Estimation of Dissolved Oxygen Concentration in El Quimbo Hydropower Plant

Víctor Manuel Hincapié Toro[1], Juan Pablo Romero Sánchez [1], Juan Esteban De la Calle[2], Amalia Avendaño Sánchez[3], Luis Daniel Benavides Navarro[3,*], and Agustín Marulanda Guerra[3]

[1]O&M Renewable, Enel Colombia, Bogotá DC. Colombia
[2]Independent consultant
[3]Universidad Escuela Colombiana de Ingeniería Julio Garavito, Bogotá DC, Colombia
Email: victor.hincapie@enel.com (V.M.H.T.); juan.romeros@enel.com (J.P.R.S.); juanes8400@gmail.com (J.E.C.); amalia.avendano@escuelaing.edu.co (A.A.S.); luis.benavides@escuelaing.edu.co (L.D.B.N); agustin.marulanda@escuelaing.edu.co (A.M.G.)
*Corresponding author
Manuscript received February 26, 2025; revised May 26, 2025; accepted June 9, 2025; published January 16, 2026

***Abstract*—Hydropower plants represent the largest category of renewable energy sources. However, they have a common environmental issue: they reduce the amount of dissolved oxygen in the rivers where they discharge the turbine outflow. Estimating dissolved oxygen levels in a practical manner represents an ongoing challenge for energy generation companies. This study presents a comprehensive model for predicting dissolved oxygen concentration using statistical regression techniques. The model is validated using data from El Quimbo hydropower plant in Colombia to predict dissolved oxygen in the water discharged into the Magdalena River. The approach uses Ordinary Least Squares regression to calibrate the river's dynamic model of oxygen concentration.**

The results show that the model explains 79.8% of the variability in dissolved oxygen concentration (R^2 = 0.798). The estimation of the required oxygen injection indicates that the highest demand occurs in September, reaching 39.99 tons per month. Model performance was assessed using statistical criteria, obtaining an Akaike Information Criterion (AIC) of 873.3 and a Schwarz-Bayesian Information Criterion (BIC) of 882.9. While this model provides a valuable initial tool for optimizing oxygen injection strategies to mitigate environmental impacts and ensure adequate water quality in hydroelectric projects, the paper also discusses future work on implementing a real-time control system for water oxygenation using more sophisticated machine-learning models.

***Keywords*—dissolved oxygen concentration, hydropower plant, ordinary least squares, water quality, estimation of dissolved oxygen levels**

I. Introduction

Several physical, chemical, and ecological effects are associated with dams in hydroelectric generation systems. Stratification of water and sediment trapping are well-documented physical changes that induce chemical alterations such as deoxygenation of deep reservoir water, phosphorus remobilization, and high hydrogen sulfide (H_2S) concentration [1]. Physical and chemical changes produce ecological changes in the impounded water and the river. Those ecological changes include altered thermal regimes, hypoxic stress, eutrophication, toxicity, oligotrophication, altered habitats, and shifts in fish communities. Understanding the dynamics of those changes in water quality and monitoring and managing their impact is essential for the environmental sustainability of hydroelectric generation systems. Several mathematical models are available to study the behavior of water quality in rivers and reservoirs [2–4]. However, the implementation of these models is hindered by their complexity. Companies and regulators require simpler models to monitor and manage water quality.

This paper presents a comprehensive model to predict water quality using statistical regression techniques, which enable the accurate modeling of river conditions, the characteristics of various water sources, and the characteristics of a hydropower reservoir. The model was validated using data from the El Quimbo hydropower plant in Colombia to predict the quality of water discharged into the Magdalena River.

A. Problem Statement

Dissolved Oxygen (DO) is a key indicator commonly used to assess water quality and analyze water contamination. It is essential in characterizing the aquatic environment and shows the equilibrium between the processes that produce or consume oxygen [4, 5]. Energy and environmental authorities, as well as electric generation companies, have explored several strategies to mitigate dissolved oxygen at hydropower dams [6].

However, despite the importance of DO management in hydroelectric dams and the construction boom of hydroelectric dams in low-latitude developing countries [7], most research efforts are concentrated in temperate zones. Furthermore, according to Winton *et al.* [1], the current literature contains ambiguity and misinformation regarding the stratification and mixing behavior of low-latitude reservoirs. In that study, it is shown that some works claim that stratification in low-latitude dams is uncommon. In contrast, others argue that the stratification behavior of tropical lakes and reservoirs is typical.

In instances where reservoir water exhibits stratification, it becomes essential to augment the dissolved oxygen levels in the discharged water and determine the most efficient method for DO injection.

B. Contribution

This paper presents two contributions. First, the paper proposes a model to capture the dynamics of oxygen concentration at a hydropower plant. The model is based on mass conservation principles and considers oxygen inflows, outflows, oxygen generation, consumption, and accumulation at the river downstream of the hydropower plant. Second, we propose Ordinary Least Squares (OLS) regression techniques to calibrate the river's dynamic model of oxygen concentration and determine the DO injection efficiently. The results allow optimization of the amount of oxygen injected when DO levels are insufficient. Finally, we present a case study in which we calibrate and assess the model using actual data from the El Quimbo hydropower plant, employing various statistical criteria.

The model achieves an uncentered coefficient of determination of 0.798 ($R^2 = 0.798$), indicating that the model explains approximately 79.8% of the variability in dissolved oxygen concentration. It also obtains an Akaike Information Criterion (AIC) of 873.3 and a Schwarz-Bayesian Information Criterion (BIC) of 882.9. The results are compared with other approaches and discussed as the first stage of implementing an automatic real-time control system to improve water quality and maintain optimal dissolved oxygen levels, thereby preventing negative impacts on the aquatic ecosystem.

C. Work Organization

In Section II, we present the literature review. In Section III, we describe the dissolved oxygen model in detail and the calibration methodology using Ordinary Least Squares techniques. Section IV discusses the characteristics of the case study at the Quimbo hydropower plant. Section V discusses the results, how the model was trained and calibrated to define the monthly DO concentration in the river downstream from a hydropower plant, compares the approach with more sophisticated machine learning proposals, and discusses lines of future work in the project. Finally, Section VI presents the conclusions.

