北京林业大学学报
JOURNAL OF BEIJING FORESTRY UNIVERSITY

基于特征选择和机器学习的森林蓄积量估算

赵娅冰 彭道黎 郭发苗 王荫 黄静娴

Estimating forest stock volume based on feature selection and machine learning

Zhao Yabing, Peng Daoli, Guo Famiao, Wang Yin, Huang Jingxian

引用本文:
赵娅冰, 彭道黎, 郭发苗, 王荫, 黄静娴. 基于特征选择和机器学习的森林蓄积量估算[J]. 北京林业大学学报, 2025, 47(4):155–167. doi: 10.12171/j.1000–1522.20240328
Zhao Yabing, Peng Daoli, Guo Famiao, Wang Yin, Huang Jingxian. Estimating forest stock volume based on feature selection and machine learning[J]. Journal of Beijing Forestry University, 2025, 47(4):155–167. doi: 10.12171/j.1000–1522.20240328

在线阅读 View online: https://doi.org/10.12171/j.1000–1522.20240328

您可能感兴趣的其他文章
Articles you may be interested in

机器学习算法在森林生长收获预估中的应用
Applications of machine learning algorithms in forest growth and yield prediction
北京林业大学学报, 2019, 41(12): 23–36. https://doi.org/10.12171/j.1000–1522.20190356

森林地上生物量遥感估算方法
Estimation of forest aboveground biomass by remote sensing
北京林业大学学报, 2021, 43(8): 137–148. https://doi.org/10.12171/j.1000–1522.20200166

参数优选支持的光学与SAR数据森林地上生物量反演研究
Inversion exploration on forest aboveground biomass of optical and SAR data supported by parameter optimization method
北京林业大学学报, 2020, 42(10): 11–19. https://doi.org/10.12171/j.1000–1522.20190389

基于机器学习的落叶松毛虫发生面积预测模型
Studies on prediction models of *Dendrolimus superans* occurrence area based on machine learning
北京林业大学学报, 2017, 39(1): 85–93. https://doi.org/10.13332/j.1000–1522.20160205

机器学习技术在风景园林中的应用进展与展望
Application progress and prospect of machine learning technology in landscape architecture
北京林业大学学报, 2021, 43(11): 137–156. https://doi.org/10.12171/j.1000–1522.20200313

东北林区10种主要森林类型的蓄积量、生物量和碳储量模型研建
Developing stand volume, biomass and carbon stock models for ten major forest types in forest region of northeastern China
北京林业大学学报, 2021, 43(3): 1–8. https://doi.org/10.12171/j.1000–1522.20200058

DOI:10.12171/j.1000−1522.20240328

基于特征选择和机器学习的森林蓄积量估算

赵娅冰　彭道黎　郭发苗　王　荫　黄静娴
（北京林业大学森林资源和环境管理国家林草局重点实验室，北京 100083）

摘要:【目的】基于多源遥感数据，评估不同特征选择方法和机器学习算法组合构建的森林蓄积量估算模型的准确性，挖掘其协同互补潜力，以期有效提高森林蓄积量的估算精度。**【方法】**以河北省第九次国家森林资源连续清查数据为基础，结合 GF-1、Sentinel-2、Sentinel-1 和 ASTER GDEM 4 种遥感数据，采用随机森林变量选择（VSURF）、递归特征消除（RFE）和 Boruta 3 种特征选择方法，以及支持向量回归（SVR）、K-最近邻（KNN）、随机森林（RF）、分类提升（CatBoost）和极端梯度提升（XGBoost）5 种机器学习算法，构建蓄积量模型，并筛选出最优模型。此外，通过方差分析量化数据集、特征选择和机器学习算法这 3 个因素对森林蓄积量估算的影响。**【结果】**（1）方差分析结果表明，数据集、特征选择和机器学习算法均对蓄积量估算性能有显著影响。（2）多源遥感数据的结合能有效提高森林蓄积量的估算性能。与其他数据集相比，联合 GF-1、Sentinel-2、Sentinel-1 和 ASTER GDEM 数据构建的模型表现出更高的估算精度。从整体来看，Boruta 特征选择方法优于 VSURF 和 RFE。CatBoost 在建模中的表现优于其他算法（SVR、KNN、RF 和 XGBoost）。（3）基于 GF-1、Sentinel-2、Sentinel-1 和 ASTER GDEM 的组合，使用 Boruta 特征选择方法和 CatBoost 机器学习算法构建的估算模型实现了最高的准确性（R^2 = 0.638 5，RMSE = 13.305 3 m^3/hm^2）。**【结论】**基于多源遥感数据估算保定市森林蓄积量时，结合特征选择和机器学习算法可显著优化模型的估算效果，得到更精准的蓄积量估算结果。研究结果不仅改进了当前应用多源遥感数据估算森林蓄积量的方法，还为大范围森林蓄积量监测提供了新的思路和参考依据。

关键词：森林蓄积量；多源遥感数据；特征选择；机器学习算法；集成学习
中图分类号：S771.8　　**文献标志码：**A　　**文章编号：**1000−1522(2025)04−0155−13
引文格式：赵娅冰，彭道黎，郭发苗，等. 基于特征选择和机器学习的森林蓄积量估算［J］. 北京林业大学学报，2025，47(4)：155−167. Zhao Yabing, Peng Daoli, Guo Famiao, et al. Estimating forest stock volume based on feature selection and machine learning［J］. Journal of Beijing Forestry University, 2025, 47(4): 155−167.

Estimating forest stock volume based on feature selection and machine learning

Zhao Yabing　Peng Daoli　Guo Famiao　Wang Yin　Huang Jingxian
(Key Laboratory of Forest Resources & Environmental Management of National Forestry and Grassland Administration, Beijing Forestry University, Beijing 100083, China)

Abstract: ［**Objective**］Based on multi-source remote sensing data, the accuracy of forest stock volume estimation models constructed by combining different feature selection methods and machine learning algorithms was evaluated, and their synergistic and complementary potentials were explored to effectively improve the estimation accuracy of forest stock volume. ［**Method**］Based on the data of the 9th National Forest Resources Continuous Inventory in Hebei Province of northern China, this study combined four types of remote sensing data, i.e., GF-1, Sentinel-2, Sentinel-1 and ASTER GDEM, and employed three types of feature selection methods, i.e., variable selection using random forests (VSURF), recursive feature elimination (RFE) and Boruta, and five types of machine learning algorithms, i.e., support vector egression

收稿日期: 2024−10−08　**修回日期:** 2024−12−03
基金项目: 国家重点研发计划项目 (2023YFD2200403)。
第一作者: 赵娅冰。主要研究方向：森林资源监测与评价。Email: z15093476806@bjfu.edu.cn　地址：100083 北京市海淀区清华东路 35 号北京林业大学林学院。
责任作者: 彭道黎，教授，博士生导师。主要研究方向：森林资源监测与评价。Email: dlpeng@bjfu.edu.cn　地址：同上。
本刊网址: http://j.bjfu.edu.cn; http://journal.bjfu.edu.cn

(SVR), K-nearest neighbor (KNN), random forest (RF), categorical boosting (CatBoost) and extreme gradient boosting (XGBoost) to construct forest stock volume model and screen the optimal model. In addition, the effects of three factors, i.e., dataset, feature selection and machine learning algorithms, on the estimation of forest stock volume were quantified by analysis of variance (ANOVA). [**Result**] (1) The results of ANOVA showed that the dataset, feature selection and machine learning algorithms all had a significant impact on performance of forest stock volume estimation. (2) The combination of multi-source remote sensing data can effectively improve the performance of forest stock volume estimation. Compared with other datasets, the model constructed by combining the GF-1, Sentinel-2, Sentinel-1 and ASTER GDEM data showed higher estimation accuracy. On the whole, the Boruta feature selection method was superior to VSURF and RFE. CatBoost outperformed other algorithms (SVR, KNN, RF and XGBoost) in modeling. (3) Based on the combination of GF-1, Sentinel-2, Sentinel-1, and ASTER GDEM, the estimation model built using Boruta for feature selection and CatBoost machine learning algorithm achieved the highest accuracy (R^2 = 0.638 5, RMSE = 13.305 3 m^3/ha). [**Conclusion**] In the estimation of forest stock volume in Baoding City of Hebei Province based on multi-source remote sensing data, the combination of feature selection and machine learning algorithm can effectively improve the model estimation effect, obtain better forest stock volume estimation results. The results of this study not only improve the current method of estimating forest stock volume based on multi-source remote sensing data, but also provide a new idea and reference basis for large-scale forest stock volume monitoring.

