

引用格式: 王彤, 张超, 周杭. 基于 Landsat 影像的曲靖市主要针叶林地上生物量估算[J]. 浙江农林大学学报, 2025, 42(6): 1132–1141. WANG Tong, ZHANG Chao, ZHOU Hang. Aboveground biomass estimation of major coniferous forests in Qujing City based on Landsat images[J]. Journal of Zhejiang A&F University, 2025, 42(6): 1132–1141.

## 基于 Landsat 影像的曲靖市主要针叶林地上生物量估算

王彤<sup>1</sup>, 张超<sup>1,2</sup>, 周杭<sup>1</sup>

(1. 西南林业大学 林学院, 云南 昆明 650233; 2. 西南林业大学 云南省山地农村生态环境演变与污染治理重点实验室, 云南 昆明 650233)

**摘要:** 【目的】森林生物量是衡量森林生态系统生产力和碳储存能力的重要指标, 研究其动态变化有助于深入理解森林生态系统功能, 为生态管理与碳循环研究提供科学依据。【方法】以云南省曲靖市为研究区, 基于 1992—2017 年以 5 a 为 1 期的一类连续清查样地数据, 结合数字高程模型 (DEM) 与 Landsat 5 TM、Landsat 8 OLI 影像, 构建地上生物量遥感估测模型。针对主要针叶树种 (云南松 *Pinus yunnanensis* 林、华山松 *P. armandii* 林及针叶混交林), 对比梯度提升回归树 (GBRT)、随机森林 (RF)、极端梯度提升 (XGBoost)、偏最小二乘法 (PLS) 和分类器梯度提升 (CatBoost) 等 5 种机器学习方法, 系统评估其拟合性能与估测精度。通过调整决定系数 ( $R^2$ )、均方根误差 (RMSE)、相对均方根误差 (rRMSE) 与平均绝对误差 (MAE) 等指标综合评价, 优选最优模型, 并用于 1992—2022 年曲靖市主要针叶林地上生物量的时序估算。【结果】不同模型在 3 类优势针叶树种上的估测性能存在差异。CatBoost 模型在针叶混交林中表现最优 ( $R^2=0.90$ , RMSE=5.66), 具有较强的非线性拟合能力; 而 GBRT 模型在云南松林与华山松林中表现更佳, 估测稳定性与解释力较强。【结论】不同机器学习模型在森林地上生物量估测中的表现因树种而异。模型选择应结合树种生长特性及其对遥感因子的响应差异, 避免统一建模策略, 以提升估测精度与适用性。图 3 参 39

**关键词:** Landsat 影像; 森林地上生物量; 机器学习; 估算模型; 曲靖市

中图分类号: S75 文献标志码: A 文章编号: 2095-0756(2025)06-1132-10

## Aboveground biomass estimation of major coniferous forests in Qujing City based on Landsat images

WANG Tong<sup>1</sup>, ZHANG Chao<sup>1,2</sup>, ZHOU Hang<sup>1</sup>

(1. College of Forestry, Southwest Forestry University, Kunming 650233, Yunnan, China; 2. Yunnan Key Laboratory of Ecological Environment Evolution and Pollution Control in Mountainous Rural Areas, Southwest Forestry University, Kunming 650233, Yunnan, China)

**Abstract:** [Objective] Forest biomass is a crucial indicator of the productivity and carbon storage capacity of forest ecosystems. Studying its dynamic changes can facilitate a profound understanding of the functions of forest ecosystems and offer a scientific basis for ecological management and carbon cycle research. [Method] Taking Qujing City of Yunnan Province as the study area, a remote sensing estimation model of aboveground biomass was constructed using data from one type of continuous clearing sample plots in every 5-year period from 1992 to 2017, in combination with DEM and Landsat 5 TM, Landsat 8 OLI images. For the main conifer species (*Pinus yunnanensis*, *P. armandii* and mixed conifer forests), five machine learning methods, namely GBRT, RF, XGBoost, PLS and CatBoost, were compared to systematically evaluate their fitting performance and estimation accuracy. The optimal model was selected through the comprehensive

收稿日期: 2025-07-27; 修回日期: 2025-10-29

基金项目: 国家自然科学基金资助项目 (32160405)

作者简介: 王彤 (ORCID: 0009-0003-4094-5164), 从事森林经理学研究。E-mail: [2076704401@qq.com](mailto:2076704401@qq.com)。通信作者: 张超 (ORCID: 0000-0003-3126-6103), 从事森林经理学研究。E-mail: [42668600@qq.com](mailto:42668600@qq.com)

evaluation of  $R^2$ , RMSE, rRMSE and MAE indicators, and used for the time series estimation of aboveground biomass of major coniferous forests in Qujing City from 1992 to 2022. [Result] The estimation performance of different models on the three dominant coniferous species varied: CatBoost model performed best in mixed coniferous forests ( $R^2=0.90$ , RMSE=5.66), with strong nonlinear fitting ability; while GBRT model performed better in *P. yunnanensis* forests and *P. armandii* forests, with stronger estimation stability and explanatory power. [Conclusion] This study indicates that the performance of different machine learning models in forest aboveground biomass estimation varies according to tree species. Model selection should consider the growth characteristics of tree species and their response differences to remote sensing factors, and avoid uniform modelling strategies to improve the estimation accuracy and applicability. [Ch, 3 fig. 39 ref.]

**Key words:** Landsat images; forest aboveground biomass; machine learning; estimation model; Qujing City

森林是地球上分布最广、结构最复杂的陆地生态系统，具备水土保持、生态平衡维持等多种生态功能<sup>[1-2]</sup>，是全球最重要的陆地碳汇和碳储库之一<sup>[3-4]</sup>，在气候调节、生态系统稳定以及“双碳”目标的实现过程中发挥着关键作用<sup>[5]</sup>。森林生物量作为森林生态系统的关键属性，承载着重要的生态功能信息，不仅构成其运行的能量基础和营养物质来源<sup>[6]</sup>，也是评估森林生态系统生产力的关键参数。森林地上生物量 (AGB) 是量化森林生态系统固碳能力的关键生态参数，其动态变化直接反映森林碳循环特征和初级生产力水平。作为表征森林生态系统结构与功能的核心指标，AGB 的精确估算及时空格局分析已成为全球气候变化研究和生态系统可持续管理领域的重要研究方向<sup>[7]</sup>。传统森林地上生物量的样地调查方法耗时费力，而遥感技术凭借其光谱信息与植被参数同 AGB 的强相关性，以及大范围、连续观测的优势，已成为区域至全球尺度森林生物量估算的重要手段，显著提升了数据获取效率并为碳储量监测提供技术支撑<sup>[8-9]</sup>。

