

王彬, 张甜, 饶路漫, 慈恩. 基于机器学习与地统计学融合的山地丘陵区土壤厚度预测[J]. 土壤学报, 2026,

WANG Bin, ZHANG Tian, RAO Luman, CI En. Prediction of Soil Thickness in Hilly and Mountainous Regions Based on the Integration of Machine Learning and Geostatistics[J]. Acta Pedologica Sinica, 2026,

基于机器学习与地统计学融合的山地丘陵区土壤厚度预测*

王彬, 张甜, 饶路漫, 慈恩[†]

(西南大学资源环境学院, 重庆 400715)

摘要: 土壤厚度是表征土壤资源状况的关键属性, 其准确的空间预测对理解土壤资源分布具有重要意义。针对传统预测模型在丘陵山区普适性差、纯机器学习模型忽略土壤空间自相关性的问题, 以渝西山地丘陵区为研究对象, 基于 117 个土壤剖面实测数据和 38 个环境协变量, 系统对比了多元线性回归 (MLR)、随机森林 (RF)、极限梯度提升 (XGBoost) 三种基准模型与随机森林残差克里格 (RFRK) 混合模型的土壤厚度预测精度。采用方差膨胀因子 (VIF)、Boruta 算法和递归特征消除 (RFE) 进行多阶段变量筛选, 从 38 个环境协变量中确定了 7 个核心变量。结果表明: (1) 集水区总面积 (TCA)、河网基准 (CNBL)、土壤增强指数 2 (SER2)、河网距离 (CND)、土壤增强指数 1 (SER1)、坡向 (ASP) 和坡度 (SLP) 是控制研究区土壤厚度空间分异的主导因素; (2) RFRK 模型预测精度最高 (决定系数 $R^2=0.68$, 林氏一致性相关系数 LCCC=0.79, 平均绝对误差 MAE=14.59 cm, 均方根误差 RMSE=19.86 cm), 显著优于三种基准模型。与性能最优的基准模型 RF 相比, RFRK 模型的 R^2 提升了 9.7%, MAE 降低了 12.3%, RMSE 降低了 6.2%。综上, 融合机器学习与地统计学的 RFRK 混合模型能更精确地刻画复杂地形区土壤厚度的空间格局, 可为区域尺度土壤资源评价与管理提供技术支撑。

关键词: 土壤厚度; 数字土壤制图; 随机森林残差克里格; 特征筛选; 空间预测

中图分类号: S151 文献标志码: A

Prediction of Soil Thickness in Hilly and Mountainous Regions Based on the Integration of Machine Learning and Geostatistics

WANG Bin, ZHANG Tian, RAO Luman, CI En[†]

(College of Resource and Environment, Southwest University, Chongqing 400715, China)

Abstract: 【Objective】The hilly region of western Chongqing is a transitional zone characterized by fragmented terrain and substantial elevation gradients, where soil thickness exhibits pronounced spatial variability driven by complex geomorphic processes. Soil thickness is a key indicator of soil resource status and plays a critical role in hydrological processes, ecosystem productivity, and geological hazard assessment. Traditional models used for soil thickness prediction often lack generalizability in hilly regions due to pronounced spatial heterogeneity, while pure machine learning approaches tend to overlook inherent spatial autocorrelation, potentially introducing local prediction biases. Therefore, accurate spatial prediction is essential for understanding soil resource distribution and supporting sustainable land management. 【Method】Based on 117 soil profiles

*国家自然科学基金项目 (41977002) 资助 Supported by the National Natural Science Foundation of China (No. 41977002)

[†] 通讯作者 Corresponding author, E-mail: cien777@163.com

作者简介: 王彬 (1999—), 男, 贵州贵阳人, 硕士研究生, 主要研究方向为数字土壤制图。E-mail: wangbin4425d@163.com

收稿日期: 2025-12-20; 收到修改稿日期: 2026-07-17; 网络首发日期 (www.cnki.net): 2026- - -

(2020–2025) and 38 environmental covariates spanning topography, climate, remote sensing, parent material, and human activities, three benchmark models were compared: multiple linear regression (MLR), random forest (RF), and extreme gradient boosting (XGBoost) against a hybrid model, random forest residual Kriging (RFRK). A multi-stage variable selection procedure incorporating variance inflation factor, boruta algorithm, and recursive feature elimination addressed multicollinearity and overfitting. Climate variables (1 km) were resampled to 12.5 m via bilinear interpolation. After excluding five outliers ($3\sigma/5\sigma$ criteria), 112 valid samples were partitioned into training (80%) and validation (20%) sets. Model performance was assessed using mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and Lin's concordance correlation coefficient (LCCC). For RFRK, variogram analysis characterized residual spatial structure, with model hyperparameters optimized through grid search. **【Result】** (1) The screening strategy reduced 38 covariates to seven core variables, ranked by importance: total catchment area (TCA) > channel network baseline (CNBL) > soil enhancement index 2 (SER2) > channel network distance (CND) > soil enhancement index 1 (SER1) > aspect (ASP) > slope (SLP). This subset (five topographic, two remote sensing) underscores terrain-driven material redistribution as the primary control on soil thickness. (2) RFRK achieved the highest accuracy ($R^2 = 0.68$, LCCC = 0.79, MAE = 14.59 cm, RMSE = 19.86 cm), significantly outperforming all benchmarks. Relative to the best benchmark (RF), RFRK improved R^2 by 9.7%, reduced MAE by 12.3%, and reduced RMSE by 6.2%. MLR performed the poorest ($R^2 = 0.37$), whereas XGBoost was intermediate ($R^2 = 0.56$). (3) Variogram analysis revealed strong residual spatial autocorrelation (nugget/sill = 0.0017; range = 179.08 m), justifying kriging for local bias correction. The predicted map showed “thick northeast/southwest, thin central,” with thick layers (>80 cm) in valleys and gentle slopes, and thin layers (0–40 cm) on steep ridges and areas with intense anthropogenic disturbance. **【Conclusion】** The RFRK hybrid model, integrating machine learning for nonlinear responses with geostatistical modeling of residual autocorrelation, characterizes soil thickness patterns in complex terrains more accurately than pure machine learning or linear models. This study validates the VIF-Boruta-RFE screening strategy and confirms RFRK's superiority in capturing both deterministic environmental controls and stochastic spatial structures. The findings provide robust technical support for soil resource assessment and management and offer a generalizable methodological reference for digital soil mapping in comparable hilly and mountainous landscapes worldwide.

Key words: Soil thickness; Digital soil mapping; Random forest residual Kriging; Feature selection; Spatial prediction

土壤厚度是表征土壤资源状况与生态功能的关键属性，也是土壤学、水文学及环境科学等多学科共同关注的基础参数，与区域水文循环、生态系统生产力及地质灾害风险等紧密相关^[1]。在土壤普查中是进行土壤类型划分和土种鉴别的关键诊断指标之一，也是耕地质量评价与保护中衡量土壤资源禀赋和可持续利用潜力的核心参量。传统的土壤厚度空间预测高度依赖专家经验，通过归纳土壤与环境变量间的知识，结合地理信息技术进行制图^[2]。然而，传统的基于专家经验的土壤厚度制图方法存在一定局限性：一方面，其高度依赖制图者的主观判断和经验知识，可重复性和标准化程度较低；另一方面，土壤厚度的形成与演化受成土母质、地形地貌、气候水文、生物作用及人类活动等多重因素的复杂耦合影响，使得基于有限经验的定性推断难以在区域尺度上准确捕捉土壤厚度的空间变异规律^[3]。

