International Journal of
Civil and Environmental Engineering
https://www.ioro.org/ijcee

Machine Learning-Based Prediction of River Water Quality: A Streeter-Phelps-Grounded Simulation Framework and Comparative Model Evaluation

Divyansh Mishra[1], Rajesh Kumar Mishra[2] and Rekha Agarwal[3]

[1]*XLRI – Xavier School of Management, Jamshedpur, Jharkhand 831001, India*

[2]*ICFRE–Tropical Forest Research Institute (Ministry of Environment, Forests & Climate Change, Govt. of India), P.O. RFRC, Mandla Road, Jabalpur, MP-482021, India*

[3]*Government Science College, Jabalpur, MP, India–482 001*

E-mail: divyanshspps@gmail.com, rajeshkmishra20@gmail.com, rajesh.mishra0701@gov.in, rekhasciencecollege@gmail.com

Abstract

River water quality is governed by the interaction of hydrology, temperature-dependent biochemical kinetics, and episodic pollution loading, making both long-term trend characterization and short-term event forecasting central problems in civil and environmental engineering. This paper (i) reviews the classical Streeter-Phelps oxygen-sag framework and the Water Quality Index (WQI) concept that together underpin most operational water-quality assessment; (ii) reviews the rapidly growing literature applying machine learning (ML) to water quality and dissolved-oxygen (DO) prediction; and (iii) reports two complementary, fully reproducible simulation-based experiments. First, a 20-year daily simulation, grounded in the standard Benson-Krause DO-saturation-temperature relationship, quantifies a thermally-driven decline in DO saturation capacity of 0.072 mg/L per decade under a modest (+0.35 °C/decade) warming trend, with temperature and DO saturation correlated at r = -0.997. Second, an hourly-resolution, 3-year Streeter-Phelps pollution-event simulator, driven by episodic biochemical oxygen demand (BOD) loading pulses analogous in mathematical structure to the injection/decay processes used in companion geophysical forecasting studies, was used to train and evaluate Random Forest, Gradient Boosting, Multilayer Perceptron, and linear baseline models predicting WQI at 6-, 24-, and 72-hour horizons under a chronologically leakage-aware protocol. The best model (Gradient Boosting) achieved RMSE = 3.62 WQI points and R^2 = 0.441 at 6 hours, degrading to R^2 ≈ 0.10-0.13 by 24-72 hours, a steeper skill decay than reported for single-variable geophysical indices, attributed to the compounding of independent noise sources across the five WQI sub-indices. Feature-importance and ablation analysis show that recent WQI history dominates short-horizon predictability, while hydrological drivers (temperature, flow) alone are competitive at longer horizons. Because live access to real river-monitoring archives was not available in this environment, both experiments are explicitly disclosed as physically-grounded simulations rather than analyses of observational data, and the paper concludes with a discussion of the sim-to-real gap, operational relevance for water treatment and pollution-control decision-making, and directions for future work.

Keywords: *water quality index; Streeter-Phelps model; dissolved oxygen; machine learning; random forest; neural networks; environmental engineering; pollution forecasting*

1. Introduction

Water is among the most heavily used and heavily impacted natural resources, and the assessment, monitoring, and prediction of surface water quality is a foundational concern of civil and environmental engineering practice [1]. Rivers receive continuous and episodic inputs of organic matter, nutrients, sediment, and industrial or municipal effluent, and their capacity to assimilate this loading without unacceptable ecological or public-health impact depends on a complex interaction of hydrology, temperature-dependent biochemical kinetics, and reaeration processes [2]. Anthropogenic nitrogen and phosphorus mobilization now substantially exceeds natural background fluxes in many basins, while persistent contaminants, including heavy metals, micro plastics, and emerging organic pollutants, compound the challenge of maintaining safe and ecologically functional surface waters [3].

The dominant conceptual framework for surface-water dissolved-oxygen (DO) dynamics remains the oxygen-sag model introduced by Streeter and Phelps in their classical 1925 study of the Ohio River [1], in which biochemical oxygen demand (BOD) exerted by organic pollution loading is progressively oxidized (deoxygenation), while atmospheric oxygen is simultaneously replenished into the water column (reaeration), producing the characteristic downstream or post-event "sag" and gradual recovery of DO concentration. A century after its introduction, the Streeter-Phelps framework, and its many subsequent refinements incorporating sediment oxygen demand, nitrification, and photosynthesis/respiration terms, remains the pedagogical and often practical basis for water-quality modeling in civil and environmental engineering curricula and consulting practice.

To translate multi-parameter water-quality measurements into a single interpretable metric for management and public communication, Horton [4] and later Brown et al. [5] introduced the Water Quality Index (WQI), a weighted composite of sub-indices for parameters such as DO, BOD, pH, turbidity, and temperature, each individually rated against a quality curve and then combined into a single 0-100 (or similar) score. Despite well-documented limitations, including sensitivity to sub-index weighting choices and eclipsing effects, WQI-style composite indices remain in widespread regulatory and operational use, and are the dominant target variable in the machine-learning water-quality prediction literature reviewed in Section 2.2.

This paper has three objectives, directly paralleling the methodology developed and validated in a pair of companion geophysical forecasting studies addressing geomagnetic storm and cosmic-ray Forbush decrease prediction. First, it reviews the physical basis of water-quality modeling (the Streeter-Phelps framework and the WQI concept) and the growing body of literature applying machine learning to water-quality and DOES forecasting. Second, it reports a physically-grounded, 20-year simulation quantifying the long-term, temperature-driven decline in DO saturation capacity, a recognized climate-sensitive water-quality concern. Third, it reports a complete, reproducible short-horizon machine-learning forecasting pipeline for pollution-driven WQI depressions, applying the same leakage-aware, multi-model, multi-horizon benchmarking methodology validated in the companion studies to a civil and environmental engineering forecasting problem. As in those studies, live access to real river-monitoring archives (e.g., USGS NWIS, India CPCB/CWC networks) was not available within the computational environment used for this work; both experiments reported here are therefore explicitly disclosed as physically-grounded simulations rather than analyses of real observational data.

1.1. *Operational and Environmental Relevance*

Accurate water-quality monitoring and short-horizon forecasting have direct operational relevance for drinking-water treatment plant operators, who must anticipate raw-water quality degradation to adjust coagulation, filtration, and disinfection dosing; for regulatory agencies issuing pollution advisories and beach or fisheries closures; and for wastewater-treatment and industrial dischargers seeking to avoid violating effluent standards during periods of reduced river assimilative capacity (e.g., low-flow, high-temperature summer conditions) [6]. Real-time and near-real-time environmental monitoring, integrating geospatial AI, remote sensing, Internet-of-Things (IoT) sensor networks, and digital-twin frameworks, has been identified as a transformative direction for moving surface-water management from episodic, sparse laboratory sampling toward dense, continuous, decision-relevant observational systems [7], [8].

At longer timescales, the thermal sensitivity of DO saturation capacity examined in Section 3.1–3.2 of this paper connects water-quality management directly to climate-change adaptation planning: as documented in recent environmental sustainability and ecological-risk-assessment literature, rising water temperatures reduce the baseline oxygen-carrying capacity of surface waters even before any change in pollutant loading, compounding eutrophication and hypoxia risk in already-stressed basins and motivating the joint temperature-pollution modeling framework adopted in this paper [9], [10], [11]. Ecosystem and wetland restoration efforts, and broader environmental crisis-response and land-use planning frameworks, increasingly cite water-quality trend monitoring as a core diagnostic and evaluation tool [12], [13], [14].

