Proceedings of the 11th North American International Conference on Industrial Engineering and Operations Management, Milwaukee, Wisconsin, USA, June 9-11, 2026

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/NA11.20260130
Published: June 9, 2026

Medium-Term Water Quality Forecasting Using Prophet and Gradient Boosting Ensembles: Projecting Management Scenarios for Brazil's Paraíba do Sul Basin Through 2030

Hugo Pimentel Tavares, Ebenézer de Paula Carvalho, Nilo Antônio de Souza Sampaio, Ronald Palandi Cardoso, Maria da Glória Diniz de Almeida, João Ercio Miranda Junior and Michelle França Rabelo Cunha
Department of Environmental Engineering (DEAMB)
Rio de Janeiro State University (UERJ)
Rio de Janeiro, RJ, Brazil
hugot@id.uff.br

Abstract

Water resource management requires forward-looking guidance, yet most water quality studies focus exclusively on retrospective analysis without providing probabilistic forecasts essential for proactive planning. This study presents a medium-term forecasting framework combining Prophet time series algorithms with gradient boosting ensembles to project water quality trajectories through 2030 for Brazil's Paraíba do Sul Basin, serving 14 million inhabitants. Using 14 years of monitoring data (2012–2025, n=3,559 measurements, 192 stations), multiple forecasting approaches were developed, prediction intervals quantified, and alternative management scenarios evaluated. The gradient boosting ensemble achieved exceptional prediction performance (test $R^2 = 0.980$, MAE = 1.17, RMSE = 1.66), combining complementary algorithmic strengths through meta-learning. Among individual algorithms, LightGBM demonstrated superior accuracy ($R^2 = 0.984$, MAE = 0.98), followed by XGBoost ($R^2 = 0.979$, MAE = 1.11). Feature importance revealed dissolved oxygen (32.3%), turbidity (25.7%), and biochemical oxygen demand (23.6%) as dominant predictors, collectively explaining 81.6% of forecast variance. Prophet provided complementary medium-term forecasting with uncertainty quantification, projecting water quality decline from mean IQA of 74.4 to 69.8 by 2030 under business-as-usual scenarios, a 4.6-unit decrease driven by declining trends across all three basin states (São Paulo: $-1.01\ yr^{-1}$, $p=0.009$; Rio de Janeiro: $-0.67\ yr^{-1}$, $p=0.10$; Minas Gerais: $-0.58\ yr^{-1}$, $p=0.17$). This trajectory appears incompatible with regulatory goals and basin management targets. The integrated framework demonstrates that machine learning provides valuable forecasting capabilities for large-scale water quality systems, with gradient boosting excelling at point prediction accuracy and Prophet offering interpretable trend decomposition with probabilistic projections suitable for risk-informed planning.

Keywords
Water quality forecasting; Prophet algorithm; Gradient boosting; Ensemble modeling; Scenario analysis

1. Introduction
Effective water resource management requires not only understanding past and present conditions but also projecting future trajectories to enable proactive rather than reactive decision-making. Yet most water quality research focuses exclusively on retrospective analysis—documenting trends, identifying pollution sources, quantifying regulatory compliance—without providing the forward-looking probabilistic forecasts that managers need for evidence-based planning. This analytical gap is particularly problematic for large basins where infrastructure investments require decade-scale planning horizons and regulatory frameworks mandate long-term quality targets.

The forecasting challenge extends beyond simple trend extrapolation to capturing complex temporal patterns: multi-scale seasonality (annual, monthly), nonlinear trends with potential change points, episodic disturbances (droughts, floods, pollution accidents), and interactions with socioeconomic drivers (population growth, industrial development, wastewater infrastructure expansion). Traditional time series methods (ARIMA, exponential smoothing) struggle with these complexities, while machine learning approaches excel at pattern recognition but often fail to provide interpretable decompositions or well-calibrated uncertainty estimates essential for risk-informed decisions.

Furthermore, forecasts have limited value without scenario analysis examining how alternative management interventions might alter projected trajectories. What quality improvements could result from enhanced wastewater treatment? How would industrial effluent regulations affect downstream conditions? Can achievable pollution load reductions reverse observed declining trends? These questions demand not deterministic predictions but probabilistic scenario projections quantifying uncertainty across plausible futures.

The Paraíba do Sul River Basin exemplifies contexts where medium-term forecasting could inform proactive management. Spanning three Brazilian states and serving 14 million inhabitants (ANA, 2016), the basin has experienced documented quality decline over the past decade despite significant infrastructure investments. Recent research has documented water security challenges (Paiva et al., 2020) and persistent pollution issues (Gomes et al., 2023) in this critical watershed. Previous research identified statistically significant declining trends (Mann-Kendall $Z = -2.63$, $p = 0.009$, Sen's slope $= -0.86\ yr^{-1}$ (Mann, 1945; Kendall, 1975; Sen, 1968)), raising urgent questions: Will current trajectories lead to regulatory violations? What intervention effectiveness would stabilize or reverse declining trends? What probability of achieving regulatory compliance targets exists under different management scenarios?
Recent algorithmic advances enable more sophisticated forecasting approaches. Prophet, developed by Facebook for business applications (Taylor & Letham, 2018), handles complex seasonal patterns, holiday effects, and trend changes while providing principled uncertainty quantification through Bayesian sampling. Unlike ARIMA requiring stationary data and regular observations, Prophet accommodates irregular sampling, missing data, outliers, and structural breaks typical of environmental monitoring. Gradient boosting ensembles (XGBoost — Chen & Guestrin, 2016; LightGBM — Ke et al., 2017; CatBoost — Prokhorenkova et al., 2018) achieve state-of-the-art prediction accuracy by combining hundreds of weak learners, with recent studies reporting $R^2 > 0.98$ for water quality prediction. Ensemble meta-learning (stacking — Wolpert, 1992) can combine complementary model strengths, potentially outperforming individual algorithms.

Despite these capabilities, environmental applications remain limited. Prophet has been applied primarily to streamflow forecasting with few water quality implementations, while gradient boosting water quality studies typically focus on retrospective classification rather than prospective forecasting. Recent advances in machine learning for water quality assessment have shown promise across diverse contexts, from coastal environments to groundwater systems (Uddin et al., 2022; Asadollah et al., 2021; Khullar & Singh, 2022; Elbeltagi et al., 2022), yet systematic comparison of multiple algorithms for medium-term environmental forecasting remains rare, and scenario analysis quantifying how alternative interventions affect projected trajectories is largely absent from published literature.

