

DOI: 10.20040/j.cnki.1000-7709.2023.20230325

武汉南湖水华生消关键因子识别及临界值

刘振旗^{1,2}, 李建¹, 江善虎², 华一安², 杨克非³

(1. 长江水资源保护科学研究所, 湖北 武汉 430000; 2. 河海大学水文水资源学院, 江苏 南京 210000;
3. 华北水利水电大学, 河南 郑州 450000)

摘要: 识别水华生消关键影响因子及其临界值对湖泊水华防控具有重要意义。以武汉市南湖为研究区, 基于水文、水质和气象监测数据, 采用皮尔逊相关分析、灰色关联分析、主成分分析方法, 筛选并确定与叶绿素 a 浓度相关性较大的关键环境因子, 利用受试者工作特征(ROC)曲线法分析关键环境影响因子的临界值。结果表明, COD_{Mn}、总磷、总氮、水温、气温与叶绿素 a 浓度具有较强的相关关系, 是水华暴发的关键因子, 其临界值分别为 6.25 mg/L、0.146 mg/L、0.725 mg/L、27.05 °C、22.35 °C。研究发现, ROC 曲线法对 COD_{Mn}、总磷、总氮等日变幅较小环境因子临界值的求解精度优于气温、水温等日变幅较大环境因子, 在湖泊水华暴发的营养盐浓度临界值确定方面具有较大优势。

关键词: 水华; 关键环境因子; 临界值; ROC 曲线法

中图分类号: TV213.3

文献标志码: A

文章编号: 1000-7709(2023)12-0041-04

1 引言

对于富营养化的湖泊和水库, 当具备适当的水温、光照等条件时极易暴发水华^[1]。水华在我国太湖、巢湖、滇池等重要湖泊频繁出现, 对湖泊生态系统健康造成重大影响。由于湖泊富营养化程度、所处的富营养化阶段, 以及所在地区气候等存在差异, 不同湖泊水华暴发的环境影响因子临界值并不一致, 导致临界值确定方法存在适用性问题^[2]。因而, 水华暴发环境影响因子的临界值一直是水华研究的热点。但已有研究多集中在实验室或某特定围隔区或以对经验值判断为主^[3-8], 这与野外实际条件相差较大, 所求临界值准确度较低。因此, 本文采用受试者工作特征(ROC)曲线分析法对比分析环境因子区间内所有数值, 最后求出临界值。该方法相较于已有求解临界值的方法, 所求环境影响因子的临界值普适性较强、准确度更高, 以期对武汉南湖水华的控制及预警提供科学依据和技术参考。

2 研究方法

利用皮尔逊相关分析、主成分分析、灰色关联

分析方法来分析环境因子与叶绿素 a 相关性的大小, 筛选出水华暴发的关键因子^[9]。其中皮尔逊相关分析、主成分分析法通过 SPSS20.0 软件分析完成, 灰色关联分析法通过 Excel2019 计算完成。

ROC 曲线法是敏感度和特异度相结合的综合评价判别准确性的一种方法^[10]。ROC 曲线分析法确定界值点时, 一个界值点对应一个敏感度和特异度, 界值点处于指标最大值与最小值之间, 选取众多界值点进行比较, 通常采用约登指数最大原则确定最优界值点, 即临界值点。

以叶绿素 a 浓度 30 μg/L 作为水华界值点, 即叶绿素 a 浓度 ≥ 30 μg/L, 同时影响因素大于界值点数值预测为水华暴发。设 A 为实际发生水华且预测也为发生水华事件总数, B 为实际未发生水华而预测发生水华事件总数, C 为实际发生水华事件而预测为未发生水华事件总数, D 为实际未发生水华事件且预测也为未发生水华事件总数。以环境因子水温和叶绿素 a 浓度分布图(图 1)为例, 对 ROC 方法作进一步解释。A 代表图 1 中右上部分点的个数, 即实际水华暴发预测也为水华暴发的事件数; B 代表图 1 中右下部分点的个数, 即预测为水华暴发, 但由于叶绿素 a 浓度小

