Sens. & Instrumen. Food Qual. (2009) 3:89–99
DOI 10.1007/s11694-008-9063-z

ORIGINAL PAPER

Improved discrimination of soft and hard white wheat using SKCS and imaging parameters

Thomas C. Pearson · Daniel L. Brabec · Hulya Dogan

Received: 8 February 2008 / Accepted: 24 March 2008 / Published online: 30 October 2008

Abstract Natural variation of hardness of wheat kernels often results in overlapping hardness indices (HI) distributions between hard and soft classes as measured with the single kernel characterization system (SKCS). This is particularly true for the case of the hard white (HW) and soft white (SW) wheat classes. To address this problem, a color camera was incorporated into the SKCS system so that color and kernel size data could be combined with SKCS measurements for classification purposes. Samples of hard red (HR), soft red (SR), HW, and SW wheat were classified using the SKCS system with and without the camera and results compared. Using the camera system, errors for separating HW from SW classes were reduced to less than 5%, as compared to 17.1% using SKCS alone. Furthermore, improved data processing applied to the low-level data currently produced by the SKCS system led to greater than 50% reduction in classification errors between SW and HR as compared to using HI data alone. Similar improvements in classification accuracies for 300-kernel sample containing mixtures of SW and HW were also achieved. The 300 kernel sample classification is usually what inspectors and grain traders use to determine sample purity rather than individual kernel results. The techniques developed should aid grain inspectors in properly identifying mixtures of these two classes. Unfortunately, for the SR and HR classes, incorporating the camera data decreased classification accuracy while increasing the complexity of the system. However, SR and HR classes can be adequately distinguished with the SKCS in its current form.

Keywords SKCS · Hardness · Image · Camera

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

T. C. Pearson (✉) · D. L. Brabec
USDA-ARS-GMPRC, 1515 College Ave., Manhattan, KS 66502, USA
e-mail: Thomas.pearson@gmprc.ksu.edu

H. Dogan
Department of Grain Science and Industry, Kansas State University, Manhattan, KS 66502, USA

Introduction

U.S. wheat standards classify grain according to several distinct features such as hardness and color [1]. In the past, certain varieties of wheat have been problematic for human inspectors to correctly classify. For example, the Arkan variety, a hard red winter (HRW) wheat grown in a hard wheat region, was frequently classified as a soft red (SR) wheat by official grain graders [2, 3]. Consequently, USDA researchers developed an instrument to objectively measure the hardness of wheat samples [4]. The single-kernel characterization system (SKCS) was designed to test 300 wheat kernels in about 3 min and determine the sample average and standard deviation for several physical parameters. The main parameters were hardness index (HI), weight, moisture, and kernel thickness. The SKCS was calibrated to give a mean HI of 75 for five reference hard samples and a mean HI of 25 for five reference soft samples [4]. However, the individual HI for the soft-wheat reference samples ranged from 24 to 37 hardness units, while the individual HI for the hard-wheat reference samples ranged from 63 to 86 hardness units [5]. However, since the development of the SKCS and selection of the NIST samples, single-kernel HI from some differing hardness classes now may overlap, making it harder to determine if a sample is pure hard, soft, or a mixture of the two classes.

Since the mid 1990's, steadily increasing production of hard white (HW) wheat has increased the importance of this class. Production of HW wheat increased nearly threefold between 2002 and 2005 (from 0.35 to 0.93 million metric tons) [6]. Grain inspectors have observed that in the U.S. Pacific Northwest (PNW) region, discriminating soft white (SW) from HW wheat has become increasingly difficult; i.e. the SKCS HI of PNW HW varieties more frequently resemble SW varieties, and vice versa. Many of the SW varieties, as well as varieties of wheat which contain both hard and soft wheat genetics, are similarly difficult to classify. Some of the soft varieties are sold into traditional soft wheat markets, although their SKCS HI scores are higher than those expected from older white wheat varieties and soft red winter (SRW) classes [7].

The SKCS HI is computed on the basis of kernel weight, moisture, and histogram features of slopes of the crush signal [4]. This method has worked well for discriminating HRW from SRW as well as for determining some milling qualities. However, the computations traditionally used to determine HI do not consider the different phases of kernel crushing as it is first compressed, fractured, and then ground into finer particles before exiting the SKCS. The forces experienced by the kernel as it is crushed exhibit three distinct phases. The first phase describes the exerted force up to the time when the kernel fractures, generally noted as a very narrow but high-magnitude peak early in the crush signal generated by the SKCS. It is expected that soft kernels crumble into smaller pieces during this initial kernel fracture, and thus their initial peak is generally smaller than that of hard kernels. During the second phase, which follows the initial fracturing, the kernel fragments undergo further crushing as the gap between the rotor and crescent of the SKCS decreases. The crush signal during this phase shows a smaller magnitude than during the first phase, indicating less resistance by the kernel to further compression. In the third phase of crushing, small fragments are slowly broken down until they exit the rotor/crescent. In this final phase, the crush force increases at a higher rate for hard classes, leading to a higher and broader peak. This is likely caused by the more moderately-sized particles that remain from the hard classes. In contrast, the small particles from soft classes do not require as much crush force to break down, leading to a lower and more lopsided peak.

