NOTE

Protein Content of Bulk Wheat from Near-Infrared Reflectance of Individual Kernels

Stephen R. Delwiche[1,2] and William R. Hruschka[1]

Cereal Chem. 77(1):86–88

ABSTRACT

Protein content of wheat by near-infrared (NIR) reflectance of bulk samples is routinely practiced. New instrumentation that permits automated NIR analysis of individual kernels is now available, with the potential for rapid NIR-based determinations of color, disease, and protein content, all on a single kernel (sk) basis. In the event that the protein content of the bulk sample is needed rather than that of the individual kernels, the present study examines the feasibility of estimating bulk sample protein from sk spectral readings. On the basis of 318 wheat samples of 10 kernels per sample, encompassing five U.S. wheat classes, the study demonstrates that with as few as 300 kernels bulk sample protein content may be estimated by sk NIR reflectance spectra at an accuracy equivalent to conventional bulk kernel NIR instrumentation.

Wheat single kernel (sk) testing for purposes of classification and quality assessment is gaining interest in both scientific and industrial communities with the advent of commercially available instruments that singulate and characterize kernels on a rapid (≈1 sec/kernel) basis. Stemming from the design of a sk instrument whose purpose was to objectively distinguish broad classes of U.S. wheat by hardness as calculated from force of crush readings (Martin et al 1993), a single-kernel characterization system (SKCS, model 4100, Perten Instruments, Springfield, IL) is now being marketed as a tool for the profiling of kernel hardness, weight, size, and moisture content. Research is currently underway at the USDA and Perten Instruments to expand the collected information by the SKCS to include visible and near-infrared reflectance data on each kernel before it is crushed. Recent studies have indicated that such an optical probe is useful in determining kernel color (Dowell 1998) and the presence of scab and levels of deoxynivalenol and ergosterol in infected kernels (Dowell et al 1999). A thorough investigation on the ability of this probe to measure protein content of individual kernels is lacking. However, previous research that utilized analytical visible/NIR spectrometry has established the feasibility of such measurement (Delwiche 1995, 1998).

Research on optical properties of single kernels has been ongoing for the past 20 years. Through studies of moisture in corn (Finney and Norris 1978) and soybeans (Lamb and Hurburgh 1991), oil in corn (Orman and Schumann 1992), and classification of rice (Osborne et al 1993) and wheat (Delwiche and Massie 1996), one constraint that has limited the commercialization of these procedures has been the time needed to handle and scan each kernel (often >1 min/kernel). With the release of the optical probe for the SKCS, as described above, sk spectral analysis in real time (≈1 sec/kernel) is attainable. Another constraint not lessened with newer equipment, however, is the time and expense of the reference methodologies. For protein content by combustion, each sk analysis is on the order of minutes, without the ability to run samples in parallel. In the event that sk protein analysis is not warranted, either because of calibration time and expense limitations or the lack of need of a sk protein profile, it would still be interesting to know whether a sk optical probe is capable of determining protein content of the bulk sample. The objective of this research was to answer this question. In the event that a technique is possible for bulk-sample protein determination by NIR reflectance of single kernels, reference chemical analyses can be reduced by a factor equal to the number of kernels used to represent a sample of wheat.

MATERIALS AND METHODS

A random selection of 318 samples was drawn from an annual market survey of U.S. grown wheat collected by the USDA Grain Inspection, Packers and Stockyards Administration (GIPSA) for the 1992 harvest. As described in detail previously (Delwiche and Massie 1996, Delwiche 1998), these samples span five market classes (hard red winter [HRW], hard red spring [HRS], soft red winter [SRW], hard white [HWW], and soft white [SWW]) and originated throughout the continental United States. Each class contained between 56 and 72 unique samples (Table I). From each sample, 10 sound kernels were drawn at random.

A custom-made reflectance attachment described in Delwiche and Massie (1996) was coupled to an NIRSystems model 6500 scanning monochromator (Silver Spring, MD). Monochromatic light (1,100–2,498 nm) was directed onto each kernel as it lay crease-side-down on the end of a vertically oriented blackened tube contained within a blackened enclosure. Reflected energy was captured by a ring of six lead sulfide detectors that were oriented 45° with respect to the vertical. Kernels were slowly revolved (1.1 rev/sec) about the tube's axis during scanning. Reflected energy was referenced to that from a pressed disk of carbon black and polytetrafluoroethylene, which has an absolute reflectance of ≈15%. Approximately 20 sec were needed to collect 32 successive scans of each kernel. Although much slower than the envisioned diode array probe described earlier, the higher signal-to-noise properties of this analytical instrument serve to establish the principle of bulk protein measurement from sk spectra.

