Simulating Dissolved Oxygen Concentrations at the Watershed Scale: A Machine Learning Approach with Physical Constraints

Pedro Pertusso[1][0009−0008−3872−4086], Martina Pou[2][0009−0000−6710−6908], Federico Vilaseca[2][0000−0002−5471−6165], Alberto Castro[1,3][0000−0002−9174−398X], and Angela Gorgoglione[2][0000−0002−2476−2339]

[1] Department of Computer Science, School of Engineering, Universidad de la República, Montevideo 11300, Uruguay
[2] Department of Fluid Mechanics and Environmental Engineering, School of Engineering, Universidad de la República, Montevideo 11300, Uruguay
[3] Department of Electrical Engineering, School of Engineering, Universidad de la República, Montevideo 11300, Uruguay

Abstract. This study focuses on simulating dissolved oxygen (DO) concentrations at the watershed scale using machine learning (ML) models, with an emphasis on incorporating domain constraints to improve prediction accuracy. The main objectives are to evaluate the performance of different ML models, assess the impact of physical and spatial dependencies, and identify the most critical features influencing DO simulation. Random Forest (RF), Extra Trees (ET), and Histogram-based Gradient Boosting (HGB) were selected for this study and trained using a set of input variables, including water and air temperature, and other hydrological information. Model performance was assessed by calculating Mean Square Error (MSE), Mean Absolute Error (MAE), and Nash-Sutcliffe Efficiency (NSE). The best model-metric combination was selected for each station, and the results were satisfactory for most monitoring stations in the basin. The feature selection analysis, run with SHapley Additive exPlanations (SHAP), was designed to capture spatial, temporal, and physical dependencies, ensuring that the models remained accurate and aligned with established physical principles. Temperature-related variables were found to be the most significant predictors of DO levels. These outcomes demonstrate the potential of ML approaches with physical constraints to effectively predict DO concentrations and contribute to better-informed water quality management in natural watersheds.

Keywords: Water quality · Dissolved oxygen · Machine learning models · Hydroinformatics.

1 Introduction

Water quality management in urban watersheds is a critical component of sustainable development. Urbanization increases impervious surfaces, alters hydrological cycles, and intensifies pollutant loads entering aquatic systems, often leading to degraded water quality [1]. Effective monitoring and predictive modeling

of water quality are essential for supporting decision-making processes that aim to protect public health, maintain ecosystem services, and ensure the resilience of urban water supplies [2]. In the context of global sustainability goals, managing urban water quality contributes directly to achieving clean water and sanitation (SDG 6), sustainable cities and communities (SDG 11), and climate action (SDG 13), emphasizing the need for robust, data-driven tools like those developed in this study.

Dissolved oxygen (DO) is a crucial indicator of water quality, and it plays a key role in maintaining the health of the aquatic ecosystem and sustainable biodiversity [3]. In fact, adequate DO levels are essential for the survival of fish, invertebrates, and microbial communities that contribute to ecosystem functioning. On the other side, low DO concentrations can lead to hypoxic or even anoxic conditions, causing fish death, modifying biogeochemical cycles, and decreasing overall water usability for human consumption, industry, and recreation [4].

In watersheds like the Santa Lucía one, located in Uruguay (South America), which serves as the primary drinking water source for more than half of the Uruguayan population, understanding and predicting DO dynamics is essential for the sustainability of water resource management [5]. This is particularly important due to the increasing pressures from agricultural runoff, wastewater discharge, and climate variability, which contribute to DO concentration variations within the Santa Lucía watershed [6].

However, DO concentrations are influenced by a complex interaction of physical, chemical, and biological processes rather than a single influencing factor [7]. Hydrological processes, such as streamflow variability and groundwater exchanges, impact oxygen diffusion and dilution capacity. Temperature significantly affects the amount of oxygen that can dissolve in water and the rates at which microbes respire. At the same time, nutrient inputs, especially nitrogen and phosphorus from agricultural and urban activities, can cause eutrophication. This process results in a reduction in oxygen levels due to algal blooms and their subsequent decomposition. Furthermore, human activities like deforestation, land use and land cover alterations, and industrial waste discharges influence these processes, creating difficulties for managing water quality [8] [9]. Considering this complexity, the capacity to accurately model DO levels is crucial for evaluating water quality threats, identifying the most significant sources of pollution, and guiding policy decisions.

Recent advances in machine learning (ML) have demonstrated very good performance in modeling water quality parameters, including DO [10] [11]. In contrast to conventional physically-based models, machine learning techniques leverage large datasets to achieve high predictive performance by capturing complex relationships between environmental variables and DO levels [12]. Studies have applied various ML techniques, such as artificial neural networks (ANNs), random forests (RF), and support vector machines (SVMs), demonstrating their potential to improve predictive precision [13] [14]. Furthermore, physics-informed ML methods, which integrate domain knowledge into data-driven models, have gained attention to improve interpretability and reliability in environmental ap-

plications [15]. Despite these advances, there are still challenges in defining the optimal model structure, selecting relevant input features, and incorporating physical constraints to improve generalizability.

