

文章编号:1004-8227(2005)02-0224-05

基于贝叶斯公式的湖泊富营养化 随机评价方法及其验证

谢平,李德,陈广才,叶爱中

(武汉大学水资源与水电工程科学国家重点实验室,湖北武汉430072)

摘要:为了准确评价湖泊所处的营养状态,进而为湖泊富营养化的防治提供科学依据,提出了基于贝叶斯公式的湖泊富营养化随机评价方法。该法通过计算单个水质指标属于某个评价级别的概率,由最大似然分类原则确定单个水质指标的评价级别,并采用加权平均法推求湖泊富营养化多个水质指标的综合评价级别,具有方法简单、计算工作量小的特点。经全国30个湖泊实测水质资料的验证,随机评价方法与模糊评价方法和灰色评价方法相比较,有29个湖泊评价结果的差异小于等于一个级别,其比例为96.6%,说明这三种富营养化评价方法的计算精度相当。根据各项评价指标与叶绿素的相关系数,计算各项水质指标在富营养化评价中所占权重的方法,适用于浮游植物型湖泊,如果采取等权重方法,则适用于任何类型的湖泊。确定湖泊富营养化评价的水质指标个数和种类的选择具有不确定性,确定湖泊富营养化的评价级别和各个级别对应的水质标准浓度也具有不确定性,因此湖泊富营养化的评价方法具有不确定性。应采用多种途径对湖泊富营养化的评价方法进行研究,并取长补短,使评价结果更趋合理。

关键词:湖泊富营养化;贝叶斯公式;随机评价方法;模糊评价方法;灰色评价方法;不确定性

文献标识码:A

富营养化的含义^[1]是指湖泊、水库、缓慢流动的河流以及某些近海水体中营养物质(一般指氮和磷的化合物)过量而引起水体植物(如藻类及大型植物)的大量生长。其结果是引起水质恶化、味觉和嗅觉变坏、溶解氧耗竭、透明度降低、渔业减产、死鱼、阻塞航道,对人和动物产生毒性。富营养化是水体由生产力较低的贫营养状态向生产力较高的富营养状态变化的一种自然现象,为了准确评价湖泊所处的营养状态,进而为湖泊富营养化的防治提供科学依据,国内一些研究者先后提出了模糊数学评价、灰色关联评价、神经网络评价等多种评价方法,在湖泊富营养化评价的应用中均取得了较好的效果。但由于影响湖泊富营养化的环境因子众多,难以根据环境因子的监测数据建立确定性的富营养化评价模型,而且相邻两个评价等级之间的界限是不明确的,评价因子在综合评价中应占多大权重也是不明确的,导致富营养化评价方法具有很强的不确定性。

到目前为止,人们已经认识到随机信息、模糊信息、未确知信息、灰色信息等多种不确定性信息^[2]。

随机信息是指由于客观条件的不充分或偶然因素的干扰,使得几种确定性结果的出现呈现偶然性,在某次试验中不能预知哪一个结果发生。模糊信息是指由于事物的复杂性,其元素特征界限不分明,使得多个事物的边界不清晰,造成其概念不能给出确定性的描述,不能给出确定的评定标准。未确知信息是指在某种决策时,所研究和处理的某些因素和信息可能既无随机性又无模糊性,但决策者由于条件的限制而对它们认识不清,即所掌握的信息不足以确定事物的真实状态和数量关系。灰色信息是指由于事物的复杂性,使得人类只能获得事物的部分信息或信息量的大致范围,而不能获得全部信息或确切信息。湖泊富营养化的评价,实质上是根据实测的各项水质指标浓度等多个不确定性信息,采用适宜的不确定性研究模型或方法,对湖泊富营养化程度作出判断。在湖泊富营养化评价的研究中,模糊评价和灰色评价研究较多^[3,4],而研究不确定性的主要方法—随机性方法却很少得到应用。

