

# 大骨节病区环境微量元素多元统计分析

白 迺 彬

(中国科学院生态环境研究中心)

我们应用自行开发的模式识别为主的多元统计分析程序包处理了典型大骨节病区环境微量元素数据<sup>[1,2]</sup>。程序包包括聚类分析、主成份分析、因子分析、模式识别与图相显示等十四个可用程序,用 FORTRAN-77 语言编写,在 VAX-11/780 机上实现。

## 一、硒与其它微量元素聚集特点

我们应用聚类分析方法<sup>[3]</sup>对水、土、粮、人发中各微量元素间相关系数矩阵实施连接,形成谱系,示为聚集生成树(图 1)。相似者相聚,相关系数是相似性的良好指标。相关系数较大者彼此聚集,形成一类,类内诸元素处于聚集状态;相关系数较小者彼此独立,形成非类,类间诸元素处于彼此独立状态。

从图 1 可见:

1. 水、土、粮各微量元素聚集行为大体类似,即难迁移元素如铁、锰、铜、铅等相聚集。而易迁移元素硒、钼等处于较为独立状态。这反映了该病区特定的黄土高原占土壤层富

含铁、锰,而又淋溶现象严重的实际情况。

2. 人发数据中铁、锰等聚集性与硒、钼等独立性均趋于均衡,说明人体与所处病区自然环境存在差异,‘水土不服’,在一定程度上表达了病因。

3. 饮用水中腐植酸与镉、铁聚类,硒不属于该类,处于较为独立状态。这点暗示:硒与腐植酸可能存在颉颃关系,为进一步实验工作提供有益的线索。

## 二、硒对病情影响的重要性

对上述相关系数矩阵进行病情加权改造,再用主成份分析法<sup>[4]</sup>压缩影响病情的因子数目,找出诸因子中影响较大者,继而应用因子分析法<sup>[5]</sup>,将全部因子负荷矩阵做方差极大正交旋转变换<sup>[6]</sup>,得到正交旋转因子负荷矩阵,以求利于得到解释。由主成份分析找出的对病情贡献较大的前几个主成份列于表 1。

表 1 主成份分析结果数值表

样 本	土		水			玉米		小麦		人发
	1	2	1	2	3	1	2	1	2	1
$\lambda_i$	0.1720	0.0966	0.1618	0.0371	0.0314	0.1835	0.1124	0.3602	0.1580	0.2426
$\Sigma\lambda_i$	0.3897		0.3406			0.4288		0.6411		0.3631
$P_j$	0.441	0.248	0.457	0.109	0.092	0.428	0.262	0.562	0.427	0.668
$P$	0.689		0.658			0.690		0.809		0.668
元 素	硒	铁	镉	腐植酸	硒	钼	硒	硒	锰	硒

表 1 中,  $G$  个主成份中第  $i$  个主成份特征值为  $\lambda_i$ , 第  $i$  个主成份病情贡献率  $P_i$  为:

$$P_i = \frac{\lambda_i}{\sum_{i=1}^G \lambda_i};$$

而前  $n$  个主成份对病情贡献率总和  $P$  为:

$$P = \sum_{i=1}^n P_i.$$

表 1 所列诸主成份之因子负荷矩阵见图 2 上半部; 变换后的方差极大正交旋转负荷矩阵见图 2 下半部. 对比之下可明显找到诸主成份所代表的微量元素. 它们亦列入表 1, 可见:

1. 所列前几个主成份特征值表明, 它们对病情的总贡献率约达 70%, 说明这些主成份基本上控制了病情. 所有表 1 列出的主成份所代表的微量元素皆离不开硒. 硒对病情

影响有相当重要的作用.

2. 人发硒的水平较其它样本硒水平对病情制约更甚, 光硒一个因素对病情贡献率已近 70%. 一般来说, 人发微量元素水平能反映其人健康水平, 特别是病区, 硒低, 患病. 就大骨节病而言, 其它元素的影响可能不大. 这一看法并不排斥某些有机毒物的致病作用.

### 三、食物链中各环节硒的作用

生态环境为统一自然综合体. 土壤元素分布影响饮用水中各微量元素含量, 它们又制约粮食的微量元素组成, 最终反映到人发, 表现为病或非病. 为了研究食物链中各环节硒的病因作用, 我们应用线性判别分析<sup>[7]</sup>, 处理了取自文献 [1] 和 [2] 的表 2 数据, 找出分类正确的最小数目变量. 若包括人发数据, 则对病情影响最大因素  $Z$  为:

$$Z = 0.002X_{水} + 0.320X_{土} - 0.033X_{玉米} + 0.185X_{小麦} + 0.929X_{人发} \approx 0.929X_{人发};$$

