浙 江 农 林 大 学 学 报, 2026, **43**(X): 1−11 https://zlxb.zafu.edu.cn
Journal of Zhejiang A&F University DOI: 10.11833/j.issn.2095-0756.20260364

引用格式: 徐上梅, 滕晨凯, 张加龙, 等. 基于 UAV-LiDAR 与 Sentinel-2 协同的森林地上碳储量估算及不确定性分析[J]. 浙江农林大学学报, 2026, **43**(X): 1−11. XU Shangmei, TENG Chenkai, ZHANG Jialong, *et al*. Estimation and uncertainty analysis of forest aboveground carbon stock based on UAV-LiDAR and Sentinel-2 collaboration[J]. *Journal of Zhejiang A&F University*, 2026, **43**(X): 1−11.

基于 UAV-LiDAR 与 Sentinel-2 协同的森林地上碳储量估算及不确定性分析

徐上梅[1]，滕晨凯[1]，张加龙[1]，姜新周[1]，钟欣玲[1]，蒋　勇[1]，年静玲[1]，罗利彬[2]

（1. 西南林业大学 林学院 (亚太林学院)，云南 昆明 650224；2. 迪庆藏族自治州林草种苗和国有林场管理工作站 云南 迪庆 674400）

摘要：【目的】分析不同类型遥感特征逐步融合对森林地上碳储量 (aboveground carbon stock，AGC) 估算精度和不确定性的影响，并识别模型预测中的关键特征。**【方法】**以云南省香格里拉市高山松 *Pinus densata* 为研究对象，融合 UAV-LiDAR 与 Sentinel-2 数据，构建由高度特征 (*H*) 逐步扩展至 *H*+密度 (*D*)+结构 (*S*)+强度 (*I*)+光学特征 (*O*) 组合，采用极端梯度提升 (Extreme Gradient Boosting，XGB)、随机森林 (Random Forest，RF) 和极端随机树 (Extra Trees，ET) 模型估算森林地上碳储量，通过嵌套 10 折交叉验证评价模型精度，并结合自助重采样-蒙特卡罗模拟 (Bootstrap-Monte Carlo，Bootstrap-MC) 方法分解及量化不确定性，利用沙普利加性解释方法 (SHAP) 分析不同特征及特征组对模型预测的贡献。**【结果】**①特征融合总体提高了森林地上碳储量估算精度，加入特征 *I* 后各模型精度提升最明显；最终组合 (*H*+*D*+*S*+*I*+*O*) 下极端梯度提升模型的精度最高，决定系数 (R^2)、均方根误差 (RMSE) 和相对均方根误差 (rRMSE) 分别为 0.723±0.197、(6.43±2.56) t·hm^{-2} 和 23.01%±8.75%。②特征融合对总不确定性的影响呈非线性，加入特征 *I* 后，3 种模型的总不确定性均下降，加入特征 *O* 后有所回升。最终组合 (*H*+*D*+*S*+*I*+*O*) 下极端梯度提升、随机森林和极端随机树模型的总不确定性分别为 (3.06±0.56)、(2.57±0.55) 和 (1.92±0.32) t·hm^{-2}，且均以模型不确定性为主。③SHAP 结果表明：强度特征组贡献率最高，达 44.4%。**【结论】**递进式融合 UAV-LiDAR 与 Sentinel-2 特征能够提高高山松林森林地上碳储量估算精度，但总不确定性并非随特征增加而持续下降。强度特征是提高模型精度和降低不确定性的关键特征；总不确定性主要来自训练样本变化引起的模型不确定性。图 4 表 7 参 37

关键词：高山松；碳储量；不确定性分析；机器学习；Bootstrap-MC；SHAP

中图分类号：S718.5　　**文献标志码：**A　　**文章编号：**2095-0756(2026)00-0001-11

Estimation and uncertainty analysis of forest aboveground carbon stock based on UAV-LiDAR and Sentinel-2 collaboration

XU Shangmei[1]，TENG Chenkai[1]，ZHANG Jialong[1]，JIANG Xinzhou[1]，ZHONG Xinling[1]，JIANG Yong[1]，NIAN Jingling[1]，LUO Libin[2]

（1. College of Forestry (Asia-Pacific Forestry College), Southwest Forestry University, Kunming 650224, Yunnan, China; 2. Diqing Tibetan Autonomous Prefecture Forest and Grass Seedling and State-owned Forest Farm Management Station, Diqing 674400, Yunnan, China）

Abstract: [**Objective**] This study aims to examine the effects of stepwise integration of different types of remote sensing features on the accuracy and uncertainty of aboveground carbon stock (AGC) estimation in

收稿日期：2026-07-17；修回日期：2026-10-25

基金项目：国家自然科学基金项目 (32260390)；云南省基础研究计划重点项目 (202501AS070047)；西南林业大学重点学科和重点实验室开放项目 (LXXK-2023Z06)

作者简介：徐上梅 (ORCID: 0009-0007-3998-1685)，从事林业遥感研究。E-mail: xu_sm1223@swfu.edu.cn。通信作者：张加龙 (ORCID: 0000-0002-6969-3656)，教授，博士，从事林业遥感研究。E-mail: jialongzhang@swfu.edu.cn

forests, and identify key features in model predictions. [**Method**] Taking *Pinus densata* forests in Shangri-La City of Yunnan Province as the research object, UAV-LiDAR and Sentinel-2 data were integrated to construct feature combinations which gradually expanded from height (*H*) to *H*+density (*D*)+structure (*S*)+intensity (*I*)+optical (*O*). Models such as Extreme Gradient Boosting (XGB), Random Forest (RF), and Extra Trees (ET) were used to estimate AGC. The accuracy of the model was evaluated through nested 10-fold cross-validation, and uncertainty was decomposed and quantified using Bootstrap-Monte Carlo (Bootstrap MC) simulation. SHAP was used to assess the contribution of different features and feature groups to model prediction. [**Result**] (1) The accuracy of estimating forest AGC generally increased with feature integration, and improved the most significantly after adding feature *I*. XGB under the final combination (*H*+*D*+*S*+*I*+*O*) achieved the highest accuracy, with R^2, RMSE, and rRMSE values of 0.723±0.197, (6.43±2.56) t·hm^{-2}, and 23.01%±8.75%, respectively. (2) The impact of feature integration on total uncertainty was non-linear. The total uncertainty of all three models decreased after adding *I*, but increased slightly after adding *O*. The total uncertainties of XGB, RF, and ET were (3.06±0.56), (2.57±0.55), and (1.92±0.32) t·hm^{-2} under the final combination (*H*+*D*+*S*+*I*+*O*), with model uncertainty being the main factor. (3) The SHAP results showed that the intensity feature group had the highest contribution rate, reaching 44.4%. [**Conclusion**] The stepwise integration of UAV-LiDAR and Sentinel-2 features improves AGC estimation in *P. densata* forests, but total uncertainty does not continuously decrease with increasing features. Intensity features are key to improving model accuracy and reducing uncertainty. Variation in training samples is the main source of total uncertainty. [Ch, 4 fig. 7 tab. 37 ref.]
Key words: *Pinus densata*; carbon stock; uncertainty analysis; machine learning; Bootstrap-MC; SHAP

