https://doi.org/10.3799/dqkx.2026.030

东南丘陵斜坡土层厚度空间分布规律与预测模型

杜静怡[1],贺海洪[2],马　骏[1],徐兴华[3],冯杭建[3],吕　庆[1,4*]

1. 浙江大学建筑工程学院,浙江杭州 310058
2. 浙江钱塘江水利建筑工程有限公司,浙江杭州 310016
3. 浙江省地质院,浙江杭州 310007
4. 重庆交通大学未来土木科技研究院,重庆 400074

摘　要: 中国东南丘陵区群发型浅层滑坡频发,其中斜坡土层厚度是决定其易发性的重要控制参数.土层厚度存在空间变异性强、沿斜坡分布规律不清、现有模型评估精度不高等问题.为揭示东南丘陵区土层厚度空间分布规律,建立高精度预测模型,提升降雨型群发滑坡风险防控能力.以浙江省诸暨市马剑镇某滑坡群发典型小流域(约2.88 km^2)为研究区,通过直接测量、探地雷达技术和无人机摄影测量构建了高质量土层厚度数据库,采用Boruta算法和递归特征消除算法揭示影响土层厚度空间分布的主要因素;利用随机森林、极端梯度提升树、支持向量回归、回归克里金、堆叠集成等机器学习方法建立7种土层厚度预测模型.结果表明,极端梯度提升树模型预测精度最高,土层厚度预测精度可达90%以上,进一步通过SHAP方法探讨了模型的可解释性.本成果可为我国东南丘陵区降雨型滑坡稳定性评价和区域风险防控提供参考.

关键词: 土层厚度;东南丘陵;浅层滑坡;机器学习;工程地质.

中图分类号: P694　　**文章编号:** 1000－2383(2026)10－3966－13　　**收稿日期:** 2025-11-10

Spatial Distribution Pattern and Prediction Model of Soil Thickness in Mountainous Area in Southeast China

Du Jingyi[1], He Haihong[2], Ma Jun[1], Xu Xinghua[3], Feng Hangjian[3], Lü Qing[1,4*]

1. *College of Civil Engineering and Architecture, Zhejiang University, Hangzhou* 310058, *China*
2. *Zhejiang Qiantang River Water Conservancy Construction Engineering Co., Ltd., Hangzhou* 310016, *China*
3. *Zhejiang Institute of Geosciences, Hangzhou* 310007, *China*
4. *Institute of Future Civil Engineering Sciences and Technology, Chongqing Jiaotong University, Chongqing* 400074, *China*

Abstract: Widespread rainfall-induced shallow landslides occur frequently in the southeastern hilly region of China, and slope soil thickness is an important controlling parameter of landslide susceptibility. However, soil thickness shows strong spatial variability, unclear distribution patterns along slopes, and relatively low assessment accuracy in existing models. This study aims to reveal the spatial distribution characteristics of soil thickness in the southeastern hilly region and to develop a high-accuracy prediction model, thereby improving risk prevention and control of rainfall-induced widespread landslides. The study was conducted in a

基金项目: 国家重点研发计划项目(No. 2024YFC3012601);国家自然科学基金面上项目(No. 42277132);浙江省自然科学基金项目(No. LTGG23D020001).

作者简介: 杜静怡(1998－),女,硕士,主要从事地质灾害风险防控理论科研工作.ORCID:0009-0003-5242-805X. E-mail: dujingyi0824@163.com

* **通讯作者:** 吕庆(1978－),男,博士,教授,博士生导师,主要从事地质灾害风险防控方面的教学与科研工作.ORCID:0000-0003-0466-2936. E-mail:lvqing@zju.edu.cn

引用格式: 杜静怡,贺海洪,马骏,徐兴华,冯杭建,吕庆,2026.东南丘陵斜坡土层厚度空间分布规律与预测模型.地球科学,51(10):3966－3978.

Citation: Du Jingyi, He Haihong, Ma Jun, Xu Xinghua, Feng Hangjian, Lü Qing, 2026. Spatial Distribution Pattern and Prediction Model of Soil Thickness in Mountainous Area in Southeast China. *Earth Science*, 51(10): 3966－3978.

representative landslide-prone headwater catchment in Majian Town, Zhuji City, Zhejiang Province, China (area≈2.88 km^2). A high-quality soil-thickness database was established using direct measurements, ground-penetrating radar (GPR), and UAV photogrammetry. The Boruta algorithm and recursive feature elimination (RFE) were applied to identify the main factors influencing the spatial distribution of soil thickness. Seven soil-thickness prediction models were developed using machine-learning methods, including random forest (RF), extreme gradient boosting (XGBoost), support vector regression (SVR), regression kriging (RK), and stacking ensemble approaches. The results show that the XGBoost model achieved the highest predictive performance, with soil-thickness prediction accuracy exceeding 90%. SHAP analysis was further applied to investigate model interpretability. The findings provide a reference for slope stability assessment and regional risk prevention and control of rainfall-induced landslides in the southeastern hilly region of China.

Key words: soil thickness; southeastern hilly region; shallow landslide; machine learning; engineering geology.

0　引言

我国东南丘陵区岩体风化强烈,残坡积层广布,受台风暴雨和短历时强降雨影响,易发生降雨型群发浅层滑坡灾害.受地质、地形、植被等因素影响,此类坡一般规模不大,属浅表破坏,在暴雨作用下常演化成坡面泥石流.土层厚度是控制这类群发浅层滑坡启动的关键(Liang *et al.*, 2022),也是滑坡易发性评价与风险防控的重要参数(Xiao *et al.*, 2023).

目前,土层厚度主要通过地质调查和理论模型预测.地质调查包括直接勘测与间接勘测两种.直接勘测主要采用手动螺旋钻、土壤探针、动态锥入度仪等工具在斜坡上开展点位测试(Chan *et al.*, 2019),具有操作便捷、精度高的优点,适用于小范围调查,但该方法成本高、获取数据离散,无法获得斜坡土层厚度的连续数据.间接勘测主要是物探方法,包括地震折射法、电阻率层析成像(Gomes *et al.*,2016)和探地雷达(Ground Penetrating Radar, GPR)等(Han *et al.*,2018).其中,GPR凭借高分辨率与高效率,在多种地质条件下被广泛用于土层厚度调查(Han *et al.*, 2018;Zhang *et al.*,2018),在坡面结构识别与厚度插值方面获得了良好效果.尽管如此,无论是直接勘测还是间接勘测,都需要投入大量人力、物力,获取的土层厚度为离散数据,难以满足大尺度范围内斜坡土层厚度数据获取需求.因此如何基于有限数据预测区域土层厚度从而揭示其空间分布规律,成为研究热点.

目前土层厚度预测方法主要包括3类:空间插值法、基于土壤演化过程的物理模型预测法、基于环境因子的经验模型与机器学习方法.空间插值法,如普通克里金法与反距离权重法,具有实现简便、可快速制图的优点,但忽略了土壤风化与迁移机制,精度有限(Dietrich *et al.*, 1995).物理模型如扩散型坡面演化模型考虑了坡面地貌对成土过程的影响,具备一定理论基础(Pelletier and Rasmussen, 2009),但依赖大量难以获取的物理参数,仅适用于小范围单坡预测,区域适应性差(Zhang *et al.*, 2022).相比之下,基于环境因子的机器学习方法兼具实用性与合理性,成为土层厚度区域预测模型的重要研究方向,已在国内外取得了不少成果,例如Shangguan *et al.*(2017)开发了全球土壤厚度数据集,Yan *et al.*(2020)构建了中国高分辨率土壤厚度图,Li *et al.*(2020)和Liu *et al.*(2022)引入集成学习与区域因子显著提升了模型稳定性.与此同时,机器学习也是滑坡易发性评价与风险研究中的常用方法之一(彭铭等,2025;贾卓等,2025).然而,国内已有模型较多针对西南地区,泛化能力和区域适用性有限.针对东南丘陵区,尤其是碎石含量高、地质结构复杂的小流域斜坡覆盖层厚度的预测模型尚很缺乏(Luo *et al.*, 2024).

