

DOI: 10.5846/stxb202201140141

谭雨欣, 田义超, 黄卓梅, 张强, 陶进, 刘虹秀, 杨永伟, 张亚丽, 林俊良, 邓静雯. 北部湾茅尾海无瓣海桑红树林地上生物量反演——基于 XGBoost 机器学习算法. 生态学报, 2023, 43(11): 4674-4688.

Tan Y X, Tian Y C, Huang Z M, Zhang Q, Tao J, Liu H X, Yang Y W, Zhang Y L, Lin J L, Deng J W. Aboveground biomass of *Sonneratia apetala* mangroves in Mawei Sea of Beibu Gulf based on XGBoost machine learning algorithm. Acta Ecologica Sinica, 2023, 43(11): 4674-4688.

# 北部湾茅尾海无瓣海桑红树林地上生物量反演 ——基于 XGBoost 机器学习算法

谭雨欣<sup>1</sup>, 田义超<sup>1,2,\*</sup>, 黄卓梅<sup>1</sup>, 张强<sup>1</sup>, 陶进<sup>1</sup>, 刘虹秀<sup>1</sup>, 杨永伟<sup>1</sup>, 张亚丽<sup>1</sup>,  
林俊良<sup>1</sup>, 邓静雯<sup>2</sup>

1 北部湾大学 资源与环境学院·北部湾海洋发展研究中心, 钦州 535000

2 北部湾大学 广西北部湾海洋环境变化与灾害研究重点实验室·海洋地理信息资源开发利用重点实验室, 钦州 535000

**摘要:** 无瓣海桑是广西从自治区外引进的外来红树林树种, 采用量化算法精确估算无瓣海桑地上生物量对红树林生态修复以及海洋蓝碳监测提供经验和方法。论文以广西茅尾海自然保护区无瓣海桑红树林为研究对象, 以野外实测无瓣海桑红树林地上生物量数据和 Sentinel-1/2 卫星提取的后向散射数据、波段数据、植被指数数据和纹理指数数据为数据源, 通过分析各遥感因子与实测红树林地上生物量之间的重要性关系, 采用极端梯度提升 (XGBoost) 机器学习算法对比了不同的变量组合对模型精度的影响, 最后基于优选的变量组合反演了无瓣海桑红树林的地上生物量。结果表明: (1) 研究区无瓣海桑红树林实测树高范围为 1.55—13.58m, 平均值为 8.37m, 胸径范围为 0.7—41cm, 平均值为 15.62cm; (2) 通过 XGBoost 算法优选的 21 个特征变量组合模型拟合效果较好, 其模型在测试阶段  $R^2 = 0.7237$ ,  $RMSE = 21.70\text{Mg}/\text{hm}^2$ 。XGBoost 算法反演研究区无瓣海桑地上生物量介于 19.14—138.46 $\text{Mg}/\text{hm}^2$  之间, 平均值为 51.92 $\text{Mg}/\text{hm}^2$ ; (3) Sentinel-1 数据衍生的交叉极化 (VH) 后向散射系数对无瓣海桑红树林地上生物量的贡献最大; (4) 无瓣海桑地上生物量高值区主要分布在北部、西北和西南部等偏西地区, 低值区主要分布在东部和东南部等偏东地区, 其反演结果与实际调查结果保持一致。总之, XGBoost 机器学习算法在无瓣海桑红树林地上生物量反演中表现出较好的应用能力。

**关键词:** XGBoost 算法; 遥感反演; 无瓣海桑红树林; 地上生物量; Sentinel-1/2; 北部湾茅尾海

## Aboveground biomass of *Sonneratia apetala* mangroves in Mawei Sea of Beibu Gulf based on XGBoost machine learning algorithm

TAN Yuxin<sup>1</sup>, TIAN Yichao<sup>1,2,\*</sup>, HUANG Zhuomei<sup>1</sup>, ZHANG Qiang<sup>1</sup>, TAO Jin<sup>1</sup>, LIU Hongxiu<sup>1</sup>, YANG Yongwei<sup>1</sup>, ZHANG Yali<sup>1</sup>, LIN Junliang<sup>1</sup>, DENG Jingwen<sup>2</sup>

1 Beibu Gulf Ocean Development Research Center, College of Resources and Environment, Beibu Gulf University, Qinzhou 535000, China

2 Guangxi Key Laboratory of Marine Environment Change and Disaster in Beibu Gulf, Key Laboratory of Marine geographic information resources development and utilization in the Beibu Gulf, Beibu Gulf University, Qinzhou 535000, China

**Abstract:** *Sonneratia apetala* is an exotic mangrove species introduced from outside Guangxi Zhuang Autonomous Region. Quantitative algorithm is used to accurately estimate the aboveground biomass (AGB) of *Sonneratia apetala*, which provides experience and methods for mangrove ecological restoration and marine blue carbon monitoring. This paper takes the *Sonneratia apetalous* mangrove of Mawei Sea Nature Reserve in Guangxi as the research object, and takes the field measured

**基金项目:** 国家自然科学基金项目 (42261024); 广西高校人文社会科学重点研究基地项目 (BHZKY2202); 北部湾大学海洋科学一流学科项目 (DRB003)

**收稿日期:** 2022-01-14; **网络出版日期:** 2023-02-07

\* 通讯作者 Corresponding author. E-mail: tianyichao1314@163.com

aboveground biomass data of *Apetalous* mangrove and the backscatter data, band data, vegetation index data and texture index data extracted by Sentinel-1/2 satellite as the data sources. EXtreme Gradient Boosting (XGBoost) machine learning algorithm was used to compare the effects of different variable combinations on the model accuracy by analyzing the importance relationship between each remote sensing variable and the measured AGB of *Sonneratia apetala* mangrove. Finally, the AGB of *Sonneratia apetala* mangrove was retrieved based on the optimal combination of variables. The results showed that: (1) The measured height of *Sonneratia apetala* mangrove in the study area ranged from 1.55m to 13.58m, with an average of 8.37m, and the diameter at breast height (DBH) ranged from 0.7 cm to 41 cm, with an average of 15.62 cm. (2) The fitting effect of the 21 feature variables combination model optimized by XGBoost algorithm was better, and its model  $R^2=0.7237$  and RMSE=21.70Mg/hm<sup>2</sup> in the testing phase. The AGB of *Sonneratia apetala* mangrove in the study area ranged from 19.14Mg/hm<sup>2</sup> to 138.46Mg/hm<sup>2</sup>, with an average of 51.92Mg/hm<sup>2</sup>. (3) Cross polarization (VH) backscattering coefficient derived from Sentinel-1 data contributed the most to AGB of *Sonneratia apetala* mangrove. (4) The high-value areas of the aboveground biomass of *Sonneratia apetala* mangrove are mainly distributed in the north, northwest and southwest regions to the west, and the low-value areas are mainly distributed in the east and southeast regions to the east. The inversion results were consistent with the actual survey results. In conclusion, XGBoost machine learning algorithm shows good application ability in AGB of *Sonneratia apetala* mangrove.

**Key Words:** XGBoost algorithm; remote sensing inversion; *Sonneratia apetala* mangrove; aboveground biomass; Sentinel-1/2; Mawei Sea of Beibu Gulf

红树林在防风固堤、促淤造陆以及维护海岸带生物多样性等方面具有重要的生态功能<sup>[1-3]</sup>,不仅能维护海岸生物多样性,还影响着海岸带地区社会经济的可持续发展<sup>[4]</sup>。红树林生态系统是地球上生产力最高的生态系统之一<sup>[5]</sup>,其储碳能力突出,尽管它只覆盖了全球表面 1% 的面积,却固定了大气中 5% 的碳<sup>[6]</sup>。因其重要的储碳能力,红树林在缓解气候变化方面存在重要作用。红树林作为海岸带“蓝碳”生态系统的重要组成部分之一,发展蓝碳有利于保护海洋生态环境并能有效应对气候变化。地上生物量作为红树林蓝碳的重要组成部分,如何准确快速地获取红树林地上生物量已成为目前红树林生态系统研究的热门课题。

红树林分布于海岸潮间带,深入根系繁茂的红树林内部开展野外调查工作比较艰难,传统的红树林野外实地调查存在工作强度大、耗费周期长,且调查效率低等问题,对于获取大范围红树林地上生物量 (AGB) 数据具有一定的局限性。随着遥感技术的发展,遥感影像在红树林地上生物量估算方面得到了广泛的应用,相比于传统的样地调查方法,遥感技术可以快速、准确、无破坏地估算红树林地上生物量<sup>[7]</sup>,成为红树林生态研究领域的重要技术手段之一。国内外学者运用遥感技术对森林生物量反演开展了大量的研究,早期的研究成果主要采用 TM、MODIS、ETM+ 等中低分辨率的光学遥感数据进行生物量估测<sup>[8-11]</sup>。随着卫星遥感影像空间分辨率的提高,IKONS-2、WorldView-2 以及 Quick Bird 等高空间分辨率的影像为红树林生物量参数反演提供了便利<sup>[12-13]</sup>。高空间分辨率遥感由于其信息丰富以及空间分辨率高等优势,在森林结构参数信息提取中发挥了很大的作用。虽然被动光学遥感影像数据在获取水平方向上的植被结构信息方面表现出了一定的优势,但由于容易受到森林植被物候和天气的影响,存在“同物异谱”和“同谱异物”现象,在获取森林生物量信息时易饱和,不适用于垂直方向上森林生物量反演<sup>[14]</sup>。主动合成孔径雷达遥感数据具有全天时、全天候工作的特点,能穿透森林植被,获取树干和主枝信息,可应用于森林植被垂直结构参数的反演,具有更高的生物量反演优势,一定程度上解决了森林生物量估算时的饱和性问题<sup>[15]</sup>。有研究表明,将主被动遥感数据进行融合,可以提高森林生物量的遥感估算精度<sup>[16]</sup>。主动合成孔径雷达和被动光学传感器虽然有其不同的优势,但将其融合起来能使森林生物量的估算精度得到提升<sup>[17]</sup>。Sentinel-1 卫星是高分辨率的主动合成孔径雷达卫星, Sentinel-2 卫星是高分辨率的被动多光谱成像卫星,已有众多学者通过研究表明,将 Sentinel-1 和 Sentinel-2 这两个卫星数据源进行融合可以显著提高红树林地上生物量的估算精度。如 Navarro 等<sup>[18]</sup>将无人机 (UAV)、

