

聚类分析方法在济宁市水质分析中的应用

刘铭¹,朱长军¹,顿珠加措²,尼玛次仁³

(1. 河北工程大学 城建学院,河北 邯郸 057150;2. 西藏自治区水文水资源勘测局日喀则水文水资源分局,西藏 日喀则 857000;3. 西藏自治区水利电力规划勘测设计研究院 西藏 拉萨 850009)

摘要:根据济宁市现有的水质监测站点的监测资料,应用聚类分析方法对济宁市水质状况进行分析评价。首先利用统计软件 SPSS 对所选取的监测站点进行聚合分类;其次用综合指数评价法对监测站点水质划分级别;最后综合聚类结果及综合指数划级的结果将济宁市各监测站点分为六大类。在区域水环境评价中应用聚类分析法对监测站点分类,符合不同特性污染源地域分布的实际,避免单一运用综合指数法产生的缺陷。

关键词:聚类分析;综合指数;SPSS;水质评价

中图分类号:X824

文献标识码:A

The application of cluster analysis methods in the analysis of water quality in Jining city

LIU Ming¹, ZHU Chang-jun¹, DUNZHU Jia-cuo², NIMA Ci-ren³

(1. College of Urban Construction, Hebei University of Engineering, Hebei Handan 056038, China; 2. Xigaze Branch of the Tibet Autonomous Region Hydrology Bureau, Tibet Xigaze 857000, China; 3. Tibet Investigation and Design Institute of Water Conservancy and Electric Power, Tibet Lhasa 850000, China)

Abstract: According to the existing water quality monitoring data from monitoring sites in Jining city, the clustering analysis method is employed to analyze and evaluate the water quality of Jining city. First of all, by using the statistical software SPSS, we aggregate and classify these monitoring sites; then, the composite index is applied in evaluating the water quality of these monitoring sites; at last, the water quality of Jining is divided into six categories by the clustering results and the comprehensive evaluation combined together. In the regional water environment evaluation, monitoring sites are classified by using clustering method conforms to the distribution actual of different pollution sources, which avoids the defect caused by only using the comprehensive index method.

Key words: clustering analysis; composite index; SPSS; water quality assessment

水质评价是指按照评价目标,选择相应的水质参数,采用相应的水质标准和评价方法,对水体的质量利用价值及水的处理要求做出评定,使用广泛的常规环境质量评价方法大致分为综合指数法、数理统计法、分级评价法和其他方法四大类^[1]。在复杂的环境系统中存在很多模糊性,例如环境影响的“大”与“小”、污染物浓度的“高”与“低”、方案的“优”与“劣”等均无法用确定的界限加以划分^[2]。在对区域水环境质量进行水质评价

时,仅将综合指数评价法应用于环境质量综合评价,其理论方法和实际结果存在着一定的缺陷,往往难以实现对区域水环境的总体评价。目前,聚类分析法已被广泛应用于环境质量评价中,聚类分析方法能够较好评价生物指标,综合考虑生化指标,有助于合理评价水质分布与特性^[3];通过聚类分析能确定海域主要污染指标和污染程度,能够较合理地反映海域的环境质量特征^[4];许多土壤研究者与环境工作者运用聚类分析在土壤质

量评价中开展了卓有成效的工作^[5]。大量的研究表明,多指标存在时应用聚类分析方法能够对环境质量做出更为客观、准确的综合评价。本次水质分析采用聚类分析的统计方法,通过对水质监测监测数据的统计分析,研究比较各监测站点水质指标数据的相关关系,对具有相近数据关系的监测站点进行归类,另外结合监测站点水质综合评价结果和区域内地理位置及当地水污染实际情况,从而对区域水环境质量做出总体初步评价。

1 监测站点资料

济宁市位于山东省西南部,京杭大运河穿过济宁市,素有“运河之都”的美誉,我国北方最大的淡水湖南四湖也位于其中。随着城市的发展,济宁市水污染情况越来越严重,同时南水北调工程的启动也对济宁市的水质提出严格要求。济宁市内河流纵横分布,各段污染状况不同,水质监测站点的布点也因此密布整个济宁市而且数量较多,这最终导致水质评价时会出现繁杂等困难。本次

分析选取济宁市水质监测数据,共选取包括 25 个地下水监测站点的水质监测结果,选取 Cl 、 SO_4 、总硬度、 BOD_5 、 COD_{Mn} 、氨氮、亚硝酸盐氮、硝酸盐氮、酚共 9 项水质监测指标,济宁市监测站点区域分布见下图 1,各个监测站点水质监测数据见表 1:



