杨 京,杨忠勇,和玉芳,等.基于机器学习的香溪河库湾水华暴发主控因子识别及贡献 [J]. 中国环境科学, 2026,46(3):1486-1498.
Yang J, Yang Z Y, He Y F, et al. Machine learning-based identification and contribution analysis of dominant factors driving algal blooms in the Xiangxi Bay [J]. China Environmental Science, 2026,46(3):1486-1498.

基于机器学习的香溪河库湾水华暴发主控因子识别及贡献

杨 京[1],杨忠勇[1,2*],和玉芳[3],熊励伟[1],胡怡玲[1],纪道斌[1,2] (1.三峡大学水利与环境学院,湖北 宜昌 443002；2.三峡大学,三峡水库生态系统湖北省野外科学观测研究站,湖北 宜昌 443002；3.河北省水利规划设计研究院有限公司,河北 石家庄 050021)

摘要：为明确香溪河库湾水华暴发的主控环境因子,基于 2024 年 8 月 9~30 日的高频监测数据,以叶绿素 a(Chl.a)为水华表征指标,采用随机森林(RF)、支持向量回归(SVR)和偏最小二乘回归(PLS)模型识别关键驱动因子.结果表明,香溪河库湾水华过程具有显著的日内节律变化特征,夜间低扰动和热分层的稳定条件显著促进水华形成,而白昼扰动和高温环境则抑制水华积累.流速(WV)和气温(AT)是水华暴发的即时驱动因子,贡献率分别为 27.1%和 18.1%,电导率(SP)作为重要调节因子贡献率为 17.1%,而溶解氧(DO)、pH 值和水温(WT)则表现为次要的反馈性因子.此外,各因子在水华过程中表现出显著的非线性与滞后响应特征.

关键词：藻类水华；香溪河；机器学习；环境因子；贡献率分析
中图分类号：X524 **文献标识码**：A **文章编号**：1000-6923(2026)03-1486-13

Machine learning-based identification and contribution analysis of dominant factors driving algal blooms in the Xiangxi Bay. YANG Jing[1], YANG Zhong-yong[1,2*], HE Yu-fang[3], XIONG Li-wei[1], HU Yi-ling[1], JI Dao-bin[1,2] (1.College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China；2.Hubei Provincial Field Observation and Research Station for the Three Gorges Reservoir Ecosystem, China Three Gorges University, Yichang 443002, China；3.Hebei Water Conservancy Planning and Design Institute Co., Ltd., Shijiazhuang 050021, China). *China Environmental Science*, 2026,46(3)：1486~1498

Abstract：Algal blooms are a major ecological and environmental issue in the tributary bays of the Three Gorges Reservoir. To identify the dominant environmental factors driving algal blooms in the Xiangxi Bay, this study utilized high-frequency monitoring data collected from August 9 to August 30, 2024, with chlorophyll-a (Chl.a) concentration as the bloom indicator. Three machine learning models—Random Forest (RF), Support Vector Regression (SVR), and Partial Least Squares Regression (PLS)—were applied to identify key driving factors. The results showed that algal blooms in Xiangxi Bay exhibited pronounced diurnal variation, with nighttime stable stratification and low disturbance favoring bloom formation, while daytime disturbances and high temperatures inhibited algal accumulation. Water velocity (WV) and air temperature (AT) were identified as immediate driving factors, with contribution rates of 27.1% and 18.1%, respectively. specific conductivity (SP) served as an important regulating factor (17.1%), while dissolved oxygen (DO), pH, and water temperature (WT) were secondary feedback factors. Moreover, the impacts of these factors on Algal blooms exhibited significant nonlinearity and lagged response characteristics.

Key words：algal blooms；Xiangxi River；machine learning models；environmental drivers；contribution analysis

水华的发生过程受多种环境因子联合作用调控,包括水温、光照、营养盐、流速、风速、水体稳定性等,且各因子的响应常具有明显的时滞特征与交互反馈机制[1-2].因此,准确识别关键驱动因子并建立有效的水华预测模型,是当前水生态研究和水质管理实践中的核心难点与热点问题[3].关于水华在生长与暴发期间的影响因素和机制等问题,大量研究普遍认为与营养盐、水温、光照和水流条件相关[4-5],但不同水体的主控因子差异显著.已有研究主要从以下 3 个视角探索了水华的发生机制与预测方法[6-7],一是从机制建模角度出发,识别与量化水华的主控因子,例如,相关研究[8]提出的"临界层理论"以及通过构建混光比(Z_mix/Z_eu)指标[9],从热力结构和光生境分布角度揭示了浮游植物竞争优势的形成机制,特别适用于三峡库区支流等具有深水、分层明显、水动力弱等特征的库湾型水体;二是基于水生态动力学的耦合模拟模型(如 CE-QUAL-W2、MIKE 等)进行动态预测,适用于尺度较大、水文过程

收稿日期：2025-07-22
基金项目：国家自然科学基金资助项目(U24A20180,42477086,52379069);湖北省中央引导地方科技发展专项(2024CFA022);水资源高效利用与工程安全国家工程中心开放基金资助项目(GJGCZX-JJ-202418)
* 责任作者, 教授, ayong0710@163.com

明确的水体[10-11];三是近年来快速发展的机器学习方法,通过挖掘多源监测数据中隐含的非线性响应关系,提升水华预测的准确性与泛化能力[12-13].典型方法包括 RF、SVM、ANN、XGBoost 等,已广泛应用于太湖、巢湖、三峡水库等典型区域,表现出良好的预测效果和变量识别能力[14-16].在三峡水库香溪河支流库湾,已有研究初步采用遥感影像与随机森林模型识别水华时空分布与营养因子关联性[17],但针对暴发期过程的动态机制识别仍较为薄弱,尚缺乏面向高频数据的建模评估框架.

香溪河作为三峡库区典型的支流型库湾,具有结构狭长、水体交换弱、热分层显著等特征[18],导致其水华暴发过程具有高度的时空异质性[19].既有研究[20]表明其蓝藻水华具有显著的年内季节性:夏季—初秋为高发窗口,呈短周期、间歇性与局地性暴发特征,且年际间强度与频次随水文调度与气候年景而波动.部分断面 Chl.a 浓度峰值超过 300μg/L,严重威胁区域水质安全与生态稳定性.然而,关于香溪河库湾水华生长及暴发的控制因素目前仍不清楚,且不同变量的响应滞后特征与协同效应亦缺乏系统分析[21].本文基于连续高频水文气象与水质数据,以 Chl.a 浓度表征水华,构建 RF、SVR、PLS 三模型;联合多相关性与滞后分析,识别主控因子、量化贡献,并通过变量筛选检验提升模型稳定性;基于暴发期的日内过程证据,提出“阈值—时窗”预警思路,构建适用于三峡库区支流库湾的水华预测框架,旨在为香溪河水质管理与暴发期预警提供参考.

1　材料与方法

1.1　监测方案

选取三峡水库距离大坝最近的一条大型支流香溪河库湾作为研究对象.香溪河上游发源于湖北省兴山县,是长江右岸一级支流,流域面积约 3200km^2,全长 184km,流域地势南高北低、地形狭长,河道弯曲,汇入三峡水库后形成典型支流库湾[8].受三峡水库水位抬升影响,香溪河库湾水动力较弱,水体交换能力差,易形成热分层和异重流结构[22],水华暴发过程呈现高频突发、短周期反复、局地高浓度积聚等特征[23].

监测工作于 2024 年 8 月 9~30 日在香溪河中游峡口断面监测平台开展(图 1).该平台处于库湾主槽段,水体扰动条件典型.各项监测指标和技术方法如下:(1)气象因子,包含气温(AT)、降雨量(RF)、风速(WS),使用香溪河野外监测平台气象站的记录,数据导出频率为 10min/次;(2)水质理化因子,包含溶解氧(DO)、pH 值、电导率(SP),采用多参数水质分析仪(EXO2,YSI,美国)进行定点连续监测,仪器垂直悬挂于表层(0.5m),监测频率为 10min/次;(3)水体流速(WV)数据,采用声学多普勒点式流速仪(Vector, Nortek,挪威)在表层定点监测,仪器采样频率 32 次/s; (4)水温(WT)数据,微型温度仪(RBR solo3,RBR,加拿大)在表层连续记录,记录频率为 5s/次;(5)光合有效辐射(PAR)数据,采用 LI−COR 192SA 光量子传感器(美国)在每日 8:00 与 14:00 人工定点记录;(6)Chl.a 浓度数据,每日 2:00、8:00、14:00、20:00 采集表层水样(500mL),现场过滤,乙酸纤维滤膜−4℃冷冻保存后在 2h 内送实验室检测,依据《水和废水监测分析方法(第四版)》[24]采用分光光度法测定.监测仪器均按标准流程进行周校准、漂移核查和清洗.

