

基于多元统计分析的黄河水质评价方法

孙国红¹, 沈跃², 徐应明², 齐爱军³

(1.天津农学院 基础科学系, 天津 300384 2.农业部环境保护科研监测所, 天津 300191 3.南开大学数学科学学院, 天津 300071)

摘要 为探讨一种适用于大尺度、多断面和长时间的水质评价方法,用系统聚类分析将 2000—2002 年黄河 6 个监测断面的 90 个水质样本分为 7 组,并用判别分析验证了结果的可靠性。其主要程序为 利用多元方差分析对各断面多年水质监测样本进行空间尺度上的显著差异性分析,识别出具有显著差异的样本,然后通过系统聚类分析把上述样本进行聚类分组,最后应用判别分析方法对各组样本进行水质评价,此方法的特点为在不损失信息的前提下能大大减轻水质评价工作量,且客观可信、分辨率高,并能综合反映总体与个别特征。结果表明,黄河流域干流从上游到下游水质总体状况呈逐渐下降趋势,上游水质一般为 I~III 类,而中游和下游水质基本为 IV~V 类和超 V 类。

关键词 水质评价 多元方差分析 系统聚类分析 判别分析 黄河流域

中图分类号 X824 **文献标志码** A **文章编号** 1672-2043(2011)06-1193-07

Water Quality Assessment of Yellow River Based on Multivariate Statistical Analysis

SUN Guo-hong¹, SHEN Yue², XU Ying-ming², QI Ai-jun³

(1.Department of Basic Science, Tianjin Agricultural University, Tianjin 300384, China; 2.Institute of Agro-environmental Protection, Ministry of Agriculture, Tianjin 300191, China; 3. School of Mathematical Sciences, Nankai University, Tianjin 300071, China)

Abstract In order to find a new approach for water quality assessment which could be useful for the monitoring of large-scale, long-term and many sections, hierarchical cluster analysis was applied to group 90 water quality samples into 7 clusters. The samples generated from five monitoring sections of Yellow River during 3 years (2000~2002). The main procedures of this approach included: (1) analyzing the spatial differences of independent samples according to multivariate analysis of variance, and recognizing the samples which were statistically significantly different between each others; (2) grouping the former samples into clusters on the basis of similarities with in a cluster and dissimilarities between different clusters based on hierarchical cluster analysis (HCA); (3) modeling the appropriate discriminatory analysis neural networks to evaluate the surface water quality of each class, then feeding back this results to every original samples. Moreover, its particular characteristics were that it could reduce the workload in assessment and comprehensively represent both holistic condition and individual and its result was objective and discriminative. The results were assigned to the five sections. Water quality of the Yellow River belonged to class I and II in Ningxia and Lanzhou section could meet the need of drinking water resource. But water quality of the Huayankou to Luokou section mostly belonged to class III and IV.

Keywords water quality assessment; multivariate analysis of variance; hierarchical cluster analysis; discriminatory analysis; Yellow River

黄河是我国第二大河,年径流量多年平均为 580 亿 m³,流域人均水资源占有量仅为全国人均的 1/4,单位耕地面积水资源量不到全国的 1/5,水资源短缺十分严重。近 10 年来,随着黄河流域社会经济的快速发展,流域废污水排放量急剧增加,加之天然来水量偏

少,流域水质污染日益加重,可利用水资源量减少,加剧了黄河流域水资源危机,也对流域水生生态环境造成巨大破坏。

黄河流域水体污染已经引起国内学者关注,相继开展了有关黄河流域河水离子变化趋势^[1]、黄河干流水体污染时空变化特征^[2]、黄河流域河水氮污染分析^[3]及水质污染状况分析与对策研究^[4]等。但是对黄河干流水质多年变化特征进行系统、详细的评价方法研究还不多见。目前,随着地表水环境质量改善技术的深

收稿日期 2011-01-21

基金项目 国家社会科学公益研究专项 (2002DIB5009)

