

文章编号:1005-3085(2005)07-0026-09

长江水质的评价和预测

李云锋, 王 勇, 杨 林

指导教师: 沈锦仁

(中国人民解放军理工大学, 南京 210007)

编者按: 本文用差分方程和回归分析的方法对问题作了正确、恰当的分析处理, 结果合理, 具有一定的创造性。

摘 要: 本文利用长江流域近两年多主要城市水质检测数据, 通过对原始数据进行归一化综合处理, 确定了水质新的综合评判指标函数 Ψ 。在对整个长江流域所有观测站的位置关系作一定的简化假设后, 得到长江综合评定函数值 $\Psi = 0.4331$, 水质为良好, 主要污染物为氨氮。

通过建立污染浓度的反应扩散方程, 本文用三种方法反演出未知的污染源强迫函数 $f(x, t)$, 并对 $f(x, t)$ 的三种数据加以综合分析, 分别给出了高锰酸钾盐和氨氮污染源的主要分布地区。

为了对长江未来水质污染发展趋势进行预测, 本文建立了回归分析模型并对回归系数进行了 F 检验, 结果是如果不采取有效的治理措施, 长江可饮用水将逐年下降, 且10年后可饮用水所占长江水总量的比例将不到 50%。根据这一预测结果, 我们进而使用二元线性回归模型, 通过对各种不可饮用水进行综合考虑, 得到如下结果: 要在未来10年内使长江干流的不可饮用水 (IV 类和 V 类水) 的比例控制在 20% 以内, 且没有劣 V 类水, 那么每年污水处理量至少为 75.195 亿吨。

关键词: 归一化; 水质综合评判指标函数; 反应扩散方程; 回归分析

分类号: AMS(2000) 76Z05

中图分类号: O212

文献标识码: A

1 长江水质的综合评价

1.1 问题的分析

附件3-1 (长江流域主要城市水质检测报告) 从多方面反映了长江近两年多的水质情况, 因此对于长江流域水质的综合评价, 主要是对水质检测报告原始数据的处理。

由于地表水环境质量标准四个主要项目指标 (溶解氧、高锰酸盐指数、氨氮、PH 值) 的原始数据的量纲各不相同, 首先对数据进行归一化和综合处理; 又由于数据量较大, 因此要对归一后数据进行综合分析, 在对数据进行处理的过程中, 考虑下面要求:

1) 归一化后的数据能够反映可饮用水与不可饮用水的区别, 为此将 III 类与 IV 类水的分界点作为界值点, 赋值为 1;

2) 对于任一个项目指标值来讲, 数据越小, 水质越好;

3) 四个项目指标综合后仍以 1 作为可饮用水与不可饮用水的分界点, 且数据越小, 水质越好。

1.2 理论分析与算法步骤

1) 数据的归一化和综合

按照上述要求, 根据附表中提供的项目标准限值, 对数据归一化处理:

以一个观测点某一时刻为例, 依次记四个项目指标的实际测量值为 p_k , 归一化的水质指标值 $P_k (k = 1, 2, 3, 4)$ 。

对于溶解氧 (DO), 定义

$$P_1 = \frac{5}{p_1}; \quad (1)$$

对于高锰酸钾指数 (CODMn), 定义

$$P_2 = \frac{p_2}{6}; \quad (2)$$

对于氨氮 (NH₃-N), 定义

$$P_3 = \frac{p_3}{1.0}; \quad (3)$$

对于 PH 值来讲, 由于 PH 值在 6-9 之间均为可饮用水, 假设当 PH 值=7 时, 水质是最好的, 作如下定义

$$P_4 = \begin{cases} \frac{1}{2} \times (p_4 - 7) & p_4 \geq 7, \\ 7 - p_4 & p_4 < 7. \end{cases} \quad (4)$$

