

DOI: 10.20103/j.stxb.202305191060

姚雨微,任鸿瑞.青藏高原草地上生物量估算.生态学报,2024,44(7):3049-3059.

Yao Y W, Ren H R. Estimation of grassland aboveground biomass on the Qinghai-Tibet Plateau. Acta Ecologica Sinica, 2024, 44(7): 3049-3059.

青藏高原草地上生物量估算

姚雨微,任鸿瑞*

太原理工大学测绘科学与技术系,太原 030024

摘要:及时准确评估草地产草量对草地资源的科学管理和可持续发展具有重要意义。青藏高原自然环境特殊,气候差异显著,地形复杂,仅依靠遥感信息准确监测草地上生物量(Aboveground Biomass, AGB)变化有较大限制。基于青藏高原草地 AGB 野外实测数据与 Landsat 遥感影像,探索了植被指数表征草地 AGB 信息的有效性,评估了气象和地形信息对准确估算草地 AGB 的影响,综合利用气象、地形和遥感信息,在新一代地球科学数据和分析应用平台(Google Earth Engine)上构建了梯度增强回归树草地 AGB 估算模型,绘制了青藏高原多年草地 AGB 空间分布图。结果表明:(1)基于单因素遥感因子的线性回归模型仅能解释 8%—40% 的草地 AGB 变化情况,其中绿色归一化植被指数(Green Normalized Difference Vegetation Index, GNDVI)对草地 AGB 解释能力较强(40%)。(2)基于遥感因子构建的梯度增强回归树模型测试集 R^2 为 0.57。分别添加气象、地形信息,模型对草地 AGB 的估测准确性有所提升,测试 R^2 为 0.62 和 0.63。(3)基于气象、地形和遥感因子的多因素估测模型能够提高草地 AGB 估测精度,经递归特征消除法优选后,基于 13 个特征变量的梯度增强回归树模型拟合效果最好(训练数据集 $R^2 = 0.79$, RMSE = 43.42 g/m², $P < 0.01$; 测试数据集 $R^2 = 0.66$, RMSE = 53.64 g/m², $P < 0.01$),可以解释 66% 草地 AGB 变化情况。(4)2010 年青藏高原平均 AGB 为 94.58 g/m², 2015 年 93.63 g/m², 2020 年 100.78 g/m²。青藏高原西北部草地 AGB 较低,东南部草地 AGB 较高,整体呈现自西北向东南逐渐增加的分布格局。研究结果为准确估算青藏高原草地产草量和碳储量等研究提供重要参考。

关键词:青藏高原;草地上生物量;梯度增强回归树;遥感

Estimation of grassland aboveground biomass on the Qinghai-Tibet Plateau

YAO Yuwei, REN Hongrui*

Department of Geomatics, Taiyuan University of Technology, Taiyuan 030024, China

Abstract: Timely and accurate assessment of grass production is of great significance to the scientific management and sustainable development of grassland resources. The Qinghai-Tibet Plateau has specially natural environment with significant climate differences and complex topography. There are major limitations in relying only on remote sensing information to accurately monitor changes in aboveground biomass (AGB) in grasslands. In this study, based on the grassland AGB field measurement data and Landsat remote sensing images on the Qinghai-Tibet Plateau, the validity of vegetation indices in characterizing grassland AGB information was explored, and the influence of meteorological information and topographic information on accurate estimation of grassland AGB on the Qinghai-Tibet Plateau was assessed. Based on the meteorological, topographic, and remote sensing image data, gradient boosting regression tree models for estimating grassland AGB were constructed on a new generation of earth science data and analysis application platform (Google Earth Engine), and the spatial distribution of multi-year grassland AGB on the Qinghai-Tibet Plateau was mapped. The results show that: (1) The linear regression model based on single remote sensing factor could only explain 8%—40% of the changes in grassland AGB, among which the green normalized difference vegetation index (GNDVI) was more capable of explaining grassland AGB (40%). (2) The gradient boosting regression tree model constructed with remote sensing factors

基金项目:第二次青藏高原综合科学考察研究项目(2019QZKK0106)

收稿日期:2023-05-19; 网络出版日期:2024-01-12

* 通讯作者 Corresponding author. E-mail: renhongrui@tyut.edu.cn

had an R^2 of 0.57 for the test dataset. With the addition of meteorological information and topographic information, respectively, the accuracy of the model for estimating the grassland AGB was improved, with the test dataset R^2 of 0.62 and 0.63. (3) The multi-factor estimation model coupled with meteorological factors, topographic factors, and remote sensing factors could effectively improve the accuracy of grassland AGB estimation. After optimization by the recursive feature elimination method, the gradient boosting regression tree model based on 13 feature variables had the best fitting effect (training dataset: $R^2 = 0.79$, RMSE = 43.42 g/m², $P < 0.01$; testing dataset: $R^2 = 0.66$, RMSE = 53.64 g/m², $P < 0.01$), which could explain 66% of grassland AGB variation. (4) The average grassland AGB on the Qinghai-Tibet Plateau was 94.58 g/m² in 2010, 93.63 g/m² in 2015, and 100.78 g/m² in 2020. The grassland AGB in the northwestern part of the Qinghai-Tibet Plateau was lower, and the grassland AGB in the southeastern part was higher. The overall distribution pattern showed a gradual increase from northwest to southeast. The research results provide important references for studies such as the accurate estimation of grassland yield and carbon storage on the Qinghai-Tibet Plateau.

Key Words: the Qinghai-Tibet Plateau; grassland aboveground biomass; gradient boosting regression tree; remote sensing

草地是陆地生态系统的重要组成部分,在生物多样性保护和畜牧业经济中发挥着关键作用^[1-2]。草地上生物量(Aboveground Biomass, AGB)是表征植被生长状况、固碳潜力的重要指标,可以有效衡量草地生产功能,是草地利用决策和资源管理的基础^[3-4]。明晰草地 AGB 现状对草地资源管理和生态保护具有重要意义。

