

Prediction of Wet-Milling Starch Yield from Corn by Near-Infrared Spectroscopy¹

R. L. WEHLING,^{2,3} D. S. JACKSON,² D. G. HOOPER,² and A. R. GHAEDIAN²

ABSTRACT

Cereal Chem. 70(6):720-723

Near-infrared reflectance (NIR) spectroscopy was used to predict the wet-milling starch yield obtainable from yellow dent corn. Spectral differences between samples with high and low starch yields were observed in two wavelength regions, 2,265–2,325 and 1,050–1,125 nm. The higher-wavelength region corresponds to known carbohydrate absorption bands; however, the origin of the shorter-wavelength differences is less clear. Different combinations of sample handling and spectral treatments were also evaluated, and the most successful instrument calibration used a

multiterm linear equation of second-derivative reflectance terms obtained from whole-kernel corn. Multiple correlation coefficients for the various calibrations ranged from 0.8 to 0.9, with reproducibility of the laboratory wet-milling procedure being a limiting factor. Application of the optimized calibration to a validation set containing samples from two crop years gave a bias-corrected standard error of prediction equal to 1.41%, a result equal to or better than the standard error of the reference method.

The amount of corn processed annually by the wet-milling industry continues to increase. As utilization of corn for starch, sweeteners, and other applications grows, it is essential that corn varieties with good wet-milling properties be selected. Also, sound postharvest practices with respect to transporting, artificially drying, and storing corn must be used to preserve milling characteristics. Development of a simple and rapid method for predicting starch yield and other wet-milling properties of corn would be beneficial both for selecting cultivars with good properties in a breeding program and for identifying corn shipments in commerce that have desirable industrial processing characteristics.

Since laboratory wet-milling is time-consuming and labor-intensive, several attempts have been made to correlate easily measured chemical or physical properties of corn with the starch yield obtained from wet milling. Lasseran (1973) reported a relationship between the amount of ethanol-soluble proteins in a kernel and starch yield. Brown et al (1979) found that a "steeping index" derived from visual examination of steeped kernel sections was highly correlated ($r = 0.97$) with starch recovery. However, a 32-hr steeping time was required to determine the "steeping index." Weller et al (1988), using four varieties and different drying treatments, developed a multiterm regression equation for predicting starch recovery from kernel starch content, test weight, and protein solubility in an ethanol-water-lactic acid solution. A coefficient of multiple determination (R^2) of 0.598 was obtained.

Fox et al (1992) related the compositions and physical properties of 27 hybrids to laboratory wet-milling characteristics. The highest correlation with starch yield was obtained by using a multiterm equation that included grain protein content and test weight ($R^2 = 0.61$). Correlation between starch content and starch yield was very low ($r = 0.20$).

Near-infrared reflectance (NIR) spectroscopy is a rapid analytical method that is widely used in the grain industry for determining composition. Near-infrared reflectance and transmission techniques have been shown to reliably measure the protein content of corn and also are useful for measuring oil and starch contents (Orman and Schumann 1991). Brumm et al (1991) have suggested that the technique has potential for measuring corn wet-milling characteristics, including starch yield. Using the sample set described by Fox et al (1992), regression equations were developed by Brumm and co-workers (1991) that related NIR measurements to starch yield. NIR correlations with starch yield (Brumm et al 1991) were equivalent in magnitude to correla-

tions between starch yield and other chemical or physical parameters (Fox et al 1992). Prediction of samples outside the calibration set was not attempted.

The purpose of our work was to further investigate the use of near-infrared spectroscopy to predict starch yield potential from different varieties of corn grown in different environments. Specific objectives included identifying spectral changes associated with differences in starch yield, investigating various methods of sample handling and various statistical treatments for calibration development, and evaluating the potential for a single calibration to predict sample properties across multiple crop years.

MATERIALS AND METHODS

Samples

Eight different genetic lines of yellow dent corn, representing both commercial varieties and experimental crosses, were grown at five different geographic locations in Nebraska during the 1990 crop year. The lines were representative of the genetic diversity and range of kernel hardness found in hybrids commonly grown in the midwestern United States. Three of the lines were also grown at three locations during the 1991 crop year. All experimental plots were irrigated.

At the end of the growing season, the corn was allowed to air dry in the field before harvest. After mechanical harvesting and field shelling, large pieces of cob or other debris in the bulk samples were removed by hand. Samples were stored at ambient temperature (15–25°C) until needed for analysis.