记 P_{ij} 为第 i 个观测点第 j 个月的水质综合指标值, $(P_k)_{ij}$ 为第 i 个观测点第 j 个月归一化后的水质指标值 ($i = 1, 2, \dots, 17; j = 1, 2, \dots, 28; k = 1, 2, 3, 4$), 考虑不同项目指标对水质影响程度的不同, 将四个项目指标的权重设为 $a_k (k = 1, 2, 3, 4)$, 水质综合指标值 P_{ij} 的表达式为

$$P_{ij} = \begin{cases} \sum_{k=1}^4 a_k \times (P_k)_{ij} & (P_k)_{ij} \leq 1, (k = 1, 2, 3, 4) \\ \max_k (P_k)_{ij} & \exists k, (P_k)_{ij} > 1. \end{cases} \quad \left(\sum_{k=1}^4 a_k = 1 \right) \quad (5)$$

2) 单个观测点水质的评估向量和长江全流域水质的综合评价

知道了一个观测点 28 个月份的水质综合指标值, 要求对该观测点水质的综合评价能够反映水质情况、是否可饮用和可饮用的程度等目标。由此构造各观测点水质综合指标向量 $Q_i = [Q_{i1}, Q_{i2}]$, 其中

$$Q_{i1} = \frac{1}{28} \sum_{j=1}^{28} P_{ij}$$

为第 i 个观测站水质的综合量值, Q_{i2} 为第 i 个观测站水质为不可饮用水的月份个数。

假设一个观测站代表一块水域, 设该水域内水质均匀, 水量为 R_i , 17 个观测站代表的水域覆盖了整个长江流域且不重复覆盖, 构造整个长江流域水质综合评价函数为

$$\Psi = \frac{\sum_{i=1}^{17} Q_{i1} R_i}{\sum_{i=1}^{17} R_i}. \quad (6)$$

1.3 水质等级标准的确定

根据式(1)-(4)式对数据的处理方法, 计算附表中四个主要项目标准限值对应指标值, 为了便于以后的讨论计算, 不妨将 PH 值按照氨氮的指标值严格的分类, 取 $a_1 = 0.2, a_2 = a_3 = 0.35, a_4 = 0.1$, 根据(5)式计算四个项目标准对应的综合指标值, 同时考虑水质污染月份的个数, 将其分别对应到水质为优质、良好、轻微污染、污染和严重污染, 则得到水质等级的划分标准见表1。

1.4 长江水质的综合评价

表 1: 水质综合指标分类限值

分类	I类	II类	III类	IV类	V类	劣V类
水质类别	良好	优质	轻微污染	污染	严重污染	强严重污染
综合指标值 $Q_i \leq$	0.317495	0.625005	1	1.6667	2.5	100
综合指标值 $Q_i \leq$	0	1	7	13	19	28

根据(1)-(5)式对28个月的观测数据进行处理求得 P_{ij} , 然后计算水质综合指标向量 $Q_i = [Q_{i1}, Q_{i2}]$, 结合表1的水质综合指标分类限值, 即可对各观测点水质情况给出相应的评价: 主要污染的为四川乐山岷江大桥和江西南昌滁槎, 主要污染物为氨氮。对于整个长江流域的水质的综合评定, 关键在于如何确定 R_i , 已知主干流和各观测点的相对位置关系如图1所示(该相对距离关系图根据地图近似取得)。

