

DOI:10.13275/j.cnki.lykxyj.2020.04.008

基于机载 LiDAR 数据估测林分平均高

赵 勋¹, 岳彩荣^{1*}, 李春干², 张丽梅³, 谷 雷¹

(1. 西南林业大学林学院, 云南 昆明 650224; 2. 广西大学林学院, 广西 南宁 530004;
3. 西南林业大学园林园艺学院, 云南 昆明 650224)

摘要: [目的] 以 2016 年 9 月广西壮族自治区高峰林场实验区获取的机载 LiDAR 点云数据为基础, 通过提取 30 m×30 m 空间林分尺度下的 LiDAR 点云特征变量实现对林分平均高的估测。[方法] 首先将 105 块实测林分平均高度的样地数据按照 3: 1 的比例随机划分为训练样本 (79) 和检验样本 (26), 采用随机森林回归 (RFR) 和支持向量回归 (SVR) 两种机器学习算法对 79 个训练样本与对应的林分 LiDAR 点云特征变量回归建模。建模方案包括随机森林模型、支持向量机模型及随机森林+支持向量机组合模型。其次利用 26 个检验样本数据评价模型预测精度。最后统计 3 个模型中训练样本和检验样本对应的精度评价指标, 以一个预测精度高、泛化能力强的模型作为最终模型进行林分平均高制图。[结果] 表明: 随机森林模型的训练样本和检验样本的决定系数 (R^2) 分别为 0.886 1 和 0.837 5, 均方根误差 (RMSE) 分别为 1.22 和 1.56; 支持向量机模型的训练样本和检验样本的决定系数 (R^2) 分别为 0.886 4 和 0.840 9, 均方根误差 (RMSE) 分别为 1.21 和 1.54; 组合模型的训练样本和检验样本的决定系数 (R^2) 分别为 0.859 8 和 0.853 2, 均方根误差 (RMSE) 分别为 1.35 和 1.48; [结论] 组合模型的泛化能力及预测精度最好, 支持向量机次之, 最后为随机森林。利用组合模型可有效完成研究区林分平均高制图。

关键词: 机载 LiDAR 点云数据; 随机森林回归; 支持向量回归; 林分平均高; 高峰林场

中图分类号: S771.8

文献标志码: A

文章编号: 1001-1498(2020)04-0059-08

林分平均高是评价森林生产力、计算森林生物量和蓄积量、研究森林碳循环的重要参考指标^[1-3]。传统的森林参数调查中, 需要花费大量人力物力, 在地形复杂、人迹罕至的地方, 甚至无法实现人工地面调查。

20 世纪 80 年代中期, 激光雷达技术被应用于森林参数的获取^[4]。机载 LiDAR 技术由于其具有测量精度高、实时性好、方便灵活等特点被广泛用于森林参数的获取。利用机载 LiDAR 点云数据获取森林参数可根据点云密度大小分为两类: 即利用高密度点云直接获取单木参数, 包括单木树高、冠幅、冠面积和材积等^[5-7]。这一类方法往往需要高

密度的激光点云, 通常点云密度需达 15 点·m⁻² 以上, 数据获取成本较高, 大区域范围难以应用。另一类利用相对稀疏的点云密度获取林分测树因子, 如平均高、郁闭度、蓄积量等林分因子。通过建立机载 LiDAR 点云数据的统计特征和野外样地的林分调查因子的相关模型完成大范围的森林参数制图^[8-9]。这一类方法对机载 LiDAR 点云数据密度要求较低, 通常利用 2 点·m⁻² 以上的低密度 LiDAR 点云即可与样地统计数据建立模型, 完成大区域下森林参数的获取^[10-12]。

近年来, 已有大量研究利用机载 LiDAR 点云数据结合样地调查数据建立模型估测林分参数。这

收稿日期: 2019-09-18 修回日期: 2020-03-05

基金项目: 云南省教育厅项目 (2018JS330); 亚太森林网络 (APFNET/2018P1-CAF)-大湄公河次区域森林可持续发展遥感监测; 国家自然科学基金 (31260156)