Laboratory Wet Milling

The laboratory wet-milling procedure was adapted from Watson et al (1951, 1955). Clean corn (300 g) was steeped in 1.0 L of aqueous solution containing 4.71 ml of 85% lactic acid and 1.48 g of sodium bisulfite. The grain was steeped for 48 hr at $51 \pm 1^\circ\text{C}$. After steeping, the grain was strained from the steepwater, weighed, and then placed in a Waring Blendor (blades reversed) containing 250 ml of tap water. The corn was ground for 90 sec at high speed.

After grinding, fiber was separated as the overs of a U.S. standard sieve No. 30 (600 μm square openings). Material passing through the No. 30 was then sieved with a U.S. standard sieve No. 230 (63 μm square openings). The starch-protein slurry that passed through the No. 230 sieve was transferred to the upper end of an aluminum beam starch table (3.05 m length \times 0.15 m width) with a 1.7% slope. The starch that collected on the table was washed with distilled water until all yellow, carotenoid-containing substances were eliminated, dried at low temperature (40–50°C) to remove most of the moisture, and then weighed. Residual moisture was measured by drying a subsample to constant weight at 103°C in a forced-draft air oven, and this value was used to calculate the dry weight of the starch. Percent starch yield was calculated by dividing the dry weight of the starch

¹Published as paper 10312, Journal Series, Agricultural Research Division, University of Nebraska, Lincoln 68583-0704.

²Department of Food Science and Technology, 143 Filley Hall, University of Nebraska, Lincoln 68583-0919.

³Author to whom correspondence should be addressed.

by the original dry weight of the corn. Starch purity was confirmed by determining the residual protein content of the isolated starch. Protein content was <0.5% ($N \times 6.25$) for all samples.

The wet-milling procedure was replicated three times for each corn sample.

Near-Infrared Spectroscopy

An NIRSystems model 6500 scanning spectrometer (NIRSystems Division of Perstorp Analytical, Silver Spring, MD) was used to obtain spectra of the corn samples. For instrument control and calibration development, the spectrometer was interfaced to an MS-DOS personal computer running the Near Infrared Spectral Analysis Software (NSAS) package (version 3.16) provided by NIRSystems.

Diffuse reflectance spectra were obtained from both ground and whole-kernel samples over a spectral range of 400–2,500 nm, with reflectance data recorded in a $\log 1/R$ format. Transmission spectra of whole-kernel samples were obtained over a 400- to 1,100-nm range, using $\log 1/T$. All spectra were collected at 2-nm intervals. Both transmission and reflectance spectra of whole-kernel corn were obtained using an NIRSystems coarse sample cell (transmission pathlength = 15 mm) and sample transport module. Each sample's spectrum represented 32 individual spectral scans collected and averaged over the length of the filled sample cell as it was moved through the infrared beam by the transport mechanism. The total sampling area was $\sim 80 \text{ cm}^2$.

Reflectance spectra of ground samples were obtained using a rotating solid sample cup. Again, each sample's spectrum consisted of the average of 32 individual scans obtained as the sample cup was rotated off-center from the infrared beam. Before packing of the sample cup, ground corn samples were prepared by cracking each sample in a burr mill, then grinding the cracked sample in a Tecator Powdertec sample mill (Tecator Division of Perstorp Analytical, Herndon, VA) equipped with a 1.0-mm screen. Ground samples were stored at ambient temperature in capped polyethylene vials for a maximum of one week before analysis.

Thirty samples from the 1990 crop year were selected for calibration development. Samples were selected for inclusion in the calibration set by a subroutine of the NSAS software, based on differences in the diffuse reflectance spectral properties of the whole-grain samples. The software based its selection of calibration samples on a standardized H statistic (Mahalanobis distance; Mahalanobis 1936), which indicates how different a sample's spectrum is from the average sample in the calibration set. One sample was later deleted from the calibration set due to a large standard deviation in the yield data of the reference starch, leaving 29 samples for calibration development. The 10 remaining 1990 crop year samples, along with the nine 1991 samples, comprised the validation set. The same calibration and validation sample sets were used with both the ground and whole-kernel reflectance data, as well as the transmission data.

Multiterm linear regression equations relating $\log 1/R$ or $\log 1/T$ values to starch yield were developed using forward stepwise and best possible regressions algorithms. The best possible regressions algorithm allowed a maximum of 70 evenly spaced wavelengths to be considered for regression. Multiterm equations using first and second derivatives of the spectral data were also developed. The optimum number of wavelengths for inclusion in the calibration equations was determined by comparing regression results for multiple correlation coefficient (R), standard error of calibration (SEC), partial F -values for each term in the equation, and the overall F of regression (Workman and Mark 1992). Additionally, these statistical parameters were used to select the optimum segment and gap values for calculating the derivatives (Williams 1987).