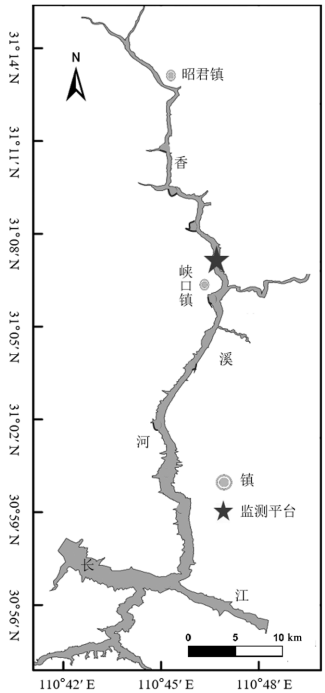


图 1　研究区域及监测平台的地理位置

Fig.1　Location of the study area and the monitoring platform

1.2　数据预处理

为确保不同监测因子的时间尺度一致并与水华指标精准对齐,以 Chl.a 浓度的采样时间点(每日

2:00、8:00、14:00、20:00)作为统一参考基准,按时间最近邻原则从各高频监测数据中提取与之匹配的气象、水文及水质变量值,构建整合后的时序数据集.在此基础上,为消除各变量量纲差异对模型训练造成的干扰,提升不同模型的训练效率与预测性能,采用最小–最大标准化方法(Min−Max Normalization)对所有输入指标数据进行归一化处理,线性转换至[0,1]区间,确保各变量在统一尺度下参与建模计算.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

式中:x 为原始数据值(某个变量在某时刻的观测值);$x_{\min}$ 为该变量在所有样本中的最小值;$x_{\max}$ 为该变量在所有样本中的最大值;x_{morm} 为归一化后的值,范围是[0,1].

1.3 相关性分析方法

为全面评估环境因子与水华演变过程(以 Chl.a 浓度表征)之间的响应关系,采用 3 种相关性分析方法:Pearson 相关系数[25]、Spearman 等级相关系数[26]与 Distance correlation(距离相关性)[27]分析,分别从线性、单调和非线性维度刻画变量间的统计关系.

1.4 贡献率分析方法

为筛选并量化此次香溪河水华暴发过程中对 Chl.a 浓度产生关键影响的因子,采用 3 种机器学习模型进行分析:随机森林(RF)、支持向量回归(SVR)和偏最小二乘回归(PLS)[28-29].以 Chl.a 浓度为目标变量,环境因子为特征变量进行建模,对各模型进行 200 次重复建模训练,结合网格搜索确定最优参数,并使用三折交叉验证以减少随机误差并防止过拟合.通过贡献率分析法,能够准确识别并量化对 Chl.a 浓度变化有重要影响的因子.

随机森林(RF):在自助采样与特征随机子集上生成多树,采用置换重要性(permutation importance)评估变量对预测误差(OOB−MSE)的增量贡献,避免仅用 MDI 的类别/尺度偏倚;对每次重复建模的贡献值做标准化与中位数汇总,并报告稳定性(IQR).为抑制冗余变量干扰,先行进行变量筛选(递归剔除/稳定筛选),再在筛选集上重训 RF 模型得到最终排序.

支持向量回归(SVR):采用 RBF 核+ε−不敏损失,对超参数(C、γ、ε)用网格搜索+k 折交叉验证确定;输入统一做对数变换/标准化.变量贡献采用 SHAP:以 KernelSHAP 在样本层面计算边际贡献并取绝对值平均得到全局重要性,同时保留 SHAP 符号解释作用方向;为降低逸值影响,对 SHAP 分布取稳健统计量(中位数、IQR).

偏最小二乘(PLS):针对强共线场景,先对 x、y 对数变换并中心化/标准化,主成分数 n_components 由交叉验证确定(最小 MSE 准则,防过拟合).变量贡献以标准化回归系数与 VIP(Variable Importance in Projection)综合评估:VIP>1 作为保留阈值,系数符号用于判读正/负效应;重复建模取系数中位数以保证稳健性.

为保证各模型结果的统一性,对 SVR 和 PLS 模型中的变量贡献率结果进行了绝对值归一化处理,使得每个模型得到的变量重要性具有相同的尺度,便于进行比较分析.为评估模型的精度和稳定性,采用决定系数(R^2)用于衡量模型对数据变异的解释能力,用均方根误差(RMSE)和平均绝对误差(MAE)评估模型的预测误差.通过这些评估指标,能直观反映出模型性能,进而选择最合适的模型进行后续分析.

2 结果与讨论

2.1 水华时序演变与环境响应特征

香溪河库湾历年在夏季—初秋阶段频繁出现短周期、高强度水华[18],因此选取 8 月作为高频过程观测窗口,旨在典型高发期内识别水华的时序演变规律及驱动机制.需要说明的是,本文的代表性限定于香溪河水华高发阶段的过程机制,并不外推为全年平均状态;对应的监测设计采用“4 次/d (2:00, 8:00,14:00,20:00)”采样与多源高频传感器联合监测,以覆盖白昼—夜间差异并捕捉扰动背景下的快速响应.

监测期(2024 年 8 月 9~30 日)落于香溪河水华的年内高发窗口,且与三峡库区防洪消落和日调峰阶段同期(图 2(b)).监测数据显示,Chl.a 浓度在此期间波动显著(图 2(a)),范围为 1.31~318.76μg/L,平均浓度为 38.70μg/L,远高于通常认定的水华暴发阈值(30μg/L)[30].其中,共监测到 28 次水华暴发过程(Chl.a≥30μg/L),9 次达到极端水华水平(Chl.a>100μg/L),表明该时段内水体处于典型的中至高强度水华活动状态.水华过程呈现明显的间歇性与短周期特征,部分高 Chl.a 浓度持续时间不超过 12h,反映出局地扰动条件与环境因子的快速耦合响应.

同期,三峡库区水位和流量条件呈现出显著变化:监测期间水位由 156.5m 下降至 152.5m,总降幅约 4.0m,平均日降幅约 0.2~0.3m;干流流量由约 17,500m^3/s 逐渐降至 14,800m^3/s,总降幅约 2,700m^3/s,且日内高低峰波动幅度约 2,000~2,500m^3/s,日调峰特征明显.这一变化主要受三峡水库防洪消落调度影响,库区水位快速下降以腾出防洪库容,同时受三峡电站发电负荷调峰控制,白昼高峰出流量大、夜间谷值出流量小.这些水位和流量的变化加强了库湾内水体扰动和水位振荡,削弱热分层稳定性,改变浮游植物的分布格局,对水华暴发和衰减过程均可能产生重要调控作用.

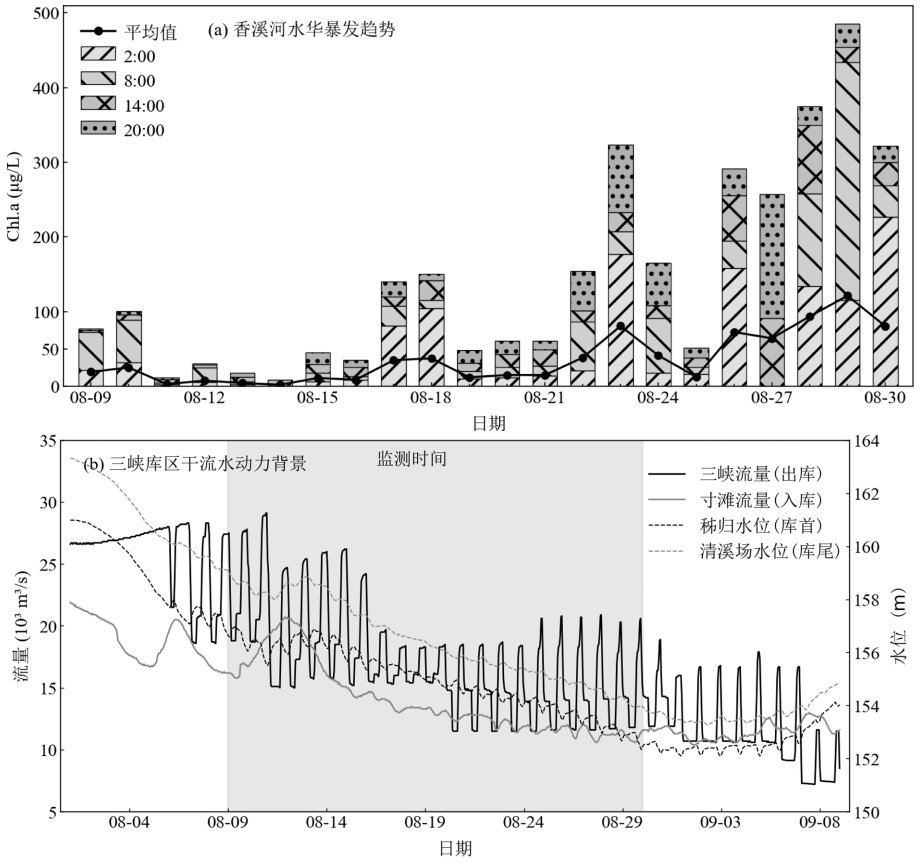


图 2　监测时段香溪河库湾水华暴发趋势和三峡库区干流水动力背景

Fig.2　Bloom outbreak trend in the Xiangxi River embayment during the monitoring period and the hydrodynamic background of the main stream of the Three Gorges Reservoir

