

一种基于样点代表性等级的土壤采样设计方法*

杨琳 朱阿兴 秦承志 李宝林 裴韬

(中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101)

摘要 采样设计是获取土壤空间分布信息的关键环节, 直接影响到土壤制图的精度。目前常用的采样设计方法大多存在着设计样本量大、采样效率不高的问题。当可投入资源难以完成一次性大量采样时, 采样往往需要多次、分批进行。然而现有分批采样方法多考虑各批采样点在地理空间的互补性, 可能造成样本点在属性空间的重叠, 影响采样资源的高效利用。鉴于此, 本研究通过对与土壤在空间分布具有协同变化的环境因子进行聚类分析, 寻找可代表土壤性状空间分布的不同等级类型的代表性样点, 建立一套基于代表性等级的采样设计方法。将该采样方法应用于位于黑龙江省嫩江县鹤山农场的研究区, 利用所采集的不同代表性等级的样点进行数字土壤制图并进行验证, 探讨采样方案与数字土壤制图精度的关系, 以评价本文所提出的采样方法。结果表明, 通过代表性等级最高的少量样点可获取研究区的大部分主要土壤类型(中国土壤系统分类的亚类级别), 且制图精度较高; 随着代表性等级较低样点的加入, 土壤图精度提高; 但当样点增加到一定数量时, 土壤图的精度变化不大。因此, 与样点数相比, 样点的代表性高低对制图精度的影响更大。该方法所提出的代表性等级可以为样点采集顺序提供参考, 有助于设计高效的逐步采样方案。

关键词 采样设计; 样点代表性等级; 模糊聚类; 数字土壤制图

中图分类号 P934 **文献标识码** A

对土壤资源及其空间分布的准确认识, 是充分、合理、持久地利用土壤资源的基础。从19世纪末俄罗斯自然地理学家 Dokuchaev 应用地理综合法对尼日格勒省黑钙土的调查开始, 土壤调查已经历了百余年的发展^[1-2]。由于不可能测得区域内所有位置的土壤性状(类型或属性), 采样调查便成为获取土壤信息及其空间分布的基本方法。采样设计直接影响到利用样点进行土壤制图的精度, 因而成为土壤采样调查中的关键环节^[3-9]。

目前, 常用于土壤性状空间分布推测制图的采样方法包括: 经典采样方法和空间采样方法。经典采样, 如随机采样、系统采样等, 简单易行、应用广泛, 但通常需要大量样点才能全面准确地获取土壤性状的空间分布特征^[10-11]。由于野外采样需要耗费大量的人力、物力和时间, 因而, 这种方法效率低、成本高。空间采样方法是在地统计学的支持下, 考虑区域土壤的空间自相关特性来设计采样样点^[12-16]。这类方法常依赖于需大量样本才能建立的空间变差函数(variogram), Webster 和 Oliver^[17]的研究表明: 需要 100~150 个样点才能建立有效的变差函数。Stein 等^[18]的研究表明, 随着样本数减少, 普通克里格法对土壤水分的预测精度显著降低。此外, 变差函数的建立需满足两个假设: 二阶平稳假设和本征假设。这两个假设在复杂的实际野外条件下很难得到满足。

近年来, 数据获取技术和现代信息技术的发展为推测土壤性状的空间分布提供了越来越多的辅助环境数据(ancillary data, 也称 secondary information), 例如遥感影像、高精度数字高程模型(DEM)及衍生出的一系列地形属性等^[19]。研究者们开始利用环境数据来辅助空间采样设计。Minasny 和 McBratney^[20]运用拉丁超立方方法将样点均匀布设于土壤环境因子的参数空间, 通过全面覆盖参数空间获取土壤的空间分布特征。该方法设计的样本量较大, 增加了野外采样的成本。Minasny 等^[21]采用变差四分法根据环境因子参数空间的变化程度来

* 国家自然科学基金项目(41001298)、科技部国际科技合作项目(2010DFB24140)、国家自然科学基金项目(40971236, 41023010)资助

作者简介: 杨琳(1982-), 女, 山东文登人, 博士, 主要从事数字土壤制图研究。E-mail: yanglin@reis.ac.cn

收稿日期: 2010-11-27; 收到修改稿日期: 2012-03-29

设计样点, 在环境因子参数变化剧烈的地区设计样点数量多, 反之则设计样点数量少。然而该方法所设计的样本也通常较多, 因此其采样成本较高。

受可投入资源(包括人力、物力)的限制, 一次性采集满足较高精度土壤制图所需的样本数量通常是不现实的, 特别是对于大尺度的土壤调查而言, 采样往往需要多次、分批、逐步进行。现有的分批采样多考虑各批采样点在地理空间的互补性, 但可能造成样本点在属性空间的重叠, 因而造成采样资源的浪费。目前, 关于分批逐步采样设计方法的研究还非常有限。

