

doi: 10.6046/zrzyyg.2021237
引用格式: 邓静雯,田义超,张强,等. 机载 LiDAR 在红树林林分平均高估算中的应用[J]. 自然资源遥感,2022,34(3):129 – 137. (Deng J W,Tian Y C,Zhang Q,et al. Application of airborne LiDAR in the estimation of the mean height of mangrove stand[J]. Remote Sensing for Natural Resources,2022,34(3):129 – 137.)

机载 LiDAR 在红树林林分平均高估算中的应用

邓静雯¹, 田义超^{1,2}, 张 强¹, 陶 进^{1,3}, 张亚丽¹, 黄升光³

(1. 北部湾大学资源与环境学院,钦州 535000; 2. 北部湾大学广西北部湾海洋灾害研究重点实验室,钦州 535000; 3. 北部湾大学海洋地理信息资源开发利用重点实验室,钦州 535000)

摘要: 构建基于激光雷达(light detection and ranging,LiDAR)数据的林分平均高反演模型可为无瓣海桑长势的动态监测提供技术支撑。以北部湾茅尾海无瓣海桑红树林湿地为对象,基于机载 LiDAR 提取的高度和强度参数变量,借助决定系数 R^2 、均方根误差 $RMSE$ 、赤池信息准则 AIC 和贝叶斯信息准则 BIC 指标对随机森林、支持向量机以及神经网络 3 种模型进行了优选,在最优模型的支持下估算了研究区的红树林平均高及其空间分布状况。结果表明,研究区无瓣海桑的林分平均高介于 3.90 ~ 11.58 m 之间,其中树高较高、胸径较大的无瓣海桑主要分布在潮沟附近以及研究区中部。在估算无瓣海桑的林分平均高时,贡献率最大的是样方点云高度最大值,其次是 75% ~ 99% 分位数高度。随机森林回归模型在估测林分平均高模型中的精度最高($R^2 = 0.938\ 1$, $RMSE = 0.58\ m$, $AIC = 80.50$ 和 $BIC = 49.05$); 支持向量机模型次之,该模型在测试阶段的 R^2 为 0.766 5, $RMSE$ 为 1.27 m; 神经网络模型的拟合效果最差。总体而言,随机森林模型是研究区无瓣海桑林分平均高反演的最优模型。

关键词: 无瓣海桑; 支持向量机; 神经网络; 随机森林; 林分平均高

中图法分类号: S 771.8; TP 79 **文献标志码:** A **文章编号:** 2097-034X(2022)03-0129-09

0 引言

红树林群落与其所在的滨海盐生沼泽湿地相互作用形成了红树林生态系统。红树林生态系统是全球最具特色的湿地生态系统,同时具备海洋生态及陆地生态的特性,具有强大的生态防护功能。林分平均高是红树林湿地生态系统模型的一个重要输入参数,是研究红树林碳储量的重要指标之一,也是估算各种红树林参数的基础。因此,采用定量化方法快速、准确地获取红树林林分平均高信息成为了近年来红树林结构参数研究的热点。

人工实地调查在林分平均高调查时费时费力,且调查效率低下。随着遥感技术的日益成熟,遥感技术已成功应用于大尺度森林制图和灾害监测等领域,效率得到了大幅度的提高,但森林高度、生物量等森林结构参数是传统光学遥感无法解决的,而激

光雷达(light detection and ranging, LiDAR)技术可以获取森林的结构参数信息^[1]。LiDAR 是传感器发出激光信号,检测目标物体的位置,接收反射信号从而测量传感器与目标物体之间距离的一种新技术。近几十年来,国内外相关学者在激光点云森林结构参数反演算法方面做了大量的研究,促进了 LiDAR 技术在林业方面的发展。其中,Næsset 等^[2]、Andersen 等^[3]、焦义涛等^[4]、洪奕丰等^[5]学者基于机载 LiDAR 点云数据,采用线性回归模型对不同区域的森林高度和生物量进行了估算。线性回归模型比较适用于建模目标之间关系相对简单且数据量不大的情况,这类线性回归模型虽然快速简单,但要求建模数据必须满足独立性、线性、正态性以及方差齐性等要求。

相对而言,随机森林、支持向量机以及人工神经网络等算法在构建复杂非线性森林参数估算模型中具有更大优势。如 Li 等^[6]和 Monnet 等^[7]基于机载

收稿日期: 2021-07-28; 修订日期: 2021-10-19
基金项目: 国家自然科学基金项目“桂西南峰丛洼地流域生态系统服务权衡关系及其驱动机制”(编号: 42061020)、广西创新驱动发展专项项目“海岸双频激光 LiDAR 探测仪产品开发与应用示范”(编号: AA18118038)、广西科技基地和人才项目“广西钦州湾红树林岛群生态系统健康评价及应用示范”(编号: 2019AC20088)和广西壮族自治区大学生创新创业训练项目“基于激光雷达数据的红树林地上生物量与碳储量反演”(编号: 202011607186)共同资助。
第一作者: 邓静雯(1999-),女,学士,主要从事遥感信息应用研究。Email: 1687606800@qq.com。
通信作者: 田义超(1986-),男,博士,副教授,主要从事资源环境遥感及海岸带生态环境监测研究。Email: tianyichao1314@yeah.net。