A major limitation in current research is the lack of in-depth evaluations of ML-based DO simulations at the watershed scale. While many studies focus on localized or site-specific modeling [16] [17], relatively few have investigated large-scale applications that incorporate spatial variability and different environmental conditions. Furthermore, the influence of domain constraints on ML performance remains underexplored, particularly in the context of integrating hydrological and biogeochemical principles into data-driven approaches. Addressing these gaps is crucial for improving the robustness and applicability of ML techniques in water quality assessments.

To bridge this gap, this study focuses on the Santa Lucía River basin and aims to (1) simulate DO concentrations at the watershed scale using different ML models, (2) evaluate the impact of domain constraints on simulation results, and (3) identify and quantify the key variables influencing model performance. By addressing these objectives, this research will contribute to advancing ML applications in water quality modeling by demonstrating the effectiveness of domain-informed feature selection and assessing the performance of multiple algorithms across varied water quality targets.

2 Materials and Methods

2.1 Study site

The Santa Lucía River Basin is a strategically important watershed in Uruguay, providing raw water for drinking and supplying over half of the country's population (Figure 1). It also holds substantial economic value, concentrating 32% of the national rural population and serving as one of the main food production hubs. Additionally, the basin supports significant industrial activity [18] [19].

However, human activities have led to notable water quality degradation. According to the Action Plan [20], diffuse pollution sources, primarily from agriculture (crop production, horticulture, forage crops, dairy farming, feedlot operations, and pig and poultry farms), account for approximately 75% of the total nitrogen load and 62% of the total phosphorus load. The remaining pollution originates from point sources, including industrial activities (meatpacking, dairy, and leather industries), agroindustry, and domestic wastewater discharges due to inadequate sanitation [21].

The basin under study spans 13,376 km^2, with a perimeter of 1,014 km and a compactness index of 2.46 (Figure 1). It is distributed across six departments: Florida (35%), San José (25%), Canelones (17%), Lavalleja (16%), Flores (6%), and Montevideo (1%). Elevation ranges from 390.5 meters in Lavalleja to -1.20 meters near the basin's outlet, with an average slope of 1.92%. The climate is temperate, with four distinct seasons, annual precipitation between 1,000 mm and 1,500 mm, and temperatures varying from 3°C to 30°C [5].

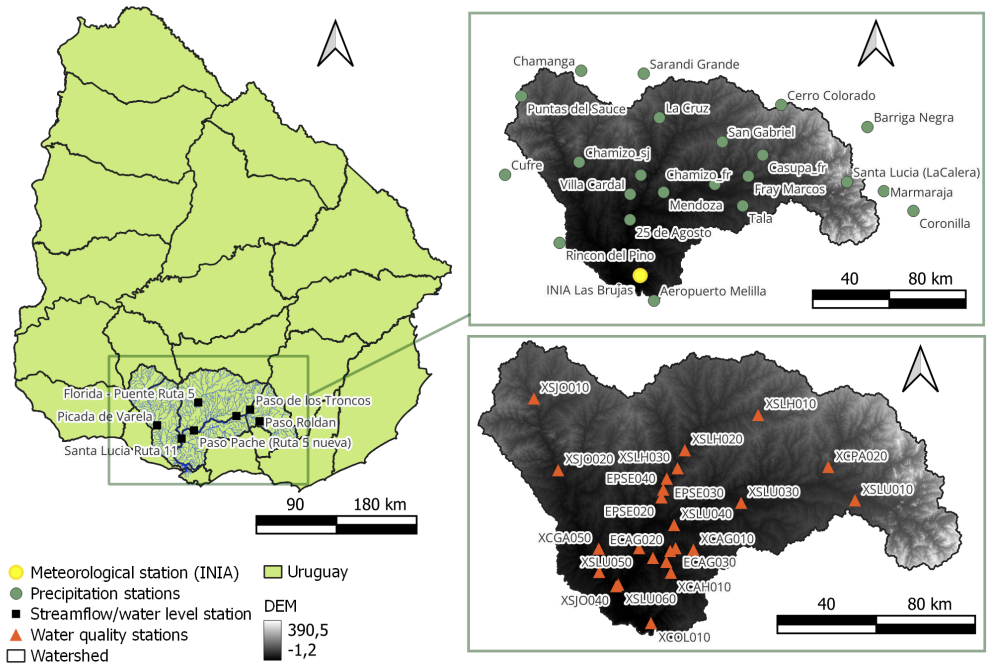


Fig. 1: Study area and location of measurement stations.

2.2 Dataset

This study utilizes hydrometric, meteorological, and water quality data obtained from national monitoring networks.

Hydrometric data consist of streamflow and water level, recorded by the Uruguayan National Water Board from January 1, 1980, to July 4, 2023. Measurements were taken at 8 monitoring stations, represented by black squares in Figure 1.

