

# 基于近红外光谱荞麦淀粉、蛋白质和总黄酮含量测定方法研究

王凤花<sup>1</sup>, 朱海龙<sup>2</sup>, 杨菊<sup>1</sup>, 席志勇<sup>1</sup>

(1. 昆明理工大学现代农业工程学院, 云南昆明 650500;

2. 昆明理工大学工程训练中心, 云南昆明 650500)

**摘要:** 淀粉、蛋白质和总黄酮是荞麦中重要的营养成分和药理保健功能因子, 实现对荞麦内部成分的快速检测对荞麦资源开发、育种和加工等环节具有重要的理论研究意义和应用价值。选用来自不同产地的 40 个荞麦样品, 测定其淀粉、蛋白质和总黄酮含量, 采用 FieldSpec 3 光谱仪采集了其光谱信息。将 40 个样本随机分成了预测集和验证集, 其中预测集 30 个, 验证集 10 个。通过比较不同预处理方法, 利用主成分分析法和人工神经网络相结合的方法分别建立了荞麦淀粉、蛋白质和总黄酮含量的预测模型, 并进行了验证。结果表明, 首先经平滑与多元散射校正预处理, 再利用主成分分析法和人工神经网络相结合的方法分别建立的荞麦淀粉含量的预测模型预测值和测量值相关性高, 相关度为 0.8719, 相对误差小, 为 1.82%, 说明该方法预测荞麦淀粉含量是可行的。但对荞麦蛋白质和总黄酮含量的预测并不理想, 还有待进一步研究。

**关键词:** 近红外光谱, 荞麦淀粉, 蛋白质, 总黄酮, 人工神经网络

## Research on determination method of starch, protein and total flavonoids content in buckwheat by Near-infrared spectroscopy

WANG Feng-hua<sup>1</sup>, ZHU Hai-long<sup>2</sup>, YANG Ju<sup>1</sup>, XI Zhi-yong<sup>1</sup>

(1. Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China;

2. Engineering Training Center, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** Starch, proteins and total flavonoids were the vital nutrients and pharmacological function factor in buckwheat, to achieve the fast detection of the buckwheat internal composition had an important theoretical significance and application value for buckwheat breeding, processing and other steps. Forty buckwheat samples from different origins had been selected. The starch, proteins and total flavonoids content of buckwheat were determined, and the near-infrared spectrum of buckwheat had been obtained using FieldSpec 3 near-infrared spectroscopy. Forty samples were randomly divided into the prediction set and validation set with 30 and 10 samples respectively. Compared with different pre-processing methods, the prediction models of buckwheat starch proteins and total flavonoids content had been established using the combination method of principal component analysis and artificial neural network, finally the models had been verified. The results showed that the correlation coefficient between the prediction value and measurement value of buckwheat starch content was 0.8719, and the relative error was smaller and its mean value was 1.82%, the method of the buckwheat starch content prediction was feasible. But the prediction for buckwheat protein and total flavonoids content was not ideal, need to be further studied.

**Key words:** Near-infrared spectroscopy; buckwheat starch; protein content; total flavonoids; BP Neural Network

中图分类号: TS207

文献标识码: A

文章编号: 1002-0306(2014)05-0281-04

荞麦又称为三角麦、乌麦、花荞, 属于双子叶蓼

科, 荞麦属, 主要分布在黄土高原高寒山区和云贵川高原山地, 具有较高的营养和保健价值, 被称为“食药两用”的粮食珍品<sup>[1]</sup>。荞麦营养元素种类丰富, 含量平衡, 荞麦中淀粉和蛋白质含量的高低, 是评价荞麦营养品质的重要指标<sup>[2]</sup>。总黄酮是荞麦中重要的药理、保健功能因子之一, 具有降血糖、降血脂、降尿酸、防便秘等功效<sup>[3]</sup>。因此, 大力发展荞麦种植产业及荞麦的精深加工技术, 提高荞麦的品

收稿日期: 2013-07-11

作者简介: 王凤花(1980-), 女, 博士, 讲师, 研究方向: 农产品品质无损检测方法的研究。

基金项目: 国家自然科学基金天元基金(11226220); 云南省应用基础面上研究项目(2010ZC028); 昆明理工大学分析测试基金(2011432, 2011436)。

