

地理空间抽样理论研究综述

姜成晟^{1,2}, 王劲峰¹, 曹志冬³

(1. 中国科学院地理科学与资源研究所, 北京 100101;

2. 中国科学院研究生院, 北京 100049;

3. 中国科学院自动化研究所, 北京 100080)

摘要: 抽样调查是地理研究、资源评估、环境问题研究和社会经济问题研究的重要手段。对于地理分布的各种资源, 由于调查数据往往具有空间相关性, 传统的抽样调查理论无法满足日益增长的空间抽样需求。空间抽样理论是对具有空间相关性的各种资源和调查对象进行抽样设计的基础。本文详细论述了空间抽样理论发展现状。首先介绍了空间抽样的产生和发展, 以及空间抽样所要研究的四个问题。然后介绍了基于设计的和基于模型的抽样统计推断方式, 以及它们适用的范围。最后本文详细论述了 Kriging 理论在抽样理论的应用、前向、后向和双向样本布局方法和六种空间抽样样本优化选择标准。

关键词: 地理空间; Kriging 抽样; 抽样调查

1 概述

1.1 空间抽样的发展

在 1895 年瑞士首都伯尔尼召开的国际统计学会 (ISI) 第五次大会上, 挪威人凯尔 (A. N. Ciel) 的报告—《对代表性调查的研究和经验》, 正式提出使用代表性样本的调查方法取代全面调查。

地统计学最早是矿物学家 D.R.krige 将其应用于南非金矿的查找, 这个方法是由 Matheron 提出来的^[1, 2]。七十年代提出了托普勒第一定律: 任何事物之间都有相关性, 相距近的事物比相距远的事物之间更加相关^[3], 对这种相关性的研究和量化构成了空间统计理论的基础, 一大批学者对空间相关性和空间变异等问题做了大量的研究^[4-9], 奠定了空间统计、空间数据分析的基础, 基于样本不独立假设的空间抽样调查技术得以迅速发展, 在生态^[10]、海洋^[11]、渔业^[12]、林业^[13]、农业^[14]、人口健康调查^[15]、环境^[16]、土壤^[17]以及水资源^[18]等方面得到了广泛的应用。在国内, 绝大多数抽样调查都是基于经典抽样^[19-22], 王劲峰研究员及其所带领的研究团队是较早地开展空间抽样技术方面研究的学者^[23-26], 也有学者在土地调查抽样设计时考虑了空间结构^[27]。

传统抽样中, 抽样对象一般不具有空间位置概念。在抽样时, 每个对象被分配一个编号以便于区别不同的抽样对象。但是对换任意 2 个编号对象, 并不影响抽样结果。空间抽样和传统抽样技术最大的差别就在于, 空间抽样调查对象具有地理空间坐标, 并且如果改变空间对象的位置, 则改变了抽样调查总体, 影响抽样调查结果。空间抽样调查对象的地理空间坐标和属性值共同构成标识空间对象的二个特征属性。空间抽样理论对这二个属性之间关系进行研究。下面将介绍空间抽样涉及的几个问题。

收稿日期: 2008-10-28; 修订日期: 2008-12-29

基金项目: 国家自然科学基金 (40471111; 70571076); 863 课题(2006AA12Z205)资助 [Foundation: National Natural Science Foundation of China, No.40471111; No70571076; 863 Program, No.2006AA12Z205]

作者简介: 姜成晟 (1981-), 男, 博士, 主要研究方向空间抽样理论、空间数据分析与建模。E-mail: jiangcs@lreis.ac.cn

通讯作者: 王劲峰, 男, 研究员, 博士生导师。E-mail: wangjf@lreis.ac.cn

1.2 空间抽样问题

无论是在样本独立的假设条件下还是在样本不独立的空间相关假设条件下，抽样主要是解决下面四个问题^[28]：

- (1) 确定样本量大小：选择样本的容量，即样本的个数；
- (2) 布样：样本如何分布，即从总体中采用什么样的方式，如何将样本从总体中选出；
- (3) 抽样估计：对抽样目标的估计，例如：均值、总值等；
- (4) 不确定性衡量：抽样估计的不确定性衡量，如方差；

本文认为应该在第一个步骤前加上：样本空间构造。样本空间是所有样本及其构成方式。在湖泊调查中，如果调查湖泊污染，那么湖泊本身就是抽样的总体，样本空间构造是通过格网化得到离散对象，亦或直接以整个湖泊作为连续的样本空间，则估算均值和总值的方法是不同的^[29]。同时构造离散化空间的方法除了普通的格网化，还有基于镶嵌单元 (tessellation) 的一种随机化分层方法^[30-35]。

在空间抽样中，核心内容之一就是样本的空间布局。样点位置选择的好坏不仅关系到资源的有效利用，而且会对后续研究结果产生重大影响^[28, 36-41]，最优样点选择问题取决于抽样目标和抽样统计推断方式^[36]。

1.3 抽样统计推断方式与抽样目标

在空间抽样领域中，推断方式主要有二种，一种是基于模型的 (model-based) 抽样统计推断，一种是基于设计的 (design-based) 抽样统计推断。Sarndal、Brus 和 Haining 详细讨论了这两种方法的本质区别^[42-44]，这两类抽样统计推断已基本为相关学者认可并广泛应用于实际研究中^[45-49]。

抽样目标是估计参数值的大小和不确定性，参数值的不确定性是源于参数本身服从某种分布^[29]。在基于模型的抽样中，参数的变异来自于模型内在的变异性。而在基于设计的抽样中，参数的变异来自于变量的随机性。Cordy 提出了一种模型辅助 (model-assisted) 的基于设计的抽样方差估计，在基于设计的方差估计中引入了基于模型的协方差方法^[50]。随后这种方法得到了广泛的研究。

1.3.1 基于设计 (Design-Based) 的抽样 基于设计的抽样技术，一般目标是估计总体的均值或总值，且抽样总体常常是有限大小的。样本的值被认为是固定的，每个样本都有入样概率，对参数的推断是基于样本分布。估计参数的变异性来自于样本的变异性，即抽样过程的变异，抽样估计值的变异是抽样中选择点的方差和点对之间的协方差。样本越具有代表性，样本量越大，抽样误差越小，当样本等于抽样总体时，不存在抽样误差，抽样必须基于设计给定的概率进行随机选择，设计抽样方法通常是基于经典抽样理论。

基于设计的抽样中，又可以分为等概率抽样和不等概率抽样。等概率抽样指的是每个样本的入样概率都是相等的，是常见的抽样方式。不等概率抽样，每个样本的入样概率不一定相等，一种常见的不等概率抽样是概率总体抽样^[51, 52] (Probability Population Sampling)。

