Data Science for Detecting the DO Elbow: Saving Energy During Aeration Cycles

Tarik El Moudden [1]*, Meriem Tahrib [2], Bahaa Eddine Elbaghazaoui[3], and Mohamed Amnai[1]

[1]Computer Science Research Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco
[2]Laboratory of Advanced Materials and Process Engineering, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco
[3]National School of Applied Sciences, Sultan Moulay Slimane University, Beni Mellal, Morocco

Abstract. Dissolved oxygen and redox potential sensors are the main tools used to control and monitor aerobic and anoxic cycles in the biological treatment tank of a wastewater treatment plant. Just running the aeration uses a lot of electricity. In this study, data analytics and machine learning were used to collect and understand the data and then model and analyse it to find the DO-Elbow, which can help to understand biological behavior and save energy in wastewater treatment.
We used data from a cycle alternating between aeration and non-aeration phases to train a linear regression model on dissolved oxygen levels. The model achieved its lowest mean squared error of 0.9362 when the data was segmented into 20 to 100 segments. Consequently, four apparent DO-Elbows were observed during the aeration phase and one or two during non-aeration when 20 segments were used. These points indicate major changes in DO behaviour, typically associated with shifts in the biological or operational processes of the treatment system.

Keywords. Data Science in Wastewater Treatment; Aeration Process Optimization; Piecewise Linear Regression; Data Preparation for Time Series Analysis; DO-Elbow Detection using Machine Learning; Dissolved Oxygen (DO) Analysis

1 Introduction

Aeration uses a lot of electricity every day in biological wastewater treatment systems like the one in Marrakesh. It is particularly important because it supplies the oxygen to the microorganisms that transform ammonia into nitrate the oxygen they need. However, if aeration is not carefully controlled, it can lead to substantial unnecessary energy consumption. We need to cycle between aerated and non-aerated phases at the proper moments to get rid of nitrogen fast [1].

* Corresponding author: tarik.elmoudden@uit.ac.ma

It's hard to control this oxygen supply because of the intricate biological interactions that place in the treatment tanks. Two basic types of microorganisms are involved. Nitrosomonas and Nitrobacter are two of the first types of bacteria that undertake nitrification, which is the process of changing ammonia into nitrate using oxygen. These bacteria demand places with a lot of oxygen (aerobic). The second group, which includes Pseudomonas, is in responsible of denitrification, which is the process of changing nitrate into nitrogen gas without using oxygen. By doing nitrification first and then denitrification, we can be confident that all the nitrogen is gone [2,3].

The "DO-Elbow," which is a sudden jump in the amount of dissolved oxygen during aeration, is one of the most important signals that this transition is happening. When bacteria are aggressively oxidising ammonia, they need a lot of oxygen. This makes the DO level climb slowly or stay the same. When ammonia runs out, the need for oxygen goes down quickly, and the DO level goes up swiftly. The change in the DO versus time graph that appears like a kink or elbow signifies the end of the nitrification phase. This is the optimal time to halt aeration and start denitrification [4][5].

The DO-elbow becomes significantly more informative when combined with data-driven analysis and real-time monitoring. DO sensing keeps a watch on DO levels all the time, so the system can observe variations in the slope of the DO curve. Data smoothing methods like moving averages and filtering get rid of noise [6]. First-order derivatives are one type of math instrument that can determine the inflection point exactly. Once this elbow is located, automatic systems can halt aeration. This saves energy and lets denitrifying bacteria start their work in low-oxygen conditions.

In addition, machine learning may look at old data on ammonia levels and DO profiles to develop models that can make predictions. Based on several forms of influent, these models can forecast when the DO elbow will happen. This technology changes a biological signal into a strong data-based control strategy. Microbiology, sensor data, and computer analysis can all be used together by operators to help wastewater treatment plants consume less energy, keep processes stable, and perform better in general [7].

This is how the rest of this document is organised: There are subsections in Section 2, "Material and Methods," that cover "Wastewater Treatment Stages in Marrakech," "Data Collection and Understanding," "Data Preparation," and "Linear Regression for Detecting the DO-Elbow." Then, in the "Results and Discussions" section, the results and discussions are explained in more depth. Finally, the section called "Conclusions and Future Works" sums up the main points.

