

基于资源一号卫星影像的天然次生林土壤有机碳反演

甄贞¹, 丁建业¹, 赵杨¹, 赵颖慧¹, 魏庆彬^{1,2,3}

1. 东北林业大学 林学院 森林生态系统可持续经营教育部重点实验室, 哈尔滨 150040;

2. 黑龙江省林业科学院, 哈尔滨 150081;

3. 哈尔滨师范大学 地理科学学院, 哈尔滨 150025

摘要: 森林土壤有机碳 SOC (Soil Organic Carbon) 是反映森林土壤质量的重要标准, 很大程度上影响森林树木的生长, 对林业的可持续发展起着十分重要的作用。探索应用高光谱影像反演天然次生林土壤有机碳含量的可能性至关重要, 可以为长时间序列、大尺度的森林土壤有机碳估测提供技术支持。本研究以天然次生林 SOC 为研究对象, 在东北林业大学帽儿山实验林场随机抽取 67 个样点, 分别采集 0—5、5—15、15—30 cm 土壤并实测 SOC 含量, 并取三层均值作为 0—30 cm SOC 含量, 然后基于资源 1F 高光谱影像, 计算光谱曲线的一阶微分、二阶微分、倒数对数、植被指数特征, 并结合数字高程模型 DEM (Digital Elevation Model)、土壤水分和森林地上生物量 AGB (Aboveground biomass) 数据集, 使用递归特征消除法进行特征筛选。应用随机森林 RF (Random Forest)、极端梯度提升 XGBoost (extreme Gradient Boosting)、支持向量机回归 SVR (Support Vector Regression)、最小二乘回归 OLS (Ordinary Least Squares Regression) 共 4 种模型, 分别对 SOC 进行估测, 并选择最佳模型反演帽儿山实验林场不同深度 SOC 含量。结果表明: 在土壤分层反演模型中均为 XGBoost 估测精度最高, 0—30 cm、0—5 cm、5—15 cm、15—30 cm 层模型 R^2 分别为 0.54、0.54、0.46 和 0.30, RMSE 分别为 21.28、44.25、15.72 和 12.56 g/kg。反演得到的帽儿山林场天然次生林 0—30、0—5、5—15 和 15—30 cm 共 4 层的 SOC 平均值分别为 67.20、88.87、46.92 和 40.12 g/kg, 森林 SOC 含量随土壤深度增加而减小。不同林分类型的 SOC 含量存在差异, SOC 含量排序从大到小为阔叶林、针阔混交林、针叶林。高光谱影像所包含的波段信息为模型的建立提供了可能, 但是过多的波段数量造成了数据冗余, 导致模型估测精度的降低, 采用递归特征消除法可以筛选出最佳特征组合, 有效降低特征数量, 提升模型估测精度。710—850 nm 波段的微分特征对天然次生林 SOC 的反演较为有效, 地形因子对于 15 cm 以上的 SOC 影响较大, 土壤水分和地上生物量对于 5—15 cm 层 SOC 影响较强。综上, 应用高光谱影像结合 DEM、土壤水分及地上生物量数据可以有效地反演天然次生林 SOC, 为基于多期高光谱影像的长时间序列及大尺度天然次生林 SOC 反演提供技术支持。

关键词: 天然次生林, 土壤有机碳 SOC (Soil Organic Carbon), 高光谱, 机器学习, 特征选择

中图分类号: S7/P2

引用格式: 甄贞, 丁建业, 赵杨, 赵颖慧, 魏庆彬. 2025. 基于资源一号卫星影像的天然次生林土壤有机碳反演. 遥感学报, 29(9): 2671–2685

Zhen Z, Ding J Y, Zhao Y, Zhao Y H and Wei Q B. 2025. Inversion of soil organic carbon in natural secondary forest based on ZY1F images. National Remote Sensing Bulletin, 29(9): 2671–2685 [DOI: 10.11834/jrs.20254333]

1 引言

森林是以乔木和其他木本植物为主体的生物生态系统, 陆地森林生态系统储存了约 24770 亿 t 碳, 在碳同化中起着至关重要的作用 (MacDicken, 2015)。在森林生态系统中, 碳被保存在 5 个主要

碳库中, 这 5 个碳库分别是森林地上生物量、地下生物量、枯木、落叶和土壤 (Sahoo 等, 2019)。在 5 个碳库中, 土壤已被确定为最大的碳库, 可以容纳的碳储量是树木和枯叶总和的 4 倍 (Broderick 等, 2015)。土壤碳库是由土壤有机碳 SOC (Soil Organic Carbon) 和土壤无机碳共同组成, 土壤无

收稿日期: 2009-01-06; 预印本: 2025-01-04

基金项目: 国家重点研发计划青年科学家项目 (编号: 2023YFF1305900); 森林生态系统可持续经营教育部重点实验室开放基金项目 (编号: KFJJ2021YB04)

第一作者简介: 甄贞, 研究方向为林业遥感、生态遥感的理论 and 应用。E-mail: zhenzhen@nefu.edu.cn

通信作者简介: 魏庆彬, 研究方向为环境生态学。E-mail: wqb0816@163.com

机碳在土壤碳库中的占比相对较小（张林，2010），而土壤有机碳的占比较大，据估计当前的全球土壤有机碳库介于1395—2200 Pg，是大气碳库或陆地植被碳库的2—3倍（罗梅等，2020）。SOC是指通过微生物作用所形成的腐殖质、动植物残体和微生物体中的碳元素（杜雪和王海燕，2022），SOC的微小变化将会对整个陆地生态系统碳循环产生显著影响（马志良和赵文强，2020）。森林SOC会影响森林土壤中的生物、化学和物理过程，在决定土壤质量和森林生物量等方面起着重要作用，也是林地质量评价、林业经营管理的重要参考指标（Stevens等，2013；游浩辰，2012）。

传统的土壤有机碳（SOC）测定依赖于对采样点土壤性质的分析，其准确性受限于采样点的数量。在大规模研究区域（如景观尺度）进行传统采样不仅耗费大量人力物力，而且难以捕捉土壤的连续性和复杂变化（李江，2008）。随着遥感技术的进步，遥感数据提供了一种替代方案，它能够覆盖更广阔的空间，提供时间一致且全面的光谱波段数据，从而实现森林SOC连续分布的反演（Broderick等，2015）。目前，遥感技术在估测森林SOC时主要依赖于多光谱数据，如Landsat系列影像（Odebiri等，2020；Bhunja等，2019）和Sentinel-2（Suleymanov等，2024）。相比之下，高光谱影像由于其波段众多、光谱分辨率高，以及能够进行精细的光谱差异分析，可以更准确地反演植被生长状况，在植被遥感研究与应用中表现出强大优势（Vane和Goetz，1993），也为量化分析植物理化特征与光谱特征的关系提供了强有力的数据源（王富强，2022）。尽管森林土壤受到地上植被的遮挡，使得直接使用被动遥感技术估测SOC具有挑战性，但研究表明，森林SOC与地表凋落物和地上生物量密切相关（Wan等，2019）。因此，可以通过提取高光谱遥感影像中反映植被生长状况的波段信息，进而有效地估测森林SOC。中国的资源一号02E卫星（ZY-1F）于2021年12月26日发射，作为高光谱业务卫星，运行在太阳同步轨道，设计寿命8 a，回归周期55 d。卫星搭载的高光谱相机能在高信噪比下捕获166个波段，支持精细地物光谱调查，为森林SOC的大尺度估测创造了条件。

森林土壤有机碳（SOC）受环境因素影响显著

（Conforti等，2016），导致其在空间分布上表现出显著的异质性（即不同区域的SOC含量差异显著）。这一特性对传统的回归模型提出了挑战。而机器学习模型在处理复杂和非线性问题方面具有突出优势。机器学习模型能够通过样本数据预测变量和响应变量间的训练与学习，发掘反演参数（响应变量）与遥感影像数据（预测变量）之间的关系，从而构建普适性较好且精度高的反演模型（王桥，2021），因此可以有效应对部分因机理不清或过程复杂而导致反演困难的情况（汪静平等，2023）。

因此，为对天然次生林的SOC进行有效估测，本研究拟利用国产星载高光谱遥感影像（ZY-1F）结合数字高程模型DEM（Digital Elevation Model）等数据集，应用随机森林RF（Random Forest）、极端梯度提升XGBoost（extreme Gradient Boosting）、支持向量机回归SVR（Support Vector Regression）、最小二乘回归OLS（Ordinary Least Squares Regression）共4种模型，分别对SOC进行估测，并探讨4种模型对于不同层深SOC的反演精度。旨在构建一套应用星载高光谱遥感数据进行森林SOC间接估测的流程与方法，评估国产高光谱卫星ZY-1F在反演天然次生林SOC方面的潜力；并探索高光谱特征对天然次生林下不同深度（0—30 cm、0—5 cm、5—15 cm和15—30 cm）SOC反演的敏感程度。

