

文章编号:1005-3085(2005)07-0065-11

长江水质综合评价与预测的数学模型

韩中庚

(解放军信息工程大学信息工程学院, 郑州 450002)

摘 要: 本文针对2005年“高教社杯”全国大学生数学建模竞赛的A题“长江水质评价和预测”问题, 首先概括地介绍了这个问题的立意与背景, 然后给出了解决这个问题的一种可行的解决方案及结果, 最后根据评卷情况、对评卷要点、问题的解决方法和答卷中存在的问题做了综合评述。

关键词: 长江水质; 水质类型; 综合评价与预测; 水质模型

分类号: AMS(2000) 62P12

中图分类号: TB114

文献标识码: A

1 问题的立意与背景

水是人类赖以生存的资源, 保护水资源等于保护我们自己, 尤其是对大江大河的水资源的保护和治理是重中之重。根据环保部门公布的数据, 我国现有的水资源符合饮用水标准(I类III类水)的只有30%, 其余的70%都不能作饮用水, 部分河流已被污染成了废水河。长江是我国第一、世界第三大河流, 长江水的污染程度日趋严重, 正面临六大危机: “森林覆盖率下降, 泥沙含量增加, 生态环境急剧恶化; 枯水期不断提前; 水质恶化, 危及城市饮用水; 物种受到威胁, 珍稀水生生物日益灭绝; 固体废物严重污染, 威胁水闸与电厂安全; 湿地面积缩减, 水的天然自洁功能日益丧失”。为此, 长江水污染问题已引起了政府和有关专家的高度重视, 国家环保局于2005年初掀起了一场“环保风暴”, 一次刮停对环境污染严重的发电厂建设项目30多项, 其中长江流域占了22项。国家主要领导人也强调指出: “要彻底改变以牺牲环境、破坏资源为代价的粗放型增长方式; 不能以牺牲环境为代价去换取一时的经济增长; 不能以眼前发展损害长远利益; 不能用局部发展损害全局利益。”在2005年3月召开的“两会”上, 许多人大代表和政协委员齐声呼吁: “以人为本, 建设文明和谐社会, 改善人与自然环境, 减少污染”。

2004年10月, 由全国政协委员、中国发展研究院执行院长章琦教授发起, 并联合组织几十名专家学者成立了“保护长江万里行”考察团, 从长江上游宜宾到下游上海, 对沿线21个重点城市做了实地考察, 揭示了一幅长江污染的真实画面, 其污染程度让人触目惊心。为此, 专家们提出“若不及时拯救, 长江生态10年内将濒临崩溃。”同时, 专家们也上书国务院, 并发出了“拿什么拯救癌变长江?”的呼唤。

正是在这样的一种背景下, 我们提出了“长江水质评价与预测”问题。目前, 长江的污染究竟达到了什么程度、主要的污染源在哪里、未来的发展趋势究竟会如何? 10年后的长江究竟会变成什么样? 会不会也像现在的淮河、海河一样变成一条最大的污水河? 如果是这样, 现在国家花巨资建设的“南水北调工程”岂不是毫无利用价值了! 为此, 我们围绕这个问题查阅了大量的相关文献资料, 并先后咨询了中国发展研究院执行院长章琦教授, 华东师范大学的陆健健教授, 长江水利管理委员会(武汉)信息中心、《长江年鉴》编辑部、水利部中国水环境研究院信息中心和水质研究所、华北水利水电学院等单位, 得到了有关专家和支持与帮助, 经过多方努力, 多种途径搜集整理得到了所需要的珍贵数据资料。同时, 我们也了解到,

目前还没有人用完全定量的方法来研究相关问题。根据多年的统计资料显示,长江流域的面积占中国版图的54%,GDP 占全国的54%;2002年长江流域的排污总量为256亿吨,即平均每秒812吨,几乎都占了全国排污总量的一半之多,2003年达到了270多亿吨,比上个世纪80年代增加了一倍多。在长江流域生活着4亿人口;如果考虑到南水北调工程的影响,那么长江水质的好坏将影响到8亿人口的生活质量,为此,研究解决长江污染的问题是多么重要,这也是我们研究这个问题的意义所在。

2 问题的解决思路

根据这个问题的实际背景和现有的观测数据,首先依据近两年17个观测点的数据对相应地区的水质情况做定量综合评价与分析;然后依据这些地点的相对地理位置、水流量和水质数据,利用简化的一维水质模型推算出相应的排污量,从而可以确定出长江干流的主要污染源所在的区段;再根据长江过去10年的总体水质检测分类数据,利用灰色系统理论和回归分析等方法,对未来长江水质发展趋势进行预测,并对可能控制水质的条件进行研究。

问题1) 按国家标准(GB 3838-2002)的规定,关于地表水的评价指标共有24项,但对水质污染最主要的是四项:PH值、溶解氧(DO)、高锰酸盐指数(CODMn)和氨氮(NH₃-N),按国标将水质可分为I类、II类、III类、IV类、V类、劣V类共六个类别,每一个类别对每一项指标都有相应的标准值(区间),只要有一项指标达到高类别的标准就算是高类别的水质,所以实际中不同类别的水质有很大的差别,而且同一类别的水在污染物的含量上也有一定的差别。为此,由过去两年多17个观测站的水质数据做综合评价时,要充分考虑这些指标不同类别的“质的差异”和同类别的“量的差异”。在这里,首先通过极差等变换将相关数据做标准化处理,然后利用动态加权法合理地构造综合评价指标函数,使能充分地体现水质的类别差异和同类别水的数量差异。最后依据综合指标值的大小对各地区的水质状况做出分析评价。

