will analyze sensor modules that embed some of these LCS. A list of these sensor modules with a cost estimate is given in Table 1. The selection criteria of these modules is determined by the related work, selecting those modules which have been considered under a similar studies as the proposed here. We must stress that these modules have different costs due to their quality, order quantity, country, etc. that we can classify in: Low (less than 10$), Mid-Low (100-200$), Mid-High (600-1000$) and High (≈<4000$). A larger selection and comparison of these LCS modules are given in (García et al. (2022)) and (Borrego et al. (2016)).

Note that LCS are designed for basic monitoring at a low cost, which compromises accuracy and durability. In this list, there are several types of LCS. Optical type sensors, such as SDSO11 (Nova Fitness Co., Ltd. (2024)) and DL-LP8P (DecentLab, Ltd. (2024)), that measure the amount of light absorbed by a given gas. Metal-oxide sensors, such as SGX, SensorTech (2024) that measure the change in electrical conductivity on a semiconductor due to the presence of certain gases. Usually this type of sensors are the cheapest and are particularly susceptible to cross sensitivities. And electrochemical sensors that have higher selectivity, good for measuring specific gases, but they are more expensive. Among these, Sensit RAMP (Sensit (2024)) and AirSensEUR (Van Poppel et al. (2023)) use this type of sensors. Finally, the ZPHS01B module (Zhengzhou Winsen Electronics Technology Co. (2024)) integrates optical, metal-oxide and electrochemical sensors and it is a Mid-Low price module with the best *price/sensor* ratio.

Since one of the key points to improve the accuracy of these LCS is the use of marginal information (such Temp, RH as well as other AQ pollutants), exploited using AI techniques (Karagulian et al. (2019); Esposito et al. (2016)) as mentioned before, it is necessary to use multi-gas modules embedding as many AQ LCS as possible.

Thus, among the different low-cost alternatives and taking into account the number of sensors and the *price/sensor* ratio, the ZPHS01B (Zhengzhou Winsen Electronics Technology Co. (2024)) is the AQ sensor module that best meets the needs and objectives of this study at the time of writing, since it embeds 9 different sensors: Temp(°C), RH (%), as well as CO, CO2, NO2, O3 that are measured in Parts Per Million (ppm), formaldehyde (CH2O) that is measured in mg/m^3, PM measured in $\mu g/m^3$ and TVOC that is measured using 4 levels according to its concentration (0-very low, 1-low, 2-intermediate and 3-high). Table 2 summarizes all this information. Notice that the O3 sensor used in this module is the electrochemical ZE27-O3 (Corp (2024)) that measures within the range 0-10 ppm with a resolution of 0.01 ppm. It operates with an accuracy of $\pm$0.1 ppm when the concentration is $\leq$1 ppm and $\pm$20% when the concentration is above 1 ppm. Also, notice that the PM readings in

Table 2. AQ information from the ZPHS01B module and units.

Parameter	[Unit]	Range of Measurement
Temperature	[ºC]	-20-65
Humidity	[%R.H.]	0-100
PM2.5	$[\mu g/m^3]$	0-1000
TVOC	levels	0-3
CH2O	$[mg/m^3]$	0-6.25
CO2	[ppm]	0-5000
CO	[ppm]	0-500
O3	[ppm]	0-10
NO2	[ppm]	0.1-10

this module are given for 2.5 (fine particles with a diameter of 2.5 μm), and PM1 and PM10 are estimated from the PM2.5 readings.

Based on this ZPHS01B module, there are several research works and projects. In (Coto-Fuentes et al. (2022)), it is shown the implementation of a device for AQ outdoor evaluation using directly this module without calibration, to map AQ pollutants in a metropolitan area. In (Felici-Castell et al. (2023)), this module is used in an AQ monitoring network, where different neural networks have been trained for forecasting of pollutant concentrations, with an estimation error of 7.2% on average and where the calibration process is done on a daily basis, but not specified. In (Vaheed et al. (2022)), this module is used for indoor AQ monitoring and calculating an AQ index. In (Antonenko et al. (2023)), the authors explain briefly the use of a neural network to determine (classify) types of air: with or without pollution. Also, in (Kennedy et al. (2021)), it is shown a prototype to measure ground to stratosphere AQ using this module in a drone. However, the variability among the individual sensors is high, stressing that the calibration process is complex and it has not been done.

The calibration process of these LCS is a challenge as mentioned before, where ML and Deep Learning (DL) models can be used. In (Zimmerman et al. (2018)), the authors show calibration models (using 16 weeks data) to improve sensor performance, highlighting that RF approach is more robust since it accounts for pollutant cross-sensitivities. Using specific LCS (RAMP system), they achieve an MRE of 15% for O3. In the study performed by (Johnson et al. (2018)), the calibration of an aerosol sensor for PM2.5 is carried out by comparing simple linear regression models with GB using the PPD42 PM sensor (Shinyei (2024)). The study concludes that gradient boosting performed better and significantly improved the performance of the sensors, reaching a coefficient of determination (R^2) of up to 0.76. In (Casey et al. (2019)), the authors show that Neural Networks (NN) generally outperform lineal models to quantify O3, CO, CO2, and CH4 in ambient air, using gas sensors integrated into

U-Pod air quality monitors. Besides, they highlight that NN capture the complex nonlinear interactions among multiple gas sensors, considering factors such as Temp, RH and atmospheric chemistry. In (Borrego et al. (2016)), the authors carry out a performance evaluation during two-weeks data of the calibration process at Aveiro (Portugal) of different LCS showing different models. In particular for O3, the best two models get R^2 (and MAE) (in ppb) of 0.77 (7.66), 0.7 (2.4), and estimation MRE between 10 and 5%. Also, in (Esposito et al. (2016)), the authors use dynamic NN for calibration achieving models with R^2 (MAE) (in ppb) of 0.69 (7.45), with a MRE of 49%.

In this context, when using AI techniques on environmental research, it is important to follow the recommendations given by (Zhu et al. (2023)) based on a review of more than 148 highly cited research papers. In this reference, it is highlighted that data preprocessing, analysis and interpretability are often overlooked, such as Feature Importance Analysis (FIA), Principal Component Analysis (PCA) and Feature Selection (FS). In addition, it is said that the process of optimizing algorithms through the selection of their hyperparameters (Hyperparameter Optimization (HPO)) is neglected in most of the environmental research studies considered. For instance, in (Johnson et al. (2018)), better results are obtained with GB, but HPO is not performed in the model, which could allow further improvements of the results. Both (Malings et al. (2019)) and (Borrego et al. (2016)) take into account some aspects related to the data analysis focused on the optimization of the problem, but they do not carry out a HPO. In (Esposito et al. (2016)), the authors carry out a kind of simple HPO, based on raw tests of different architectures and modifying hyperparameters, such as the number of hidden layers of the model, tapped delay length and feedback delay line length, concluding that a dynamic approach to these parameters improves the results with respect to a static approach without changing the value of these parameters.

