

DOI: 10.11766/trxb202308210333

高张, 马利霞, 于东升, 胡文友, 李德成, 高磊, 刘峰, 张久明, 姜军, 匡恩俊, 王鑫, 宋洁, 王桐, 丁琪洵, 王昌昆, 迟凤琴, 赵玉国. 反映样点微域空间变异的多穴位黑土层厚度快速勘察与预测方法研究[J]. 土壤学报, 2024

GAO Zhang, MA Lixia, YU Dongsheng, HU Wenyu, LI Decheng, GAO Lei, LIU Feng, ZHANG Jiuming, JIANG Jun, KUANG Enjun, WANG Xin, SONG Jie, WANG Tong, DING Qixun, WANG Changkun, CHI Fengqin, ZHAO Yugu. Study on Rapid Survey and Prediction Methods of Multi-Point Black Soil Layer Thickness Reflecting Micro-Spatial Variability of Sample Points[J]. Acta Pedologica Sinica, 2024

反映样点微域空间变异的多穴位黑土层厚度快速勘察与预测方法研究*

高张^{1,2}, 马利霞¹, 于东升^{1,2†}, 胡文友^{1,2}, 李德成^{1,2}, 高磊^{1,2}, 刘峰^{1,2}, 张久明³, 姜军^{1,2}, 匡恩俊³, 王鑫^{1,2}, 宋洁^{1,2}, 王桐^{1,2}, 丁琪洵^{1,2}, 王昌昆^{1,2}, 迟凤琴³, 赵玉国^{1,2}

(1. 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008; 2. 中国科学院大学, 北京 100049; 3. 黑龙江省黑土保护利用研究院, 哈尔滨 150086)

摘要: 黑土层厚度作为土壤质量的重要标志, 在土壤可持续发展、粮食安全和生态功能中发挥着不可替代作用。然而, 基于土壤剖面调查数据分析, 往往样本数量少、代表区域范围小, 多数仅依据点位数据统计, 缺乏空间变异预测分析, 迫切需要黑土层厚度快速调查与高性能空间预测方法。通过新建调查样点黑土层厚度多穴位“浅层挖掘+深层土钻”快速获取方法, 在黑龙江省获取 357 个样点的黑土层厚度系列样本数据; 对随机森林预测模型参数进行优化, 预测黑土层厚度空间变异及其不确定性; 分析不同穴位观测值及其均值样本对优化模型预测精度及稳定性影响, 评价模型空间预测潜能。结果表明, 研究区耕地预测平均黑土层厚度为 53.42 cm, 新建黑土层厚度快速获取与预测方法行之有效, 可替代剖面调查方法。优化随机森林模型预测黑土层厚度的空间变化解释力 R^2 达到 60%, 可精细刻画黑土层厚度空间分异; 样点单个观测穴位的随机性可改变模型预测协变量重要性数值, 影响黑土层厚度空间预测分布; 相较于样点多穴位观测均值的空间预测, 单穴位观测值预测的空间分布不确定性评估的准确性较低, 预测精度显著下降; 交叉验证指标和散点图分析表明, 优化随机森林模型对黑土层厚度具有稳定的空间预测潜能。本研究为黑土层厚度高精度快速调查与预测提供了新经验、新途径。

关键词: 黑土层厚度; 多穴位勘察; 随机森林模型; 不确定性

中图分类号: S155.27 **文献标志码:** A

Study on Rapid Survey and Prediction Methods of Multi-Point Black Soil Layer Thickness Reflecting Micro-Spatial Variability of Sample Points

GAO Zhang^{1,2}, MA Lixia¹, YU Dongsheng^{1,2†}, HU Wenyu^{1,2}, LI Decheng^{1,2}, GAO Lei^{1,2}, LIU Feng^{1,2}, ZHANG Jiuming³, JIANG Jun^{1,2}, KUANG Enjun³, WANG Xin^{1,2}, SONG Jie^{1,2}, WANG Tong^{1,2}, DING Qixun^{1,2}, WANG Changkun^{1,2}, CHI Fengqin³, ZHAO Yugu^{1,2}

*国家重点研发计划项目(2021YFD1500202)和中国科学院战略性先导科技专项(XDA28010100)资助 Supported by the National Key Research and Development Program of China (No.2021YFD1500202) and the Strategic Pioneering Science and Technology Project of Chinese Academy of Sciences (No.XDA28010100)

†通讯作者 Corresponding author, E-mail: dshyu@issas.ac.cn

作者简介: 高张(1992-), 男, 湖北潜江人, 博士研究生, 主要从事资源环境与遥感应用研究。E-mail: gaozhang@issas.ac.cn

收稿日期: 2023-08-21; 收到修改稿日期: 2024-02-26; 网络首发日期(www.cnki.net):

(1. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Heilongjiang Academy of Black Soil Conservation and Utilization, Harbin 150086, China)

Abstract: 【Objective】 As an important indicator of soil quality, black soil layer thickness plays an irreplaceable role in sustainable soil development, food security and ecological functions. However, analyses based on soil profile survey data are often based on small sample sizes and small regional scales, and most of them are based on point data statistics only. However, the studies lacked spatial variability prediction analyses, hence, there is an urgent need for rapid surveys of the thickness of the black soil layer and high-performance spatial prediction methods. 【Method】 In this paper, a series of sample data of black soil thickness at 357 sample points in Heilongjiang Province were obtained by the rapid acquisition method of "shallow excavation + deep soil drilling" for black soil thickness at multiple burrows in newly constructed sample points. The spatial variability of black soil thickness and its uncertainty were predicted through the optimisation of parameters of the Random Forest Prediction Model (RPFPM). The impacts of the different burrow observations and their mean samples on the optimization of the model's prediction accuracy and stability were analyzed, and the spatial prediction potentials of the model were evaluated. 【Result】 The predicted average thickness of the black soil layer in the arable land in the study area was 53.42 cm, and the new method of rapid acquisition and prediction of black soil layer thickness was effective and can be used as an alternative to the profiling method. The spatial variation explanatory power R^2 of the optimized random forest model for predicting black soil thickness reached 60%, which could finely depict the spatial differentiation of black soil thickness. Also, the randomness of a single observation burrow at a sample point could change the importance value of the covariates predicted by the model, and affect the spatial prediction of the distribution of the black soil thickness. Compared with the spatial prediction on the mean value of several observations, the spatial prediction on a single observation had lower accuracy for uncertainty assessment of the spatial distribution and significantly reduced prediction performance. Interestingly, the cross-validation metrics and scatterplot analyses indicated that the optimized Random Forest model had a stable spatial prediction potential of the black soil thickness. 【Conclusion】 This study provides a new perspective and new ways for high-precision and rapid investigation and prediction of black soil layer thickness.

Key words: Black soil layer thickness; Multi-site survey; Random forest model; Uncertainty

东北地区土壤黑土层特指富含腐殖质的暗沃表土层^[1], 是植被生长根基及营养重要来源, 其厚度成为土壤质量及资源禀赋的重要标志^[2-6]。

土壤剖面挖掘是黑土层厚度调查的经典方法。早在 20 世纪二三十年代起, 国内外学者利用土壤剖面挖掘调查我国的黑土层厚度^[7-11]。1925 年和 1926 年俄国 Gordeef^[7]先后调查了中东铁路沿线土壤, 将典型剖面分为腐殖质层和底土层, 其中腐殖质层包括 A 层和 B 层。1935 年美国土壤学家 Pendleton 和 Hou^[8]与中国学者在哈尔滨、海伦、安达、双城等地进行剖面调查, 得出典型剖面表土层厚度为 25 ~ 100 cm, 一般厚度在 40 ~ 50 cm。1958—1960 年, 黑龙江省农业厅^[9]在全国第一次土壤普查工作中以耕地为主, 典型剖面记录的黑土层厚度一般在 30 ~ 50 cm。1979—1994 年, 全国第二次土壤普查^[10-11]中, 以典型剖面 A 层为黑土层, 《中国土种志》^[12]等 686 个典型剖面黑土层厚度一般小于 60 cm。

近些年, 通过土壤剖面分析黑土层变薄问题也得到关注^[13-15]。陆继龙^[13]对比分析了黑龙江省和吉林省第一次土壤普查(简称“一普”)与第二次土壤普查(简称“二普”)的土壤剖面统计资料, 得出黑土层厚度在此期间每年减少 0.4 ~ 0.5 cm。沈波等^[14]的研究表明, 东北黑土区从黑土地开垦之初到开垦 20 年(“一普”)、再到开垦 40 年(“二普”), 黑土层厚度每年减少 0.7 ~ 1 cm。张之一^[3]对比分析了黑龙江省北部地区和内蒙古呼伦贝尔

盟“一普”与“二普”黑土层厚度剖面统计资料,指出并非所有黑土层厚度均变薄,变薄原因主要是坡耕地面蚀和耕地土壤由松到紧的压实过程。刘凯等^[15]通过61个“二普”剖面点的对应点位剖面调查,点对点比较得出东北黑土区的黑土层厚度平均每年减少0.32 cm。

然而,野外土壤剖面调查方法,不仅存在调查样点类型代表性问题,也因人力消耗及成本费用高导致土壤剖面样本量少,严重影响了黑土层厚度及其变化调查分析的准确性。现有调查报告也多基于点位数据统计,缺乏连续性空间分布特征分析。仅有的某小流域黑土层厚度空间预测^[16],区域覆盖范围较小(1.39 km²)也不能满足需求。因此,针对土壤剖面调查样本量少的问题,需建立黑土层厚度快速、简易调查获取方法;针对连续空间分析不足问题,需建立高性能黑土层厚度空间预测方法。