Key words: forest stock volume; multi-source remote sensing data; feature selection; machine learning algorithm; ensemble learning

森林作为陆地生态系统中最大的碳库，在应对全球气候变化方面发挥着关键作用。精确监测森林蓄积量有助于改善森林生态环境，对实现碳中和目标和调控全球碳循环机制至关重要。传统的森林蓄积量监测主要通过地面调查，虽然这种方法具有较高的准确性，但由于调查周期长、成本高，难以及时反映大面积森林资源的空间分布状况和动态变化[1]。相比之下，遥感技术具有多时相、低成本、快速宏观监测等显著优势，能够在一定程度上克服传统地面调查中所存在的局限性。

目前，森林蓄积量的遥感估算研究主要集中在多源遥感数据结合、特征选择和建模算法 3 个方面。光学数据[2]、激光探测和测距(light detection and ranging，LiDAR)[3]和合成孔径雷达(synthetic aperture radar，SAR)[4]是提取林分参数信息的主要遥感数据类型。LiDAR 能够有效地提供森林的垂直结构信息，在森林参数(如树高、生物量、碳汇)估算方面具有独特的优势，但其成本较高且缺乏多时相动态监测的历史数据，难以在大范围内推广与应用[5]。光学遥感技术数据源丰富，已广泛应用于大范围森林蓄积量估算，但光学数据受云雾等气象条件影响较大，特别是在郁闭度较高的林分中。与光学数据不同，SAR 是一种基于微波雷达辐射信号的主动遥感技术，对森林冠层具有更高的灵敏度，不易受云雾、太阳辐射等气候因素影响，因此在森林参数估算中能够实现更高的精度。然而，SAR 数据的可使用波段比较单一，且在地形崎岖地区可能会出现信号饱和的问题[6]。尽管如此，SAR 和光学数据丰富的光谱信息能够反映森林的结构和功能，在森林参数估算中显示出巨大的潜力[7]。2013 年发射的 GF-1 卫星是光学遥感技术的重要突破，结合了高分辨率、多光谱和高时间分辨率，能够获取精细的观测信息，近年来在多个领域得到了广泛应用[8−9]。然而，GF-1 数据在蓄积量估算中的潜力尚未得到充分开发，特别是在与其他遥感数据整合用于大面积估算时。此外，全球范围内免费在线访问的 Sentinel 系列影像为地球环境监测提供了理想的数据源，也为区域森林蓄积量估算研究提供了重要支持。整合光学和 SAR 遥感数据进行蓄积量估算，能够充分利用不同遥感技术的优势，从而有效提高蓄积量估算精度。

在基于遥感技术的蓄积量估算中，各种类型的特征因子已被广泛应用于森林参数的估算研究[10−11]。大量可供选择的遥感特征因子丰富了森林蓄积量遥感建模的数据源，但增加了计算负荷。此外，特征变量之间的高度相似性可能导致多重共线性和变量冗余问题，使后续的建模分析和处理变得更加复杂。因此，合理且有效的特征选择方法可以降低变量空间维数，精准识别关键特征变量，进而减少蓄积量模型估算的误差。随机森林变量选择(variable selection using random forest，VSURF)、递归特征消

除（recursive feature elimination，RFE）和 Boruta 方法已被证明可以从高维特征中筛选出最优的变量组合，增加模型可解释性，广泛应用于林业研究等领域[12−14]。但在不同遥感数据集情景下，这 3 种方法在森林蓄积量估算中的表现尚未得到充分研究，因此有必要进一步开展针对森林蓄积量估算的特征选择的相关研究，以收集更多有力证据，进而得出更加科学合理的结论。

除了变量的选择，建模算法也是影响森林蓄积量估算准确性的重要因素。过去几十年来，参数模型和非参数模型受到广泛关注。参数模型结构清晰、直观，在预测森林参数方面具有明显优势。然而，在数据分布不确定的情况下，参数模型无法有效解释变量之间的复杂关系，这也是其预测精度较低的原因。基于数据挖掘和机器学习的非参数算法，如支持向量回归（support vector regression，SVR）和 K-最近邻（K-nearest neighbor，KNN），可以处理复杂的非线性关系，其在森林蓄积量估算中的优势已经得到了认可[15−16]。然而，在实际研究中，由于森林空间分布具有高度的异质性和复杂性，实测蓄积量样本数量通常比较有限，不同模型在不同研究区的估算性能往往参差不齐，依靠单一模型进行估算容易出现系统性偏差。相比之下，集成学习算法通过综合多个模型的结果进行预测，如 Bagging 和 Boosting 算法优于其他机器学习算法，在许多研究领域都表现出了巨大的潜力。随机森林（random forest，RF）、分类提升（categorical boosting，CatBoost）和极端梯度提升（extreme gradient boosting，XGBoost）作为先进的遥感算法，在森林参数估算中具有优异的性能[17−19]。然而，这些算法在森林蓄积量估算中的具体表现仍有待进一步检验。

本研究基于样地实测蓄积量与 GF-1、Sentinel-2、Sentinel-1 数据，拟使用 3 种特征选择方法（VSURF、RFE、Boruta）和 5 种机器学习算法（SVR、KNN、RF、CatBoost 和 XGBoost）分别建立森林蓄积量估算模型，评估在 4 种数据集场景下不同的特征选择方法和机器学习算法在森林蓄积量估算中的表现，并对影响森林蓄积量估算精度的关键因素进行了分析。研究基于多源遥感数据，结合特征选择和机器学习算法开展森林蓄积量估算试验，以期有效改善森林蓄积量估算效果。这一策略可为区域尺度高精度森林蓄积量监测提供技术参考，同时也将为森林资源优化管理、碳储量评估等应用提供科学依据，进而推动“碳中和”战略目标的实现。

1　研究区概况与研究方法

1.1　研究区概况

研究地点为中国河北省保定市（38°14′～39°57′N，113°45′～116°19′E），占地约 221.12 万 hm^2（图 1）。其地处太行山北部东麓，冀中平原西部。地势由西

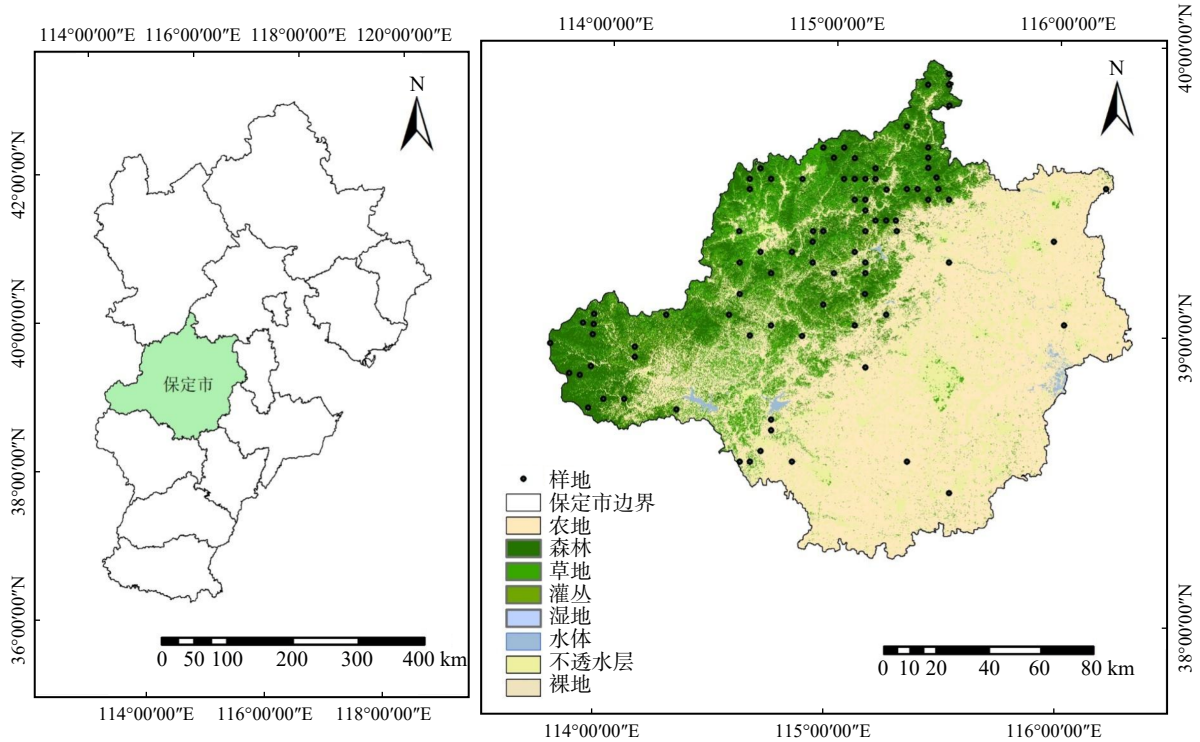


图 1　研究区位置和样地分布

Fig. 1　Location of study area and distribution of sample plots

北向东南倾斜，西部由山脉和丘陵组成，东部地区地貌平坦。保定市属暖温带大陆性季风气候带，平均降水量 489.9 mm，每年 6 — 8 月为强降水期。年平均气温 13.4 ℃，无霜期为 168 ~ 231 d。全市森林面积近 59 万 hm^2，森林覆盖率 28.25%，森林蓄积量达到 1 370 万 m^3。主要针叶树种有油松（*Pinus tabuliformis*）和侧柏（*Platycladus orientalis*），主要的阔叶树种为山杨（*Populus davidana*）、蒙古栎（*Quercus mongolica*）、白桦（*Betula platydiana*）和金合欢（*Robinia pseudoacacia*）。

