

刘振旗, 李建, 江善虎. 基于 Almon-BP 时滞神经网络模型的叶绿素 a 趋势预测[J]. 南水北调与水利科技(中英文), 2024, 22(6): 1137-1147. LIU Z Q, LI J, JIANG S H. Chlorophyll-a prediction based on the Almon-BP delayed neural network model[J]. South-to-North Water Transfers and Water Science & Technology, 2024, 22(6): 1137-1147. (in Chinese)

基于 Almon-BP 时滞神经网络模型的 叶绿素 a 趋势预测

刘振旗¹, 李建², 江善虎¹

(1. 河海大学水文水资源学院, 南京 210000; 2. 长江水资源保护科学研究所, 武汉 430000)

摘要: 基于数据驱动的机器学习方法是预测叶绿素 a 的一种重要的非机理方法, 但是现有以神经网络模型为代表的叶绿素 a 预测方法较少考虑藻类生消对环境因子的时滞效应。以三峡库区叶绿素 a 质量浓度较高的典型支流香溪河为研究区域, 采用相关性分析、主成分分析和灰色关联分析 3 种方法综合确定叶绿素 a 的主要贡献因子, 利用交叉相关分析及 Almon 分布时滞模型筛选主要贡献因子中的时滞因子, 并确定其最优滞后时间, 在此基础上构建 Almon-BP 时滞神经网络模型进行叶绿素 a 趋势预测。结果表明: 香溪河峡口主要时滞因子为气温、风速、太阳辐射、pH、溶解氧, 最优滞后时间分别为 4、2、3、7、3 d; 香溪河平邑口主要时滞因子为水温、气温、风速、降雨量、太阳辐射、浊度、pH、溶解氧、三峡水位差、氧化还原电位、三峡水位, 最优滞后时间分别为 2、2、2、4、2、10、3、2、6、10、6 d; 相比于只考虑叶绿素 a 主要贡献因子的 BP 神经网络模型(贡献因子-BP), Almon-BP 时滞神经网络模型对香溪河峡口预测结果的均方误差(E_{MS})、均方根误差(E_{RMS})、平均绝对误差(E_{MA})、平均偏差(E_{MB})等误差指标分别降低 44.4%、25.6%、31.3%、53.9%, 对香溪河平邑口预测结果的 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 等误差指标分别降低 66.7%、42.1%、37.5%、45.8%。研究模型对叶绿素 a 预测具有较高的准确度, 对叶绿素 a 早期预警具有重要意义。

关键词: 叶绿素 a 预测; 主要贡献因子; 时滞效应; 最优滞后时间; Almon-BP 模型

中图分类号: TV213.3 **文献标志码:** A **DOI:** 10.13476/j.cnki.nsbdqk.2024.0113

三峡水库建成运行后造就了一个新的生态系统。经过十多年运行, 三峡水库水质总体较好, 但部分支流库湾存在富营养化和叶绿素含量升高问题^[1-2]。2003 年水库运行以来, 典型支流叶绿素 a 含量升高影响三峡库区水质安全和水生态系统健康^[3], 开展叶绿素 a 预测研究对保障三峡库区水质和水生态安全具有重要意义。

目前研究^[4-7]普遍认为充足的营养物质、适宜的气候条件、缓慢的水流条件是叶绿素 a 含量升高的主要原因。三峡水库蓄水以来干支流温度差和泥沙含量差导致支流库湾处普遍存在分层异重流, 三峡库区蓄水导致流速变缓、干流倒灌导致支流库湾营养盐含量升高以及在分层异重流作用下混合层与临界层深度关系变化^[8-9]。库区支流环境因子(营养盐、

水动力、气象)对叶绿素 a 含量的影响一直是三峡库区研究的热点^[10-13]。叶绿素 a 升高受环境因子的影响存在一定的时间差。Jeong 等^[14]、Kim 等^[15]通过研究洛东江韩式冠盘藻和铜绿微囊藻的含量与降水量和大坝下泄量关系发现降水量对藻类含量的影响甚至可以延长到 2 a 之久。Liu 等^[16]研究得出当旬及前 1 旬的水位差和水温对叶绿素 a 含量的升高影响最为显著, 叶绿素 a 含量与水位差和水温之间存在着较为显著的时滞关系。许多学者对叶绿素 a 趋势进行预测研究。Kim 等^[17]提出一种多变量隐马尔可夫模型(MHMM), 该模型可以分析韩国洛东江叶绿素 a 含量的时空依赖性。Xia 等^[18]通过梯度提升机与 K 倍交叉验证证明前 10 d 的环境因子可以更好地预测叶绿素 a 趋势。Kim 等^[19]使用 EFDC 水动

收稿日期: 2024-03-01 修回日期: 2024-06-24 网络出版时间: 2024-07-18

网络出版地址: <https://link.cnki.net/urlid/13.1430.TV.20240715.1703.004>

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

作者简介: 刘振旗(1998—), 男, 河南濮阳人, 主要从事水文水资源研究。E-mail: 1352154392@qq.com

通信作者: 李建(1985—)男, 湖北武汉人, 正高级工程师, 博士研究生, 主要从事水文水资源研究。E-mail: lijian2750@foxmail.com

力和水质模型预测韩国汉江下游藻华。不同区域叶绿素 a 含量升高滞后时间存在差异,目前少有研究定量指出三峡库区支流环境因子具体滞后时间,为避免叶绿素 a 含量升高,有必要对三峡库区支流叶绿素 a 驱动因子滞后时间进行定量分析。同时,叶绿素 a 时滞效应的求解对野外叶绿素 a 观测具有很好的解释作用,考虑时滞效应下叶绿素 a 趋势模拟预测对浮游藻类过度生长的预警防控具有重要意义。

本研究基于神经网络算法构建一种考虑时滞效应的叶绿素 a 预测模型,提高叶绿素 a 预测精度。首先基于相关性分析、灰色关联度、主成分分析筛选出香溪河叶绿素 a 主要贡献因子,然后通过交叉相关分析确定存在时滞效应的主要贡献因子,并利用 Almon 分布时滞模型求出最优滞后时间,在此基础上构建 Almon-BP 时滞神经网络模型进行叶绿素 a 趋势预测。

1 研究区概况

香溪河是三峡水库湖北库区,同时也是靠近三峡大坝的最大支流。香溪河干流长 94 km,流域范围 $110^{\circ}25'E\sim 111^{\circ}06'E$ 、 $30^{\circ}57'N\sim 31^{\circ}34'N$ ^[20]。河口距三峡大坝 34.5 km,河流由北向南,但在与长江干流交汇区河口转向为西南方向,与由西向东的水库干流成 45° 夹角,成为流向干流上游的反向交汇出

口,其特殊的河口走向为干流水体进入支流提供充足动力,使得香溪河库湾受干流倒灌强度影响显著强于其他支流^[21]。

香溪河流域属亚热带大陆性季风气候,春季冷暖多变,夏季雨量集中,常有暴雨和伏旱,秋季多阴雨,冬季多雨雪。由于地势高差大,地形复杂,气温垂直变化明显,具有极为显著的区域小气候特征^[22]。多年平均气温 $16.6^{\circ}C$,年均降水量为 $1\ 015.6\ mm$,降水时空分布不均,冬季干旱,夏季多雨,据兴山水文站(控制流域面积 $1\ 900\ km^2$)记载,香溪河多年平均流量 $40.18\ m^3/s$ ^[23]。枯水期最小流量为 $14\ m^3/s$ 。受三峡水库回水影响,香溪河库湾最大水深为 $110\ m$,水面均宽 $670\ m$,水体断面流速不到 $0.005\ m/s$ 。流域风向以西南风、东南风、东北风为主,最大瞬时风速可达 $34\ m/s$ 。适宜的气象水文条件以及干流倒灌给支流库湾提供充足的营养盐导致香溪河多年来叶绿素 a 含量偏高。研究区域见图 1。采用 2022 年 5 月—2023 年 5 月香溪河峡口、平邑口浮标站连续监测数据,香溪河平邑口监测指标为电导率、水温、pH、氧化还原电位、溶解氧、浊度、叶绿素 a 等,峡口监测指标为气温、气压、风速、风向、降雨量、太阳辐照、电导率、水温、pH、氧化还原电位、溶解氧、浊度、叶绿素 a 等。监测频次均为 $2\ h$ 1 次。数据来源于三峡集团和长江水资源保护科学研究所。

