DOI: 10.11766/trxb202406210251 **CSTR: 32215.14.trxb202406210251**

孟可，黄魏，傅佩红，李文岳，冯玲.平原-丘陵区域数字土壤制图方法比较[J].土壤学报，2025，

MENG Ke, HUANG Wei, FU Peihong, LI Wenyue, FENG Ling. Comparison of Digital Soil Mapping Methods in Plain and Hill Mixed Regions [J]. Acta Pedologica Sinica, 2025,

平原-丘陵区域数字土壤制图方法比较*

孟 可，黄 魏† ，傅佩红，李文岳，冯 玲

（华中农业大学资源与环境学院，武汉 430070）

摘 要：构建适宜性土壤-景观关系模型是提高数字土壤制图精度的关键。由于平原-丘陵区域的多尺度复杂地形，如何充分考虑土壤-景观关系模型建立的主要环节来准确预测其土壤类型空间分布需要进一步探讨。本研究以湖北省麻城市乘马岗镇北部为研究区，将其划分为平原和丘陵 2 个地形单元，以 28 个环境变量为辅助因子，评估决策树（Decision Tree, DT）、随机森林（Random Forest, RF）、梯度提升树（Gradient Boosting Decision Tree, GBDT）和极端梯度提升（Extreme Gradient Boosting, XGBoost）在各地形下进行推理制图的精度，基于因子重要性排序筛选变量，通过对比整体与按地形分区制图的精度，探索提高平原-丘陵区域土壤类型制图精度的途径。结果表明：不同地形条件下，最优推理制图方法不同。RF 在整体和平原区域制图效果较好，XGBoost 在丘陵区域制图效果较好。通过变量筛选能够有效提升推理制图总体精度和 Kappa 系数，整体区域提升效果最好（分别提升了 4.96%和 0.06），平原区域提升效果最差（分别提升了 1.43%和 0.02）。相比较于整体制图，按地形分区制图精度最高，总体精度和 Kappa 系数分别为 73.05%和 0.69。在平原-丘陵混合区域，综合考虑制图方法优选、环境变量筛选以及制图方式能有效提升土壤类型推理制图精度，为复杂地形区域土壤类型推理制图提供了实践案例和技术支持。

关键词：土壤类型；土壤-景观关系模型；机器学习算法；环境变量；地形

中图分类号:S159 文献标志码:A

Comparison of Digital Soil Mapping Methods in Plain and Hill Mixed Regions

MENG Ke, HUANG Wei †, FU Peihong, LI Wenyue, FENG Ling

(*College of Resource and Environment, Huazhong Agricultural University, Wuhan* 430070*, China*)

Abstract: 【Objective】Digital soil mapping is a burgeoning and efficient method to express the spatial distribution of soil. Based on a data mining algorithm, this method establishes a soil-landscape relationship model to infer soil mapping by using raster data as an expression and computer-assisted. The key to improving the accuracy of digital soil mapping is constructing a suitable soil-landscape relationship model. However, the commonly used methods of digital soil mapping cannot meet the application requirements of soil mapping given the complicated nature of terrains consisting of plains and hills. How to fully consider the main links of the soil-landscape relationship model to accurately infer the spatial distribution of soil types needs further discussion. 【Method】The northern part of Chengmagang town, Macheng City, Hubei province was selected as the study area. It was divided into two terrain units, plains and hills. Based on the 28 environmental variables, Decision Tree (DT),

* 国家自然科学基金项目（41877001，42171056）资助 Supported by the National Natural Science Foundation of China (Nos. 41877001,42171056)
† 通讯作者 Corresponding author，E-mail：ccan@mail.hzau.edu.cn
作者简介：孟 可(1997-)，女，安徽宿州人，博士研究生，主要从事精细数字土壤制图研究。E-mail: Kelly1997@webmail.hzau.edu.cn
收稿日期：2024-06-21；收到修改稿日期：2024-10-16；网络首发日期（www.cnki.net）：

Random Forest (RF), Gradient Boosting Decision Tree (GBDT), and Extreme Gradient Boosting (XGBoost) were used to select optimal mapping methods for each region. Then, the optimal variables combination was selected according to the factor importance ranking of each region. Moreover, the optimal mapping methods were used to establish a soil-landscape relationship model linking soil types to the optimal variable combinations, upon which soil type mapping was inferred for each region. Soil-type mapping results for plain and hilly areas were combined as the soil-type mapping result of the terrain region. Finally, the mapping accuracy of the whole region was compared with the terrain region to further explore ways to improve the accuracy of soil-type mapping in Plain and Hill Mixed Regions. 【Result】Under different terrain conditions, the performance of each inference mapping method was different as well as the optimal inference mapping method. The performance of RF and XGBoost was superior to other algorithms. Specifically, the RF performed better in whole and plain regions while the XGBoost was the best algorithm in the hill region. The model accuracy was further effectively improved through variable screening, with the maximum increase of overall accuracy and Kappa coefficient being 4.96% and 0.059 in the whole region, respectively. However, the model accuracy improvement was not obvious in the plain region, with the increase of overall accuracy and Kappa coefficient being 1.43% and 0.018, respectively. Also, the increase in overall accuracy and Kappa coefficient was 2.82% and 0.03 in the hill region. Compared with the whole mapping method, the inference mapping method based on terrain zoning had the highest accuracy, and the overall accuracy and Kappa coefficient were 73.05% and 0.69, respectively. Meanwhile, the plain region required more remote sensing factors to participate in inference mapping compared to the whole and hill regions. 【Conclusion】The inference mapping accuracy in plain and hill regions can be effectively improved by optimizing the mapping method, selecting environment variables, and adopting appropriate mapping way. This study can provide some references for the screening of environmental variables, the selection of mapping algorithms, and the construction of mapping ways of inference mapping in plain and hill regions. It provides promising and practical examples and technical support effective for promoting the improvement of the accuracy of inference mapping in complex terrain areas.