目前土层厚度快速调查方法也在不断尝试,如土钻、雷达探测方法等。2013年和2014年张福平等^[17]在甘肃省敦煌市利用土钻法调查了124个样点1 m深土层厚度信息。张世文等^[18]采用500 MHz探地雷达调查北京市昌平区耕地土层厚度,得出土层厚度最佳探测深度小于45 cm,否则探测精度得不到保证。探地雷达接受雷达信号波,经过模型处理可以快速获取土层厚度信息,一定深度范围内模型精度高,但显然,目前雷达探测方法多数仍处于试验阶段,探测精度和稳定性问题依然没有得到解决,大范围区域应用也受到限制。而土钻快速获取方法,也还存在表层土壤松散滑落以及深层土壤压缩等问题;相对于土壤剖面调查断面,土钻调查穴位空间极小,缺乏多穴位重复调查,样点微域空间变异性势必影响调查精度。相较于探地雷达等方法存在问题,土钻法更易于满足调查急迫需求,但其调查规范性等问题需得到彻底解决。

黑土层厚度空间预测方法主要有经典统计^[19-20]、地统计学^[21-22]和机器学习^[23-24]。经典统计方法往往判定变量之间需具有较好的线性相关性,但对于变量和表达式的选择有一定主观性,不能很好拟合非线性数据^[25];地统计学方法更侧重土壤空间自相关,忽视土壤与环境之间的关系,较少考虑成土因素影响,在大面积复杂景观尺度下的空间插值结果的真实性较弱^[25];而机器学习在处理复杂景观以及非线性变量方面具有优势。黑土层厚度土钻快速调查方法,虽可满足机器学习对大样本量的数据要求,但对于大面积复杂景观地区,受限于道路通达性等条件,所能采集的样本数量往往还不够充分。在样本量受限的条件下,随机森林方法得到改进,产生了与分位数相结合的分位数回归森林方法^[26],但这种模型预测的性能及稳定性,是否随样本数据集不同而发生改变还需进一步论证。

因此,考虑到野外土壤剖面调查方法费时费力、探地雷达精度未解决等问题,土钻法更易于满足需求。为解决土钻法调查规范性问题,同时考虑微域空间变异性,本文将构建调查样点黑土层厚度多穴位土钻快速获取方法,通过优化回归森林模型参数开展空间预测,分析不同样本对优化模型预测性能及稳定性的影响,为黑土层厚度高精度快速调查与预测提供新经验、新途径。

1 材料与方 法

1.1 研究区概况

黑龙江省地处中国东北部(43°26′~53°33′ N, 121°11′~135°05′ E),辖区总面积47.3万 km²,地势西北、北部和东南部高,东北、西南部低,属寒温带与温带大陆性季风气候,年均温4.2℃,年均降雨量608.5 mm^[27]。土壤类型主要有黑土、黑钙土、栗钙土、草甸土、暗棕壤、白浆土、沼泽土、泥炭土、水稻土等;成土母质主要有黄土状黏质土、坡积物、冲积物等。研究区土地利用可概括为“五山一水一草三分田”,耕地总面积为17.19万 km²,约占全国耕地总面积的13%,集中于黑土、黑钙土、草甸土、白浆土、暗棕壤等土壤类型。

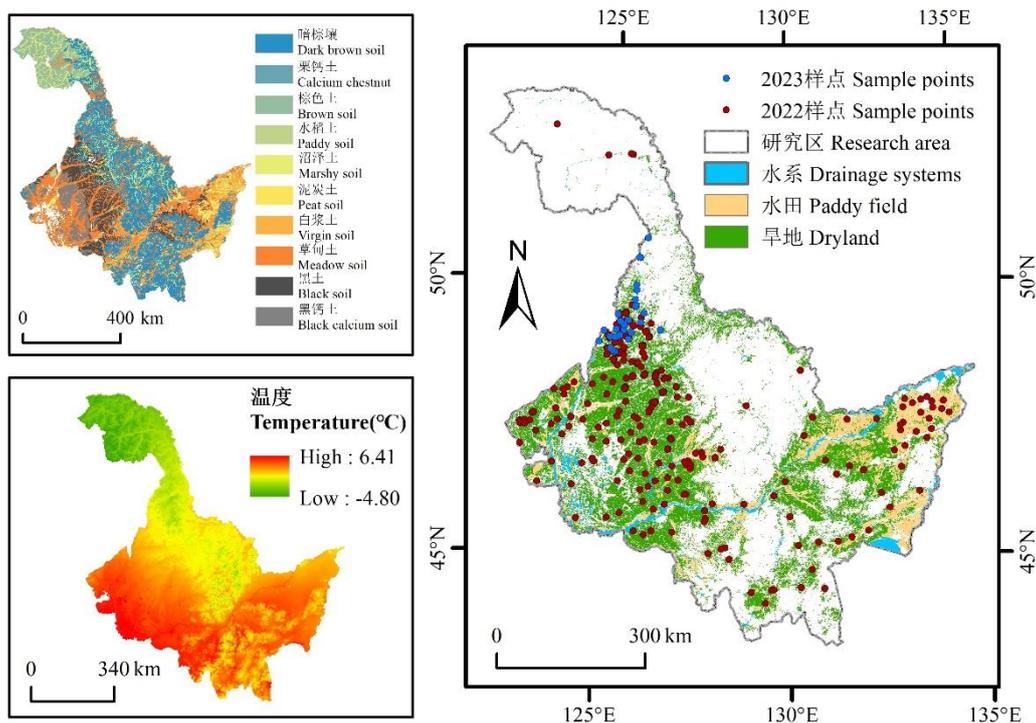


图 1 研究区位置与采样点分布

Fig. 1 Study area location and sampling point distribution

1.2 样点布设与野外调查

充分考虑土地利用、土壤及母质类型与地形因素，并兼顾坡耕地完整性，于 2022 年夏季布设和调查 311 个耕地土壤及 15 个林草地对照样点（图 1）。其中，旱地样点 283 个、水田样点 28 个；不同土壤类型样点分别为黑土 101 个、草甸土 74 个、暗棕壤 71 个、黑钙土 28 个、沼泽土 23 个、白浆土 20 个、水稻土 8 个、棕色针叶林土 1 个。

在每个调查样点 2 m×3 m 的空间范围内，随机选择 3 个土壤调查穴位 A、B、C，采用“浅层挖掘与深层土钻”相结合方法开展黑土层厚度快速调查。在每个穴位，首先挖掘 0~40 cm 简易土壤剖面（宽 40 cm×长 50 cm），如果黑土层厚度≤40 cm，可直接判断测量黑土层厚度，避免采用土钻方法因表层土壤松散沿土钻凹槽上窜或下滑带来的观测误差。当黑土层厚度≥40 cm 时，则在相应穴位中心，采用长 100 cm、直径 3 cm 的荷兰 Eijkelpamp 钻（0401SB）取土，判定黑土层厚度；为避免土壤被挤压带来观测误差，按 20 cm 等间距分别取 40~100 cm 完整土样。若黑土层厚度≥100 cm 时，以 100 cm 计。

野外调查中记录各调查样点经纬度、坡度、坡向、坡位、土地利用类型、地貌类型、土壤类型、母质类型、水土保持措施以及各穴位的表层土壤颜色、土壤质地等基本信息，同时采集表层（0~20 cm）、亚表层（20~40 cm）3 个穴位的混合土壤样品各 2 kg，带回实验室测定其理化性质。以 3 个穴位黑土层厚度的平均值作为该调查样点的黑土层厚度（图 2）。

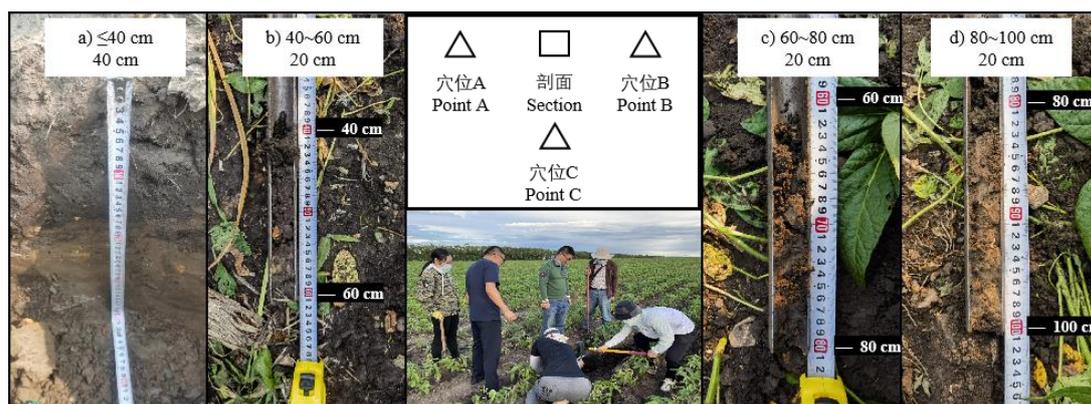


图 2 野外调查样点各观测穴位分布示意图

Fig. 2 Map of the distribution of observation points in the field survey sample sites

考虑野外调查可操作性，参考美国土壤系统分类^[1]及中国土壤系统分类^[28]关于暗沃表土层判定标准，本次黑土层调查判定指标及标准为：①土壤颜色润态明度 ≤ 3.5 ，润态彩度 ≤ 3.5 ；②土壤结构呈粒状、小角块状和小亚角块状；③石块体积占比小于 $1/2$ ^[29]。在实际操作中，先判定样点是否满足土壤结构和石块比例条件，再用芒塞尔比色卡比对土壤颜色。为减少土壤颜色判定上的主观误差，每个穴位均由 3~4 人集体判定。

