DOI: 10.13758/j.cnki.tr.202508140318 **CSTR: 32214.14.tr202508140318**
郭文蓉, 赵治东, 邱江恒, 等. 利用土壤类型典型虚点更新历史土壤图方法探索. 土壤, 2026, 58(4): 996–1007.

利用土壤类型典型虚点更新历史土壤图方法探索①

郭文蓉[1,2], 赵治东[2,3], 邱江恒[2,4], 徐胜祥[2,5], 曾 荣[1*], 刘 峰[2,5*]

(1 南京信息工程大学地理科学学院, 南京 210044; 2 土壤与农业可持续发展全国重点实验室(中国科学院南京土壤研究所), 南京 211135; 3 中国农业大学土地科学与技术学院, 北京 100083; 4 河南农业大学林学院, 郑州 450046; 5 中国科学院大学, 北京 100049)

摘 要: 通过挖掘历史土壤图中的土壤–景观关系提取虚点是当前历史土壤图更新的重要方法。然而, 现有的虚点拾取方法在代表性与空间覆盖性方面存在一定局限性。因此, 本研究基于成土环境特征, 构建了土壤典型景观部位识别算法, 用于提取具有代表性的土壤典型虚点。采用 860 个样本作为训练数据, 设置 172 个独立验证样本(包括 60 个野外校核样本和 112 个随机典型虚点), 以图斑中心点法和随机点法作为对照方法。选取高程、短波红外、母质类型、有机质含量、地形湿度指数、归一化植被指数作为关键环境因子, 构建典型虚点提取方法, 并采用多层随机森林对不同虚点拾取方法进行预测制图。结果表明, 相较于图斑中心点法和随机点法, 典型虚点法显著提升了历史土壤图更新的精度(准确率分别提升了 18% 和 8%, Kappa 系数分别提升了 0.20 和 0.10, CI 分别降低了 0.12 和 0.04)。本研究通过优化虚点选取策略, 提高虚点的代表性和质量, 为历史土壤图更新制图研究提供了方法参考。

关键词: 土壤图更新; 土壤–环境关系; 数字土壤制图; 典型景观部位; 典型虚点

中图分类号: S159 **文献标志码:** A

An Exploratory Study on Method of Updating Historical Soil Maps Based on Typical Soil Pseudo-points

GUO Wenrong[1,2], ZHAO Zhidong[2,3], QIU Jiangheng[2,4], XU Shengxiang[2,5], ZENG Rong[1*], LIU Feng[2,5*]

(1 *School of Geographical Sciences*, *Nanjing University of Information Science and Technology*, *Nanjing* 210044, *China*; 2 *State Key Laboratory of Soil and Sustainable Agriculture*, *Institute of Soil Science*, *Chinese Academy of Sciences*, *Nanjing* 211135, *China*; 3 *College of Land Science and Technology*, *China Agricultural University*, *Beijing* 100083, *China*; 4 *College of Forestry*, *Henan Agricultural University, Zhengzhou* 450046, *China*; 5 *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*)

Abstract: Extracting pseudo points by mining soil–landscape relationships embedded in historical soil maps is an important method for updating these maps. However, existing pseudo-point selection strategies suffer from limited representativeness and spatial coverage. To address this, this study developed a soil-typical landscape-position identification algorithm, grounded in pedogenetic environmental characteristics, to extract representative typical pseudo-points. A total of 860 samples were used for training and 172 independent validation samples for validation (60 field-verification samples and 112 randomly selected typical pseudo-points), with the polygon-centroid and random-point methods serving as baselines. Elevation, short-wave infrared, parent material, soil organic matter content, topographic wetness index, and normalized difference vegetation index were selected as key environmental covariates to construct the typical pseudo-point extraction scheme; predictive mapping under different pseudo-point selection methods was performed using a multi-layer random forest. Results showed that, relative to the polygon-centroid and random-point methods, the typical pseudo-point approach markedly improved historical soil map updating accuracy (overall accuracy increased by 18% and 8%, respectively; Kappa by 0.20 and 0.10; and CI decreased by 0.12 and 0.04, respectively). By optimizing pseudo-point selection, this study enhanced the representativeness and quality of pseudo points, offering a methodological reference for updating historical soil maps.

Key words: Soil map updating; Soil-environment relationships; Digital soil mapping; Typical landscape positions; Typical pseudo-points

---

①基金项目: 国家自然科学基金面上项目(42571077)和中国科学院南京土壤研究所十四五规划自主部署项目(ISSASIP2202)资助。

* 通信作者(rzeng@nuist.edu.cn; fliu@issas.ac.cn)

作者简介: 郭文蓉(1999—), 女, 山东淄博人, 硕士研究生, 主要研究方向为数字土壤制图。E-mail: guowenrong23@mails.ucas.ac.cn

20 世纪 80 年代，我国开展了第二次全国土壤普查(以下简称“二普”)[1]，系统编制了覆盖全国的土壤类型图，至今仍在国土规划、农业生产以及生态环境保护等领域发挥着重要的作用。然而，受当时制图技术、数据质量的限制，历史土壤图在空间精度与时效性方面难以全面反映当前土壤资源的实际状况。随着“3S”(遥感、地理信息系统和全球定位系统)技术的快速发展，数字土壤制图技术通过构建土壤–景观关系模型，结合多种数学方法和空间分析方法，实现土壤属性的推理与预测。凭借其高效性、低成本、高精度等优势逐渐得到广泛应用[2]。尽管从技术原理上讲，利用数字土壤制图技术更新历史土壤图已具备成熟条件，但在实际应用中仍面临一些挑战，其中样本数据的获取与代表性是最突出的问题[3]。

在土壤类型制图的研究领域内，当前普遍采用的方法是结合样点与环境变量，构建土壤–景观关系[4]，从而对未知区域的土壤类型进行空间推断和预测制图。获取样点的方式主要有两种：一是通过人工野外采样获取真实样点[5]。该方法具有时效性强、准确性高的优点，能够确保样点的真实性和代表性，但其实施过程需要投入大量的人力物力。二是基于历史土壤图信息挖掘的虚拟样点提取法[6]，即“虚点法”。虚点法是指在缺乏实地采样的区域，依据历史土壤图的空间分布和属性特征，通过规则约束或模型模拟生成虚拟样点的方法。该方法以历史土壤图为基础，结合数据挖掘技术，在目标区域内按既定规则生成具有代表性的样点，用于建模预测。该方法充分利用已有的历史土壤图数据，减少对实地采样的依赖，具有成本低、效率高等优势[7]。在开展大规模历史土壤图更新工作时，数字土壤制图技术对样本数量的需求较大，仅依靠野外采样难以满足这一需求。因此，从历史土壤图中提取虚点成为一种有效的补充方法，既能保证样本数量，又能维持样本的代表性[8]。

样点数量和空间位置的确定对于选取代表性样点至关重要。在样点数量的确定方面，研究者们提出了一些方法。例如，刘雪琦等[9-10]确定了一种按土壤类型图斑面积分级的训练点选择策略，而 Heung 等[11-12]使用土壤图斑面积加权法来确定训练点的数量。针对样点空间位置的确定，不同研究也提出了多种方法，如杨雨菲等[13]、周紫燕等[14]、孟可等[15]使用随机点法，这种方法简单易行，但可能受随机性影响较大，无法充分考虑土壤类型的空间分布特征和环境变量的影响。高鸿等[16-17]提出在每个土壤多边形内提取训练点的方法，确保所有土壤类型区域均匀覆盖，但在一定程度上忽略了环境变量对土壤类型的影响。刘京等[18]基于“个体代表性”的样本选择方法，通过其所处环境条件的相似程度来衡量代表程度，但缺乏良好的全局代表性。王子龙等[19]使用优先级指数，通过分析与目标要素相关的环境变量空间分布，赋予权重并计算优先级指数，将区域划分为不同优先级等级，但依赖权重设定的合理性。Qi 和 Zhu[7]通过环境变量的直方图分析选择典型样本，但不考虑环境变量组合对成土环境的影响，仅选择单一环境变量作为输入数据，没有考虑环境变量组合共同表征的成土环境关系，可能不适用于随机森林等数据挖掘模型[17]。综上所述，现有的样点位置选择方法在代表性、空间覆盖性等方面存在一定局限性。