Regarding the selection of parameters, in (Johnson et al. (2018)), the authors does not perform an analysis using techniques such as the aforementioned FIA and FS, but a sensitivity analysis using different meteorological variables (such as Temp and RH), determining that it is useful information for GB. In (Malings et al. (2019)), the quantification of the importance of the model variables is mentioned as a mean to understand which information is useful, concluding that for RF, to add additional information apart from AQ measurements, such as Temp and RH are very helpful. In (Borrego et al. (2016)) and (Esposito et al. (2016)), the authors do not include a specific analysis of the relative importance of different variables or features. However, a good example of FS is depicted in (Okafor et al. (2020)), where it is shown that identifying the environmental factors affecting LCS is crucial for improving data quality using data fusion and ML. These factors are then incorporated into the development of the calibration model.

In conclusion, in order to increase the resolution of city-scale AQ monitoring according to the recommendations given by (Directive 2008/50/EC (2008)) as mentioned before, it is necessary to perform a calibration process of these LCS. In this scenario, we focus on the O3 calibration by using ensemble ML techniques, comparing different techniques. For this purpose, it is necessary to carry out a thorough data treatment with a good practice criteria (Zhu et al. (2023)) including HPO, FIA and/or FS, which are usually overlooked. In this case, we achieve interesting results compared with the related work, as shown in Section 4.

3 Building the dataset and using Machine Learning algorithms

In this section we explain the process to gather AQ monitoring information from a prototyped low cost Internet of Things (IoT) node based on the ZPHS01B AQ module, how it is deployed and how the dataset is built to apply ML techniques for calibration purpose.

3.1 Building the dataset

Figure 1. Detail of the official AQ monitoring station and the AQ node with a ZPHS01B module located at Bulevar Sur (Valencia, Spain).

To calibrate the O3 sensor from the ZPHS01B module, we require a dataset (dataset-1) to train various ML models. For this purpose, we use as reference values, O3 concentration readings from the official AQ station in the Valencian AQ Monitoring Network (VAQMN), at Bulevar Sur (Valencia, Spain) managed by Generalitat Valenciana (GVA) with latitude and longitude 39.450389 and -0.396324, respectively, as shown in Figure 1. These reference values are given in $\mu g/m^3$ periodically averaging every 10-minutes. The AQ station data is retrieved from *https://rvvcca.gva.es/estatico/46250050*. The ZPHS01B module's readings are taken at a rate of 10 samples per minute, one sample every 6 seconds. Notice that, as a first approach, creating a dataset with different locations is not recommended, as it could alter environmental conditions and interfere with the training process.

Table 3 presents the structure and main statistics of the dataset. The units used for O3 concentration from the official station are in $\mu g/m^3$, meanwhile in the ZPHS01B module are in ppm. Both units are typically used in a formal and academic context, but we need to standardize them. The formula used to carry out this conversion is: "Concentration ($\mu g/m^3$) = Concentration (ppm) x 1000 x molecular mass" (Breeze Technologies (2024)).

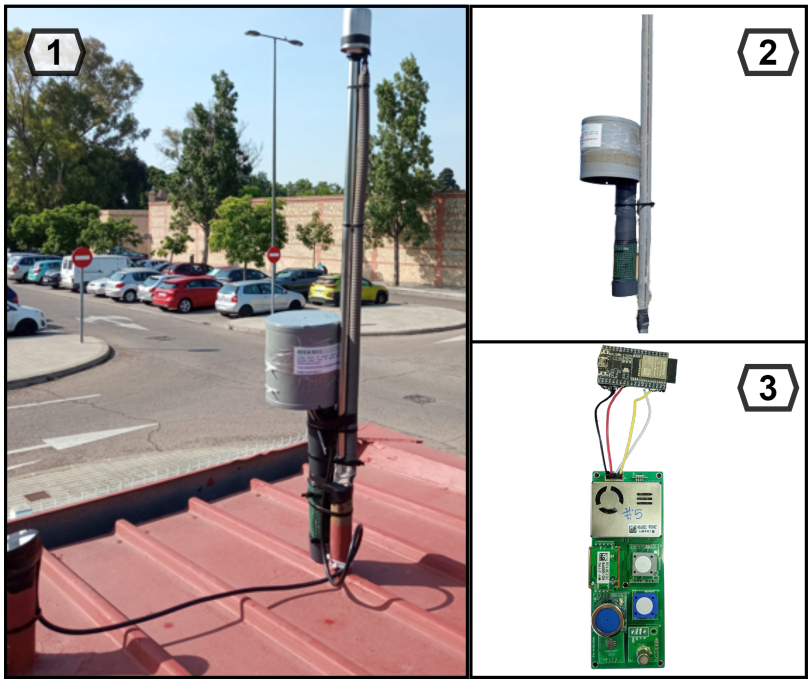

Figure 2. (1) AQ IoT node; (2) deployment detail; (3) hardware detail

In Figure 2, it is shown the IoT node (and its housing) that keeps the ZPHS01B module within a PVC pipe with a small fan at the top, to ensure air circulation. In the head of this node, it is placed the microcontroller that sends data via the LTE-M communications.

In addition, in order to test the proposed models in this paper and their generalization in Section 4, we have used another dataset (dataset-2) with two different AQ IoT nodes (Node 1 and 2), from the official AQ monitoring station called *Moli del Sol* (Valencia, Spain) with latitude and longitude 39.48113875, -0.40855865. This station is 4.1 km away from the previous one. Its data is retrieved from *https://rvvcca.gva.es/estatico/46250048*. In this case, this dataset is from May 31, 2024 till January 25, 2025, with 239 days. Now on, we will refer always to dataset-1 as the dataset, except in Section 4 where we generalize the models with dataset-2.

3.2 Analyzing the dataset

The initial data collection is based on 6-second frequency samples, including 165 days (approximately five and a half months), from June 8^{th} 2023 till November 20^{th} 2023. Based on this collection, three datasets have been created by averaging data over different time monitoring intervals: 10 min., 30 min. and 1 h. with 23496, 7843 and 3922 samples respectively. The lowest 10 min. interval is given by the official AQ station and 30 min and 1 h are common time base for AQ parameters. Although they are not large data-set, it is sufficient as shown in (Zhu et al. (2023)), due to the relationship (ratio) between sample size and

feature size, 4 in total as seen next. This ratio is called, Sample-size to Feature-size Ratio (SFR), being recommended a SFR higher than 500. More detail is given in Section 3.3.

Initially, the datasets were cleaned of invalid data. Notice that from the readings of the official AQ station, we had 275 Not a Number (NaN) during this period, that in our case were replaced using the quadratic interpolation method, since experimentally it gave better results and made the interpolation closer to the ozone signal. This explanation to prepare the dataset, also known as Missing Data Management (MDM), is recommended according to (Zhu et al. (2023)).

Table 3. Summary of main statistics of the Dataset: Minimum (Min.), Maximum (Max.), Mean (Mean), Standard Deviation, Variance (Var.), Median Absolute Deviation (MAD), percentage of samples taking Different values (Diff.), Stationarity (Stat.), Seasonality (Seas.) and High correlation (High corr.)