2023 年 7 月在黑龙江省嫩江市示范区，又布设 31 个耕地土壤样点，涉及土壤类型主要有黑土（11 个）、草甸土（7 个）、暗棕壤（12 个）、沼泽土（1 个），分别同时采用剖面法和上述多穴位土钻快速调查法获取各样点黑土层厚度数据，并以剖面调查黑土层厚度数据为真实值，以验证多穴位黑土层厚度快速勘察与预测方法的精确度和可行性。

1.3 样点黑土层厚度观测误差

由于土壤调查样点不可避免地存在微域空间上的变异性，单个穴位调查难以获取调查样点的真实值，通过在一定空间（标准剖面挖掘空间）范围内测定多个观测穴位的黑土层厚度，取其平均值作为该调查样点的观测值。观测穴位重复越多，平均值越接近真实值，但样点平均黑土层厚度计算误差依然存在。考虑到重复穴位越多采样所消耗的成本也就越大，在每个调查点设立了 3 个重复调查穴位（A、B、C）。设 Y_T 为样点黑土层厚度的真实值， Y_S 为黑土层厚度观测值，观测误差方差^[30]：

$$\delta^2(s) = \text{var}(Y_S - Y_T) \quad (1)$$

依据第 i 个样点的 3 个穴位黑土层厚度实际观测值获取该样点平均值 $\bar{Y}_T(i)$ ， $n=357$ 为样点数，调查样点观测误差 $\delta(s)$ 公式为：

$$\delta(s) = \sqrt{\frac{\sum_{i=1}^n (Y_S(i) - \bar{Y}_T(i))^2}{n}} \quad (2)$$

1.4 环境协变量数据来源

遥感因子、地形因子、气候因子、土壤因子、人为因子共同构成黑土层厚度预测的环境协变量指标体系（表 1）。其中，DEM 数据、2022—2023 年遥感影像数据从 USGS 获取，土壤属性数据从世界土壤数据库(HWSD)获取，气候数据从国家地球系统科学数据中心 (<http://www.geodata.cn>) 获取并计算年平均值，土地利用数据从 GlobeLand30 (<http://www.globallandcover.com>) 和资源环境科学与数据中心(<https://www.resdc.cn>) 下载。将 2022 年和 2023 年卫星影像通过中值合成后提取 20 个遥感因子指数值（表 1）。

表 1 预测黑土层厚度的环境协变量

Table 1 Environmental covariates for predicting black soil layer thickness

类型 Style	指标 Norm	描述 Descriptions
----------	---------	-----------------

遥感因子 Remote sensing factors	ARVI	耐大气层植被指数 Atmospherically Resistant Vegetation Index, $(B4-(2 \times B3)+B1)/(B4+(2 \times B3)+B1)$
	BAI	燃烧面积指数 Burn area index, $1/[(0.1-B3)^2+(0.06-B4)^2]$
	CMR	黏土矿物比率 Clay minerals ratio, $B5/B7$
	FMR	铁矿物质比率 Ferrous minerals ratio, $B5/B4$
	GCI	绿色叶绿素植被指标 Green chlorophyll vegetation index, $B4/B2-1$
	GDVI	绿色差异植被指数 Green difference vegetation index, $B4-B2$
	GEMI	全球环境监控指数 Global environmental monitoring index, $\eta(1-0.25\eta)-(B4-0.125)/(1-B4)$, $\eta=(2 \times [(B5)^2-(B4)^2]+1.5 \times B5+0.5 \times B4)/(B5+B4+0.5)$
	GNDVI	绿色归一化差异植被指标 Green normalized difference vegetation index, $(B4-B2)/(B4+B2)$
	GRVI	绿化率植被指数 Green ratio vegetation index, $B4/B2$
	IPVI	红外线植被百分比指数 Infrared percentage vegetation index, $B4/(B4+B3)$
	MSR	修正简单比率 Modified simple ratio, $(B4/B3-1)/[(B4/B3)^{0.5}+1]$
	NBR	归一化燃烧率 Normalized burn ratio, $(B4-B5)/(B4+B5)$
	NDVI	归一化植被指数 Normalized difference vegetation index, $(B4-B3)/(B4+B3)$
	OSAVI	优化土壤调整植被指数 Optimized soil adjusted vegetation index, $(B4-B3)/(B4+B3+0.16)$
	RDVI	归一化差异植被指数 Renormalized difference vegetation index, $(B4-B3)/(B4+B3)^{0.5}$
	RECI	红边叶绿素植被指数 Red-edged chlorophyll vegetation index, $B4/B3-1$
	RGRI	红绿比指数 Red green ratio index, $B4/B3$
	SAVI	土壤调整植被指数 Soil Adjusted Vegetation Index, $1.5(B4-B3)/(B4+B3+0.5)$
	TDVI	变换差异植被指数 Transformed difference vegetation index, $1.5[(B4-B3)/(B4+B3+0.5)^{0.5}]$
	TVI	转化植被指数 Transform the vegetation index, $\sqrt{\text{NDVI}+0.5}$
地形因子 Topographic factors	AH	山体阴影 Analytical Hillshading
	ASP	坡向 Aspect
	CS	流域坡度 Catchment Slope
	CNBL	渠道网络基础层 Channel Network Base Level
	CND	通道网络距离 Channel Network Distance
	CI	聚合指数 Convergence Index
	CON	凸面性 Convexity
	CC	曲率分类 Curvature Classification
	DAH	日各向异性加热 Diurnal Anisotropic Heating
	DC	下坡曲率 Downslope Curvature
	EAFH	有效气流高度 Effective Air Flow Heights
	FPL	流动路径长度 Flow Path Length
	FW	流动宽度 Flow Width
	GEO	地质形态 Geomorphons
	HEI	高度 Height [Mean of Upslope Area]
	LF	地形 Landform
	LC	局部曲率 Local Curvature
	LDC	局部下坡曲率 Local Downslope Curvature
	LUC	局部上坡曲率 Local Upslope Curvature
	LSF	LS 系数 LS-Factor
	MBI	质量平衡指数 Mass Balance Index
	MSP	中坡位置 Mid-Slope Position
	MCA	修正集水区 Modified Catchment Area
	MF	形态特征 Morphometric Features
	MRR	MRRTFMRRTF
	MRV	MRVBFMRVBF
	NH	归一化高度 Normalized Height
	PLC	平面曲率 Plan Curvature
	PRC	剖面曲率 Profile Curvature
	PI	保护指数 Protection Index
RSP	相对坡度位置 Relative Slope Position	
SVF	天景系数 Sky View Factor	
SLH	坡度高度 Slope Height	
SLL	坡度长度 Slope Length	
SLO	坡度 Slope	
ELE	高程 Elevation	

	STH	标准化高度 Standardized Height
	SPI	溪流动力指数 Stream Power Index
	SA	表面区域 Surface Area
	TRI	地形崎岖指数 Terrain Ruggedness Index (TRI)
	TPI	地形位置指数 Topographic Position Index
	TWI	地形湿度指数 Topographic Wetness Index
	UC	上坡曲率 Upslope Curvature
	VD	山谷深度 Valley Depth
	VRM	矢量地形坚固性 Vector Terrain Ruggedness (VRM)
	VDCN	渠道网络垂直距离 Vertical Distance to Channel Network
	VS	天空可见度 Visible Sky
	WE	风力 Wind Exposition
	WSI	避风指数 Wind Shelter Index
气候因子	TEM	年均温度 Mean annual temperature
Climatic factors	PRE	年均降雨 Mean annual precipitation
土壤因子	SOIL	土壤类型 Soil type
Soil factors	TUT	土壤质地 Soil texture
人为因子	LAND	土地利用类型 Type of land use
Human factors	PLY	种植年限 Planting of years

1.5 随机森林预测模型

随机森林模型 (Random Forest Algorithm, RF) 是一种用于分类和回归的机器学习方法, 是随机分类与回归树集合的组合。利用 Boruta 特征指标选择, 将原始 75 个变量的特征矩阵随机生成影子特征矩阵, 并与原始变量特征矩阵组合成新的特征矩阵训练 RF 模型, 计算每个变量的相对重要性。选取影子特征矩阵相对重要性的最大值, 当原始变量特征矩阵相对重要性大于这个最大值则被标记为重要变量^[31]。

对随机森林模型进行训练时, 仅利用 2/3 样本作为训练数据, 此外 1/3 作为袋外数据 (out-of-bag, OOB) 检验预测精度, 以评估随机森林性能^[32]。将袋外数据重复运行了 100 次, 计算袋外数据预测误差对模型进行校准。

随机森林模型校准主要对三项模型参数进行调整。其中, 参数 nodesize 为拆分二叉树的最小节点大小; mtry 是指定节点用于二叉树的变量个数; ntree 是形成森林的树木数量, 默认值 500 足以达到稳定预测^[33]。例如以穴位均值作为样点观测值, 针对 mtry (2 ~ 15) 和 nodesize (3 ~ 10) 参数值范围, 设置构成 112 个参数组合方案, 并采用 10 折交叉验证方法, 选择 RMSE 值最小的参数组合方案, 最终确定模型最优参数值。

单个样点或栅格像元的随机森林预测值为其所有决策树加权线性组合^[26]的平均值, 表示为:

$$\hat{Y}(i) = \sum_{m=1}^n w_m(i) \times y_m(i), (i = 1, 2, 3 \dots) \quad (3)$$

式中, $\hat{Y}(i)$ 为黑土层厚度预测值, i 为第 i 个样点或者像元, $w_m(i)$ 为第 m 棵树的权重, $y_m(i)$ 为第 m 棵树的预测值。

在随机森林预测过程中, 基于样点均值以及各穴位 A、B、C 观测值分别进行特征指标选择、模型评估、参数优化调整, 获得最优随机森林模型进行黑土层厚度空间预测; 比较四种样本预测结果, 分析不同穴位随机观测对环境变量重要性、预测空间分布、模型精度的影响。

