Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hydrological modulation of biochemical oxygen demand in sanitation-limited subtropical rivers: a case study from the Caspian Sea basin

Reza Dahmardeh Behrooz

dahmardeh_behrooz@yahoo.com

Zabol University

Sourav Ghosh

Bengaluru City University

Article

Keywords: Biochemical oxygen demand (BOD_5), Fecal coliform, River pollution, Hypoxia, Machine learning, Subtropical rivers

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11468331/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Subtropical rivers with insufficient wastewater infrastructure may experience episodic hypoxia and downstream ecological stress due to monsoon-driven organic loading. The effects of fecal contamination on oxygen consumption under different hydrological situations, however, are little understood. In order to identify hypoxia drivers and forecast biochemical oxygen demand (BOD5) in three subtropical rivers that flow into Iran's southern Caspian Sea basin, we have developed an interpretable machine learning framework that integrates Random Forest modeling with SHapley Additive exPlanations (SHAP). Based on 114 stratified observations from 18 longitudinal sites during the wet and dry seasons, the model showed high predictive accuracy (R2 = 0.89; MAE = 1.05 mg L^{-1}). It also held up well during leave-one-season-out validation (R2 = 0.87–0.91). Under well-oxygenated conditions, fecal coliform concentrations had no effect on BOD_5, but when dissolved oxygen levels were low, they dramatically increased oxygen demand, suggesting hypoxia-conditioned microbial respiration. For every river, this relationship was geographically constant. The single mid-basin station was the source of all exceedances of the 5 mg L^{-1} BOD_5 standard, according to hotspot analysis. This suggests that focused sanitation efforts at this site could result in disproportionate downstream oxygen recovery. The study shows how hydrological conditions, water quality procedures, and management-relevant decision support may all be connected through interpretable machine learning.

1. Introduction

An increasingly common stressor that damages aquatic ecosystems, interferes with water supplies, and modifies carbon cycling is hypoxia in rivers and estuaries. Oxygen depletion is driven primarily by microbial respiration of labile organic matter, although the sources and environmental controls of oxygen demand remain insufficiently understood in subtropical catchments undergoing fast urbanization and limited wastewater treatment(Balkrishna, Ghosh, Arya, et al. 2025; Behrooz et al. 2021). Seasonal high-flow events in these areas mobilize untreated sewage and agro-industrial waste, often exceeding natural reaeration capacity and causing low-oxygen episodes that endanger food security, fisheries, and biodiversity. Biochemical oxygen demand (BOD5) is the regulatory baseline for measuring organic pollution, but its five-day incubation requirement constrains both the responsiveness and spatial resolution of monitoring systems. Machine learning offers intriguing options by forecasting BOD_5 from normal water-quality monitoring(Behrooz, Esmaili-Sari, and Chakraborty 2020; Dey et al. 2022; Syeed et al. 2023). Nevertheless, the majority of solutions prioritize prediction over explanation, serving as "*black boxes*" that offer little insight into the causes of hypoxia, when drivers become more intense, or where management should concentrate(Anand et al. 2024; Patel et al. 2023). In crucial river systems, where the ecological and socioeconomic costs are quite high, this methodological gap is especially severe. The Talar, Babolrood, and Haraz rivers in northern Iran are prime examples of this problem. Originating in the Alborz Mountains and draining ~ 12,000 km^2 northward into the southern Caspian Sea, these medium-sized rivers (Talar: 150 km, 2,845 km^2; Haraz: ~185 km, 5,100 km^2; Babolrood: similar scale) traverse intensive rice plains, food-processing hubs, and growing urban centres like Qaemshahr (320,000 residents) and Amol. They support Caspian sturgeon spawning corridors, irrigate more than 200,000

hectares, produce over 40% of Iran's rice, and serve about 2 million people, making them vital to the country's economy and environment. Yet the exact monsoon-driven flushing that maintains farmland also mobilizes untreated municipal wastewater and livestock manure, producing the perfect conditions for the unresolved oxygen-demand mechanisms mentioned above (Behrooz et al. 2021, 2020). The role of fecal-derived microbial populations is an important and poorly understood mechanism in rivers with poor sanitation. Fecal coliforms are metabolically active heterotrophs that may quickly break down biodegradable carbon, but they are also monitored worldwide as markers of pollution. Their impact on oxygen dynamics is probably highly dependent on dissolved oxygen (DO) levels: under oxygen stress, microbial respiration can quicken, creating positive feedbacks that exacerbate hypoxia(Alam, Barkatullah, and Kumar 2023; Chaudhary and Walker 2019; Varma et al. 2022). This hypoxia-conditioned microbial amplification has not been quantitatively characterized in systems like the Talar–Haraz network(Behrooz et al. 2021), limiting mechanistic understanding and potentially leading to systematic underestimation of hypoxia intensity precisely where fisheries and food security are most vulnerable(Behrooz et al. 2020). Spatial decision-support frameworks for river restoration also have a deficiency. Pollution management activities are frequently diffuse and catchment-wide, despite evidence that monsoon-affected watersheds like Mazandaran's have tight longitudinal bottlenecks where untreated wastewater inputs disproportionately dictate downstream oxygen quality(Behrooz et al. 2021, 2020). Remediation efforts remain misallocated, sluggish, costly, and environmentally inefficient if these mechanistic drivers are not integrated with spatial prioritizing in high-stakes systems(Balkrishna, Banerjee, et al. 2024; Balkrishna, Singh, et al. 2024). When it comes to managing hypoxia in subtropical rivers, these interrelated knowledge gaps—mechanistic, geographical, and operational—limit both predictive capacity and intervention effectiveness. For systems like the rivers that drain the Caspian, an integrated framework that can identify threshold-dependent microbial feedbacks, resolve oxygen-depletion mechanisms, and provide actionable spatial intervention priorities is desperately needed(Behrooz et al. 2021, 2020). To satisfy these goals, we analyzed the Talar, Babolrood, and Haraz rivers, integrating Random Forest regression with SHAP (SHapley Additive exPlanations) to obtain both high predictive accuracy and mechanistic transparency. Within a single DISPER framework—Decision-Support for Integrated Pollution Evaluation and Response—we found threshold effects, predictor interactions, and intervention hotspots(He et al. 2023; Wang et al. 2024). We hypothesized that fecal contamination functions as a context-dependent BOD amplifier, having little effect at high DO but greatly raising oxygen demand at low DO.

