Analysis of Protein in Ground and Whole Field Peas by Near-Infrared Reflectance[1]

R. TKACHUK,[2] F. D. KUZINA,[2] and R. D. REICHERT[3]

Cereal Chem. 64(6):418-422

ABSTRACT

Protein content of ground and whole field peas (*Pisum sativum* L. cv. Trapper) was determined by near-infrared reflectance spectroscopy. Optimal wavelengths for prediction of protein were selected using a stepwise multiple linear regression program. Five mathematical treatments of log (1/*R*) data (smoothed, first derivative, second derivative, $\lambda x - \lambda y$, and $[\lambda x - \lambda y]/\lambda y$) were investigated. Calibration equations developed for each of these parameters were used to predict protein content in an independent set of pea samples. Protein content was predicted more accurately in ground than in whole field peas for all data treatments examined. For ground field peas, the calibration equation developed from smoothed data and incorporating four wavelengths (2,162, 2,126, 1,774, and 2,292 nm) gave a multiple correlation coefficient (*R*) of 0.996 and a standard error of estimate of 0.34%. Protein prediction for whole field peas was much poorer when smoothed data was analyzed ($R = 0.919$, standard error of estimate = 1.34%), but improved noticeably when first derivative data was used for calibration.

Field peas are used mainly for human consumption in both domestic and export markets. Pea flour is not highly processed, is high in fiber, and has relatively good protein quality and quantity (Hannigan 1979). Moreover, field peas are rich in lysine, and supplementing wheat flour with a less expensive pea flour improves the overall nutritional quality (Anonymous 1974, Bramsnaes and Olsen 1979).

Fortification of wheat flour with high-protein field pea concentrates for studies on breadmaking and nutritional characteristics of breads was reported by Fleming and Sosulski (1977) and Sosulski and Fleming (1979). Other research has been done on the incorporation of field pea products in goods such as yeast breads (Repetsky and Klein 1981), cookies (McWatters 1978), biscuits (McWatters 1980), and chemically leavened quick breads (Raidl and Klein 1983). For these and related studies, it is necessary to accurately analyze the protein content of this legume.

In recent years, the cultivar Trapper has accounted for approximately 75% of the acreage seeded to peas in Saskatchewan. Reichert and MacKenzie (1982) collected 198 samples of this cultivar in 1979 and found that protein content varied widely (13.3–27.1% on a dry moisture basis). This paper describes a procedure for selecting near-infrared (NIR) reflectance data that are highly correlated with protein content in ground and whole field peas (cultivar Trapper). These data were analyzed by a stepwise multiple linear regression technique for the purpose of selecting optimal wavelengths at which protein in peas could be best predicted.

MATERIALS AND METHODS

A total of 179 samples of smooth yellow field peas (*Pisum sativum* L. cv. Trapper) was obtained from Saskatchewan farmers after the harvests of 1977 through 1980. The number of samples collected in each year was 17 in 1977, nine in 1980, 100 in 1979, and 53 in 1980. Moisture content of the peas ranged from 8.1 to 11.7% (the 126 samples from the 1977–1979 crops were air-dried to a more limited moisture range of 9.6–10.3%). A 30-g portion of each sample was ground on a Udy cyclone grinder using a 1.0-mm screen. Protein content (N × 6.25, dry moisture basis) was determined in duplicate by the macro-Kjeldahl method of Williams (1973) using a TiO_2 catalyst, and ranged from 14.1 to 27.6%.

To obtain pea fractions, seeds were first dehulled with a Currier plate mill. Cotyledons were separated from the hull by air aspiration and sifting. Protein concentrate and starch were prepared by the method of Vose (1980). Cell wall material was isolated according to the method of Reichert (1981).

