

$$\Delta_3 = \{0, 0.191, 0.119, 0.046, 0.273, 0.242, 0.313, 0.354\}$$

于是两级差为:

$$\begin{aligned} \min_k \min_k |x_0''(k) - x_1''(k)| &= 0 \\ \max_k \max_k |x_0''(k) - x_1''(k)| &= 0.354 \end{aligned}$$

(2) 关联系数计算

令  $\rho = 0.5$ , 则有:

$$\begin{aligned} \xi_i(k) &= \frac{0 + (0.5 \times 0.354)}{\Delta_i(k) + (0.5 \times 0.354)} \\ &= \frac{0.177}{\Delta_i(k) + 0.177} \end{aligned}$$

于是得到:

$$\xi_1 = \{1, 0.50, 0.40, 0.689, 0.446, 0.827, 0.697, 0.388\}$$

$$\xi_2 = \{1, 0.63, 0.763, 0.391, 0.967, 0.656, 0.936, 0.967\}$$

$$\xi_3 = \{1, 0.481, 0.598, 0.794, 0.393, 0.422, 0.361, 0.333\}$$

4. 计算关联度

根据上述关联系数,求得  $x_i$  与  $x_0$  的关联度如下:

下:

$$r_1 = \frac{1}{8} \sum_{k=1}^8 \xi_1(k) = 0.618$$

$$r_2 = \frac{1}{8} \sum_{k=1}^8 \xi_2(k) = 0.789$$

$$r_3 = \frac{1}{8} \sum_{k=1}^8 \xi_3(k) = 0.548$$

由上述计算结果可知,在三个比较因素与分析

因素的比较中,工业产值的关联度  $r_1$  大于农业产值的关联度  $r_2$  和湖泊水位的关联度  $r_3$ , 其关联序为:  $r_1 > r_2 > r_3$ , 这表明在影响湖水化学需氧量浓度变化的三个因素中,工业产值(即工业生产过程中产生废水造成的影响)是最主要的因素,其次是农业产值(即农业生产过程施用化肥,农药造成的影响),湖泊水位影响为第三位。以上定量分析的结果与图中的直观分析是一致的,表明结果是可靠的。同时对关联度  $r_i$  作归一化处理(即  $r'_i = r_i / \sum_{i=1}^3 r_i$ ) 还可以反映出每个比较因素对湖泊化学需氧量浓度变化的影响所占的比重,这三种比较因素对分析因素的影响比重为:工业影响占的比重  $r'_1$  为 40.4%,农业影响占的比重  $r'_2$  为 31.6%,湖泊水位影响占的比重  $r'_3$  为 28%。这表明要改善该湖水中有机的污染,首先要减少工业排污对湖水的污染,其次是减少农业施用化肥、农药造成对湖水的污染。这为湖泊有机物的污染防治提供了依据。

综上所述,将灰色理论的关联度分析作为湖泊环境系统因素分析的方法,其方法简便、适用、结果可靠,且计算也并不复杂,其计算工作量较回归分析要少得多,具有实际应用的活力。

参 考 文 献

- [1] 邓聚龙,灰色控制系统,348页,华中工学院出版社:武汉,1985年.
- [2] 邓聚龙,模糊数学,(2),1(1985).
- [3] 谭继强,环境污染与防治,9(5),30(1987).

(收稿日期:1988年3月16日)

## 黄河兰州段耗氧有机污染物浓度 统计预测模型的建立

史复有                      孙 谦                      李昌迪

(甘肃省环境保护研究所) (甘肃省环境监测中心站) (甘肃省环境保护局)

黄河是一条多泥沙河流,流域内面源污染严重,在分析研究黄河的水环境问题时,必须考虑面源污染这个因素。建立河流污染断面耗氧有机污染物(以 COD 表示)浓度预测模型是河流水质规划的一