1.2 数据获取以及预处理

1.2.1 地面数据

使用 2016 年收集的河北省第九次国家森林资源连续清查数据。样地间隔为 4 km × 4 km，每个方形样地的面积为 25.82 m × 25.82 m（约 0.066 7 hm^2）。对样地内胸径≥5 cm 的活立木进行每木检尺，同时还记录了冠层高度、海拔、坡度、坡向和坡位。利用测量的单木胸径和树高数据，基于立木材积表计算单木材积，汇总样地内所有单木材积即可得到样地蓄积量，并将其转化为公顷蓄积量（m^3/hm^2）。剔除非林地样地（农田、水域、城市土地和裸地）和信息不完全的样地，最后选取保定市的 87 个样地，样地的蓄积量范围为 0.82 ~ 83.85 m^3/hm^2，平均蓄积量为 31.38 m^3/hm^2，标准偏差为 21.84 m^3/hm^2。

1.2.2 GF-1 数据

GF-1 卫星搭载 4 个宽幅（WFV）多光谱相机，能实现 4 d 的重访期。本研究所使用的覆盖研究区域的 4 景 GF-1 WFV 影像的空间分辨率为 16 m，包含 4 个光谱波段。GF-1 影像在 ENVI 软件中进行预处理。预处理步骤主要包括辐射校正、大气校正、正射校正、镶嵌和裁剪操作。预处理完成后，将影像重采样至 25.82 m 的分辨率。

1.2.3 Sentinel-2 数据

使用的 Sentinel-2 Level-1C 数据是大气表观反射率产品，利用哨兵数据应用平台（SNAP）软件中的大气校正处理器来获取 Level-2A 产品，即大气底层反射率数据。选取了空间分辨率为 10 m 和 20 m 的波段，为了匹配样地大小，将预处理后的 Sentinel-2 影像重采样至 25.82 m 的分辨率。

1.2.4 Sentinel-1 数据

使用的 Sentinel-1A 数据来源于宽幅干涉模式（IW）下的 Level-1 产品，具体为地距影像（GRD）VH + VV 双极化数据，空间分辨率为 10 m。采用 SNAP 软件对 Sentinel-1 数据进行预处理，预处理步骤主要包括轨道校正、热噪声去除、辐射定标、相干斑滤波，地形校正和分贝化。将预处理后的影像重采样至 25.82 m 的分辨率，使其与调查样地匹配。

1.2.5 地形数据

数字高程模型（digital elevation model，DEM）能够反映地表高程信息，可为森林蓄积量估算提供丰富的地形信息。使用覆盖研究区的 30 m 空间分辨率的 ASTER GDEM 数据来获取地形因子。为了确保与遥感数据保持一致性，对 ASTER GDEM 数据进行了重采样，分辨率为 25.82 m。

1.3 特征提取

从 GF-1 影像中提取常用于森林蓄积量估算的光谱特征，包括波段反射率、植被指数（VI）和纹理特征。从 Sentinel-2 数据 13 个波段中提取出分辨率为 10 m 和 20 m 的 10 个多光谱波段。从 Sentinel-1 影像中提取 VH 和 VV 后向散射系数。通过灰度共生矩阵计算纹理特征，包括具有 3 × 3、5 × 5、7 × 7 和 9 × 9 像素窗口的各个波段的均值、方差、同质性、对比度、异质性、信息熵、二阶矩和相关性特征[20]。然后，从 ASTER GDEM 影像中提取海拔、坡度和坡向变量。最后，共提取了 537 个特征作为森林蓄积量估算的输入变量（表 1）。

表 1 基于遥感影像提取的变量

Tab. 1 Variables extracted from remote sensing images

影像类别	变量类别	变量名称
GF-1	波段反射率	B1 (Blue), B2 (Green), B3 (Red), B4 (NIR)
	植被指数	改良土壤调整植被指数 (MSAVI) 归一化差值植被指数 (NDVI) 优化的土壤调整植被指数 (OSAVI) 比值植被指数 (RVI) 土壤调整植被指数 (SAVI) 增强型植被指数 (EVI)
	纹理特征	B_{i_j}Con, B_{i_j}Dis, B_{i_j}Mea, B_{i_j}Hom, B_{i_j}SeM, B_{i_j}Ent, B_{i_j}Var, B_{i_j}Cor
Sentinel-2	波段反射率	b2 (Blue), b3 (Green), b4 (Red), b5 (Red Edge1), b6 (Red Edge2), b7 (Red Edge3), b8 (NIR), b8A (Narrow NIR), b11 (SWIR1), b12 (SWIR2)
	纹理特征	b_{i_j}Con, b_{i_j}Dis, b_{i_j}Mea, b_{i_j}Hom, b_{i_j}SeM, b_{i_j}Ent, b_{i_j}Var, b_{i_j}Cor
Sentinel-1	后向散射系数	VH, VV
	纹理特征	VH/VV_$_j$Con, VH/VV_$_j$Dis, VH/VV_$_j$Mea, VH/VV_$_j$Hom, VH/VV_$_j$SeM, VH/VV_$_j$Ent, VH/VV_$_j$Var, VH/VV_$_j$Cor
ASTER GDEM	地形因子	海拔，坡度，坡面

注：*i*表示波段，*j*表示窗口大小（3 × 3，5 × 5，7 × 7，9 × 9像素），Con表示对比度，Dis表示异质性，Mea表示均值，Hom表示同质性，SeM表示二阶矩，Ent表示信息熵，Var表示方差，Cor表示相关性。

为了分析不同数据集对森林蓄积量估算的影响，设计了 4 种数据集情景：Gf1s2（GF-1 + Sentinel-2）；Gf1s1（GF-1 + Sentinel-1）；S2s1（Sentinel-2 + Sentinel-1）；Gf1s2s1dem（GF-1 + Sentinel-2 + Sentinel-1 + ASTER GDEM）。

1.4 特征选择方法

大量潜在的预测因子可用作森林蓄积量估算的候选变量，但并非所有变量都对模型预测具有显著贡献。高维特征的输入可能导致过拟合或计算效率降低，因此，特征选择是有效提高模型精度的关键。采用 VSURF、RFE 和 Boruta 3 种方法从 4 种数据集中选择特征变量用于后续建模分析。

1.4.1 随机森林变量选择

VSURF 是一种基于包装器的特征选择算法，以随机森林作为基础模型[21]。VSURF 通过评估特征的重要性来实现特征选择。具体而言，首先基于变量重要性对特征变量进行排序，消除数据集中与目标不相关的特征，以减少特征数量并提高模型准确性。最终，算法仅保留最重要的特征并输出其排序结果。这一过程在 R 语言 4.3.2 版本中使用“VSURF”包执行。

1.4.2 递归特征消除

RFE 是一种向后选择算法[22]，其通过反复构建模型，最后筛选出最佳预测特征集。具体而言，首先基于变量重要性对特征进行排序，消除得分最低的特征，剩余的特征用于构建下一个模型，重新进行特征重要性评估和特征消除，重复该过程直至所有特征都被处理过。为了进一步优化特征选择，本研究在 RFE 中引入 10 倍交叉验证来确定最优的特征数量。这个过程在 R 语言 4.3.2 版本中使用“caret”包执行。

1.4.3 Boruta

Boruta 算法类似于随机森林分类算法的基础思想[23]。该算法将原始候选特征数据集进行随机排列，随机扰动以后的特征作为原始特征变量对应的“阴影”特征，将原始特征和“阴影”特征共同输入模型进行重要性评分。通过迭代逐步剔除重要性得分低于“阴影”特征的特征变量，直至所有特征变量得到确认或拒绝。这个过程在 R 语言 4.3.2 版本中使用“Boruta”包执行。

1.5 蓄积量估算模型构建

采用支持向量回归（SVR）、K-最近邻（KNN）、随机森林（RF）、分类提升（CatBoost）和极端梯度提升（XGBoost）5 种机器学习算法对森林蓄积量进行建模。

1.5.1 支持向量回归

SVR 是通过非线性变换将样本空间转换到一个高维特征空间，核函数的引入使其在特征空间中可以应用线性学习机的方法来解决样本空间中的高度非线性回归等问题，减少了计算量。该方法基于结构风险最小化原则，在样本信息有限的情况下避免了过学习问题，具有良好的模型性能和泛化能力[24]。这个过程在 R 语言 4.3.2 版本中使用“e1071”包执行。

