

整合已有土壤样点的数字土壤制图补样方案

张淑杰^{1,2}, 朱阿兴^{1,3}, 刘 京³, 杨 琳¹

(1. 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;

2. 中国科学院研究生院, 北京 100049;

3. 美国威斯康星大学麦迪逊分校地理系, 麦迪逊, WI53706)

摘 要:我国多数地区经过两次土壤普查以及科研工作者的野外调查, 积累了一定数量的土壤样点(“已有样点”)。本文在充分整合已有样点的基础上, 提出了逐次、高效地设计补充样点的方案, 它包括3个步骤: 首先, 确定已有样点集的空间代表范围(“可推测范围”); 然后, 将已有样点集不能代表范围内的每一个栅格都看作一个候选样点, 计算每一个候选样点的可扩推范围, 选择可扩推范围最大的点作为第一个补充样点; 最后, 基于已有样点和补充样点更新样点集的空间代表范围图。重复以上过程, 直至新样点集的空间代表范围能够覆盖整个研究区。该方法在充分利用已有样点资源的基础上, 不仅能够确定补充样点的数量、位置, 而且能够给定补充样点的重要性次序, 在采样资源有限的情况下, 为采样者合理地选择样点提供了重要依据。

关 键 词:已有样点; 全局代表性; 个体代表性; 推理不确定性; 逐次补样方案

1 引言

土壤属性图是土地资源管理、环境模拟等领域所需的重要基础数据。传统的土壤图, 由野外经验丰富的土壤专家, 基于土壤—景观模型, 通过手工绘制制作。这种制图方法生产周期长, 耗费资本高。数字土壤制图利用先进的空间信息处理技术和高分辨率地形数据, 能够快速地生产高精度、高分辨率的土壤图, 是一种精细、高效、经济的土壤普查技术。

野外采集的样点是进行数字土壤制图的主要数据源, 样点的数量和布局对推测结果有较大的影响^[1-3]。因此, 现有数字土壤制图方法对样点的数量、分布或典型性都提出了一定的要求, 即只有当样点具备一定的数量并且在分布上符合某种特定的规则(如规则分布或随机分布等), 或是能够反映土壤空间变化规律的典型样点时, 才能保证样点集对研究区内土壤空间分布规律具有全局代表性, 进而保证数字土壤图产品的精度^[4]。

经过两次全国范围的土壤普查以及科研工作者的野外调查, 我国大多数地区都积累了一定数量的野外样点。然而, 在基于这些样点推测土壤属性

图, 尤其是推测大中尺度的土壤属性图时, 却由于以下两个方面的原因而无法保证样点对整个研究区土壤空间分布规律的全局代表性:

(1) 样点的数量有限且分布不符合某种规则(如规则分布、随机分布等)。受野外自然条件和采样资源的限制, 在研究区内尤其是大中尺度区域按照某种规则采集符合设计要求的样点几乎是不现实的, 所采集的样点难以涵盖全局的土壤—环境关系。

(2) 样点的典型性得不到保证。在土壤普查中, 普查者一般根据自己的经验确定样点的典型位置, 在这种情况下默认所采集到的样点对土壤空间分布规律具有全局代表性。然而, 典型样点的确定受主观因素的影响很大^[5], 难以确保所采集的样点是能够全面反映当地土壤—环境关系的典型样点。

由以上两点可知, 通常情况下已有样点的全局代表性不好, 不能满足现有数字土壤制图方法对样点的要求。基于这些样点推测土壤属性图时, 制图精度无法得到保障。因此, 必须设计补充样点以完备样点集的全局代表性, 提高土壤属性推理精度。

常见的样点设计方法主要有简单随机采样、规则采样、分层随机采样以及基于空间自相关的克里

收稿日期: 2012-01; 修订日期: 2012-05.

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

作者简介: 张淑杰(1983-), 女, 山东德州人, 博士研究生, 主要从事数字土壤制图及样点设计研究。E-mail: zhangsj@lreis.ac.cn

格采样^[6-8]等。除此之外,如何通过先验知识布设少量具有“代表性”或“平均状态”的典型样点也是样点设计领域的研究热点^[9-12]。但是,上述方法都很难兼容已有样点,造成了资源的浪费。而已有样点集尽管全局代表性不一定好,但每个样点都蕴含了采样者对当地土壤—环境关系知识的认知,补样设计过程中应充分利用这些宝贵资源。鉴于此,本文在充分整合已有样点的基础上,设计了一套逐次、高效的补样方案,并在黑龙江省鹤山农场老菜河流域进行了实验研究。