收稿日期: 2023-03-06, 修回日期: 2023-04-03

基金项目: 湖北省自然科学面上基金项目(2022CFB374); 国家自然科学基金重点项目(U2040210)

作者简介: 刘振旗(1998-), 男, 硕士研究生, 研究方向为水文水资源, E-mail: 1352154392@qq.com

通讯作者: 李建(1985-), 男, 教授级高级工程师, 研究方向为水文水资源, E-mail: lijian2750@foxmail.com

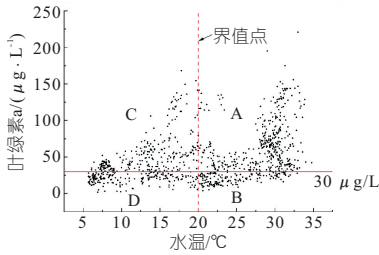


图 1 ROC 解释图 (以水温为例)

Fig. 1 ROC interpretation chart (in terms of water temperature)

于 30 µg/L, 实际未发生水华事件个数; C、D 同理。

设 T_{TPR} 为 ROC 曲线纵坐标 (敏感度), 即预测值与真实值都是水华事件数量占真实水华事件总数的比例。 T_{TPR} 越大, 预测效果越好。 T_{TPR} 计算公式为:

$$T_{TPR} = A / (A + C) \quad (1)$$

设 F_{FPR} 为 ROC 曲线横坐标 (1-特异度, 表示 1 与特异性的差值), 即预测值是水华而真实值不是水华事件占真实值无水华事件总数的比例, F_{FPR} 越小, 预测效果越好。 F_{FPR} 计算式为:

$$F_{FPR} = B / (B + D) \quad (2)$$

ROC 曲线下面积 (AUC) 代表准确率, 为判断判别效果的依据, 面积越大, 临界值准确度越高; 曲线上越靠近左上角的点所对应的约登指数越大, 判断临界值效果越好。约登指数最大点对应判断临界值点。

3 武汉南湖水华生消关键因子识别及临界值

3.1 关键因子筛选

对影响因素和叶绿素 a 浓度进行皮尔逊相关性分析, 结果见表 1。从单个因素角度分析, 与武汉南湖叶绿素 a 浓度相关性较大的驱动因子为 COD_{Mn} ($r=0.71, p \leq 0.01$)、总磷 ($r=0.63, p \leq 0.01$)、总氮 ($r=0.61, p \leq 0.01$)、水温 ($r=0.39, p \leq 0.01$)、气温 ($r=0.30, p \leq 0.01$)、溶解氧 ($r=-0.29, p \leq 0.01$)、浊度 ($r=0.28, p \leq 0.01$)、气压 ($r=-0.23, p \leq 0.01$)、电导率 ($r=-0.23, p \leq 0.01$)、氨氮 ($r=0.100, p \leq 0.01$)、湿度 ($r=0.076, p \leq 0.05$)、pH 值 ($r=0.068, p \leq 0.05$)、水位 ($r=-0.061, p \leq 0.05$)。

营养物质因素对叶绿素 a 浓度起决定性的影响, COD_{Mn} 、总磷、总氮与叶绿素 a 浓度的相关性分别为 0.71、0.63、0.61, 均大于 0.6, 呈显著正相关。此外, 气温大小影响水温的高低, 进而影响藻类的生长速率, 所以气温和水温与叶绿素 a 浓度

