

- Allen Press, 1968.
 [4] Forester, J., *World Dynamics*, Wright-Allen Press, 1971.
 [5] Meadows, D., *Dynamics of Growth in a Fi-*

nite World, Wright-Allen Press, 1972.

- [6] Meadows, D. et al., *The Limits to Growth* Wright-Allen Press, 1972.
 [7] UNESCO-MAB Programme, 5, 10(1981).

半集均方差水质评价模式

徐 鸿 楷

(青岛市环境保护科学研究所)

本文根据数理统计的基本概念,结合水质参数各自对水环境质量的影响,建立了半集均方差模式,并通过地面水质评价的应用进行了检验。

一、综合数学模式的简介

在水质评价中,国内外已提出十余种由水质参数的分指数 P_i 综合为水质指数 WQI 的数学模式。鲍尔顿 (Bolton) 等^[1] 对国外前期的部分模式做了归纳(表 1)。八十年代以来,我国的环境评价工作者对综合数学模式进行了较广泛的研究,做了有益的探索。现将国内外较典型的综合数学模式列入

表 2。

表 1 水质指数 WQI 公式

名 称	WQI 公 式	
	均权	加权
算术指数	$\frac{1}{n} \sum_{i=1}^n P_i$	$\sum_{i=1}^n W_i P_i$
苏尔维指数	$\frac{1}{100} \left(\frac{1}{n} \sum_{i=1}^n P_i \right)^2$	$\frac{1}{100} \left(\sum_{i=1}^n W_i P_i \right)^2$
几何指数	$\left(\prod_{i=1}^n P_i \right)^{1/n}$	$\prod_{i=1}^n P_i^{W_i}$

在表 1、2 中:

P_i ——水质参数 i 的分指数;

表 2 水质指数 WQI 公式

名 称	WQI 公 式	提 出 者	年 代	文 献 号
内梅罗模式	$\sqrt{\frac{(P_i)_{\max}^2 + \bar{P}^2}{2}}$	N. L. Nemerow	1974	[2]
几何均数模式	$\sqrt{(P_i)_{\max} \cdot \bar{P}}$	姚志麒	1980*	[3]
统计模式	$\left\{ \begin{array}{l} E(P_i) \triangleq \sum_{i=1}^n W_i P_i \\ \sigma^2 \end{array} \right.$	熊广政 裘小松	1980	[4]
向量模式	$\sqrt{\sum_{i=1}^n P_i^2}$	申葆诚 徐国义	1981	[5]
模糊数学模式	ΔOR (模糊矩阵复合运算)	容跃 金键	1982	[6]
标准二阶矩模式	$\sqrt{\bar{P}^2 + \sigma^2}$	张益智	1983	[7]

* 该模式于 1979 年见报道,用于大气质量评价,1980 年引入水质评价。

$(P_i)_{\max}$ ——分指数中的最大值;

$\bar{P}$ ——分指数的算术均值;

W_i ——水质参数 i 的权值;

σ^2 ——分指数的方差;

A, B ——模糊矩阵。

这些模式,大体可分为数理统计模式、模糊数学模式和空间理论模式。其中数理统计模式占较大比重。

二、半集均方差模式的建立

按照数理统计的基本概念,用来评价水质的分指数集合,构成了容量为 n 的样本。而水质指数则可认为是反映水环境质量特性的样本数字特征,它主要应包括样本观测值的集中趋势及其离散程度两个方面。

算术均值(亦即一阶原点矩)是典型的也是基本的集中趋势测度。其定义式为:

$$\bar{X} = \frac{\sum_i X_i}{n}; i = 1, 2, \dots, n \quad (1)$$

式中, X_i ——样本观测值;

n ——观测值个数。

算术均值不仅是度量数据离散程度的有效基准,而且不管总体分布如何,取自它的样本的平均数总是近似地服从正态分布,样本容量越大,其分布与正态分布越接近。这就有可能依据正态分布的规律来推断总体。因此,算术均值是水质评价中的一个重要指标。

样本观测值的离散程度,最基本的度量是用标准差(或称均方差),它是数据能变性测度。这里取定义式为:

$$\sigma = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n}}; i = 1, 2, \dots, n \quad (2)$$

由标准差的定义式可以推论: X_i 中最大值的增大或最小值的减小,均可导致标准差值的增大。联系到水质评价,也就是把远大于算术均值的分指数和远小于算术均值的分指数置于等同地位,这显然与它们的环境

