FOOD CHEMICAL CONTAMINANTS

Discrimination of Sound and Granary-Weevil-Larva-Infested Wheat Kernels by Near-Infrared Diffuse Reflectance Spectroscopy

AHMAD R. GHAEDIAN and RANDY L. WEHLING [1]
University of Nebraska, Department of Food Science and Technology, 143 Filley Hall, Lincoln, NE 68583-0919

Sound and infested wheat kernels containing late-instar granary weevil larvae, as identified by X-ray analysis, were used to evaluate the ability of near-infrared (NIR) spectroscopy to predict the presence of insect larvae in individual wheat kernels. Diffuse reflectance spectra at 1100–2500 nm were recorded from individual infested and sound kernels. Principal component analysis (PCA) of NIR spectra from sound kernels was used to construct calibration models by calculation of Mahalanobis distances. Calibration models were then applied to spectra obtained from both sound and infested kernels in a separate validation set. A 5-factor PCA model using data from a first-derivative spectral transformation was the best model for correctly classifying kernels in an expanded validation sample set, including 100% of sound, 93% of infested, 95% of sound air dried, 86% of infested air dried kernels, and 90% of sound kernels from 6 wheat varieties. Calibrations using the spectral region from 1100 to 1900 nm were least sensitive to kernel moisture differences. Similar results were obtained when discriminant analysis was applied to log 1/R data from selected discrete wavelengths of NIR spectra.

Referred to as "hidden insects" (1), internal infesters continue to damage stored grains without warning. Of all methods used to detect hidden insects, X-ray analysis is the choice of industry in the United States (2). However, this technique is time consuming and requires expensive, dedicated X-ray equipment. The cost per test is high because the film and materials used are costly. In addition, the technique becomes subjective because a highly trained technician is required to interpret radiographs and identify immature insects from grain tissues. Therefore, development of a rapid, reliable, and less subjective method for detecting insect larvae that feed internally in stored grain is desirable.

Received July 5, 1996. Accepted by AP March 15, 1997.
Published as Paper No. 11573, Journal Series, Agricultural Research Division, University of Nebraska, Lincoln, NE 68583-0704.
[1] Author to whom correspondence should be addressed.

Other methods to detect internal infesters in grains are staining of egg plugs made by female insects after laying eggs (3), visual inspection for exit holes left by emerging adult insects (4), flotation of hollowed kernels left by feeding insects (5), cracking of kernels followed by flotation of released insect parts (6), and staining crushed kernels with ninhydrin to detect amino acids corresponding to the insect (7). None of these techniques has been adopted as an official method by regulatory agencies because of questions of reliability, reproducibility, and simplicity of the method.

An infrared method for measuring carbon dioxide produced by insects (8) is unreliable as a marker of insect infestation because of the difficulty in correcting for background carbon dioxide released by respiring grains. Hackman and Goldberg (9) proposed a colorimetric procedure for determining chitin, a major structural component of insect cuticle, as an index of insect infestation in grain. This procedure, however, is not adequately specific as an index of insect infestation in grain because a high concentration of chitin also can be found in stored grain contaminated with fungi, as evidenced by the work of Donald and Mirocha (10).

A nuclear magnetic resonance (NMR) spectroscopic technique developed by Chambers et al. (11) for monitoring development of granary weevil in wheat kernels and an enzyme linked immunosorbent assay (ELISA) test (12), based on a measure of insect biomass (myosin, muscle protein) found in all insects but not in grain, have not found widespread acceptance because of their complexity or the need for further development.

Sonic detection of chewing insects also has been investigated (13, 14). Although sound detection has advantages in portability, speed, and cost, it has not been accepted as an official method, perhaps because of low sensitivity in detecting young larvae. A system for sonic detection of internal infesters was described by Vick et al. (15). They used larvae of rice weevil, lesser grain borer, and angoumois grain moth in their study and reported that the low sensitivity of the system did not allow molting larvae of the species to be detected. They also reported that infesting species have to be identified to estimate the infestation level; however, they did not find consistent acoustic differences among the 3 species that would allow species identification.

Hagstrum and Flinn (16) conducted research to distinguish 5 species of adult stored-product insects based on acoustical differences. They concluded that only some species were distinguishable from one another. Their study involved use of adult insects only and did not account for insect larvae. In an earlier study, Hagstrum et al. (17) evaluated an acoustical detection system to estimate the population of *Rhyzopertha dominica* (Coleoptera: Bostrichidae) in stored wheat. Although insect larvae were more numerous than adults, they were not acoustically detectable when adults were present.

Near-infrared (NIR) spectroscopy is a relatively fast, accurate, and economical technique available to the grain industry for compositional analysis. Wilkin et al. (18) used NIR reflectance to detect mite infestation in animal feed. However, no reports indicate use of this technology in detecting internally infesting insects. NIR spectroscopy can be used for both qualitative and quantitative analysis. To develop calibration models for quantitative analysis, a reference method of analysis is needed to determine the concentration of the constituents of interest in a training set (19, 20). NIR spectroscopy has been used for quantitative analysis of a wide variety of samples in industry. Recently, a number of statistical methods known as pattern recognition techniques have been reported for qualitative identification of samples by their NIR spectra (21–25).

One technique that has been used successfully is discriminant analysis based on Mahalanobis distances (26, 27). A clear explanation of the use of Mahalanobis distances for qualitative analysis is given by Mark and Tunnell (21). Discriminant analysis based solely on Mahalanobis distances of reflectance data from discrete wavelengths depends on proper selection of wavelengths (21). One way to avoid this difficulty is to apply principal component analysis (PCA) to the training set before calculating Mahalanobis distances. PCA analysis allows use of the entire spectrum in the discrimination method and helps to ensure that compositional differences in the unknown sample are given the chance of being detected (28).

