

中南林业科技大学学报

Journal of Central South University of Forestry & Technology

ISSN 1673-923X，CN 43-1470/S

《中南林业科技大学学报》2026年第8期

Doi:10.14067/j.cnki.1673-923X.2026.08.007

论文标题： 融合RFE-CV和Optuna-TPE超参数优化的机载LiDAR湿地松人工林碳储量估测模型

作　　者： 盛文蕾，温小荣，汪求来，等

收稿日期： 2025-07-15

引文格式： 盛文蕾,温小荣,汪求来,等.融合RFE-CV和Optuna-TPE超参数优化的机载LiDAR湿地松人工林碳储量估测模型[J].中南林业科技大学学报,2026,46(8):68-78.

Sheng W L, Wen X R, Wang Q L, et al. An integrated model for estimating carbon stock in airborne LiDAR-mapped wetland pine plantations combining RFE-CV and Optuna-TPE hyperparameter optimization[J]. Journal of Central South University of Forestry & Technology, 2026,46(8):68-78.

相关论文推荐

Recommended articles

融合无人机和手持激光雷达数据的思茅松单木地上生物量估测

Estimation of aboveground biomass of individual *Pinus kesiya* trees based on fused UAV and handheld LiDAR data

基于无人机激光雷达点云的单木生物量估测

Single tree biomass estimation based on UAV LiDARpoint cloud

地面激光雷达扫描参数与点云简化对林木参数提取的影响

Effects of terrestrial LiDAR scanning parameters and point cloud simplification on forest parameters extraction

基于无人机高密度LiDAR点云的人工针叶林单木分割算法

The single tree segmentation of UAV high-density LiDAR point cloud data based on coniferous plantations

联合UAV-LiDAR和HMLS技术的森林样地点云数据融合

Point cloud data fusion of forest plots based on UAV-LiDAR and HMLS technologies

基于星载光子计数雷达数据的森林郁闭度估测模型优化

Optimization of forest canopy closure estimation model based on spaceborne photon counting LiDAR data

Doi:10.14067/j.cnki.1673-923x.2026.08.007 http: //qks.csuft.edu.cn

融合 RFE-CV 和 Optuna-TPE 超参数优化的机载 LiDAR 湿地松人工林碳储量估测模型

盛文蕾[1a]，温小荣[1b]，汪求来[2]，叶金盛[2]

（1. 南京林业大学 a. 林草学院、水土保持学院；b. 南方现代林业协同创新中心，江苏 南京 210037；
2. 广东省林业调查规划院，广东 广州 510520）

摘 要：【目的】融合机载 LiDAR 点云特征与超参数优化算法，构建精准的湿地松 *Pinus elliottii* 人工林碳储量估测模型，为大尺度碳汇监测提供科学依据，进而制定有助于实现“双碳”目标的森林经营决策。【方法】以广东省湿地松人工林为对象，基于机载 LiDAR 数据提取 101 个点云特征变量，采用递归特征消除结合交叉验证（RFE-CV）筛选最优特征组合，并引入基于 TPE 算法的 Optuna 超参数优化框架（以 RMSE 为目标函数）构建随机森林（RF）、K- 近邻（KNN）和极端梯度提升（XGBoost）模型，模型性能通过 R^2、RMSE 和 MAE 进行评估，利用 SHAP 值解析特征对湿地松碳储量预测的贡献机制。【结果】1）运用五折 RFE-CV 筛选出 5 个与湿地松碳储量估测紧密相关的最优特征变量（A_{IH40}、A_{IH80}、H_{10}、H_{25}、I_{max}），相较于使用全部变量构建模型，特征筛选显著提高了各模型的预测精度。其中，XGBoost 模型效果最明显（R^2 从 0.78 提升至 0.89，RMSE 从 0.63 t 降至 0.46 t，MAE 从 0.48 t 降至 0.41 t）。2）基于 TPE 算法的 Optuna 超参数优化框架效果显著优于随机搜索，XGBoost 模型的 RMSE 值降低了 2.31%，误差调整百分比（13.13%）达到随机搜索的 3.57 倍；RF 和 KNN 模型的收敛迭代次数分别减少 33.33% 和 49.18%。3）经 Optuna-TPE 优化后，各模型精度显著提升，XGBoost 模型精度最优，R^2 从 0.89 显著提升至 0.93。4）SHAP 值表征不同模型的特征依赖性差异，A_{IH40} 主导 XGBoost 模型预测（SHAP 值为 0.64），而 H_{25} 对 RF 模型更具关键性（SHAP 值为 0.28）。【结论】特征选择和基于 TPE 算法的 Optuna 超参数调优对提升湿地松碳储量估测模型性能至关重要。优化后的 XGBoost 模型结合最优特征组合，实现了对湿地松人工林碳储量的高精度估测。

关键词：Optuna-TPE 超参数；碳储量；随机森林；KNN；XGBoost；贝叶斯优化

中图分类号：S718.5 **文献标志码：**A **文章编号：**1673-923X(2026)08-0068-11

An integrated model for estimating carbon stock in airborne LiDAR-mapped wetland pine plantations combining RFE-CV and Optuna-TPE hyperparameter optimization

Sheng Wenlei[1a], Wen Xiaorong[1b], Wang Qiulai[2], Ye Jinsheng[2]

(1. a. College of Forestry and Grassland, College of Soil and Water Conservation;
b. Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, Jiangsu, China;
2. Guangdong Forestry Survey and Planning Institute, Guangzhou 510520, Guangdong, China)

Abstract:【Objective】This study integrated airborne LiDAR point cloud features with a hyperparameter optimization algorithm to develop a precise carbon stock estimation model for *Pinus elliottii* plantations in wetlands. The aim is to provide a scientific basis for large-scale carbon sink monitoring, thereby supporting forest management decisions that align with the "dual carbon" goals.【Method】Taking *Pinus elliottii* plantation in Guangdong Province as the research object, 101 point cloud feature variables were extracted based on

收稿日期（Received）：2025-07-15 **修订日期**（Revised）：2025-08-21

基金项目：广东省林业科技创新项目（2021KJCX001）；国家重点研发计划项目（2016YFC0502704）；江苏省高校优势学科建设工程资助项目（PAPD）。

第一作者：盛文蕾（shengwenlei_1999@163.com），硕士研究生。

通信作者：温小荣（njw9872@163.com），副教授，博士，硕士研究生导师。

引文格式：盛文蕾 , 温小荣 , 汪求来 , 等 . 融合 RFE-CV 和 Optuna-TPE 超参数优化的机载 LiDAR 湿地松人工林碳储量估测模型 [J]. 中南林业科技大学学报 ,2026,46(8):68-78.