Meteorological data were gathered from two institutions: the National Institute of Agricultural Research and the Uruguayan Institute of Meteorology. The Las Brujas station provided daily records of Penman evapotranspiration, relative humidity, mean air temperature, maximum air temperature, minimum air temperature, and wind speed, covering the period from January 1, 1980, to July 4, 2023. Additionally, precipitation was recorded daily by 21 conventional rain gauges between January 1, 1980, and June 27, 2023. The meteorological stations are represented in green and yellow dots in Figure 1.

Water quality data were collected by the National Board for Quality and Environmental Assessment between January 18, 2011, and December 23, 2022, at 25 monitoring stations (orange triangles in Figure 1). From the initial 25 water quality monitoring stations, we selected nine for this study based on the availability of concurrent water quality and streamflow data, as well as their

strategic locations within the basin. The selected stations are XSLH020 (Florida, Puente Ruta 5), XSLU040 (Paso Pache, Ruta 5 nueva), XSLU050 (Santa Lucía, Ruta 11), XCPA020 (Paso de los Troncos), XSLU010 (Paso Roldán), EPSE020, XSJO010, XSJO020, and XSLH010. EPSE020 was included due to its critical position at the lake's outlet, serving as a key control point for water dynamics, while XSJO010, XSJO020, and XSLH010 were chosen to represent the upstream sections of the watershed. This selection ensures comprehensive spatial representation and reliable hydrological and water quality data for robust model development. The dataset includes key physicochemical and biological parameters: total phosphorus, total nitrogen, nitrate, nitrite, ammonium, phosphate, total solids, total suspended solids, turbidity, water temperature, dissolved oxygen, biochemical oxygen demand, chlorophyll-a, glyphosate, pH, and conductivity. This dataset is publicly accessible through the National Environmental Observatory.

Due to significant missing data, this dataset was previously imputed in our earlier work [15], and the resulting monthly dataset was used for this study.

2.3 Modeling

In this study, three machine learning models were implemented and compared to simulate DO concentrations in the Santa Lucía River Basin. Each model has different strengths in terms of accuracy, computational efficiency, and ability to capture complex environmental relationships.

Random Forest (RF) RF is an ensemble learning method based on decision trees, where multiple trees are trained using different subsets of the data, and their predictions are averaged to improve accuracy and reduce overfitting. RF has the ability to detect nonlinear correlations and handle noisy data, making it an adequate tool for environmental modeling [22].

Extra Trees Regressor (ET) ET is a variation of RF that adds more randomness to the decision tree building process. On the one hand, RF determines the optimal split points using information gain. On the other hand, ET randomly selects the split points, increasing the variance and helping to reduce the overfitting. This approach enhances computational efficiency compared to RF and is effective with high-dimensional datasets, making it a valuable tool for water quality modeling [23].

Histogram-based Gradient Boosting Regressor (HGB) HGB is an optimized version of Gradient Boosting. It improves computational efficiency by discretizing continuous features into discrete bins before training. This approach makes the training process much faster and reduces memory consumption, allowing scalability for large datasets. HGB also has the advantage of handling missing data effectively [24]. However, a very careful hyperparameter tuning is

required to avoid overfitting, and it may be less interpretable than tree-based models like RF and ET.

2.4 Model training and testing

To ensure robust model performance, we adopted a 5-fold cross-validation approach for hyperparameter optimization. This method divides the dataset into five subsets, using four for training and one for validation in each iteration, ensuring that every instance contributes to both training and validation sets. By averaging the outcomes from each fold, cross-validation assesses the model's performance, helping to decrease overfitting and enhance generalization while reducing data loss.

Furthermore, an "all-against-all" evaluation framework was applied, considering three different machine learning models (RF, ET, and HGB) and three different objective functions (Nash-Sutcliffe Efficiency (NSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE)). This approach ensured that the best model-metric combination was selected at each station, providing an in-depth assessment of predictive accuracy and robustness.

Once the optimal hyperparameters were selected, each model was trained with the complete training set and evaluated using the NSE as the primary performance metric. The final selection of the best-performing model for each target variable was based on the highest NSE value in the test set.

2.5 Model performance evaluation

Three objective functions were considered for model optimization: NSE, MAE, and MSE. A rigorous "all-against-all" approach was applied, where each model was evaluated using all three metrics. The best-performing model-metric pair was then selected individually for each monitoring station.

Nash-Sutcliffe efficiency (NSE) NSE measures how well the simulated values match observed data, with values closer to 1 indicating better performance. It is defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{1}$$

where O_i and P_i are observed and predicted values, respectively, and $\overline{O}$ is the mean of observed values.

Mean Absolute Error (MAE) MAE quantifies the average magnitude of errors without considering their direction. A lower MAE indicates better model accuracy. It is given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \tag{2}$$

Mean Squared Error (MSE) MSE penalizes larger errors more heavily than MAE by squaring the residuals. It is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2 \tag{3}$$

Kling-Gupta Efficiency (KGE) The KGE metric was also employed to verify the robustness of the model predictions. KGE combines correlation, bias, and variability components into a single efficiency score, providing a more holistic assessment of model performance. It is defined as:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{4}$$

where r is the Pearson correlation coefficient between observed and predicted values, $\alpha = \frac{\sigma_p}{\sigma_o}$ is the ratio of the standard deviation of predicted (σ_p) to observed values (σ_o), and $\beta = \frac{\mu_p}{\mu_o}$ is the bias ratio between the mean of predicted (μ_p) and observed values (μ_o) [25]. A KGE value closer to 1 indicates better agreement between observed and predicted data, accounting for both precision and bias in the simulation.

