

基于多源数据并顾及森林类型的森林冠层高度反演

席志^{1,2} 邢艳秋^{1,2*} 贵珍^{1,2,3} 徐华东²

(1. 东北林业大学森林作业与环境研究中心哈尔滨 150040; 2. 东北林业大学机电工程学院哈尔滨 150040; 3. 自然资源部经济管理科学研究所(黑龙江省测绘科学研究所)哈尔滨 150081)

摘要:【目的】基于随机森林方法,通过 ICESat-2 的 LiDAR 数据与 Sentinel-1 的 SAR 数据、Sentinel-2 的光学影像以及地形数据的协同构建针、阔、混三种不同森林类型对应的森林冠层高度估计模型,并对不同森林类型对应模型的精度进行验证和比较;同时,利用随机森林方法获取不同模型对应建模最优变量集及集合中各变量重要性得分,并对各变量在模型建立过程中发挥的作用进行定量地比较和分析。【方法】首先,在不同空间分辨率下,通过机载 lidar 反演得到的冠层高度验证 ICESat-2 卫星的 ATL08 产品提供的高度指标(RH95)的精度;随后,获取 SRTM 数据提供的高程和坡度数据、提取 Sentinel-1 卫星的 SAR 数据提供的 VV 和 VH 变量、Sentinel-2 卫星的光学影像提供的纹理特征、植被指数和生物物理特征、通过地理国情监测获取的研究区域内的针/阔/混森林覆盖信息,基于随机森林方法并利用以上参数构建出不区分森林类型和针/阔/混森林类型对应的冠层高度估计模型,并对模型的精度进行验证和比较;最后,利用随机森林方法获取不同森林类型对应的最优变量集,并对集合中各变量在模型构建过程中发挥的作用进行定量地比较和分析。【结果】在 250 m 空间分辨率下,ATL08 提供的 RH95 高度指数与机载 lidar 反演的冠层高度具有最好的一致性,对应的 R 和 RMSE 分别为 0.80 和 1.98 m;不区分森林类型建立的冠层高度估计模型的反演精度低于按照针/阔/混森林类型分别建立的冠层高度估计模型的精度,以上四个模型对应的两个精度指标 R 和 RMSE 分别为:0.59、0.72、0.59、0.62 和 3.68 m、3.15 m、3.37 m、3.26 m;在模型构建过程中,不同模型对应的最优变量集中包含变量种类及个数均不相同,且同一变量在不同集合中对应的重要性得分也存在明显差异。【结论】本研究通过多源数据的协同建立的森林冠层高度估计模型可以准确地获取大区域内空间连续的森林冠层高度信息,且根据不同森林类型分别建立的估计模型的精度明显高于未区分森林类型所建立估计模型的精度,证明了区分森林类型建立估计模型的必要性。不同森林类型对应的最优变量存在差异,但是基于光学数据提取的植被指数在每个集合中的重要得分之和均高于其他类型变量对应的重要性得分,证明了通过多源数据联合反演森林冠层高度时植被指数的重要性。

关键词: ICESat-2; Sentinel-1; Sentinel-2; 地形信息; 森林类型; 冠层高度; 随机森林; 最优变量

Forest canopy height inversion considering forest types based on multi-source data

Abstract: 【Objective】 Based on the Random Forest method, forest canopy height estimation models corresponding to coniferous forest, broadleaf forest and mixed forest were established by synergizing ICESat-2 lidar data, Sentinel-1 SAR data, Sentinel-2 optical images and topographic data, and then the accuracy of the established models for different forest types were validated and compared. Meanwhile, the optimal variable set corresponding to each forest type and importance score of each variable in the set was obtained with the help of Random Forest, and then the contribution of each optimal variable in the model establishment process was compared and analyzed quantitatively. 【Method】 Firstly, the accuracy of the RH95 height metric provided by ATL08 product of ICESat-2 satellite is validated by comparing with the canopy height derived from airborne lidar data at different spatial resolutions. Subsequently, elevation and slope data provided by SRTM product, VV and VH provided by Sentinel-1 SAR data, texture features, vegetation index and biophysical features provided by Sentinel-2 optical images were obtained. And then the cover information of coniferous/broadleaf/mixed forest within the study area was obtained through the NGCM (National Geographical Condition Monitoring Project) data. Based on the Random Forest method and the above derived parameters, the canopy height estimation models for whole forest (Not distinguishing forest types) and coniferous/broadleaf/mixed forest were established, and then the accuracy of each model was validated and compared. Finally, the optimal variable set corresponding to different forest types was obtained though Random Forest method and the contribution of each optimal variable in establishing the estimation model was compared and analyzed quantitatively. 【Results】 The RH95 height metric provided by ATL08 product has the best consistency with the canopy height derived

from airborne lidar data at a spatial resolution of 250 m with the R and RMSE of 0.80 and 1.98 m, respectively. The accuracy of canopy height estimation model established without distinguishing forest types is lower than that of estimation models established for coniferous/broadleaf/mixed forest. The two accuracy indicators R and RMSE corresponding to whole/coniferous/broadleaf/mixed forest are 0.59, 0.72, 0.59, 0.62 and 3.68 m, 3.15 m, 3.37 m, 3.26 m, respectively. In the process of model establishment, the optimal variable set corresponding to different models contains different types of variables and the number of variables within each set varies obviously, and there are also significant differences in the importance scores of the same variable in different sets. **【Conclusion】** Spatially continuous canopy height for a large area with high precision can be obtained by the estimation models established in this study by synergizing multi-source data. And the accuracy of estimation models considering forest types is significantly higher than that of estimation models established for whole forest, proving the necessity of establishing estimation models for different forest types. There are obvious differences in the optimal variables corresponding to different forest types, but the sum of importance scores of vegetation indices extracted from optical data in each set is higher than that of other types of variables, demonstrating the importance of vegetation indices in predicting forest canopy height through multi-source data.

Keywords:ICESat-2; Sentinel-1; Sentinel-2; topographic information; forest type; canopy height; Random Forest; optimal variable

森林冠层高度作为森林的重要属性之一，是研究生物量分配、碳储量变化、森林生产力和生物多样性的关键参数(Dong et al., 2003; Lefsky et al., 2002)，准确地获取森林冠层高度并掌握其时空变化将有助于森林管理和经营决策(Simard, et al., 2011; Zhang et al., 2016)。尽管实地调查等传统方法可以获取高精度的森林冠层高度，但该方法耗时且费力，难以在短时间内获取较大区域内空间连续的森林冠层高度信息。遥感技术是最强大的对地观测技术之一，该技术具有速度快、精度高和覆盖范围广的特点，被越来越广泛地用于植被高度提取 (Rosenqvist et al., 2003)。在所有遥感技术中，激光雷达 (light detection and ranging, lidar) 由于可以在垂直方向上对森林进行观测，从而能够直接获取森林结构的三维信息，因此在过去二十年中对森林冠层高度的提取做出了重大贡献。近年来，大量的研究使用机载 lidar 提取森林冠层高度，较高的精度证明了该技术在森林冠层高度提取中的巨大潜力 (Wulder et al., 2012; Alexander et al., 2018)。

然而，借助机载 lidar 技术虽然可以获取高精度的森林冠层高度，但较高的数据采集成本使得难以利用该技术获取较大范围内的森林冠层高度。星载 lidar 技术的出现很好地解决了上述问题，较短的重访周期和较大的覆盖范围使得利用星载 lidar 在短时间内获得区域尺度甚至全球的森林冠层高度成为了可能。在所有星载 lidar 中，美国国家航空航天局 (National Aeronautics and Space Administration, NASA) 于 2018 年发射的 ICESat-2 卫星得到了越来越多的关注，在森林冠层高度测量中得到了广泛应用。该卫星系统使用光子计数激光测高系统来测量植被高度 (Markus et al., 2017; Magruder et al., 2018)，提供了从 ATL00 至 ATL21 共 21 个产品，其中 ATL08 产品可以提供沿其轨道、不同尺度的森林冠层信息，通过与利用机载 lidar 数据提取的冠层高度进行对比，该产品的精度已经得到了证实 (Neuenschwander et al., 2019)。然而，ATL08 产品只能以 100×14 m 为基本单元提供沿轨的森林冠层高度，无法通过该产品直接获得空间连续的森林冠层高度信息(Xing et al., 2020)。

