

# 基于 GF-2 的乔木生物量估测模型研究\*

丁志丹 孙玉军<sup>†</sup> 孙 钊

(北京林业大学森林资源和环境管理国家林业和草原局重点开放性实验室, 北京, 100083)

**摘要** 以福建省将乐林场为研究区, 使用野外实测样地数据, 结合福建省二类调查数据, 获取了共 192 个样地的生物量数据, 其中杉木纯林 112 个, 马尾松纯林 80 个. 对覆盖研究区的 2 景 GF-2 影像进行预处理, 提取光谱信息、植被指数、纹理特征及地形因子, 筛选与样地生物量相关性较高的因子作为建模的自变量, 采用支持向量机、随机森林及多元逐步回归 3 种方法分别建立了杉木和马尾松生物量模型. 结果表明: 支持向量机、随机森林模型拟合效果均比多元逐步回归模型好, 其中随机森林模型决定系数  $R^2$  最高, 2 种样地的  $R^2$  分别为 0.65 和 0.72, 估计精度也最高, 分别为 65.28% 和 76.82%; 杉木样地 3 种模型的均方根误差分别为 64.27、48.16 和 77.03, 马尾松样地 3 种模型的均方根误差分别为 54.79、48.18 和 65.63, 其中随机森林模型的最低. 在 3 种模型中, 随机森林模型为乔木生物量的最优模型.

**关键词** 乔木生物量; GF-2; 支持向量机; 随机森林; 多元逐步回归

**中图分类号** S757.2

**DOI:** 10.12202/j.0476-0301.2020440

随着全球气候变暖, 森林系统的碳储量和生物量已经成为了近年来森林生态系统与全球气候变化研究最受关注的问题之一. 森林生物量是森林生态系统长期生产与代谢过程中积累的结果, 是森林生态系统运转的能量基础和物质来源, 如何精确获取森林植被信息, 计算生物量和碳储量, 成为当前研究的关键<sup>[1]</sup>. 传统的森林生物量研究方法采用实地测量法, 即在研究区内进行每木调查、树干解析、全伐样方内草本和灌木等, 这种方法耗费人力物力, 导致成本很高, 而且大多适用于地理条件较好的地区. 随着遥感技术的发展, 遥感技术逐渐在精准林业当中被运用, 并成为当今估算森林生物量和碳储量的主要方法<sup>[2-5]</sup>, 估算生物量可以实现时间和空间尺度上的一致性和连续性, 实现大范围的森林资源动态监测, 对评估碳储功能具有很好的促进作用<sup>[6]</sup>.

近几年, 随着国产高分系列影像数据的出现, 高空间分辨率数据逐渐运用于遥感估测生物量的研究之中<sup>[7-9]</sup>. 本文以高分二号(GF-2)高分辨率遥感影像为主要数据源, 结合外业样地调查数据和二类调查数据, 并提取相关的遥感因子, 基于 R 软件, 使用 3 种建模方法, 建立了将乐县国有林场杉木和马尾松乔木生物量与遥感因子间的模型, 筛选出最优模型, 为今后在该区域进行生物量或碳储量的动态监测和资源管理提供有力的科学依据.

## 1 研究材料

**1.1 研究区概况** 研究区为将乐林场, 设在将乐县国有林场内, 位于福建省三明市将乐县. 将乐林场始建于 1958 年, 地处闽江上游金溪流域两岸、将乐县城区周边, 经营区面积达 75.3 km<sup>2</sup>, 森林总蓄积量 1.16×10<sup>6</sup> m<sup>3</sup>, 森林覆盖率 85.7%, 木材年生产量 2×10<sup>4</sup> m<sup>3</sup>. 将乐县位于福建省西北部, 地处武夷山脉东南部, 海拔 180~500 m, 土壤以红壤为主, 少量分布有黄红壤. 属中亚热带季风气候, 降水充沛, 年降水量 1770 mm, 年平均气温 19.8 ℃. 将乐县山地面积 1 920 km<sup>2</sup>, 其中有林地面积 1 887 km<sup>2</sup>, 毛竹林 293 km<sup>2</sup>, 森林覆盖率达 85.2%, 林木蓄积量 1.598×10<sup>7</sup> m<sup>3</sup>, 是中国南方重点林业县、中国毛竹之乡. 该区域树种丰富, 以杉木(*Cunninghamia lanceolata* (Lamb.) Hook.)、马尾松(*Pinus massoniana* Lamb.)及毛竹(*Phyllostachys heterocyclus* (Carr.) Mitford 'Pubescens')为主要树种. 研究区位置如图 1 所示.

