

近红外光谱分析法快速测定 稻谷常规化学指标

杨学文

(江西省粮油质量监督检验中心,江西 南昌 330046)

摘要:采集150份有代表性的我国南方地区稻谷样品的近红外光谱,用偏最小二回归分析法(PLS),建立了稻谷的水分、直链淀粉、蛋白以及胶稠度的近红外定量分析模型,并对30份预测集样品进行了验证。水分、直链淀粉、蛋白以及胶稠度的校正集模型的决定系数 (R^2) 分别为0.990 3、0.560 3、0.913 2以及0.678 0,交互验证均方根误差(RMSECV)分别为0.372 8%、1.456 9%、0.305 4%以及5.031 5%;验证集标准预测偏差(RMSEP)分别为0.382 5%、1.465 0%、0.510 0%以及5.052 1%。结果表明,近红外光谱分析法可以满足快速分析的要求。

关键词:近红外光谱;偏最小二回归;稻谷;品质指标测定

中图分类号:TS 212.7 **文献标识码:**A **文章编号:**1007-7561(2016)05-0065-05

Rapid determination of rice conventional chemical index by near infrared spectroscopy

YANG Xue - wen

(Jiangxi Grain and Oil Quality Supervision and Inspection Center, Nanchang Jiangxi 330046)

Abstract:150 near infrared spectrum of rice which obtained from southern China were collected. Combined with partial least square (PLS) regression method, the quantitative analysis models of rice moisture, protein, amylose, and gel consistency were established and verified by 30 predication sets. The coefficient of determination (R^2) and the root - mean - square error of cross - validation (RMSECV) of the moisture, amylose, protein and gel consistency models were 0.990 3 and 0.372 8%, 0.560 3 and 1.456 9%, 0.913 2 and 0.305 4%, and 0.678 0 and 5.031 5%, respectively; standard error of prediction (RMSEP) were 0.382 5%, 1.465 0%, 0.510 0% and 5.052 1%, respectively. The results indicated that near - infrared analysis method can meet the demand of rapid determination.

Key words:near infrared spectroscopy; partial least square regression; rice; quality index determination

水稻是我国最主要的粮食作物之一,我国水稻的播种面积约占粮食作物总面积的1/4,产量约占全国粮食总产量的40%。目前水稻的常规化学成分如水分、直链淀粉、蛋白、胶稠度等指标主要通过常规的化学分析法进行^[1-3],样品往往需要进行去壳等前处理操作,过程繁琐,检测速度较慢,无法满足粮食收购和水稻质量快速分析的需要。因此,开发一种水稻常规化学指标的快速分析方法具有十分重要的意义。

近红外光谱分析法是一种绿色环保、简单快速且能够无损检测的一种分析方法,已广泛应用于食

品^[4]、医药^[5]、生物^[6]和石油化工^[7]等领域。目前,在水稻的近红外光谱分析方面已有部分研究^[8-10],但尚无同时测定稻谷水分、蛋白、直链淀粉以及胶稠度的相关研究。为此,本实验通过采集我国南方地区150份稻谷样品的近红外光谱,应用偏最小二乘回归(PLS),研究近红外光谱分析法同时快速测定稻谷水分、直链淀粉、粗蛋白含量以及胶稠度常规化学指标的可行性。

1 材料与方法

1.1 样品与仪器

采用来自我国南方水稻产区共150份稻谷样品,其中江西50份、广西30份、广东40份、海南30份。样品水分含量范围为10.20%~27.30%;直

收稿日期:2016-02-28

作者简介:杨学文,1973年出生,男,高级工程师。

链淀粉含量范围为 13.40% ~ 28.00%; 蛋白质含量范围为 8.30 ~ 13%; 胶稠度范围为 52 ~ 89 mm。

IM9500 近红外谷物分析仪: 瑞典 Perten 公司; Unscrambler 软件: 挪威 CAMO 公司; 3100 旋风粉碎磨: 瑞典 Perten 公司; ED 115 烘箱: 德国 BINDER; VAP300 凯氏定氮仪: 德国 Gerhardt 公司; JLGJ-45 砗谷机: 河南中主良仪器设备有限公司; GY100 高速旋转粉碎机: 五洲鼎创(北京)科技有限公司; TU-1900 紫外可见分光光度计: 北京普析通用仪器有限公司; HH.S21-Ni8 恒温水浴锅: 北京三二八科学仪器有限公司。

1.2 实验方法

1.2.1 光谱采集

在室温下直接采集稻谷样品的近红外光谱。谱区采集范围: 570 ~ 1 100 nm, 光谱数据间隔: 0.5 nm, 扫描次数: 10 次。每个稻谷样品重复测量 3 次, 采用 3 次测量的平均光谱进行建模。

1.2.2 光谱建模方法

从各省份样品中随机选取 20%, 共 30 个样品作为检验集进行外部预测, 剩余的 120 个样品作为校正集进行交互验证建立定量分析模型。采用 Unscrambler 软件建立偏最小二乘回归定量分析模型并对预测集进行预测。偏最小二乘法是目前使用最为广泛的近红外定量分析方法。

1.2.3 模型评价方法

模型的评价主要由决定系数(R^2)、交互验证均方根误差(RMSECV)、和预测集均方根误差(RMSEP)等参数来评价。 R^2 越接近于 1, 说明参照值与预测值的数值越接近。模型的 RMSECV 值越小说明模型的校正效果越好, RMSEP 值越小模型预测效果越好。一般情况下, 如果没有特别指明, R^2 一般为校正集的决定系数。

决定系数计算公式:

$$R^2 = \left(1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - y_m)^2}\right) \times 100\% \quad (6)$$

其中, y_i 为第 i 个样品的实测值, $\hat{y}_i$ 为第 i 个样品的预测值, y_m 为样品实测值的平均值。

RMSECV 和 RMSEP 计算公式:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (7)$$

其中, y_i 为第 i 个样品的实测值, $\hat{y}_i$ 为第 i 个样

品的预测值, n 为校正集或预测集的样品数。当 $\hat{y}_i$ 为校正集的预测结果时, (7) 式为 RMSECV; 当 $\hat{y}_i$ 为预测集的预测结果时, (7) 式则为 RMSEP。在实际应用中 RMSECV 常用来确定模型的维数, RMSEP 则用来对模型的实际预测效果进行评价。

1.2.4 化学指标测定

稻谷的水分含量按 GB/T 21305—2007 方法进行测定。

稻谷蛋白含量按 GB/T 5511—2008 方法进行测定。

稻谷直链淀粉按 GB/T 15683—2008 方法进行测定。

胶稠度按 GB/T 22294—2008 方法进行测定。

2 结果与分析

2.1 光谱预处理方法选择

由于近红外光谱中包含一些与待测样品性质无关的因素干扰, 如样品的状态、光的散射、杂散光及仪器响应等的影响, 使得光谱的基线发生漂移, 从而影响光谱的重复性, 因此在建模前往往需要对原始光谱进行预处理, 以消除这些不利因素对模型的影响。

本实验中采用的预处理方法是对光谱进行 Savitzky-Golay 一阶导数处理。光谱求导可以消除信号的背景, 确定光谱峰位置并改善光谱的分辨率。稻谷样品的原始光谱图如图 1 所示, 从图中可以看出原始光谱的光谱基线漂移较为严重; 经过一阶导数处理后, 光谱基线漂移得到了较好地解决。

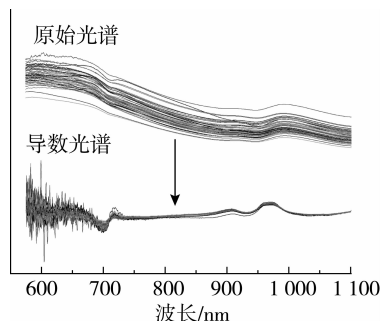


图1 稻谷样品原始光谱与一阶导数光谱对比图

在使用导数法进行预处理时, 导数平滑点数的选择十分重要, 点数过小噪声会偏大从而影响所建分析模型的预测能力; 点数过大, 平滑过度又会使得光谱失真而失去大量的有用信息。可通过不同平滑点数与模型的交互验证标准偏差(RMSECV)的变化情况来确定最佳的平滑点数。

不同平滑点数下稻谷各指标校正集模型结果见表1。

表1 不同平滑点数下稻谷常规化学指标校正集模型 RMSECV 值

参数	平滑点数	RMSECV/(%,胶稠度为mm)
水分	0 ^a	0.469 7
	3	0.398 4
	5	0.372 3
	7	0.387 7
	0	3.047 8
直链淀粉	3	1.610 1
	5	1.456 9
	7	1.663 3
	0	0.747 4
蛋白质	3	0.334 3
	5	0.305 4
	7	0.321 1
	0	8.782 2
胶稠度	3	5.111 9
	5	5.031 5
	7	5.086 4

注:a.“0”表示没有进行预处理。

由表1可以看出,经过预处理后,模型的 RMSECV 明显降低,且均在5点平滑时模型 RMSECV 值最小,故本实验采用一阶导数+5点平滑的预处理方法。

2.2 主成分数的确定

在确定模型的最佳预处理方法后,还需要确定模型的最佳主成分数。模型的主成分数过小,模型的拟合不足,交互验证和预测效果都不佳;主成分数过大,模型容易过拟合,出现交互验证结果较好但预测效果却不理想的情况。各指标模型在不同维数下的 RMSECV 值由图2所示。

从图中可以看出,水分、直链淀粉、蛋白质和胶稠度校正集模型的最佳主成分数分别为2、5、9和5。模型在达到最佳主成分数后,RMSECV 值不再明显降低。

模型在最佳预处理方法和最佳主成分数下的校正集建模结果如表2所示:

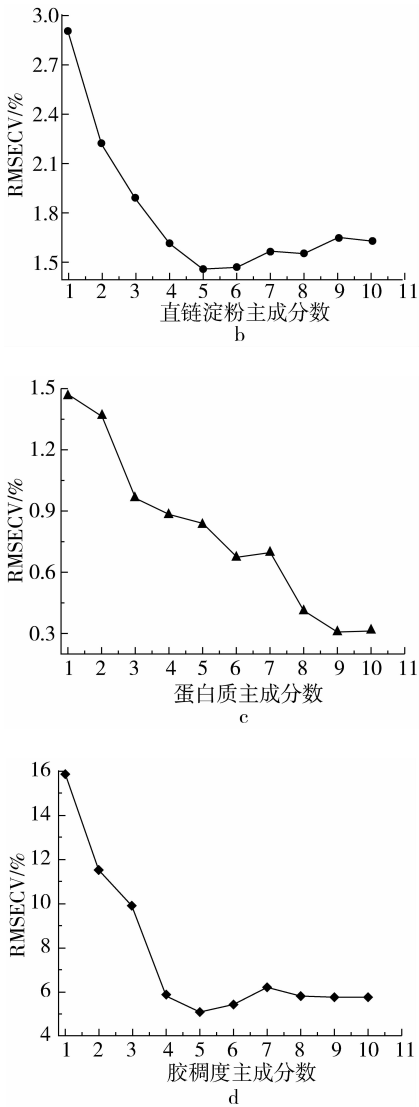
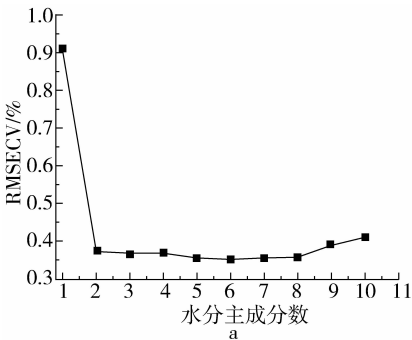


图2 各指标模型 RMSECV vs 主成分数示意图
表2 稻谷各指标校正集建模结果

参数	主成分数	RMSECV/(%,胶稠度为mm)	R ²
水分	2	0.372 8	0.990 3
直链淀粉	5	1.456 9	0.560 3
蛋白质	9	0.305 4	0.913 2
胶稠度	5	5.031 5	0.678 0

由图2和表2的结果表明,水分和蛋白质模型具有良好的自我预测效果,R²达到了90%以上,RMSECV 值也较小。而直链淀粉和胶稠度这两个指标的模型 R²较低,说明模型的校正偏差相对较大,这主要与其化学值测定方法的稳定性等因素有关。

为进一步检验模型的实际预测能力,采用所建立的定量分析模型对30个预测集样品进行预测,水分、直链淀粉、蛋白质和胶稠度模型的 RMSEP 分别为0.382 5%、1.465 0%、0.510 0%和5.052 1mm,相关预测结果如表3和图3所示。

表 3 稻谷近红外分析模型预测结果统计结果

样品 编号	水分/%		直链淀粉/%		蛋白质/%		胶稠度/mm	
	实测值	预测值	实测值	预测值	实测值	预测值	实测值	预测值
01	14.93	15.21	17.10	16.39	10.30	9.94	72.0	75.8
02	10.22	9.72	16.60	16.03	9.10	10.01	82.0	84.3
03	19.59	20.02	16.90	15.15	10.80	11.12	61.0	62.8
04	19.21	19.57	16.80	15.43	10.10	10.35	80.0	87.9
06	19.32	19.80	24.20	24.50	11.20	10.60	78.0	86.0
06	19.11	19.33	25.10	22.90	11.20	10.54	82.0	76.9
07	15.60	15.71	22.60	20.70	9.40	9.67	52.0	52.7
08	15.21	15.35	22.20	20.66	10.90	10.94	70.0	70.9
09	15.19	15.01	22.90	22.10	10.20	10.38	58.0	61.6
10	14.85	14.83	21.40	19.58	10.00	10.77	86.0	82.6
11	13.63	13.42	20.30	18.50	10.20	10.70	65.0	68.6
12	13.61	13.27	22.10	20.39	11.00	10.69	64.0	66.6
13	21.95	21.96	18.70	17.07	10.20	10.52	66.0	67.5
14	21.57	21.11	20.80	19.40	10.40	10.78	58.0	62.9
15	21.39	21.19	14.40	15.50	12.20	11.56	66.0	58.7
16	21.60	21.36	17.90	18.67	11.40	11.47	68.0	74.3
17	21.58	21.25	14.70	16.22	12.60	12.34	64.0	64.7
18	21.02	22.31	15.00	15.71	12.20	11.32	72.0	66.0
19	14.82	14.91	20.30	18.32	12.70	12.48	58.0	63.9
20	14.03	14.06	17.60	17.71	12.20	12.25	64.0	69.2
21	20.66	21.19	15.10	16.76	11.70	11.00	72.0	73.4
22	20.78	20.96	16.60	17.92	12.50	11.99	56.0	55.7
23	20.77	20.72	22.40	22.60	12.10	12.44	75.0	71.6
24	16.82	16.82	23.50	23.10	11.30	11.46	81.0	74.5
25	17.07	17.45	23.00	21.30	9.50	10.13	84.0	77.0
26	16.68	16.49	27.70	26.00	9.40	9.57	80.0	84.5
27	16.72	16.66	16.40	14.63	9.70	10.29	64.0	71.3
28	14.42	14.18	19.10	17.20	8.35	9.27	86.0	78.8
29	27.31	26.88	17.50	16.02	9.60	9.73	76.0	83.9
30	27.29	26.72	15.70	17.85	11.50	11.63	68.0	73.8

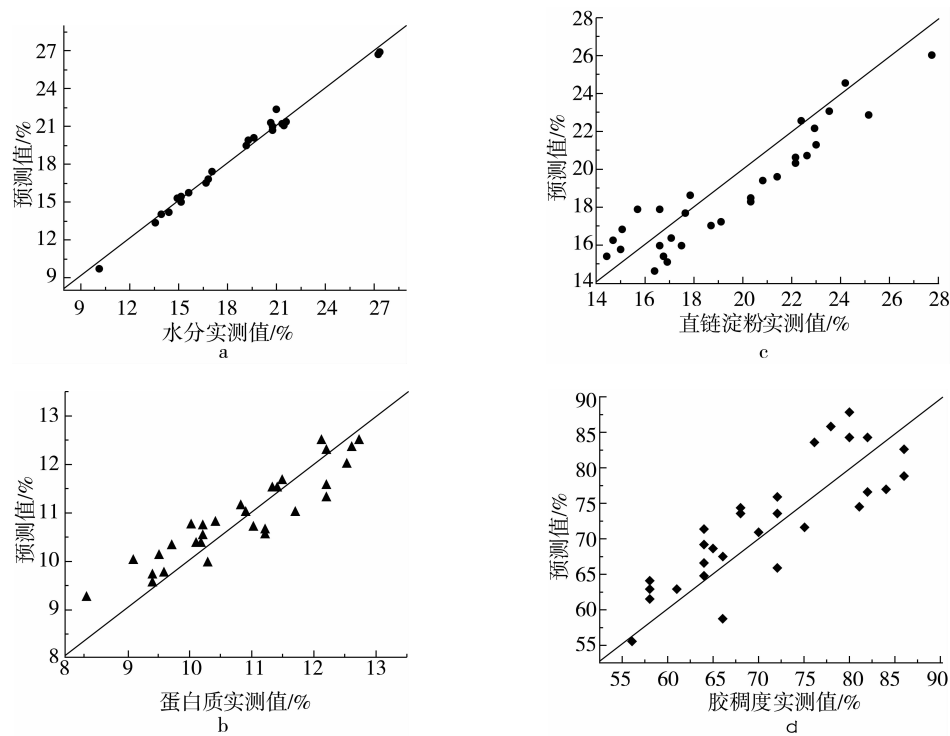


图 3 稻谷近红外各指标模型预测值 vs 实测值关系图

从表3和图3可以看出水分和蛋白质两个指标模型的预测结果比较理想, RMSEP 分别为 0.382 5% 和 0.510 0%, 能够满足相关指标分析的要求。与校正集结果类似, 直链淀粉和胶稠度这两个指标预测偏差则相对较大, RMSEP 分别为 1.465 0% 和 5.052 1mm。尽管直链淀粉和胶稠度近红外分析模型的预测效果不如水分和蛋白质的理想;但从结果上来看, 采用近红外光谱分析法能对这两个指标的预测进行大致的预测, 预测结果仍具有一定的参考价值, 可用于粮食收购环节的快速检测。

3 结论

采用近红外光谱谷物分析仪可以同时同时对稻谷常规化学指标进行预测。水分、直链淀粉、蛋白质和胶稠度模型的 RMSEP 分别达到了 0.382 5%、1.465 0%、0.510 0% 和 5.052 1mm, 基本满足了水稻常规化学指标析的要求。由于该方法无需稻谷去壳等前处理操作, 且仪器易用、效果良好, 适合进一步的应用和推广。

参考文献:

[1] 丁文平, 丁霄霖. 大米直链淀粉的研究进展[J]. 粮食与饲料工

业, 2002(1):38-40.

[2] 陈静静, 孙志高. 大米蛋白的研究进展[J]. 粮油食品科技, 2008(6):8-10.

[3] 姜松, 刘锦伟, 朱婷, 等. 籼米胶稠度测定方法的研究[J]. 中国粮油学报. 2012(11):87-91.

[4] 王玉霞, 徐荣荣, 任广鑫, 等. 绿茶茶汤中主要品质成分近红外定量分析模型的建立[J]. 茶叶科学, 2011, 31(4):355-361.

[5] Reboucas M V, Batista dos Santos J, Domingos D, et al. Near-infrared spectroscopic prediction of chemical composition of a series of petrochemical process streams for aromatics production [J]. Vib Spectrosc, 2010, 52(1):97-102.

[6] 叶旭君, 钱琼秋, 何勇, 等. 基于紫外—可见—近红外光谱技术的蔬菜细胞 ATP 含量无损检测研究[J]. 光谱学与光谱分析, 2012, 32(4):978-981.

[7] 欧阳爱国, 刘军. 基于近红外与拉曼光谱的甲醇汽油定量分析[J]. 西南师范大学学报(自然科学版), 2012, 37(9):98-100.

[8] 范维燕, 邢邯. 近红外光谱快速测定稻谷水分含量的研究[J]. 粮油食品科技. 2008, 16(5):49-52.

[9] 徐彦, 李忠海, 付湘晋, 等. 近红外光谱技术在稻米品质快速检测中的应用[J]. 食品与机械. 2011(1):158-161.

[10] 李君霞, 闵顺耕, 张洪亮, 等. 水稻糙米粗蛋白近红外光谱定量分析模型的优化研究[J]. 光谱学与光谱分析. 2006(5):833-837. 完

《粮食加工》2017 年征订启事

《粮食加工》杂志创刊于 1976 年, 1992 年为中文核心期刊, 被《中国知网》数据库长年收录, 并以其学术性与实用性结合、内容详实、信息量丰富、发行面广、印刷精美, 在全国粮食行业具有很大的影响力。刊物主要栏目有粮食经济论坛、小麦加工、稻米加工、玉米及小杂粮加工、粮食加工设备、粮食深加工、食品研究与开发、粮食物流与仓储及粮情检测分析等。

《粮食加工》为国内外公开发行人, 双月刊(逢双月 1 日出版), 8 元/期, 全年定价:48 元。国际标准 A4 开本, 全国各地邮局均可订阅。邮发代号:52-202, 国外发行代号:BM2990。

地 址: 西安市莲湖区劳动路 138 号《粮食加工》杂志社

联系人: 肖 锋

电 话: 029-88648175 传 真: 029-88631191

邮 编: 710082

E-mail: xibu98@sina.com lsjg2004@126.com

网 址: Http://www.lsjg.cn