doi: 10.6046/zrzyyg.2022173

引用格式:席磊,舒清态,孙杨,等.基于 ICESat2 的西南山地森林 LAI 遥感估测模型优化[J].自然资源遥感,2023,35(3):160-169. (Xi L,Shu Q T,Sun Y, et al. Optimizing an ICESat2 – based remote sensing estimation model for the leaf area index of mountain forests in southwestern China[J]. Remote Sensing for Natural Resources,2023,35(3):160-169.)

# 基于 ICESat2 的西南山地森林 LAI 遥感估测模型优化

席 磊, 舒清态, 孙 杨, 黄金君, 宋涵玥

(西南林业大学林学院,昆明 650224)

摘要:叶面积指数(leaf area index,LAI)是森林生态系统重要参数,如何以较小成本提升区域尺度的山地森林LAI的遥感估测精度,对于精确掌握森林LAI的情况和进一步了解森林生态系统有重要意义。本研究以星载激光雷达ICESAT - 2/ATLAS为主要信息源,以西南山地香格里拉市为研究区,基于随机森林回归(random forest,RF)遥感估测模型,结合地面51块LAI实测样地数据,在前期进行RF超参数优化基础上,采用决定系数、均方根误差、绝对平均误差和中位数绝对误差作为模型精度评价指标,对估测效果进行分析。结果表明:使用随机表面查找算法进行RF回归模型的超参数优化,能明显提升模型估测LAI精度。提取出的地面光斑特征参数在山地森林LAI估测中有较高的贡献度和极佳的效果,可用于区域尺度的山地森林物理结构参数LAI的估测。同时,利用随机表面查找算法优化后的RF回归模型,估测精度更高,估测结果与研究区森林分布现状吻合,具有一定普适性。最后,研究确定了使用ICESat - 2/ATLAS数据产品估测LAI是可行的,能为星载激光雷达估测中大范围的LAI提供一定的参考。

关键词: ICESat - 2/ATLAS; 叶面积指数; 超参数优化; 随机森林; 香格里拉 中图法分类号: TP 79; S 771.8 文献标志码: A 文章编号: 2097 - 034X(2023)03 - 0160 - 10

0 引言

叶面积指数(leaf area index,LAI)是森林生态 系统中生物地球化学循环的重要中间参数,如光合 作用、呼吸、蒸腾和碳循环的关键参数<sup>[1]</sup>。传统的 LAI通常采用地面直接或间接测量技术,只能获取 小尺度的森林 LAI,如全植物体采样和光学测定的 方法<sup>[2]</sup>。与传统地面调查数据相比,遥感技术在获 取区域尺度上森林资源信息时具有无可比拟的优 势,越来越多的遥感数据被用于森林资源监测<sup>[3]</sup>。 传统光学遥感只能获取森林冠层定性信息,存在很 多不确定性,降低了在区域尺度下的森林 LAI 估测 精度。已有研究表明,当 LAI 值超过 4.0 时,光学遥 感易发生饱和现象,导致 LAI 估计值比实际值低<sup>[2]</sup>。 激光雷达(light detection and ranging,LiDAR)是当前 发展迅速的主动遥感技术之一,已被证实在森林高 度和垂直结构的探测上具有很强的应用能力。但目前而言,机载激光雷达只能获取小尺度极高的反演 精度,但数据量大,费用昂贵<sup>[4]</sup>;星载激光雷达可以 提供覆盖全球的激光光斑数据,通过协同光学数据 或空间插值方法,可获取区域尺度高精度 LAI<sup>[5]</sup>。

用于区域尺度的森林 LAI 监测的典型星载 Li-DAR 主要有美国 NASA 发射的 ICESat – 1/GLAS 和 ICESat – 2/ATLAS。ICESat – 1 搭载地球科学激光 高度计系统(GLAS)收集的数据已成功用于检索森 林 LAI<sup>[6-7]</sup>。然而,GLAS 在进行区域尺度 LAI 反演 时,由于光斑为椭圆形,不利于地面 LAI 样地调查; 其次,GLAS 采样间隔约为 170 m,空间分布较稀 疏<sup>[8-10]</sup>,光斑直径约 70 m,回波信号容易受地形影 响<sup>[11]</sup>,降低了 LAI 估测精度。ICESat – 1/GLAS 于 2009 年 10 月 11 日退役后,美国国家航空航天局又 于 2018 年 9 月成功发射了 ICESat – 2/ATLAS。与 GLAS 相比,ATLAS 使用多光束的光子计数激光雷

收稿日期: 2022-04-28;修订日期: 2022-09-19

基金项目:国家自然科学基金项目"生态脆弱区典型森林生态系统生化参数高光谱遥感反演关键技术研究"(编号:31860205)、"基于LiDAR和 MERSI数据滇西北乔木生物量反演关键技术研究"(编号:31460194)和云南省教育厅科学研究基金项目"基于深度学习的多源遥感协同的森林生物量估测研究"(编号:2021Y249)共同资助。
 第一作者:席 磊(1997-),男,硕士研究生,研究方向为数字林业与森林资源管理。Email:swfuxilei@163.com。