Landsat 系列卫星作为全球首个长期运行的陆地观测卫星系统，自 1972 年首次发射以来，以其独特的观测优势在森林资源监测领域发挥着不可替代的作用。在区域尺度森林监测中具有广泛适用性，包括森林资源清查 (如覆盖制图、树种识别等)，动态变化监测 (如采伐迹地识别、退化评估等)，以及生物量动态反演与碳储量精确估算<sup>[10-11]</sup>。国庆喜等<sup>[12]</sup>基于专题制图仪 (TM) 遥感影像与森林资源清查样地数据，构建了多元回归模型，实现了对大兴安岭南坡地区森林生物量的反演与空间分布估算。杨伟志等<sup>[13]</sup>利用陆地成像仪 (OLI) 遥感影像提取相关遥感因子，采用逐步回归方法构建了西宁市南北山地区的森林生物量估算模型。邱布布等<sup>[14]</sup>基于杭州市 199 个样地实测数据，利用 Landsat 8 OLI 影像提取原始波段与植被指数信息，构建了绿地地上生物量估算模型，并反演区域生物量，结合同期 Landsat 7 ETM+影像结果进行对比分析。

在复杂林型和多时相环境下，如何实现区域 AGB 的高精度估算，并揭示其长期时空变化规律，是区域森林资源监测与碳储量核算的关键科学问题。本研究基于 1992—2017 年 5 a 为 1 期的一类连续清查样地数据，结合 Landsat 5 TM 与 Landsat 8 OLI 多时相遥感影像及数字高程模型 (DEM)、纹理特征和植被指数等因子，构建梯度提升回归树 (GBRT)、随机森林 (RF)、极端梯度提升 (XGBoost)、偏最小二乘法 (PLS) 和分类器梯度提升 (CatBoost) 等多种机器学习模型，系统评估不同模型及因子在 AGB 估算精度与空间异质性解释上的适用性，估测曲靖市主要针叶林地上生物量，优选性能最佳模型，估算 1992—2022 年生物量的时空格局，为森林碳储量精准监测提供技术支撑。

## 1 材料与方法

### 1.1 研究区概况

云南省曲靖市位于 24°21'00"~27°03'24"N，103°03'44"~104°32'36"E，土地总面积约 28 900.0 km<sup>2</sup>，平均海拔约 1 881.0 m。属亚热带季风气候，年平均气温为 14.8 ℃，年均降水量为 560.0~1 026.0 mm。主要植被类型为亚热带针叶林和常绿阔叶林，原生植被多已退化，现以天然次生植被及人工林为主。曲靖市森林资源较为丰富，2021 年森林覆盖率达 50.5%。优势树种主要包括云南松 *Pinus yunnanensis*、杉木

*Cunninghamia lanceolata*、华山松 *Pinus armandii*、栎类 *Quercus* spp. (如麻栎 *Quercus acutissima*、青冈 *Quercus glauca* 等) 及油茶 *Camellia oleifera* 等。

## 1.2 数据收集与处理

获取云南省曲靖市 1992—2017 年 6 期 (5 a 为 1 期) 森林资源连续清查数据。云南省森林资源连续清查采用系统抽样方法<sup>[15]</sup>, 抽样间距为  $6\text{ km} \times 8\text{ km}$ , 方形样地面积为  $0.06\text{ hm}^2$ 。以曲靖市优势树种为研究对象, 筛选出 3 类主要针叶林: 云南松林、华山松林和针叶混交林, 共计 417 个样地 (其中云南松林 264 个、华山松林 83 个、针叶混交林 70 个)。以样地蓄积量为基础计算森林生物量<sup>[16]</sup>, 蓄积量由一类连续清查数据提供, 依据样木胸径 (DBH)、树高及现有生长量方程计算得到。通过蓄积量-生物量模型<sup>[17]</sup> 计算得到森林生物量为实测数据。计算公式为:  $B=aV+b$ 。其中:  $B$  为森林生物量 ( $\text{t} \cdot \text{hm}^{-2}$ );  $V$  为森林蓄积量 ( $\text{m}^3 \cdot \text{hm}^{-2}$ );  $a$ 、 $b$  为蓄积量与生物量之间的模型转换参数<sup>[18]</sup>。

获取 2007 年云南省曲靖市森林资源二类调查数据<sup>[19]</sup>, 用于确定后续生物量制图范围。一类连续清查样地数据经几何校正后, 重采样至  $30\text{ m}$  以匹配遥感影像。遥感数据来源于谷歌地球引擎 (GEE) 平台的 Landsat TM/OLI 影像 (Collection 2 Tier 1), 时间范围为 1992—2022 年, 每年 1 月 1 日至 12 月 31 日的影像, 空间分辨率为  $30\text{ m}$ 。该数据集已完成大气校正、几何校正和地形校正, 经去云处理与时间序列插值, 最终采用中值合成生成年度影像。地形数据来自地理空间数据云的 DEM (空间分辨率为  $30\text{ m}$ )。

## 1.3 特征提取

基于 Landsat 影像的光谱响应、冠层几何特征及灰度共生矩阵 (GLCM) 描述的林分空间异质性, 提取与 AGB 显著相关的遥感因子。以 Landsat 5 TM 影像为例, 包括①单波段因子<sup>[20]</sup>: 选取 TM1~TM5 和 TM7 共 6 个原始波段作为单波段因子, TM1 区分土壤与植被, TM2 区分植被类型, TM3 (红光) 反映叶绿素吸收与冠层结构, TM4 (近红外光) 对叶片结构和水分敏感, 是生物量估算基础, TM5 (短波红外光) 反映干物质积累并辅助判别岩土, TM7 (短波红外光) 用于识别岩石与植被胁迫; ②波段组合因子: 比值指数 (如 TM42、TM53、TM73 等) 可有效削弱地形光照干扰, 增强植被信号, 显著提升森林生物量反演精度; ③DEM 地形因子: 坡度、坡向和高程, 可调节光合有效辐射 (PAR) 并作为温度与水的综合代理, 从而间接影响生物量的空间分布。④纹理特征: 对 6 个波段分别在  $3 \times 3$ 、 $5 \times 5$  和  $7 \times 7$  等 3 种窗口下提取 GLCM 的 8 类指标<sup>[16]</sup>, 包括均值、方差、均一性、对比度、相关性、相异性、角二矩和熵, 纹理指标如均值、对比度和熵可量化林分内部树冠大小、郁闭度及空间复杂性, 且已被证实与 AGB 高度相关, 最终提取 144 个纹理变量; ⑤植被指数涵盖多种典型指标<sup>[21]</sup>: 如归一化植被指数 (NDVI)、增强型植被指数 (EVI)、土壤调节植被指数 (SAVI)、垂直植被指数 (PVI)、土壤调整比值植被指数 (SARV) 等, 其中 NDVI 和 EVI 与叶面积指数 (LAI) 及光合有效辐射吸收比例 (fAPAR) 密切相关, 广泛用于 AGB 反演。⑥主成分变换因子: 对原始 6 个波段进行 K-L 变换, 保留信息量最大的前 3 个主成分 (PCA1、PCA2、PCA3), 在降维的同时保留与生物量变化相关的综合光谱信息, 从而提高模型的稳健性。变量命名规则: b1~b7 表示 Landsat TM 影像的单波段因子; 纹理特征命名规则中, 数字前缀表示窗口大小, 字母及后缀表示所用波段和纹理指标类型, 例如 R3Hom2 表示  $3 \times 3$  窗口下第 2 波段的均一性 (homogeneity, Hom) 纹理因子。