1.3.2 基于模型 (Model-Based) 的抽样 模型抽样中，抽样总体可以是无限的，用随机化的观点对待观测计量结果^[53, 54]，每个样本单元上的样本值是服从一定内在机理过程的随机变量，抽样总体中观察得到的所有样本值只是随机过程的一次实现，即使抽样总体中所有样本点全被调查，也还会存在抽样误差 (误差来源于模拟随机过程的参数值的不确定性)。样本可以不随机选择 (即某些抽样点一定被选中，而另外的点完全排除在外)，随机性由模型本身产生，由样本随机部分的分布或者协方差表示。由于各样点值的出现受统一的内在随机过程制约，因此，各样本点不独立。在常见的模型中，一般表现为自相似性。模型抽样方法通常用于具有较强空间结构规律 (样点值在空间上分布的规律) 的抽样

总体, 抽样前必须预先得到具有随机过程性质的空间结构 (通常采用地统计模型, 如半变异模型、方差—协方差模型)。

Gruijter 认为传统的认识 (设计抽样的统计推断方法不适用于样本不独立的空间抽样) 有误^[55], 因为空间抽样总体中样本的独立性与否并不是其本质属性, 它应由所采用的设计方法或模型决定。如果对于目标总体的空间分布、空间模式或者其他相关信息有所了解, 则对这种客观的现象建模, 通过模型能够更好的利用空间结构规律, 提高抽样效率。Särndal 的研究支持了这种观点^[44]。最近几年的研究则表明, 无论是基于设计的还是基于模型的抽样, 当在对总体拥有相同的认知的条件下, 二者能够获得类似的样本分布^[56]。如果对一个抽样目标分布做一个基本的假设: 抽样目标存在某种空间模式。那么对于基于模型的抽样则表示抽样目标是一个空间随机场^[57], 对于基于设计的抽样表示样本的坐标能够提供辅助信息。

2 理论研究现状

2.1 国内空间抽样理论研究

李连发提出了空间抽样框架, 以及如何在地理数据上进行抽样调查的一整套方法和理论, 阐述了空间的抽样设计思想, 同时充分利用之前的各种资料, 结合空间特征, 确定样本大小。其关键思想是在抽样中考虑了空间数据的相关性, 并且利用公式量化相关性, 提高了抽样的效率^[58, 59]。牛文杰提出了基于先验信息的 kriging 方法研究, 这种方法利用了原有的目标相关信息^[60], 这种信息相对精度较低, 但是能够反映目标的趋势变化。王劲峰提出“三明治”空间抽样模型^[25]。这种模型利用了抽样对象在空间上的变异特征, 考虑了自然和资源环境领域研究调查中数据具有相关性的特点。同时根据抽样对象特征进行分层, 采用分层抽样。但是在最后结果表达时, 通过均值和误差传递, 能够在调查者感兴趣的区域内汇报结果。这种方法既利用了抽样对象的自然属性, 也考虑到了实际工作中需要根据行政区域汇报统计结果的现状, 为此类问题提供了一个很好的解决方案。

赵雪慧介绍了小区域估计方法及其发展现状^[61], 分析了实际调查抽样中如何设计和调查采用的方法, 提出在抽样中, 要注意整体考虑抽样误差, 而不仅仅是抽样方法的误差。其提到的小区域抽样是一种新的估计方法, 从本质来说, 还是将类似的观测数据放到一起考虑, 考虑某种程度上的一致性。赵宪文研究组从利用遥感进行林业调查的角度探讨了样本分布、样本大小与林业调查结果精度的关系^[21, 62]。

2.2 国外理论研究现状

随着地理空间事物普遍存在自相似性及其相关理论的深入研究, 样点选择的研究已逐渐由专家知识主观指定过渡到通过优化决策模型自动选择。这可以有效保证其科学性, 即对同一片调查区域的多次采样, 其结果具有某一置信水平的可比性 (专家的主观判断不具有这一特性), 基于优化决策支持的样本设计可以有效避免基于主观判断选择样点的偏误。样本优化布局的目的是合理选择并分配样本点, 使所含信息量最大, 样点具有最大的代表性, 它需要充分挖掘所能提供的先验信息。Kriging 方法是地统计的核心内容, 也是空间抽样的一个重要理论基础。下面将分别介绍 Kriging 方法在空间抽样中的研究、样点布局、样点选择和样本优化标准的相关研究现状。

2.2.1 Kriging 方法在空间抽样中的研究

Kriging 方法最早是 Matheron^[63, 64]提出, 开始适用于地质矿藏的研究中, 后来 Oliver^[65]将其用到 GIS 中的插值, 并逐渐在多个领域获得了广泛的应用。Kriging 是基于随机空间变异模型的插值方法, 是一种稳定的插值方法^[66], 对现实拟合比其他纯数学方法要好, 并且有一个统计意义上的最小方差。

半变异函数是 Kriging 的核心内容之一, 它依赖于样点的数量和分布^[67]。Webster^[68]指

出如果要获得可靠的半变异值需要至少 150 个数据, 如果需要知道各向异性, 则需要更多的点。在地球科学中半变异函数的球面和指数模型是最常用的模型。研究表明 Block Kriging 与点 Kriging 一般差别不大, 但是前者更加稳定, 由于小区域的变异, 在插值 kriging 方差较小的时, 点 kriging 方差可能被掩盖, 而 block 的大小由预测的目标决定^[65]。

由于 Kriging 方法的优势, 很多学者将 Kriging 方法应用到抽样设计中。基于 kriging 方差的最优样本布局方法 20 世纪 80 年代就有学者研究^[69-71]。用先验知识估计半变异函数, 很多学者在不同领域都应用到了这一思想^[72-74]。Yfantis 发现三角形格网布样方式相比多边形或者普通格网, 能够提供最稳定的对半变异方差的估计^[75]。Webster 提出了基于多阶层样本的半变异函数估计方法^[76]。例如对于一个农业区域, 首先分为多个县, 然后分多个乡, 然后在根据地块划分底层的抽样区域, 在地块上获得样本观测值。这个观测值表达了多个阶段的变异。假设阶段之间是独立的, 这总的方差等于所有阶段的方差之和。

如图 1 所示, 样点在最下一层, 每层都有一个 kriging 的方差值, 第二层中第 j 类作为一个单元它的值等于包含的最下一层的所有样点单元的均值, 然后在第 j 类这个层次上得到 kriging 方差。

2.2.2 样本布局方法 样本优化布局的目的是合理选择并分配样本点, 充分挖掘先验信息, 使所含信息量最大, 实际中先验信息的给出形式是多种多样的, 按样本选择的逻辑顺序可以分为三类:

(1) 基于前向抽样的最优样本布局 前向抽样是指逐渐增加样本点最终形成所需最优样本的抽样方式, 这类抽样是最常见也是应用最广泛的抽样方法。基于前向抽样的最优样本布局主要是在未抽样点中探寻具有最大信息含量的样本点位置, 每增加一个样本点, 均能使由已知样本点推断出的总体表面精度有显著性的提高, 当增加的样点不能使精度提高显著或已达到所需的精度要求时, 抽样即停止。目前大多数空间抽样中的最优样本布局研究均是基于这类^[76-88]。根据起始抽样时是否存在已知样点可进一步分为两小类: a) 一阶最优样本布局, 起始抽样时没有任何已知样点, 这类抽样的先验知识较少, 抽样总体空间的结构规律通常利用预抽样获取, 即开始时的部分样本点按系统或简单随机进行抽取, 根据抽取的样本点估算抽样总体的空间结构规律(半变异函数或方差—协方差函数), 然后按一定规则逐步加密样本点, 直至形成最终的优样本布局, 使生成的总体表面精度达到给定要求或新增加的样点不能使总体表面精度有显著性提高; b) 二阶最优样本布局, 起始抽样时有部分样点已知, 抽样是为了从未知样点中寻找具有最大信息含量的加密样点, 这类抽样的先验知识较 a) 丰富, 抽样总体的空间结构规律可由已知样点直接估算得到, 后续处理过程与 a) 类同。

(2) 基于后向抽样的最优样本布局, 是指在已有的样本中逐渐消除信息含量较低的冗余样本点以形成最终样本的抽样方式。

(3) 基于双向抽样的最优样本布局, 前向抽样与后向抽样均是单向抽样方式, 即抽样过程中是单纯地添加信息量最大的样本点或单纯地消除最冗余的样本点, 如果在抽样中, 抽样起始时有部分样点已知, 且在抽样过程中样点的添加或删除均存在, 样本点的添加或删除主要是根据各样点对形成的总体表面精度的贡献大小决定。

空间抽样中, 通常假定抽样域有一定内在空间结构规律, 是稳定的随机过程。最优样本布局的任务是从域内选择某一子集, 使其所含信息量最大^[87, 89], 样本的优化布局实际上是一类组合问题, 即从有限的样本点集中抽取某一子集, 使之在某一优化准则下达到

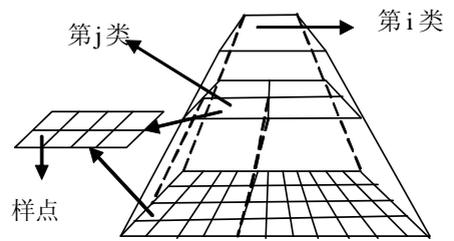


图 1 Webster 的多阶层抽样
Fig. 1 Webster's hierarchical sampling

最优^[36, 40]，这一过程主要涉及两个关键环节：样点的选择方式和样本布局的优化准则，下面分别简述之。

2.2.3 样点选择方法 组合样本点的选择问题比较复杂，在计算机学科里通常称之为完全 N-P 难题，这是一个世界性的难题，目前没有能针对大规模样本达到完全最优的行之有效的解决办法。许多学者对这类算法进行了专门研究，目前常见的解决办法是启发式(序贯选择)或计算机演化算法(模拟退火、遗传算法等)，得到的结果一般是近似最优，下面就空间抽样中前人在这方面所做的研究分别简要介绍。

(1) 随机选择法 (Narve) 随机选择法是指从所有未抽样点随机选取样点的方式，土壤调查中的一种常用方法^[42]。每次随机抽样后，将抽取后的样点与前面已抽样的样点组成样本。若由样本估算的总体表面精度达到要求或新增加的样点不能显著提升总体表面的精度，则抽样停止，所有已抽取样点(用显著提升总体表面精度作为终止条件时需扣除最后一次抽选出的样本点)的集合即为最优样本。这类方法没有考虑抽样总体的空间结构规律，效率最低，其最优样本是随机生成的，最优的真实意义已丧失，它可以用于量化评价其他最优样点选择方法的效率。

(2) 枚举法 (Enumeration) 枚举法是最直接最简单的一类最优样本点选择的解决方法，它是将基于抽样总体的样本点集所有可能的样本全部列举出来，然后从其中选取在某一优化准则下最优的样本，由于它将所有可能的组合全部进行了比对并从中进行优选，因此是全局最优的。这类方法只能适用于抽样总体规模较小的情况(受计算机计算能力限制)，在实际中应用范围较窄。Rogerson 在交通急救信号接收站点布局项目中采用枚举的方式得到了样本点的最优布局^[85]。

(3) 序贯法 (Sequential Selection) 序贯法又称为贪婪法 (Greedy)，它是利用前面已抽取的样点得到每个未抽样点的权重(权重的计算通常是 kriging variance 或 weighted kriging variance)，选取权重最大的点，如此重复，直到由所有已选样点得到的总体表面精度达到给定水平或新样点不能显著提升总体表面的精度。Journel 通过非参数 kriging 方法来逐步定位需要新添加的样本点位置，即基于已有的样本点进行 kriging 插值，获取每个未抽样点的 kriging 方差，取具有最大方差的样本点为需要新插入的样本点，如此重复多次，直到样本量达到要求^[90]。这种方法中，最终选择的样本及其分布对起始样本很敏感，不同的起始样本所形成的最终样本会截然不同。这类方法实现比较容易，时间复杂度不高，但由于样点是单向增加的，不能回溯，因此最终形成的样点集并不一定是组合意义下的最优子集，且容易陷入局部最优。有学者将其用于虫害的监测设计中^[114]。

(4) 模拟退火法 (Simulated Annealing, SA) 模拟退火法应用于空间抽样是目前研究的热点，有大量文献从自然资源调查、河口沉积物调查等各个方面对其进行了深入的研究^[36, 37, 47, 78, 81, 82, 87, 88]。这一算法最初起源于固体物质中的缓慢降温处理以使整体能量处于最低状态，它能有效避免陷入局部最优。其实施过程为：首先随机选择一个样本，随机扰动后生成新样本，对比扰动前后的两个样本，如果扰动后的样本比扰动前的样本更优(通过适应度函数评价)，则用扰动后的样本替代扰动前的样本，否则以一定概率使扰动后的样本替代之前的样本(概率选择采用 Metropolis criterion)，如此重复多次，如果连续拒绝的次数达到一定数量，则算法终止，最后一次扰动前的样本即为最优布局样本。模拟退火算法的主要特征是考虑所有抽样点，能够实现全局优化，并能适应各种抽样约束条件与边界条件，其优化准则一般是由适应度函数衡量。Groenigen 对空间模拟退火 (Simulated Spatial Annealing, SSA) 进行了详细描述^[91]。SSA 算法是 SA 算法在空间的扩展，在空间样本最优化布局中，具有重要的作用。