为了解决上述问题，常见的方法是将空间连续光学数据与 lidar 数据相结合以构建森林冠层高度估计模型 (Li et al., 2020; Lin et al., 2020)，欧洲航天局 (European Space Agency, ESA) 于 2015 年发射的 Sentinel-2 号卫星已逐渐成为一个受欢迎的选择，因为该卫星能够在较短时间内提供覆盖全球的光学影像数据。近年来，Sentinel-2 卫星的数据已被一些研究用于提取森林冠层高度，并取得了令人满意的结果。Lang et al. (2019) 通过机载激光 lidar 和 Sentinel-2 数据的联合成功获取了国家尺度的植被高度，证实了从 Sentinel-2 数据中提取的光谱指数和纹理特征对于获取植被高度的重要作用。合成孔径雷达 (Synthetic Aperture Radar, SAR) 是一种主动的对地观测技术，可以全天候的采集观测数据并不受天气的影响。因此，该技术可以弥

补光学数据采集容易受到气候条件影响的缺点。同时,由于该技术具有一定的表面穿透能力,其信号可以穿透森林冠层,从而获得更多与结构和密度相关的信息。在所有搭载 SAR 传感器的卫星中, Sentinel-1 卫星已被广泛地用于估计森林冠层高度, Nandy et al. (2021) 首次尝试通过 ICESat-2 星载 lidar 数据和 Sentinel-1 的 SAR 数据的协同进行森林冠层高度估计,证实了 Sentinel-1 卫星携带的 C 波段 SAR 提供的反向散射 (VH 和 VV) 在构建森林冠层高度估计模型中的重要作用。

机器学习 (Machine Learning) 方法具有计算复杂度低、无需参数调整、分类精度高和回归能力强等优点,被广泛地用于多源遥感数据的融合 (Liu et al., 2018; Maxwell et al., 2018)。其中,最具代表性的是随机森林 (Random Forest, RF) 方法,该方法基于装袋和随机特征选择,以提供高精度的分类结果而闻名 (Belgiu et al., 2016),作为一种重要的数学建模算法,已被广泛用于建立森林属性与多源数据中提取的变量之间的模型 (Ahmed et al., 2015; Lary et al., 2016; Tian et al., 2014)。同时,借助 RF 自身功能,可以对建模变量进行筛选,从而获取所建模型对应的最优变量集和集合中各变量的重要性分值,并以此为依据对各变量在模型构建过程中发挥的作用进行定量地比较和分析。

因此,本研究基于随机森林方法,通过 ICESat-2 星载 lidar 数据、Sentinel-1 的 SAR 数据、Sentinel-2 的光学影像及与森林生长和分布密切相关的地形数据的协同,分别构建了区分森林类型 (针/阔/混) 的森林冠层高度反演模型,并对模型的精度进行了验证和对比;同时,对模型构建过程中各变量的作用进行定量地比较和分析。本研究构建的冠层高度反演模型证实了通过多源数据联合获取大区域内空间连续的森林冠层高度的可能以及区分森林类型分别构建反演模型的必要;同时,本研究获取的不同森林类型对应的最优变量集及集合中各变量的重要性分值将对建立森林冠层高度反演模型时变量的选择具有重要的参考意义。

1 研究区概况

研究区位于黑龙江省的桦南县 ($46^{\circ}14'-46^{\circ}17'N$, $130^{\circ}34'-130^{\circ}42'E$) 范围内。桦南县位于黑龙江省东部,长白山余脉完达山麓,面积约为 $4.42 \times 10^3 km^2$,研究区域地势东北部高、西南部低。研究区域的气候属寒温带大陆性季风气候,四季分明,年平均气温和降水量分别为 $3.6^{\circ}C$ 和 $523.4 mm$,每年 5-9 月是降水集中期 ([http://http://hl.cma.gov.cn/](http://hl.cma.gov.cn/))。桦南林区是黑龙江省主要的林区之一,也是省内最重要的木材生产基地。研究区域内的森林覆盖率约为 35.1% (2017 年) 且树木种类繁多,主要包括红松、水曲柳、冷杉、白桦等;除森林外,研究区域的土地覆盖主要为农作物 (玉米和大豆) 和建设用地。研究区域的地理位置、机载激光雷达数据覆盖区域以及研究区域内的针/阔/混森林分布如图 1 所示。

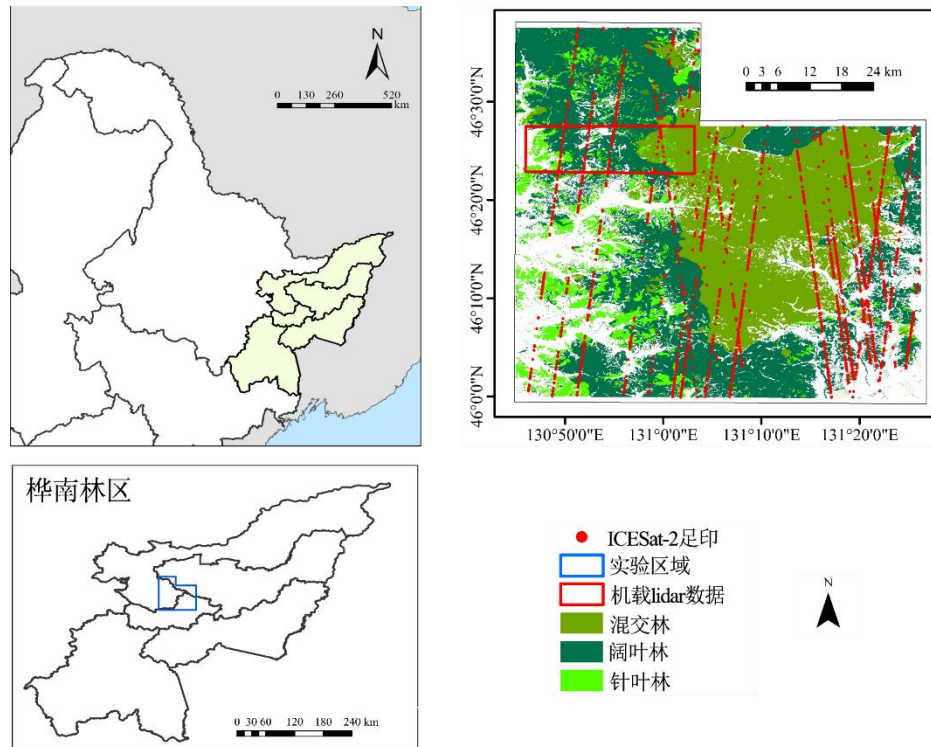


图1 研究区域位置、机载激光雷达数据覆盖区域及 ICESat-2 足印和不同森林类型空间分布情况

Fig. 1 Location of the study area and spatial distribution of airborne lidar data, ICESat-2 footprints and forest types within study area

2 研究方法

2.1 数据获取及预处理

本研究共使用了 ICESat-2 的星载 lidar 数据、Sentinel-1 的星载 SAR 数据、Sentinel-2 的星载光学影像和地形数据用于构建森林冠层高度估计模型；利用地理国情监测数据获取研究区域内不同森林类型的覆盖信息；机载激光 lidar 数据被用于验证所构建模型的精度。

2.1.1 ICESat-2 数据

发射于 2018 年的 ICESat-2 (Ice, Cloud and land Elevation Satellite) 卫星最初的设计目的是提供对海冰厚度、冰盖高度和全球植被高度的连续观测，并通过收集温带地区的测量数据为全球碳储存的调查做出贡献。ICESat-2 卫星上搭载了光子计数的激光雷达 (photon-counting lidar)，可以以较高的重复率探测单个光子，其激光被分为六束 (三对)，根据携带的能量，每对激光包含一个强光束和一个弱光束，二者对应的能量比为 4:1 (Neumann et al., 2019)。ICESat-2 提供的 ATL03 产品是一种经过地理定位的光子云产品，常被用作生成更高等级产品的输入数据。利用 DARGANN (Differential, Regressive, and Gaussian Adaptive Nearest Neighbor) 算法，对 ATL03 中的太阳背景噪声进行处理以得到信号光子，随后对信号光子进行过滤处理，并

将其划分为冠层顶部光子、冠层光子和地面光子三种类型 (Neuenschwander et al., 2019;)。基于标记的光子, 以 $100 \times 14 \text{ m}$ 为基本单元生成 ATL08 产品, 该产品不仅可以提供森林冠层高度的信息, 还可以获得其他与植被结构相关的高度指标 (表 1)。