项基础工作。本文就建立黄河兰州段污染断面耗氧有机污染物浓度多元线性回归预测模型的有关面源污染等方面问题进行分析探讨。

## 一、变量的选择

建立污染物浓度预测模型第一步工作是选择变量,特别对耗氧有机污染物这样面广量大的污染物,变量选择是否准确、完整,对多元回归模型的可靠性影响很大。

### 1. 考虑点污染源影响

工业废水和生活污水中含有大量的耗氧有机污染物,调查统计工业废水和生活污水中耗氧有机物的含量是很困难的,一般以流域的工业产值和人口数作为间接统计指标。

只考虑工业污染源(即兰州市区工业产值)这一因素可建立如下二元线性回归模型,

$$y = 0.50 + 0.09x_1 \quad (1)$$

式中  $y$  为黄河兰州段污染控制断面(包兰桥断面)的耗氧有机污染物浓度 ( $\text{mg/L}$ ),  $x_1$  为兰州市区工业产值(亿元),以下符号意义相同。

建立模型的样本数为 16(下同)。

该回归模型相关系数  $R = 0.20$ , 这说明耗氧有机污染物浓度与工业产值线性相关很差,模型的可靠性差。考虑到影响包兰桥断面耗氧有机污染物浓度变化还有其它因素,为此我们引入反映兰州市区生活污水中耗氧有机污染物的人口这一变量,建立如下二元线性回归模型。

$$y = -170.81 - 0.96x_1 + 2.43x_2 \quad (2)$$

$x_2$  为兰州市区人口(万人)

该模型的复相关系数  $R = 0.31$ , 这说明点源污染和耗氧有机污染物线性相关仍很差,只考虑点源污染建立的模型可靠性是差的。

### 2. 考虑面污染源影响

甘肃省环境保护局和兰州大学数学系等单位发现河流中泥沙含量对河水耗氧有机污染物浓度影响很大,甘肃省环境保护研究所进而发现面源污染对河水耗氧有机污染物浓度影响亦很大<sup>[1]</sup>。河流中含沙量是面源污染的一种反映,降雨量是导致面源污染发生的一个因素,将这二变量对耗氧有机污染物进行回归,建立如下模型。

$$y = 3.40 + 0.69x_3 + 0.16x_4 \quad (3)$$

$x_3$  为河水含沙量 ( $\text{kg/m}^3$ );  $x_4$  为监测前十五天内降雨量 ( $\text{mm}$ )。

该模型复相关系数  $R = 0.89$ , 这说明耗氧有机污染物浓度与面源污染相关性好,还说明面源污染对包兰桥耗氧有机污染物浓度的影响要比点污染源的大,从数学模型得出的结论和对耗氧有机污染物浓度变化规律的定性分析相吻合<sup>[1]</sup>。同时也说明

引入反映面源污染的变量的多元回归预测模型可靠性大大提高。

综合考虑点源和面源的影响,可建立如下模型。

$$y = -230.91 - 1.38x_1 + 3.29x_2 + 1.09x_3 + 0.13x_4 \quad (4)$$

该四元回归模型复相关系数  $R = 0.94$ , 比模型(2)、(3)的复相关系数都高,这说明选择的变量完全,线性相关性就好,模型的可靠性就高。

### 3. 考虑流量的影响

一般河流只考虑点源对水质的影响,这时流量常作为稀释因素出现在分母中。按这思路可建立下列模型。

$$y = 6.60 - 11.66 \times \left( \frac{\text{人口} + \text{工业产值}}{\text{流量}} \right)$$

该回归模型的相关系数  $R = 0.29$ , 比模型(2)的复相关系数还低。这解释如下:黄河这样面源污染严重的河流,流量对耗氧有机污染物浓度的影响方式复杂,不能简单当作稀释因素。可将流量作为一个影响因素来考虑,可建立如下三元回归模型。

$$y = -227.04 - 1.23x_1 + 3.12x_2 + 0.0017x_3 \quad (5)$$

$x_3$ ——河水流量 ( $\text{m}^3/\text{s}$ )

