基于无人机高光谱和机器学习的荒漠草原地上生物量估算

宋耀邦[1,2]，宣传忠[1,2,3※]，唐朝辉[1,2]，张 涛[1]，李 琦[1]

（1. 内蒙古农业大学机电工程学院，呼和浩特 010018；2. 内蒙古自治区草业与养殖业智能装备工程技术研究中心，呼和浩特 010018；3. 牧草饲料生产全程智能化装备内蒙古自治区工程研究中心，呼和浩特 010018）

摘 要：为提高对荒漠草原的监测水平和利用效率，实现无破坏性、准确估算荒漠草原地上生物量（aboveground biomass，AGB），避免过度放牧加速荒漠化，该研究选取内蒙古荒漠草原为研究区域，利用无人机采集研究区域高光谱数据，铺设样方采集地上生物量数据。对高光谱数据进行预处理并去除噪声，利用主成分分析法降维，将主成分、植被指数、全波段反射率数据与 AGB 进行相关性分析，并利用遗传算法进行特征选择。基于机器学习算法建立荒漠草原 AGB 估算模型，对比特征选择前后模型性能，选择最佳估算模型。结果表明：利用遗传算法筛选出的光谱特征构建机器学习模型可以提高模型的性能，最佳模型的决定系数为 0.81，均方根误差为 0.04 kg/m^2，平均绝对误差为 0.04 kg/m^2，归一化均方根误差为 0.11，模型可用于荒漠草原 AGB 的高精度估算制图。研究结果可为无人机高光谱的荒漠草原 AGB 精准估算提供数据支撑和理论依据。

关键词：高光谱；机器学习；荒漠草原；地上生物量；遗传算法

doi：10.11975/j.issn.1002-6819.202407068

中图分类号：S127;S252 **文献标志码：**A **文章编号：**1002-6819(2025)-04-0135-09

宋耀邦，宣传忠，唐朝辉，等. 基于无人机高光谱和机器学习的荒漠草原地上生物量估算[J]. **农业工程学报**，2025，41(4)：135-143. doi: 10.11975/j.issn.1002-6819.202407068 http://www.tcsae.org

SONG Yaobang, XUAN Chuanzhong, TANG Zhaohui, et al. Estimating aboveground biomass in desert steppe using UAV hyperspectral and machine learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(4): 135-143. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202407068 http://www.tcsae.org

0 引 言

草原不仅提供饲草饲料支撑畜牧业生产，在防风固沙、水土保持、水源涵养以及生物多样性保护中也扮演着重要角色[1]。荒漠草原是草原的重要组成部分，地处典型草原向荒漠的过渡带，被认为是草原的极限状态[2]。近年来，过度放牧、气候变化和矿产开采等生产活动导致部分草原退化、草地覆盖减少和荒漠化加剧等，对草地生态系统的自治性、物种多样性和生态平衡产生严重影响，导致草原的生产力下降，且对荒漠草原的影响尤为明显[3-4]。

作为描述群落结构和功能的综合信息载体，地上生物量（aboveground biomass，AGB）是草地监测中最常使用的指标之一[5]，传统的 AGB 估算方法通常依赖于地面采样，这种方法费时费力、难以大面积开展，且具有一定的破坏性[6-7]。利用卫星、多光谱遥感进行草地生物量的估算可以节省大量的人力物力和时间，在一定程度上弥补传统方法的不足[8]，但是由于云层和分辨率的限制，卫星多光谱遥感在家庭牧场等小范围应用场景下的精度不高。因此，使用无人机搭载高光谱传感器组成的低空遥感平台得到了研究人员的关注，这种遥感技术灵活、时效性强和分辨率高，可以快速部署、省时省力[9-10]。高光谱成像能够提供丰富的光谱信息，拥有更高的光谱分辨率，可以对植被特征进行微弱光谱差异的定量分析[11]。随着机器学习技术的发展，其强大的非线性拟合能力使得精确估算地上生物量成为可能[12]。通过无人机高光谱遥感技术，可以在不干扰生态系统的前提下，快速获得荒漠草原地上生物量的空间分布信息，进而有效监测和管理荒漠草原脆弱的生态系统或家庭牧场的草畜平衡[13-14]。

目前国内外已有学者利用高光谱遥感技术进行 AGB 估算，DANIELE 等[15] 利用卫星采集的巴西塞拉多草原高光谱遥感数据和植被生物量数据，以反射率、窄带植被指数、吸收带参数及其组合为特征输入，结合分类和回归树、Cubist 回归、偏最小二乘回归、随机森林、支持向量机回归五种机器学习算法，建立了巴西塞拉多草原 AGB 反演模型。但卫星高光谱数据获取难度大，处理过程复杂，分辨率较低，不适合家庭农场等小范围场景使用。高光谱数据拥有上百个波段，存在巨大的数据冗余，如何有效地实施高光谱数据的降维处理，从众多光谱数据中挖掘出对 AGB 敏感的特征以及建立准确的估算模型，是当前研究的难点。童新等[16] 以内蒙古科尔沁草场为研究对象，对原始高光谱反射率数据进行一至四阶微分处理，与植被指数结合构建 AGB 估算模型。该研究

收稿日期：2024-07-08 修订日期：2024-10-18

基金项目：国家自然科学基金项目 (31860666)；内蒙古直属高校基本科研业务费（BR221314，BR221032）；一流学科科研专项项目（YLXKZX-NND-009）

作者简介：宋耀邦，研究方向为农业工程测试与控制。

Email：s18336927727@163.com

※通信作者：宣传忠，教授，博士生导师，研究方向为草原畜牧业智能化技术与装备研究。Email：xcz@imau.edu.cn

中国知网 https://www.cnki.net

对于最佳波段和阶数的选取具有参考价值，但仅采用一元线性回归法建立估算模型，未尝试其他更复杂先进的建模方法，无法充分挖掘数据中的非线性关系，限制了模型精度。高宏元等[17]利用青海草地高光谱数据，提取原始光谱、一阶微分、植被指数、光谱位置面积参数四类特征变量，经连续投影算法特征选择后，采用随机森林算法构建 AGB 估算模型。变量选择全面，且结合了实际应用场景，但数据量有限，建模方法单一。

以上研究说明现有的草地 AGB 估算模型仍存在提升的空间，本研究旨在通过实地采集荒漠草原的无人机高光谱和 AGB 数据，对荒漠草原植被的光谱特性展开分析，筛选特征变量，通过多种机器学习算法，有针对性的建立荒漠草原的 AGB 预测模型，为荒漠草原生态系统的监测提供新的解决方案。

1 材料与方法

1.1 研究区概况

研究区域位于内蒙古自治区鄂尔多斯市鄂托克旗乌兰镇，北纬 39°00′20″，东经 107°36′33″附近。该地气候干燥，风多雨少，蒸发强烈，日照时数长，昼夜温差大。年平均降水量约为 222.0 mm，70% 的降水集中在 7—9 月份，年平均蒸发量约为 2 359.1 mm，是降水量的 8 倍以上，全年平均气温约为 7.0 ℃。地带性土壤为栗钙土，腐殖质层较厚，为栗色，地表普遍沙化，土层不厚。该地植被主要为荒漠化草原植被，以草本植物为主，主要的草本植物包括牛心朴子、苦豆子、沙蒿、猪毛蒿、狭叶锦鸡儿、蒺藜、费菜等。该研究区域内的植被覆盖度在 10%～25% 之间，平均不足 15%，是荒漠草原退化较为严重的地方。

