

# **Internet-Enabled Near-Infrared Analysis of Oil Seeds**

Ching-Hui Tseng, Kangming Ma, Nan Wang

AgroSolution/QTA  
Cognis Corporation  
Cincinnati, Ohio USA

## **Abstract**

Near infrared technique has been used in the agricultural industry for fast and nondestructive analysis of oil seeds. Current NIR systems have their advantages but also have disadvantages. In order to fully utilize the advantages of the current NIR systems but minimize the disadvantages, an Internet-enabled NIR system has been developed. This system includes one central processor and unlimited number of client NIR units. NIR calibration models of different applications are developed and stored in the central processor to be shared by all the client NIR units. Instrument performance of individual client NIR unit can be remotely monitored and most of the problems are possible to be solved remotely. The obtained spectra and analyzed data are stored in the central processor and can be accessed via Internet for technique application and business management. The combination of NIR and Internet is tending to be an attractive tool for oil seeds analysis.

## **Introduction**

Compositional analysis of oil seeds plays an important role for the quality control and assurance of oil seeds in both agriculture and food industries. Wet chemistry analysis methods of oil seeds are often time consuming, labor intensive, and expensive. Different analytical methods are required for each oil seed parameter or trait of interest. In

addition, each method's analysis time can last hours or days. Classically, oil seed analysis requires Kjeldahl protein analysis, extraction or pulsed nuclear magnetic resonance (NMR) for total oil analysis, oven methods or moisture meter for moisture analysis, gas chromatography (GC) or high performance liquid chromatography (HPLC) for fatty acid composition analysis, liquid-liquid extraction and subsequent spectrophotometry for chlorophyll analysis, enzymatic hydrolysis followed by colorimetric or spectrophotometric analysis for total glucosinolates analysis, or HPLC for total and individual glucosinolates analysis. Unfortunately, these methods often resulted in sample destruction during analysis process. NMR has been used for nondestructive whole seeds analysis and the technique is rapid and accurate. However, it has only found use in the analysis of total oil and moisture in seeds.

### **Near Infrared applications for oil seeds analysis**

Agricultural applications of near-infrared spectroscopy (NIR) started when Karl Norris applied the statistical regression data analysis method in NIR diffuse reflectance studies in the 1960's.<sup>1</sup> Since then, many NIR related research and applications have been reported for oil seed analysis. NIR is a rapid, nondestructive, inexpensive, and accurate method for the analysis of traits, material characteristics in seeds, grains, and other types of materials. Also, modern NIR is capable of producing multiple results from one single analysis of intact samples. Among the oil seeds, soybean is not only an important animal feed but also an important human food source due to its nutritional benefit and high oil and protein content. Much effort has been placed on NIR soybean analysis over the years. In earlier times, seed grinding was often required, and sometimes, other treatment

was needed to form a paste type of sample form. In 1968, Ben-Gera and Norris used NIR to measure the moisture content of soybean.<sup>2</sup> Hymowitz in 1974, Rinne in 1975, reported NIR oil and protein determination of soybean.<sup>3-4</sup>

In Asian countries, soybean and soybean derived products play important roles in the food market. Protein, oil, and moisture contents affect their usefulness in processing for different types of traditional foods such as tofu, miso, and natto. The intact soybean analysis for fat, protein and moisture is more preferred.<sup>5</sup> The importance of nondestructive analysis is that it can save considerable labor and time. Health benefits of soybean derived food products make it a more and more important and popular food source in other regions as well. Enhancing the quality of protein and oil content has played an important role in soybean crop improvement. NIR analysis of amino acid and fatty acid composition will benefit the breeding process and commercial soybean testing. Pazdernik compared the whole seed and ground soybean NIR analysis for 17 amino acids and 5 fatty acids.<sup>6</sup> These results showed that more accurate results were obtained with ground samples.

Sunflower is another important class of oilseeds. The nutritional benefits of sun butter make it an attractive alternate to people who are allergic to peanut butter. Studies were reported on the NIR determination of moisture, oil, protein and fiber contents.<sup>7-8</sup> The nondestructive property of NIR has allowed the seeds to be analyzed prior to germination. In contrast, the analysis of fatty acid composition of machine-husked sunflower seeds has been reported.<sup>9</sup> Perez-Vicha explored the use of intact sunflower seeds and compared it with other types of sample forms for NIR analysis of oil content and fatty acid composition.<sup>10</sup> Sunflower oil, husked seeds, and meal all gave excellent

correlations ( $r^2$  from 0.90 to 0.99), while intact seeds gave lowered NIR correlations ( $r^2$ : 0.76 to 0.85). Despite the lower correlations, NIR can be used as a pre-screening tool due to its convenience.

Rapeseed and related seeds are another important class of oil seeds. Antinutritive compounds such as glucosinolates and sinapic acid esters (SAE) in the rapeseed and related Brassicaceae family may affect the nutritional value of Brassica meal and limit its use as a high quality protein source. Nondestructive NIR was used to search for both germplasm and breeding Brassica materials with reduced SAE content.<sup>11-12</sup> The lowest SAE samples were analyzed by a reference method, which confirmed the low SAE levels. Glucosinolate and erucic acid contents were simultaneously determined using intact rapeseeds on both reflectance and transmittance NIR spectrometers.<sup>13</sup> The NIR analysis of glucosinolate content gave acceptable accuracy when compared to the wet chemistry colorimetric method. NIR analysis for oil, protein, glucosinolates and chlorophyll were developed and compared on three whole seed analyzers.<sup>14</sup> No significant differences were found between the instruments for oil (SEP 0.43-0.55%), protein (SEP 0.35-0.42%) and glucosinolates (SEP 2.4-3.8 mM/g). However, it was shown that only one instrument could effectively analyze chlorophyll. For intact rapeseed samples, NIR was used as a rapid method to estimate fatty acid composition.<sup>15</sup> Excellent correlation with GC results was obtained for oleic, linolenic and erucic acids for all sample sizes. Calibrations for the other fatty acid components were less accurate.

For intact Ethiopian mustard-seeds, NIR fatty acid composition analysis gave high accuracy and correlation for the major acids-oleic, linolenic, linoleic, and erucic.<sup>16</sup> The ability of NIR to discriminate among different fatty acid profiles was considered due

to changes within 6 spectral regions, 1140-1240, 1350-1400, 1650-1800, 1880-1920, 2140-2200, and 2240-2380 nm. All six regions are associated with fatty acid absorbers.

Over the years, other types of oilseeds have also been studied by the NIR technique. Cotton seed is a major oilseed in domestic and international markets. Products derived from cotton seeds are an important part of cotton production. A rapid method for oil content analysis of cotton seeds in cotton breeding and testing is desirable. Kohel attempted of using NIR to measure cottonseed oil content.<sup>17</sup> The calibration model gave good correlation, but when tested for unknown samples, the results were not acceptable. It has been reported that NIR has been used for ground and intact flaxseed oil analysis.<sup>18</sup> The calibration using whole flaxseed was equal in precision as that of the ground samples.

Single seed analysis is another area of interest especially for breeding programs. Crop improvement needs to evaluate large numbers of seeds in small quantity and even with single seeds. NIR as a rapid, inexpensive and non-destructive analysis technique is highly desirable in this area. It often can produce analysis results for multiple traits and properties, simultaneously. To be useful in the breeding programs, the analysis results must be precise and accurate enough for genetic segregate separations. The selected seeds with desirable traits can then be used in the germination process for the next step in development. Velasco reported in 1999 a study of simultaneous NIR analysis of seed weight, total oil content, and fatty acid composition in intact single rapeseed.<sup>19</sup> Excellent correlation was found for oleic ( $r=0.92$ ), and erucic ( $r=0.94$ ), but not for linoleic ( $r=0.75$ ) and linolenic ( $r=0.73$ ). In 1992, Orman studied the nondestructive prediction of oil content in single corn kernels using NIR transmission spectroscopy (NITS).<sup>20</sup>

## **NIR calibration model development**

Before an NIR spectrometer can be used to predict any compositional property of any oil seed, it must have a calibration model (or equation) for that property first. In order to build a good calibration model (or equation), a large set of sample seeds covering all the possible sample variances like concentration variance, variety variance, color, size, growing season, growing year, growing location, moisture level etc, with reliable chemical data obtained from a standard or an official primary method is required. After measuring the NIR spectra of the sample seeds, a calibration model can be built using any Chemometric tool such as multiple linear regression (MLR), partial least squares (PLS) regression, artificial neural network (ANN) etc. from the obtained NIR spectra and the primary data.<sup>21-23</sup> This calibration model can then be used to predict the NIR spectrum of a unknown oil seeds. For example, if a NIR calibration model has been built for the analysis of total oil in soybean, it can then be used to predict the total oil of an unknown soybean sample using the obtained NIR spectrum.