* 通讯作者: 岳彩荣, 教授, 博导, 研究方向: 林业遥感。Email: cryue@163.com

些模型主要分为参数模型和非参数模型。参数模型以线性回归为代表, Næsset、穆喜云、庞勇等^[1,4,13]分别对挪威东南部的针叶林、内蒙古大兴安岭生态站典型的森林类型、山东省泰安市徂徕山林场人工林为主的林分利用机载 LiDAR 点云数据建立线性回归模型估测林分平均高都得到了较好的结果。Kwak、Silva、刘浩、徐婷等^[14-17]基于机载 LiDAR 点云数据采用线性回归建立模型估测生物量、蓄积量、平均胸径、胸高断面积等森林参数也都取得了较好的效果。然而, 虽然线性回归模型简单明了, 但以此建立模型对数据要求较高, 它假设数据满足正态性、方差齐性、独立性和线性相关等严苛的条件才能建立模型^[18], 实际应用中数据并非能满足以上的条件。非参数模型主要有支持向量机、随机森林、K 近邻法 (K-NN)、人工神经网络 (ANN)、决策树等, 这些方法都被成功应用于森林参数估测^[19-21]。鲁林等^[22]基于机载 LiDAR 点云数据探索了随机森林算法在估测林分平均高中的适用性。Alberto 等^[23]基于机载 LiDAR 点云数据构建 K 近邻回归模型对位于巴西的火炬松 (*Pinus taeda* L.) 种植园进行了林分平均高和优势高估测, 模型的相关性 (R^2) 分别为 94% 和 90%。Monnet 等^[24]基于机载 LiDAR 点云数据构建支持向量机和普通最小二乘模型对林分优势高进行估计, 结果显示支持向量机和普通最小二乘估测精度相近, 但支持向量机模型精度更高。Tompalski 等^[25]利用最小二乘模型、随机森林、K-NN 3 个模型测试模型的可移植性, 基于不同区域、不同点云特征对断面积加权平均高、断面积平均木直径和总材积进行估测, 显示了机器学习算法在森林参数估测上具有巨大的优势

性。洪奕丰等^[26]基于机载 LiDAR 点云数据对长白落叶松 (*Larix olgensis* Henry.) 组分生物量使用随机森林算法反演建模, 随机森林模型的精度 (R^2) 均高于 0.91。由于机载 LiDAR 点云数据的特殊性, 能从三维结构上反映森林的垂直结构, 因此, 机载 LiDAR 点云数据可提供更多描绘森林的特征来估测森林参数。对于多特征参数的应用, 更好的挖掘机载 LiDAR 点云数据获取森林参数的潜力, 无疑机器学习提供了更加便利的技术手段。

国外已有多位研究者基于机载 LiDAR 点云数据使用不同的机器学习方法对森林参数进行了估测研究, 而国内相应研究较少。鉴于此, 本研究以广西壮族自治区高峰林场试验区为研究区, 采用 2016 年 9 月获取的机载 LiDAR 点云数据为基础, 结合 105 块同步获取的样地数据, 使用随机森林、支持向量机、支持向量机+随机森林的组合模型分别对研究区林分平均高进行建模估测。评价不同方法在使用机载 LiDAR 点云数据估测林分平均高应用中的适应性, 采用模型精度高、泛化能力强的模型完成研究区林分平均高制图。

1 数据与方法

1.1 研究区概况

试验区位于广西壮族自治区南宁市北部的国有高峰林场内, 为一个呈东北-西南走向的近矩形区域, 中心地理位置为 108°23'45" E, 22°58'33" N, 长 11.2 km, 宽 4.2 km, 面积约为 4 770 hm²。研究区位置见图 1。研究区属低山地貌, 海拔 90~460 m, 地形封闭, 一般坡度为 24°~34°, 坡度大于 24°的面积占研究区面积的 75%。属亚热带季风

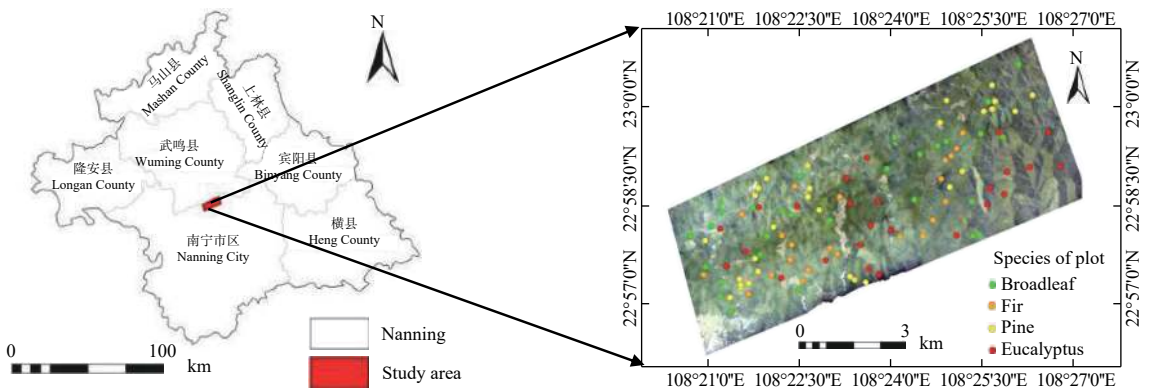


图 1 研究区位置图

Fig. 1 The location map of study area

气候, 年均温度 21.6°左右, 年均降水量 1 300 mm, 相对湿度 79%, 热量充足, 雨水充沛, 林木生长迅速。森林覆盖率为 90% 以上, 人工林约为 95%, 除桉树外, 年龄大多在 15 a 以上, 主要树种有尾叶桉 (*Eucalyptus urohylla* S.T. Blake)、巨尾桉 (*Eucalyptus grandis* × *E. urohylla*)、马尾松 (*Pinus massoniana* Lamb.)、湿地松 (*Pinus elliottii* Engelman)、八角 (*Illicium verum* Hook. f.)、红椎 (*Castanopsis hystrix* Miq.)、火力楠 (*Michelia macclurei* Dandy)、米老排 (*Mytilaria laosensis* Lec.)、厚荚相思 (*Acacia crassicaarpa* Benth.) 等。60% 的林分为人工天然混交林, 在沟谷内有少量天然杂木。

