

基于RF和EBKRP算法的新安江流域 有效土壤厚度反演

王尚晓¹, 张晓东¹, 张明¹, 牛晓楠¹, 周墨¹, 唐志敏¹, 张洁¹, 宗乐丽¹, 徐帅²

[1.中国地质调查局南京地质调查中心, 江苏南京 210016; 2.华南师范大学教育科学学院(汕尾), 广东汕尾 516600]

摘要: [目的] 快速、准确地获取区域有效土壤厚度, 分析其空间分布特征和影响因素, 为植被生长、土壤保持和粮食安全工作提供理论指导。[方法] 以新安江流域为研究区, 将野外调查数据、地形、岩性和气候等成土因素结合起来, 采用经验贝叶斯克里金回归预测(EBKRP)和随机森林(RF)算法, 得到有效土壤厚度反演结果, 并分析其与环境变量之间的关系。[结果] ①区域平均有效土壤厚度为0.2~0.3 m, 城镇建设集中和人类活动密集的盆地和平原区土壤厚度较高, 丘陵山地区则较低。②从MAE(平均绝对误差)、 R^2 (判定系数)和RMSE(均方根误差)3项精度评价指标来看, RF算法的预测结果明显优于EBKRP算法, 而且更能显示出土壤厚度空间异质性分布特征, 在一定程度上提高了土壤厚度数字制图的效果。③有效土壤厚度的估算受地形和气候变量的影响较大, 它们分别占变量重要性的46.77%和18.78%。[结论] RF算法能够有效实现对区域有效土壤厚度的反演, 克服了土壤厚度空间异质性的特点, 相较于有限采样的模型更精确, 分辨率也更高。

关键词: 有效土壤厚度; 随机森林(RF); 土壤数字制图; 经验贝叶斯克里金回归预测(EBKRP); 新安江流域

文献标识码: B

文章编号: 1000-288X(2025)01-0168-10

中图分类号: S159

文献参数: 王尚晓, 张晓东, 张明, 等. 基于RF和EBKRP算法的新安江流域有效土壤厚度反演[J]. 水土保持通报, 2025, 45(1): 168-177. Wang Shangxiao, Zhang Xiaodong, Zhang Ming, et al. Effective soil thickness inversion in Xinanjiang River basin based on random forest and empirical Bayesian Kriging regression prediction algorithms [J]. Bulletin of Soil and Water Conservation, 2025, 45(1): 168-177. DOI: 10.13961/j.cnki.stbctb.2025.01.018; CSTR: 32312.14.stbctb.2025.01.018.

Effective soil thickness inversion in Xinanjiang River basin based on random forest and empirical Bayesian Kriging regression prediction algorithms

Wang Shangxiao¹, Zhang Xiaodong¹, Zhang Ming¹, Niu Xiaonan¹,

Zhou Mo¹, Tang Zhimin¹, Zhang Jie¹, Zong Leli¹, Xu Shuai²

[1. Nanjing Geological Survey Center, China Geological Survey, Nanjing, Jiangsu 210016, China;

2. School of Education (Shanwei), South China Normal University, Shanwei, Guangzhou 516600, China]

Abstract: [Objective] The effective soil thickness of a region was rapidly and accurately obtained, and its spatial distribution and influencing factors was analyzed, in order to provide theoretical guidance for vegetation growth, soil conservation, and food security. [Methods] Taking the Xinanjiang River basin as the research area, combining field survey data, topography, lithology, climate, and other soil-forming factors, the empirical Bayesian Kriging regression prediction (EBKRP) and random forest (RF) algorithms were applied to obtain the effective soil thickness inversion results. The relationship between this data and environmental variables was also analyzed. [Results] ① The average effective soil thickness in the region ranged from 0.2 to 0.3 m. Soil thickness was higher in basin and plain areas with concentrated urban

收稿日期: 2024-09-12 修回日期: 2024-11-20 采用日期: 2024-11-21

资助项目: 中国地质调查局项目“华东地区自然资源动态监测与风险评估”(DD20230103), “华东地区国土空间用途管制技术支撑与应用服务”(DD20230495); 江苏省科技计划(“一带一路”创新合作项目)“黑土地土壤退化诊断评价与动态监测技术合作研发”(BZ2023003)

第一作者: 王尚晓(1993—), 男(汉族), 河南省南阳市人, 硕士, 工程师, 主要从事自然资源遥感调查、生态环境评价方面的研究。Email: wangshangxiao@mail.cgs.gov.cn.

通信作者: 徐帅(1992—), 男(汉族), 河南省南阳市人, 硕士, 助教, 主要从事遥感生态地质调查、地质灾害风险评价等方面的研究。Email: xushuai@m.s.cn. edu.cn.

development and intensive human activity. Meanwhile, it was lower in hilly and mountainous regions. ② Based on three accuracy evaluation indicators of MAE (mean absolute error), R^2 (coefficient of determination), and RMSE (root mean square error), the prediction results of the RF algorithm were significantly better than those of the EBKRP algorithm. It could more effectively show the spatial heterogeneity distribution of soil thickness, improving the effect of soil thickness digital mapping. ③ The effective soil thickness estimation was strongly influenced by topography and climate variables, which accounted for 46.77% and 18.78% of the variable importance, respectively. [Conclusion] The RF algorithm could effectively invert regional effective soil thickness, overcoming the spatial heterogeneity of soil thickness, and is more accurate and has a higher resolution compared to models with limited sampling.

Keywords: effective soil thickness; random forest (RF); digital soil mapping; empirical Bayesian Kriging regression prediction (EBKRP); Xinanjiang River basin

有效土壤厚度是影响植被生长、土壤养分存储和地表水文的重要属性^[1-2]。在实际应用中,由于土壤厚度的分布具有强烈的空间异质性,获取区域土壤厚度通常需要大量人力物力资源,且耗时间长^[3]。因此,基于一定的调查数据,使用算法反演区域土壤厚度具有重要意义^[4]。

绘制区域内的土壤厚度分布图有多种方法,包括基于物理模型、环境变量法以及点样本的空间插值法^[2]。物理方法根据风化、剥蚀和增加的物理速率预测土壤厚度^[5],需要较多的模型参数,难以在大面积上应用^[6]。空间插值法通过有限数量采样点来估算特定区域的土壤厚度。Laslett 等^[7]对几种土壤 pH 值的空间插值方法进行了比较,发现普通克里金法、拉普拉斯平滑样条法及反距离加权法在插值结果上表现出较高的稳定性。Kravchenko 和 Bullock^[8]通过分析土壤磷(P)和钾(K)试验数据,认为在大多数数据集中,通过仔细选择变函数模型和数据变换,克里金法优于反距离加权法。但是,点样本的空间插值得到的结果严重依靠采样调查点的分布和数量,忽略了土壤形成的过程,无法充分反映土壤厚度空间特异性的变化^[9]。随着技术发展,利用环境变量并采用机器学习算法在数字土壤制图中的应用日益普及^[10]。Kuriakose 等^[2]利用回归克里金法预测土壤厚度,并发现该方法在结合海拔高度、坡度和土地利用等环境变量后效果更好。EBKRP(经验贝叶斯克里金法)回归预测是一种结合了克里金法和回归法的地统计技术,比单独应用回归法或克里金法预测更准确^[11-12]。Sarkar 等^[6]利用回归克里金(RK)模型估算了大吉岭喜马拉雅山脉的土壤容积深度,其中 RK 模型包含了 7 个预测因子,解释了土层深度空间变异的 67%。Han 等^[13]比较了线性(多元线性回归,MLR)和非线性(人工神经网络,ANN)模型在土层深度预测中的表现,发现 MLR 模型优于 ANN 的非线性模型。