## 1.4 特征筛选

为降低高维特征带来的计算负担与过拟合风险, 提升 AGB 估测的精度与稳定性, 研究在初步提取 187 个遥感特征变量基础上, 采用二阶段特征选择策略进行优化。首先, 通过皮尔逊 (Pearson) 相关性分析筛选与 AGB 显著相关 ( $P < 0.05$ ) 的变量, 剔除无效信息; 随后结合 RF 的特征重要性, 利用递归特征消除 (RFE) 算法优选 10 个关键变量用于建模<sup>[22]</sup>。

## 1.5 估算模型选择

1.5.1 RF 模型 RF 算法是 BREIMAN<sup>[23]</sup> 于 2001 年提出的一种集成学习方法, 由多个决策树构成。该模型通过在训练过程中对样本和特征进行随机抽样, 构建多个相互独立的决策树, 并最终通过集成 (如投票或平均) 各棵树的预测结果, 提升模型的整体预测精度和稳定性。作为一种非参数模型, RF 具有出色的抗过拟合能力、良好的抗噪声性能和较强的泛化能力, 同时支持并行计算, 在计算效率和预测精度



方面均表现优异,在回归任务中,RF 通过多棵回归树的集成,有效提高对连续变量的建模能力<sup>[24]</sup>。

**1.5.2 XGBoost 模型** XGBoost 算法是 CHEN 等<sup>[25]</sup>在传统梯度提升方法的基础上提出的一种高效的集成学习算法。在模型结构与计算机制上均进行优化的增强型梯度提升算法,具备显著的训练效率与预测性能优势,尤其适用于处理大规模、高维度的数据集。在传统梯度提升决策树 (GBDT) 框架的基础上引入正则化项,有效控制模型复杂度,从而缓解过拟合问题。同时, XGBoost 采用基于二阶导数的信息进行损失函数的优化更新,进一步提升了模型的拟合精度与收敛速度。

**1.5.3 GBRT 模型** GBRT 算法与 RF 类似,是一种基于集成学习思想的回归建模方法,由 FRIEDMAN<sup>[26]</sup>首次提出,并在后续研究中不断优化与拓展<sup>[27]</sup>。GBRT 通过集成弱回归树并逐步拟合残差,基于损失函数梯度优化预测性能。相比随机森林,其更适用于处理非线性、高维和复杂特征关系,且对异常值具有鲁棒性。在遥感估测、生物量反演及碳储量建模等复杂生态问题中, GBRT 具有优越的预测与泛化能力。

**1.5.4 PLS 模型** PLS 算法是一种结合主成分分析 (PCA) 和多元线性回归思想的多变量建模方法。不仅保留主成分分析在降维处理中的优势,有效解决自变量之间多重共线性问题,还进一步考虑各主成分与因变量之间的相关性,在提取特征时同时最大化自变量与因变量之间的协方差,从而提高模型的预测能力<sup>[28]</sup>。

**1.5.5 Catboost 模型** CatBoost 是一种基于梯度提升决策树的高效机器学习框架,针对类别特征做出优化,无需手动编码,自动高效处理类别变量,提升建模效率与预测精度。采用有序目标统计方法和先验项,缓解目标泄露引起的过拟合;引入对称树结构和特有损失函数优化,提高训练速度与稳定性。核心算法基于梯度提升框架,结合有序 Boosting 策略和类别特征编码,调整样本顺序敏感性。CatBoost 凭借优异的泛化能力和计算性能,在处理复杂结构、类别变量多和数据分布不均的问题中表现突出。

**1.5.6 模型参数** 将优选的特征变量带入模型的构建,采用随机抽样的方法,以其中 80% 的样地数据作为训练数据,其余 20% 的样地数据作为检验数据。除 PLS 外,其他模型均采用网格搜索结合五折交叉验证进行超参数调优,通过遍历参数空间,筛选最优参数,确保模型具备良好的鲁棒性和泛化能力。此外,为加快模型训练与参数优化过程,所有可并行算法均设置  $n\_jobs=-1$ ,充分调用多线程计算资源。依据各模型算法特点和树种数据特征设置合理的参数范围。GBRT 模型主要调节基础学习器数量 ( $n\_estimators$ ); RF 模型进一步考虑最大树深 ( $max\_depth$ ) 与节点最小分裂样本数 ( $min\_samples\_split$ ); XGBoost 模型引入学习率 ( $learning\_rate$ ) 与树结构复杂度控制; PLS 模型采用固定成分数 ( $n\_components=5$ ), CatBoost 模型则基于迭代次数 ( $iterations$ ) 进行性能优化。

## 1.6 精度评价

采用多种评价指标对模型拟合效果进行综合评估,包括调整决定系数 ( $R^2$ )、相对均方根误差 ( $rRMSE$ )、均方根误差 ( $RMSE$ )<sup>[29]</sup>以及平均绝对误差 ( $MAE$ )<sup>[30-31]</sup>。其中  $R^2$  用于衡量解释能力并校正自变量影响;  $RMSE$  反映预测误差大小;  $rRMSE$  为标准化指标,便于不同模型误差比较;  $MAE$  表示所有预测误差的绝对值的平均值,能直观反映模型预测值与真实值之间的平均偏离程度<sup>[31]</sup>。

## 2 结果与分析

### 2.1 遥感特征变量的相关性分析

基于样地数据,采用 Pearson 相关性分析对 187 个遥感特征变量与实测森林地上生物量 AGB 之间的相关性进行检验,剔除  $P \geq 0.05$  的变量,初步筛选出显著相关因子。在此基础上,引入基于 RFE 进行进一步特征优化。结果显示 (图 1): 针叶混交林权重最高的变量为  $R7Hom2$ 、 $b2$ 、 $R7Sec2$  等; 云南松林为  $aspect$ 、 $slope$ 、 $ND54$  等; 华山松林为  $ND53$ 、 $TM54$ 、 $ARVI$  等,不同林分间特征变量组合存在差异。

