

改进的 PCA 法在疏勒河流域 水库水质评价中的应用

田晋华^{1,2}, 高雅玉²

(1. 甘肃农业大学 工学院, 甘肃 兰州 730070; 2. 甘肃省水土保持科学研究所, 甘肃 兰州 730020)

摘 要: 疏勒河流域属极端干旱荒漠气候, 水资源的形成和赋存较为困难、水生态环境脆弱。为了加强对流域水资源的定量分析, 提高水资源利用效率、更合理地开发利用有限的水资源, 研究利用结合地表水环境质量标准的主成分分析法、选取流域内上、中、下游 3 个水库的典型断面从时间和空间角度进行水质评价研究, 在去除监测项目中的重复信息得到的水质评价结果有: 1997 年之前双塔水库属 I 类水质, 1997 年后由于水库受水量减少和进入水体的污染物载荷量增大致使水质下降到 II、III 类, 2000 年以后水质基本维持在 II 类左右, 水质符合生活饮用、工业应用、农业灌溉、渔业养殖、景观娱乐等用水标准。2000 年之后党河水库水质在 I、II 类之间, 水库上游人口稀少、敦煌市排污口在水库下游致使水体受到的点源污染较少, 基本为非点源污染。2000 年以来, 赤金峡水库水质为 V 类, 因为玉门市排污口在赤金峡水库上游严重影响了水库水质, 因此, 加强石油河的污染治理及控制该河流上的点源污染刻不容缓。

关键词: 水库水质; 水质评价; 主成分分析; 疏勒河流域

中图分类号: X824

文献标识码: A

文章编号: 1672-643X(2014)01-0156-05

Application of improved PCA method in evaluation of reservoir water quality in Shulehe river basin

TIAN Jinhua^{1,2}, GAO Yayu²

(1. College of Technology, Gansu Agricultural University, Lanzhou 730070, China;

2. Institute of Soil & Water Conservation Sciences of Gansu Province, Lanzhou 730020, China)

Abstract: Shulehe river basin belongs to extremely arid desert climate, and the formation and storage of its water resources is more difficult, the environment of water ecology is fragile. In order to strengthen analysis of water resources, improve water use efficiency, develop and utilize water resources more rationally, the paper used PCA method, and selected the typical section of three reservoirs from upper, middle and lower reaches of Shulehe river basin to analyze and evaluate water quality. The results are that the water quality in Shuangta Reservoir belonged to grade I before 1997. The water quality of Shuangta reservoir dropped to grade II and III due to the amount of water flowing into reservoir decreased and pollutant load increased after 1997. Since 2000, the water quality has maintained at grade II, and the quality is good enough to be satisfied with the standards of domestic use, industrial application, agricultural irrigation, aquaculture and landscape water standards. The water quality in Danghe reservoir belonged to grade I and II, because people sparsely populate at the reservoir upstream and the sewage outfall of Dunhuang City is in the downstream of the reservoir. So the major type of pollution is non-point source pollution. The water quality in Chijinxia Reservoir is classified to trade V because the sewage outfall of Yumen City is in upstream of the reservoir. So it is urgent to strengthen pollution abatement and control the pollution of point source in Shiyou river.

Key words: water quality of reservoir; valuation of water quality; principal component analysis; Shulehe river basin

1 概 述

水环境质量评价的开展基于水质的采样调查与分析,它是水资源合理开发利用及水环境污染预防的基础和依据。虽然水环境质量由水质、水体底质的质量以及生物学质量构成^[1],但水质在很大程度上决定了水环境的好坏,因此,衡量水环境的优劣程度在很大程度上取决于水质的评价。水质的评价方法包括 3 大类:单因子分析法、水质评价指数法、利用模型分析评价法^[2-5]。随着多元统计学的发展和应用,主成分分析方法以其充分考虑各指标间的信息重叠、最大限度地保留原有信息、对高维变量进行最佳综合降维以及客观确定各个指标的权重、避免了主观随意性的优点被广泛应用于水质评价中^[6]。但传统的主成分分析方法很难直接与水质指标相联系,往往需要和聚类分析模型、灰色模糊分析模型或人工神经网络等模型耦合,共同判断水质类别,致使在增加判断复杂程度的同时也降低了模型判断的准确度。因而,改进的主成分分析法被引入到研究中来。

