

近红外透反射光谱法测定牛奶成分

杨晋辉, 卜登攀, 王加启*, 周凌云, 马 露, 张军民
(中国农业科学院北京畜牧兽医研究所, 动物营养学国家重点实验室, 北京 100193)

摘要: 为探索近红外透反射光谱法检测牛奶中主要成分的可行性, 从牧场采集150份乳样, 并用乳质分析仪测定其中的蛋白质、脂肪、乳糖、总固形物、非脂固形物含量。通过光谱数据结合偏最小二乘法以及完全交互验证建立回归模型。结果表明: 中长波(1300~2500nm)对乳成分的模型贡献较大; 蛋白质、脂肪、乳糖、总固形物、非脂固形物模型的验证集决定系数分别达到0.96、0.90、0.90、0.91、0.92, 相对分析误差分别为4.97、3.19、3.15、3.37、3.54; 对10个未知样品的预测结果表明, 该模型对蛋白质、乳糖和非脂固形物含量的预测效果较好, 相对误差均小于1.40%。

关键词: 牛奶; 近红外透反射光谱; 乳成分; 偏最小二乘法

Determination of Chemical Composition of Milk Using Near-Infrared Reflectance Spectroscopy

YANG Jin-hui, BU Deng-pan, WANG Jia-qi*, ZHOU Ling-yun, MA Lu, ZHANG Jun-min
(State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China)

Abstract: This study reports on the application of near infrared reflectance spectroscopy to determine the major chemical components in milk. Totally 150 milk samples were collected from the same dairy farm, and the contents of protein, fat, lactose, total solids and non-fat solids in milk were measured. Based on the spectral data, a regression model for each component was proposed by partial least square regression combined with cross-validation. The results showed that middle- and long- wavelength (1300–2500 nm) absorption contributed much to the regression models. The coefficient of determination for the models for protein, fat, lactose, total solids and non-fat solids in validation sets were 0.96, 0.90, 0.90, 0.91 and 0.92, respectively, with relative errors of 4.97, 3.19, 3.15, 3.37 and 3.54, respectively. The models for protein, lactose and non-fat solids were effective in predicting 10 unknown samples, with relative error less than 1.40%.

Key words: milk; near-infrared transfectance spectroscopy; milk components; partial least squares

中图分类号: TS252.7

文献标志码: A

文章编号: 1002-6630(2013)20-0153-04

doi:10.7506/spkx1002-6630-201320031

乳成分是衡量牛奶营养价值的重要指标, 不仅决定了牧场经营的效益, 也反映了奶牛机体的代谢状况。因此, 乳成分的快速测定有助于牛奶生产的经营管理。随着光谱仪器和计算机技术的发展, 近红外检测技术被逐步应用于乳品的加工和检测。近红外光谱不仅可以用于牛奶的掺假识别^[1], 也能分辨不同储存条件、不同储存时间乳品质的差异^[2]。基于近红外光谱的定量模型能检测牛奶和羊奶中的主要成分^[3-4]。结合发展的计算机技术, 近红外仪器可以实现对乳成分的实时监测^[5-6]。国内许多学者对近红外光谱的预测功能和模型优化进行了研究^[7-8]。本研究采用透反射光谱预测牛奶中蛋白、脂肪、总固形物、非脂固形物和乳糖, 筛选最优波段, 通过偏最小二

乘法建立回归模型, 以期进一步探索和优化近红外光谱法对牛奶成分的检测技术。

1 材料与amp;方法

1.1 材料与试剂

150个乳样来自于北京市顺义区的一家大型奶牛场同一天不同奶牛的同一批次牛奶。

Broad Spectrum Microtabs II 防腐剂 美国Advanced Instruments公司。

1.2 仪器与设备

DS2500近红外光谱仪(内置硅与硫化铅探测器, 配有

收稿日期: 2012-10-18

基金项目: 国家“973”计划项目(2011CB100805); “十二五”国家科技支撑计划项目(2012BAD12B08)