表 1 ATL08 产品的主要垂直结构指标

Tab.1 Main vertical structure metrics of ATL08 product

冠层高度指标 Canopy height index	描述 Description
Hcanopy _{max}	单个冠层高度的相对最大值
Hcanopy _{mean}	单个冠层高度的相对平均值
Hcanopy _{min}	单个冠层高度的相对最小值
RH90/95/98	90/95/98 分位计算的高度指数

根据之前的研究成果, 将 ATL08 产品的 RH95 指标作为构建冠层高度估计模型的因变量 (Neuenschwander et al., 2019)。本研究从美国国家冰雪数据中心 (National Snow & Ice Data Center) 下载了 2019 年和 2020 年所需的 ATL08 数据, 根据 ATL08 产品提供的不确定高度属性, 对不确定高度大于不确定高度平均值的数据进行移除处理; 随后, 借助于 NGCM 数据直接获得研究区域内森林/非森林分布信息, 并以此为依据选取了森林区域的 ATL08 数据。

2.1.2 Sentinel-1 数据

Sentinel-1 卫星系统由两颗极轨卫星 (1A 和 1B) 组成, 该系统重访周期为 6-12 天, 通过 C 波段的 SAR 获取全球范围的观测数据。由于 SAR 的穿透能力, 系统数据采集不受天气和阳光照射的影响且具有全天候获取数据的能力 (Li et al., 2020)。同时, 系统提供的后向散射 (如 VV 和 VH) 对表面和体积散射十分敏感且该技术支持单极化 (HH 或 VV) 和双极化 (HH+HV 或 VV+VH, H: 水平; V: 垂直) 作业模式 (Li et al., 2020)。本研究共使用了 2020 年生长季节 (6-9 月) 的 42 张 Sentinel-1 图像 (空间分辨率: 10 m), 所有数据均通过 Google Earth Engine (GEE) 提供的 “Sentinel-1 SAR GRD” 数据集进行下载, 并对每个波段进行了中值合成处理。

2.1.3 Sentinel-2 数据

Sentinel-2 卫星 (Sentinel-2A 和 Sentinel-2B) 是一个宽波段、高分辨率和多光谱成像系统, 该系统提供具有多种空间分辨率的多光谱数据且可以获得不同波段的表面特征信息, 被广泛地应用于对水、土壤和植被的等监测, 特别是该系统搭载的 MSI 传感器发射的 red-edge 波段对植被生长状况十分敏感, 可以提供更准确的植被生长信息 (Liu et al., 2018)。由于建模中使用的许多特征与树叶有关, 因此本研究选取了 2020 年生长季 (6-9 月) 共 186 张不受云层影响的 Sentinel-2 图像, 分别选择了 10 m 波段 (blue: 490 nm , green: 560 nm , red: 665 nm , NIR: 842 nm) 和 20 m 波段 (swir1, swir2 和 red-edge 波段)。所选择数据均通过 Copernicus Scientific Data Hub、SNAP 和 GEE 进行下载和去云、重采样等处理, 并提取了建模所需的植被指数和纹理特征等数据。

2.1.4 SRTM 数据

本研究使用 SRTM (Shuttle Radar Topographic Mission) 产品获取研究区域的高程信息, 该产品的空间分辨率为 30m, 可以提供高程、坡度和坡向等与植被的分布和生长密切相关的地形数据 (吴英等, 2012; Ying et al., 2012; Zhang et al., 2016)。在本研究中, 利用 GEE 调用并处理了 SRTM 数据, 获取了覆盖研究区域的高程信息并通过计算获得了坡度和坡向数据。

最后, 共获取了 23 个用于建模的变量, 所有变量被分成“生物物理特征”、“植被指数”、“纹理特征”、“VV&VH”和“地形信息”共五类, 所选取变量的具体信息如表 2 所示。

表 2 估计模型建立所需因变量

Tab. 2 Independent variables for establishing estimation models

变量名称 Variable name	描述 Description	分类 Category
LAI	叶面积指数	生物物理特征
FAPAR	光合有效辐射比	
FCOVER	植被覆盖率	
RVI	NIR/R	植被指数
EVI	$2.5 * ((NIR - R) / (NIR + 6 * R - 7.5 * B + 1))$	
DVI	NIR-R	
MSAVI	$(2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - R)}) / 2$	
NDVI_B84	$(NIR - R) / (NIR + R)$	
NDVI_B85	$(NIR - RE1) / (NIR + RE1)$	
NDVI_B86	$(NIR - RE2) / (NIR + RE2)$	
NDVI_B87	$(NIR - RE3) / (NIR + RE3)$	
NDVI_B8A4	$(NIR2 - R) / (NIR2 + R)$	
NDVI_B8A5	$(NIR2 - RE1) / (NIR2 + RE1)$	
NDVI_B8A6	$(NIR2 - RE2) / (NIR2 + RE2)$	
NDVI_B8A7	$(NIR2 - RE3) / (NIR2 + RE3)$	
ndvi_Contrast	用于图像分类	纹理特征
ndvi_Entropy	用于图像分类	
ndvi_GLCM_Variance	用于图像分类	
VV	后向散射值	VV&VH
VH	后向散射值	
elevation	高程	地形信息
aspect	坡度	
slope	坡向	

2.1.5 地理国情监测数据

国家地理国情监测项目 (National Geographical Condition Monitoring Project, NGCM) 于 2013 年启动, 该项目通过各级政府、不同部门间的合作完成, 旨在通过对中国境内的地理国情进行动态地测绘和统计, 从而获取反应中国地理条件的权威信息 (Zhao et al., 2016)。NGCM 产品地表类别划分为 12 个一级类、58

个二级类和 135 个三级类，利用了高分辨率遥感数据和现场实测数据对该产品的准确性进行了评估(Yang et al., 2017)。此外，通过“两级检验、一级验收”和“过程抽样检验和试验后审查”对 NGCM 产品的质量进行控制(Wu et al., 2017)。本研究选择了 2016 年收集的覆盖研究区域的 NGCM 数据，并借助该数据获取了研究区域内不同森林类型（针叶林、阔叶林和混交林）的覆盖信息。

2.1.6 机载 lidar 数据

本研究利用 SZT-R250 测量系统于 2019 年 6 月获得了高精度 ALS 数据，数据的覆盖范围如图 1 所示。SZT-R250 测量系统由中国公司独立研发，该系统集成了高精度激光扫描仪、全球导航卫星系统 (Global Navigation Satellite System, GNSS)、惯性测量单元 (Inertial Measurement Unit, IMU)、时间同步模块和控制模块，该系统上的搭载的激光雷达为 MiniVUX-1UAV，其发射频率和测距精度分别为 100 kHz 和 0.015 m。在 GNSS 和 IMU 的帮助下，系统提供了多种定位模式组合，可以以 64.5 pts/m² 的平均密度获得实时点云数据。本研究使用北京数字绿土公司提供的 LiDAR360 软件对获取的机载 lidar 数据进行预处理，通过计算冠层光子的高程与地面高程的差值获取冠层高度，并以格网为单元计算得到格网内各冠层高度的平均值，即为该格网对应的冠层高度真实值，用于验证预测得到的冠层高度值的精度。

最后，基于 WGS-84 椭球对本研究使用的所有数据进行通用横向墨卡托 (UTM) 投影 (52N)；随后，根据第 3.1 节中确定的最优分辨率，对所有指标进行合成处理并重新采样至 250 m 分辨率。

2.2 随机森林方法

本研究通过随机森林方法建立 ICESat-2 提供的冠层高度指标与基于多源数据提取的多个建模变量间的联系，构建针对不同森林类型的冠层高度估计模型；随后，利用随机森林提供的功能获取不同森林类型对应的最优变量集和集合中各变量的重要性得分。随机森林 (Random Forest, RF) 是一种基于决策树的分类方法 (Mandianpari et al., 2017)，由多个集成回归或分类树构成，使用“bagging”或“bootstrap”算法来构建大量不同的训练子集 (Tian et al., 2016; Dong et al., 2012)，集成分类的原理可表示为：

$$h(x, \theta_k), k = 1, 2, \dots \quad (1)$$

式中： x 为输入的样本向量； k 为决策树的个数； θ_k 为第 k 个决策树的参数向量，它是独立且恒等分布的随机向量。对于非训练样本，每个决策树将提供各自的分类结果，决策树将为此结果“投票”，最终的分类结果由最多票数决定 (Shao et al., 2018)。

