NIR Calibrations for Soybean Seeds and Soy Food Composition Analysis: Total Carbohydrates, Oil, Proteins and Water Contents

Jun Guo* and I.C. Baianu*, **

**AFC- NMR and NIR Microspectroscopy Facility, FSHN Dept., College of ACES, University of Illinois at Urbana, 305 Burnsides Research Laboratory, 1208 W. Pennsylvania, Ave., Urbana, IL. 61801*

*** NPRE Department, University of Illinois at Urbana, Urbana, IL. 61801*

1. Introduction

1.1. A Review of NIR Literature

Conventional chemical analysis techniques are expensive, time-consuming, and often destructive. The non-invasive Near Infrared (NIR) technology was introduced over the last decades for wide-scale, inexpensive chemical analysis of food and crop seed composition (see Williams and Norris, 1987; Wilcox and Cavins, 1995; Buning and Diller, 2000 for reviews of the NIR technique development stage prior to 1998, when Diode Arrays were introduced to NIR). NIR spectroscopic measurements obey Lambert and Beer's law, and quantitative measurements can be successfully made with high speed and ease of operation. NIR has been used in a great variety of food applications. General applications of products analyzed come from all sectors of the food industry including meats, grains, and dairy products (Shadow, 1998).

The absorption bands observed in the NIR region are primarily overtones. Therefore, the absorptions tend to be weak in intensity. However, this is actually an advantage, since absorption bands that have sufficient intensity to be observed in the NIR region arise primarily from functional groups that have a hydrogen atom attached to a carbon, nitrogen, or oxygen, which are common groups in the major constituents of food such as proteins, lipids, water and carbohydrates. Either reflectance or transmittance measurements may be made in NIR spectroscopy, depending on the type of sample.

NIR instruments can be calibrated to measure various constituents in food and agricultural commodities. The technique has found its widest use in the grain, cereal products, and oilseed processing industries. NIR techniques using reflectance measurements from ground or powdered samples have been adopted as approved methods of analysis by the American Association of Cereal Chemists (AACC 1995) for measuring protein in barely, oats, rye, triticale, and wheat (Method 39-10), protein in wheat flour (Method 39-11), and protein and oil in soybeans (Method 39-20). Techniques using measurements from whole kernel grains have also been approved for protein, oil and moisture in soybeans (Method 39-21), and protein in wheat (Method 39-25). NIR reflectance measurements also have been approved for estimating wheat hardness (Method 39-70A). These approved methods describe the instruments available for use in making these measurements, and the proper techniques for preparing samples and calibrating instruments.

NIR spectroscopy also can be used for numerous other commodities and food products. The technique has been used successfully to measure moisture, protein and fat in red meats and processed meat products (Bartholomew and Osuala 1988; Isaksson et al. 1992; Oh and

Grossklaus 1995), poultry and fish (Rasco et al. 1991). NIR spectroscopy is useful also for analyzing a number of dairy products, including measuring moisture and fat in butter; moisture, fat and protein in cheese (Pierce and Wehling 1994; Rodriquez-Otero et al. 1995); and lactose, protein and moisture in milk and whey powders (Baer et al. 1983). In addition, moisture, fat and protein have been determined in dehydrated eggs using NIR reflectance measurements (Wehling et al. 1988). NIR techniques also have shown promise for measuring total sugars and soluble solids in fruits and vegetables (Birth et al. 1985; Dull et al. 1989), and are being used commercially for monitoring the sugar content in corn sweeteners (Psotka and Shadow 1994).

However, at present, there is no established secondary method for soybean composition analysis (see, for example, **http://www.farmresearch.com/ispob**), because the older NIR technology and calibration methodology were not sufficiently reliable, by comparison with established primary methods. All secondary methods require extensive and careful calibration development, as well as optimization of analytical techniques prior to practical applications. So far, the NIR method has not been adopted as either an AOCS (American Oil Chemists' Society) Official Method, or an AOCS Recommended Practice. Rather, it is an AOCS Procedure intended to provide general guidelines for a routine instrumental method that is dependent on the instrument being used (AOCS Procedure Am 1-92). Developing NIR calibrations that are both robust and precise is, therefore, very important for a wide range of practical applications. In 1998, the optical filter or monochromator based dispersive NIR instruments equipped with Diode Array detectors became commercially available in the US. Several Diode Array dispersive NIR instruments, in conjunction with our novel calibrations were mainly employed, for bulk soybean sample analysis, in terms of total contents of protein, oil and carbohydrates.

Soybean meal is the major source of vegetable protein in the U.S. and the world. Historically, soybean meal accounts for 60-70% of the value of the soybean (Pfeffer and Gerasimowicz, 1987). It has been recently estimated that a 1% increase in protein would increase the value of the Illinois soybean crop by $56 million/yr (Iowa Agriculture and Home Economics Experiment Station. 1990).

Developing high-yielding varieties with high protein or high oil percentage has been an elusive goal for soybean breeders. Current breeding research needs include developing a better understanding of the plant genetics involved in soybean seed development; understanding the inverse relationship between protein and yield; developing new, rapid methods for measuring protein and oil concentrations in the soybeans; and expanding research on the genetic control of soybean composition (Wilson, 1991). It is difficult to increase seed protein concentration in soybean through plant breeding without adversely affecting yielding ability and oil concentration. Increasing percent protein often results in decreased seed yield, and almost always results in decreased percentage of oil (Brim and Burton, 1979; Simpson and Wilcox, 1983; Wilcox 1998; Geater et al. 2000).

Many studies have discussed the inverse relationship between protein and oil (Brim and Burton, 1979), but little is known about the relationships between protein, oil, and sugars even though sugars can contribute more than 10% of the total dry weight of the seed. Several studies indicate that total sugar concentration and oil concentration in soybean seeds are positively associated, and each is negatively correlated with protein concentration (Openshaw and Hadley, 1978). Breeding efforts to modify seed composition have concentrated on changing only oil and protein percentages while ignoring sugar concentration. Limited surveys of germplasm indicate that sugar concentration of soybean ranges from 6 to 17% and wild soybean ranges from 4 to 8%

(Kilen and Kou, 1994). In the past, screening for sugar levels in seed has been accomplished by HPLC, a time consuming and expensive procedure that strictly limits the number of lines that can be tested. The development of a rapid, low-cost screening technique would allow for a thorough survey of the genetic diversity available and for the selection of low sugar concentration soybeans in a breeding program. Reducing sugar concentration might help, in principle, to increase the total amount of either oil or protein in the seed.

Previous NIR calibrations for soybean measurements were mainly developed with soybean powder samples (Williams and Sobering 1993; Sato et al. 1994; Myoung et al., 2001A) for seed composition analysis. Individual soybean component protein or oil, and subunits of soybean protein and oil such as amino acids and fatty acids were also reported as objectives of NIR calibrations. Crude protein and protein fractions (7S, 11S) of soybean were analyzed by NIR Reflectance Spectroscopy (Cho et al., 1987; Pazdernik et al., 1996). Quantitative determination of epoxidized soybean oil was performed with NIR calibration by using the partial least squares (PLS) method (Parreira et al., 2002). The concentrations of mannosyl erythritol lipid and soybean oil in the glycolipid fermentation process were measured by using NIR spectroscopy (Nakamichi et al., 2002). A method was developed to determine the total phospholipid content in soybean oil by Fourier transform IR spectroscopy (Nzai et al., 1998). Analysis of amino and fatty acid composition in soybean seed, using NIR reflectance spectroscopy were also reported with moderately high correlation coefficients between NIR values and reference values (Pazdernik et al., 1997), and lower than those reported by Baianu et al. (2003). In addition, applications of near infrared transmittance and reflectance spectroscopy to the estimation of protein and oil contents in soybean single seeds have recently caused much research interest (Tajuddin et al., 2002; Baianu et al., 2002).

Recently, Myoung et al. developed calibrations for both whole and ground soybeans by using MPLS (Modified Partial Least Squares) algorithm with reflectance NIR spectra obtained within the range from 1100 to 2500 nm (Myoung et al. 2001A; Myoung et al. 2001B). Their calibrations were developed with protein and oil percentages on a dry matter basis, and of course, without developing moisture calibration for soybeans.

In addition to the dispersive NIR instruments we employed, a Fourier Transformed NIR instrument (FT-NIR) was also employed to provide food and soybean calibrations with better resolution, faster spectral acquisition time (compared with dispersive instruments) and wider wavelength ranges (Diem 1993; Rodriguez-Saona et al., 2000). Several FT-NIR calibrations were reported recently for the determination of major components of edible seeds and foods. A calibration method was developed to determine the total phospholipid content in soybean oil with FT-IR spectroscopy (Nzai et al., 1998). A calibration method for protein and apparent amylase contents of milled rice was developed by using NIR-FT-Raman spectroscopy (Himmelsbach et al., 2001). Determination of iodine value was performed with FT-NIR based global calibration (Cox et al., 2000). A FT-NIR reflectance spectroscopy and PLS based method was employed for rapid quantification of castor bean meal (Rodriguez-Saona et al., 2000). However, no FT-NIR calibrations for soybean seeds, soy and other health foods have been published yet. The closest work was the last report about FT-NIR reflectance spectroscopy for rapid quantification of castor bean meal components Their measurements were made on an FT-NIR system using a diffuse reflection-integrating sphere, based on multiplicative scatter correction transformed partial least-squares regression models. The calibration has a standard error of cross-validation of <0.6% and a coefficient of determination R^2 of >94%. The PerkinElmer's Spectrum One NTS FT-NIR system was employed for the research on soy and

other health foods (Guo et al., 2002), analysis of soybeans with black or brown seed coats, as well as accurate determination of soybean isoflavone concentration (You et al., 2002).

NIR analysis of soybean seeds is mainly focused on protein and oil percentages. In detail, soybean protein components are mainly the globulins, which include 2S(15%), 7S(34%), 11S(42%) and 15S(4%), on the basis of their size via ultracentrifugation. The major soy proteins are principally referred to as 7S (conglycinin) and 11S (glycinin). Both the 7S and 11S globulins are composed of multiple subunits that are assembled into distinct molecular structures that represent a very complex quaternary structure (Morr, 1988). Soybean oil, as in other oilseed crops, is primarily composed of triacylglycerols (TG), which are sequestered within seeds in discrete organelles (Wilson, 1991). In this thesis, the NIR instruments were employed to measure total protein and oil percentages of soybeans without giving their sub-component values.

Mature soybeans contain trace amounts of monosaccharides such as glucose and arabinose, and measurable amounts of di and oligosaccharides, with sucrose in the range of 2.5-8.2%; raffinose, 0.1-0.9%; and stachyose, 1.4-4.1% (Liu, 1997). The latter two are galactosylsucrose oligosaccharides and also known as alfa-galactosides of sucrose (Bach and Li, 1991). Among the soluble carbohydrates, raffinose and stachyose receive more attention, mainly because their presence has been linked to flatulence and abdominal discomfort associated with human consumption of soybeans and soy products (Liu, 1997). An ultimate solution to the flatus problem would be the genetic removal of oligosaccharides by plant breeding, since there is considerable variation in the raffinose and stachyose concentration among varieties of soybeans. Basically, the oligosaccharides in soybeans are nonreducing sugars, containing fructose, glucose, and galactose as two or more units, linked by β-fructosidic and α-galactosidic linkages. Sucrose and sucrose-related oligosaccharides are very readily hydrolyzed, so that fructose, glucose and its derivatives, are easily formed inadvertently. For NIR analysis, the small sugars including sucrose, raffinose and stachyose in the soybean samples were mainly measured.

Soybean composition analysis has been the subject of both conventional chemical analysis and instrumental analysis for several decades. Except the NIR instruments, other regularly employed instruments for such a purpose include High Performance Liquid Chromatography (HPLC), Supercritical Fluid Chromatography (SFC), Infrared Attenuated Total Reflectance (ATR), and Nuclear Magnetic Resonance (NMR), etc. (Espeja et al., 2001; Wang et al., 2001; Wilson et al., 1973; Baianu et al., 2004). Among those techniques, High Resolution Nuclear Magnetic Resonance (HR-NMR) was employed for accurate determinations of soybean protein, oil and amino acids contents (Baianu et al., 2004). The NMR composition data were compared to the corresponding NIR data, with very high correlation up to 0.99. Both high-resolution NIR and NMR calibrations and methodologies were employed in our group -- with HR-NMR employed to calibrate the NIR -- and, respectively, carried out a large number of protein and oil composition analyses of soybean seeds for breeding and selection purposes over a period of three years.

1.2. Review of NIR theory and instrumentation

As the name indicates, the term NIR refers to a portion of the electromagnetic (EM) spectrum between the visible and the infrared portion of the EM spectrum - it actually lies in between these two regions (Shadow, 1998):

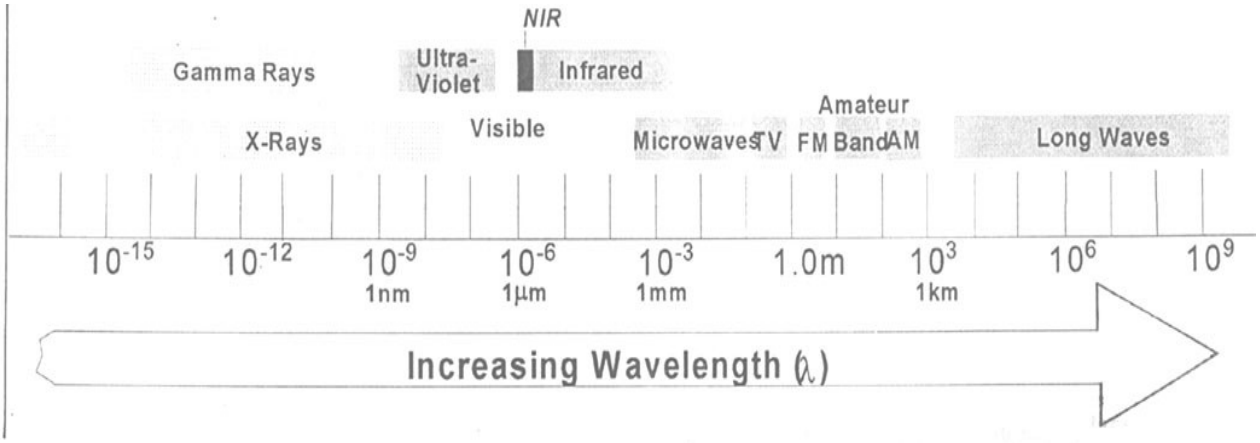

Figure 1.1. ***NIR portion in the electromagnetic spectrum (source: Shadow, 1998).***

NIR is very similar in form to visible light, and its range extends from about 780 nm to 2600 nm.

