

DOI: 10.5846/stxb202012173204

蒋馥根, 孙 华, 李成杰, 马开森, 陈松, 龙江平, 任蓝翔. 联合 GF-6 和 Sentinel-2 红边波段的森林地上生物量反演. 生态学报, 2021, 41(20): 8222-8236.

Jiang F G, Sun H, Li C J, Ma K S, Chen S, Long J P, Ren L X. Retrieving the forest aboveground biomass by combining the red edge bands of Sentinel-2 and GF-6. Acta Ecologica Sinica, 2021, 41(20): 8222-8236.

联合 GF-6 和 Sentinel-2 红边波段的森林地上生物量反演

蒋馥根^{1,2,3}, 孙 华^{1,2,3,*}, 李成杰⁴, 马开森^{1,2,3}, 陈 松^{1,2,3}, 龙江平^{1,2,3}, 任蓝翔⁵

1 中南林业科技大学 林业遥感信息工程研究中心, 长沙 410004

2 林业遥感大数据与生态安全湖南省重点实验室, 长沙 410004

3 南方森林资源经营与监测国家林业与草原局重点实验室, 长沙 410004

4 广西壮族自治区森林资源与生态环境监测中心, 南宁 530000

5 国家林业和草原局中南调查规划设计院, 长沙 410004

摘要: 光谱反射率能反映地物差异, 是森林地上生物量 (Aboveground Biomass, AGB) 遥感反演的理论基础。红边波段处于近红外与红光波段交界处快速变化的区域, 能对植被冠层结构和叶绿素含量的微小变化做出快速反应, 对植被生长状况较敏感。研究以 GF-6 和 Sentinel-2 多光谱影像作为数据源, 结合野外调查 AGB 数据, 构建落叶松和樟子松 AGB 线性和非线性估测模型, 通过比较模型精度选择最优模型进行森林 AGB 反演和空间分布制图。结果表明: GF-6 和 Sentinel-2 影像红边波段反射率与落叶松、樟子松 AGB 均呈显著相关 ($P < 0.05$), 红边波段对 AGB 估测较敏感。多变量估测模型整体估测效果优于单变量模型, 所有模型中多元线性回归模型取得了最优的决定系数 (落叶松 $R^2 = 0.66$, 樟子松 $R^2 = 0.65$) 和最低的均方根误差 (落叶松 RMSE = 31.45 t/hm^2 , 樟子松 RMSE = 54.77 t/hm^2)。相比单个数据源, 联合 GF-6 和 Sentinel-2 影像构建的多元线性回归模型估测效果得到了显著提升, 模型 RMSE 对于落叶松和樟子松 AGB 估测分别最大降低了 22.9% 和 11.2%。增加红边波段进行 AGB 估测能显著提高模型估测精度, 三组数据源分别加入红边波段信息后进行建模, 模型 RMSE 得到了显著降低。GF-6 拥有 800 km 观测幅宽和高效的重访周期, 可以快速地提供大尺度时间序列数据, 在森林地上生物量反演和动态监测方面有着很大潜力。

关键词: 森林地上生物量; 落叶松; 樟子松; 遥感反演; GF-6; Sentinel-2

Retrieving the forest aboveground biomass by combining the red edge bands of Sentinel-2 and GF-6

JIANG Fugen^{1,2,3}, SUN Hua^{1,2,3,*}, LI Chengjie⁴, MA Kaisen^{1,2,3}, CHEN Song^{1,2,3}, LONG Jiangping^{1,2,3}, REN Lanxiang⁵

1 Research Center of Forestry Remote Sensing & Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China

2 Key Laboratory of Forestry Remote Sensing Based Big Data & Ecological Security for Hunan Province, Changsha 410004, China

3 Key Laboratory of National Forestry & Grassland Administration on Forest Resources Management and Monitoring in Southern Area, Changsha 410004, China

4 Forest Resources and Ecological Environment Monitoring Center of Guangxi Zhuang Autonomous Region, Nanning 530000, China

5 Central South Investigation and Planning Institute of State Forestry and Grassland Administration, Changsha 410004, China

Abstract: Spectral reflectance can reflect the difference of ground objects, which is the theoretical basis of forest

基金项目: 国家自然科学基金面上项目 (31971578); 湖南省教育厅重点项目 (17A225); 湖南省普通高校青年骨干教师培养对象项目 (90102-7070220090001); 中南林业科技大学研究生科研创新项目 (CX202101005)

收稿日期: 2020-12-17; 采用日期: 2021-08-16

* 通讯作者 Corresponding author. E-mail: sunhua@csuft.edu.cn

aboveground biomass (AGB) in remote sensing inversion. Red edge band is located in the special wavelength and scope which change fast at the junction of near-infrared and red bands, which can quickly respond to the small changes of vegetation canopy structure and chlorophyll content. Compared with other bands, red edge band is more sensitive to vegetation growth and chlorophyll change. In this study, GF-6 and Sentinel-2 multispectral images were used as data sources, and the linear and nonlinear models were constructed on the basis of Larch and Scotch pine field survey data for AGB estimation. The estimation accuracy of all models was compared, and the model with the highest estimation accuracy was regarded as the optimal model for the final AGB mapping in the study area. The results showed that the red edge reflectance of GF-6 and Sentinel-2 images were both significantly correlated with AGB ($P < 0.05$), and red edge bands were more sensitive to AGB estimation than other bands. On the whole, the estimation effect of multivariable estimation model was better than that of univariate model. Multivariate linear regression (MLR) model obtained the highest determination coefficient ($R^2 = 0.66$ for Larch and 0.55 for Scotch pine) and the lowest root mean square error ($RMSE = 31.45 \text{ t/hm}^2$ for Larch and 54.77 t/hm^2 for Scotch pine). Compared with single data source (GF-6 or Sentinel-2), the estimation effect of multiple linear regression model constructed by combining GF-6 and Sentinel-2 images was significantly improved, and the RMSE of AGB estimation model for Larch and Scotch pine decreased by 22.9% and 11.2% in highest measure, respectively. Taking the red edge band as additional variables for AGB estimation can significantly improve the estimation accuracy and effect of the model. The RMSE of the model was significantly reduced by adding red edge band information to three groups of data sources which are GF-6, Sentinel-2 and Sentinel-2 combined with GF-6. GF-6 has an observation width of 800 km and efficient revisit period, which can provide large-scale time series data quickly. As a remote sensing data source, GF-6 has great potential in forest aboveground biomass estimation and dynamic monitoring.

Key Words: forest aboveground biomass; Larch; Scotch pine; remote sensing inversion; GF-6; Sentinel-2

森林是陆地生态系统的主体,也是生物圈中最大的碳储库,在全球碳循环和气候变化中起着至关重要的作用^[1]。森林地上生物量(Aboveground Biomass, AGB)作为评价森林服务功能和森林质量的主要指标之一,越来越受到人们的重视^[2]。准确地估测 AGB 是森林资源管理和生态系统动态监测的基础性工作^[3]。与传统的人工调查相比,遥感技术具有快速获取大尺度植被生长状况的潜力,为森林资源的监测与管理提供了有效的方法参考^[4-5]。通过抽样调查的方式获取少量野外样地数据,结合遥感影像变量进行反演已成为区域范围内生物量估测的一种有效方式^[6-7]。

光谱反射率能体现地物之间的差异,是森林参数遥感反演的理论基础。光学遥感影像是世界上覆盖范围最广、类型最多、时间序列最丰富的数据^[7]。光学遥感影像广泛应用于土地变化和森林资源监测,丰富的光谱信息能够有效反映植被的生长情况^[8-9]。然而,云层覆盖、数据条带缺失以及高植被覆盖引起的光谱饱和现象限制了光学遥感影像的进一步应用^[6,10]。与传统光学遥感成像手段相比,合成孔径雷达(SAR, Synthetic Aperture Radar)和激光雷达(LiDAR, Light Detection and Ranging)可以穿透植被冠层以获得林分或森林详细的垂直结构信息,理论上可以获得更为准确的森林参数估测结果^[11-14]。然而, SAR 必须在特定频率的波段中工作, Gleason 等^[7]发现这些波段不一定适合生物量估计。LiDAR 利用冠层高度模型(CHM, Canopy Height Model)获取研究区域的平均树高,通过构建估测模型实现对整个区域的森林生物量估算^[15]。但大范围的高分辨率数字高程模型(DEM, Digital Elevation Model)是获取准确的 CHM 的基础,这对于地形复杂的山区来说无疑是困难的^[16]。此外,主动式遥感数据成本较高,目前大区域内没有完整的高质量周期性数据覆盖,不利于开展大范围的森林监测^[10]。

目前,中分辨率光学数据仍然是大面积人工林 AGB 估测最常用的数据源^[7]。Sentinel-2 由于其携带的红边波段对植被生长状况较敏感,已被广泛应用于森林 AGB 估测^[17-18]。GF-6 作为携带红边波段的第一颗国产光学卫星,具有大幅提高对农业、林业、草原等植被资源监测的潜力^[19-21]。梁继等^[19]研究了红边特征在农作