由于 RF 方法在实现过程中随机化的引入增强了去除噪声的能力，使用该算法建模不易出现过拟合的情况，并且可以在不删除变量的情况下处理数千个输入变量。同时，通过置换的方法可以得到变量的重要性。该方法首先创建一个树，然后使用袋外 (out-of-bag, OOB) 观测计算精度 (Breiman et al., 2001)；然后，对所有输入变量进行迭代排列，计算每个变量对应的重要性得分，并将每次获得的变量重要性得分的平均值作为变量重要性得分的最终值。构建模型对应的最优变量集可以通过以下步骤获得：

- (1) 使用所有候选变量来建立估计模型，然后计算每个候选变量的 OOB 值和重要性得分；
- (2) 寻找得分最低的变量并对该变量作剔除处理，利用剩余的变量重建模型；
- (3) 重复这些步骤，计算每次获得的 OOB 值和变量重要性得分，最小 OOB 值对应的变量集合即为该

模型的最优变量集合，此时集合中各变量的重要性得分即为该变量最终的重要性得分。

3 结果与分析

3.1 ATL08 产品的精度

研究区域内的 ATL08 产品的 RH95 高度指标提供的森林冠层高度位于 2.09-36.65m 之间，且平均值为 16.07m，本研究将机载 lidar 反演得到的冠层高度作为冠层高度的真实值，利用该真值对基于 ATL08 产品获取的森林冠层高度的进行验证。为了验证研究区域 RH95 指标的准确性，以 ICESat-2 足印的地理位置为质心，计算了不同空间尺度下基于机载 lidar 点云数据反演得到冠层高度与 RH95 高度指标间的一致性。本研究选择的机载 lidar 数据包含了 145 个有效点并覆盖了针、阔、混三种森林类型，通过对机载点云数据进行后处理，共获取了包括 h_{mean} 、 h_{median} 、 h_{90} 、 h_{95} 和 h_{max} 在内的五个高度指标 (Li et al., 2015)。根据常用卫星影像的空间分辨率，选择在 10m、30m 和 250m 三个分辨率下通过机载 lidar 数据获取冠层高度，并计算其与 ATL08 提供的高度指标间的 R 和 RMSE 两个精度指标，结果如表 3 所示。

表 3 不同分辨率下 ATL08 产品和基于机载 lidar 反演高度指标间的 R 和 RMSE

Tab. 3 R and RMSE between ATL08 product and height metrics derived from airborne lidar data at different resolutions

分辨率 (Resolution)	10 m					30 m					250 m				
	h_{mean}	h_{median}	h_{90}	h_{95}	h_{max}	h_{mean}	h_{median}	h_{90}	h_{95}	h_{max}	h_{mean}	h_{median}	h_{90}	h_{95}	h_{max}
R	0.56	0.23	0.68	0.69	0.71	0.59	0.54	0.73	0.70	0.68	0.72	0.68	0.80	0.76	0.75
RMSE (m)	3.08	3.29	2.75	2.71	2.69	2.90	2.98	2.59	2.64	2.87	2.21	2.45	1.98	2.10	2.18

由表 3 可知，不同空间分辨率下对应的 R 和 RMSE 存在明显的差异，对应于 10 m、30m 和 250 m 分辨率的 R 和 RMSE 分别位于 0.23-0.71、0.54-0.73、0.68-0.80 和 2.69-3.29 m、2.59-2.98 m、1.98-2.45 m 之间，证明了机载 lidar 反演的冠层高度和 ATL08 产品提供的冠层高度之间具有较高的一致性，特别是在 250 m 空间分辨率下，二者具有最好的一致性，对应的 R 和 RMSE 分别为 0.80 和 1.98 m。相比而言，在其他两个分辨率下，二者之间的一致性相对较低，因此，本研究最终选择在 250m 分辨率下建立冠层高度估计模型，所有的数据均被重采样至该分辨率下。随后，按照以下两种情况确定每个 250×250 m²对应的冠层高度：

- (1) 如果格网中仅包含一个 ATL08 坐标，则 RH95 指标提供的高度即为该格网的冠层高度；
- (2) 如果格网中包含两个或以上 ATL08 坐标，则将 RH95 指标的平均值作为该格网的冠层高度。

最终，共选择了 2647 个样本点用于模型的训练和验证 (8:2)，不同森林类型对应的样本数量如表 4 所示。

表 4 模型建立时不同森林类型对应的训练/验证样本数量

Tab. 4 Number of training/validation samples corresponding to different forest types for establishing inversion models

森林类型	训练样本	验证样本	样本总量
Forest type	Training sample	Validation sample	Total
不区分类型	2117	530	2647

针叶林	292	74	366
阔叶林	1028	258	1286
混交林	796	199	995

3.2 森林冠层高度反演建模及精度评价

本研究利用随机森林方法建立了 ICESat-2 提供的 RH95 高度指标和基于 Sentinel-1 的 SAR 数据、Sentinel-2 的光学影像和地形数据提取的变量间的联系，构建了不区分森林类型和针、阔、混三种森林类型对应的冠层高度估计模型，并基于随机森林方法获取了不同模型对应的最优变量集。随后，利用验证样本对所构建模型的精度进行验证，并计算了不同模型对应的 R 和 RMSE 两个精度指标，结果如图 2 所示。

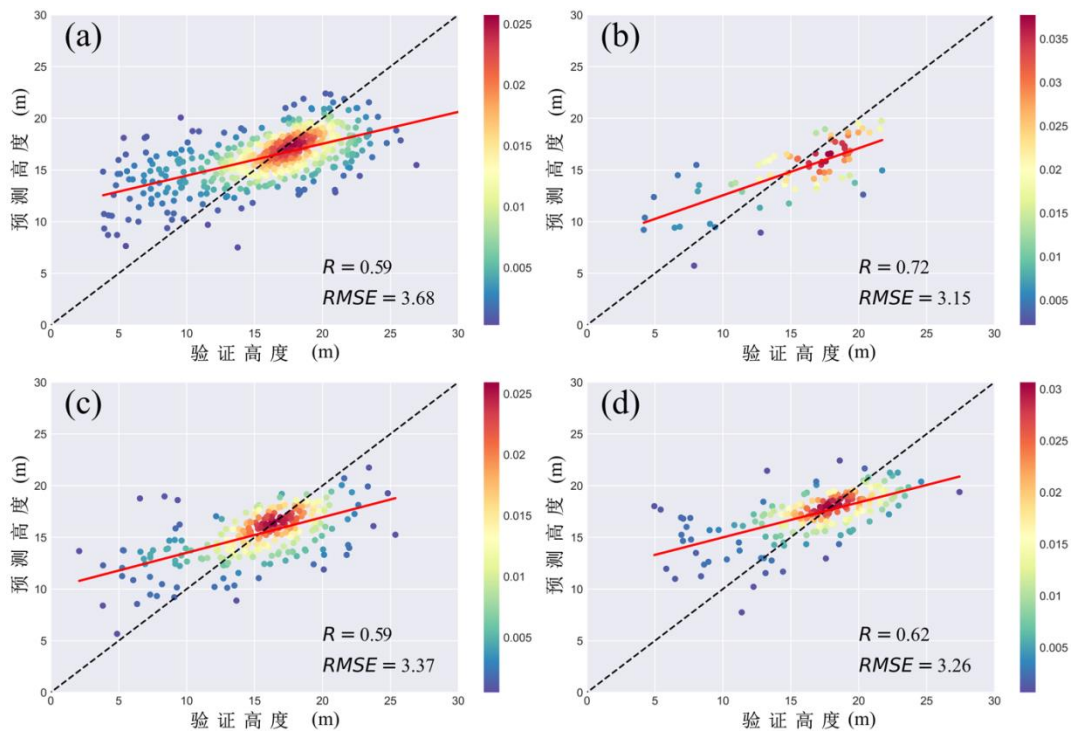


图 2 验证样本与基于 RF 预测的冠层高度之间的相关性
(a、b、c、d 分别代表不分区森林类型、针、阔、混森林类型对应的结果)。

Fig. 2 Correlations between validation samples and predicted forest canopy heights by RF

(a, b, c and d represent the results of whole forest, coniferous forest, broadleaf forest and mixed forest)