若不包括人发数据, 则

$$Z = 0.723X_{水} + 0.135X_{土} + 0.680X_{玉米} + 0.719X_{小麦} \approx 0.723X_{水} + 0.680X_{玉米} + 0.719X_{小麦}.$$

这说明: 处于食物链终端的人体营养水平硒的缺乏对病情影响是直接的, 最终的. 如果通过服药等人为途径, 补充人体中所需的硒,

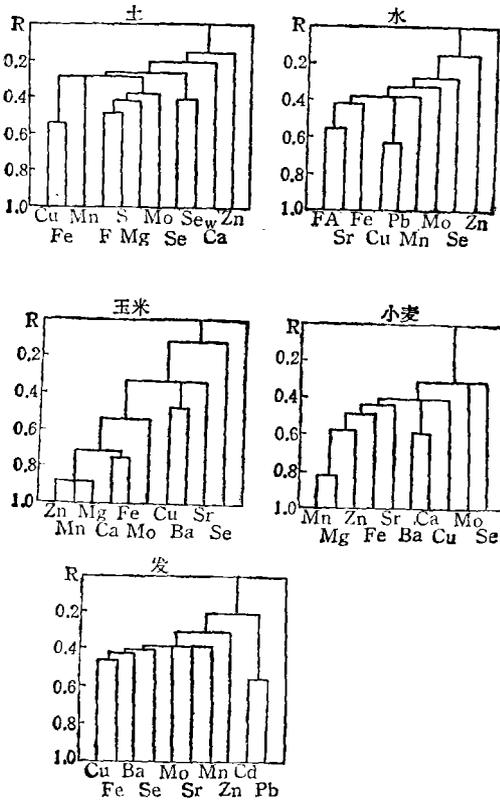


图 1 土、水、玉米、小麦、人发所含微量元素聚类生成树

表 2 水、土、粮、发所含微量元素与病情相关系数

分类	元素	样本				
		饮用水	土壤	玉米	小麦	人发
1	硒	0.1875	0.3880	0.3365	0.5627	0.4871
2	钼	0.0798	0.0480	0.3831	0.0297	0.1371
2	铜	0.1578	0.0930	0.0476	0.0785	0.0980
2	锌	0.1382	0.2110	0.1584	0.2964	0.0079
2	铁	0.2214	0.2970	0.1569	0.1575	0.0324
2	锰	0.0735	0.0340	0.1730	0.3006	0.0949

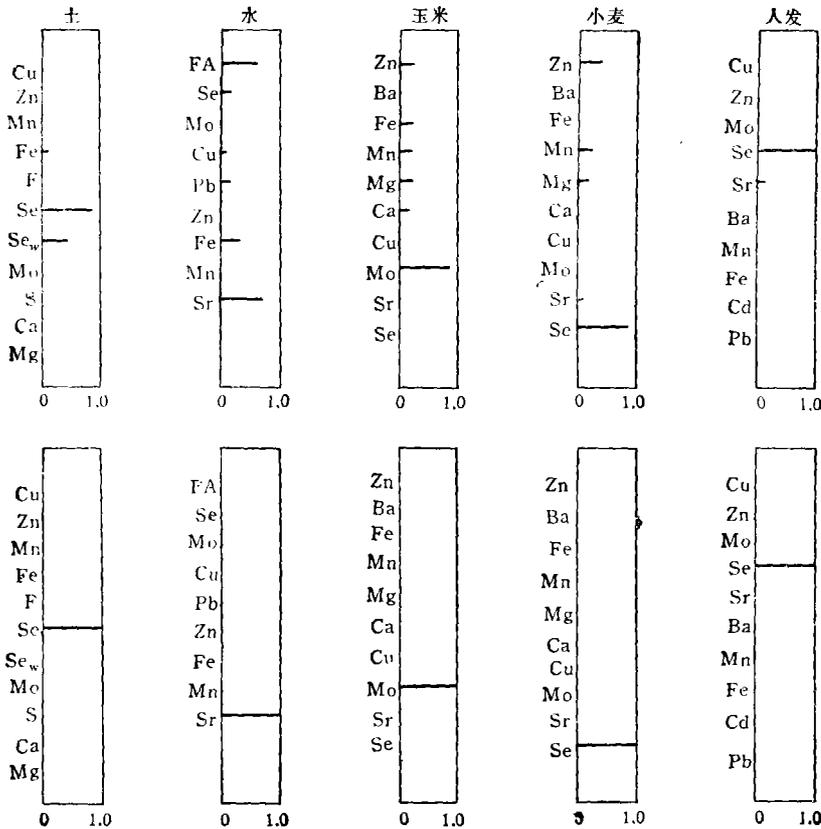


图 2 土、水、玉米、小麦、人发样本影响病情的第一主成份因子负荷  
 上半部：方差极大正交旋转变换前因子负荷  
 下半部：方差极大正交旋转变换后因子负荷

完全可以控制病情，达到防治目的。另一方面，凡是人口的各个主要途径，包括饮用水、玉米、小麦，似乎均对病情有一定的影响。环境缺硒问题应做总体考虑，改良土，增加土壤含硒量可以做为综合防治措施之一。