为此,本文以浙江省诸暨市马剑镇一典型小流域(2.88 km^2)为研究区,针对丘陵斜坡土层厚度获取困难、样本有限与空间变异性强等问题,构建“多源土层厚度获取、关键环境因子筛选、多模型预测、模型可解释性分析”的研究流程.基于直接测量、探地雷达与无人机摄影测量建立土层厚度数据库,结合Boruta与RFE筛选关键环境因子,采用多种机器学习模型预测土层厚度空间分布,并以最优模型的SHAP解释揭示主控因子作用机制.研究结果可为东南丘陵区降雨型浅层滑坡稳定性评价与区域风险防控提供参考.

1 土层厚度获取与因子重要性分析

1.1 研究区概况

研究区位于浙江省诸暨市马剑镇东北部，面积为 2.88 km^2（图 1）. 地貌属典型的构造侵蚀丘陵，地形总体呈现中部隆起、两翼低凹特征，区内高差显著，高程范围为 166～513 m. 区内山地坡度变化较大，以中－陡坡主导. 坡向覆盖 4 个方位呈斑块状分布. 地层岩性为下白垩统流纹斑岩，岩体坚硬破碎，风化程度高. 区内斜坡坡表广泛发育残坡积、坡洪积等第四系松散堆积物. 其中残坡积物为滑坡和泥石流等地质灾害的主要物源，其厚度空间差异显著，随地貌位置、坡度与基岩起伏而异. 研究区内降雨丰沛，年平均降水量高达 1 800 mm 以上，区内发育 20 余处滑坡.

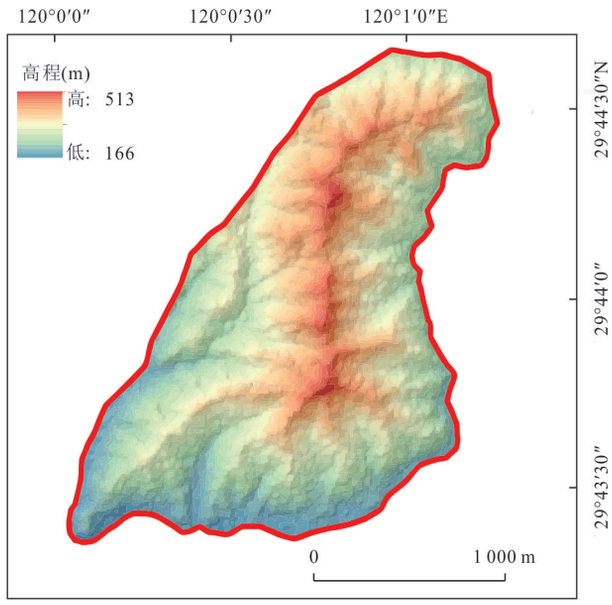


图 1 研究区地理位置及地形特征概况

Fig.1 Location and topographic setting of the study area

1.2 土层厚度数据获取

土层厚度数据的精准获取是区域土层空间分布预测的基础前提. 本文综合采用直接测量、探地雷达和无人机摄影测量 3 种方法进行土层厚度数据采集. 在地势平缓且具良好可达性的区域，优先采用通过地质露头测量或人工浅层开挖方式进行直接测量，以获取准确的土层厚度数据（图 2）.

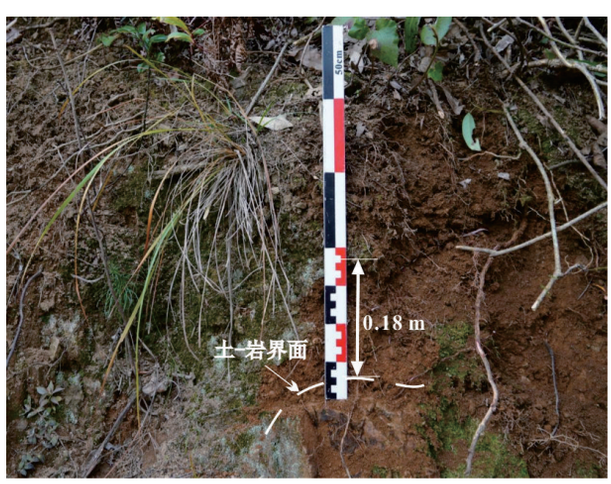


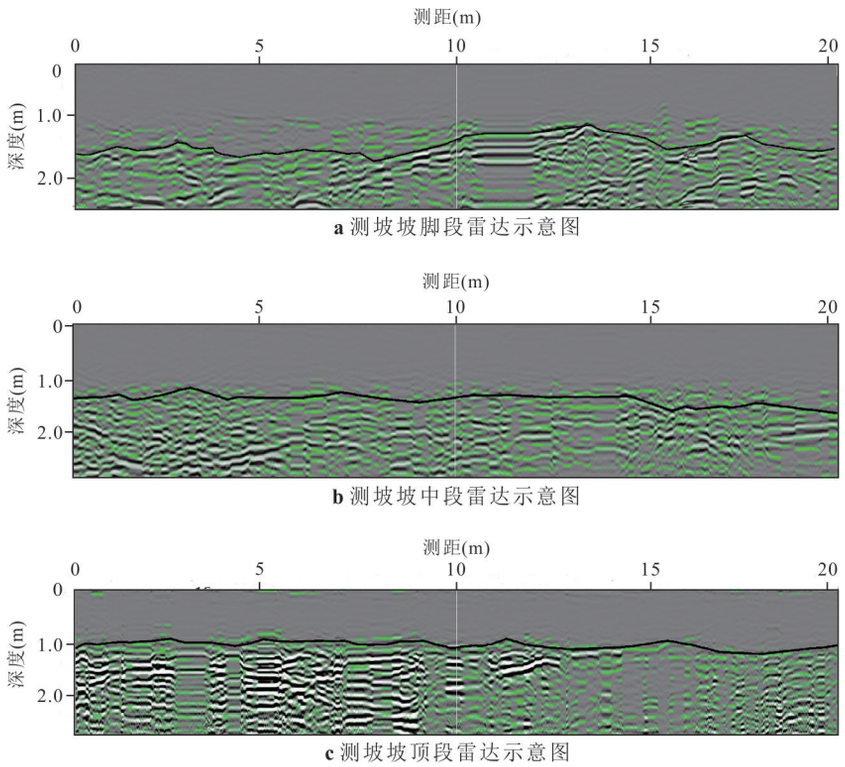
图 2 斜坡土层厚度直接测量方法示意

Fig.2 Direct measurement of slope soil thickness

对地质露头缺失或不易开展直接测量区域，本文采用探地雷达进行线状连续探测. 该方法基于高频电磁波在土壤与基岩等介质中传播速度差异，可识别土－岩界面位置. 探地雷达型号选用美国 GSSI 公司 SIR-3000 系统，搭载 400 MHz 与 900 MHz 双频天线，可兼顾不同厚度范围的探测需求. 测量流程包括外业采集与内业处理：外业方面，结合遥感影像与实地踏勘优化测线布设，清理表层植被后进行拖曳式扫描，实时监控回波图并同步记录定位信息. 内业采用 RADAN 7 软件进行数据预处理（包括距离归一、时零校正、滤波增益等），最终提取土层厚度数据，如图 3.

