

文章编号: 1673-9469(2017)04-0070-04

doi:10.3969/j.issn.1673-9469.2017.04.016

基于正态分布原理的贝叶斯水质综合评价

孙玲玲¹, 王树谦¹, 石宝红², 闫丹丹¹, 杨旭洋¹

(1. 河北工程大学 水利水电学院, 河北 邯郸 056001; 2. 河北省黄壁庄水库管理局, 河北 石家庄 050224)

摘要: 将正态分布原理与贝叶斯方法相结合来进行黄壁庄水库水质综合评价, 并给出了改进的贝叶斯(Bayes)方法的思路和计算方法。以 GB3838-2002 为评价标准, 采用改进的贝叶斯方法建立贝叶斯水质评价模型, 对黄壁庄水库坝上水质监测数据 8 项指标进行水质评价研究。评价结果表明: 改进的贝叶斯综合水质评价模型适用于黄壁庄水库的水质评价, 且黄壁庄水库综合水质处于 I—II 类, 年内水质变化不大, 丰水期水质相对较差, 为 II 级水质, 其余月份均达到 I 类水质标准。

关键词: 正态分布; 贝叶斯; 水质综合评价; 黄壁庄水库

中图分类号: X824

文献标志码: A

Comprehensive assessment of water quality based on Bayesian Method & Normal Distribution Theory

SUN Lingling¹, WANG Shuqian¹, SHI Baohong², YAN Dandan¹, YANG Xuyang¹

(1. Institute of Water Conservancy and Hydro-Electric Power, Hebei University of Engineering, Handan 056001, China;

2. Huangbizhuang Reservoir Authority in Hebei Province, Shijiazhuang, 050224, China)

Abstract: The Bayes method is based on the normal principles to analyze and evaluate the water quality of the Huangbizhuang Reservoir, the idea and calculation method of the improved Bayes method are also given. By using GB3838-2002 as evaluation criterion, the improved Bayesian model for water quality evaluation is established through selecting water quality monitoring data on dam, Huangbizhuang Reservoir. The evaluation results show that the model is suitable for Huangbizhuang reservoir water quality simulation, the water quality of Huangbizhuang Reservoir is mainly I and II grade, while it is worse in the rain season for water quality being II grade, other months achieve I standard.

Key words: Normal distribution; Bayes method; Water quality comprehensive evaluation; Huangbizhuang Reservoir

我国是一个严重的缺水型国家, 水资源紧张、水资源利用水平低下、开发不平衡、地下水超采严重、用水浪费等问题不断凸显^[1]。特别是近些年我国经济迅速发展, 工农业结构粗狂式的开发以及不合理利用使得水资源受到严重污染, 水体越来越不能满足人们正常生活饮用水水源的要求, “水质型缺水”现象愈来愈严重。随着人们对水资源水质意识的提高, 寻找出一种简便、有效、切实可行的水质评价方法是必要的。国内外研究学者在水质评价方面均做了大量研究, 出现了大量的水质评价方法^[2-7], 但

这些方法存在一定的不确定性, 或受主观影响较强, 且对基础资料要求较高、计算复杂, 因此适用性受到限制。文章选取贝叶斯方法对诱发结果的最可能原因进行概率推理^[8], 避免了因资料不全或受主观因素过重等带来的影响^[9], 并尝试将正态分布原理与贝叶斯方法相结合来进行黄壁庄水库水质评价, 在贝叶斯方法原理基础上, 增加正态分布原理来估计水质评价中的似然概率, 避免抽样带来的水质评价误差, 使评价结果更加合理、可靠。

收稿日期: 2017-09-15

特约专稿

基金项目: 国家水体污染控制与治理科技重大专项(2014ZX07203-008)

作者简介: 孙玲玲(1991-), 女, 河北衡水人, 硕士研究生, 研究方向水资源水环境。

1 水质评价指标和标准

本次水质评价依据《地表水环境质量标准》(GB3838-2002)为评价标准,结合地表水评价内容和实际监测数据情况进行黄壁庄水质综合评价。选取地表水水质评价必评内容6项,选评内容2项,共8项内容为评价指标。地表水现状评价内容分类见表1,水质类型标准见表2。

