

文章编号:1001-8166(2004)02-0268-07

地统计学方法进展研究*

孙英君¹,王劲峰¹,柏延臣²

(1. 中国科学院地理科学与资源研究所,北京 100101;2. 清华大学,北京 100084)

摘要:地统计学近年来成为空间分析的重要工具,其应用领域广泛分布在自然科学的众多领域。系统地论述了地统计学理论方法的进展,给出地统计学的总体理论框架,它包括局部估值、不确定性预测、随机模拟及多点地统计学四部分。并指出所有这些方法的理论基础是随机变量理论,其核心是利用变异函数获得研究对象的空间分布规律。同时,分别介绍了每个部分的具体方法,特别提到了新发展的多点地统计学方法。最后对地统计学的软件应用加以总结,指出将 GIS 与地统计学相结合是一种必然的趋势。

关键词:地统计学;克里格方法;随机模拟;多点地统计学

中图分类号:P7;C82 **文献标识码:**A

0 引言

地统计学起源于 20 世纪 60 年代,早期主要用于研究地质学现象的空间结构和进行空间估值。其创始人 Marheron^[20]将其简单定义为:随机函数在自然现象勘察及估计中的应用。从中可以看出,地统计学主要是利用随机函数对不确定的现象进行探索分析,并结合采样点提供的信息对未知点进行估计和模拟。地统计学最初主要用于采矿业和石油勘探中,但随着传统统计学方法在空间数据分析上的无能为力,越来越多涉及到空间分析的学科求助于地统计学的研究工具。如今,地统计学已经被广泛用于地理学、生态学、环境科学、土壤学等诸多领域的研究中。特别是 GIS 的发展带来的空间数据极大丰富,越来越多的科学家求助于地统计学来分析空间数据。

本文首先对地统计学地理论框架进行总结,在此基础上对其最新的研究进展进行分析,并对地统计学的软件应用问题进行了讨论。

1 地统计学的总体框架

1.1 区域化变量理论

地统计学处理的对象为区域化变量,即在空间

分布的变量。通常一个区域化变量具有两个性质:

在局部的某一点,区域化变量的取值是随机的;对整个区域而言,存在一个总体或平均的结构,相邻区域化变量的取值具有该结构所表达的相关关系。区域化变量的两大特点是随机性和结构性。基于此,地统计学引入随机函数及其概率分布模型为理论基础,对区域化变量加以研究。区域化变量可以看作是随机变量的一个现实(realization)。对于随机变量而言,必须在已知多个现实的前提下,才可以总结出其随机函数的概率分布。如向空中投掷一个色子,不能仅凭一次结果为 6,就可以推断每次投掷结果可能出现的值及其相应的概率。而对地学数据来讲,往往我们只有一些采样点,它们可以看作随机变量的一个现实,所以也没有办法来推断整个概率分布情况。为此,必须制定一些假设,即平稳性假设,假定在某个局部范围内空间分布是均匀的。

(1) 假定的局部范围内,变量的数学期望值为一常数,不依赖于点的空间位置

$$E\{Z(x)\} = m, \quad \forall x \quad (1)$$

(2) 协方差存在,且依赖于两点之间的距离 h

$$C(h) = E\{Z(x+h) \cdot Z(x)\} - m^2, \quad \forall x \quad (2)$$

* 收稿日期:2003-06-30;修回日期:2003-11-15.

* 基金项目:国家高技术研究发展计划(863 计划)项目“空间信息分析及原型研发”(编号:2002AA135230-1)资助.

作者简介:孙英君(1976-),女,山东省聊城市人,博士生,主要从事 GIS 空间分析方面研究. E-mail: sunyj@reis.ac.cn

这样,在空间某一局部范围内,对空间某一点 x_0 ,相距为 h 的多个点,可以看作是点 $z(x_0)$ 的多个实现,即可进行统计推断及估值预测。

1.2 理论核心——变异函数(variogram)

地统计学的主要用途,是研究对象空间自相关结构(或空间变异结构)的探测以及变量值的估计和模拟。不管哪一种用途,地统计学分析的核心是根据样本点来确定研究对象(某一变量)随空间位置而变化的规律,以此去推算未知点的属性值。这个规律,就是变异函数。样本点的变异函数计算公式为:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(u) - z(u+h)]^2 \quad (3)$$

