16

Application of Nondestructive Measurement to Improve Soybean Quality by Near Infrared Reflectance Spectroscopy

Jeong-Dong Lee[1], J. Grover Shannon[2] and Myoung-Gun Choung[3]
[1]*Division of Plant Biosciences, Kyungpook National University, Daegu 702-701*
[2]*Division of Plant Sciences, University of Missouri-Delta Center, Portageville, MO 63873*
[3]*Dept. of Herbal Medicine Resource, Kangwon National University, Samcheok 245-722*
[1,3]*Republic of Korea*
[2]*USA*

1. Introduction

Soybean [*Glycine max* (L.) Merr.] is the world's primary protein and oil source for human and animals. Soybean seeds average 40% protein, 35% carbohydrate, 20% oil, and 5% ash. Soybean is not only an essential and dominant source of nutrition for humans and animals but it also has numerous other uses and is the leading source for biodiesel. Soybean represented 56% of the world's vegetable oil seed production, and soybean oil was second only to palm (Arecaceae: *Elaeis*) oil in consumption (http://www.soystats.com). Recent studies indicate that consumption of soybean reduces cancer, blood serum cholesterol, osteoporosis and heart disease (Birt et al., 2004). Also, soybeans are a good source of minerals, vitamins, folic acid, and isoflavones which are credited with slowing the development of these diseases (Wilson, 2004). Thus, the demand for edible soybean products has increased dramatically. Also, the desire for more meat in diets among the world's population has increased, consequently the demand for soybean protein for livestock and poultry feed has increased. In addition to feed and food, soybean has numerous industrial applications such as building materials, plastics, printing inks, paints, hydraulic fluids, cosmetics, pharmaceuticals and soy-diesel fuel that burns cleaner and pollutes less than petroleum derived fuels.

The increased importance of soybeans as a world crop has led to a huge expansion in world soybean production (http://www.soystats.com). In the last twenty years, world soybean production has increased steadily from 70 million tons in 1984 to 220 million tons in 2008. About 80% of the world supply was produced in North and South America. The United States, Brazil and Argentina were the major producers and exporters of soybean. In 2008, among these countries, the United States was the leading soybean producer at 73 million tons or about 33% of the total world production. At the same time Brazil and Argentina produced about 61 (28%) and 46 (21%) million tons, respectively (http://www.soystats.com). Although soybean is native to China, China produced 14 million tons (6% of the total) and India produced about 9 million tons (4% of the total). The remaining 5% was produced in countries of Asia and Europe. In 2008 total oilseed

production was 391 million tons of which 56% (220 million tons) was from soybean making it the world's number one oil seed crop followed by rapeseed and cotton seed at 12% each.
Improving soybeans for various uses is a goal of scientists. Soybeans high in protein quantity and quality for soy-foods and animal feeds; high oil content with altered fatty acid profile for a healthier and more functional oil for food applications and biodiesel; high isolfavone content for lower cancer risk and other human health benefits; low phytic acid content seed for improved digestible phosphorus in animal feeds to reduce phosphorus pollution in the environment; high sucrose content for vegetable soybean; small seed for natto and soybean sprouts; big seed for tofu and soy paste, lipoxygenase free seed for soymilk; colored seed coat or cotyledon for cooking with rice; and other altered minor seed components will lead to greater market demand for soybeans.
Improving seed quality and agronomic traits in soybean has been and continues to be a goal of soybean research programs including soybean breeding. Measurement of soybean seed for various seed components is often difficult and expensive. Complicated, time consuming techniques to determine seed components is a limiting factor in improvement programs because the number of genotypes which can be evaluated is limited. Soybean scientists have used wet chemistry methods to measure various seed components in soybean or soybean products. Wet chemistry is the most accurate way to measure the levels of seed components. However, this method requiring destruction of soybean seed or products, is time consuming, and is too slow and labor intensive for soybean improvement programs when many samples have to be screened. Easier and effective determination of soybean genotypes such as using Near Infrared reflectance (NIR) for mensuring the promising characteristics mentioned above would greatly enhance progress in improving soybean for important seed components.
NIR was discovered by Friedrich Wilhelm Herschel in 1800 (Davies, 2000) and covers the range of the electromagnetic spectrum from 780 to 2500nm. In NIR spectroscopy, the product (such as soybean seed) is irradiated with NIR, and the reflected or transmitted radiation is measured. While the radiation penetrates the product, its spectral characteristics change through wavelength dependent scattering and absorption processes. This change depends on the chemical composition of the product, as well as on its light scattering properties which are related to the microstructure (Fig. 1).
Advanced multivariate statistical techniques, such as partial least squares regression are then applied to extract the required information from the usually convoluted spectra. The most attractive merit using NIR spectroscopy is to determine chemical composition of samples without any destruction. Since its discovery NIR spectroscopy has been applied to determine various chemical components in seed, plants, and food from many crops.
In soybean, NIR spectroscopy has been used to determine chemical composition of seed, and of other products and for evaluating soybean genotypes for various seed components. The increasing the importance of NIR spectroscopy in soybean seed composition improvement programs; the principles of NIR spectroscopy; applications for soybean seed composition improvement; and limitations of NIR technology and future research required to improve soybean seed components using NIR will be discussed in this chapter.

2. The general principles and procedures for NIR analysis in soybean

2.1 Determine soybean seed compositions that are needed by NIR application

As mentioned earlier, soybean has many useful chemical components such as protein, oil, fatty acids in soybean oil, isoflavones, sucrose, and carotenoids. If the target component has

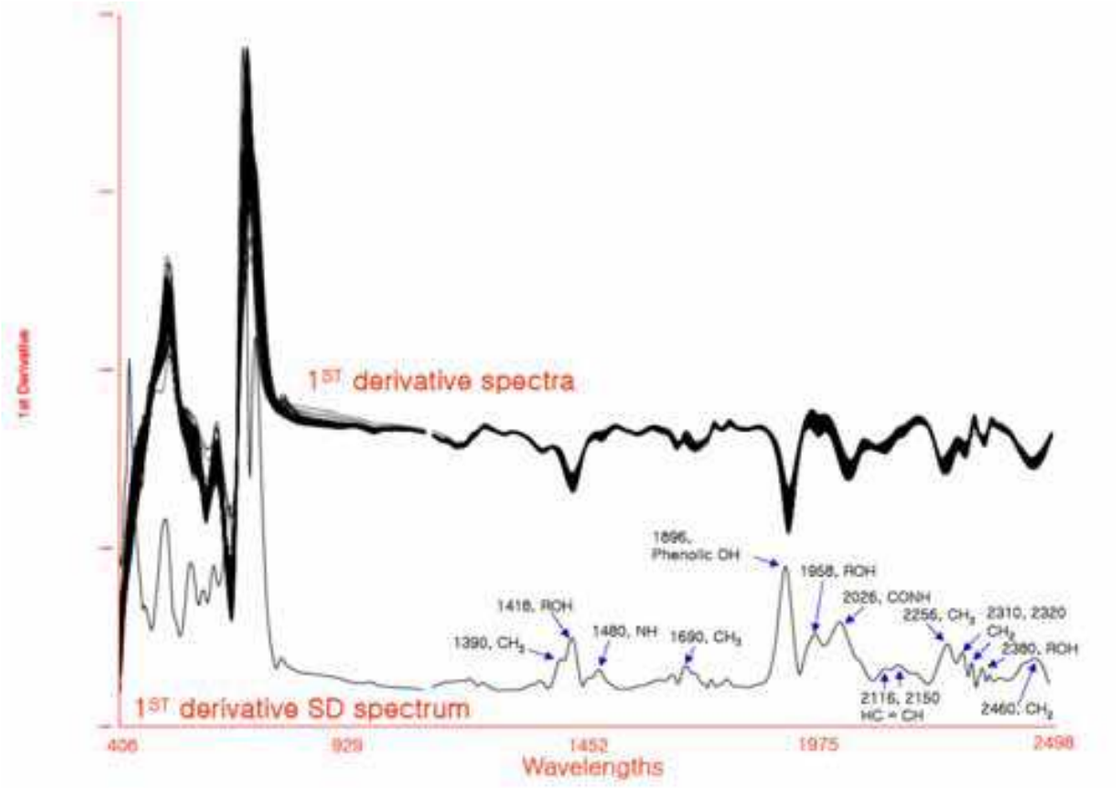


Fig. 1. First derivate (1,4,4,1) spectra and standard deviation (SD) spectrum of soybean samples (unpublished data)

A wide range of variation in soybean populations, it is relatively easy to make reliable calibration curves to estimate the amount of a target component by NIR spectroscopy. There is a wide range of genetic variation for chemical concentration among accessions of the soybean germplasm (Table 1). Two major seed components, protein and oil show a wide range of variation. However some minor seed components such as cysteine, methionine, tryptophan, and stearic acid have a relatively narrow range of variation which can limit use of NIR for measuring these traits among soybean genotypes. Good calibration curves are available for NIR spectroscopy to estimate concentration of soybean components such as protein, oil, sucrose etc. The details for measuring seed components using NIR are presented below.

