

基于机载激光雷达和机器学习算法的林分平均胸径遥感估测

唐佳俊 柴宗政

(贵州大学林学院 贵阳 550025)

摘要:【目的】针对森林调查中较大范围估测林分平均胸径精度不高的问题,提出了一种林分平均胸径遥感估测新方法。

【方法】选取贵州省桂花国有林场马厂工区为研究区域,基于机载激光雷达点云数据结合样地调查数据,提取样地水平冠层特征变量和高度特征变量,使用方差膨胀因子分析和皮尔逊相关性检验进行变量筛选,构建基于机载激光雷达和机器学习算法的林分平均胸径估测模型。【结果】研究表明:(1)通过方差膨胀因子分析和皮尔逊相关性检验,点云特征变量如平均冠层高度和高度偏态与林分平均胸径具有较好的相关性。其中,平均冠层高度对模型构建的贡献最大,与林分平均胸径呈显著正相关。(2)在比较多元线性回归和机器学习模型后,我们发现机器学习模型优于多元线性回归模型,其中随机森林模型的拟合效果最好。随机森林模型的决定系数(R^2)为0.71,均方根误差(RMSE)为2.50。(3)通过四种森林类型林分平均胸径预测值与实际值的差值进一步证实随机森林模型精度最高,拟合效果最好。【结论】综上,利用机载激光雷达点云数据提取点云特征变量,并构建基于机器学习算法的林分平均胸径估测模型是可行的。该方法的精度满足了森林资源调查的应用需求,可作为辅助林业调查工作的技术手段。关键词:机载激光雷达;林分平均胸径;随机森林回归;点云特征变量

Remote sensing estimation of average diameter at breast height of forest stands based on airborne LiDAR and machine learning algorithms

Abstract:【Objective】A new method for remotely estimating the average diameter at breast height (DBH) of forest stands was proposed to address the problem of low accuracy in large-scale estimation in forest surveys. 【Method】Selecting the Machang work area of Guihua State owned Forest Farm in Guizhou Province as the research area, based on airborne LiDAR point cloud data combined with sample plot survey data, the horizontal canopy characteristic variables and height characteristic variables of the sample plot were extracted. Variance expansion factor analysis and Pearson correlation test were used to screen variables, and a forest average diameter estimation model based on airborne LiDAR and machine learning algorithm was constructed. 【Result】The research results show that: (1) By analyzing the variance inflation factor and conducting Pearson correlation tests, we selected point cloud feature variables that have a good correlation with the DBH of forest stands, such as average canopy height and height skewness. Among them, the average canopy height contributes the most to the model construction and shows a significant positive correlation with the average DBH of forest stands, indicating its important influence on the variation of forest DBH. (2) After comparing multiple linear regression and machine learning models, we found that machine learning models outperform multiple linear regression models, with the random forest model achieving the best fit. The coefficient of determination (R^2) for the random forest model is 0.71, and the root mean square error (RMSE) is 2.50. (3) The difference between the predicted average diameter at breast height (DBH) and the actual values further confirms that the random forest model has the highest accuracy and the best fitting effect among the four forest types. 【Conclusion】In conclusion, it is feasible to use airborne LiDAR point cloud data to extract point cloud feature variables and build a model for remotely estimating the average DBH of forest stands based on machine learning algorithms. The accuracy of this method meets the application requirements of forest resource surveys and can serve as a technical means for assisting forestry survey work.

Key words: airborne LiDAR; average diameter at breast height of the forest stand; random forest regression; point cloud feature

variables

胸径是林业调查中重要的测树因子之一,在获取单木或林分尺度材积/蓄积、生物量、碳储量估算中具有广泛应用 ADDIN (李海奎等, 2011)。通常胸径测量多以人力深入林地进行地面样地抽样调查。然而,这种实地调查存在工作量大、周期长、效率低等问题且容易受到调查人员主观经验的影响。此外,由于采用抽样方式进行野外调查,只能测量有限的林分参数,无法准确描述较大空间尺度的林分结构信息 (Andersen et al., 2005)。为了解决这些问题,遥感技术被引入到森林资源的监测调查体系中,并得到广泛应用 (Balduzzi et al., 2011; 李增元等, 2021)。