2 基于已有样点的数字土壤制图补样设计方法

已有样点中含有采样者对当地土壤—环境关系知识的部分认知,在环境条件越相似土壤属性就越相似的假设下^[13],每一个样点都可以代表与之环境条件相似的地区。因此每个样点都具有一定的“个体代表性”,能够用于合理推测其所能代表区域的土壤属性值。若已有样点集蕴含的土壤—环境关系的全局代表性不好,则这些样点所能代表的空间范围就不能覆盖整个研究区。因此,需要在已有样点集不能代表的范围内,设计补充样点来完备样点集的全局代表性,扩增样点集的空间代表范围至整个研究区。

刘京提出了基于样点个体代表性,推测土壤属性并度量推理不确定性的方法^[14]。推理不确定性信息反映了已有样点集对待推测点的代表程度,即不确定性越低,说明已有样点集对待推测点的代表程度越高、推理结果的可靠程度也越高;反之,不确定性越高,推理结果的可靠程度就低。对不确定性空间分布图设置阈值,可将研究区划分为样点集能够代表的范围和不能代表的范围两部分。在不确定性低于阈值的区域,样点集对其具有较好的代表性,属于样点集能够代表的范围。在这些地区,以样点集与待推测点间的环境相似性为权重,综合

样点集的土壤属性值,可以合理地推测土壤属性值;不确定性高于阈值的区域是样点集无法代表的区域,本文在这些区域补充设计尽量少的样点,使样点集的空间代表范围覆盖整个研究区,详细流程如图1所示。

由图1可见,基于已有样点设计补充样点的方法主要包括3部分的内容:①选择与土壤属性空间特征协同变化的环境变量;②度量基于已有样点推测土壤属性的不确定性,并对不确定性设置阈值,确定已有样点集的空间代表范围;③设计尽量少的样点补充到已有样点集,使新样点集的空间代表范围能够覆盖整个研究区。

2.1 选择协同环境变量

选择与土壤属性协同变化的相关环境变量是计算待推测点与已有样点相似性、度量推理不确定

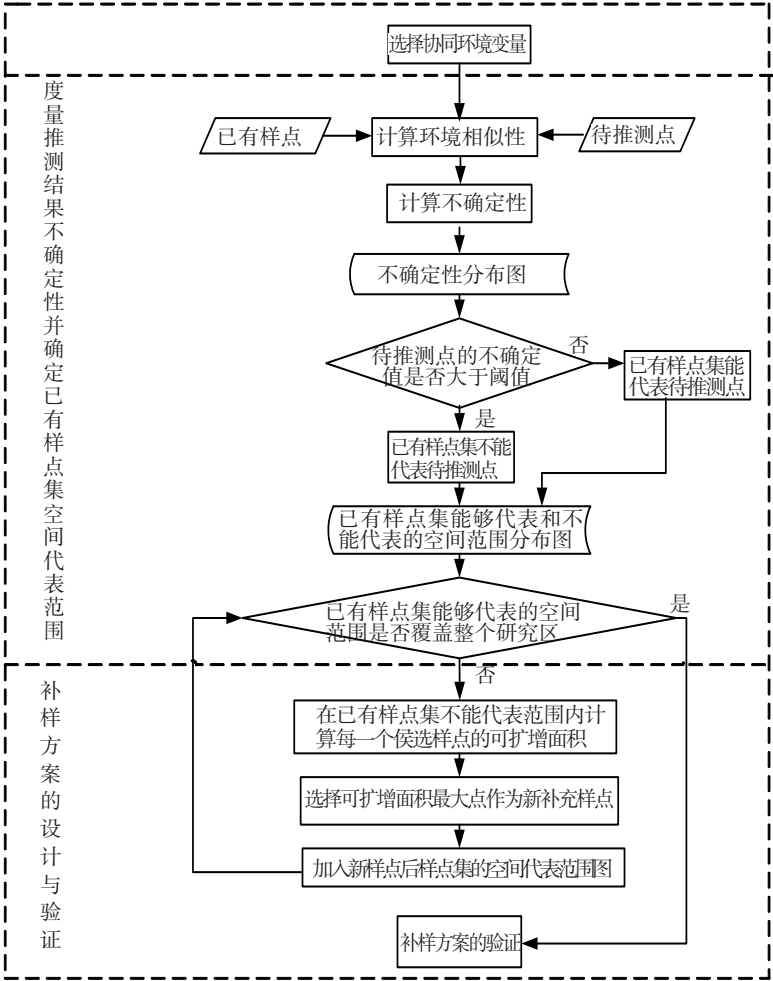


图1 基于已有土壤样点设计补充样点的流程图
Fig.1 Flow chart of the additional sampling scheme based on existing soil samples