### 2.2 生物量建模精度评价

如图 2 所示: GBRT 模型在针叶混交林、云南松林和华山松林 3 种林型的 AGB 估测中均表现良好 ( $R^2 > 0.77$ ), 其中华山松林拟合最佳 ( $R^2=0.82$ ), 云南松林误差最小 ( $RMSE=8.30 t \cdot hm^{-2}$ ); RF 模型对针阔混交林、云南松林和华山松林等 3 种林型的预测精度略低于 GBRT,  $R^2$  分别为 0.71、0.62、0.74,  $RMSE$  分别为 9.64、11.57 和  $11.27 t \cdot hm^{-2}$ ; XGBoost 模型中, 针叶混交林  $R^2=0.73$ , 估测趋势较好, 云南松林  $R^2=0.61$ 、 $RMSE=11.75$ , 部分样点偏差较大, 华山松林拟合略优但在高值区出现高低估交叉; PLS 模型的 3 种林型  $R^2$  普遍偏低, 拟合能力较差, 变异系数分别为 0.36 (针叶混交林)、0.28 (云南松

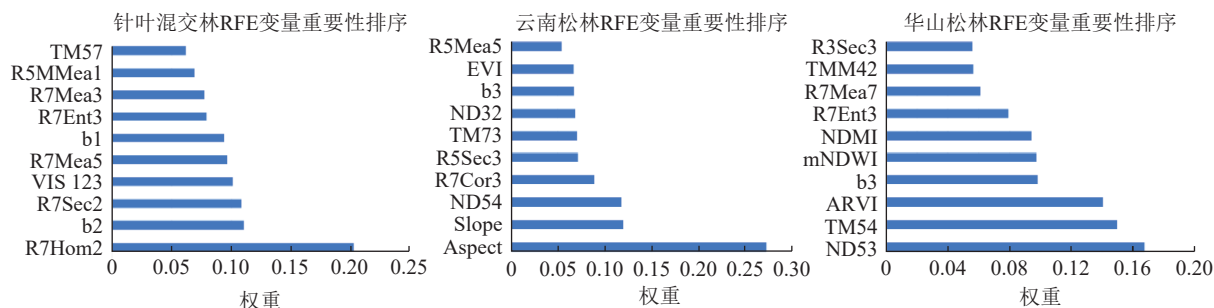


图 1 3 种针叶林遥感特征变量相关性分析

Figure 1 Correlation analysis of remote sensing characteristic variables of 3 coniferous forests

林)和 0.50 (华山松林), 表明生物量差异明显; CatBoost 模型在针叶混交林表现最佳 ( $R^2=0.90$ , RMSE=5.66), 拟合精度高, 散点集中, 优于其他模型, 云南松林与华山松林  $R^2$  分别为 0.60 和 0.69, RMSE 为 11.88 和 12.37  $\text{t} \cdot \text{hm}^{-2}$ 。模型差异可能受样地变异性、遥感特征质量及样本代表性等因素影响。综上所述, 模型适应性因树种而异, CatBoost 在针叶混交林中表现最佳 ( $R^2=0.90$ , RMSE=5.66), 具备优异的非线性拟合能力; 云南松林和华山松林则以 GBRT 效果最佳, 稳定且解释力强。可见, 模型选择需结合树种特性和数据结构, 优选最优算法以提升生物量反演准确性和可靠性。综上, CatBoost 适用于针叶混交林地上生物量估算, GBRT 更适合云南松林与华山松林的遥感建模与全域逐像元反演制图。

### 2.3 地上生物量估算结果

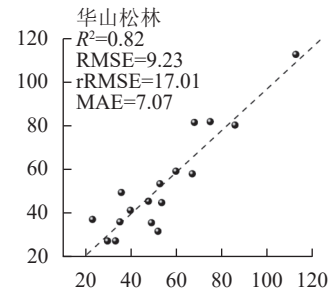
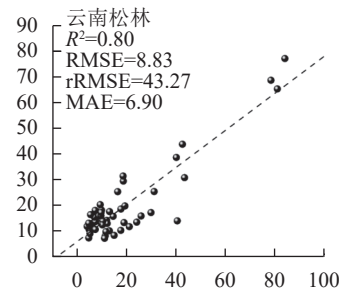
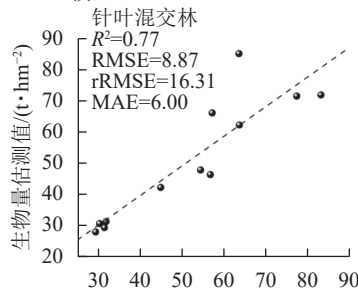
基于 Landsat 影像数据, 采用 CatBoost 和 GBRT 模型对曲靖市主要针叶林地上生物量进行 30 m×30 m 分辨率的空间估算。如图 3 所示: 将森林生物量分为低生物量区 ( $1.78 \sim 20.00 \text{ t} \cdot \text{hm}^{-2}$ )、中低生物量区 ( $> 20.00 \sim 40.00 \text{ t} \cdot \text{hm}^{-2}$ )、中等生物量区 ( $> 40.00 \sim 60.00 \text{ t} \cdot \text{hm}^{-2}$ )、中高生物量区 ( $> 60.00 \sim 80.00 \text{ t} \cdot \text{hm}^{-2}$ ) 及高生物量区 ( $> 80.00 \text{ t} \cdot \text{hm}^{-2}$ )。1992—2022 年, 曲靖市针叶林地上生物量呈持续增长态势, 空间分布格局显著优化。1992 年, 区域生物量集中于  $20.00 \sim 40.00 \text{ t} \cdot \text{hm}^{-2}$  及以下, 高生物量区分布零散, 反映当时森林质量较差, 受林地退化和过度采伐等人为干扰影响明显。1997 年, 东南部部分区域生物量回升,  $40.00 \sim 60.00 \text{ t} \cdot \text{hm}^{-2}$  区比例有所增加, 表明中幼龄林逐步成熟, 林分结构改善及天然林保护政策的实施促进碳储量提升。2002 年以后, 生物量增长加速, 中高生物量区 ( $> 60.00 \text{ t} \cdot \text{hm}^{-2}$ ) 开始连片分布, 低生物量区面积明显缩减, 显示封山育林和人工造林等生态修复措施取得显著成效。2007 年, 高生物量区 ( $> 80.00 \text{ t} \cdot \text{hm}^{-2}$ ) 面积大幅扩展, 集中于西部和南部高山区, 生态系统生产力显著增强, 该阶段变化得益于“退耕还林”和“生态公益林”等国家重点生态工程的持续推进。2012 年, 生物量分布趋于均衡, 区域内部差异缩小, 人工林进入成材期, 天然林保护成果逐步显现。2017 年, 生物量整体继续提升, 高生物量区范围进一步扩大, 尤其在西部和南部山区呈现更为明显的连片分布, 低生物量区进一步收缩, 显示生态恢复效果持续深化, 森林碳汇能力不断增强。至 2022 年, 针叶林地上生物量达到历史最高水平, 高生物量区在西部及南部山区形成大面积连片, 区域碳储能力显著提升。相比 1992 年, 生物量空间格局由“低量为主、斑块状高量”转变为“中高量为主、分布连续均衡”。

