

# 应用时间序列分析法对太湖叶绿素-*a* 含量的动态研究\*

龚绍琦<sup>1,2</sup> 黄家柱<sup>1</sup> 李云梅<sup>1</sup> 黎刚<sup>3</sup>

(1. 南京师范大学地理信息科学江苏省重点实验室 南京 210097;

2. 南京信息工程大学遥感学院 南京 210044; 3. 江苏省环境监测中心 南京 210036)

**提要** 叶绿素-*a* 是湖泊浮游植物现存量的重要指标, 其含量能反映水中浮游植物的丰度和变化规律。研究叶绿素含量的动态变化, 可以有效防止、监控湖泊富营养化的发生。以太湖为研究对象, 根据其营养水平进行分区, 使得不同区域内叶绿素的浓度不同, 而同一区域内浓度大体一致, 然后利用地理信息系统技术对每个区域内未采样点的叶绿素浓度进行插值, 取区域内所有栅格的平均值作为其最终浓度。应用时间序列分析方法对太湖 1998—2004 年每月的叶绿素含量进行动态模拟, 建立各湖区的预测模型。结果表明, 中营养区符合 MA(1)模型, 轻富营养区和中富营养区符合 AR(1)模型, 富营养区符合 MA(5)模型, 重富营养区符合 AR(6)模型。由此预测出 2005 年各湖区总叶绿素的含量, 经前 10 个月实测数据的验证说明, 所建模型在一定程度上能反映太湖各区域叶绿素浓度的变化。

**关键词** 太湖, 叶绿素-*a*, 地理信息系统, 时间序列, 模型

**中图分类号** Q958.1

水体叶绿素-*a* 是各种浮游植物最重要的色素之一, 其所占比例都比较稳定。叶绿素-*a* 含量的高低能反映水体的营养状况, 是水质调查的一个重要指标。很多学者试图通过叶绿素的研究来了解水体生态系统状况和富营养化水平, 因此叶绿素的研究最为广泛。Nikolay 等(2003)人利用 SeaWiFS 卫星资料研究了叶绿素与环境因子之间随时间变化的动态规律; 张钊(2001)分析了台湾海峡南北区域叶绿素的季节变化特征; 赵辉等(2005)用 SeaWiFS 卫星资料分析了南海不同季节叶绿素的变化及典型区域的空间分布状况; 陈晓玲等(2002)研究了香港海域叶绿素浓度的时空分布规律。纵观这些研究结果, 叶绿素含量的变化研究仅停留在定性分析上, 而且考虑的时间很短, 因此就很难对水体叶绿素的变化规律做出客观全面的定论和描述。本文中作者以我国第三大淡水湖——太湖为例, 对过去连续较长一段时间内叶绿素含量变化进行模拟。每年太湖叶绿素浓度变动都在 0.002—0.442mg/L 之间, 最高值一般出现在 7、8、9 月份, 而

最小值在 11、12、1 月份。从 1998—2004 年共 7 年的变化来看, 叶绿素浓度的年度变化不是很大, 1999 年略有下降后, 2000—2004 年并没有大幅度的上升或下降, 2000 年夏季太湖蓝藻爆发, 使得 7 月份的叶绿素浓度有明显偏高。但从每月各监测样点数值来看, 叶绿素浓度呈现无规律性变化, 数据具有较大的不确定性。因此很难用传统的多元回归方法对叶绿素变化进行模拟。美国统计学家 G. E. P. Box 和英国统计学家 G. M. JenKins 在 20 世纪 70 年代提出了处理不确定时间序列的分析模型 ARMA(p, q), 为随时间而不确定变化的数据进行定量分析提供了很好的方法。针对太湖叶绿素的变化特点, 把过去连续较长一段时间内的监测值看作是一个时间序列, 利用时间序列的分析方法, 建立能反映太湖叶绿素动态变化的数学模拟模型, 并预测未来一个时期湖内叶绿素的含量, 为防止、控制太湖富营养化发生以及初级生产力的估算提供科学依据。

\* 国家自然科学基金项目, 40571110 和 40801145 号; 江苏省高等学校研究生创新计划资助项目, 06-30 号。龚绍琦, 博士, 讲师, E-mail: gongshaoqi@hotmail.com

收稿日期: 2007-12-15, 收修改稿日期: 2008-04-15

## 1 研究区概况

太湖位于长江三角洲南缘,是我国长江中下游地区著名的五大淡水湖之一,界于北纬 30°55′40″—31°32′58″,东经 119°53′32″—120°36′10″,地跨江浙两省,湖泊面积为 2338.1km<sup>2</sup>,平均深度为 2m,湖中各种浮游植物、动物和栖息类动物种类繁多,具有良好的水生态系统。太湖不仅是国内外享有盛名的旅游胜地,而且是太湖周边地区的主要水源地,兼有航运、灌溉、养殖等多种功能,是我国经济最发达的区域之一。每年夏季太湖都有大量藻类暴发,不仅破坏湖内的生态景观,还给周边地区生产、生活带来极大不便。