Key words: Soil type; Soil-landscape relationship model; Machine learning algorithm; Environment variables; Terrain

土壤是地球生命和农业生产的重要基础资源，了解其类型及空间分布对于土壤资源的可持续利用和管理至关重要[1]。随着计算机科学，遥感、数据挖掘算法的快速发展，数字土壤制图（Digital Soil Mapping，DSM）成为一种新兴的、高效表达土壤空间分布的技术方法。该方法运用数据挖掘算法建立土壤-景观关系模型，以栅格数据为表达方式，在计算机环境下辅助制图[2-3]。

常用的数字土壤制图方法主要分为地统计法和机器学习算法，线性模型[4]、克里格法[5]和判别分析[6]为常见的地统计法，由于土壤与环境之间的关系过于复杂，为了更好地挖掘土壤-景观关系，机器学习算法被引入到构建土壤类型或土壤属性制图模型中，如决策树（Decision Tree，DT）[7]、随机森林（Random Forest，RF）[8]、梯度提升树（Gradient Boosting Decision Tree，GBDT）[9]和极端梯度提升（Extreme Gradient Boosting，XGBoost）[10]等。已有较多研究对比了多种机器学习算法在土壤属性制图方面的效果[11-12]，但在土壤类型制图中尚不多见。地形因素是土壤形成与发育的主要影响因素之一，在不同的地形条件下，土壤的形成过程和发育阶段会有所不同，从而导致土壤类型和性质差异[13]。特别是在平原-丘陵混合区域，地形的多尺度变化使得土壤类型的空间异质性难以精确捕捉，进而土壤类型推理制图精度不高。此外，土壤类型制图精度还受到多种环境因素的综合影响，这增加了平原-丘陵区域推理制图的复杂性。因此，综合考虑地形条件和土壤-景观关系模型的建立过程，比较不同数字土壤制图模型在平原-丘陵区域的土壤类型推理制图效果，并选出最优的数字制图模型，显得尤为重要。

本研究以湖北省麻城市乘马岗镇北部平原-丘陵区域为研究区，将其划分为平原和丘陵两个地形单元，通过多种机器学习算法优选各地形的制图方法，依据 RF 模型内嵌的因子重要性排序，筛选出最优环境变量集，分别构建各地形最优土壤-景观关系模型进行土壤类型推理制图。接着，对比整体

与按地形分区制图的精度，探究制图方法优选、环境变量筛选以及制图方式对平原-丘陵区域土壤类型推理制图精度的影响，以期为复杂地形区域土壤类型推理制图提供实践案例和技术支持。

1 材料与方法

1.1 研究区概况

研究区位于湖北省麻城市乘马岗镇北部（图 1），中心地理坐标为 31°25′30″N，114°58′30″E，总面积约为 28.14 km²，山地、平畈交错，属于典型的平原-丘陵混合区域，高程范围为 80~346 m，为探究制图方法和参与制图的环境变量在整体区域、平原区域和丘陵区域的差异，依据土壤志及地形起伏度分布图，以 150 m 高程为分界将整体研究区分为平原区域和丘陵区域。

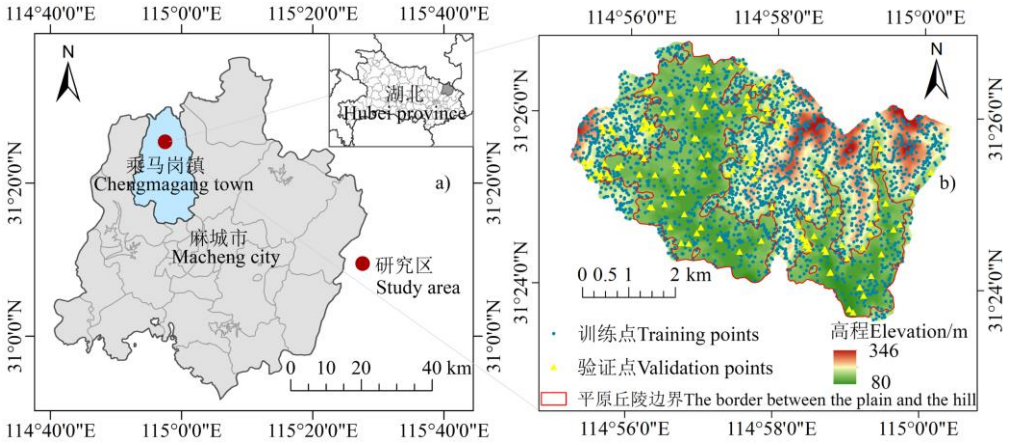

图 1 研究区概况

Fig. 1 Overview of the study area

根据第二次全国土壤普查成果的 1：50 000 传统土壤类型图可知，研究区内共有 9 种土壤类型和 3 种母质类型，其关系如表 1 所示。

表 1 研究区土壤类型、母质类型、训练点和验证点信息

Table 1 Information of soil type, parent material type, training point and validation point in the study area

序号 Number	土壤类型 Soil type	母质类型 Parent material type	土壤类型代号 Soil type code	训练点 Training points	验证点 Validation points
1	沙泥土 Silt soil	花岗岩-片麻岩的坡积或残积物 Slope deposit or remnant of granite gneiss	1	566	16
2	硅沙泥土 Silica silt soil	白云石英片岩-花岗片麻岩坡积或残积物 Dolomite quartz schist granite-gneiss slope deposit or remnant	4	607	25
3	林地泥沙土 Forest silt soil	花岗岩-片麻岩的坡积或残积物 Slope deposit or remnant of granite gneiss	1-7	97	9
4	沙泥田 Silt field	花岗岩-片麻岩的坡积或残积物 Slope deposit or remnant of granite gneiss	22	192	14
5	硅沙泥田 Silica silt field	白云石英片岩-花岗片麻岩坡积或残积物 Dolomite quartz schist granite-gneiss slope deposit or remnant	24	248	19

6	潮土田 Tidal soil field	近代河流冲积物 Modem river alluvium	30	48	13
7	硅沙土 Silica sand soil	白云石英片岩-花岗片麻岩坡积或残积物 Dolomite quartz schist granite-gneiss slope deposit or remnant	4-4	152	12
8	硅麻骨土 Silica hemp bone soil	白云石英片岩-花岗片麻岩坡积或残积物 Dolomite quartz schist granite-gneiss slope deposit or remnant	4-5	116	11
9	林地薄层硅麻骨土 A thin layer of silica hemp bone soil in woodland	白云石英片岩-花岗片麻岩坡积或残积物 Dolomite quartz schist granite-gneiss slope deposit or remnant	4-6	434	22