随着“数字土壤制图 (Digital soil mapping, DSM)”范式的发展，基于数据驱动的预测方法已成为主流，其大致可归纳为三类：基于物理过程的机理模型、基于空间自相关的地统计插值，以及基于环境协变量的经验统计模型^[4]。其中，以机器学习为代表的经验统计方法，因其强大的非线性拟合能力，能够有效揭示土壤属性与环境因子间的复杂关系，已成为应用最广泛的预测手段之一^[5-6]。这些模型通常基于土壤-气候-生物-地形-母质-成土年龄-空间位置 (SCORPAN) 框架^[7]，利用来自地形、气候、遥感影像等来源的环境协变量作为预测因子，以表征土壤形成与空间分布的主要成土要素。早期研究多采用多元线性回归 (MLR) 等线性模型，但难以刻画复杂的土壤形成过程^[8]。因此，近年来以随机森林 (RF) 为代表的机器学习算法，凭借其优秀的特征学习与抗过拟合能力，被广泛应用于土壤厚度预测中^[9-10]。然而，这些纯机器学习模型通常假设样本独立同分布，忽略了土壤

形成过程本身固有的空间依赖性,这可能导致对局部空间变异的预测出现偏差^[11]。

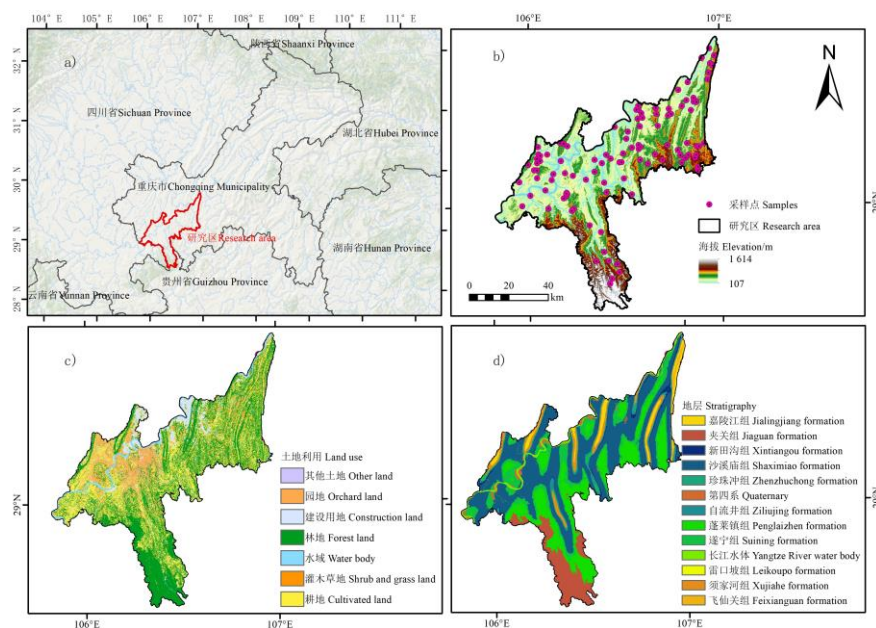
为克服上述局限,将机器学习算法与地统计方法相结合的混合模型被提出,其中,随机森林残差克里格(RFRK)通过将随机森林嵌入回归克里格框架,构建了一种能同时捕捉环境非线性效应与残差空间结构的混合模型^[12]。混合模型理论上通过整合环境与空间双重相关性,有望克服传统方法的局限。但是,其在土壤厚度预测制图领域的应用,特别是在中国西南复杂山地丘陵区的应用性与预测效能,尚缺乏充分的实证研究与系统评估。

因此,本研究以渝西山地丘陵区为研究区,旨在:(1)通过多阶段特征筛选策略,确定适用于土壤厚度预测建模的最优环境协变量组合;(2)构建并系统评估 MLR、RF、XGBoost 和 RFRK 四种模型在土壤厚度空间预测中的性能;(3)验证并阐明 RFRK 混合模型在本研究区的优越性与适用性,绘制土壤厚度空间分布图。

1 材料与方法

1.1 研究区概况

研究区位于渝西山地丘陵区,是四川盆地东南缘的川东平行岭谷与盆南丘陵过渡地带,地理坐标为 28°29′—29°45′N、105°49′—106°59′E,总面积约 5 043 km²。地处渝西长江南岸,地貌类型以低山、丘陵、河谷平坝为主,丘陵面积 2 622 km²,占比约 52%,低山占比约 38%,河谷平坝占比约 10%,区域内地貌破碎度大,地势起伏显著,为土壤厚度的空间分异提供了关键环境梯度。区域属中亚热带湿润季风气候,年均气温 18 °C,年均降水量 1 150 mm,集中且充沛的降水是影响土壤侵蚀与堆积的重要外营力。该区域主要地层单元包括侏罗系中统沙溪庙组、侏罗系上统遂宁组、侏罗系上统蓬莱镇组、白垩系下统夹关组、三叠系上统须家河组等^[13]。成土母质以泥岩、砂岩、灰岩、白云岩等风化残坡积物为主,在部分河流沿岸分布有少量第四系全新统冲积物,相应发育了紫色土、水稻土、黄壤、石灰岩土、潮土等多种类型土壤。土地利用方式多样,以林地(面积 2 168 km²,占比 43%)和耕地(面积 1 664 km²,占比 33%)为主,园地、水域、建设用地等亦有分布,其中,中低山区以林地为主,丘陵区广泛分布耕地与园地,河谷平坝区则以水田和建设用地为主。区内土壤分布不均,土壤厚度空间异质性显著,是探究山地丘陵区土壤厚度形成与分布规律的理想场所(图 1)。



注:该图基于国家测绘地理信息局标准地图服务网站下载的审图号为 GS (2024) 0650 号的标准地图制作,底图无修改。下同。Note:

This map is based on the standard map with the review number GS (2024) 0650, downloaded from the Standard Map Service website of the Ministry of Natural Resources of the People's Republic of China, with no modifications to the base map. The same below.

图 1 研究区概况图(a. 地理位置图, b. 海拔及样点分布图, c. 土地利用类型图, d. 地层分布图)

Fig. 1 Overview of the study area (a. geographical location map, b. elevation and sampling point distribution map, c. land use type map, d. stratum distribution map)

1.2 数据来源

1.2.1 土壤厚度的测量 本研究于 2020—2025 年开展了土壤厚度的野外测量工作, 基于“土地利用类型-海拔-差异化网格”三级布点法确定采样单元, 在 ArcGIS10.8 软件中进行空间叠加分析后, 以网格交点为参考, 确定地块中心为采样点。现场通过开挖土壤剖面, 分层记录 A 层、B 层的厚度并累加, 测量范围限定于母质层上方的土壤部分, 不包含母质层 (C 层) 及以下的基岩 (R) [14], 同时使用高精度 GPS 定位仪记录采样点经纬度坐标, 野外测量共获取研究区 117 个土壤剖面厚度数据。

1.2.2 环境协变量提取 本研究结合研究区实地情况, 选取地形、气候、遥感、母质特性和人类活动五类变量, 合计 38 个环境协变量, 并统一数据时点为 2023 年, 统一分辨率为 12.5 m。各环境协变量信息见表 1。

地形变量提取于 12.5 m 数字高程模型 (DEM), 数据来源于地理空间数据云平台 (<https://www.gscloud.cn/>), 利用 SAGA GIS8.1.3 软件和 ArcGIS10.8 软件计算得到 18 个常用地形协变量, 包括海拔 (HB)、山体阴影 (AH)、坡向 (ASP)、收敛指数 (CI)、河网基准 (CNBL)、河网距离 (CND)、坡长因子 (LSF)、平面曲率 (PLC)、剖面曲率 (PRC)、坡位 (PW)、相对位置指数 (RPI)、坡度 (SLP)、水动力指数 (SPI)、集水区总面积 (TCA)、地形位置指数 (TPI)、地形粗糙指数 (TRI)、地形湿度指数 (TWI) 和山谷深度 (VD)。其中, CNBL 指某一点相对于最近河网的垂直高度差, 反映该点物质侵蚀和堆积的潜在势能, 低值区 (接近河网) 代表物质汇集的洼地, 高值区代表物质输出的源地 [15]。CND 指某一点到最近河网的水平欧氏距离, 表征该点与水文网络的邻近关系, 间接反映土壤水分状况和沉积条件, 通常距离河网越近, 土壤水分越充足, 堆积作用越强 [15]。TCA 指某一点上游汇流区的累计面积, 表征该点承接的上坡来水、来沙通量, 值越大表明上游汇水、汇沙能力越强, 潜在的堆积作用也越强 [16]。