问题2) 根据干流各观测站的水质数据和相应站点的位置关系,考虑到上游的污水会对下游的水质造成一定的影响,同时江河本身都有一定的自洁能力。一般说来自洁规律与江河的水流量、流速、水流断面、水流距离等参数有关,通常上游的水质对下游的影响服从一维水质模型。为了简化计算,不妨假设在一定的时间内流速是均匀的,水流断面变化不大,则可将其一维水质模型简化为简单的常微分方程。由此可以推算出上游的污水对下游的影响程度,从而可以计算出干流各个区段的排污量,即可确定主要的污染源所在的地区。

问题3) 根据过去近10年长江的总体水污染状况的检测数据,可以看出长江总体水污染的严重程度呈现快速增长的趋势,主要是年排污总量的增加,在总水流量变化不大的情况下,使得污染河段比例的增加,即每年污染情况主要与当年的排污量和总水流量等因素有关。为此,首先可以根据过去10年排污量,利用灰色预测方法对未来的年排污量做出预测,然后利用回归分析方法确定出可饮用(或不可饮用)水的比例与总排污量和总水流量的关系式。最后根据总排污量的增长趋势来推断出可饮用(或不可饮用)水比例的变化趋势,从而可以预测出未来10年长江水质的变化情况。

问题4) 用问题3)类似的方法,首先利用过去10年的相关数据确定出的不可饮用水、劣V类水的比例与总排污量和总水流量的关系式,然后根据题目要求的条件可以求出未来10年的污水处理量。

3 长江水质的综合评价模型

3.1 数据的标准化处理

首先要对所给的水质指标进行统一的无量纲化标准处理, 使各项指标具有可比性。设四项水质指标溶解氧、高锰酸盐指数、氨氮和 PH 值的指标值分别为 x_1, x_2, x_3 和 x_4 。

1) 指标溶解氧的处理 为了与其它指标的度量标准一致性, 首先将指标数据作极小化处理, 即令倒数变换 $x'_1 = \frac{1}{x_1}$, 然后通过极差变换 $x''_1 = \frac{x'_1}{0.5}$ 将其数据标准化, 对应的分类区间变为

$$(a_1^{(1)}, b_1^{(1)}], (a_2^{(1)}, b_2^{(1)}], (a_3^{(1)}, b_3^{(1)}], (a_4^{(1)}, b_4^{(1)}], (a_5^{(1)}, b_5^{(1)}], (a_6^{(1)}, b_6^{(1)}).$$

2) 指标高锰酸盐指数的处理 对所有高锰酸盐指标数据作极差处理, 将其数据标准化, 即令 $x'_2 = \frac{x_2}{15}$, 对应的分类区间随之变为

$$(a_1^{(2)}, b_1^{(2)}], (a_2^{(2)}, b_2^{(2)}], (a_3^{(2)}, b_3^{(2)}], (a_4^{(2)}, b_4^{(2)}], (a_5^{(2)}, b_5^{(2)}], (a_6^{(2)}, b_6^{(2)}).$$

3) 指标氨氮的处理 对所有氨氮指标数据作极差处理, 将其数据标准化, 即令 $x'_3 = \frac{x_3}{2}$, 对应的分类区间随之变为

$$(a_1^{(3)}, b_1^{(3)}], (a_2^{(3)}, b_2^{(3)}], (a_3^{(3)}, b_3^{(3)}], (a_4^{(3)}, b_4^{(3)}], (a_5^{(3)}, b_5^{(3)}], (a_6^{(3)}, b_6^{(3)}).$$

4) 指标 PH 值的处理 PH 值(酸碱度)的大小反映出水质呈酸碱性的程度。通常的水生物都适应于中性水质, 即酸碱度的平衡值(PH 值略大于 7), 在这里不妨取正常值的中值 7.5。当 $PH < 7.5$ 时水质偏碱性, 当 $PH > 7.5$ 时偏酸性, 而偏离值越大水质就越坏。为此, 对所有的 PH 值指标数据作均值差处理, 即令 $x'_4 = \frac{2}{3}|x_4 - 7.5|$, 则将其数据标准化。

3.2 综合评价指标的确定方法

考虑到一个地区的污染指标的变化不仅与其所属的类型有关, 而且即便是同属一个类型也有一定的数值差异。为此, 在确定综合评价指标时, 既要能体现同类型的指标数量差异, 也要能体现不同类型指标之间的差异, 而且更要能体现不同类型等级差的差异。于是, 在这里采用动态加权法来确定相应的综合评价指标。根据实际不妨取动态加权函数为偏大型正态分布函数, 即

$$w_i(x) = \begin{cases} 0, & x \leq \alpha_i \\ 1 - e^{-\left(\frac{x - \alpha_i}{\sigma_i}\right)^2}, & x \geq \alpha_i \end{cases} \quad (i = 1, 2, 3), \quad (1)$$

