

城市湖泊水华预警模型研究 ——以北京“六海”为例

曾 勇, 杨志峰, 刘静玲

(水环境模拟国家重点实验室, 北京师范大学环境学院, 北京 100875)

摘要: 采用决策树方法和非线性回归方法建立湖泊水华预警模型。决策树方法预测水华爆发时机, 非线性回归方法预测水华爆发强度, 并运用信号灯显示方法, 划分出水华爆发的预警区间。以北京六海为例, 模型结果表明来水水量 Q , 温度 T 和总磷浓度是影响“六海”湖泊水华爆发的主要影响因子, 选择叶绿素 a (Chl- a) $< 30 \text{ ug/L}$ 的预警信号为绿色, $30 \text{ ug/L} < \text{Chl-}a < 60 \text{ ug/L}$ 为黄色, $\text{Chl-}a > 60 \text{ ug/L}$ 为红色。当每月来水量 $Q > 79.0 \text{ 万 m}^3$ 或来水量 $Q < 79.0 \text{ 万 m}^3$, 水温 < 13.4 , 预警指标为绿色; $Q < 79.0 \text{ 万 m}^3$, 水温 $T > 13.4$, 水华预警为黄色; $Q < 38.7 \text{ 万 m}^3$ 时, $T > 23.25$, $\text{TP} > 0.13 \text{ mg/L}$, 水华预警为红色。对模型结果分类进行了验证。结果表明: 模型对于限制因素发生变化时的水华预测结果更为准确, 并且结构简单, 输入输出关系明显, 结果易于解释。

关键词: 水华; 预警; 决策树; 非线性回归; 信号灯显示; 北京六海

中图分类号: X524

文献标识码: A

文章编号: 1001-6791(2007)01-0079-07

水华爆发是由水体的物理、化学和生物过程等多种因素共同作用的结果, 各要素之间关系复杂, 存在随机性、不确定性和非线性特征, 目前对于其发生的临界因素和机理不完全清楚^[1]。对水华爆发的预测主要有两种方法, 一种是确定性的机理模型, 以现实系统的基本物理、化学定律理论为基础而得到的模型, 如 QUAL- $^+$, WASP, SALMO 等水质生态模型^[2~5], 但需要大量数据确定其参数; 另一种为数据驱动方法, 包括多元统计回归方法和神经网络方法。如周勇等提出了灰色动态模型和多元线性回归模型耦合预测湖泊未来水环境变化的原理和方法^[6]。但多元统计回归模型存在着模型形式的选择问题, 一般采用线性关系进行简化, 因此在处理水花爆发的限制因素发生变化时的预测效果不佳; 而神经网络方法由于具有较强的适应能力、学习能力和真正的多输入多输出系统的特点得到人们的重视。其中 Maier 等建立的人工神经网络模型预测藻类浓度, 并对不同时期的数据输入变量作了敏感性分析^[7]。Gurbuz 采用初期终止方法训练和校正人工神经网络模型, 预测水库中藻类植物的浓度^[8]。裴洪平等通过反复的训练找到最能反映水生生态状况变化趋势的水质因子, 建立神经网络预测西湖的叶绿素 a 的浓度^[9]。而神经网络模型对于如何确定输入变量和网络结构没有很好的方法, 并且很难解释神经网络结构的功能以及它们对输出变量的影响。

近年来, Solomatine 和 Dulal 在水文预测中比较了决策树和分段线性统计回归预测方法和神经网络预测方法, 认为前者具有同样的精度, 并且模型的输入输出关系明显, 结果易于解释^[10]。Chen 和 Mynett 应用决策树和分段非线性统计回归方法预测了荷兰海岸带水华的叶绿素 a 浓度变化趋势^[11]。目前, 采取决策树和分段非线性方法对水华爆发进行预测, 特别是对水华爆发进行预警区间分类的文献并不多见。本文, 以北京六海水华爆发为例, 采用决策树和非线性统计回归方法探讨其限制因素发生变化时, 水华爆发时机、强度的预测, 并运用信号灯显示模型的方法, 划分出水华爆发的预警区间, 便于采取不同的应对措施。

收稿日期: 2005-09-16; 修订日期: 2006-02-24

基金项目: 国家自然科学基金重点资助项目 (50239020); 国家重点基础研究发展计划 (973) 资助项目 (2006CB403303)