气候变量采用 WorldClim 数据库 (<https://www.worldclim.org/data/worldclim21.html>) 中的 2000—2023 年逐年栅格数据, 并取平均值得到包括年温差 (ATD)、年蒸发量 (EVP)、年降水量 (PRE)、年平均气压 (PRS)、年太阳总辐射 (RAD)、年平均气温 (TEM) 和年平均风速 (WIN) 共 7 个指标, 表征研究区气候特征, 空间分辨率为 1 km。为消除空间分辨率差异, 利用 ArcGIS10.8 软件, 采用双线性插值方法将所有气候变量重采样至 12.5 m 空间分辨率, 以匹配地形和遥感变量。

遥感变量基于 Google Earth Engine (GEE, <https://code.earthengine.google.com/>) 平台获取的 Sentinel-2MSI 卫星遥感影像, 通过波段计算得到归一化植被指数 (NDVI)、增强植被指数 (EVI)、两个叶面积指数 (LAI_{gy}、LAI_{lc})、土壤调节植被指数 (SAVI)、三个土壤增强指数 (SER1、SER2、SER3)、地表水分指数 (LSWI) 和修正型归一化差异水体指数 (MNDWI) 共 10 个变量, 用于表征植被覆被及土壤状况 [17-18]。相关公式如表 2 所示。

表 1 研究区环境协变量的选择

Table 1 Selection of environmental covariates in the study area

序号 No.	类别 Category	缩写 Abbreviation	协变量名称 Covariate name	空间分辨率 Spatial resolution
1	地形 Topography	HB	海拔 Elevation	12.5 m
2	地形 Topography	AH	山体阴影 Mountain shadow	12.5 m
3	地形 Topography	ASP	坡向 Aspect	12.5 m
4	地形 Topography	CI	收敛指数 Convergence index	12.5 m
5	地形 Topography	CNBL	河网基准 Channel network baseline	12.5 m
6	地形 Topography	CND	河网距离 Channel network distance	12.5 m

土 壤 学 报
Acta Pedologica Sinica

序号 No.	类别 Category	缩写 Abbreviation	协变量名称 Covariate name	空间分辨率 Spatial resolution
7	地形 Topography	LSF	坡长因子 Length slope factor	12.5 m
8	地形 Topography	PLC	平面曲率 Plan curvature	12.5 m
9	地形 Topography	PRC	剖面曲率 Profile curvature	12.5 m
10	地形 Topography	PW	坡位 Position on slope	12.5 m
11	地形 Topography	RPI	相对位置指数 Relative position index	12.5 m
12	地形 Topography	SLP	坡度 Slope	12.5 m
13	地形 Topography	SPI	水动力指数 Stream power index	12.5 m
14	地形 Topography	TCA	集水区总面积 Total catchment area	12.5 m
15	地形 Topography	TPI	地形位置指数 Topographic position index	12.5 m
16	地形 Topography	TRI	地形粗糙指数 Terrain ruggedness index	12.5 m
17	地形 Topography	TWI	地形湿度指数 Topographic wetness index	12.5 m
18	地形 Topography	VD	山谷深度 Valley depth	12.5 m
19	气候 Climate	ATD	年温差 Annual temperature difference	1 km
20	气候 Climate	EVP	年蒸发量 Annual evaporation	1 km
21	气候 Climate	PRE	年降水量 Annual precipitation	1 km
22	气候 Climate	PRS	年平均气压 Annual average air pressure	1 km
23	气候 Climate	RAD	年太阳总辐射 Annual total solar radiation	1 km
24	气候 Climate	TEM	年平均气温 Annual average temperature	1 km
25	气候 Climate	WIN	年平均风速 Annual average wind speed	1 km
26	遥感 Remote sensing	EVI	增强植被指数 Enhanced vegetation index	12.5 m
27	遥感 Remote sensing	LAIgy	叶面积指数 1Leaf area index 1	12.5 m
28	遥感 Remote sensing	LAIlc	叶面积指数 2Leaf area index 2	12.5 m
29	遥感 Remote sensing	NDVI	归一化植被指数 Normalized difference vegetation index	12.5 m
30	遥感 Remote sensing	SAVI	土壤调节植被指数 Soil-adjusted vegetation index	12.5 m
31	遥感 Remote sensing	SER1	土壤增强指数 1Soil enhancement index 1	12.5 m
32	遥感 Remote sensing	SER2	土壤增强指数 2Soil enhancement index 2	12.5 m
33	遥感 Remote sensing	SER3	土壤增强指数 3Soil enhancement index 3	12.5 m
34	遥感 Remote sensing	LSWI	地表水分指数 Land surface water index	12.5 m
35	遥感 Remote sensing	MNDWI	修正型归一化差异水体指数 Modified ND water index	12.5 m
36	母质特性 Parent material properties	DC	地层 Stratigraphic layer	12.5 m
37	母质特性 Parent material properties	YX	岩性 Lithology	12.5 m
38	人类活动 Human activities	TDLY	土地利用 Land use	12.5 m

表 2 遥感变量计算公式

Table 2 Calculation formulas of remote sensing variables

缩写 Abbreviation	计算公式 Calculation formula	相关说明 Relevant description	缩写 Abbreviation	计算公式 Calculation formula	相关说明 Relevant description
NDVI	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	表征植被特性最有 效的参数	SER1	$\frac{\text{Red}}{\text{Green}}$	对土壤中碳酸岩 类成分比较敏感
EVI	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1}$	有效减少气溶胶的 影响	SER2	$\frac{\text{Red}}{\text{SWIR2}}$	对土壤的氧化还 原环境区分敏感

LAI _{gy}	$LAI = 0.57 \times \exp(2.33 \times NDVI)$	适用林地、草地和灌木	SER3	$\frac{SWIR1 - Green}{SWIR1 + Green}$	对沉积岩类矿物成分敏感
LAI _{lc}	$LAI = 3.618 \times NDVI - 0.118$	适用于耕地、园地和其他	LSWI	$\frac{NIR - SWIR1}{NIR + SWIR1}$	对叶片辐射和叶片水分敏感
SAVI	$\frac{NIR - Red}{NIR + Red + 0.5}$	能有效降低土壤背景噪声的影响	MNDWI	$\frac{Green - SWIR1}{Green + SWIR1}$	加强开放水域特征, 抑制植被、土壤、建筑物的噪音

注: NIR 为近红外波段; Red 为红光波段; Green 为绿光波段; Blue 为蓝光波段; SWIR 为短波红外波段; SWIR1 为第一短波红外波段。具体波段中心波长参见 Sentinel-2 卫星遥感影像相关技术文档。Note: NIR, near-infrared band; Red, red band; Green, green band; Blue, blue band; SWIR, short-wave infrared band; SWIR1, the first short-wave infrared band. For specific band center wavelengths, refer to the technical documentation of Sentinel-2 satellite remote sensing imagery.

