

# 水库浮游植物群落动态的人工神经网络方法\*

邬红娟 郭生练<sup>†</sup> 胡传林 刘跃<sup>‡</sup>

(水利部、中国科学院水库渔业研究所 武汉 430079)

<sup>†</sup>(武汉水利电力大学 武汉 430072)

<sup>‡</sup>(辽宁省供水局 沈阳 110003)

**提要** 根据辽宁大伙房水库 1980—1997 年的水文和湖沼学观测资料, 分别建立浮游植物丰度和蓝藻优势度人工神经网络模型。将年降雨量、7—9 月平均水温、7—8 月入库水量与 7—8 月库容之比和磷酸盐作为输入, 浮游植物生物量和丰度作为输出, 建立浮游植物群落消长的人工神经网络模型; 将 7—9 月平均水温、7—8 月入出库水量之比、磷酸盐和总氮作为输入, 蓝藻优势度作为输出, 建立浮游植物演替的人工神经网络预测模型, 并进行检验, 其模拟值与观测值平均相对误差分别为 2% 和 1%。结果表明, 人工神经网络方法优于传统的统计学模型, 可进行水库浮游植物群落动态的预测预报, 并具有较高的精度。

**关键词** 浮游植物群落动态, 人工神经网络, 预测

**中图分类号** Q141

近年来, 经济的发展加速了水体富营养化, 因浮游植物过度生长而引起的近海赤潮和湖泊水华, 已经给国民经济和人民生活带来了极大的危害, 引起了全社会的关注(黄长江等, 1999, 2000a, b)。因此, 迫切需要识别浮游植物群落变化规律并对其进行准确的预测或预报。目前有关水质及其影响因子的关系分析和现状评价研究很多, 但预测预报的研究结果甚少。人们不仅仅关注水质的现状, 更关注水质变化的原因以及将来可能发生的水质及突发事件。因此, 对浮游植物群落变化规律及其预报方法的研究显得尤为重要。本文根据辽宁大伙房水库 1980—1997 年间的可调控因子, 如水文和湖沼学观测资料(史为良等, 1994, 1995; 雷宝玲等, 1995), 采用人工神经网络方法对浮游植物群落变化与环境因子的关系进行识别和预测, 以期对水库湖沼学的研究和水资源可持续利用的科学管理提供资料。

## 1 人工神经网络(Artificial Neural Network, 简称 ANN)

### 1.1 结构

在 ANN 模型中应用较多的是具有非线性传递函数神经原构成的前馈网络中采用误差反传算法作为其学习算法的前馈网络(Back propagation algorithm, BP), 简称 BP 网络。其结构包括输入层、隐含层和输出层(图 1)。同一层神经原之间不互连, 不同层神经原

\* 国家自然科学基金资助项目, 59779008 号。邬红娟, 女, 出生于 1960 年 1 月, 博士, 副研究员, Fax: 0086-027-87803022

收稿日期: 1999-05-31, 收修改稿日期: 2000-03-03

之间则全互连。神经网络的权重是由前馈或反馈通过若干个神经原(计算元素)相互连接,这些神经原位于隐含层,并通过其连接输入层和输出层。

### 1.2 算法原理

首先分析出影响预测对象的因子,并将其作为 BP 网络的输入,每一个影响因子对应网络输入层的一个节点,将预测对象作为 BP 网络的输出,输入层和输出层的节点树由具体的预测问题来确定。当 BP 网络确定后,用该网络对样本进行监督学习,从而识别出影响因子与预测对象之间复杂的非线性映射关系。

BP 网络的核心是运用链导数法则将误差函数对连接权和阈值的一阶导数自输出层反向传播到输入层,然后由这些导数按梯度下降法修改权和阈值。由于在极小点附近误差对权和阈值的一阶梯度趋近于零,该法在接近极小点时收敛速度较慢,在局部极小问题和一定程度上妨碍了前馈网络的更深入的应用。为此,本文采用将 BP 和共轭梯度两种算法结合起来的 BP 网络联合梯度法(Hybrid Gradient Method, HGM)。共轭梯度法在目标函数二次性较强的区域中具有较好的收敛性,误差函数在极小点附近的状态又恰好近似于二次函数。

