文章编号：1671-8860(2008)05-0441-05　　文献标志码：A

野外空间采样的渐进式策略

张景雄[1]　Michael F Goodchild[2]

（1　武汉大学遥感信息工程学院，武汉市珞喻路129号，430079）
（2　加利福尼亚大学圣芭芭拉分校地理学系，美国CA 93106-4060，USA）

摘　要：介绍了有关地统计学(geostatistics)的基本理论，论述了顾及样本点位之间空间依赖性的空间采样方案的设计。基于总体上极大限度地降低克里金方差的思想，阐述了一个快速确立采样点位的序贯算法和块段克里金方法。由于野外作业的计算环境所限，数值方法和启发式搜索相结合的策略才是开发实用系统的明智选择。以假想数据为例，验证了所提算法的有效性。

关键词：空间采样；协变量；块段克里金法；协同克里金法

中图法分类号：P237.3

地统计学在空间信息科学中具有重要地位，也给空间采样方案设计提供了理论基础。传统的采样是非空间的，即不考虑样本所在的空间位置；空间采样设计时，应考虑采样点位的分布。在实际工作中，采集数据的间隔不能太大，也不能太小，间距越小，采样成本越高；反之，采样间距越大，则成本越小，但数据不一定能反映研究区的真实特征。因此，野外采样通常是以动态方式进行的，在保证成本尽量小的前提下，降低估计量的方差。对于某一个空间区域，分散在一定距离内的点位的某些特性通常会呈现空间相关性，即空间依赖性[1]。也就是说，在一定邻域范围内的采样点是相关的，它们所提供的信息有冗余。因此，在空间采样设计时，须考虑采样点间的相关性，使不确定性达到最小。

在已有一组采样点的前提下，如何确定后续采样点的位置，并且分析整个问题域的克里金方差和实施效率/成本随采样点累积的动态效应，是本文的主旨所在。在野外采样实施过程中，由于计算机软、硬件和网络传输等方面都受到很大的限制，因而需要研究一种高效、优化的算法。

1　基于地统计学的空间预测估计：克里金法(Kriging)[1]

地统计学在空间问题求解方面已经得到越来越广泛的应用。常用于空间预测的一个重要方法就是克里金法，该方法在给出最优线性无偏估值的同时，还可计算出估计值的量化评价指标——估计方差，即克里金方差(Kriging variance)。

克里金法是一种基于最优权重的空间估值方法，位于x处的变量Z的克里金估值是由周围一定区域内采样点的已有数据经线性加权组合得到的。如图1所示，假设有一样本由n个采样点组成，各采样点的观测值为$z(x_i)(i=1,2,\cdots,n)$，观测值的均值为m_z，则未采样的位置x的克里金估计值$z(x)$为：

$$z(x)=m_z+\sum_{i=1}^{n}\lambda_i(z(x_i)-m_z) \tag{1}$$

式中，$\lambda_i(i=1,2,\cdots,n)$对应邻近搜索区域内位于$x_i(i=1,2,\cdots,n)$点位的样本权重系数。将$\lambda_i(i=1,2,\cdots,n)$组成的列向量记为$\boldsymbol{W}$，则$\boldsymbol{W}$可由下式解算：

$$\mathrm{cov}_z(x_i x_j)\boldsymbol{W}=\mathrm{cov}_z(x x_i) \tag{2}$$

估计值$z(x)$的不确定性可用克里金方差来衡量：

$$\sigma_z^2(x)=\mathrm{cov}_z(0)-\boldsymbol{W}^{\mathrm{T}}\mathrm{cov}_z(x x_i) \tag{3}$$

式中，$\mathrm{cov}_z(0)$是在所求问题范围内变量Z的方差。从式(3)可以看出，克里金方差与具体采样点的观测值无关。克里金方差$\sigma_z^2(x)$主要依赖于n个观测点x_s组成的采样方案和变异函数模型。

