Effect of Mutual Interactions on the Estimation of Protein and Moisture in Wheat[1]

P. C. WILLIAMS and K. H. NORRIS[2]

ABSTRACT

Separate calibrations were set up for hard red spring wheat. Moisture content was varied from 15 to 2% by high-vacuum drying. Calibration files included 50 samples in each of five moisture ranges. Additional calibrations were run on oven-dried material. The Beltsville computerized spectrophotometer was used to select optimum and reference wavelengths for the prediction of protein and moisture. Essentially no differences were observed in the wavelengths selected for protein, but pronounced shifts occurred in the wavelengths selected for moisture determination in materials of different moisture levels. The calibrations were then used to estimate protein and moisture in "unknown" files. The most satisfactory algorithm for measuring protein in the presence of wide variations in moisture content was the normalized delta log 1/R divided by delta log $1/R_2$ algorithm. For the measurement of moisture, several algorithms were essentially equal in efficiency. The selection of optimum wavelengths for prediction based on large numbers of samples is discussed relative to wavelength selections based on smaller numbers of samples with less variance.

Cereal Chem. 60(3):202-207

The most potent factor influencing the accuracy and precision of near-infrared (NIR) spectroscopy testing of grain for protein is the nature of the surface presented to the instrumentation. One of the principal features of the surface is the packing of the particles, which is affected by the bulk density of the material, its mean particle size (MPS), its particle size distribution, and its composition, especially with regard to the presence of fibrous material and moisture. Particle size and distribution can be controlled to a fairly high degree by the employment of a suitable grinding technique; the fiber content of a grain is fairly uniform within a grain type, and neither of these features is markedly affected by growing environment. On the other hand, the moisture content of grains may vary widely even within a relatively small area, depending upon such factors as stage of maturity of the crop, rainfall, storage conditions, and relative humidity of the atmosphere. Earlier work outlined the importance of particle size in the NIR analysis of hard red spring (HRS) wheat (Hunt et al 1978, Watson et al 1977, Williams and Thompson 1978). Table I illustrates the relative influences of particle size and moisture on the prediction of protein in wheat, using commercially available instruments. The figures represent the mean standard deviation of differences between NIR and Kjeldahl protein values. Particle size ranged from 150 to 350 μm, and moisture content varied from 0 to 13.6%. In general, moisture can be regarded as a variable at least equal in importance to particle size. In practice, however, since most operators use grinders that minimize variance in particle size, moisture is a more likely source of variance. Figure 1 illustrates the influence of variation in moisture level on the NIR log 1/R traces of subsamples of a single sample of HRS wheat at four separate moisture levels. An earlier study[3] illustrated the influence of MPS on the NIR measurement of protein. The purpose of the present study was to investigate the extent to which moisture content influences the measurement of protein by NIR and vice versa and to establish a means of counteracting such influences.

MATERIALS AND METHODS

A series of Canada Western hard red spring (CWHRS) wheat, which consisted of 54 samples, was prepared so that the moisture content of the whole grain ranged from 11 to 19%. Where necessary, samples were tempered using a procedure referred to earlier (Williams and Thompson 1978). The grain samples were accurately subdivided by means of a Boerner sample divider. One portion was tested in duplicate for protein, using the Kjeldahl test and for moisture by the AACC single-stage air-oven test (AACC 1975). The protein results were reported on a moisture-free basis. The second portions (about 25 g) were carefully ground on a U-D Cyclotec laboratory grinder, using a 1.0-mm screen. This grinding procedure patterned the one recommended by NIR instrument

TABLE I
Relative Influences of Particle Size (P/S) and Moisture on Prediction of Protein by Near-Infrared Reflectance Spectroscopy

Parameter[a]	Variable	InfraAlyzer	GQA[b]-41	GQA-31
RMSD	P/S	.870	.499	.479
SEE	P/S	.329	.272	.315
RMSD	H_2O	.712	1.35	1.17
SEE	H_2O	.249	.266	.291

[a] RMSD = Root mean square of the differences between NIR and Kjeldahl protein. SEE = Standard error of calibration.
[b] GQA = Grain Quality Analyzer.

[1] Paper 509 of the Canadian Grain Commission, Grain Research Laboratory, Winnipeg, Manitoba, Canada R3C 3G9.
[2] Instrumentation Research Laboratory, USDA Agricultural Research Center, Beltsville, MD 20705.