森林是陆地生态系统的重要碳汇，森林地上碳储量 (aboveground carbon stock，AGC) 能够反映森林生态系统固碳能力，其准确估算对于森林资源管理、碳汇评估及碳达峰和碳中和 (“双碳”) 目标实现具有重要意义[1−2]。传统森林地上碳储量估算主要依赖样地调查和异速生长方程，虽然精度较高，但存在成本高、空间连续性不足和尺度扩展不确定性较大等问题[3−4]。UAV-LiDAR 可获取林分高度、冠层垂直结构和点云空间分布等三维结构信息；Sentinel-2 影像能够反映冠层光谱响应、植被生长状况和水平空间纹理差异。两者具有较强互补性，有助于提高森林生物量或碳储量估算精度[5−6]。

然而，森林生态系统结构复杂、空间异质性强，遥感观测、特征提取和机器学习建模均可能引入误差，导致估算结果存在不确定性[7]。现有研究主要采用抽样方法、模型模拟法、误差传播法、泰勒级数等，对抽样误差、模型参数、残差变异、样地尺度误差及遥感模型误差等开展了不确定性量化[8−10]。但多侧重单一误差来源或总体不确定性，训练样本变化引起的模型不确定性与输入特征误差传播形成的数据不确定性缺少同步量化和比较，不利于明确不同不确定性来源的相对贡献。不同类型特征所表征的信息及其贡献存在差异，冗余或弱相关特征可能增加模型复杂度，并导致估算性能和稳定性发生波动[11−13]。曹军等[14] 基于随机森林联立蒙特卡洛模型 (RF-MC) 分析了不同遥感变量组合对高山松 *Pinus densata* 森林地上碳储量估算精度及不确定性的影响；QIU 等[15] 通过误差方程比较了不同 Sentinel-1/2 特征组合下样地尺度和遥感估算模型的不确定性。上述研究表明，互补特征的加入能够在一定程度上降低总不确定性，但不同类型特征逐步加入后的边际增益及其对模型预测的贡献仍缺少系统分析。

基于此，本研究以云南省香格里拉市高山松林为研究对象，融合 UAV-LiDAR 与 Sentinel-2 数据，构建高度特征 (*H*)、密度特征 (*D*)、结构特征 (*S*)、强度特征 (*I*) 和光学特征 (*O*) 递进组合，采用极端梯度提升 (Extreme Gradient Boosting，XGB)、随机森林 (Random Forest，RF) 和极端随机树 (Extra Trees，ET) 模型开展森林地上碳储量估算，并结合自助重采样-蒙特卡罗模拟 (Bootstrap-MC) 方法和沙普利加性解释方法 (SHAP)，揭示不同类型遥感特征递进融合对森林地上碳储量估算精度和不确定性的影响，分解并量化模型不确定性和数据不确定性，识别关键特征及各特征组贡献，旨在为高山松林地上碳储量估算中的特征组合、模型选择和可靠性评价提供依据。

1　材料与方法

1.1　研究区概况

研究区设在云南省迪庆藏族自治州香格里拉市小中甸镇，27°20′~27°43′N，99°36′~99°59′E，地处青藏高原东南缘，横断山脉腹地，属于典型高海拔山地森林草地交错区。面积约 880.94 km^2，区域森林资源较为丰富，高山松为当地的优势树种。

1.2　数据来源和处理

1.2.1　*样地调查数据*　研究区共布设 63 个 30 m×30 m 样地，样地调查时间为 2025 年 7 月。样地内树种为高山松，调查内容包括胸径 (D_{BH})、树高 (H)、样地中心坐标、4 个方位坐标等。本研究采用香格里拉高山松单木生物量模型计算高山松地上生物量[16]，再乘以含碳系数 0.513 1[17] 得到高山松单木碳储量。将样地内各单木地上碳储量求和，并依据样地面积 (0.09 hm^2) 换算为单位面积地上碳储量 (t·hm^{-2})，作为样地尺度模型的因变量。外业调查数据见表 1。

表 1　外业调查数据

Table 1　Field survey data

指标	平均胸径/cm	平均树高/m	高山松株数/株	森林地上碳储量/(t·hm^{-2})
最大值	21.85	15.36	347.00	64.49
最小值	7.21	4.55	36.00	1.48
平均值	13.72	9.56	169.56	28.25
标准差	3.84	2.45	74.72	14.79

说明：统计对象为 63 个 30 m×30 m 样地 (n=63)。每个样地对应 1 个森林地上碳储量观测值，模型以样地为基本建模单元。

为评价样地的地形代表性，基于 Sentinel-2 特征，采用随机森林二分类模型，获得小中甸高山松预测分布范围，独立验证总体精度为 80.88%，Kappa 系数为 0.618；该范围仅用于地形比较，不参与森林地上碳储量建模与空间预测。高山松预测分布区海拔为 2 848~4 306 m (表 2)，样地海拔范围为 3 176~3 423 m，其中 60 个样地分布于 3 200~3 400 m。样地覆盖 5 个坡度等级和 8 个主要坡向，但对极端海拔、较陡坡地和部分坡向的覆盖仍有限 (表 3、表 4)。

表 2　小中甸高山松分布区与样地海拔特征比较

Table 2　Comparison of elevation characteristics between the *P. densata* distribution area and sample plots in Xiaozhongdian

高山松分布海拔范围/m	样地数量/个	高山松分布海拔范围/m	样地数量/个	高山松分布海拔范围/m	样地数量/个
2 848~2 900	0	3 100~3 200	2	3 400~3 500	1
2 900~3 000	0	3 200~3 300	39	3 500~4 306	0
3 000~3 100	0	3 300~3 400	21	合计	63

1.2.2　*UAV-LiDAR 数据*　UAV-LiDAR 数据于 2025 年 7 月 25—28 日获取。数据采集平台为飞马 V10 复合翼无人机，搭载 DV-LiDAR10 激光雷达传感器，平均点云密度 12.232 点·m^{-2}。原始点云经航带拼接和误差校正后，采用 LiDAR360 软件进行噪声点剔除、地面点分类、数字高程模型 (DEM) 构建和点云高度归一化等处理。

