Nondestructive identification of hard seeds of three legume plants using near infrared spectroscopy

Zhu Liwei[1], Huang Yanyan[1], Wang Qing[1],
Ma Hanxu[1], Sun Baoqi[1,2], Sun Qun[1,※]

(1. *Department of Plant Genetics and Breeding*, *College of Agriculture and Biotechnology*, *China Agricultural University / Beijing Key Laboratory of Crop Genetic Improvement*, *Beijing* 100193, *China*;
2. *Beijing Academy of Agriculture and Forestry Sciences*, *Beijing* 100097, *China*)

Abstract: To explore the applicability of the NIRS on nondestructive identification of hardseedness characteristics, in this study, near infrared spectroscopy (NIRS) models were established using near-infrared spectroscopy combined with Partial Least squares (DPLS) to investigate the hard seed characteristics of legume seeds, *Codariocalyx motorius Glycine max Cassia tora* L.and *Sophora alopecuroides*. 120 seeds of the three species were divided into two groups: calibration set (80 samples) and validation set (40 samples), which contained the same number of hard seeds and soft seeds. The influences of replicate times, spectral region and calibration samples on the discrimination rate were compared. The results showed that with 4 000-5 000 cm^{-1} of spectral range, vector correction, eight main components, the predictive value of the model fitted with the true value well. Discrimination rates of calibration and validation sets in NIR models were over 85% for *Glycine max*. With 4 000-8 000 cm^{-1} of spectral range, first-derivative spectroscopy, four main components, discrimination rates of calibration and validation sets were about 90% for *Cassia tora* L. With 4 000-8 000 cm^{-1} of spectral range, second derivative spectroscopy, eight main components, discrimination rates of calibration and validation sets were over 95% for *Sophora alopecuroides*.

Key words: infrared spectroscopy, seeds, technology, *Glycine max*, *Cassia tora* L, *Sophora alopecuroides*
doi：10.3969/j.issn.1002-6819.2012.z2.041
CLC number: S223.1 **Document code:** A **Article ID:** 1002-6819(2012)-Supp.2-0237-06

Zhu Liwei, Huang Yanyan, Wang Qing, et al. Nondestructive identification of hard seeds of three legume plants using near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(Supp.2): 237－242. (in English with Chinese abstract)
朱丽伟，黄艳艳，王 庆，等. 基于近红外光谱技术的三种硬实种子无损鉴定[J]. 农业工程学报，2012，28(增刊2)：237－242.

0 Introduction

Near infrared spectroscopy was developed in the late 1980s. It is a rapid detection technology, which exhibits low cost, rapid analysis, and good stability features. With the constant development of computer technology and chemo metrics in recent years, NIRS[1-7] has been increasingly employed for the quantitative analysis of seed quality traits such as moisture, protein, starch, and oil, seed authenticity identification, variety identification of rice, maize, and soybean, and seed purity identification[8-18].

In nature, there are many kinds of seeds with impermeable seed coat, thus, the seed cannot suck up water to germinate. These seeds are referred to as hard seeds. Many studies have shown that hard seeds exhibited high vigor traits that included *Codariocalyx motorius*, *Glycyrrhiza uralensis, Indigofera amblyantha*, *and Lespedeza bicolor*[19-22]. Further, under the same storage conditions, compared to the other seeds, the storage life of hard seeds is longer. For example, the germination rate of hard soybean for four years of storage was still high, about 97%, and the germination rate of non-hard seeds in the same lot dropped to 53% after 1.5 years of storage[23]. Therefore, there is important practical significance in the analysis of the physical structure and chemical composition of the hard and non-hard seed coat. It needs further study on

