

藜草粗蛋白近红外定量分析模型的建立

季晓菲, 游明鸿, 白史且*, 李达旭, 雷雄, 吴婧, 陈莉敏,
张昌兵, 鄢家俊, 闫利军, 陈丽丽, 张玉

四川省草原科学研究院, 四川 成都 611731

摘要 藜草(*Phalaris arundinacea* L.)为多年生冷季型禾本科高产牧草,粗蛋白(CP)是评价饲草品质的关键指标,但目前的化学分析方法存在诸多缺点,寻求高效、快速、准确、安全的藜草CP测定方法是现代草地畜牧业发展和草原生态恢复急需解决的实际问题。本研究旨在利用近红外光谱(NIRS)建立藜草CP的定量分析模型,为快速测定藜草CP提供有效方法。试验采集不同品种(品系)、生育期、栽培条件、干燥方式、生长年限、部位以及刈割次数的藜草样品454份,采用瑞士Buchi公司的傅里叶近红外光谱仪和Operator软件采集原始光谱,应用K-S算法剔除具有相似光谱的样品,筛选出210份用于建模和模型评价。通过凯氏定氮法测定210份样品的粗蛋白含量并在Management console软件中对光谱进行赋值,再采用软件NIRcal 5.4按照6:3的比例将样品随机分为校正集和验证集,并剔除异常样品,运用不同的光谱预处理、回归算法、建模波段和主成分数建立8个藜草CP含量的近红外定量分析模型,通过外部验证表明8个模型均可以进行实际测定。最后比较不同的统计学参数获取最佳模型。结果表明,采用4000~10000 cm⁻¹的建模光谱波段、sa3+ncl+db1(3点平滑+趋近归一化+一阶导数处理)的预处理方法、8/1-4的初/次级主成分数和偏最小二乘法(PLS)所建的模型为最佳模型,其校正决定系数(R_{cal}^2)为0.9821,验证决定系数(R_{val}^2)为0.9802,均大于0.98,表明预测性能优秀;校正标准差(SEC)和验证标准差(SEP)分别为0.7802和0.7832,均较小且非常接近,表明模型的分析精度很高并具有很好的适应性;残差(BIAS)为-0.0005,接近于0,说明模型的稳定性很高,对外界因素不敏感;预测相关系数(r)为0.99,可见样品化学值与定标模型预测值的相关度极高;相对分析误差(RPD)为7.37,RPD>4.0表明模型能够很好地进行定量分析。综上,该试验在国内首次建立了藜草CP近红外定量分析模型,该模型样品来源多、数量大、分布范围广,预测精度和准确度高,适用范围大,为快速测定藜草粗蛋白提供了有效方法,在藜草品质分析、育种、家畜日粮配置以及草产品评价流通等方面具有应用前景。

关键词 近红外光谱;藜草;粗蛋白;模型

中图分类号: O657.3

文献标识码: A

DOI: 10.3964/j.issn.1000-0593(2019)06-1731-05

引言

藜草(*Phalaris arundinacea* L.)为禾本科藜草属多年生冷季型高产牧草,具有良好的适应性和优异的生产性能被广泛用于人工草地建设、草地生态保护、草地改良等,对畜牧业发展和生态建设都起着重要作用^[1]。粗蛋白(crude protein, CP)是衡量饲草品质的关键指标,藜草粗蛋白受物候

期、发育水平、基因型、栽培措施等诸多因素影响差异较大^[2]。但目前常用的化学测定方法存在耗时、费力、成本高、步骤繁琐、污染环境等缺点^[3],因此寻求快速、高效、准确、安全的藜草粗蛋白测定方法,成为现代草地畜牧业发展和草原生态恢复急需解决的实际问题。

采用近红外光谱(NIRS)测定CP含量已成为农作物及饲草品质研究的重要手段^[4-5],具有操作方便、检测效率高、成本低、安全无污染、稳定性高和样品制备简单等优点。目前,已应用在牧草领域,NIRS已在紫花苜蓿、燕麦、羊草、老芒

