Development of Short-Wavelength Near-Infrared Spectral Imaging for Grain Color Classification

Douglas D. Archibald[a], Chi N. Thai[b], and Floyd E. Dowell[c]

[a]Quality Assessment Research Unit, Agricultural Research Service, US Department of Agriculture, Athens, GA 30604-5677
[b]Biological and Agricultural Engineering, University of Georgia, Athens, GA 30602-4435
[c]Engineering Research Unit, Grain Marketing and Production Research Center, Agricultural Research Service, US Department of Agriculture, Manhattan, KS 66502-2796

ABSTRACT

Color class of wheat is an important attribute for the identification of cultivars and the marketing of wheat, but is not always easy to measure in the visible spectral range because of variation in vitreosity and surface structure of the kernels. This work examines whether short-wavelength near infrared (SW-NIR) imaging in the range 632 – 1098 nm can be used to distinguish different cultivars. The spectral characteristics of six hard white winter (HDWW) and hard red spring (HRS) wheats were first studied by bulk-sample SW-NIR reflectance spectroscopy using regression analysis to select appropriate wavelengths and sets of wavelengths. Prediction of percent red wheat was better if C-H or O-H vibrational overtones were included in the models in addition to the tail from the visible chromophore absorbance, apparently because the vibrational bands make it possible to normalize the color measurement to the dry matter content of the samples. Next, a reflectance spectral image of 640 x 480 spatial pixels and 11 wavelengths was acquired for a mixture of the two contrasting wheat samples using a CCD camera and a liquid crystal tunable filter (LCTF). The cultivars were distinguished in the image of principal component (PC) score number two that was calculated from the spectral image. The discrimination is due to the tail from the absorbance band that peaks in the visible. PC images 3 and 6 seem to arise mainly from O-H and C-H bands, respectively, and it is speculated that these spectral features will be important for generating multivariate models to predict the color class of grain. It is shown that the contrast between the red-wheat, white-wheat and background can be increased by applying histogram equalization and segmentation of the kernels in the images.

Keywords: wheat, red color, spectral image, short-wavelength NIR, CCD camera, LCTF, PCA

1. INTRODUCTION

There are five major classes of wheat produced in significant quantities in the United States, and although several of these products are headed for distinctly different markets and end-uses, they often pass through the same grain handling and storage facilities. Some types of wheat mixtures do not present a problem and may even be desirable, but certain 'contrasting classes' should be kept separate (Table 1). A small fraction of a contrasting class of wheat can have a large effect on the milling or baking quality of the wheat and consequently its value is downgraded on the market. Because of the small percentage differences that distinguish various grades, a large subsample, typically thousands of kernels, must be analyzed grain-by-grain in order to get meaningful count statistics. Kernel morphology, mechanical properties, color and composition, are single-kernel properties that can be are measured for this purpose. Color grading is currently performed by trained visual inspection. Red wheats arise from three color-genes that produce pigmentation in the seed coat. However, the differentiation of 'red' wheat from 'white' wheat is not always obvious, as the apparent color is a result of the combination of genetic and environmental factors. Researchers hope to develop a practical instrumental system for color determination: two options under development are vis-NIR spectroscopic analysis of bulk samples,[1] and sequential single-kernel vis-NIR reflectance.[2] Spectral-imaging technology has the potential for achieving the high speed and simple sample handling attributes of the bulk method, as well as the single-kernel classification capabilities of the sequential single-kernel instruments.

A very substantial research effort has been devoted to image analysis of grain with grayscale or color digital imaging.[3] The strong point of this technology is single-kernel morphological analysis. Color analysis has had success and is most valuable in the context of morphology. However, RGB-imaging encounters difficulties with vitreosity and the many potential environmental influences on pigment concentration and apparent color.[3]

Part of the SPIE Conference on Precision Agriculture and Biological Quality • Boston, Massachusetts • November 1998
SPIE Vol. 3543 • 0277-786X/99/$10.00

The studies presented here investigate short-wave (SW) NIR imaging. In the analytical chemistry community, the region from the red end of the visible spectrum out to 1100 nm is referred to as the SW-NIR or the Herschel-NIR region. In contrast 'NIR' is usually understood to cover the range 1100 – 2500 nm. SW-NIR imaging may have some advantages over color imaging for determining grain class:

- Better light penetration through surface contaminants and damaged surfaces;
- Avoidance of interfering visible absorption due to surface contaminants;
- Potential for the compensation of any effect of water content on apparent color;
- Spectro-chemical measurement of the organic matter content as the basis rather than optical scattering density;
- Possibility of measuring non-color-gene-related compositional characteristics that are indicative of grain class.

Whether or not any of these postulates have basis in fact, NIR spectral imaging technology will undoubtedly find many applications in agriculture and elsewhere. Multivariate spectroscopy for qualitative and quantitative quality assessment is one important technology to be adapted for this new hardware, and is a significant aspect of this study.

Table 1. Definition of contrasting wheat class and its use in wheat grade determination.

CLASS	CONTRASTING CLASS*
Hard Red Winter and Hard Red Spring	Durum, Hard White, Soft White, and Unclassed Wheat
Durum	Hard Red Spring, Hard Red Winter, Soft Red Winter, Hard White, Soft White, and Unclassed Wheat
Soft Red Winter	Durum and Unclassed Wheat
Hard White and Soft White	Durum, Hard Red Winter, Hard Red Spring, Soft Red Winter, and Unclassed Wheat
GRADE*	TOLERANCE FOR CONTRASTING CLASSES
U.S. No. 1	Equal to or Less Than 1 %
U.S. No. 2	Equal to or Less Than 2 %
U.S. No. 3	Equal to or Less Than 3 %
U.S. No. 4	Equal to or Less Than 10 %
U.S. No. 5	Equal to or Less Than 10 %

*As outlined on pages 12-1 and 13-32 of the USDA/FGIS Grain Inspection Handbook, Book II, Grain Grading Procedures.