This selection process ensured that the final model configuration for each station maximized predictive accuracy while maintaining robustness.

It is important to clarify that in this study, the MSE was employed as the loss function for training the three machine learning models (RF, ET, and HGB). During the hyperparameter optimization process, the three performance metrics were utilized as objective functions in an "all-against-all" approach to identify the best model-metric combination for each station. Finally, the evaluation of the models on the testing set was conducted using NSE and KGE to assess their predictive accuracy and reliability.

2.6 Physical constraints applied to machine learning models

The process begins with the calculation of the correlation matrix for the input variables using Pearson, Spearman, and Kendall methods. Variables with a median correlation coefficient lower than 0.5 with the target variable are discarded to ensure that only the most relevant predictors are considered.

Next, both spatial and physical dependencies are taken into account. Spatial dependencies evaluate the location of monitoring stations, discarding those situated downstream of the target station to avoid information leakage. Physical dependencies consider the inherent physical relationships between variables. This means that even if a variable exhibits a correlation lower than 0.5 with the target variable, it may still be included in the model if a strong physical relationship exists. For example, water temperature (WT) is highly dependent on air temperature (AT). Therefore, even if the correlation between WT and AT is below 0.5, AT is included as an input variable in the WT prediction model due to their known physical connection [15].

Since all target variables have a monthly frequency, variables with a daily frequency are resampled to capture their monthly evolution. This is done by calculating their monthly average.

Many variables exhibit high autocorrelation. For instance, water temperature measured at a downstream station is often strongly correlated with measurements from an upstream station. To prevent information leakage and ensure model independence, input variables that contain direct or derived information about the target variable are excluded from the training process, ensuring that the models remain independent of the target variable.

Finally, additional techniques are applied to better reflect spatial and temporal relationships. The Inverse Distance Weighting (IDW) method is used to assign higher weights to observations from nearby sites, effectively reflecting spatial relationships. To address variability over time, the Exponentially Weighted Moving Average (EWMA) is implemented, assigning more importance to recent data points. These methods enhance the model's ability to identify significant patterns while ensuring physical consistency in the simulation of DO levels. A thorough description of such methods is reported in [15].

2.7 Feature importance analysis

In this study, we adopted the SHapley Additive exPlanations (SHAP) method to compute the contribution of each input feature for each ML model considered [26]. SHAP values provide an in-depth understanding of the impact that each feature has on model predictions, delivering deeper insights into how the model makes its decisions and behaves under different environmental conditions.

This analysis helps to identify not only the most influential variables but also to determine whether the model is able to capture the underlying physical processes that govern DO dynamics. By quantifying each variable's impact, SHAP enhances model interpretability and allows us to be sure that the predictions align with the real environmental processes.

3 Results and discussion

3.1 Hyperparameter optimization

To ensure optimal performance, we conducted hyperparameter optimization for the three different ML models (RF, ET, HGB). Each model was trained and evaluated using the three distinct performance metrics (NSE, MAE, MSE), resulting in a total of nine different trained models.

The optimization was performed using Optuna with a 5-fold cross-validation strategy to enhance model generalization and prevent overfitting. Table 1 summarizes the optimal hyperparameters obtained for the best models after the tuning process.

Table 1: Optimal hyperparameters for the best trained model at each station.

Station	**Trained Model**	**Hyperparameters**
EPSE020	RF (MSE)	max_depth = 22 min_samples_leaf = 5 min_samples_split = 6
XCPA020	RF (NSE)	max_depth = 25 min_samples_leaf = 3 min_samples_split = 8
XSJO010	HGB (NSE)	l2_regularization = 0.1997 learning_rate = 0.0233 max_leaf_nodes = 97
XSJO020	ET (MSE)	max_depth = 25 min_samples_leaf = 7
XSLH010	ET (MSE)	max_depth = 7 min_samples_leaf = 2 min_samples_split = 9
XSLH020	ET (NSE)	max_depth = 6 min_samples_leaf = 6 min_samples_split = 12
XSLU010	RF (MSE)	max_depth = 11 min_samples_leaf = 3 min_samples_split = 7
XSLU040	HGB (MAE)	l2_regularization = 0.6274 learning_rate = 0.0168 max_leaf_nodes = 106
XSLU050	HGB (MSE)	l2_regularization = 0.6538 learning_rate = 0.0358 max_leaf_nodes = 103

3.2 Simulation results

Once the optimal hyperparameters were identified, the models were evaluated using the best configurations obtained during the optimization process. Each model was tested under different performance metrics (NSE, MAE, and MSE) in an exhaustive pairwise evaluation approach, where all models were assessed using each metric. This allowed for a comprehensive comparison of their predictive capabilities.