目前,草地 AGB 监测方法主要包括传统的地面测量和遥感监测。地面测量即直接收割法,通过割草、干燥、称重获得 AGB,但需要耗费大量时间与人力,部分地区容易受地形和环境制约,不适合用于长时间序列、大区域尺度的生物量估算^[5]。遥感监测以其宏观性、综合性等特点,克服了地面测量的不足,被视为大范围监测的有效工具^[6]。早期的草地 AGB 遥感监测主要是构建植被指数与实测草地 AGB 的经验关系,并根据该关系反演区域尺度的草地 AGB^[7]。植被指数是通过植被在不同波段部分的反射率数据进行组合和计算获得^[8-9],可以有效反映植被生长状况和变化信息。常用的植被指数有归一化差值植被指数(normalized difference vegetation index, NDVI)^[10]、增强型植被指数(enhanced vegetation index, EVI)^[11]和土壤调整植被指数(soil adjusted vegetation index, SAVI)^[12]等。但植被指数自身局限性可能会影响 AGB 估算准确性,例如在覆盖度高的草地上,NDVI 敏感性降低,易出现饱和现象^[13]。此外,草地 AGB 的变化受外部环境因素影响较大^[14]。因此基于单因素的回归模型在估算草地生物量时存在一定的误差和不确定性。对于具有多种变量的模型,机器学习模型比传统的回归模型更适用^[15-16]。机器学习模型可以整合草地 AGB 的多种影响因素,学习高度复杂的非线性映射,从而获得更好的拟合结果^[17]。与传统的回归模型相比,梯度增强回归树是一种更强大的经验建模算法。梯度增强回归树是一种迭代决策树,利用加性模型和前向分布算法将若干个弱学习器组合成强学习器^[18]。该模型可灵活处理各种数据,预测性能好,模型复杂度较低,泛化能力较强^[19]。Yang 等^[20]探索了三种机器学习方法,分别是人工神经网络、多元自适应回归样条和梯度增强回归树来建立估算模型,结果表明梯度增强回归树是生成 1 km 空间分辨率的全球森林 AGB 分布图的最优模型,模型验证 R^2 达到 0.90。Yu 等^[18]对比了梯度增强回归树、随机森林和极端随机树三种模型在生成中国三江源地区 1982—2018 年草地 AGB 数据集方面的有效性,结果表明梯度增强回归树模型精度最高。以上研究表明梯度增强回归树模型在估算草地 AGB 方面具有很大潜力。

青藏高原是世界上最高的独特自然地理单元,有“地球第三极”之称,是我国重要的生态安全屏障,也是全球气候变化的敏感区^[21-22]。草地是青藏高原的主要生态系统,由于青藏高原特殊的自然环境以及多年来气候变化和人类活动影响,草地生态系统十分脆弱,局部退化严重^[1,17,23]。及时准确的估算青藏高原草地 AGB 能够为草地资源的可持续管理提供参考意义。然而青藏高原地形地貌复杂多样,气候差异显著^[24],基于单一植被指数的回归模型无法准确反映青藏高原生态环境的影响。环境因素与草地 AGB 密切相关^[25],地形影响土壤湿度、土壤有机质含量、接收的太阳辐射强度和温度,间接影响到物种生长和草地 AGB 分布格

局^[26]。降水气温等关键因素的变化将会影响物种的分解、蒸散和光合作用等生理生态过程从而影响草地 AGB^[27]。因此构建基于气象、地形和遥感因子的多因素模型是准确估算青藏高原草地 AGB 的可能途径。

为此,本研究将青藏高原草地作为研究区,借助 Google Earth Engine(GEE)平台,探索遥感植被指数表征草地 AGB 信息的有效性,评估气象、地形信息对草地 AGB 估算的影响,构建基于气象、地形和遥感因子的青藏高原草地 AGB 估算模型—梯度增强回归树,实现青藏高原草地 AGB 空间分布制图,为准确估算青藏高原草地产草量提供重要依据。

1 材料与方法

1.1 研究区域

青藏高原地处 26°00′12″—39°46′50″N,73°18′52″—104°46′59″E 之间,面积约为 257.24 万 km²,占中国陆地面积的 26.8%^[28],主要分布在西藏自治区、青海省、甘肃省、四川省、云南省和新疆维吾尔自治区。其地势西北高,东南低,平均海拔在 4000 m 以上,气候敏感独特,东南部温暖湿润,西北部寒冷干旱。由于气候变化和地形特征的共同作用,青藏高原植被类型复杂,其中草地为主要植被类型,草地空间分布数据(图 1)来源于全球 30 m 地表覆盖数据产品 GlobeLand30^[29—30]。

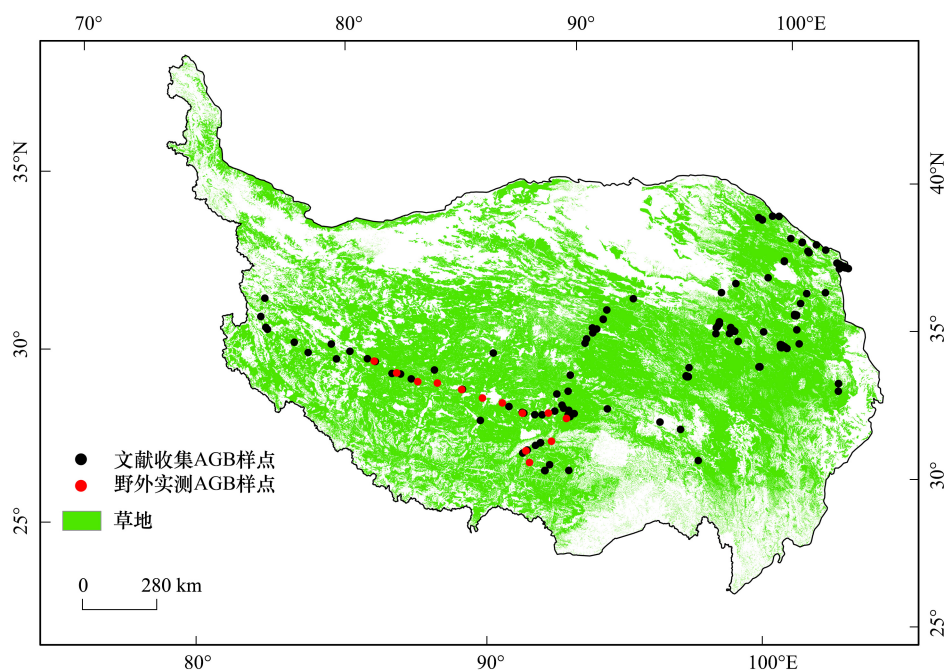


图 1 青藏高原草地 AGB 样点分布图

Fig.1 Distribution of grassland AGB sample sites on the Qinghai-Tibet Plateau

AGB;地上生物量 Aboveground Biomass

1.2 数据来源与处理

1.2.1 草地 AGB 数据

本研究中青藏高原草地 AGB 野外样点数据共 210 个(图 1),主要来源包括:(1)野外实测,共 13 个。野外调查时间在 2019 年 8 月,选择物种组成均匀、地势平整、面积足够的典型样地,记录经纬度、海拔和主要植被类型等信息,根据代表性原则在每个样地中均匀布设 3 个样方,样方大小为 1 m×1 m,将样方中的地上部分沿地面刈割,除去附着的土壤等杂物后,烘干至恒重,该样地草地 AGB 即为 3 个样方的干重取平均值。(2)文献查阅。基于中国知网(<https://www.cnki.net/>)和 Web of science(<https://www.webofscience.com/>),以“草地生物量”“生产力”“Aboveground biomass”为关键词检索,筛选样本点必须包含准确的采样时间、经纬度坐标和