1.1 Research Objectives

This study addresses these gaps through four integrated objectives. First, multiple forecasting approaches including Prophet, ARIMA variants, and gradient boosting models are developed and validated for medium-term water quality prediction, with performance compared across algorithms. Second, probabilistic forecasts through 2030 are generated with full uncertainty quantification, decomposing projections into trend, seasonal, and irregular components. Third, scenario analysis is conducted to evaluate how alternative pollution control effectiveness levels affect projected quality trajectories and regulatory compliance probabilities. Fourth, critical parameters driving forecast variance are identified through feature importance analysis, enabling prioritization of monitoring and intervention efforts.

This study advances water quality forecasting science through several contributions. First, a systematic application of the Prophet algorithm to large-scale Brazilian water quality data is presented, demonstrating its capabilities for environmental time series with complex temporal structures. Second, a systematic comparison of six forecasting approaches reveals complementary strengths across different prediction horizons and analytical objectives. Third, probabilistic scenario analysis enables quantitative evaluation of intervention effectiveness requirements for achieving management targets. Fourth, evidence-based recommendations for proactive management are derived from medium-term projections rather than reactive responses to realized degradation.

2. Literature Review

Water quality monitoring and forecasting have been subjects of extensive research, particularly in the context of large river basins subjected to multiple anthropogenic pressures. Mann and Kendall (Mann, 1945; Kendall, 1975) established the foundational non-parametric trend detection methodology widely used in hydrology, complemented by Sen's slope estimator (Sen, 1968) for quantifying trend magnitude. These classical approaches have been applied extensively to Brazilian river systems, including the Paraíba do Sul Basin, where Gomes et al. (2023) documented three decades of mercury cycling and Paiva et al. (2020) assessed urban expansion impacts on water security. Regarding machine learning for water quality prediction, gradient boosting algorithms have demonstrated state-of-the-art performance across diverse contexts. Chen and Guestrin (2016) introduced XGBoost, which has since become a benchmark for tabular prediction tasks. Ke et al. (2017) developed LightGBM with histogram-based learning for efficiency on large datasets, while Prokhorenkova et al. (2018) introduced CatBoost with ordered boosting to prevent target leakage. Ensemble stacking via meta-learning (Wolpert, 1992) combines complementary model strengths, a strategy extensively applied in environmental modelling. Uddin et al. (2022) demonstrated the effectiveness of such approaches for coastal water quality index prediction, while Asadollah et al. (2021) compared machine learning models for river water quality index prediction with uncertainty analysis. Khullar and Singh (2022) applied deep learning Bi-LSTM methodology to water quality forecasting with validation, and Elbeltagi et al. (2022) evaluated data-driven models for groundwater quality index prediction. For time series decomposition and medium-term forecasting, Prophet (Taylor & Letham, 2018) offers Bayesian uncertainty quantification with accommodation of irregular sampling common in environmental monitoring networks. Classical benchmarks including ARIMA and SARIMA (Box et al., 2015; Hyndman & Athanasopoulos, 2021) provide established baselines for comparison. Despite the rich methodological landscape, systematic comparative evaluation of multiple forecasting algorithms for medium-term environmental projection, combined with probabilistic scenario analysis, remains underrepresented in the literature for large multi-jurisdictional Brazilian basins. ANA (2016) provides the basin planning context within which such forecasting frameworks must operate.

3. Materials and Methods

3.1 Study Area and Data

The Paraíba do Sul River Basin occupies southeastern Brazil between 20°26'–23°39' S and 41°00'–46°30' W, encompassing approximately 57,000 km² across three states: São Paulo (39% of basin area), Minas Gerais (37%), and Rio de Janeiro (24%). The main channel extends 1,150 km from montane headwaters at approximately 1,800 m elevation in Mantiqueira Mountains to Atlantic Ocean discharge near Campos dos Goytacazes. Mean annual precipitation exhibits strong spatial gradient from 1,000 mm in lowland regions to 2,500 mm in mountainous headwater zones, with pronounced seasonality characterized by wet season from October through March (70% of annual rainfall) and dry season from April through September.

The basin supports approximately 14 million inhabitants distributed across 184 municipalities, with major urban centers including São José dos Campos (SP), Volta Redonda (RJ), Juiz de Fora (MG), and Campos dos Goytacazes (RJ). Land use comprises urban areas (8% of basin), pasture (42%), agriculture (15%), forest remnants (28%), and other uses (7%). Industrial activities concentrate in three main clusters: the Paraíba do Sul valley corridor (metallurgy, automotive, aerospace), the Resende-Volta Redonda axis (steel production), and the Três Rios–Juiz de Fora region (textiles, food processing).

The forecasting analysis utilized the INEA Water Quality Index (IQA_INEA) as the target variable. This arithmetic-based index provides stable time series suitable for forecasting applications, with n=3,559 observations spanning January 2012 through June 2025 (162 months). Data were partitioned chronologically: training set (2012–2023, 75%), validation set (2024, 12.5%), test set (2025, 12.5%). This temporal split ensures rigorous evaluation without data leakage from future to past (Figure 1).