Sentinel-1 和 Sentinel-2 数据融合应用于塞内加尔红树林人工林地上生物量监测,其结果表明基于 Sentinel-1 的支持向量机回归模型在  $R^2$ 、RMSE 和平均绝对误差 (MAE) 方面的结果最好,但 Sentinel-1 和 Sentinel-2 数据的集成使得在生物量较高的地区获得了更精确的估计。Pham 等<sup>[19]</sup>首次将 Sentinel-1、Sentinel-2 和 ALOS-2 PALSAR-2 数据与极端梯度提升 (XGBoost) 模型和遗传算法相结合,对越南北部红树林生物量进行了研究,研究表明结合 S-1、S-2 和 ALOS-2 PALSAR-2 数据可以估算红树林生物量,具有良好的精度 ( $R^2 = 0.683$ , RMSE = 25.08Mg/hm<sup>2</sup>)。这些研究成果表明雷达数据作为光学数据的补充,在植被生物量反演方面表现出了很大的应用潜力。但是, Sentinel-1 的 SAR 数据和 Sentinel-2 的多光谱数据在红树林地上生物量的估测研究较少,除此之外, Sentinel-1 的主动遥感合成孔径雷达 (SAR) 数据以及 Sentinel-2 的多光谱数据分别衍生出的变量对红树林地上生物量贡献如何? 不同变量组合是否会对红树林地上生物量估算精度产生影响? 这些问题需进一步验证。

除了选择高光谱以及高空间分辨率的遥感数据源之外,选择合适的模型拟合算法构建生物量反演模型也非常重要。目前,由于估算森林地上生物量的算法主要包括参数回归算法和非参数机器学习算法。其中参数回归算法主要包括一元线性回归法、多元回归法、多元逐步回归法和曲线回归法等方法<sup>[20]</sup>,这类方法在森林地上生物量估算时,各个生物量自变量因子之间可能存在严重的多重共线性问题,加之这类算法在模型拟合时可能存在过度拟合现象<sup>[20]</sup>,因此限制了该类模型在生物量反演中的应用。因此机器学习算法应运而生,常见的机器学习算法有随机森林<sup>[21]</sup>、支持向量机<sup>[22]</sup>、神经网络<sup>[23]</sup>和 K-最邻近<sup>[24]</sup>等,例如随机森林算法中涉及的 2 个“随机性”问题,分别为训练集抽取的随机性和节点候选分割特征集合的随机性,在一定程度上增加了用于特征选择的不稳健性<sup>[21]</sup>。这类算法在特征变量数据降维方面表现了一定优势<sup>[21]</sup>,但是这类算法都属于传统的机器学习算法,无法精确从大量的特征变量中提取优势特征变量,而优势特征变量的筛选往往成为生物量反演效率和精度的关键问题。为了提高模型的拟合精度和效率,本文引入了具有优势特征选择和模型拟合能力的极端梯度提升 (XGBoost) 新型机器学习方法<sup>[25]</sup>。该算法是一种在分类和回归上都拥有超高性能的先进评估器,由 Chen 和 Guestrin 提出,作为一种新型的梯度提升决策树算法 (GBDT),目前已被广泛应用于气象学<sup>[26]</sup>、能源<sup>[27]</sup>和医学<sup>[28]</sup>等领域。尽管 XGBoost 算法能够快速对相关特征变量进行可靠性筛选和识别,且该算法具备较强的预测能力,同时不易受训练数据质量的影响,但目前该算法在亚热带地区红树林地上生物量中的估算精度如何? 能否应用到北部湾茅尾海地上红树林生物量的估算? 这些问题也仍然需要实证分析。

无瓣海桑于 1985 年从孟加拉国引入我国,2002 年从广东湛江引入广西钦州茅尾海自然保护区的康熙岭片区进行规模化造林,是一种速生红树林树种,大部分乡土红树林在中低潮带无法生长,无瓣海桑却能迅速定居和建群,甚至进入原生红树林群落内部迅速生长。因其优于其他红树林树种的速生、抗寒、抗逆和防风御浪能力,成为华南沿海防护林体系工程建设中的主要树种之一<sup>[29]</sup>。康熙岭红树林自然保护区作为北部湾最北端的红树林分布区,是中国典型的南亚热带河口、港湾和海岸滩涂湿地复合型生态系统保护区。近年来入海口团和岛海岸滩涂内的无瓣海桑出现了大面积的天然扩散,但是,目前关于扩散区域无瓣海桑生物量估算方面的研究还是及其罕见和匮乏的,该区域已有红树林调查成果多以样地调查方法为主,很少有学者将样地实测数据和主被动卫星数据相结合,采用新型的梯度提升决策树 XGBoost 机器学习方法估算无瓣海桑红树林树种的地上生物量。为此,本研究以 Sentinel-1B 和 Sentinel-2A 主被动卫星遥感为数据源,在 XGBoost 机器学习算法的支持下,探索主被动遥感不同特征变量对无瓣海桑地上生物量的贡献,同时对比了不同的特征变量组合对模型精度的影响。本文可为红树林生态系统修复、生态系统健康和保护以及机器学习算法在海洋蓝碳监测中的应用提供科学依据和技术支撑。

## 1 材料和方法

### 1.1 研究区概况

研究区位于广西茅尾海康熙岭红树林保护片区,介于 108°29'16"—108°29'58"E,21°51'07"—21°51'54"N

之间,该片区属于亚热带季风气候,太阳辐射强,季风环流明显,全年平均气温在 22℃ 左右,年均降水量 1658mm,年总日照为 1673h<sup>[30]</sup>。海潮平缓、海水温度和盐度、土壤和沉积物等优越的自然地理因素为红树林提供适宜生长的资源。此区域的红树林湿地已被列为中国重要湿地之一,是广西壮族自治区级自然保护区,主要生长着无瓣海桑和桐花树,以及混生在红树林中的盐沼草本植物茳芏。本研究中的无瓣海桑红树林树种主要是由原始的人工种植无瓣海桑红树林树种扩散而来,本文的研究区主要位于团和岛的外滩地区,该区域有大面积的无瓣海桑红树林,面积大概有 350hm<sup>2</sup>。由于无瓣海桑红树林野外样地调查工作很难覆盖 350hm<sup>2</sup> 这么大区域,且越靠近外滩地区,工作难度越大,很容易陷入到淤泥和潮沟之中,因此,本文选择了团和岛近岸的红树林开展野外样地调查。具体位置如图 1 所示。

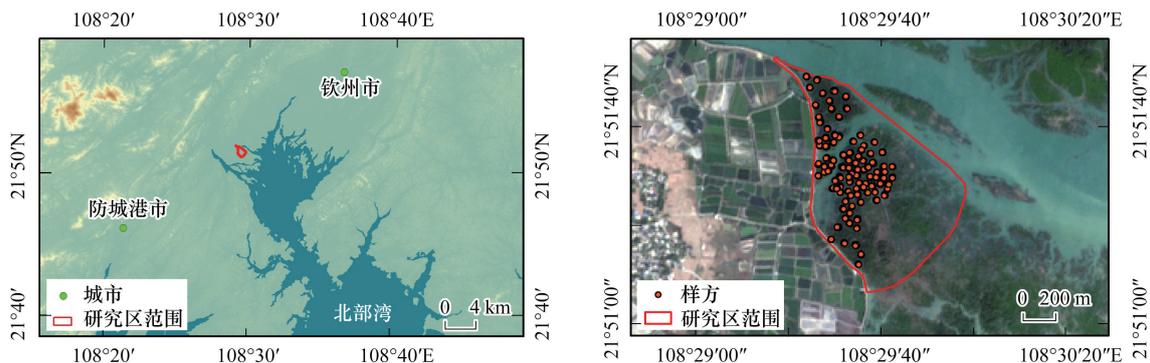


图 1 红树林研究区在广西区分布的地理位置

Fig.1 Location of mangrove research area in Guangxi Autonomous Region

## 1.2 数据来源与预处理

### 1.2.1 野外采样数据

在野外样地采样之前首先需要充分了解当地红树林树种空间分布状况和潮位信息,本团队于 2021 年 1 月 18 日和 19 日在广西茅尾海康熙岭镇红树林片区开展工作,选择退潮和低潮位时间进入无瓣海桑扩散区域开展野外样地调查工作。本文所选择的研究区主要为扩散无瓣海桑红树林群落,林中零散分布了少量茳芏,植被种群结构单一。设置正方形样地调查样方大小为 10m×10m,由于无瓣海桑属乔木,树高可达十几米,因此采用胸径法测量每棵树高度为 1.3m 处的树干直径,通过手持激光测距仪测量树高,利用全球定位系统 (GPS) 获取采样点和样方中心经纬度坐标,共采集 96 个样方。根据树高 ( $H$ )、胸径 (DBH) 计算每棵树的地上生物量,再导入 ArcGIS 10.5 中进行矢量化,得到研究区内无瓣海桑红树林样方空间分布状况。每个样方无瓣海桑的地上生物量计算方法则是采用胡懿凯等<sup>[31]</sup>构建的异速生长方程,他对广东省 10 个地区 (湛江、潮州、广州、惠州、江门、茂名、汕头、深圳、中山、珠海) 的无瓣海桑人工林固定样地群落结构进行了详细调查,获得了无瓣海桑树枝、树叶、树皮和树干等不同植株部位的实测统计数据 (表 1)。红树林气生根属于红树林地上生物量的重要组成部分,由于胡懿凯等<sup>[31]</sup>的在红树林生物量异速生长方程构建时未考虑红树林气生根这个参数,因此本文也未将这个参数纳入到地上生物量的核算之内。由于本文主要核算红树林地上生物量部分,因此,本研究主要参考了无瓣海桑的地上生物量方程及其参数,具体公式如下:

$$AGB = 0.034 \times (DBH^2 \times H)^{0.966} \quad (1)$$

式中,AGB 为生物量 (kg),DBH 为无瓣海桑胸径 (cm), $H$  为无瓣海桑树高 (m)。

### 1.2.2 遥感数据与预处理

哨兵 (Sentinel) 系列数据拥有高空间分辨率、光谱质量好、种类全等优点,是目前最强大的免费遥感数据,是欧洲航天局“哥白尼计划”对地观测卫星组成员,包括 Sentinel-1、Sentinel-2、Sentinel-3、Sentinel-4、Sentinel-5 和 Sentinel-5P 卫星,每个 Sentinel 卫星都经过两颗卫星星座来满足重访和覆盖范围要求<sup>[32]</sup>。数据从欧洲航

空局的数据中心共享网站 <https://scihub.copernicus.eu/dhus/> 获取, 本文采用 Sentinel-1B 和 Sentinel-2A 数据来对红树林地上生物量进行反演研究。

表 1 无瓣海桑生物量模型<sup>[31]</sup>

Table 1 Biomass model of *Sonneratia apetala*

植株部位 Plant parts	模型 Model	<i>a</i>	<i>b</i>	<i>T</i>		统计结果 Statistical result		
				<i>a</i>	<i>b</i>	<i>R</i> <sup>2</sup>	<i>S</i> <sup>2</sup>	<i>P</i>
树枝 Branch	$W = a \times (D^2 \times H)^b$	0.011	0.957	0.848	6.605	0.655	0.753	0.000
树叶 Leaf		0.002	0.905	0.600	4.419	0.459	1.065	0.000
树干 Trunk		0.022	0.937	2.290	17.472	0.930	0.279	0.000
树皮 Bark		0.004	0.932	1.335	10.033	0.814	0.479	0.000
地上部分 Aboveground part		0.034	0.966	2.005	15.758	0.915	0.319	0.000
地下部分 Underground part		0.003	1.119	2.257	20.552	0.948	0.283	0.000
总体 Total		0.033	1.002	2.633	21.475	0.952	0.243	0.000

*W*: 生物量 Biomass; *D*: 胸高直径 Diameter at breast height; *H*: 树高 Height; *T*: 检查统计量 Test statistic; *R*<sup>2</sup>: 决定系数 Coefficient of Determination; *S*<sup>2</sup>: 估计值标准误差 Standard error; *P*: 显著性检验值 Significant value; *a*, *b*: 模型参数

### (1) 雷达影像 Sentinel-1B 数据

Sentinel-1 为 C 波段合成孔径雷达, 由 Sentinel-1A 和 Sentinel-1B 两颗相距 180° 的卫星星座组成。该卫星处于太阳同步的近极轨道上, 轨道高度、轨道倾角和工作频率分别为 693km、98.18°、5.4GHz, 重访周期为 12d, 采用 C 波段工作模式, 运行方式为预编程的无冲突模式, 共有 4 种成像模式和 4 种极化组合方式。Sentinel-1 卫星可以从不同的角度全天时、全天候对地进行高分辨率监测成像, 提供覆盖全球的雷达观测数据, 具备重访周期短、双极化、快速产品生产的能力。本文选择 Sentinel-1B 数据(2020 年 12 月 01 日获取), 采用 Level-1 地距探测产品(GRD), 同极化(VV)、交叉极化(VH)双极化模式, 干涉宽幅模式(IW), 空间分辨率为 10m, 时间分辨率为 12d, 具体参数见表 2。为了获得与真实地物的后向散射强度符合的数据, 需要对 Sentinel-1B 数据进行预处理, 卫星数据预处理过程包括轨道校正、辐射定标、相干斑滤波、地形校正以及后向散射系数转换等, 本文的 Sentinel-1B 图像预处理过程全部借助于欧空局的 SNAP 软件来完成。

表 2 Sentinel-1B 卫星影响参数介绍

Table 2 Introduction to Sentinel-1B satellite impact parameters

参数 Arguments	数值 Value	参数 Arguments	数值 Value
极化方式 Polarization mode	VV, VH	地距分辨率 Ground distance resolution	5m×20m
成像模式 Imaging mode	IW	像元大小 Pixel size	10m×10m
幅宽 Width of cloth	250 km	成像时间 Imaging time	2020-12-01

VV: 同极化 Copolarization; VH: 交叉极化 Cross polarization; IW: 干涉宽条带 Interferometric Wide swath

### (2) 光学影像 Sentinel-2A 数据

Sentinel-2 是一种高分辨率、重访周期短、宽刈幅的多光谱成像卫星, 可用于陆地植被、土壤、水域和海岸的检测研究。该卫星采用太阳同步轨道, 轨道高度为 786km, 轨道倾角为 98.5°, 光谱分辨率范围为 15—180nm, 空间分辨率为 10、20、60m, 幅宽 290km。该卫星拥有 13 个波段, 具有独特的光谱功能, 能提取到较好的波段信息。本文下载的 Sentinel-2-Level-1C 级产品影像于 2020 年 12 月 21 日成像, 该影像已进行正射校正和几何校正处理, 需进一步进行大气校正预处理之后可得到 Sentinel-2-Level-2A 级数据, 该数据可用于红树林地上生物量特征变量的提取。

### 1.3 特征变量提取

在 ENVI5.3 和 ArcGIS10.5 软件平台的支持下, 本文选取 12 个波段信息、7 个植被指数、8 个纹理特征和 2 个后向散射系数四种类型共 29 个特征变量构建红树林地上生物量反演模型。其中, 后向散射系数由

Sentinel-1B 卫星数据提取,而波段信息、植被指数和第一主成分纹理特征(对 Sentinel-2A 数据做主成分分析(PCA)变换)则由 Sentinel-2A 卫星数据提取。

(1) 影像波段值。遥感影像的原始波段是地物识别的基础,能体现出地物的光谱特征,本文提取经预处理后的 Sentinel-2A 影像单波段光谱信息<sup>[33]</sup>,共有 12 个波段,即 B1、B2、B3、B4、B5、B6、B7、B8、B8A、B9、B11、B12,波段基本信息见表 3。

表 3 Sentinel-2A 的 12 个波段基本信息

Table 3 Basic information on 12 bands of Sentinel-2A

特征 Features	属性 Attribute	分辨率/m Spatial resolution	参考文献 References	特征 Features	属性 Attribute	分辨率/m Spatial resolution	参考文献 References
01_B1	气溶胶	60	[33]	07_B7	红边 3	20	[33]
02_B2	蓝光	10		08_B8	近红外(宽)	10	
03_B3	绿光	10		09_B8A	近红外(窄)	20	
04_B4	红光	10		10_B9	水蒸气	60	
05_B5	红边 1	20		11_B11	短波红外 1	20	
06_B6	红边 2	20		12_B12	短波红外 2	20	

(2) 植被指数。植被指数能够反映无瓣海桑红树林植物的生长状况、覆盖度和生物量等信息。本实验根据研究需求,并综合考虑康熙岭无瓣海桑红树林的分布特点,参照前人研究成果,基于 Sentinel-2A 卫星遥感数据提取了 7 种植被指数,即比值植被指数(RVI)<sup>[34]</sup>、归一化植被指数(NDVI)<sup>[35]</sup>、优化土壤调整植被指数(OSAVI)<sup>[36]</sup>、B4 和 B5 归一化垂直植被指数(NDI45)<sup>[37]</sup>、差值植被指数(DVI)<sup>[38]</sup>、绿色标准差异植被指数(GNDVI)<sup>[35]</sup>及新型倒红边叶绿素指数(IRECI)<sup>[39]</sup>,具体指数见表 4。