2 材料与方法

2.1 研究区域概况

研究区域位于黑龙江省尚志市东北林业大学帽儿山实验林场，地理位置处于长白山系张广才岭西坡小岭的余脉，地理坐标为127°29'E—127°44'E，北纬45°14'N—45°29'N，林场总面积26496 hm²，详见图1。帽儿山实验林场属于温带湿润地区、大陆性季风气候，年平均气温2.8℃左右，年降水量为723 mm左右。帽儿山实验林场植被属长白植物区系，林分以红松、落叶松、桦树为主，平均森林覆盖率为95%，森林总蓄积量为350万m³。土壤类型以暗棕壤为主，总面积23204 hm²，谷地草甸土和沼泽土只有少量分布，暗棕壤共有3个亚类，包括原始暗棕壤，典型暗棕壤和草甸暗棕壤，其中以典型暗棕壤为主，面积为21207 hm²；原始

暗棕壤主要分布于陡坡上部, 面积为 757 hm²; 草甸暗棕壤分布于缓坡下部, 面积为 1240 hm² (姜航 等, 2015)。

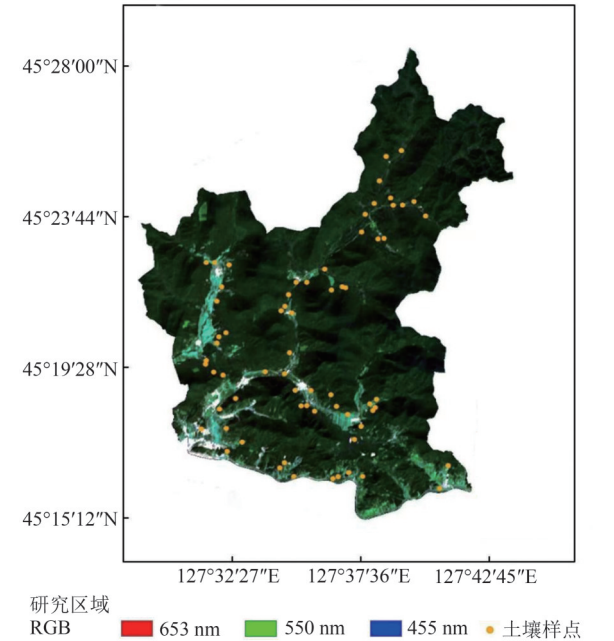


图1 研究区地理位置及采样点分布图(背景影像为2022年资源1F卫星影像标准假彩色波段显示)

Fig. 1 Location of study area and spatial distribution of sample points

2.1 土壤样本采集及处理

本研究采用随机抽样的方式确定候选土壤采样点。在实际采样的过程中, 调整、删除不可及采样点, 最终确定采样点共 67 个, 并于 2022 年 8 月采集土壤样品。表层土壤 (0—30 cm) 中的有机碳是土壤中参与全球碳循环最活跃的部分 (Chen 等, 2024), 因此选取采样深度为 0—30 cm。每个采样点设置 1 m×1 m 样方, 使用直径 38 mm 土钻按照 0—5、5—15 和 15—30 cm 分层钻取 3 个重复土样。将 3 个重复土样混合后密封保存, 带回实验室风干, 使用镊子剔除风干后土样中的植物根叶等有机物残体, 通过 1 mm 筛, 充分混合后, 从中取出 10—20 g 细磨, 全部通过 0.25 mm 筛, 使用 multi N/C 3100 总有机碳分析仪测定 SOC 浓度含量, 每个样点测三次取平均值作为该样点的土壤有机碳浓度值。将土壤样方 SOC 实测值作为所对应像元的 SOC 含量, 取 0—5 cm、5—15 cm、15—30 cm SOC 含量平均值作为该点 0—30 cm SOC 含量值。土壤分层实测 SOC 含量的基本统计结果如表 1 所示; 不同土壤分层中实测 SOC 含量的相关性分析如表 2 所示。

表1 土壤有机碳含量实测值描述性统计

Table 1 Descriptive statistics of measured soil organic carbon content

层深/cm	最小值	最大值	中值	平均值	标准差	变异系数/%
0—5	34.83	275.80	108.43	123.23	65.19	52.91
5—15	32.59	121.50	51.03	59.04	21.24	35.98
15—30	30.18	101.45	42.98	48.24	15.14	31.38
0—30	34.37	156.56	68.65	76.84	30.92	40.24

表2 不同深度土壤有机碳相关性分析

Table 2 Correlation analysis of soil organic carbon at different depths

层深/cm	0—5 cm	5—15 cm	15—30 cm
0—5	1		
5—15	0.63**	1	
15—30	0.37	0.32	1

注: **表示在 0.01 级别相关性显著; *表示在 0.05 级别相关性显著。

2.3 ZY-1F 卫星数据及预处理

本次研究选取的遥感数据为 ZY-1F 卫星高光谱相机载荷数据, 包含可见光近红外 VNIR (Visible-near-infrared) 和短波红外 SWIR (Shortwave

infrared) 166 个光谱波段, 具体参数见表 3。本研究选取的 ZY-1F 卫星 ASHI 数据, 获取时间为 2022 年 9 月 9 日, 数据质量好, 无积雪、云等干扰因素, 能够满足研究区域 SOC 反演的需要。

表3 ZY-1F 高光谱卫星 AHSI 相机参数

Table 3 The AHSI camera parameters of ZY-1F hyperspectral satellite

光谱范围/nm	光谱通道/个	光谱分辨率/nm	空间分辨率/m
VNIR: 395—1040	76	优于 10	30
SWIR: 100—2480	90	优于 20	

对 ZY-1F 高光谱影像的预处理包括辐射定标、Flaash 大气校正、几何校正以及正射校正。ZY-1F

影像由于受到外部环境、仪器自身噪音以及目标本身等多种因素的影响, 获取的原始光谱数据存在较大噪声, 导致光谱曲线反射率波动较大, 光谱平滑可以在一定程度上消除噪声的影响 (Debba 等, 2005)。本研究采用卷积平滑滤波法平滑 (Savitzky 和 Golay, 1964) 实现光谱曲线的平滑, 多项式阶数为 3, 窗口长度为 5。

2.4 特征提取

为减弱图像的背景信号、大气散射的影响, 并且有效地提升不同吸收特征的对比度, 对原始光谱曲线平滑后采用一阶微分、二阶微分、倒数对数变换这 3 种变化方式对光谱反射率进行变换。光谱反射率一阶微分及二阶微分计算公式如下 (Tsai 和 Philpot, 1998):

$$R'(\lambda_i) = \frac{(R(\lambda_{i+1}) - R(\lambda_{i-1})))}{2\Delta\lambda}$$

(1)

$$R''(\lambda_i) = \frac{(R'(\lambda_{i+1}) - R'(\lambda_{i-1})))}{2\Delta\lambda}$$

(2)

式中, $R'(\lambda_i)$ 为波段 i 的一阶微分值, $R''(\lambda_i)$ 为波段 i 的二阶微分值, λ_i 为波段 i 的波长值、 $R(\lambda_i)$ 为第 i 波段的反射率、 $\Delta\lambda$ 为波长 λ_{i+1} 与波长 λ_i 的差值。

基于不同波段提取的植被指数, 可以在一定程度上减小土壤背景对光谱曲线的影响, 提高植被生长状况的观测精度 (李粉玲 等, 2015)。本研究计算 5 种常用的植被指数, 计算公式见表 4。由于高光谱影像光谱分辨率较高, 一种植被指数存在多种波段组合方式, 故通过遍历所有可能的波段计算植被指数, 选择与 SOC 皮尔森相关系数最高的植被指数所对应的波段, 作为计算最终植被指数值的波段。本研究蓝、绿、红、近红外波段的光谱选取范围分别为: 420—490、500—590、630—720、720—867, 分别对应 ZY-1F 高光谱影像的第 1—3、5—14、20—29 和 30—47 波段。

表 4 植被指数计算公式
Table 4 Vegetation indices and their calculation formula

指数名称	计算公式	表征含义
土壤调整植被指数 SAVI (Huete, 1998)	$(1 + 0.5) \times (R_{nir} + R_{red}) / (R_{nir} + R_{red} + 0.5)$	评估土壤暴露程度
差值植被指数 DVI (Jordan, 1969)	$R_{nir} - R_{red}$	近红外与红光波段的反射率之差, 评估植被的生长和健康状况
归一化植被指数 NDVI (Tucke, 1979)	$(R_{nir} - R_{red}) / (R_{nir} + R_{red})$	近红外与红光波段的反射率之比, 评估地表植被生长状况
归一化绿度植被指数 GNDVI (Gitelson 等, 1996)	$(R_{nir} - R_{green}) / (R_{nir} + R_{green})$	近红外与绿光波段的反射率之比, 评估植被覆盖和生长状况
比值植被指数 RVI (Blackburn, 1998)	R_{red} / R_{green}	近红外与红光波段的反射率比值, 反应土壤与植被背景之间的辐射差

注: R_{blue} , R_{green} , R_{red} , R_{nir} 分别代表蓝 (ZY-1F 的光谱范围: 蓝 (420—490 nm)、绿 (500—590 nm)、红 (630—720 nm) 和近红外 (720—867 nm) 波段的光谱反射率。

地形因子作为成土过程中的一个重要自然地理因素, 通过侵蚀和堆积影响 SOC 的分布 (孙文义 等, 2010)。地形因子也会通过影响水热资源的分配间接影响森林生长状况, 从而影响森林 SOC 分布。数字高程模型 DEM (Digital Elevation Model) 数据是来自 NASA 的 EARTHDATA 网站 (<https://search.earthdata.nasa.gov/> [2023-10-26])。基于 DEM 数据应用地理信息系统软件 (ArcMap 10.4) 计算得到坡度、坡向、粗糙度等地形因子共同作为 SOC 含量的预测变量 (姜艳敏 等, 2022)。地形因子的详细信息见表 5。

