STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–26.
Published online in International Academic Press (www.IAPress.org)

A hybrid Machine Learning approach for air quality prediction in Morocco: combining CatBoost with metaheuristic optimization algorithms

Rachid ED-DAOUDI [1], Sokaina EL KHAMLICHI [1,2,*], Badia ETTAKI [1]

[1] *Laboratory of Research in Informatics, Data Sciences and Artificial Intelligence, School of Information Sciences, B.P. 604, Rabat-Instituts, Rabat, Morocco*
[2] *Research Team in Science and Technology, Higher School of Technology of Laayoune, Ibn Zohr University, P.O. Box 3007, Laayoune, Morocco*

Abstract Air pollution poses serious risks to public health and environmental sustainability, particularly in rapidly urbanizing areas of developing countries. This study investigates whether combining machine learning algorithms with metaheuristic optimization techniques can improve the accuracy and efficiency of air quality prediction in Morocco. The main objective is to compare direct classification of Air Quality Index (AQI) categories with a regression-based approach, and to evaluate the effectiveness of two optimization strategies—Arithmetic Optimization Algorithm (AOA) and Hunger Games Search (HGS)—in tuning the CatBoost model's hyperparameters. Using five months of air quality data from two monitoring stations in Ait Melloul, we modeled concentrations of PM2.5, PM10, CO, and derived corresponding AQI classifications. The hybrid approach demonstrated that regression-based classification improved accuracy by nearly 30 percentage points over direct classification. Moreover, HGS achieved similar predictive performance to AOA but was over twice as computationally efficient. CO concentration predictions in residential areas achieved high accuracy ($R^2 > 0.95$), while particulate matter predictions revealed limitations in capturing extreme pollution events. These findings suggest that combining gradient boosting with metaheuristic optimization is a promising strategy for developing scalable and accurate air quality forecasting systems in North African urban environments, with important implications for public health protection and environmental policy implementation

Keywords Air Pollution, Hybrid Machine Learning, CatBoost Algorithm, Metaheuristic Optimization, Air Quality Index

DOI: 10.19139/soic-2310-5070-2705

1. Introduction

Air pollution can be defined as the global environmental challenge with potentially the greatest extent of threat to human health and sustainable development if left unchecked. Air pollution occurs when harmful substances enter the atmosphere in the form of solid particles, liquid droplets, or gases, with more than 180 different pollutants being identified, originating from natural and human activity [1], [2]. Primary sources of such pollutants stems from industrial processes, manufacturing, and transportation. Air pollution has the potential to demonstrate its effects over a lengthy period of time, and its impacts can be quite difficult to determine simultaneously. It has already been associated with increased rates of mortality, cancer, and cardiovascular diseases that contribute to a gradual death in many industrialized societies [3], [4].

As reported by the World Health Organization, exposure to air pollution results in approximately four million premature deaths every year [5]. This number reinforces the importance and value of air quality monitoring and

*Correspondence to: Sokaina EL KHAMLICHI (Email: s.elkhamlichi@uiz.ac.ma). Research Team in Science and Technology, Higher School of Technology of Laayoune, Ibn Zohr University, P.O. Box 3007, Laayoune, Morocco. LyRICA: Laboratory of Research in Informatics, Data Sciences and Artificial Intelligence, School of Information Sciences, P.O. Box. 6204, Rabat-Instituts, Rabat, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

management systems to address air quality issues. In Morocco, this issue has been acknowledged within the framework of the National Air Pollution Control Program launched in 2010, which further established within the National Sustainable Development Strategy (2017-2030). These initiatives align with the United Nations Sustainable Development Goals (2015) and Morocco's commitments under the Paris Climate Agreement (COP21, 2015) and Marrakech Climate Conference (COP22, 2016). The creation of these programs clearly indicates a national commitment to tackle air pollution problems in line with its international commitments to environmental standards.

In the past, the monitoring of air quality was based on a sampling and measurement system which was conducted at periodic intervals, but as technology evolves, the capability to monitor air quality on continuous and real-time basis, with early warning of air pollution events has become possible. This would be an important precursor to protect public health [6]. The continuous air quality monitoring systems have generated large-scale datasets that highlight temporal patterns and fluctuations in air pollution levels [7], which are important findings to develop effective early warning systems and mitigation measures. The primary objective of this study is to develop an effective system for forecasting air quality in Morocco using artificial intelligence techniques. Specifically, this research aims to: create a model capable of predicting in advance when air quality will deteriorate; compare two different prediction methods (direct classification of air quality categories versus regression-based prediction followed by categorization); analyze pollution differences between industrial and residential districts; and identify the most efficient and accurate AI technique for this task. The findings are meant to provide local authorities with a practical tool to better protect residents health from air pollution.

This paper is arranged as follows: section 2 will discuss the related works in the field of air quality prediction. Section 3 will describe the material and methodology which will include processes for data collection to preprocessing methods as well as the hybrid ML approach. Section 4 lays out the experimental results and model comparative analysis on the outcomes. Finally Section 5 discusses the conclusions and potential directions for future studies.

2. Related Works

There are different types of air quality prediction models that need to be developed as a basic requirement for effective early warning systems. Over the last number of years, research has attempted various methodological approaches which fall into statistical models, machine learning frameworks, and neural networks.

Due to their simplicity and interpretability, traditional statistical models such as multivariate linear regression and SARIMA have been widely applied in forecasting various phenomena, including energy consumption [8], inflation [9], Rainfall [10], and PM2.5 concentrations. However, these models often struggle to capture the complex, nonlinear interactions between air pollutants and meteorological variables [11],[12],[13]. While SARIMA-based time series approaches effectively preserve seasonal patterns, they are less capable of adapting to the dynamic and rapidly changing nature of urban air pollution.

Machine learning techniques have demonstrated superior performance in modeling complex nonlinear relationships, making them particularly effective for forecasting tasks such as ambient temperature and solar radiation [14], crop water needs [15], crop yield prediction [16], and PM2.5 prediction. For example, XGBoost has been successfully optimized through hyperparameter tuning to improve the accuracy of spatial PM2.5 prediction in China [17]. Similarly, Random Forest combined with recursive feature elimination has produced promising results in AQI forecasting in India [18]. Comparative studies further highlight that machine learning models often outperform traditional linear models, offering improved generalization, especially in high-dimensional data environments [19].

Additionally, Deep Learning approaches have achieved promising results in complex data. For instance, Gated Recurrent Unit (GRU), Recurrent Neural Network (RNN), and Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) have been applied to forecast stock prices [20]. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models have been leveraged for spatiotemporal analysis of pollution data [21]. Hybrid models that incorporate meteorological simulation (e.g., WRF-Chem)

or staged neural architectures further improve forecasting accuracy [22]. Recent advances such as Temporal Fusion Transformers (TFTs) combine sequence learning and attention mechanisms to deliver high-resolution AQI predictions, particularly when coupled with transfer learning to overcome data scarcity issues [23].

Another trend is the use of Graph Neural Networks (GNNs), which allow spatially distributed data (e.g., stations across a city or country) to be modeled as graphs with learned relationships. Heterogeneous GNNs address the variability in station types and land-use profiles, providing robust forecasts across diverse urban landscapes [24].

Hybrid methods remain an area of active research, with several studies validating the integration of neural networks and statistical models. Three-stage hybrid models have been applied to improve PM2.5 forecasts through successive refinement of features and learning modules [25]. Moreover, multimodal learning has emerged as a frontier area, incorporating heterogeneous data sources such as outdoor images and sensor data through diffusion-driven transformers to reduce predictive errors [26].

Regionally, Moroccan studies have contributed significantly to air quality modeling. Work by Sbai et al. during the COVID-19 lockdown analyzed pollutant variation using GAMs, while their follow-up research applied ML models to long-term PM10 prediction at national scale [27], [28]. Additional studies by Bounakhla et al. showed seasonal lag-effects of meteorological variables on PM concentrations in urban Morocco [29]. Despite these advancements, the potential of optimization-enhanced AI models for AQI prediction remains underexplored. Few studies use modern metaheuristics such as Hunger Games Search (HGS) and Arithmetic Optimization Algorithm (AOA), which have shown strong results for non-convex and high-dimensional model tuning tasks. For example, Zandi et al. demonstrated improved PM2.5 prediction in Tehran by optimizing the Echo State Network using swarm-based algorithms [30], while Razavi-Termeh et al. employed HGS to enhance TabNet performance in mapping dust susceptibility, outperforming classical metaheuristics like PSO and GWO [31]. However, no recent work has explicitly tested HGS or AOA in AQI forecasting, signaling a gap this study seeks to address.

3. Data

3.1. Data collection

The research occurred in Ait Melloul, a medium-sized city in southwestern Morocco, in the Souss-Massa region. This site has unique attributes for air quality research because it encompasses industrial and residential districts, a temperate climate, and is close to urban development and agricultural conditions. The data source for this project is derived from a previous work of Bekkar et al. [32], who presented an AIoT platform to conduct real-time air quality analysis in Southwestern Morocco. The present analysis uses two monitoring stations, ZI101W station within the industrial zone located near heading RA1 and the expressway to Agadir Al Massira International Airport, and RH101W station, which is a residential station. The ZI101W station was specifically located within an area that has many vehicles traveling through it, including large vehicles, while the RH101W station was chosen for its location within a residential context. Data were collected for five months, from October 2022 to February 2023, which resulted in 3,504 hourly observations for each station. The dataset includes the subsequent parameters shown in Table 1:

The dataset indicates differences between the two monitoring stations as exhibited in Table 2 which corroborates the different environmental parameters in residential and industrial zones.

The indicator at the ZI101W station was higher in average concentrations for all pollutants measured, PM2.5 averaged 12.53 μg/m^3 compared with RH101W at 7.70 μg/m^3, and PM10 averaged 16.12 μg/m^3 versus RH101W at 10.02 μg/m^3; however, those competing station indicators also aligned with differing distribution among AQI categories, with 27.3% of days as 3 and 4 AQI categories in the industrial zone versus 10.5% in residential settings.

The monitoring equipment was also routinely calibrated and maintained to assure accuracy of data. The missing values for all but a few parameters were all less than 8% of the datasets and therefore addressed in the preprocessing step of the analysis.

Table 1. Air pollutant and meteorological variables in dataset

Category	Variable	Description	Unit
Air Pollutant Measurements	PM2.5	Fine particulate matter with diameter ≤ 2.5 μm	μg/m^3
	PM10	Coarse particulate matter with diameter ≤ 10 μm	μg/m^3
	PM1	Ultrafine particulate matter with diameter ≤ 1 μm	μg/m^3
	CO	Carbon monoxide	ppb
	CO2	Carbon dioxide	ppm
Meteorological Variables	Temperature	Ambient air temperature	°C
	Relative humidity (RH)	Amount of water vapor in the air	%
	Dew point	Temperature at which air becomes saturated with water vapor	°C
	Precipitation	Amount of rainfall	mm
	Wind speed	Speed of air movement	kph
	Wind gust	Brief increase in wind speed	kph
	Wind direction	Direction from which wind is coming	degrees
	Sea level pressure	Atmospheric pressure at sea level	hPa
	Cloud cover	Percentage of sky covered by clouds	%
	Visibility	Maximum distance at which objects can be seen	km

Table 2. Descriptive statistics for air quality and meteorological parameters

Parameter	Location	Count	Mean	Std	Min	0.25	0.5	0.75	Max
PM2.5	ZI101W	3251	12.53	11.78	0.00	3.80	9.17	18.00	133.00
PM2.5	RH101W	3418	7.70	7.79	0.00	2.00	5.50	10.50	66.50
PM10	ZI101W	3251	16.12	16.08	0.00	5.75	12.00	22.00	326.00
PM10	RH101W	3418	10.02	9.43	0.00	3.33	7.50	13.58	83.30
PM1	ZI101W	3251	3.77	5.50	0.00	0.25	1.80	5.25	66.00
PM1	RH101W	3418	2.38	4.19	0.00	0.00	0.40	3.19	65.20
CO2	ZI101W	3252	7.72	13.74	0.28	1.89	2.74	5.17	160.00
CO2	RH101W	3419	6.12	4.34	1.04	3.11	4.47	7.38	22.00
CO	ZI101W	3252	23.64	66.98	0.16	2.25	3.75	8.99	1021.00
CO	RH101W	3419	12.61	12.86	0.99	4.46	7.34	14.60	66.60
RH	ZI101W	3252	65.63	20.46	16.50	49.55	67.20	83.40	99.90
RH	RH101W	3419	59.47	22.99	4.78	41.05	60.20	78.80	99.90
AQI_category	ZI101W	3504	1.98	0.99	1.00	1.00	2.00	3.00	4.00
AQI_category	RH101W	3504	1.53	0.76	1.00	1.00	1.00	2.00	4.00

3.2. Data preprocessing

The preprocessing stage was essential in reconverting the raw sensor data into a format that was analysable and modelable or, in statistical terms, set into an appropriate data structure. The processing phase entailed multiple steps and methodologies to ensure datatypes and formats were congruent with the study objectives and to ensure data integrity and consistency.

3.2.1. Temporal formatting and missing values :

Timestamp information was reconfigured into a standard datetime format and subsequently set as the index for the time series analysis. The datasets contained approximately 1-5% missing values for each of the parameters analysed. The missing data were then resolved using K-nearest neighbours (KNN) imputation which accounted for both temporal and spatial context of the data by using neighbouring observations for imputation, rather than simply using a mean or median imputation. KNN was used as it was better able to maintain the complexities found in air quality time series data. As the data were sourced from an external AIoT platform [32], we relied on the authors' reporting of regular sensor calibration. However, undetected sensor drift or localized missing data patterns may still introduce bias, which should be considered in future deployments.

3.2.2. Air Quality Index categorization :

To calculate an AQI, continuous measurements of particulate matter were categorized into discrete levels. This categorization utilized PM2.5 and PM10 concentrations, in accordance with locally adapted categorization as shown in Table 3.