Received date：2012-06-17 Revised date：2012-09-20
Foundation item：National Science and Technology Support Program（2006BA106A15）, Special funds for basic research and operating expenses of the Ministry of Education（2009-3-7/2510JS052）
Biography：Zhu Liwei (1985－), female, Henan Province. Beijing College of Agriculture and Biotechnology, China Agricultural University 100193. Email: liweib0401001@163.com
※Corresponding author：Sun Qun (1971－), female, Shan Dong Province, associate professor, Ph.D. Engaged in the non-destructive identification of seeds quality. College of Agriculture and Biotechnology, China Agricultural University 100193. Email: sqcau@126.com

exploration of the impervious mechanism of hard seeds, which is important to seed vigor and germplasm storage. However, the present grading method for hard seeds is the soaking method[24], the process of which would destroy the seed coat structure of the soft seeds and is not conducive to effectively study the hard seed mechanism. Studies have shown that hard seed characteristics were under genetic control[25-28]. In 2009, we built a model with infrared spectrometry to identify single soft and hard licorice seed (Sun et al 2009). Using this model, we classified licorice seeds from the same variety into either soft or hard seeds. The discrimination rates for the calibration set, validation set, and prediction set were all greater than 95%[29]. To explore the applicability of the NIRS for the other seeds, which exhibit hard seed characteristics, hard seed identification on single seed of soybean (*Glycine max*), kudouzi (*Sophora alopecuroides L.*) and semen cassia (*Cassia obtusifolia L.*) were conducted. The study provides a new approach for the nondestructive identification of hard seed characteristics in leguminous plant seeds.

1 Materials and methods

1.1 Materials

Kejiao 00-786 soybean (*Glycine max*) seeds were harvested in 2005, at Keshan, Academy of Agricultural Sciences, Heilongjiang Province, China. Semen cassia (*Cassia obtusifolia L.*) seeds were provided by the Beijing Institute of Medicinal Plants. Kudouzi (*Sophora alopecuroides L.*) seeds were purchased in 2009 at the Anguo Herbal Medicine Market in Hebei Province, China.

1.2 Methods

1.2.1 Spectral acquisition

To collect the spectra, an MPA FT-NIR spectrometer (Bruker Instruments, Inc., Germany) was used. The resolution was 4 cm^{-1}, the scan range was 4 000-12 000 cm^{-1}, and the number of scans was 32. We selected seeds with uniform size, color, and appearance. Diffuse reflectance scanning was used for scanning both sides of the seeds. The average spectrum of both sides was selected to conduct the modeling. With collected the spectra, each seed was labeled and stored individually.

1.2.2 Determination of the hard seed characteristics

Following as spectral imaging, each seed was soaked at 25℃ for 24 h. Seeds absorbing water within 24 h were marked as soft seeds, and the remainder seeds without absorbing water within 24 h were marked as hard seeds.

1.2.3 Sample selection

With identification via soaking method, 120 seeds from each variety underwent near infrared qualitative analysis. The seeds of each variety were divided into two groups: the calibration set and the prediction set, with seed amounts of 80 and 40, respectively. The ratio of hard to soft seeds in each group was 1:1.

1.2.4 Data analysis and processing

To establish the qualitative analysis model of the hard seed, CAUNIRS near-infrared spectral analysis software were involved in this study which was developed by China Agricultural University, and the partial least squares (DPLS) was used.

2 Results and analysis

2.1 Selection of repetition time

Within the 4 000-8 000 cm^{-1} spectral range, the single spectrum and quadratic average spectral methods were used for modeling, respectively. For the three seeds, the discrimination rates of the calibraiton set and prediction set of the quadratic average spectral modeling were higher than those of the single spectral modeling(Table1). These indicated that using the average value of multiple scanning spectra for modeling could effectively improve the modeling effect.

Table 1 NIR discrimination rate from single spectral and average spectral

Seeds	Discrimination rates/%	Repetition time	
		Single spectra modeling	Quadratic average spectral modeling
Soybean	Calibration set	79.27	83.57
	Validation set	77.50	80.85
Semen cassiae	Calibration set	87.25	89.01
	Validation set	86.43	89.13
Kudouzi	Calibration set	93.75	96.53
	Validation set	94.44	95.83

2.2 NIRS modeling and verification of hard seed characteristics of single soybean seed

2.2.1 Selection of spectral range

The NIR spectra of hard seed and soft seed of soybean are similar (shown in Fig 1). Comparative results

