

袁站站, 魏卿, 陈沛沛, 等. 基于可解释机器学习的洱海北部流域溶解氧驱动因素分析与改善策略[J]. 净水技术, 2026, 45(3): 14-24.

Yuan Z Z, Wei Q, Chen P P, et al. Explainable machine learning-based analysis of driving factors and improvement solutions for DO in northern Erhai Basin [J]. Water Purification Technology, 2026, 45(3): 14-24.

基于可解释机器学习的洱海北部流域溶解氧驱动因素分析与改善策略

袁站站^{1,2}, 魏卿^{1,2}, 陈沛沛^{1,2}, 徐祖信^{1,2,*}

(1. 同济大学环境科学与工程学院, 上海 200092; 2. 同济大学水污染控制与资源绿色循环全国重点实验室, 上海 200092)

摘要 【目的】溶解氧(DO)是评价水环境质量的关键性指标之一, DO浓度过低会破坏生态平衡, 威胁水生环境的健康。本文通过建模预测DO浓度并解释驱动其变化的关键因素, 提出调控水质保护水生态的重要措施。【方法】基于洱海北部流域2011年—2020年水质水量监测数据, 本文构建了集成机器学习框架, 建立并比选出预测DO变化的最佳模型——轻量梯度提升模型(LightGBM)。此外, 本文还采用夏普利加性解释(SHAP)量化了不同特征对DO浓度变化的贡献。【结果】LightGBM的决定系数(R^2)较基准模型提高11.2%, 在预测中展现出优异性能[均方根误差(RMSE)=0.284 mg/L, 平均绝对误差(MAE)=0.226 mg/L, $R^2=0.912$]。SHAP分析显示流量是影响溶解氧浓度的最主要因素, 贡献率为35.5%, 其次是化学需氧量(COD)为17.2%。DO随流量增加先上升后下降, 适当流量有利于增氧, 大流量时携带大量耗氧物质从而降低DO。适度流量和低COD是维持较高DO的重要条件, 流量为 $0.2\text{ m}^3/\text{s}$ 左右最为适度, COD则是越低越好。【结论】研究区域夏季流量大、冬季流量小乃至断流, 均造成DO浓度低。利用流域沿线库塘净化调蓄, 可维持流量适度减少径流污染, 有利于维持流域DO含量保护水生态。本文旨在增进对DO受流量动态影响的理解, 为高原河流生态保护提供参考。

关键词 轻量梯度提升 夏普利加性解释(SHAP) 溶解氧预测 流量作用 库塘调蓄

中图分类号: X703 **文献标志码:** A **文章编号:** 1009-0177(2026)03-0014-11

DOI: 10.15890/j.cnki.jsjs.2026.03.002

Explainable Machine Learning-Based Analysis of Driving Factors and Improvement Solutions for DO in Northern Erhai Basin

Yuan Zhazhan^{1,2}, Wei Qing^{1,2}, Chen Peipei^{1,2}, Xu Zuxin^{1,2,*}

(1. College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China;

2. State Key Laboratory of Water Pollution Control and Green Resource Recycling, Tongji University, Shanghai 200092, China)

Abstract [Objective] Dissolved oxygen (DO) is a key indicator for evaluating water environmental quality, low DO concentration in rivers disrupts ecological balance and threatens aquatic ecosystem health. This paper models and predicts DO concentrations, identifies the key factors driving its variations, and proposes essential measures for regulating water quality and protecting aquatic ecosystems. [Methods] Based on water quality and quantity monitoring datas from the northern Erhai Lake Basin from 2011 to 2020, this paper constructed an integrated machine learning framework, established and selected the optimal model—light gradient boosting machine model (LightGBM) for predicting changes in DO. Furthermore, this paper employed Shapley additive explanations (SHAP) to quantify the contributions of different features to the variation in DO concentration. [Results] The coefficient of determination (R^2) for the LightGBM model improved by 11.2% compared to the baseline model, demonstrating superior predictive performance [root mean square error (RMSE)=0.284 mg/L, mean absolute error (MAE)=0.226 mg/L, $R^2=0.912$]. SHAP analysis revealed that flow rates emerged as the dominant influencing factor (35.5%), followed by chemical oxygen demand (COD) (17.2%). DO

[收稿日期] 2025-12-09

[基金项目] 云南省科技厅顶尖团队项目(202505AT350002-4)

[作者简介] 袁站站(2002—), 男, 硕士研究生, 研究方向为流域水污染治理研究, E-mail: yuanzz@tongji.edu.cn。

[通信作者] 徐祖信(1956—), 女, 教授, 研究方向为城市河流环境综合整治、污水管网建设、改造与运行调控、农业面源污染治理, E-mail: xzx@tongji.edu.cn (Z. Xu)。

levels initially increased and then decreased with rising flow rates; moderate flow rates enhanced oxygenation, while excessive flow rates introduced a large amount of oxygen-consuming pollutants, thus reducing DO. Suitable flow rates and low COD were identified as key conditions for maintaining high DO concentrations, a flow rate of $0.2 \text{ m}^3/\text{s}$ was recommended, and achieving the lowest possible COD level was crucial. [**Conclusion**] In the study areas, both high summer flows and low winter flows (or flow interruption) lead to decreased DO concentration. Regulating water storage and purification using reservoirs and ponds along the basin can stabilize flow rates, reduce runoff pollution, and sustain DO levels. This paper enhances understanding of flow-dependent DO dynamics and provides a scientific basis for ecological conservation in plateau river systems.

Keywords light gradient boosting Shapley additive explanation (SHAP) DO prediction flow rate effect reservoir storage and regulation

溶解氧(DO)含量是评价水环境质量的关键性指标之一,它反映了水生态环境的健康程度和支持水生生物生存的能力。水体中维持一定的DO对水生生物的代谢、捕食、生长和繁殖等基本生理活动至关重要^[1]。水体缺氧会影响水体污染物的生物化学反应过程,还会对鱼类、浮游生物等各类水生生物活动造成影响^[2]。水体中的DO通常来源于大气扩散、溶解和植物藻类的光合作用,DO浓度高低受多种因素耦合作用影响,一般认为有机污染物是其消耗的主要因素^[3-5],季节性的水文活动引发水体和沉积物中有机物、微生物及各种动植物之间的物质交换,使得DO浓度驱动机制更加复杂。水文过程调控DO浓度,进而影响水生环境健康,由此衍生出环境流量的概念^[6]。因此,预测水体中DO含量,并识别驱动DO变化的关键因素,对于采取有针对性的措施保护水生态环境至关重要。

不同的模型在预测DO浓度具备不同的特点。现有的模拟DO模型通常可分为物理模型、统计模型和机器学习模型3类^[7]。物理模型的构建基于DO动态变化相关的物理定律,比如斯特里特-费尔普斯模型(Streeter-Phelps)公式,通过使用微分方程表示的氧平衡进行建模^[8-9]。物理模型在理论解释上有着坚实的基础,但它们需要大量的精确测量数据,因此其在应用中非常复杂和困难。统计模型的构建基于历史数据,常用的有回归模型、时间序列分析模型等,目前应用较为广泛,通常采用历史数据探究DO的周期性与非周期性特征^[10-11]。赵海超等^[12]通过线性拟合的方式研究DO浓度和环境因子的关系,发现其与氮磷呈负相关。但在实际情况下,影响DO水平的环境因素作用通常是复杂和非线性的,这限制了相关模型的准确性和适用性。

