Classification of Hard Red Wheat by Near-Infrared Diffuse Reflectance Spectroscopy

STEPHEN R. DELWICHE[1] and KARL H. NORRIS[2]

ABSTRACT

Cereal Chem. 70(1):29–35

Various forms of discriminant analysis models have been developed and tested for distinguishing two classes of wheat—hard red winter and hard red spring. Near-infrared diffuse reflectance (NIR) spectroscopy was used to measure the intrinsic properties of ground samples of hard red winter and spring wheats grown during the 1987, 1988, 1989, and 1990 crop years, of which 100 samples from each of the first three years formed the calibration set for each model. Discriminant functions were developed by using the following parameters: NIR-predicted protein content (adjusted to 12% moisture), NIR-predicted hardness, NIR protein and NIR hardness, and the scores from principal component analysis (PCA) of full-range (1,100–2,498 nm) NIR spectra. Each function was tested on 1,325 samples (excluded from training of the models) from the 1987–1989 crop years and on 678 samples from the 1990 crop year, all of known class. Model performance, expressed as the percent of misclassified samples for each year and class, was poorest for the one-parameter models, which often had misclassification rates in excess of 25%. A five-factor PCA model was the most accurate, with an average misclassification rate of 5% for 1987, 1988, and 1989 samples. However, the misclassification rate of the PCA model rose to 8% for the 1990 samples, suggesting that model accuracy is reduced when samples grown during years excluded from calibration, such as from a new year's crop, are classified. Examination of the principal component factors indicates that hardness, protein level, and the interaction of water with protein and other constituents within wheat are responsible for correct NIR-based classification.

Wheat grown in the United States is divided into eight classes: hard red winter (HRW), hard red spring (HRS), soft red winter (SRW), durum, white (soft white winter [SWW] and club), hard white winter (HWW), mixed, and unclassed. Wheat classification is a component of the U.S. Standards and has traditionally been determined according to kernel morphology (e.g., size, shape, color, crease appearance). The price that wheat is traded in the United States is customarily based on class. The cost also may be a function of the protein content (adjusted to 12% moisture) of a wheat lot. The U.S. wheat system is different from that of Canada, Australia, and France in that wheat in the United States usually is not segregated by cultivar (Office of Technology Assessment 1989). Further, the class of wheat is often difficult to track through the market system, particularly in the case of the two dominant bread-making classes, HRW and HRS. Historically, cereal lot uniformity was maintained by human inspection of samples drawn from lots. Much of this success could be attributed to the limited number of cultivars available and to the inspector's knowledge of the geographical origin of growth, because classes generally were segregated by location. Typically, fewer than 10 cultivars per class existed, allowing the inspector to become familiar with each cultivar. Today, classification is becoming increasingly more difficult because of the presence of many more cultivars and the overlapping of growing regions. Crossbreeding between cultivars belonging to two or more classes is becoming increasingly prevalent. Of particular concern is the discrimination between HRW and HRS wheats because these are the two most common bread-making classes grown and traded (domestic and export markets) in the United States, and their morphological appearance is very similar.

Methods for objective classification include polyacrylamide gel electrophoresis, reversed-phase high-performance liquid chromatography, and digital image analysis of kernels. The first two methods are widely used for the examination of wheat endosperm proteins, and in addition to possessing the capability of distinguishing varietal differences (Wrigley et al 1982, Bietz et al 1984, Marchylo et al 1988), they have been used for classification (Endo et al 1990a,b). When digital image analysis is applied to intact kernels (Zayas et al 1985, 1986; Neuman 1987; Symons and Fulcher 1988a,b; Chen et al 1989; Thomson and Pomeranz 1991), size and shape differences among wheat classes are exploited to distinguish wheat classes and varieties. Generally, several geometric parameters are used in multivariate analysis models, yielding classification rates more than 90% accurate. The disadvantages of the existing classification methods include the length of time per analysis, the required level of operator skill, and the equipment cost.

Relying on differences between the fluorescence properties of the pericarp, aleurone, and endosperm and measuring protein content and near-infrared reflectance (NIR) hardness, Irving et al (1989) obtained near-perfect classification among five classes (HRW, HRS, SRW, SWW, and durum) and one subclass (club) of wheat. The authors noted that despite the high accuracy of their models, the application of such instrumentation at inspection sites was unlikely because of the complexity of the instrumentation and analysis. However, because these properties are intrinsic to the wheat kernel, and intrinsic properties (e.g., protein, moisture, hardness) often are easily measurable through the use of NIR spectroscopy, such instrumentation was used in the present study. Classification by NIR spectroscopy stands as an attractive alternative to the other methods because this technique is well accepted throughout the U.S. grain industry; some terminals, mills, and bakeries already possess full-wavelength range NIR scanning spectrophotometers. The objectives of this study were to examine the ability to discriminate between the two wheat classes HRW and HRS using one or both of the two parameters, NIR-predicted

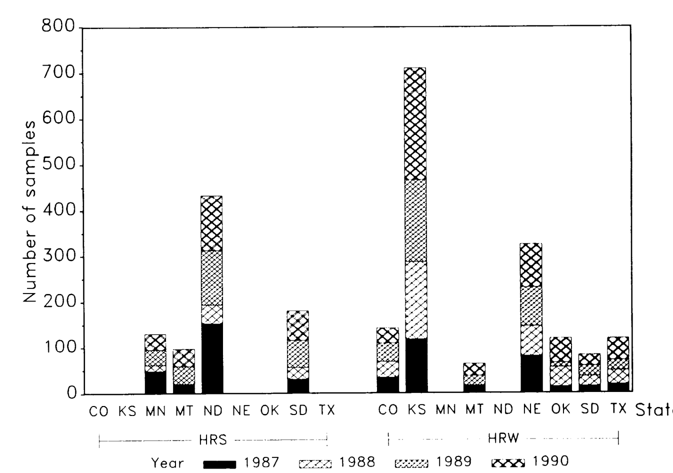

Fig. 1. Geographical origin (by state) of all wheat samples studied. HRS = hard red spring wheat; HRW = hard red winter wheat. CO = Colorado, KS = Kansas, MN = Minnesota, MT = Montana, ND = North Dakota, NE = Nebraska, OK = Oklahoma, SD = South Dakota, and TX = Texas.