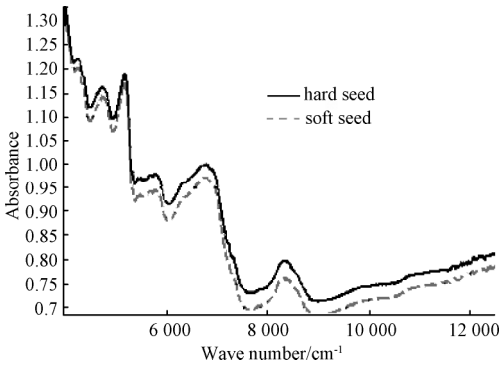

Fig.1 Near infrared spectra of hard seed and soft seed of soybean (*Glycine max*)

on the modeling effect of different spectral ranges for the soybean seed (spectra without pretreatment) are shown in Table 2. The results showed that with 4 000-10 000 cm^{-1} and 4 000-5 000 cm^{-1} of spectral regions, the modeling effects were better. With ten of the principal component, the discrimination rates of the calibration set and prediction set were 85.11% and 85.00%.

Table 2　Influences of spectral range on NIRS prediction results of soybean seeds

Spectral range/ cm^{-1}	Principal component	Discrimination rates of calibration set /%	Discrimination rates of validation set/%
4 000-5 000	10	85.11	85.00
4 000-7 000	12	82.86	80.85
4 000-8 000	13	83.57	80.85
4 000-10 000	13	84.29	85.11
5 000-9 000	13	82.98	80.00
5 000-8 000	4	79.31	76.97
8 000-10 000	7	82.98	80.00

2.2.2　Selection of pretreatment method

The pretreatments had a certain impact on the discrimination rate of the spectral model (spectral range of 4 000-5 000 cm^{-1}) (Table 3). After vector correction processing, the discrimination rates of the calibration set and prediction set were increased from 85.11% and 85.00% to 86.25% and 86.67%, respectively. Using random selection of different soybean seeds for modeling, the discrimination rates of samples of the different modeling were approximately 85 %, and the model was relatively stable.

Table 3　Influences of preprocessing methods on NIRS prediction results of soybean seeds

Preprocessing methods	Principal component	Discrimination rates of calibration set /%	Discrimination rates of validation set/%
No-preprocessing	10	85.11	85.00
Centralization	6	82.50	77.78
Range normalization	8	85.00	86.67
Vector correction	8	86.25	86.67
Scatter correction	9	82.50	84.44
First derivative (5)	4	85.00	86.67
First derivative (9)	5	83.75	82.22
First derivative (13)	5	85.00	82.22
First derivative (17)	5	85.00	84.09
Second derivative (5)	8	83.75	84.09
Second derivative (9)	9	85.00	84.09
Second derivative (13)	6	83.75	84.09
Second derivative (17)	9	85.00	86.36

2.3　NIRS modeling and verification of hard seed characteristics of single semen cassiae seed

2.3.1　Selection of spectral range

The NIR spectra of hard seed and soft seed of semen cassiae are similar (shown in Fig.2). The discrimination rates of different spectral ranges for the semen cassiae seed are shown in Table 4. The results showed that the modeling effect in the 4 000-5 000, 4 000-8 000, 4 000-10 000 and 5 000-8 000 cm^{-1} spectral regions were better than that of the other spectral regions. The discrimination rates for the calibration set and validation set of these four spectral regions were around 90.00%. For example, the discrimination rates of the calibration set and validation set for the 4 000-5 000 cm^{-1} spectral region were 90.11% and 86.96 %, respectively. Meanwhile, the modeling effect of high frequency regions, such as 8 000-10 000 cm^{-1}, were worse than those of other spectral regions.

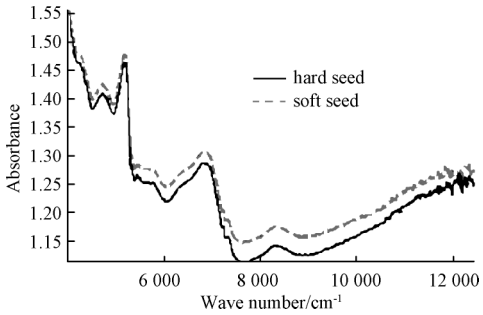

Fig.2　Near infrared spectra of hard seed and soft seed of *Cassia tora* L.