机器学习是一种快速发展的方法,作为数据驱

动模型,它能够识别输入和响应之间的潜在关系,并且在多源数据预测分析方面具有卓越的性能^[13]。已有研究表明:机器学习算法,如提升回归树、决策树算法、随机森林(RF)模型、支持向量机和遗传算法等在水质评估方面表现良好,并且在大多数情况下,预测准确性优于传统统计方法。Cao等^[14]采用多因素分解改进提高了DO预测模型的准确性。Xu等^[15]采用独立模型和基于小波变换的混合模型提高模型的预测精度。Nong等^[16]则利用特征工程优化参数提高了向量回归预测DO的精度和准确性。基于Shapley值的可解释性模型,更可以在识别结果与变量间非线性关系的同时,给出全局和局部一致的解释。Kundu等^[17]则通过可解释机器学习模型建立水质特征间的联系,并提出采用DO作为综合水质评价指标。这些机器学习技术上的进步为DO驱动因素分析带来强大的工具,结合预测的准确度和结果的可解释性,支撑管理实践策略的提出和应用。

现有的研究主要关注水质指标,往往忽视水量与水质间的关系。水量管理是水资源管理重要的组成部分,水量和水质相互影响共同塑造水环境健康的边界阈值,忽略其中任意一个维度均难以全面把握水体生态系统的真实状态。尤其在高原湖泊流域,水文过程对水质的影响更为敏感和复杂,但目前利用机器学习方法系统揭示流量特征与DO动态关系的研究仍较为缺乏。因此,集成水量要素进行DO预测与归因,不仅是方法上的创新,也对明晰水生态系统的驱动机制、制定针对性保护策略具有重要意义。

基于以上研究现状,本研究构建融合可解释机器学习的创新框架,将河道流量特征加入为变量,揭示DO的驱动因素及水动力响应,并提出相应水量

管理策略。主要研究内容:(1) 采用 4 种集成学习算法预测洱海北部流域河流的 DO 含量,并比选出最优模型;(2) 结合夏普利加性解释(SHAP)方法量化影响因子的边际贡献度,利用部分依赖图解析洱海北部流域 DO 驱动因素的阈值及交互关系;(3) 基于机器学习模型的结论并结合该地区实际情况,提出适用于该地区提高 DO 含量的水量管理策略。本研究希望通过特定地域的水质-水量耦合关系分析,定量揭示 DO 如何受多因素综合影响,以可解释的机器学习手段支撑管理决策,为维护高原湖泊水生态健康提供科学依据与实践参考。

1 方法

1.1 数据收集与处理

洱海是中国第七大淡水湖,位于中国云南省大理市境内,北部区域是洱海的主要水源补给区。洱海是高原淡水湖泊的典型代表,其生态环境敏感,治理难度较大。由于流域内人类活动加剧,洱海在 20 世纪 80 年代起水质开始下降,从贫营养状态转向富营养状态,并曾暴发 2 次蓝藻水华^[18]。洱海的污染,特别是氮、磷等富营养化关键污染物,其来源在很大程度上与北部流域密切相关^[19]。

本研究对象以洱海北部流域的河流为主,具体包含弥直河、罗时江和永安江等。本文的水文水质数据来源于云南洱海湖泊生态系统国家野外科学观测研究站对该流域的长期常规监测,监测点位具体涵盖大楼桥、绿玉池下游、跃进河下游、西湖下游、小江桥、莲河桥、湿地上游、兆邑村、黑泥沟、入湖口等 10 个主要监测点(图 1),监测时间为 2011 年—2020 年,共采集了 2 220 组数据。

10 个监测点分布于洱海北部流域的源头、上游段、中游段、下游段和入湖口,采集样品时间为每个月的月初、月中,数据包含地表水环境的瞬时流量(Q)、水温(T)、DO、总氮(TN)、氨氮、硝酸盐(NO_3^-)、总磷(TP)、可溶性磷酸盐(PO_4^{3-})和化学需氧量(COD)9 项指标。为保证模型的有效性,同时兼顾数据处理效率,考虑前人的研究,数据处理对空值进行了线性插值补充。线性插值基于相邻数据点构建直线关系,其计算复杂度低、执行速度快,在连续数据中稳定性强。异常值则采用箱型图法进行了清洗处理,箱型图通过关键统计量及异常值标注,直观呈现数据的集中趋势、离散程度和特征,其优势在

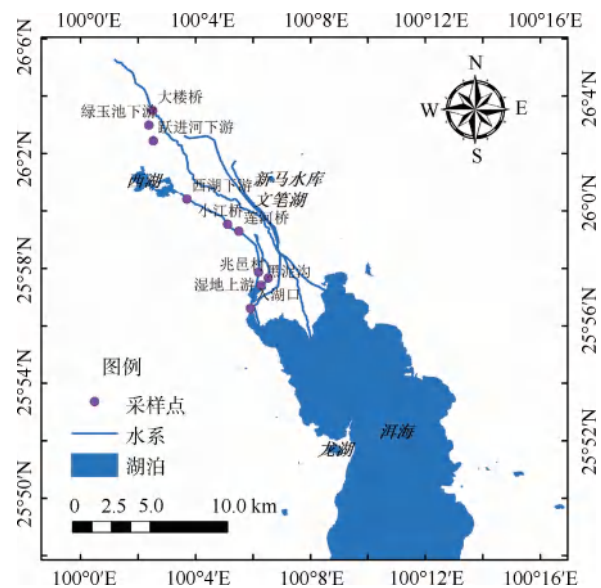


图 1 研究区域点位

Fig. 1 Location Map of the Study Areas

数据初探、异常检测及多组对比场景中尤为突出。表 1 为数据处理后所有参数的统计分析。

表 1 机器学习使用数据的参数统计分析

Tab. 1 Parametric Statistical Analysis of Data Used in Machine Learning

| 参数 | 单位 | 最大值 | 最小值 | 平均值 | 标准差 |
|--------------------|-----------------------|--------|-------|--------|-------|
| Q | m^3/s | 15.746 | 0.003 | 1.250 | 1.865 |
| T | $^{\circ}\text{C}$ | 32.700 | 8.500 | 19.429 | 4.503 |
| DO | mg/L | 13.950 | 0.410 | 5.751 | 2.609 |
| TN | mg/L | 30.556 | 0.152 | 2.261 | 2.246 |
| 氨氮 | mg/L | 4.361 | 0.001 | 0.277 | 0.343 |
| NO_3^- | mg/L | 19.505 | 0.020 | 0.845 | 1.282 |
| TP | mg/L | 0.936 | 0.007 | 0.120 | 0.088 |
| PO_4^{3-} | mg/L | 0.637 | 0.001 | 0.048 | 0.047 |
| COD | mg/L | 15.920 | 0.440 | 5.627 | 2.082 |