[1]Beltsville Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705-2350.
[2]Retired, Beltsville Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD. Mention of company or trade name is for purpose of description only and does not imply endorsement by the U.S. Department of Agriculture.

protein content and NIR hardness, and to examine the effectiveness of principal component-based discriminant analysis applied to full-range (1,100–2,500 nm) NIR spectra of HRW and HRS wheat classes.

MATERIALS AND METHODS

Sample Collection

The samples examined in the study were those of the 1987–1990 annual crop surveys conducted by the Doty Laboratories (Kansas City, MO) of U.S.-grown HRW and HRS wheats. The yearly surveys are used for the purpose of gauging the general quality of the wheat, and the information is sold to trading, milling, and baking companies. The survey samples are a very good representation of the level of quality of the HRW and HRS wheats that are grown in the United States each year.

Each crop survey consisted of more than 600 samples of commercial stock gathered from country elevators located in a nine-state region of the central United States. Figure 1 contains a summary of the geographical (i.e., state) origin by year of the HRW and HRS classes used in the present study. The proportion of spring to winter samples was approximately 1:2, with the exception of the 1988 survey, in which the number of HRS samples was small because of an insufficiency in material. Classification was performed by Doty personnel at the time of sample collection. Correctness of classification was verified with the 1987 samples by the Federal Grain Inspection Service, Board of Appeals and Review (FGIS-BAR), the official body for classification in the United States. Discrepancies between the two sources of classification amounted to 23 of 718 samples (or 3%). The disputed samples were excluded from model development. Because verification by the FGIS-BAR was not performed on the remaining three years of data, any of the classification models that have been developed in the present study cannot be expected to attain perfect classification.

Approximately 15 g of each sample was ground by a Udy cyclone grinder (model 3010-018, Fort Collins, CO) equipped with a 1-mm screen. From this, 3–5 g was loosely packed in a nylon ring (38 mm i.d., 8 mm deep) and capped on one end with an infrared transmitting quartz window 1.27 mm thick. Diffuse reflectance readings, $\log(1/R[\lambda])$, (700 readings in 2-nm increments at $\lambda = 1{,}100$–$2{,}498$ nm) were collected with an NIR Systems (Silver Spring, MD) model 6250 spectrophotometer. A ceramic disk was used as the reflectance reference. Each sample's spectrum was the average of 50 repetitive scans.

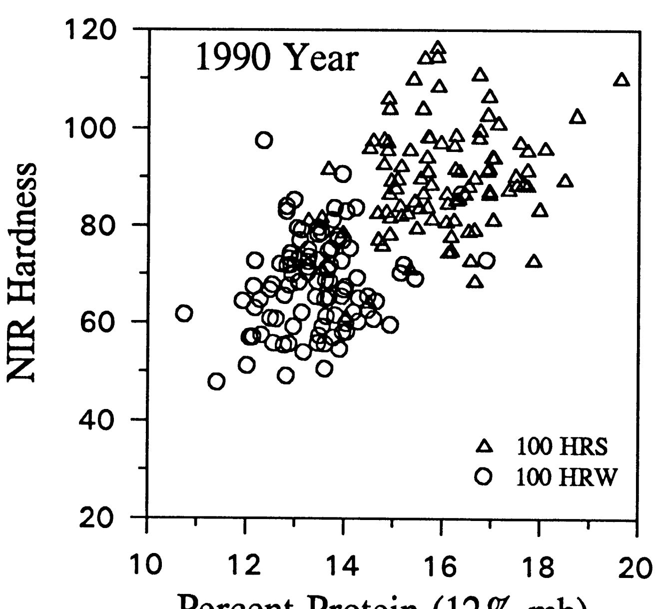

Fig. 2. Scatter plot of near-infrared reflectance (NIR) hardness vs. protein content (12% moisture basis) of 100 randomly selected samples each of hard red winter (HRW) and hard red spring (HRS) wheat from 1990.

Protein Content and NIR Hardness Determinations

Generally, spring wheats tend to have a higher protein content than do winter wheats. To some extent, the hardness of spring wheats is also greater than winter wheats (Pomeranz et al 1988, Delwiche 1991). These two observations serve as the starting point for developing objective methods for distinguishing HRW from HRS wheats. As an example, a scatter plot of NIR hardness versus NIR protein content adjusted to 12% moisture basis (mb) for a random selection of 100 samples each of 1990 HRW and HRS wheat is shown in Figure 2. HRW samples tend to be separated from HRS samples in both NIR hardness and NIR protein content, thus forming two broad clusters. The degree to which these clusters does not overlap has a direct bearing on the ability of the discriminant functions that make use of hardness, protein, or a combination of both to properly classify the wheat.

The protein content of each sample was determined from a four-wavelength spectral model that is currently used by the FGIS for official inspection. Drift in predictions due to year-to-year changes in instrument settings and grinding was minimized by standardizing the model against a set of wheat of known protein content (as analyzed by Kjeldahl assays) and applying a slope and bias adjustment to the model. The model for any given year had the form

$$\text{Protein (12\% mb)} = a_0 + a_1 \begin{bmatrix} \log(1/R)_{2,180\text{nm}} - 0.765 \log(1/R)_{2,100\text{nm}} \\ + 0.0344 \log(1/R)_{1,940\text{nm}} - 0.396 \log(1/R)_{1,680\text{nm}} \end{bmatrix} \quad (1)$$

with the values for coefficients a_0 and a_1 obtained through standardization.