Sheng W L, Wen X R, Wang Q L, et al. An integrated model for estimating carbon stock in airborne LiDAR-mapped wetland pine plantations combining RFE-CV and Optuna-TPE hyperparameter optimization[J]. Journal of Central South University of Forestry & Technology,2026,46(8):68-78.

airborne LiDAR data. The recursive feature elimination combined with cross-validation (RFE-CV) was used to screen the optimal feature combination, and the Optuna hyperparameter optimization framework based on the TPE algorithm (with RMSE as the objective function) was introduced to construct random forest (RF), k-nearest neighbor (KNN) and extreme gradient boosting (XGBoost) models. The model performance was evaluated by R^2, RMSE and MAE, and the SHAP value was used to analyze the contribution mechanism of features to the prediction of carbon storage of *Pinus elliottii.*【Result】1) Five optimal feature variables (A_{IH40}, A_{IH80}, H_{10}, H_{25}, I_{max}) closely related to the estimation of *Pinus elliottii* carbon storage were selected by 5-fold RFE-CV. Compared with the model using all variables, feature selection significantly improved the prediction accuracy of each model. Among them, the XGBoost model has the most obvious effect (R^2 increased from 0.78 to 0.89, RMSE decreased from 0.63 to 0.46 t, MAE decreased from 0.48 to 0.41 t); 2) The Optuna hyperparameter optimization framework based on the TPE algorithm is significantly better than random search: the RMSE value of the XGBoost model is reduced by 2.31%, and the error adjustment percentage (13.13%) is 3.57 times that of random search. The convergence iterations of the RF and KNN models are reduced by 33.33% and 49.18%, respectively; 3) After optimization by Optuna-TPE, the accuracy of each model has been significantly improved: the XGBoost model achieves the highest accuracy, with its R^2 significantly increasing from 0.89 to 0.93, thus becoming the optimal model; 4) SHAP value analysis revealed the difference of feature dependence in different slash pine models: A_{IH40} dominated XGBoost prediction (SHAP = 0.64), while H_{25} was more critical to the RF model (SHAP = 0.28).【Conclusion】This study showed that feature selection and Optuna hyperparameter tuning based on the TPE algorithm are crucial to improve the performance of the *Pinus elliottii* carbon storage estimation model. The optimized XGBoost model combined with the optimal feature combination realized the high-precision estimation of the carbon storage of *Pinus elliottii* plantations.
Keywords: Optuna-TPE hyperparameter; carbon storage; random forest; KNN; XGBoost; Bayesian optimization

全球城市化与工业化进程的加速增加了碳排放量，由此引发的全球气候变暖已成为当前最严峻的生态危机之一 [1-3]。森林作为陆地碳汇的核心载体，其碳储量的动态变化对全球碳循环与气候变化具有关键影响 [3-4]。在我国推进“碳达峰”“碳中和”目标的背景下，湿地松 *Pinus elliottii* 作为南方丘陵地区重要的速生经济造林树种，因固碳率高而成为碳汇造林的重要树种之一 [5-6]，对其森林碳储量的精准评估是实现“双碳”目标，应对全球气候变化的关键环节。传统碳储量估算方法（如样地清查法、微气象学法等）存在效率低下、适用范围有限等问题 [7-9]；光学遥感则受限于植被冠层穿透能力，难以获取森林垂直结构信息 [10]。机载激光雷达（LiDAR）技术通过提取点云的高度、强度、覆盖度等特征变量，为森林碳储量模型的高精度估算提供新路径 [11-14]。研究表明，融合 LiDAR 结构特征（如随机森林、XGBoost 等机器学习算法）可有效实现单木或多树种尺度碳储量的高精度反演 [15-18]，但针对湿地松人工林的大尺度、高效、高精度动态遥感监测方法仍有待完善 [19]。

与此同时，为提高碳储量模型估测效率并防止过度拟合，特征变量选择和超参数优化尤为关键 [15-16]。在特征变量选择方面，随机森林（RF）、递归特征消除（RFE）等算法能有效筛选关键变量，从而提升模型精度。例如，Cheng 等 [17] 基于遥感数据采用随机森林算法筛选特征变量，显著提升了森林碳储量模型估算精度（R^2 为 0.63 ～ 0.66，RMSE 为 36.04 Mg/hm^2）。高宇航 [20] 利用广西高峰林场桉树的遥感影像数据和实测数据，结合相关性分析和随机森林算法对变量进行重要性排序，从 31 个遥感特征变量中筛选出 10 个关键变量进行建模；Luo 等 [18] 首次将递归特征消除（RFE）与类别特征梯度提升（CatBoost）模型结合应用于温带森林地上生物量估算，通过筛选机载 LiDAR 冠层高度分位数、多光谱植被指数（如 NDRE）等关键特征，使模型精度显著提升（R^2 为 0.87，RMSE 为 28.6 Mg/hm^2）。在超参数优化方面，常用的超参数优化方法主要有随机搜索、网格搜索以及贝叶斯优化法 [21-24]。随机搜索效率高于网格搜索，但精度不稳定 [24]。贝叶斯优化则在兼顾效率与效果上更具优势，能有效减少冗余计算 [25]。树结构 Parzen 估计器（Tree-structured parzen estimator，TPE）作为改进的贝叶斯优化方法，无需假设参数独立性，在神经网络调优中已被证明优于随机搜索。Bergstra 等 [22] 首次提出树结构 Parzen 估计器（TPE），并证明其在神经网络和深度信念网络（DBNs）的调参任务中优于随机搜索和网格搜索。此外，目标函数的合理设定往往是超参数优化过程的核心 [21]。例如，在回归任务中，均方根误差（RMSE）因对异常值具有敏感性，常被用于提升模型的鲁棒性，进而实现数值的精准预估 [26-27]；在分类问题中，准确率、召回率、F1 分数等能够有效评估分类效果的指标则更为适用 [28]。尽管基于 TPE 算法的 Optuna 优化框架在深度学习领域的调参成效显著，但其在森林碳储量模型构建中的应用仍属空白 [25]，尤其是在结合 LiDAR 多维特征开展协同优化方面。当前研究存在双重局限：一是缺乏对 Optuna 处理复杂非线性碳储模型适用性的验证；二是尚未建立遥感多维特征与 TPE 算法的协同机制以适配复杂地形 [21-23]。突破超参数动

态调整与特征变量优化的协同范式，已成为提升复杂地区碳储量估算精度的关键突破口。

本研究提出“RFE-CV 与 Optuna-TPE 超参数优化”的协同框架，以湿地松人工林为对象，通过整合机载 LiDAR 数据与地面实测数据，采用结合五折交叉验证（CV）的递归消除法（RFE）筛选最优点云特征组合，进而构建基于优选特征变量的碳储量估算模型（包括 RF、KNN 及 XGBoost 模型）并开展精度评价。在此基础上，引入以 RMSE 为目标函数的 Optuna-TPE 超参数优化框架，旨在构建“特征筛选与超参数优化”的协同机制，实现自动搜索与特征选择的动态耦合，以克服现有研究中超参数优化的局限性，探索该方法在湿地松机载 LiDAR 数据处理中的可行性，为湿地松人工林碳储量的遥感估算提供新的方法论支持[17]，并为“双碳”目标的实现提供科学依据。

1　材料与方法

1.1　研究区概况

本研究随机选择广东省的 6 个林场，地理位置为 21°93′ ～ 24°49′N，109°89′ ～ 115°81′E。研究区属于亚热带季风气候，是我国光、热和水资源的丰富地区，且雨热同季，地貌类型复杂多样，有山地、丘陵、台地和平原。地势总体北高南低，北部多为山地和高丘陵，南部则为平原和台地。土壤类型主要为红壤和黄壤，植被覆盖度较高，平均覆盖度约为 75%。主要植被类型包括南亚热带常绿阔叶林、针阔混交林及针叶林；主要优势树种包括杉木 *Cunninghamia lanceolata*、马尾松 *Pinus massoniana*、桉树 *Eucalyptus*、湿地松 *Pinus elliottii* 等。

