

DOI: 10.11766/trxb202103120140

魏宇宸, 卢晓丽, 朱昌达, 张秀秀, 潘剑君. 基于地形与遥感辅助信息的小流域尺度高分辨率有机碳空间分布预测研究[J]. 土壤学报, 2023, 60 (1): 63–76.

WEI Yuchen, LU Xiaoli, ZHU Changda, ZHANG Xiuxiu, PAN Jianjun. High-resolution Digital Mapping of Soil Organic Carbon at Small Watershed Scale Using Landform Element Classification and Assisted Remote Sensing Information[J]. Acta Pedologica Sinica, 2023, 60 (1): 63–76.

基于地形与遥感辅助信息的小流域尺度高分辨率有机碳空间分布预测研究*

魏宇宸, 卢晓丽, 朱昌达, 张秀秀, 潘剑君[†]

(南京农业大学资源与环境科学学院, 南京 210095)

摘要: 土壤有机碳 (Soil organic carbon, SOC) 既是衡量土壤质量的重要指标, 也是影响全球碳氮循环的关键因素之一。作为数字土壤制图 (Digital soil mapping, DSM) 研究中起主要作用的环境变量, 地形元素在 SOC 预测制图中也是无可替代的。应用机器学习模型, 通过引入不同超参数设置下获得的高分辨率 (5 m) Geomorphons (GM) 地形分类图作为丘陵地形特征信息的补充, 结合数字高程模型 (Digital elevation model, DEM) 衍生变量和光学、合成孔径雷达 (Synthetic aperture radar, SAR) 遥感数据对句容市黄梅镇北部小流域尺度 (1: 25 000) 丘陵地貌区地表层 SOC 含量进行预测制图, 并评估不同 GM 变量在 SOC 含量预测中的表现。基于 74 个土壤样本和不同环境变量组合, 分别采用袋装决策回归树 (Bagged classification and regression tree, Bagged CART)、随机森林 (Random forest, RF) 和立体派 (Cubist) 三种方法构建 SOC 含量预测模型, 并通过四个精度验证指标, 采用十折交叉验证对生成的模型性能进行分析评价。总体上, Cubist 模型的预测表现优于 Bagged CART 和 RF 模型。分析显示, 与单独使用 DEM 衍生变量相比, 引入 GM 变量能提供更准确的 SOC 含量预测, 其中设置 20 像元 (cells) 搜索半径 (L) 与 5° 平坦度阈值 (t) 的 GM 变量表现出最高的模型贡献度, 两者与遥感类变量的组合产生了最高的预测精度 ($R^2=0.53$)。引入 GM 变量后, 使用 Cubist 模型估算 SOC 含量的 R^2 提高了 14.3%。研究表明, 在小流域尺度丘陵地貌区, 地形类变量是 SOC 预测的主要解释变量, 其中谷底平坦综合指数 (Multi-resolution index of valley bottom flatness, MRVBF) 和高程是模型中最重要的两个环境变量; 同时, 在建立 SOC 预测模型时, 高分辨率 GM 图像具有作为输入环境变量的应用潜力。

关键词: 土壤有机碳; 地形元素; 小流域尺度; 空间分布; 机器学习

中图分类号: S15 文献标志码: A

High-resolution Digital Mapping of Soil Organic Carbon at Small Watershed Scale Using Landform Element Classification and Assisted Remote Sensing Information

WEI Yuchen, LU Xiaoli, ZHU Changda, ZHANG Xiuxiu, PAN Jianjun[†]

* 国家自然科学基金项目 (41971057, 41771247) 资助 Supported by the National Natural Science Foundation of China (Nos. 41971057, 41771247)

[†] 通讯作者 Corresponding author, E-mail: jpan@njau.edu.cn

作者简介: 魏宇宸 (1997—), 男, 安徽芜湖人, 硕士研究生, 主要从事土壤属性制图研究。E-mail: weiywuhu@163.com

收稿日期: 2021-03-12; 收到修改稿日期: 2021-08-29; 网络首发日期 (www.cnki.net): 2021-12-22

(College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China)

Abstract: 【Objective】 Soil organic carbon (SOC) is an important indicator of soil fertility and plays a fundamental role in the terrestrial ecosystem carbon cycle. As one of the primary environmental factors in digital soil mapping (DSM), landform elements are irreplaceable in predicting SOC. The purpose of this study was to simulate the complex and nonlinear relationship between SOC and environmental variables and evaluate the importance of each variable to accuracy in SOC mapping. 【Method】 We applied machine learning techniques to map SOC content in a small watershed (1: 25000) of Huangmei Town, Jurong City using high-resolution landform elements classification maps known as geomorphons, digital elevation model (DEM) derivatives, optical and synthetic aperture radar (SAR) remote sensing data. The performance of all geomorphon (GM) variables under different hyperparameter settings was evaluated to predict SOC content. Three machine-learners including bagged classification and regression tree (Bagged CART), random forest (RF) and Cubist were used to construct predictive models of SOC content based on 74 soil samples and different combinations of environmental covariates. Model A, Model B, and Model D included only GM variables, DEM derivatives, and remote sensing variables, respectively. Model B was a combination of GM data and DEM derivatives, while Model E included all predictor variables. The performance of these models was evaluated based on a 10-fold cross-validation method by four statistical indicators. Concordance index (C-index), root mean square errors (RMSE), bias and coefficient of determination (R^2) of the three models were worked out for evaluation of the accuracy of their predictions. The best model was screening-out for mapping SOC in the study area based on the raster datasets of all environmental variables. 【Result】 Overall, the Cubist model performed better than RF and Bagged CART, and these models yielded similar spatial distribution patterns of SOC, i.e. an ascending trend from the northern hilly area to the southern flatter land of the study area. Our results showed that more accurate predictions of SOC content were provided with the introduction of GM variables than individual DEM derivatives. The GM map with 20 cells search radius (L) and 5° flatness threshold (t) showed the highest relative importance within four GM variables in three models. The Cubist-E model that functioned based on GM landform elements classification variables, DEM derivatives and remote sensing variables was much better than the others in performance and could explain most of the spatial heterogeneity of SOC ($R^2 = 0.53$). Also, the prediction accuracy changed with and without the GM predictors with the R^2 for estimating SOC content using the Cubist model increasing by 14.3%. The SOC contents of the hilly region predicted with the Cubist-E model ranged from 5.65 to 13.31 $\text{g}\cdot\text{kg}^{-1}$. In addition, topographic variables were the main explanatory variables for SOC predictions and the multi-resolution index of valley bottom flatness (MRVBF) and elevation were assigned as the two most important variables. 【Conclusion】 The Cubist model that functions based on GM variables, DEM derivatives, as well as remote sensing variables, is a promising approach to predicting the spatial distribution of SOC in hilly regions at a small watershed scale. The results of this study illustrate the potential of GM landform elements classification data as input when developing SOC prediction models.

Key words: Soil organic carbon; Landform elements; Small watershed scale; Spatial distribution; Machine learning

土壤作为陆地生态系统最重要的碳氮库之一, 在全球碳氮循环中发挥着关键作用^[1]。基于土壤有机碳 (Soil organic carbon, SOC) 在表征土壤质量和肥力水平等土壤属性的重要性, 在小流域尺度上对 SOC 的空间变异性进行正确的评估对于改进土壤管理措施、改善生态环境质量及维护粮食安全等方面具有重要的意义^[2]。通过传统的野外土壤调查获取这些土壤信息不仅费时费力且成本高昂^[3], 因此在一定区域内通过离散样本预测土壤属性和类

别的 DSM 技术, 由于其在一定程度上能减少采样和分析成本而得到广泛的认可和应用^[4]。

基于土壤属性与环境预测变量之间的相关性, 数字土壤制图 (Digital soil mapping, DSM) 技术通过邻域分析和数学建模方法对土壤属性与周围环境之间的关系进行量化分析并预测评估土壤-景观连续体的空间分布格局^[5]。目前使用的量化方法主要有三类: 不考虑确定性趋势的地统计学模型, 如普通克里格法; 明确随机空间相关变化并拥有确定趋

