

DOI:10.12403/j.1001-1498.20230450

基于序贯高斯条件模拟的 GEDI 数据 联合 Landsat8 反演森林地上生物量

罗绍龙, 舒清态*, 余金格, 胥 丽, 杨正道

(西南林业大学林学院, 云南 昆明 650224)

摘要: [目的] 单一遥感技术估测森林生物量存在较大局限性, 本研究旨在利用多源遥感协同技术互补激光雷达和光学遥感的优势, 提高生物量估测精度。[方法] 以星载激光雷达 GEDI 和光学遥感 Landsat8 数据为主要信息源, 采用序贯高斯条件模拟方法实现 GEDI 光斑数据由“点”到“面”的空间扩展, 结合地面 138 块生物量调查样地, 利用随机森林回归方法估测云南省香格里拉云冷杉林的地上生物量。[结果] (1) 采用序贯高斯条件模拟方法对 GEDI 光斑点进行空间插值, 模拟的 12 个生物物理指标在空间分布上呈现出随机性、破碎化的特征, 这与森林的空间分布聚集性非常相似, 参与建模的 9 个指标 OEC 均大于 0.90; (2) 利用单一 Landsat8 光学遥感数据和地形因子构建的随机森林模型精度为: $R^2=0.82$, $RMSE=35.51 \text{ t}\cdot\text{hm}^{-2}$, $P=0.77$; Landsat8 数据协同星载激光雷达 GEDI 数据构建的随机森林模型精度为: $R^2=0.86$, $RMSE=32.11 \text{ t}\cdot\text{hm}^{-2}$, $P=0.80$, 模型精度明显提升; (3) 利用多源遥感技术估测的香格里拉 2019 年云冷杉林地上的生物量总量为 37 042 605.68 t, 平均生物量为 $123.28 \text{ t}\cdot\text{hm}^{-2}$ 。[结论] 基于地统计学的序贯高斯条件模拟方法考虑到研究对象的空间异质性、能克服一定的平滑效应, 用于实现激光点由“点”到“面”的空间扩展是可行的。星载激光雷达 GEDI 与光学遥感 Landsat8 协同的多源遥感数据可有效填补单一遥感数据源的缺陷, 提高森林生物量的估测精度, 能为激光雷达联合光学遥感估测大范围、全覆盖的森林生物量提供参考。

关键词: GEDI; Landsat 8; 序贯高斯条件模拟; 随机森林; 生物量

中图分类号: X87; S757.2

文献标识码: A

文章编号: 1001-1498(2024)02-0001-12

森林生物量是森林生态系统长期物质循环与能量流动的结果, 它是反映森林生产力和森林生态系统功能的重要参数^[1], 在森林的经营、监测与评价中起到重要作用^[2]。森林生物量的估测对合理保护、管理和利用森林资源有重要意义。传统的森林生物量估测主要采取实地测量法, 虽然精度较高, 但受到时间和空间差异的限制^[3], 需要耗费大量的人力和物力, 并且对森林有一定的破坏作用^[4]。随着 3S 技术的迅速发展, 结合遥感数据估测生物量成为一种有效的技术手段^[5]。目前, 森林生物量估测大多是利用野外实测的生物量样地数据, 通过分

析样地数据与光学遥感因子之间的相关关系建立估测方程^[6]推广到区域尺度的生物量分布。然而, 光学遥感易受天气影响, 对植被的穿透性较弱, 只能获取森林冠层表面的信息^[7], 而且信号容易饱和、数据存在混合像元等问题^[8-9], 导致模型估算精度不高^[10]。

激光雷达 (LiDAR) 能快速获取目标的空间三维信息、对植被的穿透力强、可实现对森林冠层垂直结构参数的精准监测^[11-12], 在森林参数反演、生物量估测方面具有很大优势。由于激光雷达是离散采样, 很难提供全覆盖的数据, 尽管它突破了光学遥

收稿日期: 2023-11-22 修回日期: 2023-12-27

基金项目: 云南省农业联合专项-重点项目 (202301BD070001-002); 云南省教育厅科学研究基金项目 (2023Y0728)。

* 通讯作者: 舒清态, 教授, 博士生导师, 主要研究方向为林业 3S 技术应用。E-mail: shuqt@163.com

感在森林应用领域的诸多缺陷,但依然存在数据获取和处理较难等问题^[13]。此外,任何单一的遥感技术都存在局限性,激光雷达技术也不例外^[14],多传感器协同应用能够填补激光雷达在林业应用领域存在的不足^[15],因此,激光雷达协同光学遥感以获取更高精度的森林参数迅速得到诸多研究者的应用与推广^[16-17]。池弘等^[6]利用 GLAS 星载激光雷达结合 Landsat/ETM+ 数据估测了长白山地区森林地上生物量;边瑞^[18]利用无人机激光雷达结合 Landsat 数据估测了祁连山国家公园的森林生物量;巨一琳等^[19]联合机载 LiDAR 数据和多光谱数据估算了根河市大兴安岭生态观测站寒温带天然林的生物量。

全球生态系统动力学调查 (GEDI) 多波束激光雷达相比于 GLAS,采样密度有较大提高,实现对森林冠层高度、生物量等参数的测量具有重要意义^[20-21]。本研究以典型高寒山区香格里拉市为研究区,在对 GEDI 波形数据处理的基础上提取大量的森林参数,利用基于地统计学的序贯高斯条件模拟方法对 GEDI 光斑点森林参数变量实现由“点”到“面”的空间扩展^[22],获取全覆盖的 GEDI 数据,结合 Landsat8-OLI 数据和 2016 年调查的 138 个野外样地数据,利用随机森林方法估算香格里拉 2019 年云冷杉林地上生物量,评估多源遥感数据协同集成应用的优势。

1 材料与方法

1.1 研究区概况

香格里拉市位于云南省西北部“三江并流”区域,地理坐标为 26°52′~28°52′ N, 99°20′~100°19′ E,总面积 11 613 km²,最高海拔为 5 545 m,最低海拔为 1 503 m,属于山地寒温带季风气候,全年气温偏低,年平均气温仅为 5.4 ℃,四季交替不明显,夏秋时节雨水较多,冬春时节较为干旱,年均降水量为 268~945 mm。境内生态系统完好,森林资源丰富,森林覆盖率达 74.99%,优势树种有云冷杉 (*Picea asperata* Mast. and *Abies fabri* (Mast.) Craib)、高山松 (*Pinus densata* Mast.)、云南松 (*Pinus yunnanensis* Franch.)、高山栎 (*Quercus semicarpifolia* Smith) 等。研究区地理位置见图 1。