表 3　小中甸高山松分布区与样地坡向分布

Table 3　Ratio of *P. densata* distribution area and sample plots in different aspect in Xiaozhongdian

坡向	高山松分布区比例/%	样地数量/个	样地比例/%
平地	0.02	0	0
北坡	11.84	2	3.17
东北坡	13.91	3	4.76
东坡	12.76	7	11.11
东南坡	10.37	5	7.94
南坡	10.16	7	11.11
西南坡	13.94	13	20.63
西坡	14.52	19	30.16
西北坡	12.49	7	11.11
合计	100	63	100

1.2.3　*Sentinel-2 数据*　Sentinel-2 数据来源于 Copernicus 数据中心，选取 2025 年 1 月 21 日覆盖研究区，云量均低于 0.31% 的 6 景低云量 L2A 影像。利用 SNAP 将各波段采用双线性插值重采样至 30 m，并在 ENVI 5.6 中完成影像镶嵌、裁剪和波段组合。受冬季不宜开展外业调查和夏季云量较高的影响，影像与 7 月样地调查及 UAV-LiDAR 数据相差约 6 个月，根据 L2A 产品场景分类层检查，样地范围内未发现雪、冰像元；高山松为常绿针叶树，年内冠层状态相对稳定，具备样地尺度融合基础，但不同月份的光谱差异仍可能引入一定误差。

1.3 遥感特征变量的提取

本研究从 UAV-LiDAR 点云数据中提取 96 个特征变量，基于 Sentinel-2 影像提取光谱纹理特征，主要包括 7 个单波段反射率特征、6 个植被指数和 8 种纹理特征。为减小 UAV-LiDAR 与 Sentinel-2 空间尺度差异的影响，本研究以 30 m×30 m 样地为统一分析单元，分别计算样地范围内的 UAV-LiDAR 点云特征和提取重采样至 30 m 的 Sentinel-2 特征，按照样地编号匹配，实现样地尺度的特征融合。各类特征变量见表 5。

表 4 小中甸高山松分布区与样地坡度分布

Table 4 Ratio of *P. densata* distribution area and sample plots in different slops in Xiaozhongdian

坡度/(°)	高山松分布区比例/%	样地数量/个	样地比例/%
0~8	18.22	29	46.03
8~15	28.11	15	23.81
15~25	34.24	12	19.05
25~35	15.56	6	9.52
35~90	3.88	1	1.59
合计	100	63	100

表 5 特征变量

Table 5 Feature variables

特征类型	特征名	特征数量	特征说明
结构特征	CC、GF、LAI	3	分别表示冠层覆盖度、林隙分数和叶面积指数，用于表征冠层覆盖、孔隙及叶面积状况
高度特征	*H*统计量、H_p、AIH_p、CRR	41	基于地面归一化点云离地高度提取的统计量、百分位及累积指标，用于描述植被垂直高度分布
密度特征	D0~D9	10	不同高度层内回波点所占比例，用于描述点云的垂直分配
强度特征	*I*统计量、I_p、AII_p	42	基于激光回波强度提取的统计量、百分位及累积指标，用于反映冠层目标的回波强度分布
单波段	B2、B3、B4、B5、B8、B11、B12	7	分别代表蓝光、绿光、红光、红边1、近红外、短波红外1和短波红外2波段；
植被指数	NDVI、EVI、MSAVI、NDMI、CIred-edge、NBR	6	各植被指数计算方式参考文献[15, 18−20]
纹理因子	ME、VA、HO、CO、CR、DI、SM、EN	56	针对7个波段，采用3×3窗口基于灰度共生矩阵分别提取的均值、方差、均一性、对比度、相关性、相异性、角二矩和熵

说明：下标 p 表示相应百分位；*H* 统计量和 *I* 统计量包括均值、中位数、极值、标准差、方差、变异系数、四分位距、偏度和峰度等。

1.4 特征选择及组合

1.4.1 特征选择 为降低特征冗余，将特征选择嵌入外层 10 折、内层 5 折交叉验证。每个内层训练折分别对 *H*、*D*、*S*、*I* 和 *O* 组进行 Boruta 筛选，再采用 Spearman 相关性分析去冗余，其中，*H*、*D*、*S* 和 *I* 分别表示无人机激光雷达点云高度特征、垂直分布密度特征、冠层结构特征和激光回波强度特征，*O* 表示 Sentinel-2 的光学特征。当相关系数绝对值$|\rho|\geqslant 0.95$ 时保留随机森林重要性较高的特征[21−23]，并按该重要性排序使每组最多保留 2 个。特征选择仅使用相应训练子集，内层验证集和外层测试集均不参与。内层调参后，在完整外层训练集上重复该流程；全样本筛选结果仅用于最终模型训练和空间预测。

1.4.2 特征组合 为评估特征融合对模型性能和不确定性的影响，构建 *H*、*H*+*D*、*H*+*D*+*S*、*H*+*D*+*S*+*I* 和 *H*+*D*+*S*+*I*+*O* 共 5 种递进式组合。

1.5 研究方法

1.5.1 建模方法 采用极端梯度提升、随机森林和极端随机树等 3 种树模型进行森林地上碳储量估算。极端梯度提升模型通过梯度提升方式逐步拟合残差，并结合正则化和采样策略提高泛化能力[24]；随机森林模型基于 Bootstrap 抽样和随机特征选择构建多棵决策树，并以平均结果作为输出[25]；极端随机树模型在节点划分中引入随机阈值，以增强基学习器差异性并降低模型方差[26]。

嵌套交叉验证内采用随机搜索，以内层 5 折验证均方根误差 (RMSE) 均值优选参数。极端梯度提升模型随机搜索 20 组，随机森林模型和极端随机树模型各 15 组；最优参数确定后，利用完整外层训练集拟合模型。外层测试集不参与特征筛选、参数优化和模型训练，仅用于折外精度、不确定性和 SHAP 分析。3 种模型的超参数搜索范围见表 6。

表 6　3 种机器学习模型的超参数搜索范围

Table 6　Hyperparameter search ranges of the three machine-learning models

模型	超参数搜索范围
极端梯度提升模型	n_estimators：300、500、800、1 200；max_depth：2、3、4、5；learning_rate：0.01、0.02、0.03、0.05；subsample：0.70、0.80、0.90、1.00；colsample_bytree：0.60、0.70、0.80、0.90、1.00；min_child_weight：1、2、3、5；gamma：0、0.05、0.10、0.20；reg_alpha：0、0.10、0.30、0.50；reg_lambda：0.50、1.00、2.00、5.00
随机森林模型	n_estimators：300、600、900；max_depth：none、4、6、8、10、14；min_samples_split：2、4、6、8；min_samples_leaf：1、2、3、4；max_features：sqrt、0.50、0.75、1.00；bootstrap：True；max_samples：None、0.70、0.85、1.00
极端随机树模型	n_estimators：300、600、900；max_depth：none、4、6、8、10、14；min_samples_split：2、4、6、8；min_samples_leaf：1、2、3、4；max_features：sqrt、0.50、0.75、1.00；bootstrap：false