作者简介: 杨晋辉(1987—), 男, 硕士研究生, 研究方向为反刍动物营养。E-mail: 710593491@163.com

*通信作者: 王加启(1967—), 男, 研究员, 博士, 研究方向为反刍动物营养和牛奶质量改良。E-mail: wang-jia-qi@263.net

0.2cm的金反射板)、Milkoscan FT120乳质分析仪 瑞士 Foss公司。

1.3 方法

1.3.1 牛奶光谱的采集和乳成分的测定

近红外光谱扫描范围400~2500nm, 分辨率0.5nm, 每条光谱4200个数据点, 采用透反射方式扫描。在150个乳样样品中分别添加0.6mg/mL防腐剂后摇匀, 并冷藏运送至实验室用于光谱的扫描和乳成分的测定。样品在扫描前, 40℃水浴约20min左右, 并不断摇匀。扫描时确保扫描区域无气泡存在。每个样品扫描2次, 最终输出平均光谱。采用乳质分析仪测定样品中的乳成分。

1.3.2 数据分析处理

将光谱数据导入Unscrambler 9.8中进行分析, 对乳中的蛋白质、脂肪、总固形物、非脂固形物、乳糖的含量分别利用偏最小二乘法(partial least squares, PLS)建立定量模型, 根据模型中的决定系数(R^2)和均方根误差(root-mean-square error, RMSE)来确定数据的预处理方式, 根据残差值删除异常数据, 并选取作用显著的波段优化模型。数据经过完全交互验证(full cross validation)建立最优模型, 根据相对分析误差评价模型的稳定性和预测功能^[7]。选取10个未参与建模的样品对模型进行验证。

2 结果与分析

样品扫描光谱如图1所示。

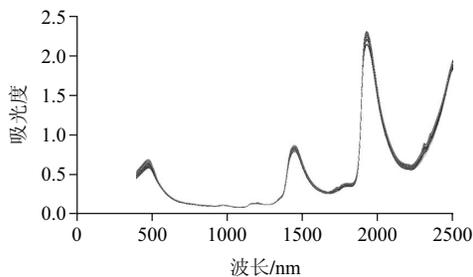


图1 150个牛奶样品的近红外透反射原始光谱

Fig.1 Original reflectance spectra of 150 raw milk samples

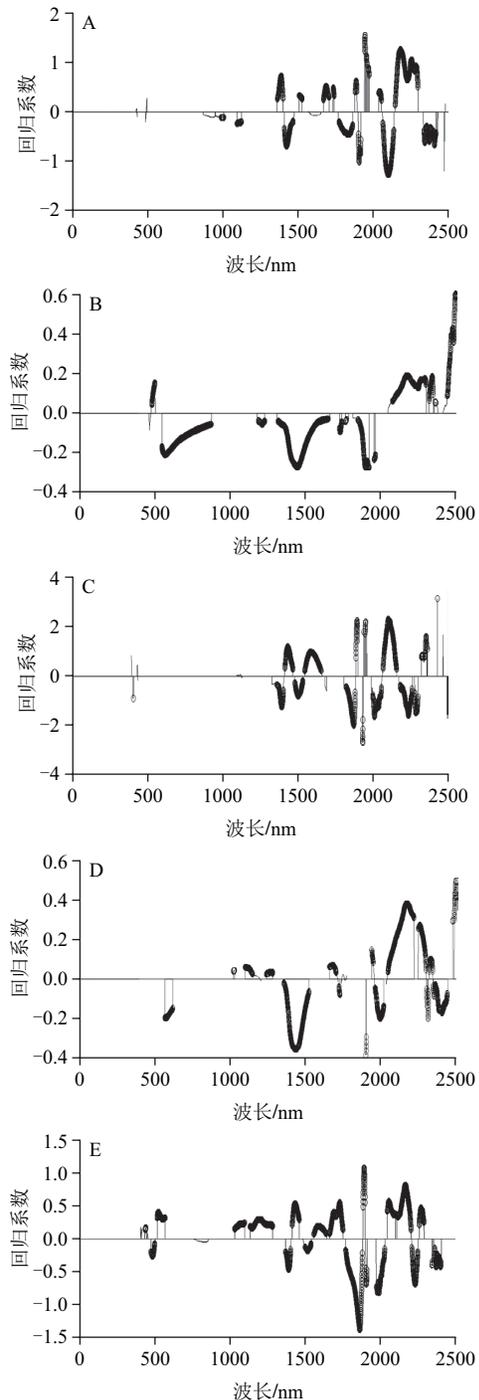
表1 乳成分统计参数

Table 1 Statistical data of milk components

指标	样品数量	最小含量/%	最大含量/%	平均含量/%	标准差/%
乳蛋白	143	3.09	4.39	3.70	0.24
乳脂肪	142	2.51	6.34	4.58	0.72
乳糖	142	4.56	5.35	5.02	0.14
总固形物	142	12.07	16.32	14.08	0.78
非脂固形物	146	8.42	10.23	9.48	0.27

