

基于熵权的未确知测度模型在湖库富营养化评价中的应用

陈娟,刘凌

(河海大学水文水资源与水利工程科学国家重点实验室,江苏 南京 210098)

摘要 利用基于熵权的未确知测度模型对 8 个湖库的富营养化状况进行了评价.为了有效解决权重计算困难的问题,根据各指标值的差异程度,利用熵权确定了各指标的权重.评价对比结果表明,基于熵权的未确知测度模型客观地考虑了指标权重问题且原理简单、思路清晰,利用该模型可以对湖库富营养化进行合理评价,但要注意置信度的选取范围.

关键词 熵权 未确知测度模型 湖泊 水库 富营养化评价

中图分类号 X142 **文献标识码** A **文章编号** 1000-1980(2008)04-0452-04

近几十年来,我国湖库富营养化问题日益严重.调查显示^[1],有 51.1% 的湖泊和 30.8% 的水库达到了富营养化程度,并对工农业生产和人们的生活造成了极大影响.湖库富营养化问题已成为当今水污染治理的一大难题.科学合理地进行湖库富营养化程度评价是湖库富营养化控制和治理过程中的一项重要基础工作.目前,我国常用的湖库富营养化评价方法有卡尔森富营养化指标法^[2]、营养度指数法^[3]、模糊评价法^[4]、物元分析法^[5]、灰色聚类法^[6]和人工神经网络模型^[7-8]等.由于湖库富营养化评价所依据的各项指标间的相关性非常复杂,评价具有模糊性、随机性及灰色性等不确定性特征,各种评价方法均有其适用条件和局限性,缺乏统一的评价模型.

本文将基于熵权的未确知测度模型应用于湖库富营养化评价,并根据各指标值的差异程度,利用熵权来确定各指标的权重.

1 模型简介

设 $x_1, x_2, \dots, x_n$ 是 n 个样本,则研究对象空间 $X = \{x_1, x_2, \dots, x_n\}$. 要评价 x_i , 需要测量 m 个指标 $I_1, I_2, \dots, I_m$, 则指标空间 $I = \{I_1, I_2, \dots, I_m\}$. x_{ij} 表示 x_i 关于 I_j 的监测值. 对于 x_{ij} , 有 k 个评价等级 $c_1, c_2, \dots, c_k$, 则评价空间 $U = \{c_1, c_2, \dots, c_k\}$. 若 $c_1 > c_2 > \dots > c_k$, 则称 $\{c_1, c_2, \dots, c_k\}$ 是评价空间 U 的一个有序分割类. 如果每个指标的分类标准已知,则可写成分类标准矩阵

$$\begin{bmatrix} c_1 & \cdots & c_k \\ a_{11} & \cdots & a_{1k} & I_1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mk} & I_m \end{bmatrix}$$

其中 a_{jk} 满足 $a_{j1} < a_{j2} < \dots < a_{jk}$ 或 $a_{j1} > a_{j2} > \dots > a_{jk}$.

1.1 单指标未确知测度

计算第 i 个样本的第 j 个指标值 x_{ij} 属于 c_k 类的未确知测度 $\mu_{ijk} = \mu(x_{ij} \in c_k)$. 这里先假定 $a_{j1} < a_{j2} < \dots < a_{jk}$.

当 $x_{ij} \leq a_{j1}$ 时,取

$$\mu_{ij1} = 1 \quad \mu_{ij2} = \cdots = \mu_{ijk} = 0 \quad (1)$$

当 $x_{ij} \geq a_{jk}$ 时,取

$$\mu_{ijk} = 1 \qquad \mu_{ij1} = \cdots = \mu_{ijk-1} = 0 \tag{2}$$

当 $a_{jl} \leq x_{ij} \leq a_{jl+1}$ 时,取

$$\mu_{ijl} = \frac{a_{jl+1} - x_{ij}}{a_{jl+1} - a_{jl}} \qquad \mu_{ijl+1} = \frac{x_{ij} - a_{jl}}{a_{jl+1} - a_{jl}} \qquad \mu_{ijk} = 0 (\text{当 } k < l \text{ 或 } k > l + 1 \text{ 时}) \tag{3}$$

1.2 指标权重的确定

指标权重是指各评价因子对方案分类的影响程度,它的确定是综合评价的一项重要内容.在信息论中,熵既可用来度量系统无序程度,还可用来度量数据本身所提供信息的有效性^[9].由于已经求出单个指标的未确知测度,即各项指标方案隶属各个等级的程度已确定,意味着该指标对方案分类的重要程度已确定,故拟采用信息熵的理论定出各评价因子的权重.