A partial least squares (PLS) algorithm was also used to develop calibration equations using $\log 1/R$, $\log 1/T$, and derivatized data. Underivatized spectra were scaled to a mean of zero and variance of one before regression. The optimum number of terms for inclusion in a PLS calibration equation was selected based on the standard errors of cross validation, which should be minimized (Workman and Mark 1992), along with the R and

SEC values obtained from the regressions.

The standard error of prediction (SEP; Workman and Mark 1992) of each calibration equation was determined by predicting the starch yield values of the validation sample set.

RESULTS AND DISCUSSION

The standard error of the laboratory wet-milling procedure was calculated from replicate millings of the 19 validation samples, as described by Workman and Mark (1992). The absolute standard error of the laboratory wet-milling procedure was found to be 2.24%.

Starch yield of the samples included in this study varied widely depending on variety and growing location, ranging from 47.2 to 64.5% recovered starch. The set of calibration samples had a mean starch yield of 56.3% and a standard deviation of 3.80%. The 19 samples in the validation set had a mean value of 58.8% and a standard deviation of 2.52%. The range of starch yields obtained was similar to that found for laboratory wet-milling of air-dried samples by other researchers (Fox et al 1992). Laboratory wet-milling procedures with batch steeping generally give starch yields lower than those achieved by commercial corn wet mills (Steinke and Johnson 1991) but may be more sensitive to differences in corn millability than are commercial practices (Fox et al 1992).

Initial regression results indicated that using near-infrared transmission data, rather than reflectance data, resulted in calibrations with substantially lower R values and higher SECs. Initial results also indicated that calibrations obtained using ground samples were no better than the calibrations obtained by measuring the reflectance of whole-kernel corn. Therefore, we chose to concentrate on developing calibrations using whole-kernel reflectance, thus eliminating the time-consuming grinding step.

Calibration results are shown in Table I. In the multiple linear regression (MLR) equations, no more than four wavelengths were used, to prevent overfitting. Multiple correlation coefficients ranged from 0.82 to 0.89. These values are lower than those normally desired for NIR methods. However, an estimation of the maximum achievable R value based on the standard error of the reference method and the range of the starch yield values for the calibration set, as described by Workman and Mark (1992),

TABLE I
Results of Various Calibration Procedures
for Measuring Starch Yield from Whole-Kernel Corn Samples

Regression Type	Spectral Treatment	Wavelengths or Number of PLS ^a Terms Used	R^b	Standard Error of Calibration (%)
Multiple linear	$\log 1/R$	1,100 nm	0.848	2.03
		1,720		
		1,760		
		1,940		
	First derivative ^c of $\log 1/R$	1,202 nm	0.892	1.86
		1,296		
		1,826		
		2,044		
	Second derivative ^d of $\log 1/R$	1,226 nm	0.872	2.01
		1,500		
		1,808		
		2,268		
Partial least squares	$\log 1/R$	6 factors ^e	0.857	2.21
		5 factors ^f		

^aPartial least squares.

^bMultiple correlation coefficient.

^cSegment = 10 points, gap = 15 points.

^dSegment = 20 points, gap = 15 points.

^eSpectral range = 1,100–2,500 nm.

^fSpectral range = 1,300–2,500 nm.

indicates that R values between 0.8 and 0.9 are the highest that can be obtained.

Prediction results for the validation set are given in Table II. Standard errors shown have been bias-corrected, as a small bias was noted between the 1990 and 1991 crop years. The overall SEP representing both crop years was calculated with the bias correction included and therefore characterizes the random error, but not the systematic error, of the NIR method. Calibrations with low SEC values (Table I) did not always yield low prediction errors for the validation set. The calibration equation using a linear combination of second-derivative terms performed substantially better with the validation set than did any of the other calibration equations. This calibration appeared to successfully predict the samples from both the 1990 and 1991 crop years; in fact, the corrected SEP for the 1991 samples was lower than for the 1990 samples. SEPs were also equal to or better than the standard error of the laboratory wet-milling procedure, which was 2.24%.

The calibration that performed second best was obtained from MLR of the $\log 1/R$ data. Prediction results were not improved when PLS regression, rather than MLR was used. For the 1990 crop year, optimized PLS calibrations using $\log 1/R$ or second-derivative data yielded SEP values nearly equal to those obtained with the corresponding MLR calibrations. However, the PLS calibrations were substantially poorer at predicting the 1991 samples, with standard errors substantially higher than those obtained from the corresponding MLR regressions. Apparently, the PLS algorithm overfit the data by including spectral characteristics unique to the 1990 data set in the calibration. Including samples from multiple crop years in the calibration set may help to overcome the latter problem.

