

变权欧式距离模型在水质综合评价中的应用*

张庆庆, 许月萍, 牛少凤, 楼章华
(浙江大学水文与水资源研究所, 浙江 杭州 310058)

摘要: 在加权欧式距离模型的基础上进行改进, 针对不同的待评价单元和不同污染性质的污染物, 根据污染因子的污染贡献率计算其权重; 以 n 个污染因子为坐标轴, 以水质标准中的 I 级水质作为 n 维空间的“原点”, 结合权重计算出每个待评价单元以及水质标准中其他各级水质与“原点”的变权欧式距离, 并将后者作为划分每个待评价单元水质级别的标准, 从而建立了适合水质综合评价的变权欧式距离模型。用该模型对 1996-2006 年钱塘江支流东阳江许村断面水质监测数据进行了综合评价, 并与灰色聚类法和模糊综合指数法的评价结果进行比较, 结果与水质实际情况符合较好, 验证了变权欧式距离模型应用于水质综合评价的科学性和合理性。

关键词: 变权欧式距离模型; 污染贡献率; 水质综合评价; 灰色聚类法; 模糊综合指数法

中图分类号: X824 **文献标志码:** A **文章编号:** 0529-6579 (2010) 05-0141-05

Application of Euclidean Distance Model with Varying Weights in Comprehensive Assessment of Surface Water Quality

ZHANG Qingqing, XU Yueping, NIU Shaofeng, LOU Zhanghua

(Institute of Hydrology and Water Resources Engineering, Zhejiang University, Hangzhou 310058, China)

Abstract: On the basis of Euclidean distance model with weights, this study improved the calculation of the weights. The weights were added to pollution factors according to different pollution contribution rates. The level I of surface water environmental quality standards was taken as the ‘coordinate origin’. The Euclidean distances with varying weights between units and ‘coordinate origin’ and between other levels of surface water environmental quality standards and ‘coordinate origin’ were calculated. The latter served as a distance level evaluation criteria. The Euclidean distance model with varying weights was established, which is suitable for assessment of surface water quality. This model was applied in comprehensive assessment of water quality, using the monitoring water quality data obtained from 1996 to 2006 in Xucun section of the Dongyang River, a tributary of the Qiantang River. The outcome was compared with that of grey clustering model and fuzzy complex index method. The results showed consistence with actual water quality situation, indicating that the Euclidean distance model with varying weights has potential application in comprehensive assessment of surface water quality.

Key words: Euclidean distance model with varying weights; grey clustering method; fuzzy complex index method; pollution contribution rate; comprehensive assessment of surface water quality

水质综合评价是以定性的方式对水环境的总体状况做出评价, 为水资源污染的控制和管理提供科学依据。由于水质指标的多样性, 选择一个合理

的综合评价方法成为首要的问题。随着新的数学方法在分类学中的广泛应用, 新的水质综合评价方法层出不穷, 如模糊数学法^[1-3]、灰色系统理论

* 收稿日期: 2009-09-22

基金项目: 浙江省自然科学基金重点资助项目 (Z5080048); 国家自然科学基金资助项目 (50809058)

作者简介: 张庆庆 (1984 年生), 女, 博士生; 通讯作者: 许月萍; E-mail: yuepingxu@zju.edu.cn

法^[4-5]、物元理论与可拓集合法等^[6-7]。较常用的有灰色聚类法以及模糊综合指数法等。然而,灰色聚类法和模糊综合指数法的计算过程都相当复杂,虽然编写程序可使运算速度加快,但由于白化函数和隶属度函数的计算均要求水质的污染程度与污染物浓度成正比,而对于像 DO 这样浓度越高水质越好的因子,则需要调整白化函数和隶属度函数的计算公式。若每次水质评价时选取的指标不完全一样,则编写的程序需作相应调整后方可使用,增加了评价工作的难度。

欧式距离是聚类分析中经常使用的概念。Kodaz 等^[8]所提出的基于信息增益的人工免疫识别系统 (IG-AIRS) 正是利用基于信息的欧式距离从众多疾病数据中识别出动脉硬化; Mohamed 等^[9]通过计算输入的英文字母与指定的相关参数之间的欧式距离建立了识别手写字的分割算法; Ruzanski 等^[10]在研究人类语音所反映的压力水平时,在包含频率特征和 Teager 能量算子特征的多属性空间中应用欧式距离实现了对压力水平的分类。同样,水质综合评价,作为一个多属性分类问题,也可以借助欧式距离来完成。如果对欧氏距离赋予权重,则变成加权欧式距离。如果权重是变化的,则变成变权欧式距离。本文将利用变权欧式距离模型对水质进行评价,并将变权欧式距离模型、灰色聚类法和模糊综合指数法应用于钱塘江支流东阳江水质评价,最后对三个评价结果进行比较和理论分析。