### 1.2 数据来源

**1.2.1 样地数据** 选取 192 个样地, 其中杉木纯林 112 个, 马尾松纯林 80 个. 数据来源包括 2 个部分: 一部分采用 2017、2018 和 2019 年 3 期的外业调查数据, 在将乐林场内设置了 102 个 20 m×30 m 的样地,

\* 林业科学技术推广资助项目([2019]06)

<sup>†</sup> 通信作者: 孙玉军(1963—), 男, 博士, 教授. 研究方向: 森林资源调查与监测, 森林结构与生长模型模拟, 林业遥感与信息技术. E-mail: sunyj@bjfu.edu.cn

收稿日期: 2020-09-11

其中杉木纯林 52 个、马尾松纯林 50 个,手持 GPS 记录每个样地的 4 个角点的经纬度坐标,在样地对每株树木进行每木检尺,记录树种、胸径、树高、冠幅、郁闭度、树龄、林分起源、密度、立地条件等调

查因子;另一部分数据来自福建省第 3 次森林资源二类调查数据,在研究区范围内选取 90 个样地,其中杉木纯林 60 个、马尾松纯林 30 个.样地分布情况如图 2 所示.

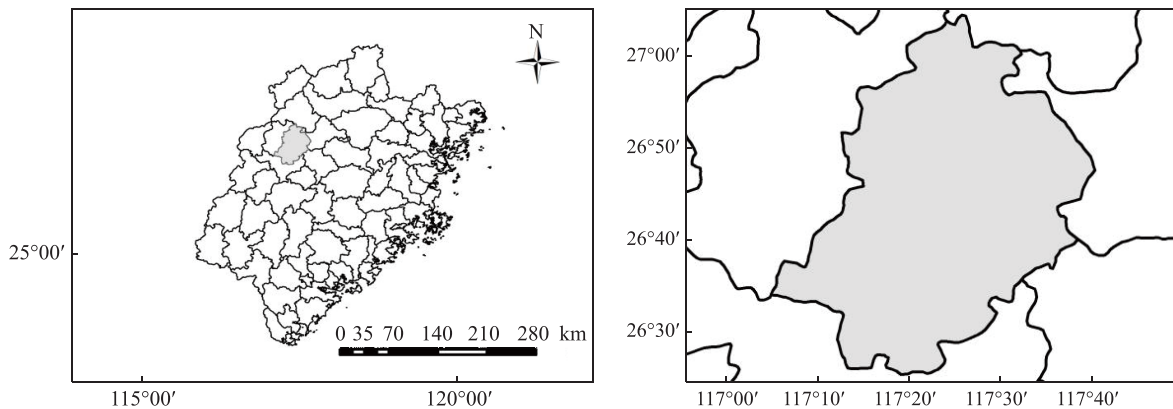


图 1 研究区位置

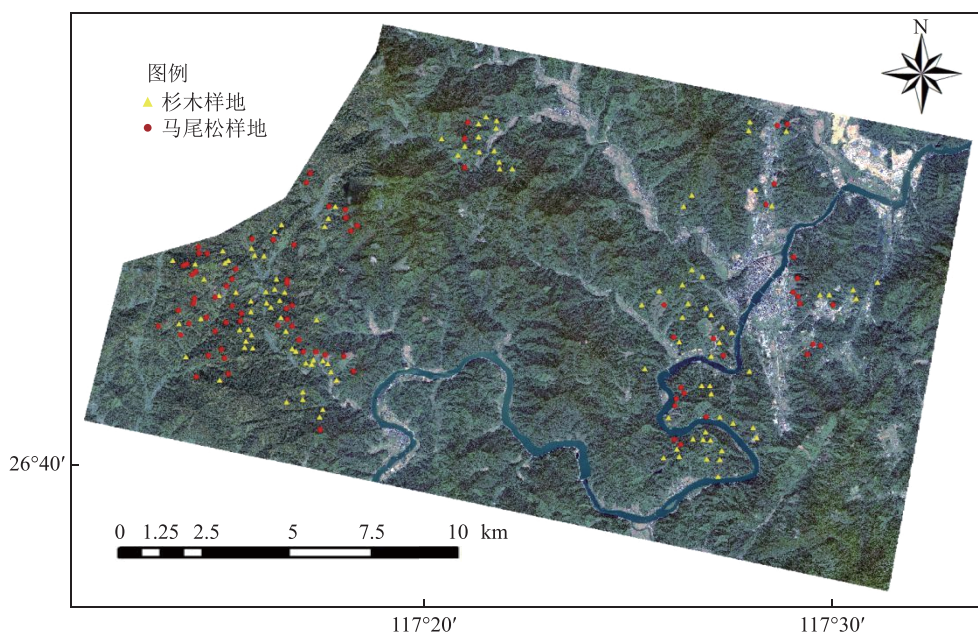


图 2 样地分布

**1.2.2 遥感数据** 遥感数据是国产高分二号(GF-2),由蓝(450~520 nm)、绿(520~590 nm)、红(630~690 nm)、近红外(770~890 nm)4个多光谱波段(4 m)和1个全色波段(1 m)组成,结合卫星的过境时间以及云量等因素选取2018年10月28日接收的GF-2卫星数据为数据源,包括2景多光谱影像和2景全色影像,覆盖整个研究区.

