

BP 神经网络水华预测模型的敏感性分析

殷高方¹, 张玉钧¹, 胡丽², 于绍惠¹, 肖雪¹,
王欢博¹, 石朝毅¹, 段静波¹, 刘文清¹

(1. 中国科学院环境光学与技术重点实验室 中国科学院安徽光学精密机械研究所, 安徽, 合肥, 230031;
2. 合肥学院 建筑工程系, 安徽, 合肥 230022)

摘要: 敏感性分析能够定量地评价模型输入变量的变化对输出结果产生的影响, 是揭示模型蕴含规律的有效途径. 本文将敏感分析方法应用于 BP 神经网络巢湖水华预测模型中, 分析结果表明巢湖水华形成受诸多环境因子共同影响, 水温、溶解氧、浊度、气温、光照强度等环境因子变化与藻类质量浓度变化相关, 其中气温是最大影响因素, 相对贡献率达到 17.01%; 气压的上升则不利于藻类质量浓度的增加; pH 值的上升对藻类质量浓度的影响有正有负.

关键词: 水华; 预测模型; BP 神经网络; 敏感性分析

中图分类号: X 824 **文献标志码:** A **文章编号:** 1001-0645(2012)12-1288-06

Sensitivity Analysis of BP Neural Network for Algal Bloom Prediction Model

YIN Gao-fang¹, ZHANG Yu-jun¹, HU Li², YU Shao-hui¹,
XIAO Xue¹, WANG Huan-bo¹, SHI Chao-yi¹, DUAN Jing-bo¹, LIU Wen-qing¹
(1. Key Laboratory of Environmental Optics & Technology, Anhui Institute of Optics and Fine Mechanics,
Chinese Academy of Sciences, Hefei, Anhui 230031, China; 2. Department of Architectural
Engineering, Hefei University, Hefei, Anhui 230022, China)

Abstract: Sensitivity analysis, which can quantitatively estimate the contribution of input variable to the output, is an effective way to reveal the inherent laws of the model. In this paper, sensitivity analysis was applied to the algal blooms forecast model based on BP neural network in Chaohu Lake. The result of the analysis indicates that algal blooms in Chaohu are affected by many factors. There was a positive correlation among the change of water temperature, dissolved oxygen, turbidity, atmospheric temperature, illumination and the change of mass concentration of algal. Among these factors, atmospheric temperature is the most important, with a relative contribution up to 17.01%; on the contrary, the rise of atmospheric pressure does harm to the algal; and the influence of high pH on the algal concentration is uncertain.

Key words: algal bloom; forecast model; BP neural network; sensitivity analysis

我国内陆湖泊和水库的水质富营养化不断加剧, 水体生态系统严重受损, 致使浮游植物群落中优势藻类过量增殖, 引发大面积的藻类水华, 影响生态环境并威胁着人体健康及饮用水安全. 分析水华形

收稿日期: 2011-12-12

基金项目: 国家“八六三”计划资助项目(2009AA063005); 合肥学院科研发展基金资助项目(12KY05ZR); 安徽光学精密机械研究所所长基金资助项目(Y03AG31144)

作者简介: 殷高方(1979—), 男, 博士生, E-mail: gfyin@aiofm.ac.cn.

通信作者: 张玉钧(1964—), 男, 博士, 研究员, E-mail: yjzhang@aiofm.ac.cn.

成规律及提高水华预测能力已经成为环境领域的研究热点,目前,这方面的主要研究方法大致可分为过程驱动方法和数据驱动方法两大类.过程驱动方法指以藻类的生态学概念为基础,利用生态变量和待定参数,通过具体数学解析式或经验公式描述水华发生的机理过程^[1].典型的过程驱动方法包括藻类种群生长和生产力的预测模型 PACGAP、浮游植物与环境因子关系模型 PROTECH-2 以及由美国国家环境保护局提出的多参数综合水质生态模型 WQASP^[2].然而,由于水华形成是一个从无机到有机、从低级生物到高级生物的复杂生态过程,不同湖库藻类生消规律差异性大,种种原因让基于生态学机理的过程驱动模型研究陷入了困境.进入 90 年代,随着计算机技术的发展和迅速普及,以现代数学理论为基础的各种数据驱动方法在水华形成过程分析中得到广泛应用,该方法是以建立输入输出数据之间的最优数学关系为目标的黑箱子方法,典型的数据驱动方法包括模糊评价^[3]、灰色评价^[4]、人工神经网络^[5]、遗传算法^[6]等.其中人工神经网络性能突出,尤其在多变量非线性关系的建模方面具有高度的非线性映射能力,能够较好地解决水华与影响因子间多元、高阶和非线性的关系问题,在水华预测的研究中得到了广泛应用^[7].