(5) 空间均衡布样 (Generalized Random Tessellation Stratified 简写 GRTS) Stevens 对此种抽样方法给予了详细的描述^[92]并应用于河谷流域的水资源和生物资源调查^[35]。严格

来说，这不是一种布样方式。因为它涉及到对样本空间的改造，将样本空间从二维平面映射到一维空间上，并且给每个样本一个空间地址坐标。通过阶层随机化 (Hierarchical randomization) 方法，将地址随机排序。GRTS 是基于 Random Tessellation Stratified (RTS) 方法，通过 GRTS 方法构造的一维样本空间，并且给每个样本在一维空间中一个对应的编码，通过这个编码可以将样本重新定位到二维表面上。GRTS 空间的构造可以分为 2 步，第一步将空间划分：假设空间 I^2 ，从右上角的点 (1, 1) 开始，选取 (1/2, 1)、(1/2, 1/2)、(1, 1/2) 三个点作为新的起始点，重复这个过程。第二步对空间中划分得到的格网编码 (图 2)。

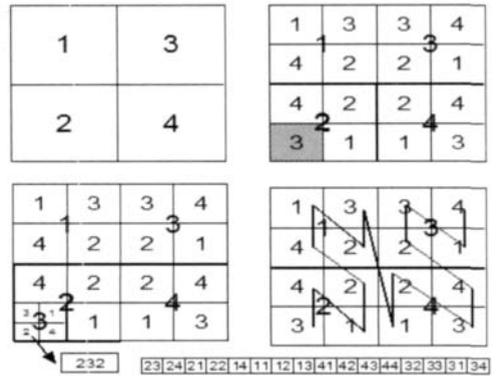


图 2 GRTS 均衡抽样

Fig. 2 GRTS Generalized Random Sampling

(6) 适应性抽样 (Adaptive Cluster Sampling) 对具有聚类特征的总体，抽样临近单元具有较高的同一性可能，因此周围单元的信息较少，在抽样时，被抽取抽样单元周围的样本都不会被抽到。Thompson 指出适应性抽样首先将抽样空间网格化，然后根据概率框架选择一定数量的样本 n_1 ，把这一步得到的样本记为初始样本。如果初始样本第 i 个样本格网中的目标数量超过预先设定的值，则以此样本 i 为中心，对其周围的格网采样，以此类推。最后计算的时候，只计算初始的样本和满足条件的扩展样本。所有达到要求的样本单元构成了一个网络，从网络中的任何一点都可以到其他的达到要求的点^[93]。与简单随机抽样相比，抽样效率还受到样本空间单元大小，初始样本大小和选取周围样本方法的影响。有学者将这种方法应用于估计鸟的数量^[94]和森林资源^[95]等。

Thompson 提出了改进的基于次序的适应性抽样方法 (adaptive cluster sampling)，假设在污染调查中一次抽样了 n 个样本，将 n 个样本排序，然后取前 10 个样本的周围继续放样。这些点称为热点^[96]。一般的适应性抽样是设定一个阈值，超过阈值就在这个值周围放样。基于次序的方法更易于使用，因为阈值不容易设定，并且设定的阈值对抽样的影响较大。

2.2.4 样本优化选择准则 样本布局的好坏需要量化衡量，这与抽样目标密切相关。对于以生成达到给定精度的总体表面为抽样目标时，其量化衡量的精度函数一般由适应度函数 (Fitness Function, FF) 进行估算即：

$$\hat{\varphi}(A|S) = E(\varphi(A|S)) \tag{1}$$

式中： $\hat{\varphi}(A|S)$ 是适应度函数， $\varphi(A|S)$ 是精度函数， A 表示抽样总体空间， S 表示样本。为了便于实际处理，一般用如下数学公式进行计算：

$$E(\varphi(A|S)) = \sum_{i=1}^{N_e} \hat{\varphi}(A|S) \tag{2}$$

式中： N_e 是评估点的总数， $x_i (i = 1, 2, 3, \dots, N_e)$ 是用于进行精度评估的点。前人对适应度函数的研究有很多，归纳起来主要有几种：

(1) Kriging 方差最小化准则 (Minimization of the Ordinary Kriging Variance, MOKV)

Kriging 方差最小化准则是目前研究最广的一类优化准则，在城市环境污染环境监测等领域得到广泛应用^[97, 98]，其目的是使抽样点的分布处于合适的位置以使 Kriging 插值后生成的表面精度最高或由已知抽样点对所有未抽样点的估计精度最高。这一思想最先由 Bertolino 提出^[99]。

Kriging 方法是一类非常重要的最优线性无偏空间插值方法, 其应用已从最初为解决地质中的矿藏蕴含量估算扩展到地理、资源、气象、环境、生态等含有空间位置信息及存在空间自相似性的许多领域^[100-108]。在具有空间二阶平稳性的对象中, Kriging 方法不仅能产生比现有其他空间插值方法 (如全局 / 局部多项式插值、IDW、RBF 等) 更好的插值结果, 而且能产生一个可对插值结果误差进行有效评定的 Kriging Variance, 插值表面的整体 Kriging 方差大小可以用来说明插值表面生成的精度, 它与样本点的布局息息相关, 通过使插值表面生成的整体 Kriging 方差最小, 可以有效规范抽样点在空间的合理布局。

(2) WM 准则 (Warrick-Myers- criterion) WM 准则^[109]的目标是为了使样本点对的分布与设计分布 (一般是均匀分布) 最大可能地保持一致, 在土壤和环境监测领域都得到应用^[82, 84], WM 优化函数数学表达式为^[82, 84]:

$$\varphi_{WM}(S) = \sum_{i=1}^{n_c} (n_i^* - n_i^S)^2 \quad (3)$$

式中: $\varphi_{WM}(S)$ 为所有距离类上 n_i^* 与 n_i^S 的差的平方和, n_c 为用于构建半变异图的距离类的数量, n_i^* 、 n_i^S 分别为设计分布与样本分布上属于第 i 个距离类的样点对的数量。

WM 准则只考虑了样本点之间的距离的分布, 因此, 为了不使样本点聚集在一起, 通常须预先将部分样本点按系统抽样的方式大致布满整个区域, 这样生成的样本点才能代表整体, 而不是局部区域。

WM 要求样本点对距离的分布尽量与设计分布一致可以提高表面生成的效率。因为用 Kriging 插值得到的表面精度受样本点的初始条件的影响极大, 各距离类内的样本点对的数量相差越远, 半变异函数的估算效率越低下, 半变异函数的效率低下亦可等同于插值表面生成效率的低下。

WM 准则主要是利用有限的样本点最大可能地模拟抽样总体真实的半变异函数。WM 准则和 Kriging 方差最小化准则都是基于对影响 Kriging 插值表面精度的关键环节进行有效控制有针对性地提出的改进抽样效率的方法。Kriging 方差最小化主要是从最终生成的表面精度进行控制, 它适应于抽样总体的空间结构规律已知的情形。而现实中, 更多的情况是抽样总体的空间变异结构预先不知道, WM 准则主要为解决这一类问题, 进行抽样设计。它从提高半变异函数的拟和精度角度来提高插值表面的整体精度, 从而间接实现抽样效率的提高。