1.5.2　不确定性量化　采用 Bootstrap-MC 方法分解模型不确定性和数据不确定性：Bootstrap 用于表征训练样本扰动引起的模型不确定性[27]；Monte Carlo 用于模拟输入特征误差传播产生的数据不确定性[28]。

在每折训练集中进行 30 次有放回 Bootstrap 重采样，并分别训练极端梯度提升、随机森林和极端随机树子模型。随后，在原始变量尺度上对输入特征进行 200 次 Monte Carlo 扰动。特征扰动幅度依据其在相应外层训练集中的标准差和预设的相对扰动比例确定，测试样本不参与扰动参数估计。随机扰动服从正态分布，限制在原值±10% 内，并根据变量物理意义进行约束。每个外层测试样本扰动 200 次。光谱与植被指数、纹理特征、强度特征、密度特征、高度特征和结构特征的扰动比例依次为 1.0%、1.5%、2.0%~2.5%、3.0%、3.0%~4.0% 和 4.0%~5.0%，分别模拟辐射与光照、窗口尺度、扫描几何、点云采样与遮挡、测距与地形归一化及冠层参数计算误差。上述比例用于不确定性情景模拟，不代表传感器的标称测量精度。

对于每个 Bootstrap 子模型，200 次扰动预测结果的离散程度用于表征输入特征误差引起的数据不确定性；先对各子模型的扰动预测结果取平均，再以 30 个子模型平均预测结果之间的离散程度表征模型不确定性。模型不确定性和数据不确定性均以预测结果的标准差表示。在 2 类不确定性近似独立的前提下，按照方差合成原则得到总不确定性。模型不确定性和数据不确定性的贡献率分别表示相应方差占总方差的百分比，两者之和为 100%，用于比较不同特征组合和不同模型中 2 类不确定性的相对贡献。

1.5.3　SHAP 分析　采用 SHAP 解释最终特征组合下精度最高模型中特征对碳储量预测的贡献[29]，计算各外层模型对相应测试样本的 SHAP 值，并按特征名称合并；未在相应外层折中入选的特征记为 0。以平均绝对 SHAP 值衡量特征的重要性，组内求和后计算 *H*、*D*、*S*、*I* 和 *O* 组的贡献率。

1.5.4　森林地上碳储量空间预测　根据嵌套交叉验证的精度选择空间预测模型，利用全部 63 个样地按相同的流程重新筛选特征，并通过内层 5 折交叉验证优化模型参数，建立全样本部署模型。以 UAV-LiDAR 覆盖区为预测范围，构建 30 m×30 m 格网提取相同的特征，获得森林地上碳储量预测值。

1.5.5　精度评价　模型精度评价采用嵌套交叉验证的外层 10 折结果。每折以 9 折样本训练、1 折样本作为独立测试集，并对未经扰动的测试数据进行预测，分别计算各外层测试折的决定系数 (R^2)、均方根误差 (RMSE) 和相对均方根误差 (rRMSE)，具体计算方法参考 QIU 等[15]，Bootstrap-MC 不参与精度指标计算。最终结果以“均值±标准差”的形式报告模型精度。

2　结果与分析

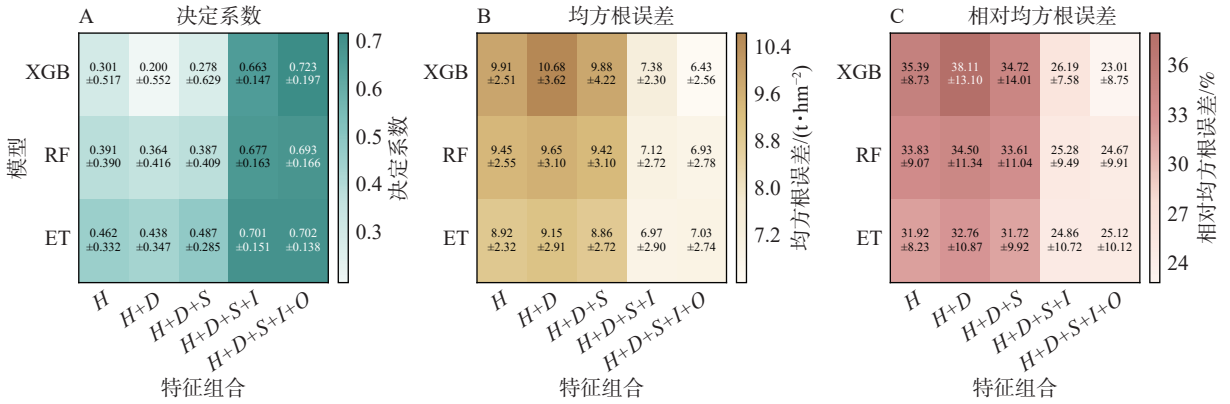
2.1　变量选择

全样本经过 Boruta-Spearman 特征筛选后共保留 9 个特征，其中包括 7 个 UAV-LiDAR 特征和 2 个 Sentinel-2 特征，分别为点云高度特征 AIH1 和 CRR，点云垂直分布密度特征 D2 和 D8，冠层结构特征 LAI，激光回波强度特征 AII1 和 AII10，以及光学特征 B11 和 ME_B11，仅用于最终模型拟合和空间预测。

2.2　不同模型与特征组合的碳储量估算精度

随着特征组合由 *H* 逐步扩展至 *H*+*D*+*S*+*I*+*O*，3 种模型总体表现为 R^2 升高，RMSE 和 rRMSE 降低趋势，但前 3 种组合间精度波动较大 (图 1)。与 *H*+*D*+*S* 组合相比，加入特征 *I* 后，3 种模型的 R^2 提高 0.214~0.385，RMSE 降低 1.89~2.5 t·hm^{-2}，精度提升最为明显；同时，折间的波动明显减小。加入特征 *O* 后，极端梯度提升模型和随机森林模型精度进一步提高，极端随机树模型则基本保持稳定。最终组合

下，极端梯度提升模型精度最高，R^2、RMSE 和 rRMSE 分别为 0.723±0.197、(6.43±2.56) t·hm^{-2} 和 23.01%±8.75%，极端随机树和随机森林模型的 R^2 分别为 0.702±0.138 和 0.693±0.166。



