

## 基于高光谱成像的牧草粗蛋白含量检测研究

高睿, 李泽东, 马铮, 孔庆明, Muhammad Rizwan, 苏中滨\*

东北农业大学电气与信息学院, 黑龙江 哈尔滨 150030

**摘要** 粗蛋白(CP)是评价牧草营养价值和品质参数的关键指标。快速、准确地对牧草中粗蛋白含量进行评估在畜牧业生产研究中具有重要意义。为确定牧草粗蛋白含量的高光谱特征波段及最优检测模型, 研究分别于2017年5月至9月间在黑龙江省杜尔伯特自治区的人工牧草场(羊草)内每月随机选取35个样本, 5个月共采集175个样本。采样时在样本点处放置1 m×1 m的样方, 将样方内所有牧草全部齐地面收割采集后称重并冷藏保存。将样本带回实验室后, 立即进行牧草叶片高光谱图像采集, 同时采用凯氏定氮法对采集的牧草样本进行粗蛋白化学值测定, 以此建立牧草粗蛋白含量高光谱数据集。研究首先通过 Savitzky-Golay 卷积平滑(SG)、多元散射校正(MSC)、变量标准化(SNV)、一阶导数(1-Der)和直接正交信号校正(DOSC)方法5种预处理方法对高光谱数据进行处理后分别建立偏最小二乘回归(PLSR)检测模型, 从中确定最优预处理方法。利用最优预处理结果, 分别采用连续投影算法(SPA)和随机蛙跳算法(RF)进行牧草粗蛋白含量的特征波段选择, 并利用选择结果分别进一步建立 PLSR 模型, 以此确定适合粗蛋白含量的特征波段选择方法, 确定最优高光谱检测模型。结果表明, 在五种高光谱预处理方法中, 基于 SNV 方法预处理后所建立的高光谱 PLSR 模型表现最优( $R^2-P=0.929$ ,  $RMSE-P=6.344 \text{ mg} \cdot \text{g}^{-1}$ ,  $RPD=4.204$ )。利用连续投影算法筛选的粗蛋白含量特征波长为30个, 分布于530~700和940~1000 nm 范围内。经随机蛙跳算法确定的粗蛋白含量特征波段为6个, 分别为826.544, 827.285, 828.766, 971.012, 972.494和973.235 nm。因此, 该研究中牧草粗蛋白含量最优高光谱检测模型为 SNV-RF-PLSR( $R^2-P=0.933$ ,  $RMSE-P=6.034 \text{ mg} \cdot \text{g}^{-1}$ ,  $RPD=4.322$ ), 模型精度较高。该研究结果为牧草粗蛋白含量的高光谱检测提供了最优模型和理论基础, 同时为指导草业生产开拓了新的技术思路。

**关键词** 牧草; 粗蛋白; 高光谱成像; 连续投影算法; 随机蛙跳算法

**中图分类号**: S431.14 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2019)10-3245-06

### 引言

牧草作为畜牧生产中重要的饲料原料, 其自身营养价值和品质参数受到多方面关注。粗蛋白(crude protein, CP)作为牧草最重要品质参数之一, 是表述牧草品质和营养价值的关键指标。在畜牧生产中, 牧草中的粗蛋白含量极大地影响了反刍动物的生产性能, 也进一步决定了动物生产的畜产品品质<sup>[1-2]</sup>。以往对该类指标的检测方法常采用实验室化学检测方法, 不仅耗时, 实验检测后的残余试剂也会对环境造成长期污染。因此, 寻找一种快速、准确且环保的牧草粗蛋白含量检测方法对于畜牧业及草地资源可持续性发展具有

重要意义。

随着光谱技术的不断发展, 研究人员开始应用各类传感器获取植物叶片和冠层光谱数据, 以期实现对植被生理参数的快速无损检测。其中高光谱成像技术可提供图像内任意像素的连续光谱数据, 应用时可有效减少光谱噪声影响, 具有准确、无损等一系列优势, 因此被广泛引用于植物生理参数的检测研究中。张亚坤等<sup>[3]</sup>在获取大豆冠层高光谱图像后, 通过分数阶微分算法预测了大豆冠层氮素含量, 最优预测模型精度达到了0.800。张爱武等<sup>[4]</sup>采用航空飞艇搭载高光谱成像设备获取图像后, 采用光谱衍生变换的方法, 用不同建模方法构建了CP含量的反演模型, 最优模型精度达到了0.918。Vigneau<sup>[5]</sup>等分别获取了三个尺度下小麦植株的高光