表 4 选取的植被指数

Table 4 Selected vegetable index characteristics

特征 Features	植被指数 Vegetation Index	公式 Formula	参考文献 References
13_RVI	比值植被指数	$\frac{B8}{B4}$	[34]
14_NDVI	归一化植被指数	$\frac{B8 - B4}{B8 + B4}$	[35]
15_OSAVI	优化土壤调整植被指数	$\frac{B8 - B4}{B8 + B4 + 0.016}$	[36]
16_NDI45	Band4 和 Band5 归一化垂直植被指数	$\frac{B5 - B4}{B5 + B4}$	[37]
17_DVI	差值植被指数	$B8 - B4$	[38]
18_GNDVI	绿色标准差异植被指数	$\frac{B7 - B3}{B7 + B3}$	[35]
19_IRECI	新型倒红边叶绿素指数	$\frac{B7 - B4}{B5/B6}$	[39]

(3) 纹理特征值。纹理特征包含了丰富的影像信息,纹理特征分析方法已被广泛运用于高分辨率影像信息识别中。主成分变换前 1 个波段包含了遥感影像的 90% 以上的信息,通过 ENVI 5.3 对 Sentinel-2A 卫星数据进行主成分变换,采用第一主成分并通过基于二阶概率统计的滤波(Co-occurrence Measures 工)具进行纹理特征提取。本研究采用灰度共生矩阵(GLCM)<sup>[40]</sup>的方法提取 Sentinel-2A 影像第一主成分(PC1)的 8 个纹理特征值,即均值(ME)、方差(VAR)、同质性(HOM)、对比度(CON)、异质性(DIS)、熵(ENT)、角二阶距(SM)、相关性(COR)。纹理特征的基本信息见表 5。

表 5 灰度共生矩阵纹理测度表

Table 5 Gray co-occurrence matrix texture measuring table

特征 Features	纹理特征 Texture Features	公式 Formula	参考文献 References
20_ME	均值	$\sum_{i,j=0}^{N-1} iP_{ij}$	[40]
21_VAR	方差	$\sum_{i,j=0}^{N-1} P_{ij}(1 - \mu_i)$	
22_HOM	同质性	$\sum_{i,j=0}^{N-1} iP_{ij}/(1 + (1 - j)^2)$	
23_CON	对比度	$\sum_{i,j=0}^{N-1} iP_{ij}(i - j)$	
24_DIS	异质性	$\sum_{i,j=0}^{N-1} iP_{ij}  i - j $	
25_ENT	熵	$\sum_{i,j=0}^{N-1} P_{ij}(-\ln P_{ij})$	
26_SM	角二阶距	$\sum_{i,j=0}^{N-1} iP_{ij}^2$	
27_COR	相关性	$\sum_{i,j=0}^{N-1} ((ijP_{ij} - \mu_i\mu_j)/\sigma_i^2\sigma_j^2)$	

$i, j$ : 灰度共生矩阵中第  $i$  行第  $j$  列的灰度值;  $P_{ij}$ : 灰度值为  $(i, j)$  的概率;  $N$ : 影像的行列数;  $\mu$ :  $P$  的均值;  $\sigma$ : 标准差

(4) 后向散射系数。雷达数据的后向散射系数是林分内地面粗糙度以及地表潮湿度综合影响的结果,其大小取决于土壤、植被覆盖度、植被含水量以及地表粗糙度<sup>[41]</sup>。数据预处理的过程主要包括几何校正、辐射定标、地理编码、斑点滤波、地形校正等,全部过程采用 SNAP 软件完成。本研究提取了 Sentinel-1B 数据的同极化 VV 和交叉极化 VH 两种极化后向散射系数作为模型的特征变量。

## 1.4 研究方法

### 1.4.1 技术路线

本文的技术路线如图 2, 主要由野外红树林样地调查与采样点数据处理、遥感数据处理、特征变量提取, 模型参数设置和红树林生物量反演五个部分构成。

### 1.4.2 XGBoost 算法

极值梯度提升算法是 Chen 和 Guestrin 提出的一种新型的梯度提升决策树算法 (GBDT)<sup>[25]</sup>, 该算法基于增强决策树, 能够通过训练过程处理机器学习中弱监督学习的分类和回归任务。XGBoost 模型旨在防止过度拟合, 同时通过简化和正则化使得预测保持最佳计算效率而降低计算成本。XGBoost 比传统的 GBDT 算法更进步的地方在于: 传统的 GBDT 只利用了一阶的导数信息, 而 XGBoost 对损失函数进行了二阶的泰勒展开, 求得模型最优解的效率更高<sup>[42]</sup>。其基本思想就是不断地添加树和进行特征分裂, 每添加一棵树就学习一个新的函数, 以每一轮的预测去拟合上一轮预测的残差, 根据样本的特征就可预测样本分数。当训练完成得到  $n$  棵树, 在每棵树中会落到对应的叶子节点上, 每个叶子节点对应一个分数, 最后将所有树的对应分数相加就得到该样本的预测值。其计算公式如下:

$$\hat{y} = \varphi(x_i) = \sum_{k=1}^k f_k(x_i) \quad f_k(x_i) \in F \quad (2)$$

式中,  $F = \{f(x) = \omega_{q(x)}\}$  ( $q: R^m \rightarrow T, \omega_{q(x)} \in R^T$ ) 为  $k$  棵树集成的函数空间,  $T$  为叶子节点个数,  $f_k$  对应一棵独立的树结构  $q$  与叶子权重  $w$ ,  $q$  为样本对应的叶子标签,  $\omega_{q(x)}$  表示叶子节点  $q$  的分数,  $f(x)$  表示其中某一可回归树。

### 1.4.3 特征选择方法

由于过多的特征会导致数据冗余, 从而降低模型反演精度, 为了提高模型的准确性, 需要筛选出有利于生物量反演的优选特征变量。特征选择能够减少算法的大量计算时间, 通过剔除不相关和冗余的数据, 避免模型过于复杂而产生过拟合现象<sup>[43]</sup>。本研究将 29 个变量导入 Python 中的特征选择器模块, 使用零重要度特征进行最优参数的筛选, 得到特征重要性值, 其范围为 0—1。将变量的重要性值从高到低排序, 重要值为 0 表

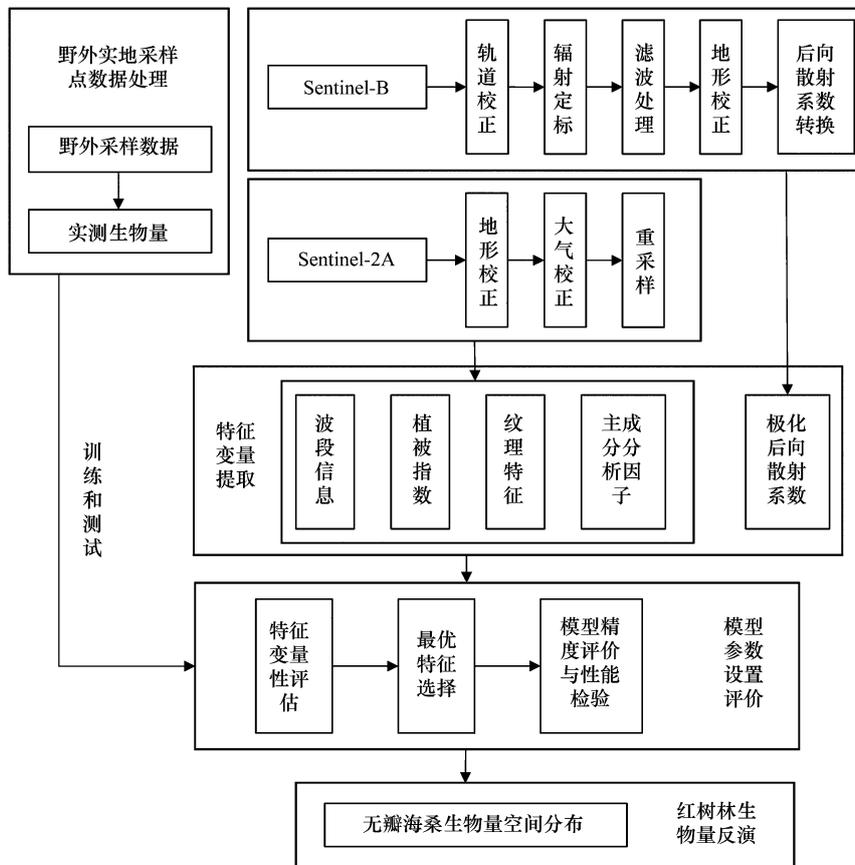


图 2 技术路线

Fig.2 Technical route

示对反演模型没有贡献。同时得到各特征对应的累积重要度,当特征累计重要性达到 95%之后,代表该特征变量的贡献率不大,将大于 95%的特征变量剔除,剩下的累计重要性小于 95%的特征作为一个变量组合。

不同类型的特征变量会对红树林地上生物量建模产生不同的影响,为了研究不同特征变量及特征组合方式对模型精度的影响程度,本文根据研究需要,使用 Sentinel-1B 和 Sentinel-2A 遥感数据设置了 6 组不同情景的变量组合方式(表 6),通过比较不同情景的预测结果,寻找能提高红树林地上生物量反演精度的最佳特征变量组合。

表 6 不同数量特征的组合情景

Table 6 A combination of different quantitative features

情景 Scenario (S)	变量个数 Number of Variables
S1	Sentinel-2 数据的 12 个波段特征
S2	Sentinel-2 数据的 7 个植被指数特征
S3	Sentinel-2 数据第一主成分的 8 个纹理特征
S4	Sentinel-1 数据的 2 个后向散射系数特征
S5	29 个变量组合(12 个波段+7 个植被指数+8 个纹理特征+2 个后向散射系数)
S6	21 个变量组合(11 个波段+3 个植被指数+5 个纹理特征+2 个后向散射系数)