BP 网络的数学表达式:设输入层、隐含层和输出层神经原节点数分别为  $N_1$ 、 $N_2$  和  $N_3$ ,隐含层和输出层的神经原的传递函数为 Sigmoid 函数,如

$$f(u_j) = \frac{1}{1 + e^{-u_j}}$$

假定有  $P$  对训练样本( $I_p, T_p, p = 1, 2, 3, \dots, P$ ),其中  $I_p \in R^{N_1}$  为第  $p$  个训练样本的输入,  $T_p \in R^{N_3}$  为第  $p$  个训练样本的期望输出,输入信号由输入层向输出层正向传播的过程可用以下公式表示:

$$net_i^I = I_{pi}; O_{pi}^I = net_i^I = I_{pi}; net_j^H = \sum_{i=1}^{N_1} W_{ji}^F O_{pi}^I - \theta_j^H; O_{pj}^H = f(net_j^H);$$

$$net_k^O = \sum_{j=1}^{N_2} W_{kj}^S O_{pj}^H - \theta_k^O; O_{pk}^O = f(net_k^O).$$

$$i = 1, 2, \dots, N_1; j = 1, 2, \dots, N_2; k = 1, 2, \dots, N_3.$$

式中,  $net_i^I$ 、 $net_j^H$  和  $net_k^O$  分别表示输入层中某一节点  $i$ 、隐含层中某一节点  $j$  和输出层中某一节点  $k$  的净输入;  $W_{ji}^F$ 、 $W_{kj}^S$  分别表示隐含层中节点  $j$  和输入层中节点  $i$  以及输出层中节点  $k$  和隐含层中节点  $j$  两两节点之间的连接权;  $\theta_j^H$ 、 $\theta_k^O$  分别表示隐含层中节点  $j$ 、输出层中节点  $k$  的阈值;  $O_{pi}^I$ 、 $O_{pj}^H$ 、 $O_{pk}^O$  分别表示前馈网络在输入第  $p$  个训练样本时由输入层节点  $i$ 、隐含层节点  $j$  和输出层节点  $k$  产生的输出。

令前馈网络参数向量

$$X(n + 1) = (W_{ji}^F, W_{kj}^S, \theta_j^H, \theta_k^O)$$

向量  $X$  的维数为  $N = N_1 \times N_2 + N_2 \times N_3 + N_2 + N_3$ , 令

$$G(X) = \frac{\partial E}{\partial X}, H(X) = \frac{\partial^2 E}{\partial X^2}$$

则多层前馈网络在极值点附近的共轭梯度学习算法可简单地用下式表示:

$$X(n+1) = X(n) + \lambda^n P(n)$$

$$\lambda^n = \frac{G(X(n))^T \cdot G(X(n))}{(P(n))^T H(X(n)) P(n)}$$

$$P(n) = -G(X(n)) + \beta^{(n-1)} P(n-1)$$

$$\beta^{(n-1)} = \frac{G(X(n))^T \cdot G(X(n))}{G(X(n-1))^T \cdot G(X(n-1))}$$

式中,  $G(X)$  为误差函数在  $X(n)$  处的梯度;  $H(X)$  是误差函数在  $X(n)$  处的 Hessian 矩阵;  $n$  为网络的训练次数。

### 1.3 特征

ANN 是生物神经网络的抽象和模拟, 它是由大量人工神经原广泛连接而成的超大规模的非线性动力学系统, 除具有一般非线性动力学系统的共性, 如不可预测性、吸引力、耗散性、非平衡性、不可逆性、高维性等特性外, ANN 还具有以下几方面的特征:

(1) 大规模并行处理能力和信息的分布式存储。由大量人工神经原构成的 ANN, 每个神经原的结构和功能都相对简单, 但其工作方式却是并行的, 信息(或学习到的知识)广泛地分布在神经网络拓扑结构和神经原之间的相互连接中。

(2) 容错性和稳健性。主要表现在神经网络中某个神经原的损坏或某两个神经原之间的联接损坏, 或某个神经原的输入有偏差时, 网络的性能(如学习性能、泛化性能、稳定性和收敛性能等)没有明显改变。

(3) 自学习功能。通过训练样本的学习, 神经网络能自动调整自己的拓扑结构及网络参数, 从而学习到蕴藏在样本中的知识, 以实现训练样本的联想记忆功能或对新的样本的泛化能力等。

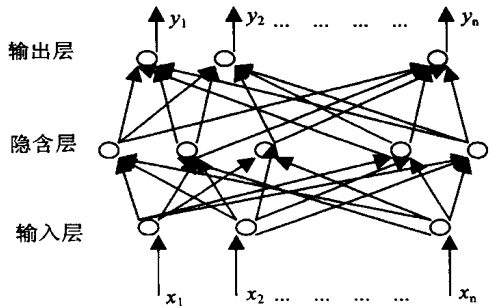


图 1 BP 网络模型结构

Fig. 1 Back propagation neural network model architecture

## 2 水库浮游植物群落演替 ANN 模型

大伙房水库位于辽宁省抚顺市东部, 建于 1958 年, 积水面积 5440km<sup>2</sup>, 平均年径流量 65 × 10<sup>8</sup> m<sup>3</sup>, 总库容 2.19 × 10<sup>8</sup> m<sup>3</sup>, 水库以防洪灌溉、城市生活及工业用水水源为主, 兼有发电、养鱼等综合利用的大型水利枢纽工程。