此外，本文采用无人机摄影测量技术进行滑坡所在区域地质露头土层厚度反演. 首先获取约 200 张多视角、高重叠的影像，进行图像对齐与点云生成，构建数字高程模型（Digital Elevation Model，DEM）与三维纹理模型（图 4a，4b）. 结合 DEM 数据与基岩出露区域影像细节，得到土层厚度的定量结果（图 4c），最后结合少量可达位置的现场测量结果与典型点对照，对反演结果进行合理性验证.

为保证不同来源土层厚度数据的一致性，对 3 类数据进行了统一与筛查：首先统一坐标系统与厚度单位，并核对采样点位置；其次以现场直接测量结果作为对照，对 GPR 解释与无人机反演结果进行合理性检查；对界面识别不清或明显异常的样本点进行复核或剔除. 最终获得土层厚度数据为 56 个（图 5），每个采样点的位置均通过全球定位系统确定. 该区域土层厚度介于 20～210 cm，平均值为 64 cm，标准差为 43 cm，变异系数为 67.57%，呈中等程度的变异水平，偏度系数为 1.15. 样本的土层厚度分布具有明显的右偏特征，表明大多数样本点土层较薄，少数点存在较厚的覆盖层，后者主要聚集在坡脚、沟谷等残积土积聚区.



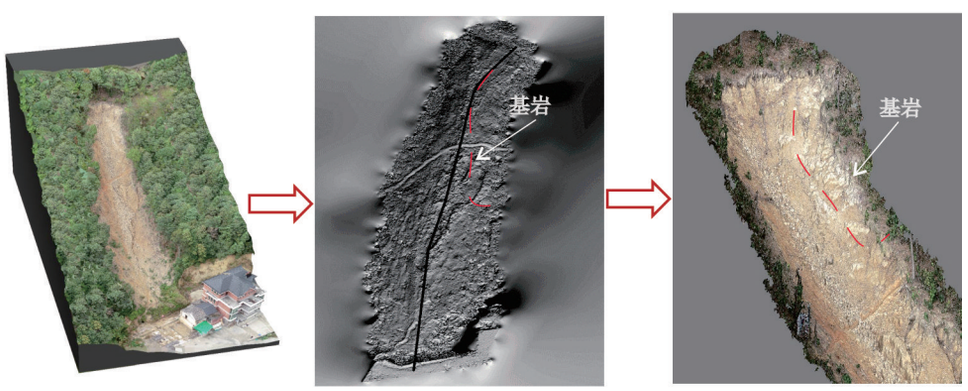
a 测坡坡脚段雷达示意图

b 测坡坡中段雷达示意图

c 测坡坡顶段雷达示意图

图 3　典型测线探地雷达剖面及土－岩界面识别结果

Fig.3　Representative GPR profile and interpreted soil-bedrock interface



a 三维网格构建　　**b** DEM生成　　**c** 土层厚度识别

图 4　基于无人机摄影测量的土层厚度反演流程与结果

Fig.4　UAV photogrammetry-based workflow and results for soil thickness inversion

1.3　样本代表性分析

为评估研究区样本在地形因子特征空间中的代表性,本文选取高程、坡度及剖面曲率等基础地形因子,采用 Kolmogorov-Smirnov 检验(K-S),对样本点与研究区背景栅格在上述因子上的分布差异进行了比较分析(表 1).为避免研究区背景栅格数量远大于样点数量对检验结果造成影响,本文从研究区栅格中随机等量抽取与样点数一致的像元作为背景样本参与检验.结果表明,样本点与研究区背景在所选地形因子上的分布差异均未达到显著水平($P>0.05$),说明现有样本在地形因子特征空间中具有良好的代表性,可为研究区尺度后续因子分析与土层厚度预测建模提供可靠数据基础.

1.4　环境影响因子提取及重要性分析

本文共选取包括地形和植被两类环境因子在内的 35 个候选因子(表 2),对其进行重要性分析以得到关键环境影响因子.地形因子基于 6 m 分辨率的 DEM 数据提取,处理过程中统一重投影至

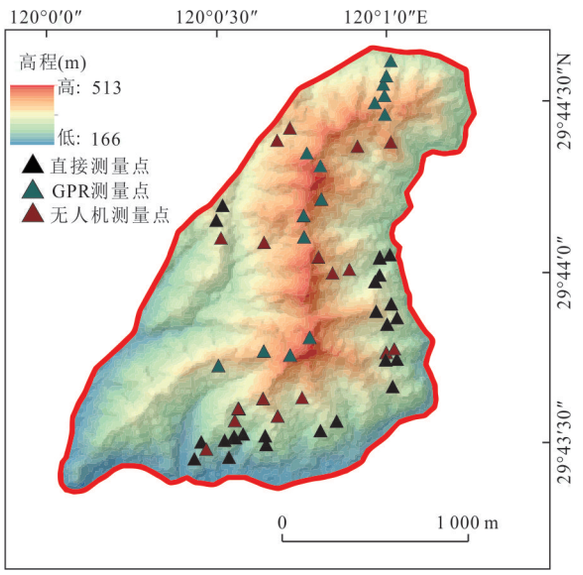


图 5 研究区土层厚度采样点空间分布

Fig.5 Spatial distribution of soil-thickness sampling points

表 1 采样点与研究区地形因子 K-S 检验统计

Table 1 K-S test results comparing topographic factors between sampling points and the study area

地形因子	K-S统计量	K-S检验结果(P值)	显著性
高程	0.33	0.93	不显著
坡度	0.29	0.96	不显著
剖面曲率	0.25	0.87	不显著

WGS84 UTM Zone 51N 坐标系，并采用 ArcMap 与 SAGA GIS 等工具进行因子提取，水文因子采用 D8 与 D∞算法计算．植被因子采用归一化植被指数表征，基于美国地质调查局(United States Geological Survey，USGS)提供的 2024 年 6 月 Landsat 8-9 OLI/TIRS 遥感影像计算．

为筛选适量且关键的环境影响因子，首先开展因子重要性分析，包括因子重要性排序及关键影响因子筛选两部分．先采用标准化(z-score)方法对数据进行预处理，使不同因子具有统一的量纲和尺度．在此基础上使用内置重要性排序机制的随机森林模型(Breiman，2001)进行 35 个因子的重要性排序，后使用 Boruta 算法和递归特征消除算法(Recursive Feature Elimination，RFE)筛选出关键影响因子，前者具备较强的特征稳定性识别能力，后者则通过迭代式识别最优影响因子组合，两种方法在地学建模中均被广泛采用(Pimenta Barbosa de Sousa *et al.*，2023)．

利用随机森林模型对可能影响土层厚度空间分布的 35 个影响因子进行重要性排序(图 6)．通过候选影响因子重要性排序可得：(1)D∞算法下计算得到的地形湿度指数(TWI)、河道网络基准面(CNBL)和上坡集水区面积(A)是随机森林重要性排序下影响研究区土层厚度空间分布的 3 个重要因子．这三者均为与水文密切相关的地形因子，其中 TWI 是表征水分积聚能力的重要参数．同时，CNBL 和 A 分别作为河道网络基准面以及上坡集水区面积，反映了地形的水文特性和流域的空间结构特征，两者对土层厚度空间分布的影响也十分显著．(2)D∞算法下的水文因子对土层厚度空间分布的重要性普遍高于 D8 算法下的地形水文因子，甚至呈两级分化趋势，D∞算法水流划分方式更为精细，更能反映真实情况．

为了筛选影响土层厚度空间分布的关键环境因子组合，本文通过随机森林模型的决定系数(R^2)和均方根误差(RMSE)对 Boruta 和 RFE 两种筛选算法的输出结果进行了对比分析．