## 2 研究方法

### 2.1 数据处理

**2.1.1 遥感图像预处理** 用ENVI 5.3.1软件对GF-2影像进行预处理.1)对GF-2原始影像中的多光谱影

像进行辐射定标和FLAASH大气校正,对全色影像进行辐射定标,目的是去除大气对影像的影响,使图像品质得到明显的提高<sup>[10]</sup>.2)利用RPC文件和30 m分辨率的DEM(数字高程图)数据,分别对多光谱影像和全色影像进行正射校正,目的是消除地形的影响或是相机方位引起的变形等;对处理好的多光谱和全色影像进行图像融合,为了使融合效果更好,采用GS(gram-schmidt pan sharpening)融合方法,融合之后的影像空间分辨率得到明显提升,光谱曲线整体升高.3)对融合后的2景影像进行镶嵌处理,根据研究区行政矢量边界进行掩膜和裁剪,最终得到研究区的遥感影像.

**2.1.2 样地数据处理** 对调查得到的样地数据内的所有乔木进行生物量计算: 1) 利用福建省二元材积表测算样地内每株树的材积, 分别进行汇总后得到杉木和马尾松各个样地的总蓄积量; 2) 根据方精云等<sup>[11]</sup>对森林蓄积量和生物量关系的研究, 得出各个样地的乔木生物量; 3) 对得到的样地乔木生物量进行统计分析, 得到研究区样地乔木生物量的统计结果, 如表 1 所示。

表 1 研究区样地乔木生物量统计结果

因变量	统计指标					
	最小值	最大值	平均值	标准差	偏度	峰度
杉木生物量/(t·hm <sup>-2</sup> )	23.26	456.85	132.91	101.75	0.89	-0.01
马尾松生物量/(t·hm <sup>-2</sup> )	16.10	390.70	81.12	81.12	0.31	0.02

## 2.2 遥感因子提取

**2.2.1 光谱因子** 利用 ArcGIS 10.3 软件, 将样地中心位置点的数据叠加在预处理之后的 GF-2 影像上, 包括外业调查得到的 GPS 数据和二类调查数据中的样地定位数据, 读取样地所在像元的 4 个波段数据, 并计算 4 个波段的反射率倒数, 将此 8 个单波段因子作为生物量建模的可选因子。将 GF-2 影像的第一波段记为  $b_1$ , 8 个单波段因子依次为  $b_1$ 、 $b_2$ 、 $b_3$ 、 $b_4$ 、 $1/b_1$ 、 $1/b_2$ 、 $1/b_3$ 、 $1/b_4$ 。由于不同波段的组合可以在一定程度上削弱非植被信息的干扰、增强植被信息<sup>[12]</sup>, 并且可以有效区分土壤和植被<sup>[13]</sup>, 所以将得到的 4 个单波段反射率进行一定的组合, 得到的波段组合信息作为生物量建模的可选因子, 分别为  $b_1+b_2$ 、 $b_1+b_3$ 、 $b_1+b_4$ 、 $b_2+b_3$ 、 $b_2+b_4$ 、 $b_3+b_4$ 、 $b_1/b_2$ 、 $b_1/b_3$ 、 $b_1/b_4$ 、 $b_2/b_3$ 、 $b_2/b_4$ 、 $b_3/b_4$ 、 $(b_1+b_2)/b_3$ 、 $(b_1+b_2)/b_4$ 、 $(b_1+b_3)/b_2$ 、 $(b_1+b_3)/b_4$ 、 $(b_1+b_4)/b_2$ 、 $(b_1+b_4)/b_3$ 、 $(b_2+b_3)/b_1$ 、 $(b_2+b_3)/b_4$ 、 $(b_2+b_4)/b_1$ 、 $(b_2+b_4)/b_3$ 、 $(b_3+b_4)/b_1$ 、 $(b_3+b_4)/b_2$ 。

**2.2.2 植被指数** 植被指数是利用卫星不同波段探测数据组合而成的, 可以增强植被某一特性或者细节, 进而更好地反映植被生长状况<sup>[14]</sup>。根据提取出的光谱波段信息计算其植被指数, 共有 9 种植被指数作为建模因子<sup>[14-23]</sup>, 分别为归一化植被指数 (NDVI)、差值植被指数 (DVI)、比值植被指数 (RVI)、土壤调节植被指数 (SAVI)、修正型土壤植被指数 (MSAVI)、转化型植被指数 (TVI)、大气阻抗植被指数 (ARVI)、增强植被指数 (EVI)、绿通道植被指数 (GNDVI)。

