RBRH
Revista Brasileira de Recursos Hídricos
Brazilian Journal of Water Resources
Versão On-line ISSN 2318-0331
RBRH, Porto Alegre, v. 31, e28, 2026
Scientific/Technical Article

https://doi.org/10.1590/2318-0331.312620250173

Modeling nutrients, algal blooms and fish mortality in a large reservoir with extreme hydrological variability using mechanistic, machine learning and hybrid approaches

Modelagem de nutrientes, florações de algas e mortandade de peixes em um grande reservatório com extrema variabilidade hidrológica usando abordagens mecanísticas, de aprendizado de máquina e híbridas

Iran Eduardo Lima Neto[1]

[1]Universidade Federal do Ceará, Fortaleza, CE, Brasil
E-mail: iran@deha.ufc.br (IELN)

Received: October 22, 2025 - Revised: February 22, 2026 - Accepted: March 25, 2026

ABSTRACT

Tropical semiarid reservoirs exhibit strong interannual and seasonal hydrological variability that greatly affects eutrophication. This study analyzed 15 years of monthly data from a large semiarid reservoir in Brazil, encompassing high-flow (2008-2011), drought (2012-2016), and intermediate (2017-2022) periods associated with massive algal blooms and fish mortality. Variables included inflow, outflow, volume, fish-cage production, total phosphorus (TP), total nitrogen (TN), chlorophyll-a (Chla), and fish mortality. A transient complete-mixing mechanistic model and four machine learning (ML) algorithms, linear regression (LR), Ridge regression (RR), decision tree (DT), and k-nearest neighbors (KNN), were calibrated using hydrological and aquaculture inputs to predict TP or TN. The mechanistic TP model achieved $R^2 \approx 0.30$ and PBIAS ≈ ±11%, while KNN performed best ($R^2 \approx 0.53$, PBIAS ≈ ±9%). For TN, the mechanistic and DT models outperformed the others ($R^2 \approx 0.28$, PBIAS = ±10%). Combining mechanistic or best ML (KNN, DT) models with an empirical power-law for Chla prediction (as a function of TP, TN, volume and rainfall) yielded high accuracy ($R^2 > 0.70$, PBIAS < 3%). Using only ML for Chla prediction produced poorer results ($R^2 = 0.22$, PBIAS = 17%). Fish mortality correlated with TN ($R^2 = 0.37$), and a hybrid mechanistic-DT model predicted mortality with high accuracy ($R^2 = 0.90$, PBIAS = –8.47%). Overall, model results revealed that water quality improves under higher storage volumes, lower inflow rates and higher outflow rates. Despite data limitations, the study offers robust and generalizable models for managing water quality and assessing ecological risks in reservoirs worldwide.

Keywords: Eco-hydrological modeling; Eutrophication; Semiarid reservoirs; Water quality modeling.

RESUMO

Reservatórios tropicais semiáridos apresentam forte variabilidade hidrológica interanual e sazonal, o que afeta significativamente a eutrofização. Este estudo analisou 15 anos de dados mensais de um grande reservatório semiárido no Brasil, abrangendo períodos de alta vazão (2008-2011), seca (2012-2016) e regime intermediário (2017-2022), associados a intensas florações de algas e mortandades de peixes. As variáveis incluíram afluência, defluência, volume, produção em tanques-rede, fósforo total (PT), nitrogênio total (NT), clorofila-a (Cla) e mortandade de peixes. Um modelo mecanístico transiente de mistura completa e quatro algoritmos de aprendizado de máquina (AM), regressão linear (RL), regressão Ridge (RR), árvore de decisão (AD) e k-vizinhos mais próximos (KNN), foram calibrados usando dados de entrada hidrológicos e de produção aquícola para prever PT ou NT. O modelo mecanístico de PT obteve $R^2 \approx 0{,}30$ e PBIAS = ±11%, enquanto o KNN apresentou o melhor desempenho ($R^2 \approx 0{,}53$, PBIAS ≈ ±9%). Para NT, os modelos mecanístico e AD superaram os demais ($R^2 \approx 0{,}28$, PBIAS = ±10%). A combinação dos modelos mecanístico ou dos melhores modelos de AM (KNN, AD) com uma relação empírica de lei de potência para prever Cla (em função de PT, NT, volume e precipitação) apresentou alta precisão ($R^2 > 0{,}70$, PBIAS < 3%). O uso apenas de AM para prever Cla resultou em desempenho inferior ($R^2 = 0{,}22$, PBIAS = 17%). A mortandade de peixes apresentou correlação com NT ($R^2 = 0{,}37$), e um modelo híbrido mecanístico-AD previu mortandade com alta precisão ($R^2 = 0{,}90$, PBIAS = –8,47%). No geral, os resultados dos modelos revelaram que a qualidade da água melhora para maiores volumes de armazenamento, menores vazões de entrada e maiores vazões de saída. Apesar das limitações de dados, o estudo apresenta modelos robustos e generalizáveis para gestão da qualidade da água e avaliação de riscos ecológicos em reservatórios.

Palavras-chave: Modelagem eco-hidrológica; Eutrofização; Reservatórios semiáridos; Modelagem da qualidade da água.

INTRODUCTION

Reservoir eutrophication is a critical issue worldwide (Paerl & Otten, 2013; Amorin & Moura, 2021; Zhao et al., 2021). In tropical semiarid regions, reservoirs are subject to extreme hydrological variability, in addition to significant nutrient emissions from external and internal sources, which pose major challenges for eutrophication control and remediation (Lacerda et al., 2018; Wiegand et al., 2020; Cortez et al., 2022; Rocha & Lima Neto, 2021, 2022; Carneiro et al., 2023; Oliveira & Lima Neto, 2025). The alternation between high-flow periods and prolonged/severe droughts directly affects nutrient transport, retention, and transformation processes, with important consequences for primary productivity and eutrophication. Understanding these dynamics is particularly critical in regions where reservoirs serve multiple purposes, such as water supply, irrigation, aquaculture, and recreation, and where climate variability and its effects on water quality are projected to intensify under climate change scenarios (Raulino et al., 2021; Lima Neto, 2025a). Nutrient dynamics, especially those involving total phosphorus (TP) and total nitrogen (TN), are central to the development of eutrophication in reservoirs (Robson, 2014; Tasnim et al., 2021; Guimarães & Lima Neto, 2023; Lima Neto, 2023). However, predicting nutrient concentrations and their potential effects, such as algal blooms and fish mortality, remains a complex task due to nonlinear interactions between hydrology, biogeochemistry, and anthropogenic pressures, including livestock grazing, agricultural activities, wastewater discharge and aquaculture (Molisani et al., 2015; Aura & Ntiba, 2024; Rocha et al., 2024a, 2024b). Mechanistic water quality models, though useful for process understanding, often require detailed input data and may underperform under high hydrological variability (Mooij et al., 2010; Chapra et al., 2016; Lira et al., 2020; Li et al., 2021; Lima Neto et al., 2022; Silva et al., 2025; Cavalcante & Lima Neto, 2026). Conversely, data-driven approaches such as machine learning (ML) or empirical models offer flexible alternatives, but their performance and interpretability in these contexts are still under investigation (Hanson et al., 2020; Carvalho et al., 2022; Goes et al., 2023; Deng et al., 2024; Park et al., 2024; Santos et al., 2024). Recent studies have emphasized the value of hybrid approaches for predicting eutrophication in water bodies. Hybrid approaches are defined here as combined models that integrate the explanatory power of process-based models with the adaptability of ML techniques and/or empirical approaches for predicting eutrophication in water bodies. Wang et al. (2020) integrated hydrological modeling and ML approaches to predict nutrient concentrations in watersheds. Deyle et al. (2022) combined 1D hydrodynamic modeling with empirical data-driven methods to predict lake eutrophication. Chen et al. (2024) combined 3D hydrodynamic modeling with ML techniques for forecasting chlorophyll-a (Chla) in lakes. However, to the author's knowledge, all previous studies on this subject have focused on water bodies characterized by relatively low variability in water levels, as compared to tropical semiarid ones. Moreover, the combination of steady-state complete-mixing TP models with classical TP-Chla empirical relationships have been restricted to temperate lakes (Chapra, 2008). However, in tropical semiarid reservoirs, not only hydrological variability is important but also the classical TP-Chla relationships do not apply (Carvalho et al., 2022). In addition, large tropical semiarid reservoirs are strongly impacted by aquaculture in fish cages, differently from reservoirs located in temperate regions (Lacerda et al., 2018; Rocha et al., 2024a, 2024b). In this sense, the combination of transient complete-mixing nutrient models (Rocha & Lima Neto, 2021, 2023) with appropriate empirical relationships for Chla (Guimarães & Lima Neto, 2023) may be particularly well-suited to tropical semiarid reservoirs, where long-term datasets are rare and hydrological conditions are highly variable. The comparison of such models with ML is also of scientific relevance. Moreover, although hydrologic and nutrient modeling in reservoirs has been widely used to predict eutrophication, phytoplankton biomass, and dissolved oxygen dynamics (Lima Neto, 2023, 2025b), which are recognized as the primary drivers of fish mortality, the explicit prediction of fish mortality directly from hydrologic and/or nutrient modeling has not yet been reported in the literature.