1.3 研究方法

本研究选取了四种模型进行对比: 基于线性假设的多元线性回归模型, 并在纯机器学习算法中选取了经典的随机森林模型和捕捉特征关系更加复杂的极限梯度提升模型, 以及融合机器学习与统计优势的随机森林残差克里格模型。其中, 前三种为基准模型, 随机森林残差克里格为混合模型, 系统比较四种模型算法在土壤厚度空间预测中的适用性与优劣。

1.3.1 模型方法 多元线性回归 (Multiple linear regression, MLR) 是一种经典的统计建模方法, 旨在通过线性组合多个自变量来预测因变量。该模型假定因变量与自变量之间存在线性关系, 并且残差独立分布。MLR 模型结构简洁、解释性强, 能够量化各环境协变量对目标变量的边际影响。本研究采用 scikit-learn 库中的线性回归 (Linear regression) 实现多元线性回归。为消除不同环境协变量量纲的影响, 所有自变量均经过标准化处理, 转换为均值为 0、标准差为 1 的分布, 模型通过最小二乘法进行参数估计。

随机森林 (RF) 是一种基于决策树的集成机器学习算法。该算法通过构建多棵决策树进行预测, 并通过自助采样法 (Bootstrap) 和随机特征选择来降低过拟合风险, 在处理高维数据时表现出色。在本研究中, 采用 scikit-learn 库中的随机森林回归器 (Random forest regressor) 实现随机森林回归。森林中决策树的数量 (n estimators) 设为 100, 最小叶子节点样本数 (min samples leaf) 设为 1。

极限梯度提升 (eXtreme Gradient Boosting, XGBoost) 是一种先进的梯度提升决策树算法。该算法通过迭代训练一系列决策树, 其中, 每棵树致力于修正前一轮预测的残差, 从而构建一个强大的集成模型。XGBoost 在目标函数中引入了正则化项以控制模型复杂度, 有效防止过拟合, 并支持并行处理, 具有极高的计算效率和优异的预测精度。本研究采用 XGBoost 库实现回归预测, 提升树的数量 (n estimators) 为 50, 学习率 (learning rate) 为 0.05, 树的最大深度 (max depth) 为 3, 目标函数 (objective) 设为 reg: squarederror, 并使用全部可用线程 (n jobs=-1) 进行并行计算, 以提升效率。

随机森林残差克里格 (Random forest with residual Kriging, RFRK) 是一种结合随机森林 (RF) 与克里格 (OK) 插值的混合空间预测方法^[19]。该方法利用 RF 捕捉目标变量与多源环境协变量之间的非线性关系, 并对 RF 的预测残差进行 OK 插值, 以量化其空间自相关性, 并对局部进行偏差修正。本研究采用 PyKrig 库中的 Regression Kriging 实现 RFRK, 决策树数量为 300, 最大树深度为 10, 最小叶子节点样本数为 2, 节点分裂时考虑的特征数 (max features) 为 auto。克里格部分选用球状变异函数模型 (spherical), 最近邻点数为 30。

1.3.2 环境协变量筛选与数据集构建 在构建土壤厚度模型时, 过多且存在多重共线性的环境协变量不仅会增加模型复杂度, 还易导致过拟合与预测性能下降。为优化模型输入、增强模型稳健性与可解释性, 本研究采用了一套多阶段的环境协变量筛选与数据清洗策略, 且研究所有模型构建与预测制图的最终输出成果的空间分辨率均为 12.5 m。

(1) 土壤样本异常值识别与处理。为确保建模数据的质量,对土壤厚度实测值进行了严格的异常值筛查。分别采用“平均值 ± 5 倍标准差”与“平均值 ± 3 倍标准差”两种阈值进行双重检验。首先,利用 5 倍标准差的标准识别并剔除极端异常值;随后,再利用 3 倍标准差的标准进行二次筛选,以移除潜在的强影响点。此流程在控制第一类错误(误删正常值)与第二类错误(遗漏异常值)之间取得了良好平衡,有效提升数据集的纯净度。

(2) 多重共线性诊断与方差膨胀因子筛选(Variance inflation factor, VIF)。为量化环境协变量间的多重共线性,计算了所有初始协变量的方差膨胀因子。VIF 量化了一个协变量被其他协变量线性解释的程度,通常以 $VIF > 10$ 作为存在严重多重共线性的阈值,逐步剔除 VIF 值最高的协变量,直至剩余协变量的 VIF 均满足要求。

(3) 基于 Boruta 算法的特征重要性评估(Boruta)。在缓解多重共线性之后,采用基于随机森林的 Boruta 算法对所有通过 VIF 检验的协变量进行重要性判定。Boruta 算法通过创建特征的阴影副本,并比较原始特征与阴影特征的重要性得分,进行统计显著性检验,从而进行特征选择^[20]。

(4) 基于递归特征消除的最终优化(RFE)。为进一步精炼特征集并优化模型性能,采用了递归特征消除算法。RFE 通过递归地构建模型(本研究以随机森林回归为基模型)并剔除最不重要的特征,最终根据交叉验证的模型精度确定最优特征数量^[21]。此步骤旨在以最少的协变量数量实现最优或接近最优的预测精度,达成了降维与性能保障的平衡。

1.3.3 模型精度评价 为评价模型在土壤厚度空间分布预测中的准确性,本研究所选用的四个统计指标为:平均绝对误差(Mean absolute error, MAE)、均方根误差(Root mean square error, RMSE)、决定系数(Coefficient of determination, R^2),以及林氏一致性相关系数(Lin's concordance correlation coefficient, LCCC)。MAE 和 RMSE 从不同角度度量预测值与实际值之间的偏差,二者值越低,代表模型的整体预测精度越高; R^2 值越高(越接近 1),表明模型对观测数据的拟合优度越好,预测值与实际值的相关性越强;LCCC 则综合衡量预测值与实际值的一致性,同时反映准确度与精密度,取值越接近 1,代表二者围绕 1:1 理想吻合线的分布越集中,模型预测结果的可靠性越强。当 MAE、RMSE 值较低,且 R^2 、LCCC 值较高时,模型性能则越优。各指标计算公式详见参考文献^[22]。

2 结果

2.1 土壤厚度描述性统计

结合数据分布特征,对 117 个土壤厚度原始样点进行异常值检测与剔除,分别采用 5 倍标准差法和 3 倍标准差法筛选无效数据,最终去除 5 个异常样本,保留 112 个有效样点,按 8:2 的比例划分训练集和测试集用于后续统计分析。土壤厚度的描述性统计分析结果详见表 3。土壤厚度的总体范围介于 13.00~133.00 cm,平均值 87.46 cm,中位数 103.00 cm,标准差 38.11 cm,变异系数 43.57%,偏度系数为-0.44,整体呈轻微左偏分布(多数样本集中在高值区域),无显著极端偏态特征,说明研究区土壤厚度以较高值为主,数据分布稳定性较好,能为后续模型构建提供可靠基础。

表 3 研究区采样点土壤厚度描述性统计

Table 3 Descriptive statistics of soil thickness at the sampling sites in the study area

样本集	样点数	最小值	最大值	平均值	标准差	变异系数	偏度
Sample set	Sample N	Minimum/ cm	Maximum/ cm	Mean/ cm	Standard deviation/ cm	Coefficient of variation/ %	Skewness
t	n	m	m	m	m	%	s
总样本 ^①	112	13.00	133.00	87.46	38.11	43.57	-0.44
训练集 ^②	89	13.00	133.00	87.55	38.30	43.75	-0.44
验证集 ^③	23	21.00	125.00	85.04	38.36	45.10	-0.02

①>Total sample, ②/Training set, ③/Validation set.