因此，本研究以江苏省仪征市为研究区，基于典型环境特征的虚点优化提取方法，通过整合历史土壤图和多源成土环境变量之间的关系，构建土壤类型与环境因子的频率分布直方图，系统识别典型环境组合特征，并筛选出具有代表性的样点。该方法旨在从样点选择的角度提升样本代表性，从而提高历史土壤图更新的精度。

1　材料与方法

1.1　研究区概况

仪征市位于江苏省中西部，隶属于扬州市，南临长江，北与安徽省天长市接壤(图 1)，属于中亚热

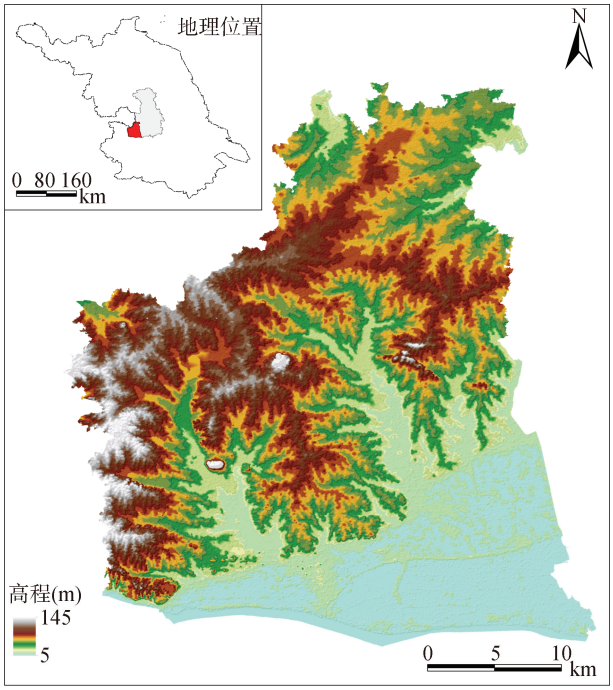


图 1　研究区地理位置与地形示意图

Fig. 1　Geographical location and topographic map of study area

带季风气候，年均温为 15.7 ℃，年均降水量为 1 058.9 mm。地形自西北向东南逐渐倾斜，地貌类型主要包括丘陵、岗地和平原。成土母质有基岩风化残积坡积物、下蜀黄土母质、黄土淤积物、长江冲积物。根据图 2 可知，仪征市的土壤划分为 4 个土类、6 个亚类、10 个土属、15 个土种，共有 890 个土种图斑。本研究基于土种级别，在土壤图斑单元中拾取虚点进行研究。

1.2 数据来源

1.2.1 历史土壤图　本研究采用的基础数据包括：1∶50 000 比例尺的第二次全国土壤类型图，以及 2022 年夏季高分一号卫星 2 m 分辨率遥感影像。在高分遥感影像支撑下通过目视解译，勾绘并剔除了非农用地地块，最终提取出土地利用类型未发生变化的区域。

1.2.2 环境变量　土壤的空间分布是五大自然成土因素(气候、生物、母质、地形和时间)和人为活动长期综合作用的结果[20]。基于 SCORPAN 土壤景观模型[21]，本研究选取了 23 个环境变量(表 1)，以获取仪征市土壤–景观知识。

在地形方面，基于 SAGA GIS 软件，使用 12.5 m 分辨率的数字高程模型计算了与研究区土壤分布最相关的地形变量。这些地形特征能够体现小气候对土壤发育的影响，并且对整个景观中的水分运动过程具有控制作用。此外，通过将经度与纬度整合为一个单一的空间变量构建了斜坐标，更有效地表征了研究区地形由西北向东南倾斜的整体趋势。在成土母质方面，采用了全国地质资料馆提供的 1∶250 000 地质图表征母质类型。此外，为较好地表征受长江冲积作用影响的母质类型，计算了研究区与长江的欧氏距离。同时，计算了与周边山体的欧氏距离，用以反映坡积物和淤积物等母质的空间分布特征。在植被和地表状态等方面，利用 GEE 云平台，基于近 5 年的 Landsat 8 时间序列影像，计算了地表温度的标准差，以反映土壤中水分运动情况[22]。同时，基于近 3 年的 Sentinel-2 时间序列影像，计算了 3 个植被指数变量，以反映土壤的生产力水平及差异。通过双线性插值法对所有环境变量进行空间

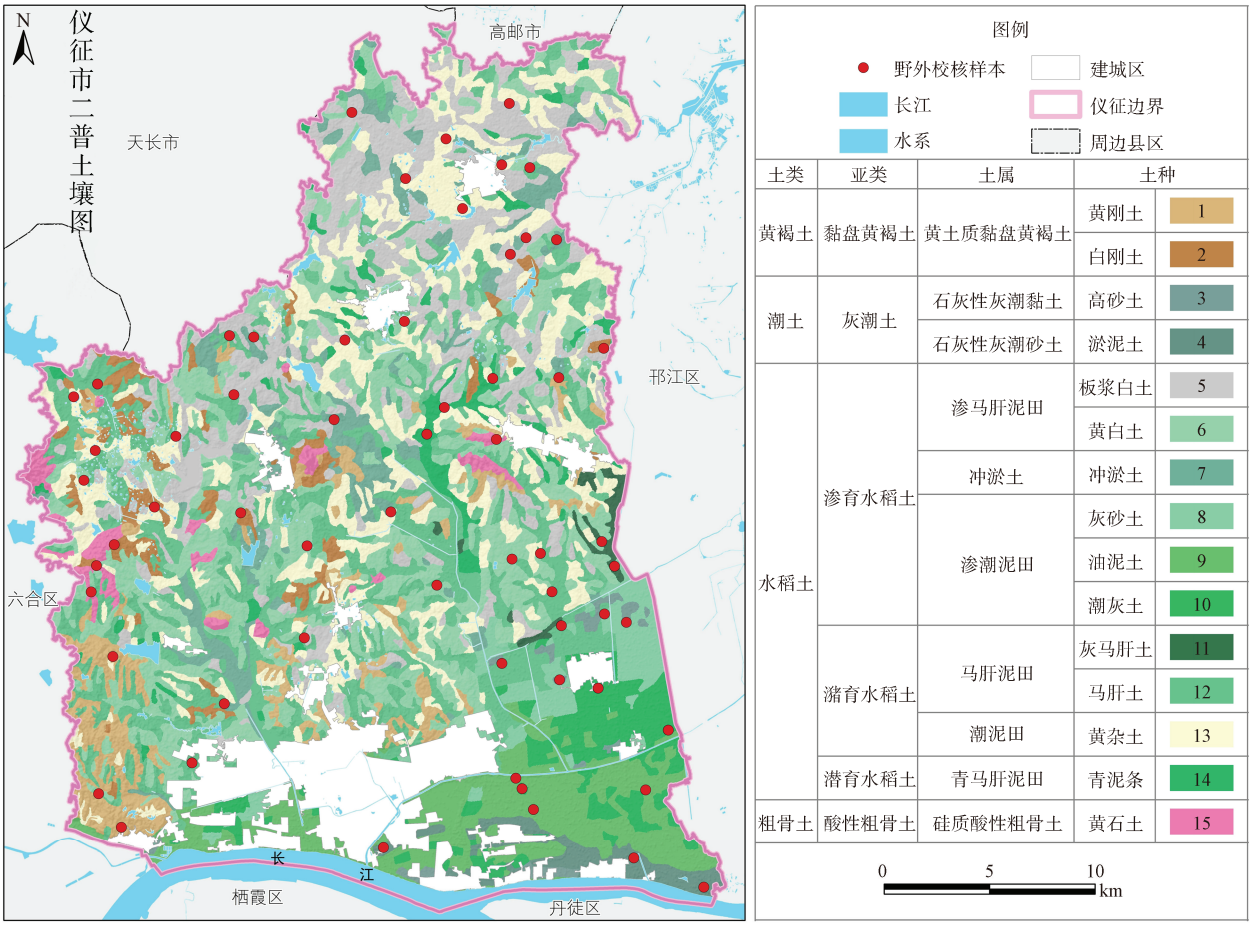


图 2　研究区历史土壤图

Fig. 2　Historical soil map of the study area

表 1　环境变量

Table 1　Environmental covariates

变量类型	变量名称	环境因素类型
地形	高程	r
	地形湿度指数	r
	坡度	r
	剖面曲率	r
	坡位	r
母质	母质类型	p
植被	归一化植被指数	o
	增强植被指数生长季峰值	o
	增强植被指数标准差	o
	增强植被指数增长速率	o
地表水热状态	地表温度标准差	c, s
	短波红外	c, s
地理位置	距内部河道距离	r, p
	距长江距离	r, p
	距山体距离	r, p
	斜坐标	r
土壤属性	pH	s
	砂粒含量	s
	粉粒含量	s
	黏粒含量	s
	粉黏比	s
	有机质含量	s
人类活动	土地利用数据	a