2.2 Spectra collection and pretreatment

An NIR spectrophotometer consists of a light source (usually a tungsten halogen light bulb), sample presentation accessory, monochromator, detector, and optical components, such as lenses, collimators, beam splitters, integrating spheres and optical fibers. Spectrophotometers are conveniently classified according to the type of monochromator (Nicolai et al., 2007). In a filter instrument, the monochromator is a wheel holding a number of absorption or interference filters. Its spectral resolution is limited. In a scanning monochromator instrument a grating or a prism is used to separate the individual frequencies of the radiation either entering or leaving the sample (Fig. 2).

The wavelength separator rotates so that the radiation of the individual wavelengths subsequently reaches the detector. Four more spectrometer types, Fourier transform (FT)

spectrophotometers, Photodiode array (PDA) spectrophotometers, Acoustic optic tunable filter (AOTF) instruments, and Liquid crystal tunable filter (LCTF) instruments are available for NIR analysis (Nicolai et al., 2007). In soybean, many NIR studies for determination of soybean components and other traits have been conducted using scanning monochromator spectrophotometer with reflectance acquisition mode (Table 5).

Seed compositions	Number of accessions	Range	Source
Protein	5530	32.5-55.9%	GRIN
Agrinine	5530	5.0-9.8%	GRIN
Cysteine	5530	1.10-2.67%	GRIN
Isoleucine	5530	2.30-6.40%	GRIN
Leucine	5530	6.50-9.20%	GRIN
Lysine	5530	2.50-7.70%	GRIN
Methionine	5530	1.50-2.80%	GRIN
Threonine	5530	2.80-4.30%	GRIN
Tryptophan	5530	0.90-1.70%	GRIN
Valine	5530	4.10-6.90%	GRIN
Oil	5530	8.9-25.6%	GRIN
Palmitic acid	5530	4.1-14.0%	GRIN
Stearic acid	5530	3.0-5.5%	GRIN
Oleic acid	5530	17.1-37.7%	GRIN
Linoleic acid	5530	46.0-58.7%	GRIN
Linolenic acid	5530	3.6-12.4%	GRIN
Sucrose	5483	0.0-11.3%	GRIN
Stachyose	5522	0.8-8.1%	GRIN
Lutein	490	1.6–14.8ug/g	(Kanamaru et al. 2006).
Total isoflavone	1296	278.4 -2,736.9µg/g	(Han et al., 2008).
Daidzein	1296	48.8 - 1,709.6µg/g	(Han et al., 2008).
Glycitein	1296	0.98 - 892.3µg/g	(Han et al., 2008).
Genistein	1296	79.8 - 1242.3µg/g	(Han et al., 2008).

USDA, ARS, National Genetic Resources Program. *Germplasm Resources Information Network - (GRIN)*. [Online Database] National Germplasm Resources Laboratory, Beltsville, Maryland. Available: http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?494186 (30 August 2010)

Table 1. Wide range of genetic variation for some seed components in soybean seed.

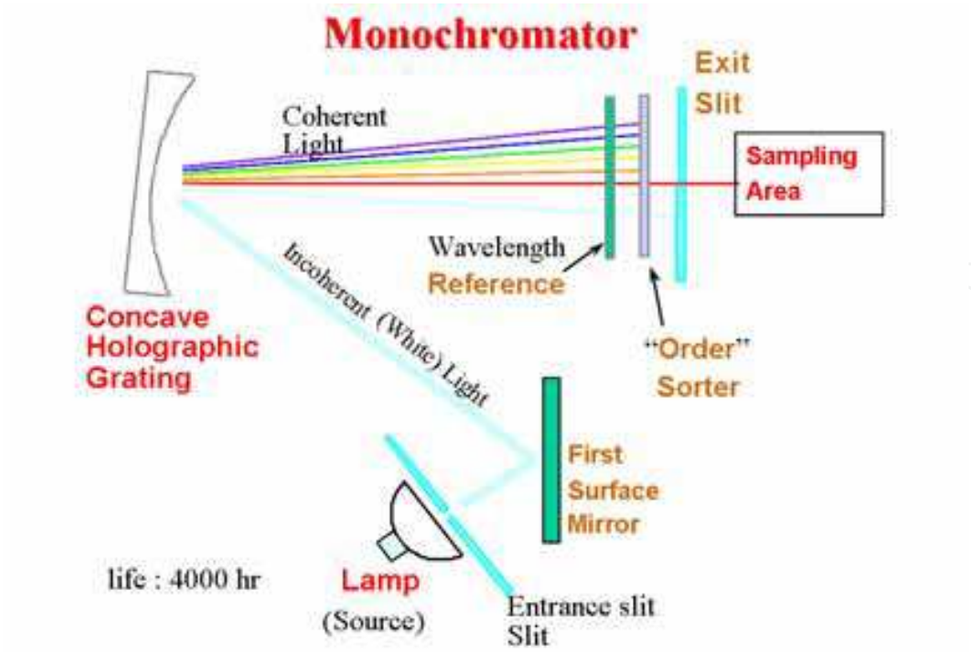


Fig. 2. General schematic of monocharomator.

2.2.1 Measurement setup

Three different measurement setups for obtaining near infrared spectra are shown in Fig. 3. In the reflectance, the light source, and detector are mounted at under a specific angle to avoid specular reflection. In the transmittance mode the light source is positioned opposite to the detector, while in the interactance mode the light source and detector are positioned parallel to each other in such a way that light due to specular reflection cannot directly enter the detector.

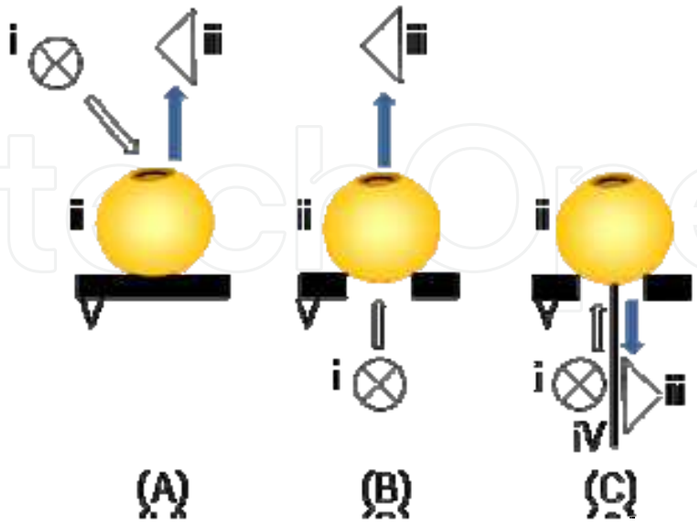


Fig. 3. Setup for the acquisition of (A) reflectance, (B) transmittance, and (C) interactance spectra, with (i) the light source, (ii) soybean, (iii) monochromator/detector, (iv) light barrier, and (v) support (modified from Nicolai et al., 2007).

2.2.2 Spectra cllection

The general procedures from collection spectra from a soybean sample to validation of results for a developed equation was described (Choung et al., 2001b; Kim et al., 2007; Choung, 2010) with minor modifications. The NIR spectroscopic analysis was performed using a NIRSystem model 6500 near-infrared scanning monochromator (Foss NIRSystems Inc., Silver Spring, MD) in the reflectance mode. Intact seed samples, single seed, bulk seeds and flour, were placed in a standard ring cup and scanned (Fig. 4).