1.2　数据收集与预处理

1.2.1　外业数据收集

本研究以广东省湿地松人工林为对象，布设 41 块 30 m×30 m 的方形样方，对样地内胸径为 5 cm 以上的林木进行每木检尺，调查并记录树种相关信息以及高程信息，并利用载波相位差分技术（Realtime kinematic，RTK）进行定位。同时，在实测相同位置、无人机不限飞的样地中，使用 DJI 经纬 M300RTK 无人机搭载 DJI L1 禅思航摄仪进行机载 LiDAR 数据的采集，飞行高度为 100 m，航线速度为 10 m/s，激光旁向重叠率为 50%，采样频率为 160 kHZ，采用三回波模式进行重复扫描。实测样地信息见表 1。

表 1　样地信息汇总
Table 1　Summary of plot information

样地号 Plot No.	株数 Tree number / 株	平均树高 Mean height /m	平均胸径 Mean DBH /cm	样地号 Plot No.	株数 Tree number / 株	平均树高 Mean height /m	平均胸径 Mean DBH /cm
1	86	17.48	11.94	22	140	12.89	13.01
2	74	18.60	13.05	23	118	14.02	11.37
3	90	13.77	9.84	24	113	13.35	11.26
4	101	12.79	10.11	25	152	12.25	11.56
5	87	17.35	11.65	26	112	11.16	8.79
6	82	15.60	11.12	27	98	9.67	8.05
7	78	17.36	11.47	28	70	16.26	10.19
8	126	13.52	8.74	29	109	13.76	13.16
9	125	12.84	8.04	30	101	16.16	13.21
10	143	12.22	9.08	31	121	16.54	15.31
11	36	21.75	13.56	32	125	7.78	4.66
12	50	25.83	17.39	33	61	9.34	4.90
13	140	13.07	9.33	34	168	10.88	6.28
14	48	28.43	19.31	35	110	14.45	11.67
15	101	16.60	14.47	36	103	14.25	12.22
16	116	15.69	14.70	37	102	15.68	10.78
17	105	15.37	14.96	38	85	16.75	11.78
18	64	17.16	15.50	39	67	19.49	13.40
19	164	12.51	10.85	40	77	18.41	13.07
20	81	14.72	8.44	41	71	22.90	16.56
21	45	24.77	14.45				

1.2.2　内业数据预处理

利用 DJI Terra 软件将无人机原始数据生成为 LAS 点云数据及影像数据，随后使用 LiDAR360 分别对各采样点的激光雷达点云数据进行去噪处理，再通过改进的渐进加密三角网滤波算法 [28] 完成地面点与非地面点的分类。采用克里金插值法生成数字高程模型（DEM），并基于该 DEM 进行归一化处理，最终提取归一化数字表面模型（nDSM）。在此基础上进一步提取森林参数，包括高度特征变量、强度特征变量、覆盖度、间隙率及叶面积指数，共得到 101 个点云特征变量，详见表 2。

表 2　点云特征变量
Table 2　Point cloud characteristic variable

点云特征变量 Point cloud feature variable	描述 Description	点云特征变量 Point cloud feature variable	描述 Description
$A_{IH1}, A_{IH5}, \ldots, A_{IH99}$	累计高度百分位数，共 15 个	$I_{A1}, I_{A5}, \ldots, I_{A99}$	累计强度百分位数，共 15 个
$H_1, H_5, \ldots, H_{99}$	高度百分位数，共 15 个	$I_1, I_5, \ldots, I_{99}$	强度百分位数，共 15 个
$D_1, D_2, \ldots, D_{10}$	密度变量，共 10 个	I_{max}	强度最大值
H_{crr}	冠层起伏率	I_{min}	强度最小值
H_{max}	高度最大值	I_{median}	强度中位数
H_{min}	高度最小值	I_{mean}	强度平均值
H_{mean}	高度平均值	I_{madmed}	强度中位数绝对偏差的中位数
H_{madmed}	高度中位数绝对偏差的中位数	I_{aad}	强度平均绝对偏差
H_{aad}	高度平均绝对偏差	$I_{skewness}$	强度偏斜度
$H_{skewness}$	高度偏斜度	I_{var}	强度方差
H_{var}	高度方差	I_{std}	强度标准差
H_{std}	高度标准差	$I_{kurtosis}$	强度峰度
$H_{kurtosis}$	高度峰度	I_{cv}	强度变异系数
H_{median}	高度中位数	I_{iq}	强度百分位数四分位数间距
H_{cv}	变异系数	C_C	覆盖度
H_{sms}	二次幂平均	G_F	间隙率
H_{cmc}	三次幂平均	L_{AI}	叶面积指数
A_{IHiq}	累计高度百分位数四分位数间距	H_{iq}	高度百分位数四分位数间距

1.3　研究方法

1.3.1　碳储量计算

基于单木实测胸径和树高数据，参考 GB/T 43648—2024 主要树种立木生物量模型与碳计量参数 [29] 计算单木生物量，将单木生物量乘以含碳系数得到单木碳储量，单木碳储量累计相加即为样地林分碳储量。具体公式如下：

$$W = 0.047\ 440 \times D^{2.103\ 59} \times H^{0.631\ 08} \times 10^{-3}。（1）$$

式中：W 为单木生物量；D 为单木实测胸径；H 为单木实测树高。

$$C = C_0 \times W。（2）$$

式中：C 为单木碳储量；C_0 为含碳系数，采用立木标准规定的 0.5。

1.3.2　基于 RFE-CV 的特征变量筛选方法

在利用机载激光雷达数据构建湿地松碳储量模型的过程中，特征选择是确保模型精准反映真实碳储量情况、提升计算效率的关键环节 [30]。本研究使用递归特征消除结合五折交叉验证（RFE-CV）的方法筛选最优特征子集，RFE 的原理是基于一个基础模型，通过迭代移除贡献度最低的特征，直到获得最佳特征子集。本研究设定 RF 为 RFE 的基础模型，并设置每次迭代剔除 1 个特征。为增强模型稳健性，本研究将特征消除过程与五折交叉验证相结合：在每次评估特征子集大小时，通过交叉验证确保所选特征组合在不同数据划分下均能保持最优性能，并以交叉验证的均方根误差（RMSE）作为评估各子集性能的指标，最终输出最优特征子集。此外，为深入探究变量组合对模型精度的影响，本研究分别基于全部点云特征变量与最优特征变量组合，独立构建 k 近邻（KNN）和极限梯度提升（XGBoost）模型对其精度进行系统的评估。

1.3.3　以 RMSE 为目标函数的 Optuna-TPE 模型超参数调优方法

针对湿地松人工林 LiDAR 数据高维度、非均匀分布的特性，为提升研究效率、减少人为误差、增强模型对复杂点云特征的适配能力，同时更精

准地反映模型估测效果，本研究引入基于 TPE 算法的 Optuna 超参数优化框架，以最小化测试集均方根误差（RMSE）为目标函数，设定优化方向为最小化该目标函数值，并开展 100 次迭代优化。此外，通过系统对比 Optuna 和随机搜索超参数优化方法在湿地松碳储量预测模型中的调优效果，评估 Optuna 的有效性。本研究以均方根误差（RMSE）为核心评价指标，同时引入收敛迭代次数（Convergence iteration，CI）与误差调整百分比（Error adjustment percentage，EAP）2 个关键性能指标。RMSE 值代表模型预测值与真实值的均方根误差，数值越低表明模型预测精度越高；CI 值指优化算法首次达到全局最优 RMSE 值（或当前迭代中的历史最优 RMSE 值）所需的试验次数，反映优化算法的搜索效率（数值越低则收敛速度越快）；EAP 值表示优化过程中模型性能的相对提升幅度，该值越高说明优化效果越好。

$$R_{\mathrm{MSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(C_i-\hat{C}_i)^2}\ ; \tag{3}$$

$$C_{\mathrm{I}}=\frac{a}{b}; \tag{4}$$

$$E_{\mathrm{AP}}=\frac{R_{\mathrm{MSE0}}-R_{\mathrm{MSEa}}}{R_{\mathrm{MSE0}}}\times 100\%\ 。 \tag{5}$$