植被类型,共收集到符合标准且公开发表的学术论文 34 篇,实测样点 197 个。文献中样点采样时间为 2004—2020 年 7—8 月,采样通常选择地势平坦且具有空间代表性的草地样地,其中布设 2—9 个重复样方,样方大小为 0.25 m×0.25 m—1 m×1 m,将样方内的草地地上部分齐地面刈割,放入烘箱烘干,称其干重。

1.2.2 数据资料

遥感资料采用空间分辨率为 30 m,时间分辨率为 16 d 的 Landsat 遥感卫星数据,包括 Landsat 5、Landsat 7 和 Landsat 8。1984 年 3 月发射的陆地卫星项目第五颗卫星,即 Landsat 5 卫星,携带 TM 传感器。1999 年 4 月发射的 Landsat 7 卫星,与 Landsat 5 在空间分辨率和光谱特性方面基本一致。该卫星携带 ETM+传感器,较 TM 传感器多了一个 15 m 空间分辨率的全色波段。2013 年 2 月发射的 Landsat 8 OLI 传感器较 ETM+传感器新增了深蓝和卷云波段,波段光谱范围划分更加精细。Landsat 5 TM、Landsat 7 ETM+和 Landsat 8 OLI 各个波段的具体信息见表 1。实测草地 AGB 样点采样时间为 2004—2020 年 7—8 月,对应的影像选取时间分别为 2004—2011(Landsat 5 TM)、2012(Landsat 7 ETM+)和 2013—2020(Landsat 8 OLI),通过 GEE 平台获取 Landsat 影像,经去云、裁剪等处理后,根据样点采样时间提取相应的单波段反射率值和植被指数数据,由于去云后导致部分影像数据不完整,选取采样时间临近的影像代替。

表 1 Landsat 5 TM、Landsat 7 ETM+和 Landsat 8 OLI 波段信息
Table 1 Landsat 5 TM, Landsat 7 ETM+ and Landsat 8 OLI band information

波段描述 Band description	Landsat 5 TM		Landsat 7 ETM+		Landsat 8 OLI	
	波段号 Band number	波段范围/μm Band range	波段号 Band number	波段范围/μm Band range	波段号 Band number	波段范围/μm Band range
蓝波段 Blue	1	0.45—0.52	1	0.45—0.52	2	0.45—0.51
绿波段 Green	2	0.52—0.60	2	0.52—0.60	3	0.53—0.59
红波段 Red	3	0.63—0.69	3	0.63—0.69	4	0.64—0.67
近红外波段 NIR	4	0.77—0.90	4	0.77—0.90	5	0.85—0.88
短波红外 1 波段 SWIR1	5	1.55—1.75	5	1.55—1.75	6	1.57—1.65
短波红外 2 波段 SWIR2	7	2.08—2.35	7	2.08—2.35	7	2.11—2.29

NIR:近红外波段 near infrared band;SWIR1:短波红外 1 波段 shortwave infrared 1 band;SWIR2:短波红外 2 波段 shortwave infrared 2 band

辅助数据包括地形(高程、坡度、坡向)和气象(年平均气温、年降水量)资料。高程数据来自 Shuttle Radar Topography Mission(SRTM),空间分辨率为 30 m,由云计算平台 GEE 提供。坡度和坡向数据通过高程数据计算生成。气象数据来自国家科技基础条件平台—国家地球系统科学数据中心(<http://www.geodata.cn>)的逐月气温、降水数据,空间分辨率为 1000 m^[31]。下载 2004—2020 年逐月气象数据,对每年 12 个月份的气温数据取平均值得到年平均气温,12 个月份的降水数据求和得到年降水量,将结果重采样至 30 m 空间分辨率,提取与各样点采样年份相对应的年平均气温和年降水量。

1.3 研究方法

1.3.1 建模变量集构建

研究构建的建模变量集共包含 20 个变量,其中遥感数据包括单波段反射率 Blue、Green、Red、NIR、SWIR1 和 SWIR2(表 1),常用植被指数 NDVI、EVI、GNDVI、NIR_v、RVI、SAVI、DVI、OSAVI 和 MSAVI(表 2)。地形数据包括高程(dem)、坡度(slope)和坡向(aspect)。气象数据包括年平均气温(temperature)和年降水量(precipitation)。

1.3.2 变量筛选

特征变量间的多重共线性会影响模型预测精度,而且过多的特征变量会使模型更加复杂,计算效率低。为提高模型准确性,本研究选用递归特征消除法(RFE)优选特征变量,将所有变量输入到模型中进行训练,根据模型给出的特征变量重要性,删除最差的特征,用剩下的特征变量继续输入模型训练,再删除最差的特征,重复上述步骤,直至所有特征都被删除,根据删除特征的顺序,得到最优特征子集。

表 2 本研究采用的植被指数

Table 2 Vegetation indices used in this study

植被指数 Vegetation indices	公式 Formulas	引用 References
归一化差值植被指数 Normalized difference vegetation index (NDVI)	$NDVI = \frac{NIR - Red}{NIR + Red}$	[10]
增强型植被指数 Enhanced vegetation index (EVI)	$EVI = \frac{NIR - Red}{NIR + 6Red - 7.5Blue + 1} \times 2.5$	[11]
绿色归一化植被指数 Green normalized difference vegetation index (GNDVI)	$GNDVI = \frac{NIR - Green}{NIR + Green}$	[32]
植被近红外反射率指数 Vegetation near-infrared reflectance index (NIR_v)	$NIR_v = \frac{NIR - Red}{NIR + Red} \times NIR$	[33]
比值植被指数 Ratio vegetation index (RVI)	$RVI = \frac{NIR}{Red}$	[34]
土壤调整植被指数 Soil adjusted vegetation index (SAVI)	$SAVI = \frac{NIR - Red}{NIR + Red + 0.5} \times (1 + 0.5)$	[12]
差值植被指数 Difference vegetation index (DVI)	$DVI = NIR - Red$	[35]
优化土壤调整植被指数 Optimized soil adjusted vegetation index (OSAVI)	$OSAVI = \frac{NIR - Red}{NIR + Red + 0.16}$	[36]
修改型土壤调节植被指数 Modified soil adjusted vegetation index (MSAVI)	$MSAVI = \frac{(2 \times NIR + 1) - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}}{2}$	[37]