1.2 数据来源与处理

1.2.1 星载激光雷达 GEDI 数据 全球生态系统

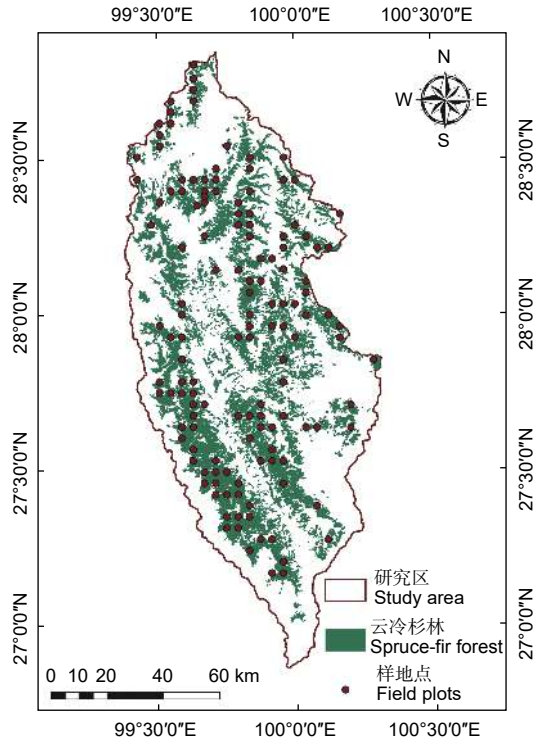


图 1 研究区地理位置

Fig. 1 Geographical location of the study area

动力学调查 (GEDI) 雷达由美国国家航天航空局 (NASA) 2018 年 12 月 5 日发射成功,自 2019 年 4 月开始收集全球 51.6° N~51.6° S 之间基于足印的测量数据,轨道高度约 400 km,光斑直径为 25 m,沿轨光斑点间隔距离为 60 m,相邻轨道间距约 600 m^[23],提供了较高的采样密度和精度。GEDI 包含的数据产品种类丰富,根据数据处理阶段的不同分为 L1、L2、L3、L4 四个级^[24]。本研究所使用的 GEDI 数据为 2019 年 4—12 月获取的 L2B 产品,它是光斑尺度的生物物理指标,包含陆地表面上每个激光足印点精确的经纬度坐标、高程、海拔、覆盖率和垂直剖面指标等,数据从 NASA 官方平台获取 (<https://search.earthdata.nasa.gov>)。

提取 37 个 GEDI 光斑数据指标,包括 quality_flag、degrade_flag、sensitivity 3 个质量指标,各指标的描述见表 1。GEDI 为波形数据,采集过程中会受到云层和其他环境因素的影响,存在部分地理定位精度差以及信号质量差的光斑,因此根据 3 个质量指标对初始的 3 858 个 GEDI 光斑点进行筛选。quality_flag 有 0 和 1 两个值,值为 0 时表示光斑信号质量差,需剔除;degrade_flag 有 0 和 1 两个值,值为 1 时表示卫星处于降轨状态,

所获取的光斑点无效, 需剔除; **Sensitivity** 是灵敏度参数, 阈值为 0 到 1, 在陆地上使用的 GEDI 光斑点灵敏度阈值一般在 0.95 以上, 剔除阈值小于 0.95 的光斑点^[21]。筛选后共有 1 302 个光斑点,

其中 80% 作为训练数据集, 用于获取各变量的空间分布结果, 剩余的 20% 作为验证数据集, 用于检验空间插值的精度。筛选前后的研究区 GEDI 光斑点分布见图 2。

表 1 本研究所提取的 GEDI 指标描述

Table 1 Description of GEDI indexes extracted in this study

指标 Indexes	描述 Description	指标 Indexes	描述 Description
cover	总的冠层覆盖率	rg_aN (N=1~6)	GEDI 6种算法所对应波形中地面分量的积分
modis_nonvegetated	modis数据的无植被标准差百分比	rv	波形中植被分量的积分
fhd_normal	叶高多样性指数	rv_aN (N=1~6)	GEDI 6种算法所对应波形中植被分量的积分
modis_treecover	modis数据的树木覆盖率	rx_energy_aN (N=1~6)	GEDI 6种算法所对应的接收波形的能量
pai	总的植被面积指数	rh100	接收波形信号开始的离地高度
pgap_theta	植被间隙	sensitivity	灵敏度参数
pgap_theta_error	植被间隙的误差	quality_flag	信号质量标志
pgap_theta_aN (N=1~6)	GEDI 6种算法所对应的植被间隙	degrade_flag	状态下降标志
rg	波形中地面分量的积分		

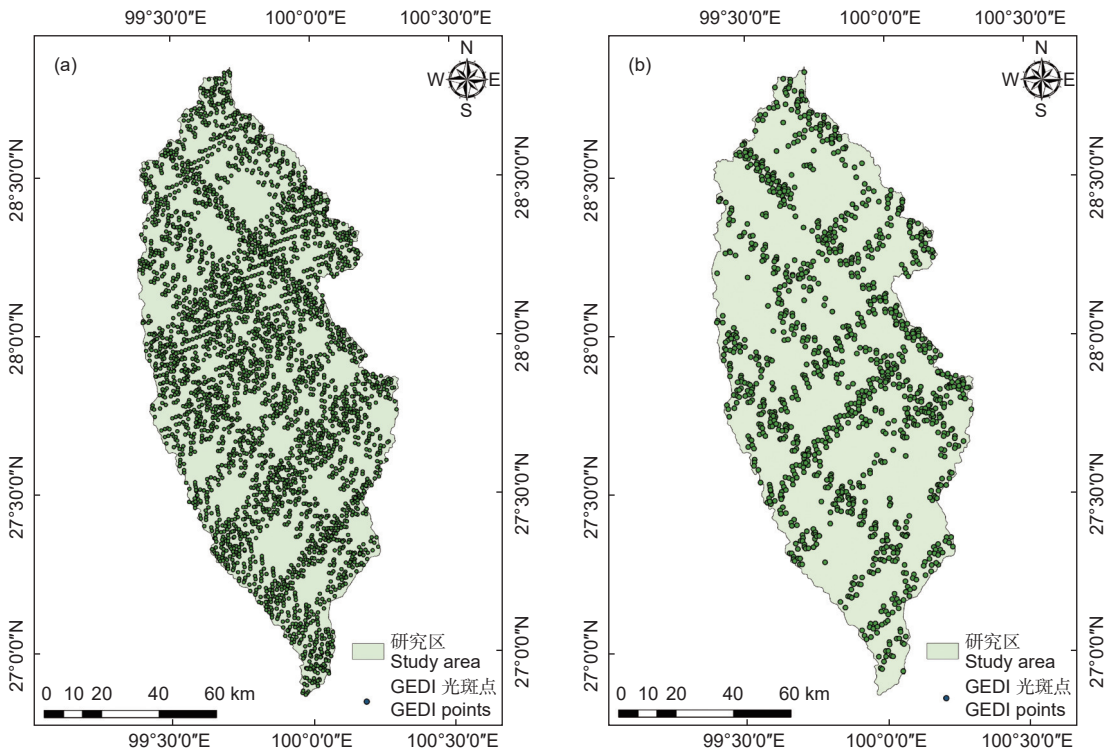


图 2 研究区筛选前 (a) 和筛选后 (b) GEDI 光斑点分布

Fig. 2 GEDI light spot distribution maps before (a) and after (b) screening in the study area

1.2.2 Landsat8 遥感影像数据 使用分辨率为 30 m 的 Landsat8-OLI 数据, 共需 3 景影像覆盖整个研究区, 数据获取时间为 2019 年 12 月。遥感影像广泛被用于森林生物量的估测研究, 在前期对影

像预处理的基础上, 提取 NDVI、RVI、DVI、PVI、EVI、ARVI、RDVI、VI3、SARVI、MSAVI、SLAVI、VIS234、B53、B345、ND43、ND563 等 16 个植被指数^[25-26]用于云冷杉林地上生物量的

估测。

1.2.3 野外调查样地数据 野外调查数据为 2016 年香格里拉市森林资源二类调查的角规控制样地数据, 共有 138 块云冷杉样地, 每块样地面积为 1 hm², 包含样地点坐标、小班地类、优势树种、平均胸径、树高等参数, 利用二元材积表计算样地的林分蓄积量。由于 Landsat8-OLI 和 GEDI 的数据采集时间均为 2019 年, 所以根据《云南省主要树种材积生长率表》^[27] 的材积净生长率 (云杉 0.85%, 冷杉 0.88%) 推算 2019 年的林分蓄积量, 然后利用林木生物量扩展因子法^[28] 计算 2019 年的样地生物量。蓄积量-生物量转换模型见