LiDAR 点云数据,在支持向量机模型的支持下对不同研究区的生物量进行估计,结果显示估测精度较好;鲁林等^[8]和赵勋等^[9]基于 LiDAR 点云数据,采用随机森林算法分别对福建省朱溪河流域的松树以及广西壮族自治区国有高峰林场的人工林林分平均高进行了估算,结果显示林分平均高与实测值具有显著的相关性,其相关系数均超过0.94。但是红树林生长在海岸滩涂地带,人们很难进入其根系繁茂的内部进行测量。虽然国外已有诸多学者对 LiDAR 反演红树林植被参数的算法开展了深入的研究,但国内学者在这方面的研究较少,仅有少量开展过红树林生物量方面的研究^[10-11]。迄今为止,采用机载 LiDAR 点云对红树林林分平均高方面的研究匮乏。因此,为了支持红树林的恢复、保护和可持续管理,需要采用一种快速高效的模型来评估红树林的林分平均高。而基于机载 LiDAR 点云数据,采用机器学习算法为红树林林分平均高的估算提供了便捷的技术手段。

无瓣海桑(*Sonneratia apetala*)天然分布于马来西亚、孟加拉国等国^[12],2002 年广西钦州市为了加快红树林恢复,从广东湛江市引进了无瓣海桑幼苗进行造林试验。在近年营造的作用下,无瓣海桑迅速占领了适宜的林地,甚至入侵了原生红树林群落^[13]。这些入侵的无瓣海桑挤占了原有的本土红树林生长空间,对当地的红树林树种生长造成了一定的影响,但是入侵的无瓣海桑高度如何,目前尚未有量化的评估数据。鉴于此,基于北部湾茅尾海康熙岭片区的 LiDAR 点云数据,借助相关数理分析指标对随机森林、支持向量机以及神经网络 3 种模型进行了优选,在最优模型的支持下估算了研究区的红树林平均高及其空间分布状况,研究方法可为今后该地区无瓣海桑长势的动态监测提供有力的技术支撑。

1 研究区概况

研究区位于北部湾茅尾海康熙岭片区,介于 N21°51′26″~21°51′42″,E108°29′27″~108°29′43″之间,该片区属于亚热带季风气候,太阳辐射强,季风环流明显。夏长冬暖,夏季时长有 6 个月之多,全年平均气温在 22℃左右,其年均降雨量为 1 658 mm,年总日照为 1 673 h^[14]。钦州市的海水温度、海水盐度、土壤及沉积物,为红树林的生长提供了有力的条件,本次主要研究树种为一种

外来木本植物——无瓣海桑(*Sonneratia apetala*),据当地渔民的调查可知,研究区的无瓣海桑扩散方向如图 1 黄色箭头所示。



图 1 研究区红树林地理位置
Fig. 1 Location of mangrove in the research area

2 数据来源与研究方法

2.1 数据来源

2.1.1 地面实测样地数据获取

2021 年 1 月在康熙岭红树林保护区共布设了 66 个 10 m×10 m 的正方形样方,样方中心点分布位置如图 1 所示,图中采用地理坐标系 GCS_WGS_1984,投影坐标系 Albers。在布设样方前确定红树林种类、种植模式和当地潮位信息。记录的信息包括样方中的胸径(在高度 1.3 m 处量测)、树高以及地理坐标(经纬度)等数量特征。

本次采样,量测 1.3 m 处树高使用钢卷尺量测,量测胸径使用皮尺量测,如图 2 所示。以样方为统计单位,结合实地量测到的树高和胸径数据,采用断面积加权法算出无瓣海桑的林分平均树高^[15],其具体计算公式为:

$$H = \frac{\sum_{i=1}^n h_i g_i}{\sum_{i=1}^n g_i}, \tag{1}$$

式中: H 为林分平均高; g_i 为第 i 棵林木的胸高断面积; h_i 为第 i 棵林木的树高; n 为样方内所有林木的棵数。



图 2 研究区野外实测样地数据获取

Fig. 2 Field measured sample site data acquisition in the study area

2.1.2 机载 LiDAR 数据获取

2021 年 1 月 19 日下午 3 点采用 DJI M600 PRO 无人机所搭载的普通数码相机和 HS40P 激光传感器在广西茅尾海康熙岭片区进行红树林飞行试验,从而获得了 LiDAR 点云数据。当天拍摄的垂直高度约为 70 m,航向及旁向重叠度为 80% 且风力小于

2 级。本文利用 Inertial Explorer(GPS – IMU)后处理软件结合惯性测量单元(inertial measurement unit,IMU)数据以及全球导航卫星系统(global navigation satellite system,GNSS)技术对原始的红树林回波信息进行解算,可以得到研究区的三维激光点云信息(图 3)。