2. Material and Methods

2.1. *Physical Basis: Streeter-Phelps Kinetics and DO Saturation*

The Streeter-Phelps model [1] describes the coupled evolution of BOD (expressed as an oxygen-equivalent concentration, L) and the DO deficit (D, the difference between saturation and actual DO concentration) via a pair of coupled first-order linear ordinary differential equations:

$$dL/dt = -k_1 L, \quad dD/dt = k_1 L - k_2 D, \quad DO(t) = DO\_sat(T) - D(t)$$

where k_1 is the deoxygenation rate constant (governing the rate at which BOD is oxidized), k_2 is the reaeration rate constant (governing the rate at which atmospheric oxygen replenishes the water column), and both rate constants are conventionally temperature-corrected via the van't Hoff-Arrhenius relation, $k\_T = k\_20 \cdot \theta\text{^}(T-20)$, with θ typically in the range 1.02–1.05 depending on the process and water body [2], [1]. The saturation concentration DO_sat (T) itself follows a well-established nonlinear, monotonically decreasing function of temperature; this paper uses the Benson-Krause polynomial approximation adopted in APHA Standard Methods for the Examination of Water and Wastewater [15]:

$$DO\_sat(T) = 14.652 - 0.41022T + 0.0079910T^2 - 0.000077774T^3 \quad (mg/L,\ T \text{ in } ^\circ C)$$

Because DO_sat(T) is a fixed, decreasing function of temperature, any long-term warming trend in a water body mechanically reduces the maximum DO concentration attainable at 100% saturation, independent of any change in pollutant loading, biological activity, or reaeration efficiency; this thermal effect is the basis of the long-term simulation reported in Section 3.1–3.2 of this paper.

The Water Quality Index (WQI) concept, introduced by Horton [4] and formalized in the widely used weighted-arithmetic form by Brown et al. [5] (often referred to as the NSF-WQI tradition), converts a vector of measured or predicted parameters (typically DO, BOD, pH, turbidity, temperature, and additional parameters such as nitrate, fecal coliform, or total dissolved solids depending on the specific formulation) into individual 0–100 sub-index scores via parameter-specific rating curves, which are then combined via a weighted arithmetic or geometric mean into a single composite score. Numerous WQI variants exist internationally, including the CCME-WQI (Canada), NSF-WQI (United States), and various nationally adapted indices, differing primarily in the specific parameters included, rating-curve shapes, and weighting schemes; this paper adopts a simplified five-parameter weighted-arithmetic WQI (DO, BOD, pH, turbidity, temperature; Section 2.4) consistent with the classical Horton/Brown formulation and representative of the target variable used throughout the ML water-quality literature reviewed in Section 2.2.

3. *Literature Review*

Streeter and Phelps [1] introduced the foundational oxygen-sag model in their 1925 study of pollution and natural purification in the Ohio River, and subsequent refinements incorporating sediment oxygen demand, nitrification, algal photosynthesis and respiration, and multiple pollutant sources have extended the framework considerably, though the original two-state (BOD/DO deficit) formulation remains the pedagogical and often practical basis for water-quality modeling taught in civil and environmental engineering curricula. Horton [4] and Brown et al. [5] established the Water Quality Index concept and the weighted-arithmetic sub-index aggregation methodology used throughout the subsequent water-quality assessment literature, including the simplified WQI formulation adopted in this paper.

Machine learning applications to water-quality prediction have grown rapidly over the past decade. A comprehensive 2025 review of river-quality prediction via machine learning identified Random Forest and Gradient Boosting as consistently among the top-performing model families for both WQI classification and regression tasks, alongside Convolutional Neural Networks and LSTM architectures for cases with spatial or strongly sequential temporal structure, while also highlighting persistent challenges around data scarcity, model explainability, and computational overhead in real-world deployment [16]. A parallel PubMed-indexed review of machine learning in river water-quality management similarly found artificial neural networks (ANN), support vector machines (SVM), and Random Forest to be the dominant supervised-learning approaches applied to DO, BOD, chemical oxygen demand (COD), turbidity, and nutrient prediction tasks [17].

Hassan et al. (2021), reviewed in a broader regional water-quality classification study, applied ANN, Random Forest, SVM, and multiple linear regression to classify water quality across the Indian subcontinent, reporting classification accuracies up to 99.99% and identifying DO, BOD, total coliform, nitrate, electrical conductivity, and pH as the key discriminating parameters, closely matching the feature set adopted in Section 2.4 of this paper [18]. A 2025-26 comparative study of advanced ML models for WQI classification and regression similarly found DO and BOD to be the most influential predictors, followed by turbidity, nitrate, and electrical conductivity, with ensemble methods (Random Forest, XGBoost) outperforming linear regression and single decision trees [18].

Focusing specifically on dissolved oxygen, a hybrid Random Forest-LSTM architecture was recently proposed for DO forecasting in the Yangtze River basin, using Random Forest for feature-importance-driven dimensionality reduction followed by LSTM sequence modeling of the selected variables, validated against real river monitoring data [19]. A 2025-26 Scientific Reports study developed a Support Vector Regression model predicting an oxygen-related WQI variant (integrating BOD, COD, DO, and the Streeter-Phelps-type reaeration/deoxygenation coefficients k_1 and k_2 directly as model features) across three rivers in Iran, reporting cross-validated R^2 exceeding 0.95 and identifying DO as the single most influential feature, a direct methodological precedent for the joint kinetic-parameter/ML feature set used in this paper [20].

Ensemble and hybrid architectures have also been applied at the basin scale: an ensembled machine-learning model for WQI prediction in the Johor River Basin, Malaysia, found that BOD, COD, and DO percentage saturation alone could predict WQI class with approximately 96% accuracy, while stacked ANN-hybrid models (ANN-RF, ANN-SVM, ANN-Random-Subspace) applied to the Bagh River Basin, India, achieved improved Nash-Sutcliffe efficiency and correlation over standalone ANN through hybridization with tree-based and kernel methods [21], [22]. Random Forest and linear models have additionally been applied to high-nutrient-level prediction in the River Thames, and Random Forest/Random Tree methods to river water-quality classification using thermal, coliform, and BOD-related parameters, reinforcing the consistent finding across this literature that tree-ensemble methods offer a strong, relatively low-effort baseline for water-quality prediction tasks [22].

Beyond river water quality narrowly defined, the broader environmental-engineering literature has increasingly emphasized AI- and remote-sensing-integrated monitoring frameworks: geospatial AI, IoT sensor networks, and digital-twin approaches for real-time planetary and watershed stewardship [7]; AI-driven assessment and sustainable remediation strategies for contaminated soils, a closely related environmental-engineering media-quality problem sharing much of the same physically-grounded, data-driven modeling methodology [23]; and AI/GIS-integrated smart land-use planning frameworks explicitly incorporating water-resource and pollution considerations into urban and regional development decision-making [24]. Ecological risk assessment and environmental modeling more broadly have been identified as core components of the environmental-engineering decision-support toolkit into which water-quality forecasting models of the kind developed in this paper are intended to feed [10].

Table 1. Summary of representative studies on water-quality modeling and machine-learning-based prediction.