通信作者: 舒清态(1970-),男,副教授,硕士生导师,研究方向为林业3S技术应用。Email: shuqt@163.com。

• 161 •

达系统代替 GLAS 使用的全波形激光雷达系统,该 系统采用低能量消耗延长激光寿命和高脉冲频率, 以获得比 GLAS 更密集的采样间隔和更小的光斑足 印,密集的足印光斑能降低山区地形对激光雷达数 据的影响<sup>[12-13]</sup>。同时由于 ATLAS 光斑小,一定程 度上降低了地形的影响。Neuenschwander 等<sup>[14]</sup> 对 ICESat-2 垂直采样误差有关的性能特征和潜在的 不确定性进行了探讨,其中包括在植被茂密的生境 内感知高度值的误差和测量精度,结果显示地形误 差为1.93 m 和2.52 m、冠层误差从0.28~1.25 m 之 间不等。Narine 等<sup>[15]</sup>在 2019 年使用 ATL08 数据产 品对芬兰的一个植被覆盖区域的 ICESat-2 断面进 行了水平和垂直方向的精度检验,得到了水平方向 偏差在5m内,同一横断面下地形和冠层高度的垂 直均方根误差 RMSE 分别为 0.85 m 和 3.2 m 的结 论。在区域尺度上,Zhang 等<sup>[16]</sup>使用 ATLAS 数据进 行了 LAI 的检索,基于 GORT 理论估计 ICESat -2 得 到亚马孙 LAI: R = 0.693, RMSE = 2.545, 并使用 MODIS 和 Sentinel -2 生成的 LAI 进行检验,但其检 验方法及精度有待进一步提升。综上,以地面实测 ATLAS 光斑 LAI 为样本数据,进行区域尺度 ATLAS 光斑足印 LAI 估测,并提升算法估测精度的研究鲜 有报道。

因此,针对新一代 ICESat - 2/ATLAS 光斑数据 特点,以西南典型山地香格里拉市为研究区,基于随 机森林超参数优化算法,以较小的地面调查样本, 实现区域尺度上 LAI 遥感估测。主要目标如下: ①基于随机森林进行特征优选,提取经过光子去 噪、光子分类算法处理后的 ATLAS 光斑参数建模 指标; ②基于超参数优化后的随机森林算法,提升 ICESat - 2/ATLAS 光斑足印内 LAI 估测精度,并进 行分析。

1 研究区概况及数据源

#### 1.1 研究区概况

以云南省西北部具有山地地形的香格里拉市 为研究区,经纬度在 E99°08′02″~100°21′15″, N26°49′07″~28°54′38″之间,如图 1 所示。研究 区属亚热带常绿阔叶林植被区向高寒植被区过渡 地带。气候易受海拔影响,昼夜温差大,海拔区间 为2 993~3 947 m<sup>[17]</sup>。按《云南植被》划分标准, 全域共有 10 种植被类型,主要类型有温凉性针叶 林、寒温性针叶林、灌丛和草甸等,主要树种有云冷 杉、高山松和高山栎等。



图 1 研究区位置及样地示意图 Fig. 1 Location and plot of the study area

# 1.2 ICESat -2 卫星数据

ICESat - 2 卫星轨道为 3 对 6 束,激光频率为 10 kHz,可得到直径约 17 m、沿轨间距0.7 m 的多波 束光子点云数据。轨道垂直间距约为 3 km,强弱光 束垂直间距约为 90 m,卫星重返周期为 91 d<sup>[18-19]</sup>。 该卫星共有 21 种数据产品,分为 3 大类,产品命名 为 ATL01 ~ ATL21。研究使用数据产品为: ATL03 数据产品(global geolocated photon data)和 ATL08 数 据产品(land and vegetation height)<sup>[20]</sup>。

ATL03 产品中包括所有光子事件的时间、经纬 度和高度等地理空间位置信息。其中各光子均以索 引的方法按层次性结构串联起来,ATLAS 中索引类 型主要有2种:一是依据光子传输时间进行编号, 二是沿轨按距离分段。即按20m划分,每个区段用 唯一的7位数编号作为区段号[21-23],区段号被存储 在 segment\_id 中。ATL08 产品是进行分区段后的 ATL03 产品<sup>[24]</sup>,主要是使用差分、回归高斯自适应 最近邻(differential, regressive, and Gaussian adaptive nearest neighbor, DRAGANN) 算法<sup>[19, 23, 25]</sup> 完成 噪声去除,再进行光子点云分类,最后分为噪声光 子、地面光子、冠层光子和冠层顶光子4大类。分类 标识以区段的形式记录在 classed\_pc\_flag 中。研究 选取香格里拉市域内 2020 年 1 月—2021 年 6 月之 间的所有 ATL03 和 ATL08 数据产品,2 类数据产品 均为118条数据、354条轨道、708条光子轨道波束。 研究数据均可在 ICESat - 2 官网下载 (https:// nside. org/data/icesat -2)

# 2 研究方法

2.1 基于光斑足印的样地设计

样地设计尽量覆盖研究区内主要植被类型,采

集样地 54 块,去除差异性较大样地 3 块,最终有效 样地为 51 块。调查样地是与 ATLAS 传感器发射的 光斑大小一致的样圆,样圆直径 17 m。样圆中心坐 标为 ATLAS 数据产品中光斑中心点的坐标,为保证 坐标复位一致,调查使用千寻星矩 SR3(Pro 版)差 分 RTK 进行点位放样,采集时保证设备为固定解状 态,5 次连续采集后取均值,所有样地中心点坐标与 光斑中心点坐标仪器误差均小于 0.02 m。