势的混合方法，如回归克里格法；并未明确考虑随机空间相关变化的基于特征空间的机器学习模型。其中，机器学习模型由于包含算法较多且自适应性较强，被广泛应用于不同尺度的 SOC 预测制图，例如多元线性回归（Multiple linear regression, MLR）^[6]、支持向量机（Support vector machine, SVM）^[7]以及随机森林（Random forest, RF）^[8]等。尽管目前并不缺乏关于 DSM 应用模型比较研究，但是如何为小流域尺度特定地理景观选择最佳的预测模型依然是 DSM 研究的一个挑战。

研究土壤属性空间分布情况首先需要确定不同来源的环境变量以及适宜的预测模型。常用的环境变量有 DEM 及其衍生变量以及遥感图像不同波段及其组合数据，年平均降雨量、积温分布等气候变量以及母质与土地利用情况等因素也可纳入模型预测^[9]，而高程变量以及提取自 DEM 的其他地形变量则更是包括 SOC 在内的土壤属性预测制图中不可或缺的重要环境变量，对于最终的预测及制图精度有重要影响^[10]。针对较大尺度上的 DSM 研究，有很多现有的数字高程模型（Digital elevation model, DEM）数据集可供使用（例如 Shuttle radar topography mission DEM 和 Aster DEM）^[11]，但是其 90 m 和 30 m 精度数据并不适合在小流域尺度上进行高精度的 DSM 研究。

数字地形模型（Digital terrain model, DTM）一般建立在数字摄影测量和激光扫描获得的 DEM 行数据的基础上，而在包括小流域尺度在内的多重尺度上，研究不同地形特征与土壤属性之间的内在关联也是 DSM 研究的重点之一^[12]。DEM 分辨率直接影响所生成的地形模型及其衍生变量的精确度，高精度的 DEM 数据可以更好地服务于土壤属性的预测制图^[13]。20 世纪 80 年代以来开发的许多用于精确地貌分类的自动化模型中^[14]，尽管在地形变量及地貌分类等方面因采用不同的模型算法而有不同的选择，但其中的多数模型均存在一个共同假设，即存在可比较的相邻地形要素且均具有在 GIS 环境中可识别的特征，例如曲率、坡度和表面粗糙度^[15]。Jasiewicz 和 Stepinski^[16]于 2013 年提出了一种名为 Geomorphons（GM）的基于像元进行地形分类的新方法，即根据高程值在指定尺度邻域搜索窗口内的相对差值来定义其空间位置所对应的地形元素类型。相比其他地形分类方法，GM 地形分类法有两

个特征：可以更加高效地利用计算机视觉工具获得地学意义明确的分类结果；能够有效避免因地形特征属性选择的主观性所导致的分类结果的不确定性，且更适合局部地形特征的识别^[17]。

近年来，GM 地形分类法在国内外地貌及土壤景观分类等领域得到了较为广泛的应用^[13]，但将 GM 地形分类图作为主要预测变量的土壤属性预测制图研究则非常稀少。本文对 GM 多参数设置的比较，包括小流域尺度 5 m 分辨率高精度 GM 制图对搜索半径参数 L 的要求以及平坦度参数 t 在丘陵地形的应用潜力进行了有意义的探讨，对其在中国东部更大尺度上的丘陵农区 DSM 的适应性研究提出了可供参考的方法，开拓了 GM 地形分类图在 SOC 预测制图方面的研究前景。根据以往的研究结论，相比平原地区，DTM 的地貌分类在有明显地形变化（导致 DEM 内部空间异质性较为显著）的区域更为有效^[18]。此外在小流域尺度上地形变化对土壤形成发育和空间分布的异质性有强烈影响，但在该尺度内的研究也相对较少^[13]。由于 GM 法基于像元的识别机制，较大的高程差异会干扰特殊土壤类型及其属性值空间分布（如高山和非高山土壤）的准确预测，而在丘陵地形中应用该方法可以减少此种误分类的发生。因此，本文将不同超参数设置下获得的不同 GM 地形分类结果作为环境变量参与到 SOC 的预测制图中，并与传统 DEM 衍生变量以及光学、合成孔径雷达遥感数据的不同组合的预测结果相比较，评价其在三种不同机器学习算法中的预测效果及最终的预测精度。

1 材料与方法

1.1 研究区概况

研究区地处长江三角洲，位于江苏省句容市黄梅镇北部区域（32°3'20"—32°5'50" N，119°10'20"—119°12'30" E），面积为 5.37 km²（图 1a）。区域平均相对高程约为 82 m，最大高程差约为 205 m，坡度的最大值、平均值分别为 69°、12°。地属北亚热带湿润气候，夏季炎热多雨，冬季寒冷干燥，年平均气温为 15.2℃，年平均干旱度约为 1.0，年均降水量为 1 060 mm。

研究区内丘陵地形突出，特别是北部区域（图 1b），而旱作和水田农业（主要是小麦，稻米和其他经济

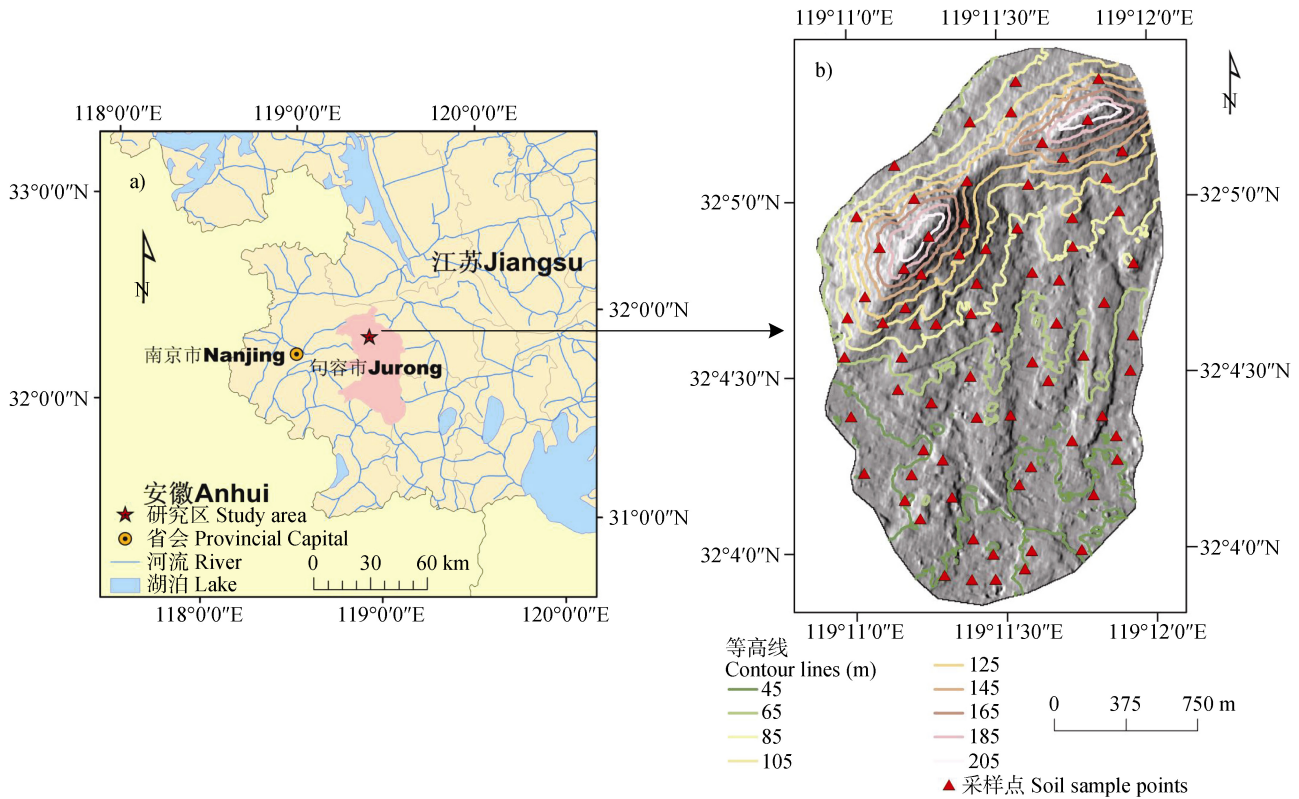


图1 研究区位置 (a)、20 m 间距等高线及采样点分布 (b)

Fig. 1 Research area within eastern China (a); field sample points on the site (b), depicted on 20 m contours