收稿日期: 2019-05-24, 修订日期: 2019-08-08

基金项目: 国家重点研发计划(2016YFD0200701)资助

作者简介: 高睿, 1989年生, 东北农业大学电气与信息学院博士研究生 e-mail: 415730327@qq.com

Muhammad Rizwan, 巴基斯坦人, 东北农业大学电气与信息学院博士研究生

\* 通讯联系人 e-mail: suzb001@163.com

谱图像数据及对应的小麦氮含量数据,并对获取的高光谱数据进行反射率校正和光谱数据提纯,在此基础上采用偏最小二乘回归(partial least squares regression, PLSR)建立了小麦氮含量的预测模型,模型决定系数均在 0.87 以上。在国内外研究中,高光谱成像技术通常应用于小麦<sup>[6]</sup>、水稻<sup>[7]</sup>、大豆<sup>[8]</sup>等大宗农作物的生理参数检测,同时研究热点多集中在氮素<sup>[9]</sup>、叶绿素<sup>[10]</sup>和生物量<sup>[11]</sup>等领域,而在叶片尺度针对牧草品质和营养参数的研究还鲜有报道。实际上,与采用高光谱技术对玉米,小麦,水稻等作物的研究不同,对于大宗农作物的根本研究目标是解决植物籽实问题,而在草业科学中动物是直接食用牧草的茎叶部分进行利用的,因此对于牧草叶片粗蛋白含量的快速检测也就具有更直接的研究意义。本研究以全生长期内的牧草为研究对象,获取了各月份牧草叶片的高光谱信息以及对应的粗蛋白含量数据,选取了最优的高光谱预处理方法及特征波长选择方法,进一步建立了牧草粗蛋白含量的高光谱检测模型。目的是为牧草粗蛋白含量的快速检测提供最优模型和理论基础,同时为指导草业生产开拓了新的技术思路。

## 1 研究区概况及数据源

### 1.1 研究区概况

本次研究所选区域位于黑龙江省大庆市杜尔伯特自治区,该区域属于大陆性季风气候,地处松嫩平原,采样中心位置为 46°46′35.89″N, 125°46′24.66″E。当地 5 月到 9 月的平均温度分别为 16.47, 21.43, 24.55, 21.82 和 15.68 °C, 平均降水量为 32.7, 150.6, 263.3, 61.1 和 93.4 mm。研究区域内的牧草主要品种为人工种植的羊草。为满足牧草光谱数据源的统一性,本研究过程中均采用羊草作为数据样本。

### 1.2 数据源

本研究于 2017 年 5 月 17 日、6 月 20 日、7 月 18 日、8 月 18 日、9 月 17 日分别每次在研究区域内的人工牧草场内随机选取 35 个样本,在选择样本点处放置 1 m×1 m 的样方。每两个样本点间距离需要至少相差 500 m,目的是尽量扩大样本范围。但为了尽量贴合牧草生长周期,每月采集样本时,均在同一片草场内采集,并尽量采集相同位置的牧草样本。将样方内所有牧草全部齐地面收割采集后称重并冷藏保存。

将样本带回实验室后,立即进行叶片高光谱数据采集。高光谱设备采用美国 Head Wall 公司生产的高光谱成像系统进行样本数据采集。高光谱传感器成像方式为线阵推扫,光谱范围 400~1 000 nm,共 812 个波段。图像采集时,设置曝光时间为 30 ms,移动平台的移动速度为 3.0 mm·s<sup>-1</sup>,每个样本中取 3 片牧草叶片放置于载物台上,镜头垂直向下,距移动平台 45 cm 进行高光谱图像测定。另一部分样本的粗蛋白化学含量测定在东北农业大学动物科学技术学院进行,测定方法采用国标《GB/T 6432—2018》凯氏定氮法测定。

## 1 实验部分

### 1.1 牧草粗蛋白含量数据处理

在对采集的 175 个牧草样本进行测定后,通过三倍标准差法对牧草粗蛋白含量样本进行剔除,得到粗蛋白含量样本 163 个。通过 Kennard-Stone 算法对牧草样本进行 3:1 比例分类,最终得到粗蛋白的建模集样本 122 个,预测集样本 41 个。经统计,牧草粗蛋白含量的建模集数据范围为 33.73~156.41 mg·g<sup>-1</sup>,平均值 93.88 mg·g<sup>-1</sup>,标准偏差 18.65 mg·g<sup>-1</sup>,具有较好的数据范围和代表性。