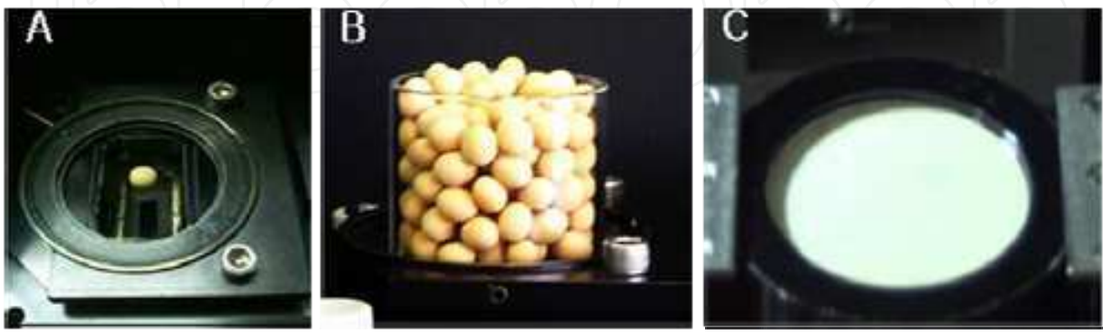


Fig. 4. General shape of soybeans, single seed (A), bulk seeds in cup (B), and soybean flour (C) for spectra collection.

Reflectance energy readings were referenced to corresponding readings from an internal ceramic disk (Fig. 5). Each spectrum was recorded from each sample, and the average of at least 16 successive scans was recorded. As a control, 16 scans over the standard ceramic disk were made before and after the samples were scanned.

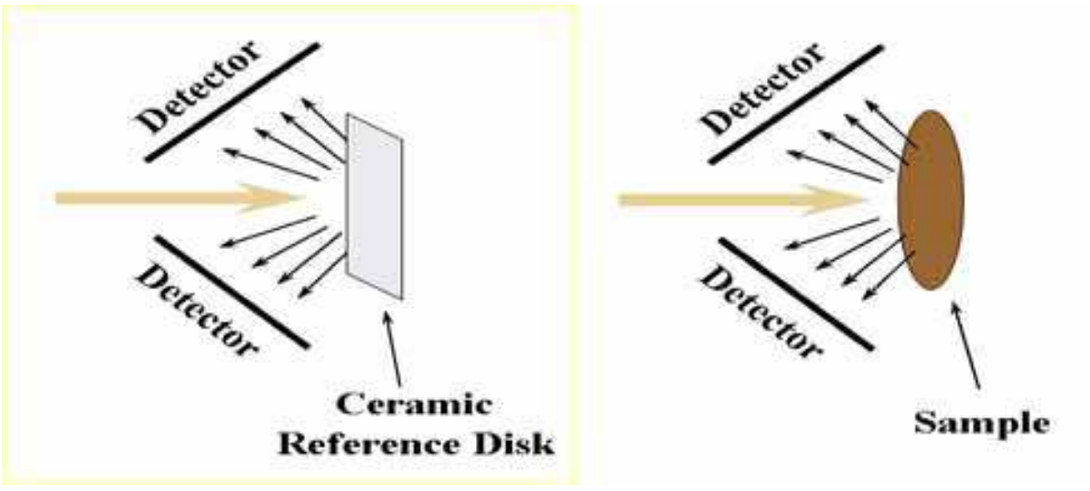


Fig. 5. Schematics for reflectance reading by the reference disk (left) and the sample (right)

Typical NIR reflectance spectra of single seed, bulk seed and flour of soybean are shown Fig. 6. All spectral data were recorded as the logarithm of the reciprocal of reflectance (log $1/R$) in the wavelength range from 400 to 2500 nm at 2 nm intervals to give a total of 1050 data points per sample. Absorption of radiation in the region of 400-2500 nm, the visible plus near-infrared region, was used to develop calibration equations related to sample properties. The scanning procedure could be completed in 1.5 min per sample, once the

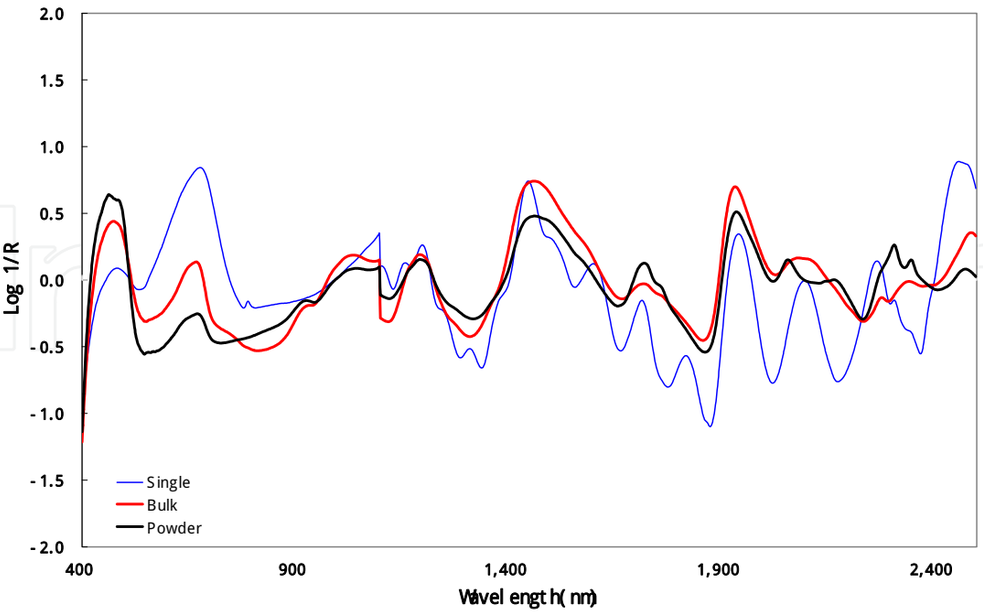


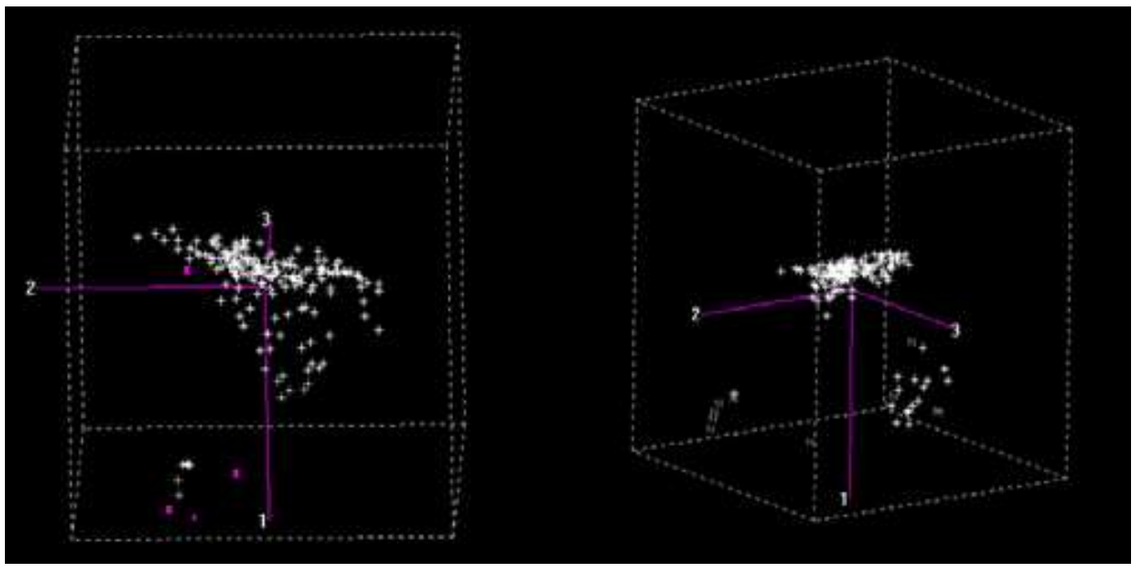
Fig. 6. Typical NIR reflectance spectra of single seed, bulk seed and flour of soybean

NIRS instrument was warmed up, and the stability of NIRS through photometric repeatability (noise test) and wavelength accuracy test was confirmed. The NIRS manipulation for scanning, mathematical processing, and statistical analysis was performed with the WinISI II software (Windows version 1.60, Foss and Infrasoft International LLC, State College, PA). In WinISI software, the Score program was used to select samples for spectrum outliers and samples to represent the entire sample set before calibration and validation. The distance between a sample and its neighbor was measured as the *H* distance and was used as a criterion for selecting those samples representing the calibration and validation sets. The Score algorithm ranks spectra according to Mahalanobis distance (*H* distance) from the average spectrum, gives spectral boundaries to eliminate outliers with $H > 3.0$, and to eliminate samples with similar spectra with $H < 0.6$ from neighboring samples for the development of an accurate and robust prediction equation (Fig. 7).