**2.2.3 纹理因子** 采用灰度共生矩阵法, 利用 ENVI 5.3.1 软件提取 GF-2 影像的纹理特征, 提取出常用的 8 个纹理特征统计量, 即均值、方差、协同性、对比度、相异性、信息熵、二阶矩及相关性<sup>[23]</sup>, 将每个波段的 8 种纹理特征提取出的信息作为建模因子, 共

32 个纹理因子。

**2.2.4 地形因子** 地形因子是生物量建模估测的重要变量因子<sup>[24]</sup>。利用 ArcGIS 10.3 软件, 根据研究区的 DEM 数据, 提取出坡向、坡度和海拔 3 种地形因子。

**2.3 遥感因子相关性筛选** 根据提取出来的遥感因子, 将其与样地生物量进行 Pearson 相关性分析, 使用皮尔森相关系数判断生物量与自变量因子的线性相关程度<sup>[25]</sup>。该过程在 SPSS 24.0 软件中进行, 分析生物量与遥感因子间的相关关系, 选择出与生物量有显著性相关且相关性较高的因子作为最终的建模因子, 如表 2 所示。

表 2 生物量与自变量因子相关性

树种	因子	相关系数
杉木	$b_4$	-0.634**
	$b_3+b_4$	-0.430*
	DVI	-0.424*
	$b_1\_Var$	-0.323*
	$b_1\_Hom$	-0.450**
	$b_1\_Dis$	0.544**
	$b_2\_Hom$	-0.341*
	$b_2\_Dis$	0.543**
	$b_3\_Var$	0.324*
	$b_3\_Con$	0.536**
	$b_3\_Dis$	0.435*
马尾松	海拔	0.582**
	$b_1$	0.422*
	$b_1+b_2$	0.421*
	$b_1+b_4$	0.455*
	$b_2+b_4$	0.437*
	$b_3+b_4$	0.429*
	$b_3\_Var$	-0.624**
	$b_3\_Hom$	0.538**
	海拔	0.332*

注:  $b_1\_Var$ 、 $b_3\_Var$ 表示第一波段和第三波段的方差,  $b_1\_Hom$ 、 $b_2\_Hom$ 和 $b_3\_Hom$ 分别表示第一波段、第二波段和第三波段的协同性,  $b_3\_Con$ 表示第三波段的对比度,  $b_1\_Dis$ 、 $b_2\_Dis$ 和 $b_3\_Dis$ 分别表示第一波段、第二波段和第三波段的相异性; \*\*表示在 0.01 水平 (双侧) 上显著相关, \*表示在 0.05 水平 (双侧) 上显著相关。

**2.4 模型构建** 根据筛选的遥感因子和样地生物量构建生物量估测模型, 通过随机抽样选取 70% 的样地数据参与模型构建, 剩余 30% 的数据作为检验样地。使用 R-3.5.3 软件构建了 3 种模型, 即支持向量机、随机森林和多元逐步回归模型。模型构建后, 进行精度评价, 选取决定系数  $R^2$ 、均方根误差 RMSE 和估测精度 EA 作为评判指标, 通过这 3 个指标筛选出最佳模型<sup>[26]</sup>。

## 3 结果与分析

### 3.1 生物量模型构建

**3.1.1 杉木样地生物量模型构建** 根据杉木样地乔木生物量与自变量因子相关性分析, 选取显著水平

在 0.01 时显著相关的自变量因子,多元线性拟合之后得到的逐步回归模型为

$$Y_1 = 208.0556 + 4.9083X_1 - 0.7368X_2,$$

式中  $Y_1$  为杉木样地的乔木生物量,  $X_1$  为杉木  $b3\_Con$  (第三波段的纹理对比度),  $X_2$  为杉木的第四波段信息.

模型的  $R^2$  值为 0.6148, 模型选入的因子包含了单波段信息和纹理因子, 说明对于杉木样地同时结合影像的光谱和纹理信息, 对乔木生物量的估测效果最好.