由于森林 SOC 与土壤水分和森林参数关系密

切, 本研究使用已有的土壤水分和森林地上生物量 AGB (AboveGround Biomass) 产品作为预测因子。基于本团队之前的研究 (Du 等, 2021), 应用帽儿山实验林场 AGB 估测结果参与 SOC 反演, 空间分辨率为 30 m。土壤水分数据为中国农业科学院农业资源与农业区划研究所研发的中国土壤水分数据集 (Meng 等, 2021), 包含 2002 年—2018 年中国陆地土壤水分信息, 单位为 m^3/m^3 , 时间分辨率为月, 空间分辨率为 0.05° , 将其重采样为 30 m 图像, 与 AGB 产品的像元大小保持一致。从该数据集中提取与实测数据较为接近的 2016 年—2018 年 3 个年份生长季 (6—8 月) 的土壤水分值,

基于皮尔逊相关系数分别计算土壤水分与SOC含量的相关性，综合选取相关性最显著的2016年7月的土壤水分值作为土壤水分变量参与建模。

表5 地形因子及计算公式
Table 5 Terrain factors and their formula

地形因子	描述	计算方法
SOS	坡度变率	地表面高程相对于水平面变化的二阶微分
SOA	坡向变率	$SOA = ((SOA1 + SOA2) - SOA1 - SOA2) / 2$ SOA:地面某点坡向变率; SOA1:正地形坡向变率; SOA2:反地形坡向变率
S	坡度	$Slo = \arctan \sqrt{f_x^2 + f_y^2} \times 180/\pi$ Slo:地面某点的坡度; f_x :该点高程x轴上的变化率; f_y :该点高程y轴上的变化率
A	坡向	$A = Swe/Ssn$ A:地面某点坡向; Swe:该点在x轴坡度; Ssn:该点在y轴坡度
H	高程	DEM数据
EV	高程变异系数	$EV = STD / Mean$ EV:地面某点高程变异系数; STD:窗口内高程标准差; Mean:窗口内平均高程
Rough	地表粗糙度	$R = 1/\cos(Slo) \times \pi / 180$ R:地面某点地表粗糙度; Slo:地面某点坡度
SCD	地表切割深度	$SCD = Mean - Min$ D:地面某点地表切割深度; Mean:窗口内平均高程; Min:窗口内最低高程
RDLS	地形起伏度	$RDLS = Max - Min$ RDLS:该栅格单元的起伏度; Max:窗口内最大高程; Min:分析窗口内最小高程值

2.5 模型建立及验证

由于本研究使用的高光谱数据特征维度高，样本数量少，是高维小样本数据，易引发“维度灾难”和“过拟合”，造成模型精度随特征数量增加而降低的状况，所以特征选择十分重要（刘笑笑，2016）。递归特征消除 RFE（Recursive Feature Elimination）是一种贪心算法，其主要思想是不断构建模型，最终选出最佳特征集合（吴辰文等，2017）。为保证最佳特征数，首先设定特征个数为1—100，步长为10进行粗筛选，选择 R^2 最高的特征数；然后，以其前后特征个数为5，步长为1进行精细筛选，最终确定最佳特征个数，以达到模型局部最优的结果。

应用RFE方法对所有建模方法筛选最优特征子集，并进行模型构建，包括极端梯度提升（XGBoost）、随机森林（RF）和支持向量机回归（SVR）3种机器学习算法和应用最小二乘回归（OLS）算法拟合的传统多元线性回归。

本研究使用网格搜索法优化参数。XGBoost以回归树 CART（Classification And Regression Tree）为基模型，通过梯度提升决策树实现多棵CART树的Boosting集成算法（Vapnik，1999）。基础学习器的数量（num_boost_round）设置为200—1000的9个数，最大深度（max_depth）设置为1—20均匀

分布的20个值。RF以决策树为基模型，利用Bootstrap抽样方法从原始训练集中抽取样本，通过逐步回归进行决策树建模，最终将多棵决策树组成随机森林的算法（Breiman，2001），进行网格搜索时，基础学习器的数量（num_boost_round）和最大深度（max_depth）的取值范围与XGBoost模型相同。SVR的主要思想是通过非线性变换将输入变量转换到高维空间；在高维空间中寻求最优的分类面（Smola和Schölkopf，2004）。本研究选用高斯径向基函数为核函数，需要对epsilon、gamma和cost参数进行优化，epsilon用于调整决策边界的宽度，设置为0.1、0.01、1；gamma是定义高斯函数宽度的参数，设为以对数均匀分布在 10^{-6} —10的15个值；cost则是每个支持向量的影响，设为以对数均匀分布在 10^{-3} —100的11个值。

由于本研究的土壤样本量较小，故使用留一交叉验证 LOO-CV（Leave-One-Out Cross-Validation），对不同模型进行模型评价。模型精度使用决定系数（ R^2 ）、均方根误差 RMSE（Root Mean Square Error）、相对均方误差 rRMSE（relative Root Mean Square Error）、平均绝对误差 MAE（Mean Absolute Error）和偏差（bias，b）4个评价指标进行评价。计算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \tag{5}$$

$$rRMSE = \frac{RMSE}{\bar{y}} \tag{6}$$

$$b = \frac{\sum_{i=1}^n |y_i - \bar{y}|}{n} \tag{7}$$

式中， $\hat{y}_i$ 为 SOC 预测值， y_i 为 SOC 实测值， $\bar{y}$ 为 SOC 实测值平均值， n 为验证样本个数。

3 结果与分析

3.1 最优特征子集选择

应用 RFE 算法得到反演天然次生林土壤各层 SOC 的最优特征子集如表 6 所示。可见：大部分最

优特征子集中均包括土壤水分特征（water），证明土壤水分对森林 SOC 的反演十分重要；不同模型所选择的特征子集具有明显差异，XGBoost 和 RF 模型最优特征子集主要为反射率微分变换（如 DD）和地形因子（如坡度和坡向），说明 XGBoost 和 RF 模型可以更好利用波段反射率微分变换和地形因子间的信息（Taghizadeh-Mehrjardi 等，2016）；XGBoost 模型最优特征子集特征大部分为红光和近红外波段反射率的微分（DD713、DD816、D713），这是由于红光和近红外波段反射率的变化更能反应植物生长状况的细微差异。其中，713 nm 波段位于可见光和近红外光谱的交界处，为 ZY-1F 的红边波段，通常被认为是植被反射特征的重要指标，红边波段的波长较长，能够更有效地绕过障碍物进行传播，与绿光相比穿透能力更强。不同土壤分层的最优特征子集也有明显的不同，例如，绿光波段反射率微分（如 D584）在 0—5 cm 土壤层建模时作用显著，而在深层土壤中则作用较小，同时地形因子（如高程和坡向等）对表层 SOC 也具有一定作用。

表 6 最优特征子集
Table 6 Optimal subset of features

土壤深度/cm	模型	最优特征子集	特征个数
0—5	XGBoost	A, water, D661, DD541, SCD, DD816, DD532, DD841, D592, EV, D824, D584, DD756, DD850, DD747, DD618, DD704	17
	RF	DD816 nm, DD859 nm, D584 nm, EV, SCD	5
	SVR	LR532 nm, LR541 nm, LR841 nm, LR850 nm, RVI(B644, B738)	5
	OLS	B575, B644, B670, B807, B824, D18, D21, D37	11
5—15	XGBoost	DD816, water, DD541, H, S, DD809, DD713, DD859, AGB	9
	RF	DD859, AGB, S, DD816, H, D618, D661	7
	SVR	B859, LR524, water, B841, RVI(B644, B747), GNDVI(B507, B841), LR541, B876	8
	OLS	B653, B687, B816, D14, D16, D22, D27, D31, DD13, DD23, DD30,	11
15—30	XGBoost	DD575, water, DD841, DD661, D549, DD781, DD524, DD854, DD764, D713	10
	RF	DD661 nm, DD841 nm	2
	SVR	LR627, LR418, water, LR618, LR704, RVI(661, 773)	6
	OLS	B824, B867, D20, D44, DD21, DD30, DD44	7
0—30	XGBoost	DD618, water, DD816, H, DD541, DD713, SCD, DD859, A, DD841, DD704, DD592, D584, DD584, D678	15
	RF	DD816, DD859, H, SOS, water	5
	SVR	B867, water, B876, B850, SAVI(B678, B721), LR736	6
	OLS	B695, B713, B773, D20, D32, DD25, DD43	7

注：B 为原始波段；D 为一阶微分；DD 为二阶微分；LR 为倒数对数。括号中数字为最佳植被指数波段组合。AGB；地上生物量；A；坡向；EV：高程变异系数；H；高程；GNDVI：归一化绿度植被指数；RVI：比植被指数；SAVI：土壤调整植被指数；SCD：地表切割深度；S；坡度；SOS：坡度变率；water：土壤水分。