2. EXPERIMENTAL

Three red and three white wheat samples were obtained that had previously been analyzed by colorimetry (personal communication, Sheldon Wishna)(Table 2).[4] SW-NIR reflectance measurements of these samples were made with an NIRSystems model 6500 spectrometer (Perstorp Analytical, Inc., Silver Spring, MD) using the small sample cup. The samples were repacked and scanned six times each. Chemometric analysis of this data was performed in MATLAB (The MathWorks, Natick, MA) programming environment using the PLS_Toolbox (Eigenvector Research, Manson, WA) and custom algorithms. Spectra images were collected with a hardware and software system designed and assembled at the University of Georgia (Figure 1).[5] The major components include the following: a Matrox Pulsar framegrabber (Matrox Electronic Systems, Ltd., Quebec, Canada); a National Instruments AT-MIO-16F-5 12-bit analog input/output board (National Instruments Corp, Austin, TX); A DVC model DVC-10 CCD monochrome camera (DVC Company, San Diego, CA); and a liquid crystal tunable filter (LCTF) for the range 632 – 1100 nm (Cambridge Research and Instrumentation, Inc.). Images were acquired at 640 by 480 pixels and 8-bits of luminosity resolution relative to white Spectralon® (Labsphere, North Sutton, NH). The system has feedback control so that the gray level of a region of interest (ROI) on the reference surface was held constant (to 200) regardless of the wavelength setting of the LCTF. Two 300 W tungsten halogen lamps with glass diffusers were used to illuminate the sample (Figure 1). For the spectral image scans, three-quarter inch strips of the wheat samples were placed on a single-strength glass plate over the Spectralon® reference with strips of empty surface and black electrical tape to serve as intensity references. All images shown are for a 50:50 mixture of HRS 61107 and HDWW 61435 (Table 2). The focus of the camera lens was adjusted with the LCTF set to 806 nm, and the 11 wavelengths of the spectral image are 632, 664, 696, 758, 834, 876, 894, 916, 950, 990 and 1024 nm. A wavelength of 1064 nm was also attempted, but the instrument was not sufficiently sensitive. Wavelength images beyond 900 nm appeared to be of lower quality than other images because of the increased noise and poor focus. Image TIF files were analyzed with the MATLAB Image Processing Toolbox.

Table 2. Wheat samples available for study.

Number	Class*	Type	% Red by Colorimetry	% White by Colorimetry	Visual Observations
61107	HRS	Red	93.3	6.7	clean, uniform, smaller,some darkened germ
60053	HRS	Red	79.9	20.7	clean, uniform, smaller
61102	HRS	Red	64.3	35.7	clean, uniform, larger
61435	HDWW	White	3.0	97.0	clean, uniform, larger, some darkened germ
59289	HDWW	White	17.3	82.7	clean, uniform, larger, many darkened germ
59286	HDWW	White	26.0	74.0	dirty, damaged, larger, many darkened germ

*HRS is hard red spring wheat and HDWW is hard white winter wheat.

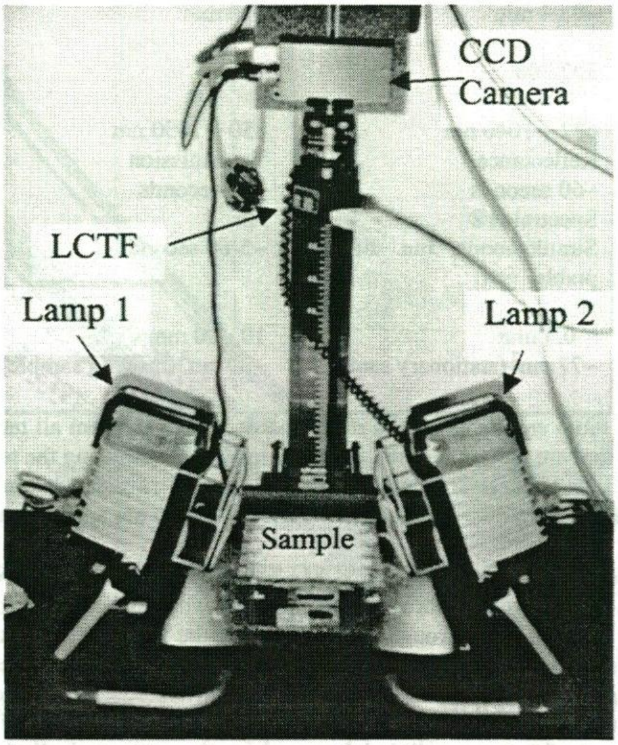

Figure 1. Photograph of the SW-NIR spectral imaging system with a wheat sample in place. Two 300-W tungsten halogen lamps illuminate the sample from oblique angles.

3. COMPARISON OF INSTRUMENTATION FOR SW-NIR IMAGING AND CONVENTIONAL NIR

It is instructive to compare the spectrometer used in this study (a spectral imaging system) to NIR spectrometers that have been successfully deployed for quantitative agricultural product analyses. Table 3 is a listing of some of the major characteristics of these systems. The low A/D resolution of he SW-NIR imager is the first difference that the spectroscopist will notice. Many quantitative NIR analyses use every bit of that range. The problem is further compounded by the fact that the SW-NIR will generally require a higher instrumental S/N and/or greater sample interactance to be able to use the very weak NIR absorbances that occur in this region. However, the total performance of an instrument is also greatly affected by sampling noise, drift between sample and reference measurements, and the degree of surface reflectance measured by the detector. In these areas the spectral imaging system may have some advantages because of the huge number of spatially resolved channels. For example, some channels can be used for a simultaneous reflectance reference; spatially differentiated surface reflectance can be identified and avoided; and the imaging system is well suited to sampling a large sample surface so that the effect of heterogeneity can be quantified. The current spectral imaging system is similar to the spectrometer systems in some important respects: spectral resolution, acquisition time, and sampling area. Moreover, it appears that many of the current limits to spectral range, detector A/D resolution, and noise are not due to physical limitations, but instead are related to engineering, manufacturing, and economic considerations.

