한작지(Korean J. Crop Sci.), 56(1): 88~93(2011)

Estimating the Important Components in Three Different Sample Types of Soybean by Near Infrared Reflectance Spectroscopy

Ho-Sun Lee*, Jung-Bong Kim**, Young-Yi Lee*, Sok-Young Lee*, Jae-Gyun Gwag*, Hyung Jin Baek*, Chung-Kon Kim*, and Mun Sup Yoon*†

*National Agrobiodiversity Center, National Academy of Agricultural Science, RDA. Suwon 441-717, Korea
**Dep. of Agrofood Resources. National Academy of Agricultural Science, RDA. Suwon 441-857, Korea

ABSTRACT This experiment was carried out to find suitable sample type for the more accurate prediction and non-destructive way in the application of near infrared reflectance spectroscopy (NIRS) technique for estimation the protein, total amino acids, and total isoflavone of soybean by comparing three different sample types, single seed, whole seeds, and milled seeds powder. The coefficient of determination in calibration (R^2) and coefficient of determination in cross-validation (1-VR) for three components analyzed using NIRS revealed that milled powder sample type yielded the highest, followed by single seed, and the whole seeds as the lowest. The coefficient of determination in calibration for single seed was moderately low(R^2 0.70-0.84), while the calibration equation developed with NIRS data scanned with whole seeds showed the lowest accuracy and reliability compared with other sample groups. The scatter plot for NIRS data versus the reference data of whole seeds showed the widest data cloud, in contrary with the milled powder type which showed flatter data cloud. By comparison of NIRS results for total isoflavone, total amino acids, and protein of soybean seeds with three sample types, the powder sample could be estimated for the most accurate prediction. However, based from the results, the use of single bean samples, without grinding the seeds and in consideration with NIRS application for more nondestructive and faster prediction, is proven to be a promising strategy for soybean component estimation using NIRS.

Keywords : calibration, cross validation, protein, total amino acids, total isoflavone

The needs of consumers for agricultural products are now widely diversified with attention focused not only on the major constituents, but also on its other physiological functions.

Soybean is one of the world's most important sources of nutritional and functional components such as protein, amino acids and isoflavone.

A simple and rapid method for estimating their contents is necessary for screening soybean varieties for breeding.

Near infrared reflectance spectroscopy (NIRS) is a rapid and nondestructive technology that does not require chemicals or reagents; it is a multi-element technique and sample measurement using this technique is being successfully implemented throughout the agricultural industry (Schmilovitch *et al.*, 2000). Since the late 1980s, NIRS has been used for measuring the internal composition of biological materials (Fontaine *et. al.*, 2001). Most published studies show that NIRS can accurately estimate the content of several internal components of plants such as dry matter and soluble solids (McGlone *et al.*, 2002), fatty acids (Velasco *et al.*, 1999; Kovalenko *et al.*, 2006), carotenoids (Berardo *et al.* 2009) as well as several inorganic components (Petisco *et al.*, 2005). Recently, the scope of NIRS has been expanded to include the determination of the physiological indices of crops (Jeong *et al.*, 2008) or the classification of normal and artificial aged corn (Min and Kang. 2008), and the evaluation of product quality of herbal medicinal extract (Mohri *et al.*, 2009). For soybean, protein(Choung *et al.*, 2001), amino acids(Fontaine *et. al.*, 2001), fatty acids (Igne *et al.*, 2008), anthocyanin (Kim *et al.*, 2008) and isoflavone (Sato *et al.*, 2008) were also predicted using NIRS.

However many instruments require a 300~500 g sample size to operate. Near-infrared analyses of intact seeds are commonly made on bulk samples of variable size, depending on the instrument and device used. Furthermore, instruments including single-seed sample holders are available.

†Corresponding author: (Phone) +82-31-299-1825
(E-mail) msyoon63@korea.kr <Received February 28, 2011>

In soybean, NIRS is currently used for the analysis of intact seeds and ground seeds powder for functional components. A further improvement in the application of NIRS technique would imply the minimization of sample size and the reduction of destructive use of samples, time and labor in measurement for samples. The analysis of single seeds for the important components using the non-destructive method is the more resonable way when dealing with soybean breeding.

