

基于机载激光雷达点云的桉树林分蓄积量估算模型构建

邓 焯^{1,2,3}, 李 斌^{1,2,3}, 范光鹏^{1,2,3}, 赵天忠^{1,2,3}, 于永辉⁴

(1. 北京林业大学 信息学院, 北京 100083; 2. 北京林业大学 国家林业草原林业智能信息处理工程技术研究中心, 北京 100083; 3. 北京林业大学 林业信息化研究所, 北京 100083; 4. 广西壮族自治区国有高峰林场, 广西南宁 530001)

摘要: 【目的】通过 2018 年 1—2 月广西国有高峰林场机载激光雷达数据及地面调查数据, 采用参数方法和非参数方法建立回归模型, 反演桉 *Eucalyptus* 树人工林森林蓄积量。【方法】通过点云提取点云高度参数、点云密度参数、林分郁闭度等点云特征变量, 采用参数方法 (逐步回归、偏最小二乘回归) 和非参数方法 (随机森林回归、支持向量机回归) 进行林分蓄积量构建, 通过与样地实测数据对比, 进行模型回归预测性能评估, 进而选择出表现最优蓄积量反演模型。【结果】采用留一法对以上 4 种模型进行验证, 结果显示: 逐步回归模型 R^2 为 0.85、均方根误差 (RMSE) 为 $23.93 \text{ m}^3 \cdot \text{hm}^{-2}$ 、平均绝对误差 (MAE) 为 $18.18 \text{ m}^3 \cdot \text{hm}^{-2}$; 偏最小二乘回归模型 R^2 为 0.81、RMSE 为 $26.52 \text{ m}^3 \cdot \text{hm}^{-2}$ 、MAE 为 $19.94 \text{ m}^3 \cdot \text{hm}^{-2}$; 核函数为 RBF 的支持向量回归模型 R^2 为 0.88、RMSE 为 $21.35 \text{ m}^3 \cdot \text{hm}^{-2}$ 、MAE 为 $16.62 \text{ m}^3 \cdot \text{hm}^{-2}$; 随机森林回归模型 R^2 为 0.84、RMSE 为 $24.53 \text{ m}^3 \cdot \text{hm}^{-2}$ 、MAE 为 $17.41 \text{ m}^3 \cdot \text{hm}^{-2}$ 。【结论】采用随机森林进行变量筛选后, RBF-SVR 模型拟合优度及泛化能力最优; 通过逐步筛选法结合方差膨胀因子 (VIF) 方法优选变量的逐步回归模型次之; 最后为随机森林回归模型与偏最小二乘回归模型。可见, 在解决林业激光雷达领域中的回归预测问题时, 采用非参数方法构建 RBF-SVR 模型更有优势。本研究建立的 4 种森林类型蓄积量模型, 各模型均有较高精度且符合森林资源调查相关技术规定要求。图 6 表 6 参 25

关键词: 机载激光雷达; 林分蓄积量; 参数回归; 非参数回归; 变量筛选

中图分类号: S757 文献标志码: A 文章编号: 2095-0756(2022)06-1330-10

Estimation model of *Eucalyptus* stand volume based on airborne LiDAR Point Cloud

DENG Zhuo^{1,2,3}, LI Bin^{1,2,3}, FAN Guangpeng^{1,2,3}, ZHAO Tianzhong^{1,2,3}, YU Yonghui⁴

(1. College of Information, Beijing Forestry University, Beijing 100083, China; 2. Engineering Research Center for Forestry-oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing Forestry College, Beijing 100083, China; 3. Institute of Forestry Information, Beijing Forestry University, Beijing 100083, China; 4. State Owned Gaofeng Forest Farm of Guangxi Zhuang Autonomous Region, Nanning 530001, Guangxi, China)

Abstract: [Objective] LiDAR, as an active remote sensing technology, has proven to be an effective and efficient means for large-scale dynamic monitoring and investigation for forest resources. With an analysis of the airborne LiDAR data and ground survey data collected of Guangxi state-owned Gaofeng forest farm from January to February 2018, this paper is aimed to establish a regression model by using parametric and nonparametric methods to inverse the forest volume of *Eucalyptus* plantation. [Method] First, point cloud characteristic variables such as point cloud height parameters, point cloud density parameters and stand canopy

收稿日期: 2022-01-10; 修回日期: 2022-06-13

基金项目: 国家重点研发计划项目 (2017YFD0600906)

作者简介: 邓焯 (ORCID: 0000-0003-2813-2198), 从事激光雷达技术、林业遥感研究。E-mail: o2008k@qq.com。通

信作者: 赵天忠 (ORCID: 0000-0003-1521-9702), 教授, 博士, 博士生导师, 从事森林资源调查研究。

E-mail: ztz@bjfu.edu.cn

density were extracted from point cloud after which parametric methods (stepwise regression, partial least squares regression) and nonparametric methods (random forest regression and support vector machine regression) were used to construct stand volume. Then the prediction performance of model regression is evaluated by comparing the estimated data of sample plot with the measured ones so that the optimal volume inversion model could be selected. [Result] After being verified by leaving one method, results are as follows: for the stepwise regression model, $R^2=0.85$, $RMSE=23.93 \text{ m}^3 \cdot \text{hm}^{-2}$ and $MAE=18.18 \text{ m}^3 \cdot \text{hm}^{-2}$; for the partial least squares regression model, $R^2=0.81$, $RMSE=26.52 \text{ m}^3 \cdot \text{hm}^{-2}$ and $MAE=19.94 \text{ m}^3 \cdot \text{hm}^{-2}$; for the support vector regression model with RBF kernel function, $R^2=0.88$, $RMSE=21.35 \text{ m}^3 \cdot \text{hm}^{-2}$, $MAE=16.62 \text{ m}^3 \cdot \text{hm}^{-2}$ whereas for the random forest regression model, $R^2=0.84$, $RMSE=24.53 \text{ m}^3 \cdot \text{hm}^{-2}$ and $MAE=17.41 \text{ m}^3 \cdot \text{hm}^{-2}$. [Conclusion] After variable screening with random forest; the RBF-SVR model has demonstrated the best fitness and generalization ability, followed by the stepwise regression model of optimizing variables by stepwise screening method combined with variance inflation factor (VIF) method whereas the random forest regression model and partial least squares regression model came last. It was also shown that nonparametric method is a better choice in the construction of RBF-SVR model in solving the problem of regression prediction in the field of forestry LiDAR and that the volume models of four forest types established in this study have high accuracy and met the requirements of relevant technical regulations of forest resources investigation. [Ch, 6 fig. 6 tab. 25 ref.]