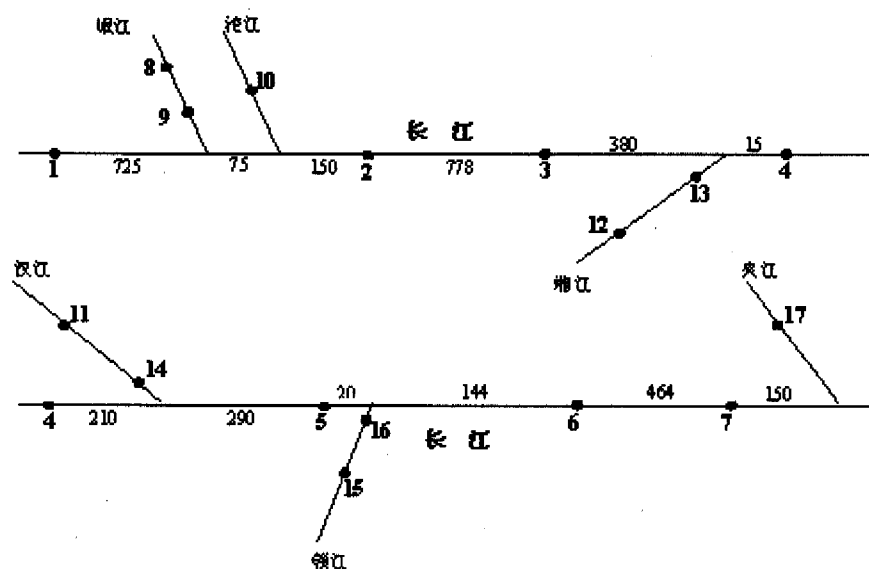


图 1: 长江各观测站相对位置图

设每一个观测点对应的水域长度为 D_i , 对应水流横截面积为 S_i , 则有

$$R_i = S_i \times D_i = \frac{L_i}{v_i} \times D_i. \quad (7)$$

考虑支流的 D_i 、 S_i , 有如下假设:

- 1) 干流相邻两个观测站的水流横截面积之差, 即为两观测站间所有支流水流横截面积之和;
- 2) 两观测站之间所有支流水流的横截面积相等;
- 3) 支流汇入干流的水量在下一个观测点处瞬间与干流水质均匀混合, 即在支流进入干流到下一个观测点水质不混合;
- 4) 对于支流对应水域的长度, 考虑观测点与干流间的距离, 根据观察适当赋值即可;
- 5) 当一个支流上有两个观测点时, 认为两者对应的水流截面积相等;

6) 任意观测站到下一个干流观测站之间水质均匀, 与该观测站水质相同。

根据上述假设和处理结果, 来确定所有观测点的 D_i 和 S_i 。

对于干流观测站对应的 D_i 、 S_i 可以根据附3-2 (长江干流主要观测站点的基本数据) 求得。假设第 i 个观测点第 j 个月的水流量为 L_{ij} , 第 i 个观测点第 j 个月的水流速度为 v_{ij} , 则第 i 个观测点的平均水流截面积为

$$S_i = \frac{1}{12} \sum_{j=1}^{12} \frac{L_{ij}}{v_{ij}} \quad (8)$$

对于观测站7, 其对应的 D_7 是该观测点到长江入海口的距离, 从图1上仅能知道 $D_7 > 150$, 不妨设 $D_7 = 350$, 可以得到长江干流个各站点间距离和水流横截面积。

对于支流上各观测点的 D_i 由观测站到干流距离和干流入口处到下一干流观测站距离共同决定, 根据图1各观测地点的相对位置关系确定。对于各支流观测点的 S_i , 根据假设(1)(2)(5), 结合图示位置关系, 则有

$$\begin{aligned} S_2 - S_1 &= S_8 + S_9 + S_{10}, & S_4 - S_3 &= S_{12} + S_{13}, & S_5 - S_4 &= S_{11} + S_{14}, \\ S_6 - S_5 &= S_{15} + S_{16}, & S_8 &= S_9, & S_{12} &= S_{13}, \\ S_{11} &= S_{14}, & S_{15} &= S_{16}, & S_{10} &= S_8 + S_9. \end{aligned}$$

结合 $S_i (i = 1, 2, \dots, 7)$ 的数值解得 $S_8 - S_{16}$, 但 S_{17} 无法定量解得, 考虑其它支流的水流横截面积情况, 不妨令 $S_{17} = S_{10}$ 。

确定了所有观测点的 D_i 和 S_i , 利用(7)式和(6)式可得 $\Psi = 0.4331$, 水质为良好。

2 长江污染源的判定

考虑流体力学的一个基本理论—一个反应扩散方程, 我们将从该方程展开讨论

$$\frac{\partial w}{\partial t} + v \frac{\partial w}{\partial x} = E \frac{\partial^2 w}{\partial x^2} + f(x, t) - \lambda w, \quad (9)$$