式 (1), 云冷杉生物量扩展因子及木材密度见表 2, 138 块样地的 2019 年生物量统计参数见表 3。为与遥感影像元尺度相匹配, 将每 hm² 样地生物量数据转换为 900 m² 的样地生物量数据, 样地点的空间分布见图 1。

表 2 香格里拉市云冷杉地上生物量扩展因子及木材密度
Table 2 Aboveground biomass expansion factor and wood density of spruce-fir in Shangri-La

树种 Tree species	木材密度 SVD	生物量扩展因子 BEF
云杉 <i>Picea asperata</i>	0.342	1.734 5
冷杉 <i>Abies fabri</i>	0.366	1.316 4

表 3 样地生物量描述统计分析

Table 3 Statistical analysis of field plots biomass

统计参数 Statistical parameters	数量/块 Number	最小值/(t·hm ⁻²) Minimum	最大值/(t·hm ⁻²) Maximum	均值/(t·hm ⁻²) Mean	标准差/(t·hm ⁻²) Standard deviation	变异系数 Coefficient of variation
值 Value	138	6.38	368.65	125.77	84.00	0.67

$$B = V \cdot SVD \cdot BEF \quad (1)$$

式中, B 为森林地上生物量, V 为蓄积量, SVD 为某一树种或树种组的木材密度, BEF 为生物量扩展因子。

1.2.4 数字高程模型数据 数字高程模型 (DEM) 数据为美国国家航空航天局 (NASA) 和日本宇宙航空研究开发机构 (JAXA) 共同开发的 ASTER GDEM V2 数字高程数据产品, 分辨率为 30 m。该数据用于 Landsat8 影像的地形辐射校

正, 以及基于栅格表面算法提取坡度、坡向和程 3 个地形因子 (图 3)。根据图 3 可知, 香格里拉海拔较高, 总体呈现出西北高东南低的趋势, 海拔落差较大, 境内地势陡峭, 多为斜坡、陡坡和峭破, 地形结构复杂。

1.3 研究方法

1.3.1 变异函数 变异函数的理论模型可通过球状模型、指数模型和高斯模型进行拟合, 用于描述区域化变量的变化规律, 变异函数的模型由模型类

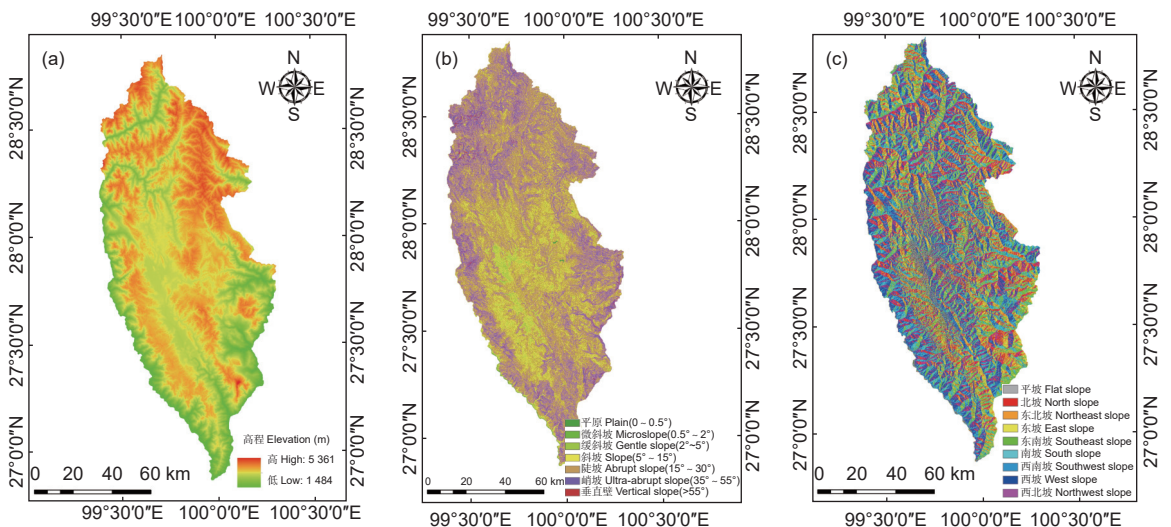


图 3 研究区高程 (a)、坡度 (b) 及坡向 (c)

Fig. 3 Elevation (a), slope (b) and aspect (c) maps of the study area

型、基台值 (Sill)、变程 (Range) 和块金值 (Nugget) 唯一确定^[29-30]。本研究利用 GS + 9 软件对 GEDI 光斑变量进行变异函数分析, 采用 3 个模型进行预测, 以模型的决定系数 (R^2), 残差平方和 (RSS) 对模型进行评价。在分析前, 对所有变量进行归一化处理以提高模型的收敛速度, 即将所有变量的数值调至 [0,1] 或 [-1,1], 计算方法见式 (2), 并对所有变量的数据结构进行正态分布性检验, 将不符合正态分布的变量进行开立方变换, 符合条件后方可进行变异函数分析。在前期对 GEDI 光斑变量进行变异函数分析和筛选的基础上, 选择最优变异函数模型用于实现序贯高斯条件模拟。

$$Y = \frac{y - y_{min}}{y_{max} - y_{min}} \quad (2)$$

式中, Y 为归一化后的值, y 为初始值, y_{min} 为初始数据的最小值, y_{max} 为初始数据的最大值。

1.3.2 序贯高斯条件模拟 序贯高斯条件模拟 (SGCS) 以变异函数理论为基础, 结合区域化变量, 用随机模拟方法估计研究变量。根据已知数据构造高斯函数, 将区域化随机变量 $Z(x)$ 的每一个取值看作是符合高斯函数 (即正态分布函数) $F(x)$ 的一次随机实现。在每一个模拟位置 x_m 处, $F(x)$ 是以 n 个已知数据 $Z(x_i)$ ($i = 1, 2, \dots, n$) 和此前的 $m-1$ 个模拟值 $Z(x_j)$ ($j = 1, 2, \dots, m-1$) 为条件的累积条件概率密度函数^[31], 在模拟过程通过构建局部累计条件概率分布来实现最终结果。该方法对所预测的空间数据可能的取值结果及其概率进行度量, 能在最大程度上正确反映区域化变量的空间波动性, 再现真实资源特性空间变异曲线的波动^[32-33]。在大多数情况下, 原始数据并不满足模拟过程的要求, 需通过正态变换后再进行模拟, 最后通过逆变换将模拟结果转换为初始值^[34]。

1.3.3 随机森林回归 随机森林 (RF) 是以决策树为基础的机器学习集成算法, 最早由 Breiman 等^[35] 提出。它利用 Bootstrap 重抽样方法从训练数据中随机抽取一部分样本, 对每个 Bootstrap 样本进行决策树建模, 然后组合多棵决策树的预测, 通过投票决策的方式得出最终预测结果^[36]。在随机森林中, 每棵决策树都是独立并在随机选择的子样本上进行训练的, 从而有效降低过拟合的风险。该方法能够进行特征重要性分析, 相比于神经网络和支持