其中 α_i 取指标 x_i 的 I 类水标准区间的中值, 即 $\alpha_i = \frac{b_i^{(i)} - a_i^{(i)}}{2}$, σ_i 由 $w_i(a_4^{(i)}) = 0.9$ ($i = 1, 2, 3$) 确定。由实际数据经计算可得 $\alpha_1 = 0.1333, \alpha_2 = 0.0667, \alpha_3 = 0.0375, \sigma_1 = 0.1757, \sigma_2 = 0.2197, \sigma_3 = 0.3048$, 则代入(1)式可以得到 DO、CODMn 和 NH₃-N 三项指标的权值函数。考虑到差异较大的是前三项指标, 以及指标 PH 值的特殊性, 这里取前三项指标的综合影响权值为 0.8, 而 PH 值的影响权值取 0.2。因此, 某地区某一时间的水质综合评价指标定义为

$$X = 0.8 \sum_{i=1}^3 w_i(x_i)x_i + 0.2x_4. \quad (2)$$

根据 2003 年 6 月到 2005 年 9 月 17 个主要观测站的 28 组实际检测数据, 经计算可得各观测站在区段的水质综合评价指标值, 即可得到一个 17×28 阶的综合评价矩阵 $(X_{ij})_{17 \times 28}$ 。

3.3 各地区水质的综合排序与评价

由 17 个观测点 28 个月的水质综合评价指标 X_{ij} ($i = 1, 2, \dots, 17; j = 1, 2, \dots, 28$), 根据其大小(即污染的程度)进行排序, 数值越大水质越差。由此可得反映 17 个观测点(地区)水质污

染程度的28个排序结果,利用决策论中 Borda 数法^[1]来确定综合排序方案。记第 j 个月的排序方案中排在第 i 个站点 S_i 后面的站点个数为 $B_j(S_i)$, 则站点 S_i 的 Borda 数为

$$B(S_i) = \sum_{j=1}^{28} B_j(S_i), \quad (i = 1, 2, \dots, 17).$$

经计算可得到各站点的 Borda 数及总排序结果如表1所示。

表1: 各观测点的 Borda 数及水污染情况总排序

| 观测站点 | S_1 | S_2 | S_3 | S_4 | S_5 | S_6 | S_7 | S_8 | S_9 | S_{10} | S_{11} | S_{12} | S_{13} | S_{14} | S_{15} | S_{16} | S_{17} |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| Borda数 | 203 | 136 | 143 | 234 | 106 | 139 | 138 | 378 | 232 | 271 | 60 | 357 | 277 | 264 | 438 | 214 | 217 |
| 总排序 | 11 | 15 | 12 | 7 | 16 | 13 | 14 | 2 | 8 | 5 | 17 | 3 | 4 | 6 | 1 | 10 | 9 |

由表1可以看出,各观测站(地区)所在的江段的水质污染的情况,水质最差的是观测点 S_{15} , 即是江西南昌赣江鄱阳湖入口地区;其次是观测点 S_8 , 即四川乐山岷江与大渡河的汇合地区;第三位的是 S_{12} , 即湖南长沙湘江洞庭湖地区;干流水质最差的是湖南岳阳段(S_4), 主要污染可能是来自于洞庭湖。干流水质最好的区段是江西九江(鄂赣交界)段(S_5), 支流水质最好的是湖北丹江口水库(S_{11})。

4 长江干流主要污染源的确定模型

针对问题2), 分析确定长江干流主要的污染物来自哪些地区。由于一个江段的水质污染, 主要来自本地区的污水和上游扩散下来污水两个部分的合成。一般说来, 对于某一江段内的水质情况与该段内的排污量和上游的水质有关, 在这里我们用排污速率(即每秒钟排污的含量)的大小来判断其排污量的多少。根据长江干流上的七个主要观测站点, 将其分为六段, 逐段分析其排污情况, 即可以找出主要污染物的污染源所在的区域。

4.1 一维水质模型

假设长江干流中的污染物的分布浓度为 $C(\text{mg/L})$, 各河段断面为均匀的, 平均流速记为 $u(\text{m/s})$, 不考虑扩散, 且自净能力使污染物的浓度也不随时间变化, 则 C 满足一维水质模型 $u \frac{dC}{dx} + kC = 0$, 其中 k 为污染物的降解系数 ($1/\text{s}$)。如果初值条件为 $C(0) = C_0$, 则可求得一维水质模型的解, 即污染物的浓度随着水流的流动自然降解的规律为 $C = C_0 e^{-k \frac{x}{u}}$ 。

对于长江干流上任一江段 AB, 即起点为 A, 终点为 B, 距离为 $d(\text{m})$, 不妨假设段内有 n 个排污口(包括支流入口和直排口), 第 i 个排污口的流量、平均流速、污染物的浓度分别为 q_i, u_i, c_i , 而用 Q_i, U_i, C_i 分别表示该江段干流的水流量、流速和污染物浓度。则

$$C_i = \frac{q_i c_i + Q_{i-1} f(C_{i-1})}{Q_{i-2q_i}}, \quad Q_i = Q_{i-1} + q_i \quad (i = 1, 2, \dots, n+1), \quad (3)$$

其中 $f(C_{i-1}) = C_{i-1} e^{-k \frac{x_{i-1}}{U_{i-1}}}$ 。如果已知起始点 A 和终点 B 的水流量、流速、污染物浓度分别为 q_0, u_0, c_0 和 $q_{n+1}, u_{n+1}, c_{n+1}$, 则在江段 AB 内的总排污量为 $w_{AB} = \sum_{i=1}^n c_i q_i$ 。即为实际上每秒钟排出的污染物的总量(排污速率)。实际中, 因为不知道一个江段内排污口的个数和相应的排污量, 要精确计算总排污量是困难的。为此, 我们只须计算一个江段内可能的最大总排污量(上界)和最小排污量(下界), 分别记为 w_{\max} 和 w_{\min} 。事实上, 当所有排污都集中在江段的源头时, 对该江段的水质影响最大。相反的, 当所有的排污都集中在江段的终点时, 对该江段的水质影响最小, 据此来确定该江段的总排污量的上界和下界值。

