DOI: 10.3969/j.issn.1008-1305.2022.12.056

基于主成分分析法和系统聚类分析法的河流水质评价研究

杜俊鹏[1]，吕　军[1]，吴计生[1]，赵立勇[2]，魏春凤[1]，张　宇[1]

(1. 松辽水资源保护科学研究所，吉林 长春 130021；2. 吉林省润佳水利工程咨询有限公司，吉林 长春 130032)

摘要：为了评价长春市经济技术开发区河流水质情况，文章通过 SPSS 软件，应用主成分分析法和系统聚类分析法进行评价。结果表明：伊通河(卫星大桥断面、自由大桥断面)水质最好，小稗子沟断面、鲶鱼沟(地表水末端)断面水质最差。主成分分析法和系统聚类分析法与单因子评价法相比较能更为详细的评价出，同一水质类别的监测断面更明确的水质优劣情况，并且能够筛选出评价的主要因子和因子之间的聚合程度，为地表水环境质量评价提供一种较优方法。

关键词：主成分分析；系统聚类分析；SPSS 水质评价

中图分类号：X824　　文献标识码：A　　文章编号：1008-1305(2022)12-0216-05

水质评价的方法有很多，从上世纪 70 年代开始，学者们通过各种不同的方法来评价河流水质情况。具体的方法有：单因子评价法、指数评价法、模糊评价法、神经网络评价法、主成分分析法、聚类分析法、灰色评价法、物元分析评价法等。其中主成分分析法是利用降维思想，在损失很少原始信息的前提下把多个指标转化为几个综合指标的多元统计方法。聚类分析是一种建立分类的多元统计分析方法，它能够将一批样本(或变量)数据根据其诸多特征，按照性质上的亲疏程度在没有先验知识的情况下进行自动分类，产生多个分类结果，类内部个体特征具有相似性，不同类间个体特征的差异性较大。[1-7]

1　研究区概况

长春市经济技术开发区简称经开区是 1992 年成立，1993 年经国务院批准的国家级经济技术开发区。地处长春市东部，位于长春向东北拓展的工业主轴线。行政面积 112km^2，分为南北两区，辖四街一镇，常住人口 40 万人，共有各类市场主体 4 万户。经开区内共有河流 16 条、人工湖 1 处。其中：伊通河流域分别为伊通河右岸(卫星路——自由大路段)，全长约 3570m；新开河右岸(东荣大路－金钱小白桥段)及东新开河(东前屯入口－洋浦大街段)，全长约 4470m；小河沿子河右岸(新城大街－伊通河段)，全长约 5120m；鲶鱼沟(绕城高速－伊通河段)，全长约 5770m。饮马河流域分别为雾开河支流干雾海河，全长约 8050m；干雾海河支流中山沟，全长约 6969m；干雾海河支流南阳沟，全长约 470m；兴隆山隆东沟，全长约 1350m。北海公园人工湖水域面积 2887.85m^2。全区共布置地表水监测点 27 个，监测项目为 PH、溶解氧、COD、高锰酸钾指数、氨氮、总磷，监测数据由经开区河长制办公室提供，为 2019 年经开区全年平均地表水监测数据。

2　研究方法

本文采用主成分分析法和系统聚类分析法相结合，具体如下。

2.1　主成分分析法具体步骤[11]