## 2 数据的收集

收集了太湖 1998 年 1 月—2005 年 10 月期间每月叶绿素的监测数据。全湖共布置 20 个监测点,其点位坐标输入 GPS 仪中,监测船只每次用 GPS 仪导航,到点后采集水样,并带回实验室分析。叶绿素 *a* 的测定是将滤过水样的滤膜,置于低温冰箱中干燥保存 6—8h,用 3 ml 90%丙酮提取,然后经过研磨、离心分离(沉淀返回作再次提取),取上清液用丙酮溶液定容。将提取液在分光光度计 750 nm、663 nm、645 nm 和 630 nm 处测定吸光值,按公式计算叶绿素 *a* 含量(金相灿等,1990)。全部的监测数据由江苏省环境监测中心站提供。

## 3 研究方法

太湖水体中叶绿素的空间分布是不均一的(刘瑞民等,2001),不同区域内其含量不一。因此根据任黎等(2004)、龚绍琦等(2005)对太湖富营养化综合评价的结果,把全湖分成 5 个营养程度不同的区域:中营养区、轻富营养区、中富营养区、富营养区和重富营养区,使得不同区域叶绿素的浓度不同,而同一区域浓度大体一致,全湖分区及监测点分布情况见图 1。

考虑到各采样点的叶绿素在空间位置上具有一定的相关性,可以利用地理信息系统中的插值方法对每个区域内未设监测点位置的叶绿素浓度进行插值。而逆距离加权(IDW)插值法是地理信息系统中最简单、最常用的、插值结果较为精确的一种确定性方法。该方法随着采样点与预测点之间距离的增大,标准样点对预测点的影响权重呈指数递减。可通过对权

重指数的优化,使得预测的均方根(Root-Mean-Square, RMS)误差最小。本研究中样点分布较少,插值的邻近点个数设定为 3,插值后的均方根误差都在 0.005mg/L 以下。考察整个湖泊叶绿素浓度的空间分布结果,并对每个湖区内所包含栅格进行统计,取所有栅格浓度的平均值作为该区叶绿素浓度的最终值。应用时间序列的分析方法分别对每个湖区 1998 年 1 月—2004 年 12 月期间叶绿素含量的变化进行动态模拟,建立其预测模型。

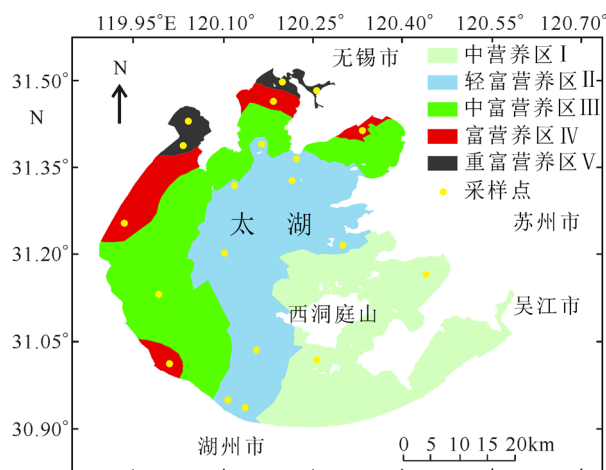


图 1 太湖营养分区及监测点分布图

Fig.1 The site of monitoring and trophic divisions in Taihu Lake

## 4 时间序列模型 ARMA(*p*, *q*)概述

时间序列是按时间顺序排列的一组数据或一有序随机数据。时间序列分析方法是假定能用过去发生方式来重现未来,即采用参数模型对所观测的有序随机数据进行分析与处理的一种方法。而 ARMA(*p*, *q*)模型是统计学家 G.E.P.Box 和 G.M.JenKins 在 20 世纪 70 年代提出的时间序列分析模型,即自回归滑动平均模型(Autoregressive Moving Average Model),也称博克斯-詹金斯(B-J)法,ARMA(*p*, *q*)模型可表示为:

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_p X_{t-p} + \varepsilon_t - \psi_1 \varepsilon_{t-1} - \psi_2 \varepsilon_{t-2} - \dots - \psi_q \varepsilon_{t-q}$$

当式中  $q=0$  称为  $p$  阶自回归模型(Autoregressive Model),记为 AR( $p$ )模型,  $\theta_k(k=1, 2, \dots, p)$  是自回归系数,  $p$  为模型的阶次,即滞后的时间周期;当式中  $p=0$  称为  $q$  阶滑动平均模型(Moving Average Model),记为 MA( $q$ )模型,  $\psi_k(k=1, 2, \dots, q)$  是滑动平均系数,  $q$  为模型的阶次,  $\varepsilon_k$  为各阶的残差(刘洪斌等,2004;徐静,2001;Paulo *et al*,2004)。