NIR hardness calculations were performed according to AACC procedures (AACC 1986, Norris et al 1989). NIR hardness is based on an empirical scale and generally ranges from about 10 (very soft) to 110 (very hard). The ability of NIR reflectance spectroscopy to measure kernel hardness arises from the property that differences in particle size distributions of ground wheat occur among the various classes (Yamazaki 1972, Meppelink 1974, Williams 1979, Wu et al 1990). Such differences affect the manner in which NIR radiation impinging on the surface of ground wheat becomes diffusely reflected. Radiation striking larger particles becomes more highly absorbed than that striking the smaller particles. Thus, class-specific particle size distributions of ground wheat are measurable from the NIR spectra. Although the greatest differences in hardness occur between soft and hard classes, smaller differences also occur within the hard wheat classes, HRW and HRS. Wu et al (1990) determined that HRS flours yield a larger fraction of small (<15 μm) particles than do HRW flours, whereas the opposite trend occurs in the 24- to 30-μm fraction.

Based on the AACC (1986) standard procedure, the NIR spectrophotometer was standardized using a set of five soft and five hard cultivars obtained from FGIS. For each year, five subsamples per cultivar from the standardization set were ground and scanned. Soft and hard cultivars were assigned the NIR hardness values 25 and 75, respectively, as prescribed by the AACC. Based on the spectra of each year's standardization set, a slope and bias correction was applied to the coefficients b_0 and b_1 of the NIR hardness equation (equation 3 of Norris et al 1989)

$$\text{Hardness} = b_0 + b_1[\log(1/R)_{2,230\text{nm}} - 0.745 \log(1/R)_{1,680\text{nm}}] \quad (2).$$

The repeatability of the NIR hardness model, as measured by the average of the standard deviations of the predicted NIR hardness scores from each of the five-subsample cultivar groups, is generally 3.

Discriminant Analyses

Fifty samples from each of the two classes and from each of the 1987, 1988, and 1989 surveys (for a total of 150 HRW and 150 HRS samples) formed the calibration (i.e., training) set. Each 50-sample group was representative (in terms of NIR protein content, NIR hardness, and state of origin) of the larger set group from which it was drawn.

Two approaches were employed for developing discriminant analysis models; in one case, the models were based on the NIR protein content (equation 1) and NIR-hardness (equation 2), and in the other case, discriminant analysis models were developed using the loadings of the spectra as derived from principal component analysis (PCA) (Pearson 1901, Joliffe 1986, Devaux et al 1988). Each of these approaches is described in more detail below.

For the first case, the SAS procedure DISCRIM (SAS Institute Inc. 1988) was used. NIR protein content and NIR hardness distributions were assumed to be multivariate normal, and a parametric method consisting of a linear discriminant function was used. Separate models, based on NIR protein content, NIR hardness, and both parameters jointly, were developed for classifying the samples into HRW or HRS groups. Each model was used to classify 1) samples from the calibration set, 2) samples from 1987 through 1989 that were not part of the calibration set, and 3) samples from the 1990 year.

For the second case, the program Discriminate, an add-on program to the spectral analysis program LAB CALC (Galactic Industries Inc., Salem, NH), was used. The calibration set was the same as that used in the first case. In application, samples from the calibration set are initially expressed in terms of their principal components. The principal components reduce the dimension of the variability space from the number of wavelengths per spectrum (700 in the present case) down to a user-selected number. Generally, between one and 10 factors (i.e., eigenvectors) are selected. Each spectrum can essentially be represented as a linear combination of these factors, in which a spectrum's unique shape is a function of the coefficients (i.e., scores) applied to the factors. Once the spectra are expressed in terms of their principal components, the scores are then expressed in a normalized Mahalanobis distance space (Mahalanobis 1936, Mark and Tunnell 1985). A linear discriminant function is developed from the normalized scores.

Before Discriminate was run, all spectra were scaled to reduce sample-to-sample absorption and scattering differences caused by variation in packing density of the material in the NIR scanning cell. Scaling was performed according to

$$\hat{A}(\lambda) = [\hat{A}(\lambda_2) - \hat{A}(\lambda_1)] \frac{A(\lambda) - A(\lambda_1)}{A(\lambda_2) - A(\lambda_1)} + \hat{A}(\lambda_1) \qquad (3)$$

where $A(\lambda)$ and $\hat{A}(\lambda)$ are the unscaled and scaled spectra, respectively, and λ_1 and λ_2 are the selected wavelengths through which all spectra are required to have common value. Values for $\hat{A}(\lambda_1)$ and $\hat{A}(\lambda_2)$ were assigned as 0.0 and 0.25, respectively; wavelengths were λ_1 = 1,100 nm and λ_2 = 2,230 nm. These two wavelengths were selected because they lack strong absorbers in wheat, yet possess a moderate difference in baseline value.

Discriminate was applied separately to the 150-sample HRW and HRS groups of the previously mentioned calibration set, yielding two unique sets of factors and scores (hereafter referred to as submodels; collectively as a model). Each submodel then was applied to each sample in trials 1–3 defined in the description of the first case. The underlying presumption is that a smaller Mahalanobis distance (*MD*) occurs when the sample's spectrum is evaluated by the submodel of the class to which the sample belongs rather than that which occurs when the submodel of the other class is used. By forming the difference

$$\Delta_{\mathrm{MD}} = [MD_{\mathrm{HRS\ submodel}} - MD_{\mathrm{HRW\ submodel}}] \qquad (4)$$

HRW samples should have $\Delta_{\mathrm{MD}} > 0$ and HRS samples should have $\Delta_{\mathrm{MD}} < 0$. A sample was deemed to be correctly classified when Δ_{MD} followed this criterion. The number of factors examined ranged from two to 10.

An additional set of analyses was performed in the second case. Rather than excluding the 1990 samples from the calibration set, 50 samples each of HRW and HRS 1990 wheat were included, and all samples from 1987 were removed. Likewise, models were developed from sets that excluded all 1988 samples, then all 1989 samples.