SKCS crush force signals have been used to predict single-kernel particle size distribution features measured from the crushed kernel [8]. The predicted particle size feature was used to help discriminate between hard and soft wheat, including HW and SW wheat. Results indicated that classification errors could be reduced by approximately 50% using these techniques. However, the sample set used in this study was small, only a narrow range of moisture contents was studied, and only SKCS parameters were used. No images of the kernels were collected to determine if they could lead to decreased classification errors.

Kernal shape features extracted from digital images have been shown useful for discriminating between wheat varieties [9]. However, computing limits at the time of this study limited the sample size to only 30 images. Image analysis of wheat has been extensively studied by many researchers [10, 11], and the speed of image analysis has increased dramatically while the cost of imaging and computing hardware has decreased.

The first objective of this work was to test alternate methods of analyzing crush-force signals and available SKCS parameters to improve discrimination between hard and soft wheat across a range of moisture contents. More specifically, the objective was to investigate features in the SKCS crush-force signal that permit the characterization of the three phases of kernel crushing. The second objective was to determine whether kernel image data combined with SKCS data could improve discrimination between wheat classes. Improved discrimination between HW and SW wheat was specifically explored.

Materials and methods

Wheat samples

The sample set (Table 1) comprised hard red (HR), SR, HW, and SW wheat. All samples were obtained from breeders or NIST samples that were known to be pure. Thus, the hardness classification for each sample was used as the "ground truth" for the training and checking classification schemes. Each of these wheat subclasses were represented by three varieties, some winter and some spring. Samples (300 kernels) from each of the 12 varieties were prepared at moisture levels of 10.0%, 12.2%, and 14.2% (all +/−0.2%), for a total of 36 samples. Moisture tempering were made by adding an appropriate amount of water to a sample and tumbling it for 24 h. Each sample was tested in the SKCS unit in two groups of 150 kernels. Of the 10,800 samples processed, data from 396 individual kernels were rejected because either the SKCS did not record the data due to its own error checking, or an image of the kernel was not captured, or both. Thus, data from a total of 10,404 kernels were analyzed.

SKCS instrument

The SKCS crush-force signals were generated by a Perten SKCS instrument (model 4100, Perten Instruments, Springfield, IL). The instrument was controlled by an

Table 1 Sample list specifying wheat subclass, variety, average hardness index, and source

Subclass	Spring/winter	Variety	Average hardness	Source
Hard red	Winter	2180	89	Kansas State Univ. 1994
	Spring	Len	89	GMPRC/NIST 1994
	Winter	Tam 105	76	GMPRC/NIST 1994
Soft red	Winter	Caldwell	5	USDA/Wooster, Ohio 1994
	Winter	Cardinal	23	GMPRC/NIST 1994
	Winter	Titan	24	GMPRC/NIST 1994
Hard white	Spring	Blanca grande	59	Washington Wheat Comm. 2005
	Spring	ID3775	65	Washington State Univ. 2004
	Spring	Klasic	59	Washington State Univ. 2004
Soft white	Winter	Eltan	23	Washington State Univ. 2004
	Winter	Madsen	39	Washington State Univ. 2004
	Winter	Tres	37	GMPRC/NIST 1994

external laptop (PC#1, Fig. 1) via cabling between the laptop com port and the SKCS internal computer. The raw SKCS crush force and conductivity signals were collected for each kernel. Basic SKCS measurements such as weight, hardness, moisture, and kernel thickness were also collected, in addition to the "low level" SKCS parameters (crush area, Gompertz a, Gompertz b, and peak force) as described in [4])

Image data were collected using a separate computer (PC#2, Fig. 1). These data were synchronized with the SKCS data using a feeder trigger signal and a USB interfacing circuit. When triggered, PC#2 collected an image from the camera of the kernel in the weighing bucket (Fig. 1). The color camera was mounted above the weighing bucket as shown in Fig. 2. Details of the image hardware, acquisition, and processing, including segmentation of the kernel from the weigher bucket background and extraction of morphological features from the kernel have been previously reported [12]). It should be noted that while two computers were used in this experiment, it is possible to perform all necessary tasks with one computer by integrating the SKCS software and image collection and processing software.