The final number of samples for calibration and validation was variable and based on the cutoff point of *H* distance, depending on the spectral and chemical variability of samples in the population used for NIRS estimation. The samples were randomly split into two sets using the WinISI program (Table 2). The calibration set was used to calibrate and cross-validate the derived equation, and the other samples were used as an external validation set to test the fit of the developed equations.

Sample set		n	Mean (%)	Range (%)	SD
Calibration	Protein	189	43.23	36.04~51.83	3.24
	Oil	189	19.15	14.57~24.05	2.05
Validation	Protein	103	42.13	36.17~49.83	3.47
	Oil	103	20.06	14.88~23.38	2.03

Table 2. Laboratory reference value statistics for protein and oil content based on ground soybean seed samples (Choung et al. 2001).



H-distance of 3.0 *H*-distance of 0.6

Fig. 7. Examples of eliminating outliers with $H > 3.0$, and $H < 0.6$ for the development of an accurate and robust prediction equations.

3. Data pocessing

The equations for NIRS prediction were developed using the Global program in WinISI software with modified partial least-squares (MPLS) regression using wavelengths of the entire visible (400-1100 nm) and near-infrared (1100-2500 nm) regions at every 8 nm. In addition to MPLS, regression methods such as PLS (partial least squares), principal component regression, and multiple linear regression were tested to develop calibration for soybean seed composition. Various mathematical treatments using the raw optical spectrum (log $1/R$), or first or second derivatives of the $1/R$ data, were applied for calibration equation development. For example, in 2, 10, 10, and 1, the first number 2 indicates the order of the derivative (two is the second derivative of log $1/R$), the second number 10 is the gap in data points over which the derivative was calculated, and the third and fourth numbers as 10 and 1, represent the number of data points used in first and second smoothings, respectively. The application of the second-derivative algorithm to the raw spectra (log $1/R$) resulted in an increase in the complexity of spectra and a clear separation between peaks, which overlapped in the raw spectra.

In addition to no scatter correction (log $1/R$), scatter corrections using the standard normal variate and detrending (SNVD) transformation were evaluated for the calibration. The SNVD was designed to remove additive baseline and multiplicative signal effects resulting in a spectrum with zero mean and a variance equal to one. Application of SNVD transformation to raw spectral data reduces the differences in spectra related to physical characteristics such as particle size and path length of samples.

Calculated calibration statistics included the standard error of calibration (SEC), the coefficient of determination (R^2), and the standard error of cross-validation (SECV). The performances of the different equations obtained in the calibration were determined from cross-validation as an internal validation method. Internal cross validation was used

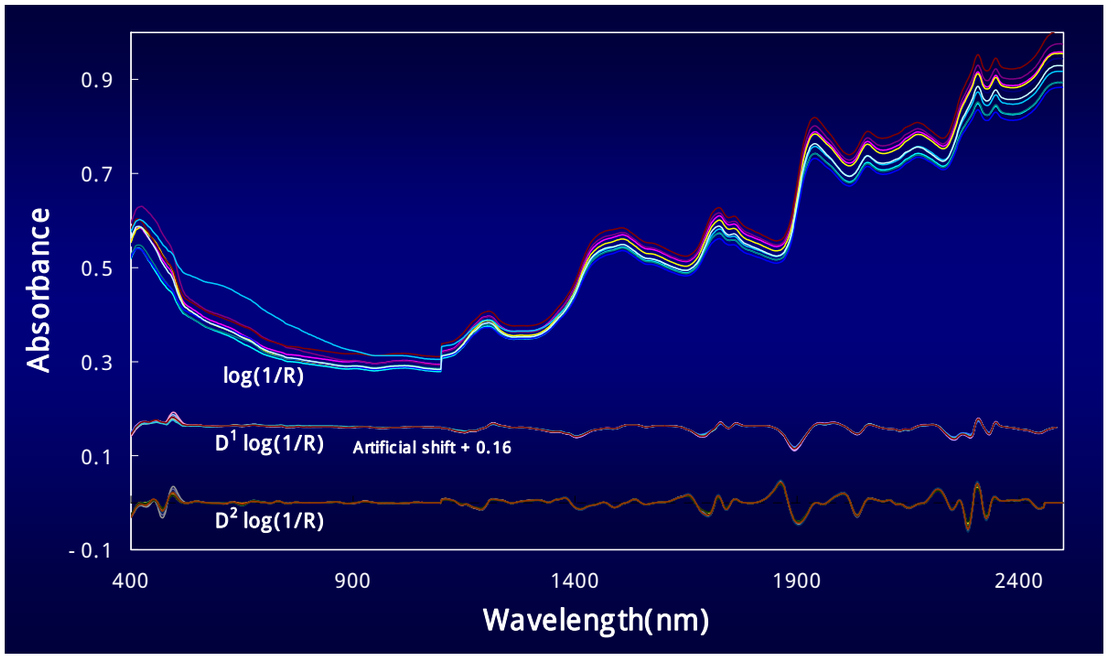


Fig. 8. The raw optical spectrum (log $1/R$), or first or second derivatives of the $1/R$ data (unpublished data)

to avoid over fitting the equations by selecting the minimum number of PLS terms in each model. The best predicted equations for each chemical component were selected on the basis of minimizing SECV and increasing R^2. Two passes to eliminate outliers were set by two outlier detection methods, t and H statistics in WinISI software. The t statistics identified outliers having residuals from reference analysis of >2.5 times the SEC. Outliers indicated that their reference values were in doubt and that the samples were in different populations due to atypical spectra. The ratio (SD/SECV) of the standard deviation of reference data (SD) to SECV, designated RSC, was calculated as a criterion for evaluating the performance of calibrations.

Constituent	Math treatment	Terms [a]	Calibration		Cross-validation		
			SEC [b]	R^2 [c]	1-VR [d]	SECV [e]	RSC [f]
Sucrose	2,10,10,1	6	0.220	0.941	0.921	0.255	3.29
Raffinose	0,0,1,1	2	0.107	0.367	0.344	0.109	1.01
Stachyose	2,8,6,1	9	0.134	0.730	0.539	0.175	1.27
TSCs	2,10,10,1	7	0.210	0.946	0.912	0.268	3.07

[a] Number of PLS loading factors in the regression model MPLS (modified partial least-squares). [b]SEC, standard error of calibration. [c]R^2, coefficient of determination of calibration. [d]1-VR, one minus the ratio of unexplained variance divided by variance. [e]SECV, standard error of cross-validation. [f]RSC, SD/SECV, the ratio of SD (standard deviation of reference data) to SECV in the calibration set.

Table 3. Equation development statistics using MPLS and scatter correction for the NIRS prediction of soluble carbohydrate contents in soybean seeds (Choung, 2010).