步骤 1：数据标准化。对原始数据进行标准化，以消除数据量纲及数量级的影响。

步骤 2：根据标准化后的数据计算相关系数矩阵。

步骤 3：计算相关系数矩阵的特征值与特征向量。相关系数矩阵的特征值 λ_i 其实就是主成分 F 的方差，一般选取特征根大于 1 的主成分进行分析。

步骤 4：计算方差贡献率并确定主成分。

步骤 5：计算主成分荷载(主成分系数矩阵)。

收稿日期：2022-07-08

作者简介：杜俊鹏(1990 年—)，男，工程师。

E-mail：1107421053@qq.com

主成分荷载值 l_{ij} 与特征向量 u_{ij} 的关系式为：$l_{ij}=u_{ij}/\sqrt{\lambda_i}$。

步骤 6：计算各主成分表达式 F_i 即主成分荷载值 l_{ij} 与对应的标准化后的指标值 x_{ij} 相乘。

步骤 7：计算主成分综合得分值 F。即各主成分得分值 F_i 与相应权重的乘积之和，对应权重为对应特征值在选取总特征值中的占比。

通过主成分分析法，可得各监测断面主成分得分值和总得分值。得分值越高，说明该断面污染越严重。

2.2　系统聚类分析法具体步骤[10]

步骤 1：首先对原始数据进行预处理，即标准化处理。

步骤 2：根据标准化后的数据计算相关系数矩阵。利用标准化后的数据，计算各变量之间相关系数，对相关系数矩阵逐层分析，步骤 1 和步骤 2 与主成分分析法一样。

步骤 3：对不同变量类型下个体距离采用平方欧氏距离计算，个体与小类、小类与小类间距离采用组间平均距离计算，逐步计算至各类对象归为一类，绘制聚类分析谱系图。

通过系统聚类分析谱系图，可以看出哪几类变量或者样本具有较大的关联性，从而对变量进行分类分析，对样本进行分类管理。

3　实例分析

通过主成分分析和系统聚类分析，利用 SPSS 软件，对经开区水质监测断面进行水质评价，具体如下。首先对监测的 27 个水质断面进行标准化处理，见表 1。

表 1　标准化数据表

采样断面	pH	溶解氧	COD	高锰酸盐指数	氨氮	总磷
☆1#伊通河(卫星大桥断面)	-1.000	2.677	-1.024	-1.139	-0.851	-1.313
☆2#伊通河(自由大桥断面)	-0.925	0.797	-1.551	-1.094	-0.906	-1.534
☆3#鲶鱼沟(地表水入境断面)	1.494	-0.370	0.062	0.266	-0.156	-0.425
☆4#鲶鱼沟(地表水末端)	1.675	-0.505	3.439	2.397	0.949	0.762
☆5#东新开河(三道镇长吉南线断面)	-0.703	0.429	-0.385	-0.528	-0.145	0.201
☆6#东新开河(洋浦大街小桥断面)	-0.884	0.508	0.082	0.060	0.785	0.452
☆7#小稗子沟(入境断面)	-0.712	0.277	0.478	0.654	2.653	1.799
☆8#小稗子沟(汇入东新开河前断面)	-0.103	-0.854	-0.375	0.393	1.812	1.561
☆9#窦开河(入境断面)	0.292	-0.730	0.873	1.093	1.657	0.875
☆10#窦开河(汇入小稗子沟前断面)	-0.012	-0.973	0.934	2.098	1.634	0.887
☆11#大稗子沟(汇入东新开河前断面)	-0.522	0.959	-0.506	-0.722	0.106	-0.001
☆12#西稗子沟(汇入东新开河前断面)	0.078	0.360	-0.628	-0.378	-0.205	0.887
☆13#西朝阳沟(入境断面)	-1.370	1.283	-0.070	-0.245	-0.289	-0.389
☆14#西朝阳沟(汇入东新开河前断面)	2.571	0.204	-0.273	-0.350	0.952	-0.085
☆15#东新开河(北远达大桥断面)	-0.736	-0.284	-0.395	-1.361	-0.501	-0.234
☆16#金钱沟(入东新河前断面)	-0.210	-0.661	-0.943	-0.722	-0.748	-1.069
☆17#安龙沟	-0.275	-0.793	0.792	0.371	-0.001	1.011
☆18#分水沟	1.041	-1.573	0.538	0.754	-0.068	1.102
☆19#小河沿子河(经开区入境断面)	1.271	0.554	-0.283	0.010	-0.784	-0.621
☆20#小河沿子河(入伊通河前，经开界内断面)	0.712	0.090	-0.770	-0.978	-0.870	-0.759
☆21#干雾海河(综保区)(中山广场断面)	0.819	-1.243	1.198	1.431	-0.823	-0.079
☆22#干雾海河(成都大路明渠汇入后断面)	0.251	-0.801	1.137	0.277	-0.391	-0.657
☆23#干雾海河(绵阳路东侧断面)	-1.181	0.868	-0.334	-0.289	-0.605	-1.009
☆24#隆东沟(绿楼小桥)	-0.934	-1.245	-1.298	-1.971	-1.028	-1.540
☆25#干雾海河(南洋沟断面)	0.399	-0.106	-0.486	-0.278	-0.954	-1.134
☆26#干雾海河(综保区吐口)	0.152	1.843	0.346	0.393	-0.544	-0.437
☆27#中山沟	-1.189	-0.712	-0.557	-0.145	-0.680	1.746

