

DOI: 10.13758/j.cnki.tr.2021.05.026

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. 土壤, 2021, 53(5): 1087–1094.

基于多时相遥感影像和随机森林算法的土壤制图^①

陈 荣, 韩浩武, 傅佩红, 杨雨菲, 黄 魏*

(华中农业大学资源与环境学院, 武汉 430070)

摘 要: 获取准确的土壤-环境关系是数字土壤制图的关键, 目前遥感影像已作为环境因子应用于土壤-环境知识的建立过程, 但单幅遥感影像所包含的光谱信息差异难以将不同土壤类型区分开来。因此本文提出了一种基于多时相遥感影像的土壤制图方法: 选取红安县澧水河流域为研究区, 以母质类型图、等高线数据和多时相哨兵二号遥感影像为基础, 提取与土壤形成有关的环境因子, 通过随机森林算法获取土壤-环境关系, 预测研究区各土壤类型的空间分布并成图, 利用野外实地分层采样点验证推理图的精度。结果表明: 推理土壤图总体分类精度高达 86%, 与原始土壤图对比, 各土壤类型的空间分布具有一定相似性, 展现了更为详细的空间细节信息, 该研究成果可为更新土壤图工作提供新方法。

关键词: 土壤-景观推理模型; 随机森林算法; 遥感影像; 数字土壤制图

中图分类号: S159 文献标志码: A

Soil Mapping Based on Multi-temporal Remote Sensing Images and Random Forest Algorithm

CHEN Rong, HAN Haowu, FU Peihong, YANG Yufei, HUANG Wei*

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

Abstract: Extracting accurate soil-environment relationship is the key to digital soil mapping. Nowadays, remote sensing images have been used as the indicators of environmental factors in the process of obtaining soil-environment knowledge. However, the spectral differences in mono-temporal image are difficult to be used to distinguish soil types. In this study, we proposed a soil mapping method based on multi-temporal remote sensing images. The Sheshui River Basin in Huajiahe Town, Honggan County, Huanggang City of Hubei Province was selected as the study area, and the parent-material-type map, the multi-temporal sentinel-2 remote sensing images, and contour data were used to extract environmental factors related to soil properties. Soil environment relationships were obtained to infer the spatial distribution of soil types using the random forest algorithm. The field sampling points in the study area were used for validation, and the confusion matrix and Kappa coefficient of inferred soil map were calculated to evaluate the map accuracy. The results demonstrated that the overall classification accuracy of the inferred soil map was as high as 86%. The soil type map obtained by inference was similar to the traditional soil map in the spatial distribution, but it could display more detailed information than the traditional soil map. This research can provide an effective alternative for updating the traditional soil map.

Key words: Soil land inference model; Random forest algorithm; Remote sensing images; Digital soil mapping

土壤普查是了解土壤资源空间分布及合理利用土壤资源的前提, 是农业生产的基础。传统的土壤普查主要是土壤学家以大比例尺航片和地形图为底图, 通过实地采样获得土壤-景观模型, 以此判断不同土壤类型的空间分布并手绘成图^[1-2]。此方法存在的精

度低、人为主观性强及时间和人工成本高^[3]等问题推动了数字土壤制图的诞生与发展。与传统土壤普查不同的是, 数字土壤制图方法是以空间分析和数学方法为技术手段, 用栅格形式来表达土壤空间分布信息, 此种方法可更好地表达出土壤分布的渐变特征^[4-6]。

①基金项目: 国家自然科学基金项目(41877001)、国家重点研发计划项目(2017YFD0202000)和中央高校基本科研业务费专项资金项目(2662019PY074)资助。

* 通讯作者(ccan@mail.hzau.edu.cn)

作者简介: 陈荣(1996—), 女, 湖北咸宁人, 硕士研究生, 主要从事精细数字土壤制图研究。E-mail: charlottechen@webmail.hzau.edu.cn

土壤-环境知识的获取和土壤空间推理方法的选择是当前精细数字土壤制图的关键。

目前,数字土壤推理制图过程常用到的协同因子多为由数字高程模型(digital elevation model, DEM)衍生的地形因子。遥感影像因其时效性强、数据分辨率高、易获取及含有丰富的下垫面信息等优势逐渐应用于土壤制图,其包含的地表光谱信息可用于反演地物属性、地物分类,提取的多种遥感指数可用于定量或定性地表植被生长状况。遥感影像可以多种形式参与到土壤推理制图过程中,如利用土壤有机质的遥感响应波段以实现其含量的估算^[7],借助影像建立特定时段的地表动态反馈模型^[8-10],在时序遥感影像数据上提取植被的轮作模式^[11-12],或直接在遥感影像数据中提取地物的分类特征^[13-15]等。但在以往的研究中,土壤制图使用的遥感影像多为单时相数据^[16],不同时间点遥感影像产生的光谱信息具有较大差异,制图结果也存在一定差异。且单一时间点的遥感数据易受天气、季节等因素的影响,在遥感影像时期的选择上也存在着人为主观性,这些环节都将直接影响制图结果。