1.6 模型预测精度检验及不确定性分析

分别采用 10 折 (T1)、留一 (T2) 以及 5 倍 10 折 (T3) 三种交叉验证方法, 同时检验四种样本数据集下随机森林模型精度, 评价随机森林回归预测潜能; 采用的评价指标为平均绝对误差 (MAE)、均方根误差 (RMSE) 和决定系数 (R^2)^[34]。

$$R^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (4)$$

$$MAE = \sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)}{n} \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}} \quad (6)$$

式中, \hat{Y}_i 为预测值, Y_i 为观测值, \bar{Y}_i 为观测值的平均值, n 为样点个数。

利用同种交叉验证方法获得的不同样本集 MAE、RMSE、 R^2 值之间比较, 分析样点多穴位及单穴位对模型预测性能的影响。利用不同交叉验证算法对同一样本数据集的 MAE、RMSE、 R^2 值差异, 分析模型预测潜能; 当这 3 种评价指标值变化较小时, 表明模型稳定性较高, 具备较好空间预测潜能。

空间预测不确定性采用定量分析方法。应用分位数回归森林模型获取研究区 100 m 空间分辨率各栅格单元 5%、50%、95% 分位数的黑土层厚度预测值; 将每个像元 90% 预测区间设定为 5% 与 95% 分位数黑土层厚度预测值之间的差值, 其与 50% 分位数黑土层厚度预测值的比值, 即为该像元不确定性指数^[35]。

$$U_c(i) = \frac{\hat{Y}_{0.95}(i) - \hat{Y}_{0.05}(i)}{\hat{Y}_{0.5}(i)}, (i = 1, 2, 3 \dots) \quad (7)$$

式中, $U_c(i)$ 为不确定性指数, i 为第 i 个像元, $\hat{Y}_{0.95}(i)$ 为 0.95 分位数黑土层厚度预测值, $\hat{Y}_{0.05}(i)$ 为 0.05 分位数黑土层厚度预测值, $\hat{Y}_{0.5}(i)$ 为 0.5 分位数黑土层厚度预测值。

同时, 利用分位数回归森林计算的预测区间覆盖率 (PICP), 对栅格单元不确定性分析的准确性进行评估^[36]。例如, 当预测区间为 90%, 记为 PICP90, 表示在 5% 与 95% 分位数之间验证数据的覆盖比例。将观测值作为验证数据, 设黑土层厚度观测值为 Y_S , α 为预测区间, $\hat{Y}_{0.5 \pm \frac{\alpha}{2}}$ 为预测区间 α 上下界预测值, 当第 i 个样点有 $Y_S(i) \in [\hat{Y}_{0.5 - \frac{\alpha}{2}}(i), \hat{Y}_{0.5 + \frac{\alpha}{2}}(i)]$ 时, C_i 记为 1, 否则记为 0, PICP 可表示为:

$$PICP = \frac{1}{n} \sum_{i=1}^n C_i \quad (8)$$

式中, n 为样点数。

为直观反映模型预测不确定性分析的准确性水平, 绘制了不同观测穴位样点预测值不确定性分析的精度图。其中, 横、纵坐标分别代表预测区间 α 及其覆盖率 PICP。当精度图的点线越靠近 1:1 线, 则说明空间预测不确定性评估良好, 高于或者低于 1:1 线则表示不确定性分别被高估或者低估。

2 结果

2.1 黑土层厚度样点统计特征及勘察精度检验

各调查样点的观测穴位 A、B、C 的黑土层厚度平均值分别为 55.52、55.29、56.86 cm, 调查样点平均值为 55.56 cm (表 2)。样点平均黑土层厚度的标准差最小, 其数据离散程度均低于观测穴位 A、B、C。样点观测穴位 A、B、C 及其平均的变异系数均介于 10% ~ 100% 之间, 属于中等变异性; 偏度均大于 0, 说明数据呈正偏态, 数据分布主体集中在左侧; 峰度均小于 0, 说明数据呈矮胖分布。此外, 样点观测穴位 A、B、C 在 95% 置信度下的均值误差均高于其样点平均的均值误差。不同土壤类型下, 黑土的黑土层厚度测定结果最高为 68.06 cm, 白浆土的黑土层厚度测定结果最低为 34.40 cm (表 3)。

表 2 土壤样点黑土层厚度描述性统计

Table 2 Descriptive statistics of black soil layer thickness at soil sample sites ($n=357$)

	最小值 Minimum value	中位数 Median	最大值 Maximum values	平均值 Average value	标准差 Standard deviation	变异系数 Variation coefficient/%	偏度 Skewness	峰度 Kurtosis	置信度 Confidence level/ (95%)
样点平均 Sample point average/cm	0	49.33	100	55.56	30.95	55.71%	0.23	-1.45	3.22
穴位 A Observation point A/cm	0	50	100	55.52	31.41	56.58%	0.24	-1.45	3.27
穴位 B Observation point B/cm	0	50	100	55.29	31.33	56.67%	0.23	-1.46	3.26
穴位 C Observation point C/cm	0	50	100	55.86	31.12	55.70%	0.21	-1.45	3.24

表 3 主要土壤类型土壤样品黑土层厚度快速调查结果统计特征

Table 3 Statistical characteristics of the results of a rapid survey of black soil layer thickness in soil samples of major soil types

土壤类型 Soil types	暗棕壤 Dark brown soil	白浆土 Virgin soil	草甸土 Meadow soil	黑钙土 Black calcium soil	黑土 Dark soil	水稻土 Paddy soil	沼泽土 Marshy soil	棕色针叶林土 Brown coniferous forest soils
平均厚度 Average thickness/cm	39.52c	34.4c	63.43ab	41.58bc	68.06a	66.42ab	56.72abc	47.33*
样点数 Number of sample points	83	20	81	28	112	8	24	1

注：不同小写字母代表不同土壤类型间具有显著性差异 ($P < 0.05$)；* 棕色针叶林土为单个样点，不进行差异分析。下同。

Note: Different letters indicate significant differences among these soil types at $P < 0.05$; * the soil of brown coniferous forest is a single sample, and its difference analysis is not carried out. The same as below.

本调查获得的各样点平均黑土层厚度小于 Zhang 等^[16]在小流域平原地带调查的平均黑土层厚度 (80 cm)，但高于“二普”调查的平均厚度 (40 cm)^[12]，原因在于研究区以及黑土层厚度判定方法均存在差异。本研究基于暗沃表层分类标准判定黑土层，而“二普”《中国土种志》确定土壤发生分类 A 层为黑土层，前者已被国内外广泛接收和认可^[1,28]。

观测穴位 A、B、C 的观测误差 $\delta(s)$ 分别为 4.94、4.02、4.74 cm，分别是样点平均的 8.91%、7.24%、8.53%；其中穴位 A 的观测误差最大，离散程度也最大。

多穴位黑土层厚度快速勘察方法与剖面法获取的黑土层厚度基本一致 ($y = 0.9517x + 3.1252$)，两者间的皮尔森相关性高达 0.9744，决定系数 R^2 为 0.9494，说明该方法精确度较高，具有可行性，可以替代剖面调查方法^[37]。

2.2 随机森林模型参数优化与检验

以穴位均值作为样点观测值，采用 boruta 方法优选出 15 个用于黑土层厚度预测的环境因子协变量，由于随机森林模型预测受共线性影响较小，较多的环境协变量可提升预测能力^[30]。其中，遥感因子 5 个、地形因子 7 个、气候因子 2 个、土壤因子 1 个。人为因子未被选中，是因为耕地主要集中在黑土、黑钙土、草甸土和白浆土 4 种土壤类型上 (图 1)，土地利用类型信息已经被土壤类型所解释。此外，观测穴位 A、B、C 分别获取了 17、15、13 个环境因子协变量。

以穴位均值作为样点观测值，袋外数据 RMSE 介于 25.25 ~ 27.96 cm 之间，平均值为 26.71 cm；MAE 介于 21.28 ~ 24.36 cm 之间，平均值为 23.09 cm。袋外数据 RMSE 和 MAE 的平均值占研究区黑土层厚度平均值的 48.11% 和 41.59%；Zhang 等^[16]构建随机森林绘制的 1.39 km² 东北小流域黑土层厚度图，袋外数据 RMSE 占流域黑土层平均厚度 46.85%，均说明随机森林对测试数据的模型构建表现良好。观测穴位 A、B、C 袋外数据 RMSE 平均值分别为 26.87 cm、27.37 cm、26.92 cm，MAE 平均值分别为 23.19 cm、23.72 cm、23.10 cm。

以穴位均值作为样点观测值，采用的随机森林最优模型参数值 ($mtry = 15$ 和 $nodesize = 4$) 与默认模型参数 ($mtry = 2$ 和 $nodesize = 5$) 相比，最优模型参数值将模型性能 R^2 、RMSE、MAE 分别提高了 5.22%、4.60%、5.42%，表明构建模型时参数调优，能够提高模型结构的预测性能。在最优模型中，10 倍交叉验证 R^2 为 0.60，即对黑土层厚度空间变化解释力达到 60%，略高于 Liu 等^[25]对祁连山地区 14 万平方千米的土壤深度的解释力 ($R^2 = 0.59$)。此外，

观测穴位 A、B、C 随机森林最优模型参数值分别为 A(mtry = 17 和 nodesize = 3)、B(mtry = 13 和 nodesize = 4)、C(mtry = 13 和 nodesize = 3)。

2.3 黑土层厚度空间分布及其不确定性

土壤样点黑土层厚度预测值按土壤类型统计结果表明, 暗棕壤、白浆土、黑钙土、棕色针叶林土黑土层厚度平均预测值较测定值略高, 草甸土、黑土、水稻土、沼泽土的平均预测值较测定值略低, 但总体差异不大(表 3—4)。不同土壤类型下, 黑土的黑土层厚度平均预测值最高, 白浆土的黑土层厚度平均预测值最低, 与实际观测结果具有一致性。

