

基于熵权决策法验证的因子分析法在水质评价中的应用

宋影¹, 王国利¹, 李伟²

(1. 大连理工大学, 辽宁 大连 116024; 2. 辽宁省水利水电科学研究院, 沈阳 110003)

摘要: 因子分析法可以通过将反映水质状况的众多指标转化为几个关键的综合性指标, 消除指标间相关性的影响, 使得水质评价结果更为客观。采用因子分析法对辽河流域铁岭站 2006 年-2012 年的水质进行了综合评价, 结果显示影响河流水质的因子主要为溶解氧、化学需氧量、高锰酸盐指数、生化需氧量、挥发酚、氟化物、石油类, 其次为 pH 值、氨氮。在此基础上对其水质变化趋势进行了分析, 发现 2006 年-2011 年水质逐渐变好, 2012 年变差, 但变差趋势幅度较小, 不影响大致变化趋势。最后, 用熵权决策法检验了因子分析法的评价结果, 两种方法的评价结果基本吻合。

关键词: 水质评价; 因子分析法; 熵权决策法; 辽河; 特征值; 载荷; 权重

中图分类号: X824 **文献标志码:** A **文章编号:** 1672-1683(2015)06-1080-04

Application of factor analysis based on entropy weight verification in water quality

SONG Ying¹, WANG Guoli¹, LI Wei²

(1. Dalian University of Technology, Dalian 116024, China;

2. Institute of Water Resources and Hydropower Research in Liaoning Province, Shenyang 110003, China)

Abstract: Factor analysis method can convert a number of water quality indexes to a few comprehensive indexes and eliminate the correlation between indexes, so it can provide more objective water quality evaluation results. The factor analysis method was used to conduct the comprehensive evaluation of water quality in the Liao River Basin from 2006 to 2012. The results showed that the main impact factors of water quality are DO, COD, permanganate value, BOD, volatile phenol, fluoride, petroleum, followed by PH and ammonia nitrogen. On the basis, the variation trend of water quality in the Liao River Basin from 2006 to 2012 was obtained, which suggested that water quality has improved gradually from 2006 to 2011, while becomes bad in 2012, but the trend of variation range is small without impacts on the general trend. Finally, the entropy weight method was used to validate the evaluation results from the factor analysis, which indicated both assessment results are similar.

Key words: water quality evaluation; factor analysis; entropy weight method; Liao River; eigen value; load; index weight

目前,水质评价方法主要有灰色聚类法、综合指数法、人工神经网络法等^[1-2]。各种方法均有各自的优缺点,如灰色聚类法过程简单,但在函数的选择时存在较大的人为性^[3];综合指数法比较简便,易于理解,但受主观影响较大,缺少严密性^[2];人工神经网络法考虑因素比较全面,但对数据的要求比较高^[1]。因子分析法是将多项指标转化为少数几项综合指标的一种统计分析方法^[3-4],综合指标起主导作用,互不相关,可避免由于数据间存在相关性而导致的分析困难的问题。因子分析法可在损失极少信息的基础上达到有效简化评价指标的目的,抓住问题的主要矛盾,还可利用因子的得分分析水质污染状况以及主要污染物的种类。熵权决策法

是一种客观的赋权方法,可根据样本中各个评价指标构成的判断矩阵计算各评价指标的权重,克服了经验确定权重的主观性,从而使评价过程更加科学公正。本文采用因子分析法,对辽河流域铁岭站的水质状况进行了分析,并利用熵权决策法对评价结果进行了验证,结果发现因子分析法可以科学准确的分析河流水质。

1 研究方法

1.1 因子分析法

假设有 m 个样本, 每个样本含有 n 个变量(r_1, r_2, r_3, \dots ,

收稿日期: 2015-02-03 修回日期: 2015-10-23 网络出版时间: 2015-11-30

网络出版地址: <http://www.cnki.net/kcms/detail/13.1334.TV.20151130.2010.012.html>

基金项目: 国家水体污染控制与治理科技重大专项“重点流域环境流量保障与容量总量控制管理关键技术与应用示范”(2013ZX07501-04)

作者简介: 宋影(1989), 女, 山东临沂人, 主要从事水文水资源方面的研究。E-mail: 459188649@qq.com

通讯作者: 王国利(1964), 男, 辽宁辽中人, 教授, 博士, 主要从事水文水资源方面研究。E-mail: wanggl@dlut.edu.cn

r_n), 这样就构成一个 $m \times n$ 阶的原始数据矩阵如下。

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \cdots & \cdots & \cdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix} \quad (1)$$