相关系数通常用于判断测量变量之间的线性关系。相关系数为 0~1,系数越大表示变量之间的关系越强,在模型解释中的负面效果就越大,因此相关系数较大的特征需要剔除。对各项指标进行相关性分析(图 2),发现 TN、TP 与其他指标存在重要的相关性,高于 0.7 的可接受范围,因此删除相关特征,删除后的数据相关性绝对值最高为 0.41,可以满足机器学习的数据需求。

1.2 机器学习模型

本文采用梯度提升决策树(GBDT)、极限梯度

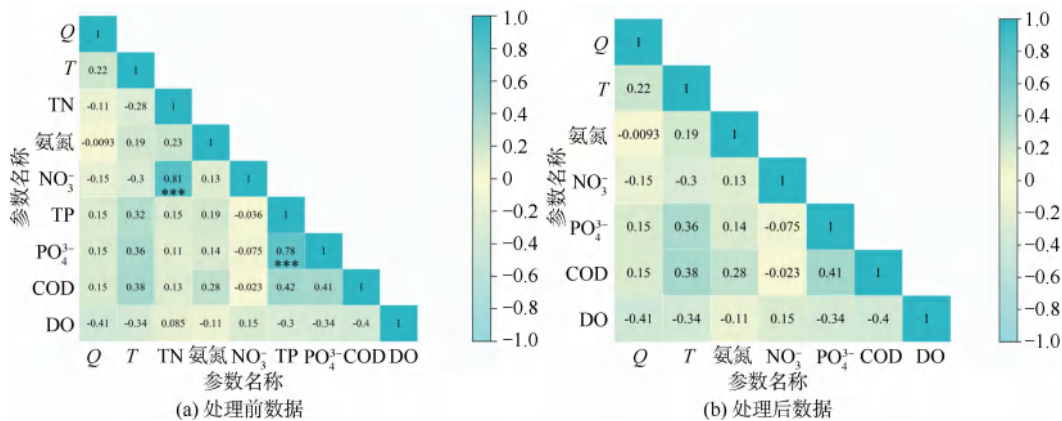


图2 数据相关系数分析

Fig. 2 Correlation Coefficient Analysis of Data

提升(XGBoost)、轻量梯度提升(LightGBM)和随机森林(RF)4类机器学习模型对水质的DO进行模拟预测,比选出最优的模型进行后续解释分析。基于3.12版本的python解释器,通过调用Python Scikit Learn包运行。

GBDT是一种基于Boosting的集成学习算法,通过迭代构建多棵决策树,逐步拟合残差来优化预测结果,其核心思想是每棵新树学习前一棵树的预测误差,最终将所有树的结果累加。XGBoost是GBDT的优化版本,引入正则化和二阶泰勒展开,提升训练效率和模型精度,它支持并行计算和缺失值处理,广泛应用于大规模数据场景^[20]。LightGBM针对XGBoost的计算效率优化,提出直方图算法、单边梯度采样(GOSS)和互斥特征捆绑(EFB),适用于高维和大规模数据。随机森林基于袋装法(Bagging)思想,通过构建多棵决策树并投票或平均预测结果,降低方差,提升泛化能力,适用于高维数据和噪声较多的场景,得到广泛引用。

1.3 模型评价与训练

决定系数(R^2)、均方根误差(RMSE)和平均绝对误差(MAE)用来评估机器学习模型的性能^[21]。MAE衡量预测值与实际值之间的平均绝对偏差,反映预测的平均误差大小。以上几种误差的值越小,模型越准确。RMSE、MAE的计算方式如式(1)~式(3)。

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (1)$$

$$R_{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$M_{AE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

其中: n ——样本量;

y_i, \hat{y}_i ——模拟值和测量值;

\bar{y}_i ——测量值的平均值。

在应用交叉验证后,使用测试数据集计算所有模型的误差指标,以确保结果可靠。

为了训练算法,研究将数据集分为训练集(80%的数据)和测试集(20%的数据)。采用 R^2 评分指标随机网格搜索来评估模型在训练子集上的性能,并找到最优超参数。初步的手动调整表明学习率(learning_rate)和决策树的数量($n_estimators$)等对机器学习性能有积极影响。模型中的其余超参数是根据Scikit-learn中提供的默认配置自动选择的。

1.4 SHAP

SHAP是一种基于博弈论的可解释性框架,通过计算每个特征对模型预测的贡献值(SHAP值),提供局部和全局解释^[22]。其核心是SHAP值,通过公平分配特征贡献(如效率性、对称性等公理)实现模型预测的透明。SHAP分析结果常由蜂巢图、柱形图、部分依赖图和个体条件期望图等形式展现。SHAP值的计算如式(4)。

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (4)$$

其中: M ——特征的总数;

ϕ_i ——特征情况;

\varnothing_0 ——特征基线值;
 z'_i ——在夏普利值计算中取 0 或 1, 表示包含或排除要素。

2 结果与讨论

2.1 环境因子特征与关联

洱海北部流域水文水质数据在月度上的分布特征如图 3 所示。采用因素月度均值代表各因素的月度情况,并根据箱型高度与颜色定位水质极值。氨氮、 PO_4^{3-} 和 COD 的均值都是在 6 月达到顶峰,DO 在 7 月达到最低值,在全年中夏季水质情况需要重点关

注。水文因子中, Q 的每个月的均值分布在 $0.46 \sim 2.91 \text{ m}^3/\text{s}$; T 的每个月的均值分布在 $12.4 \sim 23.9 \text{ }^\circ\text{C}$ 。水质因子中,COD 的每个月的均值分布在 $4.46 \sim 7.92 \text{ mg/L}$;氨氮和 NO_3^- 每个月的均值分布在 $0.15 \sim 0.56 \text{ mg/L}$, $0.36 \sim 1.83 \text{ mg/L}$; PO_4^{3-} 的每个月的均值分布在 $0.021 \sim 0.096 \text{ mg/L}$; DO 的每个月的均值则分布在 $3.98 \sim 7.61 \text{ mg/L}$, 峰值普遍在地表水 IV 类标准范围。标准化箱线图还揭示了各项水文水质数据的离散性,水文水质数据在夏季峰值离散程度相对较大,表明降雨输入对水质波动具备一定程度的影响。

