Landsat8 光谱衍生数据分类体系下的牧草生物量反演

张爱武[1,2]，张 帅[1,2]，郭超凡[1,2*]，刘路路[1,2]，胡少兴[3]，柴沙驼[4]

1. 首都师范大学三维信息获取与应用教育部重点实验室，北京 100048
2. 首都师范大学空间信息技术教育部工程研究中心，北京 100048
3. 北京航空航天大学机械工程及自动化学院，北京 100191
4. 青海大学畜牧兽医科学院(青海省畜牧兽医科学院)，青海 西宁 810016

摘 要 牧草生物量的估算对于草地资源合理利用和载畜平衡监测具有重要的意义，是评价草地生态系统与草地资源可持续发展的关键指标。基于 Landsat 遥感技术快速、无损的大面积植被生物量估算研究已广泛应用，当前大多基于单一变量或几个常用植被指数构建反演模型，这些指数往往不能从多方面反映植被理化特征。归纳了不同 Landsat8 光谱衍生数据所反映的植被理化特征及它们间的关联方式，构建了 Landsat8 光谱衍生数据的分类体系；在此基础上提出了一种基于随机梯度 Boosting(SGB)算法的多变量、非线性生物量估算模型，探讨不同类型光谱衍生数据组合对于牧草生物量反演结果的影响。以青海省海晏县为研究区进行方案可行性探讨。结果表明常用的 Landsat8 光谱衍生数据主要从植被的绿度、黄度、盖度、水分含量、纹理特征以及通过消除大气干扰和土壤背景干扰等 7 个方面反映植被的理化特征(7 个小类)，可归纳为直接因子(绿度、黄度、盖度、水分含量)、间接因子(消除大气干扰和消除土壤背景干扰)和空间因子(纹理特征) 3 大类型。在牧草生物量反演中，这些光谱衍生数据类型间具有较好的互补性，单一的直接因子模型估算结果最差，引入间接因子和空间因子均能提高模型的估算结果，而由直接因子(GNDVI，TCW，NDTI，NDSVI，TCD)、间接因子(SAVI，VARI)和空间因子(Mean_B3，Mean_B6，Hom_Ⅱ，Dis_B5)共同构建的 SGB 模型估算精度最优，R^2 达到了 0.88；RMSE 为 141.00 g·m^{-2}。与 5 种常用的生物量估算模型结果对比，该方法具有明显的优势。较单变量模型，R^2 提高了 42%～60%，RMSE 降低 47%以上，R^2_{cv} 提高了 31%～53%，$RMSE_{cv}$ 降低 29%；较多变量模型，R^2 提高了 29%～42%，RMSE 降低 35%以上，R^2_{cv} 提高了 2%～18%，$RMSE_{cv}$ 降低 2%以上。此外，所提出方法在消除反演模型过饱和方面也具一定成效。综上，利用 Landsat8 数据从反映植被不同理化特征角度构建反演模型实现了牧草生物量的精准估算，对于后期牧草生长状况实时监测以及草地资源可持续利用与管理具有重要的指导意义。研究结果还可以为今后进行大面积区域草地动态监测以及其他农业领域的研究提供参考和借鉴。

关键词 生物量；随机梯度 Boosting 算法；Landsat8 光谱衍生数据

中图分类号：TP79 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2020)01-0239-08

引 言

生物量作为草地生态系统的物质基础，是衡量草地生长状况的主要指标，代表草地初级生产力的基本水平，决定了草地的载畜能力[1]。及时、精确的掌握草地地上生物量的含量、分布及变化情况对于评估草地生态系统、计算草地载畜能力、确保草地生态安全具有重要意义[2]。

与传统实地测量方法不同，利用遥感技术可以快速、准确、无破坏的实现对草地生物量估算。Landsat 系列卫星数据被称为是最有用的遥感数据之一，已被广泛应用于区域尺度牧草生物量估产。研究发现通过对 Landsat 系列数据进行波段计算获取的光谱衍生数据比原始波段在探测生物量方面具有更好的灵敏性。例如红光谱波段对植被叶绿素敏感，近

收稿日期：2018-11-14，**修订日期**：2019-03-19

基金项目：国家自然科学基金项目(41571369)，国家重点研发计划项目(2016YFB0502500)，北京市自然科学基金项目(4162034)，青海省科技计划项目(2016-NK-138)，首都师范大学重大(重点)培育项目资助

作者简介：张爱武，女，1972 年生，首都师范大学三维信息获取与应用教育部重点实验室教授 e-mail：zhangaw98@163.com

*通讯联系人 e-mail：guochao881016@163.com

红外光谱波段对叶片组织敏感，由红光谱波段和近红外光谱波段构建的归一化植被指数(NDVI)可以反映植被的绿度特征[3]；短波红外光谱波段对植被含水量非常敏感，由短波红外构建的归一化红外指数(NDⅡ)可以反映植被水分含量[4]。这些指数能够直观的反映植被某些方面的理化特征(定义为直接因子)，因此在植被生物量反演中得到了广泛的应用。但地面植被信息的遥感获取是一个复杂的过程，还会受到大气、其他地物背景的干扰。因此相关学者推出了一些突显地面植被信息、消除背景干扰的植被指数。例如土壤调节植被指数(SAVI)能够较好的去除土壤背景对于目标信息的影响[5]。此外缨帽变换的第三分量通过影像增强的方法反映地面的土壤水份含量，(这些指数定义为间接因子)。以及纹理特征(定义为空间因子)可以从图像反映植被冠层的空间变化规律和空间相关性。这些间接因子和空间因子从不同的角度反映了地面植被的信息，但由于与直接因子具有较强的共线性，很少被应用于植被生物量反演研究。

