

# 基于遗传算法的洞庭湖富营养化指标预测模型

李广源

(湖南省水文水资源勘测局, 湖南 长沙 410007)

**摘要** 在已获取调研资料和实测数据的基础上, 分析了洞庭湖富营养化评价指标及其影响因子之间的统计特性, 据此建立了描述洞庭湖富营养化评价指标及其影响因子之间变化关系的、基于最小一乘法的多元线性回归预测模型, 并给出了统一用加速遗传算法进行洞庭湖富营养化评价指标预测模型的参数优化估计, 预测了 2010 年洞庭湖富营养化评价指标值。

**关键词** 富营养化; 预测模型; 加速遗传算法; 洞庭湖

**中图分类号** X824      **文献标识码** :A      **文章编号** :1004-693X(2004)05-0023-03

为了深入评价洞庭湖水体富营养化状态, 首先必须对洞庭湖富营养化评价指标的变化趋势进行预测。本文着重探讨洞庭湖富营养化评价指标的预测模型的建立过程, 并对未来水平年其指标值进行预测。

## 1 预测模型的建立

### 1.1 富营养化评价指标的确定<sup>[1,2]</sup>

根据调查, 洞庭湖水体中 TN、TP 污染比较严重, 为此, 将 TN、TP 作为评价洞庭湖水体富营养化状态的重要指标。此外, 与湖泊富营养化有密切联系的叶绿素  $\alpha$ 、有机污染的代表性指标  $BOD_5$  也作为评价洞庭湖水体富营养化状态的指标。

### 1.2 影响因子及其与评价指标的相关性分析

鉴于洞庭湖水质污染状况和造成污染的原因以面源为主, 决定选择以下影响因子作为研究对象:

①工农业总产值。就目前经济发展水平而言, 一般

认为, 工农业产值越高, 污染物总量也相对高, 特别是农业产值越高, 面源污染物总量也就越高。②人口。据分析, 生活污染在洞庭湖污染负荷中占有相当大的比重。③耕地面积。耕地面积的多少, 既反映了农用化学物质消耗的量, 也反应了化肥流失情况。④化肥施用量。各种作物对化肥的利用率,  $N$  为 40% ~ 50%,  $P$  为 10% ~ 20%,  $K$  为 30% ~ 40%。通常化肥施用量越高, 流失量也越大, 其中有相当大的部分流失到水体中。⑤来水量。湖泊来水量的多少, 既反映了随来水携带污染物的多少, 又反映了湖泊径流量的大小与水体交换及自净能力。⑥来沙量。湖泊含沙量的变化直接影响到水体透明度, 透明度的改变又直接影响水体中的光合作用, 进而影响到水体中藻类繁殖。⑦农药施用量。⑧环保投入。表 1 为以上各评价指标及其影响因子 1992 ~ 1999 年的统计值, 2010 年数据为预测值。

