Analysis of Oilseeds for Protein, Oil, Fiber and Moisture by Near-Infrared Reflectance Spectroscopy

J.A. Panford[a], P.C. Williams[a] and J.M. deMan[b]

[a]Canadian Grain Commission, Grain Research Laboratory, 1404-303 Main St., Winnipeg, Manitoba, R3C 3G8, Canada, and [b]University of Guelph, Food Science Department, Guelph, Ontario, Canada

Wavelength and mathematical treatments were optimized for the determination of oil, protein, moisture and crude fiber components in the ground seeds of nine oil-bearing crops [rape, flax, sunflower, safflower, sesame, palm kernel, groundnut (peanut), soybean and cottonseed] by scanning near-infrared reflectance spectroscopy. Optimum wavelengths, selected for the estimation of various components, were influenced by the algorithm (math treatment) used and differed among crops. The second derivative math appeared to be better suited for the estimation of all constituents. Methods for sample preparation and analytical results are discussed. The accuracy was quite satisfactory for routine quality control and evaluation purposes, and precision was equal to that of standard analyses.

Near infra-red reflectance (NIR) technology is based on the absorbance of light energy of a given frequency by molecules, having a permanent dipole, which vibrate at the same frequency. The difference between the incident light and light reflected from the surface of the sample is analogous to the familiar Beer-Lambert concept of absorbance/transmittance.

Near-infrared reflectance spectroscopy (NIRS) was developed in 1964 by Norris for the measurement of moisture (1). In 1965, Norris and Hart applied NIRS to the estimation of moisture in grains and seeds (2). Subsequently the technique was introduced to the grain industry in 1973 by Rosenthal as a means of rapid analysis for oil, protein and moisture (Rosenthal, R.D., lecture to American Association of cereal Chemists annual meeting, 1973). Due to its rapidity (about 20-30 seconds per test), favorable economics, simplicity of sample preparation and absence of chemicals, it has become an extremely important adjunct to the grain and food industries. At present, NIR technology is used as an analytical method for the estimation of the composition of foods, feeds, grains, oilseeds, pharmaceuticals and in medical research (3-8,19). This study was conducted to investigate the efficacy of NIR in the analysis of oilseeds composition.

MATERIALS

Seeds. Soybean (*Glycine max* L.), safflower (*Carthamus tinctorius* L.), flax (*Linum usitatissimum* L.), sunflower (*Helianthus annuus* L.) and rapeseed (*Brassica campestris* L.) were obtained in Canada. Cottonseed (*Gossypium hirsuta* L.) was obtained from Syria; groundnut (*Arachis hypogaea* L.) from the U.S. and South Africa; sesame seed (*Sesamum indicum* L.) from Mexico, and palm kernel (*Elaeis guineensis* L.) from Indonesia.

Reagents. These were anhydrous ether stabilized with 2% ethanol; concentrated sulfuric acid, 66° Baume commercial grade; 5% w/v boric acid with bromo cresol green from British Drug Houses Canada Ltd., Toronto, Canada (BDH); titanium dioxide/cupric sulfate; 0.11423N sulfuric acid; 60% w/v sodium hydroxide solution from BDH; 1N sodium chloride solution; 0.1N sodium hydroxide solution pH 10.5, and distilled/de-ionized water.

Equipment. Krups impeller-type mill, 75 ml capacity; Christie-Norris hammer mill; vacuum oven; Goldfisch oil extraction apparatus; freeze-dryer; temperature and humidity controlled cabinet; pH meter; NIR scanning monochromator, Research Composition analyzer model 6350 from Pacific Scientific Instruments Inc.; a Northstar computer with hard disk, and a muffle furnace and Fiber-tech semi-automated fiber analyzer were used.

ANALYTICAL METHODS

Samples were cleaned by hand-picking all foreign material, e.g., sand, dirt, other seeds, etc. Approximately 25 g clean sample was ground in a Krups mill using four 15-second bursts, each followed by thorough mixing to prevent clogging and expelling of oil. Palm kernel was ground in two stages, first by a Christie-Norris 8″ mill, fitted with a screen with round holes of eight mm diameter; the resultant coarse meal was ground in the Krups mill. This procedure of two-stage grinding can be applied to any types of seed, such as the faba bean which is as hard and large as plam kernel and would require reduction to a coarse meal first, then ground in the Krups mill.