作物) 主要分布在南部相对平坦的地形区域。根据中国土壤系统分类检索^[19], 共检索出人为土、淋溶土、锥形土和新成土四个土纲。由于当地有长期的水稻种植历史, 水耕人为土广泛分布在南部农耕地, 而北部丘陵则以淋溶土为主。

1.2 土壤样品采集与分析

采样设计以景观分类为基础, 综合与土壤形成过程密切相关的环境信息, 以地形数据为条件变量, 同时考虑从遥感图像获得的土地覆盖数据和通过实地调查获得的土壤母质数据。土壤采样布局主要基于以下原则: (1) 调查路线覆盖研究区主要景观类型并结合土壤景观格局的空间分布特征进行均匀布设, 研究区北部以有林地等自然景观为主, 而南部以水田、旱地等农田景观为主, 在采样设计上考虑到南北差异并尽可能均匀覆盖。(2) 采样点数量根据研究区主要景观类型及各景观内部的异质性程度确定, 如研究区北部大面积覆盖天然有林地和灌木林地, 异质性程度低, 主要剖面点的数量可相对较少; 而南部间杂分布有水田、旱地和果园等差异较大的景观类型, 因而适当增加剖面点的设置。(3)

采样点布局应与研究区景观的尺度相适应, 即服务于采样设计的景观分类结果应根据不同的空间尺度或制图比例尺的要求加以确定。根据以上原则, 研究区内共存在 7 种主要景观类型: 北部区域有丘陵天然有林地景观、丘陵灌木林地景观和丘陵人工有林地景观; 南部区域有低岗果园景观、低岗茶园景观、平地水田景观和平地旱地景观。在主要景观类型内还可根据坡度(如缓坡 $2^{\circ}\sim 6^{\circ}$ 、中缓坡 $6^{\circ}\sim 15^{\circ}$ 和中坡 $15^{\circ}\sim 25^{\circ}$) 和母质类型(如黄土母质、坡积物母质和残积物母质) 进行进一步的细分, 供布设土壤钻孔补充样点时参考。

为尽可能表征研究区内的景观异质性, 共设计了 6 条采样路线, 沿路线主径共采集 25 个土壤剖面样本, 对表层土壤 (0~20 cm) 采用全层柱状连续采样, 并在主要剖面点周围分散采集 49 个土壤钻孔样本, 共采集 74 个表层土壤样本 (图 1b)。样本自然风干后采用重铬酸钾容量法测定 SOC 含量。

1.3 遥感变量及其处理

用于建模的遥感数据包括从中国资源卫星应用中心 (<http://www.cresda.com/CN/>) 下载的高分二

号 (GF-2) 和高分三号 (GF-3) 影像。GF-2 拥有四个波段: B2 (0.45~0.52 μm)、B3 (0.52~0.59 μm)、B4 (0.63~0.69 μm) 和 B5 (0.77~0.89 μm)，成像时间选择 2019 年 9 月 19 日，无云遮盖，该时段研究区地表植被有较高的绿度覆盖，适宜提取植被指数。SAR 影像数据为 GF-3 标准条带影像，采用双极化，幅宽 130 km，空间分辨率 25 m，具有较高的植被穿透能力，成像时间为 2020 年 1 月 19 日且无云遮盖，在冬季缺少植被覆盖的情况下对裸土信息有较好的反映。

使用 ENVI 5.3 软件对 GF-2 图像进行预处理，包括辐射定标、正射校正和大气校正 (FLAASH 大气模型)^[20]。使用 PIE-SAR 6.0 对 GF-3 数据进行了以下预处理: 多视、共配准、斑点滤波 (13×13 窗口的 Lee 滤波器^[21])、地理编码和辐射定标，灰度值转换为分辨率为 25 m 的分贝标度后向散射系数。提取了 GF-2 影像的多光谱波段用于后续研究，并根据已有的波段范围计算 2 个植被光谱指数 NDVI (Normalized difference vegetation index) 和 ARI2 (Anthocyanin reflectance index 2) 作为预测环境变量，有研究显示它们与 SOC 密切相关^[22]。此外 GF-3 影像的 VH 和 VV 偏振后向散射系数也作为环境变量进行计算。

1.4 GM 地形分类及 DEM 衍生变量

使用 GRASS GIS (v 7.8) 中的 r.geomorphon 扩展工具^[17]生成 GM 模型，通过将中心像元与其 8 个相邻像元 (自初始像元向东) 进行比较而生成 8 元组模式，并继续沿逆时针方向生成三元运算符。在元组中，围绕中心像元的相对度量 (较高, 较低, 相等) 的模式分别描述为 “+” “-” 和 “0”。中心像元的直接邻位须遵守视界线 (Line of sight, LOS) 原则，即沿八个方向使用天顶角和天底角来判别地形要素并计算水平距离。最后，使用三元运算符结合 LOS 原则搜索高程变化，共生成 10 种独特的地貌类型: 平区、山峰、山脊、山肩、凸背坡、直背坡、洼地、沟谷、山脚和凹背坡^[16]。

地形特征识别过程中最重要的两个参数是搜索半径 (L) 和平坦度阈值 (t)，前者代表计算天顶角和天底角的最大距离，后者为视作水平区域的坡度阈值大小，由参考方向和搜索距离确定。应用较大的 L 值相当于以更高更宽的视角进行地形分类，并在多个尺度上同时生成地形特征模型；而较小的 L

值从局部角度进行分类，可正确识别小于 L 的二阶地形特征。在 GM 方法中使用了 5 个 L 设置 (5、10、20、30 和 40 cells)，同时考虑到丘陵地形较大的坡度变化，测试了 5 个不同的 t 值 (1°, 2°, 3°, 4° 和 5°) 以更好地适应研究区的地形起伏状况。此外对原地形分类图进行聚类后生成 GM (group) 变量，将原本的 10 种地形元素根据边界相邻且地类相似原则聚类为 4 种，分别为坡顶部 (包括原山峰、山脊、山肩)、坡中部 (包括原凸背坡、直背坡、凹背坡)、坡麓部 (包括原平区、山脚) 和低凹部 (包括原沟谷、洼地)，GM 标准地形分类图与 GM 聚类图如图 2 所示。

研究区 5 m 分辨率 DEM 通过摄影测量技术 (ENVI 5.3.1 软件 DEM 提取模块) 提取自资源三号 02 星 (ZY3-02) 前后视立体像对，过程采用 46 个地面控制点文件，生成结果经 100 个呈网格分布的实测高程值验证，精度误差为 1.59 m，符合 5 m DEM 精度要求。在生成的 DEM 的基础上，使用 ArcGIS 10.6 和 SAGA GIS 软件提取 8 个 DEM 衍生变量 (表 1) 与 GM 地形分类变量作比较，共同参与 SOC 的预测与制图。建模使用的环境变量包括 GM 地形分类图、传统 DEM 衍生变量以及光学、SAR 遥感变量，具体构成见表 1。

1.5 模型的选择及评价

采用袋装决策回归树 (Bagged CART)、随机森林 (Random forest, RF) 和立体派 (Cubist) 三种机器学习模型进行 SOC 的预测。

CART 是用于回归及分类问题的非参数数据挖掘技术，通过对训练数据集 (包括目标变量和测试变量) 进行循环二分，最终形成以二叉树为主要形式的决策树结构^[23]。Bagged CART 是一种改进后的 CART 算法，它将 CART 与 bagging 技术相结合，以提高预测模型的性能并减少过拟合发生的可能^[24]。

RF 模型同样是从 CART 发展而来的一种基于树状结构的集成机器学习技术，应用于数据分类与回归分析。RF 算法会产生大量的树和节点，而来自原始训练数据的唯一引导样本 (可替换) 会独立地构建所有树状结构中的每棵树，通过使用自举采样降低对过拟合的敏感性，并允许基于剩余测试集 (袋外样本) 估计一般误差^[25]。RF 模型通过 R 软件^[26] 中的 randomForest 包生成。