运用 ARMA( $p, q$ )模型的前提条件是建立模型的时间序列为零均值的平稳随机过程。由于湖泊受到水动力学过程的作用,不同湖区内营养盐的迁移和扩散不同,同时受湖泊生态系统中物理、化学和生物等环境因子的影响,浮游植物生物量的变化非常复杂,因此不同时期叶绿素含量具有较大的随机性,可以应用 ARMA( $p, q$ )模型对太湖叶绿素含量进行模拟。

### 5 湖泊叶绿素时间序列 ARMA( $p, q$ )模型的建立

#### 5.1 数据的预处理

由于数据在采样和分析过程中难免会带入误差,使得数据集中存在异常值。在对叶绿素浓度进行时间序列分析以前先考察每个观测值,剔除数据集中出现的异常极大或极小值,具体方法是通过对历年同期资料,并结合太湖藻类实际生长情况,估计一个真值代替。

建立 ARMA( $p, q$ )模型要求时间序列为平稳随机的零均值序列,否则会降低模型的预测精度,对于不满足条件的序列可通过一系列的转化来实现。

首先,序列的周期性和趋势性判断。作各湖区叶绿素浓度的时间序列图(图 2),从该图上可以观察到,营养区、轻富营养区和 中富营养区 III 叶绿素浓度值不存在周期性或趋势性,有很好的随机性。富营养区和重富营养区 则具有一定的周期性,通过 2 阶差分后序列变得平稳。总体来看,各区叶绿素含量在夏季 7、8 或 9 月达到最高,而在冬季 12 或 1 月出现最低。这与其它湖泊的研究结果是一致的(王飞儿等, 2004; 吴洁等, 2001)。

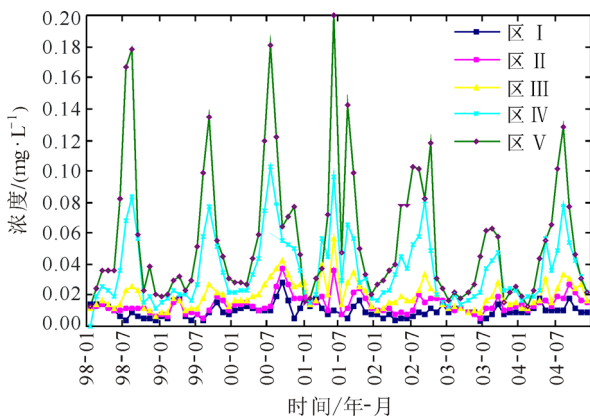


图 2 太湖各湖区叶绿素含量时间序列图

Fig.2 The temporal change in chlorophyll-a content in every trophic division of Taihu Lake

其次,序列的随机性判断。分别对各区叶绿素浓度序列作正态分布检验,结果表明,营养区和轻富营养区 很好地符合正态分布,而中富营养区、富营养区和重富营养区 并不成正态分布,将数据进行自然对数转换后,序列符合正态分布,此时序列具有随机性。

再次,序列的平稳性判断。分别计算各湖区叶绿素浓度序列前 16 阶的自相关系数和偏自相关系数(表 1),并作出自相关函数图和偏自相关函数图(图略)。从表中可以得知,各区序列数据经预处理后其自相关系数和偏自相关系数都逐渐衰减并趋近于 0,说明序列具有很好的平稳性。

最后,序列的零均值化。将上述变换后的序列取平均,把每期数据分别减去平均值,得到新序列。

#### 5.2 模型的识别

考察平稳序列各阶的自相关系数和偏自相关系数,当偏自相关系数  $r_k$ ( $k$  为滞后的阶次)为  $p$  步截尾,即当  $k > p$  时  $r_k = 0$ ,而自相关系数  $\rho_k$  拖尾,即  $\rho_k$  按负指数衰减时,则可识别序列为 AR( $p$ )模型。从数理统计上可以进一步检验  $r_k$  是否为  $p$  步截尾:若  $k > p$  时  $|r_k| > \frac{2}{\sqrt{n}}$  的个数所占百分比不超过 4.5%。