Most official methods for the determination of moisture in oilseeds (e.g., AOCS Ac 2-41, Ai 2-75) suggest the use of whole, intact seeds. However, it was found that a significant amount of moisture still remained in the seed after drying, while virtually all the moisture was removed in samples ground prior to drying. Moisture content was therefore determined by drying the ground sample at 100°C (± 2°C) for 16 hr in a vacuum oven. Percent moisture was calculated as:

$$\% \text{ Moisture in sample} = \frac{\text{Sample weight (initial)} - \text{sample weight (dry)} \times 100}{\text{Sample weight (initial)}}$$

This value was used to calculate the oil, protein and crude fiber to "as is" moisture basis.

Oil content was determined by the appropriate AOCS method (9) where applicable. However, the procedure was modified slightly. Anhydrous ether (instead of petroleum ether) was used as solvent for all oil extractions. Ground sample was vacuum-dried for 16 hr; four g were weighed (in duplicate) and extracted with solvent for 16 hr. The solvent was evaporated and the oil residue weighed to calculate the oil percentage in the sample on a moisture-free (dry) basis as:

$$\% \text{ Oil (moisture-free basis)} = \frac{\text{Weight of oil in sample}}{\text{4 g sample}} \times 100$$

and percent oil content "as is" basis is calculated as:

$$\% \text{ Oil ("as is")} = \% \text{ Oil (moisture-free)} \times \frac{100 \times x}{100}$$

where x = oven moisture content of sample as determined initially.

Total nitrogen estimation by the Kjeldahl method (10) was slightly modified as well. The procedure (16) used by the Grain Research Laboratory (GRL) of the Canadian Grain Commission was used in this study. The GRL uses the AACC modified boric acid method (10), with titanium dioxide/cupric sulfate mixture as catalyst (instead of mercury). Digestion time was increased to 50 min (from 40 min for cereals). Percent protein was calculated using a factor of 6.25, and was obtained as follows:

$$\% \text{ Protein } \frac{\text{(whole-seed and}}{\text{moisture-free)}} = \% \text{ N (nitrogen} \times 6.25$$

and percent protein "oil-free" basis was calculated as:

$$\% \text{ Protein (oil free)} = \frac{N \times 6.25 \times 100}{100 - (\%) \text{ oil content}}$$

while percent protein "as is" basis was calculated as:

$$\% \text{ Protein ("as is")} = \text{Protein (\%, moisture free)} \times \frac{100 - x}{100}$$

where x = oven moisture content of sample.

Crude fiber content was determined by the official AACC method (11). Defatted sample was boiled in concentrated sulfuric acid for 30 min and in hot sodium hydroxide for 30 min, dried, then ashed in a muffle furnace. The percentage crude fiber content was calculated as follows:

$$\% \text{ Crude fiber} = \frac{W_2 - W_1}{W_0} \times 100$$

where W_0 is the initial sample weight in g; W_1 is the weight of insoluble dry matter after boiling in acid and alkaline, and W_2 is the weight of ashed residue.

$$\text{Crude fiber "as is" basis} = \frac{\text{Crude fiber (\%)} \times 100 - x}{100}$$

where x = initial moisture of sample.

Protein extract from defatted meal was carried out using a method (14) similar to the classical protein fractionation procedure of Osborne and Mendel (12). Sample (10%) was suspended in water (pH 6.5), 1N sodium chloride (pH 7.0), 70% ethanol and 0.1N sodium hydroxide (pH 11.0) separately. The suspension was stirred for one hr at room temperature and then centrifuged. The supernatant was dialyzed against deionized water for 72 hr (with frequent change of the water) freeze-dried at 4°C for 72 hr and ground to a fine powder. Protein content of the isolates was determined by the method described above.

CALIBRATION

Thirty-five to sixty-five samples of each seed type were used as calibration sets for NIR analysis. Fresh-ground samples were scanned from 1100-2500 nm using the NIR scanning monochromator (Pacific Scientific Research Composition Analyzer Model (6250). Chemical data were entered, via a Northstar computer, for each sample scan. Wavelengths for determination of protein, oil, fiber and moisture were selected by taking measurements of the energy log l/R values followed by multiple linear regression of the reflected energy at each wavelength point against the concentration of the specific constituent. The energy signal was recorded in the form of log (l/R) (where R is the reflected energy) or optical density data, by means of a PbS detector. Log l/R data were then translated into protein, oil, etc., either by using the log l/R signal directly, or after preliminary mathematical processing (14), using the first or second derivatives of the log l/R data. Separate sets of samples with known chemical data were used for prediction to verify the reliability of the calibration constants.