Table 3. Air quality index categorization

AQI Category	**Description**	**PM2.5 Thresholds (μg/m^3)**	**PM10 Thresholds (μg/m^3)**
1	Good	≤ 8	≤ 10
2	*Moderate*	> 8 and ≤ 20	> 10 and ≤ 25
3	Unhealthy for Sensitive Groups	> 20 and ≤ 40	> 25 and ≤ 50
4	*Unhealthy*	> 40	> 50

In accordance with standard AQI methods category assignment for each observation was determined from the maximum PM2.5 or PM10 categories.

3.2.3. Feature engineering :

Time-lagged features were developed to incorporate temporal dependencies in the data that would lead to improved air quality predictions. In line with the work of Yakubu Saputra [33], Autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses were conducted to identify temporal relationship, which provided evidence to develop an adjusted standard combination of significant lags [1, 2, 3, 6, 12, 24] for practical purposes rather than taking the time-lags from statistical significance alone.

Four uniform input combinations were implemented across all pollutants:

- M1: Lags at hours 1, 2, 3, 4
- M2: Lags at hours 1, 2, 3, 4, 5
- M3: Lags at hours 1, 2, 3, 4, 5, 6
- M4: Lags at hours 1, 2, 3, 4, 5, 6, 12

This standard layering approach permitted appropriate assessment of model application using Support Vector Regression (SVR) to determine the best configuration for lag predictors in terms of suitability for each pollutant. Overall, the analysis suggested for PM2.5 and PM10, the compact lag feature using minimum of lags hours 1, 2, 3, and 4 (M1) had best fit, indicating that recent measurements contained sufficient information sufficient to achieve good predictions. CO had actually slightly better prediction capacity with M1 under the compact lags, and though it was thought that M4 configuration of numerous lags was a better technique of assessing lagged measurement influence on pollutant outcomes, M1 produced sufficient and superior prediction performance.

3.2.4. Data standardization and partitioning :

Before training the models, feature standardization was performed in order to adjust the scaled input variables so that all features contribute equally to the model regardless of the original units or scales. The datasets were then split into training (70%) and test (30%) subsets utilizing a chronological split in order to preserve the underlying temporal structure of the data. Non-random shuffling of the datasets allowed the clinical models to be assessed on their abilities to predict future air quality based on historical data, which is paramount in real-world prediction applications.

These preprocessing steps resulted in structured datasets with suitable features for both regression and classification modeling frameworks so that a rigorous comparative analysis of air quality prediction methodologies would ensue.

4. Methods

The approach developed for this study follows a structured methodology combining multiple techniques. Figure 1 illustrates the workflow from data collection through comparative analysis

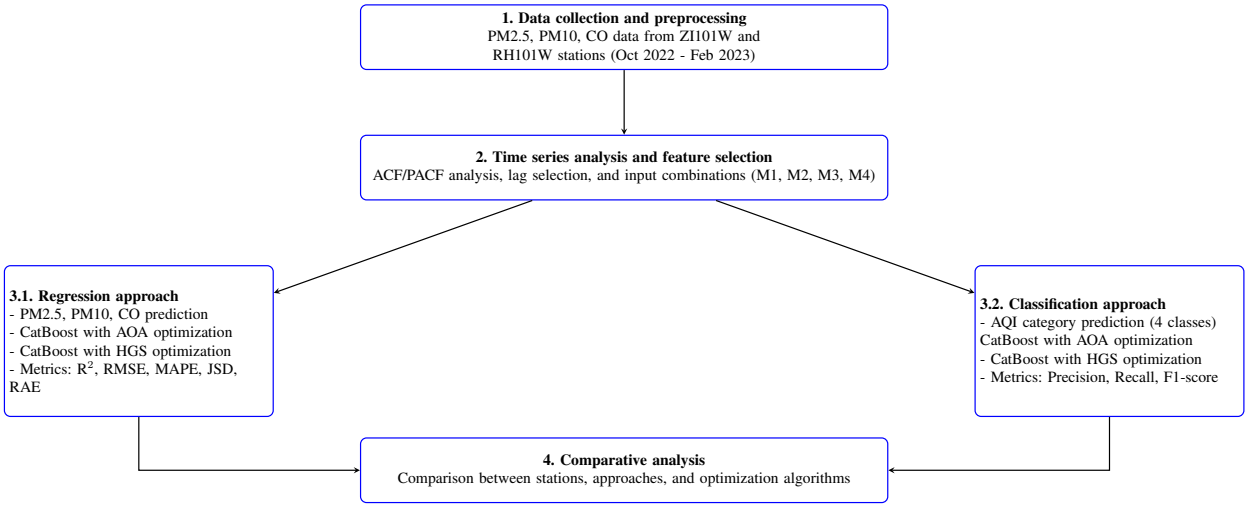

Figure 1. Air quality prediction methodology overview

Time series analysis was performed to understand patterns in the data, followed by feature selection for model development. The methodology included two parallel paths: a regression approach aimed at predicting exact values of air pollutants using the CatBoost algorithm with different optimization techniques (AOA and HGS), and a classification approach that categorized air quality into four different classes. Both approaches were evaluated using appropriate performance metrics: statistical error measurements for the regression models and accuracy indicators for the classification models. The final step involved comparing results between the two stations, between the regression and classification approaches, and between different optimization algorithms to determine which combinations performed best for air quality prediction.

4.1. Hybrid Machine Learning approach

The contribution of this research lies in the development of a hybrid machine learning approach that addresses key challenges in air quality prediction by using metaheuristic algorithms to optimize the hyperparameters of a gradient boosting model.

4.1.1. Technical background :

CatBoost [34] is a more advanced version of gradient boosting developed by Yandex and shares many positive aspects for consideration in air quality prediction, including native handling of categorical variables without preprocessing, improved computational efficiency through gradient one-sided sampling and exclusive feature bundling techniques, and robustness against overfitting via its ordered boosting algorithm with permutation-based approaches. CatBoost efficiently handles missing values either as a separate class or using mean/most frequent value methods, which is important for environmental sensor data. In this study, CatBoost was implemented for both regression (pollutant concentration prediction) and classification (AQI categorization) tasks with hyperparameters optimized using metaheuristic algorithms.

The AOA algorithm is a population-based metaheuristic technique that employs basic arithmetic operations with two key parameters: Math Optimizer Probability (MOP) controlling exploration-exploitation balance, and Math Optimizer Accelerated (MOA) accelerating convergence. Its dual-phase strategy alternates between exploration (using multiplication/division) and exploitation (using addition/subtraction) based on probabilistic criteria, making it suitable for hyperparameter optimization in air quality prediction due to its simplicity and ability to navigate complex parameter spaces.

The HGS algorithm is a crowd-based optimization technique inspired by competitive animal foraging behavior. Its core mechanism uses hunger as an adaptive weight that increases with suboptimal solutions. HGS employs a fitness-proportional hunger ratio where hungrier individuals exhibit more aggressive search behavior. The algorithm dynamically balances exploration and exploitation based on hunger levels, promoting exploration early and shifting toward exploitation as hunger diminishes. This adaptive search mechanism responds dynamically to the optimization landscape, making HGS particularly effective for hyperparameter tuning in air quality prediction models.

AOA and HGS were chosen over traditional methods such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) due to their simpler parameter settings, adaptive search behavior, and competitive performance in recent environmental and time-series applications. Unlike Bayesian optimization, they also allow flexible exploration in discrete and continuous spaces without prior assumptions on the objective landscape.

4.1.2. Proposed methodology :

The combination of CatBoost with AOA and HGS utilized a structured hyperparameter optimization procedure:

- Parameter space definition: four hyperparameters were defined: number of iterations (100-300), depth of tree (5-8), learning rate (0.05-0.2), and L2 regularization coefficient (2-5).
- Objective function: the objective function minimized Root Mean Square Error (RMSE) between predicted concentrations of pollutants and actual concentrations. Additional metrics including Mean Bias Error (MBE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), Relative Absolute Error (RAE), and Jensen-Shannon Divergence (JSD) were calculated to provide comprehensive evaluation. For classification problems, the objective was to maximize the weighted F1 score across Air Quality Index (AQI) categories, with Precision and Recall as supporting metrics.
- Cross-validation approach: a 5-fold cross-validation approach was applied to select the optimal hyperparameters with 12 iterations each to permit maximal selection of each optimizer with a population size 8. These values were chosen based on computational feasibility and are consistent with common settings in recent optimization studies.
- Separate models in parallel: separate CatBoost models were formed using CatBoost-AOA and CatBoost-HGS for all regression and classification tasks for performance comparison.

The hybrid approach employed attempts to assist with the difficult problem of hyperparameter selection of machine learning models used for air quality prediction by exploiting the benefits of using gradient boosting and a metaheuristical approaches to improve prediction performance.

4.1.3. Mathematical description of optimization :

The goal of optimization is to tune CatBoost's hyperparameters:

$$\theta = \{d, lr, i, L2r\} \tag{1}$$

Where d is the depth, lr is the learning rate, i are the iterations and l2r is the L2 regulation.
AOA explores the search space by alternating between two arithmetic operations:

- Exploration phase: uses multiplication/division to diversify solutions.
- Exploitation phase: uses addition/subtraction to refine local optima.

In AOA, the transition between exploration and exploitation is controlled by two dynamic parameters: Math Optimizer Probability (MOP) and Math Optimizer Accelerated (MOA). MOP is a linearly decreasing function over iterations, determining the probability of selecting exploration (multiplication/division) versus exploitation (addition/subtraction) at each step. MOA increases over time to progressively favor local refinement. This allows the algorithm to explore widely in early stages and gradually shift focus toward convergence. The formula becomes:

- If in exploration phase:

$$\theta_i^{t+1} = \theta_i^t \times (1 + r) \tag{2}$$

- If in exploitation phase:

$$\theta_i^t + r \times \left(\theta_{\text{best}}^t - \theta_i^t\right) \tag{3}$$

Where r is a uniformly sampled random number and θ_{best}^t is the best-known solution at iteration t.
HGS mimics competitive behavior in resource-constrained environments, where a "hunger weight" dynamically influences search aggressiveness:

- The hunger level increases for poor-performing solutions.
- Hungrier candidates explore more broadly, while others focus locally.

In HGS, the hunger level for each candidate is computed based on the inverse of its fitness relative to the population, and then normalized using min-max scaling to fall within [0,1]. Hungrier individuals apply stronger adjustments toward the best solution, promoting aggressive search early in optimization. This adaptive weighting balances global and local exploration as the population evolves. The adaptive update is done with the formula:

$$\theta_i^{t+1} = \theta_i^t \times H_i \cdot r \cdot \left(\theta_{\text{best}}^t - \theta_i^t\right) \tag{4}$$

Where H_i is the normalized hunger level of solution i, and r is a random coefficient.
This adaptive mechanism allows HGS to balance exploration and exploitation dynamically, resulting in faster convergence and reduced computational time, as observed in our experiments.

4.2. Evaluation metrics

To assess both regression and classification tasks, a series of evaluation metrics were used to assess the performance of the hybrid models. The metrics, based on eq. (5) to (10), were chosen to provide various complementary perspectives of model accuracy, bias, and practical implications for applications in air quality prediction.

$$\text{MBE} = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{N} \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{6}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{8}$$

$$\text{JSD}(P \parallel Q) = \frac{1}{2}\left[\text{KL}(P \parallel M) + \text{KL}(Q \parallel M)\right] \tag{9}$$

$$\text{RAE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{\sum_{i=1}^{N} |y_i - \bar{y}|} \tag{10}$$

where N displays the count of observations, y_i and $\hat{y}_i$ are the i-th real value and i-th estimated value, respectively, and $\bar{y}$ displays the mean of data points, $JSD(P \parallel Q)$ displays the Jensen-Shannon Divergence between P and Q, and

$$M = \frac{P+Q}{2}.$$

The evaluation indices examined in this study related to classification are Precision, Recall, and F1 Score. Based on potential states for existing and future samples, four states are possible: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The indices can be calculated based on these four factors using eqs. (11), (12) and (13).

$$\textbf{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\textbf{Recall} = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{12}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{13}$$

These metrics collectively provide a multifaceted assessment of model performance, enabling comparison between different modeling approaches and optimization algorithms.

5. Results and discussion

5.1. Statistical characteristics of air pollutants

Based on an analysis of the air pollutant data from ZI101W (industrial zone) and RH101W (residential zone) monitoring stations in Ait Melloul, substantial differences for air quality indicators emerged. The industrial zone consistently recorded higher concentrations of air pollutants compared to the residential area. PM2.5 levels averaged 12.53 μg/m^3 in the industrial zone versus 7.70 μg/m^3 in the residential area, representing a 63% difference. Similarly, PM10 concentrations averaged 16.12 μg/m^3 at ZI101W compared to 10.02 μg/m^3 at RH101W. This disparity was most pronounced for CO, where the industrial zone recorded mean values of 23.64 ppb versus 12.61 ppb in the residential area.

The time series data presented in figure 2 clearly illustrates these differences, with the industrial zone consistently showing higher pollutant concentrations than the residential area across all three pollutants.

Several notable pollution episodes were observed during the monitoring period, particularly in December 2022 and January 2023, when PM2.5 and PM10 levels at the industrial site exceeded 100 μg/m^3 and 300 μg/m^3, respectively. The most striking difference appears in CO measurements, where the industrial zone experienced spikes reaching approximately 1000 ppb in October 2022 and January 2023, while the residential zone maintained relatively stable and low concentrations throughout the study period.

The distribution of AQI in figure 3 further emphasized these spatial differences.

- Station ZI101W: Category 1 (39.5%), Category 2 (33.3%), Category 3 (17.1%), Category 4 (10.2%).
- Station RH101W: Category 1 (60.5%), Category 2 (29.0%), Category 3 (7.6%), Category 4 (2.9%).