其中, $N(h)$ 为距离相隔为矢量 h 的所有点对的个数。其核心思想是把所有的点对按照间隔距离的大小、方向进行分组,在每一个组内,计算每个点对属性值的差异,最后取平均作为该组属性值的差异(变异值)。这样,将整个空间分为不同大小和方向的组,并有相对应的属性差值。根据样本点计算某一未知点的属性值时,会考虑到多种不同距离、不同方向空间点位的相关关系。

通常,我们利用采样点及变异函数的计算公式(3)得出样本点的实验变异函数(experimental variogram),拟合后的曲线为经验变异函数。观察该变异函数的分布图像,寻找地统计学提供的某一种理论模型或者多个理论模型(basic model)的线性组合进行拟合。常见的理论模型有:线性模型、球状模型、指数模型、高斯模型、幂指数模型等。理论模型利用块金效应(nugget)、基台值(sill)以及变程(range)3个参数来描述研究对象的空间分布结构。块金效应指 h 为 0 时的变异函数值。理论上讲,该值应为 0。但由于测量误差的存在,以及我们所观测的尺度大于空间变异的细微尺度时,块金效应就不为 0 了。基台值指变异函数所达到的最大值(对某些基本变异函数,实际应用中取最大值 $\times 0.95$),即为采样点原点的方差值。变程描述了具备空间关联的范围,超出该范围,则不再具有相关关系。

变异函数的选取,不仅仅是将实验的点变异函数拟合为经验的模型变异函数曲线问题。用户需要根据自己的经验去选择变异函数的个数、类型以及基础变异函数模型各向同性的问题。

1.3 地统计学分析

运用地统计学进行空间分析基本包括以下几个步骤,即数据探索性分析,空间连续性的量化模型,未知点属性值的估计,对未知点局部及空间整体不

确定性的预测。用户可根据自己的需要截止到中间某一项。

数据探索性分析,主要是通过频率分布图、散点图、位置图等对数据的统计分布特征做一个初步的考察。这个过程最容易发现的问题就是数据的集聚,以及异常点极值的出现。通常,可利用适当的变换,如取对数变换来解决。这一阶段,本文不再赘述,将重点放在对空间连续性进行建模的诠释以及具体的发展方法上面(图 1)。

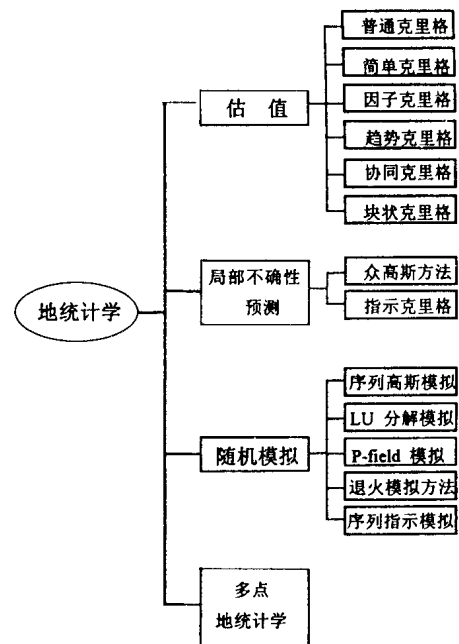


图 1 地统计学方法框架

Fig. 1 The framework of methods in Geostatistics

2 地统计学研究方法

2.1 估值(estimation)

地统计学最初应用是在矿产部门,作为矿产储量计算的基本方法取得了相当丰硕的成果。在地统计学领域,克里格(Kriging)是大家公认的估计方法的总称。实际上,它也是一种广义的最小二乘回归算法,而其最优目标定义为误差的期望值为 0,方差达到最小。所有的克里格估值可表示为:

$$Z^*(u) - m(u) = \sum_{i=1}^{n(u)} (u) [Z(u) - m(u)] \quad (4)$$

其中, (u) 为赋予样本点 $Z(u)$ 的权重,即通过变异函数计算的统计意义上的权重。 $m(u)$ 及 $m(u)$ 为 $Z(u)$ 及 $Z(u)$ 的均值。在每一个未知点 u ,根据一定的搜索半径及限制条件确定一个以 u 为中心点的邻域 $W(u)$,取该区域内所有样本点,并赋予相应权重,计算未知点的属性值。