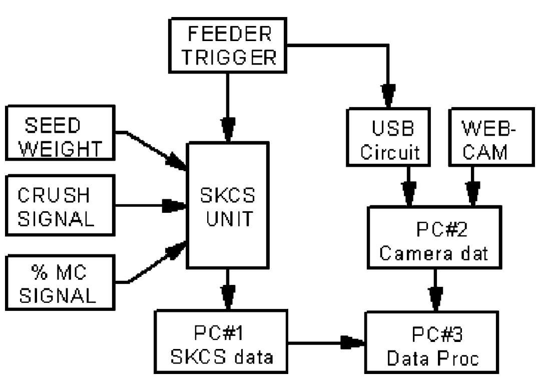

Fig. 1 Flowchart for data signal collection

After the wheat samples were tested, the crush signals were further evaluated, and additional descriptive parameters were determined. Images were analyzed to determine size and color parameters. The SKCS data and the image data were collated, and discriminate analysis was performed to classify kernels into the proper class. Figure 2 shows the flowchart for data collection and processing.

Alternative crush signal features for discrimination between hard and soft classes

The SKCS system produces a standard hardness index and parameters. These parameters do not satisfactorily discriminate between HW and SW wheat. Several signal

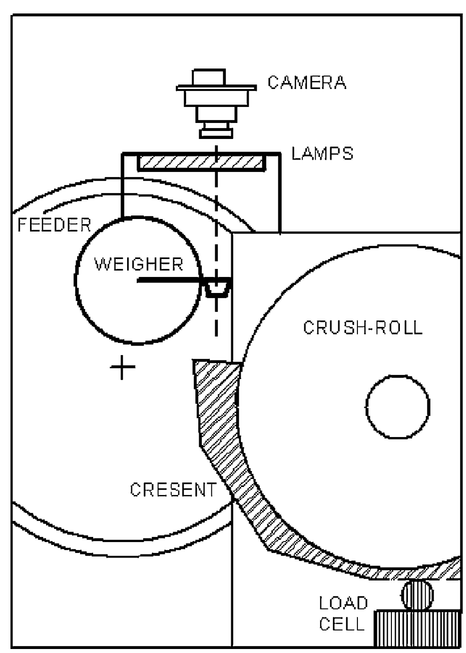

Fig. 2 Schematic of camera relative to the weigher and crush-roll

processing strategies were applied to the raw crush profile to improve discrimination. First, combinations of the normal parameters and 'low-level' parameters were evaluated. Next, the crush signals of each kernel were analyzed, and several additional signal features were extracted. These features were taken from portions of the crush signal where Gaussian features, variance features, and signal frequency features were compiled. Finally, image features were included for each kernel. Feature selection schemes were employed, and several discriminant analysis schemes were evaluated for classification accuracy. The parameters and features, the selection schemes, and the classification schemes are discussed in further detail below.

SKCS computed parameters

Initially, the normal parameters reported by the SKCS for every kernel were used as potential discriminating features. These basic parameters are hardness, moisture, weight, and diameter. Next, the "low level" parameters were included in analyses. These include features, combinations of features, and transformations of features from the basic SKCS signals. Example "low level" parameters include peak crush force, logarithm of conductance, and parameters from the derivative of the crush signal [4].

Crush signal Gaussian modeling

A Gaussian model describes the bell-shaped curve used commonly in statistics to describe normal population distributions. While the entire SKCS crush signal does not resemble a bell-shaped curve, portions of the crush signal profile have some similarity. The basic approach in this analysis was to use the summation of three bell curves to describe the crush force profile as shown in Fig. 3. The Gaussian parameters (amplitude and width) were added to all of the other low-level data generated by the SKCS 4100.

Normalization of the crush signals was performed prior to modeling. This reduced the effects of kernel size and moisture on the magnitude of the crush force signals, because larger kernels produce crush signals of higher magnitude, longer duration, and higher moisture kernels produce crush signals of comparatively lower magnitudes. The crush signals were normalized such that the signal data length was 512 points to remove variance induced by larger kernels producing crush signals of longer duration. Linear interpolation was used to normalize the length of the crush signal, reducing the effect of kernel size. Each value within the 512 points was divided by the dry weight of the kernel to compensate for the effect of kernel moisture.

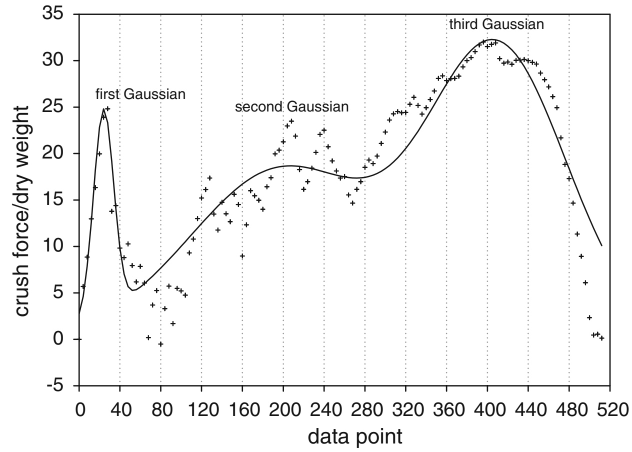

Fig. 3 Three-part Gaussian model of normalized SKCS crush signal. The solid line is the model and the +'s are the crush profile points