收稿日期：2008-03-17。

项目来源：国家973计划资助项目(2007CB714402)。

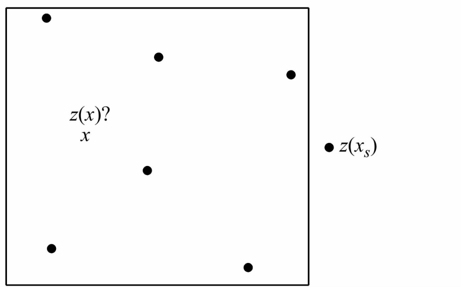


图1 用采样点集$\{z(x_s)\}$计算未采样点x的克里金估计值$z(x)$

Fig. 1 Estimating $z(x)$ by Kriging at an Unsampled Location x Using a Sample Data Set $\{z(x_s)\}$

在调查采样过程中,可以利用已有的数据资料作为辅助数据,如现存的土地覆盖图、航空影像或者卫星数据等,以减小作业成本。若将这些辅助数据作为空间协变量(covariates),可采用协同克里金法(co-Kriging)[1,2],其克里金方差应该降低。

假设有一个已知的主采样数据集$\{z_1(x_{s_1})\}$和一个次级采样数据集$\{z_2(x_{s_2})\}$,图2是这两个数据集的叠置图。在图2(a)中,主采样数据集和次级采样数据集都是不规则分布的;在图2(b)中,主采样数据集是不规则分布的,次级采样数据集是规则分布的。数据集$\{z_2(x_{s_2})\}$在两个图中分别用"+"和"□"表示。

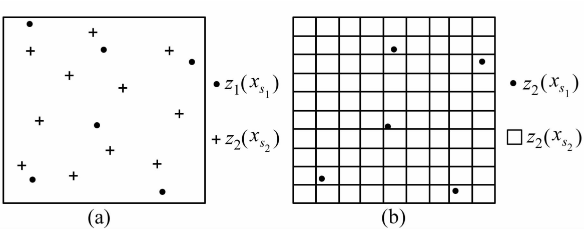


图2 用协同克里金方法,据主采样点集$\{z_1(x_{s_1})\}$和次级数据点集$\{z_2(x_{s_2})\}$采样数据估计$z_1(x)$的值

Fig. 2 Estimating $z_1(x)$ by Cokriging Primary $(z_1(x_{s_1}))$ and Secondary $(z_2(x_{s_2}))$ Sampling Data

令主采样点集和次采样点集的样本总数分别为n_1和n_2,那么在一个未采样点x处的协同克里金估计值$z_1(x)$可以用简单协同克里金法(simple co-Kriging)计算:

$$z_1(x) = m_{Z_1} + \sum_{i=0}^{n_1}\lambda_{1i}(z_1(x_{1i}) - m_{Z_1}) + \sum_{k=0}^{n_2}\lambda_{2k}(z_2(x_{2k}) - m_{Z_2}) \tag{4}$$

其中,m_{Z_1}和m_{Z_2}分别代表变量Z_1和Z_2的均值;$x_{1i}(i=1,\cdots,n_1)$、$x_{2k}(k=1,\cdots,n_2)$分别代表主采样位置和次采样位置;λ_{1i}、λ_{2k}分别对应主采样位置和次采样位置样本的权重系数。将λ_{1i}组成的列向量记为$\boldsymbol{W}_1$,将λ_{2k}组成的列向量记为$\boldsymbol{W}_2$,则它们可用下式解算:

$$\begin{bmatrix} \mathrm{cov}_{Z_1}(x_{1i}x_{1j}) & \mathrm{cov}_{Z_1Z_2}(x_{1i}x_{2l}) \\ \mathrm{cov}_{Z_2Z_1}(x_{2k}x_{1j}) & \mathrm{cov}_{Z_2}(x_{2k}x_{2l}) \end{bmatrix}\begin{bmatrix} \boldsymbol{W}_1 \\ \boldsymbol{W}_2 \end{bmatrix} = \begin{bmatrix} \mathrm{cov}_{Z_1}(x_{1i}x) \\ \mathrm{cov}_{Z_2Z_1}(x_{2k}x) \end{bmatrix} \tag{5}$$