1.2 样地数据获取

研究区划分为 4 个优势树种组, 分别设置 22 块杉木样地、25 块桉树样地 (尾叶桉、巨尾桉)、29 块松树样地 (马尾松、湿地松、杂交松)、29 块一般阔叶树样地 (红椎、火力楠、米老排、八角、厚荚相思和天然杂木林), 共计 105 块样地。在 GIS 软件中, 基于小班空间分布图上根据优势树种 (组) 分布按照相对均匀分布的原则随机布设样地, 确定样地中心点的图面坐标, 根据图面坐标采用手持双频差分 GPS 导航至样地中心, 观察周围林分情况, 若该中心四周 30 m 内林分均属于同一类型, 则该中心即为样地中心, 否则, 移动该中心至合适位置。如果无论怎样移动样地中心, 样地都无法只包含一个林分类型, 可包含两个类型, 但不能包含无林地。最后在样地中心西北方向 20 m 处附近, 随机确定样地西北角点, 用森林罗盘仪和激光测距仪测设样地边界, 样地面积为 30 m × 30 m。样地分布位置图见图 1。在测设的样地内每木检尺, 用胸径尺测量胸高直径, 起测胸高直径为 5 cm。用超声波测高仪 (Haglöf VERTEX IV) 测定 3 株平均木高和 1 株优势木高。样地计算指标包括: 平均直径、平均树高、断面积、每公顷株数、每公顷蓄积量, 其中每公顷蓄积量计算根据断面积和平均高采用形高表计算。林分平均高采用断面积加权法计算, 计算公式为式 (1)。以此方法完成 105 块样地调查。

$$H = \frac{\sum_{i=1}^n h_i g_i}{\sum_{i=1}^n g_i} \quad (1)$$

式 (1) 中: H 为林分平均高, h_i 为第 i 株树木的林木高度, g_i 为第 i 株树木的胸高断面积, n 为林分株数。

1.3 机载 LiDAR 数据获取及处理

2016 年 9 月使用 R44 直升机搭载奥地利 RIEGL 公司的 VUX-1LR 激光雷达系统获取整个研究区数据, 飞行覆盖范围 108°20'9" ~ 108°27'33" E, 23°0'28" ~ 22°56'3" N。飞行总覆盖面积约为 55 Km², 共飞行 3 驾次, 每次约为 5 h。该系统集激光测距、全球定位系统 (GPS) 和惯性导航系统 (IMS) 于一体。激光器工作波长为 1 550 nm, 激光束发散角为 0.5 mrad, 获得的平均点云密度约为 2 点每 m²。

首先对机载 LiDAR 数据进行预处理, 预处理包括去噪、滤波、分类。本研究选用阈值法消除噪点, 使用不规则三角格网滤波算法^[27], 完成整个研究区的地面点和非地面点分类; 由于森林地形的复杂性, 目前很难有一种普适性的地面滤波算法可以科学准确的分类出地面点。因此由此方法得到的地面点需进一步目视解译, 人工修正一些未分类正确的点, 最终准确得到整个研究区的地面点。对于非地面点部分, 研究区位于森林, 所以只将非地面点区分为植被点和其他点, 其他点不做处理。最后利用植被点与基于地面点插值生成的数字地面模型 (DEM) 作差得到整个研究区归一化后的植被点云。

利用归一化的植被点云提取相关点云特征统计量。具体包括: 高度百分位数 13 个 ($H_5, H_{10}, H_{20}, H_{25}, H_{30}, H_{40}, H_{50}, H_{60}, H_{70}, H_{75}, H_{80}, H_{90}, H_{95}$)、密度变量 10 个 ($D_0, D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$)、变异系数 (H_{cv})、最大值 (H_{max})、最小值 (H_{min})、平均值 (H_{mean})、中位数 (H_{median})、标准差 (H_{std})、方差 (H_{var})、高度百分位数四分位数间距 (H_{iq}) 等 8 个高度变量。由于 LiDAR 点云的首次回波对树冠表面探测更为敏感, 因此本次研究使用首次回波提取以上 LiDAR 特征。为了便于地面样地数据匹配, 统计空间林分尺度设置为 30 m × 30 m。

1.4 建模方法

基于 LiDAR 点云数据获取的 31 个特征变量结合实测样地数据采用随机森林、支持向量机回归及两种方法的组合回归建模。其中, 组合回归建模利用随机森林优选特征, 最后利用优选特征进行支持向量机回归建模。首先将 105 块样地数据按照 3:1 的比例随机划分为 79 个训练样本数据, 26 个检