本文采用 SPSS 13.0 软件进行数据处理。主要步骤^[59]为: 数据标准化, 得到标准化的数据矩阵, 再求相关系数矩阵; 然后根据相关系数矩阵求得特征值, 一般由方差累积贡献率确定主因子的个数, 建立因子载荷矩阵, 并对初始因子进行方差最大化旋转, 计算各个主因子的得分; 根据各主因子的得分, 计算综合得分。

1.2 熵权决策法

熵是系统无序程度的一种度量, 熵越小, 有效信息量越大; 熵越大, 有效信息量越小。利用熵权决策法来计算各评价指标的权重, 即是利用评价指标的有效信息量来计算, 有效信息量越大, 权重就越大。对于 m 个样本, n 个评价指标的数据系统, 熵权法的计算步骤^[10-16]如下:

步骤 1 构建原始数据矩阵, 同公式(1)。

步骤 2 原始数据矩阵标准化。

对于越大越优指标, r_{ij} 如下:

$$x_{ij} = (r_{ij} - r_{\min}) / (r_{\max} - r_{\min}) \quad (2)$$

对于越小越优指标, r_{ij} 如下:

$$x_{ij} = (r_{\max} - r_{ij}) / (r_{\max} - r_{\min}) \quad (3)$$

得到标准化矩阵为

$$X = (x_{ij})_{m \times n} \quad (4)$$

式中: x_{ij} 为标准化矩阵 X 中的元素; r_{\max} 为所有样本中同一指标的最大值; r_{\min} 为所有样本中同一指标的最小值。

步骤 3 确定熵值 H_j 。第 j 个指标的熵值定义为

表 1 相关系数矩阵

Tab. 1 The correlation coefficient matrix

污染指标	pH 值	溶解氧	氨氮	化学需氧量	高锰酸盐指数	生化需氧量	挥发酚	氟化物	石油类
pH 值	1.000								
溶解氧	0.058	1.000							
氨氮	-0.491	0.596	1.000						
化学需氧量	0.096	0.923	0.658	1.000					
高锰酸盐指数	0.152	0.879	0.673	0.974	1.000				
生化需氧量	0.114	0.894	0.693	0.978	0.999	1.000			
挥发酚	0.155	0.831	0.695	0.917	0.982	0.980	1.000		
氟化物	-0.227	0.709	0.818	0.756	0.751	0.764	0.746	1.000	
石油类	-0.056	0.826	0.591	0.727	0.790	0.805	0.835	0.577	1.000

(2) 提取主因子。

对因子载荷矩阵做方差最大化旋转得到主因子的方差

表 2 相关矩阵的特征根、贡献率和累积贡献率

Tab. 2 Eigen values, contribution rate, and cumulative contribution rate of correlation coefficient matrix

主因子	旋转前			旋转后		
	特征根	方差贡献率(%)	累积贡献率(%)	特征根	方差贡献率(%)	累积贡献率(%)
1	6.632	73.684	73.684	6.472	71.917	71.917
2	1.439	15.992	89.676	1.598	17.760	89.676

根据表 2, 第一主因子的特征值为 6.632, 方差贡献率为 73.684%, 经旋转后方差贡献率为 71.917%, 第二主因子的特征值为 1.439, 方差贡献率为 15.992%, 经旋转后方差贡献率为 17.760%, 累积贡献率 89.676% (大于 85%), 说明可

$$H_j = - \left(\sum_{i=1}^m y_{ij} \ln y_{ij} \right) / \ln m \quad (5)$$

式中: $y_{ij} = (1 + x_{ij}) / \left(\sum_{i=1}^m (1 + x_{ij}) \right)$ (为使 $\ln y_{ij}$ 有意义, 因此对 y_{ij} 进行了修正)

步骤 4 确定权重 W_j 。

$$W_j = (1 - H_j) / \left(n - \sum_{i=1}^n H_j \right) \quad (6)$$

步骤 5 计算水质良好指数综合得分。

$$WQ = \sum_{i=1}^n (W_j x_{ij}) \quad (7)$$

2 实例研究

2.1 指标的确定

河流的水质状况是由多个水质指标组成, 结合辽河(铁岭站段)的地表水环境质量特点, 选取 pH 值、溶解氧、氨氮、化学需氧量、高锰酸盐指数、生化需氧量、挥发酚、氟化物、石油类等 9 项指标评价了辽河流域铁岭站 2006 年-2012 年的水质状况, 其中 8 项评价指标属于越大污染越严重的指标, 只有溶解氧属于越大水质质量越好的评价指标。