由图 2 可知，不同模型对应的 R 和 RMSE 分别位于 0.59-0.72 和 3.15-3.68 m 之间，证明了通过所构建模型获得的冠层高度和验证样本提供的冠层高度间具有较高的一致性，但不同模型对应的 R 和 RMSE 之间存在明显的差异。如图 3(b)所示，针叶林模型具有最好的表现，对应的 R 和 RMSE 分别为 0.72 和 3.15m，且当树高低于 15m 时，存在明显高估；进一步计算了预测结果和验证结果之间差值的平均值，结果显示平均低估值约 34.4 cm。混交林对应的估计模型精度排在第二位，对应的 R 值为 0.62、RMSE 值为 3.26m，且当树高低于 17.5 m 时，存在明显的高估 (图 3(d))。统计结果表明预测结果和验证结果之间差值的平均值为 48.00 cm，意味着利用该模型估计混交林的冠层高度时约存在 48.00 cm 的高估。如图 3(c) 所示，对应于阔叶林估计模型的 R 值和 RMSE 值分别为 0.59 和 3.37 m，区分高估和低估的高度阈值为 16m，该模型对应的高估平均值约为 3.90 cm。对整个森林的整体建模获得的模型精度最差，对应的 R 值为 0.59，RMSE 值为 3.68m(图 3(a))。整体来看，按照森林类型分别建立估计模型的精度明显高于不区分森林类型建立估计模

型的精度，且针叶林模型精度>混交林模型精度>阔叶林模型精度。

为了进一步验证本研究构建模型的精度和比较不同模型间的性能，本研究选择单一森林类型覆盖的子区域，并在子区域范围内对所建模型的精度进行验证。借助 NGCM 数据和 Google 影像，在研究区内选择单独生长有针、阔、混三种森林类型的子区域 (图 3)。

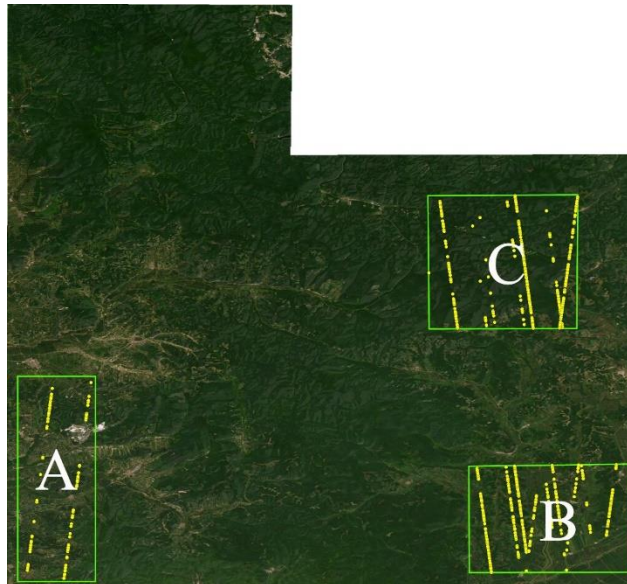


图 3 单一森林覆盖子区域 (A、B、C 分别代表针叶林、阔叶林和混交林)

Fig. 3 Sub-regions covered by single forest type (a, b and c represent coniferous forest, broadleaf forest and mixed forest, respectively)

如图 3 所示，子区域 A、B 和 C 分别仅有针叶林、阔叶林和混交林覆盖，各子区域内包含的样本数量分别为 76、325 和 340，对每个区域内通过构建模型预测得到的冠层高度和 RH95 指标进行交叉比较以获得高度差值的直方图，结果如图 4 所示。

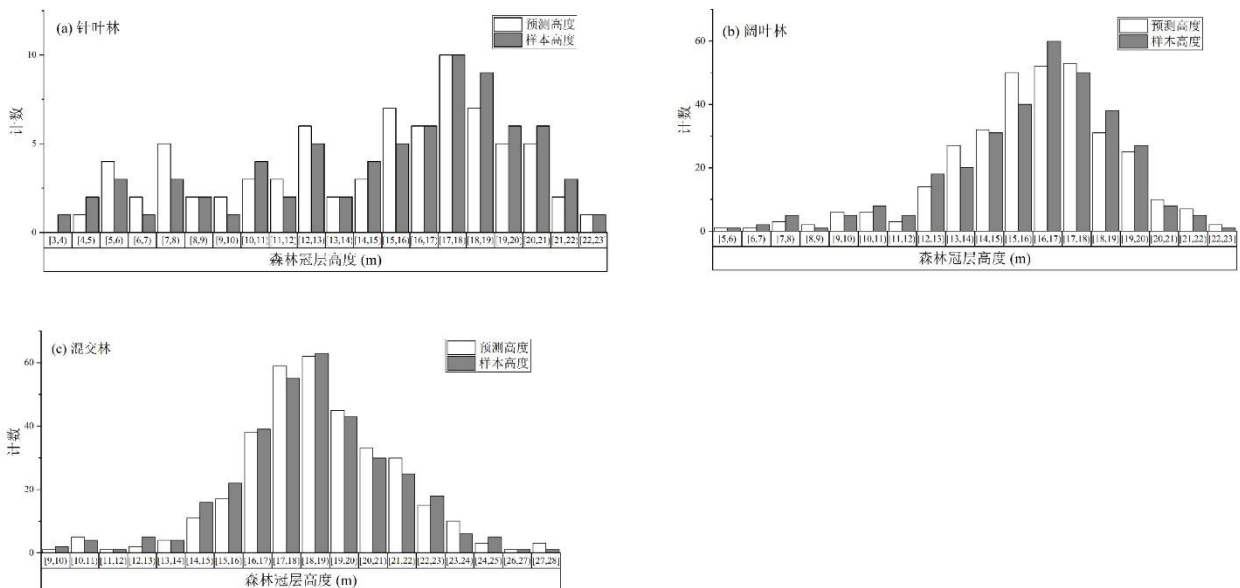


图 4 不同森林类型对应预测冠层高度和 RH95 高度差值直方图

Fig. 4 The histograms of difference between predicted canopy height and RH95 height metric for coniferous/broadleaf/mixed forest.

由图 4 可知，所有森林类型的预测森林冠层高度和样本森林冠层高度之间具有较高的一致性，为了进一步验证本研究构建估计模型的精度，计算了预测高度与样本高度之间差值的最小值、平均值和最大值，

针叶林模型对应的最小差值、平均差值和最大差值分别为 $0.01m$ 、 $1.80m$ 和 $9.23m$ ；阔叶林模型对应的差值最小值、平均值和最大值分别为 $0.29m$ 、 $2.79m$ 和 $11.37m$ ；混交林对应的差值最小值、平均值和最大值分别为 $0.01m$ 、 $1.85m$ 和 $8.92m$ 。以上结果证明了针叶林对应的冠层高度预测模型具有最好的性能。

3.3 研究区域森林冠层高度预测及精度评价

通过以上结果可知，根据不同森林类型构建的冠层高度估计模型的精度要高于不区分森林类型整体建模的精度，因此，借助本研究构建的冠层高度估计模型获取了研究区域内空间连续的森林冠层高度信息并统计了不同森林类型对应的冠层高度情况，结果如图 5 所示。

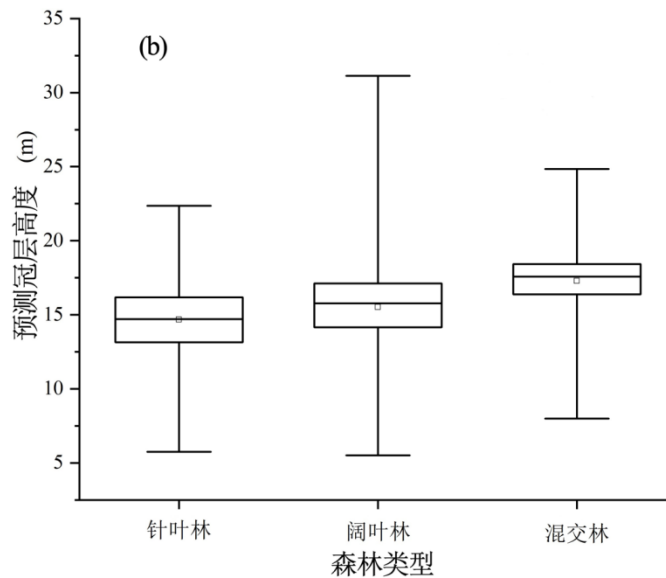


图 5 基于构建反演模型预测的研究区域森林冠层高度

Fig. 5 Forest canopy height in the study aera predicted by the established inversion models

由图 5 可知，、通过预测得到的针叶林、阔叶林和混交林冠层高度分别位于 $5.54-22.51m$ 和 $5.15-31.24m$ 和 $7.65-25.21m$ 之间。为了进一步验证通过预测得到的冠层高度的准确性，本研究将预测得到的冠层高度与 GEDI (Global Ecosystem Dynamic Investigation) 卫星提供的冠层高度进行了交叉对比，结果如图 6 所示。