作者简介: 曾 勇(1974-), 男, 福建宁化人, 博士后, 主要从事水资源与水环境管理方面研究。

通讯作者: 杨志峰, E-mail: Zfyang@bnu.edu.cn

1 研究方法

湖泊水华爆发原因非常复杂,是一个多变量、高度非线性过程。通过决策树方法将整个空间分为不同区间,对不同区间采用线性多元统计回归方法进行预测,整个过程可以看作是非线性预测的线性化过程。

1.1 决策树方法

决策树方法^[12,13]主要应用于分类和预测,回答什么条件下会得到什么值的这类问题,通过确定对象的各属性的重要程度,提取出必要的属性对研究对象进行分类。

(1) 决策树生成 当前最有影响的决策树算法是 Quinlan 于 1986 年提出的 ID3,和 1993 年提出的 C4.5。C4.5 是 ID3 的改进算法,不仅可以处理离散值属性,还能处理连续值属性;C4.5 采用了信息增益率作为选择测试属性的标准,优化选择出最能对系统进行分类的属性。

C4.5 计算步骤:在某一判断点处,如该节点的集合 S 由 N 个实例构成,对于某个连续属性 A ,分为三步进行处理:

实例排序 首先,将判定节点处所有的实例按连续属性 A 进行增序排列,得到属性值序列: $(v_1, v_2, \dots, v_n)$

生成候选分割点 任何位于两个实例之间的分割点都能同样的将 S 中所有的实例划分为两类:属性的取值属于 $(v_1, v_2, \dots, v_i)$ 的实例和属于 $(v_{i+1}, v_{i+2}, \dots, v_n)$ 的实例,这样,在 A 上有多种可能的分割点,第 i 个分割点为 $(v_i + v_{i+1})/2$

候选分割点评价 通过对第二步产生的所有分割点进行评价,从中选择一个最好的分割点 S_A 。首先计算各属性的信息熵:

$$Gain(S, A) = E(S) - \sum_v \frac{|S_v|}{|S|} E(S_v) \quad (1)$$

式中 $Gain(S, A)$ 为样本 S 在属性 A 分类上的信息熵; $E(S)$ 是样本 S 的信息熵; S_v 是符合给定输出结果标准的样本个数,定义为子集 v ; $E(S_v)$ 是属于子集 v 的信息熵;其中信息熵的计算公式为

$$E(S) = - \sum_{i=1}^N p_i \log_2 p_i \quad (2)$$

式中 p_i 为样本属于子集 i 的比例。

把集合 S 分割成 2 部分而生成的潜在信息计算为

$$split \text{ info}(S, A) = - \sum_{i=1}^2 \frac{|S_i|}{|S|} \times \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (3)$$

因此,信息增益率的计算公式为

$$gain \text{ ratio}(X) = \frac{Gain(S, A)}{split \text{ info}(S, A)} \quad (4)$$

(2) 剪枝技术 生成决策树后, C4.5 算法采取基于错误剪枝技术(EBP)来纠正过度适合问题,即剪去决策树中不能提高预测准确率的分支。

假定样本的错分率 U_{CF} 可看成是 N 次实验中某事件发生分类错误 E 次的概率, CF 为置信水平,在 C4.5 中默认为 0.25;假设错分样本的概率服从二项式分布,如下式所示:

$$f[E, N, U_{CF}(E, N)] = \frac{N!}{E! * (N - E)!} U_{CF}(E, N)^E * [1 - U_{CF}(E, N)]^{N-E} = 0.25 \quad (5)$$

从式(5)中求解计算叶节点的预测样本错分率 $U_{CF}(E, N)$ 。

接着计算叶节点的预测错分样本数:

$$E_r = N * U_{CF}(E, N) \quad (6)$$

最后, 判断是否剪枝以及如何剪枝, 分别计算3种预测错分样本数:

计算以节点 t 为根的子树 T_t 的所有叶节点预测错分样本数之和, 记为 E_{r1} ;

计算子树 T_t 被剪枝, 以叶节点代替时的预测错分样本数, 记为 E_{r2} ;

计算子树 T_t 的最大分支的预测错分样本数, 记为 E_{r3} 。

对以上3个值进行比较, E_{r2} 最小时则进行剪枝, 把子树 T_t 剪掉并代以一个叶节点; E_{r3} 最小时采用嫁接 (grafting) 策略, 即用这个最大分枝来替代子树 T ; E_{r1} 最小则不剪枝。