1.2.1. Basic Theory of IR and NIR Spectroscopy

The absorptions occurring in the NIR region are considered to be rotation-vibration combinations. In general, the overall NIR energy levels can be expressed as :

$$E_{NIR} = E_{rot} + E_{vib} + \text{anharmonic term.} \qquad \text{Eq. (1.1)}$$

The possible absorptions are quantum mechanical in nature: only discrete energy amounts or quanta can be absorbed. The vibration energy levels can be approximately calculated using this expression (Osborne et al., 1993):

$$E_n = (n+1/2)\, h\nu = (n+1/2)h(1/2\pi\sqrt{k/\mu}) \qquad \text{Eq. (1.2)}$$

Where E_n are the molecular vibration energy levels, $n=0,1,2.3,...$; h is Plank's constant, k is the elastic constant and μ is the reduced mass of the molecule, and the NIR frequencies are to a first order approximation:

$$\nu = 1/2\pi\sqrt{k/\mu}. \qquad \text{Eq. (1.3)}$$

According to the quantum selection rule for fundamental modes, the only allowed vibrational transitions are those in which $\Delta n = \pm 1$. The corresponding absorption peaks are situated in the middle infrared region, usually referred as the IR absorption bands which are linear and obtained with low power, due to fundamental modes of molecular vibration. However, real molecules do not obey exactly the laws of simple harmonic motion, and real bands, though elastic, do not obey Hooke's Law exactly. The anharmonicity weakens the selection rule, therefore, overtone bands which are non-linear and obtained with high power, given by transitions where $\Delta n = \pm 2, \pm 3$, etc, can also occur, Using Morse's empirical function, the following formula is offered by solving the wave mechanical (Schrodinger) equations for the energy levels of an anharmonic oscillator:

$$E_n = [1-x(n+1/2)]h\nu \ , \qquad \text{Eq. (1.4)}$$

where the anharmonic constant is $x \cong 0.01$

The energy associated with a transition from n to $n+\Delta n$ is $\Delta E_n = [1-(2n+\Delta n+1)x]h\nu$, and the quantum selection rules are: $\Delta n = \pm 1, \pm 2, \pm 3,...$

The first three series of quantum transitions have the highest observable intensities and the IR/NIR bands lie very close to $h\nu$, $2h\nu$, $3h\nu$, to a first approximation:

(i) $n=0$ to $n=1$, $\Delta n=+1$, $\Delta E=(1-2x)h\nu$ Eq. (1.5)

(ii) $n=0$ to $n=2$, $\Delta n=+2$, $\Delta E=(1-3x)2h\nu$ Eq. (1.6)

(iii) $n=0$ to $n=3$, $\Delta n=+3$, $\Delta E=(1-4x)3h\nu$ Eq. (1.7)

The linenear term $h\nu$ is called *the fundamental absorption*, while those near $2h\nu$ and $3h\nu$ are called respectively, *the first and second overtones*; these are the major contributions to the NIR absorption bands. However, the combinations of anharmonic vibration and rotation motions give rise to the overall NIR absorption bands. The 'pure' rotation, quantized energy levels can be expressed as:

$$E_{rot} = \mathbf{J(J+1)\,B}\,hc\,, \qquad \text{Eq. (1.8)}$$

where the *rotational quantum number* **J** takes o*nly* integer values, *B* is the rotation constant, h is Plank's constant, and c is the speed of light in vacuum. The quantum selection rules for transitions between such rotational levels are: $\Delta J=\pm 1$ for IR transitions, and $\Delta J=\pm 2, \pm 3,...,$ for NIR absorptions. In detail, the overall NIR energy levels can be then expressed as:

$$\mathbf{E_{NIR} = E_{rot} + E_{vib} + anharmonic\ term = J(J+1)Bhc + [1-x(n+1/2)]h\nu} \qquad \text{Eq. (1.9)}$$

Conventional spectroscopy is defined as a *frequency domain spectroscopy* in that the radiant power data is recorded as a function of frequency, or the inversely related wavelength. In contrast, *time domain spectroscopy*, that is observed by taking the Fourier Transform (FT) of a time domain signal, monitors the changes in radiant power ocuring over time (Skoog et al., 1998). It is important to realize that a time domain signal contains in principle the same information as the corresponding spectrum in the frequency domain; in fact, each can be converted into the other by digital computations, preferrably through a *Fast* Fourier Transform (*FFT*) algorithm. Time domain spectra can be obtained, for example, with a classical Michelson interferometer. The Michelson interferometer is used extensively to mudulate radiation in the optical region. A plot of the output power from the detector versus the path length difference between the interferng light beams in the interferometer is called *an interferogram*. By doing a FFT, the interferogram can be rapidly converted into regular spectra (in the frequency domain) that are both convenient and necessary for interpreting the data. For a continuum source, the interferogram can be represented as a sum of an infinite number of cosine terms, that is

$$\mathbf{P(\delta) = \int B(\nu)\cos(2\pi\nu\delta)\,d\nu} \qquad \text{(an integral from } -\infty \text{ to } +\infty) \qquad \text{Eq. (1.10)}$$

The Fourier transform of this integral is:

$$\mathbf{B(\nu) = \int P(\delta)\cos(2\pi\nu\delta)\,d\delta} \qquad \text{(an integral from } -\infty \text{ to } +\infty), \qquad \text{Eq. (1.11)}$$

that gives $\mathbf{B(\nu)}$ as a function of ν (in the frequency spectrum).

There are several major advantages to the use of Fourier transform instruments. The first is the throughput, or *Jaquinot advantage*, which is realized because FT instruments have few optical elements and no slits to attenuate radiation. The second advantage of FT instruments is their extremely high resolving power and wavelength reproducibility throughout the entire spectrum. These features make the analysis of complex spectra possible, in which the sheer number of lines and spectral overlap make the determination of individual spectral features difficult. The third advantage arises because all elements of the source reach the detector simultaneously. This characteristic makes it possible to obtain data for an entire spectrum in one second or less.

1.2.2. Selected examples of NIR spectra for small molecules: water, sucrose and oils

Organic molecules and water are in energy states that can absorb at NIR wavelengths (Shadow, 1998). Most metals, on the other hand, such as silver or lead in the solid state cannot absorb in the NIR region. All foods have major NIR bands since they contain organic molecules, with -C-H, -C-O-H, -C-N-H bonds, etc. Their NIR absorption obey Beer's law which states that, a component's (or *analyte*) absorption is proportional to its concentration in the sample, provided that suitable scattering conditions have already been carried out; Beer's law is the basis for both visible and NIR/IR composition analysis. Most organic molecules usually have absorption bands from 893 nm to 1045 nm, and NIR therefore can be one of the most useful spectroscopic techniques for food analysis. **Figure 1.2** shows examples of FT-NIR spectra of small molecules such as: water, sucrose, oil, for the entire NIR region; the FT-NIR spectrum of large molecules, such as proteins is shown for comparison in **Figure 1.3**.

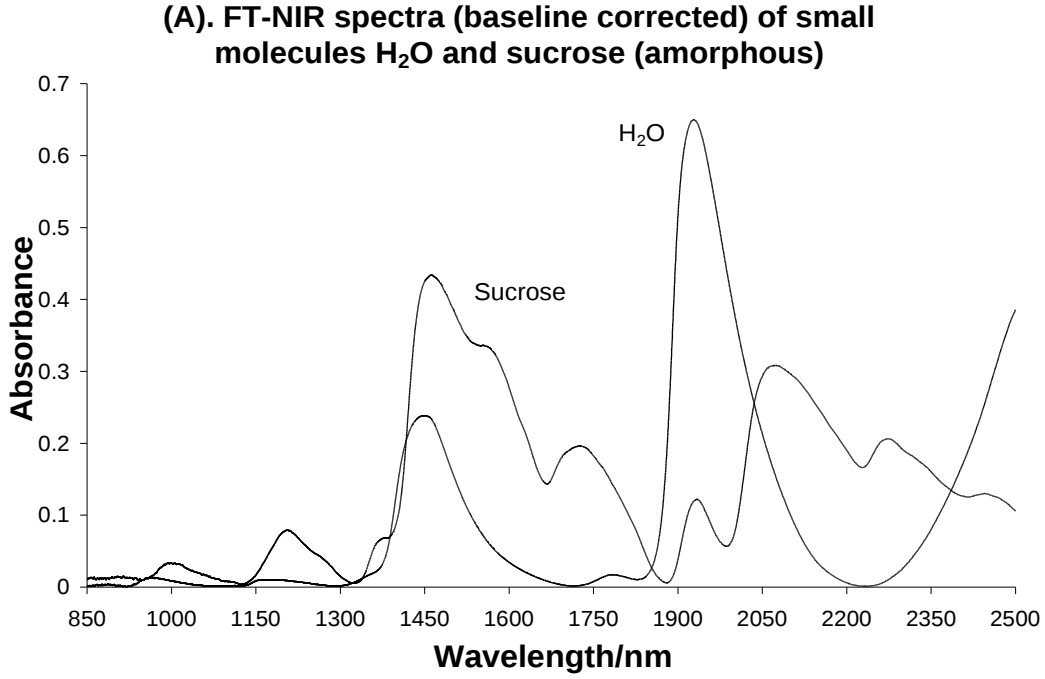

Figure 1.2. ***FT-NIR spectra (obtained with PerkinElmer's Spectrum One NTS) of (A) small molecules water and sucrose.***

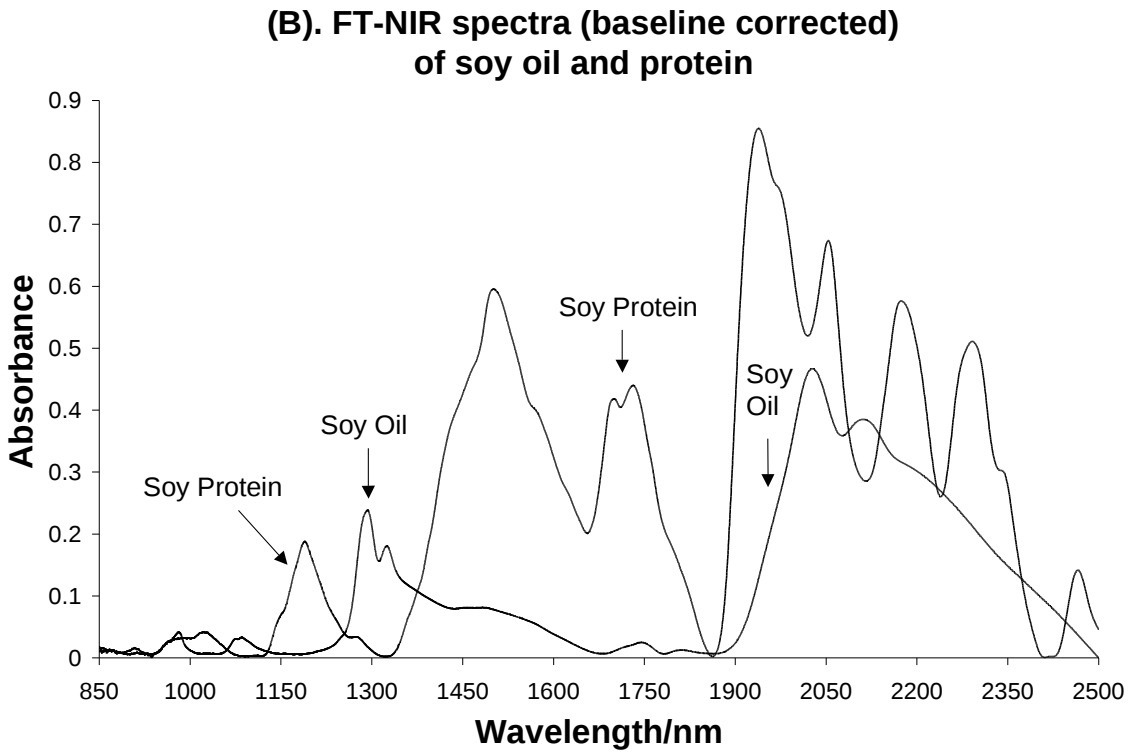

Figure 1.3. ***FT-NIR spectra (obtained with PerkinElmer's Spectrum One NTS) of oils and proteins.***

1.2.3. NIR instrumentation

The major components of NIR instruments usually contain a light source (visible and infrared), a sample compartment, an instrument cell, a signal detector and a computer control system. The NIR working modes are usually the transmittance or reflectance modes.

1.2.4. Major advantages and limitations of NIR techniques

The major advantages of NIR composition analysis are (according to Shadow,1998) as follows:

(1) Ease of operation and rapid measurement

(2) High accuracy and reproducibility

(3) No sample preparation is needed in most cases and chemicals are not used in analysis

(4) Low cost since instruments are relatively inexpensive

(5) Calibrations may be shared and compared among similar instruments.