土壤在空间的分布往往是连续变化的, 由一种典型土壤类型逐渐变化到另一种典型土壤类型, 因而在空间中存在着可代表土壤性状变化的代表性样点。这些代表性样点可根据土壤性状变化类型的差异进一步分为代表土壤性状全局变化和局域变化的样点。基于此, 本研究提出一种基于代表性等级的采样方法, 旨在寻找可代表土壤性状不同等级(尺度)变化的样点, 同时, 代表性等级可指示样点采集的优先级顺序, 为分批逐步采样的统筹规划提供所需的重要信息。本研究拟以土壤类型为例, 设计可代表土壤类型空间变化的不同代表性等级的样点, 并利用所得代表性样点进行土壤制图进而对本研究所提采样方法进行评价。

1 研究区与研究方法

1.1 研究区

研究区位于中国黑龙江省黑河市嫩江县鹤山农场老莱河流域, 面积约为 60.2 km²。该区处于寒温带季风草甸草原区, 冬季漫长而干冷, 夏季短促而湿润, 年降水量 400~550 mm。最大高程为 366 m, 高差约为 100 m, 地形起伏平缓。该区的地貌主要是不同程度切割的山前洪积台地及冲积湖积平原。整个研究区的母质基本上一致, 多为黄土状亚粘土和重壤土物质。根据第二次全国土壤普查结果, 按土壤发生学分类, 本区的主要土壤类型为黑土与草甸土^[22]。原生植被为疏林草甸、灌丛草甸和杂类草草甸, 但近四十年多被开垦为农田, 主要农作物是大豆和玉米, 人类活动剧烈。

1.2 研究方法

基于土壤景观模型理论, 土壤在空间分布的类型对应了多种环境因子在一定属性范围内的组合, 将其称之为环境因子组合, 因而可通过环境因子组合来间接代表在空间分布的不同土壤类型。本文所提出的采样设计方法的基本思路为: 通过对土壤协同环境因子采用模糊聚类的方法可以获得环境因子组合, 随着聚类数逐渐增大, 所得环境因子组合也逐渐增多、细化; 假设不同聚类数下环境因子组合可反映随尺度不同而变化的土壤类型, 即可通过多个聚类数下的环境因子组合结果设计不同代表性等级的样点。以下是具体方法:

(1) 选择并获取土壤协同环境因子。研究区母质及植被基本一致, 最大海拔高差小于 100m, 高程对土壤分布的影响不大, 因此, 选择坡度、沿等高线曲率(Planform curvature)、沿剖面曲率(Profile curvature)^[23]及地形湿度指数(Topographic wetness index)四个环境因子来体现该区土壤类型的空间分布差异。环境因子的空间变化信息通过数字地形分析方法获取。研究区的数字地形 DEM 数据是以 1:10 000 地形图为基础输入, 采用 ARC/Info 软件平台下的 TopoGrid 和 TINLATTICE 相结合的方法生成^[24]。根据本研究土壤制图的尺度(流域尺度)要求并结合地形图的比例尺, 选择 10m 分辨率作为研究区 DEM 的格网大小, 并计算坡度、沿等高线曲率、沿剖面曲率及地形湿度指数^[24-25]。

(2) 获取环境因子组合。采用模糊 *c* 均值聚类方法(Fuzzy *c*-means clustering)^[26]对所选四个环境因子进行模糊聚类。通过对该区野外考察可知, 聚类数从 8 到 13 的所得环境因子组合可体现该区不同尺度的土壤类型在空间的变化情况。因而, 聚类数选择从 8 到 13, 所得聚类结果为环境因子组合的模糊隶属度分布图(如图 1a 所示)。对于某一环境因子组合类, 其模糊隶属度图中隶属度值高的区域被认为是该类环境因子组合的中心位置, 可通过设

定隶属度阈值可确定环境因子组合的中心位置。在本研究区，隶属度阈值设定为 0.6（隶属度取值范围为 0 到 1），即隶属度大于 0.6 以上的位置为类别中心区。如图 1b 所示，图中值为 1 的像素代表该环境因子组合的中心位置。