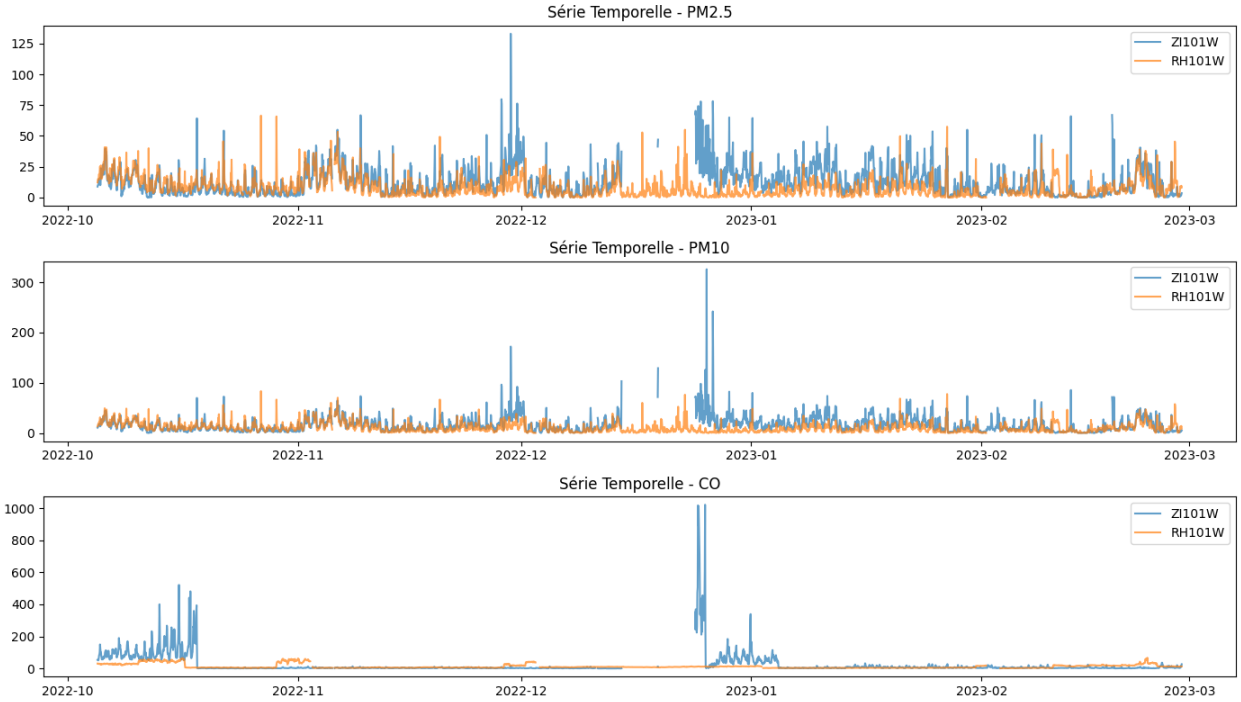

Figure 2. Time series of PM2.5, PM10, and CO concentrations

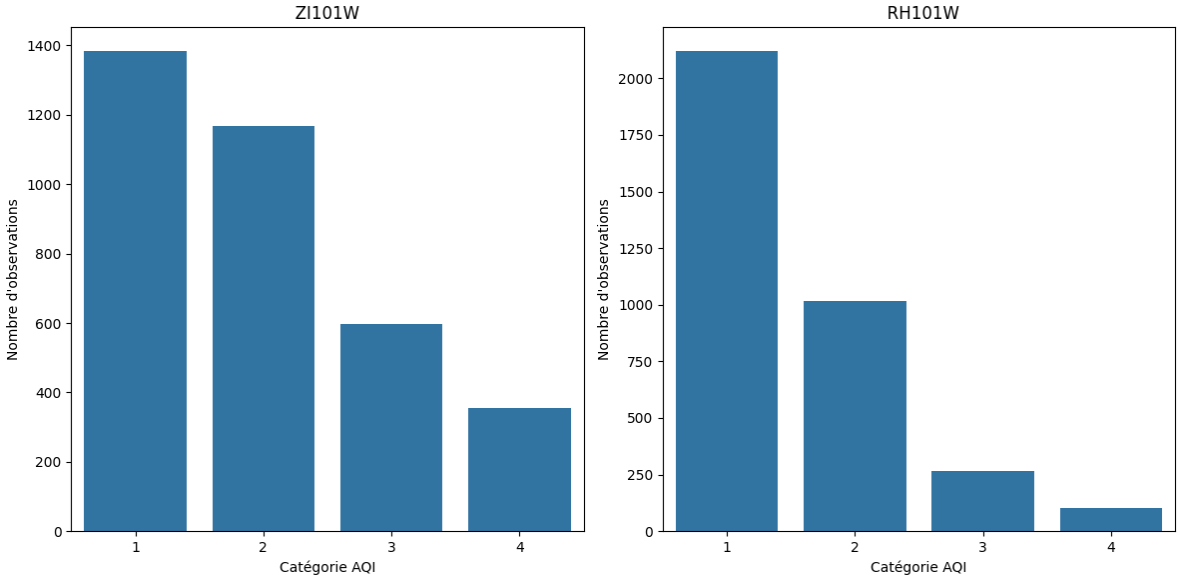

Figure 3. Distribution of air pollutants by station

The industrial zone experienced nearly three times as many days in the ”Unhealthy” category (4) and more than twice as many days in the ”Unhealthy for Sensitive Groups” category (3) compared to the residential area.

Correlation analysis in figures 4 and 5 revealed strong positive relationships between particulate matter measurements (PM2.5 and PM10) at both stations, with correlation coefficients exceeding 0.80.

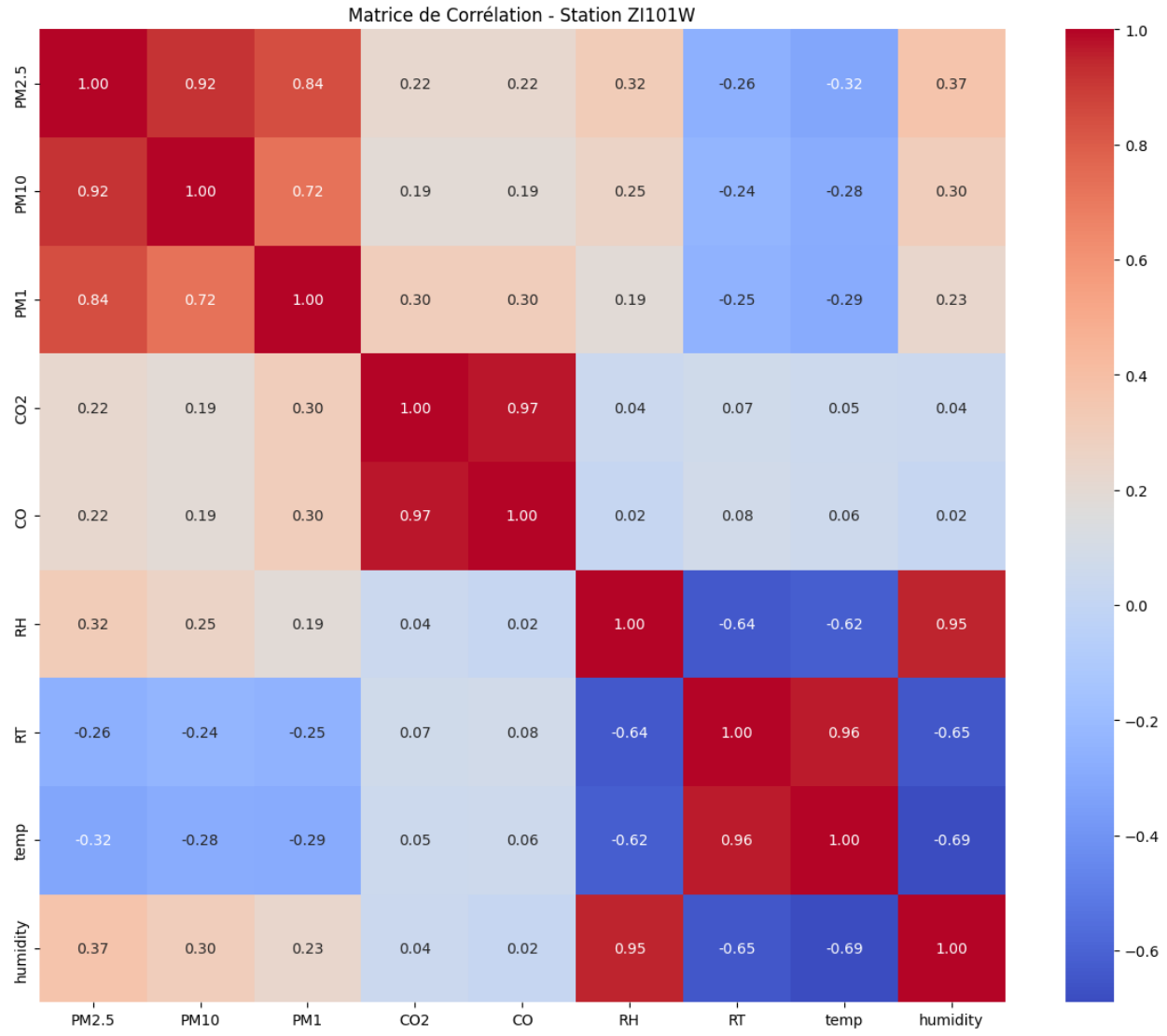

Figure 4. Correlation matrix of air pollutants - station ZI101W

The relationship between PM2.5 and CO was stronger in the industrial zone (r = 0.60) than in the residential area (r = 0.48), suggesting different emission sources between the two locations.

5.2. Performance of regression models

The regression approach aimed to predict absolute concentrations of PM2.5, PM10, and CO by hybrid CatBoost models optimized with AOA and HGS algorithms. The two models demonstrated reasonable prediction ability across all pollutants, but performance differences were large between monitoring stations and optimization techniques.

5.2.1. PM2.5 prediction performance :

In the industrial station (ZI101W), CatBoost models demonstrated reasonable prediction ability (R^2 close to 0.62 (CatBoost-HGS: 0.6191, CatBoost-AOA: 0.6111), RMSE 5.6 $\mu g/m^3$). Figure 6 shows that while the

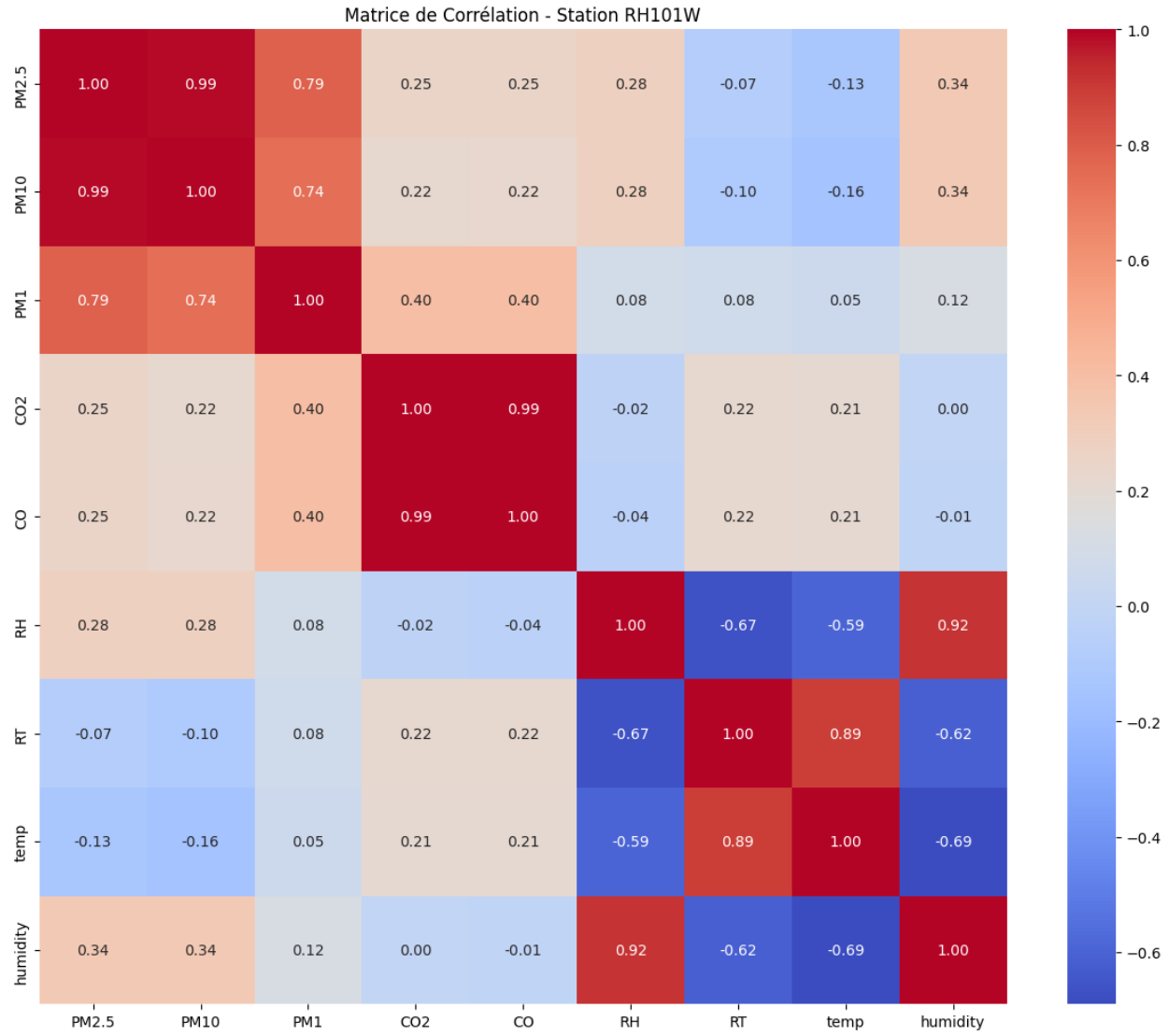

Figure 5. Correlation matrix of air pollutants - station RH101W

models adequately captured the overall patterns and moderate concentrations, the models consistently tended to underpredict the extreme peaks.