接着计算相关系数矩阵见表 2。从相关系数矩阵表可以看出，大部分相关系数大于 0.3，说明各部分变量的相关性是比较强的，它们存在信息上的重叠，因此对原始数据进行主成分分析是比较合适的。并且从表中可以看出，COD 与高锰酸盐指数的相关性最强，系数达到 0.878，氨氮与总磷的相关性也较强，系数达到 0.716。

表 2　相关系数矩阵表

项目	pH	溶解氧	COD	高锰酸盐指数	氨氮	总磷
pH	1.000	−0.303	0.438	0.429	0.142	0.068
溶解氧	−0.303	1.000	−0.337	−0.357	−0.190	−0.370
COD	0.438	−0.337	1.000	0.878	0.445	0.454
高锰酸盐指数	0.429	−0.357	0.878	1.000	0.570	0.583
氨氮	0.142	−0.190	0.445	0.570	1.000	0.716
总磷	0.068	−0.370	0.454	0.583	0.716	1.000

接下来通过 KMO 和巴特利特检验进一步说明研究方法的正确性。见表 3，从表 3 可以得出 KMO 值为 0.701，巴特利特球形度检验显著性为 0.000。通常我们认为 KMO 检验结果在 0.5～0.7 之间，同时巴特利特检验结果的显著性小于 0.05，则表示原始数据适宜进行主成分分析。KMO 检验结果大于 0.7 则非常适合主成分分析，低于 0.5 则不适合用主成分分析[8]。因此本项目是非常适合用主成分分成来进行水质评价的。

表 3　KMO 和巴特利特检验表

KMO 取样适切性量数		0.701
巴特利特球形度检验	近似卡方	75.750
	自由度	15
	显著性	0.000

计算特征值与特征向量，见表 4。查阅相关文献知，当特征值小于 1 时，表示该主成分的解释力度还不如直接引入原变量平均值的解释力度大[9-10]，因此考虑将特征值大于 1 作为纳入标准。本例中选用两个特征值，分别为 λ_1 = 3.201，λ_2 = 1.138，此时累积方差贡献率为 72.316%，也就是说通过选取两个主成分，就可以表达原始指标绝大部分的信息。并且可以进一步知道，第一主成分的影响最大，方差百分比为 53.356%。

计算主成分荷载值，即特征向量。见表 5。从主成分荷载矩阵可以看出，锰酸盐指数、COD、总磷、总氮在第一主成分荷载较大，PH 在第二主成分荷载较大。负值代表的是负相关。因此可以得出高锰酸盐指数、COD、总磷和氨氮是主要的污染因子。

表 4　总方差解释表

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积/%	总计	方差百分比	累积/%
1	3.201	53.356	53.356	3.201	53.356	53.356
2	1.138	18.960	72.316	1.138	18.960	72.316
3	0.799	13.318	85.633			
4	0.516	8.605	94.239			
5	0.240	4.008	98.246			
6	0.105	1.754	100.000			