From observation of the partial F -values associated with each

term in the MLRs (data not shown) it was evident that 2,268 nm was the wavelength of greatest significance when second-derivative terms were used, whereas 1,100 nm was most significant in the equations using $\log 1/R$. To determine whether actual spectral differences could be observed at these statistically significant wavelengths, the spectra from the three calibration samples with the lowest starch yield values were averaged, as were the spectra from the three samples with the highest values. The $\log 1/R$ and second-derivative spectra are plotted in Figure 1. Differences are observable and are more easily visualized in the second-derivative plots. Some differences in absorption band intensities, specifically at 1,432 and 1,924 nm, are most likely attributable to moisture content variation; however, other differences in both spectral shape and intensity appear to vary with differences in starch yield.

An expanded second-derivative plot of the 2,100- to 2,450-nm region is shown in Figure 2. Distinct differences in the shapes of the averaged, derivatized spectra are seen in the 2,265- to 2,325-nm region. Also, in an expanded plot of the shorter-wavelength region (Fig. 3), differences in spectral shape can be seen between the high- and low-yield samples in the 1,050- to 1,125-nm region. The visual observation of spectral differences between samples with high and low starch yields serves to confirm the statistical correlations previously observed. Additional confirmation is provided by examination of the PLS weightings, which also exhibit maxima in the 2,265- to 2,270-nm region (data not shown).

Absorption bands in the 2,250- to 2,280-nm region have been previously assigned to -O-H stretch and deformation combinations, as well as to -O-H stretch and -C-C- stretch combinations, all of which are associated with carbohydrates, including starch and cellulose (Osborne and Fearn 1986). This spectral region seems

TABLE II
Standard Errors of Prediction (SEP) Obtained from the Validation Set of 1990 and 1991 Crop Year Samples

Calibration ^a	SEP, %		
	Overall ^b	1990	1991 ^b
MLR			
$\log 1/R$	1.81	1.94	1.76
First derivative	2.07	2.16	2.08
Second derivative	1.41	1.79	0.92
PLS			
$\log 1/R$	1.95	1.98	2.03
Second derivative	1.85	1.82	2.00

^aMLR = multiple linear regression, PLS = partial least squares regression.

^b1991 samples were bias-corrected. The overall SEP includes this correction for systematic error.

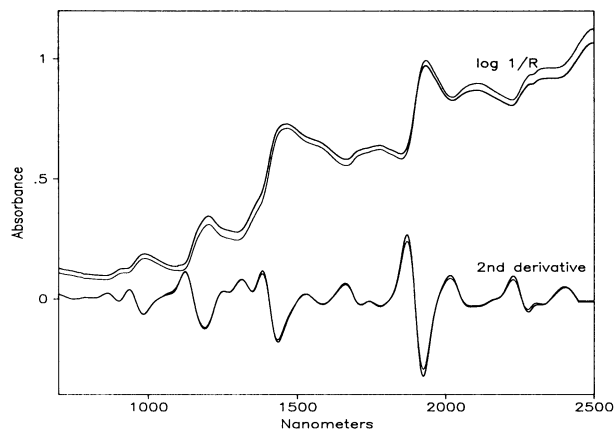


Fig. 1. The averaged spectrum from the three calibration samples with the highest starch yields, and the averaged spectrum from the three samples with the lowest starch yields. Each spectrum is shown in $\log 1/R$ and second-derivative formats.

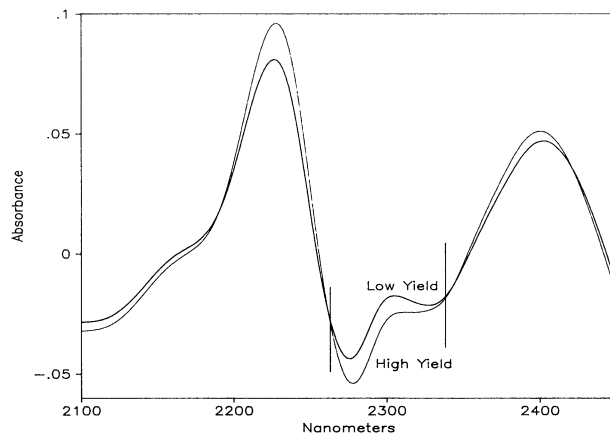


Fig. 2. Second-derivative averaged spectra of the highest and lowest samples over a spectral range of 2,100–2,450 nm. The area of spectral differences is located between the vertical lines.