基于植被指数的单变量反演模型是目前进行大面积生物量估算的主要方法，常用模型包括线性和非线性模型[6]。当生物量较低时，建立的估算模型是一元线性的，随着生物量的增加，指数模型体现出更好的拟合效果[7]。一些学者尝试通过寻求各种统计方法构建基于多变量植被指数特征的植物生物量估算模型，如高明亮[8]等基于环境卫星遥感数据和同步野外实地采样数据，进行了黄河湿地植被生物量反演研究，结果表明 MLRM(多元线性回归模型) 比 SCRM(一元曲线回归模型) 具有更好的反演精度和预测能力。

随机梯度 Boosting 算法(stochastic gradient boosting，SGB)是一种集成学习方法，在生态建模中有广泛的应用，但是在遥感中应用尚不多见。该算法的优势在于不需要预先筛选特征变量，同时可以适应复杂的非线性关系，且模型具有高度稳健型和可解释性，不容易陷入过拟合[9]。因此，提出基于随机梯度 Boosting 算法(SGB)来构建牧草生物量反演模型。以青海省海晏县为研究区，以 Landsat 8 遥感影像数据为数据源，进行方案的可行性探讨。研究的内容主要包括：(1)归纳总结植被生物量反演相关的 Landsat-光谱衍生数据，并基于它们所反映的植被理化特征及它们间的关联方式构建分类体系；(2)基于随机梯度 Boosting 算法构建多变量非线性牧草生物量反演模型，探讨不同 Landsat-光谱衍生数据类型组合对于模型的影响。以期为牧草生物量遥感监测提供理论依据，为提高牧草生物量的定量反演精度提供参考。

1 实验部分

1.1 研究区概况

研究区(图 1)位于青海省海北藏族自治州海晏县境内，地处 36°53′30″—37°5′30″N，100°47′30″—100°59′10″E；年日照时数 2 980 h，年平均温度 1.7°，年降水量 499 mm，夏秋降水多，春冬降水少，全县牧草草地面积占总面积 49.35%，草种类型多样，是全国草地生态畜牧业试验区。

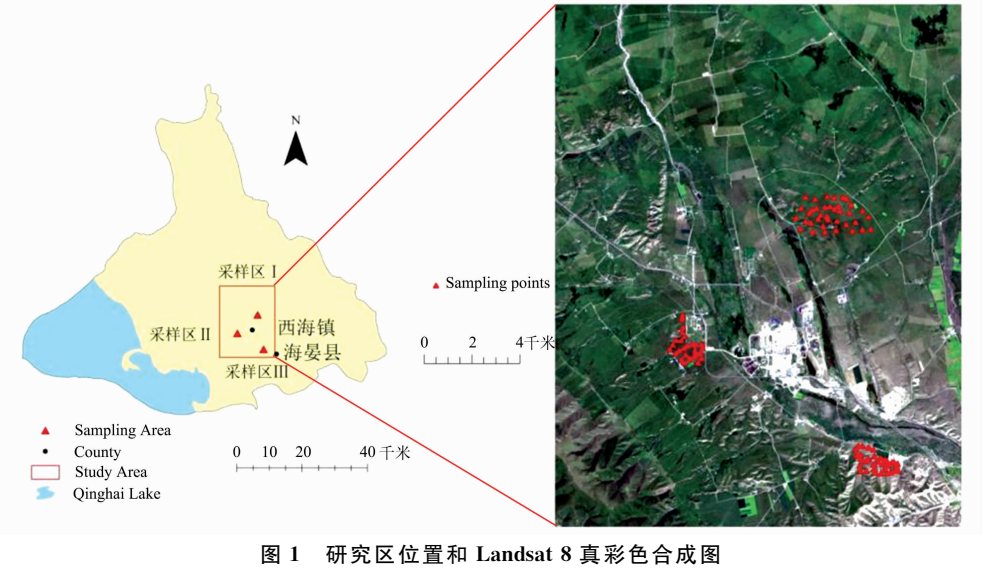

图 1 研究区位置和 Landsat 8 真彩色合成图

R：波段 4；G：波段 3；B：波段 2

Fig. 1 The location of study area and Landsat 8 true color image

R：Band 4；G：Band 3；B：Band 2

1.2 采样点设置与生物量测定

地面鲜重数据采集于 2017 年 8 月 11 日—13 日进行，根据草地类型和生物量等级高、中、低梯度选择了三个采样区(采样区Ⅰ，Ⅱ，Ⅲ)，包括两个春冬草场(Ⅰ，Ⅱ)和一个夏秋草场(Ⅲ)，三个区域分层随机采集 100 个混合样方(图 1)，剔除部分异常值后剩余 97 个采样点，样方尽可能代表整个研究区域的植被生长状况，同时用 GPS 仪测量每个样方中心点经纬度。样方规格为 0.5 m×0.5 m，齐地刈割，挑出石子和动物粪便等不可食部分称取鲜重并记录。