验样本数据,以79个训练样本数据作为建立模型的基础数据。其次利用网格搜索算法与十折交叉验证方法寻找最优参数,完成模型的建立。最后采用26个检验样本数据检验模型,完成模型的精度评价。精度评价主要采用决定系数(R^2),均方根误差(RMSE)对比随机森林、支持向量机回归及两种方法的组合模型在机载LiDAR点云数据估测林分平均高中的适应性。以上回归模型建立均使用python语言实现。

2 结果与讨论

2.1 结果

2.1.1 模型建立 基于79个训练样本数据与机载LiDAR数据提取的 $30\text{ m}\times 30\text{ m}$ 空间林分尺度下的特征变量回归建模。使用网格搜索算法结合十折交叉验证方法分别对随机森林、支持向量机及组合回归模型进行参数寻优。(1)随机森林主要对构建决策树的数量,使树节点分裂的最小特征数,最小叶子节点3个参数进行寻优。最终参数最优解的结果为:构建决策树的数量75,树节点分裂的最小特征数为2,最小叶子节点为2;(2)支持向量机模型参数寻优中选用常用的3个核函数,即线性核、RBF和多项式核函数分别进行。寻优结果为:核函数为RBF,惩罚参数(C)为10, γ 参数为0.02。(3)根据参数寻优建立的随机森林回归模型得到参与建模LiDAR特征变量的重要性,如图2。图中重要性值越高,该特征对林分平均高估测越重要。由图2可以看出, LiDAR特征变量中几个高度分位数的重要性值都较大,其中以 H_{80} 高度分为数最大;其次为点云的几个统计变量(H_{var} 、 H_{mean} 、 H_{std} 等)。说明植被点云的首次回波中75%~95%的高度分位数或最大高度分位数(H_{max})有利于拟合林分平均高或优势木平均高,这与前人研究结果一致^[22, 28-30]。密度变量重要性值较大的有 D_3 、 D_4 、 D_8 等。其中 D_3 被选中与刘浩等^[17]的研究结果一致。研究中选择大于80%信息累计贡献的特征变量使用支持向量机回归建模,随机森林回归模型优选后的特征变量具体见表1。依据网格参数寻优方法,进一步得到组合模型的参数最优解:核函数为RBF,惩罚参数(C)为10, γ 参数为0.01。

2.1.2 模型效果 使用79个训练样本数据和26个检验样本数据分别利用3个回归模型进行林分平均

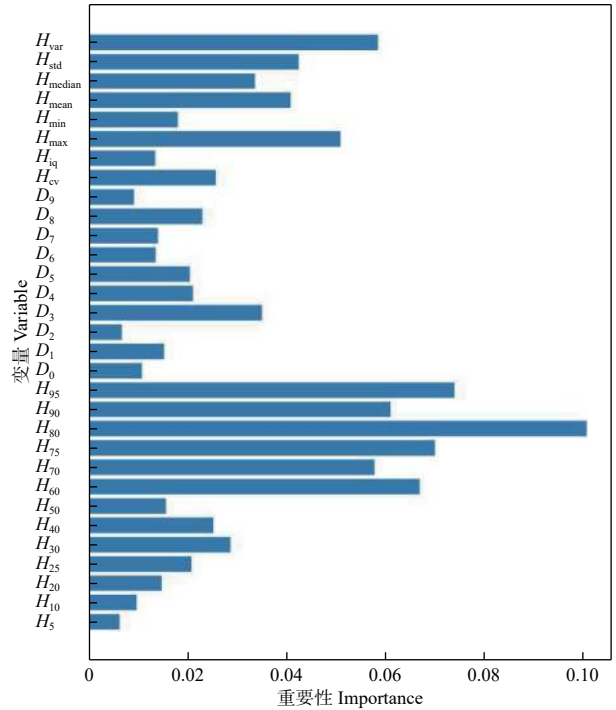


图2 特征变量重要性

Fig. 2 Importance of characteristic variables

表1 随机森林模型特征优选

Table 1 RF model feature selection

序号 Order	变量 Variable	重要性 Importance
1	H_{80}	0.100 73
2	H_{95}	0.073 89
3	H_{75}	0.069 96
4	H_{60}	0.066 89
5	H_{90}	0.061 01
6	H_{var}	0.058 46
7	H_{70}	0.057 75
8	H_{max}	0.050 81
9	H_{std}	0.042 36
10	H_{mean}	0.040 72
11	D_3	0.034 92
12	H_{median}	0.033 50
13	H_{30}	0.028 55
14	H_{cv}	0.025 59
15	H_{40}	0.025 08
16	D_8	0.022 87
17	D_4	0.020 97