[3] K. H. Norris and P. C. Williams. 1982. Unpublished data.

manufacturers. The ground samples had an MPS of 201 ± 9.2 μ, as determined by the procedure described earlier (Williams and Thompson 1978). Moisture contents ranged from 9.5 to 14%. The Beltsville Universal computerized spectrophotometer (BUCS) was used to scan the ground samples from 10,000 to 26,384 Å. The NIR signals were recorded as log 1/R (apparent reflectance). Individual readings were taken at intervals of 2 Å. For final recording, the log 1/R readings were smoothed over 20 wavelength points, and the original 8,192-point arrays shrunk to 1,024 points for convenience both of storage and of future data processing. Each data point of the final information files corresponded to 16 Å.

Next, the ground samples were dried under high vacuum, using a freeze drier for 6 hr to reduce the moisture level. The samples were again scanned and recorded by NIR. By the progressive freeze-drying and NIR recording, four series of HRS wheat samples were produced, each differing in mean moisture level. These sets of samples were named 1W2, 1W4, 2W3, and 2W5. A fifth series was produced by oven-drying the final set of samples at 130° C for 1 hr, cooling, and recording with the BUCS. This set was referred to as 1W0. A Dickey-john sample cell holding 2–3 g of ground sample was used for all NIR readings. Samples were discarded after each individual reading. The original 25 g of sample was sufficient for the complete study.

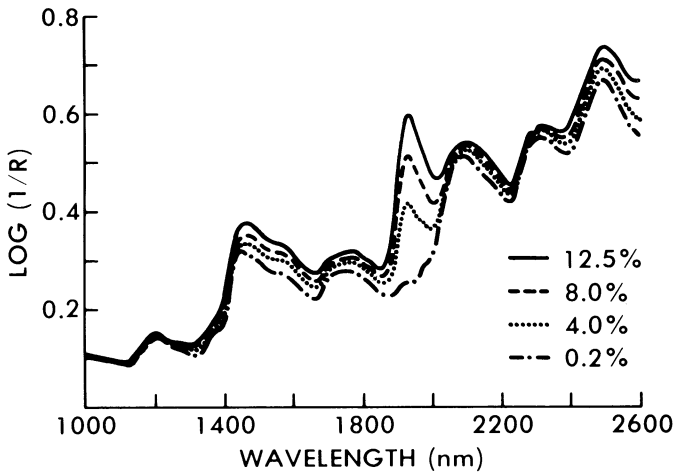

Fig. 1. Influence of moisture content on the reflectance spectra of wheat. Each curve is the average for 50 samples at the respective moisture levels.

For each series, the oven moisture contents were determined, and the "as-is" protein content recalculated, for use in subsequent computation. Because the main objective of the study was to investigate the influence of moisture level on the NIR measurement of protein, we felt that computation of the updated, "as-is" protein levels on the basis of the new moisture levels would minimize the variance in true Kjeldahl protein values, while at the same time conserving the sample. The standard deviation of moisture contents within each series was no greater than 0.8%, ie, individual series of samples did not vary greatly in moisture content within the series. Finally, a set of 48 wheat samples was prepared by culling several samples from all of the wheat series to give a very wide variance in moisture (standard deviation was over 4%). This set was named "WMX." Series 12W was composed of 25 samples of No. 1 CWHRS wheat and 25 samples of No. 2 CWHRS wheat with "normal" ranges of moisture and protein. This set was prepared for an earlier study (Norris and Williams 1978) and was included in the present study as an extra unknown series to test the accuracy of prediction.

The above seven sets of samples were specially prepared to possess a wide range of moisture between sets but a narrow range within each set. An additional set of 95 samples of Nos. 1 and 2 CWHRS wheats were assembled that possessed "normal" ranges of both moisture and protein. These represented carlots of wheat

TABLE II
Composition of Hard Red Spring Wheat Series

File	No. of Samples	Oven H_2O (%)	Standard Deviation	Kjeldahl Protein (%, "as-is")	Standard Deviation
WMX	48	6.8	4.11	13.6	1.25
1W2	50	12.5	0.80	12.6	1.17
1W0	50	0.2	0.13	14.4	1.30
2W3	50	9.7	0.40	13.6	1.24
12W	50	10.1	0.70	13.7	1.18
1W4	50	4.6	0.62	13.9	1.26
2W5	50	6.1	0.68	13.5	1.28
WHP	38	11.0	1.15	15.5	1.08
WLP	36	11.0	1.32	11.8	0.96
WHM	30	12.2	0.74	13.3	2.29
WLM	34	9.9	0.68	13.9	1.72
WMTP	48	11.2	1.45	13.7	2.05
WMTQ	47	10.8	.71	13.5	1.80