Delwiche and Norris (29) developed various forms of discriminant analysis models to classify hard red winter and hard red spring wheat based on their NIR spectra. They concluded that best results are obtained when discriminant analysis is applied to the principal components of the spectra.

Shah and Gemperline (28) used PCA to calculate Mahalanobis distances for classifying pharmaceutical raw materials and concluded that the combined method provides reliable classification models. They used PCA of the NIR spectra of raw materials to calculate Mahalanobis distances based on the significant principal components. After the most significant PC scores are determined, they can be used to represent the original data rather than the absorbance intensities at all wavelengths.

Other examples of qualitative analysis by NIR spectroscopy include the work of Rose (30) in classifying 40 pharmaceutical raw materials by discriminant analysis and the work of Shenk et al. (31), who identified different populations of forage samples from their NIR spectra. Lodder et al. (19) detected adulterants in nonprescription drugs by using NIR reflectance analysis and a pattern recognition technique called BEAST (Boostrap Error Adjusted Single Sample Technique). Gemperline et al. (32) classified pharmaceutical raw materials by using NIR spectra and a pattern recognition technique called SIMCA (Soft Independent Modeling of Class Analogy). Classification of samples was achieved by PCA of complete NIR spectra. Scott (33) used SIMCA to determine chemical classes of toxic volatile organic compounds from low-resolution mass spectra data. Derde and Massart (34) developed a classification technique based on Mahalanobis distances and used gas chromatographic data to classify olive oils according to their geographic location.

Because of the speed and availability of NIR technology, we evaluated the possibility of using it for detecting insect infestation in stored wheat.

Experimental

Preparation of Insect-Infested Wheat Kernels

A 4-week-old culture of hard red winter (HRW) wheat containing late-instar granary weevil (*Sitophilus granarius* L.) larvae was obtained from the stored-product research laboratory of the Department of Entomology at Kansas State University (Manhattan, KS). Steps for preparing granary-weevil-infested wheat have been described (35). To identify infested wheat kernels, a radiograph was made of a portion of the culture with a General Electric X-ray grain inspection unit using 20 kV at 5 mA and a 2.5 min exposure time. Kodak Type M industrial X-ray film was used. The film was placed directly on the wheat kernels; therefore images were the same size as the actual kernels. The film was developed for 5 min, washed 1 min, fixed 5 min, washed again, and dried (36). The radiograph was then viewed with a film illuminator. After infested wheat kernels were identified, 75 infested kernels containing larvae were selected and placed in a mason jar. To ensure uniformity in the source of wheat kernels used in NIR calibration, 75 sound kernels also were picked from the radiographed kernels. NIR spectra were obtained from wheat kernels within 24 h to minimize changes in the growth stage of the larvae.

Instruments

Spectral data from all wheat kernels, expressed in the form of log (1/R), were collected with an NIRSystems Model 6500 spectrometer (NIRSystems Division of Perstorp Analytical, Silver Spring, MD) from 1100 to 2498 nm. Spectra were collected at a wavelength interval of 2 nm. Individual wheat kernels were placed in a Capcell parabolic reflector (Optical Prototypes, Mars, PA). Use of this device allowed collection of radiation reflected from the entire surface of the kernel. A remote reflectance probe attached to the monochromator of the instrument by a fiber optic cable was then positioned over the Capcell to measure light reflected from the kernel. Thirty-two monochromator scans were averaged from each kernel. The NIR spectrum of the kernel was obtained by taking the ratio of the intensity of radiation reflected from the sample to that reflected from a ceramic reference plate. The remote probe and parabolic reflector were covered with a black cloth to prevent entry of stray light during collection of each spectrum.

For instrumental control and data collection, the spectrometer was interfaced to an MS-DOS personal computer running the Near Infrared Spectral Analysis Software (NSAS) package (version 3.16, NIRSystems). Spectra were stored on the harddisk drive and then converted into ASCII (JCAMP-DX) format for import into Lab Calc software (Galactic Industries, Salem, NH).

Analysis of Data

A discriminant analysis program in the Lab Calc software package was used to analyze the NIR data. This pattern recognition technique was used to measure Mahalanobis distances, expressed in standard deviations, of sound and infested kernels from the center (mean) of the training set cluster. In theory, because the Mahalanobis distance represents a measure of standard deviation, essentially all samples in a group can be expected to lie within 3× the Mahalanobis distance of their respective group mean (21, 28). A training set containing only sound kernels was used to establish the group mean and Mahalanobis distance for sound wheat kernels. Validation sets containing individual sound and infested kernels were then presented to the instrument. Kernels with a Mahalanobis distance of less than 3 standard deviations from the training set center were classified as sound, and those with a Mahalanobis distance of 3 or greater standard deviations were classified as unsound because of infestation.

Two methods of calibration were used. The first was discriminant analysis based on loadings derived from PCA of full or partial NIR spectra. The discriminant analysis program used in this study takes full advantage of both PCA and Mahalanobis distances by combining both techniques into a single method. PCA allows use of full NIR spectra and avoids the need for wavelength selection. The program uses PC scores from spectra of samples in a training set to calculate the Mahalanobis groups, rather than spectral intensities from samples at selected wavelengths. This avoids overdiscrimination that can occur when Mahalanobis distances with more than 10 wavelengths are used to develop discriminant functions.