向量机等其它算法, 它在分析变量关系上具有显著优势^[37]。

1.3.4 精度评估方法 (1) 空间插值精度评估

序贯高斯条件模拟是基于 Kriging 估计的一种不确定性空间插值方法, 会产生一定的估测误差。本研究采用总体估计值一致性 (OEC)^[38] 对 GEDI 各变量的序贯高斯条件模拟结果进行精度评估, 即以 GEDI 光斑点模拟估计的总体总值 (TG_S) 与光斑点估计的总体总值置信区间中值 (TG_P) 的比值来衡量, OEC 越接近于 1, 表明空间插值精度越高。各指标的计算公式如下:

$$TG_P = \frac{A}{a} \times \frac{1}{n} \sum_i^n gc_i \quad (3)$$

$$TG_S = \sum_j^m sc_j \quad (4)$$

$$OEC = TG_P / TG_S \quad (5)$$

式中, A 为研究区总面积, a 为光斑面积, n 为光斑点观测数量, gc_i 为光斑点实测值, m 为像元总数, sc_j 为第 j 个像元的模拟值。OEC 代表模拟总值与光斑点估计总值的接近程度, 二者相等时, $OEC=1$ 。

(2) 生物量估测模型精度评估

模型精度的高低直接影响到生物量的估测结果, 本研究基于 138 个云冷杉样地生物量数据, 采用留一交叉验证法^[39] 验证云冷杉林生物量非参数估测模型的精度, 以模型的决定系数 (R^2)、均方根误差 (RMSE)、以及总体估测精度 (P) 作为精度评价指标。

2 结果与分析

2.1 GEDI 光斑点变量筛选及序贯高斯条件模拟

2.1.1 GEDI 变量筛选 保留光斑点的 quality_flag 和 degrade_flag 两个指标只有 1 和 0 单一固定值, 参与构建生物量估测模型的意义不大, 因此仅用于光斑点的筛选, 最终采用剩余的 35 个变量进行变异函数分析, 各变量的模型拟合精度见图 4。根据变异函数的分析结果, 剔除 R^2 小于 0.5 的变量, 从而减小序贯高斯条件模拟方法对 GEDI 光斑点进行空间插值的误差, 获取精度较高的空间分布结果。筛选后保留 12 个 GEDI 变量, 其变异函数的拟合参数见表 4。

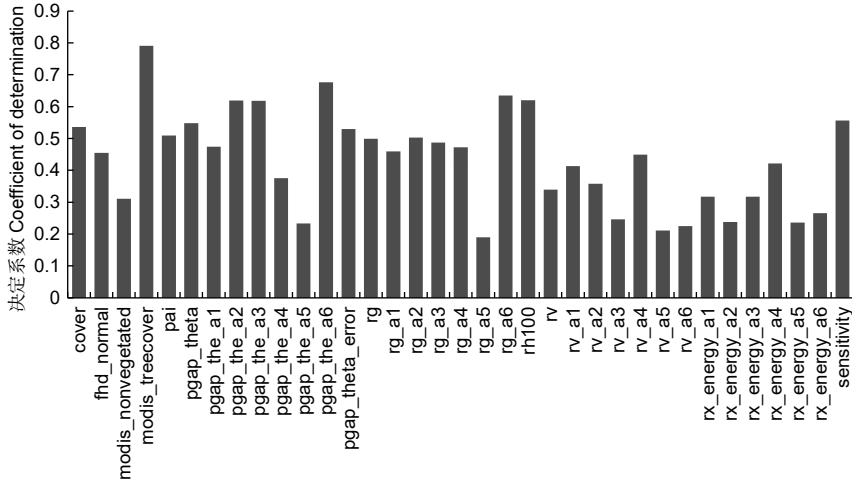


图 4 GEDI 各变量变异函数拟合精度

Fig. 4 Fitting accuracy of variogram of GEDI variables

表 4 GEDI 优选变量变异函数拟合参数

Table 4 Parameter values of variable variogram after optimization

变量 Variable	最优模型 Optimization model	R^2	RSS	C_0	$C_0 + C$	Range
cover	指数模型	0.536	1.695E-04	0.012 00	0.107 00	7 200
modis_treecover	指数模型	0.791	9.483E-05	0.010 70	0.087 80	9 900
pai	指数模型	0.509	9.028E-05	0.009 40	0.079 50	6 900
pgap_theta	指数模型	0.548	1.905E-05	0.004 40	0.037 90	7 200
pgap_theta_a2	指数模型	0.619	1.146E-04	0.011 90	0.104 80	7 200
pgap_theta_a3	指数模型	0.618	9.332E-05	0.012 90	0.107 80	6 600
pgap_theta_a6	指数模型	0.676	1.131E-04	0.013 80	0.120 60	6 900
pgap_theta_error	指数模型	0.530	4.988E-07	0.000 61	0.005 01	7 800
rg_a2	指数模型	0.503	1.900E-05	0.005 10	0.044 50	6 300
rg_a6	指数模型	0.635	1.647E-05	0.006 20	0.053 80	6 300
rh100	指数模型	0.620	3.340E-06	0.002 07	0.016 94	7 500
sensitivity	指数模型	0.556	2.612E-09	0.000 05	0.000 44	6 900

2.1.2 GEDI 变量序贯高斯条件模拟 基于 GS + 9 软件拟合的 12 个 GEDI 变量的最优变异函数, 根据模型参数 (变程、块金值、偏基台值) 进行简单 Kriging 插值, 最后采用 ArcGIS 10.5 中“高斯地统计模拟”工具完成序贯高斯条件模拟^[40]。通过反复实验, 依次设定模拟次数为 10 次、25 次、50 次、75 次、100 次、125 次, 对比分析后, 发现 50 次模拟后像元均值的方差变化趋于稳定, 因此确定模拟的次数为 50 次。为与 GEDI 光斑点的面积尺度相匹配, 输出 GEDI 各变量模拟的栅格数据分辨率为 22.15 m, 并对经过归一化处理和正态

变换后的变量进行反归一化处理和逆变换, 最终得到真实的模拟结果。序贯高斯条件模拟的流程见图 5, 12 个 GEDI 变量模拟的结果见图 6。利用预留 20% 的实测光斑数据检验模拟结果的精度, 其总体估计值一致性的检验结果见表 5。

根据总体估计值一致性检验结果, pgap_theta_error、rg_a2、rg_a6 3 个变量的 OEC 较低, 且经过检验总体估计值未在置信区间内, 因此剔除 OEC 最低的 3 个变量, 最终保留满足精度要求的 9 个变量参与构建云冷杉林生物量估测模型。

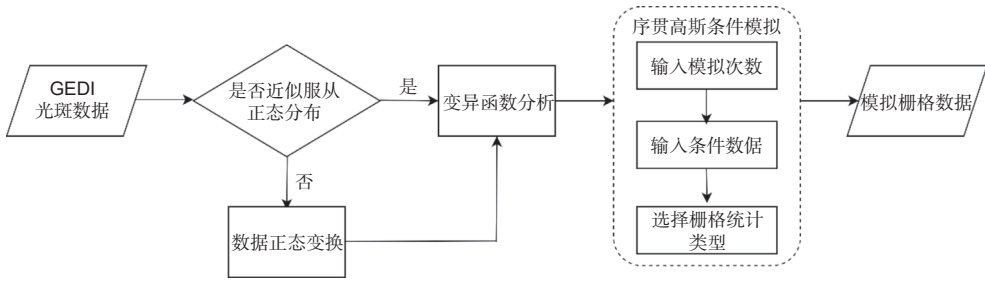


图 5 GEDI 光斑变量序贯高斯条件模拟流程

Fig. 5 Flowchart of Sequential Gaussian Condition Simulation with GEDI spot variable