2 Material and methods

In this section, a Subset of the data from a Marrakech wastewater treatment plant using alternating biological aeration processes is analysed using a SCADA system. Sensors track variables such as dissolved oxygen (DO) and dissolved oxygen potential (ORP). The collected time series data were verified and pre-processed, including replacement for missing DO values, to ensure the accuracy of the analysis. A piecewise linear regression model was applied to segment the DO curve and detect DO-Elbow point points, which are essential for improving aeration control.

2.1 Wastewater Treatment Stages in Marrakech

The wastewater treatment plant in Marrakech, which serves 1.3 million people, goes first through physical pretreatment. The next step involves primary settling to eliminate suspended solids. Next, a low-load activated sludge process is used for biological treatment to get rid of nitrogen and organic matter.

In the distribution chamber, recycled sludge from secondary settling tanks is mixed with water that has already been treated. The process uses four carousel-style aeration basins that are all next to each other. Each one handles 25% of the flow. High-capacity pumps power the fine-bubble aeration systems in these basins. With alternating aeration, we can do both nitrification (with oxygen) and denitrification (without oxygen) in the same basin. After enough time, the mixed liquor goes to clarifiers, where the treated water and settled sludge are separated [4].

2.2 Data Collection and Understanding

A Schneider Quantum CPU, which is an Industrial Programmable Logic Controller (PLC), interacts with a SCADA (Supervisory Control and Data Acquisition) system to automate the numerous treatment steps. This system gets its information from transmitters, such as the SC 100 type that works with all Hach probes (see LXV404.99.00551). It runs on Areal France's Topkapi Vision 32 software, version 5.0. These transmitters watch things like the flow of air, the redox potential (with platinum electrodes), and the amount of oxygen that is dissolved. These numbers are utilised to operate fine-bubble aeration systems and on-off valve actuators.

The SCADA system collects values of DO and ORP instantly. This arrangement lets we get important features to track the DO concentration such the date, time, and on-time readings of air flow, DO concentration (mg/L), and ORP (mV) at the output of each aeration basin (B1, B2, B3, and B4) following the biological treatment stage. Understanding data means knowing what each variable means, what its value range is, and what type of data it is. The table 1 shows metadata for all the variables used in this study to back this up. This information is the basis for good data exploration, feature engineering, and predictive modelling [8].

Table 1. Metadata of the three variables used in this study

Variable	Unit	Description
DO	mg/l	Dissolved oxygen: important for assessing water quality.
ORP	mV	Oxidation–reduction potential: important for assessingwater quality and the effectiveness of disinfection.
Air Flow	Nm^3/h	In aeration systems, it represents the volume of air moving per hour.

2.3 Data Preparation

It is important to make sure that the data is good so that we can find the "DO-Elbow" in the Bassin1 dataset. This shows how DO levels change during the aeration cycles of a biological wastewater treatment process. An initial look at the missing data showed DO column had 95 missing entries. Consequently, we had to apply a domain-specific rule and a histogram of the DO distribution to fill in the missing values. The histogram showed in Figure 1, indicates a large concentration of values near to zero. Low DO levels are uncommon during aeration and are usually the result of sensor mistakes or placeholder values. Because of this, some readings (such DO < 0.2 mg/L) were thought to be missing values and were noted for correction [9].