土壤分类常用的数据挖掘算法主要有模糊聚类^[17-18]、决策树^[19-20]、支持向量机^[21]、判别分析^[22]、随机森林^[23-24]等。其中,随机森林算法可以有效避免原始数据的缺失及噪声、异常值造成的精度低等问题,在分类和回归等多方面具有一定的优势,在国内

数字土壤制图领域中该方法被广泛应用于获取土壤类型^[23]或属性^[25-27]的空间分布信息。本研究拟将多时相遥感影像数据和随机森林算法应用于土壤分类,为土壤分类提供新的技术与方法。

本研究使用图斑面积加权法在传统土壤图上设计采样点,在母质和地形等环境协同因子的基础上,加入多时相遥感指数,通过随机森林算法建立土壤类型与环境协同因子组合的对应关系,并推理得到研究区各土壤类型的空间分布信息,进而对传统土壤图进行更新,同时探究多时相遥感影像数据在土壤制图中的可行性。

1 材料与方法

1.1 研究区概况

研究区域位于华家河镇淝水河流域(图 1),隶属于湖北省黄冈市红安县(114°23'~114°49'E, 30°56'~31°35'N),地处大别山南麓,境内地势北高南低,为低山丘陵区。土地利用类型以林地和农耕地为主。根据全国第二次土壤普查结果,研究区土壤类型共 12 种,母质类型共 6 种,其中石英片岩、基性岩分别与硅沙泥田、乌沙泥田相对应,因此这两种土壤类型不参与后续建模和推理过程。为方便后续随机森林建模和预测,数据预处理时将石灰岩、片麻岩、近代河流冲积物、泥质岩 4 种母质分别编号为 1、2、3、4。研究区各土壤类型和其对应的母质类型如表 1 所示。

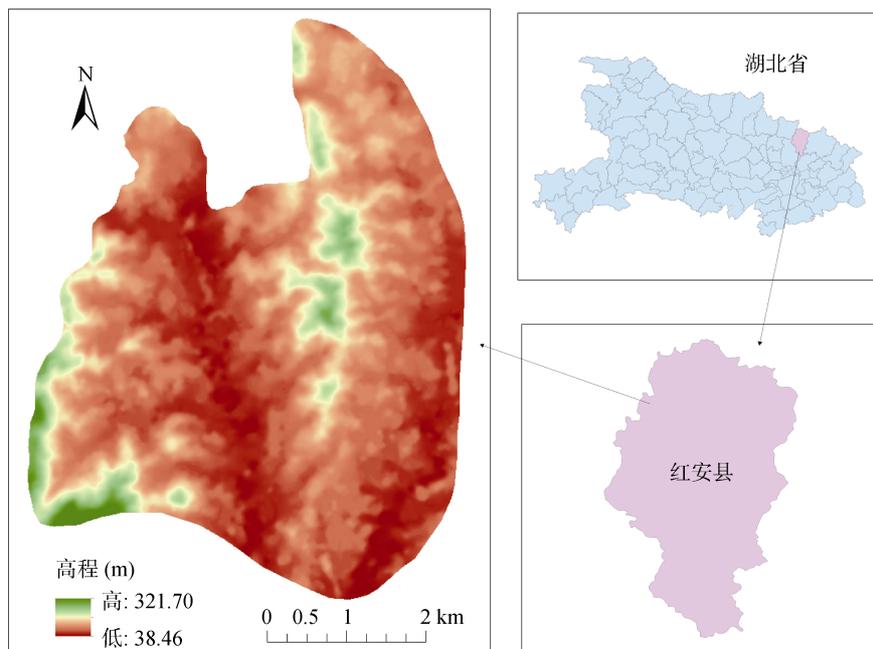


图 1 研究区域示意图及 DEM
Fig.1 Location and DEM of study area

表 1 溇水河流域土壤类型和母质类型
Table 1 Soil and parent material types in Sheshui River Basin

序号	母质类型	土壤类型名称	土壤类型代号
1	泥质岩	细沙泥土	2
2	泥质岩	林地细沙泥土	2-1
3	片麻岩	沙泥土	3
4	片麻岩	林地沙泥土	3-1
5	石灰岩	棕色石灰土	5
6	石灰岩	林地棕色石灰土	5-1
7	近代河流冲积物	浅潮沙田	13
8	石英片岩	硅沙泥田	14
9	泥质岩	细沙泥田	15
10	片麻岩	沙泥田	16
11	基性岩	乌沙泥田	18
12	近代河流冲积物	潮沙田	19