根据所研究对象的不同趋势,克里格可分为 3 种类型:

(1) 简单克里格 (Simple Kriging): 认为均值 $m(u)$ 在整个研究地区 A 是已知的一个常数。

$$m(u) = m, \text{ 已知 } \forall u \in A$$

(2) 普通克里格 (Ordinary Kriging): 认为在局部限定的一个区域 $W(u)$ 内, 均值 $m(u)$ 是一个未知的常数, 对于整个区域来讲均值有一定变动。

$$m(u) = \text{未知常数 } \forall u \in W(u)$$

(3) 趋势克里格 (Kriging with a trend model): 认为在局部限定的每一个区域 $W(u)$ 内, 均值也是平滑地变化的, 因此对于整个区域来讲均值有一定变动。这时, 利用一个多项式来对趋势进行建模。

$$m(u) = \sum_{k=0}^K a_k(u) f_k(u)$$

其中, $a_k(u)$ 为未知常数, $\forall u \in W(u)$ 。

即在一个局部区域 $W(u)$ 内, 系数 $a_k(u)$ 是未知的, 且在该局部范围内认为是一个常数。这里 $f_k(u)$ 是坐标 x, y 的函数, 同时也可以是一种变量的函数, 见趋势克里格 (Kriging with an external trend) [15]。

由于所有的插值算法都趋向于属性值空间变异局部细节的平滑, 导致小值偏大, 大值偏小的问题。这种平滑依赖于局部数据的形状: 一般高频部分都随着未知点距离样本点越来越远而被滤掉。基于这种趋势, 可利用因子克里格方法将不同尺度, 亦即不同频率的空间变异提取出来。

(1) 因子克里格 (Factorial Kriging): 以上所论述的克里格方法, 都是对某一未知点属性的估计, 而因子克里格则是用以理解各种不同尺度影响的根源。如在一个研究区内, 金属 Cd (cadmium, 镉) 的变异函数如下

$$C_d(h) = 0.3 g_0(h) + 0.30 Sph \left[\frac{h}{200\text{m}} \right] + 0.26 Sph \left[\frac{h}{1.3\text{km}} \right] \quad (5)$$

这里, 描述了金属 Cd 在 3 种不同尺度的空间上的变化: 微小尺度 (类似于块金效应, 小于第一个步长值 50 m), 局域尺度 (小的变程约 200 m), 区域尺度 (变程约 1 km)。微小尺度变程结构对应在不同的岩石类型和土地利用类型上 Cd 的空间分布。而区域的变程结构则对应于不同的地质构造上金属 Cd 的分布。

利用多种尺度的变异函数, 因子克里格可以通

过协方差剔除将各种尺度对应的部分提取出来。具体算法见参考文献 [6]。

(2) 协同克里格 (Co Kriging): 如果所研究的主变量数据与其它变量有相关关系, 且其它变量的观测比较容易实现, 则可利用相关系数将二者的关系引入, 后者称为共协变量。利用多数的共协变量及少数的主变量采样点进行主变量的估值。共协变量一般应是全覆盖的, 即在所有采样点、未知点处都是已知的, 最低要求是与主变量采样点同位的共协变量是已知的。实践表明, 只有当主变量的采样点少于共协变量的采样点时, 相对于一般的克里格方法而言, 协同克里格方法是有意义的。

(3) 块状克里格 (Block Kriging): 在实践中, 有时我们的量化目标是某个属性值在特定尺寸——“块”上的平均值。如我们以 1 hm^2 的土地为一个研究单元, 考察在该范围内, 金属污染物 Cu 的估值为多少, 以确定该污染是否超标, 从而及时监督, 施行补救。这个研究单元——“块”, 称为支集, 它也可以是一条线段, 或像本例, 1 hm^2 的土地, 是一个曲面。在计算中, 考虑的是“块”内点的平均值, 以及不同“块”之间内部点与相关“块”之间的关联关系。

2.2 局部不确定性预测 (local uncertainty)

地统计学的估计功能主要是求得一个无偏的最优估值, 同时给出每个估值的误差方差, 用以表示其不确定性。如 95% 的置信区间为

$$\text{Prob}\{Z(u) \in [z^*(u) - 2\sigma_E(u), z^*(u) + 2\sigma_E(u)]\} = 0.95 \quad (6)$$