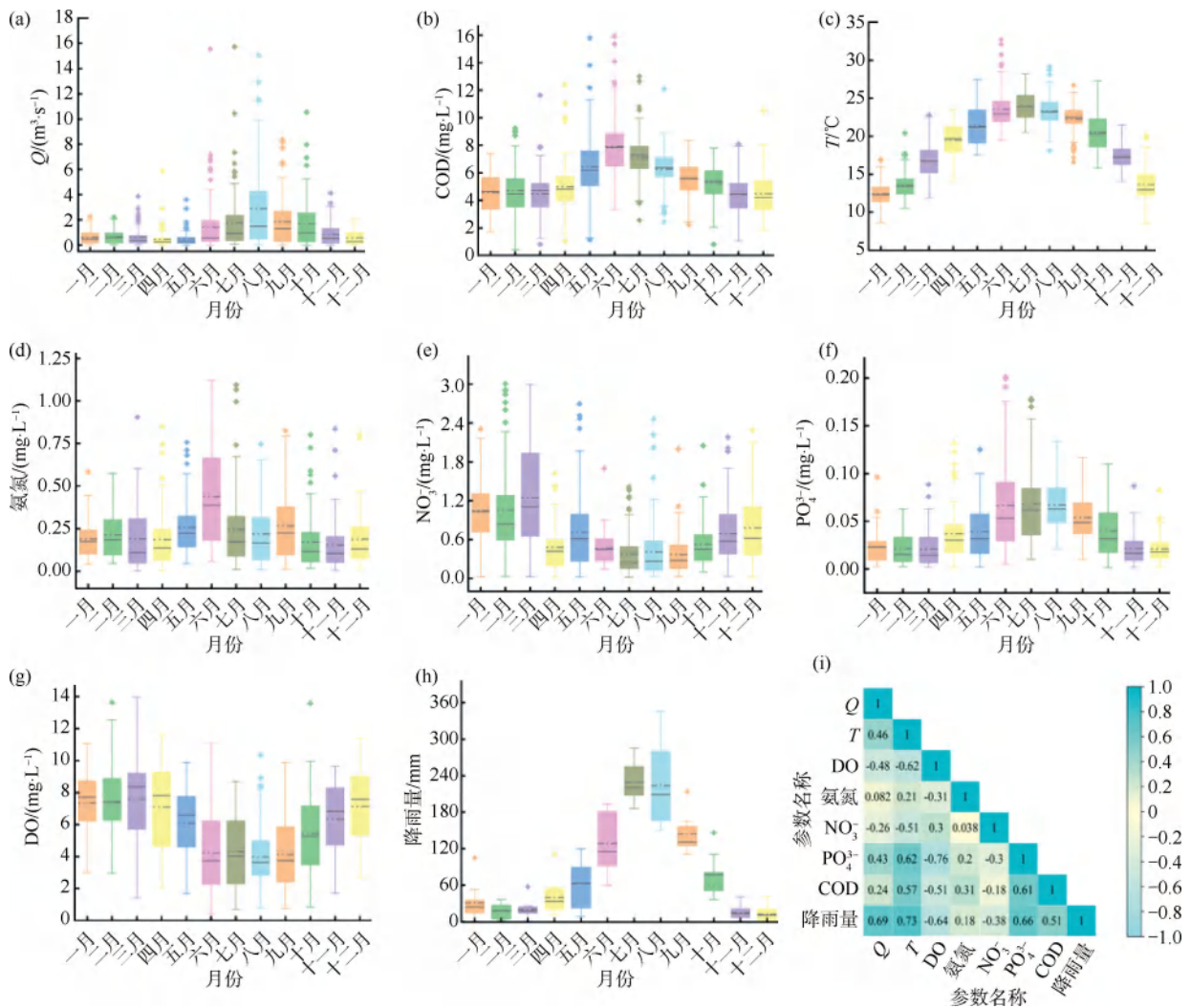


图 3 环境因子数据特征箱型图及相关性

Fig. 3 Box Plot of Environmental Factor Data Characteristics and Correlation

各因素间也存在一定的相关性。 T 和降雨量的均值在 6 月—8 月时达到最大,存在较强的正相关性,这和该区域的季风气候特征相符合。 Q 的均值在 7 月—9 月达到最大,结合相关性分析可以看出,

降雨量和 Q 存在一定的正相关性,说明 Q 有部分来源于降雨,并存在一定的滞后情况。COD 和 PO_4^{3-} 的均值也在 6 月—8 月时最大,同时也和降雨量这一因素存在正相关性。降雨冲刷一定程度上提高水

体 COD 和 PO_4^{3-} 的含量。

洱海北部流域的农业种植以水稻为主,种植面积占比为 80%,时间主要集中在每年的 6 月—9 月^[23]。夏季水文特征与农业活动具有同步性,此期间降水事件频发,与农作物营养生长及生殖旺盛期(需水需肥量最大阶段)高度重叠,夏季水质较差可能是农业面源污染物入河的原因。其他研究^[24-25]比较发现:雨季氮、磷浓度明显高于旱季,也证实了洱海北部流域水质受到面源污染的影响。

2.2 模型性能比较

通过对比不同模型在训练集和测试集上的精度和误差指标,如图 4 所示,可以看出 LightGBM 模型是预测 DO 的最佳模型。对于 DO 的模拟预测,

LightGBM 模型表现出色,模型 RMSE 为 0.284、MAE 为 0.226、 R^2 为 0.912。训练集的模型预测数据点与 1:1 线紧密对齐,偏差最小,分布均匀,表明具有出色的拟合和泛化性。这可能归功于其可以有效地处理模型数据而不会过度拟合。XGBoost 排名为第 2,测试指标达到 RMSE 为 0.298、MAE 为 0.238、 R^2 为 0.907。和 LightGBM 相反,RF 表现出最弱的性能,记录了测试 RMSE 为 0.414、MAE 为 0.331、 R^2 为 0.820。与 RF 相比,LightGBM 模型分别将 RMSE、MAE 和 R^2 优化了 31.4%、31.7% 和 11.2%。

