Classification of Wheat by Visible and Near-Infrared Reflectance from Single Kernels

STEPHEN R. DELWICHE[1,2] and DAVID R. MASSIE[1]

Cereal Chem. 73(3):399-405

ABSTRACT

Identification of wheat class is a necessary component of the official inspection of U.S. wheat, owing to differences in functionality and hence in trade value. Because of the numerous cultivars for the several U.S. wheat classes, segregation by cultivar is generally impractical during postharvest handling. Cultivars of differing wheat classes are sometimes inadvertently mixed, resulting in classification of the lot to a mixed category, thus lowering its value. Single-kernel near-infrared reflectance scans from two spectral regions (551–750 nm for distinctions based on color, 1,120–2,476 nm for distinctions based on intrinsic properties) were collected on 10 randomly drawn kernels from each of 318 unique samples obtained from commercial sources. Partial least squares and multiple linear regression analyses were used to develop binary decision models for various combinations of two wheat classes, choosing from five classes: hard white (HWH), hard red spring (HRS), hard red winter (HRW), soft red winter (SRW), and soft white (SWH). Two-class model accuracy, defined as the proportion of correctly identified kernels of a known wheat class, was greatest (99%) when red and white classes such as HRW vs. HWH were compared. Accuracies declined to typically 78–91% when the two classes were of similar color (e.g., HRW vs. SRW, HWH vs. SWH). Using a cascade of binary comparisons similar to two-class models, a five-class model structure was developed. Five-class model accuracy ranged from 65% for SRW wheat to 92% for SWH.

Because of its large geographical and climatological diversity, the United States produces several classes of wheat. Unlike governments of many of the other wheat-producing countries, the U.S. government does not regulate the development nor license the release of new wheat lines. Because of these factors, numerous wheat cultivars exist in the U.S. market. In the market channels, it is common for wheat to be segregated by class but not by cultivar, although classes may be intentionally blended to produce a specialty product. When a mixture of two or more classes occurs inadvertently, rendering a lot <90% pure of one class, that lot must be classed as mixed, which causes its price to be devalued. All export wheat and some domestic wheat is inspected by the United States Department of Agriculture, Federal Grain Inspection Service (FGIS). When assigning wheat class, inspectors rely upon their knowledge of where the wheat was grown and an examination of the size, color, and shape of individual kernels. The proliferation of new varieties, often grown outside the traditional geographic areas formerly associated with each wheat class, has placed a large burden on inspectors to maintain the accuracy of the current classification system. Research efforts are underway at several locations (Kansas City, MO; Manhattan, KS; Peoria, IL; and Beltsville, MD) to develop rapid and objective classification methods to supplement the current system.

Various approaches have been examined for identification of wheat class or cultivar. Methods include polyacrylamide gel electrophoresis (Wrigley et al 1982; Royal Australian Chemical Institute 1988), reversed-phase high-performance liquid chromatography (Bietz et al 1984, Bietz and Cobb 1985, Marchylo et al 1988, Endo. et al 1990, Huebner et al 1995), fluorescence spectrometry (Irving et al 1989), and digital image analysis (Zayas et al 1985, 1986; Neuman et al 1987; Sapirstein et al 1987; Symons and Fulcher 1988a,b; Thomson and Pomeranz 1991; Barker et al 1992a,b; Shatadal et al 1995). Despite the high abilities of many of these systems, the disadvantages are often the cost, complexity, and time-consuming aspects of operation. Ideally, an instrument that performs classification, including identification of mixtures, should be capable of such an evaluation in 5 min or less (D. Koeltzow, FGIS, *personal communication*), the time frame of existing wheat hardness instruments. Near-infrared (near-IR) spectroscopy is a possible method that meets these requirements.

Delwiche et al (1995) reported results on using near-IR reflectance spectra (1,100–2,498 nm) of bulk wheat to distinguish between pure samples of hard red spring and hard red winter wheat. Several classification algorithms were examined, including multiple linear regression (MLR), principal component analysis with Mahalanobis distance classifiers, partial least squares (PLS) regression, and artificial neural networks (ANN). Regardless of algorithm employed, identification of wheat class was successful at ≈95% accuracy on test sets consisting of 207–892 samples. Despite these high accuracies, true classification was not actually tested because the experimental design did not look at single kernels. For classification by near-IR spectra to be possible, spectral characteristics must be measured on individual kernels, thus permitting the detection of mixtures. Song et al (1995) reported successful classification accuracies (>94%) for ANN models developed from near-IR transmittance (850–1,050 nm) of single kernels. Models were developed and tested on small sample sets (96 kernels per sample × five samples × six classes), and further testing was recommended. The present study is an attempt to generalize the findings, using a larger number of unique samples and collecting spectra in the reflectance mode. The change from transmittance to reflectance was made because reflectance measurements are easier to implement in the rapid, single-kernel hardness instrument in current use by the FGIS.

MATERIALS AND METHODS

Wheat

The wheat samples used in this study (n = 318) were a subset, drawn at random, of a larger number (>3,000) collected by FGIS field offices for that agency's 1992 market survey (Table I). Between 56 and 72 samples were used for each of the U.S. wheat classes (durum excluded): hard white (HWH), hard red spring

[1]Beltsville Agricultural Research Center, Instrumentation and Sensing Laboratory, ARS, USDA, Beltsville, MD 20705.

Mention of company or trade names is for purposes of description only and does not imply endorsement by the U.S. Department of Agriculture.

[2]Corresponding author. E-mail: sdelwiche@asrr.arsusda.gov

Publication no. C-1996-0408-07R.