物识别中的表现,结果表明通过红边波段构建的红边归一化植被指数在作物样本类别区分度上比传统 NDVI (Normalized Difference Vegetation Index) 更显著,GF-6 红边特征波段在作物识别中表现优越,使得识别精度显著提高。Yu 等^[21]证明了 GF-6 两个红边波段对于土地覆盖的变化检测可以达到 89% 的整体精度,与使用同期 GF-1 数据相比有显著提升。目前,GF-6 在森林资源定量估测上的研究较少,对于森林 AGB 估测的有效性仍需进一步验证。此外,单个光学数据源由于其光谱信息的限制,应用于生物量反演的效果始终受到限制。联合多个数据源进行森林地上生物量反演的方式越来越受到欢迎^[5,13]。

本研究以旺业甸林场和塞罕坝林场为研究区,基于 GF-6 和 Sentinel-2 影像分别提取遥感变量,探索联合数据源以及红边波段对 AGB 估测的效应。同时,结合野外实测样地数据构建落叶松和樟子松 AGB 线性和非线性估测模型,比较 GF-6、Sentinel-2、联合 GF-6 和 Sentinel-2 三种数据源在 AGB 估测的效果,选择估测效果最优的模型进行 AGB 反演和空间分布制图。通过对最优模型进行适用性评价,以期为人工林 AGB 遥感估测提供方法参考。

1 数据获取与方法

1.1 研究区概况

旺业甸林场位于内蒙古自治区赤峰市西南部(118°09′—118°30′E,41°21′—41°39′N),总面积 259.58 km² (图 1)。林场属中温带大陆性季风气候,海拔范围为 500—1890 m,温度和降水随海拔升高有明显变化。年均气温 4.2℃,无霜期 117 d,年平均日照 2913.3 h,年平均降水量为 400 mm。林场森林覆盖率约为 93%,主要树种为落叶松(*Larix gmelinii* (Rupr.) Kuzen.)、油松(*Pinus tabulaeformis* Carr.)、樟子松(*Pinus sylvestris*)和云杉(*Picea asperata* Mast.)等。落叶松是林场主要的木材产品,也是北方地区重要的生态资源,及时掌握落叶松生物量分布和动态变化对于了解中国北方地区森林生态系统碳循环具有重要的价值。

塞罕坝林场成立于 1962 年,总面积 27300 hm²,是全国最大的人工林林场。林场位于河北省承德市北部(116°51′—117°39′E,42°02′—42°36′N)(图 1)。年平均气温为-1.4℃,冬季漫长,生长季节短,平均海拔为 1700 m。塞罕坝属寒温带大陆性季风气候,极端最低气温-43.2℃,年平均积雪 7 个月。年平均降水量约 490 mm,年平均无霜期 68 d。林场目前森林覆盖率已达 80%,林场树种以樟子松(*Pinus sylvestris*)和落叶松(*Larix gmelinii* (Rupr.) Kuzen.)为主,林木总蓄积量达到 1012 万 m³。

1.2 数据获取及处理

1.2.1 野外 AGB 数据获取及处理

利用 eCognition 8.7 分类软件对覆盖旺业甸林场的 GF-6 PMS(数据获取时间为 2019 年 9 月 30 日)影像(2 m 分辨率)进行多尺度分割,结合 2018 年旺业甸林场植被分布图提取植被类型,最终确定林场内落叶松分布区域(图 2)。根据不同立地条件和生长状况在林场内落叶松分布区域共设置 42 个 25 m×25 m 的样地,并于 2019 年 9 至 10 月进行野外实测调查。选用 Trimble Geo 7x GPS 全球定位仪记录样地中心坐标。对样地内胸径大于 5 cm 的样木进行每木检尺,测量树高和胸径数据,同时记录样地海拔、坡度、坡向等信息。利用塞罕坝林场 2018 年森林资源调查数据库,提取樟子松边界后计算、统计所有小班蓄积量结果,并在边界内随机设置 100 个 25 m×25 m 的样地,最终确定所有样地的森林蓄积量(图 2)。

在全国标准信息公共服务平台(<http://std.samr.gov.cn/>)获取 2018 年内蒙古自治区人工落叶松二元立木材积公式,并计算所有样地内落叶松单木材积,累加得到样地蓄积量(m³)后换算成每公顷蓄积量(m³/hm²)。樟子松单位面积蓄积量通过小班蓄积量值直接转换。利用生物量转换因子连续函数法^[22]推算出每公顷地上生物量,参数 a 和 b 采用李海奎《中国森林植被生物量和碳储量评估》^[23]中落叶松和樟子松树种/树种组参数统计结果,落叶松取 0.6096 和 33.8060,樟子松取 1.0945 和 2.0040。

$$V_l = 0.002466 - 0.000014 \times D^2 + 0.001142 \times H + 0.000036 \times H \times D^2 \quad (1)$$

式中, V_l 为落叶松材积(m³), D 为胸径(cm), H 为树高(m)。

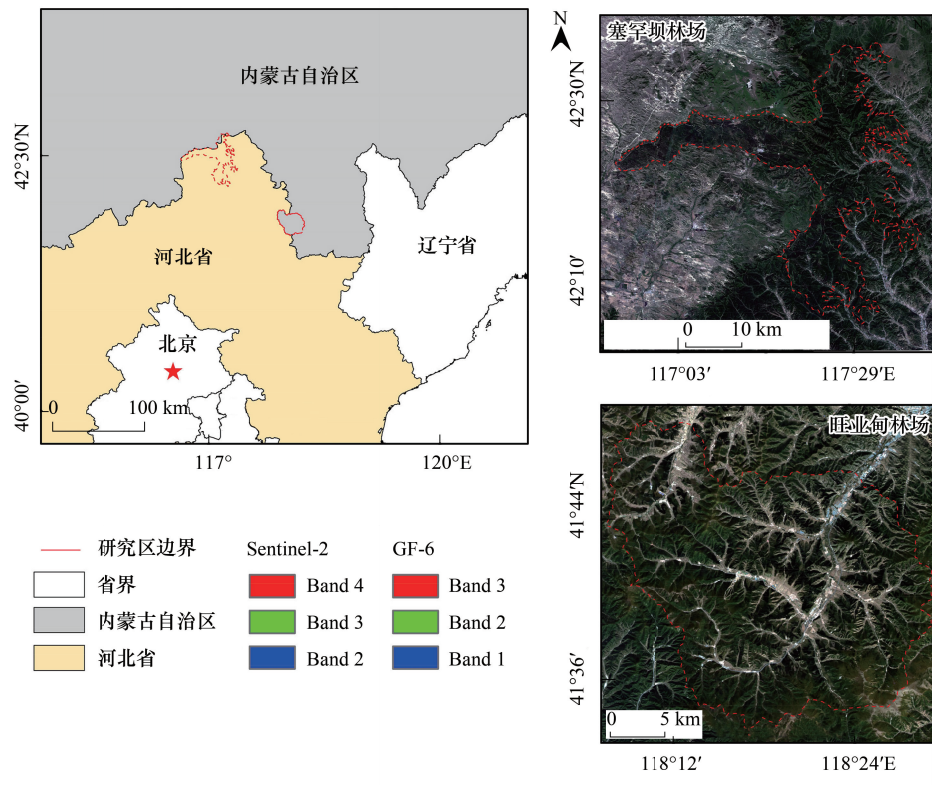


图 1 研究区位置示意图

Fig.1 Location of the study area

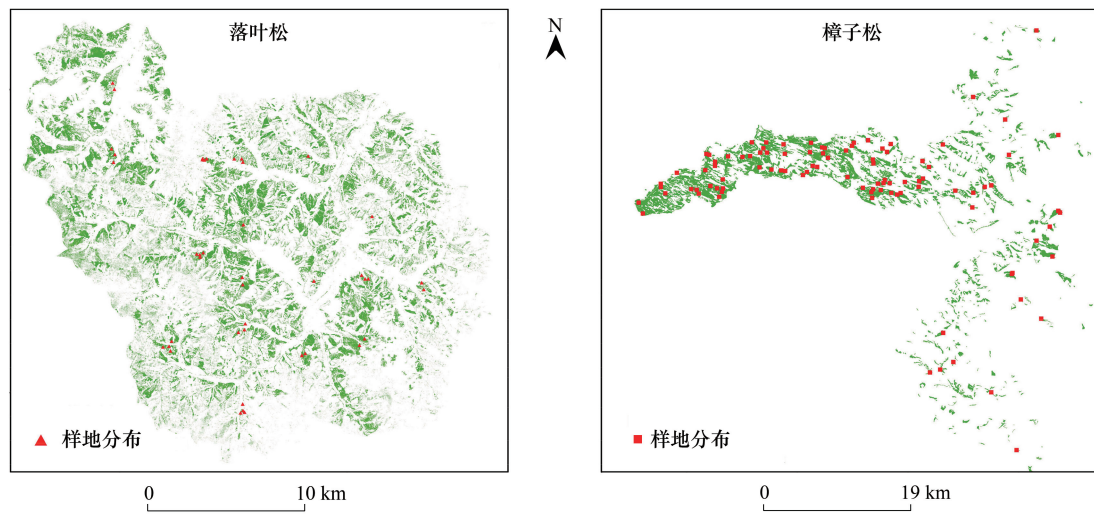


图 2 落叶松和樟子松边界及样地分布

Fig.2 Boundary and plot distribution of Larch and Scotch pine

$$B = aV_2 + b \quad (2)$$

式中, B 为每公顷生物量(t/hm^2), V_2 为每公顷蓄积量(m^3/hm^2), a 和 b 为常数。

落叶松和樟子松 AGB 分布范围分别为 $60.72\text{--}269.87 \text{ t}/\text{hm}^2$ 和 $30.06\text{--}414.23 \text{ t}/\text{hm}^2$ (表 1)。旺业甸林场落叶松样本标准差和变异系数分别为 $54.60 \text{ t}/\text{hm}^2$ 和 37.1% , 说明样本呈一定程度的离散分布。塞罕坝林场樟