注：s 表示土壤；c 表示气候；o 表示生物；r 表示地形；p 表示母质；a 表示人类活动。

标准化处理，统一重采样至 30 m 空间分辨率，以确保数据在空间尺度上的一致性。

1.3　典型虚点拾取方法

基于土壤类型在自然状态下具有时空稳定性的理论基础[23]，本研究首先剔除建设用地、水体以及地类变更区等干扰区域，进而提取土壤类型保持相对稳定的区域范围，最终构建具有时空一致性的土壤类型图作为样点提取的基础数据。

与熟悉当地土壤类型分布特征的专家深入讨论，以明确较为关键的环境变量。通过分析讨论，在典型虚点拾取阶段选取 6 个关键环境变量：高程、短波红外、母质类型、有机质含量、地形湿度指数、归一化植被指数。通过构建土壤类型与关键环境变量之间的频数直方图，实现对各成土环境特征的定量表征。本研究中，土壤图的图斑受比例尺限制，通常仅能表征主要土壤类型，单个图斑中仍可能存在面积较小、非优势的土壤类型[24]。为精准拾取到典型的优势土壤单元，本研究采用等宽离散化方法，将关键环境变量取值范围划分为 100 个均等区间(图 3)，有效降低了因变量频数分布不均导致的代表性误判风险，进而提高典型成土环境识别的准确性。

在典型数值区间识别方面，以关键环境变量频数直方图中峰值频数的 50% 作为判别阈值，将频率大于 50% 的环境范围定义为典型区间。随后，通过地

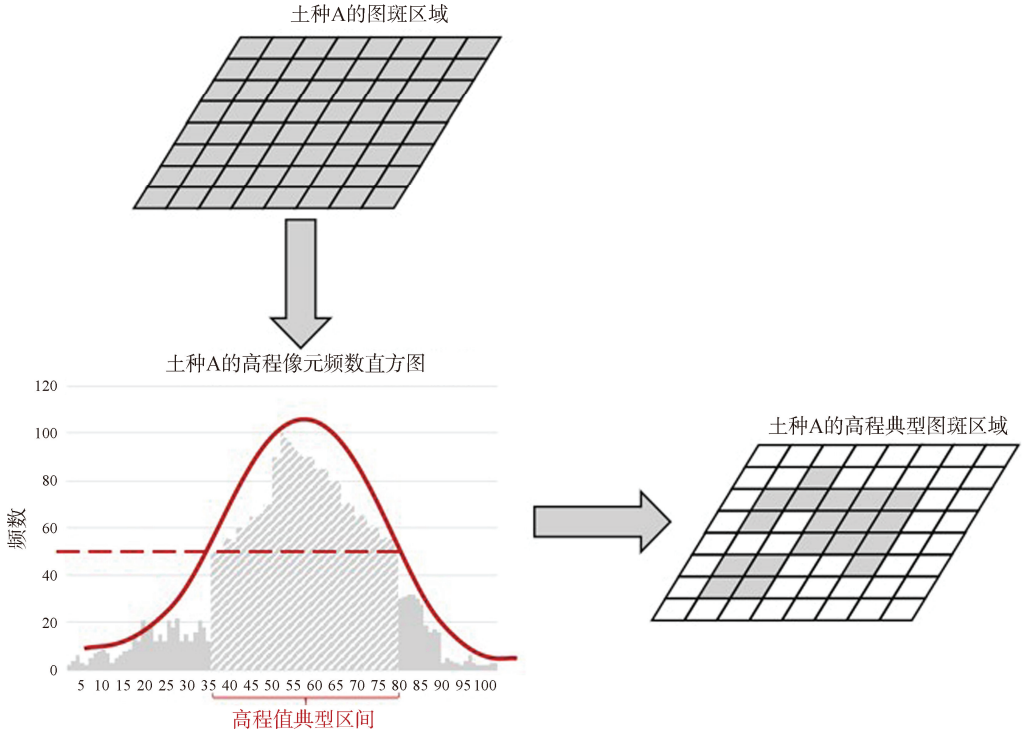


图 3　单个变量识别的典型景观区域

Fig. 3　Typical landscape area based on single variable identification

理空间映射技术，实现典型区间的可视化表达(图 3)。该方法不仅有效减少了数据噪声干扰，而且更强化了土壤类型与环境变量之间的响应关系。

基于上述方法体系，进一步开展典型成土环境组合的精准识别。首先，针对每种土壤类型计算其对应关键环境变量的频数分布特征；其次，通过空间映射生成各关键变量的典型分布图层；最后，采用空间叠置分析方法对各环境变量的典型分布图层进行空间求交，最终获得各土壤类型的关键环境变量典型分布区间。如图 4 所示，通过该方法成功识别出某土种的典型景观部位，其中灰色区域表示该土种的图斑分布范围，红色区域则表示基于环境变量筛选得到的典型景观部位。图 5 展示了油泥土与淤泥土的典型景观部位。