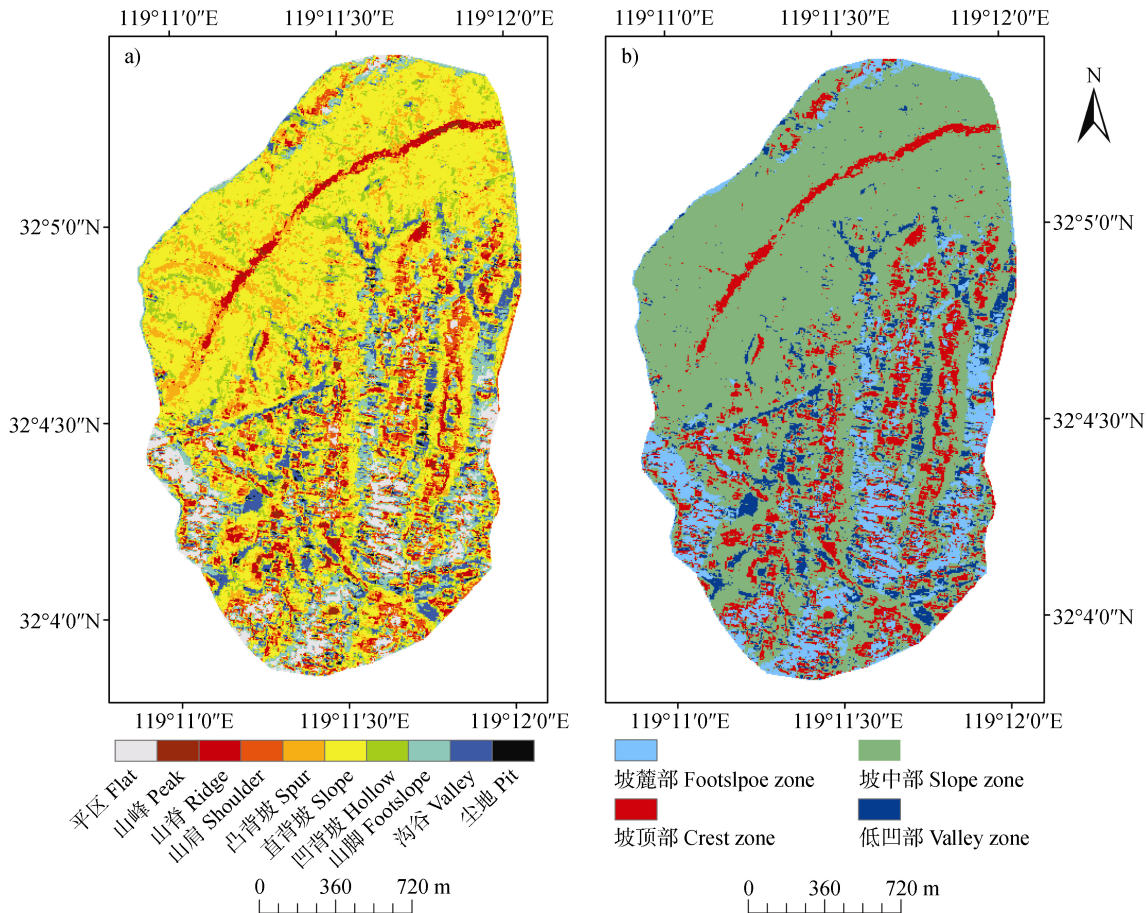


图 2 GM 标准地形分类 (20L, 5t) (a) 与 GM 地形聚类结果 (b) 示意图

Fig. 2 Images of GM standard map (20L, 5t) (a) and GM aggregation map (b)

Cubist 是另一种先进的非参数回归树算法, 适合处理 SOC 和预测变量之间的非线性关系^[27]。Cubist 建立了由多组规则组成的多元模型, 并根据这些规则选择适宜的预测模型, 也可以通过生成多个基于规则的模型 (称为 committees) 来改进预测, 而过程中生成的每一个新模型均会修正先前模型的预测, 从而使一般误差最小化。此外, Cubist 能够将环境变量重要性计算为变量对模型精度的相对贡献^[28]。Cubist 模型由 Cubist 包生成。

上述模型均通过 R 软件中的 caret 包进行参数的调节与优化。

1.6 统计分析

使用 SPSS 25 软件对 SOC 进行描述性统计分析, 并通过 Pearson 相关分析检测环境变量之间的共线性及其与 SOC 的相关性, 从模型中删除具有较高方差膨胀因子 (VIF ≥ 10) 的高度相关预测变量 ($r \geq 0.8$)^[29]。利用十折交叉验证方法评估模型的预测

性能, 通过一致性指数 (Concordance index, C-index)、均方根误差 (Root mean square error, RMSE)、偏差 (bias) 和决定系数 (R^2) 四个指标验证不同模型的表现。

为了评估预测不确定性, 从预测结果中选择在三种机器学习算法中表现最佳的模型生成 100 张 SOC 预测图逐像元计算标准差 (Standard Deviation, SD)^[30], 并计算均值作为最终的 SOC 预测图。

2 结果

2.1 有机碳含量统计特征

研究区内的 SOC 含量最小值为 $4.70 \text{ g}\cdot\text{kg}^{-1}$, 最大值为 $15.55 \text{ g}\cdot\text{kg}^{-1}$, 平均值为 $7.81 \text{ g}\cdot\text{kg}^{-1}$, 偏度系数为 1.29, 显示出 SOC 含量的偏斜分布。对 SOC 含量进行自然对数转换, 使偏度系数降低为 0.75, 且 SD 值低于平均值。通过共线性分析发现一部分环境

变量存在共线性。为避免该情况的发生，按照 VIF 值小于 10 的标准将环境变量总数减为 18 个，其中

GM 变量在满足共线性要求的基础上选择具有代表性且参数差异较大的 4 种变量作为预测变量。

表 1 环境变量的构成

Table 1 The composition of environmental variables

变量类型 Categories of variables	指标 Indexes	缩写 Abbreviation	原始分辨率 Native resolution/m	数据类型 Type
GM 地形分类 Geomorphons classification	5 cells L 与 $5^\circ t$ 设置下的 GM	GM ($5 L, 5 t$)	5	类别变量 Category
	20 cells L 与 $5^\circ t$ 设置下的 GM	GM ($20 L, 5 t$)	5	
	20 cells L 与 $1^\circ t$ 设置下的 GM	GM ($20 L, 1 t$)	5	
DEM 衍生变量 DEM derivates	聚类后的 GM	GM (group)	5	连续变量 Quantitative
	高程	DEM	5	
	收敛指数	CI	5	
	坡长因子	LSF	5	
	地表凸度	TSC	5	
	谷底平坦综合指数	MRVBF	5	
	总集水面积	TCA	5	
光学、SAR 遥感变量 Optical and SAR imagery	谷深度	VD	5	连续变量 Quantitative
	地形湿度指数	TWI	5	
	GF-2 光谱波段 2	B2	3.2	
	GF-2 光谱波段 5	B5	3.2	
	归一化植被指数	NDVI	3.2	
	花青素反射指数 2	ARI2	3.2	
	GF-3 VH 极化后向散射系数	VH	25	
GF-3 VV 极化后向散射系数	VV	25		

2.2 模型的评价与比较

基于使用的三种机器学习模型，通过环境变量的不同组合构建了小流域尺度 SOC 含量的预测模型，具体变量组合见表 2。基于不同变量组合的 Bagged CART、RF 和 Cubist 模型在预测 SOC 含量方面的性能如表 3 所示。基于模型拟合精度的比较分析表明，预测建模方法的选择和预测变量的类型（由不同环境变量构成的模组）对 SOC 值的预测效果有一定的影响。例如，对于模组 A，Bagged CART、RF 和 Cubist 模型表现出了相似的预测精度（ R^2 分别为 0.27、0.28 和 0.28），而对于模组 C 和模组 E，三种模型均显示出较为明显的差异。总体而言，三种模型针对同一模组的表现没有显著差异，其中 Cubist 模型表现出了最好的预测性能，表明不同的预测模型可能适合不同的环境变量和土壤性质。

表 2 环境变量的不同组合

Table 2 Different combinations of environmental variables

序号 No.	模组 Model	环境变量 Environmental variables
i	模组 A	GM 地形分类
ii	模组 B	DEM 衍生变量
iii	模组 C	GM 地形分类 + DEM 衍生变量
iv	模组 D	光学、SAR 遥感变量
v	模组 E	GM 地形分类 + DEM 衍生变量 + 光学、SAR 遥感变量

三种机器学习模型的 DEM 衍生变量的预测精度均高于单一类型的遥感变量或 GM 变量。对于 GM 变量而言，由于预测变量数量较少，在 CART 模型中的预测精度略低于遥感变量，在 RF 和 Cubist 模