Reflectance data were collected with a Cary model 171 spectrophotometer controlled by a PDP 11/34 minicomputer (Tkachuk 1981). A didymium glass reference standard (National Bureau of Standards, Washington DC) was used to calibrate the spectrophotometer wavelength readings (Venable and Eckerle 1979). Whole pea samples (approximately 40 g) were scanned using a cylindrical sample holder faced with an Infrasil cover (6.5 cm diameter, 2 cm deep), whereas ground pea samples (approximately 15 g) were loaded in a similar but shallower (1 cm deep) holder. A pressed sulfur pellet was used as a reflectance standard (Tkachuk and Kuzina 1978). For each sample, an average of five readings was recorded at 2.0-nm intervals over the 600 to 2,400 nm region, for a total of 900 data points. To minimize noise spikes, individual readings were automatically rejected if they varied by more than 0.004 *A'* (apparent absorbance) when compared to the average. Spectra were also plotted on a Tektronix 4027 color graphics terminal and examined visually for noise spikes or other spurious readings.

Data Processing

Reflectance values collected in transmittance mode (i.e., whole pea data) were converted to apparent absorbance (*A'*) values, where *A'* = log (1/reflectance). After subtraction of baseline spectra, five parameters were calculated from *A'* data. These were 1) smoothed spectral data, computed using a nine-point quartic convoluting function (Savitzky and Golay 1964) using a computer program adapted for the PDP 11/34 minicomputer (Jones et al 1976); 2) first derivatives, computed using a nine-point quartic first derivative function (Jones et al 1976); 3) second derivatives, computed using a 13-point quadratic second-derivative function (Jones et al 1976); 4) differences, mathematical treatment $\lambda x - \lambda y$ of smoothed spectral data for any two given wavelengths (λx, λy); 5) quotients, mathematical treatment $(\lambda x - \lambda y)/\lambda y$ of smoothed spectral data for any two given wavelengths (λx, λy).

To obtain the calibration equation that would best correlate each of the five parameters to protein content, the following procedure was adopted. Simple correlation coefficients (*r*) with Kjeldahl protein content were computed at each 2-nm interval for smoothed, first-derivative, and second-derivative data. For treatments incorporating data from two wavelengths (differences, quotients), only those combinations of wavelengths taken at a coarser 10-nm interval (e.g., 2,300 − 2,290 nm, 2,300 − 2,280 nm, ···, 850 − 840 nm) were initially correlated to Kjeldahl protein to identify promising wavelength pairs. Different subsets of wavelength (or wavelength pairs) were selected with a stepwise multiple linear regression program (PDP-11 version of the

[1]Paper no. 598 of the Grain Research Laboratory, Canadian Grain Commission, Winnipeg, Manitoba, Canada R3C 3G8.
[2]Grain Research Laboratory.
[3]Plant Biotechnology Institute, National Research Council of Canada, Saskatoon, Saskatchewan, Canada S7N OW9

BMDP-83 Statistical Program Package, Software Development Inc., Middlebury VT). The order in which variables were entered into the regression was controlled by assigning limiting values of 4.0, 3.9, and 0.01 to the program parameters F-to-enter, F-to-remove, and tolerance, respectively. Optimal subsets of wavelengths were considered to be those subsets yielding a regression equation having a high multiple correlation coefficient (R) and low standard error of estimate (SEE). For equations having similar R and SEE values, those incorporating the fewest variables and having the lowest intercorrelations among wavelengths were deemed superior. One half of the samples (evens, $n = 89$) was then used to develop a calibration equation for the selected subset of wavelengths, while the other half (odds, $n = 90$) was used as a prediction set to test the equation.

RESULTS AND DISCUSSION

Reflectance spectra of typical high-protein (26.0%) and low-protein (14.9%) field pea samples are shown in Figure 1. Ground and whole pea spectra were similar in general shape and peak positions, indicating that both contained similar information about sample composition.