图1 济宁市监测站点分布图(主要污染源已标注名称)
Fig.1 The distribution map of monitoring sites in Jinan (These main pollution sources have been marked)

表 1 监测站点水质监测数据

Tab. 1 The water quality monitoring data from monitoring sites

监测点名称	编号	县市区	Cl	SO_4	总硬度	BOD_5	COD_{Mn}	氨氮	亚硝酸盐氮	硝酸盐氮	酚
二十里铺徐家	1	任城区	25.5	5.8	262	0.3	1.0	0.05	0.003	0.40	0.002
长沟镇水利站	2	任城区	26.6	15.4	260	0.8	1.9	0.05	0.003	0.40	0.003
航运局港航工处	3	任城区	30.1	19.2	276	0.4	1.3	0.05	0.003	0.29	0.002
大象集团	4	任城区	26.6	26.9	276	0.1	0.8	0.05	0.003	0.34	0.002
运河电厂	5	任城区	85.4	188.0	573	0.2	1.0	0.05	0.003	3.18	0.002
李营镇水利站	6	任城区	18.8	9.6	302	1.4	1.2	0.05	0.003	3.25	0.002
心酒厂	7	任城区	31.2	23.1	300	0.4	1.0	0.05	0.003	1.51	0.002
液压件厂	8	任城区	17.7	30.7	288	0.9	1.9	0.05	0.003	2.37	0.002
济宁农校	9	任城区	28.7	19.2	106	0.3	3.3	0.05	0.003	0.32	0.002
实验中学	10	任城区	17.7	28.8	286	0.5	1.7	0.05	0.003	2.40	0.002
鲁抗 16 号井	11	任城区	82.2	42.3	400	1.7	0.9	0.05	0.003	0.87	0.005
接庄镇敬老院内	12	任城区	13.5	26.9	240	0.4	1.1	0.05	0.003	0.09	0.002
唐口镇水利站	13	任城区	32.3	23.1	284	0.4	1.0	0.05	0.003	0.05	0.002
化肥厂	14	任城区	64.5	117.0	362	0.4	0.7	0.05	0.003	3.29	0.002
二号井	15	任城区	51.0	110.0	1398	0.2	1.5	0.05	0.003	1.28	0.002
博古庄西	16	市中区	88.3	99.9	685	0.6	1.3	0.05	0.017	18.70	0.002
前铺村东 50 米	17	市中区	141.0	19.2	524	0.2	0.7	0.05	0.003	13.00	0.002
省水稻研究所	18	市中区	76.2	38.4	368	0.2	1.0	0.05	0.003	6.18	0.002
济宁石油库院内	19	市中区	231.0	23.1	20	0.2	1.0	0.05	0.003	7.73	0.002
香港大厦前院	20	市中区	36.5	48.0	370	0.3	1.0	0.05	0.160	0.46	0.002
电业局高新变电所	21	市中区	11.0	23.1	280	0.2	0.6	0.05	0.003	0.30	0.002
兴东居委	22	市中区	187.0	121.0	867	1.8	1.0	0.05	0.003	30.90	0.005
国茂羽绒厂区塔	23	市中区	85.4	46.1	424	0.2	1.2	0.05	0.003	6.76	0.002
市儿童医院东北	24	市中区	35.4	23.1	296	0.2	0.8	0.05	0.003	1.64	0.002
纺织器材厂院内	25	市中区	89.0	215.0	536	0.3	1.2	0.05	0.003	0.50	0.002

单位均为 mg/L

2 聚类分析与综合评价

2.1 监测站点的聚类分析

聚类是按照某个特定标准把一个数据分割成不同的类或簇,使得类内相似性尽可能的大,同时类间的差异性也尽可能的大^[6]。聚类分析方法的基本思想是研究对象之间存在着不同程度的相似性,通过分析比较各事物之间的性质,将性质相近的归为一类,将性质差别较大的归入不同的类,聚类分析中常见的有系统聚类法和 K 均值聚类法。本次数据分析选用系统聚类法,这是目前实践中使用最多的,该方法的分析步骤是:先将所有样本看成一类,选用一定的方法计算样本与样本之间的距离和类与类之间的距离;开始,每个样本自成一类,类类之间和样本之间的距离相同;然后,在所有的类中选择距离最小的合并为一个新类,并计算出其与其它类的距离;接着再将距离接近的类合并,直至将所有样本合并为一类为止。由聚类系谱图可以清楚的看出全部样本的聚集过程,进而做出样本的分类。