This study adds a new conceptual understanding for heated, sanitation-limited rivers: fecal contamination is not only a co-occurring pollutant but an active regulator of oxygen intake under hypoxia. In critical systems, operationalizing interpretable machine learning inside a DISPER framework (Balkrishna, Ghosh, Kaushik, et al. 2025; Singh et al. 2025) allows for the quick identification of high-impact wastewater intervention spots, speeding up oxygen recovery and improving resilience under extreme climatic and demographic pressures.

2. Materials and Methods

2.1 Study Area and Monitoring Design

Three subtropical rivers emptying to the southern Caspian Sea were explored to capture gradients in sanitary influence and hydrological regime. Six longitudinal stations(S1-S6) per river (18 total) were positioned to represent upstream reference zones, mid-basin urban/agro-industrial inputs, and downstream recovery reaches. Surface water samples were taken twice monthly (10th and 24th) from July–December 2022, generating 114 total samples (Figure S1). High-flow monsoon and low-flow post-monsoon conditions were obtained via sampling. All samples were collected at 0.5 m deep using sterile polyethylene vials.

2.2 Physicochemical and Microbiological Analyses

BOD_5, COD, dissolved oxygen (DO), pH, electrical conductivity (EC), nitrate, phosphate, temperature, and total coliforms were tested following conventional APHA methods. BOD_5 was quantified using the 5-day dilution method at 20°C; COD was measured by closed reflux; DO, pH, and EC were assessed in situ with calibrated multiparameter probes; nutrients were estimated using ion chromatography; and total coliforms were enumerated using multiple-tube fermentation(ISO 6878 2004; ISO 9308 2012; Public Health Association n.d.). Complete analytical and equipment data, limits of detection, calibration logs, and quality-control results are provided in Supporting Information (**Table S1**)(Balkrishna, Singh, et al. 2024; NIST 2025).

2.3 Quality Assurance and Uncertainty Control

Every sample was gathered in duplicate, shipped on ice, and examined in less than six hours. Blanks from the field and the lab verified that there was no contamination. The variability of replicate BOD5was less than 10% RSD. Probe accuracy was checked before each deployment using two-point calibration. Outliers owing to handling or equipment failure were screened using the 1.5×IQR criteria and excluded only when justified. Uncertainty related with BOD_5 prediction was assessed using 95% prediction intervals(Balkrishna, Ghosh, Kaushik, et al. 2025; Khot, Ghosh, and Shah 2021).

2.4 Data Preprocessing

All predictors (COD, DO, pH, EC, nitrate, phosphate, temperature, total coliforms) were evaluated for skewness and log-transformed where appropriate to stabilize variance (details in SI). Continuous variables were standardized (zero mean, unit variance). Station order (1–6) along each river transect was used to encode spatial position, while monsoon/post-monsoon (binary) was used to encode season(DebRoy et al. 2012; Saurav et al. 2013; Varma et al. 2022).

2.5 Machine Learning Model Development

A Random Forest regression model was designed to predict BOD_5 from the whole predictor set. To optimize predictive accuracy without overfitting, hyperparameters were adjusted using k-fold cross-validation. R^2 mean absolute error (MAE), root mean square error (RMSE), and coverage of ± 2 mg L^{-1} regulatory tolerance were used to assess performance. Leave-one-season-out validation examined

resilience across hydrological regimes. Supporting Information (**Table S2**) contains the model configuration, complete hyperparameter grid(He et al. 2023; Wang et al. 2024).

SHAP (SHapley Additive exPlanations) was employed to assess the direction, amplitude, and non-linear interactions of predictors on BOD_5. In order to identify hypoxia-conditioned microbial amplification of oxygen demand, threshold effects and conditional dependencies (such as coliform × DO) were assessed(He et al. 2023; Wang et al. 2024).

2.6. Spatial Hotspot Identification and Management Prioritization (DISPER Framework)

Predicted and measured BOD_5 distributions were examined to classify exceedance severity throughout each river transect. A DISPER decision-support framework (Decision-Support for Integrated Pollution Evaluation and Response) was used to determine station-level priorities (Burn et al. 2006; Tayebi et al. 2022). Stations were graded by size and persistence of BOD_5 exceedance and SHAP-confirmed causal drivers, providing a four-tier management action system (P1–P4): P1: Immediate wastewater intervention; P2: Targeted monitoring + flow/runoff restrictions; P3: Seasonal surveillance; P4: Routine monitoring. Criteria and thresholds used for classification are provided in Supporting Information (**Table S3**).

3. Results

3.1 Seasonal and spatial patterns in water quality

In all three of the studied rivers, water quality showed significant seasonal and longitudinal structuring. BOD_5, COD and total coliforms were non-normal (Shapiro–Wilk W = 0.85, 0.78, 0.94; $p < 0.01$), requiring non-parametric reporting. Median BOD_5 climbed from 3.5 (IQR 2.0–5.5) $mg \cdot L^{-1}$ in the post-monsoon period to 7.0 (4.0–11.0) $mg \cdot L^{-1}$ during the monsoon (Mann–Whitney $p < 0.001$, Cohen's $d = 0.85$), showing a strong monsoon effect on oxygen-demanding loading. COD and $\log_{10}$ coliforms increased in parallel (COD median 12.0 to 22.0 $mg \cdot L^{-1}$, $\log_{10}$ coliform median 1.60 to 2.00), but DO decline (8.4 to 7.3 $mg \cdot L^{-1}$), consistent with greater respiration under elevated organic and microbial load (Table 1).

Table 1
Seasonal Comparison of Physicochemical and Microbial Water Quality Indicators

Parameter	Monsoon (n = 68)	Post-Monsoon (n = 46)	p-value	Cohen's d
BOD_5 (mg/L)	7.0 (4.0–11.0)	3.5 (2.0–5.5)	< 0.001	0.85
COD (ppm)	22.0 (14.0–35.0)	12.0 (8.0–20.0)	< 0.001	0.39
DO (mg/L)	7.3 (6.5–8.1)	8.4 (7.5–9.2)	< 0.001	0.56
log_{10} Coliform	2.00 (1.30–2.70)	1.60 (1.00–2.20)	0.008	0.41
EC (µS/cm)	680 (520–900)	550 (450–700)	0.042	0.17
Temp (°C)	28.0 (26.0–30.0)	19.0 (16.0–22.0)	< 0.001	2.11

Median and interquartile range (IQR) values for biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), dissolved oxygen (DO), fecal indicator bacteria (log_{10} coliform), electrical conductivity (EC), and temperature measured during the monsoon and post−monsoon seasons.

