

土壤调查数据地域统计的最佳估值研究*

——彰武县表层土全氮量的半方差图和 块状 Kriging 估值

徐吉炎

R. 韦甫斯特

(中国科学院林业土壤研究所)

(英国洛桑试验站)

近年来国际土壤应用数学和土壤信息系统的研究趋势之一,是引用地质统计(Geostatistics)的最佳估值法处理土壤调查的数据。地质统计是目前地质部门在探矿与采矿应用的一种先进空间分析方法,它的要点是根据地面不同选点钻井所获得的不同深度的数据资料、寻求数据信息与采样点的位置和采样深度的统计相关性来对矿产进行空间结构分析与数量估计。显然,探矿与采矿中应用的地质统计是属于三维的空间统计分析。这个方法首先是由 Matheron (1965, 1971)^[6,7] 建立起来的。Journel 等^[5] (1978) 曾对此法作了全面的论述。其后 Burges 等 (1980a, 1980b)^[8,9] 以及 McBratney 等^[3] (1981) 引用此法于土壤调查。由于土壤调查,通常是在地表上进行,因此是属于二维的空间分析,可称之为“地域统计分析”。本文主要介绍辽宁省彰武县的表层土壤全氮量进行地域统计空间分析的一些新结果,并描述了土壤全氮量的数学转换以及如何从不规则采样点计算位差(Lag)的方法。

一、彰武县的情况

彰武县位于我国东北辽宁省西北部,全县面积 3,635 平方公里,年平均气温 7.2℃,无霜期 156 天,年降水量 510 毫米,其中 5—9 月降雨量为 436 毫米,约占全年降水量的 85%。林业土壤研究所曾对该县进行了土壤普查,全县共采集了 102 个土壤剖面,并对不同土壤层次进行了物理与化学分析测定。本文应用的就是这些普查与测定的数据。该县中部与南部属于辽河平原,海拔基本在 50 米以下,地势平坦,组成物质主要是沙质粘土和黄土型的沉积物。土壤以草甸土为主。北部地区属于科尔沁沙地的一部分。此区内广布沙丘,土壤大部分为半流动风沙土。东部土壤主要是棕色森林土,而西部则多为褐土。除这四种主要土壤外,还有少量的盐碱土。图 1 为彰武县主要土壤分布图。

国外应用于地域统计分析的土壤调查数据通常是等间距采样,并且采样点的密度比较大。而这里所用彰武县土壤数据在地域分布上是不规则的,土壤样品总数不多、密度较

* 本文系徐吉炎在英国洛桑试验站留学期间,与 R. Webster 博士合作对我国辽宁省彰武县的土壤调查数据进行地域统计分析研究成果的一部分。本所盛士骏、张国枢同志为我们整理了彰武县土壤调查资料, Joyce Munden 夫人为本文准备了出版所用的插图,在此一并致谢。

稀。我们采用全方位,即考虑每个土壤采样点的邻近地域所有方位的土壤样品点观测值,进行地域统计的空间分析。结果表明,应用这些数据对全方位计算其半方差,虽然它的方差基底值(Nugget)是较大,但在50公里范围内,仍然显示出所有观测值的半方差有明显的空间相关性,因此应用它进行地域统计分析和 Kriging 最佳估值是可以实现的。