The limitations are:

(1) NIR is a secondary technique, and therefore, NIR instruments must be calibrated to a reliable reference method.

(2) The calibrations must be monitored and maintained.

(3) Samples to be analyzed must be similar in composition as the calibration samples in the calibration range.

(4) Very few commercial instruments are available or designed to take into account environmental control factors such as temperature and humidity.

1.3. Basic questions for soybean seed and food analysis by NIR

Accurate, reliable and economic composition analysis of crop seeds and foods is important for food industry and agriculture. One such technique that has been requested by industry and crop breeders is the NIR method. Recent new motivation to develop the NIR methodology also comes from post-harvest food protection and defense, biomedical and biotechnology applications. In the following part, I will address several key questions faced by any extensive NIR analysis for soybean seed and food analysis that are feasible to be completed in a reasonable amount of time at moderate cost.

Although NIR technology has been widely employed for chemical analysis of seed and food composition for over 30 years, there are still basic questions related to NIR analysis of foods and crop seeds. A basic question that was asked more than 30 years ago is:

Can NIR be employed reliably and accurately for crop seed and food composition analysis?

This question has not been completely and satisfactorily answered but for a few seeds such as wheat. Although there have been preliminary reports of bulk soybean seeds, such results are not yet widely accepted. Furthermore, the older instruments available commercially prior to 1999 had low accuracy, reliability and reproducibility. There are recent claims from several NIR manufacturers that the more recent NIR instruments are capable of both high accuracy and reliability for crop seed measurements and food composition analysis. However, such claims are unsupported yet, either by independent instrument evaluation or through seed calibrations from

those NIR instrument manufacturers. None of the NIR instrument manufacturers provide high quality calibrations for composition analysis of bulk whole soybeans and corn. Furthermore, there are no commercial NIR calibrations available at present for reliable analysis of food composition. Therefore, the very important question of NIR utilization for reliable composition analysis of crop seeds and foods remains still open. A related question to the utilization of NIR for seed and food composition analysis is:

How can the NIR data pre-processing be optimized for eliminating variability in the raw NIR data which is unrelated to composition?

The first step in this direction would involve the identification of factors responsible for such variability. After NIR data pre-processing, the next question is:

What are the best NIR calibration models and procedures that would yield accurate and reliable NIR predictions of composition for bulk whole soybeans and foods?

Previously, soybean breeders employed either wet chemical analysis (Liu 1997) or generated NIR calibrations with only a small number of powder samples (Williams and Sobering 1993; Sato et al, 1994) for seed composition analysis. In several recent reports, protein and oil correlation analyses were also performed by using analytical data obtained from NIR calibrations validated with only an insufficient, small number of powder samples, within narrow composition ranges, such as a protein range from 41% to 48% (Wilcox and Cavins, 1995), or from 40% to 44% protein for commercial calibrations (Zeltex Inc., 1998). On the other hand, whole soybean analysis by NIR on a very large scale, and also for a wide range of composition, has not yet been reported. Moreover, there have been no detailed reports of comparisons between whole and ground soybean calibrations for reflectance NIR instruments.

Accelerated genetic selection and breeding programs for soybeans have at least the following stringent requirements: the rapid and accurate composition determinations for very large soybean populations, and availability of robust calibrations for many different soybean cultivars over a very wide composition ranges. NIR spectroscopy does have the potential to fulfill these stringent requirements simultaneously. The wide range composition requirement is critical because most NIR manufacturers- unfortunately for the users- do not provide any calibrations for their NIR instruments; a few manufacturers provide only limited range NIR calibrations that are totally insufficient for genetic selection and breeding research. Therefore, another important basic question is:

What are the composition ranges of the main soybean and food components that are needed for NIR calibrations to be used in genetic selection and breeding research?

Some commercial NIR instruments that come with calibrations have several limitations, such as: narrow analytical ranges, inability to measure different types of soybeans (with black, brown, or green coats), requirements for large amounts of sample for the measurement, and so on. Therefore, optimal NIR calibrations that have wide ranges of composition, high accuracy and robustness need to be developed for both the reflectance and transmittance NIR modes for soybean composition analysis in genetic selection and breeding experiments. The robustness of a NIR calibration can only be ensured through validation by testing the calibration with a very large number of soybean lines (e.g. more than 1, 000), and for a very wide range of soybean composition. No NIR calibrations have yet been tested either with a very large number of soybean samples or for a wide range of composition. This is the first attempt to develop NIR methodology with both adequate hardware and software in order to provide new and improved NIR calibrations with a wide range of soybean composition and very large number of soybean lines/accessions for genetic selection and breeding. Such development requires several NIR reflectance and transmittance instruments in order to select the few NIR instruments that satisfy the four stringent requirements of composition analysis by NIR for genetic selection and breeding programs. Such efforts need to be then followed by extensive validation as specified above.

Although NIR has been used in a great variety of food applications, FT-NIR calibrations and measurements for soy and other health foods composition have not yet been reported. As far as the composition range of soy and other health foods is concerned, generally the wider the composition range is, the more accurate calibrations will be developed. Because there is no such inverse protein-oil correlation existing in man-made foods, very wide composition ranges of protein, oil, moisture and carbohydrates are possible, for example, from 0 up to 95%.

Since 1998, various new NIR reflectance and transmittance instruments have become commercially available, such as, *dispersive* NIR instruments- DA7000 (Reflectance), IM9100 (transmittance), ZX-800/ZX-880/ZX-50 (transmittance). or *FT-NIR* instruments: Bruker's Vector 22/23, Thermal Nicolet's Antaris model, Nicolet's Nexus 670 model, Bio Rad's FTS 5000/4000, and PerkinElmer's Spectrum OneNTS model. Facing such a wide choice of instruments of different designs, the analytical chemist is forced to answer the following. important question:

Which of the new NIR instruments available commercially are suitable, or best suited, for composition analysis of crop seeds and foods?

A thorough and careful evaluation should be carried out in order to select the best tools that are available for soybean seed and food analysis.

In addition to the previously mentioned requirements for soybean composition analysis, the genetic selection and breeding programs for soybeans are likely to include different types of seeds with different seed coat colors such as yellow, green, black and brown. That creates significant problems with NIR measurements of whole soybeans. Therefore, one needs to address the question:

How can one resolve the NIR measurement problems for whole soybeans with different color coats?

For the transmittance mode NIR instruments, this coat color problem is unsolvable for whole soybeans. No reflectance NIR calibrations have been reported so far for green, black or brown soybeans. Measuring ground seeds may largely reduce the coat color effects, based on accurate calibrations for ground seed samples. However, the ground sample preparation is destructive, and powder particles increase the light scattering problems that can reduce measurement accuracy. A completely new experimental approach by dehulling soybean seeds for NIR calibrations and measurements may enable one to solve this problem completely. Thus, NIR calibrations could be developed for soybeans both with and without coats. A high quality calibration for measuring green, black or brown soybeans could be selected by comparing the accuracy and robustness between them.

1.4. Research objectives

The long-term goals and supporting objectives of this collaborative project are addressed briefly in this part, in order to resolve the important basic questions remaining in NIR analysis of soybean seeds and foods.

Research Objectives:

1. Development of novel FT-NIR calibrations for soy and other health foods, in terms of major food components protein, fat, moisture, total carbohydrates and fiber, as well as determination of soybean isoflavone content.
2. Developments and comparisons of both spectra and calibrations for intact soybean seeds and ground soybean seeds will be carried out. NIR spectral deconvolutions of ground and whole soybeans will be conducted, in order to understand the relationship between soybean individual components and entire soybean spectra.

3. Statistical analysis of NIR data and validation of novel NIR calibrations developed for soybean composition analysis on a very large, representative set of soybean accessions and developmental soybean lines for genetic breeding and selection experiments that were carried out in the Crop Science Department at UIUC.
4. A thorough evaluation of various recently designed NIR reflectance and transmittance instruments will be carried out in order to select the best tools that are available for soybean seed and food analysis: Dispersive NIR instruments include DA7000 (Reflectance), IM9100 (transmittance), ZX-800/ZX-880/ZX-50 (transmittance). FT-NIR instruments will include: Bruker's Vector 22/23, Thermal Nicolet's Antaris, Nicolet's Nexus 670, Bio Rad's FTS 5000/4000, and PerkinElmer's Spectrum One NTS.
5. Investigation of the effects of seed coat colors (yellow, black, brown, green) on NIR spectra, calibrations, and composition analysis by NIR, as well as the possible effect of limited light penetration through the soybean coat on the accuracy of NIR measurements.
6. Novel FT-NIR/mid IR spectroscopy on single soybean seeds and composition analysis of single soybean plants.

1.5. NIR calibration techniques and methods

The general strategy for current NIR calibration development can be described as:

1. Data acquisition with standard calibration samples.
2. Wavelength range selection suitable for sample composition determinations
3. Pre-processing of spectral data
4. Matrix calculations by a calibration algorithm in order to optimize the calibration parameters, such as factors, SECV and correlation of determination values
5. Generate a calibration file with the optimized calibration parameters and make it ready for sample measurement.

The GRAMS/32 software developed by ThermoGalactic, Inc. was employed to process data and provide calibration files with the DA7000 model instrument. The TQ Analyst program developed by ThermoNicolet, Inc. was employed for calibrations with PerkinElmer's Spectrum One. Most of the calibration files and deconvoluted spectra for soybeans were developed with the aid of GRAMS/32. The TQ Analyst program, which is quite similar to GRAMS/32, was

mainly employed to develop calibrations for soy and other health foods with FT-NIR, due to its flexible computation functions in the entire NIR region.

PLSplus/IQ is the specific program that is loaded onto a computer running GRAMS/32 or GRAMS/386 for calibration data processing (Galactic Industries Corporation. 1996). PLSplus/IQ is an entirely new concept in multivariant software. It was designed to build robust calibration models and distribute these applications to the real world. While PLSplus/IQ contains a wealth of tools suitable for exploratory data analysis by the statistician, its focus is in assisting the analyst in solving problems using multivariant methods. PLSplus/IQ incorporates both quantitative and qualitative methods both alone and in combination to provide the appropriate solution to the analyst's problems. PLSplus/IQ supports the most popular and powerful multivariant algorithms in use today: Partial Least Square Type 1 (PLS-1) and Type 2 (PLS-2), Principle Component Regression Analysis (PCR), etc. PLS and PCR have been proven to be superior to such methods known as Classic Least Squares (CLS) and Inverse Least Squares (ILS) for many quantitative applications. All of these techniques are typically full spectrum techniques where the data is fit to many data points thereby improving sensitivity. In addition, they are mathematically robust and can provide optimal solutions to quantitative applications.

PLSplus/IQ is designed as a modular set of programs with a menu driven interface with some key features such as:

(1) Automatically organizes experiments and data into versatile experiment databases.

(2) Extensive diagnostics for optimizing the analysis and detecting outliers with interactive graphics.

(3) Powerful selection of preprocessing algorithms including corrections for baselines, known pathlength, indeterminate pathlength, and user definable methods

(4) Fast optimized coding providing the utmost speed currently available on desktop computers.

(5) Ability to handle very large data sets.

Partial Least Square (PLS) is a spectral decomposition technique that is quite different from PCR (Principle Component Regression Analysis), though both of them employ non-linear regression methods and computer iteration (Haaland and Thomas, 1988). PLS actually uses the concentration information during the decomposition process. The main idea of PLS is to get as much as concentration information as possible into the first few loading vectors.

In actuality, PLS is simply taking advantage of the correlation relationship that already exists between the spectral data and the constituent concentration. Since the spectral data can be decomposed into its most common variations, so can the concentration data. In effect, this generates two sets of vectors and two sets of corresponding scores; one set for the spectral data, and the other for the constituent concentrations. Presumably the two sets of scores are related to each other through some type of regression, and a calibration model is constructed.

PLS performs the decomposition on both the spectral and concentration data simultaneously. As each new factor is calculated for the model, the scores are "swapped" before the contribution of the factor is removed from the raw data. The newly reduced data matrices are then used to calculate the next factor, and the process is repeated until the desired number of factors is calculated. This makes the model equations for PLS significantly more complex then those of PCR.

There are actually two versions of the PLS algorithm: PLS-1 and PLS-2. The differences between these methods are subtle but have very important effects on the results. Like the PCR method, PLS-2 calibrates for all constituents simultaneously. In other words, the results of the spectral decomposition for both of these techniques give one set of scores and one set of eigenvectors for calibration. Therefore, the calculated vectors are not optimized for each individual constituent. This may sacrifice some accuracy in the prediction of the constituents' concentrations, especially for complex sample mixtures. In PLS-1, a separate set of scores and loading vectors is calculated for each constituent of interest, in this case, the separate sets of eigenvectors and scores are specifically tuned for each constituent, and therefore, should give more accurate predictions than PCR or PLS-2.

There is, however, a minor disadvantage in using the PLS-1 technique: the speed of calculation. Since a separate set of eigenvectors and scores must be generated for every constituent of interest, the calculation will take more time. For training sets with a large number of samples and constituents, the increased time of calculation can be significant.

PLS-1 may have the largest advantage when analyzing systems that have constituent concentration that are widely varied. For example, a set of calibration spectra contains one constituent in the concentration range of 50 to 70% and a second constituent in the range of 0.1 to 0.5%. In this case, PLS-1 will almost certainly predict better than the other techniques. In most cases, PLS methods will give better results than PCR, and PLS-1 will be more accurate than PLS-2.

