# 非参数地质统计学进展

孙玉建[1,2]
(1.中国地质大学,北京 100083;2.国土资源部矿产资源储量评审中心,北京 100035)

[摘 要]地质统计学从解决固体矿产资源评价起源,应用领域迅速扩展。文章简要介绍了地质统计学的理论研究现状和应用现状。为了减轻人为因素在创建模型中的作用,非参数的地质统计学得到发展,文章重点介绍了非参数地质统计学的4种变异函数模型。此外,文章阐述了多点地质统计学的概念,这是一种用训练映像代替变异函数的地质统计学方法,使得地质现象的解释更直观。

[关键词]非参数地质统计学 克里格法 进展
[中图分类号]P628 [文献标识码]A [文章编号]0495-5331(2007)04-0079-04

## 0 引 言

地质统计学从解决金矿采矿实际问题起源[1-2]并发展起来[3]后,迅速从采矿领域应用到土壤科学、图像压缩、生物学和流行病学[4]。

地质统计学空间-时间模型正越来越多地用于解决环境问题,如监控酸沉降(acid deposition)或全球变暖问题,预测降雨量或河川径流。每个学科都从自己的角度对空间-时间模型问题提出各自的解决途径[5]。

2000年第31届世界地质大会上,和地质统计学有关的主题是“石油储层的随机模型”和“根据矿体地质特征进行地质统计学评价”。由15篇文章组成的地质统计学论文集中主要介绍了应用地质统计学在创建石油储层和矿床模型方面所提供的解决方案。论文集展示了一种趋势,就是用地质统计学的方法为相、岩石类型、构造控制应力创建模型,以及对复杂地质体和多元变量整合进行非平稳模型的创建。从论文集来看,多数人喜欢用模拟的方法解决矿床模型的创建工作,因为他们认为利用克里格模型进行项目的投资决策可能会带来偏倚的结果[6]。但实际上克里格法依然是利用条件分布的无偏估计进行随机模拟的核心。从实例上来看,将地质统计学应用于非金属矿产矿床模型的模拟得到了很大发展。

## 1 地质统计学的新发展

在实际应用过程中,利用克里格法(kriging)对未知区域估值时存在一定的“光滑”效应,即将数值较低的部分过高估算,而将数值较高的部分过低估算[7]。但这并不意味着这种方法不精确或不实用,当进行局部估值时,在最小二乘误差意义上,克里格法依然是最精确的估值方法了[8]。为了降低这种光滑效应带来的负面影响,一个基本的想法就是在估计量和估计误差之间建立一个回归函数。Guertin(1984)[9]提出采纳“等因子仲法线分布模型(an isofactorial binormal distribution model)”,同时Olea和Pawlowsky(1996)[10]利用交叉验证的方法建立了一种回归。然而,这两种方法都没有复原在克里格法估值中正确的空间变异性,即估计方差被低估的特性有所改善,但却没有再现空间变量的协方差模型。为了再现空间变量分布的全局变异性,人们发展了模拟的方法。Yao和Journel(1998)以及Yao(1998)[11-12]提出了一种建立在快速傅立叶变换基础上、应用密度谱方法进行条件模拟的方法,这种方法对克里格法估值进行严格地“后处理”从而再现正确的协方差模型,处理结果没有克里格法中人为的光滑效应,但也牺牲了局部精确性。此后,关于倾向于克里格法和赞成模拟方法的文章多有出现,Journel(2000)[8]认为全局精确性(半变异函数或结

[收稿日期]2007-01-08;[修订日期]2007-03-28。
[第一作者简介]孙玉建(1974年—),男,2002年毕业于中国地质大学(北京),获硕士学位,在读博士研究生,工程师,现主要从事矿产资源储量的估算与评价。

构再现)和局部精确性(估计方差的最小化)是相互矛盾的一对客观存在。

## 2 非参数变异函数模型

在普遍认为参数地质统计学带有较强的主观性的同时,人们在寻找各种途径减少人为因素在创建模型中的作用。这时,非参数地质统计学模型便日益盛行,如以下的非参数变异函数模型。