This study analyzes 15 years of monthly data from a large multipurpose reservoir in northeastern Brazil, covering contrasting hydrological phases, from pre- to post-drought periods. The performance of a transient complete-mixing mechanistic model and four ML algorithms in reproducing TP and TN concentrations is investigated. The use of hybrid models considering the mechanistic or ML approaches combined with empirical relationships for predicting Chla and fish mortality is also evaluated. The study aims to assess the applicability of such modeling approaches in data-limited, high-variability tropical semiarid contexts, and to identify their potential contributions to improved reservoir operation with respect to water quality.

METHODOLOGY

Study site and dataset

The study was conducted in Castanhão, a large multipurpose reservoir located in the tropical semiarid region of northeastern Brazil, as shown schematically in Figure 1. The reservoir has a capacity exceeding 6 billion m^3 and plays a critical role in regional water supply, irrigation, fish-cage aquaculture and recreation. Castanhão has a maximum depth of approximately 60 m near the dam and an average depth of approximately 15-20 m, which strongly influence stratification and nutrient cycling processes. Water transparency, represented by Secchi depth, typically ranges from 0.5 to 2.0 m, corresponding to a photic zone depth of approximately 1 to 6 m, where most phytoplankton growth occurs. Its monitoring and operation, along with that of 156 other strategic reservoirs across the state of Ceará, are managed by the Water Resources Management Company of Ceará (COGERH) (Wiegand et al., 2021; Rocha et al., 2024a, 2024b). In addition to these, the region contains over 25,000 small private reservoirs, which significantly influence not only the hydrological regime but also downstream water quality in major systems like Castanhão (Rabelo et al., 2021; Freire et al., 2023).

The catchment of Castanhão reservoir is characterized by a highly seasonal rainfall regime, with most precipitation occurring between January and May. Over the 15-year study period (2008-2022), the reservoir experienced significant hydrological variation, including a high-flow phase (2008-2011), a prolonged drought (2012-2016), and an intermediate flow regime (2017-2022),

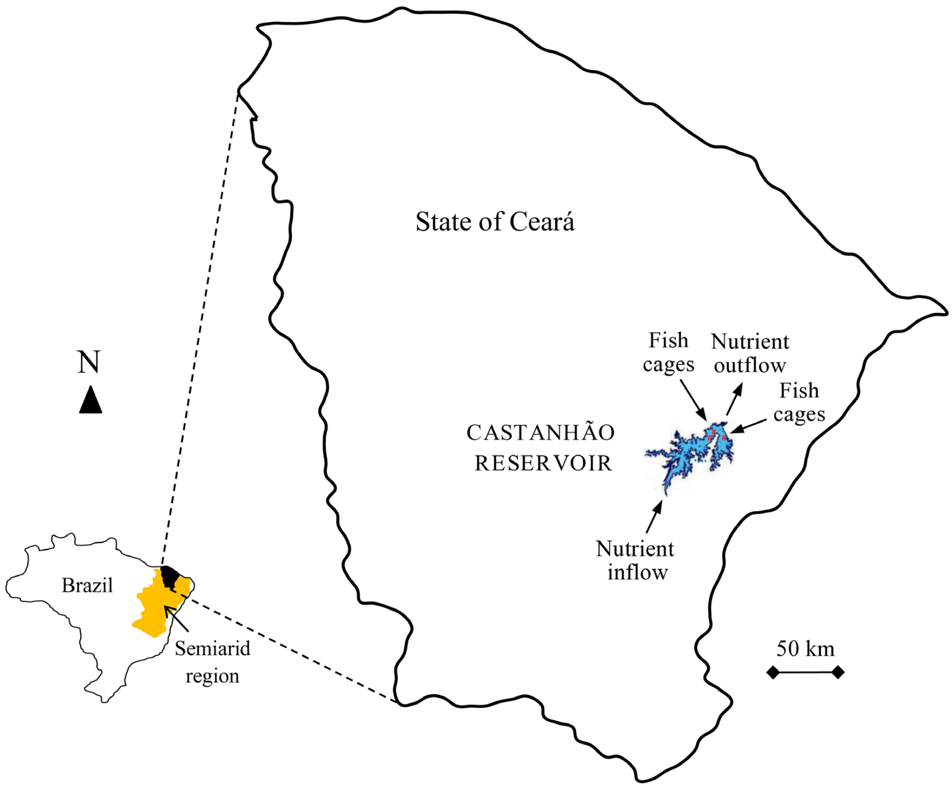

Figure 1. Location of the Castanhão reservoir in the Brazilian semiarid region, indicating external and internal nutrient loads.

as indicated in Figure 2. Monthly data were obtained from COGERH. Variables used in the analysis included inflow, outflow, water volume, fish-cage production, and concentrations of total phosphorus (TP), total nitrogen (TN), and chlorophyll-a (Chla). Figure 3 shows the fish-cage loads of TP and TN obtained from COGERH, which corresponded to 1.25 and 7.50% of total fish production, respectively.

All data provided by COGERH were used without outlier removal or artificial filtering, in order to preserve the integrity and representativeness of the observed environmental variability. The laboratory data undergo routine quality assurance and quality control procedures conducted by COGERH, following standardized sampling protocols, calibrated analytical methods, and internal consistency checks to ensure data reliability, in accordance with American Public Health Association (2017) Standard Methods.

Mechanistic models

A transient complete-mixing mechanistic model was implemented to simulate monthly TP and TN concentrations in the reservoir. The model structure followed mass balance principles (Chapra, 2008), accounting for nutrient loads from river inflow and fish cages, nutrient decay due to sedimentation and uptake minus sediment release during hypoxic conditions, and outflow losses through withdrawal and spillage. The assumption that hypoxic conditions and sediment nutrient release occur year-round is supported by Carneiro et al. (2023), who observed that bottom waters in the Castanhão reservoir remained hypoxic (< 1.5 mg/L) for more than 90% of the study period (five years). Nutrient decay was considered a first-order reaction, with TP and TN decay rates fitted to the observed data using two metrics: coefficient of determination (R^2) and percent bias (PBIAS). Equation 1, adapted from Rocha & Lima Neto (2021, 2023), was solved numerically on a monthly time step for the entire 15-year period by using a simple explicit finite difference scheme.

$$d\left(V.C\right)/dt = Q_{in}.C_{in} + W_f - k.V.C - Q_{out}.C \tag{1}$$

where V is reservoir volume (m^3), C is nutrient (TP or TN) concentration (mg/m^3), Q_{in} and Q_{out} are inflow and outflow (m^3/month), respectively, C_{in} is inflow nutrient concentration (mg/m^3), W_f is fish farming load (mg/month), and k is a global nutrient decay rate ($month^{-1}$).

The inflow concentration of TP or TN was determined for the entire time series by using the following relationship fitted by Rocha (2024) to the field and laboratory data provided by COGERH from 2020 to 2022:

$$C_{in}\left(mg/m^3\right) = f\{25.5\,[Q_{in}\left(m^3/s\right)]^{-0.92} + 59.6\,[Q_{in}\left(m^3/s\right)]^{0.32}\} \tag{2}$$

in which $f = 1$ for TP and 6 for TN. Note that Equation 2 is similar ($R^2 = 0.95$) to the relationship derived by Rocha & Lima Neto (2021) from nutrient mass balance in the Castanhão reservoir from 2008 to 2020. Therefore, it is considered suitable for the purpose of the present study.

