

## 结合温度因子估算太湖叶绿素 a 含量的神经网络模型\*

孔维娟<sup>1,2</sup>, 马荣华<sup>1</sup>, 段洪涛<sup>1</sup>

(1: 中国科学院南京地理与湖泊研究所, 南京 210008)

(2: 南京大学地理信息科学系, 南京 210093)

**摘 要:** 神经网络方法估算复杂水体水质参数的优越性已经得到证实. 基于太湖水体实测叶绿素 a 浓度, 利用 MODIS 250m 影像和反演得到的水温数据建立了估算太湖水体叶绿素 a 含量的两个单隐层 BP 神经网络模型: NN1 模型不含温度因子、NN2 模型包含温度因子, 采用 Levenberg-Marquardt 算法训练网络, 利用初期终止方法提高网络泛化能力, 均取得了较高估算精度, 其中包含温度因子的反演模型精度稍有提高, 但不显著.

**关键词:** 叶绿素 a; BP 神经网络; MODIS; 水温; 太湖

## The neural network model for estimation of chlorophyll-a with water temperature in Lake Taihu

KONG Weijuan<sup>1,2</sup>, MA Ronghua<sup>1</sup> & DUAN Hongtao<sup>1</sup>

(1: *Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, P.R.China*)

(2: *Department of Geography Information Science, Nanjing University, Nanjing 210093, P.R.China*)

**Abstract:** The advantage of neural network method for estimating water quality parameters of complex water body has been approved. Using *in-situ* measurement data of chlorophyll-a concentration, imageries of MODIS 250m and retrieval model of water temperature, we develop two single-hidden-layer BP neural network models for estimating chlorophyll-a in Lake Taihu: Model NN1 without temperature input and Model NN2 with temperature input. The training method is used by Levenberg-Marquardt algorithm, and the early-stage determination in the modeling is used to improve generalization. The results show that: the estimation precision of the two models is high, in which the estimation precision of neural network input with temperature has been improved although the test is not significant.

**Keywords:** Chlorophyll-a; BP neural network; MODIS; water temperature; Lake Taihu

二类水体水色遥感算法的一个重要应用领域是内陆湖泊水体, 其中又以叶绿素遥感算法最具代表性. 水体水质参数的遥感反演方法主要包括经验方法, 半经验方法和分析方法. 目前, 半经验方法最为常用<sup>[1-4]</sup>, 它是在已知水色/水质参数基本光谱特征的基础上, 建立遥感数据与(准)同步的实测水色/水质参数值之间的线性或非线性统计关系模型, 进而达到水色/水质参数反演的目的. 然而, 相对于大洋 I 类水体而言, II 类水体光学特性要复杂的多, 叶绿素、悬浮物、黄色物质等组分的光谱特征相互影响<sup>[5]</sup>, 水质参数与遥感数据之间的关系很难用简单的统计回归来准确描述; 另外, 目前建立的各种统计回归反演模型大都没有考虑叶绿素 a 浓度的诸多环境影响要素, 即浮游植物生长的诸多环境因子, 如水温、pH 值、透明度、总氮、总磷等.

神经网络是一种良好的非线性拟合工具, 诸多研究表明<sup>[6]</sup>, 神经网络可以很好的模拟遥感信号与水质参数之间复杂的非线性关系, 基于人工神经网络的反演结果好于多项式拟合, 因此叶绿素 a 浓度的神经网络估算模型得到了进一步的发展和较为广泛的应用, 国内外许多研究者都建立了特定水域的神经网络

\* 国家自然科学基金(40871168、40671138、40801137)和国家科技支撑项目(2007BAC26B01)联合资助. 2008-11-03 收稿; 2008-12-03 收修稿稿. 孔维娟, 女, 1983 年生, 硕士研究生; E-mail: kongweijuan2006@126.com.

