

基于 PCA 的于桥水库富营养化因子分析

杜乔乔, 颜 坤

(天津水文水资源勘测管理中心, 天津 300221)

摘 要: 为了加强城市供水安全,保障库区饮用水质量,对于桥水库富营养状况及主要影响因素进行深入的分析。根据于桥水库库区的水质监测资料,利用主成分分析的方法对影响富营养化的监测指标如叶绿素 a,总氮,总磷,氨氮,高锰酸盐指数等进行分析,10项评价指标通过主成分分析降维到4项,前4个主成分能够解释73.6%的数据内容。在主成分筛选的基础上进一步进行因子分析,并对各因子的空间分布状况进行描述,以期对于桥水库生态管理提供理论依据。结果表明:水温、pH、COD及溶解氧是影响水库水体富营养化的主要驱动因子,水库中心区域的主成分值最高。

关键词: 水质评价; 富营养化; 主成分分析; 因子分析; 于桥水库

中图分类号:X524

文献标识码:A

文章编号:1672-643X(2013)02-0212-03

Factor analysis of eutrophication in Yuqiao reservoir based on PCA

DU Qiaoqiao, YAN Kun

(Tianjin Survey and Management Center of Hydrology and Water Resources, Tianjin 300221, China)

Abstract: Eutrophication status of Yuqiao Reservoir and the main influence factors were analyzed in order to strengthen the urban water supply safety and ensure the drinking water quality. Based on the method of principal component analysis with water quality monitoring data of Yuqiao reservoir, the indicators that influence eutrophication such as chlorophyll a, total nitrogen, total phosphorus, ammonia nitrogen, potassium permanganate index were analyzed to reduce dimensions from 10 to 4 and the former 4 principal components can explain 73.6% of the whole data. On the basis of main components screening the factor analysis were further carried out, and the space distribution of every factor was also described in order to provide the theory basis for Yuqiao Reservoir ecological management. Results show that the water temperature, PH, COD and dissolved oxygen are the main driving factors which influence reservoir eutrophication, and the principal components have the highest value in the central area of the reservoir.

Key words: assessment of water quality; eutrophication; PCA; factor analysis; Yuqiao Reservoir

1 概 述

于桥水库位于天津蓟县城东,是引滦入津工程重要的调蓄水库,也是天津市重要的水源地。2005年,于桥水库藻类爆发,直接威胁到天津市的供水安全,从城市安全供水的角度出发,于桥水库内水体的生态水平直接影响到城市居民饮用水的安全程度,因此,掌握富营养化的发生机理,从机理出发对众多影响因素进行降维处理,分析主要因素的作用,以期为防止富营养化提供依据。

主成分分析(PCA)是设法将原来众多(p 个)具有一定相关性的变量或指标,重新组合成一组新的、相互无关的少数几个综合变量的指标,以此来代替原来 p 个变量或指标,即由新变量表达原变量所

持有的信息量。这种降维的特性使其广泛应用于计算机、生物、水文、地理等各类学科研究。近年来越来越多的学者采用PCA进行水质评估、水体富营养化评价等研究^[1-3],周斌等^[4]应用主成分分析研究了山东半岛近岸海域的水体富营养化状况并指出总无机氮是该海域富营养化的主要驱动因子;李一平等^[5]借助因子分析方法确定太湖富营养化的主要驱动因子为营养盐和水体的物理环境。

本文根据2009年于桥水库各采样点的监测数据,首先利用PCA方法对影响水库水体富营养化的主要指标进行主成分的筛选,然后利用因子分析方法进行主要驱动因子的研究,从而分析水库水体富营养化形成的主要驱动因素。

2 结果与分析

2.1 主成分分析

选取于桥水库 2009 年库中心、库东、库西、库南、库北 5 个监测点每月 1 次的水质资料,应用统计分析 R 软件进行主成分分析,监测项目为叶绿素 a (CHLA)、总氮 (TN)、总磷 (TP)、高锰酸盐指数 (COD)、氨氮 (NH₃-N)、正磷酸盐 (PO₃)、溶解氧 (DO)、水温、pH 值、透明度 (SD) 等 10 个水质参数。为了消除由于单位不同所带来的影响,须将原始变量按公式(1)做标准化处理。

$$X_j^k = (X_j - \mu_j) / \sqrt{\sigma_{jj}} \tag{1}$$

式中: X_j^k 表示第 j 个指标的标准化值; X_j 表示第 j 个指标的实测值; μ_j 表示第 j 个指标的平均值; σ_{jj} 表示第 j 个指标的方差。