RESULTS

Mean values and standard deviations for NIR protein content and NIR hardness of the samples, grouped according to year and class, are presented in Table I. Separate statistics are given for calibration and prediction sets for each year to emphasize the similarity in makeup of these two sets. Additionally, the statistics for all samples (calibration and prediction) within a year and class are presented in Table I. Analyses of variance (GLM procedure of SAS) performed on the combined calibration and prediction set indicated that the means of NIR protein content grouped according to wheat class, year, and the wheat class by year interaction were all significantly different ($\alpha \leq 0.001$). Within each class, Tukey's Studentized range (HSD) test indicated that all year-year comparisons of mean NIR protein content were also significantly different ($\alpha = 0.01$). For NIR hardness, the differences in the means grouped by wheat class, year, and their interaction were highly significant ($\alpha \leq 0.005$). Similarly, the HSD values indicated that all within-class year-year comparisons (with the exception of 1989–1990 HRW and 1989–1990 HRS sets) demonstrated that NIR hardness was significantly different ($\alpha = 0.01$) across any two years.

The results of the discriminant analysis model based on NIR protein content are shown in Table II. Misclassification rates on the prediction samples ranged from 0.4% for 1987 HRW to 47.4% for 1989 HRW. Because the 1987 samples from both classes had a lower NIR protein content than did the succeeding years (Table I), the discriminant model was highly successful at classifying the 1987 HRW wheat, which had the lowest average NIR protein content of any year and class. Coupled with this occurrence, however, was the relatively high (44.1%) misclassification rate for the 1987 HRS wheat. Because of the low mean NIR protein content of 1987 HRS wheat in comparison to the HRS

TABLE I
Summary of NIR[a] Protein Content and NIR Hardness of Wheat Samples, Grouped by Set (Calibration, Prediction, or All) Within Class[b] Within Crop Year (1987–1990)

				NIR Protein Content (12% moisture)		NIR Hardness	
Year	Class	Set[c]	*n*	Mean	SD	Mean	SD
1987	HRW	C	50	12.04	0.85	60.5	8.5
		P	248	12.08	0.91	60.9	8.6
		A	298	12.08	0.90	60.8	8.6
	HRS	C	50	14.86	0.82	76.7	7.1
		P	202	14.68	1.04	77.5	7.3
		A	252	14.71	1.00	77.4	7.3
1988	HRW	C	50	12.90	1.35	62.7	7.5
		P	313	13.06	1.14	63.2	7.8
		A	363	13.04	1.17	63.2	7.8
	HRS	C	50	17.08	1.08	82.6	8.1
		P	31	17.00	1.21	79.7	8.6
		A	81	17.05	1.13	81.5	8.3
1989	HRW	C	50	14.41	1.01	69.0	7.7
		P	331	14.50	1.22	67.5	8.1
		A	381	14.49	1.19	67.7	8.1
	HRS	C	50	16.51	1.57	86.3	9.2
		P	200	16.55	1.17	87.5	8.0
		A	250	16.54	1.25	87.3	8.3
1990	HRW	C	50	13.18	0.73	68.4	10.3
		P	471	13.36	0.94	68.7	10.2
		A	521	13.35	0.92	68.6	10.2
	HRS	C	50	15.80	1.28	90.1	10.9
		P	207	16.07	1.25	87.9	9.0
		A	257	16.02	1.26	88.4	9.4

[a] Near-infrared reflectance.
[b] HRW = hard red winter; HRS = hard red spring.
[c] C = calibration, P = prediction, A = all (C + P).

wheats of succeeding years, a large portion of the 1987 HRS wheat was closer in NIR protein content to the three-year mean NIR protein content for HRW wheats than it was for HRS wheats. Consequently, that portion was wrongly classified as HRW. By similar reasoning, the misclassification rate for the 1989 HRW samples was high because the winter wheats from this year contained relatively high levels of protein.

When the discriminant analysis was based on NIR hardness (Table II), misclassification rates of the prediction sets ranged from 1.5% for 1989 HRS wheat to 31.5% for 1990 HRW wheat. As in the case of NIR protein content, low misclassification rates were associated with sets having NIR hardnesses that were either HRW and lower (e.g., the 1987 set) or HRS and higher (e.g., the 1989 set) on average than the other year and class groups. Misclassification rates for the 1990 HRW and HRS groups (31.5 and 5.1%, respectively) most closely resembled those of the 1989 HRW and HRS groups (23.9 and 1.5%, respectively). This is consistent with the lack of significantly different means between these two years, as described earlier.

A summary of the misclassification rates that were produced using PCA is contained in Table III. Misclassification rates for the calibration and prediction sets are listed for the five-, six-, and eight-factor models, with these numbers of factors chosen for representation because of their relatively low misclassification rates. All three models yielded comparable misclassification rates on the 1987, 1988, and 1989 prediction samples, with the rate ranging from 0% (five-factor 1989 HRS, six-factor 1989 HRS) to 8.8% (five-factor 1989 HRW). However, when the models were applied to the 1990 samples (i.e., the year excluded from the calibration set), model performance declined as the number of factors increased. For the 1990 HRW prediction samples, the misclassification rate went from 6.6% at five factors to 16.8% at eight factors. Likewise, the misclassification rate for the 1990 HRS samples changed from 11.1 to 24.2%. Because of the better performance of the five-factor model, subsequent reported findings are based on this model.

The Δ_{MD} values for each sample yielded additional information about the performance of the PCA models. Histograms showing the distribution of Δ_{MD} for the HRW and HRS classes of the 1987 and 1988 prediction sets are displayed in Figure 3, and those of the 1989 and 1990 sets are displayed in Figure 4. Each bar represents the percentage of the population within a class and year. Two Gaussian-shaped envelopes are evident in each year's plot. Regions of overlap indicate that wheat samples from one class have been misclassified. For HRW samples, misclassification occurs when $\Delta_{MD} < 0$; the opposite is true for HRS samples.