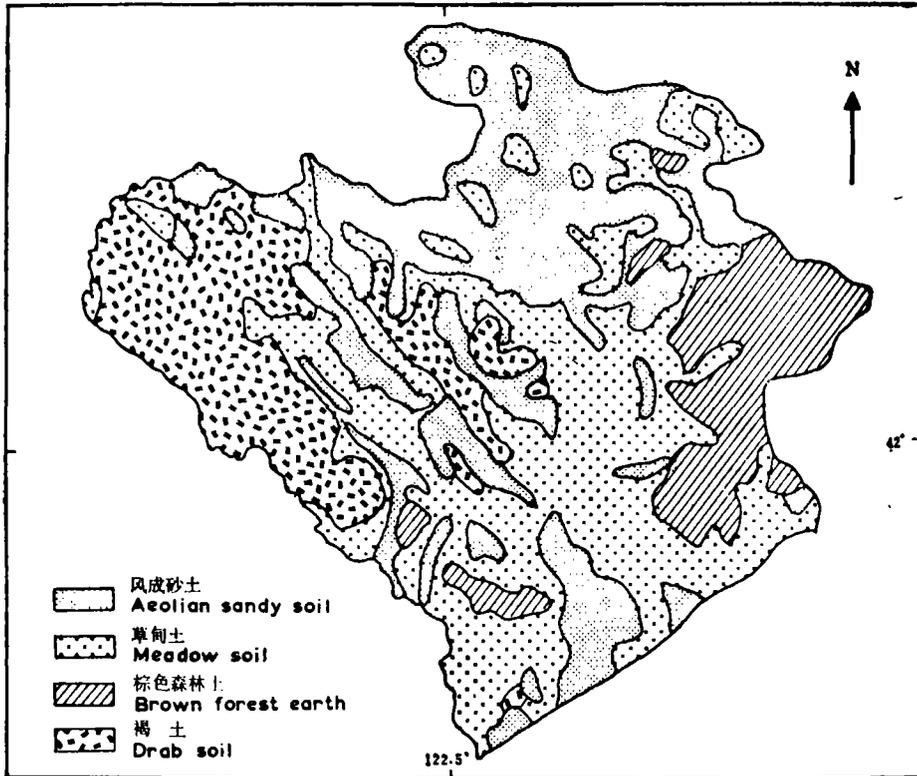


图1 彰武县主要土壤分布图

Fig. 1 Soil map of Zhangwu county

二、初步统计分析

在对土壤调查数据进行地域方差分析时,首先需要进行初步统计分析。Webster^[10](1977)系统介绍了土壤调查数据的初步统计处理方法。我们用它对彰武县的土壤调查数据进行初步统计分析表明,pH有足够好的正态分布,可直接作为地域变量应用于地域统计分析;而全氮的数据与有机质情况相同,需经对数变换,即

$$z_i = \log_{10}(P_i) \quad (1)$$

式中 P_i 为全氮量观测值, z_i 为经对数变换后的地域变量。图2所示为未经变换处理与以10为底的对数变换后全氮量的频数分布直方图。表1为其频数分布表。表2为其频数分布的统计特征参数。

表 1 全氮量变换前后频数分布比较

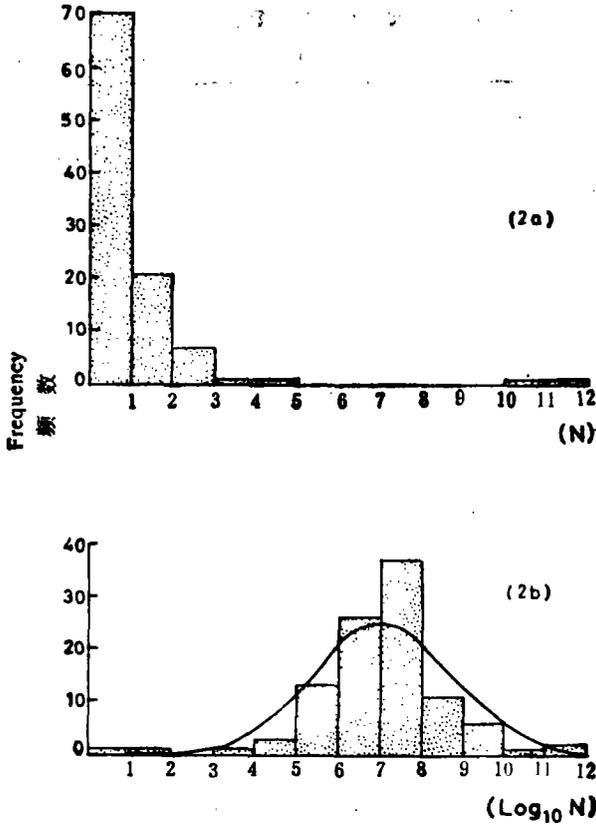
Table 1 Frequent distribution of topsoil total nitrogen before and after the transformation

组序 No.	变换前 Before the transformation			对数变换后 After the transformation		
	组界 Limit of unit	频数 Frequency	按正态分布的期望频率(%) Frequency calculated by normal distribution	组界 Limit of unit	频数 Frequency	按正态分布的期望频率(%) Frequency calculated by normal distribution
1	0.001—0.080	70	23.26	-3.000—-2.752	1	0.01
2	0.080—0.159	21	24.04	-2.752—-2.503	1	0.09
3	0.159—0.238	7	17.13	-2.503—-2.255	0	0.57
4	0.238—0.318	1	8.42	-2.255—-2.006	1	2.54
5	0.318—0.397	1	2.85	-2.006—-1.758	3	7.34
6	0.397—0.476	0	0.67	-1.758—-1.510	13	15.87
7	0.476—0.555	0	0.11	-1.510—-1.261	26	22.77
8	0.555—0.634	0	0.01	-1.261—-1.013	37	23.04
9	0.634—0.713	0	0.00	-1.013—-0.764	11	16.25
10	0.713—0.793	0	0.00	-0.764—-0.516	6	8.07
11	0.793—0.872	1	0.00	-0.516—-0.268	1	2.74
12	0.872—0.951	1	0.00	-0.268—-0.019	2	0.64
合计	0.001—0.951	102	76.49	-3.000—-0.019	102	49.43