表 1 水华暴发关键因子筛选

Tab. 1 Screening of key factors for water bloom outbreak

皮尔逊相关分析		灰色关联度		主成分分析			
指标	显著性	指标	显著性	指标	主成分 1	主成分 2	指标权重/%
COD_{Mn}	0.71**	总氮	0.998 31	气温	0.359 64	0.308 80	50.37
总磷	0.63**	总磷	0.997 83	水温	0.402 46	0.172 53	45.93
总氮	0.61**	COD_{Mn}	0.996 75	COD_{Mn}	0.301 39	0.091 55	32.12
水温	0.39**	水温	0.996 37	pH	-0.078 80	0.537 40	25.47
气温	0.30**	气温	0.996 11	总磷	0.315 74	-0.110 70	21.18
溶解氧	-0.29**	氨氮	0.995 59	总氮	0.282 92	-0.114 50	18.06
浊度	0.28**	湿度	0.994 20	浊度	0.217 15	-0.063 70	15.32
气压	-0.23**	风向	0.993 81	能见度	0.043 25	0.088 00	9.13
电导率	-0.23**	pH	0.993 73	风速	0.069 46	0.048 87	9.08
氨氮	0.100**	风速	0.993 52	降雨量	0.065 02	-0.091 70	0.21
湿度	0.076*	水位	0.993 20	风向	-0.021 40	0.027 91	-0.20
pH	0.068*	气压	0.993 18	溶解氧	-0.281 70	0.392 35	-1.18
水位	-0.061*	电导率	0.992 88	氨氮	0.187 72	-0.373 40	-5.97
降雨量	0.059	浊度	0.992 71	湿度	0.040 77	-0.407 20	-20.97
风向	-0.05	溶解氧	0.991 81	水位	-0.232 60	-0.007 90	-21.00
风速	0.033	能见度	0.990 23	电导率	-0.273 00	-0.227 50	-37.81
能见度	0.022	降雨量	0.982 18	气压	-0.358 50	-0.134 20	-39.74

注: 主成分分析中主成分 1、2 (PC1、PC2) 特征值分别为 5.137 56、2.395 39, 解释率分别为 30.20%、14.10%。

也有较强的相关性。溶解氧、浊度、气压、电导率、氨氮、湿度、pH 值、水位与叶绿素 a 浓度相关性绝对值在 0.05~0.30 之间, 相关性较低, 但仍显著相关。降雨量、风速、风向、能见度与叶绿素 a 浓度不具有显著相关性。

湖泊水华成因复杂, 影响因素众多, 通过皮尔逊相关分析法所得结果存在一定不确定性。为保证筛选结果的准确性, 再通过灰色关联分析求取南湖水华发生的关键驱动因素。由表 1 可知, 灰色关联分析结果与皮尔逊相关分析结果类似, 总氮、总磷、 COD_{Mn} 分别占据了前 3 的位置, 灰色关联度分别为 0.998 31、0.997 83、0.996 75, 营养物质对南湖水华的暴发有关键性的影响。水温、气温与叶绿素 a 浓度的关联度亦较大, 气温直接影响水温, 水温变化影响浮游植物的生长过程。氨氮、湿度、风向、pH 值、风速、水位、气压、电导率、浊度、溶解氧、能见度与叶绿素 a 浓度的关联度均达到 0.99 以上。降雨量与叶绿素 a 浓度的关联度最小为 0.982 18。综合分析可知, 各因素均与叶绿素 a 浓度关联度较大, 进一步说明是多种因素共同作用造成水华暴发。

对环境因子进行主成分分析, 结果见图 2。第 1、2 主成分分别解释了数据量的 30.2%、14.1%, 观察主成分分析中坐标较大的环境因子, 同时根据主成分的特征值和解释率计算各因素的指标权重。