1.2 数据采集与预处理

图 1 为利用无人机高光谱建立荒漠草原生物量估算模型总体方案，本研究利用奥普天成 ATH9010 无人机高光谱成像系统采集荒漠草原高光谱数据，该设备由高光谱成像仪、六旋翼无人机、机载电脑等组成，高光谱成像仪的波段范围为 380～1 000 nm，光谱分辨率为 1.3 nm，空间分辨率为 300 像素，光谱波段数为 480。高光谱数据采集完成后，在研究区域内铺设由 PVC（Polyvinyl Chloride）管组成的 1 m^2 样方，间隔相同距离，拍摄每个样方的照片并记录经纬度信息，样方铺设后将内部所有植物沿根部剪下，立即用电子天平称量 3 次求平均值作为样方的 AGB 数据。将高光谱数据导入 ENVI 5.3，进行几何校正、正射校正和波段定标，再利用白板光谱和原始光谱数据计算反射率数据，利用 Savitzky-Golay 方法对反射率数据进行滤波处理以降低噪声。根据经纬度信息定位样方位置，提取反射率数据，计算样方内每个波段反射率的均值，得到样方的反射率均值曲线。

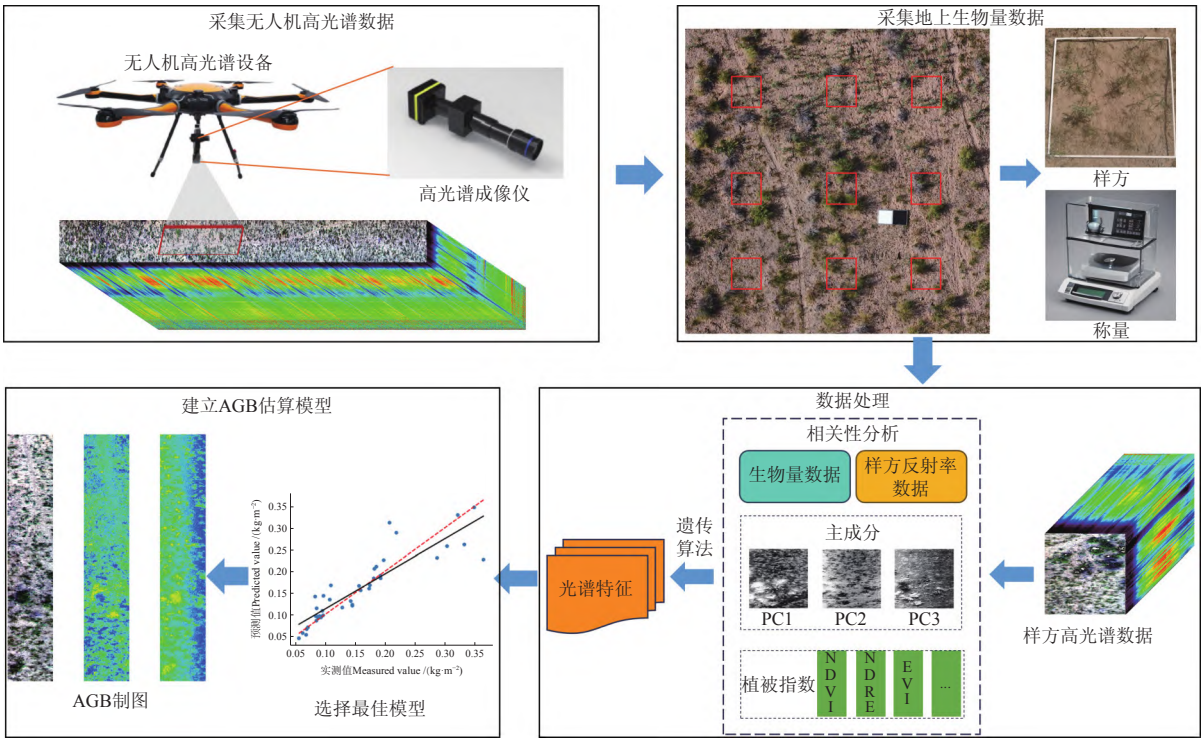


图 1 荒漠草原地上生物量估算总体方案图

Fig.1 Overall scheme of aboveground biomass (AGB) estimation in desert steppe

1.3 研究方法

1.3.1 基于主成分分析法的高光谱数据降维

高光谱数据波段多、数据量大，一些波段之间呈现较强的相关性，存在数据冗余[18-19]。高光谱降维技术利用低维数据来有效表达高维数据信息，在压缩数据量的同时提供关键的光谱特征[20]。主成分分析（principal component analysis，PCA）是最常用的降维技术之一，通过线性变换将原始数据转换到新的坐标系统中，变换后可以提取到 n 个主成分，其中第一主成分有最大的信息量，后续的主成分按信息量排列，并且与前面的主成

中国知网 https://www.cnki.net

分正交，信息量的衡量标准为主成分的方差，计算式为

$$Var(z_i) = \boldsymbol{a}_i^{\mathrm{T}} \sum \boldsymbol{a}_i \quad (1)$$

式中 $\boldsymbol{a}_i$ 是第 i 个变换向量，$\sum$是原始数据的协方差矩阵，z_i是数据降维之后的第 i 个主成分，$Var(z_i)$ 是数据降维之后的第 i 个主成分的方差[21]。当前 i 个主成分方差贡献率之和大于 85% 时，可视为包含了原数据的大部分变异信息，提取这些主成分并分别计算其平均值。

1.3.2　植被指数的计算

植被指数（vegetation index，VI）可以通过不同波段的反射率计算得出，常用来评估植物的健康、生长状况和覆盖度，是研究 AGB 的重要参数[22]，根据以往的植被 AGB 研究成果，本文选取 16 种常用的植被指数，计算式如表 1 所示。