1.2 数据来源及处理

本研究使用的原始数据主要包括母质类型图、等

高线、遥感影像和传统土壤图。根据红安县土壤图和地质图及当地土壤普查专家的意见,研究区域可以划分成 6 种母质类型。地形因子数据集是在湖北省测绘局的 10 m 等高距的等高线数据的基础上衍生而来,在 ArcGIS 中内插得到研究区的数字高程模型,提取常用的高程、坡度、坡向等地形因子。遥感影像数据来源于哨兵二号卫星(<https://scihub.copernicus.eu/>),分辨率为 10 m,选择一个完整的作物生长周期,成像时间由 2016 年 11 月至 2017 年 10 月,每月选择一景遥感影像,借助 ENVI 软件平台计算归一化植被指数,具体成像时间如表 2 所示。

研究区原始土壤类型图的比例尺为 1 : 50 000,由第二次全国土壤普查结果经数字化后得到。传统土壤图是当地土壤普查专家进行野外调查的制图成果,其往往蕴含着土壤学家的“隐性知识”^[3],即土壤-环境知识,利用决策树、随机森林等数据挖掘算法可有效提取隐性知识并加以利用,更新传统土壤图。

表 2 哨兵二号影像获取时间(年-月-日)

Table 2 Acquisition time of Sentinel-2 images

序号	成像时间	序号	成像时间	序号	成像时间
1	2016-11-26	5	2017-03-26	9	2017-07-24
2	2016-12-09	6	2017-04-15	10	2017-08-28
3	2017-01-15	7	2017-05-05	11	2017-09-17
4	2017-02-27	8	2017-06-07	12	2017-10-27

1.3 采样点设计

采样点的设计方式对数字土壤制图的结果具有重要影响。本研究中采样点数据依据传统土壤图图斑面积,采用面积加权法进行布设采样点,并保证每个图斑中最少采样点数量不低于 10 个,最后共生成采样点数量 6 686 个。

1.4 随机森林算法

随机森林(random forest, RF)分类算法是 Breiman^[28]于 2001 年提出的一种机器学习方法,是一种以决策树为基分类器的 Bagging 集成算法,广泛应用于回归和分类。随机森林算法基于自助采样法(bootstrap sampling)^[29],即有放回的采样或重复采样,从初始数据集中有放回的反复抽取数据作为训练样本集,在采样过程中约有 36.8% 的数据始终没有被抽中,这部分数据为袋外(out-of-bag, OOB)数据,用以对模型性能进行估算,称为袋外估计。OOB 误差是随机森

林模型泛化误差的无偏估计^[28]。随机森林模型中有两个重要参数:节点分裂次数 m_{try} 和决策树数量 n_{tree} ,这两个参数值的选择直接决定了随机森林模型的性能。

1.5 精度评价

本研究利用横截面采样、主观采样和均匀采样 3 种方式得到的 264 个验证点对推理制图结果的准确性进行评价。验证点集包含研究区存在的所有土壤类型,且在整个研究区均有分布,如图 2 所示。建立实地验证点与推理土壤图的混淆矩阵,可得到总体分类精度、生产精度、用户精度及 Kappa 系数等评价指标。其中生产精度是预测土壤图某土壤类型正确分类的点数与野外采样点中该土壤类型的总点数的比值,而用户精度是预测土壤图某土壤类型正确分类的点数与预测土壤图中该土壤类型的总点数之比^[30]。