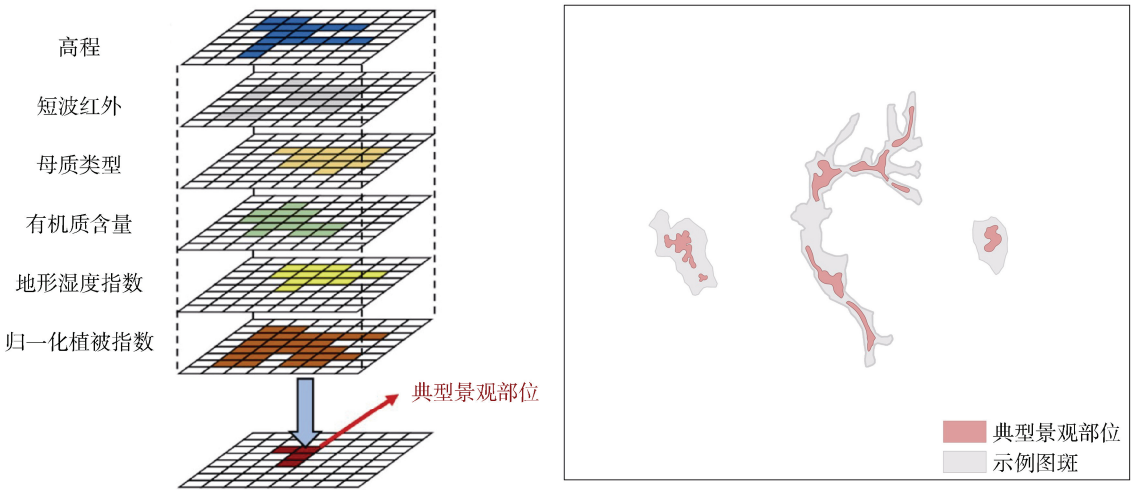


图 4　典型景观部位示例

Fig. 4　Example of typical landscape position

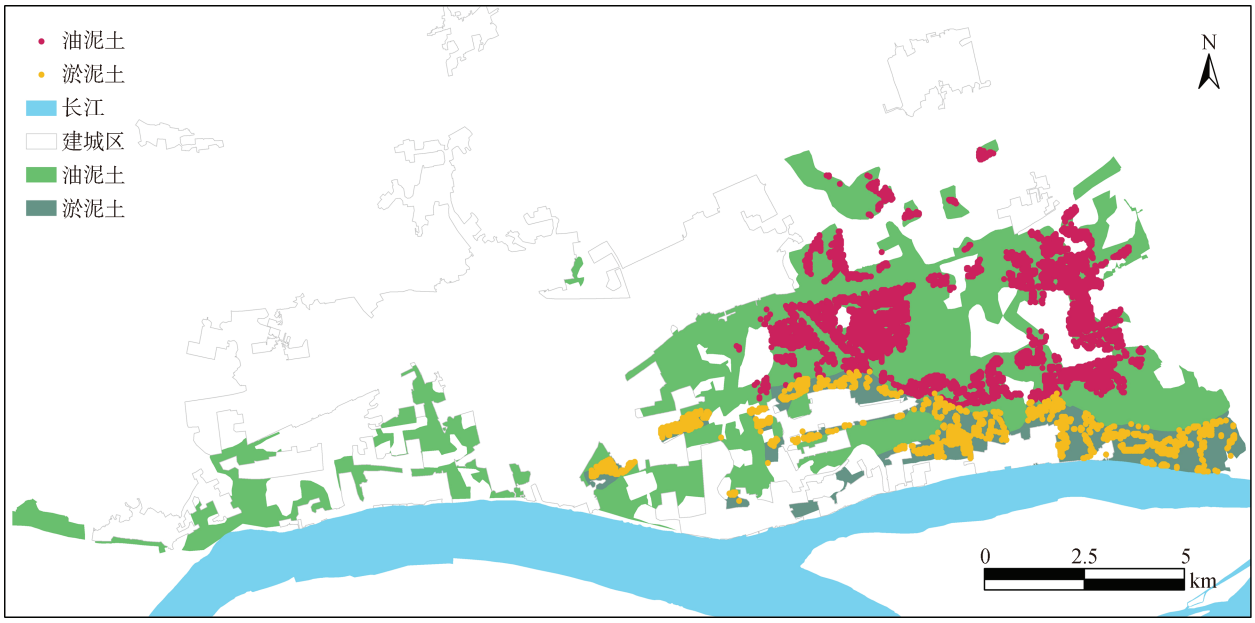


图 5　具体土壤类型典型景观部位

Fig. 5　Typical landscape positions of specific soil types

除典型虚点法外，本研究还采用图斑中心点法[25]和随机点法作为对照方法。这两种方法的特点在于它们不依赖于土壤类型与环境变量之间的先验关系[26]。随机点法通过完全随机的方式选取样点，确保其空间分布的随机性；图斑中心点法则基于土壤图斑的几何中心进行采样，体现图斑的空间代表性。使用 ArcGIS 10.8 软件中的创建随机点和要素转点工具获取。

1.4　训练样点设置

在训练样点设置方面，本研究在刘雪琦等[9-10]提出的基于土壤类型图斑面积分级的训练点拾取方

法基础上，进一步对土壤类型图内单个图斑的面积进行分级，保证样点在县域空间的均匀分布。

$$L = (\mathrm{int}(R - \lg(Np / N)))^2 \quad (1)$$

式中：L 表示各土壤图斑所在等级对应的训练样点比值；Np 表示某图斑的面积；N 表示土壤类型图的总面积；R 表示实数，其值的确定需保证最小等级 I 所对应的比值至少为 1。

为优化样点空间布局并确保采样代表性，所有样点均从已识别的典型景观部位中选取，保证其能够准确反映目标土壤类型的成土环境特征。同时，设置同一土种样点之间的最小距离为 500 m，以确保空间分布的均匀性。表 2 展示了训练集和验证集的具体组成情况。验证集包括 60 个野外实测样本和 112 个随机选取的典型虚点。该混合验证方法在保证验证数据可靠性的基础上，通过引入具有代表性的虚点，增强了空间覆盖度和景观代表性，从而全面评估模型的泛化能力。

表 2　各土种训练集和验证集数量

Table 2　Training and validation sample sizes of soil species

土种	训练集	验证集
白刚土	40	7
板浆白土	90	39
潮灰土	40	3
冲淤土	40	3
高砂土	10	2
黄白土	160	18
黄刚土	40	16
黄石土	10	7
黄杂土	160	24
灰马肝土	10	3
灰砂土	40	4
马肝土	160	17
青泥条	10	4
油泥土	40	11
淤泥土	10	14
总计	860	172

1.5　预测模型

采用多类随机森林(MRF)方法[27]进行土壤类型空间分类建模，MRF 基于随机森林(RF)的集成学习思想，将多类型土壤分类任务转变为多个二元分类子任务。分别构建随机森林模型预测各土壤类型的分布概率图，并通过最大概率硬化确定每个像元的最终类型。模型构建基于 R 语言中的 randomForest 包，ntree 设置为 500，mtry 设置为 7。

1.6　建模精度评价

基于验证集与不同虚点拾取方法预测的土壤类型图之间的混淆矩阵，计算总体精度和 Kappa 系数 2 种评价指标，用于比较不同虚点拾取方法的代表性。总体精度是指所有正确分类的验证点数占总验证点数的比例，Kappa 系数用于评估预测分类结果与实际分类之间的一致性[28-29]。

采用混淆系数(CI)来量化预测结果的不确定性，CI 反映的是预测中最可能的类型和次可能类型之间的差异程度[30]，用于衡量分类结果的不确定性。该系数计算公式如下：

$$\mathrm{CI} = u_{\max-1} / u_{\max} \quad (2)$$

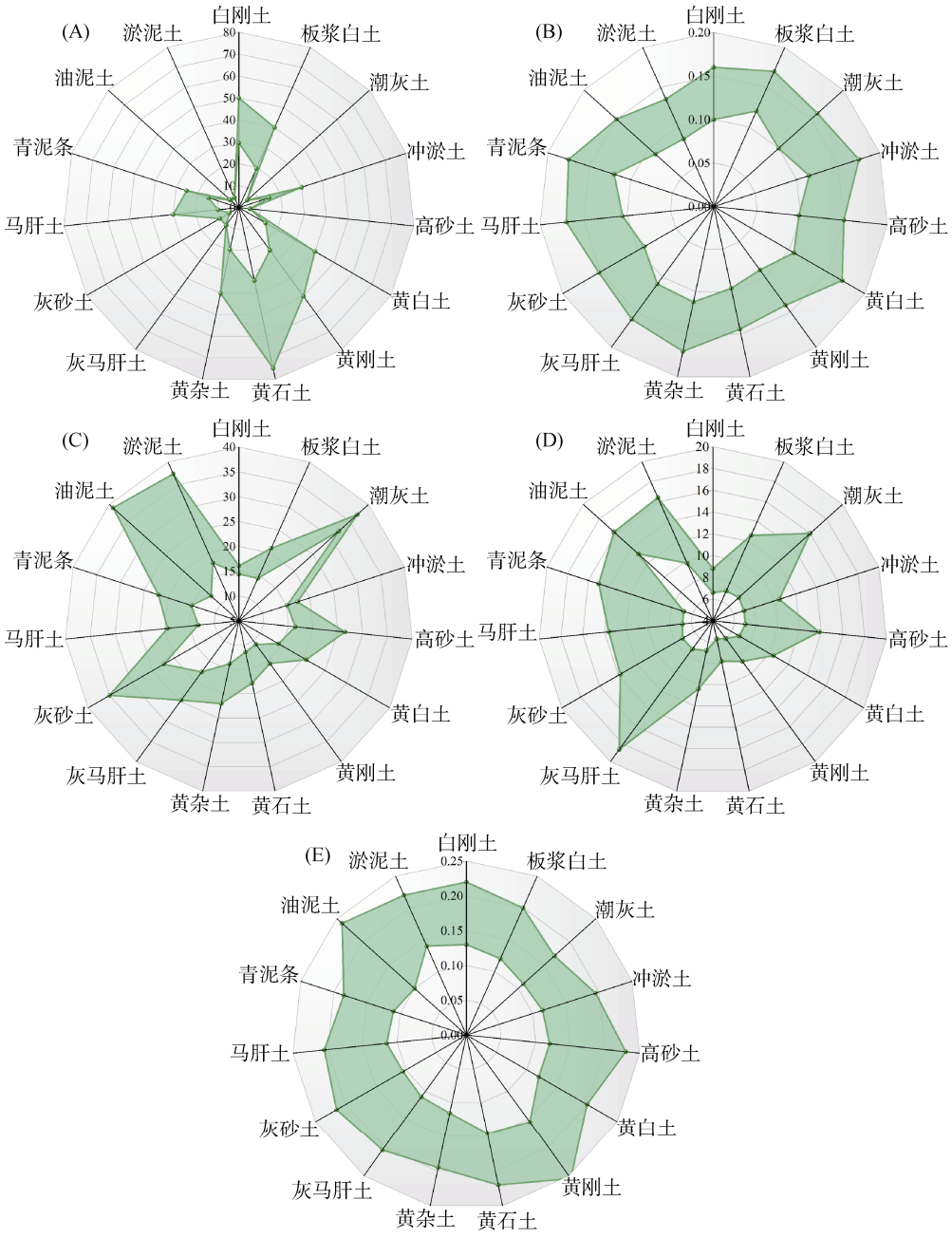
式中：$u_{\max}$ 表示某像元预测得出的最可能类型的概率值；$u_{\max-1}$ 表示次可能类型的概率值。CI 越接近 0，说明预测结果越明确，某一类型占据主导地位；CI 越接近 1，则表明各类型概率分布较为平均，预测结果越不确定。因此，CI 值与预测结果的不确定性成正比。