Figure 2 shows reflectance spectra of four pea isolates and one ground pea sample of high protein content. Spectra are arranged in order of decreasing protein content, with pea protein isolate having the highest percentage (90.9%) and pea starch the lowest (0.6%). The central wavelength of peaks in ground peas, pea starch, and pea protein isolates corresponded closely to their counterparts in wheat (Law and Tkachuk 1977). As with ground wheat, the spectra of ground peas are strongly influenced by the carbohydrate component, characterized by the starch peak at 2,100 nm. This peak was observed in the spectra of pea hulls and pea cell wall material as well as the spectrum of pea starch but was absent in the pea protein isolate. Peaks associated with the primary amine groups of 1,500, 1,980, 2,050, and 2,180 nm (Law and Tkachuk 1977) were present only in the pea protein isolate. These amino peaks could not be distinguished in the spectrum of the high-protein ground field pea sample and were supposedly masked by the predominant carbohydrate peaks.

It is apparent that attempting to quantitate a single component such as protein in a multicomponent sample (e.g., whole peas) by visually examining peaks in a complex reflectance spectrum would be difficult. NIR reflectance does not vary linearly with absorber concentration (Osborne 1981), which further complicates quantitation. However, these problems can be overcome by mathematically transforming the reflectance data and then selecting wavelengths by statistical regression analysis.

The stepwise selection of optimal wavelengths for the calibration and prediction of protein in ground peas is given in Table I. Values of R and SEE are reported for a maximum of six wavelengths. For differences and quotients of smoothed data, two wavelengths are entered into the equation at each step; therefore, wavelength selection for only three steps is reported. Smoothed A' data was considered the best predictor of protein in ground field peas, i.e., there was no distinct improvement in protein prediction when derivatives, differences, or quotients were used. There was no significant increase in R or decrease in SEE for either calibration or prediction when more than three wavelengths were entered into the smoothed A' equation. Of the five parameters examined, the second derivative of A' was particularly poor for protein determination. Williams et al (1985) examined ground field peas by NIR reflectance and similarly found that protein measurement using second-derivative data was inferior to that using first-derivative data.

Optimal wavelength selection for whole pea analysis is given in Table II. Despite the similarities in reflectance spectra for ground and whole field peas (Fig. 1), wavelengths selected for prediction of protein in whole peas were quite different from those selected for ground peas. Primary wavelengths for whole pea smoothed A' data are almost identical to the optimal wavelengths of 1,190 and 1,214 nm reported for the determination of protein in whole wheat kernels (Tkachuk 1981). It is difficult to assign a chemical structure to the 1,180–1,230 nm region for whole peas, although pea starch

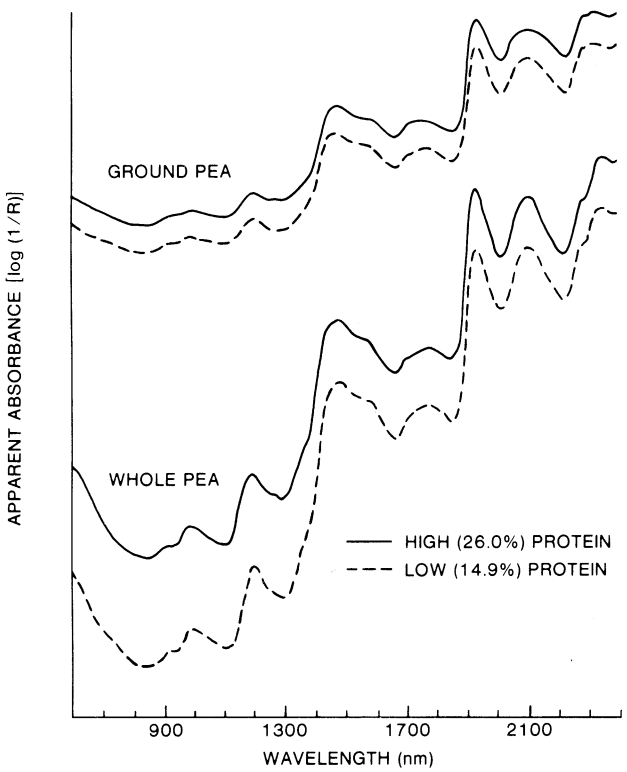

Fig. 1. Reflectance spectra of typical ground and whole Trapper field pea samples at two levels of protein content. Spectra are offset along the y-axis for clarity.