	Temp	**RH**	**PM2.5**	**CO2**	**NO2**	**CO**	**CH2O**	**TVOC**	**O3**	**O3ref**
Min	5.24	62.29	21.25	693.43	0.78	0	0.005	0	39.57	8.71
Max	42.26	118	83.69	1792.50	18.81	0.75	1.21	2.95	255.76	97.85
Mean	20.60	91.31	49.99	780.33	15.27	0.34	0.021	0.024	114.39	55.72
SD	5.70	18.12	18.14	57.16	5.65	0.28	0.02	0.13	67.11	24.83
Var.	32.57	328.41	329.34	3268.29	31.92	0.08	0.0006	0.016	4503.98	616.69
MAD	3.92	16.37	13.31	24.53	0.59	0	0.001	0	51.40	16.21
Diff.	99.1%	81.9%	87.9%	97.5%	50.6%	0.2%	81.2%	5.8%	75.0%	30.3%
Stat.	not	not	not	not	not	not	yes	yes	not	not
Seas.	yes	yes	yes	yes	yes	yes	not	not	yes	yes
High corr.	yes	yes	yes	yes	not	yes	not	not	yes	yes

Table 3 shows a summary of main statistics of the dataset. For each parameter is shown: the Minimum value (Min.), Maximum value (Max.), Mean value of all entries (Mean), Standard Deviation, Variance (Var.), Median Absolute Deviation (MAD), percentage of samples taking Different values (Diff.), Stationarity (Stat.), Seasonality (Seas.) and High correlation (High corr.) with others. Seasonality refers to recurring patterns at regular intervals, while stationarity indicates constant statistical properties over time.

From these results, it is worth mentioning that the CH2O, CO, TVOC and NO2 sensors do not seem to be working properly in the ZPHS01B module. CH2O, CO and TVOC are almost always stuck to values close to zero, seeming not to excite at normal concentrations, with very low variability. On the other hand, the NO2 sensor appears saturated, stuck at the maximum

value, 10 ppm. Thus in practice, the number of used features from Table 3 are 5, that is from the initial 9 (the reference is not included), we remove these 4 (CH2O, CO, TVOC, and NO2). Also, RH sensor has a positive offset as we can see from the maximum value, 118%.

To characterize the measurements of O3, we carry out a Discrete Fourier Transform (DFT) analysis, to see the changing patterns. The DFT is a mathematical technique that transforms a discrete signal from the time domain to the frequency domain. Figure 3 shows the peaks obtained from the O3 signal. There are two main peaks and their harmonics. The first peak appears in the frequency $f = 0.00025 \frac{1}{hour}$ which corresponds to a period of 4000 hours, 5.56 months, that is the total duration of data-set. The second peak indicates and reveals a relevant frequency component at $f = 0.04182251 \frac{1}{hour}$, which represents a period of 23.91 hours (approximately 1 day). Thus, there is an O3 pattern that it is repeated every day, as it could be expected in a city, based on how it is generated from road traffic by combustion engines as discussed in Section 1.

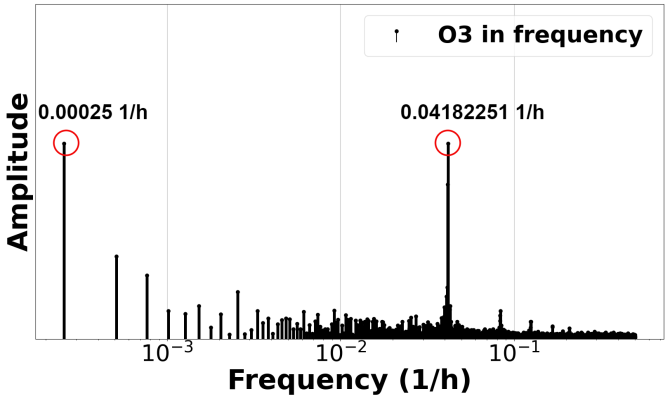

Figure 3. DFT of O3 readings from LCS

Figure 4 shows the O3 readings in $\mu g/m^3$ from the LCS and the official station (reference) for one week. It can be seen that there is an offset in the LCS readings over the ones from the reference. Also, it is clear how the O3 LCS captures the trends, useful information for the ML models.

3.3 Feature Importance Analysis and Selection

FIA and FS play crucial roles in ML models, especially in environmental research, by helping to preserve essential features (variables), reduce noise and enhance model efficiency, particularly relevant when dealing with a small set of samples or large numbers of variables (Zhu et al. (2023)).

Table 4 shows the normalized output of the FIA using the *scikit-learn* library (Pedregosa et al. (2011)), for the parameters complementary to O3, for each ML models used. In order to determine the most useful parameters for the models, a threshold is established in 8% of importance. Notice that the pattern of parameters with the highest importance, is repeated for all models. From this analysis, we conclude that Temp, RH and CO2 are the most relevant and then will be considered for the next step (FS analysis), since they show the highest values.

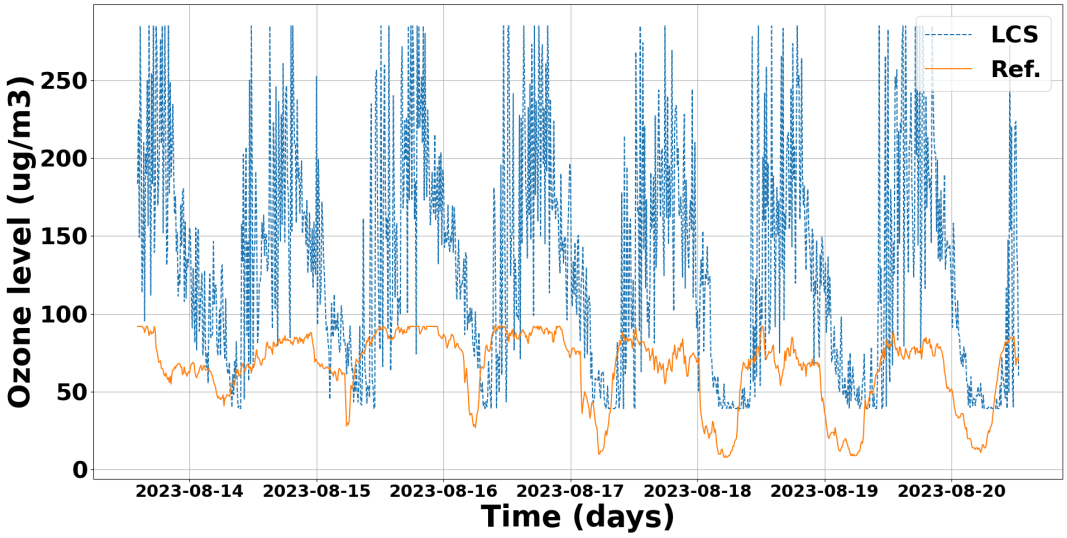

Figure 4. O3 readings in $\mu g/m^3$ from the LCS and Reference for one week

Table 4. FIA of ozone's complementary parameters for Random Forest (RF), Gradient Boosting (GB), Adaptive Boost (ADA) and Decision Tree (DT)

Model	Temp	RH	PM2.5	CO2	NO2	O3ref	CO	TVOC	CH2O
RF	0.128	0.103	0.069	0.222	0.078	0.269	0.002	0.003	0.064
GB	0.107	0.105	0.052	0.211	0.057	0.253	0.001	0.001	0.068
ADA	0.119	0.097	0.064	0.246	0.067	0.287	0.001	0.001	0.066
DT	0.115	0.088	0.070	0.232	0.061	0.276	0.001	0.002	0.061

With regard to FS, Figure 5 shows the correlation matrix among these variables. There is a high correlation among all PMx readings because all of them are calculated directly from the PM2.5 (Zhengzhou Winsen Electronics Technology Co. (2024)). Also, from this analysis, we stress that Temp and RH, are the best correlated with the rest of variables, as well as O3 LCS, O3 reference, PM2.5 and CO2, but these ones with a lower correlation. This information is very valuable to train the ML models.

