

# 长江水质评价和预测研究

张 虎, 蔡 燕, 姚海强

指导教师: 杨春德

(重庆邮电学院, 重庆 440803)

**编者按:** 本文结构完整, 表述清晰。自定义了综合污染指数, 综合评价的思路有可取之处; 分段考虑了主要污染源所在, 对结果做了尝试性的解释, 但未考虑两观测站间单位长度的污染量; 用时间序列建模及处理污水量的规划问题思路清晰, 但一次累加拟和模型中多项式指数的作用和含义不够明确。值得一提的还有, 最后的建议中与前面的结果相互印证。

**摘 要:** 本文旨在研究长江过去10年的水质变化情况, 以此来预测未来10年长江的水质情况, 并对长江的治理提供一些具有可行性的建议和意见。在第一问中, 我们从模糊数学的角度出发提出综合污染指数的概念, 并运用该指数分析各地区的水质污染情况, 得出结论: 04年较03年恶化, 05年有所好转; 在第二问中, 我们将7个长江干流观测点的水质报告表和基本数据表相结合, 用每秒流过观测点的水中所含污染物的量减去上一个观测点的水中所含污染物经过自净后残余的量, 即为两个观测点之间污染物增加的量, 得出: 高锰酸盐和氨氮的最主要污染源在湖南岳阳城及其上游地区; 在第三问中, 我们建立了针对各类水所占评价河长百分比的一次累加拟合模型和时间序列模型, 得出结论: 10年之后, I、II类水都已不存在, III类水仅存0.62%; 在第四问中, 我们采用多元线性回归, 得出长江流量, 废水排放总量, IV类, V类, VI类水占水文年全流域长度百分比之间的线性关系, 以第三问中预测数据为基础, 以未来10年总共要处理的污水为目标, 建立线性规划并求解。

**关键词:** 归一化法; 自净系数; 一次累加拟合模型; 时间序列法; 多元线性回归模型

分类号: AMS(2000) 76Z10

中图分类号: X830.3

文献标识码: A

## 1 基本假设

1) 假设主要污染物高锰酸盐指数和氨氮的自身降解系数为0.2 (单位: 1/天); 2) 假设观测点  $j$  和观测点  $j+1$  之间江水的流速是所测两点  $j$  和  $j+1$  的速度的平均值; 3) 假设忽略观测点  $j$  和观测点  $j+1$  之间增加的污染物的自净; 4) 假设2005~2014年的评价河长与2004年的评价河长相等; 5) 假设在2005~2014年间不会发生大旱大涝等自然灾害性天气; 6) 假设长江水的总流量  $z l_y$ , 废水排放总量  $w s_y$ , IV类、V类、劣V类三种类别的水在水文年全流域的百分比之间成线性关系。

## 2 问题分析

问题一中数据较多, 我们可以用统计的方法, 求出每年各类水所占的一个百分比, 以此来说明长江水质的一个整体变化趋势。针对各地区而言, 我们可以把影响水质的四个主要因素进行加权, 从而求出一个综合污染指数, 以此为标准来判断各地区水质的污染状况。在问题二中, 由于江水在流动过程中对污染物可以进行一定的自身降解, 所以水质最差的地区不一定是污染源, 如果上游污染严重, 那么经过积累作用, 到达下游的时候水质也会变得很差。仔细分析附表四, 发现每年评价的河长差别很大, 不利于我们建模求解。我们认为不妨从各类水所

占百分比入手, 预测出未来10年各类水所占的百分比即可。问题四的求解需要建立在问题三的基础上, 由于题目要求没有劣 V 类水的存在, 因此很容易得出劣 V 类水要全部进行处理。