TABLE II
Percentage of Wheat Samples Misclassified by Discriminant Analysis Models Based on NIR[a] Protein Content (12% Moisture), NIR Hardness, and the Combination of Both

		Calibration Set[b]				Prediction Set[b]			
Year	Class[c]	*n*	P	H	P + H	*n*	P	H	P + H
1987	HRW	50	0	2.0	0	248	0.4	6.0	0.4
	HRS	50	38.0	24.0	24.0	202	44.1	25.7	34.2
1988	HRW	50	8.0	8.0	6.0	313	6.7	9.3	3.8
	HRS	50	0	8.0	2.0	31	6.4	12.9	6.4
1989	HRW	50	40.0	34.0	32.0	331	47.4	23.9	27.2
	HRS	50	10.0	6.0	8.0	200	5.5	1.5	1.5
1990	HRW	...	...	...	...	521	8.1	31.5	13.8
	HRS	...	...	...	...	257	9.7	5.1	2.3

[a] Near-infrared reflectance.
[b] P = protein content, H = hardness, P + H = both.
[c] HRW = hard red winter; HRS = hard red spring.

TABLE III
Percentage of Wheat Samples Misclassified by Discriminant Analysis Models Based on Principal Component Analysis of NIR Spectra[a] Using Five, Six, or Eight Factors

		Calibration Set				Prediction Set			
Year	Class[b]	*n*	Five	Six	Eight	*n*	Five	Six	Eight
1987	HRW	50	6.0	4.0	4.0	248	5.2	6.8	5.6
	HRS	50	2.0	2.0	0	202	1.5	3.0	2.5
1988	HRW	50	4.0	0	0	313	3.8	2.6	1.9
	HRS	50	2.0	4.0	0	31	0	0	6.4
1989	HRW	50	2.0	2.0	2.0	331	8.8	7.2	5.7
	HRS	50	6.0	4.0	8.0	200	7.0	7.5	7.0
1990	HRW	...	...	...	...	471	6.6	8.8	16.8
	HRS	...	...	...	...	207	11.1	24.5	24.2

[a] Near-infrared reflectance (1,100–2,498 nm wavelength).
[b] HRW = hard red winter; HRS = hard red spring.

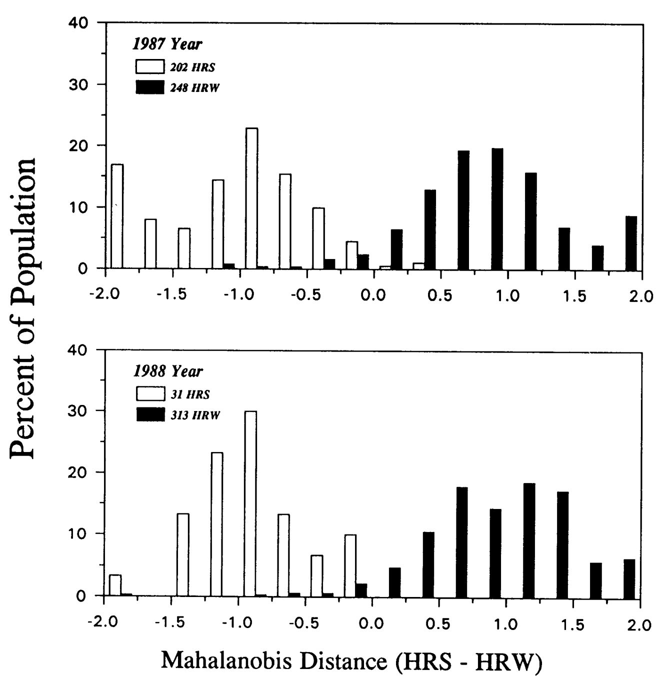

Fig. 3. Histogram of differences in the Mahalanobis distances of the samples from the 1987 and 1988 prediction sets of hard red winter (HRW) and hard red spring (HRS) wheats. The discriminant model was based on a five-factor principal component analysis, using the near-infrared reflectance spectra of 1987–1989 calibration samples.

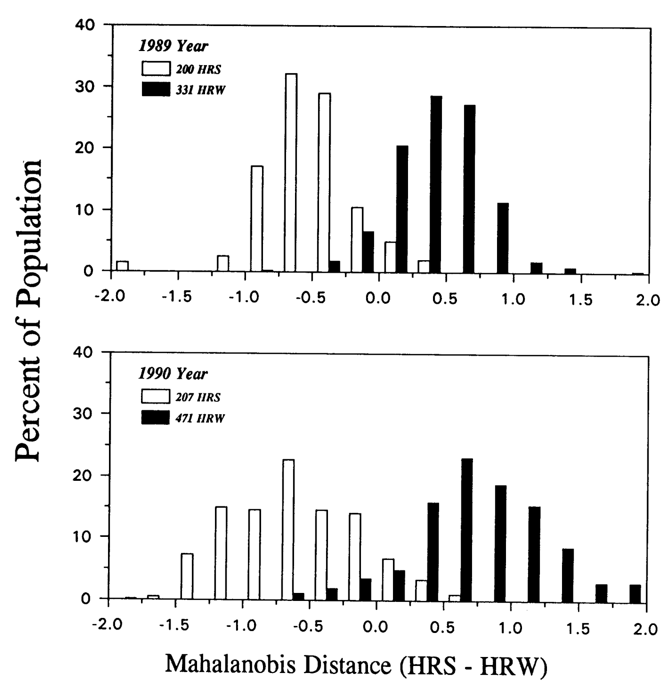

Fig. 4. Histogram of differences in the Mahalanobis distances of the samples from the 1989 and 1990 prediction sets of hard red winter (HRW) and hard red spring (HRS) wheats. The discriminant model was based on a five-factor principal component analysis, using the near-infrared reflectance spectra of 1987–1989 calibration samples.

The 1987 and 1988 histograms are generally similar in both shape and separation distance between the HRW and HRS clusters. The 1989 histograms appear to have narrower shoulders and a smaller separation distance between class medians than do those of 1987 and 1988. The 1990 histograms are most similar to the 1989 histograms, with exception that the 1990 distributions are slightly flatter and that more overlap occurs between the HRW and HRS envelopes, suggesting that proper classification becomes more difficult when discriminant models are applied to samples that were grown during years that are different from those used in calibration.