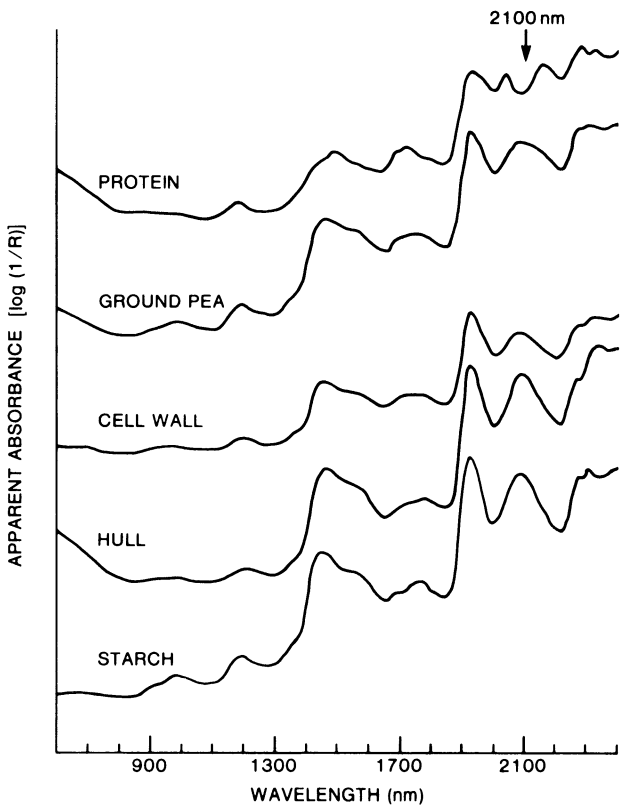

Fig. 2. Reflectance spectra of ground field pea (26.0% protein) and its major physiological components. Spectra are offset along the y-axis for clarity.

and protein isolates both absorb in this region (Fig. 2). It is interesting to note that for four pea samples varying widely in whole seed protein content (13.7–26.9%), the protein content of their hulls alone ranged only from 5.0 to 5.4%. Because NIR radiation penetrates just beneath sample surfaces, it is possible that wavelengths selected in the 1,180–1,230 nm region quantitate some component other than protein in the pea hull, which in turn is correlated with whole pea protein content. Unlike ground pea analyses, there was an obvious advantage in using derivatives, differences, or quotients when determining protein in whole peas. Among the five parameters investigated, first derivatives of A' data gave the most accurate prediction of protein in whole field peas.

For whole peas, addition of a sixth wavelength (1,266 nm) into the regression equation developed from first-derivative data (Table II) resulted in an increase in standard error (SEE) of prediction, despite the slight improvement in the R and SEE values for the calibration equation. This type of "overfitting" was discussed in detail by Hill (1979) and also by Hamid and co-workers (1978), who studied the chemical composition of tobacco by NIR analysis and noted that "a calibration equation containing too many wavelengths may not necessarily be a better prediction equation." It should be emphasized that overfitting of the calibration data must be an important consideration when deciding on the