式中：C_i 为实际碳储量，$\hat{C}_i$ 为估算碳储量，a 为首次达到最优 RMSE 的次数，b 为总迭代次数，R_{MSE0} 为初始 RMSE 值，R_{MSEa} 为最佳 RMSE 值。

1.3.4　湿地松碳储量模型构建方法

本研究将所有样本按 4∶1 的比例划分为训练集与验证集，利用训练集数据构建随机森林（RF）、K- 近邻（KNN）及极端梯度提升（XGBoost）模型。其中，RF 模型 [31] 采用集成学习策略，通过构建多棵决策树进行预测，并持续调整决策树数量、最大深度等关键参数，从而有效处理碳储量数据中的复杂关系；KNN 模型 [32] 则依据输入样本在特征空间中的距离度量，为待预测样本筛选出训练集中的 K 个最邻近样本，再利用这 K 个样本的标签或数值进行预测；XGBoost 模型 [33] 基于梯度提升决策树（GBDT）框架迭代生成一系列决策树，在每一轮迭代过程中，通过最小化损失函数的梯度（一阶导数）与海森矩阵（二阶导数）来评估当前集成模型的伪残差，并以此优化新增决策树的训练目标（即拟合残差），同时引入正则化项以控制模型复杂度。

1.3.5　模型精度评估

建立评价指标决定系数（R^2）、均方根误差和平均绝对误差（Mean absolute error，MAE）来评估模型的有效性和稳定性。R^2 值越接近 1，表示模型对数据的解释能力越强；RMSE 值越小，表示模型预测值与实际值之间的差异越小；MAE 值越小，表示模型预测值与实际值之间的差异越小。

$$R^2=1-\frac{\sum_{i=1}^{n}(C_i-\hat{C}_i)^2}{\sum_{i=1}^{n}(C_i-\bar{C}_i)^2}; \tag{6}$$

$$M_{\mathrm{AE}}=\frac{1}{n}\sum_{i=1}^{n}\left|\hat{C}_i-C_i\right|。 \tag{7}$$

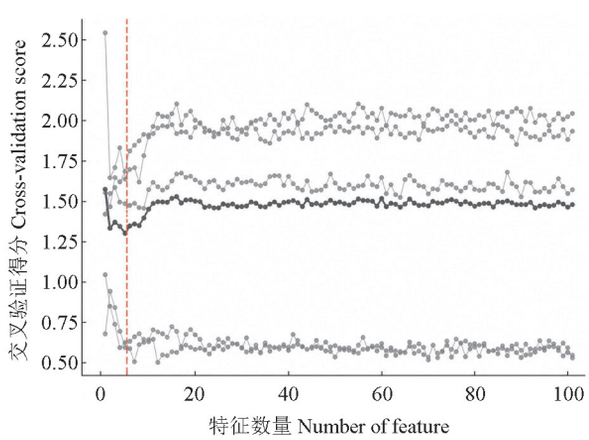
式中：C_i 为实际碳储量，$\bar{C}_i$ 为实际碳储量平均值，$\hat{C}_i$ 为估算碳储量。

2　结果与分析

2.1　基于 RFE-CV 的特征变量筛选结果

基于机载 LiDAR 数据提取的 101 个特征变量，本研究采用五折交叉验证的递归特征消除（RFE-CV）方法进行最优特征子集筛选，最终筛选出 5 个最优特征变量，其中包含 4 个高度变量（A_{IH40}、A_{IH80}、H_{10}、H_{25}）和 1 个强度变量（$I_{\max}$）。

由图 1 可知，当特征数量为 5 时，RMSE 得分达到最低（1.25），表明此时模型性能最佳。当特征数量为 0 ～ 5 时，平均 RMSE 得分逐渐降低，说明特征的引入使模型性能有所提升；而当特征数量超过 5 个后，RMSE 得分逐渐增加并趋于平稳，



每个灰色点表示 1 次交叉验证的均方根误差（RMSE）得分，每条灰色线表示一折交叉验证的变化趋势。黑色粗线表示五次交叉验证的 RMSE 得分的平均值。红色虚线标出了最优特征数量（5 个）。

Each gray dot represents the score–root mean square error (RMSE) from one cross-validation run, while each gray line shows the trend of variation across the folds. The thick black line indicates the average RMSE across all five cross-validation runs. The red dashed line identifies the optimal number of features (5).

图 1　RFE-CV 特征选择的准确性随特征数量的变化趋势

Fig. 1　The variation trend of the accuracy of RFE-CV feature selection with the number of features

这意味着继续增加特征变量已无法进一步提升模型性能。因此，本研究选择这 5 个特征作为最优特征子集是合理的。

分别采用全部变量与优选变量构建 RF、KNN 及 XGBoost 模型，各模型精度详见表 3。对于所有模型而言，使用优选变量组合建模后，其性能均得到显著提升。其中，XGBoost 模型在不同特征选择条件下均表现最优，尤其在优选变量条件下，其 R^2 值高达 0.89，RMSE 与 MAE 分别低至 0.46、0.41 t，这表明该模型对数据的拟合能力与泛化能力均较强。同时，RF 模型在优选变量条件下性能也有明显改善，R^2 达到 0.86，RMSE 和 MAE 分别为 0.51、0.41 t，展现出特征优化后的良好表现。但是，KNN 模型的精度提升幅度相对较小，性能也略逊于 RF 和 XGBoost 模型。总体来看，特征选择在提升模型性能方面发挥了关键作用，优选变量不仅提高了所有模型的 R^2 值，还降低了误差指标，有力证明了关键特征在模型训练过程中的重要性。结果表明，在实际应用中，采用特征选择结合高效模型（如 XGBoost 或 RF）能够有效提高预测精度与模型可靠性。因此，本研究采用优选特征变量组合开展后续模型构建是可行的。

表 3　模型精度评价
Table 3　Evaluation of model accuracy

变量 Variable	模型 Model	决定系数 R^2	均方根误差 RMSE /t	平均绝对误差 MAE /t
全部变量 All variables	RF	0.71	0.74	0.63
	KNN	0.66	0.81	0.69
	XGBoost	0.78	0.63	0.48
优选变量 Optimal variables	RF	0.86	0.51	0.41
	KNN	0.72	0.73	0.61
	XGBoost	0.89	0.46	0.41