张之一^[3]在全国第二次土壤普查中, 将黑土层厚度 ≤ 30 cm 定义为薄层黑土, 介于 30 ~ 60 cm 之间为中层黑土, >60 cm 为厚层黑土。研究区以穴位均值空间预测的黑土层厚度均值为 51.73 cm, 其中耕地平均黑土层厚度为 53.42 cm, 分布格局与其地形地貌特征吻合(图 3a), 平原地区黑土层厚度较高, 山地黑土层厚度较低。厚层黑土主要分布在松嫩平原和三江平原, 中层黑土分布在大兴安岭、小兴安岭以及东南山地, 薄层黑土分布在小兴安岭东南部地区以及东南山地西部地区(图 3d)。嫩江市处于毗邻内蒙古的嫩江东岸平原, 土地肥沃, 黑土地资源相对丰富, 黑土层厚度较高。此外, 松花江流域地势相对较低, 加上河流汇集泥沙堆积, 黑土层厚度也相对较高。

黑土层厚度分布的局部细节表明, 从山顶到山底黑土层厚度逐步变厚(图 3b); 从山地到平原, 由南至北黑土层厚度也逐步变厚(图 3c)。

表 4 主要土壤类型样点黑土层厚度空间预测结果统计特征

Table 4 Statistical characteristics of the results of spatial prediction of black soil layer thickness at sample sites of major soil types

土壤类型 Soil types	暗棕壤 Dark brown soil	白浆土 Virgin soil	草甸土 Meadow soil	黑钙土 Black calcium soil	黑土 Dark soil	水稻土 Paddy soil	沼泽土 Marshy soil	棕色针叶林土 Brown coniferous forest soils
平均厚度 Average thickness/cm	45.15cd	36.27d	60.05abc	50.55bcd	67.12a	64.17ab	55.03abc	57.42*
样点数 Number of sample points	83	20	81	28	112	8	24	1

注: 不同小写字母代表不同土壤类型间具有显著性差异 ($P < 0.05$); * 棕色针叶林土为单个样点, 不进行差异分析。Note: Different lowercase letters represent significant differences between soil types ($P < 0.05$); * Brown coniferous forest soils are single sample points and were not analysed for differences.

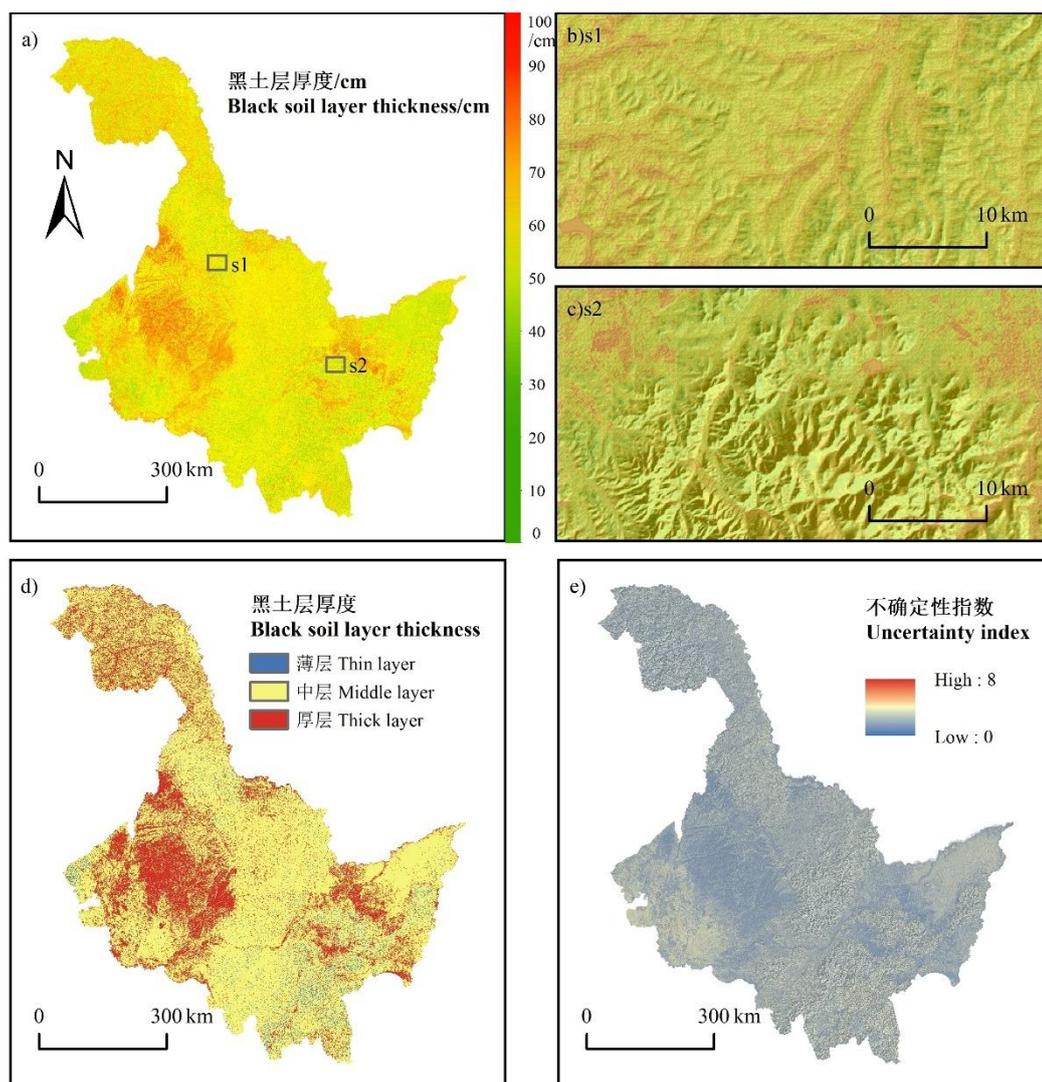


图 3 黑土层厚度预测及不确定性空间分布

Fig. 3 Spatial distribution of black soil layer thickness prediction and uncertainty

黑土层厚度空间分布不确定性指数 U_c 平均值仅为 2.13, 远低于不确定性指数中位数 4, 表明以穴位均值预测的不确定性指数总体较低(图 3e)。高不确定性指数主要分布在西北大兴安岭、北部小兴安岭以及东南山地, 一方面由于山地黑土层厚度较薄, 空间变异性高, 预测误差相对较高; 另一方面是由于山地可达性较差, 样点数目稀少。低不确定性指数分布在三江平原、松嫩平原、松花江流域以及嫩江地区, 这些区域地势开阔可达性好, 耕地分布集中连片, 样点调查密度较高。结合黑土层厚度不确定性空间信息, 黑土层厚度空间分布特征更为清晰。

3 讨论

3.1 样点观测穴位对环境协变量重要性影响

利用样点观测值, 通过 boruta 特征提取方法获取 15 个环境协变量, 不同环境变量对黑土层厚度预测的重要性有明显差异(图 4a)。而基于穴位 A、B、C 观测分别获取 17、15、13 个环境协变量(图 4b~图 4d), 其中共同的环境变量有 13 个, 重要性也因穴位不同发生变化。

以穴位均值作为样点观测值，黑土层厚度最重要的协变量有 GNDI、PRE、MRV、BAI、CND、MCA、SOIL。其中，GNDI、BAI 代表遥感因子，分别反映了黑龙江省植被指数、耕地和非耕地的分布状况，BAI 燃烧指数代表秸秆燃烧程度反映种植作物类型、生物量^[38]。燃烧明显的耕地区域黑土层较厚；未燃烧区的林地黑土层厚度较薄，Zhao 等^[38-39]研究很好论证了 BAI 等遥感指数是黑土层厚度主要协变量。PRE 代表气候因子，决定了黑土成土过程的水热条件，不仅影响土壤母质风化，也控制着植物和微生物生长，影响有机质合成、分解和迁移^[40]，进一步影响黑土层厚度。SOIL 代表土壤因子，反映土壤类型属性及其剖面结构特征，与黑土层厚度直接相关，黑土的黑土层较厚、暗棕壤的黑土层较薄^[41]。MRV、CND、MCA 代表地形因子，主要改变了黑土层成土母质、水分和热量等条件再分配，水分、母质在较低地形部位积累，在较高地形部位流失，热量的再分配表现为南坡大于北坡^[25,40]。