The objective of this work is to find suitable sample types for the more accurate prediction and non-destructive way in the application of NIRS technique in estimating the protein, total amino acids, and total isoflavone in soybean by comparing three different sample types, such as single seed, whole seeds, and milled seeds powder.

MATERIALS AND METHODS

Seed materials

This work was conducted with 142 seed samples (individual accessions) of soybean. The seed materials, conserved at the National Agrobiodiversity Center (genebank), Rural Development Administration(RDA), Korea, were selected from different geographical origins in Korea, China, other subcontinents. The moisture content of seed materials was controled 5-7% as the regulation for seed conservation of medium-term storage in the National Agrobiodiversity Center.

NIRS Analysis

Seed samples were analyzed using three different sample types, such as single seed, whole seeds(10~12 seeds) and milled seeds powder.

The analysis conditions of soybean by near infrared spectroscopy system are summarized in Table 1. Single seed samples were analyzed following the method of Font *et al.*(2004) using the NIRSystems model 6500 spectrophotometer (Foss-NIRSystems, Inc., Silver Spring, MD) equipped with a DCFA module in the reflectance mode. Intact single seed was placed on sample holder with a diameter of 3 mm. Twenty(20) single seeds per one accession were scanned and their average spectra were recorded as individual files at 2nm intervals in the 400~2,500 nm wavelength range. Whole seeds samples were analyzed by using the round cup(outer diameter 5cm, inner diameter 3.5cm) with a quartz window in the same NIRSystem with a spinning module. Milled powder type sample was grounded with a ball mill and sieved with a 1.0 mm and were analyzed with same way to whole seeds type sample.

The parameters of NIRS for equation statistics and calibration are summarized in Table 2. Using the program WINISI II v.1.50 (Infrasoft International, LLC, Port Matilda, PA), different calibration equations for three sample types, protein, total amino acids and total isoflavone were developed on the calibration set(n=142). Calibration equations were computed using the first or second derivative of raw optical data (log1/R, where R is reflectance) with several combinations of segment(smoothing) and derivative(gap) sizes [i.e., (1,4,4,1 which means 1 = number of derivative of spectra, 4 = extent of gap over which the derivative was to be calculated, 4 = the smoothing of point, 1 = second smoothing), (1,8,8,1), (2,4,4,1), (2,10,10,1)] . These parameters in the mathematical processing for different sample types

Table 1. The analysis conditions of soybean by near infrared spectroscopy system (NIR 6500, USA).

Sample type	Instrument		Measure system	Detection mode
	NIR	module		
Single seed	6500	DCFA	intact on the holder	reflectance
Whole seeds	6500	Spinning sample	half cup holder	reflectance
Milled powder	6500	Spinning sample	half cup holder	reflectance

Table 2. Parameter of NIRs for equation statistics and calibration.

Scatter correction	Program	Loding type	Regression
SNV and detrend	WINISI	PCA	Modified PLS

and law data for three components were sought through trial and error in order to maximize the coefficient determination in cross validation and to minimize the standard error of cross validation.

To correlate the spectrum data and the contents of the parameters, modified partial least squares (MPLS) was used as a regression method in the wavelength range 400~2,500 nm. In addition, standard normal variate and detrend transformations (SNV-DT) were used to correct baseline offset due to scattering effects from differences in particle size among samples. The calibration equation were optimized by removing outliers for samples on the following criteria: samples with large residuals values (the difference between the predicted and the actual values) and T-outliers (T>2.5) or H-outliers (H>3).

The different calibration equations obtained in the calibration process were then cross-validated on an internal validation set (25% of total samples randomly taken by software routine) that included outliers removed from the calibration set. The calibration statistics used were standard error of calibration (SEC), standard error of cross-validation(SECV), and coefficient determination of cross-validation(1-VR). The SEC and SECV were calculated as the square root of the mean square for the residuals on the calibration set spectra and cross-validation set, respectively. The SECV is used as the statistic for determining the best number of independent variables for the calibration equation.