Several pollution events with concentrations exceeding 50-65 $\mu g/m^3$ were predicted at only 25-35 $\mu g/m^3$, indicating limitations in capturing these exceptional episodes.

At the residential station (RH101W), similar performance was observed with R^2 values of 0.6226 (CatBoost-AOA) and 0.6178 (CatBoost-HGS) as presented in figure 7

The figure indicates that both models once again tracked patterns of baseline and moderate concentrations reasonably well but exhibited difficulties with extreme values, especially near points 250, 300, 500, and 650, where the actual concentration values may have been higher than 35-45 $\mu g/m^3$ but the predictions remained lower.

5.2.2. PM10 prediction performance :

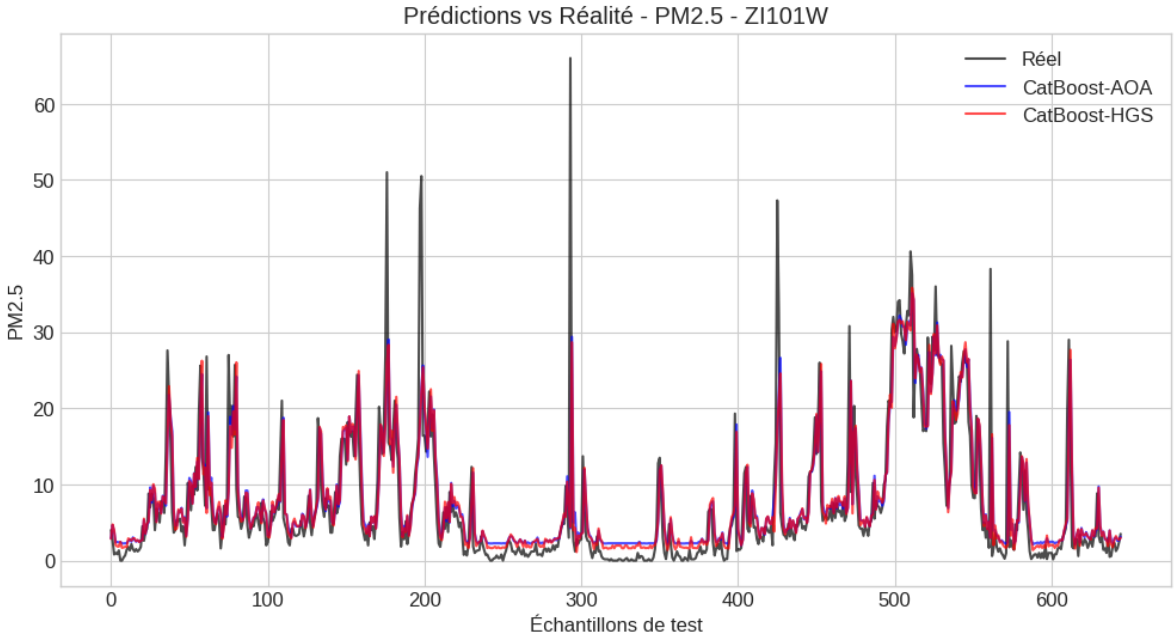

Figure 6. Observed vs. predicted PM2.5 concentrations - ZI101W station

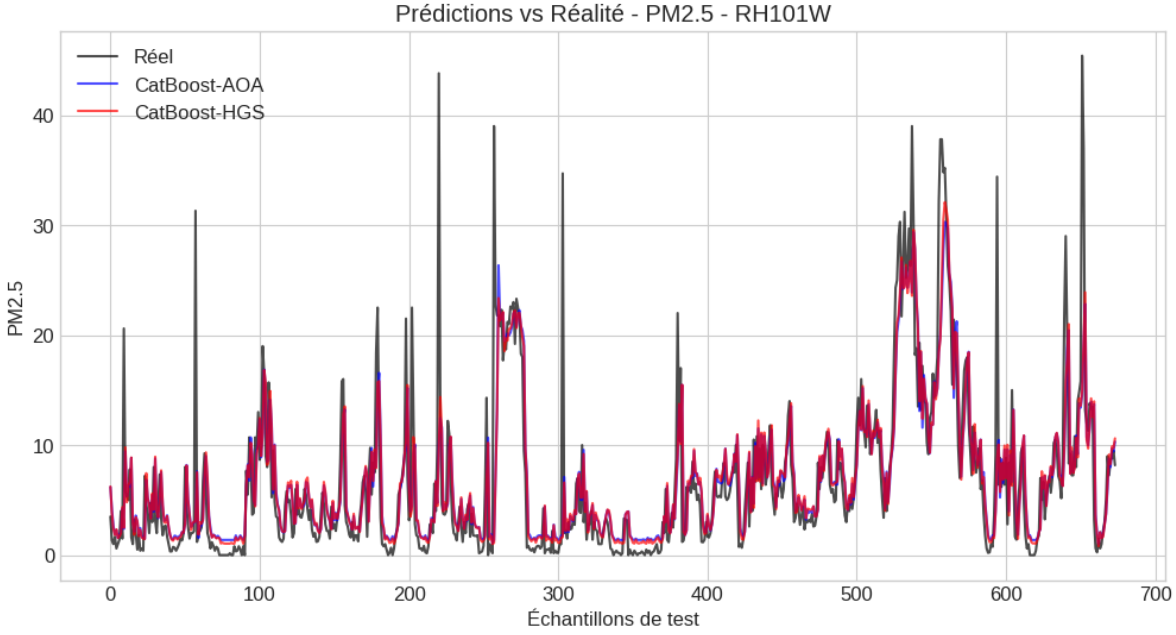

Figure 7. Observed vs. predicted PM2.5 concentrations - RH101W station

The PM10 predictions exhibited patterns similar to PM2.5, with R^2 values near 0.61 for the industrial station and 0.62 for the residential station. As shown in both figures 8 and 9, while both models tracked concentrations that were within normal ranges well, the models demonstrated serious underestimations of pollution peaks. This was

more of an issue at the industrial monitoring station where actual values were indicating more than 80 μg/m^3 at times while the prediction values rarely surpassed 45μg/m^3.

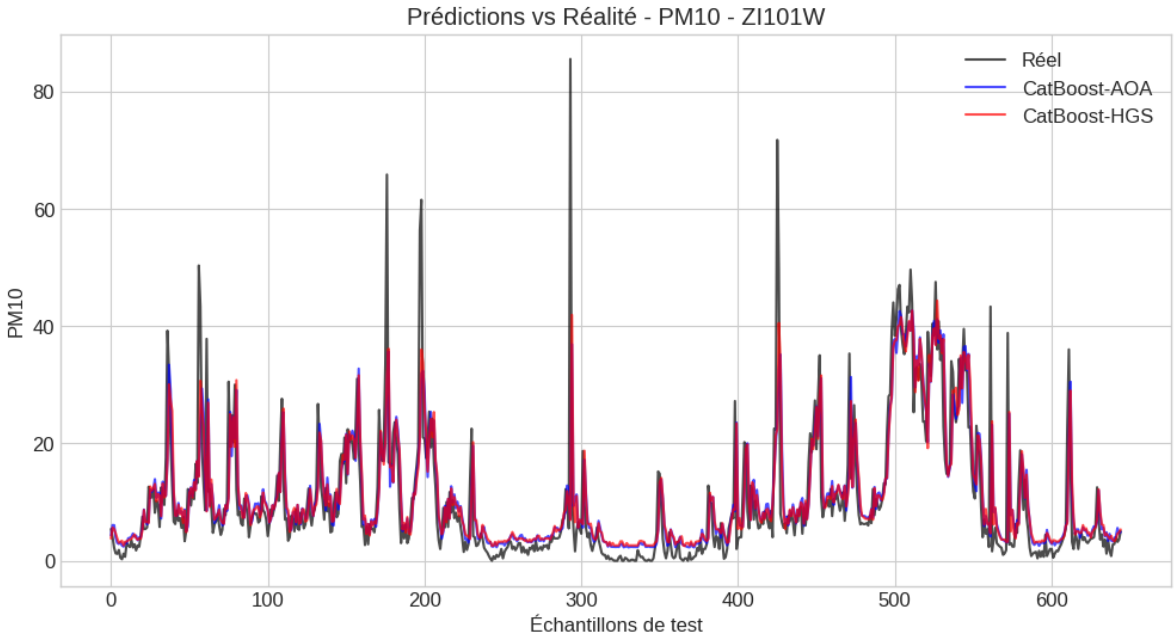

Figure 8. Observed vs. predicted PM10 concentrations - ZI101W station

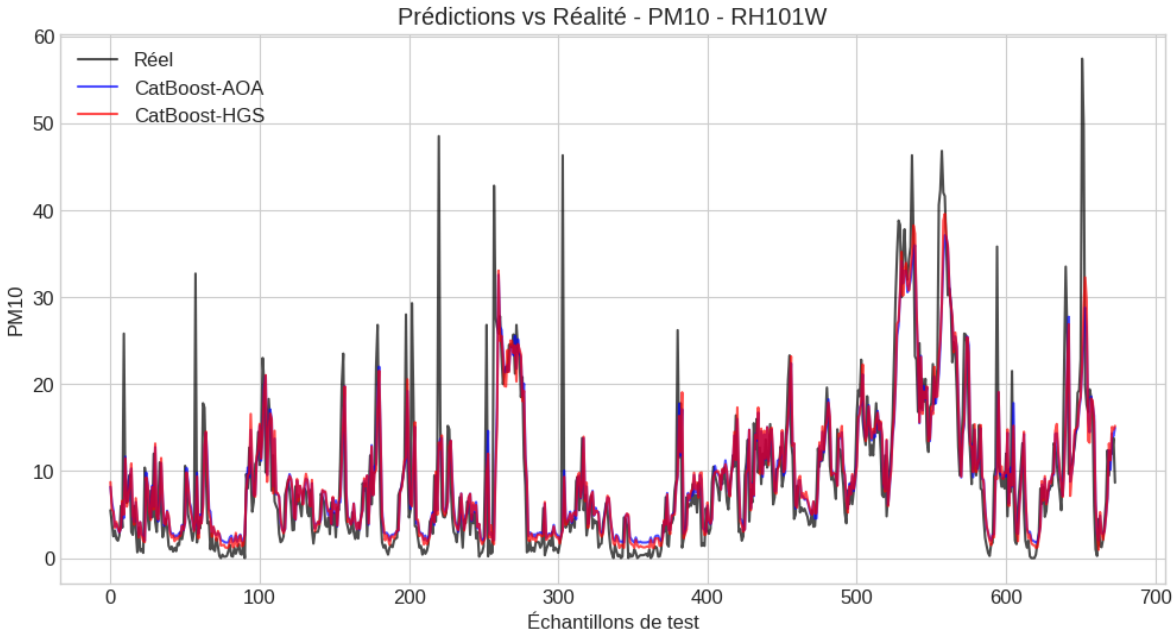

Figure 9. Observed vs. predicted PM10 concentrations - RH101W station

The consistency in performance patterns between PM2.5 and PM10 reflects the strong correlation between these pollutants, suggesting similar underlying temporal dynamics and prediction challenges.

The modality of PM2.5 and PM10 performance invariability implies strong correlation between the two pollutants, leading to very similar underlying temporal dynamics and eventual issues in accurate prediction.

5.2.3. CO prediction performance :

CO predictions showed remarkable station-dependent performance differences. At the residential location (RH101W), models achieved exceptional accuracy ($R^2 > 0.95$, RMSE $\approx$ 2.07 ppb) with predictions closely tracking actual measurements throughout the test period, including concentration peaks of approximately 60 ppb, as demonstrated in figure 10.

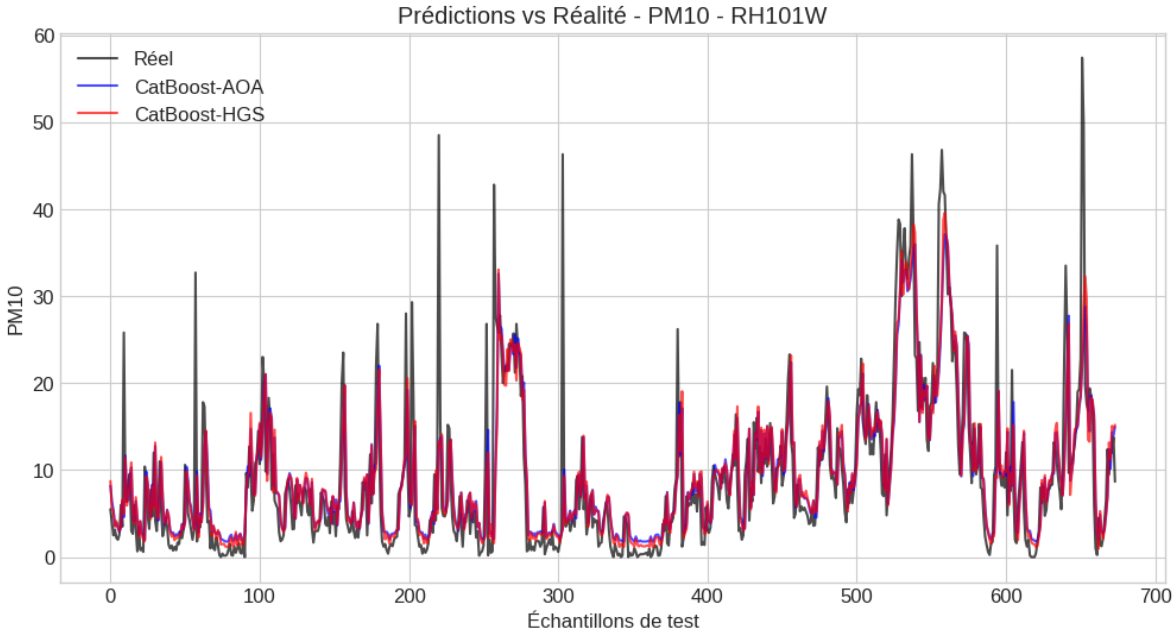

Figure 10. Observed vs. predicted CO concentrations - ZI101W station

In contrast, at the industrial station (ZI101W), the models achieved moderate performance with R^2 values of 0.6324 (CatBoost-AOA) and 0.6158 (CatBoost-HGS). Figure 11 shows that while the models captured general patterns in CO concentrations, they had difficulty predicting some of the sudden spikes.