4.2 污染物排放量的确定方法

1) 排污量的上界值 假设江段 AB 内的所有排污都集中在 A 点(源头)处,即在 A 点均匀混合后,经过 AB 段内的降解到 B 点,则 $C_B = C_1 e^{-k \frac{x}{U_B}}$, 于是 $C_1 = C_B e^{k \frac{x}{U_B}}$ 。又根据(3)式可得

$$C_1 = \frac{q_1 c_1 + Q_0 C_0}{Q_0 + q_1} = \frac{q_1 c_1 + Q_A C_A}{Q_B},$$

故有该江段内总排污量的上界值为

$$w_{\max} = q_1 c_1 = C_1 Q_B - Q_A C_A = Q_B C_B e^{k \frac{x}{U_B}} - Q_A C_A (g/s). \quad (4)$$

2) 排污量的下界值 假设江段 AB 内的所有排污都集中在 B 点(段末)处,类似地可得 $C_1 = C_B = \frac{q_1 c_1 + Q_A C_A e^{-k \frac{x}{U_A}}}{Q_B}$, 故有该江段内总排污量的下界值为

$$w_{\min} = q_1 c_1 = C_B Q_B - C_A Q_A e^{-k \frac{x}{U_B}} (g/s). \quad (5)$$

对于任意一个江段 AB, 由起始点和终点的污染浓度 C_A, C_B 、水流量 Q_A, Q_B 、流速 U_A, U_B 、距离 x (均已给定) 和降解系数 k , 则根据(4)式和(5)式就可以计算出该 AB 段的污染物总排放量的变化区间 $[w_{\min}, w_{\max}]$ 。

3) 平均相对排污量 根据所给数据, 对于每一个月每一江段都可以确定一个排污量变化区间, 记第 j 个月排污量的变化区间为 $[w_{\min}^{(j)}, w_{\max}^{(j)}] (j = 1, 2, \dots, 13)$ 。按月份取均值得 $\bar{w}_{\min} = \frac{1}{13} \sum_{j=1}^{13} w_{\min}^{(j)}, \bar{w}_{\max} = \frac{1}{13} \sum_{j=1}^{13} w_{\max}^{(j)}$, 则每一个江段都有确定的排污量区间 $[\bar{w}_{\min}, \bar{w}_{\max}]$ 。取中值 $w_{\text{med}} = \frac{1}{2}(\bar{w}_{\min} + \bar{w}_{\max})$, 即为一个江段13个月的平均排污量。如果该江段的距离总长为 $d(\text{km})$, 则一个江段每秒、每公里的排污量为 $w_0 = \frac{w_{\text{med}}}{d} (\text{kg/s} \cdot \text{km})$, 称其为平均相对排污量。对每一江段都有一个平均相对排污量指标, 它是一个可比性的指标, 由此指标的大小可以确定长江干流排量最大的区段, 即可以确定主要污染源。

4.3 长江干流主要污染源的确定方法

根据附件3中所给的“长江干流主要观测站点的基本数据”, 对于指标 CODMn 和 NH₃-N 的降解系数为 $k = 0.2(1/\text{天})$ 。按上述的方法分别计算可得结果及按排污速率大小排序如表2和表3所示。

表2: 指标 CODmn 的排放量及排序结果

| 江段 | $S_1 \sim S_2$ | $S_2 \sim S_3$ | $S_3 \sim S_4$ | $S_4 \sim S_5$ | $S_5 \sim S_6$ | $S_6 \sim S_7$ |
|------------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $[\bar{w}_{\min}, \bar{w}_{\max}]$ | [31.32, 49.03] | [30.07, 68.07] | [39.12, 78.72] | [20.29, 49.04] | [14.15, 18.58] | [23.16, 40.15] |
| 相对排污量 w_0 | 0.04229 | 0.06307 | 0.15 | 0.0693 | 0.0998 | 0.0682 |
| 排序 | 6 | 5 | 1 | 3 | 2 | 4 |

表3: 指标 NH₃-N 的排放量及排序结果

| 江段 | $S_1 \sim S_2$ | $S_2 \sim S_3$ | $S_3 \sim S_4$ | $S_4 \sim S_5$ | $S_5 \sim S_6$ | $S_6 \sim S_7$ |
|------------------------------------|----------------|----------------|-----------------|----------------|----------------|----------------|
| $[\bar{w}_{\min}, \bar{w}_{\max}]$ | [28.11, 63.12] | [30.47, 83.58] | [42.52, 113.09] | [22.52, 60.78] | [15.79, 22.05] | [8.47, 28.1] |
| 相对排污量 w_0 | 0.048 | 0.0733 | 0.197 | 0.0833 | 0.1154 | 0.0393 |
| 排序 | 5 | 4 | 1 | 3 | 2 | 6 |

从上面的结果可以看出, CODMn 和 NH₃-N 的主要污染源都在 $S_3 \sim S_4$ 段, 即在湖北宜昌到湖南岳阳之间的地区, 可能是来自于三峡水库下游和洞庭湖一带, 直观地分析此与实际数据完全相符。

5 长江水质的预测模型

要研究长江未来水质的总体变化情况,问题只需要预测10年后长江水是否还可以饮用,即I类、II类、III类水的比例总和为多少?