Table 4　Influences of spectral range on NIRS prediction results of semen cassiae seeds

Spectral range /cm^{-1}	Principal component	Discrimination rates of calibration set /%	Discrimination rates of validation set/%
4 000-5 000	9	90.11	86.96
4 000-7 000	6	85.71	86.96
4 000-8 000	7	89.01	89.13
4 000-10 000	8	89.01	86.96
5 000-9 000	8	85.71	89.13
5 000-8 000	8	90.11	89.13
8 000-10 000	5	81.32	82.61

2.3.2　Selection of pretreatment method

The modeling effects for semen cassiae seed under different pretreatment methods are shown in Table 5 (spectral range of 4 000-8 000 cm^{-1}). Pretreatment methods, such as centralization, vector correction, scatter correction, and derivative treatment, did not have a significant effect on the increase of the modeling discrimination rate (Table 5). After first derivative processing (smoothing point was 5), the discrimination rates for the calibration set and validation set were respectively 91.21% and 89.13%; these values were increased compared to the control. After first derivative processing, the principal component number significantly were decreased, and the same modeling effect could be reached with four principal components. During random selection of different semen cassiae seeds for modeling, the

average discrimination rates for the calibration set and validation set were all approximately 88.27% and 87.15%, and the model was relatively stable.

Table 5 Influences of different preprocessing methods on NIRS prediction results of semen cassiae seeds

Preprocessing methods	Principal component	Discrimination rates of calibration set /%	Discrimination rates of validation set/%
No-preprocessing	7	89.01	89.13
Centralization	6	87.91	89.13
Vector correction	8	90.11	86.96
Range normalization	8	90.11	86.96
Scatter correction	5	89.01	89.13
First derivative (5)	4	91.21	89.13
First derivative (9)	4	90.11	89.13
First derivative (13)	4	90.11	89.13
First derivative (17)	4	90.11	89.13
Second derivative (5)	7	89.13	87.91
Second derivative (9)	6	86.17	91.67
Second derivative (13)	8	89.01	86.96
Second derivative (17)	8	90.11	86.96

2.4 NIRS modeling and verification of hard seed characteristics of single kudouzi seed

2.4.1 Selection of spectral range

The NIR spectra of hard seed and soft seed of kudouzi are similar (shown in Fig 3). The modeling effects of different spectral ranges for the kudouzi seed are shown in Table 6. The modeling effects of the 4 000-8 000 and 4 000-10 000 cm^{-1} spectral regions were better, the discrimination rates for the calibration set and validation set were above 95.00%. Between the two regions, the modeling effect of the 4 000-8 000 cm^{-1} spectral region was better, and the discrimination rates of the calibration set and validation set were 96.53% and 95.83%, respectively. Meanwhile, the modeling effect of the high frequency regions, such as the 8 000-10 000 cm^{-1} spectral range, was worse than that produced by the other spectral regions. Under this spectral range, the discrimination rates for the calibration set and validation set were 93.06% and 87.50%.

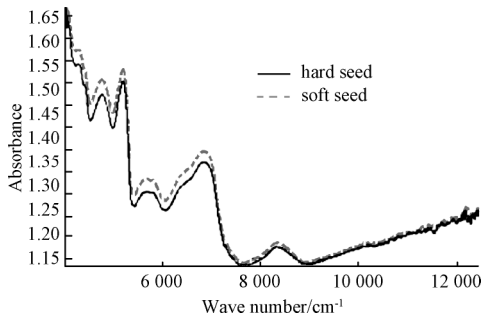

Fig3 Near infrared spectra of hard seed and soft seed of Kudouzi(*Sophora alopecuroides*)

Table 6 Influences of spectral range on NIRS prediction results of kudouzi seeds

Spectral range /cm^{-1}	Principal component	Discrimination rates of calibration set /%	Discrimination rates of validation set/%
4 000-5 000	11	93.75	91.67
4 000-7 000	11	97.92	91.67
4 000-8 000	13	96.53	95.83
4 000-10 000	10	95.14	95.83
5 000-8 000	9	93.06	92.36
5 000-9 000	1	91.67	90.97
8 000-10 000	1	93.06	87.50