作者简介 孙国红 (1967—),女,副教授,主要从事概率论与数理统计及多元统计分析等研究。E-mail: sgh1999@126.com

入研究和广泛应用,研究地表水质的时空变化特征及其原因已成为控制水体污染以及加强水质管理的重要手段^[5-6]。此类研究将主成分分析(PCA)、系统聚类分析(HCA)和判别分析(DA)等多元统计方法应用于时空模式分析,能够更好地说明研究区域的水质时空变化和生态状况^[7-8]。概括起来,这些研究对象包括内陆河流^[9-10]、近海海域^[11]、地下水^[12]和湖泊^[7]。迄今为止,基于水质评价的方法有许多种,目前国内外常用的水质评价方法主要包括模糊综合评价法^[13-15]、灰色系统理论法^[16]、投影寻踪模型法^[17]、人工神经网络法^[18]等。由于水质系统是由各种污染指标变量组成的复杂系统,各个因子之间具有不同程度的相关性,每一因子都只从某一方面反映水质情况,从而使上述方法在进行水质评价时表现出一定的局限性。

本文以黄河水质评价为例,提出基于多元统计的分析方法,利用单因素多元方差分析对各监测断面多年水质数据进行空间尺度上的显著差异性分析,识别出具有显著差异的样本,然后通过系统聚类分析(HCA)把上述样本进行聚类分组,并以各组水质数据为输入,应用判别分析得到水质综合评价结果,对黄河水质状况进行评价。此方法较之其他方法的显著特点主要表现在:①充分挖掘了水质数据的时空特征;②简化了计算过程;③结果客观可信且分辨率高,尤其适应于大尺度、多断面、长时间的水质状况评价分析,以达到降维、简化数据和提高分析结果的可靠性的目的。通过该方法在黄河水质评价中的应用,为黄河水质综合治理与水资源开发项目的总体规划与实施提供一定的理论依据。

1 水质评价方法

1.1 单因素多元方差分析^[19]

方差分析(ANOVA)是通过研究一种或多种因素的变化对试验结果的观测值是否有显著影响,从而找

出较优的试验条件或生产条件的一种常用数理统计方法。根据控制变量的个数,多元ANOVA分为单因素和多因素两种(表1),本研究的控制因素主要为空间,属单因素多元ANOVA,其基本思路是:

设有一个试验,其响应是一个 p 维向量,因素 A 有 m 个水平 $A_1, \dots, A_m$,在 A_i 水平下作 n_i 次试验,用 y_{ij} 表示在 A_i 水平下的第 j 次试验结果。因此可视在水平 A_i 下的 n_i 次试验为从总体 $N_p(\mu_i, \Sigma)$ 中的抽样,在 m 个水平下的试验,视为从 m 个不同总体 $N_p(\mu_i, \Sigma)$ 的抽样,其中 m 个总体协差阵 Σ 相同。问题是检验这 m 个水平的均值效应是否相同,即

$$H_0: \mu_1 = \dots = \mu_m \leftrightarrow H_1: \mu_1, \dots, \mu_m \text{ 不全相等。}$$

引入因子水平的效应向量 $\tau_i = \mu_i - \mu$,其中

$$\mu = \frac{1}{n} \sum_{i=1}^m n_i \mu_i, \text{ 又记 } n = \sum_{i=1}^m n_i. \text{ 于是以上假设问题等价}$$

于 $H_0: \tau_i = \dots = \tau_m = 0 \leftrightarrow H_1: \tau_1, \dots, \tau_m \text{ 不全等于 } 0$ 。

为导出该检验问题的统计量,建立如下模型

$$\begin{cases} y_{ij} = \mu + \tau_i + \varepsilon_{ij} \\ \varepsilon_{ij} \sim N_p(0, \Sigma), \text{ 且 } \varepsilon_{ij} \text{ 独立}, i=1, \dots, m; j=1, \dots, n_i \end{cases}$$

称这样的模型为单向分类模型。

显著性检验公式如下:

$$\lambda^{2n} = \frac{|W|}{|B+W|} \sim \Lambda(p, n-m, m-1)$$

式中, m 为水平数, n 为样本数, p 为向量维数, B 为组间离差平方和矩阵, W 为组内离差平方和矩阵。 Λ 统计量服从自由度为 $(p, n-m, m-1)$ 的 Λ 分布,当相应的概率值小于或等于确定的显著水平 α (通常为0.05)时,则表明该控制变量下不同水平各总体均值具有显著性差异,反之,则不存在显著性差异。

1.2 系统聚类分析法^[20]