尽管目前大伙房水库水质仍处于国家 II 级标准, 但每到夏季却连年发生水华, 1975、1977 和 1979 年在库区曾发生大面积水华, 1980 和 1981 年也发生过局部较短时间的水华, 最严重的是 1982 年, 水华覆盖全库, 持续时间长达 2 个月, 数量达 773 × 10<sup>4</sup> 个/L, 其中微囊藻占 90% 以上。

### 2.1 大伙房水库浮游植物丰度 ANN 模型

浮游植物生长和繁殖除受水体氮磷等营养元素含量影响外, 还受温度及水库水文特征的影响。调查表明, 大伙房水库的富营养化现象并不是随时间逐年加剧, 而是取决于每年径流量大小, 雨水冲刷和河流带入库内的营养盐类, 浮游植物现存量与每年的库容有关

(雷宝玲等, 1995)。

经过对大伙房水库 13 年 10 余个参数与浮游植物进行多元回归分析, 筛选出相关显著的参数构成模型结构。现以年降雨量、7—9 月水温、7—8 月入库水量/7—8 月库容、7—8 月出库水量/7—8 月库容和磷酸盐等 5 项因子为输入端, 浮游植物生物量和丰度为输出端。输入端不仅包含了直接影响浮游植物生长和繁殖的营养盐类和水温, 也包含了控制库内营养的水文因子。

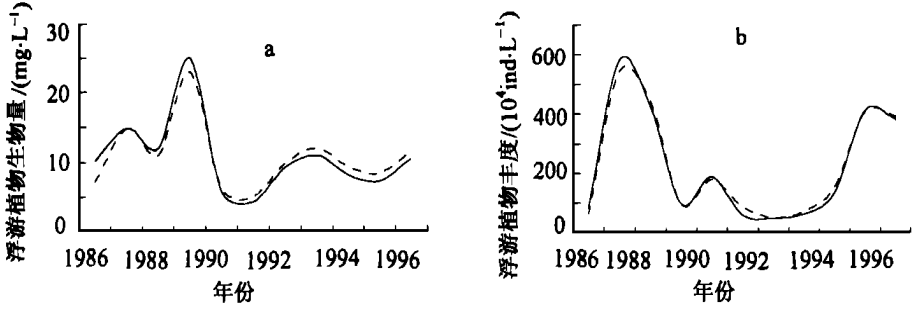


图 2 浮游植物动态的 ANN 模拟

Fig. 2 Phytoplankton dynamics modled by artificial neural network

a. 浮游植物生物量 (Phytoplankton biomass) (mg/L); b. 浮游植物丰度 (Phytoplankton abundance) ( $10^4$  ind/L)。——实测值 (Practical determination); - - - - 模拟值 (Simulation)

表 1 大伙房水库浮游植物群落变化 ANN 模型与多元回归方程预报结果误差比较

Tab. 1 The comparison of prediction for phytoplankton dynamics by ANN and multi-regression in Dahuofang Reservoir

浮游植物丰度 ( $10^4$ ind/L)				浮游植物生物量 (mg/L)			
实测值	模拟值	绝对误差	相对误差	实测值	模拟值	绝对误差	相对误差
410.95	449.10	38.15	9.2%	12.27	12.53	0.26	2.1%
349.98	349.98	0	0	10.09	10.32	0.27	2.2%

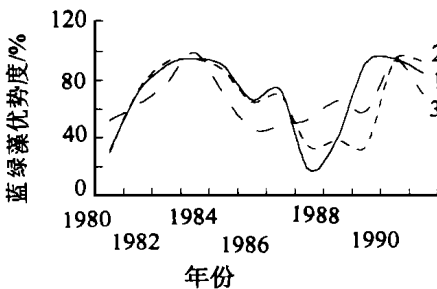


图 3 大伙房水库蓝藻优势度 ANN 模拟  
Fig. 3 The succession of Cyanophyta dominant modled by artificial neural network in Dahuofang Reservoir

1. 实测值 (Practical determination); 2. ANN 模拟值 (ANN simulation); 3. 回归模拟值 (Regression simulation)