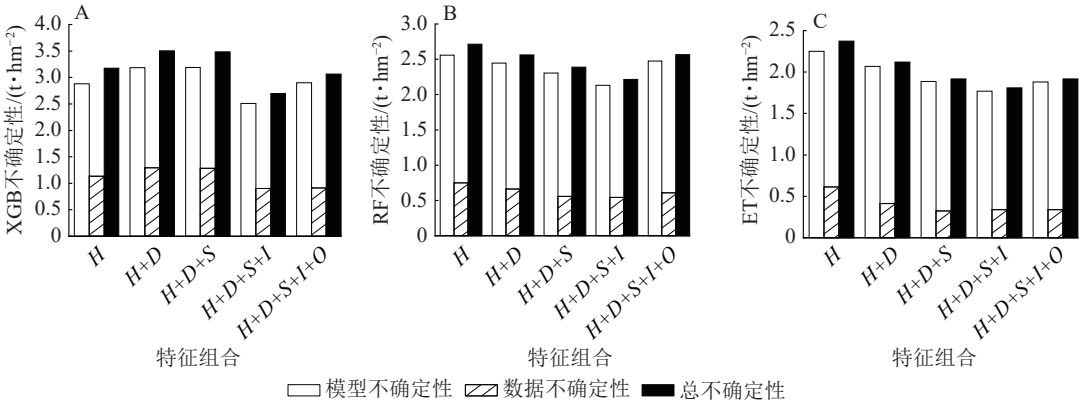
XGB. 极端梯度提升；RF. 随机森林；ET. 极端随机树。*H*、*D*、*S*、*I*分别为无人机激光雷达高度、密度、结构、强度类特征，*O*. Sentinel-2光谱纹理特征

图 1 不同特征组合下模型估算精度

Figure 1 Estimation accuracy of models under different feature combinations

2.3 不同特征组合下各模型的不确定性变化

递进式特征融合对 3 种模型总不确定性的影响呈非线性 (图 2)。由 *H* 扩展至 *H*+*D*+*S* 时，极端梯度提升模型的总不确定性先升后略降，随机森林模型和极端随机树模型持续下降。加入特征 *I* 后总不确定性均下降；加入特征 *O* 后，有所回升。最终特征组合下，极端随机树模型总不确定性最低，极端梯度提升模型最高。



XGB. 极端梯度提升树，RF. 随机森林，ET. 极端随机树。*H*、*D*、*S*、*I*为无人机激光雷达数据高度、密度、结构、强度特征，O. Sentinel-2数据光谱和纹理特征

图 2 不同模型不确定性对比

Figure 2 Uncertainty comparison among different models

为进一步明确总不确定性的主要来源，对最终特征组合下模型不确定性与数据不确定性的构成进行比较 (表 7)。3 种模型的总不确定性均以模型不确定性为主，其贡献率为 89.0%~95.5%，数据不确定性贡献率仅为 4.5%~11.0%，其中，极端梯度提升模型的数据不确定性贡献最高，极端随机树模型最低。

表 7 不同模型的不确定性及其来源贡献率

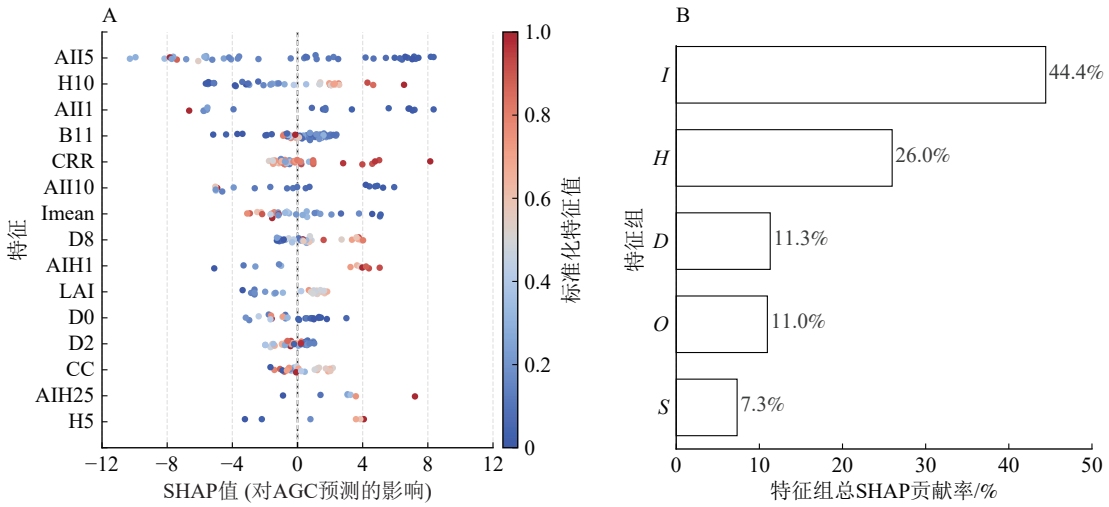
Table 7 Uncertainty and source contribution rates of different models

模型	模型不确定性/(t·hm^{-2})	数据不确定性/(t·hm^{-2})	总不确定性/(t·hm^{-2})	模型贡献率/%	数据贡献率/%
极端梯度提升模型	2.90±0.52	0.91±0.21	3.06±0.56	89.0±2.3	11.0±2.3
随机森林模型	2.47±0.53	0.61±0.14	2.57±0.55	92.7±1.3	7.3±1.3
极端随机树模型	1.88±0.32	0.34±0.06	1.92±0.32	95.5±1.5	4.5±1.5

2.4 SHAP 特征解释结果

3 种模型中极端梯度提升模型精度最高，因此采用 SHAP 方法解释其预测结果 (图 3)。在单特征层

面，AII5 重要性最高，H10、AII1 等特征也表现出较高贡献。在特征组层面，强度特征的贡献率最高，达到 44.4%。



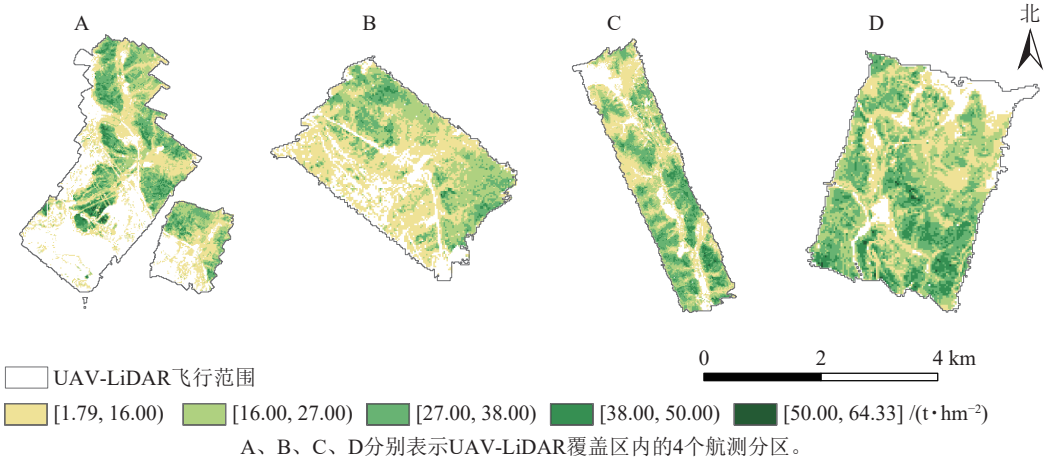
AGC. 地上碳储量。H10、CRR、AIH1、AIH25、H5. 高度类特征；D8、D0、D2. 密度类特征；LAI、CC. 结构类特征；AII5、AII1、AII10、Imean. 强度类特征；B11. Sentinel-2光谱特征。*H*、*D*、*S*、*I*. 无人机激光雷达高度、密度、结构、结构、强度类特征；*O*. sentinel-2光谱纹理特征。