表 2 全氮量变换前后的统计特征参数

Table 2 Summary of statistics for topsoil total nitrogen before and after the transformation

	变换前 Before the transformation	对数变换后 After the transformation
平均值 Mean	0.0872	-1.2530
标准差 Standard deviation	0.1299	0.4061
变差系数 Coefficient of variation	1.490	-0.3241
方差 Variance	0.0169	0.1649



2a 土壤样品全氮量的频数分布直方图

2a Before the transformation

2b 土壤样品全氮量对数变换后的频数分布直方图

2b After the transformation of common logarithm

图2 表层土全氮量的频数分布直方图

Fig. 2 Frequency histograms of total nitrogen in topsoil

三、半方差图 (Semi-Variogram)

半方差图是进行地域统计分析的一个基本性步骤。它是确定可否进行地域统计处理和最佳地域估值的前提,它也是在进行最佳估值中寻求理想的似然数学模型来拟合地域变量对于位差分布的依据。这种半方差图表示了土壤地域变量对于在距离与方位上不同的所有成对点之间的观测值的空间相关性。根据定义,地域变量的半方差为:

$$\gamma(h) = \frac{1}{2} \text{var}[z(x) - z(x+h)] \quad (2)$$

这里, $z(x)$ 和 $z(x+h)$ 是所研究土壤性质的地域变量 z 在由位差矢量 h 表示的每两个点 x 和 $x+h$ 位置的观测值。Webster 和 Burgess (1980)^[1,2] 指出,对于大多数实际土壤调查目的而言,在一定区域内土壤性质可以认为是趋向于稳定不变的,即它的一阶中心距为零。

$$E\{z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})\} = 0 \quad (3)$$

于是，地域变量的半方差就可用所有二点地域变量之差的平方的数学期望值的一半来表示。

$$\gamma(\mathbf{h}) = \frac{1}{2} E[\{z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})\}^2] \quad (4)$$

至此半方差就变成仅依赖于每两个点之间的距离与方位，而不依赖于这两点本身所在的位置 \mathbf{x} 和 $\mathbf{x} + \mathbf{h}$ 。这样可根据样品观测值来计算地域变量的半方差。在实际应用中，用若干成对样品地域变量之间的平均半方差作为相应总体半方差的估计值。估计的半方差为：

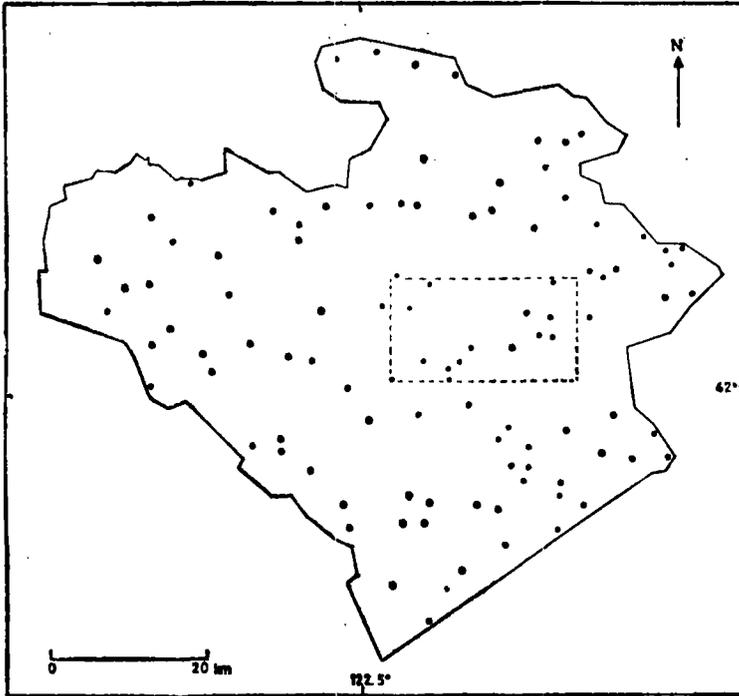


图 3 彰武县土壤调查取样点的分布图

Fig. 3 Map showing the positions of the sampling sites

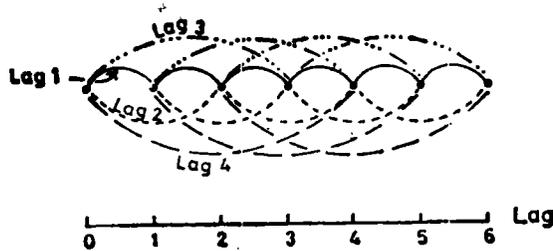


图 4 等间距样品的位差计算示意图

Fig. 4 Method computing lag of isometric samples

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z(x_i) - z(x_i + h)\}^2 \quad (5)$$

