

基于主成分分析的河流健康指标权重分配研究

周必翠¹, 吕照根¹, 舒持恺², 朱大伟³, 杨 侃²

(1. 句容水利农机局, 江苏 句容 212400; 2. 河海大学 水文水资源学院, 江苏 南京 210098;
3. 江苏省水利勘测设计研究院有限公司, 江苏 扬州 225127)

摘要: 河流健康是生态环境保护工作的重要内容, 合理评估河流健康状况需要恰当的评价体系与指标权重分配。针对权重分配中难以权衡的问题, 综合运用主客观赋权法, 先基于主成分分析得到一种权重, 再利用改进的 AHP 与熵值法得到一种权重, 在不满足一致性的条件下, 构造一种基于博弈论思想的集合赋权法得出最终权重, 可以较好的包含各种权重分配方法的信息。

关键词: 河流健康评价; 权重分配; 组合赋权方法; 主成分分析

中图分类号: X171 **文献标识码:** B **文章编号:** 1007-7839 (2017) 06-0016-04

Study on the weight distribution of river health index based on principal component analysis

ZHOU Bicui¹, LV Zhaogen¹, SHU Chikai², ZHU Dawei³, YANG Kan²

(1. Water Conservancy and Agricultural Machinery Bureau of Jurong, Zhenjiang 212400, Jiangsu;
2. College of hydrology and water resources, Hohai University, Nanjing 210098, Jiangsu;
3. Jiangsu Water Conservancy Survey and Design Institute Co., Ltd, Yangzhou 225127, Jiangsu)

Abstract: River health is an important part of the ecological environment protection. It is necessary to evaluate the health status of rivers and to evaluate the distribution of index weight. For it is difficult to balance the weight distribution problems in the comprehensive use of subjective and objective weighting method based on principal component analysis to get a weight, then get a weight by AHP and improved entropy method, in the inconsistent conditions, construct a game theory set weighting method obtains the final weight based on can good information contains a variety of weight allocation method.

Key words: river health assessment; weight distribution; combination weighting method; principal component analysis

随着我国社会经济的快速发展, 社会资源矛盾日益突出, 水资源短缺不仅仅是资源型缺水, 更存在着水质型缺水, “无水可用”与“有水难用”问题并存, 水资源供需矛盾日渐严重, 极大影响着我国社会经济的可持续发展^[1-2]。解决水质型缺水

问题的根本是提高水质, 改善水体环境, 一方面对于已受到污染的水体加以净化, 使之达到相关的水质要求, 另一方面对于未受到污染的水体进一步加强监控, 防止其受到污染^[3]。河流作为一种地表水体, 是非常重要的水资源, 水环境保护其

收稿日期: 2017-02-28

基金项目: 国家重点基础研究发展计划(973 计划)(2012CB417006)

作者简介: 周必翠(1982-), 女, 硕士, 工程师, 主要从事农田水利工作。

中一项非常重要的任务就是对河流生态系统的保护。

人类文明起源于河流, 河流提供水资源是人类社会发展的基础, 健康的河流生态系统能够有效地调节气候状况、改善生态系统环境。河流健康评价通过建立一套完整的评估体系, 能够帮助我们认识到当前河流的健康状况, 为管理者提供信息, 有效治理河流污染情况, 促进河流生态系统朝着良性方向发展。在河流健康评价工作中, 指标权重分配是非常重要的部分。

目前针对权重分配中难以权衡的问题, 多数研究采用组合赋权的方法, 山成菊等^[7]在计算河流健康指标权重时, 采用基于博弈论方法的组合赋权将层次分析法与熵值法进行组合, 一定程度上保证了赋权的准确性, 但仅仅靠2种权重进行组合, 其全面性值得商榷。本文综合运用主客观赋权法, 先基于主成分分析得到一种权重, 再利用改进的AHP与熵值法得到一种权重, 在不满足一致性的条件下, 构造一种基于博弈论思想的集合赋权法得出最终权重, 可以较好包含各种权重分配方法的信息。