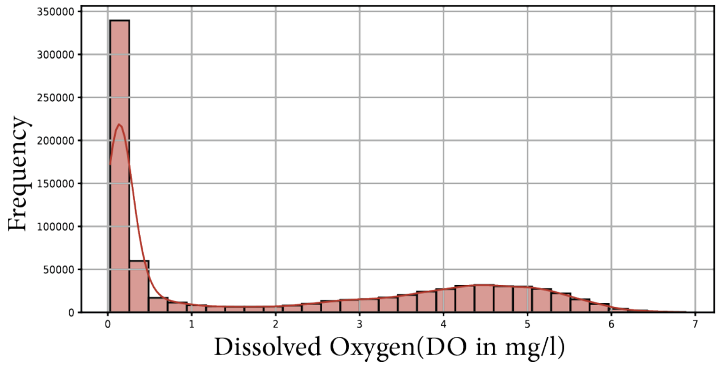

Fig.1. Distribution of Dissolved Oxygen (DO) in mg/L in Bassin_1

Indeed, from a data science perspective, such missing values in DO can severely impact the reliability of any machine learning model and data analysis, particularly curve pattern recognition and time-series modelling aimed at identifying transition points, such as the DO elbow. From a wastewater treatment standpoint, DO is a critical control parameter; missing or invalid measurements can obscure biological activity trends, delay the detection of nitrification–denitrification transitions, and lead to suboptimal aeration control. Therefore, identifying and correcting both explicit and implicit missing DO values is a necessary preprocessing step before proceeding with the identification of energy-efficient aeration phases

2.4 Linear Regression for Detecting the DO Elbow

To help determine DO-Elbows in the DO graph, we use a machine learning-based method that models how DO changes over time using piecewise linear regression [10]. The DO-Elbow is a point in the DO(t) curve where the rate of change (slope) changes significantly.

The process includes training model with different numbers of segments and preventing overfitting and instability in generalisation by monitoring the performance of the mean squared error (MSE), as shown in Figure 2. This makes it possible to find the best number of segments that show the changes in the data the best. The segmentation is not chosen at random; it is learnt from the data, which makes the decision based on data and repeatable.

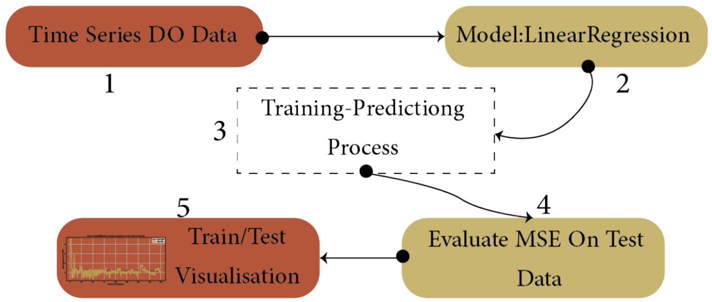

Fig.2. A Flowchart of the piecewise linear regression model. 1- The time series data of dissolved oxygen (DO) 2- the Linear Regression model. 3- Training and Prediction steps. 4- Evaluating using (MSE) on the test set. 5-visualising the optimal number of segments for DO elbow detection.

Therefore, first, the DO(t) is mathematically divided by the Equation. 1. Second from the set of elbows, we choose the slope jump and take its maximiser, as mentioned in Equation. 2.

$$DO(t) = \{a_i t + b_i,\ for\ t_{i-1} \le t < t_i, i = 1,\ldots,K\}$$

Equ.1. : a_i is the slope of linear segment, and each τ_i is a candidate DO-Elbow

$$let\ consider\ \Delta_i = \left|a_{i+1} - a_i\right|$$
$$i^* = \arg \max_{i=1\ldots.k-1} \Delta_i, \quad \tau^* = \tau_{i^*}$$

Equ. 2. τ^* is the DO-Elbow with the maximum shift in slope.

The machine learning model doesn't replace expert knowledge; it improves it. The experts can choose the best number of segments from what ML suggests and use their knowledge of the field to improve the automatically found elbows. This decision support framework for working together makes sure that changes in oxygen levels are properly recorded. This helps people decide how to run biological wastewater treatment plants and other places like them better [11].

After that, the expert picks the segmentation based on what they know and what the machine learning has learnt. Then, the next stage in generating a visualization is to put the original DO curve and the best piecewise linear approximation on top of each other. This method makes it simpler to see when the linear segment shift and helps us detect structural changes in the curve, like DO elbows. It also makes it easier to select how to regulate the aeration phase.

3 Results and Discussion

First, DO variations during *Day 11- Day 12- Day 13- Day 14* were extracted. Second, training and testing of piecewise linear models were conducted to determine the optimal number of segments for each day. Lastly, DO-Elbows were detected, reflecting biological shifts and supporting improved aeration control.

