

Principal Component Analysis of Water Quality in Lihe River

Gang Liu¹, Hui Wu¹, Zaiping Liu², Yanxi Li¹, Jingsha Song¹

¹Jingmen Academy of Environmental Sciences, Jingmen Hubei

²Hubei Jinghuan Environmental Protection Engineering Technology Co., Ltd., Jingmen Hubei

Email: 61456650@qq.com

Received: Jan. 30th, 2019; accepted: Feb. 18th, 2019; published: Feb. 25th, 2019

Abstract

Principal components analysis (PCA) was applied to assess the water quality at the end of the Lihe River in 2017, based on the monitoring data of 13 indicators of BOD₅, COD_{Mn}, NH₃-N, TP, TN, DO, pH, etc. The result showed that the water pollution score was the highest in July and lowest in January; Domestic pollution, river sediment and industrial waste water are the main causes of water pollution in the river.

Keywords

Lihe River, Community Structure, Cluster Analysis

基于主成分分析法的淠河水质评价

刘刚¹, 吴辉¹, 刘在平², 李妍羲¹, 宋荆沙¹

¹荆门市环境科学研究院, 湖北 荆门

²湖北荆环环保工程技术有限公司, 湖北 荆门

Email: 61456650@qq.com

收稿日期: 2019年1月30日; 录用日期: 2019年2月18日; 发布日期: 2019年2月25日

摘要

依据淠河入汉江口2017年常规监测数据, 选取BOD₅、COD_{Mn}、NH₃-N、TP、TN、DO等13种水质指标, 基于主成分分析法对淠河的水质情况进行综合评价。结果表明, 1月份水质较好, 7月份水质较差, 生活源、河道底泥、工业废水是造成淠河水质污染的主要原因。

关键词

浏河, 主成分分析, 水质评价

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

水环境是人类赖以生存的重要环境要素之一。随着我国经济社会的高速发展, 水环境污染已成为严重制约我国经济社会可持续发展和危害人们身体健康的重要因素。因此, 如何合理、客观地描述水质现状, 对水质污染程度进行准确评价已成为环境研究领域的一个热点和难点问题。目前, 国内外主要的水质评价方法有污染指数法[1][2]、模糊评价法[3]、神经网络评价法[4][5]、灰色评价法[6]、物元评价法[7][8]、地理信息系统评价法[9]及主成分分析法[10][11][12]等。

由于地区差异而形成的水污染的特点不同, 导致污染因子权重的确定各异, 分析方法权重值的选取主观性较强, 评价的结果不便于与其它地区水质状况进行比较, 因而也不能准确反应本地的相对污染状态。主成分分析方法是一种成功的因子降维和数据压缩的统计学方法[13][14], 它将收集到的河流水质监测数据, 在确保不损失原有主要信息的前提下, 用较少的具有代表性的主因子来概括多种水质要素所反映的水质信息, 达到找出影响其河流水质的主要污染因子, 并识别污染物来源及其成因, 为进行环境综合整治提供技术支持。本文使用 IBM SPSS Advanced Statistics 20.0 软件对浏河入汉江口 2017 年监测数据进行主成分分析, 并对水质进行评价。

2. 数据来源

浏河是汉江的主要支流之一, 流经荆门市东宝区、钟祥市, 长约 184 公里, 控制流域面积 1140 平方公里, 主要界于北纬 30°09'07"~30°27'36"、东经 112°00'17"~112°24'45"。浏河入汉江口监测断面执行《地表水环境质量标准》(GB3838-2002) III 类标准, 本文选取该监测点 2017 年常规监测数据中的 pH、电导率、水温、氨氮、溶解氧、高锰酸钾指数、生化需氧量、石油类、化学需氧量、总磷、总氮、氟化物、六价铬等 13 项指标以及浏河水文站(陈安站)的流量数据作为数据源, 见下表 1。

Table 1. Routine monitoring data of water quality from Lihe river into Hanjiang river in 2017