Due to the limited number of observations (58 for TP and 47 for TN), 80% of the data were used for calibration and

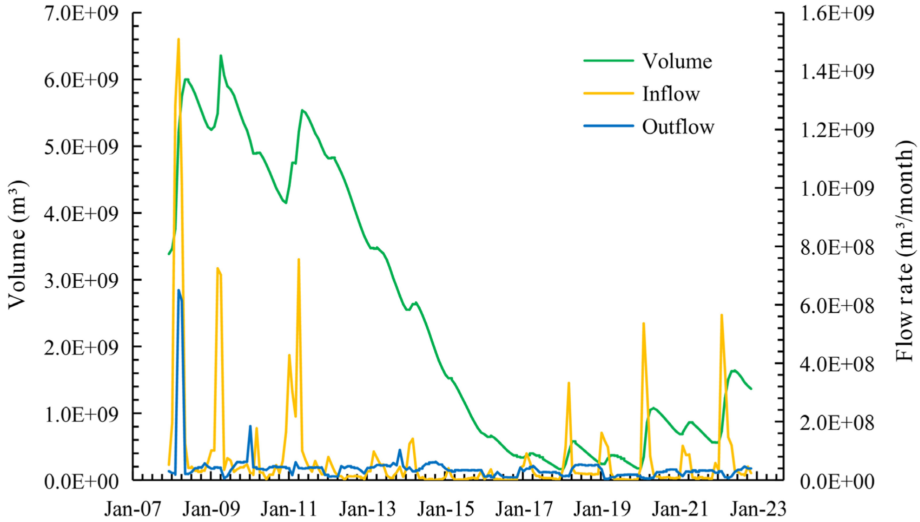

Figure 2. Interannual and seasonal variability of volume, inflow and outflow of Castanhão reservoir, including a high-flow phase (2008-2011), a prolonged drought (2012-2016), and an intermediate flow regime (2017-2022).

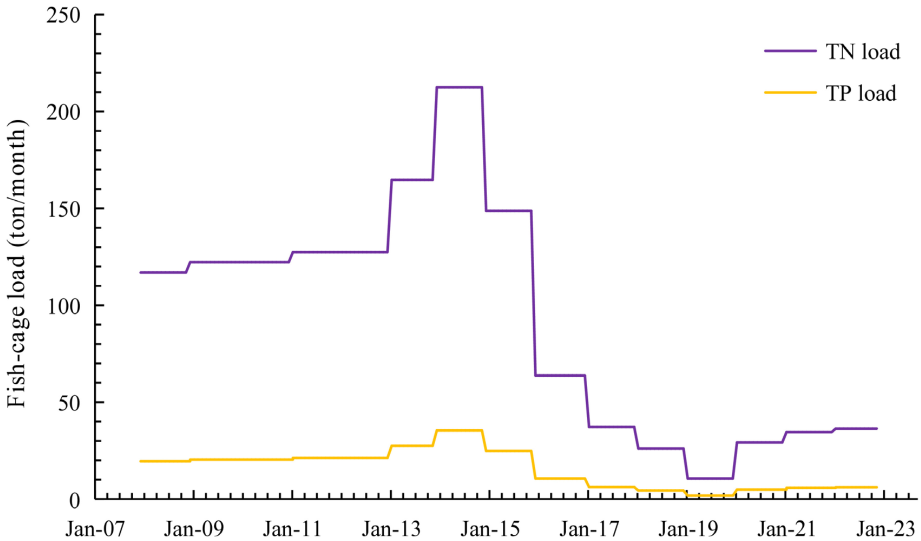

Figure 3. Fish-cage loads of TP and TN, corresponding to 1.25 and 7.50% of total fish production, respectively.

20% for validation. The first period (2008-2019) encompasses a distinct flood-to-drought transition (see Figure 2), while the second (2020-2022) corresponds to a flow recovery (intermediate) phase. The computational framework was implemented in Excel.

Machine learning models

Four machine learning (ML) algorithms were tested for simulating TP and TN concentrations: linear regression (LR), Ridge regression (RR), decision tree (DT) and k-nearest neighbors (KNN). These algorithms were selected instead of more complex ensemble methods primarily due to the limited size of the available dataset. In scenarios with small sample sizes, high-dimensional ensemble models are highly prone to overfitting, often capturing noise as if it were a signal, which compromises their generalizability. Simpler, parsimonious models are more robust under these constraints, providing more stable and interpretable results. Furthermore, these specific algorithms were chosen to maintain methodological consistency with previous studies conducted in the same semi-arid region (Carvalho et al., 2022), allowing for a direct comparison of water quality forecasting performance and ensuring that the models remain locally relevant for regional water management.

LR estimates a linear relationship between input and target variables:

$$\hat{y}_i = \beta_0 + \Sigma \beta_j \cdot x_i \tag{3}$$

where: $\hat{y}_i$ is the target variable, β_0 is the intercept, β_j are the regression coefficients, and x_i are the input variables.

RR is a variant of LR that adds a penalty term to prevent overfitting by shrinking coefficients:

$$SSR = \Sigma \left(y_i - \hat{y}_i \right)^2 + \alpha \cdot \Sigma \beta_j{}^2 \tag{4}$$

where: SSR is the sum of the squares of the residuals, $\hat{y}_i$ and y_i are the predicted and observed values, respectively, β_j are the regression coefficients, and α is the regularization parameter.

DT splits the data into branches based on feature thresholds, modeling nonlinear relationships by learning decision rules:

$$\hat{y}_i = \Sigma c_j \cdot \phi\{x \in R_j\} \tag{5}$$

where: R_j are the regions of the input space, defined by decision rules, c_j is the predicted value for the data points that fall into region R_j, $\phi\{x \in R_j\}$ is an indicator function that equals 1 if $x \in R_j$, and 0 otherwise.

KNN predicts values based on the average target values of the closest data points in the feature space:

$$\hat{y}_i = \left(1/k\right) \cdot \Sigma\left(y_i\right) \tag{6}$$

where: $i \in N_k(x)$ is the set of k nearest neighbors, and N_k is the neighborhood of x defined by the k closest points in the training sample.

The ML models were calibrated using inflow, outflow, volume and fish-cage production as input variables, and TP, TN or Chla as target variables. Due to the limited number of observations (58 for TP, 47 for TN and 30 for Chla), the same 80/20 data split for calibration and validation used for the mechanistic models was assumed here for the ML approaches, which were implemented in Python language, version 3.13.3, using primarily the Scikit-learn library. To stabilize variance and mitigate negative R^2 values driven by hydrological outliers, target variables underwent log-transformation. Predictors were standardized to zero mean and unit variance, with a global random seed (42) ensuring exact reproducibility. To prevent overfitting and enhance generalizability, hyperparameters were strictly constrained: Ridge Regression was set with $\alpha = 40$, the Decision Tree was limited to a maximum depth of 3 and a minimum of 12 samples per leaf, and the K-Nearest Neighbors (KNN) model utilized k = 15 with uniform weighting. Data partitioning was performed using the TrainTestSplit module, while the StandardScaler module was utilized for preprocessing to normalize independent variables. The predictive architecture integrated a suite of regressors, including Linear Regression (LR) for baseline analysis, Ridge Regression (RR) with L2 regularization to manage coefficient stability, K-Nearest Neighbors (KNN) for similarity-based estimation, and Decision Tree (DT) Regressor to capture non-linear relationships. These were further supported by Pandas and NumPy for data manipulation and Joblib for model persistence. A detailed description of the above-mentioned ML approaches can also be found in Carvalho et al. (2022).