效应相矛盾,因为分指数中最大值对环境质量的影响是明显的,而最小值的影响较小甚至可忽略。因此,在水质评价中若简单地按定义式引入标准差,尽管可以说明分指数的离散程度,却未能真实地反映出各自对水质的影响而不能满足水质评价的需要。因此,有必要从能够反映分指数(代表水质参数)对水环境质量影响的特点出发,寻求满足于水质评价要求的数字特征。

为便于分析和讨论问题,根据大量的水质评价实例和实际分指数数据,经过“规范化”处理并筛选出 9 组数据,分别按其量值大小为序排列绘制直方图,可归纳为不同的分布类型。

在数理统计中,分布形态主要是以偏度判别。偏度的定义式为^[8]:

$$D_e = M_3 / (M_2)^{3/2} \quad (3)$$

式中, D_e ——偏度; M_3 ——样本三阶中心矩; M_2 ——样本二阶中心矩。且有:当 $D_e > 0$ 时为正偏分布(图 1); $D_e < 0$ 时为负偏分布(图 2); $D_e = 0$ 时为对称分布(图 3)。

从各组直方图可以看出,由分指数集合综合为水质指数以反映水体质量状况,首先表现在测度数据集中趋势的算术均值($\bar{P}$)的大小。它在直方图中由细实线所示的位置,则因数据的分布形态不同而异。对于正偏分布,算术均值大于或等于该分指数数列的中位数(P_{Mc}),即有 $\bar{P} \geq P_{Mc}$ 。但从整体看来, $\bar{P}$ 的位置多偏低。因为若以中位数为界,正偏分布中小于中位数一侧的数列分布较密集,算术均值主要受较密集的小值制约。对于负偏分布,其规律正相反,此时有 $\bar{P} \leq P_{Mc}$, 且 $\bar{P}$ 的位置多偏高,它主要受大于中位数一侧较密集的大值制约。对于对称分布,则有 $\bar{P} = P_{Mc}$, 且中位数两侧数列的分布具有映象对称的特点。

实际上,在大量的水质评价实践中,对于 $\bar{P} = P_{Mc}$ 、特别是对称分布形态出现的几率

十分小, 绝大多数的分指数集合属于 $\bar{P} \neq P_{Me}$ 的偏态分布。且从分指数对水环境质量的影

响来看, 不仅数集中的最大值具有最重要的作用, 次大值等的作用也不容忽视, 甚至具有与最大值类似的影响。因此, 除了用算术均值反映该数集的整体影响外, 还应当对大于