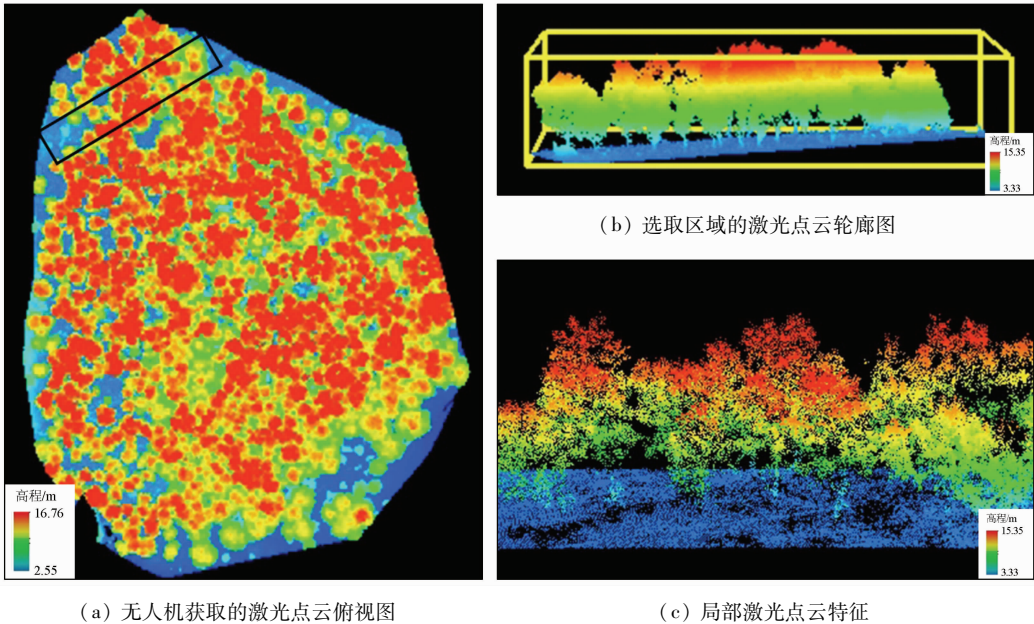


图 3 样方机载 LiDAR 数据

Fig. 3 Airborne LiDAR data at the sample site

2.2 研究方法

2.2.1 LiDAR 点云树高信息提取

机载 LiDAR 系统由激光测距装置、成像装置、IMU 姿态测量装置和全球定位系统 (global positioning system, GPS) 接收机 4 部分组成。它利用 GPS 和 IMU 分别可以获得激光源的坐标及激光脉冲的方位。激光测距的原理是测量地表采样点的激光回波脉冲对于发射主波的时间差,从而得到传感器到地面采样点的距离^[16]。其测距基本原理可表示为:

$$D = \frac{cT}{2}, \tag{2}$$

式中: D 为传感器到目标物体的距离; c 为光速; T 为激光脉冲从激光器到被测物体的来回传输时间。

使用邵为真等^[17]提出的不规则三角网滤波算法将不同密度的 LiDAR 点云数据分离出地面点云与非地面点云(植被点云),对所获取到的地面点进行插值可生成数字高程模型(digital elevation model, DEM),进而将植被点云高程减去 DEM 数据可得到植被点云的归一化值,最后利用归一化植被点云可提取相关点云变量。由于 LiDAR 点云的第一回波对植被冠层的检测更为敏感,本研究采用植被冠层的第一回波数据用于估算林分参数^[9](表 1),包括百分位数高度 15 个($hp01, hp05, \dots, hp95, hp99$)、百分位数强度 15 个($ip01, ip05, \dots, ip95, ip99$)、最

大高度\强度($h_{\max} \setminus i_{\max}$)、最小高度\强度($h_{\min} \setminus i_{\min}$)、平均高度\强度($h_{\text{mean}} \setminus i_{\text{mean}}$)、点云高度\强度标准差($h_{\text{stddev}} \setminus i_{\text{stddev}}$)、高度\强度变异系数($h_{\text{cv}} \setminus i_{\text{cv}}$)、高度\强度众数($h_{\text{mode}} \setminus i_{\text{mode}}$)以及百分位数高度\强度四分位数间距($h_{\text{iq}} \setminus i_{\text{iq}}$)。标准差一般反映用于样方内植被点云高度\强度的离散程度,标准差越大,样方内植被点云高程\强度离散程度越大,反之越小。

2.2.2 林分平均高模型的构建

随机森林算法是分类和回归问题中最受欢迎的机器学习算法之一^[18]。随机森林模型由多元回归树组成,在构建每棵回归树的过程中,每棵树节点选择分割变量时,先从所有特征变量中随机选取 k 个特征变量,然后从中找到最优分割点并生成多个回归树模型。每棵回归树的最终预测结果是节点的平均值,随机森林的最终预测结果是所有回归树预测结果的平均值。随机森林算法使用 boot - strap 采样和 bagging 集成树算法,并从弱学习者集合中获得最准确的决策。在学习过程中,大约 70% 的样本(袋内)用于训练阶段,30% 的样本(袋外)用于测试阶段。与其他基于树和增强的机器学习型类似,随机森林模型可以模拟每个变量的重要性程度。随机森林模型中最主要的超参数包括最大深度、最小样本叶数、最小样本分裂数、最大特征数和可调树数。 $ntree$ 表示随机森林所包含的决策树数目, $mtry$ 表示每次迭代的变量抽样数值,为了保证预测结果的准确性和计算结果的精度,本文采用 $ntree = 1\ 000$, $mtry$ 的取值范围为 $1 \sim M$,其中 M 为模型构建所涉及的输入变量的总数。

支持向量机作为著名的机器学习方法,常应用于各种非线性问题的求解^[19]。它的学习原理是使用一个称为超平面的边界将数据划分为具有近似值的组。通过分析输入变量与输出变量之间的定量关系,随后建立模型对新观测到的变量进行预测^[20]。对于一个分类或回归问题,支持向量机算法用支持向量定义超平面。与其他机器学习模型相比,支持向量机模型包含较少的核函数超参数(线性、多项式或径向基)和正则化参数 C ,支持向量机模型能够处理非线性数据;但是,当组成数据较为嘈杂或重叠时,支持向量机模型效率不高。支持向量机模型的性能受核函数选择的影响。为了减少偏差,本文在建模时尝试了 5 个可能的 C ,分别为: 0.01, 0.1, 1, 10 和 100。

同时,还应用人工神经网络的方法建立树高模型。它的原理是先初始化神经网络,给各个权值赋予一个范围在 $[-1, 1]$ 之间的随机数,并将样本数据通过神经网络进行正向传播,同时用均方差计算输