人工神经网络虽然能解决形成机制不明确的水华预测问题,但却很难解释由模型得到的规律,在水华形成规律的研究方面未能发挥其应有的价值.因此,本文作者在前期建立的 BP 人工神经网络巢湖水华预测模型的基础上,引入敏感性分析方法来评估模型的输入因子对输出因子的影响,深入分析影响巢湖水华形成的主要环境因子,探寻水华对环境因子的响应规律.

1 水华预测模型

1.1 数据来源

水华预测模型的建模数据来源于巢湖西半湖的浮标水质监测站点和距离浮标站点 3 km 处的合肥骆岗机场气象站,具体位置如图 1 所示.主要监测参数包括水温、pH、氧化还原电位、电导率、氨离子、硝酸根离子、氯离子、浊度、溶解氧等常规水质参数;蓝藻叶绿素 a 浓度、绿藻叶绿素 a 浓度、褐藻叶绿素 a 浓度、总叶绿素 a 浓度等浮游植物浓度表征参数;气温、大气压、能见度、风速、风向、光照强度等气象参数.常规水质参数监测设备采用美国 HACH 公

司 Hydrolab DS5 水质多参数传感器,浮游植物质量浓度监测设备采用作者单位研制的便携式水体藻质量浓度原位监测仪 AGHJ-FMA,气象参数由合肥骆岗机场气象站提供.数据来源于上述系统在 2009 年 9~11 月份期间实时监测数据,数据监测周期为 1 h,除去个别时间由于通信网络等原因造成的监测数据中断,最终获得的数据记录共计 1 400 余条.

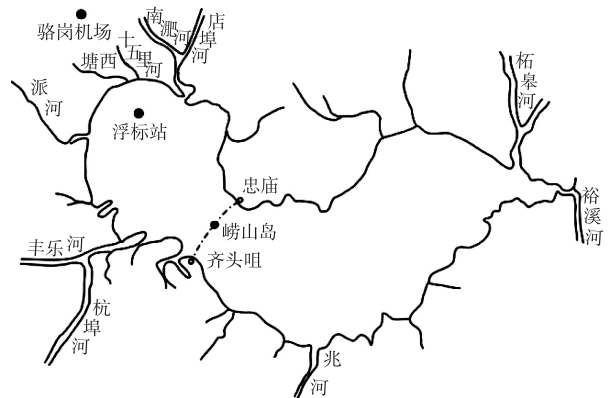


图 1 巢湖外场数据监测点
Fig. 1 Monitoring point in Chaohu lake

1.2 BP 神经网络水华预测模型

水体预测模型选用含有输入层、隐含层和输出层的三层 BP 神经网络模型.由于水体叶绿素 a 浓度是表征水体中藻类现存量的最直接指标,故将水体中总叶绿素 a 浓度作为模型的输出因子,通过对叶绿素 a 浓度的预测可以间接实现对藻类引发的水华进行预测.在确定模型输出因子后,合理选择网络的输入因子对正确应用 BP 模型和保证模型预报精度非常重要.因为输入因子中,可能存在与输出因子关联弱的噪声因子,或者重叠反映系统信息的冗余因子.经过对数据进行相关性分析和主成分分析,筛选出与输出因子相关而双彼此无关的 10 个环境因子作为模型的输入因子:水温、pH、溶氧、浊度、气温、气压、光照强度、蓝藻叶绿素 a 浓度、绿藻叶绿素 a 浓度和褐藻叶绿素 a 浓度.通过数据预处理、网络结构和训练模式设计,最终建立的模型网络结构如图 2 所示.模型的训练和测试结果表明模型的稳定性和预测能力良好;训练组决定系数分别达到了 0.61,对巢湖未来 24 h 藻类叶绿素 a 浓度连续 12 d 预测结果与实际观测值间的决定系数达到 0.608 4.描述了详细的建模过程及预测结果分析^[8].以上数据分析及建模采用美国 MathWorks 公司的数学软件 Matlab R2009a.