The second calibration method was discriminant analysis based on Mahalanobis distances applied to selected wavelengths. In this method, log 1/R values from several NIR wavelengths were used directly in discriminant analysis. Wavelengths were selected by use of best-possible-combination regression. The regression procedure was applied to a representative set that contained 30 samples each of sound and infested kernels. To perform the regression, numerical values were given to samples: 1.00 to spectra of sound kernels and 2.00 to spectra of infested kernels. Wavelengths selected by the regression algorithm were used to construct discriminant analysis models.

Acquisition of Spectra

Preliminary work was done to evaluate the effects of both positioning (crease up versus crease down) and location of wheat kernels in the sample cell. NIR spectra were obtained from a single wheat kernel in different positions. Results showed a shift in absorption depending on how and where in the sample cell the wheat kernel is placed. For this reason, it was decided to collect and study spectral data of wheat kernels positioned crease up and to scan the same kernels again with the crease being down in the sample cell. Reproducible placing of wheat kernels in the sample cell was carefully practiced. In addition, the orientation of the sample cell itself in the sample cell holder was carefully reproduced throughout this study.

The first 150 NIR spectra were acquired from 75 kernels each of infested and sound samples positioned crease up in the sample cell. Each kernel was numbered for later identification. The same kernels were used to generate the second 150 NIR spectra with kernels positioned crease down in the sample cell. To minimize the effect of any variation due to instrument drift, scanning was alternated between infested and sound kernels after every 10 samples. The numbered wheat kernels were then placed in a freezer (–18°C) for 2 days to kill the larvae. Then, both sound and infested kernels were placed in an open tray to dry at room temperature for about 70 days. They were scanned (crease down) to generate the third set of spectra. Presence or absence of larvae was confirmed by sectioning the kernels with a razor blade. The fourth set of NIR spectra was obtained from sound kernels of 6 wheat varieties (Arapahoe, Abilene, Cimarron, Karl, Tam 107, and Scout 66) grown in Nebraska. Spectra were acquired from 10 kernels of each variety, positioned crease down in the sample cell.

Training Sets

Two separate training sets were used to develop discriminant analysis models: The first training set was generated with NIR spectra of sound kernels positioned crease down. Of 75 NIR spectra, 50 were used in the training set and 25 were reserved for the test set. The second training set was developed with NIR spectra of sound kernels positioned crease up. Of 75 NIR spectra, 50 were used in the training set and 25 were used for the test set.

Validation Sets

Each model was used to classify samples in validation sets. Seven validation sets were used in this study; the number of samples used in each set is given in parentheses: (*1*) sound wheat kernels with crease down in the sample cell (25 or 75 spectra depending on the model used), (*2*) sound kernels with crease up (25 or 75 spectra), (*3*) infested kernels with crease down (75 spectra), (*4*) infested kernels with crease up (75 spectra), (*5*) sound air-dried kernels with crease down (75 spectra), (*6*) infested air-dried kernels with crease down (75 spectra), (*7*) sound kernels of 6 wheat varieties with crease down (60 spectra).

The first validation set consisted of spectra of wheat kernels from the same lot as the kernels of the first training set, in that they had the same moisture content and were positioned crease down in the sample cell. The second set was the same as the first set except that spectra were obtained by flipping the kernels over in the sample cell for scanning. The third and fourth validation sets were spectra of insect-infested kernels from the same lot. The fifth set consisted of spectra of the same sound kernels used in the first validation set after they had been air dried for about 70 days and therefore had a different moisture content compared with the training set samples. The sixth set consisted of spectra of air-dried infested wheat kernels. The kernels contained dead larvae and simulated stored fumigated kernels. The seventh set consisted of spectra of 10 sound kernels from each of 6 HRW wheat varieties (Arapahoe, Abilene, Cimarron, Karl, Tam 107, and Scout 66) grown in Nebraska. The kernels also had a lower moisture content than those used in the training set.

These validation sets were used to assess the ability of the calibration models to correctly classify wheat kernels as sound or infested that are different from the training sets with respect to moisture content and variety.

Development and Validation of Calibration Model

To create a calibration model, spectra from a training set were input into the Lab Calc DISCRIM program. Various calibration parameters, processing conditions, and diagnostics were available to optimize calibration results.

One or more regions of the spectrum could be selected for analysis. A single region encompassing the entire NIR spectrum from 1100 to 2498 nm was used initially to develop calibration models. Additionally, because the resolution of the spectra was greater than necessary for accurate analysis, the resolution of the data was reduced by averaging each 4 consecutive data points to speed up calculation. Averaging data points is better than skipping data points, because no information is discarded prior to analysis. Also, averaging 4 consecutive data points to produce a single point tended to decrease the effect of noisy data. Mean centering and spectral residuals were used for all training and validation samples. Spectral residuals are the information remaining in the secondary set of PCA vectors that contribute to the small random variation among the data. Use of spectral residuals enhances the sensitivity of the Mahalanobis distance calculations to outlier samples by using the sum squared of the spectral residuals as an additional spectral score. When needed, multiplicative scatter correction (37) and first-derivative transformation were applied to spectra of calibration and validation sets. The number of PCA factors used in analysis could be selected and was limited to the maximum number of factors recommended by a diagnostic routine performed on the training data.

In Lab Calc, the option of using wavelengths rather than PCA vectors for calculating Mahalanobis distances is also available. This method was examined for sample classification, with wavelengths selected by multiple linear regression as previously described. A maximum of 12 wavelengths was used to minimize overfitting.