表 1 用于林分平均高估测的点云特征统计量

Tab. 1 Point cloud characteristic statistics for stand mean height estimation

变量	变量意义 ^①
$hp01 \setminus ip01, \dots, hp99 \setminus ip99$	所有点云 1%, ..., 99% 分位数处对应的高度\强度值
$h_{\max} \setminus i_{\max}$	样方点云高度\强度的最大值
$h_{\min} \setminus i_{\min}$	样方点云高度\强度的最小值
$h_{\text{mean}} \setminus i_{\text{mean}}$	样方点云高度\强度的平均值 $mean$, 计算公式为: $mean = \frac{\sum_{j=1}^m x_j}{m}$
$h_{\text{stddev}} \setminus i_{\text{stddev}}$	样方点云高度\强度的标准差 $stddev$, 计算公式为: $stddev = \sqrt{\frac{1}{m} \sum_{j=1}^m (x_j - \bar{x})^2}$
$h_{\text{cv}} \setminus i_{\text{cv}}$	样方点云高度\强度的变异系数 cv , 计算公式为: $cv = \frac{stddev}{mean}$
$i_{\text{mode}} \setminus i_{\text{mode}}$	样方点云高度\强度的众数
$h_{\text{iq}} \setminus i_{\text{iq}}$	样方点云百分位数高度\强度的四分位数间距 iq , 计算公式为: $iq = p75 - p25$

① m 为每个样方对应归一化植被点云的个数; j 为每个样方对应第 j 个归一化植被点云; x_j 为第 j 植被点云对应的高度值; $\bar{x}$ 为植被点云高度均值; $p75$ 为样方内点云高度的上四分位数; $p25$ 为样方内点云高度的下四分位数。

出误差。通过逆向传播来不断调节输入层、隐含层和输出层间的权重。当神经网络误差的平方和达到最小值时就结束该算法的计算。神经网络算法不需要对数据的基本关系进行假设,但当计算量大时,训练慢,容易过拟合。由于是“黑箱”,它主要用于预测。*decay* 表示权重衰减,为了平衡计算开销和模型性能,本文在建模时尝试了 5 个可能的 *decay*,分别为: 0,0.0001,0.001,0.01 和 0.1^[20]。BP 神经网络原理如图 4 所示。

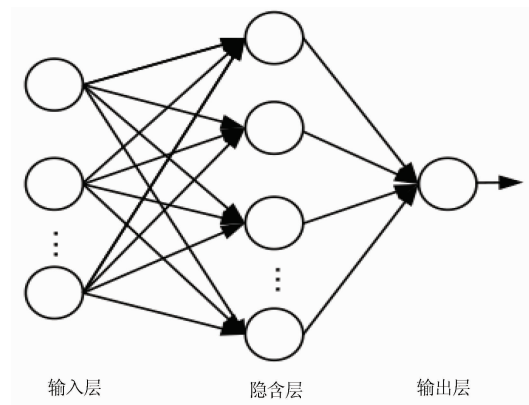


图 4 BP 神经网络原理图

Fig. 4 schematic diagram of BP neural network

2.2.3 评估标准

对于林分平均高模型的选择,利用 4 类指标来评估模型的性能,决定系数 R^2 和均方根误差 (root mean square error, RMSE) 是衡量回归模型统计误差的标准。较高的 R^2 和较低的 RMSE 表明机器学习模型的性能更好。此外,本研究还使用了赤池信息准则 (Akaike information criterion, AIC)^[21] 和贝叶斯信息准则 (Bayesian information criterion, BIC)^[22] 来检验所选模型之间的统计差异。AIC 和 BIC 越低,模型的性能越好^[23]。计算公式分别为:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}, \tag{4}$$

$$AIC = L \ln \left(\frac{RSS}{L} \right) + 2K, \tag{5}$$

$$BIC = L \ln \left(\frac{RSS}{L} \right) + K \ln(L), \tag{6}$$

式中: y_i 为真实值; $\hat{y}_i$ 为预测值; $\bar{y}$ 为真实值的平均数; N 为验证样品的总数; RSS 为残差平方和; K 为参数数量 (包括截距); L 为观测数目。

3 结果与分析

3.1 实测数据特征统计

表 2 列出了 66 个无瓣海桑红树林样方的调查结果。由表 2 可知,林分平均高的变化范围为 3.47 ~ 12.02 m,平均值为 8.19 m。红树林树高为 1.55 ~ 13.58 m,胸径为 0.70 ~ 41.00 cm。红树林实测树高的平均值相比林分平均高的平均值低了 0.1 m (表 2)。

表 2 外业测量树高、胸径和林分平均高的基本统计量
Tab. 2 Basic statistics for field measurements of tree height, diameter at breast height, and stand mean height

参数	最大值	最小值	平均值	标准差
树高/m	13.58	1.55	8.09	2.69
胸径/cm	41.00	0.70	15.38	5.55
林分平均高/m	12.02	3.47	8.19	2.28

3.2 特征选择

本文对参与红树林树高建模的 LiDAR 特征变量的重要性进行排序,选择节点不纯度 (increase in node purity, IncNodePurity) 作为判断 LiDAR 特征变量参数重要性的指标。通过比较点云特征统计变量的重要性,特征统计变量的值越大,则特征统计变量的重要性越高 (图 5)。

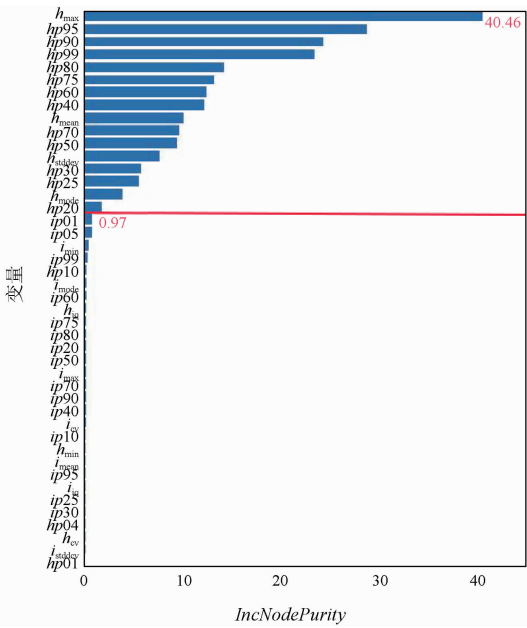


图 5 变量重要性图

Fig. 5 Variable importance chart

图 5 中重要性值越高,该特征统计变量对林分平均高估测越重要。从图 5 中可以看出,利用随机森林算法选择的 44 个特征变量中,高度统计变量参数 $h_{\max}$ 对无瓣海桑的林分平均高反演贡献率最大,其次是 LiDAR 特征变量中的百分位数高度 ($hp95, hp90, hp99$ 等),LiDAR 点云数据中提取到的强度参