2　结果与分析

2.1　不同虚点拾取方法的样本数量与分布

如图 6 所示，不同土壤类型在高程、有机质含量、地形湿度指数上表现出明显的分异特征。淤泥土与灰马肝土虽在高程梯度上的分布较为接近，呈现出相似的趋势，然而，二者在有机质含量与地形湿度指数上却存在显著差异。这种差异为区分这两种土壤类型提供了重要依据。灰砂土与油泥土在短波红外和地形湿度指数上具有一定共性，但在有机质含量、归一化植被指数与高程分布上表现出明显的不同。本研究表明，虽然单一环境变量在区分具有相似发生特征的土壤类型时存在局限性，难以完全准确地区分，但通过多个环境变量的耦合分析，能够有效提高土壤类型识别的准确性。

图 7 以白刚土为例，在高程上比较了 3 种虚点拾取方法的统计值，图中灰色区域为典型区间 (29.5~50 m)。结果显示，典型虚点法所选取的样本全部落入典型区间内；图斑中心点法有 26 个样本位于区间内，8 个低于 29.5 m (最低为 13 m)，6 个高于 50 m (最高为 70 m)；随机点法有 24 个样本落入典型区间内，7 个低于 29.5 m (最低为 10 m)，9 个高于 50 m (最高为 70 m)。典型虚点法与典型区间的一致性最高，图斑中心点法次之，随机点法偏离较多，表明基于典型虚点法筛选的样本有助于减少极值干扰且提升样本代表性。



(A. 高程；B.短波红外；C. 有机质含量；D. 地形湿度指数；E. 归一化植被指数)

图 6 不同土种在关键环境变量下的典型区间

Fig. 6 Typical ranges of different soil species under key environmental variables

图 8 展示了基于历史土壤图采用不同虚点拾取方法所获得的样点空间分布。为比较 3 种方法的效果，本研究中 3 种虚点拾取方法的样点数量均为 860 个。结果表明，典型虚点法在研究区西部低丘陵区域的样点分布相对稀疏，主要原因在于该区域土壤类型破碎化程度相对较高，环境因子梯度变化较为剧烈，缺乏连续且具有代表性的成土环境单元，难以满足典型虚点法对代表性样本的要求。相比之下，图斑中心点法和随机点法由于不依赖于环境变量的典型性判别，因此在这些区域仍能保持较为均匀的空间分布，但其样点环境代表性可能不足。

2.2 不同虚点拾取方法制图结果比较

本研究基于所有环境变量，采用 3 种虚点拾取方法获取的训练样本，对研究区进行推测制图。如表 3 所示，典型虚点法在各项指标中均表现出最佳的预测效果，其总体精度最高，为 52%，显著高于图斑中心点法(34%)和随机点法(44%)。在分类一致

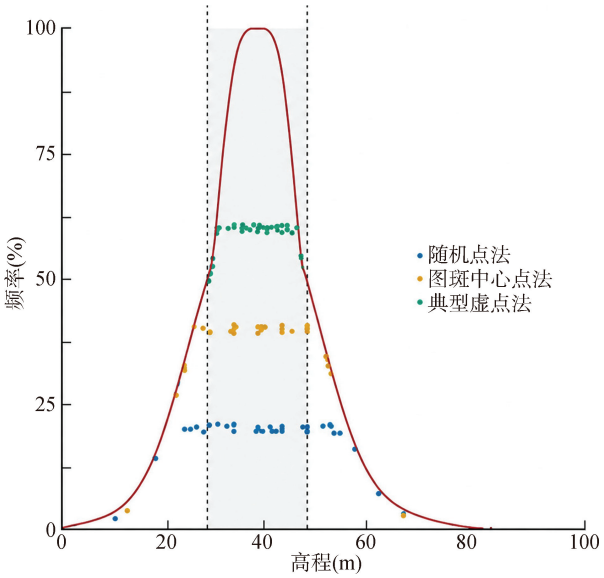
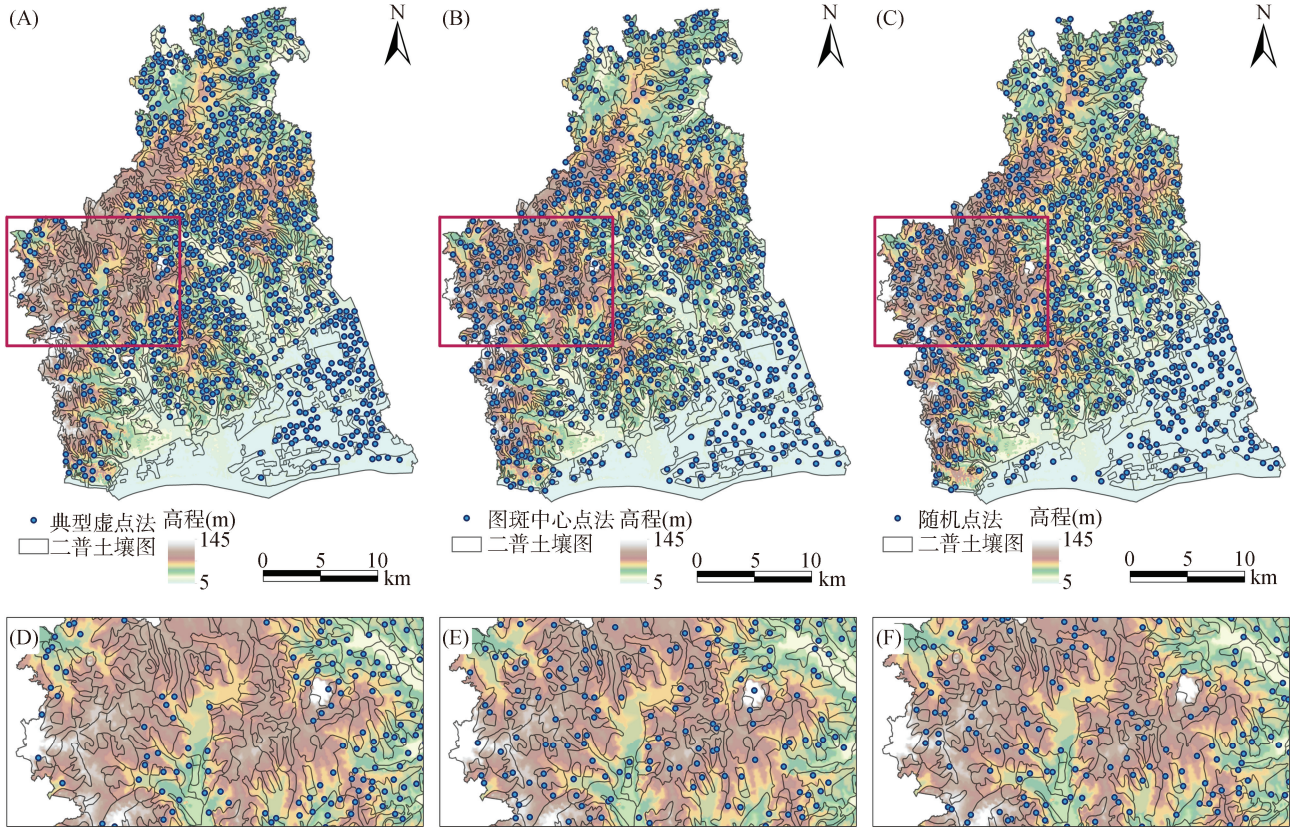