RESULTS

Tables 1 to 4 summarize the accuracy of NIR analysis. To minimize discrepancies in the oil and protein content, analytical results were converted to an "as is" basis, i.e., the oil or protein in the presence of the moisture in the meal after grinding. This is the form in which the NIR instrument "sees" the sample as well. Protein content expressed as oil free "as is" basis refers to the percentage protein in the seed with oil removed and water present. Protein content expressed as "whole-seed" basis refers to the protein in the seed with oil present and moisture removed, while protein on an "as is, whole-seed" basis refers to protein in meal with oil and moisture present. Crude fiber was expressed on an "as is" basis and also a moisture-free basis.

Figure 1 illustrates moisture band differences between "as is" and vacuum-dried rapeseed samples. It was found that some changes occurred in the ground meal of all oilseeds when stored for several weeks. Figure 2 indicates band differences between freshly-ground and stored safflowerseed meals. In order to prevent the changes during storage, which were presumed to be caused by agglomeration due to high oil content, samples were scanned by the NIR monochromator immediately after grinding and the spectra recorded prior to chemical analysis. There were marked differences among the seeds with regard to the proportions of protein fractions. Figure 3 illustrates band differences between rapeseed and palm kernel protein isolates. The optimum wavelengths for the estimation of the constituents also differed among the seed types (Tables 1-4).

DISCUSSION

For calibration and analytical purposes, ground samples were tested for moisture, oil, protein and crude fiber by use of modified official approved methods.

Tables 1 through 4 indicate the accuracy of NIR analysis by comparison with chemical analysis. The

TABLE 1

Analytical Data for Moisture Content (%) of all Seed Types Using a Scanning NIR Monochromator and Second Derivative of Log 1/R Algorithm

	Regression statistics						Wavelength (λ) in nanometers				
Commodity	N[a]	Mean	SD[b]	r[c]	SEP[d]	RPD[e]	λ_1	λ_2	λ_3	λ_4	λ_1/λ_n[f]
Cottonseed	35	7.3	0.473	0.981	0.056	8.3	1740[g]	–	–	–	–
Flax	65	6.3	0.691	0.962	0.075	9.2	2344[g]	1232	–	–	–
Groundnut	60	4.4	0.620	0.981	0.066	9.3	2114	–	–	–	–
Palm kernel	50	9.3	1.650	0.790	0.970	1.7	–	–	–	–	2028[g]/1566
Rapeseed	40	8.7	1.726	0.999	0.064	27.0	1390[g]	1958[g]	–	–	–
Safflower	40	8.0	2.802	0.852	0.291	9.6	2384[g]	1358	2074	1778[g]	–
Sesame	38	7.2	1.624	0.988	0.080	20.3	–	–	–	–	2360[g]/1780[g]
Soybean	35	6.8	0.988	0.918	0.046	21.5	2254	1980[g]	2200	1370[g]	–
Sunflower	58	5.3	0.841	0.960	0.061	13.8	2070	2342[g]	1486[g]	–	–

[a]Number of samples.
[b]Standard deviation of results of reference analyses.
[c]Correlation coefficient between NIR and reference analyses.
[d]Standard error of performance of NIR analysis.
[e]Ratio of standard deviation of reference results and standard error of performance of NIR data (i.e., SDx/SEP).
[f]Quotient mathematics.
[g]Moisture and -OH bands.

TABLE 2

Analytical Data for Oil Content (%, "as is" basis) for all Seed Types Using a Scanning NIR Monochromator and Second Derivative of Log 1/R Algorithm

	Regression statistics						Wavelength (λ) in nanometers				
Commodity	N[a]	Mean	SD[b]	r[c]	SEP[d]	RPD[e]	λ_1	λ_2	λ_3	λ_4	λ_5[f]
Cottonseed	35	23.0	1.183	0.997	0.090	8.6	1766[g]	–	–	–	–
Flax	65	40.0	1.887	0.999	0.107	17.5	2312[g]	2390[g]	1400[g]	2000[g]	–
Groundnut	60	51.2	2.849	0.992	0.299	9.5	2396[g]	–	–	–	–
Palm kernel	50	47.1	1.100	0.784	0.095	1.3	–	–	–	–	1828[g]/1156[g]
Rapeseed	40	40.7	1.810	1.000	0.051	35.4	–	–	–	–	2402/2344[g]
Safflower	40	24.4	6.663	0.969	0.719	6.8	1806[g]	2170[g]	2362[g]	1230[g]	–
Sesame	38	48.7	2.139	0.997	0.100	19.8	1804[g]	–	–	–	–
Soybean	35	21.9	1.330	0.999	0.065	20.2	–	–	–	–	2304[g]/2390[g]
Sunflower	58	47.6	1.890	1.000	0.078	24.2	2398[g]	2130[g]	2372[g]	1812[g]	–

[a]Number of samples.
[b]Standard deviation of results of reference analyses.
[c]Correlation coefficient between NIR and reference analyses.
[d]Standard error of performance of NIR analysis.
[e]Ratio of standard deviation of reference results and standard error of performance of NIR data (i.e., SDx/SEP).
[f]Quotient mathematics.
[g]Oil bands.