2.4.2 Selection of pretreatment method

Table 6 shows the modeling effects for kudouzi seed with different pretreatment methods (spectral range of 4 000-8 000 cm^{-1}). Some of the pretreatment methods have an effect on the increase of the discrimination rate (Table 7). For example, after centralization processing of the spectrum, the discrimination rates of the calibration set and validation set were increased, then decreased, but the effects were not significant. For example, when the smoothing point was 17 and the second derivative was used to preprocess the spectrum, the same modeling effect could be achieved with the principal component of eight. The average discrimination rates of the calibration set and validation set of these spectrum regions were 96.53% and 95.83%, respectively, which showed that a good effect was achieved. For the modeling of a random selection of different kudouzi seeds, the discrimination rates of the calibration set and validation set were all approximately 90%, and the model was relatively stable.

Table 7 Influences of preprocessing methods on NIRS prediction results of kudouzi seeds

Preprocessing methods	Principal component	Discrimination rates of calibration set /%	Discrimination rates of validation set/%
No-preprocessing	13	96.53	95.83
Centralization	12	97.92	95.83
Vector correction	13	96.53	93.06
Range normalization	13	96.53	95.83
Scatter correction	13	95.83	94.44
First derivative (5)	7	93.06	93.06
First derivative (9)	6	93.75	93.06
First derivative (13)	7	94.44	94.44
First derivative (17)	10	95.14	91.67
Second derivative (5)	13	97.92	93.06
Second derivative (9)	13	97.92	94.44
Second derivative (13)	13	96.53	93.06
Second derivative (17)	8	96.53	95.83

3 Discussion

In this study, we used NIRS and DPLS for the

qualitative identification of hard seed characteristics of soybean, kudouzi, and semen cassiae single seed. For soybean seeds in the 4 000-5 000 cm^{-1} spectral region, the discrimination rates of the calibration set and validation set for the quadratic average spectral model were 85.11% and 85.00%, respectively. For semen cassiae seeds in the 4 000-8 000 cm^{-1} spectral range, the highest discrimination rates for the calibration set and validation set were 89.01% and 89.13%, respectively. For the kudouzi seeds in the 4 000-8 000 cm^{-1} spectral range, the highest discrimination rates for the calibration set and validation set for the quadratic average spectral model were 96.53% and 95.83%, respectively. The random selection of samples for modeling exhibited a certain influence on the modeling effect, but the effect was not significant. For example, the discrimination rates for the calibration set and validation set for soybean seeds were approximately 85.00%, the discrimination rates for the calibration set and validation set for semen cassiae and kudouzi seeds were approximately 90%. The discrimination rate of the quadratic average spectral mode was higher than that of a single spectrum. Thus, the way to improve the discrimination rate of hard seed characteristics was to scan the samples multiple times and obtain the average value for modeling. Zhao (1999) illustrated that good measurement methods and conditions in the use of NIRS for the determination of the corn seed protein content included the use of Fourier transform NIRS. Further, an improved method could improve the signal to noise ratio in order to increase the accuracy of the test results. Sun (2009) conducted research on Ural licorice seeds, the results showed that in quartic average spectral modeling, the model discrimination rate was significantly increased compared to that of the quadratic average spectral model in[29].

Many factors influence the effectiveness of the modeling effect. From our experimental results, we found that the NIRS achieved better results in semen cassiae and kudouzi seeds than in soybean seeds. According to the observation of the morphological changes of seeds during seed soaking process, we found there were more difference between hard and soft seeds of semen cassiae and kudouzi that of soybean This difference may be the main reason for the variance in the modeling effect. The formation of hard seeds was affected by many factors, such as genetics, environmental conditions during seed maturation, and seed maturity level. Therefore, application of NIRS on the qualitative identification of the hard seed characteristics should consider all above factors in order to improve the applicability of the model. In future work, we will analyze the relationship between the NIR characteristics and hard seeds to facilitate the NIR identification mechanism of hard seeds.