高预测,进一步分析3个模型的预测林分平均高和实测林分平均高的相关关系。图3、4、5分别为随

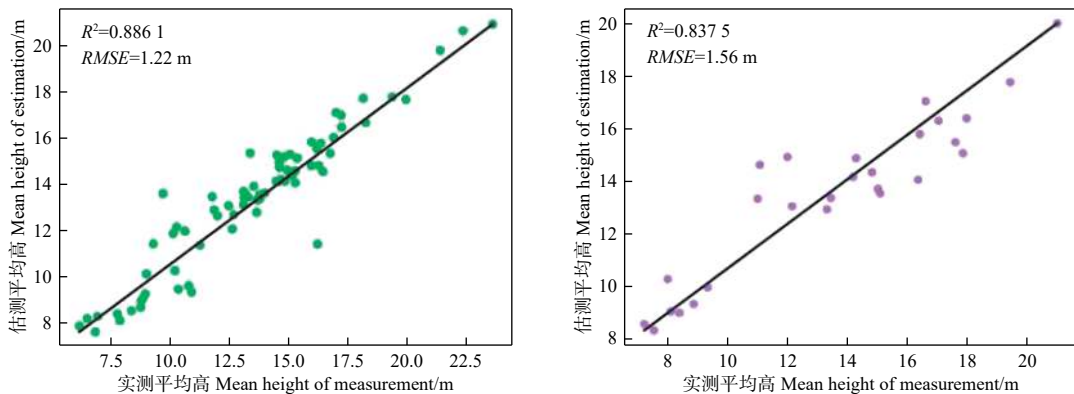


图 3 随机森林回归模型预测林分平均高和实测平均高对比图

Fig. 3 Comparison of stand mean height between and the RFR model estimations the ground measured values

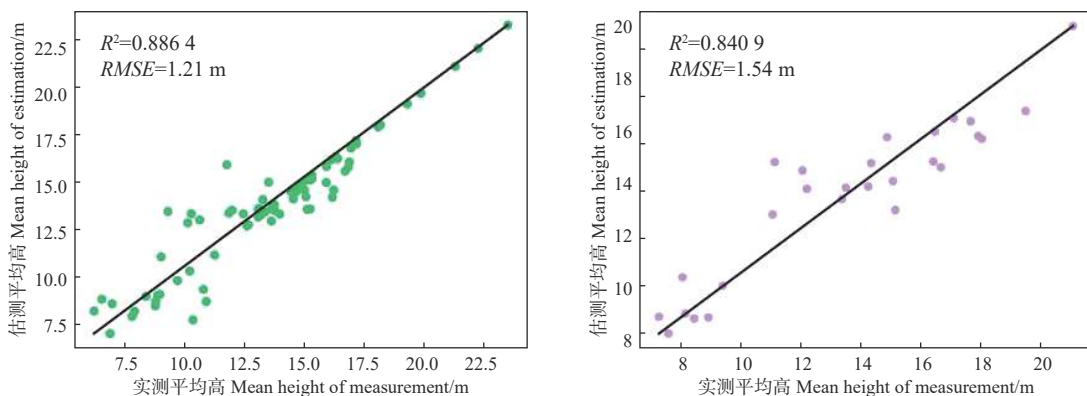


图 4 支持向量机回归模型预测林分平均高和实测平均高对比图

Fig. 4 Comparison of stand mean height between and the SVR model estimations the ground measured values

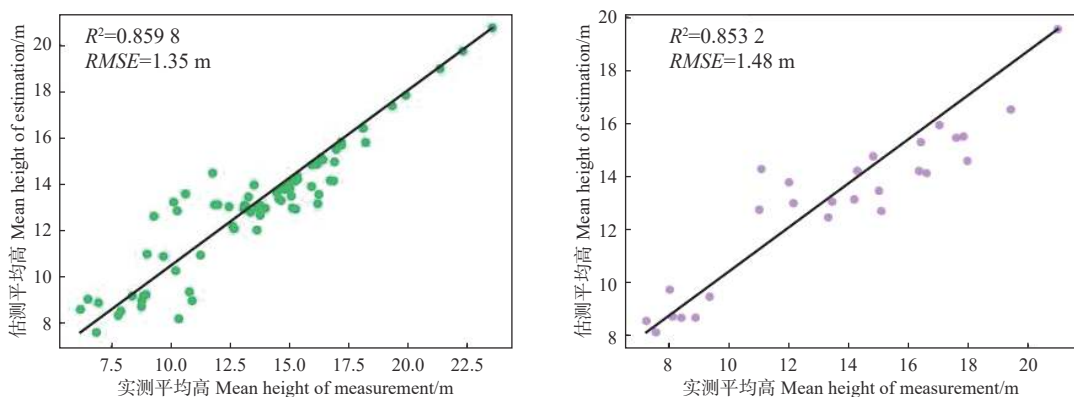


图 5 组合回归模型预测林分平均高和实测平均高对比图

Fig. 5 Comparison of stand mean height between and the RFR&SVR model estimations the ground measured values

机森林模型、支持向量机模型以及两种方法组合模型的预测林分平均高和实测平均高的相关关系。

3 个模型精度评价指标见表 2。表中差值一项反映了 3 个模型中训练数据和检验数据对应的精度指标之差, 差值越小, 模型泛化能力越强。由表 2 可知, 3 个模型的决定系数 (R^2) 均大于 0.85, $RMSE$

