

基于 Landsat 8 OLI 影像的森林生物量估测方法比较*

熊 壮¹ 胡中岳² 曹 聪¹ 刘 萍¹ 徐正春¹

(1. 华南农业大学林学与风景园林学院, 广东 广州 510642; 2. 国家林业和草原局中南调查规划设计院, 湖南 长沙 410014)

摘要 为选出精准评估森林生物量的计量方法, 文章以广东省河源市森林为研究对象, 基于 Landsat 8 OLI 影像数据和 2017 年广东省第九次国家森林资源连续清查数据, 分析对比不同特征变量选择方法, 采用多元逐步回归模型和随机森林回归模型建立河源市森林乔木层地上生物量 (Above ground biomass, AGB) 反演模型。研究表明, 利用 Pearson 相关系数法筛选变量的多元逐步回归模型 (R^2 为 0.505, E_{rmse} 为 $36.862 \text{ t} \cdot \text{hm}^{-2}$, E_{mae} 为 $30.555 \text{ t} \cdot \text{hm}^{-2}$, A_{ea} 为 58.4%) 表现最佳; 相比随机森林的特征重要性来筛选自变量, 利用 Pearson 相关系数法进行自变量选择的筛选方法更佳; 随机森林算法估算精度均优于多元逐步回归模型。采用随机森林算法进行森林地上生物量反演应用前景良好。

关键词 森林生物量; 森林资源连续清查; Landsat 8 OLI 遥感影像; 随机森林

中图分类号: S771.8 文献标志码: A 文章编号: 2096-2053 (2025) 02-0020-08

DOI: 10.20221/j.cnki.2096-2053.202502003

Discussion on the Remote Sensing Inversion of Forest Biomass in Heyuan Based on the Landsat 8 OLI Image

XIONG Zhuang¹ HU Zhongyue² CAO Cong¹
LIU Ping¹ XU Zhengchun¹

(1. College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, Guangdong 510642, China; 2. Central South Inventory and Planning Institute, National Forestry and Grassland Administration, Changsha, Hunan 410014, China)

Abstract To select the appropriate biomass estimation method to accurately evaluate forest biomass, the forest in Heyuan city, Guangdong province was taken as the research object. Based on Landsat 8 OLI image data and the data from the Ninth National Continuous Forest Inventory of Guangdong Province in 2017, the selection methods of different characteristic variables were analyzed and compared. The multiple stepwise regression models and random forest regression models were used to establish the estimation model of above ground biomass (AGB) in the forest tree layer of Heyuan city. The results showed that the multiple stepwise regression model with variable selection using the Pearson correlation coefficient method (with a coefficient of determination R^2 of 0.505, E_{rmse} of $36.862 \text{ t} \cdot \text{hm}^{-2}$, E_{mae} of $30.555 \text{ t} \cdot \text{hm}^{-2}$, and A_{ea} of 58.4%) performed the best. Compared to selecting independent variables using feature importance from the random forest, the Pearson correlation coefficient method yielded better results for independent variable selection. Nevertheless, the estimation accuracy of the random forest algorithm was superior to that of the multiple stepwise regression model. The random forest al-

* 基金项目: 广东省林业科技创新重点项目 (2021KJCX009)。

第一作者: 熊壮 (2000—), 男, 硕士研究生, 主要从事森林可持续经营研究。E-mail: nagisabear@126.com

通信作者: 徐正春 (1965—), 男, 教授, 主要从事林业经济与森林经营研究。E-mail: zcxu@scau.edu.cn

gorithm has promising application prospects for AGB estimation in forests.

Key words forest biomass; national forest inventory (NFI); Landsat 8 OLI image; random forest

以乔木为主的森林生态系统与地球的水循环、碳循环及气候环境息息相关,对于全球碳平衡及气候变化的研究具有重要意义^[1]。森林生物量作为研究碳循环、森林生物量、碳储量等方面的重要依据,直接代表着森林生态系统的健康状况^[2-3]。广东省作为国家首批低碳试点省和国家7个碳排放权交易试点省之一,减排增汇任务艰巨且重要,研究掌握广东省森林生态系统的生物量与碳储量,从而可更好地提升广东省森林经营质量和推动碳中和目标的实现^[4]。传统样地调查方式存在投入高、效率低、费时费力、不适用大面积作业等问题^[5],而遥感技术具有快速、实时、动态、大范围的特性,使森林生物量实现动态监测成为可能^[6-8]。现已有很多研究通过调查数据、遥感数据源、数字高程模型中的参数等来估测研究区域的AGB(Above ground biomass)^[9]、森林碳储量^[10]和林分立地质量^[11]等。有研究提出添加气候、地形和土壤数据辅助建模,能有效提高模型精度,但会提高模型的复杂度^[12]。非参数模型中的随机森林(Random forest, RF)机器学习算法,具有高度灵活的优点,在森林资源调查方面具有更多的适应性^[13]。陈中超等^[14]以无人机激光雷达点云数据和样地数据,分别使用多元线性模型与非线性模型对赤峰市喀喇沁旗旺业甸的人工林林分碳储量进行了估测,得到非线性模型估测出的数据比线性模型估测出的数据更加精确的结论。蒋馥根等^[15]利用Landsat 8 OLI影像结合实地调查数据,建立多元线性逐步回归、地理加权回归、kNN(K-Nearest Neighbors)回归和随机森林等4种森林地上生物量反演模型,结果表明利用随机森林模型反演得到的估测生物量空间分布与实际基本一致,反演效果较好。本文以2017年广东省森林资源连续清查中的河源市乔木林样地为例,利用Landsat 8 OLI遥感数据,采用多元逐步回归模型和随机森林回归模型进行森林地上生物量反演,并比较与评价各种模型精度。