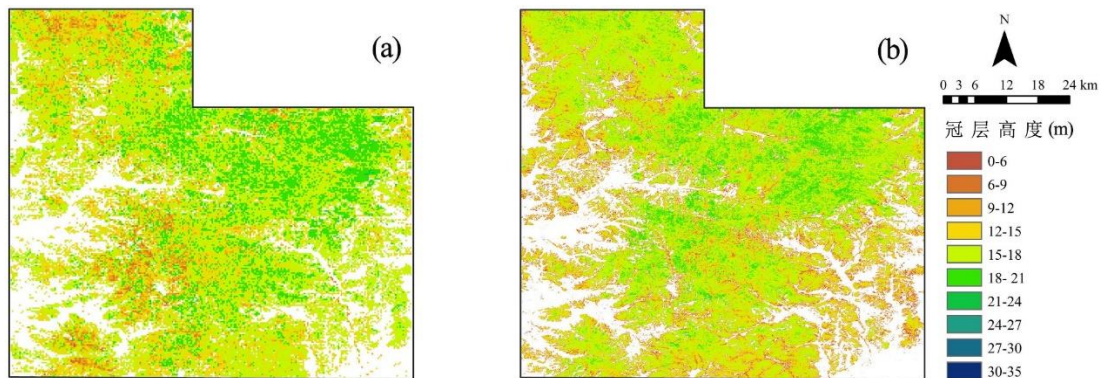


图 6 预测得到的冠层高度 (a) 及 GEDI 提供的冠层高度 (b)

Fig. 6 Forest canopy height by prediction (a) and provided by GEDI (b)

由图 6 所示，通过目视解译可知，两幅图展示的研究区域内森冠层高度的空间分布格局具有较高的一致性，即较高的冠层高度主要位于研究区域的北部及中部区域，较低的冠层高度主要位于研究区域的边缘。

同时还可以发现,对应预测得到的冠层高度大多数像元内的森林冠层高度位于 $18\text{--}24\text{ m}$ 之间,而基于 GEDI 反演得到的冠层高度,主要位于 $15\text{--}21\text{ m}$ 之间,二者之间存在较大的重叠区间。随后,本研究将 GEDI 冠层高度产品重采样至 $250\times 250\text{ m}^2$ 空间分辨率,并通过 ArcGISPro 软件对预测得到的冠层高度数据与 GEDI 冠层数据进行了配准。最后,通过“逐像元”作差的方式获取了研究区内所有像元对应的冠层高度的差值 (Ghosh et al., 2020),并统计了差值的频率直方图,结果如图 7 所示。

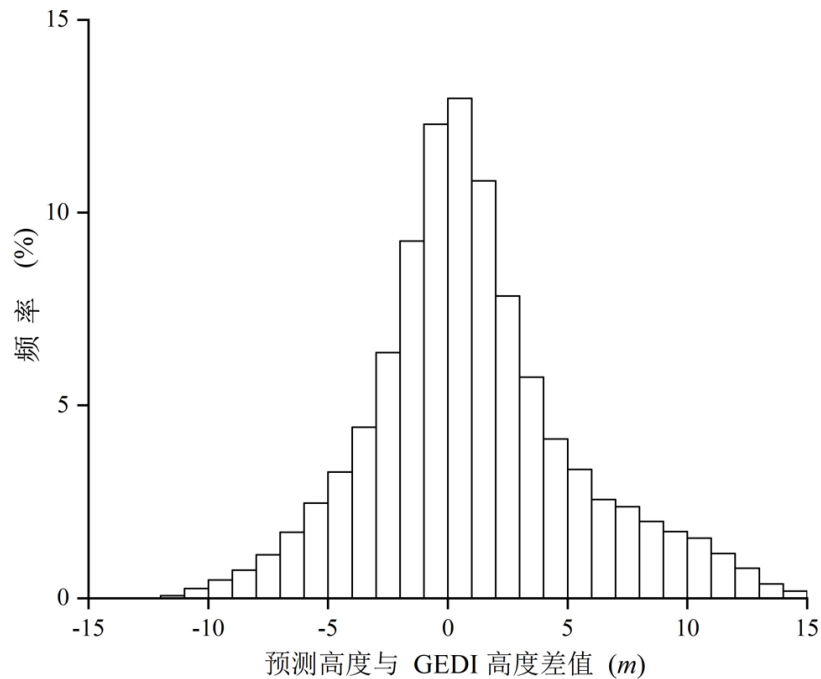


图 7 预测得到的冠层高度及 GEDI 提供的冠层高度差值的频率直方图

Fig. 7 Histogram of difference between predicted canopy height and GEDI-derived canopy height

由图 7 可知,二者之间的差值越小,计算得到的频率越高,但较大差异对应的频率显示出相反的趋势,即差异越大,对应的频率越高,且约 65.21% 的像元对应的差值位于 $[-3\text{ m}, 3\text{ m}]$ 之间,表明预测结果与 GEDI 提供的冠层高度之间的差值主要集中在该范围内,证明了二者之间较高的一致性。

3.4 不同森林类型最优变量集

本研究利用 RF 算法获得了不同估计模型对应的最优变量集,并计算了集合中每个变量的重要性得分,不同森林类型对应的最优变量集以及集合中各变量的重要性得分如图 8 所示。

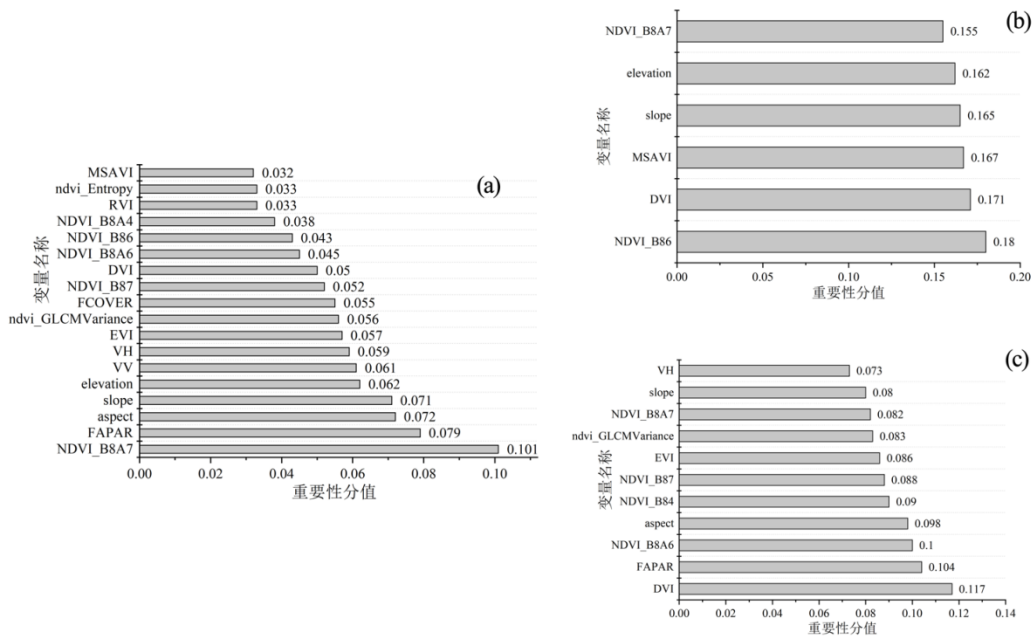


图 8 不同模型对应最优变量集及各变量重要性得分 (a, b 和 c 分别对应针叶林、阔叶林和混交林)

Fig. 8 Optimal variable set corresponding to different models and the importance scores of each variable (a, b and c correspond to coniferous/broadleaf/mixed forest, respectively)

由图 8 可知，不同模型对应的最优变量集包含的变量种类和个数均不相同，同一变量在不同集合中的重要性得分差异也较为明显。具体说来，针叶林、阔叶林和混交林对应的最优变量集中包含变量个数分别为 18、6 和 11，且仅有“NDVI_B8A7”和“Slope”两个变量出现在每个集合中。其次，同一变量的重要性得分在不同集合之间存在差异，对于“NDVI_B8A7”变量，针叶林、阔叶林和混交林的重要性得分分别为 0.101、0.155 和 0.082；而对于“slope”变量，对应的得分分别为 0.071、0.165 和 0.080。同时，本研究发现针叶林对应的最优集合内，每个变量重要性分值之间的差异是明显的，混交林的情况与针叶林非常相似；然而对于阔叶林，各变量的重要性的分布更加均匀，证明集合中每个变量在模型构建时做出的贡献几乎相同。本研究按照“生物物理特征”、“植被指数”、“纹理特征”、“VV&VH”和“地形信息”统计了不同集合中属于以上类别变量的重要性得分之和，结果如表 5 所示。