表 1　常用植被指数及计算式

Table 1　Commonly used vegetation index and calculation formula

植被指数 Vegetation index	计算式 Calculation formula	参考 Reference
归一化植被指数 Normalized difference vegetation index, NDVI	$NDVI = (R_{800} - R_{670})/(R_{800} + R_{670})$	[15]
重归一化植被指数 Renormalized difference vegetation index，RDVI	$RDVI = (R_{800} - R_{670})/(R_{800} + R_{670})^{1/2}$	[15]
植物衰老反射指数 Plant senescence reflectance index，PSRI	$PSRI = (R_{680} - R_{500})/R_{750}$	[15]
调整土壤亮度植被指数 Soil adjusted vegetation index，SAVI	$SAVI = 1.5(R_{800} - R_{670})/(R_{800} + R_{670} + 0.5)$	[23]
修正型土壤调整植被数 Modified soil adjusted vegetation index，MSAVI	$MSAVI = 1.16(R_{800} - R_{670})/(R_{800} + R_{670} + 0.16)$	[23]
增强植被指数 Enhanced vegetation index, EVI	$EVI = (R_{800} - R_{670})/(R_{800} + 6R_{670} - 7.5R_{430} + 1)$	[24]
差值植被指数 Difference vegetation index，DVI	$DVI = R_{890} - R_{670}$	[24]
比值植被指数 Ratio vegetation index，RVI	$RVI = R_{810}/R_{660}$	[24]
改进比值植被指数 Modified simple ratio index，MSR	$MSR = (R_{800}/R_{670} - 1)/(R_{800}/R_{670} + 1)^{1/2}$	[24]
红边归一化植被指数 Normalized difference red-edge index，NDRE	$NDRE = (R_{750} - R_{705})/(R_{750} + R_{705})$	[24]
绿色归一化植被指数 Green normalized difference vegetation index，GNDVI	$GNDVI = (R_{750} - R_{550})/(R_{750} + R_{550})$	[25]
红边叶绿素指数 Red edge chlorophyll index，CIre	$CIre = (R_{750}/R_{705}) - 1$	[25]
绿色叶绿素指数 Green chlorophyll index，CIg	$CIg = (R_{750}/R_{550}) - 1$	[25]
修正型叶绿素吸收反射率指数 Modified chlorophyll absorption in reflectance index，MCARI	$MCARI = R_{705} - R_{670} - 0.2(R_{750} - R_{550})\left(\frac{R_{705}}{R_{670}}\right)$	[23]
红边指数 1 Vogelmann red edge index 1，VOG1	$VOG_1 = R_{740}/R_{720}$	[25]
红边指数 2 Vogelmann red edge index 2，VOG2	$VOG_2 = (R_{734} - R_{747})/(R_{715} + R_{726})$	[25]

注：R_{430}、R_{500}、R_{550}、R_{660}、R_{670}、R_{680}、R_{705}、R_{715}、R_{726}、R_{734}、R_{747}、R_{750}、R_{800}、R_{810}、R_{890} 分别表示波长为 430、500、550、660、670、680、705、715、726、734、747、750、800、810、890 nm 处的反射率

Note: R_{430}、R_{500}、R_{550}、R_{660}、R_{670}、R_{680}、R_{705}、R_{715}、R_{726}、R_{734}、R_{747}、R_{750}、R_{800}、R_{810} and R_{890} represent the reflectance at 430、500、550、660、670、680、705、715、726、734、747、750、800、810 and 890 nm respectively.

1.3.3　利用相关系数初步筛选特征变量

输入特征变量的选择对于预测模型的表现至关重要，选择合适的特征可以提高模型的性能，特征过多将使模型更容易过拟合，影响模型对未知数据的泛化能力[26]，因此本研究对 PCA 提取的主成分、选取的植被指数、全波段反射率与 AGB 数据进行相关性分析。通过计算各个特征与 AGB 之间的相关系数，初步筛选出与 AGB 显著相关的特征变量。

1.3.4　基于遗传算法选择最佳特征组合

遗传算法（genetic algorithm，GA）是基于达尔文进化论的优化计算模型[27]，可通过模拟自然进化过程搜索出问题的近似最优解，具有强大的全局搜索能力，基于 GA 的特征选择可以较好的解决特征选择过程中特征组合数目多、难以遍历的问题[28]。在本研究中，每个个体是一个特征组合，使用二进制编码表示，在算法的初始阶段会随机生成一组个体作为种群，通过适应度函数评估每个个体的表现，在每一代中，根据个体的适应度值选择一部分用于生成下一代，选择过程会倾向于保留具有较高适应度的个体，但也会保留少量较差的个体以保持多样性。在算法运行结束时，遗传算法会输出最优个体，即在所有代中表现最好的特征组合。

1.3.5　模型构建方法与评估指标

植物叶片的结构和组成不同，使植被的反射光谱表现出复杂的非线性[29]，而传统的多元线性回归模型并不能捕捉复杂的非线性关系[30]，因此采用机器学习（machine learning，ML）算法构建荒漠草原 AGB 估算模型。本研究使用三种不同的机器学习算法建立荒漠草原 AGB 估算模型，分别为随机森林（random forest，

中国知网　https://www.cnki.net

RF）、BP 神经网络（BP neural network，BPNN）、最小绝对收缩和选择算子（least absolute shrinkage and selection operator，LASSO）。

LASSO 在线性回归损失函数的基础上，增加了对权重的限制，通过引入正则化项，建立惩罚函数，从而实现特征选择并降低模型的复杂度，避免过拟合，这个特点使其在处理具有多重共线性或特征数量大于样本数量的数据集时表现突出[31]。RF 是一种集成学习方法，其基本单元是决策树。该算法通过构建多个决策树，并对他们的预测结果进行平均，克服单个决策树过拟合问题，在众多机器学习方法中，被认为具有较高的准确性[32]。BP 神经网络是一个多层前馈神经网络，基于梯度下降法，通过反向传播不断调整网络的权重和偏置，从而使误差函数达到要求，一般的 BPNN 模型由输入层、隐藏层和输出层三部分组成，各层之间通过神经元节点连接[33]。这三种方法各有其优势，在处理非线性问题时，相较于传统线性回归方法可以提供更高的预测精度和稳定性。

在划分数据集时，将共 100 个样本的 AGB 数据以随机采样方式，按照 6:4 的比例分割为训练集和测试集。模型性能的评估指标包括决定系数（coefficient of determination，R^2）、均方根误差（root mean squared error，RMSE）、平均绝对误差（mean absolute error，MAE）和归一化均方根误差（normalized root mean square error，NRMSE），计算式见文献 [34]。

2 结果与分析

2.1 样方 AGB 统计分布与高光谱数据预处理结果

研究区域内所采集的样方 AGB 统计分布如图 2 所示，共采集了 100 个样方，均值为 0.175 kg/m^2，标准差为 0.10 kg/m^2，变异系数为 0.576，平均绝对偏差为 0.082 kg/m^2，最小值为 0.021 kg/m^2，最大值为 0.043 kg/m^2。从图 2 中可以看出，正态分布曲线的右侧尾部较长，呈现正偏态分布，样方的 AGB 集中分布在 0.05～0.3 kg/m^2，说明荒漠草原的植被较为稀疏。

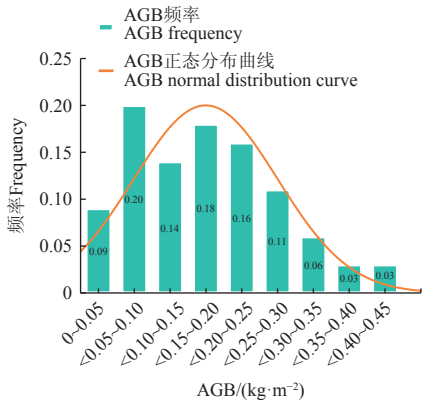


图 2 样方 AGB 统计分布直方图

Fig.2 Histogram of AGB statistical distribution in quadrats

图 3a 为选取的部分样方反射率均值曲线的原始数据，原始的高光谱数据由于大气和辐射等影响存在噪声，利用 Savitzky-Golay 方法对原始数据进行滤波处理，处理后的结果如图 3b 所示，从图中可以看出滤波处理后的样方反射率均值曲线更加平滑。样方内的 AGB 不同，其反射率均值曲线也存在差异，但总的趋势是一致的，与一般的健康绿色植物相同，在波长为 550 和 680 nm 处存在明显的反射峰和反射谷，即“绿峰”和“红谷”。

