

# Comparison of Near-Infrared Spectroscopy Calibration Methods for the Prediction of Protein, Oil, and Starch in Maize Grain

Bruce A. Orman\* and Robert A. Schumann, Jr.

Plant Breeding Division, Department of Research Specialists, Pioneer Hi-Bred International, Inc.,  
7250 Northwest 62nd Avenue, P.O. Box 1004, Johnston, Iowa 50131

This study compared the accuracy of protein, oil, and starch content prediction in maize grain, using near-infrared calibrations developed from three types of spectral data: 1100–2500-nm ground grain diffuse reflectance, 1100–2500-nm whole grain diffuse reflectance, and 680–1235-nm whole grain diffuse transmission. The best constituent predictions were by equations developed from ground grain diffuse reflectance data. Standard errors of prediction were comparable to the standard errors of the reference methods, and  $r^2$  values were all greater than 0.70. Equations developed from whole grain diffuse transmission and, to a greater extent, reflectance spectral data had higher SEP, especially for protein, and significantly lower  $r^2$  values for oil and starch prediction than corresponding ground grain equations. The performances of multiterm linear and partial least-squares regression calibration techniques were comparable for each set of spectral data.

## INTRODUCTION

Maize is the most abundantly produced crop in the United States with over 7 billion bushels being harvested in 1989 (U.S. Department of Agriculture, 1990). End uses for maize grain vary widely. Approximately 80% is consumed as animal feed for meat, poultry, and milk production. The remaining 20% is utilized in a variety of industrial processes for production of starches, oil, high fructose corn sweetener, ethanol, cereals, and alkaline cooked food products.

On average maize grain consists of 71% starch, 9% protein, and 4% oil on a dry weight basis (Hurburgh, 1989). However, genetic background and environmental effects create significant variation in constituent contents. Hurburgh (1989) reported ranges of 6.8–12.4% protein, 3.4–6.5% oil, and 67.4–75.8% starch in 1387 hybrid corn samples harvested during 1988 from throughout the mid-west.

For many applications, a direct relationship exists between the levels of chemical constituents in maize grain and relative value for feed and industrial uses (Iowa State University Cooperative Extension Service, 1989; Hurburgh, 1989). If plant breeders are to exploit existing genotypic variation in constituent levels to produce hybrids superior for various end-use applications, a rapid and accurate means of determining constituent content will be necessary. Traditional wet chemistry tests for protein, oil, and starch are too laborious and expensive for the large numbers of samples generated in a typical plant-breeding program.

Near-infrared reflectance spectroscopy (NIRS) has been used extensively to measure major constituents in wheat, maize, soybeans, and other crops (Massie and Norris, 1965; Hymowitz et al., 1974; Rinne et al., 1975; Williams, 1975; Norris et al., 1976; Shenk et al., 1976; McClure et al., 1977; Pomeranz et al., 1977a,b; Stermer et al., 1977; Watson, 1977; Williams, 1977; Finney and Norris, 1978; Saurer, 1978; Williams and Panford, 1980; Starkey, 1984; Williams et al., 1985). Hurburgh (1989) reported standard deviations of 0.4%, 0.3%, and 0.8% between reference chemistry and NIR predicted values for the determination of protein,

oil, and starch, respectively, from ground maize samples. Hurburgh (Iowa State University, Ames, IA, personal communication) and Koeltzow (FGIS, Kansas City, MO, personal communication) have reported the successful prediction of protein and oil in unground soybeans using near-infrared transmission spectroscopy (NITS).

In this study, we compared the accuracy of protein, oil, and starch content prediction in maize, using calibrations developed from three types of near-infrared spectral information; 1100–2500-nm diffuse reflectance spectra of ground grain, 1100–2500-nm diffuse reflectance spectra of whole grain, and 680–1235-nm diffuse transmission spectra of whole grain. The performances of multiterm linear regression (MLR) and partial least-squares regression (PLS) calibrations were compared for each type of spectral data.

## MATERIALS AND METHODS

**Grain Samples.** A calibration set of 156 entries was selected from a diverse group of 500 maize grain samples on the basis of their range in chemical constituents and NITS whole grain spectral diversity (Table I). The samples represented over 30 locations throughout the United States and were grown during the 1986–1988 seasons. A diversity of maize genetic backgrounds were represented by inbreds, hybrids, and F2 seed from several major seed suppliers, including normal dent, flint, white, and waxy endosperm genotypes. The grain moisture content was low, averaging 8.4%, with little variation. This is typical of samples dried and stored for breeding research use.