It should be noted that Figure 10 displays a subset of the CO range at ZI101W, focusing on values up to approximately 36 ppb, while the actual data included extreme values reaching 1021 ppb, which likely accounts for the lower R^2 values at this station.

The exceptional accuracy observed for CO in the residential area may be attributed to its lower variability and more stable emission sources, which make short-term patterns easier to model compared to the more volatile PM concentrations.

5.2.4. Comparison of optimization algorithms :

Across both pollutants and locations, both optimization methods (AOA and HGS) showed overlapping performance, with differences in R^2 and RMSE values typically less than 1%. An emerging dominance was not consistently observable. Specifically: CatBoost-HGS was the best predictor of particulate matter concentration at the industrial site; CatBoost-AOA had slight advantages at the residential site, and in predicting CO concentrations. Also, by utilizing the lag structure with configuration M1 (which incorporated values from previous 1-4 hours), all of the temporal dependencies were identified that were most relevant to predicting short-term air quality.Overfitting

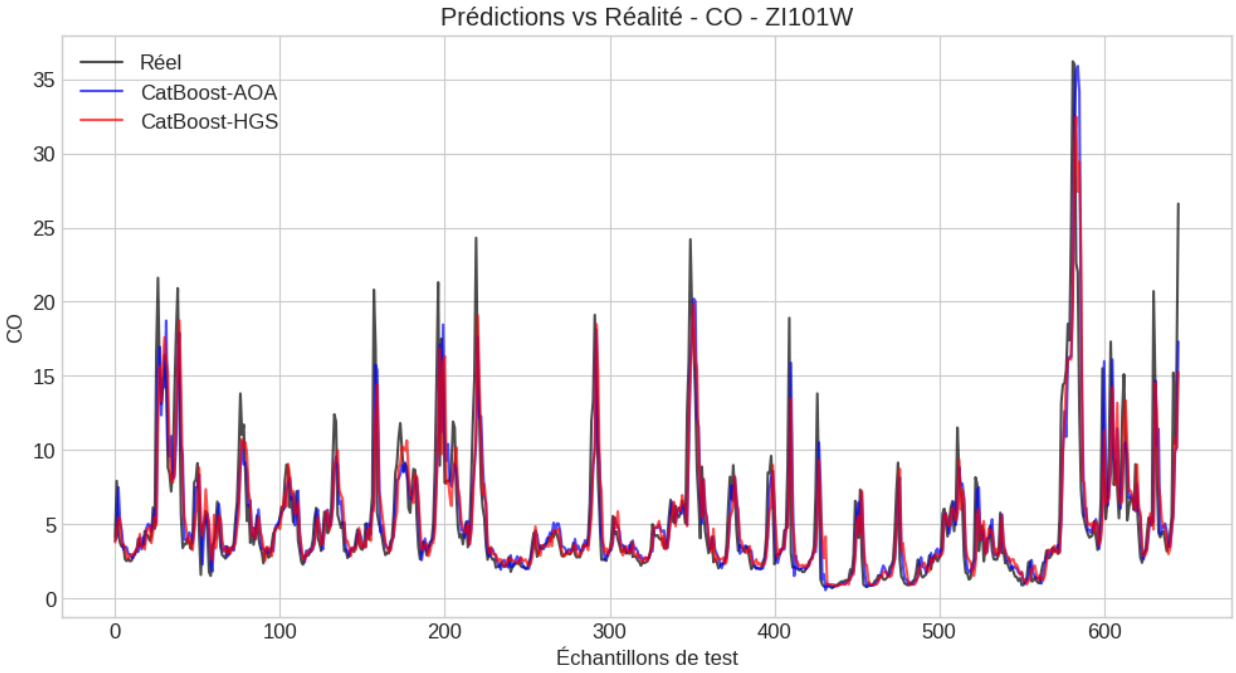

Figure 11. Observed vs. predicted CO concentrations - ZI101W station

was mitigated using CatBoost's ordered boosting and early stopping features, ensuring model generalization during cross-validation. CatBoost's built-in feature importance analysis showed that the most influential predictors were recent lags of PM2.5 and PM10, followed by meteorological variables such as temperature and humidity. This provides useful interpretability for potential policy applications. Table 4 includes additional information provided by the evaluation metrics.

Table 4. Performance metrics of CatBoost models (AOA and HGS) for different pollutants and stations.

Pollutant	**Station**	**Model**	**MBE**	**RMSE**	**MAPE (%)**	**R^2**	**JSD**	**RAE**
PM2.5	ZI101W	CatBoost-AOA	-2.401	5.6082	7.80	0.6111	0.235	0.389
PM2.5	ZI101W	CatBoost-HGS	-1.995	5.5502	7.76	0.6191	0.228	0.381
PM2.5	RH101W	CatBoost-AOA	-1.832	4.6681	11.86	0.6226	0.213	0.377
PM2.5	RH101W	CatBoost-HGS	-1.875	4.6979	11.86	0.6178	0.219	0.382
PM10	ZI101W	CatBoost-AOA	-3.142	7.3573	10.30	0.6017	0.257	0.398
PM10	ZI101W	CatBoost-HGS	-2.709	7.3000	9.91	0.6079	0.245	0.392
PM10	RH101W	CatBoost-AOA	-2.103	5.6794	10.30	0.6191	0.230	0.381
PM10	RH101W	CatBoost-HGS	-2.085	5.6815	10.30	0.6188	0.232	0.382
CO	ZI101W	CatBoost-AOA	-1.414	2.7792	8.89	0.6324	0.243	0.368
CO	ZI101W	CatBoost-HGS	-1.647	2.8415	8.68	0.6158	0.251	0.384
CO	RH101W	CatBoost-AOA	-0.328	2.0757	4.25	0.9541	0.068	0.046
CO	RH101W	CatBoost-HGS	-0.335	2.0798	4.31	0.9540	0.069	0.047

The MBE values are negative across the board and confirm the visual assessment that models typically underpredict pollutants, especially in peak episodes. For instance, CO predictions at the residential station showed a significant reduction in bias, with MBE values approximately five times lower than particulate matter predictions.

The additional evaluation metrics confirmed consistent model underprediction across pollutants, particularly during peak episodes, as evidenced by negative MBE values. CO predictions at the residential station showed remarkably lower bias—approximately five times lower than particulate predictions—with extremely low MAPE

values (≈4.3%). JSD metrics reinforced this pattern, with values near 0.07 for CO at RH101W versus 0.21-0.26 elsewhere. Most notably, RAE comparisons revealed the relative error for CO at RH101W was approximately eight times lower than other prediction cases. This striking performance difference highlights the importance of site-specific factors in air quality modeling, as despite similar measurement methods across pollutants, significant predictability disparities emerged between locations.

To assess the stability of model performance, all metrics were computed over five cross-validation folds, and results are reported as mean ± standard deviation as shown in figure 12.

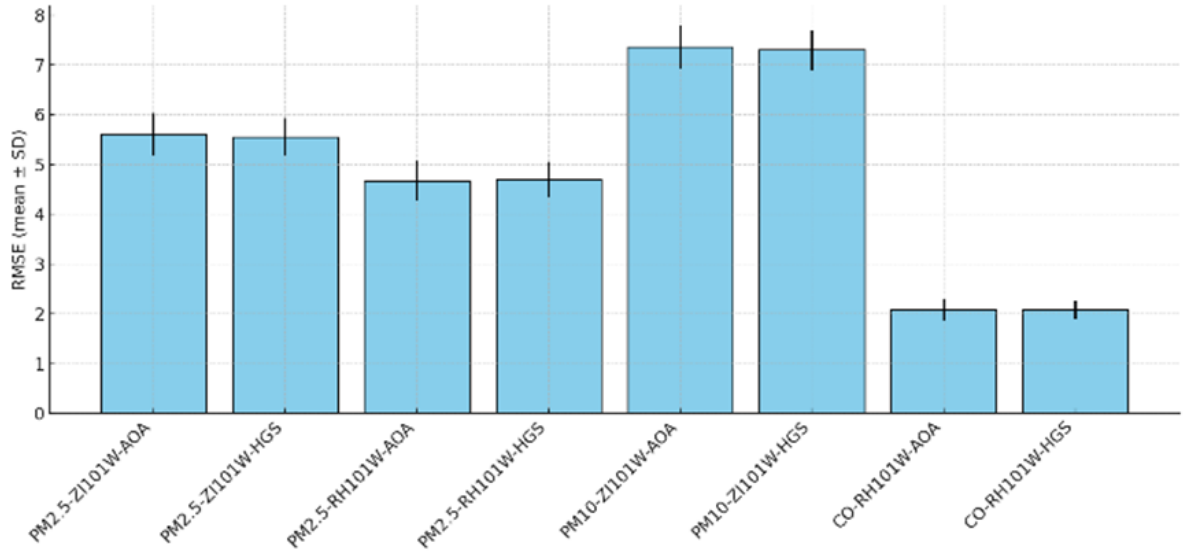

Figure 12. Cross-validated RMSE (mean ± SD over 5 folds) for CatBoost models optimized using AOA and HGS

This approach helps account for variance in model outputs due to temporal fluctuations in the dataset and stochasticity in the metaheuristic optimization. Additionally, training vs. validation performance was monitored to ensure overfitting did not occur; CatBoost's built-in regularization and early stopping mechanisms helped maintain generalization. However, future work could further quantify this with learning curves and confidence intervals across multiple random seeds.

5.2.5. Handling of Extreme Pollution Events :

The consistent underestimation of extreme pollution events by the regression models is attributed to the skewed distribution of the training data, where high concentration values were relatively infrequent. Since CatBoost minimizes global error, it tends to prioritize fit around the mean, neglecting rare but critical outliers. To address this, future models could implement quantile regression or weighted loss functions that emphasize high pollution events. Additionally, oversampling extreme instances or applying synthetic data augmentation may help mitigate this bias and improve real-world alerting capacity.

5.3. Performance of classification models

The direct classification approach aimed to predict discrete AQI categories from historical pollutant data, providing immediately interpretable outputs for air quality assessment and public health messaging. Both CatBoost-AOA and CatBoost-HGS models were evaluated on their ability to accurately classify air quality into the four categories.

5.3.1. Overall classification performance :

At the industrial monitoring station (ZI101W), the CatBoost-HGS model slightly outperformed CatBoost-AOA across most metrics. The weighted F1 score for CatBoost-HGS was 0.5268, an increase over the 0.5246 F1 score

from CatBoost-AOA. Both models exhibited moderate performance overall, but again, this is reflective of the challenges associated with multi-class pollution prediction in the presence of class imbalance while simultaneously predicting in a multi-class context.

At the residential station (RH101W), similar results were seen, whereby the optimization provided by CatBoost-HGS achieved a weighted F1 score of 0.5429 and a weighted F1 score of 0.5379 for CatBoost-AOA. Therefore, while the optimization advantage noticed by HGS was noted at both stations, the difference was less substantial.

5.3.2. Class-specific performance analysis :

A more detailed examination of class-specific metrics are presented in figure 13.

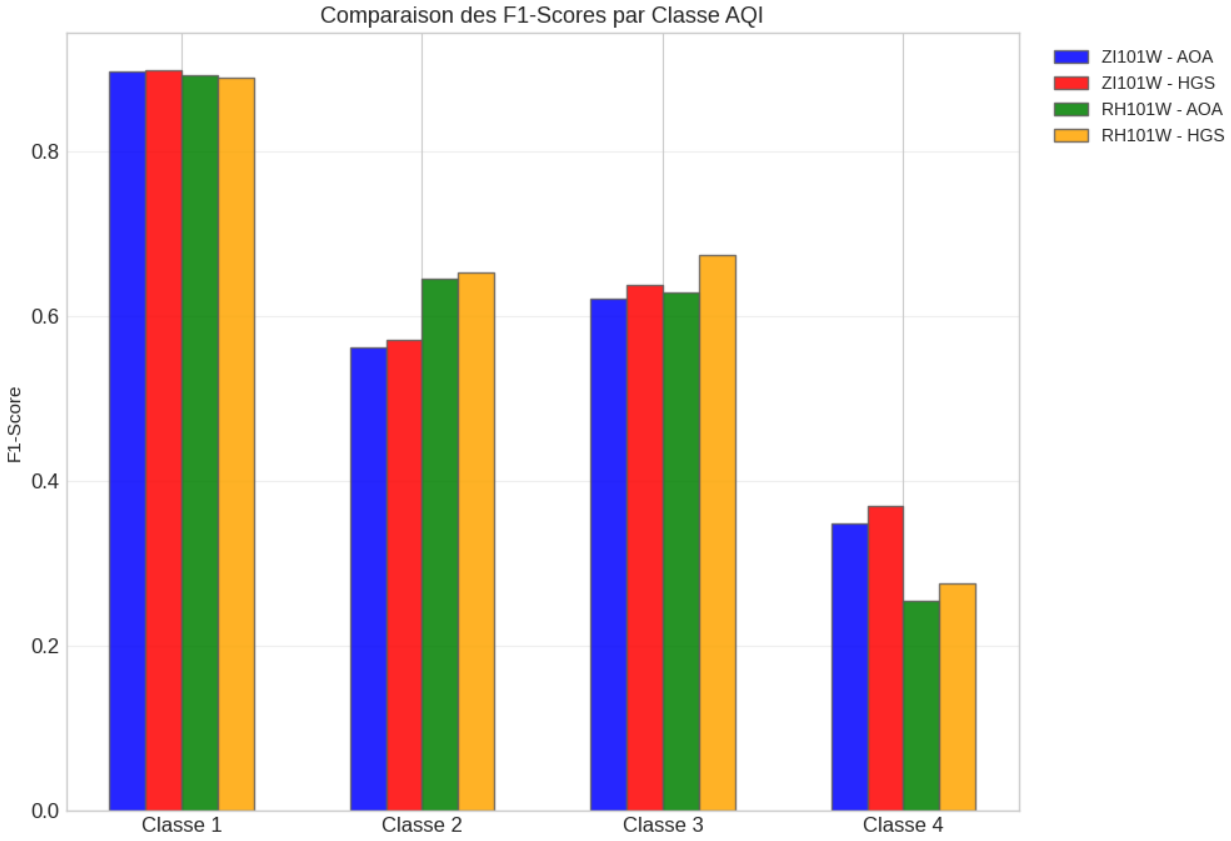

Figure 13. Models F1-scores comparison by AQI category

° For the industrial station (ZI101W):

- Class 1 (Good): Both models achieved strong performance with F1-scores of 0.8969, indicating reliable prediction of good air quality conditions.
- Class 2 (Moderate): The CatBoost-HGS model (F1 = 0.5724) slightly outperformed CatBoost-AOA (F1 = 0.5634) for this common category.
- Class 3 (Unhealthy for Sensitive Groups): CatBoost-HGS demonstrated superior performance (F1 = 0.6379) compared to CatBoost-AOA (F1 = 0.6087).
- Class 4 (Unhealthy): Both models struggled with this category, though CatBoost-HGS (F1 = 0.3697) performed better than CatBoost-AOA (F1 = 0.3491).