1.2 非线性多元统计回归

根据决策树分析结果确定出主要的输入变量以及变量的区间划分, 在各区间内, 将确定出的主要输入变量作为回归方程的自变量, chl-a 浓度为因变量, 采用 Enter 法进行分段线性回归。分断线性多元统计回归预测公式为^[14]

$$y_i = a + b_1 x_{1i} + \dots + b_n x_{ni} + e_i \quad (7)$$

式中 y_i 称为 i 区间真值 y 的估计值; x_{ni} 为 i 区间第 n 个自变量; a 为截距; b 为偏回归系数; e 为随机误差。

1.3 模型的验证

对于决策树预测水华爆发时机的准确率采取下式计算:

$$S = \frac{s}{V} \times 100\% \quad (8)$$

式中 S 为预测成功率, %; s 为预测成功次数; V 为预测样本数。

对于非线性回归方法预测水华爆发强度的准确率采取相对误差公式计算:

$$\delta = \frac{|\alpha - A|}{A} \times 100\% \quad (9)$$

式中 δ 为相对误差; α 为预测值; A 为监测值。

2 实例研究

2.1 研究区概况

北京市“六海”湖泊群是地处市中心的重要景观水域, 由西海、后海、前海、北海、中海和南海6个子湖组成。其水源主要来自官厅水库及沿途湖岸径流和湖面降水, 通过长河经铁灵闸放入, 并依次经过各子湖, 贯穿整个西城区后排入筒子河(图1), “六海”总面积 142.11 万 km^2 , 岸堤长 3.5 km, 总库容 209.34 万 m^3 。

通过 2003 - 2004 年水质监测数据分析发现, “六海”的水体已经受到严重的污染, 且呈恶化趋势。其水质问题主要是有机物污染和富营养化, 其中“六海”水体的主要污染因子包括: 总氮和总磷、 COD_{Mn} , 水质多为 Ⅲ 类水标准, 其中“六海”中的西海, 前海和后海, 监测断面处于超富营养状态, 北海, 中海和南海处于富营养状态。

研究区自 1999 年来爆发严重水华, 其水华类型为表面水华。据西海 2003 - 2004 年数据统计, 叶绿素 a 大于 30 $\mu\text{g}/\text{mL}$ 的次数为 21 次, 频率为 54%, 最长持续时间为 50 d; 叶绿素 a 大于 60 $\mu\text{g}/\text{L}$ 的次数为 12 次, 频率为 31%, 最长持续时间为 40 d。

根据研究区的生态环境问题的严重性和优先程度, 确定出控制水体的水华爆发为优先目标。其中西海水质富营养化最为严重, 其水质参数如表 1 所示。以西海水质数据作为预测样本, 可达到研究区水质控制要求。

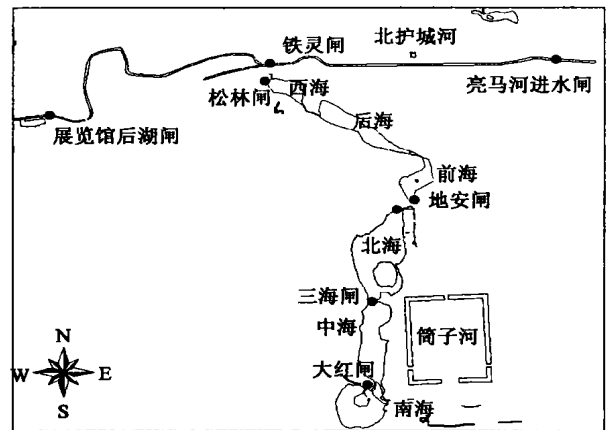


图1 研究区概况

Fig.1 Scheme of the study areas

表 1 2003 - 2004 年西海平均水质数据

Table 1 Water quality of Liu Hai lake in 2003 - 2004

数据名	T	pH	DO	COD _{Mn}	NH ₄ -N	TP	TN	Chl-a	SD
均值	24.4	8.1	12.0	8.6	0.80	0.21	1.60	71.6	0.5
范围	2.5~33.0	7.4~8.8	5.2~23.4	3.2~18.1	0.10~2.30	0.11~0.37	0.60~3.40	2.7~347.5	0.3~0.7

2.2 研究区水华爆发与各限制因素的关系

对 2003 年 4 月到 2004 年 10 月的西海 Chl-a 与各限制因素的周水质监测数据进行分析(图 2), 影响因子水温与叶绿素 a 浓度之间正相关关系。总磷在 2004 年与叶绿素 a 关系呈正相关, 但在 2003 年与叶绿素 a 呈负相关

