

# 近红外透射光谱法(NITS)分析大豆品质的研究<sup>\*</sup>

姚鑫森 张瑞英 李霞辉 程爱华 张晓波 高春霞

(农业部谷物及制品质量监督检验测试中心, 哈尔滨 150086)

**摘要** 为了研究大豆品质性状的快速测定方法,以我国东北部四省区 572 份大豆样品为材料,采用近红外透射(NITS)技术非破坏性测定大豆的粗蛋白质、粗脂肪含量。调选出的校正集样品经实验室常规分析测定,建立其吸收光谱与化学成分间的关系模型,校正并优化原有测定方程。校正方程经预测获得了较高的预测集决定系数 0.9757(蛋白质)、0.9549(脂肪)和较低的标准误差 2.18(蛋白质)、0.88(脂肪)。结果表明近红外透射光谱技术可广泛应用于大豆品种品质普查、大豆品质育种材料筛选和商品大豆质量分类分级。本研究试验用样品数量多、来源分布均匀、品种信息丰富,因此,所建立的近红外透射预测模型适用范围广。

**关键词** 大豆;近红外透射;校正;预测

中图分类号 S565.1 文献标识码 A 文章编号 1000-9841(2006)04-0417-04

近红外光谱(Near Infrared Spectroscopy)技术是利用有机物质在近红外光谱区的振动吸收而快速测定样品中多种化学成分含量的一项新技术<sup>[1]</sup>,它介于电磁波谱的 800~2500nm,根据其检测对象不同分为近红外反射光谱(Near Infrared Reflectance Spectroscopy, NIRS)和近红外透射光谱(Near Infrared Transmittance Spectroscopy, NITS)。NIRS 是根据反射与入射光强的比例关系来获得物质在近红外区的吸收光谱,电磁波谱范围是 1100nm~2500nm,这段谱区的近红外光在样品中的穿透一般不大于 1mm,所以非常适合均质良好的粉状样品分析。NITS 是根据透射与入射光强的比例关系来获得物质在近红外区的吸收光谱,电磁波谱范围是 850nm~1050nm,为分子振动吸收的二级和三级倍频吸收区。测量得到的吸收光谱符合比尔定律。同反射技术相比,透射谱区的近外光在样品的穿透最大可达 30mm,因此适合整粒或原状样品分析,是完全的无损分析技术<sup>[2]</sup>。

本试验使用的 FOSS 人工神经网络(ANN)定标,属于大样品数据库定标,可以提高定标测试的适用范围和定标的稳定性,减少定标模型的调整工作,使校正更加容易。

与常规分析技术不同,近红外光谱技术是一种间接分析技术,必须通过建立定标模型(校正模型)来实现对未知样品的定性和定量分析。分析过程包括:一是选择一组有代表性的样品并测量其近红外光谱;二是采用标准参考方法测定样品组分含量;三是将测定的光谱数据和化学法基础数据用适当的化学计量方法建立校正模型;四是利用校正模型对未知样品的组分进行测定<sup>[3]</sup>。

本研究采用近红外透射光谱(NITS)技术分析测定了黑龙江、内蒙、吉林、辽宁四省区 572 份大豆样品水分、粗蛋白质、粗脂肪含量,建立一套能够推广使用的近红外透射定标模型,为近红外透射光谱技术在我国大豆品质分析中的应用提供参考。

提高大豆含油量和蛋白质含量是大豆品质育种的主要目标。近红外透射光谱(NITS)技术具有样品无需预处理、无损耗、准确快速、重现性好、多成分同步分析、成本低、无试剂污染等特点,是一种高效、快速的现代分析技术。利用优化近红外透射光谱分析定标模型,可以从大豆样品的近红外光谱来快速检测大豆的粗蛋白质含量和粗脂肪含量,大大提高品质测试效率。

\* 收稿日期: 2006-02-07

课题来源: 2005 国家大豆良种推广补贴项目

作者简介: 姚鑫森(1980-),女,硕士,研究,从事谷物品质研究工作。

# 1 材料和方法

## 1.1 试验材料

材料取自 2005 年度黑龙江、吉林、辽宁、内蒙古四省大豆生产优势区采集的 572 份大豆样品。分别选出具有代表性的、含量梯度分布均匀的样品 23 份(粗蛋白质)和 25 份(粗脂肪)作为 ANN 方程的校正集样品,含量范围分别为 34.17%~44.27%(粗蛋白质),17.77%~23.61%(粗脂肪)。同时从未参与校正的样品中选出 26 份(蛋白质)、43 份(粗脂肪)作为预测集,对校正方程的性能进行评价,验证测定结果的准确性。其余样品作为检验集。