The advantages of PLS methods are:

1. Combine the full spectral coverage of CLS with partial composition regression of ILS.
2. Single step decomposition and regression: eigenvectors are directly related to constituents of interest rather than largest common spectral variations.
3. Calibrations are generally more robust provided that calibration set accurately reflects range of variability expected in unknown samples.
4. Can be used for very complex mixtures since only knowledge of constituents of interest is required.
5. Can sometimes be used to predict samples with constituents (contaminants) not present in the original calibration mixture.

Some of the PLS-2 disadvantages are:

1. Calculations are slower than most classical methods, especially PLS-1
2. Models are more abstract, somewhat more difficult to understand and interpret
3. Generally, a larger number of samples are required for accurate calibrations than for PLS-1
4. Collecting calibration samples can be difficult; and one must avoid colinear constituent concentrations.

1.6. Primary experimental methods

1.6.1. Standard analytical methods for protein, oil, sugars and moisture determination

The grain composition usually includes proteins, lipids, carbohydrates, moisture, vitamins and minerals, among which the first three are the principal structural components. Most analytical methods focus on proteins, lipids and carbohydrates. This section aims at introducing conventional primary techniques for the composition analyses of grains, i.e. chemical and instrumental methods other than the relatively new and fast developing Near Infrared (NIR) technology.

1.6.1.1. Protein Analysis

Proteins are abundant in grain seed cells, and are important for biological functions and cell structures. Foods of animal origin and legumes are excellent sources of proteins. They are

composed of elements including hydrogen, carbon, nitrogen, oxygen and sulfur. Twenty a-amino acids are the building blocks of proteins; the amino acid residues in a protein are linked by peptide bonds. Nitrogen is the most distinguishing element present in proteins, ranging from 13.4% to 19.1%, due to the variation in the specific amino acid composition of proteins. Generally, protein rich in basic amino acids contain more nitrogen (Jones, 1931). The analysis of proteins is complicated by the fact that some food components possess similar physico-chemical properties. Nonprotein nitrogen could come from free amino acids, small peptides, ammonium ions, nucleic acids, phospholipids, amino sugars, porphyrin, and some vitamins. Therefore, the total organic nitrogen in foods would present nitrogen primarily from proteins and to a lesser extent from all organic nitrogen-containing nonprotein substances. Depending upon methodology, other major food components, including lipids and carbohydrates, may interfere physically with analysis of food proteins.

Numerous methods have been developed to measure protein content. The basic principles of these methods include the determinations of nitrogen, peptide bonds, aromatic amino acids, ultraviolet absorptivity of proteins, free amino groups, light scattering properties, and dye-binding capacity. In addition to factors that such as sensitivity, accuracy, precision, speed and cost of analysis, what is actually being measured must be considered in the selection of an appropriate method for a particular application. In the following sections, principles and applications are described for various major protein determination methods.

1.6.1.1(A). Kjeldahl Method

In the Kjeldahl procedure (AOAC International. 1995), protein and other organic food components in a sample are digested with sulfuric acid in the presence of catalysts. The total organic nitrogen is converted to ammonium sulfate. The digest is neutralized with alkali and distilled into a boric acid solution. The borate anions formed are titrated with standardized acid, which is converted to nitrogen in the sample. The result of the analysis represents the crude protein content of the food since nitrogen also comes from nonprotein components. In the AOCS official method, soybean protein content is 5.71 X nitrogen percentage in soybeans (AOCS Official Method Ac 4-91).

1.6.1.1(B). Biuret Method

A violet-purplish color is produced when cupric ions are complexed with peptide bonds (substances containing at least two peptide bonds, i.e. biuret, large peptides, all proteins) under alkaline conditions. The absorbance of the color produced is read at 540 nm. The color intensity (absorbance) is proportional to the protein content of the sample (Robinson and Hodgen 1940). It is a rapid and simple method for analysis of proteins, but not very sensitive as compared to the Lowry method, It requires at least 2-4 mg protein for assay.

1.6.1.1(C). Lowry Method

The Lowry method (Lowry et al., 1951; Peterson 1979) combines the biuret reaction with the reduction of the Folin-Ciocalteau Phenol reagent by tyrosine and tryptophan residues in the proteins. The bluish color developed is read at 750 nm (high sensitivity for low protein concentration) or 500 nm (low sensitivity for high protein concentration). It is very sensitive, relatively simple and more specific than most other methods.

Because of its simplicity and sensitivity, the Lowry method has been widely used in protein biochemistry. However, it has not been widely used to determine proteins in food systems without first extracting the proteins from the food mixture.

1.6.1.2. Oil Analysis

Oils or lipids are a group of substances that, in general, are soluble in ether, chloroform, or other organic solvents but are sparingly soluble in water. This solubility characteristic, rather than being a common structural feature of all macromolecules, is unique to oils (Belitz and Grosch 1987).

1.6.1.2(A). Solvent Extraction Methods

The total lipid content of a food is commonly determined by organic solvent extraction methods. The accuracy of these methods greatly depends on the solubility of the lipids in the solvent used. The lipid content of a food determined by extraction with one solvent may be quite different from the determined with another solvent of different polarity. Commonly used solvents are ethyl ether and petroleum ether (Joslyn 1970). For example, soybean oil can be

extracted from ground soybean samples with Soxhlet extraction apparatus (AOCS Official Method Ac 3-44) with petroleum ether as the solvent.

1.6.1.2(B). Time Domain Pulsed NMR determination of oil content in oilseeds

As an AOCS defined International Standard (AOCS Recommended Practice AK4-95), the low resolution Time Domain Pulsed NMR (pNMR) method is specified for the rapid determination of oil and moisture contents of commercial oilseeds simultaneously. The proton resonance frequency is required to be more than 8 MHz only. This method applies to oilseed with a moisture content of less than 10%. For seeds with higher moisture contents, drying is necessary before the oil content can be determined by pNMR. The method requires the following steps: (1) Insertion of the test sample into the magnetic field of a pNMR spectrometer. (2) Application of an alternating electromagnetic field in the form of an intense 90^{o} radio frequency (RF) pulse that excites all the hydrogen nuclei. (3) Recording of the free induction decay (FID) following 90^{o} pulse. The maximum amplitude of this signal is proportional to the total number of protons from the water and oil phases of the sample. (4) Application of a second RF pulse, a so-called 180^{o} pulse, to produce a spin-echo signal when only the signal from the oil phase contributes to the FID. (5) Calculation of the difference between the two amplitudes, which is proportional to the oil or moisture content. Automatic conversion of the measured signals, after suitable calibration of the apparatus, into percentages of oil or moisture.

This method is currently used for soybeans and sunflower seeds. The precision for oil measurement is 0.6%. Calibration samples should be homogenous and free from impurities. Samples for oil-content calibration shall be of the same species, as the test samples and of similar fatty acid compositions (especially for the analysis of rapeseeds and sunflower seeds). Oil content shall be determined using the reference method described in AOCS Ai 3-75.

1.6.1.3. Sugars determination

Carbohydrates are important in foods as major source of energy, as impartners of crucial physical properties, and as modifiers of human physiological processes (Whistler and BeMiller 1997). Carbohydrates are almost exclusively of plant origin, with milk lactose being the major exception. Monosaccharides and some di or oligosaccharides are sometimes called simple or small sugars. Higher saccharides (oligo- and polysaccharides) must be first digested, i.e. hydrolyzed to monosaccharides before absorption and utilization can occur. Nondigestible

polysaccharides can be divided into soluble and insoluble classes, and along with lignin, make up dietary fiber.

According to nutrition labeling regulations of the USFDA, the "total carbohydrates" content of a food must be calculated by subtraction of the sums of the weights of crude protein, total fats, moisture and ash from the total weight of the food (Code of Federal Regulations 1997), i.e. carbohydrate is determined by difference. Sugars are defined as glucose, fructose, sucrose and lactose. The content of "other carbohydrate" (formerly called "complex carbohydrate") is obtained by calculating the difference between the amount of "total carbohydrates" and the sum of dietary fiber, sugars and sugar alcohol (sorbitol).

1.6.1.3(A). Mono and Oligosaccharides determination: ethanol extraction

For determination of any mono- (glucose, fructose), di-(sucrose, lactose, maltose), tri-(raffinose), tetra- (stachyose) or other oligo- (maltodextrins) saccharides present, the dried, lipid free sample is extracted with hot 80% ethanol (AOAC Method 922.02, 925.05). Higher oligosaccharides from added malto- or fructooligosaccharides also may be extracted. Carbohydrates are soluble in polar solvents. The aqueous alcohol of the ethanol extract is removed under reduced pressure using a rotary evaporator and a temperature of 45-50^{o}C. The residue is then dissolved in a known, measured amount of water for further measurement. Filtration should not be required, but should be used if necessary.

1.6.1.3(B). Total Carbohydrates determination: Phenol-Sulfuric Acid Method

Carbohydrates are destroyed by heat and acid. They are particularly sensitive to strong acids and high temperatures. Under these conditions, a series of complex reactions take place, beginning with a simple dehydration reaction (Whistler and BeMiller 1997). Continued heating in the presence of acid produces various furan derivatives. These products then condense with themselves and other products to produce brown and black substances. They will also condense with various phenolic compounds to produce colored compounds that are useful for carbohydrate analysis. This is the basis for Phenol-Sulfuric Acid Method (AOAC Method 44.1.30).

This method is simple, rapid, sensitive, accurate, specific for carbohydrates and widely applied. Virtually all classes of sugars, including sugar derivatives and oligo- and polysaccharides, can be determined. The reagents are inexpensive, readily available, and stable.

A stable color is produced and results are reproducible . Under proper conditions, the phenol-sulfuric method is accurate to 2%.

1.6.1.4. Moisture Analysis

Moisture assays can be one of the most important analyses performed on a food product and yet one of the most difficult from which to obtain accurate and precise data. The following listing gives some examples in which moisture content is important to the food processor (Nielsen 1994; Pomeranz and Meloan, 1994):

1. Moisture is a quality factor in the preservation of some products and affects stability.
2. Reducing moisture is used for convenience in packaging or shipping
3. Moisture content is often specified in compositional standards
4. Computations of the nutritional value of foods require that you know the moisture content
5. Moisture data are used to express results of other analytical determinations on a uniform basis (i.e. dry weight basis)
6. A change in moisture content can lead to changes in the composition of other components in foods

Several official methods have been developed for different products, such as oven drying methods, chemical method – Karl Fischer Titration, physical methods.

1.6.1.4(A). Oven drying methods

In oven drying methods, the sample is heated under specified conditions, and the loss of weight is used to calculate the moisture content of the sample. The moisture content value obtained is highly dependent on the type of oven used, conditions within the oven, and the time and temperature of drying. Various oven methods are AOAC approved for determining the moisture in many food products. AOCS Official Method Ac 2-71 is regularly used for moisture and volatile matter in soybeans. The time required may be from a few minutes to over 24 hours. Vacuum oven drying (pressure 25-100 mm Hg) and microwave oven drying are two most frequent choices.

1.6.1.4(B). A Chemical Method – the Karl Fischer Titration

The Karl Fischer Titration is particularly adaptable to food products that show erratic results when heated or submitted to vacuum. This is the method of choice for determination of water in many low-moisture foods, such as, dried fruits and vegetables (AOAC Method 967.19), candies, chocolates (AOAC Method 977.10), roasted coffee, oils and fats (AOAC Method 984.20).

This method is based on the fundamental reaction described by Bunsen in 1853 (Mitchell and Smith 1948) involving the reduction of iodine by SO2 in the presence of water. In a Karl Fischer volumetric titration, the Karl Fischer reagent (KFR) is added directly as the titrant if the moisture in the sample is accessible. However, if moisture in a solid sample is inaccessible to the reagent, the moisture is extracted from the food with an appropriate solvent. Then the extract is titrated with KFR.

1.6.1.4(C). NMR determination of moisture content in oilseeds

NMR can be used to measure moisture and oil in food materials in a nondestructive way. Two kinds of NMR may be used: time domain low resolution NMR (pulsed NMR), and frequency domain high resolution NMR (NMR spectra). High resolution frequency domain NMR of foods is a new application in which food components are distinguished by the chemical shift (resonance frequency) of their peaks in a NMR spectrum.

Low-resolution Time Domain Pulsed NMR spectrometry is an AOCS recommended Practice (AOCS Ak 4-95) for measurement of the oil and moisture contents in oilseeds simultaneously. This method applies to oilseeds with a moisture content of less than 10%. For seeds with higher moisture content, drying is necessary before the oil content can be determined by pNMR. In order to obtain a reliable calibration curve, it is recommended that the moisture content of the calibration samples be between 6% and 10% for all seeds. The moisture content of seeds can vary depending on storage conditions. It is therefore necessary to determine it in accordance with ISO 665 just prior to calibration. The precision for moisture determination is 0.1%.

Deuterium NMR measurements were previously employed to investigate the hydration properties of starches (Baianu et al., 1990). A unique hydration behavior of potato starch that involved the presence of a tightly bound water population with anisotropic motions within the potato starch granules was revealed. The studies were extended to the hydration of potato starch

suspensions (Baianu et al., 1993). Deuterium and proton NMR allowed us to resolve three water populations "bulk (or free)" water, "weakly bound" water, and trapped water in potato starch suspensions. Deuterium NMR resolved a slowly exchangeable anisotropically bound water population within the potato starch granule structure.

1.6.2. High Resolution NMR methods for soybean oil determination

As described in last section, high-resolution frequency domain NMR of foods is a new application in which food components are distinguished by the chemical shift (resonance frequency) of their peaks in a NMR spectrum. The pattern of oil resonance reflects degree of unsaturation and other chemical properties. This is useful for chemical analysis because intensities are proportional to amounts. Liquid glycerides have been detected this way in cheese, fruits, meat, oilseeds, and other food materials (Eads and Croasmun, 1988).