对于简单克里金方法来说,结合每一个估计值,简单协同克里金方法同样也可以得到一个方差估计,

$$\sigma_{Z_1}^2(x) = \mathrm{cov}_{Z_1}(0) - \boldsymbol{W}_1^{\mathrm{T}}\mathrm{cov}_{Z_1}(xx_{1i}) - \boldsymbol{W}_2^{\mathrm{T}}\mathrm{cov}_{Z_1Z_2}(xx_{2k}) \tag{6}$$

2 基于总体克里金方差最小准则的采样方案设计

Barnes和Watson[3]设计了一个高效更新克里金方差的方法。用简单克里金方法可获得所有已知n个样本的权重系数组成的权重列向量$\boldsymbol{W}$,若在点$x_{s'}$位置加入新样本s',则对未采样点x位置的值作出估计时,x位置的新克里金估计方差的计算为:

$$\sigma_Z^2(x \mid (n+1)) = \sigma_Z^2(x \mid (n)) - B_1(x,x_{s'})^2 / B_2(x_{s'}) \tag{7}$$

B_1和B_2分别为:

$$B_1(x,x_{s'}) = \mathrm{cov}_Z(xx_{s'}) - \boldsymbol{W}^{\mathrm{T}}\mathrm{cov}_Z(xx_s) \tag{8}$$

$$B_2(x_{s'}) = \mathrm{cov}_Z(0) - \boldsymbol{W}^{\mathrm{T}}\mathrm{cov}_Z(x_{s'}x_s) \tag{9}$$

式中,$\mathrm{cov}_Z(xx_{s'})$是点x和第$n+1$个加入点$x_{s'}$的协方差;$\boldsymbol{W}^{\mathrm{T}}\mathrm{cov}_Z(xx_s)$代表点$x$和已经存在的$n$个样本点的协方差加权和;$\mathrm{cov}_z(0)$代表变量$Z$的方差;$\boldsymbol{W}^{\mathrm{T}}\mathrm{cov}_Z(x_{s'}x_s)$代表新加的备选采样点$x_{s'}$和已经存在的$n$个样本点的协方差加权和。

新的最优采样点的判断准则可描述为:在空间上,若在某位置能使式(7)中的估计方差最小,或者使二次式B_1^2/B_2取最大值,则该位置即为新的最优采样点的位置。当采样点为规则格网分布时,克里金方差均值和克里金估计值都在格网节点处得到,则通过式(7)得到的一系列离散值就可以作为选择新的最优采样点的判别依据。根据式(7),可以将二次式B_1^2/B_2的总和写成一个目标函数$\Phi(x_{s'})$:

$$\Phi(x_{s'}) = -\sum_{x\in A}B_1(x,x_{s'})^2/B_2(x_{s'}) \tag{10}$$

用常规方法计算目标函数$\Phi(x_{s'})$的最小值(或者它的负数的最大值)比较困难。Powell[4]设

计了一种基于共轭梯度的高效最小化算法，但是计算一阶导数成本比较高。Press[5] 提供了一个基于共轭梯度的稳健算法，但是在应用时需要一个在式(8)、式(9)中估计克里金权重系数的附加算法，空间采样效率比较低。

模拟退火(simulated annealing)等优化算法也可用作空间优化工具，该方法非常适宜解决大尺度问题，特别是当有很多局部极值而难以判断总体极值时。但是模拟退火算法的计算量很大，这就意味着在野外用这种方法搜索采样点的花费很高。因此，需要设计一个计算效率高的空间采样算法，以满足野外作业的要求。对于一定的研究区域，使用适当的格网对其进行离散化处理，可以合理减少备选采样点的数量，从而节省用于采样设计的计算量。通常用总体克里金方差作为衡量最优空间采样的标准，因此，可以把关注重点放在克里金方差的均值上，用块段克里金(block Kriging)[1]方法，以节省搜索运算的时间。