做主成分分析之前,对上述 10 个指标进行 KMO 检验和 Bartlett 球形检验,当 KMO 检验数大于 0.5, Bartlett 球形检验显著性概率 P 值小于 0.05 时,表明该组指标不相互独立,可以进行主成分分析,本研究中 KMO 统计量为 0.628, Bartlett 球形检验值小于 0.001。对上述 10 个指标的主成分分析结果表明,前 4 个主成分累计反映了原始参数信息的 73.6%,将原来的 10 个影响指标减少为 4 个,大大降低了因子的维数,各主成分贡献率及累计贡献率见表 1。由表 1 可知,第一主成分(Z_1)对总方差的贡献率达到了 31.06%,第二主成分(Z_2)对总方差的贡献率为 19.1%, Z_3 和 Z_4 对总方差的贡献率分别为 19.45% 和 12.66%,为研究各主成分对原始变量的关系还需将主成分载荷列出以便于分析,主成分载荷也就是原先指标线性组合成各主成分时的系数,系数(绝对值)越大,说明主成分对该变量的代表性越大,详见表 2。

表 1 主成分方差贡献率

成份	特征值	方差贡献率	累积方差贡献率
1	3.106	31.060	31.060
2	1.945	19.448	50.508
3	1.266	12.662	63.169
4	1.043	10.428	73.598

2.2 因子分析

主成分 Z_1 中不难发现很多指标的荷载均超过了 0.5,为了增强解释能力,突出与各主成分相关的因子载荷,需要对因子载荷矩阵按照方差最大化进行正交旋转,从而得到公因子较为突出的因子载荷矩阵。对

4 个筛选的主成分进行因子分析结果见表 3。由表 3 中可以看出, Z_1 中 pH、氨氮和总磷占的因子载荷较大,说明第一主成分反映的信息量较丰富,不仅体现了藻类生长所需要的营养盐条件,还反映了水体的酸碱环境条件,可以概括为营养盐及化学环境因子; Z_2 中 COD 和透明度占的载荷较大,另外总氮的荷载也有所突出,说明 Z_2 主要反应的是水体中有机物及氮营养盐的多少,有机物含量高会直接促进藻类繁殖,导致水体富营养化,因此该成分归纳为水体受污染的状态; Z_3 中水温的载荷最大,适宜的水温是促进藻类生长的重要条件,因此该主成分主要反映了水体温度环境,此外磷酸盐和叶绿素相应的载荷也较大,叶绿素指标直接反应了水体富营养化的程度; Z_4 中溶解氧的载荷较大,说明 Z_4 主要反应了水库水体的自净能力,溶解氧的含量也与藻类生长密切相关。

表 2 主成分载荷

成分	1	2	3	4
TP	0.430	0.374	0.589	-0.256
TN	0.827	0.169	-0.157	0.000
CHLA	-0.640	0.439	-0.065	0.030
NH ₃ -N	-0.582	-0.632	-0.077	-0.129
COD	-0.546	0.250	0.601	-0.070
PO ₃	0.612	-0.090	0.059	-0.384
DO	0.207	-0.119	0.581	0.702
水温	-0.458	0.545	-0.332	0.374
pH	0.167	0.891	-0.146	-0.109
SD	0.732	-0.095	-0.227	0.404

表 3 因子分析

成分	1	2	3	4
TP	0.644	0.244	-0.473	0.197
TN	0.496	-0.620	-0.325	0.017
CHLA	0.076	0.460	0.598	-0.179
NH ₃ -N	-0.815	0.264	0.013	-0.161
COD	0.077	0.816	0.121	0.200
PO ₃	0.247	-0.258	-0.622	-0.141
DO	0.023	-0.035	-0.032	0.940
水温	0.174	0.072	0.846	-0.069
pH	0.835	0.003	0.299	-0.258
SD	0.174	-0.790	-0.134	0.296

为了进一步研究影响水库水体富营养化的主要及次要驱动因子,对各指标与主成分之间的相关系数进行了计算,并通过显著性检验分析该指标是否为驱动该主成分的主要因子详见表 4。表 4 中氨氮 (NH₃-N) 与 Z_1 的相关系数最高且在显著性水平为 0.01 下具有显著的相关关系,因此氨氮是第一主成分的主要驱动因子; COD 与 Z_2 的相关系数最高且都通过了显著性水平为 0.01 的检验说明 COD 为 Z_2

的主要驱动因子,类似的水温和溶解氧分别为 Z_3 和 Z_4 的主要驱动因子。总体上看,影响于桥水库富营 养化状况的主要驱动因子为氨氮、COD、水温、及溶解氧,这些因素也是控制水体富营养化的关键因素。