这里 $N(h)$ 是位差为 h 的观测值的样品成对数目。通常表示距离与方位的矢量记为 h ，称为位差 (Lag)。目前在许多国家中，土壤调查的选样点的间隔与分布是规则的，样品间距也比较小，计算半方差估计值时所取的位差通常取样品最小间距的整数倍。如图 3 所示。并可以对不同方位的样品地域变量逐个方位计算半方差。考虑本文所采用的彰武县土壤调查的地域变量，不仅在位置分布上不是等间距的 (图 4)，同时取样点密度较稀，平均间距较大，不宜对不同方位逐个计算半方差，而取全方位计算所有方向的半方差，这样其位差矢量 h 可简化为位差标量

$$h_i = |h_i|$$

这里 h_i 表示全部样品总体的不同间距的平均位差。图 5a, b 是不规则地域变量计算平均位差的示意图。

我们选用 2 公里作为计算位差的平均间距，对每一个样品点周围在此平均间距内样品点的所有位差取平均值。然后以相同平均间距对全部样品点取总平均值作为总体的位

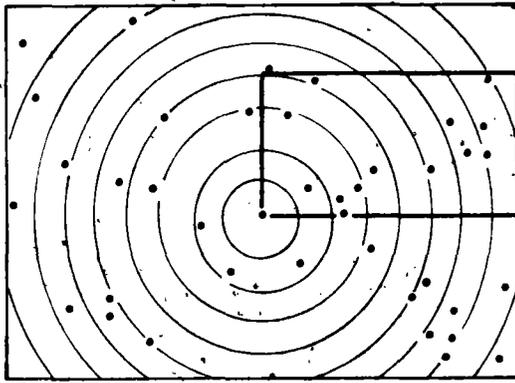


图 5a 确定不规则样品点的均等距

Fig. 5a Definition of the average isometric for irregular samples

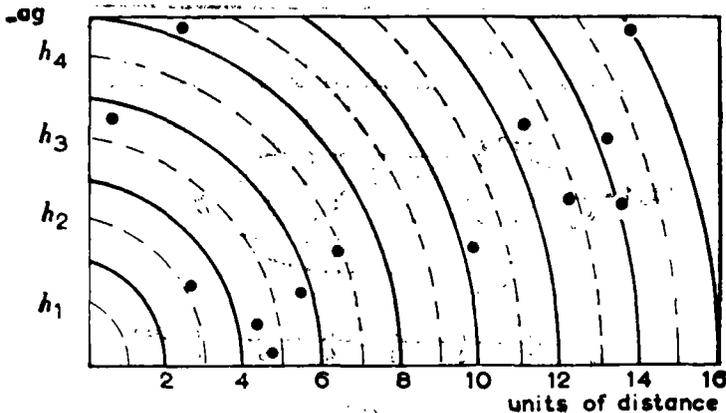


图 5b 对不同位差间距的样品点求平均位差

Fig. 5b Computing average lag of irregular samples at different average isometric

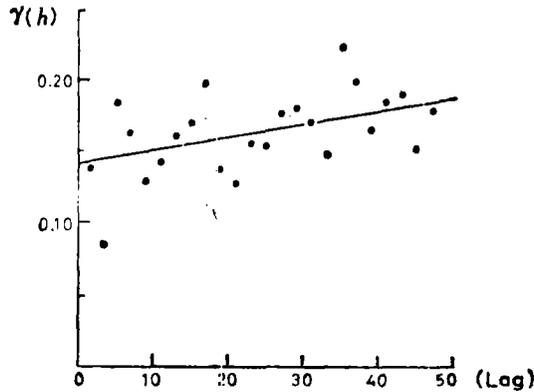


图 6 全氮量地域变量的半方差图

Fig. 6 Semi-variogram of total nitrogen in topsoil as the variable

差 h 。图 6 为彰武县表层土全氮量经对数变换后 50 公里范围内的半方差图。图中黑点是由 (4) 式计算出的对应不同位差的半方差值。

四、拟合半方差图的数学模型

在用地域统计进行最佳估值时，需要建立一个描述所研究的地域变量特征的空间结构函数模型，即寻找一个统计误差最小的半方差对位差的函数模型来表征样品地域变量半方差图。它反映了在给定期域内总体地域变量半方差 γ 对于位差 h 的分布。我们应用 Ross 等 (1980)^[8]“最大似然程序” (Maximum likelihood Program) 完成了这一工作，在用此程序的计算过程中考虑到了不同位差间隔所具有的样品成对数目上的差异，并以样品成对数目作为统计权重。对于全氮量地域变量，最小误差的拟合模型为线性函数