进行空间预测一般都要有一个假设，即隐含了假设空间样本有明确的大小。假设有一个样本集，在点 x 处的变量 Z 的克里金估计值在图 3(a) 中用圆圈表示，根据式(1)，这个变量 Z 的估值可以根据图 3(a) 中标为黑点的样本数值的线性组合求得。图 3(a) 的离散图指示了灰色遮蔽斑块的模拟点的大小范围，可以用不同的单元网格大小表示不同的分辨率，如果分辨率降低一倍时，单元网格大小可以设置为原样本大小的双倍。对于一块区域 v_x，其中心点为 x(图 3(b) 灰块中的“+”)，通常把这个区域离散化为 N_v 个规则点阵(图 3(b) 灰块中的圆圈点)。利用已有的采样点(图 3(b)中的黑点)的观测值，分别在 N_v 个离散定位点上进行克里金估计运算，然后取它们的平均值就能算出在点 x 处的块段均值的估计值：

$$z_1(v_x) = \frac{1}{N_v}\sum_{i=1}^{N_v} z_1(x_i) \tag{11}$$

对于简单的克里金方法来说，块段克里金方法如下：

$$\mathrm{cov}_{Z_1}(x_i x_j)\boldsymbol{W}_v = \mathrm{cov}_{Z_1}(v_x x_i) \tag{12}$$

式中，$\mathrm{cov}_{Z_1}(v_x x_i)$代表 N_v个离散点 s_i和数据点 x_i之间的协方差均值。如上面所谈到的基于点的简单克里金的情况，块段 v_x的克里金方差为：

$$\sigma_Z^2(v_x) = \mathrm{cov}_Z(v_x v_x) - \boldsymbol{W}_v^{\mathrm{T}}\mathrm{cov}_Z(v_x x_i) \tag{13}$$

式中，$\mathrm{cov}_Z(v_x v_x)$表示 v_x区域内任意两个点之间的协方差均；$\mathrm{cov}_Z(v_x x_i)$表示点与块的协方差均值：

$$\mathrm{cov}_Z(v_x v_x) = \frac{1}{N_v N_v}\sum_{k=1}^{N_v}\sum_{l=1}^{N_v}\mathrm{cov}_Z(x_k x_l)$$
$$\mathrm{cov}_Z(v_x x_i) = \frac{1}{N_v}\sum_{k=1}^{N_v}\mathrm{cov}_Z(x_i x_k) \tag{14}$$

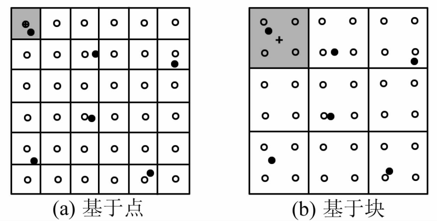


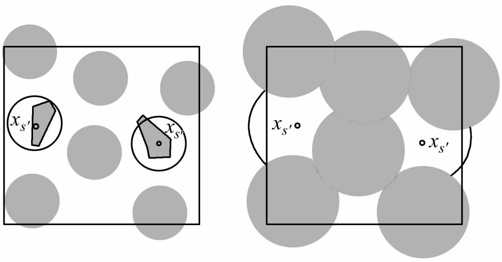
图 3　基于不同区域的克里金方法

Fig. 3　Kriging Based on Different Supports

很明显，系统地搜索备选采样点，然后执行块段克里金方法，可以极大地减小计算量，也易于计算方差的减少。当然，块段的大小要根据研究区域的特点(其几何特征、空间变化等)来决定。如可以把整个问题区域作为一个研究块来执行克里金插值，但是由于要考虑采样设计中的效率和克里金空间的预测分辨力，所以，一般都是采用中等大小的块段来执行块段克里金方法。