Boruta 算法作为一种全相关特征选择方法，具备自适应迭代机制，能够在多个训练集划分下筛选出最稳定的因子组合及其数量；而 RFE 则依赖于预设的特征数量，需通过多次交叉验证来确定最优的因子组合．考虑到研究区土层厚度数据样本数量相对有限，若选择过多的变量可能导致维度灾难和模型过拟合，而过少的变量则可能降低模型的解释力．因此，本文在 RFE 方法中设置变量数量为 3～10 个，以平衡模型的复杂性与解释力．

两种算法各因子组合性能对比统计结果中(表 3)，RFE 方法在选取 7 个影响因子时模型性能最优，其对应的 R^2 最高、RMSE 最低，优于 Boruta 筛选结果．因此，最终确定的主要环境因子共 7 项，分别为 D∞算法下的地形湿度指数(TWI_D∞)、河道网络基准面(CNBL)、上坡集水区面积(A_D∞)、归一化相对高程(NH)、地形表面凸度(TSC)、相对坡位(RSP)及高程(Elevation)．上述经筛选优化得到的的 7 个关键环境因子，构建了土层厚度预测的特征空间，将作为核心输入变量用于后续章节中的机器学习建模与对比分析．

2 土层厚度预测模型结果及分析

2.1 机器学习模型及验证指标

鉴于丘陵斜坡区土层厚度与环境因子之间的复杂关系，且在样本数量有限条件下往往表现出显

表 2　土层厚度预测候选环境影响因子属性

Table 2　Candidate environmental variables for soil-thickness prediction

环境影响因子	简称	单位	范围
高程(Elevation)	Ele	m	166～513
坡度(Slope)	Slope	°	0～75
坡向(Aspect)	Aspect	°	0～360
剖面曲率(Profile Curvature)	PRC	/	−0.11～0.09
平面曲率(Plan Curvature)	PLC	/	−4.9～3.86
D8上坡集水区面积(Upslope Contributing Area)	A_D8	m^2	3.52～13.25
D8地形湿度指数(Topographic Wetness Index)	TWI_D8	/	0.82～17.7
D8水流强度指数(Stream Power Index)	SPI_D8	/	0～52,718
D8泥沙输运指数(Sediment Transport Index)	STI_D8	/	0～1,465
D8坡度坡长因子(Length-slope factor)	LS_D8	/	0～416.4
D∞上坡集水区面积(Upslope Contributing Area)	A_D∞	m^2	3.52～11.64
D∞地形湿度指数(Topographic Wetness Index)	TWI_D∞	/	1.26～22.47
D∞水流强度指数(Stream Power Index)	SPI_D∞	/	0～9,266
D∞泥沙输运指数(Sediment Transport Index)	STI_D∞	/	0～524.86
D∞坡度坡长因子(Length-slope factor)	LS_D∞	/	0～214.56
地形位置指数(Topographic Position Index)	TPI	m	−73.81～70.54
地形粗糙度指数(Topographic Roughness Index)	TRI	m	1.32～48.09
相对坡位(Relative Slope Position)	RSP	/	−0.11～1.06
多分辨率谷底平坦度(Multiresolution Index of Valley Bottom Flatness)	MRVBF	/	0～2.98
多分辨率山脊平坦度(Multiresolution Index of Ridge Top Flatness)	MRRTF	/	0～1.93
河道网络距离(Channel Network Distance)	CND	m	−8.4～194.9
河道网络基准面(Channel Network Base Level)	CNBL	m	166.25～413.68
归一化相对高程(Normalized Height)	NH	/	0～1
归一化植被指数(Normalized Difference Vegetation Index)	NDVI	/	0.028～0.51
汇聚指数(Convergence Index)	CI	/	−88～93
水平地表流动距离(Horizontal Overland Flow Distance)	HOFD	m	0～466.45
形态保护指数(Morphometric Protection Index)	MPI	m	0～0.6
负地形开放度(Negative Topographic Openness)	NTO	/	49.7～103.12
地表流动距离(Overland Flow Distance)	OFD	/	0～511.79
正地形开放度(Positive Topographic Openness)	PTO	/	51.91～103.25
实际表面积(Real Surface Area)	RSA	m^2	33～135
地形表面凸度(Terrain Surface Convexity)	TSC	m	4.73～94.34
地形表面纹理(Terrain Surface Texture)	TST	/	0～53.98
向量粗糙度指数(Vector Ruggedness Measure)	VRM	/	0～511.79
垂直地表流动距离(Vertical Overland Flow Distance)	VOFD	/	0～219.14

著的非线性特征,本文通过多模型对比的方式,分析了不同机器学习方法在复杂地形条件下对土层厚度空间变异的刻画能力与稳定性.通过这种多模型框架,模型的适应性得到了显著提升,尤其在该地区的应用中,能够有效应对空间异质性和非线性特征.为此,本文选取了文献中广泛用于土层厚度预测的7种主流模型进行对比,包括:随机森林(Random Forest,RF)、分位数回归森林(Quantile Regression Forest,QRF)、支持向量回归(Support Vector Regression,SVR)、极端梯度提升树(Extreme Gradient Boosting,XGBoost)、基于RF/XGBoost的回归克里金(Regression Kriging,RK)和堆叠集成模型(Stacking).其中,RF、QRF、XGBoost已被多项研究证实在复杂地形区土壤属性建模中具有良好性能(Li *et al.*, 2020; Patton *et al.*, 2022; Khanifar and Khademalrasoul,2022),而RK和Stacking模型则通过引入空间残差结构或模型融合机制,用于评估其在复杂地形条件下对预测精度与稳健

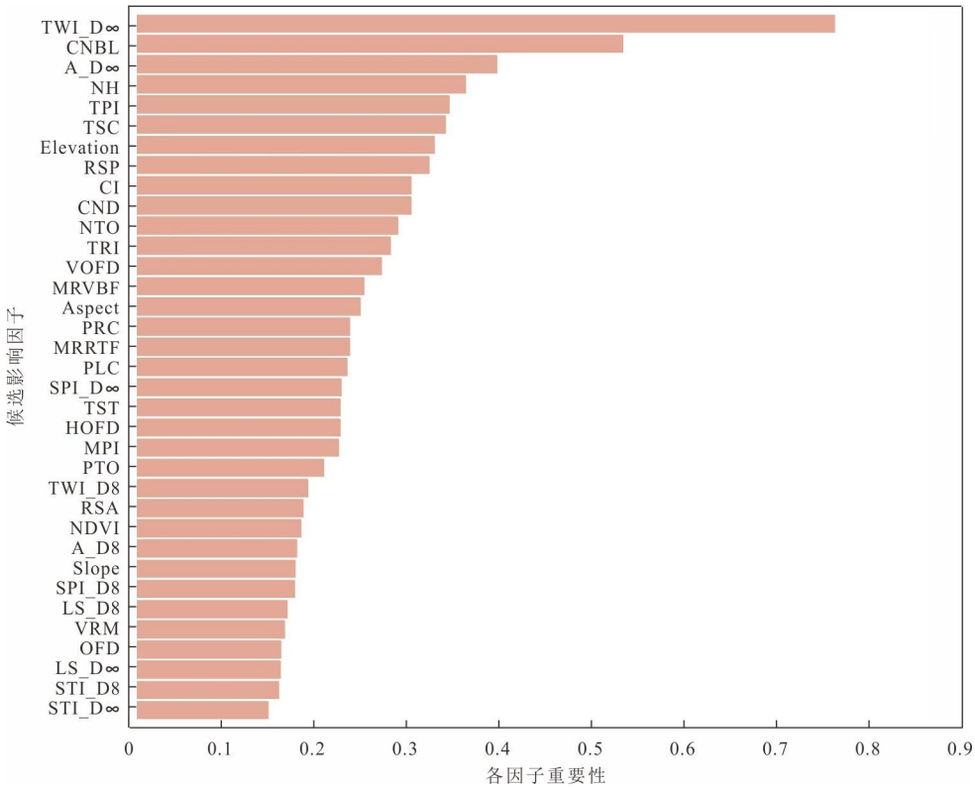


图 6 候选环境影响因子的随机森林重要性排序结果

Fig.6 Random forest importance ranking of candidate environmental factors

表 3 Boruta 与 RFE 算法下不同因子组合的预测性能对比

Table 3 Performance comparison of factor combinations for two algorithms

各因子组合	Boruta算法	RFE(10)	RFE(9)	RFE(8)	RFE(7)	RFE(6)	RFE(5)	RFE(4)	RFE(3)
R^2	0.87	0.86	0.85	0.86	0.88	0.85	0.86	0.85	0.84
RMSE	0.16	0.17	0.17	0.17	0.15	0.17	0.18	0.18	0.18

性的潜在提升效果(Li *et al.*, 2020).