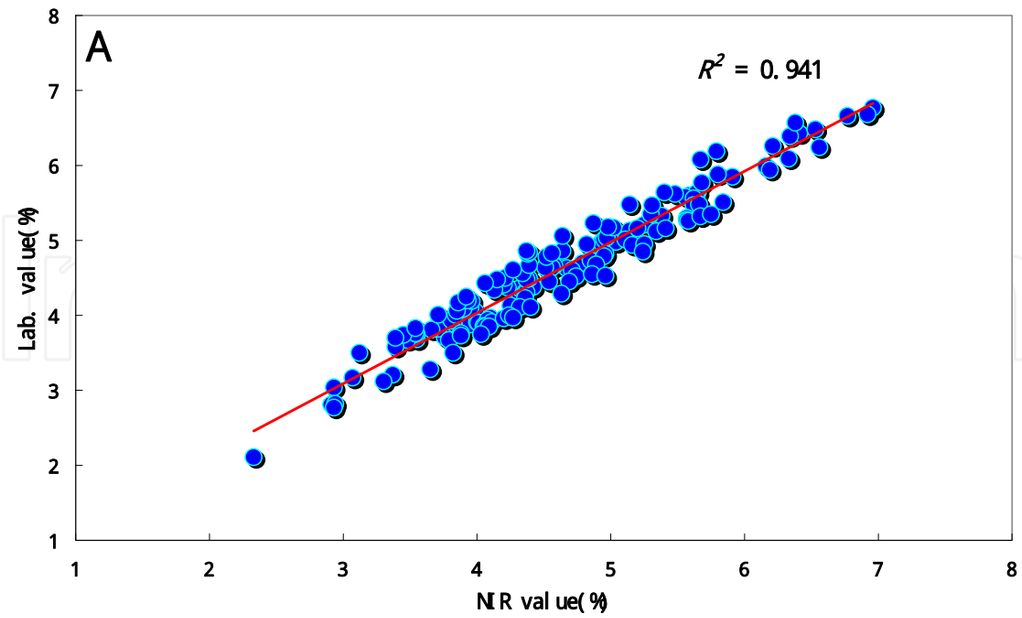


Fig. 9. Scatter plots of sucrose content in soybean seed samples by HPLC vs. by NIRs for the calibration

After calibration, the developed regression equations allowed for accurate analysis of many other samples by prediction of data based on the spectra. In addition to the internal cross-validation, the external validations of calibration models were tested for the prediction capacity on the basis of the standard error of prediction (SEP) and the coefficient of determination in prediction (r^2) (Table 4 and Fig. 9). The ratio of SD for the validation samples to the corrected SEP (designated RSP) was also used as a criterion to evaluate the accuracy of the equations. This RSP value as cutoff point was 3.0 in this study, which is the value recommended for screening purposes. The validation sample set allowed the NIRS equation to be validated for prediction accuracy, using random samples not included in the calibration sample set. The equations for selected chemical composition in intact seeds of soybean were monitored with the Monitor program in WinISI software, using the validation set.

Constituent	Mean [a]	SD [b]	Bias [c]	r^2 [d]	SEP(C) [e]	RSP [f]
Sucrose	4.59	0.899	-0.060	0.921	0.257	3.50
Raffinose	0.75	0.080	0.026	0.311	0.124	0.65
Stachyose	2.33	0.203	0.002	0.443	0.226	0.89
TSCs	7.61	0.888	0.029	0.934	0.232	3.83

[a]Samples used to monitor the model. [b]SD, standard deviation of mean. [c]Bias, average difference between reference and NIRS values. [d]r^2, coefficient of determination of cross-validation. [e]SEP(C), the corrected standard error of prediction. [f]RSP, SD/SEP(C), the ratio of SD of reference data to SEP(C) in the external validation set.

Table 4. Validation statistics for soluble carbohydrate components in soybean seeds (Choung, 2010).

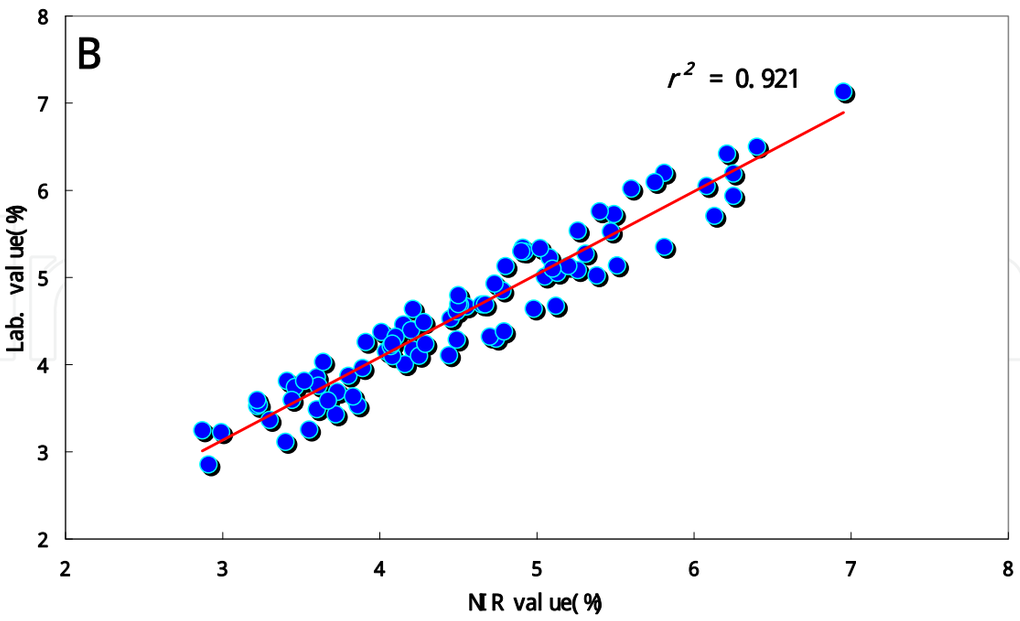


Fig. 10. Scatter plots of sucrose content in soybean seed samples by HPLC vs. by NIR for the external validation sample sets (Choung, 2010).

4. Applications

An overview of applications of NIR spectroscopy to measure the seed composition and products for soybean improvement is given in Table 5. Only information where at least an internal validation has been conducted is included in this chapter. Protein (40%) and oil (20%) are the dominant components of value in soybean seed. Therefore, NIR has been widely used for evaluation of protein and oil in comparison to other seed components (Table 5). In ground soybean flour, protein and oil contents have been accurately estimated using NIRs (Pazdernik et al., 1997; Choung et al., 2001b). However estimation of other seed components in soybean flour using NIR remains labor intensive compared to using non-destructive methods with whole seeds. Therefore, investigators continue to accurately develop and use NIR techniques for measurement of seed components by scanning bulked seed (Fig. 4). For example, Choung et al. (2001a) used 310 soybean samples to develop calibration curves for protein and oil contents in soybean. They reported that validation of NIRs equations showed very low bias (protein: 0.016%, Oil: -0.017%), stand error of prediction (protein 0.568%, oil: 0.451%) and very high coefficient of determination (r^2 Protein: 0.927, oil: 0.906). Presently, NIR spectroscopy in which whole seed can be analyzed without destruction is the method of choice for protein and oil analysis not only in soybean research (Helms & Orf, 1998; Sebolt et al., 2000; Palomeque et al., 2010, Tajuddin et al., 2003, Geater & Fehr, 2000) but also in soybean breeding programs. Estimation of seed composition with NIR using bulked soybean seed is very useful in soybean studies, however several grams of soybean seed are needed for measurements. Estimation of seed components based on analysis of single seeds would be useful. There are interesting reports using single soybean seed for prediction of seed composition in soybean. Tajuddin et al. (2002) developed calibration equations to predict protein and oil content in single soybean seed.

They used single F_8 seed to predict protein and oil content in the F_9 generation and developed good calibration equations with 0.04 – 0.07% bias, 1.32 – 1.57% for low standard error of prediction, and 0.88 – 0.87 of coefficient of determination for protein, and -0.09 – -0.14% bias, 1.06 – 1.37% for standard error of prediction, and 0.72 – 0.80 for coefficient of determination for oil. Recently, NIR spectroscopy to analyze seed protein in single soybean seed was developed by Choung et al. (2004) with high coefficient of determination for protein (r^2=0.955) and oil (r^2 = 0.920) between analyses comparing wet chemistry and NIR spectroscopy. This NIR method was applied to single seeds in segregating soybean populations by Lee et al. (2010). They developed two segregating populations from crosses between high protein and low protein parents. Protein of parents and F_2 seeds were estimated by NIR spectroscopy to select high and low protein F_2 seeds from the two populations. The selected F_2 high and low protein seeds were planted to produce F_3 seed. Means and ranges of F_3 seeds selected from high protein F_2 seeds were higher in protein than F_3 seeds from low protein F_2 seeds. This indicates that analysis of single F_2 seed for protein content using NIR spectroscopy was effective in selecting for increased protein in the F_3 generation (Fig. 11).