由于水质各项指标监测数据存在计量量纲的不同,数值大小差异较大,因此需要首先对数据进行标准化处理,本次选用极差标准化法处理数据。假设有 m 个聚类对象,每个对象有 n 个要素构成,其所对应要素数据如表 2 所示。

表 2 聚类对象与要素数据

Tab.2 The clustering objects and data elements

聚类对象	要素					
	x_1	x_2	\wedge, x_f	\wedge, x_n		
1	x_{11}	x_{12}	\wedge, x_{1j}	\wedge, x_{1n}		
2	x_{21}	x_{22}	\wedge, x_{2j}	\wedge, x_{2n}		
M	M	M	M	M	M	M
i	x_{i1}	x_{i2}	\wedge, x_{ij}	\wedge, x_{in}		
M	M	M	M	M	M	M
m	x_{m1}	x_{m2}	\wedge, x_{mj}	\wedge, x_{mn}		

极差标准化公式如下:

$$X'_{ij} = \frac{X_{ij} - \min_{1 \leq i \leq n} X_{ij}}{\max_{1 \leq i \leq n} X_{ij} - \min_{1 \leq i \leq n} X_{ij}}$$

式中 X_{ij} 为监测指标原值,其中 $1 \leq i \leq m, 1 \leq j \leq n$; X'_{ij} 为标准化后数据; $\min_{1 \leq i \leq n} X_{ij}$ 为监测指标的最小值; $\max_{1 \leq i \leq n} X_{ij}$ 为检测指标的最大值。

经过标准化处理所得的数据,极大值为 1,极小值为 0,其余的数值均在 0 与 1 之间。

系统聚类分析的依据和基础是数据之间的差

异性,差异性的测度即为距离的计算,距离较大则样本相似性越小。距离的计算有多种选择方法,本次选用 Pearson 相关性来计算各样本之间距离。

系统聚类的过程中类与类之间距离计算方法可以分为单连接法、完全连接法、平均连接法、组平均连接法和离差平方和法等,本次计算类与类之间距离选用组平均连接法,该方法采用的距离定义为两类之间的平均平方距离,如下:

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{i \in G_p} \sum_{j \in G_q} d_{ij}^2$$

$$\text{递推公式: } D_{kr}^2 = \frac{n_p}{n_q} D_{kp}^2 + \frac{n_p}{n_q} D_{kq}^2$$

根据选取的济宁市各个监测站点得到水质监测数据,利用统计分析软件 SPSS 进行聚类分析计算,计算过程如下:

- (1)用极差标准化方法对各水质指标的原始数据进行处理;
- (2)采用 Pearson 相关性测度监测站点之间的样本间距离;
- (3)选用组平均法计算类间的距离,并对样本进行归类。

经过上述聚类个步骤方法的选择得到聚类的结果如图 2 示:

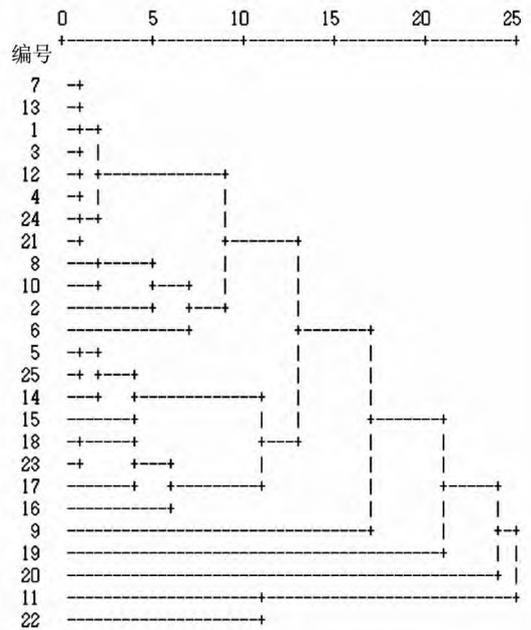


图2 监测站点聚类分析树状图

Fig.2 Monitoring sites groups classified by hierarchical cluster analysis

SPSS 软件聚类分析所得的树状图能够清楚显示出这些监测站点的聚类过程,同时根据聚类分析树状图的结果,能够将监测站点分为六类,具体

各类间的主要区别在于水质污染级别的不同或是主要污染因素的不同,结合各个监测站点以及污染源分布图可以得出:市中区总体水质质量较差,其主要原因是市区内人口以及工业企业密集分布,污染排放量大;任城区一些高污染排放企业周围水质较差外,其他地区水质良好,这与任城区河流分布较多是分不开的,污染物排放后会随河流流入湖泊,从而减轻了对地下水污染的影响。

参考文献:

- [1] 熊德琪. 环境系统模糊集分析理论与应用[M]. 大连:大连海事大学出版社,2001.
- [2] 马凤才. 项目环境影响模糊评价理论与应用研究

[D]. 北京:北京航空航天大学,2001.