2.1.1　日内变化规律与水体扰动机制　为揭示水华暴发期间 Chl.a 浓度在日内时序上的动态变化特征及其对水文气象因子的响应关系,对每日 2:00、8:00、14:00、20:00 4 个典型时段的主要指标进行了分组统计(图 3),并分析各因子的日变化节律与协同响应机制(图 4,表 1).监测结果表明,Chl.a 浓度在不同时间段存在显著差异,表现为夜间高、白昼低的变化模式,凌晨与早晨(2:00、8:00)浓度较高且波动范围大,白昼(14:00、20:00)显著下降.与此相对应,大多数环境因子如气温(AT)、水温(WT)、风速(WS)、光合有效辐射(PAR)在白昼达到高值,而夜间较低;溶解氧(DO)、pH 值则在白昼 14:00 左右出现峰值.

这种现象主要与水体动力与藻类生理生态特性两方面相关[31],白昼时段,受强光照、升温及局地风扰动等共同作用,表层水体热分层减弱,流速、风速增强(例如 8:00~14:00 时段 WV 上升 444.1%),水体垂向混合作用加剧,不利于藻类在表层稳定聚集,同时高温与强光也可能抑制藻类光合作用效率,引发光抑制效应,导致 Chl.a 浓度下降.相反,夜间至清晨水体扰动弱化,热分层与静稳层结构重新建立,水体交换与外源输入减少(表现在 2:00、8:00 时段 SP 较低),浮游植物上浮聚集,促进表层藻类生物量积累,Chl.a 浓度显著升高.

从相关性分析结果来看(表 1),多数环境因子(如 WV、AT、WT、WS、PAR、SP)与 Chl.a 变化表现出高度负相关(P<0.05),这是由于水体扰动、热力条件的增强往往会削弱表层藻类积聚,而静稳环境则有利于藻类垂向聚集与生物量上升.此外,DO、pH 值与 Chl.a 同样呈负相关.白昼时段,光合作用增强、扰动条件加强,共同导致 DO、pH 值上升,而此时藻类因扰动加剧等因素难以在表层稳定聚集,Chl.a 浓度下降[32].体现了藻类昼夜生理代谢与水体物理扰动交互调控的综合效应.

综上,香溪河库湾水华暴发过程的日内变化与环境因子变化节律密切相关,Chl.a 浓度在夜间静稳、低扰动、低光低温条件下聚集,而白昼则因水体扰动增强、光温条件变化等因素抑制藻类表层积累.

表 1 各变量与 Chl.a 浓度在时间段间百分比变化的 Pearson 相关性

Table 1 Pearson correlations between each variable and Chl.a concentration based on percent changes across time intervals

变量	WV	AT	RF	WS	PAR	WT	DO	PH 值	SP
相关性	−0.96**	−0.94*	−0.95*	−0.83	−0.94*	−0.98**	−0.94*	−0.91*	−0.86*

注: *为P<0.05; **为P<0.01.

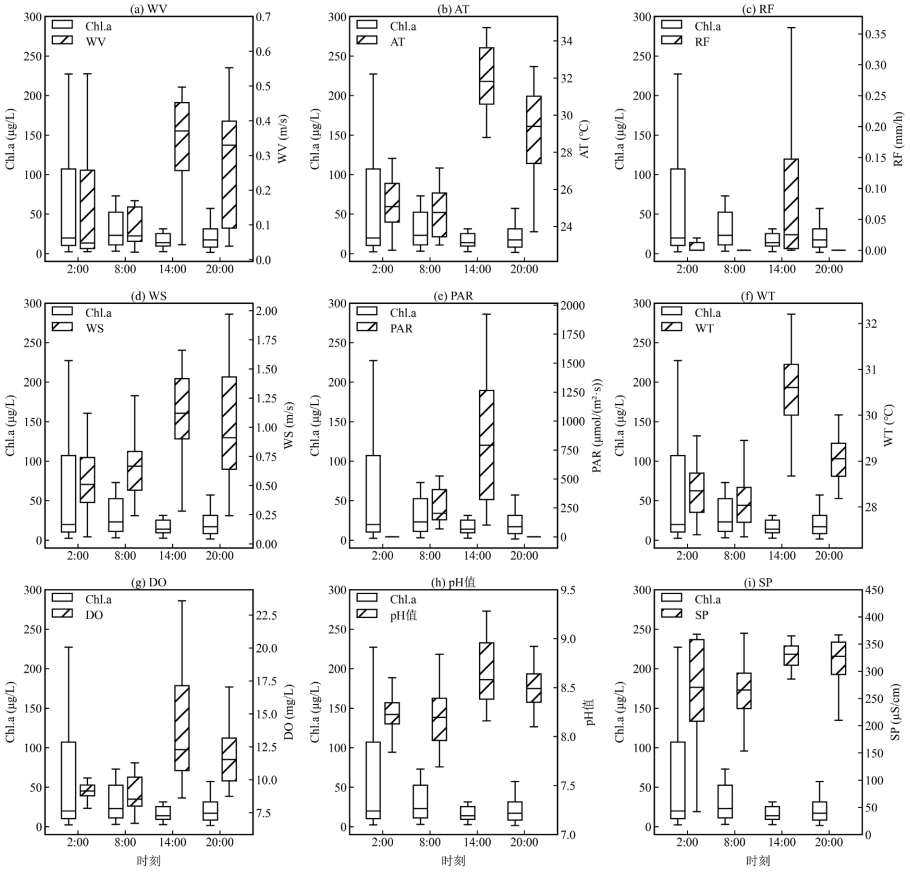


图 3 Chl.a 浓度及环境因子的日变化特征

Fig.3 Diurnal variations in Chl.a concentration and environmental factors

2.1.2 水华分级下的主控环境因子差异 将 Chl.a 浓度按照分级阈值划分为 3 类:轻度水华(Chl.a<30μg/L)、中度水华(30≤Chl.a<100μg/L)和重度水华情况(Chl.a>100μg/L)[30,33],并对各等级下的水文气象与理化因子进行比较(图 5).

从图 5 可以观察到,随着水华严重等级的上升,流速(WV)、气温(AT)、风速(WS)及光合有效辐射(PAR)均呈现显著下降趋势.特别是在重度水华阶段,WV 的中位数已降至约 0.02m/s,AT 和 PAR 也分别下降至约 25℃与近 0μmol/(m^{2}s),说明静稳、低扰动、低温低光的水体环境更利于藻类聚集与水华暴发[34].这与前文所述对夜间高水华现象的

机制解释一致,进一步说明了低扰动促进水华的　　判断.