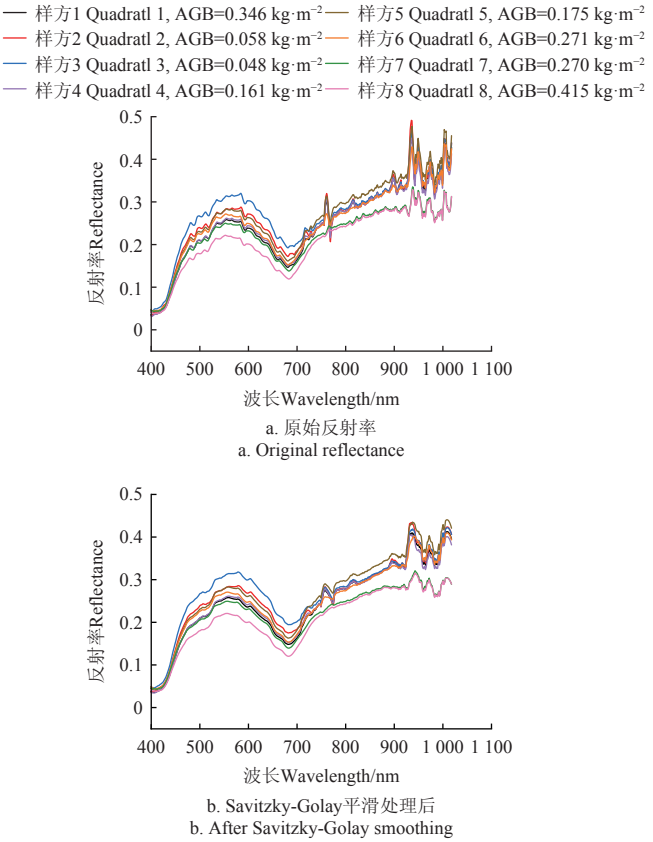


图 3 研究区域部分样方反射率均值曲线

Fig.3 Mean reflectance curve of some quadrats in study area

2.2 PCA 处理结果分析

从样方反射率数据中提取了 10 个主成分，表 2 中可以看出主成分（principal component，PC）1 的特征值和贡献率均显著大于其他主成分，说明 PC1 信息量最大。PC1、PC2、PC3 三个主成分的累计贡献率达到了 89.74%，包含原数据的大部分变异信息，可以作为估算 AGB 的特征变量。

表 2 主成分特征值和贡献率

Table 2 Principal component eigenvalue and contribution rate

主成分 Principal component	特征值 Eigenvalue	贡献率 Contribution rate/%
PC1	0.569	50.51
PC2	0.443	23.66
PC3	0.325	15.57
PC4	0.165	3.55
PC5	0.148	1.41
PC6	0.124	1.16
PC7	0.107	1.06
PC8	0.095	1.03
PC9	0.074	1.01
PC10	0.053	1.01%

中国知网 https://www.cnki.net

2.3　特征选择结果分析

2.3.1　基于相关系数的初步筛选结果

图 4 中可以看出样方内各波段反射率均值与 AGB 之间均为负相关，在可见光波段，波长为 400～687 nm 时，随着波长的增加，反射率与 AGB 的相关性呈上升趋势，在 685 nm 处达到最大，此时相关系数为-0.815，该波段位于红谷处，说明红谷附近的波段反射率与 AGB 之间存在明显的联系。波长为 685～766 nm 即红边波段范围内时，相关性随波长增加呈现下降趋势，在 766 nm 附近剧烈下降，达到最低，相关系数为−0.081。在近红外波段，波长为 766～1 018 nm 时，相关性呈现上升趋势，但总体上不如 400～687 nm 的相关性强。

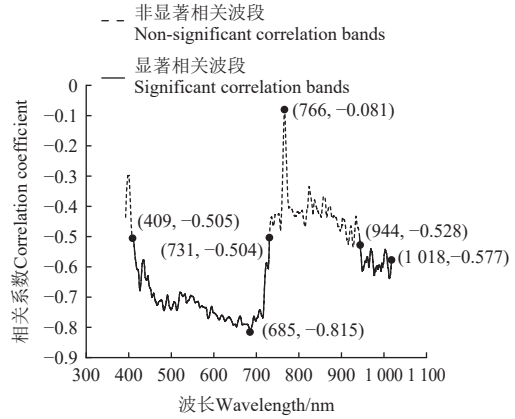
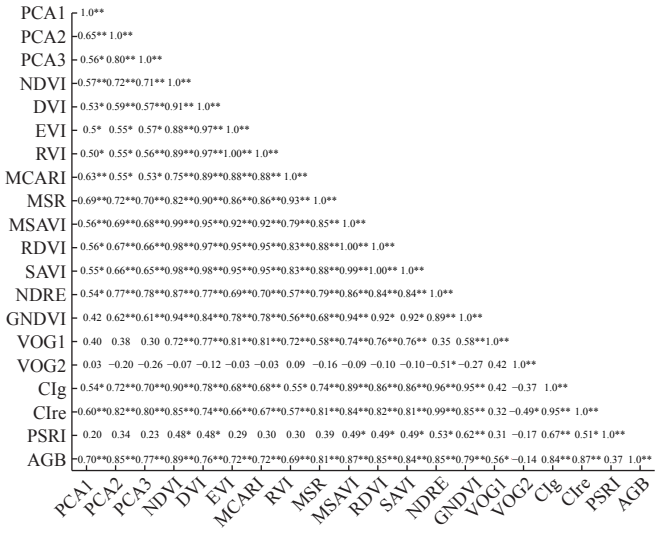


图 4　各波段与 AGB 相关系数折线图

Fig.4　Line chart of correlation coefficients between reflectance and AGB

从图 5 中可以看出 3 个主成分与 AGB 之间均呈现显著性相关，其中 PC2 的相关系数最高，PC1 最低，在选取的植被指数中除 VOG1、VOG2、PSRI 外，其余植被指数与 AGB 的相关指数均在 0.7～0.9 之间。



注：*表示 P<0.05 水平下显著相关；**表示 P<0.01 水平下显著相关。
Note: * means significant correlation at P < 0.05 level; ** means significant correlation at P < 0.01 level.