其中, w 为污染物质在长江内浓度; v 为该段河流水流速度; f 为外界输入的污染物浓度, 根据现有数据资料, 我们不妨先假设, 在任意一个分段内 f 为常数, 即 f 整个是一个分段的常数函数; E 为扩散系数, λ 为降解系数, 介于 $0.1 \sim 0.5$ 之间。

首先确定方程中速度 v , 即每一小段的水流速度 $\overline{v_i^n}$ 。取

$$\overline{v_i^n} = \overline{v} + \theta_i(v_i^n - \overline{v}) + (1 - \theta_i)(v_{i+1}^n - \overline{v}) \quad (n = 1, 2, \dots, 13; \quad i = 1, 2, \dots, 6), \quad (10)$$

其中 \overline{v} 为整条长江流域的平均流速 $\overline{v} = \frac{1}{7} \sum_{i=1}^7 v_i$, $\theta_i(x)$ 为权重且 $0 \leq \theta_i(x) \leq 1$, 由于沿长江往下, 水流速度在逐渐减少, 在上游速度比较大时可以用后一个点的流速决定该流水段的水流速度, 到下游时可以将权重逐渐转移到了 v_i 上, 由此我们取 $\theta_i(x) = x_{i+1}/3251$, 其中3251是干流上站点之间的总间隔。依据上述方程(10)可以依次解出 $\overline{v_i^n}$ 。

2.1 差分方程反演模型的建立与求解

由[3]得, 扩散系数取 $13 \text{ Km}^2/\text{天}$ 。

对微分方程两边积分, 积分区间为 (x_{i-1}, x_i) , 其中 $i = 2, 3, \dots, 7$ 。

$$\frac{\partial}{\partial t} \int_{x_{i-1}}^{x_i} w dx + v(w_i - w_{i-1}) = E \left(\frac{\partial w}{\partial x} \Big|_{x_i} - \frac{\partial w}{\partial x} \Big|_{x_{i-1}} \right) - \lambda \int_{x_{i-1}}^{x_i} w dx + f_{i-1}^n \Delta x_{i-1}$$

其中 $\Delta x_{i-1} = x_i - x_{i-1}$ 。对上式进行处理, 含积分项利用梯形公式, 含偏导数项利用一阶向前差商, $n = 13$ 时的时间偏导的差分用向后差商, $i = 7$ 时对距离偏导的差分用向后差商, 得到 f_i^n 的反演式

$$f_{i-1}^n = \frac{(w_i^{n+1} + w_{i-1}^{n+1}) - (w_i^n + w_{i-1}^n)}{2\Delta t_n} + \frac{v}{\Delta x_{i-1}}(w_i^n - w_{i-1}^n) - \frac{E}{\Delta x_{i-1}} \left(\frac{w_{i+1}^n - w_i^n}{\Delta x_i} - \frac{w_i^n - w_{i-1}^n}{\Delta x_{i-1}} \right) + \frac{\lambda}{2}(w_i^n - w_{i-1}^n), \quad (11)$$

其中 $n = 1, 2, \dots, 12; i = 1, 2, \dots, 6$ 。

根据表达式(11), 利用 Matlab 编程代入数据分别求解出污染物在各段流域的 f 的值, 进行比较后我们可以得到在一年多时间里, 主要受高锰酸钾盐污染的河段有第三段(即湖北宜昌南津关到湖南岳阳城陵矶), 主要受氨氮污染的河段有第一段(即四川攀枝花到重庆朱沱)和第三段。

