

岩土参数随机场空间最优估计精度 分析与特异值研究*

The accuracy analysis on Kriging estimation of the geotechnical parameters
random fields and a approach to the outliers processing

张 征

(北京林业大学资源与环境学院, 北京, 100083)

王思祥

(邯郸市工程质检总站, 邯郸, 056001)

程祖锋

(河北建筑科技学院, 邯郸, 056038)

刘淑春

(北京林业大学资源与环境学院, 北京, 100083)

文 摘 阐述了克立格方差的特点, 并结合实际对岩土参数随机场空间最优估计的局部估计精度进行了分析, 探讨了影响岩土参数空间最优估计精度的特异值的处理方法。研究表明, 对存在特异值的岩土参数随机场进行估计时, 应用指示克立格法能够获得符合实际的空间最优估计效果。

关键词 岩土参数, 随机场, 局部估计精度, 特异值, 克立格方差

中图法分类号 TU 471.4

作者简介 张 征, 男, 1957 年生, 工学博士, 环境科学专业博士后。1982 年毕业于长春科技大学环境与建设工程学院(原长春地质学院水文地质与工程地质系)。目前主要从事环境科学与工程、水资源、岩土工程等领域空间变异系统分析理论与方法研究。

Zhang Zheng

(College of Forest Resources and Environment, Beijing Forestry University, Beijing, 100083)

Cheng Zifeng

(Hebei Institute of Architectural Science and Technology, Handan, 056038)

Wang Enxiang

(Handan Central Engineering Quality Testing Station, Handan, 056001)

Liu Shuchun

(College of Forest Resources and Environment, Beijing Forestry University, Beijing, 100083)

Abstract In Kriging estimation on the geotechnical parameters, Kriging variance has a special meaning. This paper elaborates the characteristics of Kriging variance, and analyses systematically the locally estimated Kriging accuracy of the random fields of geotechnical parameters in the light of a practical example. The concrete procedure to deal with the outliers which exert a great influence on estimated Kriging accuracy of the geotechnical parameters has been preliminarily researched in the paper too.

Key words geotechnical parameters, random fields, locally estimated accuracy, outliers, Kriging variance

1 引 言*

在岩土参数不确定性与离散性研究中, 不可避免地会遇到岩土参数的空间最优估计问题, 而估计精度的好坏直接关系到岩土参数离散性分析的效果以及岩土工程勘察与设计方案的选择。空间最优估计方法在岩土参数不确定性与离散性研究中的应用, 为各种岩土参数的局部估计提供了衡量估计精度的定量指标, 即克立格方差。

从分析克立格方差的特点出发, 讨论岩土参数随机场空间最优估计的精度问题, 并针对岩土介质发育的不均匀性, 探讨岩土参数空间最优估计中特异值的处理方法。对完善岩土参数随机场空间变异性分析方法体系具有特殊意义。

2 局部估计精度的定量指标及其特点

2.1 局部估计精度定量指标

岩土参数随机场空间最优估计的成果主要体现在空间最优估计值与局部估计精度定量指标两个方面, 二者不可分割, 构成了一个完整的空间最优估计体系^[1]。其中局部估计精度定量指标——普通克立格方差表达式为^[2]:

$$\sigma_K^2 = \sum_{i=1}^n \lambda_i^0 Y(h_{i0}) + \mu \quad (1)$$

式中 μ 为拉格朗日乘子; λ_i^0 为克立格权系数;

* 中国博士后科学基金资助项目(No. 中博基 199711)

$\gamma(h_{i0})$ 代表岩土参数随机场待估点 x_0 与信息点 x_i 之间的变异函数值,二者均可由岩土参数空间结构分析结果获得^[3]。

2.2 局部估计精度定量指标的特点

由局部估计精度表达式(1)可见,普通克立格方差的大小是由参估信息点与待估点之间的变异函数值、拉格朗日乘子以及克立格权系数所决定的,而与参估信息值的大小无关。因此,只要参估信息的数据构型与估计构型一定,便可以在待估点取得实测值之前根据变异函数值以及普通克立格方程组^[2]的求解结果,事先计算出勘察场地内待估点的局部估计精度,从而为确定最佳观测、取样、布孔方案及选择设计方案提供依据。