数对林分平均高反演贡献较小(红色线为分割线,分割线以上的变量对红树林林分平均高的贡献率较大,分割线以下贡献率较小)。75%~99%分位数高度在研究区无瓣海桑的林分平均高估算中发挥了主要作用,这一发现与前人的研究论断契合^[8-9],他们的研究认为:植被冠层的第一回波的75%~95%分

位数高度或最大树高 $h_{\max}$ 能很好地估测平均树高。本文使用16个贡献率相对较大的特征变量参与随机森林模型的构建,随机森林模型取得了令人满意的结果,训练集和测试集的 R^2 分别为0.985 7和0.938 1, $RMSE$ 为0.58 m(表3)。

表 3 3 种 ML 模型在林分平均高检索中的性能比较
Tab. 3 Comparison of the performance of three ML models in stand mean high retrieval

模型	训练集				测试集			
	R^2	$RMSE/m$	AIC	BIC	R^2	$RMSE/m$	AIC	BIC
随机森林	0.985 7	0.28	36.59	21.42	0.938 1	0.58	80.50	49.05
支持向量机	0.985 6	0.30	37.18	22.00	0.766 5	1.27	87.85	56.40
神经网络	0.978 5	0.34	47.01	31.84	0.436 4	2.90	107.61	76.16

3.3 模型精度对比

将16个贡献率较大的特征变量带入到支持向量机、随机森林和神经网络算法模型中,本研究将70%的样本用于模型的训练,30%的样本用于模型的测试和检验,图6为不同机器学习算法在训练阶段和测试阶段所计算的精度。无瓣海桑林分平均高模型估算的 R^2 越高, $RMSE$ 越低, AIC 和 BIC 越低,预测性能越好。由图6可知,支持向量机算法

在无瓣海桑林分平均高估算中显示出可接受的结果,测试集中 R^2 为0.766 5, $RMSE$ 为1.27 m, AIC 为87.85, BIC 为56.40;而神经网络模型的结果相对较差,在测试集中 R^2 值仅为0.436 4。随机森林模型的估算结果在训练集和测试集阶段均表现出较好的结果, R^2 均达到了0.93以上,总体而言,随机森林模型在研究区域的红树林林分平均高估算中表现最佳。

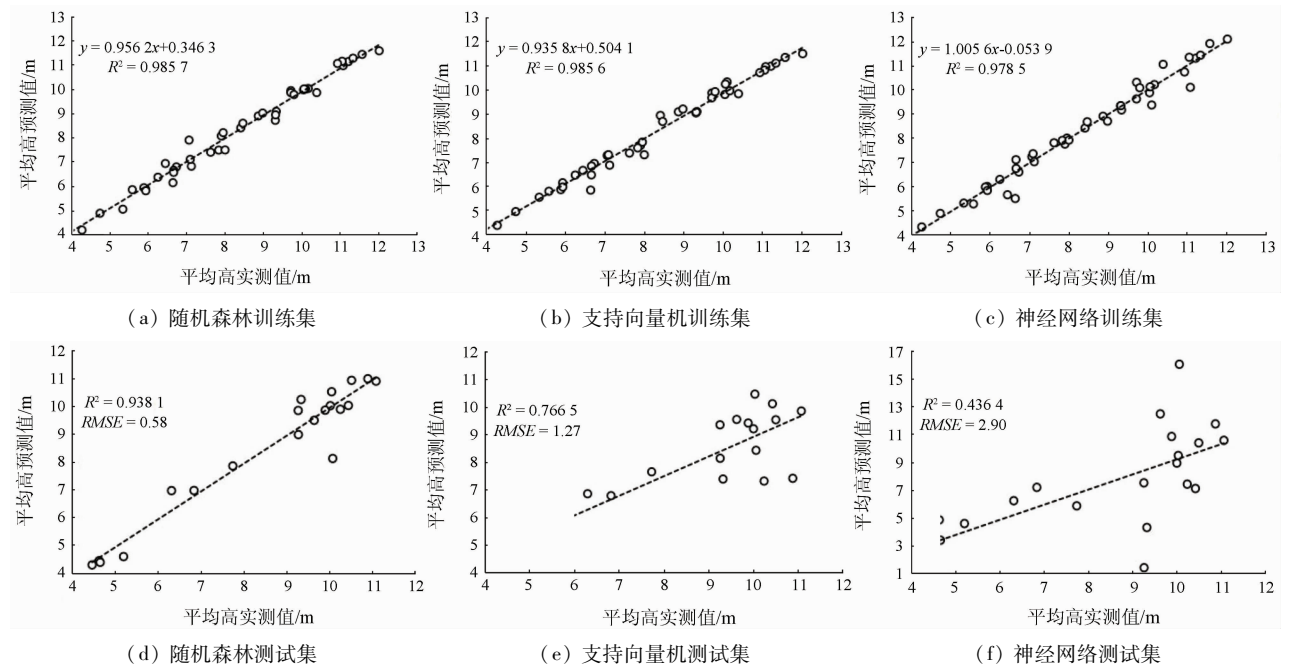


图 6 不同回归模型的训练和测试结果

Fig. 6 Training and testing results of different regression models

3.4 林分平均高的空间分布

由于随机森林模型产生了最佳的预测性能,并且优于其余的机器学习算法,因此本研究选择了随机森林算法模型对研究区无瓣海桑林分平均高进行预测。无瓣海桑林分平均高最终结果被计算成栅格数据,以Tiff格式进行存储。由地理信息系统(geographic information system, GIS)的分类工具将林分