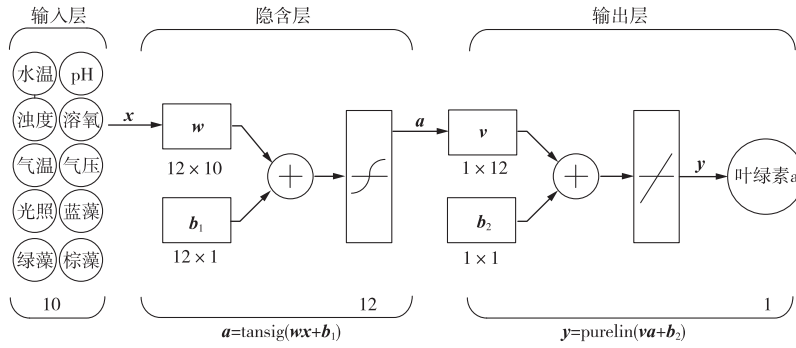


图 2 水华预测模型网络拓扑结构
Fig. 2 Topological structure of algal blooms forecast model

为使下文中的敏感性分析不失一般性, BP 神经网络描述如下: 输入层有 N 个输入变量 $x^T(x_1, x_2, \dots, x_N)$, x 为 $N \times 1$ 向量, 隐含层有 L 个神经元 $h(h_1, h_2, \dots, h_L)$, 输出层有 M 个输出变量 $y(y_1, y_2, \dots, y_M)$, y 为 $M \times 1$ 向量, 输入层与隐含层间连接权重 w 为 $N \times L$ 矩阵, 隐含层与输出层间的连接权重 v 为 $L \times M$ 矩阵. $f(x)$ 和 $g(x)$ 分别表示隐含层神经元的激活函数和输出层神经元的激活函数. 网络有 m 个训练样本, n 个测试样本, 如图 3 所示.

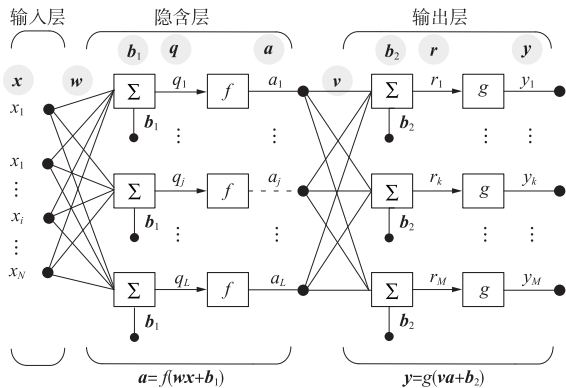


图 3 3 层 BP 神经网络
Fig. 3 Three layers BP neural network

2 敏感性分析

敏感性分析(sensitivity analysis)是一种定量描述模型输入变量对输出变量的重要性程度的方法. 令模型的每个输入变量在可能的取值范围内变动, 研究这些变量的变动对模型输出值的影响, 将影响程度的大小称为该变量的敏感性系数^[9]. 敏感性系数越大, 说明该变量对模型输出的影响越大, 实际应用中可依据敏感性系数的排序结果, 解决相应的问题. 根据实现方法的不同, 敏感性分析可以分为以下几类: ① 基于连接权的敏感性分析方法; ② 基

于输出变量对输入变量偏导的敏感性分析方法; ③ 统计方法结合的敏感性分析方法; ④ 基于输入变量扰动的敏感性分析方法^[10].