所有模型均基于Python环境,分别调用Scikit-learn、XGBoost、quantile-forest及PyKrige等标准库实现.在模型输入上,均采用经筛选后的7项关键因子作为输入,并统一进行标准化;数据集涵盖利用3种方式得到的56个土层厚度样本数据,严格按7:3比例随机划分为训练集与测试集.为降低参数选择的主观性,各模型均采用网格搜索(GridSearchCV)结合5折交叉验证确定最优参数.

最终核心参数设置:RF与QRF模型决策树数量(n_estimators)为100,节点分裂特征数(max_features)为2;XGBoost模型:学习率(learning_rate)=0.1,n_estimators=200,子采样(subsample)=0.9,列采样(colsample_bytree)=0.7,最大树深(max_depth)=7;SVR模型选用RBF核函数,惩罚系数C=10,误差容忍度epsilon=0.01,核函数系数gamma=0.01;RK模型残差部分采用泛克里金插值及球状变差模型;Stacking模型以RF、SVR和XGBoost为基学习器,岭回归(Ridge)为元学习器.

鉴于本文研究对象为连续变量的回归预测问题,模型泛化能力主要通过独立测试集进行验证,并结合训练集与测试集预测误差的对比进行综合评估,因此本文选用决定系数(R^2)、均方根误差(RMSE)、平均绝对误差(MAE)以及均方误差(MSE)4个通用回归评价指标,对模型预测性能进行综合对比与评价.

2.2 机器学习模型结果对比

从各模型在R^2、RMSE、MAE及MSE等指标上的对比结果可以看出,模型在拟合精度和预测稳定性方面表现差异显著(图7).总体而言,XGBoost模型在各项指标中表现最佳,R^2为0.904,说明该模型

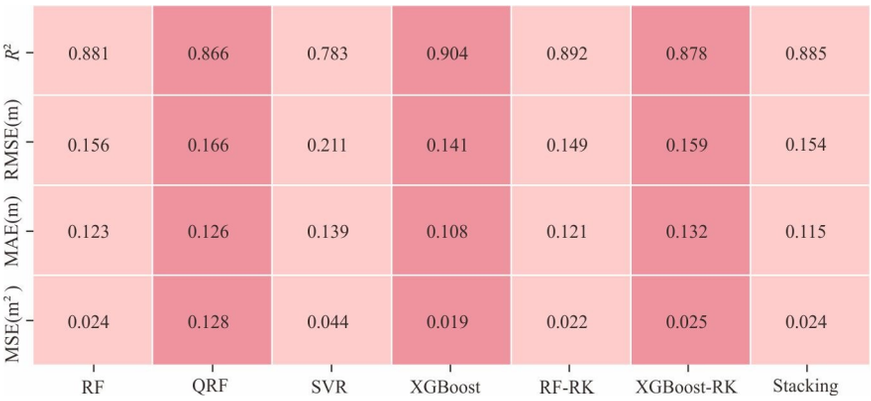


图 7　7种机器学习模型预测精度评价指标对比

Fig.7　Comparison of validation metrics across different machine learning models

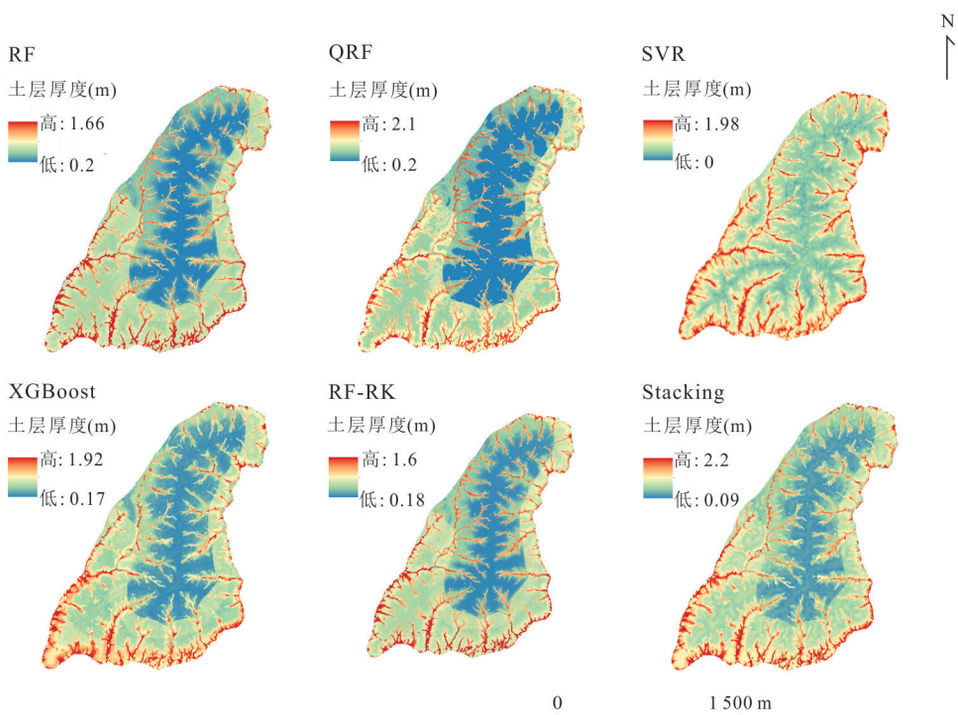


图 8　不同机器学习模型的土层厚度空间分布预测结果对比

Fig.8　Predicted soil-thickness maps derived from different machine learning models

能成功预测土层厚度,且具有较强的解释能力.同时,其MAE(0.108 m)和RMSE(0.141 m)均显著低于其他模型,表明在绝对误差等方面也具有优势,预测结果更精确且稳定.

相较之下,SVR模型表现最差,R^2仅为0.783,且其误差指标(RMSE为0.211 m,MAE为0.139 m)均为最高,反映出该模型的泛化能力较弱,不能满足高精度土层厚度建模需求.

RF-RK和Stacking模型的R^2分别为0.892和0.885,略逊于XGBoost,但仍具有较强的拟合能力.两个模型的误差指标略有上升(例如RF-RK的MAE为0.121 m,Stacking的RMSE为0.154 m),但整体误差仍在可接受范围内,具备较好的实际应用潜力.传统集成模型RF和QRF的R^2分别为0.881和0.866,预测性能稳定,但误差控制稍差,表明其在捕捉复杂地形—土壤关系时存在一定局限.