机载激光雷达 (Airborne Laser Scanning, ALS) 已被证明能够有效确定单木位置和提取树高。然而,由于其自上而下的扫描方式以及冠层的遮挡,林木下层点云数量不足,无法直接提取单木胸径。因此,单木胸径常通过树高进行推导获取 (韦雪花, 2013)。近年来,基于遥感因子的胸径模型因其能高效获得区域的单木胸径或林分平均胸径引起国内外学者广泛关注。吕金城等 (2021) 在冠层高度模型 (Canopy Height Model, CHM) 的基础上通过点云分类与克里金插值法提取天山云杉树高,并与实测胸径建立幂函数模型,精度达到 0.73,并将其用于后续生物量模型构建,拟合效果显著。Swayze 等 (2022) 提出了一种使用胸径处的点云切片进行圆形拟合提取树木胸径的方法。该方法将提取的胸径与树高进行匹配,并根据胸径与树高的关系估测缺失的单木胸径值,平均误差为 0.79 cm。Mao 等 (2023) 结合了机载激光雷达、可见光影像以及地基激光雷达 (Terrestrial Laser Scanning, TLS) 进行单木分割和单木树冠提取。他们计算了纹理、光谱参数和点云参数与胸径的相关性,并比较了不同机器学习模型估测胸径的能力。结果显示,基于遥感数据及机器学习模型在单木胸径估测方面具有较大潜力。基于区域尺度的林分参数提取适用于较低点云密度,可以在数据采集处理过程中降低成本。Lefsky 等 (1999) 使用机载激光雷达特征变量估计了花旗松样地水平的平均胸径。Hawbaker 等 (2010) 比较了使用所有激光雷达 (Light Detection And Ranging, LiDAR) 回波和仅使用首次回波估测区域林分参数的方法。研究表明,仅使用首次回波的模型对变量更具解释性,并成功估测了北美东部硬阔混交林样地水平的平均胸径。Zhang 等 (2023) 提取了激光雷达高度变量和冠层变量用于估测胸径,通过筛选最优树高胸径拟合函数,将特征变量代入反函数中进行样地水平的胸径估测。该模型的精度相较传统随机森林模型得到显著提高。

本研究以贵州省毕节市织金县桂花林场马厂工区为研究区域,基于机载激光雷达点云数据,通过对比多元线性回归及多种机器学习算法,探索了林分平均胸径最优模型的构建方法。旨在实现对区域尺度林分平均胸径精准估测,并提供一种高效、精准且低成本的森林调查资源方法,以提高森林质量。

1 研究区与数据

1.1 研究区概况

研究区位于贵州省织金县国有桂花林场,境内岩溶发育,喀斯特地貌特征突出。地理坐标分布范围在 26°38'21"—26°39'18"N, 105°44'55"—105°45'5"E,主要地貌类型为山地及丘陵,地形地势起伏较大,峰丛、槽谷、洼地相攘其间。土壤母质以砂页岩、石灰岩为主,土壤类型有黄棕壤、黄壤和石灰土,其中黄壤是主要土壤类型。气候属于亚热带东南季风气候,年平均气温 11.6~15 °C,全年日照 1170 小时左右,年降雨量 1200~1500 mm,无霜期 240~290 天,气候温和,少有冻害,夏季雨量充沛,云雾天气发生频繁。本研究以桂花林场马厂工区为研究区域,筛选森林类型多样、集中分布的区域,面积约为 2000 余亩,48 个小班。研究区地理位置及样地分布图见图 1。