## 1.2 方法

### 1.2.1 化学分析测定方法

凯氏定氮法 NY/T 4-1982<sup>[4]</sup> 测定大豆粗蛋白含量;

索氏提取法 NY/T 3-1982<sup>[5]</sup> 测定大豆粗脂肪含量。

### 1.2.2 近红外分析

近红外透射光谱法:测定大豆粗蛋白含量和粗脂肪含量

仪器:FOSS 1241 近红外谷物分析仪

#### 1.2.2.1 光谱采集

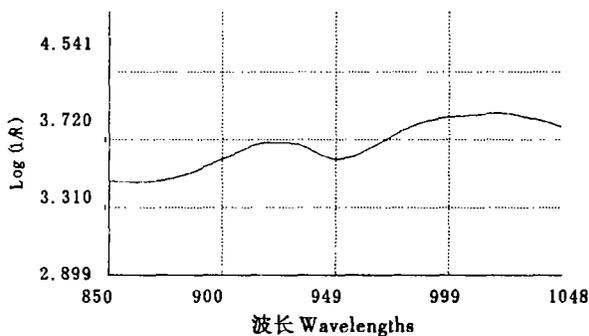


图 1 大豆籽粒样品扫描光谱图

Fig.1 Spectrum of soybean kernel

采用 FOSS1241 大豆近红外神经网络方程对全部样品进行光谱扫描。扫描温度控制在 21~25℃

之间,吸收波长为 850~1050nm,每个样品扫描两次(平行扫描一次),每次扫描设置子样品数为 10,每个子样品扫描 100 次,取其平均值。

#### 1.2.2.2 光谱散射校正

为解决光谱吸光度与样品性状间吸收非线性问题,采用标准正态变量转换法(Standard Normal Variable, SNV)结合趋势变换法(Detrend)进行预处理,之后选择导数处理(Derivative)参数,并通过改变波段 2 间隙点(Gap)和作平滑处理(Smooth)对光谱进行进一步优化。

#### 1.2.2.3 群体界定和校正集样品选择

将全部样品分为校正集、预测集和检验集。校正集和预测集样品,数量必需满足统计学要求,样品至少有 20 个左右,含量的范围也要适应检验集样品的含量范围;并且是从高到低均匀分布的<sup>[6]</sup>。

将样品的化学分析值输入定标软件 WINISI,利用 WINISI 软件计算出每份样品与平均谱带的差异,用标准化的 Mahalanobis 距离(H)表示,按 H 距离大小按比例选择,取 NH(Neighborhood H)=3.0, GH(Global H)=10.0 的样品构成校正群体。再根据交互验证(Crossing Validation)结果,剔除个别不具代表性的样品。

#### 1.2.2.4 校正方程建立

对 FOSS 神经网络方程进行斜率、截距的调整,建立大豆粗蛋白质、粗脂肪校正模型。

#### 1.2.2.5 校正方程评价指标

相关系数(r)、决定系数(R<sup>2</sup>)、标准误差(SEC/SEP)、预测值与基准法测定值残差的平均值(Bias)、残差平均方差 MSE<sup>[7]</sup>。

# 2 结果与分析

## 2.1 校正集和预测集真实值的测定

分别采用实验室常规分析方法对校正集和预测集样品进行粗蛋白质和粗脂肪含量的测定,结果见表 1。

表 1 基准方法测定结果

Table 1 Statistical data of samples collected

品质特性	分组	样品数	平均值	变幅(%)	标准误差
Quality character	Group	Sample	Mean	Range	SEC/SEP
粗蛋白质含量	校正集	23	39.66	34.17~44.27	2.83
Protein content	预测集	26	39.87	35.91~43.64	2.18
粗脂肪含量	校正集	18	20.73	17.77~23.61	1.39
Oil content	预测集	43	20.43	17.93~22.19	0.88