Once discriminant analysis models were developed and the integrity of the training data was closely examined, the models were used to predict samples from the validation sets. The number of PC factors or wavelengths was varied to obtain optimal classification of prediction set samples.

To evaluate the validity of classification of sound and infested wheat kernels based on discriminant analysis models, two terms—"correct classification" and "misclassification"—were used. Correct classification refers to percent sound kernels of a validation set classified correctly as member, or percent infested kernels of a validation set classified correctly as nonmember, of the training set. (Recall that the training set contained sound kernels only.) Misclassification, however, corresponds to percent sound kernels classified incorrectly as nonmember or percent infested kernels classified incorrectly as member of the training set. The number of PCs or wavelengths for classification was selected on the basis of optimum prediction rate or the best correct classification of the validation sets.

Results and Discussion

Figure 1 shows spectra of several sound and infested wheat kernels. There was no consistent spectral offset between larvae-infested kernels and sound kernels.

Results of Calibrations with PCA

Table 1 summarizes classification results from Mahalanobis distances based on PCA of the entire NIR spectral range from 1100 to 2498 nm. The first training set—containing sound kernels with creases positioned down in the sample cell—was used to develop these models. One sample in the training set was identified as a spectral outlier and deleted from the set. The number of PCs used to construct the models are given in the table, with the number of spectra in the validation sets shown in parentheses. The results show that the discriminant model with 7 factors was highly successful at classifying both sound and infested wheat kernels. When the 8th factor was included in constructing the model, the prediction rate was even higher, with nearly perfect correct classification. Although the number of correct classifications of kernels with crease positioned up in the sample cell was slightly lower than those of kernels with creases down, prediction rates were similar. In other words, training sets containing sound kernels with crease down could be used to predict sound or infested kernels regardless of positioning in the sample cell.

Next, calibrations were developed with multiplicative scatter correction and tested for performance by using the validation sets. Results are shown in Table 2. Using scatter-corrected spectra improved the prediction rate in 3 of 4 validation sets when a calibration with 7 factors was used, compared with results in Table 1. Results indicated that fewer factors are needed in the calibration to achieve a high prediction rate when scatter correction is used than when it is not.

Calibrations also were developed from first-derivative spectra and evaluated for performance by using the same validation sets. Results given in Table 2 show that the percentage of correct classification using the same number of factors was actually less compared with the percentage obtained with scatter correction or with use of mean centering alone.

To check the versatility of the models, validation sets containing the spectra of sound and infested kernels that had been air dried and NIR spectra of 60 sound kernels from 6 wheat varieties (10 spectra from each variety) were used. Classification results are shown in Table 3. The models performed very poorly for sound, air-dried kernels. Use of scatter-corrected spectra did not improve results, indicating that the models were useful only for wheat kernels with moisture contents similar to those of the training set. Air drying of wheat kernels altered NIR spectra to the extent that the models were no longer accurate. The moisture contents of wheat taken from a 4-week-old culture containing insect larvae before and after air drying were