5.1 可饮用水量变化规律的预测

由过去10年长江流域的总排污量分别为(174, 179, 183, 189, 207, 234, 220.5, 256, 270, 285)(亿吨)和总水流量分别为(9205, 9513, 9171.26, 13127, 9513, 9924, 8892.8, 10210, 9980, 9405)(亿立方米),其变化规律可以视为时间(年) t 的函数,不妨分别记为 $w_1 = \phi_1(t)$ 和 $w_2 = \phi_2(t)$ 。因为每年各水质类型的变化主要与总排污量和水流量有关,为此以 $\phi_1(t), \phi_2(t)$ 为解释变量,可饮用水的比例总和为响应变量,利用过去的检测数据作多元线性回归,从而可以得到可饮用水的比例与 $\phi_1(t)$ 和 $\phi_2(t)$ 函数关系。即考虑一般的线性多元回归模型为

$$y = a\phi_1(t) + b\phi_2(t) + c \quad (6)$$

由过去10年可饮用水的比例观测值 $(\phi_1(t_k), \phi_2(t_k), y_k)(k = 1, 2, \dots, 10)$,用最小二乘法求得回归系数的估计值 (a, b, c) ,代入(6)式中就可以得到可饮用水的比例与排污量和水流量的关系式。事实上,根据过去10年中枯水期、丰水期和水文年的可饮用水比例分别可得回归系数 $(a_i^{(k)}, b_i^{(k)}, c_i^{(k)})(i, k = 1, 2, 3)$ 如表4所示。

表4: 回归模型的系数

| 水期 | 枯水期 | | | 丰水期 | | | 水文年 | | |
|-----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 系数 | $a_i^{(1)}$ | $b_i^{(1)}$ | $c_i^{(1)}$ | $a_i^{(2)}$ | $b_i^{(2)}$ | $c_i^{(2)}$ | $a_i^{(3)}$ | $b_i^{(3)}$ | $c_i^{(3)}$ |
| 全流域 | -0.1405 | 0.0018 | 89.0214 | -0.1907 | 0.0008 | 113.2105 | -0.1489 | 0.0016 | 96.8396 |
| 干流 | -0.3123 | 0.0036 | 112.0766 | -0.2428 | 0.0020 | 116.0260 | -0.1846 | 0.0039 | 85.1549 |
| 支流 | -0.0419 | 0.0005 | 76.4087 | -0.1480 | 0.0001 | 109.1528 | -0.1015 | 0.0002 | 97.4869 |

将其代入回归模型(6)式中就得到各个水期的全流域、干流和支流可饮用水的百分比变化规律:

$$y_i^{(k)} = a_i^{(k)} \phi_1(t) + b_i^{(k)} \phi_2(t) + c_i^{(k)} \quad (i, k = 1, 2, 3). \quad (7)$$

利用这个模型可以对未来的水质情况(可饮用水的比例)进行预测分析。

5.2 未来10年的总排污量预测

由于过去10年的长江流域的总排污量是总体增加的趋势,为此用灰色预测模型GM(1,1)对未来10年的总排污量做出预测。依次记过去10年的总排污量为基本序列,记为

$$W^{(0)} = (w^{(0)}(1), w^{(0)}(2), \dots, w^{(0)}(10)).$$

对 $W^{(0)}$ 作AGO序列和MEAN序列,求解GM(1,1)模型可以得到未来10年的排污总量的平均值为 $\bar{w} = 303.3804$ 亿吨。因主要是考虑10年后的情况,所以中间可以认为每年是均匀增长,这里平均年增长量为 $\Delta w = \bar{w} - 285 = 18.3804$ (亿吨),则未来10年的排污量分别为 $w_k = 285 + k\Delta w(k = 1, 2, \dots, 10)$,经计算可得未来10年的排污总量(单位为亿吨),即

$$W = (303, 322, 340, 359, 377, 395, 414, 432, 450, 469). \quad (8)$$

5.3 未来10年的水质变化规律预测

根据各水期的全流域、干流和支流可饮用水的百分比的变化模型(7), 由未来10年的排污总量(8)式和相应的水流总量就可以计算出未来10年的各水期的全流域、干流和支流可饮用水的百分比的变化情况。注意到水流量没有明显的变化规律, 在这里用过去10年的平均水流量9894.1亿立方米来表示未来10年的平均年水流量, 即 $\phi_2(t) = 9894.1$ 。经计算得未来10年相应可饮用水的比例如表5所示。

表5: 未来10年的各水期可饮用水的比例预测值

| | 年份 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|-----|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 枯水期 | 全流域 | 64.1100 | 61.4397 | 58.9100 | 56.2397 | 53.7099 | 51.1802 | 48.5099 | 45.9802 | 43.4505 | 40.7802 |
| | 干流 | 53.3414 | 47.4078 | 41.7865 | 35.8529 | 30.2316 | 24.6103 | 18.6767 | 13.0553 | 7.4340 | 1.5004 |
| | 支流 | 69.0851 | 68.2897 | 67.5362 | 66.7409 | 65.9874 | 65.2339 | 64.4385 | 63.6850 | 62.9315 | 62.1362 |
| 丰水期 | 全流域 | 63.5670 | 59.9443 | 56.5123 | 52.8897 | 49.4577 | 46.0257 | 42.4030 | 38.9710 | 35.5390 | 31.9164 |
| | 干流 | 62.3132 | 57.6991 | 53.3279 | 48.7139 | 44.3427 | 39.9715 | 35.3574 | 30.9862 | 26.6150 | 22.0010 |
| | 支流 | 65.0405 | 62.2289 | 59.5652 | 56.7535 | 54.0899 | 51.4262 | 48.6145 | 45.9509 | 43.2872 | 40.4755 |
| 水文年 | 全流域 | 67.3625 | 64.5330 | 61.8525 | 59.0230 | 56.3425 | 53.6619 | 50.8325 | 48.1519 | 45.4714 | 42.6419 |
| | 干流 | 68.2090 | 64.7008 | 61.3774 | 57.8693 | 54.5458 | 51.2223 | 47.7142 | 44.3908 | 41.0673 | 37.5592 |
| | 支流 | 68.6906 | 66.7622 | 64.9353 | 63.0069 | 61.1800 | 59.3531 | 57.4247 | 55.5978 | 53.7709 | 51.8425 |