Key words: airborne LiDAR Point Cloud; stand volume; parametric regression; non-parametric regression; variables screening

森林是陆地生态系统的主体，在全球碳循环中起着十分重要的作用。森林蓄积量作为衡量森林数量的重要的指标之一，能够直观反映森林资源数量和质量，获取森林蓄积量是推算森林生物量和碳储量的必要环节^[1-3]。《“十四五”林业草原保护发展规划纲要》明确指出：将增加中国森林蓄积量作为主要目标之一，并将森林蓄积量与森林覆盖率作为两大约束性指标。因此，实现森林资源动态检测、准确获取森林蓄积量等参数信息，已成为当前森林资源调查的迫切需求。激光雷达 (light detection and ranging, LiDAR) 作为一种新兴的主动式遥感技术，能够从不同空间尺度对生态系统进行高效精准的监测^[4]。机载 LiDAR 发射的激光脉冲能穿透森林冠层并获取树冠上部枝叶的空间信息，通过林分内的空隙，测量冠层结构信息和林下地形信息^[5]。而传统方法通过样木、样地等抽样调查进行林分森林参数的推算，调查周期长、成本高且调查范围有限^[6-7]。激光雷达在获取森林空间结构因子和地形因子等信息方面具有精度高、范围广等优势。

近年来，基于机载激光雷达的林分蓄积量反演已有许多研究成果，集中在构建参数方法和非参数方法的蓄积量反演模型。PAWE 等^[8]基于机载 LiDAR 数据对波兰东南部林区建立多元线性回归蓄积量反演模型，其中均方根误差 (root mean square error, RMSE) 为 15.2%；CHIRICI 等^[9]以 Landsat 5TM、卫星 LiDAR 数据等预测变量联合气温、降水和地形等辅助变量进行意大利中部地区蓄积量的大尺度反演，其中随机森林回归模型最优，决定系数 (R^2) 为 0.69、RMSE 为 37.2%；陈松等^[10]基于 Sentinel-2 与机载 LiDAR 数据采用不同回归方法对广西高峰林场界牌、东升分场进行蓄积量反演，构建 MLR-Logistic 联立模型精度优于随机森林等机器学习方法， R^2 为 0.60、相对均方根误差 (relative root mean-squared error, RRMSE) 为 29.29%；曾伟生等^[11]基于机载 LiDAR 数据，采用线性和非线性参数回归方法对东北林区进行蓄积量反演，其中非线性回归模型 R^2 为 0.71~0.82，略优于线性回归。已有多位研究者基于机载 LiDAR 数据进行林分蓄积量反演模型研究，但对于建模方法中参数回归和非参数回归模型的比较国内研究较少。本研究以广西国有高峰林场桉 *Eucalyptus* 树人工林为研究对象，基于机载激光雷达数据及地面调查数据，采用逐步回归、偏最小二乘回归等参数回归和随机森林、支持向量回归等非参数回归进行蓄积量反演模型研建，并通过模型评价指标对以上 4 种方法进行模型评估，进而选择出拟合优度、泛化能力最优模型。

1 数据获取与预处理

1.1 研究区概况

研究区位于广西壮族自治区南宁市兴宁区的国有高峰林场, 22°51'~23°02'N, 108°06'~108°31'E, 该区地处亚热带地区, 年平均气温约 21 ℃, 年平均降水量为 1 200~1 500 mm, 相对湿度为 79%, 属丘陵地貌, 海拔为 100~460 m, 坡度为 6°~35°, 具有较厚的赤红壤, 适宜亚热带和热带树种生长, 森林覆盖率达 87%, 主要树种为杉木 *Cunninghamia lanceolata*、巨尾桉 *Eucalyptus grandis* × *E. urophylla*、马尾松 *Pinus massoniana* 等。

1.2 样地数据

调查时间为 2018 年 1—2 月, 研究区内共设置 71 块桉树样地 (图 1), 其中激光雷达覆盖范围内共 57 块样地。样地大小为 20 m×20 m、25 m×25 m 和 25 m×50 m。采用实时动态差分技术 (real-time Kinematic, RTK) 进行样地定位, 记录样地中心点及样地角点。采用每木检尺的方法, 使用胸径尺、激光测高仪和皮尺逐一测量样地内树木的胸径、树高等数据。统计样地调查数据得到样地算数平均树高、算数平均胸径、样地面积 (表 1)。通过广西地区桉树二元材积表对单木材积量进行计算^[12], 进而计算得到样地尺度的公顷蓄积量值 ($V_{\text{样地}}$): $V_{\text{样地}} = V_{\text{公顷}} S_{\text{样地}} / 10\ 000$ 。其中: $V_{\text{公顷}}$ 为通过二元材积公式计算得的每公顷蓄积量, $S_{\text{样地}}$ 为桉树样地面积。

1.3 LiDAR 数据及预处理

于 2018 年 1 月采集机载 LiDAR 数据, 使用无人机并搭载 LMS-Q680i 激光雷达扫描仪, 实际飞行高度为 1 000 m, 最大扫描角度为 30°, 波长为 1 550 nm, 激光脉冲长度为 3 ns, 采样间隔为 1 ns, 最大扫描频率为 400 KHz, 垂直分辨率为 0.15 m, 点云密度为 10 pt·m⁻²。

使用 LiDAR 360 软件对机载激光雷达点云进行点云拼接、去噪、地面点分类、基于地面点归一化等预处理, 结果如图 2 所示。根据样地角点以中心点坐标对点云进行裁剪, 提取出基于样地尺度特征变量共 48 个, 包括 37 个点云高度参数、10 个密度参数、郁闭度等点云特征变量, 具体参数见表 2。

1.4 建模数据准备

使用 MATLAB 激光雷达覆盖范围内 57 个样本进行随机抽样, 按照 3:1 的比例选取 42 个作为训练样本, 15 个作为验证样本。以样地实测公顷蓄积量为因变量, 筛选后点云特征为自变量, 采用逐步回归、偏最小二乘回归、随机森林回归、支持向量回归模型进行拟合。