在已有的敏感性分析方法中, 函数偏导在计算输入变量对模型输出的影响方面, 具有方法简单、计算快捷的优点. 因此, 本文中采用偏导敏感性分析方法来研究 BP 神经网络各输入变量的敏感性. 根据偏导敏感性分析方法原理, 在图 3 所示的三层 BP 神经网络中输入变量 x 对网络输出变量 y 的敏感性系数矩阵 s 可由下式计算得到:

$$s = \frac{\partial y}{\partial x} = \frac{\partial y}{\partial r} \cdot \frac{\partial r}{\partial a} \cdot \frac{\partial a}{\partial q} \cdot \frac{\partial q}{\partial x} = (g(r))' \cdot v \cdot (f(q))' \cdot w. \quad (1)$$

敏感性系数矩阵 s 有单样本敏感性系数和全样本敏感性系数两种表达形式: 单样本敏感性系数则能体现模型输入变量对输出变量的作用方式; 全样本敏感性系数直接反映了模型输入变量对输出变量的贡献率的大小排序情况.

2.1 单样本敏感性系数

在图 3 所示的 BP 神经网络中, 网络各层间的输入输出映射关系如下.

第 t 个样本点隐含层第 j 个神经元输入为

$$q_j^t = \sum_{n=1}^N \omega_{nj} x_n^t. \quad (2)$$

第 t 个样本点隐含层第 j 个神经元输出为

$$a_j^t = f(q_j^t) = f\left(\sum_{n=1}^N \omega_{nj} x_n^t\right). \quad (3)$$

第 t 个样本点输出层第 k 个神经元输入为

$$r_k^t = \sum_{l=1}^L v_{lk} a_l^t = \sum_{l=1}^L v_{lk} f\left(\sum_{n=1}^N \omega_{nl} x_n^t\right). \quad (4)$$

第 t 个样本点输出层第 k 个神经元输出为

$$y_k^t = g(r_k^t) = g\left[\sum_{l=1}^L v_{lk} f\left(\sum_{n=1}^N \omega_{nl} x_n^t\right)\right]. \quad (5)$$

用标量形式表示第 t 个样本的第 i 个输入变量 x_i^t 对第 k 个输出变量 y_k^t 的敏感性系数为

$$s_{i,k}^t = \frac{\partial y_k^t}{\partial x_i^t} = g'(r_k^t) \sum_{l=1}^L v_{lk} f'(q_l^t) w_{il}. \quad (6)$$

式中 $s_{i,k}^t$ 为单样本敏感性系数,它反映了单个样本点上输入变量对输出变量的贡献。

2.2 全样本敏感性系数

一般情况下,单样本敏感性系数不能全面反映该输入变量对输出变量总的影响,还需要一个综合函数将单个样本点上反映的输入变量对输出变量的敏感性综合起来,得到输入变量 x_i 总体对输出变量 y_k 的影响程度.用于全样本敏感性系数计算的综合方法有多种,这里选用均方根方法表示为^[11]

$$S_{i,k} = \sqrt{\sum_{t=1}^T (s_{i,k}^t)^2} / T, \quad (7)$$

式中 $S_{i,k}$ 为输入参数 x_i 对输出参数 y_k 的全样本敏感性系数; T 为参与敏感性分析的样本数.在神经网络训练完成,网络的权重唯一确定后,便可计算出输入层每个变量对输出层变量敏感性系数。

3 结果分析与讨论

由于水华预测模型的输入输出因子量纲不同,为使获取的各环境因子的敏感性系数有可比性,先将输入输出因子归一化,然后再利用归一化后的数据样本结合水华预测模型进行敏感性分析,分析结果如图 4,图 5 所示.从图 4 模型输入因子的全样本敏感性系数可以看出:巢湖水华形成并非由某个环境因子所决定,而是多个环境因子共同作用的结果,其中气温是最大的影响因子,它的全样本敏感性系数为 0.232,依次是蓝藻质量浓度、溶氧、水温 and pH,敏感性系数分别为 0.214, 0.213, 0.169 和 0.138.