High-resolution frequency domain NMR determination of oil content in intact soybean seeds was successfully conducted in Dr. Baianu's group as a primary method for providing standard soybean oil data (Klempir, 1999). This nondestructive method employs a GE 300 MHz ^{1}H NMR, in conjunction with Magic Angle Spinning (^{1}H MAS NMR) to examine the oil content in soybean seeds. The spectral acquisition time is much shorter than the solvent extraction method, and the spectral resolution remains very high. Recently, the High Resolution Nuclear Magnetic Resonance (HR-NMR) technique was extensively employed for accurate determinations of soybean protein, oil and amino acids contents (Baianu et al., 2003). The NMR data for soybean oil were compared to the corresponding NIR data, with very high correlation up to 0.99.

1.6.3. HPLC and Refractive Index methods for small sugar determination

HPLC is the method of choice for analysis of mono- and oligosaccharides, and can be used for analysis of polysaccharides after hydrolysis (Peris-Tortajada 2000). HPLC gives both qualitative analysis (identification of the carbohydrate) and, with peak integration, quantitative analysis. HPLC analysis is rapid, can tolerate a wide range of sample concentrations, and provides a high degree of precision and accuracy. HPLC requires no prior derivatization of carbohydrates, but does require micron-filter filtration prior to injection. Complex mixtures of mono- and oligosaccharides can be analyzed. Use of HPLC to determine soluble food carbohydrates has been tabulated (Hicks, 1988).

In addition to the chemical and spectrophotometric methods already mentioned, several physical methods have been developed to determine sugar content, including Refractive Index, Specific Gravity and Polarimetry. Among them, the Refractive Index method has been widely used in laboratories and industries.

When electromagnetic radiation passes from one medium to another, it changes direction, being bent or reflected. The ratio of the sine of the angle of incidence to the sine of the angle of refraction is termed the index of refraction, or refractive index (RI). The RI varies with the nature of the compound, temperature, wavelength of light, and concentration of the compound. By holding the first three variables constant, the concentration of the compound can be determined by measuring the RI (Whistler and BeMiller 1997). Thus, measurement of the refractive index is another way to determine total solids in solution. Use of RI to determine sugar concentration is accurate only for pure sucrose, or other pure solutions for liquid products. In this case, the solution must be clear. Refractometers that read directly in sucrose units are available from instrument manufacturers.

1.6.4. Liquid Chromatography method for soybean isoflavone determination

This method is applicable to the determination of total isoflavone content (daidzein and genistein) at minimum 0.05% and isoflavone family subtotals at minimum 0.02% in soy and foods containing soy (AOAC Official Method 2001.10). Test samples are extracted at 65 ºC for 2 hours in methanol-water (80:20), and the extracts are saponified at ambient temperature with NaOH solution. The extracts are acidified , filtered, and diluted with water to methanol-water (50:50). The extracts are then centrifuged to clarify them and analyzed by liquid chromatography (LC). Isoflavone glucosides and aglycons are separated on a C18 reversed-phase column with a methanol-water mobile phase and determined by UV detection at 260 nm. Results are expressed in aglycon units by summing the concentrations of the aglycon isoflavones (genistein, glycitein, and daidzein) and the aglycon equivalents of the corresponding glucoside forms (genistein, glycitein, and daidzein).

1.7. NIR instruments

1.7.1. Basic specifications of dispersive NIR instruments: IM9100, ZX-800, ZX-50, DA7000

1.7.1.1. IM9100

The PerCon Inframatic 9100 (IM9100 model) transmittance NIR instrument is designed to be able to measure components such as protein, oil and moisture in agricultural products. It is a

whole grain analyzer. Typically a sample of ca. 200 grams is poured into the analyzer. The analyzer automatically measures a number of small portions of the sample to compensate for inhomogeneous samples typical to whole grain analysis. To enable the analysis, the instrument is equipped with a Perten Instruments patented optical system that allows for the measurement of spectral absorption at 12 different wavelengths between 1000 and 1400 nm, within 1/15th second. The IM 9100 system is generally delivered with factory standard calibrations for soybean, wheat, barley and ripe seed. It can store up to 75 individual calibrations. The standard calibrations supplied with the instruments have a temperature correction included in the calibration. Ambient temperature should be kept within 10-40 ^{o}C. It will automatically compensate for fluctuations in the main voltage up to ± 10%. Before operating, at least 45 minutes warm-up time from room temperature is required. It is recommended to leave the unit on continuously. For the IM 9100, no sample preparation is required. Only whole grain samples can be analyzed. Before analysis it is recommended to clean samples from large impurities such as straws.

IM 9100 can automatically compensate for errors caused by changes in the sample temperature. The sample temperature is measured by a sensor in the funnel. This sample temperature coefficient is determined during calibration development and is specific for each calibration.

1.7.1.2. ZX-800 and ZX-50

ZX-800 and ZX-50 transmittance NIR instruments, based on discrete filters optics system, were bought from Zeltex Inc. (Hagerstown, Maryland 21740). Both of these instruments are designed to measure whole seed samples, including soybeans, wheat, barley etc.. ZX-800 has the following advantages: (1) Analysis in 30 seconds or less. (2) Store calibration for up to 10 products, four constituents each. (3) Accurate up to 99% of moisture, protein, oil etc. (4) Ability to analyze samples at temperatures from -20^{o}C to +50^{o}C, and constant temperature devised. (5) Totally solid state optics (no moving parts). (5) Direct readout in percentage form for both AS-IS (wet basis) and CM (Constant Moisture basis).

The wavelength range of ZX-800 for soybean measurement is from 918 to 1045 nm, representing by 12 discrete filters. The ZX-800 calibration software is all menu driven to facilitate use. The software allows one to do a multiple linear regression to get the constants (K values) for protein, moisture, oil or other reduction. The instrument has been carefully calibrated at the factory and adjusted again at our research location prior to use.

ZX-50 has a much smaller volume than ZX-800, but similar to ZX-800 in many features, except that it is accurate only up to 95% of moisture, protein, oil etc. while ZX-800 is 99% accurate. It is able to analyze samples at temperatures from -20°C to +50°C, and also constant temperature devised.

The ZX-50 instrument is conveniently designed to run without a long warm-up period required before taking a measurement. The wavelength region for soybean seeds measurement is set to be 893 nm to 1045 nm representing by 14 discrete filters, including major protein, oil and moisture peaks. The ZX-50 calibration software is all menu driven for easy use. The software uses a multiple linear regression to obtain the constants (K values) for protein, moisture, oil content in unknown samples. There should be anywhere from 150 to 300 samples in a ZX-50 calibration. The instruments have been carefully calibrated at the factory covering 38% to 43% protein range. Figures 1.4 and 1.5 show the instrument diagrams of ZX-800 and ZX-50. Figure 1.6 shows spectra of yellow soybean samples obtained with ZX-800 and ZX-50, exhibiting two major absorption bands. Referring to the spectra, typical NIR peak values used in routine measurements are: protein 1018 nm; oil 921 nm; moisture 970 nm; sugar 990 nm.

1.7.1.3. DA7000 Model

The novel and key technique for DA7000 is the Dual Diode Array detector system. **Diode Array** allows multiplexing and avoids the use of moving parts (Mark and Raghavachari, 2001). Basically, semiconductor fabrication techniques are used to create many diode detectors (most commonly silicon detectors) on a single substrate; these detectors are arranged in a line. This assembly is then placed at a suitable location to intercept the dispersed rays emanating from a diffraction grating. Thus, instead of passing the light through a slit through which only one wavelength at a time can be measured, all wavelengths are measured simultaneously.

The DA7000 Flexi-Mode NIR/VIS Spectrophotometer, abbreviated as DA7000 model, is a dispersive, diffuse reflectance VIS/NIR instrument that employs a tungsten halogen lamp as a chopped, high intensity, broadband energy (“white”) source of light. The dual diode array system consists of a state-of-the-art, Si diode array detector and a InGaAs diode array, allowing one to collect spectra simultaneously over the NIR wavelength range from 400 to 1700 nm. The spectra are digitized with a 32-bit DSP processor connected through a RS-485 (2MHz clock speed) to an IBM compatible PC computer. A turntable is employed for sample rotation during the acquisition of spectra in order to limit the “shadow” effect of soybean seeds. The entire operation

of the DA-7000 spectrophotometer is controlled by a Pentium computer. The typical data acquisition time per sample is 3 seconds, and the complete analysis time is 10 to 15 seconds.

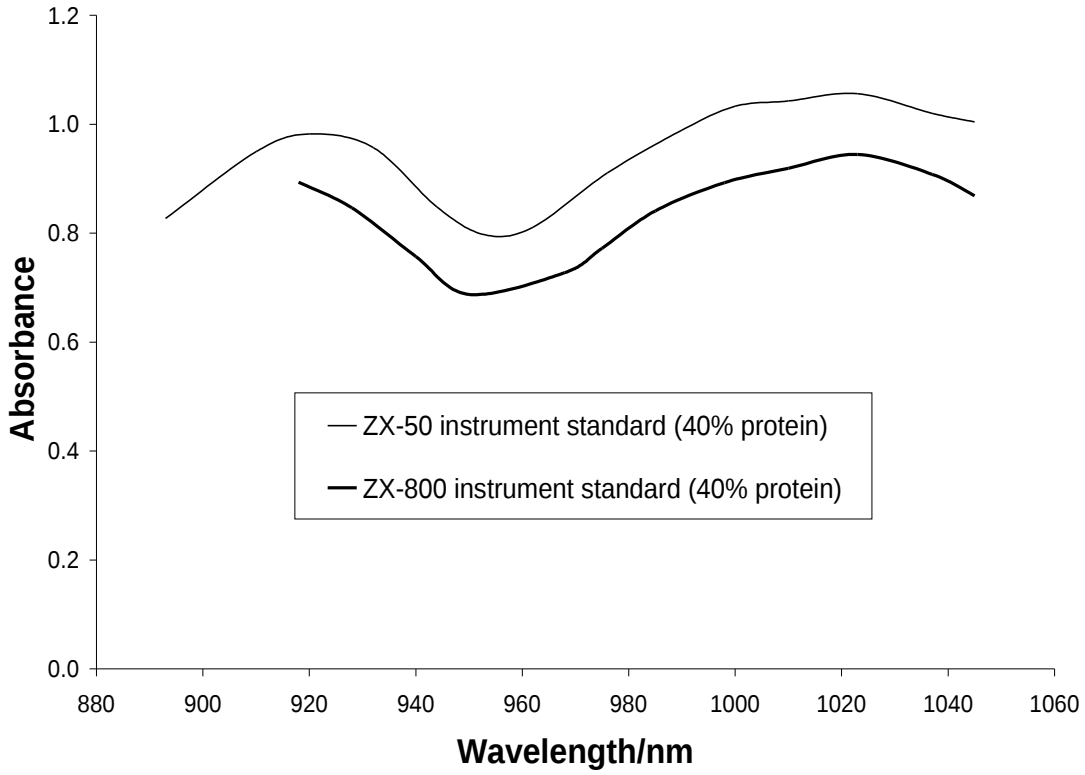

Figure 1.4. ***Spectra of yellow soybean samples obtained with ZX-800 and ZX-50.***

1.7.2. Basic specifications of PerkinElmer's Spectrum One NTS FT-NIR instrument

The PerkinElmer's Spectrum One NTS FT-IR spectrometer is a bench-top instrument that provides all the following in one self-contained unit (PerkinElmer, Inc. 2000):

1. A large, purgeable sample compartment. It can operate in ratio, single beam, or interferogram mode, i.e. dispersive or Fourier Transform mode.
2. An optical system that provides data collection over a total range of 12,000 to 4,000 cm^{-1} in the NIR region (833 to 2,500 nm) with a best resolution of 0.5 cm^{-1} (0.2 nm).
3. An electronics system based on the Motorola DSP56303 Digital Signal Processor and the Motorola 68340 Integrated Processor. It is connected to a PC, either point-to-point or over a network which utilizes Spectrum software. The nominal power consumption is 120 VA.

The entire optical system is purged and sealed at the factory. A supply of desiccant placed within the system removes water vapor and carbon dioxide that may enter. There is a range of specialized sampling accessories that have been designed specifically for the Spectrum One, including Sample Shuttle, Diffuse Reflectance, HATR, Universal ATR, Liquid Sipper and Golden Gate. Thus, it can be employed in either transmittance or reflectance mode. Attenuated Total Reflectance (ATR) mode is also available with the aid of proper ATR crystal.

Figure 1.5. shows a FT-NIR spectrum of one single yellow soybean seed (plot number 2199, protein 36.2%, oil 20.1%, moisture 9.7%) obtained with SpectrumOne NTS. It exhibits well resolved absorption bands and peaks, in which major protein, oil, moisture and sugar peaks are overlapped. It is found that PerkinElmer's SpectrumOne NTS FT-NIR Spectrometer is suitable for whole, single seed analysis. It is, of course, also reliable for bulk and powder sample analyses. Liquid samples can be also measured accurately on it.

The advantages of the FT-NIR instruments over dispersive instruments are: better resolution, faster spectral acquisition time, wider wavelength ranges, more flexibility, sensitivity towards IR range, absolute virtual instrument (AVI) function. The disadvantages are: lower sensitivity towards visible range, scattering and baseline effects in mid IR range, no automatic sample changer.