#### 2.2 样圆 LAI 测定设计

研究选择使用基于 LAI - 2000 的平台开发 LAI - 2200 测定仪来测定样圆内有效 LAI 值,该仪器主要利用鱼眼光学传感器。LAI - 2200 传感器利用 5 个不同的天顶角方向来检测植物冠层下的光照强度变化,测量时最理想的是云层均匀分布的阴天。测定模式为"ABBBB"模式,即测量一个天空空白,对比

在林下测量4个值。在每个样圆的东西南北测量4 个 B 值和中心点1个 A 值,取最后的均值来代表该 区域的有效 LAI<sup>[26-27]</sup>。最终计算得到的51 块样地 的 LAI 统计信息汇总如表1 所示,其中坡度[0, 10]°样地22 块,(10,20]°样地15 块,大于20°的样 地14 块。

表 1 样地 LAI 统计信息汇总 Tab. 1 Summary of LAI statistics for sample sites

样地 数量	均值	均值标 准差	标准差	最大值	最小值	中位数
51	0.468 1	0.038 8	0.265 8	0.967 0	$0.012\ 1$	0.5090

#### 2.3 研究路线

使用 ICESat - 2/ATLAS 估测光斑足印内 LAI 主要分为3个部分,即光子点云去噪算法、光子分类 算法和 LAI 估测模型。具体步骤如图2所示。



## Fig. 2 Study flowcharts

# 2.4 光子点云去噪算法

光子计数雷达比其他激光雷达对光子信号感知 更敏感,ATLAS 在获取地面和树冠等目标的反射光 子时,还接收到了太阳背景或大气散射引起的噪声 光子<sup>[4, 28]</sup>。因此,应该先进行噪声去除。以往的研 究基于信号光子的光子密度更大、噪声光子分布具 有随机性的特点进行了假设,开发了多种去噪算法。 当前去噪算法大致有 3 种<sup>[29-30]</sup>:基于图像处理、基 于局部统计和基于密度聚类的算法。分析 3 种算法 的优缺点后,本文使用基于密度聚类算法(different densities – based spatial clustering of applications with noise, DDBSCAN)<sup>[31]</sup>和基于局部统计算法(K – nearest neighbors – based, KNNB)<sup>[32-33]</sup>分别去除高背景 噪声的噪声光子和低背景噪声水平中信号光子周围 的少量噪声光子。该综合去噪算法可获得极佳的去 噪效果。同时,在 DDBSCAN 算法中,还计算了所有 搜索方向的光子密度,并使用最大密度差替代密度 作为最终度量参数,以此降低光子密度不一致对算 法性能产生的影响。算法流程如图 3 所示。



图 3 噪声消除算法流程

# Fig. 3 Flow chart of noise elimination algorithm

### 2.4.1 DDBSCAN 算法

综合去噪算法首先是 DDBSCAN 算法。主要 步骤为:先计算每个光子所有搜索方向的光子密 度,定义为每个光子椭圆邻域中的光子数量;设 置搜索椭圆的方向间隔为 5<sup>[34-35]</sup>;椭圆的长轴 和短轴由沿轨距离范围与高程范围的比率进行 设定<sup>[29]</sup>。其次,得到每个光子的最大密度差。 最后,根据密度差频率直方图确定阈值(阈值确 定详见文献[34])。如果密度差小于阈值,该光 子视为噪声去除。

#### 2.4.2 KNNB 噪声滤波算法

综合去噪算法第二步是 KNNB 算法,主要步骤 为:首先计算每个光子到其 k 近邻的总距离。k 代 表距离每个点最近的点数,它是计算总距离的重要 参数。该步骤中未选择原始算法中的50,而是降低 了 10 倍,使用 k = 5 的原因是 DDBSCAN 算法进行 了第一重去噪,得到了背景噪声较小的数据,对于背 景噪声水平较低的数据,k = 50 并不适用。本研究 以 5 为间隔,在1~50 之间进行多次试验分析,确定 最佳值 k = 5,结果与 Zhang 等<sup>[16]</sup>的研究保持一致。 最后根据文献[36]去除总距离大于阈值的噪声光 子。对于残留的少量孤立噪声光子,则使用目视解 译去除。

# 2.5 光子分类算法

进行 LAI 估测前,必须将去噪滤波后的信号光 子分类为地面和冠层光子。以往研究确定了几种分 类算法,Axelsson<sup>[37]</sup>提出的渐进式三角不规则网络加密(PTD)方法,但该方法不适合地形复杂区域。 为了提高地面光子提取精度,本研究使用了 Nie 等<sup>[36]</sup>提出的修正 PTD 方法,该方法对于海拔落差 较大的区域内的地面光子识别精度较高。算法流程 如图 4。



Fig. 4 Flow chart of photon classification algorithm

根据 Zhang 等<sup>[16]</sup>、Nie 等<sup>[29]</sup>和 Zhang 等<sup>[38]</sup>的研究结论,研究对原始算法中的部分参数进行了改正, 通过选择距离初始三角网最远的点下方高程最低的 点,来保证地面点不被误分类为冠层点。算法关键 步骤如下:

1)参数简述<sup>[29, 34, 36, 39]</sup>。主要包含4个参数,窗 口大小(由最大非圆形特征大小确定)预设为200 m; *D*<sub>i</sub>为未分类光子到2个初始相邻种子地面光子之间 的距离;*A*<sub>i</sub>为连接未分类光子及其种子地面光子的 线与地面区段线之间的角度;*D*<sub>s</sub>为从未分类光子到 地面曲面的距离。