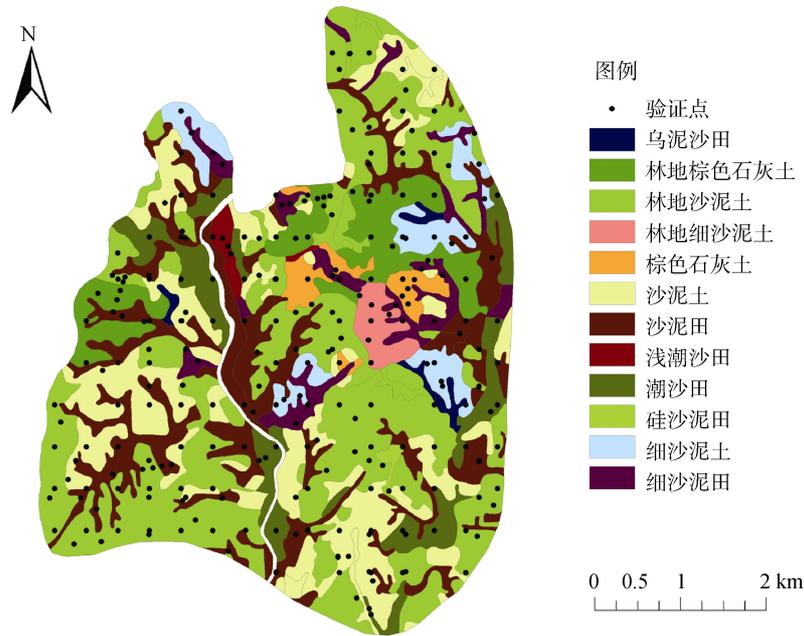


图 2 验证点分布图

Fig.2 Distribution of validation points

2 结果与分析

2.1 土壤类型与环境因子关系

本研究提取的原始环境因子如表 3 所示,包括母质类型、地形因子和遥感因子。基于采样点数据分别建立土壤类型与地形因子、遥感因子之间的关系(图 3 和图 4)。从图 3 可以看出,除高程、坡度和坡向这 3 个地形因子外,其余因子在不同土壤类型之间差异较小,因此本研究中仅选取高程、坡度和坡向参与到后续随机森林建模和推理过程中。从图 4 可以看到,不同土壤类型的全年植被指数变化规律相似,夏季 7 月和 8 月 NDVI 达到峰值,冬季则最低;林地细沙泥土(21)、林地沙泥土(31)和林地棕色石灰土(51)这 3 种土壤类型的归一化植被指数全年均处于较高水平,这可能与其土地利用类型有关;此外,4—6 月及 9—10 月,各土壤类型对应的 NDVI 值差异较大,可能是因为不同类型的植被生长及秋季叶片凋谢的速度不一致,从而导致植被指数出现较大差异。

表 3 环境因子统计
Table 3 Statistics of environmental factors

因子类别	因子名称	因子数量	分辨率(m)
母质类型	母质	1	10
地形因子	高程、坡度、坡向	3	10
遥感因子	12 月归一化植被指数	12	10

2.2 随机森林模型

本研究通过 R 语言中 random forest 包来建立随

机森林模型,环境因子为输入变量,土壤类型为输出变量,采用遍历的方式,以 OOB 误差的大小来确定模型的最优参数组合。由此得到 mtry 和 ntree 分别取 6 和 600 时,随机森林模型性能最佳,进而利用该模型进行预测。

2.3 制图结果与分析

对比原始土壤图(图 5A)和随机森林预测推理土壤图(图 5B),整体上两者的各类土壤空间分布具有一定相似性,林地沙泥土、沙泥土、细沙泥土、沙泥田、棕色石灰土等在两个土壤图中的空间分布基本一致,但棕色石灰土、浅潮沙田、沙泥土、沙泥田的面积有所减少。推理土壤图的图斑数量显著增多,展现出更精细的土壤空间细节信息,且不同土壤类型之间不再存在明显的“分界线”,过渡平滑,呈现出自然界中土壤空间分布的渐变特征。