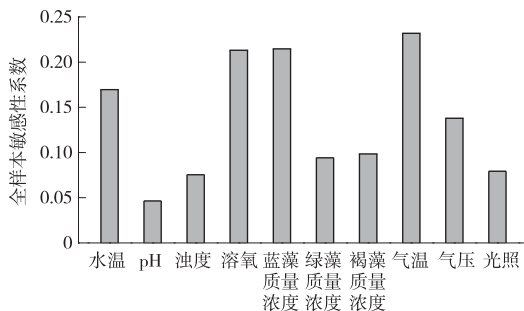


图 4 输入因子全样本敏感性系数
Fig. 4 Sensitivity coefficient of full sample

从图 5 模型的输入因子的单样本敏感性系数的

散点图可以看出:

① 如图 5(a)所示:水体温度上升对藻类贡献为正,即随着温度的升高,藻类质量浓度也随之增加,而且水温高时,水温的敏感性系数高,也就是说水温高时水温变化引起藻类质量浓度变化明显。

② 如图 5(b)所示:pH 增加对藻类质量浓度的影响有正有负,pH 较低时,随着 pH 值增加,藻类质量浓度增加;pH 值过高时,随着 pH 值的增加,藻类质量浓度却呈现下降趋势。

③ 如图 5(c)所示:浊度增加对藻类生长贡献为正,低浊度时,浊度敏感性系数变化范围较大,即影响作用大小不稳定,但浊度值超过 250NTU 后,对藻类的影响趋于稳定,且影响作用变弱。

④ 如图 5(d)所示:溶解氧的上升对藻类质量浓度的贡献为正,低溶氧时,溶氧的上升影响作用无太大变化,当溶氧超过 $12.5 \mu\text{g/L}$ 时,溶氧上升带来的影响突然增加。

⑤ 如图 5(e)所示:蓝藻质量浓度增加对未来一天的藻类贡献为正,但敏感系数变化范围大,即对藻类质量浓度影响大小不稳定。

⑥ 如图 5(f)~5(g)所示,绿藻、褐藻质量浓度增加对未来一天藻类贡献均为负,且随着高质量浓度时,这种影响大小趋于稳定。

⑦ 如图 5(h)所示:与水温一样,气温的上升对藻类贡献为正;与水温不同的是,随着气温的升高,敏感性系数呈减小趋势,也就是说低气温时气温变化引起藻类质量浓度变化明显。

⑧ 如图 5(i)所示:大气压强的增加对藻类质量浓度起负作用,在较高气压时,该负作用较小。

⑨ 如图 5(j)所示:光照强度的增加对藻类质量浓度的影响为正,高光照时光照强度增加带来的影响较大。

进一步分析巢湖水华预测模型的单样本敏感性分析结果,能够间接地获取一些巢湖水华生消规律。

① 水温、气温、浊度、光照强度等环境因子的上升是藻类质量浓度增加有利因素;水体藻类生物量的增加,由光合作用过程产生的氧气量随之增加,因此在 9~10 月巢湖藻类增长期内,水体溶解氧质量浓度也随之增加;气压的上升则不利用于藻类质量浓度的增加;pH 值的增加对藻类的影响有正有负,pH 值适当时,pH 的增加能够促进藻类生长,pH 值过高时,这种增加却会抑制藻类生长。② 巢湖水华的藻类质量浓度发展趋势与蓝藻质量浓度变化呈正