子松样本标准差和变异系数分别为 92.50 t/hm²和 92.8%,样本变异程度较大。落叶松和樟子松 AGB 均值分别为 147.11 t/hm²和 99.65 t/hm²。

表 1 研究区森林生物量统计结果
Table 1 Forest aboveground biomass statistical results of the study area

研究区 Study area	树种 Specie	样地数量 Plot number	最小值 Minimum value	最大值 Maximum value	平均值 Mean value	标准差 Standard deviation	变异系数 Coefficient of variation/%
旺业甸林场 Wangyedian forest farm	落叶松	42	60.72	269.87	147.11	54.60	37.1
塞罕坝林场 Saihanba forest farm	樟子松	100	30.06	414.23	99.65	92.50	92.8

1.2.2 遥感影像数据处理

本研究获取的遥感影像数据包括覆盖研究区且与野外调查时间相近的 Sentinel-2 和 GF-6 影像。Sentinel-2 卫星携带一枚多光谱成像仪(Multispectral imager, MSI),卫星重访周期为 10 d,幅宽为 290 km^[17]。Sentinel-2 是唯一拥有三个以上红边波段的光学卫星,这对准确获取植被生长状况非常有效^[17]。为了获取更准确的植被信息,研究选择空间分辨率较高的 Band 2—Band 8A,共计 8 个波段。GF-6 是我国第一颗携带红边波段信息的光学遥感卫星,携带 WFV(Wide Field of View)和 PMS(Panchromatic and Multi-spectral)两个传感器,能够有效反映森林和植被生长特性(表 2)^[19]。GF-6 幅宽达 800 km,与其他高分系列卫星组网运行后重返时间减少至 2 天,能持续、高效地提供高分辨率影像,大幅提高对森林、草原等植被的监测能力^[20]。

利用 Sen2cor 2.5.5 和 ENVI 5.3 遥感图像处理软件分别对 Sentinel-2 和 GF-6 影像进行辐射定标、大气校正等预处理,最终将像元亮度值转换为反射率值^[17-18]。为了将样地与遥感数据进行匹配,获得更准确的植被信息,将两种影像像元大小均重采样至与样地大小一致。

表 2 研究采用的 Sentinel-2 和 GF-6 波段信息
Table 2 The selected bands of Sentinel-2 and GF-6 in the study

遥感影像 Image	波段 Band	中心波长 Center wavelength/nm	分辨率 Resolution /m	遥感影像 Image	波段 Band	中心波长 Center wavelength/nm	分辨率 Resolution /m
哨兵 2 号 Sentinel-2	Band 2-蓝波段	490	10	高分六号	Band 1-蓝波段	485	16
	Band 3-绿波段	560	10	GF-6	Band 2-绿波段	555	16
	Band 4-红波段	665	10		Band 3-红波段	660	16
	Band 5-红边波段 1	705	20		Band 4-近红外波段	830	16
	Band 6-红边波段 2	740	20		Band 5-红边波段 1	710	16
	Band 7-红边波段 3	783	20		Band 6-红边波段 2	750	16
	Band 8-近红外波段	842	10		Band 7-紫波段	425	16
	Band 8A-红边波段 4	865	20		Band 8-黄波段	610	16

1.2.3 遥感反演变量提取及筛选

光谱反射率和植被指数是建立森林地上生物量遥感反演模型的重要变量^[10]。植被指数能对植被生长状况进行有效地度量,合适的植被指数参与建模能显著提高 AGB 估测精度^[18]。作为常用的植被指数,归一化植被指数(Normalized Difference Vegetation Index, NDVI)^[24]、增强型植被指数(Enhanced Vegetation Index, EVI)^[25]、大气抗阻植被指数(Atmospherically Resistant Vegetation Index, ARVI)^[25]、可见光抗大气指数(Visible Atmospheric Resistant Index, VARI)^[25]和红绿植被指数(Red-Green Vegetation Index, RGVI)^[25]与植物生长周期中重要的生理参数之间有良好的相关关系,已被证明对森林 AGB 估测较敏感。红边波段处于近红外与红光波段之间快速变化的区域,对植被叶绿素的微小变化较敏感^[26]。研究将同时构建红边归一化植被指数(Red-Edge Normalized Difference Vegetation Index, RENDVI)^[26]和红边叶绿素指数(Red Edge Chlorophyll Index,

RECI)^[27]作为遥感变量参与建模,以评价红边植被指数对于 AGB 估测的效应(表 3)。

表 3 植被指数计算公式
Table 3 Formulas of the vegetation index

植被指数 Vegetation index	全称 Full name	公式 Expression
单波段反射率 BR ^[24]	Band reflectance	GF-6_Band i , ($i=1, 2, \dots, 8$); Sentinel-2_Band i , ($i=2, 3, \dots, 8A$)
归一化植被指数 NDVI ^[24]	Normalized different vegetation index	$(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red})$
增强型植被指数 EVI ^[25]	Enhanced vegetation index	$2.5 \times (\text{NIR}-\text{Red})/(\text{NIR}+6 \times \text{Red}-7.5 \times \text{Blue}+1)$
大气抗阻植被指数 ARVI ^[25]	Atmospherically resistant vegetation index	$\text{NIR}-(2 \times \text{Red}-\text{Blue})/\text{NIR}+(2 \times \text{Red}-\text{Blue})$
可见光抗大气指数 VARI ^[25]	Visible atmospheric resistant index	$(\text{Green}-\text{Red})/(\text{Green}+\text{Red}+\text{Blue})$
红绿植被指数 RGV ^[25]	Red-green vegetation index	$(\text{Red}-\text{Green})/(\text{Red}+\text{Green})$
红边归一化植被指数 RENDVI ^[26]	Red-edge normalized difference vegetation index	$(\text{Red Edge2}-\text{Red Edge1})/(\text{Red Edge2}+\text{Red Edge1})$
红边叶绿素指数 RECI ^[27]	Red-edge chlorophyll index	$(\text{NIR}/\text{Red Edge1})-1$

1.3 AGB 回归模型

研究同时建立线性和非线性回归模型用于研究区森林 AGB 估测和模型精度比较。为了验证变量数量对于模型估测的影响,建立单变量线性模型、单变量非线性模型(三次多项式)、多元线性回归^[28]和多变量非参数模型(随机森林)^[29]用于比较模型估测效果。相比单变量回归,多元线性回归模型能定量地、更全面地衡量多个遥感变量对于 AGB 估测的影响。逐步回归分析用于检测并保留与 AGB 显著相关的变量,同时结合方差膨胀因子 VIF(Variance Inflation Factor)指数消除变量间的共线性,最终保留的变量应当与 AGB 显著相关且变量之间相互独立^[24]。此外,非线性模型中的非参数模型对目标函数的形式没有提前假设,更适合于复杂数据的预测^[29]。随机森林作为非参数模型的代表,已成为 AGB 估测的常用方法^[8,10,17]。随机森林通过快速构造大量的决策树来估计 AGB,模型训练过程中,决策树之间相互独立,训练速度快。另外,随机森林可以评估模型中每个变量的重要性,从而有效地判断变量对模型的贡献^[17,29]。研究将通过评价变量的重要性,通过重要性排序筛选出对模型贡献度较大的变量组合进行 AGB 估测。

1.4 精度评价

为了验证 GF-6 和 Sentinel-2 影像对 AGB 估测的有效性,研究分为 Sentinel-2 变量组、GF-6 变量组以及联合 GF-6 和 Sentinel-2 变量组,分别建立回归模型进行 AGB 估测。同时,为了充分利用样本以提高模型的可靠性,研究选择留一交叉法^[25]对估测结果进行验证,即每次只留下 1 个样地做测试集,其他样地做训练集,重复 n 次后确定最终估测结果。选择决定系数(R^2)^[30]和均方根误差(RMSE)^[30]进行模型评价,计算方式如下:

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

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

式中, y_i 表示 AGB 实测值, \hat{y}_i 表示 AGB 模型预测值, \bar{y} 表示实测 AGB 的均值, n 表示样本数量。

2 结果与分析

2.1 变量相关系数矩阵与重要性分析

计算所有遥感变量与森林 AGB 之间的 Pearson 相关系数,并将相关系数矩阵可视化。由图 3 可知,从 GF-6 和 Sentinel-2 影像中提取的遥感变量中,光谱反射率与落叶松 AGB 存在较高的相关关系,且整体相关系数高于植被指数。其中,红边波段反射率均与 AGB 显著相关($P<0.05$)且相关性系数较高。GF-6 和 Sentinel-