1.2 数据来源及处理

数字土壤制图的关键在于构建土壤-景观关系模型，该模型不仅受制图算法的影响，还需要选取适宜的环境变量以保证其稳健性。基于 SCORPAN 土壤景观模型[14]，本文选取了 28 个环境因子来获取土壤-景观知识，其中包括 1 个母质信息、12 个地形因子和 15 个遥感因子[15-16]（表 2）。母质信息来源于成土母质图及当地土壤专家所给意见；地形因子中的高程数据由湖北省测绘局提供的 10 m 等间距等高线数据插值所得，地形衍生因子利用 ArcGIS 10.5 和 SAGA GIS 9.4.1 软件计算获得；遥感因子中的植被指数通过谷歌地球引擎（Google Earth Engine，GEE）平台中的 Sentinel-2 影像（成像时间 2020 年 2 月 9 日）计算获得，同时从 GEE 获取同期的 Sentinel-2 多波段数据，包括 B2、B3、B4、B5、B6、B7、B8、B8a、B11 和 B12。为了减少数据冗余，采用 ENVI 5.3 对其进行主成分分析得到第一主成分，进而提取纹理特征。为解决各环境变量的坐标系和尺度不一致问题，所有环境变量均被标准化为 10 m 分辨率，并统一投影坐标系为 WGS_1984_UTM_Zone_50N。

表 2 环境变量基本信息

Table 2 The basic information about environment variables

	环境变量 Environment variables	空间分辨率 Resolution/m	简称 Abbreviation
母质信息 Parent material	母质 Parent material	-	PM
地形因子 Topographic factor	高程 Elevation	10	EL
	坡向 Aspect	10	Aspect
	坡度 Slope	10	Slope
	谷深 Valley depth	10	VD
	相对坡度位置 Relative slope position	10	RSP
	坡度坡长因子 Slope length and steepness factor	10	LS
	水平曲率 Horizontal curvature	10	HOC
	平面曲率 Plan curvature	10	PLC
	剖面曲率 Profile curvature	10	PRC
	地形湿度指数 Topographic wetness index	10	TWI
	地形起伏度 Topographic relief	10	TRF

	地形粗糙度 Topographic roughness	10	TR
遥感因子	第一主成分 First principal component	10	FPC
Remote sensing	均值 Mean	10	Mean
factor	方差 Variance	10	VA
	协同性 Homogeneity	10	HO
	对比度 Contrast	10	CO
	相异性 Dissimilarity	10	DI
	信息熵 Entropy	10	EN
	二阶矩 Second moment	10	SM
	相关性 Correlation	10	CR
	归一化植被指数 Normalized difference vegetation index	10	NDVI
	增强植被指数 Enhanced vegetation index	10	EVI
	差值植被指数 Differential vegetation index	10	DVI
	比值植被指数 Ratio vegetation index	10	RVI
	绿光归一化植被指数 Green normalized difference vegetation index	10	GNDVI
	土壤调整植被指数 Soil adjusted vegetation index	10	SAVI

1.3 样本点获取及处理

本研究中的样本点包括 2 460 个训练点和 141 个验证点（图 1b），其中训练点是依据传统土壤图，使用 ArcGIS 10.5 软件中的“创建随机点”工具生成，设置每种土壤类型不少于 300 个，样点最小允许距离为 50 m。为提高模型的准确性和泛化能力，本研究使用所有训练点，并充分利用数据资源，以更准确地比较不同制图方法的效果。验证点采用完全随机采样、主观采样、混淆区域边界采样和均匀（网格）采样相结合的方式进行实地采集，采样时间为 2020 年 10 月—2021 年 6 月，覆盖整个研究区所有土壤类型且空间分布均匀。研究区各土壤类型训练点及验证点数量如表 1 所示。

1.4 环境变量优选

现有研究表明，过多的环境因子不仅会导致数据冗余，而且会增加计算复杂度，降低模型精度[17]。为了降低模型的计算成本，增加模型鲁棒性，本研究采用逐步法来进行环境变量筛选。首先，基于 RF 内嵌模型计算所有环境变量的因子重要性，通过逐一添加正向排序后的环境变量进行土壤-景观关系模型的构建，并计算每次预测制图的总体精度（Overall Accuracy，OA）。最终，选择 OA 最大的环境变量组合作为最优环境变量集。

1.5 不同制图方法

本文采用决策树、随机森林、梯度提升树和极端梯度提升 4 种机器学习算法进行土壤-景观关系模型的构建。

决策树是一种由树模型构成的机器学习算法，易于理解和解释。它根据特征将信息划分成两个或多个子节点，逐步构成决策树，最终选择的结果由树最远的节点所决定[18]。决策树建立流程一般包括 3 个步骤：特征选择、决策树生成以及剪枝。信息增益和基尼系数是两种特征选择方法，本文采用基尼系数法。

随机森林是一种以决策树为基本分类器的机器学习算法，能够处理高维数据并自动估算特征重要性。该算法通过在特征集的不同随机子集中构建多个决策树，在每个决策树的每个节点上利用一

定的特征选取方法选出最优特征进行划分，取多个预测结果的平均值作为最终预测结果[19]。

梯度提升树是一种基于决策树的迭代集成学习算法，适用于处理复杂数据关系。该算法利用梯度下降法进行迭代，即将前一个树模型拟合的残差作为下一棵树的输入，以实现损失函数最小化，从而提升模型预测性能，最终将多个决策树的预测结果相加作为最终预测[20]。

极端梯度提升是在梯度提升树基础上改进的一种机器学习算法[21]。该算法与梯度提升树类似，但在损失函数后加入了正则化项，既能降低模型复杂度，提升精度，又具备准确性高、训练速度快、防止过拟合和适合大规模数据等特点[22]。

本文所有模型均在 Python 3.9 环境下实现，并在使用贝叶斯优化算法进行各模型参数优化时，设定随机搜索步数（init_points）为 15，贝叶斯优化步数（n_iter）为 500。

1.6 制图精度评价

本研究通过验证点与推理土壤图之间的混淆矩阵，计算总体精度、Kappa 系数、生产者精度（Producers' accuracy，PA）和用户精度（Users' accuracy，UA）4 种评价指标，用于评估不同制图方法的推理精度[23]。Kappa 系数用于衡量分类一致性，取值范围为 0~1，其中 0~0.20 为微弱一致性，0.21~0.40 为弱一致性，0.41~0.60 为中等一致性，0.61~0.8 为显著一致性，0.81~1 为几乎完全一致性。