2.2　以 RMSE 为目标函数的 Optuna-TPE 超参数调优结果

为验证 Optuna 在 LiDAR 数据驱动的碳储量建模中的优势，本研究对比了 Optuna 与随机搜索（Random search）的超参数优化效率。如图 2 和表 4 所示，Optuna 在森林碳储量预测模型的参数优化中展现出显著优势。具体而言，在模型精度提升方面，Optuna 在所有模型中均获得更低 RMSE：XGBoost 模型误差降低 2.31%、KNN 模型降低 0.34%、RF 模型降低 0.20%。其中，Optuna 优化的 XGBoost 模型实现了 13.13% 的误差调整百分比，显著优于随机搜索（3.68%），充分验证了 Optuna 对处理复杂模型的高度适应性。在收敛效率的提升方面，在优化 RF 模型时，Optuna 较随机搜索迭代效率提升 33.33%；Optuna 在优化 KNN 模型时迭代效率提升 49.18%；在优化 XGBoost 模型时，尽管 Optuna 与随机搜索的收敛迭代次数相同（27 次），但 Optuna 实现了更大的误差调整百

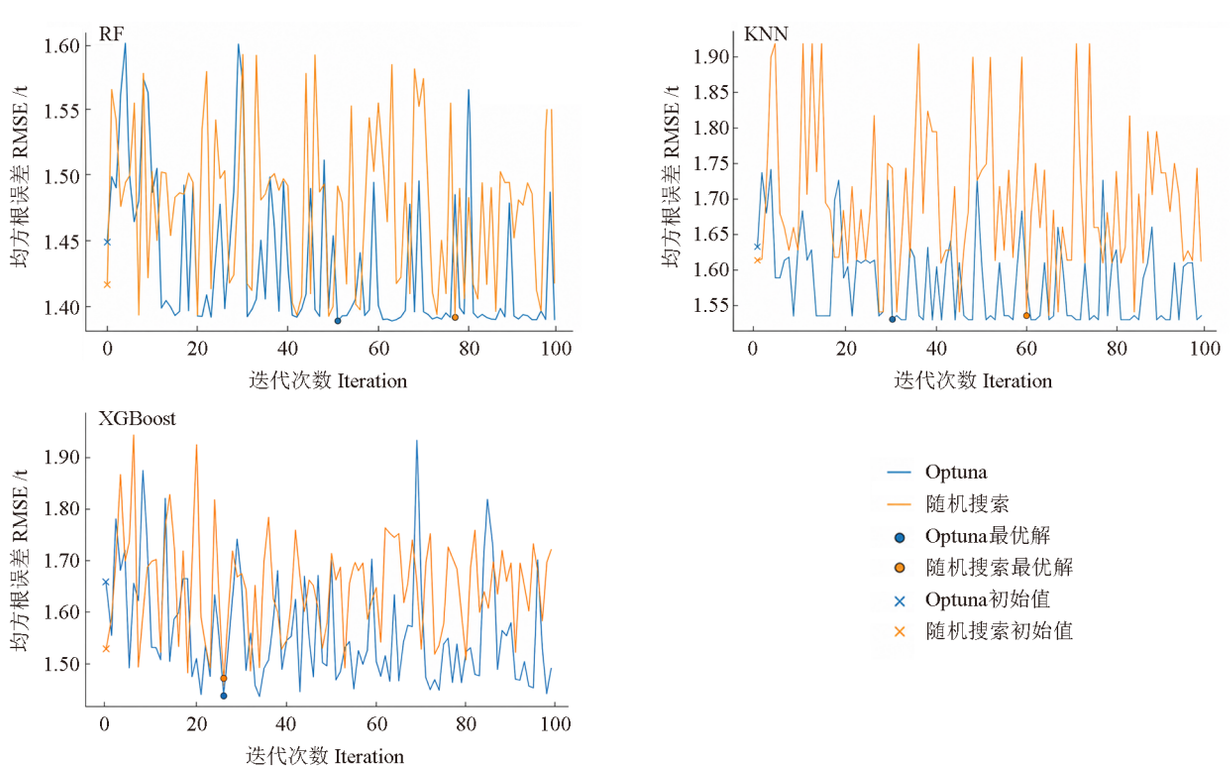


图 2　Optuna 与随机搜索在 RF、KNN 和 XGBoost 模型中的 RMSE 收敛特性对比
Fig. 2　Comparison of RMSE convergence characteristics between Optuna and random search for RF, KNN and XGBoost models

分比（13.13%），展示了其在同等收敛速度下更强的优化能力。

基于 TPE 算法的 Optuna 超参数优化框架的自动优化能力，本研究实现了特征变量最优子集与模型超参数的协同优化，显著提升了模型的预测性能。同时，利用 Optuna 内置的可视化工具，量化展示了关键超参数对模型性能的影响，有效识别关键超参数，提升调参效率，从而更有效地优化模型性能。

由表 5 可知，在 RF 模型中，叶子节点最小样本数对模型拟合的重要性最高，为 42.48%，同时自助采样设置为 True，有助于通过增加数据多样性来减少过拟合并提高模型的泛化能力，显著提高模型的稳定性和预测性能。邻居数量的设置对 KNN 模型的超参数贡献度最高，为 71.26%，说明该参数的合理设置对于 KNN 模型非常重要。对于 XGBoost 模型，L1 正则化项系数具有很高的超参数重要性，为 88.65%，证明在此模型的构建中，只需要调整该参数，就可以快速将模型达到最优性能，从而提升模型性能。

表 4　Optuna 与随机搜索在各模型超参数优化中的性能对比

Table 4　Performance comparison of Optuna and random search in hyperparameter optimization of multiple models

模型 Model	优化方法 Optimization method	最佳均方根误差 The best RMSE	收敛迭代次数 Number of convergence iteration	误差调整百分比 Error adjustment percentage /%	RMSE 相对优化 RMSE relative optimization /%	迭代效率提升 Iterative efficiency improvement /%
RF	Optuna	1.389 8	52/100	4.14	0.20	33.33
	Random search	1.392 7	78/100	1.75	—	—
KNN	Optuna	1.543 0	31/100	5.91	0.34	49.18
	Random search	1.548 3	61/100	4.56	—	—
XGBoost	Optuna	1.439 3	27/100	13.13	2.31	0
	Random search	1.473 4	27/100	3.68	—	—

表 5　模型超参数优化结果及重要性

Table 5　Model hyperparameter optimization results and importance

模型 Model	超参数 Hyperparameter	优化结果 Optimal result	重要性 Importance /%	模型 Model	超参数 Hyperparameter	优化结果 Optimal result	重要性 Importance /%
RF	学习器数量	190	8.25	XGBoost	学习器数量	434	2.34
	树的最大深度	7	6.96		树的最大深度	18	0.68
	节点最小样本数	2	4.68		学习率	0.09	3.12
	叶节点最小样本数	1	42.48		样本抽样比例	0.84	0.41
	最大特征数	None	5.25		决策树中抽样比例	0.95	0.35
	自助采样	True	32.37		最小分裂损失	0.01	3.12
KNN	加权机制	Distance	25.70		L1 正则化项系数	2.63	88.65
	邻居数量	7	71.26		L2 正则化项系数	0.01	1.89
	距离度量的阶数	1	3.04				

2.3　碳储量模型建立与评价

由表 6 可知，各模型在经过 Optuna-TPE 超参数调优后，性能均得到了显著提升。对于 XGBoost 模型，调优后 R^2 从 0.89 显著提升至 0.93，RMSE 从 0.46 t 降至 0.37 t，MAE 从 0.41 t 降至 0.33 t，表明该模型的预测精度和稳定性均有了明显提高。KNN 模型在调优后的效果尤为显著，R^2 从 0.72 上升至 0.88，RMSE 从 0.73 t 降至 0.47 t，MAE 从 0.61 t 降至 0.40 t，模型的泛化能力和准确性得到了增强。RF 模型调优后，R^2 从 0.86 提升至 0.90，RMSE 从 0.51 t 降至 0.47 t，MAE 从 0.41 t 降至 0.38 t，虽然提升幅度相对较小，但模型的预测性能也得到了改善。总体来看，调优工作有效提升