尽管有大量研究对各种插值方法和机器学习算

法进行了比较,但在选择合适的方法上仍存在争议,特别是关于地统计插值分析与机器学习方法性能的比较研究还相对较少^[14]。基于此,本文旨在研究如何快速准确的获取区域土壤厚度,并对比评估 EBKRP 和随机森林(RF)算法在土壤厚度反演上的表现。

1 研究区概况和数据源

1.1 研究区概况

新安江发源于安徽省休宁县与江西省婺源县交界处的五股尖山,位于钱塘江流域上游,流域面积约 $1.20 \times 10^4 \text{ km}^2$ ^[15]。新安江流域属亚热带季风气候,季节分明,年降水量通常介于 900~1 700 mm,区域内河流密布,水系发达。季风强度每年变化很大,这增加了干旱、洪水、持续高温或低温等自然灾害的易发性。位于新安江流域上游的黄山市,既是传统的农业区,也是新兴的旅游胜地,大部分地形为丘陵和山地,该地区森林覆盖率高,水源涵养能力强。杭州市位于新安江流域下游,是传统的经济强市,过去 10 a,杭州市 GDP 累计增长 114.0%,年均增长 7.9%,比全国平均水平高出 1.2%^[16]。

1.2 土壤数据集

土壤发生过程中随着物质的迁移和转化,会形成不同的土层,被称为“土壤发生层”^[17]。本研究的有效土壤厚度是指土壤发生层中土壤母质层(C)以上的厚度[主要包括有机质层(O)、腐殖质层(A)、淋溶层(E)、淀积层(B)],也称为“有效土体厚度”,即植物根系可以生长,土壤养分和水分可以运移的土壤层厚度^[17]。土壤厚度数据来自于 2021 和 2022 年在野外调查的土壤剖面,共 621 个调查点,其空间分布如图 1 所示。在布设样点时,已经兼顾不同土壤类型、岩性和地形组合,使其在研究区相对均匀分布。野外调查时,在坡度较小的地区或者平原区,使用冲击钻和背负式钻等工具采集土壤剖面,在坡度较大的山区选

择较新鲜的非自然剖面进行调查。取样时记录每个调查点的地形地貌特征、植被和地层岩性,并根据土壤成因特征将土壤剖面划分为不同发生层,详细记录土壤发生层厚度、边界过渡特征、颜色差异和紧实状况等特征。调查点土壤厚度范围为 0.03~2.0 m,平均值为 0.449 m,标准差和变异系数分别为 0.358 m 和 79.9%,夏皮罗-威尔克法检验($p>0.05$)表明土壤厚度样本服从正态分布。

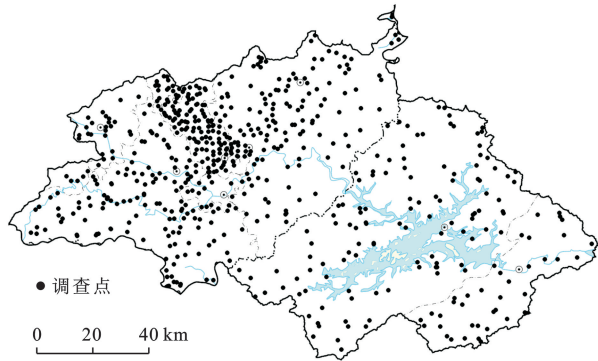


图 1 新安江流域土壤剖面野外调查点空间分布

Fig.1 Spatial distribution of soil profile survey points in field investigation in Xinanjiang River basin

1.3 环境因子

土壤的形成涉及一系列过程,包括母质、气候、生物、地形和时间这 5 个关键的成土因素^[10,18]。结合前人相关研究,选择影响土壤形成的定量和定性环境协变量(表 1,图 2)用于反演区域土壤厚度,包括地形、气候、生物、母质和地表覆被共 9 个环境因子。所有数据按研究区范围进行裁剪,采样为统一的空间分辨率。

表 1 环境因子及数据源

Table 1 Environmental factors and data sources

变量	类别	数据源
海拔高度	地形	http://www.gscloud.cn/
坡度	地形	http://www.gscloud.cn/
坡位	地形	http://www.gscloud.cn/
地形湿度指数	地形	http://www.gscloud.cn/
土地利用	地表覆被	https://data.casearth.cn/
岩性	母质	https://geocloud.cgs.gov.cn/
植被覆盖	生物	http://www.gscloud.cn/
平均降雨	气候	https://data.cma.cn/
平均温度	气候	https://data.cma.cn/

1.3.1 地形因子 海拔高度的变化会导致地貌、植被、土壤和水文等自然要素的空间变化,这种变化会影响土壤厚度的累积^[19]。本研究使用的海拔数据来源于 GDEM(基于 ASTER 卫星数据生成的数字高

程模型),研究区的海拔范围为 0~1 777 m,平均值为 375.47 m,标准差和变异系数分别为 241.13 m 和 64.22%。坡度对应力分布、地表水径流和冲刷、地下水、松散物质堆积以及人类工程活动有不同程度的影响和控制,从而影响土壤厚度的分布^[20]。坡度数据主要通过 GDEM 计算得出,研究区坡度取值范围为 0°~71.36°,平均值为 19.03°,标准差和变异系数分别为 12.65°和 66.47%。坡位为地形坡面纵剖面的上下位置,是影响太阳辐射、降水以及光、热、养分、水和其他资源空间再分配的重要地形因素,影响土壤特性的空间模式^[21]。基于 GDEM 数据,使用相对位置指数(RP)进行坡位类型的划分^[21]。研究区坡位类型最终划分为山脊、上坡、中坡、平坡、下坡和山谷 6 种类型,各类型面积占研究区总面积的比例分别为 12.10%,12.94%,40.96%,1.80%,22.35%和 9.84%。地形湿度指数(TWI)用于量化土壤湿度的空间分布。一般来说,TWI 值越高,说明土壤含水量越大,更容易饱和并产生径流,TWI 值越低,表示土壤越干燥^[22]。从 GDEM 中提取坡度和汇流量后,就可以计算出该地区的 TWI 值。研究区的 TWI 范围为 2.27~31.34,平均值为 6.99,标准差和变异系数分别为 4.57,65.38%。