疏勒河流域(图 1)位于河西走廊的最西端,属极端干旱荒漠气候,水资源的形成和赋存较为困难,水生态环境脆弱,水质的优劣程度对当地水资源状况的影响至关重要。选取流域内有代表性和典型性的上、中、下游的 3 个水库(党河水库、双塔水库、赤金峡水库)断面的水质利用改进的主成分分析方法从时间和空间角度进行评价研究,分析水质在时间和空间序列的变化趋势和原因,为水环境规划与水质向着良好的方向发展及预测防治提供较为科学的参考依据,并有针对性地提出改善水质的调控措施与对策建议。



图 1 典型水库断面示意图

2 PCA 法及其改进研究

2.1 简介(原理及算法)

主成分分析是利用线性映射的方法构造新的特征,而新的特征都是原有数据的特征函数。通过主成分分析可以在对数据信息损失最小的情况下简化原有数据,将数据中原有的重复的信息去除,取而代之的是新的一组相互间没有重复的特征信息,从而简化了复杂的数据信息。主成分分析可以使原来的多指标问题转化为几个较少的综合指标,而在这个综合指标是原来的多个指标的线性组合,并且这些综合指标之间是不相关的,在极少信息损失的情况下反映出几乎所有原来指标的信息。

2.1.1 主成分分析算法 原特征向量为 z 的数据集有 M 个样本点,新特征向量是 w (新特征向量都是原特征向量的线性组合):

$$w_j = v_j^T(z - \bar{z})$$
$$= \sum_{k=1}^m v_{jk}(z_k - \bar{z}_k), j = 1, 2, \dots, m \quad (1)$$

式中: $v_j = (v_{j1}, v_{j2}, \dots, v_{jm})^T$, v_{jk} 为常数; $\bar{z} = (\bar{z}_1, \bar{z}_2, \bar{z}_3, \dots, \bar{z}_n)^T$, $\bar{z}_j, j = 1, 2, \dots, m$ 是 M 个样本点第 j 个原特征变量 Z_j 的均值,即 $w = V(z - \bar{z})$ 。

假定训练样本集 M 个样本协方差矩阵 $U(z)$ 已经算出。 $U(z)$ 是一个 $m \times m$ 实数对称方阵,其各特征值 $\tau_1, \tau_2, \dots, \tau_n$ 及对应的特征向量 $v_1, v_2, \dots, v_n$ 由求解齐次方程组 $[U(x) - \tau_j I]v_j = 0, j = 1, 2, \dots, n$ 得到,且特征值 τ_j 与所对应的特征向量 v_j 是相互正交的,则有:

$$\begin{cases} v_j^T v_k = 0, j, k = 1, 2, \dots, m, j \neq k \\ v_j^T v_k = 1, j = k = 1, 2, \dots, m \end{cases} \quad (2)$$

令: $\tau_1 \geq \tau_2 \geq \dots \geq \tau_n$, 那么对于协方差矩阵 $U(z)$,其特征值 τ_j 和对应的特征向量 v_j 都是确定且唯一的。可得:

$$F(w_j) = F\left[\sum_{k=1}^n v_{jk}(z_k - \bar{z}_k)\right] \quad (3)$$

式中: v_{jk} 是常数,则有:

$$F(w_j) = 0 \quad (4)$$

即新特征向量 w 中各变量的均值为 0。即 $U(w) = VU(z)V^T$, 即 $V_{ln}(w) = \sum_{j=1}^m \sum_{k=1}^m v_{lj}v_{kn}U_{jk}(z), l, n = 1, 2, 3, \dots, m$ 。而 V^T 的每一列恰好是 $U(z)$ 的特征向