2 结 果

2.1 不同制图方法比较

基于所有环境变量，本文分别采用 DT、RF、GBDT 和 XGBoost 对整体、平原和丘陵区域进行推理制图。如图 2 所示，4 种算法在三种区域均可以实现推理制图，但不同算法在不同区域的制图精度有所不同。除 DT 算法外，其他三种算法在丘陵区域的制图精度均高于平原区域和整体区域。整体区域中，制图精度由高到低分别为 RF、XGBoost、GBDT 和 DT，其中 RF 算法的总体精度为 68.09%，Kappa 系数为 0.63；平原区域中，RF 算法也取得了最高的制图精度，其总体精度和 Kappa 系数分别为 68.57%和 0.62，紧接其后，制图精度由高到低分别为 DT、XGBoost 和 GBDT；丘陵区域中，RF、GBDT 和 XGBoost 算法的制图精度均高于 70%，其中 XGBoost 算法制图精度最高，其总体精度为 73.24%，Kappa 系数为 0.68。最终，分别选择各区域最优制图方法用于后续研究。

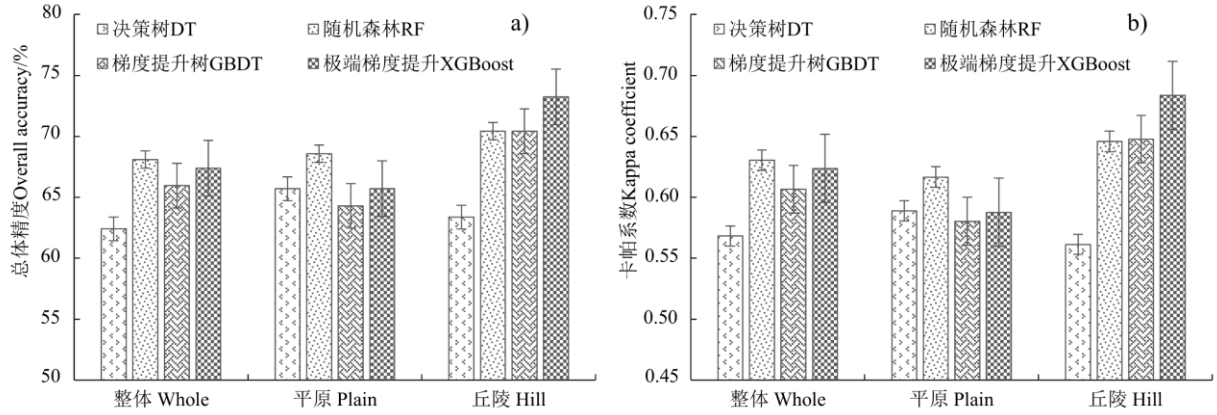

图 2 各区域推理制图方法对比

Fig. 2 Comparison of inference mapping methods in each region

2.2 环境变量选取

基于随机森林内嵌算法，分别计算整体、平原和丘陵区域所有环境变量的因子重要性，通过逐

一添加正向排序的环境变量用于各区域优选制图方法并进行精度评价。由图 3 可以看出，随着正向排序环境变量数目的增加，三种区域的优选制图方法精度均呈现先上升后平缓的趋势。整体区域制图精度在环境变量个数为 8 时最高，总体精度为 73.05%；平原区域制图精度最高为 70.00%，在环境变量个数为 17 时得到；当丘陵区域的环境变量个数为 11 时，制图精度达到最高，为 76.06%。

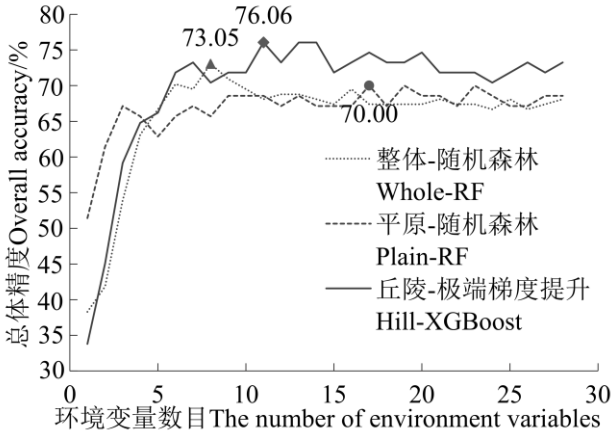

图 3 各区域优选制图方法总体精度随环境变量数目的变化

Fig. 3 Variation of overall accuracy for optimal mapping methods in each region with the number of environment variables

将各区域所有变量和筛选后变量用于各区域最优制图方法，进行土壤类型推理制图，结果如表 3 所示。与使用所有变量相比，使用筛选后变量，各区域制图精度都有了一定的提升。其中，整体区域制图精度提升效果最好，总体精度和 Kappa 系数分别提高了 4.96%和 0.06；其次是丘陵区域，总体精度提升了 2.82%，Kappa 系数增加了 0.03；平原区域制图精度的提升效果最小，总体精度和 Kappa 系数分别增加了 1.43%和 0.02。因此，环境变量筛选能够有效提升各区域土壤类型推理制图精度。本研究分别选取正向排序前 8、17 和 11 个环境变量作为整体、平原和丘陵区域的最优环境变量集。

表 3 变量筛选前后优选制图模型的性能

Table 3 The performance of optimal mapping models before and after variable selection

优选模型 Optimal model	所有变量 All variables		筛选后变量 Filtered variables	
	总体精度 Overall accuracy/%	卡帕系数 Kappa coefficient	总体精度 Overall accuracy/%	卡帕系数 Kappa coefficient
整体-随机森林 Whole-RF	68.09	0.63	73.05	0.69
平原-随机森林 Plain-RF	68.57	0.62	70	0.63
丘陵-极端梯度提升 Hill-XGBoost	73.24	0.68	76.06	0.72

2.3 最优环境变量集

根据随机森林内嵌模型计算的各区域最优环境变量集中环境变量的因子重要性，进行归一化处理，结果如图 4 所示。母质是整体、平原和丘陵区域中影响力相对最大的环境变量，但其重要性有所不同，在整体、平原和丘陵区域的重要性依次为 0.35、0.20 和 0.32。谷深、高程、相对坡度位置、坡向和坡度坡长因子为三个区域均必需的地形因子，平原和丘陵区域还需要的地形因子为地形湿度指数；遥感因子中的均值为三个区域所必需的，第一主成分为整体和平原区域所必需的因子，丘陵区域还需要的遥感因子为比值植被指数、土壤调整植被指数和归一化植被指数。与整体和丘陵区域相比，平原区域需要更多的遥感因子，包括绿光归一化植被指数、增强型植被指数、相关性、差值植被指数和方差，这可能是由于平原区域地形较为平缓，导致土壤类型分布的地带性特征不明显，