## 3 讨论

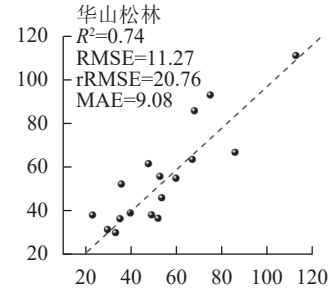
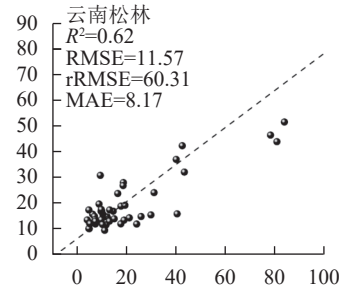
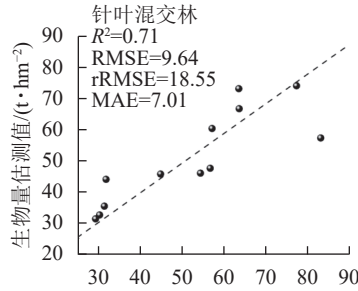
### 3.1 不同林型的模型表现与“林型-算法”匹配

本研究表明: CatBoost 模型在针叶混交林中估算性能最优 ( $R^2=0.90$ ), 该结果与 LUO 等<sup>[32]</sup>在吉林省基于 RFE 特征选择与 CatBoost 的 AGB 估测结论一致, CatBoost 的精度均优于 XGBoost 和 RF 等估测模型。CatBoost 模型能够自适应学习高维非线性特征, 并有效抑制过拟合, 尤其适用于针叶混交林冠层结构复杂和物种组成异质性高的场景<sup>[32]</sup>。在结构相对单一、冠层性状较一致的云南松林和华山松林中, GBRT 模型表现出更稳健的估算能力,  $R^2$  分别达 0.80 和 0.82, 此类林分中林木个体在高度、胸径和冠幅等表型特征差异较小, 导致遥感信号在像元尺度上内部异质性低、统计分布平稳, 有利于提升模型拟合效果。这一发现与 ZHANG 等<sup>[33]</sup>在结构单一的高山松 *Pinus densata* 林中研究结果一致, 其 GBRT 模型  $R^2$  达 0.94, 显著优于其他算法, 验证 GBRT 在结构一致性高的针叶林中具有更强的稳健性与泛化能力。进一步验证了“林型-算法”适配策略的有效性, 针对特定林型结构特征选择适宜的机器学习算法, 可

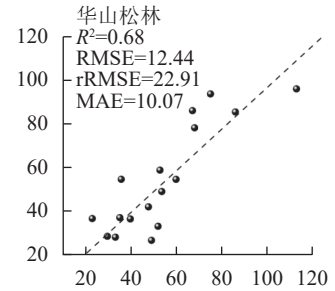
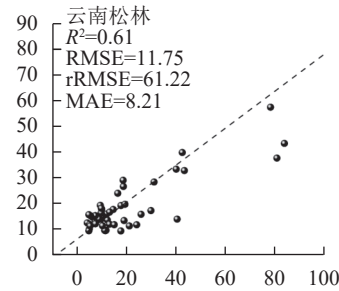
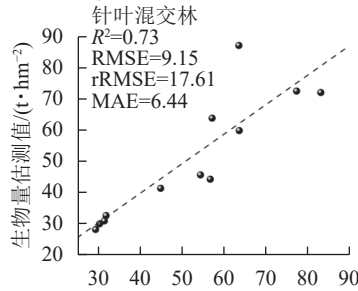
## A. GBRT模型



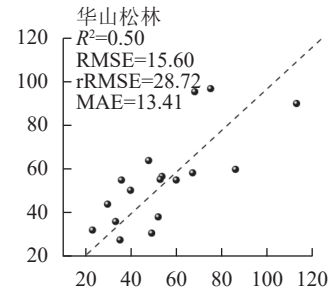
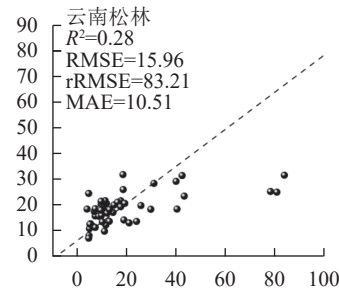
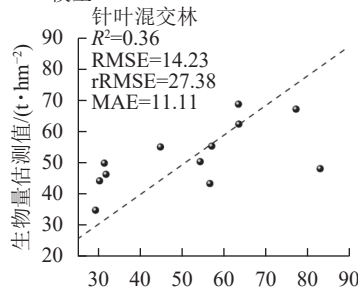
## B. RF模型



## C. XGBoost模型



## D. PLS模型



## E. CatBoost模型

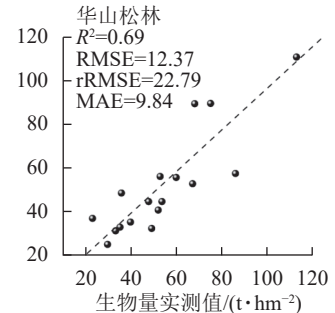
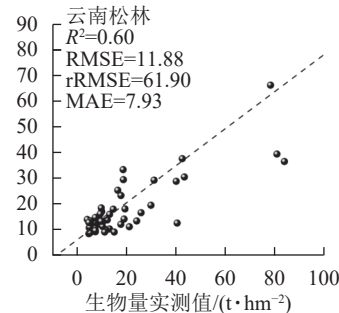
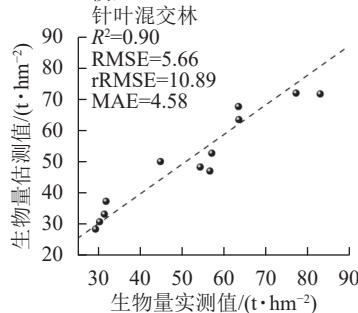