收稿日期:2018-04-26,修订日期:2018-10-16

基金项目:国家牧草产业技术体系(CARS-34),四川省“十三五”饲草育种攻关项目(2016NZ0098-1102),四川省公益性科研院所基本科研项目资助

作者简介:季晓菲,女,1986年生,四川省草原科学研究院助理研究员 e-mail: jixiaofei2012@qq.com

* 通讯联系人 e-mail: baishiqie@126.com

麦以及天然牧草等粗蛋白含量测定,但尚未见用于分析藜草粗蛋白的报道。本工作在国内首次建立了藜草粗蛋白含量的近红外定量分析模型,为快速测定藜草粗蛋白提供有效方法,为藜草品质分析、育种、家畜日粮配置以及草产品流通提供技术支撑。

1 实验部分

1.1 材料

2015年9月至10月和2016年4月至11月采集不同品种(品系)、生育期(分蘖期、拔节期、孕穗期、抽穗期、开花期、灌浆期、乳熟期、蜡熟期、完熟期、枯黄期)、栽培条件(8种行距、7种施肥梯度)、干燥方式(自然晾干、65℃烘干、杀青后65℃烘干)、生长年限(1~6年)、部位(茎、叶、穗和全株)以及不同刈割次数的藜草样品454份。除不同干燥方式的样品按要求干燥然后65℃烘干至恒重外,其余样品采集后直接经105℃杀青后65℃烘干至恒重,再采用高速粉碎机粉碎,过40目筛后装袋,放于避光处备用。

1.2 光谱采集

采用瑞士Buchi公司的傅里叶近红外光谱仪(NIRFlex N-500)和附带软件Operator采集样品光谱,谱区范围4000~10000 cm⁻¹,扫描次数32,分辨率8 cm⁻¹,扫描间隔4 cm⁻¹。在Petri培养皿中装入粉状样品,保证表面平整且装样量约在培养皿的3/4处,重复扫描3次。

1.3 粗蛋白常规化学测定

应用Kennard-Stone算法(Matlab 7.01)剔除具有相似光谱的样品^[6],挑选210份用于建模和模型评价。按照GB/T 6432-94凯氏定氮法测定所选样品的粗蛋白含量,每个样品重复测定3次取平均值。

1.4 模型建立与评价

在软件Management console(瑞士Buchi公司)中将测定的化学值赋值到光谱,再采用软件NIRcal 5.4将样品光谱按照6:3的比例随机分为校正集(Calibration-set)和验证集(Validation-set),以保证同一样品的三条光谱在同一样品集。粗蛋白含量的最大、最小值划归校正集,以避免模型外推。接着运用不同的回归算法、光谱预处理方法、建模波段和主成分数建立校正模型,并对光谱残差、属性残差进行计算以剔除异常样品。最后通过外部验证对模型的实际预测能力进行评价,采用决定系数(R^2)、校正标准差(SEC)、验证标准差(SEP)、残差(BIAS)、相对分析误差(RPD)^[6]等统计学参数验证模型效果和筛选最佳模型。

2 结果与讨论

2.1 近红外光谱扫描结果

454份藜草样品的原始光谱如图1所示,光谱曲线趋势一致,说明样品中无干扰杂质或干扰杂质的影响可忽略,而不同样品表现出不同的反射率,其差异说明各样品成分含量不同,光谱数据符合建立近红外定量分析模型的要求。

