Korean J. Crop Sci. 46(5): 401~406(2001)

Non-destructive Method for Selection of Soybean Lines Contained High Protein and Oil by Near Infrared Reflectance Spectroscopy

Myoung Gun Choung*†, In Youl Baek*, Sung Taeg Kang*, Won Young Han*, Doo Chull Shin*, Huhn Pal Moon* and Kwang Hee Kang**

**National Yeongnam Agricultural Experiment Station, RDA, Milyang 627-803, Korea*
***College of Natural Resources, Yeungnam University, Kyongsan 712-749, Kore.*

ABSTRACT : The applicability of non-destructive near infrared reflectance spectroscopic (NIRS) method was tested to determine the protein and oil contents of intact soybean [*Glycine max* (L.) Merr.] seeds. A total of 198 soybean calibration samples and 101 validation samples were used for NIRS equation development and validation, respectively. In the developed non-destructive NIRS equation for analysis of protein and oil contents, the most accurate equation was obtained at 2, 8, 6, 1(2nd derivative, 8 nm gap, 6 points smoothing, and 1 point second smoothing) and 2, 1, 20, 10 math treatment conditions with Standard Normal Variate and Detrend (SNVD) scatter correction method and entire spectrum (400~2500 nm) by using Modified Partial Least Squares (MPLS) regression, respectively. Validation of these non-destructive NIRS equations showed very low bias (protein : 0.060%, oil : -0.017%) and standard error of prediction (SEP, protein : 0.568%, oil : 0.451%) as well as high coefficient of determination (R^2, protein : 0.927, oil : 0.906). Therefore, these non-destructive NIRS equations can be applicable and reliable for determination of protein and oil content of intact soybean seeds, and non-destructive NIRS method could be used as a mass screening technique for selection of high protein and oil soybean in breeding programs.

Keywords **:** soybean [*Glycine max* (L.) Merr.], non-destructive NIRS, protein, oil

Soybean [*Glycine max* (L.) Merr.] seeds is one of the world's most important sources of protein and oil, mainly due to its high protein and oil contents (Pazdernik *et al.*, 1997). Soybean seeds contain more protein than any other cultivated commercial crops. Approximately 40% of the dry weight of the soybean seed is storage protein, and 20% is oil (Choung *et al.*, 2001).

Recently, more emphasis has been placed on breeding to improve the soybean protein and oil content, but the lack of a fast and efficient mass screening method retarded breeding for improving seed protein and oil contents.

†Corresponding author: (Phone) +82-55-350-1223 (E-mail) cmg7004@rda.go.kr <Received October 8, 2001>

For measuring protein and oil content in soybean seed, the Kjeldahl and soxhlet method are widely utilized, respectively. However, these methods and related methodologies are relatively complicate, time consuming, involved corrosive chemicals, and required elaborate laboratory facilities, which have deterred the practical utility in many breeding programs (Williams *et al.*, 1984; Pazdernik *et al.*, 1997). Due to these difficulties, rapid and less hazardous methods, such as the use of near infrared reflectance spectroscopy (NIRS), are highly demanded in soybean breeding programs.

The NIRS is a multi-trait technique that fulfills most of the requirements for rapid, accurate, and cost-effective mass screening for several seed quality traits in many crops (Velasco *et al.*, 1997; Pazdernik *et al.*, 1997; Perez-Vich *et al.*, 1998; Oh *et al.*, 2000; Choung *et al.*, 2001). The first trial of NIRS was to measure moisture content in soybean and NIRS become popular to measure moisture, protein, oil and starch contents in many cereals, legumes, forages and other food commodities over the past 20 years (Halgerson *et al.*, 1995; Hatty *et al.*, 1994; Roy *et al.*, 1993). In ground soybean powder, the protein and oil contents have been accurately estimated using NIRS (Rinne *et al.*, 1975; Hilliard and Daynard, 1976; Choung *et al.*, 2001). However, the applications and studies for non-destructive NIRS technique of soybean protein and oil were insufficient. Therefore, the objectives of this research were to study the potential of non-destructive NIR spectrocomputer system to estimate the protein and oil contents of intact soybean seeds and to provide the mass screening technique for selection of high protein and oil soybean in breeding programs.