表 4 各指标与主成分的相关系数及显著性 p 值

	Z_1		Z_2		Z_3		Z_4	
	r	p	r	p	r	p	r	p
TP	0.526	0.000	0.075	0.479	-0.399	0.000	0.231	0.028
TN	0.466	0.000	-0.582	0.000	-0.268	0.010	0.070	0.509
CHLA	0.015	0.887	0.459	0.000	0.610	0.000	-0.265	0.011
NH ₃ -N	-0.847	0.000	0.272	0.009	0.018	0.868	-0.135	0.203
COD	0.079	0.458	0.823	0.000	0.136	0.200	0.289	0.005
PO ₃	0.121	0.252	-0.282	0.007	-0.661	0.000	-0.118	0.267
DO	0.082	0.440	-0.005	0.960	-0.012	0.913	0.935	0.000
水温	0.089	0.401	0.100	0.346	0.823	0.000	-0.083	0.433
pH	0.760	0.000	0.019	0.855	0.384	0.000	-0.270	0.010
SD	0.165	0.118	-0.796	0.000	-0.226	0.032	0.159	0.132

注:表中 0.000 代表在显著性水平 0.01 下显著相关, p 值小于 0.05 代表在显著性水平 0.05 下显著相关。

2.3 因子空间分布分析

分别对于桥水库 5 个监测点位的主成分进行统计计算,以期分析各主成分在空间的分布状况(表 5)。

表 5 不同监测点位主成分均值

监测点	Z_1	Z_2	Z_3	Z_4
库中心	7.461	3.352	23.054	-1.178
库东	6.916	3.106	21.236	-1.049
库南	6.830	3.119	21.306	-1.043
库西	6.856	3.181	21.494	-1.046
库北	6.883	3.309	21.460	-1.054

由表 5 可发现,整体上 4 个主成分在库中心的值都明显高于其他监测点,说明相较其他地方,库中心的富营养化程度最高。代表着营养盐状况及水体化学环境的 Z_1 在库中心的值明显高于其他几个监测点,说明水库库中心区域的营养盐含量较高,也是最容易发生藻类爆发的地点; Z_2 在库中心和库北两个站点的值也较高,说明在水体受污染状况上这两个区域比较明显; Z_3 主要代表水体温度情况,结果表明,于桥水库库中心温度最高,库东温度略小于库西,库南温度略小于库北; Z_4 代表着水体的自净能力,库中心的数值最低说明自净能力较差,其次为库北,两个地方都比较容易形成富营养化的状态。

3 结 语

本文利用主成分分析法及因子分析法对于桥水库水体富营养化驱动因子进行了研究,将 10 个影响指标减少成 4 个互不相关的综合性指标,并且累计方差贡献率达到了 73.6%。通过对主成分的分析发现个主成分主要体现的综合指标分别为水体营养盐状况及化学环境因子、水体受污染状态因子、水体

温度环境因子以及水体自净能力因子。在因子分析中发现,影响于桥水库水体富营养化的主要驱动因子为氨氮、COD、水温、及溶解氧。相关文献也指出由于上游地区营养盐随着地表水过度流入以及库区底质等对营养盐的释放等导致库区营养盐相对很高^[6-8],并且畜禽养殖和农业用地等人类活动对其流域内氮磷等非点源污染物的贡献量是最为显著的^[9]。对各因子的空间分布分析得出,库中心是最易形成富营养化并导致藻类疯长的区域,因此需要加强对水质的监督和管理。

参考文献:

[1] 柴超,俞志明,宋秀贤,等.长江口水域富营养化特性的探索性数据分析[J].环境科学,2007,28(1):53-58.

[2] Solidoro C, Pastres R, Cossarinia G, et al. Seasonal and spatial variability of water quality parameters in the lagoon of Venice [J]. Journal of Marine Systems, 2004,51:7-18.

[3] 史小红,李畅游,贾克力.乌梁素海污染现状及驱动因子分析[J].环境科学与技术,2007,30(4):37-39.

[4] 周斌,王悠,王进河,等.山东半岛南部近岸海域富营养化状况的多元评价研究[J].海洋学报,2010,32(2):128-137.

[5] 李一平,逢勇,李俊,等.太湖富营养化的驱动因子分析[J].河海大学学报(自然科学版),2004,32(6):645-647.

[6] 陈水蓉,吴光红,苏睿先.天津市于桥水库水质变化特点及影响因素分析[J].水资源保护,2011,27(4):27-30.

[7] 王立林,杜玉凤,王鸿雁.于桥水库富营养化控制对策研究[J].水科学与工程技术,2007(6):23-26.

[8] 陈福贵,赵翠霞,王立林,等.于桥水库富营养化成因分析[J].水科学与工程技术,2008(5):56-58.

[9] 杨光,刘伟江,孙韧,等.天津市于桥水库引用水源地保护对策研究[J].水资源与水工程学报,2011,22(6):43-45.