3.4 Applying Machine Learning algorithms

As mentioned before in environmental research, the use of ML algorithms, in particular ensemble models, has increased significantly compared to DL (Zimmerman et al. (2018)). Some of the most popular ensemble algorithms are RF or GB related models (Obregon and Jung (2022)). Furthermore, based on our experience, we recognize that in AQ monitoring scenarios using

Figure 5. LCS readings and O3 reference correlation matrix

LCS such as the ZPHS01B module, datasets are often limited and constrained, which affects the use of DL techniques, as they usually tend to overfit.

This paper evaluates these ensemble ML algorithms: RF, GB and ADA algorithms, implemented in the *scikit-learn* (Pedregosa et al. (2011)) library (in the ensemble submodule), that offers efficient solutions for time series regression problems as this one. These evaluated methods exhibit the ability to handle non-linear relationships and adapt to changing patterns over time. In addition, the DT model, belonging to the tree submodule of *scikit-learn*, is also evaluated, since it is a common base of this type of ensemble algorithms.

To optimize these models as indicated in (Zhu et al. (2023)), there are different techniques and tools in order to carry out the HPO, being *GridSearch* (Pedregosa et al. (2011)) the most commonly used method to obtain a good configuration for these algorithms. *GridSearch* in *scikit-learn* is a hyperparameter tuning technique that exhaustively searches through a user-defined

hyperparameter space to find the optimal combination for a ML model. These hyperparameters are external specific model configurations settings. It systematically evaluates the model's performance across all possible user-defined hyperparameters using cross-validation, aiming to identify the configuration that maximizes estimation accuracy or minimizes a specified loss function. We choose this method due to its higher flexibility compared to other tools such as *RandomSearch* (Pedregosa et al. (2011)) that has a more random approach.

Next, we discuss the different supervised ML algorithms used and the selection of the different hyperparameters taking into account the best results of R^2, Root Mean Square Error (RMSE) and MAE.

3.4.1 Random Forest (RF)

Table 5. RF hyperparameters evaluated on *GridSearch* showing in **bold** the combination that gives the best results in terms of R^2, RMSE and MAE.

No. of estimators	**Max. depth**	**Max. features**
50, **100**, 250, 500, 900	2, 5, 7, **None**	sqrt, log2, 0.1, 0.3, 0.5, **1.0**

RF is an ensemble algorithm that relies on constructing multiple DT during training. Each tree is trained on a random subset of the dataset, and the final predictions are obtained by averaging the individual predictions for all of them. This "forest" approach helps to mitigate overfitting and improves the model's generalization. Furthermore, introducing randomness in the selection of features and samples during tree construction contributes to a more robust and accurate model for regression tasks. Table 5 shows the hyperparameters evaluated, in bold the best option. The *number of estimators* refers to the number of trees in the forest, while the maximum depth refers to the *maximum depth* of the tree. The *maximum features* variable determines the upper limit on the number of features to consider when splitting a tree into two child nodes during the tree construction process. Note that as the *number of estimators* does not have a significant role in this use case, we use the default value, 100.

3.4.2 Gradient Boosting (GB)

Table 6. GB hyperparameters evaluated on *GridSearch* showing in **bold** the combination that gives the best results in terms of R^2, RMSE and MAE.

No. of estimators	**Max. depth**	**Max. features**
50, 100, 250, 500, **900**	2, 5, 7, **None**	sqrt, log2, 0.1, 0.3, 0.5, **1.0**
Learning rate	**Subsample**	**Loss**
0.01, 0.05, **0.1**, 0.3	0.5, 0.8, **1.0**	**squared err.**, absolute err., huber

GB is an ensemble algorithm based on the iterative construction of weak DTs, which are sequentially aggregated to enhance the predictive capability of the model. In each iteration, it focuses on correcting the residual errors of the existing model by fitting a new DT to capture the deficiencies of the current model. The weighting of individual trees is determined by a learning rate, and the final output of the model is the weighted sum of predictions from all these trees. This gradual building process and the ability to handle nonlinear relationships in the data make GB effective for regression tasks. Table 6 shows the hyperparameters evaluated, in bold the best option. In addition to the previous hyperparameters, in this case, the *loss* hyperparameter refers to the loss function to be optimized, while *learning rate* reduces the contribution of each tree according to the value of the variable. The *subsample* hyperparameter represents the fraction of samples that will be used to adjust the individual base learners and if it is less than 1.0, it results in Stochastic Gradient Boosting (SGB).

3.4.3 Adaptive Boosting (ADA)

Table 7. ADA hyperparameters evaluated on *GridSearch* showing in **bold** the combination that gives the best results in terms of R^2, RMSE and MAE.

No. of estimators	Learning rate	Loss
50, 100, 250, 500, 900	**0.01**, 0.05, 0.1, 0.3	linear, square, **exponential**

ADA is an ensemble algorithm, that its primary goal is to improve the predictive accuracy by combining multiple weak regression models. When training, ADA assigns weights to data instances, giving more emphasis to observations that were poorly predicted in previous iterations. Its construction involves the sequential aggregation of regression models, each fitted to correct errors from the existing combined model. The final model is a weighted combination of individual predictions from the base models. ADA is particularly effective in enhancing generalization capability and reducing overfitting in regression tasks. Table 7 shows the hyperparameters evaluated, in bold the best option. In this model, there is a key concept to run the optimization process related to the *estimator* variable, that by default is an instance of type *DecisionTreeRegressor*, initialized with a maximum depth value of three. If the value of this hyperparameter is not modified, this model is largely constrained. Also, notice that as the number of estimators does not have a significant role on this use case, we use the default value of 50 estimators. The other hyperparameters have the same meaning in this model.

3.4.4 Decision Tree (DT)

DT is an algorithm that recursively partitions the dataset based on features, aiming to create a hierarchical structure of decision nodes to make predictions. Table 8 shows the hyperparameters evaluated, in bold the best option. The *splitter* hyperparameter indicates which strategy is used to perform the splitting at each node.

Table 8. DT hyperparameters evaluated on *GridSearch* showing in **bold** the combination that gives the best results in terms of R^2, RMSE and MAE.

Max. depth	Max. features	Splitter
2, 5, 7, **None**	sqrt, log2, 0.1, 0.3, 0.5, **1.0**	**best**, random

4 Results

We evaluated the performance metrics of these ML models under different configurations (in terms of R^2, RMSE, MAE in $\mu g/m^3$ and Mean Absolute Percentage Error (MAPE) and execution time in seconds), both with default and optimized hyperparameters, taking into account the three different datasets given by different monitoring intervals: 10 and 30 min and 1 h, as depicted in Section 3.2. We tested different training-test ratio percentages from these datasets: 60%-40%, 70%-30%, 80%-20% and 90%-10%, denoted as 60/40, 70/30, 80/20 and 90/10. From all of them, we have achieved the best results in terms of these performance metrics with 90/10 training-test ratio with a monitoring interval of 10 min, as shown in Tables 9 and 10. Besides, from the analysis carried out in Section 3.3, for the feature selection, we proceed in this section with the features that provide also the best results, based on [date, O3, Temp, RH]. We see that fewer features, better results, i.e. increasing the SFR. Then, other dimensionality reduction techniques are not required. If we add more features that are not so significant, it makes the dataset poorer.