2 影像中相关系数较大的变量主要由红边波段、近红外波段、蓝色波段和黄色波段组成。与落叶松不同的是, 樟子松植被指数相关系数整体大于光谱反射率(图 4)。以 GF-6 为数据源时, 与樟子松 AGB 相关性最高的变量为 RECI, 表现为极显著正相关($P < 0.01$)。

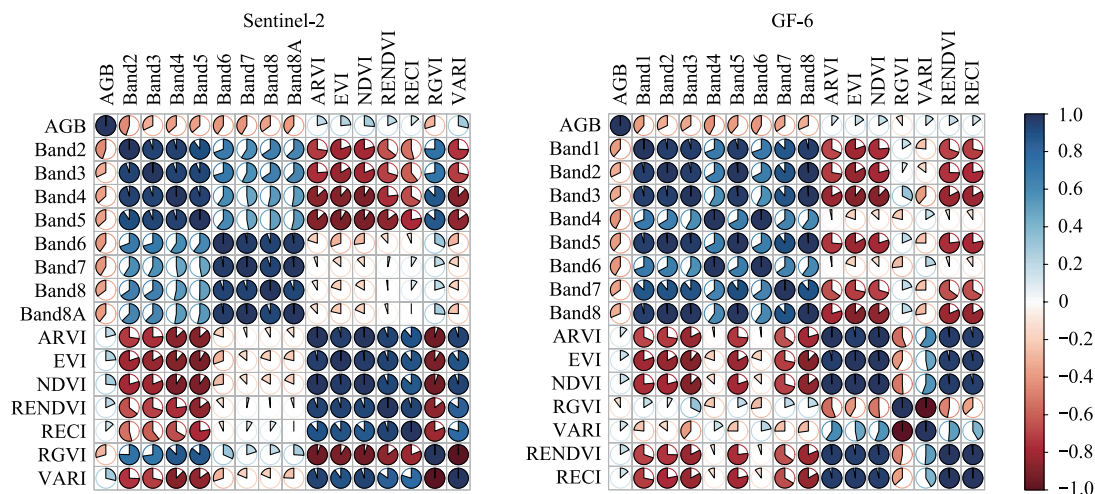


图3 Sentinel-2 与 GF-6 遥感变量与落叶松 AGB 相关系数矩阵图

Fig.3 Matrix diagram of correlation coefficient between remote sensing variables and AGB of Larch of Sentinel-2 and GF-6

AGB; 地上生物量 Aboveground Biomass; ARVI; 大气抗阻植被指数 Atmospherically Resistant Vegetation Index; EVI; 增强型植被指数 Enhanced Vegetation Index; NDVI; 归一化植被指数 Normalized Difference Vegetation Index; RGVI; 红绿植被指数 Red Green Vegetation Index; VARI; 可见光抗大气指数 Visible Atmospheric Resistant Index; RENDVI; 红边归一化植被指数 Red Edge Normalized Difference Vegetation Index; RECI; 红边叶绿素指数 Red Edge Chlorophyll Index

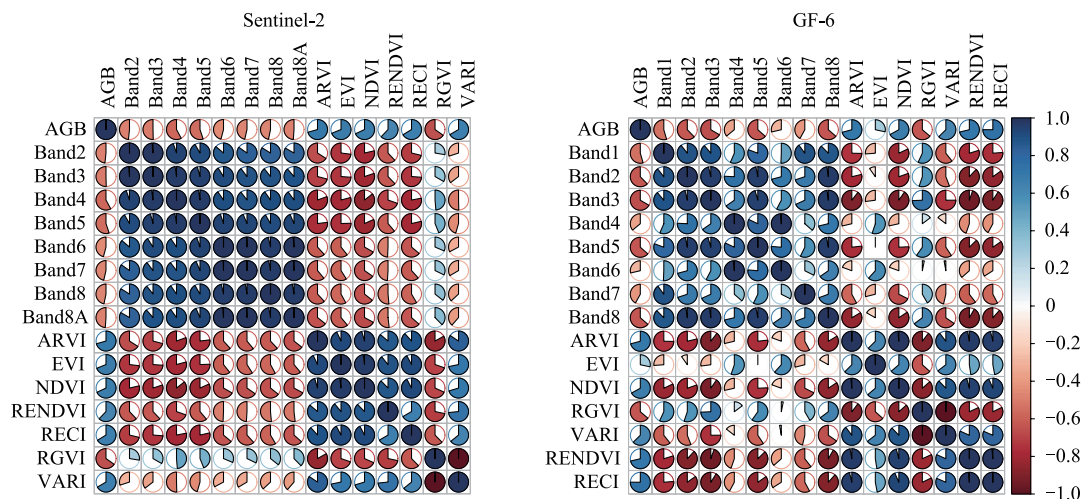


图4 Sentinel-2 与 GF-6 遥感变量与樟子松 AGB 相关系数矩阵图

Fig.4 Matrix diagram of correlation coefficient between remote sensing variables and AGB of Scotch pine of Sentinel-2 and GF-6

利用 R 语言 RandomForest 函数包分别计算变量的重要性, 按照重要性从高到低的顺序依次增加变量分别构建随机森林模型, 分析不同变量数量下模型 RMSE 的变化趋势。由图 5 和图 6 可知, 对于 GF-6、Sentinel-2 以及联合 GF-6 和 Sentinel-2 三类变量, 当变量数量分别取 3、10、2 和 5、10、10 时, 误差达到最小, 将此时的变量组合最终用于 AGB 估测。

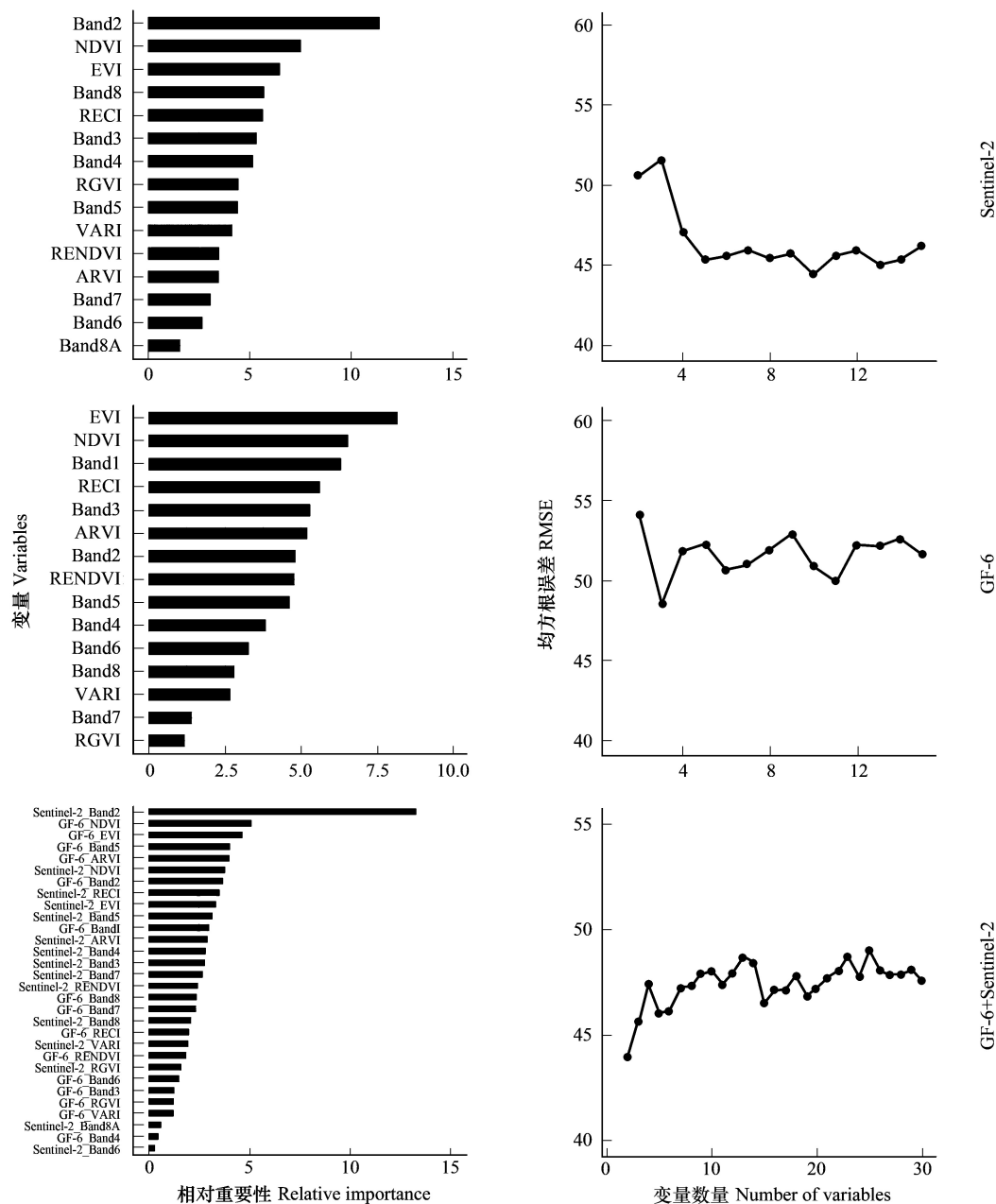


图5 旺业甸林场不同变量组的重要性排序和 RMSE 变化图

Fig.5 Importance ranking under different variable group and RMSE change in Wangyedian forest farm