After evaluating the results, the best model-metric combination was selected based on overall performance across different stations. Table 2 summarizes the final results, detailing the best-performing model, the optimal metric, and the NSE values for both training and testing at each monitoring station. Figure 2 presents boxplots comparing predicted and observed OD values across all stations, while Figure 3 presents two plots comparing simulated and observed OD time series at two stations (XSLU050 and XSLH010) as examples.

Table 2: NSE and KGE values for training and testing of the best-trained model at each station.

Station	**Best Model (Metric)**	**Train NSE**	**Train KGE**	**Test NSE**	**Test KGE**
EPSE020	RF (MSE)	0.81	0.77	0.44	0.46
XCPA020	RF (NSE)	0.82	0.80	0.81	0.90
XSJO010	HGB (NSE)	0.80	0.76	0.51	0.74
XSJO020	ET (MSE)	0.82	0.83	0.66	0.84
XSLH010	ET (MSE)	0.92	0.89	0.84	0.88
XSLH020	ET (NSE)	0.82	0.83	0.88	0.93
XSLU010	RF (MSE)	0.87	0.83	0.66	0.84
XSLU040	HGB (MAE)	0.31	0.27	-0.05	-0.15
XSLU050	HGB (MSE)	0.83	0.81	-0.02	0.43

The overall model performance across the study area was satisfactory in most cases, with 7 out of 9 stations achieving an NSE value above 0.4, indicating that the models were able to capture the variability of DO dynamics reasonably well. Among these, five stations demonstrated particularly strong performance ($NSE > 0.65$), suggesting that the chosen input variables and model configurations were well-suited for those locations.

However, two stations, XSLU040 and XSLU050, exhibited unsatisfactory results. This could be attributed to several factors, including data quality issues or unaccounted local hydrodynamic processes.

Fig. 4 shows the number of times each model and each performance metric were selected as the optimal choice across all stations.

The results indicate that no single model consistently outperformed the others, as each of the three models was selected three times. This suggests that model performance is highly dependent on the specific characteristics of each station and dataset rather than on the intrinsic superiority of one algorithm

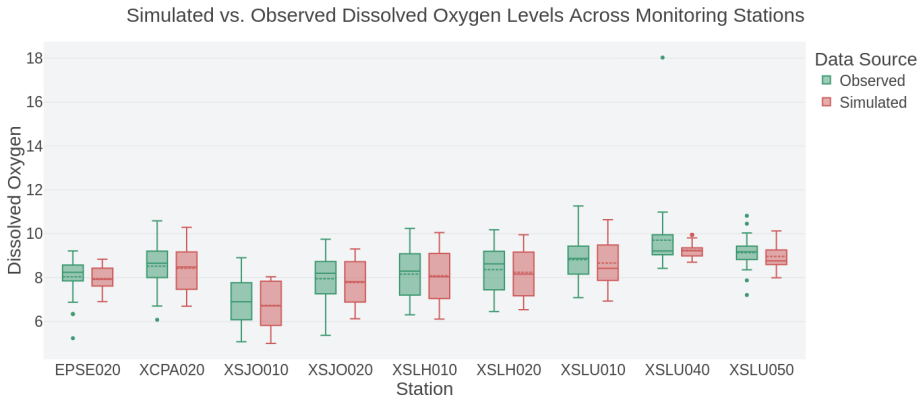


Fig. 2: Simulated and observed DO levels across monitoring stations. Green boxplots represent observed values, while red boxplots represent simulated values.

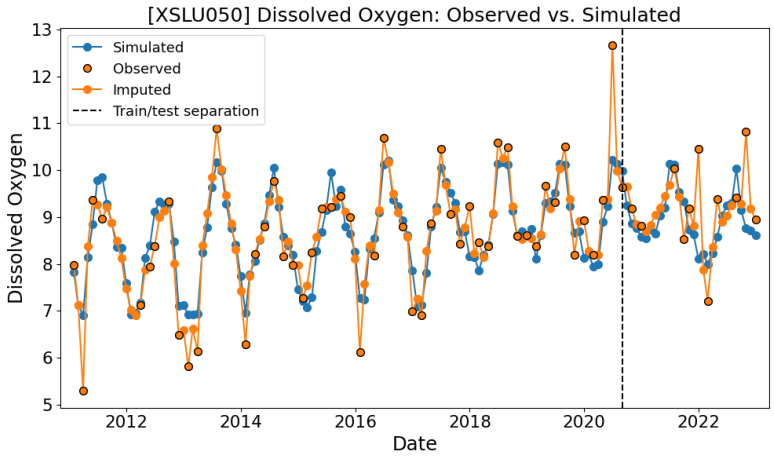
over the others. Factors such as local hydrodynamic conditions, data distribution and availability, and the influence of different input variables likely played a role in determining which model was best suited for each case. Additionally, the differences in how each algorithm handles feature interactions and nonlinearity may have contributed to this even distribution.