由此结果可见, 按目前的污染状况, 如果不采取有效的综合治理措施, 10年后到枯水期时长江干流可饮用水的比例也只剩下1.5%了, 即有98.5%的江段水都成了IV类, 或V类, 甚至劣V类水, 不可饮用了, 水中也不会有生物存在。正像专家所说: “长江生态10年内将濒临崩溃。”就是在丰水期, 也有78%的江段都变成了非饮用水, 污染的状况十分严重。

6 长江水质的控制模型

针对问题4), 如果要求未来长江干流IV类、V类和劣V类水的比例总和不超过20%, 即可饮用水比例不小于80%, 就要求(6)式中的 $y \geq 80$, 则 $a\phi_1(t) + b\phi_2(t) + c \geq 80$ 。同时要求没有劣V类水, 用类似模型(6)的方法, 可得到长江干流IV类和V类水的比例总和与总排污量 $\phi_1(t)$ 和水流量 $\phi_2(t)$ 的关系 $y' = a'\phi_1(t) + b'\phi_2(t) + c'$, 根据题目要求则应有 $y' \leq 20$, 且劣V类水的比例 $100 - y - y' = 0$ 。于是未来10年允许的排污量 $\phi_1(t)$ 则应满足

$$\begin{cases} a\phi_1(t) + b\phi_2(t) + c \geq 80, \\ a'\phi_1(t) + b'\phi_2(t) + c' \leq 20, \\ 100 - (a + a')\phi_1(t) - (b + b')\phi_2(t) - c + c' = 0. \end{cases}$$

取水流量 $\phi_2(t)$ 为平均值9894.1, 分别计算枯水期和丰水期允许的平均年排污量为217亿吨和230亿吨。再根据未来10的正常排污量预测值, 则可以得未来10年内每年需要处理的污水数量如表6所示。

表6: 未来10年的需要处理污水量预测值

| 年份 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|--------|---------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 排污量预测值 | 303 | 322 | 340 | 359 | 377 | 395 | 414 | 432 | 450 | 469 |
| 枯水期处理量 | 86 | 105 | 123 | 142 | 160 | 178 | 197 | 215 | 233 | 252 |
| 丰水期处理量 | 73 | 92 | 110 | 129 | 147 | 165 | 184 | 202 | 220 | 239 |
| 年处理区间值 | [73,86] | [92,105] | [110,123] | [129,142] | [147,160] | [165,178] | [184,197] | [202,215] | [220,233] | [239,252] |

由结果可知,因为年排污量在逐年增加,要保持一定的水质指标就必须增加污水的处理量,具体数值的多少主要取决于年水流量的大小。当水流量相对较大时,有利于水质的改善,污水处理量可以减少,也能保证水质的要求;当水流量相对较小时,会加重水质污染程度,故要增加污水处理量。为此,给出一个污水处理允许的区间值,通常情况下,根据水流量的大小在相应的区间内确定污水处理量。

7 对问题的综合评述

“长江水质的评价和预测”问题被用作2005年全国大学生数学建模竞赛A题,使得自己为此所做的工作和付出得到了专家和同行们的肯定,为此深感荣幸和欣慰,同时能够为全国大学生数学建模竞赛活动尽微薄之力而感到高兴。竞赛之后,有幸应邀参加了湖南与湖北赛区联合、河南赛区及全国的阅卷工作,在此期间很多专家们都对这个问题给予了很高的评价和赞赏,尤其是对题目的立意、题材和内容等方面,以及问题的开放性、灵活性、创造性和即时性的体现等方面都给予了好评。但也有的老师反映,这个题目太大了,要解决的问题太多了,尤其是数据量太大、太复杂了,学生在三天之内很难完成好。确实像这样一个实际的问题,数据也基本都是原始数据,短时间内解决实属不易,而且还要求做好更难了。尽管如此,在甲组参赛队中还是有60%以上队都选作了这个题目,这反映出了同学们对这个题目的偏爱和兴趣,同时也反映出广大青年学生勇于挑战,敢于拼搏的精神和气质,值得称赞。从送到全国评奖的答卷也让我们欣慰地看到很多同学完成的还是很出色的,充分地体现出了同学们的创新精神和从事实际科研的能力与素质,我想这也是我们数学建模竞赛的目的所在。2005年的赛题的突出特点是:开放性,实用性强,数据信息量大,可解方法灵活多样。我认为这是现代实际科研工作的特点之一,也应该是数学建模的一个发展方向。

7.1 解决问题的主要方法概述

由于“长江水质评价和预测”问题实际背景和特点,解决问题的可行方法灵活多样,从问题的数值结果上很难说明方法的对与错,但从实际问题来分析还是可以看出方法之间的差异,可也以判断出方法的优与劣,有些方法所得到的结果明显的不符合实际。