1.5.2 K-最近邻

KNN 是一种用于回归的典型的非参数方法，可以有效避免自变量复杂共线性问题的产生[25]。KNN 选定特征空间中的一个样本数据，通过计算不同特征值之间的距离，找出与给定样本最相邻的 k 个样本数据，并记录其属性值，然后将这些邻居样本对应的值加权平均赋给该样本数据，从而得到给定样本最终的估算值。这个过程在 R 语言 4.3.2 版本中使用“caret”包执行。

1.5.3 随机森林

RF 是 Breiman[26]在 2001 年提出的一种基于“Bagging”的集成学习方法，具有高效处理高维数据和复杂非线性关系的能力。其基本原理为采用“Bagging”思想将多个决策树集成到一个单一且强大的模型中。其随机性能够有效地增强模型的稳定性，对高维特征组合有很好的适应性，并在一定程度上实现了良好的泛化能力。这个过程在 R 语言 4.3.2 版本中使用“randomForest”包执行。

1.5.4 分类提升

CatBoost 是一种改进的类似于梯度提升决策树（GBDT）的工具包[27]。CatBoost 通过引入一种创新的梯度提升算法—order boosting，有效解决了梯度偏差和预测偏移的问题。它能够自动将分类特征转换为数字特征进行处理。与现有的 GBDT 算法相比，CatBoost 可以自动处理离散特征数据，使其适用于包含噪声样本的回归数据和具有多输入特征的回归问题。该模型具有较强的稳健性和泛化能力，算法准确性较好。这个过程在 R 语言 4.3.2 版本中使用“catboost”包执行。

1.5.5 极端梯度提升

XGBoost 是由 Chen 和 Guestrin[28]在 2016 年提出的在 GBDT 框架基础上改进的一种集成学习算法。XGBoost 基于“Boosting”思想依次生成若干决策树，将一组弱学习器的所有预测结果组合在一起，通过加性训练策略培养出一个强学习器。XGBoost 使用二阶 Taylor 展开目标函数，在训练时使用二阶导数来加快模型的收敛速度。同时，在目标函数中加入正则项来控制树的复杂度，从而减少计算量，避免了过度拟合[29]。这个过程在 R 语言 4.3.2 版本中

使用“xgboost”包执行。

1.6 模型估算结果评估

由于数据集情景的差异，不同特征选择方法和机器学习算法的估算性能可能不同。为了验证上述假设，基于 4 种数据集分别使用 3 种特征选择方法和 5 种机器学习算法共构建了 60 个模型。

使用留一交叉验证来评价和优化蓄积量估算模型，采用估算值和实测值之间的决定系数(R^2)、均方根误差(RMSE)和平均绝对误差(MAE)来评估蓄积量估算模型的预测性能[30]。R^2 越高，RMSE 和 MAE 越低，意味着预测精度越高，估计结果越好。

1.7 方差分析

为了评估不同影响因素(数据集、特征选择、机器学习算法)及其交互作用对森林蓄积量估算的影响，采用方差分析(ANOVA)来量化各个因素对估算误差的影响，并确定森林蓄积量估算中的关键因素。平方和(Sumsq)表示数据变异程度的量化指标，用于判断不同因素对结果的影响是否显著。为进一步明确各因素的相对重要性，计算 η^2 效应量，用于表示特定因素所解释的方差占总方差的比例。这个过程是在 R 语言 4.3.2 版本中执行的。

1.8 森林蓄积量制图

应用 FROM-GLC 10 产品来定义研究区的森林区域，该产品将地表覆盖分为 10 类(农地、森林、草地、灌丛、湿地、水体、苔原、不透水层、裸地、雪/冰)。研究基于 FROM-GLC 10 产品的 10 m 分辨率全球土地覆盖数据来获取保定市森林分布图，并结合多源遥感数据，利用上述得出的最佳优化模型进行森林蓄积量空间预测与制图。

2 结果与分析

2.1 特征变量选择

基于 GF-1、Sentinel-2、Sentinel-1、ASTER GDEM 多源遥感数据的 4 种组合分别为 Gf1s2、Gf1s1、S2s1、Gf1s2s1dem，4 种数据集情景下蓄积量建模的变量选择结果如表 2 所示。在不同的数据集中，基于不同方法选择的特征变量是不同的。例如，在 Gf1s2s1dem 数据集中，当使用 RFE 进行特征选择时，倾向于选择 GF-1 和 Sentinel-1 的纹理信息，而 VSURF 和 RFE 选择的特征主要来自 GF-1。此外，值得注意的是，在具有地形因素的数据集情景中，只有 RFE 选择了坡度指标，但地形的重要性不能忽略。在含有 GF-1 和 Sentinel-2 的数据集情景中，B4_3Con、B4_3Cor 和 B4_5Hom 等纹理变量被频繁选择，与 RFE 和 Boruta 相比，VSURF 倾向于选择来自 GF-1 的变量。在包含 Sentinel-1 变量的数据集情景中，VH_3Var、VH_7Mea 和 VV_9Cor 等纹理特征也被用于建模。这表明纹理因子更适合进行蓄

表 2 4 种数据集情景下森林蓄积量建模的变量选择

Tab. 2 Variable selection for forest stock volume modeling under four dataset scenarios

数据集	特征选择方法	选择的变量数量	选择的变量名称
GF-1 + Sentinel-2 (Gf1s2)	VSURF	7	B2_3SeM, B2_9Dis, B4_3Con, B4_3Co, B4_3SeM, B4_5Ent, B4_7Ent
	RFE	16	B1_9Hom, B2_3Var, B2_5Cor, B2_7Hom, B2_9Cor, B4_3Con, B4_3Cor, B4_3Dis, B4_3Hom, B4_3Var, B4_5Hom, B4_7SeM, B4_9SeM, B4_9Ent, b7_9Hom, b8A_3Var
	Boruta	8	B2_9Ent, B4_3Dis, B4_3Var, B4_5Hom, B4_7Ent, B4_7SeM, B4_9Ent, b7_3cor
Sentinel-2 + Sentinel-1 (S2s1)	VSURF	6	b4_7Mea, b6_9Hom, b7_3Con, VH_3SeM, VH_7Hom, VV_9Hom
	RFE	8	b6_5Cor, b6_7Mea, b6_9Mea, b7_3Cor, b8_3Ent, VH_3Var, VH_5Cor, VV_9Cor
	Boruta	6	b4_7Cor, b7_3Cor, b7_5Cor, b11_5Ent, VH_3Var, VH_7mea
GF-1 + Sentinel-1 (Gf1s1)	VSURF	7	B2_9Dis, B4_3Con, B4_3Cor, B4_3SeM, B4_9Hom, VH_7Hom, VV_9Hom
	RFE	18	B2_7Hom, B4_3Con, B4_3Cor, B4_3Dis, B4_3Var, B4_5Hom, B4_7Dis, B4_9Hom, VV, VH_3Var, VV_3Mea, VH_5Cor, VV_5Mea, VH_7Mea, VH_9Cor, VH_9Hom, VV_9Cor, VV_9Var
	Boruta	13	B2_3Var, B2_9Dis, B2_9Ent, B4_3Con, B4_3Dis, B4_3Var, B4_3Hom, B4_5Hom, B4_7Ent, B4_7Hom, B4_9SeM, VH_7Mea, VV_9Cor
GF-1 + Sentinel-2 + Sentinel-1 + ASTER GDEM (Gf1s2s1dem)	VSURF	9	B2_9Dis, B3_9Var, B4_3Con, B4_3Cor, B4_3Ent, B4_3SeM, B4_5Ent, B4_9Ent, VV_9Hom
	RFE	18	B2_7Hom, B3_7Hom, B4_3Con, B4_3Cor, B4_3Dis, B4_3Var, B4_5Hom, B4_9Hom, b2_9Con, b2_9Dis, b5_9SeM, VH_3Var, VV_3Mea, VV_5Mea, VH_7Mea, VH_9Cor, VV_9Cor, Slope
	Boruta	8	B4_3Con, B4_3Dis, B4_3Var, B4_3Hom, B4_5Hom, B4_9SeM, b4_7Cor, VH_7Cor

积量估算，模型中最常选择的纹理因子是同质性和相关性指标，这些纹理特征对蓄积量建模的贡献更大。此外，窗口大小也是影响蓄积量估算模型精度的一个重要因素，基于 3 × 3 和 9 × 9 窗口大小的纹理经常出现在模型中，表明这两种窗口大小更适合用于森林蓄积量估算。