本文将根据概率论中的贝叶斯公式,计算单个

收稿日期:2004-02-11;修回日期:2004-04-30

基金项目:中国科学院知识创新工程项目(CX10G-E01-08/KZCX2-SW-317)资助。

作者简介:谢平(1963~),男,湖北省松滋人,博士,教授,博士生导师,主要从事水资源及水环境研究。

水质指标属于某个评价级别的概率,由最大似然分类原则确定单个水质指标的评价级别,并采用加权平均法推求湖泊富营养化多个水质指标的综合评价级别。最后采用国内 30 个湖泊的实测水质资料,将基于贝叶斯公式的随机评价方法与模糊评价方法、灰色评价方法进行比较验证,为湖泊富营养化的防治提供科学依据。

1 基于贝叶斯公式的随机评价方法

根据概率论中贝叶斯公式的定义^[5]:如果事件组 B_1, B_2, \dots, B_n 满足 $B_i \cap B_j = \emptyset (i \neq j)$ (\emptyset 为不可能事件),且 $P(\bigcup_{i=1}^n B_i) = 1, P(B_i) > 0, (i = 1, 2, \dots, n)$, 则对于任意一事件 $A (P(A) > 0)$, 有

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{i=1}^n P(B_i)P(A | B_i)} \quad (1)$$

式中 $P(A | B_i), P(B_i | A_i)$ 分别为条件概率, $P(B_i)$ 为事件发生的概率。

令水库营养化实测浓度矩阵为 $X = (x_{jk})_{m \times n}$, 其中 n 表示水库营养化样本个数; m 表示营养化评价指标项数; x 表示实测水质指标。

令营养化标准浓度矩阵为 $J = (y_{ji})_{m \times c}$, 其中 c 表示水库营养化程度的状态或分级数; y 表示评价指标的标准浓度值; $j = 1, 2, \dots, m; k = 1, 2, \dots, n; i = 1, 2, \dots, c$ 。

设 B_i 表示某个水质指标的浓度 x_{jk} 属于水质级别 i 的事件, $i = 1, 2, \dots, c; j = 1, 2, \dots, m$ 。

水库富营养化评价的不确定性可由条件概率 $P(B_i | x_{jk})$ 来表示,它指出在划分某一个水质指标的浓度 x_{jk} 为某一类别 B_i 时的概率。条件概率 $P(B_i | x_{jk})$ 可由贝叶斯公式来确定,即

$$P(B_i | x_{jk}) = \frac{P(B_i)P(x_{jk} | B_i)}{\sum_{i=1}^c P(B_i)P(x_{jk} | B_i)} \quad (2)$$

在一个最大似然分类中,如果 $P(B_i | x_{jk}) > P(B_r | x_{jk})$, 对于所有的 $i \neq r$, 则 $\{P(B_i | x_{jk}), i = 1, 2, \dots, c\}$ 中的最大值将被选定,即该水质指标的浓度属于类别 B_i 。在最大似然分类中,每一水质指标浓度的所有概率值 $\{P(B_i | x_{jk})\}$ 称之为概率矢量,将被用作描述分类水质指标浓度的属性不确定性。

在贝叶斯公式中,假设先验概率 $P(B_i)$ 可以被正确地确定,这在许多应用中是难以达到的,但在水

质评价中,水质指标属于每一类的概率是等可能的,即先验概率 $P(B_i) = 1/c$ 。

对于浮游植物型湖泊来说,富营养化现象的发生主要是湖水中磷、氮等营养物质增加,导致藻类过量繁殖所致。因而,一般均以反映湖泊藻类数量多寡的综合指标—叶绿素 a (Chla) 作为占主导地位的评价因子。但因湖泊富营养化的演变过程十分复杂,影响富营养化程度变化的因子甚多,所以在实际工作中采用多因子的综合评价,较单一因子的评价结果更接近湖泊的实际情况。文献[1]通过全国 14 个有代表性湖泊水质、生物调查资料的统计分析,得到了叶绿素 a 与其他调查项目的相关关系,并建议在实际应用中拟选用叶绿素 a (Chla)、总磷 TP、总氮 TN、化学耗氧量 COD_{Mn} 、透明度 SD 等 5 个参数作为我国湖泊富营养化评价的基本因子。

对于富营养化评价标准^[1]中的各项水质指标(见表 1),如叶绿素 a (Chla)、总磷 TP、总氮 TN、化学耗氧量 COD、透明度 SD 等均可以看作是随机变量,它们属于某一评价类别的概率可由贝叶斯公式来计算,其具体步骤为:

表 1 富营养化评价标准

Tab. 1 Eutrophication Assessment Criteria

评价标准	Chla	TP	TN	COD	SD
	(mg/m^3)	(mg/m^3)	(mg/m^3)	(mg/L)	(m)
贫营养	1	2.5	30	0.3	10
贫中营养	2	5	50	0.4	5
中营养	4	25	300	2	1.5
中富营养	10	50	500	4	1
富营养	65	200	2 000	10	0.4
重富营养	160	600	6 000	25	0.3

(1) 计算单个水质指标 x_{jk} 属于某个级别 B_i 的概率 $P(B_i | x_{jk}) = P_{ji}$

$$P_{ji} = \frac{1/L_{ji}}{\sum_{i=1}^c 1/L_{ji}} (j = 1, 2, \dots, m; i = 1, 2, \dots, c) \quad (3)$$

其中 $P(x_{jk} | B_i)$ 可根据几何概率的概念,假设与距离 L_{ji} 成反比,如图 1 所示,距离 L_{ji} 可由下式来计算。

$$L_{ji} = |x_{jk} - y_{ji}| \quad (4)$$



图 1 水质指标与评价级别的距离示意图

Fig. 1 Sketch Map of Distance Between Water Quality Index and Assessment Order

(2) 计算单个水质指标 x_{jk} 的评价级别 A_j

根据最大似然分类的原则,取 P_{ji} 中最大值对应的级别,作为单个水质指标 x_{jk} 的评价级别 A_j 。

(3) 计算单个水质指标的权重 W_j

根据叶绿素 a、总磷 TP、总氮 TN、化学需氧量 COD、透明度 SD 共 5 项评价指标与叶绿素 a 的相关系数^[1] $R = (R_{Chl-a}, R_{TP}, R_{TN}, R_{COD}, R_{SD}) = (1, 0.93, 0.81, 0.90, -0.65)$,求得各项水质指标在富营养化评价中所占的权重^[3] $W = (W_{Chl-a}, W_{TP}, W_{TN}, W_{COD}, W_{SD}) = (0.233, 0.217, 0.189, 0.210, 0.151)$,这种权重的处理方法适用于浮游植物型湖泊,如果采取等权重方法,则适用于任何类型的湖

泊。

(4) 计算多个水质指标的综合评价级别 A

$$A = \sum_{j=1}^m W_j \times A_j \quad (5)$$

2 方法验证

采用国内 30 个湖泊的实测水质资料^[3,4],按照 6 个评价标准(见表 1),将基于贝叶斯公式的随机评价方法与模糊评价方法^[3]、灰色评价方法^[4]进行比较验证,其评价结果见表 2。