(3) 平均最短距离最小化准则 (Minimization of the Mean of the Shortest Distances, MMSD) MMSD 准则的目标是为了使样本点在整个抽样空间里最大可能地分布均匀, 在土壤调查和土壤制图中得到应用^[82, 87], 该准则主要目标是使抽样空间上任意一个空间位置点与到它最邻近的点之间的距离的期望值最小, 用数学形式表述如下:

$$\min_S \frac{1}{A} \int \|\vec{X} - V_s(\vec{X})\| \quad (4)$$

式中: S 为抽取的样本, A 为抽样总体空间, \vec{X} 为抽样总体空间上任意位置处的点 (向量), $V_s(\vec{X})$ 为 S 内与 \vec{X} 最邻近的点 (向量)。在实际中, 抽样总体空间通常是一片连续的研究区域, 公式 (4) 中的 \vec{X} 有无穷多, 无法进行操作, 因此, 通常的解决办法是将整个抽样总体空间离散成有限的栅格区域 (栅格化), 把各栅格块的中心点作为 \vec{X} 的备选点, 于是, 公式 (4) 可变为:

$$\varphi(S) = \sum_{j=1}^{n_s} \frac{\left\| \vec{X}_e^j - V_s(\vec{X}_e^j) \right\|}{n_e} \quad (5)$$

式中： j 为标识不同位置格栅格点的记号， n_e 是格栅点的总数， \vec{X}_e^j ($j = 1, 2, 3, \dots, n_e$) 为格栅点，其他符号的意义与公式 (4) 一致。

MMSD 准则生成的样本点的分布偏向于边缘，如果不考虑抽样总体空间的变异性，则生成的样本点分布与系统抽样的分布差不多，MMSD 准则方法简单、直观，易于理解，实施起来也不复杂，实际中采用较多，MMSD 存在一个很大的缺陷，对于那些空间变异复杂，抽样总体中样本点的分布不均匀 (存在强聚集性) 的环境，其抽样效率很低，有时甚至会导致有偏的结果。

Groenigen 提出来 WMSD 准则^[110]，带有权重的平均距离最小，能给不同的区域不同的权重(图 3)。其计算公式为：

$$\phi_{WMSD}(S) = \int_A w(\vec{x}) \|\vec{x} - V_s(\vec{x})\| dx \tag{6}$$

式中： $w(\vec{x})$ 是点的权重， $V_s(\vec{x})$ 是到 x 最近的点。

(4) 极大熵准则 (Maximum Entropy, ENT) 极大熵也是目前比较流行的样本优化准则^[36, 37, 79]，在河口泥沙的监测中得到应用^[37]，其目的是使样本点集所含信息量大 (信息论中熵表示信息的含量)，这对于多目标抽样尤为实用，抽样点更偏向于抽样域的边界，伸展性更强，数学表达形式如下^[36]：

$$H(S) \equiv E[-\log\{f(x_1, \dots, x_n)\}] \tag{7}$$

其中： $S = \{x_1, \dots, x_n\}$ 为抽样域的 n 个样本点位置， $f(x_1, \dots, x_n)$ 为其概率密度函数。

(5) 分形维度 (Fractal dimension) 分形维度是由 Zio 提出来的与 Ripley's- function 相关的一种标准,并将其应用到环境监测台站设计中^[111]。这种方法目标是将样本平均分布在空间中。它解决了由于采用 Krigng 方差最小化标准对数据的要求。这种方法只要知道了抽样区域的形状就可以根据样本量布样。如果点在空间中是均匀分布的，那么以点 A 为中心以 r 为半径的点数 $H(r)$ 和以 r 为半径的圆的面积是具有比例关系的 $\hat{H}(r) \propto \pi D^2$ 。如果不同尺度区域下 r 的指数都是 2，那么就说明这些区域中的点是统一模式。如果 $\hat{H}(r) \propto \pi D^2$ ，则此时的 D 就是分形维度 (fractal dimension, Mandelbrot 1982)，它描述了结构的自相似性。利用 Hastings 的方法^[112]对指数做 log 变换：

$$\log H(r) \cong \log(k) + D \log(r) \tag{8}$$

式中：目标估计 D ，当样点越均匀的分布在空间中，则 D 越接近 2。因此通过使用 (2-D) 作为算法的优化准则。

(6) 均方距离准则 (Mean squared distance to sides, vertices, and boundaries) Stevens 提出一种新的衡量空间分布均匀程度的方法^[56]。其基本思想是，如果样点分布是均匀的，那么构造出来的泰森多边形的面积变异应该为 0(图 4)，即每个样点构成的泰森多边形是一样大的。

如果样点分布不均匀那么样点的泰森多边形面积变异较大，stevnes 根据这一思想，同时考虑了多边形的形状，构造一个 SVB 判别方法：

$$SVB = \sum_{s_i \in S} \left(\int_{B(D(s_i))} (d(b, s_i) - \bar{d}) db \right) / SVB_{NOM} \tag{9}$$

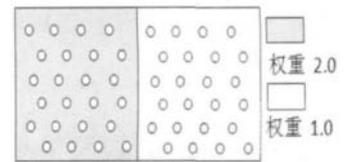


图 3 WMSD 示意图
Fig. 3 Sketch of WMSD

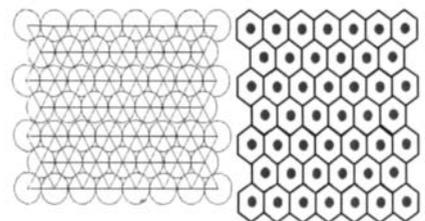


图 4 均方距离准则
Fig. 4 Mean squared distance

式中: S_i 是样本点, $D(s_i)$ 是样点的泰森多边形, $B(D(s_i))$ 是样点的泰森多边形的边 $d(b, s_i)$ 是 S_i 到边上的点 b 的距离, \bar{d} 是用总的面积除以样点个数得到的面积构成的圆的半径, SVBNOM 是一个六边形的面积和总面积除以样点个数得到面积的均方差 (Mean Square deviation)。

3 总结

本文综述了以下几个方面的内容: (1) 空间抽样的发展和空间抽样问题; (2) 抽样的二种推断模式及其适用的条件; (3) 国内外空间抽样理论发展状况。

在空间样本布局领域, 最近几年又出现了很多新的方法, 这些方法基本上可以概括为二点: 第一, 采用空间均衡的样本布设方法, 例如 GRTS, 这个也是美国国家环保局推荐的环境抽样方法。第二, 在样本获取时利用蒙特卡罗模拟以获得对总体的似然估计, 再抽取样本。总的来说, 空间抽样可以分为以 Kriging 相关理论为基础, 以提高抽样精度为目标的抽样理论; 或者是以空间优化布局为基础, 以样本的空间均衡布局为标准或者是按照某种空间布局为基础的抽样理论。

