

# 模糊聚类分析法优化城市河涌水质监测点

梁伟臻, 叶锦润, 杨 静

(广州市城市排水监测站, 广东 广州 510655)

中图分类号: X830.1

文献标识码: C

文章编号: 1006-2009(2002)03-0006-02

在进行城市河涌(小河滨、河沟等)水质监测工作的初期,一般会布设较多的监测点,以掌握水域的水质情况。经过一段时间的监测,积累了一定数量的监测数据后,便应对监测点进行优化。现用模糊数学中的聚类原理<sup>[1]</sup>,以广州市某涌为例优化水质监测点。

## 1 水质监测点的优化

### 1.1 建立标准化矩阵 $R$

该涌布设的 7 个监测点在 1997 年—2000 年间 5 个常规监测项目的平均数据见表 1。

表 1 广州市某涌 1997 年—2000 年监测数据 mg/L

监测点	COD	BOD <sub>5</sub>	NH <sub>3</sub> -N	TP	SS
淘金路	117	57.7	22.3	2.76	322
华乐路	173	81.6	19.5	2.97	207
旧管理局	204	92.9	19.8	2.87	148
第 17 中学	132	68.1	20.6	2.37	76
越秀桥	136	48.5	19.1	2.51	68
三角市	111	57.6	20.0	2.57	146
大沙头	134	40.2	21.3	2.52	176
标准值 <sup>①</sup>	100	30	15	2	70

①标准值为 GB 8978-1996《污水综合排放标准》中对“其他排污单位”的一级标准(其中 TP 为自定值)

$$\tilde{R} = \begin{bmatrix} 1 & 0.854 & 0.795 & 0.805 & 0.793 & 0.691 & 0.650 \\ 0.854 & 1 & 0.925 & 0.711 & 0.800 & 0.694 & 0.661 \\ 0.795 & 0.925 & 1 & 0.848 & 0.753 & 0.705 & 0.674 \\ 0.805 & 0.711 & 0.848 & 1 & 0.933 & 0.902 & 0.800 \\ 0.793 & 0.800 & 0.753 & 0.933 & 1 & 0.892 & 0.852 \\ 0.691 & 0.694 & 0.705 & 0.902 & 0.892 & 1 & 0.912 \\ 0.650 & 0.661 & 0.674 & 0.800 & 0.852 & 0.912 & 1 \end{bmatrix}$$

### 1.3 建立模糊等价矩阵 $R^*$

对样本空间  $X$  分类,由  $\tilde{R}$  构造出新的模糊等价矩阵  $R^*$ 。 $R^*$  可通过平方自合成法计算得到。

将表 1 中的数据标准化,标准化方法一般采用污染指数法,即用污染物浓度值除以标准值。计算后可得到标准化矩阵  $R$ 。

$$R = \begin{bmatrix} 1.17 & 1.92 & 1.49 & 1.38 & 4.60 \\ 1.73 & 2.72 & 1.30 & 1.48 & 2.96 \\ 2.04 & 3.10 & 1.32 & 1.44 & 2.11 \\ 1.32 & 2.27 & 1.37 & 1.18 & 1.09 \\ 1.36 & 1.62 & 1.27 & 1.26 & 0.97 \\ 1.11 & 1.92 & 1.33 & 1.28 & 2.09 \\ 1.34 & 1.34 & 1.42 & 1.26 & 2.51 \end{bmatrix}$$

### 1.2 建立模糊相似关系矩阵 $\tilde{R}$

用多元分析方法建立样本间的相似关系(亲疏关系)。在环境科学中,污染物在水体中呈对数正态分布或与之相近<sup>[2]</sup>,由于几何平均数能较好地反映出数据的典型特征,因此选择几何平均最小法标定,标定后得到模糊相似关系矩阵  $\tilde{R}$ 。标定公式如下:

$$R_{ij} = \frac{\sum_{k=1}^m \min(X_{ik}, X_{jk})}{\sum_{k=1}^m \sqrt{X_{ik} \cdot X_{jk}}}$$

收稿日期:2002-01-08

作者简介:梁伟臻(1975-),男,广东广州人,助理工程师,大学,从事排水水质监测与管理工作的。

计算  $R \circ R = R^2, R^2 \circ R^2 = R^4 \dots$  (式中  $\circ$  为扎德算子), 若有  $k$ , 使  $R^{2k} = R^{2k-1}$ , 则  $R^* = R^{2k}$ 。在本