1.3 Landsat OLI 光谱衍生数据提取及分类体系构建

实验所用的遥感数据为美国陆地卫星 Landsat 8 OLI 遥感影像，时间分辨率为 16 d，空间分辨率为 30 m，影像过境

时间 2017 年 8 月 10 日，使用波段包括深蓝波段(0.43～0.45 μm)在内的前 7 个波段，使用 ENVI5.1 对影像进行预处理，经过辐射定标，大气校正后得到反射率数据。

Landsat 8OLI 数据在研究农作物信息提取、叶面积指数反演、生物量估算等方面均取得较好的效果。但植被生长是一种复杂的过程，伴随着多种植被特征状态的变化，如植株高度、冠层叶面积指数、植被颜色、植被水分等，不同的特征可能产生不同的遥感信号，需要将其区别对待。根据不同衍生变量在植被生物量反演过程中所反映植被的理化特征，将常用的 Landsat-衍生变量分为 7 类(表 1)：一是反映植被绿度的绿度指数(NDVI，GNDVI，RVI，II，TCG)，由于红光谱波段对植被叶绿素敏感，近红外光谱波段对植被叶片组织敏感，两者有效结合可精确的刻画植被的绿度特征。二是反映植被衰败程度的黄度指数(NDTI，NDSVI)，该类指数常用于提取植物枯枝落叶层及农作物残余物信息，主要反映植被整体的凋萎程度及作物成熟状况。三是反映植被水分含量的衍生变量，包括水分指数(NDMI，NDII)和缨帽变化中的湿度分量(TCW)，可用于反映植株冠层水分含量和土壤湿度；四是用于反映植被覆盖度的衍生变量，包括 TCA 和 TCD，它们是经过缨帽变化中亮度分量(TCB)和绿度分量(TCG)变换到极坐标系统而获得的指数，TCA 随着植被覆盖度的增大而增大；TCD 随着阴影面积在像元中比例增加而减少，这两个变量可用于反映植被生长密集时的情况。五是用于消除大气影响因子的植被指数(ARVI，EVI，VARI)，通过增加大气修正因子，能够有效减少大气对植被的影响。六是用于消除土壤背景影响的植被指数(SAVI，PVI，MSAVI，OSVAI)，通过增加土壤调节系数，能够有效减少土壤对植被的影响[10]。七是反映植被空间特性的纹理指数，应用最广泛的是由 Haralick 等提出的灰度共生矩阵(GLCM)，主要包括均值、方差、均匀性、对比度、相异性、熵、二阶矩、相关性等 8 个指标(窗口大小为 5×5 像素)。其中类一、类二、类三和类四直接反映了植物的理化特征，定义为直接因子。类五和类六通过消除背景干扰间接的反映植被理化特性，定义为间接因子。而纹理特征则是从空间的角度反映植被的特征，定义为空间因子。

表 1　Landsat-衍生变量分类体系

Table 1　Classification system of Landsat-derived variables

特征类	光谱指数	简称	标签	特征类	光谱指数	简称	标签
绿度指数	归一化差异植被指数	NDVI[11]	Ⅰ	消除土壤背景指数	土壤调节植被指数	SAVI	Ⅱ[11]
	绿色归一化植被指数	GNDVI[11]	Ⅰ		垂直植被指数	PVI	Ⅱ[11]
	缨帽变化绿度分量	TCG[11]	Ⅰ		修改型土壤调整植被指数	MSAVI	Ⅱ[11]
	比值植被指数	RVI[11]	Ⅰ		优化型土壤调整植被指数	OSAVI	Ⅱ[11]
	红外指数	II[11]	Ⅰ	纹理特征	平均值	Mean	Ⅲ[17]
黄度指数	归一化差值耕作指数	NDTI[12]	Ⅰ		方差	Var	Ⅲ[17]
	归一化衰败植被指数	NDSVI[13]	Ⅰ		同质度	Hom	Ⅲ[17]
植被水分变化	归一化差异水体指数	NDWI[11]	Ⅰ		对比度	Con	Ⅲ[17]
	归一化红外指数	NDII[14]	Ⅰ		相异性	Dis	Ⅲ[17]
	缨帽变化湿度分量	TCW[11]	Ⅰ		熵	Ent	Ⅲ[17]
植被盖度	缨帽距	TCD[15]	Ⅰ		二阶矩	Sec	Ⅲ[17]
	缨帽角	TCA[16]	Ⅰ		相关性	Cor	Ⅲ[17]
消除大气影响指数	大气阻抗植被指数	ARVI[11]	Ⅱ				
	增强型植被指数	EVI[11]	Ⅱ				
	可见光大气阻抗植被指数	VARI[11]	Ⅱ				

注：Ⅰ表示属于直接因子，Ⅱ表示间接因子，Ⅲ表示空间因子

Note：Ⅰ represents direct fector；Ⅱ represents indirect factor and Ⅲ represents spector

1.4　随机梯度 Boosting 回归模型(SGB)

随机梯度 Boosting(SGB)是一种可用于分类和回归模型的集成学习器，具有高度稳健性和可解释性。SGB 方法对于异常值、缺失值、非平衡数据集有较好的鲁棒性，参与计算的变量不需要假设先验概率分布，并且在处理非线性关系及变量之间的存在较强自相关模型时有较大的优势。