平均高数据划分为5类,由图7可知,无瓣海桑林分平均高的模拟结果介于3.90~11.58 m之间。其中树高超过10 m的无瓣海桑主要分布在研究区潮沟附近及研究区中部,这可能与潮沟水文过程有关,潮沟所带有的营养物质可能给红树林的生长提供了充足的养分。

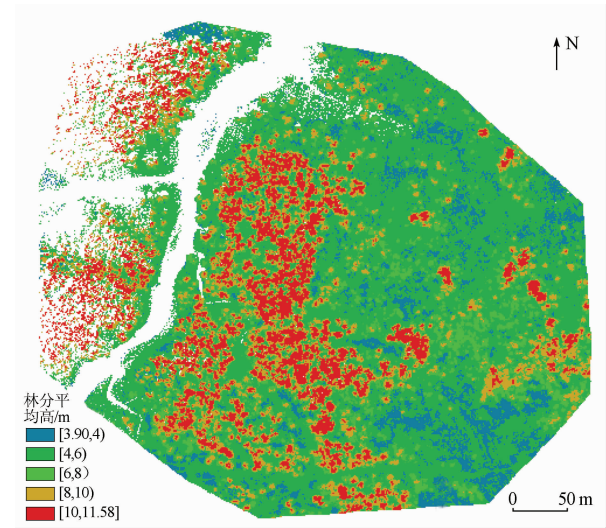


图 7 林分平均高分布

Fig. 7 Mean stand height distribution

4 讨论

本文实测得到北部湾茅尾海(康熙岭片区)的无瓣海桑树高平均值为 8.09 m,比张晓君等^[24]在广东省珠海市鹤洲北湿地地区统计到的无瓣海桑树高平均值高了 0.59 m,也比黄晓敏等^[25]在福建省厦门市集美海湾调查得到的无瓣海桑平均树高值高了 0.39 m。但是,吴瑞等^[26]在海南省东寨港的三江地段得到的无瓣海桑平均树高为 14.2 m,高于本研究区的无瓣海桑树高。无瓣海桑的树高不仅与它的生长年龄相关,同时与它生长环境的海水温度、海水盐度、土壤及沉积物、潮汐浸淹频率和海浪能量等因素更是息息相关^[27]。低温、生长环境的盐分含量过高或过低均会限制无瓣海桑的生长,并且,不同的红树林植物对潮汐浸淹频率都有相应的适应性,无瓣海桑主要分布在大潮中潮位;海浪能量太大时,会阻碍无瓣海桑的根系生长^[28]。除此之外,无瓣海桑平均高还与模型的精度有着直接的关系,模型的精度不仅受到模型本身因素的影响,而且受到其他多种外部因素的影响^[29]。本研究在进行红树林样方调查时并不是直接测量无瓣海桑的林分平均高,而是根据森林的断面积加权法计算林分平均高,断面积加权法的准确度可能会影响模型的估算效果。

随着机器学习算法的发展,林分平均高的估算有了更多的算法。在林分平均高估算算法中,机器学习算法通常比传统的参数方法获得更好的预测性能。Brice 等^[30]发现在林分平均高估算中,随机森林算法优于线性回归算法。Pourrahmati 等^[31]研究发现:在洛里高度(代表了非均匀年龄林分的平均高度)估测中随机森林优于人工神经网络。每种机

器学习算法都有自己的适用性和优势,但机器学习方法的选择、参数的确定均对预测模型的精度有一定的影响^[32]。因此,本研究采用 R 语言的“Random Forest”软件包提供的 importance 进行重要性排序。根据评价指标来获取各模型的最佳参数,利用最佳的参数进行模型模拟。由 3 种机器学习算法的模拟结果可知,随机森林模型表现出很高的模型拟合能力,其 R^2 在测试集和训练集上均达到了 0.938 1 以上, $RMSE$ 均为 0.58 m 以下,支持向量机模型次之,该模型在测试阶段的 R^2 为 0.766 5, $RMSE$ 为 1.27 m,神经网络回归模型的拟合效果最差。这些研究结果表明,在对不同的树种选择不同的特征参数时,机器学习算法均表现出一定的差异性。随机森林模型具有变量选择功能,对无信息的预测变量之间的共线性不敏感。然而,支持向量机和神经网络模型往往受到无信息和高度相关预测变量的影响,降低了模型的性能^[18]。总之,在本研究区,随机森林算法明显优于其他 2 种算法,该算法可在北部湾其他地区的红树林林分平均高估算中发挥重要的作用,但是模型最优参数的选择具有一定的不确定性,应该根据实际需要进行参数和变量的优选^[11]。

由于受样方环境的影响,样方的选择受到了限制,无法遍及更大的区域。即使模型得到了较为满意的反演结果,但是推广到更大的区域或许精度会受到一定的影响,因此对样方的采集可以选择在退潮时量测更多的红树林数据。未来的研究可能利用机载 LiDAR,在更大尺度上对不同沿海地区的无瓣海桑树种进行预测,从而提升随机森林算法在预测红树林林分平均高上的实用性,并寻求一种能获取潮位相对较高区域的实测数据的方法,使得研究范围更大。由于机载 LiDAR 数据采集成本高,试验仅在典型的小区域进行,未来可以选择尺度上推算法,将 LiDAR 点云测量的样方数据推演到哨兵数据尺度或者 Landsat 卫星尺度上。本研究中采用随机森林算法对无瓣海桑的林分平均高进行估算是否适用于其他地区的无瓣海桑或其他红树林树种,尚需进一步检验。