2 研究方法

2.1 变量筛选

2.1.1 基于逐步筛选法优选变量 对于逐步回归, 采用逐步筛选法对所有特征变量进行变量筛选; 再对筛选出特征变量进行多重共线性检验, 计算各变量间容忍度或特征变量间方差膨胀因子 (F_{VI}) 2 个统计

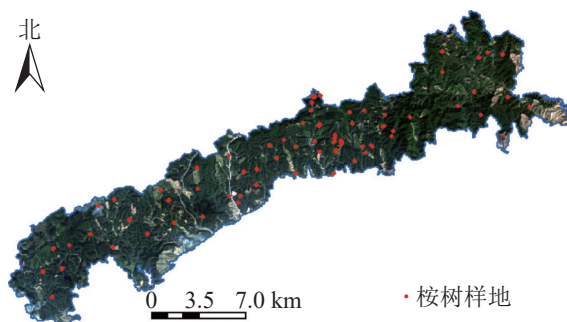


图 1 研究区样地分布示意图

Figure 1 Location of the sample plots distribution

表 1 样地参数统计

Table 1 Parameter statistics of sample plots			
项目	平均树高/m	平均胸径/cm	单位蓄积量/(m ³ ·hm ⁻²)
最大值	30.40	21.10	320.66
最小值	7.48	5.24	17.79
平均值	15.19	12.06	90.47
标准差	4.22	3.55	65.58

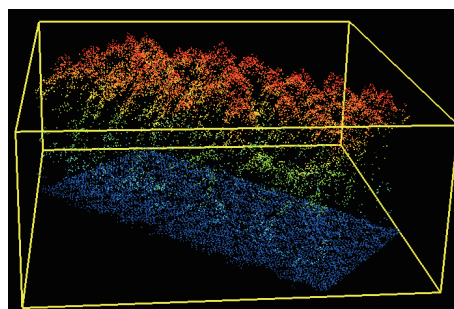


图 2 样地点云预处理结果示例

Figure 2 Example of point cloud preprocessing results of sample plots

量^[13]，对于容忍度 ≤ 0.2 或 $F_{VI} \geq 5$ 的变量进行进一步讨论，确定最优特征变量子集。 $F_{VI}=1/(1-R^2)$ 。其中： R 为特征变量间相关系数。

2.1.2 基于随机森林优选变量 对于偏最小二乘回归、支持向量机回归、随机森林回归等方法，使用随机森林中重要性排序对特征变量进行优选。其主要原理为随机森林算法在构建各决策树时，对某一特征变量进行取舍，若此时均方误差 (mean square error, MSE) 有较大变化，则该特征变量重要性高，最终得到所有特征变量重要性排序。

2.2 蓄积量模型建立方法

2.2.1 逐步回归 逐步回归 (stepwise regression, SR) 可用于筛选并剔除引起多重共线性的变量，逐步回归建立模型一般形式为： $Y=\beta_0+\beta_1X_1+\beta_2X_2+\dots+\beta_iX_i+\varepsilon$ 。逐步筛选法结合了向前选择变量法和向后选择变量法的优点，对 i 个自变量 X 分别与因变量 Y 建立一元回归模型，计算各变量所对应 F 值，其中， β_0 为常数， β_i 为回归系数， ε 为误差项。在建立逐步回归模型时，选择当前未加入模型的预测变

量中 F 的最大值所对应的 X_i 加入模型，再对已选入预测变量逐个进行 t 检验，若存在已选入预测变量不再显著，则将其剔除。重复以上步骤，直到既没有显著的解释变量选入回归方程，也没有不显著的解释变量从回归方程中剔除为止，将保留下来的最优子集作为建立模型的特征变量，并将候选变量对应系数保留。

2.2.2 偏最小二乘回归 偏最小二乘回归 (partial least squares regression, PLSR) 结合了主成分分析、典型相关分析的优点，实现了数据降维、信息综合^[14]。首先对训练样本进行数据标准化处理，调用 plsregress 函数提取主成分与自变量和因变量组合。分别计算自变量和因变量提取出成分的贡献率进而计算得到累计贡献率，当度量因子 $Q^2_h \geq 0.0975$ 停止主成分的提取^[15]。统计主成分对个数并建立各自变量与因变量的线性表达式，最后根据所建立的主成分对模型，整理得到 PLSR 模型。

2.2.3 支持向量机回归 支持向量机算法 (support vector machine, SVM) 利用内积核函数代替高维空间的非线性映射，不涉及大数定律和概率测度等有关问题，且 SVM 的决策函数仅有少数支持向量决定，该算法在解决小样本多维度回归和分类问题时泛化能力强，具备较好的“鲁棒性”。VAPNIK^[16] 在 SVM 分类的基础上引入了不敏感损失函数，得到了支持向量机回归算法 (support vector regression, SVR)。

使用 Libsvm 工具箱实现 SVR 算法，采用网格搜索法 (grid search) 对常用的 4 种核函数进行参数寻优^[17]，即线性、多项式、RBF (径向基核函数)、sigmoid (多层感知机核函数)。同时进行十折交叉验证，保证回归模型中惩罚系数 (C) 与 gamma 值 (g) 达到最优。

2.2.4 随机森林回归 随机森林回归 (RFR) 算法采用自助采样法 (bootstrap sampling)，在以决策树构建 Bagging 集成的基础上，对样本和特征变量进行随机选择^[18]。调用 TreeBagger 函数进行 RFR 算法建模，通过对决策树数量 (n_{tree}) 和最小叶子点数进行参数寻优，直到袋外 (out-of-bag, OOB) 误差的 MSE 达到最小，保证模型预测性能达到最优，将寻优结果作为 RFR 模型的建模参数，用于模型构建。