1.3.2 地表覆被因子 土壤厚度在一定程度上决定了土地利用类型,而不同的土地利用类型又会反过来影响土壤的发育和厚度。土地利用数据来自于 GLC_FCS30,该产品利用 Landsat 卫星数据生成全球 30 m 精细地表覆盖动态监测产品^[23]。研究区几种主要的土地利用类型和面积所占比例分别为林地(82.37%)、耕地(10.87%)、湿地(4.18%)、草地(0.09%)、建设用地(2.48%)和裸地(0.000 1%)。

1.3.3 母质因子 由于成土母岩的岩性、矿物成分、地球化学特征和构造环境的差异,不同构造地层的成土速率、土壤结构、质地和厚度发育都会受到很大影响^[20]。岩性数据基于研究区 1:20 万地质图,以其生态属性和资源属性对岩石分类,重新划分为 7 类岩性,包括碎屑岩、第四纪松散沉积物、火山岩、变质岩、碳酸盐岩、侵入岩和红层。研究区主要的成土母岩类型和面积所占比例分别为碎屑岩(40.5%)、变质岩(16.2%)、红层(10.83%)、碳酸盐岩(9.5%)。

1.3.4 生物因子 植被对土壤侵蚀有双重作用,首先,它能减少降雨对土壤的影响,削弱地表径流的冲刷作用,从而减少水土流失。其次,它能增强土壤的蓄水能力,巩固土壤结构,促进土壤厚度的积累,反过来增加植被覆盖度^[24]。植被覆盖度数据主要依赖于 Landsat 卫星数据基于像元二分法进行反演。研究

区的植被覆盖度的范围为 0~0.98,平均值为 0.69,标准差和变异系数分别为 0.22,31.59%。

1.3.5 气候因子 不同地区的降水和温度等气候条件导致不同的风化和矿物合成速度,以及有机物的形成和积累,影响土壤中物质的迁移、分解、合成和转化,最终影响土壤的厚度^[25]。研究区 1985—2020 年

的平均降水量和气温数据由中国气象数据网的站点数据监测数据空间插值得到。研究区的平均降雨量范围为 0~2 182.9 mm,平均值为 848.2 mm,标准差和变异系数分别为 823.8 mm 和 97.1%,平均气温的范围为 0~17.8 °C,平均值为 8.18 °C,标准差和变异系数分别为 7.96 °C 和 97.3%。

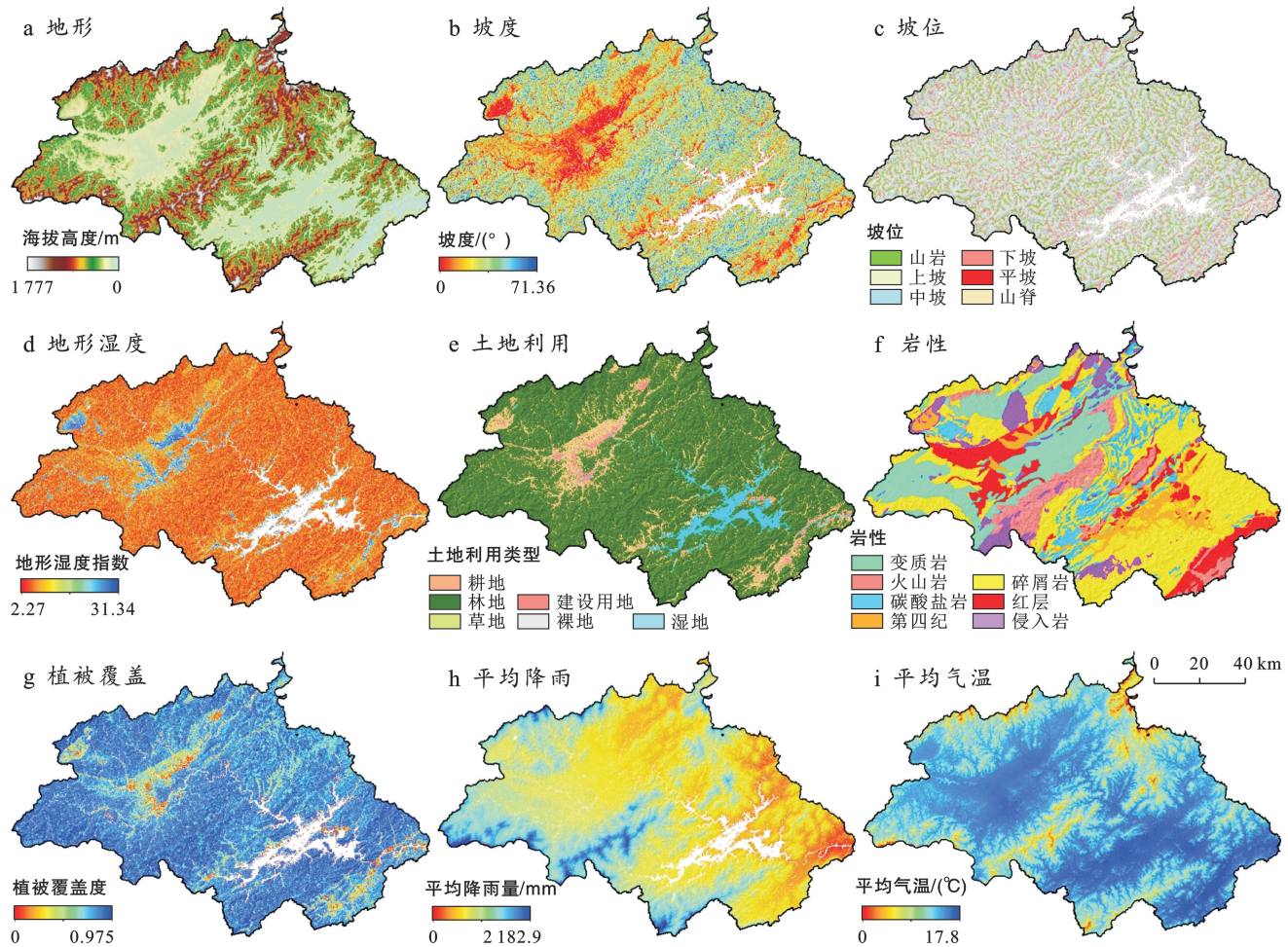


图 2 新安江流域环境变量图

Fig.2 Map of environmental variables in Xinanjiang River basin

2 研究方法

2.1 回归克里金算法

EBKRP 回归预测是一种地统计插值技术,将克里金法和回归分析相结合,利用经验贝叶斯克里金法和解释变量栅格,比单独使用其中一种方法得出的预测结果更精确^[11]。EBKRP 通过解释半变异函数中引入的误差来进行区域估计,同时考虑到了区域效应。然而,EBKRP 并未能识别出与响应变量高度相关的自变量,也未能明确指出哪些解释变量对预测结果具有显著影响。EBKRP 模型能够通过构造子集和仿真过程自动调整所需参数,自动求解克里金模型最具挑战性的特征,不需要人工调整参数。本研究中使用