Hybrid models

The combination of the mechanistic (Equations 1 and 2) or best-performing ML models (Equations 3-6) with an empirical power-law relationship given by Guimarães & Lima Neto (2023) was also used for prediction of Chla as a function of TP, TN, volume and precipitation:

$$Chla = 66.38 \cdot TP^{0.05} \cdot TN^{0.34} \cdot V\left(\%\right)^{-0.58} \cdot \left(1 + P / P_{max}\right)^{0.97} \tag{7}$$

where: $V(\%)$ is the percent volume of the reservoir, P and P_{max} are the daily and maximum observed precipitation, respectively.

Equation 7 is expected to generalize across the full hydrological range because its predictors (TP, TN, V and P) represent the dominant physical and biogeochemical drivers of phytoplankton dynamics under both low- and high-flow conditions (Guimarães & Lima Neto, 2023). To ensure consistency with the monthly modeling framework, precipitation inputs were derived from daily observations. These daily metrics were then aggregated to monthly indicators, preserving the influence of extreme rainfall events while maintaining temporal compatibility with the monthly chlorophyll-a and water quality data used in model development.

Additionally, a hybrid model based on the above-mentioned mechanistic models for nutrients and ML approaches considering TP, TN, $V(\%)$ and P/P_{max} as input variables was also used for prediction of the occurrence and magnitude of fish kill events. Finally, future scenarios of climate change were projected for TP, TN, Chla and fish mortality.

RESULTS AND DISCUSSION

Nutrient modeling

The calibration and validation performances of the nutrient mechanistic and ML models with respect to R^2 and PBIAS is summarized in Table 1. The global TP and TN decay coefficients of the mechanistic model were fitted to 0.16 and 0.09 month^{-1}, which are of the same order of the value of 0.20 month^{-1} estimated from the general relationship given by Rocha & Lima Neto (2021, 2023) ($k = 4/RT^{0.5}$) for tropical semiarid reservoirs, considering the average water residence time (RT) of Castanhão reservoir in the study period (2008 - 2022): $RT = 2.8$ year. As depicted in Figure 4a, the mechanistic model and best-performing ML approaches (DT and KNN) for TP achieved satisfactory R^2 (> 0.30) and very good PBIAS (< ±15%), indicating acceptable predictive skill and low overall bias, according to Moriasi et al. (2015). These models outperformed the others across

Table 1. Calibration and validation performances of the nutrient mechanistic and ML models.

Model	TP (calibration / validation)		TN (calibration / validation)	
	R^2	**PBIAS (%)**	**R^2**	**PBIAS (%)**
Mechanistic	0.38 / 0.26	-11.00 / 11.03	0.31 / 0.24	-0.69 / 14.77
Linear Regression (LR)	0.36 / 0.19	-8.72 / 5.40	0.26 / -0.55	-5.88 / 9.01
Ridge Regression (RR)	0.36 / 0.21	-8.91 / 5.15	0.17 / -0.01	-9.98 / -13.47
Decision Tree (DT)	0.59 / 0.36	-6.10 / 2.77	0.32 / 0.25	-7.87 / -13.94
K-Nearest Neighbors (KNN)	0.47 / 0.59	-9.31 / -8.28	0.13 / -0.02	-9.99 / -17.42

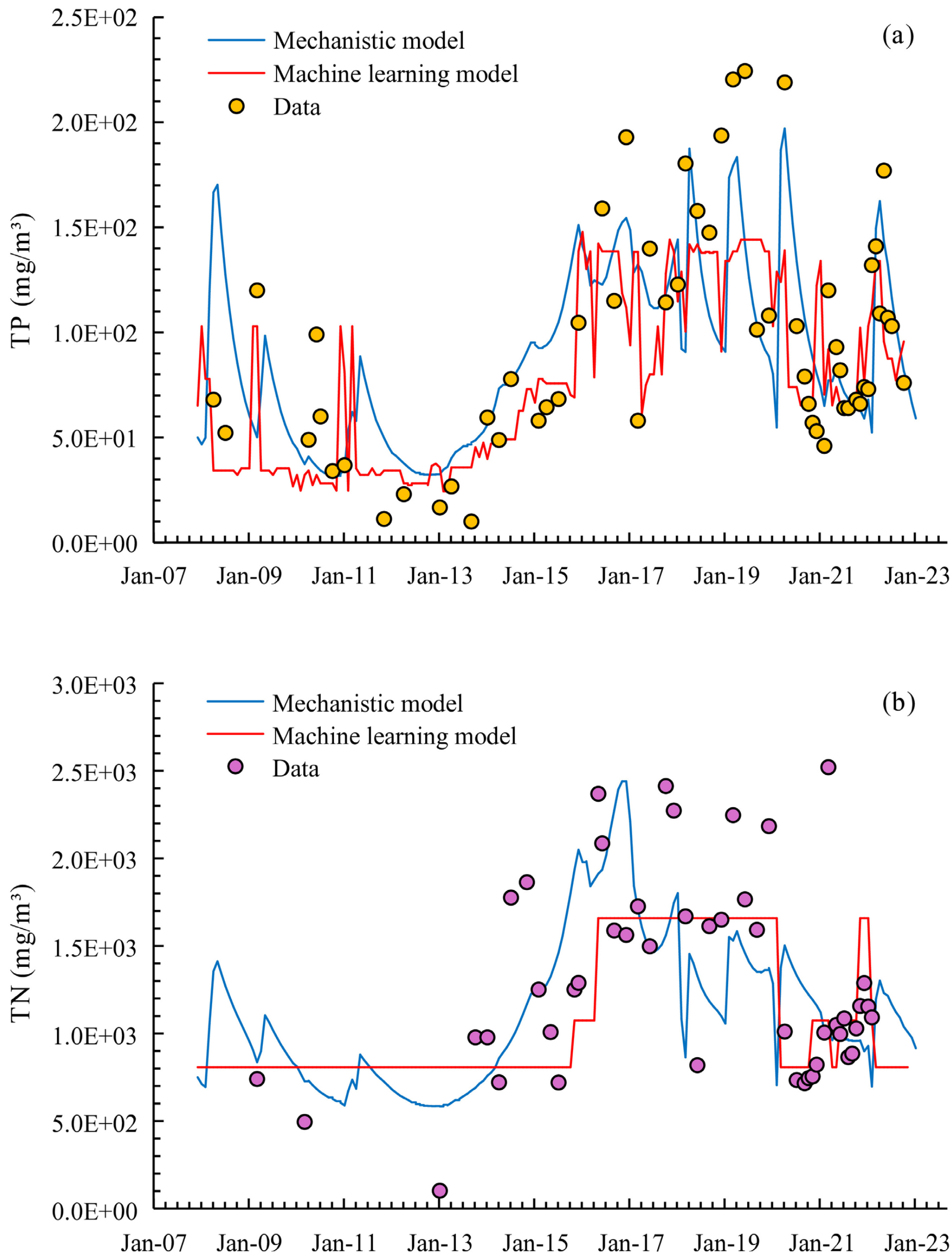

Figure 4. Comparison of nutrient model results with measured data along the study period: (a) Mechanistic and best-performing ML (KNN) models for TP, and (b) Mechanistic and best-performing ML (DT) models for TN.

most hydrological phases, especially during the drought period (2012-2016), suggesting their ability to capture complex and non-linear interactions among predictors. Nevertheless, the performance of linear ML models such as LR and RR remained lower ($R^2 < 0.30$ in the validation phase), reflecting their limited capacity to represent non-linear system behavior in such a variable environment. The lower performance of simpler models (LR and RR) is consistent with previous studies, which have shown that linear approaches have limited ability to represent the nonlinear dynamics governing nutrient behavior in tropical semiarid reservoirs (Carvalho et al., 2022). The above results demonstrate the value of both process-based and non-linear ML modeling for long-term TP trends, especially in reservoirs with highly variable hydrological regime.

TN proved more difficult to model. Most ML algorithms resulted in poor performance for R^2 (< -0.30), but satisfactory for PBIAS (< ±30%), according to Moriasi et al. (2015). On the other hand, the mechanistic and best-performing ML (DT) models for TN showed a higher R^2 (0.24 - 0.32) and very good PBIAS (< 15%), suggesting a more reliable long-term estimate compared to the other models, despite relatively weak predictive skill, as depicted in Figure 4(b). These results may reflect the greater complexity of nitrogen cycling, including biological assimilation, denitrification, and atmospheric exchange (Li et al., 2021; Rocha & Lima Neto, 2023), processes that are poorly captured by both simplified mechanistic and data-driven approaches, especially in reservoirs with highly variable hydrological regime.