LightGBM 模型是训练和测试数据集中表现最稳定的指标,可以作为后续预测 DO 含量并分析 DO 主导因素的参考模型。

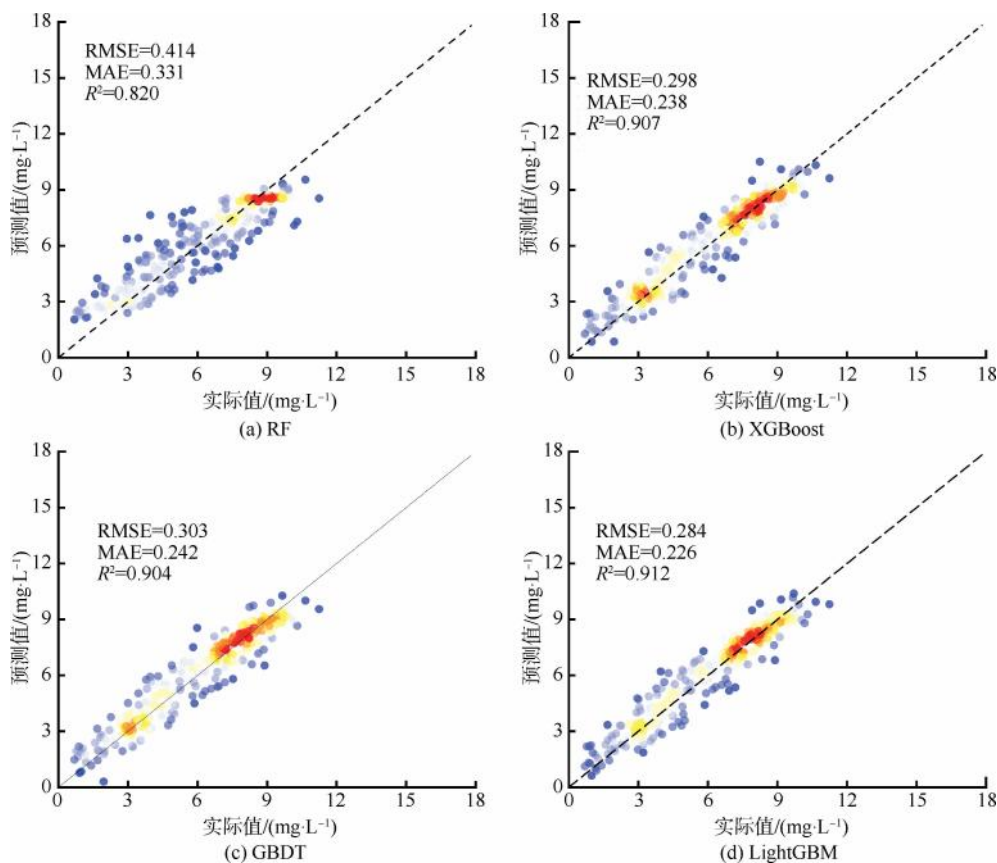


图 4 不同模型模拟结果的散点图

Fig. 4 Scatter Plot of Simulation Results for Different Models

2.3 DO 影响因素的解释分析

2.3.1 特征重要性分析

SHAP 蜂群汇总图汇总各特征与 DO 的 SHAP 值,展示了各特征影响 DO 的大趋势,其中 Q 是 DO 浓度的最重要预测因子,其次是 COD 和 T 。右侧大

于 0 的蓝色 SHAP 值表示特征增大对目标 DO 的浓度具有抑制效应。因此,除 NO_3^- 外的其他环境因子,都倾向于诱导 DO 降低。此外, T 和 COD 在右侧部分还有部分紫色与红色,意味着特殊情况下, T 和 COD 能够倾向于诱导 DO 升高。

全局柱形图是全局重要性图表,显示各因素的平均绝对 SHAP 值,可快速定位关键特征并展示具体贡献度大小。如图 5 所示,各因素对于 DO

浓度变化的贡献度情况中, Q 占比为 35.5%,COD 和 T 占比分别为 17.2% 和 15.8%。其余参数贡献较小。

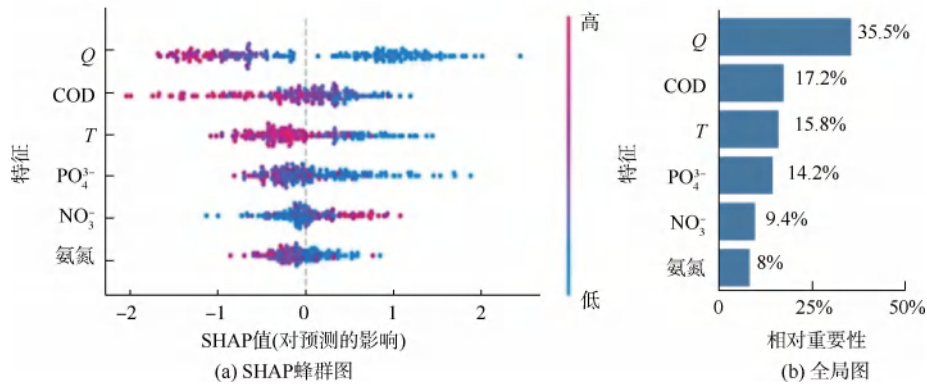


图 5 基于 LightGBM 模型预测 DO 的 SHAP 图

Fig. 5 SHAP Plot of LightGBM Model for DO Prediction

2.3.2 功能贡献的可视化

SHAP 力图是实现局部可解释性的重要工具,能够洞察单个预测的形成过程。它们通过直观展示具体特征对每个预测的影响,帮助决策者信任并理解模型。对于给定实例,力图的基线代表模型在训练数据集中的平均预测值,箭头则表示各项特征的贡献。增加预测值的特征用红色表示,

降低预测值的特征用蓝色表示,箭头长度表示影响的大小。横轴量化了这些贡献,揭示了预测值如何偏离基线。图 6 展示了 2 个随机选取的 LightGBM 模型实例。2 组中 Q 、COD 和 T 均是关键负向因素,对 DO 浓度有显著的负向贡献,其他特征影响较小。这些例子强调了 Q 和 COD 影响 DO 的主要因素。

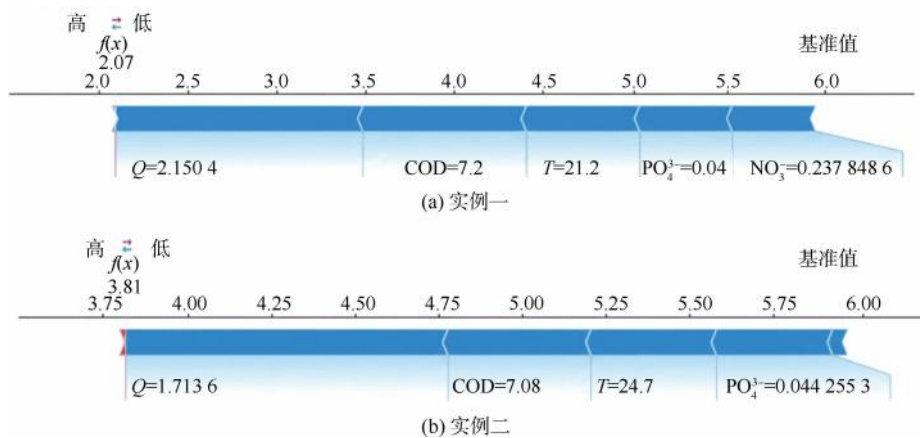


图 6 基于 LightGBM 模型的 2 组力图

Fig. 6 Force Plots of Two Random Groups of Samples Based on LightGBM Model

2.3.3 偏依赖分析

偏依赖图展示 1 个或 2 个特征对模型预测的全局平均边际效应,通过固定目标特征值并边缘化其他特征的影响,用于揭示特征与预测值的单调性或非线性关系。图 7 展示了各单因素对 DO 预测的效应。各项因素中,低 Q 阶段有诱导 DO 升高的趋势;高于一定程度后, Q 导致 DO 降低且速率先快后慢;

Q 再升高后 DO 随 Q 增大缓慢降低[图 7(a)]。COD 在前段对 DO 影响较小,高于一定阈值后稳定导致 DO 降低[图 7(b)], T 则有诱导 DO 先下降后升高的趋势[图 7(c)]。 PO_4^{3-} 驱动 DO 降低到一定程度后基本维持稳定[图 7(d)]。高 DO 含量时会强化硝化反应,驱动氨氮向 NO_3^- 的转化,水中 DO 高时 NO_3^- 较高而氨氮较低[图 7(e)~图 7(f)]。