的 EBKRP 遵循以下流程,将环境栅格因子形式改为其主成分,以减少(解释变量相互关联)多重共线性问题,将转换成主成分形式的协变量用于回归模型^[12]。选择 K-Bessel 变异函数模型,并对预测变量进行数据转换。

2.2 随机森林算法

随机森林(RF)是一种使用多棵树进行训练和预测的统计学习方法^[26]。该方法能够处理高维数据集,具有高度可靠性和低时间复杂性。随机森林模型的基本原理是将 CART 决策树纳入 Bagging 算法中,进行各种随机抽样,并进行替换,然后对其进行训练,以建立一个单一的决策树分类器。它在实现精确预测的同时避免了过度拟合,并对异常值和噪声表现

出高度的适应性和容忍度,使其优于其他传统回归模型^[26]。本研究使用 R 语言的 RF 软件包来进行相关分析。将随机森林中树的数量设置为 1 000,更多的数量会产生趋于稳定的结果。每个节点上随机选择的变量个数设置为默认值,因为默认值往往能使随机森林表现更好。最小叶子节点数则采用默认值 5。每个随机变异变量引起的随机森林误差变化表明相应变量的重要性。通过 100 次随机森林试验确定变量重要性,并将变量重要性之和标准化为 100%。

2.3 模型确认

对于每个预测模型,随机分为 75% 的训练数据和 25% 的测试数据,利用训练数据集训练空间预测模型并预测土壤厚度,使用测试数据集来验证预测模型的能力。然后使用不同的性能指标,将测试数据集的预测结果与验证数据进行比较,对模型性能进行评估。

为了评估土壤厚度反演结果的准确性,使用测试数据验证,并计算反演结果的 3 个准确性评估指标:均方根误差(RMSE)、平均绝对误差(MAE)和判定系数(R^2)。MAE 表示预测值与真实值之间的平均相对误差。RMSE 表示预测值与真实值之间的偏差程度。 R^2 表示拟合曲线对总变化的解释程度,并表明拟合曲线与实际数据的拟合程度。这些指数的计算方法如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - X_i)^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - X_i| \quad (2)$$

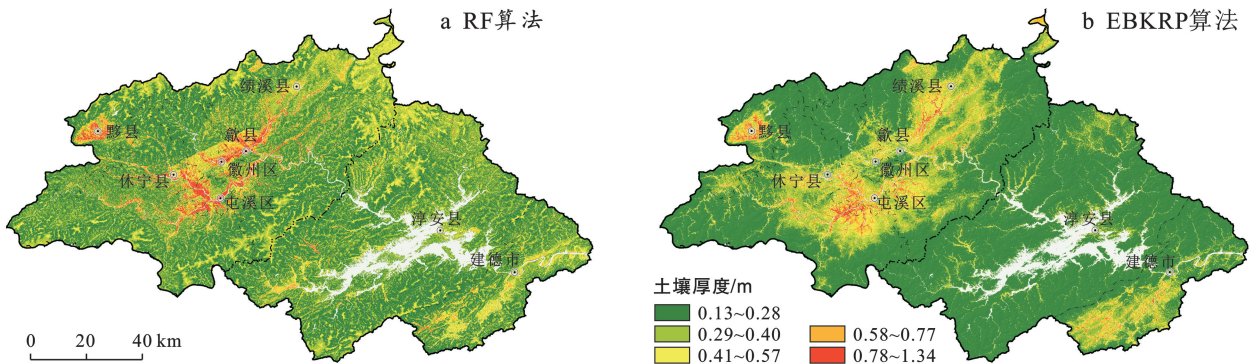


图 3 新安江流域不同模型土壤厚度反演结果

Fig.3 Soil thickness obtained from different models in Xinanjiang River basin

3.2 EBKRP 和 RF 反演结果比较

图 3—4 和表 2 分别显示了利用不同模型(RF 和 EBKRP)得到的土壤厚度值的空间分布和相应的统计信息。从反演结果来看,利用 EBKRP 得出的土壤厚度范围为 0.02 m 至 1.4 m,平均值为 0.23 m,标准偏差

$$R^2 = \frac{[\sum_{i=0}^n (Y_i - \bar{Y})(X_i - \bar{X})]^2}{\sum_{i=1}^n (Y_i - \bar{Y}) \sum_{i=1}^n (X_i - \bar{X})^2} \quad (3)$$

式中:在第 i 个观测时间, Y_i, X_i 分别代表反演土壤厚度和观测土壤厚度; n 代表样本量; \bar{Y}, \bar{X} 代表反演土壤厚度和观测土壤厚度的平均值。

3 结果与分析

3.1 土壤厚度空间分布特征

从土壤厚度反演结果的空间分布来看,EBKRP 算法和 RF 算法得到的反演结果相似,新安江流域内土壤厚度从安徽省宣城市绩溪县—黄山市休宁县、歙县和建德市等城镇建设集中的盆地和平原地区向周边丘陵区递减(图 3)。土壤厚度最高的地区位于休宁县城的南侧和屯溪区西侧,最低的地区则主要分布在安徽省和浙江省交界的山区。从两种反演结果的空间聚集分布规律来看,EBKRP 方法产生的插值结果较为平缓,多呈块状集中分布,并形成了局部极大或极小中心的分布趋势。RF 方法产生的插值结果更加分散,空间变化更加复杂,显示出一定聚集分布和变化趋势的结果。由 RF 方法在丘陵地区的反演结果显示出土壤厚度从山顶到山脚逐渐增加,而这在 EBKRP 方法的反演结果中并不明显,可知其更能描述地形的细节,更好地表达了研究区域的实际特征。EBKRP 算法在计算时,会通过变异函数插入 EBKRP,每个点都会受到相邻采样点的影响,这导致了点插值的结果趋于平缓,最大值被低估,最小值被高估^[27]。

为 0.15;利用 RF 得出的土壤厚度范围为 0.12 m 至 1.34 m,平均值为 0.3 m,标准偏差为 0.15。EBKRP 结果的最小值和最大值的集中区域都小于 RF。EBKRP 模型土壤厚度反演结果主要集中在 0.05~0.2 m 范围内,而 RF 模型的预测值略高,主要分布范围为 0.2~0.3 m。

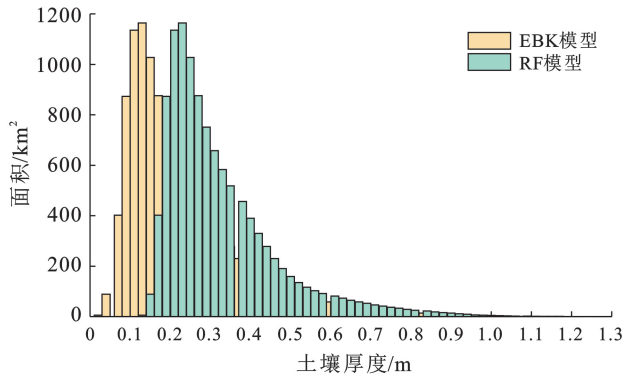


图 4 RF 和 EBKRP 不同区间的土壤厚度分布面积

Fig.4 Area of soil thickness distribution in different intervals RF of and EBKRP

表 2 不同模型反演土壤厚度的描述性统计

Table 2 Descriptive statistics of soil thickness for different model

模型	最小值/m	最大值/m	平均值/m	标准差
EBKRP	0.02	1.4	0.23	0.15
RF	0.12	1.34	0.30	0.15

图 5 显示了 EBKRP 和 RF 模型反演结果差值的空间分布。两种模型的差值范围为 -0.811 至 0.817 m, 平均值为 -0.074 m, 标准偏差为 0.123。总体而言,