### 3 模型的建立与求解

#### 3.1 问题一

首先, 我们从整体着手, 对进两年多来长江水质进行整体的定量分析, 统计每年长江水质按照水质标准分类而成的各种水类其绝对数值和相对百分比, 得出长江水质04年较03年均比较差, 但05年水质恶化现象得到了一定的控制, 其水质明显好于04年。两年来, 长江水质呈现出先恶化再好转的趋势。然后, 对于各地区水质的污染状况, 我们综合考虑对评判水质的四个因子: 溶解氧、高锰酸盐指数、氨氮和 PH 值, 采用模糊数学的方法对四者进行加权, 从而得到一个综合污染指数  $w_{ij}$ , 以这个指数来定量说明各地区水质的污染情况。我们以 I 类水: 溶氧量 7.5 mg/L, 高锰酸盐 2 mg/L, 氨氮 0.15 mg/L, ph 值取 6~9 的中值 7.5 为标准提出了一个综合污染指数的概念

$$w_{ij} = 7.5/do_{ij} + co_{ij}/2 + nh_{ij}/0.15 + (ph_{ij} - 7.5)^2$$

很明显, 按照综合污染指数函数的定义式, 综合污染指数越高, 水质越差。由以上表格分析得出各地区水质污染状况, 总体来讲, 17个观测点04年水质比03年水质明显恶化, 05年水质有较程度的好转。

#### 3.2 问题二

在问题二中, 我们把两观测点间汇入支流的主要污染物高锰酸盐指数和氨氮的自身降解忽略不计, 并把水质主要污染物高锰酸盐指数和氨氮的自身降解系数 (以下简称自净系数) 假设为 0.2 (单位 1/天)。我们根据附件三中 2004.4 ~ 2005.4 的长江主干水质监测报告中数据测算出 2004.4 ~ 2005.4 间长江水流过两相邻观测点所需时间: 第  $j$  点到  $j+1$  点距离为  $l_{j+1} - l_j, \dots, j = 1, 2, \dots, 6$ 。以两个观测站点的速度平均值

$$t_{ij} = \frac{l_{j+1} - l_j}{(v_{ij} + v_{i(j+1)})/2}, \quad i = 1, 2, \dots, 13, j = 1, 2, \dots, 6.$$

作为两站点间水流速度, 时间即为

$$t_{ij} = \frac{l_{j+1} - l_j}{(v_{ij} + v_{i(j+1)})/2}, \quad i = 1, 2, \dots, 13, j = 1, 2, \dots, 6.$$

由每个点的水流量我们可以计算出在各观测点每秒流过的水中所含有的 CODMn 的量为: CODMn 的浓度  $\times$  流量, 即

$$m_{ij} = co_{ij} \times sl_{ij}, \quad i = 1, 2, \dots, 13, j = 1, 2, \dots, 7.$$

水流经过自净到达下一观测站点时, 每秒流过的水中所含有的 CODMn 的量为: CODMn 的浓度  $\times$  流量  $\times$  自净系数的天数次幂, 即

$$m'_{ij} = m_{ij} \times (1 - x)^{t_{ij}} = co_{ij} \times sl_{ij} \times (1 - x)^{t_{ij}}, \quad i = 1, 2, \dots, 13, j = 1, 2, \dots, 7.$$

两站点间增加的 CODMn 的量

$$m_{ij} - m'_{ij} = co_{ij} \times sl_{ij} - co_{ij} \times sl_{ij} \times (1 - x)^{t_{ij}}, \quad i = 1, 2, \dots, 13, j = 1, 2, \dots, 6.$$

我们在用进行数值计算得过程中，增加量有部分为负值，而实际当中负值是不可能存在的，因此将负值全部修正为0。2004年4月~2005年4月13个月 CODMn 的平均增加值

$$\sum_{i=1}^{13} \sum_{j=1}^6 (m_{ij} - m'_{ij}) / 13.$$

同理，根据上面的推导过程，我们也可以得到 NH<sub>3</sub>-N 的平均增加值。

经分析，显然可得出以下结论：1)、高锰酸盐得主要来源为：湖南岳阳城陵矶及其上游地区，湖北宜昌南津关及其上游地区，重庆朱沱及其上游地区。2)、氨氮得主要来源为：湖南岳阳城陵矶及其上游，湖北宜昌南津关及其上游，重庆朱沱及其上游地区。即高锰酸盐和氨氮均来自于：湖南岳阳城陵矶及其上游地区。

### 3.3 问题三

#### 模型一 一次累加拟合模型

观察已知数据可知各数据具有很大的随机性和波动性，所以我们采用了百分比的一次累加序列来进行多项式的拟合，并以拟合出来的多项式为依据，计算未来10年的各类水占评价河长的百分比：我们以2004年的评价河长为准，为固定值：全流域评价河长：39412km，干流评价河长：6341km，支流评价河长为：33071km。