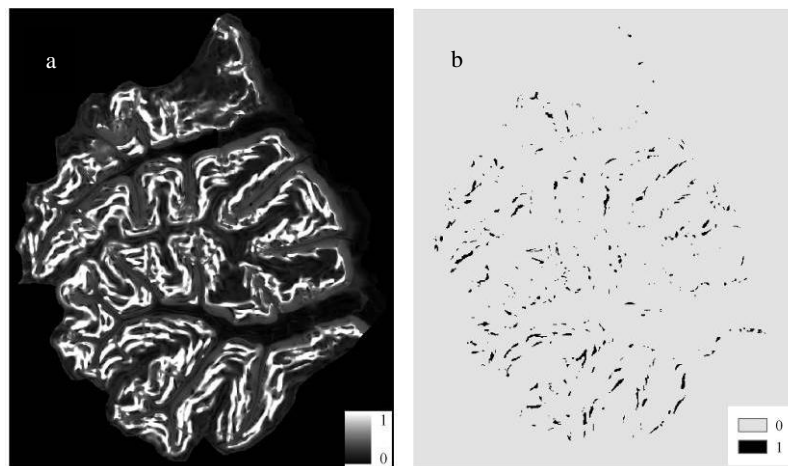


图 1 某环境因子组合模糊隶属度图 (a) 及其中心区图 (b)

Fig. 1 Fuzzy membership map (a) and typical positions map (b) of a certain combination of environmental co-variates

(3) 确定代表性等级。将多个聚类数下的环境因子组合类的中心位置进行叠加，生成环境因子组合中心位置的频率分布图（栅格格式，见图 2）。假设：环境因子组合中心位置重合的次数越多，其对全局土壤类型的代表性就越高，在其上设计的样点就更容易获得研究区土壤类型与环境之间的主要关系；反之，代表性等级也越低，主要代表了局域变化的土壤类型，在其上设计的样点获得研究区土壤类型与环境的次要关系。因此，根据中心位置出现的次数（频率）来确定代表性等级的高低，重叠次数最多的，代表性等级最高，代表性等级随重叠次数依次降低。图 2 中频率值为 6 的位置代表有 6 个环境因子组合的中心位置出现，代表性等级为 6，代表性等级最高，频率由 5 到 1 的代表性依次为 5 到 1。

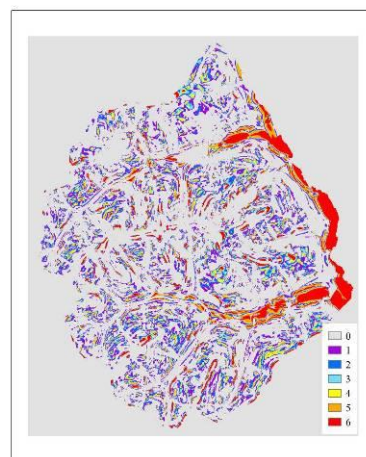


图 2 环境因子组合中心位置出现频率图

Fig. 2 Frequency of typical positions of the combination of environmental co-variates

(4) 样点设计。环境因子组合中心位置的频率分布图指示了每个像素上环境因子组合中心位置出现的次数。对于同一代表性等级的像素而言，其环境因子组合中心位置出现的次数相同，但环境因子组合类型可能不同，例如，像素 (i, j) 和 (k, l) 代表性等级同为 2，但像素 (i, j) 出现的环境因子组合类型为聚为 8 类时的第 1 类环境因子组合和聚为 10 类

时的第 4 类环境因子组合，像素 (k, l) 出现的环境因子组合类型为聚为 9 类时的第 1 类环境因子组合和聚为 10 类时的第 3 类环境因子组合。为了表达这种同一代表性等级下的不同像素以进行采样设计，将每个像素上出现的环境因子组合以链的形式来表达，将其称之为环境因子组合链。例如，像素 (i, j) 环境因子组合链的形式为 8Class1-10Class4。同种环境因子组合链的像素构成了某一景观模式在空间上的典型位置，不同代表性等级体现了不同等级的景观模式，每一代表性等级内因环境因子组合链的不同体现了不同的景观模式。

根据各像素上出现的环境因子组合，得到研究区代表性从 6 到 1 的共 322 种不同的环境因子组合链。当代表性较低的某种环境因子组合链中所出现的各聚类数下的环境因子组合与代表性较高的某种或多种环境因子组合链相应聚类数下的环境因子组合完全相同时，可认为该环境因子组合链不再具备对景观模式的指示作用，因为代表性较高的环境因子组合链包含了其代表的景观模式。例如，代表性等级为 2、ID 为 75 的环境因子组合链为 9Class7-13Class12，其出现的所有环境因子组合与代表性等级为 3、ID 为 26 的环境因子组合链 8Class4-9Class7-13Class12 中的相应环境因子组合一致，即这两种环境因子组合链都包含聚类数为 9 时的 Class7 和聚类数为 13 时的 Class12。对 ID 为 75 的环境因子组合链所对应的景观模式采样所获得信息同样也可以代表 ID 为 26 的环境因子组合链所对应的景观模式类。因此，对所有环境因子组合链进行挑选，剔除那些包含于较其代表性等级更高的环境因子组合链中的环境因子组合链，最终可得到区域内可指示采样点设计的 35 种环境因子组合链。如表 1 所示，代表性等级从 6 到 2。