图 3　极端梯度提升模型 SHAP 解释结果

Figure 3　SHAP interpretation results of the XGB model

2.5　森林地上碳储量空间分布

采用精度最高的极端梯度提升模型对 UAV-LiDAR 覆盖区高山松林森林地上碳储量进行空间预测 (图 4) 可知，森林地上碳储量为 1.79~64.33 t·hm^{-2}，整体以 16.00~38.00 t·hm^{-2} 为主，38.00 t·hm^{-2} 以上的较高值主要呈局部斑块状分布。各飞行区内不同森林地上碳储量等级交错分布，呈现出一定的空间差异。



A、B、C、D分别表示UAV-LiDAR覆盖区内的4个航测分区。

图 4　基于极端梯度提升模型的 UAV-LiDAR 覆盖区高山松林森林地上碳储量空间分布

Figure 4　Spatial distribution of AGC in *P. densata* forests within the UAV-LiDAR coverage area predicted using the XGB model

3　讨论

3.1　特征融合与关键特征的重要性

UAV-LiDAR 与 Sentinel-2 特征逐步融合总体提高了森林地上碳储量估算精度，主要是因为 LiDAR 点云能够直接获取森林三维结构信息，对碳储量差异具有较强表征能力[30]。Sentinel-2 光谱纹理特征主要反映冠层表层光谱响应和水平空间异质性。2 类数据的信息互补能够增强对样地间森林地上碳储量差异的识别能力，减少单一类型特征表征不足引起的预测偏差。这与已有多源遥感特征融合研究结论一致[6, 14]。

但密度与高度分布及冠层覆盖存在关联，加入密度特征后模型精度未明显改善，可能与新增信息有限及特征信息重叠有关，因此，特征增益取决于新增信息的互补性，而非特征数量。

进一步来看，强度特征加入后 3 种模型的精度提升最明显，其 SHAP 组贡献率达到 44.4%。已有研究将高度特征视为 LiDAR 森林地上生物量或碳储量估算的主要预测因子[31−32]，本研究中强度特征的贡献更高，这可能与研究区树种组成相对单一，冠层层次相对简单及高度特征对部分样地差异的区分能力有限有关。在相近采集条件下，强度信息反映激光脉冲与冠层目标相互作用后的回波能量差异，能够在高度、密度和冠层结构指标之外补充冠层内部组成和反射差异信息[33−34]，从而改善森林地上碳储量估算效果。GARCÍA 等[35] 研究表明：LiDAR 强度特征能够提高森林生物量及碳储量估算效果。因此，强度特征是当前研究区森林地上碳储量估算的重要信息来源。

高度特征组的贡献仅次于强度特征组，表明林分垂直尺度仍是森林地上碳储量估算的基础信息；密度特征则通过表征点云垂直分布和冠层充实程度提供补充。加入光学特征后极端梯度提升模型和随机森林模型精度提高，而极端随机树模型基本稳定，表明其增益具有模型依赖性。由于 Sentinel-2 主要反映冠层表层状态，易受到高郁闭度林分光谱饱和以及冠层遮蔽的影响，对内部生物量差异的响应有限[36]。结构特征与高度特征、密度特征存在信息重叠，导致 SHAP 贡献降低[32]，但仍能补充冠层占据状态和疏密信息从而提高估测精度。

3.2 不同模型估算性能差异与不确定性

递进式特征融合对总不确定性的影响呈非线性，说明新增特征的信息增益与模型复杂度共同影响预测稳定性。加入强度特征后，折间精度波动和总不确定性均下降；加入光学特征后总不确定性回升，说明小样本条件下新增特征及其交互可能提高模型对训练样本组成的敏感性。因此，特征增加并不必然降低不确定性，模型性能应结合平均精度、折间变异和不确定性综合评价。

3 种模型的相对表现随特征组合而变化。极端随机树模型在以 LiDAR 特征为主的前 4 种组合中平均精度较高，加入光学特征后则以极端梯度提升模型最优，说明模型优势与特征空间有关。极端梯度提升模型对新增强度信息及其交互关系的响应较强[24]，加入强度特征后总不确定性降幅最大，但其串行提升机制对 Bootstrap 训练样本变化较为敏感。随机森林模型通过样本和特征随机化降低模型依赖[25]，极端随机树模型进一步随机选择节点阈值以降低方差[26]。因此，极端梯度提升模型具有较高的平均估算精度，而极端随机树模型总不确定性较低，体现了模型拟合能力与稳定性之间的权衡。本研究基于空间预测精度选择极端梯度提升模型；若更强调稳定性，则可考虑极端随机树模型。

模型不确定性贡献率为 89.0%~95.5%，是总不确定性的主要来源。SHETTLES 等[37] 在利用 LiDAR 和地面数据估算森林生物量时也发现，模型误差是不确定性的重要来源。有限样本下，Bootstrap 重采样后训练样本组成的变化加大模型预测离散程度[27]。贡献率仅反映相对构成，极端随机树模型虽具有最高的模型不确定性贡献率，但其模型不确定性和总不确定性绝对值均最低。数据不确定性较低与输入特征扰动幅度和物理范围约束有关，其相对贡献受扰动比例和误差分布设定影响，仅反映当前参数条件[28]。