2 贝叶斯评价原理

贝叶斯方法是结合现有资料,利用概率论统计学的方法对其进行某种预测推理分析的过程,即推测某一事件发生的可能性大小。最初在使用贝叶斯方法对水质进行评价时,采用距离法计算单个指标属于某个级别的概率^[9, 11-12],在使用距离法计算似然概率时,时常会出现单指标良好而均值远次于水质标准的情况,或单指标超标而均值远优于水质标准的情况,并不能反应水质的真实情况,容易给人造成水质过好或过差的假象。因此,本文在进行贝叶斯方法计算黄壁庄水库综合水质评价时使用正态分布原理计算似然概率,更科学、客观地进行水质综合评价,为黄壁庄水库的水资源保护和监督管理提供科学依据。

Bayes公式的定义^[13]:设随机试验E的样本空间为Q, $B_i (i=1, 2, \dots, s)$ 为Q的一个有限划分且 $P(B_i) > 0$,对任何 $A \subset Q$,当 $P(A) > 0$,有

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

式中: $P(B_i|A)$, $P(A|B_i)$ 为条件概率; $P(B_i)$ 为事件 B_i 的概率。

在进行水质评价时,可将(1)式改写为:

$$P(y_{ji}|x_j) = \frac{P(y_{ji})P(x_j|y_{ji})}{\sum_{i=1}^s P(y_{ji})P(x_j|y_{ji})} \quad (2)$$

式中: y_{ji} 为水质类型的标准值,其中*i*为标准类型 $i=1, 2, \dots, s$; j 为所需评价的指标, $j=1, 2, \dots, m$; x_i 为代表月份指标值,月份总数为*n*。

$P(y_{ji})$ 为人们根据经验直觉和判断所推测的该代表站点的水质级别的可能性所推测出来的,称为先验概率。

$P(y_{ji}|x_i)$ 代表月份水质某一指标 x_i 属于该指标水质类型 y_{ji} 的可能性,与水质的级别密切相关,该值越大代表属于某一水质等级的可能性越大,可将其看作一种似然概率。根据统计理论,由于抽样缘故样本指标值和实际值总体值(标准值)往往存在一定的抽样误差,为尽可能消除这种误差,尝试使用正态分布原理来估计 $P(x_i|y_{ji})$ 。

$P(y_{ji}|x_i)$ 为后验概率,表示已知代表月份指标值 x_i 条件下,该指标水质属于*i*级别的可能性的概率。

3 黄壁庄水库水质评价

表1 地表水评价内容分类^[10]
Tab.1 Surface water evaluation content classification

必评内容	选评内容	参考内容
溶解氧(DO)、氨氮(NH ₃ -N)、高锰酸盐指数、挥发酚、化学需氧量(COD)、砷	5日生化需氧量(BOD ₅)、汞、铜、锌、氟化物、氰化物、总磷、镉、六价铬、石油类等	水温、pH值、总硬度

表2 水质类型标准
Tab.2 Water quality type standard

(单位: mg/L)

评价因子	I	II	III	IV	V
溶解氧	7.500	6.000	5.000	3.000	2.000
高锰酸盐指数	2.000	4.000	6.000	10.000	15.000
COD	15.000	15.000	20.000	30.000	40.000
氨氮	0.150	0.500	1.000	1.500	2.000
挥发酚	0.002	0.002	0.005	0.010	0.100
砷	0.050	0.050	0.050	0.100	0.100
BOD ₅	3.000	3.000	4.000	6.000	10.000
氟化物	1.000	1.000	1.000	1.500	1.500

黄壁庄水库水质评价选取黄壁庄水库坝上水质监测断面作为研究对象,监测断面设置左中右三个位置点,基本每天采样进行分析,2016 年黄壁庄水库坝上水质监测点连续监测 12 个月份,对各月份监测数据取平均值,得到各月水质监测指标。黄壁庄水库 2016 年坝上水质监测断面监测指标见表 3。

3.1 计算先验概率

在没有关于黄壁庄水库坝上断面每月水质信息条件下,认为该断面水质每月属于某类的概率相同,

即: $P(y_{j1}) = P(y_{j2}) = \dots = P(y_{j5}) = \frac{1}{5}$ 。

3.2 计算

$P(x_j|y_{ji})$ 代表月份水质某一指标 x_j 属于该指标水质类型 y_{ji} 的可能性,计算 $P(x_j|y_{ji})$ 即计算单个指标属于某级别的概率。采用正态分布原理用于估计 $P(x_j|y_{ji})$,以 j 指标水质类型标准值的均值作为该指标正态分布的 α_j , j 指标各水质类别标准差为 σ_j , C_{vj}