II. Literature Review

A. Impact of Hydropower Operations on OD Concentration

Hydropower projects frequently encounter issues related to Dissolved Oxygen (DO) concentrations in water [8]. The construction of a dam results in a reservoir that often experiences low oxygen levels due to stratification and water stagnation [1]. To maximize power generation efficiency and maintain operational capacity during periods of drought, water directed to the turbine is typically sourced from deeper reservoir layers, corresponding to the hypolimnion. Consequently, the water discharged into the outflow channel likely contains very low DO concentrations, which are then transported downstream, causing environmental impacts, particularly on aquatic life.

Several case studies conducted in low-latitude regions suggest that tropical and subtropical rivers are particularly vulnerable to thermal and hypoxic impacts resulting from dams [1]. Cold water discharges have been shown to have severe effects on fish species in the Dartmouth [9] and Keepit [10] dams in Australia. Additionally, alterations in the reproduction process and delays in fish spawning have been documented in China [11], in Brazil, up to 34 km downstream from the Tres Marías dam [12], and at the Clanwilliam dam in South Africa [13].

DO concentration is a critical parameter in water quality assessment. It is also essential for various aquatic species and serves as a fundamental environmental indicator in evaluating river ecosystems [2, 4]. Among the most significant effects of low DO concentrations on fish are reduced growth rates, increased stress, tissue hypoxia, diminished swimming activity, lowered immunity to diseases, elevated mortality rates, and a decline in the abundance, diversity, and catch rates of fish in the affected waters [14].

Regarding hypoxia issues, Higgins and Brock [15] found that 15 out of 19 dams in the southeastern United States routinely released water with dissolved oxygen concentrations lower than 5 mg/L, and seven of them with concentrations less than 2 mg/L. Hypoxic conditions with oxygen concentrations below 5mg/L were reported at dams like Bakun in Malaysia [16, 17] and Hume in Australia [18]. Although studies have shown the significant stress hypoxia causes to various fish species [6], there are few well-documented field studies on the impact of hypoxia induced by dams on downstream ecosystems [1].

Several solutions have been developed to address the issue of anoxia in reservoirs. These solutions include deep oxygen injection systems, side-stream pumping systems, submerged contact chambers, turbine ventilation, air injection into turbines, aerating turbines, surface water pumps, aerating weirs, and multi-level water intakes [1, 8, 19]. As a result, multiple management methods exist for controlling dissolved oxygen content in discharges, with their feasibility varying depending on the specific characteristics of the dam.

B. Modeling Approaches for DO Prediction

Given the need for accurate prediction of Dissolved Oxygen (DO) levels and considering that operational reservoirs are complex hybrid systems subject to high uncertainty due to natural climatic variability, dynamic modeling has become an essential tool. These models enable the simulation of hydrological system behavior and support the development of simple yet versatile decision-making instruments. Such tools allow reservoir operators to evaluate different operational scenarios and anticipate the consequences of various management strategies [4].

Recently, several proposals have investigated mathematical models to analyze the dynamics of water quality in rivers and reservoirs. These models include AEM3D, QUAL2Kw, LAKE2K, IBER, WASP, MIKE 11, MIKE 21, QUASAR, HEC-RAS, and MINTEQ [2–4]. However, the effective implementation of these models is a complex task depending on multiple factors, such as the availability and quality of input data, the structure and dynamics of the aquatic system, and the processes being simulated. Moreover, many of these models were initially designed for specific contexts, making it difficult to adapt them to other cases.

Beyond mechanistic approaches, regression methods have also been employed to predict DO levels [20]. In this regard, Nacar *et al.* [21] applied conventional regression techniques, including linear, power, and exponential models, as well as more advanced methods, such as Multivariate Adaptive Regression Splines (MARS) and the TreeNet Gradient Boosting Machine, to estimate DO concentrations in the Broad River, South Carolina. Their findings demonstrated that these models can accurately estimate DO levels across various monitoring sites and time periods.

With recent advances in data management and artificial intelligence, numerous studies have explored the prediction of Dissolved Oxygen (DO) concentrations in aquatic systems by developing AI and machine learning models [22]. These approaches offer notable advantages, including reduced input data requirements and the ability to effectively model complex, non-linear relationships [2].

Some of the machine learning techniques used for modeling water quality, including DO, are Artificial Neural Networks (ANNs) [23, 24], Recurrent Neural Networks (RNNs) [25], Random Forest, Support Vector Machines (SVMs) [26], and Adaptive-Neuro Fuzzy Inference Systems (ANFIS) [27]. Other researchers have also employed fuzzy inference systems to model water quality [2, 22].

III. Dissolved Oxygen Model and Methodology

The main objective of this study is to develop a comprehensive model that accurately captures the dynamics of DO within the Quimbo hydropower plant. This model integrates various factors influencing DO levels, including inflows, outflows, oxygen generation, consumption, and accumulation. The selected approach is grounded in mass conservation principles and is designed to provide actionable insights for managing and optimizing the oxygenation process within the plant.

The fundamental principle guiding the modeling approach is the conservation of mass. This principle states that the amount of oxygen in a closed system (or within a control volume in an open system like a river) remains constant unless influenced by external factors such as inflows, outflows, generation, and consumption [28, 29].

Moreover, under the balance of matter theory [28], the variation in oxygen concentration within the system over time, designated as β, encapsulates the dynamic characteristics of the aquatic environment and can be mathematically represented in the form of a mass balance equation for a river system, as illustrated in Eq. (1).

$$\beta = \rho_{in} - \rho_{out} + O_g - O_c \tag{1}$$

In the equation, ρ_{in} is the amount of oxygen entering the system from various sources, including atmospheric diffusion and tributary inputs. ρ_{out} represents the amount of oxygen leaving the system, primarily through water flow downstream. O_g includes oxygen injected by the plant through mechanical means and oxygen added due to natural processes like turbulence from water movement. O_c accounts for biological and chemical processes, such as the decomposition of organic and inorganic matter, which consumes oxygen.