量,则 $U(z)V^T = \begin{pmatrix} \tau_1 & & & \\ & \tau_2 & & \\ & & \ddots & \\ & & & \tau_n \end{pmatrix}$, 由 $v_1, v_2, \dots, v_n$

相互之间的正交性得到:

$$\begin{cases} U_{kk}(w) = \tau_k & k = 1, 2, \cdots, m \\ U_{kl}(w) = 0 & k \neq l, k, l = 1, 2, \cdots, m \end{cases} \quad (5)$$

由矩阵代数可知 m 阶方阵 $U(z)$ 的 m 个特征值 $\tau_1, \tau_2, \cdots, \tau_n$ 之和等于 $U(z)$ 的迹, 即:

$$\sum_{j=1}^m U_{jj}(z) = \sum_{j=1}^m U_{jj}(W) = \sum_i^m \tau \quad (6)$$

说明协方差矩阵对角元素之和经过主成分分析并不会被改变。

2.1.2 主成分分析的降维处理及其信息损失

如果新特征向量 w 的最后几个分量 $w_n, w_{n-1}, \cdots, w_{p+1}$ 的方差 $\tau_n, \tau_{n-1}, \cdots, \tau_{p+1}$ 的数值很小, 那么其对于样本点的正确分类的影响就很小, 所以 p 维向量 $\bar{w}(p \leq n)$ 就可以被用来设计分类器来完成样本的分类, 这就是主成分分析的降维方法。定义 w 的第 j 个主成分 w_j 的方差贡献率为 $\mu_j = \tau_j / \sum_{k=1}^m \tau_k$, 则 w 的 m 个成分的累计方差率就是 1。当前 p 个主成分的方差贡献率足够接近 1 时, 则前 p 个主成分就可以作为新的特征。则降维向量 $\bar{w}$ 的累计方差率为:

$$\mu_{\bar{w}} = \sum_{j=1}^k \tau_j / \sum_{j=1}^m \tau_j \quad (7)$$

方差越大, 数据包含的信息就越多, 那么 $1 - \mu_{\bar{w}}$ 的值就可作为衡量降维处理信息的损失度^[7-10]。

2.2 主成分分析的改进

在整组数据进入主成分分析之前先进行标准化。在标准化的过程中, 将标准化的过程与地表水环境质量标准(GB3838-2002)相结合, 将水质数据与地表水环境质量标准进行比较, 如果该单项指标高等于地表水环境质量 I 类标, 则标准化取值为 1; 如果该指标等于地表水环境质量 V 类标准, 则取值为 6。在这之间的水质指标就采用线性内插的方法进行取值来完成标准化过程。在计算各个主成分的综合得分时, 用标准化的数据与各个主成分的系数相乘之后再乘以各主成分的累积贡献率, 最后将各个主成分的综合得分值相加就得到最终得分。根据最终得分参照数据归一化的方法再与地表水环境质量标准进行对比最后得出水质类别, 具体的计算过程见下^[11-15]。

3 改进的 PCA 法在水库水质评价中的应用

3.1 双塔水库水质评价分析

选取 1993-2008 年双塔水库的溶解氧、高锰酸

钾盐指数、COD、BOD、氨氮、氟化物以及砷含量的年平均值作为代表双塔水库的水质指标进行评价分析。根据地表水环境质量标准(GB3838-2002)对双塔水库水质指标进行标准化处理, 结果见表 1。