TABLE 3

Analytical Data for Protein (%, whole seed basis) for all Seed Types Using a Scanning NIR Monochromator and Second Derivative of Log 1/R Algorithm

	Regression statistics						Wavelength (λ) in nanometers				
Commodity	N[a]	Mean	SD[b]	r[c]	SEP[d]	RPD[e]	λ_1	λ_2	λ_3	λ_4	λ_1/λ_n[f]
Cottonseed	35	23.1	1.387	0.983	0.149	8.2	–	–	–	–	1980[g]/ 2276[g]
Flax	65	24.4	2.303	0.999	0.095	24.2	–	–	–	–	2196[g]/ 2316
Groundnut	60	21.6	2.479	0.992	0.266	9.3	2194[g]	1308	–	–	–
Palm kernel	50	16.1	0.470	0.710	0.415	1.1	–	–	–	–	2224[g]/ 1896
Rapeseed	40	22.5	1.447	0.996	0.088	16.4	2192[g]	2260[g]	1896	2114	–
Safflower	40	16.2	1.710	0.771	0.349	4.9	2060[g]	1520[g]	–	–	–
Sesame	38	27.8	1.535	0.997	0.109	14.0	2408[g]	2196[g]	–	–	–
Soybean	35	41.4	1.730	0.994	0.094	18.4	2200[g]	1776	–	–	–
Sunflower	58	17.3	2.070	0.999	0.091	22.7	1976[g]	2390[g]	1756[g]	–	–

[a]Number of samples.
[b]Standard deviation of results of reference analyses.
[c]Correlation coefficient between NIR and reference analyses.
[d]Standard error of performance of NIR analysis.
[e]Ratio of standard deviation of reference results and standard error of performance of NIR data (i.e., SDx/SEP).
[f]Quotient mathematics.
[g]Protein bands.

TABLE 4

Analytical Data for Crude Fiber (%) of all Seed Types Using a Scanning NIR Monochromator and Second Derivative of Log 1/R Algorithm

	Regression statistics						Wavelength (λ) in nanometers				
Commodity	N[a]	Mean	SD[b]	r[c]	SEP[d]	RPD[e]	λ_1	λ_2	λ_3	λ_4	λ_1/λ_n[f]
Cottonseed	35	17.9	1.000	0.957	0.072	13.8	–	–	–	–	2396[g]/ 2428[g]
Flax	65	7.0	0.732	0.983	0.071	10.3	2194[g]	–	–	–	–
Groundnut	60	4.3	0.427	0.923	0.048	8.9	1398[g]	–	–	–	–
Palm kernel	50	9.1	0.424	0.573	0.353	1.2	1388[g]	2414[g]	2030[g]	–	–
Rapeseed	40	8.9	0.744	0.995	0.032	23.3	2274[g]	1536[g]	1498[g]	2444[g]	–
Safflower	40	37.9	3.302	0.842	0.458	7.2	1598[g]	2444[g]	1614[g]	1492[g]	–
Sesame	38	5.2	0.463	0.752	0.047	9.7	2192[g]	1696[g]	1494[g]	2030[g]	–
Soybean	35	5.0	0.256	0.763	0.027	9.3	–	–	–	–	2354[g]/ 2284
Sunflower	58	24.0	1.401	0.991	0.113	12.4	1756	1420[g]	1370[g]	–	–

[a]Number of samples.
[b]Standard deviation of results of reference analyses.
[c]Correlation coefficient between NIR and reference analyses.
[d]Standard error of performance of NIR analysis.
[e]Ratio of standard deviation of reference results and standard error of performance of NIR data (i.e., SDx/SEP).
[f]Quotient mathematics.
[g]Cellulose (fiber) bands.