3 岩土参数空间最优估计精度分析

3.1 不同方法空间估计精度对比分析

岩土参数随机场空间最优估计方法是具有不确定性的岩土参数视为区域化变量,应用克立格法对其离散性进行最优估计的方法。笔者曾对勘察现场获得的一批离散的岩石透水性指标进行了不同方法空间估计效果的对比分析,其结果表明,该法在所研究的岩土参数离散性的空间估计中确比传统估计方法有着更强的有效性和无偏性^[4]。

3.2 空间观测序列内预测精度分析

为了结合实际分析岩土参数空间最优估计的局部估计精度,本文从勘察现场选择了部分代表性区域化变量进行变异函数计算和空间最优估计,通过对其检验孔的克立格方差(σ_k^2)及估计误差的分析对比,获得了如下几点认识:

(1) 区域化变量的空间变异性直接影响着空间最优估计的精度。勘察现场部分区域化变量理论变异函数参数值见表 1。不同岩土参数区域化变量空间变异程度存在着明显的差别,这种差别必然会反映在空间最优估计的精度上,空间变异性愈强,其空间估计的总体精度愈低(表 2、表 3)。

表 1 区域化变量空间变异性与估计效果对比

Table 1 Comparison between spatial variability of regionalized variables and estimated Kriging accuracy

$Z(x)$	D_v				$\overline{D_v}$	$\overline{\sigma_k^2}$	估计相对 误差均值/%
	0°	45°	90°	135°			
Z_1	1.0000	0.0043	0.9697	0.0021	0.4940	8.4541	264.5524
Z_2	0.0026	0.0037	0.0061	0.7071	0.1799	0.0063	3.6125

注:表中 Z_1 为岩石透水性指标(渗透系数); Z_2 为岩土介质中水的酸碱度指标; D_v 为多元信息方向综合变异指标,用以表示空间变异性大小; σ_k^2 为克立格方差,用以表示空间估计精度; $\overline{D_v}$ 、 $\overline{\sigma_k^2}$ 为 D_v 、 σ_k^2 的均值。

表 2 区域化变量(Z_1)空间估计结果检验

Table 2 The tests in estimated Kriging accuracy of regionalized variables (Z_1)

点号	$Z_1/(m \cdot d^{-1})$		σ_k^2	误差	相对误差/%
	实测值	估计值			
1	0.080	0.977	12.814	-0.897	1121.250
2	1.020	3.425	8.227	-2.405	235.784
3	0.960	0.320	7.399	0.640	66.667
4	6.390	2.131	8.916	4.259	66.651
5	0.590	5.289	7.287	-4.699	796.441
6	5.000	7.358	16.551	-2.358	47.160
7	18.110	0.928	10.933	17.182	94.876
8	3.020	1.910	6.860	1.110	36.755
9	3.000	6.615	1.750	-3.615	120.500
10	3.215	5.126	3.804	-1.911	59.440

表 3 区域化变量(Z_2)空间估计结果检验

Table 3 The tests in estimated Kriging accuracy of regionalized variables (Z_2)

点号	Z_2		σ_k^2	误差	相对误差(%)
	实测值	估计值			
1	7.290	7.459	0.007	-0.169	2.318
2	7.580	7.248	0.005	0.332	4.380
3	7.460	7.322	0.005	0.138	1.850
4	7.300	7.579	0.005	-0.279	3.822
5	7.380	7.527	0.008	-0.147	1.992
6	7.900	7.258	0.008	0.642	8.127
7	7.600	7.445	0.010	0.155	2.039
8	7.300	7.590	0.007	-0.290	3.973
9	7.500	7.284	0.007	0.216	2.880
10	7.200	7.413	0.002	-0.213	2.958

(2) 克立格方差(σ_k^2)是衡量岩土参数随机场空间最优估计精度的指标,它描述了在一定数据构型与估计构型下,利用各种信息数据进行空间估计的估计方差的取值中心,即“期望值”,因而才表现出与参估信息值大小无关并可事先进行精度预测的特点^[1]。一般来讲,当估计构型表现为待估点位于周围信息点之间时,其估计精度高(σ_k^2 小);当估计构型表现为待估点位于信息点的边缘或待估点周围信息点较少时,其估计精度低(σ_k^2 大)。另外,当待估点距信息点较远时,估计精度较低(σ_k^2 较大);待估点距信息点较近时估计精度较高(σ_k^2 较小),见表 4。