2.2 环境协变量筛选与特征重要性

通过多阶段筛选策略,从 38 个初始环境变量中确定了 7 个核心变量参与建模(图 2),各环境协变量缩写对应的完整名称详见表 1。VIF 诊断结果表明,经过逐步剔除后剩余变量的 VIF 值均小于 10,变量间不存在严重多重共线性(图 2a)。Boruta 算法重要性评估从 26 个 VIF 筛选后的协变量中识别出 21 个重要变量(重要性显著高于影子特征),其中,地形变量(CNBL、TCA)、遥感变量(SER2)的重要性得分位居前列(图 2b)。RFE 交叉验证结果表明,当保留 7 个协变量时,模型性能最优(平均交叉验证 $R^2=0.33$),继续增加变量会导致模型复杂度上升而精度提升不明显(图 2c)。

最终确定的 7 个核心环境协变量及其重要性排序为: $TCA > CNBL > SER2 > CND > SER1 > ASP > SLP$ (图 2d)。该组合涵盖了地形变量(5 个)和遥感变量(2 个),这表明研究区土壤厚度的空间分异主要受控于地形主导的物质迁移过程和地表物质组成决定的土壤发育基础(图 3)。

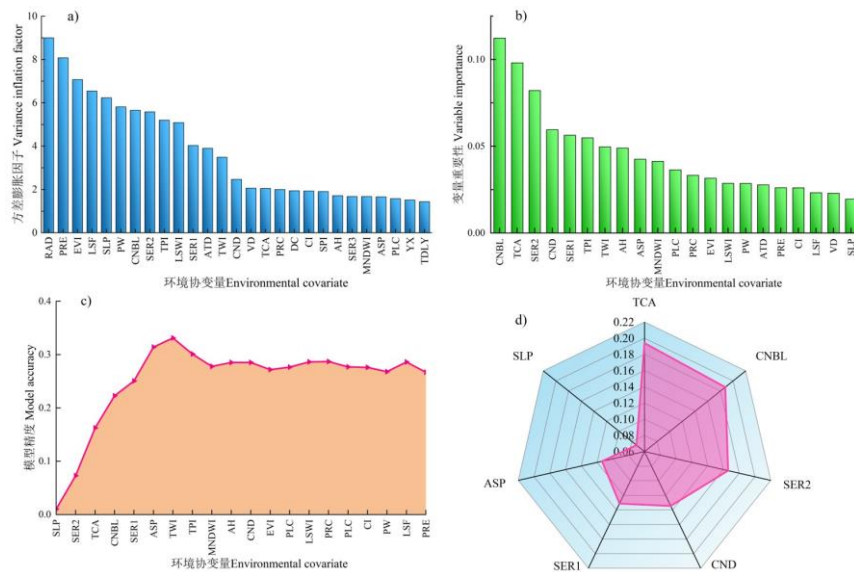


图 2 环境协变量筛选及特征重要性 (a. 方差膨胀因子 (VIF), b. Boruta 算法, c. 递归特征消除 (RFE), d. 最优组合)

Fig. 2 Environmental covariate screening and feature importance (a. Variance inflation factor (VIF), b. Boruta algorithm, c. Recursive feature elimination (RFE), d. Optimal combination)

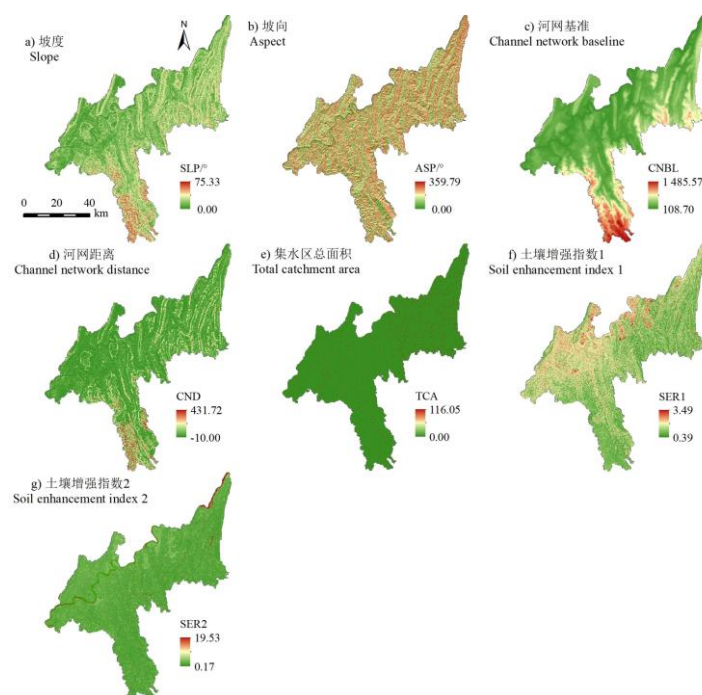


图3 研究区内7个环境协变量的空间分布

Fig. 3 Spatial distribution of the 7 environmental covariates within the study area

2.3 模型精度评估与验证

四种模型的精度评估结果汇总于表4。总体而言,所有模型均展现出一定的预测能力,但机器学习模型RF、XGBoost与RFRK的性能显著优于MLR,证实了机器学习方法在捕捉土壤厚度与环境因子间复杂非线性关系方面的优势。

RFRK模型在所有评估指标上均表现最佳,其 R^2 与LCCC分别达到0.68与0.79,而MAE与RMSE则降至14.59 cm与19.86 cm。该结果表明,RFRK模型不仅能够解释土壤厚度大部分的空间变异,其预测值与实测值也表现出高度的一致性,且预测偏差最小,综合性能最为优越。这得益于其结合了随机森林处理非线性关系的强大能力与克里格插值对空间自相关残差的修正作用。RF模型表现次之,其预测精度($R^2=0.62$)与一致性(LCCC=0.78)均低于RFRK,但显著优于MLR和XGBoost,凸显了其作为经典机器学习算法的稳健性。XGBoost模型尽管在解释方差($R^2=0.56$)上优于MLR,但其预测误差(MAE=19.62 cm, RMSE=25.09 cm)相对较大,表明其预测结果存在较高的不确定性。作为对比,MLR模型的各项指标均为最低,这进一步印证了在研究区内,土壤厚度的形成与分布由复杂的非线性过程主导,而线性模型难以有效表征该机制。

为直观展示各模型的预测性能,绘制了四种模型的预测值与实测值的散点图(图4),结果表明,MLR模型散点离散程度最大,拟合线明显偏离1:1理想线,表明线性模型难以刻画土壤厚度与环境变量间的非线性关系;RF、XGBoost和RFRK模型的散点均紧密围绕1:1理想线,其中RFRK模型拟合线与1:1理想线重合度较高。模型精度对比确认了融合机器学习和地统计学优势的混合模型——RFRK为本研究中最可靠的土壤厚度空间预测模型,展现出优越的土壤厚度预测能力。

为进一步验证RFRK模型的优越性来源,对RFRK模型预测残差的空间结构进行分析。计算RFRK模型残差的变异函数参数,结果表明,残差具有显著的空间自相关结构:块金值为195.15,基台值为114 871,变程为179.08 m,块金基台比仅为0.001 7。根据空间统计学标准,块金基台比小于0.25表明残差具有强空间自相关性^[23],说明RFRK模型在拟合环境因子主导的不确定性趋势后,剩余残差中仍有99.8%以上具有可建模的结构性成分。这表明RFRK模型中的克里格插值部分能够有效捕捉残差的空间自相关结构,从而对局部预测进行精确修正。这也解释了RFRK模型($R^2=0.68$)优于纯RF模型($R^2=0.62$)的原因,验证了在具有强空间自相关性的研究区,融合机器学习与地统计学的混合模型具有显著优势。

表4 不同预测模型精度指标特征统计

Table 4 Characteristics and statistical metrics of prediction accuracy for different models

模型 Model	MAE	RMSE	R^2	LCCC
MLR	21.85	26.31	0.37	0.48
RF	16.64	21.17	0.62	0.78
XGBoost	19.62	25.09	0.56	0.69
RFRK	14.59	19.86	0.68	0.79

注: MAE 为平均绝对误差; RMSE 为均方根误差; R^2 为决定系数; LCCC 为林氏一致性相关系数。MLR: 多元线性回归; RF: 随机森林; XGBoost: 极限梯度提升; RFRK: 随机森林残差克里格。下同。Note: MAE, mean absolute error; RMSE, root mean square error; R^2 , coefficient of determination; LCCC, Lin's concordance correlation coefficient. MLR, multiple linear regression; RF, random forest; XGBoost, extreme gradient boosting; RFRK, random forest residual kriging. The same below.