土壤厚度反演结果中差异较大的地区主要分布在地形变化相对较大的地区,而在坡度较缓或海拔较低的地区差异较小。

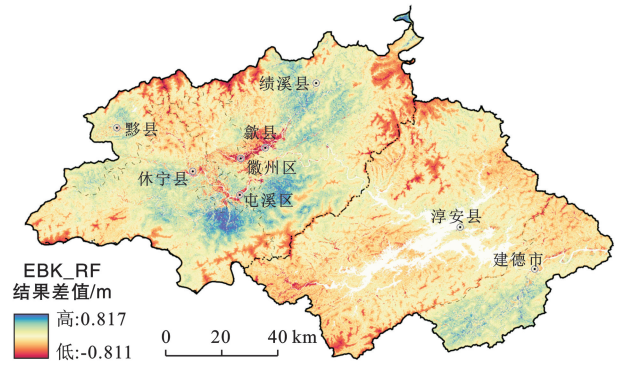


图 5 新安江流域不同模型土壤厚度反演结果空间差异分布
Fig.5 Spatial difference of soil thickness obtained from different models in Xinjiang River basin

利用从调查点获得的土壤厚度真实记录值,比较两种算法的反演结果(图 6)。从 MAE, R^2 和 RMSE3 个指标看,RF 算法的反演结果(MAE = 0.178, RMSE = 0.241, $R^2 = 0.74$)明显优于 EBKRP 算法(MAE = 0.193, RMSE = 0.267, $R^2 = 0.69$)。

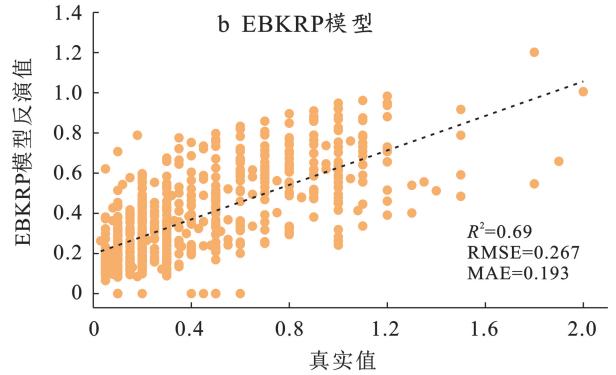
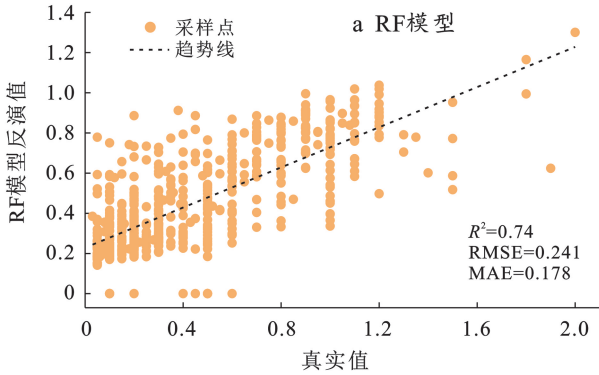


图 6 两种反演模型精度评价

Fig.6 Accuracy evaluation of different two models

3.3 垂向土层厚度反演精度

为了进一步验证两种方法反演土壤厚度的准确性,根据不同部位的土壤厚度实地调查点绘制典型的土壤剖面图(图 7—8)。现场调查发现,从山顶到山脚,土壤厚度分布不均,整体呈现出山顶薄,山脚厚,并伴有局部波动的特征。在距山顶 0~800 m 的水平范围内,由于中上部山坡和山脊的地形坡度普遍较陡,导致严重的水土流失,土层相对较薄,平均厚度约为 0.2 m。在此区间内,RF 算法在这一区间的表现更好,反演结果更准确。相比之下,EBKRP 算法的预测值较高,部分地区的预测厚度超过 0.4 m,与实际情况

存在较大偏差。距山顶 800~1 000 m 范围内,受到山脊两侧风化和坡积作用的影响,土壤厚度从 0.2 m 逐渐增加到 0.4 m。EBKRP 算法和 RF 算法的表现类似,前者预测的数值略高。距山顶 1 000~1 600 m 之间,中下坡和盆地区域的土壤厚度增加,更易于开发和利用。这些地区目前的土地利用类型主要是连片的农田。两种算法表现类似,土壤厚度逐渐增加,并受局部区域的微地貌影响而产生一定的波动。综上所述,RF 比 EBKRP 算法得到的结果更加精确,能够较准确地反映出土壤厚度不均匀分布的特征,从而在一定程度上提高了土壤厚度空间制图的效果。