表 1 经剔除后的环境因子组合链表

Table 1 Matrix of the environmental combinations chains after screening

代表性等级 Representative grade	环境因子组合链 ID Environmental combinations chain ID	聚类数 Clustering number						数量 Number
		8	9	10	11	12	13	
		环境因子组合类						
		Environmental combination classes						
6	18	1	2	5	1	11	6	9
	21	5	5	2	6	10	3	
	10	4	7	3	7	5	1	
	30	2	9	9	4	6	5	
	1	2	9	9	4	6	2	
	2	7	8	10	8	7	7	
	7	3	1	4	10	8	12	
	13	6	4	8	2	4	4	
	17	6	4	8	3	2	9	
5	8	0	3	6	5	3	13	7
	15	8	0	7	9	1	11	
	20	8	3	6	5	0	13	
	19	5	5	0	3	2	9	
	9	8	3	6	0	1	13	
	32	1	2	5	1	0	5	
	3	7	8	10	0	12	7	
4	12	3	1	4	0	0	8	4
	29	8	3	0	9	1	0	
	22	6	6	0	2	0	4	

	11	6	6	0	0	4	4	
	5	0	0	0	11	9	10	
	14	0	4	7	0	0	11	
	28	0	0	0	8	7	8	
	33	3	0	0	0	3	8	
	26	4	7	0	0	0	12	
3	16	0	4	7	2	0	0	12
	31	1	0	0	0	6	5	
	6	0	6	1	0	0	4	
	4	0	0	7	9	12	0	
	25	0	8	7	0	12	0	
	27	3	0	0	0	7	8	
	35	0	0	6	0	3	8	
	23	7	0	1	0	0	0	
2	24	0	0	0	2	12	0	3
	34	8	8	0	0	0	0	

注：每一聚类数对应列中的数字代表该聚类数下的环境因子组合类编号，例如聚类数为 8 对应的第一行数字 1 即代表聚类数为 8 下的环境因子组合 Class1, 0 代表该聚类数下的没有环境因子组合出现 Note: The digit in the column of each clustering represents the class ID of environmental combination under the corresponding clustering number, e.g. the first digit, 1, under Clustering 8 represents environmental combination Class1. The digit, 0, means that no environmental combination class occurs for the corresponding clustering number

在环境因子组合链上设计样点还需另外两个指标：环境因子组合链的面积与环境因子组合链的平均模糊隶属度。环境因子组合链的平均模糊隶属度是指该环境因子组合链包含的每一聚类数下环境因子组合的模糊隶属度的平均值。该值代表了每一像素代表该链的典型性。在得到上述指标之后，最终设计采样时按照下述原则：首先，按照代表性由高到低，先在代表性等级高的环境因子组合链（景观模式）设计样点，再在代表性等级低的环境因子组合链设计样点，所设计样点的代表性与其对应环境因子组合链的代表性等级相同。其次，对于同一种代表性等级，按照环境因子组合链的面积大小排序，依次设计。若环境因子组合链像素较少，即对应聚类数下的环境因子组合重叠面积较小，可认为该环境因子组合链代表性差，不进行采样设计，在本案例中，少于 20 个像素的环境因子组合链不进行样点设计。最后，每一环境因子组合链上根据其对应像素的平均隶属度值高低设计样点，隶属度值越高越有可能被选择为样点，将研究区的地形图叠加，避开道路、居民点、水域等，确定最终的样点位置，每一景观部位设计 3 个点。最终设计 66 个样点。通过实地野外调查，去除不可获得的样点，共获得 48 个样点，如图 3 所示。