### 2.1 密度谱模型(Density Spectrum Model)

协方差模型为地质统计学提供了空间连续性的基本测度,正定性保证了克里格系统解的存在和唯一性。传统的求解样品协方差的方法是在保证正定性的同时,利用拟合闭合型分析模型的方法完成。对于协同克里格法来说,因为线性协同区域化模型的严格限制,创建协方差模型更为困难。Bochner (1949) [13]推出定理,对于协方差函数 $C(h)$,当且仅当它能用一个有限的非减测度 $S(w)$ 的傅立叶变换表示时,它才是正定的,即:

$$C(h) = \int e^{2\pi/h \cdot w} dS(w)$$

这里,$h$ 是 $d$ 维空间中的距离向量,$S(w)$ 是频率向量 $w$ 的累积分布函数。

根据 Bochner 的这一定理, Yao 和 Journel (1998)[11]利用快速傅立叶变换(Fast Fourier Transform)将试验(交叉)协方差表转换成密度谱表,这些密度谱表在正性和单一和(positivity and unit sum)的限制下进行光滑,这种变换方法不要求在某点上进行分析模型的拟合,并且这种算法不受线性协同区域化模型的限制。

### 2.2 贝塞尔模型(Bessel Model)

利用谱理论提出的关系式[4],当且仅当:

$$C(h) = \int_0^{\infty} \Omega_d(\mid h \mid t) dM(t)$$

这里,$M(t)$ 是一个有限的非减函数,并且:

$$\Omega_d(r) = \left(\frac{2}{r}\right)^{(d-2)/2} \Gamma\left(\frac{d}{d}\right) J_{(d-2)/2}(r)$$

这里,$J_n$ 是贝塞尔函数;特别的,

$$\Omega_1(r) = \cos(r), \Omega_2(r) = J_0(r),$$
$$\Omega_d(r) = \sin(r)/3$$

Shapiro 和 Botha (1991) [14]利用这个特点,将 $M(t)$ 作为阶梯函数,在结点 $t_j$ 处使用正的阶差 $p_j$,于是有:

$$\gamma(h) = \sum_{j=1}^{m} p_j (1 - \Omega(\mid h \mid t_j))$$

这个模型可以用最小化目标函数 $Q$ 来拟合,$Q$ 是参数 $p$ 的函数,$p = (p_1, \cdots, p_m)^T, p_i \geqslant 0$:

$$Q(p) = \sum_{i=1}^{k} w_i \Big(\hat{\gamma}(h_i) - \sum_{j=1}^{m} (1 - \Omega_d(\mid h \mid_i t_j) p_j)\Big)^2$$

贝塞尔模型一个明显的缺点是不容易得到变程/基台(range/sill)。这一特点被 Cherry (1997) [15]改善,基于 $C(0) = \int_0^{\infty} dM(t)$,用算子 $\hat{C}(0) = \sum_{j=1}^{m} \hat{p}_j$ 来代替基台(sill);但是,这一算子具有很强的偏倚性和高度的不稳定性,而且这种不稳定性对于变异函数估计的影响目前还不明确[16]。

### 2.3 "黑盒子克里格法"(Blackbox Kriging)

对于每一个平方可积函数 $f: R^m \to R$,积分:

$$2\gamma(h) = \int_{R^m} (f(x) - f(x-h))^2 dx$$

是一个与空间过程 $Z(x)$ 相适应、在加权函数 f 的作用下对"白噪音(white noise)"进行移动平均的、有效的变异函数。利用在一维区间和二维区间恒定的阶梯函数 $f$,使得模型成为分段线性和分段面性的。Barry 和 Ver Hoef (1996) [17]主要考虑空间估计,他们将这些模型用简单的线性过程结合在一起来最小化估计误差,并将这种方法命名为"黑盒子克里格法"。

对这些模型的拟合基本上是借助于阶梯函数 $f$ 的权重向量 $\theta$ 的最小化:

$$Q(\theta) = \sum_{i=1}^{k} \frac{\mid N(h_i) \mid}{\hat{\gamma}^2(h_i)} (\gamma(h_i \mid \theta) + \hat{c}_0 - \hat{\gamma}(h_i))^2$$

或者:

$$Q(\theta) = \sum_{x, \tilde{x} \in D} (2\gamma(\tilde{x} - x \mid \theta) + 2\hat{c}_0 - (z(\tilde{x}) - z(x))^2)^2,$$

用哪一种形式取决于哪种模型的"变异函数云(semiva - riogram cloud)"是可行的。式中 $\hat{c}_0$ 是块金效应 $c_0$ 的估计量,是事先用某种方法计算出来的量,最简单的情况是在实际观测距离的 10% 和 15% 范围内逼近被估算的变异函数值。