Regarding the performance metrics, MSE was selected as the best metric five times, compared to the MAE, which was chosen only once, and the NSE, which was selected three times. This preference for MSE may be explained by its sensitivity to large errors, which makes it more effective in optimizing models that aim to minimize extreme deviations in DO predictions. Since water quality data can exhibit occasional high variability due to sudden changes in environmental conditions (e.g., rainfall events, pollution discharges), MSE's emphasis on penalizing larger errors likely led to better model selection in most cases. In contrast, MAE gives equal weight to all errors, which may not be ideal for capturing the nuances of DO fluctuations. Meanwhile, NSE, though widely used in hydrological modeling, balances both variance and bias, but its performance may have been influenced by the characteristics of the dataset at specific stations.

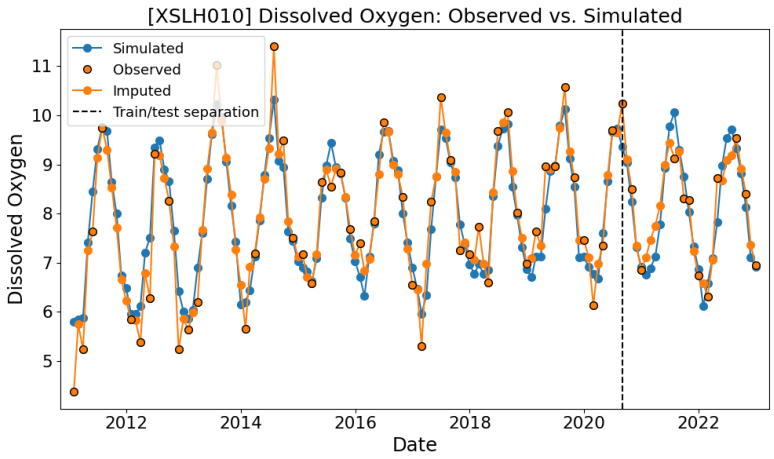
3.3 Feature importance results

The feature importance analysis was conducted using SHAP values, which were calculated based on the best-performing model at each station. This approach allowed for a detailed assessment of the contribution of each input variable to the DO predictions. For each monitoring station, the most influential variables were ranked, providing insights into the dominant drivers of DO variability across the watershed. In Figure 5, the SHAP results for the stations XSLH020 and XCPA020 are reported as an example.

The SHAP analysis revealed that the most influential variables for DO simulation were water temperature at the target station and at the nearest upstream



(a) Station XSLU050.



(b) Station XSLH010.

Fig. 3: Comparison of simulated and observed DO time series at stations (a) XSLU050 and (b) XSLH010.

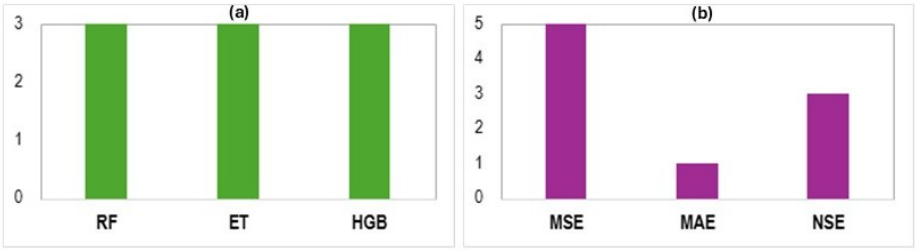


Fig. 4: Number of times each model (a) and each performance metric (b) was selected as the best.

station, along with air temperature (minimum, average, and maximum). These findings align with well-established physical and biochemical processes governing DO dynamics in freshwater systems [5].

Water temperature plays a key role in predicting DO concentration due to its direct effect on oxygen solubility and biological activity. As temperature increases, oxygen solubility decreases, leading to lower DO concentrations. Moreover, higher temperatures accelerate microbial and biochemical oxygen demand, further reducing available oxygen. The strong impact of water temperature at the target station is expected, as it directly influences local DO conditions. The importance of upstream water temperature, instead, suggests that thermal conditions propagate downstream, impacting DO levels at the target monitoring station.

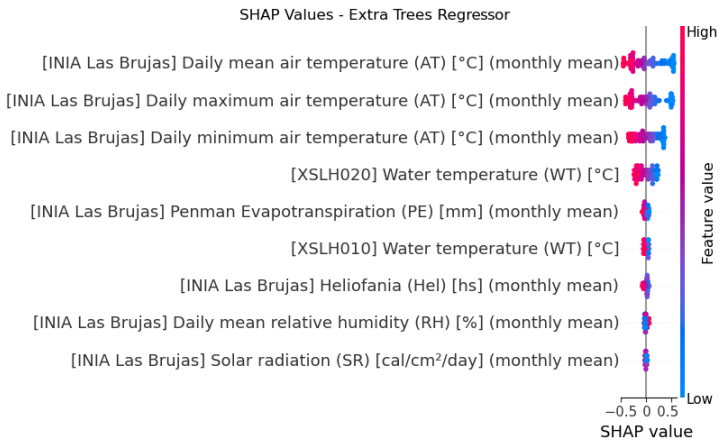
Air temperature, particularly its minimum, average, and maximum values, also emerged as a key predictor. This is consistent with its role in controlling water temperature through heat exchange processes [5]. The inclusion of multiple air temperature statistics suggests that both short-term (daily variations) and long-term (monthly trends) thermal dynamics affect DO fluctuations.

