

土壤优化采样策略研究进展

王子龙, 陈伟杰, 付强, 姜秋香, 常广义, 胡石涛

(东北农业大学 水利与土木工程学院, 黑龙江 哈尔滨 150030)

摘要: [目的] 对现有的采样方式进行系统地总结归纳, 并探寻一种优化的采样策略, 对采样强度、分析成本及其研究精度三方面进行均衡, 即以最小的经济投入换取最大化的试验精度。[方法] 广泛查阅近几年国内外的相关文献, 对土壤优化采样策略的设定进行了系统的总结。将优化采样策略的理论分为合理采样数和样点布设两方面, 就此分别介绍确定采样数的 3 种方法和确定样点布设的 4 种方式, 详细介绍其发展现状, 并对该领域今后的研究进行展望。[结果] 目前的优化采样法大多基于模型来优化采样设计的研究, 其初步采样的方式忽略了空间相关性所导致的信息损失, 必然在一定程度上造成试验结果的偏差和人力物力的浪费。而且, 研究尺度多集中在县级以下, 没有对采样方案的设计确立统一的评价体系和标准。[结论] 未来优化采样设计在样点优先级方面还应进行更加深入的研究。

关键词: 土壤特性; 采样策略; 经典抽样; 地统计学; 采样优先级; 空间采样

文献标识码: A

文章编号: 1000-288X(2017)05-0205-08

中图分类号: S159-3

文献参数: 王子龙, 陈伟杰, 付强, 等. 土壤优化采样策略研究进展[J]. 水土保持通报, 2017, 37(5): 205-212. DOI:10.13961/j.cnki.stbctb.2017.05.034; Wang Zilong, Chen Weijie, Fu Qiang, et al. Research progress of soil sampling strategy optimization[J]. Bulletin of Soil and Water Conservation, 2017, 37(5): 205-212. DOI:10.13961/j.cnki.stbctb.2017.05.034

Research Progress of Soil Sampling Strategy Optimization

WANG Zilong, CHEN Weijie, FU Qiang, JIANG Qiuxiang, CHANG Guangyi, HU Shitao

(School of Water Conservancy and Civil Engineering, Northeast Agricultural University, Harbin, Heilongjiang 150030, China)

Abstract: [Objective] To systematically summarize the existing sampling methods, and to explore an optimized sampling strategy to balance the sampling intensity, analysis cost and research precision, which was expected to maximize the accuracy of experiments with minimum economic input. [Methods] On the basis of extensive review on the domestic and foreign literatures in recent years, the optimized sampling strategy for soil properties was summarized. The theory of optimal sampling strategy was elucidated in two parts: the reasonable sampling number and the sampling point layout. Three methods and four modes were introduced to determine the reasonable sampling number and sampling points, respectively, and the future research was prospected. [Results] Most of the sampling program design used in current methods was model-based, and of which the spatial correlation of samples was ignored and relevant information was unused, these all were responsible to the deviation of experimental results and wastes of labor and materials to some extent. In addition, most researches were county scale or and/or lower scale, and there was no unified evaluation system or standards to design a formal sampling program. [Conclusion] In the future, the priority for optimal sampling design should be strengthened.

Keywords: soil properties; sampling strategy; classical sampling; geostatistics; sampling priority; spatial sampling

收稿日期: 2017-02-06

修回日期: 2017-03-10

资助项目: 国家自然科学基金项目“考虑积雪覆盖的区域土壤春墒预报机理与模型研究: 以松嫩平原黑土区为例”(51579045, 51209039); 黑龙江省自然科学基金(D201403); 黑龙江省普通本科高等学校青年创新人才培养计划(UNPYSCT-2015006); 东北农业大学“学术骨干”项目(16XG10); 东北农业大学“青年才俊”项目(14QC45); 黑龙江省博士后科研启动金(LBH-Q16017)

第一作者: 王子龙(1982—), 男(汉族), 山东省胶州市人, 博士, 副教授, 主要从事寒区农业水土资源高效利用研究。E-mail: wangzilong@neau.edu.cn.

通讯作者: 付强(1973—), 男(汉族), 辽宁省锦州市人, 博士, 教授, 主要从事农业水土资源高效利用理论、方法及应用研究。E-mail: fuqiang@neau.edu.cn.

研究土壤特性、分析其变异性必须以科学的采样策略为依托。采样策略的本质即寻求达到经济投入与试验精度间的优化平衡,以最小的经济投入所换取的离散样本集来估测连续总体集的主要信息而不失其精确性^[1]。建立集代表性、准确性和经济适用性为一体的土壤采样策略不但影响着研究工作的效率,而且也是其研究成败的关键。采样策略一般建立在统计非全面调查的基础上,根据实际工作所研究对象的性质和相关工作的目的而采用与其相适用的方法。各种采样策略会形成各异的采样结果,由此产生的抽样误差也有较大的差别^[2]。相对于实验室分析误差而言,不科学的采样策略带来的误差影响更为显著^[3],因此,对于土壤采样策略的选定及其代表性的研究是土壤特性研究的立论之本。