根据辽河流域铁岭站的特点, 采样时间设定为 2006 年 1 月-2012 年 12 月, 每年设定 12 个采样时间, 每月一次。对每年的 12 次监测数据进行平均, 得到年平均监测值。根据 2006 年-2012 年水质评价指标的年平均值建立初始矩阵 R 。

2.2 因子分析

(1) 数据标准化和相关系数矩阵。

首先对逆向评价指标溶解氧进行求倒数处理, 然后利用 SPSS13.0 对矩阵中的元素进行标准化处理, 得到标准化矩阵, 并计算出相关系数矩阵, 见表 1。

贡献率, 见表 2, 然后按特征根大于 1 和主因子累积贡献率大于 85% 的原则选择主因子进行分析。

以利用 2 个主因子来反映原始评价指标的信息。

由方差最大化旋转得到旋转后的载荷矩阵, 见表 3, 可看出各主因子主要承载哪些评价指标的信息。如果评价指标与某一个因子的载荷系数绝对值越大, 则评价指标与该因子

的关系越近。由表 3 可以看出,第一主因子在评价指标溶解氧、化学需氧量、高锰酸盐指数、生化需氧量、挥发酚、氟化物、石油类上具有比较大的载荷,旋转后的主因子载荷值分别为:溶解氧为 0.925,化学需氧量为 0.967,高锰酸盐指数 0.990,生化需氧量 0.992,挥发酚 0.977,氟化物 0.770,石油类 0.836;第二主因子在 pH、氨氮上具有比较的载荷,旋转后的主因子载荷值分别为:pH 为- 0.955,氨氮 0.681。

表 3 旋转后因子载荷矩阵

Tab. 3 Rotated component load matrix

评价指标	主因子	
	1	2
pH 值	0.158	- 0.955
溶解氧	0.925	0.051
氨氮	0.668	0.681
化学需氧量	0.967	0.038
高锰酸盐指数	0.990	0.001
生化需氧量	0.992	0.037
挥发酚	0.977	0.015
氟化物	0.770	0.446
石油类	0.836	0.132

(3) 主因子得分计算和综合评价结果。

计算主因子得分系数矩阵 P , 然后采用式 (8) 计算各年在各个主因子上的得分(见表 4)。

$$F_{ik} = \sum_{j=1}^9 (x'_{ij} \times p_{jk}) (i=1, 2, \dots, 7; k=1, 2; j=1, 2, \dots, 9) \quad (8)$$

式中: F_{ik} 为第 i 年第 k 个主因子的得分; x'_{ij} 为标准化后第 i 年第 j 个评价指标的数值; p_{jk} 为主因子得分系数矩阵 P 中的元素。

再利用式(9)计算各年的综合得分(见表 4)。

$$F_i = \sum_{k=1}^2 (F_{ik} \times \lambda_k) / \sum_{k=1}^2 (\lambda_k) (i=1, 2, \dots, 7; k=1, 2) \quad (9)$$

式中: F_i 为第 i 年的综合得分; λ_k 为第 k 个主因子的特征根。

综合得分结果(表 4)反映了辽河(铁岭站段)的污染状况,主因子得分的排序反映了各年污染性质。可以看出:2006 年综合得分最高, 该年污染状况最严重, 其次是 2007 年、2008 年、2009 年、2010 年、2012 年、2011 年; 水质污染从 2006 年至 2011 年逐渐减轻, 2012 年加重, 但加重不是很大, 不影响大致趋向。2006 年的第一主因子得分最高, 说明该年溶解氧、化学需氧量、高锰酸盐指数、生化需氧量、挥发酚、氟化物、石油类造成的污染最大, 而 2009 年的第二主因子得分最高, 说明该年 pH、氨氮造成的污染最大。

表 4 2006 年- 2012 年水质综合得分

Tab. 4 Comprehensive scores of water quality from 2006 to 2012

年份	第一主因子		第二主因子		综合	
	得分	排序	得分	排序	得分	排序
2006	1.168	1	0.151	4	0.986	1
2007	1.129	2	0.256	3	0.973	2
2008	0.892	3	- 0.310	5	0.677	3
2009	- 0.890	7	1.667	1	- 0.434	4
2010	- 0.798	5	0.565	2	- 0.555	5
2011	- 0.806	6	- 1.161	6	- 0.869	7
2012	- 0.694	4	- 1.168	7	- 0.778	6