1.3.3 模型构建

梯度增强回归树(Gradient Boosting Regression Tree, GBRT)^[38]是基于集成学习中 Boosting 算法的一种改进。基本思想是以回归树作为弱学习器,每个学习器的训练都是基于上一个学习器的预测偏差结果,沿着减少残差的梯度方向构建新的学习器,从而使预测偏差不断减小,模型精度提高,经过多次迭代后,最终结果由所有学习器互相联系得到^[19,39]。GBRT 能够适应复杂的非线性关系,产生具有竞争力、高度鲁棒、可解释的程序,尤其适用于挖掘冗杂的数据^[38]。

本研究利用 GEE 平台提供的 GBRT 算法估算草地 AGB。为提高草地 AGB 估测准确性,分别建立遥感、遥感+气象、遥感+地形、遥感+气象+地形四种 GBRT 模型。基于实测 AGB 样点,提取对应采样时间的遥感、地形和气象信息,用于模型训练。并结合递归特征消除法筛选最优变量,比较每种 GBRT 模型变量筛选前后的精度。经反复训练选取拟合效果最好的模型反演青藏高原草地 AGB 空间分布。

1.3.4 精度评价

选择样点数据的 70%作为训练集,30%作为测试集,选用五折交叉验证进行准确性评估。同时,选取决定系数 R^2 (Coefficient of Determination) 和均方根误差 RMSE (Root Mean Square Error) 对模型拟合效果进行精度评价。

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

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

式中, y_i 为第 i 个样点的实测草地 AGB; $\hat{y}_i$ 为第 i 个样点的估测草地 AGB; $\bar{y}$ 为实测草地 AGB 的平均值; n 为实测样点个数。

2 研究结果

2.1 实测草地 AGB 统计分析

青藏高原实测草地 AGB 统计分析结果表明(表 3):210 个实测样点的 AGB 值在草地生长季存在一定差

异,最小值为 2.95 g/m²、最大值为 440.46 g/m²、平均值为 135.09 g/m²,变异系数为 0.70。其中,文献收集样点共 197 个,AGB 平均值为 134.89 g/m²,变异系数为 0.71。野外实测样点共 13 个,AGB 最大值为 336.95 g/m²、最小值为 31.13 g/m²、平均值为 138.12 g/m²。

表 3 草地 AGB 观测值的统计分析

Table 3 Statistical analysis of grassland AGB observations

样点数据 Sample dataset	平均值 Average/ (g/m ²)	最大值 Maximum/ (g/m ²)	最小值 Minimum/ (g/m ²)	标准差 Standard deviation/ (g/m ²)	变异系数 Coefficient of variance
文献收集数据 Data from literature	134.89	440.46	2.95	95.30	0.71
野外实测数据 Data from field survey	138.12	336.95	31.13	77.62	0.56
总计 Total	135.09	440.46	2.95	94.31	0.70

2.2 单变量回归分析

将单变量与实测草地 AGB 进行线性回归分析,对比分析结果如下(图 2)。除坡度、坡向和年平均气温外,其余变量均通过显著性检验($P<0.01$)。遥感因子与草地 AGB 间的决定系数 R^2 在 0.08—0.40 范围内,其中,GNDVI 与 AGB 之间相关性最高, R^2 为 0.40。NDVI($R^2=0.39$) 估算 AGB 的能力略低于 GNDVI,NIR 与 AGB 相关性最低, R^2 为 0.08。气象和地形因子中,年降水量、高程与 AGB 之间相关性较高, R^2 分别为 0.28 和 0.12。经比较,GNDVI 与 AGB 的拟合效果较好,但仅能解释青藏高原 40% 的草地 AGB 变化。

2.3 模型性能评估

基于 GEE 平台构建的遥感、遥感+气象、遥感+地形、遥感+气象+地形四种 GBRT 模型性能如表 4、表 5 所

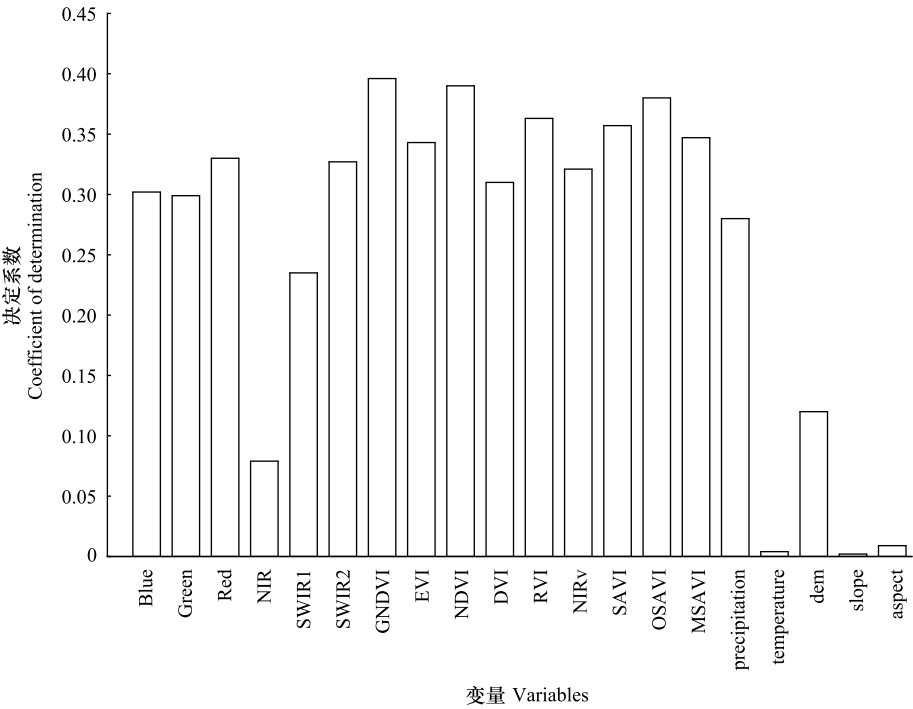


图 2 单变量与实测草地 AGB 的关系

Fig.2 Relationship between single variables and measured grassland AGB

NIR:近红外波段;SWIR1:短波红外 1 波段;SWIR2:短波红外 2 波段;GNDVI:绿色归一化植被指数;EVI:增强型植被指数;NDVI:归一化差值植被指数;DVI:差值植被指数;RVI:比值植被指数;NIR_v:植被近红外反射率指数;SAVI:土壤调整植被指数;OSAVI:优化土壤调整植被指数;MSAVI:修改型土壤调节植被指数