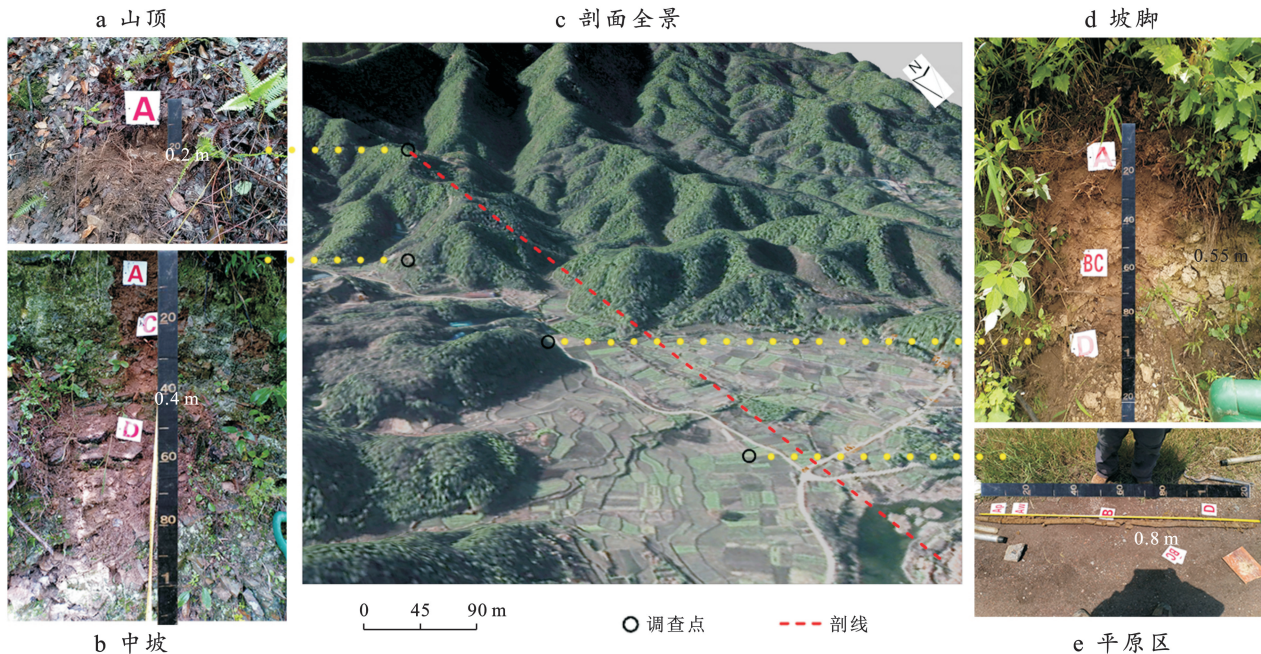


图 7 新安江流域典型土壤纵剖面
Fig.7 Typical vertical soil profiles in Xinanjiang River basin

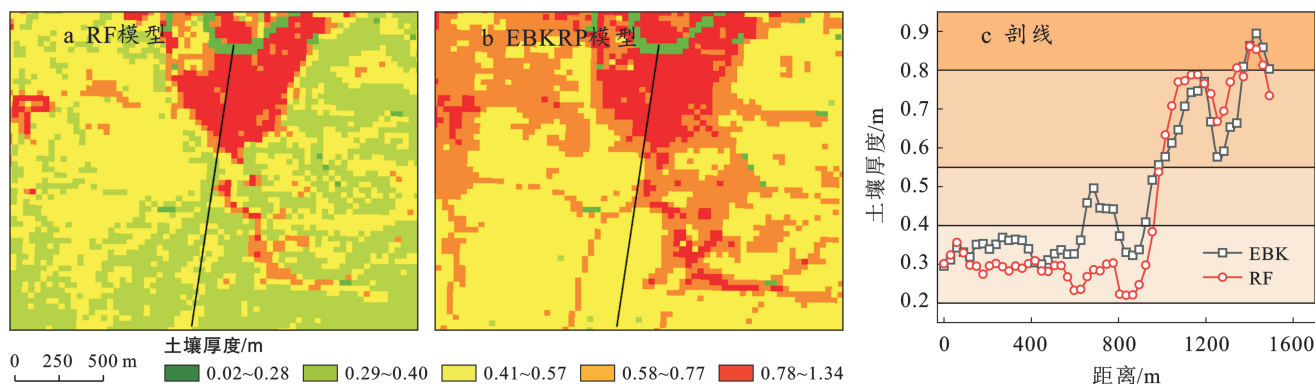


图 8 EBKRP 和 RF 的新安江流域典型土壤剖面反演结果
Fig.8 Typical soil profile predictions with EBKRP and RF in Xinanjiang River basin

3.4 变量重要性

RF 模型的变量重要性分析确定了造成新安江流域土壤厚度空间差异的主要环境因素(表 3)。地形属性和气候变量分别占总变量重要性的 46.77% 和 18.78%。坡度和植被覆盖度的重要性分别为 18.8% 和 18.06%，这与之前的研究一致，表明这些变量对估计土壤厚度至关重要^[6,28]。海拔高度、平均降雨量和 TWI 的重要性约为 10%，影响程度相似。海拔高度与降雨和温度密切相关，而降雨和温度会影响风化过程^[6]。地形湿润指数(TWI)和海拔与水流和沉积物流的运动以及侵蚀和沉积之间的平衡有关^[29]。其余因素的重要性从大到小依次为土地利用(9.91%)、平均温度(8.71%)、岩性(6.48%)和坡位(6.19%)。

4 讨论

不同模型预测的土壤厚度预测结果表现出相似

的分布规律,厚土多集中在河谷、宽广的平原和坡度较缓的地区,而薄土主要分布在坡度较陡、海拔较高的地区。土壤厚度的空间变化与流域地形具有较好的对应关系,这清楚地凸显了地形在土壤厚度空间变异中的重要性。这一结果也与其他研究一致,最厚的土壤位于在低、较平坦的沉积区,最薄的土壤在较高海拔的山区分布^[2,29]。上坡的坡度越陡,土壤侵蚀越严重,因为土壤侵蚀与坡度呈正相关,从而导致土壤变薄,坡面侵蚀的土壤堆积在洼地排水区,导致坡脚土壤厚度增加^[30]。农田区分布有一些孤立的较厚土壤点,这些孤立的点可能是由于微地貌和人类活动共同作用的结果(图 8),农田土壤厚度不仅具有显著的空间变异,而且由于人为的起垄、开沟等人为干扰,具有随机性的成分。为了进一步分析样本点对预测结果的影响,统计 621 个调查点上的各个因子值(图 9),对于连续型因子,统计样本点上各因子值的最大值、

最小值、均值和 20%~75%分布区间,反之,统计离散型因子各类别中调查点的个数。

表 3 新安江流域变量重要性统计

Table 3 Statistics on importance of environmental variables in Xinanjiang River basin

变量	类别	变量重要性
海拔高度		10.9%
坡度	地形	18.8%
坡位		6.19%
地形湿度指数		10.87%
土地利用	地表覆被	9.91%
岩性	母质	6.48%
植被覆盖	生物	18.06%
平均降雨	气候	10.06%
平均温度		8.71%

从统计结果来看,样本点的分布主要有以下几点特征:相对集中分布在坡度较小的低海拔区;植被覆盖和地形湿度指数呈双峰分布;在平均降雨和平均气温因子分布相对均匀;主要集中分布于林地、耕地等

土地利用类型内,坡位类型主要为中下坡。受制于土壤剖面的现场调查和测量的工作量,一些土壤层厚的地区,如城镇集中区和水库周期性淹没区,数据获取受到一定限制,在这些地区获取到的数据量相对较少,一定程度上影响了模型的预测精度,导致预测的土壤厚度比测量的土壤厚度变化小。类似的发现也在 Yang 等^[28]的土壤厚度制图研究中被提及。

RF 算法为每个因子提供了变量重要性度量,但简化了模型解释,没有显示目标和预测变量之间的函数关系,也没有解释预测是如何产生的。未来,将地统计学方法与机器学习方法相结合,可能会比单独的模型具有更好的预测性能。此外,本文涉及的大多环境变量都是连续的,相关操作主要是处理特定的连续数值信息。这种方法被称为“硬插值法”,在一定程度上忽略了非数值变量对土壤相关性(如土壤类型和质地)的影响,如坡位、土地利用和岩性。后续研究中,结合考虑数据不确定性的“软数据”,并将非数值变量纳入计算范围,可以更有效地提高反演结果的精度。