Soybean represented 56% of the world's vegetable oil seed production, and soybean oil was second in world oil consumption (www.soystats.com/ 2009). Soybean oil is primarily composed of five fatty acids, palmitic, stearic, oleic, linoleic and linolenic acids. Levels of

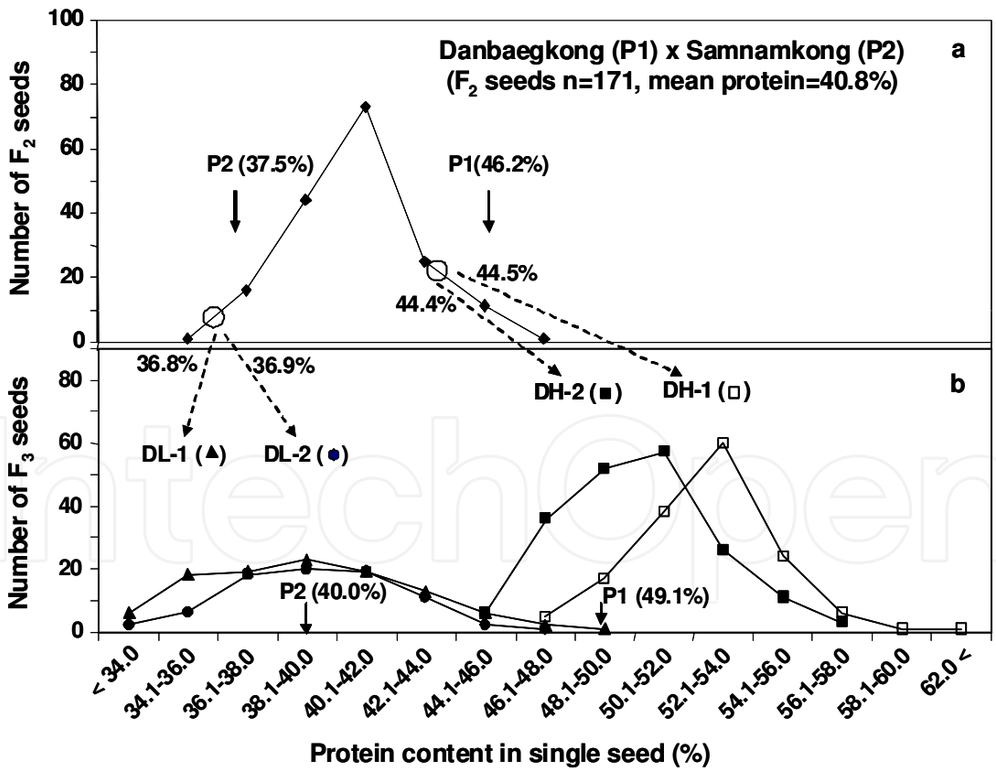


Fig. 11. Protein distribution of F_2 seeds and parents of population Danbaegkong x Samnamkong (a), and of F_3 seeds produced from a plant from a selected F_2 seed (b). F_3 seeds of DL-1, DL-2, DH-1, and DH-2 were from F_2 seed with 36.8, 36.9, 44.5, and 44.4% protein content, respectively (Lee et al., 2010).

each fatty acid in the oil have specific benefits for uses in various food applications with different consequences for human health (Lee et al., 2007). Therefore, a quick and easy metod to determine of fatty acid concentration in soybean seed is very important to modify fatty acid levels in breeding programs. Sato et al. (2002) developed reasonable calibration equations for oleic acid (r^2=0.877) and linoleic acid (r^2=0.853) to determine fatty acid concentration in soybean seed. However, Choung et al. (2005) developed good calibration equations for oleic acid (r^2=0.974), linoleic acid (r^2=0.967), and linolenic acid (r^2=0.884). They suggested that these NIRs equations have potential for use in screening for unsaturated fatty acid content in soybean seed oil. Kovalenko et al. (2006) also reported that NIR equations for total saturates had high predictive ability (r^2=0.91-0.94), however unsaturated fatty acids showed low predictive abilities, oleic (r^2=0.76-0.81), linolenic (r^2=0.73-0.76), and linolenic (r^2=0.67-0.74).

NIR spectroscopy also has been used determine soybean oil quality such as determination of degradation in frying oil (Yildize et al., 2001; Gerde et al., 2007), peroxide value in oxidized soybean oil (Yildiz et al., 2003), and *cis* and *trans* content in hydrogenated soybean oil (Li et al., 1999). Determination of minor seed composition by NIR spectroscopy in soybean seed or soybean products was also reported. Calibration equations have been developed for amino acids (Fontain et al., 2001), for inorganic phosphorus by single-seed transmittance spectra (r^2=0.276), 24-bean average transmittance spectra (r^2=0.598) and meal reflectance spectra (r^2=0.860) (Delwiche et al., 2006), determination of nonstarch polysaccharides (Hollung et al., 2005), sucrose content (r^2=0.941), raffinose content (r^2=0.344), and stachyose content (r^2=0.730) (Choung, 2010), total isoflavone content (r^2=0.95) from soybean flour (Sato et al., 2008), and anthocyanin , C3G (r^2=0.952), D3G (r^2=0.936), and Pt3G (r^2=0.8330) (Kim et al., 2008).

Interestingly, researchers used NIR spectroscopy to discriminate among levels of desease on soybean plant parts such as soybean seed, pods, etc. Wang et al. (2004) classified fungal-damaged soybean seeds by NIR spectroscopy and reported that classification accuracies of the validation sample set were 100, 99, 84, 94, and 95% for *Phomopsis*, *Cercospora kikuchii*, soybean mosaic virus, and downy mildew damaged seeds, respectively in comparison to healthy seeds. Sirisomboon et al. (2009) investigated NIR spectroscopy to identify defective pods for green soybean processing and reported that the good pod model created by primary spectra correctly classified 77.2% of samples as good pods or defective pods. Jinendra et al. (2010) reported a novel approach for rapid in vivo diagnosis of virus infected soybean by NIR spectroscopy and aquaphotomics results showed that the developed spectral calibration model can predict non-infected and infected soybean 96% and 92% of the time, respectively. Choung et al. (2009) developed a good calibration model for NIR technology that can discriminate between soybean seed with and without the Round-up ready herbicide resistance trait with high predictability (r^2=0.95). Choung et al. (2009) also developed NIR calibrations to accurately predict (r^2=0.945) seed weight of soybean.

5. Limitations of NIR and future research

Improving the quantity and quality of soybean seed composition and soybean products has long been an objective of soybean researchers. The most attractive merit using NIR spectroscopy is accurate non-destructive analysis of seed samples for various chemical traits. In soybean, the feasibility of NIR spectroscopy to measure quality attributes

Traits	Source	Spectrometer	Acquisition mode	Spectra range	References
Anthocyanin	Seed coat flour	Scanning	Reflectance	400-2500 nm	Kim et al., 2008
Fatty acids	Soybean flour	Scanning	Reflectance	400-2500 nm	Choung et al., 2005
Fatty acids	Soybean flour	Scanning	Reflectance	1100-2500 nm	Sato et al., 2002
Fatty acids	Soybean flour	Fourier Transform		4000-12500 cm^{-1}	Sun et al., 2008
Fungal-damaged seed	Single seed	Scanning	Reflectance	400-1700 nm	Wang et al., 2004
Glycine (11s) and β-congrlycinin (7s)	Soybean flour and seeds	Scanning	Transmittance Reflectance	800 - 1798 nm 1100 -2489 nm	Delwiche et al., 2007
Good soybean pod	Fresh soybean pod	Scanning	Reflectance	600-1100 nm	Sirisomboon et al., 2009
Inorganic phosphorus	Single seed and soybean meal	Scanning	Transmittance Reflectance	600-1898 nm 1100-2498nm	Delwiche et al., 2006
Isoflavone	Soybean flour and seeds	Scanning	Reflectance	1100-2500 nm	Sato et al., 2008
Lecithin	Soybean lecithin	Fourier Transform	Interactance	4000-12500 cm^{-1}	Li et al., 2009
Oil degradation	Soybean oil	Scanning	Reflectance	350-2500 nm	Gerbe et al., 2007
Oil quality	Soybean oil	Fourier Transform	Interactance	4000-10000 cm^{-1}	Li et al., 1999
Oil quality (oxidation)	Soybean oil	Scanning	Transmittance	400-2500 nm	Yildiz et al., 2001
Oligosaccharide and nonstarch polysaccharide	Soybean flour and seeds	Scanning	Reflectance	400-2498 nm	Hollung et al., 2005
Peroxide value	Soybean oil	Scanning	Transmittance	400-2500 nm	Yildiz et al., 2003
Protein	Single seed	Scanning	Reflectance	400 -2500 nm	Choung et al., 2004
Protein and Oil	Soybean flour	Scanning	Reflectance	400-2500 nm	Choung et al., 2001b
Protein and Oil	Seed	Scanning	Reflectance	400-2500 nm	Choung et al., 2001a
Protein and Oil	Single seed	Scanning	Transmittance	700-1100 nm	Tajuddin et al., 2002
Protein, amino acids	Soybean flour	Scanning	Reflectance	1100-2500 nm	Fontaine et al., 2001
Soybean meal content	Soybean meal	Scanning	Reflectance	950-1650 nm	Li et al., 2007
Sucrose	Soybean flour	Scanning	Reflectance	400-2500 nm	Choung, M.G., 2010