当序列的自相关系数  $\rho_k$  在  $q$  步截尾,而偏相关系数  $r_k$  拖尾时,则可识别序列为 MA( $q$ )模型。进一步从数理统计上检验序列  $\rho_k$  是否为  $q$  步截尾,即当

$$|\rho_k| > \frac{2}{\sqrt{n}} \sqrt{1 + 2 \sum_{i=1}^q r_i^2}$$

就认为  $\rho_k$  为  $q$  步截尾(赵玮等, 2003; 张崇甫等, 1995; 韦玉春等, 2005)。当自相关系数  $\rho_k$  和偏自相关系数  $r_k$  均拖尾时,则可识别序列为 ARMA( $p, q$ )模型。此时模型中  $p, q$  的阶数未定,通常用 ARMA(1, 1)、ARMA(1, 2)、ARMA(1, 3)、ARMA(2, 1)ARMA(2, 2)...等模型去试算,一般来说当阶数  $p+q \leq 5$  时就可以拟合绝大多数时间序列(杨振成, 2004)。ARMA( $p, q$ )模型阶数的确定可根据日本学者赤池(Akaike)提出的最小最终预报误差 FPE 准则、统计模型选择 AIC 准则以及改进的 BIC 准则进行判断,其中应用最广的是 AIC 准则。当用以上模型依次去拟合时,同时计算 FPE( $p, q$ )、AIC( $p, q$ )或 BIC( $p, q$ )的值,取 FPE、AIC 或 BIC 值最小的作为最优模拟模型(赵玮等, 2003)。

FPE、AIC 和 BIC 准则的定义为:

$$FPE(p, q) = \frac{N + p + q + 1}{N - p - q - 1} \hat{\sigma}_a^2(k)$$

表 1 太湖各湖区叶绿素时间序列的各阶自相关系数和偏自相关系数  
Tab.1 The auto- and partial-correlation coefficients of chlorophyll-a content in every trophic division of the Taihu Lake

湖区	滞后	1	2	3	4	5	6	7	8
中营养区	$\rho_k$	<b>0.462</b>	0.067	0.066	0.077	0.135	0.212	0.132	-0.119
	$r_k$	0.462	-0.186	0.152	-0.018	0.142	0.116	-0.024	-0.217
轻富营养区	$\rho_k$	0.431	0.255	0.154	0.119	0.100	0.070	0.044	0.047
	$r_k$	<b>0.431</b>	0.084	0.021	0.037	0.030	0.002	-0.004	0.022
中富营养区	$\rho_k$	0.603	0.440	0.194	0.087	0.111	0.043	0.011	0.025
	$r_k$	<b>0.603</b>	0.121	-0.180	-0.013	0.170	-0.090	-0.077	0.111
富营养区	$\rho_k$	<b>-0.540</b>	<b>0.164</b>	<b>-0.090</b>	<b>-0.183</b>	<b>0.244</b>	-0.07	-0.081	0.059
	$r_k$	-0.540	-0.176	-0.110	-0.380	-0.100	0.048	-0.199	-0.195
重富营养区	$\rho_k$	0.679	0.340	-0.031	-0.231	-0.445	-0.560	-0.471	-0.237
	$r_k$	<b>0.679</b>	<b>-0.225</b>	<b>-0.193</b>	<b>-0.208</b>	<b>-0.252</b>	<b>-0.207</b>	0.0347	0.109

湖区	滞后	9	10	11	12	13	14	15	16
中营养区	$\rho_k$	-0.118	-0.107	0.009	0.083	0.023	-0.072	-0.227	-0.175
	$r_k$	0.048	-0.163	0.166	-0.043	0.032	-0.054	-0.179	0.006
轻富营养区	$\rho_k$	0.139	0.084	0.079	0.094	0.037	-0.007	-0.190	-0.193
	$r_k$	0.129	-0.029	0.020	0.049	-0.046	-0.049	-0.230	-0.060
中富营养区	$\rho_k$	0.049	0.146	0.235	0.271	0.167	0.053	-0.074	-0.191
	$r_k$	0.068	0.091	0.147	0.068	-0.172	-0.089	-0.059	-0.185
富营养区	$\rho_k$	-0.079	0.036	0.023	0.000	0.077	-0.054	0.086	-0.258
	$r_k$	-0.108	-0.190	-0.250	-0.143	-0.005	-0.069	0.105	-0.210
重富营养区	$\rho_k$	0.053	0.262	0.445	0.465	0.433	0.258	-0.051	-0.287
	$r_k$	0.145	0.014	0.155	-0.058	0.161	-0.021	0.183	-0.021

注：时间序列模型的识别是根据自相关和偏相关系数进行判断的，表中加粗部分是作者判断的结果

$$AIC(p, q) = N \ln \hat{\sigma}_a^2(k) + 2(p+q+1) / N$$

$$BIC(p, q) = N \ln \hat{\sigma}_a^2(k) + (p+q+1) \ln N$$

其中  $N$  为样本数， $\hat{\sigma}_a^2(k)$  为残差的方差，本文参照 AIC 准则选择最优模型，各区模型的 AIC 统计量见表 2。

根据上述原则，从表 1 的结果可以看出，中营养区 I 符合 MA(1)模型、轻富营养区 和中富营养区 符合 AR(1)模型、富营养区 符合 MA(5)模型和重富营养区 符合 AR(6)模型。