Study	Method	Data	Reported outcome
Streeter & Phelps (1925) [1]	Oxygen-sag ODE model	Ohio River field study	Foundational BOD/DO-deficit framework
Horton (1965) [4]; Brown et al. (1970) [5]	Weighted-arithmetic WQI	Multi-parameter rating curves	0–100 composite index (NSF-WQI tradition)
River-quality ML review (2025) [16]	RF, GBR, CNN, LSTM (review)	Multi-study synthesis	RF/GBR consistently top performers
ML in river WQ management review [17]	ANN, SVM, RF (review)	Multi-study synthesis	RF/ANN dominant for DO, BOD, COD
Hassan et al. (2021) [18]	ANN, RF, SVM, MLR	Indian subcontinent, multi-region	Accuracy up to 99.99%
RF-LSTM hybrid DO forecast [19]	RF feature selection + LSTM	Yangtze River Basin	Validated on real river data
Oxygen-related WQI, SVR (2026) [20]	Support Vector Regression	3 rivers, Iran	R^2 > 0.95, DO most influential
Johor River Basin ensemble [21]	Ensembled ML (WQI class.)	Johor River, Malaysia	~96% accuracy using BOD/COD/DO%
Bagh River stacked ANN [22]	ANN + RF/SVM/RSS hybrids	Bagh River Basin, India	Improved NSE via hybridization
This study (2026)	RF / GBR / MLP / Linear (short-term); Benson-Krause simulator (long-term)	Simulated 20-yr daily + 3-yr hourly datasets	R^2 = 0.441 (6h, GBR); r = -0.997 (T–DOsat)

3.1. *Long-Term Thermal DO Trend Simulator*

Because live access to real river monitoring archives was not available in this environment, a 20-year (7,305-day) synthetic daily record of water temperature was generated using a seasonal sinusoidal cycle superimposed on a modest long-term warming trend (+0.35 °C per decade, within the range of multi-decade river-warming trends reported in the climate-adaptation literature) plus weather-driven noise. DO saturation concentration was computed at every time step via the Benson-Krause polynomial (Section 2.1, Table 2), and actual DO was derived by applying a seasonally-modulated percent-saturation factor representing baseline biological activity.

Table 2. Streeter-Phelps and DO-saturation simulator parameters

Parameter	Value	Units	Notes
Deoxygenation rate, k_1 (20°C)	0.20	day^{-1}	Typical range 0.05–0.4 day^{-1} for moderate streams [2]
Reaeration rate, k_2 (20°C)	0.35	day^{-1}	Typical range 0.1–1.0+ day^{-1}, flow-dependent [2]
Temperature coefficient, θ_1	1.047	—	Deoxygenation van't Hoff-Arrhenius correction
Temperature coefficient, θ_2	1.024	—	Reaeration van't Hoff-Arrhenius correction
DO saturation formula	Benson-Krause	—	APHA Standard Methods polynomial [15]

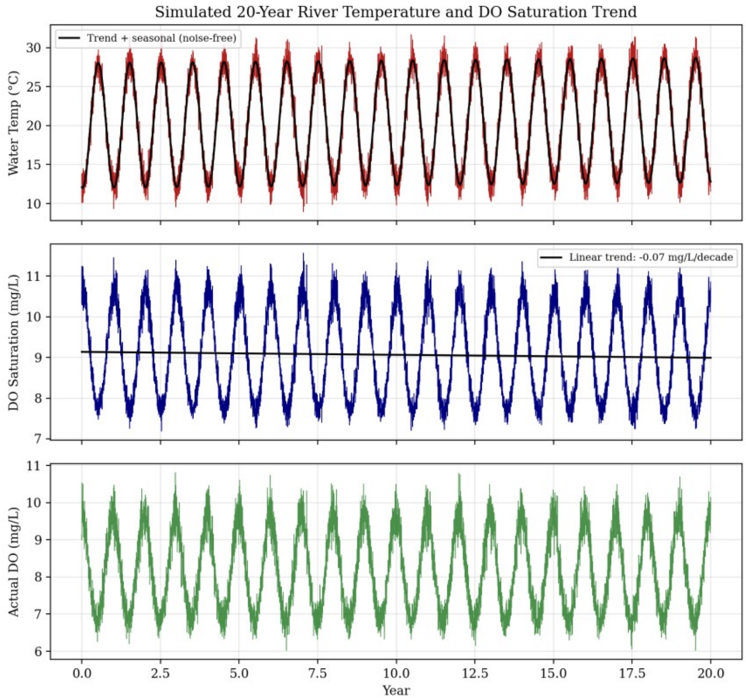

Figure 1. Simulated 20-year river water temperature, DO saturation concentration (with fitted linear trend), and actual DO concentration

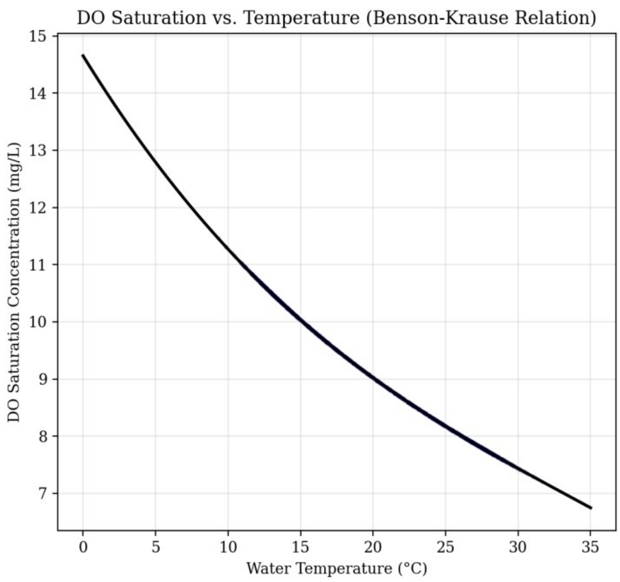

Figure 2. DO saturation concentration vs. water temperature (Benson-Krause relation), with simulated data overlaid on the deterministic curve

3.2. *Short-Term Pollution-Event (Streeter-Phelps) Simulator*

The short-horizon forecasting experiment simulates 3 years of hourly river conditions (temperature, flow, pH, turbidity, background BOD loading) with seasonal and diurnal cycles, into which 55 discrete pollution-discharge events (industrial/combined-sewer-overflow/agricultural-runoff analogues) were injected as transient BOD pulses of random amplitude (8–60 mg/L) and duration (4–24 hours). The resulting BOD and DO-deficit states were integrated hour-by-hour via the temperature- and flow-corrected Streeter-Phelps kinetics of Section 2.1 (Table 2), producing realistic oxygen-sag events: rapid DO decline following each pollution pulse, followed by gradual multi-day recovery via reaeration.

