

祝新明, 杨晓霞, 富英杰, 等. 基于机器学习的平原河网水源地叶绿素 a 含量预测[J]. 净水技术, 2026, 45(7): 163-172.

Zhu X M, Yang X X, Fu Y J, et al. Prediction of chlorophyll a content in water sources of plain river networks by machine learning[J]. Water Purification Technology, 2026, 45(7): 163-172.

基于机器学习的平原河网水源地叶绿素 a 含量预测

祝新明¹, 杨晓霞¹, 富英杰², 张颖龙¹, 马艳玲¹, 苏营营^{1,*}

(1. 浙江省嘉兴生态环境监测中心, 浙江嘉兴 314000; 2. 嘉兴市生态环境局, 浙江嘉兴 314000)

摘要 【目的】 叶绿素 a 的水平是评估水体藻类成长情况的重要指标, 利用机器学习算法开展水源地叶绿素 a 动态预测对于指导水源地水华早期预警和应急处置具有重要意义。【方法】 以 2020 年 9 月—2023 年 8 月长三角平原河网地区的千亩荡饮用水源地水质自动监测站高频监测数据为基础, 使用 Pearson 检验分析了叶绿素 a 含量与环境因子的相关性, 使用小波变化(WT)分别结合长短期记忆(LSTM)、门控循环单元(GRU)、卷积神经网络(CNN)和多层感知器(MLP)4 种机器学习算法构建叶绿素 a 的预测模型。【结果】 嘉兴市千亩荡水源地叶绿素 a 分布总体为春夏季高, 秋冬季低; 水体中水温、pH、浑浊度 3 项环境因子对该水源地藻类生长和叶绿素 a 含量变化存在重要关联性; 4 种基于时间序列的预测模型在识别叶绿素 a 的变化特征和预测效果上, WT-GRU 预测最佳, WT-LSTM 次之, WT-CNN 第三, WT-MLP 最差, WT-GRU 模型的一元线性回归拟合系数 R^2 为 0.995 5, 平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分比误差(MAPE)的值分别为 0.71、0.91 $\mu\text{g/L}$ 和 3.3%。【结论】 WT-GRU 模型在识别叶绿素 a 变化特征和预测精度上表现最优, 可有效应用于千亩荡饮用水源地叶绿素 a 动态预测及水华早期预警。

关键词 机器学习; 平原河网; 水源; 叶绿素 a; 预测模型

中图分类号: X832 文献标志码: A 文章编号: 1009-0177(2026)07-0163-10

DOI: 10.15890/j.cnki.jsjs.2026.07.019

Prediction of Chlorophyll a Content in Water Sources of Plain River Networks by Machine Learning

Zhu Xinming¹, Yang Xiaoxia¹, Fu Yingjie², Zhang Yinglong¹, Ma Yanling¹, Su Yingying^{1,*}

(1. Jiaxing Eco-Environmental Monitoring Center of Zhejiang Province, Jiaxing 314000, China;

2. Jiaxing Ecology and Environment Bureau, Jiaxing 314000, China)

Abstract [Objective] Chlorophyll a is an important index of the growth of algae in water bodies. The machine learning algorithms are used to carry out dynamic prediction of chlorophyll a in water source areas. It is of great significance for guiding early warning and emergency response of water blooms in water source areas. [Methods] This paper collected the monitoring data of Qianmu Dang drinking water source automatic monitoring station in the Yangtze River Delta plain river network area from September 2020 to August 2023. The relationship between chlorophyll a and environmental factors was explored by using Pearson's method. Wavelet transform (WT) combined with four machine learning algorithms, long short-term memory (LSTM), gated recurrent unit (GRU), convolutional neural network (CNN) and multilayer perceptron (MLP), were used to construct the prediction model of chlorophyll a. [Results] The content of chlorophyll a in the Qianmudang water source area in Jiaxing City was generally high in summer and

[收稿日期] 2024-05-03

[基金项目] 浙江省生态环境科研和成果推广项目(2022HT0044); 嘉兴市科技计划项目(2023AD31082); 长江生态环境保护修复联合研究二期项目(2022-LHYJ-02-0503-02)

[作者简介] 祝新明(1983—), 男, 工程师, 主要从事生态环境监测与评价等工作, E-mail: xinming_zhu@163.com。

[通信作者] 苏营营(1980—), 女, 高级工程师, 主要从事生态环境监测与评价等工作, E-mail: 512649763@qq.com。

spring, and low in winter and autumn. The three indices of water temperature, pH and turbidity had important correlation with the concentration of chlorophyll a in the water source. The four machine learning algorithms were obviously different in recognizing the change characteristics of chlorophyll a and predicting effect. WT-GRU had the best prediction effect. The prediction effect of WT-LSTM was the second. The prediction effect of WT-CNN was the third. The prediction effect of WT-MLP was the worst. The univariate linear regression fitting coefficient R^2 of the WT-GRU model was 0.995 5. The values of mean absolute error(MAE), root mean square error(RMSE), and mean absolute percentage error(MAPE) of the WT-GRU model were 0.71 $\mu\text{g/L}$, 0.91 $\mu\text{g/L}$, and 3.3%. [Conclusion] The WT-GRU model demonstrates superior performance in capturing chlorophyll a changing patterns and achieving high prediction accuracy. It is therefore highly suitable for dynamic chlorophyll a forecasting and early warning of algal blooms in the Qianmu Dang drinking water source area.

Keywords machine learning; plain river network; water source; chlorophyll a; prediction model

随着经济的快速增长和人口的稳步上升,江河、湖泊等水体受人类活动影响日渐显著。氮、磷等营养盐的过量排放,已成为众多水域出现富营养化问题的主要因素,不但导致水华发生频繁,更严重影响了水体质量,破坏了水生态平衡,进而增加了人类的健康风险^[1]。水体中叶绿素 a 反映浮游藻类的生长状况,也是评估水体营养状态的重要指标,与水生态环境关系密切^[2-3]。藻类水华的突发性给水体管理和治理带来了极大的挑战,因而准确预测水体中叶绿素 a 含量动态变化情况,提前研判水质的变化趋势,是快速应对潜在水华风险的有效手段之一^[4],对人类活动密切相关的饮用水水源地水华的预测更是具有十分重要的意义。

