

基于 NIRS 法快速检测荞麦蛋白质与淀粉含量的研究

郭慧敏^{1,2}, 杨修仕¹, 刘三才¹, 李进才², 任贵兴¹

(¹中国农业科学院作物科学研究所, 北京 100081; ²天津大学化工学院, 天津 300350)

摘要:为建立近红外光谱技术测定荞麦蛋白质与淀粉含量的方法,本研究以 217 份荞麦样品为试验材料,采用最小二乘回归预测和交叉验证构建近红外预测模型。分析表明:前处理采用多元散射校正法(MSC),维数(Rank)分别为 5 和 5,光谱区间 6803.9~6094.2/cm 所建立的荞麦蛋白质与淀粉含量模型的预测效果较好,其决定系数(R^2)分别为 0.9481 和 0.9167,交叉验证均方根(RMSECV)分别为 0.68 和 2.08,相对分析误差(RPD)分别为 4.39 和 3.46,均大于 3.0,外部验证相关系数均大于 0.96。本试验所建立的蛋白质与淀粉含量近红外预测模型具有较高的准确度和稳健性,可用于荞麦品质的快速测定。

关键词:荞麦;近红外光谱法;蛋白质;淀粉

Rapid Determination of Protein and Starch in Buckwheat by Near-Infrared Reflectance Spectroscopy

GUO Hui-min^{1,2}, YANG Xiu-shi¹, LIU San-cai¹, LI Jin-cai², REN Gui-xing¹

(¹Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081;

²School of Chemical Engineering and Technology, Tianjin University, Tianjin 300350)

Abstract: To establish a rapid determination method for the contents of protein and starch in buckwheat by applying Near-Infrared Reflectance Spectroscopy (NIRS), NIR spectra of 217 buckwheat samples were collected. Calibration models were established based on the spectra and chemical values by partial least-squares (PLS) algorithm with cross-validation. Analysis indicated that the models of protein and starch showed good prediction effect by using multiplicative scattering correction (MSC), with rank of 5 and 5, respectively, and frequency regions of 6803.9 – 6094.2/cm. The R^2 of them were 0.9481 and 0.9167, RMSECV were 0.68 and 2.08, RPD were 4.39 and 3.46, respectively, with the coefficients in external validation greater than 0.96. This study has demonstrated the feasibility and reliability of using NIRS to rapidly determine the content of protein and starch in buckwheat.

Key words: buckwheat; NIRS; protein; starch

荞麦栽培历史悠久,有甜荞和苦荞两个栽培种,在我国,主要分布在四川、云南、山西、内蒙古等省(自治区)。荞麦淀粉含量高达 70% 左右,且蛋白质含量高于小麦、大米等主要粮食作物,具有较高的营养价值^[1-5]。

近红外光谱技术被广泛应用于农作物品质检测

中,已成为小麦、玉米、大豆等农作物品质分析的重要手段^[6-14]。20 世纪 60 年代初,美国 K. H. Norris 等^[15]首先致力于近红外光谱技术在农产品品质分析中的应用;随着该技术不断发展,许多以近红外光谱分析技术为基础的实验方法已成为 AOCA 的标准方法^[16]。近红外技术在我国农作物品质测定中的应用

收稿日期:2016-09-20 修回日期:2016-10-26 网络出版日期:2017-04-14

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20170414.1614.004.html>

基金项目:国家科技支撑计划课题(2014BAD07B04);公益性行业(农业)科研专项(201303069);中国农业科学院科技创新工程;特色杂粮种质品质和功能成分分析(2013BAD01B05-2-5)

第一作者研究方向为 作物功能成分研究与加工。E-mail: guohuimin0208@sina.com

通信作者:任贵兴,研究方向为作物功能成分研究与利用。E-mail: renguixing@caas.cn

始于20世纪80年代^[17]。朱志华等^[18]报道了利用滤光片式近红外漫反射光谱分析仪测定我国主要粮食作物,共计7240份作物种质资源主要营养品质的进展。张晔晖等^[19]运用计算机技术与化学计量学方法,构建了完整油菜籽3种品质性状的数学模型。目前,近红外分析技术已成为我国粮油检验的标准方法^[20]。将近红外分析技术应用于农作物品质指标的测定,不仅可以克服传统化学分析方法操作繁琐、费时费力的缺陷,且可实现多组分同步测定,方便快捷。