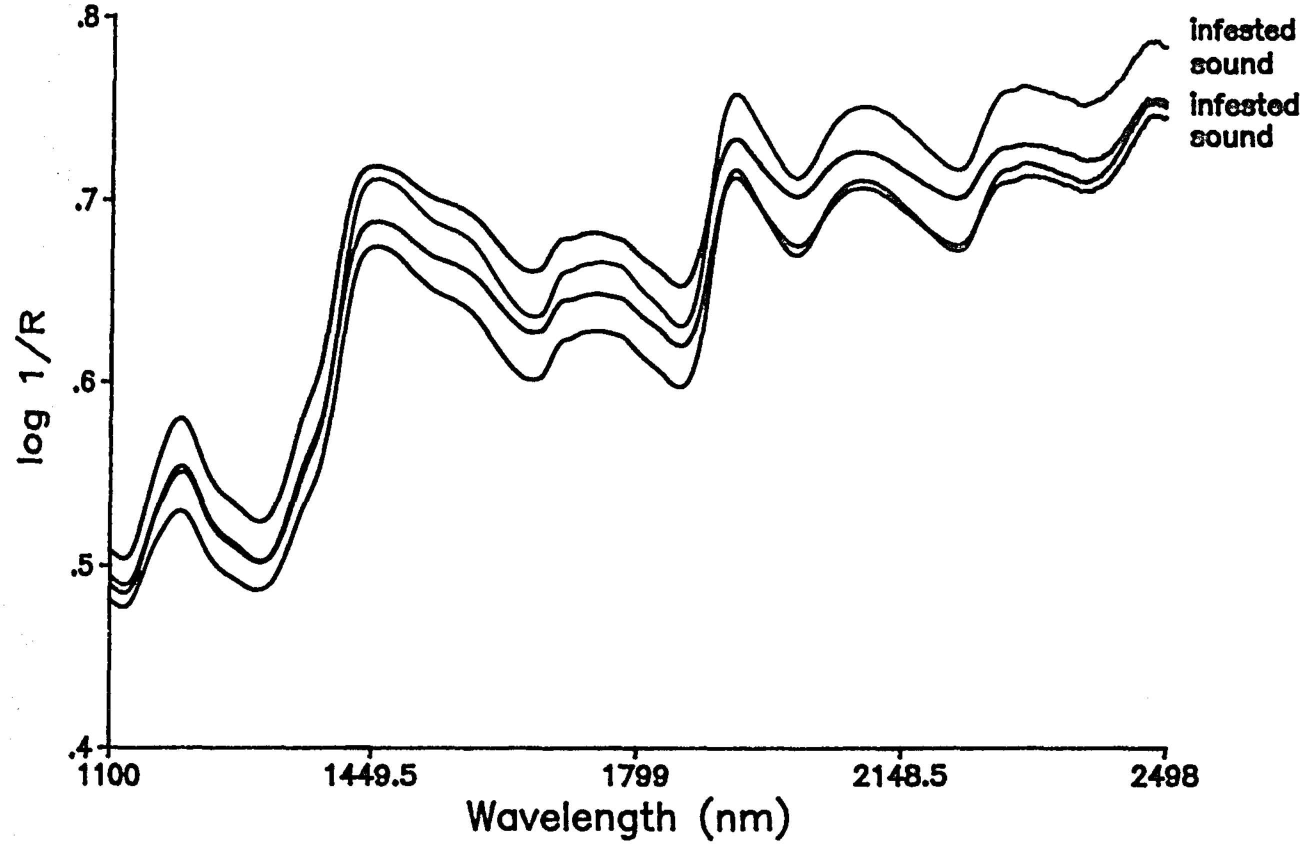

Figure 1. NIR reflectance spectra of several sound and granary-weevil-larvae-infested wheat kernels.

determined in duplicate (38). The moisture contents were 12.7 and 7.8%, respectively. Because the cultures were maintained at 70% relative humidity for 4 weeks, they had a higher moisture content. These results led us to a more detailed study to answer questions regarding the effects of moisture content and the type of wheat kernels in developing calibration models.

One major limitation of NIR spectroscopy is interference of water bands with spectral bands of other constituents (39). Although many regions in the NIR spectrum are associated with water (40), the region around 1940 nm represents the most prominent absorption region in wheat spectra. Using the same training set, wavelengths of 1901 to 1979 nm were removed from the spectra to minimize the effect of moisture differences between the training set kernels and all the validation spectra. Therefore, 2 regions of the spectra, 1100–1900 and 1980–2498 nm, were used in developing calibration models.

Table 1. Percentage of sound and infested wheat kernels classified correctly by discriminant analysis models based on principal component analysis of NIR[a] spectra

No. of factors	Sound cd[b] (25)[c]	Infested cd (75)	Sound cu[d] (75)	Infested cu (75)
7	100	93	88	92
8	100	100	92	97

[a] Near-infrared diffuse reflectance. Calibration file contained sound wheat kernels with crease down while scanning. Full spectra from 1100 to 2498 nm were used for classification.
[b] cd, crease down while scanning.
[c] *Number in parentheses is the number of spectra used in prediction set.*
[d] cu, crease up while scanning.

Table 2. Percentage of sound and infested wheat kernels classified correctly by discriminant analysis models based on principal component analysis of NIR[a] spectra using scatter correction and first-derivative spectra

No. of factors	Sound cd[b] (25)[c]	Infested cd (75)	Sound cu[d] (75)	Infested cu (75)
5 (sc)[e]	100	88	97	89
6 (sc)	96	96	99	96
7 (sc)	96	100	97	96
6 (1st der)[f]	100	76	—	—
7 (1st der)	100	80	—	—

[a] Near-infrared diffuse reflectance. Calibration file contained sound wheat kernels with crease down while scanning. Full spectra from 1100 to 2498 nm were used for classification.
[b] cd, crease down while scanning.
[c] The number in parentheses is the number of spectra used in prediction set.
[d] cu, crease up while scanning.
[e] sc, scatter-corrected spectra.
[f] 1st der, first-derivative spectra.

Table 3. Percentage of air-dried and 6 varieties of wheat kernels classified correctly by discriminant analysis models based on principal component analysis of NIR[a] spectra

No. of factors	Sound air-dried cd[b] (75)[c]	Sound 6 variety cd (60)
4	0	90
7	40	30
6 (sc)[d]	0	43

[a] Near-infrared diffuse reflectance. Calibration file contained sound wheat kernels with crease down while scanning. Full spectra from 1100 to 2498 nm were used for classification.
[b] cd, crease down while scanning.
[c] The number in parentheses is the number of spectra used in prediction set.
[d] sc, scatter-corrected spectra.

Table 4 summarizes classification rates for air-dried kernels. Correct classification rates for the validation sets are given for the 5-, 6-, 7-, and 8-factor models. Models with 5 and 7 factors gave comparable results, and the 6-factor model yielded a better classification rate. The 8-factor model demonstrated very poor performance with a high misclassification rate for sound kernels, suggesting that use of too many factors produces models that are overfitted to the training set and results in overdiscrimination.

When a first-derivative transformation was applied to the spectra, the classification rate improved for sound air-dried kernels (Table 4). A 7-factor model yielded an optimum prediction rate for both sound and infested kernels. It was chosen as an optimal model on the basis of using the least number of factors that still produced low misclassification rates. Models developed after removal of the water band from the spectra gave a very good prediction rate in classifying air-dried kernels of both sound and infested samples.

Table 4. Percentage of sound and infested air-dried wheat kernels classified correctly by discriminant analysis models based on principal component analysis of NIR[a] spectra

No. of factors	Sound air-dried cd[b] (75)[c]	Infested air-dried cd (75)
5	69	95
6	83	95
7	69	99
8	20	—
5 (1st der)[d]	96	80
6 (1st der)	97	80
7 (1st der)	91	92

[a] Near-infrared diffuse reflectance. Calibration file contained sound wheat kernels with crease down while scanning. The water band from 1900 to 1980 nm was removed from the spectra.
[b] cd, crease down while scanning.
[c] The number in parentheses is the number of spectra used in prediction set.
[d] 1st der, first-derivative spectra.