- [3] 曾淦宁,吴国权,徐晓群. 多元聚类分析方法在杭州湾水质分析上的应用[J]. 浙江工业大学学报,2009(1):14-19.
- [4] 罗薇,邵秘华,周立新. 聚类分析功能在大连港水域环境质量评价中的应用[J]. 大连海事大学学报,2004(4):51-55.
- [5] 蔡瞳,徐惠,吴群. 土壤质量聚类分析—以封丘县为例[J]. 安徽农业科学,2008,36(25):10998-10999.
- [6] 顾洪博,张继杯. 不确定数据的聚类分析研究及应用[J]. 河北工程大学学报:自然科学版,2012,29(1):109-112.

(责任编辑 王利君)

(上接第58页)

于视景角度,初步构建出节点系统和整体视景框架。滨河景观带实现整体架构是一个多元多层次设计研究的结果,本文对汉中滨河带重要视景空间节点的选择和组织是一个较为前期且类型化的概念,笔者相信在此概念的指引下,个体景观元素会组成一个有活力的整体,以高品质的视景效果传达出汉中城市形象与文化内涵。

参考文献:

- [1] 周晓娟,彭锋. 论城市滨水区景观的塑造[J]. 规划师,2002(3):37-41.

- [2] 凯文·林奇. 城市意象[M]. 北京:华夏出版社,2001.
- [3] 日本土木学会. 滨水景观设计[M]. 孙逸增译. 大连:大连理工大学出版社,2002.
- [4] 梁中效. 汉中历史文化名城的资源与开发[J]. 汉中师范学院学报,2002(2):76-82.
- [5] 程亮. 标志性建筑的视景设计研究[D]. 西安:西安建筑科技大学,2008.
- [6] 王星. 城市地标系统的合理建构[D]. 重庆:重庆大学,2008.
- [7] 西蒙·贝尔. 景观的视觉设计要素[M]. 北京:中国建筑工业出版社,2004.

(责任编辑 王利君)

(上接第66页)率为35.60%,总吸附率为17.78%,总去除率为53.38%。这表明,有接近50%的剩余残留量将随最终出水排出,因而需要增加对低去除率物质的研究。

参考文献:

- [1] 王冰,孙成,胡冠九. 环境中抗生素残留潜在危险及其研究进展[J]. 环境科学与技术,2007,30(3):108-112.
- [2] NIINA LINDQVIST, TUULA TUHKANEN, LEIF KRONBERG. Occurrence of acidic pharmaceuticals in raw and treated sewages and in receiving waters[J]. Water Research, 2005, 39:2219-2228.
- [3] QIAN SUI, JUN HUANG, SHUBO DENG, et al. Occurrence and removal of pharmaceuticals, caffeine and DEET in wastewater treatment plants of Beijing, China[J]. Water Research, 2010, 44:417-426.

- [4] 叶计朋,邹世春,张干. 典型抗生素类药物在珠江三角洲水体中的污染特征[J]. 生态环境,2007,16(2):384-388.
- [5] 徐维海,张干,邹世春,等. 典型抗生素类药物在城市污水处理厂中的含量水平及其行为特征[J]. 环境科学,2007,28(8):1779-1783.
- [6] 刘锋,陶然,应光国,等. 抗生素的环境归宿与生态效应研究进展[J]. 生态学报,2010,30(16):4503-4519.
- [7] 罗晓,王婷. 固定化真菌处理土霉素废水的研究[J]. 河北科技大学学报. 2011,32(1):98-102.
- [8] 叶常明. 水中有机污染物的生物降解性与化学结构的关系[J]. 环境化学丛刊,1986,4(6):6-10.
- [9] 杨晓凡. 水环境中药物污染、检测及去除研究进展[J]. 水资源与保护,2012,28(2):1-7.

(责任编辑 刘存英)