4 Conclusion

1) For soybean seeds: 4 000-5 000 cm^{-1},vector correction, 8 main components, the discrimination rate of calibration and validation sets are both over 86 %, even if with different modeling samples, the models have high discrimination rate (over 85%);

2) For semen cassiae seeds: 4 000-8 000 cm^{-1}, first-derivative spectroscopy, four main components, the discrimination rate of calibration and validation sets are both about 90 %, even if with different modeling samples, the models have high discrimination rate (over 85 %);

3) For Kudouzi seeds: 4 000-8 000 cm^{-1}, second derivative spectroscopy, eight main components, the discrimination rate of calibration and validation sets are both over 95 %, even if with different modeling samples, the models have high discrimination rate (over 95 %).

[References]

[1] Kang Yueqiong, Hao Feng. The Study of the Determination of Seed Moisture and Seed Vigor with Fourier Transform Near-Infrared Spectroscopy[J]. Seed, 2004, 23(7): 10－16.

[2] Fan Weiyan, Xing Han, Lin Jiayong, et al. Study on the determination of moisture content of rice by near-infrared spectr oscopy[J]. Science and Technology of Cereals, Oils and Foods, 2008, 16(5): 49－52, 69.

[3] Zhou Yan, Huang Chuanxu, Chen Bin. Fast Measuring the Moisture of Wheat by Near Infrared Spectrophotometer[J]. Modern Scientific Instruments, 2002(6): 50－52.

[4] Lu Lijun, Zhuang Shuhua. Determinating contents of moisture protein and crude fat in soybean meal by near infrared technology[J]. Journal of Molecular Science, 2001, 17(2): 115－120.

[5] Zhao Huanhuan, Hu Yaogao, Zhao Qibo, et al. Application of near infrared reflectance spectroscopy (NIRS) to analysis crude protein in grounded rye hay[J]. Acta Zoonutrimenta Sinica, 2001, 13(4): 40－43.

[6] Nicolai B M, Beullens K, Bobelyn E, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review[J]. Postharvest Biol. Technol, 2007, 46(2): 99－118.

[7] Malley D F, Findlay D L, Zippel B. Feasibility of using near-infrared reflectance spectroscopy for the analysis of C, N, P, and diatoms in lake sediments[J]. Journal of Paleolimnology, 1999, 21(5): 295.

[8] Zhao Jinghui, Liu Xu, Li Fangyuan, et al. The identification of different panax ginseng seeds by NIRS [J]. Special Wild Economic Animal and Plant Research, 2006, 28(3): 48－50.