表1 荞麦化学组分的统计参数

Table 1 Buckwheat chemical constituent's statistical parameters

	样本数	最小值	最大值	平均值	标准差	变异系数
淀粉含量(%)	40	57.2	72.4	65.5	4.4	6.7
蛋白质含量(%)	40	8.5	12.0	9.9	1.0	10.1
总黄酮(mg/100g)	40	7.1	163.5	82.9	42.2	50.9

质,对于增加农民经济收入具有十分重要的意义。

目前荞麦淀粉、蛋白质和总黄酮含量的测量仍然采用传统的化学分析方法,其具有测量精度较高的优点,但存在检测速度慢、测量成本高,需在专门的检测机构或实验室中检测等不足。近红外光谱具有分析速度快、多组分同时测定、样品不需预处理、非破坏性分析、实时分析、低分析成本以及操作简单等优点,与化学计量学相结合可实现样品组分快速无损的检测。目前近红外光谱技术已广泛应用于农产品品质检测中<sup>[4-8]</sup>,且取得了较好的分析结果。但国内外关于应用近红外光谱技术对荞麦内部成分的相关研究鲜有报道。因此,本文选取不同种植地区的8种荞麦,利用近红外光谱技术采集荞麦的光谱信息,研究光谱信息和荞麦内部品质之间的定量关系模型,建立一种快速、无损、稳定、简便的荞麦品质的评定方法,对荞麦育种、加工、检测等多个环节具有很好的理论研究意义和实际应用价值。

## 1 材料与方法

### 1.1 材料与仪器

**荞麦样品** 选用云南、山西、四川不同荞麦种植地区的7种苦荞籽粒和一种甜荞籽粒,分别是昆明本地荞、泸西苦荞、昭苦1号、昭苦2号、山西苦荞Y16号、西昌苦荞、大理苦荞和大理甜荞,共8种具有代表性的荞麦品种,每种荞麦选5个样品,共40个样品。每个样品经干燥、去杂后,装袋密封备用,要求粒径大小、均匀度一致。

**FieldSpec 3 光谱仪** 美国 ASD (Analytical Spectral Device) 公司;CS101-2D 型电热鼓风干燥箱

中外合资重庆四达实验仪器有限公司;FSD-100A 型电动粉碎机 杭州大吉光电仪器有限公司;JLML 磨米机 上海嘉定粮油检测仪器厂。

### 1.2 实验方法

**1.2.1 荞麦光谱特征的采集** 在对荞麦籽粒进行光谱测量时,实验装置由微机、光谱仪、采用内置光源的反射探头、校正白板等组成。光谱仪选用了美国 ASD (Analytical Spectral Device) 公司的 FieldSpec 3 光谱仪,光谱采样间隔为 1nm,采样范围为 350~2500nm,分辨率 3.5nm。将 40 个荞麦样品放入直径为 90mm,厚度为 15mm 的玻璃培养皿中,将内置光源的反射探头紧贴样品上表面,采用反射方式进行样品光谱采集。对每个样品做 5 个重复,每个样本扫描 10 次,然后采用 ASD 公司提供的 view Spec pro 软件进行平均处理,并将其按  $\log[1/R]$  转换为吸光度。然后将光谱数据以 ASCII 码形式导出并用 ASD View Spec Pro, Unscramble V9.7 和 DPS

(Data Procession System For Practical Statistics) 软件进行处理。

**1.2.2 荞麦内部成分含量的测定** 每个样本荞麦籽粒经脱壳、粉碎后过孔径为 0.2mm 筛,要求粒径大小、均匀度一致,采用传统化学方法测定其淀粉、蛋白质和总黄酮含量。荞麦淀粉含量采用酶水解法测量,蛋白质含量采用凯氏定氮法,荞麦总黄酮含量采用高效液相色谱法<sup>[9]</sup>测量,由云南省分析测试中心完成。