### 5.3 模型的参数估计

通过上节模型的识别，确定了各区域叶绿素浓度时间序列模型的结构和阶数，在此基础上可以进行模型参数的估计。模型参数估计的方法很多，常见的有参数的矩估计、最小二乘法估计和极大似然估计等方法。本文中作者借助 SPSS 统计软件，选择极大似然估计法来求解模型参数，得到各区叶绿素浓度的时间序列模型如下：

中营养区：  $MA(1) = \varepsilon_t - 0.5853\varepsilon_{t-1}$

轻富营养区：  $AR(1) = 0.4275X_{t-1}$

中富营养区：  $AR(1) = 0.6041X_{t-1}$

富营养区：  $MA(5) = \varepsilon_t - 0.7584\varepsilon_{t-1} + 0.1028\varepsilon_{t-2} -$

$$0.5506\varepsilon_{t-3} - 0.0429\varepsilon_{t-4} + 0.3397\varepsilon_{t-5}$$

重富营养区：  $AR(6) = 0.6664X_{t-1} - 0.0998$

$$X_{t-2} - 0.0672X_{t-3} - 0.0240X_{t-4} - 0.0952X_{t-5} - 0.2443X_{t-6}$$

### 5.4 模型的假设检验

估计出时间序列模型参数后还需要进一步检验模型是否合适，通常的做法就是检验其残差  $\varepsilon_t$  是否为白噪声。即利用所建模型对历史数据进行递推预测，计算各期的残差及其各阶自相关系数  $\rho_k$ ，并验证博克斯-皮尔斯统计量(Box-Pierce-Q-Statistic)

$$Q = N \sum_{k=1}^m \rho_k^2$$

是否服从  $\chi^2_{0.05(m-p-q)}$  分布， $m$  一般取  $\frac{N}{4}$  或  $\sqrt{N}$  附近

的值，在这里  $m$  取  $\frac{N}{4} = 21$ ，各区模型参数及统计量见表 2。从表 2 可知，各湖区叶绿素浓度时间序列模型的统计量  $Q$  值都小于  $\chi^2_{0.05}$  分布表的临界值，因此认为在显著水平 0.05 下模型残差为白噪声，所建模型可以用于短期内该湖区叶绿素浓度的预测。图 3 是根据各区所建的叶绿素浓度时间序列模型对历史数据进行模拟的结果，说明所建立的模型能很好地描述该湖区内叶绿素含量的动态变化。

表 2 太湖各湖区建立叶绿素时间序列模型的参数  
Tab.2 The parameters of chlorophyll-a time series model in every trophic region Taihu lake

区域	数据变换	适用模型	AIC	统计量 Q	$X^2_{0.05}$ 临界值
中营养区	/	MA(1)	-695.09	21.045	31.410
轻富营养区	/	AR(1)	-637.65	10.687	31.410
中富营养区	自然对数转换	AR(1)	60.45	19.433	31.410
富营养区	自然对数转换、2 阶差分	MA(5)	101.20	21.684	26.296
重富营养区	自然对数转换、2 阶差分	AR(6)	109.98	15.102	24.996