表 1. 浏河入汉江口监测断面 2017 年常规监测数据

月份	流量 (m ³ /s)	pH	水温 (°C)	电导率 (ms/m)	溶解氧 (mg/L)	COD _{Mn} (mg/L)	BOD (mg/L)	氨氮 (mg/L)	石油类 (mg/L)	六价铬 (mg/L)	COD (mg/L)	总氮 (mg/L)	总磷 (mg/L)	氟化物 (mg/L)
1	19	7.5	11	106	9.84	2.53	1.18	0.839	0.025	0.015	9.94	3.254	0.784	0.63
3	20.1	7.65	17	65.8	9.15	2.75	2.08	1.038	0.025	0.016	10.1	6.07	2.41	0.67
5	28.1	7.8	22	56.7	7.87	4.5	2.02	0.3	0.028	0.014	20	6.66	2.59	1.54
7	37.1	7.9	26	60	6.1	4.5	1.77	0.27	0.019	0.012	26	2.46	2.03	0.82
9	80	7.7	23	54.3	7	4	1.1	0.36	0.02	0.009	17	2.81	0.29	0.51
11	26	7.2	18	47	8.7	2	1.9	0.54	0.03	0.021	14	2.96	0.48	0.4

3. 研究方法

主成分分析法是利用降维思想, 把多指标变量转化为少数几个综合指标的多元统计分析方法, 且这

少数几个综合指标可以反映原来多个变量的大部分信息。本研究使用 SPSS 软件由湖北荆环环保工程技术有限公司提供,对 13 项浏河水质监测因子和流量进行主成分分析,得到综合水质评价函数;依据此评价函数计算结果,比较不同月份的水质情况,准确地掌握不同月份水体受污染的程度[15]。其分析步骤如下:

- 1) 对原始指标进行标准化处理,以消除数据之间量纲的影响;
- 2) 求标准化数据的相关关系矩阵;
- 3) 求相关系数的特征值、特征向量和方差贡献率;
- 4) 计算主成分贡献率及累计贡献率,确定主成分个数,一般要求前 n 个主成分的累计解释方差占总方差的 85%以上即可;
- 5) 选择 n 个主成分,计算相应的特征向量,求主成分分析的得分;
- 6) 计算主成分载荷得到综合函数并用其评判水质。

4. 结果与分析

4.1. 综合水质评价函数

通过主成分分析得到综合水质评价函数。运用 SPSS 软件,将各项水质指标进行标准化处理,提取相关系数矩阵(表 2),并求得各个指标解释的总方差、累计贡献率(表 3)、碎石图(图 1)及初始因子荷载(成分)矩阵(表 4)。主成分荷载矩阵需用初始因子荷载矩阵中相关数据除以主成分相对应特征值的平方根,求出 3 个主成分中每个指标对应的系数,得到相应的主成分荷载矩阵(表 5)。

Table 2. Correlation coefficient matrix

表 2. 相关系数矩阵

	水温	流量	pH	电导率	溶解氧	COD _{Mn}	BOD ₅	氨氮	石油类	COD	总氮	总磷	氟化物	六价铬
流量	0.558	1.000	0.287	-0.389	-0.630	0.464	-0.573	-0.530	-0.611	0.339	-0.403	-0.455	-0.211	-0.730
pH	0.624	0.287	1.000	-0.041	-0.645	0.908	0.037	-0.394	-0.675	0.647	0.237	0.620	0.595	-0.792
电导率	-0.740	-0.389	-0.041	1.000	0.579	-0.280	-0.470	0.540	-0.069	-0.477	-0.061	-0.083	-0.071	-0.047
水温	1.000	0.558	0.624	-0.740	-0.962	0.799	0.204	-0.806	-0.491	0.896	-0.098	0.266	0.308	-0.508
溶解氧	-0.962	-0.630	-0.645	0.579	1.000	-0.809	0.021	0.832	0.657	-0.923	0.315	-0.122	-0.197	0.621
COD _{Mn}	0.799	0.464	0.908	-0.280	-0.809	1.000	-0.004	-0.728	-0.568	0.828	0.129	0.448	0.655	-0.779
BOD ₅	0.204	-0.573	0.037	-0.470	0.021	-0.004	1.000	0.072	0.482	0.144	0.619	0.733	0.419	0.529
氨氮	-0.806	-0.530	-0.394	0.540	0.832	-0.728	0.072	1.000	0.272	-0.879	0.275	0.037	-0.377	0.421
石油类	-0.491	-0.611	-0.675	-0.069	0.657	-0.568	0.482	0.272	1.000	-0.466	0.456	0.037	0.163	0.842
COD	0.896	0.339	0.647	-0.477	-0.923	0.828	0.144	-0.879	-0.466	1.000	-0.204	0.299	0.436	-0.458
总氮	-0.098	-0.403	0.237	-0.061	0.315	0.129	0.619	0.275	0.456	-0.204	1.000	0.745	0.688	0.136
总磷	0.266	-0.455	0.620	-0.083	-0.122	0.448	0.733	0.037	0.037	0.299	0.745	1.000	0.747	-0.040
氟化物	0.308	-0.211	0.595	-0.071	-0.197	0.655	0.419	-0.377	0.163	0.436	0.688	0.747	1.000	-0.207
六价铬	-0.508	-0.730	-0.792	-0.047	0.621	-0.779	0.529	0.421	0.842	-0.458	0.136	-0.040	-0.207	1.000