Notice that the performance metrics shown in Tables 9, 10 and 11 are the weighted average of each metric over 100 different iterations by changing the content of the training and test set to obtain results with the minimum bias as possible.

Table 9. Performance metrics with the default models and 90/10 (training/testing) ratio dataset

	$\mathbf{GB_{default}}$	$\mathbf{RF_{default}}$	$\mathbf{ADA_{default}}$	$\mathbf{DT_{default}}$
$\mathbf{R^2}$	0.658	0.927	0.326	0.878
RMSE	15.341	7.092	21.543	9.158
MAE	11.604	4.179	18.123	4.691
MAPE	0.827	0.208	1.239	0.206
Time	3.829	17.187	0.342	0.213

To see the effect of HPO, Table 9 shows the results for the default ML models, while in Table 10, they are improved with HPO. Using HPO, we optimize the calibration process (higher R^2 and lower errors), but we increase significantly the execution time. In particular, the best option is given by GB with a R^2 of 0.938, RMSE of 6.492 and an execution time of 66.937 s, followed by ADA with much less execution time, 7.805 s as shown in Table 10. Notice that the execution time required for the training process is influenced by the hyperparameter optimization, in particular for GB as shown in these tables. Also we

Table 10. Performance metrics with HPO models and 90/10 (training/testing) ratio dataset

	$GB_{optimized}$	$RF_{optimized}$	$ADA_{optimized}$	$DT_{optimized}$
R^2	0.938	0.927	0.922	0.878
RMSE	6.492	7.093	7.289	9.149
MAE	4.022	4.185	3.642	4.684
MAPE	0.194	0.208	0.160	0.206
Time	66.937	18.316	7.805	0.212

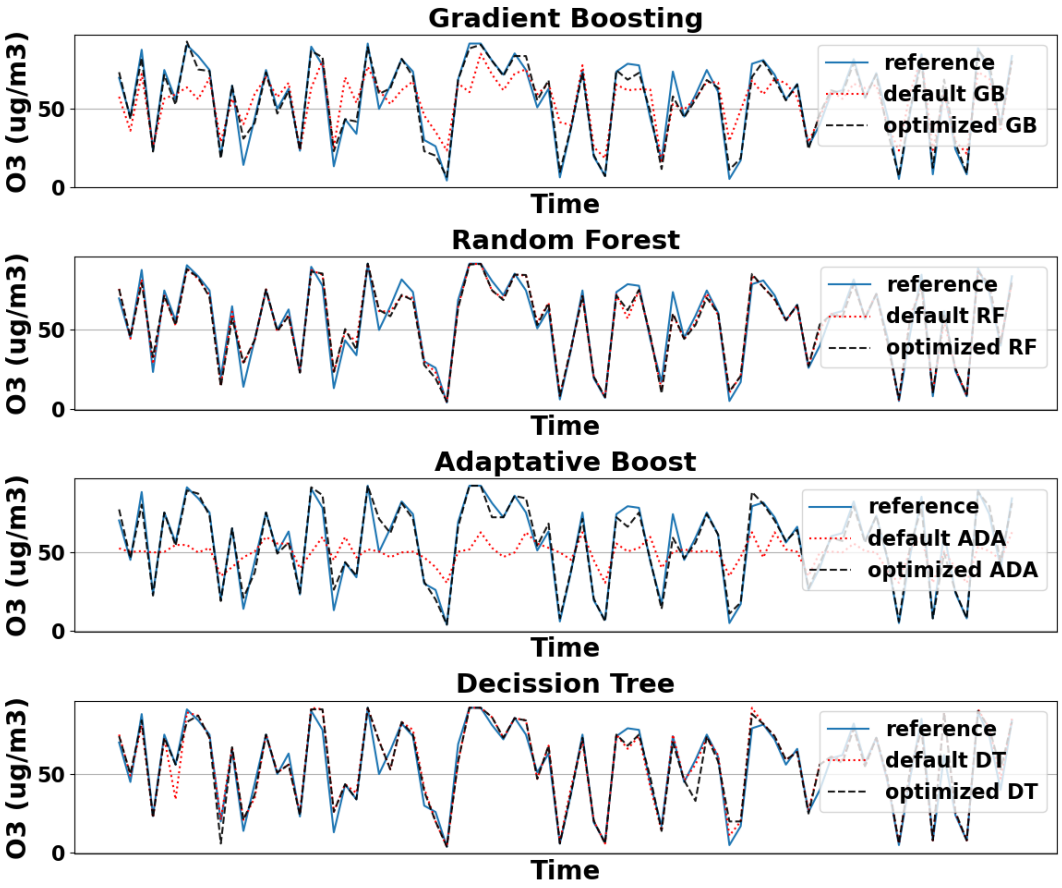

Figure 6. Ozone calibration done with default and optimized models with 90/10 (training/testing) ratio dataset

highlight that RF and DT are already well optimized and their execution times do not increase in comparison between the

default and the optimized versions. In Figure 6 it is shown the calibration process for both the default and HPO models vs O3 reference given by the different algorithms.

Table 11. Performance metrics with HPO models and 80/20 (training/testing) ratio dataset

	$\mathbf{GB_{optimized}}$	$\mathbf{RF_{optimized}}$	$\mathbf{ADA_{optimized}}$	$\mathbf{DT_{optimized}}$
$\mathbf{R^2}$	0.936	0.924	0.920	0.863
RMSE	6.664	7.253	7.416	9.735
MAE	4.221	4.415	3.833	5.104
MAPE	0.206	0.228	0.175	0.226
Time	61.054	16.618	7.078	0.194

However, it is common to use a 80/20 training/test ratio (Zhu et al. (2023)). For this ratio, Table 11 shows the results with the optimized models by HPO, where the GB model is the best one again, as it happened with previous 90/10 ratio.

A summary of these metrics (R^2 and errors) for the GB model, with different monitoring intervals and different training/test ratio percentages are shown in Figure 7. We can see how increasing the training %, the trend is to improve the accuracy of the model (R^2 getting closer to 1) and to reduce lightly the errors but increasing the training time, as it could be expected. Similar behaviors are shown by the other models, in particular with ADA model. Regarding overfitting, Figure 7 shows that the error difference between using 90% and 60% of the data for training (the maximum and minimum percentages, respectively) is approximately 2% in the worst-case scenario (1-hour dataset). This suggests that overfitting is not significant in the proposed model.

In Figure 8, it is shown the distribution error for the different models, with detail of raw, default and optimized versions. The number of samples are normalized in the Y-axis. It is appreciated with GB and ADA that their distribution errors are concentrated around zero when calibration is applied, and even more when using the HPO optimized models. This behavior is also appreciated with DT, but with lower intensity. However, RF keeps a pretty similar distribution in both versions, default and optimized, as we saw in Tables 9 and 10. Thus, in terms of error distribution, HPO significantly concentrates the error around 0 for the GB and ADA models, while practically no change is observed for the DT and RF models.

The Standard Deviations (SD) and the Confidence Intervals (CI) in $\mu g/m^3$ are shown in Table 12. This information is obtained from the error distribution statistics given in Figure 8. It can be seen how the GB adjusts better compared with the other models from the error distribution analysis. It is observed that the SD is similar to the RMSE and this is due to the fact that, as shown in this figure, the distribution is almost Gaussian.