## **Advantages and disadvantages of the current NIR systems**

NIR has many advantages to be used in oilseeds analysis as previously described, but there are also some problems that limit the capability of NIR for oilseed analysis. In order to lower the price and simplify the operation, some NIR manufacturers only provide a single component NIR analyzer such as a moisture analyzer, protein analyzer, oil analyzer, etc. Most of them are filter type NIR systems that have few filters with wavelengths related only to a specific component. Also, there are some low-cost NIR

analyzers with the capability of multi-component analysis. The NIR manufacturers usually provide calibration models for some common traits such as moisture, oil, protein, etc. of some common grains. These NIR systems use low-cost silicon detectors that detect light in the short-wavelength NIR (SWNIR) region from 850 nm to 1050 nm as shown as Figure 1. The SWNIR region usually has broad spectral features with very low absorption coefficients. Therefore, in most of the cases, they are only used to analyze total amounts of moisture, oil, and protein. If more specific structural properties or components such as iodine value, individual fatty acids, amino acids, trans-fat etc. are needed, a more powerful NIR system will be necessary. Due to the low absorption coefficients of SWNIR, the sampling method of this type of NIR system is usually designed to use transmittance measurements (shown as Figure 2) rather than reflectance measurements (shown as Figure 3) in order to enhance the spectral features. It comes with the concern that different grains may have different colors or sizes, and the color and the size of the oil seeds affect the penetrating ability of incident light. Generally, the darker the color or the smaller the size, the lower is the penetration efficiency. Therefore, a different sample container with a different sample pathlength may have to be exchanged if different grain must be analyzed.

Due to the hardware variances such as light source variance, mechanical variance, optical variance, and detector variance between different NIR systems, the calibration models provided by the instrument manufacturer usually have to be recalibrated before use. In most cases after a period of time of use, these models also have to be recalibrated. This is because the spectral quality can be affected by the light source decay or the drift

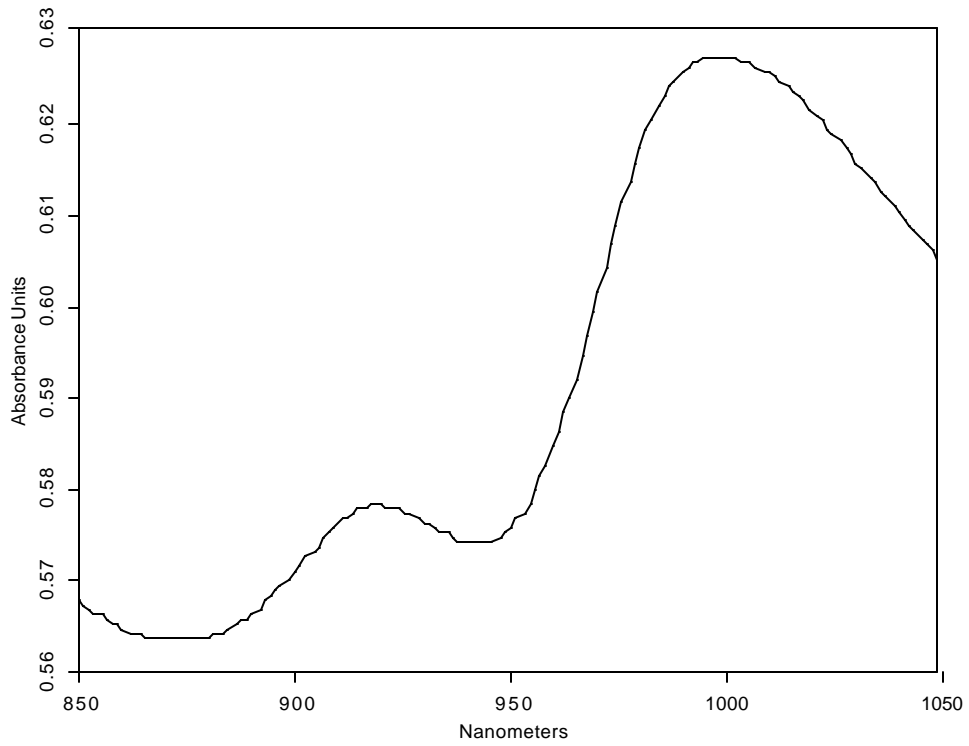


Figure 1 NIR spectrum of wheat measured from 850 nm to 1050 nm

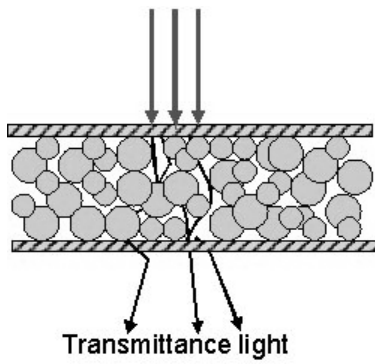


Figure 2 Transmittance measurement

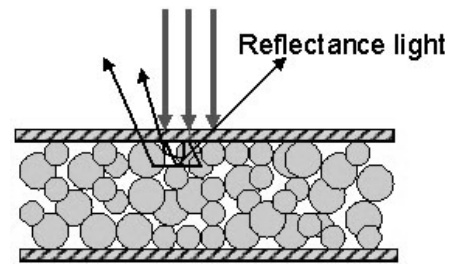


Figure 3 Reflectance measurement

of the optical alignment which in-turn affects the prediction. Users normally do not know when and how much the prediction value was shifted until a calibration expert runs the standard samples for a calibration check.

Other than the simplified NIR analyzers, there are also many full-functioned NIR systems available in the market. They usually cover wide spectral range, have good signal-to-noise ratio, good wavelength accuracy, and capable of analyzing more materials and traits. However, this kind of NIR system is usually more expensive and not suitable for the use in the field or in an adverse environment. Also, the user has to be well trained for model development and maintenance. Even if the user has been well trained or has a good background in spectroscopy and Chemometrics, there is still no guarantee that the calibration models are rugged enough without drift. In most of cases, different laboratories build their own calibration models according to the primary data from their own resources. Of course, it is inevitable that there are biases between different laboratories even the same primary method has been used. The primary method may also be different such as the oil content of an oil seed can be from different extraction processes. Also, the weight percentage of a trait can be recorded using different moisture bases, such as “as is”, dry base, 13.5% of moisture, etc. Therefore, the prediction value of the same trait in the same sample can be different from different NIR systems in different laboratories.

### **Emerging trend: NIR network and Internet-enabled NIR system**

In order to fully utilize the capability of NIR technology but still keep the simplicity of operation and maintenance, a NIR network concept has been introduced. As shown in

Figure 4, the network consists of one central processor and many NIR analyzers. Actually, there is no specific limit of the number of individual analyzers. It depends on the computing capability of the central processor. The individual NIR analyzer measures the spectrum of the oil seeds. The obtained spectrum is sent to the central processor for storage and calculation. The calculated result is then stored in the central database and sent back to the individual analyzer for display. Any authorized computer connected to the network can also access the database of the central processor to obtain the results or spectra measured by any specified NIR analyzer or a group of specified analyzers. The NIR network can be within an organization or a private company. It can also be the entire Internet and becomes an Internet-enabled NIR system.

### **Advantages of the NIR network**

Within the NIR network, a rugged and fully-functional NIR system can be used as the individual analyzer since the users of the individual analyzers do not have to develop application methods and maintain them. The only must-have function of the analyzer is to measure the NIR spectrum and communicate with the central processor. The calibration models can be developed remotely by spectroscopic experts and stored in the central processing computer. There is only one calibration model developed and stored in the central processor for the same application used by all individual analyzers. Therefore there is no primary data bias between different analyzers. However, there may still be hardware and optical bias between different systems. The accuracy of spectral wavelengths significantly affects the accuracy of NIR predictions. A Fourier transform near infrared (FT-NIR) system has better wavelength accuracy compared to other types