Calibration models also were developed from selected regions of the NIR spectra. First, the lower portion of the spectra from 1100 to 1900 nm was used for calibration development, and the performance of the models was evaluated by using the validation sets. Results obtained with the truncated spectra are given in Table 5. The 4- and 5-factor models gave similar results. As the number of factors increased beyond 5, overdiscrimination occurred, indicating that the models were overfitted to the training sets.

Use of first-derivative spectra improved performance significantly. A 5-factor model correctly classified 100% of sound, 93% of infested, 95% of sound air-dried, 86% of infested air-dried, and 90% of sound kernels from 6 wheat varieties (Table 5). These results reflect the ability of NIR spectroscopy and discriminant analysis techniques to distinguish between sound and infested wheat kernels, regardless of differences in moisture content, and to correctly classify sound kernels from different wheat varieties.

The upper region of the spectra from 1980 to 2498 nm also was evaluated and was found not to contain enough information for reliable classification. Because of strong absorptions of water, starch, and protein in this region, it may be that the IR radiation does not penetrate deeply enough into the kernel to interact with the larva.

Models were developed from the second training set that contained sound wheat kernels with creases positioned up in the sample cell while scanning. Samples from the same validation sets were used to evaluate the performance of these models. Models that could be applied to the validation sets without added mathematical treatments and still yield high classification rates were not found. (Recall that mean centering and spectral residual were used in all calibrations). Better results were obtained with multiplicative scatter-corrected spectra (Table 6).

The models were successful in classifying kernels with crease up (the same position as kernels of the training set). When only 4 factors were used, 91% of infested kernels were classified correctly. When the number of factors increased to 5

Table 5. Percentage of sound, infested, air-dried and 6 varieties of wheat kernels classified correctly by discriminant analysis models based on principal component analysis of NIR[a] spectra

No. of factors	Sound cd[b] (25)[c]	Infested cd (75)	Sound air-dried cd (75)	Infested air-dried cd (75)	Sound 6 variety cd (60)
4	100	71	88	95	78
5	100	77	91	97	78
7	—	—	65	98	—
5 (1st der)[d]	100	93	95	86	90

[a] Near-infrared diffuse reflectance. Calibration file contained sound wheat kernels with crease down while scanning. Partial spectra from 1100 to 1900 nm were used for classification.
[b] cd, crease down while scanning.
[c] The number in parentheses is the number of spectra used in prediction set.
[d] 1st der, first-derivative spectra.

and 6, correct prediction of infested kernels improved to 96 and 97%, respectively. However, the models were not successful in predicting sound kernels with the opposite positioning in the sample cell, as is evident in the table. The best result was 59% correct classification of sound kernels with 6 factors. When the results presented in Table 6 are compared with those in Table 1, the following conclusion can be made: Models constructed on the basis of spectra of sound kernels with creases positioned downward in the sample cell can predict both sound and infested kernels, regardless of how they are placed in the sample cell, with high percentages of correct classification. However, models developed on the basis of kernels with creases positioned up in the sample cell can predict only samples of the same positioning and give high misclassification rates for samples with opposite positioning.

To confirm the presence of infestation predicted by NIR spectroscopy, wheat kernels were sectioned with a razor blade for visual observation. All infested kernels that were predicted as infested were in fact infested, bearing insect larvae inside. Only one infested sample was consistently misclassified as a sound kernel. After this particular kernel was sectioned, it was found that the larva inside was clearly smaller than the larvae of other kernels. This observation was important, indicating the possible limitation of NIR spectroscopy in predicting early stages of infestation.

Biological Meaning of PCA factors

To understand PCA factors, PCs were extracted from spectra of infested wheat kernels to compare the factors of sound kernels with those of infested kernels.

Evaluation showed that factors 1, 2, and 3 of infested kernels are similar to factors of their sound counterparts when multiplied by −1. Factors 1 and 2 showed absorption from all major wheat constituents and may account for differences in overall reflectance arising from differences in kernel geometry. In general, for diffuse reflectance, the first PC is related to particle size and shape and contains little information that can be used to measure directly chemical differences among samples (40).

Table 6. Percentage of sound and infested wheat kernels correctly classified by discriminant analysis models based on principal component analysis of NIR[a] spectra using scatter-corrected spectra

No. of factors	Sound cd[b] (75)[c]	Infested cd (75)	Sound cu[d] (25)	Infested cu (75)
4 (sc)[e]	29	100	100	91
5 (sc)	33	100	100	96
6 (sc)	59	100	100	97

[a] Near-infrared diffuse reflectance. Calibration file contained sound wheat kernels with crease up while scanning. Full spectra from 1100 to 2498 nm were used for classification.
[b] cd, crease down while scanning.
[c] The number in parentheses is the number of spectra used in prediction set.
[d] cu, crease up while scanning.
[e] sc, scatter-corrected spectra.

Factor 3 of both sound and infested kernels accounted for carbohydrates, the major constituent of wheat. However, in factors 4–7, differences were observed in spectral regions associated with various chemical constituents. Factor 4 of sound kernels accounted for wheat protein, while water and phenolic structures were heavily loaded in factor 4 of infested wheat spectra. Comparing factor 5 of infested wheat with that of sound wheat showed obvious differences in the intensities of the lipid and moisture bands between the 2 factors. Lipid bands were more visible in factor 6 of infested wheat than in the same factor from sound kernels. Factor 7 of infested wheat also showed a likeness to lipid not observed with sound kernels. Therefore, on the basis of the evaluation of factors 1–7 of the full spectra, the differences between sound and infested kernels are mostly due to moisture, protein, lipid, and phenolic structures. However, when wavelengths >1900 nm are eliminated, the portion of the factors representing moisture also is eliminated, leaving protein, lipid, and phenolic structures as the major sources of variation between sound and infested wheats.