这种方法的优点是比较简单, 只需要主变量之间的关联关系。但其缺点是: 认为误差的分布是对称的, 但在实际情况中, 低值区往往被高估, 而高值区往往被低估。认为误差的方差只依赖于真实值的形状, 而不考虑具体每个值的影响, 即所谓的同方差性。但实际上被一个大值和小值包围的点, 其估值的误差一般要比被两个同规模小值包围估值点的误差要大。所以, 应确实考虑到所估计点周围样本点本身值的影响, 即利用条件概率模型

$$F(u; z | (n)) = \text{Prob}\{Z(u) \leq z | (n)\} \quad (7)$$

来推断不确定性。通常有两种方法: 参数法 (众高斯方法) 及非参数方法 (指示克里格方法)。

(1) 众高斯方法 (Multi Gaussian approach): 到目前为止, 这是应用最广泛的参数化方法。它假定所研究区域的概率分布可以用一个统一的公式表达, 最终的概率依赖于相关参数。对应于众高斯方法, 即是均值和方差。我们利用克里格方法来估计这两

个参数,同时利用光滑样本点频率分布图方式来平滑、增加其概率分布函数。

由于众高斯方法要求多点分布必须是标准正态的,且没有考虑极大值与极小值间的关联关系。对于样本点的指示变异函数不支持双高斯分布,或者作为关键的辅助信息与主变量之间不满足众高斯分布,这时需采用指示克里格方法。

(2) 指示克里格 (Indicator Kriging): 利用指示克里格方法估计未知点的不确定性,首要的一步是将各种来源的信息进行指示编码。即利用不同的阈值将原数据分为合适大小的间隔,考虑该间隔内点的关联关系及其不同的关联之间的关系。这样,就有效地解决了众高斯方法的缺点。

根据全局采用已知常数作为指示均值还是局部采用未知常数作为均值分为简单指示克里格方法和普通指示克里格方法。同时,不同阈值内的变量间也存在一定的相关性,这种关联可借助协同指示克里格方法来加入。但阈值的个数基本代表了所加入的共协变量的个数,这无疑加重了计算的负担。这时,可采用“概率克里格方法”。

(1) 概率克里格 (Probability Kriging): 即第二种信息的引入,不采用变换后的指示形式,而是原数据形式。但由于二者在尺度方面存在很大的差别,故利用阶转换形式。将源数据 z 利用其标准阶 $x(u_a) = r(u_a) / n$ 来代替,其中 $r(u)$ 为样本点累积分布中数据 z 的排序(阶)。

这样,可利用唯一的次变量信息来代替众多的阈值之间的信息,大大减少了计算量。

(2) 中值指示克里格 (Median IK): 指示克里格方法要求对 K 个指示变异函数进行估计和建模,在每个点 u 求解 K 个克里格方程。这种计算量可通过中值指示克里格方法大大的减少。但要想实行中值指示克里格,必须满足下列条件:

(a) K 个指示函数是内蕴关联的,即所有的指示直接及交叉变异函数均成一定比例,有如下关系:

$$\gamma_1(h; z_k, z_k) = \gamma_{kk} \cdot \gamma_{m1}(h) \quad \forall k, k \quad (8)$$

(b) 所有的已知精确硬数据矢量都是完整的,没有缺失而用间隔信息来替代的。这样,所有的克里格和协同克里格估值都是一样的。即

$$\begin{aligned} [F(u; z_k | (n))]_{m/K}^* &= \sum_{a=1}^{n(u)} O_K(u) I(u; z_k) = \\ &= [F(u; z_k | (n))]_{oICK}^* \quad (9) \end{aligned}$$

这样,对于中值指示克里格只需要对一个指示变异函数(一般是中间阈值处)进行建模,求解一个

克里格方程,然后按比例关系计算其它阈值范围的量。

有了以上各种求解未知点不确定性的方法,可以根据不同的目标得出符合特定要求的估值,而不是像克里格方法一样,必须是方差最小的目标。方法是利用一个损失函数,作为限制误差 $e(u) = z^*(u) - z(u)$ 的标准,由此求出最优估值。

主要包含类型:

(1) E-类估值:其标准为误差平方和最小

$$L(e(u)) = [e(u)]^2 \quad (10)$$

这种估值方法虽然其标准和克里格是一致的,但它的概率分布函数考虑到了样本点本身的数值信息。当样本点是标准正态分布的,二者是一致的,除此之外,二者是不同的。

(2) 中间估值:由于在 E 类估值中,取的是误差的平方,所以极大(小)值的影响非常显著。

为此,利用中间估值可有效消除极值的作用。它的最优估值为:

$$z_L^* = q_{0.5}(u) = F^{-1}(u; 0.5 / (n)) \quad (11)$$

(3) 分位估计:上面所述两种方法均考虑误差的大小,没将符号的影响包括在内。但在实际应用中,往往误差的分布是不对称的,其造成的后果也是不对称的。在分位估计中采用的是不对称的线性损失函数:

$$L(e(u)) = \begin{cases} 1 \cdot e(u) & \text{for } e(u) \geq 0 \text{ (高估)} \\ 2 \cdot |e(u)| & \text{for } e(u) < 0 \text{ (低估)} \end{cases} \quad (12)$$

地统计学在这一阶段的发展,表现在不确定性上。即充分考虑了各点的分布及其本身的属性值,再给出其不确定性。这对于笼统地给出一个置信区间,无疑是个很大的进步。以此为基础,可以求出不同目标函数的估值,相对于单一的克里格方法的目标函数而言,也体现了地统计学灵活性的增强。

2.3 随机模拟(simulation)

克里格方法完成了空间格局的认知,但没能使其再现。通过克里格方法,可以获得唯一的估计结果,而且极值点都被光滑下去。根据随机变量的定义,每个变量可以有多个现实,也就是说每个未知点的估值可以有多种情况,但前提是总体趋势的正确性,这种方法就是随机模拟。随机模拟可以利用各种不同类型数据(如“硬”的采样点数据,“软”的地震数据)再现已知的空间格局。“硬数据”指在采样点精确测量的变量值。“软数据”指关于该变量各种类型的间接测量值。随机模拟可以生成众多的现实,每一个现实展现同一种格局,但不同的表现方式。

在单变量分布模型中,通过随机变量的系列结果来统计其不确定性,与此类似,一系列随机产生的现实,作为模型的输入也可以表达输出结果的不确定性。这些随机现实是等概率的,即没有哪一个现实是最好的。

(1) 高斯序列模拟 (Gaussian Sequential Simulation):从数学上来讲,最初、最严格的随机模拟方法为高斯序列模拟。它克服了克里格方法平滑的效果,通过系列随机模拟现实表达由变异函数或柱状图量化的特定地质格局。克里格方法的不足之处在于单独的估计未采样点的属性值,而没有考虑该点与前面已经取得估值的各未知点的相关关系,显然,克里格方法无法再现修正后的空间关联关系。这也是其结果光滑性的原因所在。为确保这一点,需要定义所有栅格点属性值的联合概率模型,而不再像克里格方法中单个的考虑。联合分布定义为:

$$F(z_1, z_2, \dots, z_N) = Pr(Z(u_1), Z(u_2), \dots, Z(u_N) \leq z_1, z_2, \dots, z_N) \quad (13)$$

从该分布中生成一个样点要考虑所有点之间的空间相关性。

(2) LU 分解模拟 (LU Simulation):当要模拟的结点较少,所需的现实值又较多时,可采用 LU 分解模拟方法。这是一种通过对协方差矩阵进行 LU 分解以加快计算过程的模拟方法。在高斯序列模拟中,每一个现实都要计算 n 个克里格方程,以加入邻近的采样点和前面模拟过的值为条件。但 LU 分解,只需在第一次对整个协方差矩阵进行分解,以后更改模拟过的值相关关系即可。

(3) 高斯指示模拟 (Gaussian Indicator Simulation):序列高斯模拟不能考虑极值点的连续性及相关关联。在不同的属性值范围内关联是不一致的,所以采用个别范围个别变异函数来刻画其关联性的方法,即序列指示模拟。