#### 四、硒与大骨节病病情预报

如前所述，就人发数据而言，硒的水平相当满意地反映患者的病情，其它微量元素所起的作用均不明显。因此，我们觉得以人发硒的水平做为大骨节病区病情判断的重要先期指标，比起其它指标，目前理由较为充分，且简单实用。这里并不排斥某些有机毒物的致病作用，它可能在体内经过种种代谢步骤

造成体内硒水平的若干降低。这一观点，仅取自于某一典型病区数据，尚需做更大范围适用性证明。为此，我们收集了已发表的四川、陕西、东北等各主要大骨节病区的人发数据<sup>[1,8,9]</sup>，列于表 3。

我们应用模式识别法处理表 3 数据<sup>[10]</sup>，即首先找出它们的第一、二主成份  $Z_1$  和  $Z_2$ ，其中，第一主成份  $Z_1$  对病情贡献已达到 91%；而第二主成份  $Z_2$  对病情贡献为 5%，其余不足 4%。

$$\begin{aligned}
 Z_1 = & 0.9954 \times 10X_{Se} + 0.0059 \times 0.1X_{Mn} \\
 & + 0.0925 \times 0.01X_{Fe} + 0.0125 \\
 & \times 0.1X_{Cu} + 0.0203 \times 10X_{Mo} \\
 & - 0.0093 \times 0.01X_{Zn}
 \end{aligned}$$

表 3 全国大骨节病区、非病区人发数据

元素单位 类	锌 ( $\times 0.01$ )ppm	钼 ( $\times 10$ )ppm	铜 ( $\times 0.1$ )ppm	铁 ( $\times 0.01$ )ppm	锰 ( $\times 0.1$ )ppm	硒 ( $\times 10$ )ppm
病 区	1.50	1.14	0.78	0.44	0.46	0.76
	1.48	1.10	0.88	0.80	0.53	0.73
	1.59	1.20	0.74	0.28	0.44	0.63
	1.39	1.20	0.71	0.20	0.27	0.92
	1.46	0.71	1.03	0.48	0.82	0.63
	1.44	0.44	0.94	0.17	0.50	0.69
	1.50	0.46	0.89	0.18	0.54	0.70
	1.38	0.48	0.73	0.13	0.48	0.55
	1.51	0.71	0.85	0.43	1.09	0.74
	1.49	0.79	0.81	0.43	1.08	0.54
	1.51	0.59	0.82	0.28	1.61	0.58
	1.53	0.65	0.84	0.31	0.90	0.87
非 病 区	1.57	1.21	0.72	0.48	0.74	1.33
	1.56	1.10	0.71	1.10	0.86	1.90
	1.78	1.20	0.82	0.31	0.86	1.09
	1.46	1.50	0.64	0.18	0.48	1.30
	1.31	0.52	0.99	0.37	0.57	1.96
	1.43	0.51	0.86	0.15	0.47	0.88
	1.48	0.52	1.00	0.59	0.85	1.48
	1.48	0.51	0.99	0.57	0.81	1.44

$$\approx 9.954X_{Se}$$

可看出：就是硒的水平对病情有举足轻重的影响。然后，用非线性映照法将多维数据映照到  $y_1$  和  $y_2$  所张成的平面，按病与非病对本点进行分类和识别，见图 3。实际计算中， $y_1$  和  $y_2$  仅与  $Z_1$  和  $Z_2$  有个别特异点调整，所以， $y_1$  和  $y_2$  所隐含代表之元素，就是  $Z_1$  和  $Z_2$  所代表之元素。

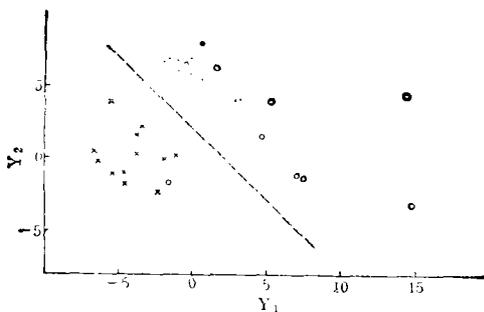


图 3 大骨节病区、非病区人发数据分类图  
× 大骨节病区 △ 非病区

令  $R = 1 - (N'/N)$  为正确分类率，其中  $N$  为样本总数， $N'$  为错分样本数。  $R > 0.80$  时，满足统计预报标准。图 3 可见，沿第一主成份方向变化较大，它正代表硒起作用。病区与非病区正确分类率达 0.95。可以说，人体硒的营养水平对大骨节病病情影响的权重足以判断病情，可以做为先期指标。实践证明：在大骨节病区增加人群的营养水平硒，对控制大骨节病的区域流行和防治确有实效<sup>[9,11,12]</sup>，这与我们的数学处理结果大致相符。