Assay of protein, total amino acids, and total isoflavone

All samples that were scanned by NIRS were the same materials used for chemical analysis.

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

Free amino acid(FAA) contents were determined by L-8500 high-speed amino acid analyzer (Hitachi, Japan). The 0.5 g of seed powder was diluted with 3% trichloroacetic acid solution. The sample was left at the room temperature for 1 h, centrifuged at 10,000g for 15 min. The collected supernatant was filtered with Millipore 0.45 μm syringe filters(Milford, USA). The filtrate was loaded on amino acid analyzer. The standard amino acid solutions were obtained from Wako (Wako-shi, Japan).

The isoflavone content was determined by HPLC method (Wong and Murphy, 1994). The respective components such as glucosides (daizdin, glycitin, and genistin), malonyl gulucosides, acetyl glucosides, and aglycons (daidzein, glycitein, and genistein) were also determined using this process. The 500 mg of the powder was extracted with 15mL of 70% (v/v) methanol and 0.1% acetic acid at 20C with vigorous shaking (80 rpm) for 5 h. The pellet obtained after centrifugation at 5000 g for 10 min was extracted again with 1 mL 80% methanol shaking for 2 h and samples were centrifuged at 5000 g for 10 min. The combined supernatants (2 mL) were centrifuged at 12 000 g for 10 min and then used for HPLC analysis. Isoflavones were separated with a Waters HPLC system (Waters Corp., Milford, MA) using a 53 by 7 mm, 3 m EPS C18 Alltech Rocket Column (Alltech Assoc., Deerfield, IL). A linear gradient composed of water included 0.1% acetic acid(solvent A) and acetonitrile included 0.1% acetic acid(solvent B) was used. Following injection of 20 u L of sample, solvent B was increased from 5 to 12% over 7 min and then isocratic elution for 3 min occurred following by an increase in solvent B to 35% over 15 min. The solvent flow was 1.0 mL min^{-1}. A Waters 996 photodiode array detector was used to measure UV absorbance at 254 nm. The 12 isoflavone standards (daidzein, genistein, genistin, glycitein, glycitin, daidzin, malonyldaidzin, malonylgenistin, malonylglycitin, acetylgenistin, acetyldaidzin, acetylglycitin) were purchased from commercial chemical company (Sigma and Fujicco Co.)

RESULTS AND DISCUSSION

Statistics Reference values

Mean values, ranges and standard deviations(SD) for protein, total amino acids and total isoflavones in the samples used in the calibration set are shown in Table 3. In the sample set, there was a comparatively wide variation in these three chemical composition. Total isoflavone content had a great variability, range 121.5-1,070.1 and a SD 183.2. The mean contents of protein and total amino acids of the calibration

Table 3. Laboratory reference value statistics of milled soybean powder for the NIRS calibration set.

Constituent	N	Mean	Range		SD
			Min	Max	
Total isoflavone(mg·100g-1)	142	430.9	121.5	1070.1	183.2
Total amino acids(%)	142	38.6	33.5	43.8	2.25
Protein(%)	142	39.5	34.2	45.5	2.28

N, numbers of accessions; SD, standard deviation of reference data

Table 4. Calibration and cross validation parameters by MPLS regression for NIRS prediction of total isoflavone, total amino acids and protein contents in the calibration set(n=142) of the three sample types in soybean.

Segment	Calibration					Cross-validation		Math treatment
	N	Mean	SD	SEC	R^2	SECV	1-VR	
Total isoflavone ($mg \cdot 100g^{-1}$)								
Single seed	140	475.68	179.0	95.36	0.72	117.58	0.57	1.4.4.1
Whole seeds	138	476.16	171.9	117.79	0.57	137.62	0.34	1.4.4.1
Milled powder	137	472.50	172.5	64.32	0.86	81.94	0.78	1.4.4.1
Total amino acids(%)								
Single seed	141	38.57	2.24	0.89	0.84	1.22	0.72	1.8.8.1
Whole seeds	139	38.54	2.22	1.15	0.73	1.25	0.69	1.8.8.1
Milled powder	139	38.27	2.23	0.69	0.90	0.85	0.87	2.4.4.1
Protein(%)								
Single seed	137	39.62	2.24	1.26	0.70	1.56	0.52	1.4.4.1
Whole seeds	141	39.55	2.35	1.78	0.43	2.06	0.25	1.8.8.1
Milled powder	137	39.52	2.27	0.94	0.83	0.98	0.82	1,4,4,1