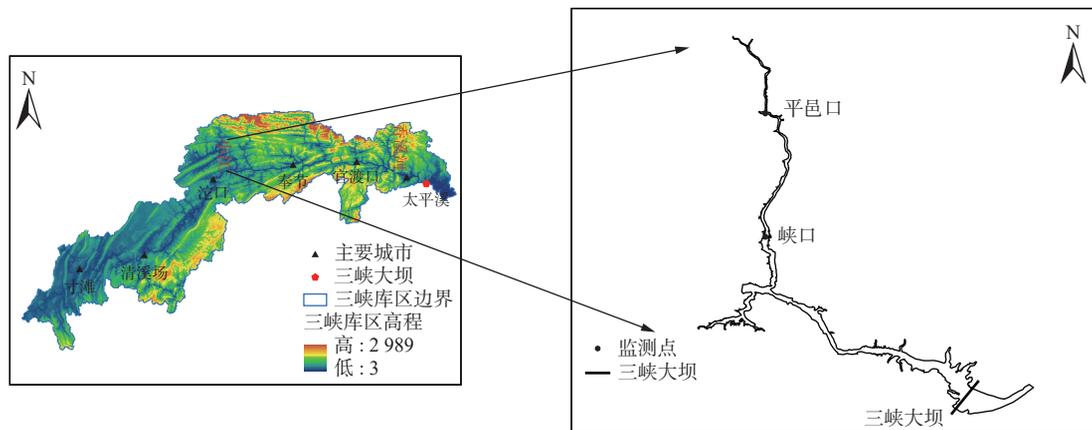


图 1 香溪河研究区域

Fig. 1 Map of the Xiangxi River study area

2 数据与方法

为消除多重共线性的影响,美国经济学家阿尔蒙(Almon)于 1965 年在数学分析中多项式逼近定理的基础上提出了多项式分布滞后的研究方法^[24-27]。该方法利用有限多项式来减少待估计参数的个数,以降低环境因子之间多重共线性。

Almon 分布滞后模型可表示为

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_k x_{t-k} + \gamma \quad (1)$$

式中: y 为因变量; $\alpha, \beta_0, \beta_1, \dots, \beta_k$ 为模型系数。 $x_t, x_{t-1}, \dots, x_{t-k}$ 为滞后某时刻的解释变量; γ 为随机干扰项; k 为滞后阶数,模型中既包括环境因子当前值,还包含环境因子过去值(滞后值)。

$$\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3 + \dots + a_m i^m \quad (2)$$

有限分布滞后模型中的参数 $\beta_i (i=0, 1, 2, \dots, k)$ 的分布近似用一个关于滞后期 i 的低阶多项式表示, 就可以利用多项式减少模型中的参数。 $\beta_i (i=0, 1, 2, \dots, k)$ 为延迟乘数, 表示过去各个时期 x 变动一个单位对 y 值影响程度。一般多项式次数 m 取 2 或者 3, 且多项式的次数 m 小于最大滞后期 k 。

$$\text{令 } Z_{mt} = \sum_{i=0}^k i^m x_{t-i}$$

$$y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + \dots + a_m Z_{mt} + \gamma \quad (3)$$

该模型仍可采用最小二乘法估计, 且 α 、 a_i 的估计值具有全部的优良统计性质。求出 a_i 值进而求出原始的 β 系数。

常用最优阶数的判定准则为 \bar{R}^2 (调整后 R^2) 修正自由度的决定系数、AIC、SC 准则, 计算公式为

$$\text{调整后 } R^2: \bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1} \quad (4)$$

$$\text{AIC 准则: } C_{AI} = -2 \frac{\ln S_{RS}}{n} + 2 \frac{k}{n} \quad (5)$$

式中: n 为个例数; S_{RS} 为残差平方和; k 为因子数。

$$\text{SC 准则: } C_S = -2 \frac{\ln S_{RS}}{n} + k \frac{\ln n}{n} \quad (6)$$

依次添加滞后项, 直到 \bar{R}^2 达到最大值, C_{AI} 、 C_S 达到最小值, 对应为最优滞后时间。

基于时滞效应的 BP 神经网络 (Almon-BP) 模型构建, 首先通过 Almon 模型对原始环境因子数据处理, 分析贡献度较大环境因子对叶绿素 a 含量升高的时滞效应, 将处理好的数据输入 BP 预测模型^[28-30]。Almon-BP 模型结构原理图见图 2。

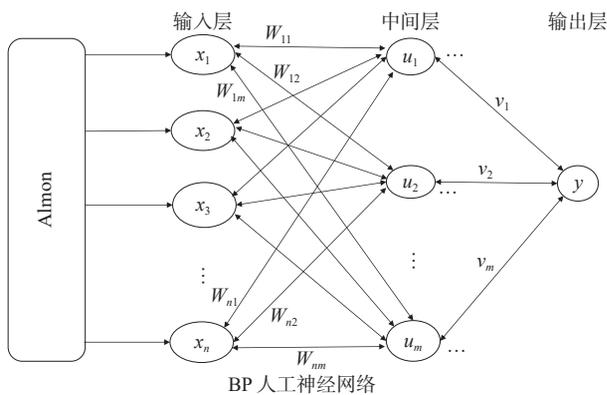


图 2 Almon-BP 模型结构原理
Fig. 2 Schematic diagram of Almon-BP model

图 2 中 x_1, x_2, \dots, x_n 为输入变量; y 为输出变量; u_i 为隐藏层神经元输出; f 为激活函数映射关系为 Sigmoid 函数; w_{ij} 为第 i 个输入变量与第 j 个隐藏层神经元的权重; θ_j^u 为隐含层 u 第 j 个神经元的偏置

项, 即相当于截距。BP 神经网络计算公式为

$$f(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

$$u_j = f\left(\sum_{i=1}^n w_{ij} x_i + \theta_j^u\right) \quad j = 1, 2, \dots, m \quad (8)$$

$$y = f\left(\sum_{j=1}^m v_j u_j + \theta^y\right) \quad (9)$$

利用 Almon-BP 模型开展叶绿素 a 趋势预测的流程见图 3。

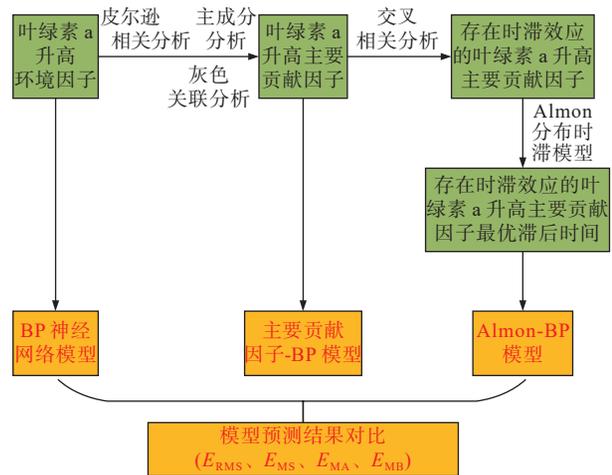


图 3 基于 Almon-BP 模型的叶绿素 a 预测流程
Fig. 3 The chlorophyll-a prediction process based on the Almon-BP model