近红外分析技术用于测定荞麦品质已有部分研究报道。王凤花等^[21]对荞麦淀粉、蛋白质与总黄酮组分进行了近红外预测模型研究,韩雍^[22]利用近红外光谱分析技术分别测定其芦丁与蛋白质含量,但上述研究存在样品数偏少、维数(主成分数)偏大,

所构建的模型在应用上存在一定的局限性。本文利用傅里叶近红外漫反射光谱技术测定我国217份荞麦样品的近红外光谱,采用偏最小二乘回归法对光谱数据进行分析,交叉验证构建荞麦蛋白质与淀粉含量近红外分析模型,为实现荞麦蛋白质与淀粉含量的快速近红外检测提供参考。

1 材料与方法

1.1 材料

试验所用材料为2014年、2015年我国山西、四川、云南、内蒙古等地区种植的217份荞麦样品,样品类型、年份和产地信息见表1。样品分析前进行脱壳、磨粉,过60目筛,于4℃保存备用。检测前于常温环境中平衡水分7d。

表1 荞麦样品的类型、年份和产地

Table 1 Type, year and origin of buckwheat samples

(份)

| 类型 Type | 年份(年) Year | 山西 Shanxi | 四川 Sichuan | 云南 Yunnan | 内蒙古 Inner Mongolia | 合计 Total |
|-------------------------|---------------|--------------|---------------|--------------|-----------------------|-------------|
| 甜荞 Common buckwheat | 2014 | 20 | 12 | 10 | 14 | 56 |
| | 2015 | 8 | 8 | 10 | 7 | 33 |
| 苦荞 Tartary buckwheat | 2014 | 22 | 22 | 10 | 16 | 70 |
| | 2015 | 16 | 20 | 14 | 8 | 58 |

1.2 主要仪器

Cyclotec 旋风式样品磨;2300型全自动定氮仪;WZZ-2S自动旋光仪;MATRIX-I傅里叶变换近红外光谱仪。

1.3 蛋白质与淀粉含量化学值测定

荞麦蛋白质与淀粉含量参照国家标准 GB/T 5511-2008 和 GB 5006-1985 测定。

1.4 近红外光谱的采集

取20g左右荞麦粉样品盛于直径50mm的旋转样品池中,谱区范围15000~3300/cm,分辨率16/cm,扫描64次。每份样品扫描2次,采集各样品的近红外光谱,将两次扫描的光谱用光谱处理软件合并成平均光谱。217份荞麦样品近红外光谱叠加图见图1。

1.5 NIRS 数学模型建立与分析

利用光谱分析软件 OPUS/QUANT 进行光谱数据的预处理,采用偏最小二乘法对校正集样品进行交叉验证建立校正模型,并用检验集样品进行外部验证,采用 SPSS 软件对其预测值与化学值进行显著性差异分析,评价模型的预测效果。

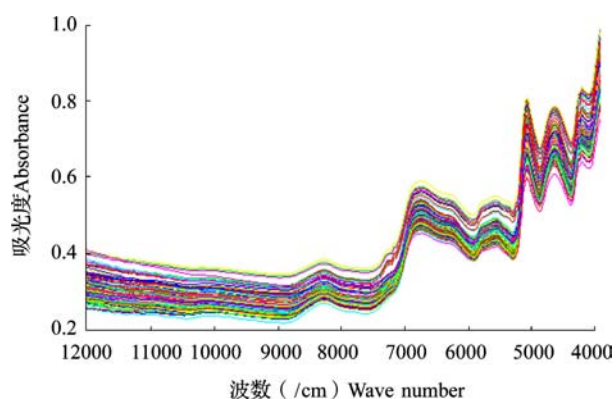


图1 217份荞麦样品近红外漫反射光谱叠加图

Fig. 1 NIRS spectra of 217 buckwheat samples

2 结果与分析

2.1 蛋白质与淀粉含量的化学测定值

采用化学方法所测得样品中蛋白质与淀粉含量的平均值、最大值和最小值见表2。蛋白质含量范围为7.73%~20.31%,淀粉含量范围为52.56%~83.24%,变幅较大,具有较好的样品代表性。经验