首先，我们对数据序列  $sj_0(k)$  按以下规则进行一次累加，得到一次累加数据序列

$$sj_1(1) = sj_0(1), \quad sj_1(k) = sj_1(k-1) + sj_0(k), \quad k \geq 2.$$

累加后的数据序列  $sj_1(k)$  成为了一个单调递增序列，克服了原始数据序列的随机性和波动性。然后我们对一次累加数据序列  $sj_1(k)$  进行多项式拟合，得出  $sj_1(k)$  的表达式

$$sj_1(k) = \alpha_n \cdot k^n + \alpha_{n-1} \cdot k^{n-1} + \cdots + \alpha_1 \cdot k + \alpha_0.$$

然后将待预测的  $k$  值带入，即得到在  $k$  时刻  $sj_1(k)$  的取值，根据以下式子即可得出待预测量  $sj_0(k)$  的值

$$sj_0(k) = sj_1(k) - sj_1(k-1), \quad k \geq 2.$$

由于从 I 到劣 V 类的水是单独预测的，故不能保证其总和为100%，因此我们应求得第  $y$  年的和  $\sum_{a=1}^6 i_{ya}$ ，用每类水所占评价河长百分比数据在第  $y$  年的和  $\sum_{a=1}^6 i_{ya}$  中所占百分比来修正：

$$i'_{ya} = i_{ya} \times 100 / \sum_{a=1}^6 i_{ya},$$

再根据2004年的评价河长，即可得到各类水所占的河长。

#### 模型二 基于时序分析法的水质总体趋势预测

将1995~2004年水文年全流域水质分类比例列为矩阵  $X$ ，令  $\begin{cases} a_{i1} = x_{i1} + x_{i2} + x_{i3} \\ a_{i2} = x_{i4} + x_{i5} + x_{i6} \end{cases}$  可得

$$A = \begin{bmatrix} 93.1 & 85.3 & 80.7 & 88.4 & 80.2 & 74.0 & 73.7 & 76.7 & 77.5 & 68.0 \\ 6.9 & 14.7 & 19.3 & 11.6 & 19.8 & 26.3 & 26.3 & 23.2 & 22.5 & 32.0 \end{bmatrix}^T$$

预测方程式如下： $\bar{a}_t = k \cdot l^t$ ，对数变换得  $\log \bar{a}_t = \log k + t \cdot \log l$

令:  $\bar{k} = \log k$ ,  $\bar{l} = \log l$  则:  $\log \bar{a}_t = \bar{k} + t \cdot \bar{l}$

$$\bar{l} = \frac{T - R}{N - 4}, \quad \bar{k} = R - \frac{7}{3} \cdot \bar{l},$$

$$\begin{cases} R = (1/10)(\log a_1 + 2 \log a_2 + 3 \log a_3 + 4 \log a_4), \\ T = (1/10)(\log a_7 + 2 \log a_8 + 3 \log a_9 + 4 \log a_{10}). \end{cases}$$

将上表有关数据带入计算得  $R = 1.9337$ ,  $T = 1.8635$ ,  $\bar{l} \approx -0.0088$ ,  $\bar{k} \approx 1.9542$ .

预测2006年长江水域可饮用水的比例为

$$\log \bar{a}_t = 1.9542 + (-0.0088) \cdot 12 \approx 1.8488, \quad \bar{a} = 10^{1.8488} \approx 70.5992.$$

预测误差指标选择均方根误差 RMSE; 令预测误差:  $e_t = a_t - \bar{a}_t$ , 则

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum e_t^2} = \sqrt{\frac{158.89}{10}} \approx 3.98,$$

预测值区间

$$\bar{a}_{t0} - f \sqrt{\frac{1}{N-1} \sum e_t^2} < \bar{a}_t \leq \bar{a}_{t0} + f \sqrt{\frac{1}{N-1} \sum e_t^2}.$$