N, number of accessions; SD, standard deviation; SEC, standard error of calibration; R^2, Coefficient of determination in calibration; SECV, standard error of cross-validation; 1-VR, coefficient of determination in cross-validation

sample set were 39.5 %(range 34.2-45.5) and 38.6%(range 33.5-43.8) with a SD of 2.28 and 2.25, respectively.

Statistics NIRS calibration

Table 4 summarizes the performance parameters obtained for the calibration equations by using single seed, whole seeds and ground seeds. The result as a whole, showed that the predicted values of total amino acids in three different samples were better correlated (R^2 0.73-0.90) than those of total isoflavone and protein(R^2 0.57-0.86 and 0.43-0.83, respectively). In the protein, whole seeds showed very poor coefficient determination values, R^2 0.43(calibration) and 0.25 (validation), respectively. The coefficient of determination in calibration (R^2) and coefficient of determination in cross-validation (1-VR) for the three components analyzed by NIRS using milled powder was highest(R^2 0.83-0.90 and 1-VR 0.78-0.87, respectively), then single seed, and then whole seeds having the lowest. The R^2 of calibration and 1-VR of cross-validation using single seed showed coefficient of determination with R^2 0.70-0.84 and 1-VR 0.52-0.72, respectively. On the other hand, calibration equation developed with NIRS data scanned with whole seeds showed the lowest accuracy and reliability with R^2 0.43-0.73 and 1-VR 0.25-0.69, respectively, compared with the other sample groups. These results were similar to those reported by other authors with whole seeds and ground powder of cereals(Lee *et al.*

2010). Fontaine *et al.*(2001) have compared calibrations obtained by the same set of ground and whole soybeans and clearly concluded that grinding improves the prediction accuracy. Tkachuk *et al.*(1987) also compared whole pea to ground powder to predict protein more accurately and stated ground pea performed much better. They also stated that the higher SEC and SEP values and the lower R^2 values for whole pea prediction may be largely attributed to sampling error arising from the large particle size of whole peas. Sato *et al.*(2008) also compared whole soybean to ground powder to predict isoflavone using NIRS and suggested that the contents of the constituents could be estimated using whole soybeans which showed quitely high coefficient determination for calibration(0.85), validation(0.82) and prediction(0.82). However, the present results showed that the use of whole seeds was clearly inferior with reference to the NIRS performance. The R^2 coefficient of single seed perimeter was higher than those of whole seeds and showed higher R^2 coefficient than 0.7. This suggests that satisfactory accuracy of NIR predictions may be achieved without grinding the seed samples.

Plots of NIRS data vs laboratory reference data for total isoflavone, total amino acids and protein for the 142 numbered samples of single seed, whole seeds and milled powder are shown in Figure 1. Scatter plots for NIRS data of whole seeds showed the widest data cloud. On the contrary, for milled powder showed flatter data cloud. The flat data cloud which is more closed to calibration curve indicate a significant coefficient between reference values and NIRS estimated values similar to those of calibration(Lee *et al.* 2010).