性的第一步。选择协同环境变量可以从两个方面考虑:一方面,在道库洽耶夫(V. V. Dokuchaev, 19世纪末)的土壤发生学理论和Jenny提出的土壤—景观概念模型的(“CLORPT”函数)指导下,从气候、地形、母质、植被等成土因素方面考虑选择哪些协同环境变量,进而通过这些协同环境变量的空间变化推测土壤属性的空间变化^[15];另一方面,从已经建立的经验关系出发来选择协同环境变量,例如推测土壤有机质含量,根据经验关系可以选择高程、坡度、曲率等变量^[16-17]。土壤形成的过程极其复杂,与土壤属性具有空间协同变化的环境变量很多,对应特定的土壤属性应在充分了解当地土壤环境关系的基础上,选择合适的协同环境变量。采用栅格模型对选取的协同环境变量进行表达,即可在每一个栅格点处形成一个环境特征向量(式1)。

$$\vec{e}_{ij} = (e_{ij}^1, e_{ij}^2, e_{ij}^3, \dots, e_{ij}^m) \quad (1)$$

式中: e_{ij}^m 表示点 (i, j) 位置处第 m 个环境变量的值。

2.2 确定已有样点集的空间代表范围

在环境条件越相似,土壤属性越相似的假设下,已有样点与待推测点的环境相似性越高,利用已有样点推测待推测点土壤属性的推理不确定性就越低,推理结果的可信程度就高,这些地区属于已有样点集能够代表的空间范围。反之,环境相似性低、推理不确定性高的地区属于已有样点集不能够代表的地区。确定已有样点集的空间代表范围主要包括以下3个步骤:

(1) 计算待推测点与已有样点之间的环境相似性。影响土壤发育的环境变量并非单一变量,因此在计算待推测点与已有样点相似性时,需要从两个层面上计算:一是计算基于单个环境变量的相似性;二是综合各环境变量的相似性,确定已有样点与待推测点间综合环境特征的相似性(式2)^[18]。

$$S_{ij}^k = \prod_{v=1}^m (E^v(e_{ij}^v, e_k^v)) \quad (2)$$

式中: S_{ij}^k 是待推测点 (i, j) 与第 k 个样点之间的环境相似度; e_{ij}^v 为待推测点 (i, j) 第 v 个环境变量的特征值; e_k^v 为第 k 个样点第 v 个环境变量的特征值; m 为协同环境变量的个数; E 为计算单个环境变量相似性的函数,该函数的形式取决于环境变量的类型。对于名称量和序数量(如土壤母质类型、植被类型等)可使用阶跃函数(布尔函数),对于间隔量和比率量变量(如温度、坡度等)可计算欧氏距离、马氏距离、GOWER距离等; P 为综合各环境变量相似

性得到两点综合环境相似性的函数,此时除需考虑相关环境变量与土壤属性之间的关系外,还要考虑相关环境变量之间的关系,可采用的综合方法有加权平均、最小限制因子法^[19]等。至此,基于公式(2)可计算待推测点与所有样点间的综合环境相似性,即在每一个栅格位置处均可形成一个相似性向量(式3):

$$\vec{S}_{ij} = (S_{ij}^1, S_{ij}^2, S_{ij}^3, \dots, S_{ij}^n) \quad (3)$$

该向量中的每一个元素 S_{ij}^n 是待推测点 (i, j) 与第 n (案例点)个已有样点的综合环境相似性^[14]。

(2) 度量推理结果的不确定性。环境相似性向量 \vec{S}_{ij} 反映了已有样点集对待推测点的代表程度,通过分析环境相似性向量,度量基于已有样点推测土壤属性的推理不确定性^[18](式4)。

$$Uncertainty_{ij} = 1 - \max(S_{ij}^1, S_{ij}^2, S_{ij}^3, \dots, S_{ij}^n) \quad (4)$$

式(4)说明,如果待推测点 (i, j) 与对其代表性最好(即环境最相似)的已有样点之间的环境相似性仍较低,那么用已有样点推测该点的土壤属性会产生较高的不确定性,根据式(4)可以产生推理不确定性空间分布图。