证,剔除 20 份异常样品,剩余样品中随机均匀选取 161 份样品为校正集,其余 36 份为检验集。校正集

和检验集样品中蛋白质与淀粉含量的统计结果见表 2。

表 2 荞麦蛋白质与淀粉含量化学测定结果统计

Table 2 Contents of protein and starch in buckwheat using chemical method

| 组别 Set | 样品数 No. | 蛋白质含量(%) Protein content | | | 淀粉含量(%) Starch content | | |
|---------------------|------------|--------------------------|-------------|-------------|------------------------|-------------|-------------|
| | | 平均值 Mean | 最大值 Max. | 最小值 Min. | 平均值 Mean | 最大值 Max. | 最小值 Min. |
| 全部样品 Samples | 197 | 12.16 | 20.31 | 7.73 | 69.29 | 83.24 | 52.56 |
| 校正集 Calibration set | 161 | 12.18 | 20.31 | 7.73 | 69.16 | 83.24 | 52.83 |
| 检验集 Test set | 36 | 12.06 | 19.50 | 7.79 | 69.88 | 83.06 | 52.56 |

2.2 校正模型的建立

采用 OPUS/QUSNT 软件对蛋白质与淀粉含量同时执行自动优化程序,以交叉验证均方差(RMSECV)和维数(Rank)为综合考虑指标,选择适宜的参数组合(包括光谱区间和预处理方法)来建立校正方程。RMSECV 越小,模型的预测精度越高。维数过少,光谱信息未得到充分利用;当 Rank > 6 时,基本上会掺杂噪音、温度波动等干扰信息,维数过多,会出现过拟合现象^[23]。表 3 为荞麦样品蛋白质与淀粉含量自动优化后满足 Rank < 6 且 RMSECV 最小的统计参数组合。采用多元散射校正(MSC, multiplicative scattering correction)进行光谱预处理可较好地消除各种因素影响。近红外模型通常根据决定系数 R^2 、交叉验证均方差 RMSECV、相对分析误差 RPD 等指标进行评价。 R^2 越接近 1,说明预测值与化学值相关性越好, R^2 小于 0.66 模型一般不可用, $0.66 < R^2 < 0.83$ 模型可用样品资源筛选, $0.83 < R^2 < 0.92$ 模型可用于实际定量分析, $R^2 > 0.92$ 模型可用于质量控制等更多运用。RPD 是用于评价模型的预测能力的重要指标, $2.4 < RPD < 3.0$ 模型可行,但精确度需进一步提高, $RPD > 3.0$ 模型可用于定量分析等实际运用^[24]。

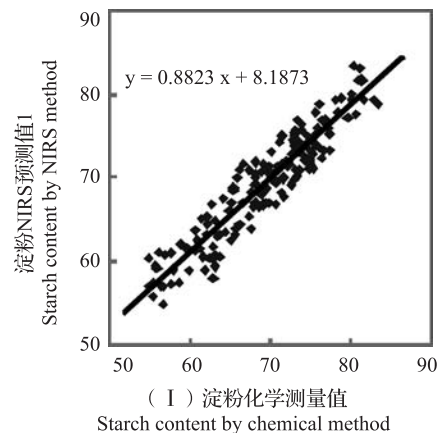
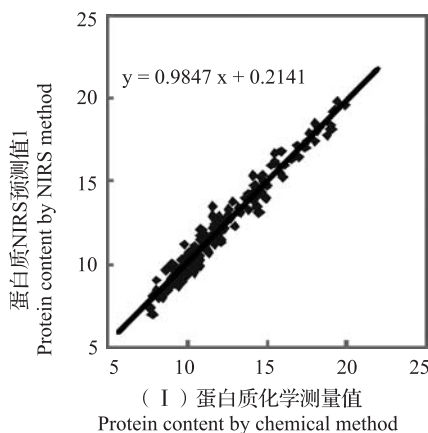
表 3 近红外光谱自动优化参数