型中差距较大。尽管在所有不同预测模型和环境变量类型之间的预测性能均存在一定差异,但与使用单一类型的地形变量相比,DEM 衍生变量和 GM 变量的组合提高了预测的精度。例如,在 DEM 衍生变量的基础上添加 GM 变量在预测 SOC 值时将 Cubist 模型的 R^2 从 0.42 提高到 0.48,表明 GM 数据包含可以提高总体预测性能的未利用地形信息,而 CART 和 RF 模型的精度提高也验证了这一点。

三种类型的环境变量共同参与预测时可以获得

最高的精度。应用 Cubist 模型获得的 R^2 与应用地形环境变量时相比(从 0.48 到 0.53)提高了 10.4%,对于其他两种预测模型也可以观察到类似的改善。这表明从多源传感器中提取的遥感变量对于土壤特性的有效建模和预测具有一定的价值。除偏差(bias)外,模组 E 的其余三项指标在三种预测模型中均有最高精确度,其中以 Cubist 模型表现最佳(C-index = 0.55, RMSE = 0.98, $R^2 = 0.53$), R^2 值表明该模型可以解释研究区内 SOC 值变异的大约 53%。

表 3 基于三种模型 SOC 模拟在不同变量模组间的精度对比

Table 3 Comparison of prediction accuracies based on three modelling techniques using different combinations of predictors

预测模型 Modelling techniques	模组 Model	C-index	RMSE	Bias	R^2
Bagged CART	A	0.31	1.51	-0.04	0.27
	B	0.43	1.47	-0.02	0.38
	C	0.48	1.29	-0.03	0.43
	D	0.45	1.45	-0.03	0.31
	E	0.49	1.25	-0.01	0.47
RF	A	0.30	1.59	-0.03	0.28
	B	0.34	1.43	-0.02	0.37
	C	0.50	1.38	-0.02	0.45
	D	0.43	1.67	-0.03	0.35
	E	0.52	1.22	-0.02	0.51
Cubist	A	0.37	1.33	-0.15	0.28
	B	0.46	1.20	-0.10	0.42
	C	0.52	1.15	-0.08	0.48
	D	0.47	1.32	-0.16	0.36
	E	0.55	0.98	-0.05	0.53

2.3 环境变量的相对重要性

对于使用模组 E 的 SOC 预测制图,按相对重要性高低进行排序的环境变量排名如图 3 所示(重要性程度被转换为百分比值)。在三种预测模型中不同类型的变量重要性略有不同,表明在这些模型中占主导地位的环境特征也有所不同。DEM 衍生变量在三种模型中均是主要的解释变量,其次是遥感变量,GM 变量所占的相对重要性百分比最低。尽管细分后各环境变量在不同模型中的排列特征有差异,但 DEM 衍生变量组中的 MRVBF 以及高程变量在三种模型中的相对重要性始终占据第一和第二的位置。

此外,GM 变量在 CART、RF 和 Cubist 模型中分别解释了 SOC 值变化的 15%、16%和 18%,表明 GM 地形分类变量具有在该研究区内对 SOC 值进行预测制图的应用潜力。

2.4 土壤有机碳空间分布

基于模组 E,通过三种机器学习方法预测的 SOC 空间分布如图 4 所示。研究区内 SOC 预测值的平均值和标准差(SD)分列如下: CART 模型为 $7.40 \text{ g}\cdot\text{kg}^{-1}$ 和 $1.62 \text{ g}\cdot\text{kg}^{-1}$, RF 模型为 $7.43 \text{ g}\cdot\text{kg}^{-1}$ 和 $1.65 \text{ g}\cdot\text{kg}^{-1}$, Cubist 模型为 $7.37 \text{ g}\cdot\text{kg}^{-1}$ 和 $1.64 \text{ g}\cdot\text{kg}^{-1}$ 。基于建模精度分析,选择 Cubist 方法在模组 E 中运行

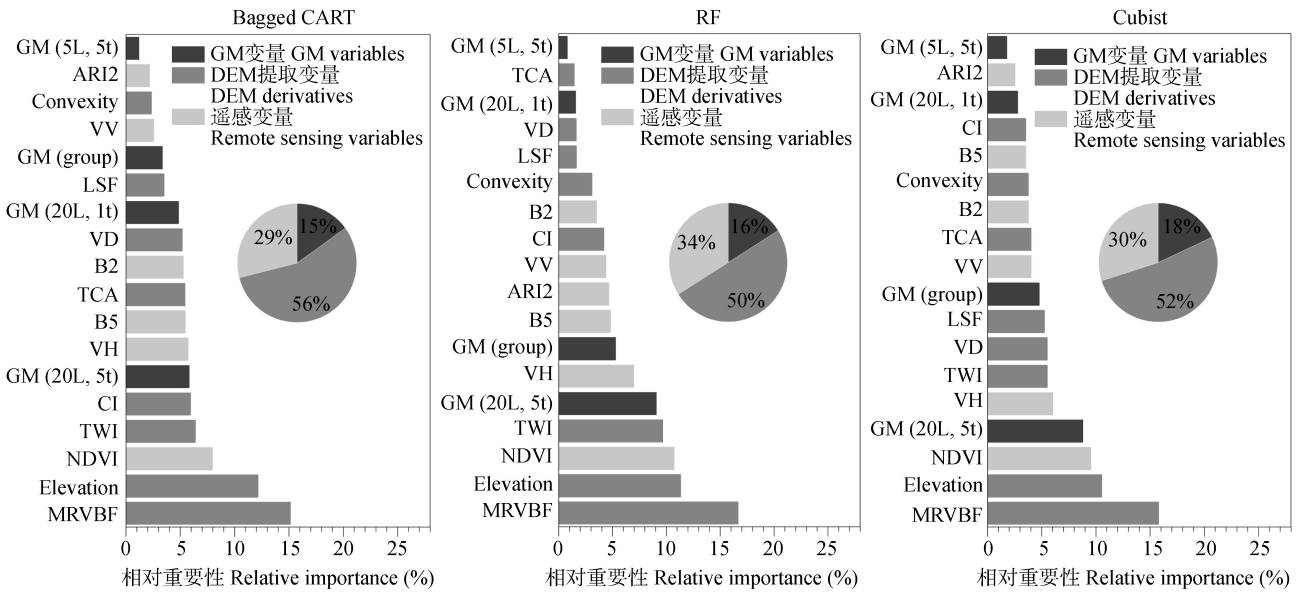


图 3 模组 E 中的环境变量对于 SOC 预测的相对重要性示意图 (模组 E: GM 变量+DEM 衍生变量+遥感变量)
 Fig. 3 The relative importance of predictor variables in Model E for predicting SOC (Model E: GM variables + DEM derivatives + Remote sensing variables)