乳成分信息如表1所示。共有150份乳样参与建模, 表中的样品数量为删除异常点后最终模型所用样品的数量。从表中可以看出, 不同样品中的乳糖含量差异不大

(标准差为0.14%), 而脂肪和总固形物含量差异较大(标准差分别为0.72%和0.78%)。



黑圈标注的为模型中的特征波段。

图2 乳中蛋白质(A)、脂肪(B)、乳糖(C)、总固形物(D)、非脂固形物(E)模型中的特征波段

Table 2 Characteristic wavelengths of protein (A), fat (B), lactose (C), total solids (D) and non-fat solids (E) in the models

如图2所示, 5种乳成分吸收峰主要出现在1300~2500nm波长处, 而在400~1000nm波段仅脂肪有少许的特征吸收波段。蛋白质的吸收峰位

于1364~1475、1683、1743、2118、2172~2290、2340~2430nm, 脂肪的吸收峰出现在500、550~876、1370~1570、1980、2130~2340、2479nm, 乳糖的吸收波段则出现在1400~1700、1870、2000~2300、2350nm, 总固形物模型中的重要波段变量主要分布在1370~1530、1970~2440nm, 非脂固形物在500nm以及1032~2408nm波段有多处吸收峰。

表2 乳成分模型参数
Table 2 Parameters of the models for milk components

指标	预处理	主成分个数	校正集			验证集			相对分析误差
			回归方程	均方根误差	决定系数	回归方程	均方根误差	决定系数	
蛋白质	未处理	9	$Y=0.969x+0.114$	0.041	0.969	$Y=0.960x+0.148$	0.048	0.959	4.97
脂肪	未处理	6	$Y=0.916x+0.383$	0.207	0.916	$Y=0.910x+0.410$	0.225	0.903	3.19
乳糖	MSC	12	$Y=0.928x+0.362$	0.038	0.928	$Y=0.911x+0.448$	0.044	0.902	3.15
总固形物	SNV	5	$Y=0.922x+1.093$	0.216	0.922	$Y=0.915x+1.188$	0.231	0.913	3.37
非脂固形物	SNV	11	$Y=0.940x+0.569$	0.066	0.940	$Y=0.927x+0.695$	0.076	0.921	3.54

表3 模型对未知样品的检测结果
Table 3 Prediction of unknown samples using the models

样品编号	蛋白质含量			脂肪含量			乳糖含量			总固形物含量			非脂固形物含量		
	真实值/%	预测值/%	相对误差/%	真实值/%	预测值/%	相对误差/%	真实值/%	预测值/%	相对误差/%	真实值/%	预测值/%	相对误差/%	真实值/%	预测值/%	相对误差/%
1	3.54	3.57	0.81	4.11	4.17	1.49	4.95	4.91	0.90	13.49	13.44	0.33	9.27	9.24	0.29
2	3.97	3.94	0.63	4.34	4.47	3.02	4.99	5.02	0.70	13.92	14.06	1.02	9.73	9.63	1.01
3	3.26	3.30	1.09	3.75	3.84	2.45	5.08	5.09	0.24	12.96	13.08	0.93	9.04	9.09	0.59
4	3.48	3.49	0.17	4.08	4.17	2.16	4.93	4.90	0.57	13.42	13.39	0.24	9.26	9.18	0.86
5	3.91	3.87	1.03	4.34	4.51	3.84	5.20	5.18	0.31	13.91	14.44	3.65	9.73	9.86	1.35
6	3.59	3.60	0.33	4.76	4.73	0.61	4.91	4.90	0.22	14.08	13.96	0.86	9.18	9.17	0.10
7	3.98	4.02	1.12	5.74	5.84	1.76	4.91	4.86	1.03	15.29	15.49	1.31	9.68	9.70	0.24
8	3.86	3.86	0.00	5.54	5.34	3.82	4.81	4.87	1.17	14.94	14.74	1.37	9.46	9.44	0.22
9	3.68	3.69	0.38	4.30	4.16	3.27	5.05	5.03	0.42	13.68	13.59	0.65	9.50	9.46	0.41
10	4.04	4.06	0.49	4.77	4.94	3.42	4.77	4.74	0.55	14.32	14.54	1.49	9.61	9.54	0.70