表 1 标准化后的双塔水库水质指标表

年份	溶解氧	高锰酸钾盐指数	COD	BOD	氨氮	氟化物	砷
1993	0.895	0.995	0.241	0.750	0.000	0.420	0.160
1994	0.880	1.005	0.221	0.670	0.000	0.430	0.180
1995	0.975	1.055	0.223	0.670	0.000	0.550	0.240
1996	0.810	1.050	0.185	0.537	0.000	0.560	0.220
1997	0.825	1.055	0.171	0.370	0.000	0.550	0.240
1998	0.890	1.005	0.135	0.233	0.000	0.680	0.220
1999	0.950	0.985	0.133	0.200	1.229	0.660	0.220
2000	1.200	1.005	0.128	0.167	1.196	0.710	0.120
2001	0.865	0.970	0.400	1.847	1.237	0.730	0.120
2002	1.355	1.035	0.497	0.987	1.134	0.740	0.060
2003	1.295	0.340	0.348	0.853	0.000	0.790	0.920
2004	1.290	0.365	0.372	0.697	0.467	0.810	0.004
2005	1.225	0.285	0.045	0.000	0.000	0.580	2.100
2006	1.275	0.425	0.045	0.000	0.000	0.430	0.056
2007	1.435	0.165	0.283	0.000	0.000	0.490	0.002
2008	0.375	0.440	0.623	0.587	0.000	0.430	0.068

将标准化好的水质指标利用 SPSS16.0 软件进行主成分分析, 得到表 2、3、4 及图 2。

表 2 双塔水库水质相关系数矩阵表

	DO	KM	COD	BOD	N	F	S
DO	1.000	-0.408	-0.248	-0.227	0.168	0.398	0.193
KM	-0.408	1.000	-0.129	0.315	0.345	0.035	-0.359
COD	-0.248	-0.129	1.000	0.616	0.169	0.169	-0.334
BOD	-0.227	0.315	0.616	1.000	0.359	0.347	-0.218
N	0.168	0.345	0.169	0.359	1.000	0.587	-0.237
F	0.398	0.035	0.169	0.347	0.587	1.000	0.094
S	0.193	-0.359	-0.334	-0.218	-0.237	0.094	1.000

表 3 双塔水库水质主成分特征表

主成分	初始特征值			平方和		
	总比例	方差	累积	总数	方差	累积
		百分率	百分率		百分率	百分率
1	2.387	34.098	34.098	2.387	34.098	34.098
2	1.771	25.307	59.405	1.771	25.307	59.405
3	1.225	17.503	76.907	1.225	17.503	76.907
4	0.816	11.654	88.561			
5	0.381	5.441	94.002			
6	0.282	4.029	98.031			
7	0.138	1.969	100.000			

表 4 双塔水库水质主成分载荷矩阵表

主成分	1	2	3
DO	-0.258	0.822	0.024
KM	0.523	-0.356	0.705
COD	0.632	-0.110	-0.718
BOD	0.820	-0.016	-0.276
N	0.680	0.463	0.347
F	0.473	0.775	0.041
S	-0.537	0.375	-0.117

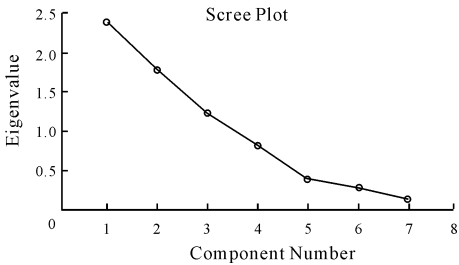


图 2 双塔水库水质主成分碎石图

从表 3、图 2 中可以看出,有 3 个主成分的特征根大于 1,第一、二、三主成分的贡献率分别为 34.098%、25.307% 和 17.503%,累计贡献率可达 76.907%,前 3 个主成分基本可代表双塔水库的水质信息,故选取前 3 个主成分来进行评价分析。根据主成分载荷矩阵表,将标准化好的水质指标与各个主成分的载荷值分别对应相乘再乘以各个主成分的贡献率之后相加就得到该年水质的综合得分。并将综合得分参照数据归一化的方法再与地表水环境质量标准进行对比最后得出水质类别,结果见表 5。