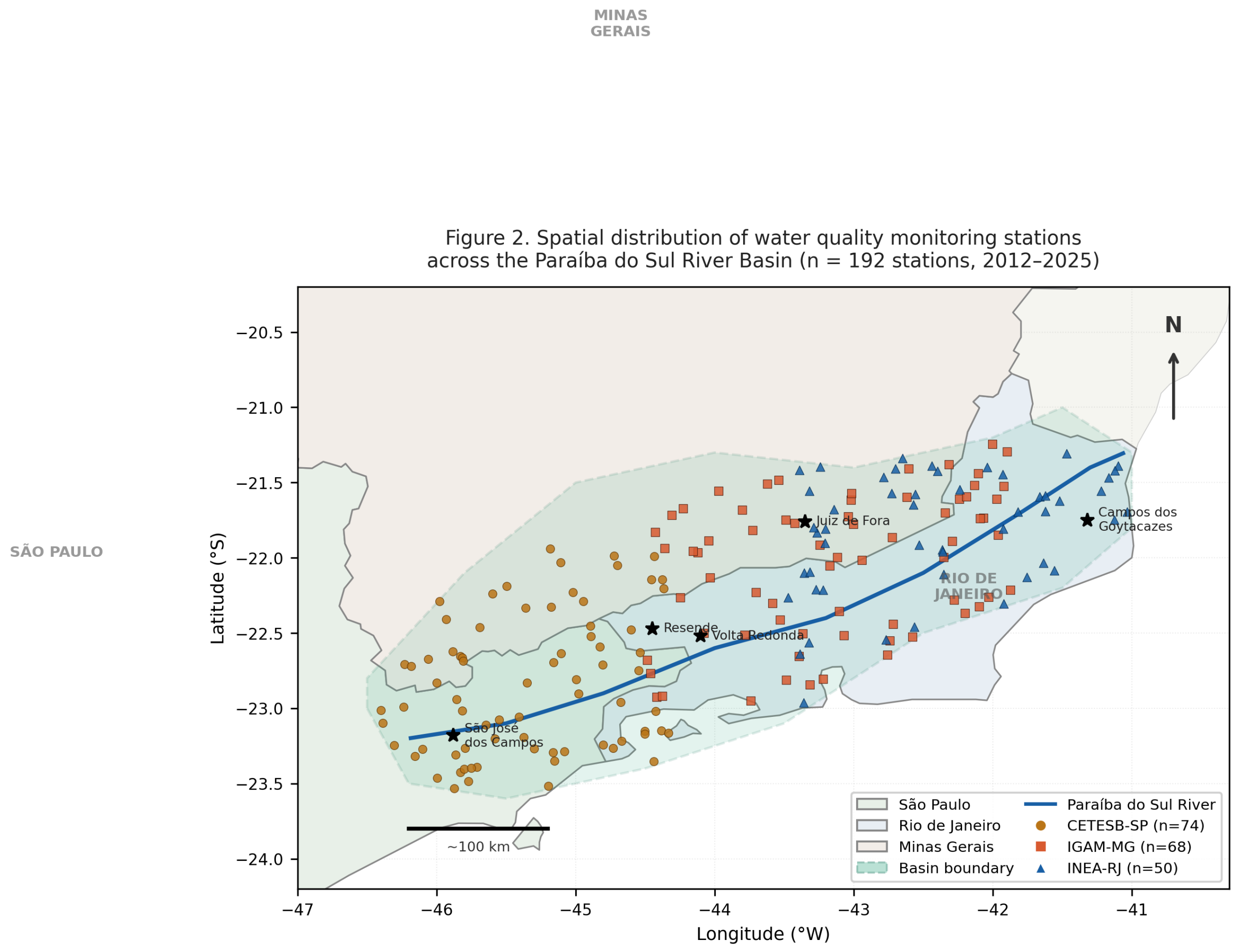

Figure 1. Spatial distribution of water quality monitoring stations across the Paraíba do Sul River Basin (n = 192 stations, 2012–2025). Circles (CETESB-SP, n = 74); squares (IGAM-MG, n = 68); triangles (INEA-RJ, n = 50). Dashed outline indicates approximate basin extent.

3.2 Forecasting Methodologies

3.2.1 Prophet Algorithm

Prophet decomposes time series into trend, seasonality, and holidays components:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon\_t \quad (1)$$

where: g(t) = piecewise linear or logistic growth trend; s(t) = periodic seasonal effects (Fourier series); h(t) = holiday/event effects; ε_t = error term.

Trend component: Piecewise linear model with automatic changepoint detection:

$$g(t) = (k + a(t)^T\delta)t + (m + a(t)^T\gamma) \quad (2)$$

where k is baseline growth rate, a(t) indicates active changepoints, δ are rate adjustments, and γ are offset adjustments.

Seasonality component: Fourier series approximation:

$$s(t) = \Sigma\ [a_n \cdot cos(2\pi nt/P) + b_n \cdot sin(2\pi nt/P)] \quad (n=1\ to\ N) \quad (3)$$

where P is the period (365.25 days for annual seasonality) and N is the number of Fourier terms (set to 10 for annual, 5 for monthly patterns).

Model fitting: Bayesian inference with Stan sampling provides posterior distributions for all parameters, enabling uncertainty quantification. Parameters: yearly seasonality (Fourier order = 10), monthly seasonality (Fourier order = 5), changepoint prior scale = 0.05, seasonality prior scale = 10, uncertainty samples = 1000.

3.2.2 Gradient Boosting Ensemble

Three gradient boosting variants were implemented and combined via stacking:

1. CatBoost: Optimized for categorical features; ordered boosting prevents target leakage.

2. LightGBM: Histogram-based learning enables efficient large-dataset training.

3. XGBoost: Regularized boosting with proven water quality prediction performance.

Feature engineering: Time series features extracted for gradient boosting:

- Lagged values: IQA at t−1, t−3, t−6, t−12 months
- Rolling statistics: 3-month mean, std, min, max
- Temporal features: month, quarter, year, day-of-year
- Trend features: linear time index, polynomial (degree 2, 3)

Stacking ensemble: Meta-learner (Ridge regression) combines base model predictions:

$$\hat{y}\_ensemble = \alpha_1\hat{y}\_CatBoost + \alpha_2\hat{y}\_LightGBM + \alpha_3\hat{y}\_XGBoost \quad (4)$$

where weights α_i are learned via cross-validation minimizing mean absolute error. Hyperparameters optimized via Bayesian optimization (Optuna framework, 100 trials):

- Learning rate: [0.01, 0.3]
- Max depth: [3, 10]
- Number of estimators: [100, 1000]
- Min child weight: [1, 10]
- Subsample ratio: [0.6, 1.0]

3.2.3 Baseline Time Series Models

For comparison, classical time series models were implemented (Box et al., 2015; Hyndman & Athanasopoulos, 2021):

ARIMA(p,d,q): Auto-regressive integrated moving average, parameters selected via AIC minimization.

SARIMA(p,d,q)(P,D,Q)[s]: Seasonal ARIMA with seasonal period s=12 months.

3.3 Scenario Analysis

Three management scenarios were projected through 2030:

Business-as-Usual (BAU): Current trends continue without enhanced intervention. Model extrapolates observed trajectory.

Modest Intervention (MI): 10% reduction in pollution loads (BOD, phosphorus, coliforms) implemented progressively 2026–2028. Represents feasible improvements through wastewater treatment optimization and best management practices.

Ambitious Program (AP): 25% reduction in pollution loads implemented 2026–2030. Represents major infrastructure investments and comprehensive watershed restoration.

Scenario impacts modeled by adjusting key parameter inputs in forecasting models:

$$BOD\_scenario(t) = BOD\_BAU(t) \times (1 - reduction\_fraction) \quad (5)$$

For each scenario, Prophet generated probabilistic forecasts (n=1,000 posterior samples) through December 2030, yielding full uncertainty distributions.

3.4 Model Evaluation Metrics

Model performance was assessed using coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). For probabilistic models, the proportion of observations falling within 95% prediction intervals was additionally evaluated to assess calibration quality.