3.1 Training and Testing Piecewise Linear Models

Figure3 shows how MSE varies on the training and testing sets as the number of segments increases. The training MSE stays about the same and goes down a little bit as the number of segments increases. This means that the model fits the training data better as the segmentation gets finer. Fortunately, the overfitting is controlled during training. On the other hand, the test MSE shows more variation due to the use of shuffled test data, especially with a small number of segments, because coarse segmentation doesn't work well for generalising [12].

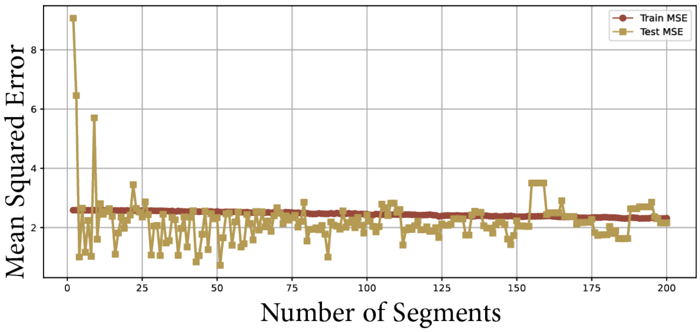

Fig.3. Train vs Test with increasing Segments in Bassin_1

The test MSE level remains at its lowest between 20 and 100 segments, where it fluctuates within a low band. This happens when the number of segments goes above about 20. This [20,100] segment shows that there is a optimal trade-off between underfitting and overfitting. After 100 segments, the test MSE starts to go up every now and then, probably because the model is overfitting small changes in the training data that don't work well for other data. So, the best number of segments seems to be between 20 and 100, where both training and test MSE are low and stable. This gives the best generalisation and keeps the underlying structure needed for reliable DO elbow detection.

The quantitative data presented in Table 2 for the DO-Elbow according to Figure 4 shows that the linear regression approach for DO-Elbow detection has high reliability, as the model successfully identified multiple elbows during both the aeration and non-aeration phases. It also demonstrated a consistent pattern of four elbows during aeration and 1-2 during non-aeration, which generally correspond to the nitrification and denitrification periods reported in the literature studies in the same context as this paper. Furthermore, this approach will help identify the likely time and pattern of nitrification and denitrification phases, which will contribute to proposing effective policies to reduce electricity consumption while maintaining the same efficiency of the wastewater treatment plant.

Table 2. Quantitative summary of DO-elbows extracted from image analysis during aeration and non-aeration phases (days 11–14).

Day	Total elbows (from the fit)	In aeration (air ON)	In non-aeration (air OFF)
11	5	4	1
12	5	4	1
13	5	4	1
14	5	3	2