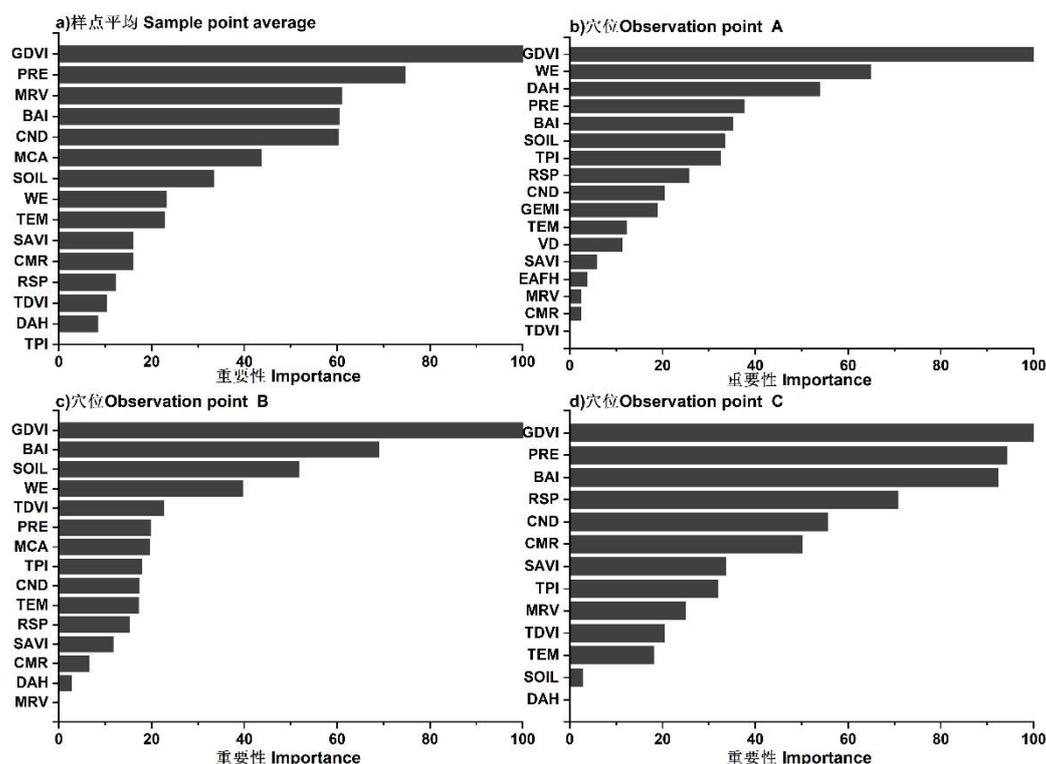


图 4 环境协变量的重要性排名：(a)样点平均；(b)观测穴位 A；(c) 观测穴位 B；(d) 观测穴位 C
Fig. 4 Importance ranking of environmental covariates: (a) sample point average; (b) observation point A; (c) observation point B; (d) observation point C

遥感因子、地形因子、气候因子、土壤因子对黑土层厚度的重要性值分别为 37.37%、38.52%、17.97%、6.14%，各因子单个变量重要性平均值分别为 7.47%、5.50%、8.99%、6.14%，说明地形因子与单变量平均气候因子解释黑土层厚度能力最高（图 4a）。但各调查样点不同观测穴位的空间变异性对变量重要性值也产生了明显影响（图 4b~图 4d），穴位 A 的遥感因子、地形因子、气候因子、土壤因子的重要性值变化量分别为 2.12%、8.14%、7.13%、1.11%；穴位 B 的各因子变化量分别为 13.67%、11.19%、8.94%、6.46%；穴位 C 各因子变化量分别为 12.45%、7.7%、0.92%、5.67%。其中，穴位 B 产生的变量重要性值变化最大，这与观测穴位随机选择有关。

3.2 样点观测穴位对黑土层厚度空间预测分布的影响

基于样点不同穴位 A、B、C 观测值与平均值预测的黑土层厚度空间分布（图 5a~图 5c；图 3a）具有明显差异，表明基于调查样点单穴位观测的随机性，会影响黑土层厚度空

间分布预测。

以调查样点多穴位均值预测的黑土层厚度空间分布（图 3a）作为对照底图，分析样点不同观测穴位 A、B、C 对空间分布预测的影响。对照结果表明，观测穴位 A 预测的区域黑土层平均厚度增加了 0.64 cm，在松嫩平原四周以及三江平原增加最明显，在大兴安岭、小兴安岭以及东南山地也被严重高估，而在松嫩平原核心区被显著低估（图 5a, 图 5d）。对观测穴位 B 而言（图 5b, 图 5e），黑土层平均厚度降低了 0.95 cm，在松嫩平原四周、小兴安岭被显著高估，在松嫩平原、大兴安岭和东南山地则被明显低估。基于观测穴位 C 预测研究区平均厚度值（图 5c）提高了 0.14 cm，在松嫩平原、大兴安岭和小兴安岭被显著高估，在松嫩平原四周、三江平原和东南山地被显著低估（图 5f）。

总之，样点不同穴位调查与预测的黑土层厚度存在差异，黑土层厚度被高估或低估具有随机性。

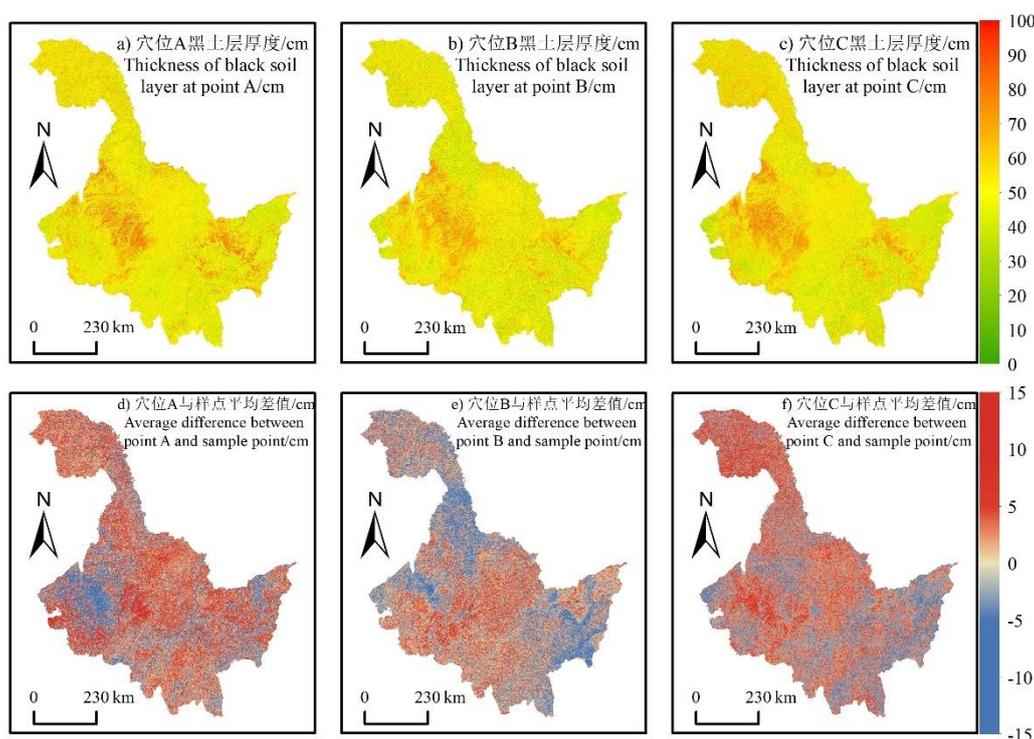


图 5 基于各观测穴位与调查样点的黑土层厚度空间预测分布差异

Fig. 5 Differences in the spatially predicted distribution of black soil layer thickness based on each observation site and the survey sample sites

3.3 样点观测穴位对黑土层厚度空间预测不确定性的影响

对比分析样点不同观测穴位 A、B、C 观测值与平均值定量分析的黑土层厚度空间分布不确定性指数（图 6a~图 6c; 图 3e）之间的差异，表明样点单穴位观测也会影响空间预测不确定性。以调查样点多穴位均值预测的黑土层厚度空间不确定性指数分布（图 3e）作为对照底图，穴位 A 预测的区域黑土层厚度不确定性指数平均值降低了 0.18（图 6d），说明以穴位 A 预测的黑土层厚度空间分布总体上可解释性增加，主要体现在大兴安岭和小兴安岭调查样点较为稀疏区域，而在调查样点稠密的松嫩平原黑土层厚度空间分布解释性却显著降低。同样，基于穴位 B 预测研究区黑土层厚度不确定性指数平均值（图 6b）下降了 0.22（图 6e），降低了松嫩平原黑土层厚度空间分布解释性，增加了大兴安岭、小兴安岭以及东南山地的黑土层厚度空间分布解释性。对于穴位 C 而言（图 6c, 图 6f），黑土层厚度不确定性指数平均值下降了 0.16，总体不确定性呈下降趋势，在山地和平原黑土层厚度

不确定性指数分别具有增加和下降趋势。

总之，样点不同穴位调查与预测的黑土层厚度不确定性指数之间存在差异，不确定性增加和下降具有较大的随机性。

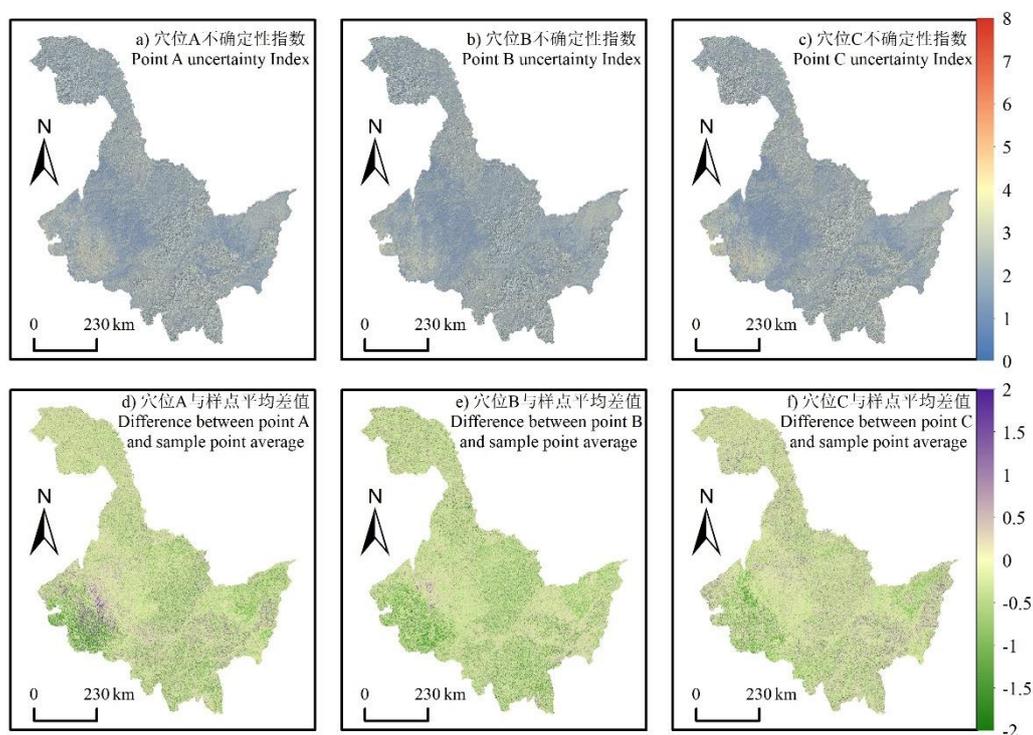


图 6 基于各观测穴位与调查样点的黑土层厚度空间不确定性指数分布差异

Fig. 6 Difference in the distribution of spatial uncertainty indices of the thickness of the black soil layer based on each observation point and the survey sample sites

