文章编号:1000-5404(2006)17-1819-02

技术方法

基于伪F统计量的水质综合评价模型

王文昌[1],易 东[1],赵增炜[1],张彦琦[1],刘 玲[1],曹 佳[2] (第三军医大学:[1]军事预防医学院卫生统计学教研室,[2]卫生毒理学教研室,重庆 400038)

提 要:目的 针对目前水质综合评价中存在的某些问题,构造出水质综合评价的一个新模型。**方法** 利用方差分析的思想,采用伪F统计量,得到评价指标的权重系数,再采用加权线性方法建立评价模型。**结果** 把该模型运用于一个实际数据,得到的结论与实际情况相吻合。**结论** 该模型对可以克服评价指标的随机性干扰,有一定的自适应性,方法简单、实用。

当前,我国各地区水质存在严重恶化趋势。对水质的监测是一个连续的动态的过程。如何对这个动态数据进行正确评价是一个急需解决的问题。目前存在的评价方法主要是参考国家有关标准并结合模糊综合评价或灰色关联分析等技术[1-4]。但是这些评价必须给出监测指标的静态权重系数或污染物参考标准,不能适应水污染中某些指标的突发性和偶发性问题。为此,我们借用统计学中的方差分析思想,基于伪F统计量,给出一个自适的指标权重的估计方法,构造了一个综合评价模型用于水质评价。

1 理论与方法

一个较好的评价体系,其权重应该有一定的自适应性,既能最大限度地区分不同监测地点在同一时段内综合表现的差异,又有一定的抗干扰能力,其评价结果既要具有一定的稳定性,又要具有一定的分辨率。评价指标的稳定性可以用指标的组内间变异来刻画,分辨率可以用组间变异来刻画,因此可以借助于方差分析思想来寻求一个自适应的权重估计。

用y_{ijk}表示第k个采样时间点、第j个采样地点的第i个指标值,其中$i=1,\cdots,m$,$j=1,\cdots,n$,$k=1,\cdots,T$;w_i为指标i的权重。

1.1 指标的标准化

与一般的综合评价类似,先采用一定的方法将监测值y_{ijk}标准化为均值为0、标准差为1的标准化变量值x_{ijk}。

1.2 权重的计算

记第i个指标的标准化值均值为$\bar{x}_{i\cdot\cdot}$,记第i个指标、第j个采样地点各采样时间点上的标准化值均值为$\bar{x}_{ij\cdot}$,则第i个指标组内变异用如下公式计算:

$$SS_{in}^{i}=\sum_{j}\sum_{k}(x_{ijk}-\bar{x}_{ij\cdot})^2 \quad (1)$$

第i个指标组间变异的计算公式为:

$$SS_{out}^{i}=\sum_{j}T(\bar{x}_{ij}-\bar{x}_{i\cdot\cdot})^2 \quad (2)$$

令:

$$F^{i}=\frac{SS_{out}^{i}/(n-1)}{SS_{in}^{i}/(nT-n)} \quad (3)$$

与方差分析类似,第i个指标的F^i刻画了各评价对象扣除随机误差干扰后在第i个指标上的总体差异。考虑到F^i是一个方差比值,为了避免夸大组间差异,对$\sqrt{F^i}$进行归一化处理得到第i个指标的权重,即:

$$w_i=\frac{\sqrt{F^i}}{\sum_{j}\sqrt{F^i}} \quad (4)$$

显然,这样计算出的w_i越大,说明各评价对象在该指标上的差异越大,因此该指标权重值是在充分考虑评价指标稳定性的基础上最大限度地暴露各评价对象的差异。

1.3 综合评价

第j个采样地点的第k个采样时间点的综合评价分值可用如下公式计算:

$$score_{jk}=\sum_{i}w_i x_{ijk} \quad (5)$$

2 实例分析

为了能够及时全面地掌握主要流域重点断面水体的水质状况,预警或预报重大(流域性)水质污染事故,从1999年9月开始至2003年12月,国家环保总局在松花江、辽河、海河、黄河、淮河、长江、珠江、太湖、巢湖、滇池等流域建设了82个水质自动监测站,其水质监测结果及4个重要的评价因子:pH值、DO(溶解氧)、CODMn(高锰酸盐指数)、NH3-N(氨氮)以周报形式定期发布在国家环境保护总局网站(http://www.zhb.gov.cn)及《光明日报》、《中国环境报》等媒体上,取得了广泛的社会关注和强烈的反响。本研究分析数据为从2005年27期到2005年52期即2005年下半年的长江干流7个监测点(D1:四川攀枝花龙洞,D2:重庆朱沱,D3:湖北宜昌南津关,D4:湖南岳阳城陵矶,D5:江西九江河西水厂,D6:安徽安庆皖河口,D7:江苏南京林山)的pH、DO、CODMn、NH3-N的监测结果。

基金项目:国家科技部攻关项目(2003DA903B03-02);国家科技部西部引导项目(2003BA869C)
Supported by the Key Sci &Tech Research Foundation of State Ministry of Science and Technology (2003DA903B032-02), and the "Guided Project for West of China" of State Ministry of Science and Technology (2003BA869C)

作者简介:王文昌(1970-),男,贵州省大方县人,讲师,主要从事卫生统计学方面的研究。电话:(023)68752343
通讯作者:曹 佳,电话:(023)68752271, E-mail: Caojia@mail.tmmu.com.cn