络估算模型, 并取得了较好的结果<sup>[7-10]</sup>, 但这些应用没有考虑影响叶绿素含量高低的其它环境要素, 如水温. 然而, 水温是藻类生物量的显著相关因子<sup>[11]</sup>, 太湖浮游藻类与水温关系密切<sup>[12]</sup>. 本文试图以太湖为实验区, 利用人工神经网络, 通过 MODIS 250m 遥感影像数据, 将温度因子和遥感反射率一起作为神经网络的输入, 建立遥感数据、水温因子与叶绿素浓度之间稳定的映射关系, 形成一个实用的叶绿素 a 浓度反演的神经网络模型, 以服务于太湖蓝藻水华的预测与预警.

## 1 数据与方法

### 1.1 影像数据

考虑到目前在轨卫星传感器的时间分辨率和空间分辨率, 选择使用 MODIS 250m 空间分辨率 2 个波段的卫星遥感影像. 首先利用 MODIS L1B 数据自带的经纬度信息进行几何校正, 然后在 ENVI4.3 软件下, 通过 FLAASH 模型进行大气校正, 已有研究表明利用 FLAASH 模型进行 MODIS 影像太湖区域大气校正可以取得良好的效果<sup>[13]</sup>, 校正时太阳高度角和方位角根据卫星影像成像时间计算得出, 海拔高程按实际约 5m 计算, 大气模式采用中纬度夏季模式, 其它参数均按实际输入, 之后转换为反射率值.

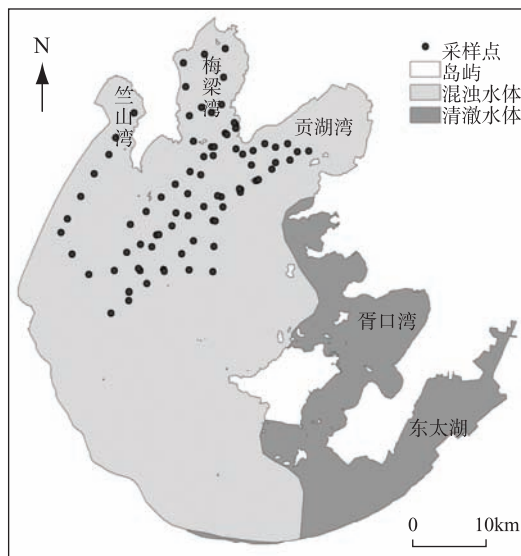


图 1 采样点位置

Fig.1 Locations of sampling points

式中,  $T_s$  是 MODIS 数据反演得到的地表温度(K);  $T_{31}$  和  $T_{32}$  分别是 MODIS 数据的第 31 和 32 波段亮温, 根据这两个波段的图像 DN 值来计算, 先把像元的 DN 值转换成相应的辐射强度值, 然后再用 Planck 函数求解星上亮温;  $A_0$ ,  $A_1$  和  $A_2$  是分裂窗算法的参数. MODIS 数据地表温度遥感反演方法及分裂窗算法基本参数估计详见文献[14]和[15]. 根据 MODIS 数据反演而得的湖泊水面温度是指水体表面的温度, 2007 年 11 月 3 日 11 时 15 分、2007 年 11 月 23 日 10 时 50 分和 2007 年 11 月 24 日 13 时 05 分的水温反演结果如图 2, 分别提取与样点位置对应的水温数值用作建模.

### 1.4 网络设计和训练

选取容错性好、应用最广泛的 BP(Back Propagation)神经网络. BP 网络可以从大量的样本中提取统计特性, 进而调整权值矩阵, 但个别样本中的误差不能左右对权值矩阵的调整<sup>[16]</sup>, 因此具有较好的容错性. 为了防止网络训练过度, 提高泛化能力, 使用初期终止方法<sup>[17]</sup>, 然后在样本数据中随机抽取 34 个作为模型的训练样本, 15 个作为确认样本, 剩余 15 个用于模型的测试.