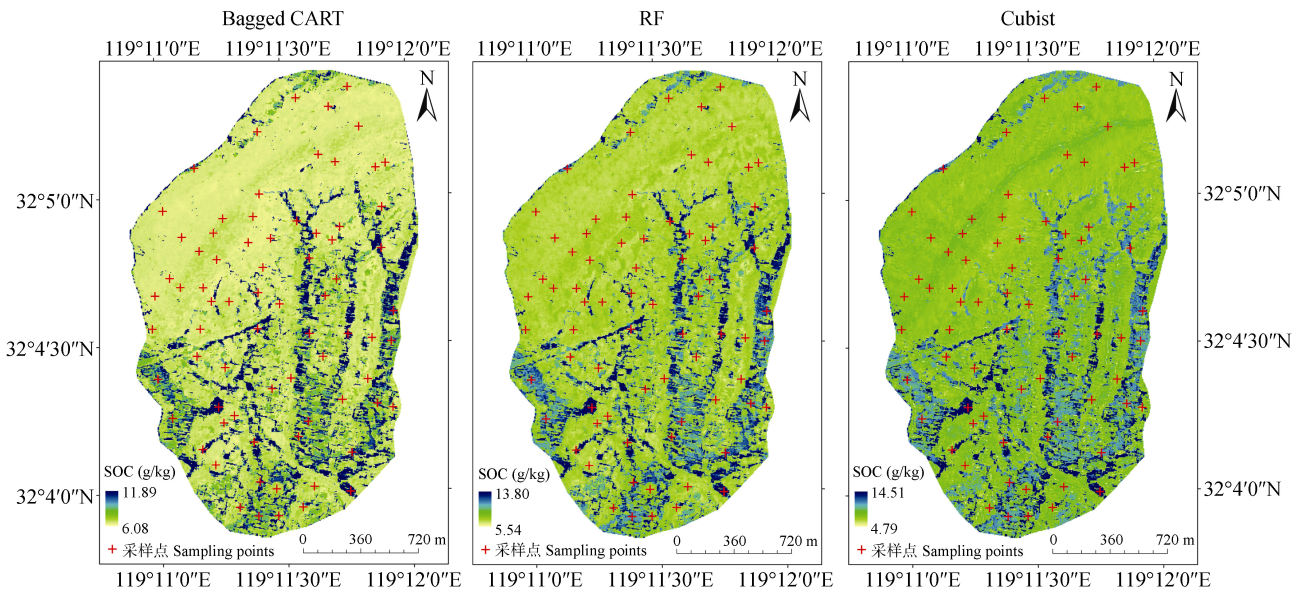


图 4 使用 Bagged CART、RF 和 Cubist 模型在模组 E 中对 SOC 值进行空间预测制图 (模组 E: GM 变量+DEM 衍生变量+遥感变量)
 Fig. 4 Maps of SOC predicted in Model E using Bagged CART, RF and Cubist (Model E: GM variables + DEM derivatives + Remote sensing variables)

100 次以评估预测不确定性，并获得这 100 次运行的平均值图 (最终制图) 和 SD 值图 (预测不确定性) (图 5)。经过 100 次运行后的 SOC 平均值为 $7.34 \text{ g}\cdot\text{kg}^{-1}$ ，而该预测模型也显示出较低的不确定性，对于 100 个预测的 SOC 分布图，平均 SD 值为

$0.03 \text{ g}\cdot\text{kg}^{-1}$ ，表明 Cubist 具有较为稳定的预测能力。尽管不同模型的总体分布特征相似，但 SOC 含量的高低却也有较为明显的差异。由 Bagged CART 模型生成的 SOC 分布图总体颜色较浅，预测所得 SOC 最大值仅为 $11.89 \text{ g}\cdot\text{kg}^{-1}$ ，为三种模型中最低，

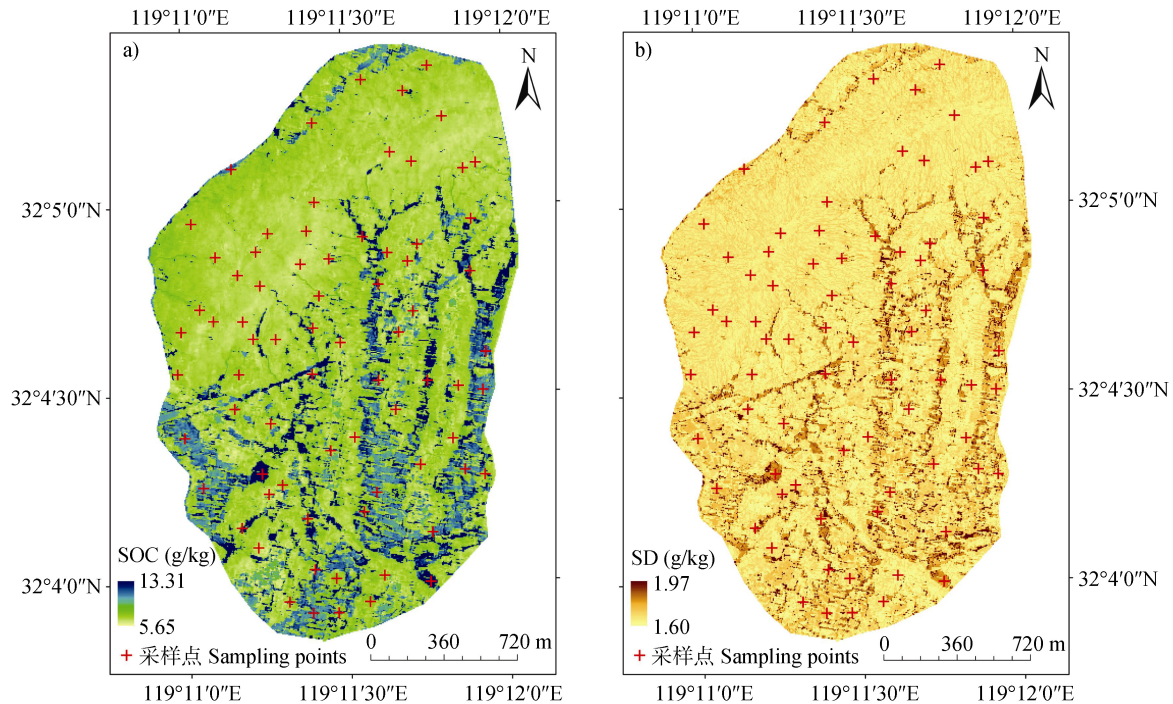


图5 以 Cubist 方法在模组 E 中运行 100 次获得的 SOC 平均值图 (a) 和标准差 (SD) 图 (b) (模组 E: GM 变量+DEM 衍生变量+遥感变量)。

Fig. 5 Mean SOC map obtained from Model E based on 100 runs of the Cubist model and their corresponding standard deviation map (Model E: GM variables + DEM derivatives + Remote sensing variables).

同时区间范围最窄。RF 模型相比前者略高, 最大值为 $13.80 \text{ g}\cdot\text{kg}^{-1}$, 而 Cubist 模型拥有最高的 SOC 预测含量 $14.51 \text{ g}\cdot\text{kg}^{-1}$, 同时拥有最宽的预测区间。

3 讨论

3.1 三种预测模型中不同类型环境变量的表现

对预测精度的比较分析表明, 机器学习方法的选择以及环境变量的类型和组合对于 SOC 的预测性能有很大的影响 (表 3)。总体而言, Cubist 在丘陵区 SOC 预测中的表现优于 RF 和 Bagged CART 模型。Fathololoumi 等^[31]在伊朗北部地形起伏剧烈的山区应用了 RF 和 Cubist 模型进行了 SOC 等土壤属性的预测制图, 结果表明预测模型的选择对于预测精度有着重要的影响, 而 Cubist 模型拥有更高的精度, 在非平坦区域表现出较好的适用性。在另一个小流域尺度研究中, Libohova 等^[32]对不同分辨率下的 GM 地形分类进行了精度验证, 发现不同超参数设置下的 GM 变量和机器学习方法的选择对土壤样点的分类精度也有影响。

精度分析结果表明, GM 地形分类数据、DEM 衍生数据和光学、SAR 遥感数据对研究区 SOC 的有效建模具有重要意义。GM 数据和传统 DEM 衍生数据的组合可以使模型的预测性能得以改进。此前的一些研究已经证明了 GM 数据对于土壤属性预测的有效性^[13], 但从已有的小流域尺度 DSM 研究来看, 高精度地形分类数据的应用仍然没有获得足够的关注, 且缺乏 GM 地形分类在 SOC 预测制图中可供比较的应用研究。Ashtekar 等^[33]在南美洲东部平原利用 GM 数据通过模糊逻辑模型进行了大尺度的 SOC、pH 等土壤属性空间分布预测研究, 但其预测过程发生了一定程度的欠拟合, 导致预测变量对 SOC 空间变异的解释程度低于 50% (文中缺乏具体的 R^2 描述), 未能提供可供分析的精度指标。Flynn 等^[34]测试了不同 GM 数据对于土壤质地、土层深度、电导率等五项数据的预测性能, 其中 GM 变量最低能独立解释电导率值变异的 10%, 对砂粒含量则最高能解释 43%, 表现出较为优异的预测性能, 但该研究并未完成实际的预测制图工作且未将其与常用的环境变量作比较并评价其精度表现, 此外由于所

处的区域地形平坦，并不能完全发挥 GM 地形分类数据作为环境变量的应用价值。

模组 E 在 Cubist 模型中解释了 SOC 空间变异的 48%，但由于国内针对小流域尺度丘陵地形的 SOC 预测制图研究案例较少，无法得出有效的精度评价结果。赖雨晴等^[35]比较了人工神经网络与地统计混合模型关于地形、植被环境变量的预测性能，但由于所选变量与 SOC 的线性关系并不明显，获得的 R^2 值仅为 0.04。尽管结果表明 GM 数据与传统地形数据、遥感数据的结合可以有效提高预测精度，但本文中对 GM 地形分类变量的探索也并不完备，例如不同分辨率下获得的 GM 地形分类图也有着显著的差别^[36]，而由此获得的不同 GM 数据作为环境变量的区别有待于进一步研究。