Agricultural products selectively absorb NIR radiation that yields information about the molecular bonds within the material being measured. Since NIR radiation are scattered between the large particle size of whole soybeans and the accurate information about material is difficult to

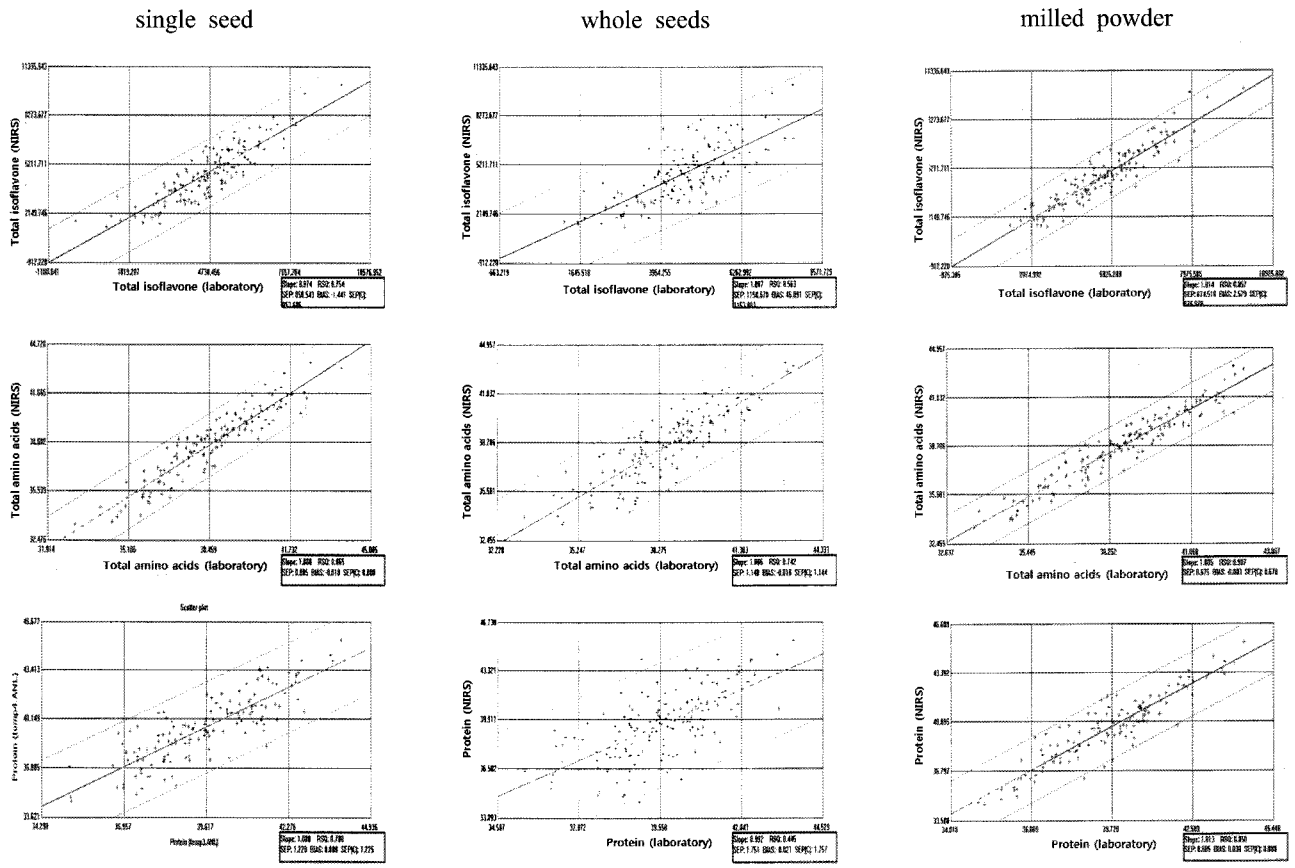

Fig. 1. Scatter plots for NIRS data vs laboratory reference data of total isoflavone, total amino acids and protein contents in the calibration set(n=142) of the three sample types in soybean.

measure particularly for the whole soybeans, thus, no satisfactory results from whole seeds samples in this research may be obtained.

The present findings of the study suggests that total isoflavone, total amino acids and protein content of soybean seeds in the powder could be estimated for the most accurate prediction. However, based from the results, the use of single seed samples, without grinding the seeds and in consideration with NIRS application for more nondestructive and faster prediction, is proven to be a promising strategy for soybean component estimation using NIRS. For the application of single seed sample for NIRS prediction, the calibration models for single seed samples developed in this study need to be updated for increasing the accuracy by applying to new samples with variable constituent characteristics as well as increasing sample size.