在渐进式采样过程中，采用启发式的搜索方法能进一步提高效率。通过启发式的方法能找到好的解决办法，但是不一定能保证找到全局最佳的解决办法。有一种启发式算法就是以最大距离法在备选采样点中搜索下一个采样点的。假设下一个定位点在点 $x_{s'}$，如图 4 中圆圈点所示，已经采样过的点用黑点表示。由地统计学原理可知，当超过变异函数模型定义的变程(range)时，空间位置的相关性就会消失，如图 4 中以黑点为中心的阴影圆所覆盖的范围所示。如果在这些圆内继续采样，它们对减小克里金方差的作用不大，因为它们和阴影圆中心点提供的信息很多是相似的。因此，为了使接下来的采样点发挥最大作用，新加的采样点应该离已有的采样点尽量远。如图 4 所示，圆圈点 $x_{s'}$ 是下一个最为经济的点，而且它能最大限度地减小克里金方差。变异函数模型范围所覆盖的影响区域的圈定意味着搜寻空间的减少，如图 4(a)中画阴影线的不规则区域。如果空间依赖性越大，则所需的新增采样点越少，在图 4(b)中仅需要再增加两个采样点即可。

当涉及多目标的最优化时，启发式搜索的优点更显著。假设备选采样点与当前采样点的估计方差和距离是用于确定下一个采样点的位置的两个准则(第二个准则主要反映从一个点移到一个



(a) 变程较小　　(b) 变程较大

图4　采用不同的变异函数模型，最大化地减小克里金方差来搜索下一个采样点 $x_{s'}$

Fig. 4　Searching for a Further Sample Location $x_{s'}$ Enabling Maximal Reduction in Kriging Variance with Variogram Models

新采样点的量测花费)，因此，有可能设计两个准则的线性组合，使采样花费最小化，同时最大化地减小估计方差。

3　试　验

为了检验本文所描述的自适应和渐进式采样策略的有效性和可行性，在FORTRAN开发平台上对GSLIB中kb2d(克里金模块)和newcokb3d(协同克里金模块)进行二次开发，设计出基于克里金和协同克里金方法的空间采样平台。编制了启发式搜索预处理子程序，可望在前期处理中节省搜索时间。需要说明的是，式(10)仅应用于简单克里金方案。

首先，使用一个假设的例子来检验搜索算法的性能。开始时，有9个采样点在研究区域内，如图5所示，用小三角标记。渐进式采样的解决方案可以由图中数字表明的顺序表述出来，这些数字表明它们依次被采样的先后顺序。所采用的球面变异函数模型的变程为5，基台(sill)为6。点克里金方法和块段克里金方法的计算所用的时间

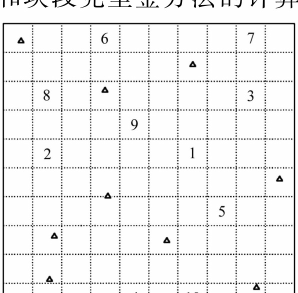


图5　一个假设的初始样点(用小三角表示)和继续采样得到的点(用数字表示)

Fig. 5　A Hypothetical Example Showing Initial Sampled and Suggested Further Sampling Locations

见表1，这些数据是在Unix系统正常的工作量情况下得出来的，是定位采样点所花时间的平均值。这里评价了两种格网系统，一种是10×10的格网平面，每个格网单元为1×1；另一个是20×20的格网平面，每个格网单元为0.5×0.5。

表1　渐进式采样的时间计算/(s·点$^{-1}$)

Tab. 1　Computation Time for Progressive Sampling

方案	点克里金法		序贯算法	块段克里金法	
	简单克里金	协同克里金	简单克里金	简单克里金	协同克里金
10/次	13	25	2	0.8	1.0
20/次	150	520	81	19	20