表 1 洞庭湖富营养化评价指标及其影响因子统计值和估计值

年份 $k$	洞庭湖富营养化评价指标的影响因子								洞庭湖富营养化评价指标			
	$x(k,1)$ 工农业总 产值/亿元	$x(k,2)$ 人口 /万人	$x(k,3)$ 耕地面积 /hm <sup>2</sup>	$x(k,4)$ 化肥施 用量/万 t	$x(k,5)$ 来水量 /亿 m <sup>3</sup>	$x(k,6)$ 来沙量 /万 t	$x(k,7)$ 农药施用 量/kt	$x(k,8)$ 环保投入 /万元	$y(k,1)$ TN (mg·L <sup>-1</sup> )	$y(k,2)$ TP (mg·L <sup>-1</sup> )	$y(k,3)$ BOD <sub>5</sub> (mg·L <sup>-1</sup> )	$y(k,4)$ 叶绿素 $\alpha$ (mg·m <sup>-3</sup> )
1992	325.3449	1027.10	6927.5	164.1940	2.382	1.0392	15.146	8402	1.12	0.027	0.91	1.903
1993	389.8480	1035.95	6867.8	167.7051	2.621	1.3687	15.052	7269	1.23	0.024	0.81	3.429
1994	536.5869	1047.02	6820.6	177.9897	3.099	1.3486	15.700	6027	1.26	0.032	0.51	3.210
1995	670.6316	1056.30	6817.7	183.6439	2.667	1.1213	17.507	7557	1.32	0.133	0.93	2.083
1996	807.3311	1059.78	6804.5	181.5028	2.571	1.2197	16.673	3824	1.23	0.115	1.06	1.662
1997	887.6329	1067.64	6761.7	188.5356	2.634	0.6192	17.114	7753	1.08	0.106	1.14	2.710
1998	455.5877	1074.19	6732.9	186.6430	3.135	1.3784	17.184	5589	0.99	0.113	1.02	2.273
1999	488.4762	1016.60	6716.1	183.1206	2.448	0.6964	17.106	6616	1.25	0.221	1.06	1.856
均值	570.1799	1050.00	6800.0	179.1668	2.695	1.0990	16.435	6630	1.18	0.100	0.93	2.390
2010 <sup>①</sup>	1667.0	1071.6	6716.0	138.6	1964.4	0.4396	17.106	36675	1.135	0.127	1.046	7.346
2010 <sup>②</sup>	1667.0	1071.6	6716.0	138.6	1964.4	0.6594	17.106	36675	1.135	0.111	1.038	7.410

注: 2010<sup>①</sup> 2010<sup>②</sup> 分别为最低与最高两种估计情况。

作者简介: 李广源(1967—), 男, 湖南邵阳人, 高级工程师, 从事水资源和水环境调查、监测评价及规划工作。

第  $j$  个富营养化评价指标  $y(k, j)$  与其第  $i$  个影响因子  $x(k, i)$  的相关系数  $r(i, j)$  为<sup>[3]</sup>

$$r(i, j) = \frac{\sum_{k=1}^n (x(k, i) - \bar{x}_i)(y(k, j) - \bar{y}_j)}{(\sum_{k=1}^n (x(k, i) - \bar{x}_i)^2 \sum_{k=1}^n (y(k, j) - \bar{y}_j)^2)^{0.5}} \quad (1)$$

$$\bar{x}_i = \frac{\sum_{k=1}^n x(k, i)}{n} \quad (2)$$

$$\bar{y}_j = \frac{\sum_{k=1}^n y(k, j)}{n} \quad (3)$$

式中:  $n$  为样本数目;  $\bar{x}_i$  为第  $i$  个影响因子  $x(k, i)$  的均值;  $\bar{y}_j$  为第  $j$  个富营养化评价指标  $y(k, j)$  的均值。利用式(1)和表1中的数据,可计算洞庭湖富营养化评价指标及其影响因子的相关系数,见表2。

表2 洞庭湖富营养化评价指标及其影响因子的相关系数

相关系数 $r(i, j)$	TN $y(k, 1)$	TP $y(k, 2)$	BOD <sub>5</sub> $y(k, 3)$	叶绿素 $\alpha$ $y(k, 4)$
工农业总产值 $x(k, 1)$	0.085	0.331	0.420	-0.153
人口 $x(k, 2)$	-0.437	-0.040	0.221	0.062
耕地面积 $x(k, 3)$	0.195	-0.796	-0.448	0.177
化肥施用量 $x(k, 4)$	-0.163	0.689	0.459	-0.214
来水量 $x(k, 5)$	-0.267	-0.201	-0.454	0.444
来沙量 $x(k, 6)$	0.056	-0.558	-0.597	0.327
农药施用量 $x(k, 7)$	-0.068	0.825	0.604	-0.474
环保投入 $x(k, 8)$	-0.043	-0.227	-0.041	0.231