### 1.2 高光谱预处理方法

高光谱数据在成像过程中会受到如信号噪声、散射光、背景环境等因素的干扰,为了降低以上干扰信息的影响,需要对高光谱图像进行牧草叶片光谱提取和光谱预处理,去除干扰信息的影响。本文中主要采用了 Savitzky-Golay 卷积平滑(SG)、多元散射校正(multiplicative scatter correction, MSC)、变量标准化(standardization normal variables, SNV)、直接正交信号校正(direct orthogonal signal correction, DOSC)和一阶导数法(derivative, 1-der)。通过 5 种预处理方法对高光谱信息进行处理后,以光谱数据为输入变量,牧草中粗蛋白含量为输出变量,利用偏最小二乘回归建模,构建模型后从中选择最优预处理方法。

### 1.3 特征变量选择方法

高光谱数据的信息量通常较大,包含的光谱波段较多,存在许多的冗余和多重共线性信息,为提高模型精度,同时降低建模时间和模型复杂度,本研究在确定牧草叶片高光谱数据的最优预处理方法后,利用最优光谱预处理结果,分别采用连续投影法(successive projections algorithm, SPA)、随机蛙跳算法(random frog, RF)对高光谱数据进行特征波段选择。SPA 是一种前向选择的变量提取方法,依据建模集内部交叉验证均方根误差(RMSECV)的最小值,以最终确定选择的变量<sup>[12-13]</sup>,而 RF 算法是依据启发群体优化而提出的解决组合优化问题的算法,其通过计算变量被选中的概率确定变量的重要性<sup>[14-15]</sup>。利用选择后的特征波段数据为输入变量,牧草中粗蛋白含量为输出变量,均采用 PLSR 进行建模,从中选择较优的牧草叶片粗蛋白含量模型特征波段选择方法。

### 1.4 算法实现与模型验证

本研究中运用 MATLAB2017b 软件编程进行异常样本剔除,样本分类,光谱预处理,特征波段选择及模型建立的算法实现。采用建模决定系数( $R^2-C$ )、预测决定系数( $R^2-P$ )、建模标准差(RMSE-C)、预测标准差(RMSE-P)和相对预测偏差(RPD)分别对模型的建模集和预测集进行检验。

## 2 结果与讨论

### 2.1 全生长期不同月份牧草叶片光谱曲线变化

本研究中牧草生长期各月份的平均原始光谱曲线如图 1 所示,在 460~500 nm 的蓝光区域,不同月份光谱曲线分散明显;在 530~570 nm 的绿光波峰区域,7 月和 8 月的光谱曲线与 5 月、6 月、9 月差异较大,并且这种差异延续到了在 620~680 nm 的红光区域。其可能原因是,在 7 月、8 月与 5 月、6 月相比,牧草冠层叶绿素含量较高,因此光谱曲线

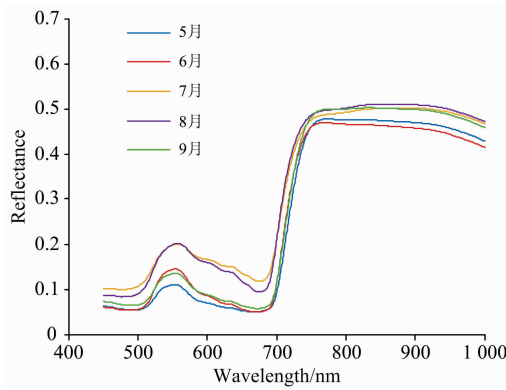


图 1 不同月份牧草叶片平均光谱曲线  
Fig. 1 Average spectrum of pasture leaves in different months

中的绿峰较为明显。在 800 nm 以后的近红外区域, 5 月、6 月的光谱曲线与 7 月、8 月、9 月的差异逐渐增加。此处近红外反射平台光谱反射率主要受叶片的内部结构(叶片间隙与细胞厚度)的影响较大, 当细胞层数越多时, 该区域的光谱反射率越高, 此外叶片细胞的形状、成分的差异也会导致该区域光谱反射率的升高, 因此干物质含量较高的 7 月, 8 月, 9 月牧草光谱反射率更高。