That five was the optimal number of principal component factors is the result of an examination of the sensitivity of misclassification to the number of factors (Fig. 5). In Figure 5, the same prediction samples from 1987 to 1989 as listed in Table I were classified by discriminant models containing two to 10 factors. A five-factor model was chosen as optimal, based on choosing the smallest number of factors that still yielded a model with low misclassification rates and also produced a bimodal shape for the histogram of Δ_{MD} (Figs. 3 and 4). However, when applied to the 1990 data (and recalling that 1990 samples were excluded from that particular calibration set), the performance of the five-factor model was slightly worse than was that for the 1987–1989 prediction sets. The eight-factor model demonstrated even poorer performance, suggesting that the use of more factors produces models that are overfitted to their calibration sets.

When calibrations included the 1990 samples and excluded samples from one of the other three years (Table IV), a highly predictable pattern for the misclassification rates of the excluded year's samples was not evident. Rather, the excluded-year misclassification rate varied from year to year. The 1987- and 1988-exclusive models yielded respective misclassification rates for 1987 (14.9% of HRW, 2.0% of HRS) and 1988 (5.4% of HRW, 0% of HRS) that were comparable to the misclassification rates for samples from the years included in calibration development. However, the 1989-excluded model demonstrated a high imbalance in misclassification rate between the 1989 winter and spring samples (93.0% of HRW, 0.5% of HRS).

Plots of factors one through six for the HRW and HRS submodels and for a submodel derived from pooling the HRW and HRS sets (1987–1989 calibration set) are shown in Figure 6. A multiplicative constant was applied to each factor (with values noted in the figure) to display more than one factor per set of axes. Each factor depicts the variation in spectral response as a function of wavelength. The contribution of each factor's explanation of variation within a submodel decreases as the factor number increases. The factor one curves (Fig. 6, upper left) resemble $\log(1/R)$ spectra, indicative that variation of sample-to-sample particle size distributions accounts for the largest variation in all three submodels. Factor two (Fig. 6, upper right), having a broad depression at a wavelength region (1,900–2,000 nm) usually ascribed to water, most likely represents the variation of sample moisture content. Factor three (Fig. 6, middle left) of the HRW submodel is noticeably different than those of the HRS and combined submodels. When multiplied by -1, factor three of the HRW submodel has a strong resemblance to the spectrum of wheat gluten (Fig. 7), suggesting that variation in protein levels among the HRW samples is the next largest variable after that of particle size and moisture content. Factor four (Fig. 6, middle right) of the HRW submodel and to a lesser extent, the other two submodels, demonstrates downward peaks at 1,420

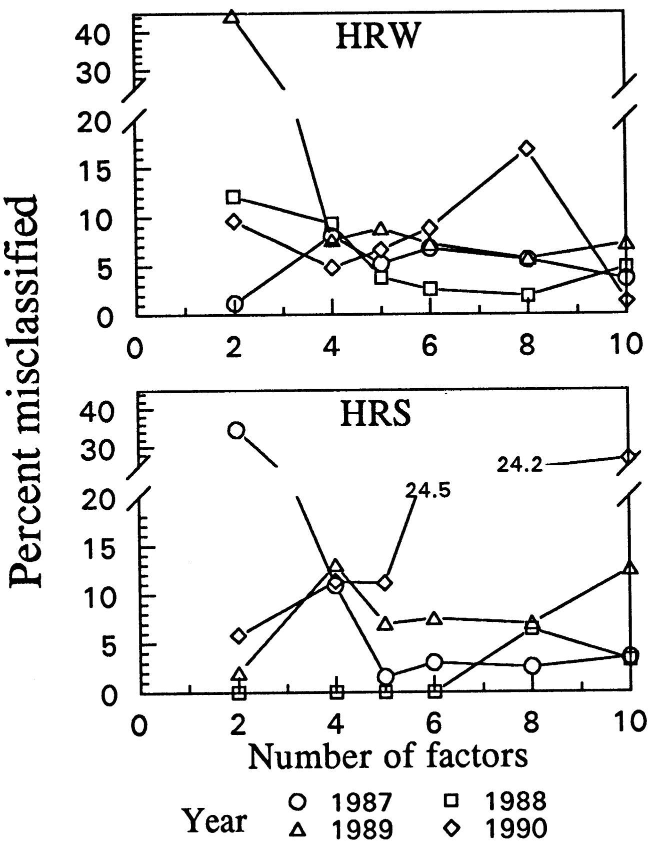

Fig. 5. Performance of the principal component analysis discriminant model on 1987, 1988, 1989, and 1990 prediction sets as a function of the number of factors. HRW = hard red winter; HRS = hard red spring.

TABLE IV
Percentage of Wheat Samples Misclassified by Five-Factor Principal Component Discriminant Analysis Models Developed Using Samples from Three of the Four Crop Years

			Year Excluded			
Year	Class[a]	*n*	1987	1988	1989	1990
1987	HRW	248	14.9	6.8	4.0	5.2
	HRS	202	2.0	3.0	4.0	1.5
1988	HRW	313	7.0	5.4	4.2	3.8
	HRS	31	0	0	3.2	0
1989	HRW	331	8.2	6.9	93.0	8.8
	HRS	200	9.0	9.5	0.5	7.0
1990	HRW	471	4.9	5.1	5.3	6.6
	HRS	207	7.7	10.1	10.1	11.1

[a]HRW = hard red winter; HRS = hard red spring.

