

DOI: 10.20103/j.stxb.202309212048

周文武,舒清态,胥丽,杨正道,高应群,吴再昆,夏翠芬,顾纯僖,李华.滇西北森林郁闭度估测模型——基于全球生态系统动力学调查多波束激光雷达数据.生态学报,2024,44(8):3525-3539.

Zhou W W, Shu Q T, Xu L, Yang Z D, Gao Y Q, Wu Z K, Xia C F, Gu C X, Li H. Construction of forest canopy closure estimation model in the northwestern Yunnan based on global ecosystem dynamics investigation multi-beam LiDAR data. Acta Ecologica Sinica, 2024, 44(8): 3525-3539.

## 滇西北森林郁闭度估测模型

### ——基于全球生态系统动力学调查多波束激光雷达数据

周文武,舒清态\*,胥丽,杨正道,高应群,吴再昆,夏翠芬,顾纯僖,李华

西南林业大学 林学院,昆明 650224

**摘要:**探究全球生态系统动力学调查(GEDI)多波束激光雷达数据估测区域森林郁闭度(FCC)的潜力,对于评估森林生态系统状态和林分环境具有重要作用。选取滇西北典型生态脆弱区香格里拉为研究区,以GEDI波形数据为信息源,提取46245个林地光斑参数,使用经验贝叶斯克里金法(EBK)获取光斑参数在研究区未知空间的连续分布,结合54块实测样地数据,采用支持向量机的递归特征消除法(SVM-RFE)、随机森林(RF)和Pearson分析分别优选特征变量,基于贝叶斯优化(BO)随机森林回归模型(BO-RFR)、贝叶斯优化梯度回归模型(BO-GBRT)和偏最小二乘法(PLSR)研建森林郁闭度最佳估测模型。结果表明:(1)EBK法预测精度高,估测结果可靠, $R^2$ :0.20—0.92, RMSE:0.004—2812.912, MAE:0.003—1996.258, MRE:0.007—4.423;(2)基于不同特征优选方法筛选的特征变量和数量略有差异,SVM-RFE法优选出6个参数(cover、pai、sensitivity、rv\_a1、rv\_a4、rg\_a4)的平均交叉验证精度达0.84,RF法以贡献度5%为阈值筛选出5个参数(cover、pai、pgap\_theta\_error、modis\_treecover、modis\_nonvegetated),Pearson法以相关性大于0.3且在0.01水平显著优选出5个参数(cover、pai、rv\_a5、rg\_a5、pgap\_theta\_error);(3)不同特征变量优选方法筛选的建模参数研建估测模型精度差异性较大,以SVM-RFE和RF方法优选参数构建估测模型的精度更佳,SVM-RFE方法优选参数研建估测模型精度变化相对稳定,以RF方法中的BO-GBRT模型为最佳FCC估测模型( $R^2=0.85$ 、RMSE=0.069, $P=86.5%$ );(4)采用BO-GBRT模型估测研究区森林郁闭度和空间制图,与GEDI pai参数预测的FCC具有较高空间相关性达0.53,FCC均值分别为0.58、0.61,主要分布在0.4—0.7,分别占比65.45%、51.79%。研究区森林郁闭度主要处于中度郁闭,北部区域主要为高度郁闭区,与研究区植被覆盖度的空间分布具有一致性,说明使用GEDI数据估测森林郁闭度的方法具有可行性、结果具有可靠性。研究为使用GEDI数据高效、及时、低成本估测大空间尺度的森林水平结构参数的相关研究奠定了基础。

**关键词:**全球生态系统动力学调查(GEDI);贝叶斯优化算法;机器学习;特征变量优选;经验贝叶斯克里金;大空间尺度

## Construction of forest canopy closure estimation model in the northwestern Yunnan based on global ecosystem dynamics investigation multi-beam LiDAR data

ZHOU Wenwu, SHU Qingtai\*, XU Li, YANG Zhengdao, GAO Yingqun, WU Zaikun, XIA Cuifen, GU Chunxi, LI Hua

College of Forestry, Southwest Forestry University, Kunming 650224, China

**Abstract:** Exploring the potential of global ecosystem dynamics investigation (GEDI) multi-beam LiDAR data to estimate regional forest canopy closure (FCC) plays an important role in assessing forest ecosystem status and stand environment.

**基金项目:**国家自然科学基金项目(31860205, 31460194);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)

收稿日期:2023-09-21; 网络出版日期:2024-01-30

\* 通讯作者 Corresponding author. E-mail: shuqt@swfu.edu.cn

The typically ecologically fragile area, Shangri-la, was selected as the study area in the northwestern Yunnan. GEDI waveform data was used as the information source to extract 46245 forest footprints parameters. The empirical Bayesian kriging (EBK) method was used to obtain the continuous distribution of footprints parameters in the unknown space of the study area. Then, combined with 54 measured samples data, the recursive feature elimination method of support vector machine (SVM-RFE), random forest (RF) and Pearson analysis were chosen to optimize the characteristic variables, respectively. The best estimation model of forest canopy closure was studied and constructed by Bayesian optimal random forest regression model (BO-RFR), Bayesian optimal gradient regression model (BO-GBRT), and partial least square method (PLSR). The results showed that: (1) The EBK method had high prediction accuracy and reliable estimation results. Its  $R^2$  was 0.20—0.92, RMSE was 0.004—2812.912, MAE was 0.003—1996.258, and MRE was 0.007—4.423. (2) There were slight differences in the method selection of characteristic variables and number based on different characteristic optimization methods. Among the three methods, the SVM-RFE method selected six parameters (cover, pai, sensitivity, rv\_a1, rv\_a4, rg\_a4) with an average cross-validation accuracy of 0.84. The RF method selected five parameters (cover, pai, pgap\_theta\_error, modis\_treecover, modis\_nonvegetated) with a contribution of 5% as the threshold. The Pearson method significantly selected five parameters (cover, pai, rv\_a5, rg\_a5, pgap\_theta\_error) with a correlation greater than 0.3 and at the 0.01 level. (3) The modeling parameters selected by different characteristic variable optimization methods had great differences in the prediction accuracy of the estimation model. Among them, the accuracy of the estimation models constructed from the parameters selected by the SVM-RFE and RF methods was better, while the accuracy of the estimation models based on the optimization parameters of SVM-RFE method was relatively stable. The BO-GBRT model in the RF method was the best FCC estimation model ( $R^2=0.85$ , RMSE=0.069,  $P=86.5\%$ ). (4) The BO-GBRT model was used to estimate the forest canopy closure and spatial mapping in the study area, which had high spatial correlation with the FCC predicted by the GEDI pai parameter of 0.53, and the mean values of the FCC were 0.58 and 0.61, which were mainly distributed in the range of 0.4—0.7, accounting for 65.45% and 51.79%, respectively. The forest canopy closure in the study area was mainly in moderate canopy closure, and the northern area was mainly in high canopy closure area, which was consistent with the spatial distribution of vegetation coverage in the study area. It indicated that the method of estimating forest canopy closure using GEDI data in this study was feasible and the results were very reliable. Our research laid a foundation for the efficient, timely and low-cost estimation of forest horizontal structural parameters at large spatial scales based on GEDI data.

**Key Words:** global ecosystem dynamics investigation; Bayesian optimization algorithm; machine learning method; characteristic variable optimization; empirical Bayesian kriging; large spatial scale

森林郁闭度(FCC)是指林分中树冠投影面积与林地面积之比<sup>[1]</sup>,作为反映林分结构和林分环境的基本参数<sup>[1-4]</sup>,是林分类型划分、地类划分、小班区划和森林质量评价的主要指标<sup>[4]</sup>,以及评估生态系统健康状态和生物多样性水平的重要生态指标<sup>[2-4]</sup>。传统的森林郁闭度直接测定方法包括树冠投影法、样线法、目测法和照片法等<sup>[1,4]</sup>。这些方法估测 FCC 需要进行大量的野外实地调查,仅能获取以样地水平为单位具有代表性的抽样数据,无法满足大空间尺度森林郁闭度空间分布和变化的研究<sup>[5-6]</sup>。以高质量的标准地采样数据为基础,结合遥感数据高效及时的高精度估测区域尺度或全球尺度的森林郁闭度成为森林资源监测或林业定量遥感反演的优先选项和有效手段<sup>[1,4,6-7]</sup>。

相较于被动遥感技术-光学遥感数据易受光谱饱和和特征影响、易受天气影响及存在难以精准获取森林垂直结构信息的局限<sup>[2,6-9]</sup>,而主动遥感技术-激光雷达(LiDAR)数据不受环境条件限制,具有较强的穿透能力和快速捕获森林植被三维信息能力<sup>[8-11]</sup>。较比机载、地基和背包激光雷达数据的获取范围受限、昂贵且不易获取<sup>[2,9,12]</sup>,星载激光雷达数据具有使用成本低、数据时空分辨率高,抗干扰能力强,能高时效完成不同时空尺度下区域范围内的森林资源连续精准监测<sup>[8-9]</sup>。全球生态系统动力学调查(GEDI)作为最新的星载多波束波形激光雷达<sup>[13-14]</sup>,在森林垂直结构参数反演(森林冠层高<sup>[14-17]</sup>、森林高度制图<sup>[18-19]</sup>等)、森林生物量<sup>[13-14,16,20]</sup>/碳储量<sup>[21-22]</sup>估测、林下地形反演及验证<sup>[15,23]</sup>等方面已有大量研究,成果颇丰。当前,已有大量