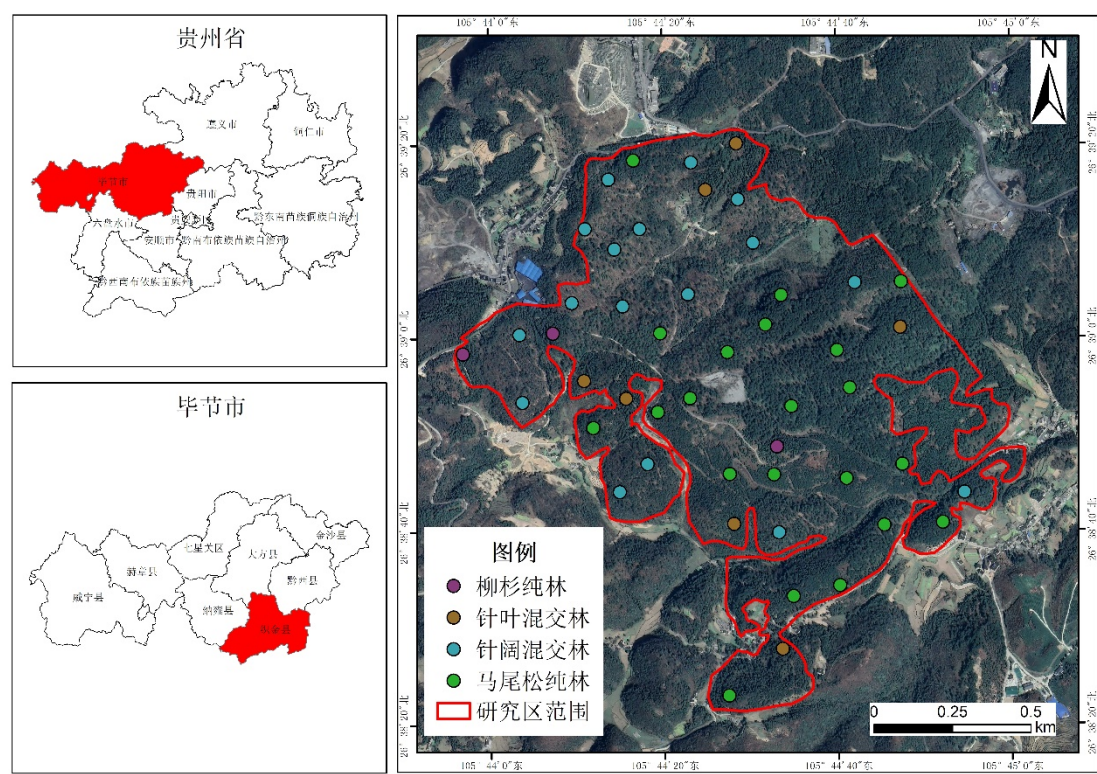


图 1 研究区地理位置及样地分布图

Fig. 1 Geographical Location and Sample Plot Distribution of the Study Area

1.2 样地布设及人工调查

野外调查数据采集于 2022 年 7 月。根据前期收集和详细了解研究区森林资源信息，按照系统布设原则，每隔 200 米布设一个样地，共布设 48 个圆形样地，样地的半径为 14.57 m，面积为 667 m²。对其中胸径（DBH）≥5 cm 的单木，依次编号挂牌并进行每木检尺。使用胸径尺测量胸径；布鲁莱斯测高器测量树高；激光测距仪站在树冠投影下部测量东南西北四个方向的冠幅；RTK 测量单木的脚点。

表 1 样地概况

Tab. 1 Overview of sample plot

森林类型	样地数量	平均胸径 (cm)	平均树高 (m)	林分蓄积 (m ³)
柳杉纯林	3	20.05	15.05	56.02
针叶混交林	7	19.84	14.35	106.03
针阔混交林	17	21.88	16.43	323.13
马尾松纯林	21	26.38	17.70	486.97

1.3 激光雷达数据获取

机载 LiDAR 数据获取时间为 2022 年 7 月，利用大疆经纬 M300RTK 无人机搭载禅思 L1 传感器进行仿地飞行获取坐标系为 CGCS2000/3-degree Gauss-Kruger CM 105E 的激光雷达点云数据，与地面调查数据坐标系一致，方便后续处理以及与地面数据匹配。在晴朗少云的天气进行，根据研究区地形特点将研究区划分为 8 个摄区，防止因地形及距离过远导致无人机与遥控器的图传信号缺失，利用 30 m 分辨率的数字高程模型（Digital Elevation Model, DEM）进行仿地飞行，相对航高 80 m，旁向重叠率 50 %，飞行速度 6 m/s，三回波模式，采样频率 160 k/s，扫描模式为重复扫描，高程精度 0.05 m，平面精度 0.1 m。

2 研究方法

2.1 LiDAR 数据处理

首先对获取到的机载激光雷达数据进行去噪处理，点云数据采用改进的渐进三角网滤波算法（Zhao e

t al., 2016) 划分地面点和植被点两大类。使用不规则三角网插值法对地面点插值生产分辨率为 0.5 m 的 DEM, 再针对 DEM 对点云数据进行归一化处理, 根据样地坐标提取所有样地的点云。

2.2 LiDAR 特征变量提取

根据现有研究表明, 第一回波更加稳定, 计算森林参数多采用首次回波 (Næsset et al., 2008; 庞勇等, 2012), 通过提取首次回波的点云高度相关特征变量及冠层相关指标能够为估测林分平均胸径提供有效的信息 (Lefsky et al., 1999; Zhang et al., 2023; Zhang et al., 2019)。由此, 使用 R 语言 lidR 包、LiDAR360 软件提取了两组样地水平林分特征变量高度变量和冠层特征变量, 其中冠层特征变量包括叶面积指数 (leaf area index, LAI)、平均冠层高度 (mean of canopy height, mCH)、冠层高度标准差 (standard deviation of canopy height, sdCH) 等 6 个变量, 高度变量包括高度平均绝对偏差、累积高度百分位数、高度偏态、高度百分位数等 46 个变量, 参数设置为消光系数为 0.5, 高度阈值为 2 m, 去除高度值低于 2 m 的灌草、枯落物点云。部分特征变量计算原理如下:

2.2.1 冠层特征变量

叶面积指数: 指单位地表面积上所有叶片表面积的一半, 是评估冠层结构最重要的指标之一。其基于 LiDAR 点云提取 LAI 的计算公式为:

$$LAI = \frac{\cos(ang) * \ln(GF)}{k}$$

式中, ang 是平均扫描角, GF 是间隙率, k 是消光系数。

平均冠层高度: 指 Z 值大于高度阈值的表层树冠点云的平均高度。

冠层高度标准差: 指 Z 值大于高度阈值的表层树冠点云高度的标准差。

2.2.2 高度特征变量

平均绝对偏差: 样地点云内点高度的离散程度

$$V = \frac{\sum_{i=1}^n (|Z_i - \bar{Z}|)}{n}$$

式中, Z_i 为样地第 i 个点的高程值, $\bar{Z}$ 为样地所有点的平均高度, n 为样地内总点数。

累积高度百分位数: 样地单元内, 所有归一化点按高度进行排序并计算所有点的累积高度。

高度百分位数: 样地单元内, 所有归一化点按高度进行排序, 然后计算样地内 X% 的点所在高度。

高度偏态: 样地内所有点 Z 值分布的对称性。

$$Skewness = \frac{\frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^3}{a^3}$$