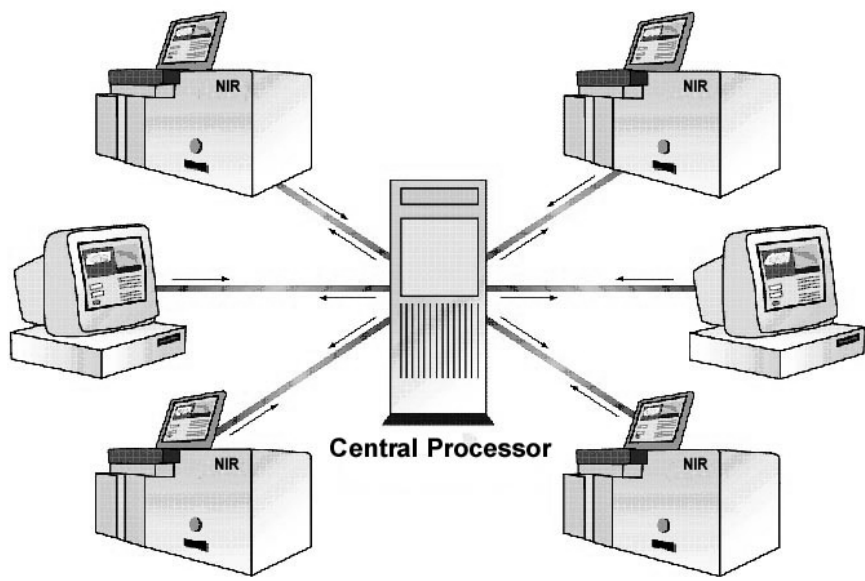


Figure 4 NIR network

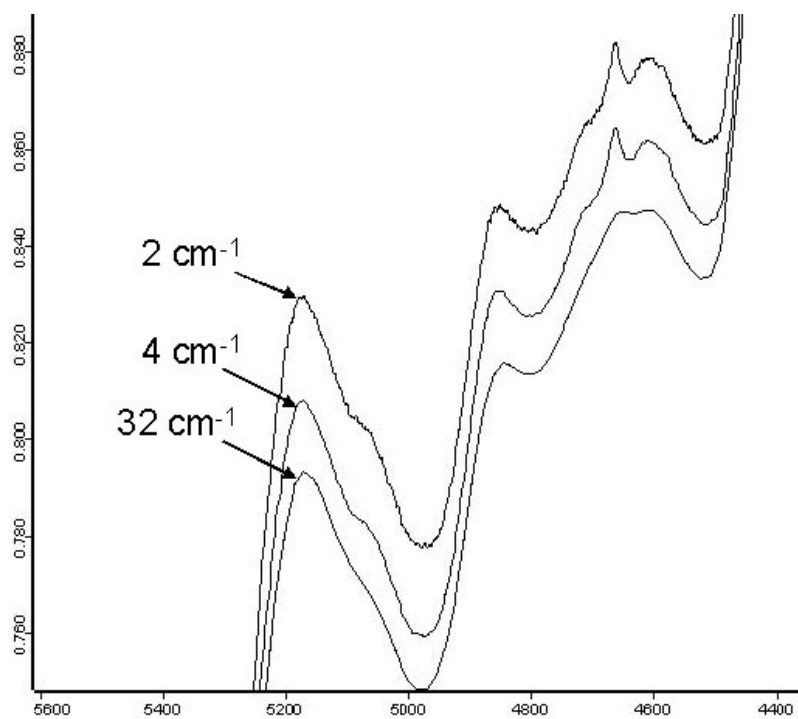


Figure 5 NIR spectra of sunflower with different resolutions

of NIR systems; hence, it is good candidates for the analyzer used in a NIR network. It also has the advantage of adjustable spectral resolutions. Figure 5 shows NIR spectra of ground sunflower seeds measured by a FT-NIR system with different spectral resolutions. It is obvious that a NIR spectrum with a lower spectral resolution exhibits a better signal-to-noise ratio and a NIR spectrum with a higher spectral resolution is noisier if the sample measuring times are the same. Therefore, if an application requires a better signal-to-noise ratio in order to distinguish minor concentration differences, a low-resolution measurement can be applied. A high resolution can also be applied if an application needs detailed spectral feature identification. Of course, it is difficult for general users to know how to choose optimum measurement parameters of a FT-NIR system. However with the NIR network, all the methods can be remotely developed by experts at the end of the central processor and used by all the individual analyzers. Since the application methods can be developed remotely and the individual analyzer is a high performance NIR system, there is no limit of the applications for each analyzer.

It is also possible for the central experts to monitor the instrument performance from the spectra transferred to the central processor. If any problem happened with the individual analyzer, remote diagnosis and possible problem-solving can be processed from the central location. As previously described, in order to build a robust oil seed calibration model, the calibration samples should cover all the possible sample variances. However, it is always hard to obtain enough representative calibration samples covering all the possible sample variances while building a calibration model. The oil seeds produced this year may have a different sample matrix than the previous year's matrix. A dry year may produce the oil seeds different from a wet year. Therefore, a calibration

model of oil seeds may have to be updated periodically or when the prediction error is not acceptable until the calibration model is robust with the analysis of all the possible analyzed samples. With the NIR network, all the models can be updated for all the analyzers simultaneously, if necessary.

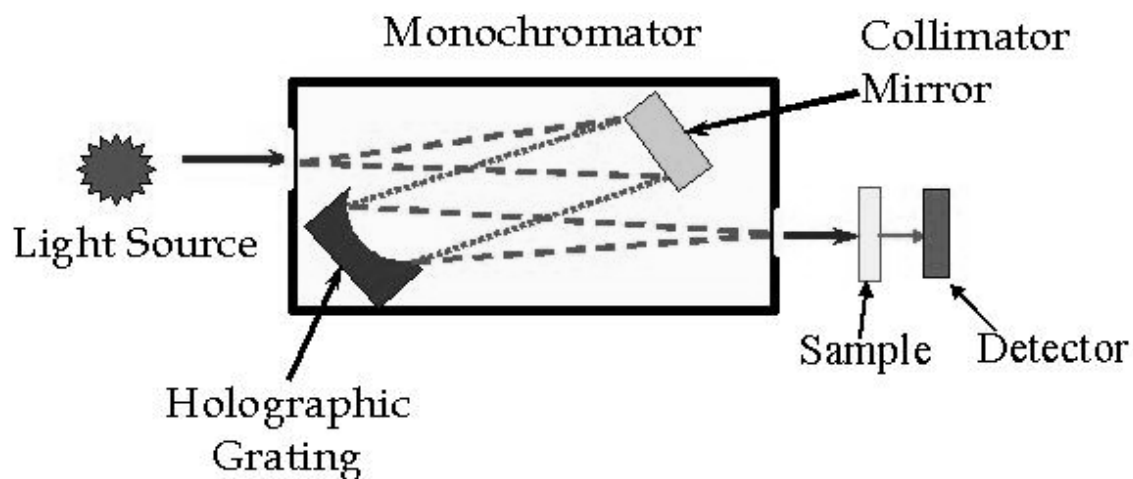
Another advantage of the NIR network is the data storage and data distribution. Since the spectra obtained by the analyzer are not stored locally, it will never run out of the storage space and there is no need to backup spectral data from the individual analyzer. With all of the data stored in the central computer, the data can be distributed to or accessed by any authorized computer connected to the network. If the central computer is connected to the Internet, the data can then be distributed to or accessed from any Internet connected computer throughout the world. In addition, once the data has been backed up from the central computer, the data from all the NIR network analyzers has been backed up also.

Of course, the central processor is the most important part of the NIR network. It should have the capability of taking care of all the model prediction works from all the analyzers and storing all the spectra and results in specified databases. PLS (Partial Least Squares), an excellent algorithm for the linear correlation modeling, and ANN (Artificial Neural Network), an excellent algorithm for the non-linear correlation modeling, are two examples of algorithms that can be used to build the calibration models which are used by all of the NIR analyzers. It is also possible that in the future a newly developed Chemometric algorithm will do a better job than all of the current existing algorithms for the calibration modeling. If the central processor will be upgraded with the capability of the new Chemometric algorithm, all the analyzers will have this

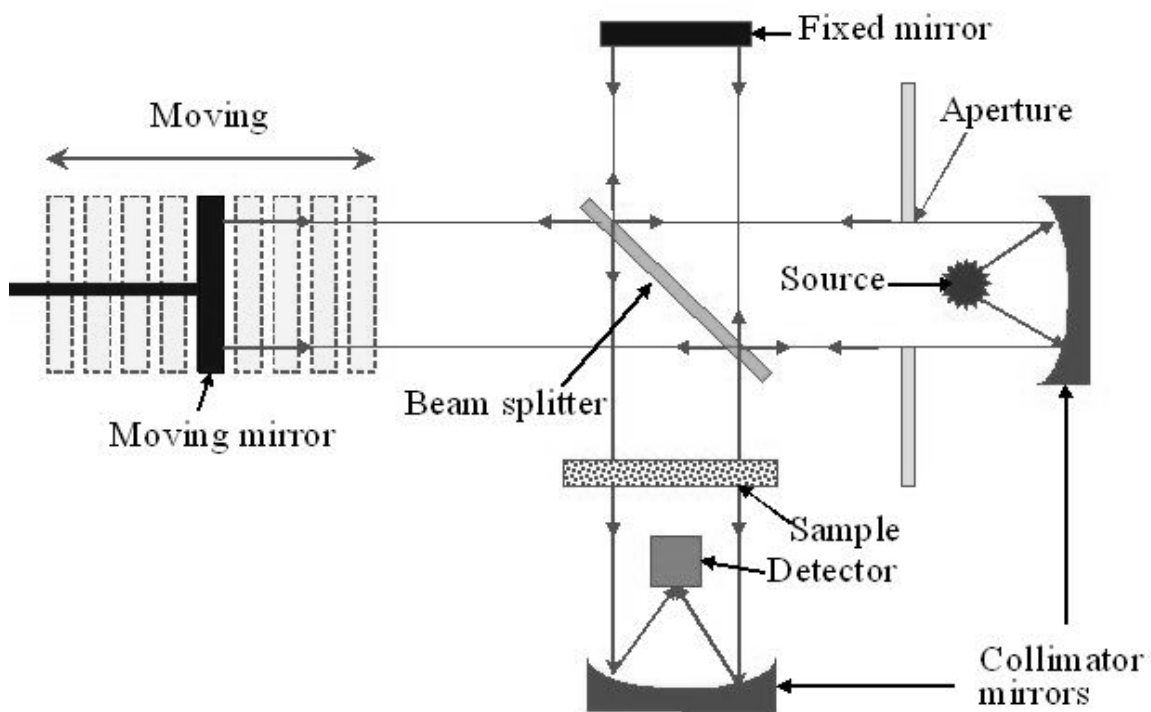
capability too. Therefore, there is no limit of the data treatment or modeling technology of the NIR network.