根据表 5 的水质评价结果,在 1997 年之前水库水质较好、属于 I 类水质。1997 年以后水库水质下降到 II、III 类,主要原因为:第一,1997 年以后水库受水量减少、进入水库及河道的污染物载荷相对提高;第二,当地大力发展经济,忽略了水环境保护,致使进入水体的污染物载荷量增大导致水质恶化。2000 年以后,当地采取了水环境污染治理和限制废污排放措施,双塔水库水质情况基本维持在 II 类左右,水质符合生活饮用、工业应用、农业灌溉、渔业养殖、景观娱乐等用水标准,由于上游来水水质、侧向泉水水质和水库蒸腾的影响,水质与上游有很大的区别,并且优于上游赤金峡水库的水质。

3.2 党河水库、赤金峡水库水质评价分析

选取 2001 - 2008 年党河水库的溶解氧、高锰酸钾盐指数、COD、BOD、氨氮、氟化物及砷含量的年均值代表水库水质指标进行评价分析。根据地表水环

境质量标准(GB3838 - 2002)对党河水库的水质指标进行标准化处理(由于处理步骤和双塔水库一致,限于篇幅,只列出最后的模型结果)。将标准化好的水质指标进行主成分分析,得到前 3 个主成分的特征根大于 1,第一、二、三主成分的贡献率分别为 50.503%、21.628% 和 18.897%,累计贡献率可达 91.028%,从而得出水质类别,见表 6。

表 5 双塔水库水质综合得分及水质类别表

年份	水质综合得分	水质级别	年份	水质综合得分	水质级别
1993	0.64346	I	2001	2.33669	III
1994	0.63462	I	2002	2.15698	III
1995	0.65890	I	2003	1.21425	II
1996	0.66094	I	2004	2.36452	III
1997	1.67863	II	2005	1.36841	II
1998	1.23658	II	2006	1.41257	II
1999	2.06986	III	2007	2.36922	III
2000	2.12369	III	2008	1.36911	II

表 6 党河、赤金峡水库水质综合得分及水质类别表

年份	党河水库		赤金峡水库	
	水质综合得分	水质级别	水质综合得分	水质级别
2001	1.99653	II	3.43592	IV
2002	1.22457	II	2.70635	III
2003	1.23568	II	2.40873	III
2004	0.98775	I	5.34754	V
2005	1.15699	II	4.22535	V
2006	0.87546	I	5.49686	V
2007	1.45692	II	5.59454	V
2008	1.23697	II	5.49571	V

由表 6 知,2000 年以来党河水库的水质基本保持在 I、II 类之间,水质情况良好,由于敦煌市排污口在水库下游,水库水质受到的点源污染较少,基本为非点源污染。水库上游为苏北蒙古族自治县,人口较为稀少,人类活动对水环境的影响较小也是水库水质较好的一个重要原因。

赤金峡水库位于疏勒河上游玉门市内污染较为严重的石油河上,2000 年以来,赤金峡水库的水质基本保持 V 类情况,主要是由于石油河是昌马河流域污染最为严重的一条河流,并且玉门市的排污口设置在赤金峡水库的上游严重的影响了水库水质,加强石油河的污染治理以及控制该河流上的点源污染已经刻不容缓。

4 结 语

结合地表水环境质量标准的主成分分析法对极

端干旱荒漠气候的疏勒河流域上、中、下游水库水质进行评价,去除监测项目中的重复信息得到的水质评价结果有:

(1)1997 年之前双塔水库属 I 类水质;1997 年以后水库水质下降到 II、III 类,主要原因为:水库受水量减少致使进入水体的污染物载荷相对提高和忽略水环境保护致使进入水体的污染物载荷量增大。2000 年以后水质情况基本维持在 II 类左右,水质符合生活饮用、工业应用、农业灌溉、渔业养殖、景观娱乐等用水标准。水库水质与上游有很大区别,且优于赤金峡水库。

(2)2000 年以来党河水库水质基本保持在 I、II 类之间,水库上游人口稀少、敦煌市排污口在水库下游致使水体受到的点源污染较少,基本为非点源污染。

(3)2000 年以来,赤金峡水库水质为 V 类,主要是由于石油河是昌马河流域污染最为严重的一条河流,并且玉门市的排污口设置在赤金峡水库的上游严重影响了水质,加强石油河的污染治理及控制该河流上的点源污染已经刻不容缓。