(HRS), hard red winter (HRW), soft red winter (SRW), and soft white (SWH). With the exception of two SWH samples, the geographical origin of each sample (by U.S. state) was known and was typical for its respective wheat class. Visual classification was performed by the FGIS Board of Appeals and Review. Ten kernels were drawn at random from each sample, examined for intactness, and visually compared to the remaining kernels. Broken, shriveled, or cracked kernels, or kernels appearing to be of another class, were discarded and replaced with kernels from the same sample; in most cases, all initially drawn kernels were accepted. Hereafter, the term *sample* refers to the 10 kernels collectively, unless noted otherwise.

Reflectance Apparatus

Diode array spectrophotometer. A 275-mm focal length spectrograph (monospec-27, Jarrell Ash, Waltham, MA), equipped with a 1,024-element, dry-air purged, thermoelectrically cooled (≈0°C) silicon array detector (Tracor Northern, Middleton, WI) was used to gather reflectance data in the visible to short-wavelength near-IR region (537–993 nm). A single grating, blazed at 600 nm with a groove density of 100 lines per millimeter, was used to disperse light onto the array element. Operating in the second order of refraction, the wavelength range (element 1 to element 1,024) was 537–993 nm, although the actual range used in nearly all subsequent analyses was 551–750 nm. Emission lines from a mercury lamp were used to establish the wavelength scale. Light from a tungsten halogen lamp was focused and conveyed through a bifurcated glass fiber-optic cable (Fig. 1). Incoming light to a kernel was contained within a center core (3.5 mm diameter), while light reflected from the kernel was captured in an annular region (4 mm i.d., 8 mm o.d.) of the fiber bundle assembly. The glass fibers that comprised the annular region were reformed into a rectangular shape (2 × 12 mm) at the cable's opposite end. The light that exited this end was then transmitted through a 550-nm long-pass interference filter and a 50-μm slit located at the entrance of the spectrograph. "Dark current" readings were collected periodically by blocking all light into the spectrograph and measuring the response of each array element.

Each kernel was placed, crease side up, on a black velvet disk situated in the bottom of a 10-mm diameter, 25-mm deep well bored into a 57-mm diameter black delrin rod. With the fiber optic probe fully inserted, a gap of 6 mm existed between the probe face and the black velvet disk. The exposure time was set to 125 msec per cycle, and the number of cycles was four. Before the spectrum of each kernel was collected, a reference spectrum was collected on a Teflon surface that had a well the size of the fiber optic probe. After each cycle, element voltages were digitized and temporarily stored in computer memory. At the end of the fourth cycle, the cycle readings for each element were averaged, the dark current signal subtracted, and the result divided by the corresponding value determined for the reference material. These resultant values were stored to a computer file for eventual classification modeling.

Samples were initially grouped in batches of five, with each batch containing all wheat classes. Within a batch, the samples were randomly ordered, although, within each sample, all 10 kernels were scanned in succession. This was done to eliminate the possible sensitivity of discriminant models to slight daily changes in instrument conditions. Such changes could cause classification to occur artificially rather than by the differences in color and intrinsic properties of the kernels. When the smallest represented class (HWH, $n = 56$) was depleted, grouping was performed in batches of four, and so on until only the class with the largest number of samples (HRW, $n = 72$) remained.

Near-infrared reflectance spectrophotometer. A custom-built reflectance assembly (Fig. 2) was used in conjunction with a scanning monochromator (model 6500, NIRSystems Inc., Silver Spring, MD) to gather log(1/*R*) data from 1,100 to 2,498 nm. Radiation from the exit slit of the monochromator was conveyed through a 7-mm diameter low-OH quartz fiber bundle that terminated 25 mm directly above the end of a vertically mounted 6.3-mm (1/4-in.) o.d. stainless steel tube. Encircling the fiber bundle terminus were six equally spaced 10-mm square lead sulfide detectors oriented 45° from the vertical. Both fiber bundle terminus and detector assembly were enclosed in a light-blocking metal housing. Each wheat kernel was placed, crease side up, on the upper end of the tube. Near the lower end, the tube passed through a stationary plate by use of a bearing, and by means of a gear mounted on the opposite end of the tube, a direct-current motor was used to spin the tube at 1.1 revolutions per second (a speed not in phase with the oscillatory motion of the spectrometer's grating). The intention of revolving the kernel was to minimize the effect of kernel orientation, assuming that, despite the consistent placement of the kernels on the tube, spectral variability due to kernel placement would be noticeable if the tube was stationary. The tube was enclosed in a cylindrical cavity 75-mm long by 60-mm diameter, which blocked out ambient light. All surfaces inside the cavity, including the tube, were sprayed with a flat black paint. Energy readings from the kernel were referenced to those from a pressed gray disk of carbon black and polytetrafluoroethylene (PTFE) (in 1:200 portions by weight), mounted at approximately the same elevation as the ker-

TABLE I
Summary of Descriptive Information on Wheat Samples[a]

Class[b]	*n*	Origin by U.S. State[c] (*n*)	Protein Content, % (12% mb)[d]		Near-Infrared (Near-IR) Hardness[e]		Kernel Weight (mg)		
			Mean	SD	Mean	SD	Mean	SD[f]	SD[g]
HWH	56	CA (48), WA(8)	12.5	1.02	63.6	8.6	41.8	3.4	10.7
HRS	62	MT (31), MN (16), ND (15)	14.1	0.98	88.1	7.4	33.2	3.0	8.0
HRW	72	TX (34), KS (19), OK (16), MT (1), ND (1), OR (1)	12.3	1.09	61.9	9.6	31.6	2.4	7.2
SRW	58	MO (23), KS (15), AL (8), VA (6), GA (4), LA (1), TN (1)	10.1	0.59	18.3	4.4	31.2	2.0	8.2
SWH	70	WA (38), OR (19), MI (11), UN (2)	10.3	1.48	24.7	6.1	38.1	4.5	10.6