定义指标熵值

$$V_j = 1 + \frac{1}{\log_2 K} \sum_{k=1}^K \mu_{ijk} \log_2 \mu_{ijk} \tag{4}$$

式中: V_j ——指标 j 所提供的信息量; K ——评价级别数目; μ_{ijk} ——单指标未确知测度.
各指标权重为

$$w_j = \frac{V_j}{\sum_{i=1}^m V_i} \tag{5}$$

式中 m 为参与评价的指标数目.

从式(5)可以看出,权重 w 与指标所提供的信息量成正比.

1.3 多指标综合测度

用 $\mu_{ik} = \mu(x_i \in c_k)$ 表示样本 x_i 属于第 k 个评价等级 c_k 的程度,则

$$\mu_{ik} = \mu(x_i \in c_k) = \sum_{j=1}^m w_j \mu_{ijk} \qquad (1 \leq i \leq n, 1 \leq k \leq K) \tag{6}$$

1.4 识别准则

由于评价等级划分是有序的,最大隶属度识别准则不适用,因此改用置信度识别准则.取置信度 λ ($1 > \lambda > 0.5$),当 $c_1 > c_2 > \cdots > c_k$ 时,令

$$K_i = \min \left\{ k : \sum_{l=1}^k \mu_{il} \geq \lambda, 1 \leq k \leq K \right\} \tag{7}$$

则 x_i 属于第 K_i 个评价等级 c_{K_i} .

可按照评分准则计算各样本得分

$$q_{x_i} = \sum_{l=1}^K n_l \mu_{il} \tag{8}$$

式中用 n_l 表示 c_k 的分值,当 $c_1 > c_2 > \cdots > c_k$ 时, n_l 为单调减小的正整数^[10].

2 模型应用

湖库营养状态受多种因素影响,各参数间有不同程度的相关关系.关于湖库的富营养化评价,目前我国还没有国家标准.卡森营养指数常被用来研究、表征湖库的富营养化程度.卡森指数(TSI(j))的计算公式为

$$TSI(j) = 10(a_j + b_j \ln \rho_j / \ln 2.5) \tag{9}$$

式中:TSI(j)——第 j 种参数的营养状态指数; a_j, b_j ——相应于第 j 种参数的 2 个常数; ρ_j —— j 种参数的质量浓度值.

本文以我国 5 个湖泊和 3 个水库作为样本,以 $\rho(\text{Chla}), \rho(\text{TP}), \rho(\text{TN})$ 和 $1/D_s$ (D_s 为透明度)4 个参数作为富营养化的评价指标,它们相应于贫、中、富 3 种营养类型的标准值取为相应于我国湖泊水质营养状态指数 TSI 为 0,50 和 100 时的各参数值,具体数据列于表 1. 样本数据来自文献^[11],列于表 2.

2.1 单指标未确知测度矩阵

以淀山湖为例:

$$\rho(\text{Chla}) \quad \mu_{111} = \frac{(0.1 - 0.0372)}{(0.1 - 0.001)} = 0.6343 \quad \mu_{112} = \frac{(0.0372 - 0.001)}{(0.1 - 0.001)} = 0.3657 \quad \mu_{113} = 0$$

表 1 评价指标的 3 类标准值

Table 1 Three standard values of evaluation indexes				
营养级别	$I_1: \rho(\text{Chla}) / (10^{-2} \text{mg} \cdot \text{L}^{-1})$	$I_2: \rho(\text{TP}) / (10 \text{mg} \cdot \text{L}^{-1})$	$I_3: \rho(\text{TN}) / (\text{mg} \cdot \text{L}^{-1})$	$I_4: D_S^{-1} / \text{m}^{-1}$
c_1 贫营养级 I	0.001	0.03	0.04	0.072
c_2 中营养级 II	0.100	0.65	0.77	0.95
c_3 富营养级 III	10.000	14.15	14.64	12.50

$$\rho(\text{TP}) \quad \mu_{121} = \frac{(0.65 - 0.29)}{(0.65 - 0.03)} = 0.5806$$
$$\mu_{122} = \frac{(0.29 - 0.03)}{(0.65 - 0.03)} = 0.4194$$
$$\mu_{123} = 0$$