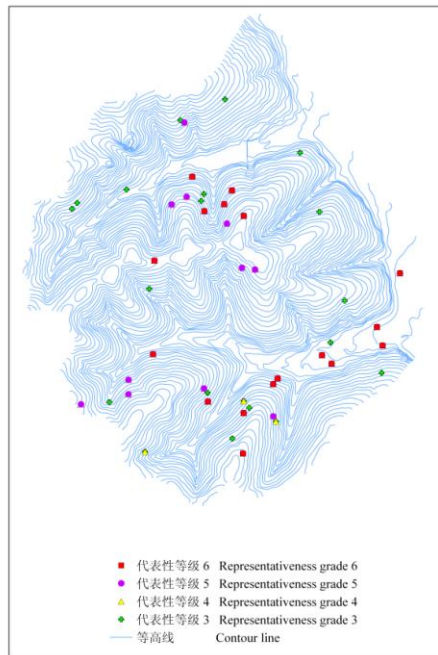


图 3 实际获得样点图

Fig. 3 Sampling point map actually obtained

2 基于代表性样点的数字土壤制图及其精度评价

采样方法效率的高低难以直接测定,本文利用所得代表性样点进行土壤制图并对土壤图进行独立样点验证,以评价所提出的采样设计方法。

2.1 数字土壤制图

经野外实地调查,通过代表性等级为 6 的样点可以获得研究区 5 种主要土壤类型:暗沃冷凉湿润锥形土、暗厚滞水湿润均腐土、普通筒育湿润均腐土、普通筒育一纤维有机滞水潜育土和石质湿润新成土,本研究中的土壤类型以中国土壤系统分类作为土壤分类系统,亚类作为基本单元;通过代表性等级为 5 的样点增加了一种土壤类型:普通冷凉湿润锥形土,并且由代表性等级 6 和 5 的样点基本可以获得研究区的大部分土壤类型;通过代表性等级为 4 的样点没有增加新的土壤类型信息,但所对应的增加环境因子组合类链细化了景观模式,特别是位于坡上的景观模式;通过代表性等级为 3 的样点增加了普通筒育湿润均腐土另一个实例(instance),即出现于不同环境状况下的同一土壤类型均可称之为该土壤类型的实例,至此,得到了研究区的所有土壤类型(案例)。

将代表性等级为 6 样点称为第 1 组样点,代表性等级为 6 和 5 的样点为第 2 组样点,代表性等级为 6、5 和 4 的样点为第 3 组样点,代表性等级为 6、5、4 和 3 的样点为第 4 组样点。通过这种方式可以模拟分批采样,即代表性等级由 6 到 3 的样点分别为第 1 批到第 4 批采集的样点。分别利用 4 组样点获取土壤-环境关系知识并通过模糊隶属度曲线来量化这种土壤与环境关系的知识^[24, 27]将环境因子数据和量化的土壤-环境关系知识输入到土壤景观推理模型(Soil-Landscape Inference Model, SoLIM)^[28],推测土壤类型在空间的分布,生成 4 套土壤类型图,如图 4 所示。