Chlorophyll-a modeling

The calibration/verification results of Chla models with respect to R^2 and PBIAS are shown in Table 2. The combination of the mechanistic or best-performing ML models with the empirical power-law relationship given by Guimarães & Lima Neto (2023) resulted in very good agreement with Chla data ($R^2 > 0.70$ and PBIAS < 3%), indicating that both modeling approaches successfully captured the key hydrological and nutrient controls on phytoplankton dynamics, as depicted in Figure 5. This resulted in a significantly better performance than using ML (KNN) with the entire dataset ($R^2 = 0.22$ and PBIAS = 17.2%). It is also important to note that this performance was also higher than that reported by Carvalho et al. (2022) for Castanhão reservoir, in which the best-performing model (random forest) yielded $R^2 = 0.52$. This probably occurred because of the combination of the mechanistic or ML approaches with the empirical model, in addition to the inclusion of relevant input variables that were not considered in the previous study: inflow, outflow and fish-cage production. On the other hand, the model performance obtained in the present study is similar to that reported by Chen et al. (2024) and Park et al. (2024) by using more complex approaches ($0.70 < R^2 < 0.95$), such as the combination of physical-based hydrodynamic models and machine learning to predict Chla in lakes with much lower hydrological variability.

These findings highlight the potential of hybrid modeling frameworks that couple nutrient simulations with simple empirical Chla models. Such approaches can serve as practical tools for predicting eutrophication under data-scarce conditions, especially when continuous water quality monitoring is unavailable. The effectiveness of the empirical Chla model also underscores the strong TP-TN co-limitation regime in this semiarid reservoir, consistent with findings from previous studies in the region (Wiegand et al., 2020).

Table 2. Calibration/verification performance of the Chla models.

Model	Chla	
	R^2	PBIAS (%)
Mechanistic (TP and TN) + Empirical	0.76	-1.61
KNN (TP) and DT (TN) + Empirical	0.73	2.10
K-Nearest Neighbors (KNN)	0.22	17.20

Fish mortality modeling

Figure 6 shows a comparison of TN mechanistic model results with measured data of fish mortality along the study period. It is evident that TN follows a trend similar to that of fish mortality. Likewise, Figure 7 shows that the magnitude of fish mortality correlated satisfactorily with TN with $R^2 = 0.37$. Possible explanations for this based on previous studies of Molisani et al. (2015), Aura & Ntiba (2024) and Rocha et al. (2024a, 2024b) are: (1) TN contributes to eutrophication and, as a consequence, oxygen depletion; (2) TN may lead to harmful algal blooms and the release of toxins that impair fish health; and (3) TN fractions such as ammonia can be directly toxic to fish under certain environmental conditions. However, as the only reliable information available during the fish kill events is the presence of strong thermal and chemical stratification, with dissolved oxygen levels often approaching zero near the surface (Carneiro et al., 2023), fish mortality likely occurred due to asphyxia caused by hypoxic conditions.

The calibration/verification results of fish mortality models with respect to R^2 and PBIAS are shown in Table 3. The combination of the mechanistic models for TP and TN with the best-fitting ML algorithm (DT) resulted in very good agreement with fish mortality magnitude data ($R^2 = 0.90$ and PBIAS = -8.47%), according to Moriasi et al. (2015), indicating that this hybrid modeling approach captured adequately the impact of hydrology and nutrient variability on fish kills. The mean absolute error was also acceptable (MAE = 14.1 ton), which is relatively low compared to the maximum observed values (up to about 3,000 ton). Figure 8 also shows that model

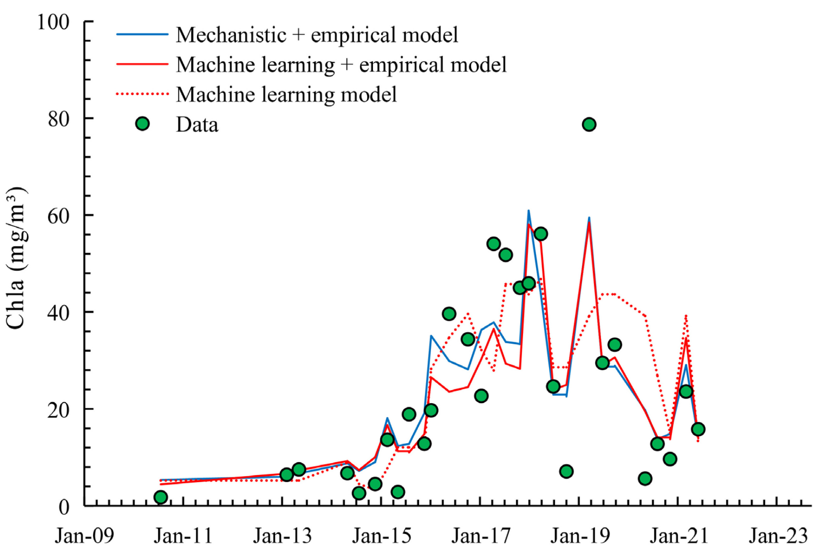

Figure 5. Comparison of Chla model results with measured data along the study period, including a hybrid mechanistic + empirical model, a hybrid ML (KNN + DT) + empirical model, and a ML (KNN) model.

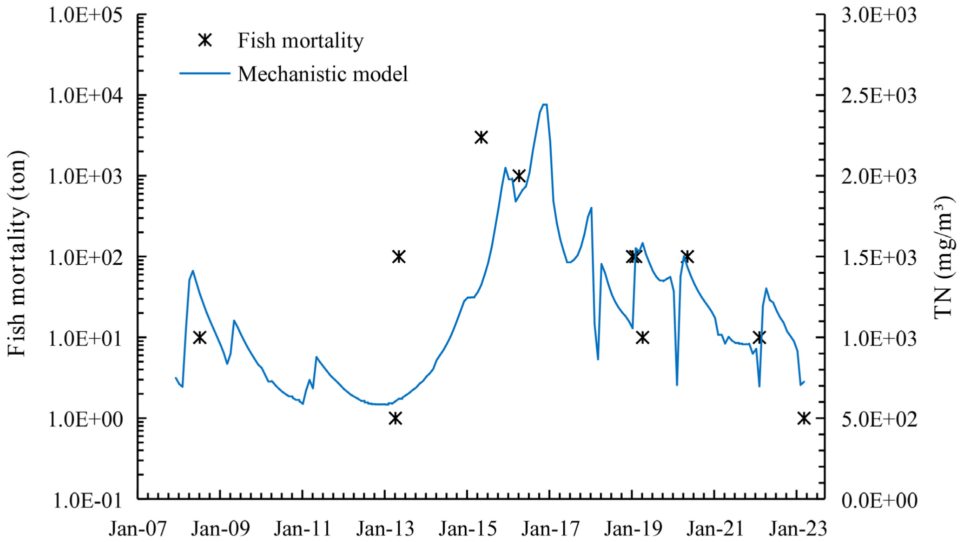

Figure 6. Comparison of TN mechanistic model results with measured data of fish mortality along the study period.

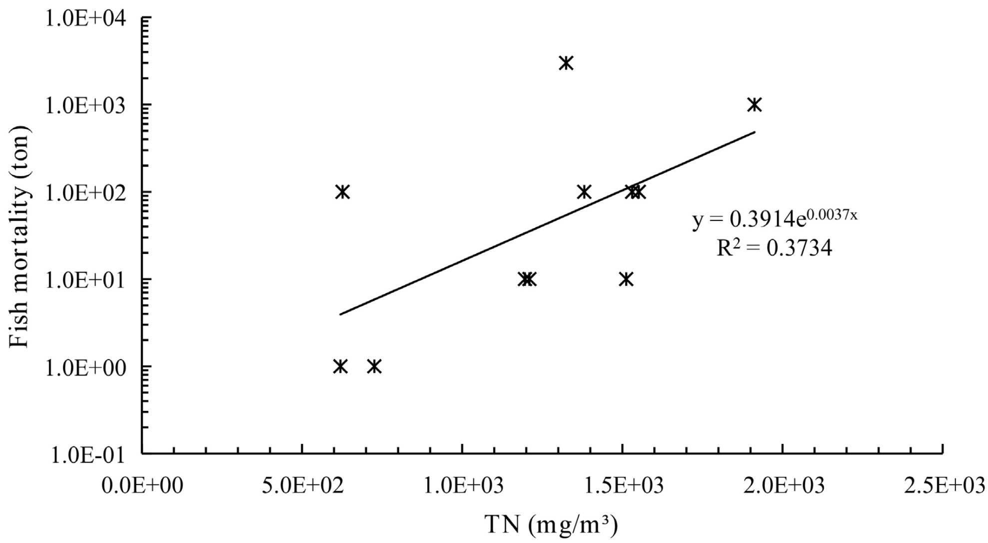

Figure 7. Correlation of the magnitude of fish mortality as a function of total nitrogen concentration.