Table 3. Comparison of the SW-NIR spectral imaging system of this study with typical characteristics of conventional NIR spectrometers used for agricultural analysis

Characteristic	SW-NIR Imaging System*	SW-NIRT Analyzer**	NIRR Analyzer**
Wavelength Selection Method	LCTF	Scanned diffraction grating	Scanned diffraction grating
Spectral Resolution	10 nm	6 nm	10 nm
Infrared Source	2x300 W Tungsten Halogen	25 W Tungsten Halogen	25 W Tungsten Halogen
Detection System	Silicon CCD Array	Silicon Photodiode	Single Element Lead Sulfide
Detector Element Size	~0.025 mm square	~1 mm square	4x (5 mm square)
Detector A/D Resolution	256	65,536	65,536
Size of Sampling Area as Projected onto the Detector	~0.04 mm^2	~10 mm^2	~30 mm^2
Number of Spatial Sample Positions Recorded	307,200	1	1
Spectral Range (practical)	632 – 1040 nm	850 – 1050 nm	1100 – 2500 nm
Mode of Operation (typical)	Reflectance	Transmission	Reflectance
Spectral Acquisition Time	~60 seconds	~60 seconds	~30 seconds
Reference Material	Spectralon®	Air	Ceramic
Elapsed Time from Reference to Sample Measurement	Simultaneous, but different pixels	~5-60 seconds	~5-60 seconds
Typical Sampling Thickness	< 0.5 mm	10 - 20 mm	0.1 – 3 mm
Typical Sampling Area (total)	~77 cm^2 (stationary sample)	~30 cm^2 (moving sample)	~60 cm^2 (moving sample)

*This is meant to describe the setup employed in this study and does not represent all technology that is available. Many of the values are approximations that are only to be used for the purpose of contrasting the technologies.
**These types of systems have been successfully applied over the last 5 to 15 years, but are not intended to represent the wide variety of NIR instruments that are currently available. Many of the values are approximated.

4. SELECTION OF WAVELENGTHS USING SW-NIR REFLECTANCE OF BULK SAMPLES

By visual inspection, 'red' wheat has a reddish-brown seed coat and 'white' wheat has a pale yellow seed coat. The apparent color saturation is very much reduced by the roughened surfaces of the kernels. This can be demonstrated by the dramatic increase in color saturation that occurs upon hydrolyzing the surface attachment of the coat by soaking kernels in 5 % NaOH for 30 – 60 minutes. The major absorbance band has its peak in the visible and the tail of this band is observed in the SW-NIR (Figure 2). A complicating factor for observing the seed-coat pigmentation in this set of samples is the presence of kernel damage due to shriveling or the occasional dark brown stains that were typically located on the wheat germ (Table 2). The major type of variance among bulk reflectance spectra is offset due to differences in light scattering (Figure 2). This effect will also occur in spectral images, because it is partially due to kernel orientation and surface roughness. As a result of this phenomenon, prediction of the slope due to color using two wavelengths (or two regression factors) is much better than using a single wavelength. Additionally, it was observed that slope by itself is not the best measure color; the error in predicting red color is lower if the NIR bands around 916 and 1000 nm are included in the model. Respectively, these bands are the C-H stretch 3rd overtone vibration (from lipid, protein, carbohydrate) and the O-H stretch second overtone (primarily hydrated starch).[6] The beneficial effect of the vibrational overtone absorptions is probably due to the normalization of the color signal to the dry matter basis. Indeed, this may explain other peaked correlations to wheat color that have been reported for the SW-NIR.[1]

Twenty-nine wavelengths were chosen for further study (Figure 3(A)); the selections were the peaks, valleys, zero-crossings and inflections observed in regression vectors for full-spectral models predicting red color of wheat. Twenty-nine is a small enough set that all-possible combinations of up to 6 wavelengths could be tested by multiple linear regression for their ability to predict the percent red color of the sample (Table 2). The results of these computations are presented in Figure 3(B), Figure 4, and Table 4. Sets of two wavelengths perform well by employing various wavelengths on the tailing of the color absorption out to 876 nm (Figure 4(A)). The best performing 3-wavelength models use a wider range of the spectrum, and tend to include two wavelengths on the strongest NIR band (e.g., 990 and 1024 nm) and the third in the color-affected region (Figure 4(B)). The larger sets of wavelengths continue this pattern, but probably don't improve performance much over a 3-wavelength model (Table 4). Three-wavelength models make sense in the following way: one for correction of the scatter offset, one for measuring the quantity of pigmentation, and the third for measuring the quantity of dry matter under observation.