### **Calibration model sharing**

NIR prediction values are determined by the NIR spectrum obtained from the sample analyzed. Figure 6 shows the configuration of a dispersive type NIR system and Figure 7 shows the configuration of a FT-NIR system. They both indicate that the NIR spectrum of a sample can be affected by many factors such as the NIR light source, the slits (or aperture), the mirrors (or lens), the grating (or beam splitter), the factors within the optical path (temperature, humidity, dust etc.), the sample material (variety, size, color, dryness etc.), the sampling device (sample container, fiber optics, or fiber probe etc.), the detector, etc. In order to have identical NIR spectra, all of the previously described factors have to be identical. Without considering the sample variances, it is always true that instrument variances still exist. NIR instrument manufacturers are trying to make their instruments of the same type more identical but it is impossible to have real identical instruments. This is why the same calibration model used by different NIR systems of the same type may produce the prediction values with bias. Normally, a set of standard samples covering the calibration concentration range is used to adjust the bias and slope of a calibration model for individual NIR system. Therefore, a standard calibration model of the same application may be different in different NIR systems. Once a modification of the model is necessary for some special reason, it is necessary to modify it for all the individual NIR systems that use the same calibration model. As previously described, the calibration models for the NIR network are stored in the central



**Figure 6** The configuration of a dispersive NIR system

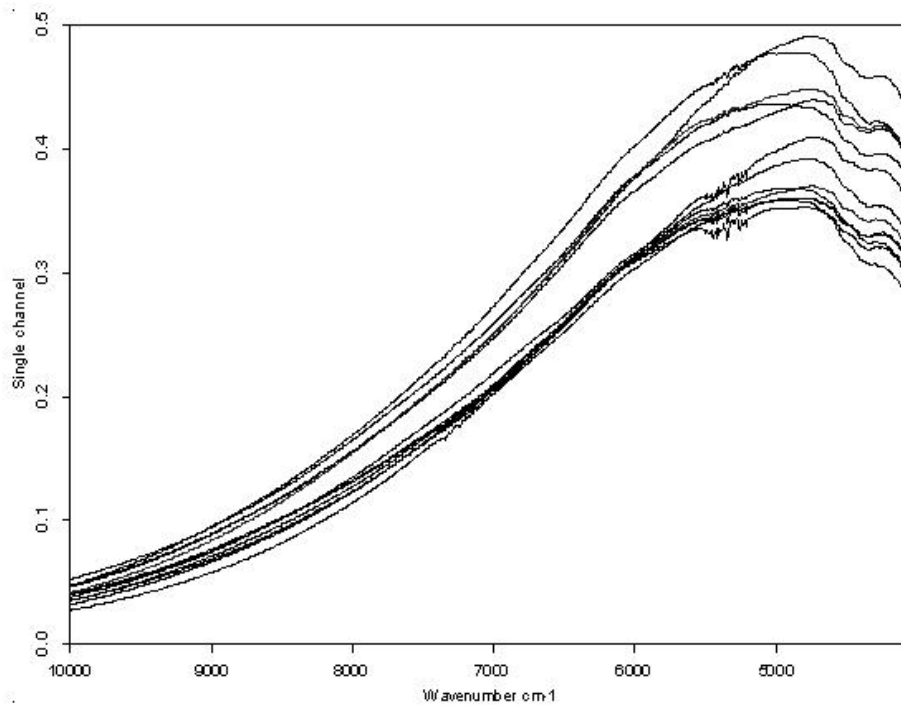


**Figure 7** The configuration of a FT-NIR system

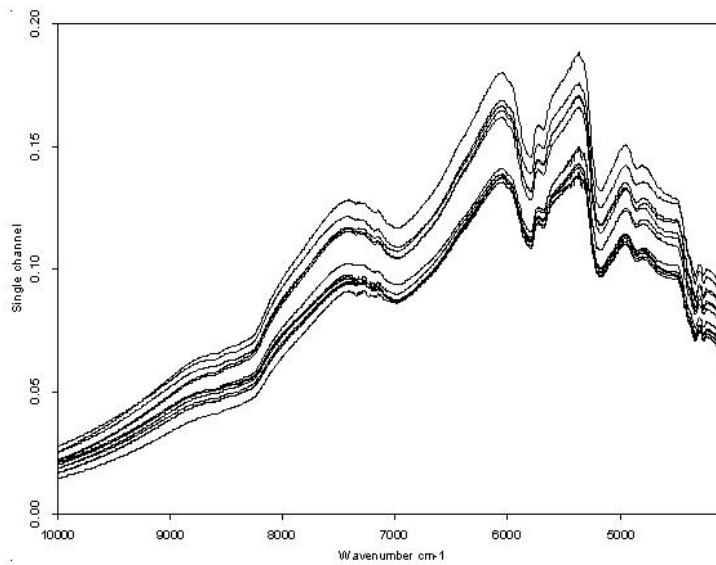
processor and shared by all of the network analyzers. It is very important that these models predict consistent results of the same sample from different analyzers. Of course, it is impossible to have identical predictions from all the individual analyzers. However, the prediction errors from any individual analyzer should be within 95% confidence level (or twice of the prediction standard error) of a calibration model.

Figure 8 shows the background spectra of 15 FT-NIR systems. These FT-NIR systems are of the same brand and the same type, and all are equipped with an integrating sphere with a lead sulfide (PbS) detector. The spectra were measured on the gold coated metal plate as the reflectance background with the spectral resolution of  $8\text{ cm}^{-1}$  and the spectral range from  $4000\text{ cm}^{-1}$  (2,500 nm or  $2.5\text{ }\mu\text{m}$ ) to  $10,000\text{ cm}^{-1}$  (1,000 nm or  $1.0\text{ }\mu\text{m}$ ). The background spectra indicates that PbS detector has the highest sensitivity around  $4800\text{ cm}^{-1}$  and the sensitivity is decreasing when the wavenumber is getting higher (or the wavelength is getting lower), but all of the background spectra are somewhat different from each other. Some of the spectra showing lower intensity than others along the whole spectral region indicate that the output of the light source is lower, the optical throughput is lower or the sensitivity of the detector is lower. Some of the background spectra exhibits obvious moisture features between  $5200\text{ cm}^{-1}$  and  $5500\text{ cm}^{-1}$  and indicates that some instruments have high humidity along the optical path.

Figure 9 shows the raw reflectance NIR spectra of the same canola seeds measured by the previously described 15 FT-NIR instruments using the same measurement parameters as the background spectra in Figure 8. Accordingly, the spectra are obviously different from each other. Therefore, if a calibration model was built by the spectra obtained only from one instrument, the prediction of the same sample measured



**Figure 8** The background spectra of 15 FT-NIR instruments



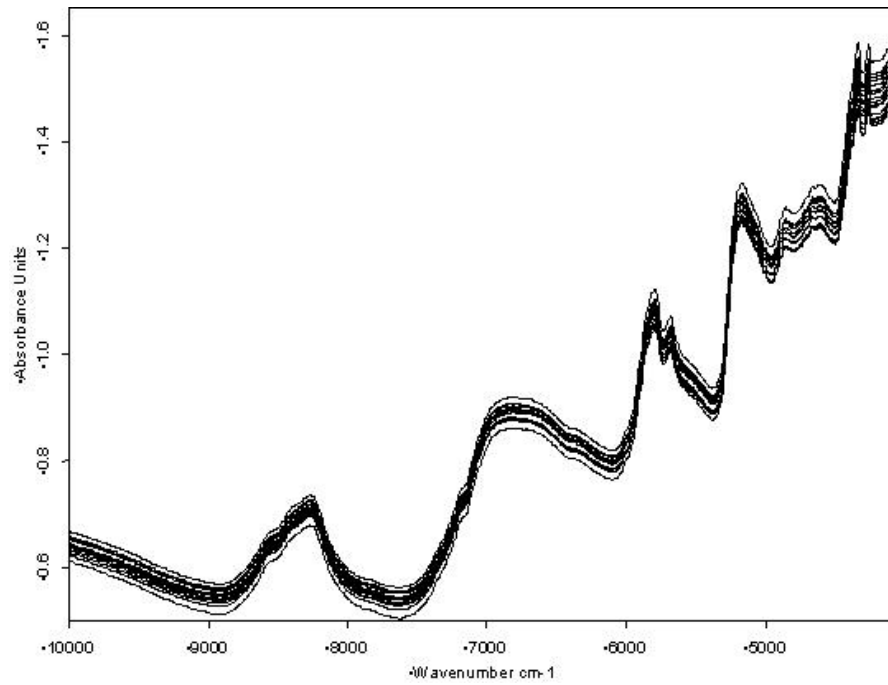
**Figure 9** Raw reflectance NIR spectra of the same canola sample obtained from 15 FT-NIR instruments

by another instrument using the same calibration model may be very different. Normally, the raw reflectance NIR spectra from a FT-NIR instrument are not directly used to build the calibration model or predict the result for quantitative analysis.