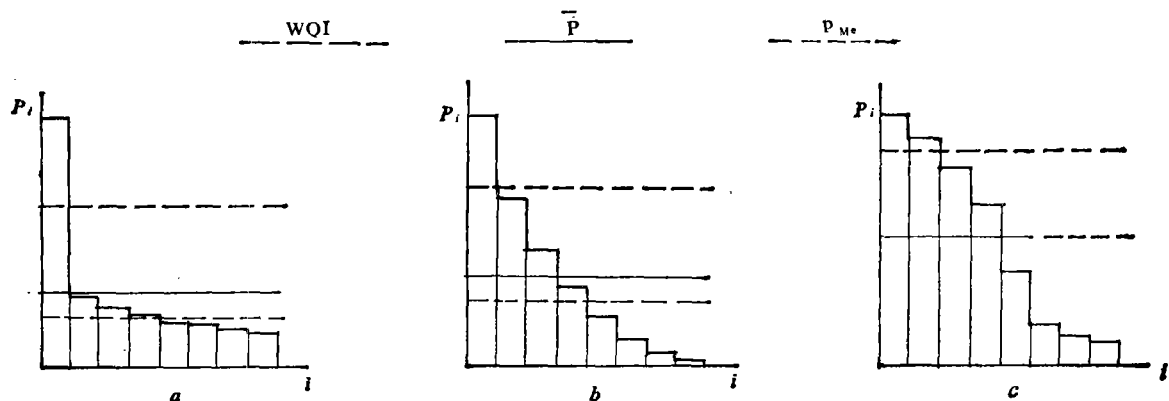


图1 正偏分布 $D_v > 0$

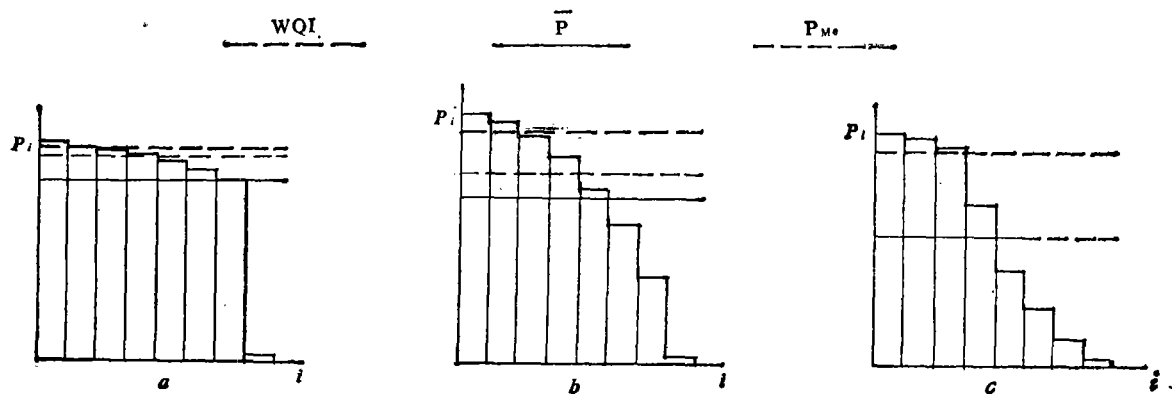


图2 负偏分布 $D_v < 0$

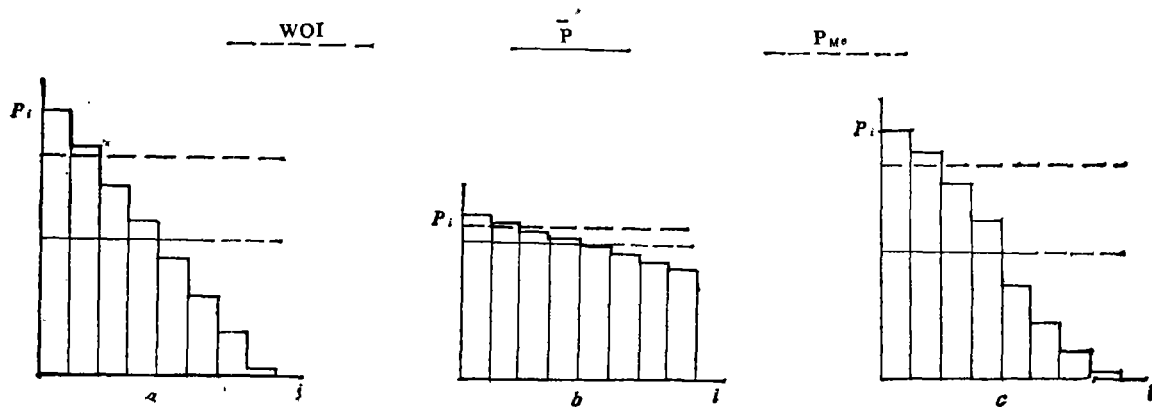


图3 对称分布 $D_v = 0$

中位数的分指数半集给予较大的权重。联系在不同的分布形态该半集数列的疏密程度不同,可通过计算它们对算术均值的均方差,以反映大于中位数半集的离散程度。据此引出一个新的数字特征,可将它定义为半集均方差,其定义式为:

$$S_h = \sqrt{\frac{\sum_{j=1}^m (P_j - \bar{P})^2}{m}} \quad (4)$$

式中 S_h ——半集均方差;

P_j ——大于中位数半集的分指数;

m ——大于中位数半集的分指数个数;

且有

$$m = \begin{cases} \frac{n}{2} & \text{当 } n \text{ 为偶数;} \\ \frac{n-1}{2} & \text{当 } n \text{ 为奇数.} \end{cases}$$

据上述分析,水质指数 WQI 作为反映水环境质量的数字特征,应在反映分指数集合集中趋势的基础上,用大于中位数半集的离散程度加以调整。据此可建立水质指数的数学模式为

$$WQI = \bar{P} + S_h \quad (5)$$

该模式可简称为半集均方差模式,其水质指数值在各组直方图中由粗虚线表示。由上述和图示可知,该模式既通过算术均值考虑了每项分指数对水环境质量的影响,也通过半集均方差对分指数中的大值给予较大的权重。据此,该模式是按各水质参数对水环境影响程度的不同,综合出水环境质量状况的数量描述。

三、综合模式的对比与讨论

为便于对比和讨论,现将对应于图 1 至图 3 的 9 组分指数值,及其分别按半集均方差模式 ($\bar{P} + S_h$) 和其它有关模式计算水质指数的结果列入表 3。表中第 1、2、3 组对应于图 1 中的 a、b、c,第 4、5、6 组对应于图 2 中的 a、b、c,第 7、8、9 组对应于图 3 中