表2数据说明,与TN线性相关程度最高的前4个影响因子是人口、耕地面积、化肥施用量和来水量;与TP线性相关程度最高的前4个影响因子是耕地面积、化肥施用量、来沙量和农药施用量;与BOD<sub>5</sub>线性相关程度最高的前4个影响因子是化肥施用量、来水量、来沙量和农药施用量;与叶绿素  $\alpha$  线性相关程度最高的前4个影响因子是来水量、来沙量、农药施用量和环保投入。必须指出,这里样本数目仅为8,随着今后样本数目的增加,上述相关系数的计算结果可望趋于更稳定。

### 1.3 预测模型的参数优化及验算结果

实际上,影响洞庭湖富营养化评价指标的各因子和评价指标的历年变化也是复杂的,这导致洞庭湖富营养化评价指标及其影响因子之间的关系非常复杂,至今尚不清楚。从期望的角度看,用多元线性回归方程来描述这种关系就目前技术水平而言仍是合适的。考虑到目前所取得的样本数目较小,回归因子的数目不宜太多,故本文只取与预测对象即富营养化评价指标线性相关程度最高的前4个影响因子作为预测模型的输入变量:

$$y(k) = \bar{y} + c_1(x(k, 1) - \bar{x}_1) + c_2(x(k, 2) - \bar{x}_2) + c_3(x(k, 3) - \bar{x}_3) + c_4(x(k, 4) - \bar{x}_4) + \varepsilon \quad (4)$$

式中:  $\varepsilon$  为误差项;  $y(k)$  为某洞庭湖富营养化评价指

标,这里取 TN、TP、BOD<sub>5</sub> 或叶绿素  $\alpha$ ;  $x(k, i)$  为与该预测评价指标相对应的影响因子,这里均取4个因子;  $\bar{x}_i$  为影响因子的样本均值;  $c_1 \sim c_4$  为模型参数。

考虑到这里最小一乘法比最小二乘法在识别预测模型参数时更为稳健<sup>[3,4]</sup>,故本文用最小一乘法进行参数识别:

$$f(c_1, c_2, c_3, c_4) = \min_{k=1}^n |y_c(k) - y(k)| \quad (5)$$

式中:  $y_c(k)$  为预测模型的计算值,即式(4)等号右边除去误差随机项的部分;  $c_1 \sim c_4$  为优化变量;  $f(c_1, c_2, c_3, c_4)$  为目标函数。该问题是一非线性优化问题,常规优化方法处理较为困难,而用基于自然选择和自然基因机制的加速遗传算法(Accelerating Genetic Algorithm,简称AGA)则较为简便且有效。下面是AGA的计算原理<sup>[5]</sup>。

设模型的参数优化问题为

$$\min f = \sum_{i=1}^m \|F(C, X_i) - Y_i\|^q \quad (6)$$

$$\text{s.t. } a_j \leq c_j \leq b_j (j = 1, 2, \dots, p)$$

式中:  $C = \{c_j\}$  为模型  $p$  个待优化参数(优化变量);  $[a_j, b_j]$  为  $c_j$  的初始变化区间(搜索区间);  $X$  为模型  $N$  维输入向量;  $Y$  为模型  $M$  维输出向量;  $F$  为一般非线性模型,即  $F: R^N \rightarrow R^M$ ;  $\{(X_i, Y_i) | i = 1, 2, \dots, m\}$  为模型输入、输出  $m$  对观测数据;  $\|\cdot\|$  为取范数;  $q$  为实常数,如当  $q$  为1时为最小一乘准则,为2时为最小二乘准则,等等,可视实际建模要求而定;  $f$  为优化准则函数。

用AGA对洞庭湖富营养化评价指标 TN、TP、BOD<sub>5</sub> 和叶绿素  $\alpha$  的预测模型即式(4)的参数进行优化估计,结果如表3~表5所示。