综合对比可见,XGBoost在测试集上取得最佳综合表现.该优势可能与研究区土层厚度受汇水条件、相对坡位等多因子共同控制有关,各环境因子之间存在明显的非线性与交互效应;同时样本数量有限、空间异质性较强,在此类情形下,提升类树模型(如XGBoost)能够更有效地刻画非线性关系,并

可通过正则化与子采样等策略在一定程度上抑制过拟合，因此在本研究的小样本与强空间异质性条件下，相较部分单一回归/核方法及简单模型，整体表现更为稳健．该认识与相关研究在复杂地形、强空间异质性区域中树模型集成方法通常表现较稳定的结论一致（Li *et al.*, 2020；Yan *et al.*, 2020）．

因 XGBoost-RK 模型各项指标均不及 XGBoost 与 RF-RK，故不进行对比分析，其余 6 种机器学习算法所预测的研究区土层厚度空间分布结果表明（图 8），不同模型在全区层面均捕捉到了相似的空间分布规律，且在局部区域具有较高一致性．各模型间的主要差异表现为预测值的绝对幅度而非空间结构，这说明尽管建模方法存在差异，各模型在识别研究区内土层厚度的主控空间因子方面具有一致性．

3 讨论

3.1 基于 SHAP 值的 XGBoost 解释与可视化分析

鉴于 XGBoost 在模型对比中表现最优，且其内部决策过程难以直接解释，本文采用 SHAP 方法对该模型进行解释分析，以量化关键因子的贡献并识别其作用方向（Lundberg *et al.*, 2020）．

图 9 展示了 7 个关键环境因子的 SHAP 点云分布（图中含义见图注）．以地形湿度指数（TWI_D∞）为例，其高值样点主要对应正的 SHAP 值，表明水分富集条件对土层厚度预测具有显著的正向促进作用．总体而言，TWI_D∞ 与上坡集水区面积（A_D∞）对土层厚度呈正向贡献，而 CNBL、RSP 与 Elevation 等因子整体呈负向贡献．

在正向贡献因子中，TWI_D∞ 的响应最为显著，随其取值增大 SHAP 值整体上升，表明水文汇聚对土层厚度具有主导促进作用；在负向贡献因子中，CNBL 最为突出，其取值增大时 SHAP 值整体降低，说明模型在坡顶或分水岭等高 CNBL 区域倾向预测较薄土层，这与坡顶和山脊地区土层较薄的实际认识一致．

7 个因子在空间维度上的 SHAP 值分布，进一步揭示了模型对地貌结构的响应机制（图 10）．CNBL 在脊线及分水岭区域以负值为主、坡脚区转为正值，表明其与坡位变化密切相关；Elevation 的整体响应较为平缓，主要在高海拔区域表现为负向调节；TWI_D∞ 与 A_D∞ 在水分富集区域（如沟谷、汇水区域）的 SHAP 值明显增大，表明模型识别了

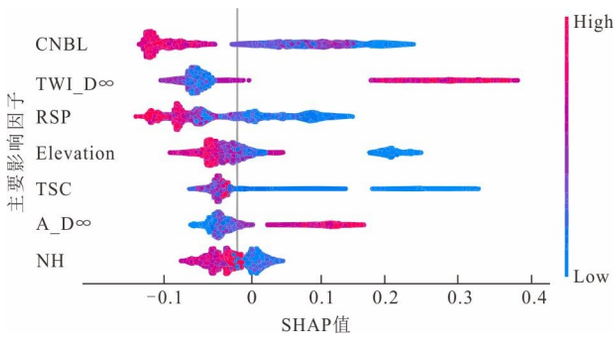


图 9 XGBoost 模型关键环境因子的 SHAP 特征贡献分布

Fig.9 SHAP bee swarm plot of key environmental factors for the XGBoost model

横轴表示因子对土层厚度预测值的贡献强度；样点颜色由蓝到红代表因子数值由小到大；散点在横轴 0.0 位右侧表示该因子对土层厚度有正向影响，左侧则为负向影响

水文汇聚对土层厚度积累的促进作用；相比之下，TSC 在局部微地形区域出现零星高值，表明模型可能在这些区域设定了阈值或受到微地形影响；NH 呈现坡脚至坡顶的平滑过渡，作为辅助调节因子，模型中其作用较为次要．整体而言，模型通过水文因子的集聚响应（如水文汇聚区土层较厚）、地形因子的层控分异（如坡顶区土层较薄）以及微地形因子的局部调节（如凹地与凸地表现出的不同厚度特征），形成了结构清晰、地貌合理的土层厚度预测机制．

从预测的空间分布特征看，研究区土层厚度整体呈现由坡顶向坡脚逐渐增厚规律，沟谷及坡脚部位形成连续的相对高值区，而脊线及高坡度区域土层厚度相对较小．该分布特征与坡面物质迁移与水文过程具有一致性：在沟谷与坡脚等汇水条件较强区域，坡面径流与重力作用促使松散物质向下游搬运并堆积，有利于土层发育；相反，在坡度较大或靠近山脊的区域，侵蚀过程更占主导，土层易减薄．结合 SHAP 解释结果，水文汇聚因子整体呈正向贡献、坡位与高程等因子呈负向调节，其作用方向与上述空间分布规律相吻合．

受研究区地形条件与野外采样条件限制，本文尚难开展高密度实测支撑下的点对点空间验证，但模型预测结果在总体趋势与局部特征上具备清晰的地貌逻辑，可作为研究区土层厚度空间分布规律的模型化表征，并为后续扩大样本与跨区检验提供依据．