[a] Values in three columns to the right apply to 10 randomly selected kernels per sample. Values in four center columns are based on subsamples of the original portions collected and measured by the Federal Grain Inspection Service.
[b] HWH = hard white, HRS = hard red spring, HRW = hard red winter, SRW = soft red winter, SWH = soft white.
[c] AL = Alabama, CA = California, GA = Georgia, KS = Kansas, LA = Louisiana, MI = Michigan, MN = Minnesota, MO = Missouri, MT = Montana, ND = North Dakota, OK = Oklahoma, OR = Oregon, TN = Tennessee, TX = Texas, UN = Unknown, VA = Virginia, WA = Washington.
[d] Protein content by near-IR (calibrated to Kjeldahl N × 5.7).
[e] AACC Approved Method 39-70A.
[f] Standard deviation of 10-kernel averages.
[g] Standard deviation of individual kernels.

nel and slid into alignment with the fiber bundle terminus during reference scans. The purpose of doping the PTFE with carbon black was to maximize the dynamic range of each kernel's transformed spectral data. A reference scan was collected and stored to computer memory before each kernel was scanned. The spectrum of a kernel was the average of 32 successive scans (i.e., grating oscillations), altogether taking ≈20 sec per kernel. Kernel spectral data were stored as log(1/*R*) to a computer file. Kernels were scanned in the order established by the diode array spectrophotometer procedure.

Classification Modeling

Kernels were divided into calibration and validation sets of equal size, with odd-numbered kernels used for developing the discriminant models and even-numbered kernels reserved for testing those models. Two-class and five-class models were developed, although the five-class models were structured as a cascade of binary comparisons similar to that of the two-class models. The two-class models examined were: HRW vs. HRS, HRW vs. SRW, HWH vs. HRW, HWH vs. SWH, and SRW vs. SWH. Separate models were developed for the diode array data (551–751 or 555–923 nm, the range used dependent on the mathematical form of the discriminant function) and the scanning monochromator data (1,120–2,476 nm). Hereafter, diode array data is referred to as *Vis* (visible but also including the short-wavelength near-IR) and scanning monochromator data as *NIR*.

For the five-class model (structure shown in Fig. 3), a first-level model was developed to discriminate white kernels (HWH and SWH) from red kernels (HRS, HRW, and SRW). At the next level, two distinct models were developed, one to discriminate between HWH and SWH, the other to discriminate between hard red (HRW and HRS combined) and SRW. The third and lowest level was used to decide whether hard red kernels were HRW or HRS.

With the exception of one five-class model that used MLR, all discriminant functions were based on PLS analysis. Our findings in a previous study on bulk wheat (Delwiche et al 1995) demonstrated that the PLS method produced classification accuracies nearly equivalent to those of more computationally intense methods such as ANN, while producing higher accuracies than a more traditional classification design based on calculation of Mahalanobis distances on principal components (MDPC). Preliminary studies (results not shown) using the present sample set indicated that the ANN and MDPC procedures on a single-kernel basis yielded equivalent and lower classification accuracies, respectively, than the PLS accuracies reported here.

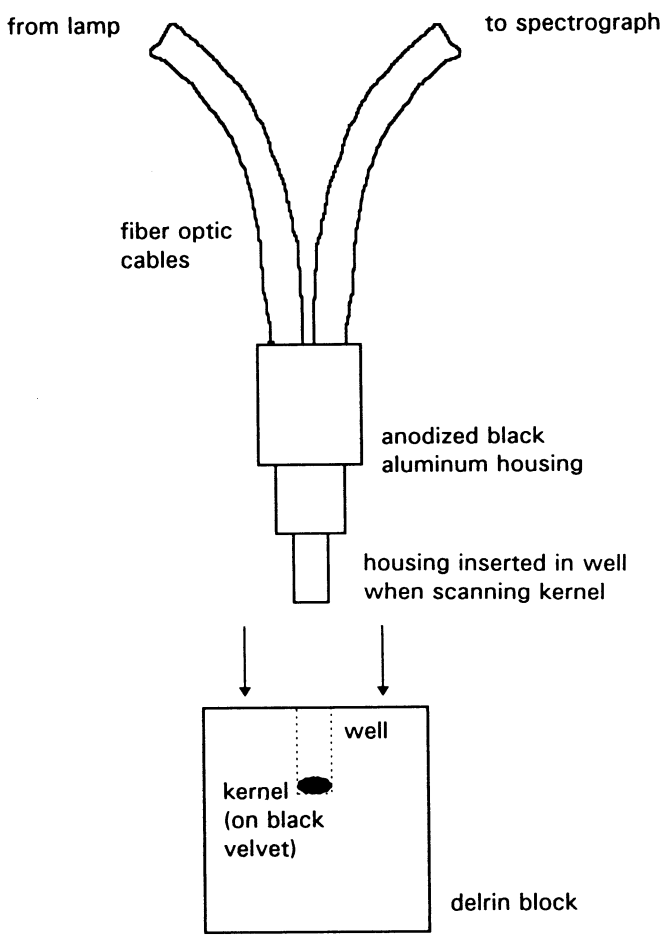

Fig. 1. Sample presentation design for the visible and short-wavelength near-IR region diode array spectrophotometer.