国内外学者已开展了大量关于叶绿素 a 预测的研究。机理型方法,如水质分析模拟程序(WASP)、土壤和水质评估工具(SWAT)、水环境与水生态模型(AQUATOX)等方法^[5-7],通过构建基于物理化学过程的数学模型来预测叶绿素 a 的变化,但在实际应用中受限于数据获取难和模型参数确定难,导致在不同水体中的预测适用性较差^[8]。传统的非机理型方法,如回归分析、时间序列分析等^[9-12],对长时序数据的需求和噪声的敏感性不强;遥感分析受气候和地形等因素影响,预测精度又不够稳定^[13]。近年来人工智能、机器学习和大数据等技术迭代升级和不断发展,基于深度学习神经网络的数据挖掘和预测分析方法为有效保护水资源提供一种可行的思路^[14]。Liang 等^[15]证明了使用长短期记忆(LSTM)网络在克服传统机理模型数据限制、计算量大和未来边界条件等限制可很好地模拟环境流体动力学水质模型的预测能力。Liao 等^[16]利用机器学习模型预测密云水库的叶绿素 a 含量,结果表明,使用随机

森林(RF)模型在数据缺失或异常时具有更高的预测精度、更稳定的结果、更少的过拟合和更强的预测能力。Huang 等^[17]利用机器学习方法对鄱阳湖叶绿素 a 开展预测,结果表明,支持向量机(SVR)模型相比其他 2 种模型具有更高的精度和准确度。任树顺等^[18]将 3 种时间序列模型用于九龙江河流库区藻华预测并开展比较分析,结果表明,LSTM 模型对总叶绿素 a 含量的短期预测优于其他 2 种模型。然而单一的机器学习模型的性能在很大程度上依赖于输入数据的质量和完整性,如数据存在偏差、不完整或存在错误,模型的预测准确性可能会受到影响^[19]。叶绿素 a 在线监测数据为非稳态数据,受观测误差、系统误差等因素影响,测试结果含有随机噪声,因此有必要对叶绿素 a 进行降噪处置。小波变换(WT)、小波分析等数据分解技术能够有效对数据降噪,从而提高机器学习数据输入的质量,已经与机器学习模型结合对水质开展预测并取得了比单一模型更好的预测效果^[20-21]。

嘉兴是典型的平原河网地区,位于水系末端,饮用水原水通过生态湿地或配套生态涵养等手段改善水质,但随着水体的流动性降低,饮用水水源地水体透明度提升,在有利气象条件和适宜的营养盐共同作用下,易引发藻类异常增殖,进而威胁水源地水质安全。因此选取嘉兴市某县级水源地——千亩荡,以水质自动监测站叶绿素 a 高频监测数据为对象,基于 WT 结合机器学习算法构建叶绿素 a 预测模型,分析不同模型对叶绿素 a 含量时序变化预测的效果,探讨最优模型,以期有效指导平原河网饮用水水源地水华风险的预警和防治,保障公众饮水安全。

1 试验材料和方法

1.1 研究区域概况

嘉兴市位于长江三角洲地区杭嘉湖平原腹地,

境内河流纵横交错,形成了密集的水系网络,受制于地理条件,饮用水原水均取自河道,近年来嘉兴市通过模仿自然的人工生态湿地等方式,改善原水的水质^[22-23]。本次选取的饮用水水源地位于嘉兴市某县级饮用水水源地,该水源地与周边联通的河道均有节制闸,水体流动性较差。

1.2 数据来源

水质监测数据来自布设在嘉兴市某县级水源地的千亩荡水站,水质监测指标包含叶绿素 a、水温、pH、溶解氧、浑浊度、氨氮、高锰酸盐指数、总氮、总磷等 11 项指标,监测频率为每 4 h 一次。数据选取 2020 年 9 月 1 日—2023 年 8 月 31 日共 3 年的日监测结果。

1.3 研究方法

1.3.1 皮尔逊(Pearson)相关系数

叶绿素 a 与其影响指标之间的关系是非线性的,建立包含所有影响指标相关模型是非必要且不切实际,因此通过 Pearson 相关性分析法识别影响叶绿素 a 含量的主要因素,为下一步数据建模做基础,Pearson 相关系数具体见参考文献^[24]。

1.3.2 WT

为优化原始数据以提高预测模型的准确性,实施了 WT 预处理,通过使用 Daubechies 小波族中的 db4 小波函数,对选定的监测数据进行了 3 层分解,滤除高频噪声,处理后的数据保留了反映长期变化趋势的低频成分^[25]。

1.3.3 LSTM 网络模型

LSTM 模型于 1997 年提出用于解决 RNN 模型梯度消失的问题^[26]。LSTM 的关键特点是它的单元结构(门控机制),包括 3 个主要部分:输入门、遗忘门和输出门,输入门决定将哪些新的信息添加进单元状态,遗忘门决定将在单元状态中保留或遗忘哪些旧的信息,输出门决定当前单元状态的输出信息^[27]。

具体如式(1)~式(6)。

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$\bar{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t C_{t-1} + i_t \bar{C}_t \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \tanh(C_t) \quad (6)$$

其中: f_t ——遗忘门;

i_t ——输入门;

o_t ——输出门;

W_f ——对应遗忘门权重矩阵;

W_i ——对应输入门权重矩阵;

W_c ——对应候选细胞状态权重矩阵;

W_o ——对应输出门权重矩阵;

b_f ——对应遗忘门偏差矩阵;

b_i ——对应输入门偏差矩阵;

b_c ——对应的候选细胞状态偏差矩阵;

b_o ——对应的输出门偏差矩阵;

x_t —— t 时刻的输入向量;

C_t —— t 时刻的候选向量;

C_{t-1} —— $t-1$ 时刻的候选向量;

$\bar{C}_t$ —— t 时刻的更新向量;

h_t —— t 时刻模型的输出向量;

h_{t-1} —— $t-1$ 时刻模型的输出向量;

σ ——sigmoid 激活函数;

$\tanh$ ——双曲正切函数。

1.3.4 门控循环单元(GRU)模型

GRU 模型被认为是 LSTM 模型的一种高效改进形式,其结构相对简单,内部参数较少,能够提高计算效率。GRU 模型主要包括重置门和更新门,这 2 个门控机制控制着信息的流动和状态的更新,重置门控制着记忆信息与新的输入信息的结合方式,更新门控制着当前状态中之前状态信息的输入程度^[28]。

1.3.5 卷积神经网络(CNN)模型

CNN 模型通常包含卷积层、激活层、池化层、全连接层等多层网络,其通过核心的卷积操作来提取输入数据的特征,进而提高模型的性能和泛化能力,是一种经典的深度学习模型^[29]。