1 数据与方法

1.1 数据来源

河源市位于广东省东北部,地处东江中上游、

韩江上游和北江上游,地形以山地、丘陵为主,属亚热带季风气候,气候温和,雨量充沛,常年平均气温20.7℃。全市总面积15 654 km²。本研究基础数据包含固定样地调查数据和遥感影像数据两部分。固定样地调查数据源自2017年广东省森林资源连续清查数据,样地为25.82 m×25.82 m、正方形,面积0.066 7 hm²,地类为乔木林地,起源以天然林为主,森林类型以阔叶混交、针阔混交为主,龄组以中幼龄林为主。遥感数据为Landsat 8 OLI遥感影像,下载自地理空间数据(www.gscloud.cn)。选择与样地调查时间同期的2017年云量低图像以保证影像质量。利用ENVI5.3软件对Landsat 8 OLI数据进行辐射定标、大气校正和镶嵌等处理。森林连续清查数据的坐标系是北京_54,而遥感数据的坐标系是WGS_84,为了使样地坐标系与遥感影像坐标系相匹配,需要将样地数据的坐标系转换为WGS_84。最后在ArcGIS软件中筛选出广东省第九次森林连续清查数据落在河源市区域的样地。

1.2 研究方法

1.2.1 乔木层地上生物量计算方法 杉木 *Cunninghamia lanceolata*、马尾松 *Pinus massoniana*、栎属树种 *Quercus* spp.、湿地松 *Pinus elliotii*、木荷 *Schima superba*、柳杉 *Cryptomeria japonica*、枫香树 *Liquidambar formosana* 的生物量参考行业标准《立木生物量模型及碳计量参数》计算^[16-22],没有行业标准的树种生物量采用通用性生物量模型计算^[23-24]。通过单木生物量模型计算得到单木生物量,再将样地内所有单木生物量相加得到对应的样地生物量,统计结果见表1。

1.2.2 遥感因子提取及自变量筛选 选取Landsat 8 OLI遥感影像中b1、b2、b3、b4、b5、b6、b7等7个单波段分辨率(Band 1-Band 7)、大气抗阻植被指数(ARVI)、差值植被指数(DVI)、扩展增强型植被指数2(EEVI2)、增强归一化差植被指数(ENDVI)、增强植被指数(EVI)、绿度差值植被指数(GDVI)、绿度归一化植被指数(GNDVI)、绿度比值植被指数(GRVI)、近红外百分比值植被指数(NIPVI)、红外植被指数2(IVI)、短红外温度指数(MVII7)、归一化植被

表 1 固定样地森林乔木层地上生物量统计结果

Table 1 Statistical results of AGB values in the forest tree layer of the sample plots

样地个数	单个样地地上生物量				
	最小值/(t·hm ⁻²)	最大值/(t·hm ⁻²)	平均值/(t·hm ⁻²)	标准差/(t·hm ⁻²)	变异系数/%
147	11.093	247.459	73.518	49.986	67.992

指数 (NDVI)、归一化植被指数 2 (NDVI43)、优化土壤调整指数 (OSAVI)、比值植被指数 (RVI)、比值植被指数 2 (RVI65)、比值植被指数 3 (RVI75)、土壤调节植被指数 (SAVI)、阴影植被指数 (SVI) 等 19 个常用植被指数, 以及窗口 3×3 的均值 (Mean)、方差 (Variance)、协调性 (Homogeneity)、对比度 (Contrast)、相异性 (Dissimilarity)、熵 (Entropy)、二阶矩 (Second moment)、相关性 (Correlation) 等 8 个纹理特征值, 每种单波段对应 8 个常用的纹理特征因子, 7 个波段共 56 个纹理特征值, 通过软件 ENVI 提取出纹理特征。

利用 SPSS 25 计算所有因子与单位地上生物量之间的 Pearson 相关系数 (Pearson correlation coefficient) 矩阵。选择与单位地上生物量在 0.01 显著水平上显著相关的变量开展变量筛选。为了避免变量之间的共线性, 引入方差膨胀因子 (Variance inflation factor, F_{VI}) 进行自变量共线性分析。对共线性诊断后 F_{VI} 值小于 10 的自变量放入逐步回归模型。