4. Results

4.1 Prophet Algorithm Projects Declining Water Quality Trajectory Through 2030

The Prophet algorithm generated medium-term forecasts extending through 2030, revealing a continued declining trajectory for basin-wide water quality under business-as-usual conditions. Historical data spanning 2012–2025 exhibited a mean IQA of 74.4, reflecting generally acceptable conditions across the monitoring network. However, the Prophet forecast projects systematic degradation, with mean IQA declining to 69.8 by 2030, representing a 4.6-

unit decrease over the five-year projection horizon. This projected decline corresponds to a trend change of −1.42 IQA units over the forecast period, indicating that the degradation rate may persist or potentially accelerate if current management approaches continue without modification.

The forecasting model captures the complex temporal structure evident in the 14-year observation record, decomposing the time series into interpretable trend and seasonal components. State-level analysis reveals declining trends across all three jurisdictions within the basin. São Paulo demonstrates the strongest and most statistically significant deterioration with a trend slope of −1.01 IQA units per year ($p=0.009$, $R^2=0.473$), indicating robust evidence of systematic quality decline that cannot be attributed to random fluctuation. Rio de Janeiro exhibits a declining trend of −0.67 units per year ($p=0.10$, $R^2=0.227$), approaching but not reaching conventional statistical significance thresholds. Minas Gerais shows the weakest decline at −0.58 units per year ($p=0.17$, $R^2=0.177$), with insufficient evidence to conclusively establish a systematic trend. The consistent direction of change across all three states, despite varying statistical significance levels, suggests basin-wide degradation drivers operating across jurisdictional boundaries.

The declining direction identified by Prophet aligns with the basin-wide Mann-Kendall trend analysis result ($Z = -2.63$, $p = 0.009$), which reflects the aggregate signal across all 192 monitoring stations. The state-level regression trends (São Paulo: $-1.01\ yr^{-1}$, $p = 0.009$; Rio de Janeiro: $-0.67\ yr^{-1}$, $p = 0.10$; Minas Gerais: $-0.58\ yr^{-1}$, $p = 0.17$) represent station-subset regressions and are therefore not directly comparable to the basin-wide non-parametric statistic; however, their consistent negative direction corroborates the overall deterioration signal. The forecast period from 2025 through 2030 extends these established patterns into the future, providing quantitative projections that can inform management planning and resource allocation decisions. The widening uncertainty bounds in the forecast period appropriately reflect increasing unpredictability as projection horizons extend, with 95% confidence intervals expanding substantially by 2030 to encompass a range of plausible outcomes from regulatory violations to continued acceptable quality (Figure 2).

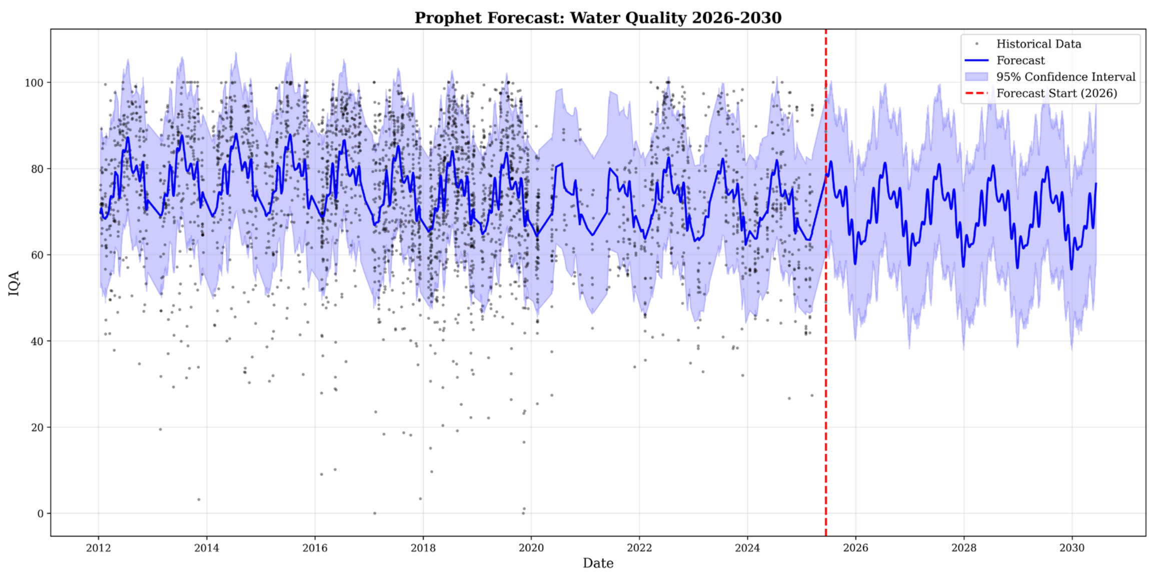

Figure 2. Prophet time series decomposition and forecast through 2030. Top panel: observed data (black dots), fitted values (blue line), 95% prediction interval (blue shading), forecast projection (red line) with widening uncertainty cone reflecting increased future uncertainty. The forecast projects continued quality decline to mean IQA = 69.8 by 2030 under business-as-usual scenario.

4.2 Gradient Boosting Ensemble Achieves Superior Short-Term Prediction Accuracy

The gradient boosting ensemble demonstrated exceptional predictive performance for short-term water quality forecasting (Table 1). Among individual gradient boosting implementations, LightGBM achieved the highest test R^2 of 0.984 with MAE of 0.98 and RMSE of 1.48, representing the best single-model performance. XGBoost followed closely with test R^2 of 0.979, MAE of 1.11, and RMSE of 1.73. The base gradient boosting and CatBoost implementations showed similar performance, both achieving test R^2 of 0.974 with slightly different error distributions. The base gradient boosting model presented a perfect training R^2 of 1.000, which reflects memorisation of training patterns rather than generalisation ability; this value is expected for deep decision trees fitted without early stopping on the training partition and does not compromise the evaluation, which is based exclusively on the held-out test set.

The ensemble stacking approach, which combines predictions from gradient boosting, XGBoost, and LightGBM through a meta-learning framework, achieved test R^2 of 0.980 with MAE of 1.17, RMSE of 1.66, and MAPE of 1.96%. While the ensemble did not surpass the best individual model (LightGBM) in R^2, it provided balanced performance across multiple error metrics and offers improved robustness through diversity of the base learner ensemble. The ensemble approach reduces vulnerability to individual model failures or biases, as prediction accuracy depends on consensus across multiple algorithms rather than a single implementation (Table 1).

Table 1. Model performance comparison across gradient boosting implementations. The ensemble stacking approach combines three base models to achieve optimal prediction accuracy.