Table 3. Eigenvalue, contribution rate and cumulative contribution rate

表 3. 特征值、贡献率及累积贡献率

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的%	累积%	合计	方差的%	累积%	合计	方差的%	累积%
1	6.663	47.596	47.596	6.663	47.596	47.596	4.708	33.631	33.631
2	3.750	26.788	74.383	3.750	26.788	74.383	4.068	29.056	62.687

Continued

3	2.049	14.634	89.017	2.049	14.634	89.017	3.686	26.330	89.017
4	0.778	5.560	94.577						
5	0.759	5.423	100.000						
6	1.201E-015	8.579E-015	100.000						
7	7.145E-016	5.103E-015	100.000						
8	1.887E-016	1.348E-015	100.000						
9	9.003E-017	6.431E-016	100.000						
10	1.702E-017	1.216E-016	100.000						
11	-2.388E-016	-1.706E-015	100.000						
12	-3.128E-016	-2.234E-015	100.000						
13	-4.262E-016		100.000						
14	-7.325E-016	-5.232E-015	100.000						

Table 4. Initial factor composition matrix
表 4. 初始因子荷载(成分)矩阵

	成份 1	成份 2	成份 3
水温	0.927		
电导率	-0.867		
溶解氧	-0.865	-0.477	
氨氮	-0.854		
化学需氧量	0.835		
六价铬		-0.951	
石油类		-0.855	
pH		0.788	0.514
高锰酸盐指数	0.588	0.693	
流量	0.481	0.529	-0.511
总磷			0.959
总氮			0.859
氟化物			0.854
生化需氧量		-0.576	0.709

Table 5. Factor load matrix
表 5. 主成分荷载矩阵

成份	水温	流量	pH	电导率	溶解氧	高锰酸盐指数	生化需氧量	氨氮	石油类	化学需氧量	总氮	总磷	氟化物	六价铬
F_1	0.214	0.084	-0.062	-0.304	-0.179	0.042	0.134	-0.210	0.060	0.175	-0.072	-0.038	-0.015	0.084
F_2	-0.037	0.083	0.229	0.243	-0.020	0.149	-0.211	0.048	-0.242	-0.002	0.001	0.044	0.060	-0.279
F_3	0.000	-0.151	0.154	0.048	0.038	0.102	0.166	0.048	0.051	0.022	0.245	0.267	0.235	-0.009

依据特征根大于 1 的原则, 由表 3 及碎石图可以看出, 有 3 个主成分的特征根大于 1, 旋转后第一主成分对总方差的贡献率为 33.631%, 3 个主成分累计解释方差为 89.017%, 符合大于 85% 的信息抽取率。

由表 4 可知, 第一主成分与水温相关性最高, 与化学需氧量、高锰酸钾指数正相关, 与电导率、溶解氧、氨氮负相关, 溶解氧含量表明水体自净能力的强弱, 一定程度反应水体的受污染程度, 可代表季节性波动明显的水体因子, 表明污染来自生活源和河流底泥; 第二主成分与六价铬高度负相关, 与石油