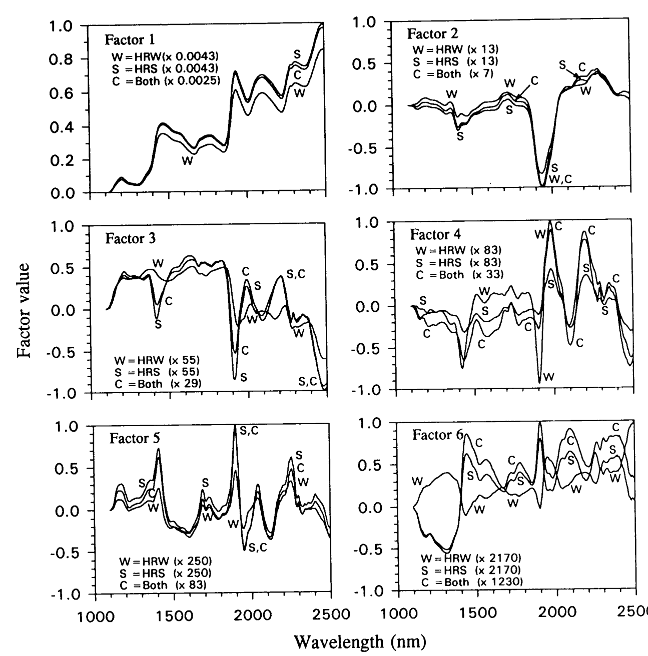

Fig. 6. Principal component factors one through six from the 1987–1989 calibration set. Three curves shown per factor, one each ($n = 150$) for the hard red winter (HRW) and hard red spring (HRS) submodels and one ($n = 300$) for the combined submodel.

and 1,900 nm that are close to the wavelengths of peak absorption caused by water. Additionally, there is an upward peak at 1,980 nm, a region usually associated with protein (Williams and Norris 1987). Therefore, factor four apparently demonstrates a molecular interaction between water and other constituents (proteins and carbohydrates). Factor five (Fig. 6, lower left) in all three submodels also presumably accounts for the same types of interactions. Finally, factor six (Fig 6, lower right), while still possessing some likeness to water, protein, and starch, has become comparatively small in magnitude when compared with the higher order factors. This factor most likely represents variability caused by varietal and environmental differences rather than by class differences.

DISCUSSION

As seen from the high misclassification rates of certain groups listed in Table II, neither NIR protein content nor NIR hardness alone is sufficient for discriminating between HRW and HRS wheats. Highly significant year-year differences in NIR protein or NIR hardness lessens the ability to develop a consistently accurate one-parameter classification model. Model accuracy could be improved by restricting a classification model to be specific to a single year; however, this would be at the expense of the generality of the model and would require a new calibration for each new crop year.

A slight improvement in model accuracy is gained when the discriminant function is based on both NIR protein content and NIR hardness. However, even with data from only one year (e.g., 1990 in Fig. 2), there is an absence of a perfect delimiter between HRW and HRS wheats that is based on both parameters.

The best classification models developed in this study were based on discriminant analysis applied to the principal components of NIR spectra. For classification of samples excluded from the 1987–1989 calibration set but grown during the same three years, the overall misclassification rate was 5%. Recalling that approximately 3% of the 1987 samples had a discrepancy in classification between that determined by the Doty Laboratories (for which the classes of all samples from 1988 through 1990 are based) and the FGIS-BAR, a 5% misclassification rate approaches the limit of experimental error.

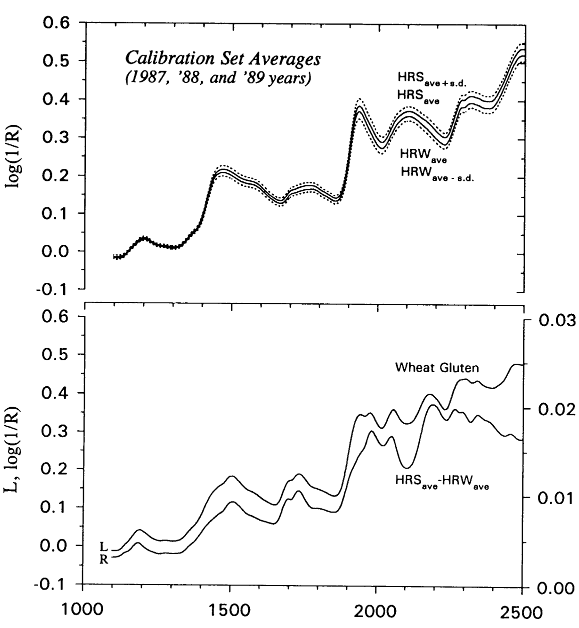

Fig. 7. The 150-curve average spectra for the 1987–1989 calibration set of hard red winter (HRW) and hard red spring (HRS) samples (top). A one-standard deviation envelope is plotted to one side of each average spectrum. The spectrum formed from the difference of the average spectra and the spectrum of wheat gluten (bottom).

In the attempt to explain the ability of the PCA models to distinguish class, the 150-curve average spectra of the HRW and HRS samples used in the 1987–1989 calibration set, along with the positive one-standard deviation envelope (for HRS) and the negative one-standard deviation envelope (for HRW), are plotted in the upper graph of Figure 7. Apparent from this graph is a lack of major differences between the spectra of the two classes. Throughout much of the wavelength region, the average spectra are within one standard deviation of each other. The spectrum formed as the difference between the average HRS and HRW spectra is shown in the lower graph of Figure 7. Also contained in the lower graph is a spectrum of wheat gluten. The similarities among the lower two plots of Figure 7 suggest that the differences between spectra of HRW and HRS classes are largely based on protein. Such protein dependency may be the reason for the poor classification performance of the 1989-excluded PCA model when applied to the 1989 HRW samples, because wheat of that class and year was abnormally high in protein, and, consequently, the HRW samples were judged as belonging to the class of higher protein.

The best three-factor combination that separated the two wheat classes consisted of factors one, four, and five. Plots of these three factor's scores (Fig. 8) of the 300-sample 1987–1989 calibration set demonstrate the degree of separation achievable. We attribute factor one to particle size variation, alluding to the differences in milling properties between the two classes, which is supportive of the observations of Endo et al (1990a). Factors four and five are more difficult to characterize other than to ascribe these to class-to-class differences in protein level and in the interaction of water, protein, and carbohydrates. It is interesting that all 1989 calibration samples form a distinct cluster away from the 1987 and 1988 samples, possibly because of a unique water-protein relationship present in the 1989 crop. That factor two is absent from the best combination is important because it indicates that water-matrix interactions are a more important feature for distinguishing class than is water alone.