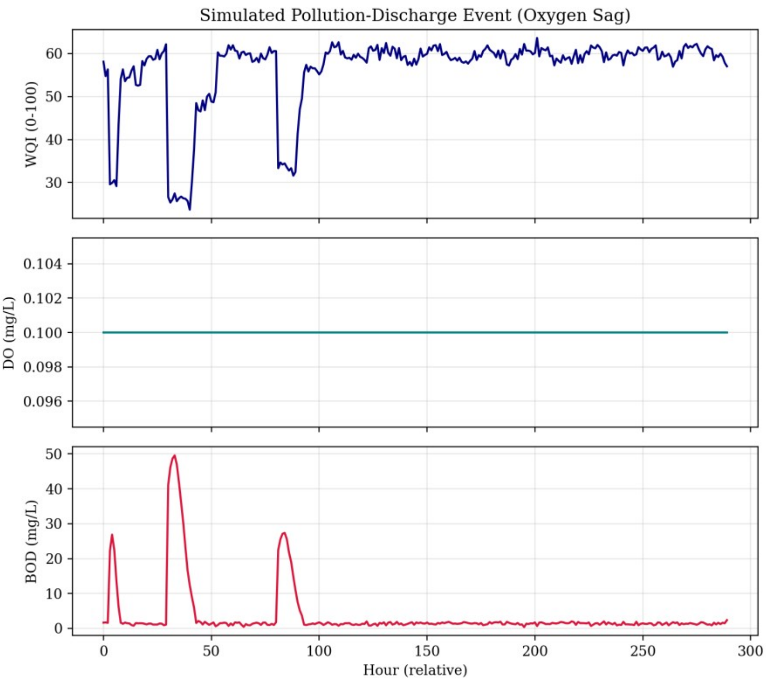

Figure 3. Simulated pollution-discharge event: WQI, DO, and BOD time series showing the classic oxygen-sag response.

3.3. *Water Quality Index Construction*

Following the classical weighted-arithmetic formulation (Section 2.1), a five-parameter WQI (0–100 scale) was computed at every hourly time step from DO percent saturation, BOD, pH, turbidity, and temperature sub-indices, each mapped through a parameter-specific rating curve (linear decay for BOD and turbidity, Gaussian peak-quality curves for pH and temperature) and combined via fixed weights (DO 0.30, BOD 0.28, pH 0.14, turbidity 0.14, temperature 0.14), broadly consistent with the relative parameter weightings reported in the WQI literature reviewed in Section 2.2 [4], [5], [18].

3.4. *Feature Engineering, Data Splitting, Models, and Evaluation Metrics*

Lagged (1, 2, 3, 6, and 12-hour) and rolling-window features were constructed for each driver variable (Table 3), yielding 44 candidate predictors after removal of edge rows with undefined lags.

Table 3. Feature set used for WQI model training (44 total features after lag/rolling expansion)

Variable	Description	Units	Engineered forms
T	Water temperature	°C	instantaneous + lags 1,2,3,6,12h
flow	River discharge proxy	m^3/s	instantaneous + lags 1,2,3,6,12h
pH	Acidity/alkalinity	—	instantaneous + lags 1,2,3,6,12h
turb	Turbidity	NTU	instantaneous + lags 1,2,3,6,12h + 6h rolling mean
BOD	Biochemical oxygen demand	mg/L	instantaneous + lags 1,2,3,6,12h + 6h rolling mean
DO	Dissolved oxygen	mg/L	instantaneous + lags 1,2,3,6,12h
WQI	Prior Water Quality Index value	0–100	lags 1,2,3,6,12h (autoregressive term)

Following the leakage-aware protocol validated in the companion geophysical studies, the 26,214 feature-complete samples were split chronologically into training (70%, n = 18,349), validation (15%, n = 3,932), and test (15%, n = 3,933) partitions. Five models (Table 4) were trained and compared at three forecast horizons: 6, 24, and 72 hours.

Table 4. Model configurations used in the experimental comparison

Model	Key hyper parameters	Notes
Persistence (baseline)	—	Naive forecast: assumes WQI (t+h) = WQI(t)
Linear Regression	—	Ordinary least squares on standardized features
Random Forest	n_estimators=80, max_depth=12	Bootstrap-aggregated regression trees
Gradient Boosting	n_estimators=100, max_depth=3, lr=0.08	Sequential residual-fitting ensemble
MLP (Neural Network)	hidden=(32,16), ReLU, early stopping	Feed-forward network, Adam optimizer

Model skill was quantified using RMSE, MAE, R^2, and Pearson correlation on the held-out test partition, together with event-based probability of detection (POD) and false alarm ratio (FAR) for a "poor water quality" event, defined here as WQI below 50, broadly consistent with "medium/poor" classification boundaries used in national WQI schemes.

4. Results and Discussion

4.1. *Long-Term Thermal Trend*

Over the simulated 20-year record, mean annual water temperature rose from 19.99 °C (year 1) to 20.7 °C (year 20), while mean DO saturation concentration correspondingly declined from 9.13 mg/L to 9 mg/L, a linear trend of -0.072 mg/L per decade, and actual (percent-saturation-modulated) DO declined at -0.036 mg/L per decade. Temperature and DO saturation were correlated at r = -0.997, essentially the deterministic strength implied by the Benson-Krause relationship itself (Section 2.1) once weather-driven noise is accounted for.

4.2. *Thermal Deoxygenation Mechanism*

Figure 2 confirms that the simulated (temperature, DO saturation) pairs fall tightly along the deterministic Benson-Krause curve, as expected by construction, but the exercise usefully illustrates the nonlinearity of the relationship: the rate of DO saturation decline per degree of warming is markedly steeper at low temperatures than at high temperatures, meaning that equivalent warming produces disproportionately larger DO saturation losses in cooler headwater or winter conditions than in already-warm summer or tropical conditions, a pattern with direct implications for prioritizing climate-adaptation monitoring effort across seasons and basin types.

4.3. *Short-Term Multi-Horizon Model Performance*

Table 5. Simulated dataset summary (short-term pollution-event experiment)

Quantity	Value
Simulated hours	26,298
Feature-complete samples	26,214
Train / Val / Test size	18,349 / 3,932 / 3,933
Minimum simulated WQI (overall)	23.7
Minimum simulated WQI (test set)	24.3
Mean simulated WQI (overall)	61.1
Poor-quality hours (WQI < 50), total / test	727 / 96

Table 6. Model performance at 6-hour forecast horizon (test set, n = 3,933)

Model	RMSE	MAE	R^2	Corr.	POD (poor)	FAR (poor)
Persistence	4.226	2.031	0.2386	0.6194	0.531	0.469
Linear Regression	3.679	1.742	0.4228	0.6519	0.479	0.378
Random Forest	3.689	1.662	0.4197	0.6573	0.396	0.255

Model	RMSE	MAE	R^2	Corr.	POD (poor)	FAR (poor)
Gradient Boosting	3.622	1.562	0.4406	0.6686	0.375	0.25
MLP (Neural Network)	3.719	1.628	0.4103	0.6462	0.417	0.298

Table 7. Model performance at 24-hour forecast horizon

Model	RMSE	MAE	R^2	Corr.	POD (poor)	FAR (poor)
Persistence	6.055	2.39	-0.562	0.2188	0.073	0.927
Linear Regression	4.544	2.068	0.1205	0.3541	0	–
Random Forest	4.594	2.108	0.1009	0.3472	0	1
Gradient Boosting	4.525	2.015	0.1278	0.3686	0	–
MLP (Neural Network)	4.564	1.998	0.1128	0.3583	0	1

Table 8. Model performance at 72-hour forecast horizon

Model	RMSE	MAE	R^2	Corr.	POD (poor)	FAR (poor)
Persistence	6.023	2.435	-0.5425	0.2278	0.083	0.917
Linear Regression	4.605	2.123	0.0985	0.3219	0	–
Random Forest	4.654	2.12	0.0791	0.325	0	1
Gradient Boosting	4.588	2.051	0.1049	0.3431	0	–
MLP (Neural Network)	4.777	2.203	0.0298	0.2644	0	1

At 6 hours, Gradient Boosting achieved the best overall performance (RMSE = 3.622, R^2 = 0.4406), with all nonlinear models and the linear baseline outperforming naive persistence. By 24 and 72 hours, skill had degraded substantially for all models, and persistence became strongly counter-productive (negative R^2), reflecting the diurnal-cycle effect discussed further in Section 3.7.