聚类分析法是一种多元统计分析方法,根据观测对象之间的彼此相似程度达到“物以类聚”的目的,其中层次聚类分析法应用最为广泛。其实质即根据观察

表1 单向多元方差分析表

Table 1 One-way multivariate analysis of variance

方差来源	离差平方和矩阵	自由度	检验统计量
组间(因素影响)	$B = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot}) (y_{ij} - \bar{y}_{i\cdot})'$	$m-1$	
组内(误差)	$W = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot}) (y_{ij} - \bar{y}_{i\cdot})'$	$n-m$	$\lambda^{2n} = \frac{ W }{ B+W }$
总合	$T = W + B = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot}) (y_{ij} - \bar{y}_{\cdot\cdot})'$	$n-1$	

值或变量之间的亲疏程度,以逐次聚合的方法,将最相似的对象结合在一起,直到聚成一类。相似程度的计算包括样本间距离和小类间距离两类,前者的测量方法有欧氏距离、Mahal 距离和 Cityblock 距离等,后者的测量方法有最短距离法、中心法、离差平方法等。本研究采用 SAS 9.1.2 实现聚类分析,样本间距离分别采用欧氏距离平方和离差平方 2 种方法。

1.3 距离判别分析^[21]

判别分析是应用性很强的一种多元统计方法。可以这样来描述一个判别问题:设有 q 个总体 $\pi_1, \dots, \pi_q$, 它们有不同的分布 $F_1(x), \dots, F_q(x)$ 或有不同的其他特征,其中每个 $F_i(x)$ 为 p 元函数。对给定的新样品 x ,需要判定 x 是来自哪一个总体。

在进行判别归类时,由假设的前提、判别的依据及处理的手法不同,可得出不同判别方法,如距离判别、Bayes 判别、Fisher 判别或典型判别、逐步判别、序贯判别等。本文采用的为多个总体情形的距离判别法。

多个总体情形的距离判别基本思想是:设有 q 个总体, $\pi_i \sim (\mu^0, \Sigma)$, $i=1, \dots, q$ 。考虑对给定的新样本 x 进行判别。

当 $\mu^0, \Sigma = \Sigma$ 未知的情形。此时要用历史样本对 μ^0 和 Σ 估计: μ^0 和 Σ 估计为

$$\bar{x}^0 = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^0 \text{ 和 } \hat{\Sigma} = \frac{1}{n-q} \sum_{i=1}^q A_i$$

$$\text{其中 } A_i = \sum_{j=1}^{n_i} (x_j^0 - \bar{x}^0) (x_j^0 - \bar{x}^0)', i=1, \dots, q.$$

本文通过考虑函数

$$\widetilde{W}_{ij}(x) = \left[x - \frac{1}{2} (\bar{x}^0 + \bar{x}^0) \right]' \hat{\Sigma}^{-1} (\bar{x}^0 - \bar{x}^0)$$

和对 R^p 的划分 $R^p = \bigcup_{i=1}^q R_i$, 其中

$R_i = \{x: \widetilde{W}_{ij} \geq 0, i \text{ 固定, 对所有 } j \neq i\}$, 得判别法则:

当 $x \in R_i$ 时, 判 $x \in \pi_i$ 。

1.4 操作程序

归纳起来,基于多元统计分析的水质评价方法的操作程序如图 1。

2 实例研究

2.1 研究区域及采样点

本文研究区域基本覆盖了整个黄河干流。采样点共设置 6 个断面,包括黄河流域上游干流循化段(S1)、兰州段(S2)、黄河流域中游郑州花园口段(S3)、吴堡段(S4)、黄河流域下游入海口山东利津段(S5)、济南泺口段(S6)。

2.2 基础数据预处理

本文采用的水质数据为黄河流域 6 个断面 2000—2002 连续 3 年 COD_{Mn} 、 BOD_5 、 $\text{NH}_3\text{-N}$ 、DO 和挥发酚 5 个污染因子年平均水质监测数据,水质样本共 90 个,具体情况见表 2。

为了从宏观角度了解黄河的水质,本文按照《地表水环境质量标准》(GB 3838—2002) 为依据进行评价与分级,取评价因子为:化学需氧量(COD_{Mn})、五日生化需氧量(BOD_5)、氨氮($\text{NH}_3\text{-N}$)、溶解氧(DO)和挥发酚共 5 个参数。

考虑到水质指标数量级上差异,对原始数据进行标准化,使其均值为 0,方差为 1。本研究的数据分析及编程采用的软件为 Mathworks 公司的 Matlab 2010 和 SAS 9.1.2^[16]。