为 j 指标各类变差系数,由 $C_{vj} = \frac{\sigma_j}{\alpha_j}$ 来估计 C_{vj} ,接

着计算第 i 类指标的标准差 σ_{ji} , $\sigma_{ji} = C_{vj} y_{ji}$ (指标各类标准差计算结果见表 4),将计算出的 σ_{ji} 带入 $t_{ji} = \frac{x_j - y_{ji}}{\sigma_{ji}}$ 中对抽样值进行标准化,用标准化正态分布对 $P(x_j|y_{ji})$ 进行计算:

$$P(x_j|y_{ji}) = 2(1 - \phi(|t_{ji}|)) \tag{3}$$

式中: $\phi(|t_j|) = \int_{-\alpha}^{|t_j|} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$, $|t_j|$ 为 t_j 的绝对值。

根据公式 (2) 计算后验概率 $P(y_{ji}|x_j)$ 。

3.3 计算多指标综合水质后验概率 P_i

$$P_i = \sum_{j=1}^m w_j P(y_{ji}|x_j) \tag{4}$$

式中: w_j 为不同水质指标 j 的权重,可根据各项水质指标对水质类型的影响及水用途来进行确定。在对黄壁庄水库进行水质综合评价时,根据《河北省水功能区划》,黄壁庄水库为 I 级保护区,且作为周边城区供水水源地,各项水质指标均应达到饮用

表 3 黄壁庄水库坝上断面监测指标 (单位: mg/L)
Tab.3 The monitoring indexes of dam upper section in the Huangbizhuang reservoir

评价因子月份	DO	高锰酸盐指数	COD	氨氮	挥发酚	砷	BOD ₅	氟化物
1	10.5	2.9	7.9	0.07	0.002	0.0002	2.0	0.50
2	11.6	2.6	8.3	0.04	0.0019	0.0002	2.0	0.50
3	13.4	2.3	8.6	0.06	0.0019	0.0002	2.0	0.50
4	10.5	2.8	8.6	0.05	0.0018	0.0002	2.7	0.50
5	8.7	1.8	8.9	0.09	0.0018	0.0002	2.0	0.50
6	8.3	3.5	9.2	0.04	0.002	0.0002	2.0	0.50
7	8.9	2.7	9.3	0.05	0.002	0.0002	3.2	0.50
8	7.8	3.2	9.5	0.09	0.0019	0.0002	3.6	0.50
9	7.3	3.4	9.0	0.05	0.002	0.0002	3.5	0.50
10	8.7	3.6	8.7	0.04	0.0018	0.0002	2.6	0.50
11	10.1	2.7	7.6	0.05	0.002	0.0002	2.7	0.50
12	10.3	3	8.1	0.05	0.0017	0.0002	2.2	0.50

表 4 黄壁庄水库各类指标标准差计算结果
Tab.4 Calculation result of various indexes of Huangbizhuang reservoir

指标	I	II	III	IV	V
DO	7.898 94	6.319 15	5.265 96	3.159 57	2.106 38
高锰酸盐指数	7.243 24	14.486 49	21.729 73	36.216 22	54.324 32
COD	73.437 50	73.437 50	97.916 67	146.875 00	195.833 33
氨氮	0.080 75	0.269 17	0.538 35	0.807 52	1.076 70
挥发酚	0.000 15	0.000 15	0.000 38	0.000 77	0.007 67
砷	0.000 54	0.000 54	0.000 54	0.001 07	0.001 07
BOD ₅	5.019 23	5.019 23	6.692 31	10.038 46	16.730 77
氟化物	0.062 50	0.062 50	0.062 50	0.093 75	0.093 75

表5 多指标综合水质后验概率
Tab.5 The posterior probability of comprehensive water quality for multiple indexes