The normalized crush profiles were non-linearly fit with the summation of three Gaussian curves (Eq. 1). Each Gaussian was restricted to a certain segment of the normalized crush data.

$$y = a_1 e^{-((x-x_1)/2b_1)^2} + a_2 e^{-((x-x_2)/2b_2)^2} + a_3 e^{-((x-x_3)/2b_3)^2} \quad (1)$$

where a_1, a_2, and a_3 are the heights of the three Gaussians, b_1, b_2, and b_3 are the widths, x_1, x_2, and x_3 are the centers of the three Gaussians, x is the data point, and y is the fitted crush profile. The center of the first Gaussian, x_1, was restricted to points 1–80. This segment physically represented the forces required to initially fracture the whole kernel. The center of the second Gaussian, x_2, was restricted to points between 100 and 200, where large fragments are broken down further. The center of the third Gaussian, x_3, was constrained to points between 350 and 450. This third segment physically represented the forces on small kernel particles exiting the narrowest gap of the SKCS. A typical plot of the three-part Gaussian model is shown in Fig. 3.

Short time standard deviations of crush signals

While the SKCS parameters and Gaussian modeling characterize the crush profile in a global sense, small differences in the abruptness of kernel fracturing crush profiles can be extracted by computing standard deviations in short time windows across the entire length of the crush profile. Hard kernels tend to fracture more abruptly, while soft kernels are crushed in a slightly smoother manner. Standard deviations of the 512-point normalized crush profile were computed in short time windows (seven data points), each window overlapping the previous one by three points. This was done to highlight regions in which the crush profile was undergoing abrupt changes This technique has been successfully used in acoustical processing for recognition and classification of sounds [13, 14]. Following the short time window computation, a cumulative sum of these data was computed and normalized to produce a signal starting at 0 and ending at 1. This signal simply shows the contributions of different time intervals to the total signal variance, without regard to the crush profile magnitude. Figure 4 displays the normalized crush profiles, deviation profiles, and cumulative deviations for both soft and hard kernels. Hard kernels tend to have high variances at the end of the crush profile, where particles start to exit the SKCS. In contrast, soft kernels tend to have higher variances at the beginning of the crush profile. The variance signal was reduced to 128 points by interpolation, and all points were saved as potential discriminating features.

Crush signal frequency spectra processing

A discrete Fourier transform (DFT) was computed on the 256 points between points 45 and 300 of the 512-point normalized crush profile. This portion of the crush profile physically represents the larger kernel particles being crushed into smaller-sized flour. Thus, this is a time period in which many small-scale fractures occur, causing cyclic changes in the crushing forces.

The magnitude of the 256-point DFT comprised 128 points. The resulting DFT magnitude was very noisy; therefore a cumulative DFT was computed by integrating from the lower frequencies up to each point. The cumulative DFT was then normalized by dividing by the magnitude of the zero frequency, or the mean magnitude. The 128-point cumulative DFT was then smoothed by averaging adjacent points and every other point was saved,