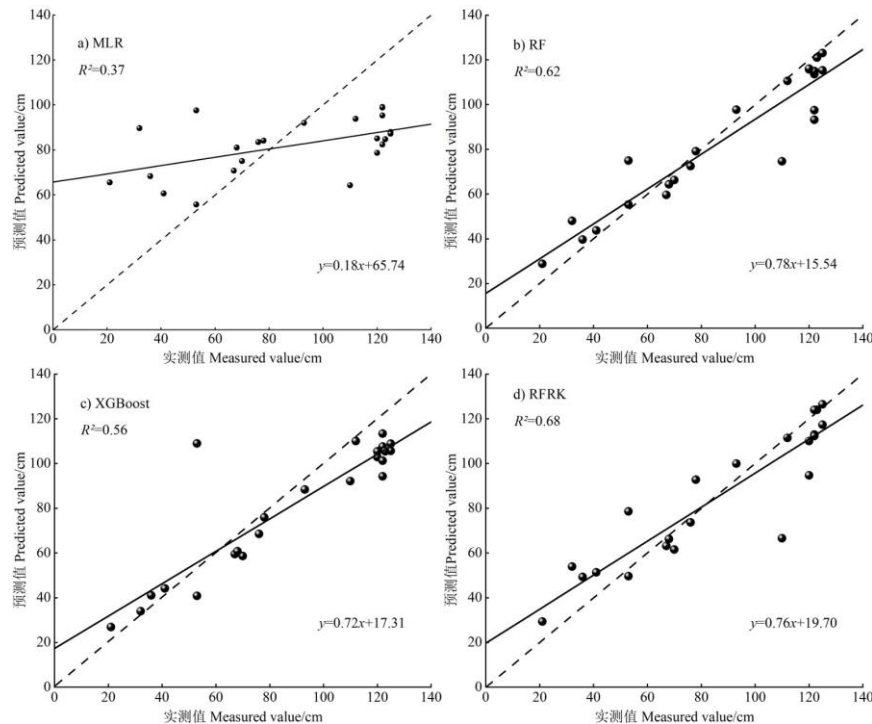


图 4 不同模型预测值与实测值散点图（虚线为 1:1 理想线，实线为实际拟合线）

Fig. 4 Scatter plots of predicted vs. measured values for different models (the dashed line represents the 1:1 ideal line, and the solid line represents the actual fitted line)

2.4 空间预测制图

基于 MLR、RF、XGBoost 三种基准模型和混合模型 RFRK 对研究区土壤厚度进行空间预测制图，并采用 0~40 cm（薄层）、40~80 cm（中层）、80 cm 以上（厚层）三级划分标准（图 5），空间分辨率为 12.5 m。就空间格局而言，RF、XGBoost 与 RFRK 三种机器学习模型的预测结果宏观趋势基本一致，均清晰反映了研究区土壤厚度“东北厚、中部薄、西南厚”的区域性特征及“沟谷厚、坡地薄”的局部地形控制规律。厚层土壤主要分布于向斜谷地、河谷平坝等低洼堆积区，薄层土壤则集中在背斜山脉的高海拔陡坡区域，中层土壤广泛分布于二者之间的过渡地带。相比之下，MLR 模型预测结果中高值和低值区域均向中值范围收缩，对山地丘陵区强烈空间异质性的刻画能力明显不足。

就三种机器学习模型的统计特征而言（表 5），预测的土壤厚度最高值为 132.48 cm，最低值 14.12 cm，均能较好地反映实际地形与植被覆盖对土壤发育的影响。三种基准模型与混合模型预测的土壤厚度均值介于 70.67 cm 至 79.00 cm 之间，与中值范围（70.46~79.64 cm）基本吻合，说明各模型在估计上未出现系统性偏差。值得注意的是，混合模型 RFRK 预测结果的标准差明显低于 RF 和 XGBoost，表明其预测值的离散程度更小，结果更为集中和稳定。该结果说明 RFRK 模型通过耦合随机森林与克里格插值，在保持预测准确性的同时，有效降低了结果的不确定性，从而在复杂景观背景下更稳健地捕捉土壤厚度的空间分布结构。

基于 RFRK 模型预测结果（图 5d），中层土（40~80 cm）为研究区主导类型，占比 52.65%（表 6），广泛分布于区域缓坡过渡带；厚层土（>80 cm）占比 46.56%，与中层土比例接近，主要集中在向斜谷地等低洼区域；薄层土（0~40 cm）占比极低（0.79%），属于零星分布类型，多局限于背斜轴部的陡峭山脊、裸岩边坡及人为破坏严重的区域。

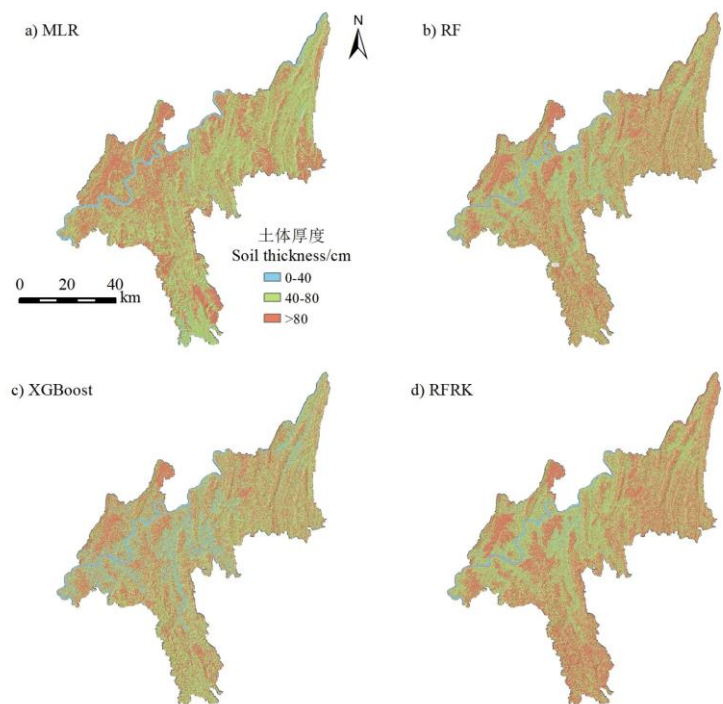


图 5 基于 MLR (a)、RF (b)、XGBoost (c) 与 RFRK (d) 预测的土壤厚度空间分布
Fig. 5 Spatial distribution of soil thickness predicted by MLR (a), RF (b), XGBoost (c), and RFRK (d)

表 5 不同预测模型预测土壤厚度的统计结果

Table 5 Statistical results of soil thickness predictions by different prediction models

模型	最小值	中值	最大值	均值	标准差
Model	Minimum/ cm	Median/ cm	Maximum/ cm	Mean/ cm	Standard deviation/ cm
MLR	0.00	72.27	125.00	75.04	15.96
RF	24.96	77.35	127.81	76.29	21.07
XGBoost	14.12	70.46	114.28	70.67	21.40
RFRK	25.12	79.64	132.48	79.00	19.54