(3) 在不确定性信息分布图的基础上,确定已有样点集的空间代表范围。对推理不确定性空间分布图设置一定的阈值,不确定性低于阈值的区域,已有样点对待推测点的代表性较好,以样点集的土壤属性推测待推测点属性可信程度较高,属于已有样点能够代表的范围。否则,不确定性高于阈值的区域,属性推理结果的可靠程度较低,属于已有样点集不能够代表的空间范围。

2.3 逐次确定补充样点

若已有样点集的全局代表性不好,则其空间代表范围不能够覆盖整个研究区,而一张不完整的土壤图不符合制图的要求。因此,必须在已有样点集不能代表的空间范围内,设计尽可能少的补充样点以完善样点集的全局代表性,在扩增可推测面积的同时降低采样成本。

为充分利用已有样点资源,本文在不确定性分布图的指导下,在已有样点集不能代表的空间范围内,采用遍历的方法依次确定补充样点的位置和数量。具体方法为:首先,将已有样点集不能代表的空间范围内每一个栅格都看作一个候选样点,按照2.2中计算已有样点与待推测点环境相似性并度量不确定性的方法,计算每一个候选样点加入到已有

样点集后,新样点集与待推测点的环境相似性和推理不确定性,通过比较不确定性与设定阈值的大小,确定每一个候选样点的可扩增面积。然后,选择使新样点集空间代表范围扩增最大的点作为第一个补充样点。最后,基于补充样点和已有样点,更新样点集的空间代表范围图。重复上述过程直至新样点集的空间代表范围能够覆盖整个研究区为止,便可设计出高效的补充样点。

相比于传统的样点设计方案,基于已有样点的数字土壤制图补样方案不仅能够提供补充样点的位置和数量,而且能够提供样点的先后顺序。这是因为先设计的样点比后设计的样点具有更好的个体代表性,这在采样资源有限的情况下,为采样者合理的选择样点提供了重要依据。

3 应用案例分析

3.1 研究区及已有样点数据

研究区位于嫩江县鹤山农场老莱河流域,地理位置:48°53'24"~48°59'24"N, 125°8'24"~125°16'12"E,面积约60.2 km²。该区处于寒温带季风草甸草原区,年降水量400~550 mm;母质类型基本一致,多为黄土状亚粘土和重壤土物质;地貌主要是不同程度切割的山前洪积台地以及冲积湖积平原,当地称之为漫岗地;地形总体起伏平缓,环境梯度小;植

被受人为影响强烈,主要农作物有小麦和大豆。在上述环境条件的综合作用下,本区的主要土壤类型为黑土与草甸土^[20]。

在老莱河流域,共搜集到102个野外样点(图2)。样点的采集方式分别为规则采样(48个)、横截面采样(20个)、目的性采样(24个)及主观采样(10个)。规则采样以0.6纬度×0.6经度为间隔分布设计,试图通过均匀的采样来获取研究区内土壤属性空间变化的整体特征;两条横截面采样,样点的平均距离为100 m。在流域中选择横穿山坡、沟谷的切线采集样点,能够在较短的距离内穿越主要的景观类型;目的性采样能够以较少的样点捕捉到研究区内土壤—环境关系的全局特征^[21]。主观采集的样点布设在上述规则采样、横截面采样、目的性采样没有采集到的景观部位,例如:山顶、河谷和陡坡位置。

3.2 选择协同环境变量

研究区面积较小,在60.2 km²的范围内气候和母质相对比较均一,天然植被遭受破坏,以玉米和大豆为主的人工植被已无法有效反映土壤属性的空间变化,地形成为有效指示土壤属性空间变化的主要因子。因此,选择能够刻画本区地形特征的环境变量作为土壤属性的协同环境变量(表1),5个环境变量均派生于1:10000地形图数字化后产生的数字高程模型^[22]。