#### 1.4.4 模型评价与检验

##### (1) 模型的配置和训练

本研究中模型的配置和检验使用 Python 编程环境中的 Scikit-learn 库进行,共获取 96 个样本,将其中 75%的样本(72 个)作为训练集,25%的样本(24 个)作为测试集。首先根据不同模型的特征设置初始值,随后采用网格搜索(GridSearch)方法进行超参数寻优<sup>[44]</sup>。通过超参数寻优后获取 XGBoost 机器学习算法的最

佳参数,本文中设置的最优参数分别为 learning rate = 0.1, gamma = 0.01, colsample\_bytree = 0.8, Gamma = 0.1, Max depth = 3, min\_child\_weight = 3, subsample = 0.6, n\_estimators = 100,最后采用最优参数对红树林地上生物量进行模拟。

## (2) 模型评价指标

为了评估模型生物量反演的准确性,本研究使用决定系数  $R^2$ (式 3)、均方根误差 RMSE(式 4)这 2 个指标来验证模型精度。 $R^2$ 为决定系数, $R^2$ 值越接近 1,表示预测值和真实值之间的相关性越大,拟合效果越好;RMSE 代表模型的均方误差的平方根,能够反映预测值与真实值之间的偏差情况,值越小,表示偏差越小,模型的效果越好<sup>[45]</sup>。

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

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

式中,  $y_i$  为实测样地的生物量,  $\bar{y}_i$  为实测红树林平均生物量,  $\hat{y}_i$  为预测的红树林生物量,  $i$  为方程中的数据个数,  $n$  为样本数量。

## 2 结果与分析

### 2.1 实测数据特征统计

表 7 列出了 96 个样方中无瓣海桑红树林的野外实测调查统计结果。由表可知,无瓣海桑树高范围为 1.55—13.58m,胸径范围为 0.7—41cm。通过异速生长方程计算得到无瓣海桑红树林地上生物量,其值变化范围为 5.46—154.83Mg/hm<sup>2</sup>,平均值为 57.7Mg/hm<sup>2</sup>。

表 7 外业测量树高、胸径和生物量的基本统计量

Table 7 Basic statistics for field measurement of tree height, DBH and biomass

参数 Arguments	最大值 Max	最小值 Min	平均值 Mean	标准差 Standard Deviation (SD)
树高 H/m	13.58	1.55	8.37	2.32
胸径 DBH/cm	41	0.7	15.62	5.86
生物量 AGB/(Mg/hm <sup>2</sup> )	154.83	5.46	57.7	32.17

### 2.2 建模结果评价

本文设计了 3 种不同变量及其组合情景,以探究不同特征变量组合对无瓣海桑红树林地上生物量反演精度的影响,并测试 XGBoost 模型在红树林地上生物量反演中的性能。6 个情景包括仅使用遥感影像波段、仅使用植被指数、仅使用纹理特征、仅使用后向散射系数、所有特征变量组合和 21 个优选特征变量组合,分别输入 Python 的 Scikit-learn 库中进行调试,不同变量组合情景的测试结果见表 8 和图 3。

表 8 具有不同数量特征的 XGBoost 模型的性能

Table 8 Performance of the XGBoost model with different numbers of features

情景 Scenario (S)	$R^2$	RMSE/(Mg/hm <sup>2</sup> )	情景 Scenario (S)	$R^2$	RMSE/(Mg/hm <sup>2</sup> )
S1	0.5456	24.95	S4	0.3039	27.31
S2	0.2019	21.18	S5	0.5802	22.56
S3	0.0886	37.57	S6	0.7237	21.70

RMSE: 均方根误差 Root mean square error

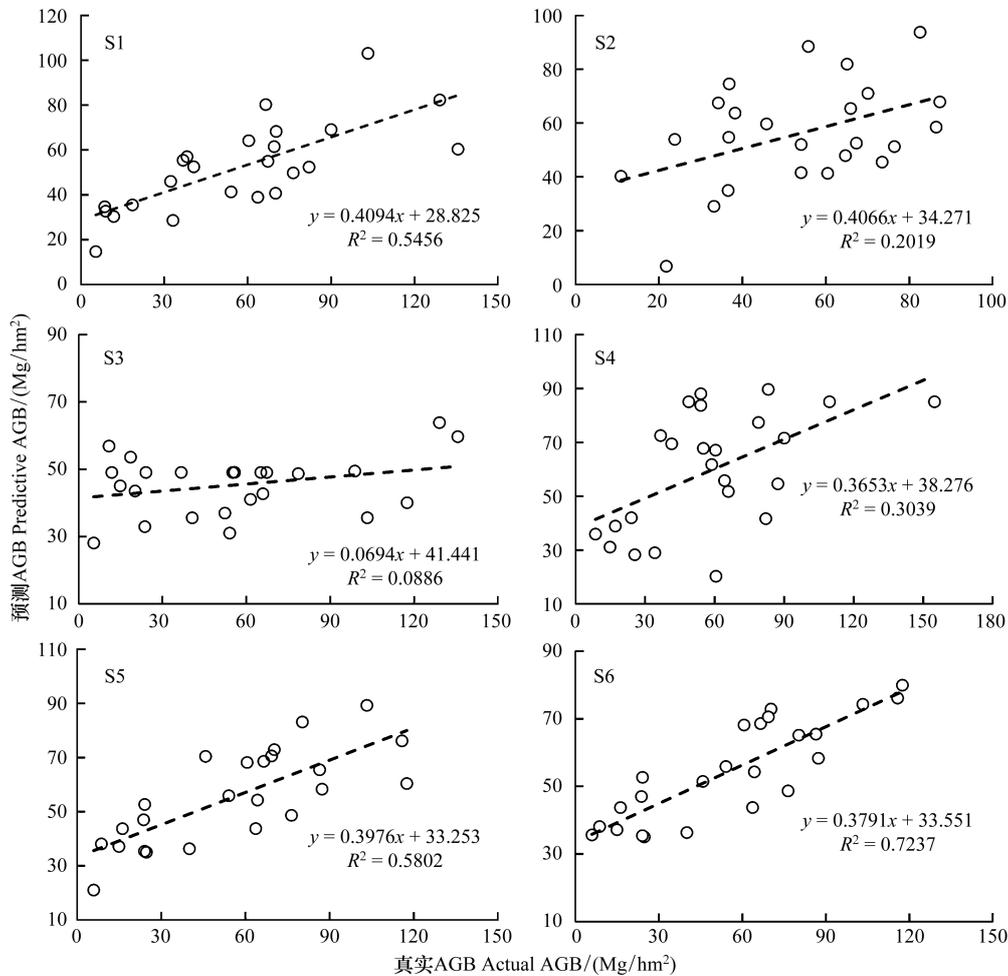


图3 各情景下 AGB 预测值与 AGB 真实值关系图

Fig.3 Relationship diagram between AGB predicted value and AGB actual value in each scenario

AGB: 地上生物量 Aboveground biomass

由图3中的散点图可知,六种特征变量及其组合的真实值和XGBoost模型的预测值具有不同的相关值,不同变量组合对模型精度的影响具有一定的差异性,其 $R^2$ 值介于0.0886—0.7237之间,说明在红树林地上生物量建模时要充分考虑变量之间的差异性以及异质性问题。表7列出了各个情景的 $R^2$ 和RMSE值,可以发现基于XGBoost的模型在S6中取得了良好的结果,优选的21个特征变量组合情景产生了最高的精度,此时 $R^2=0.7237$ , $RMSE=21.70\text{Mg}/\text{hm}^2$ ,与将所有变量组合所得到的精度( $R^2=0.5802$ , $RMSE=22.56\text{Mg}/\text{hm}^2$ )相比, $R^2$ 提高了0.1435, $RMSE$ 降低了 $0.86\text{Mg}/\text{hm}^2$ ,说明通过特征优选可以降低误差并提高模型预测精度。21个最佳特征包括11个Sentinel-2A影像波段(第1、2、3、4、5、6、7、8、9、11、12波段)、3个植被指数(RVI、NDI45和DVI)、5个纹理特征(VAR、HOM、CON、DIS、COR)以及2个极化后向散射系数(VV、VH)。相比之下,单独使用Sentinel-2影像的12个波段、单独使用7个植被指数和2个极化后向散射系数所得评估性能都不太理想,仅使用纹理特征时模型的拟合效果最低( $R^2=0.0886$ , $RMSE=37.57\text{Mg}/\text{hm}^2$ )。不同的变量组合对反演精度的影响与以下几个原因有关:第一,无瓣海桑红树林下垫面背景的差异性对Sentinel卫星影像信息产生一定影响,而不同变量在描述和模拟无瓣海桑红树林三维结构信息及下垫面时所表现出的特性不一致,从而对反演精度产生影响。第二,地形地貌对反演精度产生影响。地形改变了太阳、地表和卫星传感器三者之间构成的几何观测角度,从而造成方向性反射问题<sup>[46]</sup>,无瓣海桑红树林树下存在不同形状大小的潮沟水文单元,这种地形地貌单元在不同变量中所产生的方向性反射会发生变化,可能会对模型的精度产生影响。综上所述,S6