Spatially, BOD_5 demonstrated a distinct from upstream to mid-basin to downstream pattern with peaks at the mid-basin site (site 4). The median BOD_5 for all rivers rose from 5.0 (3.5–6.5) mg·L^{-1} at Station 1 to 10.0 (7.5–13.0) mg·L^{-1} at Station 4 before falling to 7.0 (5.5–9.0) mg·L^{-1} at Station 6. The key ecological benchmark (BOD5 > 5 mg·L^{-1}) was 100% exceeded at Station 4. River-level exceedance followed the order Talar (58% > 5 mg·L^{-1}), Babolrood (42%), and Haraz (16%), demonstrating consistent but varying sanitation impact among catchments (Table 2).

Table 2

BOD Concentrations (Median with IQR in $mg.L^{-1}$) Across Upstream, Midstream, and Downstream Stations in the Three Rivers

River	Upstream		Midstream		Downstream		River Median
	St.1	St.2	St.3	St.4	St.5	St.6	
Talar	6.5 (5.5–7.5)	7.0 (6.0–8.0)	8.0 (6.5–9.5)	12.0 (10.0–14.0)	9.5 (7.5–11.5)	8.5 (7.0–10.0)	8.0 (5.5–11.0)
Babolrood	5.5 (4.5–6.5)	6.0 (5.0–7.0)	6.5 (5.5–8.0)	10.0 (8.0–12.0)	8.5 (6.5–10.0)	7.5 (6.0–9.0)	7.0 (4.5–9.5)
Haraz	3.0 (2.5–4.0)	3.5 (3.0–4.0)	4.0 (3.0–5.0)	6.5 (5.0–8.0)	5.5 (4.5–7.0)	5.0 (4.0–6.0)	4.5 (3.0–6.0)
All	5.0 (3.5–6.5)	5.5 (4.5–7.0)	6.0 (5.0–8.0)	10.0 (7.5–13.0)	8.0 (6.0–10.0)	7.0 (5.5–9.0)	6.0 (3.5–9.5)

Median and interquartile range (IQR) values of biochemical oxygen demand (BOD) at six longitudinal stations (St.1–St.6) across the Talar, Babolrood, and Haraz Rivers. "River Median" represents the pooled median (IQR) for each river. Pairwise comparisons were conducted using Dunn's post hoc test following a Kruskal–Wallis analysis; Abbreviation: IQR, interquartile range. Pairwise (Dunn's post–hoc): Talar vs. Haraz: $p = 0.003$, $d = 1.12$; Station 4 vs. Station 1: $p < 0.001$, $d = 1.0$

3.2 Interrelationships among predictors

Rank-based correlations emphasized clear ecological linkages: BOD_5 linked most strongly with COD (Spearman $\rho = 0.87$, $p < 0.001$) and $\log_{10}$ total coliform ($\rho = 0.75$, $p < 0.001$). Season (monsoon) also linked positively with BOD_5 ($\rho = 0.62$), while temperature and EC revealed moderate positive relationships ($\rho \approx 0.57$ and 0.51, respectively). As anticipated, DO and BOD5had an inverse relationship ($\rho = -0.50$). The predictive model and subsequent interaction analysis included COD and microbiological markers in addition to physicochemical predictors due to their monotonic correlations (Fig. 1).

3.3 Predictive performance and cross-validation

Strong predictive performance was obtained on the hold-out test set by the Random Forest (RF) regression model, which was trained on the entire predictor set ($R^2 = 0.89$; MAE = 1.05 $mg{\cdot}L^{-1}$; RMSE = 1.38 $mg{\cdot}L^{-1}$). Ninety-four percent of estimates landed within ± 2 $mg{\cdot}L^{-1}$ of actual BOD_5, and 95% prediction ranges encompassed 91% of observations, suggesting reliable uncertainty quantification. Residuals indicated no systematic bias (mean residual ≈ 0.03 $mg{\cdot}L^{-1}$) and satisfactory homoscedasticity; residual normality tests returned $p = 0.21$. The RF model's better fit for non-linear, interacting predictors was demonstrated by the significantly lower performance of a linear regression baseline ($R^2 = 0.65$; MAE = 1.92 $mg{\cdot}L^{-1}$) (Table 3).

Table 3
Comparative Performance of Random Forest and Linear Regression Models for Predicting Dissolved Oxygen

Metric	Random Forest	Linear Regression	Improvement
R^2	0.89	0.65	+ 0.24
MAE (mg/L)	1.05	1.92	−45%
RMSE (mg/L)	1.38	2.61	−47%
% Within ± 2 mg/L	94%	70%	+ 24 pp
95% PI Coverage	91%	78%	+ 13 pp
Mean Residual (mg/L)	0.03	0.21	—

Summary of model performance metrics for Random Forest and Linear Regression applied to dissolved oxygen prediction. Metrics include coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), the percentage of observations predicted within ±2 mg/L, 95% prediction interval (PI) coverage, and mean residual error. "Improvement" reflects the absolute difference (for R^2 and percent coverage) or relative reduction (for MAE and RMSE) of Random Forest performance compared to Linear Regression.

Leave-one-season-out (LOSO) validation verified seasonal generalization: the RF maintained high skill (R^2 = 0.87 for monsoon test; R^2 = 0.91 for post-monsoon test) when trained on post-monsoon and tested on monsoon (and vice versa), with MAE between 0.89–1.12 mg·L^{-1} and > 94% of predictions within ± 2 mg·L^{-1}. These diagnostics indicate the model is resilient to seasonal non-stationarity and adequate for year-round early warning of BOD_5 excursions (Table 4). The associated figures have been documented in Supplementary file (**F1 & F2**)

Table 4
Leave-One-Season-Out Validation Performance

Test Season	Train n	Test n	R^2	MAE (mg/L)	RMSE (mg/L)	% Within ± 2 mg/L
Monsoon	46	68	0.87	1.12	1.45	94%
Post-Monsoon	68	46	0.91	0.89	1.19	96%
Overall Avg	—	—	0.89	1.01	1.32	95%

3.4 Mechanistic interpretation using SHAP

SHAP decomposition of RF predictions gave both global and interaction-level mechanistic insight (Table 5). Globally, COD contributed the biggest mean |SHAP| effect (≈ 2.82 mg·L^{-1}; ~32% of overall contribution), followed by $\log_{10}$ total coliform (≈ 1.91 mg·L^{-1}; ~22%) and season (≈ 1.34 mg·L^{-1}; ~15%). The projected BOD5(mean |SHAP| ≈1.12 mg·L^{-1}) was negatively impacted by DO, which is consistent with oxygen availability inhibiting net respiration.