表 5　主成分荷载矩阵表(特征向量)

项目	主成分 1	主成分 2
pH	0.274	−0.678
溶解氧	−0.300	0.251
COD	0.475	−0.191
高锰酸盐指数	0.510	−0.073
氨氮	0.410	0.465
总磷	0.425	0.468

接着从系统聚类分析法，来分析各个变量的相关性。通过 SPSS 软件绘制生成谱系图，如图 1 所示，从谱系图可以看出，COD 与高锰酸钾的关联性较强，氨氮和总磷的关联性较强，溶解氧与任何一个变量的关联性都较差，这也从另一种方法验证了相关系数矩阵表和主成分荷载矩阵表即主成分分析法的正确性，更加直观、形象的展示了各个变量之间的亲疏关系。

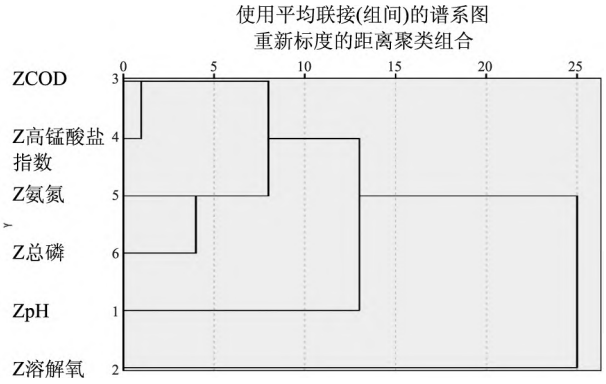

图 1　评价联接(组间)谱系图

表 6　各监测断面主成分综合得分值和单因子水质综合评价对比表

监测断面	*Y*1	*F*1 排名	*Y*2	*F*2 排名	*Y*	综合排名	单因子水质综合评价
☆1#伊通河(卫星大桥断面)	-3.050	27	0.620	8	-2.087	27	Ⅲ
☆2#伊通河(自由大桥断面)	-2.810	26	0.060	15	-2.057	26	Ⅲ
☆3#鲶鱼沟(地表水入境断面)	0.440	11	-1.410	26	-0.045	13	Ⅴ
☆4#鲶鱼沟(地表水末端)	4.180	1	-1.300	24	2.743	1	Ⅴ
☆5#东新开河(三道镇长吉南线断面)	-0.750	17	0.720	7	-0.364	15	Ⅴ
☆6#东新开河(洋浦大街小桥断面)	0.190	12	1.280	3	0.476	9	Ⅴ
☆7#小稗子沟(入境断面)	2.140	4	2.490	1	2.232	3	Ⅴ
☆8#小稗子沟(汇入东新开河前断面)	1.660	6	1.470	2	1.610	5	Ⅴ
☆9#窦开河(入境断面)	2.320	3	0.550	10	1.856	4	Ⅴ
☆10#窦开河(汇入小稗子沟前断面)	2.850	2	0.610	9	2.263	2	Ⅴ
☆11#大稗子沟(汇入东新开河前断面)	-1.000	18	0.790	6	-0.531	17	Ⅴ
☆12#西稗子沟(汇入东新开河前断面)	-0.280	14	0.500	11	-0.075	14	Ⅴ
☆13#西朝阳沟(入境断面)	-1.200	20	0.970	5	-0.631	18	Ⅴ
☆14#西朝阳沟(汇入东新开河前断面)	0.690	9	-1.210	23	0.192	11	Ⅴ
☆15#东新开河(北远达大桥断面)	-1.300	21	0.260	14	-0.891	20	Ⅴ
☆16#金钱沟(入东新河前断面)	-1.440	23	-0.640	18	-1.230	23	Ⅳ
☆17#安龙沟	1.160	8	0.280	13	0.929	7	Ⅴ
☆18#分水沟	1.840	5	-0.770	19	1.155	6	Ⅴ
☆19#小河沿子河(经开区入境断面)	-0.530	15	-1.320	25	-0.737	19	Ⅴ
☆20#小河沿子河(入伊通河前，经开界内断面)	-1.380	22	-1.000	20	-1.280	24	Ⅳ
☆21#干雾海河(综保区)(中山广场断面)	1.530	7	-1.620	27	0.704	8	Ⅴ
☆22#干雾海河(成都大路明渠汇入后断面)	0.550	10	-1.100	21	0.117	12	Ⅴ
☆23#干雾海河(绵阳路东侧断面)	-1.570	24	0.350	12	-1.066	21	Ⅳ
☆24#隆东沟(绿楼小桥)	-2.580	25	-0.490	17	-2.032	25	Ⅳ
☆25#干雾海河(南洋沟断面)	-1.100	19	-1.160	22	-1.116	22	Ⅳ
☆26#干雾海河(综保区吐口)	-0.550	16	-0.190	16	-0.456	16	Ⅴ
☆27#中山沟	0.010	13	1.250	4	0.335	10	Ⅴ