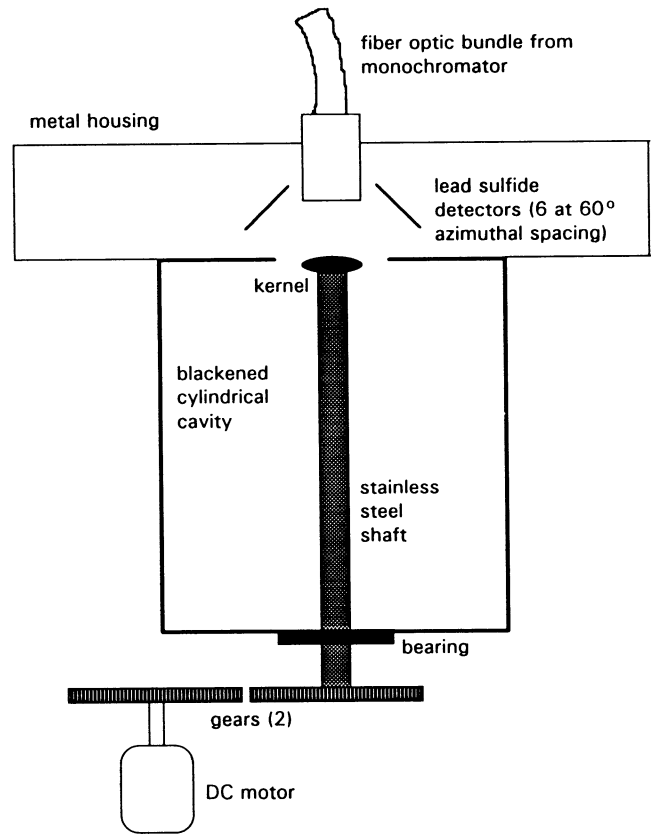

Fig. 2. Sample presentation design for the near-infrared reflectance spectrophotometer.

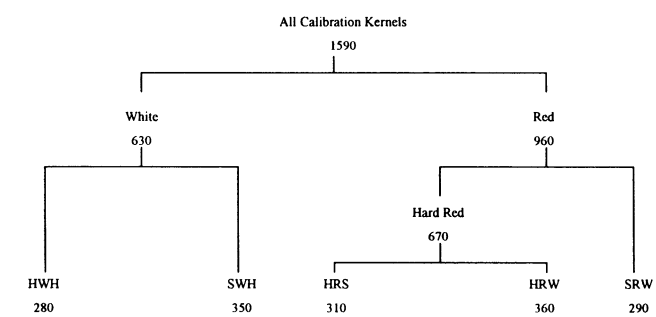

Fig. 3. Cascade structure of calibration kernels (n = 1,590) used to develop five-class models in either partial least squares or multiple linear regression analyses. The number of calibration kernels available for each stage is indicated. (HWH = hard white, SWH = soft white, HRS = hard red spring, HRW = hard red winter, SRW = soft red winter.)

The one model based on MLR was developed to allow a direct comparison to a PLS model of similar sample structure. If classification accuracies were comparable between the two methods, the hardware design for an MLR-based classification system, conceivably consisting of fixed diodes or filters, would be less expensive and therefore preferable to that of a PLS-based system. The purpose for basing all other reported classification results on PLS models was to demonstrate the upper limit of accuracy that could conceivably be attained by proper selection of wavelengths in MLR models.

Mathematics of PLS are well developed (Lindberg et al 1983) and are therefore omitted. Likewise, the MLR procedure on spectral data is thoroughly described in Hruschka (1987). The only spectral pretreatments performed before application of either procedure are as follows. To each spectrum of the Vis data, a running mean smooth (Hruschka, 1987) with a half-width, or gap, of 10 points was applied, followed by a log(1/*R*) transformation, a three-point central second difference (gap = 20 points), and the keeping of every fourth value (discarding the other three values). For the NIR data, each log(1/*R*) spectrum received a running mean smooth (gap = 1 point), whereupon every third value was kept.

For either procedure and for any model, all kernels of one group (wheat class) were assigned a constant value (0.0); the other group's kernels were assigned a different value (1.0). The MLR procedure consisted of a stepwise search of the best combination of five wavelengths in the equation,

$$\text{class} = k_0 + k_1 f_1 + \ldots + k_5 f_5,$$

that would produce the highest value for the multiple coefficient of determination (R^2), where f_is are spectral values and k_is are constants.

During model testing, a kernel was deemed to be correctly categorized if the predicted value lay on the same side of the midpoint (0.5 in the present example) as its assigned value. In the five-class cascade structure, only correctly categorized kernels from one level's model were allowed to be tested at the next level, such that in the end, the overall classification accuracy for each wheat class could be determined.