Results of Calibrations Based on Discrete Wavelengths

Qualitative NIR reflectance analysis using Mahalanobis distances based on discrete wavelengths is conceptually simple (21) and requires simpler and less expensive instrumentation. Detection of internal insect infestation based on selected discrete wavelengths was, therefore, attempted.

Results of discriminant analysis using discrete wavelengths are shown in Table 7. The training set that proved to be the best when using PCA was also used to develop discrete-wavelength calibration models. Several analyses were made to find wavelength combinations that could be used to construct calibration models that would give a high classification rate when applied to the validation set samples. With 10 wavelengths, >90% of the samples in the 7 validation sets, other than infested air-dried, were correctly classified. None of the wavelength combinations was successful in classifying infested air-dried kernels as nonmembers of the training set. Overall best results were obtained at 12 wavelengths. This combination of wavelengths gave a model that correctly classified 100% of sound, 93% of infested, 99% of sound air-dried, and 90% of sound kernels from 6 wheat varieties into their respective classes, but correctly classified only 55% of infested air-dried kernels. These results were very similar to classification results obtained with a 5-factor PCA model based on first-derivative spectra from 1100 to 1900 nm, except that the model with 5 PC factors gave 86% correct classification of infested air-dried kernels, compared with only 55% correct classification when the discriminant analysis was based on selected wavelengths.

The previous comparison of PCA factors of sound and infested kernels showed differences in spectral regions associated with lipids, protein, and phenolic compounds. Several of the discrete wavelengths selected (1200, 1360, 1440, and 1660 nm) are related to -C-H stretching or combination bands, such as those arising from lipids. Also, the 1420 nm wavelength is associated with an -OH stretch first overtone arising from phenolic compounds. Therefore, a number of the discrete

Table 7. Percentage of sound, infested, air-dried, and 6 varieties of wheat kernels classified correctly by discriminant analysis models based on Mahalanobis distances of selected wavelengths of NIR[a] spectra

No. of wavelengths	Sound cd[b] (25)	Infested cd (75)[c]	Sound cu[d] (75)	Infested cu (75)	Sound air-dried cd (75)	Infested air-dried cd (75)	Sound 6 variety cd (60)
6[e]	100	60	—	50	—	—	100
9[f]	100	63	100	50	99	56	88
10[g]	100	63	100	55	100	—	88
10[h]	100	90	100	93	97	48	90
11[i]	100	85	100	92	99	45	90
12[j]	100	93	100	93	99	55	90

[a] Near-infrared diffuse reflectance. Calibration file contained sound wheat kernels with crease down while scanning.
[b] cd, crease down while scanning.
[c] The number in parentheses is the number of spectra used in prediction set.
[d] cu, crease up while scanning.
[e] 6 selected wavelengths: 1200, 1300, 1320, 1360, 1660, 1880 nm.
[f] 9 selected wavelengths: 1240, 1300, 1320, 1360, 1400, 1420, 1440, 1680, 2220 nm.
[g] 10 selected wavelengths: 1240, 1300, 1320, 1360, 1400, 1420, 1440, 1680, 2220, 1200 nm.
[h] 10 selected wavelengths: 1240, 1300, 1320, 1360, 1400, 1420, 1440, 1680, 2220, 1880 nm.
[i] 11 selected wavelengths: 1240, 1300, 1320, 1360, 1400, 1420, 1440, 1680, 2220, 1880, 1660 nm.
[j] 12 selected wavelengths: 1240, 1300, 1320, 1360, 1400, 1420, 1440, 1680, 2220, 1880, 1660, 1200 nm.

wavelengths chosen appear to respond to the same chemical components as those observed in the PCA factors. That a higher classification rate for infested air-dried kernels was achieved with the PCA technique indicates that there must be a wavelength or wavelengths in the spectra that are responsible for recognizing these kernels. Therefore, the challenge now is to identify these wavelengths and include them in discrete-wavelength calibration models.

Initial attempts at NIR spectroscopic detection of internal insect infestation in wheat kernels were successful. PCA of NIR diffuse reflectance spectra from sound kernels was used to construct calibration models by calculation of Mahalanobis distances. A 5-factor PCA model using data from a first-derivative spectral transformation was the best model for correctly classifying sound kernels and kernels infested with late-instar granary weevil larvae. Calibrations using the spectral region from 1100 to 1900 nm were least sensitive to kernel moisture differences. Infested wheat kernels containing insect larvae at earlier stages of growth now need to be evaluated to find the earliest stage at which insects can be detected. The ability to detect larvae of other internal infesters, such as rice weevil and lesser grain borer, should also be evaluated. Finally, from a commercialization standpoint, it is advantageous to use filter-based instruments, which are much less expensive than scanning monochromator instruments, for this measurement. Rapid scanning systems are commercially available that allow a spectrum to be collected in less than 1 s, making the full-spectrum method practical from a time standpoint, but these systems are quite expensive. Therefore, continued emphasis should be given to identification of combinations of discrete wavelengths that are more robust. Also, automated sample presentation systems need to be developed to make this technique practical.

Acknowledgments

We thank John R. Pedersen of Kansas State University for providing the granary weevil larvae-infested wheat kernels and X-ray analysis equipment for this work.