2.3 数据空间尺度多元方差分析

考虑到样本的空间分异性与相关性,采用单因素多元方差分析对表 2 中数据进行空间尺度方差分析(见表 3)。结果表明,从 4 种统计量的值来看,6 个断面 S1、S2、S3、S4、S5 和 S6 在 2000—2002 3 年内在变量 COD_{Mn} 、 BOD_5 、 $\text{NH}_3\text{-N}$ 、DO 和挥发酚的联合分布下

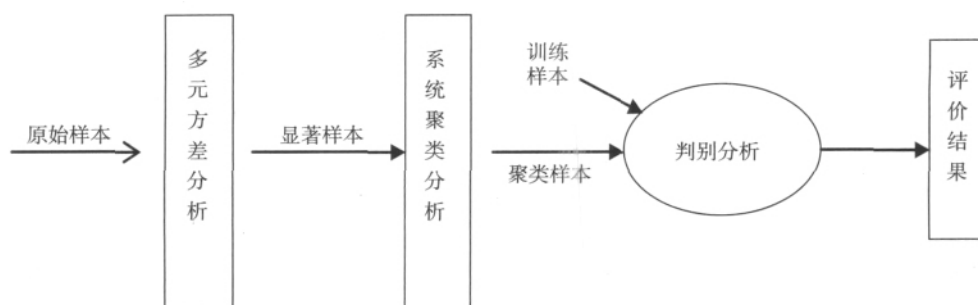


图 1 操作程序

Figure 1 Operating procedures

表 2 2000—2002 年黄河各断面水质监测数据统计描述 (mg·L⁻¹)

Table 2 Statistical descriptive of different water quality parameters at different siteson the Yellow River during 2000~2002 (mg·L⁻¹)

断面	年份	COD _{Mn}		BOD ₅		NH ₃ -N		挥发酚		DO	
		平均值	标准差	平均值	标准差	平均值	标准差	平均值	标准差	平均值	标准差
S1	2000	2.45	1.08	1.28	0.82	0.45	0.23	0.003	0.001	9.80	2.29
	2001	2.08	1.74	1.03	1.47	0.29	0.41	0.002	0.001	9.45	1.93
	2002	1.80	1.53	0.80	0.96	0.37	0.32	0.002	0.001	8.60	1.63
S2	2000	2.57	1.68	1.48	2.30	0.62	0.56	0.004	0.002	8.70	2.82
	2001	2.26	1.29	1.98	1.91	0.56	0.28	0.003	0.002	8.93	3.17
	2002	2.29	1.72	1.88	1.42	0.64	0.39	0.002	0.001	8.84	1.59
S3	2000	3.77	2.28	3.56	1.78	1.09	0.47	0.003	0.002	7.65	3.74
	2001	5.29	1.84	3.43	1.94	0.90	0.35	0.002	0.001	7.87	1.81
	2002	4.73	1.71	4.58	2.73	0.85	0.58	0.001	0.001	8.25	2.40
S4	2000	3.60	1.08	2.00	1.59	0.69	0.59	0.001	0.00	8.33	3.67
	2001	3.70	1.93	2.04	1.08	0.86	0.30	0.001	0.00	7.73	2.80
	2002	3.58	2.52	1.77	1.29	1.07	0.64	0.001	0.00	8.54	1.64
S5	2000	3.80	1.61	2.30	0.92	1.08	0.76	0.002	0.001	5.43	2.39
	2001	4.15	1.39	3.08	1.56	1.26	0.72	0.003	0.002	5.02	1.82
	2002	4.72	2.82	3.70	1.74	0.97	0.68	0.006	0.001	6.00	2.04
S6	2000	4.59	1.30	3.41	1.03	1.24	0.92	0.005	0.002	5.36	1.36
	2001	4.56	1.61	3.53	2.20	1.35	0.95	0.004	0.001	7.60	1.60
	2002	5.38	1.85	5.10	1.26	1.04	0.61	0.007	0.001	4.10	1.09

表 3 采样点空间多元方差分析

Table 3 One-way multivariate analysis of variance

Statistic	Value	F Value	Num DF	Den DF	Pr>F
Wilks' Lambda	0.000 2	11.94	25	31.22	<.000 1
Pillai's Trace	3.17	4.17	25	60.00	<.000 1
Hotelling-Lawley	96.53	27.74	25	12.13	<.000 1
Roy's Greatest Root	85.20	204.48	5	12.00	<.000 1