近红外光谱所建立的乳成分模型参数如表2所示。蛋白质的模型精度最好(校正集和验证集的决定系数分别为0.969和0.959, 均方根误差分别为0.041和0.048); 脂肪和总固形物的模型较为简明, 分别仅需要6个和5个主成分便可以建立模型, 但均方根误差相对较大(校正集和验证集的均方根误差均大于0.20); 各成分模型的相对分析误差均大于3, 模型的稳定性和预测能力较好。

利用模型对未知样品的乳成分进行预测, 结果如表3所示。模型对蛋白质、乳糖、非脂固形物的预测结果较好, 3种成分的相对误差均小于1.40%, 而对脂肪和总固形物的预测效果较差, 特别是脂肪的预测效果相对误差为0.61%~3.84%。

3 讨论

牛奶中防腐剂的添加少于1mg/mL并不会对脂肪和蛋白近红外模型的测定造成太大影响^[9]。为避免光程变化、仪器噪声、基线漂移、谱峰重叠以及大分子引起的散射效应对光谱所造成的影响, 需要通过对光数据进行数据

归一化、平滑、SNV、求导、MSC等预处理^[10]。本研究中, 样品的原始光谱便可以对乳脂肪和乳蛋白建立较好的模型, 而乳糖、总固形物和非脂固形物模型则需要对光谱进行预处理。乳脂肪和总固形物含量之间存在显著的线性关系^[7], 研究中乳脂和总固形物模型的校正集和验证集均方根误差较大, 这可能是由于未均质样品中粒径不等的脂肪球对光线的不规则散射和透射所引起的, 也可能与这两种成分的变异范围较大有关。

近红外分为短波(800~1100nm)和中长波(1100~2500nm)两段, 短波段有较好的穿透能力, 适合做透射分析^[9,11]; 而中长波穿透能力较弱, 适合做反射分析^[12-14]。但有研究证明, 基于中长波的透射光谱也可得到较好乳脂肪和乳蛋白的模型^[4,9]。而近红外-可见光的透射光谱(400~1000nm)对于牛奶中蛋白、脂肪和乳糖均未能建立良好的模型, 其反射光谱则可以建立较好的脂肪和蛋白模型^[14]。中波(1100~1800nm)的漫反射光谱的模型参数要优于透射光谱的模型^[15]。由于牛奶中脂肪和蛋白质分子特性的不同, 长波(1700~2500nm)漫反射光谱对于脂肪建模效果优于蛋白质模型^[14,16-17]。本研究采用透射装置, 牛奶的扫描厚度为0.2cm, 透射光在其中的光程为0.4cm, 较好的结合了透射光谱和漫反射光谱的优点。

蛋白肽键中N—H键对光谱的吸收形成了蛋白特有的吸收峰, 1400~1500nm波长处为肽键和氨基中N—H键伸缩振动一倍频区, 2050~2200nm波长处为蛋白或氨基中N—H键伸缩振动或弯曲振动的倍频区和合频区, 2300~2400nm波长处则为蛋白中亚甲基和次甲基弯曲振动的二倍频区^[13,18-19]。脂肪对于近红外光谱的吸收源于脂肪中C=O键和脂肪酸链中C—H键的伸缩振动, 1365~1395nm波长处为C—H键的合频区, 1620~1695nm波长处为C—H键的伸缩振动一倍频区, 2140~2380nm波长处主要为C—H键伸缩振动和—CH₂变形振动的合频区, 2470nm波长处为甲基中C—H键的合频区^[13,18-19]。本研究中这些波段都对乳蛋白和乳脂肪模型构建有显著的贡献作用。乳糖是由葡萄糖和半乳糖缩合而成的二糖, 其中的C—OH键能够形成特定吸收峰, 1450nm和2000nm波长处分别为羟基振动的一倍频区和合频区^[18-19], 本研究中1350~1690nm以及2000~2300nm波长处为乳糖的主要响应区, 这与文献报道的研究结果相近^[19]。

本研究中乳糖模型的参数要优于其他研究^[4,14,20], 并能够对乳中乳糖含量进行准确预测(相对误差为0.22%~1.17%), 而且首次利用牛奶的透射光谱对乳中的总固形物和非脂固形物进行了预测, 模型的验证集决定系数和相对分析误差分别都大于0.91和3.30, 模型验证的结果也显示, 光谱对乳中的总固形物和非脂固形物含量预测准确率达到了90%和100%(相对误差小于1.50%)。