$$\rho(\text{TN}) \quad \mu_{131} = 0 \quad \mu_{132} = \frac{(14.64 - 1.086)}{(14.64 - 0.77)} = 0.9772 \quad \mu_{133} = \frac{(1.086 - 0.77)}{(14.64 - 0.77)} = 0.0228$$

$$1/D_S \quad \mu_{141} = 0 \quad \mu_{142} = \frac{(12.5 - 1.488)}{(12.5 - 0.95)} = 0.9534 \quad \mu_{143} = \frac{(1.488 - 0.95)}{(14.64 - 0.77)} = 0.0466$$

依此类推,可得出矩阵

$$\begin{pmatrix} \mu_{1jk} \end{pmatrix}_{4 \times 3} \quad \begin{pmatrix} \mu_{2jk} \end{pmatrix}_{4 \times 3} \quad \begin{pmatrix} \mu_{3jk} \end{pmatrix}_{4 \times 3} \quad \begin{pmatrix} \mu_{4jk} \end{pmatrix}_{4 \times 3} \quad \begin{pmatrix} \mu_{5jk} \end{pmatrix}_{4 \times 3} \quad \begin{pmatrix} \mu_{6jk} \end{pmatrix}_{4 \times 3} \quad \begin{pmatrix} \mu_{7jk} \end{pmatrix}_{4 \times 3} \quad \begin{pmatrix} \mu_{8jk} \end{pmatrix}_{4 \times 3}$$

2.2 熵权重矩阵的确定

由式(4),(5)可得各样本的熵权重矩阵为

$$\begin{aligned} w_1 &= (0.1601, 0.1516, 0.3586, 0.3297) & w_2 &= (0.1706, 0.2600, 0.3967, 0.1727) \\ w_3 &= (0.2469, 0.2433, 0.3334, 0.1764) & w_4 &= (0.2368, 0.3945, 0.1974, 0.1713) \\ w_5 &= (0.3635, 0.1492, 0.3180, 0.1693) & w_6 &= (0.3446, 0.2405, 0.2063, 0.2085) \\ w_7 &= (0.3566, 0.2335, 0.2018, 0.2080) & w_8 &= (0.3171, 0.2424, 0.2179, 0.2226) \end{aligned}$$

2.3 多指标综合测度

根据式(6)可得多指标综合测度为

$$\begin{aligned} \mu_1 &= (0.19, 0.79, 0.02) & \mu_2 &= (0.16, 0.77, 0.06) & \mu_3 &= (0.76, 0.24, 0.00) \\ \mu_4 &= (0.00, 0.47, 0.53) & \mu_5 &= (0.14, 0.85, 0.01) & \mu_6 &= (0.57, 0.43, 0.00) \\ \mu_7 &= (0.52, 0.48, 0.00) & \mu_8 &= (0.00, 0.74, 0.26) \end{aligned}$$

2.4 识别和排序

由置信度识别准则,取置信度 $\lambda = 0.85$. $\mu_1 \cap 0.19 + 0.79 = 0.98 > 0.85$,根据式(7)可得 x_1 为中营养级 II,类似可得出 x_2, x_3, x_5, x_6, x_7 为中营养级 II, x_4, x_8 为富营养级 III.

取 $n_l = 8 - 2l$ 代入式(8),得

$$q = (4.33, 4.20, 5.52, 2.93, 4.27, 5.15, 5.04, 3.49)$$

3 结果分析

将基于熵权的未确知测度模型评价结果与 B-P 法评价结果、水质实况列于表 3.

表 3 基于熵权的未确知测度模型评价结果与 B-P 法评价结果、水质实况的比较

Table 3 Comparison of evaluation results from the unascertained measure model based on entropy weight with those from a B-P model and actual water quality					
评价对象	湖库名称	分值	基于熵权的未确知测度模型评价结果	B-P 法评价结果	水质实况
x_1	淀山湖	4.33	II	II	II
x_2	镜泊湖	4.20	II	III	II
x_3	密云水库	5.52	II	II	II
x_4	玄武湖	2.93	III	III	III
x_5	于桥水库	4.27	II	II	II
x_6	高州水库	5.15	II	II	II
x_7	紫碧湖	5.04	II	II	I
x_8	东山湖	3.49	III	III	III