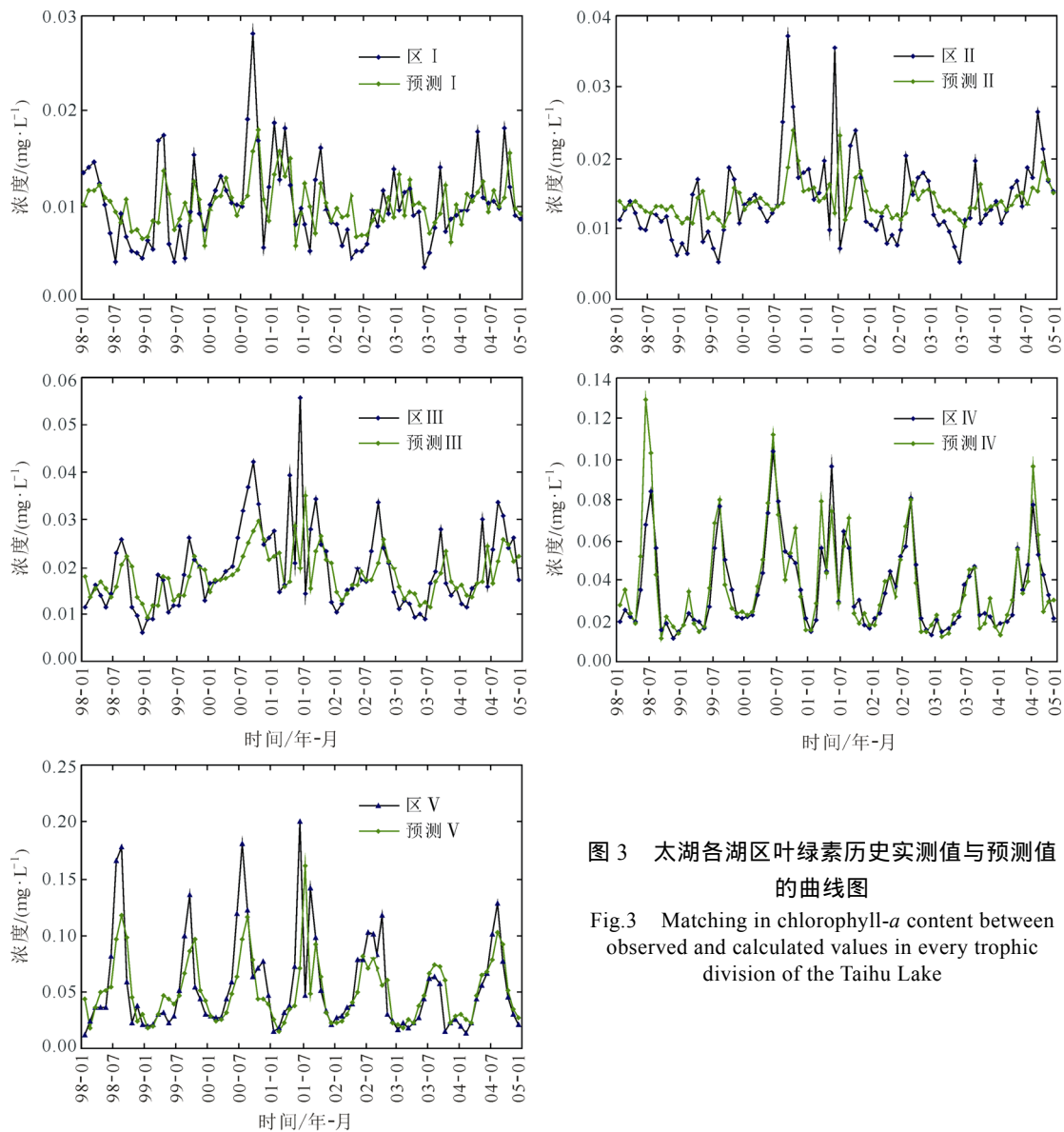


图 3 太湖各湖区叶绿素历史实测值与预测值的曲线图

Fig.3 Matching in chlorophyll-a content between observed and calculated values in every trophic division of the Taihu Lake

### 5.5 模型的预测及其精度评价

利用各区域所建的模型预测 2005 年全年叶绿素的含量, 并与前 10 个月的实测值相比较, 计算预测的相对误差。由表 3 可看出, 模型对 1、2、3 月份预测的相对误差较大, 可能是冬季温度较低、光照减弱,

不利于藻类的生长, 叶绿素含量变化的差异性较大, 模型预测困难。除了重富营养区的模型预测误差较大, 其它各区模型的预测精度大部分都在 15% 以下, 说明所建模型在一定程度上能反映太湖各区域叶绿素浓度的变化。中营养区 I、轻富营养区 和中富营

养区 建立的时间序列模型中只有一个变量,对历史数据的考察较少,前一个数据的变化对后期预测结果有很大的影响,其预测的长度和稳健性会稍差些。由表 3 可以看到,这三个模型预测的时期越长,其预测值越接近。而其它两区建立的模型中变量较多,综合了多个历史数据的影响,后期预测的结果仍然很好。Clai 等(1998)在用时间序列模型来预测大气污染物时通过引入了 Theil 指数来判断模型所预测的长度,由于 Theil 指数是一个从信息和熵的角度来考察样本差异性的指标,广泛应用于经济学领域,其数学机理不清楚以及准确性无法验证,故本文不作借鉴。

### 5.6 模型的分析

通过对不同湖区叶绿素浓度序列的分析,建立了各湖区叶绿素变化的模拟模型。模拟和预测的结果

表明,该方法能够以一定的精度来模拟和预测叶绿素含量动态变化过程,对各区域叶绿素的趋势预测有一定的适用性。随着新的实测数据的不断补充,各模型的验证及预测精度将会得到进一步提高。本研究是依据各区域叶绿素含量自身的历史变动过程而建立模型的,没有引入直接影响叶绿素变动的环境因子或生物因子,而是将这些因子的影响作为模型中的随机干扰或误差来对待。中营养区和富营养区属于移动平均模型  $MA(q)$ ,它是用过去各时期的随机干扰或预测误差的线性组合来表达当前预测值,说明这两湖区叶绿素含量变化主要受外界随机因素的影响,藻类的数量分布有更大的不确定性。中营养区为 1 阶模型,只与前一时期测量值有关,这主要是有吴淞江和太浦河流经该区,成了太湖主要的出湖河流。随着湖水的流动带走部分营养物质,不利于藻类的生