研究使用光学遥感数据<sup>[3-6,24-25]</sup>、地基或机载激光雷达数据<sup>[2,7,12]</sup>估测森林郁闭度;少量研究使用合成孔径雷达(SAR)数据<sup>[26-27]</sup>和星载 ICESat-1/2(ice, cloud and land elevation satellite-1/ice, cloud and land elevation satellite-2)数据<sup>[5-6,8,28]</sup>反演 FCC,但 SAR 数据的高散斑噪声和植被间可分离性有限,导致 FCC 预测精度受限<sup>[25]</sup>,较比大光斑全波形激光雷达系统(GLAS)数据,光子计数激光测高仪(ATLAS)数据光斑更小,拥有更高的采样率<sup>[18-19]</sup>,但原始数据的去噪、分类等预处理达到较高精度是一大难点;目前,使用 GEDI 数据估测森林水平结构参数(森林郁闭度、覆盖度等)的相关研究鲜有。GEDI 数据在空间上呈现出条带状的非连续性分布,参数指标需要选用地统计学中的空间插值方法<sup>[13]</sup>或空间回归方法<sup>[29]</sup>预测以获取覆盖整个研究区连续性的面状属性数据<sup>[5-6,13]</sup>,进而实现目标参数的遥感制图。

本研究以“三江”并流核心区香格里拉为实验区,以星载激光雷达 GEDI 数据为信息源提取建模参数,结合 54 块实测样地数据,采用经验贝叶斯克里金法(EBK)法获取 GEDI 参数在研究区未知空间内的连续性分布,使用支持向量机的递归特征消除法(SVM-RFE)、随机森林(RF)和 Pearson 法分别优选特征变量,基于偏最小二乘法(PLSR)、贝叶斯优化随机森林回归模型(BO-RFR)、贝叶斯优化随机梯度回归模型(BO-GBRT)模型研建森林郁闭度遥感估测模型,以反演区域尺度的 FCC 和空间制图;利用 GEDI 的 pai、cover 参数分别绘制森林郁闭度、植被覆盖度的空间分布图。本研究旨在利用经验贝叶斯克里金法获取特征参数在未知区域的空间连续性分布,探索贝叶斯优化(BO)算法对机器学习模型的优化提升能力,探究不同特征变量优选方法及估测模型对 FCC 反演结果的影响,从而探究 GEDI 数据快速估测森林水平结构参数的潜力,为森林生态系统智慧管理和科学规划提供重要参考和技术手段。

## 1 材料与方法

### 1.1 研究区概况

研究区香格里拉市(26°49'7"—28°54'38"N,99°8'2"—100°21'15"E)是“三江并流”核心区(图 1),地形总趋势为西北高、东南低,海拔相对高差达 4042 m,平均海拔 3459 m,多年平均气温介于 4.7—16.5℃,年平均降雨量 649.4 mm,属山地寒温带季风气候<sup>[6]</sup>。研究区介于云南省亚热带常绿阔叶林植被区与青藏高原高寒植被区,低纬高原的地理特征使其南北植被在空间分布上极具垂直分异特征,主要植被类型有冷杉(*Abies fabri*)、高山松(*Pinus densata*)、云杉(*Picea asperata*)、落叶松(*Larix gmelinii*)等寒温带性针叶林,林地总面积占土地面积的 83.3%,森林覆盖率达 76%,是云南省重点保护林区<sup>[6,8-9]</sup>,开展森林郁闭度研究对于低纬度高海拔地区有效保护生态脆弱区极具参考价值。

### 1.2 样地数据

本研究使用的 54 个郁闭度实测数据是在香格里拉市小中甸镇和建塘镇布设 20 m×30 m(约为 0.06 km<sup>2</sup>)的标准样方调查所得(图 1),由于采样对象主要为温性针叶林,为避免雨、雪季节对野外调查带来的困难,最佳采样时间定为 2021 年 11 月。而研究区为低纬度高寒山区,海拔落差大,布设样方时综合考虑交通条件、地形地势、立地条件和林分情况等因素,涵盖了不同坡度、海拔处的主要森林植被类型,包括冷杉、高山松、云杉、落叶松、高山栎(*Quercus semecarpifolia*)、云南松(*Pinus yunnanensis*)等优势树种;然后使用南方测绘 T66 RTK 处于固定解状态下,使用千寻星矩 SR3(Pro 版)差分定位仪器测定样方中心经纬度坐标(含对角)和高程信息,保证误差在 2‰以内后记录;最后利用树冠投影法<sup>[1]</sup>测算 54 个样地的郁闭度值(表 1)。

表 1 郁闭度统计信息描述性分析

Table 1 Descriptive analysis of canopy closure statistics

样本数 Statistical value	最大值 Maximum value	最小值 Minimum value	平均值 Mean value	标准差 Standard deviation	方差 Variance	中位数 Median
54	0.83	0.20	0.50	0.176	0.031	0.5

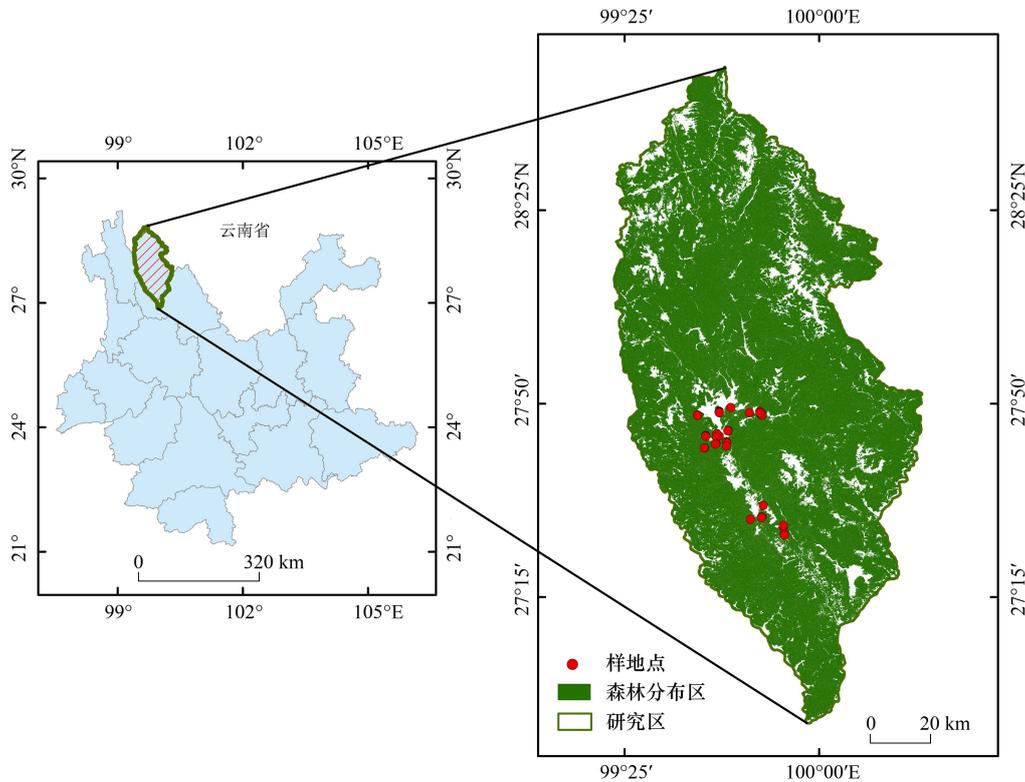


图1 研究区位置、森林分布及样地示意图

Fig.1 Location, forest distribution area and plot diagram of the study area

该图基于审图号为云 S(2019)206 号的标准地图制作,底图无修改

### 1.3 GEDI 数据获取及预处理

#### 1.3.1 GEDI 数据产品

全球生态系统动力学调查研究卫星是一种新的多波束全波形激光雷达传感器<sup>[13–24]</sup>,其搭载在美国国际空间站(ISS)上,由美国国家航天航空局(NASA)于2018年12月5日在美国肯尼迪航天中心成功发射,通过全波形激光探测和测距(激光雷达)激光系统收集全球51.6°N—51.6°S之间的LiDAR数据<sup>[14,17,18–22]</sup>。GEDI系统由3个激光器组成,工作频率为242 HZ,四束波长为1064 nm,其中一束被分成两束能量较弱的激光,共发射8条地面轨道对地观测(4条全功率光束和4条覆盖光束),沿轨间距约为600 m,沿轨纬向间距约为735 m,沿轨大脚印光斑间距约为60 m,脚印光斑间距约为25 m,单个光斑面积约为0.06 km<sup>2</sup>,地理坐标系和高程基准面均为WGS84<sup>[16,18–23]</sup>。GEDI Level 2级产品有Level 2A和Level 2B两类,Level 2A数据产品主要提供地面高程信息及植被的相对高程矩阵信息<sup>[18–19,21–22,30]</sup>;Level 2B数据主要提供森林结构多样性指标,如植被覆盖度、叶面积指数、垂直叶剖面数据等<sup>[21–22,30]</sup>。本文研究对象为森林郁闭度,因此选择GEDI Level 2B Version 2级产品数据,单条数据含8个条带,包括BEAM0000、BEAM0001、BEAM0010、BEAM0011、BEAM0101、BEAM0110、BEAM1000及BEAM1011,数据从Earthdata Search免费获取(<https://www.earthdata.nasa.gov>)<sup>[16,22,30]</sup>。为使GEDI数据均匀覆盖研究区,使用NASA提供的框选工具以香格里拉为边界选取研究区内2021年4月—2021年12月之间所有的GEDI Level 2B数据,含54条数据,216条轨道,432条轨道波束。

#### 1.3.2 GEDI 参数指标提取、筛选及有林地光斑分布

本研究使用Python语言提取和处理LiDAR数据集,共提取43个参数指标,参数简介详见表2和GEDI用户指南<sup>[30]</sup>。首先提取参数shot\_number创建足印标号索引,利用lon\_lowestmod和lat\_lowestmode参数确定光斑的经纬度信息,实现精准定位。而落在研究区内的初始光斑数量达数千万,但部分足印光斑内信息缺失或质