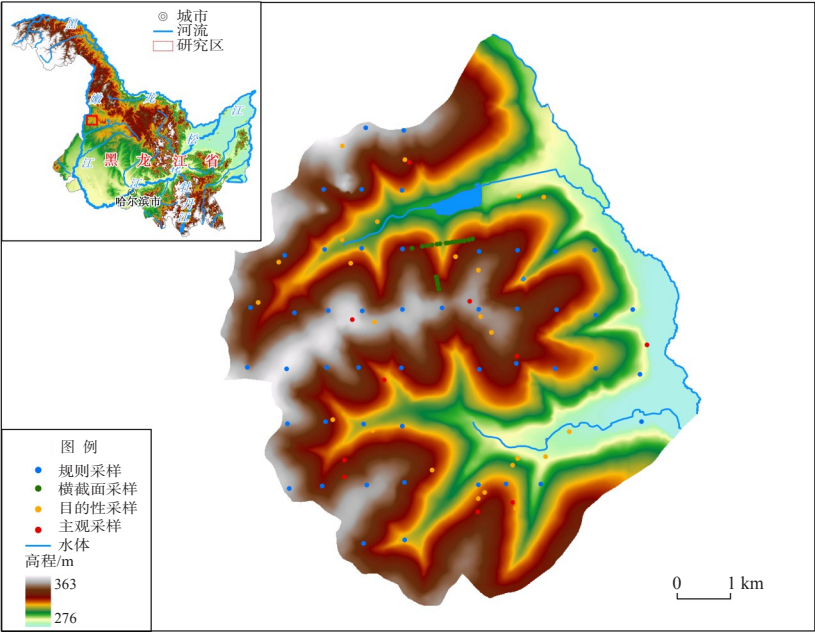


图2 研究区土壤样点分布图
Fig.2 Spatial distribution of all existing soil samples

3.3 设计补充样点

将已有样点集不能代表范围内每一个栅格都看作候选样点,然后计算每一个候选样点加入已有样点集后新样点集的空间代表范围,最后选择能够使新样点集的空间代表范围扩增最大的候选样点作为补充设计的第一个样点。重复上述过程直至新样点集的空间代表范围覆盖整个研究区,即可设计出具有先后重要性次序的补充样点,具体方法详见2.3。

为便于研究,本文通过以下方式验证基于已有样点的数字土壤制图补样方案的有效性。将已有的102个样点分成3部分,5个作为已有样点(图

3a),47个作为验证样点(图 3b),50个作为候选的补充样点(图 3c)。3分样点通过以下方式划分:首先从102个样点中随机挑选5个作为已有样点,然后将横截面采集的20个样点与隔列选择的27个规则

表1 与土壤属性协同变化的环境变量
Tab.1 Environment co-variables with soil spatial variation

环境变量		分辨率/m
地形	坡度	10
	沿等高线曲率	10
	沿剖面曲率	10
	地形湿度指数	10
	坡位指数	10

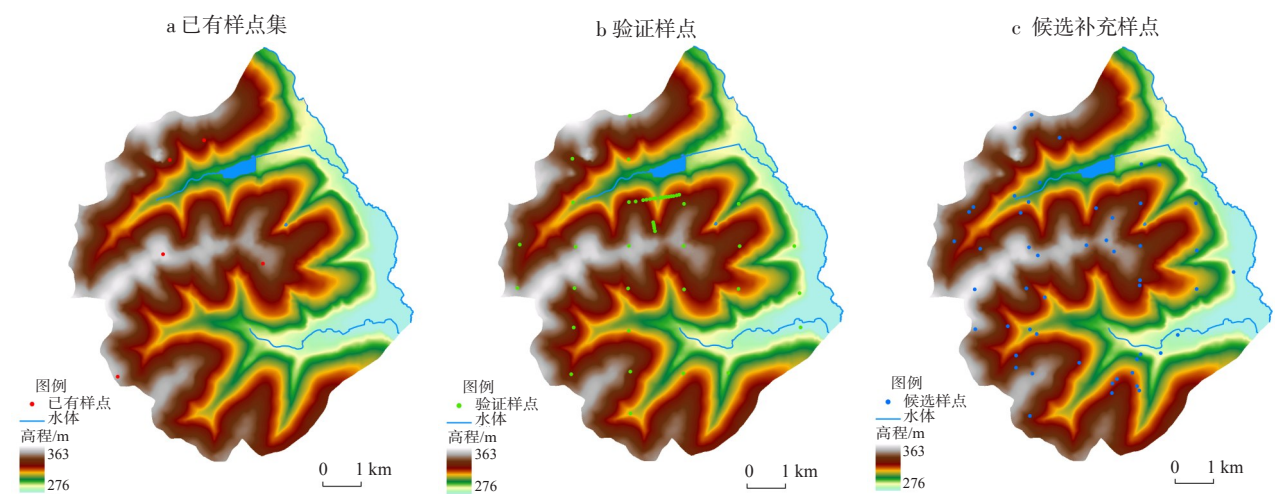


图3 各样点空间分布图
Fig.3 Spatial distribution of existing validation and candidate samples