Table 5
SHAP-Based Ranking of Predictor Contributions in the Random Forest BOD Model

Rank	Predictor	Mean |SHAP| (mg/L)	% Total Contribution	Direction
1	COD	2.82	32.1%	Positive
2	$\log_{10}$ Coliform	1.91	21.7%	Positive
3	Season (Monsoon = 1)	1.34	15.3%	Mixed
4	DO	1.12	12.8%	Negative
5	Temperature	0.78	8.9%	Positive
6	EC	0.61	6.9%	Positive
7	Station	0.19	2.2%	Positive
8	NO_3	0.04	0.5%	Weak
9	pH	0.02	0.2%	Neutral
10	PO_4	0.01	0.1%	Neutral

Predictor importance estimated using mean absolute SHAP values (mean |SHAP|), expressed in mg L^{-1} contribution to BOD predictions. Percentage contribution refers to each variable's share of total model influence. "*Direction*" indicates the dominant monotonic effect observed in SHAP main effects: Positive (increasing BOD), Negative (decreasing BOD), Mixed (context-dependent), Weak, or Neutral. Variables include physicochemical parameters, microbial indicators, season, and station index. The single strongest interaction, according to interaction analysis, was the $\log_{10}$ coliform × DO pair (mean |interaction SHAP| ≈0.68 mg·L^{-1}; ~38% of overall interaction contribution). Dependence graphs indicated a clear threshold behaviour: the SHAP contribution of coliforms to anticipated BOD_5 was near-zero for DO levels above ≈ 5 mg·L^{-1} but climbed dramatically when DO went below this threshold (Table 5). This pattern shows a hypoxia-conditioned microbial amplification: fecal-derived microbial activity contributes disproportionately to oxygen demand when oxygen is already constrained.

Local force-plot case studies reveal that COD, coliforms, and the coliform×DO interaction dominate high-BOD_5 predictions at hotspot samples (Table 6). The SI contains interaction matrices and complete SHAP graphs have been documented in supplementary file (**Figure F3**).

Table 6
Pairwise SHAP Interaction Effects Revealing Drivers of BOD Dynamics

Interaction Pair	Mean |Interaction SHAP| (mg/L)	% of Total Interaction	Ecological Interpretation
log_{10} Coliform × DO	0.68	38.4%	Microbial BOD amplified under hypoxia
COD × Season	0.51	28.8%	Monsoon concentrates organic load
COD × Station	0.31	17.5%	Downstream accumulation
Coliform × Season	0.19	10.7%	Fecal wash-off in monsoon
DO × Temperature	0.08	4.5%	Temperature-driven respiration

Mean absolute interaction SHAP values for the strongest feature pairs contributing to Random Forest BOD predictions. Values represent the magnitude of interaction effects relative to total interaction contributions. The final column summarizes the ecological interpretation associated with each interaction, based solely on mechanistic patterns identified in SHAP analyses. Abbreviation: SHAP, Shapley Additive Explanations.

3.5 Spatial prioritization and DISPER action tiers

Significant regional differences in pollution status were found throughout the examined areas when river-level BOD5 compliance was evaluated against the ecological benchmark of 5 mg/L (Table 7).

Table 7
River-Level WHO Compliance

River	Median (IQR)	% > 5 mg/L	WHO Status
Talar	8.0 (5.5–11.0)	58%	Moderate → Polluted
Babolrood	7.0 (4.5–9.5)	42%	Moderate
Haraz	4.5 (3.0–6.0)	16%	Clean → Moderate

The Talar River demonstrated the highest systemic pollution burden, with 58% of its samples above the threshold and a median BOD_5 of 8.0 mg/L resulting in a categorization of Moderate to Polluted WHO grade. The Babolrood River was classed as Moderate 42% exceedance), whereas the Haraz River displayed the best overall quality (> 16% exceedance), attaining a Clean to Moderate grade.

Building upon this river-level context, a station-level analysis identified extremely different pollutant profiles, necessitating site-specific prioritizing within the DISPER framework (Table 8).

Table 8
Station-Level WHO Hotspots

Station	Median (IQR)	% > 5 mg/L	WHO Class	Hotspot Severity	Priority
St.4	10.0 (7.5–13.0)	100%	Polluted	Extreme Critical	P1
St.5	8.0 (6.0–10.0)	79%	Moderate → Polluted	Critical	P2
St.6	7.0 (5.5–9.0)	68%	Moderate	Moderate Critical	P3
St.3	6.0 (5.0–8.0)	63%	Moderate	Moderate Critical	P3
St.2	5.5 (4.5–7.0)	50%	Moderate	Less Critical	P4
St.1	5.0 (3.5–6.5)	42%	Clean → Moderate	Less Critical	P4

P1 (Extreme): Immediate treatment plant; P2 (Critical): Enhanced monitoring + runoff control; P3 (Moderate): Seasonal surveillance; P4 (Less): Routine checks

The BOD_5 standard was uniformly exceeded at Station 4 (100% exceedance), which reported the highest median value (10.0 mg/L). This level of organic contamination exceeded the strict Extreme Critical severity standards (Median Deviation > 5.0mg/L), necessitating its classification as the highest action tier: P1 (Immediate wastewater intervention)(Burn et al. 2006). On the other hand, Stations 1 and 2 received the lowest action tier, P4 (Routine checks), because they had the lowest exceedance rates (< 50%) and small median deviations from the benchmark, placing them in the Less Critical severity class. Intermediate action tiers P2/P3 were systematically allocated to Stations 5, 6, and 3, consistent with their Critical and Moderate Critical BOD_5 pollution burdens, respectively. This integrated classification thus generates a spatially explicit and severity-graded management hierarchy across all monitoring locations, transforming the observed pollution gradients into actionable intervention priorities. Although the absolute number of exceedances increases with the stricter criterion, the DISPER ranking is robust, with Station 4 continuing to be the highest-priority intervention in all tested scenarios.

4. Discussion

The combined pressures of wastewater discharge, microbial pollution, and climate unpredictability make freshwater rivers in rapidly urbanizing areas more susceptible to hypoxia. Fecal pollution exacerbates oxygen depletion nonlinearly, especially when dissolved oxygen is already limited, according to this study's mechanistic and geographical data(Roslev, Bjergbæk, and Hesselsoe 2004). This interaction, which can be quantified using interpretable machine learning, validates a positive microbial feedback loop in which respiration rises at the exact moment when the aquatic system is least able to support functions that require oxygen(Singh et al. 2021). Such feedbacks decrease recovery windows, amplify diel stress on aquatic biota, and contribute to the persistence of hypoxic conditions(Bjergbæk and Roslev 2005).