A. Input Data

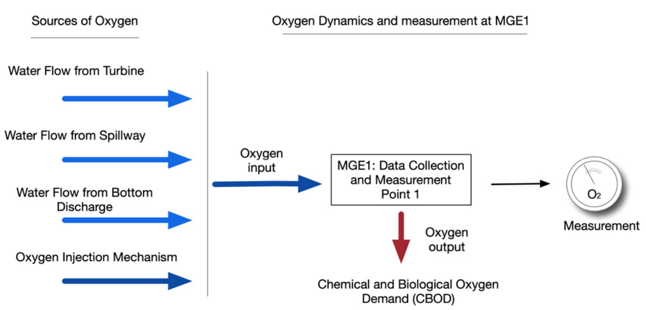

Fig. 1. Oxygen dynamics and measurement at MGE1.

Based on the principles of mass balance and mass conservation, the input data for the model can be derived from various sources, including the consumption of oxygen within the system and the corresponding flow rates. These may encompass the water within the reservoir, the water at the spillway, the turbine outflow, the downstream river, and any artificial source of oxygen injection, as illustrated in Fig. 1. The data must be filtered and classified according to a typical time window.

B. Dissolved Oxygen Model

The model for the DO dynamics in the water in a hydropower plant can be formulated as a partial differential equation based on the rate of oxygen concentration change over time, as shown in Eq. (2).

$$\begin{aligned}\frac{dC}{dt} = \beta_1 \textstyle\sum_{t=1}^{T} \sigma_t Q_t + \beta_2 \sum_{b=1}^{B} \sigma_b Q_b +\\ +\beta_3 \textstyle\sum_{s=1}^{S} \sigma_s Q_s + \beta_4 \gamma - CDO\end{aligned} \tag{2}$$

where, $\beta_1, \cdots, \beta_4$ the calibrated coefficients represent the efficiency of oxygen transfer for the turbine, bottom discharge, spillway, and artificial oxygen injection, respectively. The parameters σ_t, σ_b, and σ_s represent the oxygen concentration (mg/L) of the water flow from the turbine, the bottom discharge, and the spillway of the hydropower plant, respectively. Q_t, Q_b, and Q_s are water flow rates (m^3/s) from the turbine, the bottom discharge, and the spillway. CDO is the Chemical Demand of Oxygen (kg/d) required by biological and non-biological processes within the system, and γ is the oxygen plant's injection rate (kg/d).

Eq. (2) models the dynamic interaction between the inflow and outflow of oxygen, its generation and consumption within the system, and the resulting accumulation or depletion of DO over time.

Each variable exerts a distinct influence on the overall oxygen concentration. For instance, turbine flow rates contribute to the mixing of water and oxygen. Higher flow rates typically enhance mixing, potentially leading to increased oxygen levels. It is important to note that the bottom discharge flow rate influences the oxygen content at the bottom of the reservoir, where oxygen levels are often lower due to decomposition processes. Similarly, the spillway flow rate affects the oxygen levels at the surface and contributes to the overall oxygen dynamics through mixing and aeration.

Conversely, the oxygen concentrations at turbines indicate the oxygen levels in the water being released from these turbines. These measurements are essential for calibrating the model. While bottom discharge oxygen concentration provides insights into the oxygen levels in the deeper parts of the reservoir, spillway oxygen concentration reflects the surface oxygen levels.

Chemical and Biological Oxygen Demand represents the total oxygen required by chemical and biological processes within the water body. Higher CDO values indicate higher consumption of oxygen, which can result in lower overall oxygen levels if not adequately compensated for by generation and inflow. Efficiency coefficients are specific to the system's components (turbines, spillways, and plant) and represent the effectiveness with which oxygen is transferred into the water. Calibration of these coefficients is essential for accurate modeling.

C. Model Calibration

Calibration can be conducted using OLS regression to guarantee the model's precision. Concretely, we adjust the model coefficients based on observed data. The process has the following steps: data preparation, initial parameter estimation, regression analysis, and validation. We now explain each step in detail.

1) Data preparation

It is well known that data collection processes are susceptible to systematic and random errors. As a result, the raw data collected may not be entirely accurate and should not be used without careful consideration. Data preparation is critical in the regression algorithms to ensure data accuracy and quality.

Data preparation is collecting, labeling, and cleaning raw data to ensure its suitability for utilization in linear regression [30]. We have conducted a series of tests to confirm that the input data falls within the specified range, is in the correct format, and has been entered consistently and logically. We also verified that each field in the database has a unique entry. Fig. 2 shows the data distribution for each variable used in the model. These histograms show the average value in red, the 25th percentile in black, and the 75th percentile in yellow.

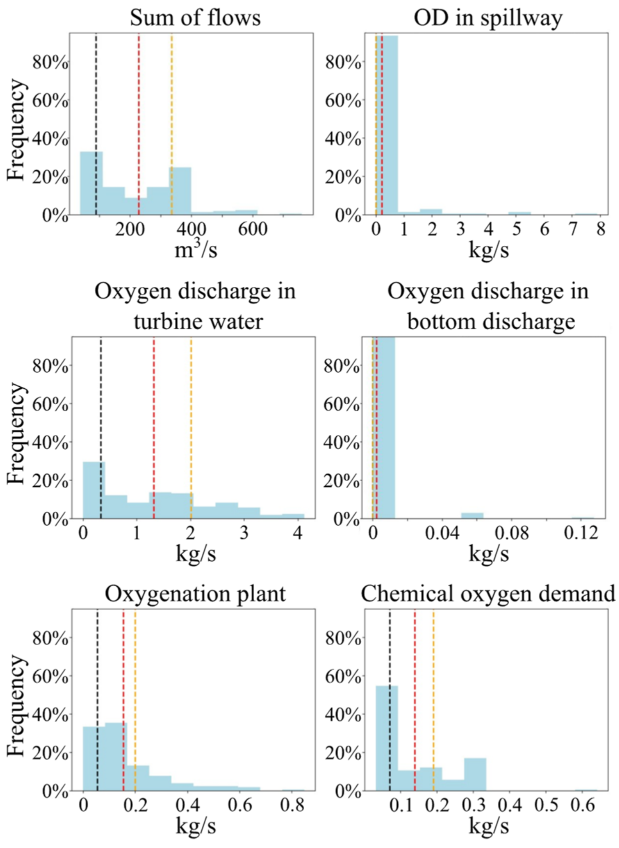

Fig. 2. Histograms of input variables.