图 5　光谱变量与 AGB 之间相关系数

Fig.5　Correlation between spectral variables and AGB

在 0.05 显著性水平下，选择与 AGB 显著相关的特征和波段，一共选择了 PC1、PC2、PC3、NDVI、RDVI、EVI、DVI、RVI、MSR、SAVI、MSAVI、NDRE、GNDVI、CIre、CIg、MCARI、VOG1 共 17 个特征变量，410～730 和 942～1 018 nm，共 301 个波段。

2.3.2　基于 GA 的特征选择结果

图 6 为迭代过程中 RMSE 的变化曲线。随着种群迭代次数增加，RMSE 较高的个体逐渐被淘汰，种群的 RMSE 逐渐降低，在经过 15 次迭代后，PC1、PC3、NDVI、NDRE 作为最优特征被遗传给后代，此时对于最佳波段的选择并未结束，在第 38 代后，大量波段被淘汰，最终 536～557、673～690、703～715 nm 范围内的波段被保留，此后种群中的最优个体特征基本没有变化，RMSE 没有改善，算法在第 56 代结束。在 GA 所选择的波段当中，536～557 nm 位于反射率曲线绿峰附近，673～690 nm 位于红谷附近，703～715 nm 位于红边处。

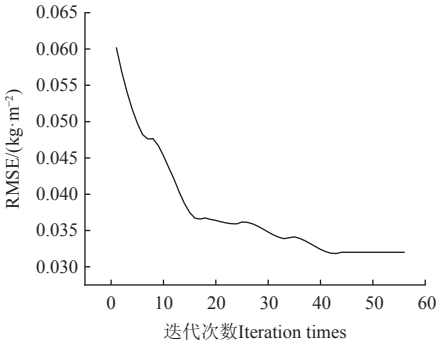


图 6　迭代过程 RMSE 变化曲线

Fig.6　RMSE change curve of iterative process

2.4　荒漠草原 AGB 估算模型

图 7a 为基于全波段反射率数据训练的荒漠草原 AGB 估算模型散点图，在使用全波段反射率数据训练的估算模型中，LASSO 模型的 R^2=0.46，RMSE=0.07 kg/m^2，MAE=0.05 kg/m^2，NRMSE=0.19，在所有模型中的性能表现最差。BPNN 模型的 R^2=0.74，RMSE=0.05 kg/m^2，MAE=0.04 kg/m^2，NRMSE=0.14，性能表现最好。对比图 7b，使用经过相关系数筛选及 GA 特征选择后的数据训练的三种估算模型在性能上均得到了不同程度的提高，其中 LASSO 模型的性能提升最大，R^2=0.76，提高了 65.2%，RMSE=0.05 kg/m^2，提高了 28.6%，MAE=0.04 kg/m^2，提高了 20%，NRMSE=0.12，提高了 36.8%。在所有的估算模型中 BPNN 模型的性能最好，拟合曲线最接近 1:1 线，R^2=0.81，RMSE=0.04 kg/m^2，MAE=0.04 kg/m^2，NRMSE=0.11，评价指标均为最高水平，可作为荒漠草原 AGB 的最佳估算模型。

从图 7 中可以看出，基于全波段数据训练得到的 BPNN 模型在 AGB 制图方面的表现不佳，与图 8a 原数据中植被的分布情况存在较大的偏差，特别是裸土区域的反演结果以及生物量较大的区域。在经过特征选择后，BPNN 模型的反演结果得到提升，与植被的实际分布情况较为一致，但在 AGB 较高的区域仍存在一些偏差。

中国知网　https://www.cnki.net

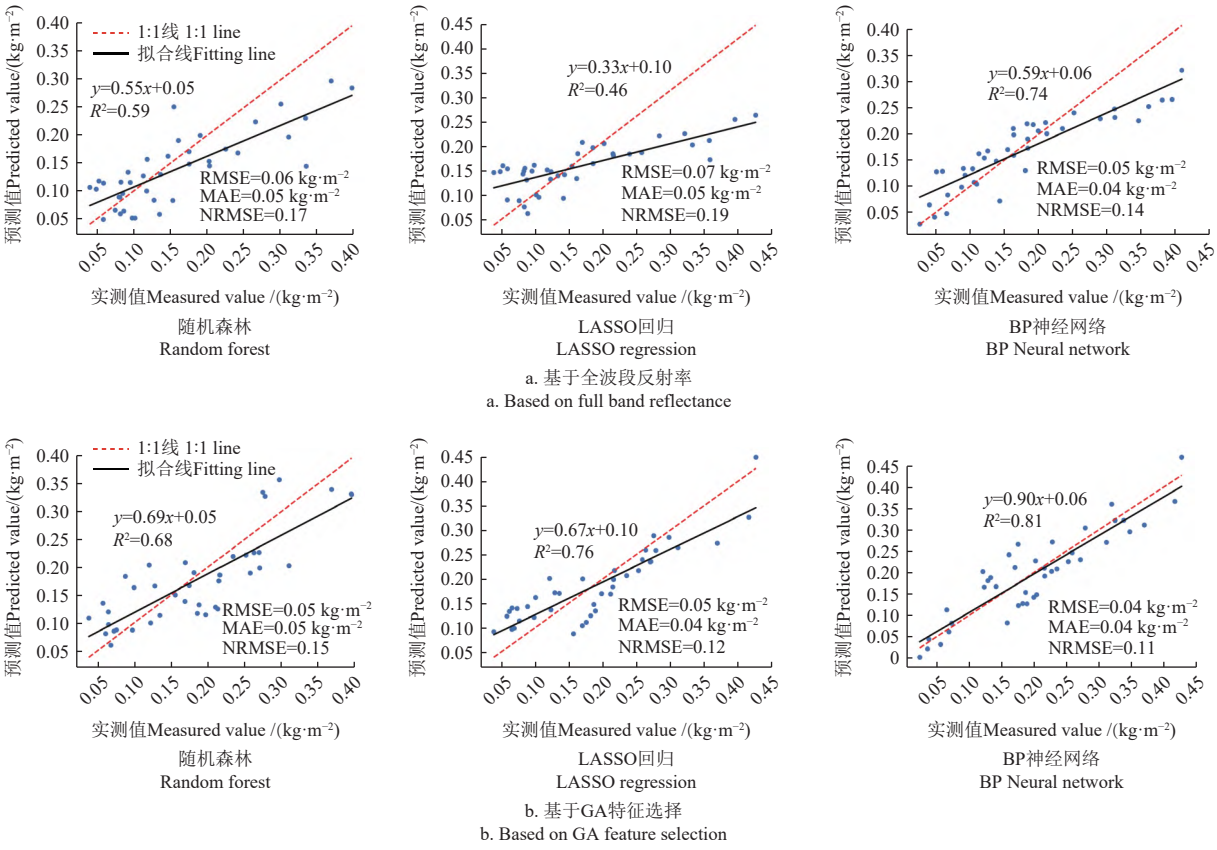


图 7 荒漠草原 AGB 估算模型散点图

Fig.7 Scatter plot of AGB estimation model in desert steppe

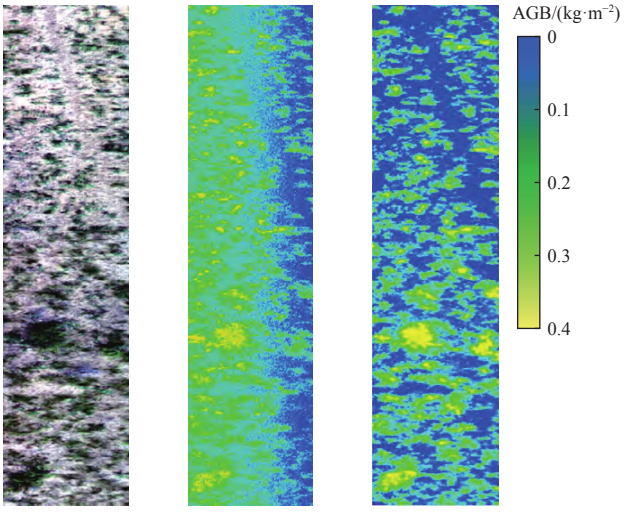


图 8 荒漠草原 AGB 空间分布

Fig.8 Spatial distribution of AGB in desert steppe

3 讨 论

与多光谱相比，高光谱技术可以在连续的光谱范围内获取数百个窄波段的光谱数据，更高的光谱分辨率使其能够敏锐地捕捉到光谱之间的细微差异。利用这些微小的差异，研究人员可以更加精准地检测植被的健康状况、氮含量、含水率等信息[35-36]。但高光谱技术也存在明显的不足，高光谱数据通常包含数百个波段，很多波段之间的信息高度相关，存在大量的数据冗余。同时这么多的波段也会带来维度灾难，在进行模型训练时需要进行波段选择和降维处理，增加了数据和处理的复杂性。