表3 4种指标下预测模型的优化估计

指标	影响因子 $\bar{x}_1 \sim \bar{x}_4$	最高影响 因子均值	模型 参数	优化估计值
TN	人口/万人	1050	$c_1$	-2.4634
	耕地面积/hm <sup>2</sup>	6800	$c_2$	1.7579
	化肥施用量/万 t	179.1668	$c_3$	-0.0262
	来水量/亿 m <sup>3</sup>	2695	$c_4$	-0.0175
TP	耕地面积/hm <sup>2</sup>	6800	$c_1$	-1.1498
	化肥施用量/万 t	179.1668	$c_2$	0.1262
	来沙量/万 t	1.099	$c_3$	-0.0712
	农药施用量/kt	16.435	$c_4$	0.0316
BOD <sub>5</sub>	化肥施用量/kt	179.1668	$c_1$	0.6545
	来水量/亿 m <sup>3</sup>	2695	$c_2$	-0.4445
	来沙量/万 t	1.099	$c_3$	-0.0367
	农药施用量/kt	16.435	$c_4$	0.0490
叶绿素 $\alpha$	来水量/亿 m <sup>3</sup>	2695	$c_1$	0.7785
	来沙量/万 t	1.099	$c_2$	0.2940
	农药施用量/kt	16.435	$c_3$	-0.4465
	环保投入/万元	6630	$c_4$	0.2003

表 4 4 种指标下预测模型的验算结果

年度	TN				TP			
	实测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	预测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	绝对误差 ( $\text{mg}\cdot\text{L}^{-1}$ )	相对误差 /%	实测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	预测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	绝对误差 ( $\text{mg}\cdot\text{L}^{-1}$ )	相对误差 /%
1992	1.120	1.267	0.147	13.161	0.027	0.027	0.000	0.125
1993	1.230	1.230	0.000	0.000	0.024	0.012	-0.012	-50.424
1994	1.260	1.183	-0.077	-6.081	0.032	0.052	0.020	63.163
1995	1.320	1.166	-0.154	-11.660	0.133	0.133	0.000	-0.025
1996	1.230	1.157	-0.073	-5.900	0.115	0.098	-0.017	-14.416
1997	1.080	1.128	0.048	4.407	0.106	0.169	0.063	59.348
1998	0.990	1.098	0.108	10.922	0.113	0.118	0.005	4.415
1999	1.250	1.250	0.000	-0.001	0.221	0.162	-0.059	-26.892

  

年度	BOD <sub>5</sub>				叶绿素 $\alpha$			
	实测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	预测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	绝对误差 ( $\text{mg}\cdot\text{L}^{-1}$ )	相对误差 /%	实测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	预测值 ( $\text{mg}\cdot\text{L}^{-1}$ )	绝对误差 ( $\text{mg}\cdot\text{L}^{-1}$ )	相对误差 /%
1992	0.910	0.910	0.000	-0.002	1.903	3.060	1.157	60.823
1993	0.810	0.810	0.000	0.004	3.429	3.158	-0.271	-7.891
1994	0.510	0.697	0.187	36.737	3.210	2.987	-0.223	-6.961
1995	0.930	1.023	0.093	10.030	2.083	2.083	0.000	0.002
1996	1.060	1.007	-0.053	-4.957	1.662	1.662	0.000	-0.000
1997	1.140	1.069	-0.071	-6.217	2.710	2.124	-0.586	-21.607
1998	1.020	0.810	-0.210	-20.626	2.273	2.273	0.000	0.000
1999	1.060	1.113	0.053	5.013	1.856	1.778	-0.078	-4.193

表 3 和表 4 的计算结果说明,用式(4)作为洞庭湖富营养化评价指标 TN、TP、BOD<sub>5</sub> 和叶绿素  $\alpha$  的预测模型是较为满意的。