的 a、b、c。在所列各模式中,分指数值与水质指数值均以 1 作为水环境是否污染的基本分级界限,大于 1 者标志已受到污染。且由表中 σ 、 S_h 的数据推断,建立半集均方差的概念并使之与标准差对比,可定性判别数集的分布形态。此时有: $S_h > \sigma$ 属正偏分布; $S_h < \sigma$ 属负偏分布; $S_h = \sigma$ 属对称分布。

由表中计算结果可知,多维空间向量模式(表中 E 列)的水质指数值,均大于各组分指数集合中的最大值。对于半集均方差偏小的负偏分布,此特征尤为明显(如第 4 组),这在实际上说明该模式未能真实地反映水质状况。

与均方根等价的标准二阶矩模式(表中 D 列)和几何指数模式(表中 C 列),其水质指数值均小于半集均方差模式,且随着标准差的减小渐趋于算术均值(如第 8 组)。因为它们也是分指数集合的某种平均状况,从而仅反映了样本观测值的集中趋势测度。用此指数值代表水环境质量的数字特征,显然不够完整。

内梅罗模式(表中 A 列)和几何均数模式(表中 B 列)采用最大值与算术均值再次平均(均方根与几何平均)的方法,均难作出合理的环境影响解释。以内梅罗模式与半集均方差模式对比: 当 $S_h < \sigma$ 时,说明大值较密集,次大值等的影响较明显,故而有后者水质指数大于前者;但当 $S_h > \sigma$ 时,大值较离散,次大值等的影响较小,却出现后者小于前者的异常状况(如第 1、2 组)。这正说明内梅罗模式由于单一强调最大值而带来的弱点。

双指标统计模式(表中 $\bar{P}$ 与 σ^2) 在评价中的主要问题是二次综合的不确定性。如根据表中第 4、5 组分指数的 $\bar{P}$ 、 σ^2 数据(2.463、0.821)和(2.275、1.164),很难判定那一组的水质污染更重些。但按半集均方差模式,第 5 组(3.18)明显重于第 4 组(2.93)。分析相应的分指数集合数据构成,此结论是合理的。

表 3 不同数学模式的水质指数对比

序号	分 指 数	$\bar{P}$	P_{Me}	σ^2	σ	S_h	WQI					
							$\bar{P}+S_h$	A	B	C	D	E
1	3.4、1.0、0.85、0.75、0.65、0.6、0.55、0.5	1.0375	0.7	0.821	0.906	1.194	2.23	2.51	1.88	0.83	1.38	3.90
2	3.4、2.3、1.6、1.1、0.7、0.4、0.2、0.1	1.225	0.9	1.164	1.079	1.229	2.45	2.56	2.04	0.73	1.63	4.62
3	3.4、3.1、2.7、2.2、1.3、0.55、0.4、0.35	1.75	1.75	1.385	1.177	1.188	2.94	2.70	2.44	1.26	2.11	5.96
4	3.0、2.95、2.9、2.85、2.75、2.65、2.5、0.1	2.4625	2.8	0.821	0.906	0.466	2.93	2.74	2.72	1.84	2.62	7.42
5	3.4、3.3、3.1、2.8、2.4、1.9、1.2、0.1	2.275	2.6	1.164	1.079	0.904	3.18	2.89	2.78	1.64	2.52	7.12
6	3.15、3.1、2.95、2.2、1.3、0.8、0.4、0.1	1.75	1.75	1.385	1.177	1.165	2.91	2.55	2.35	1.13	2.11	5.96
7	3.6、3.1、2.6、2.1、1.6、1.1、0.6、0.1	1.85	1.85	1.313	1.146	1.146	3.00	2.86	2.58	1.26	2.18	6.15
8	2.2、2.1、2.0、1.9、1.8、1.7、1.6、1.5	1.85	1.85	0.052	0.229	0.229	2.07	2.03	2.02	1.84	1.86	5.27
9	3.4、3.1、2.7、2.2、1.3、0.8、0.4、0.1	1.75	1.75	1.412	1.188	1.188	2.94	2.70	2.44	1.13	2.12	5.98