4　结语

通过相关系数矩阵表、主成分荷载矩阵表和谱系图可以得到，高锰酸盐指数、COD、总磷和氨氮为主要的污染因子，并且高锰酸钾指数和 COD 的关联度较大，总磷和氨氮的关联度较大。通过计算各个监测断面的主成分综合得分值，在与单因子水质评价进行比较，见表 6，可以看出经开区河流水质总体较差，大多数为Ⅴ类水体。水质最好的监测断面为伊通河(卫星大桥断面)、伊通河(自由大桥断面)水质为Ⅲ类。鲶鱼沟(地表水末端)断面、小稗子沟(入境断面)、小稗子沟(汇入东新开河前断面，窦开河为小稗子沟支流，汇入小稗子沟断面)水质最差。本文的研究思路，可以为其他河流的水质评价提供参考依据，为水资源管理、水污染防治提供科学方法。接下来作者将用神经网络法、灰色评价法等多种方法对水质进一步评价，从更多方面完善其工作。

参考文献

[1] 赵朋飞，刘俊，胡少敏. 基于 SPSS 软件的主成分分析法在水质评价中的应用[J]. 应用技术，2016(10)：119-121.

[2] 邱顺凡. 村镇地表水体水质监测点优化布置与水质评价方法研究[D]. 湖南大学，2014.

[3] 周志军，潘三军，杨培慧. SPSS 模糊聚类分析法在水质监测断面聚类分析中的应用[J]. 仪器仪表与分析监测，2007(4)：32-34.

[4] 张亚娟，牛珊珊，孙亚乔，等. SPSS 软件在渭河流域(陕西段)水质主成分分析评价中的运用[J]. 安徽农业科学，2012，40(29)：14414-14416.
[5] 刘清园，李永，蒲迅赤，等. 改进的主成分分析法在水库水质评价中的应用研究[J]. 四川环境，2017，36(6)：117-122.
[6] 张莹，刘硕，王宏. 基于 SPSS 的主成分分析法在松花江哈尔滨段的水质评价[J]. 哈尔滨师范大学自然科学学报，2015，31(3)：132-135.
[7] 钱程，穆文平，王康，等. 基于主成分分析的地下水水质模糊综合评价[J] 水电能源科学，2016，34(11)：31-35.
[8] 薛东玮. 吕梁市三川河水质及预测分析[D]. 山西财经大学，2020.
[9] 姜厚竹. 松花江流域省界缓冲区水质监测指标与断面优化[D]. 东北林业大学，2017.
[10] 郑泽豪. 基于聚类分析水质指标相关性研究[J]. 广东水利水电，2020(5)：59-62.
[11] 杨芳，杨盼，卢路，等. 基于主成分分析法的洞庭湖水质评价[J]. 人民长江，2019，50(增刊 2)：42-46.
[12] 张文睿，孙栋元，武兰珍，等. 基于主成分分析的甘肃省流域分区水质评价[J]. 水利规划与设计，2021(8)：72-78.