样点共47个样点作为验证点,最后将剩余的50个样点作为可供选择的候选补充样点。

在计算单个环境变量相似性时,式(2)中的 E 函数采用GOWER距离,该距离范围为0~1,不需要进行标准化处理即可作为样点间单个环境变量的相似性。在综合各环境变量的相似性获取两点间综合环境相似性时,式(2)中的 P 函数选择最小限制因子法。然后,通过公式(4)度量基于5个已有样点推测土壤属性时的推理不确定性。最后,设置不确定性阈值为0.2,在推理不确定性分布图的基础上确定已有样点集的空间范围(图4)。其中,绿色部分为已有样点集能够代表的部分,即用已有样点推测该区土壤属性时,推理不确定性小于0.2,推理结果的可靠性较高。粉色部分为已有样点集无法代表的区域,该区主要位于沟谷部位。这是由于5个已有样点都分布在山坡上,样点位置的环境特征与沟谷相差很

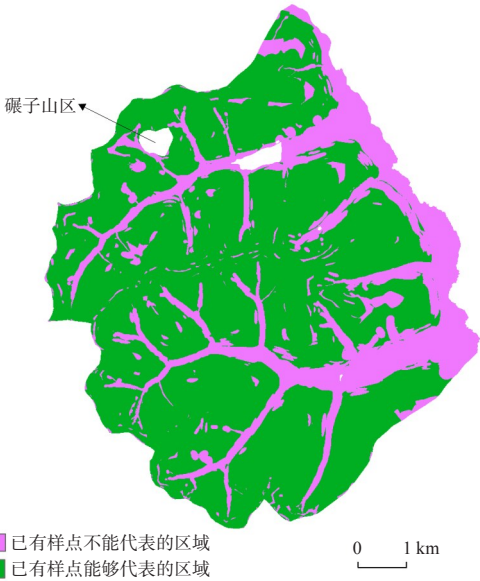


图4 已有样点集不能代表的空间范围图
Fig.4 Locations that existing samples cannot represent

大,因此无法利用已有样点推测沟谷位置的土壤属性。另外,碾子山地区是矿区,地形特征与其他区域明显不同,因此对其进行掩模处理。

确定了已有样点集不能代表的空间范围之后,即可进行补充样点的设计。需要注意的是,应用实验中,不是将已有样点集不能代表范围内的所有栅格都看作候选样点,而是只计算可供选择的50个候选样点分别加入已有样点集后新样点集的空间代表范围,然后选择使空间代表范围扩增最大的点作为第一个补充样点。之后,将该补充样点加入到已有样点集中,基于新样点集更新空间代表范围图。重复上述过程,直至样点集的空间代表范围能够覆盖整个研究区或空间代表范围不再扩增为止。

图5中,横坐标为补充样点的次序,纵坐标为对应样点的可扩增面积。应用实验中只确定了33个补充样点的位置和先后次序,这是因为加入33个补充样点后新样点集的空间代表范围包含了剩余17个候选补充样点的空间代表范围。也就是说,如果再将17个候选补充样点加入到样点集中,样点集的空间代表范围不会增加。从图5中可见,次序较高的样点,即首先被选为补充样点的候选样点,其可扩增面积要远大于次序较低的样点。据此,在采样资源有限的情况下,样点次序为采样者合理地选择补充样点提供了重要依据,即应优先选取样点次序高(重要性高)的样点。

3.4 补样方案的验证

在5个已有样点的基础上,按照补充样点的先后次序每次只加入1个样点到已有样点集中,形成样点个数分别为6个,7个,8个,...,38个的33套新样点集。基于每套新样点集,利用样点个体代表性推测土壤属性的方法,推测了33张土壤表层有机质空间分布图^[14]。然后,利用47个独立验证点对33张土壤表层有机质空间分布图分别进行验证。选用均方根误差(RMSE)为验证指标,该指标度量了推测值相对于实际观测值的偏差,其计算公式如式(5)。RMSE越低,说明推测值与实际观测值越接近,推测结果越好。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (V_{pre}^i - V_{obs}^i)^2}{n}}$$
 (5)

式中: V_{pre}^i 为第 i 个验证点的推测值; V_{obs}^i 为第 i 个验证点的实际测量值; n 为验证点的个数。图6显示了随补充样点的增加,RMSE的整体趋势在降低,