## 2 结果与分析

### 2.1 荞麦光谱特征和内部成分含量的测定结果

经上述实验方法得到荞麦原始光谱曲线如图 1 所示,其化学测量结果统计参数如表 1 所示。

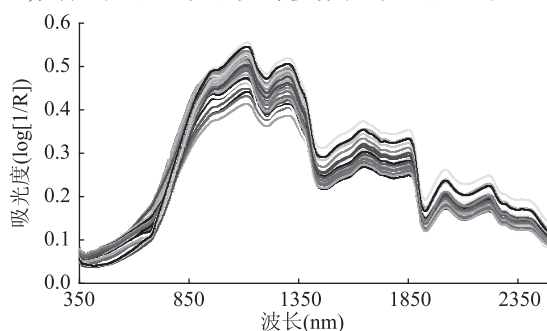


图1 荞麦近红外光谱吸光度

Fig.1 Absorbance graph of buckwheat near-infrared spectroscopy

### 2.2 数据预处理

采集的近红外光谱往往受到高频随机噪音、基线漂移、样本不均匀、光散射等影响,所以需要进行光谱预处理。经过对多种不同预处理方法的比较,最终采用平均平滑法对实验所得的荞麦原始近红外光谱进行预处理,选用平滑窗口大小为 3,再进行多元散射校正(multiplicative scatter correction,简称 MSC)处理,能够很好滤除各种因素产生的高频噪声。

### 2.3 基于主成分分析的主成分数提取

将 40 个样本随机分成预测集与验证集,预测集为 30 个,验证集为 10 个。样本光谱波段从 350~2500nm 共有 2151 个点,采用全谱段计算时,计算量大,而且由于有些区域样品的光谱信息很弱,与样品的组成或性质间缺乏相关关系。主成分分析是一种非常有效的数据挖掘方法,它可将原来多个波长变量转化为较少的新变量,并且这些新变量既彼此互不相关,又能综合反映原来多个波长变量的信息<sup>[10-11]</sup>。所以分别对预测集和验证集进行主成分分析,其主成分累计可信度见表 2。

表2 前10个主成分累计可信度

Table 2 Accumulative reliabilities of the first 10 PCs

主成分数	预测集累计可信度(%)	验证集累计可信(%)
PC01	82.467	81.467
PC02	91.286	89.787
PC03	96.865	95.940
PC04	98.924	98.540
PC05	99.378	99.052
PC06	99.61	99.342
PC07	99.756	99.501
PC08	99.841	99.677
PC09	99.892	99.734
PC10	99.935	99.851

## 2.4 基于BP神经网络的荞麦淀粉预测模型

将30个预测集由PCA提取的10个主成分的分作为BP神经网络的输入变量,输出层为荞麦淀粉含量的测量值,经依次实验得到荞麦淀粉含量的最佳隐含层节点数为6,网络设定训练迭代次数为1000次,目标误差设定为0.0001,建模样本的拟合残差为 $4.1594 \times 10^{-3}$ 。建立荞麦淀粉含量的校正模型,对10个验证集进行预测,预测值与测量值的相关分析见图2。误差分析结果见表3。

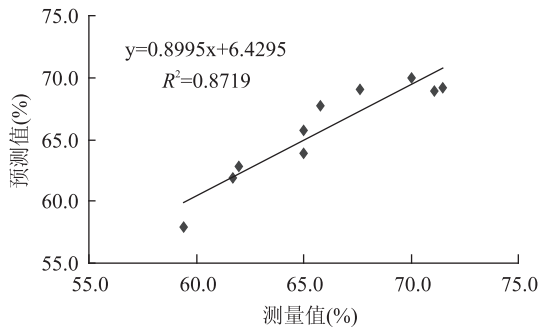


图2 荞麦淀粉含量预测值与测量值的相关图

Fig.2 Relationship between chemical analysis values and prediction values of Buckwheat starch

表3 荞麦淀粉含量BP预测模型预测结果

Table 3 Prediction results for Buckwheat starch by BP model

样本数	化学测量值(%)	预测值(%)	绝对误差(%)	相对误差(%)
1	59.4	57.91	-1.49	-2.51
2	65.0	65.71	0.71	1.09
3	71.1	68.95	-2.15	-3.03
4	67.6	69.03	1.43	2.12
5	71.5	69.24	-2.26	-3.17
6	62.0	62.82	0.82	1.31
7	61.7	61.85	0.15	0.24
8	65.8	67.78	1.98	3.01
9	65.0	63.89	-1.11	-1.70
10	70.0	69.96	-0.04	-0.06

由图2可知,荞麦淀粉含量BP神经网络对验证集10个样本的预测值与测量值相关度为0.8719,两

者相关性较高。由表3可知,荞麦淀粉含量预测值与测量值的最大绝对误差为-2.26%,相对误差绝对平均值为1.82%,相对误差绝对平均值较小,说明该方法预测荞麦淀粉含量是可行的。