(4) P-field 模拟 (P-field Simulation):利用分位数算法 $z^{(l)}(u) = F^{-1}(u; p^{(l)} | (n))$, $l = 1, \dots, L$, 对应不同的分位数,只要有概率分布函数 $F(u; z | (n))$,便可以得到一系列的 L 个模拟结果 $z^{(l)}(u)$, $l = 1, \dots, L$ 。如果采用的概率分布函数只以采样点的已知数据为条件,则未必能得出真实的 $C(u-u')$, 因为 u 点得到的概率分布函数没有考虑前面模拟过,但与它很相近的值 $z^{(l)}(u')$ 的关联。序列高斯模拟中,将原始的采样点及每点前面模拟过的未知点的模拟值作为估计概率分布函数的先决条件,这样在每个点,都需要计算概率分布函数,这样

会加大计算量,延长整个模拟的时间。而 P-field 模拟方法首先利用原始采样点数据估计出概率分布函数。然后利用系列随机自相关数 P 将相邻两点模拟值的关联关系引入到最后利用概率值从概率分布函数中抽取模拟值的计算中,这样只需要计算一次概率分布函数,大大节省了计算时间。

(5) 模拟退火方法 (Simulation Annealing) 模拟退火方法不再涉及到随机函数模型,而是一个最优化的过程。对某个问题来讲,一般都有一个近似的答案,以此为起点,逐步修正,得到满足约束条件的最优解。

可将其它模拟方法的结果作为该方法的原始图像的输入,利用一定的扰动机制进行模拟,直到满足目标函数为止。常用的目标函数有单点的概率分布函数、变异函数模型、指示变异函数模型、多点统计、相关系数、交叉变异函数。目标函数可以由多个部分构成,每个部分赋予不同的权重。在扰动修正过程中,每一部分及其权重都可以变化。

在这一阶段,地统计学最大的进步就是从全局出发,充分考虑了整个空间的不确定性,而不局限于某个子域。多个现实的结果,与克里格方法单一的结果对比,更方便评价结果的不确定性。特别是在作为某个模型的输入时,不同的现实得到的结果代表了模型描述事件各种可能出现的情况,在实践中颇为有用。

2.4 多点地统计学 (multi-point geostatistics)

多点地统计学的发展主要得益于地统计学在石油领域的应用。早期,地统计学多用于煤炭问题,通过块状估值得出可开采储量。但在对石油储区的研究中,人们发现单纯的某个点的渗透性是没有意义的,而应该以流的观点来看待渗透性问题。这就使得对渗透性的连通性或其空间格局的量化比得到某局部点的精确值更为重要,而不是光滑的估计。传统的地统计学借助于煤炭科学的思想,利用变异函数来量化空间格局。但变异函数只能度量空间上两个点之间的关联,所以表现空间格局有很大的局限性。对于关联性很强的情况,或所研究对象具备较为明显的曲线特征,这时要想量化其空间格局需要包含多个空间点。在图像分析中,通过多点模板或者窗口来量化其格局。意识到变异函数在表达地质连续性上的局限性后,地统计学家将图像分析中的思路借鉴过来,一个新的领域在地统计学中升起:多点地统计学。

原本地统计学模拟包括认知和再现两部分。认

知通过变异函数来完成,而再现通过序列高斯模拟的多个现实来完成。多点地统计学进一步改善了认知部分,即通过多个点的训练图像来取代变异函数,更有效地反映了研究目标的空间分布结构。而对于图像分析而言,它只注重认知部分,但没有再现功能。

多点地统计学的核心是训练图像。由于在地统计学中也出现过多点信息,但从未被量化过,而一般是将信息隐含的应用到具体问题模型中去。但如通过图像的方式,可全面量化原数据各阶的信息,因此我们可采用非条件的布尔方法得到训练图像再进行分析^[3]。

这种方法主要是在由于石油领域的问题引出,因此也主要应用在这个领域。包括理论本身,还有待于进一步完善。

3 地统计学软件进展

地统计学理论上的发展、应用远远要超过其软件的发展,使得新方法的推广由于软件的局限性受到了很大的限制。如目前地统计理论热点,正则化、模拟或随机模拟等已相对成熟,但其软件却相对落后。现有软件如 Geovariance、GsLib、Gstat、GS⁺等,涉及到模拟分析的仅限于极少数,且大都基于 Dos 或者 Unix 系统,不具有通用性。而涉及到的正则化,至今还没有发现。此外,作为一种空间数据的分析工具,地统计学软件在高质量图像表达方面还远远不够^[30]。