MATERIALS AND METHODS

Soybean samples

The total of 310 soybean germplasms were used in this study. The soybeans were grown at the experimental field of National Yeongnam Agricultural Experiment Station, Milyang, Korea in 2000. Seed samples of 310 soybean germ-

plasms were measured as intact soybeans by a NIRS system, and then soybean seeds were ground with a ball mill and sieved with a 1.0 mm screen. The ground samples were analyzed protein and oil contents by standard chemical methods.

Measuring protein, oil, and moisture content

Auto-Kjeldahl system was applied to determine the protein contents of soybean seeds. 0.2 g of ground sample was digested by Buchi B-435 digestion system and Buchi B-412 scrubber with 20 ml of sulfuric acid and 3 g of catalyst ($CuSO_4$: K_2SO_4=1 : 9). Percent nitrogen was calculated by Buchi B-339 auto-kjeldahl system and then converted to percent protein by multiplying 6.25.

The oil contents were determined by auto-soxhlet method with Buchi B-811 extracted system. 2 g of ground samples was extracted by hexane for 2 hours, pre-heated for 10 minutes, and then dried 1 hour at 105°C.

The moisture contents were analyzed by oven-dry method with 105°C for 2 hours, and then all protein and oil contents were estimated on the basis of dry matter.

Scanning and pretreatment of NIRS spectra

The intact soybean spectra in the visible-near infrared region were measured on a NIRSystem Model 6500 (Silver Springs, MD, USA) monochrometer near infrared reflectance spectrophotometer using a small reflectance vessel (Photo 1).

The NIRS spectral data were recorded between 400 nm and 2500 nm at 2-nm intervals and stored as the reciprocal logarithm (log 1/R) of the reflected energy. NIRS instrument control as well as all graphics, and NIRS specific calculations were performed with the software package WinISI (version 1.05) released by Infrasoft International (Port Matilda, PA). In WinISI software package, two programs, *Center* and *Select*, were used to screen samples for spectral outliers and to choose samples that represented the 310 soybean samples, respectively. The *Center* program defined spectral boundaries that eliminated outliers, which were defined as having a maximum standardized Mahalanobis' distance (*H*-distance) of 3.0 from the samples mean, and the *Select* program eliminated samples having similar spectra (minimum standardized *H*-distance of 0.6 from their nearest neighboring samples) (Shenk and Westerhaus, 1991a). From the results of pre-treatment of NIRS spectra, one calibration set (198 samples) and one validation set (101 samples) were randomly selected from the log 1/R spectra of 310 soybean germplasms.

Photo 1. Small reflecta.nce vessel for measuring NIRS spectrum of intact soybean seeds (about 15 g samples needed).

Calibration

The NIRS calibration equations of protein and oil content were developed for intact seed calibration sample set using the WinISI program *Calibrate* with the modified partial least squares (MPLS) regression of 3 different derivative math treatment (log 1/R, D^1 log 1/R and D^2 log 1/R). Three additional regression methods, such as partial least squares (PLS), principle component regression (PCR) and multiple linear regression (MLR) were tested on the calibration sample set. The standard normal variate and detrend (SNVD) and "none" transformations were implemented for scatter correction (Shenk and Westerhaus, 1991b). And the wavelengths at every 2 nm across the entire visible (408~1092 nm) plus near infrared (1108~2492 nm) spectrum were used for calibration. A trimmed spectrum including only the near infrared range was tested against the entire spectrum. The Standard error of calibration (SEC), coefficient of determination(R^2), standard error cross-validation (SECV) and/or one minus the ratio of unexplained variance to total variance (1-VR) statistics were used to select the best calibration equation (Windham *et al.*, 1989).

Validation

The protein and oil NIRS equations of intact soybean seeds were monitored with the WinISI program *Monitor*, using the validation set of 101 samples. The standard error of prediction (SEP), R^2, bias, standard deviation of residual (R.SD), and SEP/Mean (Standard Error of Prediction per Mean) statistics were analyzed to determine the accuracy of prediction (Windham *et al.*, 1989). The 101 validation samples had the standardized *H*-distance of 3.0 or less from the mean of the calibration sample set.