Table 3. Calibration/verification performance of the fish mortality models.

Model	Fish Mortality	
	R^2	PBIAS (%)
Mechanistic (TP and TN) + ML (LR)	0.01	-20.29
Mechanistic (TP and TN) + ML (RR)	0.01	-20.37
Mechanistic (TP and TN) + ML (DT)	0.90	-8.47
Mechanistic (TP and TN) + ML (KNN)	0.33	-19.95

predictions responded to 82% of the fish-kill events, including all the relevant mortalities (> 100 ton).

Model applications

Extreme events, such as heatwaves and prolonged droughts, can substantially affect reservoir water quality by increasing water temperature, enhancing stratification, and reducing inflows and reservoir volumes. Observed data from reservoirs in Ceará show that heatwaves are associated with surface water temperature increases of 2-5 °C and chlorophyll-a concentrations rising by 20-50%, while drought periods have led to inflow reductions of 30-80%, storage volume decreases of up to ~2% of total capacity, and nutrient and chlorophyll-a increases of 20-60%, driven by longer residence times and reduced dilution capacity.

Empirical relationships based on the temperature coefficient using Arrhenius-type equations (Chapra, 2008; Toné & Lima Neto, 2020) indicate that an increase of 5 °C in water temperature (from 30 to 35 °C) can accelerate nutrient decay by 30%, increase chlorophyll-a concentrations by 40%, and reduce dissolved oxygen saturation by 10%, which may affect fish mortality.

The modeling approaches used in this study, particularly the ones including mechanistic components, can capture these

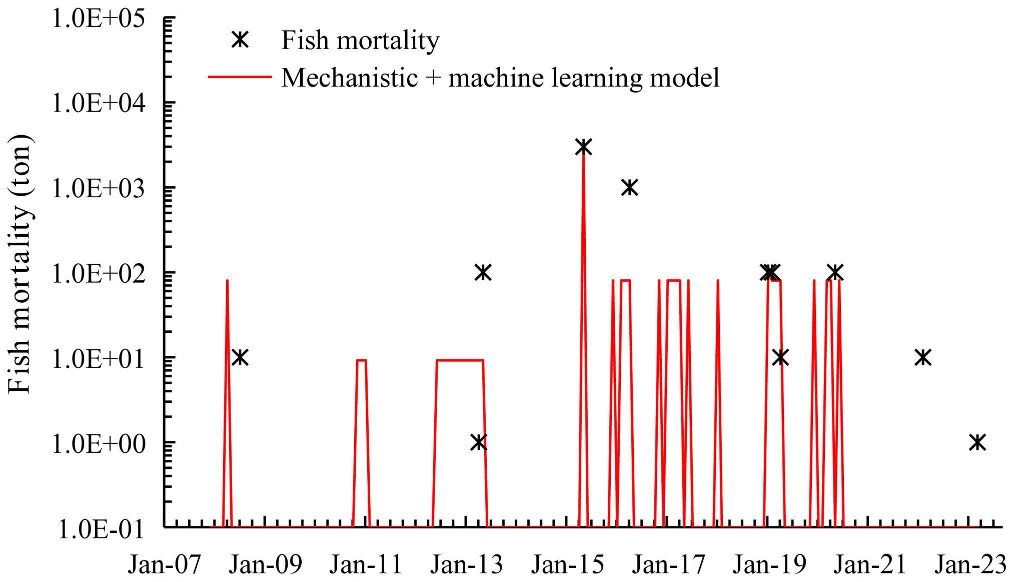

Figure 8. Prediction of fish mortality occurence and magnitude with a hybrid mechanistic + ML (DT) model.

impacts, as well as the effects of extreme floods and droughts, climate change, and variations in fish production, by adjusting key input variables such as inflows/outflows, storage volume, fish biomass, nutrient decay and algal growth coefficients. Consequently, the models enable explicit simulation of extreme scenarios, supporting risk assessment and adaptive management strategies under increasing hydroclimatic and ecological variability.

Model simulations for different reservoir operational conditions revealed that water quality consistently improves under higher storage volumes, lower inflow rates, and higher outflow rates. Larger storage volumes enhance dilution capacity, reducing nutrient concentrations and improving overall water quality. Lower inflow rates decrease external nutrient loading from upstream sources, thereby limiting eutrophication processes. At the same time, higher outflow rates promote more effective flushing of nutrients and phytoplankton biomass, reducing water residence time and preventing the accumulation of pollutants within the reservoir. These combined hydrological conditions contribute to more favorable water quality by enhancing dilution, reducing nutrient inputs, and increasing flushing efficiency.

It is also important to mention that the present study was conducted in a reservoir characterized by high hydroclimatic variability, including pronounced fluctuations in inflow, storage volume, and nutrient concentrations driven by intermittent rainfall patterns typical of semi-arid tropical regions. Despite these challenging conditions, the models demonstrated satisfactory predictive performance and robustness during both calibration and validation periods. This indicates that the modeling framework is capable of capturing the dominant physical and biogeochemical processes governing water quality dynamics under highly variable environmental conditions. Therefore, the demonstrated robustness under such highly dynamic conditions suggests that the proposed modeling approaches are also applicable, in principle, to reservoirs with lower hydroclimatic variability, where system behavior is typically more stable and predictable.

Model uncertainties and limitations

All environmental models are inherently subject to uncertainties associated with input data, parameter estimation, measurement errors, and structural limitations. In the present study, these uncertainties were evaluated considering measurement limitations, monitoring frequency, analytical precision, and structural simplifications. Typical uncertainty levels were assumed to be approximately 10% for precipitation, 20% for discharge, 15% for reservoir volume, and 20% for nutrient concentrations. Biological variables, such as chlorophyll-a and fish mortality, present higher uncertainty levels, typically around 30%, due to the greater variability and complexity of ecological processes. Considering the propagation of these independent uncertainty sources using the root-sum-of-squares (RSS) method, the overall prediction uncertainty was estimated to be approximately 40%, which is consistent with previous hydrological and water quality modeling studies (Chapra, 2008; Harmel et al., 2006; McMillan et al., 2012; Moriasi et al., 2015).

To further strengthen the evaluation of model robustness and predictive capability, an independent validation strategy was adopted, using 80% of the data for calibration and 20% for validation. This approach provides a more rigorous assessment of model generalization and reduces the risk of overfitting. The results demonstrated consistent performance between calibration and validation, particularly for the hybrid models, which exhibited greater stability due to the incorporation of mechanistic constraints. Despite the inherent uncertainties, the models showed satisfactory predictive performance, indicating that the results are robust and suitable for long-term trend analysis and for supporting water quality management decisions.

It is important to stress that both mechanistic and machine learning models are subject to inherent limitations related to structural simplifications and dependence on the quality, quantity, and temporal resolution of available data. The mechanistic model adopted in this study represents key nutrient cycling processes but does not explicitly resolve more complex mechanisms such as detailed sediment-water interactions, internal loading dynamics under varying redox conditions, and spatial heterogeneity associated with hypoxia. More sophisticated total phosphorus (TP) models that incorporate sediment exchange processes and different hypoxia regimes have demonstrated satisfactory to very good agreement with observations in both the water column and sediments (Lima Neto, 2025b). Similarly, total nitrogen (TN) and machine learning models may achieve improved predictive performance when more detailed process representations and larger, higher-frequency datasets are available. Therefore, model performance is inherently constrained by both process representation and data availability, and future work should focus on incorporating additional mechanistic complexity and expanding monitoring programs to further improve predictive accuracy and process representation.