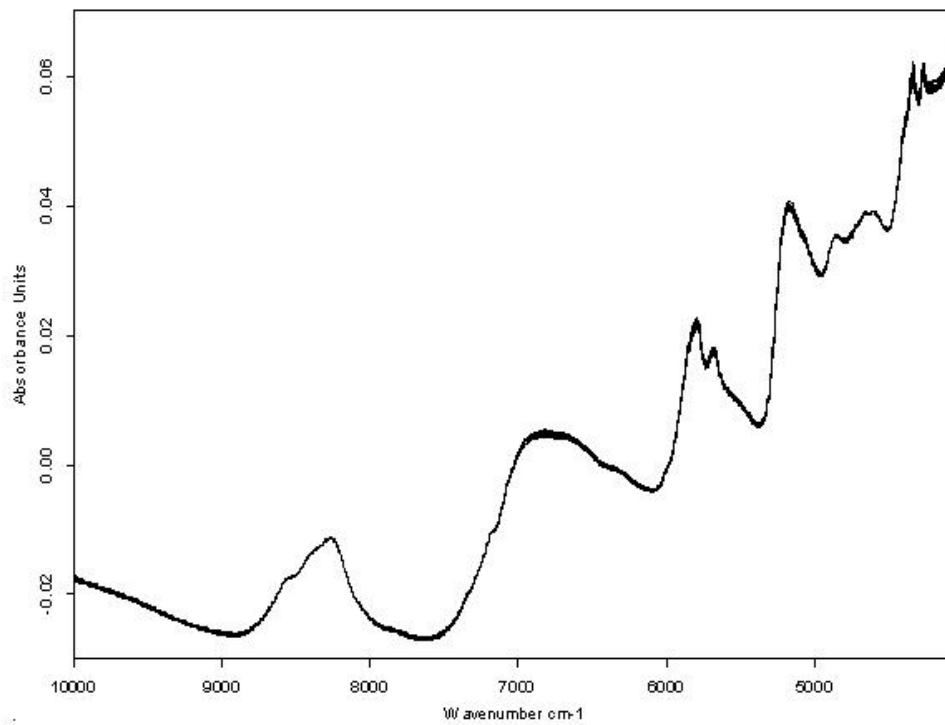
In most of cases, absorbance spectra are used to build the calibration model or predict by a calibration model. The absorbance spectrum is defined as:

$$\text{Absorbance spectrum} = -\log (\text{Raw spectrum} / \text{Background spectrum})$$

It indicates that some of the background variances from different instruments can be ratioed out and removed. Figure 10 shows the absorbance spectra converted from the raw reflectance spectra in Figure 9. It is obvious that much of the spectral variances have been removed or reduced. For example, the moisture features between  $5200 \text{ cm}^{-1}$  and  $5500 \text{ cm}^{-1}$  have been removed and the shapes of all the spectra have become similar. However, the offset and ramp between the spectra still obviously exist. Some data treatment techniques such as 1st derivative, 2<sup>nd</sup> derivative, multiplicative scatter correction (MSC), vector normalization, etc. can be used to reduce the offset and ramp.<sup>24</sup> Since vector normalization is going to be used in the following discussion, a short description of this technique is presented. During the process of vector normalization, the average y-value of the spectrum is calculated first. This average value is then subtracted from the spectrum so that the middle of the spectrum is pulled down to  $y = 0$ . The sum of the squares of all y-values is then calculated and the spectrum is divided by the square root of this sum. The vector norm of the resultant spectrum is defined to be 1. Figure 11 shows the spectra in Figure 10 after the treatment of vector normalization. It is obvious that these spectra look much more similar compared to the spectra in Figure 9 and 10.



**Figure 10** Absorbance NIR spectra converted from Figure 9

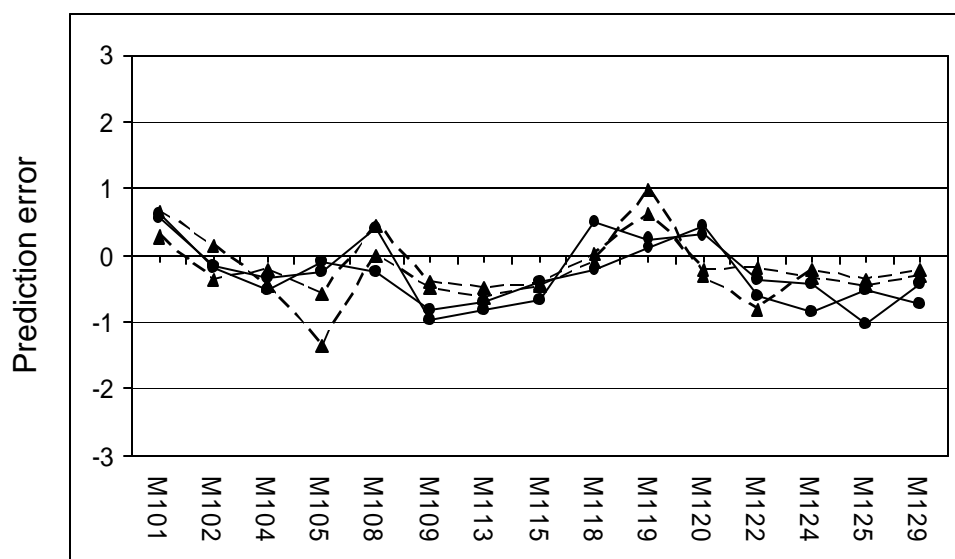


**Figure 11** Spectra in Figure 10 after vector normalization

However, they are still not identical and the prediction values of a calibration model may still be different. Actually, this situation may be good enough in some cases, such as when the spectral change obvious changes with the concentration change of a specific trait, the prediction error due to the instrument variation becomes not critical; or if the analysis is only for the screening purpose, the prediction accuracy may not be a big issue. For example, a PLS calibration model of oil content in canola seeds was built using 150 NIR spectra obtained from a FT-NIR system without any data pretreatment except the mean center calculation. The calibration result shows the  $R^2$  (determination coefficient) of 97.19% and RMSECV (Root Mean Squared Error of Cross Validation) of 0.43% with the oil content ranged from 40.9% to 52.0%. Two samples with the oil content of 47.5% and 45.2% were measured in duplicate by 14 other FT-NIR systems. The prediction errors are shown in Table 1 and illustrated in Figure 12. Most of the prediction errors, except for four predictions, are still within twice of the standard error of the calibration. This calibration model can be shared by different FT-NIR instruments if such prediction errors are acceptable. However, if the spectra are pretreated with vector normalization, a new PLS calibration model can be built with the  $R^2$  of 96.85% and RMSECV of 0.46%. In this case, the calibration appears to look not as good as the calibration model without the data treatment, but it can be used to predict the spectra obtained from different FT-NIR instruments with much smaller prediction errors. Table 2 shows the prediction errors of the same spectra in Table 1 and Figure 13 illustrates the errors with the graph. It is clear that all the prediction errors are within twice of the standard error of the calibration. The model can then be shared by different FT-NIR systems without any problem. However, it is not always true. If the spectral change is not obvious with the

**Table 1** Prediction errors of the oil content of two samples measured by 15 FT-NIR systems using the PLS model without data pretreatment

Instrument	Prediction errors			
	Sample 1(47.5% oil)		Sample 2 (45.2% oil)	
	Measure 1	Measure 2	Measure 1	Measure 2
M101	0.66	0.28	0.56	0.63
M102	0.15	-0.37	-0.16	-0.19
M104	-0.45	-0.22	-0.34	-0.53
M105	-1.35	-0.58	-0.26	-0.10
M108	0.00	0.44	0.39	-0.23
M109	-0.48	-0.41	-0.98	-0.82
M113	-0.63	-0.49	-0.84	-0.70
M115	-0.46	-0.43	-0.68	-0.39
M118	-0.09	0.03	0.50	-0.22
M119	0.99	0.60	0.25	0.11
M120	-0.31	-0.23	0.31	0.41
M122	-0.84	-0.20	-0.38	-0.61
M124	-0.21	-0.33	-0.44	-0.85
M125	-0.37	-0.47	-1.04	-0.52
M129	-0.21	-0.31	-0.44	-0.72