该模型复相关系数  $R = 0.48$ , 比模型(2)的复相关系数要高,说明引入流量后,多元回归的线性相关性有所提高,流量对河流耗氧有机污染物有一定影响。

### 4. 综合各因素影响

从以上分析可看到影响包兰桥断面耗氧有机污染物浓度的因素有:工业产值、人口、含沙量、降雨量以及流量等五个因素,我们可建立一个五元线性统计模型。

$$y = -218.86 - 1.36x_1 + 3.16x_2 + 1.46x_3 + 0.13x_4 - 0.0008x_5 \quad (6)$$

该模型复相关系数  $R = 0.98$ , 说明相关性很好,该模型的可靠程度很高。

### 5. 关于模型可靠性分析

我们将建立多元回归模型的 16 个实测数据与模型 2、3、4、6 的计算值及相对偏差列于表 1。

此表可以说明复相关系数高的多元线性回归模型的计算值与实测数据的相对偏差小,也就是说复相关系数高的多元回归模型的可靠程度高。

## 二、几个问题讨论

### 1. 样本相关矩阵

对于模型(5)来说,我们可得如下样本相关矩

表 1 多元回归模型与计算值的相对偏差

| 时间<br>(年)     | 模型 (6)       |             | 模型 (4)       |             | 模型 (3)       |             | 模型 (2)       |             | 实测数据<br>(ppm) |
|---------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|---------------|
|               | 计算值<br>(ppm) | 相对偏差<br>(%) | 计算值<br>(ppm) | 相对偏差<br>(%) | 计算值<br>(ppm) | 相对偏差<br>(%) | 计算值<br>(ppm) | 相对偏差<br>(%) |               |
| 1981          | 3.4          | 78.95       | 2.9          | 34.48       | 3.45         | 81.58       | 4.3          | 126.32      | 1.9           |
|               | 4.0          | 6.98        | 3.5          | 18.6        | 3.8          | 11.63       | 4.3          | 0           | 4.3           |
|               | 2.7          | 8.0         | 3.2          | 28.0        | 3.6          | 44.0        | 4.3          | 41.86       | 2.5           |
| 1982          | 8.2          | 14.58       | 7.6          | 20.83       | 7.2          | 25.0        | 5.4          | 43.75       | 9.6           |
|               | 4.85         | 8.49        | 5.05         | 47.17       | 3.7          | 30.19       | 5.4          | 1.89        | 5.3           |
|               | 3.65         | 25.86       | 4.3          | 48.28       | 3.6          | 24.13       | 5.4          | 86.21       | 2.9           |
| 1983          | 12.5         | 1.91        | 12.6         | 11.5        | 13.5         | 19.47       | 4.8          | 57.52       | 11.3          |
|               | 2.3          | 11.54       | 2.8          | 7.69        | 3.4          | 30.77       | 4.8          | 84.52       | 2.6           |
| 1984          | 2.3          | 43.75       | 1.9          | 18.75       | 3.45         | 115.63      | 4.4          | 175.0       | 1.6           |
|               | 7.1          | 6.58        | 7.1          | 6.58        | 8.1          | 6.58        | 4.4          | 84.52       | 7.6           |
|               | 1.9          | 34.48       | 1.9          | 34.48       | 3.4          | 17.24       | 4.4          | 51.72       | 2.9           |
| 1985          | 4.7          | 56.67       | 4.3          | 43.33       | 3.4          | 13.33       | 6.3          | 110.0       | 3.0           |
|               | 9.2          | 7.07        | 8.8          | 11.11       | 7.8          | 23.21       | 6.3          | 36.36       | 9.9           |
|               | 5.9          | 11.94       | 6.2          | 7.46        | 5.6          | 16.42       | 6.3          | 5.97        | 6.7           |
| 1986          | 6.9          | 11.29       | 7.1          | 14.52       | 5.6          | 9.68        | 6.7          | 8.06        | 6.2           |
|               | 4.8          | 0           | 4.7          | 2.08        | 3.5          | 27.08       | 6.7          | 39.58       | 4.8           |
| 平均相对偏差<br>(%) | 20.5         |             | 22.18        |             | 30.87        |             | 58.84        |             |               |
| 复相关系数<br>R    | 0.98         |             | 0.94         |             | 0.89         |             | 0.31         |             |               |