仅依靠地形因子和少量遥感因子无法提供足够的信息来进行推理制图。

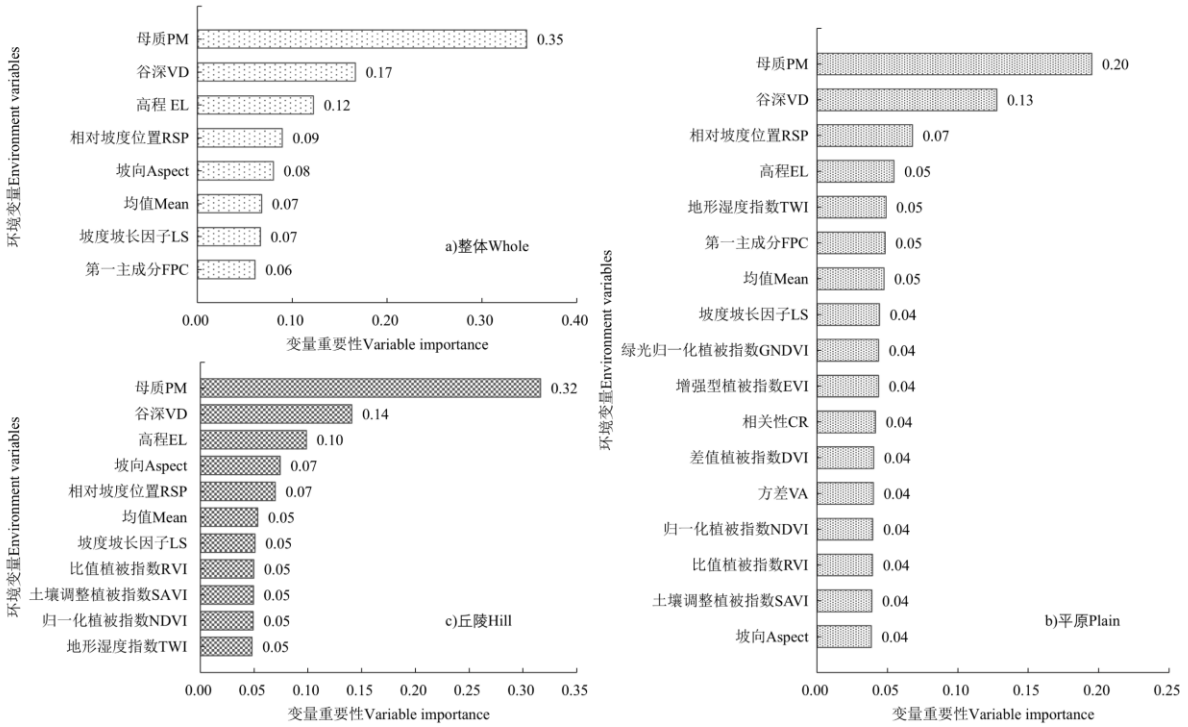

图 4 各区域筛选后环境变量的因子重要性

Fig. 4 Feature importance of filtered environment variables in each region

2.4 推理制图结果及精度评价

基于各区域最优制图方法和最优环境变量集，构建各区域最优土壤-景观关系模型，进行各区域土壤类型推理制图。将平原区域和丘陵区域土壤类型推理制图结果合并为按地形分区推理制图结果，并与整体区域土壤类型推理制图结果进行对比，结果如图 5 所示。在使用所有变量进行推理制图时，整体推理制图的总体精度为 68.09%，Kappa 系数为 0.6305，按地形推理制图的总体精度和 Kappa 系数分别为 70.92%和 0.6643，各自提高了 2.83%和 0.0338；在使用筛选后变量进行推理制图时，虽然按地形推理制图和整体推理制图的总体精度一致，皆为 73.05%，但 Kappa 系数由整体推理制图中的 0.6895 提升到按地形推理制图中的 0.6897。综上，按地形分区的制图方式也能有效提升平原-丘陵区域土壤类型推理制图精度，但在进行环境变量筛选后，其提升效果不明显。本研究选取基于筛选后变量进行的按地形推理制图为研究区最优推理制图结果。

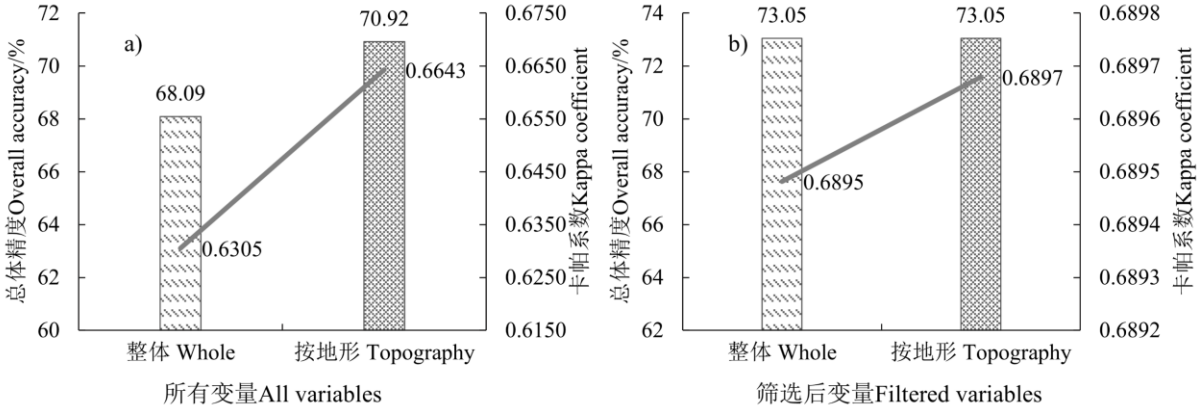

图 5 不同推理制图方式之间的对比

Fig. 5 Comparison of different inference mapping ways

将原始土壤图与研究区最优推理土壤图进行对比分析，根据图 6 可知，相比较于原始土壤图，推理土壤图的空间分布更加破碎，展现出更多的细节信息。整体上，两者空间分布大体一致，其中潮土田在两者中的空间分布完全一致，这可能是因为该土壤是唯一由近代河流冲积物发育而来。然而，部分土壤类型在推理土壤图中的空间分布范围与原始土壤图相比，发生了较大变化，如硅沙土和硅麻骨土，该两种土壤类型在推理土壤图中的分布面积明显减少，并且很大一部分被推理为硅沙泥土，这可能是由于硅沙土、硅麻骨土和硅沙泥土均由白云石英片岩花岗片麻岩坡积或残积物发育而来，成土环境类似，且在研究区中往往相伴而生，具有较强的相似性。