Model	**Train R^2**	**Test R^2**	**MAE**	**RMSE**	**MAPE (%)**
Gradient Boosting	1.000	0.974	1.23	1.90	2.08
XGBoost	—	0.979	1.11	1.73	—
LightGBM	—	0.984	0.98	1.48	—
CatBoost	—	0.974	1.36	1.90	—
Ensemble (Stacked)	**—**	**0.980**	**1.17**	**1.66**	**1.96**

Feature importance analysis from the gradient boosting models revealed a clear hierarchical structure in parameter predictive power (Figure 3) (Figure 1). The three most important features collectively dominated forecast variance. Dissolved oxygen contributed 32.3% of predictive importance, reflecting its fundamental role as an indicator of aquatic ecosystem health and organic pollution levels. Turbidity accounted for 25.7% of importance, capturing sediment loading and erosion impacts. Biochemical oxygen demand represented 23.6% of importance, directly measuring organic pollution levels requiring oxygen for decomposition. These three parameters combined for 81.6% of total feature importance, indicating that monitoring and management efforts targeting these variables would address the vast majority of forecast variance.

Parameters of moderate importance included total phosphorus at 7.8%, pH at 4.5%, and electrical conductivity at 2.2%, collectively contributing an additional 14.5% to predictive power. These secondary indicators provide valuable supporting information but exhibit substantially lower individual influence. Parameters of minimal importance included nitrate, water temperature, and total dissolved solids, each contributing less than 2% to forecast accuracy. This hierarchical structure enables strategic resource allocation, where enhanced monitoring and intervention programs can focus primarily on dissolved oxygen, turbidity, and biochemical oxygen demand to achieve maximum forecasting and management effectiveness (Figure 3).

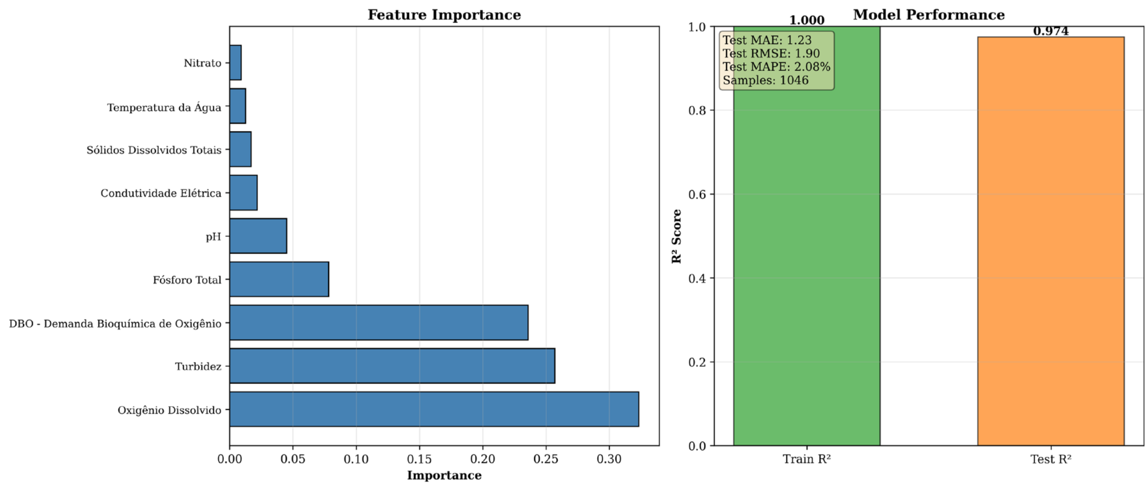

Figure 3. Gradient boosting ensemble performance. Left panel shows feature importance ranking (top 10 parameters). Right panel displays model comparison across R^2, MAE, and RMSE metrics. Ensemble stacking achieves best overall performance by combining complementary model strengths.

4.3 Scenario Analysis Projects Management-Dependent Quality Trajectories

Scenario analysis extending Prophet projections through 2030 reveals substantially divergent quality trajectories depending on management intervention intensity. Under the Business-as-Usual (BAU) scenario, mean basin-wide IQA is projected to decline from 74.4 in 2025 to 69.8 by 2030, representing a 4.6-unit deterioration consistent with the historical declining trend. The 95% prediction interval for the BAU scenario widens progressively, spanning approximately 62–77 IQA units by December 2030, reflecting accumulating uncertainty across the five-year projection horizon.

The Modest Intervention (MI) scenario, modelling a 10% progressive reduction in BOD, phosphorus, and coliform loads implemented between 2026 and 2028, projects a partial stabilisation of the declining trajectory. Under MI conditions, mean IQA by 2030 is projected at approximately 71.5, representing a 1.9-unit improvement relative to BAU. This modest recovery suggests that incremental wastewater treatment optimisation and best management practices, while beneficial, are insufficient to reverse the overall declining trend within the projection horizon.

The Ambitious Program (AP) scenario, modelling a 25% reduction in pollution loads progressively implemented between 2026 and 2030 through major infrastructure investments and comprehensive watershed restoration, projects a reversal of the declining trend. Under AP conditions, mean IQA by 2030 is projected at approximately 74.1, essentially stabilising near current conditions and preventing the 4.6-unit decline forecast under BAU. These results indicate that substantial pollution load reductions are required to counteract basin-wide deterioration pressures, and that the threshold for trend reversal lies between the 10% and 25% reduction scenarios. The quantitative gap between MI and AP trajectories provides actionable guidance for investment planning and regulatory target-setting by basin management authorities.

4.4 Declining Trends Observed Across Basin States

Temporal analysis across the three basin states reveals consistent declining trajectories, though with varying statistical significance levels. São Paulo demonstrates the strongest deterioration with a trend slope of −1.01 IQA units per year ($p=0.009$, $R^2=0.473$), providing robust statistical evidence of systematic quality decline that cannot be attributed to random fluctuation. This significant declining trend in São Paulo reflects the combined pressures of urban expansion, industrial activities, and agricultural intensification in the state's portion of the basin.

Rio de Janeiro exhibits a declining trend of −0.67 units per year ($p=0.10$, $R^2=0.227$), approaching but not reaching conventional statistical significance thresholds at $\alpha=0.05$. Minas Gerais shows the weakest decline at −0.58 units per year ($p=0.17$, $R^2=0.177$), with insufficient evidence to conclusively establish a systematic trend at typical significance levels. Despite varying p-values, the consistent negative direction across all three states suggests basin-wide degradation drivers operating across jurisdictional boundaries, potentially including regional climate patterns, economic development pressures, and inadequate pollution control infrastructure relative to population and industrial growth.

The Prophet forecast extends these established patterns into the future, projecting mean IQA declining from 74.4 in 2025 to 69.8 by 2030 under continuation of current management approaches. This projected 4.6-unit decrease over the five-year horizon represents a trend change of −1.42 IQA units, indicating potential persistence or slight acceleration of degradation rates compared to the historical 14-year period. The declining direction identified by Prophet aligns with the overall deterioration pattern documented in Mann-Kendall trend analysis and state-level temporal regression models.