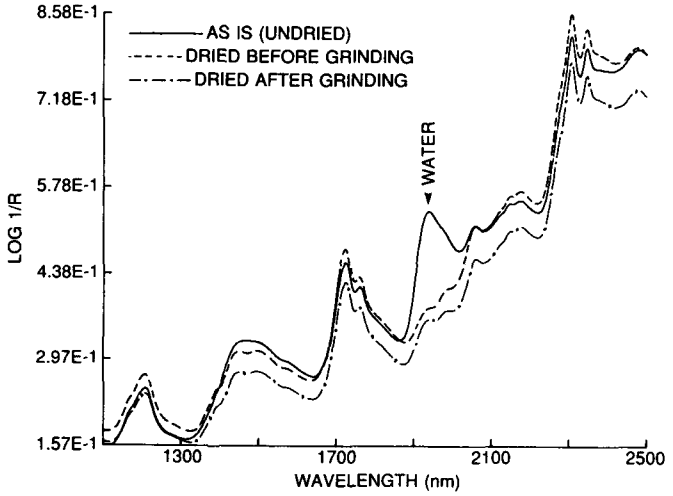

FIG. 1. Log 1/R traces of rapeseed samples of different moisture levels.

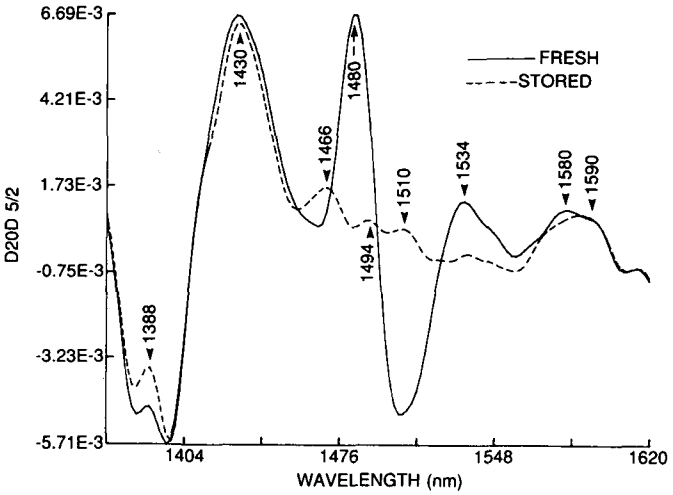

FIG. 2. Second derivative spectra of effects of storage on safflower meal.

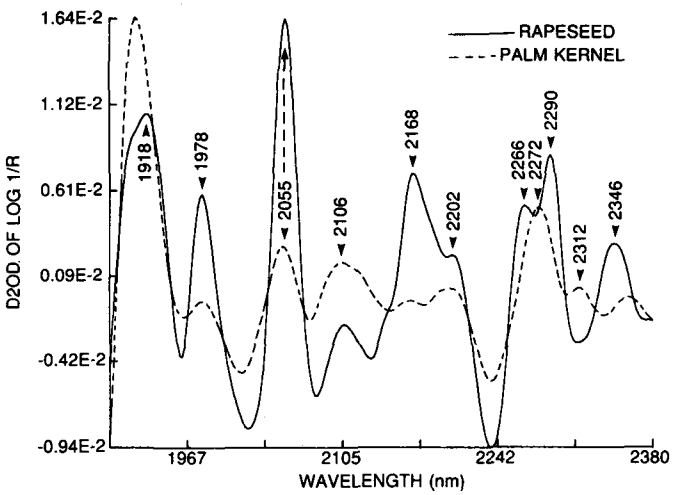

FIG. 3. Second derivative spectra of rapeseed and palm kernel proteins.

high correlation coefficients between the two techniques for most seeds demonstrate the efficacy of NIR technology. These tables show a new statistic, the RPD, as well. This is the ratio of the standard deviation of results of reference analysis and the standard error of performance of NIR data (i.e., SDx/SEP). For example, for rapeseed moisture (Table 1). SEP (standard error of performance or standard deviation of differences between NIR and reference analysis) = SD/RPD = 1.726/27.0 = 0.064% moisture, while SEP for flaxseed oil (Table 2) is 1.887/17.5 = 0.107% oil. The RPD is used to illustrate the efficiency of calibration in terms of the original standard deviation of the percentages of the constituent in the series of samples as determined by wet chemistry reference methods, when analyzed by NIR. When the value of the standard error of performance (SEP) approaches that of the standard deviation (SD), the calibration is not measuring/predicting anything. Therefore, the higher the RPD value, the more efficient or better the calibration. Ideally, RPD value of 10 or higher indicates a very good calibration, while values lower than 10 may reflect a poor calibration or too narrow a range in the constituent in the calibration samples. An RPD value of 10 indicates that the error of prediction by NIR is only one-tenth of the standard deviation of the reference result.