例中可计算出, 若  $R^8 = R^{16}$ , 则  $R^* = R^8$ 。

$$R^* = \begin{bmatrix} 1 & 0.854 & 0.854 & 0.848 & 0.848 & 0.848 & 0.848 \\ 0.854 & 1 & 0.925 & 0.848 & 0.848 & 0.848 & 0.848 \\ 0.854 & 0.925 & 1 & 0.848 & 0.848 & 0.848 & 0.848 \\ 0.848 & 0.848 & 0.848 & 1 & 0.933 & 0.902 & 0.902 \\ 0.848 & 0.848 & 0.848 & 0.933 & 1 & 0.902 & 0.902 \\ 0.848 & 0.848 & 0.848 & 0.902 & 0.902 & 1 & 0.912 \\ 0.805 & 0.848 & 0.848 & 0.902 & 0.902 & 0.912 & 1 \end{bmatrix}$$

在实际运算过程中, 计算  $R^*$  的工作量较大, 而使用直接聚类的方法, 步骤简单, 运算结果也在使用  $R^*$  聚类的结果接近, 因此建议直接用  $R$  聚类, 以简化运算。

1.4 动态聚类

用置信水平  $\lambda$  截割  $R^*$ , 对被分类对象动态聚类。从  $R^*$  可以看出, 可选择的  $\lambda$  为 {0.933, 0.925, 0.912, 0.854}。

1.5 画动态聚类图

根据所确定的  $\lambda$  值画出动态聚类图, 见图 1。

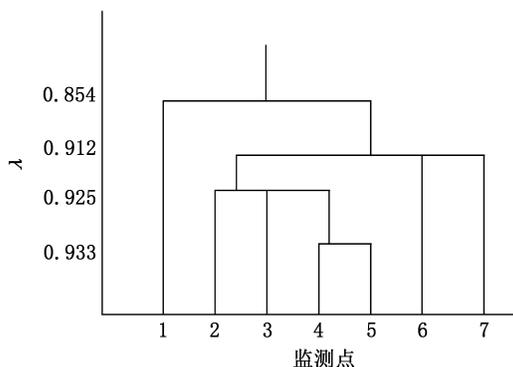


图 1 广州市某涌监测点动态聚类

1.6 选择最佳阈值

在用  $R^*$  聚类分析时, 最佳  $\lambda$  值往往凭经验确定, 缺乏科学性和客观性, 现提出将  $\lambda$  的变化率  $C$  作为  $\lambda$  的定量选择标准。

$$C = \frac{\lambda_{i-1} - \lambda_i}{n_i - n_{i-1}}$$

式中  $i$  为  $\lambda$  从高到低的聚合序次数,  $n_i$  和  $n_{i-1}$  分别是第  $i$  次和第  $i-1$  次聚类的元素个数。

假设当  $C_i = \max(C_j)$  时, 认为第  $i$  次聚类的置

信水平  $\lambda_i$  为最佳阈值。根据上式可计算出:

$$C_1 = \frac{1 - 0.933}{2 - 0} = 0.0335$$

同理可算出  $C_2 = 0.004, C_3 = 0.0065, C_4 = 0.058$ 。由于将所有的监测点归为一类没有实际意义, 所以将  $C_4$  舍弃, 则

$$C_i = \max(C_1, C_2, C_3) = C_1 = 0.0335$$

即选取  $\lambda = 0.933$  为最佳置信水平。用该  $\lambda$  值截割  $R^*$ , 从图 1 可见, 将监测点 4 和 5 归为一类最合适, 可信度达 93.3%。

2 结论

应用模糊聚类分析方法, 可在不能明确各污染物相互作用过程的情况下, 确定其主要联系和相互关系及影响的主要因素, 并在累积影响的基础上进行综合分析<sup>[3]</sup>。应用该方法优化水质监测点克服了以往单凭经验优化时过于依赖主观分析的不合理因素, 提供了客观分析的依据, 提高了水质监测点优化的准确性和工作效率。实际应用情况也表明, 用模糊聚类分析法优化水质监测点既具有较强理论性, 又切实可行。

[参考文献]

[1] 贺仲雄. 模糊数学及其应用[M]. 天津: 天津科学技术出版社, 1983. 23- 52.  
 [2] 吴聿明. 环境统计学[M]. 北京: 中国环境科学出版社, 1990. 60- 71.  
 [3] 毛文锋, 吴仁海. 累积影响评价的理论[J]. 环境科学研究, 1998, (5): 37- 41.

本栏目责任编辑 姚朝英