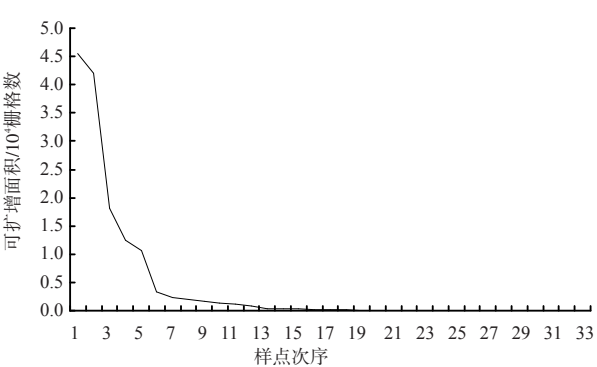


图5 样点集空间代表范围的扩增面积随补充样点的变化
Fig.5 Variation of additional represented area with additional samples

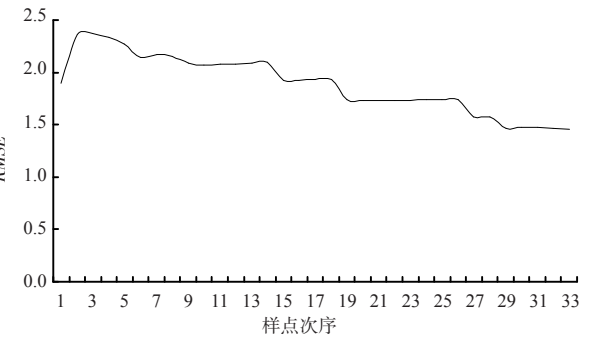


图6 RMSE随补充样点的变化
Fig.6 Variation of RMSE with additional samples

即土壤表层有机质的推理精度在提高。由图可见,在第二个补充样点加入后,RMSE突然升高,这是因为第一个补充样点加入到已有样点集时,样点集的空间代表范围仍较小,只能代表47个验证样点中的39个,此时实际参与的验证点个数为39。当第二个补充样点加入之后,样点集的空间代表范围进一步扩大,能够代表所有验证样点,实际参与的验证点个数为47。验证点个数不同时,RMSE不具有可比性,这是造成RMSE开始陡然增高的原因。之后,在验证点数量相同的情况下,RMSE随补充样点个数的增加呈不断下降的趋势,而且先补充加入的样点能使RMSE更快程度的降低,这进一步说明了补充样点次序的重要性。

4 结论与讨论

在黑龙江省嫩江县老莱河流域样点设计实验中,基于已有样点的数字土壤制图补样方案能够以较少的样点最大程度地扩增样点集的空间代表范

围,而且能够给出样点的先后次序,这说明该补样方案是逐次地、高效地。与规则布样、随机布样及克里格布样等常用样点布设方案相比,该补样方案的优势主要体现在以下两个方面:

(1) 基于已有样点的数字土壤制图补样方案,充分整合了数量有限、分布不符合某种设计规则、典型性得不到保证的已有样点资源。这为样点设计过程中,有效利用研究区内不同采样目的和不同来源的样点数据提供了可借鉴的方法。

(2) 基于已有样点的数字土壤制图补样方案,每一个新样点的布设都是建立在上一次推理结果的基础上,所以该补样方案是逐次地、渐进地推选;这些具有先后补充顺序(样点的重要程度而非实际采样顺序)的样点,在采样资源有限的情况下,为合理地选择样点提供了重要依据。

在环境条件越相似、土壤属性越相似的假设下,可以通过度量已有样点与待推测点间综合环境条件的相似性,确定样点集的空间代表范围。因此,该方法应用的关键在于选择合适的协同环境变量,尽可能详细准确地描述土壤发生地的环境组合特征。另外,通过逐次补充样点可确保新样点集的空间代表范围能够覆盖整个研究区,但当样点集的空间代表范围覆盖整个研究区以后,应如何设计尽可能少的样点最大程度地降低推理结果的不确定性、提高土壤属性推理精度将是下一步研究的重点。