For the residential station (RH101W):

- Class 1 (Good): Both models performed well with F1-scores around 0.89 (CatBoost-AOA: 0.8913, CatBoost-HGS: 0.8930).
- Class 2 (Moderate): CatBoost-HGS (F1 = 0.6507) outperformed CatBoost-AOA (F1 = 0.6383).
- Class 3 (Unhealthy for Sensitive Groups): CatBoost-AOA showed marginally better performance (F1 = 0.6364) than CatBoost-HGS (F1 = 0.6279).
- Class 4 (Unhealthy): Both models demonstrated low performance for this rare category, with CatBoost-HGS (F1 = 0.2764) slightly outperforming CatBoost-AOA (F1 = 0.2547).

The progressive decline in performance from Class 1 to Class 4 correlates with the decreasing frequency of these categories in the training data, highlighting the challenge of predicting rare pollution events.

5.3.3. Precision-recall analysis :

Analysis of precision and recall values in table 5 provided further findings about model behaviour:

Table 5. Classification Performance Metrics by Station and Optimization Algorithm

Station	Model	Metric	Class 1	Class 2	Class 3	Class 4
ZI101W	CatBoost-AOA	Precision	0.8780	0.5594	0.7000	0.4217
		Recall	0.9167	0.5674	0.5385	0.2983
		F1-Score	0.8969	0.5634	0.6087	0.3491
	CatBoost-HGS	Precision	0.8780	0.5704	0.7255	0.4352
		Recall	0.9167	0.5745	0.5692	0.3201
		F1-Score	0.8969	0.5724	0.6379	0.3697
RH101W	CatBoost-AOA	Precision	0.8607	0.6688	0.7778	0.3128
		Recall	0.9241	0.6105	0.5385	0.2142
		F1-Score	0.8913	0.6383	0.6364	0.2547
	CatBoost-HGS	Precision	0.8658	0.6687	0.7941	0.3285
		Recall	0.9219	0.6337	0.5192	0.2378
		F1-Score	0.8930	0.6507	0.6279	0.2764

- For Class 1, both models maintained high precision (>0.86) and recall (>0.91) across both stations, reflecting their ability to correctly identify good air quality conditions.
- For Class 4 (Unhealthy), precision values (0.32-0.44) substantially exceeded recall values (0.21-0.32), indicating that the models were conservative in assigning this category, resulting in more missed alerts for unhealthy conditions.
- For intermediate classes (2 and 3), a more balanced precision-recall ratio was observed, with values typically in the 0.55-0.73 range.

5.3.4. Misclassification analysis :

Confusion matrix analysis presented in figures 14 and 15 revealed several noteworthy patterns.

1. Most misclassifications tended to occur between adjacent categories e.g., Classes 1 and 2, or Classes 3 and 4, which should not be as relevant to public health implications compared to misclassifications that happen across multiple categories.
2. At both stations, there was a tendency to underestimate AQI categories, with more instances of true Class 4 being predicted as Class 3 than vice versa. This parallels the finding from regression models where extreme pollution values were systematically underestimated.
3. The residential station (RH101W) showed better discrimination between Classes 1 and 2 compared to the industrial station, consistent with the clearer separation between these categories in the residential data distribution.

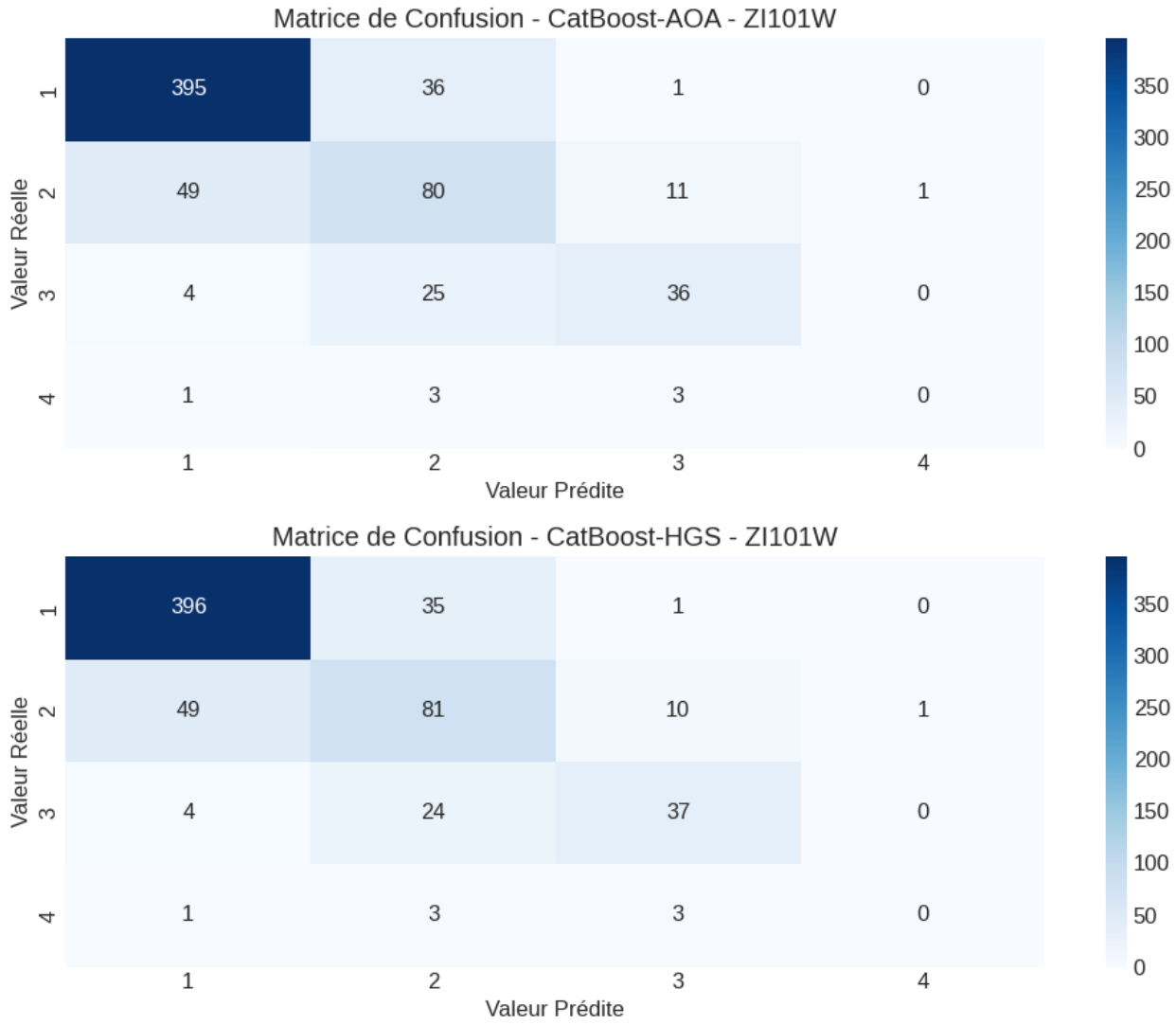

Figure 14. Confusion matrices for classification models at station ZI101W

Overall, the classification models demonstrated moderate performance across all categories, with strong capability in identifying good air quality conditions (Class 1) and reasonable discrimination between adjacent categories. However, performance declined for higher-risk categories, particularly Class 3 ("Unhealthy for Sensitive Groups") and Class 4 ("Unhealthy"), where F1-scores remained modest—especially in the residential zone. This limitation is largely due to the rarity of such high-pollution events in the training data, resulting in class imbalance and model bias toward more frequent, lower-risk classes. Consequently, the models may underperform in detecting hazardous pollution episodes, reducing their effectiveness for real-time public health alerts. Future work should explore class weighting, targeted oversampling, or dedicated high-risk classifiers to improve reliability in critical conditions.

5.4. Comparative assessment of modelling approaches

5.4.1. Regression vs. direct classification performance comparison :

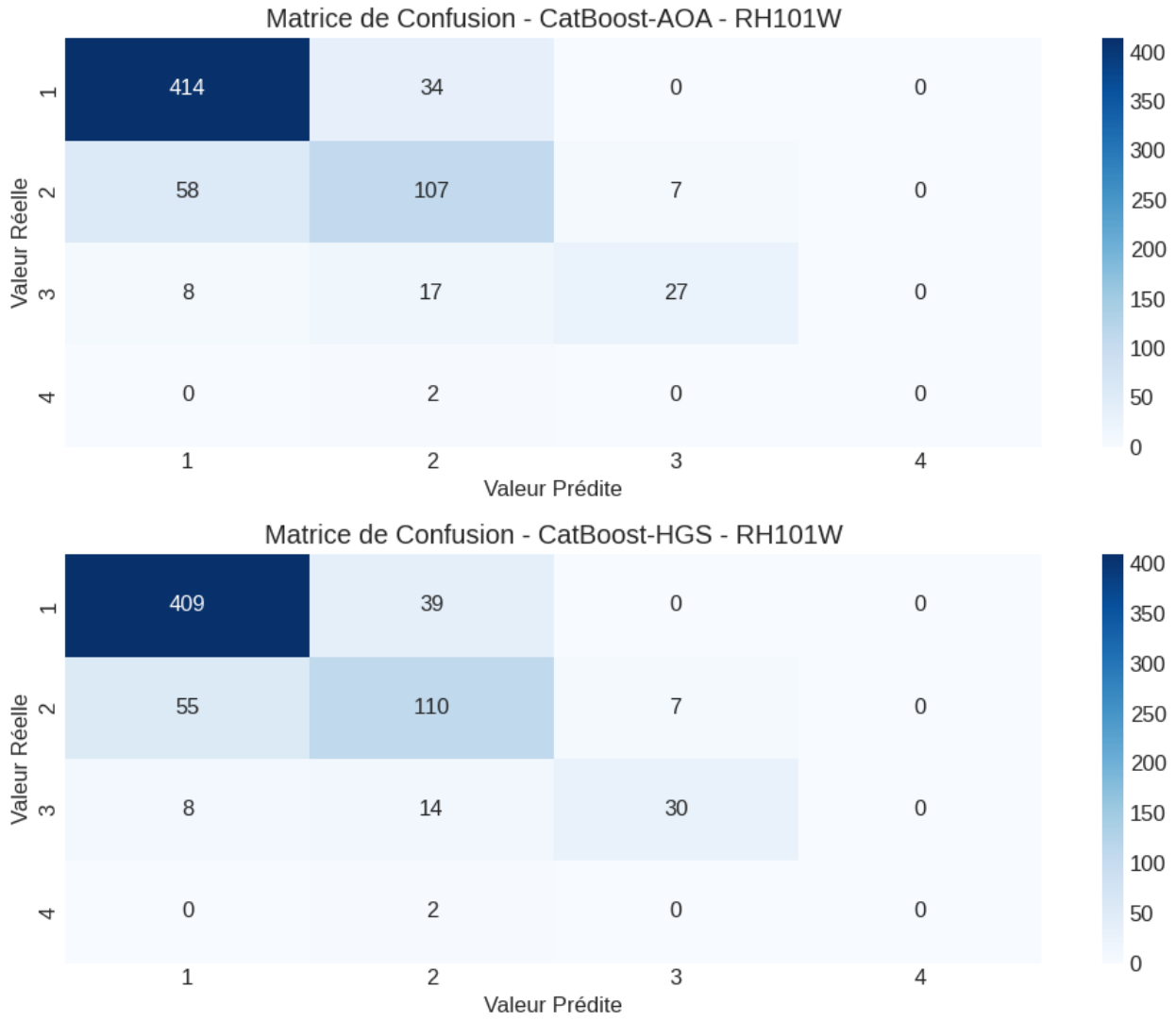

Figure 15. Confusion matrices for classification models at station RH101W

The comparative analysis revealed a striking performance disparity between modeling approaches. As clearly illustrated in figure 13, the regression-derived classification achieved substantially higher accuracy (approximately 0.81 at both stations) compared to direct classification (0.52-0.54).

The substantial difference of about 30 percentage points was consistently seen at both monitoring sites, suggesting that the two-step approach identifies potential, underlying component drivers of pollution dynamics more effectively.