由表 3 可知,基于熵权的未确知测度模型评价结果与 B-P 法评价结果、水质实况基本相符.评价结果不

同的是镜泊湖和紫碧湖. 镜泊湖测度模型评价结果与水质实况一致,均为中营养;B-P 法评价结果为富营养. 从原始数据和测度模型计算过程可以看出,镜泊湖 $\rho(\text{Chla})$ 和 $1/D_s$ 监测值均低于 II 级标准值,权重分别为 17.06% 和 17.27%, $\rho(\text{TP})$ 和 $\rho(\text{TN})$ 监测值虽然高于 II 级标准值且权重较大(26.00% 和 39.67%),但两者属于 II 级标准值的未确知测度比较大,分别为 0.8141 和 0.9640,最后综合测度为 II 级是合理的. 紫碧湖测度模型评价结果与 B-P 法评价结果一致,均为中营养. 水质实况为贫营养. 从表 2 数据可知,紫碧湖 4 项指标值均大于 I 级标准值而小于 II 级标准值. 4 项指标中只有 $\rho(\text{Chla})$ 指标属于 I 级标准值的未确知测度大,权重为 35.66%,其他 3 项指标均属于 II 级标准值的未确知测度大,总权重占了 64.34%. 由于水质实况也是根据一定方法计算出来的,因此紫碧湖采用测度模型所得评价结果——II 级是可以接受的.

4 结 语

将基于熵权的未确知测度模型应用于湖库富营养化评价是有效可行的. 这种评价方法利用熵权来计算权重,可以有效地解决指标权重计算困难的问题,并且可以得到合理的评价结果. 但必须指出,该方法没有考虑主观因素的影响,并且置信度是在一定的范围内选取的,如果置信度超出选取范围,则可能会导致不同的评价结果.

参考文献:

[1] 陈守煜. 工程模糊集理论与应用[M]. 北京: 国防工业出版社,1998.
[2] 郑曦,刘登义. 镜湖富营养化污染及其治理研究 II 镜湖水体污染现状评述[J]. 安徽师范大学学报:自然科学版,2000,23(3): 269-272.
[3] 张思冲,张雪萍,廖永丰. 营养度指数法在寒地湖泊富营养化评价中的应用[J]. 哈尔滨工业大学学报,2003,35(4): 416-418.
[4] 胡著邦,徐建民,全为民. 模糊评价法在湖泊富营养化评价中的应用[J]. 农业环境保护,2002,21(6): 535-536, 539.
[5] 饶清山,张江山. 熵权物元分析模型在湖泊富营养化评价中的应用[J]. 环境工程,2006,24(6): 80-82.
[6] 许文杰,陈为国. 应用灰色聚类法评价湖泊水体富营养化[J]. 山东建筑大学学报,2007,22(1): 49-51, 56.
[7] 任黎,董增川,李少华. 人工神经网络评价模型在太湖富营养化评价中的应用[J]. 河海大学学报:自然科学版,2004,32(2): 147-150.
[8] 韩涛,李怀恩,彭文启. 基于 MATLAB 的神经网络在湖泊富营养化评价中的应用[J]. 水资源保护,2005,21(1): 24-26.
[9] 孟庆生. 信息论[M]. 西安: 西安交通大学出版社,1989: 19-36.
[10] 刘开第,庞彦军,张博文. 水环境质量评价的未确知测度模型[J]. 环境工程,2000,18(2): 58-60.
[11] 李祚泳,丁晶,彭荔红. 环境质量评价原理与方法[M]. 北京: 化学工业出版社,2004.

Application of unascertained measure model based on entropy weight to evaluation of eutrophication of lakes and reservoirs

CHEN Juan, LIU Ling

(State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering Hohai University, Nanjing 210098, China)

Abstract: An unascertained measure model based on entropy weight was set up. In order to effectively solve the problem of weight calculation, entropy weight was used to calculate index weight based on the differences among evaluation indexes. The model was applied to the evaluation of eutrophication of 5 lakes and 3 reservoirs, and the results were compared with those of a B-P model and actual water quality. It is shown that the model objectively considers the problem of index weight, and that the principle is simple. It is also concluded that this model is feasible and effective in evaluating the eutrophication of lakes and reservoirs, but more attention should be given to the selection range of confidence.

Key words: entropy weight; unascertained measure model; lake; reservoir; evaluation of eutrophication