示。未经过 RFE 特征优选之前(表 4),基于全部遥感因子构建的 GBRT 模型测试 R^2 为 0.55, RMSE 为 61.75 g/m^2 。较基于单一植被指数的回归模型精度有明显提高,可解释 55% 草地 AGB 变化情况。在遥感因子的基础上分别添加气象、地形因子,两种模型的测试 R^2 分别为 0.59 和 0.60。综合遥感、气象和地形数据后,模型测试精度达到 0.62, RMSE 为 56.49 g/m^2 ,模型性能得到了有效提升。

基于 RFE 优选的特征变量构建的 GBRT 模型精度优于输入全部变量的模型。由表 5 可知,以 GNDVI、Green、SWIR2、EVI、NDVI、DVI、Red、SWIR1、NIR、Blue、SAVI 和 MSAVI 为输入变量的遥感模型测试 R^2 为 0.57, RMSE 为 60.27 g/m^2 。基于筛选后特征变量构建的遥感+气象、遥感+地形模型性能提高,测试 R^2 分别为 0.62 和 0.63, RMSE 分别降低 3.93 g/m^2 和 4.49 g/m^2 。以 GNDVI、precipitation、slope、temperature、dem、aspect、SWIR2、EVI、Blue、Green、NIR、SAVI 和 SWIR1 为基础构建的遥感+气象+地形 GBRT 模型,测试集 R^2 为 0.66, RMSE 为 53.64 g/m^2 ,比经过特征筛选后的遥感 GBRT 模型估测精度提高了 0.09, RMSE 降低了 6.63 g/m^2 ,模型训练集 R^2 也从 0.70 提高至 0.79。

表 4 未经过特征筛选的 GBRT 模型评估

Table 4 Evaluation of GBRT models without feature filtering

输入变量 Input variables	测试数据 Testing dataset			训练数据 Training dataset		
	决定系数 R^2	均方根误差 RMSE/(g/m^2)	P	决定系数 R^2	均方根误差 RMSE/(g/m^2)	P
遥感 R	0.55	61.75	<0.01	0.67	54.72	<0.01
遥感+气象 R+M	0.59	58.99	<0.01	0.70	51.78	<0.01
遥感+地形 R+T	0.60	57.66	<0.01	0.70	52.60	<0.01
遥感+气象+地形 R+M+T	0.62	56.49	<0.01	0.74	48.37	<0.01

R:遥感 Remote sensing; R+M:遥感+气象 Remote sensing+Meteorological; R+T:遥感+地形 Remote sensing+Topographic; R+M+T:遥感+气象+地形 Remote sensing+Meteorological+Topographic; R^2 :决定系数 coefficient of determination; RMSE:均方根误差 root mean square error; P :拟合优度 P 值 P -value

表 5 经过特征筛选的 GBRT 模型评估

Table 5 Evaluation of GBRT models with feature filtering

变量类别 Variable type	输入变量 Input variables	测试数据 Testing dataset			训练数据 Training dataset		
		决定系数 R^2	均方根误差 RMSE/ (g/m^2)	P	决定系数 R^2	均方根误差 RMSE/ (g/m^2)	P
遥感 R	GNDVI, Green, SWIR2, EVI, NDVI, DVI, Red, SWIR1, NIR, Blue, SAVI, MSAVI	0.57	60.27	<0.01	0.70	52.16	<0.01
遥感+气象 R+M	GNDVI, EVI, precipitation, SWIR2, temperature, Green	0.62	56.34	<0.01	0.75	47.91	<0.01
遥感+地形 R+T	GNDVI, slope, Green, EVI, dem, aspect, SWIR2, MSAVI	0.63	55.78	<0.01	0.77	46.10	<0.01
遥感+气象+地形 R+M+T	GNDVI, precipitation, slope, temperature, dem, aspect, SWIR2, EVI, Blue, Green, NIR, SAVI, SWIR1	0.66	53.64	<0.01	0.79	43.42	<0.01

经综合考虑气象、地形及遥感数据和通过 RFE 特征优选后,模型拟合精度逐步提高。以 GNDVI、precipitation、slope、temperature、dem、aspect、SWIR2、EVI、Blue、Green、NIR、SAVI 和 SWIR1 为基础构建的 GBRT 模型具有最佳性能,可解释青藏高原草地 AGB 66% 的变化(测试数据集 $R^2 = 0.66$, RMSE = 53.64 g/m^2 , $P < 0.01$)。由该模型估测的草地 AGB 与实测草地 AGB 的线性关系以及数据分布情况(图 3),可得两者线性关系较好,但实测 AGB 中有些较高的值被估计的偏低,存在低估情况。从数据分布而言,实测 AGB 分布范围大于估测草地 AGB,实测 AGB 数据四分位数范围在 50—180 g/m^2 ,估测 AGB 数据四分位数范围在 80—160 g/m^2 。两者在中位数、平均数上具有相似的分布特征。