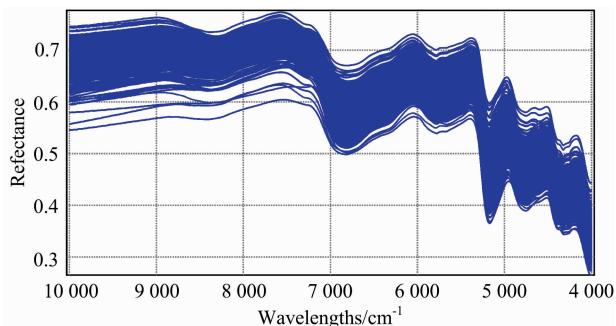


图1 藜草样品的原始光谱图

Fig. 1 The raw spectra of *Phalaris arundinacea* samples

2.2 粗蛋白测定结果

从所选的样品中剔除了6个异常值,共204份样品,其化学测定结果如表1所示。藜草样品粗蛋白含量变幅为2.70%~29.08%。陈光吉等^[8]测得不同生育期变幅为7.77%~10.83%;张永亮等^[2]报道分蘖期、抽穗期、开花期和结实期的CP分别为23.8%,13.6%,9.6%,5.1%;贾有余等^[9]研究表明不同栽培条件下的CP变幅为3.05%~20.64%。骆秀梅等^[10]发现不同行距与施氮量下藜草各器官CP变幅为5.61%~25.69%。相比之下,本试验用于建模的粗蛋白变幅更大,具有广泛的适用性,同时校正集和验证集的最大值、平均值、最小值和标准差都非常接近,表明校正集和验证集的分组合理,满足建立近红外分析模型的基础。

表1 校正集和验证集藜草样品的粗蛋白化学分析结果(%DM)

Table 1 CP of *Phalaris arundinacea* samples used in calibration and validation sets (%DM)

样品集	样品数	最小值	最大值	平均值	标准差
校正集	136	2.70	29.08	12.72	5.83
验证集	68	3.21	27.09	12.95	5.56

2.3 粗蛋白定标模型的建立

表2为采用软件NIRcal 5.4通过不同的预处理方法、建模波段、回归算法和主成分数组合得到的CP定标模型。试验所建立的8个定标模型的校正决定系数(R_{cal}^2)均大于0.98,表现出较好的预测性能^[11],校正标准差(SEC)较小,均在2%以内,满足分析精度的标准,可以进一步筛选最优模型。

2.4 模型的外部验证

建立校正模型后,通过外部验证对模型的实际预测能力进行评价,验证结果见表3。验证决定系数(R_{val}^2)在0.9580~0.9881之间,验证标准差(SEP)都在2%DM以内,表明预测结果的精确性较高。残差BIAS范围在-0.1012~0.1018,均接近0;斜率范围0.9455~1.0051,均接近1,表明预测结果的准确性较高。采用成对数据双尾t检验分析化学值和预测值的差异性,表明8个模型的化学测定值与预测值的差异均不显著($p>0.05$);相关系数(r)在0.9788~0.9941,说明样品化学值与定标模型预测值的相

表 2 建立的 8 个粗蛋白近红外定标模型的条件及对应结果

Table 2 Conditions of establishing NIRS models and results of 8 calibrations for CP

编号	回归算法	光谱预处理方法	建模波段/cm ⁻¹	初/次级主成分	R ² _{cal}	SEC
1	PLS	ncl	5 000~10 000	8/1-8	0.978 8	0.849 2
2	PLS	mf	5 000~10 000	8/1-8	0.979 1	0.840 4
3	PLS	mf	5 000~7 144, 7 404~10 000	6/1-6	0.979 5	0.835 3
4	PCR	mf+db1+nle	4 000~10 000	3/1-3	0.959 0	1.156 3
5	PLS	ncl+db1	4 000~10 000 cm ⁻¹	9/1-4	0.983 0	0.759 7
6	PLS	ds2	4 000~10 000	9/1-7	0.991 6	0.528 1
7	PLS	sa3+ncl+db1	4 000~10 000 cm ⁻¹	8/1-4	0.982 1	0.780 2
8	PLS	mf+db1+nle	4 000~10 000	8/1-4	0.985 3	0.709 0

注: PLS: 偏最小二乘法; PCR: 主成分回归; ncl: 趋近归一化; mf: 整体多元散射校正; nle: 结合单位长度归一化; db1: 一阶导数处理; sa3: 3 点平滑; ds2: Segment 5 Gap5 的 3 点二阶泰勒求导

Note: PLS: Partial least squares; PCR: Principal component regression; ncl: Normalization by closure; mf: multiple scatter correction (full); nle: Normalization (to unit length); db1: Derivatives (1st BCAP); sa3: Smoothing average 3 points; ds2: 2nd Taylor 3 points Segment 5 Gap5

表 3 藜草粗蛋白含量的外部验证参数

Table 3 Parameters of external validation for CP of *Phalaris arundinacea* samples