2)初始种子地面光子选择和区段线生成。选择每个窗口中高程最低的光子作为初始种子地面光子,然后连接初始种子地面光子以生成初始区段线。 剩余光子均标记为未分类光子。

3)地面光子的迭代致密化。对地面曲线进行2次迭代加密,提高地面曲线精度,便于地面和植被光子的分离。具体流程为,计算每个未分类光子的距

离 *D*<sub>i</sub>后,若 *D*<sub>i</sub>的最大值大于阈值,则从未分类光子 中提取高程低于 *D*<sub>i</sub>最大值的光子作为地面光子,以 保证提取的光子中无冠层光子。同时,为确定最佳 地面区段线,设置以10 为间隔,测试10~50 之间的 数值,阈值为20,多次迭代调用自身,直到无新的种 子地面光子为止。最终结果与 Zhang 等<sup>[16]</sup>的研究 保持一致。生成初始区段线后,采用 Douglas – Peucker 算法生成改进的区段线。

4)地面和植被光子的区分。为了提取最终地 面光子,使用三次样条插值拟合地面光子并建立地 面曲面。仅有 *D*,低于 1 m 的光子被识别为地面光 子,其余均为植被光子<sup>[16, 34, 39]</sup>。

5)最终,研究区 ATL03 的有效光子数量达到了 千万级以上,根据 ATL08 进行 100 m 的抽稀,得到 研究区内有效光斑数量为 94 086 个,再与最新的林 地范围叠加分析,得到有效林地光斑数量为 74 873 个,非林地有效光斑数量为 19 213 个,具体光斑分 布如图 5 所示。





# 2.6 随机森林算法超参数优化估测 LAI

随机森林通过自助采样法,将提供的样本数据 集经过随机采样后,再基于每个采样集进行训练构 建决策树;在结点处,先从该结点的属性集合中随 机选择一个包含多个属性的子集,然后再从这个子 集中选择最优的一个属性进行划分<sup>[38,40]</sup>。研究使 用 Python 语言在 Pycharm 中调用 scikit - learn 库<sup>[41]</sup> 实现超参数确定,基于51 块样地数据使用随机表面 查找算法抽取一定数量的参数组合,进行多折交叉 验证,以确定模型最优超参数。研究进行超参数优 化的参数有(n\_estimators,min\_samples\_split,min\_ samples\_leaf,max\_features,max\_depth,bootstrap)。 各参数含义如表2所示。

表 2	随机森林算法参数说明
Tab. 2	Description of the parameters
of th	e random forest algorithm

	e	
参数名	描述	类型
n_estimators	决策树的数量	整数型
min_samples_split	节点可分的最小样本数	整数或浮点型
min_samples_leaf	叶子节点含有的最少样 本	整数或浮点型
max_features	构建决策树最优模型时 考虑的最大特征数	整数或浮点型
max_depth	决策树最大深度	整数型
bootstrap	样本集是否放回抽样	布尔型

# 2.7 精度评价标准

为探究随机森林算法超参数优化后估测 LAI 能力,以及拟合估测模型的精度,研究结果分别对随机森林默认参数、超参数优化后模型的决定系数 R<sup>2</sup>、绝对平均误差 MAE、均方误差 MSE、均方根误差

RMSE 和中位数绝对误差 MedAE 进行定量评价<sup>[21,42]</sup>。决定系数指模型拟合优度,是表示回归 直线对观测值的拟合程度,数值为[0,1]之间,数 值越接近于1,代表模型拟合效果越好。绝对平均 误差是所有单个观测值与算术平均值的偏差的绝 对值的平均,数值越小,误差越小。均方误差是预 测数据和原始数据对应点误差的平方和的均值, 数值越小,误差越小。均方根误差也叫回归系统 的拟合标准差,是均方误差的平方根,数值越小, 误差越小。中位数绝对误差非常适合含有离群点 的数据集,属于评价模型好坏的指标之一,数值越 小、误差越小,则模型越好<sup>[41]</sup>。各项精度评价指标 公式分别为:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \widehat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y})^{2}}, \qquad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n-1}} , \qquad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |(y_i - \hat{y}_i)| , \qquad (3)$$

 $MedAE(y,\hat{y}) = median(|y_1 - \hat{y_1}|, \dots, |y_n - \hat{y_n}|),$ (4)

式中: $y_i$ 为实际值; $\hat{y}_i$ 为估计值; $\bar{y}$ 为估计值均值; n为样本量。 3 结果与分析

# 3.1 随机森林算法中默认参数估测 LAI

对 51 块现地测量 LAI 的有效样地进行随机森 林模型拟合,构建默认参数设置的随机森林模型。 同时进行特征参数重要性评价<sup>[41]</sup>,参与拟合的参数 共计 48 个,所有参数均有一定贡献率,48 个参数参 与模型构建的贡献率百分比如图 6 所示。总体贡献 度阈值绘制区间为(0,20],其中最高贡献率的参数 为 photon\_rate\_can(19.38%),最低贡献率的参数为 h\_min\_canopy\_abs(0.18%),中位数贡献率为 1.41%,建模中贡献率排前十的参数及含义见表 3。



Fig. 6 Proportion of importance contribution of modelling parameters

表 3 未优化随机森林模型建模参数贡献率统计 Tab.3 Statistics on the contribution of modeling parameters of the unoptimized random forest model