表 3 太湖各湖区叶绿素浓度(mg/L)预测结果及误差(%)分析  
Tab.3 Results and the bias (%) in calculation of chlorophyll-a content (mg/L) in every trophic division of the Taihu Lake

时期 (年-月)	中营养区			轻富营养区			中富营养区		
	实测值	预测值	相对误差	实测值	预测值	相对误差	实测值	预测值	相对误差
2005-1	0.0082	0.00898	9.50	0.0082	0.01454	77.01	0.0080	0.01731	116.43
2005-2	0.0074	0.01005	35.85	0.0102	0.01417	38.91	0.0110	0.01744	58.58
2005-3	0.0063	0.01005	59.57	0.0091	0.01401	84.45	0.0107	0.01752	63.77
2005-4	0.0136	0.01005	26.08	0.0192	0.01395	27.36	0.0198	0.01757	11.26
2005-5	0.0083	0.01005	21.12	0.0151	0.01392	7.83	0.0161	0.01760	9.32
2005-6	0.0125	0.01005	19.58	0.0173	0.01391	19.62	0.0202	0.01762	12.78
2005-7	0.0092	0.01005	9.27	0.0123	0.01390	13.01	0.0225	0.01763	21.65
2005-8	0.0109	0.01005	7.77	0.0146	0.01390	4.81	0.0185	0.01764	4.68
2005-9	0.0122	0.01005	17.60	0.0148	0.01390	6.10	0.0177	0.01764	0.35
2005-10	0.0116	0.01005	13.34	0.0198	0.01390	29.81	0.0212	0.01764	16.79
2005-11		0.01005			0.01390			0.01764	
2005-12		0.01005			0.01390			0.01764	

时期 (年-月)	富营养区			重富营养区		
	实测值	预测值	相对误差	实测值	预测值	相对误差
2005-1	0.0101	0.03013	198.32	0.0137	0.01993	45.47
2005-2	0.0152	0.02929	92.70	0.0179	0.02054	14.75
2005-3	0.0168	0.02932	74.52	0.0250	0.02601	4.04
2005-4	0.0277	0.03254	17.47	0.0411	0.03625	11.80
2005-5	0.0323	0.03236	0.19	0.0396	0.05051	27.55
2005-6	0.0373	0.03215	13.81	0.0464	0.06580	41.81
2005-7	0.0323	0.03193	1.15	0.0492	0.07464	51.71
2005-8	0.0386	0.03168	17.93	0.0572	0.07447	30.19
2005-9	0.0341	0.03141	7.89	0.0459	0.06544	42.57
2005-10	0.0314	0.03112	0.89	0.0457	0.05286	15.67
2005-11		0.03081			0.04165	
2005-12		0.03048			0.03392	

长,同时水流也加速了藻类的迁移和向外运送,使得叶绿素浓度变化较大。富营养区为 5 阶模型,即预测值与过去 5 个时期测量值相关,该区有直湖港、武进港、官渎河和阊江口,这些都是严重的污染源,其各种污染物输入太湖后一部分在底泥中积累起来,不断地为藻类的生长提供营养物质,导致某时刻叶绿素含量对前 5 个时期产生的随机干扰都有依赖性。其它三区均属于自回归 AR( $p$ )模型,它是直接通过变量自身历史观测值来反映对当前预测值的影响和作用。轻富营养区和中富营养区为 1 阶模型,也只与前一时期测量值有关,这是由于这两湖区依次位于太湖的东西沿岸和湖心区,受水动力学的作用,浮于湖表层的藻类流动迅速,使得不同时期叶绿素浓度变化较快。重富营养区为 6 阶模型,即预测值与过去 6 个时期测量值有很好的相关性,该区有梁溪河、漕桥河和太湖运河流入,都是严重的纳污河流,其各种污染物、营养盐类经河道水体的输送流入太湖,同时北部的五里湖与外界水流交换缓慢,当地盛行季风不利于污染物的扩散,使得该湖区营养程度高,有利于各种藻类的不断生长,各个时期叶绿素浓度的延续性大。

## 6 结论

(1) 按照太湖富营养化程度大小对全湖进行分区,利用地理信息系统技术对每个时期全湖未采样点位置上叶绿素浓度进行了插值,将每个湖区内所有栅格的浓度值取平均,以此作为该湖区叶绿素的浓度值。该方法充分利用了各监测点的空间位置信息,同时考虑了太湖叶绿素的实际空间分布状况,具有更好的准确性。

(2) 太湖水体中叶绿素的空间分布是不均一的,不同时期湖内含量也不一样,具有较大的随机动态性,因此很难用传统的多元回归分析方法来模拟。本文中作者应用时间序列分析方法模拟了全湖各区叶绿素含量的变化,建立了各营养区的时间序列模型,并预测出各区 2005 年叶绿素的含量。经前 10 个月实测数据的验证,说明所建模型在一定程度上能反映太湖各区域叶绿素浓度的变化。