2) Parameters estimation

Initial parameter estimation for the model coefficients is derived from empirical observations and theoretical considerations.

3) Multiple regression approach

Simple linear regression is an effective method for forecasting a response based on a single predictor variable. In many cases, such as DO estimation, the dependent variable is influenced by more than one independent variable. Rather than creating a separate simple linear regression model for each predictor, it is preferable to extend the simple linear regression model so that it can directly accommodate multiple predictors. This can be achieved by assigning a distinct slope coefficient to each predictor in a single model. In general, if there are p distinct predictors, the multiple linear regression model can be formulated as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \tag{3}$$

In this context, x_j represents the jth predictor, while β_j quantifies the association between that variable and the dependent variable y. The coefficient β_j is interpreted as the average effect on y of a one-unit increase in x_j.

According to Eqs. (2) and (3), the estimated dissolved oxygen in the downstream region of the hydropower plant can be formulated as a multiple linear regression problem, as illustrated in Eq. (4).

$$\begin{aligned} y = \textstyle\sum_{n=1}^{N} \frac{1}{Q_n} \{ \beta_1 \sum_{t=1}^{T} \sigma_t Q_t + \beta_2 \sum_{b=1}^{B} \sigma_b Q_b + \\ + \beta_3 \textstyle\sum_{s=1}^{S} \sigma_s Q_s + \beta_4 \gamma - CDO \} \end{aligned} \tag{4}$$

where, $\sum_{t=1}^{T} \sigma_t Q_t$, $\sum_{b=1}^{B} \sigma_b Q_b$, $\sum_{s=1}^{S} \sigma_s Q_s$, and γ represent the predictors p_1, p_2, p_3, and p_4, respectively. The regression coefficients β_1, β_1, β_3, and β_4 can be estimated using ordinary least squares regression, as shown in Eq. (5).

$$\begin{aligned} \min_{\beta_1, \beta_2, \beta_3, \beta_4} \Big\| y - \textstyle\sum_{n=1}^{N} \frac{1}{Q_n} \{ \beta_1 \sum_{t=1}^{T} \sigma_t Q_t + \beta_2 \sum_{b=1}^{B} \sigma_b Q_b + \\ + \beta_3 \textstyle\sum_{s=1}^{S} \sigma_s Q_s + \beta_4 \gamma - CDO \} \Big\|^2 \end{aligned} \tag{5}$$

IV. Case Study

The proposed methodology has been applied to the El Quimbo Hydropower Plant, located in Colombia, in the upper Magdalena River basin. The dam that has generated the reservoir of the El Quimbo Hydropower Power Plant is situated in the canyon formed by the Magdalena River at the rocky ridge of the Gualanda and Superior Formation at the El Quimbo site, 1300 m upstream of the confluence of the Magdalena and Paez rivers, as shown in Fig. 3.

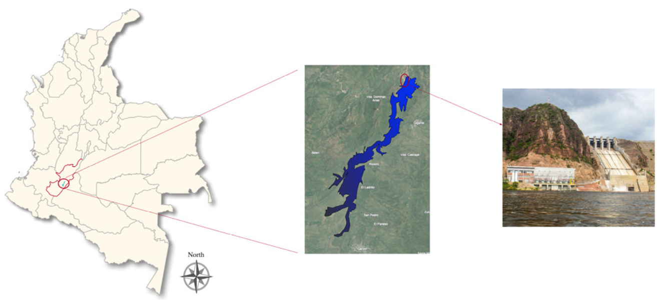

Fig. 3. Location of the El Quimbo Hydropower Plant in Huila, Colombia.

The hydropower consists of a dam, an auxiliary closure dam, a detour system, a spillway, a conveyance system, and a powerhouse at the foot of the dam. The reservoir has a length of 55 km at the maximum normal operating level (elevation 720 m above sea level), a maximum width of 4 km, and an average width of 1.4 km. The inundation area would be 8250 ha, and the total reservoir volume 3205 hm^3 and the useful storage volume is 1824 hm^3. The powerhouse is shallow at the foot of the dam, with two vertical-axis Francis turbines and a rated power of 200 MW per machine, working with a net head of 122 m and a flow of 187.5 m^3/s. The total installed capacity is about 400 MW.

V. Results and Discussion

A. Input Data

The veracity of a model depends on the quality and availability of the input data. It is, therefore, essential to ensure that data is collected continuously and with precision to maintain the model's reliability.

The data utilized in the modeling process were obtained from various monitoring points, including Spillway, Turbine Flow, and Bottom Discharge. Fig. 4 shows the locations of the MGE1 points where CDO is measured.

Fig. 4. Measurement and data collection points.

The Quimbo reservoir exhibits spatial heterogeneity in its DO levels, influenced by depth, distance from inflow points, and proximity to the plant's discharge areas. The model accounts for these spatial variations by segmenting the reservoir into distinct zones and applying localized calibration parameters. This approach ensures that the model can accurately predict oxygen levels across different parts of the river.

B. Normality Tests

Normality tests can determine whether a normal distribution fits a data set and calculate the probability that a random variable underlying the data set is normally distributed. Two types of normality tests have been used: normality analysis by hypothesis testing and quantile comparison plots.

Shapiro-Wilk, D'Agostino's K-squared, Jarque-Bera, Lilliefors, and Anderson-Darling methods were used to test the normality hypothesis. In all these methods, the p-value is greater than the predefined significance level, so the null hypothesis cannot be rejected, and the data have a normal distribution pattern.