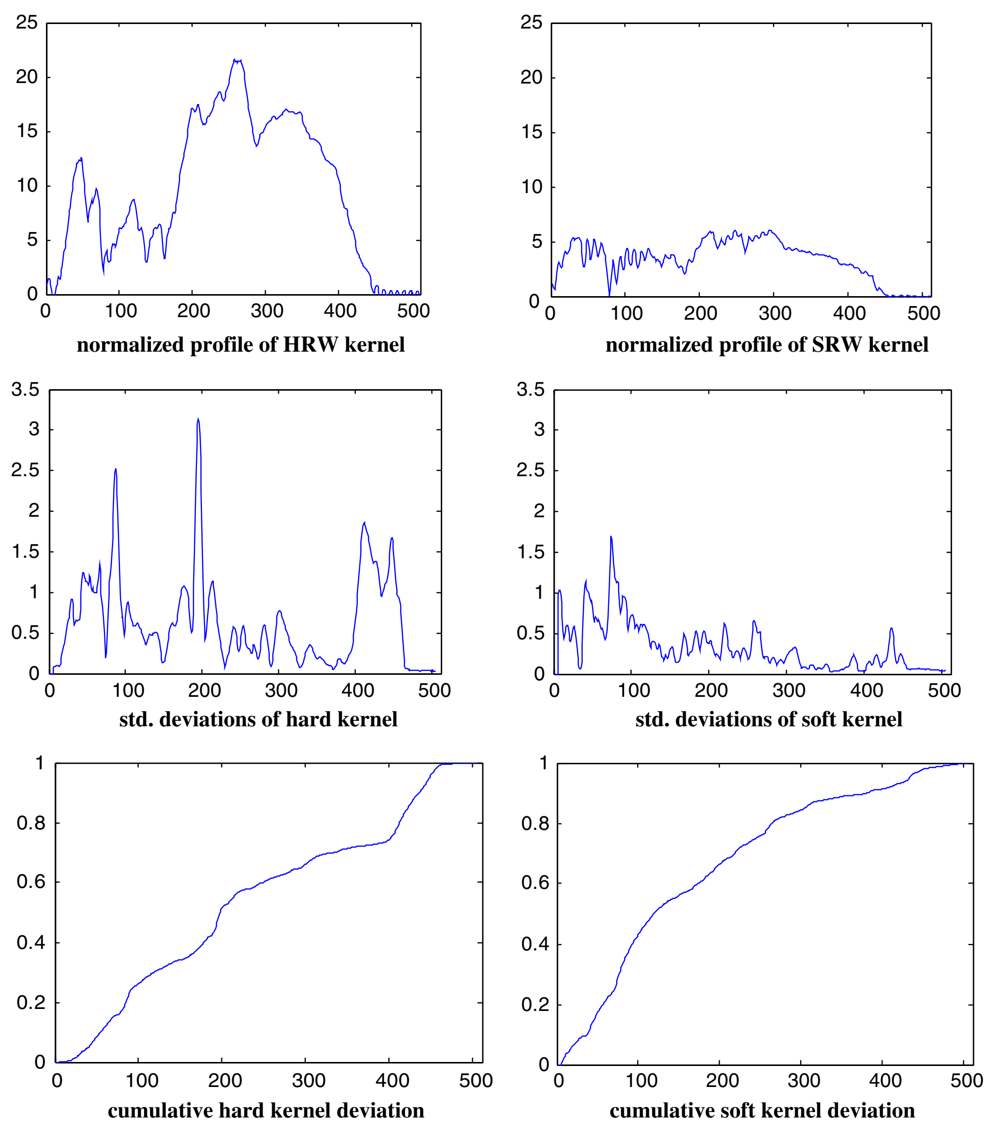

Fig. 4 Example crush profiles, deviation profiles, and cumulative deviations. Note that most of the variance in the HRW kernel is at the end of the profile while it is at the beginning of the SRW kernel. While the maximum short time standard deviations were considerably higher for the HRW than the SRW, this was not always a distinguishing feature for discriminating hardness classes, particularly for HW and SW classes

producing 64 features representing the frequencies in this region of the crush profile.

Image features

Images of each kernel were segmented from the background as described in [12]. The image was converted from RGB representation to hue, saturation, and lightness (HSL) representation using the Matlab software package (v7.04.365, The Mathworks Inc., Natick, MA). Hue and saturation represent the color of the kernel without regard to lightness and are somewhat immune to fluctuations in intensity of the lighting. After segmenting the kernel from the background, the maximum, average, standard deviation, median, and cumulative histograms of the hue and saturation components of the kernel pixels were computed. The hue histogram comprised 40 bins of intensities 0 through 40, as all hue values fell within this range. Both hue and saturation were represented with 8-bit resolution, so that the maximum intensity was 255. Hue values from 0 to 40 represent colors between pure red (hue = 0) and orange (hue = 40). The saturation histogram comprised 90 bins, as all saturation values fell below intensities of 90. A saturation value of 0 represents a mix of all colors, while a saturation value of 90 represents the mixture with greatest color purity. Morphological features of the kernels were also extracted using the Matlab library; these included kernel cross-sectional area, length, width, length/width ratio, and perimeter.

Parameter selection, discriminant analysis and single kernel classification

The HW and SW data were randomly divided into training and validation sets of equal size of 2,600 kernels for each set. The HR and SR data was used in it is entirety in the validation set in order to gain the best classification improvements between HW and SW kernels. Stepwise discriminant analysis ($P_{\mathrm{entry}} = 0.05$, $P_{\mathrm{exit}} = 0.05$) was used as the selection scheme. After parameters were selected, a discriminant function was developed using the training set and tested on the validation set. The discriminant analysis was performed using commercial statistical software [15]. Features and discriminant functions were developed using only the SW and HW data from the training set. Reported results are those from the validation set predictions applied to all data not in the training set. Four separate parameter sets and functions were developed:

1. only the main SKCS crush parameters: HI, weight, moisture, and kernel thickness;
2. the main SKCS crush parameters combined with low-level parameters such as crush area, Gompertz a, Gompertz b, and peak force;
3. all main and low-level SKCS crush parameters combined with the Gaussian model, short time standard deviations, and DFT features;
4. all SKCS features combined with the image data.

300 kernel sample classification

The SKCS classifies a 300-kernel sample as ‘soft’, ‘mixed’ (hard and soft), or ‘hard’ based on the average hardness index of the 300-kernel sample and a four-bin histogram of the hardness values [16]. Samples having an average hardness value less than or equal to 46 are classified as ‘soft’ or ‘mixed’, depending on the histogram of hardness values from the sample. Conversely, samples having average hardness values greater than 46 are classed as ‘hard’ or ‘mixed’. The histogram contains four bins comprising counts of hardness values in the ranges of: less than or equal to 33 (bin I), 33 to 46 (bin II), 46 to 59 (bin III), and greater than 59 (bin IV). These four bins are used to classify a sample as mixed or pure. While the exact procedure to classify a sample as mixed or pure is fairly elaborate, the most common route to classifying a sample as mixed is if bin I is greater than 10% for samples having an average hardness greater than 46; and bin IV greater than 10% for those samples having an average hardness less than or equal to 46.