TABLE III
Accuracy of NIRS Moisture Determination Using Different Algorithms

Mathematical Treatment	Points Summed	Wavelength (angstrom units) 1	2	3	4	5	6	SEP[a] (%)	Mean Bias (%)	Comments
Log 1/R	2	18,912	23,008	19,536	17,200	22,496	22,304[b]	0.111	−0.006	BUCS selected
Log 1/R	2	19,536	23,008	17,200				0.116	−0.016	BUCS selected
Log 1/R	2	19,392	16,800	21,008	22,304	21,792	23,104	0.123	−0.001	Simulated I:GAC
Log 1/R	2	19,392	23,104	22,304				0.128	−0.005	Simulated I:GAC
Log 1/R	2	19,392	23,104	22,304				0.201	0.063	Simulated I:GAC
Log 1/R	2	19,392	23,104					0.128	−0.014	Simulated I:GAC
KM	2	19,392	16,800	21,008	22,304	21,792	23,104	0.151	−0.009	Kubelka/Munk
KM	2	19,392	23,104	22,304				0.155	−0.021	Kubelka/Munk
ΔLog 1/R	6	18,496						0.169	0	
d^2 (Log 1/R)	6	19,568						0.228	0	Second derivative
d (Log 1/R)	2	(19,320 −	18,690)	(21,630 −	21,080)	(23,040	− 22,640)	0.182	0.005	Simulate GQA 31 EL
dR/R	4	18,896	21,488	20,800	22,912			0.208	−0.037	Simulated GQA 41
dR/R	6	18,608[c] ÷	17,204					0.128	−0.020	Normalized dR/R
dR/R	6	18,496						0.169	0	
d (Log 1/R)	6	18,608[c] ÷	17,024					0.128	−0.020	Normalized delta log 1/R
d^2 (Log 1/R)	6	19,552[c] ÷	14,048					0.146	0.010	Normalized second derivative
KM	2	19,312[c] ÷	20,624					0.121	−0.024	Normalized KM
KM	2	19,312[c] ÷	21,440	19,952 ÷	22,544			0.100	−0.006	Normalized KM

[a] Standard error of prediction.
[b] Selected by BUCS during calibration, but not used by computer in prediction.
[c] Division symbols represent normalized treatments.

unloaded at the ports of Vancouver, British Columbia, and Thunder Bay, Ontario. They were ground on the U-D Cyclotec grinder, tested for Kjeldahl protein and oven moisture in the Grain Research Laboratory, Winnipeg, and used in Beltsville to generate sets of samples to investigate further the influence of mathematical treatments on the NIR measurement of moisture, and the interactions between protein and moisture on NIR moisture measurement. The composition of these additional sets reflected various combinations of the 95 samples, assembled to provide high and low mean protein at similar mean moisture levels (WHP and WLP), high and low mean moisture at similar mean protein (WHM and WLM), and two additional sets with similar mean protein and moisture levels but with high (WMTP) and low variances in moisture (WMTQ). The number of samples, mean protein, and moisture and variance (standard deviation) of all sets of samples are summarized in Table II.

The selection of the optimum wavelengths, the use of the optimum mathematical treatment of log 1/R signals, the inclusion of maximum anticipated variance in the calibration, and the accuracy of the chemical testing upon which the calibration is based are the most important factors in determining the efficiency of an NIR measuring device. The virtues of an NIR calibration are usually evaluated by means of the standard error of estimate (SEE) and multiple correlation accruing from the computing of the calibration constants. These figures can be misleading because the SEE is based upon the computing of residuals from the samples used in the calibration, and do not necessarily indicate the accuracy that will be attained in the NIR analysis of future "unknown" samples not used in the calibration. The true efficacy of an NIR calibration can best be evaluated by the analysis of a series of "unknown" samples, followed by computation of the mean deviation (d) from standard chemical or physical values, and the standard deviation of these deviations, ie, the standard deviation of the differences (SDD). Accordingly, in this study after wavelength optimization and calibration on each individual series of samples, the BUCS was used to predict protein in selected series of "unknown" samples. The standard deviation of differences between NIR and standard results for protein is referred to in this article as the standard error of prediction (SEP). Collections of samples are referred to as sets or series. The log 1/R data recorded on magnetic tape from each individual set of samples are collectively referred to as "files." Since biases cause more problems under operational conditions than high standard errors of difference, bias values are also quoted for all comparisons. The SEE is referred to in this article as the standard error of calibration (SEC). The procedure was repeated using each of 14 different mathematical treatments of the log 1/R signal. Three of these treatments were similar to the math treatments employed in commercial instruments. These were the six simple log 1/R signals taken at wavelengths used in the Technicion InfraAlyzer and the Dickey-john Grain Analysis Computer, the three delta log 1/R points similar to those of the Neotec GQA model 31, and the four delta R/R points similar to the Neotec GQA 41. We recognize that a computer can only simulate rather than reproduce the operation of a commercial instrument, and the terms "I:GAC," "G31," and "G41" are used only to indicate the algorithm employed. The first samples, totaling 348, were then scanned by the BUCS to select the optimum wavelengths and reference wavelengths for the prediction of protein in HRS wheat in the presence of maximum variance in moisture. Individual files ranged from 0.4 to nearly 13% in average moisture content and possessed a range of protein of 9.8 to more than 17%. This procedure was called mass wavelength selection.