表 2 模型精度评价指标

Table 2 Model accuracy evaluation index

回归模型 Regression model	训练样本 Training sample		检验样本 Test sample		差值 Difference	
	R^2	$RMSE/m$	R^2	$RMSE/m$	R^2	$RMSE/m$
	RFR	0.886 1	1.22	0.837 5	1.56	0.048 6
SVR	0.886 4	1.21	0.840 9	1.54	0.045 5	0.33
RFR&SVR	0.859 8	1.35	0.853 2	1.48	0.006 6	0.13

也都较小。从检验样本的精度指标及其与模型精度的差值看,组合模型(RFR&SVR)的差值最小,泛化能力最强,说明利用组合模型更适合推演到区域尺度完成林分平均高制图。最后,采用python

语言读取整个研究区的对应的LiDAR变量特征,利用组合模型完成整个研究区林分平均高制图,图6为研究区林分平均高分布图。

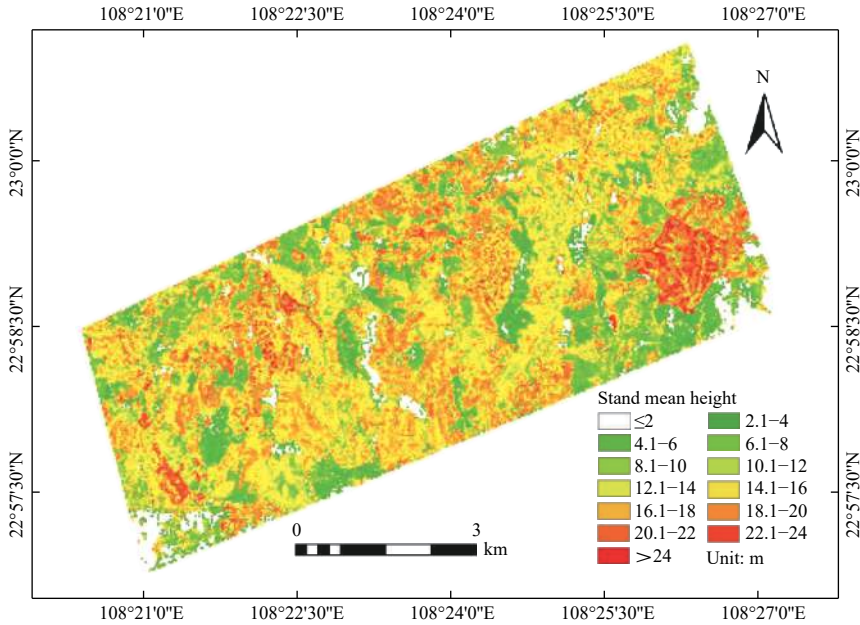


图6 组合模型(RFR&SVR)林分平均高制图

Fig. 6 Stand mean height map based on combined model (RFR&SVR)

2.2 讨论

(1) 随机森林回归模型中,从训练样本的随机抽样到特征选择的随机抽样,两层随机抽样保证了模型的稳定性,在预测林分平均高中具有较强的抗噪声能力,使得随机森林模型在预测林分平均高中表现较好,训练数据和检验数据的预测精度均高于80%。同时,每棵决策树的建立,投票选择出对林分平均高解释性最强的特征变量,取每个特征的均值作为最终在模型中的贡献程度,相比于支持向量机模型而言可以更好的解释特征变量。

(2) 使用支持向量机预测林分平均高,通过核函数由低维到高维映射,解决了高维空间计算复杂的问题。同时,支持向量机寻找最少的支持向量构建支撑面,对噪声异值敏感度降低。回归过程中引入误差损失函数,使得误差容忍度提高,模型精度有所提高。对于小样本数据,支持向量机表现较优,研究结果也证明了这一点。较随机森林而言,支持向量机在利用105块实测样地数据建模预测林分平均高中精度优于随机森林模型。

(3) 组合模型利用了随机森林模型优选特征的可解释性和科学性,同时结合了支持向量机模型

对小样本数据回归预测的优势,更好的完成林分平均树高的预测。研究结果得到组合模型的泛化能力最强,估测精度也与前两种方法相差较小,更适合于模型的推演与制图。

(4) 随机森林和支持向量机这一类机器学习算法,在模型的构建中,模型参数的优化是精确的预测林分平均高的关键。使用格网搜索算法进行参数的选择,是一种高效快捷的方式,它利用交叉验证的方法对每一组参数进行打分,最终选取出最优的参数组合,避免了在预测林分平均高建模中盲目调节参数的目的。

(5) 利用105块实测数据对研究区内的林分平均高建模回归预测,虽然取得了较好的结果,但由于本次研究中受样地数量的限制,没有分树种进行估测,一定程度会影响树高的估测精度。因此,后期需要分树种建模预测林分平均高,进一步改善预测精度。