借助 Matlab 软件, 建立了单隐层 BP 神经网络模型(图 3). 网络输入为 MODIS 250m 影像 Band1、Band2 及 MODIS 影像反演得到的水温因子, 输出为叶绿素 a 浓度. 隐含层和输出层的传递函数分别为双

### 1.2 实测数据

根据 MODIS 卫星过境时间(2007 年 11 月 3 日 11 时 15 分、2007 年 11 月 23 日 10 时 50 分和 2007 年 11 月 24 日 13 时 05 分), 分别于 2007 年 11 月 3 日、23 日和 24 日每天 10 时到 13 时左右在太湖北部湖区采集水样点共 80 个(图 1), 采样当天天气晴朗无云, 每个采样点均采集表层水样, 且利用 GPS 精确定位, 采样后立即返回实验室过滤, 然后用丙酮萃取, 避光 24h 后, 采用分光光度计法测定叶绿素 a 浓度. 剔除异常及与卫星过境时间相差大于 2h 的点, 剩余 64 个样点数据用于建模.

### 1.3 水温数据

MODIS 卫星遥感是通过探测地表面的热辐射强度来推算地表温度, 以 MODIS 两个相邻红外波段(第 31 和 32 波段)为基础进行星上亮温计算, 进而考虑大气和地表面的影响来反演地表温度, 对于 MODIS 数据的某个图像像元地表温度为:

$$T_s = A_0 + A_1 T_{31} - A_2 T_{32} \quad (1)$$

曲正切 S 型函数和线性函数, 神经网络性能函数选用均方误差  $MSE$ :

$$MSE = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (2)$$

其中,  $N$  为样本个数,  $t_i$  为目标输出,  $a_i$  为神经网络输出.

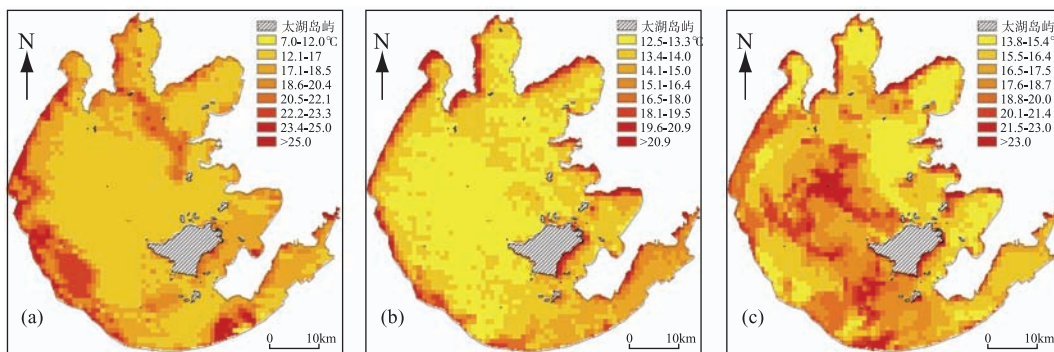


图 2 水温分布图: 2007 年 11 月 3 日 11 时 15 分(a); 2007 年 11 月 23 日 10 时 50 分(b); 2007 年 11 月 24 日 13 时 05 分(c)

Fig.2 Spatial distribution of temperature in Lake Taihu: (a) 11:15 AM on 3 Nov. 2007; (b) 10:50 AM on 23 Nov. 2007; and (c) 13:05 PM on 24 Nov. 2007

为了加快网络的收敛速度, 模型的输入和目标输出均归一化到 0 均值, 1 标准差, 网络训练采用批处理模式, 训练方法采用快速学习算法 Levenberg-Marquardt, 网络的初始权值和偏置取 (-0.5, 0.5) 之间的随机数, 隐含层神经元数目的确定没有统一明确的规定, 一般依靠反复尝试完成, 通过逐渐调整神经元个数, 对比测试集反演精度确定最佳隐含层神经元个数.