1.3.6 多层感知器(MLP)模型

MLP 模型是一种前馈神经网络,它由 3 个或更多的层组成:输入层、一个或多个隐藏层及输出层。每一层由多个神经元或节点组成,而不同层之间的神经元通过权重连接相互影响。MLP 模型能够学习和模拟复杂的非线性关系,因此广泛应用于分类、回归和特征学习等任务^[30]。

1.4 模型评价

不同预测模型预测结果的精度比较应用平均绝

对误差(MAE)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)等3个指标^[31]。各指标的具体如式(7)~式(9)。

$$M_{AE} = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (7)$$

$$R_{MSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (8)$$

$$M_{APE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

其中: M_{AE} ——MAE, $\mu\text{g/L}$;
 R_{MSE} ——RMSE, $\mu\text{g/L}$;

M_{APE} ——MAPE;
 $\hat{y}_i$ ——预测值, $\mu\text{g/L}$;
 y_i ——实际值, $\mu\text{g/L}$;
 n ——样本个数。

2 结果与讨论

2.1 叶绿素 a 季节变化特征及与环境因子的关系

千亩荡饮用水源叶绿素 a 的质量浓度变化如图 1 所示,叶绿素 a 质量浓度为 3.4~34.2 $\mu\text{g/L}$,平均质量浓度为 13.9 $\mu\text{g/L}$ 。叶绿素 a 季节性变化明显,以 2022 年为例,春季、夏季、秋季、冬季叶绿素 a 的质量浓度分别为 16.0、21.3、6.1、7.1 $\mu\text{g/L}$,呈现出夏季>春季>秋季>冬季,这与汪婷婷等^[32]在香溪河库湾不同季节叶绿素 a 质量浓度变化规律基本一致。

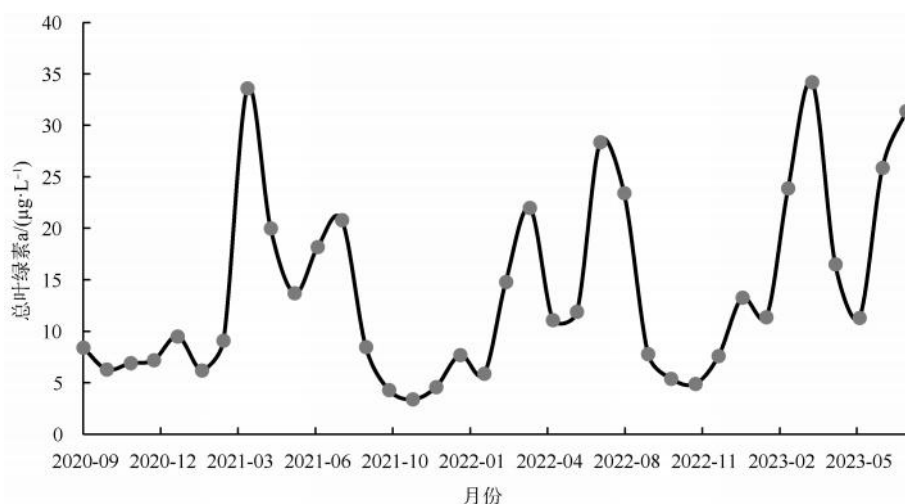


图 1 千亩荡水源地叶绿素 a 质量浓度变化特征

Fig. 1 Variation of Characteristics of Chlorophyll a Mass Concentration in Qianmu Dang Water Source

叶绿素 a 与相关监测指标的 Pearson 相关性分析结果如表 1 所示。千亩荡水源地水体中叶绿素 a 与 pH、水温均呈显著正相关,相关系数分别为 0.362 和 0.406,表明 pH 和水温与叶绿素 a 存在重要的关联。藻类的光合作用吸收了水体中的二氧化碳,释放了氧气,进而增加了水体的 pH^[33]。水温的季节性变化对藻类群落结构的演替有明显的影响,间接或直接影响叶绿素 a 质量浓度的变化^[34]。千亩荡水源地水体中叶绿素 a 与浑浊度呈显著负相关,相关系数为-0.466。较高水体浑浊度会增加光的散射和吸收,这可能会影响浮游植物初级生产,进而导致叶绿素 a 质量浓度的变化^[35]。千亩荡水源地水体中叶绿素 a 与溶解氧和营养盐均无显著性关系。冬季溶解氧含量高而叶绿素 a 含量低,夏季溶

表 1 叶绿素 a 与环境因子的相关性分析

Tab. 1 Correlation Analysis between Chlorophyll a and Environmental Factors

指标	相关系数
pH	0.362 **
水温	0.406 **
溶解氧	0.052
浑浊度	-0.466 **
高锰酸盐指数	0.059
氨氮	-0.172
总磷	-0.280
总氮	0.186

注:**表示相关性在 $P<0.01$ 层上显著(双尾)

解氧含量低而叶绿素 a 含量高,这种季节性变化可

能干扰两者之间长期关联性的观察。气温升高导致叶绿素 a 质量浓度增加,而藻类生长消耗了大量氮、磷营养元素,导致其质量浓度下降,这 2 种相反的变化趋势可能削弱了氮、磷质量浓度与叶绿素 a 质量浓度之间的相关性^[36]。

2.2 不同模型预测效果比较分析

2.2.1 模型构建

本文采用了 Tensorflow 的 Keras 平台,通过以下步骤构建并校正了 4 种机器学习模型,以预测叶绿素 a 的变化:(1)数据预处理,对原始数据进行归一化处理,确保不同监测指标数值范围一致性;(2)特征选择,基于叶绿素 a 与其他监测指标的相关性,选择 pH、水温、浑浊度和叶绿素 a 作为输入特征;(3)数据集划分,将 2020 年 9 月 1 日—2023 年 8 月 31 日相关指标的监测数据集划分为训练集(前 80%)和测试集(后 20%),为模型训练和评估提供基础;(4)模型建立,构建 CNN、MLP、LSTM、GRU 4 种模型,每种模型都针对时间序列数据的特点进行了特别设计;(5)模型训练,使用 Adam 优化器,设置学习率为 0.001,通过随机网格搜索确定时间窗口为 32,样本批量大小为 16,隐藏层设计为 2 个,分别包含 32 个和 16 个神经元,进行 50 次迭代训练;(6)模型校正,采用早停法避免过拟合,并优化隐藏层神经元数量和网络结构,提高模型泛化能力;(7)性能评估,在测试集上进行预测,通过反归一化处理恢复原始数值范围,并使用 MSE、 R^2 等指标评估模型性能。