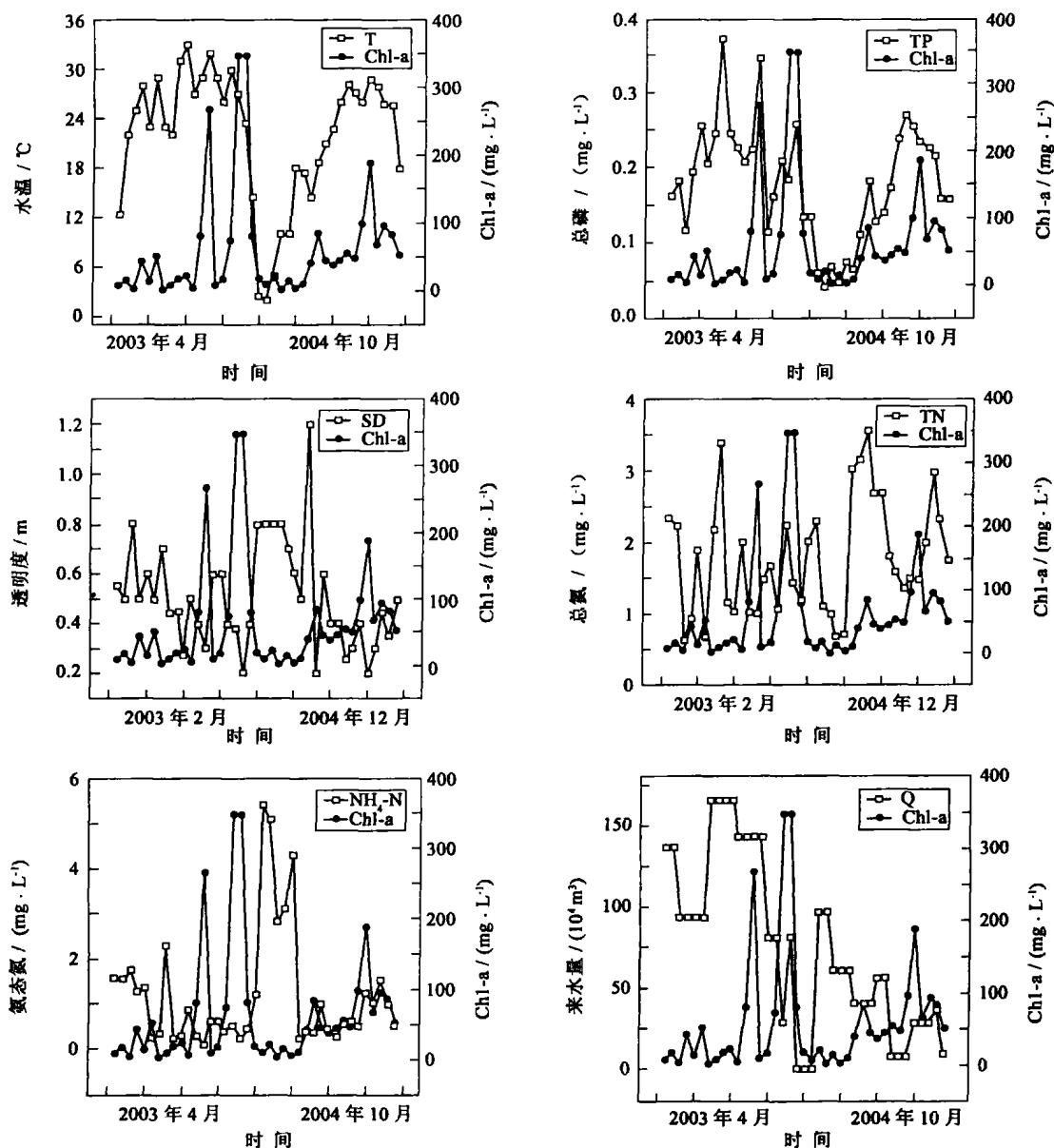


图 2 西海叶绿素 a(Chl-a)浓度与水温(T),总磷(TP),透明度(SD),总氮(TN),氨氮(NH₄-N),来水水量(Q)等因素之间的相关关系

Fig. 2 Relations of Chl-a among T, TP, SD, TN, NH₄-N and water quantity of inflow