注 *P<0.05,以下同。

皆存在显著差异性。

2.4 聚类分析

通过聚类分析发现(图 2),研究断面间的水质样本至多可聚类成相似的 7 组(分别编号为 G1~G7),其中各组数据特征用组内水质样本均值与标准差表示(表 4)。研究结果表明,上游循化断面 2000 年和 2001 年水质相近,2002 年水质与甘肃兰州断面 2001、2002 年水质相近;中游郑州花园口断面和吴堡断面 2000—2002 年 3 年水质均相近,说明水质变化起伏不大;黄河下游入海口山东利津断面 2001 年和 2002 年水质相近,济南冻口断面 2001 年和 2002 年水质与利津断面 2000 水质相近。

2.5 基于判别分析模型的水质评价

首先,在《地表水环境质量标准》(GB 3838—2002)(表 5) 各级评价标准间随机正态分布内插入

500 个样本,共形成 3 000 个训练样本。此途径解决了以往仅用各级评价标准作为训练样本而导致样本数过少和无法构建检测样本的问题。

其次,依据上述方法生成检测样本,各级标准间随机正态分布内插入 100 个样本,共形成 600 个检测样本,经过检验错判概率为 0.17%,所以判别模型的泛化能力很强。对实际测得的样本进行判别归类,可以获得 6 个类别的线性判别函数(表 6)。

由表 6 可得 6 个类别的线性判别函数为:

$$Y1=-270.139\ 9+2.058\ 6x_1+4.997\ 2x_2+8.623\ 7x_3-18.172\ 4x_4+49.661\ 4x_5.$$

$$Y2=-129.684\ 6+2.121\ 6x_1+4.817\ 9x_2+59.750\ 9x_3-4.212\ 4x_4+32.064\ 1x_5.$$

$$Y3=-186.433\ 7+4.901\ 6x_1+11.115\ 7x_2+140.786\ 2x_3+28.401\ 1x_4+25.926\ 6x_5.$$

$$Y4=-313.326\ 5+7.035\ 9x_1+15.720\ 3x_2+237.836\ 4x_3+80.811\ 1x_4+18.476\ 7x_5.$$

$$Y5=-594.016\ 1+9.799\ 3x_1+25.152\ 1x_2+334.533\ 9x_3+588.826\ 5x_4+11.026\ 5x_5.$$