2.2.2 WT 前后模型预测性能的评价结果

WT 前后 4 种模型对叶绿素 a 预测评价结果如表 2 所示,WT 后 4 种模型预测性能均有提高,其中 GRU 预测准确度提升最为显著,LSTM 和 CNN 次之,MLP 的提升效果最差。LSTM 的 MAE、RMSE 和 MAPE 三项评价指标,分别比 WT 前提升 57.6%、61.3%和 9.7 百分点,GRU 的 MAE、RMSE 和 MAPE 三项评价指标,分别比 WT 前提升 75.4%、78.9%和 9.8 百分点,CNN 的 MAE、RMSE 和 MAPE 三项评价指标,分别比 WT 前提升 38.9%、45.3%和 7.0 百分点,MLP 的 MAE、RMSE 和 MAPE 三项评价指标,分别比 WT 前提升 10.8%、11.4%和 6.8 百分点。这种改进主要归功于 WT 预处理有效地去除了数据中的高频噪声,从而提高了模型对水质变化趋势的捕捉能力^[25]。

表 2 4 种模型叶绿素 a 预测评价结果

Tab. 2 Evaluation Results of Four Models for Chlorophyll a Prediction

模型	MAE/($\mu\text{g}\cdot\text{L}^{-1}$)	RMSE/($\mu\text{g}\cdot\text{L}^{-1}$)	MAPE
LSTM	3.28	4.52	16.8%
GRU	2.89	4.31	13.1%
CNN	4.50	6.14	20.3%
MLP	4.74	6.43	23.3%
WT-LSTM	1.39	1.75	7.1%
WT-GRU	0.71	0.91	3.3%
WT-CNN	2.75	3.36	13.3%
WT-MLP	4.23	5.70	16.5%

WT 后的 4 种模型中,WT-GRU 模型在预测叶绿素 a 的预测结果的评价指标上均表现出最佳性能,MAE、RMSE 和 MAPE 值分别为 0.71、0.91 $\mu\text{g}/\text{L}$ 和 3.3%,比 WT-LSTM 模型分别降低 48.9%、48.0%和 3.8 百分点,表明其在捕捉时间序列数据的复杂模式方面具有良好的优势^[28]。WT-LSTM 模型在 3 项预测结果的评价指标上表现出预测性能次之,MAE、RMSE 和 MAPE 值分别为 1.39、0.91 $\mu\text{g}/\text{L}$ 和 3.3%,比 WT-CNN 模型分别降低 49.5%、47.9%和 6.2 百分点,显示出良好的预测能力。WT-CNN 模型在本文中的预测性能居中,MAE、RMSE 和 MAPE 值分别为 2.75、3.36 $\mu\text{g}/\text{L}$ 和 13.3%,可能是由于其卷积层结构对水质信息的时间顺序并不敏感,不能很好地完成预测任务^[37]。WT-MLP 模型在本文中的预测性能最差,MAE、RMSE 和 MAPE 值分别为 2.75、3.36 $\mu\text{g}/\text{L}$ 和 13.3%,MLP 模型不具备循环结构的记忆功能,处理非平稳性问题困难,特征提取能力有限^[38],而叶绿素 a 质量浓度随时间的非稳态变化,导致其在预测结果中表现不佳。

2.2.3 模型预测值与实际值一致性

4 种机器学习模型测试集中预测值与实际值的对比如图 2(散点)和图 3(折线)所示,由图 2 可知,WT-GRU 在散点图中的点非常紧密地围绕对角线分布,显示出最小的预测误差,WT-LSTM 在散点图中的点相对集中在对角线附近,但比 WT-GRU 稍有分散,WT-CNN 在散点图中的点显示出一定的分散性,表明预测值与实际值之间的偏差略大,WT-MLP 在散点图中的点表现出较大的分散,这可能意味着模型在预测上存在较大的不确定性或偏差。WT-