气温所占指标权重最大为 50.37%; 水温受气温的直接影响, 指标权重也较大, 占比为 45.93%, 营养物质指标权重相对次之, COD_{Mn} 、总

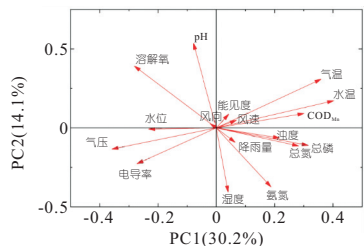


图 2 环境因子主成分分析结果

Fig. 2 Results of principal component analysis of environmental factors

磷、总氮指标权重分别为 32.12%、21.18%、18.06%。与灰色关联分析、皮尔逊相关分析不同的是,营养物质因素并非最重要的因素,气温成为影响水华暴发最重要的因素;与前两种方法相比,pH 值在影响因素重要性排名中明显靠前,指标权重为 25.47%。浊度指标权重为 15.32%,能见度和风速指标权重相近,都在 9%左右。以上指标权重均大于 9%,相对较大,而降雨量、风向、溶解氧、氨氮、湿度、水位、电导率、气压指标权重较小,都小于 1%。计算结果见表 1。

虽然 3 种方法影响因素按大小排序各不相同,但皮尔逊相关分析、灰色关联分析、主成分分析结果中 COD_{Mn}、总磷、总氮、水温、气温均占影响因素排名前 6 位,重要性排名相差不大。溶解氧通过皮尔逊相关分析重要性排第 6,而通过灰色关联分析和主成分分析分别排第 15、12 位,3 种方法分析所得结果不稳定性较大,不作为水华暴发的关键因子,同理氨氮、pH 值、浊度、气压、电导率、湿度、水位、降雨量、风向、风速、能见度与溶解氧相同,通过不同方法得出因子重要性排名相差较大,不稳定性较强。同时其他因素与 COD_{Mn}、总磷、总氮、水温、气温相比与叶绿素 a 相关性相对较低。综上所述,筛选后水华暴发的关键因子为 COD_{Mn}、总磷、总氮、水温、气温。

3.2 ROC 曲线求临界值

采用 ROC 曲线法求解水华暴发关键因子 COD_{Mn}、总磷、总氮、水温、气温的临界值。给定 COD_{Mn}、总磷、总氮、水温、气温不同的值,分别求取各因子的敏感度(TPR)、1-特异度(FPR)见表 2,绘制 ROC 曲线见图 3。ROC 曲线分析各因子

表 2 ROC 曲线分析关键因子结果

Tab. 2 Results of key factors of ROC curve analysis

指标	AUC	95%置信区间	临界值	敏感度	特异度	显著性
水温	0.681	0.648~0.713	27.05	0.455	0.882	0.00
COD _{Mn}	0.770	0.741~0.798	6.25	0.594	0.802	0.00
总氮	0.720	0.689~0.751	0.725	0.633	0.681	0.00
总磷	0.841	0.818~0.863	0.146	0.671	0.912	0.00
气温	0.632	0.597~0.667	22.35	0.51	0.723	0.00

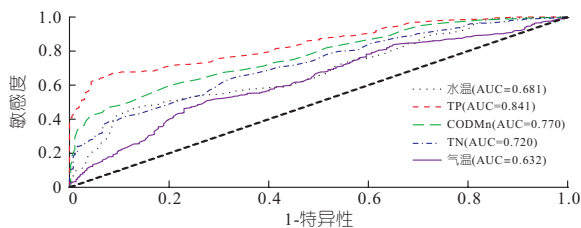


图 3 关键因子 ROC 曲线图

Fig. 3 ROC curves chart of key factors

AUC 值均大于 0.5 且显著性小于 0.05,因而各因素临界值具有统计学意义。通过敏感度和 1-特异度求取各影响因素的约登指数,选取约登指数最大点对应的数值即为该影响因素的临界值。