A-Nemerow 模式^[2]; B-姚志麒模式^[3]; C-几何指数模式^[1]; D-张益智模式^[7]; E-申葆诚模式^[4]。

表 4 不同数学模式的水质指数对比

序号	分 指 数	$\bar{P}$	P_{Me}	σ^2	σ	S_h	WQI					
							$\bar{P}+S_h$	A	B	C	D	E
1	3.46、3.28、3.09、3.08、2.11、 2.06、0.84、0.26	2.273	2.595	1.240	1.114	0.968	3.24	2.93	2.80	1.78	2.53	7.16
2	3.37、2.60、2.37、2.16、2.11、 2.05、1.77、0.93	2.17	2.135	0.432	0.650	0.645	2.82	2.83	2.70	2.06	2.26	6.41
3	2.54、2.35、2.25、2.01、1.26、 1.03、0.73、0.70	1.609	1.635	0.505	0.711	0.705	2.31	2.13	2.02	1.43	1.76	4.97
4	3.04、2.40、2.12、1.81、1.50、 0.69、0.47、0.12	1.519	1.655	0.910	0.954	0.940	2.46	2.40	2.15	1.06	1.79	5.07
5	2.96、1.66、1.56、1.49、1.18、 1.15、1.01、0.69	1.463	1.335	0.409	0.639	0.757	2.22	2.33	2.08	1.35	1.60	4.51
6	2.52、2.28、2.21、1.84、1.38、 0.85、0.09、0.02	1.399	1.61	0.850	0.922	0.850	2.25	2.04	1.88	0.69	1.68	4.74
7	1.70、1.55、1.30、1.10、0.63、 0.51、0.31、0.06	0.895	0.865	0.318	0.564	0.566	1.46	1.36	1.23	0.62	1.06	2.99
8	1.05、0.96、0.92、0.80、0.45、 0.35、0.21、0.19	0.616	0.625	0.110	0.331	0.329	0.94	0.86	0.80	0.51	0.70	1.98
9	1.04、0.82、0.72、0.61、0.50、 0.22、0.14、0.02	0.509	0.555	0.112	0.334	0.329	0.84	0.82	0.73	0.32	0.61	1.72

(下转第39页)

的比值在人为输送量中为 0.12, 在自然背景输送量中为 0.09。可见人为加入的影响加大了可溶态的数量。

3. 洛阳市进入伊洛河的污染金属量占伊洛河输送总量的 24.6%, 7 种元素所占的该元素总量的百分比顺序为: Cr (91%) > Zn (31.4%) > Cd (26.9%) > Co (22.9%) > Cu (16.2%) > Pb (6.7%) > Ni (5.6%)。这和伊洛河的污染状况基本相同。

三、几点看法

1. 计算通量是一项长期的艰巨任务, 瞬间取样的方法常会引入较大的误差, 今后在

(上接第 59 页)

为进一步验证上述对比与讨论, 表 4 中也列出 9 组水质评价的实际分指数数据, 计算内容全部与表 3 同。由该表可知, 其结果无一例外地重现了上述对比与讨论的规律和结论。

根据上述分析、对比和讨论, 建立水质指数的综合数学模式, 不仅要求在数学方法上准确、严谨、使用方便, 更要求其环境效应合理。从数理统计分析、实际环境影响以及计算结果可知, 半集均方差模式基本上满足了上述要求, 与所讨论的其它模式比较更为合理。通过水质评价实践也证明了它能更确切地反映水质状况。但由于我国地域辽阔, 地面水质状况差异甚大, 该模式还需通过更

有条件的情况下应采用连续取样或多次取样及同一水团的追踪取样等多种方法来进行。

2. 为了全面掌握伊洛河向黄河输送量的情况, 应对巩县的废水排放进行监测, 可在巩县入伊洛河前的排污渠上的断面进行监测。

3. 在伊洛河入黄河口的黄河干流上应设上下两个对照断面, 以便研究金属总量进入黄河后的变化。

参 考 文 献

- [1] 张书权等, 第二次河流泥沙国际学术讨论会论文集, 139—147 页, 水利电力出版社, 1983 年 10 月。
- [2] 王晓蓉等, 环境化学, 1(2), 140(1982)。

广泛的应用加以多方面的验证, 以求更加完善。

参 考 文 献

- [1] Bolton, P. W., *Water Pollut. Control*, 77 (2), 271 (1978).
- [2] Nemerow, N. I., *Scientific Stream Pollution Analysis*, Scripta Book Co, 210—231, New York, 1974.
- [3] 姚志麒, 环境科学, 1(6), 53(1980).
- [4] 熊广政、裴小松, 环境科学, 1(4), 53(1980).
- [5] 申葆诚、徐国义, 环境科学, 2(4), 57(1981).
- [6] 容跃、金键, 环境科学, 3(2), 69(1982).
- [7] 张益智, 环境科学, 4(6), 80(1983).
- [8] 高玉堂, 环境监测常用统计方法, 14—15, 原子能出版社, 1980 年。