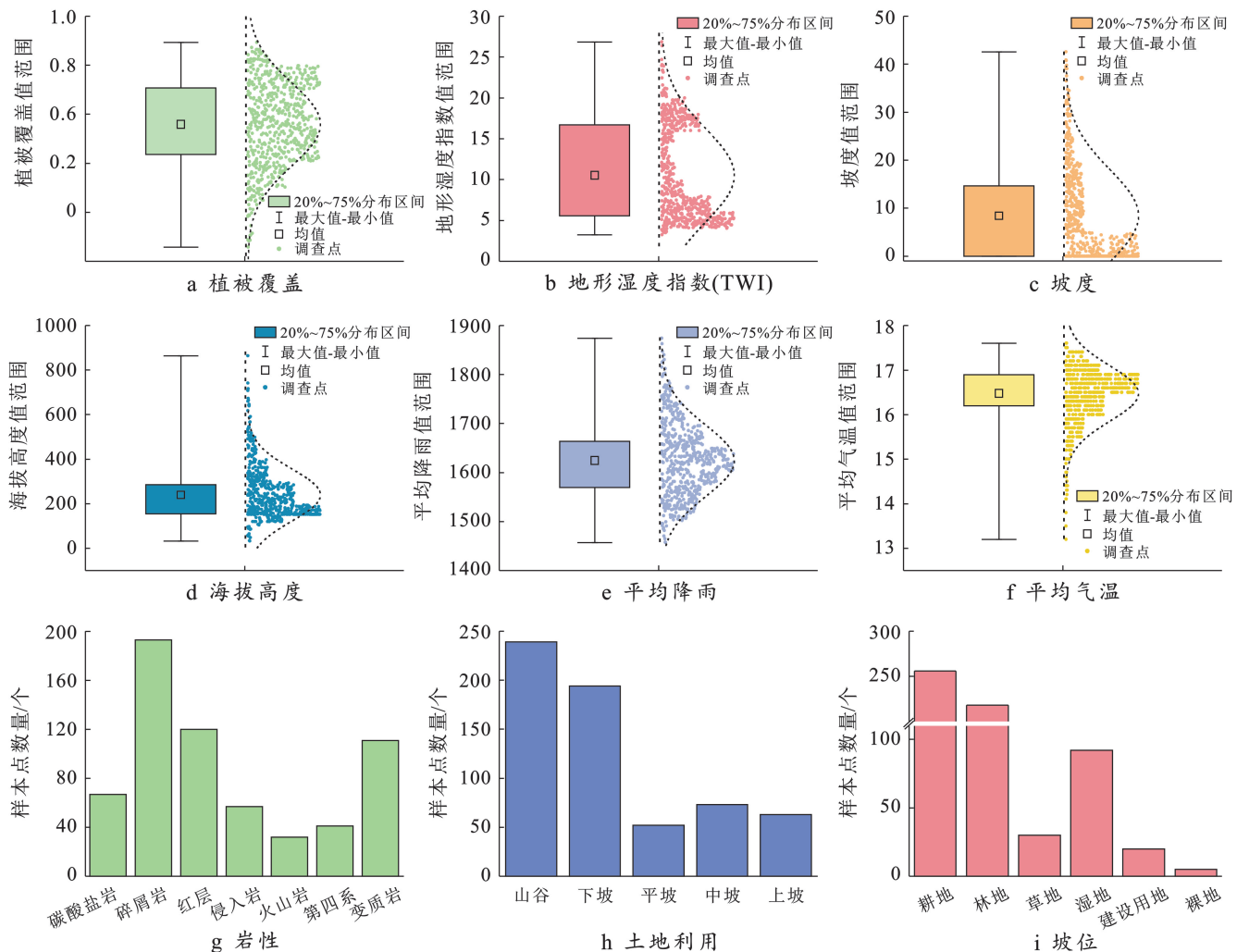


图 9 新安江流域样本点因子值统计结果

Fig.9 Statistical values of sample point factors in Xinanjiang River basin

5 结论

(1) 新安江流域内有效土壤厚度为 0.02 m 至 1.4 m, 平均厚度约为 0.23~0.3 m。流域土壤厚度具有以下明显的空间分布模式: 土壤较厚的区域主要集中在山谷和低海拔地带, 而土壤较薄的区域则多见于陡峭的山坡和高海拔地区。

(2) 两种算法对比, RF 在反演土壤厚度方面优于 EBKRP, 具体表现为 RF 算法的 RMSE(0.241 m) 和 MAE(0.178 m) 值更低, R^2 (0.74) 更高, 且能更好地保留细节, 能够反映出土壤异质性的空间分布特征。

(3) 地形属性对土壤厚度的贡献最大, 其中坡度、海拔和地形湿度指数等因子是决定新安江流域土壤厚度的关键因素。

参考文献 (References)

- [1] Gochis D J, Vivoni E R, Watts C J. The impact of soil depth on land surface energy and water fluxes in the North American Monsoon region [J]. *Journal of Arid Environments*, 2010, 74(5): 564-571.
- [2] Kuriakose S L, Devkota S, Rossiter D G, et al. Prediction of soil depth using environmental variables in an anthropogenic landscape: A case study in the Western Ghats of Kerala, India [J]. *CATENA*, 2009, 79(1): 27-38.
- [3] 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.
Liu Jintao, Zhao Wei, Liu Yu. Modelling soil thickness evolution: Advancements and challenges [J]. *Acta Pedologica Sinica*, 2024, 61(2): 319-330.
- [4] 周云龙. 福建南安市石井: 东田地区土壤厚度空间分布特征及主要影响因素 [J]. *华东地质*, 2022, 43(1): 124-132.
Zhou Yunlong. Spatial distribution characteristics and main influencing factors of soil thickness in Shijing-Dongtian area, Nan'an City, Fujian Province [J]. *East China Geology*, 2022, 43(1): 124-132.
- [5] Heimsath A M, Dietrich W E, Nishiizumi K, et al. Cosmogenic nuclides, topography, and the spatial variation of soil depth [J]. *Geomorphology*, 1999, 27(1/2): 151-172.
- [6] Sarkar S, Roy A K, Martha T R. Soil depth estimation through soil-landscape modelling using regression Kriging in a Himalayan terrain [J]. *International Journal of Geographical Information Science*, 2013, 27(12): 2436-2454.
- [7] Laslett G M, McBratney A B, Pahl P J, et al. Comparison of several spatial prediction methods for soil pH [J]. *Journal of Soil Science*, 1987, 38(2): 325-341.
- [8] Kravchenko A, Bullock D G. A comparative study of interpolation methods for mapping soil properties [J]. *Agronomy Journal*, 1999, 91(3): 393-400.
- [9] Zhang Shuai, Liu Gang, Chen Shuli, et al. Assessing soil thickness in a black soil watershed in Northeast China using random forest and field observations [J]. *International Soil and Water Conservation Research*, 2021, 9(1): 49-57.
- [10] Zhang Ganlin, Liu Feng, Song Xiaodong. Recent progress and future prospect of digital soil mapping: A review [J]. *Journal of Integrative Agriculture*, 2017, 16(12): 2871-2885.
- [11] Krivoruchko K, Gribov A. Evaluation of empirical Bayesian Kriging [J]. *Spatial Statistics*, 2019, 32: 100368.
- [12] Gribov A, Krivoruchko K. Empirical Bayesian Kriging implementation and usage [J]. *Science of the Total Environment*, 2020, 722: 137290.
- [13] Han Xiaole, Liu Jintao, Mitra S, et al. Selection of optimal scales for soil depth prediction on headwater hillslopes: A modeling approach [J]. *Catena*, 2018, 163: 257-275.
- [14] Mallik S, Bhowmik T, Mishra U, et al. Mapping and prediction of soil organic carbon by an advanced geostatistical technique using remote sensing and terrain data [J]. *Geocarto International*, 2022, 37(8): 2198-2214.
- [15] 唐志敏, 张晓东, 张明, 等. 新安江流域土壤元素地球化学特征: 来自岩石建造类型的约束 [J]. *华东地质*, 2023, 44(2): 172-185.
Tang Zhimin, Zhang Xiaodong, Zhang Ming, et al. Geochemical characteristics of soil elements in Xin'an River basin: Constraints from rock formation types [J]. *East China Geology*, 2023, 44(2): 172-185.
- [16] 王尚晓, 牛晓楠, 张洁, 等. 新安江流域 2000—2021 年 NDVI 时空变化特征及其影响因素分析 [J]. *华东地质*, 2023, 44(3): 273-281.
Wang Shangxiao, Niu Xiaonan, Zhang Jie, et al. Spatial and temporal variation features of NDVI and its influence factors in the Xin'an River basin from 2000 to 2021 [J]. *East China Geology*, 2023, 44(3): 273-281.
- [17] 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.
Yi Chen, Li Decheng, Zhang Ganlin, et al. Criteria for partition of soil thickness and case studies [J]. *Acta Pedologica Sinica*, 2015, 52(1): 220-227.
- [18] McBratney A, Santos M L M, Minasny B. On digital soil mapping [J]. *Geoderma*, 2003, 117: 3-52.
- [19] 邱思慧, 林少颖, 王维奇. 中国东部地区不同海拔梯度土