3.2 土层厚度空间分布规律及其主控因素分析

从整体空间分布看，研究区整体表现为由坡顶

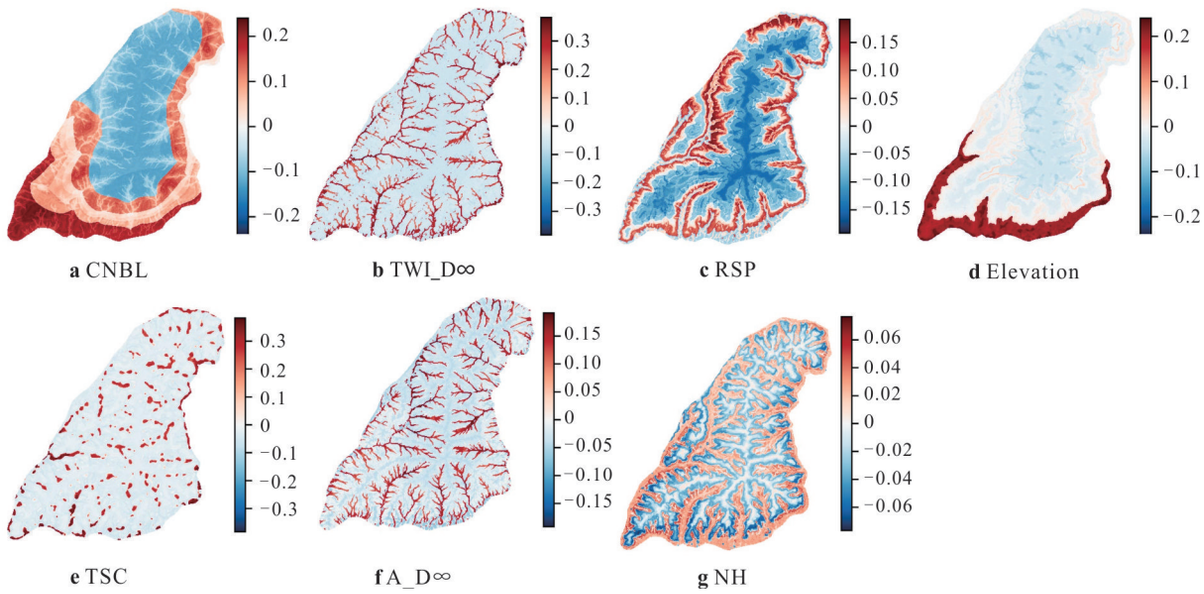


图 10　XGBoost模型关键因子 SHAP 值的空间分布特征

Fig.10　Spatial distribution of SHAP values for key factors in the XGBoost model

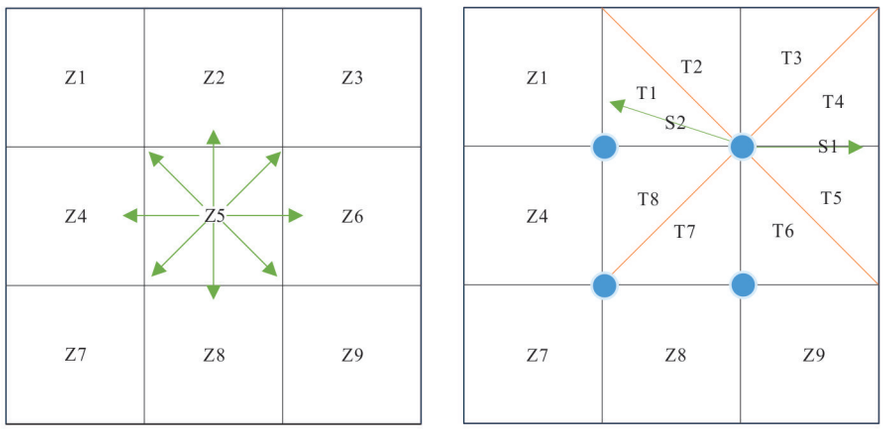
颜色代表 SHAP 值的大小.正值(红色系)表示该因子在该空间位置对土层厚度起增厚作用;负值(蓝色系)表示起减薄作用

向坡脚逐渐增厚的分布趋势.同时沟谷及坡脚部位形成连续的高值区,而脊线及高坡度区域则普遍表现为较薄的土层厚度.该空间分布规律在不同模型的预测图上表现一致(图 8),说明模型能够稳定识别研究区土层厚度的空间分布特征.在前述基于SHAP的模型解释分析基础上,本节从地形因子及其派生的水文特征的物理意义出发,进一步探讨研究区土层厚度空间分布规律及其主控因素,并重点围绕不同流向算法下水文因子差异、地形与植被因子的相对作用以及关键控制因子的作用机制进行讨论.

(1)流向是水文分析中的关键地形因子,其计算结果影响径流路径及汇水区的识别精度,从而决定地形湿度指数、上坡集水区面积等水文相关因子的表达.常用的流向算法包括D8和D∞,其差异如图 11所示,D8计算效率高且实现简单;D∞允许连续流向并可在相邻栅格间分配汇流(Tarboton,1997),更能刻画坡面分散汇流过程.两类算法各有适用性:在地形相对简单或强调计算效率的场景中,D8具有优势;而在沟谷发育、地形起伏较强的区域,D∞可能更能反映实际汇流过程,但D∞算法也存在计算量大、对DEM分辨率更为敏感等问题.本文在同一DEM与统一参数条件下分别提取两种算法下的水文因子并进行对比,结果表明,基于D∞算法得到的TWI_D∞与A_D∞与土层厚度的对应关系更清晰,模型解释与预测表现也更稳定.因此,在本研究区条件下,D∞算法更适合作为水文因子提取的基础方法.

(2)在本研究区,相较于植被因子,地形特征与研究区土层厚度的关联性更强,尤其低地到高地过渡的区域,河道网络基准面和地形湿度指数等地形指标表现出显著的影响.这一趋势与国内外实地研究结果一致(Li *et al.*, 2020;Ferrell *et al.*, 2023).总体而言,地形相关参数是预测土壤特性时最常用且相对稳定的环境因子之一.

(3)地形因子中与水文相关的因子对土层厚度空间分布起主导作用.在D∞算法下计算得到的地形湿度指数(TWI_D∞)、河道网络基准面(CNBL)和上坡集水区面积(A_D∞)是影响研究区土层厚度空间分布的前3个重要因子.这三者均为与水文密切相关的地形因子,其中TWI为表征水分积聚能力的重要参数,对土层厚度具显著影响,该结果与相关研究中“地形湿度指数常为土层厚度的重要控制因子”的认识一致(Li *et al.*, 2020;Zhang *et al.*, 2022;Ferrell *et al.*, 2023).同时,CNBL和A_D∞分别作为河道网络基准面和上坡集水区面积,反映了地形的水文特性和流域的空间结构特征.以归一化相对高程(NH)、相对坡位(RSP)为代表的地形影响因子也对土层厚度空间分布起一定调节作用,NH值和RSP值越大,表示越靠近坡顶或山脊区域,



a D8算法流向示意图 **b** D∞算法流向示意图

图 11 基于 D8 与 D∞算法的地表水流方向提取结果对比

Fig.11 Comparison of flow-direction results derived from the D8 and D∞ algorithms

a.D8算法在3×3邻域(Z1～Z9)内,将中心栅格(Z5)的水流指向坡降最大的相邻栅格,箭头表示 8 个离散流向;b.D∞ 算法将邻域划分为 8 个三角形单元(T1～T8),依据坡度方向确定连续流向(S1、S2 为示意),并可将汇流按夹角比例分配至相邻两个栅格

更易受剥蚀作用影响而表现为较薄土层.

研究区土层厚度空间分布主要受控于坡面水文积聚特征与微地形(如相对坡位、归一化相对高程等)两个维度.水文因子决定了物质沉积与水分汇聚的空间分布,而微地形则通过反映剥蚀与沉积的地貌过程分异,进一步影响土层厚度的空间分布;二者的协同作用构建了土层厚度的空间异质性主控框架.

土层厚度是降雨型浅层滑坡稳定性评价与易发性分析中的关键输入参数之一.本文获得的土层厚度空间预测结果可为区域滑坡风险评估与防控提供参数基础,并为隐患排查与监测布设提供参考.受样本密度与研究区尺度限制,上述应用仍需在更大范围与更多样本条件下进一步检验与完善.

4 结论

本文针对东南山地丘陵区斜坡土层厚度预测不准、分布规律不清的问题,以诸暨市马剑镇某滑坡高发的小流域山区为研究区,开展了土层厚度空间分布规律与预测模型研究.通过多方法土层厚度数据采集、环境变量提取与筛选、多模型对比验证以及可解释性分析,构建了集“测量－筛选－建模－解释”于一体的研究流程,为该地区滑坡预测模型提供了土层厚度参数支撑.主要结论如下:

(1)综合应用直接测量、探地雷达与无人机摄影测量3种技术,构建了具有代表性的土层厚度样本数据库.3种技术各具优势互为补充,为后续特征筛选与模型训练提供了数据基础.

(2)基于土层厚度数据及地形与植被因子,采用基于随机森林的Boruta特征选择和递归特征消除(RFE)算法,识别了土层厚度的主要控制因子.结果表明,在本研究区地形因子相较植被因子更为重要,土层厚度空间分布主要受水文条件与微地形特征的耦合控制.

(3)采用多种机器学习方法进行预测分析,结果表明,XGBoost模型在各项评估指标中表现最佳(R^2=0.904,MAE=0.108 m,RMSE=0.141 m),能够有效解释90%的土层厚度空间变异性,说明其在本研究区对土层厚度具有较强的拟合与预测能力;其他模型亦表现良好,其中支持向量回归模型相对较弱.

(4)通过SHAP方法对最优模型进行解释分析.结果显示,地形湿度指数与上坡集水区面积对土层厚度总体呈正向贡献,河道网络基准面与坡位等因子整体呈负向调节,揭示了水文汇聚与坡位分异对土层厚度形成的共同控制作用.