注:SEC,校正集标准误差;SEP,预测集标准误差

该校正集样品梯度分布比较均匀,见图2、3。

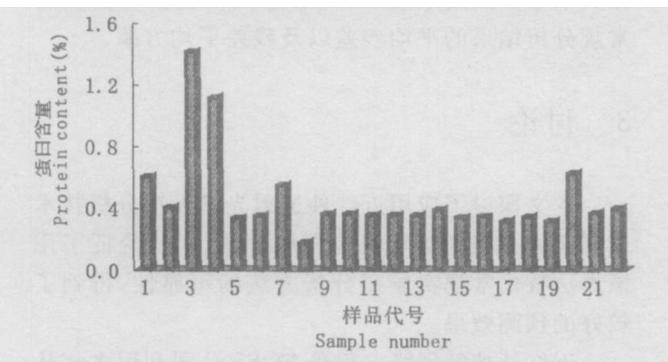


图2 校正集样品蛋白含量梯度分布示意图

Fig.2 Protein content distribution of calibration samples

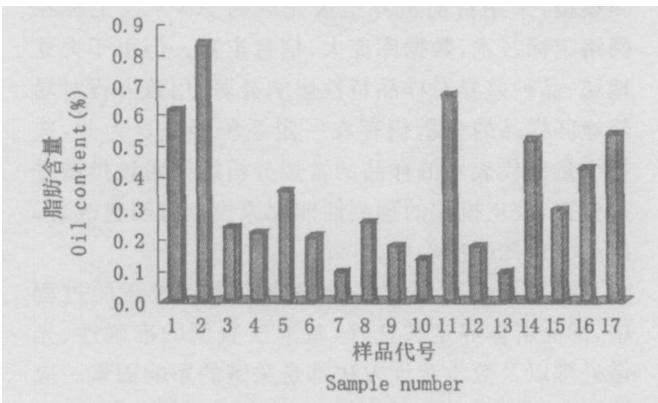


图3 校正集样品脂肪含量梯度分布示意图

Fig.3 Oil content distribution of calibration samples

## 2.2 校正方程的建立

粗蛋白质含量、粗脂肪含量采用 FOSS1241 近红外谷物品质分析仪(大豆神经网络方程 SO 090711)测定,测定结果采用实验室常规分析方法粗蛋白质含量、粗脂肪含量进行斜率和截距校正。见表2。

表2 校正方程结果

Table 2 Results of calibration

品质特性	样品数	近红外	实验室	斜率	截距
Quality character	Sample	ANL	LAB	Slope	Bias
粗蛋白质含量	23	39.87	39.66	0.9748	0.7969
Protein content					
粗脂肪含量	18	20.95	21.05	0.8643	2.9424
Oil content					

注: ANL, 近红外预测平均值; LAB, 实验室常规分析值

近红外分析结果与实验室常规分析结果的相关系数粗蛋白质、粗脂肪分别为 0.9905、0.9737, 两种方法具有很高的相关性; 从近红外值 NIR 与常规分析值的相关性图中也可以看出, 二者线性关系良好, 具有同等的准确性和测定精度; 且近红外结果的标准偏差均小于常规法结果的标准偏差, 且经过 F 检验和 t 检验, 两种方法之间无显著性差异。见图4、5。

和表3、4。

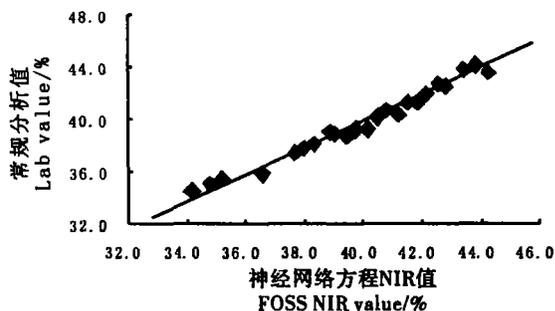


图4 蛋白校正集样品的神经网络方程 NIR 结果与常规分析值相关性

Fig. 4 Protein correlation of calibration samples between NIR value and chemical value

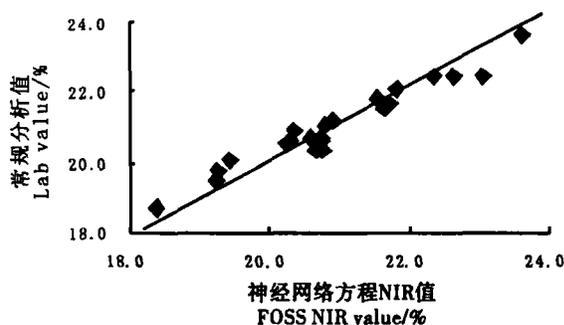


图5 脂肪校正集样品的神经网络方程 NIR 结果与常规分析值相关性

Fig. 5 Oil correlation of calibration samples between NIR value and chemical value