在传统抽样中,只考虑抽样对象的不同(通常赋予其编号),而抽样对象的排列组合不影响抽样结果。但在空间抽样中,加入了抽样对象空间位置这一因素,与其属性值共同参与到空间采样中来,并对最终的结果产生极大的影响。在空间抽样中,抽样方法大致划分为:基于设计的抽样方法和基于模型的抽样方法。目前,已有国内外学者均对其间的本质区别进行了系统的讨论^[4-6]。

1 合理采样数研究

在设计采样方案中,为使有限的样本值能够尽可能精确地代表总体样本的期望值,采样数应当达到一定的数量,即合理采样数。合理采样数是一个非常重要的因素,它不但决定了采样调查过程的效率和耗费,而且也极大程度上影响着采样结果的准确性。

在估计研究区样本容量时,首先要对其标准差做出估计,同时,还应提出预定的准确度和置信系数的要求,然后据以下几种方式来确定样本容量。一般而言,标准差和准确度可由过去的相关资料来估计,但仅有部分地域范围内的采样数据,也可以通过少量小样本的平均极差来估算标准差 σ 。

$$\sigma \approx \bar{w}/c \quad (1)$$

式中: \bar{w} —— a 个容量为 b 的小样本极差的平均值; c ——折算系数;可通过查由平均极差折算标准差的折算系数和近似自由度值表得到。

1.1 Cochran 采样数法

当采样相互独立且采样数目足够多时,可以认为中心极限定理成立^[7]。普遍使用的是 Cochran^[8]应用于区域纯随机采样而提出满足 t 分布的最佳采样数量计算公式,在一定置信区间和相对误差的条件下,求得相应的采样数。Cochran 公式如下:

$$n = \lambda_{\alpha, f}^2 \left(\frac{\sigma}{\Delta} \right)^2 \quad (2)$$

式中: n ——采样数; $\lambda_{\alpha, f}$ —— t 分布特征值; α ——显著水平, α 通常设为 90%,95%; f ——自由度, $f=N-1$, N ——研究区总样本容量; Δ ——采样精度,通常按 $\Delta=k\mu$; μ ——总体样本均值,取 $k=5\%,10\%,15\%$ 等,其中相对相对误差 k 的确定应根据研究对象的实际平均值和临界水平来确定^[9]。

当样本容量较小或总体方差 σ^2 未知时,由经典统计理论可知,随机变量服从 t 分布,可以用样本方差 S^2 代替总体方差 σ^2 。则采样数为:

$$n = \lambda_{\alpha, f}^2 \left(\frac{S}{\Delta} \right)^2 \quad (3)$$

当总体样本均值 μ 难以得到时,也可用样本均值 \bar{Z}_N 近似代替,此时:

$$\Delta = k \bar{Z}_N \quad (4)$$

则采样数为:

$$n = \lambda_{\alpha, f}^2 \left(\frac{C_v}{k} \right)^2 \quad (5)$$

通过 t 分布法样本合理采样数 n 值表,选用不同的 C_v 可以得到不同的采样数目。

以上样本容量估算的方法将总体看作为无限性的。若总样本量有限,估计的抽样误差应通过抽样分数矫正,因此,可得下式:

$$n' = \lambda_{\alpha, f}^2 \left(\frac{\sigma}{\Delta} \right)^2 \left[1 + \frac{1}{N} \left(\frac{\lambda_{\alpha, f} \sigma}{\Delta} \right)^2 \right] \quad (6)$$

如果 N 很大时,可近似看作总样本容量具有无限性,不必进行矫正,直接计算其近似值。如果,抽样分数 n_0/N 不大,可采用 n_0 作为采样数的样本量,否则,必须计算 n' 值。

$$n' = n_0' / \left(1 + \frac{n_0'}{N} \right) \quad (7)$$

从式(7)可知 n' (采样样本容量)分别与 s^2 (样本方差)和 $1-\alpha$ (置信概率)呈正相关,同时,与 Δ (采样精度)呈负相关。

Cochran 合理采样数法是将整个研究区域土壤设定为一个性质随机分布的均质整体,在忽略土壤的空间结构变化的情况下,通过变异系数和标准差等参数推出所求置信区间和误差条件下的合理采样数。谢宝妮等^[10]认为由此得出的采样数量偏少,无法得出局部变异情况,只能获得变化的大致趋势。

1.2 分层采样法

在分层随机抽样中,抽样单位在各层次的分配方法可分为比重分配法和最适分配法。

1.2.1 比重分配法 若各区层比例为 $p_i = N_i/N$,其中 N_i 为区层的总样本容量,则当研究区总采样样

本容量为 n' 时,各层区采样样本容量可通过 $p_i * n'$ 来分配,故而,若要确定各层区的采样样本容量,只需估计出研究区采样样本容量 n' 。比例配置法的分层采样样本容量的近似公式为:

$$n_0' = \sum(p_i s_i^2) / (d^2 / t^2) \quad (8)$$

式中: s_i ——第 i 层区的标准差; d ——允许的误差; t —— t 分布的统计数,可通过查表(由平均极差折算标准差的折算系数和近似自由度值表),由其中的自由度估计值以及置信系数决定。