降维是研究高光谱数据的一个关键步骤，不仅可以节省算力，还能提高后续分析的效率和准确性。本文所使用的降维方法为主成分分析法，该方法能够通过线性变换将高维数据转换为低维数据，可以降低数据量，提供更明显的特征，该方法虽然是一种实用的降维方法，但是在线性变换的过程中难免忽略掉部分重要的信息，影响数据的完整性，通过改进降维算法，可得到更能代表整个光谱数据的特征。此外，利用不同的降维方法得到的特征也会存在差异，可以将多种降维方法得到的特征进行组合，与 AGB 建立回归模型，寻找对 AGB 更加敏感的特征。

植被指数是研究 AGB 的重要参数，目前已经得到了广泛应用，利用植被指数建立 AGB 预测模型是一种十分方便且实用的方法，但这种方法仅仅利用了部分波段的反射率数据，光谱数据的利用率并不高，并不能充分发挥高光谱技术的优势。从不同 AGB 样方的反射率均值之间的差异可以看出部分波段与 AGB 之间存在相关性，因此本研究计算了全波段反射率数据与 AGB 的相关系数，筛选了与 AGB 显著相关的植被指数和波段。PSRI 是一种主要用来评估植被衰老程度的光谱指数，由于本研究在采集生物量数据时并未考虑植被的不同时期，死去的植物没有被视为 AGB 采集的对象，导致该指数与 AGB

中国知网 https://www.cnki.net

的相关性不强，但这并不代表该指数没有预测 AGB 的潜力。

在特征的选择方法部分，本研究所采用的算法为 GA 遗传算法，遗传算法的优势在于其强大的全局搜索能力，通过模拟自然选择的过程来逐步优化特征组合，寻找能够使预测模型表现最好的特征子集。但是该算法也存在一些局限性，例如 GA 选择的特征组合依赖于适应度函数，因此选择具有代表性的适应度函数至关重要。此外，该算法只能根据适应度函数选择出一组最佳特征，无法解释这组特征与 AGB 之间的关系。当数据较多时，GA 迭代的速度较慢，为了确定最佳参数，需要进行复杂的预试验。

本研究使用了 RF、BPNN、LASSO 三种机器学习算法构建估算模型，分别使用全波段反射率数据和经过特征选择后的数据训练模型，结果表明当数据的维度过高时，RF 的性能不佳，因为其无法自动识别出对目标变量最敏感的波段或特征，即使在进行特征选择后，RF 仍然由于数据过于复杂和存在噪声导致性能提升不明显。LASSO 的正则化项能够有效抑制不相关特征的影响，在进行了特征选择后，LASSO 能够抓住关键特征并避免噪声干扰，从而大幅提升模型性能。BPNN 在处理高维数据时也具有较强的抗干扰能力，即使是面对全波段反射率数据也能获得较好的预测效果。所有 AGB 估算模型在 AGB 较大的样本点处均存在明显误差，其原因可能是研究区域内植被较为稀疏，采集到的样本 AGB 集中在 0.05～0.3 kg/m^2，AGB 大于 0.3 kg/m^2 的样本比较少，导致 AGB 估算模型的泛化能力弱，在估算生物量较大的样本时能力差。

尽管本研究取得了一些进展，但仍存在一些局限性。仅单期数据构建的模型仍存在许多不足，受限于特定的研究区域和植被类型，其适用性于其他区域尚需进一步验证。因此，未来研究需要在更广泛的区域内收集更多样本，采集更多时期的 AGB 和高光谱数据，针对荒漠草原植物的多个生长时期进行深入研究，以提高模型的适用性和稳定性。

4　结　论

本研究利用无人机搭载高光谱传感器组成采集系统，采集荒漠草原高光谱数据，对数据进行去噪、计算反射率、PCA 等处理，计算植被指数，结合实地采集的荒漠草原 AGB 数据，进行相关性分析，利用 GA 选择光谱特征，利用机器学习方法，建立荒漠草原地上生物量估算模型，结论如下：

1）通过 PCA 提取的三个主成分与 AGB 之间均存在较强的相关性。NDVI（normalized difference vegetation index，NDVI）、NDRE（normalized difference red-edge index，NDRE）以及 536～557、673～690、703～715 nm 范围内的波段与荒漠草原 AGB 之间存在密切的关系。

2）使用 GA 特征选择后的数据训练机器学习模型，可以提高模型的性能，LASSO 模型性能提升最大，R^2 提高 65.2%，RMSE 提高 28.6%，平均绝对误差提高 20%，归一化均方根误差提高 36.8%。BPNN 模型性能最佳，R^2=0.81，RMSE=0.04 kg/m^2，MAE=0.04 kg/m^2，NRMSE=0.11。

3）进行特征选择后的 BPNN 模型制图效果与原图植被的分布情况基本一致，验证了基于无人机高光谱的荒漠草原地上生物量估算模型的有效性。

［参　考　文　献］

[1] 赵越，徐大伟，范凯凯，等. Landsat 8 和机器学习估算蒙古高原草地地上生物量[J]. 农业工程学报，2022，38(24)：138-144.
ZHAO Yue, XU Dawei, FAN Kaikai, et al. Estimating above-ground biomass in grassland using Landsat 8 and machine learning in Mongolian Plateau[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(24): 138-144. (in Chinese with English abstract)

[2] 赵天启，古琛，王亚婷，等. 内蒙古不同草地类型群落物种多度分布格局分析[J]. 中国草地学报，2017，39(5)：68-75.
ZHAO Tianqi, GU Chen, WANG Yating, et al. Analysis on the species-abundance distribution patterns of different grassland communities in Inner Mongolia[J]. Chinese Journal of Grassland, 2017, 39(5): 68-75. (in Chinese with English abstract)

[3] 陈林，曹萌豪，宋乃平，等. 中国荒漠草原的研究态势与热点分析：基于文献计量研究[J]. 生态学报，2021，41(24)：9990-10000.
CHEN Lin, CAO Menghao, SONG Naiping, et al. Research situation and hotspot analysis of desert grassland in China based on bibliometric research[J]. Acta Ecologica Sinica, 2021, 41(24): 9990-10000. (in Chinese with English abstract)

[4] ZHANG Y Z, WANG Q, WANG Z Q, et al. Impact of human activities and climate change on the grassland dynamics under different regime policies in the Mongolian Plateau[J]. The Science of the Total Environment, 2020, 698: 134304.