对于问题1)是多因素多属性的综合评价分析问题,可用的方法有综合加权法、模糊综合评判法、主成份分析法、灰色聚类、多目标决策、神经网络等方法,无论用哪一种方法都应该有一个合理的综合评价指标(函数),作为综合评价的依据。有一部分参赛队没有用任何综合评价的方法,只是对所给的数据做了一些简单的统计计算,即统计各类水出现的次数、频率,或取平均值即得到一个评价结果,这显然是不可取的。

对于问题2)是由历史数据来确定长江干流的主要污染源,事实上这是一个微分方程的反问题。大多数参赛队都是利用简化的一维水质模型(连续形式或差分形式)研究污染物的降解作用,从而可以确定各地区污染物的浓度变化。有一部分参赛队考虑了干流上排污源的影响,并

假设排污点在江段内均匀分布，或者所有的排污源都集中在某一点处等处理方法，这些都是合理的方法。最后根据各江段排污量的多少确定主要的污染源在哪里。

对于问题3)是对未来10年水质变化的预测问题，即是一个典型的小样本的预测问题，可用的方法有很多，如灰色预测、时间序列、回归分析、拟合（指数拟合、线性拟合、分段非线性拟合、二次拟合）等等。只是这些方法如何来使用，使用方法不当可能会导致错误，有的会得到荒唐的结果。事实上，对于没有明显变化规律（或趋势）的数据样本做预测，用任何方法都是不可靠的。因为各单项比例都没有明显的变化趋势，所以有很多的参赛队都是对Ⅰ类、Ⅱ类、Ⅲ类、Ⅳ类、Ⅴ类和劣Ⅴ类水的比例分别做预测的，所得到的结果都是不可靠的，比如各类水的比例预测值之和大于100%是个明显的错误。

对于问题4)是在问题3)的基础上进一步研究对水质的控制问题。事实上，该问题与问题3)紧密相关的问题，但有些参赛队把它们看成两个不相关的问题，使得问题人为的复杂化了。如果在问题3)中找出了可饮用水（或不可饮用水）的比例与总排污量和水流的关系，通过预测总排污量和水流量来计算出可饮用水的比例，而不是直接对各类水的比例进行预测，则问题4)就是水到渠成的事情了。否则问题4)就变成了与问题3)独立的问题，只能重复做上述的工作，使工作量大大增加。值得我们注意的是，在有多种可能的方法（或模型）时，要学会综合比较、优化选择合理有效的方法（或模型），而不是简单地套用某一种方法（或模型），这正向人们所说的“不要光重视最优化模型，更要重视模型的最优化。”

7.2 答卷中出现的主要问题概述

由于该问题的开放性和解决方法的非惟一性等特点，所用知识和方法在以往的赛题中也并不多见，这就使得有些学生感到困惑和陌生，在处理问题时不知所措，因此出现了一些问题或错误。

1) 对问题1)的理解。很多人对“长江水质情况做出定量的综合评价”的理解不当，甚至有人认为在理解上有歧义，不知道该评什么，用什么数据来评，如何来评等？其实出现这些模糊概念的关键是不清楚什么是“综合评价”，让我们看一下相关解释。“综合”——把事物的各个部分、方面、因素结合成一个统一整体加以考察；即把各种不同而互相关联的事物或现象组合在一起。“评价”——泛指衡量人物、事物的作用或价值。因此，“综合评价”的基本含义就不言而喻了。要构成“综合评价”的问题必须要有五个要素^[2]：“被评价对象；评价指标；权重系数；综合评价模型（函数）和评价者”。解决综合评价问题的一般步骤是：“明确评价目的；确定被评价对象；建立评价指标体系；评价指标的标准化处理；确定相对权重系数；选择或构造综合评价模型；计算各系统的综合评价值，并进行排序或分类”。这是任何一个综合评价问题都必须要做的工作。于是，问题1)要评什么，用什么数据来评，如何来评等问题就不言自明了吧。

2) 由于某些人对定量的综合评价方法不了解，就出现了仅对水质的类型做简单的统计计算，以各类水出现的次数或概率来评价，或以各类水出现的平均次数来评价。此类作法的主要问题是：缺少科学性，不符合综合评价的要求，不能区分同类水质中的数量差异。还有队说国标中关于水的分类方法不合理，自己试图给出一种自认为合理的分类标准。也有的队硬给某个地区或整个长江的水质进行分类，这就等于是修改了国标或自定分类标准，这也是不合适的。

3) 有相当多的参赛队在评价中，不考虑PH值在综合评价中的影响是不合适的。虽然PH值大小对水的分类无影响，但对水质是有很大影响的，过酸或过碱性的水质对生态环境的影响是很大的。

4) 有些队在建立综合评价指标时，对原始数据没有作标准化处理（无量纲化，或归一化等），这在综合评价中是一个低级的错误，违背了综合评价指标“可公度性”的要求。还有一

些仅用所给数据简单的作图描点分析,缺少定量依据。很多答卷没有给出明确的分析结果。

5) 对于问题2),有些队没有考虑长江的自然降解功能,仅对17个观测点的数据做统计,依据浓度的大小得到结果是不合适的。有相当多的一些参赛队都没有考虑干流上的排污源(包括支流和直排口等)的排污影响。有的参赛队是以两观测站之间的观测数据计算出来的排污总量的大小来确定污染源是不太合理的,因为站点之间的距离不同,排污总量不是可比指标。个别的队对降解系数为0.2不理解,认为是有20%的污染物被降解掉了,剩下的80%流到下游。也有的队通过一维水质模型计算出来的排污量为负值的情况没有做出合理的解释说明。