LITERATURE CITED

Berardo N., G. Mazzinelli, P. Valoti, P. Lagana, and R. Redaelli. 2009. Characterization of maize germplasm for the chemical composition of the grain. J. Agric. Food Chem. 2009, 57, 2378-2384.

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: 106-111.

Font, R., M.D. Rio, J.M. Fernandez-Martinez, and A. de Haro-Bailon.2004. Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). Journal of Agricultural and Food Chemistry 52: 3563-3569.

Fontaine, J., J. Hrr, and B. Schirmer. 2001. Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal. J. Agric. Food Chem. 49: 57-66.

Igne, B., G.R. Rippke, and C.R. Hurburgh. Measurement of whole soybean fatty acids by near infrared spectroscopy. 2008. J. of Ame. Oil Chemists' Society 85: 1105-1113.

Jeong, J.C., H.C. Ok, O.S. Hurr, and C.G. Kim. 2008. Prediction of sprouting capacity using near-infrared spectroscopy in potato tubers. Am. J. Pot. Res. 89: 309-314.

Kim, Y.H., H.K. Ahn, E.S. Lee, and H.D. Kim. 2008. Development of prediction model by NIRS for anthocyanin contents in black colored soybean. J. Crop. Sci. 53: 15-20.

Kovalenko, I.V., G.R. Rippke, and C.R. Hurburgh. 2006. Measurement of soybean fatty acids by near-infrared spectroscopy: linear and nonlinear calibration methods. J. Am. Oil Chem. Soc. 83: 421-427.

Lee, Y.Y., J.B. Kim, H.S. Lee, S.Y. Lee, J.G. Gwag, H.C Ko, Y.C. Huh, D.Y. Hyun, and C.K. Kim. 2010. Application of near0infrared reflectance spectroscopy (NIR) method to rapid determination of seed protein in coarse cereal germplasm. J. Crop Sci. 55: 357-364.

McGlone, V.A., R.B. Jordan, R. Seelye, and P.J. Martinsen. 2002. Coparing density and NIR methods for measurement of Kiwifruit dry matter and soluble solids content. Postharv. Biol. Technol. 26: 191-198.

Min, T.G. and W.S. Kang. 2008. Nondestructive calssification between normal and artificially aged corn(*Zea mays* L.) seeds using near infrared spectroscopy. J. Crop Sci. 53: 314-319.

Mohri, T. Y. Sakata and M. Otsuka. 2009. Quantative evaluation of glycyrrhizic acid that affects the product quality of kakkonto extract, a traditional herbal medicine, by a chemometric near infrared spectroscopic method. J. Near Infrared Spectrosc. 17: 84-100.

Petisco, C., B. Garcia-Criado, B.R. Vazquez de Aldana, I. Zabalgogeazcoa, S. Mediavilla, and A. Garcia-Ciudad. 2005. Use of near-infrared reflectance spectroscopy in predicting nitrogen, phosphorus and calcium contents in heterogeneous woody plant species. Anal. Bioanal Chem. 382: 458-465.

Sato, T., K. Eguchi, T. Hatano, and Y. Nishiba. Use of near-infrared reflectance spectroscopy for the estimation of the isoflavone contents of soybean seeds. 2008. Plant Prod. Sci. 11: 481-486.

Schmilovitch, Z., A. Mizrach, A. Hoffman, H. Egozi, and Y. Fuchs. 2000. Determination of mango physiological indices by near-infrared spectrometry. Postharv. Biol. Technol. 19: 245-252.

Tkachuk, R., F.D. Kuzina, and R.D. Reichert. 1987. Analysis of protein in ground and whole field peas by near-infrared reflectance spectroscopy. Cereal Chem. 64: 418-422.

Velasco, L., M. Christian, and C. Heiko. 1999. Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. Becker. Euphytica 106: 79-85.

Wong, H. and P.A. Murphy. 1994. Isoflavone content in commercial soybean food. J. Agric. Food Chem. 42: 1674-1677.