RESULTS

Optimization of Mathematical Treatment

The BUCS was used to optimize wavelengths for prediction and, where necessary, reference wavelengths for 12 mathematical treatments, using all 95 samples used in the composition of WMTP and WMTQ. The BUCS was calibrated to WMTQ (47 samples) and used to predict moisture in the WMTP series. The mathematical treatments used included discrete log 1/R signals, difference calculations (Δ log 1/R), first and second derivatives of log 1/R, Kubelka/Munk treatment, and normalized versions of the last three treatments. The mathematical treatments used, including details of wavelengths, are described in an earlier paper (Norris and Williams 1978), and summarized in Table III. Normalized treatments are indicated by division symbols. The "points summed" column in Table III gives the optimum number of wavelength points to be summed by the computer in order to smooth the signal before computing the result. The BUCS makes only one pass of a sample and records a reading every 16 Å in single

TABLE IV
NIRS Prediction of Moisture and Protein in High- and Low-Protein Files of Hard Red Spring Wheat Using Different Mathematical Treatments

Mathematical Treatment	No. of Wavelength Points	High-Protein Calibration SEP (%)	High-Protein Calibration Bias (%)	Low-Protein Calibration SEP	Low-Protein Calibration Bias
Log 1/R	6 (selected)	0.125[a,b]	0.020	0.115[a]	0.023
Log 1/R	3 (selected)	0.122[a]	−0.036	0.115[a]	−0.003
Log 1/R I	6 (InfraAlyzer)	0.144[a]	0.069	0.144[a]	−0.048
Log 1/R I	3 (InfraAlyzer)	0.128[a]	0.012	0.146[a]	−0.075
Log 1/R I	2(1) (InfraAlyzer)	0.255[b]	0.199	0.322[c]	−0.276
Log 1/R I	2(2) (InfraAlyzer)	0.128[a]	0.012	0.146[a]	−0.075
KM	6	0.165[a]	0.003	0.204[b]	−0.120
KM	3	0.168[a]	−0.034	0.161[a]	−0.065
Δ log 1/R	1	0.248[b]	−0.177	0.244[b]	0.182
d^2 log 1/R	1	0.331[c]	0.132	0.233[b]	−0.137
Δ Log 1/R	3 (GQA 31)	0.231[b]	−0.126	0.222[b]	0.118
dR/R	4 (GQA 41)	0.313[b]	−0.205	0.238[b]	0.040
dR/R	1	0.248[b]	−0.177	0.244[b]	0.182
d Log 1/R	1(3)	0.113[a]	0.022	0.132[a]	−0.022
d R/R	1(3)	0.121[a]	0.040	0.142[a]	−0.038
d^2 Log 1/R	1(3)	0.204[b]	−0.053	0.157[a]	0.054
KM	1(3)	0.124[a]	0.061	0.137[a]	−0.059
KM	2(3)	0.124[a]	−0.011	0.112[a]	0.011
Mean standard deviation		0.183 ±	0.071	0.179 ±	0.059

[a] Wavelengths = 19,392 and 22,304 Å.
[b] Wavelengths = 19,392 and 23,104 Å.
[c] Normalized treatments.
[d] Means with different subscripts differ significantly at $P = .05$.

precision from 10,000 to 26,384 Å. To reduce system noise, it was necessary to sum a certain number of signals and divide the total by the number of points summed before computing. Commercial instruments achieve smoothing by taking multiple readings of the sample (usually 80–120, depending on the instrument).