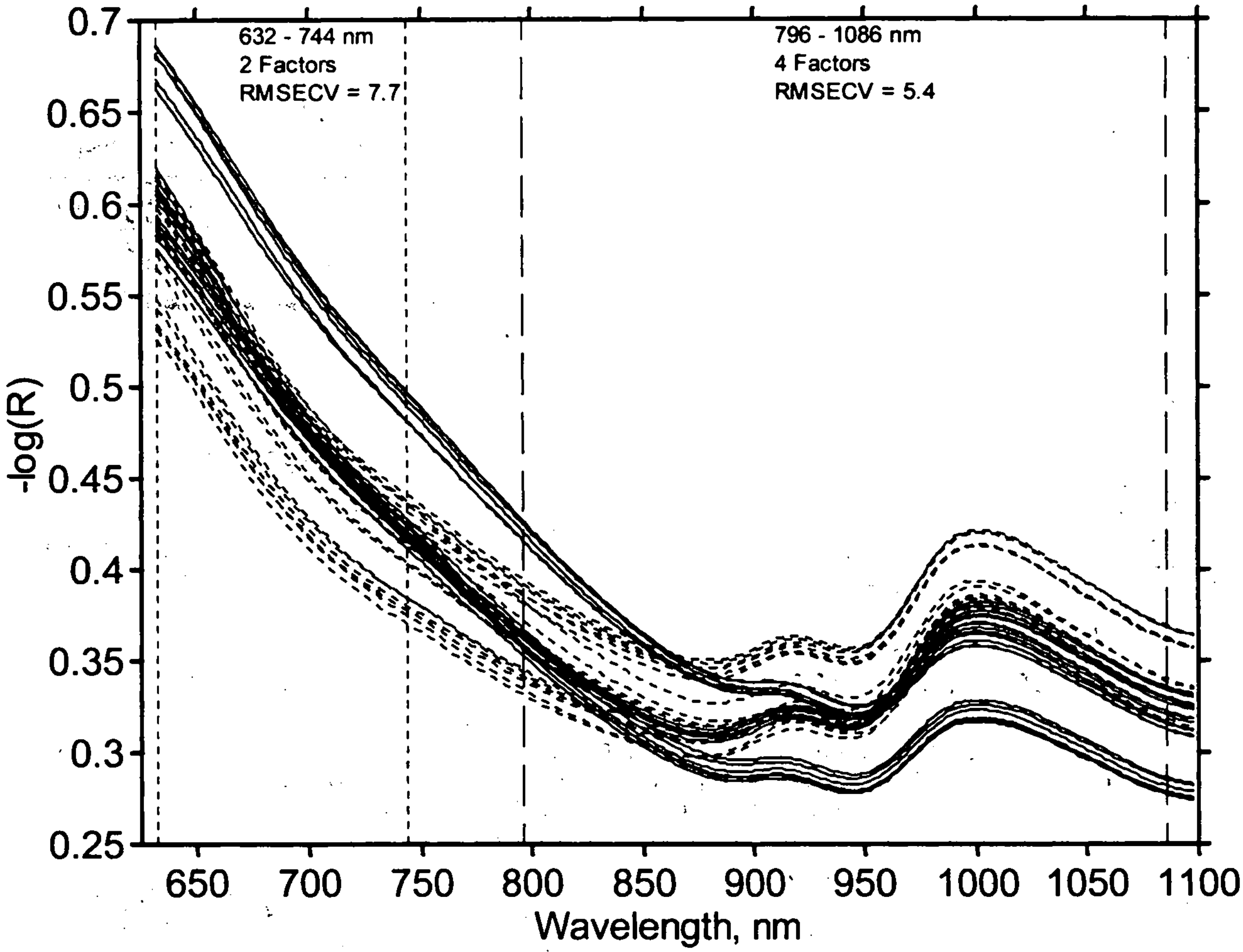

Figure 2. NIR reflectance spectra of three HRS (solid lines) and three HDWW (dotted lines) wheats: the samples were scanned on an NIRS model 6500 spectrometer and were repacked six times each. The vertical lines and annotation indicate the optimal spectral regions for prediction of red color using either a two-factor (dotted lines) or four-factor partial least-squares regression model (dashed lines). Cross-validation for estimation of the root mean-squared error (RMSECV) of prediction of percent red color was done with one sample (6 spectra) for each data split.

Table 4. Optimal wavelength sets for determination of % red wheat from bulk sample SW-NIR by applying multiple linear regression to all possible combinations of a subset of 29 wavelengths across the spectral range 632 – 1098 nm.

# Wavelengths per Combination	RMSEC, % Red	Best Wavelength Sets
1	21.7	684
2	6.8	722, 834
3	2.8	684, 990, 1024
4	2.7	684, 990, 1030, 1064
5	2.5	696, 906, 916, 990, 1024
6	2.3	722, 894, 906, 916, 990, 1030