3.2 不同模型的森林SOC估测精度检验

应用4种模型(XGBoost、RF、SVR、OLS)对不同土壤深度SOC进行估测,精度如表7所示。可见:对于不同的土壤深度,4种模型的拟合效果基本相同,均为XGBoost>RF>SVR>OLS。从土壤深度来看,虽然应用XGBoost估测0—30 cm SOC的模型 R^2 与估测表层土壤(0—5 cm) SOC的 R^2 相同(均为0.54),但估测0—30 cm SOC的模型rRMSE、MAE和*b*均<0—5 cm;不同土层SOC的模

型拟合 R^2 、RMSE、rRMSE、MAE和*b*均呈现出随深度增加而逐渐减小的趋势,这种变化趋势与土壤深度的值域变化有关。根据表1 SOC实测值描述性统计可知,SOC实测值标准差和变异系数随土壤深度的增加而逐渐减小,说明SOC含量值域变化减小,稳定性增大,因此,随土壤深度增加估测模型的RMSE、rRMSE、MAE和*b*更小,但估测精度 R^2 并没有明显增加。

表7 基于4种模型的不同深度SOC估测精度
Table 7 Different depth SOC estimation accuracy based on four models

土壤深度/cm	模型	R^2	RMSE/(g/kg)	rRMSE	MAE/(g/kg)	<i>b</i> /(g/kg)
0—5	XGBoost	0.54	44.25	0.36	35.98	39.15
	RF	0.22	59.99	0.49	50.48	46.30
	SVR	0.17	69.73	0.57	57.25	49.62
	OLS	0.15	78.74	0.64	59.57	50.33
5—15	XGBoost	0.46	15.72	0.27	12.54	14.21
	RF	0.19	19.73	0.33	16.28	16.72
	SVR	0.15	21.08	0.36	14.90	18.00
	OLS	0.11	23.47	0.37	15.52	18.71
15—30	XGBoost	0.30	12.56	0.26	9.67	10.71
	RF	0.21	13.31	0.28	10.69	11.63
	SVR	0.18	16.79	0.36	11.10	12.11
	OLS	0.12	17.83	0.39	12.50	12.87
0—30	XGBoost	0.54	21.28	0.28	17.29	19.24
	RF	0.25	26.70	0.35	22.15	21.95
	SVR	0.13	29.67	0.39	22.93	25.85
	OLS	0.11	30.22	0.41	23.21	26.21

图2为应用最优模型(XGBoost)估测不同深度SOC含量与实测值的拟合曲线。总体来看,不同深度的SOC均有“高值低估”和“低值高估”的现象。由图2(a)可见,表层土壤(0—5 cm)的SOC反演精度较好,预测点更接近1:1直线,但由于表层土壤(0—5 cm)的SOC值域较高、变异系数较大(表1),因此其它指标(RMSE等)要远高于其他层的SOC。由图2(b)和(c)可见,5—15 cm和15—30 cm SOC值域变化范围小,散点在低值区域分布集中,且有较明显的“高值低估”现象。这可能是由于本研究不仅应用了光谱波段组合,还应用了植被指数、地上生物量等森林参数间接反演森林SOC,而研究区域较高的森林郁闭度使植被指数及森林生物量出现饱和(Du等,

2021),从而导致了森林SOC含量高值低估的现象。由图2(d)可见,0—30 cm土壤的SOC反演精度较好,与表层土壤(0—5 cm)的 R^2 相当,但RMSE、MAE以及*b*均远远小于表层土壤。

结合皮尔森相关性分析发现(表2),0—5 cm与其他土层相关性随着层深的增加而减弱,表层(0—5 cm)和底层(15—30 cm)土壤的SOC含量相关性较低。因此,虽然应用ZY-1F数据和XGBoost能较好的反演表层土壤的SOC含量,但这个反演方法是基于植被指数和光谱信息的间接反演,随着底层SOC含量和表层SOC含量之间相关性变小,这种方法也对深层SOC预测能力逐渐减弱,因此,对15—30 cm的SOC含量估测效果相对较差。

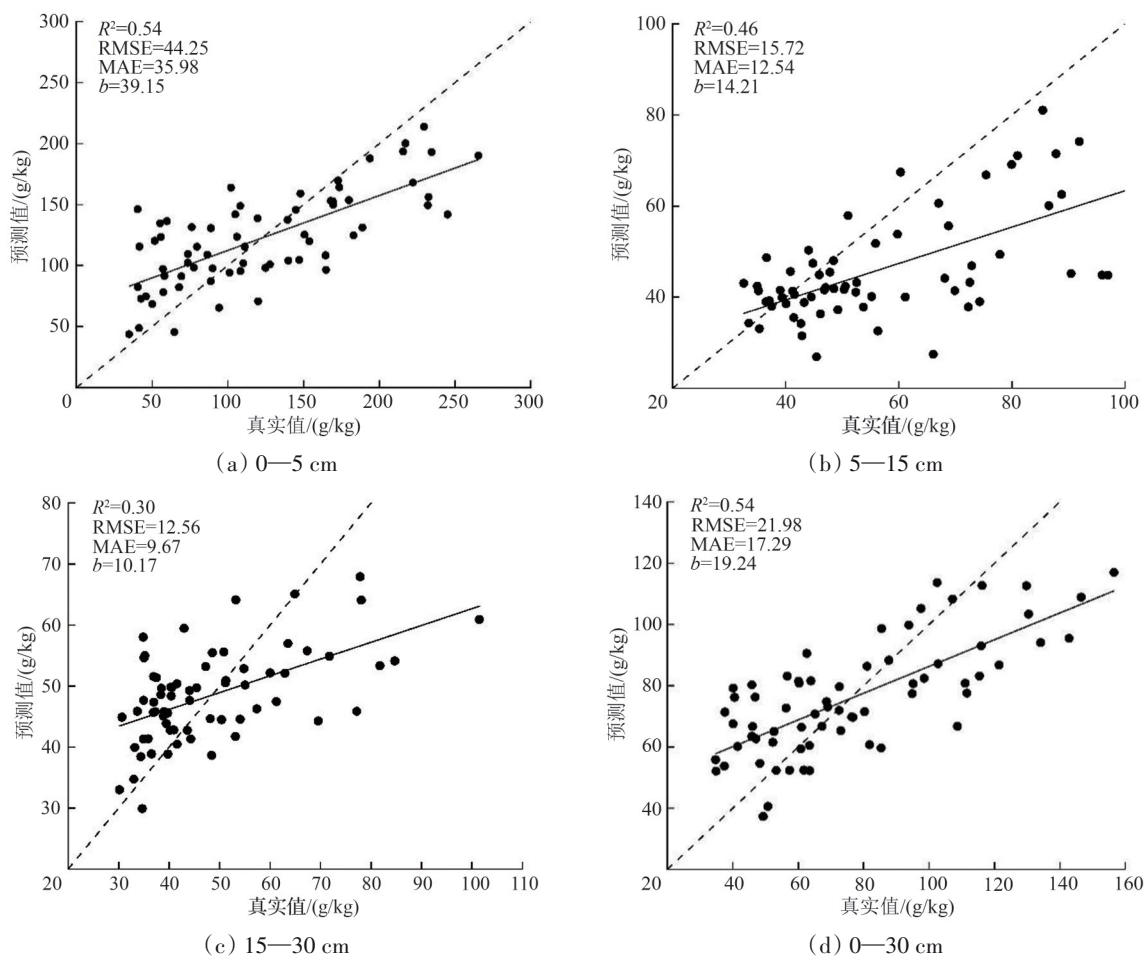


图2 应用XGBoost模型估测的不同土层SOC含量与实测值的拟合曲线

Fig. 2 The fitting curves of measured and predicted SOC content at different depths

3.3 不同土壤分层的森林SOC反演图

应用最佳模型XGBoost反演帽儿山天然次生林SOC如图3所示。由图3(a)可见, 0—5 cm的表层土壤中森林SOC含量主要集中在80—160 g/kg, 研究区中部高, 四周低, 河流附近的SOC含量较少。由图3(b)可见, 5—15 cm土壤中森林SOC含量集中在40—70 g/kg, 其分布与0—5 cm土壤相似, 中部高, 四周低。由图3(c)可见, 15—30 cm土壤中SOC含量主要集中在20—60 g/kg, 研究区分布呈现东部低, 西部高, 河流附近的分布与0—5 cm、5—15 cm相反, 即河流附近SOC含量较高。由图3(d)可见, 0—30 cm的森林SOC含量集中在40—100 g/kg, 研究区中部略低, 四周较高, 与河流分布具有相关性, 河流附近的森林SOC含量相对较高。

不同土壤深度SOC含量预测值描述性统计如表8所示。总体而言, 帽儿山实验林场0—30 cm

的森林SOC预测值取值范围为20.67—118.23 g/kg, 与真实值的取值范围(34.47—156.56 g/kg)相比明显缩小, 虽然预测值和真实值的平均值相差不大(67.20 v.s. 76.84 g/kg), 但中值差异较大(68.68 v.s. 117.85 g/kg), 标准差的差异也较大, 分别为11.88 g/kg和30.92 g/kg, 说明预测值比真实值更集中。0—5 cm、5—15 cm和15—30 cm的森林SOC变化范围分别为27.64—217.74 g/kg、22.53—83.84 g/kg和25.00—70.58 g/kg, 即随着土层加深, SOC含量不断减少, 具有明显的表聚现象(王文静等, 2013); 变异系数分别为21.59%、14.61%和12.11%, 数据的离散程度逐渐减小。这是由于森林表层土壤具有大量的植被凋落物, 同时表层土壤松软更利于透水透气, 进而提高了表层SOC的含量; 随着土层加深, 土壤环境变封闭, 土壤密度等其他性质变化幅度降低, 使深层SOC含量减少且比表层稳定(赵俊勇等, 2016)。