表 6 RFRK 土壤厚度分级统计结果

Table 6 Statistical results of soil thickness classification by the RFRK model

分级	范围	像元数	像元大小	面积	占比
Classification	Range/ cm	Pixel count	Pixel size/ m	Area/ km ²	Proportion/ %
薄层 Thin layer	0-40	211 573	12.5×12.5	39.87	0.79
中层 Medium layer	40-80	16 782 223	12.5×12.5	2655.80	52.65
厚层 Thick layer	>80	15 287 317	12.5×12.5	2348.20	46.56

3 讨论

根据 Jenny^[24]提出的成土因素方程，结合 McBratney 等^[5]扩展的 SCORPAN 框架，研究选取了地形、气候、生物、母质特性及人类活动五类环境协变量。合理的环境协变量组合有助于提取关键土壤信息，提升土壤厚度异质性的解释能力和建模精度。张晓婷^[25]研究同样表明，经过协变量筛选的模型较包含全部环境协变量的模型具有更强健的性能。通过“VIF-Boruta-RFE”多阶段筛选策略，

有效实现了从“去除冗余”到“辨识重要”再到“优化组合”的变量精炼过程。基于 7 个核心环境协变量的特征重要性分析,坡度 (SLP)、坡向 (ASP)、河网基准 (CNBL)、河网距离 (CND)、集水区总面积 (TCA)、土壤增强指数 1 (SER1) 和土壤增强指数 2 (SER2) 被识别为控制土壤厚度空间分异的主导因子 (图 2)。这些因子通过不同的土壤发生学过程共同调控土壤厚度的形成与演化。

就地形因子的作用机制而言, TCA 在本研究中重要性较高 (图 2d), 表明研究区土壤厚度的宏观格局主要由汇流势能决定。汇水面积越大, 上坡来水来沙通量越高, 物质堆积潜力越强, 这与坡地土壤演化模型中的“汇流面积-土壤厚度”正相关关系吻合^[16]。CNBL 与 CND 共同表征了物质运移的“源-汇”关系^[26]: CNBL 低值区代表接近侵蚀基准面的洼地, 是物质汇集的“汇”; CND 低值区代表靠近河网的区域, 土壤水分充足、沉积作用占优^[27]。二者重要性排序靠前, 印证了研究区土壤厚度受控于“坡地侵蚀-沟谷堆积”的水沙再分配过程。SLP 重要性排序靠后 (图 2d), Fang 等^[10]也发现坡度可能不是预测区域土壤厚度的关键因素, 可能因其信息已部分被 TCA、CNBL 等水文地形参数所吸收, 且综合的汇流势能 (CNBL+TCA) 较单纯的局地动力 (SLP) 能更好地解释土壤厚度的堆积模式。但就土壤发生学角度而言, 坡度仍是控制坡面物质稳定性的基础因子, 坡度越大, 重力侵蚀潜力越强, 土壤滞留时间越短。此外, 在相同地形条件下, 地表物质组成的差异 (源于母岩类型、风化程度或地球化学过程) 会显著调节土壤的发育速率与保存潜力^[28]。Chen 等^[29]指出, 遥感变量是土壤厚度预测的重要预测因子, SER1 与 SER2 作为遥感衍生的地表物质敏感指数, 分别对碳酸盐成分和土壤氧化还原环境具有指示意义^[30]。SER2 的重要性高于 SER1 (图 2d), 表明在本研究区, 由水分条件驱动的氧化还原状态差异对土壤厚度的反映较碳酸盐含量更为显著, SER2 高值区土壤水分充足, 有利于风化产物的保存和土壤物质积累, 实质上反映了由水分驱动的“残积-坡积”耦合过程对山地丘陵区土壤厚度空间格局的关键塑造作用。

上述因子并非独立作用, 而是通过“地形驱动水沙运移-地表物质决定土壤发育基础”的耦合机制共同塑造土壤厚度格局^[31]。该机制在土壤发生学上可概括为: 地形创造了物质再分配的动力条件, 地表物质提供了土壤发育的物质基础, 二者的耦合决定了特定地貌位置上土壤形成速率与侵蚀速率的平衡点, 最终表现为土壤厚度的空间分异。在川东平行岭谷的构造地貌背景下, 这种耦合机制尤为显著, 该发现与 Yan 等^[32]研究结果一致, 进一步证实了地形主导的物质迁移是控制土壤厚度宏观格局的普遍机制。

模型精度呈现 $MLR < XGBoost < RF < RFRK$ 的递进关系 (表 4), 该结果深刻反映了不同算法对土壤形成过程复杂性的刻画能力差异。MLR 作为线性模型, 难以表征土壤厚度与环境因子间复杂的非线性关系, 以 RF 和 XGBoost 为代表的机器学习模型能够有效学习这种非线性关系, 预测能力显著优于 MLR^[33-34]。但是, 这些纯机器学习模型通常基于样本独立同分布的假设, 未能有效抓取土壤形成过程中固有的空间依赖性。RFRK 模型的优越性在于其理论框架的完整性^[19]: 其随机森林部分负责拟合环境因子主导的确定性趋势, 捕捉宏观非线性格局; 而克里格插值则对残差中具有空间自相关的随机成分进行建模和局部修正。Wang 等^[35]指出, 这种“趋势面拟合+空间残差校正”的协同机制, 使 RFRK 能够更全面地刻画土壤厚度空间分异中受环境控制和受局部随机过程共同作用的双重特征。通过对比四种模型预测的散点图和空间分布图 (图 4), 发现 MLR 对土壤厚度的线性拟合有限, 而 XGBoost、RF 和 RFRK 预测的土壤厚度空间分布总体趋势一致, 但局部细节上存在差异, 这与 Li 等^[36]的研究结论一致。根据 RFRK 预测的土壤厚度分级空间布局, 厚层土壤主要分布于平台处及坡地前沿, 薄层土则分布在陡峭山脊处, 符合 Xiao 等^[37]山坡系统土壤厚度随坡度梯度升高而降低的研究结果。

当前研究仍存在一定的局限性, 未来可从以下方面深入探索。首先, 环境协变量的表征维度可进一步拓展。当前使用的多为静态、近地表的环境指标, 未能充分纳入时间动态 (如长期土地利用变迁、植被演替) 和地下三维结构 (如基岩埋深、岩溶网络) 信息。其次, 空间预测场景可向“无样点”或“少样点”延伸。当前模型的训练与验证高度依赖研究区内有限的实测样点, 对于完全缺乏土壤样点的未知区域, 其预测结果的迁移性与可靠性面临挑战。因此, 发展不依赖本地样本或仅

需少量样本即可进行高精度预测的迁移学习或元学习框架, 将是今后土壤厚度预测研究的重点。

4 结 论

本研究通过多阶段特征筛选确定了 7 个核心环境协变量, 证实了研究区土壤厚度主要受地形主导的物质迁移过程控制。在对比的四种模型中, 融合机器学习与地统计学的随机森林残差克里格 (RFRK) 混合模型表现最优, 表明同时考虑环境因子非线性效应与残差空间自相关的建模框架具有显著优势。该方法在复杂山地丘陵区具有良好的适用性, 可为区域土壤资源精准管理提供技术支撑。

参考文献 (References)