参考文献 (References)

- [1] Matheron G. Principles of geostatistics. *Economic Geology*, 1963, 58: 1246-1266.
- [2] Matheron G. Kriging, or polynomial interpolation procedures. *Canadian Mining and Metallurgical Bulletin*, 1967, 60: 1041-1045.
- [3] Tobler W. A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, 1970, 46 (2): 234-240.
- [4] Cliff A, Ord J. *Spatial Autocorrelation*. London: Pion, 1973.
- [5] Fisher M, Scholten H J, Unwin D. *Spatial Analytical Perspectives on GIS*. London: Taylor & Francis, 1996.
- [6] Anselin L. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic, 1988.
- [7] Haining R P. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University, 2003.
- [8] Ripley B D. *Spatial Statistics*. New York: John Wiley & Sons. Inc., 1981.
- [9] Griffith D A. *Spatial Autocorrelation and Spatial Filtering*. Springer, 2003.
- [10] Stark K E, Arsenault A, Bradfield G E. Variation in soil seed bank species composition of a dry coniferous forest: Spatial scale and sampling considerations. *Plant Ecology*, 2008, 197(2): 173-181.
- [11] Nakamoto S, Fang Z, Matsuura T. Spatial sampling requirements for tropical Pacific sea surface temperature variability. *Journal of Geophysical Research*, 1994, 99(C9): 18363-18370.
- [12] Nick G, Howard C J, McCormick Mark I. Spatial variability in reef fish distribution, abundance, size and biomass: A multi-scale analysis. *Marine Ecology*, 2001, 214: 237-251.
- [13] Dessard H, Bar-Hen A. Experimental design for spatial sampling applied to the study of tropical forest regeneration. *Canadian Journal of Forest Research*, 2005, 35(5): 1149-1155.
- [14] Lianfa L, Wang Jinfeng. Optimal decision-making model of spatial sampling for survey of China's land with remotely sensed data. *Science in China (Series D)*, 2005, 48(6): 752-764.
- [15] Kumar N. Spatial sampling design for a demographic and health survey. *Population Research and Policy Review*, 2007, 26(5): 581-599.
- [16] Fuentes M, Chaudhuri A, Holland D. Bayesian entropy for spatial sampling design of environmental data. *Environmental and Ecological Statistics*, 2007, 14(3): 323-340.
- [17] Lark R M. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma*, 2002, 105(1/2): 49-80.
- [18] Christakos G, Killam B R. Sampling design for classifying contaminant level using annealing search algorithms. *Water Resources Research*, 1993, 29(12): 4063-4076.
- [19] Feng Shiyong. Some hotspot issues on the application and theory of survey sampling. *Statistics & Information Forum*, 2007, 22(1): 5-31. [冯士雍. 抽样调查应用与理论中的若干前沿问题. *统计与信息论坛*, 2007, 22(1): 5-31.]
- [20] Feng Shiyong, Shi Xiquan. *Sampling Surveys Analysis: Theory, Methods and Case Analysis*. Shanghai: Shanghai Scientific & Technical Publishers, 1996. [冯士雍, 施锡铨. 抽样调查: 理论、方法与实践. 上海: 上海科学技术出版社, 1996.]

- [21] Zhao Xianwen. Quantitative Methods by Remote Sensing in Forestry. Beijing: China Forestry Publishing House, 1997. [赵宪文. 林业遥感定量估测. 北京: 中国林业出版社, 1997.]
- [22] Wu Bingfang. Operational remote sensing methods for agricultural statistics. *Acta Geographica Sinica*, 2000, 55(1): 25-35. [吴炳方. 全国农情监测与估产运行化遥感方法. *地理学报*, 2000, 55(1): 25-35.]
- [23] Li Lianfa, Wang Jinfeng, Liu Jiyuan. Optimal decision-making model of spatial sampling for survey of China's land with remotely sensed data. *Science in China (Series D)*, 2005, 48(5).
- [24] Li Lianfa, Wang Jinfeng. Integrated spatial sampling modeling of geospatial data. *Science in China (Series D)*, 2004, 47(3): 201-208.
- [25] Wang Jinfeng, Zhuang Dafang, Li Lianfa. Spatial sampling design for monitoring the area of cultivated land. *International Journal of Remote Sensing*, 2002, 13(2): 263-284.
- [26] Wang Jinfeng, Robert H, Wise S. Spatial sampling design for monitoring drought, flood and earthquake in China. *Progress in Natural Science*, 1999, (4). [王劲峰, Robert H, Wise S. 中国干旱洪水地震灾害监测空间采样设计. *自然科学进展*, 1999, (4).]
- [27] Wang X J, Qi F. The effects of sampling design on spatial structure analysis of contaminated soil. *Science of the Total Environment*, 1998, 223(1-3): 29-41.
- [28] Delmelle E. Optimization of second-phase spatial sampling using auxiliary information. In: Department of Geography, State University of New York at Buffalo, 2005. 108.
- [29] Cooper C. Sampling and variance estimation on continuous domains. *Environmetrics*, 2006, 17(6): 539-553.
- [30] Dalenius T, Hájek J, Zubrzycki S. On plane sampling and related geometrical problems. In: Proceedings of the 4th Berkeley Symposium on Probability and Mathematical Statistics, 1961. 125-150.
- [31] Olea R A. Sampling design optimization for spatial functions. *Mathematical Geology*, 1984, 16: 369-392.
- [32] Overton W, Stehman S. Properties of designs for sampling continuous spatial resources from a triangular grid. *Communications in Statistics Part A: Theory and Methods*, 1993, 22: 2641-2660.
- [33] Stevens Jr D L, Olsen A R. Spatially Restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics*, 1999, 4: 415-428.
- [34] Stevens Jr D L, Olsen A R. Spatially-restricted random sampling designs for design-based and model-based estimation. In: Accuracy 2000: Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. The Netherlands: Delft University Press, 2000. 609-616.
- [35] Stevens Jr D L, Olsen A R. Spatially-balanced sampling of natural resources. *Journal of the American Statistical Association*, 2004, 99: 262-277.
- [36] Banjevic M. Optimal network designs in spatial statistics. In: Department of Statistics, Stanford University, 2004. 114.
- [37] Nunes L M. Optimal estuarine sediment monitoring network design with simulated annealing. *Journal of Environmental Management*, 2006, 78: 294-304.
- [38] Wagner B J. Sampling design methods for groundwater modeling under uncertainty. *Water Resources Research*, 1995, 31(10): 2581-2591.
- [39] Yan S Q, Minsker B. Optimal groundwater remediation design using an Adaptive Neural Network Genetic Algorithm. *Water Resources Research*, 2006. 42.
- [40] Zhu Z Y. Optimal sampling design and parameter estimation of Gaussian random fields. In: Department of Statistics, The University of Chicago, 2002. 132.
- [41] Zimmerman D L. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 2006, 17: 635-652.
- [42] Brus D J, de Gruijter J J. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 1997, 80: 1-44.
- [43] Haining R P. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University, 2003.
- [44] Särndal C E, Swensson B, Wretman J. *Model Assisted Survey Sampling*. New York: Springer, 1992.
- [45] Defeo O, Rueda M. Spatial structure, sampling design and abundance estimates in sandy beach macroinfauna: Some warnings and new perspectives. *Marine Biology*, 2002, 140(6): 1215-1225.
- [46] Flores L A, Martinez L I, Ferrer C M. Systematic sample design for the estimation of spatial means. *Environmetrics*, 2003, 14(1): 45-61.
- [47] Lark R M, Cullis B R. Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science*, 2004, 55(4): 799-813.
- [48] Little R J. To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, 2004, 99: 546-556.
- [49] Papritz A, Webster R. Estimating temporal change in soil monitoring: Sampling from simulated fields. *European*