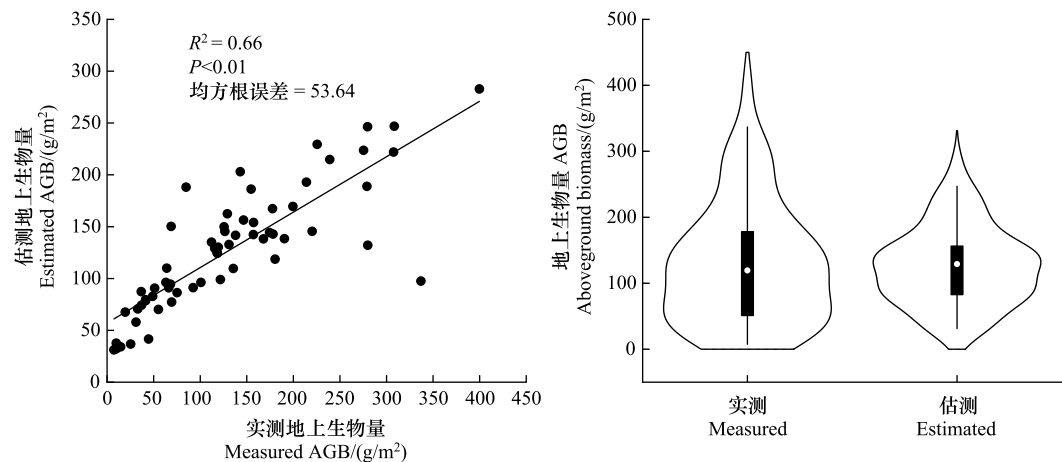


图3 实测 AGB 与最优 GBRT 模型估测 AGB 之间的线性关系和小提琴图

Fig.3 Linear and violin plots between the measured AGB and the estimated AGB from the optimal GBRT model

2.4 青藏高原草地 AGB 估算

基于最优 GBRT 模型反演了青藏高原 2010、2015 和 2020 年草地 AGB 空间分布图(图 4),草地 AGB 平均值范围为 93.63—100.78 g/m²,分别是 2010 年 94.58 g/m²,2015 年 93.63 g/m²,2020 年 100.78 g/m²。青藏高原草地 AGB 整体呈现东南高,西北低的空间格局。其中,西藏自治区阿里地区草地 AGB 较低,2010 年基本小于 60 g/m²,2015、2020 年 AGB 有明显上升;青海省果洛藏族自治州南部、黄南藏族自治州和四川省西北部草地 AGB 相对较高,与实际空间分布基本一致。

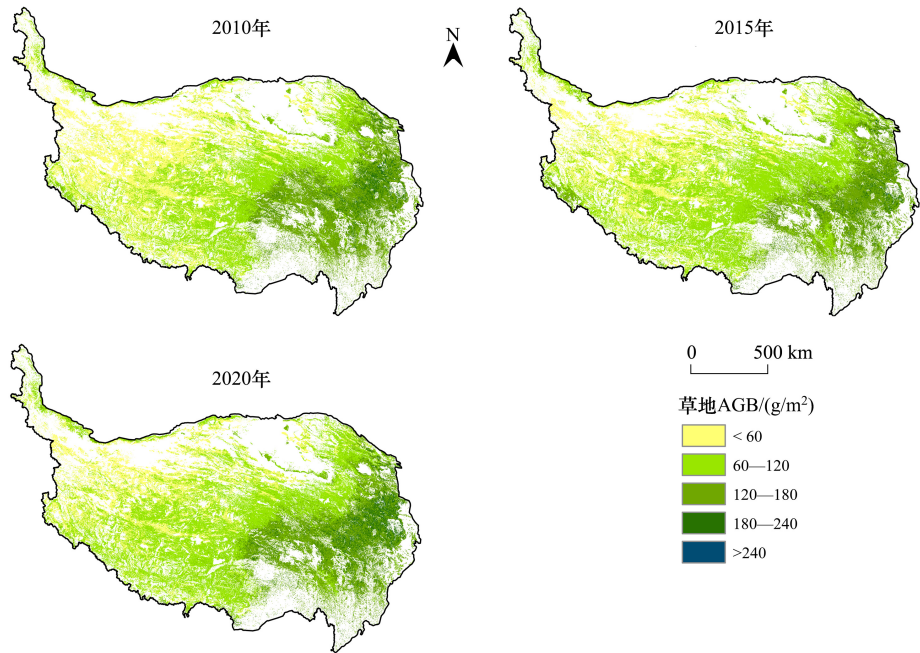


图4 青藏高原 2010、2015 和 2020 年草地 AGB 空间分布图

Fig.4 Spatial distribution of grassland AGB on the Qinghai-Tibet Plateau in 2010, 2015 and 2020

3 讨论

3.1 最优 GBRT 模型分析

研究^[40—41]表明草地 AGB 受到多种环境因素影响,其中,气象、地形因素与草地生长状况密不可分。青藏

高原气候敏感、地形险峻、生态环境复杂,为提高草地 AGB 估算准确度,基于遥感、气象和地形信息构建了 GBRT 模型,测试集 R^2 达到 0.66,表明构建多因素模型能够更好的解释草地 AGB 变化情况。这一研究结果与 Liang 等^[42]、Tang 等^[16]一致,Liang 等^[42]基于地理位置(经度、纬度)和草地覆盖度构建了三江源地区多因素非线性回归模型,与基于 NDVI 的最佳单因素模型相比,模型 RMSE 降低了 20%。Tang 等^[16]基于经度、植被指数、光照和土壤因子等环境变量构建随机森林模型估算黄河流域草地 AGB,结果表明,该模型性能高于 NDVI 构建的单变量回归模型。

图 5 为最优 GBRT 模型中特征变量重要性排序,重要性得分越高,说明该特征对模型估算结果贡献越大。遥感因子特征重要性占比 62.06%,其中,GNDVI 最高,该指数使用绿波段代替了 NDVI 中的红波段,对叶绿素含量更加敏感,赵翊含等^[43]研究表明由绿波段构建的植被指数对高寒草地 AGB 有重要作用。气象与地形因子特征重要性占比 37.94%,年降水量特征重要性最高,说明年降水量与 AGB 密切相关。大量研究也表明^[17,44]在青藏高原地区,降水比气温对草地 AGB 影响更大。气象、地形信息是影响草地 AGB 的重要环境因素,降水影响土壤含水量、空气湿度,温度影响植被的光合作用、蒸腾作用、呼吸作用^[27],地形主要影响气温、湿度和光照度^[26],以及影响土壤中物质、能量的再分配影响植被生长。因此,综合考虑了气象、地形和遥感多个因素的最优 GBRT 模型计算效率高,预测准确性强,在估算青藏高原草地 AGB 方面具有良好应用前景。