量不佳,研究提取 quality\_flag、degrade\_flag 和 sensitivity 三个参数作为筛选指标优选出高质量足印光斑<sup>[30]</sup>。其中,quality\_flag 用 0 和 1 表示,1 表示足印光斑质量佳须保留,0 则删除;degrade\_flag 也用 0 和 1 表示,1 表示卫星处在降级中,数据性能差须删除,0 则保留;sensitivity 则与地表覆盖类型有关,范围在 0—1,越接近 1 说明植被覆盖度越高,质量越佳,本研究区森林覆盖率高,阈值<sup>[13—24]</sup>设为 sensitivity  $\geq 0.95$ 。通过高标准筛选后留取有效足印光斑 49469 个,再使用 2021 年森林资源林地变更数据对有林地(郁闭度大于 0.2 的为有林地,小于 0.2 为非林地<sup>[1]</sup>)范围内的足印光斑进行空间叠置分析,获得研究区有林地光斑 46245 个(图 2),非林地光斑 3224 个。

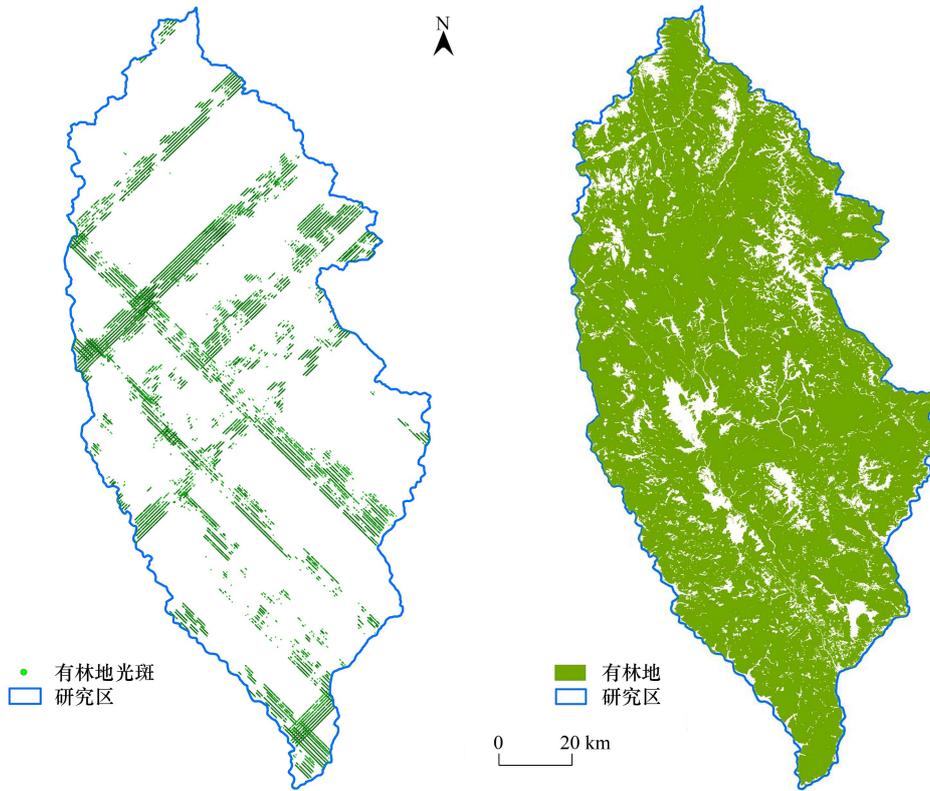


图 2 有林地光斑分布图和有林地林班图

Fig.2 Effective forest land footprints distribution map and forest class map

表 2 GEDI L2B 参数介绍

Table 2 Parameters introduction of GEDI L2B

参数 Parameters	描述 Descriptions	参数 Parameters	描述 Descriptions
lon_lowestmod	最低模式中心的经度	lat_lowestmode	最低模式中心的纬度
quality_flag	光斑质量综合评价指标	modis_nonvegetated	modis 数据的非植被百分比
degrade_flag	卫星是否存在降级地理位置信息	modis_treecover	modis 数据的树木覆盖率
shot_number	光斑足印标号	pgap_theta_aN	森林冠层间隙概率
rx_energy_aN	去除平均噪声后接收到的波形总能量	pgap_theta_error	森林冠层间隙概率的总误差
rv_aN	波形中植被分量的积分	rh100	接收波形信号开始的离地高度
rg_aN	波形中地面分量的积分	height_lastbin	相对森林冠层间隙误差的地面高度
leaf_on_doy	有植被信息的数据	flhd_normal	叶高多样性指数
leaf_off_doy	没有植被信息的数据	sensitivity	灵敏度
pai	植被面积指数	cover	植被总覆盖度

GEDI: 全球生态系统动力学调查;\_aN(N=1—6) 对应 GEDI 的 6 种算法 which means 6 algorithms for GEDI

#### 1.4 研究方法

实测样本数和遥感数据集的选择<sup>[6]</sup>、遥感因子的设置与筛选<sup>[31]</sup>、估测模型的选型及模型参数<sup>[32]</sup>的确定决定了森林结构参数反演结果的精度。为获得最佳的估测模型和结果,本研究选择 EBK 法<sup>[33]</sup>获取 GEDI 的 38 个参数在研究区未知空间的连续性分布,较比普通克里格插值、反距离权重插值、径向基函数插值等,EBK 法插值误差最小、预测精度最高,插值结果最理想<sup>[34-35]</sup>;为了探究不同特征变量优选方法及估测模型对 FCC 反演结果的影响,分别选择 SVM-RFE(通过逐步训练建模来移除权重得分最低的特征变量,以达到从整体上消减冗余特征和降低特征变量之间共线性的目的)<sup>[36-37]</sup>、RF(通过大量决策树得到特征变量因子的重要程度进行综合评分,然后根据其重要性得分排序)<sup>[37]</sup>、Pearson(相关系数的绝对值在 0—1,越靠近 1 说明相关性越强,反之,则越弱)<sup>[38]</sup> 优选出最佳特征变量作为建模参数,基于 PLSR<sup>[39]</sup>、BO-RFR<sup>[40-43]</sup>、BO-GBRT<sup>[40-42,44]</sup> 模型构建最佳 FCC 遥感估测模型。为探究 GEDI 参数直接绘制区域尺度森林郁闭度的潜力,本研究基于比尔定律<sup>[45]</sup>使用森林郁闭度与叶面积指数(LAI)的数学函数关系<sup>[45-47]</sup>,通过组合公式(1)求解 FCC。

$$\begin{cases} \text{LAI} = 93\% \times \text{PAI} \\ \rho(\theta) = e^{-G(\theta) \times \frac{\Omega(\theta) \text{LAI}}{\cos(\theta)}} \\ \text{FCC} = 1 - \rho(\theta) \end{cases} \quad (1)$$

式中,PAI 表示植被面积指数; $\rho(\theta)$  为观测天顶角为  $0^\circ$  时的森林冠层间隙概率; $G(\theta)$  为叶倾角投影系数,取值为 0.5; $\Omega(\theta)$  为聚集指数,取值为 1。

研究为探索贝叶斯优化算法对机器学习方法的优化提升能力,以 RFR 和 GBRT 模型作为基础模型。较比粒子群算法、遗传算法和差分算法,贝叶斯优化算法能够在很少的评估代价下得到一个全局近似最优解<sup>[42]</sup>,因此模型优化模拟次数更少、模型运算速率更快且模型估测精度更佳<sup>[40]</sup>。其算法核心是使用概率模型代理原始评估代价高昂的复杂目标函数<sup>[40-41]</sup>;使用代理模型的后验信息构造主动选择策略,即采集函数<sup>[41-42]</sup>。这就使得概率模型能够更准确地满足黑箱函数的行为,有效地减少了不必要的采样,并在理论上保证最终收敛到全局最优解<sup>[41-42]</sup>,以此减小模型计算量和优化目标模型参数,提高模型估测精度。BO 算法是以“贝叶斯定理”(式 2)为基础的不断迭代优化的过程<sup>[42]</sup>。

$$p(f | D_{1:t}) = \frac{p(D_{1:t} | f) p(f)}{p(D_{1:t})} \quad (2)$$

式中, $f$  表示优化模型中的初始参数; $D_{1:t} = \{(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t)\}$  表示已观测集合, $x_t$  表示决策向量, $y_t = f(x_t) + \varepsilon_t$  表示观测值, $\varepsilon_t$  表示观测误差; $p(D_{1:t} | f)$  表示  $y$  的似然分布; $p(f)$  表示  $f$  的先验概率分布; $p(D_{1:t})$  表示边缘化  $f$  的边际似然分布,在贝叶斯优化中主要用于超参数; $p(f | D_{1:t})$  表示  $f$  的后验概率分布。

贝叶斯优化算法过程共分 6 步(图 3),有 3 个核心步骤:(1)根据最大化采集函数选择出下一个最具“潜力”的评估点  $x_t = \arg\max_{x \in X} \alpha(x | D_{1:t-1})$ ; (2)根据评估点  $x_t$  计算目标函数值  $y_t = f(x_t) + \varepsilon_t$ ; (3)将新获得的输入-观测值对  $\{x_t, y_t\}$  加入到历史观测集中  $D_{1:t-1}$  中,不断更新概率代理模型为下次模型迭代做准备。本研究对 RFR 的 5 个重要参数和 GBRT 的 6 个重要参数(表 3)进行 1000 次范围内的模拟优化,以寻找到最佳模

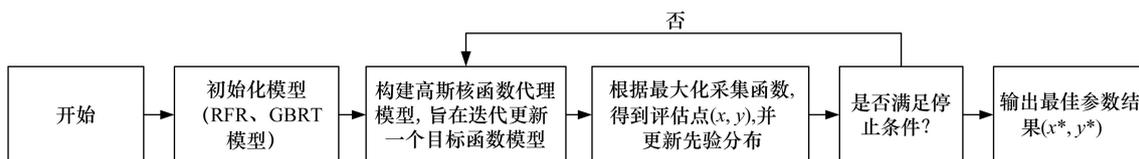


图 3 实现贝叶斯优化随机森林回归模型(BO-RFR)、贝叶斯优化梯度回归模型(BO-GBRT)的算法流程图

Fig.3 Bayesian optimal random forest regression model (BO-RFR) and Bayesian optimal gradient regression model (BO-GBRT) algorithm flow chart