1 评价方法简介

1.1 变权欧式距离模型

为了将欧式距离应用于水质综合评价,必须考虑不同污染因子对于水质影响程度的不同,即污染因子的权重。利用变权欧式距离模型对水质进行综合评价的原理为:对于某一个待评价单元(即某个监测断面在某个监测时间的各污染因子的监测值),根据污染因子的污染贡献率计算其权重,该权重对于每个监测值都有所不同,所以称为变权。另外,在根据污染贡献率计算权重时,应将 DO 这种浓度越高污染越小的污染因子与其他浓度越高污染越重的污染因子区分开来。建立以各个污染因子为坐标轴的 n 维空间,以水质标准中的 I 级浓度限制为“原点”,然后结合污染因子的权重系数计算每个待评价单元以及水质评价标准中其他各级水质与“原点”间的变权欧式距离,并将后者作为划分每个待评价单元水质级别的标准。其计算步骤如下:

1.1.1 建立欧式空间 记 $i=1, 2, \dots, m$ 为待评价单元; $j=1, 2, \dots, n$ 为参与评价的污染因子; $k=1, 2, \dots, 5$ 为根据地表水质量标准划分的水质等级; 设实测样本矩阵为 C , c_{ij} ($i=1, 2, \dots, m$; $j=1, 2, \dots, n$) 为第 i 个待评价单元第 j 个污染因子的实测浓度值。根据国家规定的水质分级标准,设含有 n 个污染因子的 5 级水质标准矩阵为 B , 其中 b_{jk} ($j=1, 2, \dots, n$; $k=1, 2, \dots, 5$) 为第 j 个污染因子的第 k 级标准浓度限值。将 C 中的每一行和 B 中的每一列均看作以 n 个污染因子为坐标轴的欧式空间中的点。

1.1.2 欧式空间坐标转换 将水质标准矩阵 B 的 I 级浓度限值作为欧式空间的原点,则 C 和 B 的坐标变为:

$$c'_{ij} = \begin{cases} 0 & c_{ij} \leq b_{j1} \\ \frac{c_{ij} - b_{j1}}{b_{j5} - b_{j1}} & c_{ij} \geq b_{j1} \end{cases} \quad (1)$$

$$b'_{jk} = \frac{b_{jk} - b_{j1}}{b_{j5} - b_{j1}} \quad (2)$$

由公式 (2) 可知 b'_{j1} ($j=1, 2, \dots, n$) 全为零, b'_{j5} ($j=1, 2, \dots, n$) 全为 1。

1.1.3 权重的确定 根据污染贡献率计算各个监测值的权重,即某监测值超标倍数越多,其权重越大^[2]:

$$\omega_{ij} = (c_{ij}/b_{j1}) / \left(\sum_{j=1}^n c_{ij}/b_{j1} \right) \quad (3)$$

式中, ω_{ij} , c_{ij} 分别为第 i 个待评单元中第 j 个污染因子的权重和实测浓度; c_{ij}/b_{j1} 反映了污染因子的超标量。

对于像 DO 这样浓度越高,水质越好的因子,上述权重系数的公式应当作出调整。因此,假设第 λ ($1 \leq \lambda \leq n$) 个因子浓度越高,水质越好,则第 i 个待评价单元中 DO 实测值的权重为:

$$\omega_{i\lambda} = \frac{b_{\lambda 1}/c_{i\lambda}}{b_{\lambda 1}/c_{i\lambda} + \sum_{j=1}^{\lambda-1} c_{ij}/b_{j1} + \sum_{j=\lambda+1}^n c_{ij}/b_{j1}} \quad (4)$$

1.1.4 变权欧式距离的确定 利用前述得出的 ω_{ij} 分别计算第 i 个待评价单元转换后的坐标和第 k 级水质标准转换后的坐标到原点的变权欧式距离^[11]:

$$\delta_i = \sqrt{\sum_{j=1}^n (\omega_{ij} c'_{ij})^2} \quad (5)$$

$$\Delta_{ik} = \sqrt{\sum_{j=1}^n (\omega_{ij} b'_{jk})^2} \quad (6)$$

可见 $0 \leq \Delta_{ik} \leq 1$ 。将 δ_i 与 Δ_{ik} 比较即可判断第 i 个待评价单元的水质级别。

根据评价因子的污染贡献率,由式(3)和式(4)计算所得的权重系数见表 3。

根据表 2 和 3,公式(6)计算出各个待测单元的

各级水质标准与 I 级水质标准的变权欧式距离 Δ , 见表 4。

表 3 各待评价单元中各污染因子的权重系数

Table 3 Weighting coefficients of each pollution factor in each unit under review