式中: Z_i 为样地第 i 个点的高程值, $\bar{Z}$ 为样地所有点的平均高度, n 为样地内总点数, a 为样地内点云高度分布的标准差。

2.3 LiDAR 特征变量选取

本研究通过进行方差膨胀因子 (variance inflation factor, VIF) 分析来避免特征变量间存在的多重共线性问题, 保留 VIF 值低于 10 的指标 (常晓敏, 2021)。分别计算保留特征变量间的皮尔逊 (Pearson) 相关系数, 皮尔逊相关系数证实, VIF 确定的保留特征变量对的 |r| 均小于 0.8。再计算实地测量胸径平均值与保留变量的相关系数, 去除 |r| 小于 0.2 的特征变量。用筛选所得特征变量进行后续模型构建。特征变量选取及后续模型构建均基于 R 完成。

2.4 林分平均胸径估测模型

多元线性回归 (Multiple Linear Regression, MLR): 多元线性回归在遥感数据定量估测森林参数问题中被广泛应用, 通过提取点云数据的特征变量为自变量, 以样地实测所得到的森林参数为因变量建立多元线性回归方程。多元线性回归的一般公式为:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_ix_i + \dots + b_nx_n$$

式中, Y 为样地平均胸径, a 为回归常数, b_i 为回归系数, $i=1, 2, 3, \dots, n$, x_i 是特征变量。

随机森林模型(Random forest, RF): 随机森林由众多决策树构成, 是一种能够同时胜任分类与回归任务的非参数方法, 因为比基于模型的方法更稳定而被广泛使用 (Belgiu et al.,2016)。随机森林的构建有两个关键参数需要设定, 一个是决策树数量(ntree), 另一个是节点处随机变量数量(mtry), 决策树数量一般设置为默认值 500, 随机变量数量是指节点中用于二叉树的变量个数, 一般取值为变量个数的三分之一。

支持向量机(Support vector machine, SVM): 支持向量机的核心思想即通过定义不同核函数将输入空间非线性映射在一个高维特征空间中, 在高维空间构造最优线性回归超平面作为决策面 (林卓等, 2015)。在模型构建过程中, 选择不同核函数类型会直接影响到模型的预测性能。

最近邻算法 (K-nearest neighbors, KNN): 最近邻算法的思想非常简单直观, 即如果某一样本在特征空间中的 k 个最邻近的样本大多数属于同一类别, 则该样本也属于这一类别 (曹庆先, 2010)。需要设置参数有最近邻个数 k 以及计算数据的距离类型, 包括欧氏距离、曼哈顿距离等, 确定适当的 k 值可以使模型更加精准。

2.5 模型评价

本研究通过十折交叉验证分析评估模型的性能, 其将数据集划分十折, 使用其中一折作为验证, 其余作为训练数据反复迭代, 在多次评估结果中求取平均值作为模型最后评分。本研究采用决定系数(R²)和均方根误差(RMSE)评价胸径估测模型的精度。计算公式如下:

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

式中: $\hat{y}_i$ 为模型预测值, y_i 为样地测量值, $\bar{y}_i$ 为样地测量均值, n 为样地数量。

3 结果与分析

3.1 变量选取

为提高模型精度, 使模型更加精简, 将点云数据提取的 52 个特征变量进行方差膨胀因子分析去除变量间多重共线性, 筛选 VIF<10 的变量再与样地平均胸径进行相关性分析 (图 2)。结果表明: LAI、CC 与 DBH 相关性低且不显著, DBH 与 sdCH、mCH、H_m、H_k、H_{A1}呈显著正相关, 与H_s呈显著负相关。因此, 选择H_m、H_k、H_{A1}、H_s、sdCH、mCH 作为模型输入变量, DBH 作为输出变量。

表 2 经方差膨胀因子分析筛选后的特征变量汇总表
Tab. 2 Summary table of characteristic variables screened through square difference expansion factor analysis

变量类别	变量名称	变量描述	简写
高度变量	1%分位数累计高度	激光返回点云 1%的点所在累计高度	H _{A1}
	高度中位数	所有点高度值的中位数	H _m
	高度峰度	所有点高度值的平坦程度	H _k
	高度偏态	所有点高度分布的对称程度	H _s
	平均冠层高度	大于高度阈值的所有点高度平均值,指树冠平均高度	mCH
冠层特征变量	冠层高度标准差	林分冠层高度的标准差	sdCH
	郁闭度	冠层垂直投影占林地面积百分比	CC
	叶面积指数	叶片表面积的一半	LAI