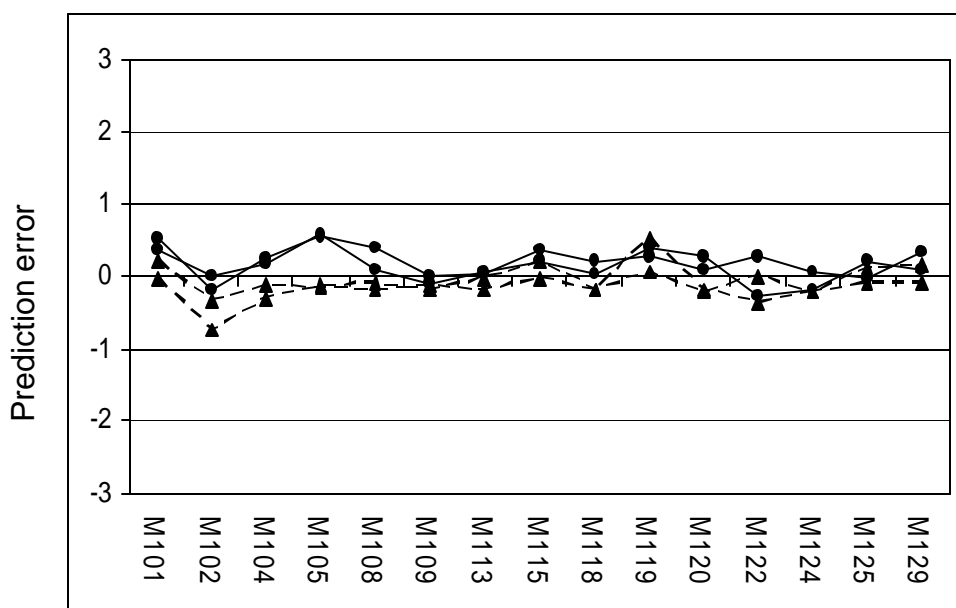


**Figure 12** Graph of the prediction errors in Table 1

concentration change of a specific trait, the prediction error due to the instrument variation becomes critical. For example a PLS calibration model of linolenic acid content in canola seeds was built using 150 NIR spectra obtained from a FT-NIR system without any data pretreatment except the mean center calculation. The calibration result shows the  $R^2$  of 96.69% and RMSECV of 0.22% with the linolenic acid content ranged from 1.88% to 7.82%. Two samples with the linolenic content of 1.89% and 5.22% were measured in duplicate by 14 other FT-NIR systems. The prediction errors are shown in Table 3 and illustrated in Figure 14. There are many predictions with the errors much more than twice of the standard error of the calibration. This calibration model should not be shared by different FT-NIR instruments. As previously described, if a PLS model were built using the spectra with the data pretreatment of vector normalization, the prediction errors of the spectra obtained from different NIR systems are possibly reducible. It is also not true in this case. Table 4 shows the prediction errors, predicted by the PLS model with vector normalization of the same spectra in Table 3. Figure 13 illustrates the errors with a graph. It is obvious that vector normalization does not help at all and even makes the analysis reproducibility worse. However, if the instrument variance were considered during the modeling process, it would be possible to build a robust calibration model that can be shared by many NIR systems of the same type. After considering the instrument factor into the modeling process, the new calibration model is able to predict the same spectra in Table 3 and 4 with much lower prediction errors. Table 5 lists the prediction errors. It is also obvious by comparing Figure 14, 15 and 16, that the latest calibration model is much more rugged than the others when they are used by different instruments.

**Table 2** Prediction errors of the oil content of two samples measured by 15 FT-NIR systems using the PLS model with vector normalization pretreatment

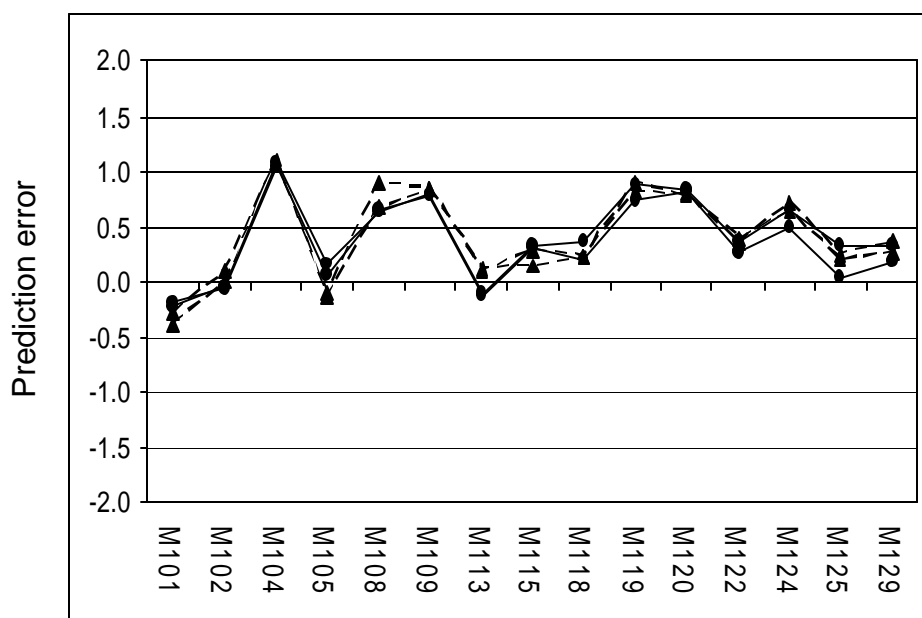
Instrument	Prediction errors			
	Sample 1(47.5% oil)		Sample 2 (45.2% oil)	
	Measure 1	Measure 2	Measure 1	Measure 2
M101	0.22	-0.03	0.53	0.36
M102	-0.34	-0.73	-0.20	0.00
M104	-0.13	-0.30	0.24	0.17
M105	-0.14	-0.12	0.56	0.59
M108	-0.18	-0.08	0.38	0.09
M109	-0.14	-0.18	-0.01	-0.09
M113	-0.19	-0.04	0.02	0.06
M115	-0.04	0.20	0.35	0.23
M118	-0.17	-0.18	0.21	0.02
M119	0.53	0.07	0.28	0.38
M120	-0.18	-0.20	0.08	0.26
M122	-0.37	0.00	0.27	-0.27
M124	-0.21	-0.23	0.06	-0.17
M125	0.14	-0.10	-0.03	0.22
M129	0.14	-0.10	0.33	0.08



**Figure 13** Graph of the prediction errors in Table 2

**Table 3** Prediction errors of linolenic acid content of two samples measured by 15 FT-NIR systems using the PLS model without data pretreatment

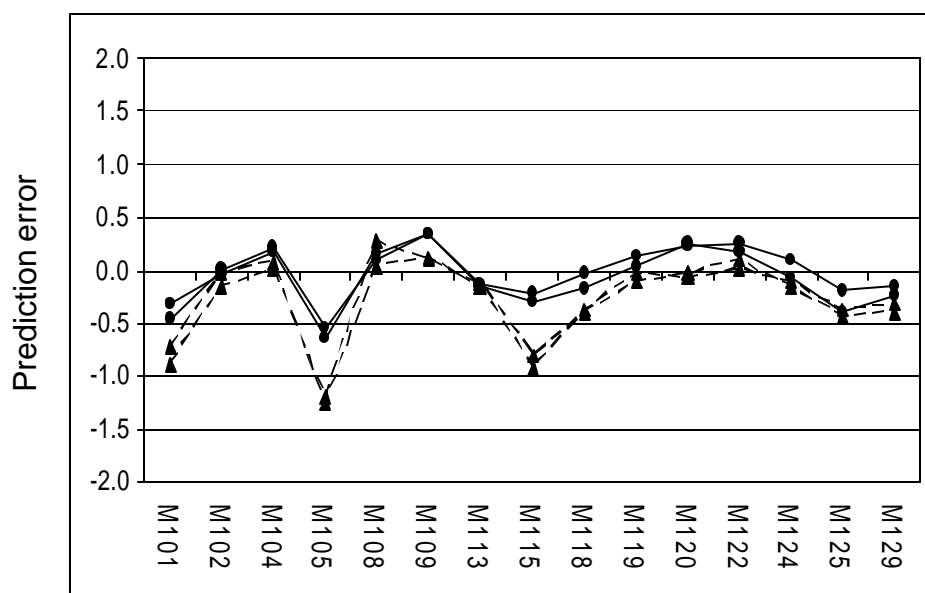
Instrument	Prediction errors			
	Sample 1(1.89% linolenic)		Sample 2 (5.22% linolenic)	
	Measure 1	Measure 2	Measure 1	Measure 2
M101	-0.39	-0.28	-0.22	-0.19
M102	0.01	0.11	-0.03	-0.07
M104	1.10	1.08	1.09	1.03
M105	-0.10	-0.14	0.17	0.07
M108	0.91	0.68	0.63	0.65
M109	0.85	0.85	0.81	0.78
M113	0.10	0.11	-0.10	-0.13
M115	0.28	0.14	0.33	0.31
M118	0.23	0.23	0.36	0.21
M119	0.81	0.89	0.87	0.73
M120	0.79	0.77	0.84	0.82
M122	0.39	0.35	0.37	0.27
M124	0.62	0.72	0.66	0.49
M125	0.19	0.25	0.32	0.05
M129	0.27	0.37	0.33	0.19