RMSE: 均方根误差 Root Mean Square Error

2.2 AGB 反演

以 GF-6 和 Sentinel-2 作为数据源,结合旺业甸林场和塞罕坝林场 AGB 实测数据建立单变量线性模型、单变量三次多项式模型、多元线性回归和随机森林模型进行 AGB 估测和模型精度比较。结果如表 4 所示,单变量线性模型和三次多项式模型估测效果较差,不能有效地进行 AGB 估测。多变量模型中,多元线性回归在三组数据源中均优于随机森林模型,在落叶松和樟子松构建的模型中 RMSE 分别降低了 8.0%、19.3%、28.5% 和 11.6%、5.0%、4.4%。

利用逐步回归分别对三组变量进行变量筛选,最终获得的变量主要由红边波段、蓝色波段、近红外波段和黄色波段构成,说明这些波段信息对 AGB 具有较好的敏感性。相比 Sentinel-2,基于 GF-6 建立的多元线性回

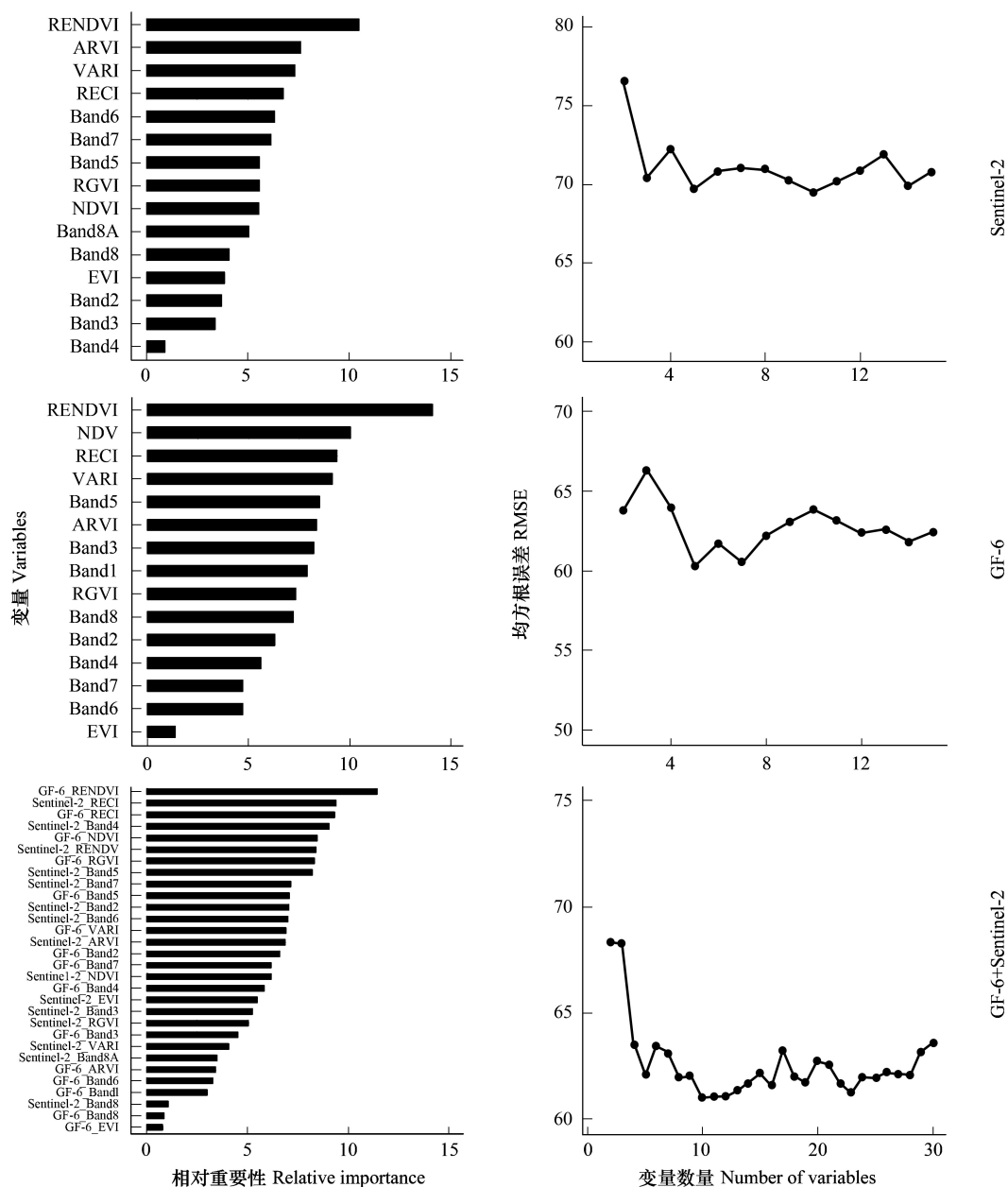


图6 塞罕坝林场不同变量组的重要性排序和 RMSE 变化图

Fig.6 Importance ranking under different variable group and RMSE change in Saihanba forest farm

归模型决定系数 R^2 有一定提高, 估测误差略有下降, 估测效果较优。联合 GF-6 和 Sentinel-2 影像构建的多元线性回归模型估测效果最好, 决定系数 R^2 分别达到了 0.66 和 0.65, RMSE 为 31.45 t/hm^2 和 54.77 t/hm^2 。多元线性回归模型在所有模型中估测效果最好, 将作为最优模型进行旺业甸林场落叶松和塞罕坝林场樟子松 AGB 反演和空间分布制图。

为了验证红边波段对落叶松、樟子松 AGB 估测的有效性, 将三组数据源剔除红边波段信息之后, 结合多元线性回归再次建立估测模型进行比较。结果如表 5 所示, 在剔除红边波段之后, 三组数据源在估测落叶松 AGB 模型精度显著下降, 模型 RMSE 分别提高了 9.6%、17.6% 和 28.7%, 樟子松分别提高了 6.2%、7.9% 和 8.2%, 说明加入红边波段能显著提高 AGB 估测精度。此外, 在剔除三组数据源的红边波段信息后, 联合 GF-6 和 Sentinel-2 的数据源相比 GF-6、Sentinel-2 估测效果仍然有显著提高, RMSE 显著降低。

表 4 AGB 回归估测结果
Table 4 AGB regression estimation results

研究区 Study area	数据源 Data source	模型 Model	变量 Variable	决定系数 R^2	均方根误差 RMSE/ (t/hm^2)
旺业甸林场	哨兵 2 号	单变量线性回归	Sentinel-2_Band2	0.20	48.23
		单变量非线性回归	Sentinel-2_NDVI	0.23	47.34
		随机森林	Sentinel-2 _ RENDVI, Sentinel-2 _ Band5, Sentinel-2 _ RECI, Sentinel-2_Band7, Sentinel-2 _ EVI, Sentinel-2 _ Band6, Sentinel-2 _ Band8A, Sentinel-2 _ RGVI, Sentinel-2 _ NDVI, Sentinel-2_VARI	0.32	44.39
		多元线性回归	Sentinel-2 _ RENDVI, Sentinel-2 _ VARI, Sentinel-2 _ Band2, Sentinel-2 _ Band3, Sentinel-2 _ Band7, Sentinel-2 _ Band8, Sentinel-2_Band8A	0.43	40.84
	高分六号	单变量线性回归	GF-6_Band4	0.18	48.83
		单变量非线性回归	GF-6_EVI	0.23	47.25
		随机森林	GF-6_RENDVI, GF-6_RECI, GF-6_Band5	0.21	48.59
		多元线性回归	GF-6 _ EVI, GF-6 _ RENDVI, GF-6 _ RGVI, GF-6_Band1, GF-6_Band6, GF-6_Band7, GF-6_Band8	0.47	39.23
	哨兵 2 号+ 高分六号	单变量线性回归	Sentinel-2_Band2	0.20	48.23
		单变量非线性回归	GF-6_EVI	0.23	47.25
		随机森林	Sentinel-2_RENDVI, GF-6_RENDVI	0.34	43.96
		多元线性回归	GF-6_RENDVI, GF-6_RGVI, GF-6_Band1, GF-6 _ Band5, GF-6 _ Band8, Sentinel-2 _ ARVI, Sentinel-2_NDVI, Sentinel-2_Band7, Sentinel-2_Band8	0.66	31.45
塞罕坝林场	哨兵 2 号	单变量线性回归	Sentinel-2_ARVI	0.49	65.16
		单变量非线性回归	Sentinel-2_ARVI	0.53	62.74
		随机森林	Sentinel-2 _ RENDVI, Sentinel-2 _ ARVI, Sentinel-2_VARI, Sentinel-2_RECI, Sentinel-2 _ Band6, Sentinel-2 _ Band7, Sentinel-2 _ Band5, Sentinel-2_RGVI, Sentinel-2_NDVI, Sentinel-2_Band8A	0.43	69.72
		多元线性回归	Sentinel-2_RECI, Sentinel-2_ARVI, Sentinel-2_VARI, Sentinel-2_Band2	0.55	61.65
	高分六号	单变量线性回归	GF-6_RECI	0.56	60.72
		单变量非线性回归	GF-6_RECI	0.59	58.35
		随机森林	GF-6_RENDVI, GF-6_NDVI, GF-6_RECI, GF-6_VARI, GF-6_Band5	0.57	60.26
		多元线性回归	GF-6_RECI, GF-6_RENDVI, GF-6_ARVI, GF-6_EVI	0.62	57.24
	哨兵 2 号+ 高分六号	单变量线性回归	GF-6_RECI	0.56	60.72
		单变量非线性回归	GF-6_RECI	0.59	58.35
		随机森林	GF-6 _ RENDVI, Sentinel-2 _ RECI, GF-6 _ RECI, Sentinel-2 _ Band4, GF-6 _ NDVI, Sentinel-2_RENDVI, GF-6_RGVI, Sentinel-2 _ Band5, Sentinel-2_Band7, GF-6_Band5	0.61	57.31
		多元线性回归	GF-6 _ RECI, GF-6 _ RENDVI, Sentinel-2 _ Band2, Sentinel-2_Band3, Sentinel-2_Band8, Sentinel-2_ARVI, Sentinel-2_RECI, Sentinel-2_RENDVI	0.65	54.77