TABLE I
Wavelength Selection for Protein in Ground Field Peas

		Wavelengths Selected (nm)		Calibration (evens, $n = 89$)		Prediction (odds, $n = 90$)	
A' Treatment	Step	λx	λy	R	SEE	R	SEE
Smoothed	1	2,162	···	0.580	2.528	0.661	2.531
	2	2,126	···	0.990	0.439	0.993	0.433
	3	1,774	···	0.993	0.365	0.995	0.344
	4	2,292[a]	···	0.994	0.348	0.996	0.334
	5	2,208[a]	···	0.994	0.337	0.996	0.309
	6	1,148[a]	···	0.995	0.324	0.996	0.308
First derivative	1	2,158	···	0.985	0.530	0.988	0.543
	2	1,664	···	0.988	0.484	0.992	0.455
	3	1,204	···	0.990	0.436	0.994	0.393
	4	1,756	···	0.991	0.430	0.995	0.363
	5	1,552	···	0.991	0.417	0.995	0.333
	6	1,008	···	0.992	0.409	0.996	0.327
Second derivative	1	1,740	···	0.934	1.113	0.945	1.081
	2	2,052	···	0.961	0.859	0.970	0.815
	3	2,314	···	0.970	0.765	0.979	0.688
	4	1,690	···	0.979	0.639	0.980	0.676
	5	2,128	···	0.982	0.603	0.984	0.615
	6	1,442	···	0.984	0.565	0.986	0.571
Differences $(\lambda x - \lambda y)$	1	2,162	2,126	0.990	0.439	0.992	0.434
	2	2,206	1,774	0.992	0.384	0.995	0.352
	3	1,814	1,688	0.993	0.367	0.996	0.333
Quotients $(\lambda x - \lambda y)/\lambda y$	1	2,162	2,126	0.988	0.479	0.992	0.484
	2	2,206	1,774	0.993	0.359	0.995	0.330
	3	1,212	1,190	0.994	0.345	0.996	0.321

[a] Entered into equation by lowering tolerance to 0.0001.

TABLE II
Wavelength Selection for Protein in Whole Field Peas

		Wavelengths Selected (nm)		Calibration (evens, $n = 89$)		Prediction (odds, $n = 90$)	
A' Treatment	Step	λx	λy	R	SEE	R	SEE
Smoothed	1	1,194	···	0.053	3.099	0.136	3.281
	2	1,208	···	0.918	1.241	0.906	1.447
	3	1,232[a]	···	0.921	1.225	0.898	1.506
	4	1,532[a]	···	0.932	1.147	0.919	1.341
	5	1,316[a]	···	0.939	1.094	0.924	1.314
	6	1,552[a]	···	0.940	1.087	0.926	1.310
First derivative	1	1,202	···	0.894	1.391	0.907	1.462
	2	1,012	···	0.923	1.203	0.930	1.300
	3	1,596	···	0.941	1.064	0.943	1.167
	4	924	···	0.953	0.961	0.947	1.136
	5	1,106	···	0.958	0.912	0.948	1.121
	6	1,266	···	0.962	0.872	0.946	1.127
Second derivative	1	1,144	···	0.926	1.174	0.890	1.586
	2	1,182	···	0.935	1.108	0.908	1.472
	3	928	···	0.951	0.968	0.927	1.311
	4	1,162	···	0.954	0.947	0.935	1.262
	5	1,258	···	0.955	0.946	0.938	1.232
	6	962	···	0.960	0.896	0.941	1.215
Differences $(\lambda x - \lambda y)$	1	1,208	1,194	0.912	1.270	0.904	1.464
	2	1,234	1,174	0.928	1.165	0.928	1.294
	3	1,552	1,532	0.935	1.116	0.936	1.232
Quotients $(\lambda - \lambda y)/\lambda y$	1	1,208	1,194	0.900	1.362	0.895	1.533
	2	1,236	1,172	0.920	1.220	0.924	1.329
	3	1,552	1,532	0.934	1.122	0.934	1.249

[a] Entered into equation by lowering tolerance to 0.0001.