3.2 环境变量重要性评价

由于土壤性质和植被覆盖之间的密切关系，各种各样的植被指数和遥感反射率也均是被广泛使用的预测变量。SAR 数据对模型的贡献度则取决于后向散射系数对地表条件和土壤湿度变化的敏感性^[37]。尽管在国际上如哨兵 1 号等 SAR 数据在 SOC 制图中的应用已有了探索^[38]，但利用中国雷达卫星数据进行相关实践的研究仍非常少见。本文的分析结果表明高分三号 SAR 数据在中国东部丘陵区的 SOC 预测中也表现出了较高的适应性。

地形是土壤形成过程中最重要的因素之一，基于 DEM 的衍生地形变量通常被用作 DSM 的关键预测变量，因为地形控制着溶质、水和沉积物的流动，进而影响土壤发育和土壤属性的空间分布。前人研究表明，高程、坡度和 TWI 等地形变量与 SOC 值具有高度显著的相关性，并对其他土壤属性的空间分布产生影响^[39]。与本文的结果相似，Nabiollahi 等^[9]发现在伊朗西南部的土壤属性预测制图中，DEM 衍生变量中的 MRVBF 同样是最重要的环境变量。除 MRVBF 之外，如高程值、TWI 等 DEM 衍生变量也是较为重要的地形因子，这些不同的 DEM 衍生变量也都曾被认定为预测土壤属性的关键因子^[40]。

GM 地形分类变量作为地形特征信息的另一种反映，在分析结果中也体现出与模型中其它 DEM 衍生变量不同的贡献度，四种具有较大差异的 GM 变量在 Cubist 模型中能独立解释 SOC 值变异的 18%。在这四种 GM 变量中，20 cells L 与 $5^\circ t$ 设置下获得

的 GM 地形分类图在三种机器学习模型中均表现出了最高的相对重要性，且相比 GM (20 L , 5 t) 和 GM (5 L , 5 t) 在同一模型中的贡献度均有较大的差距。这表明在小流域尺度的高精度 GM 制图中，20 像元左右的搜索半径较为合适，能捕捉较多的地形差异，而 5 像元设置下则可能导致生成的地形分类图破碎化程度较高，包含较多的无用或错误信息。搜索半径的设置在一定程度上受制于制图尺度以及对精度的需求，如 Ngunjiri 等^[41]在肯尼亚西部高原应用 GM 地形分类信息进行 1: 250 000 土壤类型预测制图，在 30 m 的空间分辨率基础上 20 像元的搜索半径同样有最佳的预测效果，表明搜索半径的适宜大小受制于制图尺度以及空间分辨率，但更细致的规律有待进一步研究。在相同的搜索半径下，丘陵地形对于较大的平坦度阈值 ($5^\circ t$) 相比默认值 ($1^\circ t$) 明显有更高的适应性，即放宽对“平坦地形”的识别要求，获得更大面积的“平地”对应于丘陵区实际的田块分布。因为丘陵地形起伏较大，该区域有较多的农田并不是建立在完全的平坦地，而是略有起伏的缓坡，这一差异在流域尺度上可能并不显著，但在运用精细 DEM 的小流域尺度上却较为突出，因此较大的平坦度阈值设置生成的 GM 变量相比常规设置的 $1^\circ t$ 在模型中拥有更高的相对重要性。综合来看，对于有精细地形分类或相关制图需求的应用研究，20 像元左右的搜索半径配合对应尺度的高精度 DEM 可以捕捉足够多的地形特征并获得较好的分类结果。平坦度阈值 t 的设置则应考虑研究区域的地形起伏程度，丘陵区 and 山区适用于较大的阈值如 $5^\circ t$ ，而平原地区则以不超过 2° 为宜^[13]。除此之外，经过聚类后的 GM 地形分类图 (图 2b) 在 RF 和 Cubist 模型中也均显示出一定的贡献度。尽管损失了部分特征信息，这种异质性较低的 GM 聚类图在更大尺度上的 DSM 研究中可能会起到更好的作用^[42]。

3.3 土壤有机碳含量空间分布

总体而言，三个预测模型产生了相似的 SOC 空间分布格局 (图 4)。较高的 SOC 含量集中于研究区的南部和中部的部分地区，主要为地势低洼区域。由于该地降雨较为频繁，水力作用导致的土壤侵蚀时有发生，较细的土壤颗粒在水流的搬运作用和地形高差导致的重力作用下向低处迁移，而黏粒含量较高的土壤对于养分的吸附和维持较强，因此有更

高的 SOC 含量。该结果反应了地形特征对 SOC 含量空间分异的影响。齐雁冰等^[43]及袁玉琦等^[44]在利用 RF 模型分别对陕西省土壤有机质和福建亚热带复杂地貌区有机碳进行预测的研究中得出了相似的结论,即在地貌复杂多变且耕地较少的区域,地形及其相关环境变量往往对 SOC 的空间分布有关键性的影响。而在小流域尺度内,没有了气候因素的作用(如降雨量和积温的空间分布不均等),高程和地形对 SOC 空间分布的影响则更为突出。

SOC 的准确预测是制订农业政策、改进农耕措施、缓解气候变化、落实环境保护等工作的基础和前提。而高分辨率 GM 地形分类数据的应用为改进土壤属性预测和环境监测提供了更多有效的助益。因此,如何为多尺度 DSM 研究提供适宜的地形分类数据并与多源环境变量相结合以提高预测制图精度也就有了更加现实的意义。

4 结 论

本文使用 GM 地形分类变量、DEM 衍生变量和光学、SAR 遥感变量作为模型输入,采用三种机器学习方法作为模型的实现,并比较不同类型变量组合的预测精度差异来研究小流域尺度上丘陵区 SOC 含量的空间分布,结果表明适宜超参数设置下获得的 GM 变量可以应用于丘陵地形的 SOC 预测制图研究。主要结论可以概括如下:(1) Cubist 模型在小流域尺度上预测丘陵区 SOC 含量方面优于 Bagged CART 模型和 RF 模型,表现出最佳性能。三种模型的 SOC 预测图具有相似的空间分布格局和显著的空间异质性,在地势低洼处 SOC 含量较高。(2) GM 地形分类数据的应用有助于提高 SOC 含量预测精度,其中 GM (20 L, 5 t) 的表现最佳,说明丘陵地形对较大的搜索半径和平坦度阈值有更高的适应性。GM (group) 也表现出了在大尺度 DSM 研究中的应用潜力。(3) GM 地形分类变量、DEM 衍生变量和光学、SAR 遥感变量的组合拥有最高的预测精度,比较有无 GM 变量的精度变化时,使用 Cubist 模型预测 SOC 含量的 R^2 提高了 14.3%。(4) DEM 衍生变量是 SOC 预测的主要解释变量,其次是遥感类变量、GM 地形分类变量。MRVBF 和高程值是影响丘陵区 SOC 空间分布的最重要的两个环境变量。

参考文献 (References)

- [1] Meersmans J, De Ridder F, Canters F, et al. A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium) [J]. *Geoderma*, 2008, 143 (1/2): 1—13.
- [2] Schillaci C, Acutis M, Lombardo L, et al. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling [J]. *Science of the Total Environment*, 2017, 601/602: 821—832.
- [3] Yang R M, Zhang G L, Liu F, et al. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem [J]. *Ecological Indicators*, 2016, 60: 870—878.
- [4] McBratney A B, Mendonça Santos M L, Minasny B. On digital soil mapping [J]. *Geoderma*, 2003, 117 (1/2): 3—52.
- [5] Zhang G L, Liu F, Song X D. Recent progress and future prospect of digital soil mapping: A review [J]. *Journal of Integrative Agriculture*, 2017, 16 (12): 2871—2885.
- [6] Lamichhane S, Kumar L, Wilson B. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review [J]. *Geoderma*, 2019, 352: 395—413.
- [7] Guo L, Fu P, Shi T Z, et al. Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images [J]. *Soil and Tillage Research*, 2020, 196: 104477.
- [8] Were K, Bui D T, Dick Ø B, et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape [J]. *Ecological Indicators*, 2015, 52: 394—403.
- [9] Nabiollahi K, Taghizadeh-Mehrjardi R, Shahabi A, et al. Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests [J]. *Geoderma*, 2021, 385: 114858.
- [10] Song X D, Liu F, Zhang G L, et al. Mapping soil organic carbon using local terrain attributes: A comparison of different polynomial models [J]. *Pedosphere*, 2017, 27 (4): 681—693.
- [11] de Morisson Valeriano M, de Fátima Rossetti D. Delineation of main relief subdomains of central Amazonia for regional geomorphometric mapping with SRTM data [J]. *Journal of South American Earth Sciences*, 2020, 104: 102842.
- [12] Mueller T G, Pierce F J. Soil carbon maps [J]. *Soil Science Society of America Journal*, 2003, 67 (1):