类、生化需氧量负相关，与流量、高锰酸盐指数正相关，可表明污染来自工业源；第三主成分与总磷、总氮高度相关，与氟化物、生化需氧量正相关，与流量负相关，可表明浏河水质有富营养化加剧的风险。分析方差贡献率可知 3 个主成分方差贡献率接近，表明浏河水体受污染不是某方面原因，是多因素作用的结果。

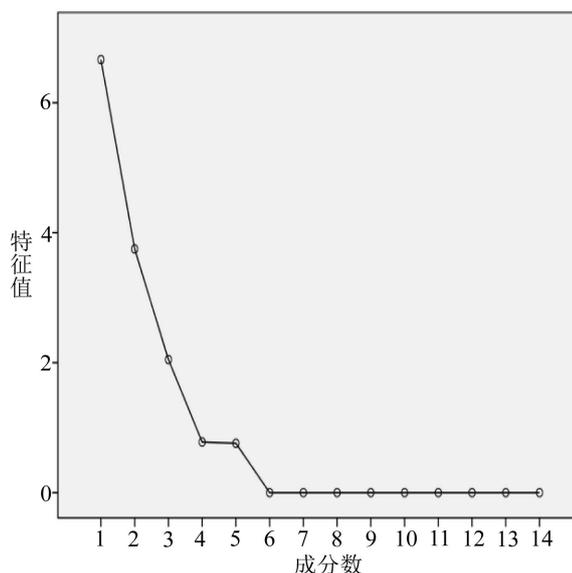


Figure 1. Scree plot

图 1. 碎石图

由上表 5 可知, F_x 是特征向量与相对应系数的累和; 由表 3 可知, 三个主成分特征根分别为: $\lambda_1 = 6.663$, $\lambda_2 = 3.75$, $\lambda_3 = 2.049$ 。

综上, 可以得到浏河综合水质评价函数:

$$F = \left[\frac{\lambda_1}{(\lambda_1 + \lambda_2 + \lambda_3)} \right] F_1 + \left[\frac{\lambda_2}{(\lambda_1 + \lambda_2 + \lambda_3)} \right] F_2 + \left[\frac{\lambda_3}{(\lambda_1 + \lambda_2 + \lambda_3)} \right] F_3$$

$$F = 0.534665F_1 + 0.300915F_2 + 0.16442F_3 \quad (1)$$

4.2. 水质等级评价

将所有月份采样点的因子值分别代入主成分分析得到的综合水质评价函数(1), 统计得到不同时期每个采样点的综合水质评价函数值 F , 其中污染严重程度随着得分的增大而增大。

根据图 2 对 2017 年浏河水质污染严重程度进行排序(7 月 > 9 月 > 5 月 > 11 月 > 3 月 > 1 月), 其中 7 月份污染最严重, 1 月份污染最轻。呈现出明显的季节性变化, 与地表水环境质量标准评价结果一致。

5. 结论

本文从浏河入汉江口 2017 年的采样数据分析中选取 13 个主要水质因子和流量进行主成分分析, 得到综合水质评价函数; 依据综合水质评价函数, 得到水质污染程度趋势图。