总体而言,本文在典型小流域尺度上揭示了东南丘陵区土层厚度的空间分异特征,并提出了可用于该类区域的样本获取与建模分析流程,为降雨型浅层滑坡稳定性评价与风险防控中土层厚度参数的空间赋值提供参考.

虽然本文构建的模型在研究区取得了较理想的预测结果,同时初步揭示了水文与微地形影响土层厚度空间分布的规律,但仍存在以下局限性及可

进一步深化的方向:

(1)地形因子的尺度敏感性.本文统一采用6 m分辨率DEM提取地形因子,不同分辨率DEM可能影响部分地形因子(如TWI、相对坡位)的提取结果及模型稳定性.未来可结合多精度DEM,评估关键因子在不同尺度下的稳定性及其对模型解释结构的影响,为区域尺度适配提供依据.

(2)"数据驱动"与"地学知识"融合.当前研究仍以数据驱动建模为主,对坡面风化、物质迁移等过程的直接表达有限.未来可探索将机器学习与具明确地貌过程意义的模型(如地貌索引模型)相结合,构建"过程－结构－数据"的分析体系,以进一步提升跨区泛化能力与物理可解释性.

References

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1): 5－32. https://doi. org/10.1023/A:1010933404324

Chan, H. C., Chang, C. C., Chen, P. A., et al., 2019. Using Multinomial Logistic Regression for Prediction of Soil Depth in an Area of Complex Topography in Taiwan. *CATENA*, 176: 419－429. https://doi. org/10.1016/j.catena.2019.01.030

Dietrich, W. E., Reiss, R., Hsu, M. L., et al., 1995. A Process-Based Model for Colluvial Soil Depth and Shallow Landsliding Using Digital Elevation Data. *Hydrological Processes*, 9(3－4): 383－400. https://doi. org/10.1002/hyp.3360090311

Ferrell, R. M., Devine, S., O'Geen, A. T., 2023. Trends in Regolith Thickness in a Headwater Catchment, Sierra Nevada, California. *Vadose Zone Journal*, 22(4): e20259. https://doi.org/10.1002/vzj2.20259

Gomes, G. J. C., Vrugt, J. A., Vargas, E. A., 2016. Toward Improved Prediction of the Bedrock Depth underneath Hillslopes: Bayesian Inference of the Bottom-up Control Hypothesis Using High-Resolution Topographic Data. *Water Resources Research*, 52(4): 3085－3112. https://doi.org/10.1002/2015wr018147

Han, X. L., Liu, J. T., Mitra, S., et al., 2018. Selection of Optimal Scales for Soil Depth Prediction on Headwater Hillslopes: A Modeling Approach. *CATENA*, 163: 257－275. https://doi. org/10.1016/j. catena.2017.12.026

Jia, Z., Cheng, Z. J., Chang, Z. L., et al., 2025.. Modeling and Uncertainty in Landslide Susceptibility Prediction Considering Coupling Mode of Landslide Types. *Earth Science*, 50(6): 2311－2329 (in Chinese with English abstract).

Khanifar, J., Khademalrasoul, A., 2022. Utilization of Multi-Scale Geomorphometric Algorithms and Third-Order Morphometric Variables for Soil Thickness Modeling. *Transactions in GIS*, 26(2): 887－901. https://doi.org/10.1111/tgis.12891

Li, X. C., Luo, J. H., Jin, X. L., et al., 2020. Improving Soil Thickness Estimations Based on Multiple Environmental Variables with Stacking Ensemble Methods. *Remote Sensing*, 12(21): 3609. https://doi. org/10.3390/rs12213609

Liang, X., Segoni, S., Yin, K. L., et al., 2022. Characteristics of Landslides and Debris Flows Triggered by Extreme Rainfall in Daoshi Town during the 2019 Typhoon Lekima, Zhejiang Province, China. *Landslides*, 19(7): 1735－1749. https://doi. org/10.1007/s10346-022-01889-5

Liu, F., Wu, H. Y., Zhao, Y. G., et al., 2022. Mapping High Resolution National Soil Information Grids of China. *Science Bulletin*, 67(3): 328－340. https://doi.org/10.1016/j.scib.2021.10.013

Lundberg, S. M., Erion, G., Chen, H., et al., 2020. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence*, 2(1): 56－67. https://doi. org/10.1038/s42256-019-0138-9

Luo, Z. D., Lian, J. J., Nie, Y. P., et al., 2024. Improving Soil Thickness Estimations and Its Spatial Pattern on Hillslopes in Karst Forests along Latitudinal Gradients. *Geoderma*, 441: 116749. https://doi.org/10.1016/j.geoderma.2023.116749

Patton, N. R., Shulmeister, J., Ellerton, D., et al., 2022. Measuring Landscape Evolution from Inception to Maturity: Insights from a Coastal Dune System. *Earth and Planetary Science Letters*, 584: 117448. https://doi.org/10.1016/j.epsl.2022.117448

Pelletier, J. D., Rasmussen, C., 2009. Geomorphically Based Predictive Mapping of Soil Thickness in Upland Watersheds. *Water Resources Research*, 45(9): 2008WR007319. https://doi. org/10.1029/2008wr007319

Peng, M., Wang, Y., Ma, C. Y., et al., 2025.. Review of Risk Assessment and Prevention for Valley Landslide Disaster Chains. *Earth Science*, 50(10): 3723－3760 (in Chinese with English abstract).

Pimenta Barbosa de Sousa, G., Tayebi, M., Rabelo Campos, L., et al., 2023. Improvement of Spatial Prediction

of Soil Depth *via* Earth Observation. *CATENA*, 223: 106915. https://doi.org/10.1016/j.catena.2023.106915

Shangguan, W., Hengl, T., Mendes de Jesus, J., et al., 2017. Mapping the Global Depth to Bedrock for Land Surface Modeling. *Journal of Advances in Modeling Earth Systems*, 9(1): 65－88. https://doi.org/10.1002/2016ms000686

Tarboton, D. G., 1997. A New Method for the Determination of Flow Directions and Upslope Areas in Grid Digital Elevation Models. *Water Resources Research*, 33(2): 309－319. https://doi.org/10.1029/96wr03137

Xiao, T., Segoni, S., Liang, X., et al., 2023. Generating Soil Thickness Maps by Means of Geomorphological-Empirical Approach and Random Forest Algorithm in Wanzhou County, Three Gorges Reservoir. *Geoscience Frontiers*, 14(2): 101514. https://doi. org/10.1016/j.gsf.2022.101514

Yan, F. P., Shangguan, W., Zhang, J., et al., 2020. Depth-to-Bedrock Map of China at a Spatial Resolution of 100 Meters. *Scientific Data*, 7(1): 2. https://doi. org/10.1038/s41597-019-0345-6

Zhang, X. L., Dao, L. G., Zhang, C. S., et al., 2018. Mapping the Spatial Distribution of Soil Depth in a Grassland Ecosystem with the Aid of Ground Penetrating Radar and GIS (Northwestern Sichuan, China). *Grassland Science*, 64(4): 217－225. https://doi. org/10.1111/grs.12201

Zhang, Y. H., Xu, X. L., Li, Z. W., et al., 2022. Modelling Soil Thickness Using Environmental Attributes in Karst Watersheds. *CATENA*, 212: 106053. https://doi.org/10.1016/j.catena.2022.106053

中文参考文献

贾卓，程志金，常志璐，等，2025.. 考虑滑坡类型耦合方式的滑坡易发性预测建模与不确定性分析 . 地球科学，50(6): 2311－2329.

彭铭，王悦，马晨议，等，2025.. 河谷滑坡灾害链风险评估及防控研究进展 . 地球科学，50(10): 3723－3760.