中最优特征的融合提高了 XGBoost 机器学习算法对研究区无瓣海桑红树林地上生物量的预测能力。

### 2.3 变量的重要性

XGBoost 算法可以识别特征变量的重要性,将 29 个特征变量放入 XGBoost 模型进行重要性排序,检验不同特征变量对于红树林地上生物量反演的重要程度,特征变量的重要性程度以标准化重要性 (normalized\_importance) 指标来表征。结果重要性值越高,表示该特征变量对红树林地上生物量的反演贡献越高。由图 4 和 5 可知,对于红树林地上生物量反演贡献最高的特征是 28\_VH,该变量表示 Sentinel-1B 影像的 VH 极化后向散射系数,其值远大于其他特征的重要性值,次重要的特征是 27\_COR、02\_B2、22\_HOM、08\_B8,分别表示 Sentinel-2A 影像第一主成分纹理特征的相关性、Sentinel-2A 的蓝光波段 (B4)、影像第一主成分纹理特征的同质性和 Sentinel-2A 的近红外波段 (B8)。在 Sentinel-2A 多光谱波段中,蓝光波段 (B4)、近红外波段 (B8) 和红边波段 2 (B6) 对红树林地上生物量的敏感性较大;在植被指数中,只有 NDI45 和 DVI 的作用相对较大;在纹理特征中,Correlation、Homogeneity 和 Dissimilarity 对红树林地上生物量最敏感;两个后向散射系数的贡献都很理想。然而,纹理特征值 Second Moment 和植被指数 NDVI 和 OSAVI 等对红树林生物量反演的作用较差。

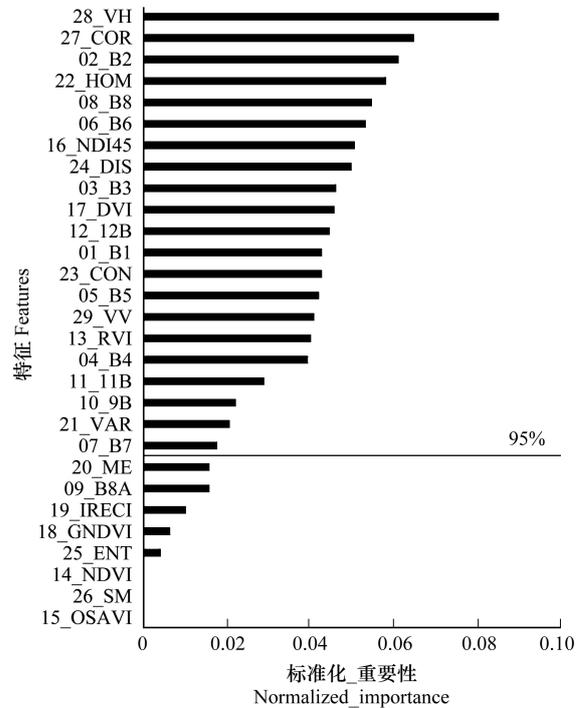


图 4 变量重要性图

Fig.4 Variable importance chart

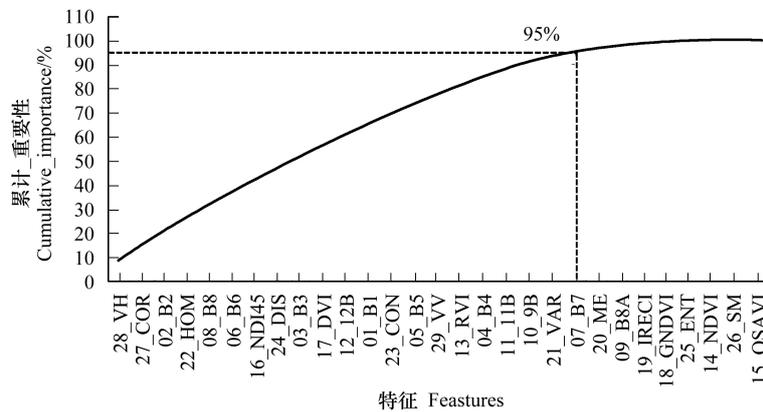


图 5 累计的变量重要性

Fig.5 Cumulative variable importance

### 2.4 无瓣海桑生物量空间分布

由 2.2 节可知,将由波段值、植被指数、纹理特征和后向散射系数组成的 21 个优选特征变量组合作为变量放入到 XGBoost 模型中,模型表现出了最高的预测性能。因此,本研究借助于 XGBoost 模型对研究区的红树林地上生物量进行了反演,结果以 Tiff 栅格数据格式存储,将拟合结果导入 ArcGIS10.5 中进行可视化 (见图 6)。由图 6 可知,无瓣海桑红树林地上生物量最低预测值为 19.14Mg/hm<sup>2</sup>,最高预测值为 138.46Mg/hm<sup>2</sup>,平均值为 51.92Mg/hm<sup>2</sup>。从生物量空间分布看,研究区无瓣海桑自西北至东南呈现出斑块不均匀状分布

特征,这种不均匀分布特征可能与主潮沟和次潮沟的空间分布特征有关,潮沟携带的营养物质是红树林地上生物量呈现出空间分异的原因之一。西北部的无瓣海桑红树林地上生物量值较高;西部地区处于主潮沟的入口处,主潮沟将研究区分隔为两部分,潮沟区域红树林地上生物量值偏小;中部及东南部地区由于荳蔻群落的影响,生物量拟合结果较低;而东部及东北部地区的无瓣海桑地上生物量也低于北部地区,该地区的无瓣海桑是从近陆人工种植区的无瓣海桑种子沿潮沟漂流扩散而来,驻扎在这里的无瓣海桑多以幼小的树苗为主,因此其地上生物量普遍偏低。

### 3 讨论

#### 3.1 模型反演结果与精度评价

由本文的研究结果可知,XGBoost 模型估测的无瓣海桑红树林 AGB 值的平均值为  $51.92\text{Mg}/\text{hm}^2$ ,略低于实地调查的平均值  $57.7\text{Mg}/\text{hm}^2$ ,但是与实地调查值非常接近,这反映了 XGBoost 模型在无瓣海桑红树林地上生物量模拟方面表现出了一定的优势。在本研究中无瓣海桑野外样地调查的地上生物量真实值范围为  $5.46\text{—}154.83\text{Mg}/\text{hm}^2$ ,XGBoost 模型估测的预测值范围为  $19.14\text{—}138.46\text{Mg}/\text{hm}^2$ ,由真实值与预测值的关系可知,在低生物量范围时,XGBoost 模型高估了红树林 AGB;在高生物量范围时,XGBoost 则低估了红树林 AGB。尽管采用了较为先进的机器学习算法,但机器学习算法中不同的特征变量所表现的特征不一致,因此,也不能完全消除不同特征变量和波段值饱和所带来的影响<sup>[47]</sup>。研究区无瓣海桑 AGB 平均值为  $51.92\text{Mg}/\text{hm}^2$ ,低于海南省东北部红树林 AGB 平均值  $119.26\text{Mg}/\text{hm}^2$ ,原因是由于纬度和水热条件不同,红树林地上生物量也表现出区域分异特征。此外,无瓣海桑的生长和扩散也受盐度、土壤肥力、动物采食和潮汐等诸多因素的制约<sup>[48]</sup>,这也使得无瓣海桑红树林地上生物量在不同地区表现出一定的差异性。从空间分布格局上来看,钦州市的无瓣海桑起初为人工引种,主要分布在茅尾海西部沿岸和西北部康熙岭河口、沿岸边缘区,研究区中无瓣海桑是经多年的演化,逐渐扩散至此定居,从团和岛海岸线的纯无瓣海桑生长区由西向东至出海方向逐渐过渡为无瓣海桑与桐花树和荳蔻混生区,这解释了为什么研究区北部和西南部沿岸地区的无瓣海桑地上生物量偏高,而偏东部地区的地上生物量偏低。

在以往的研究中,学者们曾采用简单线性回归<sup>[49]</sup>和多元线性回归<sup>[50]</sup>来进行红树林地上生物量估算,但模型的拟合效果较低,如  $R^2 = 0.51$ <sup>[51]</sup>, $R^2 = 0.65$ <sup>[13]</sup>。但是,随着机器学习算法的快速发展,采用机器学习方法估算红树林地上生物量的成果也相继出现,由本研究的结果可知,机器学习算法方法比多元线性回归或其他参数回归方法取得了更好的拟合效果。Tian 等<sup>[52]</sup>采用不同的机器学习算法对红树林的地上生物量进行了对比研究,研究结果发现新型的 XGBoost 机器学习算法往往比传统的机器学习算法如随机森林以及支持向量机等获得更好的拟合效果,如 Tian 等<sup>[52]</sup>采用 XGBoost 新型机器学习拟合的  $R^2$  可达到 0.8319, RMSE 为  $22.7638\text{Mg}/\text{hm}^2$ 。本文基于 Sentinel-1/2 卫星数据,也采用的新型 XGBoost 算法估测了团和岛外滩地区的无瓣海桑地上生物量,模型取得了良好的拟合效果( $R^2$  为 0.7237, RMSE 为  $21.70\text{Mg}/\text{hm}^2$ ),但比 Tian 等<sup>[52]</sup>的模型拟合效果偏低,原因与模型参数的选择以及卫星遥感数据源的选择有关。Tian 等<sup>[52]</sup>在对红树林地上生物量

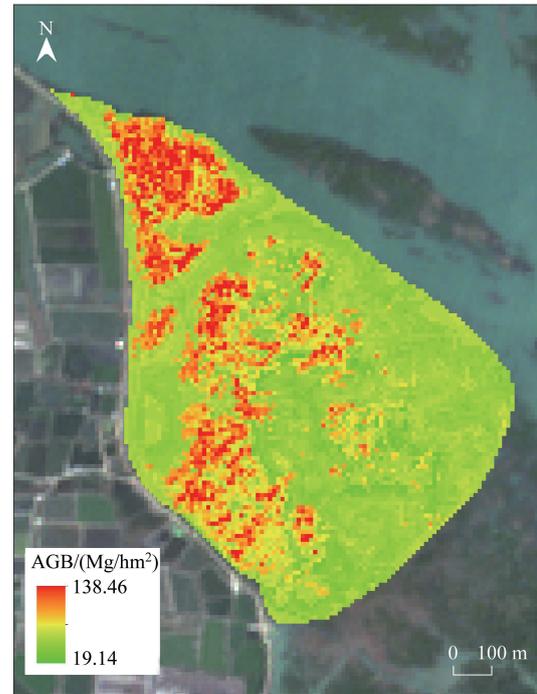


图 6 研究区红树林生物量空间分布图

Fig.6 Spatial distribution map of mangrove biomass in the study area

估算时,采用的是无人机激光雷达技术,与本文采用的传统光学遥感相比,无人机激光雷达技术在红树林三维地物信息的采集方面具有一定的优势,其模型估算红树林的地上生物量的精度也有所提升。此外,模型预测的精度也依赖于样地实测数据的精确性,由于红树林下面潮沟以及潮沟地形地貌的影响,在野外很难保证实测样地都符合大小相同的标准样地,同时遥感卫星接收的图像数据受红树林地形地貌的影响,野外样地的位置很难与图像完全匹配,这对反演的验证精度有一定的影响。

### 3.2 特征变量重要性

由本研究的结果可知,在 XGBoost 模型选择的特征变量中,对于红树林地上生物量敏感性最大的是极化后向散射系数,其次是纹理特征和 Sentinel-2 的波段信息,植被指数的作用效果较其他变量差。极化后向散射特征对红树林的地上生物量估算贡献最大,这是由于 Sentinel-1 的 SAR 具有穿透顶层树冠的能力,不仅与叶生物量有关,也与树枝生物量和树干生物量有关。SAR 数据可以穿透云雾,能够减少大气对定量估测的影响,将卫星光学数据和 SAR 数据结合之后,他们在森林地上生物量估测中就表现出一定的优势<sup>[53]</sup>。例如,韩宗涛等<sup>[24]</sup>联合 P 波段机载 SAR 数据和 Landsat-8 OLI 数据,利用基于快速迭代特征选择的 k 最近邻法(KNN-FIFS)方法估测大兴安岭根河森林保护区的森林地上生物量,证明了结合 SAR 数据和 Landsat-8 OLI 多光谱数据采用 KNN-FIFS 方法估测森林地上生物量的精度较高,本研究提出的将 Sentinel-1 和 Sentinel-2 数据结合,为无瓣海桑红树林地上生物量的估算提供了一种新的思路和方法。重要性筛选结果中 NDVI 对红树林地上生物量估测的精度贡献较少,原因可能是由于传统的 NDVI 指数主要反映植被冠层的信息,不能有效反映垂直方向上的特征(如树高和树干直径),无法全面地获取红树林的立体信息<sup>[54]</sup>。本文选择的无瓣海桑异速生长方程是根据红树林 1.3m 的树高和树干直径来计算研究区的生物量,而 NDVI 值并不能表征红树林的树高以及树干等三维结构。本文的研究结果还发现红树林地上生物量除与 NDVI 不相关之外,还与 SM 和 OSAVI 等特征变量不相关,但与多种纹理特征表现出了极大的相关性,这与以往的研究结果相符,如刘霜<sup>[55]</sup>采用 Sentinel-1/2 数据估测了重庆市南川区森林生物量,研究结果表明 Sentinel-1 的雷达数据和 Sentinel-2 光学数据中的纹理因子均与生物量显示出较大的相关性。Li 等<sup>[56]</sup>的研究也证明了这一点,他采用 SPOT5 影像的纹理特征,分别对 5 种森林类型的生物量进行了估算和验证,结果发现纹理特征在回归方程中的贡献率较大。在本文设计的几种特征组合方案中,单独使用纹理特征预测性能较低,但将纹理特征的变量组合到其他的特征变量时,模型的拟合  $R^2$  会得到显著的提高。Lu<sup>[57]</sup>曾通过研究热带雨林,发现纹理特征与遥感影像的波段值或植被指数结合时,森林生物量的估算精度比单独使用波段值或植被指数的精度得到提高,这也印证了上述的观点。此外值得注意的是,并不是模型的特征变量越多,红树林生物量反演的精度就越高,过多的特征数量容易导致数据的冗余及过拟合,如本实验中将全部的 29 个变量进行组合时,模型拟合的  $R^2$  比优选的 21 个变量组合所得到的  $R^2$  偏低,因此,采用机器学习算法选择合适的特征变量对红树林地上生物量的估算起到了至关重要的作用。

### 3.3 结果的不确定性和研究展望

除了使用较为新型的机器学习反演算法之外,生物量遥感估算的精度也受到多种因素影响,红树群落的纯度是其中之一,曹庆先等<sup>[58]</sup>通过研究表明,使用单一树种的生物量遥感预估精度( $P$ )都远高于混合树种,均方根误差(RMSE)远低于混合树种,这说明单一红树的生物量遥感反演效果要优于混合树种。此外,由人为造成的调查样地坐标方位不准确、多源遥感数据的传感器和分辨率不同、遥感影像校正误差的存在、样地区域范围以及退涨潮时无瓣海桑红树林土壤和水体反射等因素也会对红树林地上生物量的精度产生一定的影响。本文目前尚未考虑到以上因子对于红树林地上生物量的影响,在后续的研究中,将考虑以下几个方面进行改进,以提高估测精度,降低估测误差,得到更精确的估测结果:(1)应尽可能多地增加红树林调查样本数量,并严格按照红树林调查规范进行红树林采样;(2)尽可能选择与获取样地调查数据同期的遥感影像数据,避免时间差;(3)考虑加入气候、土壤、环境等可能影响估测精度的非遥感因子;(4)尝试结合不同传感器的多源数据如激光雷达数据,或采用蚁群算法/遗传算法等优化 XGBoost 机器学习模型来获取精度更高的红树林

地上生物量。此外,由于本研究发现 SAR 数据的后向散射系数对红树林的地上生物量估算精度结果较高,但是其物理机制和作用尚不明确,未来可继续探索不同极化方式如特征向量(H/A/Alpha)分解、特直推式支持向量机(TSVM)分解以及三分量(VanZyl3)分解等极化分解方法对红树林地上生物量的作用机制,以筛选出更适合红树林地上生物量反演的参数和方法,其研究结果可为红树林生态修复以及海洋蓝碳监测提供经验和 technical 方法。