$$Y6=-1\ 140+12.679\ 0x_1+39.384\ 7x_2+431.100\ 0x_3+1\ 614\ x_4+3.303\ 3x_5.$$

应用线性判别函数对 S1~S6 断面的 18 个观测进行判别归类结果见表 7。

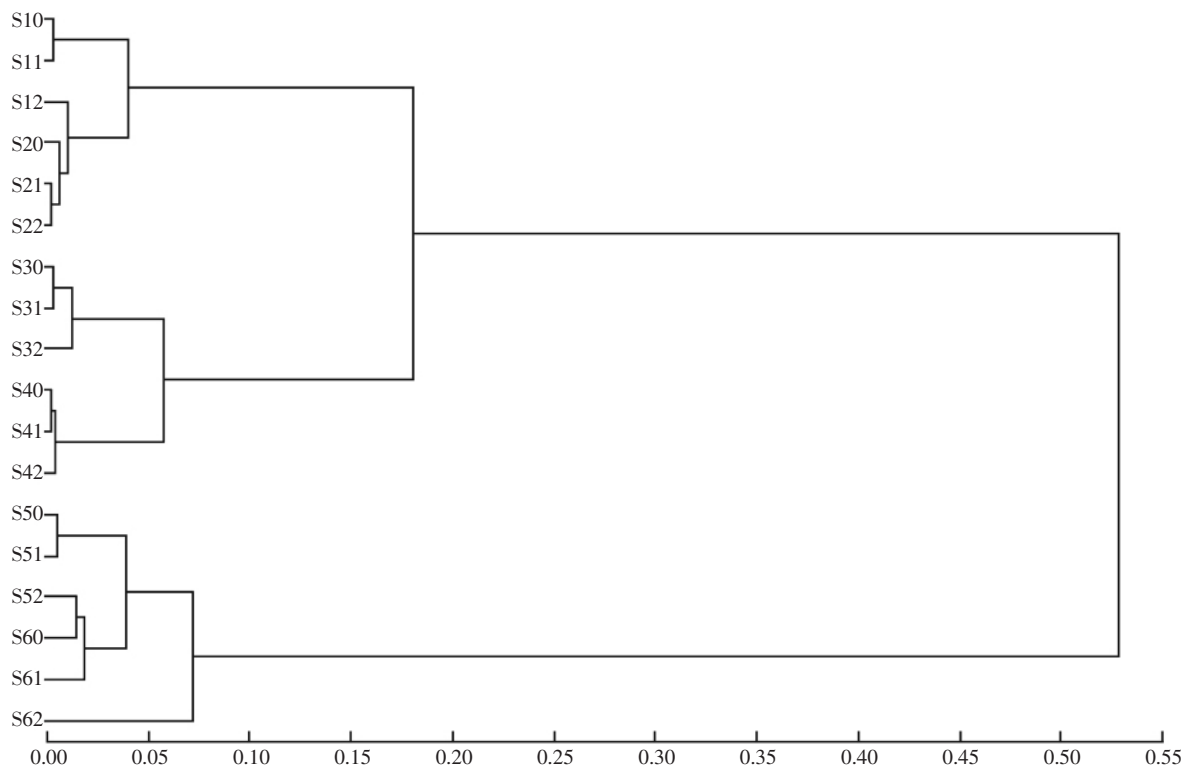


图 2 基于欧氏距离平方和离差平方方法的聚类分析树形图

Figure 2 Dendrogram from cluster analysis based on Euclidean distance and Ward's method

表 4 7 组样本的数据特征及对应水质样本

Table 4 Characteristics of 7 groups from CA and their corresponding samples

组号	COD _{Mn} /mg·L ⁻¹		BOD ₅ /mg·L ⁻¹		NH ₃ -N/mg·L ⁻¹		挥发酚/mg·L ⁻¹		DO/mg·L ⁻¹		对应水质样本
	平均值	标准差	平均值	标准差	平均值	标准差	平均值	标准差	平均值	标准差	
G1	2.23	0.32	1.54	0.54	0.55	0.12	0.003	0.001	8.77	0.15	S21,S22,S20,S12
G2	3.62	0.51	1.94	0.15	0.87	0.19	0.001	0	8.20	0.42	S40,S41,S42
G3	2.27	0.26	1.16	0.18	0.37	0.11	0.006	0	9.63	0.25	S10,S11
G4	4.59	0.48	3.86	0.63	0.95	0.13	0.002	0.001	7.92	0.30	S30,S31,S32
G5	3.98	0.96	2.69	0.55	1.10	0.07	0.003	0.001	5.23	0.29	S50,S51
G6	4.62	1.67	3.55	0.15	1.21	0.32	0.003	0.001	6.32	1.15	S52,S60,S61
G7	5.38	—	5.10	—	1.04	—	0.007	—	4.10	—	S62

注 表中 S_{xy} 表示第 x 个断面 200y 年水质样本。

应用此训练好的检测样本对 7 组样本进行评价，其污染程度为 G7>G5>G6>G4>G3>G2>G1（表 8），从污染趋势来看，以上判别分析与传统的单项指数法

（GB 3838—2002）基本一致，但从计算结果来看，此方法不仅体现了样本总体特征，还能高分辨率的区分相近的样本（如 G1~G3），同时避免了单项指数法以偏差全所造成的失误。

将上述评价结果分解到各个断面可知，黄河流域干流从上游循化段到下游济南泺口段水质总体状况呈逐渐下降趋势，上游水质一般为Ⅰ~Ⅲ类，可满足作为饮用水水源水的要求，而中游和下游水质基本为Ⅳ~Ⅴ类和超Ⅴ类，可以满足水产养殖等渔业用水和一般工业用水、娱乐用水区等的要求。由上述分析可以看出，聚类分析与水质综合标识指数相结合的水质分

表 5 GB 3838—2002 地表水环境质量标准 (mg·L⁻¹)

Table 5 National standard of surface water quality (mg·L⁻¹)

指标	I	II	III	IV	V
COD _{Mn}	15	15	20	30	40
BOD ₅	3	3	4	6	10
NH ₃ -N	0.15	0.5	1.0	0.5	2.0
DO	7.5	6	5	3	2
挥发酚	0.002	0.002	0.005	0.01	0.1

表 6 线性判别函数

Table 6 Linear discriminant function

Variable	1	2	3	4	5	6
Constant	-270.139 9	-129.684 6	-186.433 7	-313.326 5	-594.016 1	-1 140
x ₁	2.058 6	2.121 6	4.901 6	7.035 9	9.799 3	12.679 0
x ₂	4.997 3	4.817 9	11.115 7	15.720 3	25.152 1	39.384 7
x ₃	8.623 7	59.750 9	140.786 2	237.836 4	334.533 9	431.100 0
x ₄	-18.172 4	-4.212 4	28.401 1	80.811 1	588.826 6	1 614
x ₅	49.661 4	32.064 1	25.926 6	18.476 8	11.026 5	3.303 3