监测断面	月份	I	II	III	IV	V
坝上	1	0.278 7	0.254 5	0.175 2	0.148 8	0.142 7
	2	0.260 1	0.238 5	0.184 0	0.161 9	0.155 5
	3	0.263 7	0.233 8	0.180 1	0.164 0	0.158 3
	4	0.252 3	0.231 4	0.194 5	0.164 9	0.156 9
	5	0.267 5	0.227 6	0.193 5	0.162 2	0.149 3
	6	0.268 8	0.270 4	0.171 9	0.151 3	0.137 6
	7	0.264 1	0.266 0	0.181 1	0.152 0	0.136 8
	8	0.224 8	0.229 7	0.228 0	0.170 5	0.146 9
	9	0.256 4	0.268 5	0.190 1	0.149 8	0.135 2
	10	0.246 4	0.237 0	0.200 8	0.163 5	0.152 4
	11	0.276 6	0.256 9	0.176 9	0.149 1	0.140 5
	12	0.246 3	0.225 6	0.198 9	0.168 5	0.160 6

表6 黄壁庄水库综合水质评价结果
Tab.6 The comprehensive water quality assessment in the Huangbizhuang reservoir

坝上断面	月份	1	2	3	4	5	6	7	8	9	10	11	12
	水质类型	I	I	I	I	I	II	II	II	II	I	I	I

水指标,在进行黄壁庄水库水质综合评价时采取平均权重,即 $w_{j1}=w_{j2}=\dots=w_{j12}=\frac{1}{8}$ 。

多指标均匀权重下黄壁庄水库坝上断面 2016 年各月份的综合水质后验概率见表 5。

3.4 确定最终各月份水质级别

采用最大概率原则进行确定黄壁庄水库 2016 年不同月份的水质类型 h :

$$P_h = \max_{i=1, 2, \dots, 12} P_i \quad (5)$$

从而确定黄壁庄水库 2016 年坝上监测断面 12 个月份的综合水质评价结果,得到表 6。

3.5 黄壁庄水库综合水质评价结果分析

本文基于正态分布原理估计似然概率来改进贝叶斯方法,构建贝叶斯综合水质评价模型,对黄壁庄水库坝上断面 2016 年 12 个月份的监测数据进行水质综合评价研究,得到以下结论:

黄壁庄水库水质的年内变化不大,均达到 II 级水质标准以上,年内 6~9 月份水质相对较差,为 II 级水质,其余月份均达到 I 类水质标准。

水质较差月份集中在年内丰水期,且处于夏季,受夏季生物繁殖较快影响,水库生物大量消耗溶解氧,伴随着温度的升高,各生物间及各污染物质之间的相互作用增加,各种复杂反应所释放物质均可对水库水质造成一定的影响,造成水库水质恶化。

此外,由于丰水期降水增加,大气污染物质、化肥农药及其他有毒物质伴随着降雨径流均直入水库,造成对水库水体水质的影响。

9 月份过后,水库水质恢复 I 类水质,一方面是黄壁庄水库容量较大,水体具有较强的水体自净能力;另一方面,由于灌溉黄壁庄水库向下游排水使水库污染物总量减少,同时上游岗南水库来水,淡化了库区内污染物浓度,使水质恢复到 I 类水。

选取的 8 个水质评价指标中,通过表 2 和表 3 的分析即可看出,对水库水质影响较大的是高锰酸盐指数和 BOD₅,此次对黄壁庄水库水质进行综合评价时,采用各类物质同权重进行计算,倘若增大高锰酸盐指数和 BOD₅ 的权重,水质结果与表 5 相比会更差。

4 结论

采用贝叶斯公式计算水质评价具有综合性强、计算简便、快捷等特点,是可以应用到水质评价中的一个新方法,同时本文结合正态分布原理估计似然概率,避免了抽样误差的影响,可以更科学、合理的用于评价水质。但值得注意的是:采用最大概率法对结果进行判定过于敏感,如 6、7 月份水质 I 类和 II 类仅差 0.001 6 和 0.002,差分比分别为 0.595%、0.75%;虽然使用正态分布原理计算贝叶斯公式中的似然概率,一定程度上(下转第 83 页)