表 5 剔除红边波段后多元线性回归估测结果

Table 5 Multiple linear regression estimation results after eliminating red edge band				
研究区 Study area	数据源 Data source	变量 Variable	决定系数 R^2	均方根误差 RMSE/ (t/hm^2)
旺业甸林场	哨兵 2 号	Sentinel-2_B2, Sentinel-2_B3	0.29	45.19
	高分六号	GF-6_B4, GF-6_EVI, GF-6_RGVI	0.22	47.61
	哨兵 2 号+高分六号	GF-6 _ Band2, GF-6 _ Band4, Sentinel-2 _ Band2, Sentinel-2_Band8	0.33	44.14
塞罕坝林场	哨兵 2 号	Sentinel-2_Band2, Sentinel-2_Band8, Sentinel-2_NDVI	0.49	65.82
	高分六号	GF-6_ARVI, GF-6_EVI	0.54	62.15
	哨兵 2 号+高分六号	GF-6_ARVI, Sentinel-2_VARI	0.57	59.67

2.3 模型拟合图及 AGB 空间分布

通过 Sentinel-2 和 GF-6 影像分别构建的多元线性回归模型均存在一定程度的高估和低估,但将两种影像联合后建立的回归模型误差显著降低,拟合效果达到最优(图 7、图 8)。联合 GF-6 和 Sentinel-2 建立的估测模型预测 AGB 值与实测 AGB 值具有良好的线性关系,绝大部分预测值随机分布在回归线两侧。

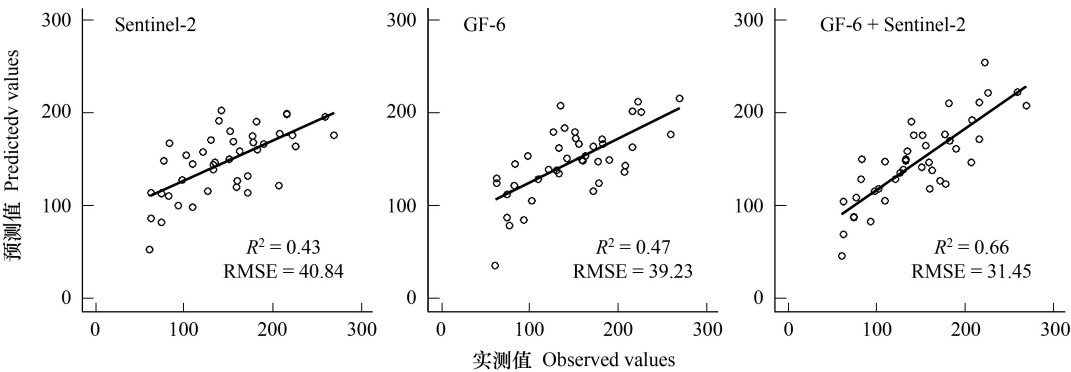


图 7 基于 Sentinel-2、GF-6 和联合数据源的落叶松 AGB 回归模型拟合图

Fig.7 Fitting diagram of AGB regression models of Larch based on Sentinel-2, GF-6 and federated data source

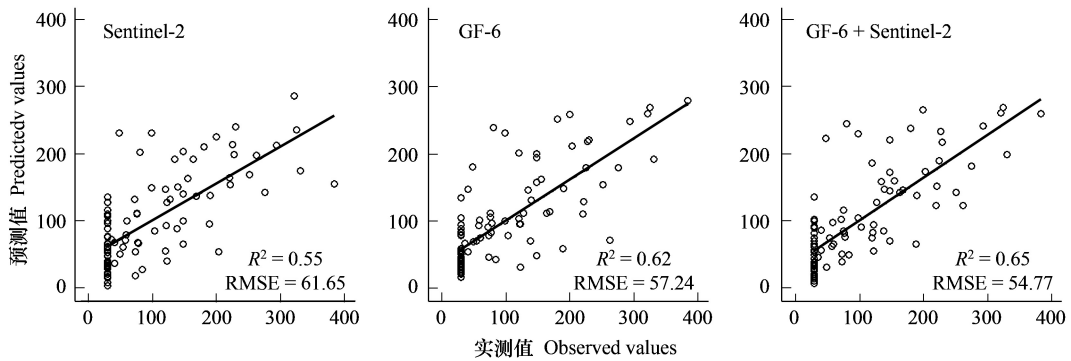


图 8 基于 Sentinel-2、GF-6 和联合数据源的樟子松 AGB 回归模型拟合图

Fig.8 Fitting diagram of AGB regression models of Scotch pine based on Sentinel-2, GF-6 and federated data source

基于 Sentinel-2、GF-6、联合 GF-6 和 Sentinel-2 筛选的变量通过构建多元线性回归分别得到了旺业甸林场落叶松 AGB 的空间分布图(图 9)。图 9 所示的 AGB 空间分布相似。在三种 AGB 空间分布中,落叶松 AGB

均值分别为 136.8 t/hm^2 、 128.7 t/hm^2 和 140.98 t/hm^2 , 与使用单个数据源相比, 联合 GF-6 和 Sentinel-2 数据源得到的 AGB 均值有一定的增加, 且最接近实测样地的 AGB 均值 147.11 t/hm^2 。较大的 AGB 预测值主要分布在西部和东南地区, 而北部和西南地区的预测值相对较小。由于靠近建筑物和道路, 人为干扰活动较多, 中部和西北地区的 AGB 值偏小。对于塞罕坝林场, 生物量值较大的值主要分布在西部和西北部地区, 联合数据源构建的模型预测效果较好 (图 10)。以联合 GF-6 和 Sentinel-2 为数据源预测的 AGB 空间分布与林场实际 AGB 分布基本一致, 可以作为有效的数据源进行 AGB 反演。