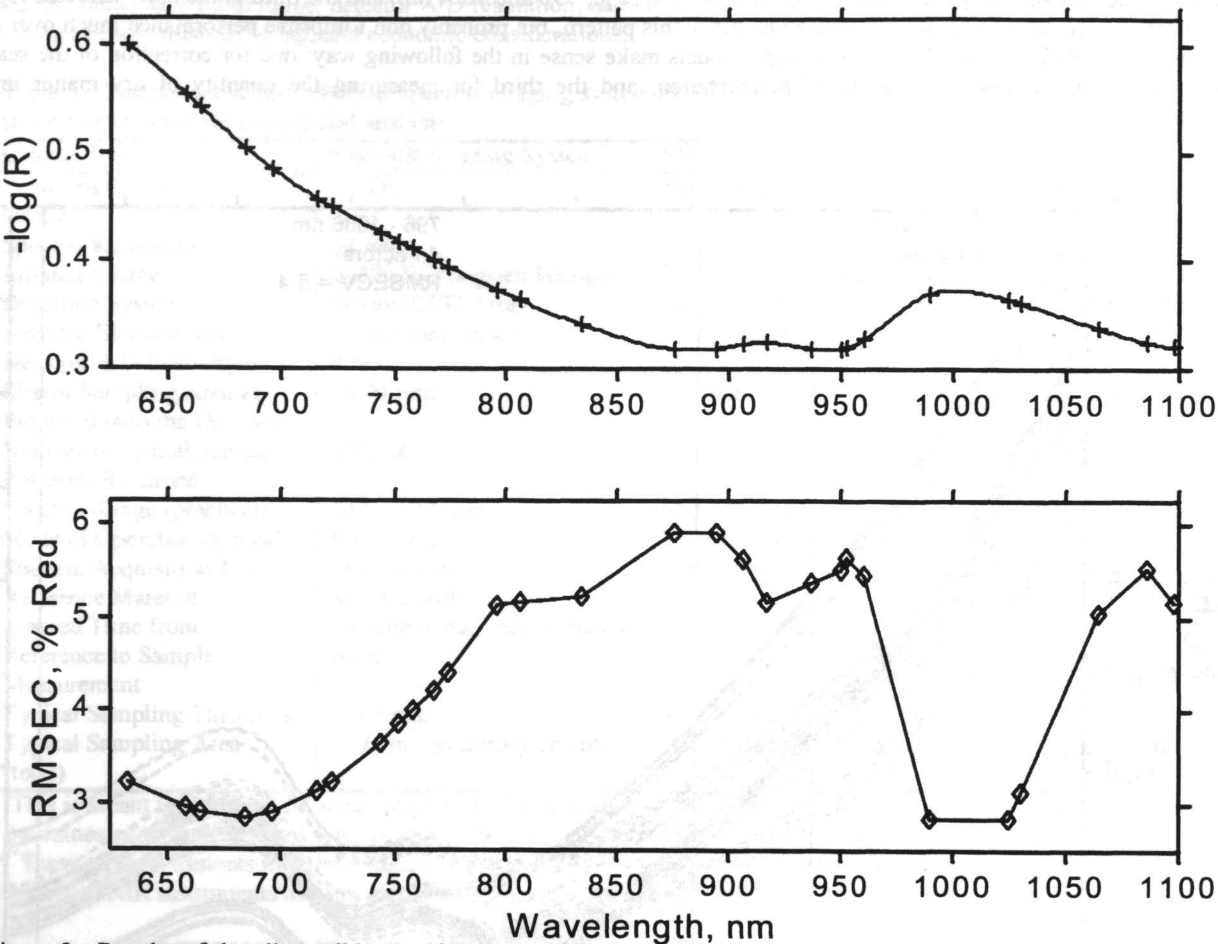

Figure 3. Results of the all-possible combinations search for wavelengths to predict the red color of wheat samples: (A) mean reflectance spectrum of the bulk wheat sample with markers indicating the set of wavelengths chosen for the search; (B) minimum model error for 3-wavelength sets containing the indicated wavelength.

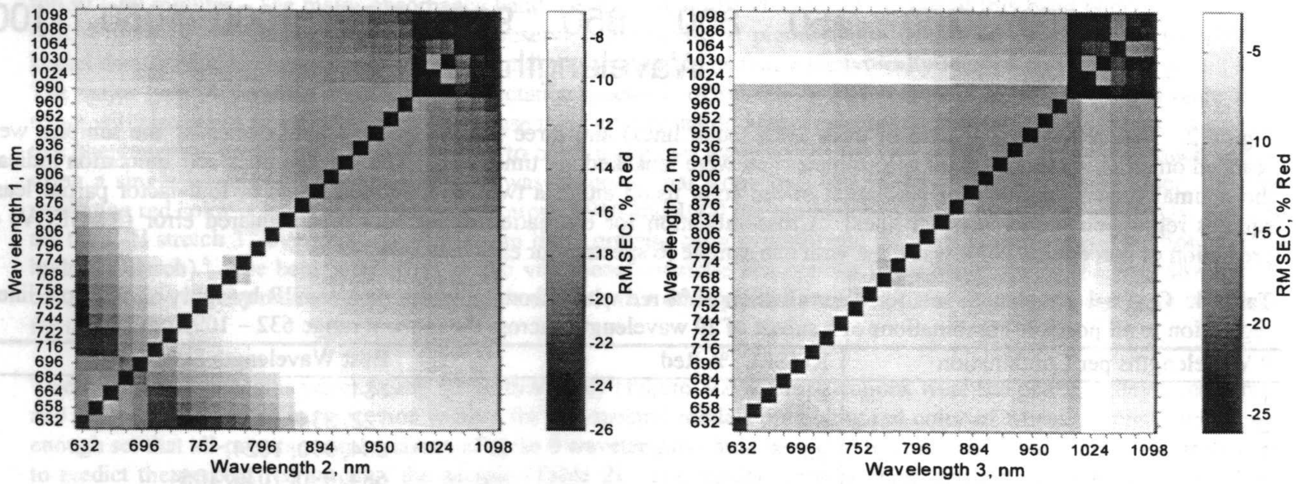

Figure 4. Error surfaces for use of multiple linear regression to predict percentage red color from all-possible combinations of the select set of wavelengths with different set sizes: (A) two-wavelength sets, and (B) three-wavelength sets that included 990 nm. The minimum RMSEC for two-wavelength sets was 6.8, while that for the three wavelength set was 2.8.

5. ANALYSIS OF SW-NIR SPECTRAL IMAGES

The spectra from a set of 11 wavelength images from 632 nm to 1024 nm were analyzed with principal component analysis (PCA) without any preprocessing of the data. PCA decomposes the set of 307,200 11-point spectra into linear combinations of orthogonal PCs. Eight PCs were calculated, and the PC scores were examined as images, four of which are displayed in Figure 5. The scores themselves have no units, but the relative score values are an indication of factors that are varying in the data. PC-factor loading vectors can be examined to understand the spectral changes that produce the image contrast (Figure 6). Loading vector plots can be though of as multivariate difference spectra for the underlying phenomenon. The first PC loading is the average spectrum and the first score image is the average of all wavelength images. The pixel values in the wavelength images are fractional reflectance (R) values; if transformed as –log(R), PC1 of Figure 6 is similar to the average bulk sample spectrum in Figure 3(A). The lesser attenuation at 1000 nm may be due to the image reference having included O-H attenuation by the supporting glass plate.