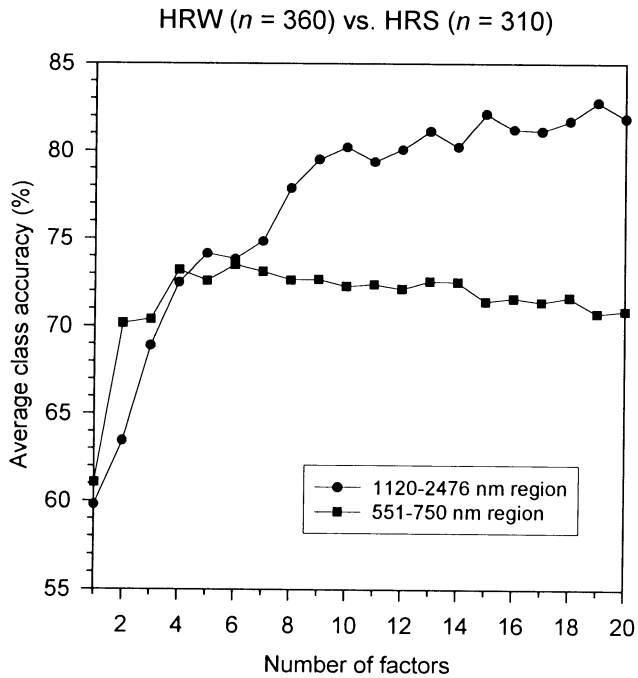

Fig. 4. Classification accuracy as a function of the number of partial least squares factors for the two-class comparison, hard red winter (HRW) vs. hard red spring (HRS). Classification models were developed on odd-numbered spectra and tested on even-numbered spectra. Two models are shown: one for the 1,120- to 2,476-nm region spectra and one for the 551- to 750-nm region spectra. Accuracies plotted above apply to the test set.

NIR Model Repeatability

One sample from each wheat class was arbitrarily chosen for repeatability trials. From the five even-numbered kernels (the test kernels), the smallest, largest, and an intermediate size kernel were each scanned 25 times on the reflectance spectrophotometer. Between scans, the kernel was removed from its perch and reseated in the same position. This was performed to simulate the variability in a classification model that could be caused by a near-perfect kernel placement device. All appropriate two-class models developed from the NIR data were subsequently applied to the 25 spectra of each repeatability kernel, with repeatabilities recorded as the number of times the kernel was correctly classified.

RESULTS AND DISCUSSION

Values for protein content, near-IR hardness, and kernel weight of the samples used in this study are presented in Table I. As is often typical, HRS wheats were highest in both protein content and near-IR hardness, followed by, in decreasing order, HWH, HRW, SWH, and SRW. As shown previously (Delwiche and Norris 1993), protein and hardness when used together to form a discriminant function are not sufficient for differentiating two similar classes such as HRS and HRW. Rather, near-IR spectral data, until now evaluated only on ground (Delwiche and Norris 1993) and bulk (Delwiche et al 1995) samples, produced high (≈95%) accuracies for differentiation. Further, at the level of individual kernels, the wide variability in kernel-to-kernel protein content (with a typical standard deviation of 1.4–2.4 percentage

TABLE II
Summary of Accuracies of Two-Class Partial Least Squares (PLS) Models

Class$_1$ vs. Class$_2$[a]	n_1	n_2	Wavelength Region[b]	Correctly Classified, %		
				Class$_1$	Class$_2$	Avg$_{1\&2}$
HRW vs. HRS	360	310	NIR	87.2	73.2	80.22
			Vis	81.4	64.8	73.11
HRW vs. SRW	360	290	NIR	83.3	73.4	78.39
			Vis	78.9	66.6	72.72
HWH vs. HRW	280	360	NIR	91.4	93.3	92.38
			Vis	98.6	99.4	99.01
HWH vs. SWH	280	350	NIR	89.6	92.9	91.25
			Vis	80.4	81.4	80.89
SRW vs. SWH	290	350	NIR	82.1	84.9	83.46
			Vis	100.0	97.7	98.86

[a] HRW = hard red winter, HRS = hard red spring, SRW = soft red winter, HWH = hard white, SWH = soft white.
[b] NIR = 1,120–2,476 nm range [= 227 log(1/R) points], 10 PLS factors; Vis = 551–750 nm range (= 113 d2(log(1/R)) points), 7 PLS factors.

units per sample [Delwiche 1995]) prohibits accurate classification models that are solely reliant on protein content.

Two-class model classification accuracies are summarized in Table II. All PLS models developed from the Vis data were based on seven factors, whereas 10 factors were used for such models developed from the NIR data. These values were selected after careful observation of the influence of the number of PLS factors on the test-set classification accuracy, as shown for one class pair (HRW vs. HRS) in Figure 4. Although model accuracy for this pair continued to increase beyond 10 factors for the NIR data, it did so at a much smaller rate, and the increase was probably due to overfitting caused by the likeness in calibration and test kernels. Whereas four factors produced accuracy equivalent to that of seven factors for the Vis data of the current example, in tests of accuracy vs. number of factors for other two-class classification pairs (not shown), seven factors appeared to be the best compromise between model accuracy and overfitting to the closed population.

As seen in Table II, the highest classification accuracies occurred when red wheats were compared to white wheats (HWH vs. HRW, SRW vs. SWH), using the models based on the Vis data. Pigmentation (principally caused by carotenoids and flavones, as shown by Kruger and Reed [1988]) was sufficiently different in the two types of kernels over the 551- to 750-nm region (yellow-green to very dark red barely visible to the human eye) to be measurable by the diode array spectrophotometer. In the SRW vs. SWH model based on the Vis region, all soft red kernels and nearly 98% of the soft white kernels in the test set were correctly classified. In contrast, classification accuracies were no larger than 85% for these same wheat classes when the NIR data were used. However, when classification models involved classes of the same color (HRW vs. HRS, HRW vs. SRW, or HWH vs. SWH), the NIR data produced models that averaged higher accuracies than those based on Vis data. For instance, the average accuracy of the NIR HWH vs. SWH model was more than 10 percentage points greater than that of the Vis model of the same two classes. Presumably, differences in protein content between the two classes (Table I), a constituent easily measured in the traditional near-IR region (1,000–2,500 nm), aided in improving classification accuracy beyond that attainable from the lower wavelength Vis region. Hardness, through association with vitreousness, affects the level of short-wavelength near-IR radiation transmitted through single kernels (Delwiche 1993). Therefore, it is assumed that hardness also influenced the single-kernel reflectance spectra, which then aided in the discrimination of hard and soft kernels.