Each kernel was weighed on an electronic balance (0.01 mg resolution) before scanning, after drying in a convection oven (130°C, 19 hr), and immediately before combustion (of the intact kernel) within a Leco model FP-428 nitrogen analyzer (St. Joseph, MI). The latter two weights were used to express protein content (N × 5.7) on a 12% moisture basis, while the former weight was used to determine the kernel moisture content at the time of scanning. Repeatability error for the nitrogen analyzer, measured as the standard deviation of 200 repeated analyses on 40-mg portions

[1] USDA-ARS, Beltsville Agricultural Research Center, Beltsville, MD 20705. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable.

[2] Corresponding author. E-mail: sdelwiche@asrr.arsusda.gov Phone: 301-504-8450. Fax: 301-504-9466.

Publication no. C-1999-1207-03R.

(i.e., equivalent to the mass of a typical wheat kernel) of a ground wheat sample over a two-month period, was 0.25% protein content.

Based on previous trials of wavelength region selection (Delwiche, 1998), only log(1/*R*) values from 1,100 to 1,398 nm were used in regression modeling. Longer wavelengths tended to increase the number of regression terms without giving the benefit of lower model error. Partial least squares (PLS) regression models (Geladi and Kowalski 1986) were developed correlating mass-averaged NIR spectra with mass-averaged protein determinations. The averaging of samples from the spectral perspective was mathematically performed at every wavelength as:

$$\log(1/R)_{\text{average}} = \frac{\sum_{i=1}^{j} \log(1/R)_i m_i}{\sum_{i=1}^{j} m_i} \tag{1}$$

where, m is the sk mass and i is the summation index over j kernels. Likewise, the reference protein content, P_{average}, of the j-kernel composite was calculated from the sk protein contents as:

$$P_{\text{average}} = \frac{\sum_{i=1}^{j} P_i m_i}{\sum_{i=1}^{j} m_i} \tag{2}$$

Note that, because protein content is the mass average of the same kernels used to form the corresponding mass-averaged spectrum, the modeling error from the PLS regressions does not include sampling error. This latter term may be estimated by application of the Central Limit Theorem to the reference protein determinations of individual kernels.

The PLS regressions for each wheat class were developed first on true samples (10 kernels per sample). Second, wheat classes were pooled to form the superclasses RED, WHITE, and ALL, which represented the three red classes, the two white classes, and all five classes, respectively. PLS regression models were developed on true samples for these superclasses. Third, kernels from more than one true sample were combined to form new samples consisting of more than 10 kernels. (Combustion protein values were available for only 10 kernels from each of the original 318 samples.)

To form these combination samples from the true samples, the true samples were ordered within each class by protein values furnished by GIPSA (NIR four-wavelength bulk determinations calibrated to Kjeldahl values). They were then grouped by 3's, 5's, or 10's to give combination samples of 30, 50, and 100 kernels, respectively. The grouping was started at the lowest protein value within each class, and any remaining true samples at the highest protein level within the class were not used. The spectral and protein averaging within each combination sample was then done using the formulas above. Because of the resulting low number of combination samples, PLS regression models were developed only on the ALL superclass in this part of the experiment.

Commercial software (PLSPlus/IQ for GRAMS/32, Galactic Industries, Salem, NH) was used to perform PLS regression modeling. Because of the limited number of samples after mass averaging, samples were not subdivided into calibration and test sets. Rather, a one-sample-out cross validation procedure was implemented. In this case, each sample is temporarily withdrawn from the pool of calibration samples, a PLS regression equation is developed, then that equation is used to predict the protein content of the removed sample. This procedure is repeated for all samples within the pool, whereupon the residuals for protein content prediction (modeled – reference) are tallied and a first-order linear regression is applied to determine the coefficient of determination (R^2) of the cross-validation. The optimal number of PLS factors is determined by an F test on the residuals (Galactic 1996).

RESULTS AND DISCUSSION

Descriptive statistics of the protein contents of the bulk samples, as well as the performance statistics of the PLS regressions are shown in Table I. The standard error of cross validation (SECV = the root mean square of the residuals) for the 10-kernel composites was essentially equivalent across wheat classes, ranging from 0.36% protein (SRW) to 0.38% protein (SWW). With the exception of SRW, R^2 values were in excess of 0.90. The reduced range of actual protein contents for SRW (7.6–11.5%) was likely the cause for the lower R^2 value. When wheat classes were pooled

TABLE I
Performance Summary of Near-Infrared (NIR) Models Where Each Sample is theWeighted Composite of 1 to 100 Individual Wheat Kernels