表 4 不同位置区域化变量预测精度对比

Table 4 Kriging variance of regionalized variables at different locations

变量	Z_1		Z_2		Z_3		Z_4	
预测点号	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
预测点相对坐标/km	x	y	x	y	x	y	x	y
预测值	3.789	5.253	7.487	7.492	0.573	0.573	12.564	13.006
σ_k^2	1.615	2.827	0.007	0.008	7.604	7.604	1.394	1.469

注: Z_3 为岩土介质中的方解石饱和度; Z_4 为岩土介质中水的氯离子浓度, mg/L。

(3) 当岩土参数区域化变量 $Z(x)$ 具有各向同性结构时, 在几何形状、位置上与待估点 x_0 有对称性的信息样品在没有其它效应影响的条件下, 具有相同的权系数。但自然界中的岩土介质是各向异性的, 在这种客观条件下即使信息点位置与待估点有对称性, 也不能得出克立格权系数有对称性的结论。

(4) 在野外实际工作中, 岩土工程勘察取样、测试点的分布通常是不均匀的。在这种情况下, 即使空间变异结构是各向同性的, 位于与待估点对称位置上的两个信息样品的克立格权系数, 也会因受到分布不均匀的其它信息样品的影响而失去对称性。因此, 用克立格法对岩土参数区域化变量进行估值时, 不会因某个区域的信息样品丛聚在一起而过分增大这一区域信息样品总的权系数, 这就是所谓减弱丛聚效应^[5]。此外, 克立格权系数还具有屏蔽效应, 即当块金常数为零或很小时, 距离待估点最近一圈样品的权系数明显地要比待估点周围第二圈样品的权系数为大。但这种屏蔽效应随着块金常数值增大而逐渐减弱。

总之, 影响岩土参数随机场克立格估计精度的基本因素主要包括岩土参数空间结构的连续性及变异性

$\gamma(h)$ 、信息样品与待估点间的几何关系 $\sum_{i=1}^n \lambda_i^0 \gamma(h_{i0})$ 、数据构型特征 $\sum_{i=1}^n \sum_{j=1}^n \lambda_i^0 \lambda_j^0 \gamma(h_{ij})$ 、估计块段时待估点所位于的待估块段本身的几何特征 $\bar{\gamma}(V, V)$ 等 4 种因素。为提高估计精度, 在可能条件下, 可由调整后 3 项因素获得。这也为岩土工程勘探网的合理布局提供了依据。

(5) 在对岩土参数进行局部估计时, 估计邻域内参数信息值的大小及局部数值波动幅度对估计精度亦有些影响, σ_k^2 与估计误差间并不存在完全对应的确定性变化关系, 这点在衡量岩土参数局部估计精度时应加以注意(表 5、表 6)。事实上, 把克立格方差视为局部估计精度量度(指一次局部估计)的观点仅适合于多正态随机域, 而这种随机域与实际模型间存在一定差距。比较而言, 把克立格方差当做数据构型指数似乎更恰当^[6]。

表 5 区域化变量(Z_3) 空间估计 σ_k^2 与估计误差对比

Table 5 Comparison between spatial estimated error and $\sigma_k^2(Z_3)$

Z_3			σ_k^2	误差	相对误差 /%	$\frac{1}{\sigma_k^2}$	相对误差 均值/%
点号	实测值	估计值					
1	0.580	0.600	8.554	-0.020	3.448		
2	0.490	0.667	7.983	-0.177	36.122		
3	0.820	0.614	8.211	0.206	25.122		
4	0.580	0.595	7.698	-0.015	2.586	8.285	21.905
5	0.490	0.660	9.125	-0.170	34.694		
6	0.930	0.570	8.212	0.360	38.710		
7	0.490	0.552	8.211	-0.062	12.653		

表 6 区域化变量(Z_4) 空间估计 σ_k^2 与估计误差对比

Table 6 Comparison between spatial estimated error and $\sigma_k^2(Z_4)$

$Z_4/(mg \cdot L^{-1})$			σ_k^2	误差	相对误差 /%	$\frac{1}{\sigma_k^2}$	相对误差 均值/%
点号	实测值	估计值					
1	9.900	12.600	1.062	-2.700	27.273		
2	8.500	13.368	0.791	-4.868	57.271		
3	14.200	10.420	0.751	3.780	26.620		
4	14.200	10.267	0.756	3.933	27.697	1.251	26.566
5	9.900	13.259	1.848	-3.359	33.929		
6	12.800	12.354	1.412	0.446	3.484		
7	11.300	12.395	1.885	-1.095	9.690		