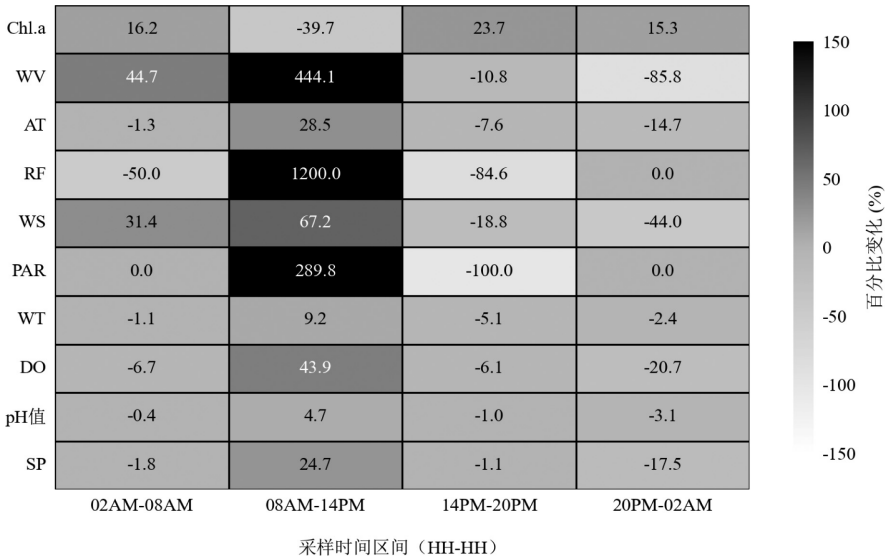


图 4　Chl.a 浓度与环境因子日内百分比变化特征

Fig.4　Diurnal percent changes in Chl.a concentration and environmental factors

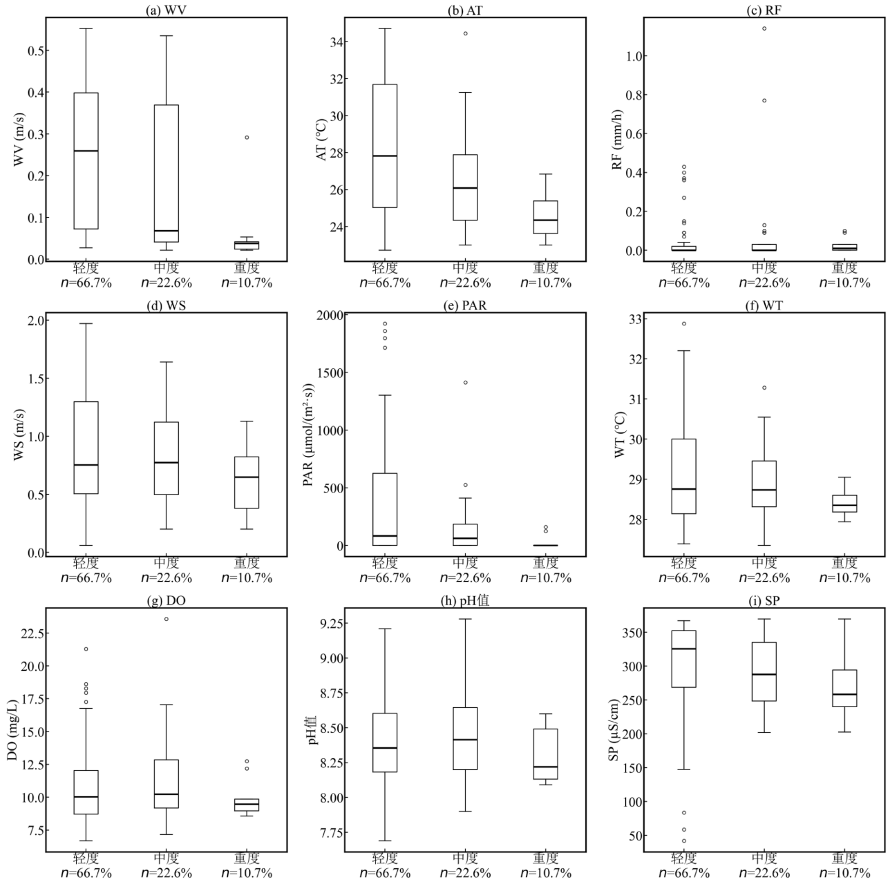


图 5　不同水华等级下环境因子的分布特征

Fig.5　Distributions of environmental factors across different bloom severity levels

溶解氧(DO)与 pH 值亦呈现下降趋势,重度水华期间 DO 浓度集中在 8~10mg/L,低于轻度水华和中度水华期;pH 值中位数则维持在 8.2 左右,波动收窄,表明藻类密集生长期间水体的生物代谢活动趋于

平稳,并可能抑制了氧气释放与 CO_2 消耗效率.电导率(SP)在重度水华阶段略低于前两阶段,说明在高浓度水华形成过程中,水体离子强度趋于稳定、水体交换减弱与外源输入减少有关,这种低电导状态有助于维持表层水体的稳定性,为藻类聚集创造有利条件;水温(WT)虽随水华级别变化幅度不大,但在重度水华期同样表现出较低的集中水平,进一步支持藻类聚集更适于低水温的假设[35].

总体来看,不同水华等级下的环境因子分布差异明确,水华暴发的形成与低扰动、低光照、弱风、低气温等环境条件密切相关,表明水体稳定性与热分层结构可能是诱发高强度水华的关键背景因素.

2.1.3 环境变量相关性与滞后响应分析 为全面评估水华演变(以 Chl.a 浓度表征)与各环境因子之间的响应关系,对各因子与 Chl.a 浓度的即时相关性和滞后效应进行量化(图 6).结果显示,流速(WV)与气温(AT)在相关性分析中对 Chl.a 浓度均表现出显著的负相关关系(P<0.01),其中 Distance 方法揭示出较强的非线性关系,表明水体扰动与热力条件是显著的水华抑制因子.PAR、WT 与 SP 对 Chl.a 浓度亦表现出负相关趋势(P<0.05),而 DO 与 pH 值则表现出较强的正相关性(P<0.05);同时,以上 5 个理化因子在 Distance 分析中也均呈现出统计显著的非线性联系.

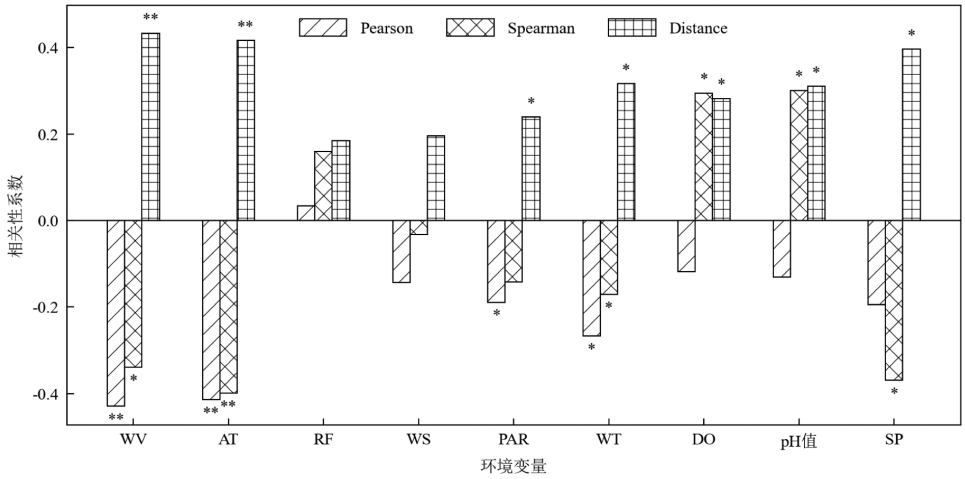


图 6 Chl.a 浓度与环境变量的相关性分析结果

Fig.6 Results of correlation analyses between Chl−a concentration and environmental variables

*为 P<0.05; **为 P<0.01

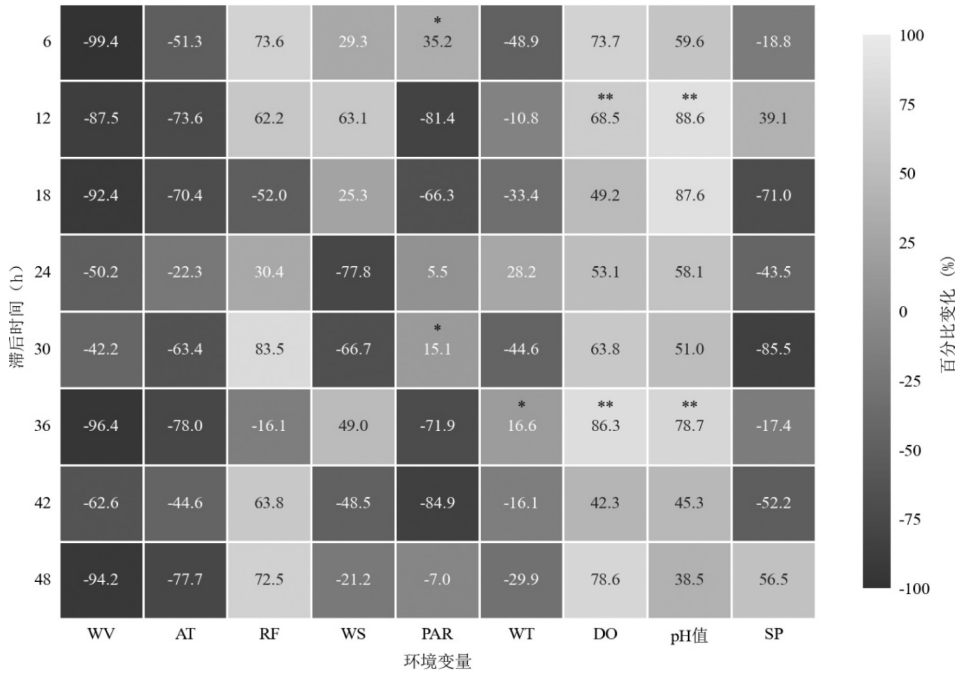
为更全面地揭示环境因子对水华过程的非同步影响机制,特别是捕捉某些因子在时间上对 Chl.a 浓度的滞后响应效应,进一步构建了 6~48h、以 6h 为步长的相关性滞后热图(图 7).该方法有助于识别那些在当前时刻虽未显著相关,但在一定时间延后后对水华具有驱动作用的变量,从而揭示潜在的累积效应、间接影响路径或缓释性机制.图中色块所示数值代表变量在相应滞后时间点相较于即时响应(滞后 0h)的相关性增幅百分比.