由表1可以看出，点克里金方法速度最慢，同时，点克里金中，协同克里金要比简单克里金方法花费的时间更多。对简单的克里金来说，序贯算法在两种格网平面上都能很好地减少运算所用的时间。最快的解决方案是块段克里金方法，主要是因为它把整个问题域当作一个单独的块段来处理。更有趣的是，块段协同克里金方法和块段简单克里金方法所用的时间是差不多的，尽管协同克里金方法要复杂些。可以推断，系统地搜索备选采样点，然后执行块段克里金方法，可以极大地减小计算量。在较大范围的区域应用中，搜集采样点较多，应用这个方法更有意义[8]。应当指出的是，块段克里金方法仅仅适合于全局性的克里金方差估计和方差减小的计算，而且在划分的块内，每一个离散化的点有相同的数据格局。

4　结　语

本文提出的方法适用于计算设备和数据传输环境受局限的空间采样设计。试验证明了空间渐进式策略的正确性和有效性，在野外作业环境下具有较高的使用价值和应用潜力。为了研究的方便，本文仅考虑了方差的问题。在实际操作中，可能还需要考虑经济因素、后勤保障等因素。在数据和知识对确定稳健的变异函数模型来说不是很充足时，变异函数模型的构建和更新更重要，特别是在野外采样的早期。即使现有数据较充足，所拟合的变异函数模型较可靠，在采样和知识累积过程中，仍然需要更新和调整变异函数模型。

致谢：感谢李德仁院士、刘经南院士和同事一如既往的支持和帮助，感谢Phaedon Kyriakidis教授和Jiefeng Zhou博士的帮助。

参　考　文　献

[1]　Journel A G, Huijbregts C J. Mining Geostatistics

[M]. London: Academic Press, 1978

[2] McBratney A B, Webster R, Burgess T M. The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables: 1. Theory and Method [J]. Computers and Geosciences, 1981(7): 331-334

[3] Barnes R J, Watson A G. Efficient Updating of Kriging Estimates and Variances [J]. Mathematical Geology, 1992, 24(1): 129-133

[4] Powell M J D. An Efficient Method for Finding the Minimal of a Function of Several Variables Without Calculating Derivatives [J]. Computer Journal, 1964 (7): 155-162

[5] Press W H, Teukolsky S A, Vetterling W T. Numerical Recipes in C: the Art of Scientific Computing[M]. Cambridge: Cambridge University Press, 1993

[6] Hammersley J M, Handscomb D C. Monte Carlo Methods [M]. London: Methuen, 1964

[7] Besag J E. Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion) [J]. Journal of the Royal Statistical Society, 1974, 36: 192-236

[8] Goodchild M F, Zhou Jiefeng. Finding Geographic Information: Collection-level Metadata [J]. GeoInformatica, 2003, 7(2): 95-112

第一作者简介：张景雄，教授，博士生导师。现从事地球空间信息科学与技术、摄影测量与遥感、地统计学等领域的科研与教学工作。主要研究兴趣包括空间信息集成处理、尺度模型、遥感信息真实性检验、数据同化等。

E-mail: jxzhang@whu.edu.cn

Towards Progressive Strategies for Spatial Sampling in the Field

ZHANG Jingxiong[1]　*Michael F Goodchild*[2]

(1　School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2　Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA))

Abstract: The geostatistical basis for adaptive/progressive sampling is discussed, following an introduction to the necessary statistical background and developments in geographic information technologies. Where computational resources are limited, as they are in the field, strategies that combine heuristic and numerical approaches are the key to successful field implementation. A sequential algorithm for rapid location of further samples is formulated, using the criterion of maximum global reduction in Kriging variance. Results from a test confirm the effectiveness of the proposed algorithms.

Key words: spatial sampling; covariate; block-Kriging; co-Kriging

About the first author: ZHANG Jingxiong, professor, Ph.D supervisor. He is engaged in scientific research and teaching on geographical information science & technology, photogrammetry and remote sensing, and geostatistics. His main research interests include spatial information fusion, scale, validation in remote sensing, and data assimilation.

E-mail: jxzhang@whu.edu.cn