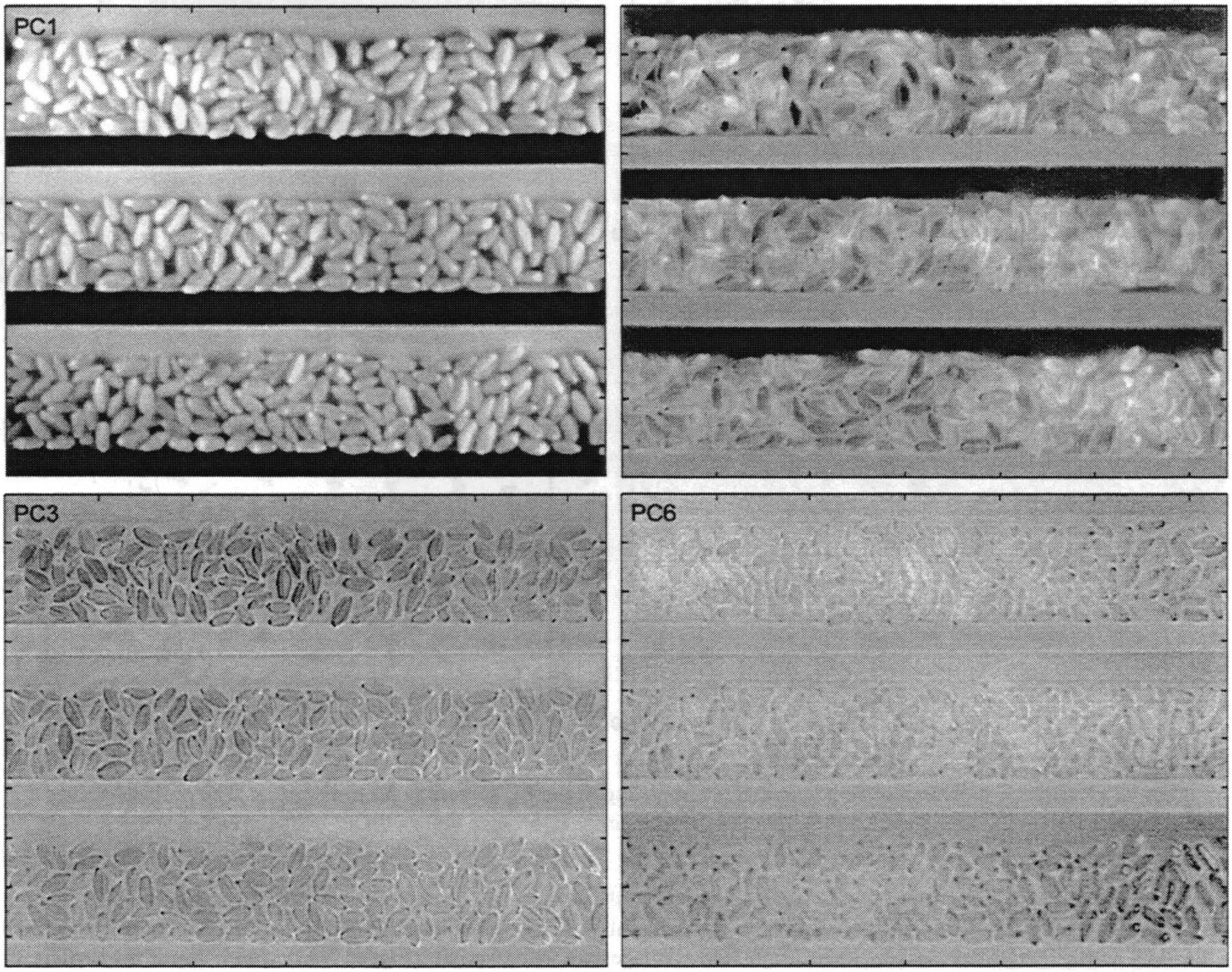

Figure 5. Principal component score images for an 11-wavelength SW-NIR spectral image of a 50:50 mixture of HRS and HDWW wheat. The horizontal stripes between the grain stripes are black and white reference surfaces.

The PC2 score image differentiates many of the HRS and HDWW kernels (Figure 5), and the loading plot shows that this is primarily a result of a slope in the near-visible end of the spectrum (Figure 6). PC3 appears to have a substantial contribution from the O-H overtone band (Figure 6), while PC6 contrast is generated from variation in the C-H overtone region (Figure 6). Neither the PC3 score image nor the PC6 score image shows much contrast between the two kinds of wheat, and moreover, the surface features are more subdued than many of the wavelength images (not shown). PC4 and PC5 seemed to be due to uneven illumination along the two spatial dimensions (data not shown). PC7 appeared to highlight saturated pixels, and PC8 was white noise (data not shown). Very similar results were obtained if the reference surfaces were omitted from the spectral image prior to PC analysis. Thus, C-H absorption from the black reference surface did not have any effect because the black attenuation was even across the spectra range.