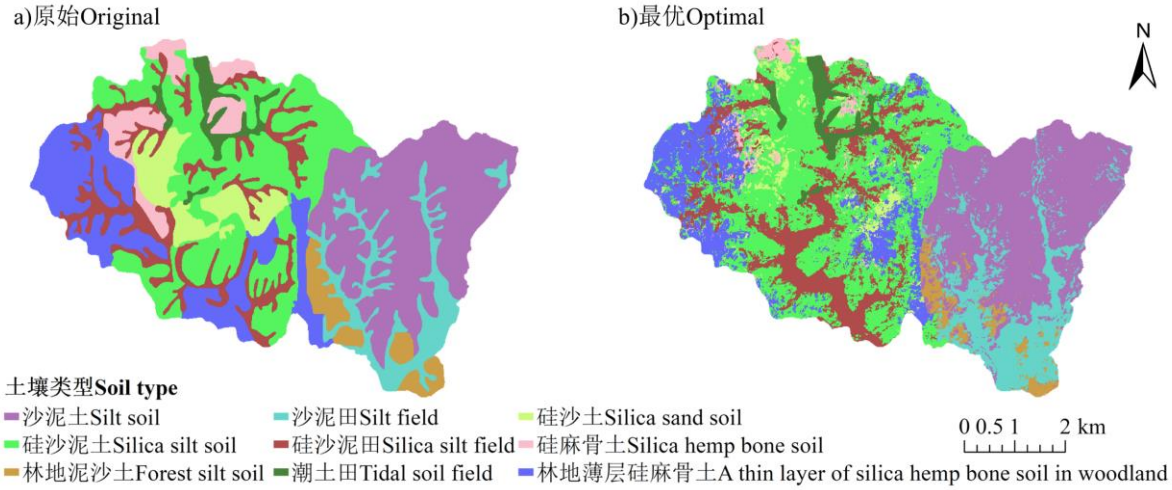

图 6 原始土壤图与最优推理土壤图

Fig. 6 Original soil map and optimal inference soil map

为了进一步评估最优推理土壤图的分类准确性，表 4 给出了各土壤类型在推理土壤图中的生产精度和用户精度。生产精度表示验证点土壤类型在推理土壤制图中被正确分类的概率，用户精度则表示在推理土壤制图中，落在该土壤类型上的验证点被判定为该土壤类型的概率。根据推理制图的原理，土种级别的推理制图相比较于土属级别更为复杂，因此难以取得较高的精度。通过混淆矩阵可以看出，所有土壤类型的用户精度均较为理想（均超过 60%）；除林地泥沙土、硅沙土和硅麻骨土外，其他土壤类型的生产精度均超过 70%。尽管林地泥沙土和硅麻骨土的生产精度仅为 33%和 27%，但它们的用户精度均为 100%，这表明在推理土壤图中，其他土壤类型不会被错误分类为这两种类型。

表 4 实地验证点与最优推理土壤图间的混淆矩阵

Table 4 Confusion matrix between validation points and optimal inference soil map

实地验证点 Validation points	最优推理土壤图 Optimal inference soil map										
	沙泥土 Silt soil	硅沙泥土 Silica silt soil	林地泥沙土 Forest silt soil	沙泥田 Silt field	硅沙泥田 Silica silt field	潮土田 Tidal soil field	硅沙土 Silica sand soil	硅麻骨土 Silica hemp bone soil	林地薄层硅麻骨土 A thin layer of silica hemp bone soil in woodland	合计 Total	生产精度 Producers' accuracy /%
沙泥土 Silt soil	13	0	0	3	0	0	0	0	0	16	81

硅沙泥土 Silica silt soil	0	22	0	0	2	0	1	0	0	25	88
林地泥沙土 Forest silt soil	2	0	3	4	0	0	0	0	0	9	33
沙泥田 Silt field	4	0	0	10	0	0	0	0	0	14	71
硅沙泥田 Silica silt field	0	1	0	0	17	0	0	0	1	19	89
潮土田 Tidal soil field	0	0	0	0	0	13	0	0	0	13	100
硅沙土 Silica sand soil	0	5	0	0	0	0	5	0	2	12	42
硅麻骨土 Silica hemp bone soil	0	3	0	0	5	0	0	3	0	11	27
林地薄层硅麻骨土 A thin layer of silica hemp bone soil in woodland	0	1	0	0	4	0	0	0	17	22	77
合计 Total	19	32	3	17	28	13	6	3	20	141	
用户精度 Users' accuracy /%	68	69	100	59	61	100	83	100	85		

3 讨 论

本研究首先对比了 DT、RF、GBDT 和 XGBoost 算法在整体、平原和丘陵区域的推理制图效果，然后基于因子重要性进行了环境变量筛选，并比较了整体和按地形推理制图结果。最后，结合优选的制图方法、筛选后的环境变量和按地形制图方式，进行了研究区土壤类型推理制图。研究结果为平原-丘陵区域推理制图中的环境变量筛选、制图算法选择和制图方式构建提供了一定的技术参考，有助于提升推理制图精度。此外，利用该方法生成的土壤类型图可以作为参考，与第二次全国土壤普查成果的土壤类型图进行对比，以识别潜在的错分区域，并为实地调查提供建议。

目前已有许多研究基于机器学习算法开展数字土壤制图，但仍存在一些不足，例如，杨雨菲等[24]利用多种机器学习算法进行推理制图时未考虑环境变量的重要性。张晓婷等[25]虽然考虑了环境变量在推理制图中的重要性，但未探讨不同机器学习算法在不同区域的表现差异。本文通过对比分析发现，RF 算法在整体和平原区域的推理制图效果最好，丘陵区域制图精度最高的为 XGBoost 算法。尽管四种机器学习算法在不同地形中的推理制图能力不同，相对而言，RF 与 XGBoost 算法的推理制图能力均较好（图 2），与先前所得结论一致[10, 16, 22]。此外，多种机器学习算法在丘陵区域的推理制图精度明显高于整体和平原区域，这是因为地形因素是土壤形成发育的主要影响因素之一，丘陵区域地形起伏度大，基于地形差异进一步引起其他环境变量的空间分布差异，从而导致土壤类型和性质的差异[26]。因此，在进行土壤类型推理制图时，需要根据不同的地形条件，选择合适的环境变量来进行推理制图，以提高制图精度。