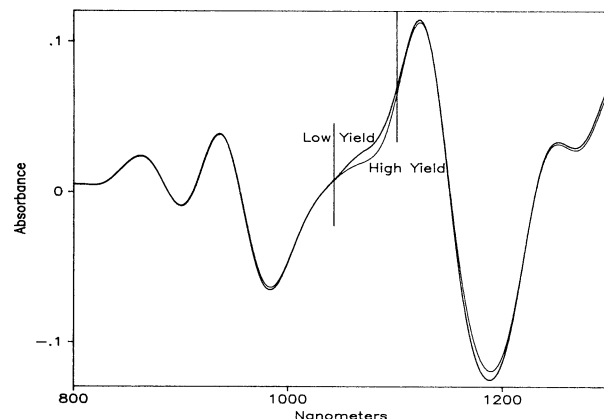


Fig. 3. Second-derivative averaged spectra of the highest and lowest samples over a spectral range of 800–1,300 nm. The area of spectral differences is located between the vertical lines.

to be sensitive to differences in starch structure, or interactions of starch with other constituents, that affect starch yield. No spectral differences were noted in the starch absorption band centered at 2,100 nm. The differences in the 1,050- to 1,125-nm range are more difficult to interpret, as most absorptions in this region are characterized by -C-H and -C-C- stretching motions associated with lipid or other hydrocarbons. It is possible that differences in this region represent a starch-lipid interaction, but additional evidence is needed before this can be confirmed.

In summary, our research has shown that NIR spectroscopy can predict the starch yield obtainable from corn and that a single calibration should be able to successfully predict samples from more than one crop year. To develop calibrations applicable on a global scale, calibration sets with a large number of samples grown over multiple crop years and with a wide geographical distribution are needed. Also, our data show that lack of reproducibility of the laboratory wet-milling reference method, rather than instrument reproducibility, is likely the limiting factor in the performance of the NIR method. This limitation has also been suggested by Brumm et al (1991). Development of a more reproducible reference procedure would help to improve the reliability of the NIR method.

ACKNOWLEDGMENT

This work was supported in part by a grant from the Nebraska Corn Development, Utilization and Marketing Board.

LITERATURE CITED

BROWN, R. B., FULFORD, G. N., DAYNARD, T. B., MEIERING, A. G., and OTTEN, L. 1979. Effect of drying method on grain corn quality. *Cereal Chem.* 56:529.

- BRUMM, T. J., JOHNSON, L. A., and HURBURGH, C. R. 1991. The potential of near infrared spectroscopy for determining maize wet milling characteristics. (Abstr.) *Cereal Foods World* 36:706.
- FOX, S. R., JOHNSON, L. A., HURBURGH, C. R., Jr., DORSEY-REDDING, C., and BAILEY, T. B. 1992. Relations of grain proximate composition and physical properties to wet-milling characteristics of maize. *Cereal Chem.* 69:191.
- LASSERAN, J. C. 1973. Incidences of drying and storing conditions of corn (maize) on its quality for starch industry. *Staerke* 25:257.
- MAHALANOBIS, P. C. 1936. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* 2:49.
- ORMAN, B. A., and SCHUMANN, R. A. 1991. Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain. *J. Agric. Food Chem.* 39:883.
- OSBORNE, B. G., and FEARN, T. 1986. Theory of near infrared spectrophotometry. Pages 37-41 in: *Near Infrared Spectroscopic Analysis*. Longman: Essex, UK.
- STEINKE, J. D., and JOHNSON, L. A. 1991. Steeping maize in the presence of multiple enzymes. I. Static batchwise steeping. *Cereal Chem.* 68:7.
- WATSON, S. A., WILLIAMS, C. B., and WAKELY, R. D. 1951. Laboratory steeping procedures used in a wet milling research program. *Cereal Chem.* 28:105.
- WATSON, S. A., SANDERS, E. H., WAKELY, R. D., and WILLIAMS, C. B. 1955. Peripheral cells of the endosperms of grain sorghum and corn and their influence on starch purification. *Cereal Chem.* 32:165.
- WELLER, C. L., PAULSEN, M. R., and STEINBERG, M. P. 1988. Correlation of starch recovery with assorted quality factors of four corn hybrids. *Cereal Chem.* 65:392.
- WILLIAMS, P. C. 1987. Variables affecting near-infrared reflectance spectroscopic analysis. Page 143 in: *Near-Infrared Technology in the Agricultural and Food Industries*. P. C. Williams and K. H. Norris, eds. Am. Assoc. Cereal Chem., St. Paul, MN.
- WORKMAN, J., and MARK, H. 1992. Auxiliary statistics for the calibration model. *Spectroscopy* 7(9):14.

[Received April 7, 1993. Accepted August 3, 1993.]