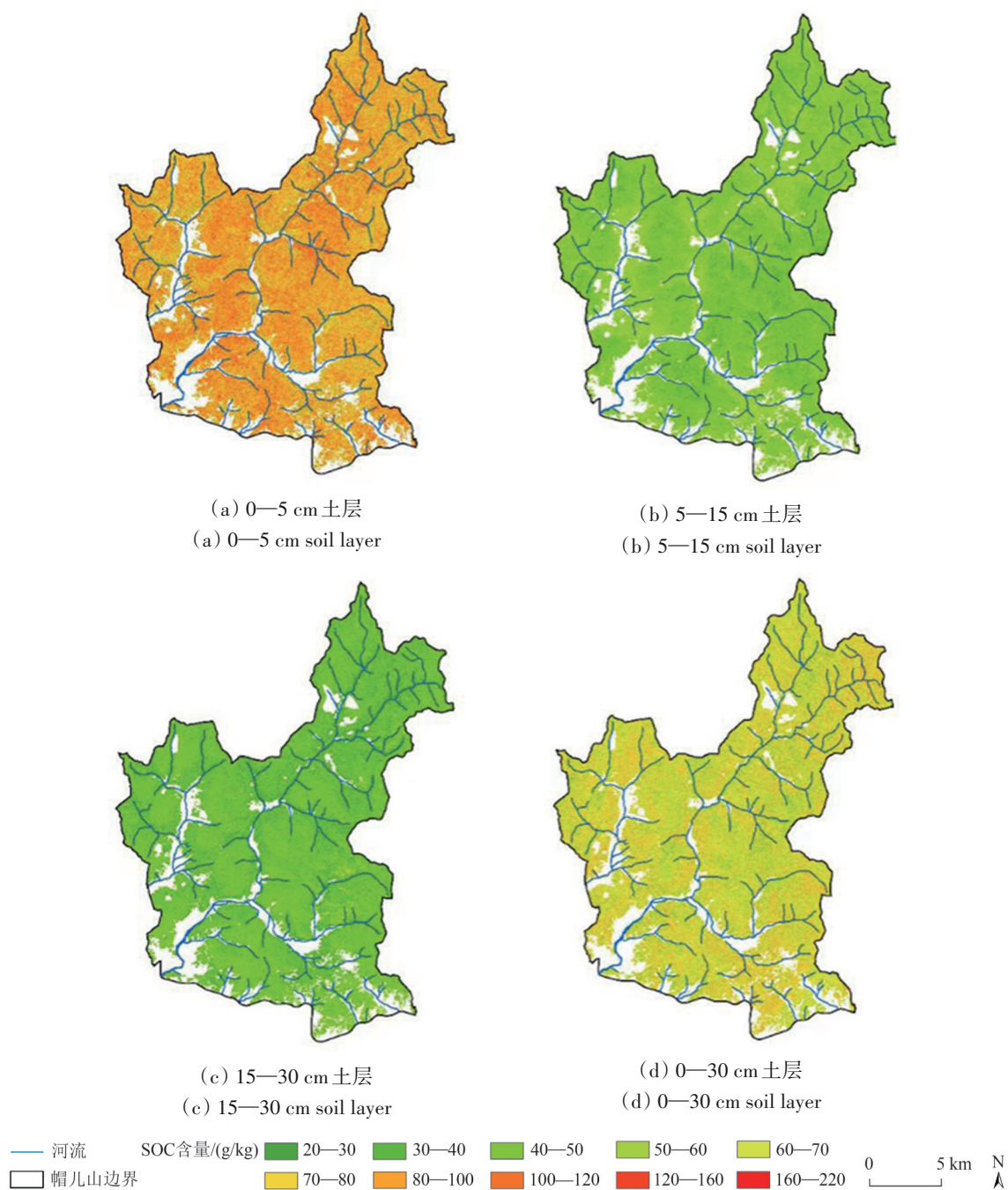


图3 帽儿山实验林场天然次生林不同SOC分布图

Fig. 3 The SOC(g/kg)distribution map of the natural secondary forests in Maorshan experimental forest farm

表8 应用XGBoost模型的帽儿山实验林场森林SOC含量预测结果

Table 8 Descriptive results of the predicted SOC content for Maorshan experimental forest farm using XGBoost						
层深/cm	SOC 含量/(g/kg)					变异系数/%
	最小值	最大值	中值	平均值	标准差	
0—5	27.64(34.83)	217.74(275.80)	122.70(108.43)	88.87(123.23)	19.19(65.19)	21.59(52.91)
5—15	22.53(32.59)	83.84(121.50)	53.18(51.03)	46.92(59.04)	5.86(21.24)	14.61(35.98)
15—30	25.00(30.18)	70.58(101.45)	47.80(42.98)	40.12(48.24)	5.68(15.14)	12.11(31.38)
0—30	20.67(34.47)	118.23(156.56)	68.68(117.85)	67.20(76.84)	11.88(30.92)	17.68(40.24)

注: 括号中为样点测量的真实值。

3.4 不同林分类型SOC含量的比较

由于森林SOC大多来源于植被凋落物，所以林分类型会对SOC的含量产生影响（杜雪和王海燕，2022）。本研究根据帽儿山林相图和SOC反演结果，对比分析了不同林分类型（阔叶林、针叶林和针阔混交林）中SOC含量的差异，结果如表9所示。可见3种林分类型各个层深的SOC含量趋势相似，均表现为阔叶林>针阔混交林>针叶林。

表9 不同林分类型SOC含量对比
Table 9 The statistics of SOC content of different stand types

层深/cm	SOC 含量/(g/kg)		
	阔叶林	针叶林	针阔混交林
0—5	89.53±20.68	87.62±18.89	89.38±19.78
5—15	47.71±5.68	45.65±5.25	47.40±5.51
15—30	42.23±6.48	38.07±5.76	40.16±5.67
0—30	68.87±12.56	65.21±11.77	67.54±12.36

4 讨论

4.1 不同遥感特征对SOC含量影响

为去除噪音、增加波段信息，本研究进行了光谱一阶微分、二阶微分、倒数对数变换。研究发现，光谱一阶微分和二阶微分对SOC估测作用显著，与魏丹萍和郑光辉（2022）、叶勤等（2017）研究结果一致，这是因为光谱的低阶微分对噪声的敏感性较低，因此可以在一定程度上消除基线和背景噪声的影响，提升光谱的灵敏性，并且光谱微分使用相邻两波段计算，反映了相邻波段反射率的变化情况，从某种意义上包含了两个波段的信息，对SOC更敏感（Tian等，2011）。从特征重要性排序来看，红光波段和近红外波段对于SOC估测起到重要作用，这与Gholizade（2018）的实验结果相似，原因可能是红色波段和近红外波段对于某些植被特征，例如生物量特别是植物叶绿素含量十分敏感（Mutanga和Skidmore，2004），叶绿素含量与土壤有机质和SOC密切相关，为植被的生理状况提供了有价值的信息（Odebiri等，2020）。红边波段位于可见光和近红外光谱的交界处，通常被认为是植被反射特征的重要指标，红边波段的波长较长，能够更有效地绕过障碍物进行传播，与绿光相比穿透能力更强。因此，在遥感应用中能够更有效地捕捉植被的生长动态和健康状况（Jiang等，2021）。绿光波段反射率微分

（如D584）在0—5 cm土壤层建模时作用显著，而在深层土壤中则作用较小，这是由于绿光波段位于绿色植被的反射峰附近，不同植被类型冠层的绿光波段反射率产生变化进而对表层SOC含量产生影响（Wang等，2018），

在本研究中也发现地形因子在森林SOC的预测中起到十分重要的作用，这是因为地形因子影响表层土壤的温度、湿度等因素，使表层SOC发生下渗等物理过程，使得地形因子对表层SOC作用显著（郭亮娜等，2023）。田风霞（Tian等，2010）对祁连山东段亚高山灌丛地带SOC含量研究发现：地形因子对SOC的分布产生影响，特别是坡向对于表层SOC具有显著作用，这与本研究结果一致，坡向作为重要的地形因子，通过影响光照强度与光照时间使土壤水热条件产生较大差异，进而间接影响SOC的积累（Zhang等，2015）。