		Percent Protein (12% moisture basis)						
		Reference Method (combustion)				NIR Model[a]		
Wheat Class[b]	*n*[c]	Minimum	Maximum	Mean	SD	Factors	SECV[d]	R^2
10 kernels/sample								
HRW	72	8.43	16.38	10.60	1.22	10	0.365	0.909
HRS	62	10.28	15.99	13.18	1.46	10	0.383	0.931
SRW	58	7.59	11.53	8.87	0.72	10	0.363	0.748
HWW	56	9.94	15.78	11.97	1.27	10	0.374	0.912
SWW	70	7.56	14.68	10.12	1.52	10	0.385	0.935
RED	192	7.59	16.38	10.91	2.09	12	0.344	0.973
WHITE	126	7.56	15.78	10.94	1.68	10	0.371	0.951
ALL	318	7.56	16.38	10.92	1.94	13	0.348	0.968
1 kernel/sample								
ALL	3,180	4.66	21.85	10.91	2.73	14	0.634	0.946
30 kernels/sample								
ALL	104	7.99	15.67	10.86	1.77	10	0.223	0.984
50 kernels/sample								
ALL	62	8.03	14.77	10.86	1.76	10	0.177	0.990
100 kernels/sample								
ALL	30	8.42	14.62	10.83	1.72	8	0.162	0.991

[a] Partial least squares on 1,100–1,398 nm region [150 log(1/*R*) values].
[b] HRW = hard red winter, HRS = hard red spring, HWW = hard white wheat, SRW = soft red winter, SWW = soft white wheat, RED = HRW + HRS + SRW, WHITE = HWW + SWW, ALL = all wheat classes.
[c] Number of mass-averaged spectra.
[d] Standard error of cross validation, which is the root mean square of residuals (modeled – reference) from a one-sample-out cross validation.

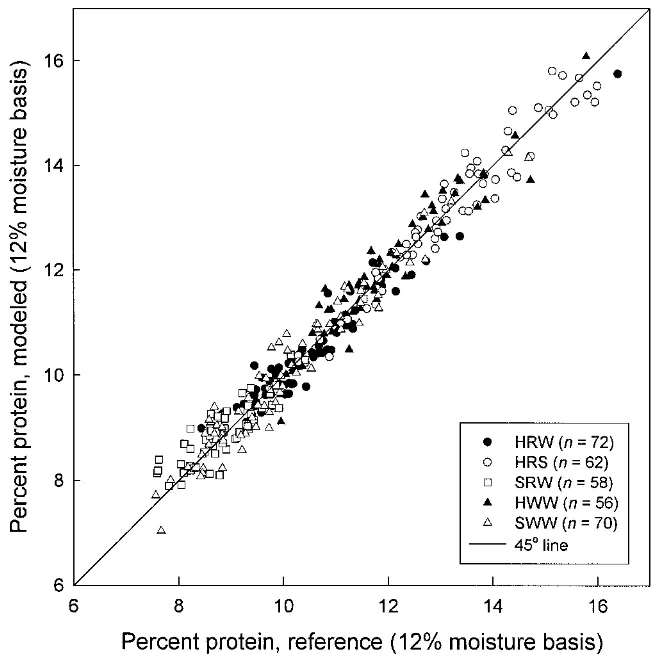

Fig. 1. Cross-validation predictions of ALL model (10 kernels/sample). ALL = all wheat classes.

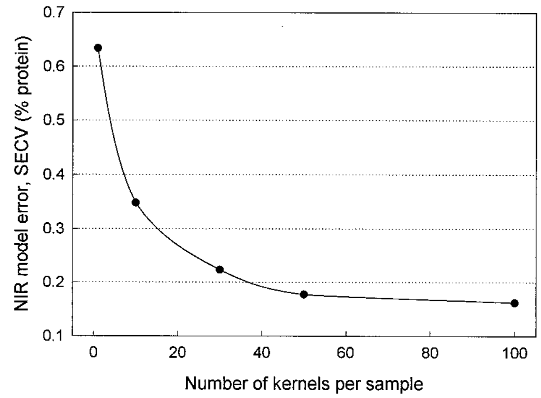

Fig. 2. Improvement in bulk sample near-infrared (NIR) protein content model error with increase in number of individual kernels used to form a mass-averaged spectrum. SECV = standard error of cross validation.

to form the superclasses RED, WHITE, and ALL, the SECV became actually lower than those of the single class models, though the number of PLS factors increased by as much as three. Thus, it appears that segregation by wheat class is not necessary for protein content of bulk wheat based on sk spectral mass averaging. A plot of the predicted vs. actual values from the ALL superclass cross-validation is shown in Fig. 1. From this plot, it is evident that the dispersion of the values from the line of perfect fit is invariant with wheat class.