A Monte Carlo simulation was used to randomly select 300-kernel samples from the HW and SW populations to create HW/SW mixtures of pre-determined proportions. Each simulation generated 1,000 simulated 300-kernel samples. Blends were simulated in 1% increments from 0% to 20% for both SW and HW wheat. The classification based on the hardness index and the histogram of hardness indices was performed and used as a benchmark. The single-kernel classifications were used to enhance the sample classification. If the predicted class of the wheat matched the hardness index class (e.g. if the hardness index was greater than 46 and the single kernel class was ‘hard’), then nothing was done. However, if the class did not match, then the hardness histogram bin containing the hardness value in question was decremented by one, and the next adjacent bin in the direction of the kernel’s predicted class was increased by one. For example, if a kernel had a hardness value of 20 but a predicted class of ‘hard’, then the histogram bin containing counts of hardness values less than 33 (bin I) was decremented by one, and the bin containing counts between 33 and 46 (bin II) was increased by one. This was done in order to utilize the single-kernel classifications in the existing hardness classification system. Data from the single-kernel classifications utilizing all main and low-level SKCS crush parameters combined with the Gaussian model, short time standard deviations, and

DFT features, as well as all SKCS parameters combined with the image data, were studied.

Results and discussion

Table 2 displays results for classifying kernels into their own classes using hardness index only. Kernels were classified as 'hard' if their HI was above 46, and as 'soft' if it was below or equal to 46. While errors for HR were below 2%, and errors for SR were below 6%, the errors for HW and SW were much higher, averaging over 17%. These results illustrate the difficulty in discriminating HW from SW using the current configuration of the SKCS system. When discriminant analysis was performed on the feature set restricted to the main SKCS parameters of HI, moisture, weight, and thickness, the validation set error rates for HW and SW were both 17%. This result is, on average, equivalent to that obtained when using HI alone. Thus, adding moisture, weight and kernel thickness to HI does not aid in classifying HW and SW kernels. This is not surprising, as the computation of HI uses these other parameters.

When the stepwise discriminant analysis procedure was applied to the data set that included SKCS low-level features (i.e. crush area, Gompertz a, Gompertz b, and peak force) along with the four SKCS main parameters, the validation set error rates for HW and SW white were reduced slightly to 13.2% and 14.7%, respectively. These error rates, approximately 3% lower (overall) than those obtained with HI alone, represent a 17% reduction in error rate over what is accomplished with hardness index alone. This classification scheme would be very simple to implement, since the SKCS already computes these features. The stepwise selection procedure chose the following four features in order: HI, $(\text{Gompertz b})^9$, crush area, and kernel weight. Again, since these additional features are used for the HI computation, it is somewhat expected that they would not greatly improve discrimination power over HI alone. This result demonstrates the need for other features in order to improve classification accuracy.

Table 2 Classification results using HI alone. Note that the largest errors occur for SW and HW classes

Class	Class predicted by HI		n
	Hard	Soft	
HR	98.6%	1.4%	2568
HW	83.6%	16.4%	2646
SR	5.4%	94.6%	2638
SW	18.1%	81.9%	2552
Total	61.6%	38.4%	10404

Table 3 displays validation set results for discriminating SW from HW using all of the SKCS features computed. Error rates for discrimination of SW from HW averaged 7.7%, which is 55% lower than that from HI alone. These results indicate that classification rates might be substantially improved using current SKCS hardware, as only software changes are needed to compute these features and implement this classification scheme. The error rate of 7.7% for distinguishing SW from HW is nearly the same as the previously developed method using predicted PSD along with other SKCS low level data [8]. However, the study by Pearson et al. [8] did not include different moisture levels and had a fairly small sample size. The two studies confirm that errors in distinguishing HW from SW wheat can be reduced by over 50% with additional signal processing in the SKCS software.

Average error rates for kernels at the low, medium, and high moisture levels averaged 6.5%, 9.0%, and 7.6%, respectively. While there does not appear to be a trend between classification accuracy and moisture content, it appears that more errors are made at moderate moisture levels, which represent normal storage conditions.