$$\bar{\gamma}(h) = 0.1409 + 0.000939h \quad (6)$$

图 6 中的直线就是对样品点平均间隔为 2 公里分布的半方差按最小二乘法拟合的此线性函数。

五、块状 Kriging 最佳估值

Kriging 技术提供了一个具有最小方差对任何给定的几何正方区间的平均特征进行无偏估值的方法。Burgess 和 Webster^[1,2] (1980) 详细地描述了它在土壤调查中的应用。这里概述它的基本要点如下。设 x_0 为所研究范围内的任一点。在位置 x_0 上的 Kriging 估计值，记为 $\hat{z}(x_0)$ ，它是一组以 x_0 为中心的一定邻域内所有观测值 $z(x_i)$ 的加权平均值

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (7)$$

其中 λ_i 为权重系数, $i = 1, 2, \dots, n$ 。对邻域内任一点进行 Kriging 估值, 它有两个基本的要求。第一, 这种估值必须是无畸偏的, 即邻域内任一点的估值须与该点期望值相一致, 因此权重系数之和必须为 1, $\sum_{i=1}^n \lambda_i = 1$ 。第二个基本要求是这种估值要具有最小的估计方差, 或称为 Kriging 方差, 通常记为 σ_k^2 。在 x_0 点进行估值的方差是该点的期望值 $\hat{z}(x_0)$ 和其真值 $z(x_0)$ 之差的平方, 即方差的数学期望值:

$$\begin{aligned} \sigma_k^2 &= E\{[z(x_0) - \hat{z}(x_0)]^2\} \\ &= 2 \sum_{i=1}^n \lambda_i \gamma(x_0, x_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i, x_j) \end{aligned} \quad (8)$$

这里 $\gamma(x_i, x_j)$ 为若干样品点 i 和 j 点之间的观测值的半方差, 它可从半方差图得到, $\gamma(x_0, x_i)$ 为若干样品点 i 和被估值点 x_0 之间的半方差。当它们满足下式

$$\sum_{j=1}^n \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_i, x_0) \quad (9)$$

时, 它的方差 σ_k^2 具有最小值。其中 μ 为位差参数。此时最小方差为

$$\sigma_k^2 = \sum_{i=1}^n \lambda_i \gamma(x_0, x_i) + \mu \quad (10)$$

由于所选用彰武县的样品平均间距较大, 不宜采用严格的点 Kriging 估值, 而采用块状 Kriging 估值方法, 即不再对某一点 x_0 进行估值, 而代之以估计以 x_0 为中心的一小方块 B 的平均值, 其估值的 Kriging 方差记为 σ_{kb}^2

$$\begin{aligned} \sigma_{kb}^2 &= E\{[z(B) - \hat{z}(B)]^2\} \\ &= 2 \sum_{i=1}^n \lambda_i \bar{\gamma}(B, x_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i, x_j) - \bar{\gamma}(B, B) \end{aligned} \quad (11)$$

这里 $\gamma(x_i, x_j)$ 仍为若干样品点 i 和点 j 之间的半方差, 而 $\bar{\gamma}(B, x_i)$ 则为样品点 i 与被估值小方块 B 之间的平均半方差, $\bar{\gamma}(B, B)$ 则为小方块 B 内点之间的平均半方差。此时, 当所有点满足

$$\sum_{i=1}^n \lambda_i \gamma(x_i, x_j) + \phi = \bar{\gamma}(B, B) \quad (12)$$

时, 估计值方差最小, 并为

$$\sigma_{kb}^2 = \sum_{i=1}^n \lambda_i \gamma(B, x_i) + \phi - \sigma^2(B) \quad (13)$$