2.2 微分方程反演模型对污染源的判定

对差分方程中的扩散项 $E \frac{\partial^2 w}{\partial x^2}$ 及浓度随时间变化率 $\frac{\partial w}{\partial t}$ 进行数据分析及其定性分析, 容易发现这两项对整个等式影响很小, 可以忽略不计, 因此可将方程式(9)简化为常微分方程

$$v \frac{dw}{dx} = f(x) - \lambda w. \quad (12)$$

根据17个观察点在长江流域的分布情况, 把支流也看作是污染源, 即我们反演出的将浓度随时间的变化率项 f 中会有支流的影响, 将干流的七个观察点相邻两点之间作为一小段。假设 CD 两点之间的距离为 A , 以某一种污染物质为例, 假设该物质在两点的浓度分别为: w_0, w_A , 以这三个数为初始条件解出 $w(x)$ 以及该段的污染源排出的污染物浓度 f , 其数学模型为

$$\begin{cases} v \frac{dw}{dx} = f(x) - \lambda w, \\ w|_0 = w_0, \\ w|_A = w_A. \end{cases} \quad (13)$$

由该微分方程积分得到长江各段中 f 的反演公式

$$f = \frac{\lambda}{1 - e^{-\frac{\lambda}{v}A}}(w_A - w_0 e^{-\frac{\lambda}{v}A}).$$

利用 Matlab 编程代入数据分别求解出污染物在各段流域的 f 的值, 进行比较后我们可以得到在一年多时间里, 主要受高锰酸钾盐污染的河段为湖北宜昌南津关到湖南岳阳城陵矶段, 主要受氨氮污染的河段为四川攀枝花到重庆朱沱段和湖北宜昌南津关到湖南岳阳城陵矶段。

2.3 含支流的微分方程反演模型

考虑两站点之间出现支流情况, 根据地图描述的相邻干流站点间出现支流数量将这类情况分为两种, 并可以依据地图近似得出支流入口处距相邻两干流站点的距离(参考问题I):

1) 两站点间只有一个支点情况

当支流汇入长江时, 支流入口处干流上的点 E 的浓度发生突变, 假设在 E 点处突变后三处的单位时间水流量分别为 p_1, p_2, p_3 , 则有

$$w_{21} = \frac{w_{11}p_1 + w_3p_3}{p_1 + p_3}. \quad (14)$$

由 C 到 E 、 C 到 D 点, 两段距离可以依据地图近似测量, 分别记为 E 、 D , 根据两点间无支流情况的微分方程式(13)和方程式(14)得出含一个支流的 f 反演公式。

2) 两站点间有两个支点情况

记干流上的点 E 、 F 、 D 与段始点 C 的距离分别用 E 、 F 、 D 表示。可以将两支流情况转化为两个单支流的组合, 根据微分方程式(13)和方程(14), 与单支流相比仅增加一个未知数和一个方程式, 得出两个支流的 f 反演公式。

综合三种情况下 f 表达式, 得到整个长江流域六段的 f 反演公式如下:

① 没有支流

$$f = \frac{\lambda}{1 - e^{-\frac{\lambda}{v}A}}(w_A - w_0 e^{-\frac{\lambda}{v}A});$$

② 有一个支流

$$f = \frac{\lambda[w_1 p_1 e^{-\frac{\lambda}{v}E} - w_2(p_1 + p_3)e^{\frac{\lambda}{v}(D-E)} + w_3 p_3]}{p_1 e^{-\frac{\lambda}{v}E} - (p_1 + p_3)e^{\frac{\lambda}{v}(D-E)} + p_3};$$

③ 有两个支流

$$f = \lambda \frac{w_1 p_1 + w_3 p_3 e^{\frac{\lambda}{v}E} + w_4 p_4 e^{\frac{\lambda}{v}F} - w_2(p_1 + p_3 + p_4)e^{\frac{\lambda}{v}D}}{p_1 + p_3 e^{\frac{\lambda}{v}E} + p_4 e^{\frac{\lambda}{v}F} - (p_1 + p_3 + p_4)e^{\frac{\lambda}{v}D}}.$$