GIS 是对空间数据进行搜集、存储、检索、转换、显示及分析的一门技术。它可以将具有地理坐标的数据信息作为一个专题层,或地图文档来进行管理。作为一个强大的数据库系统,它可以存储具有同样空间范围的多种专题信息。编辑、操作这些空间数据,对于现有的 GIS 软件已不成问题,但对空间数据分布格局进行建模,抽取其特征还很欠缺。这就需要像地质统计 (geostatistics) 这类空间分析的统计软件包。地统计学近年来在国际上发展迅猛,特别是 GIS 的发展,对空间分析功能提出了一个新的要求,使得地统计学成为多个学科重视的焦点。但到目前为止,二者之间的结合还很少,或非常欠缺。如大型软件 ArcGIS,从 8 版本以后加入了扩展模块,其中即有地统计学。但内容仅限于克里格系列方法,而对于模拟方法还是一片空白。所以,未来将两者结合起来将是一种必然的趋势。一种比较快捷的方式是利用组件式思想,将地统计学软件内嵌到 GIS 软件内部。这种结合方式要考虑到两个原本不同系统的融合,所

以稍显繁琐,且二者关系较为松散。但针对目前强大的需求,这无疑是一种多快好省的方法。

4 结 论

(1) 地统计学的研究方法包括局部估值、不确定性预测、随机模拟及多点地统计学四部分。从单一的估计局部未知点的最优值,到考虑整体空间关联的随机模拟,从特定的目标求解估值,到根据需要确定目标函数求解估值,从两点的变异函数到多点的训练图像,地统计学在向着更完美的展现空间分布,更贴切地满足用户需求方向发展。

(2) 地统计学软件的发展需要进一步的可视化研究,高质量的图形输入、输出,以及空间分析的需求,使得地统计学和 GIS 的结合成为一种必然的趋势。

参考文献 (references) :

- [1] Andrew M Liebhold, et al. Geostatistics and geographic information systems in applied insect ecology[J]. *Annual Review of Entomology*, 1993, 38: 303-327.
- [2] Bhatti A U, Mulla D, Frazier B. Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and Thematic Mapper images [J]. *Remote Sensing of Environment*, 1991, 37: 181-191.
- [3] Caers J. Geostatistics: From pattern recognition to pattern reproduction[A]. In: Nikravesh M, Aminzadeh F, Zadeh L, eds. *Soft Computing and Intelligent Data Analysis in Oil Exploration*[C]. Elsevier Publisher, 2001. 1-22.
- [4] Caers J, Zhang T. Multiple-point Geostatistics: A Quantitative Vehicle for Integrating Geologic Analogs into Multiple Reservoir Models [M]. *Integration of Outcrop and Modern Analogs in Reservoir Modeling*, AAPG Memoir, 2002. 1-24.
- [5] Caers J. History matching under training-image based geological model constraints[J]. *SPE*, 2002, 77429: 1-16.
- [6] Deutsch Journel. *GSlib-Geostatistical Software Library and User's Guide* (2nd) [M]. New York: Oxford University Press, 1998.
- [7] Dungan J. Spatial prediction of vegetation quantities using ground and image data[J]. *International Journal of Remote Sensing*, 1998, 19: 267-285.
- [8] Frykman P, Deutsch C V. Geostatistical scaling laws applied to core and log data[J]. *SPE Paper Number 56822*, Houston, TX, 1999: 887-898.
- [9] Goovaerts P. Factorial Kriging analysis: A useful tool for exploring the structure of multivariate spatial soil information[J]. *Journal of Soil Science*, 1992, 43: 597-619.
- [10] Goovaerts P. Study of spatial relationships between two sets of variables using multivariate geostatistics[J]. *Geoderma*, 1994, 62: 93-107.
- [11] Goovaerts P. Geostatistical tools for characterizing the spatial