Tables 1-4 also illustrate the different wavelengths selected by computer for the NIR analysis of the four constituents in the nine oilseed types studied. Differences in these wavelengths were attributed to several factors, including:

(i) Interactions between constituents within a commodity;
(ii) Interactions between sample and instrument;
(iii) Variations in the particle size distribution of each seed type;
(iv) In the case of oil content, variation in fatty acid composition;
(v) Variation in degree of unsaturation in fatty acids;
(vi) Molecular associations resulting in the development of crystalline structures, dimers, hydration and other hydrogen bonding complexes;
(vii) In the case of proteins, differences in the distribution of protein fractions,
(viii) Differences in the amino acid composition of the proteins of different commodities.

Variations in composition gave wide differences in the physical morphology of the different commodities. For example, high fiber (average 37.9%) and low oil (average 24.4%) content of safflowerseed produced a meal completely different in structure from that of groundnut, which contained an average of 4.3% crude fiber and 51.2% oil. Cellulosic fiber has NIR absorption bands in the same region, (2260-2360 nm), as oil. Therefore, high fiber content may force an alternate band selection for oil in safflower, while high oil can force alternate band selection for fiber in groundnut.

Differences in the physical nature of each commodity affect the pattern by which each seed type is ground. Particle size and shape can cause small differences in the path lengths of the radiation which may, in turn, affect band selection for the constituents present. It was impossible to determine the actual particle size distribution of the ground meals on an as is" basis.

Ordinary sieving could not be used as the meals agglomerated and stuck to the sides, due to the presence of oil. The Coulter Counter instrument could not be used either, as the meal conglomerated in the water present. A solvent could not be used either, because it extracted the oil in the meal. An air separation system, which might work in the determination of the particle size distribution in oilseeds, was not available.

The wide range in fatty acid composition could also affect band selection. For example, rapeseed oil contains a high proportion of long chain, unsaturated fatty acids including erucic acid (C22:1), compared to palm kernel oil, which contains mostly short chain, saturated fatty acids including palmitic acid. Holman and Edmondson (18) found that in the region of first overtones of C-H stretching, 1.69-1.73, $-CH_3$ absorptions were overshadowed by stronger bands at 1.74 and 1.77 μm due to CH_2 absorptions as the chain length of the fatty acid increased. They also observed that absorptions at 2.15 and 2.19 μm increased with increasing *cis* unsaturation, while *trans* isomers of the acids did not exhibit strong absorption in the region nor elsewhere in the NIR range.

The effects of different water content levels are demonstrated in Figure 1. The water band at 1940 nm is virtually eliminated after drying the ground sample. However, the water band in the 1440 region remained the same before and after grinding and drying. These conditions may cause differences in band selection during the estimation of water. The development of products

TABLE 5

Wavelengths Selected for the Estimation of Oil Content of Three Oilseeds Using a Scanning NIR Monochromator and Second Derivative of Log 1/R Algorithm

Commodity/ order of terms		Wavelength (in nm)	Tentative assignment	Reference
Rapeseed	1	2402	Fatty acid (3rd overtone coupled C-O/O-H stretch, carboxylic acids)	18,19
			Fatty acid [2nd overtone $-C(CH_3)_3$; CH deformation]	18,19
	2	2344	Fatty acid (1st overtone C-H stretching, methylene groups)	18,19,20
			Conjugated triene (C-H deformation)	18
			Fatty acid (2nd overtone, C-O/O-H stretch coupled)	18,19
			Fatty acid (2nd overtone, CH stretch)	19
Soybean	1	2304	Fatty acid (1st overtone, C-H stretch, methylene groups, combination)	18,19,20
			Conjugated triene (CH deformation)	18
	2	2390	Fatty acid (1st overtone, C-H stretch, methylene groups, combination)	18,19,20
			Amino acid (1st overtone COO^- stretch or combination)	19
			Fatty acid [2nd overtone $-C(CH_3)_3$:]	19
Sunflower	1	2398	Fatty acid [2nd overtone $-C(CH_3)_3$	19
			Fatty acid (1st overtone, C-H stretch)	18,19,20
	2	2130	Fatty acid (1st overtone C-H stretch, combination)	18,19
			Long chain fatty acid (C-H, 3rd overtone)	18,19
	3	2372	Fatty acid (2nd overtone, C-H deformation CHO groups)	19
			Acid (2nd overtone C-O/O-H stretch coupled)	19
	4	1812	Fatty acid (1st overtone C-H stretch, carbonyl compounds)	19
			Acid (2nd overtone, C=O vibrations)	19
			Carboxylic acids (CO/OH stretch coupled)	18,19