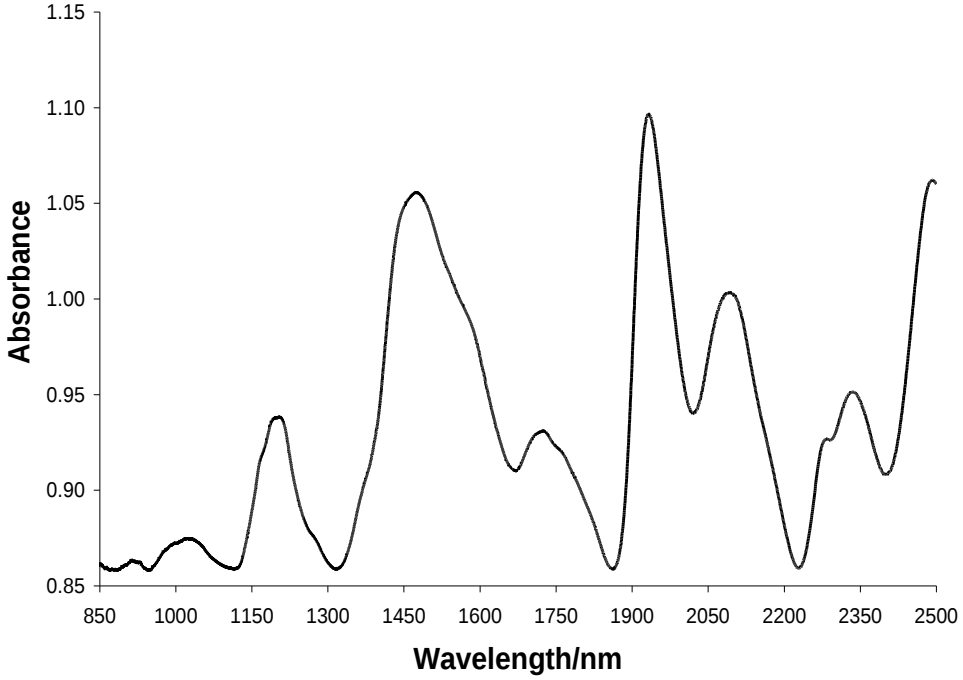

Figure 1.9. ***FT-NIR spectrum of one single yellow soybean seed (plot number 2199, protein 36.2%, oil 20.1%, moisture 9.7%) obtained with Spectrum One NTS.***

CONCLUSIONS

1. Novel calibrations for rapid, reliable and accurate composition analysis of a variety of several soy based foods and bulk soybean seeds were developed and validated in this project with a large number of different samples (N >~12, 000). The availability of such calibrations is important for establishing NIR as a secondary method for composition analysis of foods and soybeans both in applications and fundamental research.

2. Rapid, accurate, and cost-effective composition analyses of soy and other health foods are essential for improving the efficiency and quality of health foods production. The first attempts at developing Fourier Transform Near Infrared Reflectance Spectroscopy (FT-NIRS) calibrations for soy and other health foods were carried out successfully in this project. For such calibrations, food samples were employed that have a very wide range of composition for protein, fat and moisture. Novel calibrations were also developed for soy tofu and soymilk. All major food components protein, fat, moisture, total carbohydrates and fiber were quantitated. These calibrations for soy /health foods are characterized by low standard errors (< ~1.0% for protein, fat and moisture) and also by high degrees of correlation (~99%) between the NIR predicted values and laboratory reference values. The first calibrations for soybean isoflavones reported in this thesis are also characterized by low standard errors (<0.02%) and high degrees of correlation (~99%).

3. Accurate and robust calibrations were also developed for whole and ground yellow soybean samples, as well as single seed and three seeds, with the DA7000 Dual Diode Array Reflectance NIR instrument, for a wide range of compositions. The protein range was from 32.9% to 50.9, and for oil from 12.6% to 21.1% (on a wet basis). The moisture range of standard samples was from 6.1% to 11.7%. Moisture determination errors do significantly affect the accuracy of any calibration in the NIR region. Furthermore, the moisture calibrations are significantly less accurate for soybean powders in comparison with whole soybean seeds. The accuracy of whole yellow soybean calibration is at least as high as that of coarsely ground soybean calibration for both protein and oil. The difference of measurement precision between whole and ground soybeans appears not to be significant. For practical purposes, NIRS analysis of whole soybeans may be preferable to coarsely ground samples for major components, such as: protein, oil and

moisture. The content of small sugars in bulk whole soybean seeds could be measured accurately only with the whole soybean calibration. The calibrations reported in this thesis establish the DDA-NIR methodology for practical uses of measuring soybeans with high accuracy. This NIR methodology can be also reliably used in soybean genetic selection and breeding programs.

4. Four new dispersive NIR instruments ZX-50, ZX-800, IM9100 and DA7000 (a Dual Diode-Array spectrometer) were extensively evaluated with bulk soybean samples, and were found to be suitable for soybean composition analysis. These four instruments are based on technological improvements that occurred since 1998. The small measurement variances, as well as the long-term instrument which are here reported for the first time show that, the novel NIR results obtained over the last four years in this research project have established the suitability of these four NIR instruments and methodologies for analytical use in agriculture and food industry. The ZX-50 instrument can be also reliably employed for analytical use in composition analysis of single soybean plants. The measurement variances for protein and oil determined with ZX-50 are also small compared with the single plant composition variances. Wide composition range calibrations (protein from 33 to 53%, oil from 12 to 23%, and moisture from 2 to 16%) are necessary for transmittance NIR instruments such as ZX-800 and ZX-50 for soybean seed analysis.

The Dispersive Transmittance NIR instruments came with manufacturer's calibrations for soybeans, corn and wheat samples. They are simple, inexpensive, self-contained for use in practical environments and easy for operation with bulk seed samples. The Diode Array Reflectance NIR instrument DA7000, though not commercially calibrated, is superior to the dispersive instruments with: (1) higher speed, (2) higher sensitivity at constant resolution, (3) less sample amounts (1/2 or 2/3 less) for whole seed measurements, and (4) the capability to measure ground seeds. On the other hand, FT-NIR instruments offer still higher sensitivity and much higher spectral resolution than the DA7000 dispersive instrument. Such instruments also have considerably shorter spectral acquisition time in comparison with either filter-based or dispersive instruments that employ moving gratings. The FT-NIR instruments have not only the advantage of being convenient (with little or no sample preparation required), but also high speed, low cost per sample analysis and high reproducibility when calibrated correctly. These are indeed considerable advantages of FT-NIR to be weighted against the higher cost of most FT-NIR instruments.

5. Data analysis for 17 different soybean groups that included 5,000 different lines indicate that there is a high level of inverse correlation between mean protein and oil contents of soybeans (-R > 0.90). The highest protein content was found to be as high as 55%, on a dry basis. The total number of soybean samples included in this analysis is over 10,000. The level of inverse protein-oil correlation is very significant for breeding experiments for soybean protein and oil composition improvements. The distributions of protein values for such groups were found to be close to normal, thereby justifying the use of statistical methods that are standard for sampling of large populations. The protein-oil inverse correlations investigated in this research project (-R between 0.80 and 0.97) are also high for soybean groups harvested over ten years. This high correlation can be considered as another indicator of consistent and accurate measurements for soybean composition.

6. There appears to be only a very weak correlation between sugar concentration and protein or oil concentrations in the MAPIII 98 soybean variety. The trend goes up for sugar vs. oil (R= 0.47) and goes down for sugar vs. protein (R= -0.38). Such trends for sugar concentration against protein or oil concentration were not found for other developmental soybean collections and soybean accessions with high genetic diversity, i.e. Germplasm selections. This is the first extensive effort for measuring small sugars of soybean selection groups by NIR. The sugar contents in this study are found to have narrow ranges.

7. The calibrations for single soybean seed and three soybean seeds were found to be marginally acceptable for preliminary measurements. Additional, accurate calibrations for single soybean seed could be, in principle, developed with the more sensitive instruments such as the PerkinElmer's Spectrum One NTS FT-NIR spectrometer. Black seed coat effects on soybean composition were almost eliminated by measuring ground soybean samples. An alternative approach would be to measure half soybean seeds with the Spectrum One NTS FT-NIR spectrometer.

8. Improvements of whole soybean NIR analysis are possible, at least in principle, by separating out the constituent peaks from the rather complex spectra of bulk whole soybeans and by comparing the measured NIR peaks with the deconvoluted peaks of the major soybean constituents. Detailed studies aimed at NIR spectral deconvolution for bulk whole soybeans were

carried out; thus the results obtained to be shown here are consistent with the peaks assigned to be major soybean constituents that were investigated (protein, oil and moisture). The main absorption peaks in the NIR soybean seed spectra are dominated by protein and oil; certain characteristic moisture peaks are overlapped with those of protein and oil. At 5 nm resolution, deconvoluted NIR spectra of whole and ground soybeans are very similar.

9. Many other NIR applications may become possible in the future. NIRS may be used in applications that have stringent requirements such as: food quality control, pharmaceutical and biomedical applications. NIR analysis of liquid samples can be performed with some decent FT-NIR instruments. Improved calibration data processing procedures, novel calibration models and improved high quality calibrations developed in this research can be further employed for accurate analysis of other types of crop seeds, and biomedical applications such as early diagnosis of age related diseases and cancer, related nutrition aspects with animal models. Some other minor components such as fatty acids, amino acids and soy phytate may be also reliably measured with NIR instruments based on high quality calibrations.

REFERENCES

AACC. 1995. Approved Methods of Analysis, 9th ed. The American Association of Cereal Chemists. St. Paul, Minnesota.

AOAC International. 1995. Official Methods of Analysis, 16th ed., AOAC International, Gaithersburg, Maryland.

AOCS. 1998. Official Methods and Recommended Practices of the AOCS - 5th ed. The American Oil Chemists' Society. Champaign, Illinois.

Avram, M. and Mateescu, G.D. 1966. In *Infrared Spectroscopy Applications in Organic Chemistry*, 41-47.

Bach, K.E. and Li, B.W. 1991. Determination of oligosaccharides in protein-rich feedstuffs by gas-liquid chromatography and high-performance liquid chromatography. J. Agric. Food Chem. 39(4), 689-694.

Baer, R.J., Frank, J.F. and Loewenstein, M. 1983. Compositional analysis of whey powders using near infrared reflectance spectroscopy. *J. Food Sc*i. 48, 959.

Baianu, I.C. 1993. *Value Added Soybean Summi*t, Washington, D.C.

Baianu, I.C. and Kumosinski, T.F. 1994. Physical Chemistry of Food Processes. Vol. 2, 338-420.

Baianu, I.C., Cho, S.I., Stroshine, R.L. and Kruntz, G.W. 1993. Nondestructive sugar content measurements of intact fruit using spin spin relaxation time (T2) measurements by pulsed 1H magnetic resonance. Trans. ASAE. 36, 1217-1221.

Baianu I.C., Yakubu, P.I., Iozu, E.M. and Orr, P.H. 1993. Hydration of potato starch in aqueous suspensions determined from nuclear magnetic relaxation studies by oxygen, deuterium and proton NMR: Relaxation mechanisms and quantitative analysis. J. Agric. Food Chem. 41(2), 162-167.

Baianu I.C., Yakubu P.I. and Orr, P.H. 1990. Unique hydration behavior of potato starch as determined by deuterium NMR. J. Food Sci. 55(2), 458-461.

Baianu, I.C., You T., Guo, J. and Nelson, R.L. 2002. Calibration of Dual Diode Array and Fourier
Transform NIR Spectrometers for Composition Analysis of Single Soybean Seeds in Genetic Selection, Cross Breeding Experiments, *Proceedings for the 9th Biennial Conference of the Cellular and Molecular Biology of the Soybean*, August 11-14, 2002, 508.

Baianu, I. C., You T., Costescu, D.M., Lozano, P.R., Prisecaru, V. and Nelson, R.L. 2004. Chapter 11, "High Resolution Nuclear Magnetic Resonance and Near Infrared Determination of Soybean Oil, Protein and Amino Acid Residues in Soybean Seeds", D. Luthria, ed., AOCS Publications. 193-248.

Babic, M., Turan, J., Babic, L. J. and Lazic, V. 2000. The influence of drying and storing technology onto soybean mechanical properties. PTEP (Yugoslavia). 4(3-4), 57-59.

Bartholomew, D.T., and Osuala, C.I. 1988. Use of the InfraAlyzer in proximate analysis of mutton. *J. Food Sci*. 53, 379.

Belitz, H.D. and Grosch, W. 1987. *Food Chemistry*, Springer-Verlag, Berlin.

Birth, G. S. 1986. The Light Scattering Characteristics of Ground Grains. Intern. Agrophys. 2(1), 59-67.

Birth, G.S., Dull, G.G., Renfroe, W.T. and Kays, S.J. 1985. Nondestructive spectrophotometric determination of dry matter in onions. J. Am. Horticultural Soc. 110, 297.

Brim, C.A. and Burton, J.W. 1979. Recurrent selection in soybeans (Glycine max): 2. Selection for increased percent protein in seeds. *Crop Sci*. 19(4), 494-498.

Buning, P.H. and Diller, M. 2000. Rapid analysis of foods using near-infrared spectrometry (NIRS): development, use and new possibilities. *Ernahrungs Umschau*. 47(1), 15-20.

Buslov, D.K. and Nikonenko, N.A. 1997. Regularized Method of Spectral Curve Deconvolution. *Appl. Spectrosc*. 51, 666.

Buslov, D.K., Nikonenko, N.A., Sushko, N.I. and Zhbankov, R.G. 1999A. Analysis of the results of α-D-glucose Fourier transform infrared spectrum deconvolution: comparison with experimental and theoretical data. *Spectrochimica Acta Part A*. 55, 229-238.

Buslov, D.K., Nikonenko, N.A., Sushko, N.I. and Zhbankov, R.G. 1999B. Deconvolution of Fourier transform infrared spectrum of α-D-galactose: comparison with experimental and theoretical data. *Spectrochimica Acta Part* A. 55, 1101-1108.

Cameron, A.G. 1967. An assessment of the potential application of the method of attenuated total reflectance (ATR) infrared qualitative analysis to food materials. *J. Food Technol.* 2, 223.