根据地图描述的17个站点的相对位置, 可以确定任意两干流站点间需要哪一类计算, 即确定该段水域使用的 f 的表达式。然后分别根据 f 的表达式, 利用 Matlab 编程求出各污染物的 f 的结果。综合该六段水域在近一年多内受两种污染物污染的情况, 可以得出主要受高锰酸钾盐污染的河段为湖北宜昌南津关到湖南岳阳城陵矶段, 主要受氨氮污染的河段为四川攀枝花到重庆朱沱段和湖北宜昌南津关到湖南岳阳城陵矶段。

单独从三个模型的数据反映的结果进行分析, 这三个模型得出的结论是基本一致的。现综合三个模型得出的结果, 将每一段中三个模型的各种污染物的年度和求和进行分析可以得到, 在这六段流域中高锰酸钾盐污染源主要分布在湖北宜昌南津关到湖南岳阳城陵矶段; 氨氮污染源主要分布在四川攀枝花到重庆朱沱段和湖北宜昌南津关到湖南岳阳城陵矶段, 两者 f 值的差距比较小, 但与其它四段相比差距较大。

3 回归模型对水质的预测分析

3.1 回归模型对问题的求解^[1]

可以通过回归模型对水质进行预测分析。考虑水文年长江流域水质变化, 记近10年的观测数据为 $(t_i, y_{ij})(i = 1, 2, \dots, 10, j = 1, 2, \dots, 6)$, 其中 y_{ij} 表示第 i 年第 j 类水所占百分比, 可以根据附件4的观测数据找到。分别令 $j = 1, 2, \dots, 6$ 设六类水百分比 y_j 随年份 t 变化的线性方程为

$$y_j = A_j t + B_j, \quad (17)$$

其中 (A_j, B_j) 是参数。

1) 利用最小二乘法进行参数估计, 确定 A_j 和 B_j , 得到线性回归方程;

2) 给定显著性水平 $\alpha = 0.05$, 对回归系数进行 F 检验, 确定回归分析的效果。效果好, 回归方程确定; 效果不好, 转向步骤(3)。

3) 剔除数据中的偶然因素, 保持长远趋势利用三项移动平均法对数据进行平滑处理, 令

$$y_i = \frac{y_{i-1} + y_i + y_{i+1}}{3} \quad (i = 1, 2, \dots, n-1),$$

得到一组新的数据, 转向步骤1)。

经过三次循环后, $j = 4$ 时的回归效果仍然很差, 因此认为没有理想的线性回归方程, 来描述 IV 类水的变化趋势。由于六类水的总百分比之和为100, 定义 $y_4 = 100 - y_1 - y_2 - y_3 - y_5 - y_6$ 。考虑到各类水质百分比是正值, 最终得到的线性回归方程为

$$y_j = \max(0, y_j) \quad j = 1, 2, \dots, 6.$$

令 $t = 11, 12, \dots, 20$, 利用 y_j 的表达式即可预测未来10年水质情况, 但是为了保证六类水质的总的百分比之和为100, 需要对预测值进行标准化处理, 即将预测值乘以系数 $\lambda = 100 / \sum_{j=1}^6 y_j$ 。

3.2 模型的改进和预测结果

上述回归分析中 II、III、IV 类水质在首次回归分析中不满足对回归系数的假设检验, 而且在预测中出现了较多的0值, 这是不符合实际情况的。对于水质污染发展趋势的预测分析, 我们主要考虑的是 IV、V、劣 V 类水质百分比的变化趋势, 为此, 对模型作如下改进: 1) 将可饮用水, 即 I、II、III 类水综合考虑; 2) 由于 IV 类水没有符合要求的线性回归方程, 同样用取余的方法求得。