阵:

|       |        |        |       |       |        |
|-------|--------|--------|-------|-------|--------|
| 1     | 0.200  | 0.221  | 0.821 | 0.881 | 0.261  |
| 0.200 | 1      | 0.995  | 0.145 | 0.109 | -0.264 |
| 0.221 | 0.995  | 1      | 0.129 | 0.104 | -0.280 |
| 0.821 | 0.149  | 0.129  | 1     | 0.873 | 0.534  |
| 0.881 | 0.109  | 0.104  | 0.873 | 1     | 0.455  |
| 0.261 | -0.264 | -0.280 | 0.534 | 0.455 | 1      |

从中可看出:与包兰桥断面耗氧有机污染物关系最密切的变量是含沙量和降雨量,它们与  $y$  的相关系数分别为 0.821 与 0.881,是优度相关,可解释为面源污染对黄河兰州段 COD 值影响很大.兰州市区工业产值、人口、流量与  $y$  的相关系数分别为 0.200、0.221 与 0.261,可解释为点源污染及流量对 COD 值影响很小.从样本相关矩阵得出的结论也和定性分析<sup>[1]</sup>是吻合的.

2. 水期监测平均值与每日监测值

通常环保部门每个水期连续监测五日,每日监测一次.水质规划工作常用每个水期五日监测平均值,由于平均值减少了数据的离散性,所以用平均值建立的多元线性回归模型可靠性要高.我们曾用

十六个水期对应的 79 个日监测值建立的多元回归模型如下:

$$y = -152 - 0.94x_1 + 2.21x_2 + 1.06x_3 + 0.17x_4 - 0.001x_5$$

该模型的复相关系数  $R = 0.76$ , 比起模型(5)的复相关系数低,上述结论正确.

3. 模型中负系数问题

模型中会出现某些变量系数为负现象,例如模型(2)、(4)、(5)中工业产值( $x_1$ )的系数为负值,说明该变量与因变量呈负相关,在模型其它变量值不变的情况下,工业产值越高,包兰桥断面的耗氧有机污染物浓度越低,这是不合理的.从样本相关矩阵中可看到  $x_1$  与  $x_2$  (人口)之间的相关系数为 0.995 大于  $x_1$  与  $y$ ,  $x_2$  与  $y$  的相关系数,这样的两个变量同时引入多元回归线性模型中往往会导致其中一个变量的系数为负值,因此  $x_1$  和  $x_2$  这两变量要删掉一个,保留对规划工作意义大的工业产值这个变量.这样可以建立如下四元线性回归模型.

$$y = 4.25 + 0.0065x_1 + 1.10x_2 + 0.15x_3 - 0.001x_4 \quad (\text{下转第82页})$$

表 6 三峡库区及其上游水环境中放射性核素所致居民内照射剂量

|                               | <sup>238</sup> U     | <sup>226</sup> Ra    | <sup>232</sup> Th    | <sup>137</sup> Cs    | <sup>40</sup> K      |
|-------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 饮江水年食入量(Bq)                   | 2.20                 | 2.20                 | 1.46                 | 0.235                | 34.8                 |
| 吃虾年食入量(Bq)                    | 0.201                | 1.30                 | 0.521                | 0.033                | 80.4                 |
| 吃蟹螺蚌食入量(Bq)                   | 2.20                 | 7.80                 | 4.14                 | 0.112                | 77.3                 |
| 总年食入量(Bq)                     | 4.60                 | 11.3                 | 6.12                 | 0.38                 | 192.5                |
| 食入转换因子 <sup>[6]</sup> (Sv/Bq) | $6.3 \times 10^{-8}$ | $3.1 \times 10^{-7}$ | $7.4 \times 10^{-7}$ | $1.4 \times 10^{-8}$ | $5.1 \times 10^{-9}$ |
| 待积剂量当量(μSv)                   | 0.29                 | 3.50                 | 4.53                 | 0.0053               | 0.98                 |