3 结果与分析

3.1 叶绿素 a 含量升高主要贡献因子筛选

藻密度越大叶绿素 a 质量浓度越高, 由于藻类密度需要通过显微镜进行观测, 通过在线自动监测无法获取准确数据, 因此选取叶绿素 a 来表征浮游植物的生物量及聚生过程。根据以往监测经验, 当叶绿素 a 质量浓度达到 $30 \mu\text{g/L}$ 时, 藻密度接近 10^7L^{-1} 的藻类异常现象临界值。对环境因子和叶绿素 a 进行皮尔逊相关性分析, 其分析结果见图 4、5。从单个因素角度分析, 与香溪河峡口叶绿素 a 相关性较大的驱动因子为溶解氧 ($r=0.518^{**}$)、pH ($r=0.425^{**}$)、氧化还原电位 ($r=-0.251^{**}$)、气压 ($r=-0.159^{**}$)、气温 ($r=0.146^{**}$)、太阳辐照 ($r=0.142^{**}$)、风速 ($r=0.132^*$)。与香溪河上游平邑口叶绿素 a 相关性较大的驱动因子为氧化还原电位 ($r=-0.556^{**}$)、气温 ($r=0.491^{**}$)、三峡水位 ($r=-0.476^{**}$)、水温 ($r=0.463^{**}$)、pH ($r=0.415^{**}$)、溶解氧 ($r=0.404^{**}$)、气压 ($r=-0.387^{**}$)、太阳辐照 ($r=0.323^{**}$)、浊度 ($r=0.298^{**}$)、电导率 ($r=-0.266^{**}$)、

三峡水位差($r=0.179^{**}$)、风速($r=0.162^{**}$)、降雨量($r=0.137^*$)。

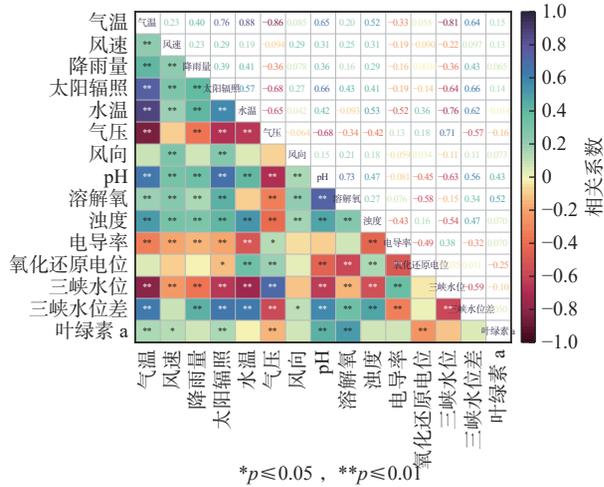
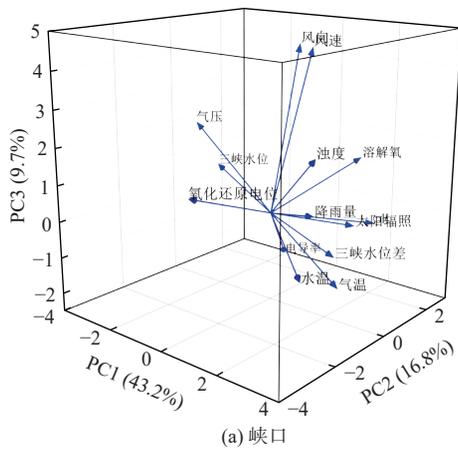


图 4 香溪河峡口环境因子和叶绿素 a 相关性分析

Fig. 4 Correlation analysis between environmental factors at Xiangxi River Xiaokou and chlorophyll-a mass concentration

对环境因子进行主成分分析, 分析结果见图 6, 峡口第一和第二、第三主成分分别解释了数据量的 43.2% 和 16.8%、9.7%, 累计贡献度为 69.7%。主成分 1 特征值为 6.05, 主成分 2、3 特征值分别为 2.36、1.35, 主成分特征值均大于 1。香溪河平邑口第一和第二、第三主成分分别解释了数据量的 50.0% 和 10.3%、8.4%, 累计贡献度为 68.7%。主成分 1 特征值为 6.996, 主成分 2、3 特征值分别为 1.44、1.18, 主



成分特征值均大于 1。观察主成分分析中坐标较大的环境因子, 根据主成分的特征值和解释率以及因子在主成分上的坐标分量得出各因子的指标权重, 计算结果见图 7。

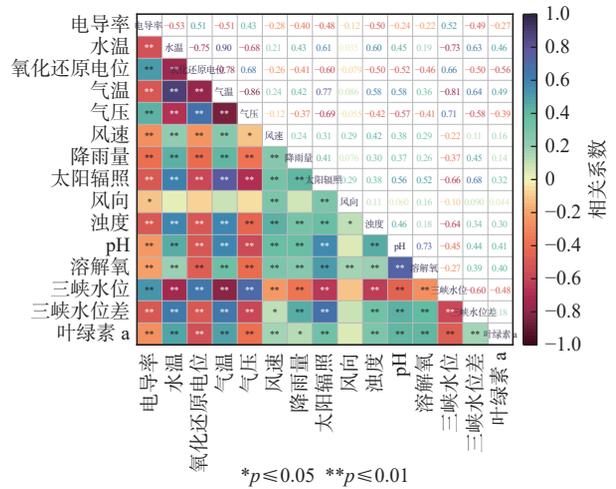


图 5 香溪河平邑口环境因子和叶绿素 a 相关性分析

Fig. 5 Correlation analysis between environmental factors at Xiangxi River Pingyikou and chlorophyll-a mass concentration

三峡库区支流叶绿素 a 含量升高成因复杂, 影响因素众多, 为保证筛选结果的准确性, 再通过灰色关联分析求取香溪河叶绿素 a 含量升高贡献度较大驱动因素。由图 7 可知, 各因素均与叶绿素 a 质量浓度关联度较大, 更加说明了叶绿素 a 含量升高不是单一因素影响的结果而是多种因素共同作用造成的。

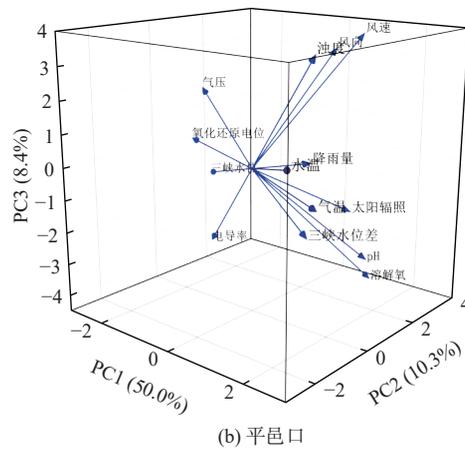


图 6 香溪河主成分分析载荷

Fig. 6 Principal component analysis loading plot of the Xiangxi River

香溪河峡口电导率、氧化还原电位、气压、三峡水位通过主成分分析其指标权重均小于 0, 且相关系数均小于 0.4, 属于弱相关。香溪河平邑口电导率、气压通过主成分分析其指标权重均小于 0, 且相关系数均小于 0.4, 属于弱相关, 不作为叶绿素 a 含量升高的主要贡献因子。香溪河峡口叶绿

a 含量升高主要贡献因子为溶解氧、pH、气温、太阳辐射、风速、风向、浊度、降雨量、三峡水位差、水温。香溪河平邑口叶绿素 a 含量升高主要贡献因子为溶解氧、pH、气温、太阳辐射、风速、风向、浊度、降雨量、三峡水位差、水温、氧化还原电位、三峡水位。