2.2 最优光谱预处理方法的选择

不同光谱预处理方法的处理结果图 2 所示, 由图可知, 经由 SG 平滑方法处理后的光谱曲线与原始光谱无明显差异; 而经 MSC 和 SNV 方法预处理后, 原始分散的光谱曲线更为集中, 光谱曲线中的微小变化有所放大; 经 1-Der 方法处理后, 有效放大了光谱曲线中的细微变化, 同时也分离了部分重叠的光谱信号; DOSC 方法使光谱曲线分散趋势更为明显, 放大了数据间的差异。

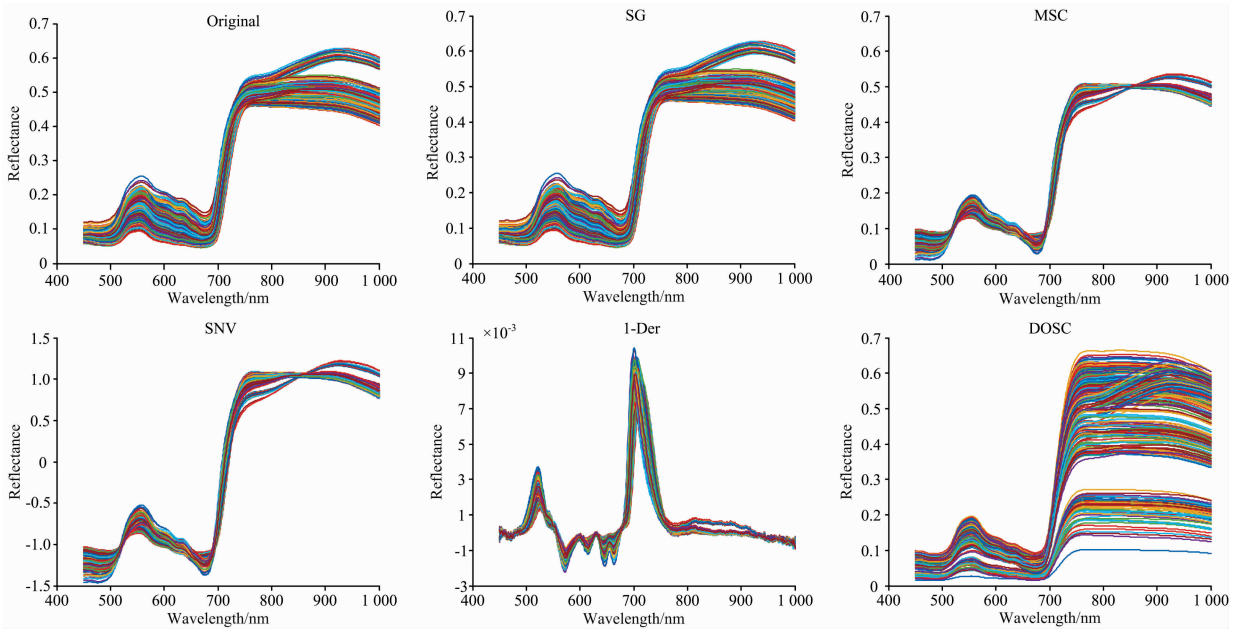


图 2 不同光谱预处理方法的处理结果  
Fig. 2 Results of spectrum with different preprocessing

根据不同高光谱预处理方法处理后所建立的牧草粗蛋白含量 PLSR 模型结果如表 1 所示。不同光谱预处理方法建立的粗蛋白含量检测模型的  $R^2-C$  和  $R^2-P$  均在 0.9 以上, 均可用于牧草粗蛋白含量的检测。其中经 SG 和 1-der 处理后建立的模型性能不及原始光谱建模。而经 MSC, SNV 和 DOSC 方法处理的光谱数据建立的模型较原始光谱模型得到了有效提高,  $R^2-P$  分别提高了 0.012, 0.015, 0.006, RMSE-P 分别降低了 0.172, 0.247, 0.106, RPD 分别提高了 0.121, 0.167, 0.080。在几种提高模型性能的预处理方法中, SNV 方法将  $R^2-P$ , RMSE-P 和 RPD 值分别优化了 1.64%, 3.75%, 1.98%, 为本研究中光谱最优预处理方法, 说明 SNV 预处理方法可以有效剔除原始光谱中与粗蛋白含量的无关信息, 提升模型的整体性能。