Journal of Soil Science, 1995, 46(1): 13-27.

- [50] Cordy C B, Thompson C M. An application of the deterministic variogram to design-based variance estimation. *Mathematical Geology*, 1995, 27: 173-205.
- [51] Jiang Baofa, Xu Xiaofei, Wang Jichuan. Time/space location sampling and its application in survey of outpatients visiting sexually transmitted diseases clinic. *Chinese Journal of Health Statistics*, 2003, 20(4): 502-702. [姜宝法, 徐晓菲, 王济川. 时间/空间定位抽样设计及其在性病门诊病人调查中的应用. *中国卫生统计*, 2003, 20(4): 502-702.]
- [52] Jin Yongjin, Hou Zhiqiang. Self-weighting sample design for the China's population changes survey. *Statistics & Information Forum*, 2007, 22(4): 11-13. [金勇进, 侯志强. 中国人口变动调查的自加权抽样设计. *统计与信息论坛*, 2007, 22(4): 11-13.]
- [53] Cochran W G. *Sampling techniques*. 3rd edn. New York: John Wiley & Sons, 1977.
- [54] Sukhatme B V. Testing the hypothesis that two populations differ only in location. *The Annals of Mathematical Statistics*, 1958, 29(1): 60-78.
- [55] De Grujter J J, Braak C J F T. Model free estimation from spatial samples: A reappraisal of classical sampling theory. *Mathematical Geology*, 1990, 22(4): 407-415.
- [56] Stevens Jr. D L. Spatial properties of design-based versus model-based approaches to environmental sampling. In: Caetano M, Painho M (eds.). *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 2006.
- [57] Cressie N A C. *Statistics for Spatial Data*. New York: Wiley-Interscience, 1993.
- [58] Li Lianfa, Wang Jinfeng, Liu Jiyuan. Optimization of decision-making for spatial sampling in remote sensing survey of land. *Science in China (Series D)*, 2004, (10): 975-982. [李连发, 王劲峰, 刘纪远. 国土遥感调查的空间抽样优化决策. *中国科学(D 辑)*, 2004, (10): 975-982.]
- [59] Li Lianfa, Wang Jinfeng. Spatial sampling model for geographic data. *Progress in Natural Science*, 2002, (5). [李连发, 王劲峰. 地理数据空间抽样模型. *自然科学进展*, 2002, (5).]
- [60] Niu Wenjie. Prior information residual Kriging. *Computer Engineering and Applications*, 2004, (35). [牛文杰. 基于先验信息的残余克里金法的研究. *计算机工程与应用*, 2004, (35).]
- [61] Zhao Xuehui. Sample survey methods and theory progress. *Statistics & Information Forum*, 2003, 18(5): 42-72. [赵雪慧. 抽样调查理论和方法的最新进展. *统计与信息论坛*, 2003, 18(5): 42-72.]
- [62] Li Chonggui, Zhao Xianwen. Determining the optimum sample plots for establishing canopy density estimating equation in monitoring area. *Journal of Beijing Forestry University*, 2005, 27(6): 24-27. [李崇贵, 赵宪文. 确定监测区域建立森林郁闭度估测方程最优样地的研究. *北京林业大学学报*, 2005, 27(6): 24-27.]
- [63] Matheron G. *Les Variables Régionalisées et Leur Estimation*. Paris: Masson, 1965.
- [64] Matheron G. *The Theory of Regionalized Variables and Its Applications*. Fontainebleau: Ecole de Mines, 1971.
- [65] Oliver M A, Webster R. Kriging: A method of interpolation for geographical information systems. *International Journal of Geographical Information Science*, 1990, 4(3): 313-332.
- [66] Laslett G M, McBratney A B, Pahl P J. Comparison of several spatial prediction methods for soil pH. *European Journal of Soil Science*, 1987, 38: 325-341.
- [67] Cressie N, Hawkins D M. Robust estimation of the variogram I. *Mathematical Geology*, 1980, 12(2): 115-125.
- [68] Webster R, Oliver M A. How large a sample is needed to estimate the regional variogram adequately? In: Soares A (ed). *Geostatistics Troia '92*. Dordrecht: Kluwer Academic, 1993, 155-165.
- [69] Hughes J P, Lettenmaier D P. Data requirements for Kriging: Estimation and network design. *Water Resources Research*, 1981, 17: 1641-1650.
- [70] Russo D. Design of an optimal sampling network for estimating the variogram. *Soil Science Society of America Journal*, 1984, 48: 708-716.
- [71] Warrick A W, Myers D E. Optimization of sampling locations for variogram. *Calculations Water Resources Research*, 1987, 23: 496-500.
- [72] Hammond L C, Pritchett W L, Chew U. Soil sampling in relation to soil heterogeneity. *Soil Science Society of America Proceedings*, 1958, 22: 548-552.
- [73] Olson J S, Potter P E. Variance components of crossbedding direction in some basal Pennsylvanian sandstones of the Eastern Interior Basin: statistical methods. *Journal of Geology*, 1954, 62: 26-49.
- [74] Webster R, Butler B E. Soil survey and classification studies at Ginninderra. *Australian Journal of Soil Research*, 1976, 14: 1-24.
- [75] Yfantis E A, Flatman G T, Behar J V. Efficiency of Kriging estimation for square, triangular, and hexagonal grids. *Mathematical Geology*, 1987, 19(3): 183-205.
- [76] Webster R et al. Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis of