关系，表明总磷浓度过高，已不再是水华爆发的限制因素。透明度基本上和叶绿素 a 呈负相关关系，表明水华越严重，透明度就越低。总氮和氨氮与叶绿素 a 之间呈负相关关系，表明氮已不成为研究区水华爆发的限制因素。来水量与叶绿素 a 负相关关系明显，表明来水量越少，叶绿素 a 浓度越高。

2.3 决策树分类结果

选择 2003 年 4 月到 2004 年 10 月西海的水温、总氮、总磷和氨氮指标的每周监测数据，同时选择西海入水口(铁灵闸)来水量作为反映研究区水体流动性特征，来水量以半个月的日监测数据总和作为为分类指标，共 39 个样本。输出指标取叶绿素 a 浓度作为研究区水华爆发的指示。根据《中国水资源公报编制大纲》中的《湖泊、水库富营养化评分与分类方法》，评分值 ≥ 60 为富营养化水体，当评分值为 60 分时，Chl-a 的浓度为 26.0 ug/L，评分值为 70 分时，浓度为 64.0 ug/L。Chen 和 Mynett 以 Chl-a 浓度 30 ug/L 和 60 ug/L 作为预测荷兰海岸带水华发生的判断标准^[11]。因此,本文选择 Chl-a 浓度大于 30 ug/L 和 60 ug/L 为水华发生标准是适当的。

根据公式计算的分类结果如表 2 所示。

表 2 指标分类结果
Table 2 Classify results of indexes

指 标	输入变量属性					输出变量属性
	<i>T</i> /	IN / (mg L ⁻¹)	NH ₄ -N / (mg L ⁻¹)	TP / (mg L ⁻¹)	Inflow / (10 ⁴ m ³)	Chl-a / (ug L ⁻¹)
最佳分类点	13.4	0.68	1.30	0.09	39.50	30
信息增益率	0.315	0.201	0.123	0.315	0.343	
最佳分类点	23.25	1.01	1.56	0.13	19.33	60
信息增益率	0.330	0.134	0.122	0.198	0.416	

根据表 2 的信息熵的增益率结果，当选择 Chl-a 分类标准为 30 ug/L 时，对于系统的分类首先选择铁灵闸来水流量作为第一分类标准。然后选择水温作为第二分类标准，TP 作为第三分类标准建立决策树；当选 Chl-a 分类标准为 60 ug/L，对于系统的分类首先选择铁灵闸来水流量 (*Q*) 作为第一分类标准；然后选择水温 (*T*) 作为第二分类标准；总磷(TP)作为第三分类标准建立决策树；同时根据剪枝技术简化建立的决策树,最终得到的决策规则(表 3)。

其次，应该确定水华警戒水平，根据警戒水平设计预警区间，即确定预警分界点。本文中采用类似交通管制信号红、黄、绿的标志，发出预警信号，进行预警。预警信号为绿灯时，说明水环境状况良好，几乎没有水华现象发生；当叶绿素 a 浓度超过 30 ug/ mL，预警信号为黄灯时，说明水生态环境系统不正常，有水华出现；叶绿素 a 浓度超过 60 ug/ mL，预警信号为红灯，说明水华问题十分严重，有危险性。

综合表 3 不同 Chl-a 浓度下的分类规则，得出研究区水华爆发预警决策规则，同时根据多元回归式(7)，应用 SPSS 软件中 regress 功能，建立研究区水体水华影响要素分段区间的预警规则体系：

表 3 不同 Chl-a 浓度下的分类规则
Table 3 Rules in the different classify criterion of Chl-a

规则集	输入变量	输出变量Chl-a/ (ug L ⁻¹)	预警指标
规则 1	如果来水量 $Q > 39.5 \times 10^4 \text{ m}^3$	< 30	绿色
规则 2	$Q < 39.5 \times 10^4 \text{ m}^3$ ；水温 $T > 13.4$	> 30	黄色
规则 3	$Q < 39.5 \times 10^4 \text{ m}^3$ ；水温 $T < 13.4$	< 30	绿色
规则 4	来水量 $Q < 19.33 \times 10^4 \text{ m}^3$ ；水温 $T > 23.25$ ；TP $> 0.13 \text{ mg/L}$	> 60	红色
规则 5	来水量 $Q < 19.33 \times 10^4 \text{ m}^3$ ；水温 $T < 23.25$	< 60	黄色
规则 6	来水量 $Q > 19.33 \times 10^4 \text{ m}^3$	< 60	黄色