2.2　不同数据集情景对蓄积量估算模型的影响

应用 3 个统计指标（R^2、RMSE、MAE）来评估从不同数据集情景构建的蓄积量估算模型。在 4 种数据集情景下，其最优模型由不同的特征选择方法和建模方法构建（图 2）。在数据集 Gf1s2、Gf1s1 和 Gf1s2s1dem 中，基于 Boruta 和 CatBoost 构建的模型具有相对最高的蓄积量估算性能；对于 S2s1 数据集，基于 RFE 和 CatBoost 的蓄积量估算性能优于其他模型。在已建立的所有模型中，基于 Gf1s2s1dem 数据集使用 Boruta 和 CatBoost 算法的蓄积量估算结果散点图具有最好的拟合度和最低的预测误差（R^2 = 0.638 5，RMSE = 13.305 3 m^3/hm^2）。此外，模型中都存在高值低估和低值高估的问题，即当蓄积量值较大时，观察到模型的预测值被显著低估；当蓄积量值较小时，表现出明显的高估现象（图 2）。对比不同数据集的模型评价指标（图 3）发现，与其他 3 个数据集相比，数据集 S2s1 的估算性能相对较弱。添加了 Gf1 变量的 Gf1s2 和 Gf1s1 数据集在估算精度方面表现更好，整体而言，Gf1s1 的蓄积量估算性能优于 Gf1s2 数据集；而同时包含所有类型变量的 Gf1s2s1dem 数据集又进一步提高了整体建模精度。

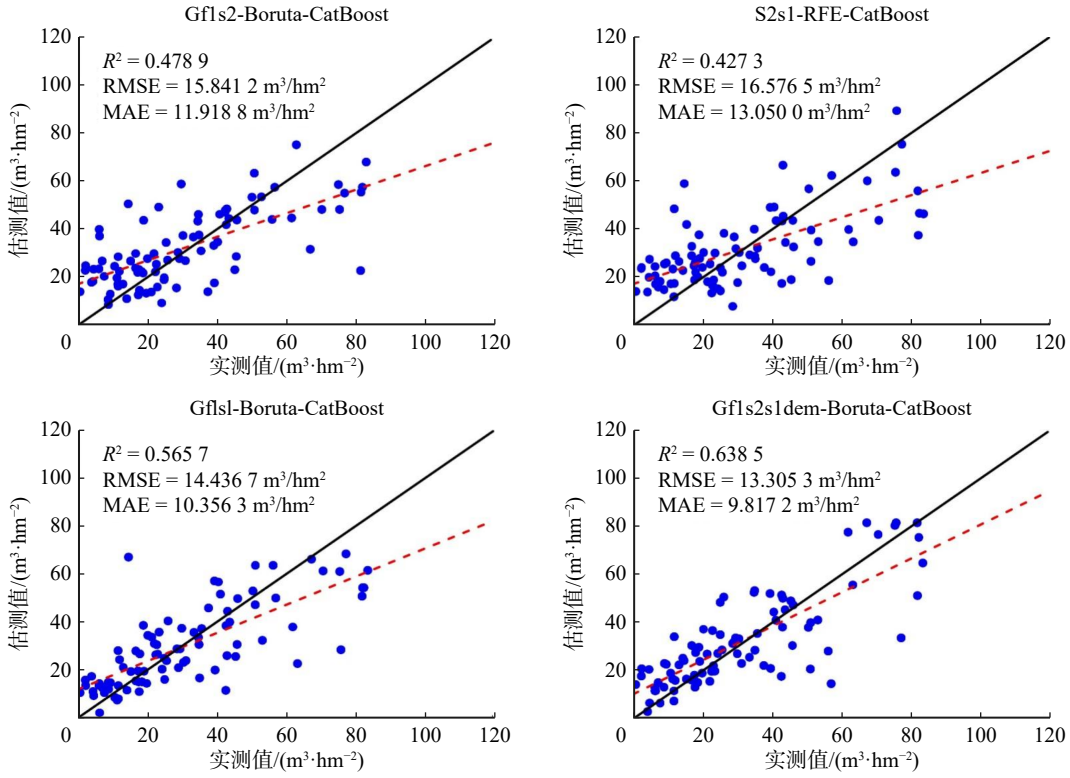
2.3　不同特征选择方法和机器学习算法对蓄积量估算模型的影响

使用 3 种特征选择方法和 5 种机器学习算法，在 4 种数据集情境下分别构建模型，模型评估结果如图 4 所示。在 3 种特征选择方法中，基于 RFE 或 Boruta 特征选择方法构建的模型整体比使用 VSURF 方法的模型表现更出色，其中，基于 Boruta 的模型性能略高于 RFE。当使用 RFE 或 Boruta 进行特征选择时，CatBoost 模型在 5 种机器学习算法中总体上表现最佳。就模型评估指标而言，基于 CatBoost 的估算模型在多数情况下均优于 RF 和 XGBoost。

综合分析得出，基于 Gf1s2s1dem 数据集（GF-1 + Sentinel-2 + Sentinel-1 + ASTER GDEM），使用 Boruta 特征选择方法和 CatBoost 机器学习算法构建的估算模型准确性最高（R^2 = 0.638 5，RMSE = 13.305 3 m^3/hm^2），为森林蓄积量的最佳优化估算模型。

2.4　不同特征选择方法中不同特征变量对蓄积量估算的影响

为进一步分析各方法选择的特征变量对蓄积量估算能力的影响，研究基于整体表现最佳的



黑色实线表示 1∶1 验证线，红色虚线表示拟合线。

图 2　4 种数据集场景下的最佳估算模型

Fig. 2　Best estimation models for four dataset scenarios

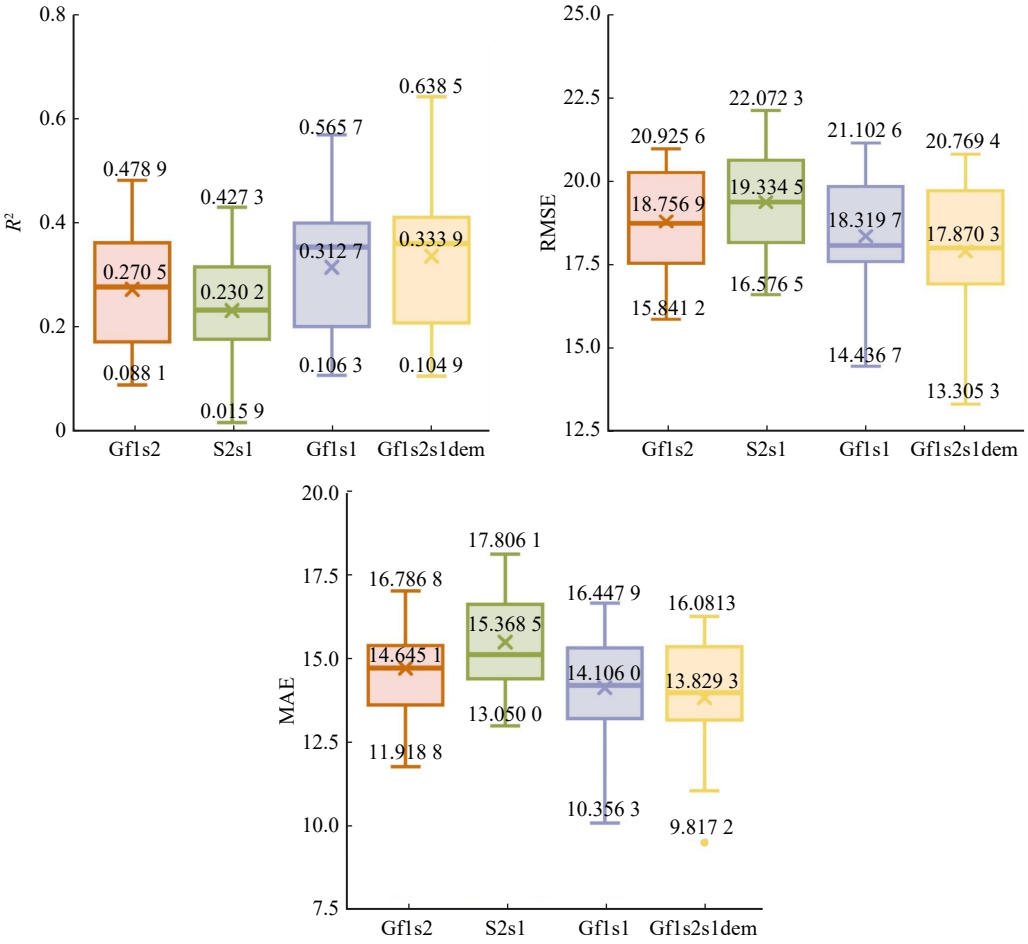


图 3 4 种数据集场景下的模型准确性评估

Fig. 3 Model accuracy evaluation for four dataset scenarios

Gf1s2s1dem 数据集，通过 Pearson 相关系数计算了所选变量之间的相关性，同时基于最优建模算法 CatBoost 评估了蓄积量估算模型的变量重要性，相关性和重要性排名最高的 8 个变量如图 5 所示。其中，B4_3Hom 与蓄积量的相关性最高，相关系数为 0.31，其次是 B4_5Ent。纹理特征 VV_9Cor、B2_9Dis 和 B4_3Dis 在 VSURF、RFE 和 Boruta 中分别具有最高的重要性。虽然各个模型选择的变量并不相似，但主要变量是纹理特征。其中，B4_3Dis 变量在 Boruta 特征选择方法中重要性最高，而在 VSURF 特征选择方法中重要性排第 2。当使用 VSURF 方法时，就模型重要性而言，Sentinel-1 表现良好；而对于 RFE 和 Boruta 方法，GF-1 比其他特征更重要。综合变量相关性和重要性结果可以看出，GF-1 和 Sentinel-1 的纹理指标对蓄积量估算有很大贡献，与 Sentinel-2 相比，GF-1 的变量在建模中具有更高的相关性和重要性。