表 5 最优集合中不同类别重要性得分

Tab. 5 Importance scores belonging to different kinds of variable in the optimal set

	针叶林 (Coniferous forest)	阔叶林 (Broadleaf forest)	混交林 (Mixed forest)
生物物理指数	0.13	0	0.10
植被指数	0.45	0.67	0.57
纹理特征	0.09	0	0.08
VV&VH	0.12	0	0.07
地形信息	0.21	0.33	0.18
Σ	1.00	1.00	1.00

4 讨论

本研究首先利用机载 lidar 反演的冠层高度验证了 ICESat-2 的 RH95 高度指标的精度，并确定了两个冠层高度匹配的最佳空间分辨率；随后，基于随机森林方法，通过 RH95 高度指标与 Sentinel-1 的 SAR 数据、Sentinel-2 的光学影像和地形数据的协同，构建了针对不同森林类型的冠层高度估计模型；最后，通过验证样本和 GEDI 提供的冠层高度对所构建模型的精度进行了验证，证明了所提出的方法在较大区域内获取空间连续的森林冠层高度的潜力，获取的冠层高度将有助于森林动态监测和其他利用冠层高度进行的研究工作。

4.1 不同森林类型对应估计模型精度对比

本研究基于机器学习算法建立了整林、针叶林、阔叶林和混交林的森林冠层高度预测模型，与不分区森林类型直接建模相比，所建立的针叶林、阔叶林和混交林冠层高度预测模型的精度都有所提高，证明了本文提出的方法在森林冠层高度估计中的优势。以上结论与 Liu et al., (2019) 的研究的结果一致，证明了通过 Sentinel-1 和 Sentinel-2 的联合获取东北地区林分尺度森林高度时的优势。同时，第 3.2 和 3.3 节的结果还表明，针叶林对应的预测模型性能最好，预测结果精度高于阔叶林和混交林对应的预测结果的精度。上述情况可能归因于不同森林类型对应的因变量 (ATL08 产品的 RH95 指标) 的准确性间的差异，Alexander et al., (2018) 已经在之前的研究中证明了无论点云数据的密度如何，通过点云数据提取得到的阔叶林的冠层高度的精度均低于针叶林对应的精度，可能是由于阔叶林的叶面积比针叶林大，导致到达地面的激光雷达点更少，因此很难使用插值方法获得真实的 DEM，直接影响了所获取的冠层高度的精度。Mielcarek et al., (2018) 选择了七种树木，通过不同的激光雷达反演树木的冠层高度，并将预测的结果与现场测量的结果进行比较以验证了不同插值方法在预测森林冠层高度的准确性，结果表明利用不同的激光雷达进行森林冠层提取时，针叶林对应的结果的误差最小。因此，本研究推断针叶林覆盖区域的 ATL08 产品具有更高的精度，而针叶林对应的预测模型在估计森林冠层高度时具有最高的精度可能与针叶林覆盖区域的 RH95 精度较其他地区该指标的精度高有关。通过对本研究的结果和上述文献中获得的结论分析，再次证实了建立估计不同森林类型的冠层高度预测模型的必要性，而不仅是简单地将所有森林类型放在一起估计森林冠层高度，特别是对于森林类型较多或者面积较大的区域，区分森林类型分别建立估计模型尤为重要。

4.2 不同森林类型对应最优变量集及变量重要性对比

本研究最重要的目的之一是确定在构建冠层高度估计模型时不同森林类型对应最优变量集，并对集合中各变量在模型构建过程中发挥的作用进行定量地比较和分析。图 8 中的结果表明，最优变量主要包括植被指数、高程、坡向、坡度、VV 和 VH，且植被指数在森林冠层高度估计中的重要性已经得到证明。Pascual et al., (2008) 分析了利用激光雷达获得的森林冠层高度与 Landsat 影像之间的关系，证实了植被指数在森林冠层高度估计时的重要作用。同时，气温和降水等气象因子与高程、坡度和坡向密切相关，而气象因子又是影响植被空间分布与生长状况的主要因素之一，因此，森林的冠层高度与高程、坡度和坡向等地形因素

关系密切。VV 和 VH 在森林冠层高度估计中的重要作用也在本研究中得到了证实，本研究的结论与 Matasciet al., (2018) 的研究结论一致，他们的研究证明了 VH 与林分尺度森林平均高度的相关性较高，而对应 VV 相关性较低甚至相反。

由表 6 可知,在模型构建过程中本研究所选择的所有变量都包含在针叶林和混交林对应的最优集合中,但对应阔叶林的最优集合仅包含“植被指数”和“地形信息”两类变量。同时,本研究还发现在所有模型对应的最优变量集中,属于“植被指数”的变量对应的重要性得分分别为 0.45、0.67 和 0.57,证明了植被指数是构建森林冠层高度估计模型时最重要的变量类型。排在第二位的变量类别是“地形信息”,针叶林、阔叶林和混交林对应的最优变量集中属于地形信息变量的重要性得分之和分别为 0.21、0.33 和 0.18。因此,“植被指数”和“地形信息”包含变量重要性得分之和分别占针叶林、阔叶林和混交林最优变量集中所有变量得分之和的 66%、100%和 74%,表明两个类变量在估计森林冠层高度方面的重要作用。之前的研究也表明了森林冠层高度与植被指数之间存在很强的正相关关系,植被指数在森林冠层高度估计中的潜力也已得到证明(Pascual et al., 2008; Matasci et al., 2018)。同时,森林的类型、分布和生长与位置、高程、坡度等地形因素密切相关,因此,基于 SRTM 数据获取的地形变量在构建森林冠层高度估计模型时十分重要。本研究进一步计算了每个最优变量集中变量间的相关性,结果如图 9 所示。

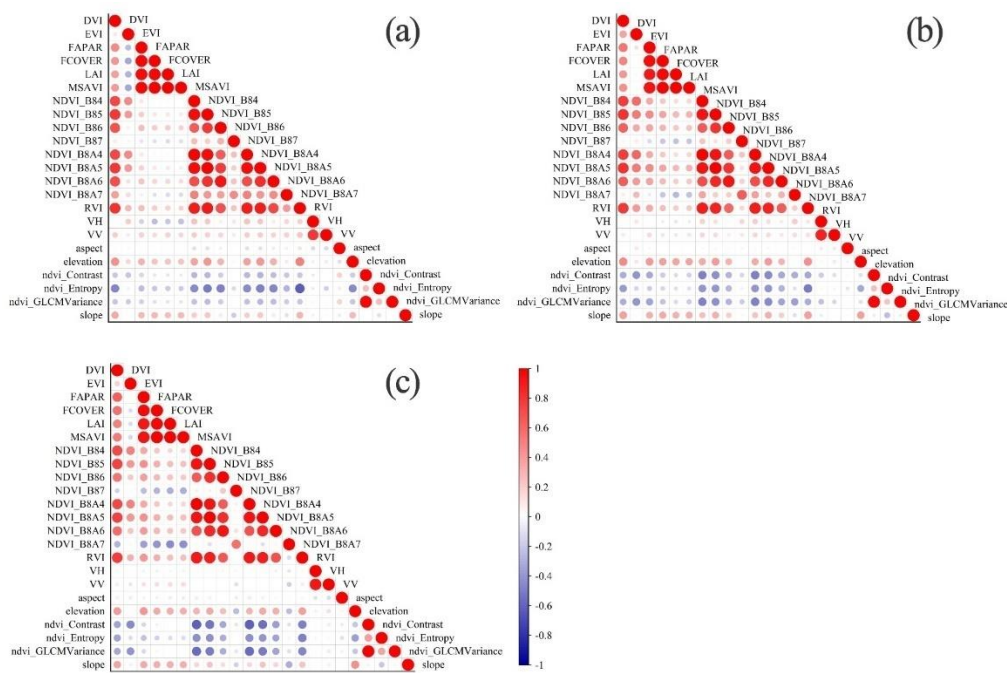


图 9 不同模型对应最优变量集各变量间相关性 (a、b、c 分别对应针叶林、阔叶林和混交林的结果)
Fig. 9 Correlations between different variables in each optimal set (a, b and c represent the result of coniferous/broadleaf/mixed forest, respectively)