2.3 模型评估方法

本研究使用决定系数 (R^2)、均方根误差 (RMSE) 和平均绝对误差 (MAE) 对林分蓄积量估测模型进行定量的精度验证和评价。其计算公式为：

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

表 2 提取点云特征变量

变量类型	变量名	特征描述
高度变量	$H_{\max}$	归一化后所有点的Z值的最大值
	$H_{\min}$	归一化后所有点的Z值的最小值
	H_{median_z}	高度平均偏差
	H_{stddev}	高度标准差
	H_{kurtosis}	高度峰度
	H_{sqr_m}	高度二次幂均值
	H_{cur_m}	高度三次幂均值
密度变量	$H_{p1}, H_{p5}, \dots, H_{p99}$	归一化点云高度分布的百分位数，共15个
	$H_{A1}, H_{A2}, \dots, H_{A99}$	归一化点云累计高度的百分位数，共15个
	$D_1, D_2, \dots, D_9$	将点云从低到高分成10个相同高度的切片，该层回波数点在所有返回点的所占比例，共9个
郁闭度	C_C	首次回波中，植被点数与所有点的比值

$$E_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}};$$

$$E_{\text{MA}} = \frac{1}{n} \sum_{i=1}^n |(y_i - \hat{y}_i)|。$$

其中： y_i 、 $\bar{y}_i$ 、 $\hat{y}_i$ 分别为实测蓄积量值、实测蓄积量均值、模型预测蓄积量值； E_{RMS} 为均方根误差； E_{MA} 为平均绝对误差； n 为验证样本数量。进行评估模型时， R^2 越趋近于1，代表拟合程度越高； E_{RMS} 越小，代表真实值与模型预测值之间离散程度越小； E_{MA} 越小，代表真实值与模型预测值之间的误差越小。

3 结果与分析

3.1 变量筛选结果

3.1.1 逐步筛选法 在采用逐步筛选法进行偏 F 检验时，取偏 F 检验拒绝域的临界值为 $F_{\text{进}} > F_{\text{出}}$ ， $F_{\text{进}}$ 为选入变量时的临界值、 $F_{\text{出}}$ 为删除变量时的临界值。本研究设定 $F_{\text{进}}$ 为0.10， $F_{\text{出}}$ 为0.11，筛选特征变量结果为 D_9 、 H_{p95} 、 H_{max} 、 H_{kurtosis} 。

对筛选出的变量进行双变量相关性分析，根据相关系数(R)计算各变量间方差膨胀因子(F_{VI})。

计算结果如表3， H_{max} 与 H_{p95} 之间 F_{VI} 为250.25， F_{VI} 大于5，认为两者之间存在共线性。因此分别以 D_9 、 H_{max} 、 H_{kurtosis} 和 D_9 、 H_{p95} 、 H_{kurtosis} 为建模因子。

3.1.2 基于随机森林筛选变量 对所有的48个特征变量进行重要性排序，调用随机森林中OOBPermuted VarDeltaError参数，得到所有特征变量的重要性。选择重要性大于0.2的变量作为建模因子，筛选结果如图3，包括8个高度变量、2个密度变量，其中 H_{curt} 重要性最高，达0.52； D_6 重要性相对最低，为0.22。

3.2 模型建立与评价

3.2.1 逐步回归法 根据特征变量优选结果，分别建立以 D_9 、 H_{max} 、 H_{kurtosis} 和 D_9 、 H_{p95} 、 H_{kurtosis} 为建模因子的多元线性模型：

$$Y_1 = 470.232D_9 + 4.897H_{\text{p95}} + 1.453H_{\text{kurtosis}} - 56.903;$$

$$Y_2 = 482.214D_9 + 4.855H_{\text{max}} + 1.461H_{\text{kurtosis}} - 65.01。$$

使用SPSS 26对 Y_1 、 Y_2 模型进行初步评估，计算2种模型的相关系数(R)、 R^2 以及标准估算误差(SE)，结果如表4。结果表明 Y_2 中 R^2 、SE均优于 Y_1 ，因此选定 Y_2 作为逐步回归模型。

3.2.2 偏最小二乘回归 利用PLSR提取的主成分 F_1 、 F_2 与因变量 Y 关系分别为 $F_1=5.840Y+b$ 和 $F_2=0.271Y+b$ ， b 为常数，得到最小二乘回归模型为：

$$Y = -42.571 - 1.344H_{\text{A25}} - 0.513H_{\text{A30}} + 1.450H_{\text{curt}} + 1.134H_{\text{mean}} + 0.705H_{\text{p20}} + 1.088H_{\text{p25}} + 1.337H_{\text{p30}} + 1.523H_{\text{p40}} + 24.647D_6 + 445.634D_9。$$

3.2.3 支持向量机回归 调用meshgrid函数对 C 、 γ 进行参数寻优并采用十折交叉验证，调用svmtrain函数，分别构建4种不同核函数的SVR模型并进行模型训练，训练结果如表5。得到拟合结果

表3 各特征变量间的方差膨胀因子

变量	D_9	H_{max}	H_{p95}	H_{kurtosis}
D_9		2.46	2.69	1.00
H_{max}	2.46		250.25	1.01
H_{p95}	2.69	250.25		1.01
H_{kurtosis}	1.00	1.01	1.01	

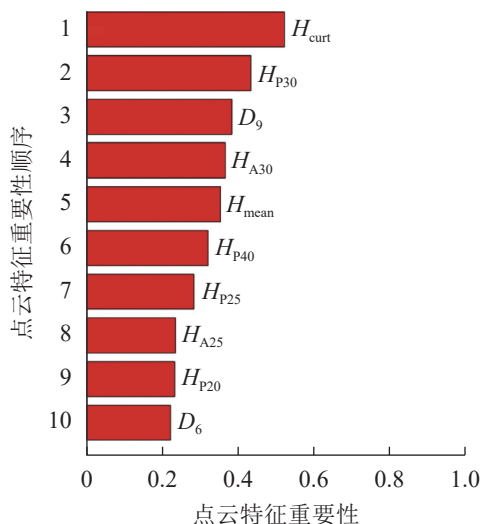


图3 点云特征变量重要性排序

Figure 3 Importance ranking of point cloud characteristic variables

表4 逐步回归模型初步评估

模型	R	R^2	调整后 R^2	SE
Y_1	0.950	0.902	0.894	18.460
Y_2	0.953	0.908	0.900	17.904

最优模型为 RBF-SVR，其 C 为 8， γ 为 0.125。RBF-SVR 模型 R^2 为 0.85，RMSE 为 $29.24\text{ m}^3\cdot\text{hm}^{-2}$ ，MAE 为 $94.98\text{ m}^3\cdot\text{hm}^{-2}$ ，选定该模型作为本研究的 SVR 模型。

3.2.4 随机森林回归 本研究采用穷举法对决策树数量和最小叶子点数进行参数寻优。分别设置决策树数量和最小叶子点数最小值为 50 和 2，每次递增 10 和 1，对寻优过程循环，通过观察 OOB 误差的 MSE 变化，直到寻找到本模型最优参数，寻优结果如图 4。最终确定决策树数量为 90，叶子节点数量为 2，将寻优结果作为建模参数，输入训练样本，完成 RFR 模型构建。