图 2 生物量估测各模型结果散点分布

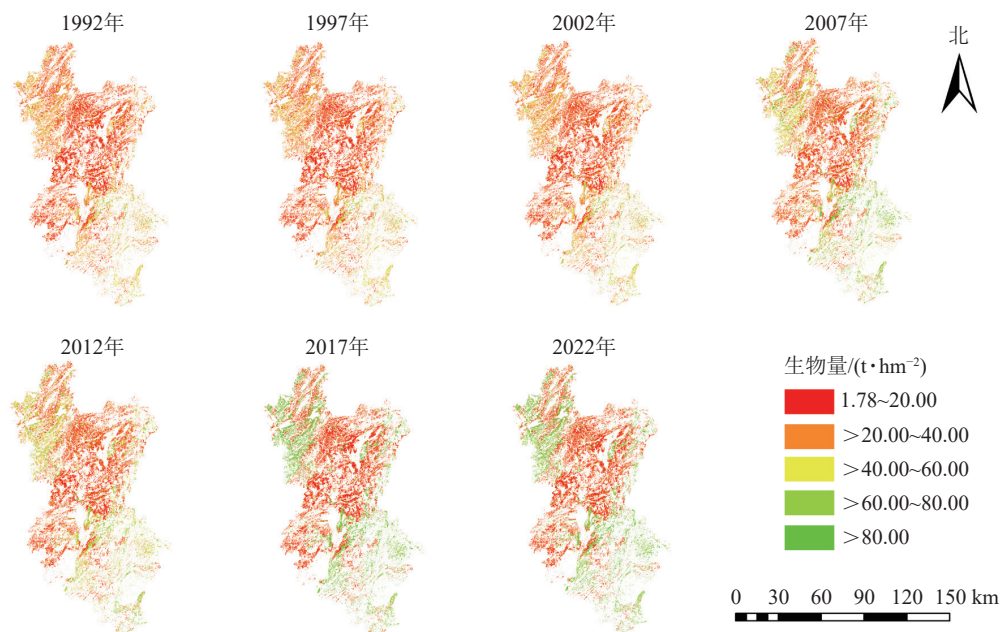
Figure 2 Scatter plots of the results of each model for biomass estimation

显著提升区域森林生物量估算的精度与可靠性，为构建差异化生物量反演模型提供关键的方法学依据。

## 3.2 AGB 空间格局演化

1992—2022 年，曲靖市针叶林 AGB 总体持续增长，空间格局由零散低值逐步演变为中高值连片分





基于云南省地理信息公共服务平台云S(2025)78号的标准地图制作，底图边界无修改。

图3 曲靖市主要针叶林地上生物量估算结果

Figure 3 Estimation results of aboveground biomass of major coniferous forests in Qujing City

布。低生物量区明显缩减，中高值斑块自 2000 年起快速扩展并趋于连通，与史川等<sup>[34]</sup>基于 2002—2017 年多期遥感和样地数据的研究结果一致，滇中地区(含曲靖市)针叶林碳储量总体持续上升，处于显著的碳汇增强阶段。2012 年后高生物量区进一步扩展并趋于连通，体现出持续的管护与经营增汇效应；但陡坡及立地条件较差区域，增幅有限，呈现明显的空间异质性。2009—2012 年极端干旱事件曾导致中龄林生长停滞和高生物量区出现短期回落，揭示了森林恢复过程存在显著的空间异质性和气候敏感性<sup>[35]</sup>。

### 3.3 特征选择与时序数据的优势及研究局限性

本研究采用二阶段特征选择策略，借鉴相关性初筛结合 RFE 的集成特征选择策略<sup>[32]</sup>，先按 Pearson 初筛将 187 个原始特征缩减至 106 个，显著缓解高维共线性问题<sup>[33]</sup>；随后在此基础上引入 RFE，筛选出与各林型显著相关的 10 个关键变量，在实现高变量压缩率(>70%)与增强模型稳健性的同时，也表明该策略具有良好的跨区域、跨林型普适性与可迁移性。其次，时间序列遥感数据通过整合多时相信息，有效削弱单一时相数据中不确定性因素的干扰，提升生物量估算模型的准确度与鲁棒性<sup>[36]</sup>。然而，生物量空间异质性显著，其估算结果仍存在一定不确定性，与岳彩荣<sup>[37]</sup>关于香格里拉高山松林生物量估算的研究结论一致，生物量的空间分布属性随时间动态变化，这种动态特征成为不确定性的来源，在长时序研究中尤为突出。本研究存在若干局限性：所使用的 Landsat TM/OLI 影像在传感器差异、空间分辨率偏低以及云处理后时间插值等方面存在局限，对预测结果会产生一定影响<sup>[17]</sup>；森林资源清查样地数据年代偏旧且样本有限，后续研究可引入更新、更丰富的数据源；当前研究仅估算针叶林乔木层的 AGB，未涵盖其他植被类型及土壤碳库，导致区域碳储量结果偏低。这一偏差得到已有研究的支持，如张加龙等<sup>[38]</sup>指出，忽略地下生物量与枯死木碳库可能导致碳储量被低估约 10%~15%。此外，GBRT 与 CatBoost 模型在森林碳储量估算中的研究较少<sup>[39]</sup>，其可行性与适用性仍需进一步探讨。后续应综合考虑树种差异与地形异质性，整合 GF-6、LiDAR 等多源高分辨率遥感数据，及多源遥感特征变量对模型性能的影响，以期提升碳储量空间分布估算的准确性。

## 4 结论

不同树种针叶林对遥感因子的响应存在明显差异，呈现特异性变量组合，其中，针叶混交林主要依赖高阶纹理和原始光谱，云南松林更敏感于地形因子和植被指数，华山松林则体现植被指数与纹理的复合响应，揭示其光谱结构和地形依赖的异质性。基于 GBRT、RF、XGBoost、PLS 和 CatBoost 等 5 种模

型构建生物量估测, 针叶混交林以 CatBoost 表现最佳, 云南松林和华山松林则以 GBRT 模型效果优异, 可见需结合树种特性选择相应合适的模型。1992—2022 年, 曲靖市针叶林生物量持续增长, 空间分布显著优化, 早期低值斑块广泛, 碳汇能力弱, 2002 年起生态修复促使中高值区扩展, 2022 年生物量达峰, 显示森林生态系统由退化向恢复转变, 体现了政策与自然恢复共同提升碳汇能力。