RESULTS AND DISCUSSION

Protein and oil contents and NIRS spectra

The mean protein and oil contents of the calibration sample set were 44.4% (range : 39.9% to 50.9%) and 18.4% (range : 14.3% to 21.8%) with a standard deviation of 1.98% and 1.35% as determined by auto-Kjeldahl and auto-soxhlet system, respectively (Table 1). And the means, ranges and standard deviations of protein and oil contents in validation sample set were similar to the calibration sample set (Table 1). The log 1/R spectra of the intact soybean seeds with high and low contents of protein and oil are shown in Fig. 1.

Fig. 2 shows the D^2 log 1/R spectra and mean standard deviation spectrum of calibration samples that are obtained by 900~2500 nm. Several high standard deviation peaks are closely connected with the functional groups (C-H, CH_3, O-H, C-O, N-H and C=O), and those peaks contribute to the NIRS calibration of protein and oil (Osborne and Fearn, 1988).

Table 1. Laboratory reference value statistics for protein and oil content based on intact soybean seed samples.

Sample set		n†	Mean	Range†	SD
			%		
Calibration	Protein	198	44.36	39.92~50.90	1.98
	Oil	198	18.39	14.32~21.84	1.35
Validation	Protein	101	44.75	40.66~51.13	2.09
	Oil	101	18.37	14.45~21.42	1.48

†Sample number
‡Protein and oil contents were estimated on the basis of dry matter.

Calibration and validation analysis for protein content

The optimal NIRS equation of intact soybean seeds protein are shown in Table 2. The difference of scatter correction method, wavelength, and math treatment effect did not highly improve the MPLS regression performance, thus the optimal equation condition was obtained at 2, 8, 6, 1 (2nd derivative, 8 nm gap, 6 points smoothing, and 1 point second smoothing) math treatment condition with SNVD scatter correction method and entire spectrum (Table 2 and Fig. 3, left panel). In a different regression method with same 2, 8, 6, 1 math treatment and SNVD scatter correction, the equation of MPLS method showed the lowest SEC and the highest R^2 among other regression methods (data not shown).

One important criterion for evaluating NIRS equations involves the test of prediction accuracy with unknown samples. Validation sample set allows NIRS equation to be validated for prediction accuracy based on random samples not used in calibration sample set (Pazdernik *et al.*, 1997). Based on the SEP, R^2, bias, residual of standard deviation and SEP/Mean, the best equation (2, 8, 6, 1; SNVD; 400~2500 nm) using MPLS method was accurately predicting the protein contents of validation sample set (Table 2). The right panel of Fig. 3 demonstrates the accuracy of the intact soybean seeds equation for protein on the basis of the relationship between the actual protein value calculated from auto-Kjeldahl and the predicted protein values from the NIRS. Fig. 4 showed the histogram of differences between NIR and Kjeldahl protein contents with a fitted normal distribution curve. This result indicates that the non-destructive NIRS analysis can be used as a mass screening method for