蛋白质模型的参数与其他研究相近^[7,14,20],但模型对于脂肪的预测能力相对较弱,仍需进一步改进。

本研究利用近红外的透反射光谱,通过不同的预处理,建立乳成分的定量模型,乳成分的特征吸收波段集中在1300~2500nm,牛奶中的5种乳成分模型均具有较好的稳定性(相对分析误差>3)和预测精度(决定系数>0.90),并对乳中蛋白质、乳糖和非脂固形物含量有较好的预测效果。

参考文献:

- [1] 李亮,丁武. 掺有植物性填充物牛奶的近红外光谱判别分析[J]. 光谱学与光谱分析, 2010, 30(5): 238-242.
- [2] AL-QADIRI H M, LIN M, AL-HOLY M A, et al. Monitoring quality loss of pasteurized skim milk using visible and short wavelength near-infrared spectroscopy and multivariate analysis[J]. Journal of Dairy Science, 2008, 91(3): 950-958.
- [3] DRÁČKOVÁ M, HADRA L, JANŠTOVÁ B, et al. Analysis of goat milk by near-infrared spectroscopy[J]. Acta Veterinaria Brno, 2008, 77(3): 415-422.
- [4] TSENKOVA R, ATANASSOVA S, ITOH K, et al. Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a spectral region from 1,100 to 2,400 nanometers[J]. Journal of Animal Science, 2000, 78(3): 515-522.
- [5] KAWAMURA S, KAWASAKI M, NAKATSUJI H, et al. Near-infrared spectroscopic sensing system for online monitoring of milk quality during milking[J]. Sensing and Instrumentation for Food Quality and Safety, 2007, 1(1): 37-43.
- [6] KAWASAKI M, KAWAMURA S, TSUKAHARA M, et al. Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot[J]. Computers and Electronics in Agriculture, 2008, 63(1): 22-27.
- [7] 李晓云,王加华,黄亚伟,等. 便携式近红外仪检测牛奶中脂肪、蛋白质及干物质含量[J]. 光谱学与光谱分析, 2011, 31(3): 665-668.
- [8] 彭丹,徐可欣,宋杨. 用多维校正法提高近红外牛奶成分校正模型稳健性的研究[J]. 光谱学与光谱分析, 2009, 29(4): 913-917.
- [9] 鲁超. 近红外光谱技术在液态奶质量评定中的应用研究[D]. 北京: 中国农业大学, 2006: 24-26.
- [10] 王硕. 近红外光谱分析中牛奶模型优化方法的初步研究[D]. 天津: 天津大学, 2005: 37-41.
- [11] 皮付伟,王艳岭,鲁超. CCD短波近红外光谱仪测定牛奶成分的可行性研究[J]. 现代科学仪器, 2006(4): 34-36.
- [12] 吕丽娜,张玥,周定文. 采用近红外漫反射光谱法分析牛奶成分[J]. 天津大学学报, 2004, 37(12): 1093-1096.
- [13] 王丽杰,徐可欣,郭建英. 采用近红外光谱技术检测牛奶中脂肪、蛋白质及乳糖含量[J]. 光电子·激光, 2004, 15(4): 468-471.
- [14] AERNOUTS B, POLSHIN E, LAMMERTYN J, et al. Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: reflectance or transmittance[J]. Journal and Dairy Science, 2011, 94(11): 5315-5329.
- [15] 褚鹏蛟,吕丽娜,常敏,等. 复合近红外光谱法在牛奶测定中的应用[J]. 理化检验: 化学分册, 2007, 43(2): 116-121.
- [16] 王云,徐可欣,常敏. 近红外光谱技术检测牛奶中脂肪及蛋白质含量校正模型的建立[J]. 光学仪器, 2006, 28(3): 3-7.
- [17] 常敏. 牛奶质量光学检测的作用机理及影响因素的研究[D]. 天津: 天津大学, 2006: 17-25.
- [18] 王丽杰. 快速检测牛奶成分的近红外光谱测量方法及系统研究[D]. 哈尔滨: 哈尔滨理工大学, 2006: 21-23.
- [19] 张玥. 近红外牛奶成分分析方法及仪器研究[D]. 天津: 天津大学, 2004: 34-39.
- [20] JANKOVSKÁ R, ŠUSTOVÁ K. Analysis of cow milk by near-infrared spectroscopy[J]. Czech Journal of Food Science, 2009, 21(4): 123-128.