RFR: 随机森林回归模型 Random forest regression model; GBRT: 梯度回归模型 Gradient regression model

拟优化次数和最佳模型参数用于建模。

表 3 随机森林回归 (RFR) 和梯度回归 (GBRT) 模型参数说明

Table 3 Description of random forest regression (RFR) and gradient boosting regression tree (GBRT) model parameters

模型 Model types	参数 Parameters	描述 Descriptions	类型 Types
随机森林回归模型 RFR	max_depth	树深	整数型
梯度回归模型 GBRT	n_estimators	决策树数目	整数型
	min_samples_split	节点可分的最小样本数	整数型或浮点型
	min_samples_leaf	叶子节点包含的最少样本	整数型或浮点型
随机森林回归模型 RFR	max_features	构建最优模型时考虑的最大特征数	整数型或浮点型
梯度回归模型 GBRT	subsample	用于构建最优模型时输出的最佳样本比例	浮点型
	learning_rate	模型学习率	浮点型

RFR: 随机森林回归模型 Random forest regression model; GBRT: 梯度回归模型 Gradient regression model

### 1.5 模型精度评价

本研究使用地统计学中的交叉验证来评估经验贝叶斯克里金插值模型的拟合精度和插值结果的准确性, 该法在数据集中移除一个点, 使用剩余的所有其他点来预测被移除点的位置, 将预测值与实测值进行比较, 并生成大量统计数据来确定预测的准确性<sup>[33]</sup>。本研究采用留一交叉验证法 (LOOCV) 来验证 FCC 遥感估测模型的预测精度和估测结果, LOOCV 对于小样本数据采用逐一参与训练建模和验证, 解决了建模集和验证集相同的问题, 有效避免了拟合模型出现局部最优的问题<sup>[6,9]</sup>, 与 K 折交叉验证相比, 验证结果具有可复制性且不受随机因素的影响, 拥有更强的鲁棒性<sup>[9]</sup>, 可有效解决估测模型过拟合或欠拟合的问题。其中, 以决定系数 ( $R^2$ )、均方根误差 (RMSE)、平均绝对误差 (MAE)、平均相对误差 (MRE) 作为 EBK 法的评估指标; 以  $R^2$ 、RMSE 和总体预测精度 ( $P$ ) 作为 FCC 遥感估测模型的综合评价指标。相关计算公式如下:

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

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

$$\text{MAE} = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (5)$$

$$\text{MRE} = \frac{\sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{N} \quad (6)$$

$$P = \left( 1 - \frac{\text{RMSE}}{\bar{y}} \right) \quad (7)$$

式中,  $y_i$  为郁闭度实测值;  $\hat{y}_i$  为郁闭度预测值;  $\bar{y}$  为模型预测平均值;  $N$  为郁闭度总样本数。

## 2 研究结果与分析

### 2.1 GEDI 参数指标插值结果

本研究采用 EBK 法插值出 GEDI 参数在研究区的未知空间分布, 根据图 4, 38 个参数使用 EBK 方法预测精度  $R^2$  在 0.20—0.92, RMSE 在 0.004—2812.912, MAE 在 0.003—1996.258, MRE 在 0.007—4.423,  $R^2$  和 RMSE

波动较大,MAE 和 RMSE 变化趋势基本一致。其中,modis\_treecover 和 modis\_nonvegetated 参数插值结果最佳,rx\_energy 系列参数的预测精度佳,sensitivity、pai、cover、rh100 等单参数及 rv、rg、pgap 等系列参数的估测精度高,leaf\_off\_doy 和 leaf\_on\_doy 参数的预测精度最低。总体呈现出,不同参数的估测精度和预测结果差异性较大,而同类型参数的估测结果具有一致性和相似性,因为 EBK 法插值尊重原始实测数据值的基础属性,不同类型间的数据分布相对离散,而同类型数据分布相对集中,这体现了数据分布的波动性<sup>[35]</sup>。

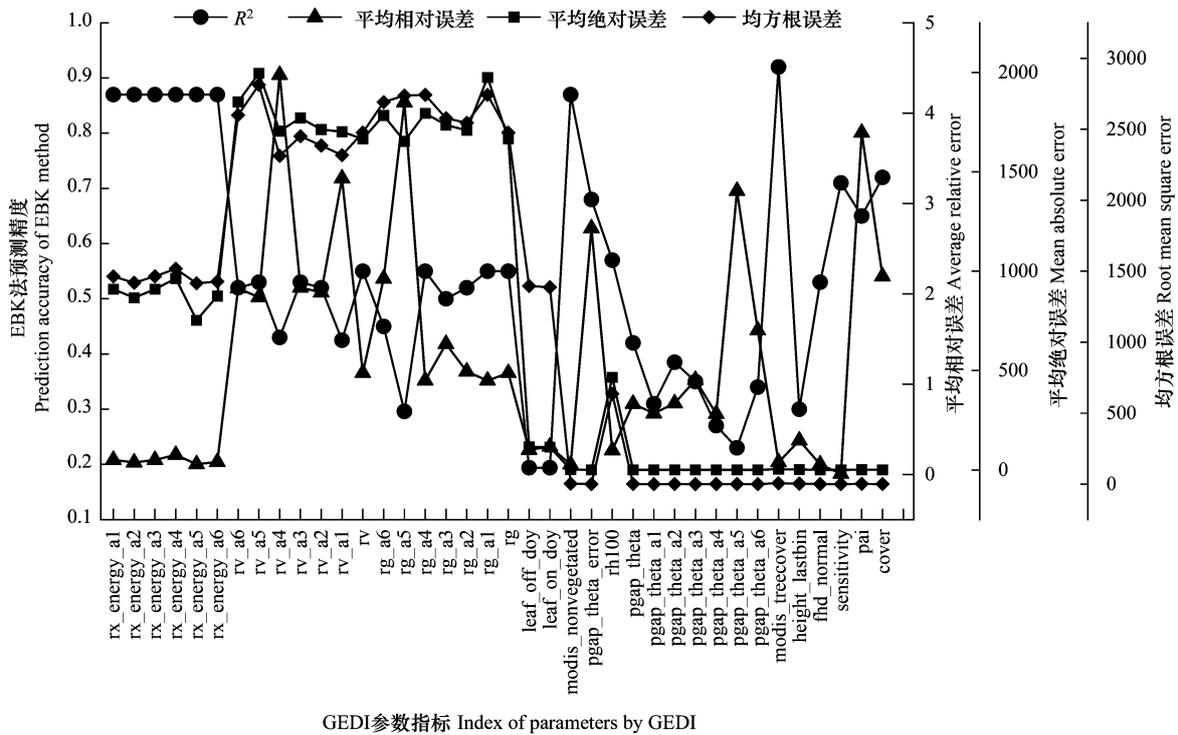


图 4 基于经验贝叶斯克里金插值的相关参数及评价指标

Fig.4 Relevant parameters and evaluation indexes based on empirical Bayesian kriging interpolation

EBK:经验贝叶斯克里金 Empirical Bayesian kriging; GEDI:全球生态系统动力学调查 Global ecosystem dynamics investigation;  $R^2$ :决定系数 Determination coefficient;参数指标描述见表 2

### 2.2 特征变量优选

为优选出最佳建模参数集,本研究采用 Pearson、RF、SVM-RFE 法优选最佳特征变量因子。根据图 5,研究选择相关性较强的参数作为郁闭度估测的自变量,在 38 个参数中,Pearson 系数绝对值在 0.002—0.52,将相关性大于 0.30 设为阈值且在 0.01 水平显著,优选出 5 个特征变量因子作为自变量,分别为 cover、pai、rv\_a5、rg\_a5、pgap\_theta\_error,相关系数从大到小分别为 0.52、0.36、0.30、-0.35、-0.37。

本研究将提取的 38 个 GEDI 参数使用 RF 进行特征重要性评价和排序(图 6),贡献度在 0.20%—22.54%,为筛选出高质量建模参数,研究将 5%设置为阈值,共优选出 5 个参数,特征重要性从高到低分别为 cover、pai、pgap\_theta\_error、modis\_treecover、modis\_nonvegetated,贡献度分别为 22.54%、9.38%、5.88%、5.65%、5.43%。

根据图 7,经过 SVM-RFE 方法特征变量优选,38 个特征参数减少到 6 个(cover、pai、sensitivity、rv\_a1、rv\_a4、rg\_a4),同时 SVM-RFE 交叉验证精度达到 0.84,在保留有效参数组合的情况下极大的降低数据维度,以此优选出最佳建模参数组和提高估测模型精度。

根据三种特征变量优选方法筛选建模参数结果发现,使用不同的特征变量优选方法筛选出的建模参数及数量略有差异,但 cover、pai 参数是共有参数,与 FCC 有着紧密关系。cover 参数表示植被覆盖度(FVC),在林