随机森林可以根据因变量对提供的自变量进行重要性排序, 对于筛选合适的自变量具有非常好的参考价值。利用 Python 语言中 Random Forest 函数对 161 个特征变量进行重要性排序, 选取重要性高的变量同时用于地上生物量反演模型建模, 将建模结果与 Pearson 相关系数特征变量选择法进行比较。

1.2.3 随机森林模型 随机森林作为非参数模型, 对比其他常用模型在分析数据、误差平衡以及变量重要性评估等方面具有优势, 且训练速度快, 能够得到变量重要性排序。研究选择 Python 中的 Random Forest 函数, 利用软件 Pycharm Community Edition 编写基于随机森林算法的森林地上生物量反演程序, 导入样地数据与遥感数据进行学习分析, 并采用了十折交叉验证方法 (Ten-fold Cross-Validation), 将数据集划分为 10 个子集, 每次用 9 个子集进行训练, 剩下 1 个子集进行测试, 循环 10 次以获取平均性能指标。这种方法有效地利用了全部数据, 减少了模型对单一数据集划分的依

赖, 提高了结果的可靠性。

1.2.4 模型精度检验 随机森林回归中, 本文随机抽取 80% 的数据作为训练集训练模型, 数据剩下的 20% 作为验证集以评估模型的泛化能力。采用的模型精度评价指标为决定系数 (R^2)、调整后的决定系数 (R^2_{adj})、均方根误差 (E_{rmse})、平均绝对误差 (E_{mae}) 和估测精度 (Estimation accuracy, A_{ea})。

2 结果与分析

2.1 基于 Pearson 相关系数法筛选变量估测地上生物量

通过 SPSS 25 的 Pearson 相关系数矩阵计算, 得到在 0.01 级别 (双尾) 上的因子共 35 个, 包含不同的单波段、植被指数以及纹理特征。从中选择相关性较高的因子, 同时根据 F_{VI} 的数值筛选出合适自变量, 避免共线性, 采用多元逐步回归模型建立线性模型进行变量筛选, 得到变量为 Mean1 (对应单波段 Band 1 的纹理特征: 均值, 下同)、Second_moment1、Mean2、Dissimilarity4、Second_moment4、Mean6、Dissimilarity7、Second_moment7、b3、b7、NDVI43、ARVI、IVI。变量及多元逐步回归模型统计结果如表 2 所示。

2.1.1 多元逐步回归模型 根据 F_{VI} 筛选后, 得到多元线性回归模型方程为:

$$Y = -56.655 - 3.285 \times x_1 + 19.781 \times x_2 + 33.003 \times x_3 + 45.883 \times x_4 + 47.402 \times x_5 - 6.477 \times x_6 - 7.86 \times x_7 + 1.401 \times x_8 - 0.14 \times x_9 + 0.105 \times x_{10} - 392.518 \times x_{11} - 126.493 \times x_{12} + 66.463 \times x_{13}$$

式中: x_1 为 Mean1, x_2 为 Second_moment1, x_3 为 Mean2, x_4 为 Dissimilarity4, x_5 为 Second_moment4, x_6 为 Mean6, x_7 为 Dissimilarity7, x_8 为 Second_moment7, x_9 为 b3, x_{10} 为 b7, x_{11} 为 NDVI43, x_{12} 为 ARVI, x_{13} 为 IVI, Y 为地上生物量。

由上式可知, 地上生物量与 Mean1、Mean6、Dissimilarity7、b3、NDVI43、ARVI 呈负相关, 与 Second_moment1、Mean2、Dissimilarity4、Second_moment4、Second_moment7、b7、IVI 呈正相关, 其中 NDVI43 对地上生物量估测值影响最大。

表 2 基于 Pearson 相关系数法的多元逐步回归模型统计数据

变量	未标准化系数	标准错误	标准化系数	T 检验	F_{VI}
常量	-56.655	77.395		-0.732	
Mean1	-3.285	10.238	-0.037	-0.321	3.621
Second_moment1	19.781	14.255	0.098	1.388	1.352
Mean2	33.003	12.275	0.354	2.689	4.643
Dissimilarity4	45.883	21.410	0.286	2.143	4.767
Second_moment4	47.402	22.162	0.260	2.139	3.960
Mean6	-6.477	4.320	-0.218	-1.499	5.652
Dissimilarity7	-7.860	11.859	-0.078	-0.663	3.719
Second_moment7	1.401	21.245	0.007	0.066	2.867
b3	-0.140	0.040	-0.658	-3.501	9.494
b7	0.105	0.028	0.699	3.759	9.273
NDVI43	-392.518	68.175	-0.696	-5.757	3.922
ARVI	-126.493	32.455	-0.508	-3.897	4.562
IVI	66.463	12.124	0.545	5.482	2.652