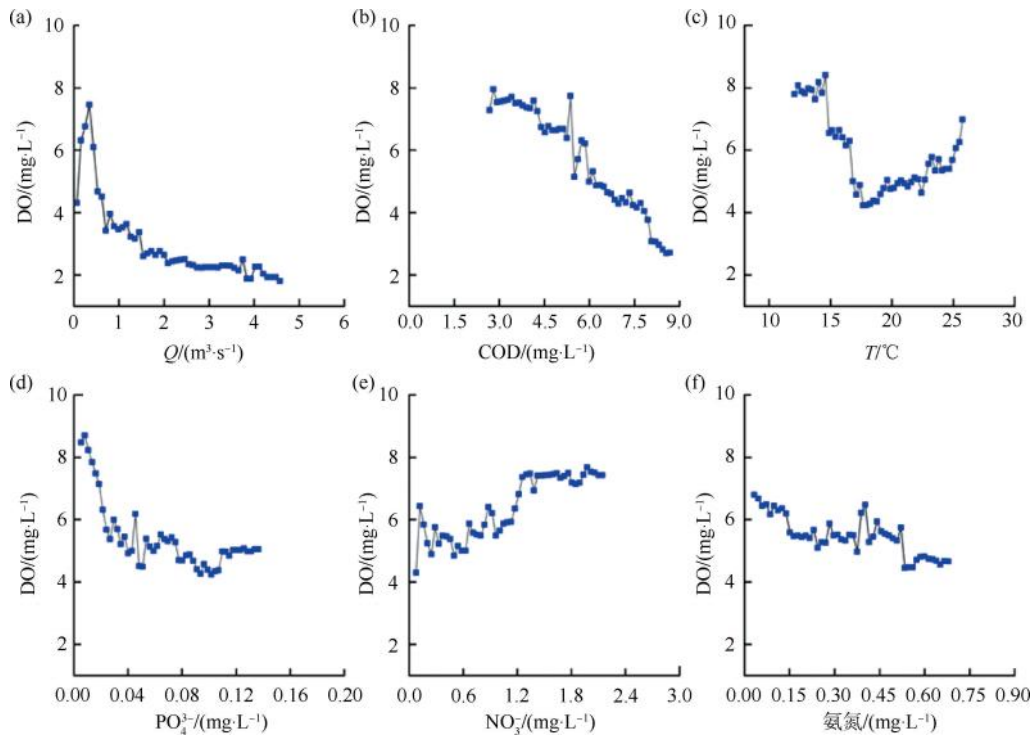


图7 基于 LightGBM 模型的 DO 六特征偏依赖图

Fig. 7 Partial Dependence Plots of Six Features Based on LightGBM Model for DO

影响水中 DO 含量变化的机制,主要可以归纳为大气复氧、水生植物产氧、DO 逸散和生物消耗 4 个部分。SHAP 分析一致表明, Q 是驱动洱海北部流域 DO 动态的最关键因子,这一发现与大多数河流生态系统研究中水文过程主导水体理化性质的观点相符:在一定流量下,水流流速强化了水体的大气复氧,提高了水体 DO 含量。本研究进一步揭示了在高原流域, Q 与 DO 的关系呈现出独特的非线性特征[图 7(a)],即低 Q 时复氧为主,高 Q 时耗氧物质冲刷效应占主导,这补充了现有模型多在平原河流中建立的简单线性或单调关系。COD 和降雨量之间的关系,佐证了降雨道中雨水径流带入大量有机物进入水体的观点^[26-27]。这与其他研究^[28]较为一致,间歇性洪水冲击会增大入河耗氧物质含量,降低河流 DO 含量,引发水体缺氧现象。

T 的升高最先引起的是 DO 逸散加剧,降低饱和 DO 浓度,同时一方面会诱导微生物活动加剧,进而强化生物消耗 DO 量,另一方面还会强化水体产氧光合生物的光合作用,增强 DO 的产生^[29]。 T (15.8%) 的贡献及其与 DO 先负后正的非典型关系,暗示了在洱海北部流域,光合作用产氧的效应可

能在高温下凸显,这一发现对传统上认为 T 升高仅通过降低溶解度和加速呼吸作用来减少 DO 的认知提出了挑战。氨氮较高时通常会发生硝化反应,降低水中 DO; NO_3^- 则由硝化反应产生,通过反硝化作用去除,其浓度较 DO 含量存在一定的滞后性,但基本上呈正相关。

2.3.4 与 Q 的交互效应

为进一步探讨 Q 与其他因素之间的相互作用,构建了特征交互偏依赖图,以考察其他因素与 Q 关于 DO 浓度的相互作用效应(图 8)。这些图可视化了 2 个变量的综合影响如何影响模型输出,同时保持其他特征不变。由图 8(b)可知,当 Q 超过 $1.5 \text{ m}^3/\text{s}$ 时,DO 等值线几乎与横轴平行,反映出此时调节 Q 作用有限。这暗示着在强降雨径流期间,流域内可能存在一个“污染负荷饱和”效应,即无论 Q 如何再增加,单位水体所承载的耗氧有机物浓度已达上限,此时单纯增加稀释能力对提升 DO 的效果有限。这为管理实践提供了重要启示:必须在源头控制 COD 负荷,而非仅仅依赖水文调控。整体来看, Q 对 DO 的影响始终是最大的,各交互偏依赖图均在 Q 为 $0.15 \sim 0.21 \text{ m}^3/\text{s}$ 存在一个峰值,适度 Q 是

维持较高 DO 的重要条件,该地区 Q 在 $0.2 \text{ m}^3/\text{s}$ 左右最为适度。COD 与 PO_4^{3-} 的峰值集中在低浓度

区,这与整体负影响 DO 的趋势一致; T 的峰值分裂成低温和高温的 2 个,也与偏依赖图的趋势相同。

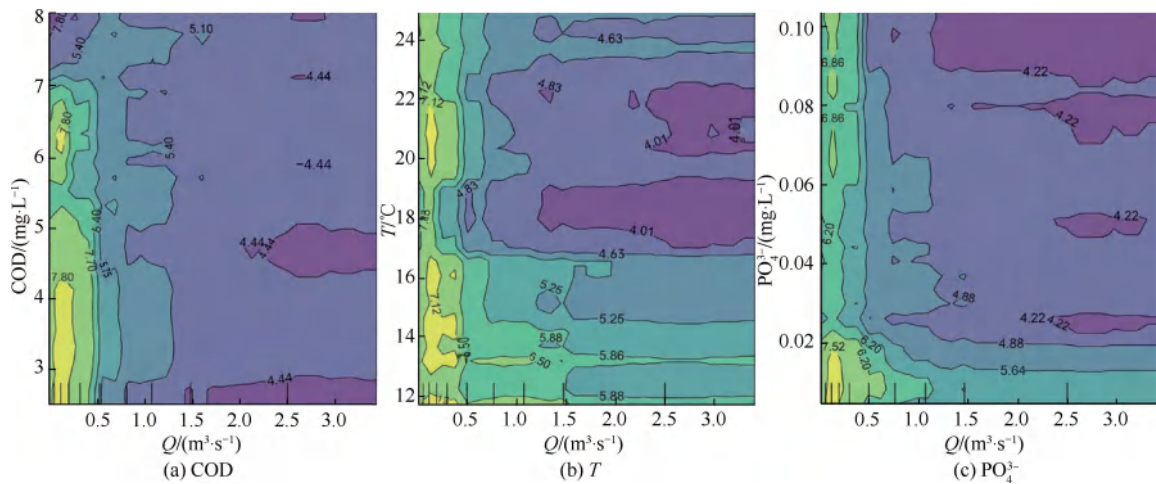


图 8 基于 LightGBM 模型的流量与其他因素特征二维偏依赖图

Fig. 8 2D Partial Dependence Plots between Flow Rate and other Factors Based on LightGBM Model