**Figure 14** Graph of the prediction errors in Table 3

**Table 4** Prediction errors of linolenic acid content of two samples measured by 15 FT-NIR systems using the PLS model with vector normalization pretreatment

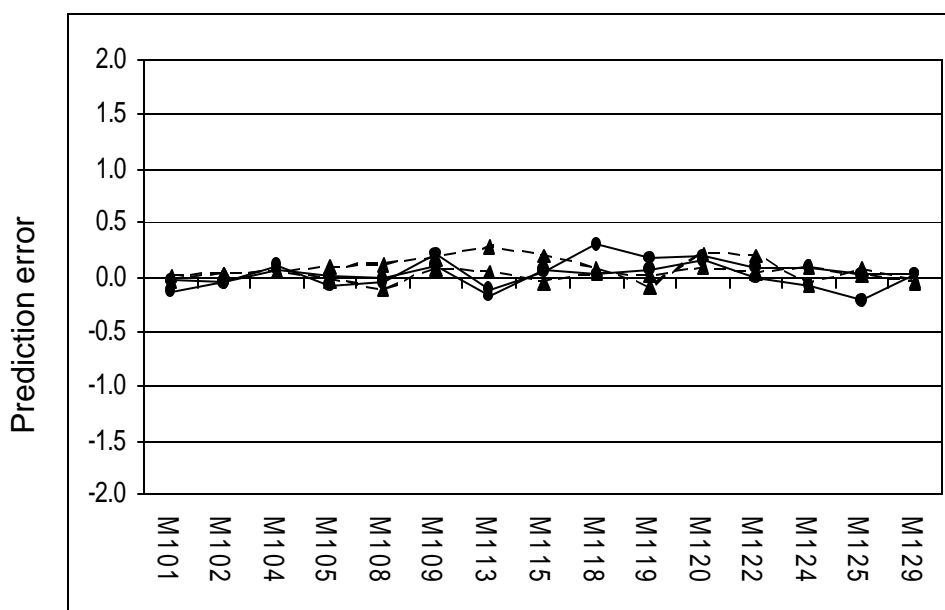
Instrument	Prediction errors			
	Sample 1(1.89% linolenic)		Sample 2 (5.22% linolenic)	
	Measure 1	Measure 2	Measure 1	Measure 2
M101	-0.90	-0.72	-0.45	-0.33
M102	-0.15	-0.02	0.00	-0.03
M104	0.00	0.07	0.21	0.17
M105	-1.21	-1.27	-0.53	-0.64
M108	0.28	0.04	0.09	0.15
M109	0.09	0.11	0.33	0.34
M113	-0.13	-0.15	-0.14	-0.15
M115	-0.80	-0.94	-0.21	-0.29
M118	-0.40	-0.38	-0.02	-0.17
M119	-0.10	-0.03	0.13	0.03
M120	-0.03	-0.07	0.24	0.25
M122	0.09	0.02	0.26	0.18
M124	-0.16	-0.11	0.09	-0.06
M125	-0.44	-0.38	-0.19	-0.41
M129	-0.39	-0.33	-0.16	-0.23



**Figure 15** Graph of the prediction errors in Table 4

**Table 5** Prediction errors of linolenic acid content of two samples measured by 15 FT-NIR systems using the PLS model with instrument factor consideration

Instrument	Prediction errors			
	Sample 1(1.89% linolenic)		Sample 2 (5.22% linolenic)	
	Measure 1	Measure 2	Measure 1	Measure 2
M101	-0.03	0.02	-0.14	-0.02
M102	0.03	0.02	-0.06	-0.06
M104	0.05	0.06	0.10	0.08
M105	0.09	-0.03	-0.06	0.00
M108	0.11	-0.11	-0.05	-0.02
M109	0.17	0.08	0.22	0.11
M113	0.28	0.06	-0.11	-0.17
M115	0.19	-0.04	0.04	0.07
M118	0.06	0.02	0.30	0.02
M119	-0.09	0.01	0.17	0.07
M120	0.21	0.09	0.19	0.16
M122	0.20	0.03	0.10	0.00
M124	-0.06	0.09	0.10	-0.07
M125	0.07	0.02	0.04	-0.21
M129	-0.03	-0.06	0.02	0.02



**Figure 16** Graph of the prediction errors in Table 5

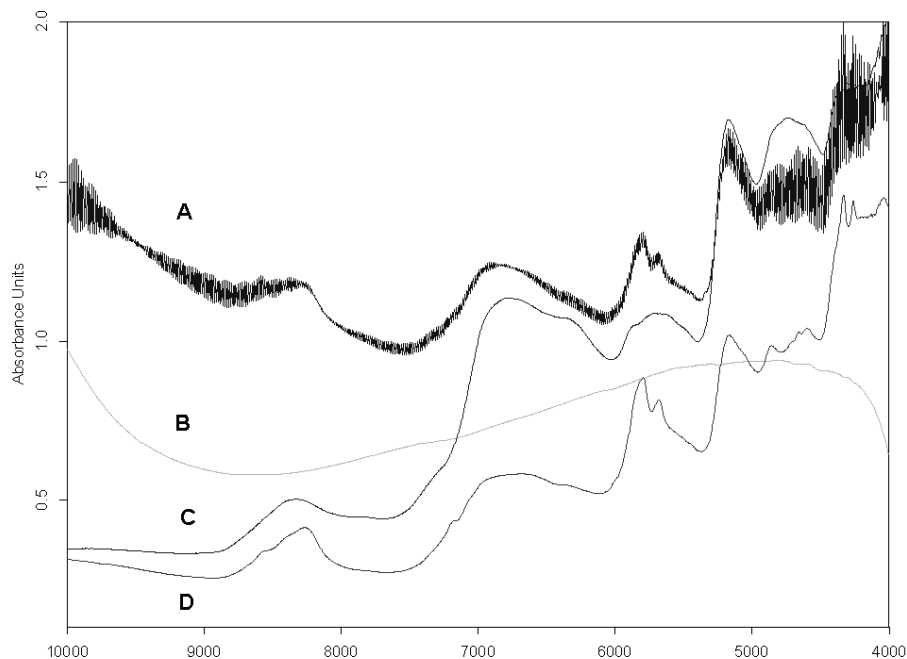
## **Remote spectral monitoring**

As described before, near infrared spectroscopy is a powerful tool for rapid and non-destructive analysis of oil seeds. Samples are scanned with NIR spectrometer and the spectra generated are predicted by the calibration model built in advance, particularly by comparing the absorbance or the shapes of the spectra in the wavelength range or ranges used for the calibration. In order to generate accurate results, it is critical that sample spectra are collected with the same conditions as the calibration standards. The prediction could be reasonably well assumed that there is no significant overall difference between the spectra of the calibration standards and the samples to be analyzed.

However, in real situations, many factors could make the spectra collected from the analyzed samples differ from the standard samples. Mainly there are three categories of sources that affect the quality of the spectra: performance of spectrometer, environmental changes, and human errors. Figure 17 includes some typical NIR spectra. Figure 17A is a noisy spectrum generated due to the unstable performance of the spectrometer. Figure 17B is a spectrum obtained by human error without sample within the optical path. Figure 17C is a spectrum of other grains but not the oil seed to be analyzed. It can be due to the selection of wrong material from the analysis menu or the position of wrong material in the sampling device. Figure 17D show the spectrum of canola seeds, which is the correct oil seed to be analyzed. Inherently these factors are in a random fashion or may not be reproduced. Even they are very different from the spectra in the calibration model, the model still predicts them with some values if there is no other conditional limit.

As described in the previous sections, to ensure the normal performance and eliminate the drifts resulting from hardware changes, current practice suggests periodic calibrations of the spectrometer using check samples. After comparing with the reference data, necessary adjustments, such as bias offset, are then applied to individual spectrometers. Although this process does help to correct the errors resulted by hardware drift, there is no precautionary action that can be taken. Therefore, a system which can monitor the quality of the spectra instantly from each analyzer is of great benefit to ensure the integrity of the analysis results.