表 6　调优前后碳储量模型评价

Table 6　Evaluation of carbon stock models before and after tuning

模型 Model	指标 Index	优化前 Before optimization	优化后 After optimization
XGBoost	R^2	0.89	0.93
	RMSE /t	0.46	0.37
	MAE /t	0.41	0.33
KNN	R^2	0.72	0.88
	RMSE /t	0.73	0.47
	MAE /t	0.61	0.40
RF	R^2	0.86	0.90
	RMSE /t	0.51	0.47
	MAE /t	0.41	0.38

了所有模型的碳储量预测能力，其中 KNN 模型的优化效果最佳，XGBoost 模型在优化后展现出最高的预测精度，而 RF 模型则在优化后保持了较为稳定的性能提升。

从图 3 可以看出，所有模型的预测值与实际值都呈正相关，且数据大致沿着拟合线分布，说明模型具有较好的预测能力。其中，XGBoost 模型表现最佳。综合模型评价指标和模型结果拟合效果可以看出，模型优度表现为 XGBoost 模型＞RF 模型＞ KNN 模型。

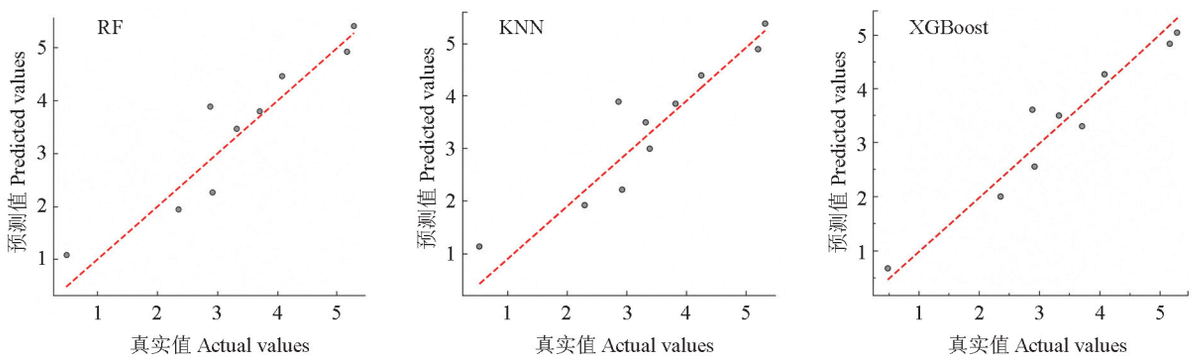


图 3　RF、KNN、XGBoost 模型碳储量实测值和预测值的散点图

Fig. 3　Scatter plots of measured and predicted carbon stock values for the RF, KNN, and XGBoost models

2.4　基于 SHAP 值的最优特征组合中特征变量的重要性评估

为研究各变量对模型的贡献程度，针对机器学习模型的“黑箱”特性，本研究基于 SHAP[34]（SHapley Additive exPlanations）来统一量化解释关键变量对不同模型预测结果的影响。从表 7 可以看出，不同模型对同一特征的重要性评估存在明显差异。例如，特征 A_{IH40} 在 RF 中的 SHAP 值为 0.35，在 KNN 中为 0.27，而在 XGBoost 中高达 0.64，这一差异表明 A_{IH40} 在 XGBoost 模型中具有更高的预测价值。相反，特征 H_{25} 在 XGBoost 中的 SHAP 值仅为 0.01，而在 RF 和 KNN 中的 SHAP 值分别为 0.28 和 0.20，说明该特征在 RF 和 KNN 模型中发挥的作用相对较大。此外，特征 I_{max} 在 KNN 中的 SHAP 值（0.33）显著高于 RF（0.10）和 XGBoost（0.22），可见 KNN 模型对这一特征的依赖性更强。总体来看，不同模型由于自身算法机制的不同，在捕捉特征与目标变量关系的方式上也有所不同，进而导致在特征选择和权重分配方面表现不同。

表 7　特征变量重要性 SHAP 值

Table 7　The SHAP value of the importance of characteristic variables

变量 Variables	SHAP 值 SHAP value		
	RF	KNN	XGBoost
A_{IH40}	0.35	0.27	0.64
A_{IH80}	0.30	0.31	0.14
H_{10}	0.26	0.27	0.36
H_{25}	0.28	0.20	0.01
I_{max}	0.10	0.33	0.22

3　讨论与结论

3.1　讨　论

3.1.1　基于 RFE-CV 的特征筛选与模型性能提升的关联性

研究发现，使用优选特征变量组合建立的模型性能显著优于使用全部变量建立的模型：XGBoost 模型的 R^2 从 0.78 提升至 0.89（提升 14.1%），随机森林模型的 R^2 从 0.71 提升至 0.86（提升 21.1%），验证了特征选择在降低维度噪声、增强模型泛化能力等方面的关键作用。这一结论与高宇航 [20] 和 Luo 等 [18] 的研究结果相一致，即合适的变量组合可以显著提高模型对森林碳储量估测的精度。这说明特征选择能够有效去除冗余信息，减少模型复杂度，提升模型的泛化能力和预测精度。在实际应用中，特征选择能够帮助研究者聚焦对目标变量具有重要影响的关键特征，从而提高模型的解释性和实用性。除此之外，通过 RFE-CV 筛选出 4 个高度变量（A_{IH40}、A_{IH80}、H_{10}、H_{25}）和 1 个强度变量（I_{max}），证明了高度变量在模型中的主导地位，这与穆喜云等 [35] 的研究结果相似，表明高度变量与森林碳储量的相关性较强。

3.1.2　以 RMSE 为目标函数的 Optuna-TPE 调优框架的适用性

与传统网格搜索或随机搜索相比，以 RMSE 为目标函数的 Optuna-TPE 调优框架的优势是其“自适应采样”特点 [21]，动态聚焦最优超参数区

域，更加适配湿地松 LiDAR 数据的特性。例如，XGBoost 模型的 RMSE 值降低了 2.31%（Optuna 误差调整百分比达随机搜索的 3.57 倍），RF 和 KNN 模型的收敛迭代次数分别减少 33.33% 和 49.18%。基于 TPE 算法的超参数重要性分析展示了关键超参数的贡献程度，为高效调优提供了理论支撑。XGBoost 模型对 L1 正则化项系数的高敏感性（重要性 88.65%）表明其依赖正则化约束缓解过拟合；KNN 模型则通过优化邻居数量使 R^2 提升 0.16，体现其在小样本高异质性场景中的潜力。这一结果与王可月等 [15] 和李海霞等 [23] 关于自动化调参可降低人为经验偏差的结论一致，但需结合模型特性差异化设计优化目标。比如 XGBoost，因集成学习机制更适合复杂非线性建模 [33]。此外，超参数优化效果在一定程度上依赖于目标函数的设定和搜索空间的定义。如果目标函数不恰当或搜索空间不合理，可能导致优化结果不理想。因此，本研究选取 RMSE 为目标函数，不仅是因为其对异常值敏感，更因其能适配湿地松机载 LiDAR 数据高维度与非均匀分布的特性，以降低估测偏差，这对湿地松森林碳储量的精确估测具有实践意义 [1]。同时，Optuna 在处理高维度、大规模数据时可能会面临计算资源和时间成本的挑战。正如 Bergstra 等 [21] 指出，Optuna 在高维空间中可能不如遗传算法等进化策略有效，尤其是当搜索空间中存在复杂的相互作用时。Yu[24] 在研究中也提到，面对大规模数据集时，Optuna 的计算成本可能会显著增加，从而限制其在实际应用中的效率。因此，在未来研究中需进一步探索混合式优化方法（如 Optuna 融合遗传算法、粒子群优化等）[25,36]，并遵循目标函数精准化、搜索空间结构化、算力评估可行性原则，以实现碳储量模型效能与计算成本的动态平衡。