The quantile comparison plot (qqplot) represents the quantiles of the data distribution against the theoretical quantiles that follow a normal distribution with the same mean and standard deviation as the measured data. Fig. 5 shows this comparison, where the data are aligned close to the diagonal of the plot, which allows for confirmation that the data follow a normal distribution.

C. Ordinary Least Squares Regression

Ordinary least squares regression was used to calibrate the river's dynamic model of oxygen concentration. This process involved fitting the model coefficients to the observed data to optimize the model's accuracy in predicting oxygen concentration.

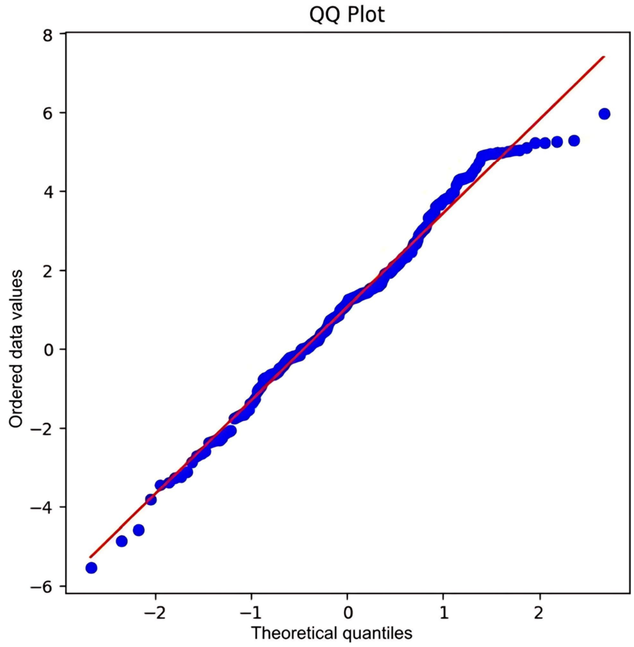

Fig. 5. Normal Quantile-Quantile plot.

Table 1 presents the model results using OLS regression with a dataset of 183 observations. The uncentered coefficient of determination (R^2 = 0.798) indicates that the model explains approximately 79.8% of the variability in dissolved oxygen concentration, suggesting an acceptable fit to the data. Additionally, the Akaike Information Criterion (AIC) and the Schwarz Bayesian Information Criterion (BIC) have values of 873.3 and 882.9, respectively.

Table 1. OLS regression results

Parameter	Value
Dep variable	y
Model	OLS
Method	Least Squares
No. Observations	183
DF Residuals	180
DF Model	3
Covariance Type	Non-robust
R-squared (uncentered)	0.798
Adj. R-squared (uncentered)	0.798
F-statistic	237.7
Prob (F-statistic)	2.4E-62
Log-Likelihood	-433.6
AIC	873.3
BIC	882.9

An R^2 of 0.798 indicates that the model explains a significant portion of the variance in dissolved oxygen concentration. Still, it also suggests that approximately 20.2% of the variability remains unexplained by the included predictors. This unexplained variation could be attributed to external factors, such as seasonal climate fluctuations, changes in oxygen demand, or measurement inaccuracies that are not directly addressed in the model. Future work may incorporate more sophisticated machine learning techniques into the project, as well as other improvements to the current model.

The evaluation metrics comprehensively assess the model's fit to the observed data. The high R-squared values indicate that the model explains a sizable portion of the variability in oxygen levels. Furthermore, the statistical significance of the model coefficients and the F-statistics provide additional support for the robustness of the model.

Table 2 shows the value of the coefficients β_1, β_3, and β_4 obtained from the least squares regression. Note that all the coefficients are statistically significant, as indicated by their p values, $p < 0.05$. This suggests that the data provide sufficient evidence to reject the null hypothesis for the entire population, thereby indicating a non-zero correlation. Consequently, changes in the independent variable are associated with changes in the dependent variable at the population level. In the case of β_2, a value of 1 is assumed because of its negligible impact on the dissolved oxygen concentration.

Table 2. Coefficient from OLS regression

Coef	Value	Std err	T	$p > \|t\|$	$\|0.025$	$0.975\|$
β_1	1.3905	0.206	6.746	0.000	0.984	1.797
β_3	1.9359	0.129	14.972	0.000	1.681	2.191
β_4	11.6412	1.009	11.536	0.000	9.650	13.632

When comparing our regression model to approaches used in previous studies, we found no dissolved oxygen (DO) models based on actual data applied to hydroelectric projects that consider both the river and the reservoir. Moreover, many studies focus on machine learning techniques, which, unlike our regression approach, assume a linear relationship between predictors and dissolved oxygen (DO) and are better suited for capturing complex non-linear interactions.

Csábrági *et al.* [10] estimated the oxygen concentration in the Tisza River using Radial Basis Function Neural Networks (RBFNN) and General Regression Neural Networks (GRNN). The results showed that machine learning techniques, particularly neural networks, achieved superior predictive performance when trained on carefully selected, homogeneous data groups. The three configurations evaluated reported RMSE values ranging from 0.73 to 1.88 mg/L, with coefficients of determination (R^2) ranging from 0.57 to 0.9. In comparison, our regression model obtained an R^2 of 0.798, indicating strong explanatory power but with limitations in capturing more complex non-linear dynamics.

While previous research suggests that neural networks offer higher accuracy, our regression model remains advantageous due to its simplicity and interpretability. It makes it more practical in scenarios where explainability is essential. Future improvements could include integrating spatially optimized training data and further exploring advanced machine learning techniques to enhance predictive accuracy.

D. Dissolved Oxygen Estimation

Two levels of certainty, 95% and 100%, are used in calculating the maximum monthly capacity required. Fig. 6 shows the quantity of oxygen required for injection in tons per month to sustain the desired dissolved oxygen levels in the river, with a 95% confidence level. The most significant demand for oxygen injection is observed in July, August, September, and October.

Fig. 7 shows the results at the 100% confidence level. This is equivalent to the highest level ever recorded in history, occurring at a rate of one occurrence per month. Note that the major requirement for oxygen injection is estimated at 39.99 tons per month in September. As the data indicates, July, August, and October are the most necessary months for oxygen injection.