CONCLUSION

This study demonstrates the value of integrating different modeling approaches to better understand and predict eutrophication dynamics in tropical semiarid reservoirs. Despite the challenges posed by highly variable hydrological regimes and data scarcity, both mechanistic and machine learning modeling strategies produced satisfactory predictions for total phosphorus and total nitrogen. The hybrid use of mechanistic or machine learning models combined with empirical relationships significantly improved chlorophyll-a predictions, highlighting the importance of combining complementary modeling techniques. Furthermore, fish mortality was effectively predicted through a hybrid model combining mechanistic and machine learning approaches that captured both its occurrence and magnitude. Beyond the specific case of the Castanhão reservoir, the proposed hybrid framework represents a transferable and scalable workflow, not only for reservoirs with high hydrological variability. Nevertheless, some limitations should be acknowledged, including the relatively small number of observations for key variables, the limited validation scope for the machine learning and hybrid components due to data constraints. Despite these limitations, the results provide a robust and generalizable modeling framework with strong potential to support water-quality management and ecological risk assessment in data-limited reservoirs worldwide.

DATA AVAILABILITY STATEMENT

Research data is only available upon request.

ACKNOWLEDGEMENTS

The author thanks the Water Resources Management Company of Ceará (COGERH) for providing the data necessary for the analyses and acknowledges the financial support from the National Council for Scientific and Technological Development - CNPq (#307680/2023-1) and the Coordination for the Improvement of Higher Education Personnel - CAPES (#2160/2024).

REFERENCES

American Public Health Association – APHA. (2017). *Standard methods for the examination of water and wastewater* (23rd ed). Washington, D.C.: APHA.

Amorin, A. M., & Moura, A. N. (2021). Ecological impacts of freshwater algal blooms on water quality, plankton biodiversity, structure, and ecosystem functioning. *The Science of the Total Environment*, *758*, 143605. https://doi.org/10.1016/j.scitotenv.2020.143605.

Aura, C. M., & Ntiba, M. J. (2024). Possible lessons on rapid assessment of fish kills in cages in Lake Victoria, Kenya for informed decision making. *Aquatic Ecosystem Health & Management*, *27*(3), 65-73. https://doi.org/10.14321/aehm.027.03.65.

Carneiro, B. L. D. S., Rocha, M. J. D., Barros, M. U. G., Paulino, W. D., & Lima Neto, I. E. (2023). Predicting anoxia in the wet and dry periods of tropical semiarid reservoirs. *Journal of Environmental Management*, *326*, 116720. https://doi.org/10.1016/j.jenvman.2022.116720.

Carvalho, T. M. N., Lima Neto, I. E., & Souza Filho, F. A. (2022). Uncovering the influence of hydrological and climate variables in chlorophyll-a concentration in tropical reservoirs with machine learning. *Environmental Science and Pollution Research International*, *29*(49), 74967-74982. https://doi.org/10.1007/s11356-022-21168-z.

Cavalcante, J. S., & Lima Neto, I. E. (2026). Hydrological-hydraulic and water quality modeling of the Maceió Stream in Fortaleza/CE: a case study. *Revista Brasileira de Recursos Hídricos*, *31*, e15. https://doi.org/10.1590/2318-0331.312620250138.

Chapra, S. C., Dolan, D. M., & Dove, A. (2016). Mass-balance modeling framework for simulating and managing long-term water quality for the lower Great Lakes. *Journal of Great Lakes Research*, *42*(6), 1166-1173. https://doi.org/10.1016/j.jglr.2016.04.008.

Chapra, S. C. (2008). *Surface water-quality modeling* (835 p.). Long Grove: Waveland Press.

Chen, C., Chen, Q., Yao, S., He, M., Zhang, J., Li, G., & Lin, Y. (2024). Combining physical-based model and machine learning to forecast chlorophyll-a concentration in freshwater lakes. *The Science of the Total Environment*, *907*, 168097. https://doi.org/10.1016/j.scitotenv.2023.168097.

Cortez, F., Monicelli, F., Cavalcante, H., & Becker, V. (2022). Effects of prolonged drought on water quality after drying of a semiarid tropical reservoir, Brazil. *Limnologica*, *93*, 125959. https://doi.org/10.1016/j.limno.2022.125959.

Deng, Y., Zhang, Y., Pan, D., Yang, S. X., & Gharabaghi, B. (2024). Review of recent advances in remote sensing and machine learning methods for lake water quality management. *Remote Sensing (Basel)*, *16*(22), 4196. https://doi.org/10.3390/rs16224196.

Deyle, E. R., Bouffard, D., Frossard, V., Schwefel, R., Melack, J., & Sugihara, G. (2022). A hybrid empirical and parametric approach for managing ecosystem complexity: water quality in Lake Geneva under nonstationary futures. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(26), e2102466119. https://doi.org/10.1073/pnas.2102466119.

Freire, L. L., Costa, A. C., & Lima Neto, I. E. (2023). Effects of rainfall and land use on nutrient responses in rivers in the Brazilian semiarid region. *Environmental Monitoring and Assessment*, *195*(6), 652. https://doi.org/10.1007/s10661-023-11281-y.

Goes, M. C. A., Barros, M. U. G., & Lima Neto, I. E. (2023). Prediction of total phosphorus in reservoir cascade systems. *Environmental Monitoring and Assessment*, *195*(12), 1550. https://doi.org/10.1007/s10661-023-12155-z.

Guimarães, B. M. D. M., & Lima Neto, I. E. (2023). Chlorophyll-a prediction in tropical reservoirs as a function of hydroclimatic variability and water quality. *Environmental Science and Pollution Research International*, *30*(39), 91028-91045. https://doi.org/10.1007/s11356-023-28826-w.

Hanson, P. C., Stillman, A. B., Jia, X., Karpatne, A., Dugan, H. A., Carey, C. C., Stachelek, J., Ward, N. K., Zhang, Y., Read, J. S., & Kumar, V. (2020). Predicting lake surface water phosphorus dynamics using process-guided machine learning. *Ecological Modelling*, *430*, 109136. https://doi.org/10.1016/j.ecolmodel.2020.109136.

Harmel, R. D., Cooper, R. J., Slade, R. M., Haney, R. L., & Arnold, J. G. (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Transactions of the ASABE*, *49*(3), 689-701. https://doi.org/10.13031/2013.20488.

Lacerda, L. D., Santos, J. A., Marins, R. V., & Silva, F. A. T. F. (2018). Limnology of the largest multi-use artificial reservoir in NE Brazil: The Castanhão Reservoir, Ceará State. *Anais da Academia Brasileira de Ciências*, *90*(2 Suppl 1), 2073-2096. https://doi.org/10.1590/0001-3765201820180085.

Li, Q., Yu, Q., Wang, F., Yan, W., & Wang, J. (2021). Nitrogen removal in the Chaohu Lake, China: implication in estimating lake N uptake velocity and modelling N removal efficiency of large lakes and reservoirs in the Changjiang River network. *Ecological Indicators*, *124*, 107353. https://doi.org/10.1016/j.ecolind.2021.107353.

Lima Neto, I. E. (2023). Modeling water quality in a tropical reservoir using CE-QUAL-W2: handling data scarcity, urban pollution and hydroclimatic seasonality. *Revista Brasileira de Recursos Hídricos*, *28*, e8. https://doi.org/10.1590/2318-0331.282320230003.

Lima Neto, I. E. (2025a). Modelling phosphorus inputs and dynamics in a large tropical semiarid basin. *Environmental Science and Pollution Research International*, *32*(11), 7070-7084. https://doi.org/10.1007/s11356-025-36181-1.

Lima Neto, I. E. (2025b). Phosphorus dynamics in water and sediments in a large multi-use reservoir under extreme volume variation. *Ecological Modelling*, *510*, 111316. https://doi.org/10.1016/j.ecolmodel.2025.111316.