首先以 MODIS 250m 影像 Band1 和 Band2 为网络输入建立估算模型, 称为 NN1; 然后再加入温度因子作为输入建立估算模型, 称为 NN2. 反复尝试后, 确定了两个网络的隐含层神经元数分别为 5 和 6, 网络训练过程中的误差曲线分别如图 4 所示, 确认集误差开始增加, 训练停止, 测试集和确认集误差曲线基本一致, 说明样本集划分基本合

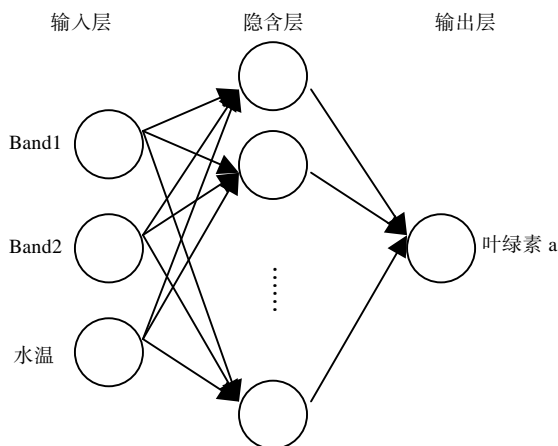


图 3 神经网络结构

Fig.3 The structure of neural network

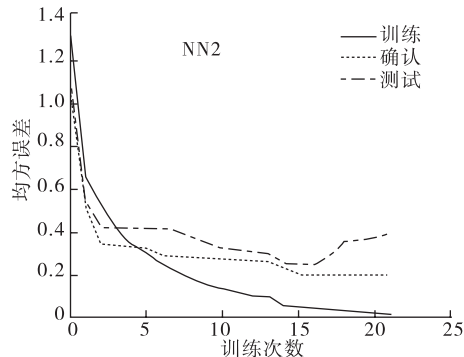
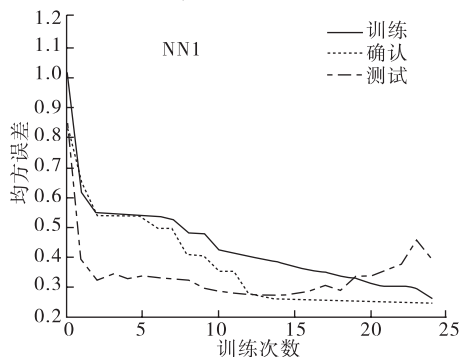


图 4 训练误差曲线

Fig.4 Errors curve in training process

理, NN1 训练误差略大于 NN2. 采用两个指标反映网络模型的估算值和实测值之间的差异: 线性 Pearson 相关系数的平方  $R^2$  和  $RMSE$ .

## 2 结果与分析

NN1 和 NN2 各集合估测值和实测值之间的差异如表 1 所示, 可以发现两个模型均取得了较高的反演精度, 各数据集  $R^2$  在 0.6882 和 0.9587 之间,  $RMSE$  最小值 7.8385, 最大值 21.5174, 对于太湖这种复杂的混浊水体两个模型均取得了较好的反演精度. 对比两个模型的反演精度可以发现, 在输入中加入温度因子的 NN2 模型相关系数  $R^2$  略微大于 NN1,  $RMSE$  略小于 NN1, 可见加入温度因子后反演精度有所提高, 但提高并不显著.

对比模型 NN2 的估测值与实测值(图 5)发现, 叶绿素 a 浓度变化范围为 1.1695–138.2556 $\mu\text{g/L}$ , 在小于 20 $\mu\text{g/L}$  的低浓度区估测值与实测值相差较小, 反演精度最好, 其次是大于 100 $\mu\text{g/L}$  的高浓度区, 中等浓度区反演精度不及低、高浓度区, 可能与中等浓度样本点较少有关, 在低浓度区出现估测值为负的情况, 这种异常情况在下文讨论.