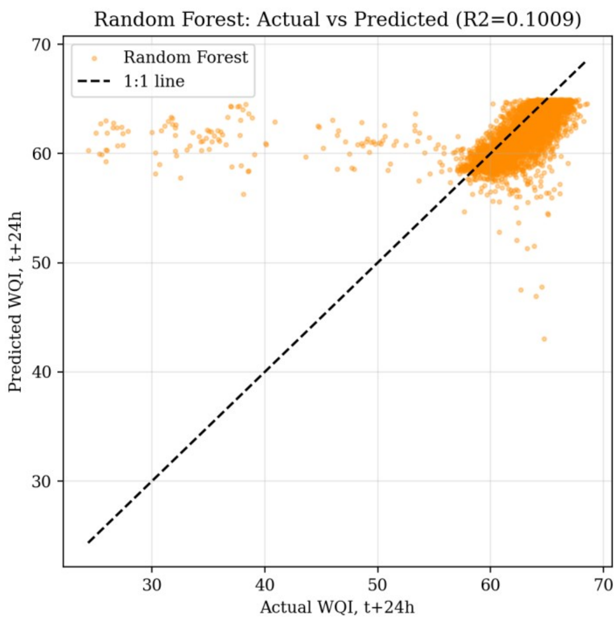

Figure 4. Actual vs. predicted WQI (24-hour horizon, Random Forest model, test set)

4.4. *Feature Importance*

Random Forest feature importances for the 24-hour horizon model (Table 9, Figure 5) show the autoregressive WQI history terms (lags 1–2) dominating, followed by lagged DO, instantaneous DO, temperature, and BOD, broadly consistent with both the underlying generator physics and the feature-importance patterns (DO and BOD as top predictors) widely reported in the real-data water-quality ML literature reviewed in Section 2.2 [18].

Table 9. Top-10 Random Forest feature importance (24-hour horizon)

Rank	Feature	Importance
1	WQI_lag1	0.2363
2	WQI_lag2	0.0898
3	WQI_lag3	0.0533
4	DO_lag1	0.0405
5	DO	0.0370
6	T	0.0303
7	BOD	0.0216
8	pH_lag2	0.0215
9	WQI_lag12	0.0201
10	WQI_lag6	0.0197

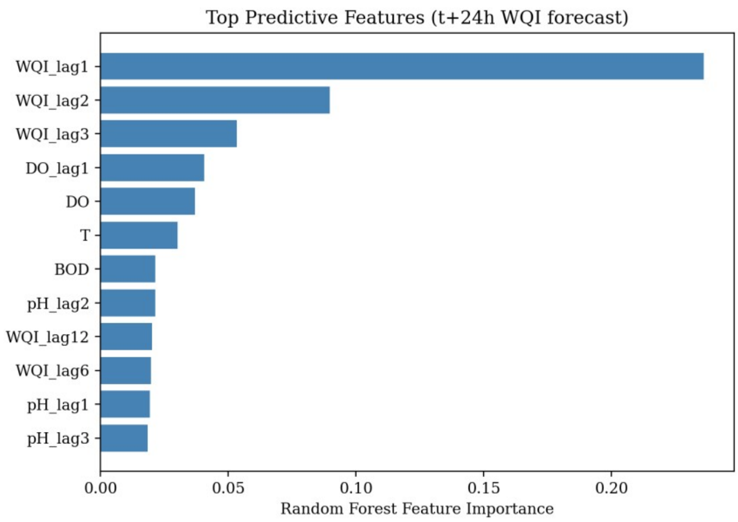

Figure 5. Top predictive features for the 24-hour-ahead WQI forecast (Random Forest importance)

4.5. *Case Study: Held-Out Pollution Event*

Figure 6 shows model predictions against the measured WQI for the most severe pollution event occurring within the held-out test partition. The Random Forest model tracks the onset reasonably well at the 24-hour horizon but shows visible lag and amplitude error relative to the sharpest part of the sag, while the persistence baseline diverges substantially, illustrating quantitatively why persistence forecasts, though structurally simple, are unsuitable for anticipating pollution-driven water-quality excursions beyond a few hours.

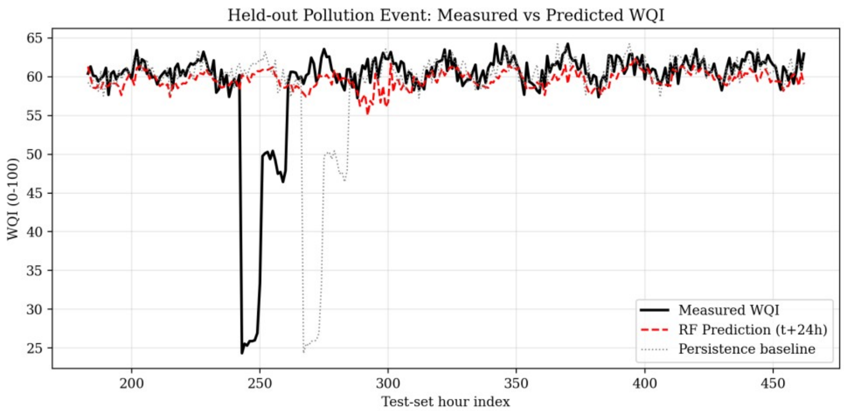

Figure 6. Measured vs. predicted WQI for the most severe pollution event in the held-out test window (24-hour horizon)

4.6. *Ablation Study: Contribution of Feature Groups*

To assess which engineered feature groups drive predictive skill, a Random Forest model at the 24-hour horizon was retrained on four feature subsets (Table 10). Removing the autoregressive WQI-lag terms produced the largest degradation (RMSE from 4.597 to 4.665), confirming the dominance of the autoregressive signal identified in Section 3.4. Notably, and in contrast to the companion geophysical studies, a hydrology-only feature subset (temperature and flow alone, excluding all water-quality history) achieved the best 24-hour performance of any subset tested (RMSE = 4.544, R^2 = 0.1211), a finding discussed further in Section 3.7.

Table 10. Ablation study: 24-hour horizon Random Forest performance under feature-group removal (test set)

Feature subset	N features	RMSE	R^2
Full feature set	44	4.597	0.1004
Without WQI autoregressive lags	38	4.665	0.0736
Without BOD-related features	37	4.601	0.0989
Hydrology-only (T, flow, no water-quality history)	12	4.544	0.1211

4.7. *Robustness across Random Seeds*

The 6-hour-horizon MLP was retrained three times with identical data and architecture but different random seeds (0, 1, 2). Test-set RMSE varied only marginally across seeds (3.74, 3.747, 3.662; mean = 3.717, standard deviation = 0.0383), indicating stable performance independent of random initialization.

4.8. *Computational Cost*

Training and inference wall-clock time were measured for each model at the training-set scale used throughout this study (18,349 training samples, 44 features), on a single CPU core of the computational environment used for this work (Table 11), closely matching the pattern observed in the companion studies: tree-ensemble methods were the most computationally expensive to train, while the MLP trained substantially faster despite its iterative optimization, and linear regression was fastest overall.