Based on the results of the two-class tests, both spectral regions were chosen for the five-class models, with the nature of the comparison (white vs. red, hard vs. soft) determining which wavelength region to use. At the level of the first comparison, white vs. red, the Vis data were used. All subsequent comparisons were based on the NIR data.

The five-class PLS and MLR model test results are summarized in Figures 5 and 6, respectively. For the red vs. white comparison, kernels were correctly categorized, on average, 98.2 and 97.4% of the time for the PLS and MLR models, respectively. Such rates are comparable to the red vs. white models involving two wheat classes (Table II). Likewise, when the correctly categorized white kernels were subsequently tested with the preexisting HWH vs. SWH PLS model, the rate of correct assignment (89.3% for HWH, 93.6% for SWH) was nearly identical to that obtained when all possible test kernels of these classes were tested (Table II). The overall classification rates of 86.1% for HWH and 92.0% for SWH, which are smaller than their two-class counterparts, reflect the fact that two binary classification tests were applied in succession, with each prone to some inaccuracy.

Whereas white kernels required only one additional test beyond the red vs. white test, red kernels, consisting of three possible classes, required as many as two tests, depending on whether a kernel was categorized as hard (two tests) or soft (one test). It is interesting to note that all SRW test kernels were identified as red at the primary stage of the five-class PLS model, such that the diminishment in classification accuracy to 65.5% for SRW was the result of the secondary test (hard red vs. soft red). For the HRW and HRS classes, the greatest losses in classification accuracies occurred at the third stage (HRW vs. HRS), such that the

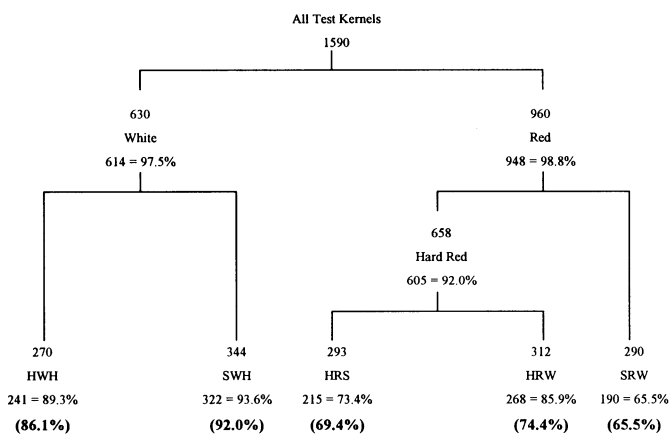

Fig. 5. Results of classification of all test kernels using two-group partial least squares (PLS) models, starting with white vs. red categorization (indicated by uppermost branch) and proceeding with further testing of correctly categorized samples, until the appropriate test for wheat class occurred. A seven-factor PLS model on 551- to 750-nm data was used for the white vs. red test; 10-factor PLS models on 1,120- to 2,476-nm data were used for all other tests. Numbers of kernels correctly classified at each stage appear below each category, together with the corresponding percentage relative to the number available for classification at that stage (appearing above each category). Larger bold-face percentages are relative to the total number of kernels in each class.

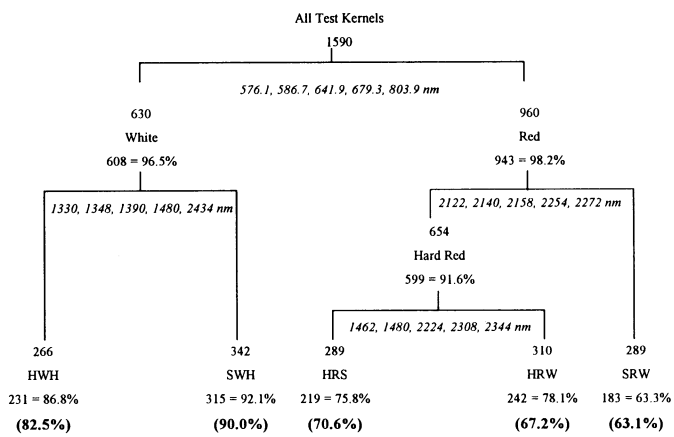

Fig. 6. Results of classification of all test kernels using two-group multiple linear regression (MLR) models, starting with white vs. red categorization (indicated by uppermost branch) and proceeding with further testing of correctly categorized samples, until the appropriate test for wheat class occurred. Five-wavelength MLR models were used at all stages, where the wavelengths (identified in italics) were selected from stepwise regression on the calibration kernels. Visible region spectral data (555–923 nm search region) were used for the white vs. red test; near-infrared region spectral data (1,120–2,476 nm search region) were used for all other tests. Numbers of kernels correctly classified at each stage appear below each category, together with the corresponding percentage relative to the number available for classification at that stage (appearing above each category). Larger bold-face percentages are relative to the total number of kernels in each class.

overall classification rate for either class was <75%. Similar results were obtained for the five-class MLR model (Fig. 6), with four classes demonstrating a decrease in overall classification accuracy and one class an increase in accuracy when compared to the PLS model. Although the total number of correctly classified kernels was higher for the PLS model than the MLR model (1,236 and 1,190, respectively), the mathematical simplicity of the MLR model and a less complicated hardware design for measuring reflectance at fixed wavelengths may outweigh this loss in accuracy.