表 2 湖泊富营养化评价结果表

Tab. 2 Results of Water-body Eutrophication Assessment

序号	湖泊名称	Chla (mg/m ³)	TP (mg/m ³)	TN (mg/m ³)	COD (mg/L)	SD (m)	评价结果(评价级别)			是否一致
							随机评价	模糊评价	灰色评价	
1	邛海	0.88	130	410	1.43	2.98	中营养级(3.16)	中营养级(3.28)	中营养级(3.00)	√
2	洱海	4.33	21	180	3.38	2.4	中营养级(3.21)	中营养级(3.20)	中营养级(3.00)	√
3	博斯腾湖	4.91	50	969	5.42	1.46	中富营养(3.62)	中富营养(3.75)	中富营养(4.00)	√
4	于桥水库	16.2	26	1 020	5.16	1.16	中富营养(3.78)	中富营养(3.84)	中富营养(4.00)	√
5	磁湖	15.38	87	1 540	4.4	0.65	中富营养(4.34)	中富营养(4.25)	中富营养(4.00)	√
6	巢湖	14.56	140	2 270	4.34	0.27	富营养级(4.71)	富营养级(4.69)	中富营养(4.00)	×
7	甘棠湖	77.7	135	2 140	6.96	0.36	富营养级(4.79)	富营养级(4.95)	富营养级(5.00)	√
8	蘑菇湖	82.4	332	2 660	14.6	0.49	富营养级(5.00)	富营养级(4.98)	富营养级(5.00)	√
9	杭州西湖	95.94	136	2 230	10.18	0.37	富营养级(5.00)	富营养级(4.95)	富营养级(5.00)	√
10	南京玄武湖	202.1	708	6 790	8.86	0.31	重富营养(5.64)	重富营养(5.80)	重富营养(6.00)	√
11	武汉墨水湖	262.4	500	16 050	13.6	0.15	重富营养(5.79)	重富营养(5.88)	重富营养(6.00)	√
12	广州东山湖	185.1	670	7 200	14.8	0.26	重富营养(5.79)	重富营养(5.93)	重富营养(6.00)	√
13	青海湖	14.6	20	220	1.4	4.5	中营养级(3.08)	贫中营养(2.15)	贫中营养(2.00)	×
14	太湖	100	20	900	2.83	0.5	中富营养(3.96)	富营养级(4.53)	中富营养(4.00)	×
15	呼伦湖	11.6	80	130	8.29	0.5	中富营养(3.98)	中富营养(4.08)	中富营养(4.00)	√
16	洪泽湖	11.5	100	460	5.5	0.3	中富营养(4.30)	中富营养(3.97)	中富营养(4.00)	√
17	巢湖	25.3	30	1 670	6.26	0.25	中富营养(4.28)	中富营养(4.17)	中富营养(4.00)	√
18	滇池	189.3	20	230	10.13	0.5	中富营养(4.42)	富营养级(4.99)	中营养级(3.00)	×
19	抚仙湖	19	20	210	1.61	7.03	中营养级(3.08)	贫中营养(2.10)	贫中营养(2.00)	×
20	武汉东湖	1 913.7	105	2 000	10.7	0.4	富营养级(5.02)	富营养级(5.26)	富营养级(5.00)	√
21	杭州西湖(1)	6 920	130	2 760	10.3	0.35	富营养级(5.23)	富营养级(5.27)	富营养级(5.00)	√
22	莘湖	7 060	177	2 860	51	0.41	富营养级(5.44)	重富营养(5.57)	富营养级(5.00)	×
23	外海北	945	56	1 080	3.37	0.87	中富营养(4.47)	富营养级(5.09)	中富营养(4.00)	×
24	外海中	916	40	870	3.53	0.92	中富营养(4.47)	富营养级(5.13)	中富营养(4.00)	×
25	外海南	1 011	42	960	3.68	0.9	中富营养(4.47)	富营养级(5.11)	中富营养(4.00)	×
26	洱海(1)	22.36	34	490	2.11	0.3	中富营养(3.88)	中富营养(3.90)	中富营养(4.00)	√
27	藏碧湖	8.9	49	520	2.05	1.5	中富营养(3.64)	中营养级(3.13)	中富营养(4.00)	×
28	西湖(2)	8.01	33	503	3.7	1.6	中富营养(3.63)	中营养级(3.28)	中富营养(4.00)	×
29	海西海	7.32	45	318	1.28	3.5	中营养级(3.30)	中营养级(2.68)	中营养级(3.00)	√
30	云南天池	51.63	23	620	4.05	1.2	中富营养(4.02)	中富营养(3.99)	中富营养(4.00)	√

注:表中序号 1~12 的资料来源于文献[3], 13~30 的资料来源于文献[4]。“√”表示评价结果一致,“×”表示评价结果相差一个级别,“××”表示评价结果相差两个级别。

从表2可以看出,在30个湖泊中有19个湖泊的评价结果是完全一致的,其比例为63.3%;有10个湖泊的评价结果相差一个级别,其比例为33.3%;只有一个湖泊的评价结果相差两个级别,其比例为3.4%,说明用这三种方法进行富营养化评价都是可行的。比较而言,随机评价方法与灰色评价方法的结果比较一致,两者精度相当,而与模糊评价方法的结果差异略大一些。但由于湖泊富营养化评价的不确定性,特别是当湖泊评价级别的计算值介于两个标准评价级别之间时,其结果可能取上限值,也可能取下限值,如评价级别为2.4、2.5、2.6的湖泊,其水质差异很小,而其评价结果必然相差一个级别,从这个意义上来说,在30个湖泊中有29个湖泊评价结果的差异小于等于一个级别,其比例为96.6%,说明这三种富营养化评价方法的计算精度基本相当。