[9] Liang Liang, Liu Zhixiao, Yang Minhua, et al. Discrimination of variety and authenticity for rice based on visual/near infrared reflection spectra[J]. Journal of Infrared and Millimeter Waves, 2009, 28(5): 353－356, 391.
[10] Huang Yanyan, Zhu Liwei, Li Junhui, et al. Rapid and nondestructive discrimination of hybrid maize seed purity using near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2011, 31(3): 661－664.
[11] Lidia E A, David D E, Susan D, et al. Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels[J]. Journal of Cereal Science, 2012, 55(2): 160－165.
[12] Leon L, Varo A G, Downey G. Parent and harvest year effects on near-infrared reflectance spectroscopic analysis of Olive (Olea europaea L.) fruit traits[J]. J. AgricFood Chem, 2004, 52(16): 4957－4962.
[13] Paul W, Paul G, Glen F, et al. Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis[J]. Analytica Chimica Acta, 2009, 653(2): 121－130.
[14] Gokhan H, Bismark L, Mark S. Near Infrared Reflectance Spectroscopy Predicts Protein, Starch, and Seed Weight in Intact Seed of Common Bean (*Phaseolus vulgaris L.*)[J]. J. AgricFood Chem, 2010, 58(2): 702－706.
[15] Akiko T, Takuya S, Sumio K. Nondestructive determination of tannin content in dividual acorns by near-infrared spectroscopy[J]. Ecol Res, 2011, 26(3): 679－685.
[16] Marcal P, Joan S, Francesc C, et al. Near-Infrared spectroscopy Analysis of Seed Coat of Commom Beans (*Phaseolus vulgaris L.*): a potential tool for breeding and quality and evaluation[J]. J. AgricFood Chem, 2012, 60(3): 706－712.
[17] Aldo R, Luis G, Ezequiel O, et al. Near-infrared reflectance spectroscopy (NIRS) for Protein, Tryptophan, and Lysine evaluation in quality protein maize (QPM) breeding programs[J]. J. AgricFood Chem, 2011, 59(20): 10781－10786.
[18] Roland W, Willi G, Bernhard R, et al. Near-Infrared Spectroscopy on measure maize forage quality parameters online[J]. Crop Sci, 2003, 43(4): 1407－1413.
[19] Tian Juan, Sun Qun, Wang Jianhua, et al. Differences of seed vigor among different levels of hard seeds of glycyrrhiza uralensis fisch[J]. Plant Physiolgogy Communications, 2007, 43(2): 235－240.
[20] Xu Benmei, Sun Yuntao, Sun Chao, et al. Studies on higher vigour of hard seeds[J]. Seed, 2005, 24(8): 44－48.
[21] Bai Chunxia, Han Jianguo, Sun Yan, et al. Study on the relationship between hard-seededness and seed vigor of indigofera amblyantha and lespedeza bicolor[J]. Acta Prataculturae Sinica, 2006, 15(5): 82－87.
[22] Xu Benmei, Sun Yuntao, Li Ruili, et al. Detection of higher vigour of hard seed of codariocalyx motorius[J]. Scientia Silvae Sinicae, 2006, 42(10): 54－58.
[23] Wang Jinlong. Study on preservation of soybean germplasm using soybean hard seed[J]. Soybean Science, 1999, 18(4): 351－354.
[24] Yang Qihe, Yin Xiaojuan, Ye Wanhui. Dormancy mechanism and breaking methods for hard seeds[J]. Chinese Bulletin of Botany, 2006, 23(1): 108－118.
[25] Zhang Taiping, Boquet D J and Moore S H. Inheritance of hard-seed of soybean and its relationship with other traits[J]. Soybean Science, 1992(11): 88－92.
[26] Kilen T C, Hartwig E E. An inheritance study of impermeable seed in soybean[J]. Field Crops Research. 1978, (1): 65－70.
[27] Marjushkin V F, Sichkar V F, Michailov V G, et al. Inheritance of hardseededness in soybean[J]. Soybean Genetic News, 1987, 14: 294－297.
[28] Shahi T P, Pandey M P. Inheritance of seed permeability in soybean[J]. Indian Journal of Genetics and Plant Breeding, 1982, 42（2）: 196－199.
[29] Sun Qun, Li Junhui, Wang Jianhua, et al. Identification of hardness of licorice single seed using near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2009, 29(10): 2669－2672.

基于近红外光谱技术的三种硬实种子无损鉴定

朱丽伟[1]，黄艳艳[1]，王　庆[1]，马晗煦[1]，孙宝启[2]，孙　群[1※]

（1. 中国农业大学农学与生物技术学院植物遗传育种学系/北京市作物遗传改良重点实验室，北京 100193；
2. 北京市农林科学院，北京 100097）

摘要：为探讨近红外光谱技术在鉴定种子硬实特性上的普遍性，该文采用近红外光谱法结合偏最小二乘法建立了大豆、苦豆子和决明子单粒种子硬实特性的定性分析模型，每种种子均选择 120 粒种子进行近红外定性分析，种子分为建模集、检验集 2 组，建模集 80 粒，检验集 40 粒，各组中硬实与非硬实种子的比例为 1:1，比较了光谱重复次数、光谱范围以及不同建模样品的建模效果。结果表明：采用二次平均光谱所建模型的鉴别率优于单次光谱；大豆采用 4 000～5 000 cm^{-1} 光谱范围，矢量校正预处理，主成分为 8 时，建模集与检验集鉴别率均在 85%以上；决明子采用 4 000～8 000 cm^{-1} 光谱范围，一阶导数预处理，主成分为 4 时，模型建模集与检验集鉴别率均在 90%左右；苦豆子采用 4 000～8 000 cm^{-1} 光谱范围，二阶导数预处理，主成分为 8 时，模型的建模集与检验集鉴别率均在 95%以上。以上结果表明近红外光谱技术可以很好地应用于单粒种子硬实特性的判断，有助于硬实机理的深入研究。

关键词：红外光谱，种子，技术，大豆，苦豆子，决明子