表 1 不同光谱预处理方法建立的牧草粗蛋白含量 PLSR 检测模型

Table 1 Modeling results of crude protein content with different preprocessing

预处理方法	主成分数 (LVs)	建模集		预测集		
		$R^2-C$	RMSE-C	$R^2-P$	RMSE-P	RPD
Original	18	0.908	6.523	0.914	6.591	4.037
SG	17	0.881	6.867	0.913	6.698	3.926
MSC	16	0.915	6.456	0.926	6.419	4.158
SNV	16	0.918	6.283	0.929	6.344	4.204
1-der	10	0.903	6.441	0.914	6.647	3.952
DOSC	2	0.917	6.429	0.920	6.485	4.117

2.3 基于 SPA 算法的特征波段选择

本研究中设置 SPA 算法的最大变量数为 50 个, 采用

SPA 进行牧草粗蛋白含量特征波长选择时,图 3 中显示了模型 RMSE 随特征变量数的变化情况,可以看出 RMSE 随着变量数的增加快速下降,当 RMSE 值为 6.506 时,此时运算结果达到最优,共筛选出 30 个与牧草粗

蛋白含量相关的特征变量,全部建模变量的 4.24%。采用 SPA 算法筛选出的与氮含量相关的特征波长个数较多,特征波长分布于 530~700 和 940~1 000 nm 范围内。

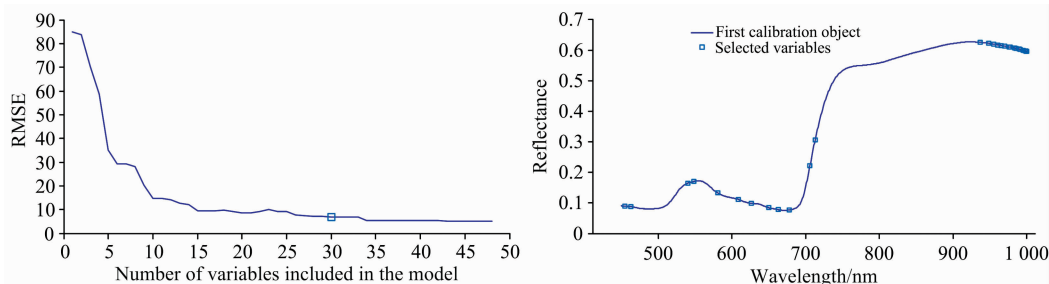


图 3 基于 SPA 算法的粗蛋白含量特征变量选择过程与结果

Fig. 3 Process and results of crude protein characteristic variables based on SPA algorithm

#### 2.4 基于 RF 算法的特征波段选择

本研究中 RF 算法运行时所设参数分别为:初始设定蛙群个数为 5,迭代次数 1 000,  $CA_{MAX}$  为 10,采用多次随机计算取平均值方法进行,运行次数设定为 20。如图 4 所示,在 RF 算法对牧草粗蛋白含量的特征波段进行计算时,当特征

变量数为 5 时, RMSECV 为最小值 6.334。此时对应的选择概率为 0.32(绿色横线位置),将选择概率高于 0.32 的波长作为特征变量选择结果。采用 RF 算法筛选出与牧草粗蛋白含量相关的特征波长共有 6 个,分别为 826.544, 827.285, 828.766, 971.012, 972.494 和 973.235 nm。

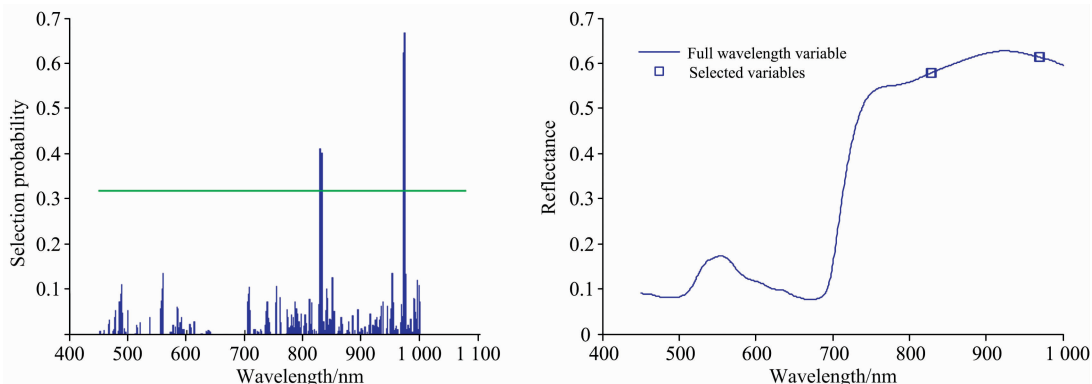


图 4 基于 RF 算法的粗蛋白含量特征变量选择过程与结果

Fig. 4 Process and results of crude protein characteristic variables based on RF algorithm