Table 5. Overview of applications of NIR spectroscopy to evaluate soybean for levels of various important chemical components and traits.

has been shown for many products (Table 5) such as soybean seed composition, discrimination of pest infected or non-infected soybean seeds or pods, and various soybean products. It is clear that the accuracy of the NIR calibration models should be sufficient for analysis of various seed components and soybean; however there are several important issues associated with NIR spectroscopy for accurately measuring important traits for soybean improvement.

First, it is hard to develop good calibration equations for minor seed components with relatively low phenotypic variation. For example, depending on researchers, two saturated fatty acids, palmitic acid and stearic acid showed relatively low coefficient of determination than unsaturated fatty acids. Among unsaturated fatty acids linolenic acid also showed a lower coefficient of determination than oleic acid and linoleic acid (Choung et al., 2005; Kovalenkon et al., 2006; Sun et al., 2008).

Second, the threshold for coefficient of determination of 0.95 or more is important for assuring accurate NIR estimations for various traits. Calibration equations with less than 0.95 coefficient of determinations can be useful for rough screens for measuring soybean genotypes or soybean products for various components of interest. If soybean scientists use the NIR with calibration equations with relatively low coefficient of determinations to measure traits, it is necessary to use other tools such as wet chemistry, or molecular markers to confirm screening data. However, it is clear that even though calibration equations with low coefficient of determination, NIR is still useful to screen samples where a high level of accuracy is not required.

Third, NIR spectroscopy has been widely and accurately used for measuring protein and oil. There is evidence where NIR is useful for measuring levels of amino acids in protein, unsaturated fatty acid in oil and other useful traits. However, prediction equations for traits other than protein and oil need refining for broader use of NIR for measuring levels of other important seed characteristics. Therefore, calibration models should be developed and optimized from large datasets to develop highly accurate calibration equations. Protocols are needed to update calibration models with minimal effort, and collaborative efforts are needed to improve and share NIR spectroscopy technology leading to broader use of NIR technology to efficiently measure more traits.

6. References

Birt, D.F.; Hendrick, S. & Alekel, D.L. (2004) Soybean and the prevention of chronic human disease, In: *Soybeans: Improvement, Production, and Uses*, Boerma, H.R. & Specht, J.E. (Ed.), pp.1047-1117, ASA-CSSA-SSSA, ISBN 0-89118-154-7, Madison, WI, USA.

Choung, M.G.; Baek, I.Y.; Kang, S.T.; Han, W.Y.; Shin, D.C.; Kwack, Y.H. & Moon, H.P. (2004). Non-destructive analysis method of one seed grain by near infrared reflectance spectroscopy. Korean Intellectual Property Office Registration No.(Date) 10-0433263-000 (2004.05.18) (Korean with English abstract).

Choung, M.G. (2010). Determination of sucrose content in soybean using near-infrared reflectance spectroscopy. *J Korean Soc. Appl. Biol. Chem.*, Vol.(53), No.(4), 478-484, ISSN 1738-2203.

Choung, M.G.; Baek, I.Y.; Kang, S.T.; Han, W.Y.; Shin, D.Ch.; Moon, H.P. & Kang, K.H. (2001a). Non-destructive method for selection of soybean lines contained high protein and oil by near infrared reflectance spectroscopy. *Korean J Crop Sci.*, Vol.(46) , No. (5), 401-406, ISSN 0252-9777.

Choung, M.G.; Baek, I.Y.; Kang, S.T.; Han, W.Y.; Shin, D.Ch.; Moon, H.P. & Kang, K.H. (2001b). Determination of protein and oil contents in soybean seed by near infrared reflectance spectroscopy. *Korean J Crop Sci.*,Vol.46, No.(2), 106-111, ISSN 0252-9777.

Choung, M.G.; Hwang, Y.S. & Baek, I.Y. (2009a). Simple determination of seed weights in crops using near infrared reflectance spectroscopy. Korean Intellectual Property Office Registration No.(Date) 10-934410 (2009.02.18).

Choung, M.G.; Hwang, Y.S.; Kang, S.T.; Baek, I.Y. & Kim, H.T. (2009), Non-destructive discriminant analysis of herbicide tolerance soybean. Korean Intellectual Property Office Registration No.(Date) 10-0883664 (2009.02.18).

Choung, M.G.; Kang, S.T.; Han, W.Y.; Baek, I.Y.; Kim, H.K.; Kim, K.S. & Park, S.H. (2005). Determination of fatty acid composition in soybean seed using near infrared reflectance spectroscopy. *Korean J Breed.*, Vol.37, No.(4), 197-202, ISSN 0250-3360.

Davies, A.M.C. (2000). William Herschel and the discovery of near infrared. *Spectrosc. Eur.*,Vol.12, 10-16, ISSN 0966-0941.

Delwiche, S.R.; Pordesimo, L.O.; Panthee, D.R. & Pantalone, V.R. (2007). Assessing glycinin(11S) and β-conglycinin (7S) fractions of soybean storage protein by near-infrared spectroscopy. *J Am. Oil Chem. Soc.*, Vol.(84), 1107-1115, ISSN 0003-021X.

Delwiche, S.R.; Pordesimo, L.O.; Scaboo, A.M. & Pantalone, V.R. (2006). Measurement of inorganic phosphorus in soybean with near-infrared spectroscopy. *J Agric. Food Chem.*,Vol.54, No. (19), 6951-6956, ISSN 1520-5118.

Fontaine, J.; Horr, J.; & Schirmer, B. (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.*,Vol.49, No.(1), 57-66, ISSN 1520-5118.

Geater, C. & Fehr, W.R. (2000). Association of total sugar content with other seed traits of diverse soybean cultivars. *Crop Sci.*, Vol. (40), No. (6), 1552-1555, ISSN 1435-0653.

Gerde, J.A.; Hardy, C.L.; Hurburgh Jr, Ch.R. & White P.J. (2007). Rapid determination of degradation in frying oils with near-infrared spectroscopy *J Am. Oil Chem. Soc.*, Vol.(84), 519-522, ISSN 0003-021X.

Han, W.Y.; Park, K.Y.; Choung, M.G.; Kim, H.T.; Ko, J.M.; Baek, I.Y. & Lee, C.Y. (2008). Growth characteristics and qualities of Korean soybean landraces. *Korean J Crop Sci.*, Vol.(53), No.(:89-95, ISSN 0252-9777.

Helms, T.C. & Orf, J.H. (1998). Protein, oil an yield soybean lines selected for increased protein. *Crop Sci.*, Vol.(38), No.(3), 707–711, ISSN 1435-0653.

Hollung, K.; Overland, M.; Hrustic, M.; Sekulic, P.; Miladinovic, J.; Martens, H.; Narum, B.; Sahlstrom, S.; Sorensen, M.; Storebakken, T. & Skrede, A. (2005). Evaluation of nonstarch polysaccharides and oligosaccharide content of different soybean varieties (*Glycine max*) by near-infrared spectroscopy and proteomics. *J Agric. Food Chem.*, Vol.(53), No.(23), 9112-9121, ISSN 1520-5118.