图 6、图 7 结果显示,WV 与 AT 在滞后 0h 相关性最强,随后随滞后时间增长显著衰减,呈现典型的即时扰动型驱动特征.相较之下,DO、PH 值和 SP 等变量在滞后时间中(12,18,30,36h)相关性显著增强,其中 DO 与 pH 值在 12 与 36h 达到最大相关性提升(P<0.01),呈现出典型的迟滞性响应,表明其调控作用更依赖于累积性代谢或内源过程.WT 与 SP 亦在多个滞后时间窗口中表现出稳定的正相关增强趋势,特别是在 Spearman 分析中,SP 在滞后 42h 相关性提升超过 45%(P<0.01),WT 在 24,30,42h 的相关性提升均超过 60%(P<0.05),显示其通过水中离子结构稳定性、水体交换和水体热稳定性影响浮游植物分布具有时间积累特征.PAR 在滞后 6h 处在 3 种相关性分析中均呈现出超过 30%的正向提升(P<0.05),说明其对藻类光合作用的影响具有短时滞后特性[36].

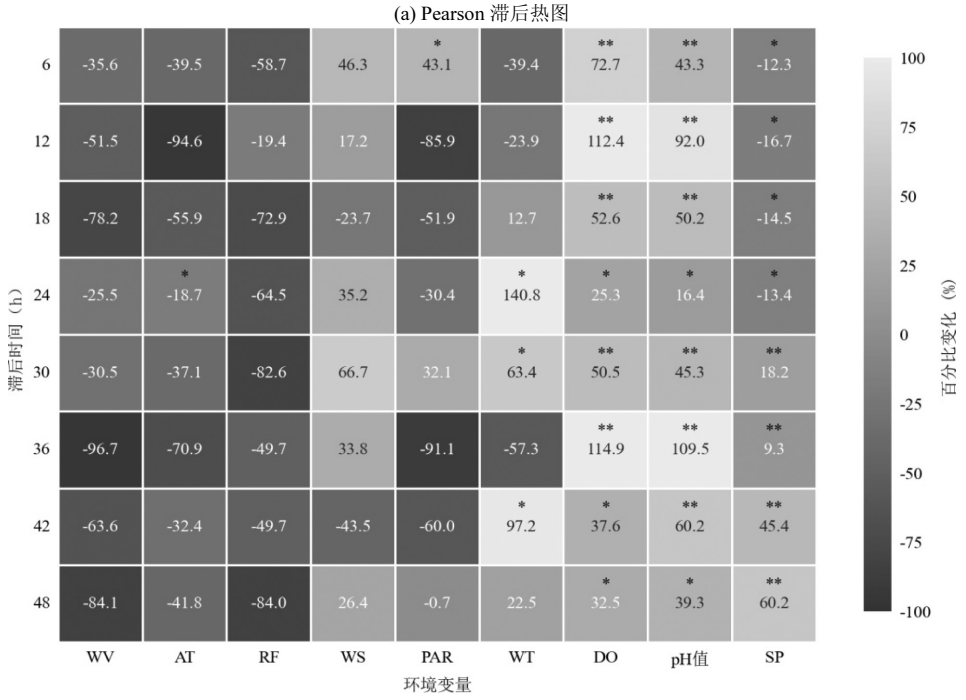
值得注意的是,RF 与 WS 在 3 种相关性方法下相关系数均较低,且未达到显著性水平(图 6),滞后热图亦未显示出明显响应增强.分析认为,该现象与研究期内实际观测条件密切相关:降雨过程整体强度

弱、频次低,主要集中于 8 月 17 日与 28 日的两次短时降雨(强度均约为 1mm/h),其余时段降雨量接近于零,导致对水体营养状态影响有限,难以引起 Chl.a 浓度的实质性响应.已有研究亦表明,在降雨事件稀疏或背景营养充足条件下,RF 对水华演变的贡献通常不显著.此外,WS 虽在日间波动范围较大(0~2.0m/s),但受香溪河峡口地形狭窄与高差岸线遮挡影响,外源风力传导受限,形成局地风屏蔽效应,无法在水体表层形成有效扰动;同时,风速扰动多表现为短时、间歇性变化,其调控作用具有高度空间与季节依赖性,在本研究所处的"微风–弱扰动"背景下难以对水华形成产生显著驱动[37].因此,RF 与 WS 相关性整体偏弱,滞后热图中亦缺乏统计意义上的响应变化,符合其机制上的间接性与局限性特征.

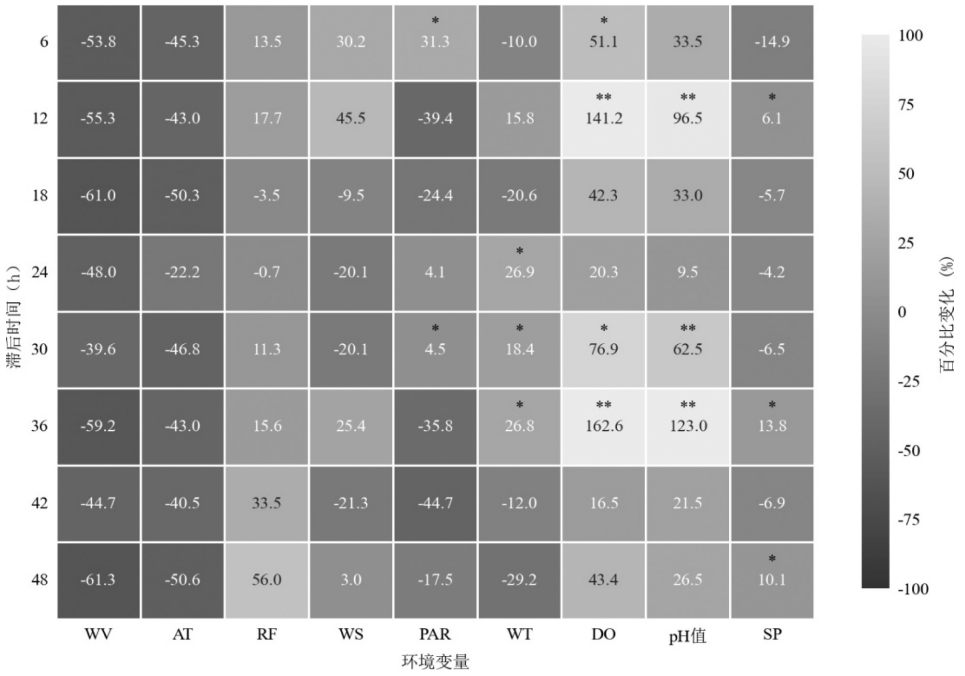
综上,Chl.a 浓度变化对不同环境变量存在显著的响应节律与时间尺度差异,WV 与 AT 可视为典型的即时扰动因子,WT、DO、pH 值与 SP 则体现出较强的迟滞调节效应.



(a) Pearson 滞后热图



(b) Spearman 滞后热图



(c) Distance 滞后热图

图 7 Chl.a 与环境变量的相关性滞后热图

Fig.7 Lagged correlation heatmaps between Chl.a and environmental variables

2.2 模型识别结果与驱动机制解释

2.2.1 主导因子识别与模型性能评估 本文基于随机森林(RF)、支持向量回归(SVR)与偏最小二乘回归(PLS) 3 种机器学习模型的建模结果筛选出影响水华生长过程的关键因子,以 Chl.a 浓度为目标变量表征水华演变情况,其他监测指标为特征变量输入模型作为响应因子.如图 8,将模型每次训练的变量贡献率结果进行排序统计[38],在各位次中累计频次最高的变量作为本位次最重要的变量(排除此前已选变量),得到各变量的重要性位次的排序,分别为 1~9;再将每个变量在 RF、SVR、PLS 模型的名次相加,得到综合排名(Total).分数越高则表示重要性越低.

影响因子在 RF 和 SVR 模型结果和综合排名中表现出了相似的规律:WV、AT 对 Chl.a 浓度的贡献率最大,占据第 1 和第 2 位;WT、DO、PH 值、SP 的贡献率次之,分别在第 3~第 6RF、WS、PAR 的贡献率最低,重要性排名在第 7~第 9.在 PLS 模型中,AT 和 DO 的贡献率最大,在第一和第二位,WT、WV、PAR、PH 值的重要性次之,RF、WS、SP 的贡献率最低.

模型分析结果与相关性及滞后效应分析中的发现相互吻合:WV 与 AT 表现出显著负相关趋势在这两类非线性建模方法中对 Chl.a 变化具有强解释力;WT、DO、PH 值与 SP 表现出部分相关性,验证了其在模型中的长期调节作用;RF、WS 与 PAR 在 RF 与 SVR 模型中的重要性排序均靠后,表明其对模型训练贡献较低,这与相关性低或滞后不显著的结果相吻合.