3 结论

基于LiDAR点云数据提取的31个特征变量,以105块样地数据为基础,使用随机森林、支持向

量机、两种方法的组合模型对广西壮族自治区高峰林场实验区的林分平均高进行预测。3种模型的模型精度和检验精度都较高。其中随机森林模型的决定系数 (R^2) 和均方根误差 ($RMSE$) 为 0.886 1 和 1.22, 独立样本检验结果 R^2 和 $RMSE$ 为 0.837 5 和 1.56; 支持向量机模型的决定系数 (R^2) 和均方根误差 ($RMSE$) 分别为 0.886 4 和 1.21, 独立样本检验结果 R^2 和 $RMSE$ 分别为 0.840 9 和 1.54; 两种方法的组合模型的决定系数 (R^2) 和均方根误差 ($RMSE$) 分别为 0.859 8 和 1.35, 独立样本检验结果 R^2 和 $RMSE$ 分别为 0.853 2 和 1.48。总体上组合模型在基于机载 LiDAR 点云数据估测林分平均高中模型泛化能力最强、制图精度最高, 更适合研究区完成森林林分平均高制图。

由于本次研究区仅限于高峰林场, 研究范围有限, 研究中的 3 种估测方法及其效果差异是否适用广西其它地区还有待进一步试验。基于机载 LiDAR 点云数据开展可移植性的森林参数估测方法是今后深入研究的重要内容, 也是笔者今后努力的方向。

致谢: 本次实验的机载 LiDAR 点云数据及野外调查地面数据由广西省壮族自治区林业勘测设计院提供, 在此表示衷心感谢。

参考文献:

- [1] Næsset E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data[J]. Remote Sensing of Environment, 2002, 80(1): 88-99.
- [2] Magnussen S, Boudewyn P. Derivations of stand heights from airborne laser scanner data with canopy-based quantileestimators[J]. Canadian Journal of Forest Research, 2011, 28(7): 1016-1031.
- [3] Garcia M, Saatchi S, Ferraz A, et al. Impact of data model and point density on aboveground forest biomass estimation from airborne LiDAR[J]. Carbon Balance and Management, 2017, 12(1): 4.
- [4] 穆喜云, 张秋良, 刘清旺, 等. 基于机载LiDAR数据的林分平均高及郁闭度反演[J]. 东北林业大学学报, 2015, 43(9): 84-89.
- [5] Popescu S C, Wynne R H, Nelson R F. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size[J]. Computers and Electronics in Agriculture, 2002, 37(1-3): 71-95.
- [6] Popescu S C. Estimating biomass of individual pine trees using airborne lidar[J]. Biomass and Bioenergy, 2007, 31(9): 646-655.
- [7] Feng L, Chang T, Gui Z, et al. Estimation of forest parameter and biomass for individual pine trees using airborne LiDAR[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(9): 219-224+242.
- [8] 穆喜云, 张秋良, 刘清旺, 等. 基于机载激光雷达的寒温带典型森林高度制图研究[J]. 北京林业大学学报, 2015, 37(7): 58-67.
- [9] Urbazaev M, Thiel C, Cremer F, et al. Estimation of forest above-ground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico[J]. Carbon Balance and Management, 2018, 13(1): 5.
- [10] Bouvier M, Durrieu S, Fournier R A, et al. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data[J]. Remote Sensing of Environment, 2015, 156: 322-334.
- [11] Montealegre A L, Lamelas M T, La Riva J D, et al. Use of low point density ALS data to estimate stand-level structural variables in Mediterranean Aleppo pine forest[J]. Forestry, 2016, 89(4): 373-382.
- [12] 高 婷, 李卫忠, 赵鹏祥, 等. 基于ArboLiDAR的大野口林区森林参数估测[J]. 西北林学院学报, 2017, 32(4): 172-177+223.
- [13] 庞 勇, 赵 峰, 李增元, 等. 机载激光雷达平均树高提取研究[J]. 遥感学报, 2008, (1): 152-158.
- [14] Kwak D A, Cui G, Lee W K, et al. Estimating plot volume using lidar height and intensity distributional parameters[J]. International Journal of Remote Sensing, 2014, 35(13): 4601-4629.
- [15] Silva C A, Carine K, Hudak A T, et al. Modeling and mapping basal area of *Pinustaeda L.* plantation using airborne LiDAR data[J]. Anais da Academia Brasileira de Ciências, 2017, 89(3): 1895-1905.
- [16] 徐 婷, 曹 林, 申 鑫, 等. 基于机载激光雷达与Landsat 8 OLI数据的亚热带森林生物量估算[J]. 植物生态学报, 2015, 39(4): 309-321.
- [17] 刘 浩, 张峥男, 曹 林. 机载激光雷达森林垂直结构剖面参数的沿海平原人工林分特征反演[J]. 遥感学报, 2018, 22(5): 872-888.
- [18] Osborne J W, Waters E. Four assumptions of multiple regression that researchers should always test[J]. Practical Assessment, Research & Evaluation, 2002, 8(2).
- [19] Dalponte M, Bruzzone L, Gianelle D. Estimation of tree biomass volume in alpine forest areas using multireturn Lidar data and support vector regression[C]// Image and Signal Processing for Remote Sensing XIV. International Society for Optics and Photonics, 2008. doi: 10.1117/12.801669
- [20] Gleason C J, Im J. Forest biomass estimation from airborne LiDAR data using machine learning approaches[J]. Remote Sensing of Environment, 2012, 125: 80-91.
- [21] Li M, Im J, Quackenbush L J, et al. Forest biomass and carbon stock quantification using airborne LiDAR data: A case study over hunting-wildlife forest in the Adirondack Park[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(7): 3143-3156.
- [22] 鲁 林, 周小成, 余治忠, 等. 随机森林算法在机载LiDAR数据林分平均树高估算中的应用研究[J]. 地球信息科学学报, 2016, 18(8): 1133-1140.
- [23] Alberto S C, Carine K, Hudak A T, et al. Estimating stand height and