In terms of generalization as mentioned in Section 3.1, we have checked the same proposed models with dataset-2 under the same conditions, with 90/10 (training/testing) ratio. In Table 13, we summarize the metrics given by the best model based on GB for dataset-1 and for Node 1 and 2 from dataset-2 respectively. In particular, if we focus on MAE, we see that Node 2

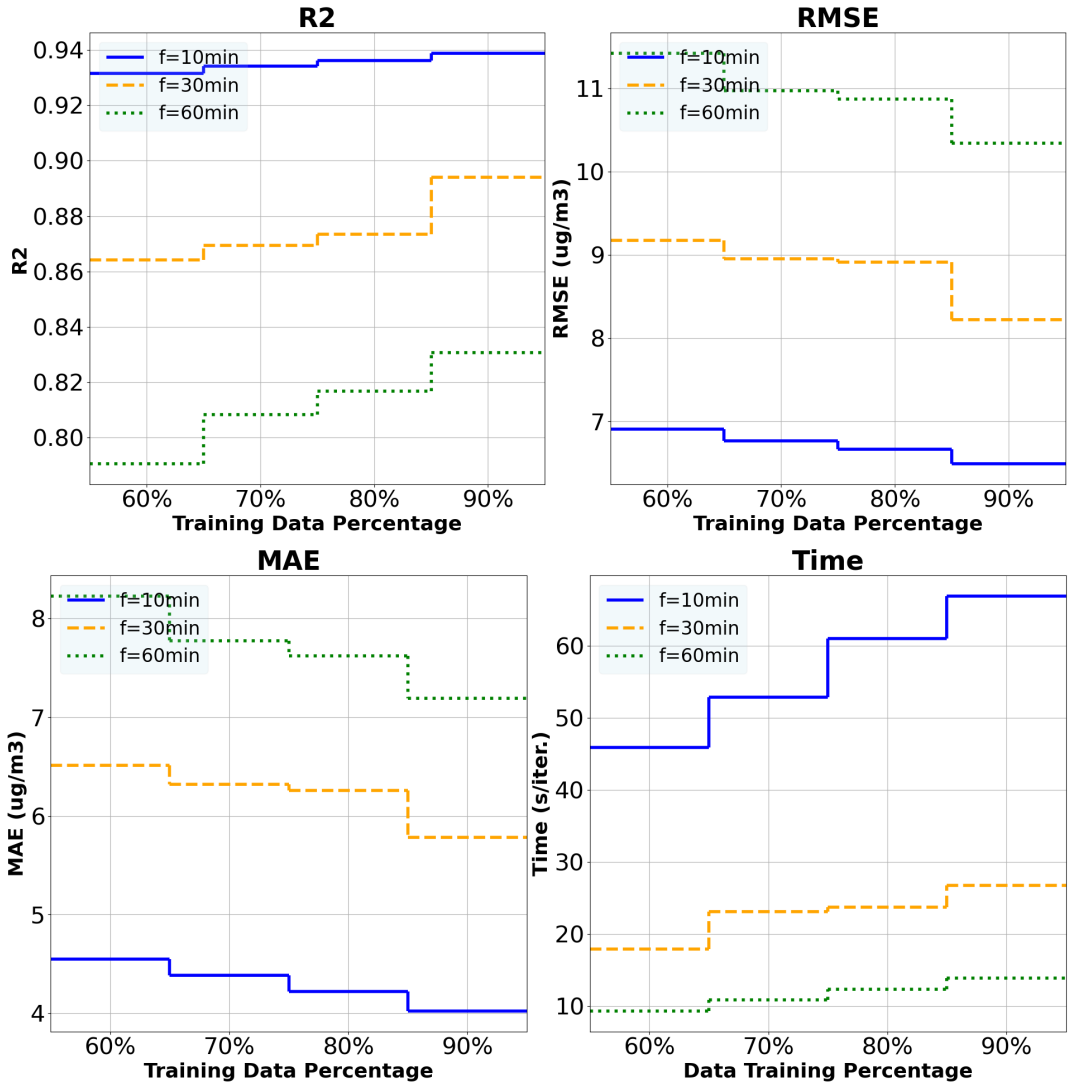

Figure 7. Ozone estimation analysis for GB default and optimized models with different % training datasets and monitoring intervals

performs slightly better than Node 1 in dataset-2, likely due to manufacturing variations associated with their low cost, as well as the results from dataset-1 are between these two, validating its generalized behavior.

Finally, in Table 14, we show the improvement in % using the different ML models for the calibration process from the LCS raw readings of the module, highlighting the better performance of GB model compared to the other models.

In Table 15, we compare our models for O3 calibration for LCS, against the related work with a similar approach. We must stress that the starting point is slightly different compared to ours, since these studies have used more reliable and expensive LCS, approximately ten times more expensive that the ZPHS01B module. As mentioned before, our model reduces the estimation error up to 94.05% from raw readings based on MAE measurements, with a MRE of 7.21% (given by MAE