"黑盒子"模型在概念上容易理解,并且有一个物理解释(移动平均过程),是合理的非参数模型。当 $d = 1$ 时,他们可以任意精确度接近任意有效的变异函数。另外一个好处是可以用一种真正的非参数的方式处理各向异性体,不同的方向用不同的权系数 $\theta_i$。而且,用接近于原点的、更多更小的距离区间,模型具有更多的灵活性;但是,$d \neq 1$ 时的灵活性还有待于评估,而且块金常数的估算在大多数情况下需要额外的计算和精心观察。

### 2.4 样条函数光滑法(Spline Function Smoothing)

Lele (1995) [18]介绍了一种令人感兴趣的方法:

协方差增量的矩阵 $\Psi$:

$$\Psi_{ij} = Cov(Z(x_i) - Z(x_1), Z(x_j) - Z(x_1))$$

对于 $i,j=2,\cdots,n$,有:

$$\Psi_{ij} = \gamma(x_i - x_1) + \gamma(x_j - x_1) - \gamma(x_i - x_j),$$

$$\gamma(x_i - x_j) = \Psi_{i1} + \Psi_{j1} - 2\Psi_{ij}$$

这种算法始于对试验半变异函数图形的样条光滑,Lele 没有用一个有效的正定的函数来替换此光滑函数,而是计算了此样条光滑的矩阵 $\tilde{\Psi}$,并且用一个正定的接近值 $\hat{\Psi}$ 代替 $\tilde{\Psi}$(例如借助于频谱分析)。这个矩阵用来计算新的估计量 $\hat{\gamma}(x_i - x_j)$,新的试验半变异函数图形再次绘制出来并进行样条光滑,这个过程一直重复直至所估算的半变异函数达到"直观的光滑"。利用传统的样条光滑法来解决变异函数的拟合很吸引人,但这种方法的具体实施步骤还很粗略,方法的完善还有待于进一步工作。

## 3 地质统计学的新思路

虽然基于变异函数的地质统计学是应用最早也是应用范围最广的方法,它的局限性促使了"多点地质统计学"的提出。

作为数学科学的一个分支,地质统计学对物理概念的定量化和形式化,是建立在严格的数学法则和等式基础之上的,这就不可避免地将所研究的现象或自然过程过度简单化,这尤其体现在以变异函数为工具表达地质体的各向异性和连续性。变异函数描述的是空间两点变量在统计学意义上的相异性,而不是地质意义上的变化。当地质统计学家试图用变异函数来解释它所包含的意义时,发现它与实际的地质现象只有有限的联系。在固体矿产的开采过程中,采矿工程师关心某一局部位置品位的精确值,所以通过钻孔等手段取得大量的"硬数据"。而油气工作者往往不太关心局部孔隙率情况,他们更关心整体的渗透率,意图将储层内部的连通性和复杂的曲线特征(如运移通道或交错层理)用数学模型表达出来,因为变异函数只反映了单一地层单元或单一相内的某一个方向或某几个方向的变异性,应用起来就显得捉襟见肘了。为了利用地质工作中含有丰富信息的露头数据和多点数据,Caers (2002)[19] 提出"多点地质统计学(multiple - point geostatistics)。这是一种以训练映像(training image)为基础的不使用变异函数(Variogram - free geostatistics)的方法。"训练映像"实质上是地质模式的数据库,在此基础上,可以进行象变异函数这样的2阶统计,也可以进行更高阶的多点统计。一旦从训练映像中提取出所需的模式后,就需要依附于从地下工程(如钻孔编录、地震和开采生产)获得大量数据。在描述地质体的各向异性时,用训练映像代替了变异函数,因为它包含了多点信息,更重要的是在进行地质统计学的估计或模拟前,训练映像让人们更直观地看出一系列的多储层模型的模式。Feyen(2003)[20] 的例子说明了在具有复杂地质体体系的水文地质应用中描述地质体的各向异性是一个有利的工具,象砂道或泥质透镜体往往优先构成流体的通道或封闭体,所以在预测地下水的流动和运移时,精确地识别和定位这些地质结构显得十分重要。因为已知数据稀少,所以借用训练映像得出地质各向异性的预期模式并进行多点统计,然后将此多点统计导入地质统计学模型与大量的硬数据或软数据结合起来。多点地质统计学被认为是流体地质各向异性体模拟的一大进步。

特别感谢蒋小伟和陈劲松在资料收集过程中给予的帮助。

[参考文献]

[1] Krige D G. A statistical approach to some mine valuations and allied problems at the Witwatersrand, Master's Thesis[M]. University of Witwatersrand. 1951.