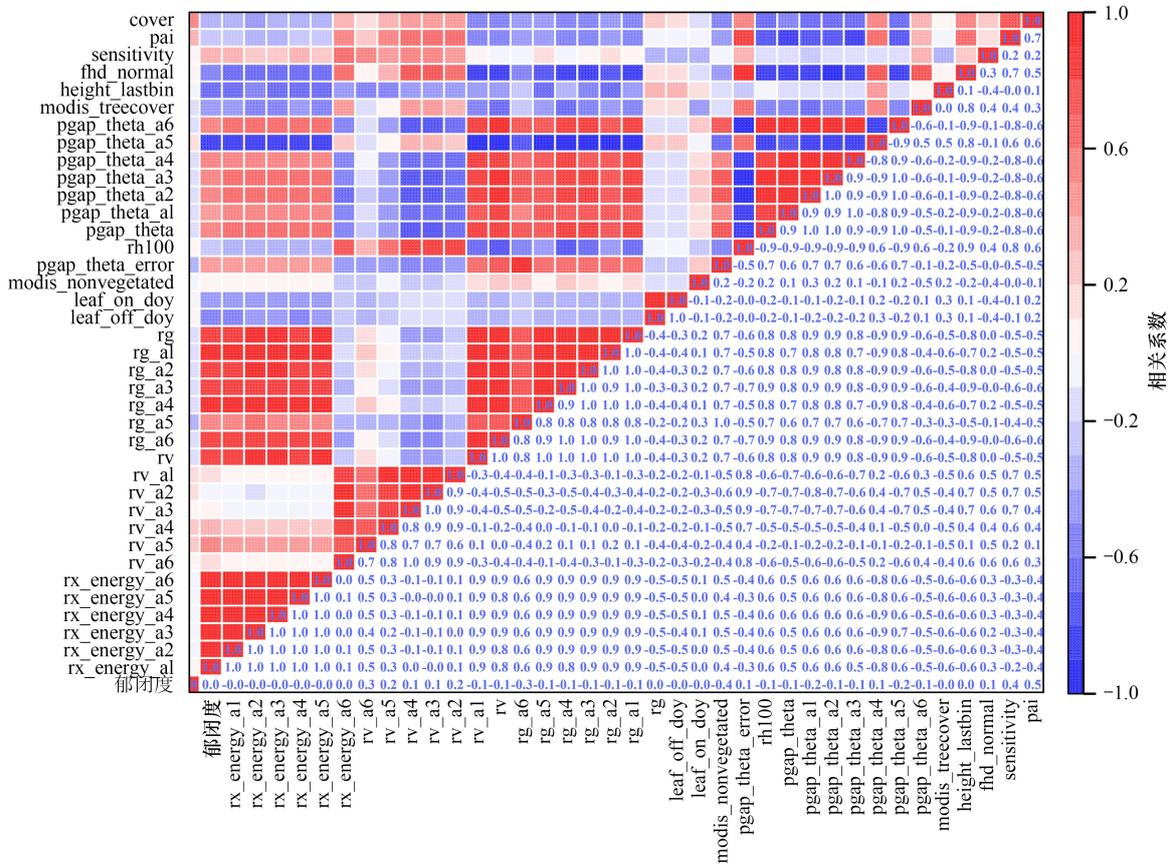


图5 GEDI各参数与郁闭度的相关系数矩阵

Fig.5 The correlation coefficient matrix of GEDI parameters and canopy closure

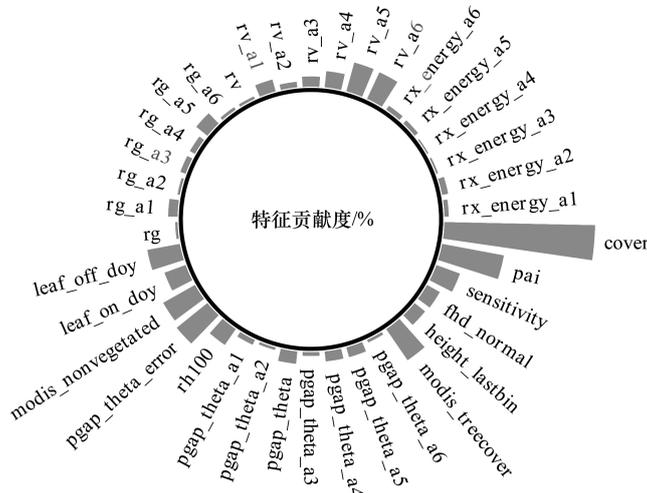


图6 GEDI各参数特征重要性贡献比

Fig.6 The characteristic importance contribution ratio of GEDI parameters

分中常用郁闭度的概念表示植被的覆盖情况,郁闭度近似等于覆盖度<sup>[48]</sup>;pai 参数表示植被面积指数,通过式(1)可求解出 FCC<sup>[45-47]</sup>。

### 2.3 遥感建模结果

使用 Pearson、RF 和 SVM-RFE 方法分别优选出不同数量的最优特征变量因子研建 FCC 估测模型。经过 1000 次参数寻优,确定 100 次、300 次分别为 BO-RFR、BO-GBRT 模型的最佳模拟优化次数。根据表 4 和图 8, RFR 和 GBRT 模型初始估测精度 ( $R^2$  为 0.23—0.35、RMSE 为 0.1533—0.1404、 $P$  为 69.86%—71.06%) 与 PLSR 模型的预测精度 ( $R^2$  为 0.20—0.28、RMSE 为 0.1581—0.1503、 $P$  为 68.34%—69.82%) 基本一致,说明了留一交叉验证法评估模型的精度更严格、具有更强的鲁棒性<sup>[9]</sup>,但预测结果准确性较低,还有待提升;使用 BO 算法改进 RFR 和 GBRT 模型后,模型精度有效提升,BO-RFR 和 BO-GBRT 模型的  $R^2$ 、RMSE 和  $P$  分别在 0.70—0.85、0.1006—0.069、80.63%—86.50%,较比未优化前  $R^2$  平均提高 0.48, RMSE 平均降低 0.06,  $P$  平均提高 12.42%。

根据表 4 和图 8,从纵向分析,同种特征变量优选方法筛选出的变量对构建不同估测模型精度会产生显著影响,基于优化算法改进后的机器学习模型 (BO-GBRT、BO-RFR) 比传统统计回归模型 (PLSR) 估测精度更高,预测结果更准确,其中 BO-GBRT 模型在总体精度上更佳;从横向分析,不同特征变量优选方法筛选出的变量对构建相同估测模型精度略有差异,较比传统方法 (Pearson),采用机器学习方法 (RF、SVM-RFE) 筛选的参数构建估测模型精度更佳。本研究采用 RF 优选的 5 个特征变量研建的 BO-GBRT 模型作为 FCC 最佳估测模型,  $R^2$  最大为 0.85、RMSE 最小为 0.069,  $P$  最高为 86.5%。

表 4 郁闭度初始估测模型研建结果

Table 4 The results of the initial estimation model of canopy closure

模型 Model	特征变量优选方法 Optimal selection method of characteristic variables	决定系数 $R^2$	均方根误差 RMSE	总体预测精度 $P/\%$
梯度回归模型 GBRT	Pearson	0.23	0.1533	69.88
	RF	0.35	0.1404	71.06
	SVM-RFE	0.28	0.1495	70.27
随机森林回归模型 RFR	Pearson	0.25	0.1512	69.86
	RF	0.26	0.1504	70.01
	SVM-RFE	0.30	0.1462	70.68

RF: 随机森林 Random forest; SVM-RFE: 支持向量机的递归特征消除法 The recursive feature elimination method of support vector machine;  $R^2$ : 决定系数 Determination coefficient; RMSE: 均方根误差 Root mean square error;  $P$ : 总体预测精度 Overall prediction accuracy

### 2.4 研究区森林郁闭度的空间分布

根据林分密度划分标准<sup>[1]</sup>,郁闭度在 0.7 以上为密,0.4—0.7 为中,0.2—0.4 为疏。本研究采用 FCC 最佳估测模型 (BO-GBRT)、式 (1) 预测研究区森林郁闭度,使用 EBK 法预测植被覆盖度的空间分布 (图 9),总体呈现出:研究区内植被覆盖总体水平较高,北部区域为 FCC 和 FVC 高值主要分布区,FCC 高度郁闭区由东南向北贯穿分布,中度郁闭区占比最大,低度郁闭区主要分布在东南部城镇区域、河流及雪山附近等人口聚集区或边缘地区,存在总体分布不均、区域差异较大的现象。其中,基于 BO-GBRT 模型估测 FCC 并进行空间制图,均值为 0.58,主要为中度郁闭,占比 65.45%,其次是高度郁闭,占比 28.42%,最后是低度郁闭,占比 6.13%;

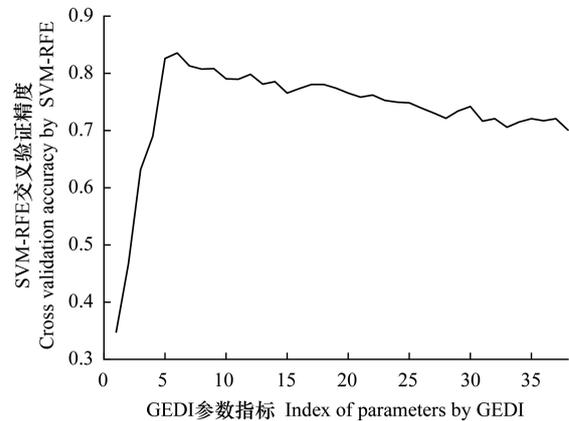


图 7 基于 SVM-RFE 的特征变量优选

Fig.7 Feature variable preferred selection based on SVM-RFE method

SVM-RFE: 支持向量机的递归特征消除法 The recursive feature elimination method of support vector machine