3.3 局限性与展望

本研究 63 个样地主要分布于海拔 3 200~3 400 m，对极端海拔、较陡坡地和部分坡向的覆盖不足。尽管嵌套交叉验证降低了信息泄漏和过拟合风险，但外层每个测试折仅包含约 6~7 个样地，折间波动表明模型评价仍受样本划分影响，有限的样本数量和空间代表性可能影响模型区域泛化能力。森林地上碳储量空间预测仅限于具有完整 UAV-LiDAR 和 Sentinel-2 特征的飞行覆盖区，且缺少独立区域验证，不能代表整个小中甸高山松林。此外，数据获取相差约 6 个月，Bootstrap-MC 扰动参数主要依据经验设定，尚未充分涵盖地面测量、森林地上碳储量参考值及遥感获取与处理误差。后续应扩大并分层均衡布设样地，开展独立区域或空间交叉验证；同时扩大 LiDAR 覆盖范围，利用重复观测和实测误差校准扰动参数，完善地面与遥感误差传播，提高高山松林地上碳储量估算结果及其区域适用性评价的可靠性。

4 结论

本研究融合 LiDAR 点云特征与 Sentinel-2 光谱纹理特征，构建递进式特征组合，结合极端梯度提升模型、随机森林模型、极端随机树模型、Bootstrap-MC 和 SHAP 方法，开展森林碳储量估算、不确定性

分析及特征解释。主要结论如下：①递进式特征组合能够整体提高森林地上碳储量估算精度，其增益取决于新增信息的互补性。强度特征加入后精度提升最明显。*H*+*D*+*S*+*I*+*O* 组合下极端梯度提升模型精度最高，决定系数、均方根误差和相对均方根误差分别为 0.723±0.197、(6.43±2.56) t·hm^{-2} 和 23.01%±8.75%。②特征融合对总不确定性的影响呈非线性。加入强度特征后 3 种模型的总不确定性均下降，加入光学特征后有所回升。模型不确定性贡献率为 89.0%~95.5%，是总不确定性的主要来源。③不同特征组对森林地上碳储量预测的贡献存在明显差异，强度特征是高山松林地上碳储量估算的主要驱动信息。SHAP 结果表明，强度特征组贡献率最高，为 44.4%。

5　参考文献

[1] WALKER W S, GORELIK S R, COOK-PATTON S C, *et al*. The global potential for increased storage of carbon on land[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2022, **119**(23): e2111312119. DOI: 10.1073/pnas.2111312119.

[2] 刘世荣, 王晖, 李海奎, 等. 碳中和目标下中国森林碳储量、碳汇变化预估与潜力提升途径[J]. 林业科学, 2024, **60**(4): 157−172. LIU Shirong, WANG Hui, LI Haikui, *et al*. Projections of China's forest carbon storage and sequestration and ways of their potential capacity enhancement[J]. *Scientia Silvae Sinicae*, 2024, **60**(4): 157−172. DOI: 10.11707/j.1001-7488.LYKX20230206.

[3] CHAVE J, RÉJOU-MÉCHAIN M, BÚRQUEZ A, *et al*. Improved allometric models to estimate the aboveground biomass of tropical trees[J]. *Global Change Biology*, 2014, **20**(10): 3177−3190. DOI: 10.1111/gcb.12629.

[4] LU Dengsheng, CHEN Qi, WANG Guangxing, *et al*. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems[J]. *International Journal of Digital Earth*, 2016, **9**(1): 63−105. DOI: 10.1080/17538947.2014.990526.

[5] WANG Yueting, JIA Xiang, CHAI Guoqi, *et al*. Improved estimation of aboveground biomass of regional coniferous forests integrating UAV-LiDAR strip data, Sentinel-1 and Sentinel-2 imageries[J]. *Plant Methods*, 2023, **19**(1): 65. DOI: 10.1186/s13007-023-01043-9.

[6] ZHANG Lingjun, ZHAO Yinyin, CHEN Chao, *et al*. UAV-LiDAR integration with Sentinel-2 enhances precision in AGB estimation for bamboo forests[J]. *Remote Sensing*, 2024, **16**(4): 705. DOI: 10.3390/rs16040705.

[7] JUCKER T, ASNER G P, DALPONTE M, *et al*. Estimating aboveground carbon density and its uncertainty in Borneo's structurally complex tropical forests using airborne laser scanning[J]. *Biogeosciences*, 2018, **15**(12): 3811−3830. DOI: 10.5194/bg-15-3811-2018.

[8] 傅煜, 雷渊才, 曾伟生. 区域尺度杉木生物量估计的不确定性度量[J]. 林业科学, 2014, **50**(12): 79−86. FU Yu, LEI Yuancai, ZENG Weisheng. Uncertainty assessment in regional-scale above ground biomass estimation of Chinese fir[J]. *Scientia Silvae Sinicae*, 2014, **50**(12): 79−86. DOI: 10.11707/j.1001-7488.20141211.

[9] 秦立厚, 张茂震, 钟世红, 等. 森林生物量估算中模型不确定性分析[J]. 生态学报, 2017, **37**(23): 7912−7919. QIN Lihou, ZHANG Maozhen, ZHONG Shihong, *et al*. Model uncertainty in forest biomass estimation[J]. *Acta Ecologica Sinica*, 2017, **37**(23): 7912−7919. DOI: 10.5846/stxb201609281973.

[10] 黄屹杰, 张加龙, 胡耀鹏, 等. 高山松地上生物量遥感估算的不确定性分析[J]. 浙江农林大学学报, 2022, **39**(3): 531−539. HUANG Yijie, ZHANG Jialong, HU Yaopeng, *et al*. Uncertainty analysis of estimating aboveground biomass of *Pinus densata* by remote sensing[J]. *Journal of Zhejiang A&F University*, 2022, **39**(3): 531−539. DOI: 10.11833/j.issn.2095-0756.20210473.

[11] 徐川, 林辉, 吴胜义, 等. 基于多层级协同融合策略的乔木林地上碳储量遥感估算研究[J/OL]. 西北林学院学报, 2026-04-30. https://link.cnki.net/urlid/61.1202.s.20260429.1451.006. XU Chuan, LIN Hui, WU Shengyi, *et al*. Multi-level collaborative (MLC) fusion for remote sensing estimation of aboveground carbon in arbor forests[J]. *Journal of Northwest Forestry University*, 2026-4-30. https://link.cnki.net/urlid/61.1202.s.20260429.1451.006.

[12] DU Chunyu, FAN Wenyi, MA Ye, *et al*. 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[J]. *Sensors*, 2021, **21**(17): 5974. DOI: 10.3390/s21175974.

[13] ZHANG Danhua, NI Hui. Inversion of forest biomass based on multi-source remote sensing images[J]. *Sensors*, 2023,

23(23): 9313. DOI: 10.3390/s23239313.

[14] 曹军, 张加龙, 肖庆琳, 等. 基于随机森林和蒙特卡洛的高山松地上碳储量估测及不确定性分析[J]. 林业科学研究, 2023, **36**(5): 131−139. CAO Jun, ZHANG Jialong, XIAO Qinglin, *et al*. Estimation and uncertainty analysis of aboveground carbon storage of *Pinus densata* based on random forests and Monte Carlo[J]. *Forest Research*, 2023, **36**(5): 131−139. DOI: 10.12403/j.1001-1498.20220603.