3.2 青藏高原草地 AGB 比较

本研究使用最优 GBRT 模型反演了三年青藏高原草地 AGB,空间变化呈现自西北向东南方向增加的趋势,研究结果与 Zhang 等^[41]、金哲人等^[45]一致。AGB 平均值范围为 93.63—100.78 g/m²,略高于 Zeng 等^[17] (77.12 g/m²),与 Zhang 等^[41] (102.40 g/m²) 最为接近。研究^[17,41]表明过去十几年青藏高原草地 AGB 呈上升趋势,本研究中 AGB 平均值虽有较小波动,总体仍呈上升趋势。不同研究中存在 AGB 估算差异的原因可能是:数据来源不同,包括实测 AGB 数据、遥感影像数据等。Zeng 等^[17]研究所用遥感数据是空间分辨率为 250 m 的 MODIS 数据,与本研究中数据空间分辨率不同。机器学习模型效果易受样点数量和分布情况影响^[46]。而且,不同研究所用模型不同,也会导致估算差异^[47]。

本研究中最优 GBRT 模型具有较高的精度,但机器学习模型性能依赖于训练数据。由于交通不便和海拔限制,青藏高原西北部样点较少,这可能会影响该区域草地 AGB 估算的准确性;此外,AGB 样点大多收集于不同文献,尽管操作人员在测量时严格遵守了相关规范,但不同研究的测量方法、样本选择可能存在不同,实测数据中的不确定性可能会导致估算误差。未来可继续扩大样点数量,尝试不同的深度学习算法及特征优选方法,提升草地 AGB 遥感监测和建模的准确性。

4 结论

本研究利用 GEE 平台和机器学习算法,结合详实的 2004—2020 年期间青藏高原实测草地 AGB 数据,评估了基于不同特征变量构建的 GBRT 模型的性能,估算了青藏高原草地 AGB,得到主要结论如下:

(1) 基于单一遥感因子的线性回归模型性能较差,仅能解释草地 AGB 变化的 8%—40%。基于全部遥感因子构建的 GBRT 模型测试数据集 R^2 为 0.55。

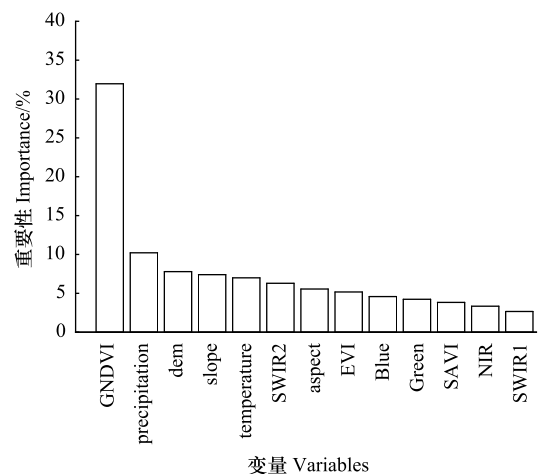


图 5 最优 GBRT 模型中特征变量重要性

Fig.5 Importance of feature variables in the optimal GBRT model

(2) 基于遥感+气象、遥感+地形的 GBRT 模型拟合效果较好, 经 RFE 特征筛选后, 测试数据集 R^2 分别为 0.62 和 0.63, 对草地 AGB 的解释能力有所增强。

(3) 在青藏高原地区, 基于气象、地形和遥感信息构建 GBRT 模型能够有效提高草地 AGB 估算精度, 模型验证表明, $R^2=0.66$, $RMSE=53.64\text{ g/m}^2$, $P<0.01$, 估测值与实测值吻合性较好, 适用于青藏高原草地 AGB 模拟。

(4) 利用最优 GBRT 模型反演青藏高原草地 AGB 分布图, 呈现东南高, 西北低的空间格局。

参考文献 (References):

- [1] Yang S X, Feng Q S, Liang T G, Liu B K, Zhang W J, Xie H J. Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. *Remote Sensing of Environment*, 2018, 204: 448-455.
- [2] Zhou W, Li H R, Xie L J, Nie X M, Wang Z, Du Z P, Yue T X. Remote sensing inversion of grassland aboveground biomass based on high accuracy surface modeling. *Ecological Indicators*, 2021, 121: 107215.
- [3] 焦翠翠, 于贵瑞, 陈智, 何念鹏. 基于遥感反演的 1982—2015 年中国北方温带和青藏高原高寒草地上生物量空间数据集. *中国科学数据*, 2019, 4(1): 35-49.
- [4] Jobbágy E G, Sala O E. Controls of grass and shrub aboveground production in the Patagonian steppe. *Ecological Applications*, 2000, 10(2): 541-549.
- [5] Jobbágy E G, Sala O E, Paruelo J M. Patterns and controls of primary production in the Patagonian steppe: a remote sensing approach. *Ecology*, 2002, 83(2): 307-319.
- [6] 修晓敏, 周淑芳, 陈黔, 蒙继华, 董文全, 杨广斌, 李晓松. 基于 Google Earth Engine 与机器学习的省级尺度零散布草地生物量估算. *测绘通报*, 2019(3): 46-52, 75.
- [7] 葛静. 中国北方地区草地上生物量遥感估测及变化分析研究[D]. 兰州: 兰州大学, 2022.
- [8] Ren H R, Zhou G S, Zhang X S. Estimation of green aboveground biomass of desert steppe in Inner Mongolia based on red-edge reflectance curve area method. *Biosystems Engineering*, 2011, 109(4): 385-395.
- [9] 姜怡忻. 协同多源卫星遥感数据的草地生物量反演方法研究[D]. 成都: 电子科技大学, 2022.
- [10] Tucker C J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 1979, 8(2): 127-150.
- [11] Huete A, Didan K, Miura T, Rodriguez E P, Gao X, Ferreira L G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 2002, 83(1/2): 195-213.
- [12] Huete A R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 1988, 25(3): 295-309.
- [13] Ren H, Feng G. Are soil-adjusted vegetation indices better than soil-unadjusted vegetation indices for above-ground green biomass estimation in arid and semi-arid grasslands? *Grass and Forage Science*, 2015, 70(4): 611-619.
- [14] Ali I, Cawkwell F, Dwyer E, Barrett B, Green S. Satellite remote sensing of grasslands: from observation to management. *Journal of Plant Ecology*, 2016, 9(6): 649-671.
- [15] Idowu S, Saguna S, Åhlund C, Schelén O. Applied machine learning: forecasting heat load in district heating system. *Energy and Buildings*, 2016, 133: 478-488.
- [16] Tang R, Zhao Y T, Lin H L. Spatio-temporal variation characteristics of aboveground biomass in the headwater of the Yellow River based on machine learning. *Remote Sensing*, 2021, 13(17): 3404.
- [17] Zeng N, Ren X L, He H L, Zhang L, Zhao D, Ge R, Li P, Niu Z E. Estimating grassland aboveground biomass on the Tibetan Plateau using a random forest algorithm. *Ecological Indicators*, 2019, 102: 479-487.
- [18] Yu R Y, Yao Y J, Wang Q, Wan H W, Xie Z J, Tang W J, Zhang Z P, Yang J M, Shang K, Guo X Z, Bei X Y. Satellite-derived estimation of grassland aboveground biomass in the three-river headwaters region of China during 1982-2018. *Remote Sensing*, 2021, 13(15): 2993.
- [19] 张宏鸣, 刘雯, 韩文霆, 刘全中, 宋荣杰, 侯贵河. 基于梯度提升树算法的夏玉米叶面积指数反演. *农业机械学报*, 2019, 50(5): 251-259.
- [20] Yang L, Liang S L, Zhang Y Z. A new method for generating a global forest aboveground biomass map from multiple high-level satellite products and ancillary information. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 2587-2597.
- [21] 李红英, 张存桂, 汪生珍, 马伟东, 刘峰贵, 陈琼, 周强, 夏兴生, 牛百成. 近 40 年青藏高原植被动态变化对水热条件的响应. *生态学报*, 2022, 42(12): 4770-4783.
- [22] 周广胜, 任鸿瑞, 刘通, 周莉, 汲玉河, 宋兴阳, 吕晓敏. 一种基于地形-气候-遥感信息的区域植被制图方法及其在青藏高原的应用. *中*