如果,抽样分数 n_0/N 极小,则 $n_0' \approx n'$ 。如果计算所得采样样本容量大于总体样本容量的 10%,则采用不重复抽样公式,进一步计算精确 n' 值^[11]。

$$n' = n_0' / \left(1 + \frac{n_0'}{N}\right) \quad (9)$$

比重分配法与 Cochran 法相似,利用样本的标准差和相对误差计算出总的采样数量,再依据各区层的比例来分配,因此造就忽略客观存在于各区层间的变异程度,导致其层区所占比例越大,所需的采样数量就会越多。

1.2.2 最适分配法 最适分配法的着眼点在于根据各层次的区层和误差大小来配置相应的抽样单位,其数量与区层和误差成正相关。Neyman 证实若按区层大小与相应标准差乘积相应地分配抽样单位,可得到最小方差的样本平均数,故而称其为最适分配法^[12]。各区层抽样数计算公式为:

$$n_i' = n' N_i s_i / \sum(N_i s_i) \quad (10)$$

最适分配法样本容量估算的近似公式:

$$n_0' = (\sum p_i s_i)^2 (d^2 / t^2) = (\sum p_i s_i)^2 / V \quad (11)$$

$$n' = n_0' / \left(1 + \frac{1}{NV} \sum p_i s_i^2\right) \quad (12)$$

抽样分数 n_0/N 极小时,可令 $n_0' \approx n'$ 。

式中: V ——一定置信系数下的样本必需方差。

姜城等^[13]发现最适分配法不但可以取得较为理想的采样数,而且也可以而且也可以根据实际情况需要的调节局部区域采样点分布的密度,提高采样效率。姚荣江等^[14]通过经典统计学和地统计学法对黄河三角洲地区不同层次土壤盐分的合理采样数研究中发现采用分层采样法可以很好地降低采样数,采用最适分配法相对于比重分配法可最大降低 35% 的采样数。

1.3 基于地统计学法

地统计学是由 Matheron^[15]提出,并最早由矿物学家 D. R. krige 应用于南非金矿查找过程中。基于地统计学的合理采样数研究是在原土壤研究指标的半方差函数结构和分布格局等研究的基础上分别抽

样,通过比较各抽样方式下研究对象的空间变异特性和插值精度来确定合理采样数目。其一般研究方式如下:首先,将各研究对象的采样数据集划分为训练样本和验证样本,在此基础上不断对训练样本随机抽取数据减少其样点数量,然后,探讨分析研究对象的空间结构特征和克里格插值精度的变化,由此,确定其合理采样数。阎波杰等^[16]研究发现,地统计学法分析和克里格插值法相比经典统计法而言考虑了空间结构性和随机性,能够据此推求较为科学的采样数。

克里格插值法也称作空间局部估计,是一种线性无偏估计的方法,是通过考虑研究对象的空间变异性,得到对待插点属性值的影响范围,利用该范围的已有采样点来估计待插点值^[17]。其中,应用最为广泛的当属普通克里格法,但由于其只需考虑目标变量的空间信息,故而其估测精度需要大量的采样数据和合理的采样密度为依托^[18]。协同克里格法是普通克里格法的拓展,一般涉及多个变量,即主变量和其他的辅助变量。利用主变量的自相关性和具有良好协同性的辅助变量来提高其估测精度,在精度相同的条件下能够有效地优化采样数。与线性回归相结合的克里格法是利用建立主变量与现存的辅助变量间的线性关系,从而求得主变量的数据集,据此,可通过普通克里格法插值求得研究区域的估值分布。

李润林等^[19]研究表明,基于辅助变量的协同克里格插值的精度高于普通克里格插值,同时,得出了该研究在保证精度条件下的采样数下限。庞凤等^[20]将有机质作为辅助变量来研究农田土壤铜含量进行插值分析和采样数量优化研究。李楠等^[21]以有机质作为辅助变量,也通过协同克里格法对速效磷、速效钾等进行采样数优化。李艳等^[22]将土壤电导率作为辅助变量来分析土壤的全盐含量,并发现通过协同克里格法,深层的采样数减少 40% 仍能够满足精度要求,据此可以优化采样数目,节省成本花费。Wu 等^[23]发现采用线性回归相结合的克里格法显著优于协同克里格法。

基于地统计学的合理采样数研究具有诸多优势,但也同样在应用上有一定程度的不足。其主要问题在于,合理采样数设定依赖于大量采样后所得的变异函数,因此,对于不具备区域先验知识的条件下,难以利用该法指导采样。此外,建立变异函数还需要满足二阶准平稳假设和准本征假设。由于该假设在野外采样时很难满足,故而,使得这样优化得到的采样方案有一定的局限性。

2 样点优化布设研究

自 Tobler^[24] 提出了普遍存在于地理空间事物间的空间自相关性的概念开始,许多研究人员投入了对地理空间对象相关性和变异性特征的研究。正是基于空间自相关性的存在,Haining 等^[4] 指出处于相关性影响范围内的点具有相近的性质,样本信息的重叠造成有效样本信息的损失。已有研究表明,采样点的优化布设比单单添加采样数更为重要^[25]。故而,对样点的空间布设的研究意义重大。所谓样点的优化布设,可理解为在研究范围中确定 m 个样点的最优位置,使其具有最佳的全局代表性,以满足研究目的所需。目前确定样点空间布设的方法有经典采样法、目的性采样法、辅助变量法和样点代表性等级法 4 种。