2.3 熵权决策法

利用式(2)至式(4)对初始矩阵 R 进行归一化处理, 其中溶解氧属于越大越优指标, pH 值、氨氮、化学需氧量、高锰酸盐指数、生化需氧量、挥发酚、氟化物、石油类属于越小越优指标。数据经过处理后, 利用熵权决策法的式(5)至式(6)计算出各个评价指标的熵值和权重(表 5), 进而利用式(7)计算出水质良好指数历年综合评分, 见表 6。从表 6 可以看出, 2006 年水质最差, 2011 年最好, 2012 年次之, 水质良好状况排序为 2011 年、2012 年、2010 年、2009 年、2008 年、2007 年、2006 年。水质从 2006 年至 2011 年逐渐变好, 2012 年变差, 但变化幅度不大, 不影响大致趋势。

表 5 各因素熵值和权重

Tab. 5 The entropy and weight values of each factor

因素	pH	溶解氧	氨氮	化学需氧量	高锰酸盐指数	生化需氧量	挥发酚	氟化物	石油类
熵 H_j	0.988	0.983	0.978	0.985	0.980	0.980	0.980	0.986	0.989
权重	0.078	0.111	0.145	0.097	0.133	0.133	0.135	0.093	0.070

表 6 辽河(铁岭站段)水质良好指数历年综合得分

Tab. 6 Comprehensive scores of water quality goodness index in the Liao River Basin (Tieling section) in past years

年份	2006	2007	2008	2009	2010	2011	2012
综合评分	0.109	0.153	0.193	0.777	0.821	0.890	0.824
排序	7	6	5	4	3	1	2

2.4 合理性分析

为了验证因子分析法的有效性, 利用熵权决策法对河流水质进行评价, 与因子分析法评价结果进行对比, 结果表明, 因子分析法与熵权决策法得到的评价结果一致。2008 年以后, 辽河流域一直在开展水质水量联合调度, 旨在保障不影响原有各项供水指标的前提下改善河流水质, 同时相关执法部门进一步加强了环境管理和检查力度, 达到了在控制污染源的基础上改善河流水质的目的。因子分析法的评价计算结果符合上述流域水资源管理模式的实施。

3 结论

河流水质评价系统是一个由多个因子构成的复杂系统, 评价指标多, 利用因子分析法可以简化评价指标并且损失较少信息, 抓住主要矛盾, 根据各主因子的得分对水质污染状况以及重要污染物的种类进行分析。本研究结果表明, 通过因子分析法得到的评价结果与熵权决策法的评价结果吻合, 因此因子分析法能够从整体上准确分析辽河(铁岭站段)的水质状况, 为水环境管理决策提供技术支持。

参考文献(References):

[1] 赵晓慎, 张超, 王文川. 基于熵权赋权的贝叶斯水质评价模型[J]. 水电能源科学, 2011, 29(6): 33-35. (ZHAO Xiaoshen, ZHANG Chao, WANG Wenchuan. Water quality evaluation model using bayesian method on entropy weight[J]. Water Resources and Power, 2011, 29(6): 33-35. (in Chinese))