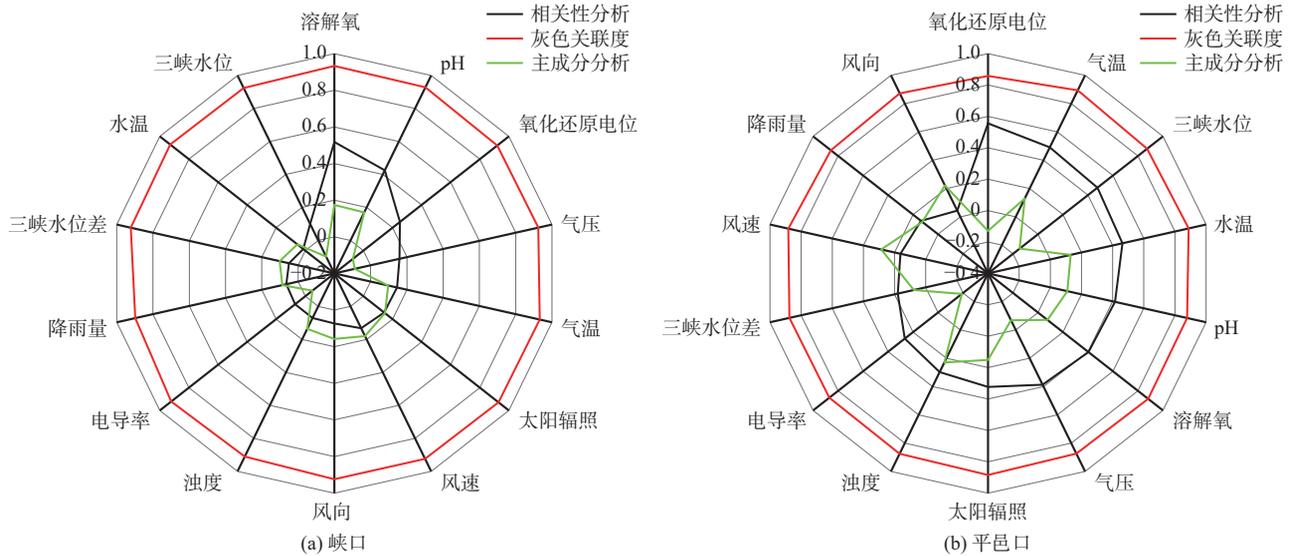


图 7 香溪河主要贡献因子筛选

Fig. 7 The Xiangxi River contributed factors screen

3.2 叶绿素 a 含量升高主要贡献因子时滞分析及最优滞后时间确定

3.2.1 时滞分析

选取前一句的时间通过交叉相关分析进行峡口时滞因子筛选。由图 8 可知:随着滞后时间增加,溶解氧、pH 与叶绿素 a 质量浓度的相关程度逐渐降低,但二者相关系数均超过置信区间上限 0.1,二者均存在时滞效应;太阳辐射、气温随着滞后时间增加与叶绿素 a 质量浓度相关程度呈先升高后降低的趋势,太阳辐射在滞后 9 d 时,相关系数小于置信区间上限 0.1,最大存在 8 d 的时滞,气温在第 5 d 时与叶绿素 a 质量浓度相关关系未超过置信区间上限 0.1,最大存在 4 d 时滞;风速在滞后 3 d 时相关关系未超过置信区间,最大存在 2 d 时滞;风向、浊度、降雨量、三峡水位差、水温均未超过置信区间,不存在时滞效应。

同理选取前一句的时间通过交叉相关分析进行平邑口时滞因子筛选。由图 9 可知:太阳辐射、水温、气温随滞后时间增加与叶绿素 a 质量浓度相关程度整体呈降低趋势,但滞后时间为 10 d 时相关程度仍超过置信区间上限 0.1,太阳辐射、水温、气温对香溪河平邑口叶绿素 a 质量浓度存在滞后影响,最大滞后时间为 10 d;风速在滞后第 3 d 时未超过置信区间,最大存在 2 d 的时滞;降雨量、pH、溶解氧与叶绿素 a 浓度相关关系随滞后时间呈波动变化趋势,降雨量在滞后 6 d 时,相关关系未超过置信区间,最大滞后时间为 5 d;pH 在滞后期 4 d 时相关关系小于置信区间上限 0.1,最大滞后时间为 3 d;溶解

氧最大滞后时间为 2 d;浊度随滞后时间增加,与叶绿素 a 质量浓度相关程度整体呈升高趋势,超过置信区间,浊度对香溪河平邑口叶绿素 a 质量浓度存在滞后影响;三峡水位、三峡水位差、氧化还原电位与叶绿素 a 质量浓度相关程度均大于置信区间上限 0.1,存在滞后时间;在滞后一句时间内,风向与叶绿素 a 质量浓度相关关系均小于置信区间上限 0.1,不存在滞后。

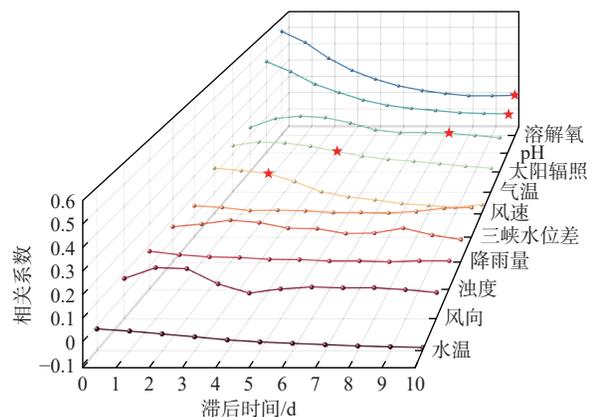


图 8 峡口叶绿素 a 升高主要贡献因子中时滞因子筛选 (图中标记五角星处为最大滞后时间)

Fig. 8 Selection of time lag factors among the primary contributing factors to the elevation of chlorophyll-a at the Xiakou (the maximum lag time is marked in the figure)

3.2.2 最优滞后时间确定

根据 Almon 分布时滞模型求出存在时间滞后的主要贡献因子各个滞后时间的评价指标见图 10。当调整后 R^2 最大时, C_{AI} 、 C_S 最小时对应时间为最优滞后时间。当调整后 R^2 最大与 C_{AI} 、 C_S 最小对应滞后时间不一致时,根据三者的变化剧

烈程度,选择波动较大的指标对应时间作为最优滞后时间,指标波动相对较小,对最优滞后时间选取影响不明显,评价指标变化较大对最优滞后时间选取影响较大。香溪河峡口溶解氧不同滞后时间对应调整后 R^2 呈先升高后降低的变化趋势,调整后 R^2 在滞后时间为 3 d 时达到最大值,而 C_{Al} 、 C_s 呈先降低后上升的趋势,在滞后时间第 4 d 达到最小值, C_{Al} 、 C_s 波动相对较小,波动变化范围在 0.7% 以下,变化较小,调整后 R^2 变化较大波动达到 10.8%。综合以上分析香溪河峡口溶解氧最优滞后时间为 3 d。同理根据调整后 R^2 最大、 C_{Al} 、 C_s 最小准则以及当评价指标对应滞后时间不一致时选取波动较大指标对应滞后时间作为该因子最优滞后时间,综合以上分析香溪河峡口环境因子气温、风速、太阳辐照、pH、溶解氧对峡口叶绿素 a 质量浓度升高存在滞后效应,最优滞后时间分别为 4、2、3、7、3 d。香溪河平邑口环境因子水温、气温、风速、降雨量、太阳辐照、浊度、pH、

溶解氧、三峡水位差、氧化还原电位、三峡水位对平邑口叶绿素 a 质量浓度升高存在滞后效应,最优滞后时间分别为 2、2、2、4、2、10、3、2、6、10、6 d。

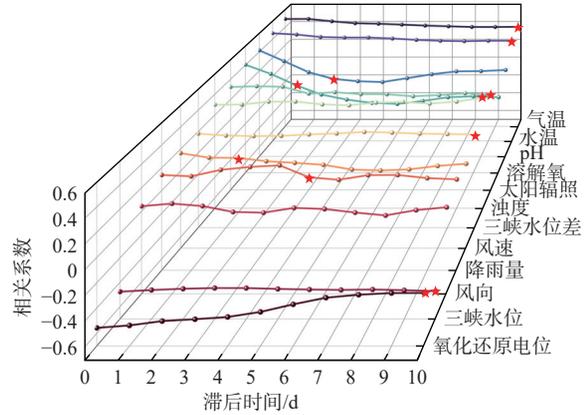
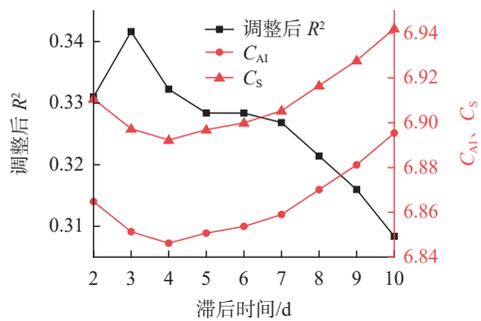
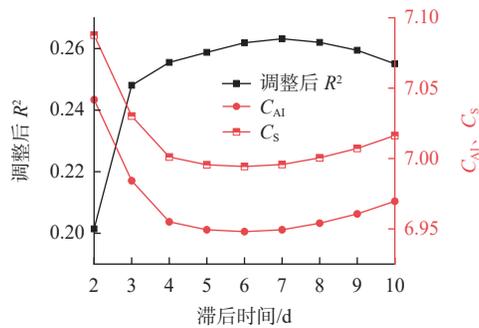