- 国科学: 地球科学, 2023, 53(2): 227-235.
- [23] 王采娥, 黄梅, 王文银, 李子好, 张涛, 马林, 白彦福, 王彦龙, 施建军, 龙瑞军, 刘玉, 王晓丽, 马玉寿, 尚占环. 三江源区高寒坡地退化植物群落多样性和地上生物量沿海拔梯度的变化特征. 生态学报, 2022, 42(9): 3640-3655.
- [24] 牛慧兰, 任鸿瑞. 青藏高原草地非光合植被覆盖度时空动态. 草业科学, 2022, 39(8): 1521-1530.
- [25] Ge J, Hou M J, Liang T G, Feng Q S, Meng X Y, Liu J, Bao X Y, Gao H Y. Spatiotemporal dynamics of grassland aboveground biomass and its driving factors in North China over the past 20 years. *Science of the Total Environment*, 2022, 826: 154226.
- [26] Shoko C, Mutanga O, Dube T. Remotely sensed C3 and C4 grass species aboveground biomass variability in response to seasonal climate and topography. *African Journal of Ecology*, 2019, 57(4): 477-489.
- [27] 郭亚奇. 气候变化对青藏高原植被演替和生产力影响的模拟[D]. 北京: 中国农业科学院, 2012.
- [28] 张德铨, 李炳元, 郑度. 论青藏高原范围与面积. 地理研究, 2002, 21(1): 1-8.
- [29] Jun C, Ban Y F, Li S N. Open access to Earth land-cover map. *Nature*, 2014, 514(7523): 434.
- [30] 陈军, 陈晋, 廖安平. 全球地表覆盖遥感制图. 北京: 科学出版社, 2016.
- [31] Peng S Z, Ding Y X, Liu W Z, Li Z. 1km monthly temperature and precipitation dataset for China from 1901 to 2017. *Earth System Science Data*, 2019, 11(4): 1931-1946.
- [32] Gitelson A A, Kaufman Y J, Merzlyak M N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 1996, 58(3): 289-298.
- [33] Badgley G, Field C B, Berry J A. Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances*, 2017, 3(3): e1602244.
- [34] Jordan C F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 1969, 50(4): 663-666.
- [35] Richardson A J, Wiegand C L. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, 1977, 43(12): 1541-1552.
- [36] Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 1996, 55(2): 95-107.
- [37] Qi J, Chehbouni A, Huete A R, Kerr Y H, Sorooshian S. A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 1994, 48(2): 119-126.
- [38] Friedman J H. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 2001, 29(5): 1189-1232.
- [39] 李建更, 张尹, 刘迎春. 基于核梯度提升树的森林高度估测方法. 北京工业大学学报, 2021, 47(10): 1113-1121.
- [40] Meng B P, Liang T G, Yi S H, Yin J P, Cui X, Ge J, Hou M J, Lv Y Y, Sun Y. Modeling alpine grassland above ground biomass based on remote sensing data and machine learning algorithm: a case study in east of the Tibetan Plateau, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 2986-2995.
- [41] Zhang X Z, Li M, Wu J S, He Y T, Niu B. Alpine grassland aboveground biomass and theoretical livestock carrying capacity on the Tibetan Plateau. *Journal of Resources and Ecology*, 2022, 13(1): 129-141.
- [42] Liang T G, Yang S X, Feng Q S, Liu B K, Zhang R P, Huang X D, Xie H J. Multi-factor modeling of above-ground biomass in alpine grassland: a case study in the Three-River Headwaters Region, China. *Remote Sensing of Environment*, 2016, 186: 164-172.
- [43] 赵翊含, 侯蒙京, 冯琦胜, 高宏元, 梁天刚, 贺金生, 钱大文. 基于 Landsat 8 和随机森林的青海门源天然草地上生物量遥感估算. 草业学报, 2022, 31(7): 1-14.
- [44] Zhu X Y, Hou J H, Li M X, Xu L, Li X, Li Y, Cheng C J, Zhao W Z, He N P. High-resolution spatial distribution of vegetation biomass and its environmental response on Qinghai-Tibet Plateau: intensive grid-field survey. *Ecological Indicators*, 2023, 149: 110167.
- [45] 金哲人, 冯琦胜, 王瑞涇, 梁天刚. 基于 MODIS 数据与机器学习的青藏高原草地上生物量研究. 草业学报, 2022, 31(10): 1-17.
- [46] Gao X X, Dong S K, Li S, Xu Y D, Liu S L, Zhao H D, Yeomans J, Li Y, Shen H, Wu S N, Zhi Y L. Using the random forest model and validated MODIS with the field spectrometer measurement promote the accuracy of estimating aboveground biomass and coverage of alpine grasslands on the Qinghai-Tibetan Plateau. *Ecological Indicators*, 2020, 112: 106114.
- [47] 宋柯馨, 蒋馥根, 胡宗达, 吕云兵, 龙依, 邓目丽, 陈松, 孙华. 西藏自治区草地地上生物量遥感反演研究. 生态学报, 2023, 43(13): 5600-5613.