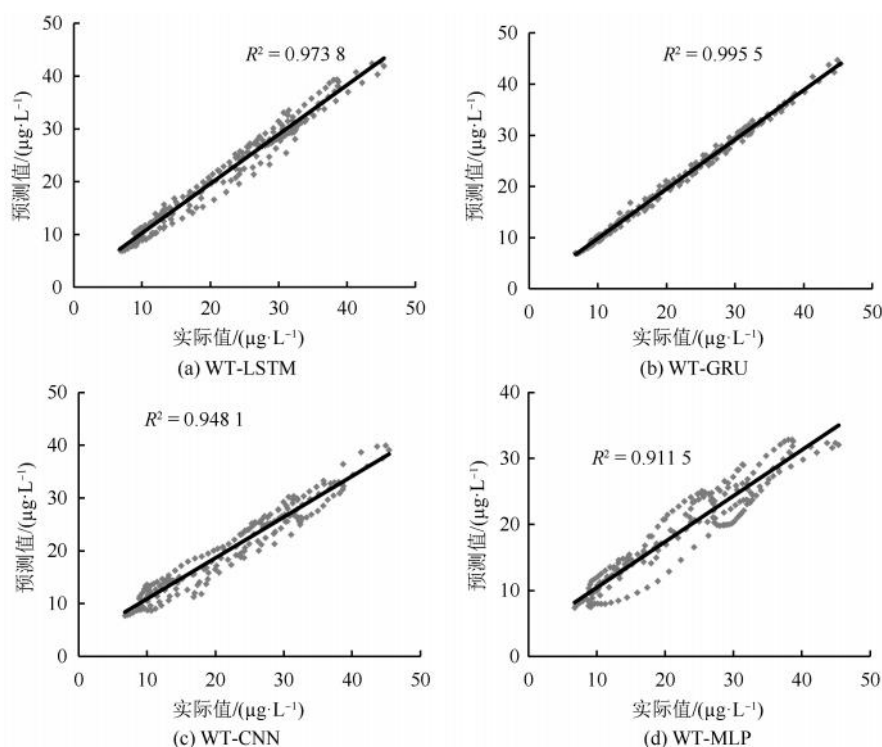


图 2 4 种机器学习模型测试集回归预测拟合

Fig. 2 Regression Prediction Fitting of Test Set of Four Machine Learning Models

GRU、WT-LSTM、WT-CNN、WT-MLP 的相关系数 R^2 值分别为 0.995 5、0.973 8、0.948 1、0.911 5, 预测值与实际值的拟合结果为 WT-GRU>WT-LSTM>WT-CNN>WT-MLP。这一结果表明, WT-GRU 模型能够有效地处理时间序列数据中的长期依赖问题, 并且能够适应数据中的非线性和复杂模式, 这与 WT-GRU 模型在叶绿素 a 的预测结果的 3 项评价指标上均表现出最佳性能一致。WT-LSTM 模型通过引入门控机制, 能有效地捕捉时间序列数据中的长期依赖关系, 这使得它在处理具有时间连续性的数据时同样表现出色。WT-CNN 模型通过卷积层能够捕捉局部特征, 但在处理时间序列数据时, 可能无法充分利用数据的全局信息和长期依赖性。相比之下, WT-MLP 模型 R^2 值最小, 预测值与实际值相对松散, 这意味着模型的预测结果还存在较大的误差。MLP 缺乏对时间序列数据中的记忆和序列依赖性的捕捉能力, 无法有效处理非稳态数据^[39], 在处理如水源地叶绿素 a 质量浓度变化这样的时间序列问题时, MLP 模型的这一局限性可能导致其无法准确预测其未来的质量浓度变化, 从而影响到对水源地藻类水华等环境问题的及时响应和处置。

由图 3 可知, WT-GRU 预测值与实际值的吻合程度最好, WT-LSTM 次之, WT-CNN 和 WT-MLP 较差。特别是在高值区和数据突变情况下, WT-CNN 和 WT-MLP 难以准确捕捉叶绿素 a 质量浓度的快速变化, 整体预测值较为平缓, 无法反映实际的变化趋势; WT-LSTM 在某些时刻预测值低于实际值, 显示出一定的局限性, 相比之下 WT-GRU 模型的预测结果要明显好于其他 3 个模型, 这可能是因为 GRU 模型独特的门控机制能够更好地学习数据中的长期和短期模式, 在处理具有复杂时间依赖性的数据时, 能够提供更为精确和适应性强的预测结果。特别是在数据突变的情况下, GRU 模型能够迅速适应新的模式, 而不会出现预测滞后或偏差。

季节性差分自回归移动模型(SARIMA)在预测鄱阳湖水体中的叶绿素 a 质量浓度时表现出较高的准确度, 但预测值与实测值之间在某些特定月份仍然存在较大差距, 表现出一定的不足^[40], 而 WT-GRU 模型的优势在于其能够处理时间序列数据的非线性特征, 这在处理复杂的水环境系统时尤为重要。支持向量回归模型通过与 RF 递归特征消除方法结合, 显著提高了水体中叶绿素 a 预测精度, 相关

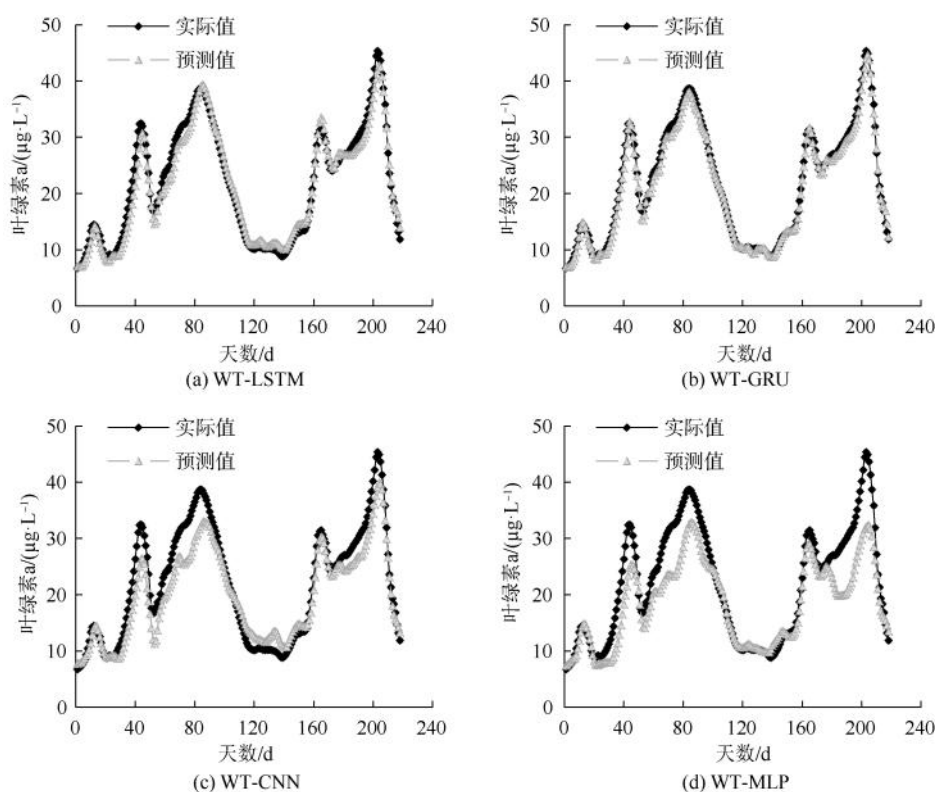


图3 4种机器学习模型预测值与实际值对比

Fig. 3 Comparison between Predicted and Actual Values of Four Machine Learning Models

系数达到 0.94^[17],但低于 WT-GRU 模型的相关系数,表明由于 WT-GRU 深度学习的特性,在泛化能力和适应性方面表现更好。基于 NARX 神经网络短期预测千岛湖叶绿素 a 含量,相关系数保持在 0.8~0.9, RMSE 在 3.87~5.48 $\mu\text{g/L}$ ^[12]。相比之下,WT-GRU 模型表现出更低的 RMSE 和更高的相关系数,这表明 WT-GRU 模型在捕捉时间序列数据的动态变化方面可能更为敏感和准确。利用 LSTM 模型在预测三峡水库叶绿素 a 质量浓度变化时, RMSE 为 0.049~0.221 $\mu\text{g/L}$ ^[25]。相比之下 WT-GRU 模型的 RMSE 稍高,但整体趋势预测得非常准确。本文的 WT-GRU 模型利用 WT 对数据进行预处理,显著提升了对叶绿素 a 质量浓度预测的准确性。WT 通过分解信号,有效地提取了关键的时间尺度特征,这对于时间序列分析至关重要。GRU 模型的门控单元能够自适应地调节信息流,增强了对长期依赖性的捕捉能力,相较于 LSTM,GRU 通过简化门控结构,减少了参数量,从而降低了过拟合风险,提升了模型的泛化性。这些改进使得 WT-GRU 在预测叶绿素 a 含量方面具有显著优势。因此 WT-