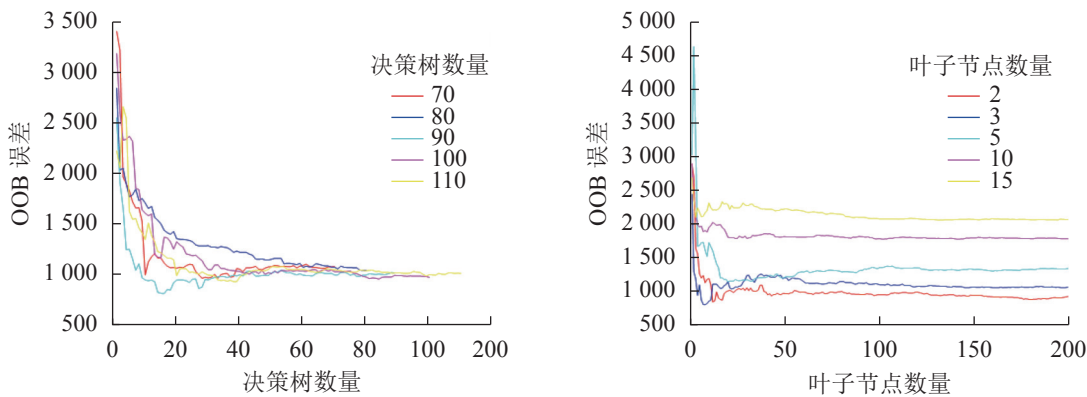


图 4 随机森林参数寻优结果
Figure 4 Evaluation of growing stock volume inversion model

3.2.5 模型评价 使用验证样本对各模型的预测性能进行评估，结果如表 6、图 5。可以看出不同的回归模型均表现出较好的拟合结果，其中拟合结果最优模型为 RFR 模型，模型评价结果 R^2 为 0.95，RMSE 为 $12.64\text{ m}^3\cdot\text{hm}^{-2}$ ，MAE 为 $8.00\text{ m}^3\cdot\text{hm}^{-2}$ ，RBF-SVR 模型其次， R^2 为 0.94，RMSE 为 $13.09\text{ m}^3\cdot\text{hm}^{-2}$ ，MAE 为 $11.65\text{ m}^3\cdot\text{hm}^{-2}$ 。将验证样本带入模型，检验模型的预测能力，结果显示各模型预测能力与模型的拟合效果一致。通过传统方法的划分训练样本与验证样本，存在一定的偶然性与不确定性，为了进一步确保各模型的稳定性与泛化能力，采用留一法交叉验证 (leave-one-out cross validation, LOOCV) 对本研究中各模型进行再次评估^[19]，如图 6 所示：RBF-SVR 模型表现最优， R^2 为 0.88，RMSE 为 $21.35\text{ m}^3\cdot\text{hm}^{-2}$ ，MAE 为 $16.62\text{ m}^3\cdot\text{hm}^{-2}$ ，与其他模型相比， R^2 高出 0.03~0.07，RMSE 减少 2.58~5.17 $\text{m}^3\cdot\text{hm}^{-2}$ ，MAE 减少 0.79~3.32 $\text{m}^3\cdot\text{hm}^{-2}$ 。

4 讨论

本研究采用 SR、PLSR 等参数回归方法与 RFR、SVR 等非参数回归方法进行广西高峰林场桉树人工林的蓄积量反演模型研建，通过对点云特征变量筛选、建模因子优选、参数寻优等方法保证各模型达到最优性能，在此基础上采用留一法对各模型进行交叉验证，保证了模型的稳定性与泛化能力，并将该结果作为本研究各模型性能评估的最终结果。研究结果如下：①核函数为 RBF 的 SVR 在 4 种模型中预测精度最高 (ΔR^2 为 0.03~0.07、 ΔRMSE 为 $2.58\sim5.17\text{ m}^3\cdot\text{hm}^{-2}$ 、 ΔMAE 为 $0.79\sim3.32\text{ m}^3\cdot\text{hm}^{-2}$)，采用 SVR 模

表 5 SVR 不同核函数拟合结果

Table 5 Fitting results of different kernel functions of SVR				
核函数		R^2	RMSE/($\text{m}^3\cdot\text{hm}^{-2}$)	MAE/($\text{m}^3\cdot\text{hm}^{-2}$)
线性	训练样本	0.89	18.62	14.49
	验证样本	0.83	30.43	24.73
多项式	训练样本	0.74	34.34	23.06
	验证样本	0.78	52.33	33.54
RBF	训练样本	0.95	13.09	11.65
	验证样本	0.85	29.24	23.96
sigmoid	训练样本	0.80	24.91	17.79
	验证样本	0.77	35.91	25.71

表 6 蓄积量反演模型评估

Table 6 Evaluation of growing stock volume inversion model				
模型	样本	R^2	RMSE/($\text{m}^3\cdot\text{hm}^{-2}$)	MAE/($\text{m}^3\cdot\text{hm}^{-2}$)
SR	训练样本	0.91	17.03	13.27
	验证样本	0.82	33.33	24.03
PLSR	训练样本	0.90	17.63	13.49
	验证样本	0.80	34.53	27.76
RBF-SVR	训练样本	0.94	13.09	11.65
	验证样本	0.85	29.24	23.96
RFR	训练样本	0.95	12.64	8.00
	验证样本	0.88	28.11	19.48