## 2 2010 年三峡工程运行初期洞庭湖富营养化评价指标变化趋势的预测

根据洞庭湖区各县、市、区“十五”发展规划资料以及国家有关方针、政策,确定了三峡工程运行初期(2010 年)各影响因子的估计值,参见表 1。表 1 中:①工业总产值、农业总产值、人口 3 项指标年平均增长率取各县市的中值,分别为 +12%、+4.7% 和 +0.48%。②耕地面积按土地利用总量动态平衡的规定维持在现有水平上,由于湖区内化肥施用量已经相当高,播种面积施用化肥量是全省的 1.26 倍,安乡县最高曾达到 1563 kg/hm<sup>2</sup>,随着农业产业结构的调整,化肥施用量将会下降,今后年平均增长率约为 -2.5%。③环保投入:1999 年为 6616 万元(据湖南省环境统计资料),占工农业总产值的 0.14%,2010 年湖区环保投入按工农业总产值的 0.25% 计算。④关于 2010 年的来沙量,既要考虑三峡水库建成后,输入洞庭湖的沙量减少后的值,又要在此值的基础上再减去“四水”中上游实施水土保持规划后向洞庭湖少排的沙量。按规划,到 2010 年“四水”每年向洞庭湖少输沙 1360 万 t,建库 13 年后,入湖平均沙量为 5977 万 t,减去 1360 万 t,后期为 4617 万 t,与 1999 年比较减少 57.10%。因此,三峡工程运行初期,洞庭湖的来沙与 1999 年比较,按减少 40%~60% 计算比较接近实际,因此,本预测采用最低与最

高情况分别进行,以寻求变化范围。⑤农药施用量取 1999 年水平。

把上述 2010 年各影响因子的估计值代入预测模型式(4),可得 2010 年洞庭湖富营养化各评价指标的预测值,参见表 1。从预测情况来看,与 1999 年洞庭湖富营养化情况相比,2010 年叶绿素  $\alpha$  有了明显的提高,而其他评价指标则变化不大,这与客观发展情况也是相符的。在坚持可持续发展的战略原则下,随着国家对环境保护管理及防治措施等方面的加强,入湖的污染物总量应会得到基本控制,因而 TP、TN、BOD<sub>5</sub> 这些指标浓度也会相应维持在一定水平,但洞庭湖水体中 TP、TN 浓度目前水平较高,已具备了富营养化的潜在能力,且在今后年份中,氮、磷等污染物还会存在一定的总量继续入湖及部分滞留湖中。在这种营养物质继续存在的条件下,随着三峡工程运行后,入湖泥沙大量减少,水体透明度增大,水体中光合作用增强,影响富营养化的藻类繁殖能力增强,叶绿素  $\alpha$  将会明显提高。

## 3 结 语

a. 分析了洞庭湖富营养化评价指标及其影响因子之间的统计特性,据此建立了描述洞庭湖富营养化评价指标及其影响因子之间变化关系的、基于最小一乘法的多元线性回归预测模型,并给出了统一用加速遗传算法进行洞庭湖富营养化评价指标预测模型的参数优化估计。上述建模方法简便、通用,且精度高。

(下转第 28 页)

平均处理成本 64.19 元/t,平均经营成本 33.19 元/t,则年垃圾处理成本节约 193.67 万元,年垃圾处理经营成本节约 100.14 万元。

综上所述,由于该工艺使大部分垃圾有机成分得到发酵处理,大大减少了填埋场甲烷等易燃、易爆、有害气体的产生,有效地降低了填埋场沼气燃烧、爆炸的可能性和有害气体泄入大气,使填埋场的运行更加安全<sup>[4]</sup>。

由于发酵处理,减少了垃圾在填埋场的生化分解,有效地降低了填埋场废弃物的沉降,使填埋保护系统开裂、填埋场形成凹塘造成运行困难的可能性降低。分选预处理车间与一次发酵车间为垃圾进一步处理的资源化(焚烧、堆肥等)创造了良好条件。

(上接第 22 页)买入或卖出排污权。实行排污权交易的条件:一是合理分配排污权,二是完善的市场条件,三是政府部门的有效管理<sup>[4]</sup>。

在政府管理机构没有增加排污权的供给,即总的环境状况没有恶化的前提下,通过排污权交易,边际治理成本比较高的污染者将买进排污权,而边际治理成本比较低的污染者将出售排污权,其结果是全社会总的污染治理成本最小化。在不考虑其他因素影响的条件,排污权交易应该会降低污水处理费水平。