除母质外，本文选取了 12 个地形因子和 15 个遥感因子作为土壤类型推理制图的环境变量集。为了降低数据冗余，减少数据计算量，本文对环境变量集信息进行了充分挖掘，优选出各区域环境变量。在各区域筛选后的环境变量中，地形因子在整体和丘陵区域的贡献率分别为 52%和 48%，而在平原区域的贡献率仅为 38%；遥感因子在整体、丘陵和平原区域的个数分别为 2、4 和 10，贡献率分别为 13%、20%和 42%，再次证明了平原区域地形差异较小，土壤类型分布的地带性特征明显减弱或不明显，需要更多的遥感因子参与推理制图以提升精度[27]。同时，与使用所有环境变量进行推理制图相比，使用筛选后环境变量进行推理制图时，土壤类型制图精度均得到了有效的提升，虽然提升效果在不同区域有所不同（表 3）。例如，整体区域推理制图的总体精度提升了 4.96%，而平原区域推理制图精度仅提升了 1.43%。此外，本研究还发现，按地形推理制图精度高于整体推理制图精度，与王苏放[28]研究结果一致。因此，本研究表明考虑制图方法优选、环境变量筛选以及地形分区制图方式均能有效提升平原-丘陵区域土壤类型推理制图精度。

但本研究尚有一些不足之处，已有研究表明不同的样点布设方式会产生不同的推理制图效果[29]，本文仅采用随机采样的方式布设训练点，未考虑训练点与图斑边界距离，导致图斑边缘训练点存在错分问题。因此若能够布设空间代表性高且准确分类的样点，有望进一步提高土壤类型推理制图精度。其次，随着计算机科学的发展，当前深度学习算法在推理制图中的应用逐渐受到关注，在今后的研究中可以做进一步的探索。同时，已有研究表明人类活动因子有利于提高土壤类型推理制图精度[30]，但本文中未考虑人类活动因子，后续的研究中可以引入该类因子以进一步提升推理制图精度。最后，本文构建的各区域最优土壤-景观关系模型仅适用于本研究区，其是否具有较好的可迁移性，还有待进一步探讨。

4 结 论

本研究以平原-丘陵区域为整体研究区，将其划分为平原区域和丘陵区域两个地形单元，探讨了综合考虑制图方法优选、环境变量筛选以及制图方式对平原-丘陵区域土壤类型推理制图精度的影响。利用 DT、RF、GBDT 和 XGBoost 四种机器学习算法进行各区域土壤类型推理制图，各制图方法均能实现土壤类型分类，但它们在不同地形下的制图效果不同，应根据具体地形来选择适宜的制图方法。基于 RF 模型内嵌的因子重要性进行的最优环境变量筛选，证明了环境变量筛选不仅能有效减少数据冗余，还能提高推理制图精度，但不同地形下的提升效果存在差异，整体区域的精度提升最为显著，总体精度和 Kappa 系数分别提高了 4.96%和 0.06。遥感因子在平原区域的贡献度远高于整体和丘陵区域。此外，采用地形分区制图方式能有效缓解多尺度地形复杂性，进而提高平原-丘陵区域的土壤类型推理制图精度。依据综合考虑制图方法优选、环境变量筛选以及制图方式构建的最优土壤-景观关系模型绘制的研究区土壤类型图，发现其总体精度和 Kappa 系数分别为 73.05%和 0.69。因此，本研究为构建平原-丘陵区域适宜性土壤-景观关系模型，从而提高土壤类型推理制图精度提供了一定的理论参考和技术支撑。

参考文献（References）

[1] Zhang G L, Shi Z, Wang Q B, et al. Development of soil geography in the New Era and its future[J]. Acta Pedologica Sinica, 2023, 60(5): 1264-1276. [张甘霖, 史舟, 王秋兵, 等. 新时代土壤地理学的发展现状与趋势[J]. 土壤学报, 2023, 60(5): 1264-1276.]

[2] Zhu A X, Yang L, Fan N Q, et al. The review and outlook of digital soil mapping[J]. Progress in Geography, 2018, 37(1): 66-78. [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] Zhang G L, Shi Z, Zhu A X, et al. Progress and perspective of studies on soils in space and time[J]. Acta Pedologica Sinica, 2020, 57(5): 1060-1070. [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.]

[4] Sun X L, Zhao Y G, Qin C Z, et al. Effects of DEM resolution on multi-factor linear soil-landscape models and their application in predictive soil mapping[J]. Acta Pedologica Sinica, 2008, 45(5): 971-977. [孙孝林, 赵玉国, 秦承志, 等. DEM 栅格分辨率对多元线性土壤—景观模型及其制图应用的影响[J]. 土壤学报, 2008, 45(5): 971-977.]

[5] Deng J, Shi H D, Zhao J, et al. Soil selenium distribution and its influencing factors in Zunyi city[J]. Soil and Fertilizer Sciences in China, 2019(3): 49-55. [邓军, 师华定, 赵建, 等. 遵义市土壤硒分布及其影响因素研究[J]. 中国土壤与肥料, 2019(3): 49-55.]

[6] Qiu L, Li A B, Zhao Y G. Digital soil mapping based on Fisher discriminant analysis[J]. Chinese Journal of Soil Science, 2012, 43(6): 1281-1286. [邱琳, 李安波, 赵玉国. 基于 Fisher 判别分析的数字土壤制图研究[J]. 土壤通报, 2012, 43(6): 1281-1286.]

[7] Han H W, Xu W, Huang W, et al. Soil mapping based on remote sensing images and decision tree algorithm[J]. Chinese Journal of Soil Science, 2019, 50(1): 8-14. [韩浩武, 许伟, 黄魏, 等. 基于遥感影像和决策树算法的土壤制图[J]. 土壤通报, 2019, 50(1): 8-14.]

[8] Li K, Huang W, Fu P H, et al. Research on Soil Type Inference Based on Combinatorial Cartography Method[J]. Acta Pedologica Sinica, 2024: DOI:10.11766/trxb202402030056 [李坤，黄魏，傅佩红，等. 基于结合型制图方法的土壤类型推理研究[J]. 土壤学报, 2024: DOI:10.11766/trxb202402030056]

[9] Zhu Y L, Feng X Y, Yan Q W, et al. Spatial distribution and main controlling factors of soil organic carbon under cultivated land based on GBDT model in black soil region of Northeast China[J]. China Environmental Science, 2024, 44(3): 1407-1417. [祝元丽, 冯向阳, 闫庆武, 等. 基于 GBDT 的望奎县农田土壤有机碳主控因子研究[J]. 中国环境科学, 2024, 44(3): 1407-1417.]