Importantly, our data reveal that hypoxia creation is not equally distributed over the river continuum. Rather, the mid-basin reaches, where hydraulic dilution decreases but urban expansion and wastewater load peak, are where the greatest oxygen stress is concentrated. These midstream limitations function as biogeochemical tipping zones: neither pristine headwaters nor tidal-influenced downstream parts demonstrated comparable ecological stress. This geographical disparity is consistent with global data from Asia(Alam and Singh 2023; Dey et al. 2022), Africa(Yavuz 2025), and South America(Baptista et al. 2007; González et al. 2022), where peri-urban and mid-catchment population surges lag infrastructural development. Our findings thus highlight a structural global challenge: wastewater systems are constructed for cities' cores, whereas water quality collapse arises outside them(Eriksen et al. 2022; Li et al. 2020).

In a number of ways, the study's integrated methodology increases our knowledge of anthropogenic hypoxia(Eriksen et al. 2022; Paul 2017). Traditional pollution evaluation frequently ends at threshold exceedances; in contrast, our SHAP-enabled research exposes why certain conditions cause excessive BOD? We demonstrated that coliform activity emerges as a dominating oxygen sink only when DO decreases below an ecological WHO safety limit (~ 5 $mg \cdot L^{-1}$)(BIS 2023; WHO 2022). This mechanistic transparency is critical for evidence-based intervention, as it suggests that improving cleanliness alone may be insufficient once hypoxia is established—oxygen recovery must be emphasized in parallel to interrupt microbial growth(El-Ghannam et al. 2023; Tayebi et al. 2022).

By integrating these mechanistic insights into the DISPER priority(Burn et al. 2006), we move from science to practical decision-making. The most effective wastewater interventions can be applied to P1 hotspots, which are the areas that cause the majority of hypoxic load, rather than dispersing scarce resources thinly over entire basins. This aligns the study with national strategies aimed at river rejuvenation, climate-adaptive water governance, and global river-restoration goals under SDG 6(UN 2025). In many places, investment plans are hampered by economic and logistical restrictions; consequently, focusing interventions where ecological returns are greatest becomes both a scientific and societal imperative.

Furthermore, the resilience of our model throughout monsoon and non-monsoon seasons suggests that machine learning may assist ongoing, anticipatory water management in data-constrained contexts. Instead of reacting to seasonal degradation, predictive insights allow utilities and regulators to prepare for hypoxia-prone periods, responding at earlier warning signals(Singh et al. 2021). This capability is crucial as climate change intensifies storm-driven microbial loads and affects natural oxygen-regeneration mechanisms across the Global South(Abd El Moniem 2009; Aryal, Kim, and Lakshmi 2025; Behrooz et al. 2021). Our workflow, which includes priority mapping, mechanistic interpretation, predictive modelling, and data preprocessing, may be immediately applied to river systems with comparable developmental histories.

Nevertheless, important challenges continue that dictate future research orientations. First, flow estimation and health risk assessment are limited by measurement gaps in river discharge and microbial

functional activity. Second, geographical monitoring networks need to be densified in order to capture micro-scale discharge plumes, even though they are adequate for identifying hotspots. Third, comprehending infrastructure weaknesses and governance capacity—aspects outside the purview of this study—is essential to the viability of interventions. Fully integrated river-basin management that integrates research, engineering, politics, and community dynamics will be possible if these issues are resolved.

Overall, this study provides a generalizable conceptual development to river ecology and environmental management: hypoxia risk originates not merely from excessive organic pollution, but from a synergistic coupling between insufficient dilution and respiration amplification by microbial pollutants(Baptista et al. 2007; Bjergbæk and Roslev 2005; Li et al. 2020; Roslev et al. 2004). By disclosing both the mechanism and the geography of this link, and by embedding the findings inside a decision-making framework, we present a timely model for concentrating limited restoration resources where oxygen recovery will be fastest and most permanent(Mandour et al. 2021). Our work demonstrates how water quality integrated with machine learning approach can accelerate environmental protection by bridging the gap between complex system dynamics and actionable interventions.

5. Recommendations and Policy Implications

In mid-basin urban communities, where untreated or partially treated sewage causes hypoxia through a microbial amplification mechanism, this study clearly and urgently indicates the need for focused wastewater management. Based on the spatial and mechanistic findings, the following management recommendations are proposed:

Prioritize mid-basin hotspots for immediate wastewater improvement Midstream areas pose the greatest risk and the greatest opportunity: they receive the highest residential sewage loads while possessing little natural diluting capability. DISPER-based prioritization promotes incremental, high-return investment, commencing with P1 hotspots where hypoxia originates. Near-term alternatives like as decentralized treatment and microbial load reduction technology should be promptly deployed to disrupt oxygen-depletion cascades.

Combine active oxygen-restoration techniques with sanitation interventions. Once DO drops below the ~ 5 $mg \cdot L^{-1}$ ecological safety margin, when microbial proliferation accelerates, reducing pollution loads alone might not be enough. Managers should employ dual-action recovery: improved aeration, environmental flow releases, and nature-based remedies (riparian buffers, floodplain reconnection) to increase resistance against oxygen collapse.

Establish predictive early warning systems for seasonal stress. Monsoon-linked hydrological variability induces acute surges in microbial contamination. Predictive machine-learning algorithms should prompt pre-event operational alarms for utilities and regulators, enabling heightened surveillance during storm periods and early activation of mitigating solutions before hypoxia becomes prevalent.

Modernize regulatory frameworks to reflect microbial–oxygen coupling Environmental standards should migrate away from pollutant-specific limits and toward interaction-aware criteria that concurrently track fecal contamination and oxygen stress. Adoption of dual-trigger criteria will better safeguard both aquatic life and human health, while combining with SDG-6 and climate adaption targets to ensure sustainable compliance.

Strengthen governance capacity and financing for equitable implementation Coordinated governance across administrative boundaries is necessary since hypoxia hotspots are frequently located outside the core city jurisdictions. In areas with tight budgets, scant data, and severe sanitation disparities, development will be accelerated through performance-linked finance, industrial co-responsibility, and community-based monitoring. Because of its intrinsic scalability, the DISPER architecture facilitates resilient river recovery in rapidly urbanizing basins worldwide.

6. Conclusion

Rivers sustain life, culture, and economies, yet their resistance is progressively eroded by human-driven oxygen stress. This study illustrates that restoring these systems is not about addressing every source of degradation equally, but about understanding where hypoxia originates and why it persists. By identifying mid-basin sections as the main leverage points of recovery and exposing the underlying connection between microbial pollution and decreased oxygen conditions, we offer a clear roadmap that is both practically feasible and supported by science. This study provides communities and policymakers with the knowledge and guidance they need to safeguard the living rivers that sustain their future.