References

(1) U.S. Department of Agriculture, Agricultural Research Service (1986) *Stored Grain Insects*, Agriculture Handbook No. 500, National Technical Information Service, Springfield, VA

(2) Arteman, R.L. (1981) *The Detection of Internal Insect Infestation in Grain*, USDA Fed. Grain Insp. Serv. Standardization Div. Inspect. Methods Res. Branch Rep. No. 81-22-1, The Service, Kansas City, MO

(3) Goossens, H.J. (1949) *Cereal Chem.* **26**, 419–420

(4) Nicholson, J.F., Akers, J.C., Harris, K.L., & Kurtz, O.L. (1953) *J. Assoc. Off. Agric. Chem.* **36**, 146–159

(5) Apt, A.C. (1952) *Northwest. Miller* **247** (23), 24

(6) Harris, K.L., Nicholson, J.F., Randolph, L.K., & Trawick, J.L. (1952) *J. Assoc. Off. Agric. Chem.* **35**, 115–158

(7) Dennis, N.M., & Decker, R.W. (1962) *J. Econ. Entomol.* **55**, 199–203

(8) Street, M.W., Jr, & Bruce, W.A. (1976) *Food Eng.* **48** (2), 94–96

(9) Hackman, R.H., & Goldberg, M. (1981) *Anal. Biochem.* **110**, 277–280

(10) Donald, W.W., & Mirocha, C.J. (1977) *Cereal Chem.* **54**, 466–474

(11) Chambers, J., McKevitt, N.J., & Stubbs, M.R. (1984) *Bull. Entomol. Res.* **74**, 707–724

(12) Kitto, G.B. (1991) *Assoc. Oper. Millers Tech. Bull.* **3**, 5835–5838

(13) Adams, R.E., Wolf, J.E., Milner, M., & Shellenberger, J.A. (1954) *Cereal Chem.* **31**, 271–276

(14) Webb, J.C., & Landolt, P.J. (1984) *J. Environ. Sci. Health, Part A, Environ. Sci. Eng.* **19**, 367–375

(15) Vick, K.W., Webb, J.C., Weaver, B.A., & Litzkow, C. (1988) *J. Econ. Entomol.* **81**, 1489–1493

(16) Hagstrum, D.W., & Flinn, P.W. (1993) *J. Econ. Entomol.* **86**, 1271–1278

(17) Hagstrum, D.W., Vick, K.W., & Webb, J.C. (1990) *J. Econ. Entomol.* **83**, 625–628

(18) Wilkin, D.R., Cowe, I.A., Thind, B.B., McNicol, J.W., & Cuthbertson, D.C. (1986) *J. Agric. Sci. Camb.* **107**, 439–448

(19) Lodder, R.A., Selby, M., & Hieftje, G. (1987) *Anal. Chem.* **59**, 1921–1930

(20) Gemperline, P.J., & Boyer, N.R. (1995) *Anal. Chem.* **67**, 160–166

(21) Mark, H.L., & Tunnell, D. (1985) *Anal. Chem.* **57**, 1449–1456

(22) Mark, H.L. (1987) *Anal. Chem.* **59**, 790–795

(23) Devaux, M.F., Bertrand, D., Robert, P., & Qannari, M. (1988) *Appl. Spectrosc.* **42**, 1015–1019

(24) Devaux, M.F., Bertrand, D., Robert, P., & Qannari, M. (1988) *Appl. Spectrosc.* **42**, 1020–1023

(25) Downey, G., Bertrand, D., & Kelly, P.M. (1990) *Appl. Spectrosc.* **44**, 150–155

(26) Lanza, E. (1983) *J. Food Sci.* **48**, 471–474

(27) Ciurczak, E.W., & Maldacker, T.A. (1986) *Spectroscopy* **1**, 36–39

(28) Shah, N.K., & Gemperline, P.J. (1990) *Anal. Chem.* **62**, 465–470

(29) Delwiche, S.R., & Norris, K. (1993) *Cereal Chem.* **70**, 29–35

(30) Rose, J.R. (1982) in Second Annual Symposium on Near-Infrared Reflectance Analysis, Technicon Instrument Corp., Tarrytown, NY

(31) Shenk, J.S., Landa, I., Hoover, M.R., & Westerhaus, M.O. (1981) *Crop. Sci.* **21**, 355–358

(32) Gemperline, P.J., Webber, L.D., & Cox, F.O. (1989) *Anal. Chem.* **61**, 138–144

(33) Scott, D.R. (1986) *Anal. Chem.* **58**, 881–890

(34) Derde, M.P., & Massart, D.L. (1986) *Anal. Chim. Acta* **184**, 33–51

(35) Wehling, R.L., Wetzel, D.L., & Pedersen, J.R. (1984) *J. Assoc. Off. Anal. Chem.* **67**, 644–647

(36) Pedersen, J.R. (1995) personal communication, Kansas State University, Department of Grain Science and Industry, Manhattan, KS

(37) Martens, H., Jensen, S.A., & Geladi, P. (1983) in *Proceedings of the Nordic Symposium on Applied Statistics,* O.H.J. Christie (Ed.), Stokkland Forlag, Stavanger, Norway, pp. 205–234

(38) *Official Methods of Analysis* (1984) 14th Ed., AOAC, Arlington, VA, sec **14.004**

(39) Rasco, B.A., Miller, C.E., & King, T.L. (1991) *J. Agric. Food Chem.* **39**, 67–72

(40) Osborne, B.G., Fearn, T., & Hindle, P.H. (1993) *Practical NIR Spectroscopy,* Longman, Essex, UK