式中:  $N-1$  为自由度, 本次预测自由度为  $10-1=9$ .  $f$  为概率度, 为避免与时间  $t$  混淆, 因此将概率度由  $t$  换为  $f$ . 本次预测样本数未超过20项, 为小样本, 用  $t$  分布概率度, 查  $t$  分布表, 按80%信度计, 自由度为9,  $t$  值为1.38, 则

$$\bar{a}_{t0} - 5.8 \leq \bar{a}_t \leq \bar{a}_{t0} + 5.8.$$

基于时间序列法研究得出的结论可知, 长江水质虽然有时某一年的情况较前一年可能会有所好转, 但是它的总体趋势是在逐年恶化。其中好转的情况体现在随机误差较大这一特征之上。这与前面所提到的一次累加序列模型相统一的, 同时也是对一次累加序列模型不能体现水质情况随机性的一个有力补充。

### 3.4 问题四

本问要求我们计算如果未来10年内每年长江干流的 IV 类和 V 类水的比例控制在20%以内, 且没有劣 V 类水, 那么每年需要处理的污水量。即以控制污水总量的方式来控制 IV 类、V 类、劣 V 类水的比例, 因此我们需要知道污水总量与 IV 类、V 类、劣 V 类水类水所占比例之间的关系。在此, 我们建立了多元线性回归模型来求解:

从整体考虑, 在过去的10年中, 长江水的总流量  $zl_y$ , 废水排放总量  $ws_y$ , IV、V、劣 V 类三类别的水在水文年全流域的百分比之间的关系可以用以下式子来表示

$$ws_y = \beta_1 \cdot zl_y + \beta_2 \cdot i_{y4} + \beta_3 \cdot i_{y5} + \beta_4 \cdot i_{y6}, \quad y = 2005, 2006, \dots, 2014$$

其中

$$y = ws_y = \begin{bmatrix} 174 \\ 179 \\ 183 \\ 189 \\ 207 \\ 234 \\ 220.5 \\ 256 \\ 270 \\ 285 \end{bmatrix}, \quad X = i_{ya} = \begin{bmatrix} 9250 & 3.9000 & 3.0000 & 0 \\ 9513 & 9.7000 & 1.9000 & 3.1000 \\ 9171.26 & 13.300 & 2.6000 & 3.4000 \\ 13127 & 8.3000 & 1.7000 & 1.6000 \\ 9513 & 9.5000 & 6.2000 & 4.1000 \\ 9924 & 16.600 & 4.4000 & 5.3000 \\ 8892.8 & 14.000 & 5.5000 & 6.8000 \\ 10210 & 10.000 & 3.2000 & 10.0000 \\ 9980 & 6.4000 & 5.8000 & 10.3000 \\ 9405 & 14.8000 & 5.9000 & 11.3000 \end{bmatrix}$$

代入以上数据，关系式变为

$$ws_y = 0.0132 \cdot zl_y + 0.8208 \cdot i_{y4} + 8.3156 \cdot i_{y5} + 8.3657 \cdot i_{y6}, \quad y = 2005, 2006, \dots, 2014.$$

在2005~2014年间，我们每年都要处理一定量得污水  $cl_y$ ，上式变为

$$ws_y - cl_y = 0.0132 \cdot zl_y + 0.8208 \cdot i_{y4} + 8.3156 \cdot i_{y5} + 8.3657 \cdot i_{y6},$$

变换得

$$cl_y = ws_y - [0.0132 \cdot zl_y + 0.8208 \cdot i_{y4} + 8.3156 \cdot i_{y5} + 8.3657 \cdot i_{y6}], \quad y = 2005, 2006, \dots, 2014.$$

因为要处理得污水应尽可能得少，故建立以下目标函数

$$\min \sum_{y=2005}^{2014} cl_y \quad st : \dots \begin{cases} i_{y4} + i_{y5} \leq 20 \\ i_{y6} = 0. \end{cases}$$