Declarations

Author contributions

Reza Dahmardeh Behrooz: Software, Resources, Project administration, Methodology,

Investigation, Funding, acquisition. Sourav Ghosh: Formal analysis, Data curation, Writing – review

Funding Declaration

This article was supported by the University of Zabol (Iran) with grant code (IR-UOZ-GR-0088)

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Ethical approval: N/A

Consent to participate: N/A

Consent to Publication: All authors have given their consent to publish the manuscript in current form.

Competing interests: Authors declare no competing interests

Clinical trial number: N/A

References

1. Abd El Moniem, Ahmed. 2009. 'Overview of Water Resources and Requirements in Egypt'. *Environmental Studies* 2(1):82–97. doi:10.21608/jesj.2009.183681.
2. Alam, Aftab, Md Barkatullah, and Amit Kumar. 2023. 'Water Quality Status of Different Ghats of River Ganga in Patna Urban Area'. *Engineering Proceedings 2023, Vol. 56, Page 85* 56(1):85. doi:10.3390/ASEC2023-15408.
3. Alam, Aftab, and Anshuman Singh. 2023. 'Groundwater Quality Assessment Using SPSS Based on Multivariate Statistics and Water Quality Index of Gaya, Bihar (India)'. *Environmental Monitoring and Assessment* 195(6). doi:10.1007/S10661-023-11294-7.
4. Anand, Akshay, Amit Kumar, Jayant Kumar, Bihari Singh, and Tripti Gangwar. 2024. 'Analysis of Impact of Cremation on Water Quality of River Ganga at Digha Ghat in Patna, Bihar, India'. *International Journal of Environmental Sciences* 7(3):40–56. doi:10.47604/IJES.2791.
5. Aryal, Aashutosh, Kyung Y. Kim, and Venkataraman Lakshmi. 2025. 'Understanding Flood Dynamics in the Indus River Basin: Lessons from the 2022 Pakistan Deluge'. *Journal of Hydrology: Regional Studies* 59:102362. doi:10.1016/J.EJRH.2025.102362.
6. Balkrishna, Acharya, Srimoyee Banerjee, Sourav Ghosh, Divya Chauhan, Ilika Kaushik, Vedpriya Arya, and Sumit Kumar Singh. 2024. 'Reuse of Sewage Sludge as Organic Agricultural Products: An Efficient Technology-Based Initiative'. *Applied and Environmental Soil Science* 2024(1):1433973. doi:10.1155/2024/1433973.
7. Balkrishna, Acharya, Sourav Ghosh, Vedpriya Arya, Divya Joshi, Shelly Singh, Kumud Kannojia, Adity Saxena, Aswani Nagose, Sanyam Taneja, and Durgesh Kumar Jaiswal. 2025. 'Accessing the Differential Suitability of Ganga River Water for Drinking, Irrigation, and Industrial Use: A First-Time Report Based on Index Analysis and Unsupervised Machine Learning, Highlighting the Need for Region-Specific Water Management Approaches'. *ACS ES&T Water*. doi:10.1021/ACSESTWATER.5C00574.
8. Balkrishna, Acharya, Sourav Ghosh, Ilika Kaushik, Ved Priya Arya, Divya Joshi, Diksha Semwal, Aditi Saxena, and Shelly Singh. 2025. 'Sequential Distribution, Potential Sources, and Health Risk Assessment of Persistent Toxic Substances in Sewage Sludge Used as Organic Fertilizer in Indo-Gangetic Region'. *Environmental Science and Pollution Research* 32(5):2324–58. doi:10.1007/S11356-024-35706-4/METRICS.
9. Balkrishna, Acharya, Sumit Kumar Singh, Sourav Ghosh, Srimoyee Banerjee, Sonam Verma, and Vedpriya Arya. 2024. 'An Analytical Review on the Integrated Management of River Resources through Namami Gange'. *Water Policy* 26(5):462–79. doi:10.2166/WP.2024.209.

10. Baptista, Darcilio F., Daniel F. Buss, Mariana Egler, Alexandre Giovanelli, Mariana P. Silveira, and Jorge L. Nessimian. 2007. 'A Multimetric Index Based on Benthic Macroinvertebrates for Evaluation of Atlantic Forest Streams at Rio de Janeiro State, Brazil'. *Hydrobiologia* 575(1):83–94. doi:10.1007/S10750-006-0286-X.
11. Behrooz, Reza Dahmardeh, Abbas Esmaili-Sari, and Paromita Chakraborty. 2020. 'Distribution and ECo-Toxicological Risk Assessment of Legacy Persistent Organic Pollutants in Surface Water of Talar, Babolrood and Haraz Rivers'. *Water 2020, Vol. 12, Page 3104* 12(11):3104. doi:10.3390/W12113104.
12. Behrooz, Reza Dahmardeh, Abbas Esmaili-Sari, Magdalena Urbaniak, and Paromita Chakraborty. 2021. 'Assessing Diazinon Pollution in the Three Major Rivers Flowing into the Caspian Sea (Iran)'. *Water 2021, Vol. 13, Page 335* 13(3):335. doi:10.3390/W13030335.
13. BIS. 2023. 'Indian Standard Specifications for Drinking Water Is : 10500 - 1983. | Pollution Control Board | Government Of Assam, India'. https://pcb.assam.gov.in/information-services/indian-standard-specifications-for-drinking-water.
14. Bjergbæk, L. A., and Peter Roslev. 2005. 'Formation of Nonculturable Escherichia Coli in Drinking Water'. *Journal of Applied Microbiology* 99(5):1090–98. doi:10.1111/j.1365-2672.2005.02706.x.
15. Burn, S., D. Desilva, M. Ambrose, S. Meddings, C. Diaper, R. Correll, R. Miller, and L. Wolf. 2006. 'A Decision Support System for Urban Groundwater Resource Sustainability'. *Water Practice & Technology* 1(1). doi:10.2166/WPT.2006010.
16. Chaudhary, Meenakshi, and Tony R. Walker. 2019. 'River Ganga Pollution: Causes and Failed Management Plans (Correspondence on Dwivedi et al. 2018. Ganga Water Pollution: A Potential Health Threat to Inhabitants of Ganga Basin. Environment International 117, 327–338)'. *Environment International* 126:202–6. doi:10.1016/J.ENVINT.2019.02.033.
17. DebRoy, S., S. Das, S. Ghosh, S. Banerjee, D. Chatterjee, A. Bhattacharjee, I. Mukherjee, and Shaon RayChaudhuri. 2012. 'Isolation of Nitrate and Phosphate Removing Bacteria from Various Environmental Sites'. *OnLine Journal of Biological Sciences* 12(2):62–71. doi:10.3844/OJBSCI.2012.62.71.
18. Dey, Uttiya, Soumyajit Sarkar, Srimanti Duttagupta, Animesh Bhattacharya, Kousik Das, Sumedha Saha, and Abhijit Mukherjee. 2022. 'Influence of Hydrology and Sanitation on Groundwater Coliform Contamination in Some Parts of Western Bengal Basin: Implication to Safe Drinking Water'. *Frontiers in Water* 4:875624. doi:10.3389/FRWA.2022.875624/BIBTEX.
19. El-Ghannam, Mohamed K., Fatma Wassar, Sabah Morsy, Mohamed Hafez, Chiter M. Parihar, Kent O. Burkey, and Ahmed M. Abdallah. 2023. 'Controlled Drainage in the Nile River Delta of Egypt: A Promising Approach for Decreasing Drainage off-Site Effects and Enhancing Yield and Water Use Efficiency of Wheat'. *Journal of Arid Land 2023 15:4* 15(4):460–76. doi:10.1007/S40333-023-0095-3.
20. Eriksen, Tor Erik, Dean Jacobsen, Benoît O. L. Demars, John E. Brittain, Geir Søli, and Nikolai Friberg. 2022. 'Effects of Pollution-Induced Changes in Oxygen Conditions Scaling up from Individuals to

