

因子分析法在汤河水库富营养化研究中的应用

杨瑞坤¹, 宓永宁¹, 胡志民², 方广玲¹

(1. 沈阳农业大学水利学院, 沈阳 110161; 2. 沈阳市水利工程质量监督站, 沈阳 110032)

摘 要: 水体富营养化是湖泊和水库面临的主要水质问题。运用综合营养状态指数法对汤河水库各断面营养水平进行评价, 得出其目前的富营养化程度比较轻。因子分析法的应用表明, 氮、磷和水温是汤河水库水体富营养化的主要影响因素。

关键词: 因子分析; 富营养化; 汤河水库

中图分类号: X524

文献标识码: A

文章编号: 1005-3409(2007)03-0324-02

Application of Factor Analysis Method to the Eutrophication Study of Tanghe Reservoir

YANG Rui-kun¹, MI Yong-ning¹, HU Zhi-min², FANG Guang-ling¹

(1. College of Water Conservancy, Shenyang Agricultural University, Shenyang 110161;

2. Shenyang Supervision Station of Water Conservancy Project, Shenyang 110032, China)

Abstract: Eutrophication is most important to the water quality of lake and reservoir. TLI method is used to estimate Tanghe reservoir's eutrophication level, and it indicates that the level of Tanghe reservoir's eutrophication is light. Application of factor analysis show that the nitrogen, phosphorus and water temperature are the main factors which can affect the reservoir's eutrophication.

Key words: factor analysis; eutrophication; Tanghe reservoir

1 引 言

湖泊和水库都属于静水环境, 水流速度很低, 进入湖泊和水库的营养物质在其中不断的积累, 这就使富营养化成为湖泊和水库面临的主要水质问题。所谓富营养化是指这样一种过程: 低浮游生物生产率(低营养物输入)的湖泊与水库变成高浮游生物生产率(高营养物输入)的湖泊与水库; 结果本来是蓝色的水体由于藻类的大量生长会变成绿色的水体, 先是有少量的硅藻和绿藻, 接着就是大量的蓝绿藻, 这时水体开始出现臭味, 水质变坏。^[1]

从湖泊的发展过程看, 由贫营养向富营养的过渡是一个正常过程, 在自然状态下, 这个过程进展非常缓慢, 通常需要用地质年代的时间尺度来衡量。但由于人类大规模的生产活动, 此过程的演化速度大大加快。影响水体富营养化的除 N、P 等营养物质外, 还包括气候因子(如光、温和降水等)、地理因子(如湖盆、河底的地质状况等)和生物因子等。本文运用因子分析法, 对水体富营养化评价过程中的几项常用指标进行分析, 明确影响汤河水库富营养化的主要影响因素, 为其富营养化的防治提供决策建议。

2 因子分析法概述

因子分析法是研究一组变量之间相关关系的一种多元统计方法, 对已知的数据进行分析计算, 分析和识别出对高维度变量有公共影响的公因子, 并对其加以评析, 得出科学合理的评价, 通过分析所得信息为决策提供科学的依据。^[2]

因子分析数学模型^[3]:

$$\begin{cases} x_1 = \alpha_{11}F_1 + \alpha_{12}F_2 + \dots + \alpha_{1m}F_m + \alpha_1\varepsilon_1 \\ x_2 = \alpha_{21}F_1 + \alpha_{22}F_2 + \dots + \alpha_{2m}F_m + \alpha_2\varepsilon_2 \\ \dots \\ x_p = \alpha_{p1}F_1 + \alpha_{p2}F_2 + \dots + \alpha_{pm}F_m + \alpha_p\varepsilon_p \end{cases} \quad (1)$$

式中: $x_1, x_2, x_3 \dots x_p$ 为 p 个原有变量, 是均值为 0、标准差为 1 的标准化变量, $F_1, F_2, F_3 \dots F_m$ 为 m 个因子变量(或公共因子), m 小于 p , 表示为矩阵的形式为:

$$X = AF + \alpha\varepsilon$$

式中: F ——因子变量或公共因子, A ——因子载荷矩阵, α_j ——因子载荷, 是第 i 个原有变量在第 j 个因子变量上的负荷, ε ——特殊因子, 表示了原有变量不能被因子变量所解释的部分, 相当于多元回归分析中的残差部分。