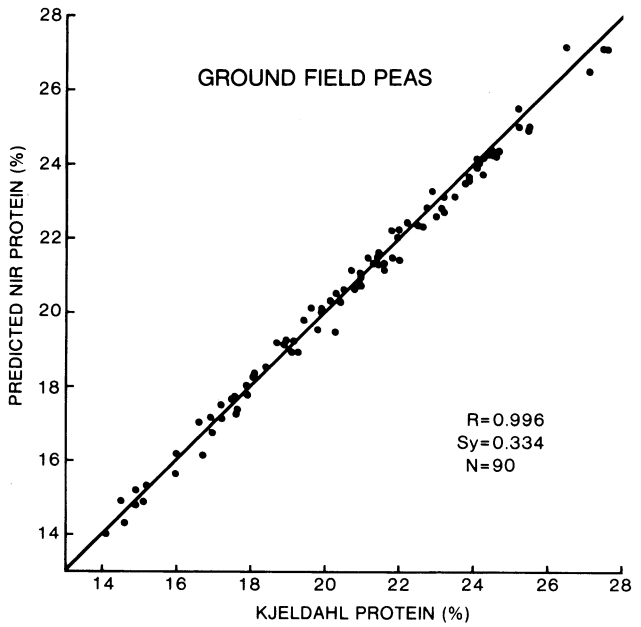

Fig. 3. Prediction plot for protein content (N × 6.25, db) of ground field peas.

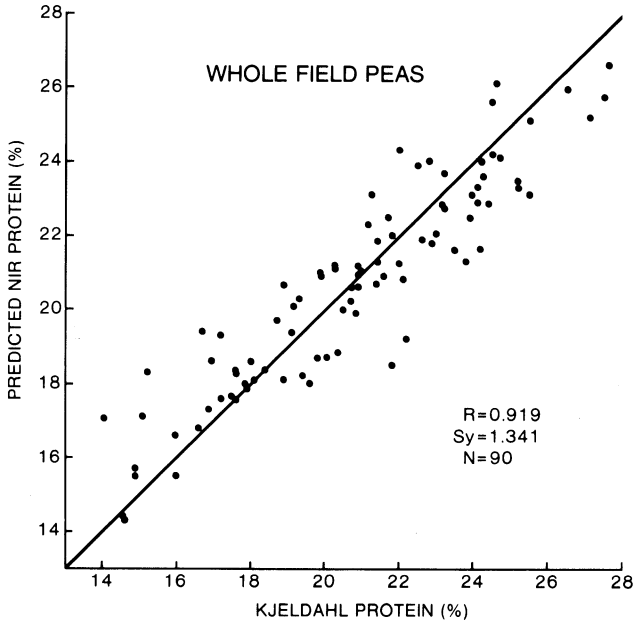

Fig. 4. Prediction plot for protein content (N × 6.25, db) of whole field peas.

TABLE III
Regression Constants of Calibration Equations for Protein Determination in Field Peas Using Smoothed and First Derivative Data

Regression Coefficients	A′ Smoothed		First Derivative	
	λx (nm)	Constant	λx (nm)	Constant
		Ground field peas		
b_0		25.75		28.22
b_1	2,162	1,104.52	2,158	−192.55
b_2	2,126	−807.91	1,664	−182.55
b_3	1,774	−150.24	1,204	192.42
b_4	2,292	−169.96	1,756	97.58
		Whole field peas		
b_0		23.78		18.17
b_1	1,194	3,000.12	1,202	257.51
b_2	1,208	−3,800.25	1,012	−349.85
b_3	1,232	706.35	1,596	93.61
b_4	1,532	86.52	924	156.97

optimum number of wavelengths for inclusion in the regression equation.

Regardless of the data treatment, prediction of protein was more accurate in ground than whole peas. The higher SEE values for whole pea prediction may be attributed in large part to sampling error arising from the large particle size of whole peas. Norris and Williams (1984) found that errors caused by particle size or surface changes were reduced by using first or second derivatives of the log $(1/R)$ signal when computing protein concentrations in hard red spring wheat. As previously mentioned, protein prediction in whole field peas was similarly improved by using such data transformations.