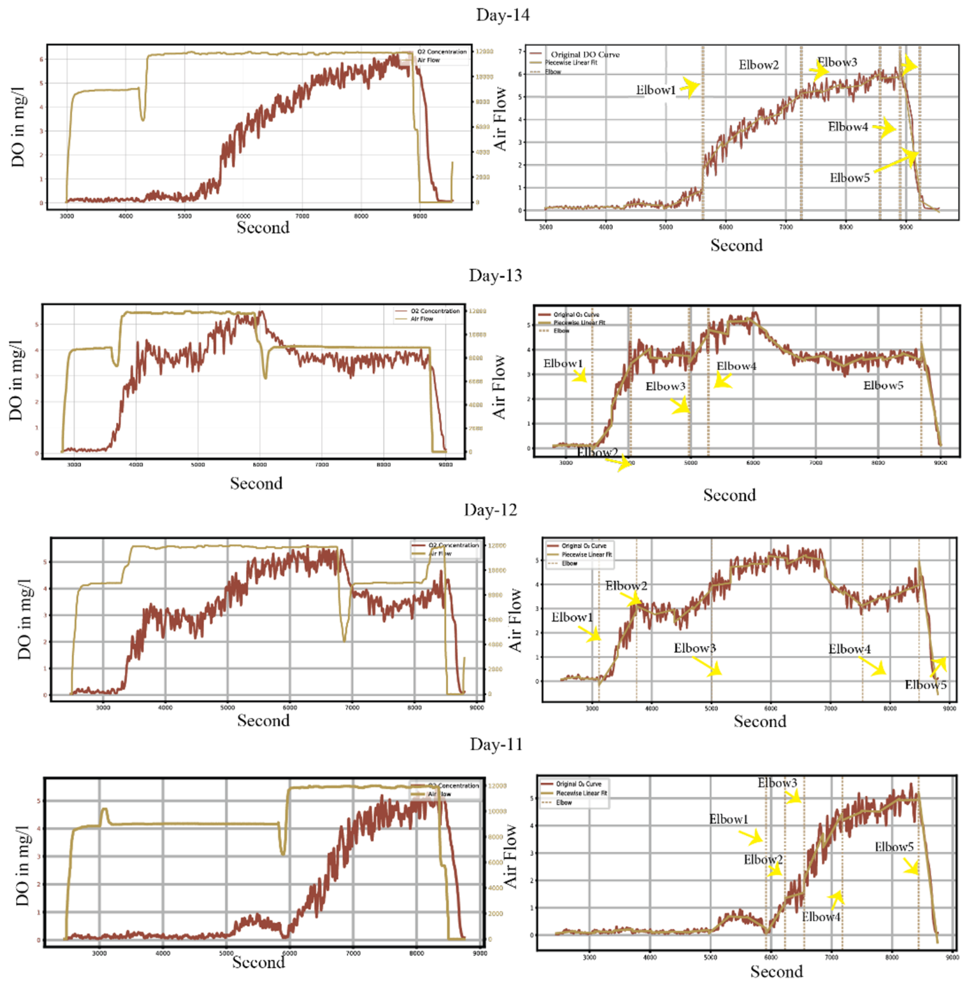

Fig. 4. DO-elbows reveal oxygen consumption variations across aeration and non-aeration phases (days 11–14) for energy optimization.

3.2 Aeration and Non-Aeration Phases and DO Trends

We chose the one-cycle from Day 11 to Day 14 to test our approach for detecting DO-Elbow, as it clearly illustrates how the aeration control system can alter the DO concentration, and detecting the elbow can significantly aid in managing electricity consumption. It's crucial to talk about these kinds of things so that elbow detection methods work better, especially when we want to spend less energy and make sure that phase transitions in the process are correct.

The subfigures in left column of Figure 4, demonstrate a connection between air flow and DO over time. The air flow curve shows patterns in on-off-on behavior pattern during the aeration phase. This could signify that the blower was shut off for a brief time or that the control system was modified, like when the dissolved oxygen reached a certain level or when the energy-saving mode was turned on. Consequently, this irregularity in lower airflow, particularly when dissolved oxygen (DO) values are between 0.3 and 3 mg/L, can make it more difficult to locate the DO-elbow visually because it causes the transition of the oxygen curve to be less steep or slows it down. This could make it harder to deploy automated detection methods that need regular aeration. However, the fact that the restarted air flow and the ongoing DO growth were in sync validates the expected biological processes. The flattening of the DO curve, which indicates that there are four elbows during aeration as mentioned in Table 2 , is a good sign that nitrification has been completed. Unfortunately, the air-off (anoxic) to launch denitrification for day 14 has two elbows, as mentioned in Table 2, which can be caused by many possibilities or reasons, such as:

- The first elbow is explained by there still being some dissolved oxygen left in the water. Heterotrophic and nitrifying bacteria quickly consume it. The second elbow is likely associated with nitrate depletion: as denitrification slows, the rate of DO decrease levels off.
- The position of the dissolved oxygen probe near a diffuser or surface area may affect its reliability by showing a second elbow when in fact the biological process is intact.
- Alternatively, the anomaly may simply reflect sensor latency or contamination, causing abnormal readings. In that case, the DO sensor should be cleaned and recalibrated.