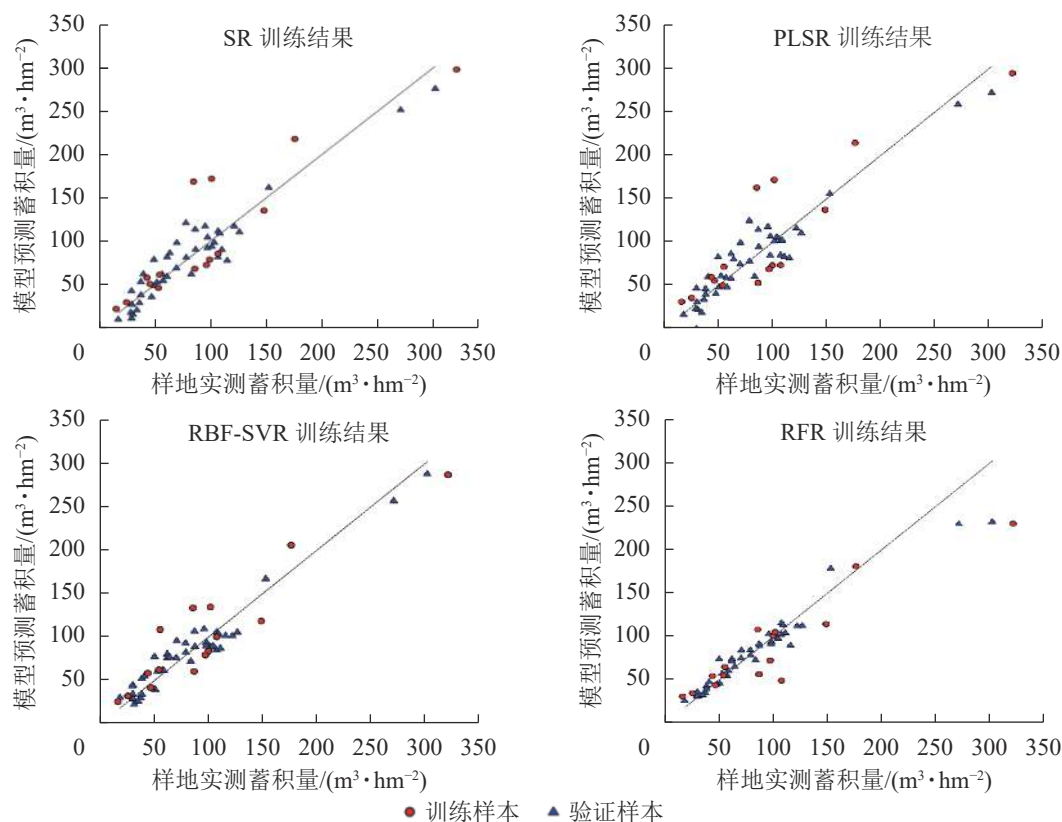


图5 各模型训练结果与验证结果散点图

Figure 5 Scatter diagram of training results and verification results of each model

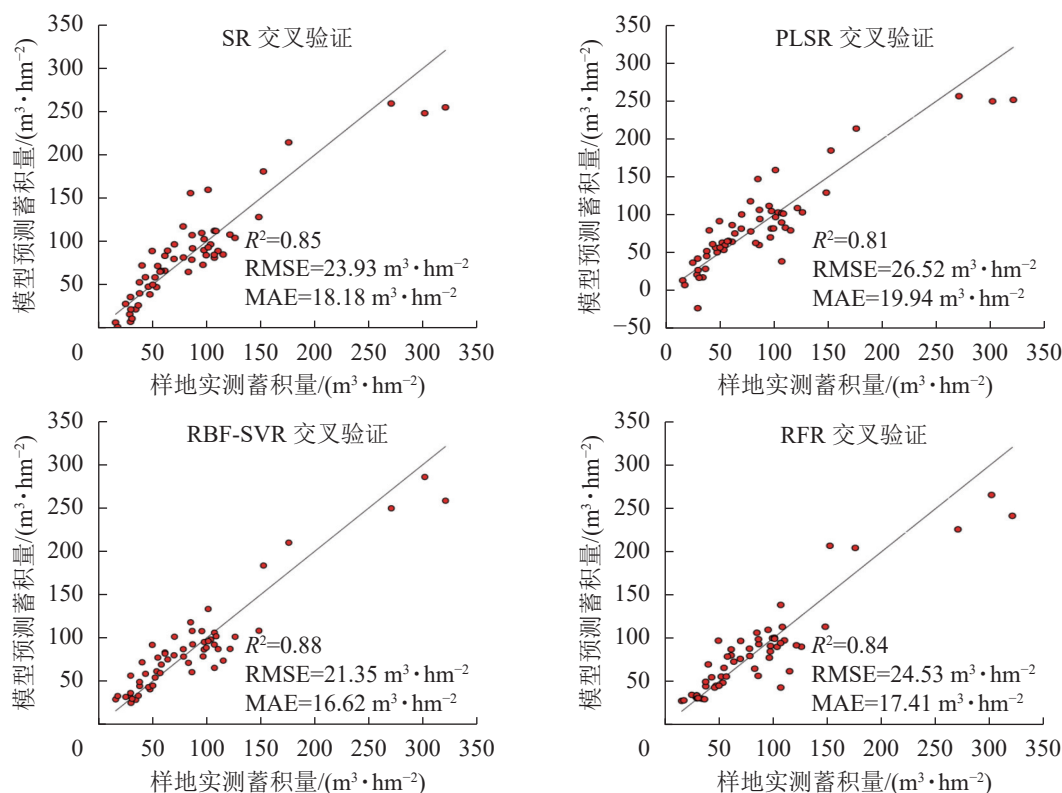


图6 各模型十折交叉验证结果

Figure 6 Ten-fold cross validation results of each model

型在解决林业激光雷达领域的回归预测问题已有广泛应用,肖越^[20]结合 Landsat 8、高分 2 号 (GF-2) 共 2 种光学遥感数据与极化 SAR 数据进行旺业甸林场尺度的蓄积量反演,采用 SVR 模型拟合精度最高 (R^2 为 0.48,

RMSE 为 $57.27 \text{ m}^3 \cdot \text{hm}^{-2}$), 与 MLR、RFR 等方法相比, ΔR^2 为 0.10~0.11、 ΔRMSE 为 $5.30 \sim 6.00 \text{ m}^3 \cdot \text{hm}^{-2}$ 。赵勋等^[21] 基于机载激光雷达进行广西高峰林场林分平均树高估测中, 通过随机森林特征变量筛选, 采用 SVR 相比较 RFR 方法, R^2 增加 0.01、RMSE 减少 0.06 m。以上研究结果与本研究一致, 进一步表明解决基于 LiDAR 点云特征变量的林分蓄积量反演问题时, 结合随机森林筛选特征变量与支持向量机回归可作为有效建模方法。