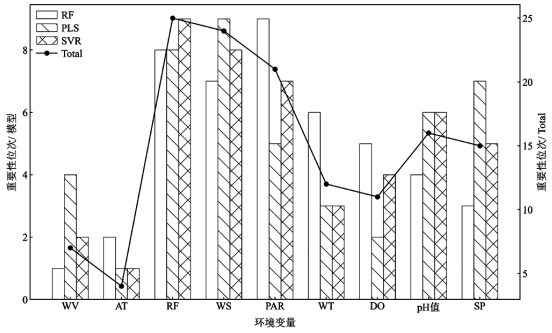


图 8 各环境变量在模型分析中的重要性统计排序

Fig.8 Ranking of environmental variables by importance in the model analyses

基于模型综合表现和相关分析,确认 WV、AT 为关键影响因子,WT、DO、PH 值、SP 居中,具备一定保留价值;而 RF、WS 与 PAR 得分显著较高,预测贡献性相对较弱.初步筛选出 RF、WS 与 PAR 为冗余变量,后续建模将其剔除以提升模型精度与简洁性.

如图 9 所示,在全变量建模情况下,3 种模型在性能指标上的表现差异显著.RF 拟合更优但重复波动较大;PLS 稳定性最好但解释力弱;SVR 介于两者之间.

从 R^2 来看,RF 模型在拟合优度方面表现最佳.R^2 的分布范围(Q1~Q3)为 0.43~0.69,中位数 Q2=0.66,但 R^2 波动范围较大,表现出较强的模型拟合能力和不稳定性.SVR 模型的 R^2 变化范围为 0.15~0.35,Q2=0.35,表明其拟合效果相对稳定,但整体拟合优度较低.PLS 模型的 R^2 仅在 0.05~0.08 之间,Q2=0.06,显示出其在解释 Chl.a 浓度变化方面的能力弱,未能有效捕捉数据中的复杂模式.

在误差指标 MAE、RMSE 方面,RF 模型的 MAE 和 RMSE 数值相对较小,但波动范围较大,表明 RF 模型虽然预测误差较小,但在重复建模时误差的波动较为显著.相较之下,SVR 和 PLS 模型的误差指标值相对较高,但数据波动比 RF 模型小,表明尽管 SVR 和 PLS 模型的误差较大,但它们在重复建模中的稳定性优于 RF 模型.

在剔除贡献率最低的 4 个变量(RF、WS、PAR、ORP)后,模型在重复建模中所得到的评估指标结果如图 9 (b)所示.RF 模型的性能得到了显著的提升,主要体现在模型稳定性提升了约 15.54%,表明变量筛选有效提高了 RF 模型的稳定性.SVR 和 PLS 模型在筛选前后的性能变化不大.

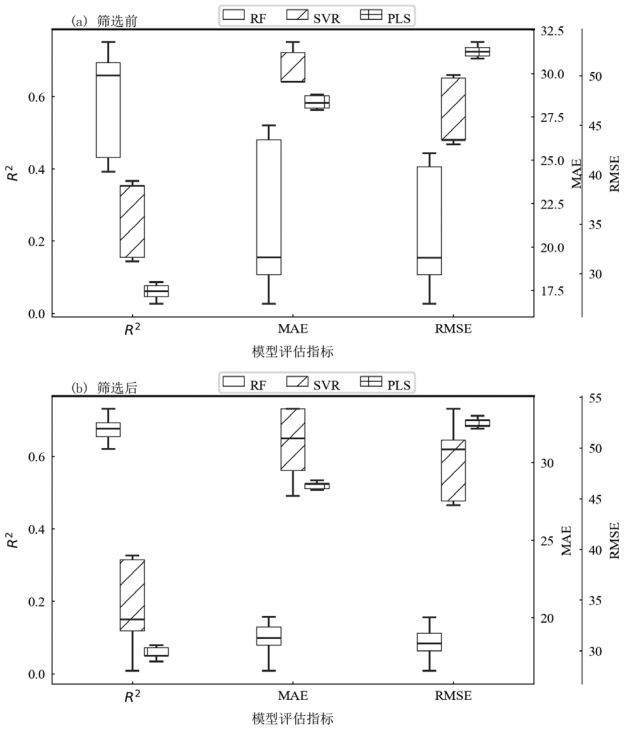
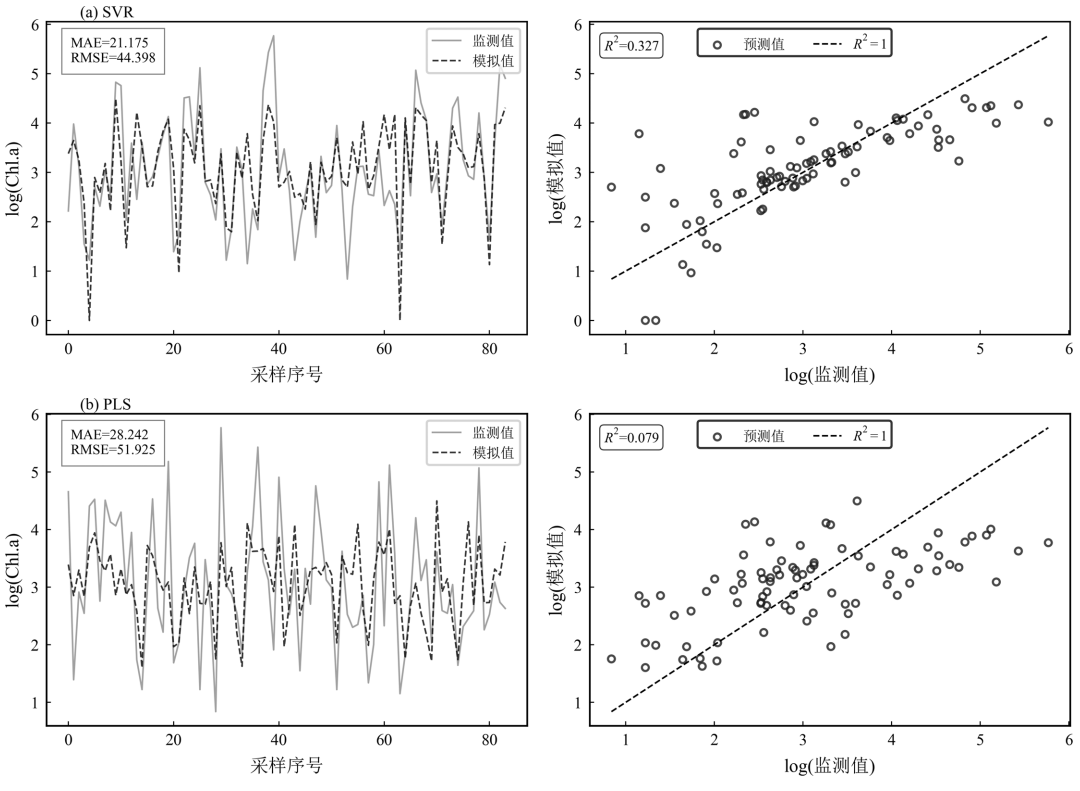


图 9　变量筛选前后模型性能评估情况

Fig.9　Model performance before and after variable selection



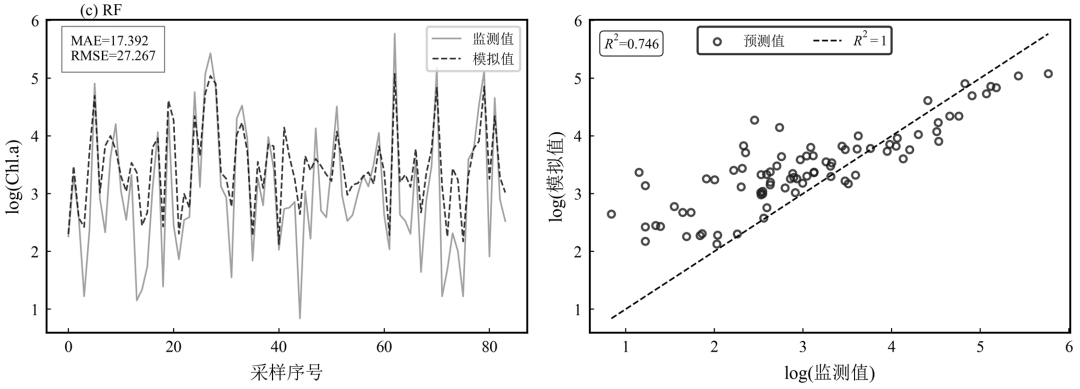


图 10 模型评估指标对比情况

Fig.10 Comparison of model evaluation metrics

从图 10 中可以看出,RF 模型在所有评估指标上都优于 SVR 和 PLS 模型,有最小的误差范围和最大的拟合优度,能较好地适应多维数据驱动下的复杂非线性响应关系,是水华贡献率研究的有效工具.