3 汤河水库富营养化研究

3.1 汤河水库自然概况

汤河水库位于辽河流域太子河支流汤河干流上, 坐落在辽阳市弓长岭区汤河乡境内。汤河是太子河中游左岸的一条较大支流, 全长 90.9 km, 流域面积 1 460 km², 汤河水库控制流域面积 1 228 km², 占汤河流域面积的 84%, 占太子河流域面积的 9%。汤河水库流域属于辽阳市东南山区, 有长白山余脉、千山山脉穿过, 山地面积占 90%, 平均山岭高

* 收稿日期: 2006-06-16

作者简介: 杨瑞坤(1981-), 男, 硕士研究生, 通讯作者: 宓永宁(1959-), 男, 教授, 博士, 主要从事岩土工程材料与水环境方面的教学与科研工作。

度在 400~ 450 m,属中低山区。流域内山地林木丛生,植被良好,水资源丰富。汤河流域属寒温带季风气候,早晚温差较大,四季冷暖、干湿分明。多年平均降雨量为 774 mm,多年平均径流量为 2. 89 亿 m³。

3.2 汤河水库营养化评价

湖泊水库富营养化评价,就是通过与湖泊水库营养状态有关的一系列指标及指标间的相互关系,对湖泊水库的营养状态作出切合实际的判断。本文运用营养状态综合指数法^[4],选取 2005 年各断面水质监测数据的年平均值(见表 2)对汤河水库 2005 年各断面的营养状态进行评价,见表 1:

表 1 汤河水库 2005 年各断面营养化水平

| 断面名称 | 坝前 | 库中 | 东叉中 | 东叉头 | 西叉中 | 西叉头 |
|----------|-------|-------|-------|-------|-------|-------|
| 营养状态综合指数 | 29.32 | 29.56 | 31.15 | 32.48 | 31.35 | 32.88 |
| 营养化水平 | 贫营养 | 贫营养 | 中营养 | 中营养 | 中营养 | 中营养 |

分析表 1,汤河水库坝前和库中断面的营养状态为贫营养水平,其他断面均处于中营养水平。由于各断面的营养状态指数均分布在 30 附近,所以营养程度均比较轻。

4 汤河水库富营养化因子分析

水体富营养化是因 N、P 等营养物质的过量输入而引起的一种水体效应。本文选用水体富营养化评价中经常使用的各项指标,以 2005 年各断面监测值的年平均值(见表 2)为依据,使用 DPS 数据处理系统软件做因子分析,探究影响汤河水库水体富营养化的主要影响因素。

表 2 汤河水库 2005 年各断面监测结果平均值

| 断面名称 | COD _{Mn} / (mg·L ⁻¹) | BOD ₅ / (mg·L ⁻¹) | TN/ (mg·L ⁻¹) | TP/ (mg·L ⁻¹) | Chla/ (mg·m ⁻³) | 水温 /℃ | SD /m | DO/ (mg·L ⁻¹) |
|------|--|---|------------------------------|------------------------------|--------------------------------|----------|----------|------------------------------|
| 坝前 | 2.9175 | 1.2000 | 0.6958 | 0.0055 | 1.9992 | 19 | 3.4875 | 7.8500 |
| 库中 | 2.8814 | 1.0000 | 0.5914 | 0.0050 | 2.8871 | 19 | 3.5143 | 7.8714 |
| 东叉中 | 2.8857 | 1.1571 | 0.6329 | 0.0050 | 3.3757 | 19 | 2.7143 | 7.3857 |
| 东叉头 | 2.9400 | 1.1714 | 0.5814 | 0.0050 | 3.9929 | 19 | 2.0429 | 7.3571 |
| 西叉中 | 3.0643 | 1.0000 | 0.6257 | 0.0050 | 3.2686 | 19 | 2.6857 | 8.4857 |
| 西叉头 | 2.9186 | 1.1714 | 0.6871 | 0.0050 | 3.7014 | 19 | 1.9571 | 7.7857 |

4.1 建立相关系数矩阵

在实际问题中,往往变量的量纲不同,不同的量纲会引起变量取值的分散程度差异很大。为了消除由于量纲的不同可能带来的影响,将原始数据 Z_{ij} ($i = 1, 2, \dots, n, j = 1, 2, \dots, p$, 其中 i 为变量个数, p 为样本点)标准化为均值为 0、方差为 1 的变量,有:

$$x_{ij} = \frac{z_{ij} - \bar{z}_j}{S_{jj}} \tag{2}$$

$$\bar{z}_j = \frac{1}{n} \sum_{i=1}^n z_{ij} \tag{3}$$

$$S_{jj}^2 = \frac{1}{n-1} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2 \tag{4}$$