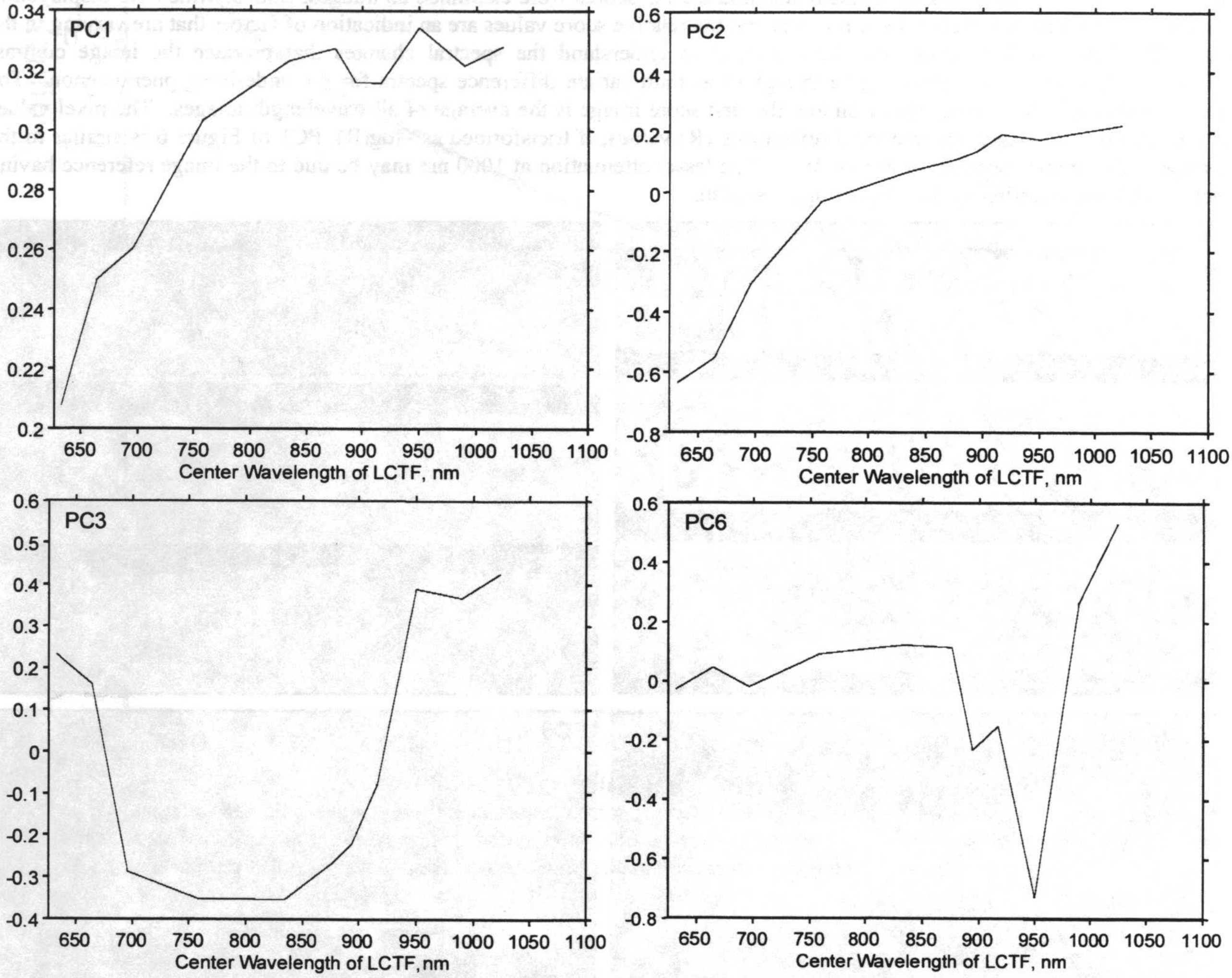

Figure 6. Principal component loading vectors corresponding to the PC score images in Figure 5. The ordinate values are loading weights, which are linear combinations of the measured reflectance values for all image pixels.

PC score images do not necessarily have the most natural polarity. The negative of PC2 scores produces an image in which the dark kernel images are red kernels and light kernel images are white kernels (Figure 7(A)). For many of the kernels, the red-white contrast can be further improved by histogram equalization Figure 7(B). Finally, binarization was applied to one of the wavelength images (Figure 7(C)) and this was used to code the background as gray (Figure 7(D)). The quality of the segmentation is not great at this date. Improvements in this area will surely come when: 1) a better background is used, 2) interpolation is used to correct the uneven illumination, and 3) morphological algorithms are used to isolate kernels. However, a simple sum of luminosity values for each isolated kernel image may not be the best way to measure color class. This is because the dark germ stains were detected on some white kernels (Figure 7(B)). Therefore, a classification algorithm may need to use the kernel morphology to select the most advantageous site to sample the kernel color. This capability is a distinct advantage over bulk or single-kernel spectroscopy, where it is difficult or impossible to control the measurement site on the kernel.

The wavelength sets in Table 4 have yet to be analyzed. Moreover, more sophisticated analysis needs to be done. These preliminary results indicate that SW-NIR can generate contrast to distinguish two classes of wheat. However, this contrast is primarily due to one type of spectral feature. Multivariate training algorithms need to be applied to utilize multiple spectral features to classify or quantify the color of each kernel. The PC analysis also demonstrated spectral sensitivity and discrimination of features related to dry matter content, C-H content, dark stains, surface roughness, and mirror-like surface reflections. Multivariate models will be able to utilize the signatures of these features to better predict the color class of each kernel in a spectral image.