②本研究采用 2 种方法对特征变量进行筛选, 逐步回归法保证入选的特征变量显著且各变量之间不存在共线关系, 随机森林则通过计算各特征变量对建模时的贡献度排序进而筛选变量。肖越^[20] 利用逐步回归法从遥感特征变量中筛选出建模因子用于蓄积量回归模型构建; 周蓉等^[22] 在基于 Landsat 8 遥感影像反演地上生物量的研究中, 采用随机森林重要性排序方法从遥感因子中筛选出特征变量用于构建模型。这 2 种方法均广泛应用于构建模型前的变量筛选, 合理地选择模型筛选方法, 更有利于回归模型构建。

③本研究旨在选取最优桉树蓄积量反演模型, 参数方法 SR 中逐步筛选法已对变量进行筛选, 故 SR 模型不再采用随机森林进行特征变量筛选。在本研究模型评估中, SR 模型评估结果均优于 PLSR 模型, 袁钰娜等^[23] 对东北林区 4 种不同针叶林蓄积量反演所建立的回归模型中, PLSR 模型的拟合结果和预测精度均优于 SR (ΔR^2 为 0.05~0.15, ΔRMSE 为 2.6%~4.2%), 与本研究结果不一致, 但由于 2 种回归方法并未使用同一变量筛选方法, 因此, 不能直接认为在解决预测回归问题时前者更加可靠, 对于以上 2 种回归方法的模型性能与预测能力需进一步讨论。

④本研究在林场尺度通过较少样本进行桉树林分蓄积量反演, 非参数方法总体预测精度较高且 RBF-SVR 模型略优于 RFR 模型。SVR 核心思想为将低维空间中的向量用非线性函数映射到一个高维特征空间, 进而寻求线性回归超平面并解决低维空间中的非线性问题^[24], 支持向量机回归解决多维度小样本回归预测问题相较随机森林回归更有优势。在反演尺度较大且训练样本较多的林分蓄积量反演问题时, 采用支持向量机回归方法训练效率将明显下降, 此时选择随机森林回归将更适用。

⑤本研究采用逐步筛选法筛选出的 4 个特征变量中, 点云密度变量被选择 1 次; 在采用随机森林筛选的 10 个变量中, 点云密度变量被选择 2 次, 且密度变量 D_9 在重要性排序中位于第 3 位。点云密度变量作为本研究中重要解释变量之一, 该类型变量能够描述林分中树木的水平结构信息, 与树高(垂直结构变量)等参数结合可解释林分空间结构信息。孙忠秋等^[25] 加入林分郁闭度作为水平结构的解释变量, 与仅用点云高度参数相比较, RMSE 下降 $0.27 \text{ m}^3 \cdot \text{hm}^{-2}$ 、MAE 下降 $0.08 \text{ m}^3 \cdot \text{hm}^{-2}$, 该变量对模型预测性能有一定提升。本研究将郁闭度作为候选变量, 但在逐步筛选法中又将其剔除, 因为在基于随机森林法的重要性排序中, 林分郁闭度参数在 48 个候选变量中排名第 44 位。由此可见, 在林业激光雷达应用中, 将林分密度用于衡量林分水平结构信息更加可靠。

5 结论

本研究通过 2 种参数回归方法(逐步回归、偏最小二乘回归)与 2 种非参数方法(随机森林回归、支持向量机回归)对林分蓄积量进行估测。在采用传统方法划分训练样本与验证样本的结果中, 非参数方法模型精度均优于参数回归方法; 采用留一法对各模型预测精度进行交叉验证时, 表现最优模型 RBF-SVR 属于非参数回归方法, 表明解决林业激光雷达领域中的回归预测问题时, 非参数方法相较参数方法更有优势。本研究中蓄积量反演模型已在广西高峰林场内取得较好模型精度, 但对于大尺度反演森林蓄积量的适用性有待进一步论证。

6 参考文献

- [1] 张瑜, 陈存友, 胡希军. 应用投影寻踪分类技术的森林生态功能评价[J]. 浙江农林大学学报, 2020, 37(2): 243 - 250.
ZHANG Yu, CHEN Cunyou, HU Xijun. Evaluation of forest ecological function based on projection pursuit classification [J]. J Zhejiang A&F Univ, 2020, 37(2): 243 - 250.
- [2] 程武学, 杨存建, 周介铭, 等. 森林蓄积量遥感定量估测研究综述[J]. 安徽农业科学, 2009, 37(16): 7746 - 7750.
CHEN Wuxue, YANG Cunjian, ZHOU Jieming, et al. Research summary of forest volume quantitative estimation based on remote sensing technology [J]. J Anhui Agric Sci, 2009, 37(16): 7746 - 7750.