GRU 模型由于其深度学习架构,能够更好地处理复杂的非线性关系和长期依赖性,在平原河网水源地利用 WT-GRU 模型开展藻类的预测,提升水华的早期预警是一种较好的选择。

3 结论

(1)嘉兴市千亩荡水源地叶绿素 a 质量浓度呈季节性变化,夏春季高,秋冬季低;叶绿素 a 质量浓度与 pH 和水温呈显著正相关,与浑浊度呈显著负相关,表明这些环境因子与该水源地藻类生长和叶绿素 a 质量浓度变化存在重要关联性。

(2)WT 预处理的引入显著提高了各模型预测叶绿素 a 的性能,WT-GRU 模型的一元线性回归拟合系数 R^2 为 0.995 5, MAE、RMSE 和 MAPE 分别为 0.71、0.91 $\mu\text{g/L}$ 和 3.3%,在千亩荡水源地叶绿素 a 的预测中明显优于其他 3 种模型,显示出其在捕获时间序列数据的长期依赖和复杂模式方面的显著优势。

参考文献

[1] 苏玉萍,赖寿辉,林佳,等.富营养化饮用水源地山仔水库

- 限制性营养元素研究[J]. 环境科学学报, 2015, 35(10): 3107-3113.
- Su Y P, Lai S H, Lin J, et al. Research on the limiting nutrient in Shanzi Reservoir, an eutrophied drinking water source[J]. *Acta Scientiae Circumstantiae*, 2015, 35(10): 3107-3113.
- [2] 姜庆宏, 潘彤, 李卫平, 等. 南海湖营养盐季节特征及其与叶绿素 a 的相关性研究[J]. 水生态学杂志, 2020, 41(6): 74-80.
- Jiang Q H, Pan T, Li W P, et al. Nutrients in Nanhai Lake; Seasonal variation and correlation with chlorophyll-a[J]. *Journal of Hydroecology*, 2020, 41(6): 74-80.
- [3] 邹伟, 朱广伟, 蔡永久, 等. 综合营养状态指数(TLI)在夏季长江中下游湖库评价中的局限及改进意见[J]. 湖泊科学, 2020, 32(1): 36-47.
- Zou W, Zhu G W, Cai Y J, et al. The limitations of comprehensive trophic level index (TLI) in the eutrophication assessment of lakes along the middle and lower reaches of the Yangtze River during summer season and recommendation for its improvement[J]. *Journal of Lake Sciences*, 2020, 32(1): 36-47.
- [4] Ding W X, Zhang C Y, Shang S P, et al. Optimization of deep learning model for coastal chlorophyll a dynamic forecast[J]. *Ecological Modelling*, 2022, 467: 109913. DOI: 10.1016/j.ecolmodel.2022.109913.
- [5] Mbuhi M J, Mbiri R, Wendi C. Water quality modeling and sensitivity analysis using Water Quality Analysis Simulation Program (WASP) in the Shenandoah River watershed[J]. *Physical Geography*, 2019, 40(2): 127-148.
- [6] Orlińska-Wóznia P. Modeling chlorophyll a with use of the SWAT tool for the nielba river (west-central Poland) as an example of an unmonitored watercourse[J]. *Water*, 2022, 14(10): 1528.
- [7] 李建茹, 李兴, 李卫平, 等. 基于 AQUATOX 模型的乌梁素海富营养化模拟及控制研究[J]. 生态环境学报, 2020, 29(6): 1215-1224.
- Li J R, Li X, Li W P, et al. Simulation and control of eutrophication in Wuliangsu Lake using aquatox model[J]. *Ecology and Environment Sciences*, 2020, 29(6): 1215-1224.
- [8] Cen H B, Jiang J H, Han G Q, et al. Applying deep learning in the prediction of chlorophyll-a in the East China Sea[J]. *Remote Sensing*, 2022, 14(21): 5461.
- [9] 李娜, 李勇, 冯家成, 等. 太湖水体 Chl-a 预测模型 ARIMA 的构建及应用优化[J]. 环境科学, 2021, 42(5): 2223-2231.
- Li N, Li Y, Feng J C, et al. Construction and application optimization of the chl-a forecast model ARIMA for Lake Taihu[J]. *Environmental Science*, 2021, 42(5): 2223-2231.
- [10] Eze E, Kirby S, Attridge J, et al. Time series chlorophyll-a concentration data analysis: A novel forecasting model for aquaculture industry [C]//Halifax: The 7th International Conference on Time Series and Forecasting, 2021: 27.
- [11] Papenfus M, Schaeffer B, Pollard A I, et al. Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs[J]. *Environmental Monitoring and Assessment*, 2020, 192(12): 808.
- [12] 李未, 朱梦圆, 王裕成, 等. 基于 NARX 神经网络的千岛湖藻类短期预测模型构建[J]. 环境科学研究, 2022, 35(4): 918-925.
- Li W, Zhu M Y, Wang Y C, et al. Short-term forecasting model for algae based on NARX neural network in Qiandaohu Reservoir[J]. *Research of Environmental Sciences*, 2022, 35(4): 918-925.
- [13] Maier P M, Keller S. Estimating chlorophyll a concentrations of several inland waters with hyperspectral data and machine learning models[J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, IV-2/W5: 609-614. DOI: 10.5194/isprs-annals-IV-2-W5-609-2019.
- [14] Zhu M Y, Wang J W, Yang X, et al. A review of the application of machine learning in water quality evaluation[J]. *Eco-Environment & Health*, 2022, 1(2): 107-116.
- [15] Liang Z Y, Zou R, Chen X, et al. Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach[J]. *Journal of Hydrology*, 2020, 581: 124432. DOI: 10.1016/j.jhydrol.2019.124432.
- [16] Liao Z M, Zang N, Wang X, et al. Machine learning-based prediction of chlorophyll-a variations in receiving reservoir of world's largest water transfer project—a case study in the Miyun Reservoir, North China[J]. *Water*, 2021, 13(17): 2406.
- [17] Huang H D, Zhang J. Prediction of chlorophyll a and risk assessment of water blooms in Poyang Lake based on a machine learning method[J]. *Environmental Pollution*, 2024, 347: 123501. DOI: 10.1016/j.envpol.2024.123501.
- [18] 任树顺, 高萌, 王煦雯, 等. 基于 3 种时间序列模型的九龙江河流库区藻华预测[J]. 环境科学学报, 2022, 42(11): 172-183.
- Ren S S, Gao M, Wang X W, et al. Algal bloom prediction in the Jiulong River Reservoir based on three types of time series models[J]. *Acta Scientiae Circumstantiae*, 2022, 42(11): 172-183.
- [19] Yan X H, Zhang T Q, Du W Y, et al. A comprehensive review of machine learning for water quality prediction over the past five years[J]. *Journal of Marine Science and Engineering*, 2024, 12(1): 159.
- [20] Lin J N, Liu Q Q, Song Y, et al. Temporal prediction of coastal water quality based on environmental factors with machine learning[J]. *Journal of Marine Science and Engineering*, 2023, 11(8): 1608.