式中: \bar{z}_j 、 S_{jj} ——原始变量的均值和样本标准差。

对标准化数据计算 $R = \frac{1}{n} X'X$, 得相关矩阵 R , 见表 3:

表 3 各指标相关矩阵

| | COD _{Mn} | BOD ₅ | TN | TP | Chla | 水温 | SD | DO |
|-------------------|-------------------|------------------|---------|---------|---------|---------|---------|---------|
| COD _{Mn} | 1.000 | - 0.420 | - 0.059 | - 0.124 | 0.143 | 0.124 | - 0.225 | 0.740 |
| BOD ₅ | - 0.420 | 1.000 | 0.510 | 0.446 | 0.009 | - 0.446 | - 0.338 | - 0.670 |
| TN | - 0.059 | 0.510 | 1.000 | 0.620 | - 0.467 | - 0.620 | 0.055 | 0.179 |
| TP | - 0.124 | 0.446 | 0.620 | 1.000 | - 0.849 | - 1.000 | 0.549 | 0.072 |
| Chla | 0.143 | 0.009 | - 0.467 | - 0.849 | 1.000 | 0.849 | - 0.884 | - 0.319 |
| 水温 | 0.124 | - 0.446 | - 0.620 | - 1.000 | 0.849 | 1.000 | - 0.549 | - 0.072 |
| SD | - 0.225 | - 0.338 | 0.055 | 0.549 | - 0.884 | - 0.549 | 1.000 | 0.279 |
| DO | 0.740 | - 0.670 | 0.179 | 0.072 | - 0.319 | - 0.072 | 0.279 | 1.000 |

4.2 特征值和特征向量的计算

解特征方程 $|M - R| = 0$, 求出特征值 λ 及特征向量 u , 见表 4 和表 5:

表 4 特征值和方差累计贡献率

| No | 特征值 | 百分率/% | 累计百分率/% |
|----|---------|----------|----------|
| 1 | 3.81873 | 47.73418 | 47.73418 |
| 2 | 2.369 | 29.61245 | 77.34663 |
| 3 | 1.33805 | 16.72565 | 94.07228 |
| 4 | 0.44083 | 5.51038 | 99.58266 |
| 5 | 0.03339 | 0.41734 | 100 |
| 6 | 0 | 0 | 100 |
| 7 | 0 | 0 | 100 |
| 8 | 0 | 0 | 100 |

表 5 单位特征向量

| | 因子 1 | 因子 2 | 因子 3 | 因子 4 | 因子 5 | 因子 6 | 因子 7 | 因子 8 |
|-------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| COD _{Mn} | - 0.113 | 0.442 | 0.530 | - 0.487 | 0.510 | - 0.001 | - 0.034 | - 0.115 |
| BOD ₅ | 0.178 | - 0.563 | 0.288 | - 0.183 | 0.218 | 0.101 | 0.183 | 0.666 |
| TN | 0.342 | - 0.108 | 0.485 | 0.687 | 0.252 | 0.116 | - 0.102 | - 0.275 |
| TP | 0.496 | - 0.038 | 0.093 | - 0.314 | - 0.302 | 0.228 | 0.592 | - 0.390 |
| Chla | - 0.475 | - 0.193 | 0.188 | - 0.053 | - 0.223 | 0.777 | - 0.170 | - 0.130 |
| 水温 | - 0.496 | 0.088 | - 0.093 | 0.314 | 0.302 | - 0.003 | 0.745 | - 0.012 |
| SD | 0.343 | 0.278 | - 0.519 | 0.041 | 0.447 | 0.553 | - 0.045 | 0.161 |
| DO | 0.063 | 0.598 | 0.281 | 0.236 | - 0.444 | 0.124 | 0.136 | 0.522 |

根据方差累计贡献率的要求,即 $\sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i \geq 85$, 取前 m 个因子作为因子变量。由表 4 可知,由前 3 个主因子代替原来 8 因子的累积贡献率达 94.07%,因此取前三个主因子做因子载荷分析。