图 7　不同虚点拾取方法样点统计分布

Fig. 7　Statistical distribution of sample points under different pseudo-point selection methods

性方面，典型虚点法的 Kappa 系数较图斑中点法和随机点法分别提高了 0.2 和 0.1。从模型不确定性分析，典型虚点法的 CI 均值最低，为 0.62；而图斑中心点法的 CI 均值最高，达到 0.74；随机点法的 CI 均值居中，为 0.66。3 种方法的 CI 均值反映典型虚点法能够降低模型预测的不确定性，增强预测的可信度。进一步分析发现，尽管随机点法不如典型虚点法表现优异，但其预测精度仍显著高于图斑中心点法。主要原因在于县域尺度下土壤类型图斑边界形态复杂，存在大量狭长、破碎等特殊几何形态的图斑。图斑中心点法由于严格依赖图斑几何中心进行采样，难以有效捕捉复杂图斑的空间特征，从而导致样点的环境代表性降低，易产生分类偏差。而随机点法通过空间均匀采样，能够在一定程度上捕捉图斑内部的异质性，从而获得更多具有代表性的训练样本。

不同虚点拾取方法建模结果的 CI 分布特征如图 9 所示。结果表明，基于典型虚点法推测制图的结果表现出最低的预测不确定性。图斑中心点法仅在研究区东南部呈较低不确定性，随机点法则在研究区南部取得较好结果，而二者在其他区域的 CI 值普遍高于 0.75。值得注意的是，在 CI 值为 0.75~1 的高不确定性等级中，典型虚点



(A. 典型虚点法；B. 图斑中心点法；C. 随机点法；D. 典型虚点法局部细节；E. 图斑中心点法局部细节；F. 随机点法局部细节)

图 8　不同虚点拾取方法的样点分布

Fig. 8　Distribution of sample points under different pseudo-point selection methods

表 3 不同虚点拾取方法的制图精度

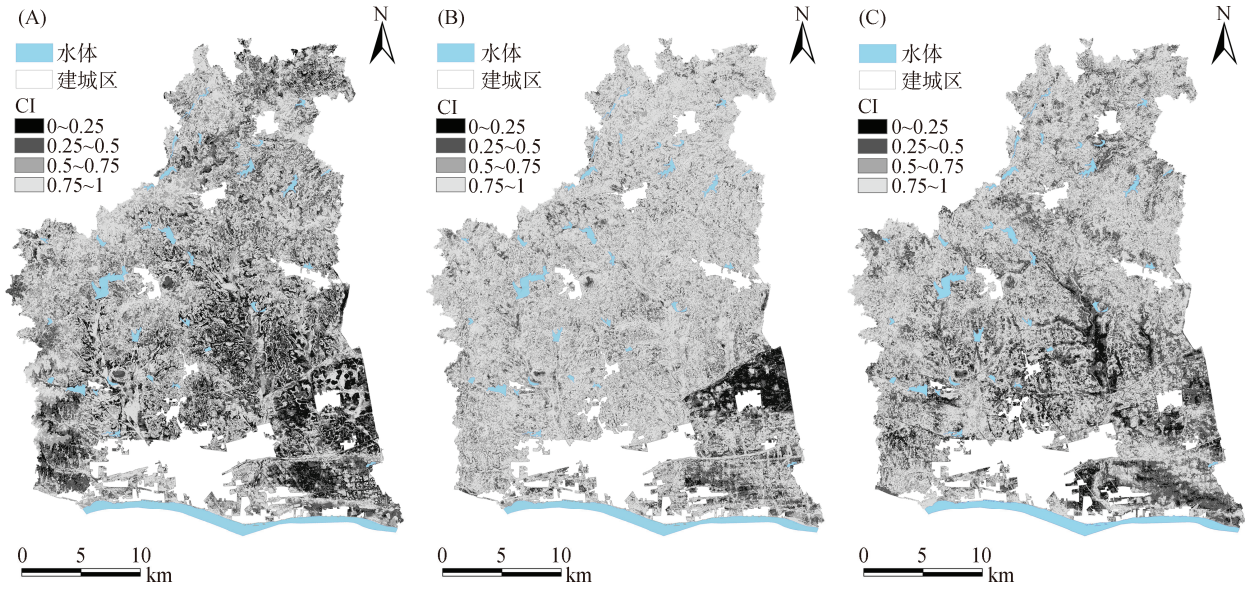
Table 3 Mapping accuracies of different pseudo-point selection methods

方法	总体精度(%)	Kappa 系数	CI 均值
典型虚点法	52	0.47	0.62
图斑中心点法	34	0.27	0.74
随机点法	44	0.37	0.66

法面积占比为 34%，显著低于图斑中心点法(55%)和随机点法(40%)，这一结果充分表明了典型虚点法具有更优的空间预测稳定性。

2.3 更新后土壤类型图分布格局对比

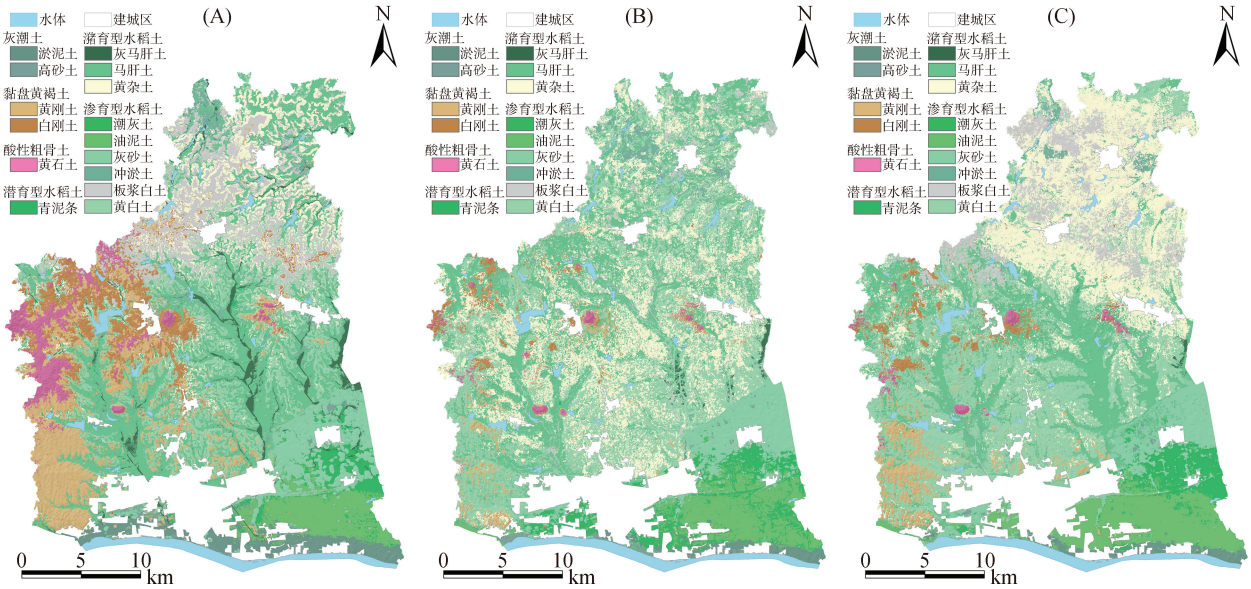
图 10 为不同虚点拾取方法更新后的土壤类型空间分布图。从宏观空间格局分析，3 种预测结果均能反映区域土壤类型的基本空间分布规律：潮土集中分布于长江冲积平原，黄褐土主要出现在低丘陵和高岗地区，粗骨土多分布于西部丘陵高岗地和中部突起的低山坡地，水稻土则广泛分布于整个县域。在微观尺度的空间表达方面，不同方法存在显著差异。图斑中心点法因严格遵循图斑几何中心采样，预测结果与历史土壤图保持高度空间一致性，