5 结论

本研究以北部湾茅尾海红树林湿地为研究对象,基于机载 LiDAR 数据采用 3 种机器学习算法对研究区无瓣海桑的林分平均高进行建模,并反演了研究区的红树林树高空间分布。其主要研究结论如下:

1) 研究区林分树高范围为 3.47 ~ 12.02 m,平

均值为 8.19 m。

2)样方点云高度最大值 $h_{\max}$ 对无瓣海桑的林分平均高的贡献率最大,其次是 75% ~ 99% 分位数高度。

3)随机森林回归模型的预测性能最好 ($R^2 = 0.938\ 1, RMSE = 0.58, AIC = 80.50, BIC = 49.05$),其次是支持向量机回归模型 ($R^2 = 0.766\ 5, RMSE = 1.27, AIC = 87.85, BIC = 56.40$),BP 神经网络的模型预测性能最差 ($R^2 = 0.436\ 4, RMSE = 2.90, AIC = 107.61, BIC = 76.16$)。

4)随机森林反演结果可知,树高较高、胸径较大的单木红树林主要分布在研究区潮沟附近以及研究区中部。

参考文献(References):

[1] 庞勇,李增元,陈尔学,等.激光雷达技术及其在林业上的应用[J].林业科学,2005(3):129-136.
Pang Y, Li Z Y, Chen E X, et al. LiDAR remote sensing technology and its application in forestry[J]. Scientia Silvae Sinicae, 2005(3):129-136.

[2] Næsset E, Økland T. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve[J]. Remote Sensing of Environment, 2002, 79(1):105-115.

[3] Andersen H E, Mcgaughey R J, Reutebuch S E. Estimating forest canopy fuel parameters using LiDAR data[J]. Remote Sensing of Environment, 2005, 94(4):441-449.

[4] 焦义涛,邢艳秋,霍达,等.基于机载 LiDAR 点云估测林分的平均树高[J].西北林学院学报,2015,30(3):170-174.
Jiao Y T, Xing Y Q, Huo D, et al. Study on mean canopy height estimation from airborne LiDAR point cloud data[J]. Journal of Northwest Forestry University, 2015, 30(3):170-174.

[5] 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.
Hong Y F, Zhang S G, Chen W, et al. Inversion of biomass components for *Larix olgensis* plantation using airborne LiDAR[J]. Forest Research, 2019, 32(5):83-90.

[6] Li M, Im J, Quackenbush L J, et al. Forest biomass and carbon stock quantification using airborne LiDAR data: A case study over Huntington Wildlife Forest in the Adirondack Park[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 7(7):3143-3156.

[7] Monnet J M, Chanussot J, Berger F. Support vector regression for the estimation of forest stand parameters using airborne laser scanning[J]. IEEE Geoscience and Remote Sensing Letters, 2011, 8(3):580-584.

[8] 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.
Lu L, Zhou X C, Yu Z Z, et al. Plot-level forest height inversion using airborne LiDAR data based on the random forest[J]. Journal of Geo-Information Science, 2016, 18(8):1133-1140.

[9] 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 X, Yue C R, Li C G, et al. Estimation of forest stand mean height based on airborne LiDAR point cloud data[J]. Forest Research, 2020, 33(4):59-66.

[10] 王德智.结合 UAV-LiDAR 和卫星遥感数据的红树林多尺度观测方法研究[D].武汉:中国地质大学(武汉),2020.
Wang D Z. Integration of UAV-LiDAR data and satellite imagery for multi-scale mangrove observation[D]. Wuhan: China University of Geosciences(Wuhan), 2020.

[11] Tian Y C, Huang H, Zhang G Q, et al. Aboveground mangrove biomass estimation in Beibu Gulf using machine learning and UAV remote sensing[J]. Science of the Total Environment, 2021, 781:146816-146834.

[12] 廖宝文,郑松发,陈玉军,等.外来红树植物无瓣海桑生物学特性与生态环境适应性分析[J].生态学杂志,2004(1):10-15.
Liao B W, Zheng S F, Chen Y J, et al. Biological characteristics and ecological adaptability for non-indigenous mangrove species *Sonneratia apetala*[J]. Chinese Journal of Ecology, 2004(1):10-15.

[13] 邓必玉.广西钦州沿海无瓣海桑入侵性初探[J].湖北林业科技,2020,49(5):46-52.
Deng B Y. A preliminary exploration on the ecological invasion of *Sonneratia apetala* in Qinzhou area of Guangxi[J]. Hubei Forestry Science and Technology, 2020, 49(5):46-52.

[14] 田义超,黄远林,张强,等.北部湾典型海岛生态系统服务价值空间异质性对比研究[J].海洋科学,2019,43(2):60-68.
Tian Y C, Huang Y L, Zhang Q, et al. A comparative study of spatial heterogeneity of ecosystem service value in typical islands in Beibu Gulf[J]. Marine Sciences, 2019, 43(2):60-68.

[15] 周生祥.林分平均树高的求算与分析[J].华东森林经理,1989(3):37-45.
Zhou S X. Calculation and analysis of mean height of tree stand[J]. East China Forest Management, 1989(3):37-45.

[16] Baltsavias E P. Airborne laser scanning: Basic relations and formulas[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 1999, 54(2-3):199-214.

[17] 邵为真,赵富燕,梁周雁.基于不规则三角网的渐进加密滤波算法研究[J].北京测绘,2016(6):17-21.
Shao W Z, Zhao F Y, Liang Z Y. The research of improved progressive triangulated irregular network densification filtering algorithm[J]. Beijing Surveying and Mapping, 2016(6):17-21.

[18] Breiman L. Random forest[J]. Machine Learning, 2001, 45:5-32.