2.4 管理策略建议

本研究通过可解释机器学习方法 (SHAP) 揭示: Q 与 COD 是洱海北部流域 DO 季节性波动的主要驱动因素。这一发现不仅印证了水文过程与污染物输入对高原湖泊水质的耦合影响机制,也为针对性的流域精细化管理提供了直接的科学依据。

模型分析表明:DO 在雨季(6月—9月)的恶化是 Q 增大、外源 COD 输入增加及 T 升高等因素协同作用的结果。这揭示了传统关注点源治理的局限性:在农业主导的流域,非点源污染的水文驱动特征尤为突出。农田退水随降雨径流溢流,形成了“流量增加—污染冲释—DO 下降”的典型路径。与之相对,旱季(3月—5月)虽然外源输入减少,水质总体较好,但水体流动性不足可能导致局部区域出现低 DO 风险,提示在静态或缓流水体中,内源释放与复氧能力的平衡同样需要关注。因此,流域管理需从“污染控制”转向“水文—污染协同调控”,核心在于平抑流量峰值、削减污染负荷,并保障水体的适度流动。这要求管理策略具备时空动态适应性。

基于以上认识,广泛分布于流域内的库塘与湿地,不应仅被视为静态的蓄水或净化单元,而应发挥好动态调蓄作用^[30]。旱季时,可逐步释放库塘存水以补充河道生态基流,增强水体自净能力,缓解静止导致的低 DO 风险,同时为雨季腾出调蓄容量。在雨季,则优先蓄滞前期降雨径流与农田退水,直接削减入河 Q 与污染峰荷。以罗时江片区为例,现有库

塘 133 座、总库容约为 152 万 m^3 、湿地为 62 处、总调蓄容量约为 65 万 m^3 ;该区域灌溉需水量为 896 万 m^3 、退水量为 805 万 m^3 ,基本平衡。全年地表降雨径流量约为 219 万 t ,主要集中在 6 月和 8 月,在未利用库塘时,降雨径流高峰混合农田退水溢流直接排入河道。通过现有库塘—湿地系统的互联与智能调度,有能力在 5 年一遇降雨下实现无溢流,此为策略的可行性提供了实证支持。

需注意的是,库塘调蓄需配合净化措施,降低 COD 排放量。库塘容量有限,在降雨量大的时期有快速排水需求,通过一定的处理方法处理排河污水,可有效降低入河污染通量^[31]。生态滤坝与滞留池组合工艺,可对库塘溢流污水进行快速净化,有效截留径流中的 COD,减少污染排入河道。研究^[32]表明,该类组合工艺可提升 COD 截留效率分别达 60%。大部分降雨径流与农业退水用于农业灌溉,少量通过组合工艺净化达标的排放入河,将从根本上改变污染水直排模式,构建出流域健康水循环新模式。

3 结论

通过对洱海北部流域 2011 年—2020 年水质水量监测数据分析,这项工作构建了 4 种预测 DO 的机器学习算法并进行了比较。模型对比结果表明,LightGBM 模型 R^2 较基准模型提高了 11.2%,在预测中展现最优性能 ($\text{RMSE} = 0.284 \text{ mg/L}$, $\text{MAE} =$

0.226 mg/L, $R^2 = 0.912$), 适用于水体 DO 的预测。基于 SHAP 的可解释分析, 进一步揭示了环境因子对 DO 的非线性影响及多因子间的协同机制。结果表明, Q 和 COD 是影响 DO 浓度的关键驱动因素, 维持适度 Q 和低含量 COD 有利于水体 DO 含量升高。该地区夏季多雨期与农业活动高峰期重叠, 流域农业面源污染负荷输入风险大。对此, 提出水量调控与污染治理相结合的管理策略, 包括建设库塘调蓄系统与生态滤坝等截污设施, 以有效削减洪峰 Q 与径流污染负荷, 维持水环境长期健康。本文从 DO 变化的驱动因素及水动力响应出发, 结合当地实际情况提出了因地制宜的“水量-水质”协同管理策略, 为高原河流水环境保护与管理提供参考。

参考文献

- [1] Doroudi S, Kheyrui Y, Sharafati A, et al. Enhancing prediction of dissolved oxygen over Santa Margarita River: Long short-term memory incorporated with multi-objective observer-teacher-learner optimization [J]. *Journal of Water Process Engineering*, 2025, 70: 106969. DOI: 10.1016/j.jwpe.2025.106969.
- [2] Roman M R, Brandt S B, Houde E D, et al. Interactive effects of hypoxia and temperature on coastal pelagic zooplankton and fish [J]. *Frontiers in Marine Science*, 2019, 6: 139. DOI: 10.3389/fmars.2019.00139.
- [3] Friedrich J, Janssen F, Aleynik D, et al. Investigating hypoxia in aquatic environments: Diverse approaches to addressing a complex phenomenon [J]. *Biogeosciences*, 2014, 11 (4): 1215-1259.
- [4] Liu H J, Cai F X, Huang Z Q, et al. Seasonal hydrological variation impacts nitrogen speciation and enhances bioavailability in plateau lake sediments [J]. *Water Research*, 2025, 271: 122990. DOI: 10.1016/j.watres.2024.122990.
- [5] 黎睿, 汤显强, 胡艳平, 等. 四湖总干渠溶解氧季节性异常特征与成因分析 [J]. *中国环境科学*, 2025, 45(5): 2816-2826.
Li R, Tang X Q, Hu Y P, et al. Spatial-temporal distribution characteristics of dissolved oxygen and its causes in Sihui Canal [J]. *China Environmental Science*, 2025, 45(5): 2816-2826.
- [6] Wei Q, Xue L Q, Liao S M, et al. Ecohydrological indicators and environmental flow assessment in the middle and lower reaches of the Huai River, China [J]. *Science of the Total Environment*, 2024, 940: 173639. DOI: 10.1016/j.scitotenv.2024.173639.
- [7] Merabet K, Di Nunno F, Granata F, et al. Predicting water quality variables using gradient boosting machine: Global versus local explainability using SHapley Additive Explanations (SHAP) [J]. *Earth Science Informatics*, 2025, 18(3): 298-298.
- [8] Wu J X, Yu X X. Numerical investigation of dissolved oxygen transportation through a coupled SWE and streeter-phelps model [J]. *Mathematical Problems in Engineering*, 2021, 2021: 6663696. DOI: 10.1155/2021/6663696.
- [9] Peng X X, Qian H H, Wang Y. On a spatiotemporal model for pollutants and dissolved oxygen in a river [J]. *Discrete and Continuous Dynamical Systems-Series B*, 2024, 29 (12): 4859-4873.
- [10] Shi P, Li G H, Yuan Y M, et al. Prediction of dissolved oxygen content in aquaculture using clustering-based softplus extreme learning machine [J]. *Computers and Electronics in Agriculture*, 2019, 157: 329-338. DOI: 10.1016/j.compag.2019.01.004.
- [11] Xu C, Luo P C, Wu P, et al. Detection of periodicity, aperiodicity, and corresponding driving factors of river dissolved oxygen based on high-frequency measurements [J]. *Journal of Hydrology*, 2022, 609: 127711. DOI: 10.1016/j.jhydrol.2022.127711.
- [12] 赵海超, 王圣瑞, 赵明, 等. 洱海水体溶解氧及其与环境因子的关系 [J]. *环境科学*, 2011, 32(7): 1952-1959.
Zhao H C, Wang S R, Zhao M, et al. Relationship between the DO and the environmental factors of the water body in Lake Erhai [J]. *Environmental Science*, 2011, 32(7): 1952-1959.
- [13] Singh R B, Olbert A I, Samantra A, et al. AI-driven modelling approaches for predicting oxygen levels in aquatic environments [J]. *Journal of Water Process Engineering*, 2024, 66: 105940. DOI: 10.1016/j.jwpe.2024.105940.
- [14] Cao W J, Huan J, Liu C, et al. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction [J]. *Aquacultural Engineering*, 2019, 84: 50-59. DOI: 10.1016/j.aquaeng.2018.12.003.
- [15] Xu C, Chen X H, Zhang L L. Predicting river dissolved oxygen time series based on stand-alone models and hybrid wavelet-based models [J]. *Journal of Environmental Management*, 2021, 295: 113085. DOI: 10.1016/j.jenvman.2021.113085.
- [16] Nong X Z, Lai C, Chen L H, et al. Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China [J]. *Ecological Indicators*, 2023, 146: 109845. DOI: 10.1016/j.ecolind.2022.109845.
- [17] Kundu S, Datta P, Pal P, et al. Unveiling the hidden connections: Using explainable artificial intelligence to assess water quality criteria in nine giant rivers [J]. *Journal of Cleaner Production*, 2025, 492: 144861. DOI: 10.1016/j.jclepro.2025.144861.
- [18] Lin S S, Shen S L, Zhou A N, et al. Sustainable development and environmental restoration in Lake Erhai, China [J]. *Journal of Cleaner Production*, 2020, 258: 120758. DOI: 10.1016/j.jclepro.2020.120758.