4.3 因子载荷阵

因子载荷 $a_{ij} = u_{ij} \cdot \sqrt{\lambda_j}$, 是第 i 个原有变量在第 j 个因子变量上的负荷。在因子载荷矩阵中,往往某一行或某一列中可能多个 a_{ij} 比较大,说明某个变量 x_i 可能同时与几个因子有比较大的相关关系或某个因子变量可能解释多个原变量的信息。在实际分析中,希望对因子变量的含义有比较清醒的认识,让它只解释一小部分原变量的信息。因此,在因子分析时经常对因子载荷矩阵进行旋转,目的是使每个变量仅在一个公共因子上有较大的载荷,而在其余公共因子上的载荷比较小。本文采用方差最大法对因子载荷阵进行旋转,结果见表 6:

表 6 因子载荷阵方差最大旋转结果

| | 因子 1 | 因子 2 | 因子 3 |
|-------------------|-----------|-----------|-----------|
| COD _{Mn} | 0.01471 | 0.90842 | 0.24794 |
| BOD ₅ | 0.73276 | - 0.60117 | 0.29213 |
| TN | 0.87988 | 0.09855 | - 0.07528 |
| TP | 0.77071 | - 0.05498 | - 0.59709 |
| Chla | - 0.42301 | - 0.08954 | 0.90009 |
| 水温 | - 0.77071 | 0.05498 | 0.59709 |
| SD | - 0.04324 | 0.02541 | - 0.99502 |
| DO | - 0.00204 | 0.94976 | - 0.25847 |

通过对表 6 的分析可以得出,生化需氧量(BOD₅)、总氮(TN)、总磷(TP)和水温在第一主因子上的载荷较大,对其贡献明显,表明氮、磷作为营养物质和水温是汤河水库富营养化的主要影响因素。对于第二主因子,高锰酸钾指数(COD_{Mn})和溶解氧(DO)的贡献突出,反映了水体中有机物和溶解氧对富营养化的影响。而对第三主因子有重要贡献的叶绿素(Chla)和透明度(SD)是表征水体富营养化的重要指标。

(下转第 329 页)

4 结 论

(1) 研究地点 29 年生油松林的生物量为 92 623 kg/hm², 在不同器官的分配是叶为 12 762 kg/hm²; 干为 49 946 kg/hm²; 枝为 17 659 kg/hm²; 根为 12 260 kg/hm², 依次是干> 枝> 叶> 根。

(2) 油松林各器官中养分元素含量差异较大, 5 种元素在地上器官中的含量顺序均为叶> 枝> 干, 在根系中的含量顺序为小根> 大根> 根颈。针叶中的 N、P、K、Mg 的含量在一龄叶中的最高, 二、三龄叶中的含量有所降低, 二龄叶中的含量均大于三龄叶中的含量, 唯独 Ca 的含量与此不符, Ca 参考文献:

[1] 吴征镒. 中国植被[M]. 北京: 科学出版社, 1980. 212.

[2] 余新晓, 于志民, 王礼先, 等. 水源保护林培育 经营 管理 评价[M]. 北京: 中国林业出版社, 2001. 24.

[3] 董鸣, 王义凤, 孔繁志, 等. 陆地生物群落调查观测与分析[M]. 北京: 中国标准出版社, 1996. 152– 159.

[4] 刘光崧, 蒋能慧, 张连第, 等. 土壤理化分析与剖面描述[M]. 北京: 中国标准出版社, 1996. 121– 192.

[5] 谢贤群, 王立军. 水环境要素观测与分析[M]. 北京: 中国标准出版社, 1998. 279– 297.

[6] 沈国舫, 董世仁, 聂道平, 等. 油松人工林营养循环研究 I: 营养元素的含量及分布[J]. 北京林业大学学报, 1985, 7 (4): 1– 14.

[7] 陈灵芝, 胡肆慧, 孔繁志, 等. 人工油松林的化学元素特征[J]. 植物学报, 1987, 29(3): 302– 308.

[8] 刘茜, 项文化, 蔡宝玉, 等. 湿地松人工林养分循环及密度关系的研究[J]. 林业科学, 1998, 34(3): 11– 17.

[9] I J Fernander, R A Strcht emeyer. Correlation between element concentration in spruce foliage and forest soils[J]. Commun Soil Sci Plant Anal, 1984, (15): 1243– 1255.

[10] 高甲荣, 肖斌. 桥山林区油松人工林营养元素分配与积累的研究[J]. 应用生态学报, 2001, 12(5): 667– 671.

[11] 刘广全, 倪文近, 刘慧芳, 等. 秦岭南坡锐齿栎林的生态环境及其营养积累[J]. 应用生态学报, 2002, 13(5): 513– 518.