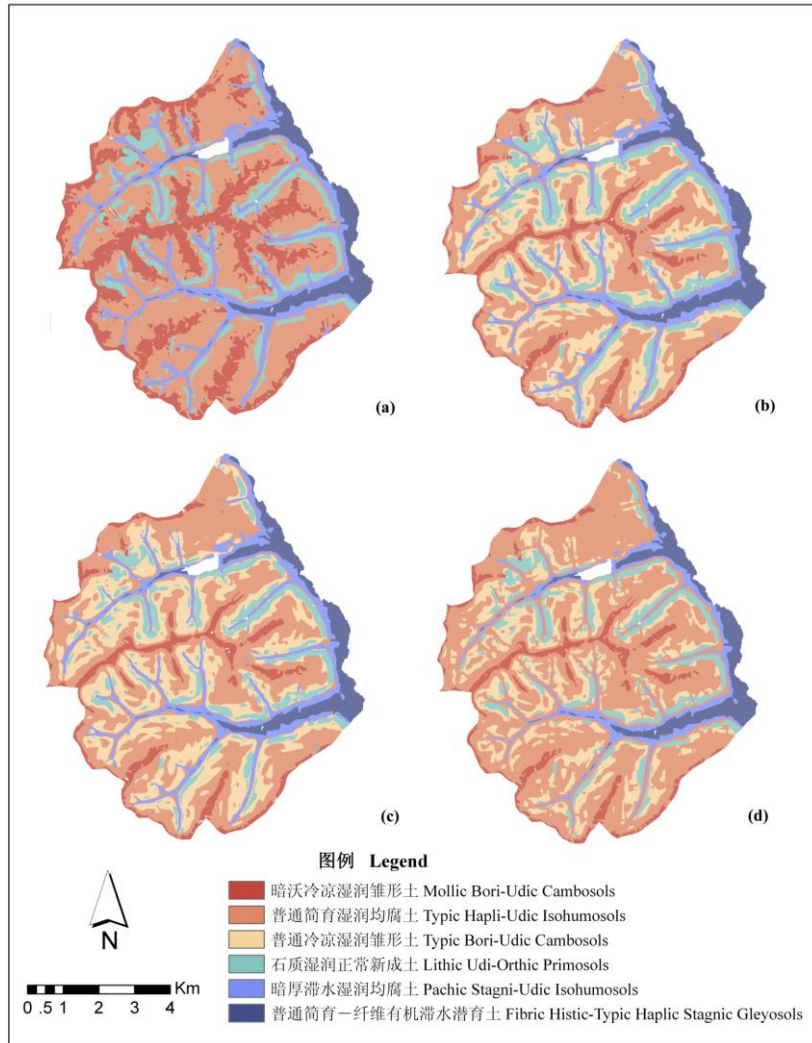


图 4 土壤类型图：(a) 为代表性等级 6 的样点所得，(b) 为代表性等级 6 和 5 的样点所得；(c) 为代表性等级 6、5、4 所得；(d) 为代表性等级 6、5、4 和 3 的样点所得

Fig. 4 Soil maps: (a) using samples of Representativeness Grade 6; (b) using samples of Representativeness Grade 6 and 5; (c) using samples of Representativeness Grade 6, 5 and 4; (d) using samples of Representativeness Grade 6, 5, 4 and 3

由图 4 可以看出：由代表性等级 6 样点所得土壤图（图 4a）到第二组样点（代表性等级为 6 和 5 的样点）所得土壤图（图 4b）因增加了一种土壤类型而变化最大，而图 4b 到图 4c 以及图 4c 到图 4d 的变化主要是土壤类型在空间分布的细节变化。

2.2 数字土壤制图的精度评价

利用 50 个独立野外验证样点对土壤类型图的精度进行评价。验证点集通过系统采样（规则采样）、横切线采样和主观采样三种方案获得（图 5）。系统采样以 $0.6' \times 0.6'$ （纬度 \times 经度）为间隔分布（图 5 中圆形点），用以检验土壤图获取的土壤类型空间变化信息的整体情况。横切线采样是在研究区中选择一条横穿山坡、沟谷的切线，沿着这条切线设计样点，使其在较短的距离穿越主要的景观模式变化，共布设 10 个样点（图 5 中十字点），样点间的平均距离为 100 m。主观采样是将样点主观布设在均匀采样和横切线采样没有布设到的景观部位上，例如：山顶，河谷和陡坡位置，共 7 个样点（图 5 中五角星点）。

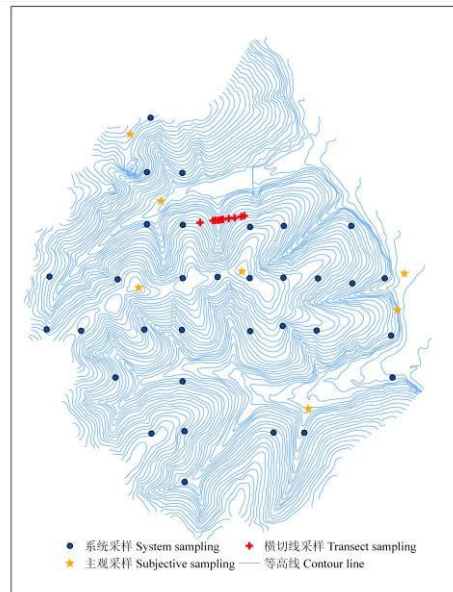


图 5 验证点集

Fig. 5 Validation point set

经上述验证点集的验证，四套土壤图的总体精度见表 2。此处的样点数量为制图有效样点数量，即包括去除获取土壤环境知识时解释为过渡类型的环境组合类链所对应的样点，及解释类内纯度不等于 100% 中不纯的样点。由表 2 看出，代表性等级为 6 的样点（12 个）可以获得较高精度的土壤图，当样点数量由 12 增加到 20，总体精度有一个较大的提高，然后由 20 个样点到 22 个样点总体精度提高较少，由 22 个样点到 35 个样点总体精度变化不大。由此可以认为当增加的样点不能增加土壤制图有效的信息时，土壤制图的精度的提高不大。

表 2 采用不同代表性等级样点生成土壤图的总精度

Table 2 Overall accuracies of soil maps based on sampling points of different representiveness grade

代表性等级 Representativeness grade	样点数量 Number of samples	总精度 Overall accuracy (%)
6	12	62
6, 5	20	70
6, 5, 4	22	72
6, 5, 4, 3	35	76