2.4 制图精度验证

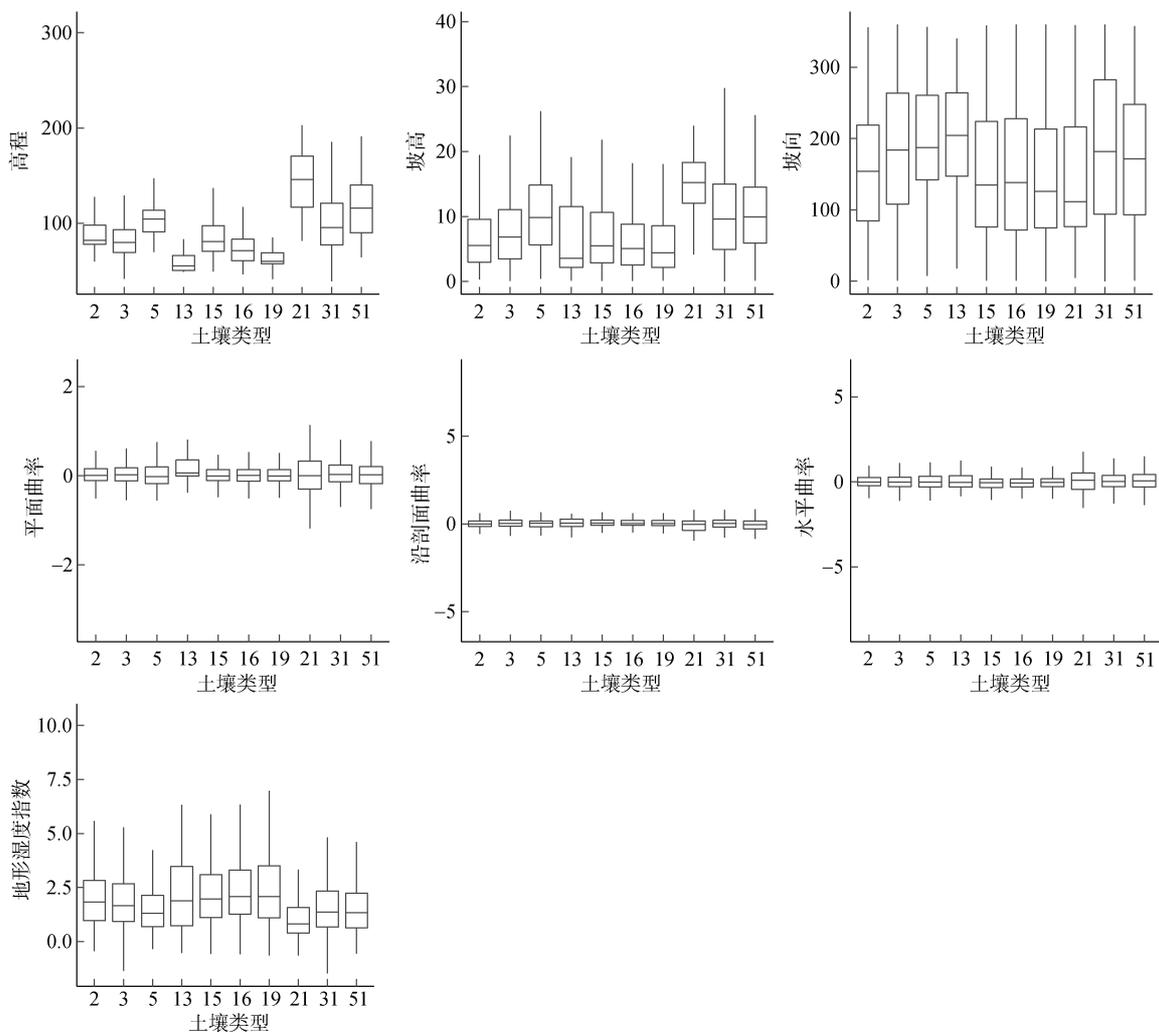
为进一步评价土壤分类的准确性,我们采用野外独立样本点对推理后土壤图进行评价。评价结果以混淆矩阵的形式展示,其中包括总体分类精度、各土壤类型的生产精度和用户精度及 Kappa 系数等精度评价指标,如表 4 所示。总体来说,推理结果图的整体分类精度达到了 86%,分类结果较好;Kappa 系数为 83%,表现出高度一致性,这说明验证样点与土壤图之间的契合程度高,推理土壤图所表达的土壤类型空间分布信息是可信的。大部分土壤类型的分类精度较高:除浅潮沙田外,其余土壤类型均有良好的用户精

度($\geq 70\%$)；除棕色石灰土和细沙泥田外，其余土壤类型的生产精度良好($\geq 70\%$)。因此，基于多时相植被指数的土壤推理方法具有较好的制图结果。

从混淆矩阵中可以看到，首先，林地沙泥土、林地棕色石灰土的生产精度和用户精度均较高。这是由于这两种土壤类型主要分布于研究区北部高程较大的丘陵处，土地利用类型为林地，与同母质发育而来的其他土壤类型相比，其归一化植被指数值较大，容易区分；沙泥土和潮沙田与研究区内河流分布关系密切，土壤养分较丰富，肥力水平高，宜耕性强，作物长势稳健，亦具有较高的植被指数，故分类精度较高。其次，浅潮沙田的生产精度和棕色石灰土的用户精度均较低，混淆矩阵中可以看出，这两种土壤类型容易被错分为同种母质下的潮沙田和林地棕色石灰土，其

原因可能是同一母质发育而来的两种土壤类型分布区域间地形起伏程度差异较小，且土地利用类型相近，作物生长规律相似，在时序遥感影像上表现为数值相近、无明显差异性的归一化植被指数，因此利用地形因子和多时相的遥感植被指数均不能有效区分，从而产生分类误差。此外，细沙泥田的生产精度较低，分类时易与细沙泥土混淆，从而影响到整体分类精度。