将观测数据 (t_i, y_{ij}) 的可饮用水百分比进行合并记新的数据为 (t_i, z_{ij}) , $(j = 1, 4, 5, 6)$, z_{ij} 依次表示第 i 年可饮用水、IV 类水、V 类水、劣 V 类水所占百分比。首先对 (t_i, z_{ij}) , $(j = 1, 4, 5, 6)$ 进行平滑处理, 然后用最小二乘法求出回归方程并进行 F 检验, 得到

$$\begin{aligned} z_1 &= 1.9056t + 88.067, & z_5 &= 0.41627t + 2.08937, \\ z_6 &= 1.2496t - 0.13571, & z_4 &= 100 - z_1 - z_5 - z_6. \end{aligned}$$

令 $t = 11, 12, \dots, 20$, 根据根据前10年数据回归得到的 z_j 的表达式对未来10年水质污染发展进行预测, 有以下结果: 可饮用水呈逐年递减的趋势, 10年以后将低于50而且劣 V 类递增的速率较快。长江流域水质呈整体下降的趋势, 需要即时治理。

4 基于回归模型的预测控制

4.1 理论分析

考虑影响污水处理量的因素: 污水处理量 = 污水排放量 - 长江允许排污量。其中, 污水排放量可以由预测得到, 长江允许排污量与干流 IV + V 类水百分比和劣 V 类水百分比有关。

记长江总流量为 L (亿立方米)、污水排放总量为 M (亿吨)、相对排污量为 $H = \frac{M}{L}$ (吨/立方米) 和 IV + V 类水百分比为 N_1 、劣 V 类水百分比为 N_2 。从附表4中可以得到近10年的 L 、 M 、 N_1 和 N_2 的观测数据。

1) 设相对排污量随时间的变化关系式为 $H = g(t)$;

2) 同时考虑 L 和 M 对 N_1 、 N_2 的影响, 设 IV + V 类水百分比为 N_1 随相对排污量 H 的变化关系式为 $N_1 = f_1(H)$, 劣 V 类水百分比为 N_2 随相对排污量 H 的变化关系式为 $N_2 = f_2(H)$;

3) 根据近10年的观测数据回归分析得到 g 、 f_1 和 f_2 的线性表达式;

4) 如果 $f_1(H)$ 和 $f_2(H)$ 已知, 如果同时满足 $N_1 \leq 20$ 和 $N_2 = 0$ 的取值要求, 给定 t 值, 则需要处理的相对污水量为

$$\Delta H_t = \max\{g(t) - f_1^{-1}(20), g(t) - f_2^{-1}(0), 0\}. \quad (18)$$

则污水处理量

$$\Delta M_t = \Delta H_t \times \bar{L}, \quad (19)$$

其中 $\bar{L} = \frac{1}{10} \sum_{k=1}^{10} L_k$ 为10年长江总流量的平均值。

4.2 回归模型对污水处理量的预测

根据4.1的理论分析预测排污量。首先按照问题三的回归分析模型求解 $H = g(t)$ 、 $N_1 = f_1(H)$ 、 $N_2 = f_2(H)$ 并进行 F 检验。在求解回归方程的过程中, 为了使结果更符合普遍情况, 适当剔除了一些点, 在求 $N_1 = f_1(H)$ 时, 剔除了2003年的点(该点是一个突变量), 在求 $N_2 = f_2(H)$ 时, 剔除了1998年的点(该点的自变量太小), 得到了回归效果较好的方程

$$H = g(t) = 0.0014t + 0.015,$$

$$N_1 = f_1(H) = 1891.9H - 26.6,$$

$$N_2 = f_2(H) = 702.7222H - 13.817.$$

令 $t = 11, \dots, 20$, 由(18)式和(19)式可以得到未来10年每年要处理的污水量 ΔM_t 。

4.3 二元线性回归模型的建立与求解[2]

在上面对 $N_1 = f_1(H)$ 和 $N_2 = f_2(H)$ 进行回归分析的过程中, 都剔除了特殊点, 而且在对回归方程 $N_2 = f_2(H) = 702.7222H - 13.817$ 进行 F 检验时发现, F 检验刚好满足要求, 说明回归的效果并不是太好。对原始数据进行观察, 由于劣V类所占百分比出现了大量的0值, 仅三个非零点, 所以考虑将IV、V类和劣V类综合考虑, 建立二元线性回归模型, 这样可以减少0值的影响。