2001 年，Friedman[18] 提出 Gradient Boosting 算法，该算法将每次迭代的组合分类器在 x 上的值作为损失函数空间在 x 上的负梯度，将组合分类器的系数作为步长，来近似逼近组合分类器的损失函数的最小值。令 $\boldsymbol{X}=[x_1, x_2, \cdots, x_n]^{\mathrm{T}}$，经 M 次迭代后，得到最终的回归树模型

$$F(x) = F_0(x) + v\beta_1 h_1(\boldsymbol{X}) + v\beta_2 h_2(\boldsymbol{X}) + \cdots + v\beta_M h_M(\boldsymbol{X}) \quad (1)$$

其中，$F_0(x)$是用于估计损失函数最小化的常数值；收缩性参数 v 称为“学习率”，决定了每棵树对最终模型的贡献率；β 是模型权重。

2002 年，Friedman[19] 结合 Breiman 的 bagging 思想，在 Gradient boosting 算法基础上引入随机化参数，提出了 SGB 算法，即在每一次迭代过程中，随机抽取训练样本的一部分来拟合分类器。

以决策树为基础分类器的 SGB 算法，可以计算每个变量在减少整体模型总误差平方和的贡献来对其变量重要性进行评价。总误差平方和减少量 $\hat{J}_j^2(T)$表达式为

$$\hat{J}_j^2(T)=\frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N-1}\hat{I}_n^2 I \tag{2}$$

其中，j 为分裂变量，N 为叶子节点的数量，决策节点数量为 $N-1$，在 n 节点内误差平方和减少量为 $\hat{I}_n^2$，M 为决策树数量。

该方法实施后项特征消除来确定生物量预测所需要的 Landsat-光谱衍生数据从而实现变量选择。更准确的说，根据式(2)可以计算各个变量的误差平方和减少量，误差平方和减少量越小，特征变量对模型的贡献越大，逐步消除变量贡献率小的变量实现变量选择[20]。

1.5 模型建立与精度评价

基于统计分析软件 R 的"gbm"包，通过随机梯度 Boosting 变量选择，选择直接因子、直接因子-间接因子、直接因子-空间因子和直接因子-间接因子-空间因子组合中最优特征组合，探讨不同数据类型组合对于估算结果的影响。

为了验证该模型的有效性，设计了 5 种常用模型进行对比分析，包括 1 种一元线性回归模型、2 种非线性回归模型(指数模型和对数模型)，1 种多元线性回归模型(逐步线性回归)和 1 种多元非线性模型(随机森林模型)。采用均方根误差(RMSE)和决定系数(R^2)对模型精度进行评价；并使用十折交叉[21]验证方法对最优模型进行精度验证。十折交叉验证将数据集划分为 10 个子数据集，将每个子集数据分别做一次验证集，其余 9 组子集数据作为训练集，从而避免模型过拟合。

2 结果与讨论

2.1 模型构建

随机梯度 Boosting 方法进行特征选择与其他特征选择方法的不同之处在于该方法的特征选择是嵌入在训练过程中的，是面向于最终模型性能的。也就是说 SGB 算法各个特征对模型的影响是通过每个变量对模型的误差平方和减少量来计算得到的，减少量越大，变量对模型的贡献越大。采用 SGB 算法对由 12 个直接因子、7 个间接因子和 56 个空间因子(7 个波段，每个波段 8 个特征纹理，共 56 个)共构建的 4 个数据集(直接因子、直接因子-间接因子、直接因子-空间因子、直接因子-间接因子-空间因子)进行模型构建。