参数名	描述	数值/%
photon_rate_can	计算后每 100 m 段内冠层光 子的光子率	19.38
n_toc_photons	区段内冠层顶部光子数	7.81
n_ca_photons	区段内冠层光子数	5.66
asr	表观反射率	4.20
h_median_canopy	区段内个体相对冠层高的中 位数	3.62
solar_azimuth	太阳方位角	3.34
solar_elevation	太阳高度角	2.92
toc_roughness	区段内冠层顶光子相对高度 的标准偏差	2.73
h_min_canopy	区段内冠层高度的最小值	2.56
dem_h	地理定位点处的最佳可用 DEM值	2.55

通过对 ICESat – 2/ATLAS 提取的所有参数进行随机森林建模,真实值和预测值情况如图 7 所示, 各项评价指标分别为:  $R^2$  = 0.872 7, *MAE* = 0.075 5, *MSE* = 0.008 8, *RMSE* = 0.093 8, *MedAE* = 0.061 4, 结果表明, 提取的参数在使用随机森林算法建模估计 LAI 中呈现很好的效果。



Fig. 7 Point line diagram of the random forest model fit

#### 3.2 随机森林超参数优化后估测 LAI

针对使用随机森林算法默认参数拟合的估测模 型,研究又对精度的进一步提升做出了思考。最终 选择使用随机表面查找算法对随机森林算法进行超 参数优化<sup>[40-41, 43]</sup>。抽取 50 000 组参数组合,进行 10 折交叉验证,最终确定模型最优超参数为('n\_estimators': 1 340, 'min\_samples\_split': 2, 'min\_samples\_leaf': 2, 'max\_features': 'auto', 'max\_depth': None, 'bootstrap': TRUE)。使用得到的超参数再次 对 51 块现地测量 LAI 进行随机森林模型拟合,构建 基于超参数优化后的随机森林模型,同时进行特征 参数重要性评价,参与模型拟合的参数共计48个, 去除贡献率为0的参数 h\_te\_mode 和 h\_dif\_ref,剩 余46个参数的贡献率如图8所示。总体贡献度绘 制区间为(0,35],其中最高贡献率的参数为 photon\_ rate\_can(32.22%),最低的参数为 h\_median\_canopy (0.01%),中位数贡献率为0.23%。优化后建模中 贡献率排前十的参数及含义见表4所示。



图 8 优化后建模参数重要性贡献比例 Fig. 8 Proportion of importance contribution of modelling parameters after optimization



		• 1		
参数	名		描述	数值/%
photon_rate_can			计算后每 100 m 段内冠层光 子的光子率	32.22
toc_roughness			区段内冠层顶光子相对高度 的标准偏差	13.39
asr			表观反射率	6.41
snr			定位光子的信噪比	6.19
h_te_median			在 WGS84 椭球体上方的光子 高度的中值(分类为地形区段 内)	4.86
h_te_max			在 WGS84 椭球体上方的光子 高度的最大值(分类为地形区 段内)	4.73
h_te_min			在 WGS84 椭球体上方的光子 高度的最小值(分类为地形区 段内)	4.61
canopy_openness		s	区段内冠层光子与段内所有 光子的标准差(可推断冠层开 放度)	4.41
h_te_rh25			25%分位数高度处的地形高度值	4.25
segment_landcover		ver	IGBP 地表覆盖类型	4.17
		1.0 0.8	$R^2 = 0.872\ 7$ $MSE = 0.008\ 8$ $MAE = 0.075\ 5$ $RMSE = 0.093\ 8$	
	真实值	0.6		
	LAI	0.4		

使用超参数优化后的随机森林模型对 ICESat – 2/ATLAS 提取的所有参数进行建模,对比真实值和预测值的效果如图 9 所示,各项评价指标分别为:  $R^2 = 0.907 1$ , MAE = 0.056 6, MSE = 0.006 4, RMSE = 0.080 1, MedAE = 0.042 0。对比未优化的模型可得出: 超参数优化后的随机森林模型估测 LAI 优于未优化的模型。总体呈现更高的  $R^2$ 和更低的 RMSE, 详见优化前后散点图(图 10)。







图 10 随机森林模型拟合散点图 Fig. 10 Scatterplot of random forest model fit

1.0

# 3.3 超参数优化后拟合 LAI 制图

0.2

04

(a) 优化前

0.6

LAI预测值

0.8

0.2

使用超参数优化后的模型对研究区内 74 873 个林地有效光斑进行估测,得到了所有光斑对应的 LAI 预测值,进行区域制图后,得到了研究区内所有 光斑 LAI 的空间分布如图 11 所示。进行空间制图 分级后,研究区 LAI 最大值为 0.954,最小值为 0.016,均值为 0.525。且从图 11 中可以看出整个 研究区内植被覆盖总体较高,LAI 较低区域主要分 布在研究区边缘,多为河流或常年积雪区域周边; LAI 较高区域总体呈现西北东南贯穿的趋势,主要 因为中部地区人工林种植比例逐年提高,同时东北 地区为普达措国家森林公园分布区域<sup>[44]</sup>,这也从侧 面印证了 LAI 预测结果具有一定可靠性。



图 11 研究区内 ICESat -2 光斑 LAI 空间分布 Fig. 11 Spatial distribution of ICESat -2 spot LAI in the study area