参考文献:

- [1] 郑建华,王世梅,王卓娟. 长江水质的综合评价及其预测[J]. 贵州水力发电,2006,20(6):7-12.
- [2] 李文杰. 梁子湖流域土地利用变化对流域水环境的影响[D]. 武汉:华中师范大学,2009:23-45.
- [3] 岳隽,王仰麟,李正国,等. 河流水质时空变化及其受土地利用影响的研究——以深圳市主要河流为例[J]. 水科学进展,2006,17(3):359-364.
- [4] Troiano J, Garreton C, Krauter C, et al. Influence of a-

mount and method of irrigation water application on leaching of Atrazine[J]. Journal of Environmental Quality, 1993, 22(2):290-298.

- [5] 肖长来,梁秀娟. 水环境监测与评价[M]. 北京:清华大学出版社,2008:50-62.
- [6] 邹海明,蒋良富,李粉茹. 基于主成分分析的水质评价方法[J]. 数学的实践与认识,2008,38(8):85-90.
- [7] 朱永生. 实验数据多元统计分析[M]. 北京:科学出版社,2009:13-20.
- [8] 任若恩. 多元统计数据分析——理论、方法、实例[M]. 北京:国防工业出版社,1997:31-52.
- [9] Trevor Hastie, Robert Tibshirani, Jerome Friedman. The elements of statistical learning[M]. Springer Series in Statistics, Springer,2009.
- [10] CERN Program Library. Detector description and simulation tool[M]. CERN, Geneva, Switzerland, 1994.
- [11] Jove S B, Connell T L, Miller L G, et al. Oxidation of ammonia and methane in an alkaline, saline lake[J]. Limnology and Oceanography, 1999, 44(1):178-188.
- [12] 张德刚,汤利,陈永川,等. 滇池流域典型城郊村镇排放污水氮、磷特征分析[J]. 农业环境科学学报,2007, 26(6):2245-2250.
- [13] 陈述云,张崇甫. 对多指标综合评价的主成分分析方法的改进[J]. 统计研究,1995(1):35-39.
- [14] Patrick Debels, Ricardo Figueroa, Roberto Urrutia, et al. Evaluation of water quality in the Chill'an river(Central Chile) using physicochemical parameters and a modified water quality index[J]. Environmental Monitoring and Assessment, 2005, 110(1-3):301-322.
- [15] Ouyang Ying. Evaluation of river water quality monitoring stations by principal component analysis[J]. Water Research, 2005, 39(12):2621-2635.

(上接第 155 页)

拟合曲线符合实际情况,对于研究混凝土的抗冻耐久性有一定的参考价值。然而,试验只进行了同一种掺量的纤维混凝土的抗冻试验,对于纤维掺量参数对混凝土的抗冻性有待进一步的研究。

参考文献:

- [1] 朱江,苏健波,李士恩. 聚丙烯纤维混凝土的力学性能研究[J]. 广西工学院学报. 2000,11(2):60-64.
- [2] 黄承逵. 纤维混凝土结构[M]. 北京:机械工业出版社,2004.
- [3] 钱庆炜,张经双. 纤维混凝土特性研究及应用前景[J].

西部探矿工程,2005,17(10):166-168.

- [4] 陈少峰,孙立,李振宝. 混凝土盐冻破坏的试验研究[J]. 公路,2006,47(4):216-219.
- [5] Kohn S D, Darter M I. Structural design of composite concrete pavement[J]. Proc. Second International Conference on Concrete Pavement Design, 1981(1):57-61.
- [6] 杨全兵,吴学礼,黄士元. 去冰盐对混凝土剥蚀的物理机理[J]. 上海建材学院学报,1991,4(4):341-346.
- [7] 蔡昊. 混凝土抗冻耐久性预测模型[D]. 北京:清华大学,1998.