As stated earlier, the wavelengths used in each MLR equation were those that produced the highest coefficient of determination during stepwise regression on the calibration set. While a complete interpretation of why these wavelengths were selected is not possible, some insight can be drawn based on published knowledge of near-IR absorbers (Williams and Norris 1987). Several other combinations of wavelengths (throughout the 555- to 923-nm search region) for the red vs. white separation yielded performances in calibration nearly equivalent to those of the combination shown in Figure 6. This confirms that categorization in red or white groups is achieved because of the spectrally broad pigmentation differences between the two groups. At the HWH vs. SWH classification level, class differences in protein and oil were spectrally sensed at 1,480 and 1,390 nm, respectively. The remaining three wavelengths (1,330, 1,348, and 2,434 nm) of this MLR equation did not directly correspond to any of the broadly defined food constituents (protein, starch, oil, water, and cellulose) and therefore could have represented an interaction of two or more of these constituents. For hard red vs. soft red categorization, the wavelengths 2,158, 2,140, and 2,272 nm corresponded to near-IR absorption bands for protein, oil, and cellulose, respectively. The other two wavelengths (2,122 and 2,254 nm) were also probably responding to constituent interactions. At the HRW vs. HRS classification level, one of the selected wavelengths attributed to protein (1,480 nm), was also selected in the HWH vs. SWH classification. Of the four remaining wavelengths, one did not correspond to a constituent absorption band. Two wavelengths (2,308 and 2,344 nm) corresponded to absorption bands for oil, and the last wavelength (1,462 nm) appeared to be sensitive to class differences in cellulosic compounds. Noticeably absent from all MLR equations was a wavelength directly attributable to water absorption. We previously observed this phenomenon and postulated that water-matrix interactions were useful in the spectral detection of class differences, but water alone was not (Delwiche and Norris 1993).

TABLE III
Repeatability Performance of Two-Class Models

Actual Class[a]	Two-Class Model Applied[b]	Out of 25 Scans, Number of Times Correctly Classified[c]		
		Small Kernel[d]	Medium Kernel	Large Kernel
HRW	HRW vs. HRS	25***	6**	25***
	HRW vs. SRW	22***	25***	25***
	HWH vs. HRW	25***	24***	25***
HWH	HWH vs. HRW	19**	25***	25***
	HWH vs. SWH	2***	25***	25***
SRW	HRW vs. SRW	19**	25***	25***
	SRW vs. SWH	21***	25***	25***
SWH	HWH vs. SWH	25***	25***	25***
	SRW vs. SWH	21***	21***	13
HRS	HRW vs. HRS	0***	2***	7*

[a] HRW = hard red winter, HRS = hard red spring, SRW = soft red winter, HWH = hard white, SWH = soft white.
[b] Ten-factor partial least squares models developed from 1,120–2,476 nm wavelength region.
[c] Chi-square test rejected null hypothesis (H_0) that a kernel would be misclassified the same number of times as correctly classified at the following significance levels: * = P <0.05; ** = P <0.01; *** = P <0.005.
[d] One sample selected from each wheat class, from which a small, medium, and large kernel was chosen from test set kernels.

Repeatability Trials

Repeatability readings, reported as the number of times out of 25 scans that a kernel was correctly classified by a suitable two-class PLS model, are summarized in Table III. Repeatability was interpreted as the tendency for a model to classify a trial of 25 scans into one wheat class, regardless of whether this was obtained by correct (ideal number = 25) or incorrect (ideal number = 0) assignment. In 17 of the 30 trials, the models correctly classified all 25 scans. One trial, a HRW vs. HRS model applied to a small HRS kernel, resulted in the misclassification of all scans. Thus, full repeatability was attained for 18 of 30 trials. A chi-square test was performed on each trial, with the null hypothesis being that for each scan the probability of misclassification was equal to that of correct classification. In all except one trial, repeatability was deemed significant at the $P < 0.05$, $P < 0.01$, or $P < 0.005$ level. Only in the trial involving an SRW vs. SWH model applied to the large SWH kernel was repeatability not significant. The general trend was for repeatability to improve as the kernel size increased. Hence, potentially deleterious effects on classification accuracy caused by kernel presentation did not seem apparent.

CONCLUSIONS

Classification of individual kernels of wheat by PLS or MLR models developed from near-infrared reflectance spectra is possible. Model accuracy, defined as the proportion of kernels from a known wheat class correctly assigned during model application, is dependent on the spectral region (visible and short-wavelength near-IR vs. near-IR), the number of wheat classes built into a model, and which classes they are. The highest accuracies (>98%) occur for models that have been developed to distinguish red from white wheats. A rapid (≈0.5 sec per kernel) diode array instrument, operating over a 555- to 923-nm range, was demonstrated to be effective in sensing the differences in pigmentation between these two groups. Such accuracies should be acceptable for official grain grading purposes. However, classification within the red or white wheats (e.g., HRW vs SRW) declines to an average accuracy of 78–91%, using spectral data in the near-IR region (1,120–2,476 nm). Although useful for screening purposes, accuracies of two-class-within-one-color models are not sufficiently high for official grading. Model repeatability, defined as the degree of consistency on repeated measurements of the same kernel, is excellent for all two-class models.

Model accuracy is diminished when more classes are included. For a five-class model developed from a set of 318 unique wheat samples, SRW wheats were most difficult to correctly classify (accuracy 63–66%), followed by HRS (69–71%), HRW (67–74%), HWH (82–86%), and SWH (90–92%). These accuracies suggest that a multiclass wheat classification system that is based on spectral data alone using a binary tree technique is probably not sufficient for official inspection. However, when used in conjunction with other single-kernel measurements related to the physical measurement of hardness, near-IR reflectance data may allow for the development of an acceptable, rapid, and objective wheat classification system. Such a combination is the target for future research.