水温、COD_{Mn}、总氮、总磷、气温的敏感度分别为 0.455、0.594、0.633、0.671、0.51,特异度分别为 0.882、0.802、0.681、0.912、0.723。总磷临界值准确度最高,最大约登指数为 0.583,对应的临界值为 0.146 mg/L,曲线下面积 AUC 为 0.841,即临界值的准确度为 84.1%。COD_{Mn} 对应最大约登指数为 0.396,水华暴发时临界值为 6.25 mg/L,准确度为 77%。总氮最大约登指数为 0.314,约登指数对应的临界值为 0.725 mg/L,准确度为 72%。水温最大约登指数为 0.337,最大约登指数对应临界值水温为 27.05 °C,曲线下 AUC 面积为 0.681,相应的其准确度为 68.1%。气温约登指数最大值为 0.233,临界值气温为 22.35 °C,准确率为 63.2%。

对于不同富营养化程度的湖泊或同一湖泊的不同富营养化阶段、湖泊所在地区气候、污染物集中输入、风速风向和波浪扰动等,这些因素都对水华临界值确定有影响。试验分析多采用控制条件法,环境条件较稳定,这与实际野外环境条件存在较大差异,水华与环境因子的响应关系有所不同。因而,水华暴发临界值与实验室确定临界值存在较大差别。

为验证所求各影响因素临界值准确性,以叶绿素 a 浓度 30 μg/L 为水华暴发的临界值,选取监测数据中叶绿素 a 浓度处于 20~40 μg/L 的数据,即水华处于即将暴发和刚暴发时的监测数据,求取各影响因素的平均值与所求临界值进行对比。叶绿素 a 浓度处于 20~40 μg/L 时,总氮、总磷、COD_{Mn} 浓度的平均值分别为 0.68、0.12、5.59 mg/L。而所求总氮、总磷、COD_{Mn} 临界值浓度为 0.725、0.146、6.25 mg/L。通过比较,二者相差不大,进一步证实所求总氮、总磷、COD_{Mn} 临界值具有较高的准确性。而叶绿素 a 浓度处于 20~40 μg/L 时水温平均值为 18.19 °C,气温平均值

为 16.25 °C，所求水温、气温临界值分别为 27.05、22.35 °C。水温和气温所求临界值与实际值差别较大，对比结果见表 3。

表 3 ROC 曲线法与 20~40µg/L 叶绿素 a 浓度区间对应环境因子平均值对比

Tab.3 Comparison of the ROC curve method with the mean values of environmental factors corresponding to the 20~40 µg/L chlorophyll a concentration interval

指标	ROC 曲线 临界值	20~40 µg/L 叶绿素 a 浓度 区间对应环境因子平均值
COD _{Mn} /(mg·L ⁻¹)	6.250	5.59
总氮/(mg·L ⁻¹)	0.725	0.68
总磷/(mg·L ⁻¹)	0.146	0.12
气温/°C	22.350	16.25
水温/°C	27.050	18.19

考虑到气温不可控性较强且季节、日变幅较大，水温受气温变化的直接影响，变化幅度也较大。武汉市南湖属于亚热带季风气候，四季分明、温差相对较大，导致所求气温和水温准确度较低，临界值与平均值相差亦较大。而 COD_{Mn}、总磷、总氮受人为影响较大，可控性较强，COD_{Mn}、总磷、总氮浓度全年变化幅度不大，通过 ROC 曲线所求 COD_{Mn}、总磷、总氮临界值浓度准确度较高，与 20~40µg/L 叶绿素 a 浓度区间对应的平均值相差不大，人为控制可使湖泊营养盐浓度低于水华暴发的临界值，进而控制水华的暴发。

4 结论

- a. 采用皮尔逊相关分析、灰色关联分析、主成分分析 3 种分析方法相互验证筛选出南湖水华暴发的关键因子为 COD_{Mn}、总磷、总氮、水温、气温。
- b. 基于 ROC 曲线得出防止水华暴发的关键因子 COD_{Mn}、总磷、总氮临界值分别为 6.25、0.146、0.725 mg/L，准确度分别为 77.0%、84.1%、72.0%，准确度较高。关键因子气温和水温的临界值分别为 22.35、27.05 °C，准确度分别为 63.2%、68.1%，准确度略低，主要原因是气温