从空间预测不确定性评估的准确性水平看，当预测值在预测区间的占比曲线越贴近 1:1 线，即预测区间 α 与预测占比 PICP 值相当，表明黑土层厚度空间预测不确定性的准确性水平越高。基于调查样点平均观测值的预测区间 α 介于 0 ~ 0.5 之间，略微低于 1:1 线，说明空间不确定性的准确性水平被低估，即预测区间宽度相比于预测值略窄；在预测区间 α 介于 0.5 ~ 1 之间相对贴合 1:1 线，空间不确定性的准确性水平较高。穴位 A 在预测区间 α 介于 0 ~ 1 之间离 1:1 线稍远，空间不确定解释能力稍弱，其准确性水平被明显低估。而穴位 B 离 1:1 线最远，表明其空间不确定性的准确性水平最低，即预测区间宽度太窄，不确定性能力被低估。穴位 C 在预测区间 α 介于 0 ~ 1 之间总体略低于 1:1 线，空间不确定性的准确性水平也被低估。

总之，基于调查样点平均观测值的空间预测不确定性的准确性水平相对较高，而基于单穴位观测值预测的空间不确定性的准确性水平均被低估，总体表现为穴位 C 优于穴位 A，进一步优于穴位 B。

3.4 样点观测穴位对黑土层厚度空间预测精度的影响

通过对比分析 T1、T2、T3 三种交叉验证方法、4 种样本数据集下的随机森林空间预测精度差异，表明样点多穴位观测较任何单穴位观测的模型预测精度高（表 5），从空间预测角度证明了样点多穴位观测的必要性。如在 T1 交叉验证方法下，观测穴位 A、B、C 相比于样点平均 RMSE 值分别提高了 30.11%、32.19%、30.80%，MAE 值升高明显， R^2 降幅均较大，极大降低了模型预测精度。同样，在 T2、T3 交叉验证方法下均得出相似结果。

显然，多穴位快速调查获取样点均值法，能够有效提高模型预测精度，得到更准确的黑土层厚度空间分布。

多穴位快速调查获取样点均值法，能代表微域范围内黑土层厚度整体水平，而单穴位观测则只能反映微域范围内某个点的实际水平，不能代表微域整体水平。因为栅格化的环境是样点所在栅格的平均值，均值与均值的配对更能反映实际存在的规律。所以，在模型预测精度上，多穴位快速调查获取样点均值法预测精度较高。

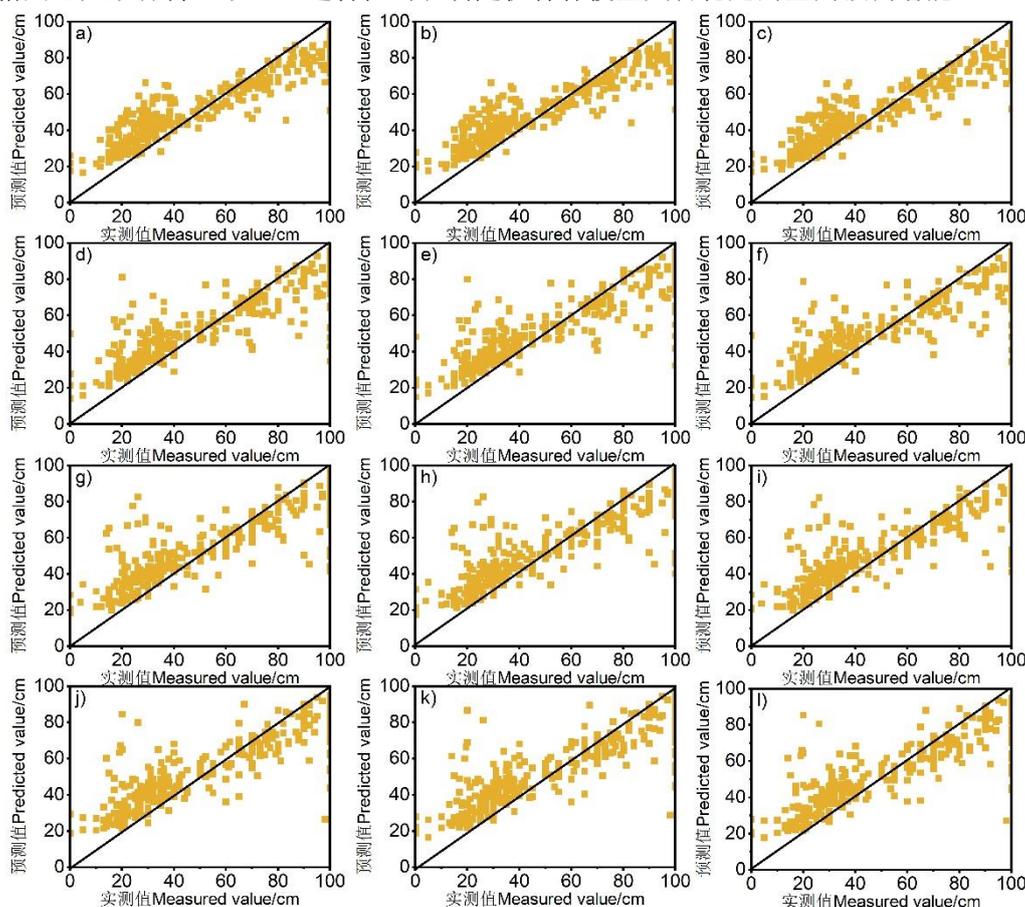
表 5 三种交叉验证方法获取 4 种样本随机森林模型精度值

Table 5 Three cross-validation methods to obtain four-sample random forest model accuracy values

验证方法 Validation methods	样点平均 Sample point average			穴位 A Observation point A			穴位 B Observation point B			穴位 C Observation point C		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
	T1/cm	0.60	21.59	19.28	0.24	28.09	24.63	0.18	28.54	23.98	0.16	28.24
T2/cm	0.61	21.58	19.31	0.23	28.20	24.70	0.18	28.61	23.94	0.16	28.21	23.26
T3/cm	0.59	21.57	19.33	0.25	27.94	24.48	0.18	28.54	24.04	0.17	28.07	23.09

3.5 基于随机森林模型的黑土层厚度空间预测潜能稳定性

基于样点均值数据集在不同交叉验证方法下的模型拟合精度值差异较小（表 5），说明模型相对稳定，即模型具备较好的空间预测潜能。以 RMSE 为例，样点平均从 10 折交叉验证（T1）到留一交叉验证（T2）再到 5 倍 10 折交叉验证（T3），其值从 21.59 cm 到 21.58 cm 再到 21.57 cm，基本稳定。同理，观测穴位 A、B、C 在不同交叉验证方法下的预测精度具有较好稳定性。观察 T1、T2、T3 三种交叉验证方法在 4 种样本下的预测值与实测值之间的散点图（图 7），点散射均未明显出现；同一样本下不同交叉验证方法的散点图格局也大致保持一致。上述特征均表明随机森林模型具备稳定的空间预测潜能。



注：基于样点平均观测值：(a) T1, (b) T2, (c) T3；基于穴位 A 观测值：(d) T1, (e) T2, (f) T3；基于穴位 B 观测值：(g) T1, (h) T2, (i) T3；基于穴位 C 观测值：(j) T1, (k) T2, (l) T3。Note: Based on sample mean observations (a) T1, (b) T2, (c) T3; based on point A observations: (d) T1, (e) T2, (f) T3; based on point B observations: (g) T1, (h) T2, (i) T3; based on point C observations: (j) T1, (k) T2, (l) T3.