Ecosystems in a Tropical River Network'. *Science of The Total Environment* 814:151958. doi:10.1016/J.SCITOTENV.2021.151958.

21. González, Padilla, P.;. Bautista-Capetillo, C.;. Ruiz-Canales, A.;. González-Trinidad, J.;. Júnez-Ferreira, H. E.;. Contreras Rodríguez, Robles Rovelo, Pedro Padilla González, Carlos Bautista-Capetillo, Antonio Ruiz-Canales, Julián González-Trinidad, Hugo Enrique Júnez-Ferreira, Ada Rebeca Contreras Rodríguez, and Cruz Octavio. 2022. 'Characterization of Scale Deposits in a Drinking Water Network in a Semi-Arid Region'. *International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 3257* 19(6):3257. doi:10.3390/IJERPH19063257.
22. He, Zheyu, Yuanjian Yang, Runzhuo Fang, Shaohui Zhou, Wenchuan Zhao, Yingjie Bai, Junsheng Li, and Bo Wang. 2023. 'Integration of Shapley Additive Explanations with Random Forest Model for Quantitative Precipitation Estimation of Mesoscale Convective Systems'. *Frontiers in Environmental Science* 10:1057081. doi:10.3389/FENVS.2022.1057081/BIBTEX.
23. ISO 6878. 2004. *ISO 6878:2004 - Water Quality — Determination of Phosphorus — Ammonium Molybdate Spectrometric Method*. https://www.iso.org/standard/36917.html.
24. ISO 9308. 2012. *ISO 9308-2:2012(En), Water Quality — Enumeration of Escherichia Coli and Coliform Bacteria — Part 2: Most Probable Number Method*. https://www.iso.org/obp/ui/#iso:std:iso:9308:-2:ed-2:v1:en.
25. Khot, Siddiqua, Sourav Ghosh, and Janvi Shah. 2021. 'Elucidating the Role of Plant Growth Promoting Bacteria for Nitrate and Phosphate Bioremediation: A Sustainable Approach Towards Crop Productivity and Environmental Protection'. *Journal of Applied Biotechnology Reports* 8(4):385–94. doi:10.30491/JABR.2020.240413.1260.
26. Li, Jianhua, Xiaochen Chen, Xinliang Zhang, Zhenjia Huang, Liang Xiao, Liangliang Huang, Yuichi Kano, Tatsuro Sato, Yukihiro Shimatani, and Chunpeng Zhang. 2020. 'Fish Biodiversity Conservation and Restoration, Yangtze River Basin, China, Urgently Needs "Scientific" and "Ecological" Action'. *Water (Switzerland)* 12(11):1–8. doi:10.3390/W12113043.
27. Mandour, Ahmed, Mahmoud Kh El-Sayed, Ayman A. El-Gamal, Ahmed M. Khadr, and Ahmed Elshazly. 2021. 'Temporal Distribution of Trace Metals Pollution Load Index in the Nile Delta Coastal Surface Sediments'. *Marine Pollution Bulletin* 167:112290. doi:10.1016/J.MARPOLBUL.2021.112290.
28. NIST. 2025. 'Standard Reference Materials | NIST'. https://www.nist.gov/srm.
29. Patel, Abhinav, Shive Prakash Rai, Kossitse Venyo Akpataku, Nijesh Puthiyottil, Abhinesh Kumar Singh, Neeraj Pant, Rajesh Singh, Prashant Rai, and Jacob Noble. 2023. 'Hydrogeochemical Characterization of Groundwater in the Shallow Aquifer System of Middle Ganga Basin, India'. *Groundwater for Sustainable Development* 21:100934. doi:10.1016/J.GSD.2023.100934.
30. Paul, Dipak. 2017. 'Research on Heavy Metal Pollution of River Ganga: A Review'. *Annals of Agrarian Science* 15(2):278–86. doi:10.1016/J.AASCI.2017.04.001.
31. Public Health Association, American. n.d. 'APHA Method 4500-NO3: Standard Methods for the Examination of Water and Wastewater'.