由图 9 可知,最优变量集中各变量间存在显著的正相关关系,如针叶林对应模型中的“NDVI_B8_A4”变量与“MSAVI”变量、阔叶林对应模型中的“NDVI_B8_A5”变量与“NDVI_B84”变量以及混交林对应模型中的“MSAVI”变量与“LAI”变量。然而,在不同变量之间也存在负相关的关系,如属于纹理特征的变量与几乎所有其他变量均呈负相关的关系。值得注意的是,图 9 中的变量间相关性结果表明重要性较低的变量仍然存在,但它们在模型建立过程中的作用不容忽视。

本研究基于随机森林方法,通过多源数据的联合建立了顾及森林类型的冠层高度反演模型,验证了通过多源遥感数据的协同估计森林冠层高度的可行性。同时,星载遥感数据的使用使得获取大区域内空间连

续的森林冠层高度成为可能。本研究对 ICESat-2 提供的高度指标的精度进行了验证, 为将该数据用于森林冠层高度估计建模提供了依据, 极大地降低了野外实测数据采集时耗费的时间。然而, 卫星数据的获取很容易受到天气的影响, 将在一定程度上影响模型的构建同时, 由于受到样本数据覆盖范围的限制, 本研究构建的冠层高度模型的准确性需要在更大区域内进行验证。上述情况可能会影响模型的应用, 未来应探索使用其他数据或基于新的机器学习方法构建精度可靠、实用性强的森林冠层高度模型的可能性。本研究构建的估计模型验证了通过多源数据的协同获取空间连续的森林冠层高度的可能, 同时, 本研究的结论对模型构建时变量的选择具有重要的参考意义。

参考文献

- [1]Ahmed O S, Franklin S E, Wulder M A, et al. 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101: 89-101.
- [2]Alexander C, Korstjens A H, Hill R A, et al. 2018. Influence of micro-topographic and crown characteristics on tree height estimations in tropical forests based on LiDAR canopy height models. *International journal of applied earth observation and geoinformation*, 65: 105-113.
- [3]Belgiu M, Drăgut L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114: 24-31.
- [4]Breiman L. 2001. Random forests. *Machine learning*, 45: 5-32.
- [5]Dong J, Kaufmann R K, Myneni R B, et al. 2003. Remote sensing estimates of boreal and temperate forest woody biomass: Carbon pools, sources, and sinks. *Remote sensing of Environment*, 84(3): 393-410.
- [6]Dong J W, Xiao X M, Sheldon S, et al. 2012. A comparison of forest cover maps in Mainland Southeast Asia from multiple sources: PALSAR, MERIS, MODIS and FRA. *Remote Sensing of Environment*, 127: 60-73.
- [7]Ghosh S M, Behera M D, Paramanik S. 2020. Canopy Height Estimation Using Sentinel Series Images through Machine Learning Models in a Mangrove Forest. *Remote Sensing*, 12(9): 1519.
- Heilongjiang Meteorological Bureau. Available online: <http://hl.cma.gov.cn/>.
- [8]Lang N, Schindler K, Wegner J D. 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sensing of Environment*, 233: 111347.
- [9]Lary D J, Alavi A H, Gandomi A H, et al. 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1): 3-10.
- [10]Lefsky M A, Cohen W B, Harding D J, et al. 2022. Lidar remote sensing of above-ground biomass in three biomes. *Global ecology and biogeography*, 11(5): 393-399.
- [11]Li W, Niu Z, Liang X, et al. 2015. Geostatistical modeling using LiDAR-derived prior knowledge with SPOT-6 data to estimate temperate forest canopy cover and above-ground biomass via stratified random sampling. *International Journal of Applied Earth Observation and Geoinformation*, 41: 88-98.
- [12]Li W, Niu Z, Shang R, et al. 2020. High-resolution mapping of forest height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation*, 92:102163.
- [13]Lin X, Xu M, Cao C, et al. 2020. Estimates of Forest height Using a Combination of ICESat-2/ATLAS Data and Stereo-Photogrammetry. *Remote Sensing*, 12(21): 3649.
- [14]Liu T, Abd-Elrahman A, Morton J, et al. 2018. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience & remote sensing*, 55(2): 243-264.
- [15]Magruder L A, Brunt K M. 2018. Performance analysis of airborne photon-counting lidar data in preparation for the ICESat-2

- mission. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5): 2911-2918.
- [16]Mandianpari M, Salehi B, Mohammadimanesh F, et al. 2017. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130: 13-31.
- [17]Markus T, Neumann T, Martino A, et al. 2017. The ice, cloud, and land elevation satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote sensing of environment*, 190: 260-273.
- [18]Matasci G, Hermosilla T, Wulder MA, et al. 2018. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote sensing of environment*, 209: 90-106.
- [19]Maxwell A E, Warner T A, Fang F. 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International journal of remote sensing*, 39(9): 2784-2817.
- [20]Mielcarek M, Stereńczak K, Khosravipour A. 2018. Testing and evaluating different LiDAR-derived canopy height model generation methods for tree height estimation. *International Journal of Applied Earth Observation and Geoinformation*, 71: 132-143.
- [21]Nandy S, Srinet R, Padalia H. 2021. Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. *Geophysical Research Letters*, 48(14): e2021GL093799.
- [22]Neuenschwander A L, Magruder L A. 2019. Canopy and terrain height retrievals with ICESat-2: A first look. *Remote sensing*, 11(14): 1721.
- [23]Neuenschwander A, Pitts K. 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote sensing of environment*, 221: 247-259.
- [24]Neumann T, Martino A, Markus T, et al. 2019. The Ice, Cloud, and Land Elevation Satellite-2 Mission: A global geolocated photon product derived from the advanced topographic laser altimeter system. *Remote sensing of environment*, 233: 111325.
- [25]Pascual C, Cohen W, Garcíaabril A, et al. 2008. Mean height and variability of height derived from lidar data and Landsat images relationship. *Silvilaser*, 2008: 8th.
- [26]Rosenqvist A, Milne A, Lucas R, et al. 2003. A review of remote sensing technology in support of the Kyoto protocol. *Environmental Science & Policy*, 6(5): 441-455.
- [27]Shao Z, Yang K, Zhou W. 2018. Performance Evaluation of Single-Label and Multi-Label Remote Sensing Image Retrieval Using a Dense Labeling Dataset. *Remote Sensing*, 10(6): 964.
- [28]Simard M, Pinto N, Fisher J B, et al. 2011. Mapping forest height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, 116(G4).
- [29]Tian S, Zhang X, Tian J, et al. 2016. Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sensing*, 8(11): 954.
- [30]Tian X, Li Z, Su Z, et al. 2014. Estimating montane forest above-ground biomass in the upper reaches of the Heihe River Basin using Landsat-TM data. *International journal of remote sensing*, 35(21): 7339-7362.
- [31]Wu C, Venevsky S, Sitch S, et al. 2017. Present-day and future contribution of climate and fires to vegetation composition in the boreal forest of China. *Ecosphere*, 8(8): e01917.
- [32]Wulder MA, White J C, Nelson R F, et al. 2012. Lidar sampling for large-area forest characterization: A review. *Remote sensing of environment*, 121: 196-209.
- [33]Xing Y, Huang J, Gruen A, et al. 2020. Assessing the Performance of ICESat-2/ATLAS Multi-Channel Photon Data for Estimating Ground Topographic in Forested Terrain. *Remote Sensing*, 12(13): 2084.
- [34]Yang R, Dong C, Zhang Y. 2017. Method of population spatialization under the support of geographic national conditions data. *Sci. Surv. Mapp*, 42:76-81.
- [35]吴英, 张万幸, 张丽琼, 等. 2012. 基于 DEM 的地形与植被分布关联分析. *东北林业大学学报*, 4(2): 22-40.
- [36]Wu Y, Zhang W X, Zhang L Q, et al. 2012. Analysis of correlation between terrain and forest spatial distribution based on DEM. *Journal of Northeast Forestry University*, 40(11): 96-98. [in Chinese]
- [37]Zhang L, Du B. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and remote sensing magazine*, 4(2): 22-40.

[38]Zhang J, Liu J, Zhai L, et al. 2016. Implementation of Geographical Conditions Monitoring in Beijing-Tianjin-Hebei, China. ISPRS International Journal of Geo-Information, 5(6): 89.

[39]Zhao Y, Cai Y. 2016. The application of the national geographic census results in quality inspection of basic surveying and mapping. Geomat. Spat. Inform. Technol, 39: 139-145.