图 7 三种不同方法交叉验证散点图

Fig. 7 Cross-validation of scatterplots by three different methods

4 结论

(1) 建立了调查样点黑土层厚度多穴位“浅层挖掘+深层土钻”快速勘察新方法，检验结果表明，与剖面法获取的黑土层厚度基本一致，精度达到 94.94%，可以替代剖面调查方法。(2) 基于样点的空间预测证明了多穴位观测的科学性及必要性。原因在于土壤样点微域空间依然存在变异性，单穴位观测随机性改变了模型变量重要性值，也影响了黑土层厚度空间分布预测；相比于调查样点多穴位平均值的模型预测，基于单穴位观测值的预测精度显著下降，不确定性评估的准确性也降低。(3) 构建及优化的随机森林模型，具有稳定的黑土层厚度空间预测潜能。通过特征选择、模型性能评估、模型参数调优等步骤，确定了随机森林最优模型。随机森林最优模型与默认模型相比，将模型性能指标 R^2 提高了 5.22%，RMSE、MAE 分别降低了 4.60%、5.42%，黑土层厚度空间变异解释力 R^2 达到 60%。对于不同样本集在相同规模下的黑土层厚度空间预测，随机森林模型具有极强稳定性。(4) 基于调查样点与空间预测统计的当前黑龙江省平均黑土层厚度分别为 55.56cm 和 51.73 cm，预测耕地平均黑土层厚度为 53.42 cm。其中，厚层黑土主要分布在东北三江平原和西南松嫩平原，中层黑土分布在大兴安岭、小兴安岭以及东南山地，薄层黑土分布在小兴安岭东南部地区以及东南山地西部地区。黑土层厚度空间分布格局，可为研究揭示黑土层变薄特征、强化黑土资源保护等提供科学依据。

参考文献 (References)

- [1] Nachtergaele F. Soil taxonomy—a basic system of soil classification for making and interpreting soil surveys: Second edition, by Soil Survey Staff, 1999, USDA–NRCS, Agriculture Handbook number 436, Hardbound[J]. Geoderma, 2001, 99: 336-337.
- [2] Sadras V O, Calviño P A. Quantification of grain yield response to soil depth in soybean, maize, sunflower, and wheat[J]. Agronomy Journal, 2001, 93(3): 577-583.
- [3] Zhang Z Y. The thickness changes of Ah horizon after the phaeozems cultivated[J]. Journal of Heilongjiang August First Land Reclamation University, 2010, 22(5):1-3. [张之一. 黑土开垦后黑土层厚度的变化[J]. 黑龙江八一农垦大学学报, 2010, 22(5) :1-3.]
- [4] Duan X W, Xie Y, Ou T H, et al. Effects of soil erosion on long-term soil productivity in the black soil region of northeastern China[J]. Catena, 2011, 87(2): 268-275.
- [5] Gu Z J, Xie Y, Gao Y, et al. Quantitative assessment of soil productivity and predicted impacts of water erosion in the black soil region of northeastern China[J]. Science of the Total Environment, 2018, 637/638: 706-716.
- [6] Pieters A, Baruch Z. Soil depth and fertility effects on biomass and nutrient allocation in jaraguagrass[J]. Journal of Range Management, 1997, 50(3): 268-273.
- [7] Gordeef T P. Description of the soils and rocks amid which has been found a mammoth - task. The society for the study of Manchuria, Harbin Bul.6, 1926.

- [8] Pendleton R L, Hou K. Reconnaissance soil survey of the Harbin region[J]. Soil Bulletin, 1935, 11: 42-106.
- [9] Heilongjiang Province Agriculture Department. Heilongjiang soil[M]. Printed by the Department of Agriculture of Heilongjiang Province, 1962:15-20. [黑龙江省农业厅. 黑龙江土壤[M]. 黑龙江省农业厅印, 1962:15-20.]
- [10] Heilongjiang Provincial Land Management Bureau, Heilongjiang Provincial Soil Census Office. Heilongjiang soil [M]. Beijing: Agriculture Press, 1992. [黑龙江省土地管理局, 黑龙江省土壤普查办公室. 黑龙江土壤[M]. 北京: 农业出版社, 1992.]
- [11] National Soil Census Office. Soil of China[M]. Beijing: China Agriculture Press, 1998. [全国土壤普查办公室. 中国土壤[M]. 北京: 中国农业出版社, 1998.]
- [12] National Soil Census Office. Soil species of China [M]. Beijing: China Agriculture Press, 1995. [全国土壤普查办公室. 中国土种志[M].北京: 中国农业出版社, 1995.]
- [13] Lu J L. Phaeozem degradation and sustainable agriculture in China[J]. Journal of Soil and Water Conservation, 2001, 15(2): 53-55, 67.[陆继龙. 我国黑土的退化问题及可持续农业[J]. 水土保持学报, 2001, 15(2): 53-55, 67.]
- [14] Shen B, Fan J R, Pan Q B, et al. General situation of pilot project of comprehensive prevention and control of soil and water loss of chernozem region in northeast of China[J]. Soil and Water Conservation in China, 2003(11): 7-8. [沈波, 范建荣, 潘庆宾, 等. 东北黑土区水土流失综合防治试点工程项目概况[J]. 中国水土保持, 2003(11): 7-8.]
- [15] Liu K, Wei M H, Dai H M, et al. Spatiotemporal variation of black soil layer thickness in black soil region of northeast China[J]. Geology and Resources, 2022, 31(3):434-442, 394. [刘凯, 魏明辉, 戴慧敏, 等. 东北黑土区黑土层厚度的时空变化[J]. 地质与资源, 2022, 31(3):434-442, 394.]
- [16] Zhang S, Liu G, Chen S L, 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.
- [17] Zhang F P, Gao Z, Ma Q Q, et al. Construction of minimum data set for soil quality assessment in the Dunhuang oasis[J]. Chinese Journal of Soil Science, 2017, 48(5): 1047-1054. [张福平, 高张, 马倩倩, 等. 面向敦煌绿洲土壤质量评价的最小数据集构建研究[J]. 土壤通报, 2017, 48(5): 1047-1054.]
- [18] Zhang S W, Song X X, Wang Y, et al. Estimation of effective soil layer thickness of newly increased cultivated land based on GPR envelope detection[J]. Journal of Anhui Science and Technology University, 2022, 36(5): 54-64. [张世文, 宋孝心, 王阳, 等. 基于探地雷达包络检波的新增耕地有效土层厚度估算[J]. 安徽科技学院学报, 2022, 36(5): 54-64.]
- [19] Ziadat F M. Prediction of soil depth from digital terrain data by integrating statistical and visual approaches[J]. Pedosphere, 2010, 20(3): 361-367.
- [20] Gessler P E, Moore I D, McKenzie N J, et al. Soil-landscape modelling and spatial prediction of soil attributes[J]. International Journal of Geographical Information Systems, 1995, 9(4): 421-432.
- [21] 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.
- [22] Scarpone C, Schmidt M G, Bulmer C E, et al. Modelling soil thickness in the critical zone for Southern British Columbia[J]. Geoderma, 2016, 282: 59-69.
- [23] Lacoste M, Mulder V L, Richer-de-Forges A C, et al. Evaluating large-extent spatial modeling approaches: A case study for soil depth for France[J]. Geoderma Regional, 2016, 7(2): 137-152.
- [24] Lu Y Y, Liu F, Zhao Y G, et al. An integrated method of selecting environmental covariates for predictive soil depth mapping[J]. Journal of Integrative Agriculture, 2019, 18(2): 301-315.
- [25] Liu F, Yang F, Zhao Y G, et al. Predicting soil depth in a large and complex area using machine learning and

- environmental correlations[J]. *Journal of Integrative Agriculture*, 2022, 21(8): 2422-2434.
- [26] Meinshausen N, Ridgeway G. Quantile regression forests[J]. *Journal of Machine Learning Research*, 2006, 7(6): 983-999.
- [27] Zhang Z Y, Zhai R C, Cai D L. Introduction to soil series in Heilongjiang Province[M]. Harbin: Harbin Cartographic Publishing House, 2006. [张之一, 翟瑞常, 蔡德利. 黑龙江土系概论[M]. 哈尔滨: 哈尔滨地图出版社, 2006.]
- [28] Gong Z T, Chen Z C, Shi X Z, et al. Chinese Soil Taxonomy: Theory, methodology and practice [M]. Beijing: Science Press, 1999. [龚子同, 陈志诚, 史学正, 等. 中国土壤系统分类: 理论·方法·实践[M]. 北京: 科学出版社, 1999.]
- [29] Xie Y, Gao Y, Gu Z J, et al. Potential hazard assessment of slope farmland soil erosion in black soil region of northeastern China[J]. *Science of Soil and Water Conservation*, 2020, 18(6):105-114.[谢云, 高燕, 顾治家, 等. 东北黑土区坡耕地水土流失危险程度评价[J]. *中国水土保持科学*, 2020, 18(6):105-114.]
- [30] Takoutsing B, Heuvelink G B M. Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors[J]. *Geoderma*, 2022, 428: 116192.
- [31] Chen Y, Ma L X, Yu D S, et al. Comparison of feature selection methods for mapping soil organic matter in subtropical restored forests[J]. *Ecological Indicators*, 2022, 135: 108545.
- [32] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [33] Kuhn M. Building predictive models in R using the caret package[J]. *Journal of Statistical Software*, 2008, 28(5): 1-26.
- [34] Armstrong J S, Collopy F. Error measures for generalizing about forecasting methods: Empirical comparisons[J]. *International Journal of Forecasting*, 1992, 8(1): 69-80.
- [35] Tian Z Y, Liu F, Liang Y, et al. Mapping soil erodibility in southeast China at 250 m resolution: Using environmental variables and random forest regression with limited samples[J]. *International Soil and Water Conservation Research*, 2022, 10(1): 62-74.
- [36] Malone B P, McBratney A B, Minasny B. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes[J]. *Geoderma*, 2011, 160(3/4): 614-626.
- [37] Zhang G L, Li D C. Manual of soil description and sampling[M]. Beijing: Science Press, 2016. [张甘霖, 李德成. 野外土壤描述与采样手册[M]. 北京: 科学出版社, 2016.]
- [38] Yang G Y, Zhao H M, Tong D Q, et al. Impacts of post-harvest open biomass burning and burning ban policy on severe haze in the Northeastern China[J]. *Science of the Total Environment*, 2020, 716: 136517.
- [39] Zhao H M, Yang G Y, Tong D Q, et al. Interannual and seasonal variability of greenhouse gases and aerosol emissions from biomass burning in Northeastern China constrained by satellite observations[J]. *Remote Sensing*, 2021, 13(5): 1005.
- [40] Geng Z C, Dai W. Agrology[M]. Beijing: Science Press, 2011. [耿增超, 戴伟. 土壤学[M]. 北京: 科学出版社, 2011.]
- [41] Bei'an Soil Census Office. Bei'an soil[M]. Printed by Bei'an Soil Census Office, 1985. [北安市土壤普查办公室. 北安土壤[M]. 北安市土壤普查办公室印, 1985.]

(责任编辑: 檀满枝)