CONCLUSIONS

Various discriminant analysis models have been developed for objectively distinguishing HRW from HRS wheat. A calibration set, consisting of equal numbers of HRW and HRS samples from three crop years (300 samples total) was used to develop discriminant functions. The discriminant functions were based on NIR-predicted protein content (12% mb), NIR hardness, the

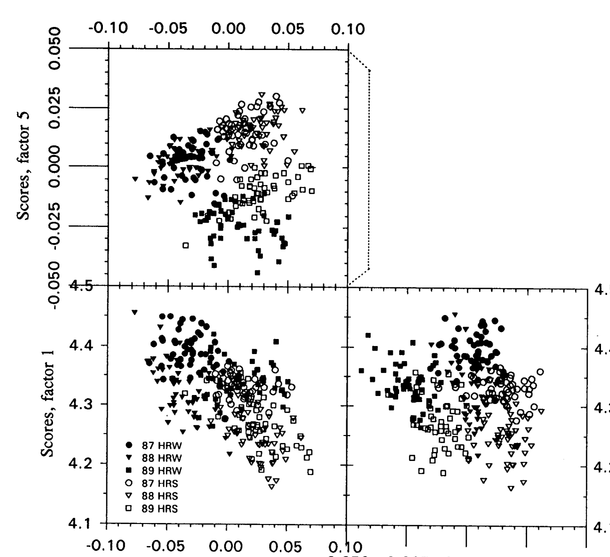

Fig. 8. Principal component scores belonging to factors one, four, and five of the 1987–1989 calibration set. Determined from the combined ($n = 300$) submodel. HRW = hard read winter; HRS = hard red spring.

combination of NIR protein and NIR hardness, and the principal components of the NIR spectra. Each model was used to predict the class of 2,103 samples (excluded from the calibration set).

The following conclusions are offered. First, NIR protein content or NIR hardness alone are insufficient for correctly distinguishing HRS and HRW wheats. Second, when NIR protein content and NIR hardness are jointly used to form a discriminant function, a slight improvement is observed in classification; however, year-to-year differences in NIR protein and NIR hardness may severely limit the performance of models when applied to samples from crop years not included in calibration. And third, discriminant analysis applied to the principal components of the NIR spectra yields a highly consistent and accurate (>95%) classification model. However, even with this technique, accuracy decreases when the model is applied to samples grown during years that are absent from the model's calibration set, especially when the excluded-year's crop has characteristics (protein, hardness, carbohydrate) markedly different from those of the included years.

ACKNOWLEDGMENT

We thank Joyce Shaffer of USDA-ARS for sample analysis and data file management and Eurvin Williams of USDA-FGIS Board of Appeals and Review for official classification services.

LITERATURE CITED

AMERICAN ASSOCIATION OF CEREAL CHEMISTS. 1983. Approved Methods of the AACC, 8th ed. Method 39-70A, approved October 1986, revised November 1989. The Association: St. Paul, MN.

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.

CHEN, C., CHIANG, Y. P., and POMERANZ, Y. 1989. Image analysis and characterization of cereal grains with a laser range finder and camera contour extractor. Cereal Chem. 66:466-470.

DELWICHE, S. R. 1991. Guide to wheat hardness: 1989 crop year. Paper ARS-79-1. U.S. Agric. Res. Serv. ARS.

DEVAUX, M. F., BERTRAND, D., ROBERT, P., and QANNARI, M. 1988. Application of multidimensional analyses to the extraction of discriminant spectral patterns from NIR spectra. Appl. Spectrosc. 42:1015-1019.

ENDO, S., OKADA, K., NAGAO, S., and D'APPOLONIA, B. L. 1990a. 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.

ENDO, S., OKADA, K., NAGAO, S., and D'APPOLONIA, B. L. 1990b. Quality characteristics of hard red spring and winter wheats. II. Statistical evaluations of reversed-phase high-performance liquid chromatography and milling data. Cereal Chem. 67:486-489.

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.

JOLIFFE, I. T. 1986. Principal Component Analysis. Springer-Verlag: New York.

MAHALANOBIS, P. C. 1936. On the generalized distance in statistics. Proc. Natl. Inst. Sci. India 2:49-55.

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.

MARK, H. L., and TUNNELL, D. 1985. Qualitative near-infrared reflectance analysis using Mahalanobis distances. Anal. Chem. 57:1449-1456.

MEPPELINK, E. K. 1974. Untersuchungen über die methodik der kornhärtebestimmung bei weizen. Getreide Mehl Brot 28:142-149.

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.

NORRIS, K. H., HRUSCHKA, W. R., BEAN, M. M., and SLAUGHTER, D. C. 1989. A definition of wheat hardness using near infrared reflectance spectroscopy. Cereal Foods World 34:696-705.

OFFICE OF TECHNOLOGY ASSESSMENT. 1989. Enhancing the quality of U.S. grain for international trade. OTA-F-399. U.S. Government Printing Office: Washington, DC.

PEARSON, K. 1901. On lines and planes of closest fit to systems of points in space. Philos. Mag. 6:559-572.

POMERANZ, Y., CZUCHAJOWSKA, Z., SHOGREN, M. D., RUBENTHALER, G. L., BOLTE, L. C., JEFFERS, H. C., and MATTERN, P. J. 1988. Hardness and functional (bread and cookie-making) properties of U.S. wheats. Cereal Foods World 33:297-304.

SAS INSTITUTE, INC. 1988. SAS User's Guide: Statistics. Release 6.03 Ed. SAS Institute Inc.: Cary, NC.

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. 1979. Screening wheat for protein and hardness by near infrared reflectance spectroscopy. Cereal Chem. 56:169.

WILLIAMS, P. C., and NORRIS, K. H., eds. 1987. Page 246 in: Near-Infrared Technology in the Agricultural and Food Industries. 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. Adv. Cereal Sci. Technol. 5:211-259.

WU, Y. V., STRINGFELLOW, A. C., and BIETZ, J. A. 1990. Relation of wheat hardness to air-classification yields and flour particle size distribution. Cereal Chem. 67:421-427.

YAMAZAKI, W. T. 1972. A modified particle-size index test for kernel texture in soft wheat. Crop Sci. 12:116.

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 March 31, 1992. Accepted July 8, 1992.]