表 1 模型 NN2 和 NN1 估算精度比较

Tab.1 The estimation precision comparison between NN1 and NN2

模型	隐含层神经元数	测试集		确认集		训练集	
		$R^2$	$RMSE$	$R^2$	$RMSE$	$R^2$	$RMSE$
NN1	5	0.7115	19.9024	0.7035	18.5251	0.6882	21.5174
NN2	6	0.7424	18.3435	0.7738	16.5834	0.9587	7.8385

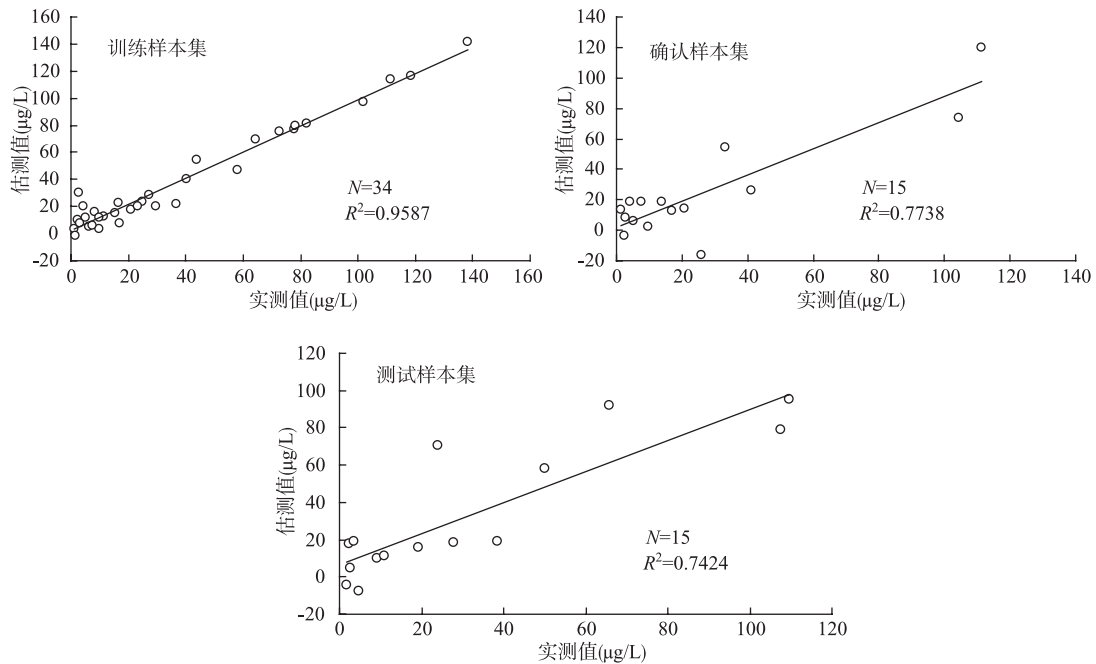


图 5 模型 NN2 叶绿素 a 浓度实测值与估测值对比

Fig.5 Relationship between *in situ* Chl.a concentration and neural network NN2 result

应用模型 NN2 对 2007 年 11 月 3 日的 MODIS 遥感影像进行叶绿素 a 浓度估算(图 6), 与原图对比发现, 估算结果基本符合蓝藻水华空间分布, 与图 2 相对比, 在蓝藻水华覆盖区和叶绿素 a 高浓度区水温高

于其他区域. 由于东部湖区水体清澈, 水质较好, 属于光学浅水区<sup>[18]</sup>, 底质对水体遥感反射比有着一定影响<sup>[19]</sup>, 而且到目前为止, 还没有发生蓝藻水华的记录, 因此这些区域水体本文暂不考虑. 图 5 和图 6 均表明叶绿素 a 浓度估算结果出现负值, 由于不符合常理, 我们将负值定义为异常值, 通过对比分析发现异常值有以下 2 种情况: (1) 出现在低浓度区(图 5、6), 在浓度的小值区域估测值可能低于实测值, 从而使得估测值为负, 就模型本身而言不算异常, 却不符合常理, 这只能寄希望于进一步提高模型精度来解决. (2) 出现在高浓度水华区(图 6), 蓝藻完全覆盖水面, 使得可见光不能进入水体, 光谱特征类似植被, 传感器获得的信号不包含任何水体信息, 因此不能用遥感数据准确估算水质参数, 所以叶绿素 a 浓度结果图在水华区要么异常要么不可信, 而且高浓度水华区叶绿素 a 浓度极高, 在这些区域探讨叶绿素 a 浓度已经没有意义.