Table 3 Parameters of NIRS optimization

| 组分 Component | 维数 Rank | 光谱区间(/cm) Frequency regions | 预处理方法 Preprocessing method |
|-----------------|------------|------------------------------------|-------------------------------|
| 蛋白质 Protein | 5 | 6101.9 ~ 5770.2 4428.0 ~ 4242.8 | MSC |
| 淀粉 Starch | 4 | 6803.9 ~ 6094.2 | MSC |

MSC: multiplicative scattering correction

采用所选择的预测模型对蛋白质与淀粉含量同时建立校正方程(图 2);其交叉验证结果见表 4。从表 4 可知,NIRS 方法所预测的蛋白质与淀粉含量的平均值与化学方法所得结果相差很小。模型 I 中蛋白质与淀粉含量的校正决定系数 R^2 分别为 0.9626 和 0.8840, RMSECV 分别为 0.58 和 2.45, RPD 均大于 2.4,说明该模型的预测分析是可行的;模型 II 中蛋白质与淀粉含量的校正决定系数分别为 0.9481 和 0.9167, RMSECV 分别为 0.68 和 2.08, RPD 均大于 3,说明该模型的预测分析效果良好。



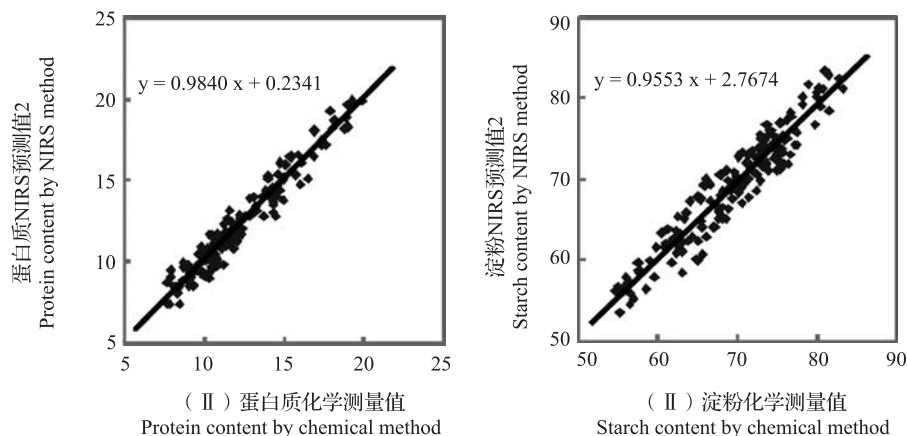


图2 交叉验证集近红外预测值与化学测量值的比较

Fig. 2 Comparison of measured values with the values predicted by the calibration model in cross validation set

表4 荞麦蛋白质与淀粉含量近红外同步校正模型的交叉验证结果

Table 4 Cross validation parameters of NIRS calibration equations developed to synchronously estimate protein and starch content in buckwheat

| 模型 Model | 组分 Component | 平均值 Mean | 最大值 Max. | 最小值 Min. | 维数 Rank | 决定系数 R ² | 交叉验证均方差 RMSECV | 相对分析 误差 RPD |
|-------------|-----------------|-------------|-------------|-------------|------------|------------------------|-------------------|----------------|
| I | 蛋白质 | 12.17 | 19.85 | 6.96 | 5 | 0.9626 | 0.58 | 5.17 |
| | 淀粉 | 69.40 | 83.44 | 54.93 | 4 | 0.8840 | 2.45 | 2.94 |
| II | 蛋白质 | 12.18 | 19.89 | 7.38 | 5 | 0.9481 | 0.68 | 4.39 |
| | 淀粉 | 69.04 | 83.28 | 53.59 | 5 | 0.9167 | 2.08 | 3.46 |

RMSECV: Root-Mean-Squares Error of Cross-Validation, RPD: Relative Percent Difference