## 2.5 基于BP神经网络的荞麦蛋白质含量预测模型

将30个预测集由PCA提取的10个主成分的分作为BP神经网络的输入变量,输出层为荞麦蛋白质含量的测量值,经依次实验得到蛋白质含量的最佳隐含层节点数为8,网络设定训练迭代次数为1000次,目标误差设定为0.0001,建模样本的拟合残差为 $8.0101 \times 10^{-3}$ 。建立蛋白质含量的校正模型,对10个验证集进行预测,预测值与测量值的相关分析见图3。误差分析结果见表4。

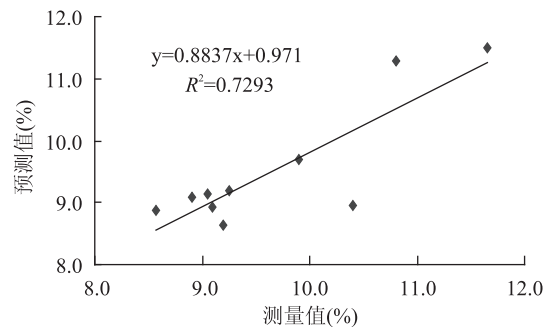


图3 荞麦蛋白质含量预测值与测量值的相关图

Fig.3 Relationship between chemical analysis values and prediction values of Buckwheat protein

表4 荞麦蛋白质含量BP预测模型预测结果

Table 4 Prediction results for Buckwheat protein by BP model

样本数	化学测量值(%)	预测值(%)	绝对误差(%)	相对误差(%)
1	10.8	11.28	0.48	4.44
2	10.4	8.95	-1.45	-13.93
3	9.2	8.64	-0.56	-6.04
4	8.9	9.10	0.20	2.21
5	9.1	9.14	0.09	0.97
6	8.6	8.87	0.29	3.35
7	9.9	9.70	-0.20	-2.00
8	9.1	8.92	-0.18	-2.02
9	9.3	9.20	-0.05	-0.57
10	11.7	11.48	-0.17	-1.42

由图3可知,荞麦蛋白质含量BP神经网络对验证集10个样本的预测值与测量值相关度为0.7293。由表4可知,荞麦蛋白质含量预测值与测量值的最大绝对误差为-1.45%,相对误差绝对平均值为3.695%,其相对误差绝对平均值较小,但其预测值与测量值的相关度较低,荞麦蛋白质的预测方法还有待进一步研究。

## 2.6 基于BP神经网络的荞麦总黄酮含量预测模型

将30个预测集由PCA提取的10个主成分的分作为BP神经网络的输入变量,输出层为荞麦总黄酮含量的测量值,经依次实验得到总黄酮含量的最佳隐含层节点数为6,网络设定训练迭代次数为1000次,目标误差设定为0.0001,建模样本的拟合残差为

$3.547 \times 10^{-3}$ 。建立总黄酮含量的校正模型,对 10 个验证集进行预测,预测值与测量值的相关分析见图 4。误差分析结果见表 5。

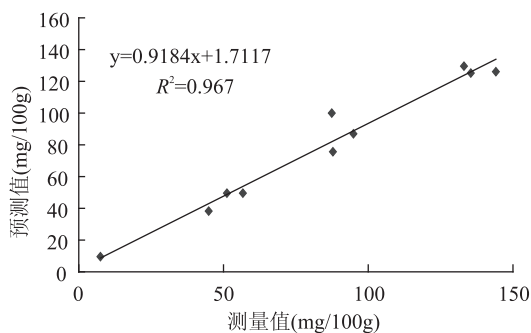


图 4 荞麦总黄酮含量预测值与测量值的相关图

Fig.4 Relationship between chemical analysis values and prediction values of Buckwheat total flavonoids

表 5 荞麦总黄酮含量 BP 预测模型预测结果

Table 5 Prediction results for Buckwheat total flavonoids by BP model

样本数	化学测量值 (mg/100g)	预测值 (mg/100g)	绝对误差 (%)	相对误差 (%)
1	144.1	126.19	-17.91	-12.43
2	87.3	100.19	12.87	14.74
3	133.0	129.78	-3.17	-2.39
4	135.4	125.40	-10.03	-7.40
5	94.9	87.04	-7.86	-8.28
6	87.8	75.84	-11.96	-13.62
7	51.1	49.15	-1.95	-3.82
8	56.6	49.60	-6.95	-12.29
9	44.9	38.15	-6.75	-15.03
10	7.4	9.57	2.20	29.77