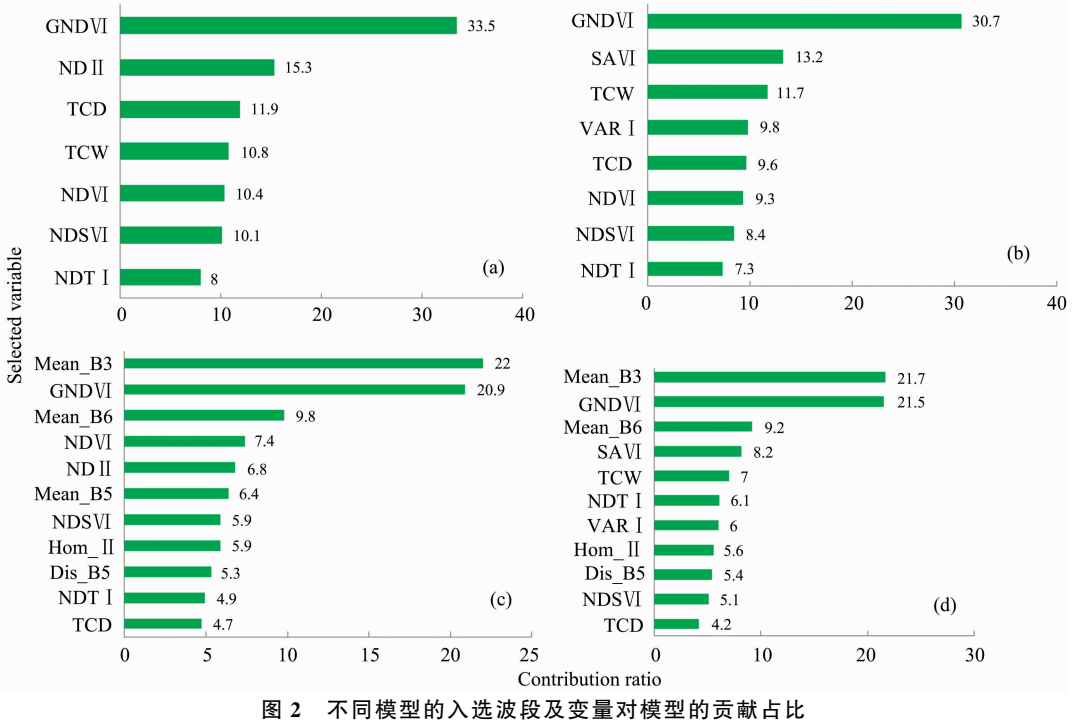

图 2 不同模型的入选波段及变量对模型的贡献占比

(a)：直接因子模型；(b)：直接因子-间接因子模型；

(c)：直接因子-空间因子模型；(d)：直接因子-间接因子-空间因子模型

Fig. 2 The selected bands of different models and the contribution ratio of variables to models

(a)：Direct factor model；(b)：Direct factor-indirect factor model；

(c)：Direct factor-space factor model；(d)：Direct factor-indirect factor-spatial factor model

基于 SGB 对四个数据集进行特征波段选择，选择的特征波段及变量所对应的模型贡献占比如图 2 所示。直接因子模型中，共有 7 个特征变量入选，其中 GNDVI 占比最大，达到 33.5%。说明植被绿度在该模型中起关键作用。同时研究表明与其他绿度指数相比，GNDVI 对于植被叶绿素含量的变化更加敏感。叶绿素反映植被的生长状况，进而反映在生物量方面。其次植被水分(TCW，NDII)和植被盖度(TCD)、植被黄度(NDSVI，NDTI)等因素也对生物量反演具有重要

的意义。直接因子-间接因子模型中，共有 8 个特征变量入选。其中 GNDVI 同样占比最大，达到 30.7%。说明植被绿度和叶绿素在该模型中起关键作用。除了反映水分植被盖度和黄度的指数，还新增加了大气消除指数和土壤消除指数，且均在模型中占有重要的比重。说明在牧草生物量反演中会受到这两个因素的影响。直接因子-空间因子模型中，共有 11 个特征变量入选，有 5 个是纹理特征，且平均值 Mean_B3（第三波段的均值特征）成为占比最大的特征，占到了 22.0%。说明纹理特征在生物量反演模型中具有非常重要的作用。GNDVI 同样占比较高，说明植被绿度在生物量反演中的重要性。直接因子-间接因子-空间因子模型中共有 11 个特征变量入选。其中纹理因子 5 个，占比 41.9%。且 Mean_B3 在所有特征中占比最大 21.7%。直接因子 5 个，占比 43.9%。间接因子 2 个，占比 14.2%。总的来说这些常用数据类型组合从各个方面反映了植被的理化特征，进而反映出生物量。它们之间不仅仅是高相关性，还具有较好的互补性，随机梯度 Boosting 模型可以较好的克服其共线性问题。

表 2 为各个模型选择后的变量与样地生物量的建模结果。如表所示，仅采用直接因子与生物量拟合时精度最低，R^2 为 0.80，RMSE 为 185.85 g·m^{-2}。通过增加间接因子和空间因子均可增加模型的拟合精度。直接因子和空间因子模型的拟合结果表现为 R^2 为 0.83，RMSE 为 158.15 g·m^{-2}；直接因子和间接因子模型的拟合结果为 R^2 为 0.84，RMSE 为 157.63 g·m^{-2}；相较于直接因子模型 R^2 均有所增加，RMSE 均更低。而直接因子、间接因子和空间因子所组合的特征集进行回归建模 R^2 最高，达到了 0.88；RMSE 最低，为 141.00 g·m^{-2}，是拟合生物量的最优模型。总的来说四个模型都能够较好的拟合草原的生物量。拟合模型的各个因子之间是具有兼容性的，通过因子组合可以更好的刻画生物量与这些特征之间的关系。

表 2　模型及精度

Table 2　Model and precision

模型编号	特征组合	入选特征数量	R^2	RMSE/(g·m^{-2})
Ⅰ	直接因子	7	0.80	185.85
Ⅱ	直接因子-间接因子	8	0.83	158.15
Ⅲ	直接因子-空间因子	11	0.84	157.63
Ⅳ	直接因子-间接因子-空间因子	11	0.88	141.00

2.2　模型对比及偏差分析

我们提出了一种多变量、非线性生物量模型，相比于传统的方法，一个比较明显的区别在于该模型更加的复杂化。为了探索本模型与其他模型在普及方面的区别，我们设计了 5 个对比模型，1 个单变量线性模型，2 个单变量非线性模型，1 个多元线性模型（逐步线性回归）和一组多元非线性回归模型（随机森林）进行模型的对比分析。分别采用模型精度和交叉验证精度作为评价指标对不同模型与生物量的估算效果进行评价。此外，由于大量的文献提出过饱和问题是遥感反演中的一个制约因素，我们绘制了 6 种不同模型的残差结果与 NDVI 的关系图，以便能够直观的观察不同模型对于过饱和问题的效果。