Table III summarizes the efficiency of calibrations incorporating 18 different treatments for the prediction of moisture. The SEP data contained the earlier observations of Law and Tkachuk that the Kubelka/Munk algorithm was satisfactory for the prediction of moisture. These workers did not report extensive testing of other algorithms, and the present study revealed several algorithms capable of predicting moisture with excellent accuracy. A test originally described by Cochrane (1941) was used to test the homogeneity of the SEP values. The square of the highest SEP value was compared with the sum of the squares of the remaining 17 SEP values. The highest SEP value from Table III (0.228) was found significantly different from the remainder. By eliminating the highest SEP from subsequent comparisons, the next six highest SEPs were all found to be significantly higher than the remaining SEPs. When the value of 0.155 was reached for the SEP (between standard oven and BUCS moisture, using the Kubelka/Munk algorithms with three wavelength points), this value and all lower SEPs were found to be not significantly different from the remaining 11 SEP values. In general, the measurement of moisture in HRS wheat showed less dependence on mathematical treatment than did the measurement of protein. This is probably because the signal for moisture in the area 1,840–1,950 nm is much stronger than the interfering absorbers in that area than is the signal for protein relative to the interfering absorbers in the area 2,130–2,190 nm, which is usually associated with protein prediction.

Interactions Between Protein and Moisture in the Measurement of Moisture

The two series WHP and WLP differed by about 4% in mean protein level but had fairly low standard deviations in protein content (1.08 and 0.96). Their mean moisture contents were the same. The BUCS was calibrated using WHP and all algorithms used in the first experiment, and the calibrations were used to predict moisture in WLP, and vice versa. The optimum wavelengths selected were the same as in the first experiment, and the results are summarized in Table IV in terms of SEP and bias. Calibrations based on WLP and WHP, which had lower protein variance than WMTP and WMTQ, gave generally higher SEP and bias values than when the calibrations were based on the samples with higher protein variance. Also, when calibrations were based on samples with high protein, in 15 out of 18 comparisons the signs of the biases reversed when the calibrations were based on the lower protein samples. When calibrations were based on the same samples, but rearranged in series that gave equal protein distribution, no significant biases in moisture were observed.

Analysis of variance showed that there was no overall significant difference in the accuracy of predicting moisture whether the calibrations were based on high- or low-protein samples. Several algorithms (11 out of 18) were equally effective in the prediction of moisture, including the log 1/R, and the normalized d(log 1/R) and dR/R, and Kubelka/Munk algorithms. The protein influence did emphasize the fact that certain other mathematical treatments may become less effective in measuring moisture when protein variance is low in the calibration samples, although protein had less effect on the more efficient algorithms. For example, the SEPs for unnormalized (simple dR/R, Δ(log 1/R), d^2 (log 1/R), and the simulated GQA 31 and 41-type algorithms all increased significantly when protein variance was reduced by calibrating to the WHP or WLP samples. Although this applied to both high- and low-protein calibrations, it was slightly more pronounced when the high-protein calibration was used for the prediction of moisture in the lower-protein file. Analysis of variance showed that the mean SEP of the moisture predictions from calibrations based on high-protein variance samples was significantly lower than the mean SEPs of the two reduced-protein variance series. When the high-protein variance and the two low-protein variance calibrations were compared, 12 of the math treatments were not significantly different from each other (Table V). These data indicate that for the greatest accuracy in predicting moisture in HRS wheat, the samples used in calibration should be uniformly distributed with respect to protein as well as to moisture.

Influence of Protein on the Measurement of Protein

Table VI illustrates the influence of protein on the prediction of protein in using a number of different mathematical treatments. Protein was predicted in the low-protein file (WLP) using calibrations based on the high-protein file (WHP), and vice versa, using several mathematical treatments. Lower protein variance had the overall effect of increasing the error of prediction in all cases, as indicated by the standard error of prediction, and bias data. The normalized d(log 1/R) algorithm, and the combination of four delta R/R wavelengths used in the GQA 41 appeared to be influenced to the greatest extent by the protein effect. The log 1/R wavelengths used in the original InfraAlyzer and Dickey-john GAC showed least bias, although the standard deviation data were comparable with those of other algorithms. The overall effect for all algorithms suggested that calibration on high protein tended to cause the average results to be biased slightly upwards by about 0.06%. Calibration on low-protein wheat had a more significant effect in bringing the results of analyzing higher protein wheats downwards (overall bias was −0.17). Application of the Cochrane test (1941) to the SEPs for the high-protein calibration showed no significant difference between any of the treatments, whereas in the case of the low-protein calibration the SEPs for the log 1/R (computer selected) and normalized d(log 1/R) treatments were significantly higher than those of the remaining treatments. These observations verified those of earlier work (Williams and Thompson 1978) where calibrations were used in which the population mean protein ranged from 10.5 to 15.5, and the conclusion was that calibration on low-protein samples biased the results of NIR testing for protein in higher protein wheats downwards, although the reverse was not necessarily true.