日变幅较大，水温直接受气温的影响。说明 ROC 曲线法对 COD_{Mn}、总磷、总氮等日变幅较小环境因子临界值的求解精度优于气温、水温等日变幅较大环境因子，ROC 曲线法在湖泊水华暴发的营养盐浓度临界值确定方面具有较大优势。

参考文献：

- [1] 刘文杰. 内陆湖泊蓝藻水华的遥感监测与评价研究[D]. 北京:中国地质大学(北京), 2013.
- [2] 曹晶,田泽斌,储昭升,等. 太湖蓝藻水华暴发的氮磷控制阈值分析[J]. 湖泊科学,2022,34(4):1075-1089.
- [3] 贺能琴,聂菊芬. 滇池北部围隔区蓝藻水华暴发主要影响因素及交互作用研究[J]. 环境科学导刊, 2017, 36(4): 18-21.
- [4] 杨柳燕,揣小明,刘涛,等. 太湖北部草藻转换围隔试验研究[C]//首届中国湖泊论坛论文集. 2011: 371-374.
- [5] WU Y L,XU H,YANG G J, et al. Developing the critical phosphorus threshold for spring algal growth in Lake Taihu, China[J]. China environmental science,2013,33(9):1622-1629.
- [6] XU H, PAERL H W, QIN B Q, et al. Determining critical nutrient thresholds needed to control harmful cyanobacterial blooms in eutrophic lake Taihu, China[J]. Environmental science and technology, 2015, 49(2): 1051-1059.
- [7] 温成成,黄廷林,孔昌昊,等. 北方富营养分层型水库藻类季节性暴发机制及其阈值分析[J]. 环境科学,2023,44(3):1452-1464.
- [8] 田晶,郭生练,王俊,等. 汉江中下游干流水华关键环境因子识别及阈值分析[J]. 水资源保护,2022,38(5):196-203.
- [9] 纪德洋,金锋,冬雷,等. 基于皮尔逊相关系数的光伏电站数据修复[J]. 中国电机工程学报, 2022, 42(04): 1514-1523.
- [10] 任晶晶. 药物性胆汁淤积肝胆生化指标的对比研究及 ROC 曲线分析[D]. 晋中:山西医科大学,2021.

Identification of Key Factors and Critical Values for Production and Elimination of Water Bloom in Wuhan South Lake

LIU Zhen-qi^{1,2}, LI Jian¹, JIANG Shan-hu², HUA Yi-an², YANG Ke-fei³

(1. Changjiang Water Resources Protection Institute, Wuhan 430000, China; 2. College of Hydrology and Water Resources, Hohai University, Nanjing 210000, China; 3. North China University of Water Resources and Electric Power, Zhengzhou 450000, China)

Abstract: The identification of key influencing factors and their critical values for the prevention and control of water bloom in lakes is of great significance. Taking Wuhan South Lake as the study area, based on hydrological, water quality and meteorological monitoring data, Pearson Correlation Analysis, Gray Correlation Analysis and Principal Component Analysis were used to screen and identify key environmental factors with high correlation with chlorophyll a concentration, and the critical values of key environmental impact factors were analyzed by the Receiver Operating Characteristic (ROC) curve method. The results show that the COD_{Mn}, total phosphorus, total nitrogen, water temperature and air temperature had strong correlations with chlorophyll a concentration and were the key factors for water bloom outbreak, with the critical values of 6.25 mg/L, 0.146 mg/L, 0.725 mg/L, 27.05 °C and 22.35 °C, respectively. It was found that the accuracy of the ROC curve method for solving the critical values of COD_{Mn}, total phosphorus, total nitrogen and other environmental factors with small daily variation was better than that of environmental factors with large daily variation such as air temperature and water temperature, and the ROC curve method has a greater advantage in determining the critical values of nutrient concentrations for water blooms in lakes.

Key words: water bloom; key environmental factors; critical value; ROC curve method