Lima Neto, I. E., Medeiros, P. H. A., Costa, A. C., Wiegand, M. C., Barros, A. R. M., & Barros, M. U. G. (2022). Assessment of phosphorus loading dynamics in a tropical reservoir with high seasonal water level changes. *The Science of the Total Environment*, *815*, 152875. https://doi.org/10.1016/j.scitotenv.2021.152875.

Lira, C. C. S., Medeiros, P. H. A., & Lima Neto, I. E. (2020). Modelling the impact of sediment management on the trophic state of a tropical reservoir with high water storage variations. *Anais da Academia Brasileira de Ciências*, *92*(1), e20181169. https://doi.org/10.1590/0001-3765202020181169.

McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes*, *26*(26), 4078-4111. https://doi.org/10.1002/hyp.9384.

Molisani, M. M., Monte, T. M., Vasconcellos, G. H., Barroso, H. S., Moreira, M. O. P., Becker, H., Rezende, C. E., Franco, M. A. L., Farias, E. G. G., & Camargo, P. B. (2015). Relative effects of nutrient emission from intensive cage aquaculture on the semiarid reservoir water quality. *Environmental Monitoring and Assessment*, *187*(11), 707. https://doi.org/10.1007/s10661-015-4925-4.

Mooij, W. M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D. B. R., Degermendzhy, A. G., DeAngelis, D. L., De Senerpont Domis, L. N., Downing, A. S., Elliott, J. A., Fragoso Junior, C. R., Gaedke, U., Genova, S. N., Gulati, R. D., Hakanson, L., Hamilton, D. P., & Hipsey, M. (2010). Challenges and opportunities for integrating lake ecosystem modelling approaches. *Aquatic Ecology*, *44*(3), 633-667. https://doi.org/10.1007/s10452-010-9339-3.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: performance measures and evaluation criteria. *Transactions of the ASABE*, *58*(6), 1763-1785. https://doi.org/10.13031/trans.58.10715.

Oliveira, L. C. S., & Lima Neto, I. E. (2025). The impacts of the water intake operation on the hydraulic transients, sediment resuspension and water quality of the largest multi-use reservoir in Latin America. *Revista Brasileira de Recursos Hídricos*, *30*, e6. https://doi.org/10.1590/2318-0331.302520240093.

Paerl, H. W., & Otten, T. G. (2013). Harmful cyanobacterial blooms: causes, consequences, and controls. *Microbial Ecology*, *65*(4), 995-1010. https://doi.org/10.1007/s00248-012-0159-y.

Park, J., Patel, K., & Lee, W. H. (2024). Recent advances in algal bloom detection and prediction technology using machine learning. *The Science of the Total Environment*, *938*, 173546. https://doi.org/10.1016/j.scitotenv.2024.173546.

Rabelo, U. P., Dietrich, J., Costa, A. C., Simshäuser, M. N., Scholz, F. E., Nguyen, V. T., & Lima Neto, I. E. (2021). Representing a dense network of ponds and reservoirs in a semi-distributed dryland catchment model. *Journal of Hydrology*, *603*, 127103. https://doi.org/10.1016/j.jhydrol.2021.127103.

Raulino, J. B., Silveira, C. S., & Lima Neto, I. E. (2021). Assessment of climate change impacts on hydrology and water quality of large semiarid reservoirs in Brazil. *Hydrological Sciences Journal*, *66*(8), 1321-1336. https://doi.org/10.1080/02626667.2021.1933491.

Robson, B. J. (2014). State of the art in modelling of phosphorus in aquatic systems: Review, criticisms and commentary. *Environmental Modelling & Software*, *61*, 339-359. https://doi.org/10.1016/j.envsoft.2014.01.012.

Rocha, M. A. M. (2024). *Water quality dynamics in a multi-purpose reservoir in the Brazilian semiarid: responses to hydro-climatic variability and external and internal nutrient loadings* (Tese de doutorado). Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza. Retrieved in 2025, October 22, from https://repositorio.ufc.br/handle/riufc/76966

Rocha, M. A. M., Barros, M. U. G., Costa, A. C., Souza Filho, F. A., & Lima Neto, I. E. (2024a). Understanding the water quality dynamics in a large tropical reservoir under hydrological drought conditions. *Water, Air, and Soil Pollution*, *235*(1), 76. https://doi.org/10.1007/s11270-024-06890-3.

Rocha, M. A. M., Barros, M. U. G., Souza Filho, F. A., & Lima Neto, I. E. (2024b). Diel and seasonal mixing patterns and water quality dynamics in a multipurpose tropical semiarid reservoir. *Environmental Science and Pollution Research International*, *31*(30), 43309-43322. https://doi.org/10.1007/s11356-024-34044-9.

Rocha, M. J. D., & Lima Neto, I. E. (2021). Modeling flow-related phosphorus inputs to tropical semiarid reservoirs. *Journal of Environmental Management*, *295*, 113123. https://doi.org/10.1016/j.jenvman.2021.113123.

Rocha, M. J. D., & Lima Neto, I. E. (2022). Internal phosphorus loading and its driving factors in the dry period of Brazilian semiarid reservoirs. *Journal of Environmental Management*, *312*, 114983. https://doi.org/10.1016/j.jenvman.2022.114983.

Rocha, M. J. D., & Lima Neto, I. E. (2023). Nitrogen mass balance and uptake velocity for eutrophic reservoirs in the Brazilian semiarid region. *Environmental Science and Pollution Research International*, *30*(42), 95621-95633. https://doi.org/10.1007/s11356-023-29136-x.

Santos, V. O., Guimarães, B. M. D. M., Lima Neto, I. E., Souza Filho, F. A., Rocha, P. A. R., Thé, J. V. G., & Gharabaghi, B. (2024). Chlorophyll-a estimation in 149 tropical semi-arid reservoirs using remote sensing data and six machine learning methods. *Remote Sensing*, *16*(11), 1870. https://doi.org/10.3390/rs16111870.

Silva, E. M. R., Almeida, L. G., Medeiros, P. H. A., Lima, G. D., Andrade, E. M., & Araújo, J. C. (2025). Reservoir eutrophication in the Brazilian semiarid: modeling of sediment removal and control of external loads as remediation measures. *Environmental Science and Pollution Research International*, *32*(12), 7663-7679. https://doi.org/10.1007/s11356-025-36183-z.

Tasnim, B., Fang, X., Hayworth, J. S., & Tian, D. (2021). Simulating nutrients and phytoplankton dynamics in lakes: model development and applications. *Water*, *13*(15), 2088. https://doi.org/10.3390/w13152088.

Toné, A. J. A., & Lima Neto, I. E. (2020). Modelagem simplificada do fósforo total em lagos e reservatórios brasileiros. *Revista DAE*, *68*(221), 142-156. https://doi.org/10.36659/dae.2020.012.

Wang, B., Hipsey, M. R., & Oldham, C. (2020). ML-SWAN-v1: A hybrid machine learning framework for the concentration prediction and discovery of transport pathways of surface water nutrients. *Geoscientific Model Development*, *13*(9), 4253-4272. https://doi.org/10.5194/gmd-13-4253-2020.

Wiegand, M. C., Nascimento, A. T. P., Costa, A. C., & Lima Neto, I. E. (2020). Evaluation of limiting nutrient of algal production in reservoirs of the Brazilian semiarid. *Brazilian Journal of Environmental Sciences*, *55*(4), https://doi.org/10.5327/Z2176-947820200681.

Wiegand, M. C., Nascimento, A. T. P., Costa, A. C., & Lima Neto, I. E. (2021). Trophic state changes of semi-arid reservoirs as a function of the hydro-climatic variability. *Journal of Arid Environments*, *184*, 104321. https://doi.org/10.1016/j.jaridenv.2020.104321.

Zhao, K., Wang, L., You, Q., Pan, Y., Liu, T., Zhou, Y., Zhang, J., Pang, W., & Wang, Q. (2021). Influence of cyanobacterial blooms and environmental variation on zooplankton and eukaryotic phytoplankton in a large, shallow, eutrophic lake in China. *The Science of the Total Environment*, *773*, 145421. https://doi.org/10.1016/j.scitotenv.2021.145421.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Fernando Mainardi Fan