本工作表明：我们提出的处理大骨节病区环境微量元素数据的多元统计分析系统，可以对硒在致病中所起的重要作用，做出定量研究。

#### 参 考 文 献

- [1] 中共中央地方病防治领导小组办公室，永寿大骨节病科学考察文集（1979—1982），80 页、86 页、89

页、92 页、101 页，人民卫生出版社，1984 年。

[ 2 ] 永寿大骨节病科学考察队环境专题组，环境科学学报，5(1)，1(1985)。

[ 3 ] Anderberg, M. R., Cluster Analysis for Application, Academic Press, New York, 1973.

[ 4 ] Morrison, D. E., Multivariate Statistical Methods, McGraw-Hill, New York, 1967.

[ 5 ] Horst, P., Factor Analysis of Data Matrices, Holt, Rinehart and Winston, Inc., New York, 1965.

[ 6 ] Kaiser, H. F., Psychometrika, 23, 187(1958).

[ 7 ] Green, P. E. et al., Mathematical Tools for Applied Multivariate Analysis, Academic Press, New York, 1976.

[ 8 ] 盛士骏等，中国环境科学，2(6)，46(1982)。

[ 9 ] 李继云等，环境科学，2(5)，18(1981)。

[ 10 ] Kowalski, B. R., Chemometrics: Theory and Application, American Chemical Society, Washington, D. C., 1977.

[ 11 ] 闻芝梅等，中国地方病学杂志，4(1)，73(1985)。

[ 12 ] 王凡等，中国地方病学杂志 1(1)，35(1982)。

# 用灵敏度分析污水稳定塘设计裕量

许晓鸣 潘南鹏 曹维勤 朱新源 段振渤

(北京市环境保护科学研究所)

迄今为止，污水稳定塘的各种设计方法都假定所依据的参数和边界条件是稳定不变的。而在实际情况中，不论是水质、水量还是环境因素和运行条件，都有极大的可变性。为了保证出水水质的相对稳定，有必要在设计计算结果中加入一定的裕量，即安全量。但目前尚无求解污水稳定塘设计裕量的合理方法。本文分析了主要的变化因素对稳定塘设计结果的影响。在此基础上，运用过程系统工程中的灵敏度分析<sup>[1]</sup>，提出了求稳定塘设计裕量的方法。

## 一、影响处理效果的不确定因素及其变化范围

污水稳定塘净化效果受环境因素、运行情况、水质条件等的综合影响。这些影响有的较为确定，如温度，一般已在设计中加以考虑。有的影响因素却是不确定的。各种不确定因素从两方面导致设计结果与实际情况相偏离。

1. 引起设计参数的变化。污水稳定塘最重要的设计参数是 BOD<sub>5</sub> 一级降解速率常数  $K$  和无量纲扩散系数  $d$ 。 $K$  值一般由动态试验得出。它主要取决于污水的降解性质，同时还受取得实验数据的试验方法和数据统计

方法的影响。这当中任一环节的不确定性都能导致  $K$  值的误差。故通常设计中所用的  $K$  值实际上只是一个具有较大变化范围的统计平均值。其变化范围因系统而异。对美国 Corinne 市稳定塘运行数据<sup>[2]</sup>进行统计，其结果表明，当置信度为 0.1 时， $K$  的置信区间宽度为  $\pm 0.03$ 。

$d$  值综合表征稳定塘内流动与混合情况。由于设计先于建塘，故  $d$  值一般根据经验公式估值。 $d$  值除受估值误差影响外，还受到稳定塘的几何形状、进出口位置、风的搅动、流量变化等因素的影响。用 Corinne 市稳定塘实测的数据<sup>[2]</sup>分析，当置信度为 0.1 时， $d$  的置信区间宽度为  $\pm 0.356$ 。

2. 引起设计边界条件的变化。在实际运行中，进水水质和水量都是高度可变的。但根据现场实测数据，可求得二者的均值置信区间。水质的变化可影响  $K$  值，水量的变化则影响停留时间。

综合以上的分析可知，不确定因素对处理效果的影响主要体现在对速率常数  $K$ 、扩散系数  $d$ 、进水浓度  $c_0$  和水量  $Q$  的影响上。这些因素的变化范围可取其均值的置信区间宽度。