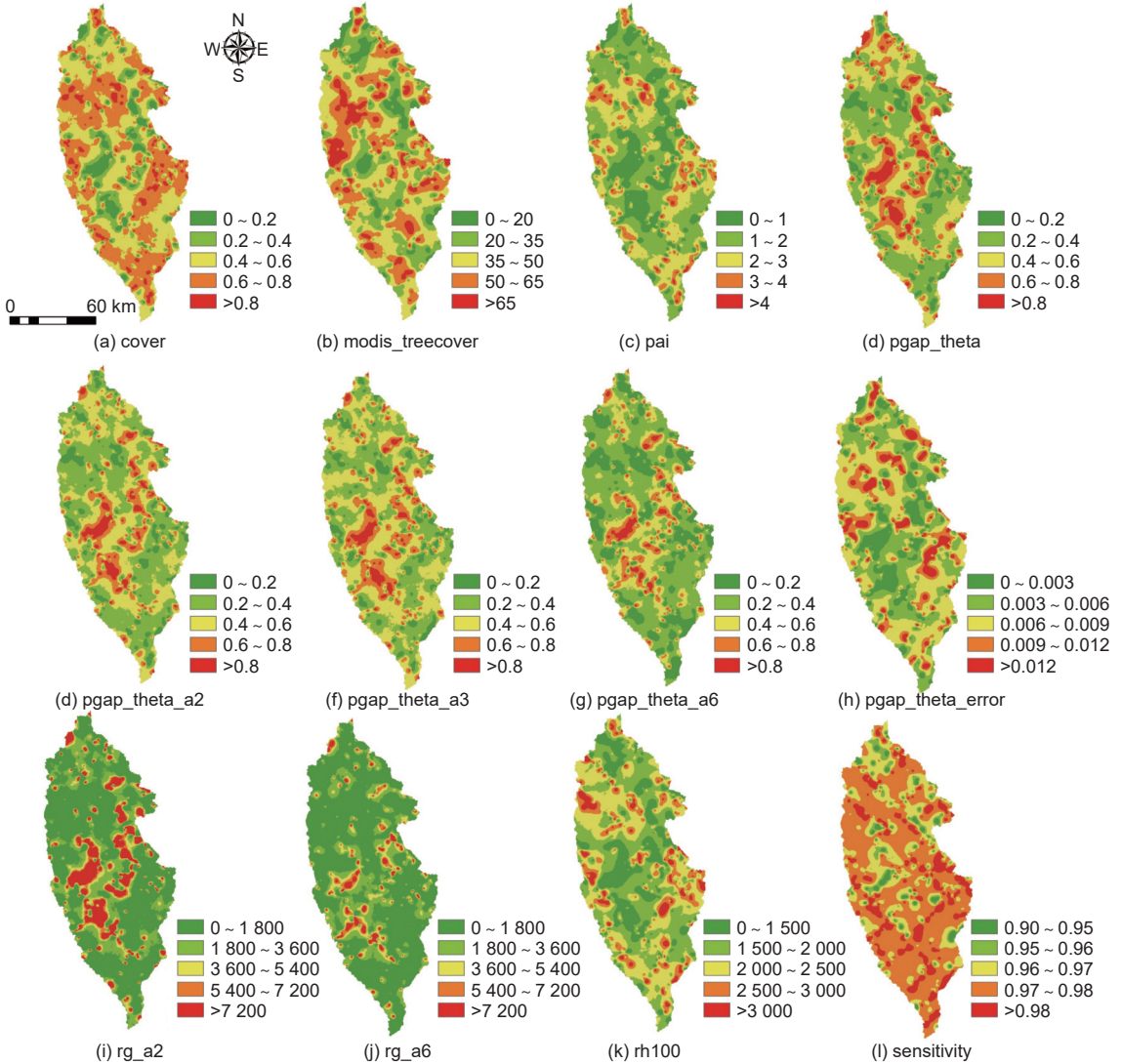


图 6 GEDI 变量序贯高斯条件模拟结果

Fig. 6 The results of Sequential Gaussian Condition Simulation for GEDI variables

2.2 随机森林建模

2.2.1 利用 Landsat8 光学数据构建估测模型 利用 Python 3.7 软件, 基于 Landsat8 影像的 16 个植被指数和 3 个地形因子, 采用随机森林特征重要

性分析方法对参与建模的 19 个特征进行特征重要性贡献度分析, 其中对建模贡献度最大的变量是 VI3, 贡献率为 12.97%; 贡献率最小的变量是 VIS234, 贡献率为 1.52%, 各特征的建模贡献度

表 5 GEDI 变量序贯高斯条件模拟结果总体估计值一致性检验

Table 5 The test of overall estimate consistency for Sequential Gaussian Conditional Simulation results of GEDI variables

变量 Variable	OEC	变量 Variable	OEC	变量 Variable	OEC
cover	0.965	pgap_theta_a2	0.937	rg_a2	0.713
modis_treecover	0.993	pgap_theta_a3	0.979	rg_a6	0.634
pai	0.907	pgap_theta_a6	0.960	rh100	0.962
pgap_theta	0.930	pgap_theta_error	0.849	sensitivity	0.995

占比见图 7。采用所有特征构建随机森林回归模型，模型的 $R^2=0.82$ 、 $RMSE=35.51 \text{ t}\cdot\text{hm}^{-2}$ 、 $P=0.77$ ，模型精度拟合的散点图见图 8。

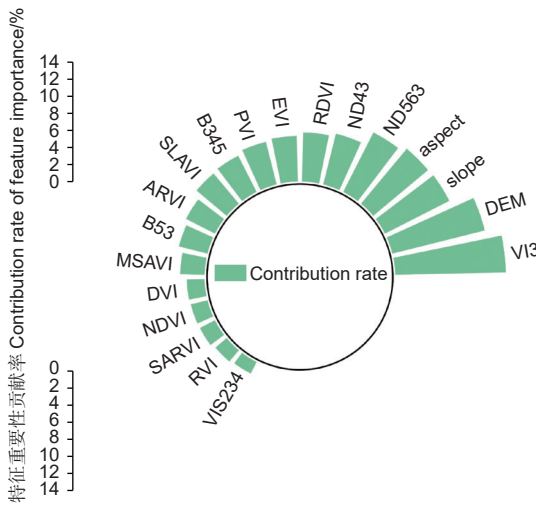


图 7 建模特征重要性贡献度占比

Fig. 7 Contribution proportion of modeling features' importance

栅格数据。针对保留的 9 个 GEDI 光斑变量、Landsat8-OLI 影像的 16 个植被指数和 3 个地形因子，采用随机森林特征重要性分析方法对参与建模的所有特征进行特征重要性贡献度分析，其中对建模贡献率最大的变量是 VI3，贡献率为 12.05%；贡献率最小的变量是 RVI，贡献率为 0.96%，各特征的建模贡献度占比见图 9。

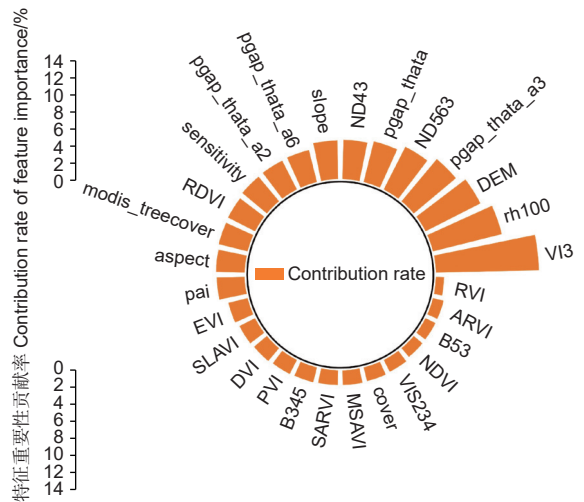
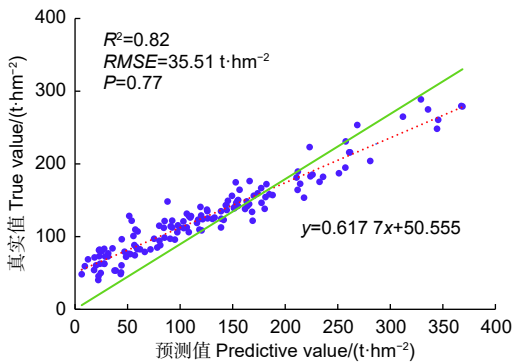