**Spectral Data.** All spectral data were obtained by using a Pacific Scientific (Silver Spring, MD) Model 6250 scanning monochromator infrared spectrophotometer and stored on an IBM AT using ISI (Infrasoft International, State College, PA) software. Reflectance values were collected by using a lead sulfite detector and transmission values by using a silicon detector. Ground samples were reduced to a 1-mm particle size by using a Magic Mill Model III+ and packed into 3.5 cm diameter sample cups. Whole grain samples were scanned by using a 2 cm path length cuvette capable of holding 100 g of grain. The actual spectrum representing each sample was an average of 64 spectra taken as the sample was rotated or transported through the light path. Reflectance values were collected at 2-nm intervals between 1100 and 2500 nm and transmission values at 1-nm intervals between 680 and 1235 nm. All reflectance and transmission values were normalized to  $\log 1/R$  or  $\log 1/T$  values.

The original set of 500 maize grain samples was narrowed to a calibration set of 156 by collecting whole grain diffuse

\* To whom correspondence should be addressed.

**Table I. Range of Chemical Composition of the Calibration Set**

constituent	mean	low	high
protein, %	10.69	7.94	16.82
oil, %	3.97	2.09	4.94
starch, %	69.43	62.67	73.25
moisture, %	8.42	6.11	10.30

**Table II. Math Options Used by MLR Calibration Algorithms**

spectral type	der <sup>a</sup>	gap <sup>b</sup>	boxcar <sup>c</sup>	spectral type	der <sup>a</sup>	gap <sup>b</sup>	boxcar <sup>c</sup>
transmission	1	5	5	reflectance	1	5	5
	1	10	5		1	10	5
	1	15	5		1	15	5
	2	10	5		2	10	10
	2	15	10		2	20	10
					2	30	10

<sup>a</sup> Derivative. <sup>b</sup> Derivative gap (Hruschka, 1987). <sup>c</sup> Boxcar smooth (Hruschka, 1987).

transmission spectra on all samples and eliminating those samples with similar spectra by use of the ISI subset program. Critical squared correlation match limits of 0.9500 and 0.9974 were used as the criterion for subsetting the population.

**Reference Chemistry Procedures.** All samples in the calibration set were tested for moisture, protein, oil, and starch by approved reference procedures: moisture, oven drying at 130 °C for 24 h (FGIS, 1986); protein, Kjeldahl (CRA, 1980); oil, petroleum ether extraction (AACC, 1982); and starch, enzymatic hydrolysis (CRA, 1986).

**Construction of Calibration Equations.** Calibration equations to predict protein, oil, and starch on a dry weight basis from each of the three sets of infrared spectral data were developed by two procedures. First, the ISI unical software program was utilized to construct multiterm linear regression (MLR) equations for each constituent. Specific wavelengths and math treatments (Table II) were determined stepwise by the software algorithms to construct calibration equations that best predicted constituent values determined by reference chemistry procedures.

The second calibration software program used, partial least-squares (PLS), was also developed by ISI. The PLS algorithms enable information from all spectral data points to be incorporated into the calibration model. Zero-, first-, second-, and third-derivative math treatments were explored for each PLS calibration. Since every eighth wavelength data point was used, the derivative gap and boxcar smoothing values (Hruschka, 1977) were each held constant at five over all PLS calibrations.

Both calibration algorithms selected every third sample and eliminated those from the actual calibration set. These samples served as an independent validation set.

## RESULTS

The performance of constituent prediction equations was assessed by comparing the standard errors of prediction (SEP) and coefficients of determination ( $r^2$ ) of near-infrared predicted values with those determined by reference chemistry procedures for the validation set (Table III). Both the MLR and PLS algorithms produced equations with similar statistical performance for all constituents/spectral combinations, the only exception being for oil values predicted from whole grain diffuse reflectance spectra. The PLS-derived equation yielded an  $r^2$  of 0.25, while the MLR equation had an  $r^2$  of only 0.15.