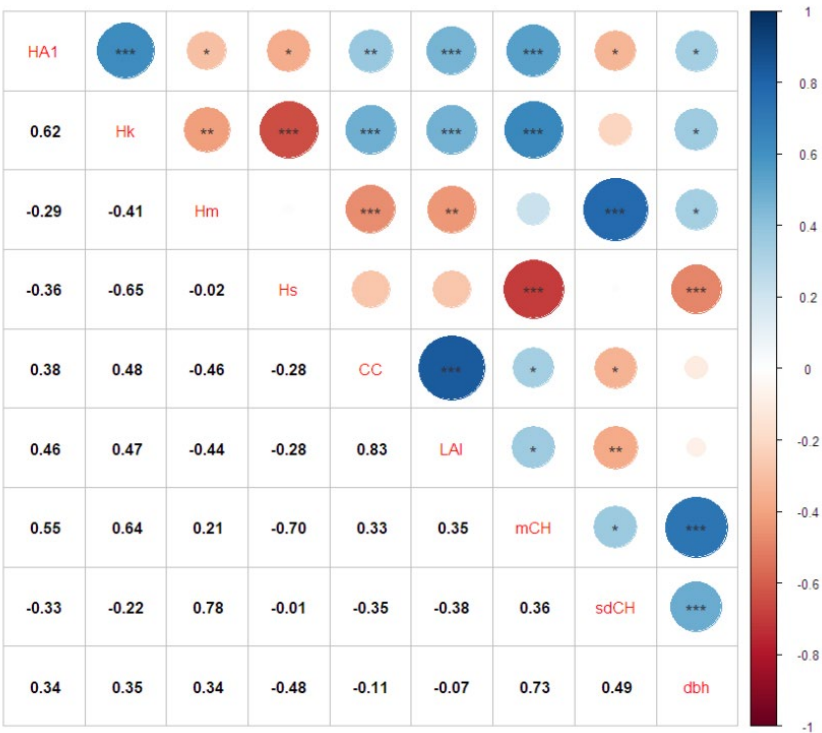


图 2 点云特征变量与林分平均胸径相关分析图

Fig. 2 Correlation analysis between point cloud characteristic variables and average diameter at breast height of forest stands

3. 2 模型预测结果

多元线性回归：调查圆形样地 48 个，平均胸径处于 10.7819~43.3684cm，均值 23.4348cm。使用实地测量平均胸径拟合多元线性回归模型最终形式为 $DBH=2.87+1.12\cdot H_{A1}-0.53\cdot H_k-1.15\cdot H_m-2.92\cdot H_s+3.70\cdot sdCH+0.51\cdot mCH$ ，模型决定系数 R^2 为 0.56，均方根误差RMSE为 2.98。

随机森林：对变量数量及决策树数量进行寻优，ntree 的增加会提高模型精度但也会使模型运算效率降低，mtry 默认设为变量数目的三分之一，由于具体数据不同，仍需对其进行寻优。经过测试，ntree=800、mtry=4 建立随机森林回归模型。基于随机森林模型得到各自变量对 dbh 影响的相对重要性得分如下表所示，冠层特征变量能够提供 73.62%的相对贡献率，高度特征变量贡献了 26.38%。随机森林模型决定系数 R^2 为 0.71，均方根误差RMSE为 2.50。

表 3 随机森林各影响因子的重要值
Tab. 3 Important values of various influencing factors in random forests

变量	重要值/%	排序
mCH	60.94	1
sdCH	12.68	2
H _{A1}	8.90	3
H _k	7.12	4
H _m	5.48	5
H _s	4.88	6

支持向量机：本研究选取了线性内核、多项式内核和径向基内核 3 种核函数构建支持向量机模型。惩罚系数 C 过大或过小时，会有过拟合或欠拟合的风险，通过尝试可能的 C 值构建模型。模型构建结果如下表所示，具有径向基内核的支持向量机模型的确定系数最大，均方根误差最小。

表 4 不同核函数的支持向量机精度评价
Tab. 4 Accuracy evaluation of support vector machines with different kernel functions

核函数类型	决定系数 (R ²)	均方根误差/cm
线性内核	0.66	2.48
多项式内核	0.66	2.49

径向基内核0.672.32

最近邻算法：采用欧氏距离作为距离度量，k 是最近邻算法中的一个超参数，KNN 通过寻找与目标对象欧氏距离最近的 k 个点，取 k 个点的平均值作为目标对象的预测值，k 取值范围一般在 3~10，本文选取 k=10 进行模型构建。最近邻算法结果，决定系数 R^2 为 0.65，均方根误差 RMSE 为 2.31。

3.3 模型拟合精度

将样地平均胸径实测值和预测值进行回归拟合，由图 1 可以看出，基于样地水平提取的特征变量与样地平均胸径间存在较好的相关性。多元线性回归模型、支持向量机模型以及最近邻算法模型的拟合结果较为分散，模型精度较低，其中支持向量机模型的 DBH 预测值与实测值间存在较大误差。通过对比 4 种方法构建的模型精度，随机森林模型的 R^2 为 0.71，均方根误差RMSE为 2.50，其精度较高、拟合效果较好，说明机器学习算法能够较准确地估测样地平均胸径（表 5 和图 3）。

表 5 不同模型拟合效果
Tab. 5 Fitting effects of different models

模型类型	决定系数 (R^2)	均方根误差/cm
多元线性回归 (MLR)	0.56	2.98
随机森林 (RF)	0.71	2.50
支持向量机 (SVM)	0.67	2.32
最近邻算法 (KNN)	0.65	2.31