图 9 建模特征重要性贡献度占比

Fig. 9 Contribution proportion of modeling features' importance

利用 9 个 GEDI 光斑变量联合 Landsat8 影像的 16 个植被指数特征、以及 3 个地形因子建立随机森林回归模型，模型的 $R^2=0.86$ 、 $RMSE=32.11 \text{ t}\cdot\text{hm}^{-2}$ 、 $P=0.80$ ，两种遥感数据联合后构建的模型精度明显提升。模型精度拟合的散点图见图 10。

2.2.3 生物量估测模型地形因子贡献度分析 根据图 7 和图 9 可知，两个随机森林模型中 3 个地形因子的建模贡献度占比均较大，尤其以 DEM 较为显著，前后两个模型中的贡献度分别达到 11.18% 和 7.21%。香格里拉地处高寒山区，海拔落差较大，云冷杉为喜阴树种，主要分布在海拔较高的阴坡和半阴坡地带，生长趋势严重受到海拔高

图 8 利用 Landsat8 数据的随机森林模型拟合散点图
Fig. 8 Scatterplot of random forest model fit for using Landsat8 data

2.2.2 联合 GEDI 和 Landsat8 数据构建估测模型

考虑到与样地面积和 Landsat8 数据的栅格大小相匹配，将基于 2.1 节 GEDI 光斑数据 9 个变量的序贯高斯条件模拟结果重采样分辨率为 30 m 的

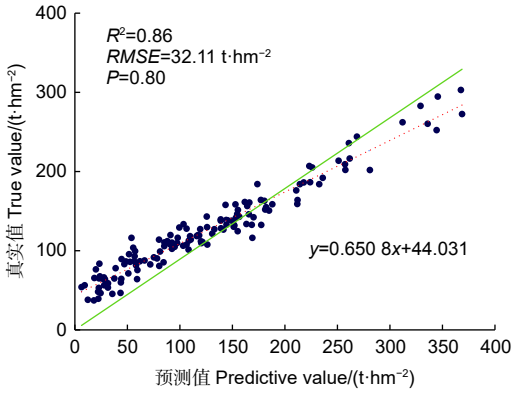


图 10 联合两种遥感数据的随机森林模型拟合散点图

Fig. 10 Scatterplot of random forest model fit for combining two types of remote sensing data

度的影响, 垂直地带差异性较强, 因此 DEM 对模型的贡献度较大。

2.3 多源遥感数据协同的生物量估测

根据两个随机森林模型精度的对比分析结果, 利用星载激光雷达 GEDI 数据协同 Landsat8 光学遥感数据构建模型估测香格里拉 2019 年云冷杉林的地上生物量。估测的 2019 年云冷杉林地上生物量的总量为 37 042 605.68 t, 平均生物量为 123.28 t·hm⁻², 生物量空间分布结果见图 11。香格里拉地处高寒山区, 平均海拔约为 3 500 m, 总体呈现出西北高东南低的趋势。云冷杉林主要集中于南部地区, 根据图 11 可知, 高生物量值并非全部集中分布于南部地区, 而是分布在海拔 3 600 m 以上的高寒地区, 这是因为冷杉林主要分布于海拔 3 500~3 800 m 的阴坡、半阴坡地带, 云杉林主要分布于海拔 3 200~3 700 m 的阴坡、半阴坡以及沟谷地带。此外, 香格里拉人口较多的乡镇主要分布在南部的建塘镇、虎跳峡镇、金江镇和三坝乡, 受人文因素和自然因素的影响, 云冷杉林地上生物量总体分布不均, 区域差异较大。

3 讨论

3.1 序贯高斯条件模拟方法对激光点空间扩展的应用分析

地统计学是研究空间变异及格局的有效方法, 最初被应用于地质研究领域, 随着统计学的发展, 以及森林资源变化监测的重要性日益显著, 该方法也被广泛应用于林业研究领域^[41]。星载激光雷达 GEDI 只能提供离散的采样点, 不具有成像性, 所以研究采用基于地统计学的序贯高斯条件模拟方法

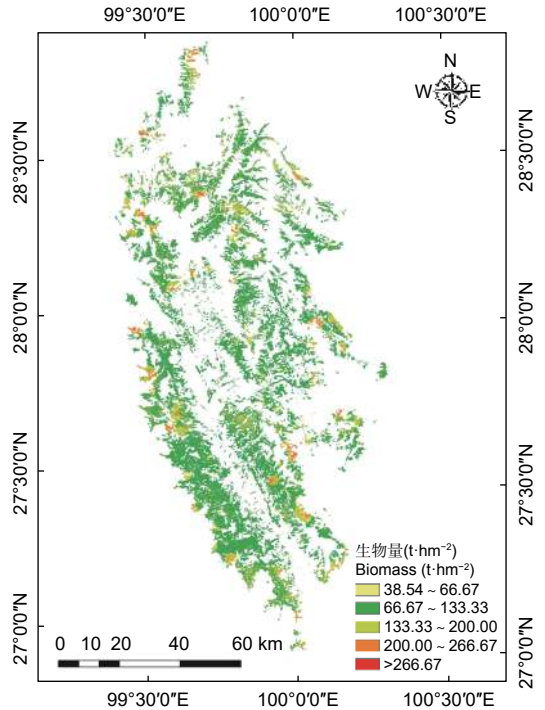


图 11 云冷杉地上生物量估测结果空间分布

Fig. 11 Spatial distribution of aboveground biomass estimation results of spruce-fir forest

实现了 GEDI 光斑由“点”到“面”的空间扩展, 获取与地面调查样地对应的变量指标数据。相比于 Kriging 插值, 该方法克服了一定的平滑效应, 能够在最大程度上真实地展现研究对象的空间分布格局及变化^[31-32]。根据 12 个 GEDI 变量的模拟结果 (图 6) 可知, 每个生物物理指标都呈现出破碎化、随机性的特征, 这与香格里拉森林空间分布聚集性非常相似, 且参与建模的 9 个变量 OEC 均在 0.9 以上, 精度满足要求, 因此, 序贯高斯条件模拟方法可用于激光点的空间扩展, 为离散的“点”数据转向大区域尺度的“面”数据的应用提供了可靠的技术支持。

3.2 云冷杉材积生长率不确定性及驱动因素分析

本研究参考《云南省主要树种材积生长率表》云杉和冷杉的材积净生长率推算 2019 年云冷杉林地上生物量, 该表按 15 个龄级分别计算各龄级的材积净生长率, 其中云杉 I 龄级的材积净生长率为 8.84%, X V 龄级的净生长率为 0.35%; 冷杉 I 龄级的材积净生长率为 14.64%, X V 龄级的净生长率为 0.12%^[27], 两个树种不同龄级的材积净生长率差距较大, 因减少成本和受其他因素限制, 统一采用总平均净生长率, 相关研究表明, 利用该表