- variance and residual maximum likelihood. *Computers & Geosciences*, 2006, 32(9): 1320-1333.
- [77] Arbia G, Lafratta G. Anisotropic spatial sampling designs for urban pollution. *Journal of the Royal Statistical Society (Series C): Applied Statistics*, 2002, 51: 223-234.
- [78] Brus D J, Jansen M J W, Gruijter J J D. Optimizing two- and three-stage designs for spatial inventories of natural resources by simulated annealing. *Environmental and Ecological Statistics*, 2002, 9(1): 71-88.
- [79] Bueso M C et al. A study on sensitivity of spatial sampling designs to a priori discretization schemes. *Environmental Modelling & Software*, 2005, 20(7): 891-902.
- [80] Chao C T, Thompson S K. Optimal adaptive selection of sampling sites. *Environmetrics*, 2001, 12: 517-538.
- [81] Van Groenigen J W, Pieters G, Stein A. Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics*, 2000, 11(2): 227-244.
- [82] Van Groenigen J W, Siderius W, Stein A. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, 1999, 87: 239-259.
- [83] Hoeting J A et al. Model selection for geostatistical models. *Ecological Applications*, 2006, 16(1): 87-98.
- [84] Lin Y P, Rouhani S. Multiple-point variance analysis for optimal adjustment of a monitoring network. *Environmental Monitoring and Assessment*, 2001, 69(3): 239-266.
- [85] Rogerson P A, Delmelle E, Batta R. Optimal sampling design for variables with varying spatial importance. *Geographical Analysis*, 2004, 36(2): 177-194.
- [86] Saito H et al. Geostatistical interpolation of object counts collected from multiple strip transects: Ordinary Kriging versus finite domain kriging. *Stochastic Environmental Research and Risk Assessment*, 2005, 19(1): 71-85.
- [87] Simbahan G C, Dobermaan A. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma*, 2006, 133: 345-362.
- [88] Wiens D P. Robustness in spatial studies II: Minimax design. *Environmetrics*, 2005, 16(2): 205-217.
- [89] Papritz A, Webster R. Estimating temporal change in soil monitoring: 2. Sampling from simulated fields. *European Journal of Soil Science*, 1995, 46(1): 13-27.
- [90] Journel A G. Nonparametric geostatistics for risk and additional sampling assessment. In: *Principles of Environmental Sampling*. American Chemical Society, 1988: 45-72.
- [91] Van Groenigen J W, Stein A. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality*, 1998, 27: 1078-1086.
- [92] Stevens Jr D L. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, 1997, 8: 167-195.
- [93] Thompson S K. Factors influencing the efficiency of adaptive cluster sampling, in technical report 94-0301. *Technical Reports and Reprint Series*. Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University, University Park, PA, 1994.
- [94] Smith D R, Conroy M J, Brakhage D H. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, 1995, 51: 777-788.
- [95] Roesch F A Jr. Adaptive cluster sampling for forest inventories. *Forest Science*, 1993, 39: 655-669.
- [96] Thompson S K. Adaptive cluster sampling based on order statistics. *Environmetrics*, 1996, 7: 123-133
- [97] van Groenigen J W, Pieters G, Stein A. Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics*, 2000, 11(2): 227-244.
- [98] Zimmerman D L, Holland D M. Complementary co-kriging: Spatial prediction using data combined from several environmental monitoring networks *Environmetrics*, 2005, 16(3): 219-234.
- [99] Bertolino F, Luciano A, Racugno W. Some aspects of detection networks optimization with the kriging procedure. *Metron*, 1983, 41(3): 91-107.
- [100] Cattle J A, McBratney A B, Minasny B. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *Journal of Environmental Quality*, 2002, 31(5): 1576-1588.
- [101] Chica-Olmo M, Luque-Espinar J A. Applications of the local estimation of the probability distribution function in environmental sciences by kriging methods. *Inverse Problems*, 2002, 18(1): 25-36.
- [102] Emery X. A disjunctive Kriging program for assessing point-support conditional distributions. *Computers & Geosciences*, 2006, 32(7): 965-983.
- [103] Fuentes M. A high frequency Kriging approach for non-stationary environmental processes. *Environmetrics*, 2001, 12 (2): 469-483.
- [104] Lyon S W et al. Defining probability of saturation with indicator kriging on hard and soft data. *Advances in Water Resources*, 2006, 29(2): 181-193.
- [105] Muller T G et al. Map quality for ordinary Kriging and inverse distance weighted interpolation. *Nutrient Management*

- & Soil & Plant Analysis, 2004, 68: 2042-2047.
- [106] Triantafyllis J et al. Mapping of salinity risk in the lower Namoi valley using non-linear kriging methods. *Agricultural Water Management*, 2004, 69(3): 203-229.
- [107] Wang X J et al. Kriging and PAH pollution assessment in the topsoil of Tianjin area. *Bulletin of Environmental Contamination and Toxicology*, 2003, 71(1): 189-195.
- [108] Yeh M S, Lin Y P, Chang L C. Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms. *Environmental Geology*, 2006, 50(1): 101-121.
- [109] Warrick A W, Myers D E. Optimization of sampling locations for variogram calculations. *Water Resources Research*, 1987, 23(3): 496-500.
- [110] Groenigen J W. Spatial simulated annealing for optimizing sampling. In: Soares A, Gomez-Hernandez J, Froidevaux R (eds). *GeoENV I: Geostatistics for Environmental Applications*. Lisbon, Portugal: Kluwer Academic Publishers, 1997. 351-361.
- [111] Zio S D, Fontanella L, Ippoliti L. Optimal spatial sampling schemes for environmental surveys *Environmental and Ecological Statistics*, 2004, 11(4): 397-414.
- [112] Hastings H M, Sugihara G. *Fractals: A User's Guide for the Natural Sciences*. Oxford University Press, 1993.

A Review of Geo-Spatial Sampling Theory

JIANG Chengsheng^{1,2}, WANG Jinfeng¹, CAO Zhidong³

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

2. *Graduate University of Chinese Academy of Sciences, Beijing 100039, China;*

3. *Institute of Automation, CAS, Beijing 100080, China)*

Abstract: Sample survey is the groundwork of the studies of natural resources, environmental problems and socio-economy. The geo-related characteristic of spatial data limits the application of classic sampling theory, which is essentially based on independent assumption. Spatial sampling theory is the foundation of sample survey of spatial related resources. Firstly, this paper introduces the history of spatial sampling theory and presents four main issues addressed by this theory. Then it reviews the theory and the applications of model-based and design-based statistics inference. Finally, this paper gives a detailed description of (1): Kriging theory application in spatial sampling; (2): forward and backward samples distribution methods, and the combination of the above two; (3): six criteria for optimization of sample selection.

Key words: geography spatial; Kriging sampling; sample survey