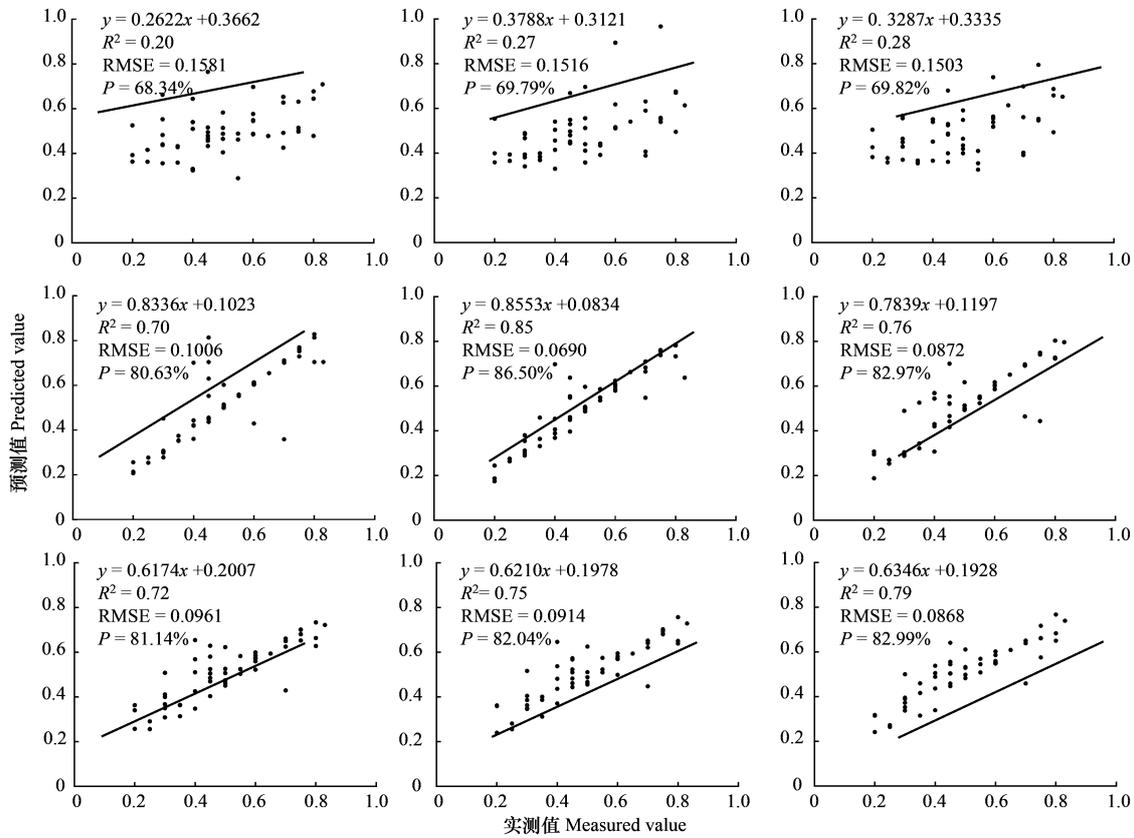


图8 基于不同特征变量优选方法下的估测模型研建结果

Fig.8 The results of the estimation model based on different characteristic variable optimization methods

纵向:从左至右为 Pearson、RF、SVM-RFE;横向:从上往下为 PLSR、BO-GBRT、BO-RFR;P: 总体预测精度 Overall prediction accuracy; R<sup>2</sup>: 决定系数 Determination coefficient;RMSE: 均方根误差 Root mean square error

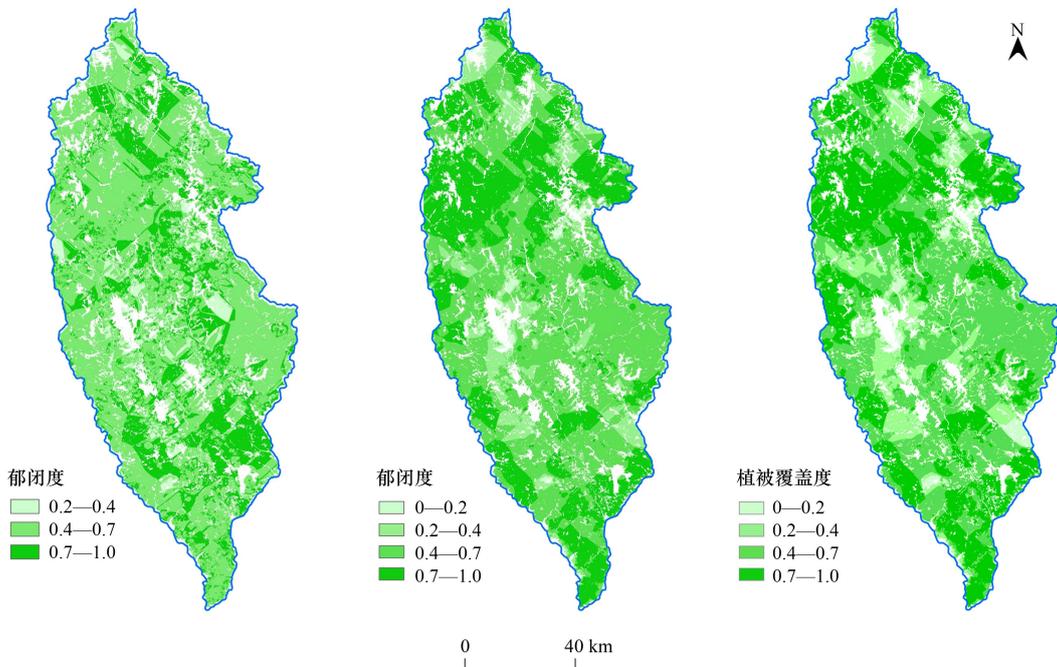


图9 基于不同方法预测研究区森林郁闭度和覆盖度的空间分布

Fig.9 The spatial distribution of forest canopy density and forest vegetation coverage in the study area was predicted based on different methods

从左至右为:BO-GBRT、式(1)、EBK

使用式(1)求解 FCC 并进行空间制图,均值为 0.61,主要为中度郁闭,占比 51.79%,其次是高度郁闭,占比 35.53%,最后是低度郁闭,占比 11.04%,较比前者,中度郁闭区间减少,高、低度郁闭区间增加;利用研究区有林地光斑内的 cover 参数基于 EBK 法绘制 FVC 的空间分布图,均值为 0.62,主要分布在 0.4—0.7,占比 47.26%,其次是 0.7—1.0,占比 37.92%,最少是 0.2—0.4 和 0—0.2,分别占 12.11%、2.71%。对比周文武等<sup>[6]</sup>、魏治越<sup>[8]</sup>等研究结果,研究区 FCC 总体分布趋势与之基本一致,主要分布区间基本相同,与森林资源分布基本吻合,但采用式(1)预测的结果与之相比,FCC 中值区域明显降低,高值和低值区域有增加趋势。根据图 9,BO-GBRT 模型估测的 FCC 空间分布图与式(1)预测的 FCC 空间分布图、植被覆盖度空间分布图的空间相关性分别为 0.53、0.50,且后两者的空间相关性达 0.97,具有较高的一致性,相互印证估测结果具有一定的可靠性;研究使用 2021 年森林资源林地变更数据裁剪出研究区有林地 FCC 和 FVC 的空间分布,所以在三个预测结果中,分布在 0—2 区间的占比极少或没有,该法为利用 GEDI 参数快速获取大尺度区域的 FCC 和 FVC 提供一种参考。

### 3 讨论

#### 3.1 EBK 法插值结果分析

森林郁闭度作为评估生态系统健康状态和生物多样性水平的重要生态指标,本研究利用 GEDI 数据基于插值的方法探索了一种更加及时、高效、低成本、高精度获取区域尺度 FCC 的方法,突破了 GEDI 激光雷达数据分布不连续,需要联合多源遥感数据才能获取区域尺度内连续性面状属性产品的局限,该方法为大尺度或全球范围内刻画森林郁闭度提供了一种新视野。空间插值就是利用已知位置的空间样本数据预测未知地理空间的特征值<sup>[13,34]</sup>,与普通克里格、协同克里格等传统的地统计插值方法相比,经验贝叶斯克里金通过估计最佳半变异函数作为未知区域的半变异函数进行目标参数的插值预测<sup>[34,49]</sup>,即通过对半变异函数的基础预测模拟,计算拟合度,对于空间分布不连续数据或空间突变数据的预测稳定性及精度都较高<sup>[35,49]</sup>,这为充分利用 GEDI 高密度、高精度光斑数据插值预测大空间尺度的 FCC 提供便利。研究表明,GEDI 各参数预测精度佳, $R^2$ 最高达 0.92,最低也有 0.2, RMSE 在 0.004—2812.912, MAE 在 0.003—1996.258, MRE 在 0.007—4.423;采用 Pearson、RF 和 SVM-RFE 方法分别优选出的建模参数的 EBK 法预测精度高( $R^2$ 为 0.43—0.92、RMSE 为 0.008—2812.912、MAE 为 0.007—1996.258、MRE 为 0.007—4.12),这在很大程度上降低了误差传递,与张仁平等<sup>[34]</sup>、陈海生等<sup>[35]</sup>、高浩然等<sup>[49]</sup>研究结果具有一致性。较比 Xu 等<sup>[13]</sup>研究结果,EBK 法预测精度明显比普通克里格插值精度高,EBK 法预测结果的条带效应较弱,这与 GEDI 数据沿轨道均匀分布有关<sup>[19—23]</sup>,也符合地理学第一定律(空间自相关性)及第二定律(空间异质性),但随着预测精度提高,条带效应随之减弱。为增强不同空间插值方法间的对比性以进一步提高空间插值预测精度,可选择 ANUSPLIN 软件法、三次样条法或“3G”(即“GIS&GP(遗传规划算法)&GA(遗传算法)”)方法进行预测对比<sup>[34]</sup>;为进一步减弱条带效应,对同一条带或相邻条带上的光斑数据以 100 m 为分段进行抽稀<sup>[13]</sup>,但这样会丢失大量的样本数据导致插值精度下降,会影响估测结果精准性,相比之下,本研究选择高标准筛选高质量光斑数据用以提高插值精度<sup>[16,23]</sup>,以此减弱条带效应的方法更可取。