Plots of determined (Kjeldahl) versus predicted (reflectance) protein for the 90 odd-numbered samples of ground (Fig. 3) and whole (Fig. 4) peas also illustrate the higher SEE (i.e., the wider scatter of points) for protein determination in whole peas. Calibration equations used for prediction were derived from smoothed A' data of the 89 even-numbered samples and include only the first four wavelengths (2,162, 2,126, 1,774, and 2,292 nm for ground peas; 1,194, 1,208, 1,232, and 1,532 nm for whole peas) as selected by stepwise regression (Tables I and II). Each calibration equation is in the form $y = b_0 + b_1x_1 + \cdots + b_4x_4$, where y is predicted protein, b_0 is the intercept, b_1 to b_4 are the regression coefficients at wavelengths selected by the BMDP regression program, and x_1 to x_4 are the values for the transformed A' data at those wavelengths. Regression constants used in these equations are given in Table III.

Commercial NIR reflectance instruments such as the Dickey-john Instalab 600 are typically equipped with six narrow band pass filters having central wavelengths of 2,310 nm (oil), 2,230 nm (reference wavelength), 2,180 nm (protein), 2,100 nm (starch), 1,940 nm (moisture), and 1,680 nm (reference wavelength). When smoothed A' data of the 89 even-numbered samples were regressed against these six wavelengths, it was found that the subset of 2,180, 2,100, 1,680 and 1,940 nm gave an excellent calibration for protein in ground peas (R = 0.994, SEE = 0.356) but a very poor calibration for whole peas (R = 0.300, SEE = 3.012) (Table IV). This indicates that commercial NIR instruments could be calibrated to accurately analyze protein in ground field peas, but that standard filters supplied with such instruments would have to be replaced with filters having central wavelengths corresponding to those given in Table II to analyze protein in whole field peas. The total number of replacement filters required would depend on the level of accuracy desired.

SUMMARY

The results of this study indicate that NIR spectroscopy can be used to determine protein in ground field peas with reasonable accuracy. Although NIR measurement of protein in whole field peas would not be satisfactory for routine laboratory analyses, such a nondestructive method could still be of practical use for screening material when only limited quantities of seed are available. It should be possible to convert NIR instruments, now used commercially for analysis of oil, moisture, and protein in a wide variety of grains and oilseeds, to measure protein in whole field peas. This would require installation of narrow bandpass filters having central wavelengths corresponding to those recommended in this paper.

TABLE IV
Selection of Instalab 600 Wavelengths for Protein Determination in Field Peas Using Smoothed Data

Step	Wavelengths Selected (nm)	Calibration (evens, n = 89)		Prediction (odds, n = 90)	
	λx	R	SEE	R	SEE
		Ground field peas			
1	2,180	0.648	2.363	0.724	2.336
2	2,100	0.988	0.491	0.990	0.498
3	1,680	0.993	0.376	0.993	0.388
4	1,940	0.994	0.356	0.993	0.392
5	2,310[a]	0.994	0.354	0.993	0.390
6	2,230[a]	0.994	0.356	0.993	0.395
		Whole field peas			
1	2,180[a]	0.151	3.068	−0.064	3.357
2	2,100[a]	0.206	3.054	0.130	3.291
3	1,680[a]	0.273	3.020	0.206	3.272
4	1,940[a]	0.300	3.012	0.209	3.291
5	2,310[a]	0.301	3.030	0.203	3.318
6	2,100[a]	0.306	3.043	0.213	3.336

[a] Entered into equation by lowering tolerance to 0.0001.

ACKNOWLEDGMENTS

The authors thank V. J. Mellish and L. J. Macri for their assistance with data analysis and preparation of this manuscript.

LITERATURE CITED

ANONYMOUS. 1974. Pea flour and pea protein concentrates. PFPS Bull. 1. University of Saskatchewan, Saskatoon, Canada.

BRAMSNAES, F., and OLSEN, H. S. 1979. Development of field pea and faba bean proteins. J. Am. Oil Chem. Soc. 56:450.