- [1] Liu J T, Zhao W, Liu Y. Modelling soil thickness evolution: Advancements and challenges[J]. *Acta Pedologica Sinica*, 2024, 61(2): 319-330. [刘金涛, 赵薇, 刘彧. 土壤厚度演化模型理论方法研究进展[J]. *土壤学报*, 2024, 61(2): 319-330.]
- [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] Malone B, Searle R. Improvements to the Australian national soil thickness map using an integrated data mining approach[J]. *Geoderma*, 2020, 377: 114579.
- [4] Zhu Q L, Han Z, Fang F Y, et al. Soil thickness prediction models: Types, accuracy and influencing factors[J]. *Catena*, 2025, 258: 109282.
- [5] McBratney A B, Santos M L M, Minasny B. On digital soil mapping[J]. *Geoderma*, 2003, 117(1/2): 3-52.
- [6] 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.]
- [7] Minasny B, McBratney A B. Digital soil mapping: A brief history and some lessons[J]. *Geoderma*, 2016, 264: 301-311.
- [8] McKenzie N J, Ryan P J. Spatial prediction of soil properties using environmental correlation[J]. *Geoderma*, 1999, 89(1/2): 67-94.
- [9] Meng K, Huang W, Fu P H, et al. Comparison of digital soil mapping methods in plain and hill mixed regions[J]. *Acta Pedologica Sinica*, 2025, 62(4): 958-969. [孟可, 黄魏, 傅佩红, 等. 平原-丘陵区域数字土壤制图方法比较[J]. *土壤学报*, 2025, 62(4): 958-969.]
- [10] Fang F Y, Zi R Y, Han Z, et al. Estimation of soil thickness in karst landforms using a quantile regression forests approach[J]. *Geoderma*, 2025, 460: 117416.
- [11] Chen Y L, Liang T B, Zhang Y L, et al. Spatial prediction of soil thicknesses in Sichuan Province based on feature-ensemble learning[J]. *Soils*, 2023, 55(4): 894-902. [陈玉蓝, 梁太波, 张艳玲, 等. 基于特征集成学习的四川省土壤厚度预测[J]. *土壤*, 2023, 55(4): 894-902.]
- [12] De Benedetto D, Barca E, Castellini M, et al. Prediction of soil organic carbon at field scale by regression Kriging and multivariate adaptive regression splines using geophysical covariates[J]. *Land*, 2022, 11(3): 381.
- [13] Zhang W, Zhang Y, Fu L Y, et al. Study of the hydraulic response models in Middle Jurassic mudstone fluid wells of Chongqing district[J]. *Progress in Geophysics*, 2019, 34(5): 1826-1834. [张旺, 张艳, 符力耘, 等. 重庆地区侏罗系中统泥岩流体井的水力响应模型研究[J]. *地球物理学进展*, 2019, 34(5): 1826-1834.]
- [14] Yi C, Li D C, Zhang G L, et al. Criteria for partition of soil thickness and case studies[J]. *Acta Pedologica Sinica*, 2015, 52(1): 220-227. [易晨, 李德成, 张甘霖, 等. 土壤厚度的划分标准与案例研究[J]. *土壤学报*, 2015, 52(1): 220-227.]
- [15] Guo Z X, Adhikari K, Chellamy M, et al. Selection of terrain attributes and its scale dependency on soil organic carbon prediction[J]. *Geoderma*, 2019, 340: 303-312.
- [16] Li Z Y, Yang T, Wang C, et al. Assessing the precision of total contributing area (TCA) estimated by flow direction algorithms based on the analytical solution of theoretical TCA on synthetic surfaces[J]. *Water Resources Research*, 2021, 57(4): e2020WR028546.
- [17] Lin Y Z, Qiu B W, Chen F X, et al. Remote sensing monitoring method for plant stress resistance under drought stress on large scale[J]. *Journal of Geo-Information Science*, 2022, 24(11): 2225-2233. [林艺真, 邱炳文, 陈芳鑫, 等. 干旱胁迫下植物抗逆性大尺度遥感监

- 测方法[J]. 地球信息科学学报, 2022, 24(11): 2225-2233.]
- [18] Ao D, Yang J H, Ding W T, et al. Review of 54 vegetation indices[J]. Journal of Anhui Agricultural Sciences, 2023, 51(1): 13-21, 28. [敖登, 杨佳慧, 丁维婷, 等. 54 种植被指数研究进展综述[J]. 安徽农业科学, 2023, 51(1): 13-21, 28.]
- [19] Guo P T, Li M F, Luo W, et al. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals Kriging approach[J]. Geoderma, 2015, 237: 49-59.
- [20] Pohjankukka J, Räsänen T A, Pitkänen T P, et al. Digital mapping of peat thickness and extent in Finland using remote sensing and machine learning[J]. Geoderma, 2025, 455: 117216.
- [21] 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.]
- [22] Li X C, Luo J H, Jin X L, et al. Improving soil thickness estimations based on multiple environmental variables with stacking ensemble methods[J]. Remote Sensing, 2020, 12(21): 3609.
- [23] Addis H K, Klik A, Strohmeier S. Spatial variability of selected soil attributes under agricultural land use system in a mountainous watershed, Ethiopia[J]. International Journal of Geosciences, 2015, 6(6): 605-613.
- [24] Jenny H. Factors of soil formation: A system of quantitative pedology[M]. New York: McGraw-Hill, 1941: 1-270.
- [25] Zhang X T. Soil mapping in mixed region based on feature screening and machine learning method[D]. Wuhan: Huazhong Agricultural University, 2023. [张晓婷. 基于特征筛选及机器学习方法的混合区域土壤制图[D]. 武汉: 华中农业大学, 2023.]
- [26] Rosin N A, de Mello D C, Bonfatti B R, et al. Mapping soil thickness using a mechanistic model and machine learning approaches[J]. Catena, 2025, 249: 108621.
- [27] Huang Y P, Luo G W, Gao L H, et al. Spatial variation of soil thickness and soil moisture in the Northeast Plain[J]. Research of Soil and Water Conservation, 2025, 32(3): 108-118. [黄艳平, 罗贵文, 高林海, 等. 东北平原土体厚度和土壤水分的空间变异[J]. 水土保持研究, 2025, 32(3): 108-118.]
- [28] Zhang W C, Li J W, Zhang X, et al. Environmental variables controlling soil aggregate stability across spatial scales and locations in a karst region of southwestern China[J]. Geoderma, 2025, 456: 117240.
- [29] Chen S C, Richer-de-Forges A C, Mulder V L, et al. Digital mapping of the soil thickness of loess deposits over a calcareous bedrock in central France[J]. Catena, 2021, 198: 105062.
- [30] Pan X Y, Chen Y, Cui J H, et al. Accuracy analysis of remote sensing index enhancement for SVM salt inversion model[J]. Geocarto International, 2022, 37(8): 2406-2423.
- [31] Luo B L, Li J W, Tang J H, et al. Microtopography effects on pedogenesis in the mudstone-derived soils of the hilly mountainous regions[J]. Scientific Reports, 2024, 14: 11998.
- [32] Yan Q N, Wainwright H, Dafflon B, et al. A hybrid data-model approach to map soil thickness in mountain hillslopes[J]. Earth Surface Dynamics, 2021, 9(5): 1347-1361.
- [33] Pradhan A M S, Kim Y T. Development and evaluation of relative relief based soil thickness model: A comparative study in hilly terrain, South Korea[J]. KSCE Journal of Civil Engineering, 2021, 25(6): 2186-2198.
- [34] Xue Z H, Yi X Y, Feng W K, et al. Prediction and mapping of soil thickness in alpine canyon regions based on whale optimization algorithm optimized random forest: A case study of Baihetan Reservoir area in China[J]. Computers & Geosciences, 2024, 191: 105667.
- [35] Wang L, Wu W, Liu H B. Digital mapping of topsoil pH by random forest with residual Kriging (RFRK) in a hilly region[J]. Soil Research, 2019, 57(4): 387-396.
- [36] Li A D, Tan X, Wu W, et al. Predicting active-layer soil thickness using topographic variables at a small watershed scale[J]. PLoS One, 2017, 12(9): e0183742.
- [37] Xiao T, Segoni S, Liang X, et al. Generating soil thickness maps by means of geomorphological-empirical approach and random forest algorithm in Wanzhou County, Three Gorges Reservoir[J]. Geoscience Frontiers, 2023, 14(2): 101514.

(责任编辑: 陈荣府)