Jinendra, B.; Tamaki, K.; Kuroki, S.; Vassileva, M.; Yoshida, S. & Tsenkova, R. (2010). Near infrared spectroscopy and aquaphotomics: Novel approach for rapid in vivo diagnosis of virus infected soybean. *Biochemical and Biophysical Research Communications*, Vol.(397), No.(4), 685-690, ISSN 0006-291X.

Kanamaru, K.; Wang, S.; Abe, J.; Yamada, T. &Kitamura, K. (2006). Identification and characterization of wild soybean (*Glycine soja* Sieb. et Zecc.) strains with high lutein content. *Breed. Sci.*, Vol.(56), No.(3), 231-234, ISSN 1347-3735.

Kim, K.S.; Park, S.H. & Choung, M.G. (2007). Nondestructive determination of oil content and fatty acid composition in perllia seed by near-infrared spectroscopy. *J Agric. Food Chem.*, Vol.(55), No.(5), 1679-1685, ISSN 1520-5118.

Kim, Y.H.; Ahn, H.K.; Lee, E.S. & Kim, H.D. (2008). Development of prediction model by NIRS for anthocyanin contents in black colored soybean. *Korean J Crop Sci.*, Vol.(53), No.(1), 15-20, ISSN 0252-9777.

Kovalenko, I.V.; Rippke, G.R. & Hurburg, C.R. (2006). Measurement of soybean fatty acid by near-infrared spectroscopy: Liner an nonlinear calibration methods. *J Am. Oil Chem. Soc.*, Vol.(83), 421-427, ISSN 0003-021X.

Lee, J.D.; Bilyeu, K.D. & Shannon, J.G. (2007). Genetics and breeding for modified fatty acid profile in soybean seed oil. *J Crop Sci. Biotech.*, Vol.10, No.(4), 201-210, ISSN 1975-9479.

Lee, J.D.; Shannon, J.G. & Choung, M.G. (2010). Selection for protein content in soybean from single F_2 seed by near infrared reflectance spectroscopy. *Euphytica*, Vol.(172), No.(1), 117-123, ISSN 0014-2336.

Li, H.; Goulden, M.; Cocciardi, R. & Hughes, G. (2009). Fourier transform near infrared spectroscopy as a quality control tool for the analysis of lecithin and by-products during soybean oil processing. *J Am. Oil Chem. Soc.*, Vol.(86), 835-841, ISSN 0003-021X.

Li, H.; Lv, X.; Wang, J.; Li, J.; Yang, H. & Qin Y. (2007). Quantitative determination of soybean meal content in compound feeds: comparison of near-infrared spectroscopy and real-time PCR. *Anial. Bioanal. Chem.*, Vol.(389), No.(7-8), 2313-2322, ISSN 1618-2642.

Li, H.; Voort, F.R.; Sedman, J. & Ismail, A.A. (1999). Rapid Determination of *cis* and *trans* content,Iodine value, and saponification number of edible oils by fourier transform near-infrared spectroscopy. *J Am. Oil Chem. Soc.*, Vol.(76), No.(4), 491-497, ISSN 0003-021X.

Nicolai, B.M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K.I. & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, Vol.(46), No.(2), 99-118. ISSN 0925-5214.

Palomeque, L.; Liu, L.J.; Li, W.; Hedges, B.R.; Cober, E.R.; Smid, M.P.; Lukens, L. & Rajcan, I. (2010). Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. *Theor. Appl. Genet.*, Vol.(120), No.(5), 997-1003, ISSN 0040-5752.

Pezdernik, D.L.; Killam, A.S. & Orf, J.H. (1997). Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. *Agron. J*, Vol.(89), No.(4), 679-685, ISSN 1435-0645.

Sato, T.; Eguchi, K.; Hatano, T. & Nishiba, Y. (2008). Use of near-infrared reflectance spectroscopy for the estimation of the isoflavone contents of soybean seeds. *Plant Prod. Sci.*, Vol.11, No.(4), 481-486, ISSN 1349-1008.

Sato, T.; Takahashi, M. & Matsunaga, R. (2002). Use of NIR spectroscopy for estimation of FA composition of soy flour. J. *Am. Oil Chem. Soc.*, Vol.(79), No.(6), 535-537, ISSN 0003-021X.

Sebolt, A.M.; Shoemaker, R.C. & Diers, B.W. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.*, Vol. (40), No.(5), 1438–1444, ISSN 1435-0653.

Sirisomboon, S.; Hashimoto, Y. & Tanaka, M. (2009). Study on non-destructive evaluation methods for defect pods for green soybean processing by near-infrared spectroscopy. *Journal of Food Engineering*, Vol.(93), No.(4), 502-512, ISSN 0260-8774.

Sun, J.M.; Han, F.X.; Yan, S.R.; Yang, H. & Tetsuo, S. (2008). Rapid determination of fatty acids in soybeans (*Glycine max* (L.) Merr.) by FT-near-infrared reflectance spectroscopy. *Spectroscopy and Spectral Analysis*, Vol.(28), No.(6), 1290-1295, ISSN 1000-0593.

Tajuddin, T.; Watanabe, S.; Masuda, R.; Harada, K. & Kawano, S. (2002). Application of near infrared transmittance spectroscopy to the estimation of protein and lipid contents in single seeds of soybean recombinant inbred for quantitative trait loci analysis. *J Near Infrared. Spectrosc.*, Vol.(10), No.(4), 315-325, ISSN 0967-0335.

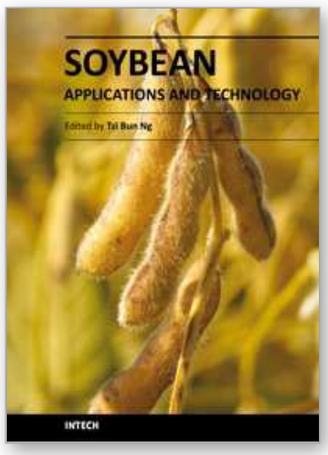
Tajuddin, T.; Watanabe, S.; Yamanaka, N.; Harada, K. (2003) Analysis of quantitative trait loci for protein and lipid contents in soybean seed using recombinant inbred lines. *Breed. Sci.*, Vol.(53), No.(2), 133-140, ISSN 1347-3735.

Wang, D.; Dowell, F.E.; Ram, M.S.; & Schapaugh, W.T.(2004). Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. *International journal of food properties*, Vol.(7), No.(1), 75-82, ISSN 1094-2912.

Wilson, R.F. (2004). Seed compostion, In: *Soybeans: Improvement, Production, and Uses*, Boerma, H.R. & Specht, J.E. (Ed.), pp. 624-677, ASA-CSSA-SSSA, ISBN 0-89118-154-7, Madison, WI, USA

Yildiz, G.; Wehling, R.L. & Cuppett, S.L. (2001). Method for determination oxidation of vegetable oils by near-infrared spectroscopy. *J Am. Oil Chem. Soc.*, Vol.(78), No.(5), 495-502, ISSN 0003-021X.

Yildiz, G.; Wehling, R.L.; & Cuppett, S.L. (2003). Comparison of four analytical methods for the determination of peroxide value in oxidized soybean oils. *J Am. Oil Chem. Soc.*, Vol.(80), No.(2), 103-107, ISSN 0003-021X.



Soybean - Applications and Technology
Edited by Prof. Tzi-Bun Ng

ISBN 978-953-307-207-4
Hard cover, 402 pages
Publisher InTech
Published online 26, April, 2011
Published in print edition April, 2011

Soybean is an agricultural crop of tremendous economic importance. Soybean and food items derived from it form dietary components of numerous people, especially those living in the Orient. The health benefits of soybean have attracted the attention of nutritionists as well as common people.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jeong-Dong Lee, J. Grover Shannon and Myoung-Gun Choung (2011). Application of Nondestructive Measurement to Improve Soybean Quality by Near Infrared Reflectance Spectroscopy, Soybean - Applications and Technology, Prof. Tzi-Bun Ng (Ed.), ISBN: 978-953-307-207-4, InTech, Available from: http://www.intechopen.com/books/soybean-applications-and-technology/application-of-nondestructive-measurement-to-improve-soybean-quality-by-near-infrared-reflectance-sp

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen