https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

Improving raw readings from ozone low cost 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 oxidising 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 serves as an indicator of the activity of other air pollutants and plays a key role in Air Quality monitoring systems in smart cities. To increase its spatial sampling resolution over the city map, ozone low cost sensors are an interesting alternative, but they have a lack of accuracy. In this context, artificial intelligence techniques, in particular ensemble machine learning methods, can improve the raw readings from these sensors taking into account additional environmental information. In this paper, we analyse, propose and compare different techniques, reducing the estimation error in around 94%, achieving the best results using the Gradient Boosting algorithm and outperforming the related work using sensor approximately 10 times less expensive.

1 Introduction

Air Quality (AQ) is a fundamental aspect of environmental health that addresses the composition and purity of gases in the atmosphere, in terms of fine Particulate Matter (PM), Nitrogen Oxides (such as NO, NO2 and totals NOx), Volatile Organic Compounds (VOC) and ground level Ozone (O3). AQ has a direct impact on both human health and the environment Manisalidis et al. (2020).

According to the World Health Organization (WHO), 99% of the world's population breathes air that exceeds the limit values of the recommended safety guidelines H. Adair-Rohani (2024). The WHO also provides global guidelines to help governments and authorities establish and implement policies to protect human health from the adverse effects of air pollution. These guidelines specify recommended levels for these pollutants taking into account both short-term and long-term exposure Organization et al. (2021) and are regularly reviewed and updated to incorporate the latest scientific evidence on the health effects of air pollution.

Among these pollutants, we focus on ground level Ozone, a highly oxidising gaseous pollutant, that has very reactive properties and is harmful at high levels 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 the "primary" pollutants from combustion of fossil materials (such as NO) react with sunlight Seinfeld and Pandis (2016). Thus, its concentration serves as

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

an indicator of the activity of other air pollutants and plays a key role in AQ programs, providing crucial information 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 Directive 2008/50/EC (2008). And Low-Cost Sensor (LCS) are becoming increasingly important, an interesting alternative, but they do not have good accuracy et al (2016).

In this context, Artificial Intelligence (AI) techniques, in particular ensemble Machine Learning (ML), can improve the raw readings from these sensors taking into account additional environmental information, such as Temperature (T), Relative Humidity (RH) as well as other AQ pollutants Zimmerman et al. (2018).

We propose and compare different techniques, reducing the estimation error up to 94.05% based on Mean Absolute Error (MAE) measurements, that is a 5.95% of Mean Relative Error (MRE), achieving the best results with the Gradient Boosting (GB) algorithm and outperforming the related work, using sensors approximately 10 times less expensive.

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 calibration of the O3 LCS. The results are shown in Section 4 and finally, the conclusions of the experiment and future work are presented in Section 5.

2 Related work

With regard to 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	Relative cost
SDSO11 Nova Fitness Co., Ltd. (2024)	T, RH, PM, PA	Low
DL-LP8P DecentLab, Ltd. (2024)	T, RH, CO2, PA	Low
MiCS-6814 SGX, SensorTech (2024)	CO, NO2, C2H5OH, NH3, CH4	Low
ZPHS01B Zhengzhou Winsen Electronics Technology Co. (2024)	T, 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 systems and/or modules with approximate cost.

A list of these sensor systems and sensor modules with a relative cost are given in Table 1. Notice that these modules have different prices based on the quality of their embedded sensors. At the time of writing, the first three products cost around 10 euros or less, the ZPHS01B module costs around 150 euros and the RAMP module costs more than 10 times the previous cost.

Since one of the key points to improve the accuracy of these LCS is the use of marginal information (such T, RH as well as other AQ pollutants) as mentioned before, it is necessary to use modules embedding as many AQ LCS as possible.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

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

Among these options, the ZPHS01B Zhengzhou Winsen Electronics Technology Co. (2024) AQ module is the best solution at the time of writing, since it embeds 9 different sensors: T (°C), RH (%), as well as PM, CO, CO2, NO2, O3 measured in Parts Per Million (ppm), formaldehyde (CH2O) measured in mg/m^3 and Total Volatile Organic Compounds (TVOC) that are measured using 4 levels according to their concentration (very low, low, intermediate and high). Table 2 summarises 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 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 based directly on the readings from this module 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, 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

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024


Atmospheric Measurement Techniques Discussions

Random Forest (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 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 behaviours and models, in particular for O3 the best two models get a coefficient of determination (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 Neural Network (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. There it is highlighted that is common to overlook the data preprocessing, analysis and interpretability, such as Feature Importance Analysis (FIA), Principal Component Analysis (PCA) and Feature Selection (FS). Besides it is said that the process of optimising algorithms through the selection of their hyperparameters (Hyperparameter Optimization (HPO)) is neglected in most of the environmental research studies considered.

In conclusion, we see that to increase the AQ monitoring resolution at a city scale, LCS are required. However due to their poor performance, further processing of their raw readings is necessary following a criteria of best-practice, by leveraging the information provided by the additional embedded LCS, in particular for the ground level Ozone estimation.

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

In order to carry out the calibration of the O3 sensor from the ZPHS01B module, we need a dataset to train the different 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 data structure provided by the VAQMN is shown in Table 3. The ZPHS01B module's readings are taken at a rate of 10 samples per minute.

The structure and main statistics of the dataset are shown in Table 4. 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 are typically used in a formal and academic context but we need to standardise 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).

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

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

Table 3. Information provided by the official VAQMN station and [units]

Parameter	**[Units]**	**Parameter**	**[Units]**
Temperature	[ºC]	NO	[$\mu g/m^3$]
Humidity	[%R.H.]	NOX	[$\mu g/m^3$]
PM10	[$\mu g/m^3$]	NO2 ECO	[$\mu g/m^3$]
PM10 S/C	[$\mu g/m^3$]	NO ECO	[$\mu g/m^3$]
CO	[mg/m^3]	SO2	[$\mu g/m^3$]
O3	[$\mu g/m^3$]	NH3	[$\mu g/m^3$]
NO2	[$\mu g/m^3$]	HH:MM	[hour:minute]

In Figure 2, it is shown the IoT node, that keeps the ZPHS01B module within a PVC pipe with a small fan at the top, to ensure air movement over the board's sensors. This figure shows the detail of this module and its housing. At the upper site of this IoT node (the head of the node), it is placed the microcontroller that sends data via the LTE-M communications.

3.2 Analysing the dataset

The data collection includes 165 days, approximately five and a half months, from June 8^{th} 2023 till November 20^{th} 2023. Based on this collection, three different datasets have been created, using different monitoring intervals: 10 and 30 min. and

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

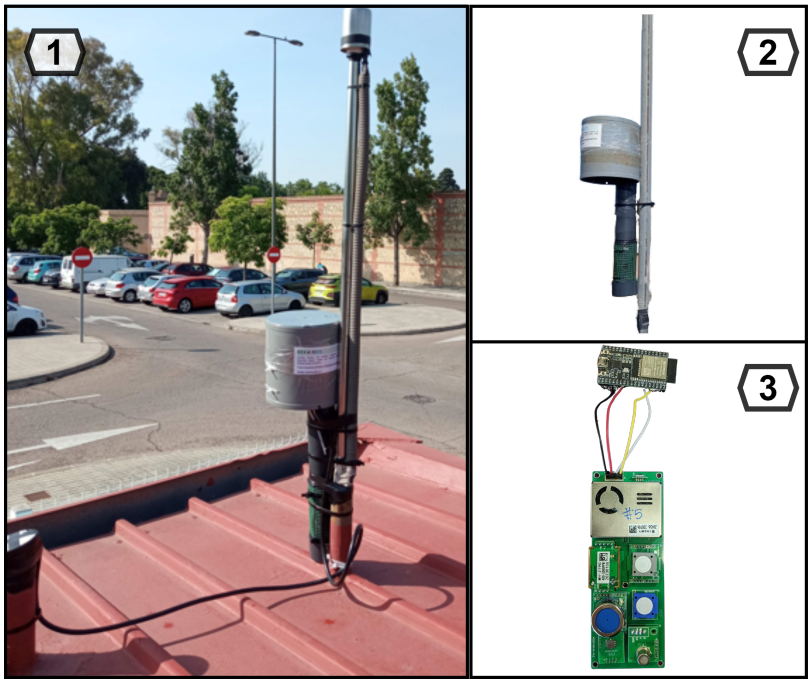

Figure 2. (1) AQ IoT node; (2) deployment detail; (3) hardware detail

1 h. with 23496, 7843 and 3922 samples respectively. Although it is not a large dataset, it is sufficient, as shown in Zhu et al. (2023).

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 4 shows a summary of main statistics of the dataset. From these results, it is worth mentioning that the CH2O, CO, NO2 and TVOC sensors are not very reliable in the ZPHS01B module. Also, the RH sensor has a positive offset as we can see from the maximum value, 118%. The other sensors have a normal behaviour, although with low accuracy.

To characterise the measurements of the ground ozone, we carry out a DFT analysis, to see the ground ozone changing patterns. Figure 3 shows the peaks obtained from the ground ozone 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 $T = 4000 hours = 5.56 months$, that is the total period of data capture. The second peak indicates the other relevant frequency component at $f = 0.04182251 \frac{1}{hour}$, which represents a period $T = 23.91 hours \approx 1 day$, that is, there is a pattern that it is repeated every day, as it could be expected.

Figure 4 and 5 show the O3 readings in $\mu g/m^3$ from the LCS and the official station (reference) for different time duration, the whole dataset (165 days) and one week respectively. 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.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

Table 4. Summary of main statistics of the Dataset

	Temp	Hum	PM2.5	CO2	NO2	O3	O3ref
Min	5.24	62.29	21.25	693.43	0.78	39.57	8.71
Max	42.26	118	83.69	1792.50	18.81	255.76	97.85
Mean	20.60	91.31	49.99	780.33	15.27	114.39	55.72
SD	5.70	18.12	18.14	57.16	5.65	67.11	24.83
Var.	32.57	328.41	329.34	3268.29	31.92	4503.98	616.69
MAD	3.92	16.37	13.31	24.53	0.59	51.40	16.21
Diff.	99.1%	81.9%	87.9%	97.5%	50.6%	75.0%	30.3%
Stat.	not	not	not	not	not	not	not
Seas.	yes	yes	yes	yes	yes	yes	yes
High corr.	yes	yes	yes	yes	not	yes	yes

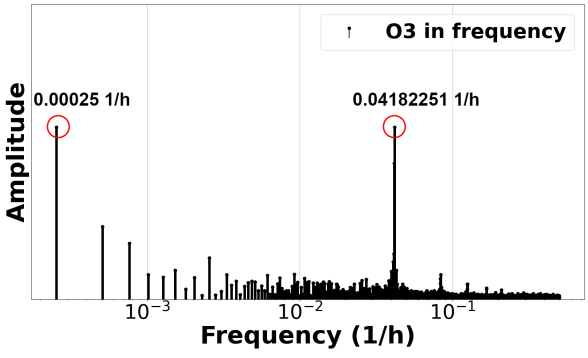

Figure 3. Discrete Fourier Transform (DFT) of O3 readings from LCS

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

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

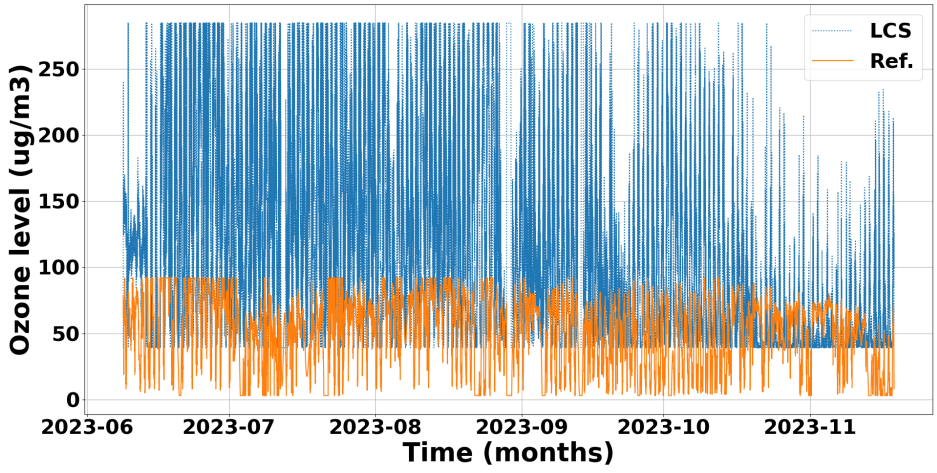

Figure 4. O3 readings in $\mu g/m^3$ from the LCS vs Reference for the whole dataset, 165 days

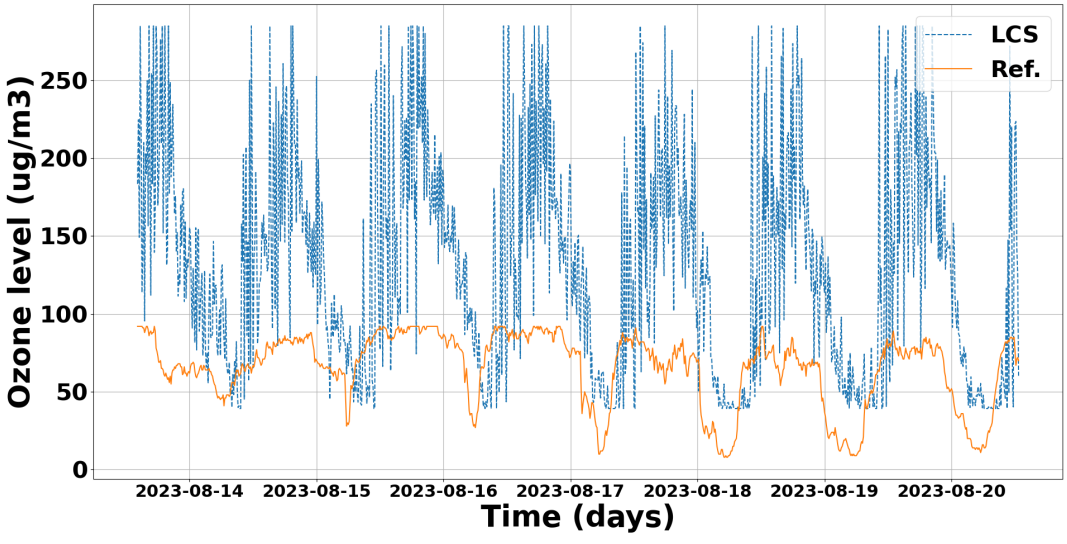

Figure 5. Zoom for a week from Figure 4: O3 readings vs Reference in $\mu g/m^3$

numbers of variables Zhu et al. (2023). These techniques also help to improve the Sample-size to Feature-size Ratio (SFR), that is the ratio between the total number of samples vs the number of features, being recommended a SFR higher than 500, in the same previous reference.

Table 5 shows the normalised output of the FIA using the *scikit-learn* library Pedregosa et al. (2011), for the parameters complementary to O3, for each ML models used. For clarity it is not included the importance of date and ozone itself from LCS values, that complete the rest. From this analysis we conclude that T, RH and CO2 are the most relevant and then will be considered for the next step, FS since they show the highest values.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

Table 5. Feature Importance Analysis (FIA) of ozone's complementary parameters

Model	Temp.	Hum.	PM2.5	CO2	NO2
RF	0.128	0.103	0.069	0.222	0.078
GB	0.107	0.105	0.052	0.211	0.057
ADA	0.119	0.097	0.064	0.246	0.067
DT	0.115	0.088	0.070	0.232	0.061

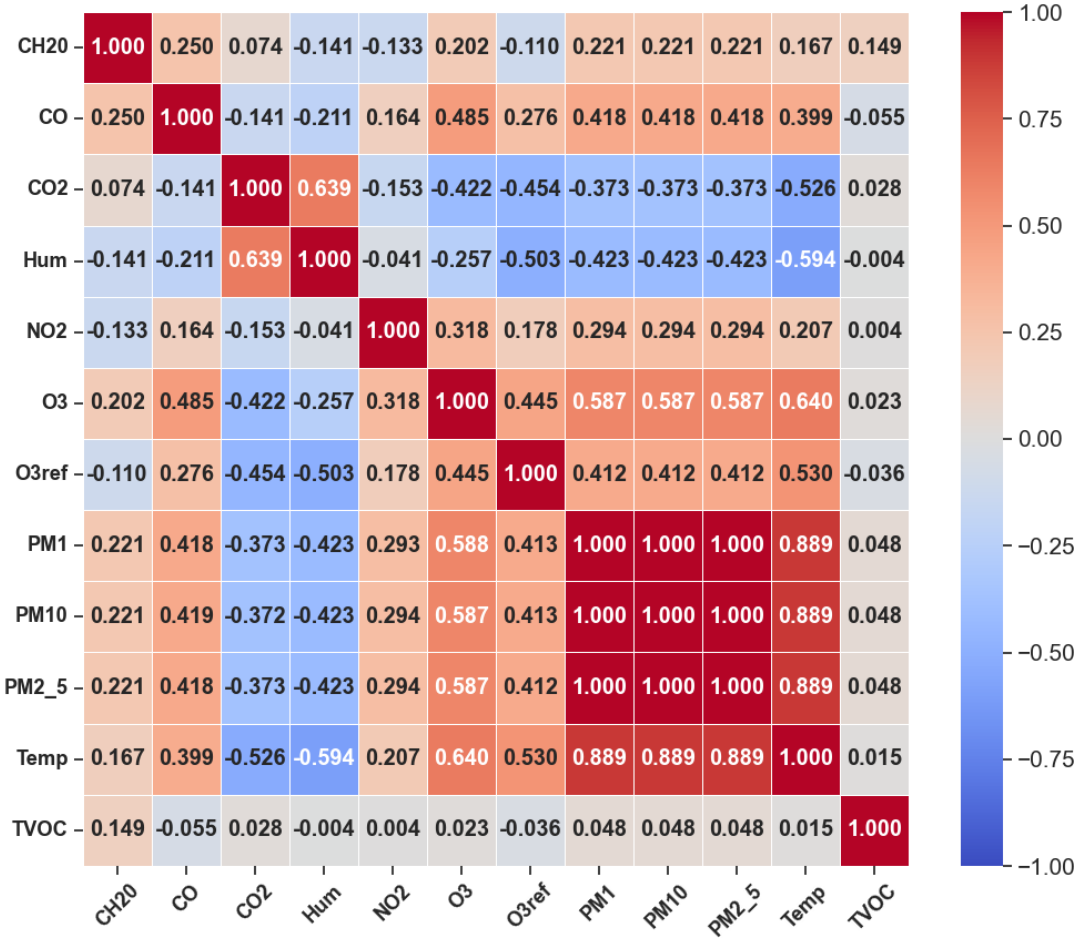

Figure 6. LCS readings and O3 reference correlation matrix

With regard to FS, Figure 6 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).

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

Also, from this analysis, we stress that T 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.

Table 6. Error assessing O3 with [date, O3, T, RH PM2.5, NO2, CO2]

Error/Model	RF	GB	ADA	DT
MSE	71.71	48.46	67.51	152.30
MAPE	0.3036	0.2217	0.2119	0.3082

Table 7. Error assessing O3 with [date, O3, T, RH]

Error/Model	RF	GB	ADA	DT
MSE	52.46	44.52	55.77	94.50
MAPE	0.2286	0.2065	0.1755	0.2264

Following with the FS analysis, the ML models have been trained with their optimal configurations and with different datasets. In particular, among these datasets, two of them showed better results, as we can see in Table 6 and 7, based on [date, O3, T, RH PM2.5, NO2, CO2] and [date, O3, T, RH] respectively, showing Mean Square Error (MSE) and Mean Absolute Prediction Error (MAPE) over 20 different iterations. In this case, we see that fewer features (Table 7), better results, i.e. increasing the SFR. Thus, if we add more features that are not so significant, it makes the dataset poorer. Also, if we include CO2, we do not improve neither. Thus using this low number of features (only four), other dimensionality reduction techniques are not required.

Thus, as a conclusion of this section, the dataset used for ozone LCS calibration is [date, O3, T, RH].

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).

This paper evaluates these ensemble ML algorithms: RF, GB and Adaptive Boosting (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 Decision Tree (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.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

To optimise 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 maximises estimation accuracy or minimises 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.

3.4.1 Random Forest (RF)

Table 8. RF hyperparameters evaluated on *GridSearch*.

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**

It 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 generalisation. 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 8 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* doesn't have a significant role in this use case, we use the default value, 100.

3.4.2 Gradient Boosting (GB)

Table 9. GB hyperparameters evaluated on *GridSearch*.

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

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

It 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 focus 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 9 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 optimised, 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 10. ADA hyperparameters evaluated on *GridSearch*.

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

It 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 generalisation capability and reducing overfitting in regression tasks. Table 10 shows the hyperparameters evaluated, in bold the best option. In this model, there is a key concept to run the optimisation process related to the *estimator* variable, that by default is an instance of type *DecisionTreeRegressor*, initialised 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)

Table 11. DT hyperparameters evaluated on *GridSearch*.

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

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

It is an algorithm that recursively partitions the dataset based on features, aiming to create a hierarchical structure of decision nodes to make predictions. Table 11 shows the hyperparameters evaluated, in bold the best option. The *splitter* hyperparameter indicates which strategy is used to perform the splitting at each node.

4 Results

We have evaluated the performance metrics of these ML models with different configurations (in terms of R^2, Root Mean Square Error (RMSE), MAE and MAPE in $\mu g/m^3$ and execution time in seconds), both with default and optimised hyperparameters, taking into account three different datasets given by different monitoring intervals: 10 and 30 min and 1 h, as depicted in Section 3.2. From these datasets, we have tested different training-test (ratio) percentages: 60%-40%, 70%-30%, 80%-20% and 90%-10%.

Table 12. Performance metrics with the default models and 90% training 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

Table 13. Performance metrics with HPO models and 90% training dataset

	$\mathbf{GB_{optimized}}$	$\mathbf{RF_{optimized}}$	$\mathbf{ADA_{optimized}}$	$\mathbf{DT_{optimized}}$
$\mathbf{R^2}$	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

From all of them, we have achieved the best results with 90/10 training-test ratio with a monitoring interval of 10 min. These results are shown for the default ML models in Table 12 and the ones improved with HPO in Table 13. Using HPO, we optimise the calibration process (higher R^2 and lower errors), but we increase significantly the execution time. In particular,

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

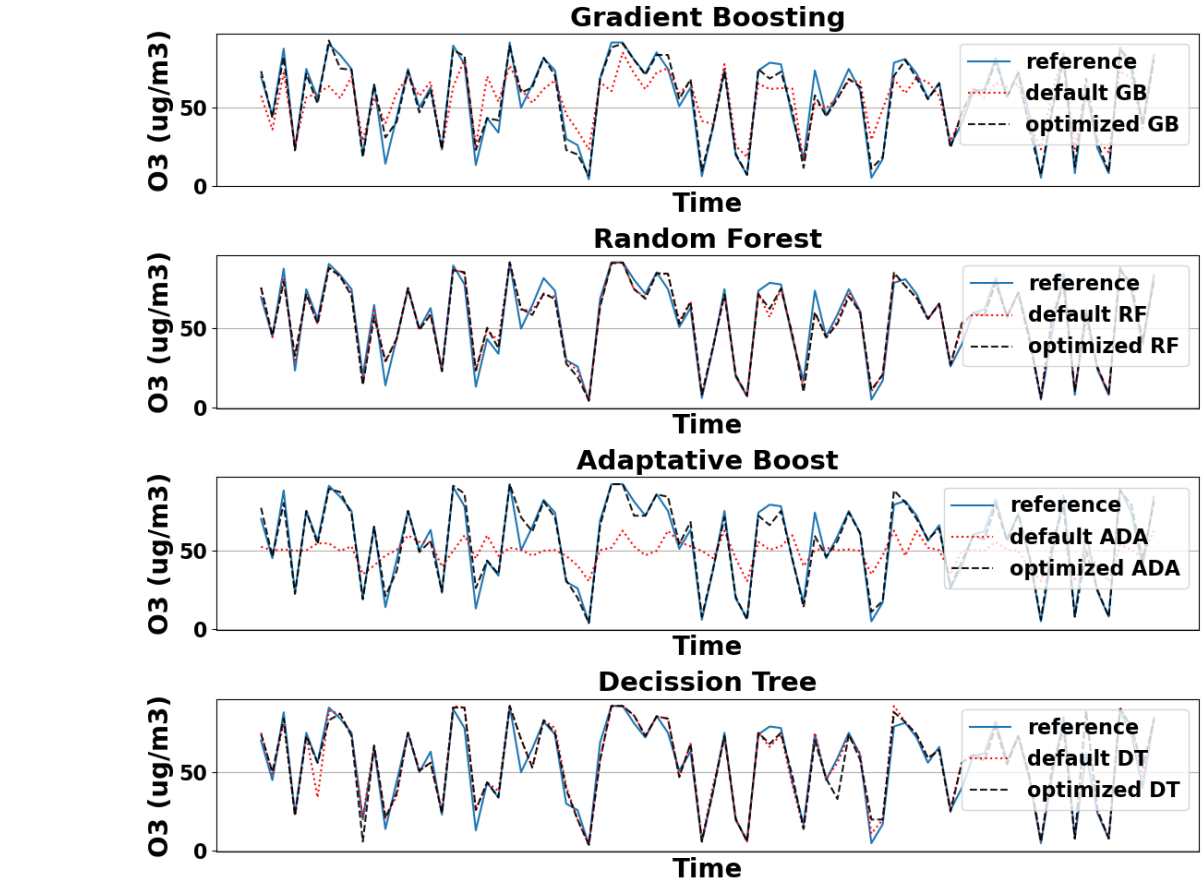

Figure 7. Ozone calibration done with default and optimised models with 90% training dataset

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 13. Notice that the execution time required for the training process is influenced by the hyperparameter optimisation, in particular for GB as it is shown in these tables. Also we highlight that RF and DT are already well optimised and their execution times do not increase in comparison between the default and the optimised versions. In Figure 7 is shown the calibration process for both the default and HPO models vs O3 reference given by the different algorithms.

However, it is common to use a 80/20 training/test ratio Zhu et al. (2023). For this ratio, Table 14 shows the results with the optimised models by HPO, where the GB model is the best one, 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 8. 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 behaviours are shown by the other models, in particular with ADA model.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

Table 14. Performance metrics with HPO models and 80% training dataset

	$GB_{optimized}$	$RF_{optimized}$	$ADA_{optimized}$	$DT_{optimized}$
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

In Figure 9, it is shown the distribution error for the different models, with detail of raw, default and optimised versions. The number of samples are normalised. 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 optimised models. This behaviour is also appreciated with DT, but with lower intensity. However, RF keeps a pretty similar distribution in both versions, default and optimised as we can see in Tables 12 and 13.

Table 15. Standard Deviation (σ) and Confidence Interval (CI) for the error estimation with raw, default and optimised models

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

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

The Standard Deviations (SD) and the Confidence Intervals (CI) in $\mu g/m^3$ are shown in Table 15. This information is obtained from the error distribution statistics given in Figure 9. In the same line as before, once again we can see how the GB adjusts better compared with the other models.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