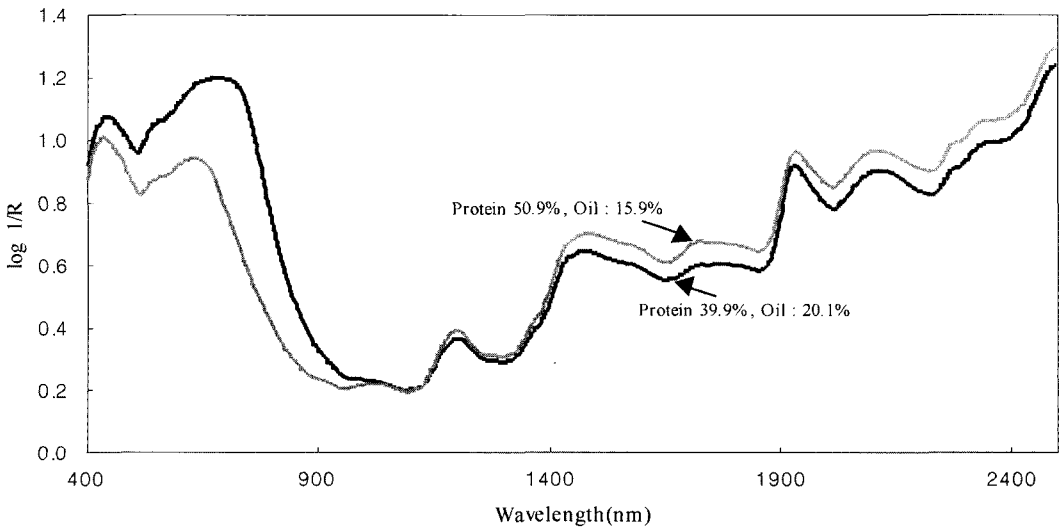

Fig. 1. Raw spectra of NIRS with different protein and oil contents in intact soybean seed samples.

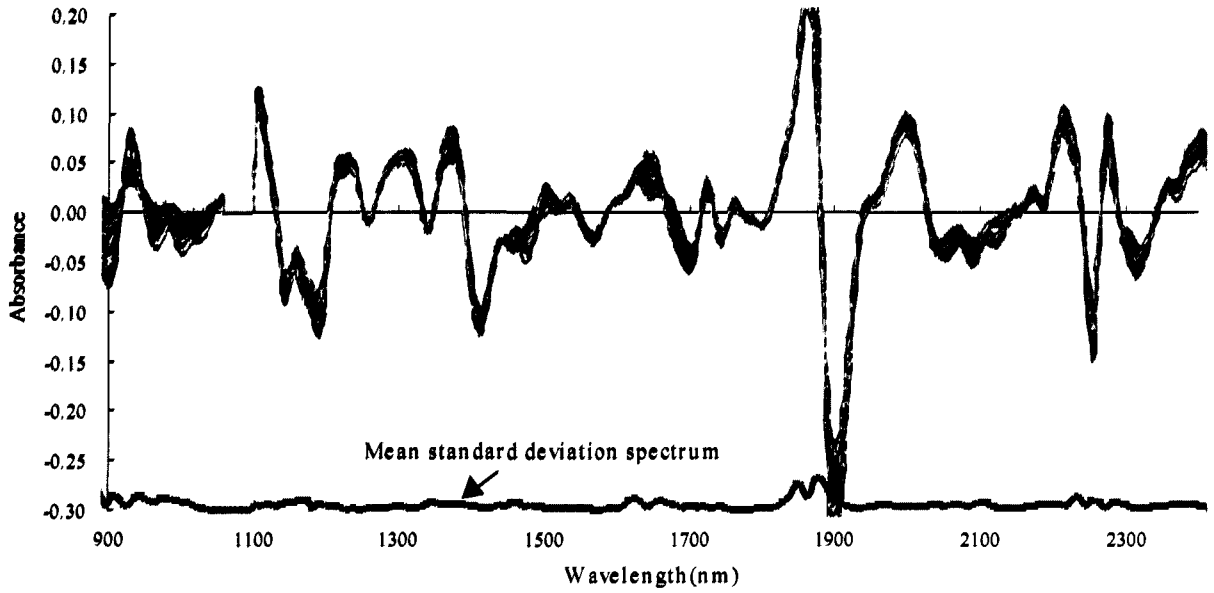

Fig. 2. Second derivative spectra and mean standard deviation spectrum of intact soybean seeds in calibration sample set.

Table 2. Statistics for optimal protein calibration and validation obtained from MPLS regression method.

	Calibration† (n=198)			Validation (n=101)				
Term	SEC	R^2	1-VR	SEP	R^2	Bias	R. SD(%)	SEP/M(%)
9	0.442	0.950	0.926	0.568	0.927	0.060	0.57	1.27

†Math treatment condition: 2, 8, 6, 1; Scatter correction: SNVD; Wavelength range: 400~2,500 nm.
SEC: standard error of calibration; R^2: coefficient of determination; 1-VR: one minus the ratio of unexplained variance to total variance; SEP: standard error of prediction; Bias: difference between reference method and predicted mean; R.SD: residual of standard deviation; SEP/M: standard error of prediction/predicted mean.

Table 3. Statistics for optimal oil calibration and validation obtained from MPLS regression method.