2.1 经典采样法

经典采样方法一般被包含于基于设计的抽样范畴。该方法以有限样本总体的随机抽样为理论基础。其随机性产生于从总体样本抽样的过程中,而不必要对推测研究对象的总体分布进行假设。该方法通过随机采样所得的子样本对整体样本的参数进行估测,一定程度上适用于大尺度采样调查的参数估计。用于抽样的样本量越大,其对总体样本参数的估计就越接近,直到抽样样本等同于总体样本时,误差为零。但由于该方法没有充分考虑样本的空间相关性,仅通过随机性产生布设结果,通常与其他确定样点空间布设的方法相结合使用,也常作为参照进行比较评估。目前,常涉及的经典采样法有如下 6 种。

2.1.1 随机采样法 随机采样法,即从研究区域中对各个采样单元都赋予同等概率被选取的采样方法,其过程具有随机性,可以避免主观影响,具有最小的采样方差,此前已被证明其样本均值和样本方差均可较好地代表总体的最佳估计。但是,因为随机抽样其特性使得采样点分布并不均匀,甚至会出现局部集中和局部空缺的情况。

2.1.2 规则栅格采样法 规则栅格采样法是将研究区域划分成规则的栅格小区后规律进行采样的方式,具有系统的成分,是土壤特性时空变异分析的基础。一般做法是在研究区地图上根据所研究对象的变异性和其他因素而确定的网格尺寸进行叠加,在网格的交点或者网格内部进行采样,通常,会选取半径 3~10 m 范围内的 5~10 钻通过四分法进行混合为一份样本。栅格形状除正方形外,还包括矩形、三角形、菱形和六角形,有研究表明,采用正三角形的栅格相对于正方形而言得到的样本估值更精细一些^[26]。

虽然规则栅格采样技术原理简单,但在实际操作中由于受到地形复杂、交通条件等多种因素的影响,很难将计划采样点在研究区域中准确定位实施采样。此外,由于规则栅格采样带有系统采样的性质,与纯随机采样相比,在方差估算方面会有偏差,对于样本总量平均值的估算则近似随机采样。很多情况下,规则栅格采样技术会与地统计学方法相结合对区域土壤特性进行描述,这样可以很好地体现其连续性和变异性。Franzen 和 Peck 在比较了不同网格尺度下的土质估算精度后,发现以 200 英寸网格经行采样最为合适,并由此提出了一种以 200 英寸为基准根据数据变化差异来调整网格尺寸继而选定最尺寸的方法^[27]。

2.1.3 嵌套采样法 嵌套采样法是将研究区划分为若干个一级单元,在每个一级单元里继续划分出若干二级单元,在所得结果中,可以随机选取或根据研究需要进一步细分,此过程可以根据实际情况持续到可分辨的最小单位。在大尺度的研究区域中,通过整体的经验推断结合局部的嵌套采样能够获得较为精准的采样结果。

2.1.4 分层采样 分层采样,即在对调查对象性质深入分析的基础上,将研究区域划分为性质较为均匀的类型区层,在各个区层中独立进行随机等量抽样构成样本总体。因此,亦称为类型采样。相对于纯随机采样法,分层采样法集统计分组和抽样原理于一体应用于研究变量的均值在各区层间存在明显差异的情况下,能够得到较为精确的估值,此外,在大尺度研究区域下的估测精度也有较好的改善。分层采样的精髓在于其高效的分层方式,由此提出分层效率的概念,即通过相同条件下随机抽样的方差与分层抽样所得方差的比值来定义,比值越大即分层效率越高;只有当分层效率大于 1 时,相对于简单随机抽样才更具有代表性。朱爽等^[28] 研究表明,选取不同的分层方法得到的分层效率差异较大。曹志冬等^[29] 认为先验知识越丰富其指导下的分层抽样所得的估算精度越高,而在没有先验知识的条件下随意分层,其估算精度相较于简单随机抽样并没有明显差异。张锦水等^[30] 研究表明,随着分层层数的增加抽样量在降低,抽样精度也在降低。

近年来,随着地理信息产业的发展,传统的分层采样方法也有了相关的改进研究。连健等^[31] 在分层抽样的基础上利用 GIS 和 ESDA 方法与空间信息相结合,对基于空间“动态抽样框”编织技术的空间分层抽样法进行了研究和应用。翟佩璇等^[32] 则借鉴聚类分析 K-Means 算法的思想将聚类算法和分层抽样相组合,并较为客观地评定了聚类分层抽样的可行性

和有效性,为传统分层采样方法的改进提供了借鉴。王劲峰^[33]在分层采样的基础上首次提出了“三明治空间抽样模型”。三明治空间模型不但完成了样本层、知识层和报告层的信息和误差的传递,而且报告单元可由用户设定的灵活性也突破了经典抽样理论默认其唯一性的局限,解决了多个报告单元布设带来的费用问题。