2.1.2 随机森林回归模型 本研究首先通过 Pearson 相关系数矩阵计算, 获得了经 VIF 筛选前后的 2 组变量数据集。鉴于随机森林算法作为非参数模型具有处理高维数据、避免过拟合等优势, 研究将 2 组变量数据分别导入模型进行训练。通过对比分析, 最终获得了自变量经 VIF 筛选与未经 VIF 筛选的 2 种随机森林回归模型的估测精度结果, 为后续变量筛选方法比较和模型比较提供了实证依据。

2.2 基于随机森林算法筛选变量估测地上生物量

利用 Python 中 Random Forest 函数计算, 可以得到 56 个特征变量的重要性数值并排序, 选取重要性最高的前 13 个特征变量, 如表 3 所示。

表 3 变量重要性数值

变量	重要性数值
NDVI43	0.275 4
IVI	0.194 3
b4	0.025 3
ENDVI	0.021 8
Mean4	0.020 5
Variance5	0.019 3
OSAVI	0.016 1
b3	0.014 4
NDVI	0.014 4
Correlation5	0.014 3
b2	0.013 5
RVI	0.013 0
Contrast6	0.012 9

2.2.1 多元逐步回归模型 根据表 3 中随机森林重要性数值排序筛选出自变量放入多元逐步回归模型, 得到统计数据如表 4 所示。

由表 4 可知, 随机森林回归的结果中, 有 4 个自变量 F_{VI} 值显著高于 10, 存在共线性过高的现象, 因此并不适合用于多元逐步回归。

2.2.2 随机森林回归模型 根据随机森林重要性数值排序筛选合适自变量放入随机森林模型中, 得到模型估测精度结果, 见表 5。

2.3 两种筛选变量方法对地上生物量估测精度比较

本文通过 Pearson 相关系数法和随机森林法筛选出用于建模的自变量因子, 结合多元逐步回归和随机森林回归分别建立反演模型, 最后利用 R^2 、均方根误差、平均绝对误差和估测精度以评价各模型精度。结果如表 5 所示。

根据表 5 可以看出, 利用 Pearson 相关系数法筛选因子后建立的最终模型结果精度优于随机森林法。在两种因子筛选方式中, 随机森林回归均为最优模型, 但 Pearson 相关系数法结合随机森林回归 ($F_{VI} < 10$) 模型精度 ($R^2 = 0.542$, $R_{adj}^2 = 0.497$, $E_{rmse} = 33.827 \text{ t} \cdot \text{hm}^{-2}$, $E_{mae} = 26.490 \text{ t} \cdot \text{hm}^{-2}$, $A_{ca} = 56.9\%$) 要更优于随机森林自变量筛选与随机森林回归结合的模式精度 ($R^2 = 0.513$, $R_{adj}^2 = 0.465$, $E_{rmse} = 334.863 \text{ t} \cdot \text{hm}^{-2}$, $E_{mae} = 30.043 \text{ t} \cdot \text{hm}^{-2}$, $A_{ca} = 56.9\%$)。利用 Pearson 相关系数法筛选因子建立的 3 种模型中, 经过自变量因子共线

表 4 基于随机森林的多元逐步回归模型统计数据

Table 4 Multiple Stepwise Regression model statistics based on random forest

变量	未标准化系数	标准错误	标准化系数	T 检验	F_{VI}
常量	-25.959	51.909		-0.500	
NDVI43	-469.434	108.758	-0.832	-4.316	10.096
IVI	46.392	12.145	0.380	3.820	2.691
ENDVI	-133.596	144.088	-0.395	-0.927	49.204
b4	0.321	0.079	1.744	4.043	50.537
Mean4	1.814	11.520	0.027	0.157	8.198
Variance5	1.248	1.155	0.097	1.081	2.203
b3	-0.274	0.073	-1.290	-3.742	32.272
NDVI	-0.238	86.331	-0.001	-0.003	34.315
Correlation5	5.805	7.185	0.051	0.808	1.095
Contrast6	-0.728	1.352	-0.048	-0.538	2.158
b2	-0.014	0.037	-0.062	-0.372	7.513
RVI	6.657	2.882	0.426	2.310	9.219

表 5 不同模型预测结果

Table 5 Prediction results based on different multiple models

因子筛选方式	模型	R^2	R^2_{adj}	$E_{rmse}/(t \cdot hm^{-2})$	$E_{mae}/(t \cdot hm^{-2})$	$A_{ca}/\%$
Pearson 相关系数法	多元逐步回归	0.505	0.456	36.862	30.555	58.4
	随机森林回归 ($F_{VI} < 10$)	0.542	0.497	33.827	26.490	56.9
	随机森林回归	0.523	0.372	34.520	28.252	60.4
随机森林法	多元逐步回归	0.507	0.462	36.653	73.407	47.2
	随机森林回归	0.513	0.465	34.863	30.043	56.9

性筛选后的随机森林回归模型最优, R^2 为 0.542, 比其他模型的 R^2 都要高, 多元逐步回归模型次之, 自变量没有经共线性分析的随机森林回归虽然 R^2 数值高于多元逐步回归, 但 R^2_{adj} 仅为 0.372, 说明模型拟合优度并不高。

将遥感数据带入 4 个模型中, 得到预测值与实测值对比, 如图 1 所示。散点图展示了实际值与预测值之间的关系。4 张图表中的数据点分布较为一致, 回归线均呈正斜率, 说明预测值随着实际值增加而增加。3 种随机森林回归都相近且回归线通过大多数点, 表明随机森林回归模型的预测值与实测值之间有较好的线性关系。多元逐步回归的预测精度较差, 数据点偏离回归线较远。

2.4 河源市地上生物量反演

通过上文可知, 基于 Pearson 相关系数法和共线性分析的随机森林模型为最佳模型, 以 Landsat 8 OLI 影像为数据源, 结合广东省森林资源连续清查数据, 利用随机森林模型对研究区进行地上生物量反演, 反演结果及卫星图像如图 2 所示。反演得到河源市单位面积生物量为 $72.175 t \cdot hm^{-2}$, 与

广东省森林资源连续清查数据计算出的单位面积生物量 $73.518 t \cdot hm^{-2}$ 接近。河源市样地的森林类型多为阔叶混交林, 林龄以中幼林为主, 结合河源市的卫星地图观察, 研究区西北部、中部、东部及南部森林地上生物量偏低, 这些地区以城镇为主, 裸地面积大, 因此植被覆盖度较低; 研究区东北部、西部、北部偏西及中部偏南的人为活动少, 森林覆盖率和郁闭度高, 因此森林地上生物量也相对更高。随机森林反演结果显示, 研究区西北部、中部、东部及南部的森林地上生物量分布处于较低的水平, 研究区东北部、西部、北部偏西及中部偏南的地区森林地上生物量较高, 与研究区的森林地上生物量实际分布情况基本一致, 整个研究区的反演结果较理想, 进一步证明了随机森林在森林地上生物量遥感反演中的可行性和适用性。

3 讨论与分析

3.1 本文基于 Landsat 8 OLI 影像数据构建模型, 并反演河源市森林地上生物量。研究结果表明,

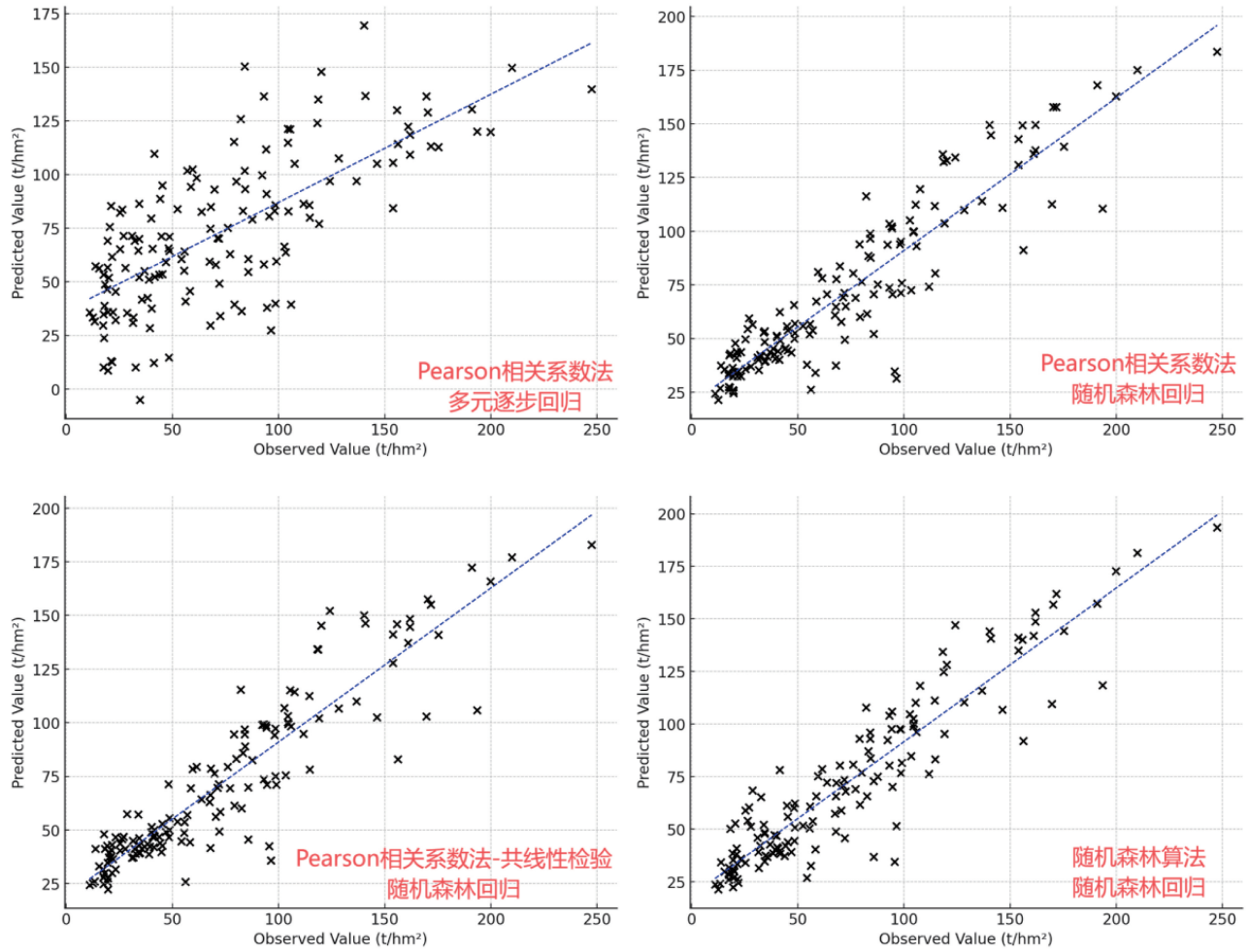
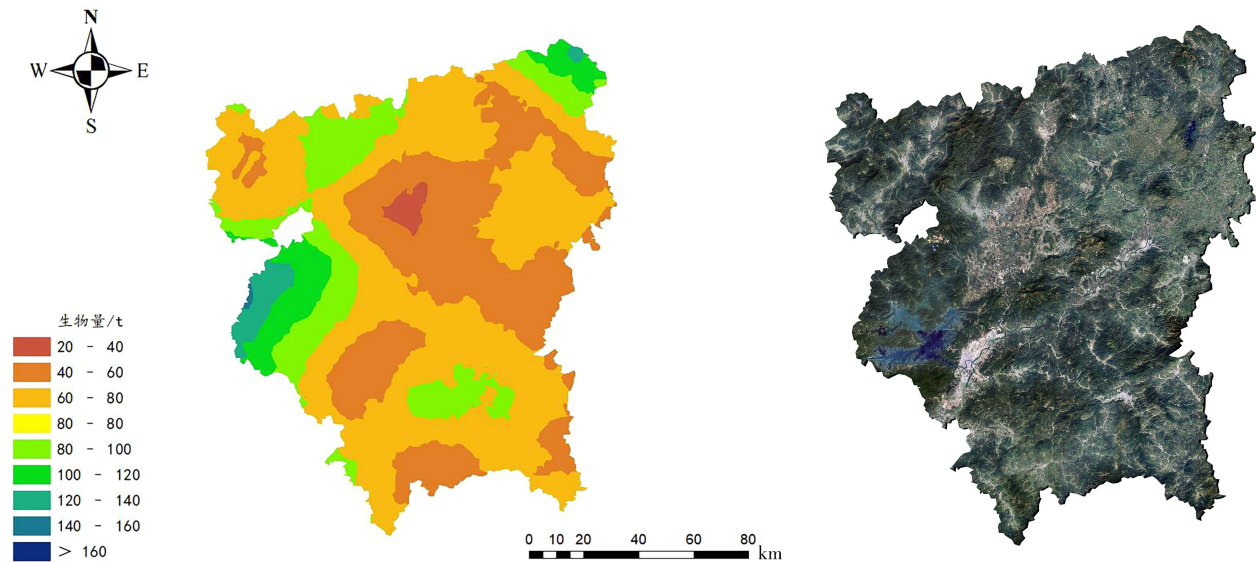


图 1 随机森林回归模型和多元逐步回归模型的预测值与实测值之比

Figure 1 The ratio of predicted value to observed value in Random Forest and Multiple Stepwise Regression Models



注：左图为生物量反演图，右图为卫星图。

图 2 河源市森林乔木层生物量反演图及卫星图像

Figure 2 Biomass inversion and satellite imagery in the forest tree layer of Heyuan city

与李康杰^[25]、郑伟楠^[26]等研究利用 Landsat 8 OLI 影像数据, 并采用多元逐步回归、BP 神经网络及随机森林模型所得的估测结果一致, 证明 Landsat 8 OLI 影像数据在森林地上生物量估测中的适用性。本文与李康杰^[25]等的研究对象以天然林为主, 主要树种为阔叶混交林, 林龄以中幼龄林为主, 而郑伟楠^[26]等和蒋馥根^[15]等的研究则以人工林为主, 树种主要为松林。研究发现, 人工林的生物量遥感反演精度较高, 可能由于其受人为经营管理的影响较大, 从而减少了其他环境因子(如气候、土壤水分及地形因素)对反演结果的干扰。因此, 在未来的研究中, 除了关注森林起源的差异, 还需综合考虑树种组成、林龄及立地质量等因素, 以进一步提高生物量估测的准确性。本文通过 Pearson 相关系数法和随机森林法筛选出相关系数高的变量, 提高特征变量选择的多样性, 更全面地分析比较不同建模方式的优劣。