- 及应用:以北部湾盆地福山凹陷为例[J].地学前缘, 2015, 22(1): 215-222.
- [2] 刘文超. 码头庄油田储层非均质性及其与剩余油分布的关系[J]. 岩性油气藏, 2012(2): 111-116.
- [3] 裴亦楠, 薛叔浩. 油气储层评价技术[M]. 北京: 石油工业出版社, 2001: 290-324.
- [4] 马庆林, 赵淑娥, 廖远涛, 等. 北部湾盆地福山凹陷古近系流沙港组层序地层样式及其研究意义[J]. 地球科学(中国地质大学学报), 2012(4): 667-678.
- [5] 孙鸣, 王华, 廖远涛, 等. 福山凹陷白莲地区流沙港组一段扇三角洲沉积体系与油气成藏条件分析[J]. 中南大学学报: 自然科学版, 2013(10): 4150-4160.
- [6] 刘恩涛, 王华, 李媛, 等. 北部湾盆地福山凹陷构造转换带对层序及沉积体系的控制[J]. 中国石油大学学报: 自然科学版, 2013(3): 17-22.
- [7] 石彦民, 刘菊, 张梅珠, 等. 海南福山凹陷油气勘探实践与认识[J]. 华南地震, 2007(3): 57-68.
- [8] 张庆, 朱玉双, 郭兵等. 杏北区长6储层非均质性对含油性分布的影响[J]. 西北大学学报: 自然科学版, 2009(2): 277-282.
- [9] 田景春, 刘伟伟, 王峰, 等. 鄂尔多斯盆地高桥地区上古生界致密砂岩储层非均质性特征[J]. 石油与天然气地质, 2014(2): 183-189.
- [10] 李凯强, 张景军, 李静伟, 等. 鄂尔多斯盆地西南部长6储层岩石学特征[J]. 河北工程大学学报: 自然科学版, 2016, 33(4): 93-97.
- [11] 马淑英. 鄂尔多斯盆地白马南地区长81储层非均质性研究[D]. 西安: 西北大学, 2006.
- [12] 李金燕. 乔庄油田沙三中储层特征研究[D]. 长沙: 中南大学, 2007.
- [13] 谢青, 王建民, 时圣彪. 鄂尔多斯盆地太阳湾油区长2储层非均质性研究[J]. 岩性油气藏, 2015, 27(4): 61-67.

(上接第73页)解决了抽样所带来的误差,同时水质权重也是评价中的关键因子,但在水质指标权重的确定方面没有确定的标准,对黄壁庄水库水质评价结果仍有待进一步探讨。

参考文献:

- [1] 联合国环境规划署. 全球环境展望 2000[M]. 北京: 中国环境科学出版社, 2001.
- [2] 石宝红. 黄壁庄水库生态环境现状及修复措施探讨[C]//2015第七届全国河湖治理与水生态文明发展论坛论文集, 2015: 323-325.
- [3] 曾永, 樊引琴. 水质模糊综合评价法与单因子指数评价法比较[J]. 人民黄河, 2007, 29(2): 64-65.
- [4] 汪尚朋, 李江云, 郑旭荣, 等. 水质模糊评价的探讨[J]. 中国农村水利水电, 2005(1): 49-51.
- [5] 章新, 贺石磊, 张雍照, 等. 水质评价的灰色关联分析方法研究[J]. 水资源与水工程学报, 2010(5): 117-119.
- [6] 王晓萍, 孙继洋, 金鑫. 基于BP神经网络的钱塘江水质指标的预测[J]. 浙江大学学报: 工学版, 2007, 41(2): 361-364.
- [7] 王海峰, 王慧勇, 李春燕, 等. 邯郸市地下水水化学类型分布及污染物成因分析[J]. 河北工程大学学报: 自然科学版, 2017, 34(1): 53-56.
- [8] 孙鹏程, 陈吉宁. 基于贝叶斯网络的河流突发性水质污染事故风险评估[J]. 环境科学, 2009, 30(1): 47-51.
- [9] 侯佑泽, 秦天玲, 严登华, 等. 基于贝叶斯理论的武烈河水质综合评价[J]. 南水北调与水利科技, 2012, 10(1): 94-97.
- [10] 万咸涛. 水资源质量评价综述[J]. 水资源研究, 2005, 26(4): 32-37.
- [11] 廖杰, 王文对, 丁晶. 贝叶斯公式在河流水质综合评价中的应用[J]. 四川师范大学学报: 自然科学版, 2007, 30(4): 519-522.
- [12] 扈震, 杨之江, 齐培培. 基于贝叶斯方法的饮用水水质预测及可视化研究[J]. 中国给水排水, 2012, 28(5): 53-56.
- [13] 何镇邦, 李桂荣. 概率论与数理统计[M]. 北京: 北京理工大学出版社, 1988.