1) 设相对排污量 H 随 N_1 和 N_2 的变化关系式为 $H = f(N_1, N_2) = AN_1 + BN_2 + C$, 利用最小二乘法求得回归方程

$$f(N_1, N_2) = 0.002N_1 + 0.005N_2 + 0.0187; \quad (20)$$

2) 对回归模型进行 F 检验: 给定 $\alpha = 0.05$, 查表有 $F_{1-\alpha}(2, 7) = 4.74$, 计算得 $F = 8.0168$, 满足 $F > F_{1-\alpha}(2, 7)$, 回归效果较好;

3) 考虑 $N_1 \leq 20, N_2 = 0$, 得到允许相对排污量的临界值为 $H_0 = f^{-1}(20, 0) = 0.0227$, 给定 t 值, 则需要处理的相对污水量为

$$\Delta H_t = \max\{h(t) - 0.0227, 0\}; \quad (21)$$

4) 令 $t = 11, \dots, 20$, 结合(19)式得到未来10年需要处理的污水量, 结果见表2。

表 2: 未来10年每年需处理的污水量

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
$\Delta M(\text{亿吨})$	75.195	89.047	102.9	116.75	130.6	143.46	157.32	171.17	185.02	198.87

对于两种方法对污水处理量的预测, 结合已知的原始数据进行分析: 对于方法一, 有 $H_0 = \min\{f_1^{-1}(20), f_2^{-1}(0)\} = 0.0197$, 然而对于1997年和1999年来讲, 虽然相对排污量大于0.0197, 但是IV、V类百分比小于20, 劣V类百分比为0, 不需要进行污水处理; 对于方法二, $H_0 = 0.0227$, 观察已知数据, 当相对排污量大于0.0227时, IV、V类百分比均大于20, 当相对排污量小于0.0227时, IV、V类百分比均小于20, 对于劣V类百分比, 临界值 $H_0 = 0.0227$ 恰好出现在百分比不为0的前面附近, 这是符合实际情况的。因此认为用二元线性回归对污水处理量的预测结果更优, 最终预测结果如表2作示。

参考文献:

- [1] 傅鹏, 刘琼荪, 何中市. 数学实验[M]. 北京: 科学出版社, 2003
- [2] 梁之舜, 邓集贤, 杨维权, 司徒荣, 邓永录. 概率论与数理统计[M]. 北京: 高等教育出版社, 2003
- [3] 徐祖信, 廖振良, 张锦平. 基于数学模型的苏州河上游和支流水质对干流水质的影响分析[J]. 水动力学研究与发展, A辑, 2004, 19(6): 738

The Estimation and Forecast of the Yangtze River's Water Quality

LI Yun-feng, WANG Yong, YANG Lin

Advisor: SHEN Jin-ren

(The PLA University of Science and Technology, 210007)

Abstract: In this paper we integrate the original data and get a new estimate function $f(x, t)$. After predigesting the positions of all the observation stations along the Yangtze River, we obtain the estimate function $f(x, t)$, and find that the water quality is good, and the mostly contamination is ammonia and nitrogen.

By establishing a reactive diffusion equation on the concentration of the pollute, we deduce the function of polluting source in three ways, then analyse its value and find the areas where potassium permanganate and ammonia and nitrogen distribute.

For forecasting the tendency of the water quality, we establish a regress model and do F -test on the coefficients in the regress model, then conclude that if we don't adopt any methods to control contaminations, the quantity of drinking water would decline year by year, and the gross quantity of Yangtze River's drinking water will be half of it ten years later. So we can use bi-element linear regress model to consider all kinds of water which is not able to drink and make following conclusion: in ten years if we make the proportion of nondrinking water to be less than 20, then the quantity of sewage must be treated with more than 75.195 hundred million tons per year.

Keywords: normalization; evaluation function of water quality; reaction diffusion equation; Regression analysis