编号	R ² _{val}	SEP/%	BIAS	Slope	r	t	t _{0.05}
1	0.975 4	0.888 4	-0.101 2	0.945 5	0.987 6	1.622 8	1.960 0
2	0.976 1	0.874 2	0.093 0	0.947 6	0.988 0	1.514 5	1.960 0
3	0.976 6	0.867 3	0.101 8	0.946 4	0.988 2	1.671 8	1.960 0
4	0.958 0	1.157 5	-0.033 5	1.005 1	0.978 8	0.408 2	1.960 0
5	0.980 7	0.772 1	0.004 5	0.979 9	0.990 3	0.083 2	1.960 0
6	0.988 1	0.607 9	0.035 1	0.978 3	0.994 1	0.822 1	1.960 0
7	0.980 2	0.783 2	-0.000 5	0.981 1	0.990 0	0.009 2	1.960 0
8	0.982 1	0.744 0	0.000 8	0.986 1	0.991 0	0.156 6	1.960 0

关度极高。由此可见, 所建的 8 个粗蛋白定标模型均可以对藜草粗蛋白进行实际检测。

2.5 最佳模型的确定和分析

根据 R²_{cal}, R²_{val}, SEC, SEP, BIAS, r 等指标确定最优模型。综合来看, 模型 7 具有较高的 R²_{cal}, R²_{val} 和 r, 以及较低的 SEC 和 SEP, 并且 SEC 和 SEP 十分接近, BIAS 最小, 可确定为最优模型。其中 R²_{cal} 和 R²_{val} 分别为 0.982 1 和 0.980 2, 均大于 0.98; SEC 和 SEP 分别为 0.780 2% 和 0.783 2%, 均较小且非常接近, 表明模型的分析精度很高并具有很好的适应性; 残差(BIAS)为 -0.000 5, 接近于 0, 说明模型的稳定性很高; 预测相关系数(r)为 0.990 0, 说明样品化学值与定标模型预测值的相关度极高; 最佳模型的相对分析误差(RPD)为 7.37, RPD>4.0 进一步验证了藜草 CP 定标模型能够很好地进行定量分析^[10]。最佳模型的化学值与预测值的相关分析如图 2 所示, 校正集方程式为 $f(x) = 0.982 1x + 0.228 2$, 验证集方程式为 $f(x) = 0.981 1x + 0.245 6$, 可见模型的预测值与化学分析值非常相近。

由于近红外光谱技术是一种间接分析技术, 其检测准确性与样品的代表性、化学分析误差、样品状态、装样条件、光谱预处理方法等诸多因素相关。本研究采集了不同品种(系)、部位、栽培条件、干燥方式、生育期、生长年限以及刈割次数的藜草样品 454 份, 样品来源广泛, 所测粗蛋白含量变幅为 2.70%~29.08%, 包含了目前科研与生产上可能出

现的含量水平, 使建立的藜草粗蛋白模型能够满足实际生产和科研的要求。同时化学测定采用国家标准凯氏定氮法, 并三次重复测定, 能够有效减少化学分析误差。在样品含水量、样品温度、样品粒度和装样条件上都采用统一干燥标准、重复装样的方法来消除干扰。并且通过比较不同的光谱预处理、回归算法、建模波段和主成分数建立了较优的 8 个模型, 再通过统计学评价参数综合分析选出最优模型, 保证了建模质量。

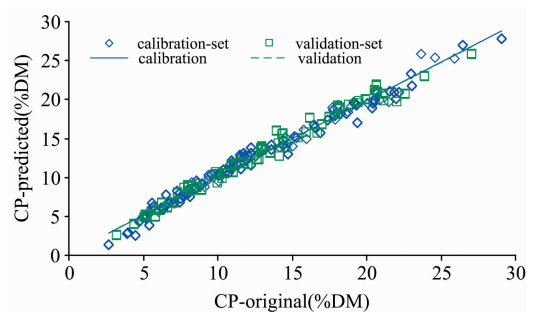


图 2 校正集与验证集化学测量值与预测值的相关性
Fig. 2 Correlation between original and predicted values of CP in calibration and validation sets