Carrao, P.M., de-Goes, F.S. and Kikuchi, A. 2002. Extraction time for soybean isoflavone determination. *Brazilian Archives Bio. Technol*. 45 (4), 515-518.

Cho, R.K., Iwamoto, M. and Saio, K. 1987. Determination of 7S and 11S Globulins in Ground Whole Soybeans by Near Infrared Reflectance Spectroscopy Analysis. *Nippon Shokuhin Kogyo Gakkaishi.* 34(10), 666-672.

Choi, Y.S., Lee, B.H., Kim, J.H. and Kim, N.S. 2000. Concentration of phytoestrogens in soybeans and soybean products in Korea. *J. Sci. Food Agric*. 80(12), 1709-1712.

Code of Federal Regulations. 1997. Title 21, Part 101- Food Labeling, *US Government Printing Office*, Washington DC.

Cox, R.J., Lebrasseur, E., Michiels, H. B., Li, H., Voort van-de, F.R., Ismail, A., Sedman, J. and Li, H. 2000. Determination of iodine value with a Fourier transform-near infrared based global calibration using disposable vials: an international collaborative study. *J. Am. Oil Chem. Soc.* 77(12), 1229-1234.

Cregan, P., Zhu, Y.L., Song, Q.J., Hofmann, N.E., Yaklich, R.W., Specht, J.E. and Nelson, R.L. 2002. Linkage disequilibrium and association analysis for QTL discovery. Program and Proceedings for the 9th Biennial Conference of the Cellular and Molecular Biology of the Soybean, August 11-14, 2002, P.106

de-Rijke, E., Zafra, G.A., Ariese, F., Brinkman, U.A. and Gooijer, C. 2001. Determination of isoflavone glucoside malonates in Trifolium pratense L. (red clover) extracts: Quantification and stability studies. *J. Chromatography A*. 932 (1-2), 55-64.

Delwiche, S.R. 1995. Single wheat kernel analysis by near-infrared transmittance:
Protein content. *Cereal Chem*. 72, 11-16.

Dev, S.B., Rha, C.K. and Walder, F. 1984. Secondary structural changes in globular protein induced by a surfactant: Fourier self-deconvolution of FT-IR spectra. J. Biomolecular Structure & Dynamics. 2(2), 431.

Diem, M. 1993. *Introduction to Modern Vibrational Spectroscopy*, Chapter 6, John Wiley & Sons, Inc.

Diers, B.W., Keim, P., Fehr, W.R. and Shoemaker, R.C. 1992. RFLP Analysis of Soybean Seed Protein and Oil Content. *Theor. & Appl. Genetics*. 83, 608-612.

Dull, G.G., Birth, G.S., Smittle, D.A. and Lefler, R.G. 1989. Near infrared analysis of soluble solids in intact cantaloupe. *J. Food Sci*. 54, 393.

Espeja, E., Garcia, M., Marina, M. and Luisa, C. 2001. Fast detection of added soybean proteins in cow's, goat's, and ewe's milk by perfusion reversed-phase high-performance liquid chromatography. *J. Separation Sci*. 24(10/11), 856-864.

Fahrenfort, J. 1961. Attenuated total reflection. A new principle for the production of useful infrared reflection spectra of organic compounds. *Spectrochim. Acta* 17, 698.

Fox, J.D. and Robyt, J. F. 1990. Miniaturization of the three carbohydrates analyses using a microsample plate reader. Anal. Biochem. 195, 93-96.

Frankel, E.N., Nash A.M. and Snyder, J.M. 1987. A methodology study to evaluate quality of soybeans stored at different moisture levels. J. Am. Oil Chem. Soc. 64(7), 987-992.

Galactic Industries Corporation. 1996. GRAMS/32 User's Guide: PLS plus/IQ for GRAMS/32 and GRAMS/386, chapters 1-4.

Gandhi, A.P., Nenwani, M.M. and Ali, N. 1984. Investigations on the trypsin inhibitor, urease and cooking behavior of soybean Glycine max. Merr. Food Chem. 15(3), 215-218.

Garcia-Ruiz, C., Torre, M. and Marina, M.L. 1999. Analysis of bovine whey proteins in soybean dairy-like products by capillary electrophoresis. J. Chromatography A 859(1), 77-86.

Geater, C.W., Fehr, W.R. and Wilson, L.A. 2000. Association of Soybean Seed Traits with Physical Properties of Natto. Crop Sci. 40, 1529-1534.

Geater, C.W., Fehr, W.R., Wilson, L.A. and Robyt, J.F. 2001. A More Rapid Method of Total Sugar Analysis for Soybean Seed. Crop Sci. 41, 250-252.

Geladi, P., MacDougall, D. and Martens, H. 1985. Linearization and Scatter-Correction for Near -Infrared Reflectance Spectra of Meat. Appl. Spectrosc. 39(3), 491-500.

Greenspan, L., 1977. Humidity fixed points of binary saturated aqueous solutions. J. Res. of NBS. A. Phys. Chem., 81A(1), 89.

Guo J. and Baianu, I.C. 2002A. Determination of Soy and Other Health Foods Composition by Fourier Transform Near Infrared Reflectance Spectroscopy. Proceedings for the 9th Biennial Conference of the Cellular and Molecular Biology of the Soybean, August 11-14, 2002, P506.

Guo, J., You T., Baianu, I.C., Nelson, R.L. 2002B. Evaluation of Four Dispersive NIR Instruments and Methodology Development for Soybean Composition Analysis in Genetic Selection and Breeding Programs, Proceedings of the China and International Soy Conference and Exhibition 2002 (CISCE 2002), November 6-9, 2002, 412-413.

Haaland, D. M, and Thomas, E.V. 1988. Partial Least Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. Anal. Chem. 60, 1193-1202.

Harrick, N.J. 1960. Surface Chemistry from spectral analysis of totally internal reflected radiation. J. Phys. Chem. 64, 1110.

Haswell, S.J. 1992. Practical Guide to Chemometrics. Marcel Dekker, Inc., New York.

Hartwig E.E., Kuo, T.M. and Kenty, M.M. 1997. Seed Protein and Its Relationship to Soluble Sugars in Soybean, Crop Sci. 37, 770-773.

Hicks, K.B. 1988. HPLC of carbohydrates. Advances in Carbohydrate Chemistry and Biochemistry. 46, 17.

Himmelsbach, D. S., Barton II, F.E., McClung, A.M. and Champagne, E.T. 2001. Protein and apparent amylose contents of milled rice by NIR-FT/Raman spectroscopy. Cereal Chem. 78(4), 488-492.

Howard, M. 2001. Near-Infrared Applications in Biotechnology, (Biochemical and Pharmaceutical). Chapter 11 “Fundamentals of Near-Infrared Spectroscopy”, page 312, Edited by Raghavachari, R. Promega Corporation, Madison, Wisconsin.

Hymowitz, T., Collins, F.I., Panczner, J. and Walker, W.M. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64, 613-616.

Hymowitz, T., Dudley, J.W., Collins, F.I. and Brown, C.M. 1974. Estimations of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance. Crop Sci. 14, 713-715.

Iowa Agriculture and Home Economics Experiment Station. 1990. Special Report 92.

Isaksson, T., Miller, C.E. and Naes, T. 1992. Nondestructive NIR and NIT determination of protein, fat and water in plastic wrapped, homogenized meats. Appl. Spectrosc. 46, 1685.

Isaksson, T. and Naes, T. 1988. The effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. Appl. Spectrosc. 42(7), 1273-1284.

Jasansky, A and Bilanski, W.K. 1973, Spectral absorbance of soybeans and its importance for efficient heat processing. Can. Inst. Food Sci. Technol. J. 6(3), 151-155.

Jenke, D.R. 1997. Chromatographic method validation; A review of current practices and procedures I. General concepts and guidelines. Instrum. Sci. Technol. 25, 345-359.

Jones, D.B. 1931. Factors for converting percentages of nitrogen in foods and feeds into percentages of nitrogen , USDA Circular No. 183, Washington DC.

Joslyn, M.A. 1970. Methods in Food Analysis, 2nd ed. Academic Press, New York.

Kennedy, I.R., Mwandemele, O.D. and McWhirter, K.S. 1985. Estimation of sucrose, raffinose and stachyose in soybean (Glycine max) seeds. Food Chem. 17, 85-94.

Kilen, T.C. and Kou, T.M. 1994. Soybean quantitative trait loci and marker assisted breeding. Agron. Abstr. p. 107.

Klempir J.M. 1999. Nuclear Magnetic Resonance Characterization of Soybeans For Oil Content Measurement. MS Thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Knudsen, I.M. 1986. High Performance Liquid Chromatography Determination of Oligosaccharides in Leguminous Seeds. J. Sci. Food Agric. 37, 560-566.

Koga, T., Inoue, N., Asai, T. and Simizu, S. 1992. Selection of wavelength in near IR spectrometry affects accuracy of tofu lee components estimation. Chikusan no Kenkyu (in Japanese) 46(7), 773-778.

Kumosinski, T.F., Unruh, J.J. 1994. Global-secondary-structure analysis of proteins in solution. Resolution-enhanced deconvolution Fourier transform infrared spectroscopy in water. ACS Symposium Series, Philadelphia, PA. 576 (Molecular Modeling), 71.

Kyncl, J. 1982. The role of cereal production and processing Czechoslovakia, Progress in Cereal Chemistry and Technology, 1-8, Elsevier 1983.

Lamb, D. and Hurburgh, C.R. 1991. Moisture determination in single soybean seeds by near-infrared transmittance. Trans. ASAE. 34, 2123-2129.

Lanser, A.C., List, G.R., Holloway, R.K. and Mounts, T.L. 1991. FTIR estimation of free fatty acid content in crude oils extracted from damaged soybeans. J. Am. Oil Chem. Soc. 68(6), 448.

Lee, S.J., Ahn, J.K., Kim, S.H., Kim, J.T., Han, S.J., Jung, M.Y. and Chung, I.M. 2003. Variation in isoflavone of soybean cultivars with location and storage duration. J. Agric. Food Chem. 51(11), 3382-3389.

Liu, K. 1997. Soybeans Chemistry, Technology and Utilization. Chapman and Hall, Chapters 1 and 2.

Liu, K.S., Orthoefer, F. and Thompson, K. 1995. The case for food-grade soybean varieties. INFORM 6(5), 593-599.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. 1951. Protein measurement with the Folin phenol reagent. J. Bio. Chem. 193, 265-275.

Madden, H.M., 1978. Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data. Anal. Chem. 50, 1383.

Mahalanobis, P.C. 1936. On the generalized distance in statistics. Proc. Natl. Inst. Of Science of India, 2, 49.

Mansur, L.M., Orf, J.H., Chase, K., Jarvik, T., Cregan, P.B. and Lark, K.G. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36, 1327- 1336.

Mark, H. 2001. Near-Infrared Applications in Biotechnology, Chapter 11 “Fundamentals of Near Infrared Spectroscopy”, p.312, Editor Raghavachari, R. Promega Corporation, Madison, Wisconsin.

Martens, H., Jensen, S.A. and Geladi, P. 1983. Proceedings, Nordic Symposium on Applied Statistics, Stavanger, June 1983, pp.205-234.

Martin, S.K., Dye, B.W. and McBlain, B.A. 1990. Use of Hill and Short-Row Plots for Selection of Soybean Genotypes. Crop Sci. 30 (1), 74-79.

Mitchell, J. Jr., and D. M. Smith. 1948. Aquametry. John Wiley & Sons, New York.

Mora-Gutierrez, A. and Baianu, I.C. 1991. Comparisons with potato starch and potato maltodextrins. J. Agric. Food Chem. 39, 1057-1062.

Morr, C.V. 1988. Soybean utilization alternatives: a symposium sponsored by the Center for Alternative Crops and Products. pp 107-115.

Mujoo, R., Trinh, D.T. and Ng, K.W. 2003. Characterization of storage proteins in different soybean varieties and their relationship to tofu yield and texture. Food Chem. 82(2), 265-273.

Myoung, G.C., In, Y.B., Sung, T.K., Won, Y.H., Doo, C.S., Huhn, P.M. and Kwang, H.K. 2001A. Determination of protein and oil contents in soybean seed by near infrared reflectance spectroscopy. Korean J Crop Sci 46(2), 106-111.

Myoung, G.C., In, Y.B., Sung, T.K., Won, Y.H., Doo, C.S., Huhn, P.M. and Kwang, H.K. 2001B. Non-destructive method for selection of soybean lines contained high protein and oil by near infrared reflectance spectroscopy. Korean J Crop Sci 46(5), 401-406.

Nakamichi, K, Suehara, K.I., Nakano, Y., Kakugawa, K., Tamai, M., and Yano, T. 2002. Measurement of the concentrations of mannosyl erythritol lipid and soybean oil in the glycolipid fermentation process using near infrared spectroscopy. Japan. J. Near Infrared Spectrosc. 10(1), 53-61.

Nieh C.D. and Snyder, H.E. 1991. Solvent extraction of oil from soybean flour: I. Extraction rate, a countercurrent extraction system, and oil quality. J. Am. Oil Chem. Soc. 68(4), 246-249.

Nielsen, S.S. 1994. Introduction to the Chemical Analysis of Foods. Jones and Bartlet Publishers, Boston, London, Chapters 7, 10, 14.

Norris, K.H. and Williams, P.C. 1984. Optimization of mathematical treatment of raw near-IR signal in the measurement of protein in hard red spring wheat: 1. Influence of particle size. Cereal Chem. 62, 158-165.

Norris, K.H. 1983. Using gap derivatives as pre-processing for quantitative models, Food Research & Data Analysis, Proceedings of the 1982 IUFST Symposium, Martens, H. Ed., Applied Science Publishers, Oslo, 46-47.