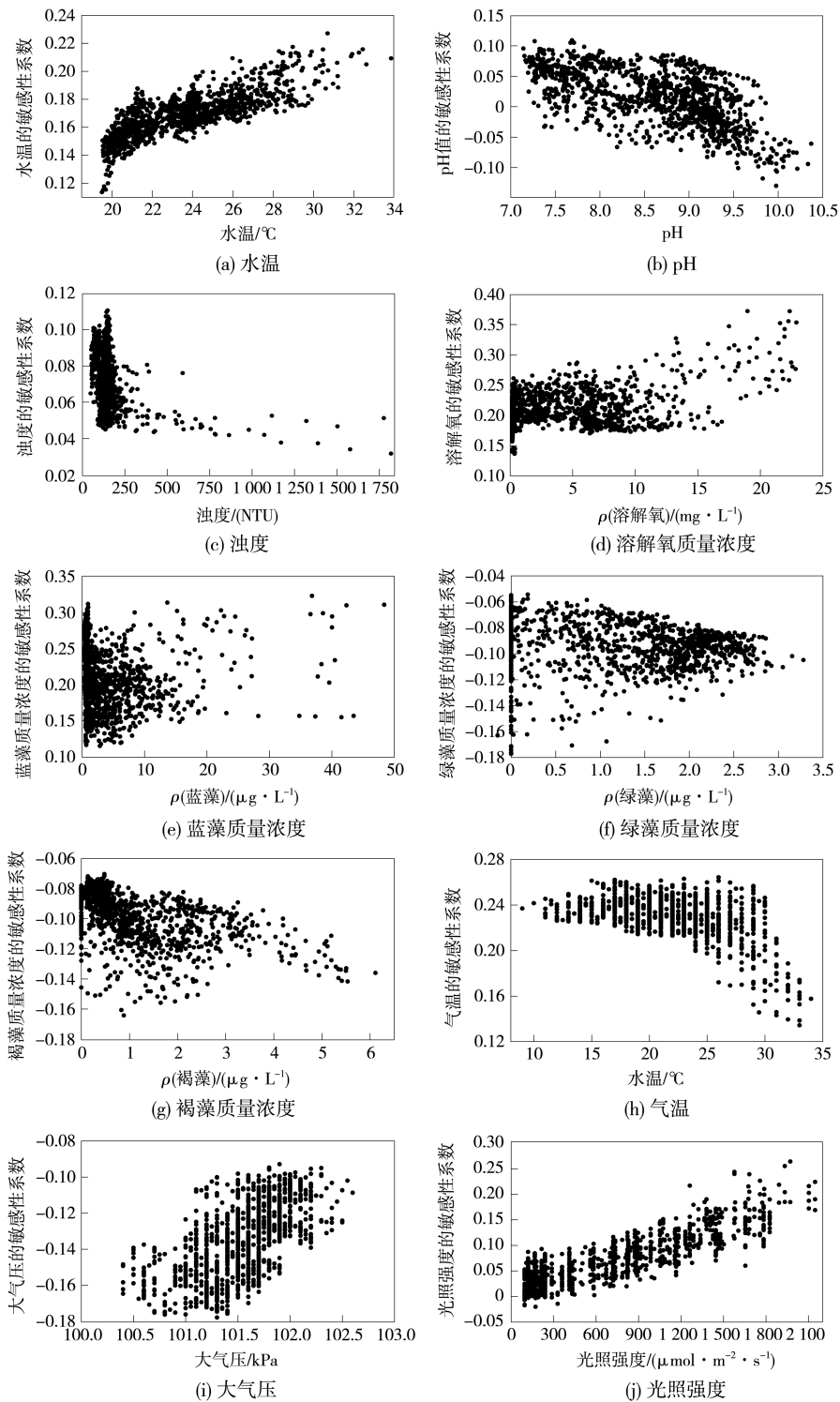


图 5 输入因子单样本敏感性系数的散点图
 Fig. 5 The scatter diagram sensitivity coefficient of one sample

相关,而与绿藻、褐藻质量浓度变化呈负相关,说明形成巢湖水华的优势藻类是蓝藻,并且在蓝藻质量浓度上升时,其它藻类质量浓度减小. ③ 虽然蓝藻是巢湖水华的优势藻类,但蓝藻的质量浓度对巢湖水华的贡献并不稳定,如图 5(e)所示,分析认为,可

能是夏秋季节巢湖水体中主要藻类为铜绿微囊藻,它具有一种调节细胞沉降的结构——伪空胞,极易受到外界光照强度、温度、风浪等环境因子的影响而在水体中迁移聚集,因而仅仅依据蓝藻质量浓度很难推测未来水华的发展趋势. 以上结论与已有文献

研究结果相符^[12-15]。除此之外,敏感性分析结果还给出一些现有文献尚未提及的信息,如当溶氧质量浓度超过 12.5 $\mu\text{g/L}$ 时,溶氧变化对藻类质量浓度的影响突然增加,大气压的增加对藻类质量浓度有负作用,藻类质量浓度对水温和气温响应规律的差异等。

4 结 论

基于巢湖水华的 BP 神经网络模型,利用偏导敏感性分析方法对巢湖水华成因进行了分析,克服神经网络模型不可解释的缺陷,给出单个环境因子与藻类质量浓度变化间的关系,为水华形成规律的研究提供数据参考。另外,作者已意识到水华的生消伴随着各种物理、化学和生物过程,各种环境因子之间相互作用并互为因果,绝非由单一环境因子扰动而引起的。因此,作者建议今后可利用多变量共同作用下的敏感性分析方法,进一步分析模型的输入变量间的相互作用对模型输出的影响,更全面地了解多个环境因子共同作用支配下水华的生消规律。