The dominance of temperature-related variables in DO predictions indicates that the model successfully captures the thermal dependency of DO dynamics.

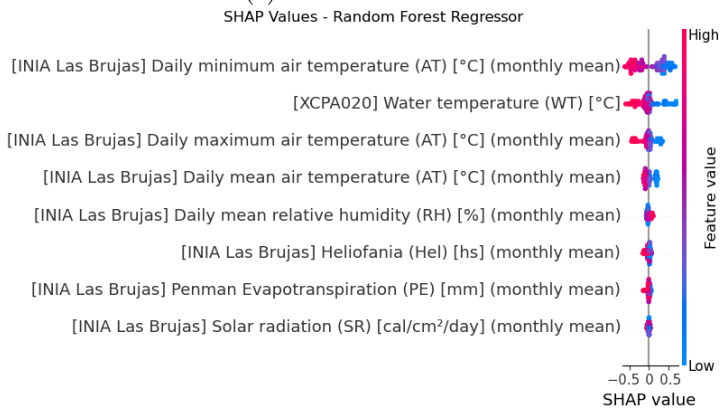
4 Conclusions

This study aimed to simulate DO concentrations at the watershed scale using ML models, assess the impact of physical constraints on the simulation results, and identify the key variables driving model performance. By incorporating domain knowledge through physical constraints, the models were able to capture essential environmental relationships, improving the realism and accuracy of DO predictions.

The models were optimized using hyperparameter tuning, and performance was evaluated through a rigorous "all-against-all" approach, selecting the best model-metric pair for each station. The results revealed no clear preference for any single model, as RF, ET, and HGB were each chosen three times. Among the



(a) Station XSLH020.



(b) Station XCPA020.

Fig. 5: SHAP values for the best-performing models at stations XSLH020 and XCPA020.

evaluation metrics, MSE was selected more frequently than the other metrics, highlighting its greater sensitivity to large errors during model training.

The models performed satisfactorily in most stations, with seven out of nine stations achieving NSE values above 0.4 and five stations yielding very good results. However, two stations (XSLU040 and XSLU050) exhibited unsatisfactory performance, possibly due to data limitations or the absence of key predictors.

The feature importance analysis shows that water temperature, both at the target station and an upstream station, along with air temperature, were the most influential variables in the DO simulations. These results align with the known physical processes governing DO dynamics, where temperature plays a crucial role in oxygen solubility and biological activity.

In conclusion, the study demonstrates the effectiveness of applying ML models with physical constraints to simulate DO concentrations at the watershed scale. The results emphasize the importance of incorporating domain-specific knowledge into model design while also pointing to the potential for future improvements by expanding the range of input variables and refining the model's physical assumptions.

Acknowledgments. This work was supported by the National Research and Innovation Agency (ANII) [grant numbers FMV-3-2022-1-172720].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. A. Gorgoglione, C. Russo, A. Gioia, V. Iacobellis, A. Castro: Comparing neural network architectures for simulating pollutant loads and first flush events in urban watersheds: Balancing specialization and generalization. Chemosphere **379**, 144395 (2025). https://doi.org/10.1016/j.chemosphere.2025.144395.
2. T.H. Boratto, D.E. Campos, D.L. Fonseca, W.A. Soares Filho, Z.M. Yaseen, A. Gorgoglione, L. Goliatt: Hybridized machine learning models for phosphate pollution modeling in water systems for multiple uses. Journal of Water Process Engineering **64**, 105598 (2024). https://doi.org/10.1016/j.jwpe.2024.105598.
3. A. Banerjee, M. Chakrabarty, N. Rakshit, A. R. Bhowmick, S. Ray: Environmental factors as indicators of dissolved oxygen concentration and zooplankton abundance: Deep learning versus traditional regression approach. Ecological Indicators **100**, 99–117 (2019). https://doi.org/10.1016/j.ecolind.2018.09.051.
4. C. Lucas, G. Chalar, E. Ibarguren, S. Baeza, S. De Giacomi, E. Alvareda, E. Brum, M. Paradiso, P. Mejía, M. Crossa: Nutrient levels, trophic status and land-use influences on streams, rivers and lakes in a protected floodplain of Uruguay. Limnologica, **94**, 125966 (2022). https://doi.org/10.1016/j.limno.2022.125966.
5. A. Gorgoglione, J. Gregorio, A. Ríos, J. Alonso, C. Chreties, M. Fossati: Influence of land use/land cover on surface-water quality of Santa Lucía river, Uruguay. Sustainability **12**(10), 4692 (2020). https://doi.org/10.3390/su12114692.
6. L. E. Aubriot, L. Delbene, S. Haakonsson, A. Somma, F. Hirsch, S. Bonilla: Evolución de la eutrofización en el Río Santa Lucía: Influencia de la intensificación productiva y perspectivas. INNOTEC (14 jul-dic), 07–16 (2017).