	Calibration† (n=198)			Validation (n=101)				
Term	SEC	R^2	1-VR	SEP	R^2	Bias	R. SD(%)	SEP/M(%)
9	0.418	0.904	0.795	0.451	0.906	-0.017	0.45	2.45

†Math treatment condition: 2, 1, 20, 10; Scatter correction: SNVD; Wavelength range : 400~2,500 nm.
SEC: standard error of calibration; R^2: coefficient of determination; 1-VR: one minus the ratio of unexplained variance to total variance; SEP: standard error of prediction; Bias: difference between reference method and predicted mean; R.SD: residual of standard deviation; SEP/M: standard error of prediction/predicted mean.

improving protein content in number of soybean breeding programs.

Calibration and validation analysis for oil content

Table 3 shows the best NIRS equation statistics of oil content obtained from intact soybean seed. The best equation condition was obtained at 2, 1, 20, 10 (2nd derivative, 1 nm gap, 20 points smoothing and 10 points second smoothing) math treatment condition with SNVD scatter correction method and entire spectrum (400~2500 nm) region (Table 3 and Fig. 5, left panel). Based on several prediction statistics, the best non-destructive NIRS equation of oil analysis using MPLS method equation (2, 1, 20, 10; SNVD; 400~2500 nm) was well predicting the oil contents of validation sample set, and the SEP value and R^2 of prediction were 0.451% and 0.906, respectively (Table 3 and Fig. 5, right panel). Also, the histogram of differences between NIR and soxhlet oil contents with a well fitted normal distribution curve was showed in Fig. 6. This result indicates that the non-destructive NIRS analysis can be used as an effective method for measuring soybean oil contents.

In conclusion, the main purpose of creating these non-destructive NIRS equations was to develop a rapid and pre-

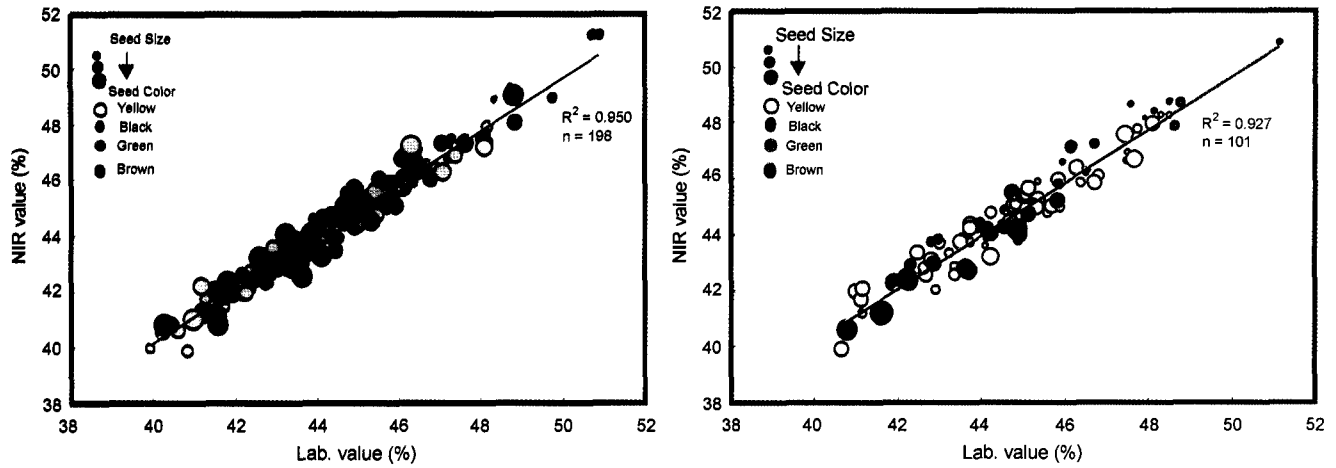

Fig. 3. Scatter plots of protein content by Kjeldahl versus protein content by NIRS for the calibration (left) and validation (right) sample set.