目前国内外尚未见报道近红外在藜草品质分析上的应用, 国内学者对其他饲草的粗蛋白近红外模型已有研究, 严

旭等^[6]以 110 份样品建立的老芒麦粗蛋白近红外定量模型的 R_{cal}^2 和 R_{val}^2 分别为 0.994 5 和 0.993 8, SEC 和 SEP 分别为 0.322 9 和 0.326 1, 相关系数 r 为 0.996 9, 相对分析误差 (RPD) 为 12.63; 刘哲等^[12]采集内蒙古锡林郭勒盟草原主要牧草及混合牧草样品 407 份建立的天然牧草粗蛋白近红外定量模型的 SEP 为 0.571, RPD 为 5.949。由于不同种牧草组成成分差异较大, 并且不同光谱仪、制样方法和化学计量学方法具有一定差别, 因此不同种牧草间的建模效果难以比较, 但从定标样品和验证参数相比较而言, 本研究建立的 CP 含量定标模型, 具有样品数量大、来源多、分布范围广的特点, 同时建模效果相似。

因此, 本研究结果表明藜草 CP 模型预测效果良好, 可用于科研和生产的实际分析。为综合评价藜草的营养价值后期将尝试建立藜草的其他品质指标模型。同时, 为了在生产

中广泛应用 NIRS 模型, 将进一步在实际应用中添加不同年份的样品数据, 对方程进行纠正和优化^[13-14], 以提高模型的适应性和预测能力。

3 结 论

藜草粗蛋白最优模型为采用 $4\ 000\sim 10\ 000\ \text{cm}^{-1}$ 的光谱波段、sa3+ncl+dbl 的预处理方法、8/1-4 的初/次级主成分和偏最小二乘法 PLS 所建的模型, 其 R_{cal}^2 和 R_{val}^2 分别为 0.982 1 和 0.980 2, SEC 和 SEP 分别为 0.780 2% 和 0.783 2%, 相关系数 r 为 0.99, RPD 为 7.37, 表明该模型能够准确的进行定量分析, 为快速测定藜草粗蛋白提供了有效方法, 在藜草品质分析、育种、家畜日粮配置以及草产品流通等方面具有应用前景。

References

- [1] Zhang C, Zhang D W, Sun Y N, et al. *Journal of Plant Physiology*, 2017, 215: 11.
- [2] ZHANG Yong-liang, LUO Xiu-mei(张永亮, 骆秀梅). *Acta Agrestia Sinica(草地学报)*, 2008, 16(6): 659.
- [3] Ye D, Sun L, Zou B, et al. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2018, 189: 463.
- [4] López A, Arazuri S, Jarén C, et al. *Procedia Technology*, 2013, 8(1): 488.
- [5] Ercioglu E, Velioglu H M, Boyaci I H. *Talanta*, 2018, 178: 716.
- [6] YAN Xu, BAI Shi-qie, YAN Jia-jun, et al(严旭, 白史且, 鄢家俊, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2015, 35(8): 2103.
- [7] Magwaza L S, Naidoo S I M, Laurie S M, et al. *LWT-Food Science and Technology*, 2016, 72: 63.
- [8] CHEN Guang-ji, SONG Shan-dan, GUO Chun-hua, et al(陈光吉, 宋善丹, 郭春华, 等). *Acta Prataculturae Sinica(草业学报)*, 2015, 24(9): 63.
- [9] JIA You-yu, LUO Xiu-mei, ZHANG Yong-liang(贾有余, 骆秀梅, 张永亮). *Pratacultural Science(草业科学)*, 2016, 33(11): 2312.
- [10] LUO Xiu-mei, ZHANG Yong-liang(骆秀梅, 张永亮). *Journal of Inner Mongolia University for the Nationalities • Natural Sciences(内蒙古民族大学学报 • 自然科学版)*, 2017, 32(5): 418.
- [11] Fagan C C, Everard Colm D, McDonnell K. *Bioresource Technology*, 2011, 102(8): 5200.
- [12] LIU Zhe, WANG Yu-qin, XUE Shu-yuan, et al(刘哲, 王玉琴, 薛树媛, 等). *Acta Agrestia Sinica(草地学报)*, 2018, 26(1): 249.
- [13] Gislum R, Mick lander E, Nielsen J P. *Field Crops Research*, 2004, 88: 269.
- [14] WANG Jia-hua, TANG Zhi-hui, HAN Dong-hai(王加华, 汤智辉, 韩东海). *Journal of Food Safety and Quality(食品安全质量检测学报)*, 2014, 5(3): 742.