这里 ϕ 为最佳块状估值位差参数, $\sigma^2(B)$ 为小方块内估计方差。解方程 (12), 它实际是一组线性方程组, 即可求出各样品点 x_i 观测值的地域变量对方块 B 中心点 x_0 估值的一套权重系数 λ_i 和位差参数 ϕ 。然后用它们解方程 (7) 和 (13) 就可求出 x_0 点的估计值 $\hat{z}(x_0)$ 与估计方差 σ_{kb}^2 。在实践中, 我们选用 1 平方公里作为方块 B , 计算离方块中心点 x_0 最近的 50 公里范围内 10 到 25 个样品观测值的统计权重, 更远或更多的点不予考虑。图 7 所示为某点实际估值的统计权重分布图, 为了图示的方便仅取最近的 15 个点对该点进行块状 Kriging 估值, 用这 15 个观测值的地域变量所得的估值为 -1.2914, 估计方差为 0.0151。Kriging 估值方法比其他地域估值方法的主要优点之一是它可以计算估值的

估计方差,并使其为最小。但尽管如此,如果距离 x_0 最近的观测点很远,或者半方差图的基底方差很大,或者两者皆是,则方差值仍然可能是大的。

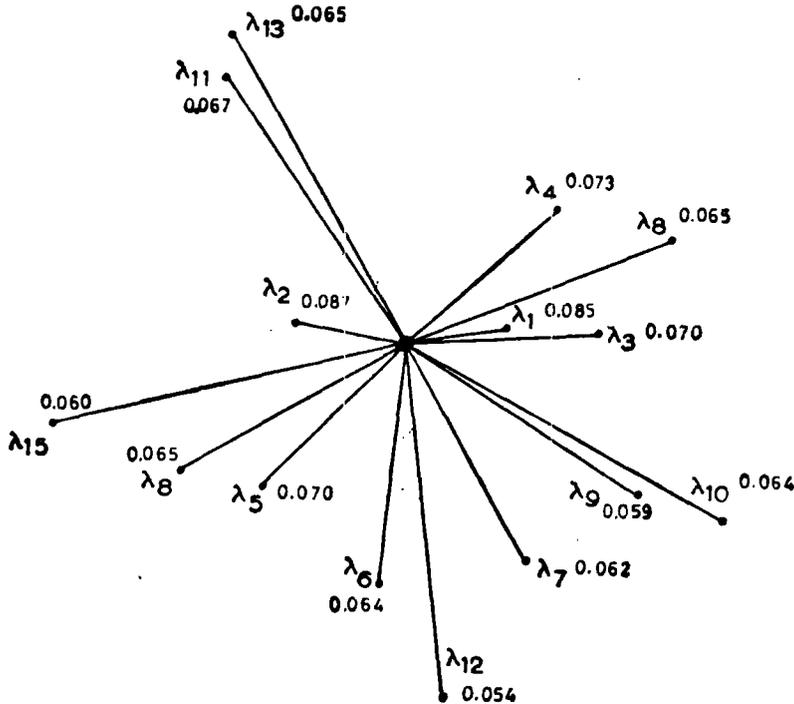


图7 估值点的权重分配图

Fig. 7 Distribution of the weights of estimated point

六、等值线图与立体透视图

为了用计算机自动绘制彰武县表层土全氮量的等值线图与立体透视图,我们用1平方公里块状 Kriging 方法,对全县以1.25公里为地图坐标方格逐点进行估值,得到了全县3,475个数据。在获得了地域各点的最佳估值之后,就可应用计算机制图软件绘制该项目更为精细坐标方格的等值线图和立体透视图。我们应用洛桑试验站计算机室的 SURFACE II^[9] 和 GHOST BENSON 系统^[4]绘制了彰武县全氮量等值线分布图与立体透视图(图8和图9),还绘制了进行块状 Kriging 最佳估值的方差等值分布图与立体图(图10和图11)。应用全县全氮量等值线图(图8)与土壤调查分类图(图1)相比较时,不难发现,尽管这两幅图是完全不同的项目,但土壤与全氮量的分布确存在着一定的相关性。东部的棕色森林土的全氮量含量最高,都在0.07%以上。西部褐土较低,普遍在0.05%以下。北部的风沙土最低。而中部和南部的草甸土,大多居于中间,在0.05—0.07%范围内。这表明用全县102个表层土全氮量观测值进行 Kriging 最佳估值确实能在一定程度上反映出不同土壤中全氮量的数量规律,或者说为土壤分类提供了一定的数量依据。若仅用102个原始土类全氮量数据作这种精确的数量估计却是难于实现的。