#### 2.5 基于不同光谱特征变量选择方法的检测结果

根据不同特征变量选择方法建立的牧草粗蛋白含量 PLSR 模型结果如表 2 所示,相比于经过 SNV 方法处理的全变量建模,采用 SPA 算法筛选后建立的模型性能不及全波段建模, RMSE-P 增加了 0.230, RPD 降低了 0.039,可能原因是 SPA 算法在筛选过程中剔除了与牧草粗蛋白含量相关的变量,使模型性能有所降低。而采用 RF 算法建立模型的  $R^2-P$ , RMSE-P 和 RPD 均有所优化, RMSE-P 降低了

0.279, RPD 增加了 0.118。RF 算法将变量数从 812 个减少到 6 个,建模变量减少了 99.2%,使模型复杂度得到了有效降低,该算法使模型  $R^2-P$ , RMSEP 和 RPD 分别优化了 0.4%, 4.62% 和 2.8%。经过 RF 算法筛选出的特征变量建立模型的各项指标均较优,不仅提高了模型精度,同时降低牧草粗蛋白检测模型的复杂度,减少了建模时间,最优模型预测结果如图 5 所示。

表 2 不同特征变量选择方法建立的牧草粗蛋白含量 PLSR 检测模型

Table 2 Modeling results of CP with different characteristic variable selection algorithms

特征变量选择方法	特征变量个数	主成分数 (LVs)	建模集			预测集		
			$R^2-C$	RMSE-C	$R^2-P$	RMSE-P	RPD	
SNV	812	16	0.918	6.283	0.929	6.344	4.204	
SPA	30	1	0.906	6.434	0.928	6.574	4.163	
RF	6	3	0.924	6.183	0.933	6.034	4.322	

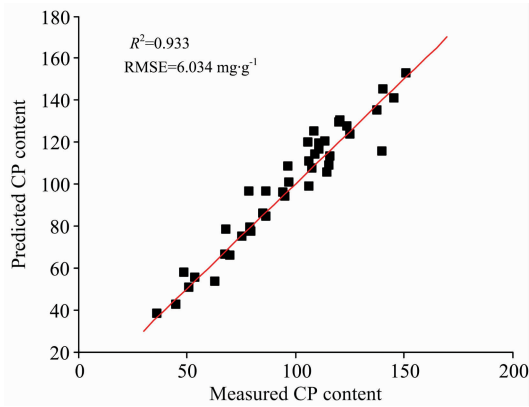


图 5 牧草粗蛋白含量最优模型预测结果

Fig. 5 Optimal prediction result of crude protein content

### 3 结 论

探索了 5 种高光谱数据预处理方法对牧草粗蛋白含量检测的效果, 同时对比了连续投影算法和随机蛙跳算法对牧草粗蛋白含量特征波段选择的效果, 得出以下结论: (1) 在五种高光谱预处理方法中, 针对牧草粗蛋白含量的最优高光谱预处理方法为 SNV。基于 SNV 方法预处理后所建立的高光谱 PLSR 模型表现最优 ( $R^2-P = 0.929$ ,  $RMSE-P = 6.344 \text{ mg} \cdot \text{g}^{-1}$ ,  $RPD = 4.204$ )。 (2) 相比于连续投影算法, 随机蛙跳算法更适用于牧草粗蛋白含量的特征波段选择。经随机蛙跳算法确定的粗蛋白含量特征波段为 826.544, 827.285, 828.766, 971.012, 972.494 和 973.235 nm。 (3) 本研究中确定的牧草最优高光谱检测模型为 SNV-RF-PLSR ( $R^2-P = 0.933$ ,  $RMSE-P = 6.034 \text{ mg} \cdot \text{g}^{-1}$ ,  $RPD = 4.322$ ), 模型精度较高。综上所述, 本研究通过采集牧草全生长期粗蛋白样本, 确定了牧草叶片粗蛋白含量的最优高光谱检测模型, 该研究结果为指导草业生产开拓了新的技术思路, 在精准农业领域应用前景广泛。