FLEMING, S. E., and SOSULSKI, F. W. 1977. Breadmaking properties of four concentrated plant proteins. Cereal Chem. 54:1124.

HAMID, A., McCLURE, W. F., and WEEKS, W. W. 1978. Rapid spectrophotometric analysis of the chemical composition of tobacco. Part 2. Total alkaloids. Beitr. Tabakforsch. 9:267.

HANNIGAN, K. J. 1979. Flour from peas. Food Eng. Int. 4(2):22.

HILL, M. A. 1979. Annoted computer output for regression analysis: A supplement to writeups for the BMDP2R and BMDP9R computer programs. Tech. Rep. 48. BMDP Statistical Software. University of

JONES, R. N., BACH, T. E., FUHRER, H., KARTHA, V. A., PITHA, J., SASHADRI, K. S., VENKATARAGHAVAN, R., and YOUNG, R. P. 1976. A program for smoothing spectra and obtaining first and second derivatives. Pages 65-97 in: Computer Programs for Infrared Spectrophotometry, Vol. 1. Bull. 11. Natl. Res. Council Canada: Ottawa.

LAW, D. P., and TKACHUK, R. 1977. Near-infrared diffuse reflectance spectra of wheat and wheat components. Cereal Chem. 54:256.

McWATTERS, K. H. 1980. Replacement of milk protein with protein from cowpea and field pea flours in baking powder biscuits. Cereal Chem. 57:223.

NORRIS, K. H., and WILLIAMS, P. C. 1984. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size. Cereal Chem. 61:158.

OSBORNE, B. G. 1981. Principles and practice of near infra-red (NIR) reflectance analysis. J. Food Technol. 16:13.

RAIDL, M. A., and KLEIN, B. P. 1983. Effects of soy or field pea flour substitution on physical and sensory characteristics of chemically leavened quick breads. Cereal Chem. 60:367.

REICHERT, R. D. 1981. Quantivative isolation and estimation of cell wall material from dehulled pea (*Pisum sativum*) flours and concentrates. Cereal Chem. 58:266.

REICHERT, R. D., and MacKENZIE, S. L. 1982. Composition of peas (*Pisum sativum*) varying widely in protein content. J. Agric. Food Chem.

30:312.
REPETSKY, J. A., and KLEIN, B. P. 1981. Partial replacement of wheat flour with yellow field pea flour in white pan bread. J. Food Sci. 47:326.
SAVITZKY, A., and GOLAY, M. J. E. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627.
SOSULSKI, F. W., and FLEMING, S. E. 1979. Sensory evaulation of bread prepared from composite flours. Bakers Dig. 53(3):20.
TKACHUK, R. 1981. Protein analysis of whole wheat kernels by near infrared reflectance. Cereal Foods World 26(10):584.
TKACHUK, R., and KUZINA, F. D. 1978. Sulfur as a proposed near infrared reflectance standard. Appl. Optics 17(17):2817.
VENABLE, W. H., and ECKERLE, K. L. 1979.Standard reference materials: Didymium glass filters for calibrating the wavelength scale of spectrophotometers—SRM 2009, 2010, 2013, and 2014. Natl. Bur. Stand. U.S. Spec. Publ. 260-66. U.S. Government Printing Office: Washington, DC.
VOSE, J. R. 1980. Production and functionality of starches and protein isolates from legume seeds (field peas and horsebeans). Cereal Chem. 57:406.
WILLIAMS, P. C. 1973. Use of titanium dioxide as a catalyst for large-scale Kjeldahl determination of the total nitrogen content of cereal grains. J. Sci. Food Agric. 24:343.
WILLIAMS, P. C., MacKENZIE, S. L., and STARKEY, P. M. 1985. Determination of methionine in peas by near-infrared reflectance spectroscopy (NIRS). J. Agric. Food Chem. 33:811.

[Received January 16, 1987. June 16, 1987. Accepted June 16, 1987.]