从上述分析得出,RF 模型的表现最佳,能够更好地处理水环境中多因子驱动下的复杂非线性响应关系,是构建 Chl.a 浓度解释框架的优选工具.

2.2.2 不同因子的驱动作用与对比 为进一步定量评估关键环境因子对 Chl.a 浓度变化的驱动强度,基于构建后的优化 RF 模型,统计各变量在多轮训练中的贡献率分布,并计算其平均贡献率(Aevr)以反映其综合影响能力(图 11).

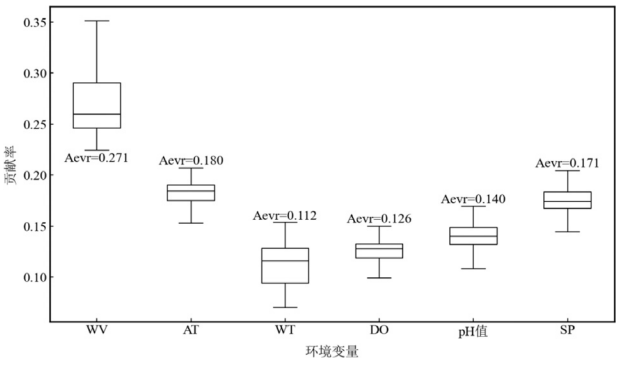


图 11 变量筛选后 RF 模型的贡献率结果

Fig.11 Contribution rates from the RF model after variable selection

从图 11 中可以发现,流速(WV)对 Chl.a 浓度的贡献率最高,Aevr=0.271,在所有变量中表现最为突出.这一结果与江明顺等在奥基乔比湖的研究相符,其指出水体扰动强度是影响藻类聚集和垂向迁移的重要因子,低流速状态更容易导致水体分层、藻类上浮并形成富集区[39].同时,WV 的贡献率分布在 0.246~0.29 之间,虽然有一定波动,但维持在较高水平,说明其对水华形成的持续性影响显著.

气温(AT)以 Aevr=0.181 位居第 2,也与已有研究高度一致.有关研究[40]表明,气温作为热力条件的核心变量,不仅影响水体表层温度分布,还能通过控制热分层强度与水体垂直交换来间接调控藻类生理状态和营养盐再悬浮,进而促进水华的发展.本研究中,AT 贡献率显著也验证了其在热驱动型水华机制中的主导地位.

电导率(SP)的贡献率同样较高(Aevr=0.171),其显著性可能源于其对水体矿化程度与离子交换强度的反映,相关研究指出 SP 的变化可指示水体外源输入与交换强度[41],在封闭或半封闭水体中,高 SP 往往伴随水体稳定性增强和混合作用减弱,促使浮游植物在表层形成富集态势.此外,结合临界层理论[8]可解释 SP 对垂向结构的影响——温盐梯度形成的界面常常是藻类聚集带,电导率在此过程中发挥间接调节作用.

相较之下,DO、pH 值和 WT 的平均贡献率较低,分别为 0.126、0.140 和 0.112,表明这些变量虽对水体生态过程有反馈作用,但在预测水华演变过程中并非首要驱动因子.类似的结果也出现在 Chen 等[42]的研究中,在对太湖蓝藻水华的研究结果表明溶解氧和 pH 值更倾向于作为水华结果的反映变量,而非诱发因子.

综上所述,本研究贡献率分析结果与已有文献的实证规律吻合,也从理论机制上验证了流速、气温等因子的主导性作用,SP、pH 值等因子的调节性作用,以及 DO、WT 等变量的次要反馈性角色.据此,面向库区防洪消落与日调峰情景,可构建阈值—时

窗预警策略:当夜间低流速、低气温且静稳增强时判定高风险窗口;当 SP/WT/DO/pH 值出现 12~42h 的先行信号时提前实施混合/曝气或生态调度,以降低表层富集风险.上述结论既为机理认识提供过程证据,也为预测建模与管理决策提供可操作的触发条件与提前量.

3　结论

3.1　水华在日内尺度呈夜间积聚、白昼衰减.白昼 WV、AT 升高对表层聚集起抑制作用;夜间低扰动、较低温有利于聚集,体现典型的扰动–热力驱动模式.

3.2　相关性与滞后分析表明,WV、AT 与 Chl.a 显著负相关,滞后 0h 效应最强,属即时驱动型;WT、DO、pH 值、SP 在滞后 12~42h 相关性增强,属迟滞调节型;PAR 在滞后 6h 增强,表现为短时延后.

3.3　三模型对比显示 RF 模型在 R^2、MAE、RMSE 等指标上整体优于 SVR/PLS 模型,适于表征多因子驱动下的非线性关系;基于 RF 模型的平均贡献率排序为 WV 0.271、AT 0.180、SP 0.171、pH 值 0.140、DO 0.126、WT 0.112,与相关/滞后结果一致.

参考文献:

[1] 张中天,袁海钰,许　强,等.太湖叶绿素 a 质量浓度年际变化特征及其对多因子协同的响应 [J]. 水利水电科技进展, 2025,45(3):93-98.
Zhang Z T, Yuan H Y, Xu Q, et al. Interannual variation of chlorophyll-a concentration in Lake Taihu and its response to multiple factors [J]. Advances in Science and Technology of Water Resources and Hydropower, 2025,45(3):93-98.

[2] Sungjin K, Sewoong C.Causal impact analysis of weir opening on cyanobacterial blooms and water quality in the Yeongsan River, Korea: A bayesian structural time-series analysis and median difference test [J]. Science of the Total Environment, 2024,924.

[3] Llanos, Alex A H, Meza, et al. An artificial neural network model for water quality prediction in the amoju hydrographic subbasin, cajamarca-peru [J]. International Journal of Advanced Computer Science and Applications, 2024,15,(9):1021-1032.

[4] 胡怡玲,杨忠勇,冉　薇,等.水流扰动环境中微囊藻群体生长及成群分析 [J]. 中国环境科学, 2025,45(6):3280-3288.
Hu Y L, Yang Z Y, Ran W, et al. Growth and colony formation analysis of Microcystis populations under flow-disturbance conditions [J]. China Environmental Science, 2025,45(6):3280-3288.

[5] Conley D J, Paerl H W, Howarth R W, et al. Ecology controlling eutrophication: Nitrogen and phosphorus [J]. Science, 2009,323(5917): 1014-1015.

[6] 朱广伟,许　海,朱梦圆,等.我国城市水源水库水质风险成因及对策 [J]. 湖泊科学, 2024,36(1):1-16.
Zhu, G W, Xu H, Zhu M Y, et al. Causes and countermeasures of water quality risks in urban drinking water reservoirs in China [J]. Journal of Lake Sciences, 2024,36(1):1-16.

[7] 王思琪.基于多源数据的太湖蓝藻水华风险评估及日内变化时空规律挖掘研究 [D]. 武汉:武汉大学, 2023.
Wang S Q. Risk assessment of cyanobacterial blooms in Lake Taihu based on multi-source data and spatiotemporal patterns of diurnal variation [D]. Wuhan: Wuhan University, 2023.

[8] 刘德富,黄钰铃,纪道斌,等.三峡水库支流水华与生态调度 [M]. 北京:中国水利水电出版社, 201311.429.
Liu D F, Huang Y L, Ji D B, et al. Algal blooms and ecological regulation in the tributaries of the Three Gorges Reservoir [M]. Beijing: China Water & Power Press, 2013:429.

[9] 杨正健,刘德富,马　骏,等.三峡水库支流水华与生态调度新进展 [M]. 北京:中国水利水电出版社, 2021:351.
Yang Z J, Liu D F, Ma J, et al. New advances in tributary algal blooms and ecological regulation in the Three Gorges Reservoir [M]. Beijing: China Water & Power Press, 2021:351.

[10] 崔玉洁.三峡水库香溪河藻类生长敏感生态动力学过程及其模拟 [D]. 武汉:武汉大学, 2017.
Cui Y J. Ecological dynamic processes sensitive to algal growth and their simulation in the Xiangxi River of the Three Gorges Reservoir [D]. Wuhan: Wuhan University, 2017.

[11] Nan L J, Chen L, Guan W, et al.Spatio-Temporal simulation analysis on water environment of typical eutrophic tributaries of Three Gorges Reservoir based on MIKE 21 [C]//4th International Conference on Green Energy, Environment and Sustainable Development, 2023:102-116.

[12] Jungsu P, Keval P, Lee W H. Recent advances in algal bloom detection and prediction technology using machine learning [J]. Science of the Total Environment, 2024,938,173546.