#### 3.2 特征变量优选

特征变量因子的优选与建模参数的组合对于估测模型和预测结果的精准性具有至关重要的作用<sup>[6,31]</sup>。相比 GEDI Level 2A 数据产品,Level 2B 数据产品主要提供了冠层覆盖度、叶面积指数和垂直叶剖面数据等<sup>[21—22,30]</sup>,以此作为数据源提取特征变量估测森林郁闭度更具有合理性,预测结果更佳。根据 Pearson、RF 和 SVM-RFE 方法优选结果可知,虽然不同的特征变量优选方法筛选出的建模参数及数量都会有差异,但对目标参数解释性更强,对模型贡献度更大的特征变量因子都会被保留,cover 和 pai 参数作为保留的共有特征变量,两者通过比尔定律的数学形式变换计算或定义泛化均可得到郁闭度<sup>[48]</sup>,与郁闭度具有较强相关性和较高的重要性。采用三种方法分别优选出不同数量的最优特征变量构建最佳 FCC 估测模型,较比 Pearson 方法,

以 RF 和 SVM-RFE 方法优选的参数构建遥感估测模型精度更佳,以 SVM-RFE 方法优选的参数构建估测模型精度变化相对 RF 方法更小,因为 SVM-RFE 可以从整体上消减冗余特征和降低特征变量之间的共线性,使得估测模型具有更强的模型泛化能力和稳定性<sup>[36-37]</sup>;但在 RF 方法中寻到了最佳 FCC 估测模型( $R^2 = 0.85$ 、 $RMSE = 0.069$ ,  $P = 86.5\%$ ),这可能与 RF 方法将特征变量因子进行重要性排序时,在有效消除整体冗余特征的同时具有随机性有关<sup>[37]</sup>,说明这一组参数适用于 BO-GBRT 模型。下一步可增加 Boruta 算法、快速迭代特征选择的 K-NN(KNN-FIFS)模型等<sup>[31]</sup>方法筛选特征变量,尝试优选出更佳建模参数以提高估测模型预测精度<sup>[6]</sup>。

### 3.3 模型误差传递及贝叶斯优化算法

本研究在尺度转化过程中,存在插值方法和机器学习模型间的误差传递,实测数据样本量、遥感模型及模型参数的不确定性是影响估测模型精度的重要因素,而预测模型精度对于估测结果的精准性具有至关重要的作用<sup>[32]</sup>。由于模型的代表性与建模样本数有关,建模样本数越多,估测模型越具有代表性,同时不确定也会随之降低,但随着模型样本数量的增加,达到一定临界值时,再增加样本量,估测模型精度也不再发生明显变化,因此为了节省人力物力财力,同时满足野外调查大样本原则(50 个)和精度需要<sup>[3,9]</sup>,本研究调查了 54 块实测样地数据用以构建模型,但研究样地分布相对集中,可能会增大预测结果的不确定性。其中,郁闭度分布在 0.2—0.4、0.4—0.7、0.7—0.9 区间的样本数分别为 14、28、12,呈现出正态分布;郁闭度在坡度为 0°—10°、10°—20°、>20°的样本数分别占比 42.6%、29.6%、27.8%,坡度分布相对均匀。由此,说明了采样的代表性与建模结果的合理性,以此降低了采样带来的不确定性和误差传递。为了进一步减弱模型误差传递对估测结果的影响,研究选择贝叶斯优化算法对 RFR、GBRT 模型进行性能优化提升,在同一非参数优化模型中,BO 算法比粒子群优化算法、遗传算法、差分算法更能以较少的优化次数达到更高的模型运算速率和模型估测精度<sup>[41-42]</sup>。结果表明,RFR 模型和 GBRT 模型初始预测精度与 PLSR 模型精度具有一致性,使用 BO 算法可有效提高机器学习模型的预测精度,提高估测结果精准性,较比未优化前,优化后的 RFR、GBRT 模型, $R^2$ 平均提高 0.48, RMSE 平均降低 0.06,  $P$  平均提高 12.42%;较比固定参数次数对基础模型进行模拟优化提升<sup>[6,40]</sup>,本研究选择在 1000 次范围内寻找 BO-RFR(100)、BO-GBRT(300)模型的最佳模拟优化次数,以较少的优化次数取得了较高的模型估测精度,极大地节省了模拟优化时间成本;研究选择 BO-GBRT( $R^2 = 0.85$ 、 $RMSE = 0.069$ ,  $P = 86.5\%$ )作为 FCC 最佳估测模型以反演研究区森林郁闭度,BO-GBRT 模型是基于原有模型不断进行迭代改进后使得下次新模型较比上次模型产生更小误差,并朝着残差减小的梯度方向构建新的组合模型<sup>[6,43,50]</sup>。为了进一步提高估测模型精度,相较于只优化 RFR 和 GBRT 模型的主要四个参数<sup>[6,40-41]</sup>,本研究在此基础上拓展了主要优化参数,逐渐向全模型参数优化延伸;研究仅在 1000 次范围内寻找最佳模拟优化次数,如果不计时间成本,后期可尝试扩大范围搜索,或可引入深度森林算法让小样本数据也可进行深度神经网络学习拟合<sup>[51]</sup>。本研究中未对 GEDI 强弱波束进行区分研究,而在已有研究中,使用夜间强波束反演林下地形精度高于覆盖波束<sup>[23]</sup>,鉴于此,未来可区分强、弱波束以探究对 FCC 估测精度的影响。

### 3.4 基于 GEDI 数据绘制大空间尺度森林郁闭度的展望

本文建立了一套从源数据预处理和特征变量集成化提取、多种特征变量优选、多种非参数模型主要参数优化及拟合最佳模型的流程化模块处理系统,研究者只需要输入实测样本数据和按规定将预处理好的遥感数据进行建模,并根据自身需求增加或挑选出最佳估测模型后即可得到研究区内森林郁闭度。而且 GEDI 收集了 51.6°N—51.6°S 之间的所有 LiDAR 数据<sup>[14,17,18-22]</sup>,覆盖了全球主要的森林植被区,不仅可以满足不同研究区选取的灵活性以达到模型可移植性检验的需求<sup>[11,18-19]</sup>,还可为大区域或全球范围内刻画森林郁闭度提供数据源。本研究基于叶面积指数与森林郁闭度之间的数学函数关系<sup>[45-47]</sup>预测了研究区内的 FCC,均值为 0.61,主要分布在 0.4—0.7,占比 51.79%,呈现出中度郁闭,与 BO-GBRT 模型估测 FCC 的空间相关性达 0.53,和植被覆盖度之间的空间相关性达 0.97,使用该法快速预测 FCC 结果具有一定可靠性,这为大尺度及时、高效、低成本的估测 FCC 提供了一种参考,也为绘制全球主要的森林植被区覆盖度的空间分布提供了可能,这

有利于满足大空间尺度生态系统健康状态的评估和生物多样性水平的研究,为后续使用 GEDI 数据估测大区域的森林水平结构参数的相关研究奠定基础。

#### 4 结论

为评估多波束激光雷达数据估测森林郁闭度的能力,本研究以 GEDI 数据为信息源提取建模参数,结合 54 块实测样地数据,采用 EBK 法获取特征变量在研究区未知空间内的连续性分布,使用三种特征变量优选方法(SVM-RFE、RF、Pearson)和三种模型(PLSR、BO-RFR、BO-GBRT)研建森林郁闭度遥感估测模型。结果显示:使用经验贝叶斯克里金法将 GEDI 数据实现由“点”到“面”的预测,EBK 法预测精度佳,估测结果相对可靠,解决了空间尺度转移的问题;不同的特征变量优选方法筛选出的建模参数及数量都有差异,cover、pai 参数为共有参数,与 FCC 有着紧密关系;使用贝叶斯优化算法对机器学习模型主要参数进行改进优化可有效提高 FCC 估测模型精度,能有效降低模型误差传递,以 RF 和 SVM-RFE 方法优选的参数构建遥感估测模型精度更佳,SVM-RFE 方法筛选的参数研建估测模型精度变化相对稳定,以 RF 方法中估测精度最佳的 BO-GBRT 模型( $R^2=0.85$ 、 $RMSE=0.069$ 、 $P=86.5\%$ )预测研究区的森林郁闭度及空间制图,与 GEDI pai 参数预测的 FCC 具有较高的空间相关性。研究区森林郁闭度主要处于中度郁闭,高度郁闭区由东南向北贯穿分布,北部为主要分布区,FCC 空间分布与当地林木资源分布情况基本一致,因此使用 GEDI 数据估测森林郁闭度的方法具有可行性、结果具有可靠性。本研究提出的方法为快速遥感估测大空间尺度的森林水平结构参数奠定了基础,有利于满足大空间尺度生态系统健康状态的评估和生物多样性水平的研究。