[12] 高甲荣. 秦岭火地塘林区油松人工林营养元素生物循环的研究[J]. 西北林学院院报, 1987, 2(1): 23– 35.

[13] 陈灵芝, 任继凯, 鲍显诚, 等. 北京西山(卧佛寺附近) 人工油松群落特征及生物量研究[J]. 植物生态学与地植物学丛刊, 1984, 8(3): 173– 181.

[14] 高甲荣, 张东升, 肖斌, 等. 黄土区油松人工林生态系统营养元素分配格局和积累的研究[J]. 北京林业大学学报, 2002, 24(1): 26– 30.

(上接第 325 页)

4.4 计算样本因子得分

在对实际问题的分析研究中, 往往希望得到每个样本在不同因子上的具体得分 F_{ij} , 从而计算各样本的综合得分 F_i , 结果见表 7。

$$F_{ij} = \sum_{k=1}^8 x_{ik} \cdot p_{jk} \tag{5}$$

$$F_i = \sum_{j=1}^3 F_{ij} \cdot \lambda_j / \sum_{j=1}^3 \lambda_j \tag{6}$$

式中: x_{ik} —— 标准化后的第 i 个断面第 k 个变量的数值($i = 1, 2, \dots, 6; j = 1, 2, 3$), p_{jk} —— 第 k 个变量在第 j 个主因子上经旋转的载荷。

表 7 样本因子得分

| No | Y(i, 1) | Y(i, 2) | Y(i, 3) | 综合得分 |
|-----|------------|------------|------------|------------|
| 坝 前 | - 3. 66988 | - 0. 02872 | 4. 23215 | - 1. 11875 |
| 库中 | 0. 19812 | 0. 3649 | - 6. 28357 | - 0. 9018 |
| 东叉中 | - 11. 4546 | - 3. 71147 | 25. 73485 | - 2. 40504 |
| 东叉头 | 5. 6084 | 0. 22519 | - 10. 0909 | 1. 122579 |
| 西叉中 | - 0. 98237 | 1. 90749 | 1. 84827 | 0. 430589 |
| 西叉头 | 10. 30029 | 1. 24261 | - 15. 4408 | 2. 872424 |

参考文献:

[1] 郑丹, 陈春云. 环境系统数学模型[M]. 北京: 化学工业出版社, 2003. 80.

[2] 于秀林, 任雪松. 多元统计分析[M]. 北京: 中国统计出版社, 1999. 154– 162.

在二龄叶中的含量最高, 一龄叶中的含量最小, 三龄叶中的含量居中。

(3) 油松林养分元素贮存总量为 695. 17 kg/hm², 叶中贮存的养分总量为 291. 94 kg/hm², 树枝中贮存的养分总量为 206. 33 kg/hm², 树干中贮存的养分总量为 110. 09 kg/hm², 根系中贮存的养分总量为 86. 81 kg/hm²。各器官中养分贮存量大小排序为叶> 枝> 干> 根。

(4) 油松林中 N 元素是 5 种养分元素中富集能力最强的元素, 不同器官富集系数总的趋势是叶> 枝> 根> 干。油松林生态系统每积累 1 t 干物质需 N、P、K、Ca 和 Mg 元素共计 7. 51 kg, 所需的养分元素的顺序为 N > Ca > K > Mg > P。

分析表 8 可以得出, 在汤河水库的 6 个监测断面中, 西叉头和东叉头断面的得分最高, 西叉中断面的得分次之, 而坝前和库中断面的得分较低, 从而也反映了各断面营养水平排名。除了东叉中断面, 其他断面的营养水平排序与综合营养状态指数法的评价结果一致。

5 结 论

因子分析的出发点是用较少的相互独立的因子变量来代替原有变量的大部分信息。本文运用因子分析法, 对水体富营养化评价过程中经常使用的指标进行了分析研究, 在该几项指标中, 氮、磷、生化需氧量和水温在第一主因子上的载荷较大, 化学需氧量和溶解氧对第二主因子的贡献较大, 而第三主因子则体现了水体富营化的表征指标, 透明度和叶绿素 a 含量。通过对以上各指标的因子分析, 表明汤河水库富营养化的主要影响因素为营养物质氮、磷的输入和水温, 与一般认为氮、磷是水体富营养化主要影响因素的结论相同, 而样本因子得分也符合综合营养状态指数法对各断面的评价结果。因此认为因子分析法在汤河水库富营养化的研究中较为适用, 可以为其富营养化的防治提供决策建议。