[13] Ma J, Duan H, Chen C, et al. Projected response of algal blooms in global lakes to future climatic and land use changes: Machine learning approaches [J]. Water Research, 2025,271,122889.

[14] Pan X, Yuan J, Yang Z, et al. Spatio-Temporal variation of cyanobacteria blooms in Taihu Lake using multiple remote sensing indices and machine learning [J]. Remote Sensing, 2024,16,(5):889.

[15] Zhang Z j, Zhang H, Jin Y, et al. Analysing the spatiotemporal variation and influencing factors of Lake Chaohu's CDOM over the past 40years using machine learning [J]. Ecohydrology, 2024,17,(3): e2639-e2639.

[16] 成再强.香溪河库湾倒灌异重流演变与水流振荡的混合效应研究 [D]. 宜昌:三峡大学, 2022.
Cheng Z Q. Study on the evolution of backflow density currents and mixed effects of flow oscillations in Xiangxi Bay [D]. Yichang: China Three Gorges University, 2022.

[17] 张家璇.香溪河消落带土壤重金属的生态与健康风险评价及预测 [D]. 宜昌:三峡大学, 2024.
Zhang J X. Ecological and health risk assessment and prediction of soil heavy metals in the drawdown zone of Xiangxi River [D]. Yichang: China Three Gorges University, 2024.

[18] 纪道斌,方　娇,龙良红,等.三峡水库不同支流库湾蓄水期溶解氧分层特性及差异性 [J]. 环境科学, 2022,43(7):3543-3551.
Ji D B, Fang J, Long L H, et al. Stratification characteristics and differences of dissolved oxygen in tributary bays of the Three Gorges

Reservoir during impoundment period [J]. Environmental Science, 2022,43(7):3543−3551.

[19] 杨忠勇,钱门亮,纪道斌,等.香溪河初冬降温过程中垂向混合结构特征分析 [J]. 中国环境科学, 2021,41(6):2862−2870.
Yang Z Y, Qian M L, Ji D B, et al. Characteristics of vertical mixing structure during early winter cooling in the Xiangxi River [J]. China Environmental Science, 2021,41(6):2862−2870.

[20] 龙良红,黄宇擘,徐 慧,等.近 20 年来三峡水库水动力特性及其水环境效应研究:回顾与展望 [J]. 湖泊科学, 2023,35,(2):383−399.
Long L H, Huang Y B, Xu H, et al. Hydrodynamic characteristics of the Three Gorges Reservoir and their water environmental effects over the past 20 years: review and outlook [J]. Journal of Lake Sciences, 2023,35(2):383−397.

[21] Zhang C X, Lei G C, Zhao F X, et al. Functional trait−based phytoplankton biomass and assemblage analyses in the pre−growing season for comprehensive algal bloom risk assessment [J]. Water Research, 2024,257,121755.

[22] 刘德富,郑丙辉,黄应平,等.三峡水库支流水华机理及防控技术 [Z]. 三峡大学, 2014.
Liu D F, Zheng B H, Huang Y P, et al. Mechanisms and control technologies of algal blooms in tributaries of the Three Gorges Reservoir [Z]. China Three Gorges University, 2014.

[23] Yu Z Z, Wang L L. Factors influencing thermal structure in a tributary bay of Three Gorges Reservoir [J]. Journal of Hydrodynamics, 2011, 23(4):407−415.

[24] 国家环保总局.水和废水监测分析方法(第四版) [M]. 北京:中国环境科学出版社, 2002:670−672.
National Environmental Protection Agency. Standard methods for the examination of Water and wastewater (4th ed) [M]. China Environmental Press, 2002:670−672.

[25] Delgado A, Vidal R, Jayme I, et al. Applying grey clustering method and pearson correlation to assess water quality [J]. International Journal of Emerging Technology and Advanced Engineering, 2022,12(8):19−29.

[26] 吴珺玥,赵二刚,郭增良,等.基于 Spearman 系数和 TCN 的光伏出力超短期多步预测 [J]. 太阳能学报, 2023,44(09):180−186.
Wu J Y, Zhao E G, Guo Z L, et al. Ultra−short−term multi−step prediction of photovoltaic power output based on Spearman correlation and TCN [J]. Acta Energiae Solaris Sinica, 2023,44(9):180−186.

[27] Li W, Wang Q W, Yao J F. Distance correlation test for high−dimensional independence [J]. Bernoulli, 2024,30,(4):3165−3192.

[28] 佘星源,何术锋,林育青,等.汉江中下游多目标生态流量时空耦合计算 [J]. 中国环境科学, 2025,45(9):5093−5103.
She X Y, He S F, Lin Y Q, et al. Spatiotemporal coupled calculation of multi−objective ecological flows in the middle and lower reaches of the Hanjiang River [J]. China Environmental Science, 2025,45(9):5093−5103.

[29] Zhao S B, Shi N F, Sun S B , et al. Prediction of tobacco moisture content based on improved support vector regression model [J]. Journal of Sensors, 2024,2024,(1):2167−2178.

[30] 龙良红,黄宇擘,关文海,等.水动力过程驱动的三峡水库澎溪河水华生消机制 [J]. 中国环境科学, 2025,45(6):3245−3255.
Long L H, Huang Y B, Guan W H, et al. Mechanisms of algal bloom dynamics in Pengxi River of the Three Gorges Reservoir driven by hydrodynamic processes [J]. China Environmental Science, 2025,45 (6):3245−3255.

[31] 陈家炜.水源水库扬水曝气运行对藻类和有机物的影响研究 [D]. 西安:西安建筑科技大学, 2018.
Chen J W. Study on the effects of pumped aeration operation on algae and organic matter in source reservoir [D]. Xi'an: Xi'an University of Architecture and Technology, 2018.

[32] Baho D L, Leu E, Pomati F, et al. Resilience of natural phytoplankton communities to pulse disturbances from micropollutant exposure and vertical mixing [J]. Environ Toxicol Chem, 2019,38,(10):2197−2208.

[33] DB44/T2261−2020 水华程度分级与监测技术规程 [S].
DB44/T2261−2020 Technical specification for classification and monitoring of algal blooms [S].

[34] Gao C, Liang Z J, Xin P L, et al. Identification of key water environmental factor contributions and spatiotemporal differential characteristics for eutrophication in Dianchi Lake [J]. Environmental Monitoring and Assessment, 2024,196,(12):1217.

[35] Jiang Y Y, Song Y Y, Liu J L, et al. Machine learning assisted precise prediction of algae bloom in large−scale water diversion engineering [J]. Desalination, 2025,610.

[36] Jokiel, Paul L J, Christopher P R, Ku'ulei S. Coral−algae metabolism and diurnal changes in the CO_2−carbonate system of bulk sea water [J]. Peerj, 2014,2,(1):378.

[37] 薛宗璞,朱 伟,柏 松,等.太湖微囊藻漂移—聚集规律——基于风与光共同作用的模拟 [J]. 湖泊科学, 2024,36,(6):1660−1671.
Xue Z P, Zhu W, Bai S, et al. Simulation of drift–aggregation patterns of Microcystis in Lake Taihu under combined effects of wind and light. [J]. Journal of Lake Sciences, 2024,36(6):1660−1671.

[38] 罗晓春,杭 鑫,曹 云,等.太湖富营养化条件下影响蓝藻水华的主导气象因子 [J]. 湖泊科学, 2019,31(5):1248−1258.
Luo X C, Hang X, Cao Y, et al. Dominant meteorological factors affecting cyanobacterial blooms under eutrophic conditions in Lake Taihu [J]. Journal of Lake Sciences, 2019,31(5):1248−1258.

[39] Jiang Mingshun C, Jennifer M. Modeling water quality and cyanobacteria blooms in Lake Okeechobee: II. Dynamics of diurnal cycles and impacts of cyanobacteria diel vertical migration [Z]. Ecological Modelling, 2025,505.

[40] Madyouni Haifa M, Pol B. Simulation of thermal stratification and water temperature dynamics in the Joumine reservoir (Tunisia) [Z]. Ecological Informatics, 2025,86.

[41] 张 群,许旭明,陈 倩.潮白河流域水库浮游动物群落季节变化特征及其驱动因素 [J]. 北京大学学报(自然科学版), 2023,59(2): 290−300.
Zhang Q, Xu X, Chen Q. Seasonal variation and driving factors of zooplankton communities in reservoir of the Chaobai River Basin [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2023,59(2): 290−300.

[42] Chen Y C,Yue Y H, Wang J, et al. Microbial community dynamics and assembly mechanisms across different stages of cyanobacterial bloom in a large freshwater lake [J]. Science of the Total Environment, 2024,907,168207.

作者简介: 杨 京(2000−),男,四川南充人,三峡大学硕士研究生,主要从事湖库水华发生机制及其控制研究.发表论文 3 篇.Jing6171@qq.com.