- jclepro. 2020. 120758.
- [19] 华兆晖, 李锐, 杨智, 等. 2017—2022 年洱海水体营养状态的时空变化趋势及其成因分析[J]. 湖泊科学, 2024, 36(6): 1639–1649.
- Hua Z H, Li R, Yang Z, et al. The spatiotemporal variation and potential causes for the nutrient status of water in Lake Erhai from 2017 to 2022 [J]. Journal of Lake Science, 2024, 36(6): 1639–1649.
- [20] Zhao Y B, Chen M. Prediction of river dissolved oxygen (DO) based on multi-source data and various machine learning coupling models [J]. PLoS One, 2025, 20(3): e0319256. DOI: 10.1371/journal.pone.0319256.
- [21] Lu H F, Ma X. Hybrid decision tree-based machine learning models for short-term water quality prediction[J]. Chemosphere, 2020, 249: 126169. DOI: 10.1016/j.chemosphere.2020.126169.
- [22] Singh R B, Olbert A I, Samantra A, et al. AI-driven modelling approaches for predicting oxygen levels in aquatic environments [J]. Journal of Water Process Engineering, 2024, 66: 105940. DOI: 10.1016/j.jwpe.2024.105940.
- [23] Shi X M, Li Y, Yao B, et al. Impact of high temporal resolution data on water quality modeling: Insights from Erhai case study [J]. Processes, 2025, 13(6): 1726–1726.
- [24] 于超, 储金宇, 白晓华, 等. 洱海入湖河流弥苴河下游氮磷季节性变化特征及主要影响因素[J]. 生态学报, 2011, 31(23): 7104–7111.
- Yu C, Chu J Y, Bai X H, et al. Seasonal variation of nitrogen and phosphorus in Miju River and Lake Erhai and influencing factors [J]. Acta Ecologica Sinica, 2011, 31(23): 7104–7111.
- [25] 项颂, 庞燕, 窦嘉顺, 等. 不同时空尺度下土地利用对洱海入湖河流水质的影响[J]. 生态学报, 2018, 38(3): 876–885.
- Xiang S, Pang Y, Dou J S, et al. Impact of land use on the water quality of inflow river to Erhai Lake at different temporal and spatial scales [J]. Acta Ecologica Sinica, 2018, 38(3): 876–885.
- [26] Cheng M J, Liu X, Xiao H, et al. Enhancing rainfall-runoff pollution modeling by incorporation of neglected physical processes [J]. Frontiers of Agricultural Science and Engineering, 2023, 10(4): 553–65. DOI: 10.15302/j-fase-2023519.
- [27] Xu Y T, Li P, Zhang M H, et al. Quantifying seasonal variations in pollution sources with machine learning-enhanced positive matrix factorization [J]. Ecological Indicators, 2024, 166: 112543. DOI: 10.1016/j.ecolind.2024.112543.
- [28] Henry C, Minier J P, Brambilla S. Particle resuspension: Challenges and perspectives for future models [J]. Physics Reports, 2023, 1007: 1–98. DOI: 10.1016/j.physrep.2022.12.005.
- [29] Zhi W, Ouyang W Y, Shen C P, et al. Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers [J]. Nature Water, 2023, 1(3): 249–260.
- [30] Tian Y S, Shen J, Feng J M, et al. Research advancements on agricultural non-point source pollution in major lake and reservoir watersheds of China: Status, sources, monitoring, and prospects [J]. Ecological Indicators, 2025, 178: 113981. DOI: 10.1016/j.ecolind.2025.113981.
- [31] Peng J Y, Hou Z Y, Yuan J, et al. The storm runoff management strategy based on agricultural ditch nutrient loss characteristics in Erhai Lake, China [J]. Journal of Contaminant Hydrology, 2024, 261: 104305. DOI: 10.1016/j.jconhyd.2024.104305.
- [32] Biswal B K, Vijayaraghavan K, Tsen-Tieng D L, et al. Biochar-based bioretention systems for removal of chemical and microbial pollutants from stormwater: A critical review [J]. Journal of Hazardous Materials, 2022, 422: 126886. DOI: 10.1016/j.jhazmat.2021.126886.

(上接第 13 页)

- [20] 朱秀芳, 李原. 基于孤立森林、支持向量机和空间邻接关系的温排水自动提取方法[J/OL]. 中国环境科学: 1–8 [2025–09–18]. <https://doi.org/10.19674/j.cnki.issn1000-6923.20230809.001>.
- Zhu X F, Li Y. An automatic extraction method for thermal discharge water based on isolation forest, support vector machine and spatial adjacency [J/OL]. China Environmental Science: 1–8 [2025–09–18]. <https://doi.org/10.19674/j.cnki.issn1000-6923.20230809.001>.
- [21] 田锦, 赵修良, 屈国普, 等. 核电站流出液 γ 核素在线监测系统优化设计 [J]. 核电子学与探测技术, 2025, 45(8): 1173–1178.
- Tian J, Zhao X L, Qu G P, et al. Optimization design of an online gamma nuclide detection system for nuclear power plant effluents [J]. Nuclear Electronics & Detection Technology, 2025, 45(8): 1173–1178.
- [22] Valkovic V. Radioactivity in the environment: Physicochemical aspects and applications [M]. 2rd ed. New York: Elsevier, 2000: 377–379.