由图 4 可知,荞麦总黄酮含量 BP 神经网络对验证集 10 个样本的预测值与测量值相关度为 0.9670。由表 5 可知,荞麦总黄酮含量预测值与测量值的最大绝对误差为 -17.91%,相对误差绝对平均值为 11.978%。其预测值与测量值的相关度较高,但其相对误差绝对平均值较大,预测效果不理想,分析主要原因为所选荞麦品种中大理甜荞的总黄酮含量较小,导致实验样本总黄酮含量变异系数为 50.9,差异太大,应优选实验样本和增加样本数量进一步优化建模方法,还需进一步的研究。

### 3 结论与讨论

3.1 文章采用主成分分析和人工神经网络方法建立了荞麦淀粉、蛋白质和总黄酮含量人工神经网络预测模型,并对其预测结果进行验证,结果表明荞麦淀粉含量人工神经网络预测模型对验证集 10 个样本

的预测相关度为 0.8719,相对误差绝对平均值为 1.82%,说明该方法预测荞麦淀粉含量是可行的。

3.2 文中建立的荞麦蛋白质含量预测模型对 10 个样本预测相关度为 0.7293,相对误差绝对平均值为 3.695%。建立的荞麦总黄酮含量预测模型对 10 个样本预测相关度为 0.9670,相对误差绝对平均值为 11.978%。研究表明,利用主成分分析和人工神经网络方法对荞麦蛋白质和总黄酮含量的预测效果不理想,分析主要原因为所选荞麦品种的蛋白质和总黄酮含量差异较大,样本数量较少,在以后的研究中应增加荞麦品种和样本数量进一步优化建模方法。本文对荞麦的营养因子蛋白质和重要药理因子总黄酮含量的快速无损检测方法的进一步研究提供了一定的基础。

### 参考文献

- [1] Kayashita J, Shimaoka I, Nakajoh M, et al. Consumption of a buckwheat protein extract retards 7, 12 - Dimethylbenz [ alpha ] anthracene - induced mammary carcinogenesis in rats. [ J ]. Bioscience Biotechnology Biochemistry, 1999, 63 ( 10 ): 1837-1839.
- [2] SiuMei Choi, Chingyung Ma. Structural characterization of globulin from common buckwheat ( Fagopyrum esculentum Moench ) using circular dichroism and Raman spectroscopy [ J ]. Food Chemistry, 2007, 102: 150-160.
- [3] 杨政水. 苦荞麦的功能特性及其开发利用 [ J ]. 食品研究与开发, 2005, 26 ( 1 ): 100-103.
- [4] 王冬, 马智宏, 潘立刚, 等. 近红外光谱法定量测定小麦粉中的石灰类添加物的研究 [ J ]. 光谱学与光谱分析, 2013, 33 ( 1 ): 69-73.
- [5] 张巧杰, 张军. 基于小波变换的大米直链淀粉波长选择方法 [ J ]. 农业机械学报, 2010, 41 ( 2 ): 138-142.
- [6] 马东红, 王锡昌, 刘利平, 等. 近红外光谱技术在食品产地溯源中的研究进展 [ J ]. 光谱学与光谱分析, 2011, 31 ( 4 ): 877-881.
- [7] 刘星, 单杨, 张欣, 等. 近红外光谱法鉴别奶牛饲料中三聚氰胺甲醛树脂的可行性研究 [ J ]. 食品科学, 2012, 33 ( 12 ): 154-158.
- [8] 席志勇, 王凤花. 谷物品质无损检测方法研究进展 [ J ]. 食品工业科技, 2012, 33 ( 15 ): 394-396, 400.
- [9] 张琪, 刘慧灵, 朱瑞, 等. 苦荞麦中总黄酮和芦丁的含量测定方法的研究 [ J ]. 食品科学, 2003, 24 ( 7 ): 113-116.
- [10] 陆婉珍. 现代近红外光谱分析技术 [ M ]. 北京: 中国石化出版社, 2006: 105-164.
- [11] 张淑娟, 王凤花, 张海红, 等. 基于主成分分析和 BP 神经网络的土壤养分近红外光谱检测 [ J ]. 山西农业大学学报, 2009, 29 ( 6 ): 483-487.