## 5 参考文献

- [1] 汤旭光, 刘殿伟, 王宗明, 等. 森林地上生物量遥感估算研究进展[J]. 生态学杂志, 2012, **31**(5): 1311–1318.  
TANG Xuguang, LIU Dianwei, WANG Zongming, *et al.* Estimation of forest aboveground biomass based on remote sensing data: a review [J]. *Chinese Journal of Ecology*, 2012, **31**(5): 1311–1318.
- [2] 张少伟, 惠刚盈, 韩宗涛, 等. 基于光学多光谱与 SAR 遥感特征快速优化的大区域森林地上生物量估测[J]. 遥感技术与应用, 2019, **34**(5): 925–938.  
ZHANG Shaowei, HUI Gangying, HAN Zongtao, *et al.* Estimation of large-scale forest above-ground biomass based on fast optimizing remotely sensed features from optical multi-spectral and SAR data [J]. *Remote Sensing Technology and Application*, 2019, **34**(5): 925–938.
- [3] 谢福明, 字李, 舒清态. 基于优化 k-NN 模型的高山松地上生物量遥感估测[J]. 浙江农林大学学报, 2019, **36**(3): 515–523.  
XIE Fuming, ZI Li, SHU Qingtai. Optimizing the k-nearest neighbors technique for estimating *Pinus densata* aboveground biomass based on remote sensing [J]. *Journal of Zhejiang A&F University*, 2019, **36**(3): 515–523.
- [4] MOHD ZAKI N A, LATIF Z A, SURATMAN M N. Modelling above-ground live trees biomass and carbon stock estimation of tropical lowland Dipterocarp forest: integration of field-based and remotely sensed estimates [J]. *International Journal of Remote Sensing*, 2018, **39**(8): 2312–2340.
- [5] 于贵瑞. 全球变化与陆地生态系统碳循环和碳蓄积[M]. 北京: 气象出版社, 2003.  
YU Guirui. *Global Change, Carbon Cycle and Storage in Terrestrial Ecosystem*[M]. Beijing: China Meteorological Press, 2003.
- [6] 张志, 田昕, 陈尔学, 等. 森林地上生物量估测方法研究综述[J]. 北京林业大学学报, 2011, **33**(5): 144–150.  
ZHANG Zhi, TIAN Xin, CHEN Erxue, *et al.* Review of methods on estimating forest above ground biomass [J]. *Journal of Beijing Forestry University*, 2011, **33**(5): 144–150.
- [7] LU Dengsheng, CHEN Qi, WANG Guangxing, *et al.* Aboveground forest biomass estimation with landsat and LiDAR data and uncertainty analysis of the estimates[J/OL]. *International Journal of Forestry Research*, 2012, **2012**: 436537[2025-07-05]. DOI: [10.1155/2012/436537](https://doi.org/10.1155/2012/436537).
- [8] CHEN Qi. LiDAR remote sensing of vegetation biomass[M]//WANG Guangxing, WENG Qihao. *Remote Sensing of Natural Resources*. Boca Raton: CRC Press, 2013: 424–445.
- [9] 徐小军, 周国模, 杜华强, 等. 基于 Landsat TM 数据估算雷竹林地上生物量[J]. 林业科学, 2011, **47**(9): 1–6.  
XU Xiaojun, ZHOU Guomo, DU Huaqiang, *et al.* Estimation of aboveground biomass of *Phyllostachys praecox* forest based on Landsat thematic mapper image [J]. *Scientia Silvae Sinicae*, 2011, **47**(9): 1–6.
- [10] 徐婷, 曹林, 余光辉. 基于 Landsat 8 OLI 的特征变量优化提取及森林生物量反演[J]. 遥感技术与应用, 2015, **30**(2): 226–234.  
XU Ting, CAO Lin, SHE Guanghui. Feature extraction and forest biomass estimation based on Landsat 8 OLI [J]. *Remote Sensing Technology and Application*, 2015, **30**(2): 226–234.
- [11] 井学辉, 曹磊, 刘云生, 等. 阿尔泰山小东沟乔木生物量空间分布规律[J]. 干旱区研究, 2016, **33**(3): 511–518.  
JING Xuehui, CAO Lei, LIU Yunsheng, *et al.* Spatial distribution pattern of biomass of *Arbor* species in Xiaodonggou in the Altay Mountains, China [J]. *Arid Zone Research*, 2016, **33**(3): 511–518.
- [12] 国庆喜, 张锋. 基于遥感信息估测森林的生物量[J]. 东北林业大学学报, 2003, **31**(2): 13–16.  
GUO Qingxi, ZHANG Feng. Estimation of forest biomass based on remote sensing [J]. *Journal of Northeast Forestry University*, 2003, **31**(2): 13–16.
- [13] 杨伟志, 赵鹏祥, 薛大庆, 等. 基于 Landsat-8 影像的西宁市南北山森林生物量估测模型研究[J]. 西北林学院学报, 2016, **31**(2): 33–37, 97.



- YANG Weizhi, ZHAO Pengxiang, XUE Daqing, *et al.* Estimation of forest biomass model of north and south mountains based on Landsat-8 remote sensing image data in Xining [J]. *Journal of Northwest Forestry University*, 2016, **31**(2): 33–37, 97.
- [14] 邱布布, 徐丽华, 张茂震, 等. 基于 Landsat OLI 和 ETM+ 的杭州城市绿地地上生物量估算比较研究[J]. 西北林学院学报, 2018, **33**(1): 225–232.
- QIU Bubu, XU Lihua, ZHANG Maozhen, *et al.* Estimation of above-ground biomass of urban green land in Hangzhou based on Landsat OLI and ETM+ data [J]. *Journal of Northwest Forestry University*, 2018, **33**(1): 225–232.
- [15] 魏安超, 张大为. 云南松单木材积生长率模型研究[J]. 林业资源管理, 2020(6): 40–46.
- WEI Anchao, ZHANG Dawei. Study on the volume growth rate model of *Pinus yunnanensis* of individual tree [J]. *Forest Resources Management*, 2020(6): 40–46.
- [16] 李成荣, 王妍, 张超, 等. 基于 GBRT 模型的滇东地区森林碳储量遥感估算研究[J]. 三峡生态环境监测, 2023, **8**(3): 96–106.
- LI Chengrong, WANG Yan, ZHANG Chao, *et al.* Remote sensing estimation of forest carbon storage in Eastern Yunnan based on GBRT model [J]. *Ecology and Environmental Monitoring of Three Gorges*, 2023, **8**(3): 96–106.
- [17] 方精云, 刘国华, 徐嵩龄. 我国森林植被的生物量和净生产量[J]. 生态学报, 1996, **16**(5): 497–508.
- FANG Jingyun, LIU Guohua, XU Songling. Biomass and net production of forest vegetation in China [J]. *Acta Ecologica Sinica*, 1996, **16**(5): 497–508.
- [18] 罗恒春, 魏安超, 黄田, 等. 云南松生物量和碳储量动态变化分析[J]. 林业资源管理, 2016(6): 37–43.
- LUO Hengchun, WEI Anchao, HUANG Tian, *et al.* Analysis on dynamic change of biomass and carbon stock of *Pinus yunnanensis* in Yunnan, China [J]. *Forest Resources Management*, 2016(6): 37–43.
- [19] 卢腾飞, 胥辉, 欧光龙. 基于混合效应模型的曲靖市云南松林地上生物量遥感估测[J]. 西南林业大学学报(自然科学), 2020, **40**(1): 104–115.
- LU Tengfei, XU Hui, OU Guanglong. Remote sensing estimation on aboveground biomass for *Pinus yunnanensis* forests in Qujing City using mixed effect models [J]. *Journal of Southwest Forestry University (Natural Sciences)*, 2020, **40**(1): 104–115.
- [20] 赵盼盼. 基于 Landsat TM 和 ALOS PALSAR 数据的森林地上生物量估测研究[D]. 杭州: 浙江农林大学, 2016.
- ZHAO Panpan. *Aboveground Forest Biomass Estimation Based on Landsat TM and ALOS PALSAR Data*[D]. Hangzhou: Zhejiang A&F University, 2016.
- [21] 孙雪莲. 基于 Landsat 8-OLI 的香格里拉高山松林生物量遥感估测模型研究[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.
- [22] 马俊鹏. 基于多源哨兵卫星影像的湿地遥感制图研究[D]. 哈尔滨: 黑龙江大学, 2024.
- MA Junpeng. *Wetland Remote Sensing Mapping Based on Multi-source Sentinel Satellite Imagery*[D]. Harbin: Helongjiang University, 2024.
- [23] BREIMAN L. Random forests [J]. *Machine Learning*, 2001, **45**(1): 5–32.
- [24] XU Ruize, ZHANG Jiahua, WANG Jingwen, *et al.* Quantitative assessment of factors influencing the spatiotemporal variation in carbon dioxide fluxes simulated by multi-source remote sensing data in tropical vegetation[J/OL]. *Remote Sensing*, 2023, **15**(24): 5677[2025-07-05]. DOI: [10.3390/rs15245677](https://doi.org/10.3390/rs15245677).
- [25] 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.
- [26] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine [J]. *The Annals of Statistics*, 2001, **29**(5): 1189–1232.
- [27] 翁剑成, 付宇, 林鹏飞, 等. 基于梯度推进决策树的日维度交通指数预测模型[J]. 交通运输系统工程与信息, 2019, **19**(2): 80–85, 93.
- WENG Jiancheng, FU Yu, LIN Pengfei, *et al.* GBDT method based on prediction model of daily dimension traffic index [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2019, **19**(2): 80–85, 93.