给出的净生长率计算的材积生长量结果偏小^[42]。根据2016年二类调查的林木材积推算2019年的林木材积,期间有3年的时间跨度,近年来气候变化多样,时间跨度导致的气候因素会影响林木的生长趋势,进而影响生物量的生长量,贾勃等^[43]的研究发现,生物量的生长量受到年均温度的正向影响;何潇等^[44]的研究也表明气候因子对林分生物量有显著影响。此外,研究利用生物量扩展因子法计算2019年云冷杉林的地上生物量,三重不确定性因素会导致估测结果存在微弱的偏差。胥丽等^[22]利用该方法获取了香格里拉2019年的栎类森林地上生物量精度较高的估测结果,但后续研究中,若条件允许建议采用实地验证法对估测结果进行检验,并对林木材积生长率变化气候驱动因素进行分析,降低生长率不确定性因素导致的生物量估测偏差。

3.3 生物量估测精度分析及模型优化建议

针对生物量的估测结果,本研究与前人的结果做了一些对比。谢福明等^[45]基于KNN模型估测了香格里拉2016年云冷杉林地上生物量, $RMSE$ 为 $41.64 \text{ t}\cdot\text{hm}^{-2}$,而本研究基于随机森林模型的 $RMSE$ 为 $32.11 \text{ t}\cdot\text{hm}^{-2}$,估测精度明显提升。根据云冷杉材积净生长率,以及生物量扩展因子法,基于2016年香格里拉森林资源二类调查的小班数据计算了2019年云冷杉林地上生物量,其总生物量为 $39\,961\,196.39 \text{ t}$,138块样地的2019年生物量平均值为 $125.77 \text{ t}\cdot\text{hm}^{-2}$ 。相比于二者的结果,本研究的估测结果具有一定可靠性,但在精度方面依然有可提升的空间。非参数模型是一种黑箱操作原理,它是直接或间接地从实际系统的实验分析中得到响应,机械式的系统化难以提高模型的预测精度,为提高模型的预测精度,获取更加精准的生物量空间分布制图,可对模型进行优化。宋涵玥等^[46]采用随机搜索法(Random Searching)对模型参数进行优化,优化后的生物量估测精度提高2.01%。此外,还可采用遗传优化算法、粒子群优化算法和贝叶斯优化方法^[47]对模型进行优化。

4 结论

本研究采用星载激光雷达GEDI数据协同Landsat8-OLI影像估测了滇西北生态系统较为完善的高寒山区香格里拉云冷杉林的地上生物量,评估多源遥感数据协同估测森林生物量的优势。主要结论如下:

(1) 利用地统计学GS+9软件对GEDI观测点做变异函数分析,经第一次筛选的12个GEDI变量变异函数的最优模型均为指数模型,模型决定系数均大于0.5。基于变异函数的分析结果,获取12个变量的序贯高斯条件模拟空间分布,采用OEC检验其模拟精度,并进行第二次筛选,筛选后9个GEDI变量的OEC均高于0.9,较接近于1,且模拟估计的总体总值均在置信区间内,满足精度要求,基于地统计学的序贯高斯条件模拟方法可实现GEDI光斑点由“点”到“面”的空间扩展。

(2) 利用单一光学遥感Landsat8数据的16个植被指数和3个地形因子构建生物量估测模型,模型的 $R^2=0.82$, $RMSE=35.51 \text{ t}\cdot\text{hm}^{-2}$, $P=0.77$;星载激光雷达GEDI数据协同光学遥感Landsat8数据构建的生物量估测模型 $R^2=0.86$, $RMSE=32.11 \text{ t}\cdot\text{hm}^{-2}$, $P=0.80$,精度明显提升。多源遥感数据协同能够互补不同的数据优势,提高森林生物量的估测精度。

(3) 利用2019年的两种遥感数据,结合2016年野外调查数据,根据材积净生长率和生物量扩展因子法获取138块云冷杉样地2019年的生物量,利用随机森林回归方法估测的香格里拉2019年云冷杉林的地上生物量总量为 $37\,042\,605.68 \text{ t}$,平均生物量为 $123.28 \text{ t}\cdot\text{hm}^{-2}$,主要分布在 $38.54\sim 282.19 \text{ t}\cdot\text{hm}^{-2}$ 之间。

参考文献:

- [1] 李卓凡. 兴安落叶松林生物量与碳储量的研究[D]. 呼和浩特: 内蒙古农业大学, 2013.
- [2] 曹 养, 岳彩荣, 李天骄, 等. 森林生物量估测进展[J]. 林业调查规划, 2015, 40(6): 22-25.
- [3] 郑冬梅. 三峡库区森林生物量和碳储量的遥感估测研究[J]. 遥感信息, 2013, 28(5): 95-98.
- [4] Potapov P, Li X, Hernandez-Serna A, et al. Mapping global forest canopy height through integration of GEDI and Landsat data[J]. Remote Sensing of Environment, 2021, 253: 112-165.
- [5] Nelson R, Margolis H, Montesano P L, et al. Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations[J]. Remote Sensing of Environment, 2017, 188: 127-140.
- [6] 池 泓, 黄进良, 邱 娟, 等. GLAS星载激光雷达和Landsat/ETM+数据的森林生物量估算[J]. 测绘科学, 2018, 43(4): 9-16.
- [7] 秦 磊. 基于ICESat-2星载激光雷达光子云数据反演森林冠层高度方法研究[D]. 哈尔滨: 东北林业大学, 2020.
- [8] Chopping M, Moisen G G, Su L, et al. Large area mapping of southwestern forest crown cover, canopy height, and biomass using the NASA Multiangle Imaging Spectro-Radiometer[J].