3 结语

(1) 确定湖泊富营养化评价的水质指标个数和种类的选择具有不确定性,确定湖泊富营养化的评价级别和各个级别对应的水质标准浓度也具有不确定性,因此湖泊富营养化的评价方法具有不确定性。

(2) 湖泊富营养化评价的实质是根据实测的各项水质指标浓度等多个不确定性信息,采用适宜的不确定性研究方法或模型,对湖泊富营养化程度作出判断。

(3) 不确定性研究可以采用随机信息、模糊信息、未确知信息、灰色信息等多种不确定性研究方

法,应采用多种途径对湖泊富营养化的评价方法进行研究。

(4) 基于贝叶斯公式的湖泊富营养化随机评价方法,具有方法简单、计算工作量小的特点,经全国30个湖泊实测水质资料的验证,并与模糊评价方法和灰色评价方法的结果进行比较,发现这三种方法的计算精度相当。

(5) 采用各项评价指标与叶绿素的相关系数,计算各项水质指标在富营养化评价中所占权重的方法,适用于浮游植物型湖泊,如果采取等权重方法,则适用于任何类型的湖泊。

(6) 用多种方法进行湖泊富营养化评价可以取长补短,使评价结果更趋合理,更加符合实际要求。如何对多种方法的富营养化评价结果进行综合是一个值得研究的课题。

参考文献:

- [1] 舒金华. 我国湖泊富营养化程度评价方法的探讨[J]. 环境污染与防治, 1990, 12(5): 2~7.
- [2] 王清印, 刘志勇. 不确定性信息的概念、类别及其数学表述[J]. 运筹与管理, 2001, 10(4): 9~15.
- [3] 陈守煜. 湖库水体富营养化评价级别特征值与识别模型[J]. 黑龙江水专学报, 1999, 26(1): 1~8.
- [4] 冯玉国. 湖泊富营养化灰色评价模型及其应用[J]. 系统工程理论与实践, 1996, 8: 43~47.
- [5] 《数学手册》编写组. 数学手册[Z]. 北京: 高等教育出版社, 1979. 785~786.

A LAKE EUTROPHICATION STOCHASTIC ASSESSMENT METHOD BY USING BAYESIAN FORMULA AND ITS VERIFICATION

XIE Ping, Li De, CHEN Guang-cai, YE Ai-zhong

(State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China)

Abstract: In order to correctly evaluate the trophic state of a lake and then to establish scientific basis for the prevention of lake eutrophication, a lake eutrophication stochastic assessment method (LESAM) by using Bayesian formula is proposed. In this method, according to the maximum likelihood ordering principle, the assessment order of each water quality index of an actual lake was firstly derived through calculating the probability of each of those water quality indices belonging to the assessment order used in the lake eutrophication assessment criteria. Secondly, the assessment order of lake eutrophication is derived by using the weighing average method. The characteristics of LESAM are that the method is simple, and the computation work is moderate. LESAM, together with other two methods, i. e. the lake eutrophication fuzzy assessment method (LEFAM) and the lake eutrophication gray assessment method (LEGAM), has been tested against the measured data of 30 lakes all over the country. It is found that the difference of assessment results among those three methods is not greater than 1 order in 29 lakes, i. e. 96.6% of the total lakes. It is concluded that all these three methods (LESAM, LEFAM, and LEGAM) have comparable computation precision. Though in lake eutrophication assessment we can calculate the weight of each water quality index according to the correlation coefficient between each water quality index and chlorophyll, this method only suits the lakes of phytoplankton type. As to the equally weighting average method, it can be used for any kind of lakes. For there is a lot of uncertainty involved both in determining the number and type of water quality index in lake eutrophication assessment and in determining the assessment order and its corresponding standard water quality concentration, the lake eutrophication assessment methods also have uncertainty. It is suggested that the lake eutrophication assessment should be studied with various approaches in order to overcome the weakness of one particular method and to make assessment results more reasonable.

Key words: lake eutrophication; Bayesian formula; stochastic assessment method; fuzzy assessment method; grey assessment method; uncertainty