收稿日期:2005-05-23;修回日期:2006-05-31

2.1 指标标准化

记第 i 个指标的均值为 $\bar{y}_{i..}$，标准差为 $S_{i..}$。DO 为高优指标，其标准化值采用如下公式计算：

$$x_{ijk}=\frac{y_{ijk}-\bar{y}_{i..}}{S_{i..}} \tag{6}$$

CODMn、NH3-N 为低优指标，其标准化值采用如下公式计算：

$$x_{ijk}=\frac{\bar{y}_{i..}-y_{ijk}}{S_{i..}} \tag{7}$$

pH 值为中优指标，理论上应该为 7，因此我们先将原始数据减去 7 取绝对值后按低优指标进行处理。

2.2 权重值的计算见表 1

表 1 权重计算

污染物	SS_{in}	SS_{out}	F	w
pH	128.917	52.083	65.520	0.299
DO	133.589	47.411	74.586	0.319
CODMn	70.811	110.189	17.011	0.152
NH3-N	107.473	73.527	38.692	0.230

由表 1 可看出，长江干流 2005 年下半年的水质差异在扣除随机波动后主要来自于 DO、pH 和 NH3-N，其次才是 CODMn。

2.3 评价结果

长江干流各采样点 2005 年下半年的水质评价结果如图 1。

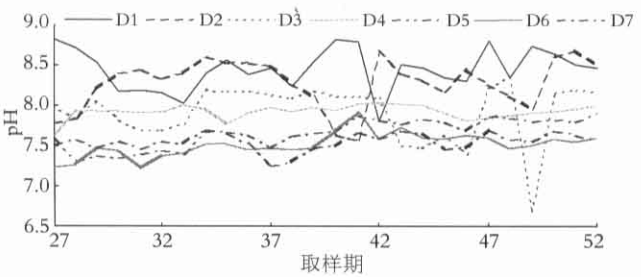


图 1 各取样点的综合评价结果

由图 1 可以看出：重庆朱沱在 27、28 及 38 期水质较差，其主要污染物为 NH3-N（超过国家Ⅲ类水质标准）；三峡水库出口湖北宜昌南津关在 32、33 期水质较差，其主要污染物还是 NH3-N；从 39 期开始的相当长一段时间内湖南岳阳城陵矶的水质表现均较差，其主要污染物为 CODMn 和 NH3-N；江西九江河西水厂、安徽安庆皖河口及江苏南京林山 3 个取样点的水质表现比较平稳，监测结果显示大多为Ⅱ类水质；四川攀枝花龙洞的 CODMn 在 34 期一度高于其他监测点，检测结果显示属Ⅲ类水质，但从监测结果可以看出，四川攀枝花龙洞的 CODMn 除开 34 期外表现比较平稳，其他指标在所有观察时间点上表现也比较平稳，从统计观点来看，该点表现应该是一种随机干扰，并且 CODMn 对综合评价结果影响较小，因此其总体评价结果未出现较大的波动。

3 讨论

为了使评价过程有一定的适应性，克服指标随机扰动对评价结果的影响，文献[5]在考虑稳定性和分辨率时是对如下模型(8)采用遗传算法进行优化得到权系数的。

$$\min \quad K_1W+\frac{K_2}{A+F} \tag{8}$$

其中，K_1、K_2、A 为常数，W 反映了评价指标的稳定性，即

$$W=\frac{1}{nT}\sum_{i=1}^{m}\sum_{j=1}^{T}\sum_{k=1}^{T}w_i\,|x_{ijk}-\bar{x}_{ij.}|;$$

F 反映了评价指标的分辨率，即

$$F=\frac{1}{nT}\sum_{i=1}^{m}\sum_{j=1}^{T}\sum_{k=1}^{T}w_i\,|x_{ijk}-\bar{x}_{i..}|。$$

该模型求解比较麻烦，在某些极端情形下，很可能求解出某几个指标的权值非 0，而其余值指标的权值为 1 的结果，不具备普适性。

由于本方法采用伪 F 统计量来构造权重，根据伪 F 统计量的统计性质，本方法能够在克服随机干扰基础上分辨出各评价对象的差异主要表现在什么指标上，权重具有一定的自适性，使得评价结果具有一定的稳定性和分辨率。由实际应用结果来看本方法的评价结果与实际情况比较吻合。

在进行水质评价时，一般采用国家相应的水质标准再使用模糊综合评价或灰关联分析进行，但是这些评价方法不能处理污染的突发性和偶发性问题，也不可能找出监测点间的最大差异在什么指标上，从而不能给污染物的防控提供更进一步的详细信息。本方法构造出的权重能够充分地暴露评价对象间的差异，权重大小反应了各评价对象在各指标上的差异大小，能为污染物的防控提供更进一步的信息。

关键词：水质评价；伪 F 统计量

中图法分类号：R195.1 **文献标识码**：B

参考文献：

[1] 吕义清，武胜忠，孙建星. 灰色关联分析在水质综合评价中的应用[J]. 太原工业大学学报，1997，28(3)：31－36.

[2] 路明娟. 模糊数学在水质评价中的应用[J]. 甘肃科技，2000，(4)：23－24.

[3] LIOU S M，LO S L，HU C Y. Application of two-stage fuzzy set theory to river quality evaluation in Taiwan[J]. Water Res，2003，37(6)：1406－1416.

[4] CHANG N B，CHEN H W，NING S K. Identification of river water quality using the fuzzy synthetic evaluation approach[J]. J Environ Manage，2001，63(3)：293－305.

[5] 倪苏云，攀 登，吴冲锋. 基于遗传算法的基金绩效综合评价研究[J]. 系统工程，2003，21(2)：1－6.

[6] 郭祖超. 医用数理统计方法[M]. 第 3 版. 北京：人民卫生出版社，1988：259－263.

（编辑 薛国文）