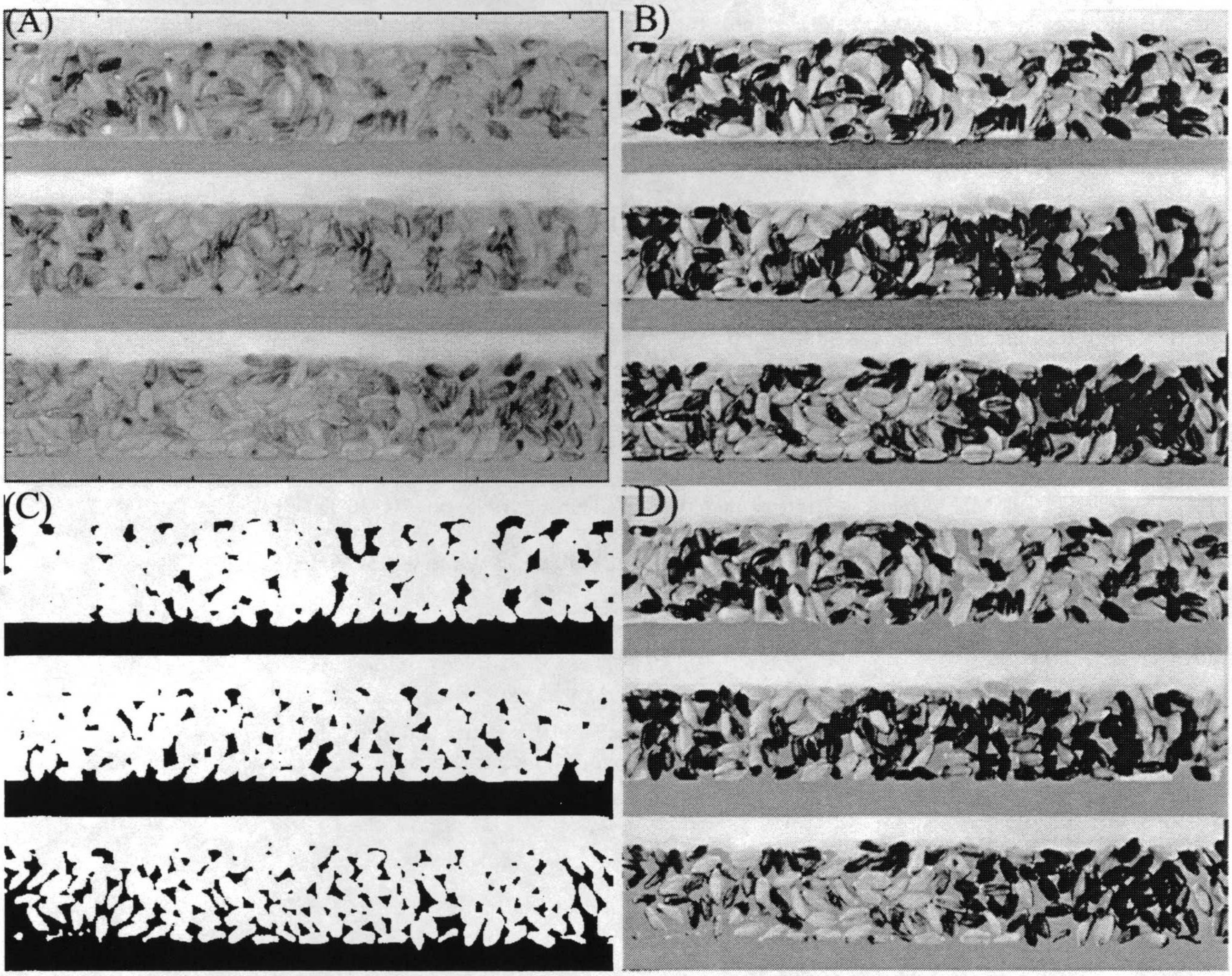

Figure 7. Enhancement and processing of images to improve discrimination of red wheat, white wheat and background: (A) negative image of PC2 scores; (B) contrast enhancement of (A) by histogram equalization to match a flat histogram with 64 bins; (C) segmentation of a portion of the background by binarization of the 834 nm image frame using a grayscale threshold of 30 %; and (D) contrast enhanced and segmented image with the background gray level set at 50 %.

6. CONCLUSION

Viewed as just a spectrometer, the performance of a current short-wavelength NIR spectral imaging system is substantially poorer than conventional (non-imaging) NIR spectrometers. Nonetheless, useful spectroscopic discriminations can be performed with the system. Hard red spring and hard white winter wheats were contrasted by multivariate analysis of reflectance spectra in an 11-wavelength spectral image spanning the range 632 – 1024 nm. Image processing further improved the differentiation of the cultivars, and patchy discoloration of white kernels can also be detected. Multivariate analyses also demonstrated contrast due to O-H and C-H vibrational overtones, uneven illumination, and saturated pixels.

Analysis of bulk samples showed that vibrational overtones improve models for prediction of color class. This suggests that a training approach to multivariate classification will produce a greater ability to differentiate color class of wheat kernels by use of SW-NIR spectral imaging.

7. ACKNOWLEDGEMENTS

We thank Sheldon Wishna of USDA-GIPSA-TSD for providing wheat samples with colorimetry values, and Jason C. Grant, formerly of UGA, for help with operation of the UGA SW-NIR imaging system. This research was supported by USDA-ARS CRIS project 6612-4400-017-00D, "Spectroscopic Sensors and Multivariate Calibration Methods for Quality Assessment in the Agricultural and Food Industries."

8. DISCLAIMER

Reference to company or trade names is for descriptive purposes only and does not imply endorsement of the United States Department of Agriculture.

9. REFERENCES

1 J.A. Ronalds and A.B. Blakeney. Determination of Grain Colour by Near Infrared Reflectance and Near Infrared Transmittance Spectroscopy. In: *Leaping Ahead with Near Infrared Spectroscopy*, eds. G.D. Batten, P.C. Flinn, L.A. Welsh, and A.B. Blakeney. North Melbourne, Victoria, Australia: NIR Spectroscopy Group, Royal Australian Chemical Institute, 1995. pp. 148-153.

2 D. Wang, F.E. Dowell, and R.E. Lacey, Single Wheat Kernel Color Classification Using Near-Infrared Reflectance Spectra *Cereal Chem.* (submitted 1998).

3 H.D. Sapirstein. Variety Identification by Digital Image Analysis. In: *Identification of Food-Grain Varieties*, ed. C.W. Wrigley. St Paul, MN: American Association of Cereal Chemists, 1995. pp. 91-130.

4 S. Wishna, Using Single Kernel Colorimetry to Differentiate Red and White Wheat Varieties : ASAE Paper 973024, ASAE, 1997, St. Joseph, MI.

5 M.D. Evans, C.N. Thai, and J.C. Grant, Development of a Spectral Imaging System based on a Liquid Crystal Tunable Filter *Trans.ASAE*, 1998 (In Press).

6 B.G. Osborne, T. Fearn, and P.H. Hindle. Theory of near infrared spectrophotometry. In: *Practical NIR spectroscopy with applications in food and beverage analysis.*, Anonymous Longman Scientific & Technical, 1993. pp. 13-35.