2.5 方差分析

为进一步确定森林蓄积量建模过程中的关键因素，使用 ANOVA 来量化各因素对蓄积量估算误差的影响。表 3 显示：机器学习算法对估算误差的影响最显著，解释了 R^2 总方差的 76%，RMSE 的 74%、和 MAE 的 66%。虽然数据集和特征选择对估算精度的影响相对较小，但它们对 R^2、RMSE 和 MAE 结果均有显著影响。因此，在估算森林蓄积量时，必须综合考虑这 3 个因素对建模的影响。

2.6 森林蓄积量的预测和制图

基于上述建模结果，选取最优模型来预测研究区的森林蓄积量值，即使用 GF1s2s1dem 数据集，基于 Boruta 方法确定的遥感特征变量和 CatBoost 算法来建立蓄积量估算模型，并进行森林蓄积量分布制图。整个研究区森林蓄积量的反演结果如图 6 所示。蓄积量区域主要分布于研究区中西部平原地带，分布较为分散；高蓄积量区域在空间分布上较为集中，主要在研究区北部和西部山区。造成这种分布特征的原因在于研究区西北部山区海拔高，植被覆盖率较高且受人为干扰程度低。此外，保定市蓄积量预测值大多在 80 m^3/hm^2 以下，高蓄积量范围的样地较少，存在较多的低蓄积量林地，这与样地调查得到的林龄偏小结果一致。

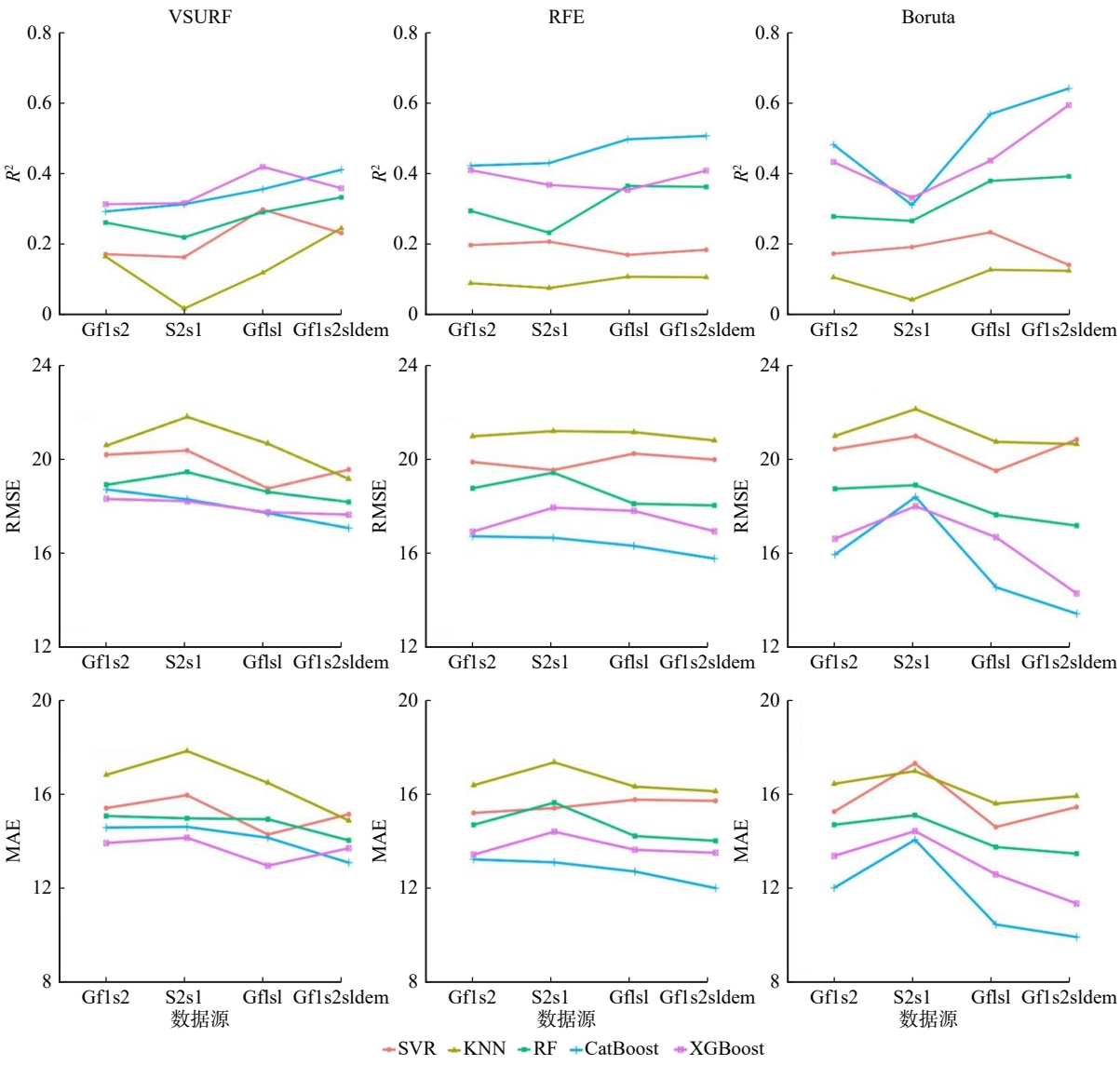


图 4　4 种数据集场景下基于 3 种特征选择方法和 5 种机器学习算法的模型准确性评估

Fig. 4　Model accuracy evaluation based on three feature selection methods and five machine learning algorithms for four dataset scenarios

3　讨　　论

森林蓄积量作为衡量森林生态系统功能的重要指标，其准确估算对于实现森林可持续经营与管理以及制定气候变化应对策略具有至关重要的作用。随着遥感技术的发展，将多源遥感数据与实地调查数据联合进行森林蓄积量的估算，已成为当前的研究热点。特征选择方法和机器学习算法能够有效提高蓄积量估算的准确性。本研究基于 4 个数据集场景，使用 3 种特征选择方法（VSURF、RFE 和 Boruta）和 5 种机器学习算法（SVR、KNN、RF、CatBoost 和 XGBoost）构建了 60 个模型。研究表明，基于 GF-1、Sentinel-2、Sentinel-1 和 ASTER GDEM 数据组合的建模性能在所有数据集场景中表现最佳。由变量选择结果和重要性排序结果可知，雷达数据在估算森林蓄积量方面的潜力低于光学传感器数据，在亚马逊潮湿热带地区的森林分类研究中也得到了类似的结论[31]。之前的研究表明，SAR 数据受云、雾等气候因素的影响较小，具有穿透林冠获取树干信息的优势，与光学数据相比，SAR 数据估算性能更好[6]。然而，在本研究中，从 Sentinel-1 提取的变量对森林蓄积量估算精度的提升效果较为有限。这可能源于两方面原因：一方面，C 波段 SAR 影像在数据处理过程中较为复杂，且受地形因素影响较大，这降低了森林结构参数提取的准确性；另一方面，可能是因为缺乏对蓄积量估算敏感的特征变量，导致建模性能较差。在本研究中，从 ASTER GDEM 数据中提取的坡度指标在蓄积量估算中表现良好，这可能是因为坡度中蕴含了部分植被信息。地形差异会导致水分、土壤和日照条件的不同，进而影响植被的生长速率和林分结构。但本研究尚未探讨不同地形条件下森林蓄积量估算的差异与联系。未来研究需要从这一