For all constituents the best near-infrared prediction results were obtained from equations developed by using ground grain diffuse reflectance spectral data. SEP were comparable to the reference chemistry procedures. The SEP for equations developed from whole grain diffuse reflectance and transmission data were greater than those obtained from ground grain diffuse reflectance spectra (Table III).

Protein values predicted by equations developed from all three types of spectral data correlated well with reference values, having  $r^2$  values ranging from 0.83 to 0.98 (Table III). For starch, correlation with reference values ranged from an  $r^2$  of 0.58 to 0.78 which, while not as good as protein, were considered acceptable. Oil content predicted by equations developed from ground grain diffuse reflectance spectral data correlated well with reference values, having  $r^2$  values of 0.76 and 0.74 for the PLS and MLR calibrations, respectively. However,  $r^2$  values of only 0.22 for PLS and 0.41 for MLR were obtained by using equations developed from whole grain diffuse transmission spectra. Whole grain diffuse reflectance equations produced poor correlations with an  $r^2$  of only 0.25 for PLS and 0.15 for MLR, relative to reference values.

Tables IV–VI contain math treatments selected for equations used to predict protein, oil, and starch, respectively, from the three types of spectral information. Second-derivative mathematics gave the best results for MLR calibrations developed from ground grain diffuse reflectance and whole grain diffuse transmission spectral data, except for starch, in which case first-derivative math was superior. The best equations from whole grain diffuse reflectance spectral data were obtained by using first-derivative math for all constituents. Most PLS-derived equations performed best when first-derivative math was utilized. For the two exceptions, protein and oil prediction from whole grain diffuse reflectance spectra, third-derivative math was superior. However, in both cases first derivatives gave very similar performance relative to third derivatives. For protein, the SEP and  $r^2$  relative to reference values were, respectively, 0.57 and 0.81 for the equation using first derivatives and 0.54 and 0.83 for the equation utilizing third-derivative math. Similarly, for oil, third-derivative math yielded SEP and  $r^2$  values of 0.32 and 0.25, respectively, relative to 0.32 and 0.20 for calibrations based on first-derivative mathematics.

A variety of derivative gap and boxcar smoothing values (Hruschka, 1977) were chosen by the MLR algorithm. However, trials at varying these values for MLR and PLS equations had little overall effect on equation performance and appear secondary in importance relative to the effect of wavelength and derivative mathematics selection.

## DISCUSSION

Use of data from all wavelengths throughout the sample spectra by PLS algorithms did not improve calibration performance. MLR and PLS calibration algorithms produced equations for each constituent/spectral type with similar performance, as measured by the standard errors of prediction (SEP) and coefficients of determination ( $r^2$ ) between near-infrared and reference values for the validation set (Table III).

Protein, oil, and starch predictions made from ground grain diffuse reflectance spectral data correlated best with reference values. For all constituents, the SEP were similar to those for reference procedures, and  $r^2$  values were all above 0.70. These results are similar to those obtained by Hurburgh (1989) for NIRS analysis of ground maize grain.

Calibrations developed from whole grain diffuse transmission spectral data, while not as accurate as those made from ground grain diffuse reflectance spectral data, were considered acceptable. The increases in SEP and decreased  $r^2$  values (Table III) were more than offset by the increased speed with which samples could be tested if grinding was not necessary. It is estimated that the number of samples which could be processed per day could be increased 3–4 times by eliminating grinding. For most

**Table III. Statistical Comparison of Near-Infrared Predicted Protein, Oil, and Starch Values Relative to Reference Chemistry Procedures for the Validation Set**

constituent	spectral type	SE lab <sup>a</sup>	SEP <sup>b</sup>		r <sup>2</sup>	
			MLR <sup>c</sup>	PLS <sup>d</sup>	MLR <sup>c</sup>	PLS <sup>d</sup>
protein	whole grain, transmission	0.40	0.38	0.41	0.90	0.92
	whole grain, reflectance		0.52	0.54	0.90	0.83
	ground grain, reflectance		0.28	0.26	0.98	0.97
oil	whole grain, transmission	0.33	0.26	0.32	0.41	0.22
	whole grain, reflectance		0.32	0.32	0.15	0.25
	ground grain, reflectance		0.21	0.18	0.74	0.76
starch	whole grain, transmission	1.21	1.18	1.27	0.58	0.59
	whole grain reflectance		1.14	1.21	0.69	0.62
	ground grain, reflectance		1.13	1.02	0.78	0.73

<sup>a</sup> Standard error of laboratory reference method. <sup>b</sup> Standard error of prediction. <sup>c</sup> Multiterm linear regression. <sup>d</sup> Partial least-squares regression.