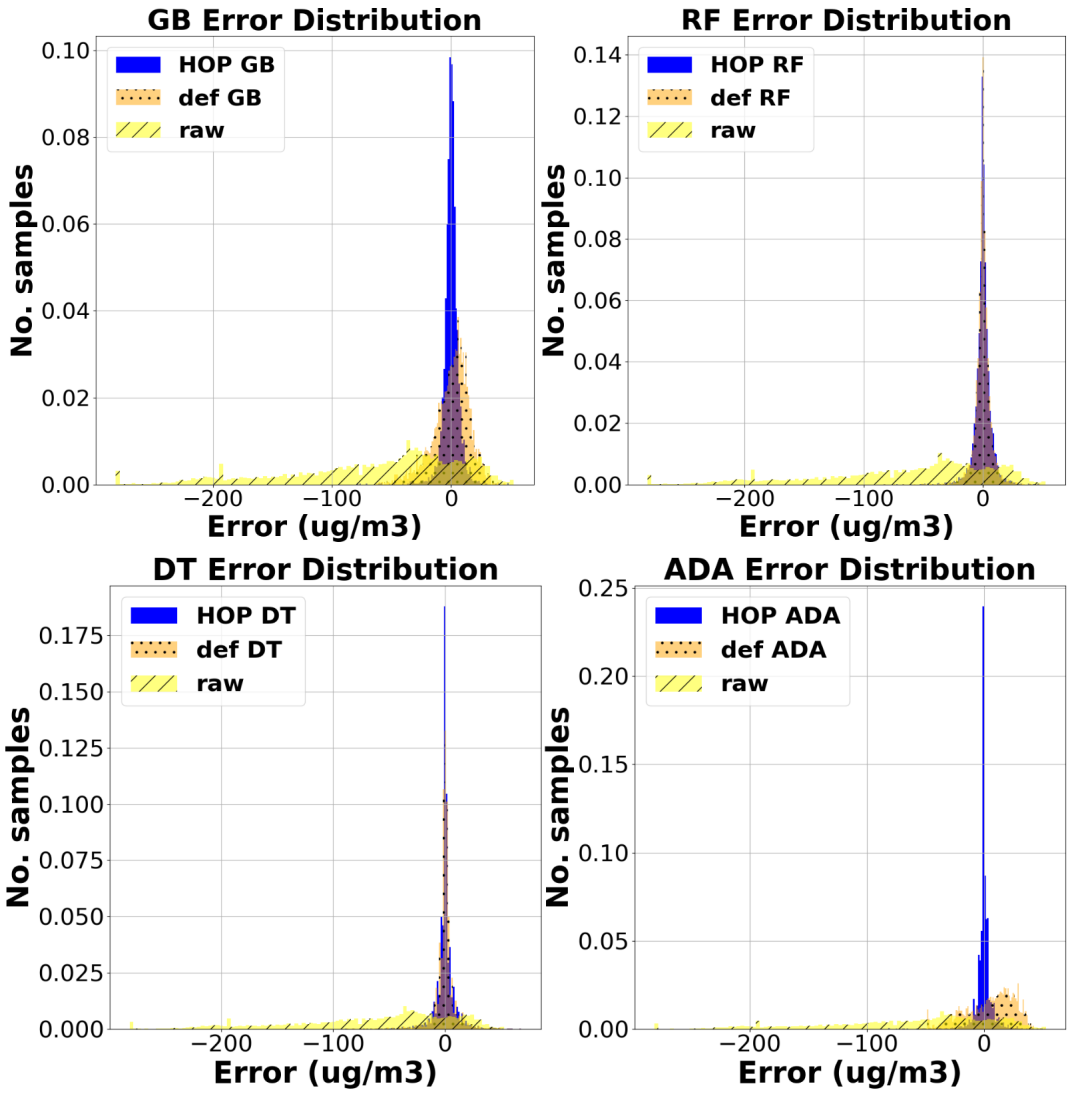

Figure 8. Distribution error for the different models, with detail of raw, default and optimized versions

4.022 with 90/10 dataset and with O3 mean value of 55.72 $\mu g/m^3$ as shown in Table 3, using GB with only 4 features, as shown in Section 3.3.

5 Conclusions

This paper focuses on ground-level ozone (O3), as it serves as an indicator of other pollution levels in urban areas using LCS nodes based on the ZPHS01B module. These nodes will permit to increase the spatial sampling of AQ. Given their low accuracy, we employed ML methods after thorough analysis, particularly DT and the ensemble algorithms (GB, RF and ADA),

Table 12. Standard Deviation (σ) and Confidence Interval (CI) for the error estimation with raw, default and optimized models

Data	σ
Raw	72.53
$GB_{default}$	15.39
$GB_{optimized}$	6.74
$RF_{default}$	7.25
$RF_{optimized}$	7.23
$ADA_{default}$	21.29
$ADA_{optimized}$	7.50
$DT_{default}$	9.73
$DT_{optimized}$	9.65

Data	CI
Raw	[-65.05,-60.91]
$GB_{default}$	[-0.17,0.7]
$GB_{optimized}$	[-0.11,0.27]
$RF_{default}$	[-0.1,0.3]
$RF_{optimized}$	[-0.05,0.36]
$ADA_{default}$	[3.91,5.12]
$ADA_{optimized}$	[-0.38,0.05]
$DT_{default}$	[-0.26,0.29]
$DT_{optimized}$	[-0.31,0.24]

Table 13. Generalization test with dataset-1 and dataset-2 (Node 1 and 2) using $GB_{optimized}$ algorithm with 90/10 (training/testing) ratio.

	Dataset-1	Dataset-2 (Node 1)	Dataset-2 (Node 2)
$\mathbf{R^2}$	0.938	0.940	0.954
RMSE	6.492	6.107	5.332
MAE	4.022	4.336	3.741
MAPE	0.194	0.167	0.141

Table 14. Improvement (in%) of O3 calibration from the raw readings with the different optimized models.

	GB	RF	ADA	DT
$\mathbf{R^2}$	258.13%	256.27%	256.1%	246.27%
RMSE	93.05%	92.43%	92.29%	89.85%
MAE	94.05%	93.82%	94.59%	92.79%
MAPE	62.75%	58.8%	68.35%	59.12%

taking into account additional environmental information, reducing the estimation error in around 94.05% from the LCS raw

Table 15. Comparison with the similar related work

Study	Location	Sensor	R^2	MRE [%]
(Borrego et al. (2016))	Aveiro (PT)	many	0.70-0.77	10-5%
(Zimmerman et al. (2018))	Pittsburg (USA)	RAMP	0.86	15%
(Esposito et al. (2016))	Cambridge (UK)	SnaQ	0.69	42%
Our model	Valencia	ZPHS01B	0.938	7.21%

readings with a MRE of 7.21% using GB, and more than 89% in the other models, outperforming the related work. Thus, the raw readings from this O3 LCS, after the proposed calibration process are then adjusted with higher accuracy.

In particular, we have used a data set of 165 days, with different monitoring intervals, giving the best results when we use 10 min monitoring interval, as it could be expected. If we use higher monitoring intervals (30 min or 1 hour), we see that we start losing details, smoothing the dataset and overlooking different behaviors that in the ML process helps to reduce the prediction error. For the training process, we have carried out several techniques (FIA and FS) in order to select the most relevant features, applying HPO within the different models, with different percentages for training and testing.

Besides, we checked that for the ZPHS01B module and O3 calibration, 165 days of dataset-1 provided sufficient information to generalize the proposed models comparing with a dataset-2 of 239 days. This aligns with the SFR recommended values according to (Zhu et al. (2023). Thus, given the features and characteristics of this module, the original dataset (165 days) contains enough information to generalize the behavior of the O3 sensor and their response.

As future work, we plan to expand the dataset and include complementary parameters, such as wind speed or road traffic density, to increase the accuracy of these models. In addition, we focus on the design of new calibration and forecasting algorithms for the different sensors embedded in the low-cost ZPHS01B module in order to improve AQ monitoring resolution.

Data availability. Please feel free to contact to the authors for further information: http://www.uv.es/eco4rupa/dataset.html

Author contributions. G.M.F and S.F.C contributed equally to air quality gathering process and calibration techniques, as well as coding and manuscript writing. E.M.A prepared the dataset. J.J.P.S prepared the hardware infrastructure. S.F.C and J.S.G managed the funding and external collaborations.

Competing interests. No competing interests are present.

Acknowledgements. This paper is partially funded by the Grant PID2021-126823OB- I00 MCIN funded by MCIN/AEI/ 10.13039/ 501100011033 and by the European Union NextGeneration EU/ PRTR; by the Generalitat Valenciana with grant references CIAICO/2022/ 179, CIACIF/2023/ 416 and CIAEST/2022/ 64 as well as the Spanish Ministry of Education in the call for Senior Professors and Researchers to stay in foreign centers for the grant with reference PRX23/00589.

We are grateful with the Generalitat Valenciana and its AQ monitoring networm, in particular with Rafael Orts Bargues from the Atmospheric Protection Service

References

Antonenko, A., Boretskij, V., and Zagaria, O.: Classification of Indoor Air Pollution Using Low-cost Sensors by Machine Learning, in: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, pp. EGU–14 856, https://doi.org/10.5194/egusphere-egu23-14856, 2023.