4 岩土参数随机场空间最优估计中特异值处理方法探讨

岩土介质的非均质各向异性是客观存在的, 由于受取样条件的限制, 在评价岩土参数离散性时, 往往遇到某些“特异值”, 这给岩土参数空间变异结构分析及空间最优估计理论模型计算带来了相当大的困难。因此, 特异值处理方法的研究, 对岩土参数随机场空间估计方法的应用, 具有十分重要的实际意义。

4.1 岩土参数特异值识别及处理的一般方法

对特异值进行识别和处理的一般方法包括常规数理统计方法、影响系数法和估计邻域法等^[7]。其中常规数理统计方法主要依据样品均值(m)、均方差(σ)倍数、变异系数(σ/m)及分布密度函数曲线上拐点对应值来识别和处理特异值。影响系数法是在研究空间分布变量变异程度基础上, 对可能出现的特异值的影响系数(k)人为赋值, 以适当地抑制其影响程度的一种特异值识别和处理的方法。该法对样品组观测值需进行多次识别后方能识别出所有特异值。估计邻域法是 D·G·克立格与 D·M·霍金斯将空间分布变量地质统计分析思想用于识别及处理特异值的方法。该法将所提出的识别特异值的统计量与邻域内观测值的平均方差(σ^2)及变异系数联系起来, 使得特异值的识别及处理同时考虑了观测值和样品的空间环境。经上述方法识别出的特异值均采用“下限化”、“均值化”或“正常值最优化”等方式加以处理, 极端差异性信息被“平滑”掉了。在岩土参数离散性研究中, 应更加注重应用那种既能保证足够的空间估计精度, 又能刻画岩土介质不均匀性(有条件地保留特异值)的处理方法。

4.2 一种非参数地质统计学方法——指示克立格法

(1) 指示函数与指示变异函数

在研究域每一信息样品点上定义一个关于某一区域化变量 $Z(x)$ 信息值的如下阶梯函数:

$$i(x; z) = \begin{cases} 1, & Z(x) \leq z \\ 0, & Z(x) > z \end{cases} \quad (2)$$

式中 z 为区域化变量 $Z(x)$ 的边界信息值。在给定 z 的条件下, 随机函数 $I(x; z)$ 服从二项分布。当 $I(x+h; z)$ 和 $I(x; z)$ 为被 h 分隔的代表指示函数的两个区域化变量时, 则可定义如下指示变异函数及实验指示变异函数^[7]:

$$\gamma_1(h; z) = \frac{1}{2} E [I(x+h; z) - I(x; z)]^2 \quad (3)$$

$$\gamma_1^*(h; z) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [I(x_\alpha+h; z) - I(x_\alpha; z)]^2 \quad (4)$$

式中 $N(h)$ 为数据对数目; $E\{\cdot\}$ 代表数学期望。

(2) 指示克立格方程组

设 $\{Z(x_\alpha), x_\alpha \in D, \alpha = 1, 2, \dots, n\}$ 为研究域 D 内 n 个区域化变量的观测信息值, 在给定边界信息值后, 得到样品的指示函数空间: $\{i(x_\alpha; z), \alpha = 1, 2, \dots, n\}$, 则 $I(x_0; z)$ 的空间估计值可表示为

$$I^*(x_0; z) = \sum_{\alpha=1}^n \lambda_\alpha^0(z) \cdot i(x_\alpha; z) \quad (5)$$

式中待定权系数 $\lambda_\alpha^0(z)$ 的求解要求满足无偏和估计方差最小的条件, 为此, 应该解如下指示克立格方程组^[7]:

$$\left. \begin{aligned} \sum_{\beta=1}^n \lambda_\beta^0(z) \gamma_1(x_\alpha, x_\beta; z) + \mu &= \gamma_1(x_\alpha, x_0; z) \\ (\alpha &= 1, 2, \dots, n) \\ \sum_{\alpha=1}^n \lambda_\alpha^0(z) &= 1 \end{aligned} \right\} \quad (6)$$