ACKNOWLEDGMENTS

We thank NIRSystems, Inc. (Silver Spring, MD) for detectors and amplification circuitry and J. Shaffer (ARS, Beltsville) for data collection.

LITERATURE CITED

BARKER, D. A., VUORI, T. A., HEGEDUS, M. R., and MYERS, D. G. 1992a. The use of ray parameters for the discrimination of Australian wheat varieties. Plant Varieties Seeds 5:35-45.

BARKER, D. A., VUORI, T. A., and MYERS, D. G. 1992b. The use of slice and aspect ratio parameters for the discrimination of Australian wheat varieties. Plant Varieties Seeds 5:47-52.

BIETZ, J. A., BURNOUF, T., COBB, L. A., and WALL, J. S. 1984. Wheat varietal identification and genetic analysis by reversed-phase high-performance liquid chromatography. Cereal Chem. 61:129-135.

BIETZ, J. A., and COBB, L. A. 1985. Improved procedures for rapid wheat varietal identification by reversed-phase high-performance liquid chromatography of gliadin. Cereal Chem. 61:332.

DELWICHE, S. R. 1993. Measurement of single-kernel wheat hardness using near-infrared transmittance. Trans. ASAE 36:1431-1437.

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

DELWICHE, S. R., and NORRIS, K. H. 1993. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. Cereal Chem. 70:29-35.

DELWICHE, S. R., CHEN, Y. R., and HRUSCHKA, W. R. 1995. Differentiation of hard red wheat by near-infrared analysis of bulk samples. Cereal Chem. 72:243-247.

ENDO, S., OKADA, K., NAGAO, S., and D'APPOLONIA, B. L. 1990. Quality characteristics of hard red spring and winter wheats. I. Differentiation by reversed-phase high-performance liquid chromatography and milling properties. Cereal Chem. 67:480-485.

HRUSCHKA, W. R. 1987. Data analysis: Wavelength selection methods. Pages 35-55 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.

HUEBNER, F. R., NELSEN, T. C., and BIETZ, J. A. 1995. Differences among gliadins from spring and winter wheat cultivars. Cereal Chem. 72:341-343.

IRVING, D. W., FULCHER, R. G., BEAN, M. M., and SAUNDERS, R. M. 1989. Differentiation of wheat based on fluorescence, hardness, and protein. Cereal Chem. 66:471-477.

KRUGER, J. E., and REED, G. 1988. Enzymes and color. Pages 441-500 in: Wheat: Chemistry and Technology, 3rd ed. Vol. I. Y. Pomeranz, ed. Am. Assoc. Cereal Chem.: St. Paul, MN.

LINDBERG, W., PERSSON, J. A., and WOLD, S. 1983. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate. Anal. Chem. 55:643-648.

MARCHYLO, B. A., HATCHER, D. W., and KRUGER, J. E. 1988. Identification of wheat cultivars by reversed-phase high-performance liquid chromatography of storage proteins. Cereal Chem. 65:28-40.

NEUMAN, M., SAPIRSTEIN, H. D., SHWEDYK, E., and BUSHUK, W. 1987. Discrimination of wheat class and variety by digital image analysis of whole grain samples. J. Cereal Sci. 6:125-132.

ROYAL AUSTRALIAN CHEMICAL INSTITUTE. 1988. Official testing method E2. Electrophoretic identification of cereal varieties. The Institute: Melbourne, NSW.

SAPIRSTEIN, H. D., NEUMAN, M., WRIGHT, E. H., SHWEDYK, E., and BUSHUK, W. 1987. An instrumental system for cereal grain classification using digital image analysis. J. Cereal Sci. 6:3-14.

SHATADAL, P., JAYAS, D. S., and BULLEY, N. R. 1995. Digital image analysis for software separation and classification of touching grains. II. Classification. Trans. ASAE 38:645-649.

SONG, H., DELWICHE, S. R., and CHEN, Y. R. 1995. Neural network classification of wheat using single kernel near-infrared transmittance spectra. Optical Eng. 34:2927-2934.

SYMONS, S. J., and FULCHER, R. G. 1988a. Determination of wheat kernel morphological variation by digital image analysis: I. Variation in Eastern Canadian milling quality wheats. J. Cereal Sci. 8:211-218.

SYMONS, S. J., and FULCHER, R. G. 1988b. Determination of wheat kernel morphological variation by digital image analysis: II. Variation in cultivars of soft white winter wheats. J. Cereal Sci. 8:219-229.

THOMSON, W. H., and POMERANZ, Y. 1991. Classification of wheat kernels using three-dimensional image analysis. Cereal Chem. 68:357-361.

WILLIAMS, P. C., and NORRIS, K. H. 1987. Qualitative applications of near-infrared reflectance spectroscopy. Pages 241-246 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.

WRIGLEY, C. W., AUTRAN, J. C., and BUSHUK, W. 1982. Identification of cereal varieties by gel electrophoresis of the grain proteins. Pages 211-259 in: Advances in Cereal Science and Technology, Vol. 5. Am. Assoc. Cereal Chem.: St. Paul, MN.

ZAYAS, I., POMERANZ, Y., and LAI, F. S. 1985. Discrimination between Arthur and Arkan wheats by image analysis. Cereal Chem. 62:478-480.

ZAYAS, I., LAI, F. S., and POMERANZ, Y. 1986. Discrimination between wheat classes and varieties by image analysis. Cereal Chem. 63:52-56.

[Received September 15, 1995. Accepted January 18, 1996.]