[15] QIU Bo, LI Sha, CAO Jun, *et al*. Uncertainty analysis of forest aboveground carbon stock estimation combining Sentinel-1 and Sentinel-2 images[J]. *Forests*, 2024, **15**(12): 2134. DOI: 10.3390/f15122134.

[16] 孙雪莲. 基于 Landsat 8-OLI 的香格里拉高山松林生物量遥感估算模型研究[D]. 昆明: 西南林业大学, 2016. SUN Xuelian. *Biomass Estimation Model of Pinus densata Forests in Shangri-La City Based on Landsat 8-OLI by Remote Sensing*[D]. Kunming: Southwest Forestry University, 2016.

[17] 王金亮, 王小花, 岳彩荣, 等. 滇西北香格里拉森林 4 个建群种的含碳率[J]. 生态环境学报, 2012, **21**(4): 613−619. WANG Jinliang, WANG Xiaohua, YUE Cairong, *et al*. Carbon content rate in dominant species of four forest types in Shangri-la, northwest Yunnan Province[J]. *Ecology and Environmental Sciences*, 2012, **21**(4): 613−619. DOI: 10.16258/j.cnki.1674-5906.2012.04.010.

[18] GITELSON A A, VIÑA A, ARKEBAUER T J, *et al*. Remote estimation of leaf area index and green leaf biomass in maize canopies[J]. *Geophysical Research Letters*, 2003, **30**(5): 2002GL016450. DOI: 10.1029/2002GL016450.

[19] WILSON E H, SADER S A. Detection of forest harvest type using multiple dates of Landsat TM imagery[J]. *Remote Sensing of Environment*, 2002, **80**(3): 385−396. DOI: 10.1016/S0034-4257(01)00318-2.

[20] 吴超, 徐伟恒, 黄邵东, 等. 林火监测中遥感应用的研究现状[J]. 西南林业大学学报, 2020, **40**(3): 172−179. WU Chao, XU Weiheng, HUANG Shaodong, *et al*. Research progress of remote sensing for forest-fire monitoring[J]. *Journal of Southwest Forestry University*, 2020, **40**(3): 172−179. DOI: 10.11929/j.swfu.201910060.

[21] DING Zhidan, SUN Zhao, XIE Yunhong, *et al*. Optimizing crown density and volume estimation across two coniferous forest types in Southern China *via* Boruta and Cubist methods[J]. *Journal of Plant Ecology*, 2024, **17**(5): rtae039. DOI: 10.1093/jpe/rtae039.

[22] KURSA M B, RUDNICKI W R. Feature selection with the boruta package[J]. *Journal of Statistical Software*, 2010, **36**(11): 1−13. DOI: 10.18637/jss.v036.i11.

[23] LIANG Mengyu, DUNCANSON L, SILVA J A, *et al*. Quantifying aboveground biomass dynamics from charcoal degradation in Mozambique using GEDI Lidar and Landsat[J]. *Remote Sensing of Environment*, 2023, **284**: 113367. DOI: 10.1016/j.rse.2022.113367.

[24] CHEN Tianqi, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//KRISHNAPURAM B, SHAH M. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 2016: 785−794. DOI: 10.1145/2939672.2939785.

[25] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, **45**(1): 5−32. DOI: 10.1023/A:1010933404324.

[26] GEURTS P, ERNST D, WEHENKEL L. Extremely randomized trees[J]. *Machine Learning*, 2006, **63**(1): 3−42. DOI: 10.1007/s10994-006-6226-1.

[27] EFRON B. Bootstrap methods: another look at the jackknife[M]//KOTZ S, JOHNSON N L. *Breakthroughs in Statistics*. New York: Springer, 1992: 569−593. DOI: 10.1007/978-1-4612-4380-9_41.

[28] WANG Guoqing, DONG Wenquan, ZHANG Huaiqing, *et al*. Transferability of aboveground biomass estimation using Sentinel-1/2 and GEDI data in subtropical forests of complex terrain, China[J]. *iScience*, 2026, **29**(4): 115346. DOI: 10.1016/j.isci.2026.115346.

[29] LUNDBERG S M, ERION G, CHEN H, *et al*. From local explanations to global understanding with explainable AI for trees[J]. *Nature Machine Intelligence*, 2020, **2**(1): 56−67. DOI: 10.1038/s42256-019-0138-9.

[30] LEFSKY M A, COHEN W B, HARDING D J, *et al*. Lidar remote sensing of above-ground biomass in three biomes[J]. *Global Ecology and Biogeography*, 2002, **11**(5): 393−399. DOI: 10.1046/j.1466-822x.2002.00303.x.

[31] ZHANG Xiaofang, LI Xiaoyao, SHARMA R P, *et al*. Forest above-ground biomass estimation based on strongly collinear variables derived from airborne laser scanning data[J]. *Ecological Indicators*, 2024, **166**: 112517. DOI: 10.1016/j.ecolind.2024.112517.

[32] CHENG Xiaofang, ZHOU Lai, LIU Shaoyu, *et al*. Improving forest above-ground biomass estimation using UAV LiDAR and RGB with machine learning algorithm[J]. *Forests*, 2025, **16**(12): 1819. DOI: 10.3390/f16121819.

[33] YOU Haotian, WANG Tiejun, SKIDMORE A K, *et al*. Quantifying the effects of normalisation of airborne LiDAR intensity on coniferous forest leaf area index estimations[J]. *Remote Sensing*, 2017, **9**(2): 163. DOI: 10.3390/rs9020163.

[34] KIM S, MCGAUGHEY R J, ANDERSEN H E, *et al*. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data[J]. *Remote Sensing of Environment*, 2009, **113**(8): 1575−1586. DOI: 10.1016/j.rse.2009.03.017.

[35] GARCÍA M, RIAÑO D, CHUVIECO E, *et al*. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data[J]. *Remote Sensing of Environment*, 2010, **114**(4): 816−830. DOI: 10.1016/j.rse.2009.11.021.

[36] CAO Luodan, PAN Jianjun, LI Ruijuan, *et al*. Integrating airborne LiDAR and optical data to estimate forest aboveground biomass in arid and semi-arid regions of China[J]. *Remote Sensing*, 2018, **10**(4): 532. DOI: 10.3390/rs10040532.

[37] SHETTLES M, HILKER T, TEMESGEN H. Examination of uncertainty in per unit area estimates of aboveground biomass using terrestrial LiDAR and ground data[J]. *Canadian Journal of Forest Research*, 2016, **46**(5): 706−715. DOI: 10.1139/cjfr-2015-0265.