Establishment of Quantitative Model for Analyzing Crude Protein in *Phalaris arundinacea* L. by Near Infrared Spectroscopy (NIRS)

Ji Xiao-fei, YOU Ming-hong, BAI Shi-qie*, LI Da-xu, LEI Xiong, WU Qi, CHEN Li-min, ZHANG Chang-bing, YAN Jia-jun, YAN Li-jun, CHEN Li-li, ZHANG Yu
Sichuan Academy of Grassland Science, Chengdu 611731, China

Abstract Reed canary grass (*Phalaris arundinacea* L.) is a perennial cool-season gramineae grass with a high yield. Crude protein (CP) is a key indicator in the evaluation of forage quality, but the use of chemical analytical methods to determine the CP content is disadvantageous. Therefore, a fast, efficient, accurate, and safe determination method is required in the development of modern grassland agriculture, animal husbandry and grassland ecological restoration. The purpose of this study is to use near-infrared spectroscopy (NIRS) techniques to develop a quantitative model for the analysis of CP in reed canary grass and provide an effective method for a rapid determination. We collected 454 samples of reed canary grass from various resources, including different cultivars (or strains), different growth stages, different cultivation conditions, different drying methods, different growth years, different parts and different harvest times. The original spectra of all of the samples were obtained using a near-infrared spectrometer (NIRFlex N-500) and Operator software of the Swiss Buchi company. A total of 210 samples were selected for the development and evaluation of models after deleting samples with similar spectra by a K-S algorithm, and were assayed using the Kjeldahl nitrogen method to obtain the chemical values of CP; we then assigned them to spectra in a Management console software. The samples were randomly divided into a calibration set and a validation set according to the proportion of 6 : 3, using the NIRcal 5.4 software; the outliers were then eliminated. We established 8 quantitative analysis models for the CP content of reed canary grass by applying different spectral data pretreatments, primary/secondary principal components, spectral regions, and regression algorithms. We revealed that all of the 8 models can be used in the determination of CP by performing an external validation. The best model was chosen by comparing statistic parameters. The results showed that the best calibration model was developed by the spectral data pretreatment of sa3+ncl+db1 (smoothing average 3 points+ normalization by closure+first derivative BCAP), choosing the primary/secondary principal component of 8/1-4 and spectral region of 4 000~10 000 cm^{-1} in combination with the partial least square (PLS) regression algorithm. Its calibration coefficient of determination (R_{cal}^2) and external validation coefficient of determination (R_{val}^2) were 0.982 1 and 0.980 2, respectively; both were larger than 0.98, suggesting an excellent predictive ability. The standard errors of calibration (SEC) and prediction (SEP) were 0.780 2 and 0.783 2, respectively; both were very small and similar, which demonstrated the high analytical accuracy and robust fitting. The bias value of $-0.000 5$, close to 0, demonstrated the model's stability and robustness, i. e., its insensitivity to the external factors. The correlation coefficient of validation (r) of 0.99 indicated a very high correlation between the predicted and chemical values. The residual predictive deviation (RPD) was 7.37 (above 4.0), further confirming that the CP model can be used for a high-quality quantitative analysis. Therefore, in this study, a quantitative model for a CP analysis of *Phalaris arundinacea* L. was developed using NIRS for the first time in China with a large data collection from different sources and high accuracy, which guaranteed the reliability and practicability. The model provides an effective method to quantify CP of reed canary grass for a rapid screening of germplasm in breeding programs, optimization of the allocation of livestock diets, and classification of forage products in the supply chain.

Keywords NIRS; *Phalaris arundinacea* L.; Crude protein; Quantitative model

(Received Apr. 26, 2018; accepted Oct. 16, 2018)

* Corresponding author