1 指标权重分配计算方法

目前, 权重的计算方法主要分为主观赋权与客观赋权两大类^[4-5]。其中, 主观赋权法主要有层次分析法(AHP)、专家调查法(Delphi法)等, 权重主要根据专家学者经验与主观意愿进行权衡, 其原始数据由专家学者主观判断而得到; 客观赋权法主要有主成分分析法、熵值法等, 权重由指标数据进行分析得出, 其原始数据由各指标在被评价单位中的实际数值形成^[6]。无论是哪种方法计算权重, 都有其局限性, 如主观赋权法虽然考虑到了专家学者的偏好, 却忽视了指数据本身的效用值, 得到的权重可能是偏离实际的。客观赋权法虽然较好地考虑到了指标数据本身价值, 但无法照顾到决策者的主观偏好, 得到的权重可能是不尽人意的。为解决这种问题, 本文采用一种组合赋权, 引入主成分分析到河流健康评价指标权重分配计算中, 结合改进的AHP与熵值法, 然后基于博弈论思想进行综合赋权, 以改善权重分配计算的准确性与全面性。

2 基于主成分分析的指标权重分配步骤

通过集合不同的权重分配方法, 可以兼顾各种方法的优点, 弥补各自的不足, 提高权重分配精度。对于同一对象不同计算方法其结果一致性的检查, 可以用Kendall-W协和系数进行检验^[7], 利用博弈论思想将单一的主成分分析法和AHP与熵值法集成得到综合权重。

2.1 主成分分析法

(1) 标准化处理

在进行计算前, 需要对数据进行标准化处理, 主要包括数据的同趋化处理与无量纲化处理。同趋化处理是使正向指标与逆向指标都具有相同的数值效应, 即都是越大越好或越小越好, 可按下式处理:

$$y_{ij} = \begin{cases} \frac{x_{ij}}{x_j}, & \text{第}j\text{个指标为正向指标} \\ \frac{x_j}{x_{ij}}, & \text{第}j\text{个指标为逆向指标} \end{cases} \quad (1)$$

式中:

y_{ij} —经过同趋化处理的第*i*个样本第*j*个指标值;

x_{ij} —第*i*个样本第*j*个指标原始数值;

x_j —第*j*个指标平均值。

无量纲化主要是使不同指标数据处于同一量级, 可以进行两两对比, 可按下式处理:

$$z_{ij} = \frac{y_{ij} - y_j}{\sigma_j} \quad (2)$$

式中:

z_{ij} —经过无量纲处理的第*i*个样本第*j*个指标值;

y_j 、 σ_j — y_{ij} 的均值、均方差。

(2) 主成分分析

由标准化的数据进行分析(SPSS软件处理), 可得到主成分的成分矩阵及解释的总方差, 用成分矩阵的系数除以相应特征值的平方根得到主成分的得分系数, 各个主成分得分系数的加权平均即为指标权重 w_1 ^[2]。

2.2 改进的AHP与熵权法

在对一个样本河流进行评价工作中, 建立了三层河流健康评价指标体系(见表1), 目标层A为河流健康综合指数, 准则层B分为自然环境子系统B1, 其指标层含有6个指标(D1~D6), 社会服务子系统B2, 其指标层含有5个指标(D7~D11)。

在指标权重分配计算中, 指标层 D 对准则层 B 的权重 w_a 采用熵值法进行计算, 准则层 B 对目标层 A 的权重 w_b 采用 AHP 进行计算。则指标层 D 对目标层 A 的总层次权重 w_2 为:

$$w_2 = w_a \cdot w_b \quad (3)$$

表 1 河流健康评价体系层次划分

目标层 (A)	准则层 (B)	指标层 (D)
河流 健康 综合 指数 (A)	自然环境 子系统 (B1)	河岸稳定性指数 (D1)
		河流流动性指数 (D2)
		生态流量满足程度指数 (D3)
		水质综合指数 (D4)
		岸坡植被结构完整性指数 (D5)
		河流生物多样性指数 (D6)
	社会服务 子系统 (B2)	防洪工程达标率 (D7)
		岸线利用管理指数 (D8)
		公众满意度 (D9)
		水功能区水质达标率 (D10)
		供水水量保证率 (D11)