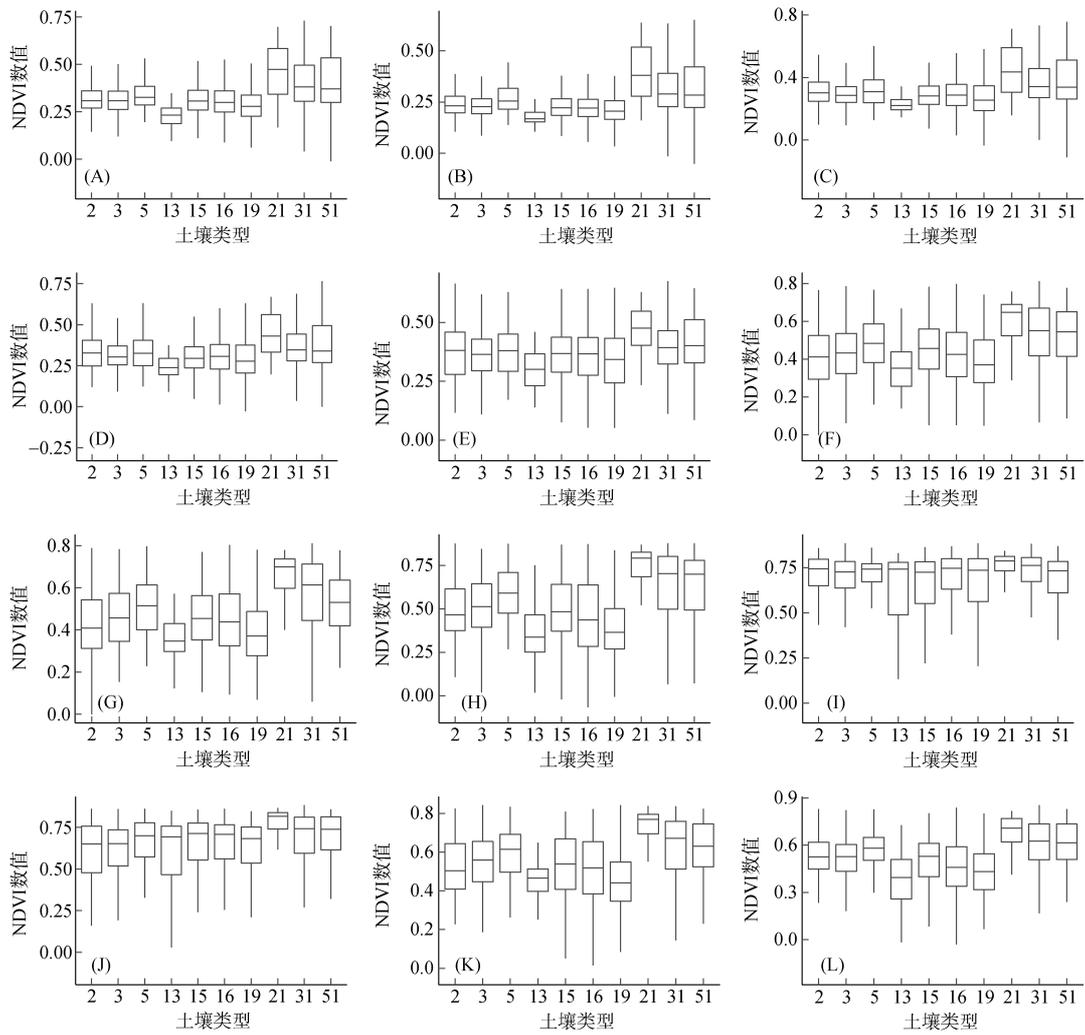
综上所述，经 264 个野外独立样点验证结果表明，基于多时序遥感影像推理得到的土壤图分类精度达到了 86%，土壤图斑更为破碎，数量变多，空间详细程度更高，Kappa 系数为 83%，验证样点与推理之间具有显著的一致性水平，因此利用这种方法来推理得到研究区的土壤类型空间分布信息是可行的。



(图中横坐标的土壤类型分别为 2: 细沙泥土; 3: 沙泥土; 5: 棕色石灰土; 13: 浅潮沙田; 15: 细沙泥田; 16: 沙泥土; 19: 潮沙田; 21: 林地细沙泥土; 31: 林地沙泥土; 51: 林地棕色石灰土, 下图同)

图 3 土壤类型与地形因子的关系

Fig.3 Relationships between soil types and terrain factors



(A~L 分别表示 1—12 月)