4 讨论

### 4.1 超参数优化对模型精度和参数筛选的作用

研究使用的超参数优化可以提高模型构建时各 个参数的贡献率,结合图 6 和图 8 综合分析超参数 优化前后的效果可以看出,优化后的模型中参数贡 献率最大值可以从 19.38% 提高到 32.22%,同时, 未优化模型中所有参数均有贡献率;优化后去除贡 献率为 0 的参数 2 个。对去除低贡献率的参数有较 为明显的作用。再结合图 10 可以看出,进行参数优 化后的模型得出的预测值更接近于真实值,误差更 小,预测结果更稳定。

# 4.2 ATLAS 数据估测精度分析及优化建议

本研究所有样地均使用差分 RTK 设备测定,且 保证误差在 0.02 m 内,能保证坐标采集精度较高。 同时,样地洗择与光斑足印重叠的样圆(半径为8.5 m, 面积为176.625 m<sup>2</sup>),更能保证测定的 LAI 值与该 光斑足印的真实值误差较小。虽然现场 LAI 的测定 使用了精度极高的 LAI - 2200 设备,但香格里拉市 位于高海拔落差的区域<sup>[17,44]</sup>,林中测定作业较为不 便,且数据处理也较繁杂,如考虑时间成本,可替换 简单鱼眼镜头采集,但精度可能会受到影响。同时, 研究使用随机表面查找算法是考虑了时间效率的决 定,并不能算作该模型的最优参数集合,只能算作 50 000 组参数组合中的最优参数。如果后期不考虑 时间成本,可使用 scikit - learn 库中的全面查找算 法<sup>[41]</sup>对模型进行全参数优化。除此之外,还可以引 入深度学习[43,45-47]等其他算法,进一步提高拟合 精度。

# 4.3 基于 ATLAS 数据估测 LAI 模型可移植性展望

本文建立了一套数据预处理、参数集成化提取、 超参数自动适应性查找、最佳模型拟合的流程化系 统。只需要输入测定的真实值数据进行建模,即可 得到研究区中所有光斑内的 LAI 预测值。同时,因 为 ICESat - 2/ATLAS 数据采集特性,基本可以实现 全球覆盖<sup>[19]</sup>,保证了研究区选取的灵活性。

# 5 结论

为评估星载光子计数雷达估测 LAI 的能力,研 究基于 ICESat -2 卫星先进地形激光测高系统(AT-LAS)获取的光子点云数据对香格里拉市内所有光 斑进行了数据预处理、参数提取和模型优化等步骤。 总结出了一套自动查找随机森林模型超参数优化后 估测 LAI 的程序方法,并进行了优化前后的差异性 分析。确定了基于 ICESat - 2/ATLAS 数据经过去 噪、分类算法处理后提取的参数在对 LAI 估测中展 现了极好的效果,模型优化前后的决定系数 R<sup>2</sup>均大 于 0.8。同时使用随机表面查找算法进行随机森林 模型的超参数优化,可以较为明显地提升模型估测 精度。最后将 LAI 估测值进行空间制图,总体呈现 四周低、中间高的特点;同时,LAI 较高的区域总体 呈现西北向东南贯穿的趋势,与研究区的森林分布 现状极为吻合。综上,使用 ATLAS 数据产品估测 LAI 是可行的。

#### 参考文献(References):

- [1] Chen J M, Menges C H, Leblanc S G. Global mapping of foliage clumping index using multi – angular satellite data [J]. Remote Sensing of Environment, 2005, 97 (4):447 – 457.
- [2] Friedl M A, Sulla Menashe D, Tan B, et al. MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets[J]. Remote Sensing of Environment, 2010, 114(1): 168 – 182.
- [3] 林晓娟. 基于 ICESat 2 和 GEDI 森林冠层高度和森林地上生物量遥感诊断[D]. 北京:中国科学院大学(中国科学院空天信息创新研究院),2021.

Lin X J. Remote sensing diagnosis of forest canopy height and forest aboveground biomass based on ICESat – 2 and GEDI[D]. Beijing; University of Chinese Academy of Sciences (Aerospace Information Research Institute, Chinese Academy of Sciences), 2021.

- [4] Moussavi M S, Abdalati W, Scambos T, et al. Applicability of an automatic surface detection approach to micro – pulse photon – counting LiDAR altimetry data: Implications for canopy height retrieval from future ICESat – 2 data[J]. International Journal of Remote Sensing, 2014, 35(13):5263 – 5279.
- [5] Liu M S, Xing Y Q, Wu H B, et al. Study on mean forest canopy height estimation based on ICESat – GLAS waveforms [J]. Forest Research, 2014, 27(3):309 – 315.
- [6] García M, Popescu S, Riaño D, et al. Characterization of canopy fuels using ICESat/GLAS data[J]. Remote Sensing of Environment, 2012,123:81 – 89.
- [7] Luo S, Wang C, Li G, et al. Retrieving leaf area index using ICE-Sat/GLAS full – waveform data [J]. Remote Sensing Letters, 2013,4(8):745-753.
- [8] Neumann T A, Martino A J, Markus T, et al. The ice, cloud, and land elevation satellite - 2 mission: A global geolocated photon product derived from the advanced topographic laser altimeter system[J]. Remote Sensing of Environment, 2019, 233:111325.
- [9] Huang X, Xie H, Liang T, et al. Estimating vertical error of SRTM and map – based DEMs using ICESat altimetry data in the eastern Tibetan Plateau [J]. International Journal of Remote Sensing, 2011,32(18):5177-5196.
- [10] 董佳臣,倪文俭,张志玉,等. ICESat 2 植被冠层高度和地表 高程数据产品用于森林高度提取的效果评价[J]. 遥感学报, 2021,25(6):1294-307.