#### 参考文献(References):

- [1] 孟宪宇. 测树学. 3 版. 北京: 中国林业出版社, 2006: 76-77.
- [2] Pu Y H, Xu D D, Wang H B, An D S, Xu X A. Extracting canopy closure by the CHM-based and SHP-based methods with a hemispherical FOV from UAV-LiDAR data in a poplar plantation. *Remote Sensing*, 2021, 13(19): 3837.
- [3] Hua Y Y, Zhao X S. Multi-model estimation of forest canopy closure by using red edge bands based on sentinel-2 images. *Forests*, 2021, 12(12): 1768.
- [4] 张瑞英. 基于多源遥感数据的森林郁闭度估测方法研究[D]. 呼和浩特: 内蒙古师范大学, 2015.
- [5] 王蕊, 邢艳秋, 王立海, 尤号田, 邱赛, 王爱娟. 联合星载 ICESat-GLAS 波形与多光谱 Landsat-TM 影像的森林郁闭度估测. *应用生态学报*, 2015, 26(6): 1657-1664.
- [6] 周文武, 舒清态, 王书伟, 杨正道, 罗绍龙, 胥丽, 肖劲楠. 基于多源遥感数据协同的滇西北森林郁闭度估测. *应用生态学报*, 2023, 34(7): 1806-1816.
- [7] Gao T, Gao Z H, Sun B, Qin P Y, Li Y F, Yan Z Y. An integrated method for estimating forest-canopy closure based on UAV LiDAR data. *Remote Sensing*, 2022, 14(17): 4317.
- [8] 魏治越, 李浩, 舒清态, 席磊, 宋涵玥, 邱霜, 杨泽至. 基于星载激光雷达 ICESat-2/ATLAS 数据的森林郁闭度估测研究. *西南林业大学学报: 自然科学版*, 2024, 44(2): 1-9.
- [9] Song H Y, Xi L, Shu Q T, Wei Z Y, Qiu S. Estimate forest aboveground biomass of mountain by ICESat-2/ATLAS data interacting cokriging. *Forests*, 2022, 14(1): 13.
- [10] Lin X J, Xu M, Cao C X, Dang Y F, Bashir B, Xie B, Huang Z B. Estimates of forest canopy height using a combination of ICESat-2/ATLAS data and stereo-photogrammetry. *Remote Sensing*, 2020, 12(21): 3649.
- [11] Zhu X X, Nie S, Wang C, Xi X H, Li D, Li G Y, Wang P, Cao D, Yang X B. Estimating terrain slope from ICESat-2 data in forest environments. *Remote Sensing*, 2020, 12(20): 3300.
- [12] 段祝庚, 吴凌霄, 江学良. 无人机激光雷达点云密度对森林遥感反演指数提取的影响. *武汉大学学报: 信息科学版*, 2022, 47, doi: 10.13203/j.whugis20210719.
- [13] Xu L, Shu Q T, Fu H Y, Zhou W W, Luo S L, Gao Y Q, Yu J G, Guo C S, Yang Z D, Xiao J N, Wang S W. Estimation of *Quercus* biomass in Shangri-La based on GEDI spaceborne lidar data. *Forests*, 2023, 14(5): 876.
- [14] 韩明辉. 基于星载激光雷达 GEDI 数据反演森林结构参数的研究[D]. 哈尔滨: 东北林业大学, 2022.
- [15] 刘丽娟, 王成, 聂胜, 朱笑笑, 习晓环, 王金亮. GEDI L2A 不同算法对地面高程和森林冠层高度精度的影响分析. *中国科学院大学学报*, 2022, 39(4): 502-511.
- [16] 韩明辉, 邢艳秋, 李国元, 黄佳鹏, 蔡龙涛. 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. *中南林业科技大学学报*, 2022, 42(10): 72-82.
- [17] Ghosh S M, Behera M D, Kumar S, Das P, Prakash A J, Bhaskaran P K, Roy P S, Barik S K, Jeganathan C, Srivastava P K, Behera S K.

- Predicting the forest canopy height from LiDAR and multi-sensor data using machine learning over India. *Remote Sensing*, 2022, 14(23): 5968.
- [18] 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.
- [19] Zhu X X, Nie S, Wang C, Xi X H, Lao J Y, Li D. Consistency analysis of forest height retrievals between GEDI and ICESat-2. *Remote Sensing of Environment*, 2022, 281: 113244.
- [20] Chen L, Ren C Y, Bao G D, Zhang B, Wang Z M, Liu M Y, Man W D, Liu J F. Improved object-based estimation of forest aboveground biomass by integrating LiDAR data from GEDI and ICESat-2 with multi-sensor images in a heterogeneous mountainous region. *Remote Sensing*, 2022, 14(12): 2743.
- [21] Liang M Y, González-Roglich M, Roehrdanz P, Tabor K, Zvoleff A, Leitold V, Silva J, Fatoyinbo T, Hansen M, Duncanson L. Assessing protected area's carbon stocks and ecological structure at regional-scale using GEDI lidar. *Global Environmental Change*, 2023, 78: 102621.
- [22] Crockett E T H, Atkins J W, Guo Q F, Sun G, Potter K M, Ollinger S, Silva C A, Tang H, Woodall C W, Holgerson J, Xiao J F. Structural and species diversity explain aboveground carbon storage in forests across the United States: Evidence from GEDI and forest inventory data. *Remote Sensing of Environment*, 2023, 295: 113703.
- [23] 董瀚元, 于颖, 范文义. 星载激光雷达 GEDI 数据林下地形反演性能验证. *南京林业大学学报: 自然科学版*, 2023, 47(2): 141-149.
- [24] 杨存建, 倪静, 周其林, 程武学, 韩沙鸥. 不同林分郁闭度与遥感数据的相关性. *生态学报*, 2015, 35(7): 2119-2125.
- [25] Xie B, Cao C X, Xu M, Yang X W, Duerler R S, Bashir B, Huang Z B, Wang K M, Chen Y Y, Guo H Y. Improved forest canopy closure estimation using multispectral satellite imagery within google earth engine. *Remote Sensing*, 2022, 14(9): 2051.
- [26] 宁凯. 基于 SAR 影像反演山区森林郁闭度的方法研究[D]. 成都: 西南交通大学, 2014.
- [27] Cui L, Jiao Z T, Zhao K G, Sun M, Dong Y D, Yin S Y, Zhang X N, Guo J, Xie R, Zhu Z D, Li S J, Tong Y D. Retrieving forest canopy elements clumping index using ICESat GLAS LiDAR data. *Remote Sensing*, 2021, 13(5): 948.
- [28] 邱赛, 邢艳秋, 田静, 丁建华. 基于 ICESat-GLAS 波形数据估测森林郁闭度. *南京林业大学学报: 自然科学版*, 2016, 40(5): 99-106.
- [29] 李明泽, 郭鸿郡, 范文义, 甄贞. 基于 GWR 的大兴安岭森林立地质量遥感分析. *林业科学*, 2017, 53(6): 56-66.
- [30] Beck J, Wirt B, Armston J, Hofton M, Lutchke S, Tang H, Blair J B, Dubayah R. Global ecosystem dynamics investigation (GEDI) level 2 user guide Version 2. Washington: University of Maryland, 2021.
- [31] Zhang W F, Zhao L X, Li Y, Shi J M, Yan M, Ji Y J. Forest above-ground biomass inversion using optical and SAR images based on a multi-step feature optimized inversion model. *Remote Sensing*, 2022, 14(7): 1608.
- [32] 秦立厚, 张茂震, 钟世红, 于晓辉. 森林生物量估算中模型不确定性分析. *生态学报*, 2017, 37(23): 7912-7919.
- [33] Berger J O, De Oliveira V, Sansó B. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 2001, 96(456): 1361-1374.
- [34] 张仁平, 张云玲, 郭靖, 冯琦胜, 梁天刚. 新疆地区降水分布的空间插值方法比较. *草业科学*, 2018, 35(3): 521-529.
- [35] 陈海生, 金玮佳. 基于经验贝叶斯克里金的微尺度植烟田土壤有机质空间变异性. *西南农业学报*, 2020, 33(2): 363-368.
- [36] 张睿, 马建文. 一种 SVM-RFE 高光谱数据特征选择算法. *武汉大学学报: 信息科学版*, 2009, 34(7): 834-837.
- [37] 吴辰文, 梁靖涵, 王伟, 李长生. 基于递归特征消除方法的随机森林算法. *统计与决策*, 2017, (21): 60-63.
- [38] Duncanson L, Neuenschwander A, Hancock S, Thomas N, Fatoyinbo T, Simard M, Silva C A, Armston J, Lutchke S B, Hofton M, Kellner J R, Dubayah R. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sensing of Environment*, 2020, 242: 111779.
- [39] Lei C L, Ju C Y, Cai T J, Jing X, Wei X H, Di X Y. Estimating canopy closure density and above-ground tree biomass using partial least square methods in Chinese boreal forests. *Journal of Forestry Research*, 2012, 23(2): 191-196.
- [40] 仇文岗, 唐理斌, 陈福勇, 杨甲锋. 基于 4 种超参数优化算法及随机森林模型预测 TBM 掘进速度. *应用基础与工程科学学报*, 2021, 29(5): 1186-1200.
- [41] Cho H H, Kim Y J, Lee E J, Choi D Y, Lee Y J, Rhee W J. Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks. *IEEE Access*, 2020, 8: 52588-52608.
- [42] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述. *软件学报*, 2018, 29(10): 3068-3090.
- [43] Breiman L. Random forests. *Machine Learning*, 2001, 45: 5-32.
- [44] Friedman J H. Greedy function approximation; a gradient boosting machine. *The Annals of Statistics*, 2001, 29(5): 1189-1232.
- [45] Nilson T. Inversion of gap frequency data in forest stands. *Agricultural and Forest Meteorology*, 1999, 98/99: 437-448.
- [46] Tang H, Dubayah R, Swatantran A, Hofton M, Sheldon S, Clark D B, Blair B. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sensing of Environment*, 2012, 124: 242-250.
- [47] Armston J D, Scarth P F, Phinn S R, Danaher T J. Analysis of multi-date MISR measurements for forest and woodland communities, Queensland, Australia. *Remote Sensing of Environment*, 2007, 107(1/2): 287-298.
- [48] 梁顺林, 李小文, 王锦地. 定量遥感: 理念与算法. 2 版. 北京: 科学出版社, 2019: 478-479.
- [49] 高浩然, 周勇, 刘甲康, 程晓明, 郭嵩, 江衍, 谭恒鑫. 基于 EBK 插值预测和 GDM 模型的襄州区耕地土壤重金属时空分布及来源变化分析. *环境科学*, 2022, 43(11): 5180-5191.
- [50] Zhang J L, Lu C, Xu H, Wang G X. Estimating aboveground biomass of *Pinus densata*-dominated forests using Landsat time series and permanent sample plot data. *Journal of Forestry Research*, 2019, 30(5): 1689-1706.
- [51] 夏恒, 汤健, 乔俊飞. 深度森林研究综述. *北京工业大学学报*, 2022, 48(2): 182-196.