表 3　不同模型精度对比

Table 3　Accuracy comparison of different models

模型类型		最优入选变量	模型精度		交叉验证精度	
			R^2	RMSE/(g·m^{-2})	R^2_{cv}	RMSE$_{cv}$/(g·m^{-2})
单变量线性	$Y=a+bx$	GNDVI	0.51	269.10	0.47	281.01
单变量非线性	$Y=a+b\ln x$	Mean_B6	0.52	267.36	0.49	274.05
	$Y=ax^b$	GNDVI	0.62	266.26	0.55	289.76
多元线性回归模型	逐步线性回归	B2，ARVI，NDMI，RVI，TCD，TCA，ENT_B7	0.68	217.07	0.61	252.89
多元非线性回归模型	随机森林	SAVI，VARI，TCW，NDSVI，TCD，Mean_B3，Mean_B6	0.87	141.20	0.70	204.22
	SGB	TCD，NDSVI，Dis_B5，Hom_Ⅱ，VARI，NDTI，TCW，SAVI，Mean_B6，GNDVI，Mean_B3	0.88	141.00	0.72	198.98

6 组不同模型的模型精度和交叉验证精度如表 3 所示，其中单变量线性模型生物量反演模型精度和交叉验证精度均最低。单变量非线性模型次之。研究表明在进行大面积生物量反演时指数模型具有更好的拟合效果[7]。相比于单变量模型，3 组多变量模型具有更好的反演结果。尤其是两组非线性模型具有更好的拟合效果，R^2 达到了 0.85 以上，RMSE 低于 142 g·m^{-2}；R^2_{cv} 达到了 0.70 以上，RMSE$_{cv}$ 低于 205 g·m^{-2}。此外研究结果表明本方法具有最优的回归精度。总的来说在生物量反演模型中多变量、非线性模型具有较大的潜力。

残差反映了模型观测值与估算值之间的偏差。NDVI 是一种使用最为广泛的植被指数，但是研究表明，在生物量较高时会出现过饱和问题。因此采用残差-NDVI 关系图直观的展示不同模型对于过饱和问题的响应。6 组不同模型的残差-NDVI 结果如图 3 所示。总体上讲，6 种模型的残差趋势是一致的，当 NDVI 值小于 0.7 的时候残差较小，当 NDVI 值大于 0.7 时残差突然增大。说明这些模型均受到了过饱和问题的干扰。但图 3(e)和(f)，尤其是本模型无论是总体残差还是当 NDVI 大于 0.7 后的残差均较小，说明本方法是可行的，能够在一定程度上消除过饱和的影响。

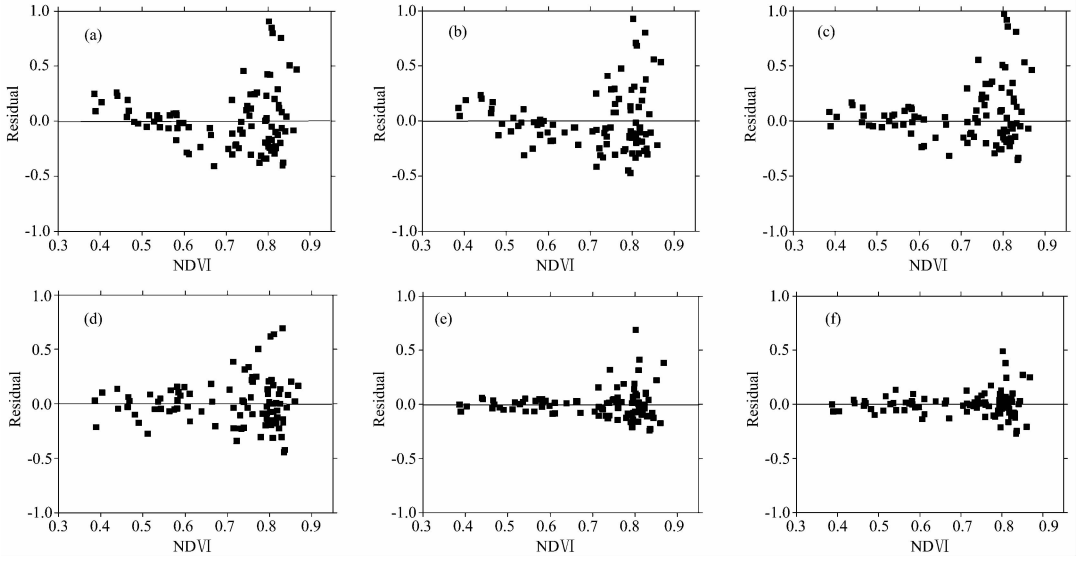

图 3 残差结果与 NDVI 的关系图

(a)：单变量线性模型；(b)：单变量对数模型；(c)：单变量指数模型；
(d)：逐步线性回归模型；(e)：随机森林模型；(f)：SGB 模型

Fig. 3 The relationship between residual and NDVI

(a)：Univariate linear model；(b)：Univariate logarithmic model；(c)：Univariate exponential model；
(d)：Stepwise linear regression model；(e)：Random forest model；(f)：Stochastic gradient boosting model