3.3 DO-Elbow Detection and Electric Energy Consumption

During aeration with four elbows, the days in the surface, 11-14, in Figure 4, suggest that there is a switch off and switch on of the aeration, which can burn electricity with unnecessary re-aeration after the nitrification. It is noticed also, each ramp wastes a bit, and frequent restarts shorten equipment life, which can prolong more time aeration for each cycle in the day, which adds more electric energy consumption. Therefore, we propose many recommendations to reduce energy consumption, such as:

- When d(DO)/dt falls toward zero near the plateau (nitrification done or near done), it should stop aeration. This eliminates the inefficient final aeration period during which DO is merely increased toward saturation without improving treatment performance.
- Predicting the occurrence of the elbow by approximately 50 seconds allows stopping aeration at this elbow to maintain the efficiency of the treatment (full nitration) while reducing the time of excess air.
- Tracking and classifying the elbow as normal or an outlier to diagnose aeration efficiency in automatic settings of aeration. This may trigger alarms or activate adaptive control to correct the problem before energy consumption increases.

4 Conclusion and Future Works

This study focuses on identifying and understanding the DO-Elbow. We found that a segmentation range of 20 to 100 yielded the lowest MSE using piecewise linear regression. 20 segments with an MSE of 0.9362 gave us a fit that was easy to understand. The segmentation clearly revealed two distinct DO-Elbow points. These elbows indicate biological changes, likely corresponding to the start, stop, and continuation of nitrification and denitrification, or to variations in aeration operation. The information gained helps operators learn more about how microbes work and helps them control aeration in a smarter way to save energy.

However, the current work only looks at the past and doesn't try to guess how many DO-Elbows will happen in the future. The method works to find elbow points using historical data, but it doesn't yet include predictive models that can show where DO-Elbows will be in real time. In future work, we will develop a predictive machine learning framework that incorporates additional features. This will enable proactive aeration planning, making wastewater treatment plants more efficient to control.

References

1. N. Tsalas, S. K. Golfinopoulos, S. Samios, G. Katsouras, K. Peroulis, Optimization of energy consumption in a wastewater treatment plant: an overview, Energies 17 (12) (2024) 2808.
2. N.-S. Fan, Y.-H. Bai, J. Wu, Q. Zhang, J.-J. Fu, W.-L. Zhou, B.-C. Huang, R.-C. Jin, A two-stage anammox process for the advanced treatment of high-strength ammonium wastewater: Microbial community and nitrogen transformation, Journal of Cleaner Production 261 (2020) 121148.
3. Y. Liu, H. H. Ngo, W. Guo, L. Peng, D. Wang, B. Ni, The roles of free ammonia (fa) in biological wastewater treatment processes: A review, Environment international 123 (2019) 10-19.
4. M. Tahri, T. El Moudden, B. Bachiri, O. Elrhaouat, K. Elmejdoubi, M. Taky, M. El Amrani, A. Elmidaoui, Nutrient removal performance within the biological treatment of the marrakech wastewater treatment plant and characterization of the aeration and non-aeration process, Desalination and Water Treatment 257 (2022) 117-130.
5. E. Burnashova, S. Y. Semyonov, A. Sibataev, Controlled nitrification in wastewater treatment, in: IOP Conference Series: Earth and Environmental Science, Vol. 201, IOP Publishing, 2018, p. 012002.
6. A. Raudys, V. Len£iauskas, E. Mal£ius, Moving averages for financial data smoothing, in: International conference on information and software technologies, Springer, 2013, pp. 34-45.
7. Y. Xie, Y. Chen, Q. Lian, H. Yin, J. Peng, M. Sheng, Y. Wang, Enhancing real-time prediction of euent water quality of wastewater treatment plant based on improved feedforward neural network coupled with optimization algorithm, Water 14 (7)(2022) 1053.
8. N. Gerhart, R. Torres, L. Giddens, Challenges in the model development process: Discussions with data scientists, Communications of the Association for Information Systems 53 (1) (2023) 591-611.
9. J. Brownlee, Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python, Machine Learning Mastery, 2020.
10. S. Moreno-Carbonell, E. F. Sánchez-Úbeda, A piecewise linear regression model ensemble for large-scale curve tting, Algorithms 17 (4) (2024) 147.
11. P. Gerbert, J. Justus, M. Hecker, Competing in the age of articial intelligence, BCG Perspectives 48 (April) (2017) 1-7.
12. X. Ying, An overview of overfitting and its solutions, in: Journal of physics: Conference series, Vol. 1168, IOP Publishing, 2019, p. 022022.