[5] CHEN A, XU C, ZHANG M, et al. Cross-scale mapping of above-ground biomass and shrub dominance by integrating UAV and satellite data in temperate grassland[J]. Remote Sensing of Environment, 2024, 304: 114024.

[6] 刘艳，聂磊，杨耘. 基于植被指数估算天山牧区不同利用类型草地总产草量[J]. 农业工程学报，2018，34(9)：182-188.
LIU Yan, NIE Lei, YANG Yun. Estimation of total yield of different grassland types in Tianshan pastoral area based on vegetation index[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(9): 182-188. (in Chinese with English abstract)

[7] GUIMARÃES‐STEINICKE C, WEIGELT A, EBELING A, et al. Diversity effects on canopy structure change throughout a growing season in experimental grassland communities[J]. Remote Sensing, 2022, 14(7): 1557.

[8] FILHO M G, KUPLICH T M, QUADROS F L F D. Estimating natural grassland biomass by vegetation indices using Sentinel-2 remote sensing data[J]. International Journal of Remote Sensing, 2020, 41(8): 2861-2876.

[9] 李琦，宣传忠，唐朝辉，等. 基于遥感技术的植被生物量反演研究进展[J]. 内蒙古农业大学学报（自然科学版），2024，45(4)：88-100.
LI Qi, XUAN Chuanzhong, TANG chaohui, et al. A review on vegetation biomass inversion based on remote sensing technology[J]. Journal of Inner Mongolia Agricultural University (Natural Science Edition), 2024, 45(4): 88-100. (in Chinese with English abstract)

[10] 王震，李映雪，吴芳，等. 冠层光谱红边参数结合随机森林机器学习估算冬小麦叶绿素相对含量[J]. 农业工程学报，

中国知网　https://www.cnki.net

2024，40(4)：171-182.
WANG Zhen, LI Yingxue, WU Fang, et al. Estimation of winter wheat chlorophyll relative content combing with canopy spectrum red edge parameters and random forest machine learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(4): 171-182. (in Chinese with English abstract)

[11] ZHONG Y F, WANG X Y, XU Y, et al. Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications[J]. IEEE Geoscience and Remote Sensing Magazine, 2018, 6(4): 46-62.

[12] 张继贤，刘飞，王坚. 轻小型无人机测绘遥感系统研究进展[J]. 遥感学报，2021，25(3)：708-724.
ZHANG Jixian, LIU Fei, WANG Jian. Review of the light-weighted and small UAV system for aerial photography and remote sensing[J]. National Remote Sensing Bulletin, 2021, 25(3): 708-724. (in Chinese with English abstract)

[13] LYU X, LI X B, GONG J R, et al. Remote-sensing inversion method for aboveground biomass of typical steppe in Inner Mongolia, China[J]. Ecological Indicators, 2021, 120: 106883.

[14] 赵春江. 农业遥感研究与应用进展[J]. 农业机械学报，2014，45(12)：277-293.
ZHAO Chunjiang. Advances of research and application in remote sensing for agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(12): 277-293. (in Chinese with English abstract)

[15] DANIELE J A, SOARES G L, RICARDO D, et al. Aboveground biomass estimates over Brazilian savannas using hyperspectral metrics and machine learning models: Experiences with Hyperion/EO-1[J]. GIScience & Remote Sensing, 2021, 58(7): 1112-1129.

[16] 童新，杨震雷，张亦然，等. 基于不同阶微分高光谱植被指数的牧区草场地上生物量估算[J]. 草地学报，2022，30(9)：2438-2448.
TONG Xin, YANG Zhenlei, ZHANG Yiran, et al. Estimation of pasture aboveground biomass using different orders of differential hyperspectral vegetation indices[J]. Acta Agrestia Sinica, 2022, 30(9): 2438-2448. (in Chinese with English abstract)

[17] 高宏元，侯蒙京，葛静，等. 基于随机森林的高寒草地地上生物量高光谱估算[J]. 草地学报，2021，29(8)：1757-1768.
GAO Hongyuan, HOU Mengjing, GE Jing, et al. Hyperspectral estimation of aboveground biomass of alpine grassland based on random forest algorithm[J]. Acta Agrestia Sinica, 2021, 29(8): 1757-1768. (in Chinese with English abstract)

[18] 杨红艳，杜健民，阮培英，等. 基于无人机遥感与随机森林的荒漠草原植被分类方法[J]. 农业机械学报，2021，52(6)：186-194.
YANG Hongyan, DU Jianmin, RUAN Peiying, et al. Vegetation classification of desert steppe based on unmanned aerial vehicle remote sensing and random forest[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(6): 186-194. (in Chinese with English abstract)

[19] 刘浩，杨锡震，张蓓，等. 基于分数阶微分光谱指数的冬小麦根域土壤含水率估算模型[J]. 农业工程学报，2023，39(13)：131-140.
LIU Hao, YANG Xizhen, ZHANG Bei, et al. Estimation model of soil moisture content in root domain of winter wheat using afractional-order differential spectral index[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(13): 131-140. (in Chinese with English abstract)

[20] 童庆禧，张兵，张立福. 中国高光谱遥感的前沿进展[J]. 遥感学报，2016，20(5)：689-707.
TONG Qingxi, ZHANG Bing, ZHANG Lifu. Current progress of hyperspectral remote sensing in China[J]. Journal of Remote Sensing, 2016, 20(5): 689-707. (in Chinese with English abstract)

[21] 金文玲，曹乃亮，朱明东，等. 基于近红外超连续激光光谱的水稻种子活力无损分级检测研究[J]. 中国光学，2020，13(5)：1032-1043.
JIN Wenling, CAO Nailiang, ZHU Mingdong, et al. Nondestructive grading test of rice seed activity using near infrared super-continuum laser spectrum[J]. Chinese Optics, 2020, 13(5): 1032-1043. (in Chinese with English abstract)

[22] 孙世泽，汪传建，尹小君，等. 无人机多光谱影像的天然草地生物量估算[J]. 遥感学报，2018，22(5)：848-856.
SUN Shize, WANG Chuanjian, YIN Xiaojun, et al. Estimating aboveground biomass of natural grassland based on multispectral images of unmanned aerial vehicles[J]. Journal of Remote Sensing, 2018, 22(5): 848-856. (in Chinese with English abstract)

[23] 冯惠芬，李映雪，吴芳，等. 机器学习结合高光谱植被指数与 SPAD 值估算冬小麦氮含量[J]. 农业工程学报，2024，40(1)：235-245.
FENG Huifen, LI Yingxue, WU Fang, et al. Estimating winter wheat nitrogen content using SPAD and hyperspectral vegetation indices with machine learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(1): 235-245. (in Chinese with English abstract)

[24] 严海军，卓越，李茂娜，等. 基于机器学习和无人机多光谱遥感的苜蓿产量预测[J]. 农业工程学报，2022，38(11)：64-71.
YAN Haijun, ZHUO Yue, LI Maona, et al. Alfalfa yield prediction using machine learning and UAV multispectral remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(11): 64-71. (in Chinese with English abstract)

[25] 王进，周广胜，何奇瑾，等. 内蒙古克氏针茅草原净生态系统碳交换的物候特征及遥感监测[J]. 应用生态学报，2024，35(3)：659-668.
WANG Jin, ZHOU Guangsheng, HE Qijin, et al. Phenological characteristics of net ecosystem carbon exchange of Stipa kryloi steppe in Inner Mongorlia, China and its remote sensing monitoring[J]. Chinese Journal of Applied Ecology, 2024, 35(3): 659-668. (in Chinese with English abstract)

[26] XU D W, WANG C, CHEN J, et al. The superiority of the normalized difference phenology index (NDPI) for estimating grassland aboveground fresh biomass[J]. Remote Sensing of Environment, 2021, 264: 112578.