表3 近红外分析结果(校正集)与常规分析结果

Table 3 Contrast between NIR analysis and chemical analysis by calibration samples

地区	蛋白质 Protein		脂肪 Oil	
	近红外 Nir	常规 Lab	近红外 Nir	常规 Lab
标准误差 S	2.81	2.83	1.11	1.14
相关系数 r	0.9905		0.9737	
F 检验	0.98		0.95	
t 检验	0.07		0.09	
备注	F <sub>0.05</sub> = 1.93		t <sub>0.05</sub> = 2.06	

表4 近红外分析结果(预测集)与常规分析结果

Table 4 Contrast between NIR analysis and chemical analysis by prediction samples

地区 District	蛋白质 Protein	脂肪 Oil
相关系数 r	0.9878	0.9772
决定系数 R <sup>2</sup>	0.9757	0.9549
平均残差 Bias	-0.11	-0.04
残差平均方差 MSE	0.093	0.038

注: 残差为近红外预测值与常规分析结果的差值

## 2.3 校正模型的检验和评价

校正模型建立后, 需要用预测集样品对新建立的校正模型进行检验。大豆样品粗蛋白质、粗脂肪

校正方程的预测值与常规分析值相关关系分别见图6、7。

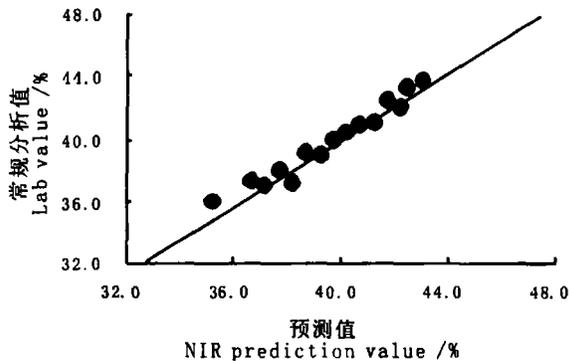


图6 蛋白预测集样品的 NIR 值与常规分析值相关性

Fig. 6 Protein correlation of prediction samples between NIR value and chemical value

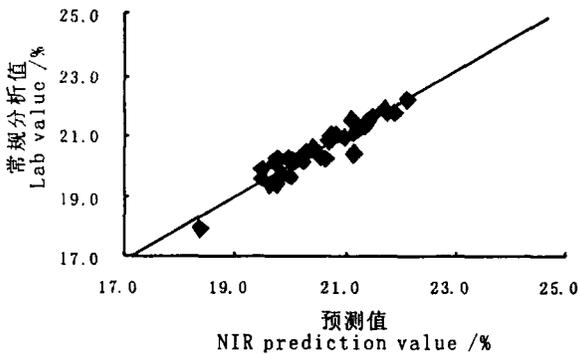


图7 脂肪预测集样品的 NIR 值与常规分析值相关性

Fig. 7 Oil correlation of prediction samples between NIR value and chemical value

本试验粗蛋白质、粗脂肪校正模型的检验结果见表5。

表5 预测集对校正模型质量的评价效果

Table 5 Evaluate result of calibration model by prediction samples

品质特性	相关系数	决定系数	平均残差	残差平均方差
Quality character	r	R <sup>2</sup>	Bias	MSE
FOSS 蛋白方程	0.9736	0.9479	0.16	0.27
FOSS Protein Model				
蛋白质校正方程	0.9735	0.9478	-0.06	0.25
Protein Calibration Model				
FOSS 脂肪方程	0.9503	0.9031	-0.19	0.13
FOSS Oil Model				
脂肪校正方程	0.9504	0.9033	0.01	0.07
Oil Calibration Model				