多元逐步回归作为参数模型, 虽然具有较强的可解释性, 但在处理大规模和复杂数据时, 其估测偏差较大, 适应性较差。相比之下, 随机森林法的估测精度更优于多元逐步回归, 表明在遥感数据与森林地上生物量反演中, 非参数模型相比参数模型具有更高的灵活性和适应性。

随机森林法能够对变量进行重要性排序, 但这并不意味着重要性较高的变量组合在一起就一定能够构建出最优模型。本研究结果表明, 相较于基于随机森林法变量重要性排序的随机森林回归模型, 基于 Pearson 相关系数筛选变量的随机森林回归模型表现更优。这表明 Pearson 相关系数法结合多元逐步回归分析能够筛选出拟合度更高、误差更小的变量组合, 从而提升模型的估测精度。

遥感建模估测技术极大便利了森林地上生物量估测, 随机森林算法在处理高维度、多变量且存在非线性关系的数据时表现出色, 同时, 可以结合其他机器学习算法, 如支持向量机 (Support vector machine, SVM)、神经网络等, 进一步提高模型的精度和适应性。对于多元逐步回归模型, 需加强对变量共线性的处理, 通过引入主成分分析 (Principal component analysis, PCA) 或偏最小二乘回归 (Partial least squares regression, PLSR) 等方法, 减少共线性对模型性能的影响。在未来的研究中, 可以尝试其他遥感影响如雷达遥感^[27], 提高遥感影像的穿透能力, 或是引入更多的辅助

数据, 如气候、地形、土壤等环境因子^[28], 提高模型对森林地上生物量估测的精度。可以在不同季节、不同年份开展多时相研究, 分析森林地上生物量的时空动态变化规律, 为森林资源的长期监测和管理提供支持。随着遥感技术和数据分析方法的不断发展和进步, 利用遥感数据进行森林地上生物量反演和监测将具有更加广阔的应用前景和发展空间。

本研究结果为河源市乃至广东省的森林资源监测和管理提供了科学依据和技术支持, 有助于推动区域碳中和目标的实现。同时, 本研究方法和模型可推广应用于其他地区, 为全国森林生态系统的监测和管理提供参考。

3.2 本文结合 Landsat 8 OLI 遥感数据和 2017 年广东省森林资源连续清查数据, 利用 Pearson 相关系数法和随机森林法进行自变量因子选择, 利用多元逐步回归模型和随机森林模型两种方法反演河源市的森林地上生物量, 评估不同模型的精度, 选出最优方法。研究结果如下:

(1) 利用 Pearson 相关系数法选出的变量组合要优于随机森林法根据重要性排序得到的变量组合。通过 Pearson 相关系数矩阵计算, 得到在 0.01 级别 (双尾) 上的因子, 选择相关性较高的因子, 同时根据方差膨胀因子的数值来判断筛选因子。同时通过随机森林法对变量因子的重要性排序, 选取重要性较高的特征变量。采用多元线性逐步回归和随机森林模型分别建立反演模型。

(2) 基于 Pearson 相关系数法筛选变量时, 随机森林回归模型变量为 F_{VI} 数值筛选前时, R^2 为 0.523, E_{rmse} 为 $34.520 \text{ t} \cdot \text{hm}^{-2}$, E_{mae} 为 $28.252 \text{ t} \cdot \text{hm}^{-2}$, A_{ea} 为 60.4%。变量为 F_{VI} 数值筛选后时, R^2 为 0.542, E_{rmse} 为 $33.827 \text{ t} \cdot \text{hm}^{-2}$, E_{mae} 为 $26.490 \text{ t} \cdot \text{hm}^{-2}$, A_{ea} 为 56.9%。多元逐步回归模型的 R^2 为 0.505, E_{rmse} 为 $36.862 \text{ t} \cdot \text{hm}^{-2}$, E_{mae} 为 $30.555 \text{ t} \cdot \text{hm}^{-2}$, A_{ea} 为 58.4%。基于随机森林法筛选变量时, 随机森林回归模型的 R^2 为 0.513, E_{rmse} 为 $34.863 \text{ t} \cdot \text{hm}^{-2}$, E_{mae} 为 $30.043 \text{ t} \cdot \text{hm}^{-2}$, A_{ea} 为 56.9%。筛选出来的变量间存在较高的共线性, 所以不推荐用于多元逐步回归模型。综上, 基于 Pearson 相关系数法和共线性分析的随机森林回归为最佳模型。

(3) 随机森林模型的河源市森林地上生物量

反演结果与研究区森林地上生物量实际分布大致相同。研究区西北部、中部、东部及南部有城镇分布,裸地面积相对较广,因此森林地上生物量值偏低,研究区东北部、西部、北部偏西及中部偏南的森林覆盖率较高,因此森林地上生物量值更高。反演结果理想,且单位面积生物量的估测值与实际值接近,证明在地上生物量遥感反演中使用随机森林模型的可行性和适用性高,值得提倡。