- [21] Wu J H, Wang Z C. A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory[J]. Water, 2022, 14(4): 610.
- [22] 魏捷, 朱红, 沈利清, 等. 构筑生态湿地改善桐乡市西部水源水质探讨[J]. 供水技术, 2018, 12(2): 34-39.
Wei J, Zhu H, Shen L Q, et al. Discussion on improvement of water source quality in western Tongxiang City by constructing the ecological wetland[J]. Water Technology, 2018, 12(2): 34-39.
- [23] 苏营营, 胡文凌, 叶朝霞, 等. 平原河网型水源地水质改善的实践和思考——以嘉兴市石臼漾人工生态湿地工程为例[J]. 环境工程, 2013, 31(S1): 216-219.
Su Y Y, Hu W L, Ye (C/Z)X, et al. Practice and thinking of water quality improvement with drinking water source in plain river network—An Example of Shijiuyang Artificial Ecosystem Wetland Project in Jiaying City[J]. Environmental Engineering, 2013, 31(S1): 216-219.
- [24] 胡添翼, 杨光, 陈波, 等. 基于 Pearson 相关性检验的 ARIMA 边坡位移监测模型[J]. 水利水电技术, 2016, 47(1): 71-75.
Hu T Y, Yang G, Chen B, et al. Pearson correlation test-based ARIMA model of displacement prediction[J]. Water Resources and Hydropower Engineering, 2016, 47(1): 71-75.
- [25] 欧阳添, 闪锐, 周博天, 等. 基于 LSTM 网络的在线藻类时序数据预测研究: 以三峡水库为例[J]. 湖泊科学, 2021, 33(4): 1031-1042.
Ouyang T, Shan K, Zhou B T, et al. Research on the online forecasting of algal kinetics based on time-series data and LSTM neural network: Taking Three Gorges Reservoir as an example[J]. Journal of Lake Sciences, 2021, 33(4): 1031-1042.
- [26] 翟梦梦, 王旭春, 任浩, 等. 基于 Python 的 LSTM 模型对流感预测的研究[J]. 中国卫生统计, 2022, 39(2): 162-166, 171.
Zhai M M, Wang X C, Ren H, et al. Study on LSTM model based on Python in prediction of influenza[J]. Chinese Journal of Health Statistics, 2022, 39(2): 162-166, 171.
- [27] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [28] 鲁言波, 陈湛峰, 李晓芳. 基于粒子群优化的 GRU 广东省跨境断面水质预测模型研究[J]. 生态环境学报, 2023, 32(9): 1673-1681.
Lu Y B, Chen Z F, Li X F. A study on water quality prediction model of cross-boundary sections in Guangdong Province based on GRU improved with particle swarm optimization[J]. Ecology and Environment Sciences, 2023, 32(9): 1673-1681.
- [29] 吴鑫俊, 赵晓东, 丁茜, 等. 基于数据驱动的 CNN 洪水演进预测方法[J]. 水力发电学报, 2021, 40(5): 79-86.
Wu X J, Zhao X D, Ding X, et al. CNN flood routing method based on data-driven training[J]. Journal of Hydroelectric Engineering, 2021, 40(5): 79-86.
- [30] 刘亚鑫, 贾建和, 李洪波, 等. PSO-MLP 模型预测降雨对府河氨氮的影响[J]. 科学技术与工程, 2022, 22(32): 14511-14517.
Liu Y X, Jia J H, Li H B, et al. PSO-MLP model predicts the effect of rainfall on ammonia nitrogen in Fuhe River[J]. Science Technology and Engineering, 2022, 22(32): 14511-14517.
- [31] 陈仟. 武汉市空气质量指数影响因素分析与预测[D]. 武汉: 中南财经政法大学, 2020.
Chen Q. Analysis of influencing factors and forecast of air quality index in Wuhan[D]. Wuhan: Zhongnan University of Economics and Law, 2020.
- [32] 汪婷婷, 杨正健, 刘德富. 香溪河库湾不同季节叶绿素 a 浓度影响因子分析[J]. 水生态学杂志, 2018, 39(3): 14-21.
Wang T T, Yang Z J, Liu D F. Seasonal distribution of chlorophyll-a in Xiangxi Bay of Three Gorges Reservoir and relationship to environment factors[J]. Journal of Hydroecology, 2018, 39(3): 14-21.
- [33] 罗宜富, 李磊, 李秋华, 等. 阿哈水库叶绿素 a 时空分布特征及其与藻类、环境因子的关系[J]. 环境科学, 2017, 38(10): 4151-4159.
Luo Y F, Li L, Li Q H, et al. Spatial and temporal distribution of chlorophyll a and its relationship to algae and environmental factors in Aha reservoir[J]. Environmental Science, 2017, 38(10): 4151-4159.
- [34] 贡丹丹, 刘德富, 张佳磊, 等. 周期性的温度扰动对藻类群落结构演替的影响[J]. 环境科学, 2016, 37(6): 2149-2157.
Gong D D, Liu D F, Zhang J L, et al. Influence of periodic temperature disturbance on the succession of algal community structure[J]. Environmental Science, 2016, 37(6): 2149-2157.
- [35] 李嘉皓, 田波, 曹芳, 等. 上海城市水体富营养化关键参量的遥感反演与时序分析[J]. 华东师范大学学报(自然科学版), 2022(1): 135-147.
Li J H, Tian B, Cao F, et al. Remote sensing inversion and time-series analysis of critical parameters for eutrophication assessment of urban waters in Shanghai[J]. Journal of East China Normal University (Natural Science), 2022(1): 135-147.
- [36] 祝新明, 张怡平, 李莉, 等. 嘉兴饮用水源地叶绿素 a 与环境因子的关系分析[J]. 四川环境, 2022, 41(6): 178-184.
Zhu X M, Zhang Y P, Li L, et al. Analysis on the relationship between chlorophyll-a and environmental factors in drinking water sources of Jiaxing[J]. Sichuan Environment, 2022, 41(6): 178-184.
- [37] 王军, 高梓勋, 朱永明. 基于 CNN-LSTM 模型的黄河水质预测研究[J]. 人民黄河, 2021, 43(5): 96-99, 109.
Wang J, Gao Z X, Zhu Y M. Research on Yellow River water