4.2 森林SOC含量分布

预测结果显示，森林SOC 0—5 cm、5—15 cm与15—30 cm含量分布状况相反，河流附近表层SOC含量较少，而深层SOC含量较多，这与谢蕾蕾发现的典型黑土防护林带SOC含量与年降水呈显著负相关的结果类似，可能是由河流附近表层土壤水分较多，土壤水分使SOC下渗导致（谢雷雷等，2024）。预测结果表现出明显的表聚现象，表层SOC含量明显高于深层SOC，这是由于下层土壤植被凋落物和根系分泌物明显减少，使下层土壤养分和水分减少，进而导致SOC含量的减少（陈美玲等，2022）。针对不同林分类型SOC含量表现为阔叶林>针阔混交林>针叶林，这可能是由于阔叶林凋落物中含较多易分解糖类化合物，而针叶林中含有较多木质素等抗腐烂化合物导致凋落物难以分解SOC（杨渺等，2007）；另一方面，阔叶林郁闭度一般高于针叶林，林下环境较为潮湿、阴冷，有利于大团聚体的形成和稳定，并且也有利于微生物分解植被凋落物和土壤的碳化过程（Taghizadeh-Mehrjardi等，2016），而且帽儿山天然次生林的阔叶林主要以白桦、水曲柳等高固碳树种为主，阔叶林的树种组成也是影响阔叶林SOC含量较高的主要原因。

4.3 森林SOC含量反演

与多光谱数据相比，高光谱数据提供更高的光谱分辨率以及连续和窄带图像信息，这有助于

捕获光谱特性的细节 (Bao 等, 2021)。尽管高光谱遥感技术在细节捕获上更为优秀, 但利用高光谱遥感数据并不能直接反演 SOC 含量。因此, 进行森林 SOC 反演时须依靠相应的替代指标 (如植被指数、土壤光谱特征等), 这是目前森林 SOC 反演的主要难点, 限制了反演精度的提高 (Merry 和 Levine, 1995)。这些替代指标的选取对 SOC 反演精度至关重要, 植被指数等指标可以提供与 SOC 含量的间接相关信息, 但 SOC 分布具有复杂性, 这些指标无法充分捕捉其细微变化, 导致反演精度受限, 往往会产生较低的 R^2 值 (dos Santos 等, 2023; Shafizadeh-Moghadam 等, 2022)。

结合高光谱遥感数据与其他遥感数据 (如 SAR 数据、激光雷达等), 可以弥补单一数据源的不足, 以捕捉到更多关于 SOC 含量细微变化的信息, 从而提高森林 SOC 反演精度。例如, 与光学遥感相比, SAR 传感器成像受大气条件的影响要小得多, 这一优势确保了 SAR 图像在森林 SOC 反演中的适用性 (Filgueiras 等, 2019; Woodhouse, 2017)。随着遥感技术的不断进步, 未来会应用更多高空间、高光谱分辨率和主动遥感卫星投入使用, 这将极大地促进森林 SOC 反演技术的发展, 并在林业领域有着广泛的应用前景。

5 结 论

为了对天然次生林的 SOC 进行有效估测以满足天然次生林可持续经营的需求, 本研究以帽儿山天然次生林为研究对象, 利用国产星载高光谱遥感影像 (ZY-1F) 结合机器学习模型, 构建一套应用星载高光谱遥感数据进行森林 SOC 间接估测的流程与方法, 并评估中国国产高光谱卫星 ZY-1F 在反演天然次生林 SOC 方面的潜力。基于 ZY-1F 与 XGBoost 模型可建立有效的 SOC 估测流程, 模型拟合精度最高, 估测的表层 (0—5 cm) 土壤 SOC 含量最高, 随土层加深呈显著递减趋势, 且河流对 SOC 的空间分布具有显著影响。对森林 SOC 估测较敏感的是 ZY-1F 数据的波段微分, 波段间的细微变化与 SOC 含量十分相关。地形因子、土壤水分对 SOC 估测作用较大。同时, 不同林分类型 SOC 含量存在显著差异, 表现为阔叶林 (68.87 g/kg) > 针阔混交林 (67.54 g/kg) > 针叶林 (65.21 g/kg), 且该排序在各土层均保持一致。综上, 本研究验证了基于国产星载高光谱数据 (ZY-1F) 和 XGBoost

算法间接反演森林 SOC 的可行性, 有利于实现长时间序列的森林 SOC 连续预测。未来的研究中, 还需要在不同年份、不同季节进行测试和验证, 以实现帽儿山天然次生林 SOC 进行连年观测的目标。

志 谢 感谢国家科技资源共享服务平台—国家地球系统科学数据中心—土壤分中心 (<http://soil.geodata.cn>) 提供参考数据支撑。

参考文献 (References)

- Bao Y L, Ustin S, Meng X T, Zhang X L, Guan H X, Qi B S and Liu H J. 2021. A regional-scale hyperspectral prediction model of soil organic carbon considering geomorphic features. *Geoderma*, 403: 115263 [DOI: 10.1016/j.geoderma.2021.115263]
- Bhunja G S, Kumar Shit P and Pourghasemi H R. 2019. Soil organic carbon mapping using remote sensing techniques and multivariate regression model. *Geocarto International*, 34(2): 215-226 [DOI: 10.1080/10106049.2017.1381179]
- Blackburn G A. 1998. Quantifying chlorophylls and carotenoids at leaf and canopy scales: an evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, 66(3): 273-285 [DOI: 10.1016/s0034-4257(98)00059-5]
- Breiman L. 2001. Random forests. *Machine Learning*, 45(1): 5-32 [DOI: 10.1023/A:1010933404324]
- Broderick D E, Frey K E, Rogan J, Alexander H D and Zimov N S. 2015. Estimating upper soil horizon carbon stocks in a permafrost watershed of Northeast Siberia by integrating field measurements with Landsat-5 TM and WorldView-2 satellite data. *GIScience and Remote Sensing*, 52(2): 131-157 [DOI: 10.1080/15481603.2015.1010434]
- Chen B, Lu Q K, Wei L F, Fu W Q, Wei Z Y and Tian S. 2024. Global predictions of topsoil organic carbon stocks under changing climate in the 21st century. *Science of the Total Environment*, 908: 168448 [DOI: 10.1016/j.scitotenv.2023.168448]
- Chen M L, Zhao Y P, Zhang J C, Wang Y H, Meng, M J, Liu X, Li C and Xie D J. 2022. Soil organic carbon characteristics of different forest types in Fengyang mountain. *Journal of Northeast Forestry University*, 50(10): 69-75 (陈美玲, 赵友朋, 张金池, 王宇浩, 孟苗婧, 刘鑫, 李翀, 谢德晋. 2022. 凤阳山典型林分土壤有机碳特征. *东北林业大学学报*, 50(10): 69-75) [DOI: 10.13759/j.cnki.dlxb.2022.10.014]
- Conforti M, Lucà F, Scarciglia F, Matteucci G and Buttafuoco G. 2016. Soil carbon stock in relation to soil properties and landscape position in a forest ecosystem of southern Italy (Calabria region). *CATENA*, 144: 23-33 [DOI: 10.1016/j.catena.2016.04.023]