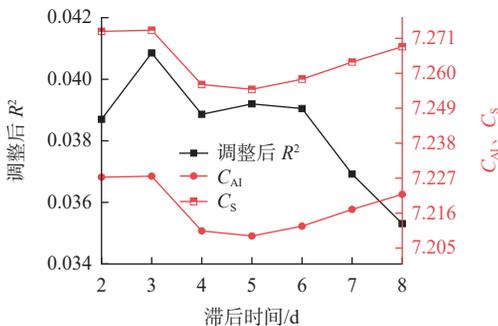
图 9 平邑口叶绿素 a 升高主要贡献因子中时滞因子筛选 (图中标记五角星处为最大滞后时间)
Fig. 9 Selection of time lag factors among the primary contributing factors to the elevation of chlorophyll-a at the Pingyikou (the maximum lag time is marked in the figure)



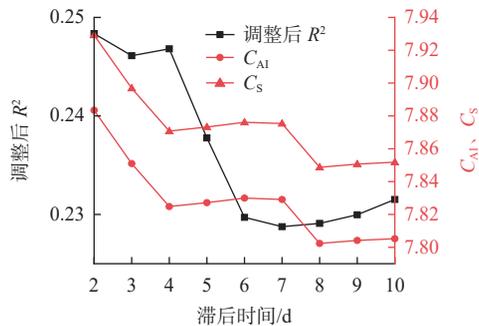
(a1) 溶解氧时滞(峡口)



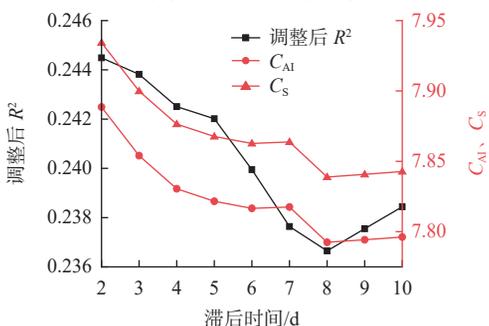
(a2) pH时滞(峡口)



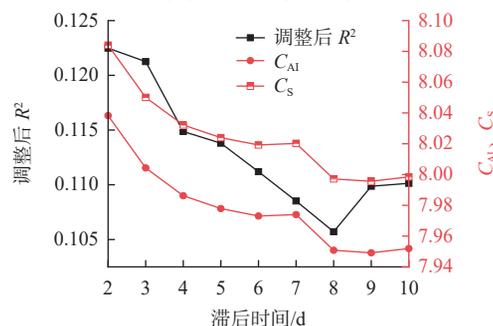
(a3) 太阳辐照时滞(峡口)



(b1) 水温时滞(平邑口)



(b2) 气温时滞(平邑口)



(b3) 太阳辐照时滞(平邑口)

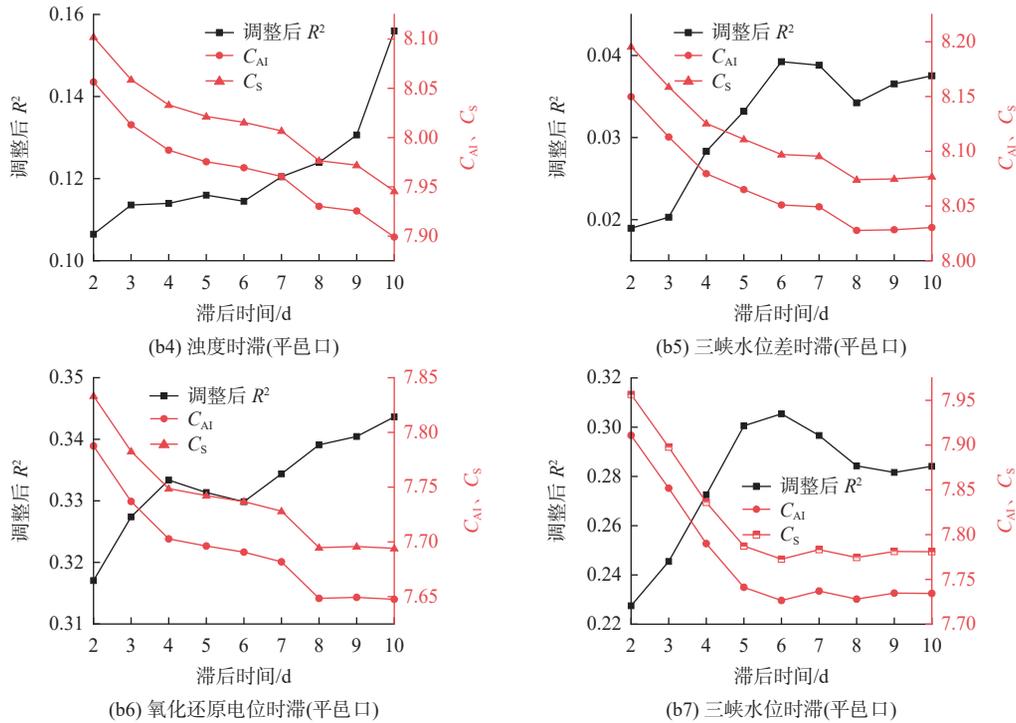


图 10 最优滞后时间的确定

Fig. 10 Determination of optimal time lag time

3.3 基于时滞效应叶绿素 a 趋势预测

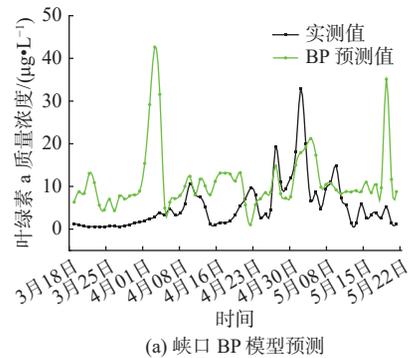
在上文求出叶绿素 a 质量浓度升高主要贡献因子及存在时滞效应的主要贡献因子最优滞后时间的基础上,构建 Almon-BP 时滞神经网络模型对叶绿素 a 趋势进行预测。采用逐日尺度数据对叶绿素 a 趋势进行预测,数据系列长度前 80% 作为模型的训练集,数据系列长度后 20% 作为模型预测集。预测时间序列为 2023 年 3 月 19 日至 5 月 22 日。

基于神经网络模型对香溪河叶绿素 a 趋势预测结果见表 1、图 11、图 12。在考虑所有环境因子的基础上对香溪河进行 BP 神经网络预测,其峡口均方误差(E_{MS})、均方根误差(E_{RMS})、平均绝对误差(E_{MA})、平均偏差(E_{MB})分别为 122.60、11.06、7.80、6.71。平邑口 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 分别为 332.80、18.23、14.67、14.20。基于皮尔逊相关分析、灰色关联分析、主成分分析筛选出的叶绿素 a 含量升高主要贡献因子进行贡献因子-BP 神经网络联合分析,香溪河峡口 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 分别为 59.97、7.74、6.59、4.82,香溪河平邑口 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 分别为 219.49、14.77、11.40、10.91。采用 Almon 模型求取存在时滞效应的主要贡献因子最优滞后时间的基础上,进行 Almon-BP 时滞神经网络模型联合对叶绿素 a 趋势预测,香溪河峡口 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 分别为 33.35、5.75、4.53、2.22,香溪河平邑口 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 分别为 73.16、8.55、7.13、5.91。