3 结 论

本文提出了一种利用土壤协同环境因子来设计可代表不同空间变化尺度的土壤类型的代表性样点的采样方法，样点的代表性等级同时还可指示样点的采集顺序。将该采样方法应用于位于黑龙江省嫩江县鹤山农场的研究区，利用所采集的不同代表性等级的样点进行数字土壤制图并进行验证，以评价本文所提出的采样方法。结果表明，通过代表性等级最高的少量样点可获取研究区的大部分主要土壤类型，随着代表性等级较低样点的加入，再获得其他土壤类型；同时，随着样点数量增加，土壤类型图精度提高，但当样点增加到一定数量，土壤图的精度变化不大，由此也可验证，并非采集越多的样点就可以提高制图精度，样点的代表性高低对制图精度的影响更大。

本研究所提出的采样方法一方面可以克服现有采样设计方法的局限性，提高采样效率、节约采样成本，另一方面可以为调查者有效地利用有限的采样资源并进行统筹规划提供重要的样点采集顺序信息，因而可作为现有采样方法的拓展。除了在本文中应用于土壤类型采样

设计中以外, 该采样方法也可以应用于其他地理要素的采样设计。

尽管所提出的采样方法在本文案例中取得了较好的效果, 但仍然存在一些有待于进一步探讨的方面, 例如方法参数(如聚类数、确定环境因子组合中心位置的隶属度阈值)的选择、方法敏感性研究以及该方法与经典采样或空间采样方法的对比等。本研究的后续工作将在这些方面展开, 同时将在较大面积的研究区应用以探讨该方法在大尺度地理要素调查中的效率。

参考文献

- [1] USDA. Soil Taxonomy: A basic system of soil classification for making and interpreting soil survey. <http://soils.usda.gov/technical/classification/taxonomy/>. 1999
- [2] 庄卫民. 土壤调查与制图技术: 理论 方法 应用. 北京: 中国农业科技出版社, 1995. Zhuang W M. Techniques of soil survey and mapping: Theories, methodology, application (In Chinese). Beijing: China Agricultural Science and Technology Press, 1995
- [3] Webster R. Quantitative and numerical methods in soil classification and survey. Oxford: Clarendon Press, 1979
- [4] Kish L. Survey sampling. USA: John Wiley & Sons, 1985
- [5] 王珂, 沈掌泉, Bailey J S, et al. 精确农业田间土壤空间变异与采样方式研究. 农业工程学报, 2001, 17(2): 33-36. Wang K, Shen Z Q, Bailey J S, et al. Spatial variants and sampling strategies of soil properties for precision agriculture (In Chinese). Transactions of The Chinese Society of Agricultural Engineering, 2001, 17(2): 33-36
- [6] 任振辉, 吴宝忠. 精细农业中最佳土壤采样间距确定方法的研究. 农机化研究, 2006, 6: 82-85. Ren Z H, Wu B Z. The research of determine method for the best soil sampling space in precision agriculture (In Chinese). Journal of Agricultural Mechanization Research, 2006, 6: 82-85
- [7] Gregoire T G, Valentine H T. Sampling strategies for natural resources and the environment. Taylor & Francis Group, 2007
- [8] Brus D J, Noij I G A M. Designing sampling schemes for effect monitoring of nutrient leaching from agricultural soils. European Journal of Soil Science, 2008, 59(2): 292-303
- [9] Heim A, Wehrli L, Eugster W, et al. Effects of sampling design on the probability to detect soil carbon stock changes at the Swiss CarboEurope site Lageren. Geoderma, 2009, 149: 347-354
- [10] Cochran W G. Sampling techniques. 3d ed. USA: John Wiley & Sons, 1977
- [11] 胡键颖, 孙山泽. 抽样调查的理论、方法和应用. 北京: 北京大学出版社, 2006. Hu J Y, Sun S Z. Sampling theories, methodology and applications (In Chinese). Beijing: Peking University Press, 2006
- [12] Cressie N. Statistics for spatial data. New York: Wiley & Sons, 1991
- [13] 李世清, 高亚军, 李生秀. 土壤养分的空间变异性及确定样本容量的研究. 土壤与环境, 2000, 9(1): 56-59. Li S Q, Gao Y J, Li S X. Study on spatial variability of soil nutrient and determining number of sample (In Chinese). Soil and Environmental Sciences, 2000, 9(1): 56-59
- [14] Wang J F, Liu J Y, Zhuang D F. Spatial sampling design for monitoring the area of cultivated land. International Journal of Remote Sensing, 2002, 23(2): 263-284
- [15] Haining R. Spatial data analysis. Cambridge: Cambridge University Press, 2003
- [16] 王劲峰, 姜成晟, 李连发, 等. 空间抽样与统计推断. 北京: 科学出版社, 2009. Wang J F, Jiang C S, Li L F, et al. Spatial sampling and statistical inference (In Chinese). Beijing: Science Press, 2009
- [17] Webster R, Oliver M A. Sample adequately to estimate variograms of soil properties. European Journal of Soil Science, 1992, 43(1): 177-192
- [18] Stein A, van Dooremolen W, Bouma J, et al. Corriging point data on moisture deficit. Soil Science Society of America Journal, 1988, 52: 1 418-1 423