- Remote Sensing of Environment, 2008, 112: 2051-2063.
- [9] 李德仁, 王长委, 胡月明, 等. 遥感技术估算森林生物量的研究进展[J]. 武汉大学学报(信息科学版), 2012, 37(6): 631-635.
- [10] Zhang J K, Rivard B, Sanchez-Azofeifa A, *et al.* Intra and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery[J]. Remote Sensing of Environment, 2006, 105: 129-141.
- [11] Korhonen L, Korpela I, Heiskanen J, *et al.* Airborne discrete-return LiDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index[J]. Remote Sensing of Environment, 2011, 115: 1065-1080.
- [12] Peduzzi A, Wynne R H, Fox T R, *et al.* Estimating leaf area index in intensively managed pine plantations using airborne laser scanner data[J]. Forest Ecology and Management, 2012, 270: 54-65.
- [13] 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, 41(3): 129-136.
- [14] 曹林, 余光辉, 代劲松, 等. 激光雷达技术估测森林生物量的研究现状及展望[J]. 南京林业大学学报(自然科学版), 2013, 37(3): 163-169.
- [15] Straub C, Dees M, Weinacker H, *et al.* Using airborne laser scanner data and CIR orthophotos to estimate the stem volume of forest stands[J]. Photogrammetrie, Fernerkundung, Geoinformation, 2009, 30: 277-287.
- [16] Lefsky M A, Cohen W B, Spies T A. An evaluation of alternate remote sensing products for forest inventory monitoring and mapping of Douglas-fir forests in western Oregon[J]. Canadian Journal of Forest Research, 2001, 31: 78-87.
- [17] Hese S, Lucht W, Schimmler C, *et al.* Global biomass mapping for an improved understanding of the CO₂ balance—the Earth observation mission Carbon3D[J]. Remote Sensing of Environment, 2005, 94: 94-104.
- [18] 边瑞. 基于多源遥感数据的祁连山国家公园森林生物量估算研究[D]. 兰州: 兰州大学, 2021.
- [19] 巨一琳, 姬永杰, 黄继茂, 等. 联合LiDAR和多光谱数据森林地上生物量反演研究[J]. 南京林业大学学报(自然科学版), 2022, 46(1): 58-68.
- [20] Juan G H, Adrián P. Using GEDI Lidar data and airborne laser scanning to assess height growth dynamics in fast-growing species: A showcase in Spain[J]. Forest Ecosystems, 2021, 8: 182-198.
- [21] 韩明辉, 邢艳秋, 李国元, 等. GEDI不同算法组数据反演森林最大冠层高度和生物量精度比较[J]. 中南林业科技大学学报, 2022, 42(10): 72-82.
- [22] Xu L, Shu Q, Fu H, *et al.* Estimation of Quercus Biomass in Shangri-La Based on GEDI Spaceborne Lidar Data[J]. Forests, 2023, 14: 876.
- [23] Dubayah R, Blair J B, Goetz S, *et al.* The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth's forests and topography[J]. Science of Remote Sensing, 2020, 1: 100002.
- [24] 朱笑笑. 基于ICESat-2和GEDI数据的中国30米分辨率森林高度反演研究[D]. 北京: 中国科学院大学, 2021.
- [25] 王金亮, 程鹏飞, 徐申, 等. 基于遥感信息模型的香格里拉森林生物量估算[J]. 浙江农林大学学报, 2013, 30(3): 325-329.
- [26] 曹军, 张加龙, 肖庆琳, 等. 基于随机森林和蒙特卡洛的高山松地上碳储量估测及不确定性分析[J]. 林业科学研究, 2023, 35(3): 131-139.
- [27] 云南省森林资源调查管理处[M]. 云南省主要树种材积生长率表. 1982.
- [28] 胥辉, 张子翼, 欧光龙. 云南省森林生物量和碳储量估算及分布研究[M]. 昆明: 云南科技出版社, 2019.
- [29] Yan P, Peng H, Yan L, *et al.* Spatial variability of soil physical properties based on GIS and Geo-Statistical methods in the Red Beds of the Nanxiong Basin, China[J]. Pol J Environ Stud, 2019, 28: 2961-2972.
- [30] 王铭鑫, 范超, 高秉博, 等. 融合半变异函数的空间随机森林插值方法[J]. 中国生态农业学报(中英文), 2022, 30(3): 451-457.
- [31] 赵彦锋, 化全县, 陈杰. Kriging插值和序贯高斯条件模拟的原理比较及在土壤空间变异研究中的案例分析[J]. 土壤学报, 2011, 48(4): 856-862.
- [32] Qu M, Li W, Zhang C. Spatial distribution and uncertainty assessment of potential ecological risks of heavy metals in soil using sequential gaussian simulation[J]. Human and Ecological Risk Assessment, 2014, 20: 764-778.
- [33] Huang J, Liu W, Zeng G, *et al.* An exploration of spatial human health risk assessment of soil toxic metals under different land uses using sequential indicator simulation[J]. Ecotoxicology and Environmental Safety, 2016, 129: 199-209.
- [34] 吕晓光, 王德发, 姜洪福. 储层地质模型及随机建模技术[J]. 大庆石油地质与开发, 2000, 19(1): 10-16.
- [35] Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author)[J]. Statistical Science, 2001, 16: 199-231.
- [36] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
- [37] 张雷, 王琳琳, 张旭东, 等. 随机森林算法基本思想及其在生态学中的应用——以云南松分布模拟为例[J]. 生态学报, 2014, 34(3): 650-659.
- [38] 张茂震, 王广兴, 葛宏立, 等. 基于空间仿真的仙居县森林碳分布估算[J]. 林业科学, 2014, 50(11): 13-22.
- [39] 岳继博, 杨贵军, 冯海宽. 基于随机森林算法的冬小麦生物量遥感估算模型对比[J]. 农业工程学报, 2016, 32(18): 175-182.
- [40] 徐辰星, 濮励杰, 朱明, 等. 基于序贯高斯条件模拟的土壤重金属含量预测与不确定性评价——以宜兴市土壤Hg为例[J]. 土壤学报, 2018, 55(4): 999-1006.
- [41] Hock B K, Payn T W, Shirley J W. Using a Geographic Information System and Geostatistics to Estimate Site Index of *Pinus radiata* for Kaingaroa Forest[J]. N. Z. J. For. Sci., 1993, 23: 264-277.
- [42] 韩文洪. 云南省森林资源二类调查中的生长量计算分析[J]. 中南林业调查规划, 2009, 28(4): 9-11.
- [43] 贾勃, 王新杰. 东北针阔混交林生物量动态过程及稳定性研究[J]. 林业科学研究, 2023, 35(5): 41-49.
- [44] 何潇, 徐奇刚, 雷相东. 气候敏感的落叶松人工林分生物量模型研究[J]. 林业科学研究, 2021, 34(6): 20-27.
- [45] 谢福明. 基于优化k-NN模型的森林地上生物量遥感估测[D]. 昆

明: 西南林业大学, 2019.

[46] 宋涵玥, 舒清态, 席磊, 等. 基于星载ICESat-2/ATLAS数据的森林地上生物量估测[J]. 农业工程学报, 2022, 38(10): 191-199.

[47] Victoria A H, Maragatham G. Automatic tuning of hyperparameters using Bayesian optimization[J]. Evolving Systems, 2021, 12(1): 217-223.

Combined GEDI Data and Landsat 8 for Inversion of Forest Aboveground Biomass Based on Sequential Gaussian Condition Simulation

LUO Shao-long, SHU Qing-tai, YU Jin-ge, XU Li, YANG Zheng-dao

(College of Forestry, Southwest Forestry University, Kunming 650224, Yunnan, China)

Abstract: [Objective] There are significant limitations in estimating forest biomass using a single remote sensing technology. The research aims to utilize the advantages of multi-source remote sensing collaboration technology to complement LiDAR and optical remote sensing, and to improve the accuracy of biomass estimation. [Methods] Based on the two main information sources, including spaceborne LiDAR GEDI and optical remote sensing Landsat8 data, Sequential Gaussian Conditional Simulation (SGCS) method was used to achieve spatial expansion of GEDI data from "points" to "polygons". Combining with aboveground biomass data of 138 plots, the random forest method was used to estimate the aboveground biomass of *Picea asperata* and *Abies fabri* forests in Shangri-La, Yunnan Province. [Result] (1) The SGCS method was used to perform spatial interpolation on GEDI footprints, and the simulated spatial distribution maps of 12 biophysical indicators showed random and fragmented characteristics, which were very similar to the spatial distribution and clustering of forests, and the OEC of 9 indexes involved in modeling were greater than 0.90. (2) The accuracy of the random forest model constructed using a single optical remote sensing Landsat8 data was: $R^2=0.82$, $RMSE=35.51 \text{ t}\cdot\text{hm}^{-2}$, $P=0.77$; The accuracy of the random forest model constructed by combining two types of remote sensing data was: $R^2=0.86$, $RMSE=32.11 \text{ t}\cdot\text{hm}^{-2}$, $P=0.80$. It could be obviously found that the accuracy of the model was improved. (3) The total aboveground biomass of *Picea asperata* and *Abies fabri* forests in Shangri-La in 2019 estimated by multi-source remote sensing technology was 37 042 605.68 t, and the average biomass was $123.28 \text{ t}\cdot\text{hm}^{-2}$. [Conclusion] The SGCS method based on geostatistics has some advantages, including considering the spatial heterogeneity of the research object and being able to overcome smoothing effects. It is feasible to achieve spatial expansion of GEDI footprints from "point" to "polygon". The multi-source remote sensing data based on the combination of spaceborne LiDAR GEDI and optical remote sensing Landsat8 can effectively fill the defects of a single remote sensing data source, improve the estimation accuracy of forest biomass, and provide a reference for the combination of LiDAR and optical remote sensing to estimate large-scale and fully covered forest biomass.

Keywords: GEDI; Landsat 8; Sequential Gaussian Conditional Simulation; random forest; biomass

(责任编辑: 彭南轩)