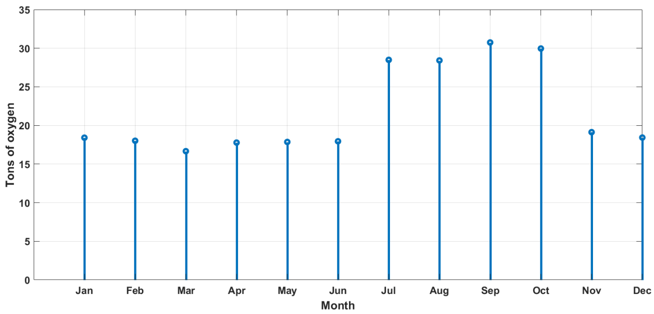

Fig. 6. Oxygen required for injection with a 95% confidence level.

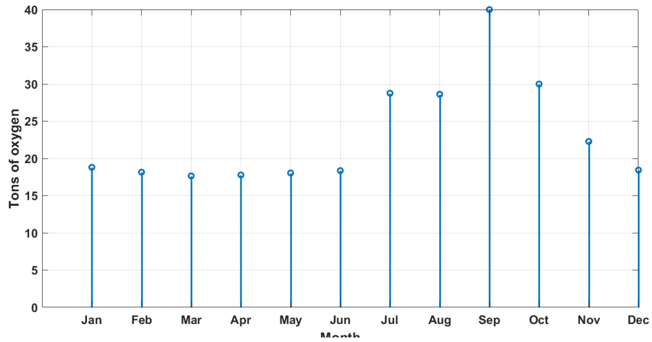

Fig. 7. Oxygen required for injection with a 100% confidence level.

The El Quimbo reservoir demonstrates spatial heterogeneity in its dissolved oxygen levels, which are influenced by depth, distance from inflow points, and proximity to the plant's discharge areas. To account for these spatial variations, the model segments the reservoir into distinct zones and applies localized calibration parameters to each zone. This approach ensures that the model can accurately predict oxygen levels across different parts of the reservoir.

E. Comparison with Other ML Models

While OLS presents a simple, explainable model, it does not account for the complex, non-linear, time-related, and seasonal dependencies found in actual water ecosystems. Recent results indicate that more advanced machine learning models may enhance the modeling of these non-linear and time-dependent relationships.

Consider, for example, models using Deep Learning Architectures, as proposed by Ma *et al.* in [31], where the authors investigate a combination of Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRUs) with a temporal attention mechanism to predict DO in aquaculture settings. This model achieved a high coefficient of determination ($R^2 = 0.9682$) and a low RMSE of 0.0249 mg/L, indicating better performance in capturing spatial and temporal features of water quality data as compared to our model. Similarly, Li *et al.* [32] introduced a hybrid model integrating CNN and Temporal Convolutional Networks (TCN) with Symplectic Geometric Mode Decomposition (SGMD) to predict DO concentrations. This approach effectively modeled temporal trends and seasonal information, enhancing prediction accuracy.

Other authors have explored Ensemble and Hybrid Models to predict DO in water systems. Zhao and Chen [33] proposed a hybrid model combining Discrete Wavelet Transform (DWT), Kernel Principal Component Analysis (KPCA), Grey Wolf Optimization (GWO), and Extreme Gradient Boosting (XGBoost) for river DO prediction. This model demonstrated significant improvements in prediction accuracy across multiple evaluation metrics. Similarly, Granata *et al.* [34] developed ensemble models, including Additive Regression with Radial Basis Function (AR-RBF) and a stacked Multilayer Perceptron with Random Forest (MLP-RF), for forecasting DO in the Mississippi River. These models provided accurate predictions, outperforming traditional methods.

Finally, some other techniques have been explored to predict DO in water. For example, Yu *et al.* [35] introduced a framework that integrates physical models with recurrent neural networks to predict lake DO concentrations. This approach dynamically adjusts timesteps to handle significant DO fluctuations, enhancing prediction robustness even with limited training data. Wu *et al.* [36] explore explainable artificial intelligence techniques for CNN models in predicting DO in the Dianchi River basin. By incorporating prior physical knowledge, the model provided interpretable insights into the factors influencing DO levels.

While the use of Ordinary Least Squares (OLS) regression for predicting Dissolved Oxygen (DO) may appear simple when compared to state-of-the-art Machine Learning (ML) techniques, the current study must be understood within the context of its real-world application. This work represents the initial phase of a larger deployment effort at the Quimbo Reservoir in Colombia, where the priority is to establish a functional and explainable baseline model as part of a real-time DO monitoring system.

VI. Conclusions

This study estimated the monthly oxygen demand for the Magdalena River in Colombia, downstream of El Quimbo's hydropower plant, using a multiple regression approach to estimate dissolved oxygen concentrations in the river. The model is the first approach to predict the oxygen demand and thus calculate the required injection rates from oxygen production plants.

The model's accuracy is highly dependent on the quality, availability, and speed of the input data. The model is fed with temporal data, but the information is not loaded in real time. Like all models, the DO dynamics model relies on certain assumptions and simplifications to make the problem tractable. In our case, the model only considers sources of oxygen, such as the turbine, bottom discharge, spillway, and artificial oxygen injection. Similarly, the model incorporates chemical and biological oxygen demand as a single variable. Thus, several factors are simplified, including seasonal considerations and ecological characteristics.

This research is the first step in a multi-staged project. As data availability and system stability increase, future project phases will incorporate and evaluate more complex ML models, such as Convolutional Neural Networks (CNNs), ensemble models (e.g., XGBoost), and process-guided learning frameworks. These will be tested and compared against the baseline in terms of predictive performance, interpretability, and deployment feasibility. Furthermore, the computational models are expected to be integrated with real-time monitoring systems to enhance the predictive capabilities of the model. Continuous data from sensors located at various points in the reservoir can be fed into the model, allowing for dynamic updates and real-time predictions. This integration facilitates proactive management of oxygenation processes, allowing timely intervention to maintain optimal DO levels.