- 规则 1 如来水量 $Q > 39.5 \text{ 万 m}^3$ ，则预警指标 = 绿色；
- 规则 2 如来水量 $Q < 39.5 \text{ 万 m}^3$ ，水温 $T < 13.4$ ，则预警指标 = 绿色；
- 回归方程为

$$\text{Chl-a} = 12.865 + 0.626 T - 0.225 Q$$

其中 R 为 0.768;

规则 3 如来水量 $Q < 39.5$ 万 m^3 , 水温 $T > 13.4$, 则预警指标 = 黄色;

预测回归方程为

$$\text{Chl-}a = 18.708 + 4.549 T - 2.75 Q$$

其中 R 为 0.409;

规则 4 如来水量 $Q < 19.33$ 万 m^3 , 水温 $T > 23.25$, $TP > 0.13$ $\mu\text{g/L}$, 则预警指标 = 红色;

预测回归方程为

$$\text{Chl-}a = -337.7 + 12.839 T - 2.93 Q + 506.1 TP$$

其中 R 为 0.749。

表 4 模型的训练和验证

Table 4 Training and test of the model

模型集合	水华发生标准 / (mg L^{-1})	样本数 / 个	预测成功率 / %	预测精度 / %
训练集	< 30	18	92	7.30
	> 30	21	77	13.6
	其中 > 60	12	83	6.50
验证集	< 30	10	90	15.6
	> 30	10	88	6.15
	其中 > 60	3	100	12.2

应用 2005 年研究区水质监测部分数据对模型预测效果进行验证。根据式 (8) 计算水华爆发时机的预测成功率, 而爆发强度预测精度则根据式 (9) 计算。经过检验, 采取以上规则, 训练集和验证集成功率及其相对误差结果如表 4 所示。

从以上模型中可以得出以下的规律: 来水量、水温和总磷浓度是影响水华爆发的重要

因素。其中提高来水水量是防止研究区水华爆发的最有效措施, 当每月来水量 $Q > 38.66$ 万 $\text{m}^3/\text{月}$ 时, 就能避免研究区水华的严重爆发。水温也是研究区水华爆发的重要限制因子。结果表明, 水华爆发时温度一般高于 23.25 。总磷对水华爆发也有影响, 它排在最后, 暗示研究区水体营养盐浓度过高, 已经不成为水华爆发的最主要限制因子。

3 结 论

(1) 根据决策树分类模型和分段线性回归法将影响水华爆发的因素进行优选, 并对优选出的因素取值范围进行区间的划分, 寻找各区间的局部回归模型, 因此模型对于处理限制因素发生变化时水华预测结果更为准确, 且结构简单, 输入输出关系明显, 结果易于解释; 采取信号灯模型系统, 使得模型结果更加直观, 便于管理部门针对不同水华警戒水平采取不同的应对措施。

(2) 在春夏秋季每月来水量 $Q < 79.0$ 万 m^3 , 水温 $T > 13.4$, 水华预警为黄色; 有关部门应加强水质的监控, 防止水华突然爆发; 同时增加再生水或来水量, 通过闸门的优化调度, 控制水量和水力要素, 优先满足研究区的最小生态需水量; 同时恢复水生生态系统结构和功能的完整性。如滨水岸带生态构建技术, 水生植被和生物恢复技术等^[15~17]。

(3) 当夏季每月来水量 $Q < 38.7$ 万 m^3 时, 水温 $T > 23.25$, $TP > 0.13$ mg/L , 水华预警为红色; 必须采取应急措施来降低研究区水华爆发的强度和历时, 如增加来水水量, 改善水动力学条件; 削减水体的外源污染负荷; 通过对水体本身直接进行物理、化学、生物净化工程方法等应急技术措施, 就地改善和净化水体水质; 如可采用人工浮岛净化技术以及气浮、机械收藻、投放安全的微生物制剂等应急技术等。

参考文献:

- [1] 卢小燕, 徐福留, 詹 巍, 等. 湖泊富营养化模型的研究现状与发展趋势[J]. 水科学进展, 2003, 14(6): 792 - 798.
- [2] 夏 军, 奚 明. 水体富营养化综合水质模型及其应用研究[J]. 上海环境科学, 2000, 19(7): 302 - 308.
- [3] 彭 虹, 郭生练. 汉江下游河段水质生态模型及数值模拟[J]. 长江流域资源与环境, 2002, 11(4): 363 - 369.