- Debba P, van Ruitenbeek F J A, van der Meer F D, Carranza E J M and Stein A. 2005. Optimal field sampling for targeting minerals using hyperspectral data. *Remote Sensing of Environment*, 99(4): 373-386 [DOI: 10.1016/J.RSE.2005.05.005]
- dos Santos E P, Moreira M C, Fernandes-Filho E I, Demattê J A M, Dionizio E A, da Silva D D, Cruz R R P, Moura-Bueno J M, dos Santos U J and Costa M H. 2023. Sentinel-1 imagery used for estimation of soil organic carbon by dual-polarization SAR vegetation indices. *Remote Sensing*, 15(23): 5464 [DOI: 10.3390/rs15235464]
- Du C Y, Fan W Y, Ma Y, Jin H I and Zhen Z. 2021. The effect of synergistic approaches of features and ensemble learning algorithms on aboveground biomass estimation of natural secondary forests based on ALS and Landsat 8. *Sensors*, 21(17): 5974 [DOI: 10.3390/s21175974]
- Du X and Wang H Y. 2022. Active components of forest soil organic carbon and its influencing factors in China. *World Forestry Research*, 35(1): 76-81 (杜雪, 王海燕. 2022. 中国森林土壤有机碳活性组分及其影响因素. *世界林业研究*, 35(1): 76-81) [DOI: 10.13348/j.cnki.sjlyyj.2021.0068.y]
- Filgueiras R, Mantovani E C, Althoff D, Fernandes Filho E I and da Cunha F F. 2019. Crop NDVI monitoring based on sentinel 1. *Remote Sensing*, 11(12): 1441 [DOI: 10.3390/rs11121441]
- Gholizadeh A, Žižala D, Saberioon M and Borůvka L. 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, 218: 89-103 [DOI: 10.1016/j.rse.2018.09.015]
- Gitelson A A, Kaufman Y J and Merzlyak M N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3): 289-298 [DOI: 10.1016/S0034-4257(96)00072-7]
- Guo L N, Li J R, Zhang B, Zhu S J and Fu F W. 2023. Influencing factors of forest soil organic carbon and its research progress. *Journal of Hunan Ecological Science*, 10(3): 85-91 (郭亮娜, 李江荣, 张波, 朱思洁, 付芳伟. 2023. 森林土壤有机碳的影响因子及其研究进展. *湖南生态科学学报*, 10(3): 85-91) [DOI: 10.3969/j.issn.2095-7300.2023.03.011]
- Huete A R. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3): 295-309 [DOI: 10.1016/0034-4257(88)90106-X]
- Jiang F G, Sun H, Li C J, Ma K S, Chen S, Long J P and Ren L X. 2021. Retrieving the forest aboveground biomass by combining the red edge bands of Sentinel-2 and GF-6. *Acta Ecologica Sinica*, 41(20): 8222-8236 (蒋馥根, 孙华, 李成杰, 马开森, 陈松, 龙江平, 任蓝翔. 2021. 联合 GF-6 和 Sentinel-2 红边波段的森林地上生物量反演. *生态学报*, 41(20): 8222-8236) [DOI: 10.5846/stxb202012173204]
- Jiang H, Gao F and Cui X Y. 2015. Soil organic carbon storage and effects of topographical factors of the secondary forest region of Mao, Er mountains. *Forest Engineering*, 31(3): 15-20 (姜航, 高菲, 崔晓阳. 2015. 帽儿山次生林区土壤有机碳储量及地形因子的影响. *森林工程*, 31(3): 15-20) [DOI: 10.3969/j.issn.1001-005X.2015.03.004]
- Jiang Y M, Shi H J, Zhang S W, Shui J F, Cao X P and Zheng C. 2022. A dataset of topographic factors with 12.5 m spatial resolution in the Yellow River Basin. *China Scientific Data*, 7(4): 127-143 (姜艳敏, 史海静, 张少伟, 税军峰, 曹晓萍, 郑诚. 2022. 黄河流域 12.5 米分辨率地形因子数据集. *中国科学数据*, 7(4): 127-143) [DOI: 10.11922/11-6035.ncdc.2021.0017.zh]
- Jordan C F. 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 50(4): 663-666 [DOI: 10.2307/1936256]
- Li F L, Wang L, Liu J and Chang Q R. 2015. Remote sensing estimation of SPAD value for wheat leaf based on GF-1 data. *Transactions of the Chinese Society for Agricultural Machinery*, 46(9): 273-281 (李粉玲, 王力, 刘京, 常庆瑞. 2015. 基于高分一号卫星数据的冬小麦叶片 SPAD 值遥感估算. *农业机械学报*, 46(9): 273-281) [DOI: 10.6041/j.issn.1000-1298.2015.09.040]
- Li J. 2008. Study on Soil Carbon Stock and its Spatial Distribution, Influence Factors in the Forest of China. Yaan: Sichuan Agricultural University (李江. 2008. 中国主要森林群落林下土壤有机碳储量格局及其影响因子研究. 雅安: 四川农业大学)
- Liu X X. 2016. Study on the Remote Sensing Feature Selection Method for Forest Biomass Estimation Based on RF-RFE. Taian: Shandong Agricultural University (刘笑笑. 2016. 基于 RF-RFE 算法的森林生物量遥感特征选择方法研究. 泰安: 山东农业大学)
- Luo M, Guo L, Zhang H T, Wang S Q and Liang P. 2020. Characterization of spatial distribution of soil organic carbon in China based on environmental variables. *Acta Pedologica Sinica*, 57(1): 48-59 (罗梅, 郭龙, 张海涛, 汪善勤, 梁攀. 2020. 基于环境变量的中国土壤有机碳空间分布特征. *土壤学报*, 57(1): 48-59) [DOI: 10.11766/trxb201812110454]
- Ma Z L and Zhao W Q. 2020. Research progress on input of plant community-derived soil organic carbon and its responses to climate warming. *Chinese Journal of Ecology*, 39(1): 270-281 (马志良, 赵文强. 2020. 植物群落向土壤有机碳输入及其对气候变暖的响应研究进展. *生态学杂志*, 39(1): 270-281) [DOI: 10.13292/j.1000-4890.202001.009]
- MacDicken K G. 2015. Global forest resources assessment 2015: what, why and how?. *Forest Ecology and Management*, 352: 3-8 [DOI: 10.1016/j.foreco.2015.02.006]
- Meng X J, Mao K B, Meng F, Shi J C, Zeng J Y, Shen X Y, Cui Y K, Jiang L M and Guo Z H. 2021. A fine-resolution soil moisture dataset for China in 2002—2018. *Earth System Science Data*, 13(7): 3239-3261 [DOI: 10.5194/essd-13-3239-2021]
- Merry C J and Levine E R. 1995. Methods to assess soil carbon using remote sensing techniques//Lal R, Kimble J, Levine E and Stewart B A, eds. *Assessment Methods for Soil Carbon*. Soils and

- Global Change. Boca Raton: CRC Press: 265-374
- Mutanga O and Skidmore A K. 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19): 3999-4014 [DOI: 10.1080/01431160310001654923]
- Odebiri O, Mutanga O, Odindi J, Peerbhay K and Dovey S. 2020. Predicting soil organic carbon stocks under commercial forest plantations in KwaZulu-Natal province, South Africa using remotely sensed data. *GIScience and Remote Sensing*, 57(4): 450-463 [DOI: 10.1080/15481603.2020.1731108]
- Sahoo U K, Singh S L, Gogoi A, Kenye A and Sahoo S S. 2019. Active and passive soil organic carbon pools as affected by different land use types in Mizoram, Northeast India. *PLoS One*, 14(7): e0219969 [DOI: 10.1371/journal.pone.0219969]
- Savitzky A and Golay M J E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8): 1627-1639 [DOI: 10.1021/ac60214a047]
- Shafizadeh-Moghadam H, Minaei F, Talebi-Khiavi H, Xu T T and Homae M. 2022. Synergetic use of multi-temporal Sentinel-1, Sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. *CATENA*, 212: 106077 [DOI: 10.1016/j.catena.2022.106077]
- Smola A J and Schölkopf B. 2004. A tutorial on support vector regression. *Statistics and Computing*, 14(3): 199-222 [DOI: 10.1023/B:STCO.0000035301.49549.88]
- Stevens A, Nocita M, Tóth G, Montanarella L and van Wesemael B. 2013. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLoS One*, 8(6): e66409 [DOI: 10.1371/journal.pone.0066409]
- Suleymanov A, Abakumov E, Nizamutdinov T, Polyakov V, Shevchenko E and Makarova M. 2024. Soil organic carbon stock retrieval from Sentinel-2A using a hybrid approach. *Environmental Monitoring and Assessment*, 196(1): 23 [DOI: 10.1007/s10661-023-12172-y]
- Sun W Y, Guo S L and Song X Y. 2010. Effects of topographies and land use on spatial distribution of surface soil organic carbon in hilly region of Loess Plateau. *Journal of Natural Resources*, 25(3): 443-453 (孙文义, 郭胜利, 宋小燕. 2010. 地形和土地利用对黄土丘陵沟壑区表层土壤有机碳空间分布影响. *自然资源学报*, 25(3): 443-453) [DOI: 10.11849/zrzyxb.2010.03.010]
- Taghizadeh-Mehrjardi R, Nabiollahi K and Kerry R. 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma*, 266: 98-110 [DOI: 10.1016/J.GEODERMA.2015.12.003]
- Tian F X, Zhao C Y and Wang Y. 2010. Spatial and temporal variations of soil moisture and its relationship with environmental factors in the east section of the Qilian Mountains. *Agricultural Research in the Arid Areas*, 28(6): 23-29 (田凤霞, 赵传燕, 王瑶. 2010. 祁连山东段土壤水分时空分布特征及其与环境因子的关系. *干旱地区农业研究*, 28(6): 23-29)
- Tian Y C, Yao X, Yang J, Cao W X and Zhu Y. 2011. Extracting red edge position parameters from ground- and space-based hyperspectral data for estimation of canopy leaf nitrogen concentration in rice. *Plant Production Science*, 14(3): 270-281 [DOI: 10.1626/pps.14.270]
- Tsai F and Philpot W. 1998. Derivative analysis of hyperspectral data. *Remote Sensing of Environment*, 66(1): 41-51 [DOI: 10.1016/S0034-4257(98)00032-7]
- Tucker C J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2): 127-150 [DOI: 10.1016/0034-4257(79)90013-0]
- Vane G and Goetz A F H. 1993. Terrestrial imaging spectrometry: current status, future trends. *Remote Sensing of Environment*, 44(2/3): 117-126 [DOI: 10.1016/0034-4257(93)90011-L]
- Vapnik V N. 1999. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5): 988-999 [DOI: 10.1109/72.788640]
- Wan Q Z, Zhu G F, Guo H W, Zhang Y, Pan H X, Yong L L and Ma H Y. 2019. Influence of vegetation coverage and climate environment on soil organic carbon in the Qilian Mountains. *Scientific Reports*, 9(1): 17623 [DOI: 10.1038/s41598-019-53837-4]
- Wang F Q. 2022. Application of hyperspectral technique in monitoring vegetation characteristics. *South Agricultural Machinery*, 53(18): 76-78 (王富强. 2022. 高光谱技术在植被特征监测中的应用研究. *南方农机*, 53(18): 76-78) [DOI: 10.3969/j.issn.1672-3872.2022.18.021]
- Wang J P, Wu X D, Ma D J, Wen J G and Xiao Q. 2023. Remote sensing retrieval based on machine learning algorithm: uncertainty analysis. *National Remote Sensing Bulletin*, 27(3): 790-801 (汪静平, 吴小丹, 马杜娟, 闻建光, 肖青. 2023. 基于机器学习的遥感反演: 不确定性因素分析. *遥感学报*, 27(3): 790-801) [DOI: 10.11834/jrs.20221172]
- Wang Q. 2021. Progress of environmental remote sensing monitoring technology in China and some related frontier issues. *National Remote Sensing Bulletin*, 25(1): 25-36 (王桥. 2021. 中国环境遥感监测技术进展及若干前沿问题. *遥感学报*, 25(1): 25-36) [DOI: 10.11834/jrs.20210572]
- Wang S, Zhuang Q L, Jia S H, Jin X X and Wang Q B. 2018. Spatial variations of soil organic carbon stocks in a coastal hilly area of China. *Geoderma*, 314: 8-19 [DOI: 10.1016/j.geoderma.2017.10.052]
- Wang W J, Wang B T, Lv Z, Ren B C, Wang N, Nie S and Cao Y B. 2013. Soil organic carbon reserve of different forests in Taiyue Mountain. *Journal of Arid Land Resources and Environment*, 27(1): 81-85 (王文静, 王百田, 吕钊, 仁本超, 王宁, 聂森, 曹远博. 2013. 山西太岳山不同林分土壤有机碳储量研究. *干旱区资源与环境*, 27(1): 81-85) [DOI: 10.13448/j.cnki.jalre.2013.01.007]
- Wei D P and Zheng G H. 2022. Estimation of soil total phosphorus