As the next stage of this research, advanced models based on neural networks are proposed to optimize oxygen injection at the El Quimbo hydroelectric power plant. These models will enable more accurate predictions of dissolved oxygen demands, taking into account hydrological and environmental variables, as well as energy efficiency in oxygen generation. Additionally, the implementation of a Model Predictive Control (MPC) system will be explored to dynamically adjust injection rates, minimizing energy consumption while ensuring compliance with water quality standards. Integrating these advanced techniques with the current regression model will lay the foundation for an intelligent and autonomous management system capable of adapting to seasonal variations and external events, ultimately improving the energy and environmental efficiency of the project.

Following the development of the neural network-based model, Explainable Artificial Intelligence (XAI) techniques are also proposed to interpret and justify the system's predictions. This approach enables an understanding of the influence of each variable on the results, thereby strengthening the model's traceability and its integration into operational processes. Developing an early warning system for critical dissolved oxygen conditions is also proposed based on model predictions and real-time data. This will enable the anticipation of hypoxia events and the activation of preventive actions, thereby improving response capacity and reducing environmental risks.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Víctor Manuel Hincapié Toro: project administration, conceptualization, methodology, funding acquisition, review, and editing. Juan Pablo Romero Sánchez: data curation, review, and editing. Juan Esteban De la Calle: original draft preparation, review, and editing. Amalia Avendaño Sánchez: Formal analysis and investigation. Luis Daniel Benavides Navarro: Results discussion, machine learning discussion, resources, document preparation, review, and editing. Agustin Marulanda Guerra: original draft preparation, methodology, review, and editing. All authors had approved the final version.

Funding

This research was partially funded through the project "Optimización y mejoramiento del proceso de oxigenación de la central hidroeléctrica El Quimbo por medio de la implementación de tecnologías sostenibles y de un sistema de control inteligente en tiempo real para la preservación de los ecosistemas en el área de influencia", registered in the

Colombian Ministry of Science, Technology and Innovation under file 93264.

REFERENCES

[1] R. S. Winton, E. Calamita, and B. Wehrli, "Reviews and syntheses: Dams, water quality and tropical reservoir stratification," *Biogeosciences*, vol. 16, no. 8, pp. 1657–1671, 2019. doi: 10.5194/bg-16-1657-2019

[2] M. Zhong, S. Liu, K. Li, H. Jiang, T. Jiang, and G. Tang, "Modeling spatial patterns of dissolved oxygen and the impact mechanisms in a cascade river," *Frontiers in Environmental Science*, vol. 9, 2021. doi: 10.3389/fenvs.2021.781646

[3] S. Lin, L. Boegman, S. Shan, and R. Mulligan, "An automatic lake-model application using near-real-time data forcing: Development of an Operational Forecast Workflow (COASTLINES) for Lake Erie," *Geoscientific Model Development*, vol. 15, no. 3, pp. 1331–1353, 2022. doi: 10.5194/gmd-15-1331-2022

[4] O. Abouelsaad, E. Matta, M. Omar, and R. Hinkelmann, "Numerical simulation of dissolved oxygen as a water quality indicator in artificial lagoons—Case study El Gouna, Egypt," *Regional Studies in Marine Science*, vol. 56, p. 102697, 2022. doi: https://doi.org/10.1016/j.rsma.2022.102697

[5] J. Alzate, C. Duran, J. Escobar-Vargas, C. Piedrahita, and L. Montoya, "On the spatial-temporal behavior, and on the relationship between water quality and hydrometeorological information to predict dissolved oxygen in tropical reservoirs. Case Study La Miel, Hydropower Dam," Researchgate, Jan. 2023.

[6] R. S. Winton *et al.*, "Patterns and drivers of water quality changes associated with dams in the Tropical Andes," *Hydrol. Earth Syst. Sci.*, vol. 27, no. 7, pp. 1493–1505, Apr. 2023. doi: 10.5194/hess-27-1493-2023

[7] C. Zarfl, A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner, "A global boom in hydropower dam construction," *Aquatic Sciences*, vol. 77, no. 1, pp. 161–170, Jan. 2015. doi: 10.1007/s00027-014-0377-0

[8] M. S. Bevelhimer and C. C. Coutant, *Assessment of Dissolved Oxygen Mitigation at Hydropower Dams Using an Integrated Hydrodynamic/Water Quality/Fish Growth Model*, Jul. 2006. doi: 10.2172/934797

[9] C. Todd, T. Ryan, S. Nicol, and A. Bearlin, "The impact of cold water releases on the critical period of post-spawning survival and its implications for Murray Cod (Maccullochella Peelii Peelii): A case study of the Mitta Mitta River, Southeastern Australia," *River Research and Applications*, vol. 21, pp. 1035–1052, Nov. 2005. doi: 10.1002/rra.873

[10] R. Preece and H. Jones, "The effect of Keepit Dam on the temperature regime of the Namoi River, Australia," *River Research and Applications*, vol. 18, pp. 397–414, Jul. 2002. doi: 10.1002/rra.686

[11] Y. Zhong and G. Power, "Environmental impacts of hydroelectric projects on fish resources in China," *Regulated Rivers-Research & Management*, vol. 12, pp. 81–98, 1996

[12] Y. Sato, N. Bazzoli, E. Rizzo, M. Boschi, and M. Miranda, "Influence of the Abaeté River on the reproductive success of the neotropical migratory teleost Prochilodus argenteus in the São Francisco River, downstream from the Três Marias Dam, southeastern Brazil," *River Research and Applications*, vol. 21, pp. 939–950, Oct. 2005. doi: 10.1002/rra.859

[13] J. King, J. A. Cambray, and N. D. Impson, "Linked effects of dam-released floods and water temperature on spawning of the Clanwilliam yellowfish Barbus capensis," *Hydrobiologia*, vol. 384, pp. 245–265, 1998.