其指示克立格方差为 σ_{IK}^2 为

$$\sigma_{IK}^2 = \sum_{\alpha=1}^n \lambda_\alpha^0(z) \gamma_1(x_\alpha, x_0; z) + \mu \quad (7)$$

式中 μ 为拉格朗日乘子; $\gamma_1(x_\alpha, x_\beta; z)$, $\gamma_1(x_\alpha, x_0; z)$ 分别代表在给定边界信息值 z 条件下, 各信息点 x_α 与 x_β 之间以及待估点 x_0 与信息点 x_α 之间的指示变异函数值 ($\alpha, \beta = 1, 2, \dots, n$)。

当给定一系列边界信息值 $z_l (l = 1, 2, \dots, L)$ 时式 (5) 可以改写为

$$I^*(x_0; z_l) = \sum_{\alpha=1}^n \lambda_\alpha^0(z_l) \cdot i(x_\alpha; z_l) \quad (8)$$

为了求得 $I^*(x_0; z_l)$ 就应该解 l 个指示克立格方程组。其相应的计算域待估点 x_0 的估计值可由下式求得^[7]:

$$Z^*(x_0) = \sum_{l=1}^L I_l^*(x_0) \cdot [Z(x_0) | x_0 \in z_l]^* \quad (9)$$

(3) 指示克立格法的计算步骤及应用实例

a) 原始数据预分析及指示值转换

以勘察现场岩土参数区域化变量 Z_1 原始数据为例, 该批数据中最大值 18.11 (Z_{117} 孔), 已大大超过其它钻孔该区域化变量的平均值 2.815。据式 (2) 可方便地确定任一有意义的边界信息值, 将特异值与原始数据“分离”开来, 其转换后的指示值见表 7。

b) 实验指示变异函数计算及理论变异模型拟合

据表 7 指示值, 利用式 (4) 进行实验指示变异函数计算, 并运用加权多项式回归法对变异曲线进行球状模型的最优拟合^[5], 其拟合结果揭示了勘察现场岩土参数区域化变量小于等于某一边界信息值的指示函数值空间变化的结构性(表 8)。表中变异函数理论模型及其各种参数, 在岩土介质空间变异性分析中均有着特定的物理涵义^[3]。

c) 求待估点属于边界信息值下的概率估值

利用式 (5) 估计式, 其中待定权系数由式 (6) 指示克立格方程组解出, 可获得勘察现场计算域岩土参数属于边界信息值下的空间分布概率。

表 7 岩土参数原始数据指示值转换表

Table 7 Indicator value transformed from primary data of geotechnical parameter

孔号	Z_{112}	Z_{84}	Z_{113}	Z_{85}	Z_{137}	Z_{116}	Z_{142}	Z_{73}	Z_{117}	Z_{138}	Z_{74}	Z_{86}	Z_{87}	Z_{102}	Z_{90}	Z_{89}	Z_{144}	Z_{76}	Z_{103}
$Z_1/(m \cdot d^{-1})$	0.08	1.02	0.96	6.39	0.78	0.59	1.76	5.00	18.11	0.43	3.02	0.86	2.69	3.00	5.11	3.22	2.47	4.15	9.15
$i(x_\alpha; z)$	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1

表 8 实验指示变异函数计算及理论变异模型拟合成果表

Table 8 Calculated results of experimental indicator variogram and fitted spherical model

方向	变程 a	拱高 C	块金常数 C_0	基台值 B	指示实验函数理论模型	误差标准偏差 σ_e^*
0°	33.5459	0.0096	0.0	0.0096	$\gamma_1(h; z) = 0.0096\text{Sph}(33.5459), 0 < h \leq 33.5459$	0.2684
45°	37.3576	0.0150	0.0	0.0150	$\gamma_1(h; z) = 0.0150\text{Sph}(37.3576), 0 < h \leq 37.3576$	
90°	34.2981	0.0111	0.0	0.0111	$\gamma_1(h; z) = 0.0111\text{Sph}(34.2981), 0 < h \leq 34.2981$	
135°	50.0348	0.0170	0.0	0.0170	$\gamma_1(h; z) = 0.0170\text{Sph}(50.0348), 0 < h \leq 50.0348$	