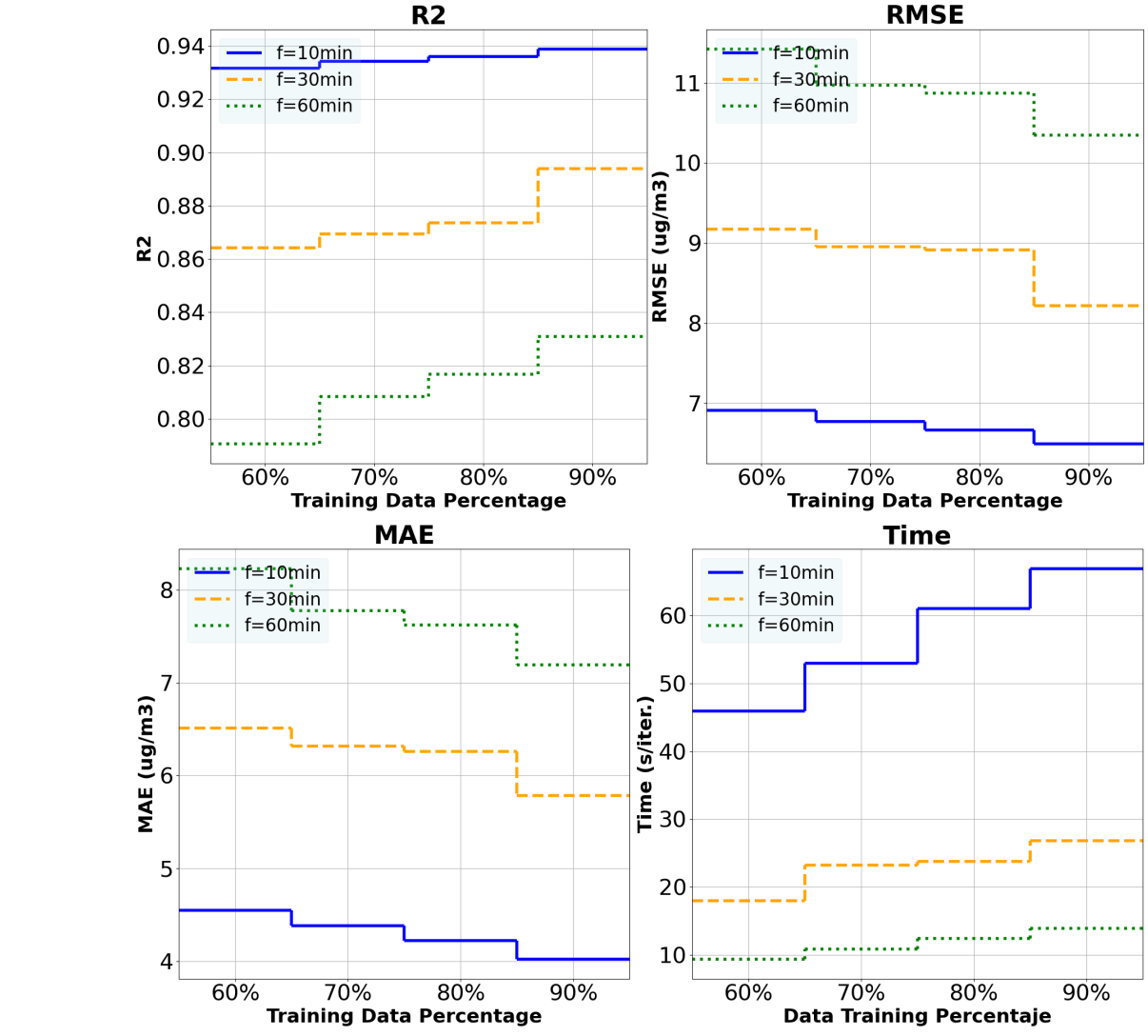

Figure 8. Ozone estimation analysis for GB default and optimised models with different % training datasets and monitoring intervals

Finally, in the Table 16, we show the improvement in % using the different ML models for the calibration process from the LCS raw readings of the module.

In Table 17, we compare our models for ground level Ozone 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, the one we used. Nevertheless, our model reduces the estimation error up to 94.05% based on MAE measurements, that is a 5.95% of MRE approximately, using GB with only 4 features, as shown in Section 3.3.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