图 4 土壤类型与遥感因子的关系

Fig.4 Relationships between soil types and remote sensing factors

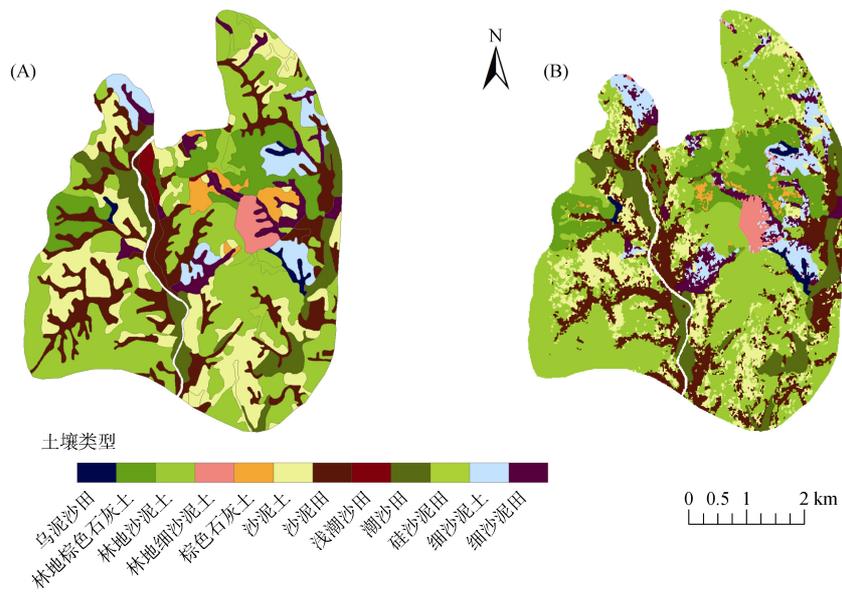


图 5 原始土壤图(A)和推理土壤图(B)

Fig.5 Original (A) and inferred (B) soil maps

表 4 实地验证点与推理土壤图间的混淆矩阵
Table 4 Confusion matrix between filed validation points and inferred soil map

实地验证点	推理土壤图									合计	用户精度(%)	
	细沙 泥土	沙泥土	棕色石 灰土	浅潮 沙田	细沙 泥田	沙泥田	潮沙田	林地细沙 泥土	林地沙 泥土			林地棕色 石灰土
细沙泥土	21				5						26	81
沙泥土		23				2			8		33	70
棕色石灰土			3							1	4	75
浅潮沙田				3			3				6	50
细沙泥田					12						12	100
沙泥田		2				37			4		43	86
潮沙田							20				20	100
林地细沙泥土	2				1			7			10	70
林地沙泥土		7							76		83	92
林地棕色石灰土			3							24	27	89
合计	23	32	6	3	18	39	23	7	88	25	264	
生产精度(%)	91	72	50	100	67	95	87	100	86	96		

注：总 264 个验证点，其中 226 个点分类正确，总体精度 86%，Kappa 系数为 83%。

3 讨论

本研究提出了一种基于多时相遥感影像和随机森林算法的土壤推理制图方法。利用时序哨兵二号遥感影像数据提取归一化植被指数，作为遥感因子与常用的高程等地形因子、母质类型组成完整的环境因子集，并通过随机森林算法提取土壤-环境知识后推理制图，经野外独立验证点验证其结果分类精度高达 86%，更新后土壤图表达出的研究区内各土壤类型的分布信息可信度较高。

已有研究大多采用单时相遥感影像数据，通过提取植被指数、纹理信息、主成分等遥感光谱指数用于土壤分类及土壤属性制图，并取得了较好的结果，但同时我们也应看到不同季节影像的制图结果具有较大的差异，且在影像时间的选择上多存在主观性。相比于单时相遥感影像数据易受天气、季节等因素的影响，本研究采用多时相遥感影像数据参与土壤制图，避免了依赖单时相遥感影像数据的局限性。周紫燕等^[23]在同一区域的研究中，同样采用了随机森林算法进行数据挖掘，但不同的是其使用的环境因子中除了常用的地形因子外，只加入了单时相的遥感因子，制图精度为 76%。与其相比，本研究加入的是多时相遥感因子，推理土壤图的准确性有了明显提升，精度提高了 10%。本研究结果具有较高的预测精度，各土壤类型空间分布的详细程度有较大提升，可以满足未来对高精度数字土壤制图的要求，为土壤普查工作提供参考依据。

本研究虽取得了较好的分类精度，但仍有一些需要改进的地方。在遥感影像的数据基础上，仅提取了时序归一化植被指数，未对其他可从影像提取的光谱信息加以利用，而归一化植被指数实际上由植被和土壤两部分组成，土壤背景对归一化植被指数干扰强烈，影响较大，且在植被稀疏的地方，归一化植被指数的表征能力较差。未来可引入更多能间接反映土壤发生发育的遥感光谱指数，同时，为避免环境因子较多带来的数据冗余现象，可采用主成分分析对因子进行降维处理，以减少不必要的信息冗余；此外，可对研究区进行地貌分区，针对不同区域采用不同的推理协同因子组合和权重，分区制图。