3.1.3 模型选择与数据适配性

经特征筛选和超参数优化全流程优化后，各模型性能显著提升。XGBoost 模型的 R^2 达 0.93（提升 19.2%），RMSE 降至 0.37 t（降幅 41.3%），成为最优模型；KNN 模型的 R^2 提升 33.3%，实现较大改进；RF 模型保持稳健（MAE 降低 39.7%）。三大模型在优化后均展现良好性能，证实集成学习模型在碳储量预测中的优势，同时表明多种机器学习模型的适用性。这与 Wang 等 [37] 的结果一致，其融合无人机 LiDAR 与卫星遥感数据预测红树林碳储量时发现，XGBoost 在特征筛选与复杂地形建模中表现最优。傅志强 [38] 研究表明，XGBoost 算法在进行碳储量反演时精度最高。Cheng 等 [17] 研究显示，RF 在精度和可靠性方面优于其他算法。这可能与模型的算法原理和数据特征有关：XGBoost 的梯度提升机制适合处理湿地松机载 LiDAR 特征间的非线性关系；RF 则基于决策树的集成策略来实现对数据的处理和预测，适合处理具有复杂关系的数据。因此，在实际应用中，应根据研究对象的特性和数据分布选择合适的模型。对高维非线性数据（如融合激光雷达与多光谱特征）优先选择 XGBoost；当样本量有限或存在测量噪声时，RF 的随机空间策略更能保障泛化能力。此外，本研究仅依赖 LiDAR 数据，未考虑土壤条件、地形等，但湿地松的固碳能力受多种因子影响 [35]，未来可扩大样地范围，融合多源数据（如多光谱数据、土壤数据）来增强模型的生态学解释性，提升模型普适性。

3.1.4 基于 SHAP 值的特征重要性解释差异

本研究引入 SHAP[34] 来统一解释 RF、KNN 和 XGBoost 模型对湿地松碳储量的预测效果。通过计算优选特征变量对预测结果的边际贡献（SHAP 值），并取绝对平均值来量化其对森林碳储量估算的整体重要性：该值越高表明特征贡献度越大，对模型输出的影响越显著。SHAP 值量化分析表明，不同模型由于算法机制的不同（如 RF 的决策树集成、XGBoost 的梯度提升机制），对相同特征的依赖度存在显著差异。例如，特征 A_{IH40} 在 XGBoost 模型中具有更高的预测价值特征，而特征 H_{25} 在 RF 和 KNN 模型中的作用相对较大。这说明模型选择直接影响特征变量的贡献度，需结合算法特性解析碳储量驱动机制。这与 Chen 等 [39] 的研究结果一致，他们整合无人机和多光谱数据，通过 SHAP 值评估不同尺度下特征的重要性差异，为草原碳汇监测的多源数据融合提供了优化策略。此外，Li 等 [30] 利用 SHAP 值分析得出不同树种对特征的依赖性也存在差异：针叶林模型中，冠层高度的贡献更显著；阔叶林模型中，光谱反射率（如近红外波段）的贡献更高。因此，在实际应用中，选择合适的模型和理解其对特征的利用方式对于准确评估特征的重要性至关重要，未来可增加不同植被类型下的模型比较研究，以探索特征重要性在多尺度、多类型环境中的变化规律，为多物种森林碳储量估算的特征优化提供支持，为碳储量估算提供更具普适性和适应性的方法支撑。

3.2 结 论

本研究以湿地松人工林为对象，提出“RFE-CV 特征筛选与 Optuna-TPE 超参数优化”的协同框架，显著提升了湿地松人工林碳储量模型的估

测精度。结果表明，该框架能有效降维去噪、提升模型泛化性并显著优化模型性能。其中，优化后的 XGBoost 模型展现出最优性能，充分证实了集成学习算法在预测湿地松人工林碳储量上的优越性。此外，本研究发现高度变量在模型构建中占主导地位，5 个关键变量中有 4 个为高度相关变量。同时，基于 SHAP 值的分析结果表明，由于算法机制不同，模型对不同特征的响应存在显著差异，为明确湿地松碳汇的关键驱动特征提供了重要的理论与方法学依据。本研究所构建的全流程优化框架，不仅提升了湿地松碳储量估测的精度和效率，为湿地松人工林碳储量的精准估算提供了可靠的解决方案，也为相关森林参数反演及生态过程建模研究提供了参考性的概念框架。未来可进一步扩大样地覆盖范围（如纳入不同林龄、立地条件的林分），并融合多源数据以增强模型的生态学解释能力，从而进一步提升模型在区域尺度碳储量估测中的适用性。

参考文献：

[1] 张邦文 , 金苏蓉 , 邵虹 , 等 . 基于小班尺度的乔木林碳储量和碳密度估算 : 以宁都县为例 [J]. 南方林业科学 ,2023,51(4):31-35.
Zhang B W, Jin S R, Shao H, et al. Estimation of carbon storage and carbon density based on the subcompartment scale: a case study of Ningdu County, Jiangxi Province[J]. South China Forestry Science,2023,51(4):31-35.

[2] Koch A, Kaplan J O. Tropical forest restoration under future climate change[J]. Nature Climate Change,2022,12:279-283.

[3] Peduzzi P. The disaster risk, global change, and sustainability nexus[J]. Sustainability,2019,11:957.

[4] Auffhammer M, Gong Y Z. China’s carbon emissions from fossil fuels and market-based opportunities for control[J]. Annual Review of Resource Economics,2015,7:11-34.

[5] 王辽阔 , 何介南 . 湘北湿地松人工林生物量分配格局与碳汇潜力研究 [J]. 湖南林业科技 ,2025,52(2):10-18.
Wang L K, He J N. Study on biomass allocation pattern and carbon sink potential of *Pinus elliottii* plantation in northern Hunan[J]. Hunan Forestry Science & Technology,2025,52(2):10-18.

[6] Fei Q, Chen Y M, Ke X, et al. Somatic embryogenesis of slash pine (*Pinus elliottii* Engelm.): initiation, maturation, germination and mycorrhization of regenerated plantlets[J]. Plant Cell, Tissue and Organ Culture (PCTOC),2024,157:65.

[7] 张兴锐 , 齐国辉 , 陈利英 , 等 . 森林植被和土壤碳储量研究进展 [J]. 青海农林科技 ,2023(1):56-61.
Zhang X R, Qi G H, Chen L Y, et al. Research progress of forest vegetation and soil carbon storage[J]. Science and Technology of Qinghai Agriculture and Forestry,2023(1):56-61.