- tree density in *Pinus taeda* plantations using in-situ data, airborne LiDAR and k-Nearest neighbor imputation[J]. Anais da Academia Brasileira de Ciências, 2018, 90(1): 295-309.
- [24] Monnet J M, Chanussot J, Berger, Frédéric. Support vector regression for the estimation of forest stand parameters using airborne laser scanning[J]. IEEE Geoscience & Remote Sensing Letters, 2011, 8(3): 580-584.
- [25] Tompalski P, White J C, Coops N C, et al. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data[J]. Remote Sensing of Environment, 2019, 227: 110-124.
- [26] 洪奕丰, 张守攻, 陈 伟, 等. 基于机载激光雷达的落叶松组分生物量反演[J]. 林业科学研究, 2019, 32(5): 83-90.
- [27] Axelsson P. DEM generation from laser scanner data using adaptive TIN models[J]. International Archives of Photogrammetry and Remote Sensing, 2000, 33(B4/1, PART 4): 111-118.
- [28] Naesset E, Bjerknæs K. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data[J]. Remote Sensing of Environment, 2001, 78(3): 328-340.
- [29] Tesfamichael S G, Ahmed F B, Van Aardt J A N. Investigating the impact of discrete-return lidar point density on estimations of mean and dominant plot-level tree height in *Eucalyptus grandis* plantations[J]. International Journal of Remote Sensing, 2010, 31(11): 2925-2940.
- [30] 庞 勇, 李增元, 谭炳香, 等. 点云密度对机载激光雷达林分高度反演的影响[J]. 林业科学研究, 2008, 21(Z1): 14-19.

Estimation of Forest Stand Mean Height Based on Airborne LiDAR Point Cloud Data

ZHAO Xun¹, YUE Cai-rong¹, LI Chun-gan², ZHANG Li-mei³, GU Lei¹

(1. School of Forestry, Southwest Forestry University, Kunming 650224, Yunnan, China; 2. Forestry College, Guangxi University, Nanning 530004, Guangxi, China; 3. School of Landscape Architecture and Horticulture Sciences, Southwest Forestry University, Kunming 650224, Yunnan, China)

Abstract: [Objective] Based on the airborne LiDAR point cloud data obtained in the experimental area of the Gaofeng Forest Farm in Guangxi Zhuang Autonomous Region in September 2016 to estimate the stand mean height by extracting the LiDAR point cloud characteristic variables at the spatial stand scale of 30 m×30 m. [Method] The data of mean height obtained from 105 stands were randomly divided into training samples (79) and test samples (26) according to a ratio of 3:1. Two machine learning algorithms, namely random forest regression and support vector regression, were adopted to conduct regression modeling of 79 training samples and corresponding LiDAR point cloud feature variables. The modeling schemes include random forest model, support vector machine model and random forest + support vector combination model. Secondly, the 26 test sample data were used to evaluate the prediction accuracy of the model. Finally, the accuracy evaluation indexes corresponding to the training samples and the test samples in the three models were counted, and a model with high prediction accuracy and strong generalization ability was taken as the final model to map the stand mean height of forest. [Result] The result showed that the determination coefficient (R^2) of the training samples and the test samples of the random forest model were 0.8861 and 0.8375, respectively, and the root mean square errors (RMSE) were 1.22 and 1.56, respectively. The R^2 of the training samples and the test samples of support vector machine model were 0.8864 and 0.8409, respectively, and the RMSE were 1.21 and 1.54, respectively. The R^2 of the training samples and the test samples of the combined model were 0.8598 and 0.8532 respectively, and the RMSE were 1.35 and 1.48 respectively. [Conclusion] The combined model has the best generalization ability and prediction accuracy, followed by support vector machine and the random forest. The combined model can effectively complete the mapping of stand mean height in the study area.

Keywords: Airborne LiDAR Point Cloud Data; random forest regression; support vector regression; stand mean height; Gaofeng Forest Farm