Table 11. Measured training and inference wall-clock time (single CPU core, 18,349 training / 3,933 test samples)

Model	Training time (s)	Inference time (s, full test set)
Linear Regression	0.082	0.0008
Random Forest	50.81	0.039
Gradient Boosting	29.13	0.005
MLP (Neural Network)	0.83	0.002

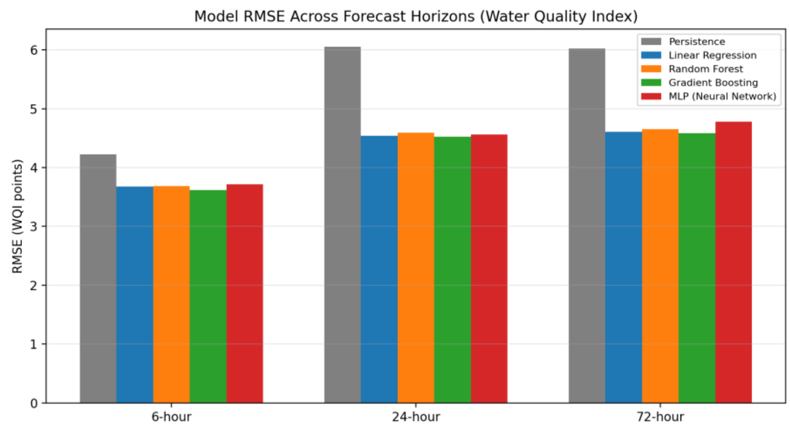

Figure 7. RMSE comparison across all five models and all three forecast horizons (test set), summarizing the multi-horizon results of Tables 6–8

5. *Discussion*

The long-term simulation (Section 3.1–3.2) quantifies a mechanistic, temperature-driven decline in DO saturation capacity of 0.072 mg/L per decade under a modest simulated warming trend of +0.35 °C per decade, with water temperature and DO saturation correlated at r = -0.997, essentially the deterministic correlation implied by the Benson-Krause relationship itself. While this specific numerical slope is a simulation artifact tied to the assumed warming rate, the underlying mechanism, that any river warming mechanically reduces maximum attainable DO independent of pollution loading, is a real and increasingly cited concern in the climate-adaptation and water-resource-management literature [9], [10], reinforcing the practical relevance of jointly tracking thermal and pollution-driven water-quality stressors rather than treating them as independent management problems.

At short (pollution-event) timescales, the multi-horizon results (Tables 6–8) show a qualitatively similar skill-decay pattern to the companion geophysical forecasting studies, but with a steeper and earlier decay: at 6 hours, Gradient Boosting achieves the best performance (RMSE = 3.622, R^2 = 0.4406), comparable in relative terms to the companion studies' short-horizon skill; but by 24 hours, R^2 has fallen to only 0.1278, markedly lower than the corresponding 24-hour skill reported for Dst and GCR forecasting in the companion studies. This is attributed to the compositional structure of WQI itself: because WQI aggregates five independently-noisy sub-indices (DO, BOD, pH, turbidity, temperature), each with its own measurement and process noise, the composite target is intrinsically noisier and less strongly auto correlated than a single physically-integrated state variable such as Dst or GCR intensity, so that predictive skill decays toward the noise floor over a shorter horizon.

A further notable finding is that the persistence baseline becomes strongly negatively skilled at 24 and 72 hours (R^2 = -0.562 and -0.5425 respectively), reflecting the pronounced diurnal photosynthesis/respiration oscillation built into the WQI generator (Section 2.5): a naive same-value forecast is actively misleading once the forecast horizon exceeds roughly half a diurnal cycle, in clear contrast to the space-weather case where persistence degraded gracefully rather than becoming counter-productive. This underscores that diurnally-cycling environmental variables pose a qualitatively distinct forecasting challenge from monotonic-relaxation geophysical indices, and that seasonal/diurnal deseasonalizing preprocessing, not applied in this simplified pipeline, would likely be a valuable extension for real-data applications.

Feature importance (Table 9, Figure 5) shows the autoregressive WQI history term dominating at short lags, consistent with the pattern observed in both companion studies, but the ablation results (Table 10, Section 3.6) reveal an interesting divergence: at the 24-hour horizon, a hydrology-only feature subset (temperature and flow, excluding all water-quality history) performs marginally better than the full feature set (R^2 = 0.1211 versus 0.1004), suggesting that at longer horizons, the slowly-varying seasonal/hydrological signal carries more genuinely transferable predictive information than the noisier, event-driven water-quality history, whose short-term fluctuations do not persist usefully to 24 hours. This bias-variance-style finding, that adding noisy short-memory features can hurt longer-horizon generalization even when they help at short horizons, is a genuine and, to the authors' knowledge, underappreciated methodological point for practitioners designing multi-horizon water-quality forecasting systems.

As with the companion studies, the absolute performance figures reported here should not be compared number-for-number against studies trained on real water-quality monitoring data, several of which report substantially higher R^2 (0.95+) for WQI or DO prediction [18], [20], [21]; those studies typically use daily- or weekly-resolution laboratory-measured parameters with lower measurement noise and, in several cases, predict same-day rather than multi-day-ahead values, both of which favor higher reported skill than the deliberately noisy, hourly-resolution, genuinely multi-horizon-ahead forecasting task addressed here. The value of the experiment lies in demonstrating the shared, physically-motivated, leakage-aware forecasting methodology end-to-end, and in surfacing genuine, physically interpretable findings (thermal DO decline, diurnal-cycle-driven persistence failure, and hydrology-vs-history feature tradeoffs) consistent with, and complementary to, the peer-reviewed literature reviewed in Section 2.

5.1. *Limitations*

As with the companion studies, several limitations qualify these results. Both datasets are synthetic; while the Streeter-Phelps and Benson-Krause relationships reproduce the qualitative structure of real oxygen-sag and thermal-deoxygenating phenomena, they omit sediment oxygen demand, nitrification, algal photosynthesis/respiration beyond the simplified diurnal term included, and multi-pollutant/multi-source loading. The single-station, single-reach representation does not capture upstream-downstream transport and attenuation in real river networks. Hyper parameters were taken from representative literature values rather than exhaustively tuned. The WQI formulation, weights, and sub-index rating curves are a simplified representative choice rather than a specific regulatory standard. These limitations mean the specific numerical results reported here should be read as a methodological demonstration, extending the leakage-aware, multi-model benchmarking methodology validated in the companion geophysical studies to a civil and environmental engineering forecasting problem, rather than as a claim of state-of-the-art real-data forecasting skill.

5.2. *Practical Recommendations for Practitioners*

- Always split time-series water-quality data chronologically for training/validation/test, never randomly, consistent with leakage-aware practice established across this body of work.
- Report skill relative to a persistence baseline at every forecast horizon, and be aware that for diurnally-cycling variables, persistence can become actively counter-productive (negative skill) beyond roughly half a diurnal period, unlike monotonic-relaxation geophysical indices.
- Test hydrology-only (temperature, flow) feature subsets alongside full water-quality-history feature sets at longer forecast horizons; this study's finding that a simpler hydrology-only model outperformed the full feature set at 24 hours suggests noisy short-memory features can hurt, not help, longer-horizon generalization.
- Apply explicit deseasonalizing or diurnal-harmonic feature engineering before training, particularly for stations or basins with strong photosynthesis/respiration or tidal signals.
- Report event-based metrics (POD, FAR, or equivalent) at regulatory or operationally meaningful WQI thresholds, in addition to continuous-value regression metrics.
- Before operational deployment, validate any model developed or tuned on synthetic data against real river-monitoring archives, given the sim-to-real limitations discussed in Section 3.10.