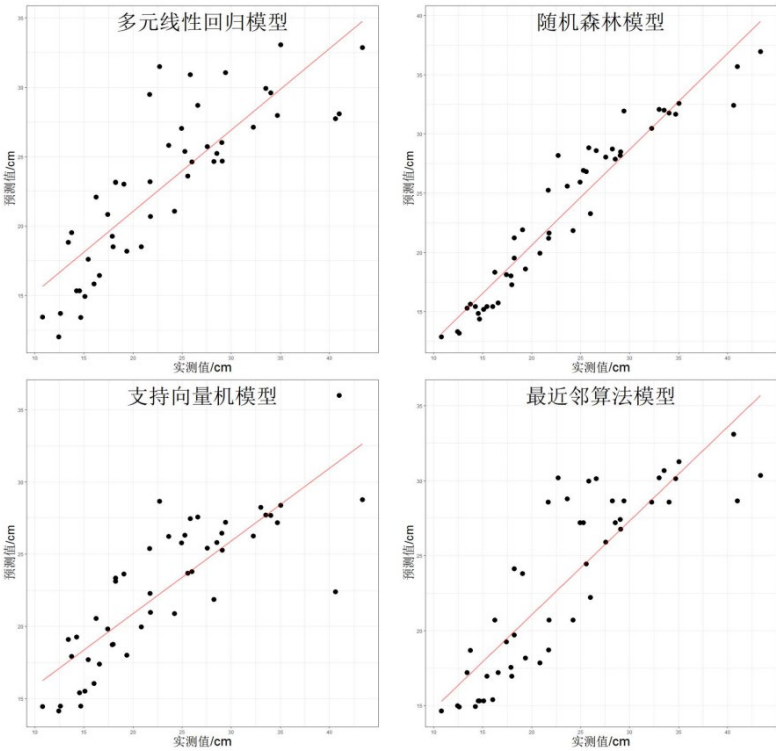


图 3 不同模型预测值与实测值的关系

Fig. 3 The relationship between predicted values of different models and measured values

3.4 基于森林类型的精度验证

从表 6 不同森林类型模型拟合结果表看出，四种模型中，随机森林模型对柳杉纯林、针叶混交林、马尾松纯林的预测结果差值均最小，对针阔混交林预测差值略高于支持向量机；多元线性回归模型在马尾松纯和针阔混交林预测中表现较差，对于另两种森林类型拟合结果较好；支持向量机与最近邻算法对马尾松纯林的预测值与实测值差值较大，对另三种森林类型的拟合效果较好。对于四种森林类型而言，各模型对柳杉纯林与针叶混交林预测结果较好，实测值与预测值间差值较小。通过对比四种模型对四种森林类型预

测值与实际值的差值发现，随机森林模型精度最高，拟合效果最好。

表 6 不同森林类型模型拟合结果
Tab. 6 Fitting results of different forest type models

森林类型	样地数量	实测值(cm)	MLR		RF		SVM		KNN	
			预测值(cm)	差值(cm)	预测值(cm)	差值(cm)	预测值(cm)	差值(cm)	预测值(cm)	差值(cm)
柳杉纯林	3	20.05	19.47	-0.58	20.13	0.08	19.54	-0.51	19.59	-0.46
针叶混交林	7	19.84	20.80	0.96	19.83	-0.01	20.30	0.46	20.10	0.26
针阔混交林	17	21.88	23.31	1.43	22.70	0.82	22.53	0.65	23.21	1.33
马尾松纯林	21	26.38	24.26	-2.12	25.56	-0.82	23.87	-2.51	24.74	-1.64

4 讨论

本研究基于机载激光雷达数据提取特征变量进行相关性分析，发现冠层特征变量和高度特征变量与林分平均胸径呈现不同程度的相关性。其中，选取参与建模的冠层特征变量与林分平均胸径呈极显著正相关，且对模型具有较强的解释能力。此外，同类研究也证实了引入冠层相关参数以精确预测胸径的必要性（付逍遥，2021；黄昕晰，2020；张崢男，2023）。有研究表明，胸径与树高应符合异速生长关系，由于树木的遗传特征和生理特性的影响（Tian et al.,2022）。虽然通过提取树高已经能够满足预测胸径模型的构建，但较多的研究表明，胸径与树高的关系受林分结构和气候等因素的影响较大（Lines et al.,2012；Hulshof et al.,2015）。在树木的幼龄阶段，其主要特征是高度快速生长，随着年龄的增长，高度变化不再显著。此时，树木更多地进行胸径生长。为确保供给胸径生长的光照和水源，在树木高生长放缓后，树冠会随着胸径生长不断横向扩张。因此，在构建胸径预测模型时加入冠层参数是可行的。

在模型中，多元线性回归模型表现最差，其决定系数（R²）和均方根误差（RMSE）分别为 0.56 和 2.98。相比之下，随机森林模型的拟合效果最好，R² 为 0.71，RMSE 为 2.50。这一结果与区分森林类型后各模型预测值与实测值的差异相符（表 6）。在随机森林模型中，平均冠层高度对模型构建贡献最大，与林分平均胸径呈显著正相关，是影响林分胸径变化的重要因素。这与陈玉玲等（2020）基于机器学习方法构建胸径模型的研究结果一致。然而，本研究的模型精度相对较低，可能是因为样本量相对较少且未对树种进行区分。之前的研究已经证明了胸径预测模型在不同树种之间存在显著差异（Cysneiros et al.,2020；Mensah et al.,2018）。胸径与冠层、树高之间的关系并不仅仅是简单的线性关系。相比传统模型构建方法，机器学习不对数据分布做严格要求，能够准确应对高维复杂数据进行模型构建。在机器学习算法中，模型参数的寻优是精确预测林分平均胸径的关键。本研究使用网格搜索穷举了可能的参数组合进行寻优，以期选取到最优模型。