Influence of Moisture Level of Calibration Samples on the Prediction of Protein

The first seven series of wheat of different moisture levels and variance described in the experimental section were individually used to calibrate the BUCS, which was then used to predict protein in each of the remaining six files. In view of the superiority of

TABLE V
Mean Standard Errors of Prediction and Bias Illustrating Influence of Protein Variance on Prediction of Moisture in Hard Red Spring Wheat

	Mean SEP[a]	Standard Deviation	Mean Bias	Standard Deviation
High variance (WMTQ)	0.150	0.036	0.013	0.017
Low variance (WHP)	0.183	0.071	−0.014	0.106
Low variance (WLP)	0.179	0.059	−0.017	0.112

[a]Standard error of prediction.

TABLE VI
NIRS Prediction of Protein in High- and Low-Protein Files of Hard Red Spring Wheat Using Different Mathematical Treatments

Mathematical Treatment	No. of Wavelength Points	High-Protein Calibration SDD[a]	High-Protein Calibration Bias	Low-Protein Calibration SDD[a]	Low-Protein Calibration Bias
Log 1/R	6 (selected)	0.207	0.002	0.455	−0.288
Log 1/R I	6 (InfraAlyzer)	0.238	−0.024	0.315	−0.054
d Log 1/R	3 (GQA 31)	0.354	0.200	0.280	−0.030
dR/R	4 (GQA 41)	0.349	0.224	0.355	−0.257
d log 1/R	1 (normalized)	0.286	−0.130	0.453	−0.358
d^2 Log 1/R	1 (normalized)	0.312	0.213	0.205	0.015
KM	3 (normalized)	0.243	−0.048	0.338	−0.194

[a]Standard deviation of differences.

TABLE VII
Influence of Moisture Level of Calibration Samples on Accuracy of Prediction of Protein in Wheat Using Different Algorithms

Calibration File	SEP[a] (bias corrected) I:GAC	G31	G41	DR/R÷DR/R	D Log $1/R_1$÷ D Log $1/R_2$ (one term)	D_2 log $1/R_1$÷ D_2 log $1/R_2$ (one term)	M KM/KM (three terms)
WMX[b]	.303	.377	.313	.249	.262	.290	.274
1W2	1.015	1.521	2.023	.262	.260	.358	.310
1W4	1.259	3.222	2.538	.253	.254	.313	.319
1W0	.427	.477	.443	.251	.261	.318	.367
2W3	.945	2.321	2.556	.240	.240	.316	.657
2W5	.314	2.418	1.388	.238	.240	.291	.476
12W	.874	1.162	1.407	.261	.259	.298	.392
Overall SEP[a]	.734	1.642	1.524	.251	.254	.312	.399
Overall SEC[c]	.250	.294	.266	.243	.245	.287	.260

[a]SEP = Standard error of prediction, ie, standard deviation of differences between Kjeldahl and NIR protein.
[b]Calibration based on WMX. Individual standard errors of prediction (SEPs) represent mean SEPs for prediction of protein in remaining six files from the WMX calibration, using the different math treatments. Other SEPs based on calibrations to 1W2, etc.
[c]SEC = Standard error of calibration, or standard error of estimate.

TABLE VIII
Influence of Individual Wavelength Selection in Files of Different Moisture on Protein Prediction Compared to Mass-Selected Wavelength

File	WMX	1W2	1W4	1W0	2W3	2W5	12W	Overall
Individual	21504 Å	21376	21600	21584	21632	21424	21504	···
SEP (bias)	.282	1.640	.786	1.324	.997	.837	.282	.871
SEP (no bias)	.252	.616	.375	.418	.425	.385	.252	.382
Mass	21504	21504	21504	21504	21504	21504	21504	···
SEP (bias)	.282	.290	.284	.268	.266	.263	.296	.273
SEP (no bias)	.252	.266	.266	.256	.242	.242	.265	.256

TABLE IX
Influence of Varying Wavelength Used for Prediction of Protein in Presence of Wide Variations in Moisture[a]