- 壤有机碳库特征及其影响因素综述[J].中国水土保持科学(中英文),2022,20(3):142-150.
- Qiu Sihui, Lin Shaoying, Wang Weiqi. Characteristics of soil organic carbon pool at different altitude gradients in Eastern China and its influencing factors: A review [J]. Science of Soil and Water Conservation, 2022,20(3):142-150.
- [20] 赵玉国,张甘霖,龚子同,等.海南岛不同地质背景下的土壤类型、质量特征和作物适宜性[J].第四纪研究,2005,25(3):389-395.
- Zhao Yuguo, Zhang Ganlin, Gong Zitong, et al. Soil type, soil quality and crop suitability of soils developed from different geology environments in Hainan Island [J]. Quaternary Sciences, 2005,25(3):389-395.
- [21] 刘鹏举,夏智武,唐小明.基于 DEM 和坡面特征的坡位生成方法[J].北京林业大学学报,2016,38(2):68-73.
- Liu Pengju, Xia Zhiwu, Tang Xiaoming. Slope position extraction method based on DEM and slope features [J]. Journal of Beijing Forestry University, 2016,38(2):68-73.
- [22] O'Loughlin E M. Prediction of surface saturation zones in natural catchments by topographic analysis [J]. Water Resources Research, 1986,22(5):794-804.
- [23] Zhang Xiao, Liu Liangyun, Chen Xidong, et al. GLC-FC30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery [J]. Earth System Science Data, 2021,13(6):2753-2776.
- [24] 王晶,赵文武,刘月,等.植物功能性状对土壤保持的影响研究述评[J].生态学报,2019,39(9):3355-3364.
- Wang Jing, Zhao Wenwu, Liu Yue, et al. Effects of plant functional traits on soil conservation: A review [J]. Acta Ecologica Sinica, 2019,39(9):3355-3364.
- [25] 常海钦,付亚龙,林鑫,等.流域盆地化学风化强度空间分布及控制因素研究:以长江和珠江为例[J].地球科学进展,2019,34(1):93-102.
- Chang Haiqin, Fu Yalong, Lin Xin, et al. Spatial distribution and controlling factors of chemical weathering intensity in drainage basins: A case study in the Yangtze River basin and Pearl River basin [J]. Advances in Earth Science, 2019,34(1):93-102.
- [26] Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author) [J]. Statistical Science, 2001,16(3):216-218.
- [27] Molla A, Zhang Weiwei, Zuo Shudi, et al. A machine learning and geostatistical hybrid method to improve spatial prediction accuracy of soil potentially toxic elements [J]. Stochastic Environmental Research and Risk Assessment, 2023,37(2):681-696.
- [28] Yang Qiyong, Zhang Fawang, Jiang Zhongcheng, et al. Relationship between soil depth and terrain attributes in karst region in Southwest China [J]. Journal of Soils and Sediments, 2014,14(9):1568-1576.
- [29] Moore I D, Grayson R B, Ladson A R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications [J]. Hydrological Processes, 1991,5(1):3-30.
- [30] 刘宝元,阎百兴,沈波,等.东北黑土区农地水土流失现状与综合治理对策[J].中国水土保持科学,2008,6(1):1-8.
- Liu Baoyuan, Yan Baixing, Shen Bo, et al. Current status and comprehensive control strategies of soil erosion for cultivated land in the Northeastern black soil area of China [J]. Science of Soil and Water Conservation, 2008,6(1):1-8.

(上接第 157 页)

- [25] 罗玉婷,唐川,熊江,等.四川省汶川县下庄沟“8·20”泥石流成因分析及堵江范围预测[J].水土保持通报,2020,40(6):193-199.
- Luo Yuting, Tang Chuan, Xiong Jiang, et al. Cause analysis of “8·20” debris flow and forecast of river-blocking range in Xiazhuang gully of Wenchuan County, Sichuan Province [J]. Bulletin of Soil and Water Conservation, 2020,40(6):193-199.
- [26] O'Brien J S, Julien P Y, Fullerton W T. Two-dimensional water flood and mudflow simulation [J]. Journal of Hydraulic Engineering, 1993,119(2):244-261.
- [27] 倪化勇,吕学军,刘宇杰,等.低频泥石流特征及其预测初步探讨[J].工程地质学报,2007,15(5):612-620.
- Ni Huayong, Lv Xuejun, Liu Yujie, et al. Discussions on characteristics and prediction of debris flows with low occurrence frequency [J]. Journal of Engineering Geology, 2007,15(5):612-620.
- [28] Yokoya N, Yamanoi K, He Wei, et al. Breaking the limits of remote sensing by simulation and deep learning for flood and debris flow mapping [J]. Materials Science and Engineering, 2021,1051(1):48-55
- [29] Baggio T, Martini M, Bettella F, et al. Debris flow and debris flood hazard assessment in mountain catchments [J]. Catena, 2024,245:108338.
- [30] Stolle A, Langer M, Blöthe J H, et al. On predicting debris flows in arid mountain belts [J]. Global and Planetary Change, 2015,126:1-13.