表 7 基于判别分析的年度水质样本评价结果 (平均值)

Table 7 Results of water quality assessment based on discriminant analysis for 2000~2002 years (mg·L⁻¹)

断面	年份	COD _{Mn}	BOD ₅	NH ₃ -N	挥发酚	DO	类别
S1	2000	2.45	1.28	0.45	0.003	9.80	1
	2001	2.08	1.03	0.29	0.002	9.45	1
	2002	1.80	0.80	0.37	0.002	8.60	2
S2	2000	2.57	1.48	0.62	0.004	8.70	2
	2001	2.26	1.98	0.56	0.003	8.93	2
	2002	2.29	1.88	0.64	0.002	8.84	2
S3	2000	3.77	3.56	1.09	0.003	7.65	3
	2001	5.29	3.43	0.90	0.002	7.87	3
	2002	4.73	4.58	0.85	0.001	8.25	3
S4	2000	3.60	2.00	0.69	0.001	8.33	2
	2001	3.70	2.04	0.86	0.001	7.73	2
	2002	3.58	1.77	1.07	0.001	8.54	3
S5	2000	3.80	2.30	1.08	0.002	5.43	4
	2001	4.15	3.08	1.26	0.003	5.02	4
	2002	4.72	3.70	0.97	0.006	6.00	4
S6	2000	4.59	3.41	1.24	0.005	5.36	3
	2001	4.56	3.53	1.35	0.004	7.60	3
	2002	5.38	5.10	1.04	0.007	4.10	3

表 8 判别分析与传统的单项指数法比较

Table 8 Results of water quality assessment based on discriminant analysis

方法	G1	G2	G3	G4	G5	G6	G7
国家标准方法	II	III	II	III	IV	IV	V
本文方法	II	II	II	III	IV	III	IV

析方法符合实际情况,证明了此方法的科学性和实用性。

3 结论

(1) 将系统聚类分析方法用于大尺度、多断面、长时间的水质分析,利用单因素多元 ANOVA 对各断面多年水质监测样本进行空间尺度上的显著差异性分析,识别出具有显著差异的样本。然后,通过聚类分析

把上述样本进行聚类分组,各组样本特征用组内样本均值表示,并以此为输入,应用判别分析得到各组水质评价结果。聚类分析和判别分析的结果得到了相互验证,结果客观可信、分辨率高,且能综合反映水质变化总体与个别特征。

(2) 选择合适的聚类分析方法,将存在显著差异的 90 个样品分为 7 组,并通过方差分析验证了聚类分析结果的可靠性,通过聚类分析大大减轻了水质评价的工作量。

(3) 在判别分析中,各级标准间随机正态分布内插 100 个样本,共形成 600 个检测样本,经过检验错判概率为 0.17%,所以判别模型的泛化能力很强,对实际测得的样本进行判别归类。评价结果为 G7>G5>G6>G4>G3>G2>G1,不仅与传统的单项指数法结果基

本一致,而且证实了该方法的科学性与高分辨能力。

(4) 基于判别分析模型对黄河流域干流水质状况进行分析表明,黄河流域干流从上游循化段到下游济南冻口段水质总体状况呈逐渐下降趋势,上游水质一般为Ⅰ~Ⅲ类,而中游和下游水质基本为Ⅳ~Ⅴ类和超Ⅴ类。

参考文献:

- [1] 陈静生,李荷碧,夏星辉,等.近30年来黄河水质变化趋势及原因分析[J].环境化学,2000,19(2):97-102.
CHEN Jing-sheng, LI He-bi, XIA Xing-hui, et al. A study on water-quality trend in the Yellow River system from 1960'S to 1990'S[J]. *Environmental Chemistry*, 2000, 19(2):97-102.
- [2] 李春晖,杨志峰.黄河干流水体污染时空变化特征[J].水资源与水工程学报,2004,15(2):10-14,20.
LI Chun-hui, YANG Zhi-feng. Changes of water pollution in the main channels of the Yellow River [J]. *Journal of Water Resources & Water Engineering*, 2004, 15(2):10-14, 20.
- [3] 夏星辉,周劲松,杨志峰,等.黄河流域河水氮污染分析[J].环境科学学报,2001,21(5):563-568.
XIA Xing-hui, ZHOU Jing-song, YANG Zhi-feng, et al. An analysis of nitrogen pollution in river water of the Yellow River basin [J]. *Acta Scientia Circumstantia*, 2001, 21(5):563-568.
- [4] 吴青,张青云,李立阳,等.近十年黄河流域水质状况及变化趋势分析[J].安阳工学院学报,2006,4(1):1-5.
WU Qing, ZHANG Qing-yun, LI Li-yang, et al. Study on quality and development tendency of water resources in the Yellow River basin in decade years[J]. *Journal of Anyang Institute of Technology*, 2006, 4(1):1-5.
- [5] 周丰,郭怀成,刘永,等.基于多元统计分析和RBFNNs的水质评价方法[J].环境科学学报,2007,27(5):846-853.
ZHOU Feng, GUO Huai-cheng, LIU Yong, et al. A new approach for water quality assessment based on multivariate statistical analysis and Radial Basis Function Neural Networks[J]. *Acta Scientia Circumstantia*, 2007, 27(5):846-853.
- [6] Chen D J, Lu J, Yuan S F, et al. Spatial and temporal variations of water quality in Cao River of eastern China [J]. *Journal of Environmental Sciences*, 2006, 18(4):680-688.
- [7] Kazi T G, Arain M B, Jamali M K, et al. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study[J]. *Ecotoxicology and Environmental Safety*, 2009, 72:301-309.
- [8] Paul Bierman, Megan Lewis a, Bertram Ostendorf, et al. A review of methods for analysing spatial and temporal patterns in coastal water quality[J]. *Ecological Indicators*, 2011, 11:103-114.
- [9] Singh K P, Malik A, Mohan D, et al. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India): A case study[J]. *Water Research*, 2004, 38:3980-3992.
- [10] Kowalkowski T, Zbytniewski R, Szejna J, et al. Application of chemometrics in riverwater classification[J]. *Water Research*, 2006, 40:744-752.
- [11] Reghunath R, Murthy T R S, Raghavan B R. The Utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka[J]. *India Water Research*, 2002, 36:2437-2442.
- [12] Liu C W, Lin K H, Kuo Y M. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan [J]. *Science of the Total Environment*, 2003, 31(3):77-89.
- [13] 李莲芳,曾希柏.利用模糊综合评判法评价潮白河流域水质[J].农业环境科学学报,2006,25(2):471-476.
LI Lian-fang, ZENG Xi-bai. Water quality assessment in Chaobai River by fuzzy synthetic evaluation method [J]. *Journal of Agro-Environment Science*, 2006, 25(2):471-476.
- [14] Elisabetta Giusti, Stefano Marsili-Libelli. Spatio-temporal dissolved oxygen dynamics in the Orbetello lagoon by fuzzy pattern recognition[J]. *Ecological Modelling*, 2009, 220:2415-2426.
- [15] 周贵忠,张金恒,王军强,等.利用模糊数学评价大沽河干流水质的研究[J].农业环境科学学报,2010,29(增刊):191-195.
ZHOU Gui-zhong, ZHANG Jin-heng, WANG Jun-qiang, et al. Application of the fuzzy mathematics in evaluation Dagur River water quality[J]. *Journal of Agro-Environment Science*, 2010, 29(Suppl):191-195.
- [16] W C Ip, B Q Hu, H Wong, et al. Applications of grey relational method to river environment quality evaluation in China [J]. *Journal of Hydrology*, 2009, 379:284-290.
- [17] ZHANG Chi, DONG Sihui. A new water quality assessment model based on projection pursuit technique[J]. *Journal of Environmental Sciences Supplement*, 2009: S154-S157.
- [18] Sundarambal Palani, Shie-Yui Liong, Pavel Tkach. An ANN application for water quality forecasting[J]. *Marine Pollution Bulletin*, 2008, 56:1586-1597.
- [19] 张润楚.多元统计分析[M].北京:科学出版社,2003.
ZHANG Run-chu. Multivariate statistical analysis[M]. Beijing: Science Press, 2003.
- [20] 高惠璇.应用多元统计分析[M].北京:北京大学出版社,2006.
GAO Hui-xuan. Application of multivariate statistical analysis[M]. Beijing: Beijing University Press, 2006.