2.1.5 系统采样 系统采样又称为等距抽样或机械采样,其基本思想是在样本总体中以某种既定顺序排列,随机采取第一个样本点,然后按照一定距离进行采样,直到形成所需的样本总体。在此基础上,继而发展形成系统分层采样,在对总体样本所规划的栅格小区中,继续细分二级栅格,并赋予每个二级栅格一个坐标值 (X,Y) (其中 X,Y 取正整数),在第一个一级栅格中随机选取一个二级栅格,通过行列间的既定规律变化,继而依次得到各个一级栅格中的二级栅格,由此构成整个样本总体。王迪等^[34]在对玉米种

植面积空间抽样调查方案的研究中,发现系统分层采样相对于简单随机采样、分层采样和系统采样而言更为合理。

2.1.6 整群采样 整群采样又称之为区域抽样,适用于主要变异来源在区域间或区域内各单元间。通常,整个研究区域可划为单元数近似相等的组群,在此基础上,随机选取采样组群,最后,在所选组群的各个单元内随机采样得最终的采样结果。其中,群的作用是扩大抽样单位。由于整群采样具有均匀性差和误差较大的缺点,故而一般在空间采样中很少利用,通常与分层抽样组合使用,以此加强其代表性。吴炳方等^[35]将整群采样和样条采样相配合来计算作物种植面积,发现利用整群采样可以克服遥感监测大面积区域的技术瓶颈,同时也节约成本保证精度。

综上所述,对随机采样法、规则栅格采样法、嵌套采样法、分层采样法、系统采样法和整群采样法的特点比较,总结可得如表 1 所示^[33]。

表 1 基本经典采样方法间比较

类别	优点	缺点	误差来源	适用情况
随机采样法	简单易行、对样本均值和方差具有较好的估计	样点分布不均匀	随机性	样本分布均匀、变异性小,样点间独立性强的地区
规则栅格采样法	图上选取简单、适用范围较广	实际定位难,成本较高	网格尺度、不完全随机性	小尺度的区域
嵌套采样法	极大地减少采样经费	样点难以在实地定位,样点分布不均	分布不均匀、各级嵌套网格尺度、嵌套的层数	组成复杂且分层明显的大尺度区域
分层采样法	精度高、误差小、剪表性剪	需要较为丰富的先验知识	先验知识、分层方式、分层层数	层区内性质均匀而层区间差异明显
系统采样法	简便易行,样本在总体中分布比较均匀	精度随总体标志值变动较大	总体排列顺序、总体标志值	波动较大或自然总体相关系数呈凹形分布及负相关
整群采样法	省时省力	均匀性差、误差较大、代表性不强	群间方差、群间差异	变异来源在区域间或区域内各单元间

2.2 目的性采样法

目的性采样即根据专家的先验知识或数据挖掘算法,通过采集被判定为具有区域代表性或者能代表区域均值的典型样点来估算整个研究区域特性的采样方式。目的性采样属于早期应用较多的采样方法,由于其本身极其依赖具有多年实践经验积累的专家和先验知识的指导,故而自身的局限性导致其不易推广。此外,目的性采样的精度是建立在专家主观选择的基础上,很难得到客观的误差评价。随着地统计学的不不断发展,空间采样不断进步,目的性采样已经渐渐淡出历史舞台。

2.3 辅助变量法

20 世纪 70 年代,Huggett^[36]详尽地阐述了土壤

成因与景观类型间的关系,在土壤发生学理论的基础上提出了土壤景观模型。在世界各国的大量研究下,逐步形成了一种基于土壤景观模型,结合空间分析和数学方法的土壤调查法,并被广泛地应用于空间采样点的合理布设。土壤属性的空间分布是在多种地形因子共同作用下形成的,故而,将地形因子作为辅助变量来估测研究对象的空间分布,不但更容易借助遥感等先进技术获得大尺度研究区域的辅助信息,而且在一定程度上考虑了地形因子对土壤性质空间变异的影响。所谓地形因子就是描述空间地理特征的指标,常见的有高程、坡度、平面曲率和植被覆盖等。其中定量环境因子如高程、坡度等相对容易获得,应用较为普遍,此外,定性环境因子如成土母质、土地利用

类型等对土壤属性也就较为显著的影响。将定量和定性环境因子结合作为辅助变量,对于提高土壤性质空间分布估测精度有较好的效果。另有研究表明,基于地形因子作为辅助信息来对研究对象空间属性估计的精度相较于仅基于采样点的方法而言有显著的提高^[37]。Hengl 等^[38]、Kerry 等^[39]、Simbahan 等^[40]和 Minasny 等^[41]均利用辅助变量法进行采样点的布设并成功获取了所研究对象的空间信息。

在已有研究的基础上,可总结其一般研究方式是通过辅助因子与土壤性质的相关性来推求其空间分布,利用优化算法,以不确定性最低为目标函数,在研究区域内搜索一定数量点的最佳位置。对于一定数量采样点的最优布设方式,有枚举法、贪婪法、模拟退火法 3 种。