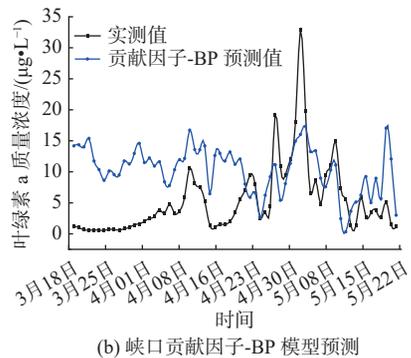
表 1 香溪河 3 种预测模型预测误差对比

Tab. 1 Comparison of prediction errors of three prediction models for the Xiangxi River

监测点	模型	E_{MS}	E_{RMS}	E_{MA}	E_{MB}
峡口	BP	122.61	11.06	7.80	6.71
	贡献因子-BP	59.97	7.74	6.59	4.82
	Almon-BP	33.35	5.75	4.53	2.22
平邑口	BP	332.80	18.23	14.67	14.20
	贡献因子-BP	219.49	14.77	11.40	10.91
	Almon-BP	73.16	8.55	7.13	5.91



(a) 峡口 BP 模型预测



(b) 峡口贡献因子-BP 模型预测

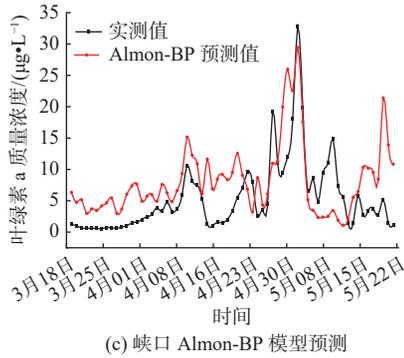


图 11 峡口 3 种模型预测结果与实测值对比

Fig. 11 Comparison of three model predictions and measured values at the Xiakou

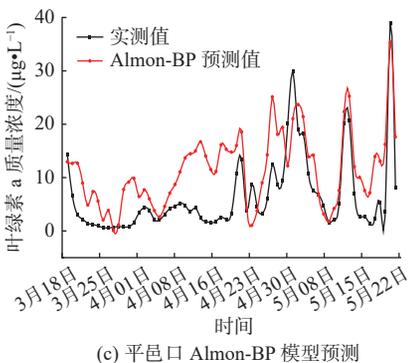
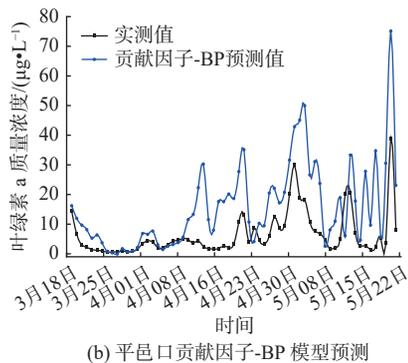
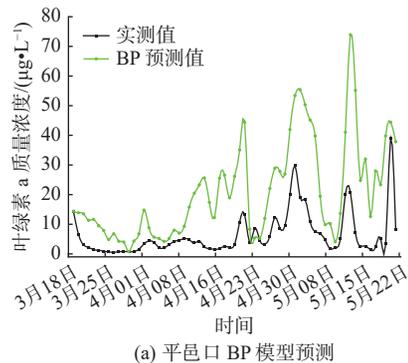


图 12 平邑口 3 种模型预测结果与实测值对比

Fig. 12 Comparison of three model predictions and measured values at the Pingyikou

4 讨论

根据香溪河 2022 年 5 月—2023 年 5 月连续监测数据,运用皮尔逊相关分析、灰色关联分析、主成

分分析筛选叶绿素 a 质量浓度升高主要贡献因子并按大小排序。3 种方法影响因素按大小排序存在部分差异,主成分分析方法和皮尔逊相关性分析、灰色关联分析原理不同。皮尔逊相关性分析、灰色关联分析 2 种方法都是通过环境因子与叶绿素 a 质量浓度关联性大小筛选出主要贡献因子,主成分分析与前 2 种方法有本质区别,主成分分析不通过叶绿素 a 质量浓度,直接从环境因子本身筛选出主要贡献因子。所以主成分分析与皮尔逊相关性分析和灰色关联分析结果之间存在差异较大。同时皮尔逊相关性分析通过环境因子与叶绿素 a 质量浓度相关性大小筛选出主要贡献因子,灰色关联分析根据环境因子和叶绿素 a 质量浓度的时间序列曲线的相似程度来判断其关联程度的大小:若 2 条曲线的形状彼此相似,则关联度大;反之关联度就小。所以三者分析结果存在差异,3 种方法互为补充综合考虑分析筛选出叶绿素 a 质量浓度升高主要贡献因子。

由于水生生态系统的复杂性,时滞效应在藻类繁殖、叶绿素 a 含量升高广泛存在。环境条件与叶绿素 a 含量升高之间存在时滞效应, Lee 等^[31]基于人工神经网络模型研究表明叶绿素 a 含量对环境因子变量滞后时间通常为 7~14 d。刘成建等^[32]研究指出叶绿素 a 含量对环境因子的最大响应滞后时间为 10 d。也有研究湖泊叶绿素 a 含量与驱动因子之间滞后时间为 2~4 d^[33]。叶绿素 a 含量滞后时间受当地水文、水质、气象等因素的综合影响,不同区域所处地理位置不同,同一地理位置不同季节气象因素相差较大,环境因子变量如三峡水位波动与监测点的距离远近都会导致叶绿素 a 含量滞后时间发生变化^[34-35]。基于香溪河小时尺度数据采用 Almon 分布时滞模型对叶绿素 a 含量滞后时间进一步量化。香溪河峡口、平邑口叶绿素 a 含量升高主要贡献因子不同,存在时滞效应的主要贡献因子最优滞后时间也存在较大差异。峡口受长江干流倒灌影响,水文情况复杂导致气温与太阳辐照滞后时间不同,峡口距离三峡水库较近与三峡水位波动不存在时滞效应。平邑口位于香溪河上游受长江干流影响较小,太阳辐照影响气温,气温影响水温进而影响到浮游植物的生长繁殖,太阳辐照、气温、水温 3 者滞后时间一致都为 2 d,平邑口距离三峡水库较远,受三峡水位、三峡水位差滞后时间为 6 d,而峡口不受三峡水位和三峡水位差的滞后影响,同一环境因子

在不同地理位置滞后时间存在部分差异。

采用 3 种模型对香溪河叶绿素 a 趋势进行预测, 基于皮尔逊相关分析、灰色关联分析、主成分分析筛选出的叶绿素 a 质量浓度升高主要贡献因子进行贡献因子-BP 神经网络模型联合预测, 主要贡献因子的筛选降低了模型参数输入的维度, 同时消除了变量之间的冗余性和相关性, 与仅使用 BP 神经网络模型相比香溪河峡口 E_{MS} 降低了 51.1%, E_{RMS} 降低了 30.0%, E_{MA} 、 E_{MB} 分别降低了 15.5%、28.2%, 香溪河平邑口 E_{MS} 降低了 34.0%, E_{RMS} 降低了 19.0%, E_{MA} 、 E_{MB} 分别降低了 22.2%、23.2%, 模型预测误差相对降低。在采用 Almon 模型求取存在时滞效应的主要贡献因子最优滞后时间的基础上, 进行 Almon-BP 时滞神经网络模型联合叶绿素 a 预测, 其预测结果相对于贡献因子-BP 神经网络联合预测误差进一步降低, 模型预测性能进一步提升, 香溪河峡口 Almon-BP 时滞神经网络模型相对于贡献因子-BP 神经网络联合叶绿素 a 预测 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 分别降低了 44.4%、25.6%、31.3%、53.9%, 香溪河平邑口 E_{MS} 、 E_{RMS} 、 E_{MA} 、 E_{MB} 分别降低了 66.7%、42.1%、37.5%、45.8%, Almon-BP 时滞神经网络模型对叶绿素 a 趋势预测误差进一步降低。通过求出环境因子滞后时间可以更好地进行叶绿素 a 趋势的预测。