1) 浏河水质污染呈现明显季节性特征, 随着气温升高而污染程度加深, 春冬季的 12 月至 2 月污染较轻, 夏季的 6 月至 8 月污染最重。

2) 第一主成分解释总方差值较大, 第二主成分及第三主成分解释总方差值较接近, 说明浏河水质污染来源复杂多样, 主要包括生活源、河道底泥和工业源。

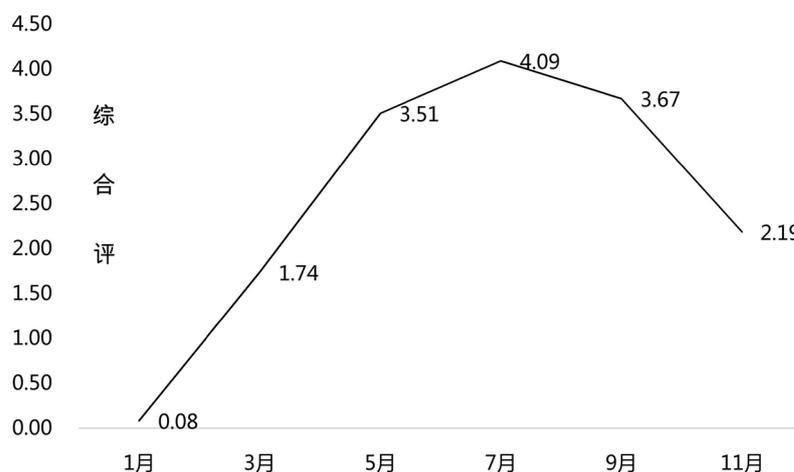


Figure 2. Principal component analysis comprehensive score
图 2. 主成分分析综合得分

3) 第一主成分中水温、电导率、溶解氧、流量代表水体环境的物理状况; 氨氮与水温负相关($R^2 = 0.65$), 可能因为冬春季节流域内人口密度增加, 生产活动加大引起的; COD 与水温正相关, 考虑到温度同样也能影响有机质的释放, 随着温度升高而增大, 当超过 20°C 时 COD 释放量最大[16], 因此, 浏河水质中的有机物污染主要为河道底泥引起的。

4) 第二主成分、第三主成分中六价铬、石油类、总磷、总氮、氟化物与流量负相关, 表明其属于点源污染。浏河流域主要分布磷化工生产企业, 六价铬、总磷、氟化物是其生产废水的特征因子, 表明点源污染主要源于工业废水。

参考文献

- [1] 梁德华, 蒋火华. 河流水质综合评价方法的统一和改进[J]. 中国环境监测, 2002, 18(2): 63-66.
- [2] 薛巧英. 水环境质量评价方法的比较分析[J]. 环境保护科学, 2004, 30(4): 64-67.
- [3] 陈广桐. 模糊聚类方法在水体环境单元分类中的应用[J]. 环境工程, 1994, 12(2): 61-63.
- [4] 王李管, 贾明涛. 水质评价及预测的神经网络方法[J]. 环境工程, 1998, 16(2): 62-65.
- [5] 刘国东, 黄川友, 丁晶. 水质综合评价的神经网络模型[J]. 中国环境科学, 1998, 18(6): 514-517.
- [6] 陆洲, 夏秋颖, 周琳, 等. 等斜率灰色聚类法在地面水环境质量评价中的应用[J]. 环境保护科学, 2000(5): 43-46.
- [7] 王国平, 王洪光. 物元分析法用于水环境质量的评价比较[J]. 干旱环境监测, 1997(2): 65-67.
- [8] 冯玉国. 用物元分析法综合评价环境质量[J]. 环境保护, 1995(9): 30-32.
- [9] 方丽玲, 吴立新, 周广明, 等. GIS 支持下水质模型的应用[J]. 新疆环境保护, 1999, 21(3): 1-8.
- [10] 刘德林, 刘贤赵. 主成分分析在河流水质综合评价中的应用[J]. 水土保持研究, 2006, 13(3): 124-125.
- [11] 吴浩东, 胡衡生. 基于主成分分析法的明江河水质评价[J]. 湖北农业科学, 2010, 49(10): 247-249.
- [12] 王红, 姚君兰, 李艳蕾, 等. 基于主成分分析法的梁子湖水水质评价分析[J]. 湖北大学学报, 2017, 39(6): 602-608.
- [13] 吉祝美, 方里, 张俊, 等. 主成分分析法在 SPSS 软件中的操作及在河流水质评价中的应用[J]. 环保科技, 2012, 18(4): 38-43.
- [14] 朱琳, 王雅南, 韩美, 等. 武水河水水质时空分布特征及污染成因的解析[J]. 环境科学, 2018, 38(6): 2151-2155.
- [15] 李晓丰, 李兴华, 黄涛, 等. PCA 水质分类法在通济堰水质评价中的应用[J]. 四川环境, 2016, 35(3): 41-45.
- [16] 黄飞. 三岔湖底泥污染物释放规律研究[D]: [硕士学位论文]. 四川: 西南交通大学, 2011: 6-7.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2164-5485，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：aep@hanspub.org