参考文献

- [1] 徐伟义,金晓斌,杨绪红,等. 中国森林植被生物量空间网格化估计[J]. 自然资源学报,2018,33(10):1725-1741.
- [2] 邱书志,薄乖民,丁骞,等. 白龙江林区森林植被碳储量和碳汇功能研究[J]. 中南林业科技大学学报,2018,38(1):88-93.
- [3] 沈希,张茂震,祁祥斌. 基于回归与随机模拟的区域森林碳分布估计方法比较[J]. 林业科学,2011,47(6):1-8.
- [4] HU H, WANG S, GUO Z, et al. The stage-classified matrix models project a significant increase in biomass carbon stocks in China's forests between 2005 and 2050 [J]. Scientific Reports, 2015, 5(1):11203.
- [5] 张志,田昕,陈尔学,等. 森林地上生物量估测方法研究综述[J]. 北京林业大学学报,2011,33(5):144-150.
- [6] SOMOGYI Z, CIENCIALA E, MÄKIPÄÄ R, et al. Indirect methods of large-scale forest biomass estimation[J]. European Journal of Forest Research, 2007, 126(2):197-207.
- [7] CROSBY M K, MATNEY T G, SCHULTZ E B, et al. Consequences of Landsat image strata classification errors on bias and variance of inventory estimates: a forest inventory case study[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(1):243-251.
- [8] 何兴元,任春颖,陈琳,等. 森林生态系统遥感监测技术研究进展[J]. 地理科学,2018,38(7):997-1011.
- [9] 王楠. 基于 Landsat8OLI 影像的森林地上生物量估算和空间尺度转换研究[D]. 南京:南京林业大学,2023.
- [10] 郭泽鑫,胡中岳,曹聪,等. 广东主要森林类型林分生物量和碳储量模型研建[J]. 林业科学,2023,59(12):37-50.
- [11] 陈敏,叶金盛,刘萍. 基于森林连续清查数据的马尾松立地质量评价[J]. 林业资源管理,2020(6):85-89.
- [12] 余超,宋立奕,李明阳,等. 河南西峡县森林地上生物量时空动态分析[J]. 南京林业大学学报(自然科学版),2017,41(6):93-101.
- [13] BELGIU M, DRĂGU L. Random forest in remote sensing: A review of applications and future directions[J]. Isprs Journal of Photogrammetry and Remote Sensing, 2016, 114:24-31.
- [14] 陈中超,刘清旺,李春干,等. 基于无人机激光雷达的人工林碳储量线性与非线性估测模型比较[J]. 北京林业大学学报,2021,43(12):9-16.
- [15] 蒋馥根,孙华,林辉,等. 旺业甸林场人工林生物量遥感反演研究[J]. 中南林业科技大学学报,2019,39(10):88-94.
- [16] 国家林业和草原局. 立木生物量模型及碳计量参数: 马尾松:LY/T 2263-2014[S]. 北京:中国标准出版社,2014.
- [17] 国家林业和草原局. 立木生物量模型及碳计量参数: 湿地松:LY/T 2261-2014[S]. 北京:中国标准出版社,2014.
- [18] 国家林业和草原局. 立木生物量模型及碳计量参数: 杉木:LY/T 2264-2014[S]. 北京:中国标准出版社,2014.
- [19] 国家林业和草原局. 立木生物量模型及碳计量参数: 柳杉:LY/T 2657-2016[S]. 北京:中国标准出版社,2016.
- [20] 国家林业和草原局. 立木生物量模型及碳计量参数: 栎树:LY/T 2658-2016[S]. 北京:中国标准出版社,2016.
- [21] 国家林业和草原局. 立木生物量模型及碳计量参数: 木荷:LY/T 2660-2016[S]. 北京:中国标准出版社,2016.
- [22] 国家林业和草原局. 立木生物量模型及碳计量参数: 枫香:LY/T 2661-2016[S]. 北京:中国标准出版社,2016.
- [23] 曾伟生. 基于木材密度的34个树种组一元立木生物量模型建立[J]. 林业资源管理,2017(6):41-46.
- [24] 曾伟生,唐守正. 一个新的通用性相对生长生物量模型[J]. 林业科学,2012,48(1):48-52.
- [25] 李康杰,胡中岳,刘萍,等. 基于 Landsat-8OLI 影像的珠三角森林生物量估测[J]. 中南林业科技大学学报,2023,43(3):73-81.
- [26] 郑伟楠,吴勇,欧光龙. 基于 Landsat 8OLI 的昆明市主要森林类型生物量遥感估测与反演[J]. 西南林业大学学报,2023,43(11):107-116.
- [27] 阮兰君,杨燕琼. 遥感在森林生物量监测的应用[J]. 林业与环境科学,2016,32(2):94-98.
- [28] 李超. 中亚热带森林地上生物量遥感估测精度提升方法及影响因素分析[D]. 南京:南京林业大学,2021.