[8] Gibbs H K, Brown S, Niles J O, et al. Monitoring and estimating tropical forest carbon stocks: making REDD a reality[J]. Environmental Research Letters,2007,2:045023.

[9] Xie X K, Wang Q L, Dai L M, et al. Application of China’s national forest continuous inventory database[J]. Environmental Management,2011,48:1095-1106.

[10] 孙华 , 龙依 , 王储 , 等 . 森林储量遥感估测的不确定性研究进展 [J]. 中南林业科技大学学报 ,2026,46(2):1-12.
Sun H, Long Y, Wang C, et al. Research progress on uncertainty in remote sensing-based forest reserves estimation[J]. Journal of Central South University of Forestry & Technology, 2026,46(2):1-12.

[11] Xu W F, Cheng Y F, Luo M Y, et al. Progress and limitations in forest carbon stock estimation using remote sensing technologies: a comprehensive review[J]. Forests,2025,16:449.

[12] Gong Z, Ge W Y, Guo J Q, et al. Satellite remote sensing of vegetation phenology: progress, challenges, and opportunities[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2024, 217:149-164.

[13] 刘腾艳 . 遥感结合过程模型的浙江省森林碳储量时空演变研究 [D]. 杭州 : 浙江农林大学 ,2019:2-3.
Liu T Y. Temporal-spatial evolution of forest carbon stock in Zhejiang based on remote sensing combined with process model[D]. Hangzhou: Zhejiang A & F University,2019:2-3.

[14] 陈明 , 陈松 , 王彬彬 , 等 . 星载激光雷达足印点质量对获取森林冠层高度的影响 [J]. 中南林业科技大学学报 ,2026,46(3):197-207.
Chen M, Chen S, Wang B B, et al. The impact of spaceborne LiDAR footprint quality on canopy height extraction[J]. Journal of Central South University of Forestry & Technology,2026,46(3):197-207.

[15] 王可月 , 王轶夫 , 陈馨 , 等 . 基于集成学习算法和 Optuna 调优的江西省森林碳储量遥感估测 [J]. 生态学报 ,2025,45(2):685-700.
Wang K Y, Wang Y F, Chen X, et al. Remote sensing estimation of forest carbon storage in Jiangxi Province based on ensemble learning algorithm and Optuna tuning[J]. Acta Ecologica Sinica,2025,45(2):685-700.

[16] Jiang F G, Li M D, Chen S W. Adaptive Gaussian-PSO XGBoost model for Alpine forests aboveground biomass estimation using spaceborne PolSAR and LiDAR data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2025,18:10157-10171.

[17] Cheng F Y, Ou G L, Wang M, et al. Remote sensing estimation

of forest carbon stock based on machine learning algorithms[J]. Forests,2024,15:681.

[18] Luo M, Wang Y F, Xie Y H, et al. Combination of feature selection and CatBoost for prediction: the first application to the estimation of aboveground biomass[J]. Forests,2021,12:216.

[19] Masek J G, Collatz G J. Estimating forest carbon fluxes in a disturbed southeastern landscape: integration of remote sensing, forest inventory, and biogeochemical modeling[J]. Journal of Geophysical Research: Biogeosciences,2006,111:G01006.

[20] 高宇航 . 基于 Landsat 8 和机器学习的高峰林场桉树地上碳储量模型研究 [D]. 北京 : 北京林业大学 ,2022:51-53.
Gao Y H. Study on aboveground carbon storage model of *Eucalyptus* in Gaofeng forest farm based on Landsat 8 and machine learning[D]. Beijing: Beijing Forestry University,2022: 51-53.

[21] Bergstra J, Bengio Y. Random search for hyper-parameter optimization[J]. Journal of Machine Learning Research,2012,13: 281-305.

[22] Bergstra J, Bardenet R, Bengio Y, et al. Algorithms for hyper-parameter optimization[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems,ACM,2011:2546-2554.

[23] 李海霞 , 宋丹蕾 , 孔佳宁 , 等 . 传统机器学习模型的超参数优化技术评估 [J]. 计算机科学 ,2024,51(8):242-255.
Li H X, Song D L, Kong J N, et al. Evaluation of hyperparameter optimization techniques for traditional machine learning models[J]. Computer Science,2024,51(8):242-255.

[24] Yu T C. Optuna hyperparameters optimization for discriminating life-threatening ventricular arrhythmia problem[D]. New York: State University of New York at Buffalo,2023:1-10.

[25] Lin L, Zhang J, Zhang N, et al. Optimized LightGBM power fingerprint identification based on entropy features[J]. Entropy,2022,24:1558.

[26] Hyndman R J, Koehler A B. Another look at measures of forecast accuracy[J]. International Journal of Forecasting,2006, 22(4):679-688.

[27] Poorazimy M, Shataee S, Mcroberts R E, et al. Integrating airborne laser scanning data, space-borne radar data and digital aerial imagery to estimate aboveground carbon stock in Hyrcanian forests, Iran[J]. Remote Sensing of Environment,2020, 240:111669.

[28] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks[J]. Information Processing & Management,2009,45:427-437.

[29] GB/T 43648—2024 主要树种立木生物量模型与碳计量参数 [S]. GB/T 43648-2024 Tree biomass models and related parameters to carbon accounting for major tree species[S].

[30] Li Q, Yin J, Zhang X X, et al. Tree-level carbon stock estimations across diverse species using multi-source remote sensing integration[J]. Computers and Electronics in Agriculture,2025,231:109904.

[31] Sarkar S, Sagan V, Bhadra S, et al. Soybean seed composition prediction from standing crops using Planet Scope satellite imagery and machine learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2023,204:257-274.

[32] Fu Y Y, He H S, Hawbaker T J, et al. Evaluating k-nearest neighbor (kNN) Imputation models for species-level aboveground forest biomass mapping in northeast China[J]. Remote Sensing, 2019,11:2005.

[33] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM,2016:785-794.

[34] Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions[J]. Knowledge and Information Systems,2014,41:647-665.

[35] 穆喜云 , 刘清旺 , 庞勇 , 等 . 基于机载激光雷达的森林地上碳储量估测 [J]. 东北林业大学学报 ,2016,44(11):52-56.
Mu X Y, Liu Q W, Pang Y, et al. Forest aboveground carbon storage using RF algorithmic model and airborne LiDAR data[J]. Journal of Northeast Forestry University,2016,44(11):52-56.

[36] Chinembiri T S, Mutanga O, Dube T. Carbon stock prediction in managed forest ecosystems using Bayesian and frequentist geostatistical techniques and new generation remote sensing metrics[J]. Remote Sensing,2023,15:1649.

[37] Wang Z Y, Zhang Y, Li F L, et al. Regional mangrove vegetation carbon stocks predicted integrating UAV-LiDAR and satellite data[J]. Journal of Environmental Management,2024,368:122101.

[38] 傅志强 . 基于机器学习的林地碳储量遥感反演 [D]. 南京 : 南京信息工程大学 ,2024:29-41.
Fu Z Q. Remote sensing inversion of forest land carbon storage based on machine learning[D]. Nanjing: Nanjing University of Information Science & Technology,2024:29-41.

[39] Chen A, Xu C, Zhang M, et al. Cross-scale mapping of above-ground biomass and shrub dominance by integrating UAV and satellite data in temperate grassland[J]. Remote Sensing of Environment,2024,304:114024.

[本文编校：戴欧琳]