(1) 枚举法是处理该类型最直接的方法,通过列举全部可能的样本点集组合来比对得到全局最优样本。但是,在大尺度的采样区域内,面对数量较多的采样数,其位置的组合方式会异常复杂,采用枚举法往往会造成“组合爆炸”问题,故而在实际中应用范围有限。

(2) 贪婪法是利用已知采样点通过 Kriging variance 或 weighted kriging variance 等方法来推求未知采样点的权重,从而选定权重最大的点,继而推算下一个点直到满足总体精度无法再随样点数的增加而显著增大时为止。这类方法便于操作,但是整个过程是通过单向推求,无法回溯,最终形成的样点集并不一定是组合形式下的全局最优,容易形成局部最优的结果。

(3) 模拟退火法在大量的优化算法研究中非常热门,它参考了所有采样点,能够近似得到一个全局优化的样点布置。其一般应用过程如下:在已有样本集中随机选采样本子集,随机进行扰动,然后通过适应度函数评价选择较优的那个样本子集,如果扰动后的子集无法代替原有子集,则设定一定的概率使新生的样本子集替代原有子集,经过一定反复后得到的子集,即为最优的布局样本。Vasát 等^[42]在模拟退火法中通过克里格法和加权克里格法对采样点进行优化布局,对不同土壤属性得到全局或局部最优布设。韩宗伟等^[43]通过多种地形因子在道路周边设置采样尺度,通过模拟退火法对各样点的空间布局进行了优化,获取了基于路网的土壤采样优化布局,为降低成本、提高效率提供了有效手段。

2.4 样点代表性等级法

样点代表性等级是对样点全局代表性强弱的等级划分。其中,样点全局代表性是指样点具备的土壤空间特性能够很好地契合研究区实际的变异分布。

全局代表性高的样点不但在数值空间内很好地囊括了目标区域土壤特性的典型值,而且在地理空间内,也可以极大限度地代表土壤属性的空间变异。而样点代表性等级法,即是由具有全局代表性样点构成的集合中,在有限样本容量的条件下,可以优先选择代表性较高的样点,由此获得最大的研究效益。其一般方法是,在土壤景观模型理论的基础上,将所选环境因子的分布图叠加,通过赋予的权重计算而得到各位置的代表性等级,根据一定的优化规则剔除冗余信息,最后确定可供采样点布设的代表性等级序列。

杨琳等^[44]发现,通过少量的代表性较高的样点可以较好地代表研究区域的相关特性,而随着低代表性等级样点的加入,其精度并没有太大变化。这种采样方法不但可以极好地避免数据冗余,提高采样效率,降低研究成本,为后续研究提供了很好的数据支撑,而且,由此得到的采样点能够更好地保证插值结果的精度,即具有 Mcbratney 等^[45]指出的现有插值方法所要求样点须具备研究区空间变化的全局代表性。在此基础上,孙孝林等^[46]在相关研究中证实了样点代表性等级法在成本、精确度上有明显的优点,同时提出了推广此法的关键,即探寻较好的土壤协同环境因子。此外,早在 2009 年郭丽娜等^[47]按照观生态学理论引入了等别多样性指数,以此提出了一种利用等别质量代表性来确定农用地等别质量检测带的方法。

3 存在的问题及展望

尽管国内外学者对于优化采样设计做了大量的研究,提出了多种方法,但在应用中仍存在诸多问题:

(1) 当下对合理采样数的研究,多热衷于利用地统计学通过克里格法进行优化,这种优化采样设计的方式更多的是为多次采样和日后的监测点布置提供参考,然而,并不适用于没有先验知识情况下的初步采样设计。

(2) 多数基于模型来优化采样设计的研究,支撑其理论的初步采样数据,往往是通过简单地遵循随机抽样的方法甚至是通过目的性采样得到的。这种忽略空间相关性所导致的信息损失,必然在一定程度上造成试验结果的偏差和人力物力的浪费。

(3) 建立在样本个体完全独立基础上经典抽样方法虽然广泛运用于各个领域,但是,随着 3S 技术的成熟和推广,对具有相关性的地理空间对象而言,这种传统的方法自带的弊病已无法满足于当下的研究所需。

(4) 合理采样设计的研究尺度多集中在县级以

下的区域,在大尺度上的研究尚少。

(5) 对合理采样方案设计没有统一的评价体系和标准。

通过对国内外相关研究和发展的总结可知,样点代表性等级法开拓的新思路,可能会使今后相关研究的采样策略更倾向于带有采样优先级的弹性设计。在利用辅助因子来推断优先级的过程中,可引入相关的评估机制。其中,对于研究对象和辅助因子的相关性研究以及寻找一套契合该方法的评价机制将成为相关研究的重点。

[参 考 文 献]