- content in coastal areas based on hyperspectral reflectance. *Spectroscopy and Spectral Analysis*, 42(2): 517-523 (魏丹萍, 郑光辉. 2022. 高光谱反射率的滨海地区土壤全磷含量反演. *光谱学与光谱分析*, 42(2): 517-523) [DOI: 10.3964/j.issn.1000-0593(2022)02-0517-07]
- Woodhouse I H. 2017. *Introduction to Microwave Remote Sensing*. Boca Raton: CRC Press: 400 [DOI: 10.1201/9781315272573]
- Wu C W, Liang J H, Wang W and Li C S. 2017. Random forest algorithm based on recursive feature elimination. *Statistics and Decision*, (21): 60-63 (吴辰文, 梁靖涵, 王伟, 李长生. 2017. 基于递归特征消除方法的随机森林算法. *统计与决策*, (21): 60-63) [DOI: 10.13546/j.cnki.tjyjc.2017.21.014]
- Xie L L, Pan L, Li W Z, Shi L P, Liu X Y, Wang X Y and Wang X W. 2024. Characteristics and influencing factors of organic carbon distribution in shelterbelts and cultivated soils in typical black soil areas. *Journal of Northeast Forestry University*, 52(7): 71-76 (谢雷雷, 潘理, 李伟珍, 师立鹏, 刘星雨, 王雪誉, 王秀伟. 2024. 典型黑土区防护林带和耕地土壤有机碳分布特征及其影响因素. *东北林业大学学报*, 52(7): 71-76) [DOI: 10.13759/j.cnki.dlxb.2024.07.005]
- Yang M, Li X W, Zhang J and Xue B. 2007. Review of soil organic carbon dynamic progress and the impact factors during herbaceous and woody vegetation types conversation. *Acta Prataculturae Sinica*, 16(4): 126-138 (杨渺, 李贤伟, 张健, 薛波. 2007. 植被覆盖变化过程中土壤有机碳库动态及其影响因素研究进展. *草业学报*, 16(4): 126-138) [DOI: 10.3321/j.issn:1004-5759.2007.04.019]
- Ye Q, Jiang X Q, Li X C and Lin Y. 2017. Comparison on inversion model of soil organic matter content based on hyperspectral data. *Transactions of the Chinese Society for Agricultural Machinery*, 48(3): 164-172 (叶勤, 姜雪芹, 李西灿, 林怡. 2017. 基于高光谱数据的土壤有机质含量反演模型比较. *农业机械学报*, 48(3): 164-172) [DOI: 10.6041/j.issn.1000-1298.2017.03.021]
- You H C. 2012. *Research on the Remote Sensing Inversion and Spatial Distribution of Soil Organic Carbon in Forest*. Fuzhou: Fujian Agriculture and Forestry University (游浩辰. 2012. 林地土壤有机碳遥感反演及空间分异研究. 福州: 福建农林大学)
- Zhang L. 2010. *Qualitative and Quantitative Research of Soil Organic Carbon Transfer to Soil Inorganic Carbon in Desert Grassland*. Beijing: Beijing Forestry University (张林. 2010. 荒漠草原土壤有机碳向土壤无机碳酸盐转移的定性定量研究. 北京: 北京林业大学)
- Zhang S P, Qiao J, Sun X Y, Wang B P, Wang W W, Cui L J and Yu X. 2015. Effects of slope aspect and slope position on spatial distribution of soil nutrients of *Paulownia fortunei* plantation. *Journal of Central South University of Forestry and Technology*, 35(1): 109-116 (张顺平, 乔杰, 孙向阳, 王保平, 王伟炜, 崔令军, 于鑫. 2015. 坡向、坡位对泡桐人工林土壤养分空间分布的影响. *中南林业科技大学学报*, 35(1): 109-116) [DOI: 10.14067/j.cnki.1673-923x.2015.01.020]
- Zhao J Y, Sun X Y, Li S Y, Liu Y, Zhang J D and Fan J G. 2016. Soil organic carbon storage and vertical distribution of carbon and nitrogen under different forest types in the Laotudingzi mountain of Liaoning. *Journal of Northeast Forestry University*, 44(10): 65-68, 78 (赵俊勇, 孙向阳, 李素艳, 刘艳, 张骏达, 范俊岗. 2016. 辽宁省老秃顶子不同林分类型土壤有机碳储量和碳氮垂直分布特征. *东北林业大学学报*, 44(10): 65-68, 78) [DOI: 10.3969/j.issn.1000-5382.2016.10.013]

Inversion of soil organic carbon in natural secondary forest based on ZY-1F images

ZHEN Zhen¹, DING Jianye¹, ZHAO Yang¹, ZHAO Yinghui¹, WEI Qingbin^{1,2,3}

1. Key Laboratory of Sustainable Forest Ecosystem Management, College of Forestry, Northeast Forestry University, Harbin 150040, China;

2. Heilongjiang Academy of Forestry, Harbin 150081, China;

3. College of Geographical Science, Harbin Normal University, Harbin 150025, China

Abstract: Forest Soil Organic Carbon (SOC) is a critical indicator of forest soil quality, significantly affecting the growth of forest trees and playing an essential role in the sustainable development of forestry. Investigating the potential of utilizing hyperspectral images to accurately determine the SOC in natural secondary forests is imperative. This investigation would aid in providing technical assistance for estimating forest SOC on a long-term and large-scale basis. With a focus on the SOC of natural secondary forests, this study randomly selected a total of 67 samples in the Maoershan Experimental Forest Farm of Northeast Forestry University. Soil samples were collected from three different depths: 0—5 cm, 5—15 cm, and 15—30 cm. The SOC content of each sample was measured, and the mean of the three layers was calculated as the SOC content for 0—30 cm depth. The hyperspectral image of ZY-1F was analyzed to calculate the first-order differential, second-order differential, reciprocal logarithm of the spectral curve, and vegetation indices. The Recursive Feature Elimination

(RFE) method was then employed to screen the features, taking into account the Digital Elevation Model (DEM), soil moisture, and forest Aboveground Biomass (AGB) datasets. Three machine learning models, namely, random forest, Extreme Gradient Boosting (XGBoost), and support vector regression, and ordinary least squares regression were employed to estimate the SOC content, and the best model was chosen to estimate the SOC at various depths. Results showed that XGBoost had the highest accuracy in various soil depths: the R^2 for the soil depths of 0–30 cm, 0–5 cm, 5–15 cm, and 15–30 cm were 0.54, 0.54, 0.46, and 0.30, respectively, and the RMSEs were 21.28, 44.25, 15.72, and 12.56 g/kg, respectively. The average SOC values in the natural secondary forest of Maoershan Forest Farm were estimated to be 67.20, 88.87, 46.92, and 40.12 g/kg for the 0–30 cm, 0–5 cm, 5–15 cm, and 15–30 cm soil layers, respectively. The SOC concentration in the forest declined as the soil depth increased. Variations in SOC content exist across various forest types, and the SOC is ordered in descending order as follows: broad-leaved forest, mixed coniferous and broad-leaved forest, and coniferous forest. The band information from hyperspectral images enables the estimation of the SOC in forests. However, the large number of bands leads to data redundancy, which in turn reduces the accuracy of the model's estimates. The RFE method can be employed to identify the optimal combination of features, thereby reducing the amount of features and enhancing the accuracy of model estimate. The differential characteristics of the 710–850 nm bands in hyperspectral images are extremely beneficial for accurately estimating the SOC in natural secondary forests. Topographic factors exert a more significant influence on the SOC at depths above 15 cm, while soil moisture and AGB have a more pronounced effect on SOC in the 5–15 cm layer compared with other factors. The integration of hyperspectral images with DEM, soil moisture, and AGB data facilitates the accurate estimation of the SOC content of natural secondary forests. This approach offers valuable support for estimating long-term and large-scale SOC of natural secondary forests on the basis of multiperiod hyperspectral images.

Key words: natural secondary forest, SOC, hyperspectral, machine learning model, feature selection

Supported by National Key Research and Development Program of China for Young Scientists (No. 2023YFF1305900); Open Fund Project of Key Laboratory of Sustainable Forest Ecosystem Management, Ministry of Education (No. KFJJ2021YB04)