TABLE 6

Effects of Derivatization on NIR Analysis for Oil and Protein in All Seed Types

				Log 1/R		1st Derivative of Log 1/R		2nd Derivative of Log 1/R	
Commodity	N[a]	Mean	SD[b]	r[c]	SEP[d]	r	SEP	r	SEP
A. Oil (% "as is" basis).									
Cottonseed	35	23.0	1.183	0.955	0.151	0.998	0.077	0.997	0.090
Flax	65	40.0	1.887	0.999	0.133	0.998	0.136	0.999	0.107
Groundnut	60	51.2	2.849	0.993	0.309	0.981	0.482	0.992	0.299
Palm kernel	50	47.1	1.100	0.663	0.916	0.713	0.729	0.784	0.095
Rapeseed	40	40.7	1.810	0.988	0.119	0.998	0.087	1.000	0.051
Safflower	40	24.4	6.663	0.942	0.724	0.987	1.220	0.969	0.719
Sesame	38	48.7	2.139	0.996	0.419	0.998	0.167	0.997	0.100
Soybean	35	21.9	1.330	0.990	0.083	0.998	0.100	0.999	0.065
Sunflower	58	47.6	1.890	0.999	0.160	0.999	0.088	1.000	0.078
B. Protein (%, whole seed basis).									
Cottonseed	35	23.1	1.387	0.999	0.074	0.989	0.159	0.983	0.149
Flax	65	24.4	2.303	0.999	0.114	0.998	0.143	0.999	0.095
Groundnut	60	21.6	2.479	0.977	0.427	0.991	0.315	0.992	0.266
Palm kernel	50	16.1	0.470	0.627	0.161	0.568	0.498	0.710	0.415
Rapeseed	40	22.5	1.447	0.995	0.089	0.996	0.590	0.996	0.088
Safflower	40	16.2	1.710	0.738	0.305	0.907	1.820	0.771	0.348
Sesame	38	27.8	1.535	0.989	0.289	0.994	0.175	0.997	0.109
Soybean	35	41.4	1.730	0.960	0.196	0.948	0.214	0.994	0.094
Sunflower	58	17.3	2.070	0.997	0.135	0.999	0.109	0.999	0.091

[a]Number of samples.
[b]Standard deviation of results of reference analyses.
[c]Correlation coefficient between NIR and reference analyses.
[d]Standard error of performance of NIR analyses.

from oil hydration and oxidation may also affect band selection. Figure 2 illustrates changes in bands of freshly ground and stored safflowerseed meals. The stored sample clearly showed a different spectrum with different band positions, suggesting the presence of constituents (absent in the fresh ground meal) as a result of chemical changes that occurred during storage.

In addition, variations in the distribution of protein fractions as well as in the amino acid composition of the seed types influence band positions.

Band assignments traditionally have been made using isolated pure compounds. They may be rather different when substances such as amino acids or fatty acids exist as complex matrices in agricultural materials. The computer tends to select bands for a given constituent at which there is a combination of higher absorbance values, with least interference from other constituents or absorbing molecular groups. While the wavelengths selected in the present study were not exactly the same for different commodities, several similar wavelengths occurred in different sequences for many of the constituents. Due to overtone combinations, etc., the differences in selected wavelengths indicate that detection of the hydrogen functional groups, O-H, C-H, N-H, etc., may occur over the entire NIR region (1100-2500 nm) and wavelength shifts are not uncommon in different commodities. The system selects the wavelength(s) at which the concentrations of the constituents can be estimated with the highest accuracy. Accuracy is defined as the highest correlation coefficient combined with the lowest standard error of performance (SEP). In general, the fewer the number of wavelenghts required to produce the highest accuracy, the better. The diversity in selected wavelength is also indicative of the complexity of the constituents found in the seed types.

The algorithm (mathematical treatment) used also affects the wavelength points selected for each constituent, as indicated in Tables 1-4. Law and Tkachuk (17) found that up to six wavelengths were required for quantitative analysis, depending on the mathematical treatment used when they worked with wheat components. It was evident in the present work that the second derivative treatment was generally better suited for the estimation of all constituents. During derivatization, absorption peaks appear as minima in the spectrum but the individual bands are more clearly defined, which simplifies assignment of wavelengths.