- [2] 杨磊磊, 卢文喜, 黄鹤, 等. 改进内梅罗污染指数法和模糊综合法在水质评价中的应用[J]. 水电能源科学, 2012, 30(6): 41-44. (YANG Leilei, LU Wenxi, HUANG He, et al. Application of improved nemerow pollution exponential method and fuzzy comprehensive evaluation method used in water quality assessment[J]. Water Resources and Power, 2012, 30(6): 41-44. (in Chinese))
- [3] 郝昊, 王晓昌, 张琼华, 等. 宜兴城市内河污染物时空分布及解析[J]. 环境工程学报, 2015, 9(1): 125-130. (HAO Hao, WANG Xiaochang, ZHANG Qionghua, et al. Spatiotemporal distribution and analysis for pollutants in inland Rivers of Yixing, China[J]. Chinese Journal of Environmental Engineering, 2015, 9(1): 125-130. (in Chinese))
- [4] 陈一, 唐飞, 李铁刚, 等. 因子分析法在质谱成像数据分析中的应用[J]. 分析化学研究报告, 2014, 42(8): 1099-1103. (CHEN Yi, TANG Fei, LI Tiegang, et al. Application of factor analysis in imaging mass spectrometric data analysis[J]. Chinese Journal of Analytical Chemistry, 2014, 42(8): 1099-1103. (in Chinese))
- [5] 张德同. 因子分析法在白城市水质评价中的应用[J]. 东北水利水电, 2014(12): 39-40. (ZHANG Detong. Application of factor analysis method in water quality evaluation of Baicheng City[J]. Water Resources & Hydropower of Northeast China, 2014(12): 39-40. (in Chinese))
- [6] 伊燕平, 卢文喜, 辛欣, 等. 因子分析法在金泉工业区地下水水质评价中的应用[J]. 中国环境监测, 2012, 22(1): 10-12. (YI Yanping, LU Wenxi, XI Xin, et al. Application of factor analysis method on the groundwater quality evaluation in Jinquan Industrial Area[J]. Environmental Monitoring of China, 2012, 22(1): 10-12. (in Chinese))
- [7] 张斓子, 陆必应, 周智敏, 等. 基于因子分析法和图像对比度的穿墙雷达杂波抑制[J]. 电子与信息学报, 2013, 35(11): 2686-2692. (ZHANG Lanzi, LU Biying, ZHOU Zhimin, et al. The clutter suppression based on factor analysis and image contrast in through the wall application[J]. Journal of Electronics & Information Technology, 2013, 35(11): 2686-2692. (in Chinese))
- [8] 宁连举, 李萌. 基于因子分析法构建大中型工业企业技术创新能力评价模型[J]. 科研管理, 2011, 32(3): 351-358. (NING Lianju, LI Meng. Evaluation model for large and medium sized industrial enterprises, technological innovation capability based on factor analysis method[J]. Science Research Management, 2011, 32(3): 351-358. (in Chinese))
- [9] 王维国, 冯云. 基于因子分析法的中国城市人居环境现状综合评价及影响因素分析[J]. 生态经济, 2011(5): 174-177. (WANG Weigu, FENG Yun. A comparative research on dalian, s human living environment with the other 36 domestic cities[J]. Ecological Economy, 2011(5): 174-177. (in Chinese))
- [10] 张鹤志, 马传明, 王江思. 基于层次分析法-熵权法的中原城市群生态环境评价[J]. 安全与环境工程, 2014, 21(1): 87-92. (ZHANG Huzhi, MA Chuaming, WANG Jiangsi. Environment condition assessment of urban agglomeration in the central plain based on AHP entropy method[J]. Safety and Environmental Engineering, 2014, 21(1): 87-92. (in Chinese))
- [11] 孙培学, 赵坤鹏. 基于熵权法的组合模型在大坝渗流预测中的应用[J]. 水电能源科学, 2013, 31(12): 70-73. (SU N Peixue, ZHAO Kunpeng. Application of combination model based on entropy method in dam seepage forecasting[J]. Water Resources and Power, 2013, 31(12): 70-73. (in Chinese))
- [12] 陈南祥, 苏荣, 曹文庚. 基于熵权的集对分析法在土默特左旗地下水水质评价中的应用[J]. 干旱区资源与环境, 2013, 27(6): 30-34. (CHEN Nanxiang, SU Rong, CAO Wengeng. Application of the set pair analysis method to evaluation of shallow groundwater quality based on entropy weight[J]. Journal of Arid Land Resources and Environment, 2013, 27(6): 30-34. (in Chinese))
- [13] 梁桂兰, 徐卫亚, 谈小龙. 基于熵权的可拓理论在岩体质量评价中的应用[J]. 岩土力学, 2010, 31(2): 535-540. (LIANG Guilan, XU Weiya, TAN Xiaolong. Application of extension theory based on entropy weight to rock quality evaluation[J]. Rock and Soil Mechanics, 2010, 31(2): 535-540. (in Chinese))
- [14] 孙鸿鹤, 程先富, 倪玲, 等. 基于云模型和熵权法的巢湖流域防洪减灾能力评估[J]. 灾害学, 2015, 30(1): 222-227. (SUN Honghu, CHENG Xianfu, NI Ling, et al. Capacity evaluation of flood disaster prevention and reduction in Chaohu basin based on cloud model and entropy weight method[J]. Journal of Catastrophology, 2015, 30(1): 222-227. (in Chinese))
- [15] 龚艳冰. 基于正态云模型和熵权的河西走廊城市化生态风险综合评价[J]. 干旱区资源与环境, 2013, 26(5): 169-174. (GONG Yanbing. Comprehensive Assessment on ecological risk of Hexi corridor urbanization based on normal cloud model and entropy weight[J]. Journal of Arid Land Resources and Environment, 2013, 26(5): 169-174. (in Chinese))
- [16] 龚艳冰, 张继国. 基于正态云模型和熵权的人口发展现代化程度综合评价[J]. 中国人口·资源与环境, 2012, 22(1): 138-143. (GONG Yanbing, ZHANG Jiguo. Comprehensive assessment on population development degree of modernization based on normal cloud model and entropy weight[J]. China Population·Resources and Environment, 2012, 22(1): 138-143. (in Chinese))