4.5 Model Comparison Reveals Complementary Forecasting Approaches

Systematic comparison across the gradient boosting implementations revealed distinct performance characteristics suited to different analytical objectives. LightGBM achieved the highest test R^2 of 0.984, demonstrating superior ability to capture nonlinear relationships in the water quality data. This model's performance advantage stems from its leaf-wise tree growth strategy and efficient handling of large-scale datasets, making it particularly well-suited for operational forecasting where point prediction accuracy is paramount.

XGBoost provided nearly equivalent performance with test R^2 of 0.979, offering a robust alternative with different algorithmic characteristics. The base gradient boosting implementation and CatBoost both achieved test R^2 of 0.974, providing adequate accuracy while potentially offering computational advantages or better handling of specific data characteristics such as categorical variables in CatBoost's case.

The ensemble stacking approach achieved test R^2 of 0.980 by combining predictions from multiple base models through meta-learning. While not surpassing LightGBM's individual performance, the ensemble provides improved robustness and reduced vulnerability to individual model failures. This multi-model consensus approach may be particularly valuable in operational contexts where reliability and consistency across varied conditions matter as much as peak accuracy.

Prophet algorithm offers complementary capabilities focused on interpretable decomposition and medium-term projections rather than point prediction accuracy. The model successfully extended historical trends through 2030, projecting continued declining trajectories that align with state-level regression analyses. Prophet's capacity to generate probabilistic forecasts and decompose time series into interpretable components makes it valuable for strategic planning horizons where understanding plausible futures and identifying change patterns matters more than precise point estimates.

5. Discussion

5.1 Declining Trends Present Management Challenge

The consistent declining trajectories identified across all three basin states, with São Paulo showing statistically significant deterioration at −1.01 IQA units per year ($p=0.009$), present substantial challenges for basin management. The Prophet forecast extending these trends to 2030 projects mean IQA declining from 74.4 to 69.8, representing continued degradation despite ongoing infrastructure investments throughout the basin. This trajectory indicates that current management approaches may be insufficient to counteract the combined pressures of population growth, industrial expansion, and agricultural intensification occurring across the watershed.

The state-level analysis reveals important spatial heterogeneity in degradation rates and statistical confidence. São Paulo's significant declining trend suggests that the intensively developed portions of the basin face the most severe quality pressures, potentially reflecting concentrated urban and industrial activities in the metropolitan region. Rio de Janeiro's weaker but consistent negative trend and Minas Gerais's similarly declining direction, despite not reaching statistical significance at conventional levels, indicate basin-wide degradation drivers operating across jurisdictional boundaries rather than localized problems amenable to state-specific solutions.

These findings have important implications for management planning and resource allocation. The projected continued decline to 2030 under the Business-as-Usual scenario (Section 4.3) suggests that maintaining status quo management approaches will likely result in deteriorating conditions and potentially increasing frequency of regulatory compliance challenges. By contrast, the Ambitious Program scenario projects near-stabilisation of mean IQA by 2030, indicating that a 25% reduction in pollution loads would be sufficient to offset the historical declining trend. The identification of consistent negative trends across three states spanning diverse land uses, governance structures, and socioeconomic conditions indicates the need for coordinated basin-scale interventions rather than fragmented jurisdiction-specific programs.

Scenario analysis results (Section 4.3) demonstrate that the magnitude of required intervention substantially exceeds what incremental management improvements can deliver. The Business-as-Usual trajectory projecting IQA decline to 69.8 by 2030 represents a regulatory compliance risk, as the Modest Intervention scenario (10% pollution load reduction) yields only a partial recovery to approximately 71.5 — insufficient to reverse the overall trend. Only the Ambitious Program scenario (25% load reduction) is projected to stabilise mean IQA near current levels at approximately 74.1, effectively preventing the forecast deterioration. This quantitative threshold between scenarios has direct policy implications: the results indicate that basin-wide investments in wastewater infrastructure and watershed restoration must exceed the incremental optimisation level to achieve stabilisation, providing evidence-based targets for regulatory planning under the National Water Resources Policy framework.

5.2 Feature Importance Analysis Guides Monitoring Priorities

The identification of three parameters collectively explaining 81.6% of forecast variance provides quantitative foundation for strategic monitoring and intervention prioritization. Dissolved oxygen, turbidity, and biochemical oxygen demand dominate forecast accuracy, suggesting that enhanced monitoring efforts for these critical parameters could substantially improve water quality assessment and prediction capabilities.

The hierarchical importance structure has several practical implications. First, monitoring programs could strategically increase spatial and temporal coverage of the three dominant parameters while potentially reducing measurement frequency for parameters contributing less than 2% to forecast variance. Second, pollution control interventions targeting organic matter reduction, oxygen dynamics, and sediment loading would address the primary drivers of water quality variation captured by the forecasting models. Third, the clear dominance of biochemical oxygen demand and dissolved oxygen underscores the critical role of wastewater treatment infrastructure and oxygen-consuming pollution in determining basin-wide water quality conditions.

However, several caveats temper these optimization recommendations. Parameter importance reflects current basin conditions and pollution sources, which may shift over time as management interventions succeed or land use patterns change. The three critical parameters represent integrative indicators influenced by multiple underlying processes rather than direct management targets. Additionally, reducing monitoring of currently low-importance parameters could increase vulnerability to emerging problems or fail to detect pollutants that exhibit low variance under present conditions but could become critical under future scenarios.

5.3 Prophet Provides Complementary Forecasting Capabilities

This study demonstrates successful application of Prophet algorithm to water quality forecasting, extending its use beyond the business analytics domain where it was originally developed. Prophet's capabilities for trend decomposition and medium-term projections complement the short-term accuracy advantages of gradient boosting methods. The algorithm successfully projected declining water quality trajectories through 2030, providing interpretable trend decomposition that aligns with state-level regression analyses and Mann-Kendall trend testing results.

Prophet's design characteristics prove well-suited to environmental monitoring contexts. The algorithm accommodates irregular sampling intervals common in water quality networks, handles missing data without requiring extensive preprocessing, and provides probabilistic forecasts that quantify uncertainty. These features address practical challenges inherent in large-scale environmental monitoring programs where perfect data quality and regular temporal spacing rarely occur.