- variability of microbiological and physics-chemical soil properties [J]. *Biology and Fertility of Soils*, 1998, 27(4):315-334.
- [12] Goovaerts P. Accounting for scale-dependent correlation in the spatial prediction of soil properties[A]. In: Gomez-Herenandez J, Soares A, Froidevaux R, eds. *geoENV - Geostatistics for Environmental Applications* [C]. Dordrecht: Kluwer Academic Publishers, 1999. 405-416.
- [13] Goovaerts P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall[J]. *Journal of Hydrology*, 2000, 228:113-129.
- [14] Goovaerts P. Geostatistical mapping of satellite data using p-field simulation with conditional probability fields[A]. In: Heuvelink GB M, Lemmens MJ P M, eds. *Proceeding of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*[C]. Amsterdam:2000. 253-260.
- [15] Goovaerts P. *Geostatistics for Natural Resources Evaluation* [M]. New York: Oxford University Press, 1997.
- [16] Isaaks E H, Srivastava R M. *Applied Geostatistics*[M]. New York: Oxford University Press, 1989.
- [17] Journel A G. *Fundamentals of geostatistics in five lessons. Volume 8 Short Course in Geology*[C]. American Geophysical Union, 1989.
- [18] Journel A G. Conditioning geostatistical realization to non-linear volume averages[J]. *Mathematical Geology*, 1999, 31: 955-964.
- [19] Journel A G. Combining knowledge from diverse information sources: An alternative to traditional data independence hypotheses[J]. *Mathematical Geology*, 2002, 34(5): 573-596.
- [20] Journel, Huijbregts. *Mining Geostatistics*[M]. London: Academic Press, 1978.
- [21] Matheron B. *The Theory of Regionalized Variables* [M]. Fontainebleau: Centre de Morphologie Mathematique, 1971.
- [22] Rao S, Journel A. Deriving conditional distributions from ordinary Kriging[A]. In: Baafi E Y, Schofield N A, eds. *Geostatistics Wollongong '96 (Vol. 1)* [C]. Dordrecht, The Netherlands, Kluwer:1997. 92-102.
- [23] Strebelle S. *Sequential Simulation Drawing Structure from Training Images*[D]. Stanford University, 2000.
- [24] Strebelle S. Conditional simulation of complex geological structures using multiple-point statistics[J]. *Mathematical Geology*, 2002, 34: 1-22.
- [25] Wakernagel H. Geostatistical techniques for interpreting multivariate spatial information[A]. In: Chung C F, Fabbri A G, Sinding-Larsen R, eds. *Quantitative Analysis of Mineral and Energy Resources*[C]. Redel, Dordrecht, 1998. 393-409.
- [26] Bai Yanchen(柏延臣), Li Xin(李新). Spatial data analysis and spatial models[J]. *Geographical Research(地理研究)*, 1999, 18(2):185-190(in Chinese).
- [27] Wang Jinfeng(王劲峰), Li Lianfa(李连发). A theoretic framework for spatial analysis[J]. *Acta Geographica Sinica(地理学报)*, 2000, 55(1):92-103(in Chinese).
- [28] Wang Jinfeng(王劲峰), Bai Yanchen(柏延臣). Research and development of spatial analysis functions in GIS[J]. *Journal of Image and Graphics(中国图象图形学报)*, 2001, 6(A)(9): 849-853(in Chinese).
- [29] Xiao Bin(肖斌), Zhao Pengda(赵鹏大), Hou Jingru(侯景儒). New development of geostatistics[J]. *Advance in Earth Sciences(地球科学进展)*, 2000, 3: 293-296(in Chinese).

STUDY ON PROGRESS OF METHODS IN GEOSTATISTICS

SUN Ying-jun¹, WANG Jin-feng¹, BAI Yan-chen²

(1. *Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;*

2. *Tsinghua University, Beijing 100084, China*)

Abstract: Geostatistics has been an important tool for spatial analysis recently, it has been applied many fields of natural science, such as geography, ecology, soil science and so on. The paper discusses the progress in the theory method of Geostatistics. The whole framework has been pointed out, including estimation, local uncertainty, stochastic simulation and multi-point simulation. Especially the new methods, for example Sequential Gaussian Simulation, LU Simulation, P-field simulation, Annealing Simulation are described. The paper analyzes the difference between multi-point simulation with the traditional geostatistics methods. That is the breakthrough of two-point method-variogram to training image to describe the spatial structure of objects. Meanwhile, the paper makes a comment on the software of Geostatistics. The need of graphic user's interface, high quality image input and output, together with the absence of spatial analysis function, caused the combination of Geostatistics and GIS is an inevitable trend.

Key words: Geostatistics; Kriging method; Stochastic Simulation; Multi-point Simulation.