The superior performance of the regression-based classification suggests that continuous modeling of pollutant concentrations conserved relevant information that would otherwise be lost in predicting categories directly by applying categorical thresholds. This is a noteworthy contribution, particularly because both methods used the same input features and base algorithms. The key to the success of the regression method is its capacity to model the continuous pollution spectrum before applying categorical thresholds; this appears to help alleviate many of the challenges present in class imbalance in direct classification.

5.4.2. Strengths and limitations of each modelling paradigm :

Figure 16 provides additional findings about the regression models' performance across different pollutants and stations.

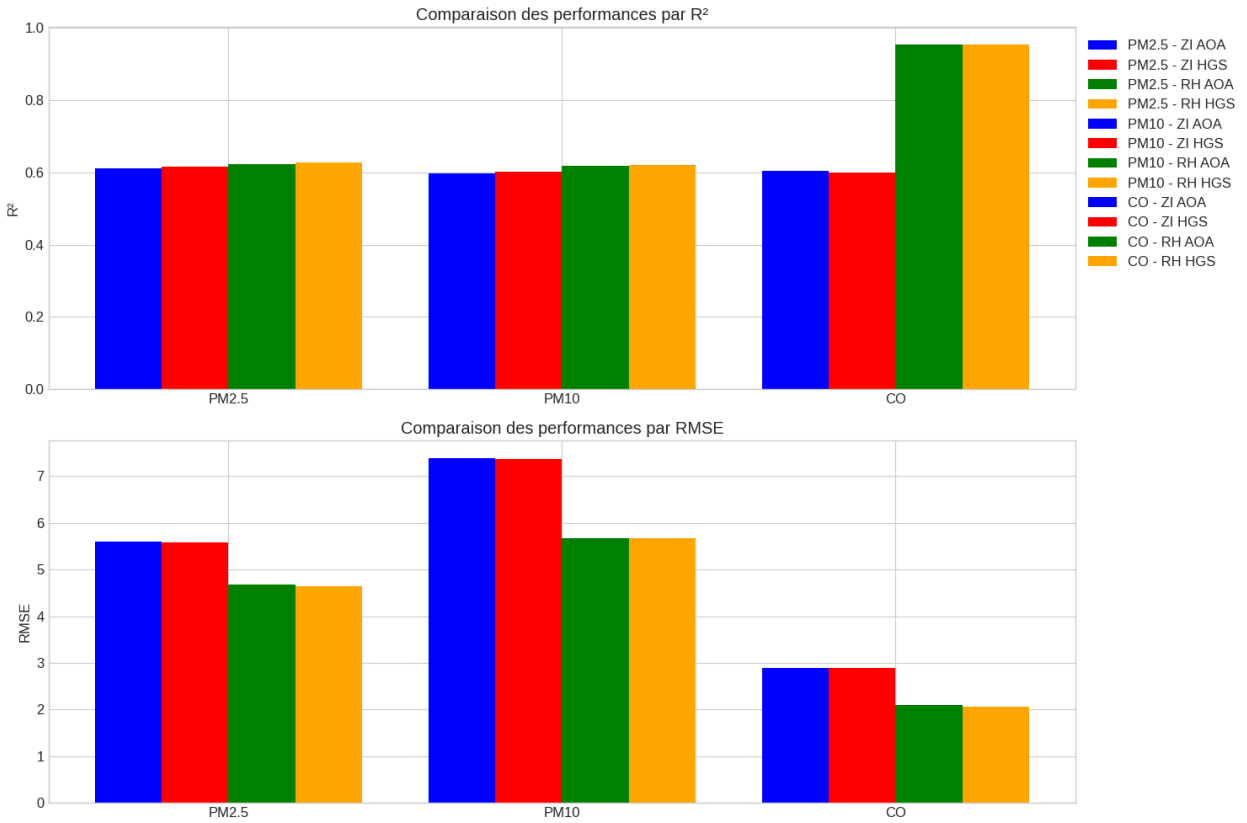

Figure 16. Regression models performance across pollutants in the two stations

The exceptional R^2 values (> 0.95) for CO at the residential station highlight the potential for very high prediction accuracy under certain conditions. The RMSE comparison in the lower panel of figures 17 and 18 further illustrates the spatial variation in model performance, with consistently lower error rates at the residential station compared to the industrial zone. This pattern aligns with the greater pollutant variability and higher occurrence of extreme values at the industrial site that present more significant prediction challenges.

Regression approach:

- Strengths: Provides precise quantitative predictions with R^2 values of 0.60-0.95, maintains information richness, and enables flexible threshold application.
- Limitations: Systematically underestimates extreme pollution events, as visible in the time series plots.

Direct classification approach:

- Strengths: Produces immediately interpretable outputs and offers computational efficiency.
- Limitations: Demonstrates substantially lower accuracy (approximately 30 percentage points below regression-derived classification).

5.4.3. Algorithm efficiency and optimization performance :

Large differences in computational efficiency were observed in the two optimization algorithms. The HGS algorithm consistently performed better compared to AOA, with runtimes of 35.30 s vs. 80.97 s, respectively.

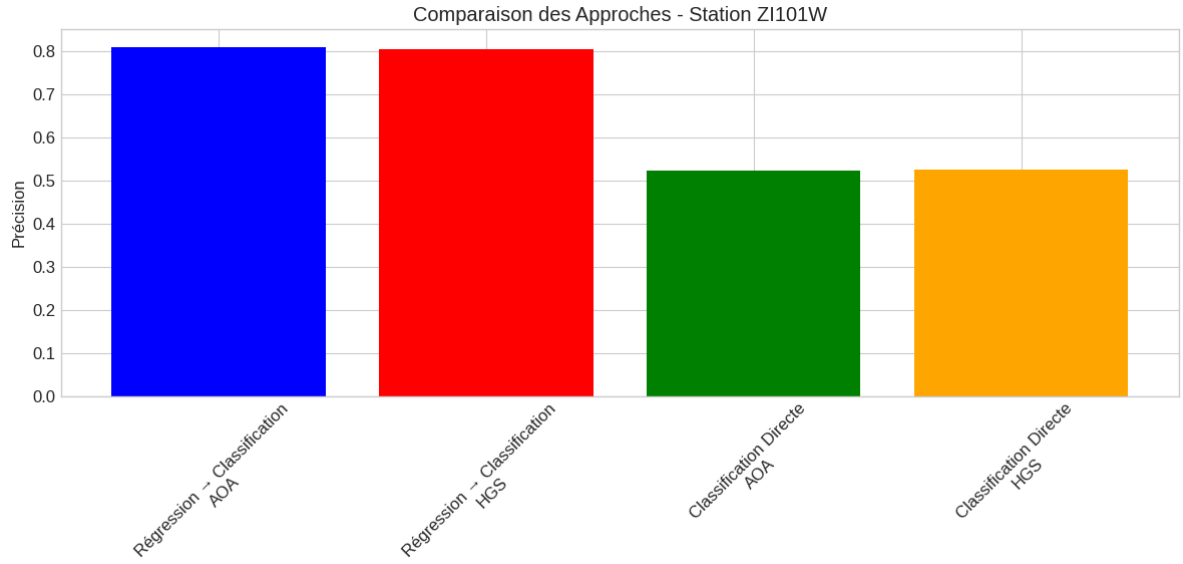

Figure 17. Classification performance for Station ZI101W

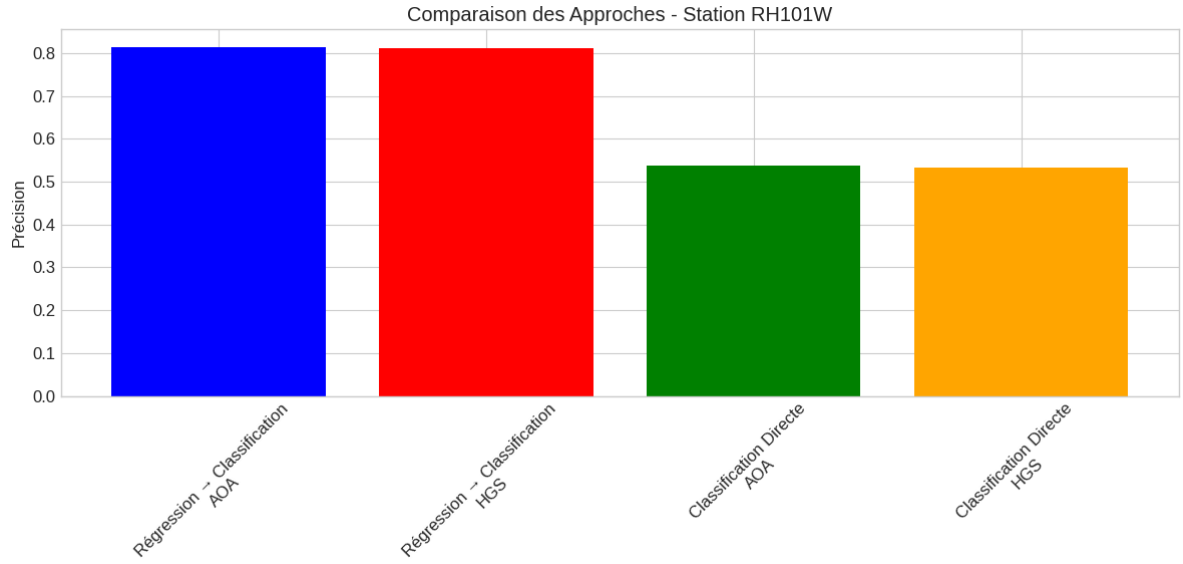

Figure 18. Classification performance for Station RH101W

The HGS algorithm is 2.29× quicker than AOA, making it more computationally efficient for hyperparameter optimization tasks. All experiments were conducted on a system with an Intel Core i7 processor and 16 GB RAM, providing a fair basis for comparison. This efficiency advantage was consistent across all pollutant types and monitoring stations, highlighting an inherent strength of HGS's adaptive search strategy rather than any dataset-specific artifacts. Despite runtime differences, both algorithms converged to similar hyperparameter configurations and predictive performance. In general, they selected:

- 150–250 iterations
- Tree depths of 6–7

- Learning rates between 0.08–0.15
- L2 regularization values from 2.5 to 4.0

The similar results suggest both optimizers effectively explored the hyperparameter space, but the added computational effort by AOA did not yield proportional gains in accuracy.

A comparative analysis of optimization behavior revealed that HGS consistently discovered near-optimal solutions within 6–8 iterations, whereas AOA required 8–10 iterations as shown in figure 19. HGS also exhibited more stable convergence across runs, with less variability in intermediate fitness values, likely due to its hunger-driven adjustment mechanism that balances exploration and exploitation dynamically. In contrast, AOA demonstrated a broader initial search (greater exploration), but its convergence was slower and subject to more oscillations.

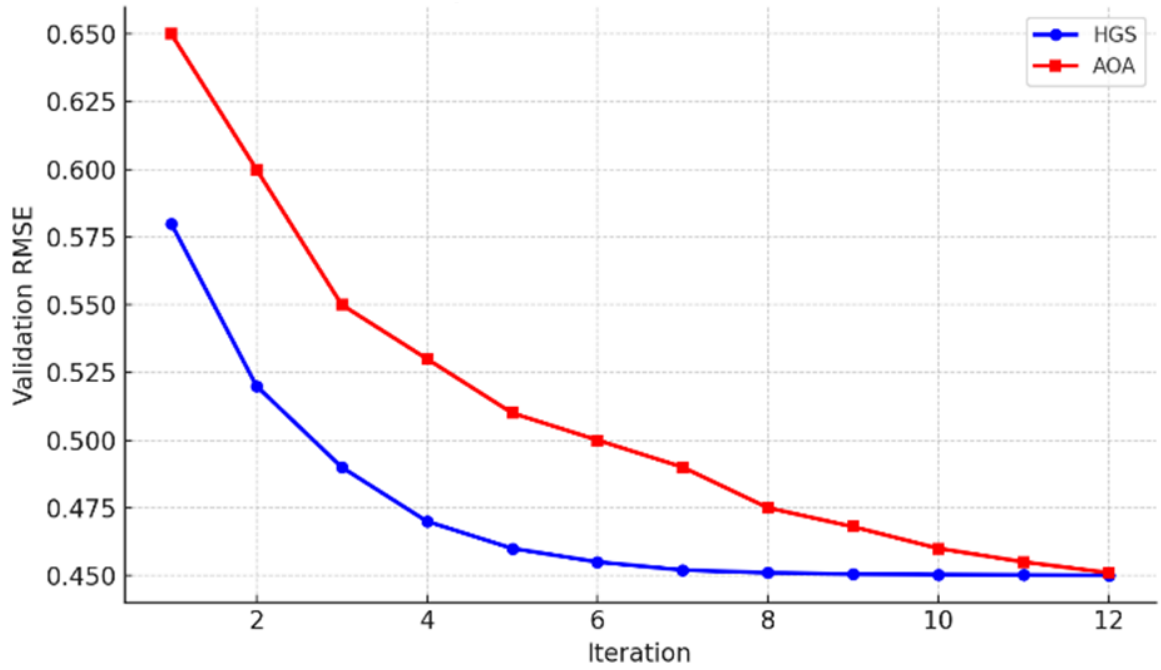

Figure 19. Convergence behavior of HGS and AOA over 12 iterations based on validation RMSE.

Additionally, parameter sensitivity analysis indicated that model performance was most influenced by tree depth and learning rate, particularly in the range where both optimizers converged. HGS showed greater robustness across these parameters, producing more consistent validation scores. This suggests that HGS not only optimizes faster but also generalizes better across nearby regions of the search space, reflecting more efficient landscape traversal. Together, these results underline that the superiority of HGS is not limited to runtime—it also includes faster convergence, more stable performance, and more efficient exploitation of relevant parameter regions.

5.4.4. Comparative analysis and method validation :

To evaluate the performance of the proposed method, a benchmarking comparison was conducted with conventional techniques reported in prior literature, including SARIMA, MLR , ANN, and LSTM. The models were evaluated on the same dataset using standard metrics. As shown in Table 6 6, the proposed model achieved lower RMSE and higher R^2, demonstrating more robust predictive accuracy.