参考文献：

- [1] Hudson B D. The soil survey as paradigm-based science[J]. Soil Science Society of America Journal, 1992, 56(3): 836-841.
- [2] Zhu A X, Hudson B, Burt J, et al. Soil mapping using GIS, expert knowledge, and fuzzy logic[J]. Soil Science Society of America Journal, 2001, 65(5): 1463-1472.
- [3] Shi X, Zhu A X, Burt J E, et al. A case-based reasoning approach to fuzzy soil mapping[J]. Soil Science Society of America Journal, 2004, 68(3): 885-894.
- [4] McBratney A B, Mendonça Santos M L, Minasny B. On digital soil mapping[J]. Geoderma, 2003, 117(1/2): 3-52.
- [5] 孙福军, 雷秋良, 刘颖, 等. 数字土壤制图技术研究进展与展望[J]. 土壤通报, 2011, 42(6): 1502-1507.
- [6] 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.

- [7] 王琪, 吴成永, 陈克龙, 等. 基于多光谱遥感图像的青海湖流域土壤有机质估算初探[J]. 土壤, 2019, 51(1): 160–167.
- [8] 刘峰, 朱阿兴, 李宝林, 等. 利用陆面反馈动态模式来识别土壤类型的空间差异[J]. 土壤通报, 2009, 40(3): 501–508.
- [9] Zhu A X, Liu F, Li B L, et al. Differentiation of soil conditions over low relief areas using feedback dynamic patterns[J]. Soil Science Society of America Journal, 2010, 74(3): 861–869.
- [10] Zeng C Y, Zhu A, Liu F, et al. The impact of rainfall magnitude on the performance of digital soil mapping over low-relief areas using a land surface dynamic feedback method[J]. Ecological Indicators, 2017, 72: 297–309.
- [11] 宋敏, 杨琳, 朱阿兴, 等. 轮作模式在农耕地土壤有机质推测试制中的应用[J]. 土壤通报, 2017, 48(4): 778–785.
- [12] Yang L, Song M, Zhu A, et al. Predicting soil organic carbon content in croplands using crop rotation and Fourier transform decomposed variables[J]. Geoderma, 2019, 340: 289–302.
- [13] Mulder V L, de Bruin S, Schaepman M E, et al. The use of remote sensing in soil and terrain mapping—A review[J]. Geoderma, 2011, 162(1/2): 1–19.
- [14] Sreenivas K, Dadhwal V K, Kumar S, et al. Digital mapping of soil organic and inorganic carbon status in India[J]. Geoderma, 2016, 269: 160–173.
- [15] 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.
- [16] 刘焕军, 杨昊轩, 徐梦园, 等. 基于裸土期多时相遥感影像特征及最大似然法的土壤分类[J]. 农业工程学报, 2018, 34(14): 132–139, 304.
- [17] 朱阿兴, 李宝林, 杨琳, 等. 基于 GIS、模糊逻辑和专家知识的土壤制图及其在中国应用前景[J]. 土壤学报, 2005, 42(5): 142–149.
- [18] 杨琳, Fahmy S, Hann S, 等. 基于土壤–环境关系的更新传统土壤图研究[J]. 土壤学报, 2010, 47(6): 1039–1049.
- [19] 周斌, 王繁, 王人潮. 运用分类树进行土壤类型自动制图的研究[J]. 水土保持学报, 2004, 18(2): 140–143, 147.
- [20] 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.
- [21] 石伟, 南卓铜, 李韧, 等. 基于支持向量机的典型冻土区土壤制图研究[J]. 土壤学报, 2011, 48(3): 461–469.
- [22] 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.
- [23] 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.
- [24] 韩杏杏, 陈杰, 王海洋, 等. 基于随机森林模型的耕地表层土壤有机质含量空间预测——以河南省辉县市为例[J]. 土壤, 2019, 51(1): 152–159.
- [25] 姜赛平, 张怀志, 张认连, 等. 基于三种空间预测模型的海南岛土壤有机质空间分布研究[J]. 土壤学报, 2018, 55(4): 1007–1017.
- [26] 卢宏亮, 赵明松, 刘斌寅, 等. 基于随机森林模型的安徽省土壤属性空间分布预测[J]. 土壤, 2019, 51(3): 602–608.
- [27] 袁玉琦, 陈翰阅, 张黎明, 等. 基于多变量与 RF 算法的耕地土壤有机碳空间预测研究——以福建亚热带复杂地貌区为例[J]. 土壤学报, 2021, 58(4): 887–899.
- [28] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.
- [29] Efron B, Tibshirani R J. Introduction//An Introduction to the Bootstrap[M]. Boston, MA: Springer US, 1993: 1–9.
- [30] 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.