2.3 外部检验

为进一步测试所建模型的可靠性,分别采用上述两种模型对36份检验集样品的蛋白质与淀粉含量进行预测,将模型预测平均值与其化学测量平均值(蛋白质与淀粉含量分别为12.06%与69.88%)进行比较,如表5所示。

表5 近红外模型对检验集样品预测效果评价

Table 5 Estimation and evaluation of the NIRS calibration equations on test set

| 组分 Component | 预测平均值 (%) Mean | 绝对偏差范围 Absolute deviation | 相关系数 r^2 | t 检验 t -test |
|-----------------|-------------------|------------------------------|---------------|---------------------|
| 蛋白质 Protein | I 11.94 | -1.04 ~ 1.21 | 0.98 | 0.16 |
| | II 11.98 | -0.88 ~ 1.37 | 0.98 | 0.42 |
| 淀粉 Starch | I 70.03 | -8.62 ~ 5.48 | 0.94 | 0.75 |
| | II 69.64 | -4.43 ~ 5.15 | 0.96 | 0.56 |

结果表明,2个模型对蛋白质与淀粉含量的预测值和化学测量值之间相关系数为0.94~0.98,同时 t 检验结果显示两者之间的差异无统计学意义($P > 0.05$)。由此可见2组参数所建模型均具有较好的预测效果,可以用于荞麦蛋白质与淀粉含量的同步定量测定。相比较而言,2个模型对荞麦蛋白质含量的预测值与化学测量值相关性均达0.98以上,其准确性高于对淀粉含量的预测。在对淀粉含量的预测方面,模型I的预测结果绝对偏差范围较大,预测效果需要进一步优化。

3 讨论

3.1 荞麦蛋白质与淀粉含量 NIRS 模型评价

本试验以荞麦粉末为研究对象,采集样品近红外光谱数据,并利用QUANT软件建立定量分析模型。从校正模型的预测参数 R^2 、RMSECV、RPD以及外部验证结果来看,荞麦样品中蛋白质与淀粉含

量都得到较好的预测效果。模型 I 为荞麦蛋白质近红外预测的最佳模型, R^2 为 0.9626, RMSECV 为 0.58, RPD 为 5.17; 模型 II 为荞麦淀粉近红外预测的最佳模型, R^2 为 0.9167, RMSECV 为 2.08, RPD 为 3.46。近红外光谱分析对荞麦蛋白质的预测准确性比淀粉含量的预测较高。同时,本研究在 2 个模型基础上,建立了 2 组分同步分析近红外模型,其中模型 II 对蛋白质与淀粉含量的交叉验证 R^2 均大于 0.90, RPD 均大于 3.0, 外部检验绝对偏差相对小, 相关系数高, 为最优同步分析近红外模型。

3.2 荞麦 NIRS 模型自动优化参数比较

韩雍^[22]研究得出荞麦蛋白质的近红外优化光谱区间为 9000 ~ 4000/cm, 比本试验中光谱区间要广; V. Roussel 等^[25]报道淀粉的特征光谱区间为 6896 ~ 4273/cm (1450 ~ 2340 nm), 与本试验中淀粉的光谱区间 (6803 ~ 6094/cm) 有部分重叠, 均符合 C-H、N-H、O-H 的一倍频的光谱区间 (6840 ~ 5384/cm)^[26]。本研究以 MSC 进行光谱预处理, 与韩雍^[22]甜荞蛋白质光谱预处理方法一致, 苦荞蛋白质则以一阶导数 + 矢量归一化进行光谱预处理。构建近红外模型时, 样品物理状态对光谱有较大影响, MSC 主要消除颗粒大小及颗粒分布不均匀产生的散射影响, 猜测本试验中样品的颗粒状态对光谱有一定影响。