32. Roslev, P., L. A. Bjergbæk, and M. Hesselsoe. 2004. 'Effect of Oxygen on Survival of Faecal Pollution Indicators in Drinking Water'. *Journal of Applied Microbiology* 96(5):938–45. doi:10.1111/J.1365-2672.2004.02209.X.
33. Saurav, Ghosh, Indranil Mukherjee, Ashoke Ranjan Thakur, and Shaon Ray Chaudhuri. 2013. 'Assay for Rapid Screening of Phytochemicals as Antimicrobial Agents'. *American Journal of Biochemistry and Biotechnology* 9(1):8–11. doi:10.3844/AJBBSP.2013.8.11.
34. Singh, Rupanjali, Anuj Singh, Chandrajit Balomajumder, and Ajit Kumar Vidyarthi. 2025. 'Assessment of Industrial Effluent Discharges Contributing to Ganga Water Pollution through a Multivariate Statistical Framework: Investigating the Context of Indian Industries'. *Environmental Science and Pollution Research International*. doi:10.1007/S11356-024-35823-0.
35. Singh, Vijay Kumar, Devendra Kumar, Sudhir Kumar Singh, Quoc Bao Pham, Nguyen Thi Thuy Linh, Safwan Mohammed, and Duong Tran Anh. 2021. 'Development of Fuzzy Analytic Hierarchy Process Based Water Quality Model of Upper Ganga River Basin, India'. *Journal of Environmental Management* 284:111985. doi:10.1016/J.JENVMAN.2021.111985.
36. Syeed, M. M. Mahbubul, Md Shakhawat Hossain, Md Rajaul Karim, Mohammad Faisal Uddin, Mahady Hasan, and Razib Hayat Khan. 2023. 'Surface Water Quality Profiling Using the Water Quality Index, Pollution Index and Statistical Methods: A Critical Review'. *Environmental and Sustainability Indicators* 18:100247. doi:10.1016/J.INDIC.2023.100247.
37. Tayebi, Mohsen, Akram Bemani, Abdolvahhab Fetanat, and Masoud Fehresti-Sani. 2022. 'A Decision Support System for Sustainability Prioritization of Air Pollution Control Technologies in Energy and Carbon Management: Oil & Gas Industry of Iran'. *Journal of Natural Gas Science and Engineering* 99:104416. doi:10.1016/J.JNGSE.2022.104416.
38. UN. 2025. 'THE 17 GOALS | Sustainable Development'. https://sdgs.un.org/goals.
39. Varma, Kriti, Piyush Tripathi, Stuti Upadhyaya, Atul Srivastava, Nirdesh Kumar Ravi, Anjali Singhal, and Pawan Kumar Jha. 2022. 'Assessment of Mass Bathing Event (Kumbh- 2019) Impact on the River Water Quality by Using Multivariate Analysis and Water Quality Index (WQI) Techniques at Sangam (Prayagraj), India.' *Groundwater for Sustainable Development* 17:100750. doi:10.1016/J.GSD.2022.100750.
40. Wang, Zhiyuan, Huihui Liu, Muhammad Nasir Amin, Kaffayatullah Khan, Muhammad Tahir Qadir, and Suleman Ayub Khan. 2024. 'Optimizing Machine Learning Techniques and SHapley Additive ExPlanations (SHAP) Analysis for the Compressive Property of Self-Compacting Concrete'. *Materials Today Communications* 39:108804. doi:10.1016/J.MTCOMM.2024.108804.
41. WHO. 2022. 'Water Sanitation and Health'. https://www.who.int/teams/environment-climate-change-and-health/water-sanitation-and-health/water-safety-and-quality/drinking-water-quality-guidelines.
42. Yavuz, Veysel Süleyman. 2025. 'Impact of Temperature and Flow Rate on Oxygen Dynamics and Water Quality in Major Turkish Rivers'. *Scientific Reports* 15(1):1–16. doi:10.1038/S41598-025-06433-8;SUBJMETA.

Figures

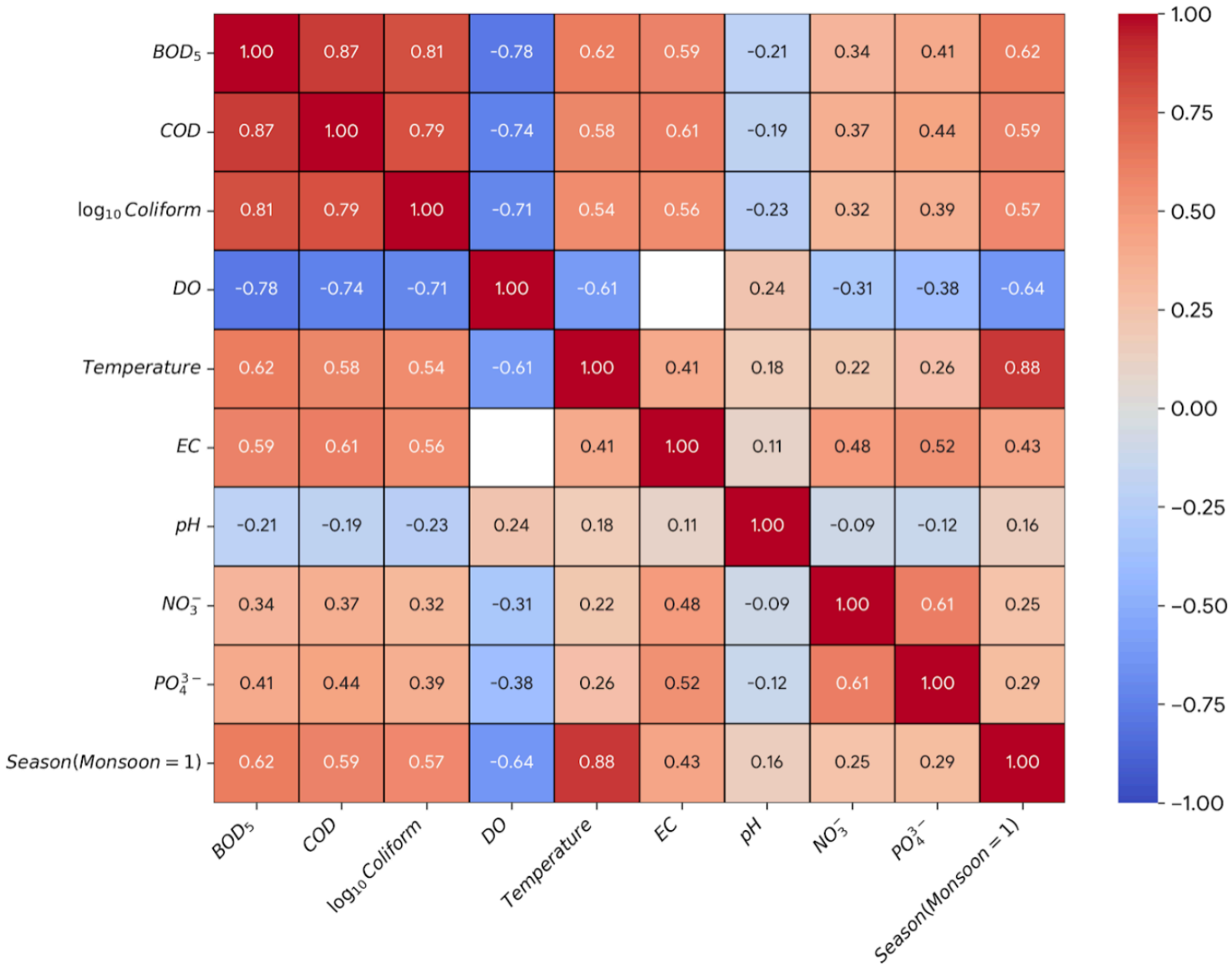


Figure 1

Correlations among the tested variables.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supportinginfo.docx