Using the NIR network (or Internet-enabled NIR) system illustrated in Figure 4, the spectral quality of each individual analyzer can be remotely monitored. With certain spectral screening strategies, it is possible to detect some problems and respond instantly to the issues for correction. The spectral screening process consists of a computing of the Mahalanobis distance (MD) <sup>25</sup> of each spectrum and a pattern recognition process. Mahalanobis distance is an indication of the similarity between the analyzed spectrum and the spectra of the calibration standards. It is used to serve for quantifying outliers. During the PLS calculation the Mahalanobis distances of each calibration spectrum is determined. From these values the threshold of the Mahalanobis distance is derived. Spectra of unknown samples can be reliably analyzed using a calibration function if their Mahalanobis distance is within this threshold. The threshold can be set tight to sensitively identify even slight variances. However, it is not suitable for the NIR network applications since many uncertain factors can happen while many analyzers are used by not well-trained users. Therefore, the Mahalanobis distance of the NIR network models usually have to be set higher otherwise a large portion of the spectra will be considered as



**Figure 17** Typical NIR spectrum of: A) instrument malfunction, B) no seeds, C) wrong seeds, D) canola seeds

**Table 6** Mahalanobis distance in correlation to the NIR spectra

Mahalanobis	Spectral Quality
0.0 ~ 1.0	Spectra are similar to standards used to build calibration
1.0 ~ 10.0	Spectra is different from the standards, may caused by <ul style="list-style-type: none"> <li>(1) Wrong material</li> <li>(2) Severe sample contamination</li> <li>(3) spectrometer malfunction</li> </ul>
>10.0	Spectra is extremely different from the standards, may caused by <ul style="list-style-type: none"> <li>(1) significant hardware problem</li> <li>(2) Human errors</li> </ul>

outliers. In this case, the Mahalanobis distance can also be used as a general screening tool to identify the possible problems. Table 5 shows a general rule to categorize the spectra. Actually, in most of cases, the Mahalanobis distance should be less than 0.5. Slight contamination is very possible for field measurement and it can increase the Mahalanobis distance to be higher. Because many field applications are for screening purposes and very accurate results may be necessary, a Mahalanobis distance up to 1.0 may still be acceptable. When a Mahalanobis distance is over 1.0 and less than 10.0, some problems may happen such as the wrong material was selected from the menu, the wrong material was analyzed, a severely contaminated sample was analyzed, or the obtained spectrum was distorted due to an instrument malfunction. Of course, the prediction value should not be displayed in this case. In some rare cases, the Mahalanobis distance can be over 10.0, which would happen if the light source was off, the light path was blocked, or the user forgot to put the sample in, etc.

If per chance the Mahalanobis distance screening process and pattern recognition process of an Internet-enabled NIR system does not work well, the NIR experts still can download the problem spectra, identify and solve the problem remotely from anywhere of the world.

## **Conclusion**

NIR technology becomes a very important tool for oil seed analysis because it is rapid, non-destructive, and can be operated by non-experts. Even though current NIR systems have a few problems, most of the problems can be solved by using the network strategy. Internet technology is one of the greatest and most popular technologies in this

century. The Internet is reaching most of the organizations and families all over the world. A NIR network can be used for only a small area but it can also be combined with the Internet and used all over the world. An Internet-enabled NIR system can then be used by non-experts in the laboratories, plants, elevators, control rooms, cargo stations, etc. at any location in the world, but the model prediction, model development, data storage and spectral monitoring can be processed by only one or several central locations. Since the calibration models are developed and shared at the central location, there is no limit of the capability of any individual NIR analyzer. After a long time of development, the central processor will have a large number of applications. The NIR network users only have to access the central processor website and purchase the applications they need. Their NIR analyzers will be able to do the required analysis, immediately. We believe that Internet-enabled NIR is the trend of this new century for oil seed quality analysis.

## **Reference:**

1. Norris, K. H.; Hart, J.R. **Direct spectrophotometric determination of moisture content of grain and seeds**. Intern. Symp., Washington, D.C., 1963(4), 19-25 (1965).
2. Ben-Gera, I.; Norris, K. H. **Determination of moisture content in soybeans by direct spectrophotometry**. Isr. J. Agric. Res. 18:125-132 (1968).
3. Hymowitz, T.; Dudley, J. W.; Collin, F. E.; Brown, C. M. **Estimation of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance**. Crop Sci. 14:713-715 (1974).
4. Rinne, R. W.; Gibson, S., Bradley, J.; Seif, R.; Brim, C. A. **Soybean protein and oil percentages determined by infrared analysis**. U.S. Dep. Agric. Res. Publ. ARS-NC-26 (1975).
5. Takahashi, M.; Hajika, M.; Igita, K.; Sato, T. Editor(s): Davies, Anthony M. C.; Williams, P.C. **Rapid estimation of protein, oil and moisture contents in whole-grain soybean seeds by near infrared reflectance spectroscopy**. Near Infrared Spectroscopy: The Future Waves, Proceedings of the International Conference on Near Infrared Spectroscopy, 7th, Montreal, Aug. 6-11, 1995, 494-497 (1996).
6. Pazdernik, D. L.; Killam, A. S.; Orf, J. H. **Analysis of amino and fatty acid composition in soybean seed, using near infra red reflectance spectroscopy**. Agronomy Journal 89(4), 679-685 (1997).
7. Kaffka, K. J.; Norris, K. H.; Peredi, J.; Balogh, A. **Attempts to determine oil, protein, water and fiber content in sunflower seeds by the NIR technique**. Acta Alimentaria, 11(3), 253-69 (1982).

8. Robertson, J. A.; Barton, F. E. **Oil and water analysis of sunflower seed by near-infrared reflectance spectroscopy.** J. Am. Oil Chem. Soc. 61(3), 543-7(1984).
9. Sato, T.; Takahata, Y.; Noda, T.; Yanagisawa, T.; Morishita, T.; Sakai, S.  
**Nondestructive determination of fatty acid composition of husked sunflower (Helianthus annua L.) seeds by near-infrared spectroscopy.** J. Am. Oil Chem. Soc. 72(10), 1177-83 (1995).
10. Perez-Vicha, B.; Velasco, L.; Fernandez-Martinez, J. M. **Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near-infrared reflectance spectroscopy.** J. Am. Oil Chem. Soc. 75(5), 547-555 (1998).
11. Velasco, L.; Matthaus, B.; Mollers, C. **Nondestructive assessment of sinapic acid esters in Brassica species: I. analysis by near infrared reflectance spectroscopy.** Crop Science 38(6), 1645-1650 (1998).
12. Velasco, L.; Mollers, C. **Nondestructive assessment of sinapic acid esters in Brassica species: II. evaluation of germplasm and identification of phenotypes with reduced levels.** Crop Science 38(6), 1650-1654 (1998).
13. Salgo, A.; Fabian, Z.; Ungar, E.; WeinbrennerVarga, Z. Editor(s): Murray, Ian; Cowe, Ian A. **Determination of anti-nutritive factors by near infrared techniques in rapeseed.** Making Light Work: Adv. Near Infrared Spectrosc., Int. Conf. Near Infrared Spectrosc. 336-41 (1992).
14. Daun, J. K.; Clear, K. M.; Williams, P. **Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed.** J. Am. Oil Chem. Soc. 71(10), 1063-8 (1994).

15. Velasco, L.; Becker, H. C. **Estimating the fatty acid composition of the oil in intact-seed rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy.** *Euphytica* 101(2), 221-229 (1998).
16. Velasco, L.; Fernandez-Martinez, J. M.; De Haro, A. **Determination of the fatty acid composition of the oil in intact-seed mustard by near-infrared reflectance spectroscopy.** *J. Am. Oil Chem. Soc.* 74(12), 1595-1602 (1997).
17. Kohel, R. J. **Evaluation of near infrared reflectance for oil content of cottonseed.** *Journal of Cotton Science* 2(1), 23-26 (1998).
18. Bhatta, R. S. **Measurement of oil in whole flaxseed by near-infrared reflectance spectroscopy.** *J. Am. Oil Chem. Soc.* 68(1), 34-8 (1991).
19. Velasco, L.; Mollers, C.; Becker, H. C. **Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy.** *Euphytica* 106(1), 79-85 (1999).
20. Orman, B. A.; Schumann, R. A., Jr. **Nondestructive single-kernel oil determination of maize by near-infrared transmission spectroscopy.** *J. Am. Oil Chem. Soc.* 69(10), 1036-8 (1992).
21. Geladi, P.; Dabakk, E., **An overview of Chemometrics applications in near infrared spectrometry,** *J. Near Infrared Spectroscopy*, 3, 119-132 (1995).
22. Haaland, D. M.; Thomas, E. V., **Partial least-squares methods for spectral analysis. 1. Relation to other quantitative calibration methods and the extraction of qualitative information,** *Anal. Chem.*, 60, 1193-1202 (1988).
23. Næs, T; Kvaal, K; Isaksson, T.; Miller, C., **Artificial neural networks in multivariate calibrations,** *J. Near Infrared Spectroscopy*, 1, 1 (1993)

24. Geladi, P.; MacDougall, D.; Martens, H., **Linearization and scatter correction for near infrared reflectance spectra of meat**, Applied Spectroscopy, 39, 491 (1985).
25. Gemperline, P. J.; Boyer, N. R., **Classification of near-infrared spectra using wavelength distances: comparison to the Mahalanobis distance and residual variance methods**, Anal. Chem., 67, 160-166 (1995).