**3.1.2 马尾松样地生物量模型构建** 同样进行多元线性拟合, 得到的马尾松逐步回归模型为

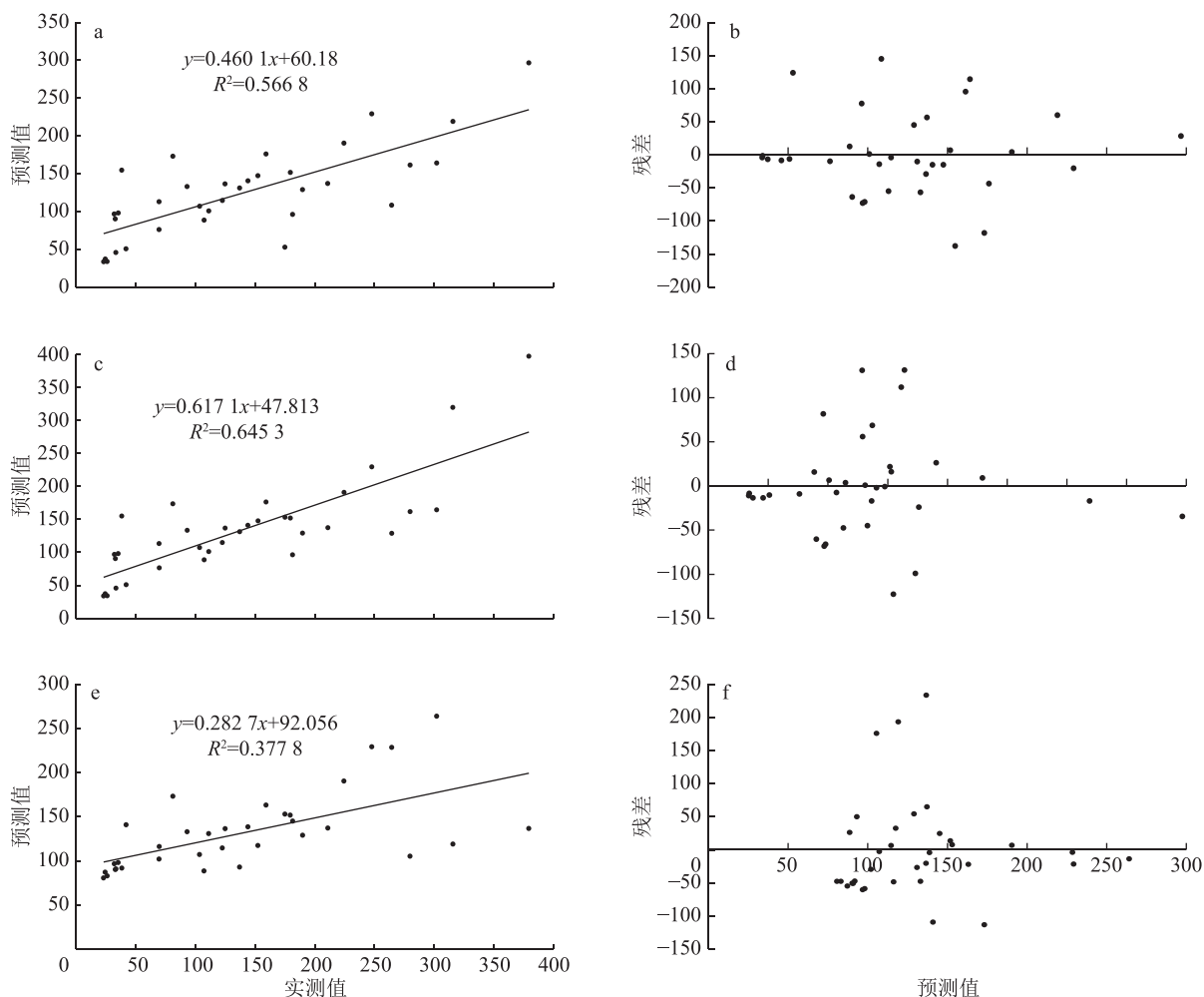
$$Y_2 = 133.9319 - 11.5619X_3 + 0.3181X_4,$$

式中  $Y_2$  为马尾松样地的乔木生物量,  $X_3$  为马尾松  $b3\_Var$  (第三波段的纹理方差),  $X_4$  为马尾松 2 个波段的组合  $b2+b4$ .