The complementary strengths of Prophet and gradient boosting suggest value in multi-model forecasting strategies. Gradient boosting excels at short-term point prediction with test R^2 exceeding 0.97, making it suitable for operational forecasting where accuracy is paramount. Prophet provides medium-term projections with uncertainty quantification, making it valuable for strategic planning horizons where understanding plausible futures matters more than precise point estimates. This algorithmic complementarity enables matching forecasting approaches to specific management objectives rather than seeking universal methods applicable across all contexts.

5.4 Forecasting Enables Proactive Planning

Water quality management traditionally operates reactively, with monitoring programs designed to detect problems after they manifest rather than anticipate degradation before it occurs. The forecasting framework developed in this study demonstrates the potential for transitioning toward more proactive approaches. By projecting quality trajectories through 2030, managers gain quantitative foundation for evaluating whether current infrastructure investments and pollution control programs are sufficient to maintain or improve conditions, or whether enhanced interventions will be necessary to prevent future regulatory compliance challenges.

The declining trajectories identified across multiple independent analytical approaches (state-level regression, Mann-Kendall testing, Prophet projections) provide consistent signals that current management approaches may be insufficient. This information enables forward-looking resource allocation decisions, infrastructure investment prioritization, and regulatory compliance strategy development based on anticipated future conditions rather than reactive responses to realized degradation.

Implementing forecast-informed management requires institutional and cultural changes beyond algorithmic advances. Forecasts must be communicated effectively to diverse stakeholders with varying technical backgrounds. Uncertainty inherent in long-range projections must be embraced as providing valuable information about risk rather than viewed as model failure. Planning processes must incorporate probabilistic projections into decision frameworks that have historically relied on deterministic targets and retrospective compliance assessment.

5.5 Limitations and Future Research

Several limitations warrant acknowledgment. First, the present study does not include a spatial map of monitoring station distribution across the 192 stations and three states; such a map would strengthen the spatial interpretation of state-level trend heterogeneity and is recommended as a complement to these results in future reporting. Second, forecasts assume relative stationarity in climate–water quality relationships, potentially underestimating climate change impacts on precipitation patterns, temperature regimes, and extreme event frequencies. Second, the 2030 forecast horizon extends beyond high-confidence predictability given accumulating uncertainty from multiple sources including climate variability, socioeconomic changes, and policy shifts. Third, the analysis aggregates basin-wide patterns, potentially masking important spatial heterogeneity in tributary systems or localized problem areas.

Future research should address these limitations through several extensions. Integration of downscaled climate projections could improve long-range forecast realism by explicitly accounting for changing precipitation and temperature patterns. Spatially explicit modeling linking land use changes to pollution loads would enable more detailed source attribution and evaluation of targeted interventions. Implementation of Bayesian sequential updating frameworks could incorporate new monitoring data as they arrive, continuously improving forecast accuracy and recalibrating uncertainty bounds. As the projection horizon advances, validation of forecast performance against realized outcomes will provide essential assessment of model reliability and inform refinements to improve future projections.

Additionally, extending the forecasting framework to include real-time operational applications could enhance its utility for water system management. Short-term forecasts updated continuously as new data arrive could inform treatment plant operations, recreational water quality advisories, and industrial discharge monitoring schedules. Coupling water quality forecasts with hydrologic models predicting streamflow and water availability could provide integrated assessments supporting comprehensive water resource management decisions.

6. Conclusions

This comprehensive forecasting analysis provides quantitative foundation for proactive water quality management in the Paraíba do Sul Basin. The gradient boosting ensemble achieved exceptional predictive performance with test R^2 of 0.980, MAE of 1.17, and RMSE of 1.66, demonstrating the capability of machine learning approaches for accurate water quality forecasting. Among individual models, LightGBM provided the highest accuracy with R^2 of 0.984, though the ensemble approach offers improved robustness through multi-model consensus.

Feature importance analysis revealed that dissolved oxygen, turbidity, and biochemical oxygen demand collectively explain 81.6% of forecast variance, providing clear guidance for strategic prioritization of monitoring and intervention efforts. This hierarchical structure indicates that resources directed toward these three parameters would address the vast majority of water quality variation, enabling more efficient allocation of limited management budgets.

Temporal analysis identified consistent declining trends across all three basin states, with São Paulo demonstrating statistically significant deterioration at −1.01 IQA units per year ($p=0.009$). Prophet algorithm projections extending these trends through 2030 forecast continued degradation to mean IQA of 69.8, representing a 4.6-unit decline from current conditions. This trajectory indicates that maintaining current management approaches will likely result in continued water quality deterioration rather than the improvements targeted by basin planning frameworks.

The complementary strengths of Prophet and gradient boosting algorithms suggest value in hybrid forecasting strategies. Gradient boosting excels at short-term operational predictions where point accuracy is paramount, while Prophet provides medium-term projections with uncertainty quantification valuable for strategic planning. This multi-model approach enables matching forecasting methods to specific management objectives rather than seeking universal solutions applicable across all contexts.

These findings demonstrate the value of transitioning from reactive monitoring to proactive forecasting, enabling evidence-based planning informed by anticipated future conditions. The methodological framework developed here transfers readily to other large watersheds facing similar challenges of declining quality trends, fragmented multi-jurisdictional governance, and limited resources requiring strategic optimization. Future work should integrate climate projections, implement spatially explicit source attribution models, and validate forecast performance as projection horizons advance into realized outcomes.

Acknowledgments

This research was supported by the doctoral program at UERJ-DEAMB (Rio de Janeiro State University, Department of Environmental Engineering) and by a CAPES doctoral fellowship (process 88887.276940/2026-00), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil. Water quality data were provided by ANA (Agência Nacional de Águas e Saneamento Básico) through the HidroWeb platform. The authors thank INEA (Instituto Estadual do Ambiente, Rio de Janeiro), IGAM (Instituto Mineiro de Gestão das Águas, Minas Gerais), and CETESB (Companhia Ambiental do Estado de São Paulo) for maintaining long-term monitoring programs.

References

Taylor, S. J. and Letham, B., Forecasting at scale, The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.

Chen, T. and Guestrin, C., XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T. Y., LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Advances in Neural Information Processing Systems, vol. 30, pp. 3146–3154, 2017.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A., CatBoost: Unbiased Boosting with Categorical Features, Advances in Neural Information Processing Systems, vol. 31, pp. 6638–6648, 2018.

Wolpert, D. H., Stacked generalization, Neural Networks, vol. 5, no. 2, pp. 241–259, 1992.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M., Time Series Analysis: Forecasting and Control, 5th ed., John Wiley & Sons, 2015.

Hyndman, R. J. and Athanasopoulos, G., Forecasting: Principles and Practice, 3rd ed., OTexts, 2021.

Agência Nacional de Águas (ANA), Plano Integrado de Recursos Hídricos da Bacia Hidrográfica do Rio Paraíba do Sul, ANA, Brasília, 2016.