- quality prediction based on CNN-LSTM model [J]. *Yellow River*, 2021, 43(5): 96–99, 109.
- [38] 赵博. 深度异步残差网络及在路网交通流预测中的应用[D]. 北京: 北京交通大学, 2018.
- Zhao B. An asynch-res deep learning model and its application to road network traffic flow prediction [D]. Beijing: Beijing Jiaotong University, 2018.
- [39] 毛远宏, 孙琛琛, 徐鲁豫, 等. 基于深度学习的时间序列预测方法综述[J]. *微电子学与计算机*, 2023, 40(4): 8–17.
- Mao Y H, Sun C C, Xu L Y, et al. A survey of time series forecasting methods based on deep learning[J]. *Microelectronics & Computer*, 2023, 40(4): 8–17.
- [40] 钱春龙, 曾一川, 袁伟皓, 等. 基于时间序列的鄱阳湖 Chl-a 预测模型优化构建[J]. *长江科学院院报*, 2023, 40(10): 14–21.
- Qian C L, Zeng Y C, Yuan W H, et al. Optimal construction of chl-a prediction model for Poyang Lake based on time series[J]. *Journal of Changjiang River Scientific Research Institute*, 2023, 40(10): 14–21.
-
- (上接第 77 页)
- [12] Zhang X, Zhang Y, Shi P, et al. The deep challenge of nitrate pollution in river water of China [J]. *Science of the Total Environment*, 2021, 770: 144674. DOI: 10.1016/j.scitotenv.2020.144674.
- [13] Zhang R Y, Zhu L, Liu B Y, et al. Uncoupled nitrification-denitrification reduces nitrous oxide emissions in canals affected by ship disturbance[J]. *Water Research*, 2026, 289: 124927. DOI: 10.1016/j.watres.2025.124927.
- [14] 王水, 屈健, 李冰, 等. 太湖流域典型区域纺织行业污染特征分析[J]. *环境科技*, 2012, 25(5): 5–7.
- Wang S, Qu J, Li B, et al. Analysis on pollution characteristics of the textile industry in typical region of Taihu Lake Basin[J]. *Environmental Science and Technology*, 2012, 25(5): 5–7.
- [15] 牛小丹. 基于上下游合作的跨界河流水环境保护研究[D]. 上海: 华东师范大学, 2018.
- Niu X D. Study on transboundary river environment protection based on the cooperation of the upstream and downstream[D]. Shanghai: East China Normal University, 2018.
- [16] 朱永杰, 苏翔, 徐晓黎, 等. 太浦河丰水期上下游水环境现状分析及水质评价[J]. *给水排水*, 2024, 50(s1): 191–195.
- Zhu Y J, Su X, Xu X L, et al. Analysis and water quality assessment of water environment of Taipu river basin during high water period[J]. *Water & Wastewater Engineering*, 2024, 50(s1): 191–195.
- [17] 殷燕, 张运林, 时志强, 等. 太湖化学耗氧量和生化需氧量的时空分布特征[J]. *环境科学学报*, 2010, 30(12): 2544–2552.
- Yin Y, Zhang Y L, Shi Z Q, et al. Temporal-spatial variations of chemical oxygen demand and biochemical oxygen demand in Lake Taihu [J]. *Acta Scientiae Circumstantiae*, 2010, 30(12): 2544–2552.
- [18] Tang S Y, Zhang Z H, Zhang X H. Coupling *in-situ* ozonation with ferric chloride addition for ceramic ultrafiltration membrane fouling mitigation in wastewater treatment: Quantitative fouling analysis[J]. *Journal of Membrane Science*, 2018, 555: 307–317. DOI: 10.1016/j.memsci.2018.03.061.
- [19] Avalon N E, Reis M A, Thornburg C C, et al. Leptochelins A-C, cytotoxic metallophores produced by geographically dispersed *Leptothoe* strains of marine cyanobacteria [J]. *Journal of the American Chemical Society*, 2024, 146(27): 18626–18638.
- [20] 任俊宏. 太湖藻型和草型湖区 2-MIB 时空分布特征及驱动因子研究[D]; 无锡: 江南大学, 2024.
- Ren J H. Spatial and temporal characteristics and driving factors of 2-MIB in algae- and macrophyte-dominated zones in Lake Taihu[D]. Wuxi: Jiangnan University, 2024.
- [21] 童俊. 金泽水库及其上游来水水质特征及变化规律[J]. *净水技术*, 2020, 39(1): 58–68.
- Tong J. Characteristics and variation rules of raw water quality for jinze reservoir and upstream water [J]. *Water Purification Technology*, 2020, 39(1): 58–68.
- [22] 赵肥西, 秦赫, 彭晓然, 等. 丹江口水库氮磷浓度特征和营养阈值分析[J/OL]. *水电能源科学*, 2026–06–09. <https://link.cnki.net/urlid/42.1231.TK.20260121.1135.022>.
- Zhao F X, Qin H, Peng X R, et al. Analysis of nitrogen and phosphorus concentration characteristics and nutrient thresholds in Danjiangkou Reservoir [J]. *Water Resources and Power*, 2026–06–09. <https://link.cnki.net/urlid/42.1231.TK.20260121.1135.022>.
- [23] 朱宜平. 近十年来青草沙水库取水口水质变化趋势分析[J]. *华东师范大学学报(自然科学版)*, 2022(3): 50–60.
- Zhu Y P. Analysis of changes in water quality at the Qingcaosha Reservoir water intake over a period of 10 years[J]. *Journal of East China Normal University(Natural Science)*, 2022(3): 50–60.