**Table IV. Wavelengths and Math Treatments for Best Near-Infrared Protein Prediction Equations**

spectral type	MLR			PLS		
	der <sup>a</sup>	gap	box-car	der <sup>a</sup>	gap	box-car
whole grain, diffuse transmission	2	15	10	1	5	5
whole grain, diffuse reflectance	1	15	5	3	5	5
ground grain, diffuse reflectance	2	10	10	1	5	5

<sup>a</sup> Derivative.

**Table V. Wavelengths and Math Treatments for Best Near-Infrared Oil Prediction Equations**

spectral type	MLR			PLS		
	der <sup>a</sup>	gap	box-car	der <sup>a</sup>	gap	box-car
whole grain, diffuse transmission	2	15	10	1	5	5
whole grain, diffuse reflectance	1	15	5	3	5	5
ground grain, diffuse reflectance	2	30	10	1	5	5

<sup>a</sup> Derivative.

**Table VI. Wavelengths and Math Treatments for Best Near-Infrared Starch Prediction Equations**

spectral type	MLR			PLS		
	der <sup>a</sup>	gap	box-car	der <sup>a</sup>	gap	box-car
whole grain, diffuse transmission	1	15	5	1	5	5
whole grain, diffuse reflectance	1	15	5	1	5	5
ground grain, diffuse reflectance	1	15	5	1	5	5

<sup>a</sup> Derivative.

breeding programs, testing of early generation material for constituent content would be used to select only the top 10–20% of a segregating population. Therefore, some accuracy can be sacrificed for greater speed and decreased cost of analysis. With the use of winter nurseries, turnaround time between harvest of one crop and planting of the next generation is often less than 1 month. In such situations rapid analysis is essential.

Calibrations developed from whole grain diffuse reflectance spectral data performed worse than either ground grain diffuse reflectance or whole grain diffuse transmission derived calibrations. In most cases, the SEP were higher and r<sup>2</sup> values lower, especially for oil prediction. For whole seed analysis, equations developed from transmission data are superior to those derived from reflectance spectral data.

It is apparent from this study and those conducted by others (Massie et al., 1965; Hymowitz et al., 1974; Finney and Norris, 1978) that NIRS and NITS have useful applications for the rapid analysis of protein, oil, and starch

content of maize grain. While constituent predictions made by equations developed from ground grain diffuse reflectance spectral data have the greatest accuracy relative to reference values, greater speed of analysis can be achieved by use of calibrations derived from whole grain diffuse transmission spectral data, with only a minor sacrifice in accuracy.

#### ACKNOWLEDGMENT

The technical assistance of D. Phillips and B. Hobart in obtaining the spectral data is appreciated. Useful suggestions on the preliminary drafts of the manuscript were made by Drs. T. Brumback, A. Cavaliere, and O. Smith. Seed samples were supplied by D. Nepl, and the manuscript was prepared by N. Wiig.

#### LITERATURE CITED

- AACC. Crude fat in grain and stock feeds. *Approved Methods of the American Association of Cereal Chemists*; AACC: St. Paul, 1982; Method 30-20.
- CRA. Protein. In *Standard Analytical Methods of the Member Companies*; Corn Refiners Association: Washington, DC, 1980; Method A-18.
- CRA. Starch. In *Standard Analytical Methods of the Member Companies*; Corn Refiners Association: Washington, DC, 1986; Method A-20.
- FGIS. Air-Oven Methods. In *Moisture Handbook*; Federal Grain Inspection Service, U.S. Department of Agriculture: Washington, DC, 1986; Chapter 4.
- Finney, E. E.; Norris, K. H. Determination of Moisture in Corn Kernels by Near-Infrared Transmittance Measurements. *Trans. ASAE* 1978, 21, 581–584.
- Hruschka, W. R. Data analysis: Wavelength selection methods. In *Near-Infrared Technology in the Agricultural and Food Industries*; Williams, P. C., Norris, K. H., Eds.; American Association of Cereal Chemists: St. Paul, 1987; Chapter 3.
- Hurburgh, C. R. The Value of Quality to New and Existing Corn Uses. Agricultural Engineering Staff Papers Series FPR 89-2; ASAE/CSAE Meeting Presentation June 25–28, 1989.
- Hymowitz, T.; Dudley, J. W.; Collins, F. I.; Brown, C. M. Estimation of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance. *Crop Sci.* 1974, 14, 713–715.
- Iowa State University Cooperative Extension Service. *Nutrient Content and Feeding Value of Iowa Corn*; Extension Publication AS 598; Iowa State University: Ames, 1989.
- Massie, D. R.; Norris, K. H. Spectral reflectance and transmittance properties of grain in the visible and near infrared. *Trans. ASAE* 1965, 8, 598–600.
- McClure, W. F.; Norris, K. H.; Weeks, W. W. Rapid spectrophotometric analysis of the chemical composition of tobacco. Part I. Total reducing sugars. *Beitr. Tabakforsch. Int.* 1977, 9, 13–18.
- Norris, K. H.; Barnes, R. F.; Moore, J. E.; Shenk, K. S. Predicting forage quality by infrared reflectance spectroscopy. *J. Anim. Sci.* 1976, 43, 889–897.