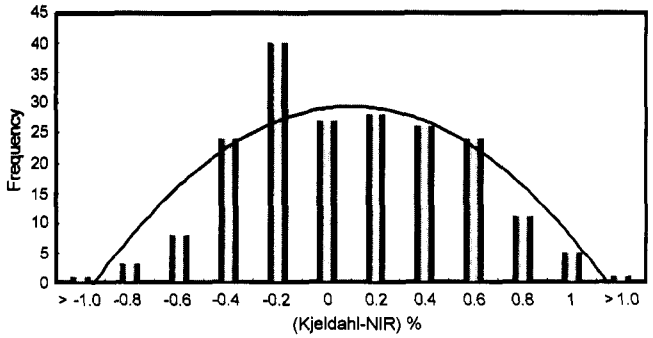

Fig. 4. Histogram of differences between Kjeldahl and NIR protein contents with a fitted normal distribution curve.

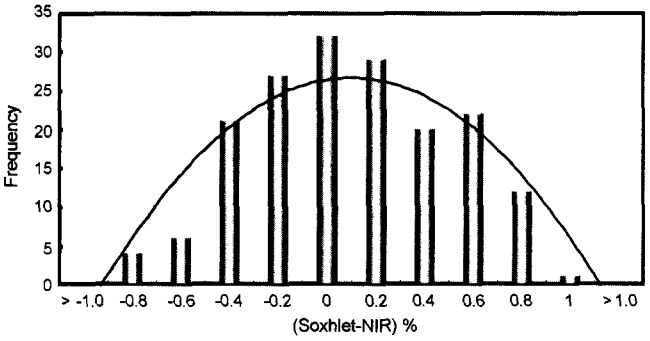

Fig. 6. Histogram of differences between soxhlet and NIR oil contents with a fitted normal distribution curve.

cise screening method for protein and oil analysis of soybean seeds. However, the non-destructive NIRS protein and oil analysis equation in this study did not predict the calibration set so accurately as the destructive NIRS (using ground seeds) protein and oil equations in previously reported result (Choung *et al.*, 2001), but non-destructive NIRS protein and oil equations have demonstrated standard errors less than 0.5% of protein and oil contents in intact soybean seed sam-

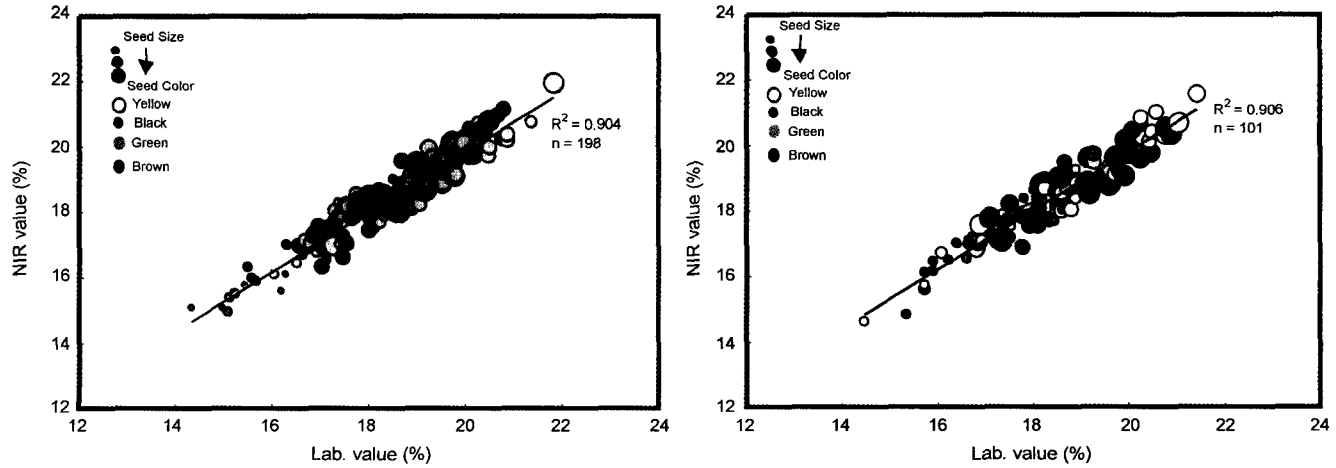

Fig. 5. Scatter plots of oil content by soxhlet versus oil content by NIRS for the calibration (left) and validation (right) sample set.

ples. Therefore, non-destructive NIRS method can be used as a mass screening technique to quickly evaluate a large number of soybean lines and breeding populations for improving protein and oil content. Future research should aim to develope NIRS technique of non-destructive one-seed analysis and to improve the accuracy and sample range.