### References

- [ 1 ] Huang Y, Shelby S, Wang X, et al. *Journal of Animal Science*, 2018, 96: 29.
- [ 2 ] WANG Bing, JIANG Lin-shu, LIU Jian-xin(王炳, 蒋林树, 刘建新). *Chinese Journal of Animal Nutrition(动物营养学报)*, 2017, 29(6): 1921.
- [ 3 ] ZHANG Ya-kun, LUO Bin, PAN Da-yu, et al(张亚坤, 罗斌, 潘大宇, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2018, 38(10): 3221.
- [ 4 ] ZHANG Ai-wu, YAN Wen-yan, GUO Fan-chao(张爱武, 鄢文艳, 郭超凡). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2018, 34(3): 188.
- [ 5 ] Vigneau N, Ecartot M, Rabatel G, et al. *Field Crops Research*, 2011, 122(1): 25.
- [ 6 ] Bauriegel E, Giebel A, Geyer M, et al. *Computers and Electronics in Agriculture*, 2011, 75(2): 304.
- [ 7 ] Gnyp M L, Miao Y, Yuan F, et al. *Field Crops Research*, 2014, 155: 42.
- [ 8 ] Ma Y, Huang M, Yang B, et al. *Computers and Electronics in Agriculture*, 2014, 106: 102.
- [ 9 ] Inoue Y, Sakaiya E, Zhu Y, et al. *Remote Sensing of Environment*, 2012, 126: 210.
- [ 10 ] Darvishzadeh R, Skidmore A, Schlerf M, et al. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2008, 63(4): 409.
- [ 11 ] John R, Chen J, Giannico V, et al. *Remote Sensing of Environment*, 2018, 213: 34.
- [ 12 ] Cheng J H, Sun D W, Pu H. *Food Chemistry*, 2016, 197: 855.
- [ 13 ] Liu K, Chen X, Li L, et al. *Analytica Chimica Acta*, 2015, 858: 16.
- [ 14 ] Hu M H, Dong Q L, Liu B L, et al. *Postharvest Biology and Technology*, 2015, 106: 1.
- [ 15 ] Zhang C, Ye H, Liu F, et al. *Sensors*, 2016, 16(2): 244.

# Research on Crude Protein of Pasture Based on Hyperspectral Imaging

GAO Rui, LI Ze-dong, MA Zheng, KONG Qing-ming, Muhammad Rizwan, SU Zhong-bin\*

Academy of Electric and Information, Northeast Agricultural University, Harbin 150030, China

**Abstract** Crude protein (CP) is the key parameter for evaluating nutritive value and quality of pasture. It has a great significance for evaluating crude protein content of pasture quickly and accurately in animal husbandry. For confirming the hyperspectral characteristic bands and optimal detection model of crude protein content in pasture, we randomly selected thirty-five sample plots each month from May to September, 2017 in Dorbet, Heilongjiang Province, one hundred and seventy-five samples for all. A 1 m×1 m quadrangle was placed at the sample point during sampling, and all the aboveground pastures in the quadrangle were collected, weighed and stored in cold storage. After carrying the samples the laboratory, we collected the hyperspectral information immediately and determined the chemical values of crude protein by Kjeldahl determination, establishing the hyperspectral dataset of crude protein content. We used five pre-processing methods including SG, MSC, SNV, 1-Der, DOSC to process the hyperspectral data and then, built the PLSR models for confirming the optimal pre-processing method. Based on the optimal pre-processing result, the characteristic bands of crude protein were selected by successive projections algorithm and random frog algorithm, then the PLSR models were built for confirming the optimal selection method of characteristic variables and the optimal hyperspectral detection model. The results showed that the hyperspectral detection model based on SNV was the best in the five pre-processing methods. Thirty bands were selected by SPA and distributed in 530 to 700 nm and 940 to 1 000 nm. Six bands were selected by RF, and respectively were 826.544, 827.285, 828.766, 971.012, 972.494 and 973.235 nm. Therefore, the optimal hyperspectral detection model was SNV-RF-PLSR in this research, and the accuracy of model was good. The results of this research provided an optimal model and theoretical basis for hyperspectral detection of crude protein in pastures and in addition, developed new technique solutions for guiding the production of grassland industry.

**Keywords** Pasture; Crude protein; Hyperspectral; Successive projections algorithm; Random frog

(Received May 24, 2019; accepted Aug. 8, 2019)

\* Corresponding author