Wavelength Variation		0	16	32	48	64	80A
Above 21,504 Å	bias	+0.08	+0.157	+0.203	+0.201	+0.281	+0.306
	SEP[b]	0.256	0.263	0.295	0.339	0.383	0.427
Below 21,504 Å	bias	+0.08	−0.129	−0.208	−0.338	−0.489	−0.616
	SEP[b]	0.256	0.271	0.295	0.303	0.291	0.337

[a]Mass wavelength for prediction = 21,504 Å.
[b]Standard error of prediction.

normalized math treatments demonstrated earlier[3] for the prediction of protein in the presence of variable particle size, only the normalized math treatments and simulated instruments were tested in this part of the study. The results are summarized in Table VII in the form of mean SEP for the prediction of six files. The algorithm and wavelength combinations similar to those used in the simulated commercial instruments displayed a highly significant influence of moisture level in the calibration samples on the subsequent prediction of protein in HRS wheat. When the BUCS was calibrated with samples carrying a wide range in moisture (file WMX) the results were acceptable. But when the variance in moisture was restricted, predictions were not as good, particularly with the simulated commercial instruments. An interesting observation was that predictions based on calibration to 1W0, the low moisture set, were fairly satisfactory. This was believed due to the presence of only very small amounts of water, which reduced interference.

The four normalized treatments were all superior to any of the simulated instruments. The d R/R and d(log 1/R) normalized treatments were generally superior to either the second derivative or Kubelka/Munk normalized treatments. The overall mean standard errors of calibration (standard errors of estimate) were all satisfactory. The SEC and SEP values for d R/R and d(log 1/R) normalized treatments were similar in magnitude, which indicated that most of the error in the prediction was attributable to sampling and sample preparation rather than analysis. Differences between SEC and SEP became progressively larger as the SEP increased.

The Mass Selection System of Wavelength Optimization for Calibration

Individual calibrations were set up on the BUCS using each of the first seven individual files carrying samples of different mean moisture levels and standard deviations. The BUCS selected different primary wavelengths for prediction and normalization. Each individual calibration was based on a single d(log 1/R) prediction wavelength point divided (normalized) by a second wavelength point. A multiple calibration was then established using all of the seven files collectively (348 samples), again using a single wavelength point for prediction and a second point for normalization. The calibrations based on each individual file were used to predict protein in the remaining files using the wavelengths selected. The predictions of protein were compared to predictions of protein in the same files using the wavelengths and calibrations of the multiple calibration (mass selection). The results are summarized in Table VIII. The mass-selected wavelength calibration was clearly superior to calibrations based on wavelengths selected for individual files.

The derivatization for the d(log 1/R) algorithm was based on wavelengths six points above and six points below the central wavelength point. For example, for a primary central prediction wavelength point of 21,504 Å, the actual wavelengths used in the d(log 1/R) algorithm were separated from 21,504 by 6 × 16 Å above and below the central point, ie, 21,600 and 21,408 Å respectively, and the wavelengths for normalization were 22,656 ± 96 Å, ie, 22,752 and 22,560 Å. To test the tolerance of a calibration to variation in the central wavelength point, the BUCS calibrations were set up using file WMX, the high-moisture variance file, and the calibrations used to predict protein in the remaining six files.

The first calibration used the primary wavelength, 21,504 Å. The central wavelength point was then varied in increments of 16 Å above and below the central point, and calibrations were set up based on these wavelengths. The wavelength for normalization was held constant at 22,656 Å. Each calibration was used to predict protein in the remaining six files. The results of predictions in terms of SEP and bias are summarized in Table IX. Both the mean SEP and bias for predicting protein in the six files increased progressively as the primary wavelength point was moved away in either direction from the optimum point of 21,504 Å. Significant increases in both SEP and bias increased when the central wavelength was varied by 32 Å or more from the optimum. When the experiment was repeated, holding the optimum central wavelength for prediction constant but varying the normalizing or dividing wavelength, the calibration showed less sensitivity, and the reference wavelength could be varied by ±80 Å without causing significant increases in SEP or bias. The normalized log 1/R traces for two subsamples of the same wheat with widely different moisture contents revealed that, in the area of the prediction wavelength, the two curves coincided over a much shorter area than at the area of the normalizing wavelength (Fig. 2). This would account for the greater stability of the normalizing wavelength.