5 结论

香溪河峡口叶绿素 a 质量浓度升高主要贡献因子为溶解氧、pH、气温、太阳辐照、风速、风向、浊度、降雨量、三峡水位差、水温, 平邑口叶绿素 a 质量浓度升高主要贡献因子为溶解氧、pH、气温、太阳辐照、风速、风向、浊度、降雨量、三峡水位差、水温、氧化还原电位、三峡水位。

香溪河峡口环境因子中气温、风速、太阳辐照、pH、溶解氧对峡口叶绿素 a 质量浓度升高存在滞后效应, 最优滞后时间分别为 4、2、3、7、3 d, 其他环境因子不存在时间滞后。香溪河平邑口环境因子中水温、气温、风速、降雨量、太阳辐照、浊度、pH、溶解氧、三峡水位差、氧化还原电位、三峡水位对平邑口叶绿素 a 质量浓度升高存在滞后效应, 最优滞后时间分别为 2、2、2、4、2、10、3、2、6、10、6 d。风向不存在时滞效应。

相比于 BP 神经网络模型和贡献因子-BP 神经网络模型, 采用考虑时滞效应的 Almon-BP 时滞神

神经网络模型对香溪河峡口、平邑口叶绿素 a 质量浓度趋势进行预测, 可以有效提高叶绿素 a 质量浓度预测精度, 研究验证了 Almon-BP 时滞神经网络模型预测叶绿素 a 质量浓度的有效性。

参考文献:

- [1] 黄恒粤, 陈垚, 刘臻, 等. 基于知识图谱的三峡库区水环境研究热点与展望 [J]. *人民长江*, 2022, 53(5): 53-61. DOI: 10.16232/j.cnki.1001-4179.2022.05.009.
- [2] YANG Z J, WEI C Y, LIU D F, et al. The influence of hydraulic characteristics on algal bloom in Three Gorges Reservoir, China: A combination of cultural experiments and field monitoring [J]. *Water Research*, 2022, 211: 118030. DOI: 10.1016/j.watres.2021.118030.
- [3] 姚金忠, 范向军, 杨霞, 等. 三峡库区重点支流水华现状、成因及防控对策 [J]. *环境工程学报*, 2022, 16(6): 2041-2048. DOI: 10.12030/j.cjee.202112199.
- [4] XIANG R, WANG L J, LI H, et al. Water quality variation in tributaries of the Three Gorges Reservoir from 2000 to 2015 [J]. *Water Research*, 2021, 195: 116993. DOI: 10.1016/j.watres.2021.116993.
- [5] JI D B, WELLS S A, YANG Z J, et al. Impacts of water level rise on algal bloom prevention in the tributary of Three Gorges Reservoir, China [J]. *Ecological Engineering*, 2017, 98: 70-81. DOI: 10.1016/j.ecoleng.2016.10.019.
- [6] LAIO N, ZHANG L L, CHEN M, et al. The influence mechanism of water level operation on algal blooms in canyon reservoirs and bloom prevention [J]. *Science of the Total Environment*, 2024, 912: 169377. DOI: 10.1016/j.scitotenv.2023.169377.
- [7] 李建, 尹炜, 贾海燕, 等. 汉江中下游水华防控生态调度研究 [J]. *湖泊科学*, 2022, 34(3): 740-751. DOI: 10.18307/2022.0304.
- [8] 刘德富, 杨正健, 纪道斌, 等. 三峡水库支流水华机理及其调控技术研究进展 [J]. *水利学报*, 2016, 47(3): 443-454. DOI: 10.13243/j.cnki.slxb.20151304.
- [9] 陈怡, 姚亚芹, 李一平, 等. 夏季香溪河库湾溶解氧与叶绿素 a 日间垂向分布特征及影响因子初探 [J]. *水电能源科学*, 2023, 41(4): 69-72. DOI: 10.20040/j.cnki.1000-7709.2023.20221029.
- [10] 杨凡, 杨正健, 纪道斌, 等. 三峡库区不同河段支流丰水期叶绿素 a 和营养盐的空间分布特征 [J]. *环境科学*, 2019, 40(11): 4944-4952. DOI: 10.13227/j.hjks.201905182.
- [11] 方海涛, 宋林旭, 纪道斌, 等. 香溪河夏季水华暴发差异性研究及其机制分析 [J]. *环境科学与技术*, 2019, 42(9): 68-74. DOI: 10.19672/j.cnki.1003-6504.2019.09.011.
- [12] 潘晓洁, 黄一凡, 郑志伟, 等. 三峡水库小江夏初水