年	月	COD _{Mn}	BOD ₅	NH ₃ -N	COD _{Cr}	TN	TP	DO
1996	7	0.098 5	0.036 3	0.039 7	0.103 3	0.444 1	0.164 5	0.113 5
1999	4	0.092 3	0.025 6	0.291 1	0.078 4	0.414 2	0.015 9	0.082 6
2000	12	0.196 5	0.145 9	0.064 5	0.093 9	0.283 0	0.096 0	0.120 2
2004	3	0.029 5	0.022 7	0.395 5	0.064 9	0.439 8	0.019 2	0.028 5
2004	7	0.076 4	0.036 4	0.417 3	0.119 2	0.239 4	0.021 7	0.089 6
2005	3	0.057 2	0.133 3	0.155 2	0.149 1	0.383 0	0.060 0	0.062 2
2005	7	0.064 2	0.036 4	0.251 4	0.100 0	0.366 5	0.049 9	0.131 6
2006	7	0.113 8	0.069 9	0.196 2	0.141 0	0.236 9	0.102 8	0.139 4
2006	11	0.044 8	0.054 0	0.260 1	0.105 6	0.431 7	0.050 1	0.053 6

表 4 变权欧式距离 Δ

Table 4 Euclidean distance with varying weights Δ

年	月	I (Δ_1)	II (Δ_2)	III (Δ_3)	IV (Δ_4)	V (Δ_5)
1996	7	[0,0.075 9]	[0.075 9,0.201 6]	[0.201 6,0.333 9]	[0.333 9,0.469 6]	[0.469 6,1]
1999	4	[0,0.089 4]	[0.089 4,0.229 9]	[0.229 9,0.374 4]	[0.374 4,0.521 1]	[0.521 1,1]
2000	12	[0,0.057 3]	[0.057 3,0.145 4]	[0.145 4,0.256 3]	[0.256 3,0.391 1]	[0.391 1,1]
2004	3	[0,0.105 2]	[0.105 2,0.267 8]	[0.267 8,0.432 3]	[0.432 3,0.597 2]	[0.597 2,1]
2004	7	[0,0.092 8]	[0.092 8,0.225 8]	[0.225 8,0.368 5]	[0.368 5,0.511 2]	[0.511 2,1]
2005	3	[0,0.074 7]	[0.074 7,0.192 9]	[0.192 9,0.326 2]	[0.326 2,0.470 7]	[0.470 7,1]
2005	7	[0,0.086 7]	[0.086 7,0.211 7]	[0.211 7,0.349 4]	[0.349 4,0.482 5]	[0.482 5,1]
2006	7	[0,0.072 2]	[0.072 2,0.166 5]	[0.166 5,0.285 5]	[0.285 5,0.402 9]	[0.402 9,1]
2006	11	[0,0.089 5]	[0.089 5,0.230 1]	[0.230 1,0.376 6]	[0.376 6,0.524 9]	[0.524 9,1]

表 5 给出了利用式 (1) 对实测样本表 1 进行坐标转换后由式 (5) 计算出的每个待测单元与 I 级水质标准的变权欧式距离 δ 然后利用式 (7) 对每个待测单元的水质进行评价得到的水质评价结果。表 5 同时也给出了灰色聚类法和模糊综合指数法的水质评价结果。

2.2 评价结果

通过比较可以看出,变权欧式距离模型的评价结果与灰色聚类法、模糊综合指数法的评价结果基本一致,这说明了变权欧式距离模型应用于水质综合评价的合理性。根据 3 种评价方法的评价过程以及评价结果可得到以下结论:

1) 模糊综合指数方法以及变权欧式距离模型都是变权的评价方法,其权重的计算方法充分体现了超标污染物对水质的影响。灰色聚类法则认为某个污染因子在水质标准中的标准值越小,则水质对其越敏感,该因子所占权重也就越大,并根据不同

表 5 1996-2006 年东阳江许村断面水质评价结果

Table 5 Water quality evaluation results of Xucun section of the Dongyang River from 1996 to 2006

年	月	距离 δ_i	级别 H	本法评 价结果	灰色聚 类法	模糊综合 指数法
1996	7	0.353 1	4.141 4	4	3	3
1999	4	0.349 3	3.826 7	3	2	2
2000	12	0.112 7	2.628 7	2	2	2
2004	3	2.946 8	-	5	5	5
2004	7	0.816 5	5.624 7	5	4	5
2005	3	0.651 0	5.340 8	5	3	4
2005	7	0.616 5	5.259 0	5	4	4
2006	7	1.864 7	-	5	4	5
2006	11	1.917 4	-	5	5	5