2.3 基于博弈论的组合赋权

在进行组合赋权前, 需要对两种方法所得权重进行一致性检验。Kendall-W 协和系数是检验 m 评价方法对 n 个对象的评判结果之间是否一致的常用方法, 协和系数 W 是表示不同方法之间的一致性程度, 其值在 0 ~ 1 之间, 越接近 1 表示不同方法之间一致性越高。其计算公式为:

$$W = \frac{12 \sum_{i=1}^n R_i^2 - 3m^2n(n+1)^2}{m^2n(n^2-1)} \quad (4)$$

其中, R_i 为第 i 种评价方法各评价对象权重排序之和。统计量 $m(n-1)W$ 在大样本情况下近似服从卡方分布 $\chi^2 = (n-1)$, 因此可利用卡方分布进行统计量 W 的假设检验。

若两种方法得到的权重具有一致性, 则表明两种权重分配在每个指标上是一致的, 此时可通过简单的加权平均组合得到最终权重。但如果一致性不好, 则表明两种权重分配具有较大差异, 需要一种合适的方法进行组合赋权。博弈论思想就是在不同方法之间寻找妥协, 使组合权重与各个

权重之间的偏差最小化, 让不一致趋于妥协统一。基于博弈论思想的组合赋权就是要极小化不同权重分配方法之间的差异, 尽可能保留各种方法的优势。因此, 在不满足一致性情况下, 可采用博弈论思想集成赋权, 具体步骤如下:

假设有 m 种赋权方法对评价指标进行权重分配, 则任意一种线性组合赋权可表示为:

$$w = \sum_{k=1}^m \beta_k w_k \quad (\beta_k \geq 0) \quad (5)$$

式中:

w —组合赋权权重向量;

β_k —线性组合赋权中第 k 种权重向量 w_k 系数。

为了得到一个满意的组合赋权结果, 达到其与各个权重分配方法之间的差异极小化, 需要优化各个线性组合系数 β_k , 使 w 与个 w_k 的离差最小, 即有:

$$\min \left\| \sum_{k=1}^m \beta_k w_k - w_i \right\|_2 \quad (i=1, 2, \dots, m) \quad (6)$$

其中, w_i 为第 i 种权重分配方法得到权重分配向量。

由矩阵微分性质, 可知式 (6) 成立的一阶导数

条件为: $\sum_{k=1}^m \beta_k w_i w_k^T = w_i w_i^T$, 对应方程组:

$$\begin{bmatrix} w_1 w_1^T & w_1 w_2^T & \cdots & w_1 w_m^T \\ w_2 w_1^T & w_2 w_2^T & \cdots & w_2 w_m^T \\ \vdots & \vdots & \cdots & \vdots \\ w_m w_1^T & w_m w_2^T & \cdots & w_m w_m^T \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} = \begin{bmatrix} w_1 w_1^T \\ w_2 w_2^T \\ \vdots \\ w_m w_m^T \end{bmatrix} \quad (7)$$

可解式 (7) 得到解向量 $(\beta_1, \beta_2, \dots, \beta_m)^T$, 再进行归一化处理, 并代入式 (5) 即可得到基于博弈论思想的集合权重 w 。

3 计算结果

对样本河流健康建立的评价指标体系 (表 1), 利用 SPSS 软件进行主成分分析计算, 得到权重分配向量 w_1 ; 再运用改进的 AHP 与熵值法得到权重向量 w_2 。结果见表 2。