Nzai, J. M. and Proctor, A. 1998. Determination of phospholipids in vegetable oil by Fourier transform infrared spectroscopy. J. Am. Oil Chem. Soc. 75(10), 1281-1289.

Oh, E.K., and Grossklaus, D. 1995. Measurement of the components in meat patties by near infrared reflectance spectroscopy. Meat Sci. 41, 157.

Openshaw, S.J. and Hadley, H.H. 1978. Maternal Effects on Sugar Content in Soybean Seeds. Crop Sci. 18, 581-584.

Openshaw, S.J. and Hadley, H.H. 1981. Selection to Modify Sugar Content of Soybean Seeds. Crop Sci. 21, 805-808.

Openshaw, S.J. and Hadley, H.H. 1984. Selection Indexes to Modify Protein Concentration of Soybean Seeds. Crop Sci. 24, 1-4.

Orman, B.A. and Schumann, R.A. 1992. Nondestructive single-kernel oil determination of maize by near-infrared transmission spectroscopy. J.Am.Oil Chem. Soc. 69(10), 1036-1038.

Osborne, B. G., Fearn, T. and Hindle, P.H. 1993. Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. pp 13-21, 99-117.

Ozaki Y., Sasic, S. and Jiang, J. 2001. How can we unravel complicated near infrared spectra? – recent progress in spectral analysis methods for resolution enhancement and band assignment in the near infrared region. J. Near Infrared Spectrosc. 9, 63-95.

Parreira, T.F., Ferreira, M.C., Sales, H.J. and DeAlmeida, W.B. 2002. Quantitative determination of epoxidized soybean oil using near-infrared spectroscopy and multivariate calibration. Appl. Spectrosc. 56(12), 1607-1614.

Pazdernik, D.L., Killam, A.S. and Orf, J.H. 1997. Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. Agron. J. 89(4), 679-685.

Pazdernik, D.L., Plehn, S.J., Halgerson, J.L. and Orf, J.H. 1996. Effect of Temperature and Genotype on the Crude Glycinin Fraction (11S) of Soybean and Its Analysis by Near-Infrared Reflectance Spectroscopy (Near-IRS). J. Agric. Food Chem. 44 (8), 2278 –2281.

Peris-Tortajada, M. 2000. HPLC Determination of Carbohydrates in Foods, Chapter 7, Food Analysis by HPLC, 2nd ed. Edited by Leo M. L. Nollet, Marcel Dekker Inc., New York

PerkinElmer Instruments. 2000. Spectrum One User's Guide, Shelton, CT 06484, USA.

Peterson, G.L. 1979. Review of the Folin phenol protein quantitation method of Lowry, Rosebrough, Farr and Randall. Analytical Biochemistry. 100, 201-220.

Pierce, M.M. and Wehling R.L. 1994. Comparison of sample handling and data treatment methods for determining moisture and fat in Cheddar cheese by near infrared spectroscopy. J. Agric. Food Chem. 42, 2830.

Pfeffer, P. and Gerasimowicz, W. (Eds.). 1987. NMR in Agriculture. CRS Publishers: Boca Raton, FL.

Pomeranz Y. and Meloan, C.E. 1994. Food Analysis Theory and Practice, third edition, chapter 33, 567-574. Chapman & Hall, New York, NY.

Psotka, J. and Shadow, W. 1994. NIR analysis in the wet corn refining industry – A technology review of methods in use. Intern. Sugar J. 96, 358.

Rasco, B.A., Miller, C.E. and King, T.L. 1991. Utilization of NIR spectroscopy to estimate the proximate composition of trout muscle with minimal sample pretreatment. J. Agric. Food Chem. 39, 67.

Reffer, J.A. and Martoglio, P.A. 1995. Uniting Microscopy and Spectroscopy, Practical Guide to Infrared Microspectroscopy, Chapter 2, p.41, Edited by Howard J. Humecki, Marcel Dekker, Inc.

Rinne, R.W., Gibbons, S., Bradley, J., Seif R. and Grim, C.A. 1975. Soybean Protein and Oil percentages Determined by Infrared Analysis. USDA-ARS-NC-26, 1-4.

Robinson, H.W. and Hodgen, C.G. 1940. The biuret reaction in the determination of serum protein. 1. A study of the conditions necessary for the production of the stable color which bears a quantitative relationship to the protein concentration. J. Bio. Chem. 135, 707-725.

Rodriquez-Otero, J.L., Hermida, M. and Cepeda, A. 1995. Determination of fat, protein and total solids in cheese by near infrared reflectance spectroscopy. J. AOAC International. 78, 802.

Rodriguez-Saona, L. E., Fry F.S. and Calvey, E.M. 2000. Use of Fourier Transform Near-Infrared Reflectance Spectroscopy for Rapid Quantification of Castor Bean Meal in a Selection of Flour-Based Products. J. Agric. Food Chem. 48(11), 5169-5177.

Sato, T, Abe, H., Kawano, A., Ueno, G., Suzuki, K. and Iwamoto, M. 1994. Near-Infrared Spectroscopic Analysis of Deterioration Indices of Soybeans for Process Control in Oil Milling Plant. J. Am. Oil Chem Soc. 71(10), 1049-1055.

Savitsky, A. and Golay, M.J. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627-1639.

Sehern, N.A. and Lambert, J.W. 1984. Effect of stratification for percent protein in two soybean populations. Crop Sci. 24, 225-228

Shadow, W. 1998. Rapid Analysis for the Food Industry Using Near-Infrared Spectroscopy. Perten Instruments North America, Inc., 1-17.

Silvela, L., Rodgers, R., Barrera, A. and Alexander, D.E. 1989. Effect of selection intensity and population size on percent oil in maize, Zea mays L. Ther. Appl. Genet. 78(2), 298-304.

Simpson, A.M. Jr. and Wilcox, J.R. 1983. Genetic and phenotypic associations of agronomic characteristics in four high protein soybean populations [Glycine max]. Crop Sci. 23, 1077.

Smith, K.J. and Huyser, W. 1987. In Soybeans: Improvement, Production, and Uses (J.R. Wilcox ed.), p.19.

Society for The Study of Natto. 1990. Methods of natto research. 1st ed. (In Japanese). National Food Research Institute, Ministry of Agriculture, Forestry and Fisheries, Japan.

Song, T., Barua, K., Buseman, G. and Murphy, P.A. 1998. Soy isoflavone analysis: quality control and a new internal standard. Am. J. Clinical Nutrition. 68 (6), supplement, 1474S-1479S.

Specht, J., Nelson, R.L., Webster, R., Naidu, S., Ainsworth, E. and Ort, D. 2002. A genomic perspective on the soybean protein(+)/oil(-)/yield(-) enigma. Program and Proceedings for the 9th Biennial Conference of the Cellular and Molecular Biology of the Soybean, August 11-14, 2002, P.202.

Steiner, J., Termonia, Y. and Deltour, J. 1972. Comments on smoothing and di€erentiation of data by simplified least square procedure. Anal. Chem. 44, 1906-1909.

Stuber, C.W. 1992, Biochemical and Molecular Markers in Plant Breeding. Plant Breeding Rev. 9, 37-61.

Tajuddin, T., Watanabe, S., Masuda, R., Harada, K. and Kawano, S. 2002. Application of near infrared transmittance spectroscopy to the estimation of protein and lipid contents in single seeds of soybean recombinant inbred lines for quantitative trait loci analysis. J. Near Infrared Spectrosc. 10(4), 315-325.

Tekel, J., Daeseleire, E., Heeremans, A. and Peteghem, C.V. 1999. Development of a simple method for the determination of genistein, daidzein, biochanin A, and formononetin (biochanin B) in human urine. J. Agric. Food Chem. 47(9), 3489-3494.

Thanh, V.H. and Shibasaki, K.J. 1976. Major proteins of soybean seeds. A straight-forward fractionation and their characterization. Agric. Food Chem. 24(6), 1117-1121.

Tomomatsu, H. 1994. Health effects of oligosaccharides. Food Technol. Oct, 61-65.

Wang, D., Dowell, F.E. and Lacey, R.E. 1999. Single wheat kernel color classification by using near-infrared reflectance spectra. Cereal Chem. 76(1), 30-33.

Wang, X., Zhao, S. and Wang, R. 2001. Analysis of soybean lecithin by supercritical fluid chromatography. Sepu. 19(4), 344-346 (Journal written in Chinese).

Wehling, R.L., Pierce, M.M. and Froning, G.W. 1988. Determination of moisture, fat and protein in spray dried whole egg by near infrared reflectance spectroscopy. J. Food Sci. 53, 1356.

Wesley, I.J., Osborne, B.G., Anderssen, R.S and Skerritt, J.H. 1999. Curve fitting applied to near infrared deconvolution of wheat functional proteins in flour. Near Infrared Spectroscopy, Proceedings of the International Conference, 9th, Verona, Italy, p.215.

Wehrmann, V.K., Fehr, W.R., Cianzio, S.R. and Cavins, J.F. 1987. Transfer of high seed protein to high yielding soybean cultivars. Crop Sci. 25, 927-931.

Whistler, R.L. and BeMiller, J.N. 1997. Carbohydrates Chemistry for Food Scientists, Eagen Press, St. Paul, MN.

Wilcox, J.R. 1998. Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Sci. 38, 1536-1540.

Wilcox, J.R. and Cavins, J.F. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35, 1036-1041.

Wilcox J.R. and Shibles, R.M. 2001. Interrelationship among Seed Quality Attributes in Soybean, Crop Sci. 41, 11-14.

Wilson, J.M., Kramer, A. and Ben-Gera, I. 1973. Quantitative Determination of Fat, Protein and Carbohydrates of Soy Products with Infrared Attenuated Total Reflectance. J. Food Sci. 38(1), 14-17.

Williams, P. and Norris, K. 1987. Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, Inc. St.Paul, MN. pp 43-47.

Williams, P.C. and Sobering, D.C. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. J. Near Infrared Spectrosc. 1, 25-32.

Wilson, J.M., Kramer, A. and Ben-Gera, I. 1973. Quantitative determination of fat, protein, and carbohydrates of soy products with infrared attenuated total reflectance. J. Food Sci. 38(1), 14-17.

Wilson, R.F. 1991. Designing Value-Added Soybeans for Markets of the Future. *American Oil Chemists' Society Publication*, Champaign, Illinois, pages 17, 53, 80, 88, 109, 115.

Xu, H. and Wilcox, J.R. 1992. Recurrent selection for maturity and percent seed protein in Glycine max based on S0 plant evaluations. *Euphytica 62*, 51-57.

Yazdi-Samadi B., Rinne, R.W., and Seif, R.D. 1977. Components of Developing Soybean Seeds: Oil, Protein, Sugars, Starch, Organic Acids and Amino Acids. *Agron. J*. 69, 481-486.

You, T., Guo, J., Baianu, I.C. and Nelson, R.L. 2002A. Determination of Isoflavones Contents for Selected Soybean Lines by Fourier Transform Near Infrared Reflectance Spectroscopy". *Proceedings for the 9th Biennial Conference of the Cellular and Molecular Biology of the Soybean*, August 11-14, 2002, P505.

You, T., Guo, J., Baianu, I.C. and Nelson, R.L. 2002B. Calibration of dual-diode array and fourier transform near infrared reflectance spectrometers for composition analysis of single

soybean seeds in genetic selection, cross-breeding experiments. *Program and Proceedings for the 9th Biennial Conference of the Cellular and Molecular Biology of the Soybean*, August 11-14, 2002, page 508.

You T., Guo, J., Baianu, I.C., Nelson, R.L. 2002C. Rapid Determination of Protein, Oil, Moisture, and Isoflavones Contents of Single Soybean Seeds by Fourier Transform Near Infrared Reflectance Spectroscopy. Proceedings of the China and International Soy Conference and Exhibition 2002 (CISCE 2002), November 6-9, 2002, 414-415.

You, T., Baianu, I.C., Guo, J., Nelson, R.L. 2011a. Diode-Array Near Infrared Spectroscopy for Rapid Analysis of Soybeans: Light Scattering Corrections for Intact and Ground Soybean Seeds. (*submitted*).

You T., Guo, J., Baianu, I.C. and Nelson, R.L. 2011b, The impact of soybean composition ranges on calibration development with PLS-1 algorithm by theoretical simulations, . (*submitted*)

APPENDIX

Description of soybean groups

1. MAPIII developmental soybean groups, selected by Dr. R. L. Nelson's group, are F3 BC2 lines that were developed using Williams 82 soybean line as the recurrent parent and three sources of high protein as the donor parents. The numbers after the group names such as 97 and 98 indicate years of harvest. Bell, Fisher, EG1000, Hume and MIV are the locations for soybean planting.

2. OP MAP groups contain several populations of F3 generation lines, which carried 50% genes from each parent.

3. The Germplasm soybean samples belong to a big type of voted genetically variable materials and are genetically distinct. Their protein, oil, sugar and other components are controlled by different genes and they all vary independently.

4. The ProGen lines are F5 experimental lines developed by crossing between high protein lines.

5. The ProSel group includes experimental lines that were developed by crossing US varieties with high protein lines introduced from other foreign countries.

6. The 2000 Yield Map group involves a set of random, full-sib lines. One parent is a commercial soybean variety and the other parent was an experimental line that Dr. R. L. Nelson developed by crossing two primitive soybean lines from China.

7. The "BC High Protein" group is part of a breeding project to developed back crossed derived high protein lines to help study the genetics of high protein. The "High Protein" lines are part of the breeding project to develop high protein lines that will also be relatively high yielding.

8. The 2002 Proteintest-single plant group contains single plants that were selected for high protein in an experiment to combine high protein genes from *G. max* and *G. soja*.