3.3 荞麦异常材料剔除

异常材料的剔除是构建近红外预测模型中常见的问题^[27]。本研究剔除了 20 份异常材料, 占试验材料的 11.8%, 与其他主要粮食作物同类研究相比还是偏高。分析其原因可能与荞麦子粒的结构有关。荞麦, 特别是苦荞子粒种皮占种子重量比较大, 种皮坚硬, 且包裹较紧密, 不易脱壳, 一些品种要脱壳多次, 且一些碎壳混在粉中无法分离出来, 可能会干扰试验。尽管多元校正散射法能减少这种干扰, 但不排除混杂严重的仍会成为异常材料。另一个原因可能与非蛋白氮有关。现用的蛋白质含量测定方法实际是测定的 N 含量, 再通过换算间接转换为蛋白质含量。然而, 非蛋白氮含量高的材料具有与普通材料不同的近红外光谱的反应特性, 容易成为异常材料。为了提高所构建模型的稳健性和预测效果, 应该针对形成异常材料的可能原因及解决方法开展进一步的研究。

3.4 荞麦 NIRS 模型有待进一步优化

本试验以山西、四川、云南、内蒙古 4 个种植地收集的 217 份荞麦样品为原料进行近红外光谱学分

析, 试样来源性广, 所建模型准确度高。但理想的预测模型需要大量的种质资源, 不同品种不同产地的荞麦受基因型和环境条件的影响, 品质存在差异^[28], 本研究所建模型尚不足以满足所有荞麦资源的品质分析; 且近红外光谱分析以化学测量值为基础, 其精确度对所建模型的稳定性有一定影响^[29]。今后有必要在现有基础上进一步扩大荞麦样品数量, 同时加强化学测量操作的准确性, 提高模型的质量和适用范围, 为荞麦资源品质鉴定提供准确高效的方法。

4 结论

本试验以 217 份荞麦样品为试验材料, 采用化学方法测定其蛋白质与淀粉含量, 采集样品近红外光谱数据, 并采用最小二乘回归预测和交叉验证构建近红外预测模型。所建荞麦蛋白质与淀粉含量的近红外预测模型准确度高, 稳健性好, 具有一定的应用价值, 可用于荞麦蛋白质与淀粉含量的快速测定。

参考文献

- [1] 林汝法, 周小理, 任贵兴, 等. 中国荞麦的生产与贸易、营养与食品[J]. 食品科学, 2005, 26(1): 259-263
- [2] 王红育, 李颖. 荞麦的研究现状及应用前景[J]. 食品科学, 2004, 25(10): 388-391
- [3] Ikeda K. Buckwheat composition, chemistry, and processing[J]. Advan Food Nut Res, 2002, 44(44): 395-434
- [4] 章华伟, 刘邻渭. 荞麦淀粉研究进展[J]. 粮食与油脂, 2002(7): 32-33
- [5] 杜双奎, 李志西, 于修焯. 荞麦蛋白研究进展[J]. 食品科学, 2004, 25(10): 409-414
- [6] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005: 1-9
- [7] Bellon V, Vigneau J L, Sévila F. Infrared and near-infrared technology for the food industry and agricultural uses; Online applications[J]. Food Control, 1994, 5: 21-27
- [8] 李伟, 肖爱平, 冷鹃. 近红外光谱技术及其在农作物中的应用[J]. 中国农学通报, 2009, 25(3): 56-59
- [9] 金华丽, 许春红, 徐泽林. 近红外光谱法测定小麦籽粒中的蛋白含量[J]. 河南工业大学学报: 自然科学版, 2010, 31(6): 21-24
- [10] 魏良明, 姜海鹰, 李军会, 等. 玉米杂交种品质性状的近红外光谱分析技术研究[J]. 光谱学与光谱分析, 2005, 25(9): 1404-1407
- [11] 孙君明, 韩粉霞, 闫淑荣, 等. 傅里叶近红外反射光谱法快速测定大豆脂肪酸含量[J]. 光谱学与光谱分析, 2008, 28(6): 1290-1295
- [12] Gamsworthy P C, Wiseman J, Fegeros K. Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy[J]. J Agricul Sci, 2000, 135(4): 409-417
- [13] Kovalenko I V, Rippke G R, Hurburgh C R. Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy[J]. J Agr Food Chem, 2006, 54(10): 3485-3491
- [14] Sohn M, Barton F E, McClung A M, et al. Near-infrared spectroscopy for determination of protein and amylose in rice flour through use of derivatives[J]. Cereal Chem, 2004, 81: 341-344