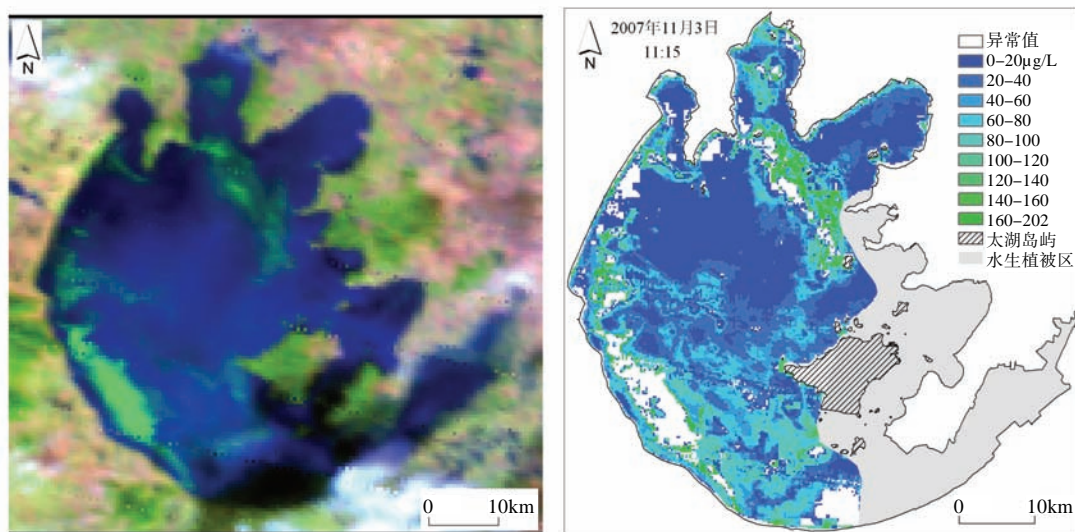


图 6 2007 年 11 月 3 日 11 时 15 分叶绿素 a 浓度估算结果

Fig.6 Chl.a concentration distribution map of Lake Taihu at 11:15 AM, on 3 Nov. 2007

### 3 结论与讨论

神经网络对太湖水体水质参数与遥感数据之间的复杂关系有较强的模拟能力. 利用 MODIS 250m 影像 Band1、Band2 与反演得到的温度数据建立了输入含温度因子的 NN2 模型和输入不含温度因子的 NN1 模型, 两者均用于估算太湖混浊水体区的叶绿素 a 含量. 结果表明: (1) NN1 和 NN2 均取得了较高的反演精度,  $R^2$  在 0.6882 和 0.9587 之间,  $RMSE$  最小值 7.8385, 最大值 21.5174; (2) 模型加入温度因子估算精度有所提高但没有明显提高; (3) 将模型 NN2 应用于 2007 年 11 月 3 日的 MODIS 遥感影像取得了较好的效果, 估算结果与蓝藻水华空间分布基本相符.

本研究建立的模型存在一定的不足之处: (1) 由 MODIS 数据反演得到水温数据本身带有一定误差, 在提高精度的同时也给模型增加了一个误差源; (2) 受数据限制本研究仅考虑了对藻类生物量影响较大的温度因子, 影响叶绿素 a 含量其他重要因子没有考虑, 这可能也是精度没有明显提高的原因之一; (3) 叶绿素 a 浓度变化范围大却仅有 64 个样点, 且分布不均匀, 低浓度点多, 中高浓度点少, 样本波动大, 这些都给建立高精度估算模型带了困难.