模型的  $R^2$  值为 0.6869, 模型选入的因子包含了纹理因子和多波段组合信息, 说明对于马尾松样地

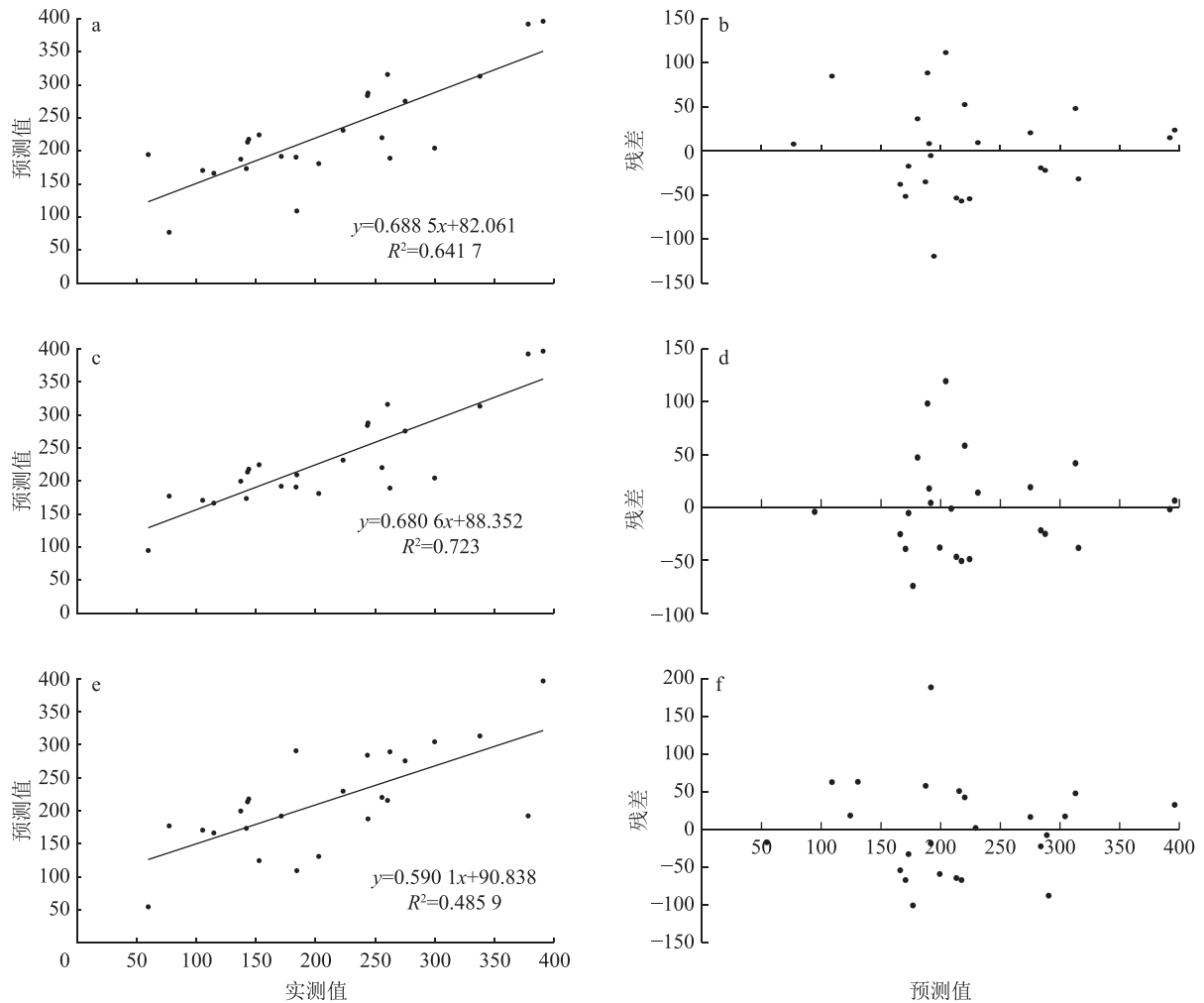
同时结合影像的光谱和纹理信息, 对乔木生物量的估测效果最好.

**3.2 模型精度评价** 根据样地实际分布情况, 采用随机抽样在 112 个杉木纯林中选择 78 个样地, 在 80 个马尾松纯林中选择 56 个样地, 分别构建了支持向量机模型、随机森林模型以及多元逐步回归模型, 并均取得较好的结果. 为了进一步判定哪个模型的拟合效果最佳、检验出最优估测模型, 以及模型是否具有推广性, 使用剩余 30% 的样地即 34 个杉木纯林样地和 24 个马尾松纯林, 对模型进行检验<sup>[27]</sup>. 根据 3 种方法所建模型来算出检验样地的生物量估测值, 通过实测值和估测值的拟合效果的决定系数 ( $R^2$ ) 和所建模型的残差, 检验模型预测值和实际观测值是否存在较好的拟合效果, 并计算它们的均方根误差和估测精度. 杉木样地 3 种模型估测结果如图 3 所示, 马尾松样地 3 种模型估测结果如图 4 所示. 将 3 种方法建立的模型检验结果汇总, 结果见表 3.



a-b. 支持向量机; c-d. 随机森林; e-f. 多元逐步回归.

图 3 杉木样地 3 种模型精度评价



a~b. 支持向量机; c~d. 随机森林; e~f. 多元逐步回归.

图 4 马尾松样地 3 种模型精度评价

表 3 3 种模型精度评价比较

树种	模型	$R^2$	RMSE	EA/%
杉木	支持向量机	0.57	64.27	53.67
	随机森林	0.65	48.16	65.28
	多元逐步回归	0.38	77.03	44.47
马尾松	支持向量机	0.64	54.79	73.64
	随机森林	0.72	48.18	76.82
	多元逐步回归	0.49	65.63	68.42

由图 3 可知,除了少数几个样点的预测值与实测值相差较大外,大部分实测值与预测值相近,可能是由于 GF-2 影像与样地调查数据有匹配误差<sup>[28]</sup>.对于杉木样地,3 种模型的实测值和预测值的决定系数  $R^2$  分别为 0.57、0.65 和 0.38,其中,随机森林的决定系数最高,并且其残差图的散点分布较均匀,即随机森林模型的拟合效果最好<sup>[29]</sup>.由图 4 可知,对于马尾松样地,3 种模型的实测值和预测值的决定系数  $R^2$  分别为 0.64、0.72 和 0.49,同样为随机森林模型的

拟合效果最好.