- 258—267.
- [13] Flynn T, Rozanov A, Ellis F, et al. Farm-scale soil patterns derived from automated terrain classification[J]. *Catena*, 2020, 185: 104311.
- [14] Brabyn L. Landscape classification using GIS and national digital databases[J]. *Landscape Research*, 1996, 21 (3): 277—300.
- [15] Iwahashi J, Pike R J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature[J]. *Geomorphology*, 2007, 86 (3/4): 409—440.
- [16] Jasiewicz J, Stepinski T F. Geomorphons—A pattern recognition approach to classification and mapping of landforms[J]. *Geomorphology*, 2013, 182: 147—156.
- [17] Kang X, Wang Y W, Qin C Z, et al. A new method of landform element classification based on multi-scale morphology[J]. *Geographical Research*, 2016, 35 (9): 1637—1646. [康鑫, 王彦文, 秦承志, 等. 多分析尺度下综合判别的地形元素分类方法[J]. *地理研究*, 2016, 35 (9): 1637—1646.]
- [18] Gdulová K, Marešová J, Moudrý V. Accuracy assessment of the global TanDEM-X digital elevation model in a mountain environment[J]. *Remote Sensing of Environment*, 2020, 241: 111724.
- [19] Chinese Soil Taxonomy Research Group, Institute of Soil Science, Chinese Academy of Sciences, Cooperative Research Group on Chinese Soil Taxonomy. Keys to Chinese Soil Taxonomy[M]. 3rd ed. Hefei: Press of University of Science and Technology of China, 2001: 25—191. [中国科学院南京土壤研究所土壤系统分类课题组, 中国土壤系统分类课题研究协作组. 中国土壤系统分类检索 (第三版) [M]. 合肥: 中国科学技术大学出版社, 2001: 25—191.]
- [20] Cai W T, Zhao S H, Zhang Z H, et al. Comparison of different crop residue indices for estimating crop residue cover using field observation data[C]//2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics). August 6-9, 2018, Hangzhou, China. IEEE, 2018: 1—4.
- [21] Lee J S. Speckle suppression and analysis for synthetic aperture radar images[C]//29th Annual Technical Symposium. Proc SPIE 0556, Intl Conf on Speckle, San Diego, USA. 1985, 0556: 170—179.
- [22] Gholizadeh A, Žižala D, Saberioon M, et al. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging[J]. *Remote Sensing of Environment*, 2018, 218: 89—103.
- [23] Aerts W, Kint V, van Orshoven J, et al. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests[J]. *Ecological Modelling*, 2010, 221 (8): 1119—1130.
- [24] Deng H X, Diao Y F, Wu W, et al. A high-speed D-CART online fault diagnosis algorithm for rotor systems[J]. *Applied Intelligence*, 2020, 50 (1): 29—41.
- [25] Wiesmeier M, Barthold F, Spörlein P, et al. Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany) [J]. *Geoderma Regional*, 2014, 1: 67—78.
- [26] Team C R. Team RDC.R: A language and environment for statistical computing. R foundation for statistical computing: Vienna, Austria[EB/OL]. 2012
- [27] Pouladi N, Møller A B, Tabatabai S, et al. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging[J]. *Geoderma*, 2019, 342: 85—92.
- [28] Lacoste M, Minasny B, McBratney A, et al. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape[J]. *Geoderma*, 2014, 213: 296—311.
- [29] Lombardo L, Saia S, Schillaci C, et al. Modeling soil organic carbon with Quantile Regression: Dissecting predictors' effects on carbon stocks[J]. *Geoderma*, 2018, 318: 148—159.
- [30] Jeong G, Oeverdieck H, Park S J, et al. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain[J]. *Catena*, 2017, 154: 73—84.
- [31] Fatholoulumi S, Vaezi A R, Alavipanah S K, et al. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran[J]. *Science of the Total Environment*, 2020, 721: 137703.
- [32] Libohova Z, Winzeler H E, Lee B, et al. Geomorphons: Landform and property predictions in a glacial moraine in Indiana landscapes[J]. *Catena*, 2016, 142: 66—76.
- [33] Ashtekar J, Owens P, Brown R, et al. Digital mapping of soil properties and associated uncertainties in the Llanos Orientales, South America[M]//GlobalSoilMap. Boca Raton, Florida: CRC Press, 2014: 367—372.
- [34] Flynn T, Rozanov A, de Clercq W, et al. Semi-automatic disaggregation of a national resource inventory into a farm-scale soil depth class map[J]. *Geoderma*, 2019, 337: 1136—1145.
- [35] Lai Y Q, Sun X L, Wang H L. Mapping of soil organic carbon using neural network and its mixed model with geostatistics in a small area of typical hilly region[J]. *Chinese Journal of Soil Science*, 2020, 51 (6): 1313—1322. [赖雨晴, 孙孝林, 王会利. 人工神经网络及其与地统计的混合模型在小面积丘陵区土壤有机碳预测制图上的应用研究[J]. *土壤通报*, 2020, 51 (6): 1313—1322.]
- [36] Atkinson J, de Clercq W, Rozanov A. Multi-resolution soil-landscape characterisation in KwaZulu Natal: Using

- geomorphons to classify local soils for improved digital geomorphological modelling[J]. *Geoderma Regional*, 2020, 22: e00291.
- [37] Kasischke E S, Melack J M, Craig Dobson M. The use of imaging radars for ecological applications—A review[J]. *Remote Sensing of Environment*, 1997, 59(2): 141—156.
- [38] Zhou T, Geng Y J, Chen J, et al. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms[J]. *Science of the Total Environment*, 2020, 729: 138244.
- [39] Obu J, Lantuit H, Myers-Smith I, et al. Effect of terrain characteristics on soil organic carbon and total nitrogen stocks in soils of Herschel island, western Canadian arctic[J]. *Permafrost and Periglacial Processes*, 2017, 28(1): 92—107.
- [40] Adhikari K, Hartemink A E, Minasny B, et al. Digital mapping of soil organic carbon contents and stocks in Denmark[J]. *PLoS One*, 2014, 9(8): e105519.
- [41] Ngunjiri M W, Libohova Z, Owens P R, et al. Landform pattern recognition and classification for predicting soil types of the Uasin Gishu Plateau, Kenya[J]. *Catena*, 2020, 188: 104390.
- [42] Silva S H G, de Menezes M D, de Mello C R, et al. Geomorphometric tool associated with soil types and properties spatial variability at watersheds under tropical conditions[J]. *Scientia Agricola*, 2016, 73(4): 363—370.
- [43] Qi Y B, Wang Y Y, Chen Y, et al. Soil organic matter prediction based on remote sensing data and random forest model in Shaanxi Province[J]. *Journal of Natural Resources*, 2017, 32(6): 1074—1086. [齐雁冰, 王茵茵, 陈洋, 等. 基于遥感与随机森林算法的陕西省土壤有机质空间预测[J]. *自然资源学报*, 2017, 32(6): 1074—1086.]
- [44] Yuan Y Q, Chen H Y, Zhang L M, et al. Prediction of spatial distribution of soil organic carbon in farmland based on multi-variables and random forest algorithm —A case study of a subtropical complex geomorphic region in Fujian as an example[J]. *Acta Pedologica Sinica*, 2021, 58(4): 887—899. [袁玉琦, 陈瀚阅, 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.]

(责任编辑: 檀满枝)