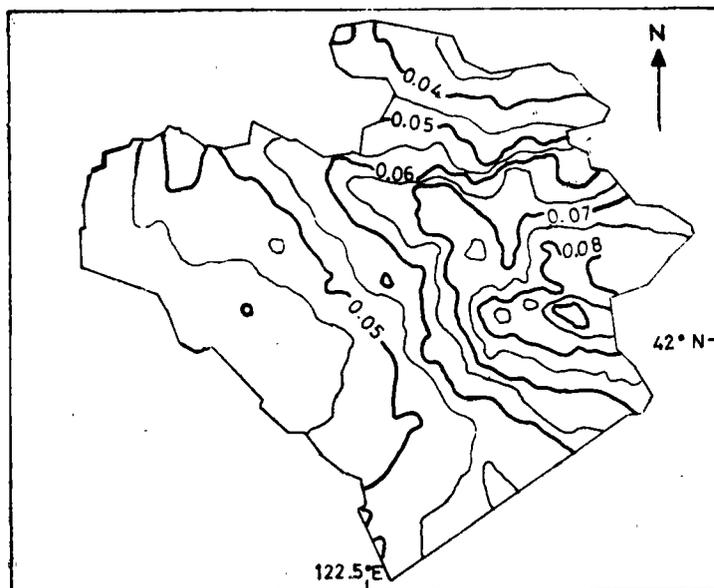


图 8 全氮量块状 Kriging 估值等值线图

Fig. 8 Map of total nitrogen in topsoil made by contouring kriging estimates over 1 km×1km blocks

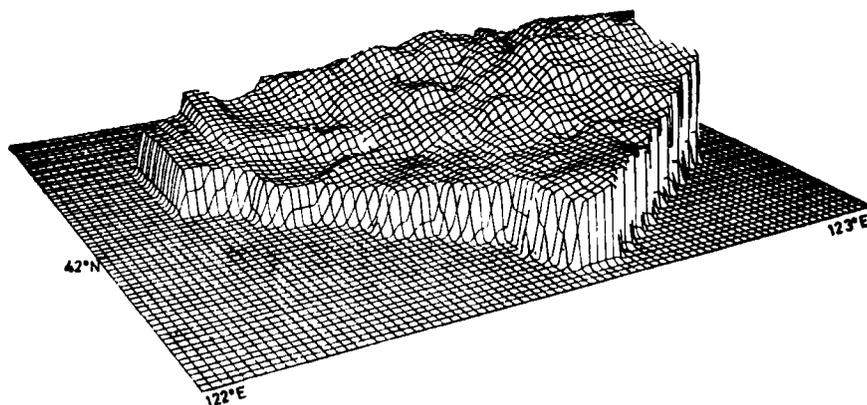


图 9 全氮量块状 Kriging 估值立体透视图

Fig. 9 Perspective graph of total nitrogen in topsoil associated with figure 8

七、讨 论

由以上对彰武县表层土全氮量的地域统计分析看出,尽管所采用的数据是少了一些,但在 50 公里范围内的半方差图确实存在着明显地域分布相关性,用它进行块状 Kriging 最佳估值,仍然可以相当好地反映出全县全氮量的地域分布趋势,为土壤分类、土地利用、改良以及施肥提供一定的数量依据。但须指出,由于样品采样密度小,样品点之间的间距较大,必然会增大 Kriging 估值方差。从图 10、11 估值方差图不难看出,县内地区估值方差要比沿县边界周围各点的要小,这也是显而易见的,因为沿县边界的点只有县内的土壤样品观测值用于地域统计分析,而县外无点提供分析,这自然增大了它们的统计方差。

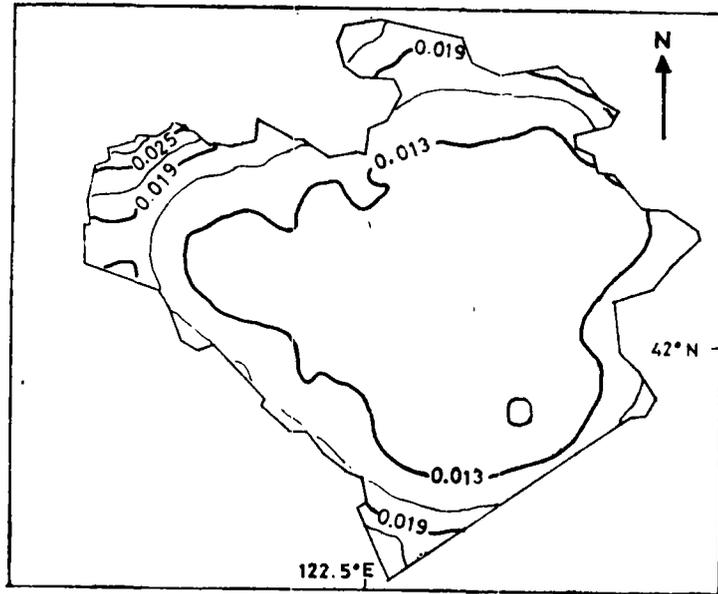


图 10 块状 Kriging 估值方差等值线图

Fig. 10 Map of the kriging estimation variance of total nitrogen in topsoil associated with figure 8