REFERENCES

Choung, M. G., I. Y. Baek, S. T. Kang, W. Y. Han, D. C. Shin, H. P. Moon and K. H. Kang. 2001. Determination of protein and oil contents in soybean seed by near infrared reflectance spectroscopy. *Korean J. Crop Sci.* 46(2) : 106-111.

Halgerson, J. M., C. C. Sheaffer, O. B. Hesterman, T. S. Griffin, M. D. Stern, and G. W. Randall. 1995. Prediction of ruminal protein degradability of forages using near infrared reflectance spectroscopy. *Agron. J.* 87 : 1227-1231.

Hatty, J. A., W. E. Sabbe, G. D. Basten and A. B. Blakeney. 1994. Nitrogen and starch analysis of cotton leaves using near infrared reflectance spectroscopy (NIRS). *Commun. Soil Sci. Plant Anal.* 25 : 1855-1863.

Hilliard, J. H. and T. B. Daynard. 1976. Measurement of protein and oil in grains and soybean with reflected near infrared light. *Canadian Institute of Food Science and Technology. Journal* 9 : 11-14.

Oh, K. W., M. G. Choung, S. B. Pae, C. S. Jung, B. J. Kim, Y. C. Kwan, J. T. Kim and Y. H. Kwack. 2000. Determination of seed lipid and protein contents in perilla and peanut by near-infrared reflectance spectroscopy. *Korean J. Crop Sci.* 45(5) : 339-342.

Osborne, B. G. and T. Fearn. 1988. Near infrared spectroscopy in food analysis. *Longman Scientific & Technical. John Wiley & Sons, Inc.*

Pazdernik, D. L., A. S. Killam, and J. H. Orf. 1997. Analysis of amino acid and fatty acid composition in soybean seed, using near infrared reflectance spetroscopy. *Agron. J.* 89 : 679-685.

Perez-Vich, B., L. Velasco and J. M. Fernandez-Martinez. 1998. Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near-infrared reflectance spectroscopy. *J. Amer. Oil Chem. Soc.* 75(5) : 547-555.

Probst, A. H. and R. W. Judd. 1973. Origin, US history and development and world distribution. In: ed. Caldwell, B. E. Soybeans: Improvement, production and uses. Agron. mongr. 16 1st ed. ASA, CSSA and SSSA, Madison, WI.

Rinne, R. W., S. Gibbons, J. Bradley, R. Sief, and C. A. Brim. 1975. Soybean protein and oil percentage determined by infrared analysis. *Agric. Res. Pub. ARS-NC-26 USDA*: Washington DC.

Roy, S., R. C. Anantheswaran, J. S. Shenk, M. O. Westerhaus and R. B. Beelman. 1993. Determination of moisture content of mushrooms by vis-NIR spectroscopy. *J. Sci. Food Agric.* 63 : 355-360.

Shenk, J. S. and M. O. Westerhaus. 1991a. Population definition, sample selection and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci.* 31 : 469-474.

Shenk, J. S. and M. O. Westerhaus. 1991b. Population structuring of near infrared spectra and modified partial least squares regression. *Crop Sci.* 31 : 1548-1555.

Velasco, L., J. M. Fernandez-Martinez and A. De Haro. 1997. Determination of the fatty acid composition of the oil in intact seed mustard by near-infrared reflectance spectroscopy. *J. Amer. Oil Chem. Soc.* 74(12) : 1595-1602.

Williams, P. C., K. R. Preston, K. H. Norris and P. M. Starkey. 1984. Determination of amino acids in wheat and barley by near-infrared reflectance spectroscopy. *J. Food Sci.* 49 : 17-20.

Windham, W. R., D. R. Mertens and F. E. Barton. 1989. Protocol for NIRS calibration: Sample selection and equation development and validation. In G. C. Marten et al. (ed.) Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality. *Agric. Handb.* 643. USDA-ARS, Washington, DC.