2.3 牧草生物量反演及制图

基于上述分析，构建的牧草生物量反演模型较传统的方法具有明显优势，因此将该方法应用于整个研究区生物量反演制图。通过 K-Means 方法将研究区分为非植被(城区、道路和水域)和植被两类，非植被在制图中予以剔除。结果如图 4 所示，可以看出研究区牧草生物量分布具有明显的空间差异性。远离城区的牧草生物量较高，而城区周边的牧草生物量明显较低，可能是由于城区周围多为夏季牧场，牛羊放牧制约了牧草生物量的累积，此外旅游开发以及人为活动也会在一定程度影响牧草的生长。

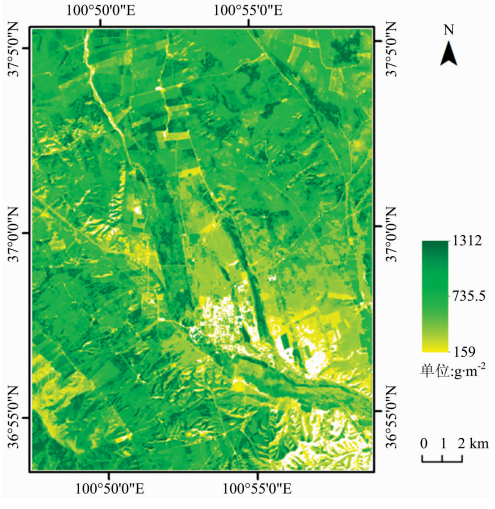

图 4 研究区牧草生物量估算结果图

Fig. 4 Estimation results of biomass in study area

3 结 论

采用 Landsat8 遥感影像结合地面实测数据进行牧草生物量反演研究。首先通过 Landsat8 光谱衍生数据所反映的植被理化特征及它们间的关联方式，构建了不同光谱衍生数据的分类体系；并在此基础上提出了一种基于随机梯度 Boosting 算法的多变量非线性生物量估算模型，探讨不同光谱衍生数据分类组合对于估算结果的影响。以青海省海晏县为研究区进行方案可行性研究。结论如下：

(1)共收集了 27 个与生物量相关的 Landsat8 光谱衍生数据，根据它们所反映的植被理化特征，可以划分为 7 个小类，它们分别反映了植被的绿度、黄度、水分、植被盖度、纹理特征、消除大气干扰和消除土壤背景干扰。根据它们与植被理化特征的关联方式，7 个小类可以合并为 3 个大类：直接因子(绿度指数、黄度指数、水分指数、植被盖度)、间接因子(消除大气干扰指数和消除土壤背景干扰指数)和空间因子(纹理特征)。

(2)基于随机梯度 Boosting 算法探讨了不同光谱衍生数据类型组合对于估算结果的影响，结果表明在生物量估算模型中直接因子、间接因子和空间因子具有互补性。基于直接因子-间接因子-空间因子构建的估算模型优于其他组合模型，R^2 达到了 0.88；RMSE 为 141.00 g·m^{-2}。

(3)通过与 5 种生物量估算模型结果对比表明，本文提出的模型具有更好的估算结果。较单变量模型，R^2 提高了 42%～60%，RMSE 降低了 47%以上，R^2_{cv} 提高了 31%～53%，RMSE$_{cv}$降低 29%；较多变量模型 R^2 提高了 29%～42%，RMSE 降低 35%以上，R^2_{cv}提高了 2%～18%，RMSE$_{cv}$降低 2%以上。总的来说非线性模型、多变量模型具有更好的估算结果，是今后研究的重点。此外，提出的模型在消除过饱和方面也具有明显优势。

综上，提出了一种利用 Landsat 数据进行牧草生物量估算的有效方法，一定程度上满足了畜牧业可持续发展的需求，并且该方法可以扩展到其他植被类型和更多生物参量的估算研究。为今后进行大面积区域草地动态监测以及其他农业领域的研究提供了参考和借鉴。

References

[1] ZHAO Ming-wei, YUE Tian-xiang, SUN Xiao-fang, et al(赵明伟，岳天祥，孙晓芳，等). Acta Ecologica Sinica(生态学报), 2014, 34(17): 4891.

[2] YAN Rui-rui, TANG Huan, DING Lei, et al(闫瑞瑞，唐　欢，丁　蕾，等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2017, 33(15): 210.

[3] Gouveia C M, Trigo R M, Beguería S, et al. Global and Planetary Change, 2017, 151: 15.

[4] Wilson N R, Norman L M. International Journal of Remote Sensing, 2018, 39(10): 3243.

[5] Candiago S, Remondino F, De Giglio M, et al. Remote Sensing, 2015, 7(4): 4026.

[6] QIAN Yu-rong, YANG Feng, LI Jian-long, et al(钱育蓉，杨　峰，李建龙，等). Pratacultural Science(草业科学), 2013, 30(9): 1330.

[7] ZHANG Yan-nan, NIU Jian-ming, ZHANG Qing, et al(张艳楠，牛建明，张　庆，等). Acta Prataculturae Sinica(草业学报), 2012, 21(1): 229.

[8] GAO Ming-liang, ZHAO Wen-ji, GONG Zhao-ning, et al(高明亮，赵文吉，宫兆宁，等). Acta Ecologica Sinica(生态学报), 2013, 33(2): 542.