- Pomeranz, Y.; David, G. D.; Stoops, J. L.; Hubbard, J. D. Estimation of protein content and quality in grain sorghum. *Cereal Foods World* 1977a, 22, 472 (Abstract).
- Pomeranz, Y.; Moore, R. B.; Lai, F. S. Reliability of five methods for protein determination in barley and malt. *Am. Soc. Brew. Chem. J.* 1977b, 35, 86-93.
- Rinne, R. W.; Gibbons, S.; Bradley, J.; Seif, R.; Brim, C. A. Soybean protein and oil percentages determined by infrared analysis. Agricultural Research Service No. 26; USDA: Washington, DC, 1975; pp 1-3.
- Saurer, W. Use of infra-red reflectance measurement for the determination of protein and moisture content and kernel hardness in wheat. *Getreide, Mehl Brot.* 1978, 32, 272-276 (in German).
- Shenk, J. S.; Mason, W. N.; Risius, M. L.; Norris, K. H.; Barnes, R. F. Application of infrared reflectance analysis to feedstuff evaluation. In *Proceedings, 1st International Symposium, Feed Composition, Animal Nutritional Requirements, and Computerization of Diets*; Fonnesbeck, P. V., Harris, L. E., Kearl, L. C., Eds.; Utah Agricultural Experiment Station: Logan, 1976; pp 242-248.
- Starkey, P. M. Determination of amino acids in wheat and barley by near-infrared reflectance spectroscopy. *J. Food Sci.* 1984, 49, 17-20.
- Stermer, R. A.; Pomeranz, Y.; McGinty, R. J. Infrared reflectance spectroscopy for estimation of moisture of whole grain. *Cereal Chem.* 1977a, 54, 345-351.
- Stermer, R. A.; Watson, C. A.; Dikeman, E. Infrared spectra of milled rice. *Trans. ASAE* 1977b, 76-3030, 347-549, 557.
- U.S. Department of Agriculture. *Crop Production, 1989 Summary*; Annual Crop Summary CrPr 2-1 (90); USDA: Washington, DC, 1990.
- Watson, C. A. Near infrared reflectance spectrophotometric analysis of agricultural products. *Anal. Chem.* 1977, 49, 835a-840a.
- Williams, P. C. Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. *Cereal Chem.* 1975, 52, 561-576.
- Williams, P. C. Technical aspects of instituting a protein grading system at a grain elevator, using near infrared reflection spectroscopy (NIRS). In *Advances in Automated Analysis*; Technicon International Congress, Industrial Symposia; Mediad: Tarrytown, NY, 1976; Vol. 2, pp 118-121.
- Williams, P. C.; Panford, J. A. The development and use of universal calibrations for the measurement of protein and moisture in HRS wheat by near-infrared reflectance spectroscopy (NIR). *Cereal Foods World* 1980, 25, 530 (Abstract).
- Williams, P. C.; Norris, K. H.; Sobering, D. C. Determination of protein and moisture in wheat and barley by near-infrared transmission. *J. Agric. Food Chem.* 1985, 33, 239-244.

Received for review August 2, 1990. Accepted January 2, 1991.

Registry No. Starch, 9005-25-8.