神经网络模型的输入层、隐含层、输出层的节

点数  $N_1$ 、 $N_2$ 、 $N_3$  分别为 5、6、2, 训练精度为 0.05, 转入联合梯度算法精度为 0.02, 11 年系列数据进行学习, 2 年数据预测检验。模拟和预测结果见图 2 和表 1。

### 2.2 蓝藻优势度 ANN 模型

水库浮游植物种类的演替不仅受营养和温度的影响, 还受环境扰动的影响(主要是水库的水文因子) (Harris, 1983; Sommer, 1981; Reynolds, 1997)。根据大伙房水库 12 年相关资料分析, 筛选出 7—9 月水温、8 月入水量/8 月出水量、磷酸盐和总氮(TN)为输入, 8 月蓝藻占藻类总数的百分比为输出。其 ANN 模型输入层、隐含层、输出层的节点

数  $N_1$ 、 $N_2$ 、 $N_3$  分别为 4、5、1, 训练精度为 0.05, 转入联合梯度计算精度为 0.02, 10 年系列数据进行学习训练, 2 年系列数据预测检验。模拟结果见图 3 和表 2。

表 2 大伙房水库蓝藻优势度 ANN 模型与多元回归方程预报结果误差比较

Tab. 2 The comparison of prediction for Cyanophyta dominant by ANN and multi-regression in Dahuofang Reservoir

年份	实测值(%)	ANN 模拟(%)	相对误差	回归模拟(%)	相对误差
1980	30	28.5676	0.05	51.6	-0.72
1981	70	71.2682	-0.02	61.8	0.12
1982	90	92.2146	-0.02	77.2	0.14
1983	95	94.4568	0.01	99.1	-0.04
1984	90	86.6706	0.04	70.4	0.23
1985	66	64.5701	0.02	45.8	0.31
1986	73	68.5664	0.06	48.1	0.34
1987	18	33.8164	-0.889	52.4	-1.91
1988	40	38.3112	0.04	65.5	-0.64
1989	92	34.3503	0.63	58.3	0.37
1990	95	93.9637	0.01	93.5	0.02
1991	85	93.2477	-0.1	70	0.18
平均相对误差			-0.01		-0.14

### 3 讨论

每一种生物在其生境中都是一个超维体(hypervolume),其生长繁殖和演替受多种生物和非生物因子的影响,它们之间的关系表现出复杂的高度非线性和不确定性。ANN 具有良好的非线性映射逼近性能,通过神经原作用函数的简单复合就能逼近有限子集的任意非线性函数(胡铁松,1997),试验表明,ANN 模拟精度较高,其预测值与实测值误差比多元回归误差要小,说明能够用于水库浮游植物群落演替的模拟和预测。

尽管已有人将 ANN 模型用于环境因子与叶绿素关系模拟(Karul *et al.*, 1998)以及环境因子与湖泊浮游植物种类关系模拟(Recknagel, 1998),但他们均未对模型进行检验。ANN 是对人类大脑结构和功能的一种模仿,使其具有许多类似人类智能的特点,如果信息输入不正确,也能通过不断学习和训练得出较好的模拟结果,但其泛化或外延效果却可能很差。因此参数的选择对于 ANN 模型的精度很重要。

本文的实验结果表明,选择年降雨量、7—9 月水温、7—8 月入库水量/7—8 月库容、7—8 月出库水量/7—8 月库容、磷酸盐对水库浮游植物丰度的映射,以及选择水库入出流之比,水温、磷酸盐和总氮作为对蓝藻优势度的映射是正确的。同时也表明水库水文情势、水温和营养是影响浮游植物的主要因子。一方面,水文情势决定了水环境的扰动频率和稳定性,直接影响浮游植物群落的丰富度、优势度和结构(Harris, 1983; Sommer, 1981; Reynolds, 1997; Kar+Erich *et al.*, 1998);另一方面,水文情势也影响水库的水温和营养,间接影响浮游植物群落的丰富度、优势度和结构。调查表明,大伙房水库浮游植物丰度不随时间变化,而与水库的库容有关(雷宝玲等,1995);藻类水华(蓝藻优势度为 85% 以上)并不是年年夏季出现,经分析发现,藻类水华与每年夏季入出流之比关联度最大。因此,通过水库对水文情势的调控,即可控制藻类水华的发生。

### 4 结语

由于水库水文变化和营养状况以及浮游植物群落变化所表现出的高度非线性,用传

统的回归分析方法建立的模型既不能对同一水体进行检验,更不能用于其它水体。初步试验表明:用ANN识别浮游植物群落与多种不同因子间的复杂关系并进行其动态预报是可行的。与传统的回归模型相比,ANN方法不要求监测数据具有很强的规律性,就可用训练后的网络模型对其进行预报,具有一定的实用性。