- [1] 李连发,王劲峰. 地理数据空间抽样模型[J]. 自然科学进展,2002,12(5):545-548.
- [2] 郝向东. 浅谈系统抽样和分层抽样[J]. 统计与信息论坛,1996(4):62-65.
- [3] Havlin J L, Jacobsen J S. Soil testing: Prospects for improving nutrient recommendations. [M]. Madison, Wisconsin, USA: Soil Science Society of America, Inc. American Society of Agronomy, Inc. 1994.
- [4] Haining R, Zhang J. Spatial Data Analysis: Theory and Practice[M]// Spatial Data Analysis: Theory and practice. Cambridge University Press,2003:1077.
- [5] Brus D J, De Gruijter J J. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil[J]. Geoderma, 1997,80(1/2):1-44.
- [6] 贺本岚. 复杂抽样中基于设计与基于模型推断的比较研究[J]. 北京市第十六次统计科学研讨会获奖论文集, 2011.
- [7] 史海滨,陈亚新. 土壤水分空间变异的套合结构模型及区域信息估值[J]. 水利学报,1994(7):70-77,89.
- [8] Cochran W G. Sampling Techniques (3rd ed.) [M]. John Wiley and Sons, Inc, New York, 1977.
- [9] Cameron D R, Nyborg M, Toogood J A, et al. Accuracy of field sampling for soil tests[J]. Canadian Journal of Soil Science, 1971,51(2):165-175.
- [10] 谢宝妮,常庆瑞,秦占飞. 县域土壤养分离群样点检测及其合理采样数研究[J]. 干旱地区农业研究,2012,30(2):56-61.
- [11] Sachs L. Applied Statistics[J]. Springer, 1984,17(3): 389-390.
- [12] 盖钧镒. 试验统计方法[M]. 北京:中国农业出版社, 2003:339-346.
- [13] 姜城,杨俐苹,金继运,等. 土壤养分变异与合理取样数量[J]. 植物营养与肥料学报,2001,7(3):262-270.
- [14] 姚荣江,杨劲松,姜龙. 黄河三角洲土壤盐分空间变异性与合理采样数研究[J]. 水土保持学报,2006,20(6): 89-94.
- [15] Matheron G. Principles of geostatistics[J]. Economic Geology, 1963,58(8):1246-1266.
- [16] 阎波杰,潘瑜春,赵春江. 区域土壤重金属空间变异及合理采样数确定[J]. 农业工程学报,2008,24(S2):260-264.
- [17] 陈飞香,戴慧,胡月明,等. 区域土壤空间抽样方法研究[J]. 地理与地理信息科学,2012,28(6):53-56.
- [18] Cattle J A, McBratney A, Minasny B. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination[J]. Journal of Environmental Quality, 2002,31(5):1576-1588.
- [19] 李润林,姚艳敏,唐鹏钦,等. 县域耕地土壤锌含量的协同克里格插值及采样数量优化[J]. 土壤通报,2013,44(4):830-838.
- [20] 庞凤,李廷轩,王永东,等. 县域农田土壤铜含量的协同克里格插值及采样数量优化[J]. 中国农业科学,2009, 42(8):2828-2836.
- [21] 李楠,徐东瑞,吴杨洁. 土壤养分含量的协同克里格插值研究[J]. 浙江农业学报,2011,23(5):1001-1006.
- [22] 李艳,史舟,王人潮,等. 海涂土壤剖面电导率的协同克里格法估值及不同取样数目的比较研究[J]. 土壤学报,2004,41(3):434-443.
- [23] Wu J, Norvell W A, Hopkins D G, et al. Improved prediction and mapping of soil copper by kriging with auxiliary data for cation-exchange capacity[J]. Soil Science Society of America Journal, 2003,67(3):919-927.
- [24] Tobler W R. A Computer Movie Simulating Urban Growth in the Detroit Region[J]. Economic Geography, 1970,46(2):234-240.
- [25] 赵业婷,李志鹏,常庆瑞,等. 基于 Cokriging 的耕层土壤全氮空间特征及采样数量优化研究[J]. 土壤学报, 2014,51(2):415-422.
- [26] Weindorf D C, Zhu Y. Spatial variability of soil properties at Capulin Volcano, New Mexico, USA: Implications for sampling strategy[J]. Pedosphere, 2010,20(2):185-197.
- [27] Franzen D W, Peck T R. Field soil sampling density for variable rate fertilization. [J]. Journal of Production Agriculture, 1996,8(4):568-574.
- [28] 朱爽,张锦水. 面向省级农作物种植面积遥感估算的分层方法[J]. 农业工程学报,2013,29(2):184-191.
- [29] 曹志冬,王劲峰,李连发,等. 地理空间中不同分层抽样方式的效率与优化策略[J]. 地理科学进展,2008, 27(3):152-160.
- [30] 张锦水,潘耀忠,胡潭高,等. 冬小麦种植面积空间抽样效率影响因子分析[J]. 农业工程学报,2009,25(8): 169-173.
- [31] 连健,李小娟,宫辉力,等. GIS支持下的空间分层抽样方法研究:以北京市人均农业总产值抽样调查为例