Assignment of wavelengths in this study was based on verification and comparisons with correlation charts and other information in the literature (Table 5). Functional group absorptions usually occur near or at the same frequencies (19,20). Minor shifts are usually attributed to factors such as temperature, the size of the rest of the molecule, other absorbing groups present in the molecule and some or all of the other factors listed earlier.

The consistently poor results for palm kernel may be attributed to several factors:

(i) Over 60% of the fatty acids in the oil were saturated short chain (C6:0 to C14:0) acids;

(ii) More than 80% of the protein was water soluble or albuminous type (Fig. 3), and
(iii) There was only a small range in the concentration of the constituents. For example, the range in moisture and crude fiber were 1% and 2.1% respectively, which accounted for the high SEE in palm kernel.

In general, the statistical evaluation of calibrations are much improved with wider ranges of constituents. This was apparent in the other seeds.

Quantitative analysis of agricultural products by NIR spectroscopy is an empirical process and governed by statistics. Evaluation of the validity of calibrations and subsequent predictions is always dependent on the accuracy of the chemical method(s) against which the NIR instrument is calibrated (Table 6), in addition to the factors involved in wavelength and mathematical treatment selection.

ACKNOWLEDGMENTS

The authors are grateful to the Canadian Grain Commission for providing funding for this work. Gratitude is also expressed to L. Kawchuk, C. Ashley and Y. Lim for their technical assistance in the crude fiber and protein analyses. This is Paper No. 608 of the Canadian grain Commission, Grain Research Laboratory.

REFERENCES

1. Norris, K.H., *Agric. Eng. 45*:370 (1964).
2. Norris, K.H., and J.R. Hart, in *Principles and Methods of Measuring Moisture in Liquids and Solids, Vol. 4.* edited by A. Wexler, Reinholt, New York, 1965, pp. 19-25.
3. Hymowitz, T., J.W. Dudley, F.I. Collins and C.M. Brown, *Crop Sci. 14*:713 (1974).
4. Williams, P.C., *Cereal Chem. 52*:561 (1975).
5. Shenk, J.S., *Forage and Grassland Progress XVIII*:1 (1977).
6. Tkachuk, R., *Cereal Foods World 26*:584 (1981).
7. Peuchant, E., C. Salles and R. Jensen, *Anal. Chem. 59*:1816 (1987).
8. Lodder, R.A., M. Selby and G.M. Hieftje, *Ibid. 59*:1921 (1987).
9. *Official Methods and Recommended Practices of the American Oil Chemists' Society, 3rd edn., Vol. 1,* edited by R.C. Walker, AOCS, Champaign, IL, 1987, Methods Aa 4-38; Ab 3-49; Ac3-44; Af 3-54; Ag 1-65, and Ai 3-75.
10. *Approved Methods of the American Association of Cereal Chemists, Vol. 2,* compiled by the Approved Methods Committee, AACC Inc., St. Paul, MN, 1976, Modified Kjeldahl-Boric acid method #46-12.
11. *Official Methods and Recommended Practices of the American Oil Chemists' Society, 3rd edn., Vol. 1,* edited by R.C. Walker, AOCS, Champaign, IL, 1987, Method Ba 6-84.
12. Osborne, T.B., and L.B. Mendel, *J. Biol. Chem. 18*:1 (1914).
13. Norris, K.H., and P.C. Williams, *Cereal Chem. 61*:158 (1984).
14. Betschart, A.A., *J. Food Sci. 40*:1010 (1975).
15. Williams, P.C., and B.N. Thompson, *Ibid. 55*:1014 (1978).
16. *Kjeldahl Protein Methods,* Grain Research Laboratory Division, Canadian Grain Commission, edited by P.C. Williams, Candian Grain Commission, Winnipeg, Canada, 1967.
17. Law, D.P., and R. Tkachuk, *Cereal Chem. 54*:256 (1977).
18. Holman, R.T., and P.R. Edmondson, *Analytical Chem. 28*:1533 (1956).
19. Murray, I., and P.C. Williams, in *Near-infrared Technology in the Agriculture and Food Industries,* edited by P. Williams and K. Norris, Amer. Assoc. Cereal Chem. Inc., St. Paul., Minnesota, 1987, pp. 17-34.
20. Coblentz, W.W., *Investigations of IR spectra,* Parts I and II, Carnegie Institute, Washington, Publ. 35, Washington, D.C., 1905.

[Received August 10, 1987; accepted May 16, 1988]