[10] Qiu S Q, Zhao M S, Lu Y Y, et al. Spatial modeling of soil pH in Anhui Province based on XGBoost model[J]. Science Technology and Engineering, 2023, 23(4): 1472-1480. [邱士其, 赵明松, 芦园园, 等. 基于 XGBoost 模型的安徽省土壤 pH 空间建模[J]. 科学技术与工程, 2023, 23(4): 1472-1480.]

[11] Guo F, Xu Z, Ma H H, et al. Estimating chromium concentration in arable soil based on the optimal principal components by hyperspectral data[J]. Ecological Indicators, 2021, 133: 108400.

[12] Feng F, Wang Y H, Zuo Y F. A study on factors that influence the spatial distribution of soil cadmium pollution based on RF-XGBoost[J]. Journal of Agro-environment Science, 2023, 42(4): 811-819. [冯锋, 王育红, 左雨芳. 基于 RF-XGBoost 的土壤镉污染影响因子及空间分布研究[J]. 农业环境科学学报, 2023, 42(4): 811-819.]

[13] Ye H C, Nie C J, Zhang Y, et al. Research on digital mapping of soil types in different topographical units assisted by environmental variables[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, https://link.cnki.net/urlid/11.1964.s.20240723.1730.007. [叶回春，聂超甲，张越，等. 基于环境变量辅助的不同地形单元土壤类型数字制图研究[J/OL] . 农业机械学报, https://link.cnki.net/urlid/11.1964.s.20240723.1730.007.]

[14] Mcbratney A, Mendonça santos M, Minasny B. On digital soil mapping[J]. Geoderma, 2003, 117(1): 3-52.

[15] Chen R, Han H W, Fu P H, et al. Soil mapping based on multi-temporal remote sensing images and random forest algorithm[J]. Soils, 2021, 53(5): 1087-1094. [陈荣, 韩浩武, 傅佩红, 等. 基于多时相遥感影像和随机森林算法的土壤制图[J]. 土壤, 2021, 53(5): 1087-1094.]

[16] Guo J H, Liu F, Xu S X, et al. Comparison of digital mapping methods for the thickness of black soil layer of cultivated land in typical black soil area of Songnen Plain[J]. Journal of Geo-Information Science, 2024, 26(6): 1452-1468. [郭俊辉, 刘峰, 徐胜祥, 等. 松嫩典型黑土区耕地黑土层厚度数字制图方法比较[J]. 地球信息科学学报, 2024, 26(6): 1452-1468.]

[17] Han W T, Cui J W, Cui X, et al. Estimation of farmland soil salinity content based on feature optimization and machine learning algorithms[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023,54(3):328-337. [韩文霆，崔家伟，崔欣，等. 基于特征优选与机器学习算法的农田土壤含盐量估算研究[J]. 农业机械学报, 2023,54(3):328-337.]

[18] Zebari D A, Haron H, Zeebaree D Q, et al. A simultaneous approach for compression and encryption techniques using deoxyribonucleic acid[C]//2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA). August 26-28, 2019, Island of Ulkulhas, Maldives. IEEE, 2019: 1-6.

[19] Schonlau M, Zou R Y. The random forest algorithm for statistical learning[J]. The Stata Journal: Promoting Communications on Statistics and Stata, 2020, 20(1): 3-29.

[20] Liang W Z, Luo S Z, Zhao G Y, et al. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms[J]. Mathematics, 2020, 8(5): 765.

[21] Chen T Q, Guestrin C. XGBoost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on

Knowledge Discovery and Data Mining. San Francisco California USA. ACM, 2016: 785-794.

[22] Yan H B, Liang Y H, Lu X J, et al. Remote sensing modeling and applications in drought monitoring based on XGBoost and fusion of multi-dimensional spatiotemporal data[J]. Journal of Geo-Information Science, 2024, 26(6): 1531-1546. [晏红波, 梁雨豪, 卢献健, 等. 基于 XGBoost 融合多维度时空数据的干旱遥感建模及应用研究[J]. 地球信息科学学报, 2024, 26(6): 1531-1546.]

[23] Wu Y N, Guo C E, Yu D P, et al. Spatial stratification and evaluation method of remote sensing classification based on uncertainty analysis[J]. Journal of Geo-Information Science, 2022, 24(9): 1803-1816. [吴亚楠, 郭长恩, 于东平, 等. 基于不确定性分析的遥感分类空间分层及评估方法[J]. 地球信息科学学报, 2022, 24(9): 1803-1816.]

[24] Yang Y F, Han H W, Chen R, et al. Inference methods of digital soil mapping[J]. Chinese Journal of Soil Science, 2020, 51(5): 1016-1023. [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.]

[25] Zhang X T, Huang W, Fu P H, et al. Research on digital soil mapping based on feature selection algorithm[J]. Acta Pedologica Sinica, 2024, 61(3): 635-647. [张晓婷, 黄魏, 傅佩红, 等. 基于特征筛选算法的数字土壤制图研究[J]. 土壤学报, 2024, 61(3): 635-647.]

[26] Guo P T. Study on predictive soil mapping in hilly and mountainous areas— Taking topographic factors as auxiliary variables[D]. Chongqing: Southwest University, 2009. [郭澎涛. 丘陵山地预测性土壤制图研究——以地形因子为辅助变量[D]. 重庆: 西南大学, 2009.]

[27] Liu X B. Comparative study on regional digital soil mapping methods[D]. Zhengzhou: Zhengzhou University, 2013. [刘晓冰. 区域数字土壤制图方法对比研究[D]. 郑州: 郑州大学, 2013.]

[28] Wang S F. Study on digital soil mapping in mixed region of plain and hilly areas[D]. Wuhan: Huazhong Agricultural University, 2022. [王苏放. 平原丘陵混合区域数字土壤制图研究[D]. 武汉: 华中农业大学, 2022.]

[29] Huang W, Luo Y, Wang S Q, et al. Research on obtaining soil-environment relationship and reasoning mapping based on traditional soil map[J]. Acta Pedologica Sinica, 2016, 53(1): 72-80. [黄魏, 罗云, 汪善勤, 等. 基于传统土壤图的土壤—环境关系获取及推理制图研究[J]. 土壤学报, 2016, 53(1): 72-80.]

[30] Liu C, Dong C, Wang Z R, et al. Effects of human activity factors on the accuracy of soil type digital mapping model in hilly areas[J]. China Agricultural Informatics, 2023, 35(3): 58-74. [刘成, 董超, 王卓然, 等. 山丘区人类活动因子对土壤类型数字制图模型精度的影响[J]. 中国农业信息, 2023, 35(3): 58-74.]

（责任编辑：檀满枝）