- [28] LIANG Xiao, XIE Qiwei, LUO Chunyan, *et al.* Principal model analysis based on bagging PLS and PCA and its application in financial statement fraud [J]. *Journal of Systems Science and Information*, 2024, **12**(2): 212–228.
- [29] 李春华. 升金湖枯水期草滩植被碳储量遥感估算研究[D]. 合肥: 安徽大学, 2021.  
LI Chunhua. *Remote Sensing Estimation of Grassland Carbon Storage During Dry Season at Shengjin Lake*[D]. Hefei: Anhui University, 2021.
- [30] 娄明华, 张会儒, 雷相东, 等. 天然栎类阔叶混交林林分平均高与平均胸径关系模型[J]. 北京林业大学学报, 2020, **42**(9): 37–50.  
LOU Minghua, ZHANG Huiru, LEI Xiangdong, *et al.* Relationship model between stand mean height and mean DBH for natural *Quercus* spp. broadleaved mixed stands [J]. *Journal of Beijing Forestry University*, 2020, **42**(9): 37–50.
- [31] 徐兴. 智能算法在 PID 失调时间预测中的应用[J]. 石化技术, 2025, **32**(4): 209–211.  
XU Xing. Application of intelligent algorithm in PID maladjustment time prediction [J]. *Petrochemical Industry Technology*, 2025, **32**(4): 209–211.
- [32] LUO Mi, WANG Yifu, XIE Yunhong, *et al.* Combination of feature selection and CatBoost for prediction: the first application to the estimation of aboveground biomass[J/OL]. *Forests*, 2021, **12**(2): 216[2025-07-05]. DOI: [10.3390/f12020216](https://doi.org/10.3390/f12020216).
- [33] ZHANG Jialong, LU Chi, XU Hui, *et al.* Estimating aboveground biomass of *Pinus densata* -dominated forests using Landsat time series and permanent sample plot data [J]. *Journal of Forestry Research*, 2019, **30**(5): 1689–1706.
- [34] 史川, 胥辉, 张成程. 滇中地区针叶林碳储量遥感反演及动态变化分析[J]. 现代农业研究, 2025, **31**(3): 83–85.  
SHI Chuan, XU Hui, ZHANG Chengcheng. Remote sensing inversion and dynamic change analysis of carbon storage in coniferous forests in Central Yunnan [J]. *Modern Agriculture Research*, 2025, **31**(3): 83–85.
- [35] 陶云, 张万诚, 段长春, 等. 云南 2009—2012 年 4 年连旱的气候成因研究[J]. 云南大学学报(自然科学版), 2014, **36**(6): 866–874.  
TAO Yun, ZHANG Wancheng, DUAN Changchun, *et al.* Climatic causes of continuous drought over Yunnan Province from 2009 to 2012 [J]. *Journal of Yunnan University (Natural Sciences Edition)*, 2014, **36**(6): 866–874.
- [36] ZHANG Yuzhen, LIU Jingjing, LI Wenhao, *et al.* A proposed ensemble feature selection method for estimating forest aboveground biomass from multiple satellite data[J/OL]. *Remote Sensing*, 2023, **15**(4): 1096[2025-07-05]. DOI: [10.3390/rs15041096](https://doi.org/10.3390/rs15041096).
- [37] 岳彩荣. 香格里拉县森林生物量遥感估测研究[D]. 北京: 北京林业大学, 2012.  
YUE Cairong. *Forest Biomass Estimation in Shangri-La County Based on Remote Sensing*[D]. Beijing: Beijing Forestry University, 2012.
- [38] 张加龙, 胥辉. 基于遥感的高山松连清固定样地地上生物量估测模型构建[J]. 北京林业大学学报, 2020, **42**(7): 1–11.  
ZHANG Jialong, XU Hui. Establishment of remote sensing based model to estimate the aboveground biomass of *Pinus densata* for permanent sample plots from national forestry inventory [J]. *Journal of Beijing Forestry University*, 2020, **42**(7): 1–11.
- [39] 孙梦莲, 余坤勇, 张晓萍, 等. 基于机载激光雷达点云数据和 Catboost 算法的杉木单木蓄积量估测研究[J]. 西南林业大学学报(自然科学), 2024, **44**(3): 157–165.  
SUN Menglian, YU Kunyong, ZHANG Xiaoping, *et al.* Research on estimation of single wood accumulation of chinese fir based on airborne Lidar point cloud data and Catboost algorithm [J]. *Journal of Southwest Forestry University (Natural Sciences)*, 2024, **44**(3): 157–165.