When the number of kernels used in spectral mass averaging increased from 10 to 30, the SECV dropped to 0.223% protein, a value that is typical of bulk sample (>100 g) NIR calibrations (Delwiche et al 1998). A further increase to 50 and 100 kernels per mass average resulted in a decline of the SECV to 0.18% and 0.16% protein, respectively (Fig. 2). Perhaps more important than SECV alone when considering the acceptability of NIR protein calibration based on sk spectra is an estimation of error that includes that due to sampling. Assuming the independence of sampling and modeling errors, and their conformance to normal distribution, overall variance is stated as the sum of the NIR model variance (which is inclusive of instrument error) and sampling variance:

$$\begin{aligned}\sigma^2_{\text{overall}} &= \sigma^2_{\text{NIR}} + \sigma^2_{\text{sampling}} \\ &= \sigma^2_{\text{NIR}} + \frac{\sigma^2_{\text{sk protein reference}}}{n}\end{aligned} \tag{3}$$

where the first term is approximated by the square of the SECV and the numerator of the second is the square of the SD of individual kernel protein content. By substituting values from Table I into this expression (SD = 2.73% protein and SECV = 0.177% protein for n = 50 [a number found suitable for characterizing weight, size, hardness, and % moisture]) (Osborne et al 1997), overall error is estimated to be 0.42% protein. For n = 100, this value becomes 0.32% protein. Based on 300 kernels, which is the number used in a profile assessment of wheat hardness as measured on a commercial instrument (SKCS 4100), overall error is estimated to be <0.25% protein.

Thus, provided a fast NIR spectrometer (such as a diode array-based instrument) can collect spectra that are comparable to the analytical instrument described here, but on the same time scale as that used in hardness measurements (≈2 kernels/sec), a single kernel instrument for bulk sample protein estimation could be on a par with those of dedicated whole grain NIR instruments. One benefit of such a system is that an operator would not need a second instrument for measurement of bulk sample protein content.

A second benefit is strongly suggested: Although this study used averaged protein values from several single-kernel combustion determinations as the calibration reference, these averages could presumably be obtained by bulk determinations (combustion or Kjeldahl). That is, the sk spectral averages could be calibrated to bulk protein determinations, reducing the number, time, and expense of reference determinations considerably. There remains the question of whether these calibrations could then be reapplied to the individual kernel spectra, recovering the individual kernel protein values. This and the utility of such information is the subject of current research.

LITERATURE CITED

Delwiche, S. R. 1995. Single wheat kernel analysis by near-infrared transmittance: Protein content. Cereal Chem. 72:11-16.

Delwiche, S. R. 1998. Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. J. Cereal Sci. 27:241-254.

Delwiche, S. R., and Massie, D. R. 1996. Classification of wheat by visible and near-infrared reflectance from single kernels. Cereal Chem. 73:399-405.

Delwiche, S. R., Pierce, R. O., Chung, O. K., and Seabourn, B. W. 1998. Protein content of wheat by near-infrared spectroscopy of whole grain: Collaborative study. J. Assoc. Off. Anal. Chem. Intl. 81:587-603.

Dowell, F. E. 1998. Automated color classification of single wheat kernels using visible and near-infrared reflectance. Cereal Chem. 75:142-144.

Dowell, F. E., Ram, M. S., and Seitz, L. M. 1999. Predicting scab, vomitoxin, and ergosterol in single wheat kernels using near-infrared spectroscopy. Cereal Chem. 76:573-576.

Finney, E. E., and Norris, K. H. 1978. Determination of moisture in corn kernels by near-infrared transmittance. Trans. ASAE 21:581-584.

Geladi, P., and Kowalski, B. R. 1986. Partial least-squares regression: A tutorial. Anal. Chem. Acta 185:1-17.

Lamb, D. T., and Hurburgh, C. R. 1991. Moisture determination in single soybean seeds by near-infrared transmittance. Trans. ASAE 34:2123-2129.

Martin, C. R., Rousser, R., and Brabec, D. L. 1993. Development of a single-kernel wheat characterization system. Trans. ASAE 36:1399-1404.

Orman, B. A., and Schumann, R. A. 1992. Nondestructive single-kernel oil determination of maize by near-infrared transmission spectroscopy. J. Am. Oil Chem. Soc. 69:1036-1038.

Osborne, B. G., Mertens, B., Thompson, M., and Fearn, T. 1993. The authentication of Basmati rice using near infrared spectroscopy. J. Near Infrared Spectrosc. 1:77-83.

Osborne, B. G., Kotwal, Z., Blakeney, A. B., O'Brien, L. O., Shah, S., and Fearn, T. 1997. Application of the single-kernel characterization system to wheat receiving testing and quality prediction. Cereal Chem. 74:467-470.

[Received August 25, 1999. Accepted November 5, 1999.]