目标函数及约束化简为

$$\min \left[ 2402.4 - 0.8208 \cdot \sum_{y=2005}^{2014} i_{y4} - 8.3156 \cdot \sum_{y=2005}^{2014} i_{y5} \right]$$

$$st : \begin{cases} i_{y4} \leq \{7.25, 13.55, 14.38, 15.52, 17.07, 18.25, 17.75, 17.22, 16.66, 16.07\}^T \\ i_{y5} \leq \{7.62, 7.88, 8.87, 10.11, 11.71, 13.14, 13.38, 13.55, 13.67, 13.72\}^T \\ i_{y4} + i_{y5} \leq 20, y = 2005, 2006, \dots, 2014. \end{cases}$$

线形规划求解得：min 1390.621 亿吨。

将  $z$  输出整理即得出  $i_{y4}$ ,  $i_{y5}$  的数据，然后根据下式即可求出每年应处理得污水量。

$$cl_y = ws_y - [0.0132 \cdot zl_y + 0.8208 \cdot i_{y4} + 8.3156 \cdot i_{y5} + 8.3657 \cdot i_{y6}].$$

### 3.5 问题五

通过对附件一、二的认真研读,以及查阅相关的资料我们从政策层面方面提出以下两点建议:1、加强宣传力度,唤醒民众正确认识长江的现状。2、立即制定相应的法律法规,加大对违法排污行为的处罚力度。在具体治理长江方面的几点意见:1、加强对中小支流的治理。2、加强主要污染城市的污水处理力度,污水达不到标准,则由政府出面买单。3、对污染严重的企业实行关停并转,集中建设实力强、能治理好污染的大企业。4、大力发展低污染农业,控制农药化肥的使用量。5、提高警惕,坚决防止江水治了又污。

#### 4 模型的评价与改进

在本文中,我们没有使用单一的模型对问题进行分析求解,而是综合运用了多种数学模型,并且增加了一定的检验环节。这样就增加了我们所求数据的合理性,同时也增加了论文的实际参考价值。在运用归一化法综合评价长江水质情况时,我们很巧妙的将多个指标转化为单一指标,这样就很好的解决了多指标难以比较的难题。在解决问题三时,通过一次累加序列拟合法和时间序列法的有机结合,解决了长江水质情况随机性和波动性较大从而难以准确预测的难题。

在改进方面我们认为:在问题一中模糊数学确定的污染指数可以改进为用模糊神经网络进行判断的指数,可以更好的反映污染程度,以及不同点的关系;问题二中,可以改进为灰色评价,以更好的反映污染物的来源;问题三中,我们用以预测的一次累加拟合模型虽然在数据点以内误差较小,但在数据点以外,进行预测时,误差将是不可知的,可以改进为使用时间序列法、灰色系统理论、BP神经网络进行预测,或者将其三者有机的结合在一起,通过优化组合形成最优化权值组合法,这样既可以达到较高的精度要求又可以避免应用大量的参数,达到最小误差预测的目的。

### The Water Quality Evaluation in Yangtze River and Estimate Research

ZHANG Hu, CAI Yan, YAO Hai-qiang

Advisor: YANG Chun-de

(Chongqing University of Posts and Telecommunications, chongqing 440803)

**Abstract:** This paper aims at researching the change of the water quality in Changjiang River over the past ten years. So as to forecast the water quality in the next ten years as well as putting forward some constructive suggestions on renovation. In the first question, we bring forward the concept of colligating pollution exponent from the aspect of fuzzifying mathematics idea and use this concept to analyze the situation of water pollution. The conclusion we draw is: the water quality in 2004 is worse than that in 2003, but to the year 2005, the situation is becoming better. In the second question, we combine the water quality observation (which was gained from seven locations along the Changjiang River) and basic data report. we use mathematical way to obtain the amount of the increasing pollution. Our conclusion is: Yueyang city in Hunan province and its upper drainage areas are the main sources of permanganate and nitric ammonia pollution. In the third question, we establish the once accumulate the fit pattern self-purification factor and the time queue law pattern and arrive at one conclusion: ten years later, the first two types of water will no longer exist, and the third type water will only remain 0.62% of the whole. In the fourth question, we use plurality linear regression pattern to conclude the linearity relationships among the flux of Changjiang River, the total amount of waste water, the percentage of the fourth/ fifth/ sixth type of water and put up our way to solve the problem.

**Keywords:** normalization law; self-purification factor; once accumulate the fit pattern self-purification factor; time queue law; plurality linear regression pattern