[14] D. Breitburg, "Effects of hypoxia, and the balance between hypoxia and enrichment, on coastal fishes and fisheries," *Estuaries*, vol. 25, pp. 767–781, Aug. 2002. doi: 10.1007/BF02804904

[15] J. M. Higgins and W. G. Brock, "Overview of Reservoir Release Improvements at 20 TVA Dams," *Journal of Energy Engineering-Asce*, vol. 125, pp. 1–17, 1999.

[16] N. Lee, L. T. Yee, and J. Grinang, "Physico-chemical characteristics in the filling phase of Bakun hydroelectric reservoir, Sarawak, Malaysia," *International Journal of Applied*, vol. 2, no. 6, 2012.

[17] F.-A. Wera, T.-Y. Ling, L. Nyanti, S. Sim, and J. Grinang, "Effects of opened and closed spillway operations of a large tropical hydroelectric Dam on the Water quality of the downstream river," *Journal of Chemistry*, vol. 2019, pp. 1–11, Jan. 2019. doi: 10.1155/2019/6567107

[18] K. Walker, T. Hillman, and W. Williams, "The effects of impoundments on rivers: An Australian case study," *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie*, vol. 20, pp. 1695–1701, Jan. 1978.

[19] M. W. Beutel and A. J. Horne, "A review of the effects of hypolimnetic oxygenation on lake and reservoir water quality," *Lake and Reservoir Management*, vol. 15, no. 4, pp. 285–297, 1999. doi: 10.1080/07438149909354124

[20] A. Banerjee, M. Chakrabarty, N. Rakshit, A. Bhowmick, and S. Ray, "Environmental factors as indicators of dissolved oxygen concentration and zooplankton abundance: Deep learning versus traditional regression approach," *Ecological Indicators*, vol. 100, Oct. 2018. doi: 10.1016/j.ecolind.2018.09.051

[21] S. Nacar, B. Mete, and A. Bayram, "Estimation of daily dissolved oxygen concentration for river water quality using conventional regression analysis, multivariate adaptive regression splines, and TreeNet techniques," *Environ Monit Assess*, vol. 192, no. 12, p. 752, Nov. 2020. doi: 10.1007/s10661-020-08649-9

[22] M. Lowe, R. Qin, and X. Mao, "A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring," *Water*, vol. 14, no. 9, 2022. doi: 10.3390/w14091384

[23] A. Csábrági *et al.*, "Estimation of dissolved oxygen in riverine ecosystems: Comparison of differently optimized neural networks," *Ecological Engineering*, vol. 138, pp. 298–309, 2019. doi: https://doi.org/10.1016/j.ecoleng.2019.07.023

[24] S. Zhu and S. Heddam, "Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: Extreme Learning Machines (ELM) versus Artificial Neural Network (ANN)," *Water Quality Research Journal of Canada*, vol. 55, Jul. 2019. doi: 10.2166/wqrj.2019.053

[25] S. V. Moghadam, A. Sharafati, H. Feizi, S. M. S. Marjaie, S. B. H. S. Asadollah, and D. Motta, "An efficient strategy for predicting river dissolved oxygen concentration: Application of deep recurrent neural network model.," *Environ Monit Assess*, vol. 193, no. 12, p. 798, Nov. 2021. doi: 10.1007/s10661-021-09586-x

[26] W. Li *et al.*, "Concentration estimation of dissolved oxygen in Pearl River Basin using input variable selection and machine learning techniques," *Sci Total Environ*, vol. 731, 139099, Aug. 2020. doi: 10.1016/j.scitotenv.2020.139099.

[27] A. N. Ahmed *et al.*, "Machine learning methods for better water quality prediction," *Journal of Hydrology*, vol. 578, 124084, 2019. doi: https://doi.org/10.1016/j.jhydrol.2019.124084

[28] N. Nuggenhalli, *Chemical Engineering Principles and Applications*. London: Springer, 2023.

[29] X. Du, J. Wang, V. Jegatheesan, and G. Shi, "Dissolved oxygen control in activated sludge process using a neural network-based adaptive PID algorithm," *Applied Sciences*, vol. 8, no. 2, 2018. doi: 10.3390/app8020261

[30] R. Croft, Y. Xie, and M. A. Babar, "Data preparation for software vulnerability prediction: A systematic literature review," *IEEE Transactions on Software Engineering*, vol. 49, no. 3, pp. 1044–1063, 2023. doi: 10.1109/TSE.2022.3171202

[31] Y. Ma, Q. Fang, S. Xia, and Y. Zhou, "Prediction of the dissolved oxygen content in aquaculture based on the CNN-GRU hybrid neural network," *Water*, vol. 16, no. 24, 2024. doi: 10.3390/w16243547

[32] W. Li, Z. Dong, T. Chen, F. Wang, and F. Huang, "Enhanced prediction of dissolved oxygen concentration using a hybrid deep learning approach with sinusoidal geometric mode decomposition," *Water Air Soil Pollut*, vol. 235, no. 7, Jul. 2024. doi: 10.1007/s11270-024-07242-x

[33] Y. Zhao and M. Chen, "Prediction of river dissolved oxygen (DO) based on multi-source data and various machine learning coupling models," *PLoS ONE*, vol. 20, no. 3, p. e0319256, Mar. 2025. doi: 10.1371/journal.pone.0319256

[34] F. Granata, S. Zhu, and F. Di Nunno, "Dissolved oxygen forecasting in the Mississippi River: advanced ensemble machine learning models," *Environ. Sci.: Adv.*, vol. 3, no. 11, pp. 1537–1551, 2024. doi: 10.1039/d4va00119b

[35] R. Yu *et al.*, "Adaptive process-guided learning: An application in predicting Lake DO concentrations," in P*roc. 2024 IEEE International Conference on Data Mining (ICDM)*, Abu Dhabi, United Arab Emirates: IEEE, Dec. 2024, pp. 580–589. doi: 10.1109/icdm59182.2024.00065

[36] J. Wu *et al.*, "Dissolved oxygen prediction in the Dianchi River basin with explainable artificial intelligence based on physical prior knowledge," *Environmental Modelling & Software*, vol. 188, p. 106412, Apr. 2025. doi: 10.1016/j.envsoft.2025.106412

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0)