(上接第 102 页) 驳岸点，临时码头上面结构采用贝雷梁钢结构搭设面板，下部采用钢管桩作为支撑结构，每跨渡槽槽身采用起吊质量 100t 起重船进行水上吊运至临时码头，30t 装自卸汽车拉至岸边，再采用液压破碎机现场破碎，将钢筋、砼分离，最后用挖机或装载机配合装 8t 自卸汽车将其拉运至指定弃渣场进行堆放处理。

(2) 槽身下部结构

槽身下部结构采用自上而下的拆除顺序，即首先拆除排架横梁，继而拆除排架墩柱，最后是排架基础。采用船载挖掘破碎机对槽身下部结构各个部位进行破碎拆除，再用船载挖机或装载机挖运废料到装载船送到岸边临时堆放点进行堆放，最后用挖机或装载机配合装 8t 自卸汽车将其拉运至指定弃渣场进行堆放处理[6]。

3.3 方案比选

通过上述比较，3 种拆除方案各有其优缺点，综合考虑设计采用方案二，用舟梁合一的形式搭设临时浮桥进行拆除。

渡槽(长槽)拆除总体施工流程为：管线及其他附属设备迁改保护→渡槽桥面系→搭设临时浮桥→拆除渡槽槽身→拆除排架横梁、排架墩柱及排架基础→装运→拆除临时浮桥→场地清理恢复。

表 2　拆除工程数量表

	施工要点	工程造价	优点	缺点
方案一	采用贝雷梁搭设临时钢便桥	造价约 460.15 万元	具有结构简单、运输方便、架设快捷、载重量大、互换性好、适应性强的特点	投资较高，工期较长
方案二	采用舟梁合一的形式搭设临时浮桥	造价约 332.95 万元	设备简单，吃水浅，载货量大，成桥迅速，架拆方便，抗冲性能好，稳定性较强，相对于其他方案投资较低	本身无自航能力，需拖船或顶推船拖带的货船
方案三	采用 100t 起重船吊运至临时码头	造价约 371.85 万元	机动性较强	效率较低，工期较长，受水位影响较大

4　结语

连环水库南干渠槽身和立架混均已老化，有崩塌的危险，严重危及过往船只的安全，不利于社会稳定，同时存在一定的安全隐患。该渡槽拆除难度高，风险大，渡槽位于漠阳江东支流河道上，河道交通较繁忙，为确保既能安全拆除，又尽可能减少对河道交通的影响，同时兼顾起吊、运输设备对于现场环境的要求和工程造价等因素，设计出本工程可针对短、长槽的不同周边环境，选用最佳方案进行拆除方法。

本文较全面地分析了该渡槽拆除工程设计方案并对拆除方式选用有了更多认识，达到了节约工期的同时也保证了施工质量，对同类工程有一定的借鉴意义。

参考文献

[1] 孙丽玲. 海城市下坎灌区渡槽改造工程设计及计算[J]. 水利科技与经济，2022(4)：31-35.
[2] 郭妍. 韶山灌区现代化改造中楠竹长虹渡槽拆除重建方案分析[J]. 陕西水利，2022(4)：133-135.
[3] 严冬青. 南干渡槽除险加固工程设计[J]. 中国科技信息，2021(22)：43-44.
[4] 张帅. 灯塔灌区东沙浒渡槽拆除重建工程设计方案分析[J]. 黑龙江水利科技，2020(12)：87-89.
[5] 刘波，郑斌. 南桥渡槽拆除重建施工技术[J]. 水科学与工程技术，2017(2)：88-90.
[6] 曹武，罗平，陈崇德. 永圣渡槽拆除重建工程施工组织评价研究[J]. 小水电，2019(2)：64-68.