Dong J C, Ni W J, Zhang Z Y, et al. Performance of ICESat - 2

2023 年

ATL08 product on the estimation of forest height by referencing to small footprint LiDAR data[J]. National Remote Sensing Bulletin, 2021,25(6):1294-1307.

- [11] Huang J P, Xing Y Q, Qin L. Review of noise filtering algorithm for photon data [J]. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020, 42:105 – 110.
- [12] Gwenzi D, Lefsky M A, Suchdeo V P, et al. Prospects of the ICESat -2 laser altimetry mission for savanna ecosystem structural studies based on airborne simulation data [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 118:68 - 82.
- [13] 张 欢,李弘毅,李浩杰,等. 基于机载 LiDAR 的高寒山区遥感 高程数据精度评估[J]. 遥感技术与应用,2021,36(6):1311 – 1320.

Zhang H, Li H Y, Li H J, et al. Accuracy evaluation of remote sensing elevation data in alpine mountains based on airborne LiDAR [J]. Remote Sensing Technology and Application, 2021, 36(6): 1311 – 1320.

- [14] Neuenschwander A L, Magruder L A. The potential impact of vertical sampling uncertainty on ICESat - 2/ATLAS terrain and canopy height retrievals for multiple ecosystems [J]. Remote Sensing, 2016,8(12):1039.
- [15] Narine L L, Popescu S, Neuenschwander A, et al. Estimating aboveground biomass and forest canopy cover with simulated ICESat – 2 data [J]. Remote Sensing of Environment, 2019, 224:1 – 11.
- [16] Zhang J, Tian J, Li X, et al. Leaf area index retrieval with ICESat 2 photon counting LiDAR [J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 103:102488.
- [17] Shi L, Zhao H, Li Y, et al. Evaluation of Shangri La County's tourism resources and ecotourism carrying capacity [J]. International Journal of Sustainable Development and World Ecology, 2015,22(2):103-109.
- [18] 俄相颖,戴光耀,吴松华. ICESat 2 ATL03 数据预处理及校正 方法[J]. 红外与激光工程,2021,50(6):92-102.
  E X Y, Dai G Y, Wu S H. ICESat - 2 ATL03 data preprocessing and correction method[J]. Infrared and Laser Engineering,2021, 50(6):92-102.
- [19] Magruder L, Neumann T, Kurtz N. ICESat 2 early mission synopsis and observatory performance [J]. Earth and Space Science, 2021,8(5):e2020EA001555.
- [20] Neuenschwander A, Guenther E, White J C, et al. Validation of IC-ESat – 2 terrain and canopy heights in boreal forests [J]. Remote Sensing of Environment, 2020, 251:112110.
- [21] Parrish C E, Magruder L A, Neuenschwander A L, et al. Validation of ICESat – 2 ATLAS bathymetry and analysis of ATLAS's bathymetric mapping performance[J]. Remote Sensing, 2019, 11(14): 1634.
- [22] Wang C, Zhu X, Nie S, et al. Ground elevation accuracy verification of ICESat - 2 data: A case study in Alaska, USA [J]. Optics Express, 2019, 27(26):38168 - 38179.
- [23] Ghosh S M, Behera M D, Paramanik S. Canopy height estimation using Sentinel series images through machine learning models in a mangrove forest[J]. Remote Sensing, 2020, 12(9):1519.
- [24] 马山木,甘甫平,吴怀春,等. ICESat 2 数据监测青藏高原湖 泊 2018—2021 年水位变化[J].自然资源遥感,2022,34(3):

164 - 172. doi:10.6046/zrzyyg.2021329.

Ma S M, Gan F P, Wu H C, et al. Monitoring lake level changes on the Tibetan Plateau from 2018 to 2021 using ICESat - 2 data[J]. Remote Sensing for Natural Resources, 2022, 34(3):164 - 172. doi:10.6046/zrzyyg.2021329.

- [25] Shen X, Ke C Q, Yu X, et al. Evaluation of ice, cloud, and land elevation satellite - 2 (ICESat - 2) land ice surface heights using airborne topographic mapper (ATM) data in Antarctica[J]. International Journal of Remote Sensing, 2021, 42(7):2556-2573.
- [26] Chen Y, Ma L, Yu D, et al. Improving leaf area index retrieval using multi – sensor images and stacking learning in subtropical forests of China[J]. Remote Sensing, 2021, 14(1):148.
- [27] Fang H, Baret F, Plummer S, et al. An overview of global leaf area index (LAI): Methods, products, validation, and applications [J]. Reviews of Geophysics, 2019, 57(3):739-799.
- [28] Chen B, Pang Y, Li Z, et al. Ground and top of canopy extraction from photon – counting LiDAR data using local outlier factor with ellipse searching area [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(9):1447 – 1451.
- [29] Nie S, Wang C, Xi X, et al. Estimating the vegetation canopy height using micro – pulse photon – counting LiDAR data[J]. Optics Express, 2018, 26(10): A520 – A540.
- [30] An M, Xing W, Han Y, et al. The optimal soil water content models based on crop – LAI and hyperspectral data of winter wheat[J]. Irrigation Science, 2021, 39(6):687 – 701.
- [31] Zhang J, Kerekes J. An adaptive density based model for extracting surface returns from photon – counting laser altimeter data [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 12(4): 726 – 730.
- [32] Xia S, Wang C, Xi X H, et al. Point cloud filtering and tree height estimation using airborne experiment data of ICESat - 2 [J]. Remote Sensing, 2014, 18(4):1199 - 1207.
- [33] 陆大进,黎 东,朱笑笑,等. 基于卷积神经网络的 ICESat 2 光子点云去嗓分类[J]. 地球信息科学学报,2021,23(11):2086 2095.
  Lu D J,Li D,Zhu X X, et al. Denoising and classification of ICESat 2 photon point aloud based on convolutional neural network [J].