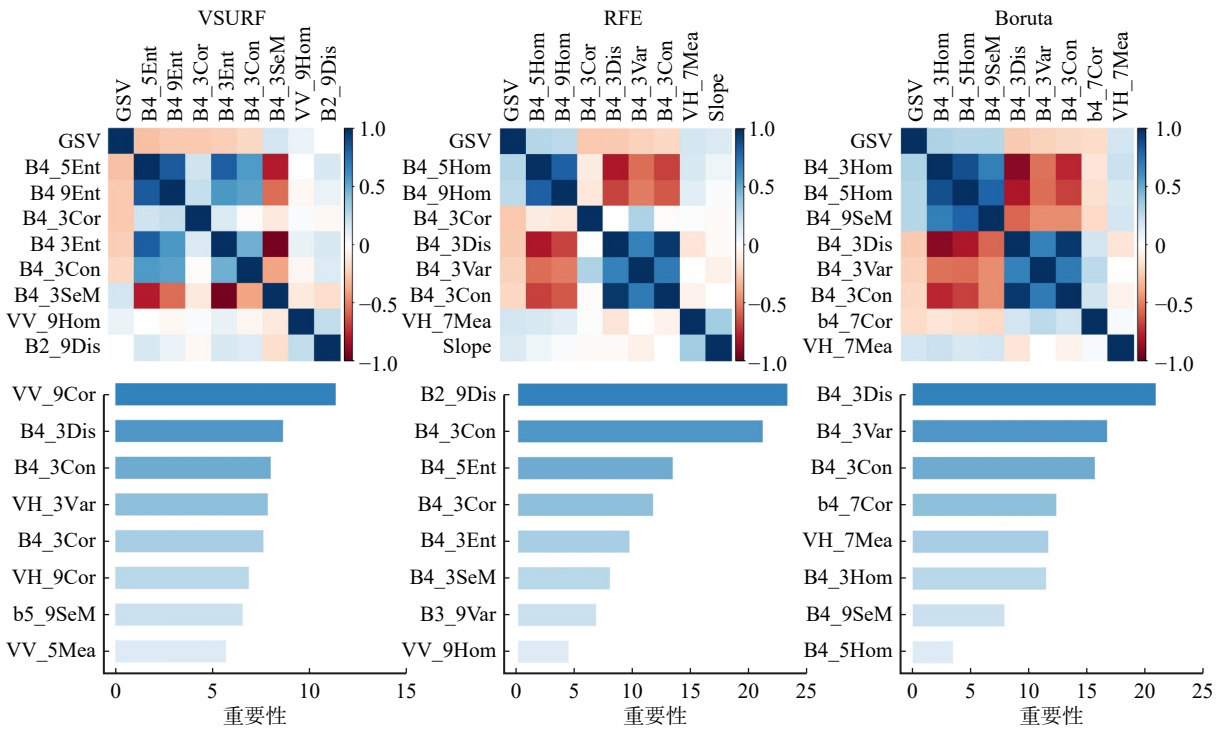


图 5 基于 3 种特征选择方法的 CatBoost 模型的相关性和重要性分析

Fig. 5 Correlation and importance analysis of CatBoost models for three feature selection methods

表 3 3 个因素的 R^2、RMSE 和 MAE 的方差分析

Tab. 3 ANOVA of R^2, RMSE, and MAE for three different factors

因素	df	R^2			RMSE/(m³·hm⁻²)			MAE/(m³·hm⁻²)		
		Sumsq	η^2	P	Sumsq	η^2	P	Sumsq	η^2	P
数据集	3	0.10	0.08	3.120×10^{-5}	17.57	0.08	5.010×10^{-6}	20.70	0.13	3.720×10^{-7}
特征选择	2	0.02	0.02	2.610×10^{-2}	4.02	0.02	9.782×10^{-3}	6.12	0.04	7.000×10^{-4}
机器学习算法	4	0.88	0.76	4.490×10^{-14}	158.26	0.74	3.990×10^{-15}	103.03	0.66	9.400×10^{-14}
数据集 × 特征选择	6	0.02	0.01	0.407 4	5.39	0.03	4.862×10^{-2}	4.60	0.03	5.070×10^{-2}
数据集 × 机器学习算法	12	0.03	0.03	0.422 0	5.75	0.03	0.256 717	3.17	0.02	0.592 62
特征选择 × 机器学习算法	8	0.06	0.05	1.280×10^{-2}	15.37	0.07	5.910×10^{-4}	10.05	0.06	3.39×10^{-3}
残差	24	0.06			8.53			7.36		

角度入手，以实现多源遥感数据的高精度森林蓄积量估算，使研究更具现实意义。

多种遥感变量可作为蓄积量建模的潜在预测因子，但由于这些变量之间存在高度相关性，同时包含这些变量并不会提高建模性能。因此，选择合适的变量选择方法尤为重要。本研究采用的多种特征选择算法在变量筛选结果上表现出明显的差异性，这主要归因于各方法的内在原理和机制不同。其中，与 GF-1 相关的变量在数量上具有明显优势，并且在蓄积量估算中发挥了重要作用。在大多数模型中，影像丰富的纹理特征信息在蓄积量预测中所起的作用比光谱波段更为显著，它能够有效描述森林的空间分布和结构信息，增强其与森林蓄积量的相关性。目前的研究结果还表明，与 VSURF 和 RFE 方法相比，使用 Boruta 方法进行特征选择时，构建的森林蓄积量模型展现出更优的估算性能。这可能是因为 Boruta 方法基于随机森林的集成学习架构，通过多棵决策树的共识决策提高了模型的稳定性，并且能够准确识别非线性关系和交互作用特征，使得所选变量具有较强的模型适用性。这一结果与 Agjee 等[12]基于多时相高光谱数据比较 RFE 和 Boruta 在生物防治剂功效检测中的研究结论相似。这些结果表明，Boruta 能够有效地筛选出不同类型的变量，并且其选择的变量集在原始特征集中保留了更多有

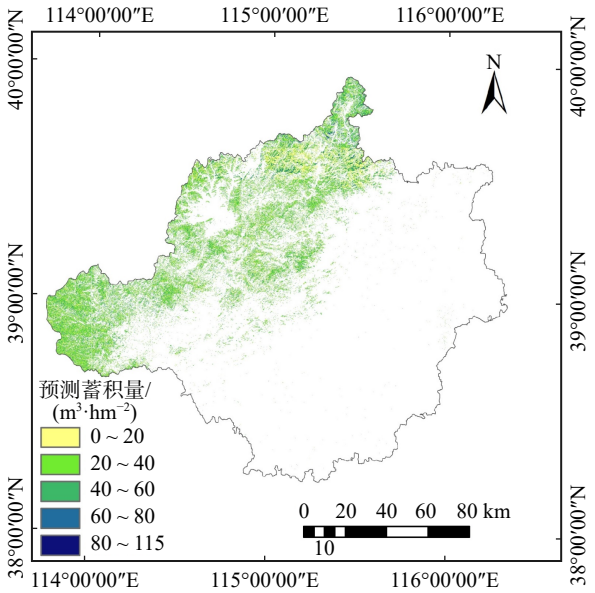


图 6　研究区森林蓄积量空间分布

Fig. 6　Spatial distribution of forest stock volume in the study area

价值的信息，证明了该方法在蓄积量预测中的有效性。

选择合适的算法对于构建蓄积量模型至关重要。在本研究中，比较了 SVR、KNN、RF、CatBoost 和 XGBoost 的估算性能。结果表明，与其他机器学习算法相比，3 种集成学习算法（RF、CatBoost 和 XGBoost）具有更高的蓄积量估算精度。这种优势主要源于集成学习算法能够在蓄积量和遥感因子之间建立复杂的非线性关系，尤其在生物物理环境复杂的林分中表现突出，并且能够灵活组合不同来源的数据。本研究发现，在 3 种集成学习算法中，与 RF 和 XGBoost 相比，CatBoost 具有更好的建模性能，并在一定程度上缓解了高值低估和低值高估问题。具体而言，CatBoost 的预测性能略微优于 XGBoost，这可能因为 CatBoost 在选择树结构时添加了计算叶节点的算法，从而有效避免了过拟合现象。类似的结论在森林地上生物量估算的研究[14]中得到证实。而 Li 等[32]在探讨机器学习算法对生物量估算潜力时报告了 XGBoost 更好的结果。这种差异可能源于本研究中使用的不同遥感数据集和特征选择方法。尽管在本研究中 CatBoost 模型的精度取得了较好的结果，但森林蓄积量建模中高值低估和低值高估的问题仍未得到彻底解决。一方面是由于光学传感器和雷达数据饱和度的限制。另一方面，也可能与研究区域采样点不足有关。收集足够数量的样本对于改进极值蓄积量估算的偏差至关重要，未来的研究需要补充地面采样点数据，基于更高分辨率的遥感数据开展估算研究，并考虑森林分层的影响，以进一步改进蓄积量估算的现有工作。本研究证明了 CatBoost 在改进蓄积量估算方面的潜力，同时也凸显了未来研究中解决森林蓄积量估算饱和问题的必要性。

传统的基于野外测量的森林蓄积量调查方法虽然较为准确，但其在大面积的森林调查中面临着周期长、成本高的问题。因此，基于多源遥感数据的森林蓄积量估算是一种非常重要的方法。国家森林资源连续清查数据在发布前经过了国家和省级森林部门的多次检查和修订，是目前省级范围内质量最高的数据之一。研究结果表明，将国家森林资源连续清查数据与遥感数据结合，可以成功地应用于蓄积量估算。然而，在保定市的地面数据中，低蓄积量样地较多，高蓄积量样地分布较少，这种不均衡的分布可能会对模型的建立以及森林蓄积量的反演精度产生一定的影响。本研究通过系统比较不同特征选择方法和机器学习算法在森林蓄积量估算中的应用效果，为提升大范围森林资源监测精度提供了重要的方法借鉴。特别是当该方法体系与国家森林资源连续清查体系相结合时，可构建一个更加完善的森林资源监测框架，为实现更精准、高效的森林资源动态监测开辟新的途径。