### 4 参考文献

- [1] 李素菊, 吴 倩, 王学军等. 巢湖浮游植物叶绿素含量与反射光谱特征的关系. 湖泊科学, 2002, 14(3): 228-223.
- [2] 马荣华, 戴锦芳. 应用实测光谱估测太湖梅梁湾附近水体叶绿素浓度. 遥感学报, 2005, 9(1): 78-86.



- [3] Luoheng H. Estimating chlorophyll-a concentration using first-derivative spectra in coastal water. *International Journal of Remote Sensing*, 2005, **26**(13): 5235-5244.
- [4] 段洪涛, 张 柏, 宋开山等. 长春市南湖叶绿素 a 浓度高光谱估测模型研究. *水科学进展*, 2006, **17**(3): 323-328.
- [5] 马荣华, 戴锦芳. 太湖水色遥感中叶绿素和悬浮物的高光谱分析. 见: 庄逢甘, 陈述彭主编. 中国遥感应用协会 2005 年年会论文集. 北京: 中国宇航出版社, 2005: 236-245.
- [6] Keiner LE, Yan XH. A Neural Network Model for Estimation Sea Surface Chlorophyll and Sediments from Thematic Mapper Imagery. *Remote Sens Environ*, 1998, **66**(2): 153-165.
- [7] Schiller H, Doerffer R. Neural network for emulation of an inverse model-operational derivation of Case II water properties from MERIS data. *International Journal of Remote Sensing*, 1999, **20**(9): 1735-1746.
- [8] Zhang Y, Koponen S, Pulliainen J *et al.* Application of empirical neural network to chlorophyll-a estimation in coastal water using optosensors. *IEEE Sensors Journal*, 2003, **3**(4): 376-382.
- [9] 张亭祿, 贺明霞. 基于人工神经网络的一类水域叶绿素 a 浓度反演方法. *遥感学报*, 2002, **6**(1): 40-44.
- [10] 唐军武, 丁 静, 田纪伟等. 黄东海二类水体三要素浓度反演的神经网络模型. *高技术通讯*, 2005, **15**(3): 83-88.
- [11] 陈宇炜, 秦伯强, 高锡云. 太湖梅梁湾藻类及相关环境因子逐步回归统计和蓝藻水华的初步预测. *湖泊科学*, 2001, **13**(1): 63-71.
- [12] 刘元波, 高锡云. 太湖北部梅梁湾水域水质藻类分析. *湖泊科学*, 1997, **9**(3): 255-260.
- [13] 李国砚, 张仲远, 郑艳芬等. MODIS 影像的大气校正及在太湖蓝藻监测中的应用. *湖泊科学*, 2008, **20**(2): 160-166.
- [14] 覃志豪, 高懋芳, 秦晓敏等. 农业旱灾监测中的地表温度遥感反演方法——以 MODIS 数据为例. *自然灾害学报*, 2005, **14**(4): 64-71.
- [15] 高懋芳, 覃志豪, 徐 斌. 用 MODIS 数据反演地表温度的基本参数估计方法. *干旱区研究*, 2007, **24**(1): 113-119.
- [16] 韩力群. 人工神经网络教程. 北京: 北京邮电大学出版社, 2006: 65-65.
- [17] Cuneyt K, Selcuk S, Ahmet FC *et al.* Case studies on the use of neural networks in eutrophication modeling. *Ecological Modelling*, 2000, **134**: 145-152.
- [18] MA RH, TANG JW, Dai JF. Bio-optical model with optimal parameter suitable for Lake Taihu in water colour remote sensing. *International Journal of Remote Sensing*, 2006, **27**(19): 4305-4328.
- [19] 马荣华, 段洪涛, 张寿选等. 太湖草型湖区底质对遥感反射比的贡献. *遥感学报*, 2008, **13**(3): 483-489.