注: 表中 Sph 表示变异函数理论模型中的标准球型模型, 其物理意义见文献[4]。

d) 计算待估点岩土参数随机场区域化变量的估计值
首先计算待估点 x_0 处小于或等于各边界信息值的岩土参数区域化变量估计值, 然后根据式(9)即可计算出待估点岩土参数区域化变量的估计值。勘察场地计算域岩土参数随机场空间估值结果见图 1。图中岩土参数随机场空间最优估计值的计算、叠加由自编程程序完成, 图件形成由二维三维科技绘图软件 SURFER 实现。分析式(9)可知, 在岩土参数随机场区域化变量结构分析和空间估计过程中, 指示克里格法限制了特异值的影响范围, 但仍然以概率的方式考虑了特异值的存在。

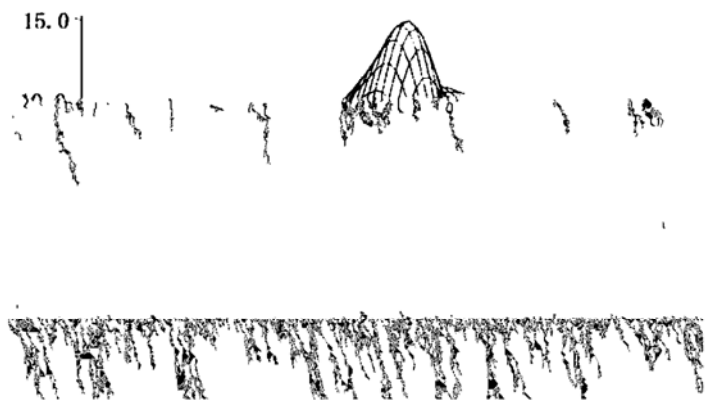


图 1 岩土参数随机场指示克里格法估值立体图

Fig. 1 Stereogram for estimated results of the random field of geotechnical parameter by indicator Kriging

5 结 论

(1) 克里格方差的大小是由参估信息点与待估点之间的空间配置关系、参估信息点相互间的空间配置关系以及相应的变异函数值所决定的, 与参估信息值的大小无关, 因而在勘察场地待估点处取得岩土参数实测值之前便可求出克里格方差。实践表明, 岩土介质空间变异性愈强, 岩土参数信息值数据构型与估计构型愈不理想, 其克里格方差愈大。据此可为岩土工

程勘察中取样、测试及监测点的合理布置、调整与优化提供定量依据。

(2) 在对岩土参数随机场区域化变量进行局部估计时, 估计邻域内参估信息值的大小及局部数值波动幅度对估计精度亦有影响, 用克里格方差反映局部估计精度是有条件的。这一点在衡量岩土参数局部估计精度时应加以注意。克里格方差给出的是在一定数据构型与估计构型下估计方差的取值中心, 因而表现出与参估信息值大小无关的特点。事实上, 克里格方差更多的是具有一种数据构型与估计构型指数的涵义, 在衡量岩土参数随机场局部估计精度时应正确使用。

(3) 属于非参数地质统计学范畴的指示克里格法, 既限制了特异值的影响范围, 又以概率的方式考虑了特异值的存在, 不失为一种对岩土参数随机场在边界信息值下的空间概率和空间分布数值进行最优估计的可行方法。

参 考 文 献

- 1 王俊庸等. 局部估计精度和最佳补钻位置. 见: 中国地质学会数学地质专业委员会主编. 中国数学地质(2). 北京: 地质出版社, 1990. 12~ 16
- 2 A·G·儒尔奈耳, CH·J·尤日布雷格茨著. 矿业地质统计学. 侯景儒, 黄竞先等译. 北京: 冶金出版社, 1982. 300~ 303
- 3 张 征等. 岩土参数的变异性及其评价方法. 土木工程学报, 1995, 28(6): 43~ 51
- 4 张 征等. 岩土参数空间变异性分析原理与最优估计模型. 岩土工程学报, 1996, 18(4): 40~ 47
- 5 王仁铎等. 线性地质统计学. 北京: 地质出版社, 1988. 139~ 141, 68~ 71
- 6 Kostov C, Journel A G. 储层描述中专家信息的外推和编码. 李希宗译. 见: 许仕策等译. 油藏描述译文集. 北京: 石油工业出版社, 1994. 102~ 111
- 7 侯景儒等. 矿床统计预测及地质统计学的理论与应用. 北京: 冶金工业出版社, 1993. 435~ 437, 416~ 420