4　结　　论

本研究以河北省保定市为例，利用 GF-1、Sentinel-2、Sentinel-1、ASTER GDEM 和 NFCI 数据进行蓄积量遥感建模，并对 3 种特征选择方法和 5 种机器学习算法在蓄积量估算中的应用效果进行了评估，同时分析了数据集、特征选择和机器学习算法这 3 个因素对蓄积量建模性能的影响。研究得到以下结论。

（1）多源遥感数据的结合可有效提高森林蓄积量的估算精度。与其他 3 种数据集构建的模型相比，结合 GF-1、Sentinel-2、Sentinel-1 和 ASTER GDEM 数据的模型能够为蓄积量估算提供更为丰富的信息，具有更好的估算性能。其中，GF-1 为蓄积量估算提供了重要的光谱反射率信息，特别在纹理特征方面。

（2）特征选择方法可显著影响森林蓄积量的估算性能。与 VSURF 和 RFE 特征选择方法相比，基于 Boruta 构建的模型表现出更好的性能，表明使用 Boruta 方法进行特征选择在提高蓄积量模型精度方面更具有优势。CatBoost 算法与其他机器学习算法相比具有明显的改进效果，在蓄积量估算中显示出更好的潜力。

（3）方差分析结果表明，数据集、特征选择和机器学习算法均会对蓄积量估算性能产生影响。其

中，机器学习算法对模型估算性能的影响最为显著。

（4）基于 Gf1s2s1dem 数据集（GF-1 + Sentinel-2 + Sentinel-1 + ASTER GDEM），使用 Boruta 特征选择方法和 CatBoost 机器学习算法构建的估算模型准确性最高（R^2 = 0.638 5，RMSE = 13.305 3 m^3/hm^2），为森林蓄积量的最佳优化估算模型。

参 考 文 献

[1] Lindberg E, Hollaus M. Comparison of methods for estimation of stem volume, stem number and basal area from airborne laser scanning data in a hemi-boreal forest[J]. Remote Sensing, 2012, 4(4): 1004.

[2] Li X Y, Zhang M, Long J P, et al. A novel method for estimating spatial distribution of forest above-ground biomass based on multispectral fusion data and ensemble learning algorithm[J]. Remote Sensing, 2021, 13(19): 3910.

[3] Kumar L, Sinha P, Taylor S, et al. Review of the use of remote sensing for biomass estimation to support renewable energy generation[J]. Journal of Applied Remote Sensing, 2015, 9(1): 097696.

[4] Sinha S, Jeganathan C, Sharma L K, et al. A review of radar remote sensing for biomass estimation[J]. International Journal of Environmental Science and Technology, 2015, 12(5): 1779−1792.

[5] Laurin G V, Chen Q, Lindsell J A, et al. Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 89: 49−58.

[6] Soja M J, Quegan S, D'Alessandro M M, et al. Mapping above-ground biomass in tropical forests with ground-cancelled P-band SAR and limited reference data[J]. Remote Sensing of Environment, 2021, 253: 112153.

[7] López-Serrano P M, López-Sánchez C A, Alvarez-González J G, et al. A comparison of machine learning techniques applied to landsat-5 TM spectral data for biomass estimation[J]. Canadian Journal of Remote Sensing, 2016, 42(6): 690−705.

[8] Fu B L, Wang Y Q, Campbell A, et al. Comparison of object-based and pixel-based random forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data[J]. Ecological Indicators, 2017, 73: 105−117.

[9] Feng L, Li J, Gong W S, et al. Radiometric cross-calibration of Gaofen-1 WFV cameras using Landsat-8 OLI images: a solution for large view angle associated problems[J]. Remote Sensing of Environment, 2016, 174: 56−68.

[10] Liu Y A, Gong W S, Xing Y Q, et al. Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 151: 277−289.

[11] 张方圆, 吴胜义, 乔海亮, 等. 基于 Landsat 8 OLI 和资源 3 号立体像对数据的桉树森林蓄积量估测 [J]. 中南林业科技大学学报, 2024, 44(5): 93−101.
Zhang F Y, Wu S Y, Qiao H L, et al. *Eucalyptus* forest stock estimation study based on Landsat 8 OLI and Resource 3 stereo image[J]. Journal of Central South University of Forestry & Technology, 2024, 44(5): 93−101.

[12] Agjee N H, Ismall R, Mutanga O. Identifying relevant hyperspectral bands using Boruta: a temporal analysis of water hyacinth biocontrol[J]. Journal of Applied Remote Sensing, 2016, 10: 042002.

[13] Ma L, Li M C, Ma X X, et al. A review of supervised object-based land-cover image classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 277−293.

[14] Luo M, Wang Y F, Xie Y H, et al. Combination of feature selection and CatBoost for prediction: the first application to the estimation of aboveground biomass[J]. Forests, 2021, 12(2): 216.

[15] 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): 79−84, 118.
Liu Z H, Lin H, Long J P, et al. Study on volume estimation model of Wangyedian Forest Farm based on GF-2[J]. Journal of Central South University of Forestry & Technology, 2020, 40(3): 79−84, 118.

[16] 王海宾, 彭道黎, 高秀会, 等. 基于 GF-1 PMS 影像和 k-NN 方法的延庆区森林蓄积量估测 [J]. 浙江农林大学学报, 2018, 35(6): 1070−1078.
Wang H B, Peng D L, Gao X H, et al. Forest stock volume estimates in Yanqing District based on GF-1 PMS images and k-NN method[J]. Journal of Zhejiang Agriculture and Forestry University, 2018, 35(6): 1070−1078.

[17] 刘美艳, 聂胜, 王成, 等. 基于 ICESat-2 和 Sentinel-2A 数据的森林蓄积量反演 [J]. 自然资源遥感, 2024, 36(1): 210−216.
Liu M Y, Nie S, Wang C, et al. Forest stock volume inversion based on ICESat-2 and Sentinel-2A data[J]. Remote Sensing for Natural Resources, 2024, 36(1): 210−216.

[18] Zhang C Y, Denka S, Cooper H, et al. Quantification of sawgrass marsh aboveground biomass in the coastal everglades using object-based ensemble analysis and Landsat data[J]. Remote Sensing of Environment, 2018, 204: 366−379.

[19] Wang J, Xu J, Peng Y, et al. Prediction of forest unit volume based on hybrid feature selection and ensemble learning[J]. Evolutionary Intelligence, 2020, 13(1): 21−32.

[20] Haralick R M, Shanmugam K, Dinstein I. Textural features for image classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1973, 3(6): 610−621.

[21] Genuer R, Poggi J M, Tuleau-Malot C. Variable selection using

random forests[J]. Pattern Recognition Letters, 2010, 31(14): 2225−2236.

[22] Zhang Y Z, Liang S L, Sun G Q. Forest biomass mapping of Northeastern China using GLAS and MODIS data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(1): 140−152.

[23] Kursa M B, Jankowski A, Rudnicki W R. Boruta：a system for feature selection[J]. Fundamenta Informaticae, 2010, 101(4): 271−286.

[24] Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: a review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(3): 247−259.

[25] Chirici G, Mura M, McInerney D, et al. McRoberts：a meta-analysis and review of the literature on the k-nearest neighbors technique for forestry applications that use remotely sensed data[J]. Remote Sensing of Environment, 2016, 176: 282−294.

[26] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5−32.

[27] Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features [Z]. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc, 2018: 6639–6649.

[28] Chen T, Guestrin C. XGBoost: a scalable tree boosting system[Z]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: Association for Computing Machinery, 2016: 785–794.

[29] Zhang N, Chen M J, Yang F, et al. Forest height mapping using feature selection and machine learning by integrating multi-source satellite data in Baoding City, North China[J]. Remote Sensing, 2022, 14(18): 4434.

[30] Chen M J, Qiu X C, Zeng W S, et al. Combining sample plot stratification and machine learning algorithms to improve forest aboveground carbon density estimation in Northeast China using airborne LiDAR Data[J]. Remote Sensing, 2022, 14(6): 1477.

[31] Guiying L, Dengsheng L, Emilio M, et al. Comparative analysis of classification algorithms and multiple sensor data for land use/land cover classification in the Brazilian Amazon[J]. Journal of Applied Remote Sensing, 2012, 6(1): 061706.

[32] Li Y C, Li C, Li M Y, et al. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms[J]. Forests, 2019, 10(12): 1073.

（责任编辑　赵田芸
责任编委　张会儒）