由表 3 可知:对于杉木样地,支持向量机模型的估测精度为 53.67%,均方根误差为 64.27;随机森林模型的估测精度为 65.28%,均方根误差为 48.16;多元逐步回归模型的估测精度为 44.47%,均方根误差为 77.03.对于马尾松样地,支持向量机模型的估测精度为 73.64%,均方根误差为 54.79;随机森林模型的估测精度为 76.82%,均方根误差为 48.18;多元逐步回归模型的估测精度为 68.42%,均方根误差为 65.63.

从结果来看,虽然随机森林模型对于乔木生物量估测的效果最好,但在本研究中精度并不高:1)可能与样地生物量的计算有关,本研究是基于蓄积量-生物量转换方程计算的,其准确性不如标准木法高,在今后研究中可以尝试样地调查时采伐过的标准木数据;2)与影像的纹理特征提取有关,本研究所选取的建模因子较多,但对于纹理特征没有进行深入研究,关于不同窗口能否有效提高模型的精度,还有待进一步研究.

## 4 结论与讨论

以福建省将乐林场为研究区,使用 GF-2 遥感数据,结合野外实测样地数据和福建省二类调查数据,采用支持向量机、随机森林以及多元逐步回归 3 种方法分别建立了杉木和马尾松生物量模型,并对模型进行精度评价,筛选出最优模型.研究结论如下:

1)充分利用了影像信息,提取了 GF-2 的光谱信息、植被指数、纹理特征以及地形因子等遥感信息,其中,在杉木样地中使用了 2 个波段信息、1 个植被指数、8 个纹理特征以及 1 个地形因子,在马尾松样地中使用了 5 个波段信息、2 个纹理特征以及 1 个地形因子它们参与建模,均取得了较好的拟合效果.

2)不论是杉木样地还是马尾松样地,2 种机器学习模型拟合效果均比多元逐步回归模型好,其中随机森林模型决定系数  $R^2$  最高、估计精度最高,并且其均方根误差最低.

从以上结论可知,利用 GF-2 影像信息进行乔木生物量估测是可行的,且随机森林模型的拟合度最高.但在研究过程中也有一些不足之处:在构建多元逐步回归模型时,在备选因子中选取了与乔木生物量呈显著相关的几个遥感因子,但最后只选了 2 个自变量因子参与建模,其原因可能是由于模型在筛选因子的条件基于诸多假设之上,而在现实中,较多条件都不易被满足;只使用了单一模型进行生物量建模,随着技术的发展,在今后的研究中可以加入集成学习的思想<sup>[30]</sup>,将多种模型进行结合,以便得到更好的精度,为后续的研究提供便利.