2.8kg。那么沿江居民个体年食入放射性核素总量、剂量转换因子和估算得的待积剂量当量列于表6。数据表明,沿江居民由于饮江水和吃水生动物,年食入放射性核素所接受的待积剂量当量最大为9.31μSv(成人),其中<sup>232</sup>Th的贡献最大,达4.53μSv,其次是<sup>226</sup>Ra,为3.50μSv。因此,当前三峡库区及其上游放射性水环境的现状,尚不构成对沿江居民健康的危害。

#### 四、小 结

通过对Ge(Li)反康普顿γ谱仪计算机系统分析三峡库区及其上游水环境各类样品中放射性核素所获得的数据,进行了各种统计学分析、转移系数及沿江居民所受内照射剂量的估算,结论是:

1.河底泥中主要天然放射性核素和裂片核素<sup>137</sup>Cs,除<sup>232</sup>Th浓度的频数分布服从对数偏态分布外,均服从偏态分布。河水中核素浓度的频数分布看不出规律性。

2.水藻、水草和蟹中主要天然核素浓度的频数分布均服从对数正态分布。蚌中主要核素均服从正态分布,而螺中主要核素只有<sup>226</sup>Ra和<sup>232</sup>Th服从正态分布。极差与均值之比,河底泥、河水、水草特别是水藻的核素都大于西安土壤小麦的相应值。而中位数与均值之比,底泥、水草、特别是水藻又低得多。

(上接第74页)

流量前面系数为正或负,从道理上讲都能解释得通。该模型复相关系数 $R = 0.91$ ,其可靠性能满足规划预测工作要求。

#### 三、结 论

1.变量选择对多元线性回归模型可靠性(即复相关系数)影响很大。变量选择的准确完全是建立一个满意的多元回归模型的前提。

2.用水期监测平均值建立多元回归模型既便于

3.从核素的转移来看,<sup>232</sup>Th和<sup>226</sup>Ra容易沉积于河底泥中,而<sup>137</sup>Cs不容易沉积。水藻容易从河水和底泥中富集<sup>137</sup>Cs和<sup>40</sup>K。螺、蚌和蟹容易从河水和底泥中富集<sup>226</sup>Ra,蝌蚪则容易从河水和底泥中富集<sup>137</sup>Cs和<sup>40</sup>K。水生动植物都较易从水中浓集<sup>232</sup>Th,但不易从底泥中获取该核素。船底附生物不宜作为水环境监测介质。

4.从三峡库区及其上游水环境中放射性的现状来看,白龙江下游广元点受到了轻微放射性污染,其次是嘉陵江合川点和金沙江宜宾点。沿江居民由于饮用江水和吃水生动物而年食入放射性核素,所接受的待积剂量当量约9.31μSv,尚不会危害居民身体健康。

致谢:全部样品均由四川省工业卫生研究所杨开祥、向长兴同志组织采集和制备,在此表示衷心的感谢!

#### 参 考 文 献

- [1] 高玉堂,环境监测常用统计方法,32页,北京原子能出版社,1979年。
- [2] 数学手册编写组,数学手册,291页,北京人民教育出版社,1979年。
- [3] 庞巨丰等,环境科学,9(1),1(1988)。
- [4] 同[2],7-15页,57页。
- [5] 中华人民共和国国家标准,GB 4792-84,(1984)。

应用,也比用日监测值建立的回归模型可靠性高。

3.模型中某些变量前面系数为负值往往是不合理的,可通过分析样本相关系数阵,删除对规划工作意义小的有关变量来消除。

致谢:兰州第三毛纺织厂林亨同志参加本文讨论,在此表示感谢。

#### 参 考 文 献

- [1] 史复有,人民黄河,(3),21(1988)。