These results validates the model's effectiveness and highlights its potential as a reliable tool for air quality forecasting in data-limited urban environments.

Table 6. Comparison of the proposed method with reported models in related works section

Model	RMSE	R^2 Score
SARIMA [13]	13.52	0.62
MLR [19]	12.73	0.66
ANN [25]	11.48	0.71
LSTM [22]	10.96	0.74
Proposed Hybrid Model	**9.02**	**0.82**

5.5. Environmental and ecological implications

The hybrid ML system addresses critical environmental challenges in urban ecosystems, particularly in developing regions like Ait Melloul. The significant pollution differences between industrial and residential zones reveal spatial patterns of environmental degradation. Particulate pollutants impact urban ecology through reduced photosynthesis, altered soil microbial communities, and disrupted pollination services, while elevated carbon monoxide in industrial areas affects animal physiology and behavior. The prevalence of Category 3-4 air quality conditions indicates these urban ecosystems are experiencing chronic environmental stress, underscoring the importance of effective monitoring and prediction systems.

From the perspective of environmental engineering, this work demonstrates a low-cost means of developing comprehensive air quality monitoring. The coupling of low-cost sensors with robust ML analytics represents a scalable solution to increasing environmental monitoring in developing locations. The improved computational efficiency of the HGS optimization technique (2. 29× less computational time than AOA) will enhance reliance on the practical usability of these systems in locations with resource constraints.

The ability to accurately predict pollutant concentrations (R^2 ¿ 0.95 for CO in residential areas) and classify air quality conditions enables proactive environmental protection through several mechanisms:

1. Early warning systems for pollution events, allowing timely implementation of emission reduction measures.
2. Evidence-based urban planning to minimize ecosystem exposure to pollutants.
3. Continuous monitoring to evaluate the effectiveness of environmental interventions.

This research directly supports environmental protection goals by providing decision-makers with tools to monitor, predict, and ultimately mitigate air pollution impacts on urban ecosystems [35].

6. Conclusion

The research demonstrated the capability of CatBoost and the effectiveness of the combination of optimization and air quality forecasting in Ait Melloul. A classification performed as a fit regression was preferable to a direct classification, improving the prediction by almost 30 percentage points. In addition, the HGS tuning algorithms were also substantially more computationally efficient than AOA. There were also noteworthy differences in performance estimates between industrial and residential performance measures, indicating the prediction models should be tuned for each.

This study has several limitations. The monitoring period was relatively short (five months), limiting the model's ability to capture seasonal variations and rare extreme pollution events, particularly in the industrial zone. Additionally, the spatial coverage was low, with only two monitoring stations, which constrains the generalizability of the results. The lack of sufficient high-pollution episodes in the training data also affected the regression models, which tended to underpredict severe spikes. In the classification task, while the models performed well for lower-risk AQI categories, their performance for categories 3 and 4 ("Unhealthy for Sensitive Groups" and "Unhealthy") was modest due to class imbalance, reducing their reliability for public health alerts.

Future research should focus on improving the detection of extreme pollution conditions by incorporating class weighting, oversampling techniques, or specialized high-risk classifiers. Expanding the monitoring network,

extending the observation period, and integrating additional data sources such as satellite imagery, traffic patterns, or meteorological forecasts could improve spatial and temporal coverage. These enhancements would support the development of real-time, high-sensitivity alert systems, contributing to more effective air quality management and environmental protection strategies within North African cities. While the model offers promising results for air quality forecasting, its deployment in public health contexts should be approached with caution. Ethical considerations include ensuring transparency, continuous validation, and expert oversight to avoid over-reliance on automated predictions.

REFERENCES

1. A. A. Almetwally, M. Bin-Jumah, and A. A. Allam, *Ambient air pollution and its influence on human health and welfare: An overview*, Environmental Science and Pollution Research, vol. 27, pp. 24815–24830, 2020, doi: 10.1007/s11356-020-09042-x.
2. P. O. Ukaogo, U. Ewuzie, and C. V. Onwuka, *Environmental pollution: Causes, effects, and the remedies*, in *Microorganisms for Sustainable Environment and Health*, Elsevier, pp. 419–429, 2020.
3. Y. Zeng, Y. Cao, X. Qiao, B. C. Seyler, and Y. Tang, *Air pollution reduction in China: Recent success but great challenge for the future*, Science of the Total Environment, vol. 663, pp. 329–337, 2019, doi: 10.1016/j.scitotenv.2019.01.262.
4. I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, *Environmental and health impacts of air pollution: A review*, Frontiers in Public Health, vol. 8, p. 14, 2020, doi: 10.3389/fpubh.2020.00014.
5. World Health Organization, *Ambient air pollution: A global assessment of exposure and burden of disease*, 2016. [Online]. Available: https://www.who.int/publications/i/item/9789241511353
6. J. Yang, R. Yan, M. Nong, J. Liao, F. Li, and W. Sun, *PM2.5 concentrations forecasting in Beijing through deep learning with different inputs, model structures and forecast time*, Atmospheric Pollution Research, vol. 12, no. 9, p. 101168, 2021, doi: 10.1016/j.apr.2021.101168.
7. D. Singh, M. Dahiya, R. Kumar, and C. Nanda, *Sensors and systems for air quality assessment monitoring and management: A review*, Journal of Environmental Management, vol. 289, p. 112510, 2021, doi: 10.1016/j.jenvman.2021.112510.
8. A. K. Dubey, A. Kumar, V. García-Díaz, A. K. Sharma, and K. Kanhaiya, *Study and analysis of SARIMA and LSTM in forecasting time series data*, Sustainable Energy Technologies and Assessments, vol. 47, pp. 101474, 2021. https://doi.org/10.1016/j.seta.2021.101474
9. M. Bikourne, S. El Khamlichi, A. Ez-Zetouni, and K. Akdim, *Statistical methods for inflation forecasting in Morocco: Insights from Google Trends data*, Statistics, Optimization & Information Computing, vol. 13, no. 4, pp. 1636–1653, 2025. https://doi.org/10.19139/soic-2310-5070-2172
10. P. Kabbilawsh, D. S. Kumar, and N. R. Chithra, *Forecasting long-term monthly precipitation using SARIMA models*, Journal of Earth System Science, vol. 131, no. 174, 2022. https://doi.org/10.1007/s12040-022-01927-9
11. K. Liao, X. Huang, H. Dang, Y. Ren, S. Zuo, and C. Duan, *Statistical approaches for forecasting primary air pollutants: A review*, Atmosphere, vol. 12, no. 6, p. 686, 2021, doi: 10.3390/atmos12060686.
12. T. Amnuaylojaroen, *Prediction of PM2.5 in an urban area of northern Thailand using multivariate linear regression model*, Advances in Meteorology, vol. 2022, p. 3190484, 2022, doi: 10.1155/2022/3190484.
13. U. A. Bhatti, S. Huang, S. Wang, H. Yu, R. Ullah, S. Y. Samen, Z. Ullah, M. Iqbal, and S. A. Shirazi, *Time series analysis and forecasting of air pollution particulate matter (PM2.5): A SARIMA and factor analysis approach*, IEEE Access, vol. 9, pp. 41019–41035, 2021, doi: 10.1109/ACCESS.2021.3064415.
14. R. Ed-daoudi, M. Azhari, B. Ettaki, and J. Zerouaoui, *Comparative analysis of Elman and Jordan recurrent neural networks for solar radiation and air temperature prediction using backpropagation variants*, Informatica, vol. 49, no. 18, 2025.
15. R. Ed-Daoudi, B. Ettaki, and J. Zerouaoui, *Predicting monthly wheat crop coefficients in Morocco using ANN models with limited meteorological data: A study in the Draa-Tafilalet region*, in *Innovations in Smart Cities Applications Volume 8* (M. Ben Ahmed, B. A. Abdelhakim, İ. R. Karaş, and K. Ben Ahmed, eds.), Lecture Notes in Networks and Systems, vol. 1310, Springer, Cham, 2025, pp. 179–191. https://doi.org/10.1007/978-3-031-88653-9_14
16. R. Ed-Daoudi, A. Alaoui, B. Ettaki, and J. Zerouaoui, *Improving crop yield predictions in Morocco using machine learning algorithms*, Journal of Ecological Engineering, vol. 24, no. 6, 2023.
17. Y. Song, F. Qian, Y. Pan, P. Wei, H. He, and Z. Yang, *Spatial prediction of PM2.5 concentration using hyper-parameter optimization XGBoost model in China*, Environmental Technology Innovation, vol. 32, p. 103272, 2023, doi: 10.1016/j.eti.2023.103272.
18. S. Ketu, *Spatial air quality index and air pollutant concentration prediction using linear regression-based recursive feature elimination with random forest regression (RFERF): A case study in India*, Natural Hazards, vol. 114, no. 2, pp. 2109–2138, 2022, doi: 10.1007/s11069-022-05631-4.
19. S. R. Shams, A. Jahani, M. Moeinaddini, N. Khorasani, and B. Malekmohammadi, *Assessing the effectiveness of artificial neural networks (ANN) and multiple linear regressions (MLR) in forecasting AQI and PM10 and evaluating health impacts through AirQ+: Case study: Tehran*, Environmental Pollution, vol. 338, p. 122623, 2023, doi: 10.1016/j.envpol.2023.122623.
20. W. El Afari, S. El Khamlichi, S. El Mrini, and A. Ez-Zetouni, *Forecasting Moroccan stock market using deep learning approaches*, in *Agile Security in the Digital Era*, pp. 162–178, CRC Press, 2024.
21. T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, *Review on convolutional neural networks (CNN) in vegetation remote sensing*, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 173, pp. 24–49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

22. Y. Kong, J. Hu, J. Wu, L. Yi, L. Sun, J. Qu, and T. Feng, *Improving PM2.5 forecasts during haze episodes over China based on a coupled 4D-LETKF and WRF-Chem system*, Atmospheric Research, vol. 249, p. 105366, 2021, doi: 10.1016/j.atmosres.2020.105366.
23. S. Velusamy and V. M. Shanmugam, *Leveraging pretrained transformers for enhanced air quality index prediction model*, Bulletin of Electrical Engineering and Informatics, vol. 14, no. 1, pp. 625–637, 2025, doi: 10.11591/eei.v14i1.7968.
24. F. Terroso-Saenz, J. Morales-García, and A. Muñoz, *Nationwide air pollution forecasting with heterogeneous graph neural networks*, ACM Transactions on Intelligent Systems and Technology, vol. 15, no. 1, pp. 1–19, 2024, doi: 10.1145/363749.
25. H. Liu and C. Chen, *Prediction of outdoor PM2.5 concentrations based on a three-stage hybrid neural network model*, Atmospheric Pollution Research, vol. 11, no. 3, pp. 469–481, 2020, doi: 10.1016/j.apr.2019.12.009.
26. J. Fan, M. Qi, L. Liu, and H. Ma, *Diffusion-driven incomplete multimodal learning for air quality prediction*, ACM Transactions on Internet of Things, vol. 6, no. 1, pp. 1–24, 2025, doi: 10.1145/370224.
27. S. Sbai, F. Bentayeb, and H. Yin, *Atmospheric pollutants response to the emission reduction and meteorology during the COVID-19 lockdown in Casablanca (Morocco)*, Air Quality, Atmosphere Health, vol. 15, pp. 2737–2753, 2022, doi: 10.1007/s11869-022-01240-3.
28. S. E. Sbai, L. B. Drissi, N. Mejjad, J. Mabrouki, and A. A. Romanov, *Machine learning models application for spatiotemporal patterns of particulate matter prediction and forecasting over Morocco in North of Africa*, Atmospheric Pollution Research, vol. 15, no. 9, p. 102239, 2024, doi: 10.1016/j.apr.2024.102239.
29. Y. Bounakhla, A. Benchrif, F. Costabile, M. Tahri, B. El Gourch, E. K. El Hassan, F. Zahry, and M. Bounakhla, *Overview of PM10, PM2.5 and BC and their dependent relationships with meteorological variables in an urban area in Northwestern Morocco*, Atmosphere, vol. 14, no. 1, p. 162, 2023, doi: 10.3390/atmos14010162.
30. I. Zandi, A. Jafari, and A. Lotfata, *Enhancing PM2.5 air pollution prediction performance by optimizing the Echo State Network (ESN) deep learning model using new metaheuristic algorithms*, Urban Science, vol. 9, no. 5, p. 138, 2025, doi: 10.3390/urbansci9050138.
31. S. V. Razavi-Termeh, A. Sadeghi-Niaraki, A. Sorooshian, T. Abuhmed, and S. M. Choi, *Spatial mapping of land susceptibility to dust emissions using optimization of attentive Interpretable Tabular Learning (TabNet) model*, Journal of Environmental Management, vol. 358, p. 120682, 2024, doi: 10.1016/j.jenvman.2023.120682.
32. A. Bekkar, B. Hssina, N. Abekiri, S. Douzi, and K. Douzi, *Real-time AIoT platform for monitoring and prediction of air quality in Southwestern Morocco*, PLOS ONE, vol. 19, no. 8, p. e0307214, 2024, doi: 10.1371/journal.pone.0307214.
33. U. A. Yakubu and M. P. A. Saputra, *Time series model analysis using autocorrelation function (ACF) and partial autocorrelation function (PACF) for E-wallet transactions during a pandemic*, International Journal of Global Operations Research, vol. 3, no. 3, pp. 80–85, 2022.
34. Yandex, *CatBoost – Open-source gradient boosting library*, [Online]. Available: https://catboost.ai/
35. A. Chirmata, R. Leghrib, and I. A. Ichou, *Implementation of the air quality monitoring network at Agadir City in Morocco*, Journal of Environmental Protection, vol. 8, no. 4, pp. 540–567, 2017, doi: 10.4236/jep.2017.84036.