- [J]. 地理与地理信息科学, 2008, 24(6): 30-3338.
- [32] 翟佩璇, 高飞, 吴兆福, 等. 基于聚类分层的地理空间数据抽样研究[J]. 信息通信, 2016(4): 26-28.
- [33] 王劲峰. 空间抽样与统计推断[M]. 北京: 科学出版社, 2009.
- [34] 王迪, 周清波, 陈仲新, 等. 玉米种植面积空间抽样调查方案优化设计[J]. 农业工程学报, 2014, 30(8): 117-125.
- [35] 吴炳方, 李强子. 基于两个独立抽样框架的农作物种植面积遥感估算方法[J]. 遥感学报, 2004, 8(6): 551-569.
- [36] Huggett R J. Soil landscape systems: A model of soil Genesis[J]. Geoderma, 1975, 13(1): 1-22.
- [37] Li Qiquan, Yue Tianxiang, Wang Changquan, et al. Spatially distributed modeling of soil organic matter across China: An application of artificial neural network approach[J]. Catena, 2013, 104(2): 210-218.
- [38] Hengl T, Rossiter D G, Stein A. Soil sampling strategies for spatial prediction by correlation with auxiliary maps[J]. Australian Journal of Soil Research, 2003, 41(8): 1403-1422.
- [39] Kerry R, Oliver M A. Variograms of ancillary data to aid sampling for soil surveys[J]. Precision Agriculture, 2003, 4(3): 261-278.
- [40] Simbahan G C, Dobermann A. Sampling optimization based on secondary information and its utilization in soil carbon mapping[J]. Geoderma, 2006, 133(3): 345-362.
- [41] Minasny B, McBratney A B. A conditioned Latin hypercube method for sampling in the presence of ancillary information[J]. Computers & geosciences, 2006, 32(9): 1378-1388.
- [42] Vašát R, Heuvelink G B M, Boruvka L. Sampling design optimization for multivariate soil mapping[J]. Geoderma, 2010, 155(3): 147-153.
- [43] 韩宗伟, 黄魏, 罗云, 等. 基于路网的土壤采样布局优化: 模拟退火神经网络算法[J]. 应用生态学报, 2015, 26(3): 891-900.
- [44] 杨琳, 朱阿兴, 秦承志, 等. 一种基于样点代表性等级的土壤采样设计方法[J]. 土壤学报, 2011, 48(5): 938-946.
- [45] Mcbratney A B, Santos M L M, Minasny B. On digital soil mapping[J]. Geoderma, 2003, 117(1/2): 3-52.
- [46] 孙孝林, 王会利, 宁源. 样点代表性等级采样法在丘陵山区土壤表层有机质制图中的应用[J]. 土壤, 2014, 46(3): 439-445.
- [47] 郭力娜, 张凤荣, 马仁会, 等. 基于标准样地的国家级农用地等别质量监测点设置方法探讨: 以冀豫鄂三省为例[J]. 资源科学, 2009, 31(11): 1957-1966.

(上接第 204 页)

- [3] 郭洪峰, 许月卿, 吴艳芳. 基于地形梯度的土地利用格局与时空变化分析: 以北京市平谷区为例[J]. 经济地理, 2013, 33(1): 160-166.
- [4] 徐新良, 刘纪远, 邵全琴, 等. 30 a 来青海三江源生态系统格局和空间结构动态变化[J]. 地理研究, 2008, 27(4): 829-838.
- [5] 朱瑜馨, 赵军, 曹静, 祁连山山地生态系统稳定性评估模型[J]. 干旱区研究, 2002, 38(4): 33-37.
- [6] 温晓金, 杨海娟, 刘焱序. 基于地形因子的 km 尺度景观生态廊道布局研究[J]. 地理科学进展, 2013, 32(2): 298-307.
- [7] 田瑞云, 王玉宽, 傅斌, 等. 基于 DEM 的地形单元多样性指数及其算法[J]. 地理科学进展, 2013, 32(1): 121-129.
- [8] 孙丕苓, 许月卿, 王数. 环京津贫困带土地利用变化的地形梯度效应分析[J]. 农业工程学报, 2014, 30(14): 277-288.
- [9] 李升发, 董玉祥. 广州市土地利用垂直分层模式及其演变过程[J]. 自然资源学报, 2013, 28(6): 1019-1029.
- [10] 龚文峰, 袁力, 范文义. 基于地形梯度的哈尔滨市土地利用格局变化分析[J]. 农业工程学报, 2013, 29(2): 250-259.
- [11] Mottet A, Ladet S, Coqué N, et al. Agricultural land-use change and its drivers in mountain landscapes: A case study in the Pyrenees[J]. Agriculture, Ecosystems & Environment, 2006, 114(2): 296-310.
- [12] 郭洪峰, 许月卿, 吴艳芳. 基于地形梯度的土地利用格局与时空变化分析: 以北京市平谷区为例[J]. 经济地理, 2013, 33(1): 160-166.
- [13] 谢高地, 鲁春霞, 冷允法, 等. 青藏高原生态资产的价值评估[J]. 自然资源学报, 2003, 18(2): 189-196.
- [14] 李超, 杜哲, 陈亚恒, 等. 环京津地区土地生态服务价值时空分异特征[J]. 土壤通报, 2015, 46(1): 42-47.
- [15] 谢高地, 甄霖, 鲁春霞, 等. 一个基于专家知识的生态系统服务价值化方法[J]. 自然资源学报, 2008, 23(5): 911-919.
- [16] 石晓丽, 王卫. 生态系统功能价值综合评估方法与应用: 以河北省康保县为例[J]. 生态学报, 2008, 28(8): 3998-4006.