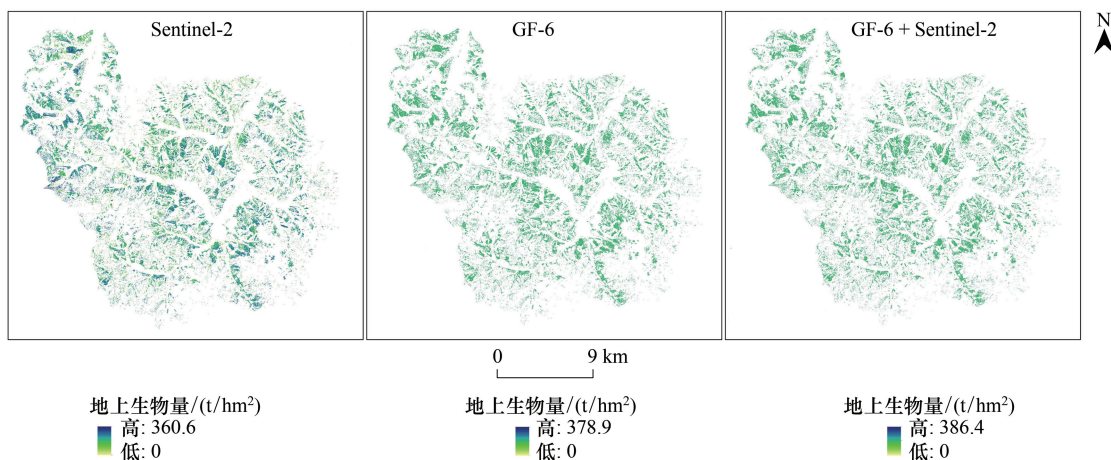


图 9 旺业甸林场落叶松 AGB 空间分布图

Fig.9 AGB Spatial distribution of Larch in Wangyedian forest farm

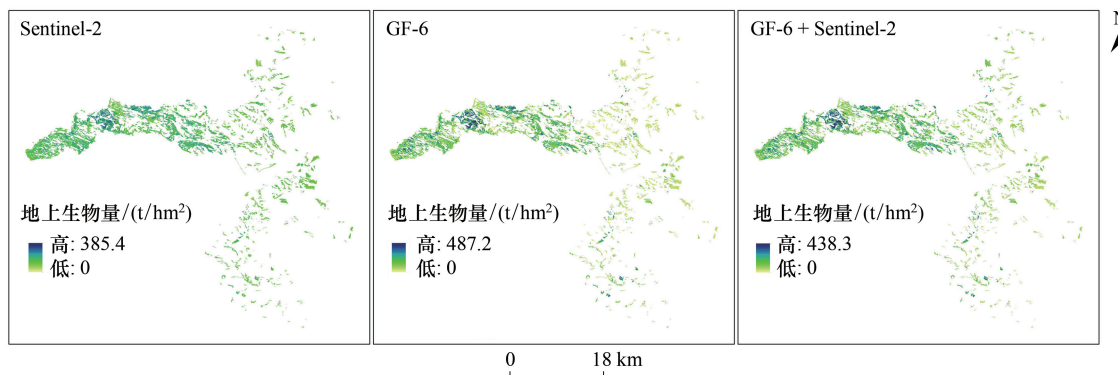


图 10 塞罕坝林场樟子松 AGB 空间分布图

Fig.10 AGB Spatial distribution of Scotch pine in Saihanba forest farm

2.4 模型适用性评价

研究通过建模变量与最优模型产生的残差之间的显著性检验以及残差的空间自相关检验进行模型的适用性评价。对于落叶松和樟子松, 所有建模变量与残差之间均不存在显著相关 ($P > 0.1$)。以落叶松为例, 构建 AGB 残差与建模变量之间的散点图 (图 11), 结果表明变量对 AGB 估测的不确定性没有显著影响, 变量筛选方法和结果是可行的。此外, 残差之间的空间自相关通过莫兰指数 (Moran's I) 来检验, 最终落叶松和樟子松的结果分别为 0.21 和 0.05 ($P > 0.1$), 表明残差之间没有显著的空间自相关现象, 呈随机分布。

3 讨论

光谱反射率和植被指数与森林生长状况密切相关, 可以作为森林 AGB 估测基本的建模变量^[10,17]。相比可见光波段, 红边波段能对植被冠层结构和叶绿素含量的微小变化做出快速反应, 有效提高 AGB 估测精

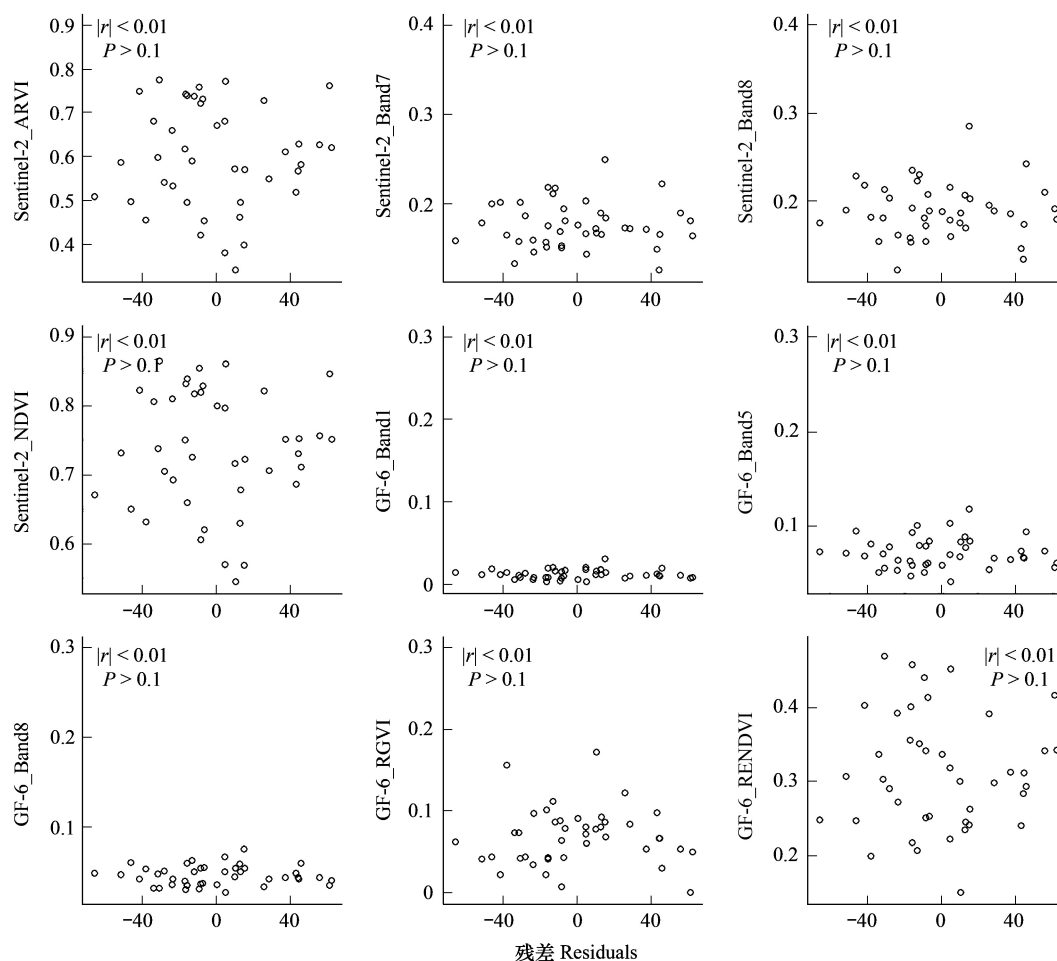


图 11 残差与建模变量散点图

Fig.11 Scatter plots of residuals and modeling variables

度^[18-20]。同时,基于红边波段构建的植被指数可以减弱 NDVI 等传统植被指数的饱和现象,从而更准确地预测和评价植被的生长状况^[26]。由旺业甸林场和塞罕坝林场 GF-6 和 Sentinel-2 提取的遥感变量中,红边波段的反射率均与 AGB 显著相关,且经过变量筛选之后保留的变量中也均包含红边波段反射率及红边波段构建的植被指数。联合 GF-6 和 Sentinel-2 建立的多元线性回归模型取得了最高的决定系数和最低的 RMSE,相比单个数据源,模型 RMSE 分别得到了显著降低。Li 等^[31]以资源三号高分辨率影像为数据源结合多元线性回归实现了旺业甸落叶松 AGB 估测,模型 RMSE 相比本研究中的 Sentinel-2 和 GF-6 数据源分别降低了 17.0% 和 13.6%,然而 GF-6 在联合 Sentinel-2 后建立的模型 RMSE 相比资源三号数据源下降了 7.2%,说明联合 GF-6 和 Sentinel-2 影像结合多元线性回归能有效地对落叶松 AGB 进行估测。

相比 Sentinel-2 影像,GF-6 拥有的黄色波段(590—630 nm)在植被生长监测中也展现了优越性。胡秀娟等^[32]通过 WorldView-2 卫星影像的黄色波段构建了新型植被指数,有效地对水土流失区的植被健康状况进行了监测。曾庆伟等^[33]证明了基于 GF-6 近红外和黄色波段建立的 YNDVI 指数比传统的 NDVI 能更有效地反映森林扰动信息。但黄色波段对森林 AGB 估测的敏感性仍需进一步验证,研究发现 GF-6 黄色波段与旺业甸林场落叶松 AGB 显著相关,且 GF-6、联合 GF-6 和 Sentinel-2 变量组在多元线性回归中筛选的结果均保留了黄色波段,说明黄色波段对 AGB 估测较敏感。

旺业甸林场和塞罕坝林场均为人工林林场,森林状况分布较均匀,光学遥感获取的冠层光谱差异有限,这导致通过构建大量决策树进行预测的适用于更复杂情况的非线性随机森林模型估测效果有限。此外,样地数

量和特征变量数量能显著影响模型的估测效果,对于小样本人工林,多元线性回归能取得比非线性模型更好的估测效果,这与 Li 等^[31]在旺业甸林场人工林得到的研究结果相似。由于森林类型、树种、林龄的差异,遥感影像提取的地表植被信息存在同物异谱、同谱异物和其他能显著影响估测精度的因素。常用的遥感变量包括波段反射率、植被指数、地形因子、纹理指数等,这些变量与不同树种 AGB 之间的相关性及其显著性并不一致。所以在进行 AGB 建模和估测时,需要同时考虑变量与 AGB 之间的线性和非线性关系,利用不同的特征变量选择方法构建相对应的模型探索更多的估测方式以提高估测精度和模型的可解释性。