### 参 考 文 献

- 史为良,夏得昌,1994. 大伙房和柴河水库的环境因子同鲢、鳙生长和产量的关系. 海洋与湖沼, 25(1): 77—86
- 史为良,夏得昌,1995. 对用大伙房水库环境因子推定鲢、鳙生长和产量模型的检验. 大连水产学院学报, 10(1): 10—13
- 胡铁松,1997. 神经网络预测与优化. 大连:大连海事大学出版社, 48—56
- 黄长江,董巧香,1999. 1997年底中国东南沿海大规模赤潮原因生物的形态分类与生态学特征. 海洋与湖沼, 30(6): 197—204
- 黄长江,董巧香,2000a. 1998年春季珠江口海域大规模赤潮原因生物的形态分类和生物特征I. 海洋与湖沼, 31(2): 197—204
- 黄长江,董巧香,2000b. 1998年春季珠江口海域大规模赤潮原因生物的形态分类和生物特征II. 海洋与湖沼, 31(33): 233—238
- 雷宝玲,李开国,1995. 大伙房水库. 见:金相灿编. 中国湖泊环境(第三册). 北京:海洋出版社, 405—434
- Harris G P, 1983. Mixed layer physics and phytoplankton populations: studies on equilibrium and non-equilibrium ecology. Prog Phycol Res, 2: 1—52
- Karl-Erich L, Chorus I, 1998. The effect of water column mixing on phytoplankton succession, diversity and similarity. J Plankton Res, 20(4): 1927—1951
- Kanul C, Soyupak S, Germen E, 1998. A New Approach to Mathematical Water Quality Modeling in Reservoirs: Neural Networks. International Review of Hydrobiology, 83: 689—696, Special Issue 3rd International Conference on Reservoir Limnology and Water Quality. Ceske Budejovice, Czech Republic. August 11—15, 1997
- Recknagle F, 1998. Modeling and prediction of phyto- and zooplankton dynamics in Lake Kasumigaura by artificial neural networks. Lakes & Reservoirs: Research and Management, 3(2): 123—133
- Reynolds C S, 1997. Vegetative Procession in Pelagic: A model for Ecosystem Theory. Excellence in Ecology 9. Ecology Institute, Nordbunte 23, D-21385 Oldendorf/Luhe
- Sommer U, 1981. The role of r- and K-selection in the succession of phytoplankton in Lake Constance. Acta Oecologica, 2: 327—342

## MODELING AND PREDICTION OF PHYTOPLANKTON DYNAMICS IN RESERVOIRS BY ARTIFICIAL NEURAL NETWORK

WU Hong-Juan, GUO Sheng-Lian<sup>1</sup>, HU Chuan-Lin, LIU Yue<sup>2</sup>

(*Institute of Reservoir Fisheries, Ministry of Water Resources, The Chinese Academy of Sciences, Wuhan, 430079*)

<sup>1</sup>(*Wuhan University of Hydraulic & Electrical Engineering, Wuhan, 430072*)

<sup>2</sup>(*Liaoning Water Resources Bureau, Shenyang, 110003*)

**Abstract** Five series of physico-chemical data (1986—1996), including annual precipitation ( $p$ ), average water temperature from July to August ( $T$ ), the ratio of inflow and storage in July and August, the ratio of outflow and storage in July and August and phosphorus ( $PO_4$ ) from Dahuofang Reservoir was used to develop for predicting timing and magnitudes for phytoplankton and four series (1980—1989) of the ratio of inflow and outflow in July and August, average water temperature from July to August ( $T$ ), phosphorus ( $PO_4$ ) and total nitrogen (TN) from Dahuofang Reservoir was trained to develop for predicting timing for Cyanophyta dominant by artificial neural network model, respectively. These models were successful in estimating the output in two years (model 1 in 1997—1998 and model 2 in 1990—1991), with the average relative errors of 2% and 1% for calculated and observed data, respectively. The study indicates the potential of artificial neural network as predictive tool for highly non-linear phenomena, such as phytoplankton dynamics in reservoirs, better than classical statistical models.

**Key words** Phytoplankton dynamics, Artificial neural network, Predictive modeling

### 刊物动态

《海洋与湖沼》(中文版)荣获2000年度中国科学院优秀期刊一等奖,这是本刊有史以来获得的最高奖次。获得一等奖的有15种期刊,在水科学领域的科技期刊中,《海洋与湖沼》(中文版)是惟一获得此项荣誉的刊物。