(3) 按照太湖富营养化程度大小对全湖进行分区,建立了各营养区叶绿素浓度的时间序列模型,在今后的湖泊环境保护过程中,则可根据历史的监测

资料来预测未来一段时间内叶绿素含量的变化情况,为湖泊藻类水华爆发的预测预报提供理论依据。

## 参 考 文 献

- 王飞儿,吕映春,陈英旭等,2004. 千岛湖叶绿素  $a$  浓度动态变化及其影响因素分析. 浙江大学学报(农业与生命科学版), 30(1): 22—26
- 韦玉春,陈锁忠,2005. 地理建模原理与方法. 北京:科学出版社, 210—240
- 任黎,董增川,李少华,2004. 人工神经网络模型在太湖富营养化评价中的应用. 河海大学学报(自然科学版), 32(2): 147—150
- 刘洪斌,武伟,魏朝富等,2004. AR 模型在土壤水分动态模拟中的应用. 山地学报, 22(1): 121—125
- 刘瑞民,王学军,王翠红等,2001. 应用地统计学方法研究湖泊中叶绿素  $a$  的空间分布. 农业环境保护, 20(5): 308—310
- 杨振成,2004. 自回归滑动平均模型中阶数及参数的确定. 统计与决策, 12: 21—22
- 吴洁,钱天鸣,虞左明,2001. 西湖叶绿素- $a$  周年动态变化及藻类增长潜力试验. 湖泊科学, 13(2): 143—148
- 张钊,2001. 台湾海峡叶绿素  $a$  含量的季节变化. 台湾海峡, 20(3): 314—318
- 张崇甫,陈述云,胡希玲等,1995. 统计分析方法与应用. 重庆:重庆大学出版社, 114—124
- 陈晓玲,李毓湘,李志林,2002. 香港海域叶绿素- $a$  浓度的时空分布规律. 地理学报, 57(4): 422—428
- 金相灿,屠清瑛,1990. 湖泊富营养化调查规范(第二版). 北京:中国环境科学出版社, 270—288
- 赵辉,齐义泉,王东晓等,2005. 南海叶绿素浓度季节变化及空间分布规律. 海洋学报, 27(4): 46—52
- 徐静,2001. ARMA 模型及其应用. 立信会计高等专科学校学报, 15(3): 21—24
- 龚绍琦,黄家柱,李云梅等,2005. 基于 GIS 下的太湖水质富营养化模糊综合评价. 环境科学, 26(5): 34—37
- Clai G, Kerschbaumer A, Tosi E *et al*, 1998. Analysis of urban atmospheric pollution data in the bologna area. Environment Monitoring and Assessment, 52: 149—157
- Nikolay P N, Li Bailian,2003. Time-series analysis of remote-sensed chlorophyll and environmental factors in the Santa Monica-San Pedro Basin off Southern California. Journal of Marine Systems, 39: 185—202
- Paulo Cortez, Miguel Rocha, Jose Neves,2004. Evolving time series forecasting ARMA models. Journal of Heuristics, 10: 415—429

## THE TIME SERIES OF CHLOROPHYLL-*a* CONTENT IN TAIHU LAKE

GONG Shao-Qi<sup>1,2</sup>, HUANG Jia-Zhu<sup>1</sup>, LI Yun-Mei<sup>1</sup>, LI Gang<sup>3</sup>

(1. Jiangsu Provincial Key Laboratory of Geographic Information Sciences, Nanjing Normal University, Nanjing 210097;

2. School of Remote Sensing, Nanjing University of Information Science & technology, Nanjing, 210044;

3. Jiangsu Province Environmental Monitoring Centre, Nanjing, 210036)

**Abstract** Chlorophyll-*a* concentration is an important index to phytoplankton biomass in lake, showing the abundance and distribution of phytoplankton in lake water, and reflecting the change in lacustrine chlorophyll-*a* content, which may help indicate the state of eutrophication. Lake water was sampled with chlorophyll-*a* concentration measured or interpolated lake-wise in grid pattern using GIS (geographic information system) technique. In terms of chlorophyll-*a* concentration as well as nutrition level, Taihu Lake was divided into five divisions from mesotrophic to hypertrophic state. The historical data of observational chlorophyll-*a* content were simulated in month from 1998 to 2004 in time series. Processes including data pretreatment, parameterization and presumption test were carried out for simulating raw data for model establishment and validation. For each division, different models are built up capable of simulating chlorophyll-*a* content in 2005. The results show that these models are able to well-describe the temporal and spatial changes in trophism of the lake.

**Key words** Taihu lake, Chlorophyll-*a*, GIS, Eutrophication, Model