- [3] 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 Ecol Sin*, 1996, **16**(5): 497 – 508.
- [4] 郭庆华, 刘瑾, 陶胜利, 等. 激光雷达在森林生态系统监测模拟中的应用现状与展望[J]. 科学通报, 2014, **59**(4): 459 – 478.
GUO Qinghua, LIU Jin, TAO Shengli, *et al.* Perspectives and prospects of LiDAR in forest ecosystem monitoring and modeling [J]. *Chin Sci Bull*, 2014, **59**(4): 459 – 478.
- [5] 李增元, 刘清旺, 庞勇. 激光雷达森林参数反演研究进展[J]. 遥感学报, 2016, **20**(5): 1138 – 1150.
LI Zengyuan, LIU Qingwang, PANG Yong. Review on forest parameters inversion using LiDAR [J]. *J Remote Sensing*, 2016, **20**(5): 1138 – 1150.
- [6] 徐文兵, 高飞, 杜华强. 几种测量方法在森林资源调查中的应用与精度分析[J]. 浙江林学院学报, 2009, **26**(1): 132 – 136.
XU Wenbing, GAO Fei, DU Huaqiang. Application and precision analysis of several surveying methods in forest resources survey [J]. *J Zhejiang For Coll*, 2009, **26**(1): 132 – 136.
- [7] 李春干, 代华兵. 中国森林资源调查: 历史, 现状与趋势[J]. 世界林业研究, 2021, **34**(6): 72 – 80.
LI Chungan, DAI Huabing. Forest management inventory in china: history, current status and trend [J]. *World For Res*, 2021, **34**(6): 72 – 80.
- [8] PAWE H, PIOTR T, PIOTR W. Area-based estimation of growing stock volume in Scots pine stands using ALS and airborne image-based point clouds [J]. *Forestry*, 2017(5): 686 – 696.
- [9] CHIRICI G, GIANNETTI F, MCROBERTS R E, *et al.* Wall-to-wall spatial prediction of growing stock volume based on Italian National Forest Inventory plots and remotely sensed data[J/OL]. *Int J Appl Earth Observation Geoinf*, 2019, **84**: 101959[2021-12-30]. doi: 10.1016/j.jag.2019.101959.
- [10] 陈松, 孙华, 吴童, 等. 基于 Sentinel-2 与机载激光雷达数据的误差变量联立方程组森林蓄积量反演研究[J]. 中南林业科技大学学报, 2020, **40**(12): 44 – 53.
CHEN Song, SUN Hua, WU Tong, *et al.* Study on the forest volume inversion based on the simultaneous equations of error variables of Sentinel-2 and airborne Lidar data [J]. *J Cent South Univ For Technol*, 2020, **40**(12): 44 – 53.
- [11] 曾伟生, 孙乡楠, 王六如, 等. 基于机载激光雷达数据的森林蓄积量模型研建[J]. 林业科学, 2021, **57**(2): 31 – 38.
ZENG Weisheng, SUN Xiangnan, WANG Liuru, *et al.* Development of forest stand volume models based on airborne laser scanning data [J]. *Sci Silv Sin*, 2021, **57**(2): 31 – 38.
- [12] 黄道年, 廖泽钊. 广西桉树二元材积表编制的研究[J]. 广西农学院学报, 1986(1): 53 – 60.
HUANG Daonian, LIAO Zezhao. Study on compiling binary volume table of *Eucalyptus* in Guangxi [J]. *J Guangxi Agric Coll*, 1986(1): 53 – 60.
- [13] 肖武, 陈佳乐, 笪宏志, 等. 基于无人机影像的采煤沉陷区玉米生物量反演与分析[J]. 农业机械学报, 2018, **49**(8): 169 – 180.
XIAO Wu, CHEN Jiale, DA Hongzhi, *et al.* Inversion and analysis of maize biomass in coal mining subsidence area based on UAV images [J]. *Trans Chin Soc Agric Mach*, 2018, **49**(8): 169 – 180.
- [14] 于雷, 洪永胜, 耿雷, 等. 基于偏最小二乘回归的土壤有机质含量高光谱估算[J]. 农业工程学报, 2015, **31**(14): 103 – 109.
YU Lei, HONG Yongsheng, GENG Lei, *et al.* Hyperspectral estimation of soil organic matter content based on partial least square regression [J]. *Trans Chin Soc Agric Eng*, 2015, **31**(14): 103 – 109.
- [15] 刘琼阁, 彭道黎, 涂云燕. 基于偏最小二乘回归的森林蓄积量遥感估测[J]. 中南林业科技大学学报, 2014, **34**(2): 81 – 84, 132.
LIU Qiongge, PENG Daoli, TU Yunya. Estimation of forest stock volume based on partial least squares regression [J]. *J Cent South Univ For Technol*, 2014, **34**(2): 81 – 84, 132.
- [16] VAPNIK V, IZMAILOV R. Intelligent learning: similarity control and knowledge transfer [J/OL]. *Ann Math Artif Intell*, 2017(1/2)[2022-01-01]. doi: 10.1007/978-3-319-17091-6_1.
- [17] BAO Yukun, LUI Zhitao. A fast grid search method in support vector regression forecasting time series[R]// [s.l.] *Intelligent Data Engineering and Automated Learning Vol 4244*. Berlin Heidelberg: Springer, 2006: 504 – 511.
- [18] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, **26**(3): 32 – 38.

- FANG Kuangnan, WU Jianbin, ZHU Jianping, *et al.* A review of random forest method research [J]. *Stat Inf Forum*, 2011, **26**(3): 32 – 38.
- [19] MOLINARO A M, SIMON R, PFEIFFER R M. Prediction error estimation: a comparison of resampling methods[J/OL]. *Bioinformatics*, 2005, **21**(15): 3301-7[2021-012-31]. doi: 10.1093/bioinformatics/bti499.
- [20] 肖越. 基于多源遥感数据的旺业甸林场森林蓄积量估测方法研究[D]. 长沙: 中南林业科技大学, 2021.
XIAO Yue. *Research on Estimation Method of Forest Volume of Wangyedian Forest Farm based on Multi-Source Remote Sensing Data* [D]. Changsha: Central South University of Forestry & Technology, 2021.
- [21] 赵勋, 岳彩荣, 李春干, 等. 基于机载 LiDAR 数据估测林分平均高[J]. 林业科学研究, 2020, **33**(4): 59 – 66.
ZHAO Xun, YUE Cairong, LI Chungan, *et al.* Estimation of forest stand mean height based on airborne lidar point cloud data [J]. *For Res*, 2020, **33**(4): 59 – 66.
- [22] 周蓉, 赵天忠, 吴发云. 基于 Landsat 8 遥感影像的地上生物量模型反演研究[J]. *西北林学院学报*, 2022, **37**(2): 186 – 192.
ZHOU Rong, ZHAO Tianzhong, WU Fayun. Aboveground biomass model based on landsat 8 remote sensing images [J]. *J Northwest For Univ*, 2022, **37**(2): 186 – 192.
- [23] 袁钰娜, 彭道黎, 王威, 等. 利用机载激光雷达技术估测东北林区典型针叶林的蓄积量[J]. 应用生态学报, 2021, **32**(3): 836 – 844.
YUAN Yuna, PENG Daoli, WANG Wei, *et al.* Estimating standing stocks of the typical conifer stands in Northeast China based on airborne LiDAR data [J]. *Chin J Appl Ecol*, 2021, **32**(3): 836 – 844.
- [24] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. *电子科技大学学报*, 2011, **40**(1): 2 – 10.
DING Shifei, QI Bingjuan, TAN Hongyan. An overview on theory and algorithm of support vector machines [J]. *J Univ Electron Sci Technol China*, 2011, **40**(1): 2 – 10.
- [25] 孙忠秋, 高金萍, 吴发云, 等. 基于机载激光雷达点云和随机森林算法的森林蓄积量估测[J]. *林业科学*, 2021, **57**(8): 68 – 81.
SUN Zhongqiu, GAO Jinping, WU Fayun, *et al.* Estimating forest stock volume via small-footprint LiDAR point cloud data and random forest algorithm [J]. *Sci Silv Sin*, 2021, **57**(8): 68 – 81.