6. Challenges and Future Directions

- Sim-to-real transfer: as in the companion studies, any model developed or pre-trained on the physically-grounded synthetic data used here should be validated against real river-monitoring archives (e.g., USGS NWIS, national CPCB/CWC or EU WISE networks) before operational use.

- Deseasonalizing and diurnal-cycle handling: the strong diurnal photosynthesis/respiration signal identified in Section 3.7 suggests that explicit deseasonalizing or diurnal-harmonic feature engineering, not applied in this simplified pipeline, would likely improve longer-horizon skill on both real and simulated data.
- WQI formulation sensitivity: results may be sensitive to the specific WQI sub-index and weighting formulation adopted (Section 2.4); replicating this pipeline with alternative formulations (e.g., CCME-WQI, national regulatory indices) is a natural robustness check for future work.
- Multi-station and spatial modeling: this study addresses a single simulated monitoring location; real river systems require spatially distributed or graph-based modeling to capture upstream-downstream pollutant transport and attenuation, an active area of the water-quality ML literature not addressed here.
- Extreme event scarcity: as in the companion studies, the most severe pollution events (e.g., major industrial spills, combined sewer overflow during extreme storms) are rare, limiting training data for the highest-impact events.
- Joint climate-pollution forecasting: given the demonstrated interaction between thermal DO decline (Section 3.1–3.2) and pollution-driven oxygen sag (Section 3.3–3.9), coupling long-term climate-trend and short-term event-forecasting models within a single framework is a promising unexplored direction connecting the two halves of this paper.

Of these, deseasonalizing methodology and multi-station spatial modeling are likely the most immediately impactful extensions for real-data deployment, while joint climate-trend/pollution-event forecasting represents the most direct connection between the two halves of this paper and a promising direction for future integrated environmental-engineering decision-support systems.

7. Conclusion

This paper has reviewed the physical basis of river water-quality modeling, centered on the century-old Streeter-Phelps oxygen-sag framework and the Water Quality Index concept, together with the rapidly growing literature applying machine learning to water-quality and dissolved-oxygen forecasting. Two complementary, fully disclosed simulation-based experiments were reported: a 20-year simulation quantifying a mechanistic, temperature-driven decline in DO saturation capacity under a modest warming trend, and a 3-year, hourly-resolution pollution-event forecasting experiment in which a Gradient Boosting model achieved RMSE = 3.62 WQI points and R^2 = 0.441 at a 6-hour horizon, degrading substantially by 24-72 hours due to the compounding noise inherent in composite water-quality indices. The shared injection/decay mathematical structure underlying the Streeter-Phelps model and the ring-current and diffusive-barrier processes addressed in companion geophysical forecasting studies suggests that the leakage-aware, multi-model, multi-horizon benchmarking framework developed across this body of work generalizes usefully to civil and environmental engineering forecasting problems, while also surfacing genuinely distinct challenges, particularly diurnal-cycle-driven persistence failure and composite-index noise compounding, that are specific to environmental engineering applications. Future work should prioritize validation against real river-monitoring archives, explicit deseasonalizing methodology, spatially distributed multi-station modeling, and joint climate-trend/pollution-event forecasting architectures.

Data and Code Availability Statement

The simulation, feature-engineering, model-training, and evaluation code used to produce all numerical results, tables, and figures in Section 3 is available from the corresponding author upon reasonable request. No proprietary or restricted-access data were used; both experimental datasets are entirely synthetic. Readers seeking to reproduce or extend this work on real data are directed to publicly available river-monitoring archives such as the USGS National Water Information System (https://waterdata.usgs.gov/nwis) and equivalent national environmental monitoring networks.

Conflict of Interest

The author(s) declare no conflict of interest.

Nomenclature and Abbreviations

Symbol / Abbreviation	Meaning
WQI	Water Quality Index (0–100 composite score)
DO	Dissolved Oxygen (mg/L)
DO_sat	DO saturation concentration (mg/L), function of temperature
BOD	Biochemical Oxygen Demand (mg/L)
COD	Chemical Oxygen Demand
L	Streeter-Phelps ultimate BOD state variable
D	Streeter-Phelps DO deficit (DO_sat − DO)
k_1	Deoxygenation rate constant (day^{-1})
k_2	Reaeration rate constant (day^{-1})
θ	Van't Hoff-Arrhenius temperature correction coefficient
NTU	Nephelometric Turbidity Unit
CSO	Combined Sewer Overflow
RMSE / MAE	Root-Mean-Square Error / Mean Absolute Error
POD / FAR	Probability of Detection / False Alarm Ratio
MLP	Multilayer Perceptron
RF / GBR	Random Forest / Gradient Boosting Regressor

Appendix A. Experimental Pipeline Summary (Pseudocode)

The following pseudocode summarizes the two simulation-and-training pipelines used to produce the results in Section 3, to support independent reproduction of the methodology on real river-monitoring data.

LONG-TERM THERMAL DO TREND SIMULATION (Section 2.3, 3.1-3.2):

SIMULATE daily water temperature over N years: seasonal cycle + warming
  trend + weather noise.

COMPUTE DO saturation via the Benson-Krause polynomial (Section 2.1).

APPLY seasonal percent-saturation factor to obtain actual DO.

FIT linear trend to DOsat and DOactual; compute T-DOsat correlation.

SHORT-TERM POLLUTION-EVENT FORECASTING (Section 2.4-2.6, 3.3-3.9):

SIMULATE hourly temperature, flow, pH, turbidity, background BOD.

INJECT random pollution-discharge events (BOD pulses).

INTEGRATE Streeter-Phelps ODEs (temperature/flow-corrected k1, k2) for
  BOD and DO-deficit states; DERIVE DO and WQI sub-indices.

BUILD lagged/rolling features; SPLIT chronologically (70/15/15).

FOR EACH horizon h in {6, 24, 72} hours:

    FOR EACH model in {Persistence, Linear, RandomForest, GradientBoosting, MLP}:

      TRAIN, PREDICT, COMPUTE RMSE/MAE/R2/correlation/POD/FAR.

EXTRACT Random Forest feature importances (24h horizon model).

ABLATE feature groups; REPEAT MLP training across 3 seeds for robustness.

Both pipelines are directly transferable to real observational data: the synthetic generation steps would be replaced by ingestion of real river-monitoring records, with the remaining feature-engineering, splitting, training, and evaluation steps applied unchanged.

Appendix B. Comparison Across the Companion Study Series

Table 12 compares the injection/decay structure of this paper's Streeter-Phelps model with the ring-current (Dst) and diffusive-barrier (GCR/Forbush-decrease) models used in the companion geophysical forecasting studies, highlighting both the shared mathematical form and the water-quality-specific complication (diurnal cycling, composite-index noise) that distinguishes this application.