传统的地面胸径调查方法存在许多缺点（Chave et al.,2005；Mokroš et al.,2018），在茂密或者是情况复杂的林分中，地基激光雷达作为能够高精度测量胸径的手段，受到了全球卫星导航系统（global navigation satellite system, GNSS）信号接收困难的限制，从而限制了其应用范围。此外，由于林内遮挡物较多，地基激光雷达也难以胜任大范围的森林参数提取任务。相比之下，机载激光雷达具有更高的效率，受到的约束也较少，因此具有更大的作业范围。作为一种高效提取大范围森林参数的有效方法，机载激光雷达需要优化算法提高模型精度，并采用分树种建模方法来提取林分平均胸径，以构建更具普适性的模型。这将成为进一步研究的重要方面。

5 结论

本研究以贵州省毕节市织金县桂花林场马厂工区为研究区域。利用机载激光雷达点云数据，提取了 4

6 个高度特征变量和 6 个冠层特征变量。以 48 块样地数据为基础,通过变量优选进行模型构建。使用多元线性回归、随机森林、支持向量机和最近邻算法对贵州省毕节市织金县桂花林场林分平均胸径进行预测。研究表明:平均冠层高度和高度偏态等点云特征变量与林分胸径具有显著的相关性。其中,平均冠层高度对模型构建贡献最大,是影响林分胸径变化的重要因素。与多元线性回归模型相比,机器学习模型的效果更好。其中,随机森林模型的拟合效果最佳。因此,基于机载激光雷达点云数据提取点云特征变量估测林分平均胸径的方法具有可行性,能够作为一种技术手段辅助林业调查工作。随着数据来源的不断丰富,未来研究可以侧重于多种机器学习组合模型的应用,以更好地服务森林调查与经营工作。

参考文献

- 李海奎, 法蕾. 2011. 基于分级的全国主要树种树高-胸径曲线模型. 林业科学, 47(10):83-90.
(LI H K, FA L. Height-Diameter Model for Major Tree Species in China Using the Classified Height Method. Scientia Silvae Sinicae, 47(10):83-90. [in Chinese])
- Andersen H, McGaughey R J, Reutebuch S E. 2005. Estimating forest canopy fuel parameters using LIDAR data. Remote sensing of Environment, 94(4):441-449.
- Balduzzi M A, Van der Zande D, Stuckens J, *et al.* 2011. The properties of terrestrial laser system intensity for measuring leaf geometries: a case study with conference pear trees (*Pyrus Communis*). Sensors, 11(2):1657-1681.
- 李增元, 陈尔学. 2021. 中国林业遥感发展历程. 遥感学报, 25(01):292-301.
(LI Z Y, CHEN E X. 2021. Development course of forestry remote sensing in China. National Remote Sensing Bulletin, 25(01):292-301. [in Chinese])
- 韦雪花. 2013. 轻小型航空遥感森林几何参数提取研究. 北京林业大学.
(WEI X H. Research of Forestry Geometrical Parameter Extraction with Light and Small Airborne Remote Sensing System. Beijing Forestry University. [in Chinese])
- 吕金城, 王振锡, 杨勇强, 等. 2021. 基于无人机影像的天山云杉林树高提取及蓄积量的反演. 新疆农业科学, 58(10):1838-1845.
- (LV J C, WANG Z X, YANG Y Q, *et al.* 2021. Height Extraction and Growing Stock Inversion of *Picea schrenkiana* var. *tianshanica* in Tianshan Mountain Based on UAV Image. Xinjiang Agricultural Sciences, 58(10):1838-1845. [in Chinese])
- Swayze N C, Tinkham W T. 2022. Application of unmanned aerial system structure from motion point cloud detected tree heights and stem diameters to model missing stem diameters. MethodsX, 9:101729.
- Mao Z, Lu Z, Wu Y, *et al.* 2023. DBH Estimation for Individual Tree: Two-Dimensional Images or Three-Dimensional Point Clouds?. Remote Sensing, 15(16):4116.
- Lefsky M A, Cohen W B, Acker S A, *et al.* 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. Remote sensing of environment, 70(3):339-361.
- Hawbaker T J, Gobakken T, Lesak A, *et al.* 2010. Light detection and ranging-based measures of mixed hardwood forest structure. Forest science, 56(3):313-326.
- Zhang Z, Wang T, Skidmore A K, *et al.* 2023. An improved area-based approach for estimating plot-level tree DBH from airborne LiDAR data. Forest Ecosystems, 10:100089.
- Zhao X, Guo Q, Su Y, *et al.* 2016. Improved progressive TIN densification filtering algorithm for airborne LiDAR data in forested areas. ISPRS Journal of Photogrammetry and Remote Sensing, 117:79-91.
- Næsset E, Gobakken T. 2008. Estimation of above-and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sensing of Environment, 112(6):3079-3090.
- 庞勇, 李增元. 2012. 基于机载激光雷达的小兴安岭温带森林组分生物量反演[J]. 植物生态学报, 36(10):1095-1105.
(PANG Y, LI Z Y. Inversion of biomass components of the temperate forest using airborne Lidar technology in Xiaoxing'an Mountains, Northeastern of China. Chinese Journal of Plant Ecology, 36(10):1095-1105. [in Chinese])
- Zhang Z, Cao L, Mulverhill C, *et al.* 2019. Prediction of diameter distributions with multimodal models using LiDAR data in subtropical planted forests. Forests, 10(2):125.
- 常晓敏. 2021. 华北北部防护林主要树种立地-生长-结构-功能多元耦合关系研究. 北京林业大学.
(CHANG X M. 2021. Multivariate Coupling Relationship of Site-Age-Structure-Function of Main Tree in Species in Shelter