排污权交易充分运用市场作用,用最小的社会污染治理成本,实现总量控制的目标,鉴于其类似的调控作用,对浙江省这种经济发达、市场化程度较高、需水和排污量大且污染较严重的地区来说,在污水处理费保证下的排污权交易确实值得考虑。

浙江省自改革开放以来,经济发展迅速,市场化程度较高,近年来各级政府对于水污染问题非常重视,而且我国曾经有过排污收费和排污权交易配套使用的尝试,所以浙江省已经初步具备实行排污收费保证下的排污权交易的条件。

(上接第 25 页)

b. 本文建立的预测模型属于统计模型,因此用于建模的样本数据的代表性对模型的实际预测效果的影响较大。随着样本数据的不断积累,预测模型按照本文的建模方法可继续优化估计。

c. 鉴于洞庭湖富营养化评价指标及其影响因子之间的关系较为复杂,可以从不同的角度提取样本信息,从而建立不同的预测模型(洞庭湖富营养化评价指标的影响因子及其数目、洞庭湖富营养化评价指标及其影响因子之间的关系都可以不同),然后建立组合预测模型,以综合利用多种预测模型所提供的预测信息来提高预测的精度和稳定性,而各单个预测模型的权重仍可用本文给出的加速遗传算法

## 参考文献:

- [1] 林援朝. 城市垃圾管理与处理处置技术标准规范应用实务全[ M ]. 北京:光明日报出版社, 2002. 760~772.
- [2] 中华人民共和国建设部. 城市生活垃圾堆肥处理工程项目建设标准[ M ]. 北京:中国计划出版社, 2001. 6~7, 10~11.
- [3] 中国国际工程咨询公司. 投资项目经济咨询评估指南[ M ]. 北京:中国经济出版社, 1998. 61~81.
- [4] 钱学德, 郭志平, 施建勇. 现代卫生填埋场的设计与施工[ M ]. 北京:中国建筑工业出版社, 2001. 159~173, 203~205.

(收稿日期 2004-04-15 编辑:高渭文)

## 4 结 语

本文对浙江省水价组成和水价执行机制进行了分析。在对环境水价的着重分析中,重点讨论了政府投资、污水回用对污水处理费的影响和污水处理费对排污权交易的保障作用。最终认为在江南水乡区域应该实行以污水处理费为保证金的排污权交易制度,同时减少政府对污水处理厂建设、运营的投资,加大污水回用力度。

## 参考文献:

- [1] 秦翠霞. 征收污水处理费解决污水厂运行费用[ J ]. 中国给水排水, 1999(3): 32~33.
- [2] 石进国. 从资金角度谈城市污水集中处理如何步入良性循环[ J ]. 城市发展研究, 1998(5): 31~35.
- [3] 张晓昕. 浅谈北京市污水处理综合利用的可行性[ J ]. 北京节能, 1999(3): 6~9.
- [4] 马中. 环境与资源经济学概论[ M ]. 北京:高等教育出版社, 1999. 204~211.

(收稿日期 2003-06-05 编辑:傅伟群)

来最佳确定。

## 参考文献:

- [1] 饶群, 芮孝芳. 富营养化机理及数学模拟研究进展[ J ]. 水文, 2001, 21(2): 15~19, 24.
- [2] 丁晶, 邓育仁. 随机水文学[ M ]. 成都:成都科技大学出版社, 1988. 1~387.
- [3] 金菊良, 杨晓华, 储开凤, 等. 加速基因算法在海洋环境预报中的应用[ J ]. 海洋环境科学, 1997, 16(4): 7~12.
- [4] 金菊良, 杨晓华, 丁晶. 标准遗传算法的改进方案——加速遗传算法[ J ]. 系统工程理论与实践, 2001, 21(4): 8~13.
- [5] 金菊良, 丁晶. 水资源系统工程[ M ]. 成都:四川科学技术出版社, 2002. 34~72.

(收稿日期 2003-05-19 编辑:高渭文)