按式 (4) 进行两种权重向量的一致性检验, 得出 Kendall 协和系数 $W=0.2836$, $\chi^2=5.672$, 给定显著性水平 $\alpha=0.05$, 查卡方分布表的 $\chi_{0.05}^2(10)=18.307 > 5.672$, 因此两种权重分配方法不是一致的, 不能进行简单的加权平均组合赋权, 可以采

用基于博弈论思想的集合赋权方法。

表 2 两种权重分配方法结果

指标层 (D)	w_1	w_2
河岸稳定性指数 (D1)	0.1010	0.0419
河流流动性指数 (D2)	0.1013	0.0419
生态流量满足程度指数 (D3)	0.1016	0.1048
水质综合指数 (D4)	0.0947	0.1482
岸坡植被结构完整性指数 (D5)	0.0936	0.0933
河流生物多样性指数 (D6)	0.0931	0.1153
防洪工程达标率 (D7)	0.1023	0.0421
岸线利用管理指数 (D8)	0.0873	0.0495
公众满意度 (D9)	0.0703	0.1014
水功能区水质达标率 (D10)	0.0982	0.1165
供水水量保证率 (D11)	0.0566	0.1451

将上述两种权重分配向量 w_1 与 w_2 代入式 (7), 由式 (7) 可以求得线性组合系数 $\beta_1=0.2392$, $\beta_2=0.8043$, 归一化得 $\beta_1=0.2292$, $\beta_2=0.0.7708$ 。求得最终集合权重向量 $w=(0.0555, 0.0555, 0.1040, 0.1359, 0.0934, 0.1102, 0.0559, 0.0582, 0.0943, 0.1123, 0.1248)$ 。可以看出 w 的每个指标权重都在 w_1 与 w_2 之间, 表明 w 是 w_1 与 w_2 一种均衡优化, 且是基于博弈论思想的偏差极小化。

4 结论

根据样本河流建立的评价指标体系, 从得出的最终集合权重分配来看, 11 个指标中权重最大的是水质综合指数 (D4), 达到 0.1359, 说明河流

水质情况在河流健康评价中至关重要, 此外指标权重达到 0.1 以上的还有生态流量满足程度指数 (D3)、河流生物多样性指数 (D6)、水功能区水质达标率 (D10)、供水水量保证率 (D11), 表明在这些方面河流的表现对河流整体的健康状况影响是比较大的。11 个指标权重中没有出现极端小的情况, 都在 0.05 以上, 表明每个指标都有其作用, 在河流整体健康表现中都是不可或缺。

河流健康评价工作中, 指标权重分配是非常重要的部分, 如何合理全面地去衡量各个指标的重要程度需要慎重考虑, 本文采用的基于主成分分析的权重分配方法是一种比较好权重计算方法, 当然在今后的生态评估工作中还有待进一步深入研究。

参考文献:

[1] 董哲仁. 河流健康的内涵 [J]. 中国水利, 2005 (4):1-4.

[2] 吴振, 邵东国, 顾文权. 基于主成分分析的区域初始水权分配 [J]. 南水北调与水利科技, 2016, 14 (4):1792.

[3] 山成菊, 董增川, 樊孔明, 等. 组合赋权法在河流健康评价权重计算中的应用 [J]. 河海大学学报: 自然科学版, 2012, 40 (6):622-628.

[4] 陈琼华. 综合评价中的赋权方法 [J]. 统计与决策, 2004, 174 (2):118-119.

[5] 胡永宏, 贺思辉. 综合评价方法 [M]. 北京: 科学出版社, 2000.

[6] 叶义成. 系统综合评价技术及其应用 [M]. 北京: 冶金工业出版社, 2006.

[7] 宋光兴, 杨德礼. 基于决策者偏好及赋权法一致性的组合赋权法 [J]. 系统工程与电子技术, 2004, 26 (9):1226-1230.

(责任编辑: 华智睿)