[27] 王梦迪，何莉，刘潜，等. 基于小麦冠层无人机高光谱影像的农田土壤含水率估算[J]. 农业工程学报，2023，39(6)：120-129.
WANG Mengdi, HE Li, LIU Qian, et al. Estimating soil moisture contents of farmland using UAV hyperspectral images of wheat canopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(6): 120-129. (in Chinese with English abstract)

[28] SHAHLA N, EHSAN B, NASSER G A, et al. A novel ACO-GA hybrid algorithm for feature selection in protein function prediction[J]. Expert Systems with Applications, 2009, 36(10): 12086-12094.

[29] LI B, XU X M, ZHANG L, et al. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162: 161-172.

[30] MENG B P, LIANG T G, YI S H, et al. 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[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 2986-2995.

[31] GUO R, GAO J L, FU S, et al. Estimating aboveground

中国知网 https://www.cnki.net

biomass of alpine grassland during the wilting period using in Situ Hyperspectral, Sentinel-2, and Sentinel-1 data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-16.

[32] LI Z H, ZHAO Y, TAYLOR J, et al. Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data[J]. Remote Sensing of Environment, 2022, 273: 112967.

[33] 付波霖，孙军，李雨阳，等. 基于多光谱影像和机器学习算法的红树林树种 LAI 估算[J]. 农业工程学报，2022，38(7)：218-228.
FU Bolin, SUN Jun, LI Yuyang, et al. Mangrove LAI estimation based on remote sensing images and machine learning algorithms[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(7): 218-228. (in Chinese with English abstract)

[34] 陶惠林，徐良骥，冯海宽，等. 基于无人机高光谱遥感数据的冬小麦产量估算[J]. 农业机械学报，2020，51(7)：146-155.
TAO Huilin, XU Liangji, FENG Haikuan, et al. Winter wheat yield estimation based on UAV hyperspectral remote sensing data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 146-155. (in Chinese with English abstract)

[35] 彭涛，赵丽，张爱军，等. 土壤全氮的无人机高光谱响应特征及估测模型构建[J]. 农业工程学报，2023，39(4)：92-101.
PENG Tao, ZHAO Li, ZHANG Aijun, et al. UAV hyperspectral response characteristics and estimation model construction of soil total nitrogen[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 92-101. (in Chinese with English abstract)

[36] 王新云，郭艺歌，何杰. 基于多源遥感数据的草地生物量估算方法[J]. 农业工程学报，2014，30(11)：159-166.
WANG Xinyun, GUO Yige, HE Jie. Estimation of above-ground biomass of grassland based on multi-source remote sensing data[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(11): 159-166. (in Chinese with English abstract)

Estimating aboveground biomass in desert steppe using UAV hyperspectral and machine learning

SONG Yaobang[1,2] , XUAN Chuanzhong[1,2,3※] , TANG Zhaohui[1,2] , ZHANG Tao[1] , LI Qi[1]

(1. *College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot* 010018, *China*; 2. *Inner Mongolia Engineering Research Center for Intelligent Facilities in Prataculture and Livestock Breeding, Hohhot* 010018, *China*; 3. *Inner Mongolia Engineering Research Center of Intelligent Equipment for the Entire Process of Forage and Feed Production, Hohhot* 010018, *China*)

Abstract: Grassland can provide the abundant forage and feed resources for the livestock industry, in order to maintain the ecological balance and biodiversity. One special types of the grassland, desert steppe is located at the transitional zone between grasslands and deserts, particularly with the relatively fragile ecosystem at risk of desertification. Aboveground biomass is one of the most important indicators to monitor the community structure and function in the grassland. However, the traditional estimation of aboveground biomass cannot fully meet the monitoring needs of desert grassland in a large area, due to the time-consuming, destructiveness and laborious. Satellite multi-spectral remote sensing can be expected to serve as such application against the clouds and resolution. In this study, non-destructive and accurate estimation was performed on the aboveground biomass in the desert grassland, in order to improve the monitoring level and utilization efficiency. The study area was taken from the desert grassland at Wulan Town, Etuoke Banner, Ordos City, Inner Mongolia, China. A drone was utilized to capture the hyperspectral data in the study area. Sample plots were established to collect the aboveground biomass data. A series of preprocessing was carried out on the hyperspectral datasets. Firstly, the average reflectance of each sample plot was precisely calculated to standardize the data. Subsequently, the advanced techniques of noise reduction were applied to eliminate any potential noise interference, particularly for the integrity and reliability of data. Principal component analysis (PCA) was then utilized to reduce the dimensionality. Three principal components were extracted successfully, including PC1, PC2, and PC3. These principal components were effectively condensed to highlight the key spectral features in the high-dimensional hyperspectral data. In parallel, 16 vegetation indices were calculated using the reflectance data. These indices were widely recognized in the field of vegetation research, in order to characterize the physiological and ecological status of the vegetation. After that, the correlation analysis was conducted among the principal components, vegetation indices, full-band reflectance data, and AGB. The relationships among these variables were then obtained to identify the most significant influencing factors on the AGB. A genetic algorithm was employed to further optimize the feature set. This algorithm was inspired by the principles of natural evolution, indicating the strong global searching. Finally, an optimal combination of optimal features was selected after continuous iteration and evaluation, including PC1, PC3, NDVI, NDRE. Among them, the specific bands were in the ranges of 536-557 nm, 673-690 nm, and 703-715 nm, respectively. According to these selected features, different estimation models of aboveground biomass were developed for the desert grasslands using Random Forest, BP Neural Network, and LASSO regression. A ten-fold cross-validation was applied to evaluate the performance of the improved model. The results demonstrate that the better performance of the improved model was achieved in the machine learning with the spectral features that selected by the genetic algorithm. Among them, the LASSO model exhibited the greatest performance, indicating R^2=0.76, an increase of 65.2%, RMSE=0.05 kg/m^2, an improvement of 28.6%, MAE (mean absolute error)=0.04 kg/m^2, an increase of 20%, NRMSE (normalized root mean square error)=0.12, and an improvement of 36.8%. The BP Neural Network model performed the best, with R^2=0.81, RMSE=0.04 kg/m^2, MAE=0.04 kg/m^2, and NRMSE=0.11. A consistence was found between the estimation and mapping of desert grassland AGB with the BPNN model and the actual distribution of vegetation. Therefore, the UAV-based hyperspectral data can be expected to construct an AGB estimation model for the desert grasslands. The finding can provide a scientific basis to formulate the grazing plans, in order prevent the overgrazing from the desertification in sustainable ranches.

Keywords: hyperspectral; machine learning; desert steppe; aboveground biomass; genetic algorithm

中国知网　https://www.cnki.net