7. B. A. Cox: A review of dissolved oxygen modelling techniques for lowland rivers. Science of The Total Environment **314-316**, 303–334 (2003).
8. S. M. Greig, D. A. Sear, P. A. Carling: A review of factors influencing the availability of dissolved oxygen to incubating salmonid embryos. Hydrological Processes, **21**(3), 323–334 (2006).
9. M. A. Peña, S. Katsev, T. Oguz, D. Gilbert: Modeling dissolved oxygen dynamics and hypoxia. Biogeosciences, **7**(3), 933–957 (2010).
10. M. Valera, R. K. Walter, B. A. Bailey, J. E. Castillo: Machine Learning Based Predictions Of Dissolved Oxygen In A Small Coastal Embayment. Journal of Marine Sciences and Engineering, **8**(12), 1007 (2020). https://doi.org/10.3390/jmse8121007.
11. B.F. Ziyad Sami, S.D. Latif, A.N. Ahmed, M.F. Chow, M.A. Murti, A. Suhendi, B.H. Ziyad Sami, J.K. Wong, A.H. Birima, A. El-Shafie: Machine learning algorithm as a sustainable tool for dissolved oxygen prediction: a case study of Feitsui Reservoir, Taiwan. Sci Rep, **12**, 3649 (2022). https://doi.org/10.1038/s41598-022-06969-z.
12. M. Pou, M. Pastorini,J. Alonso, and A. Gorgoglione: Exploring the nexus between water quality and land use/land cover change in an urban watershed in Uruguay: a machine learning approach. Environ Sci Pollut Res, **31**, 48687-–48705 (2024). https://doi.org/10.1007/s11356-024-34414-3.
13. C. Russo, A. Castro, A. Gioia, V. Iacobellis, A. Gorgoglione: Improving the sediment and nutrient first-flush prediction and ranking its influencing factors: An integrated machine-learning framework. Journal of Hydrology **616**, 128842 (2023).
14. F. Vilaseca, A. Castro, C. Chreties, A. Gorgoglione: Assessing influential rainfall-runoff variables to simulate daily streamflow using random forest. Hydrol Sci J, **68**(12), 1738-–1753 (2023). https://doi.org/10.1080/02626667.2023.2232356.
15. M. Pastorini, R. Rodríguez, L. Etcheverry, A. Castro, A. Gorgoglione: Enhancing environmental data imputation: A physically-constrained machine learning framework. Science of The Total Environment **926**, 171773 (2024).
16. P. Barreto, S. Dogliotti, C. Perdomo: Surface Water Quality of Intensive Farming Areas Within the Santa Lucia River Basin of Uruguay. Air, Soil and Water Research **10** (2017). https://doi.org/10.1177/1178622117715446
17. I. Díaz, P. Levrini, M. Achkar, C. Crisci; C. Fernández Nion; G. Goyenola; N. Mazzeo: Empirical Modeling of Stream Nutrients for Countries without Robust Water Quality Monitoring Systems. Environments **8**(11), 129 (2021). https://doi.org/10.3390/environments8110129
18. DINOT/MVOTMA. Atlas de la cuenca del río santa lucía, 2015, https://www.gub.uy/ministerio-vivienda-ordenamiento-territorial/comunicacion/publicaciones/atlas-cuenca-del-rio-santa-lucia, last accessed 2025/2/11.
19. MVOTMA. Plan Nacional de Aguas, 2017, www.gub.uy/inisterio-ambiente/politicas-y-gestion/planes/plan-nacional-aguas, last accessed 2025/2/11.
20. SNAAC. Plan de acción para la protección de la calidad ambiental de la cuenca del río Santa Lucía. Medidas de segunda generación, 2018, https://www.gub.uy/ministerio-ambiente/sites/ministerio-ambiente/files/documentos/publicaciones/PLAN_DE_ACCION_RIO_SANTA_LUCIA_MEDIDAS_DE_2da_GENERACION.pdf, last accessed 2025/2/11.
21. R. Navas, J. Alonso, A. Gorgoglione, R.W. Vervoort: Identifying Climate and Human Impact Trends in Streamflow: A Case Study in Uruguay. Water, **11**(7), 1433 (2019). https://doi.org/10.3390/w11071433.
22. L. Breiman: Random forests. Machine learning, **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324.

23. P. Geurts, D. Ernst, L. Wehenkel: Extremely randomized trees. Machine Learning, **63**(1), 3–42 (2006). https://doi.org/10.1007/s10994-006-6226-1
24. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T. Y. Liu: LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems (NIPS 2017), **30**, 3149–3157 (2017).
25. A. Izzaddin, A. Langousis, V. Totaro, M. Yaseen, and V. Iacobellis: A new diagram for performance evaluation of complex models. Stochastic Environmental Research and Risk Assessment, **38**(6), 2261–2281 (2024).
26. S. M. Lundberg and S.-I. Lee: A unified approach to interpreting model predictions. Advances in neural information processing systems, 4768–4777 (2017)