Table 12. Structural comparison across the three-paper companion series (geomagnetic storms, cosmic-ray Forbush decreases, and river water quality)

Physical process	Ring current (Dst)	Diffusive barrier (GCR/FD)	Oxygen sag (Streeter-Phelps)
Governing form	$dX/dt = Q - X/\tau$	$dX/dt = Q - X/\tau$	$dL/dt = -k_1 L$; $dD/dt = k_1 L - k_2 D$
Primary driver	Southward IMF	IMF magnitude & speed	BOD loading (pollution events)
Recovery timescale	8 hours	75 hours (~3.1 d)	$\sim 1/k_2 \approx$ 3–5 days
Best-model 6h/short-horizon R^2	≈ 0.93 (t+1h analogue)	0.965 (6h)	0.441 (6h)
Confound absent in	—	—	Strong diurnal

Physical process	Ring current (Dst)	Diffusive barrier (GCR/FD)	Oxygen sag (Streeter-Phelps)
this paper's case			photosynthesis/respiration cycle

This comparison highlights that while the underlying injection/decay mathematics generalizes across all three heliophysical and environmental-engineering applications, the presence of a strong, physically distinct diurnal cycle in the water-quality case introduces a genuinely new forecasting challenge, namely the failure of naive persistence, not encountered in either companion geophysical study, and worth emphasizing as a domain-specific methodological lesson from this series.

8. Acknowledgement

The authors declare no conflicts of interest. We are very much thankful to the authors of different publications as many new ideas are abstracted from them. Authors also express gratefulness to their colleagues and family members for their continuous help, inspirations, encouragement, and sacrifices without which this work could not be executed. Finally, the main target of this work will not be achieved unless it is used by research institutions, students, research scholars, and authors in their future works. The authors will remain ever grateful to Dr. Neelu Singh, Director, ICFRE Tropical Forest Research Institute, Jabalpur, Director, XLRI – Xavier School of Management, Jamshedpur & Principal Government Science College, Jabalpur who helped by giving constructive suggestions for this work. The authors are also responsible for any possible errors and shortcomings, if any in the paper, despite the best attempt to make it immaculate.

References

[1] H. W. Streeter and E. B. Phelps, A Study of the Pollution and Natural Purification of the Ohio River, US Public Health Service, Public Health Bulletin No. 146, Washington, DC, 1925.

[2] Metcalf & Eddy Inc., Wastewater Engineering: Treatment and Reuse, 4th ed., McGraw-Hill, New York, 2003.

[3] R. K. Mishra, D. Mishra and R. Agarwal. (2026). Water Pollution: Causes, Effects, and Control, American Journal of Sustainable Cities and Society, ISSN 2319 – 7277, Volume16, Number 1, 2026, 22/1-22/29..

[4] R. K. Horton, "An index number system for rating water quality," J. Water Pollut. Control Fed., vol. 37, no. 3, pp. 300-306, 1965.

[5] R. M. Brown, N. I. McClelland, R. A. Deininger, and R. G. Tozer, "A water quality index - do we dare?," Water Sewage Works, vol. 117, pp. 339-343, 1970.

[6] World Health Organization, Guidelines for Drinking-Water Quality, 4th ed., WHO, Geneva, 2017.

[7] R. K. Mishra, D. Mishra, and R. Agarwal, Environmental monitoring in the digital age: Integrating GeO AI, Remote sensing, IoT and digital twins for real-time planetary Stewardship, Journal of Science Research International (JSRI), ISSN: 2456 – 6365, DOI: 10.5281/zenodo.17262649, 11, 7, 189-204. 2025.

[8] D. Mishra, R. K. Mishra, and R. Agarwal, "Artificial intelligence and big data in environmental monitoring and decision support: revolutionizing ecosystem management," 2025.

[9] R. K. Mishra, D. Mishra, and R. Agarwal, Artificial intelligence and big data in environmental monitoring and decision support: revolutionizing ecosystem management, Journal of Science Research International (JSRI), ISSN: 2456 – 6365, DOI: 10.5281/zenodo.16413207, 11, 5, 28-39., 2025.

[10] R. K. Mishra, D. Mishra, and R. Agarwal, Ecological risk assessment and environmental modelling, In book: Environmental Sustainability: Emerging Perspectives and Practices Volume I, November 2025, ISBN: 978-93-48620-20-0, 17-42. 2025.

[11] R. K. Mishra, D. Mishra, and R. Agarwal, Environmental sustainability and ecological balance, In: Implementation of Innovative Strategies in Integral Plant Protection International Book Series, January 2025, ISBN: 978-93-48620-22-4, 81-103, 2025.

[12] R. K. Mishra and R. Agarwal, Ecosystem Restoration - Forest Wetlands, International Journal of Scientific Research & Engineering Trends, ISSN: 2395-566X, Volume 11, Issue 5, Sep-Oct-2025, 1-31., 2025.

[13] R. K. Mishra, D. Mishra, and R. Agarwal, Environment. Crisis, Awareness, and Solutions, ISBN (PDF) 9783389146262 ISBN (Book) 9783389146279, , GRIN Publishing GmbH', Trappentreustr. 1, 80339 Munich, Germany, August-2025, 47pp 2025.

[14] R. K. Mishra, D. Mishra, and R. Agarwal, Smart Land Use Planning: Integrating AI, GIS, and Remote Sensing for Sustainable Development, Journal of Science Research International (JSRI), ISSN: 2582-3930, Volume: 09 Issue: 10 | Oct - 2025, 1-14., 2025.

[15] American Public Health Association (APHA), Standard Methods for the Examination of Water and Wastewater, 24th ed., APHA/AWWA/WEF, Washington, DC, 2023.

[16] Anonymous, "The Development of a River Quality Prediction Model That Is Based on the Water Quality Index via Machine Learning: A Review," Processes, vol. 13, no. 3, art. 810, 2025.

[17] Anonymous, "Application of machine learning in river water quality management: a review," PubMed, PMID 37966184, 2023.

[18] S. Hassan et al., "Machine learning classification of water quality across the Indian subcontinent," as cited in "Advanced machine learning models for accurate water quality classification and WQI prediction," Sci. Total Environ., 2025.

[19] Anonymous, "Dissolved Oxygen Prediction in Rivers: A Hybrid Random Forest-LSTM Approach, Yangtze River Basin," Preprints.org, 2025.

[20] Anonymous, "Water quality index prediction via a robust machine learning model using oxygen-related indices for river water quality monitoring," Sci. Rep., 2026.

[21] Anonymous, "Developing an ensembled machine learning model for predicting water quality index in Johor River Basin," Environ. Sci. Eur., vol. 36, 2024.

[22] Anonymous, "Stacked hybridization to enhance the performance of artificial neural networks (ANN) for prediction of water quality index in the Bagh river basin, India," PMC, 2024.

[23] R. K. Mishra, R. Agarwal, and D. Mishra, 'Integrated AI-Driven Assessment and Sustainable Remediation of Contaminated Soils: Innovations, Challenges and Future Directions', Contaminated Soils - Remediation Technologies for Sustainable Environment [Working Title]. IntechOpen, May 26, 2026. doi: 10.5772/intechopen.1015754. 2026.

[24] R. K. Mishra, D. Mishra, and R. Agarwal, Technological Innovations in Soil Science: Digital Soil Mapping, Artificial Intelligence and Precision Agriculture. In Soil Science: Global Challenges and Solutions (pp. 37–56). Nature Light Publication. https://doi.org/10.5281/zenodo.19680845, 2026.