#### 参考文献(References):

- [ 1 ] Murdiyarso D, Purbopuspito J, Kauffman J B, Warren M W, Sasmito S D, Donato D C, Manuri S, Krisnawati H, Taberima S, Kurnianto S. The potential of Indonesian mangrove forests for global climate change mitigation. *Nature Climate Change*, 2015, 5(12): 1089-1092.
- [ 2 ] 何斌源, 范航清, 王瑁, 赖廷和, 王文卿. 中国红树林湿地物种多样性及其形成. *生态学报*, 2007, 27(11): 4859-4870.
- [ 3 ] 张乔民, 隋淑珍. 中国红树林湿地资源及其保护. *自然资源学报*, 2001, 16(1): 28-36.
- [ 4 ] Polidoro B A, Elfes C T, Sanciangco J C, Pippard H, Carpenter K E. Conservation status of marine biodiversity in Oceania: an analysis of marine species on the IUCN red list of threatened species. *Journal of Marine Sciences*, 2011, 2011: 247030.
- [ 5 ] 范航清. 红树林: 海岸环保卫士. 南宁: 广西科学技术出版社, 2000.
- [ 6 ] Duarte C M, Middelburg J J, Caraco N. Major role of marine vegetation on the oceanic carbon cycle. *Biogeosciences*, 2005, 2(1): 1-8.
- [ 7 ] 赵天舸, 于瑞宏, 张志磊, 白雪松, 曾庆奥. 湿地植被地上生物量遥感估算方法研究进展. *生态学杂志*, 2016, 35(7): 1936-1946.
- [ 8 ] Zheng D L, Rademacher J, Chen J Q, Crow T, Bresee M, Le Moine J, Ryu S R. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment*, 2004, 93(3): 402-411.
- [ 9 ] Hall R J, Skakun R S, Arsenault E J, Case B S. Modeling forest stand structure attributes using Landsat ETM+ data: application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, 2006, 225(1/3): 378-390.
- [ 10 ] 黄国胜, 夏朝宗. 基于 MODIS 的东北地区森林生物量研究. *林业资源管理*, 2005, (4): 40-44.
- [ 11 ] 曹庆先, 徐大平, 鞠洪波. 基于 TM 影像纹理与光谱特征的红树林生物量估算. *林业资源管理*, 2010, (6): 102-108.
- [ 12 ] Zhu Y H, Liu K, Liu L, Wang S G, Liu H X. Retrieval of mangrove aboveground biomass at the individual species level with worldview-2 images. *Remote Sensing*, 2015, 7(9): 12192-12214.
- [ 13 ] Hirata Y, Tabuchi R, Patanaponpaiboon P, Pongpam S, Yoneda R, Fujioka Y. Estimation of aboveground biomass in mangrove forests using high-resolution satellite data. *Journal of Forest Research*, 2014, 19(1): 34-41.
- [ 14 ] 张少伟, 惠刚盈, 韩宗涛, 孙珊珊, 田昕. 基于光学多光谱与 SAR 遥感特征快速优化的大区域森林地上生物量估测. *遥感技术与应用*, 2019, 34(5): 925-938.
- [ 15 ] 李兰, 陈尔学, 李增元, 冯琦, 赵磊. 合成孔径雷达森林树高和地上生物量估测研究进展. *遥感技术与应用*, 2016, 31(4): 625-633.
- [ 16 ] 张颖. 基于高分辨率遥感和极化雷达数据的大兴安岭地区森林地上生物量估测[D]. 北京: 北京林业大学, 2016.
- [ 17 ] 汤旭光, 刘殿伟, 王宗明, 贾明明, 董张玉, 刘婧怡, 徐文明. 森林地上生物量遥感估算研究进展. *生态学杂志*, 2012, 31(5): 1311-1318.
- [ 18 ] Navarro J A, Algeet N, Fernández-Landa A, Esteban J, Rodríguez-Noriega P, Guillén-Climent M L. Integration of UAV, Sentinel-1, and Sentinel-2 data for mangrove plantation aboveground biomass monitoring in Senegal. *Remote Sensing*, 2019, 11(1): 77.
- [ 19 ] Pham T D, Yokoya N, Xia J S, Ha N T, Le N N, Nguyen T T T, Dao T H, Vu T T P, Pham T D, Takeuchi W. Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam. *Remote Sensing*, 2020, 12(8): 1334.
- [ 20 ] 谭清梅, 刘红玉, 张华兵, 王聪, 侯明行. 盐城海滨湿地植被地上生物量遥感估算研究. *自然资源学报*, 2013, 28(12): 2044-2055.
- [ 21 ] 杨明星, 徐天蜀, 牛晓花, 霍鹏, 岳彩荣. 基于 Sentinel-1A 雷达影像的思茅松林蓄积量估测. *西部林业科学*, 2019, 48(2): 52-58.
- [ 22 ] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, 1995, 20(3): 273-297.
- [ 23 ] 张媛, 王玲, 包安明, 刘海隆. 基于神经网络的玛纳斯河流域植被地上生物量反演. *干旱区研究*, 2019, 36(4): 863-869.
- [ 24 ] 韩宗涛, 江洪, 王威, 李增元, 陈尔学, 闫敏, 田昕. 基于多源遥感的森林地上生物量 KNN-FIFS 估测. *林业科学*, 2018, 54(9): 70-79.
- [ 25 ] Chen T Q, Guestrin C. XGBoost: A scalable tree boosting system//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016: 785-794.
- [ 26 ] Fan J L, Wang X K, Wu L F, Zhou H M, Zhang F C, Yu X, Lu X H, Xiang Y Z. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. *Energy Conversion and Management*, 2018, 164: 102-111.
- [ 27 ] Torres-Barrón A, Alonso Á, Dorronsoro J R. Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing*, 2019, 326-327: 151-160.
- [ 28 ] 谷鸿秋, 王春娟, 李子孝, 王伊龙, 王拥军, 姜勇. 基于 Logistic 回归与 XGBoost 构建缺血性卒中院内复发风险预测模型的初步比较研究.

- 中国卒中杂志, 2020, 15(6): 587-594.
- [29] 李云, 郑德璋, 陈焕雄, 廖宝文, 郑松发, 陈相如. 红树植物无瓣海桑引种的初步研究. 林业科学研究, 1998, 11(1): 42-47.
- [30] 田义超, 黄远林, 张强, 陶进, 张亚丽, 黄鹄, 梁铭忠, 周国清. 北部湾典型海岛生态系统服务价值空间异质性对比研究. 海洋科学, 2019, 43(2): 60-68.
- [31] 胡懿凯, 徐耀文, 薛春泉, 罗勇, 廖宝文, 朱宁华. 广东省无瓣海桑和林地土壤碳储量研究. 华南农业大学学报, 2019, 40(6): 95-103.
- [32] 龚燃. “哨兵”卫星家族概览. 国际太空, 2014, (7): 23-28.
- [33] Jordan C F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 1969, 50(4): 663-666.
- [34] Dash J, Curran P J. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Advances in Space Research*, 2007, 39(1): 100-104.
- [35] Blackburn G A. Quantifying chlorophylls and carotenoids at leaf and canopy scales: an evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, 1998, 66(3): 273-285.
- [36] Venkatesh B, Anuradha J. A review of feature selection and its methods. *Cybernetics and Information Technologies*, 2019, 19(1): 3-26.
- [37] Daughtry C S T, Walthall C L, Kim M S, De Colstoun E B, McMurtrey III J E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 2000, 74(2): 229-239.
- [38] Mcfeeters S K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 1996, 17(7): 1425-1432.
- [39] Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 1996, 55(2): 95-107.
- [40] Kadiyala A, Kumar A. Applications of python to evaluate the performance of decision tree-based boosting algorithms. *Environmental Progress & Sustainable Energy*, 2018, 37(2): 618-623.
- [41] 蒙良莉. 基于哨兵多源遥感数据的红树林信息提取算法研究[D]. 南宁: 南宁师范大学, 2020.
- [42] 沙靖岚. 基于 LightGBM 与 XGBoost 算法的 P2P 网络借贷违约预测模型比较研究[D]. 大连: 东北财经大学, 2017.
- [43] Powell S L, Cohen W B, Healey S P, Kennedy R E, Moisen G G, Pierce K B, Ohmann J L. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. *Remote Sensing of Environment*, 2010, 114(5): 1053-1068.
- [44] Lin G, Fu J Y, Jiang D, Wang J H, Wang Q, Dong D L. Spatial variation of the relationship between PM<sub>2.5</sub> concentrations and meteorological parameters in China. *BioMed Research International*, 2015, 2015: 1-15.
- [45] 杨福芹, 冯海宽, 李振海, 高林, 杨贵军, 戴华阳. 基于赤池信息量准则的冬小麦叶面积指数高光谱估测. 农业工程学报, 2016, 32(3): 163-168.
- [46] 王新云, 郭艺歌, 过志峰, 覃文汉, 孙国清. 地形因子对森林冠层 BRF 模拟的影响分析. 遥感信息, 2012, 27(4): 82-85.
- [47] 刘雪莲, 欧绍龙, 陆双飞, 岳彩荣. 基于 Sentinel-1A 微波遥感数据的森林蓄积量估测. 西部林业科学, 2020, 49(6): 128-136.
- [48] 文玉叶. 不同纬度无瓣海桑的繁殖和扩散特性研究[D]. 厦门: 厦门大学, 2014.
- [49] Darmawan S, Sari D K, Takeuchi W, Wikantika K, Hernawati R. Development of aboveground mangrove forests' biomass dataset for Southeast Asia based on ALOS-PALSAR 25-m mosaic. *Journal of Applied Remote Sensing*, 2019, 13(4): 044519.
- [50] Hamdan O, Aziz H K, Hasmadi I M. L-band ALOS PALSAR for biomass estimation of Matang Mangroves, Malaysia. *Remote Sensing of Environment*, 2014, 155: 69-78.
- [51] Pham T D, Yoshino K. Aboveground biomass estimation of mangrove species using ALOS-2 PALSAR imagery in Hai Phong City, Vietnam. *Journal of Applied Remote Sensing*, 2017, 11(2): 026010.
- [52] Tian Y C, Huang H, Zhou G Q, Zhang Y, Tao J, Zhang Y L, Lin J L. Aboveground mangrove biomass estimation in Beibu Gulf using machine learning and UAV remote sensing. *Science of the Total Environment*, 2021, 781: 146816.
- [53] 潘婧靓, 邢艳秋, 黄佳鹏, 汪献义. 联合 GF-3 PolSAR 数据与 Landsat-8 OLI 数据的森林地上生物量估测. 中南林业科技大学学报, 2020, 40(8): 83-90.
- [54] 黎夏, 叶嘉安, 王树功, 刘凯, 刘小平, 钱峻屏, 陈晓越, 何执兼, 覃朝锋. 红树林湿地植被生物量的雷达遥感估算. 遥感学报, 2006, 10(3): 387-396.
- [55] 刘霜. 基于 Sentinel-1/2 的重庆市南川区森林生物量估算研究[D]. 成都: 成都理工大学, 2020.
- [56] Li M S, Tan Y, Pan J, Peng S K. Modeling forest aboveground biomass by combining spectrum, textures and topographic features. *Frontiers in Forestry in China*, 2008, 3(1): 10-15.
- [57] Lu D S. Estimation of Forest Stand Parameters Using Landsat TM Images in the Brazilian Amazon Basin.//30th International Symposium on Remote Sensing of Environment: Information for Risk Management and Sustainable Development. Center for the Study of Institutions, Population, and Environmental Change, Indiana University, Bloomington, IN 47408, 2003: 423-426.
- [58] 曹庆先, 徐大平, 鞠洪波. 北部湾沿海 5 种红树林群落生物量的遥感估算. 广西科学, 2011, 18(03): 289-293.