注: MSE, 预测误差平方的加权平均值,  $MSE = \sum (Y_i - y_i)^2 / N$

校正后的方程, 在不仅能够保持原方程较高的

相关系数和决定系数, 还能够降低预测值与实验室常规分析结果的平均残差以及残差平均方差。

### 3 讨论

本文探讨了应用近红外透射光谱无损分析技术来测定大豆粗脂肪和粗蛋白质含量, 初步论证了用该方法替代常规实验室分析方法的可靠性, 得到了较好的预测效果。

近红外神经网络方程是 FOSS 公司利用多年从世界各地收集的大豆样品的光谱信息建立的一套预测模型, 采用目前世界上最先进的 ANN 神经网络定标技术, 数据库庞大, 信息丰富。但由于大豆地域、品种差异对样品特性影响显著, 用该方程对局部地区样品的分析仍存在一定系统误差。为此, 选择少量有代表性的样品的常规分析结果对该模型进行校正, 校正模型的预测性能比原模型有明显改善, 提高了预测的准确度。

在对原有大豆神经网络方程进行校正的过程中, 样品群体界定和选择、基准法数据的准确性、光谱处理以及数学统计方法都是关键的影响因素。应该不断研究和完善, 提高近红外方法预测的准确性, 并拓展其应用范围。

### 参 考 文 献

- 1 张子仪, 陈学秀. 近红外光谱分析技术[M]. 北京: 中国农业出版社, 1992, 1-67.
- 2 赵武善, 黄斌, 陈哲, 等. 近红外透射技术分析大豆籽粒[J]. 光谱实验室, 2004(6): 1199-1202.
- 3 王友林, 王建军, 张冬青, 等. 近红外漫反射光谱法对粮油作物品质无损分析的研究[J]. 浙江农业学报, 2002(2): 105-111.
- 4 中国科学院油料作物研究所. NY/T 4-1982 谷类、油料作物种子粗脂肪测定方法[S]. 中华人民共和国国家标准, 1983. 01. 01
- 5 中国农科院分析室. NY/T 3-1982 谷类、豆类作物种子粗蛋白质测定方法(半微量凯氏法)[S]. 中华人民共和国国家标准, 1983. 01. 01
- 6 Near Infrared Reflectance Method for Whole Grain Analysis in Soybeans[S]. AACC Method 39.21, reapproval November 3, 1999.
- 7 杨曙明, 宋荣, 张瑜, 等. 饲料中水分、粗蛋白质、粗纤维、粗脂肪、赖氨酸、蛋氨酸快速测定 近红外光谱法[S]. GB/T 18868-2002.

(下转 424 页)

## NaCl TOLERANCE OF DIFFERENT GENOTYPES IN SOYBEAN UNDER TISSUE CULTURE

Wang Ping<sup>1</sup> Wang Gang<sup>2</sup> Ji Jing<sup>2</sup>

(1. *Ocean School, Huaihai Institute of Technology, Lianyungang 222005;*  
2. *Agriculture & Bioengineering College, Tianjin University, Tianjin 300072*)

**Abstract** NaCl tolerance was researched by observing emergence rate, seedling height and length of hypocotyl under solid medium with three genotypes in soybean. The results showed that there was a difference in NaCl tolerance among the different genotypes of soybean. Dongnong 46 was weak for NaCl tolerance. Among three characters of seedling stage, seedling height and length of hypocotyl showed to be more sensitive to NaCl. The increase of seedling height and length of hypocotyl was inhibited when the medium was added with 35~65 mmol/L NaCl.

**Key words** Soybean; Genotype; NaCl tolerance

---

(上接 420 页)

## STUDY ON ANALYSIS OF SOYBEAN QUALITY BY NEAR INFRARED TRANSMITTANCE SPECTROSCOPY

Yao Xinmiao Zhang Ruiying Li Xiahui Cheng Aihua Zhang Xiaobo Gao Chunxia

(*Inspection and Testing Center For Quality of Cereals and Their Products,*  
*Ministry of Agriculture, Harbin 150086*)

**Abstract** To study the rapid analysis method for soybean quality, 572 soybean variety samples were collected from four provinces in Northeast China. Soybean protein and oil content were analyzed by near infrared transmittance spectroscopy without destroying. Calibration model were established with data from chemical analysis and absorbed spectrum of calibration samples to improve the accuracy of prediction results. Satisfied  $r^2$  0.9757 (protein), 0.9549 (oil) and SEP 2.18 (protein), 0.88 (oil) were achieved. Results showed that near infrared transmittance spectrum techniques could be applied in soybean quality determination, generation selection in soybean breeding program and soybean quality classification. Abundant samples and varieties applied in this research made the prediction model more practicable.

**Key words** Soybean; Near infrared transmittance spectroscopy; Calibration; Prediction