6) 对于问题3),有很多的队是对各类水质的比例直接作预测的,明显的是不合理的,主要是各单类比例没有明显的变化规律可寻,用任何方法预测都是不可靠的,从而出现了各类水的比例总和大于10%的错误结果。各类水的比例变化应与总的年排污量和总水流量有关,有的队没有考虑总水流量的影响,只考虑排污量的影响不合适。很多的队都是用拟合的方法来做预测的,方法虽可行,但用指数拟合和二次函数拟合增长过快,线性拟合效果欠佳,三次以上的高次拟合是拟合问题的一大忌。对长江的年总流量用任何方法预测都不太合理,因为没有明显的趋势,所以取平均值较好。

7) 对于问题4),大多数队都是在问题3)中直接对各类水比例做的预测,没有找出各类水比例与年排污总量和水流总量之间的关系,为此要解决问题4)必须先解决这个问题,使得问题复杂化。有的队对题意的理解有误,如题目中要求“四类水和五类水的比例不超过20%”,他认为是两类水分别小于等于20%。还有个别的队把所给长江长度的百分比理解成长江水浓度的百分比,于是对总排污量直接处理80%,即得到污水的处理量,显然是错误的。还有少数队直接对长江水总流量进行处理80%,从而得到了污水处理量达到几仟亿吨的荒唐结果。八九仟立方米的滔滔长江水如何来处理呢?这应该是一个常识性的问题。

8 结束语

2005年“高教社杯”全国大学生数学建模竞赛工作圆满的结束了,作为“长江水质评价和预测”问题的命题人,同全国的25446名参赛大学生一样也是一个参与者,同样也有深刻的感受和非常的收获。从开始考虑这个问题到搜集数据,直到最后形成题目,在这大半年的时间里让我感慨万千,酸甜苦辣可以说是一言难尽,在这里也不想赘述。在这个题目发布以后,引起的国内多家媒体和十多家网站的关注报道,这也说明了这个题目的时效性。在几次评卷过程中,接触到了很多专家教授和同行们,从他们对赛题的不同看法和评价中,也都给了我很多的启迪。我个人觉得中国的数学建模竞赛已经历十几个年头了,可以说是从当初的初级阶段发展到中级阶段,目前应该是正处在从中级往高级迈进的一个时期。总结历史展望未来,中国的大学生数学建模竞赛应该向何处发展,尤其是竞赛题应该如何发展?在开放性问题上个人觉得应该寻求适合中国国情的“中国式开放”,而不要套学“美国式开放”。一个理由是在中国可以共享的信息资源还很有限,虽然解决问题的必要数据资料客观存在,但是事实上很多都是属于主管部门的专利,所以我们无法获取。有人说:“如果把长江水质评价和预测问题中所给的数据都去掉就像美国的赛题了。”不错,我也相信这一点。我用了近几个月时间,动用了网络、电话、人力、物力、财力等资源才收集到这些宝贵的数据资料,让参赛学生三天内如何能做到呢?简直是天方夜谭。因此,我们的数学建模竞赛应向何处发展,赛题应如何开放?是值得我们共同关心、思考和研究的课题。

虽然今年的竞赛已落下帷幕,但围绕着这个问题也留给我们很多的思考,从题目本身还有很多有待于进一步讨论研究的问题。同时由于评卷规则的限制,没能看到所有参赛队的答卷,

可能还有很多好的答卷和很好的做法没有看到, 这里讲的不一定全面。因此, 也希望有更多的同行和学生一起来讨论、交流和研究相关的问题, 本人的 E-mail: zhghan@163.com。

致谢: 对全国组委会的专家、中国发展研究院的章琦教授、华东师大的陆健健教授和《长江年鉴》编辑部、水利部中国水环境研究院信息中心和水质研究所的有关专家, 以及多年给予我关心和帮助的同行们一并表示衷心的感谢。

参考文献:

- [1] 岳超源. 决策理论与方法[M]. 北京: 科学出版社, 2003
- [2] 郭亚军. 综合评价理论与方法[M]. 北京: 科学出版社, 2002
- [3] 韩中庚. 数学建模方法及其应用[M]. 北京: 高等教育出版社, 2005
- [4] 韩中庚. 招聘公务员问题的优化模型与评述[J]. 工程数学学报, 2004,21(7):147-154
- [5] 韩中庚. 研究生录取问题的优化模型与评述[J]. 数学的实践与认识, 2005,35(7):126-135
- [6] 国家环保总局. 长江流域水质检测数据[R/OL]. <http://ww2.sepa.gov.cn/apps/Queryqgzdwaterenv.jsp>

The Mathematical Model of Water Quality Comprehensive Evaluation and Prediction of the Yangtze River

HAN Zhong-geng

(Institute of Information Engineering, Information Engineering University, PLA, zhengzhou 450002)

Abstract: In this paper, according to the problem A of 2005 HIGHER EDUCATION PRESS CUP CUMCM, the motivation and background of the problem of water quality evaluation and prediction of the Yangtze River are introduced. Then the practicable solution methods and its results of the problem are given. Finally, according to the grading papers process, the comprehensive comments of the outline of grading, the solution methods and the existing problems are given.

Keywords: water quality of the Yangtze River; water quality classification; comprehensive evaluation and prediction; water quality model