SAR 和 LiDAR 等主动遥感方式能穿透森林冠层,获取更准确的森林结构信息,从而提高 AGB 估测精度。但由于地形影响和数据覆盖有限等限制,利用主动遥感数据源进行大范围 AGB 估测效率有限^[15]。此外,由于高分辨率光学影像数据量大、覆盖不完全、数据昂贵等因素,目前与国家森林资源连续清查样地大小基本一致的中分辨率光学数据仍是森林 AGB 反演的主流数据源。研究利用 GF-6 和 Sentinel-2 影像结合多元线性回归实现了旺业甸林场落叶松和塞罕坝林场樟子松 AGB 估测,能对森林 AGB 遥感监测研究提供一定的参考。然而,本研究所用的野外观测样地数量相对偏少且树种单一,采用线性回归所建立的生物量模型的适用性需要进一步验证。在今后的研究中,可以考虑通过增加实测样地数量和其他研究区对比进行方法验证。

4 结论

以 Sentinel-2 和 GF-6 作为数据源,结合野外实测 AGB 数据建立线性和非线性回归模型,对旺业甸林场落叶松和塞罕坝林场樟子松 AGB 进行反演和空间分布制图。结果表明:(1)由 Sentinel-2 和 GF-6 影像中提取的红边波段信息均与 AGB 显著相关,且经过筛选用于建模的变量主要由红边、蓝色、近红外和黄色波段等对 AGB 较敏感的波段信息构成;(2)GF-6 相比 Sentinel-2 拥有对植被参数较敏感的黄色波段,且作为数据源建立的多元线性回归模型 RMSE 相比 Sentinel-2 显著降低,表明国产 GF-6 可以代替 Sentinel-2 用于 AGB 估测。GF-6 幅宽达到 800 km,能更高效地监测大范围森林动态变化;(3)增加红边波段进行落叶 AGB 估测能显著提高模型估测精度。三组数据源分别加入红边波段信息后进行建模,落叶松估测模型 RMSE 分别下降了 9.6%、17.6%和 28.7%,樟子松则分别下降了 6.2%、7.9%和 8.2%;(4)联合 GF-6 和 Sentinel-2 建立的多元线性回归模型取得了最高的决定系数和最低的 RMSE。相比单个数据源,模型 RMSE 得到了显著降低。联合 GF-6 和 Sentinel-2 数据源结合多元线性回归能有效地对森林 AGB 进行估测,为森林资源管理和遥感动态监测提供新的数据源参考。

参考文献(References):

- [1] 黄金龙, 居为民, 郑光, 康婷婷. 基于高分辨率遥感影像的森林地上生物量估算. 生态学报, 2013, 33(20): 6497-6508.
- [2] 刘迎春, 高显连, 付超, 于贵瑞, 刘兆英. 基于森林资源清查数据估算中国森林生物量固碳潜力. 生态学报, 2019, 39(11): 4002-4010.
- [3] 蒙诗栋, 庞勇, 张钟军, 李增元, 王雪琼, 李世明. WorldView-2 纹理的森林地上生物量反演. 遥感学报, 2017, 21(5): 812-824.
- [4] 魏雪梅. 多源数据支持下的森林地上生物量估算方法. 武汉大学学报: 信息科学版, 2019, 44(9): 1385-1390.
- [5] 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.
- [6] 周律, 欧光龙, 王俊峰, 胥辉. 基于空间回归模型的思茅松林生物量遥感估测及光饱和点确定. 林业科学, 2020, 56(3): 38-47.
- [7] Gleason C J, Im J. A review of remote sensing of forest biomass and biofuel: options for small-area applications. GIScience & Remote Sensing, 2011, 48(2): 141-170.
- [8] Cui Y L, Sun H, Wang G X, Li C J, Xu X Y. A Probability-based spectral unmixing analysis for mapping percentage vegetation cover of arid and semi-arid areas. Remote Sensing, 2019, 11(24): 3038.
- [9] Hu Y, Xu X L, Wu F Y, Sun Z Q, Xia H M, Meng Q M, Huang W L, Zhou H, Gao J P, Li W T, Peng D L, Xiao X M. Estimating forest stock volume in Hunan province, China, by integrating in situ plot data, sentinel-2 images, and linear and machine learning regression models. Remote Sensing, 2020, 12: 186.
- [10] Lu D S, Chen Q, Wang G X, Liu L J, Li G Y, Moran E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. International Journal of Digital Earth, 2016, 9(1): 63-105.
- [11] 李兰, 陈尔学, 李增元, 任冲, 赵磊, 谷鑫志. 森林地上生物量的多基线 InSAR 层析估测方法. 林业科学, 2017, 53(11): 85-93.

- [12] 刘晓亮, 隋立春, 白永飞, 赵旦, 赵玉金, 刘艳书, 翟秋萍. 地基激光雷达灌丛化草原小叶锦鸡儿生物量估算. 遥感学报, 2020, 24(7): 894-903.
- [13] 邱赛, 邢艳秋, 徐卫华, 丁建华, 田静. 星载大光斑 LiDAR 与 HJ-1A 高光谱数据联合估测区域森林地上生物量. 生态学报, 2016, 36(22): 7401-7411.
- [14] 王新云, 郭艺歌, 何杰. 基于 HJ1B 和 ALOS/PALSAR 数据的森林地上生物量遥感估算. 生态学报, 2016, 36(13): 4109-4121.
- [15] 陈松, 孙华, 吴童, 蒋馥根. 基于 Sentinel-2 与机载激光雷达数据的误差变量联立方程组森林蓄积量反演研究. 中南林业科技大学学报, 2020, 40(12): 44-53.
- [16] García-Gutiérrez J, Martínez-Álvarez F, Troncoso A, Riquelme J C. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing*, 2015, 167: 24-31.
- [17] Vafaei S, Soosani J, Adeli K, Fadaei H, Naghavi H, Pham T D, Bui D T. Improving accuracy estimation of forest aboveground biomass based on incorporation of ALOS-2 PALSAR-2 and Sentinel-2A imagery and machine learning: a case study of the Hyrcanian forest area (Iran). *Remote Sensing*, 2018, 10(2): 172.
- [18] Chen Y Y, Li L W, Lu D S, Li D Q. Exploring bamboo forest aboveground biomass estimation using Sentinel-2 data. *Remote Sensing*, 2019, 11(1): 7.
- [19] 梁继, 郑镇伟, 夏诗婷, 张晓彤, 唐媛媛. 高分六号红边特征的农作物识别与评估. 遥感学报, 2020, 24(10): 1168-1179.
- [20] 梁志国, 隋傲, 于颖, 赵戈榕, 谢秋, 刘代超. 高分六号遥感数据在有林地识别中的应用. 东北林业大学学报, 2020, 48(5): 35-39.
- [21] Yu J X, Liu Y L, Ren Y H, Ma H J, Wang D C, Jing Y F, Yu L J. Application study on double-constrained change detection for land use/land cover based on GF-6 WFV imageries. *Remote Sensing*, 2020, 12(18): 2943.
- [22] 方精云, 陈安平. 中国森林植被碳库的动态变化及其意义. 植物学报, 2001, 43(9): 967-973.
- [23] 李海奎, 雷渊才. 中国森林植被生物量和碳储量评估. 北京: 中国林业出版社, 2010.
- [24] Yu X H, Ge H L, Lu D S, Zhang M Z, Lai Z X, Yao R T. Comparative study on variable selection approaches in establishment of remote sensing model for forest biomass estimation. *Remote Sensing*, 2019, 11(12): 1437.
- [25] Sun H, Wang Q, Wang G X, Lin H, Luo P, Li J P, Zeng S Q, Xu X Y, Ren L X. Optimizing kNN for mapping vegetation cover of arid and semi-arid areas using Landsat images. *Remote Sensing*, 2018, 10(8): 1248.
- [26] Lin S R, Li J, Liu Q H, Li L H, Zhao J, Yu W T. Evaluating the effectiveness of using vegetation indices based on red-edge reflectance from Sentinel-2 to estimate gross primary productivity. *Remote Sensing*, 2019, 11(11): 1303.
- [27] Gitelson A A, Gritz Y, Merzlyak M N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 2003, 160(3): 271-282.
- [28] 孙华, 鞠洪波, 张怀清, 林辉, 凌成星. 三种回归分析方法在 Hyperion 影像 LAI 反演中的比较. 生态学报, 2012, 32(24): 7781-7790.
- [29] Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5-32.
- [30] Willmott C J, Ackleson S G, Davis R E, Feddesma J J, Klink K M, Legates D R, O'Donnell J, Rowe C M. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, 1986, 90(C5): 8995-9005.
- [31] Li G Y, Xie Z L, Jiang X D, Lu D S, Chen E X. Integration of ZiYuan-3 multispectral and stereo data for modeling aboveground biomass of larch plantations in North China. *Remote Sensing*, 2019, 11(19): 2328.
- [32] 胡秀娟, 徐涵秋, 郭燕滨, 张博博. 水土流失区生态修复后植被健康的遥感判别. 应用生态学报, 2017, 28(1): 250-256.
- [33] 曾庆伟, 武红敢. GF-6 WFV 黄边波段在森林扰动监测中的应用潜力研究. 北京测绘, 2020, 34(2): 208-213.