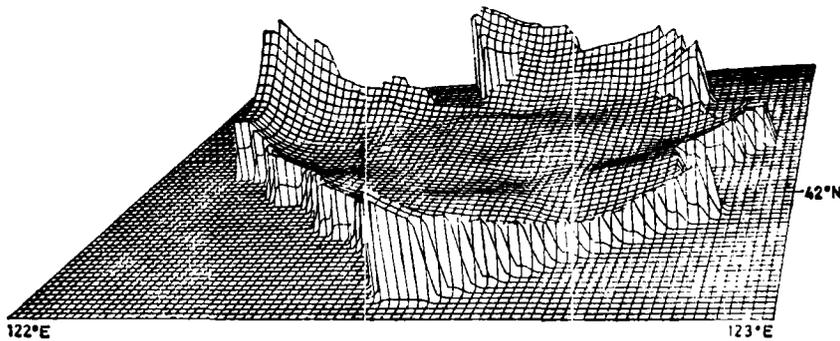


图 11 块状 Kriging 估值方差立体透视图

Fig. 11 Perspective graph of the block kriging estimation variance associated with figure 10

同时用 102 个点对全县土壤性质进行 Kriging 估值,仅只能用块状 Kriging 估值法,作地域分布趋势估计,即只能估计出一定区域内的平均值,而不能反映点或小面积地块的土壤性质特征。若要用严格的点 Kriging 方法对点或小面积土壤性质特征进行最佳估值,则必须适当增加土壤采样密度,缩小样品的平均间距,同时尽可能地收集边缘以外一定距离范围内的土壤样品观测值提供进行地域统计分析。

总之,地域统计分析方法,作为土壤数学和电子计算机技术在土壤调查和土壤科学中的应用,对于地域辽阔的我国来说,不仅具有重要的理论意义,而且有很大的实际应用价

值。当然它不仅可应用于土壤科学,从广义上说,凡是在地域上有相对稳定不变的地域分布特征的地质、探矿、地理、林业、植物和环境调查等研究领域,都可用它来研究各自的地域变量在空间分布的数量规律,因此,它具有广阔的前途。

参 考 文 献

- [1] Burgess, T. M. and Webster, R., 1980a: Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. *Journal of Soil Science*, 31: 315—331.
- [2] Burgess, T. M. and Webster, R., 1980b: Ditto, II. Block kriging, 31: 331—341.
- [3] Burgess, T. M., Webster, R. and McBratney, A. B., 1981: Ditto, IV. Sampling Strategy, 32: 643—659.
- [4] Calderbank, V. J. and Prior, W. A. J., 1977: The ghost graphical output system. User Manual, United Kingdom. Atomic Energy Authority, Culham, Oxford.
- [5] Journel, A. G. and Hvijsbregts, Ch. J., 1978: *Mining Geostatistics*, London: Academic Press.
- [6] Matheron, G., 1965: *Les variables regionalisees et leur estimation* Paris: Masson.
- [7] Matheron, G., 1971: *The theory of regionalized variables and its applications*, Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau No. 5.
- [8] Ross, G. J. S., 1980: MLP Maximum likelihood program, Rothamsted Experimental Station.
- [9] Sampson, R. J., 1978: The surface II graphics program (Revision one) No. 1, Series on Spatial Analysis, Kansas Geological Survey, Lawrence.
- [10] Webster, R., 1977: *Quantitative and Numerical Methods in Soil Classification and survey*, Oxford University Press.

OPTIMAL ESTIMATION OF SOIL SURVEY DATA BY GEOSTATISTICAL METHOD—SEMI-VARIOGRAM AND BLOCK KRIGING ESTIMATION OF TOP- SOIL NITROGEN OF ZHANGWU COUNTY

Xu Jiyan

(Institute of Forestry and Soil Science, Academia Sinica)

R. Webster

(Rothamsted Experimental Station, U. K.)

Summary

Kriging is a method of estimating values of properties at places without bias and with minimum error from observed values in the neighbourhood. The method was applied to measurements of total nitrogen in the top layers of the soils in Zhangwu County in China. After transforming the measurements to logarithms to stabilize their variances, the semi-variogram was computed from them. The semi-variogram showed spatial dependence extending to 50 hm, although it had a large non-spatial, or "nugget" component. Mathematical models were fitted, and the one giving the best fit in a least square sense was used to estimate values and their estimation variances on a fine grid (1.25 km) from which isarithmic maps of the county were drawn.