These observations are analagous to the improvement in efficiency of commercial NIR instruments when large numbers of samples are used in the calibration. The wider the variance in samples introduced to the calibration, the more capable the instrument will be of accurate analysis. The development of "universal" calibration constants or "K values" for all commercial on-line NIR instruments developed at the Grain Research Laboratory (GRL), Winnipeg, has shown to be effective in the prediction of protein in wheat. Universal constants are now a routine feature of the InfraAlyzer, GAC, and GQA series of NIR instruments. The constants developed at the GRL used over 2,000 individual samples drawn from several seasons and many locations, whereas the more usual procedure is to use 30–40 samples in a calibration. These smaller individual calibrations eventually fall susceptible to variations in season, growing location, variety, or some other variable, and are generally less efficient than the "universal" constants. In the present experiment, selection of wavelengths for prediction based on large numbers of samples showing very wide variance in moisture were found to be considerably more effective in subsequent predictions than wavelengths selected on smaller numbers of samples with less variance in individual series.

Using a combination of a normalized mathematical treatment and the mass selection system for wavelength optimization, a highly efficient calibration system was developed that involved only a single normalized wavelength point, and was capable of providing accurate predictions of protein in wheat despite wide variations in the moisture contents of the samples. When the system was used to calibrate a NIR instrument using samples with an average moisture content of 0.2%, accurate protein predictions were made in files of samples with average moisture levels of over 12%, and vice versa. The simplicity of the algorithm, and its use of optical data from only a single normalized wavelength point (four filters: two for prediction and two for normalization in a filter instrument), together with the efficiency in protein prediction over a wide range of moisture variance should make normalized Δ(log 1/R) a very useful algorithm for use in future NIR instrumentation. Normalization of mathematical treatments of raw log 1/R optical data is equally applicable to d R/R, d^2(log 1/R) Kubelka/Munk and simple log 1/R algorithms. The principle of normalization, or dividing the signal at a prediction wavelength by a second signal at a reference wavelength, constitutes the most significant end-result of the experiments described above.

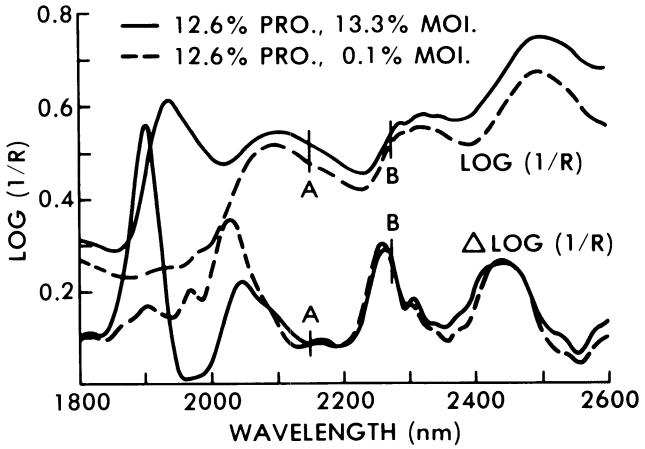

Fig. 2. Delta log 1/R spectra for two subsamples of wheat with widely different moisture contents. A = Protein prediction wavelength; B = protein normalization wavelength.

ACKNOWLEDGMENTS

S. Stevenson, H. M. Cordeiro, and Joyce Shaffer are acknowledged for their competent technical assistance in the preparation of the samples and in data processing, respectively. The authors are grateful to Ron Moen for his helpful advice with some of the statistical treatments applied.

LITERATURE CITED

AMERICAN ASSOCIATION OF CEREAL CHEMISTS. 1975. Approved Methods of the AACC. Method 44-15A, approved May 1968; and Method 45-16, approved May 1975. The Association, St. Paul, MN.

COCHRANE, W. G. 1941. The distribution of the largest of a set of estimated vacancies as a fraction of their total. Ann. Eugen. 11:47.

HUNT, W. H., FULK, D. W., THOMAS, J. C. and NOLAN, T. 1978. Effect of type of grinder on protein values of hard red winter wheat when analyzed by infrared reflectance devices. Cereal Foods World 23:143.

LAW, D. P. and TKACHUK, R. 1977. Determination of moisture content in wheat by near-infrared diffuse reflectance spectrophotometry. Cereal Chem. 54:874.

WILLIAMS, P. C. and THOMPSON, B. N. 1978. Influence of whole meal granularity on analysis of HRS wheat for protein and moisture by near-infrared reflectance spectroscopy (NRS). Cereal Chem. 55:1014.

WATSON, C. A., SHUEY, W. C., BANASIK, O. J. and DICK, J. W. 1977. Effect of wheat class on near-infrared reflectance. Cereal Chem. 54:1264.

[Received June 26, 1981. Accepted November 24, 1982]