Borrego, C., Costa, A., Ginja, J., Amorim, M., Coutinho, M., Karatzas, K., Sioumis, T., Katsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Smith, P., André, N., Gérard, P., Francis, L., Castell, N., Schneider, P., Viana, M., Minguillón, M., Reimringer, W., Otjes, R., von Sicard, O., Pohle, R., Elen, B., Suriano, D., Pfister, V., Prato, M., Dipinto, S., and Penza, M.: Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise, Atmospheric Environment, 147, 246–263, https://doi.org/https://doi.org/10.1016/j.atmosenv.2016.09.050, 2016.

Breeze Technologies: Air pollution – How to convert between mg/m^3, μg/m^3 and ppm, ppb, https://www.breeze-technologies.de/blog/air-pollution-how-to-convert-between-mgm3-Âμgm3-ppm-ppb/, 2024.

Casey, J. G., Collier-Oxandale, A., and Hannigan, M.: Performance of artificial neural networks and linear models to quantify 4 trace gas species in an oil and gas production region with low-cost sensors, Sensors and Actuators B: Chemical, 283, 504–514, https://doi.org/https://doi.org/10.1016/j.snb.2018.12.049, 2019.

Corp, Z. W. E. T.: Ozone detection module ZE27-03, https://www.winsen-sensor.com/d/files/manual/ze27-o3.pdf, [Accessed 27/11/2024], 2024.

Coto-Fuentes, H., Valdés-Perezgasga, F., Guevara-Amatón, K., Limones-Ríos, K., and Calderón-Ibarra, C.: Integración de estaciones KNARIO con un sistema de información geográfico para el monitoreo de la calidad del aire en la zona metropolitana de La Laguna, Revista Ciencia, Ingeniería y Desarrollo, 1, 109–114, 2022.

DecentLab, Ltd.: Air quality sensor DL-LP8P, https://www.catsensors.com/media/Decentlab/Productos/Decentlab-DL-LP8P-datasheet.pdf, accessed: 27/11/2024, 2024.

Directive 2008/50/EC: of the European Parliament and of the Councils of 21 May 2009 on ambient air quality and cleaner air for Europe., Official Journal of the European Communities, L 152, 1–44, 2008.

Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R., and Popoola, O.: Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems, Sensors and Actuators B: Chemical, 231, 701–713, https://doi.org/https://doi.org/10.1016/j.snb.2016.03.038, 2016.

Felici-Castell, S., Segura-Garcia, J., Perez-Solano, J. J., Fayos-Jordan, R., Soriano-Asensi, A., and Alcaraz-Calero, J. M.: AI-IoT Low-Cost Pollution-Monitoring Sensor Network to Assist Citizens with Respiratory Problems, Sensors, 23, https://doi.org/10.3390/s23239585, 2023.

Garcia, M. A., Villanueva, J., Pardo, N., Perez, I. A., and Sanchez, M. L.: Analysis of ozone concentrations between 2002–2020 in urban air in Northern Spain, Atmosphere, 12, 1495, 2021.

García, M. R., Spinazzé, A., Branco, P. T., Borghi, F., Villena, G., Cattaneo, A., Gilio, A. D., Mihucz, V. G., Álvarez, E. G., Lopes, S. I., Bergmans, B., Orłowski, C., Karatzas, K., Marques, G., Saffell, J., and Sousa, S. I.: Review of low-cost sensors for indoor air quality: Features and applications, Applied Spectroscopy Reviews, 57, 747–779, https://doi.org/10.1080/05704928.2022.2085734, 2022.

H. Adair-Rohani: Air pollution responsible for 6.7 million deaths every year, https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants, accessed: 27/11/2024, 2024.

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, Atmospheric Environment, 184, 9–16, https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.04.019, 2018.

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, https://doi.org/10.3390/atmos10090506, 2019.

Kennedy, Z., Huber, D., Xie, H. R., Sohl, J. E., Page, J., and Dowell, W.: Miniature Multi-Sensor Array (mini-MSA) for Ground-to-Stratosphere Air Measurement, Phase II, Mechanical Engineering Commons, https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1600&context=spacegrant, 2021.

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

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., and Bezirtzoglou, E.: Environmental and Health Impacts of Air Pollution: A Review, Frontiers in Public Health, 8, https://doi.org/10.3389/fpubh.2020.00014, 2020.

Nova Fitness Co., Ltd.: Air quality sensor SDS011, https://cdn-reichelt.de/documents/datenblatt/X200/SDS011-DATASHEET.pdf, accessed: 27/11/2024, 2024.

Obregon, J. and Jung, J.-Y.: Chapter 4 - Explanation of ensemble models, in: Human-Centered Artificial Intelligence, edited by Nam, C. S., Jung, J.-Y., and Lee, S., pp. 51–72, Academic Press, ISBN 978-0-323-85648-5, https://doi.org/https://doi.org/10.1016/B978-0-323-85648-5.00011-6, 2022.

Okafor, N. U., Alghorani, Y., and Delaney, D. T.: Improving Data Quality of Low-cost IoT Sensors in Environmental Monitoring Networks Using Data Fusion and Machine Learning Approach, ICT Express, 6, 220–228, https://doi.org/https://doi.org/10.1016/j.icte.2020.06.004, 2020.

Organization, W. H. et al.: Air Quality Guidelines-Update 2021, Copenhagen, Denmark: WHO Regional Office for Europe, 2021.

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

Seinfeld, J. H. and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, 2016.

Sensit: anatrac.com, https://www.anatrac.com/wp-content/uploads/2021/04/sensit-ramp-brochure.pdf, [Accessed 27-11-2024], 2024.

SGX, SensorTech: Air quality sensor MiCS-6814, https://www.sgxsensortech.com/content/uploads/2015/02/1143_Datasheet-MiCS-6814-rev-8.pdf, accessed: 21/11/2024, 2024.

Shinyei: PPD42 sensor by Shinyei Tech. Co., https://www.shinyei.co.jp/stc/eng/products/optical/ppd42nj.html, [Accessed27/11/2024], 2024.

Vaheed, S., Nayak, P., Rajput, P. S., Snehit, T. U., Kiran, Y. S., and Kumar, L.: Building IoT-Assisted Indoor Air Quality Pollution Monitoring System, in: 2022 7th International Conference on Communication and Electronics Systems (ICCES), pp. 484–489, https://doi.org/10.1109/ICCES54183.2022.9835822, 2022.

Van Poppel, M., Schneider, P., Peters, J., Yatkin, S., Gerboles, M., Matheeussen, C., Bartonova, A., Davila, S., Signorini, M., Vogt, M., Dauge, F., Skaar, J., and Haugen, R.: SensEURCity: A multi-city air quality dataset collected for 2020/2021 using open low-cost sensor systems, Scientific Data, 10, https://doi.org/10.1038/s41597-023-02135-w, 2023.

Zhengzhou Winsen Electronics Technology Co., L.: Multi-in-One Sensor Module (Model: ZPHS01B) Manual, https://www.winsen-sensor.com/d/files/zphs01b-english-version1_1-20200713.pdf, [Accessed 27/11/2024], 2024.

Zhu, J.-J., Yang, M., and Ren, Z. J.: Machine Learning in Environmental Research: Common Pitfalls and Best Practices, Environmental Science & Technology, 57, 17 671–17 689, https://doi.org/10.1021/acs.est.3c00026, pMID: 37384597, 2023.

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, Atmospheric Measurement Techniques, 11, 291–313, https://doi.org/10.5194/amt-11-291-2018, 2018.