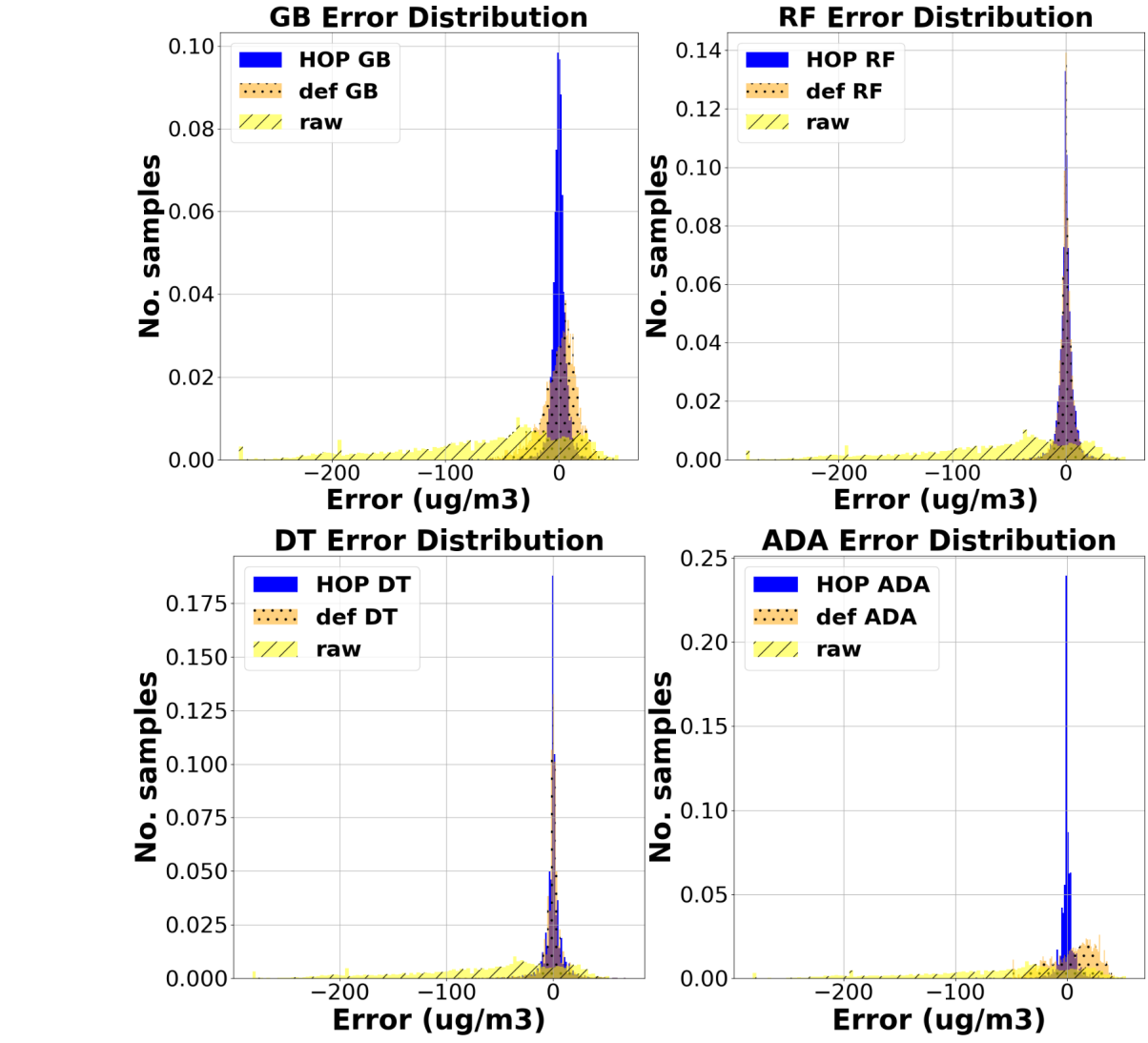

Figure 9. Distribution error for the different models, with detail of raw, default and optimised versions

5 Conclusions

In this paper we have focused on ground level Ozone (O3) since in the cities is an indicator of other pollution levels using LCS nodes based on the ZPHS01B module. These nodes permit to increase the spatial sampling of AQ. Due to their low accuracy, after a thorough analysis, we have used machine learning methods, in particular DT and the ensemble algorithms (GB, RF and ADA), taking into account additional environmental information, reducing the estimation error in around 94% with 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 adjusted with higher accuracy.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

Table 16. Improvement (in%) of O3 calibration from the raw readings with the different optimised 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%

Table 17. Comparison with the similar related work

Study	**Location**	**Sensor**	$\mathbf{R^2}$	MRE [%]
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	0.93	5.95%

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. 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.

As future work, we plan to enlarge the dataset and to add complementary parameters, such as wind speed, traffic density to increase the accuracy of this model.

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.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

Acknowledgements. This paper is partially funded by the Grant PID2021-126823OB-I00, funded by MCIN/AEI/10.13039/501100011033; by ERDF, a way of making Europe and the Grant TED2021-131040B-C33, funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR. Also, Generalitat Valenciana with grant references CIAICO/2022/179 and CIAEST/2022/64.

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

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

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.

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.

Corp, Z. W. E. T.: Ozone detection module ZE27-03, https://www.winsen-sensor.com/d/files/manual/ze27-o3.pdf, [Accessed 27-01-2024], 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: 27/04/2024, 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.

et al, C. B.: 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.

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.

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: 25/01/2024, 2024.

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.

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: 27/04/2024, 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.

https://doi.org/10.5194/amt-2024-127
Preprint. Discussion started: 16 October 2024

Atmospheric Measurement Techniques Discussions

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 27-01-2024], 2024.

SGX, SensorTech: Air quality sensor MiCS-6814, https://www.sgxsensortech.com/content/uploads/2015/02/1143_Datasheet-MiCS-6814-rev-8.pdf, accessed: 21/05/2024, 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 27-01-2024], 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.