[2] Krige D G. Two - dimensional weighted moving average trend surfaces for ore valuation, in Proceedings of the Symposium on Mathematics, Statistics and Computer Applications in Ore Valuation, Johannesburg, 1966 [M]. South African Institute of Mining and Metallurgy, Johannesburg. 1966.

[3] Matheron G. Traité de Géostatistique Appliquée[M]. Vols 1 and 2, Technip, Paris. 1962 - 1963.

[4] Cressie N. Statistics for Spatial Data, Revised Edition [M]. Wiley, New York. 1993, 85.

[5] Kyriakidis Phaedon C, Andre G. Journel. Geostatistical Space - Time Models: A Review, Mathematical Geology[J]. 1999, 31 (6), 651.

[6] Armstrong M, C Bettini, N Champigny, A Galli and A Remacre (eds.). Geostatistics Rio 2000, Quantitative Geology and Geostatistics[M]. Kluwer Academic Publishers, Dordrecht. 2002.

[7] Goovaerts P. Geostatistics for natural resources evaluation[M]. Oxford University Press, 1997, 261.

[8] Journel A G, Phaedon C. Kyriakidis, and Shuguang Mao, Correcting the smoothing effect of estimators: A spectral postprocessor. Mathematical Geology[J]. 2000, 32(7), 787.

[9] Guertin K. Correcting conditional bias, in Verly, G, David, M., Journel, A. G, and Marechal, A, eds., Geostatistics for natural resources characterization [M]. Part 1: D. Reidel Publishing Co., Dordrecht, Holland, 1984, 245 - 260.

[10] Olea R A, Pawlowsky V. Compensating for estimation smoothing in kriging, Mathematical. Geology[J]. 1996, 28, (4), 407 - 417.

[11] Yao Tingting A G. Journel, Automatic Modeling of (Cross) Covariance Tables Using Fast Fourier Transform, Mathematical Geology[J]. 1998,30(6), 589.

[12] Yao T. Conditional spectral simulation with phase identification: Mathematical. Geology[J]. 1998,30(3),285 -308.

[13] Bochner S. Fourier transform[M]. Princeton Univ. Press, London, 1949,219.

[14] Shapiro A, Botha J . Variogram fitting with a general class of conditionally non - negative definite functions. Comput. Statist [J]. Data Anal. 1991,11, 87 -96.

[15] Cherry S. Non - parametric estimation of the sill in geostatistics. Environmetrics[J]. 1997, 8,13 - 28.

[16] Ploner A, D Rudolf, New directions in geostatistics, Journal of statistical planning and inference[J]. 2000, 91(2000), 502 - 505.

[17] Barry R, J Ver Hoef . Blackbox kriging: spatial prediction without specifying variogram models. Journal of Agriculture Biology Environmental Statistics[J]. 1996,1, 297 -322.

[18] Lele S. Inner product matrices, kriging, and non - parametric estimation of variogram. Mathematical Geology[J]. 1995,27,673 -692.

[19] Caers J, Zhang T. Multiple - point geostatistics: a quantitative vehicle for integration geologic analogs into multiple reservoir models. In: "Integration of outcrop and mod ern analog data in reservoir models" [M]. AAPG memoir. 2002.

[20] Feyen L, J Caers. Multiple - point geostatistics: a powerful tool to improve groundwater flow and transport predictions in multi - modal formations[M]. Fund for Scientific Research - Flanders (Belgium) , Postdoctoral Fellowship thesis. 2003.


## ADVANCES OF NON - PARAMETRICAL GEOSTATISTICS


SUN Yujian[1,2]

(1. *China University of Geosciences*, *Beijing* 100083;

2. *Mineral Resources and Reserves Evaluation Center*, *Ministry of Land and Resources*, *Beijing* 100035)



**Abstract**: Geostatistics came from solid mineral resources evaluation and is used in other fields quickly. Status of theoretical geostatistics and applied geostatistics is introduced. To reduce subjective factor in geostatistical modeling, non - parametrical geostatistics has become popular. This paper presents four kinds of variogram model for non - parametrical geostatistics. Furthermore, the concept of multiple - point geostatistics is introduced, which can represent geological phenomena intuitively via training image instead of variogram.


**Key words**: non - parametrical geostatistics, kriging, advance