参考文献:

- [1] Seip K L. The ecosystem of a mesotrophic lake-I, simulating plankton biomass and the timing of phytoplankton blooms [J]. *Aquatic Sciences*, 1991, 53(2/3):239-262.
- [2] Ferguson A J D. The role of modelling in the control of toxic blue-green algae [J]. *Hydrobiologia*, 1997, 349: 1-4.
- [3] Marsili-Libelli S. Fuzzy prediction of the algal blooms in the Orbetello lagoon [J]. *Environmental Modelling & Software*, 2004, 19:799-808.
- [4] Ni Haier, Zhou Ruijuan. Grey evaluation and forecast of sustainable use of fisheries resources in Zhoushan fisheries ground [J]. *Oceanologia et Limnologia Sinica*, 2009, 40(3):319-324.
- [5] Recknagel F, Bobbin J, Whigham P, et al. Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes [J]. *Journal of Hydroinformatics*, 2002, 4(2):125-133.
- [6] Bobbin J, Recknagel F. Inducing explanatory rules for the prediction of algal blooms by genetic algorithms [J]. *Environment International*, 2001, 27(2):237-242.
- [7] Scardi M. Advances in neural network modeling of phytoplankton primary production [J]. *Ecological Modelling*, 2001, 146:33-45.
- [8] 殷高方,张玉钧,胡丽,等. BP 神经网络在巢湖水华预测中的应用研究 [J]. *北京理工大学学报:自然科学版*, 2012, 32(6):655-660.
Yin Gaofang, Zhang Yujun, Hu Li, et al. The application of BP neural network in algal blooms short-term forecast [J]. *Transactions of Beijing Institute of Technology*, 2012, 32(6):655-660. (in Chinese)
- [9] Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models [J]. *Ecological Modelling*, 2003, 160:249-264.
- [10] Dimopoulos Y, Bourret P, Lek S. Use of some sensitivity criteria for choosing networks with good generalization ability [J]. *Neural Processing Letters*, 1995, 2(6):1-4.
- [11] 蔡毅,邢岩,胡丹. 敏感性分析综述 [J]. *北京师范大学学报:自然科学版*, 2008(1):9-16.
Cai Yi, Xing Yan, Hu Dan. On sensitivity analysis [J]. *Journal of Beijing Normal University: Natural Science*, 2008(1):9-16. (in Chinese)
- [12] 鱼京善,崔国庆. 北京动物园水体水华发生的生态学机理 [J]. *环境工程*, 2004(4):62-65.
Yu Jingshan, Cui Guoqing. The ecological mechanism of algal bloom in bodies of water at the Beijing Zoo [J]. *Environmental Engineering*, 2004(4):62-65. (in Chinese)
- [13] 王志红,崔福义,安全,等. pH 与水库水富营养化进程的相关性研究 [J]. *给水排水*, 2004(5):37-41.
Wang Zhihong, Cui Fuyi, An Quan, et al. Study on influence of pH on the advance of eutrophication in reservoir [J]. *Water & Wastewater Engineering*, 2004(5):37-41. (in Chinese)
- [14] 姜霞,王书航,钟立香,等. 巢湖藻类生物量季节性变化特征 [J]. *环境科学*, 2010(9):2056-2062.
Jiang Xia, Wang Shuhang, Zhong Lixiang, et al. Seasonal variation characteristics of algae biomass in Chaohu Lake [J]. *Environmental Science*, 2010(9): 2056-2062. (in Chinese)
- [15] 孔繁翔,高光. 大型浅水富营养化湖泊中蓝藻水华形成机理的思考 [J]. *生态学报*, 2005(3):589-595.
Kong Fanxiang, Gao Guang. Hypothesis on cyanobacteria bloom-forming mechanism in large shallow eutrophic lakes [J]. *Acta Ecologica Sinica*, 2005(3):589-595. (in Chinese)