- [19] McBratney A B, Mendonca Santos M L, Minasny B. On digital soil mapping. *Geoderma*, 2003, 117: 3-52
- [20] Minasny B, McBratney A B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, 2006, 32: 1 378-1 388
- [21] Minasny B, McBratney A B, Walvoort D J J. The variance quadtree algorithm: Use for spatial sampling design. *Computers and Geosciences*, 2007, 33: 383-392
- [22] 刘兴山, 蓝宏. 黑龙江省第二次土壤普查数据册. 黑龙江省土壤普查办公室, 黑龙江省土地勘测利用技术中心, 1990. Liu X S, Lan H. The second soil survey data book of Heilongjiang Province (In Chinese). Soil Survey Office of Heilongjiang Province, Land Survey and Use Technology Center of Heilongjiang Province, 1990
- [23] Shary P A, Sharaya L S, Mitusov A V. Fundamental quantitative methods of land surface analysis. *Geoderma*, 2002, 107: 1-32
- [24] 杨琳, 朱阿兴, 李宝林, 等. 应用模糊 c 均值聚类获取土壤制图所需土壤-环境关系知识的方法研究. *土壤学报*, 2007, 44(5): 16-23. Yang L, Zhu A X, Li B L, et al. Extraction of knowledge about soil-environment relationship for soil mapping using fuzzy c -means (FCM) clustering (In Chinese). *Acta Pedologica Sinica*, 2007, 44(5): 16-23
- [25] 秦承志, 杨琳, 朱阿兴, 等. 平缓地区地形指数的计算方法. *地理科学进展*, 2006, 25(6): 87-93. Qin C Z, Yang L, Zhu A X, et al. Computation method of topographic index in low relief area (In Chinese). *Progress in Geography*, 2006, 25(6): 87-93
- [26] Bezdek J C, Ehrlich R, Full W. FCM: The fuzzy c -means clustering algorithm. *Computers and Geosciences*, 1984, 10(2/3): 191-203
- [27] Zhu A X, Yang L, Li B L, et al. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma*, 2010, 155(3/4): 166-174
- [28] Zhu A X, Hudson B, Burt J, et al. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 2001, 65: 1 463-1 472

A SOIL SAMPLING METHOD BASED ON REPRESENTATIVENESS GRADE OF SAMPLING POINTS

Yang Lin Zhu A-xing Qin Chengzhi Li Baolin Pei Tao

(State Key Laboratory of Environment and Resources Information System, Institute of Geographical Sciences and Resources Research, Chinese Academy of Sciences, Beijing 100101, China)

Abstract Sampling design plays a critical role in capturing information about spatial distribution of soils, influences directly the accuracy of samples-based soil mapping. The commonly used methods are often confronted with such problems as the need for large amounts of samples and low sampling efficiency. When the resources available are inadequate for accomplishing a large volume of sampling once for all, additional sampling has to be done many times and in batches. However, the present in-batches sampling methods mostly take into account supplementality of the batches of sampling points in geographic space, which lead to overlapping of the sampling points in attribute space, affecting efficient utilization of the resources. In view of this, clustering analysis of soil environmental co-variates was performed to explore for sampling points representative of different grades of soil spatial distribution and to formulate a sampling designing method based on representativeness grade. This proposed method was applied to the study area located in Heshan Farm, Nenjiang County, Heilongjiang Province, and digital soil mapping was done using selected sampling points of different representativeness grades for verification of the methods. Results show that only a small number of sampling points of the highest grade could obtain most of the main soil types (sub-order in Chinese Soil Taxonomy) and soil mapping was high in accuracy; and addition of sampling points lower in representative grade could raise the accuracy of soil mapping; but the

accuracy did not improve much when the number of sampling points added exceeded a certain level. Therefore, compared with the number of sampling points, representativeness grade of the points was more important to accuracy of soil mapping. The representativeness grade system proposed in this paper can be used as reference for sequencing of sampling points and is contributive to formulation of high-efficiency stepwise sampling design.

Key words Sampling design; Representativeness grade; Fuzzy clustering; Digital soil mapping