2 photon point cloud based on convolutional neural network [ J ] Journal of Geo – Information Science, 2021, 23(11): 2086 – 2095.

- [34] Zhu X, Nie S, Wang C, et al. A ground elevation and vegetation height retrieval algorithm using micro – pulse photon – counting Li-DAR data[J]. Remote Sensing, 2018, 10(12):1962.
- [35] Tang H, Dubayah R, Swatantran A, et al. Retrieval of vertical LAI profiles over tropical rain forests using waveform LiDAR at La Selva, Costa Rica [J]. Remote Sensing of Environment, 2012, 124 (9):242 - 250.
- [36] Nie S, Wang C, Dong P, et al. A revised progressive TIN densification for filtering airborne LiDAR data [J]. Measurement, 2017, 104:70-77.
- [37] Axelsson P. DEM generation from laser scanner data using adaptive TIN models[J]. International Archives of Photogrammetry and Remote Sensing, 2000, 33(4):110-117.
- [38] Zhang L,Zeng Y,Zhuang R, et al. In situ observation constrained global surface soil moisture using random forest model[J]. Remote Sensing, 2021, 13 (23):4893.
- [39] Zhu X, Wang C, Nie S, et al. Mapping forest height using photon -

counting LiDAR data and Landsat8 OLI data: A case study in Virginia and North Carolina, USA [J]. Ecological Indicators, 2020, 114:106287.

- [40] Zhou R, Yang C, Li E, et al. Object based wetland vegetation classification using multi – feature selection of unoccupied aerial vehicle RGB imagery[J]. Remote Sensing, 2021, 13(23):4910.
- [41] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit learn: Machine learning in Python [J]. The Journal of Mmachine Learning Research, 2011, 12:2825 – 2830.
- [42] Narine L L, Popescu S C, Malambo L. Using ICESat 2 to estimate and map forest aboveground biomass: A first example [J]. Remote Sensing, 2020, 12(11):1824.
- [43] Narine L L, Popescu S C, Malambo L. Synergy of ICESat 2 and Landsat for mapping forest aboveground biomass with deep learning [J]. Remote Sensing, 2019, 11(12):1503.
- [44] Su T, Spicer R A, Wu F X, et al. A middle Eocene lowland humid

subtropical "Shangri – La" ecosystem in central Tibet [J]. Proceedings of the National Academy of Sciences, 2020, 117(52): 32989 - 32995.

- [45] Li W, Niu Z, Shang R, et al. High resolution mapping of forest canopy height using machine learning by coupling ICESat – 2 Li-DAR with Sentinel – 1, Sentinel – 2 and Landsat8 data[J]. International Journal of Applied Earth Observation and Geoinformation, 2020,92:102163.
- [46] Silva C A, Duncanson L, Hancock S, et al. Fusing simulated GEDI, ICESat – 2 and NISAR data for regional aboveground biomass mapping[J]. Remote Sensing of Environment, 2021, 253:112234.
- [47] 夏恒,汤健,乔俊飞. 深度森林研究综述[J]. 北京工业大学 学报,2022,48(2):182-196.
   Xia H, Tang J, Qiao J F. Review of deep forest[J]. Journal of Beijing University of Technology,2022,48(2):182-196.

# Optimizing an ICESat2 – based remote sensing estimation model for the leaf area index of mountain forests in southwestern China

XI Lei, SHU Qingtai, SUN Yang, HUANG Jinjun, SONG Hanyue (College of Forestry, Southwest Forestry University, Kunming 650224, China)

Abstract: The leaf area index (LAI) is a critical parameter for the forest ecosystem. Improving the remote sensing estimation accuracy of the regional LAI of mountain forests at a low cost is of great significance for accurately determining the LAIs of forests and for further understanding the forest ecosystem. With spaceborne LiDAR ICESat -2/ATLAS data as a primary information source, this study investigated Shangri - La City in mountainous areas in southwestern China. Based on the remote sensing estimation model using random forest (RF) regression, RF hyperparameter optimization, and the data of 51 measured sample plots of LAI, this study analyzed the estimation effects of the model using accuracy evaluation indicators such as coefficient of determination  $(R^2)$ , root mean square error (RMSE), mean absolute error (MAE), and median absolute error (MedAE). The results are as follows: The hyperparameter optimization of the RF regression model using a random surface search algorithm can significantly improve the estimation accuracy of LAI. The extracted characteristic parameters of ground spots showed high contribution and excellent effects in the LAI estimation of mountain forests. Therefore, they can be applied to the estimation of regional LAI of mountain forests. The RF regression model optimized using the random surface search algorithm yielded higher estimation accuracy. The estimation results were consistent with the forest distribution in the study area, indicating certain generality. Finally, this study determined that it is feasible to employ ICESat - 2/ATLAS data products for LAI estimation, providing a reference for medium - to large - scale LAI estimation based on spaceborne LiDAR.

Keywords: ICESat - 2/ATLAS; leaf area index; hyperparameter optimization; random forest; Shangri - La (责任编辑: 李 瑜)