参考文献

- [1] 王珂,沈掌泉, Bailey J B, 等. 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.
- [2] Mueller T G, Pierce F J. Soil carbon maps: Enhancing spatial estimates with simple terrain attributes at multiple scales. *Soil Science Society of America Journal*, 2003, 67(1): 258-267
- [3] 朱会义, 贾绍凤. 降雨信息空间插值的不确定性分析. 地理科学进展, 2004, 23(2): 34-42.
- [4] Goovaerts P. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 1999, 89: 1-45.
- [5] Webster R, Oliver M A. *Statistical Methods in Soil and Resource Survey*. Oxford: Oxford University Press, 1990: 29-30
- [6] 王劲峰, 李连发, 孟斌, 等. 空间分析. 北京: 科学出版社, 2006: 129-136.
- [7] 姜成晟, 王劲峰, 曹志冬. 地理空间抽样理论研究综述. 地理学报, 2009, 64(3): 368-380.
- [8] 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. 地理科学进展, 2008, 27(3): 152-160.
- [9] Zhu A X, Yang L, Li B L, et al. Construction of Membership Function for Predictive Soil Mapping under Fuzzy Logic. *Geoderma*, 2010, 155(3-4): 164-174..
- [10] Zhu A X, Yang L, Li B L, et al. Purposive sampling for digital soil mapping for areas with limited data//Hartemink A E, McBratney A B, Mendonca Santos M L. *Digital Soil Mapping with Limited Data*. New York: Springer-Verlag, 2008: 233-245.
- [11] 杨琳, 朱阿兴, 秦承志, 等. 基于典型点的目的性采样设计方法及其在土壤制图中的应用. 地理科学进展, 2010, 29(3): 279-286.
- [12] 赵量, 赵玉国, 李德成, 等. 基于模糊集理论提取土壤: 地形定量关系及制图应用. 土壤学报, 2007, 44(6): 961-967.
- [13] Hudson B D. The soil survey as paradigm-based science. *Soil Science Society of America Journal*, 1992, 56(3): 836-841.
- [14] 刘京. 基于样点个体代表性的大范围土壤属性制图方法研究[D]. 北京: 北京师范大学, 2010: 10-15.
- [15] 朱阿兴, 李宝林, 裴韬, 等. 精细数字土壤普查模型与方法. 北京: 科学出版社, 2008: 22-23.
- [16] 连刚, 郭旭东, 傅伯杰, 等. 黄土丘陵沟壑区县域土壤有机质空间分布特征及预测. 地理科学进展, 2006, 25(2): 112-122.
- [17] 孙孝林, 赵玉国, 赵量, 等. 应用土壤—景观定量模型预测土壤属性空间分布及制图. 土壤, 2008, 40(5): 837-842.
- [18] Zhu A X. Measuring uncertainty in class assignment for natural resource maps using a similarity model. *Photogrammetric Engineering & Remote Sensing*, 1997, 63(10): 1195-1202.
- [19] Zhu A X, Band L E. A knowledge-based approach to data integration for soil mapping. *Canadian Journal of Remote Sensing*, 1994, 20(4): 408-418.
- [20] 杨琳. 目的性采样下样本设计与制图精度的关系研究: 以数字土壤制图为例[D]. 北京: 中国科学院地理科学与资源研究所, 2009: 15-16.
- [21] 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. 土壤学报, 2011, 48(5): 938-946.
- [22] 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): 784-791.

An Integrative Sampling Scheme for Digital Soil Mapping

ZHANG Shujie^{1,2}, ZHU Axing^{1,3}, LIU Jing³, YANG Lin¹

(1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China;

3. Department of Geography, University of Wisconsin-Madison, Madison, WI53706, USA)

Abstract: In most areas there exist some soil field samples (“existing samples data”) which were accumulated through two historical soil surveys and/or specific field studies. The accuracy of soil property map predicted based on this kind of samples is lower, because the size of these samples is limited and the global representativeness is poor. However, these samples are valuable resource because they do capture the relationship between soil and environment conditions. This paper presents a stepwise and effective method to design additional samples by integrating existing samples. The aim of the method is to design as few additional samples as possible to complement the global representativeness of the existing samples and improve the accuracy of predictive soil mapping. The detail process is as follows. First, we determine the spatial extent which can be represented by the existing samples. Second, each location (grid) can be considered as a candidate for additional sampling in the area which the existing samples do not cover and the area which can be covered by each candidate was then calculated. Third, the location which can represent the largest additional area was chosen as the first additional sample. Finally, the above procedure was iterated until the new sample set covered the whole study area. This integrative sampling scheme can not only determine the number and the locations of additional samples, but also give the order of the samples. It provides samplers with the important information on how many points can be sampled reasonably when the available resources are limited.

Key words: existing samples; global representativeness; individual representativeness; prediction uncertainty; stepwise additional sampling designing scheme

本文引用格式:

张淑杰, 朱阿兴, 刘京, 等. 整合已有土壤样点的数字土壤制图补样方案. 地理科学进展, 2012, 31(10): 1318-1325.