(A. 典型虚点法；B. 图斑中心点法；C. 随机点法)

图 9 不同虚点拾取方法建模结果的不确定性分布

Fig. 9 Uncertainty distribution of modeling results based on different pseudo-point selection methods



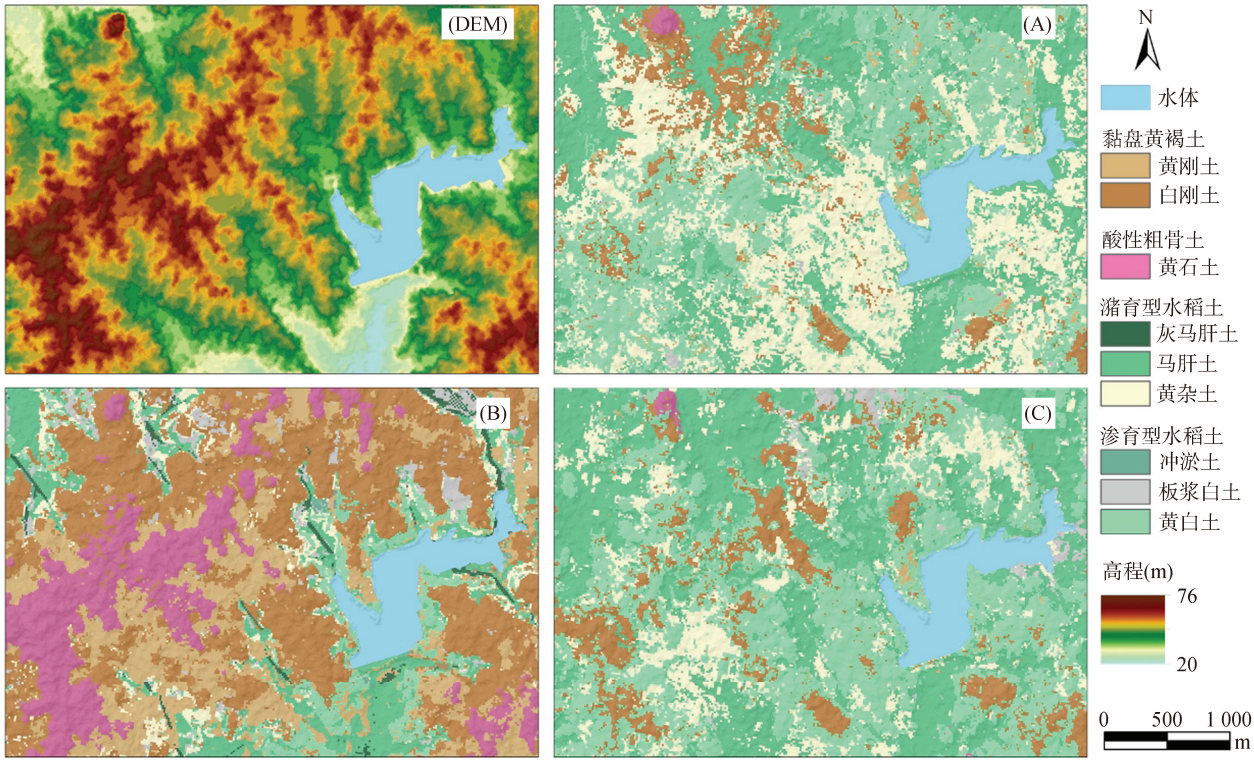
(A. 典型虚点法；B. 图斑中心点法；C. 随机点法)

图 10 不同虚点拾取方法数字制图结果

Fig. 10 Digital mapping results based on different pseudo-point selection methods

但受限于固定采样点位，缺乏对环境异质性的响应。随机点法虽然通过均匀采样提高了空间覆盖度，但由于采样位置的随机性，导致预测结果的空间随机性较高。

对研究区西北部低丘陵地带不同虚点拾取方法制图结果的细节对比分析(图 11)表明，该区域土壤空间分布格局受地形梯度影响显著。不同虚点拾取方法的预测模型变量重要性排序进一步证明了这一点：高程、地形湿度指数、坡度等地形变量对黄石土、黄刚土等的影响较大。其中，典型虚点法的预测结果与土壤–景观关系最为一致，能够清晰呈现出自坡顶至坡脚的土壤类型地形顺序，即黄石土位于坡顶，黄刚土处于上坡，经坡中的白刚土过渡到坡下的水稻土。这一空间分布特征与区域成土环境的变化规律高度一致。相比之下，图斑中心点法和随机点法的预测结果在该区域出现了大量水稻土分布，而研究区内水稻土主要分布于地势相对平坦的地区，这一预测结果明显偏离了实际土壤分布规律。典型虚点法凭借其对环境变量代表性的有效提取，提升了样点的质量与代表性，从而增强了制图结果对土壤空间变异特征与成土过程的表达能力。



(A. 图斑中心点法；B. 典型虚点法；C. 随机点法)

图 11　不同虚点拾取方法数字制图细节

Fig. 11　Details of digital mapping based on different pseudo-point selection methods

3　讨论

样本的代表性在数字土壤制图中具有关键作用[31-32]，具有代表性的样本在模型应用、制图精度等方面展现出显著优势[33]。本研究对比了典型虚点法、图斑中心点法、随机点法 3 种虚点拾取方法，发现 3 种方法的本质差别在于对土壤–景观关系的利用程度。从样本的统计特征看，典型虚点法依据关键环境变量的频数分布识别各个土壤类型的典型区间，并据此在代表性位置选取样本，降低图斑内部非优势土壤类型的干扰，使样本在典型区间内取得更高覆盖率；相比之下，图斑中心点法和随机点法所选样本在环境因子的取值上较为分散，容易落入土壤类型过渡地带或环境特征辨识度较弱的范围，难以精确刻画特定土壤类型的环境特征[34]。典型虚点法通过减少非典型区样本选取，增强同一土壤类型样本在环境变量上的一致性，并降低不同土壤类型在环境因子上的取值交叠，从而提高样本的代表性。在模型预测中，该方法表现出更稳定的判别边界与更强的泛化能力。因此，在土壤图精度验证中，典型虚点法获得更高的总体精度、Kappa 系数，并表现出更低的 CI 值。在土壤地带性分布规律上，典型虚点法在研究区西部区域清晰地再现了坡顶黄石土、上坡黄刚土、坡中白刚土、坡脚水稻土的土壤类型地形序列，而另外两种方法在

该区域产生了对水稻土的误判。这表明样本代表性对空间格局预测至关重要。综上，将关键环境因子的典型区间作为样本选取的约束，能够有效提升样本代表性，并将其转化为模型精度的实质改进与地带性格局的正确预测，从而提升空间分布的一致性。

本研究结果与张仁铎[35]认为的随机点法总体表现较好的观点存在差异，但与初玉婷等[36]、Cheng 等[37]关于随机点法存在样本代表性不足的观点一致。造成差异的关键在于先验知识的可获得性，当缺乏先验知识或难以掌握预测对象的分布规律时，随机点法通过大量样本数据训练模型，确实具有一定的可行性和效果。然而，当可借助土壤–景观模型或依托专家知识获得相应的先验知识时，随机点法的适用性随之下降，其对模型精度和空间匹配性的贡献也随之减弱[38-39]。因此，在具备必要的先验知识与环境变量的支撑下，以典型性为核心的样本选择是提升数字土壤制图精度与可靠性的关键方法。

本研究仍存在一定的局限性。首先，尽管典型虚点法在具体点位强调代表性，但是“点–点”之间环境变量在特征空间中的梯度变化以及“点–线–面”构成的采样网络所具有的全局空间覆盖能力，在本研究中尚未被充分考虑。未来研究可通过引入空间梯度分析与采样网络优化方法，提高样点空间代表性，从而为高精度的土壤类型图提供更有利的支撑。其次，该典型虚点法可能不适用于面积较大的人为改土区域。该方法的前提是土壤类型图上表征的土壤类型稳定无变化。因此，如果某些土壤图斑受到广泛的人为改土影响，仅通过典型虚点法拾取虚点来更新土壤类型图可能无法达到预期效果。最后，本研究的目标在于通过优化虚点选取来提高其代表性，并未对多种预测模型进行对比，也未系统分析样本数量对结果的影响。为进一步验证结论的稳健性，后续将尝试采用不同的预测模型，全面评估模型选择可能带来的影响；同时，探讨样本数量对预测精度的作用，以优化样本量并提升制图精度。

4 结论

1) 典型虚点法在样点选取、模型精度和预测稳定性方面表现出显著优势，其总体精度和 Kappa 系数均高于图斑中心点法和随机点法，CI 均值则低于其他两种方法，表现出更好的预测稳定性。

2) 在区域尺度上，西部丘陵岗地以地形因子为主导环境变量。典型虚点法的预测结果与该区域土壤–景观关系的契合度最高，而图斑中心点法和随机点法对该关系的刻画相对不足。

3) 虽然典型虚点法在土地利用类型发生大规模变化的区域存在一定局限性，但其通过整合历史土壤图与环境变量的关系，提高了样点的环境代表性，为历史土壤图数字化更新提供了可靠的虚点拾取方法，同时也为土壤资源信息化管理提供了理论参考和技术支持。

参考文献：

[1] 周明枞, 杜国华, 王浩清. 调整全国土壤普查工作的建议[J]. 土壤, 1981, 13(2): 41–46.

[2] 朱阿兴, 杨琳, 樊乃卿, 等. 数字土壤制图研究综述与展望[J]. 地理科学进展, 2018, 37(1): 66–78.

[3] 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.

[4] Huggett R J. Soil landscape systems: A model of soil Genesis[J]. Geoderma, 1975, 13(1): 1–22.

[5] Pham T H, Osterloh K, Nguyen K D. Mapping of soil sampling sites using terrain and hydrological attributes[J]. Artificial Intelligence in Agriculture, 2025, 15(3): 470–481.

[6] 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[J]. 土壤学报, 2007, 44(5): 784–791.

[7] Qi F, Zhu A X. Knowledge discovery from soil maps using inductive learning[J]. International Journal of Geographical Information Science, 2003, 17(8): 771–795.

[8] 朱阿兴. 精细数字土壤普查模型与方法[M]. 北京: 科学出版社, 2008.

[9] 刘雪琦, 朱阿兴, 杨琳, 等. 土壤图更新中基于土壤类型面积分级的训练样点选择方法[J]. 土壤学报, 2017, 54(1): 36–47.

[10] Liu X Q, Zhu A X, Yang L, et al. A graded proportion method of training sample selection for updating conventional soil maps[J]. Geoderma, 2020, 357: 113939.

[11] Heung B, Ho H C, Zhang J, et al. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping[J]. Geoderma, 2016, 265: 62–77.

[12] Heung B, Bulmer C E, Schmidt M G, et al. Provincial-scale digital soil mapping using a random forest approach for British Columbia[J]. Canadian Journal of Soil Science, 2022, 102(3): 597–620.

[13] 杨雨菲, 韩浩武, 陈荣, 等. 数字土壤制图的推理方法对比研究[J]. 土壤通报, 2020, 51(5): 1016–1023.

[14] 周紫燕, 黄魏, 许伟, 等. 基于随机森林算法的原始土壤图更新研究[J]. 华中农业大学学报, 2019, 38(3): 53–59.

[15] 孟可, 黄魏, 傅佩红, 等. 平原–丘陵区域数字土壤制图方法比较[J]. 土壤学报, 2025, 62(4): 958–969.

[16] 高鸿, 朱娟, 王良杰, 等. 基于成土环境地理邻域分析的历史土壤图训练样本筛选[J]. 土壤学报, 2018, 55(3): 585–594.

[17] Gao H, Zhang X Y, Wang L J, et al. Selection of training samples for updating conventional soil map based on

spatial neighborhood analysis of environmental covariates[J]. Geoderma, 2020, 366: 114244.
[18] 刘京, 朱阿兴, 张淑杰, 等. 基于样点个体代表性的大尺度土壤属性制图方法[J]. 土壤学报, 2013, 50(1): 12–20.
[19] 王子龙, 陈伟杰, 付强, 等. 基于优先级指数的土壤采样设计方法研究[J]. 农业机械学报, 2018, 49(7): 244–251.
[20] 张甘霖, 史舟, 朱阿兴, 等. 土壤时空变化研究的进展与未来[J]. 土壤学报, 2020, 57(5): 1060–1070.
[21] McBratney A B, Mendonça Santos M L, Minasny B. On digital soil mapping[J]. Geoderma, 2003, 117(1/2): 3–52.
[22] Lamichhane S, Kumar L, Adhikari K. Updating the national soil map of Nepal through digital soil mapping[J]. Geoderma, 2021, 394: 115041.
[23] Herrmann D L, Schifman L A, Shuster W D. Widespread loss of intermediate soil horizons in urban landscapes[J]. Proceedings of the National Academy of Sciences of the United States of America, 2018, 115(26): 6751–6755.
[24] Liu F, Geng X Y, Zhu A X, et al. Soil polygon disaggregation through similarity-based prediction with legacy pedons[J]. Journal of Arid Land, 2016, 8(5): 760–772.
[25] 孟祥添, 鲍依临, 刘焕军, 等. 基于高分 5 号影像的东北典型黑土区土壤分类[J]. 农业工程学报, 2020, 36(16): 231–238.
[26] 孙越琦, 孙笑梅, 巫振富, 等. 样点规模与采样方法对表层土壤 pH 空间预测精度的影响[J]. 土壤学报, 2023, 60(6): 1595–1609.
[27] Bao Y L, Yao F M, Meng X T, et al. A fine digital soil mapping by integrating remote sensing-based process model and deep learning method in Northeast China[J]. Soil and Tillage Research, 2024, 238: 106010.
[28] Zhan Q Q, Zhao W, Yang M J, et al. A long-term record (1995–2019) of the dynamics of land desertification in the middle reaches of Yarlung Zangbo River basin derived from Landsat data[J]. Geography and Sustainability, 2021, 2(1): 12–21.
[29] Wang L P, Wang X, Kooch Y, et al. Improvement of data imbalance for digital soil class mapping in Eastern China[J]. Computers and Electronics in Agriculture, 2023, 214: 108322.
[30] Burrough P A, van Gaans P F M, Hootsmans R. Continuous classification in soil survey: Spatial correlation, confusion and boundaries[J]. Geoderma, 1997, 77(2/3/4): 115–135.
[31] 李坤, 陈宇昊, 李文岳, 等. 基于核密度估计的土壤样本代表性修正研究[J]. 华中农业大学学报, 2025, 44(1): 94–104.
[32] 张自刚, 王迪, 李炜琪, 等. 区域土壤调查采样点布设方法研究进展[J]. 中国农业信息, 2024, 36(6): 81–96.
[33] An Y M, Yang L, Zhu A X, et al. Identification of representative samples from existing samples for digital soil mapping[J]. Geoderma, 2018, 311: 109–119.
[34] Wan C J, Kuzyakov Y, Cheng C X, et al. A soil sampling design for arable land quality observation by using SPCOSA–CLHS hybrid approach[J]. Land Degradation & Development, 2021, 32(17): 4889–4906.
[35] 张仁铎. 空间变异理论及应用[M]. 北京: 科学出版社, 2005: 27–35.
[36] 初玉婷, 李晓岚, 廉海荣, 等. 基于特征代表性的土壤环境质量监测点布局优化方法[J]. 农业环境科学学报, 2023, 42(11): 2430–2439.
[37] Cheng W, Zhu A X, Qin C Z, et al. Updating conventional soil maps by mining soil–environment relationships from individual soil polygons[J]. Journal of Integrative Agriculture, 2019, 18(2): 265–278.
[38] 王劲峰, 姜成晟, 李连发, 等. 空间抽样与统计推断[M]. 北京: 科学出版社, 2009.
[39] 汪训平, 赵安. 一种基于代表性样点的钉螺密度抽样调查设计方法[J]. 长江流域资源与环境, 2018, 27(2): 395–404.

(责任编辑：毛小芳)