[9] Lawrence R, Bunn A, Powell S, et al. Remote Sensing of Environment, 2004, 90(3): 331.

[10] LI Wei, MU Meng, CHEN Guan-bin, et al(李　微，牟　蒙，陈官滨，等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(5): 1418.

[11] Xue J, Su B. Journal of Sensors, 2017, 2017: doi.org/10.1155/2017/1353691.

[12] Van Deventer A P, Ward A D, Gowda P H, et al. Photogrammetric Engineering and Remote Sensing, 1997, 63: 87.

[13] Qi J, Marsett R, Heilman P, et al. Transactions American Geophysical Union, 2002, 83(51): 601.

[14] Sriwongsitanon N, Gao H, Savenije H H G, et al. Hydrology and Earth System Sciences, 2016, 20(8): 8419.

[15] Duane M V, Cohen W B, Campbell J L, et al. Forest Science, 2010, 56(4): 405.

[16] Powell S L, Cohen W B, Healey S P, et al. Remote Sensing of Environment, 2010, 114(5): 1053.

[17] Haralick R M, Shanmugam K. IEEE Transactions on Systems, Man, and Cybernetics, 1973, (6): 610.

[18] Friedman J H. Annals of Statistics, 2001, 29: 1189.

[19] Friedman J H. Computational Statistics & Data Analysis, 2002, 38(4): 367.

[20] Dube T, Mutanga O, Elhadi A, et al. Sensors, 2014, 14(8): 15348.

[21] MENG Shi-li, PANG Yong, ZHANG Zhong-jun, et al(蒙诗栎，庞　勇，张钟军，等). Journal of Remote Sensing(遥感学报), 2017, 21(5): 812.

Grass Biomass Inversion Based on Landsat 8 Spectral Derived Data Classification System

ZHANG Ai-wu[1,2], ZHANG Shuai[1,2], GUO Chao-fan[1,2*], LIU Lu-lu[1,2], HU Shao-xing[3], CHAI Sha-tuo[4]

1. Key Laboratory of 3D Information Acquisition and Application, Ministry of Education, Capital Normal University, Beijing 100048, China
2. Engineering Research Center, Ministry of Education, Capital Information Technology, Capital Normal University, Beijing 100048, China
3. School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China
4. Qinghai University, College of Animal Husbandry and Veterinary Medicine (Qinghai Academy of Animal Science and Veterinary Medicine), Xining 810016, China

Abstract Estimation of forage biomass is of great significance for the rational use of grassland resources and monitoring of livestock load balance, and it is a key indicator for evaluating the sustainable development of grassland ecosystems and grassland resources. The rapid and non-destructive study of large-area vegetation biomass estimation based on Landsat remote sensing technology has been widely used. Most of the current researches are based on single variable or several commonly used vegetation indices to construct inversion models. These indices often cannot reflect the physical and chemical characteristics of vegetation in-many aspects. In this paper, the classification systems of different Landsat8-derived data were constructed by their corresponding physicochemical characteristics of vegetation andintersectional pattern with plants. A multivariable nonlinear biomass estimation model based on stochastic gradient boosting algorithm was proposed and the model estimation results were discussed with different combinations of derived data categories. The program feasibility study was carried out with Haiyan County in Qinghai Province as the study area. The results showed that the Landsat8-derived data reflected the physical and chemical characteristics of vegetation mainly from the aspects of vegetation greenness, yellowness, coverage, moisture content, texture characteristics and elimination of atmospheric disturbance and soil background interference(7 subcategories). On the other hand, these data can also be summarized into three categories: direct factors (greenness, yellowness, coverage, moisture content), indirect factors (eliminating atmospheric interference and eliminating soil background interference), and spatial factors (texture characteristics). The derived data categories have obvious complementarity. The direct factor (GNDVI, TCW, NDTI, NDSVI, TCD)-indirect factor (SAVI, VARI)-space factor(Mean_B3, Mean_B6, Hom_Ⅱ, Dis_B5) model had the best estimation accuracy, and R^2 reached 0.88; the RMSE was 141.00 g·m^{-2}, however the single direct factor model estimates result was the worst. Compared with the results of six typical biomass estimation models, the proposed method had obvious advantages. Compared with the univariate models, R^2 increased by 42%～60%, RMSE decreased by more than 47%, R^2_{cv} increased by 31%～53%, and RMSE$_{cv}$ decreased by more than 29%; Compared with the multivariate models, R^2 increased by 29%～42%, RMSE decreased by more than 35%; and R^2_{cv} increased by 2%～18%, RMSEcv decreased by more than 2%. In addition, the proposed model also had some effect in eliminatingoversaturation problem. In summary, this paper uses Landsat8 data to construct an inversion model from the perspective of reflecting different physical and chemical characteristics of vegetation to achieve accurate estimation of forage biomass, which has important guiding significance for the real-time monitoring of pastage growth and the sustainable use and management of grassland resources. The research results can also provide reference and reference for future large-area regional grassland dynamic monitoring and other agricultural research.

Keywords Biomass; Stochastic gradient boosting algorithm; Landsat-derived data

(Received Nov. 14, 2018; accepted Mar. 19, 2019)

* Corresponding author