[19] Gholami R, Fakhari N. Support vector machine: Principles, parameters, and applications[J]. Handbook of Neural Computation, 2017, 27:515-535.

[20] 欧强新.基于机器学习的吉林天然针阔混交林生长建模[D].北京:中国林业科学研究院,2019.
Ou Q X. Modeling stand growth of natural conifer and broadleaf mixed forests in Jilin Province based on machine learning[D]. Beijing: Chinese Academy of Forestry, 2019.

[21] Akaike H T. A new look at the statistical model identification[J]. IEEE Transactions on Automatic Control, 1974, 19(6):716-723.

[22] Schwarz G E. Estimating the dimension of a model[J]. The Annals of Statistics, 1978, 6(2):461-464.

[23] Vrieze, Scott I. Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [J]. Psychological methods, 2012, 17(2): 228 – 243.

[24] 张晓君, 管伟, 廖宝文, 等. 珠海人工红树林与天然红树林群落特征比较研究 [J]. 生态科学, 2014, 33(2): 321 – 326.

Zhang X J, Guan W, Liao B W, et al. Comparison on the community characteristics between artificial and natural mangroves in Zhuhai City [J]. Ecological Science, 2014, 33(2): 321 – 326.

[25] 黄晓敏, 卢昌义. 厦门海湾典型无瓣海桑人工种群特征和幼苗更新扩散现状研究 [J]. 生态科学, 2018, 37(5): 1 – 6.

Huang X M, Lu C Y. Studies on the characteristics of artificial population of *Sonneratia apetala* and the current recruitment and dispersal of the seedlings in Xiamen Bay [J]. Ecological Science, 2018, 37(5): 1 – 6.

[26] 吴瑞, 陈丹丹, 王道儒, 等. 海南省东寨港红树林资源现状调查分析 [J]. 热带农业科学, 2016, 36(9): 62 – 71.

Wu R, Chen D D, Wang D R, et al. Survey of mangrove resources in Dongzhai Harbour of Hainan Province [J]. Chinese Journal of Tropical Agriculture, 2016, 36(9): 62 – 71.

[27] 郑俊鸣, 舒志君, 方笑, 等. 红树林造林修复技术探讨 [J]. 防护林科技, 2016(1): 99 – 103.

Zheng J M, Shu Z J, Fang X, et al. Discussion on techniques of mangrove afforestation and rehabilitation [J]. Protection Forest Science and Technology, 2016(1): 99 – 103.

[28] 廖宝文. 中国红树林恢复与重建技术 [M]. 北京: 科学出版社, 2010: 20 – 44.

Liao B W. Restoration and teconstruction techniques of mangrove forests in China [M]. Beijing: Science Press, 2010: 20 – 44.

[29] Lu D S, Chen Q, Wang G X, et al. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates [J]. International Journal of Forestry Research, 2012(2): 436537 – 436553.

[30] Brice M, Michael A W, Geordie W H, et al. Forest inventory stand height estimates from very high spatial resolution satellite imagery calibrated with LiDAR plots [J]. International Journal of Remote Sensing, 2013, 34(12): 4406 – 4424.

[31] Pourrahmati M R, Baghdadi N, Darvishsefat A A, et al. Mapping Lorey's height over Hyrcanian forests of Iran using synergy of ICE-Sat/GLAS and optical images [J]. European Journal of Remote Sensing, 2018, 51(1): 100 – 115.

[32] Waske B, Benediktsson J A, Árnason, et al. Mapping of hyperspectral AVIRIS data using machine – learning algorithms [J]. Canadian Journal of Remote Sensing, 2009, 35(1): 106 – 116.

Application of airborne LiDAR in the estimation of the mean height of mangrove stand

DENG Jingwen¹, TIAN Yichao^{1,2}, ZHANG Qiang¹, TAO Jin^{1,3}, ZHANG Yali¹, HUANG Shengguang³

(1. Collage of Resources and Environment, Beibu Gulf University, Qinzhou 535000, China; 2. Guangxi Key Laboratory of Marine Disaster in the Beibu Gulf, Beibu Gulf University, Qinzhou 535000, China; 3. Key Laboratory of Marine Geographic Information Resources Development and Utilization, Beibu Gulf University, Qinzhou 535000, China)

Abstract: This study constructed a stand mean height inversion model based on LiDAR data, aiming to provide dynamic monitoring of the growth of *Sonneratia apetala*. Taking the mangrove wetland of *Sonneratia apetala* in the Maowei Sea of Beibu Gulf as the study object and based on the height and intensity parameters extracted using airborne LiDAR data, this study compared three models, namely random forest, support vector machine, and neural network, based on the coefficient of determination (R^2), root mean square error ($RMSE$), Akachi information criterion (AIC), and Bayesian information criterion (BIC) and then estimated the mean height and spatial distribution of Mangrove in the study area using the optimal model. The results are as follows. The stand mean height of *Sonneratia apetala* in the study area is 3.90 ~ 11.58 m, and *Sonneratia apetala* with a higher tree height and a larger diameter at breast height is mainly distributed near the tidal trench and the middle part of the study area. In the estimation of the stand mean height of *Sonneratia apetala*, the maximum percentile height (h_{max}) had the highest contribution rate, followed by 75% ~ 99% percentile height. The random forest regression model yielded the highest precision ($R^2 = 0.938\ 1$, $RMSE = 0.58$ m, $AIC = 80.50$, and $BIC = 49.05$), followed by the support vector machine model ($R^2 = 0.766\ 5$ and $RMSE = 1.27$ m in the test stage), and the neural network regression model yielded the worst fitting effects. Overall, the random forest model is the optimal model for the inversion of the stand mean height of *Sonneratia apetala* in the study area.

Keywords: *Sonneratia apetala*; support vector machine; neural network; random forest; stand mean height

(责任编辑: 陈 理)