- 华暴发特征及原因分析 [J]. *长江流域资源与环境*, 2015, 24(11): 1944-1952. DOI: [10.11870/cjlyzyyhj201511019](https://doi.org/10.11870/cjlyzyyhj201511019).
- [13] 姜伟, 周川, 纪道斌, 等. 三峡库区澎溪河与磨刀溪电导率等水质特征与水华的关系比较 [J]. *环境科学*, 2017, 38(6): 2326-2335. DOI: [10.13227/j.hjcx.201610183](https://doi.org/10.13227/j.hjcx.201610183).
- [14] JEONG K S, KIM D K, JOO G J. Delayed influence of dam storage and discharge on the determination of seasonal proliferations of *Microcystis aeruginosa* and *Stephanodiscus hantzschii* in a regulated river system of the lower Nakdong River (South Korea)[J]. *Water Research*, 2007, 41(6): 1269-1279. DOI: [10.1016/j.watres.2006.11.054](https://doi.org/10.1016/j.watres.2006.11.054).
- [15] KIM M C, JEONG K S, KANG D K, et al. Time lags between hydrological variables and phytoplankton biomass responses in a regulated river (the Nakdong River)[J]. *Journal of Ecology and Environment*, 2009, 32(4): 221-227. DOI: [10.5141/JEFB.2009.32.4.221](https://doi.org/10.5141/JEFB.2009.32.4.221).
- [16] LIU C, CHEN Y, ZOU L, et al. Time-lag effect: River algal blooms on multiple driving factors[J]. *Frontiers in Earth Science*, 2022, 9: 8132387. DOI: [10.3389/feart.2021.813287](https://doi.org/10.3389/feart.2021.813287).
- [17] KIM K B, JUNG M K, TSANG Y F, et al. Stochastic modeling of chlorophyll-a for probabilistic assessment and monitoring of algae blooms in the Lower Nakdong River, South Korea[J]. *Journal of Hazardous Materials*, 2020, 400: 123066. DOI: [10.1016/j.jhazmat.2020.123066](https://doi.org/10.1016/j.jhazmat.2020.123066).
- [18] XIA R, WANG G, ZHANG Y, et al. River algal blooms are well predicted by antecedent environmental conditions[J]. *Water Research*, 2020, 185: 116221. DOI: [10.1016/j.watres.2020.116221](https://doi.org/10.1016/j.watres.2020.116221).
- [19] KIM J, LEE T, SEO D. Algal bloom prediction of the lower Han River, Korea using the EFDC hydrodynamic and water quality model[J]. *Ecological Modelling*, 2017, 366: 27-36. DOI: [10.1016/j.ecolmodel.2017.10.015](https://doi.org/10.1016/j.ecolmodel.2017.10.015).
- [20] LUO L, LIANG X, LUO M, et al. Characterizing the hierarchical groundwater flow systems in Karstic Xiangxi River basin, West Hubei, Central China[J]. *Applied Geochemistry*, 2022, 143: 105371. DOI: [10.1016/j.apgeochem.2022.105371](https://doi.org/10.1016/j.apgeochem.2022.105371).
- [21] 陈聿奇, 崔玉洁, 黄浩昇, 等. 分层异重流背景下降雨对三峡水库香溪河水华消退的影响 [J]. *湖泊科学*, 2023, 35(2): 519-529. DOI: [10.18307/2023.0210](https://doi.org/10.18307/2023.0210).
- [22] 谭路, 申恒伦, 王岚, 等. 三峡水库干流与香溪河库湾水体营养状态及其对水文条件的响应 [J]. *长江流域资源与环境*, 2021, 30(6): 1488-1499. DOI: [10.11870/cjlyzyyhj202106023](https://doi.org/10.11870/cjlyzyyhj202106023).
- [23] JI L, ZHAO L, OUYANG T, et al. Influence and driving of environmental heterogeneity on the epilithic diatom community in Xiangxi River, a tributary of the Three Gorges Reservoir area[J]. *Environmental Science*, 2023, 44(4): 2083-2092. DOI: [10.13227/j.hjcx.202206113](https://doi.org/10.13227/j.hjcx.202206113).
- [24] LUKMAN A F, KIBRIA G B. Almon-KL estimator for the distributed lag model[J]. *Arab Journal of Basic and Applied Sciences*, 2021, 28(1): 406-412. DOI: [10.1080/25765299.2021.1989160](https://doi.org/10.1080/25765299.2021.1989160).
- [25] MAJID A, ASLAM M, ALTAF S, et al. Addressing the distributed lag models with heteroscedastic errors[J]. *Communications in Statistics-Simulation and Computation*, 2021, 50(12): 4464-4482. DOI: [10.1080/03610918.2019.1643884](https://doi.org/10.1080/03610918.2019.1643884).
- [26] DUAN Y F, REN Z Y, SUN Y J. Time-lay effects of climate on water use efficiency in the Loess Plateau of northern Shaanxi[J]. *Acta Ecologica Sinica*, 2020, 40(10): 3408-3419. DOI: [10.5846/stxb201905110963](https://doi.org/10.5846/stxb201905110963).
- [27] ZBAY N. Two-parameter ridge estimation for the coefficients of Almon distributed lag model[J]. *Springer International Publishing*, 2019, 43(4): 1819-1828. DOI: [10.1007/s40995-018-0634-5](https://doi.org/10.1007/s40995-018-0634-5).
- [28] 殷高方, 张玉钧, 胡丽, 等. BP 神经网络在水华短期预测中的应用 [J]. *北京理工大学学报*, 2012, 32(6): 655-660. DOI: [10.15918/j.tbit1001-0645.2012.06.006](https://doi.org/10.15918/j.tbit1001-0645.2012.06.006).
- [29] 张克鑫, 陆开宏, 朱津永, 等. 基于 BP 神经网络的藻类水华预测模型研究 [J]. *中国环境监测*, 2012, 28(3): 53-57. DOI: [10.19316/j.issn.1002-6002.2012.03.012](https://doi.org/10.19316/j.issn.1002-6002.2012.03.012).
- [30] 李世明, 王小艺, 许继平, 等. 基于人工神经网络的河湖蓝藻水华状态评价研究 [J]. *水资源与水工程学报*, 2017, 28(4): 93-96. DOI: [10.11705/j.issn.1672-643X.2017.04.16](https://doi.org/10.11705/j.issn.1672-643X.2017.04.16).
- [31] LEE J H, HUANG Y, DICKMAN M, et al. Neural network modelling of coastal algal blooms [J]. *Ecological Modelling*, 2003, 159(2-3): 179-201. DOI: [10.1016/S0304-3800\(02\)00281-8](https://doi.org/10.1016/S0304-3800(02)00281-8).
- [32] 刘成建, 夏军, 宋进喜, 等. 汉江中下游水质时空特征与土地利用类型响应识别研究 [J]. *环境科学研究*, 2021, 34(4): 910-919. DOI: [10.13198/j.issn.1001-6929.2020.08.15](https://doi.org/10.13198/j.issn.1001-6929.2020.08.15).
- [33] RECKNAGEL F, OSTROVSKY I S, CAO H, et al. Ecological relationships, thresholds and time-lags determining phytoplankton community dynamics of Lake Kinneret, Israel elucidated by evolutionary computation and wavelets[J]. *Ecological Modelling*, 2013, 255: 70-86. DOI: [10.1016/j.ecolmodel.2013.02.006](https://doi.org/10.1016/j.ecolmodel.2013.02.006).

- [34] 王慧梅, 孙伟, 祖元刚, 等. 不同环境因子对兴安落叶松树干液流的时滞效应复杂性及其综合影响[J]. *应用生态学报*, 2011, 22(12): 3109-3116. DOI: 10.13287/j.1001-9332.2011.0465.
- [35] 马云飞, 何全军, 张月维, 等. 时滞与累积效应下气候因子和人类活动对松辽流域植 LAI 动态的影响[J]. *生态学报*, 2024, 44(5): 1-15. DOI: 10.20103/j.stxb.202305060947.

Chlorophyll-a prediction based on the Almon-BP delayed neural network model

LIU Zhenqi¹, LI Jian², JIANG Shanhu¹

(1. College of Hydrology and Water Resources, Hohai University, Nanjing 210000, China;

2. Changjiang Water Resources Protection Institute, Wuhan 430000, China)

Abstract: Since the impoundment of the Three Gorges Reservoir, the nutrient enrichment status of tributary water bodies in the reservoir area significantly deteriorated, transitioning from riverine to lacustrine ecosystems. The reservoir impoundment resulted in reduced flow velocities of tributary water bodies, leading to increased water transparency and nutrient concentrations conducive to the proliferation of increasing in chlorophyll-a mass concentration. Consequently, typical tributaries such as the Xiangxi River in the Three Gorges Reservoir area experienced frequent outbreaks of increasing in chlorophyll-a mass concentration. These recurrent increase in chlorophyll-a mass concentration events not only degraded water quality and aquatic ecosystems but also posed constraints on the sustainable development of society. Correlation analysis, principal component analysis, and grey relational analysis were utilized to identify and validate significant contributors to increase in chlorophyll-a mass concentration. Subsequently, cross-correlation analysis and the Almon distributed lag model were employed to ascertain factors among the major contributors exhibiting time lags and to determine the optimal lag time. Building upon this analysis, an Almon-BP neural network model was developed to forecast the trends of chlorophyll-a mass concentration. The major contributing factors to increase in chlorophyll-a mass concentration at the Xiangxi River Xiakou included dissolved oxygen, pH, air temperature, solar radiation, wind speed, wind direction, turbidity, rainfall, Three Gorges water level difference, and water temperature. Similarly, significant factors at the Pingyikou of the Xiangxi River included dissolved oxygen, pH, air temperature, solar radiation, wind speed, wind direction, turbidity, rainfall, Three Gorges water level difference, water temperature, redox potential, and Three Gorges water level. Among environmental factors at the Xiakou of the Xiangxi River, air temperature, wind speed, solar radiation, pH, and dissolved oxygen exhibited lag effects on increased in chlorophyll-a mass concentration, with optimal lag times ranging from 2 to 7 days, while other environmental factors did not display time lags. Conversely, at the Pingyikou of the Xiangxi River, factors such as water temperature, air temperature, wind speed, rainfall, solar radiation, turbidity, pH, dissolved oxygen, Three Gorges water level difference, redox potential, and Three Gorges water level exhibited lag effects on chlorophyll-a mass concentration, with optimal lag times ranging from 2 to 10 days. Wind direction did not show lag effects. Comparative analysis of three prediction models the BP neural network model considering all environmental factors, the BP neural network model considering only major contributing factors, and the Almon-BP neural network model considering the optimal lag time of major contributors revealed that the Almon-BP neural network model outperformed the corresponding BP models in predicting chlorophyll-a mass concentration in the Xiangxi River, with lower prediction errors. This underscored the efficacy of the Almon-BP neural network model in enhancing the accuracy of chlorophyll-a mass concentration prediction, which was crucial for early warning and mitigating harmful algal bloom occurrences.

Key words: Chlorophyll-a prediction; main contribution factor; time-lag effect; optimal lag time; Almon-BP model