The stepwise selection procedure selected eight features, listed in Table 4, for discrimination between SW and HW classes. The features are listed in the order of Wilks' λ statistic for the discriminant model, which includes the feature and all those above it. A lower Wilks' λ statistic indicates a better group of features for discrimination [15]. Not shown in Table 4 is HI, which was selected as the single best feature for discrimination but was later removed from the model, as its significance was reduced in the presence of the other features. The two most significant features in the final discriminant model were the height of the Gaussian curves on the third and first peaks. After dividing all the crush forces by the dry weight, the HW kernels tend to have higher normalized forces for the first and third peaks. However, as can be seen from the means and standard deviations, these features for HW and SW

Table 3 Average classification results using low level SKCS features alone to discriminate SW from HW, then applied to the other classes

Class	Classification results using all SKCS features		n
	Hard	Soft	
HR	97.2%	2.8%	2546
HW	93.3%	6.7%	1242
SR	4.5%	95.5%	2633
SW	8.7%	91.3%	1295
Total	61.1%	38.9%	7716

Results are from the validation set and averaged across the low, medium, and high moisture levels

Table 4 Selected features for discriminating HW from SW

Feature	Wilks' λ	Group means		Group std dev	
		HW	SW	HW	SW
Third Gaussian height	0.64	19.21	13.82	3.88	3.27
First Gaussian height	0.54	9.13	8.17	3.37	3.45
Cumulative std dev #4	0.52	0.03	0.04	0.011	0.013
Cumulative DFT 250 Hz	0.49	6.27	5.34	4.25	3.1
Cumulative DFT 1625 Hz	0.47	14.06	8.99	4.14	2.7
$(\text{Gompertz } b)^9$	0.46	0.5	0.43	0.09	0.1
Weight	0.42	353.1	404.38	85.23	94.17

Note that all types of SKCS features were chosen (SKCS, Gaussian model parameters, short time cumulative standard deviations, and cumulative DFT features). Features are arranged by their contribution to the Wilks' λ statistic for the group of selected features

groups are not significantly different at the 95% confidence level.

Note that the cumulative crush standard deviation at window #4 was chosen as the third most important feature. This represents cumulative standard deviations from short time windows early in the crush signal. As shown in Fig. 5, the crush profiles from soft kernels have a greater percentage of variance in short time windows at the beginning and middle of the crush signal. In contrast, crush forces due to breaking down smaller particles from hard kernels contribute more variance to the end of the crush signal (Fig. 4). While more force is usually required to initially fracture hard kernels than soft kernels, the short-time variance from these fractures does not contribute as much to the overall variance, since variances due to the breaking down of smaller hard kernel particles as the material exits the SKCS are higher for hard kernels. Thus, the crush force characteristics from fracturing smaller particles as the kernel exits the SKCS are important phenomena for distinguishing between SW and HW kernels. This is not a feature that is distinctly measured during the computation of HI.

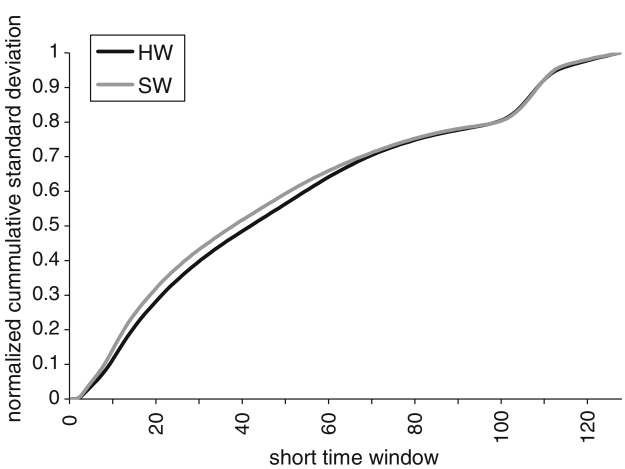

Fig. 5 Average normalized short time standard deviations of crush profiles. Note that the HW kernels tend to have a slightly less proportion of variance in their crush profiles at the beginning and middle of the crush. Even though the magnitude of the first breakage of hard kernels is usually higher than that of soft kernels, fracturing of kernel particles cause greater force variance than they do for soft kernels. For clarity, curves for SR and HR are not shown. The SR curve is similar to that of SW but is slightly higher than SW below window #80. The HR curve closely resembles HW but is slightly lower than HW below window #80

From the DFT features, the cumulative values at 250 and 1625 Hz were selected. As can be seen in Fig. 6, the normalized cumulative frequency spectra of HW and SW wheat are most different at 1625 Hz. Note that both HR and SR followed the same trends as HW and SW in the normalized frequency spectra, but the differences between HR and SR are more pronounced. Since the harder kernels fracture in a more abrupt manner, these kernels create more high-frequency energy than soft kernels. While the crush force slopes used to compute HI do capture some of these characteristics, they are likely convoluted with crush force slopes from other portions of the signal. By restricting the DFT window to the middle portion of the signal, the crush forces are better characterized. Additionally, the forces in this portion of the signal are likely less prone to variation due to the orientation of the kernel as it enters the crushing rotor/crescent of the SKCS, since the kernel has already been initially fractured.