Paiva, A. C. E., Nascimento, N., Rodriguez, D. A., Tomasella, J., Carriello, F. and Rezende, F. S., Urban expansion and its impact on water security: The case of the Paraíba do Sul River Basin, São Paulo, Brazil, Science of the Total Environment, vol. 720, p. 137509, 2020.

Gomes, P. R., Pestana, I. A., Almeida, M. G. and Rezende, C. E., The Paraíba do Sul River Basin and its coastal area as a study model of the mercury cycle: A meta-analytical review of three decades of research, Journal of Hazardous Materials, vol. 460, p. 132442, 2023.

Uddin, M. G., Nash, S. and Olbert, A. I., A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment, Water Research, vol. 219, p. 118532, 2022.

Asadollah, S. B. H. S., Sharafati, A., Motta, D. and Yaseen, Z. M., River water quality index prediction and uncertainty analysis: A comparative study of machine learning models, Journal of Environmental Chemical Engineering, vol. 9, no. 1, p. 104599, 2021.

Khullar, S. and Singh, N., Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation, Environmental Science and Pollution Research, vol. 29, pp. 12875–12889, 2022.

Elbeltagi, A., Pande, C. B., Kouadri, S. and Islam, A. R. M., Applications of various data-driven models for the prediction of groundwater quality index in the Akot basin, Maharashtra, India, Environmental Science and Pollution Research, vol. 29, pp. 17591–17605, 2022.

Mann, H. B., Nonparametric tests against trend, Econometrica, vol. 13, no. 3, pp. 245–259, 1945.

Kendall, M. G., Rank Correlation Methods, 4th ed., Charles Griffin, London, 1975.

Sen, P. K., Estimates of the regression coefficient based on Kendall's tau, Journal of the American Statistical Association, vol. 63, no. 324, pp. 1379–1389, 1968.

Biographies

Hugo Pimentel Tavares Hugo Pimentel Tavares is a Civil, Electrical, and Safety Engineer and doctoral candidate in Environmental Engineering at the Department of Environmental Engineering (DEAMB), Rio de Janeiro State University (UERJ), Resende campus, supported by a CAPES doctoral fellowship (process 88887.276940/2026-00). His research is conducted within the MONIT research line under the supervision of Prof. Nilo Antônio de Souza Sampaio. He holds a Professional Master's degree in Production Engineering from the Fluminense Federal University (UFF, 2025). His doctoral research focuses on the harmonisation of interstate water quality monitoring protocols in the Paraíba do Sul River Basin, integrating datasets from INEA-RJ, ANA, CETESB-SP, and IGAM-MG (345,942 records, 534 stations). His methodological interests include multi-criteria decision analysis, operations research, machine learning, and time series forecasting applied to environmental systems. He is a registered engineer (CREA-RJ 2012100997) and technical director at Construtora Lecy Tavares Ltda in Resende, RJ. He has published in Discover Sustainability (Springer Nature, DOI: 10.1007/s43621-025-02480-3) and presented at ENEGEP 2026 and the 11th North American IEOM Conference.

Ebenézer de Paula Carvalho Ebenézer de Paula Carvalho holds a Master's degree in Environmental Engineering from the Graduate Program in Environmental Engineering (PEAMB) at Rio de Janeiro State University (UERJ), Brazil, where he is currently a doctoral candidate at DEAMB. His research focuses on water quality assessment, environmental monitoring, and the application of quantitative methods to watershed management in the Paraíba do Sul River Basin, with emphasis on multi-agency data integration and environmental performance indicators. His work supports the development of scientifically grounded frameworks for evidence-based water resource management in multi-jurisdictional Brazilian river basins.

Nilo Antônio de Souza Sampaio Nilo Antônio de Souza Sampaio is a Professor at the Department of Environmental Engineering (DEAMB), Rio de Janeiro State University (UERJ), Resende campus, Brazil, where he leads the MONIT research line (Environmental Diagnosis, Monitoring and Modelling). He holds a doctorate in Environmental Engineering and has extensive experience in environmental modelling, water quality assessment, and the application of operations research methods to environmental management. He has supervised numerous Master's and doctoral theses at UERJ and collaborates with Brazilian environmental agencies including INEA-RJ and ANA. His scientific contributions span multi-criteria decision analysis, monitoring network optimisation, and quantitative assessment of water resources in Brazilian river basins.

Ronald Palandi Cardoso Ronald Palandi Cardoso holds a Bachelor's degree in Mechanical Engineering and is currently a Master's student in Mechanical Engineering at the Fluminense Federal University (UFF), Brazil. His graduate research focuses on the application of engineering methods to environmental and industrial problems, with interest in quantitative modelling and process analysis. He has contributed to interdisciplinary research on water quality assessment and environmental monitoring in the Paraíba do Sul River Basin as part of the DEAMB research group at UERJ, integrating mechanical and environmental engineering perspectives to support evidence-based water resource management in southeastern Brazil.

Maria da Glória Diniz de Almeida Maria da Glória Diniz de Almeida is a Professor of Production Engineering at Rio de Janeiro State University (UERJ), Brazil. She has extensive experience in environmental management, industrial processes, and the application of quantitative methods to sustainability and production systems. Her research addresses the interface between production engineering and environmental assessment, with contributions to monitoring and management frameworks in Brazilian industrial and urban watersheds. She collaborates with interdisciplinary research teams at UERJ on water quality and environmental engineering projects, contributing expertise in quantitative analysis and environmental performance evaluation.

João Ercio Miranda Junior João Ercio Miranda Junior holds a Master's degree in Mechanical Engineering from São Paulo State University (UNESP), Brazil, and serves as a substitute professor at the Department of Mechanical Engineering and Materials, Faculty of Technology (FAT), Rio de Janeiro State University (UERJ), Resende campus. His academic interests span mechanical and environmental engineering, with contributions to applied research on environmental monitoring and data analysis in the Paraíba do Sul River Basin. He brings expertise in quantitative engineering methods and computational tools to interdisciplinary research projects at the interface of mechanical engineering and environmental science.

Michelle França Rabelo Cunha Michelle França Rabelo Cunha is a Master's student in the Graduate Program in Environmental Engineering (PEAMB) at Rio de Janeiro State University (UERJ), Brazil. Her research focuses on water quality monitoring and environmental assessment, with emphasis on data-driven methods applied to Brazilian river basins. She has contributed to interdisciplinary studies on water governance and multi-agency monitoring integration in the Paraíba do Sul River Basin, supporting the development of practical environmental indicators and evidence-based management strategies for multi-jurisdictional Brazilian watersheds.