- rbelts of Northern North China. Beijing Forestry University. [in Chinese])
- Belgiu M, Drăguț L. 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114:24-31.
- 林卓, 吴承祯, 洪伟, 等. 2015. 基于 BP 神经网络和支持向量机的杉木人工林收获模型研究. 北京林业大学学报, 37(01):42-47.
- (LIN Z, WU C Z, HONG W, *et al.* Yield model of *Cunninghamia lanceolata* plantation based on back propagation neural network and support vector machine. Journal of Beijing Forestry University, 37(01):42-47. [in Chinese])
- 曹庆先. 2010. 北部湾沿海红树林生物量和碳贮量的遥感估算. 中国林业科学研究院.
- (CAO Q X. 2010. The estimation of mangrove biomass and carbon storage using remote sensing data in Beibu Gulf coast. Chinese academy of forestry. [in Chinese])
- 付逍遥. 2021. 基于机载和无人机点云数据的平原人工林结构参数动态监测. 南京林业大学.
- (FU X Y. 2021. Dynamic Monitoring of Plain Plantation Structure Parameters Based on Airborne and UAV Point Cloud Data. Nanjing forestry university. [in Chinese])
- 黄昕晰. 2020. 基于无人机影像与 Mask R-CNN 的城市单木检测与参数提取研究. 浙江农林大学.
- (HUANG X X. 2020. Research on Urban Single Tree Detection and Parameters Extraction Based on UAV Image and Mask R-CNN. Zhejiang Agriculture and Forestry University. [in Chinese])
- 张峥男. 2023. 杉木和桉树人工林胸径结构估测和采伐优先等级划分. 南京林业大学.
- (ZHANG Z N. 2023. Estimation of Tree Diameter Structure and Logging Priority Level Zoning in Chinese fir and Eucalyptus Plantations. Nanjing forestry university. [in Chinese])
- Tian D, Jiang L, Shahzad M K, *et al.* 2022. Climate-sensitive tree height-diameter models for mixed forests in Northeastern China. Agricultural and Forest Meteorology, 326:109182.
- Lines E R, Zavala M A, Purves D W, *et al.* 2012. Predictable changes in aboveground allometry of trees along gradients of temperature, aridity and competition. Global Ecology and Biogeography, 21(10):1017-1028.
- Hulshof C M, Swenson N G, Weiser M D. 2015. Tree height - diameter allometry across the United States. Ecology and evolution, 5(6):1193-1204.
- 欧强新. 2019. 基于机器学习的吉林天然针阔混交林生长建模. 中国林业科学研究院.
- (OU X Q. 2019. Modeling Stand Growth of Natural Conifer and Broadleaf Mixed Forests in Jilin Province Based on Machine Learning. Chinese academy of forestry. [in Chinese])
- 朱兆廷, 孙玉军, 梁瑞婷, 等. 2023. 基于树冠和竞争因子的杉木胸径估测. 北京林业大学学报, 45(09):42-51.
- (ZHU Z T, ZHU Y J, LIANG R T, *et al.* 2023. Estimating DBH of *Cunninghamia lanceolata* based on crown and competition factors. Journal of Beijing Forestry University, 45(09):42-51. [in Chinese])
- 陈玉玲. 2020. 人工林适地适树与生长收获效益评估研究. 北京林业大学.
- (CHEN Y L. 2020. Research on Matching Tree Species with Site and Growth Yield Benefit Assessment of Plantation. Beijing Forestry University. [in Chinese])
- Cysneiros V C, Pelissari A L, Gaudi T D, *et al.* Modeling of tree height - diameter relationships in the Atlantic Forest: effect of forest type on tree allometry[J]. Canadian Journal of Forest Research, 2020,50(12):1289-1298.
- Mensah S, Pienaar O L, Kunneke A, *et al.* 2018. Height - Diameter allometry in South Africa's indigenous high forests: Assessing generic models performance and function forms. Forest Ecology and Management, 410:1-11.
- Chave J, Andalo C, Brown S, *et al.* 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia, 145:87-99.
- Mokroš M, Liang X, Surový P, *et al.* 2018. Evaluation of close-range photogrammetry image collection methods for estimating tree diameters. ISPRS International Journal of Geo-Information, 7(3):93.