- [4] 金腊华, 徐峰俊. 水环境数值模拟与可视化技术[M]. 北京: 化学工业出版社, 2004.
- [5] Walter M. Predicting eutrophication effects in the Burrinjuck reservoir (Australia) by means of the deterministic model SALMO and the recurrent neural network model ANNA[J]. *Ecological modeling*, 2001, 146: 97 - 113.
- [6] 周 勇, 刘 凡, 吴 丹, 等. 湖泊水环境预测的原理和方法[J]. *长江流域资源与环境*, 1999, 8(3): 305 - 311.
- [7] Maier H R, Dandy G C, Burch M D. Use of artificial neural networks for modeling cyanobacteria *Anabaena* spp. In the river murray, south Australia[J]. *Ecological modeling*, 1998, 105: 257 - 272.
- [8] Gurbuz H. Predicting dominant phytoplankton quantities in a reservoir by using neural networks[J]. *Hydrobiologia*, 2003, 504: 133 - 141.
- [9] 裴洪平, 罗妮娜, 蒋 勇. 利用 BP 神经网络方法预测西湖叶绿素 a 的浓度[J]. *生态学报*, 2004, 24 (2): 246 - 251.
- [10] Solomatine D P, Dulal K N. Model tree as an alternative to neural network in rainfall-runoff modeling[J]. *Hydrological Science Journal*, 2003, 48: 399 - 411.
- [11] Chen Q, Mynett A E. Predicting phaeocystis globosa bloom in Dutch coastal waters by decision trees and non-linear piecewise regression[J]. *Ecological Modeling*, 2004, 176: 277 - 290.
- [12] Mingers J. An Empirical comparison of selection measures for decision - tree induction[J]. *Machine learning*, 1989, 3: 319 - 342.
- [13] Quinlan J R. C4.5: Programs for machine learning[M]. San Mateo, California: Morgan Kaufmann, 1993.
- [14] 张文彤. SPSS11 统计分析教程高级篇[M]. 北京: 希望电子出版社, 2002.
- [15] 韩秀娣. 非点源污染最佳管理实践[J]. *上海环境科学*, 2000, 19(3): 102 - 104.
- [16] Webb A A, Erskine W D. A practical scientific approach to riparian vegetation rehabilitation in Australia[J]. *Journal of environmental management*, 2003, 68: 329 - 341.
- [17] 全为民, 沈新强, 严力蛟. 富营养化水体生物净化效应的研究进展[J]. *应用生态学报*, 2003, 14(11): 2057 - 2061.

Algalbloom prediction models for Liuhai-lake in Beijing city^{*}

ZENG Yong, YANG Zhi-feng, LIU Jing-ling

(State Key Laboratory of Water Environment Simulation, School of Environmental, Beijing Normal University, Beijing 100875, China)

Abstract: The algalbloom prediction models are constructed by using the decision trees to qualitatively predict bloom timing and use the nonlinear piecewise regression to quantitatively predict bloom intensity. The traffic light systems are used as the indicator for degree of algal bloom. Liuhai-lake in Beijing city is used an example. The constructed model indicates that the water inflow, the temperature and total phosphorus are the most impact factors on the algalbloom in Liuhai-lake. The concentration of Chl-a < 30 ug/L is labeled as green, 30 ug/L < Chl-a < 60 ug/L as yellow and Chl-a > 60 ug/L as red. When water inflow $Q > 79.0 \times 10^4 \text{ m}^3$ or $Q < 79.0 \times 10^4 \text{ m}^3$ and water temperature < 13.4, the indicator is green; when $Q < 79.0 \times 10^4 \text{ m}^3$, water temperature $T > 13.4$, the indicator is yellow; and when $Q < 38.7 \times 10^4 \text{ m}^3$, water temperature > 23.25, TP > 0.13 ug/L, the indicator is red. The model is test by an independent dataset from the same area, the predicted blooming time error rate and the error of predicted bloom intensity are presented in the paper. The model has great advantages to deal with the common problem in algal-blooms. It's more accurate when the limiting factor is changing. And the structure is understandable and easy to interpret.

Key words: algalbloom; prediction; decision trees; nonlinear regression; traffic light system; Liuhai-lake in Beijing city

^{*} The study is financially supported by the National Basic Research and Development Program of China (2006CB403303) and the Key Program of National Natural Science Foundation of China (No. 50239020).