的水质类别确定相应级别的各污染指标的权重。这种权重计算方法显然没有考虑超标污染物对水质的损害,所以,灰色聚类法评价水质容易偏轻。

2) 变权欧式距离模型按照污染贡献率计算污染因子的权重有助于对水质进行单因子分析, 确定主要致污因子。比较表3的数据, TN和 $\text{NH}_3\text{-N}$ 的污染贡献率要明显高于其他5项污染因子, 是东阳江许村断面在1996-2006年期间的主要致污因子。

3) 从表5来看, 虽然用3种方法得到2004年7月和2005年7月的水质评价结果均为V类, 但变权欧式距离模型能得到更进一步的结论: 前者的 H 值大于后者的 H 值, 故前者的水质要劣于后者; 二者的 H 值都远远超过了5, 所以二者均属于劣V类水质。但是, 对于2004年3月、2006年7月和11月这种 $\delta_i > 1$ 的情况, 只能判断均为V类水质却不能判断污染程度孰重孰轻, 但一般情况下, δ_i 越大则污染程度越重的可能性比较大。

3 结 论

通过变权欧式距离模型的应用, 以及将其评价结果与灰色聚类法、模糊综合指数法的评价结果进行比较, 验证了变权欧式距离模型应用于水质综合评价的科学性和合理性。综上分析, 变权欧式距离模型具有如下优点:

1) 变权欧式距离模型不仅全面考虑了各污染因子, 而且强调了超标污染物对水质的影响, 便于对水质进行因子分析。

2) 当 $\delta_i > 1$ 时, 变权欧式距离模型评价结果中的 H , 不仅给出水质级别, 而且能够给出待评价单元对于某级别水质的贴近度, 详细地体现了水质的变化趋势。

3) 变权欧氏距离模型形象、直观、方便、准确率较高, 能够较完整地反映水环境污染程度, 适用于水质综合评价。

参考文献:

[1] ADEBOWALE K O, AGUNBIADE F O, OLU-OWOLABI B I. Fuzzy comprehensive assessment of metal contamination of water and sediments in Ondo Estuary, Nigeria

- [J]. *Chemistry and Ecology*, 2008, 24(4): 269-283.
- [2] 朱雷, 陈威. 模糊综合指数法在水质评价中的应用[J]. *武汉理工大学学报*, 2001, 23(8): 61-65.
- [3] SASIKUMAR K, MUJUMDAR P P. Application of fuzzy probability in water quality management of a river system [J]. *Systems Science*, 2000, 31(5): 575-597.
- [4] 许文杰, 陈文国. 应用灰色聚类法评价湖泊水体富营养化[J]. *山东建筑大学学报*, 2007, 22(1): 49-51.
- [5] 刘金英, 杨天行, 李明, 等. 一种加权绝对灰色关联度及其在密云水库水质评价中的应用[J]. *吉林大学学报: 地球科学版*, 2005, 35(7): 54-58.
- [6] SUN Xiuling, TAN Yongming, XU Xiaochi. Matter Element Extension Evaluating Model Based on Synthesis Weights and Approach Degree and its Application [C]. *The 2nd International Conference on Bioinformatics and Biomedical Engineering*. Shanghai, China: Inst of Elec and Elec Eng Computer Society, 2008: 3661-3665.
- [7] 张龙云, 曹升乐. 物元可拓法在黄河水质评价中的改进及其应用[J]. *山东大学学报: 工学版*, 2007, 37(6): 91-96.
- [8] KODAZ H, KARA S, LATIOGLU F, et al. A new hybrid classifier system: Information gain-based artificial immune recognition system [J]. *Experimental Techniques*, 2007, 31(6): 36-43.
- [9] MOHAMED S, ABDELKRIM L, ALI B H, et al. A segmentation method to handwritten word recognition [J]. *Neural Network World*, 2007, 17(3): 225-236.
- [10] RUZANSKI E, HANSEN J H L, MEYERHOFF J, et al. Stress level classification of speech using Euclidean distance metrics in a novel hybrid multi-dimensional feature space [C]//2006 IEEE International Conference on Acoustics, Speech, and Signal Processing. Toulouse: Institute of Electrical and Electronics Engineers Inc, 2006: 1425-1428.
- [11] KURNIAWATI R, JIN J S, SHEPHERD J A. An efficient nearest-neighbour search while varying Euclidean metrics [C]//Proceedings of the sixth ACM international conference on Multimedia. New York: Association for Computing Machinery, 1998: 411-418.