## 5 参考文献

- [1] 邹文涛,陈绍志,赵荣.森林生态系统碳储量及碳通量遥感监测研究进展[J].世界林业研究,2017,30(5):1
- [2] BROWN S. Estimating biomass and biomass change of tropical forests: a primer[J]. Fao Forestry Paper, 1997, 18: 45
- [3] 薛巍,张秋良,赵鹏翔,等.基于遥感技术的森林生物量估算应用[J].水土保持研究,2009,16(2):209
- [4] SCOTT G, RALPH D. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change[J]. Carbon Management, 2011, 2(3): 231
- [5] 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
- [6] SUN W Y, MU X M, SONG X Y, et al. Changes in extreme temperature and precipitation events in the Loess Plateau (China) during 1960-2013 under global warming[J]. Atmospheric Research, 2016, 168: 33
- [7] 曾晶,张晓丽.高分一号遥感影像下崂山林场林分生物量反演估算研究[J].中南林业科技大学学报,2016,36(1):46
- [8] 苟睿坤,陈佳琦,段高辉,等.基于GF-2的油松人工林地上生物量反演[J].应用生态学报,2019,30(12):4031
- [9] 徐梦伶,林辉,孙华,等.3种模型在GF-2影像的生物量估测中的比较[J].中南林业科技大学学报,2018,38(1):62
- [10] 黄绍霖,徐涵秋,林娜,等.亚热带地区马尾松林碳储量的遥感估算:以长汀河田盆地为例[J].生态学报,2013,33(10):2992
- [11] 方精云,刘国华,徐嵩龄.我国森林植被的生物量和净生产量[J].生态学报,1996,16(5):497
- [12] ZHOU Y C, ZHANG Y Q, VAZE J, et al. Improving runoff estimates using remote sensing vegetation data for bushfire impacted catchments[J]. Agricultural & Forest Meteorology, 2013, S182/183: 332
- [13] 徐梦伶.基于GF-2影像的森林生物量估测研究[D].长沙:中南林业科技大学,2018
- [14] HASHEMI S A, CHAI M M F, BAYAT S. An analysis of vegetation indices in relation to tree species diversity using by satellite data in the northern forests of Iran[J]. Arabian Journal of Geosciences, 2013, 6(9): 3363
- [15] HUETE A R. A soil-adjusted vegetation index (SAVI)[J]. Remote Sensing of Environment, 1988, 25(3): 295
- [16] Miura T, Huete A, Yoshioka H. An empirical investigation of cross-sensor relationships of NDVI and red/near-infrared reflectance using EO-1 Hyperion data[J]. Remote Sensing of Environment, 2006, 100(2): 223
- [17] ROUSE J W J R, HAAS R H, SCHELL J A, et al. Monitoring vegetation systems in the great plains with ERTS[C]//Third ERTS Symposium. Washington D C: NASA, 1973: 309
- [18] QI J, HUETE A R. Interpretation of vegetation indices derived from multi-temporal SPOT images[J]. Remote Sensing of Environment, 1993, 44: 89
- [19] KAUFMAN Y J, TANRE D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS[J]. IEEE Transactions on Geoscience & Remote Sensing, 1992, 30(2): 261
- [20] LYON J G, YUAN D, LUNETTA R S, et al. A change detection experiment using vegetation indices[J]. Photogrammetric Engineering and Remote Sensing, 1998, 64(2): 143
- [21] SCHLERF M, ATZBERGER C, HILL J. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data[J]. Remote Sensing of Environment, 2005, 95(2): 177

- [22] JORDAN C F. Derivation of leaf-area index from quality of light on the forest floor[J]. *Ecology*, 1969, 50(4): 663
- [23] 刘丽, 匡纲要, 等. 图像纹理特征提取方法综述[J]. *中国图像图形学报*, 2009, 14(4): 622
- [24] 李粉玲, 李京忠, 张琦翔, 等. DEM提取坡度·坡向算法的对比研究[J]. *安徽农业科学*, 2008, 36(17): 7355
- [25] 房秀凤, 谭炳香, 刘清旺, 等. 基于纹理信息的森林蓄积量估计[J]. *东北林业大学学报*, 2017, 45(11): 21
- [26] 孙华, 鞠洪波, 张怀清, 等. 三种回归分析方法在 Hyperion影像LAI反演中的比较[J]. *生态学报*, 2012, 32(24): 7781
- [27] 杨妍婷. 基于高分遥感数据的森林郁闭度估测方法研究[J]. *智能计算机与应用*, 2019, 9(2): 67
- [28] 陈日东, 林什全, 潘国英, 等. 天堂山林场森林地上生物量及碳储量的遥感估算模型构建[J]. *林业与环境科学*, 2019, 35(3): 71
- [29] 邹琪. 深圳市植被碳储量遥感估算研究[D]. 南京: 南京林业大学, 2017
- [30] 刘兆华, 林辉, 龙江平, 等. 基于高分二号的旺业甸林场蓄积量估测模型研究[J]. *中南林业科技大学学报*, 2020, 40(3): 79

## Estimation of tree biomass with GF-2

DING Zhidan SUN Yujun<sup>†</sup> SUN Zhao

(State Forestry & Grassland Administration Key Laboratory of Forest Resources & Environmental Management, Beijing Forestry University, 100083, Beijing, China)

**Abstract** Biomass data from a total of 192 plots (112 pure forests of Chinese fir, 80 pure forests of *Pinus massoniana*) in Jiangle State Forest Farm in Sanming City, Fujian Province were obtained from field measured sample plot data and second-class survey data of Fujian Province. Two scene GF-2 images from the study area were preprocessed, spectral information, vegetation index, texture features and topographic factors were extracted, factors highly-correlated with the biomass as independent variables were screened out. Biomass models of fir and *Pinus massoniana* were established from support vector machine, random forest and multiple stepwise regressions. Fitting of the two machine learning models was found to be better than the multiple stepwise regression model. The random forest model showed the highest determination coefficient  $R^2$  (0.65 and 0.72 for the 2 plots), and the highest estimation accuracy (65.28% and 76.82%). The mean root square errors in the 3 models for the Chinese fir plot were 64.27, 48.16 and 77.03. The mean root square errors in the three models for the *Pinus massoniana* plot were 54.79, 48.16 and 65.63, with the random forest model showing the lowest value. It is concluded that the random forest model is the most optimal among all three models.

**Keywords** tree biomass; GF-2; support vector machine; random forest; multiple stepwise regression