The middle portion of the crush profile was also previously identified as important for predicting single kernel particle size distributions (PSD) and distinguishing between hard and soft wheat kernels [8] However, in that study, DFT features and Gaussian curve fitting parameters were not found useful for estimating particle size distributions of single kernels but this study found these features

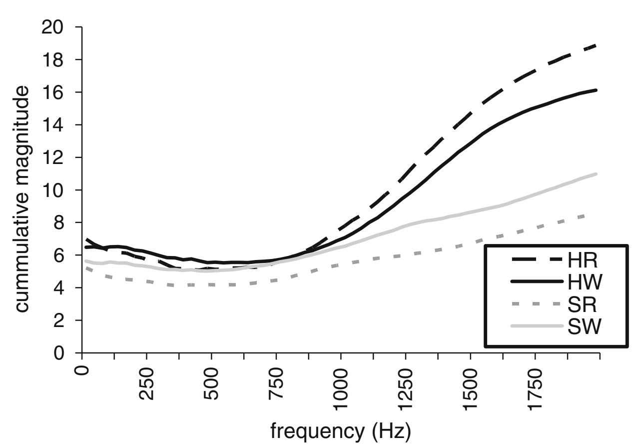

Fig. 6 Normalized cumulative frequency spectra of the mid crush profile. Note that soft kernels have a much lower energy at higher frequencies than hard kernels do

very useful in discriminating between HW and SW wheat. This discrepancy suggests that particle size distribution of resulting flour is not entirely dependent on the forces required to crush a kernel. This phenomenon was also observed in [8] as the highest classification accuracies between hard and soft classes required PSD to be combined with HI and other crush profile features.

Table 5 displays validation set classification results for selecting features to discriminate between SW and HW using combined SKCS and image features. Note that the discrimination of SW from HW is greatly improved over HI alone, but at the cost of reduced accuracy for SR. The reduced accuracy of SR is likely due to the training set only including HW and SW classes. Nevertheless, in areas where SR is not grown or handled, the use of the SKCS with a camera attachment may markedly improve an inspector's ability to determine the purity of HW and SW wheat. Average error rates when discriminating between SW and HW improved from a level over 17% for HI alone to below 5% for low-level SKCS features combined with image features. The error rates for HW and SW using SKCS features combined with image features are approximately 50% of those obtained when the feature set is restricted to SKCS-generated features. However, the image hardware adds complexity and cost to the system. Furthermore, further study is required to test the robustness of calibrations on a long-term basis and across different instruments.

Average error rates for SW and HW varied slightly for the three moisture-dependent calibrations. The average error rates for kernels at the low, medium, and high moisture levels were 2.5%, 4.6%, and 6.1%, respectively. This indicates a slightly decreased ability to distinguish HW from SW as moisture level increases when using combined image and SKCS parameters.

A total of 15 features were selected from the pool of combined SKCS and image features to discriminate HW from SW wheat. The features and averages for the SW and HW classes are listed in Table 6. Note that all types of features were selected: image, Gaussian parameters, short time standard deviations, and cumulative DFT. The most significant feature was the median hue of kernel pixels, followed by the height of the third Gaussian. During the stepwise selection process, HI was never chosen; median hue was a more significant discriminating feature than HI, and median hue combined with HI was not as strong a discriminating pair as median hue combined with the height of the third Gaussian. The median hue from HW kernels was slightly less than that for SW kernels, as HW kernels tend to appear redder than SW kernels. Hue histogram values from bins 22 and 31 were chosen, as the hard kernels tend to have slightly more red than the soft kernels. The SW kernels tend to have less variance in color, or more color purity, than HW kernels; this is indicated by the average saturation and saturation histogram values selected. The average saturation for SW kernels was slightly higher than for HW kernels. Among the saturation histogram bins selected, bins 54 and 67 had higher counts for SW kernels and lower counts for saturation histogram bin 22 than HW. Finally, kernel length was selected, as soft kernels tend to be slightly larger than hard kernels.

Figures 7 and 8 display the percentages of samples that were classified as 'mixed' from the Monte Carlo simulations using various mixtures of HW and SW. For mixtures comprising mostly soft or HW wheat, the single kernel classifications have steeper curves, indicating that the

Table 5 Average classification results using low level SKCS features combined with image features to discriminate SW from HW, then applied to the other classes

Class	Classification results using combined SKCS and image features		n
	Hard	Soft	
HR	100.0%	0.0%	2546
HW	96.2%	3.8%	1242
SR	54.6%	45.4%	2633
SW	5.0%	95.0%	1295
Total	71.0%	29.0%	7716

Results are from the validation set and averaged across the low, medium, and high moisture levels. Note that the improved accuracy for SW and HW classes but lowered accuracy for SR

Table 6 Selected features and averages of the features for the two groups used to discriminate HW from SW. Features are arranged by their contribution to the Wilks' λ statistic for the group of selected features

Feature	Wilks' λ	Group means		Standard deviations	
		HW	SW	HW	SW
Median hue	0.46	24.59	26.34	0.51	1.04
Third Gaussian height	0.37	19.21	13.82	3.88	3.27
Sat hist bin #54	0.34	400.63	857.39	307.68	411.29
Sat hist bin #22	0.32	175.61	151.03	112.72	105.46
Hue hist bin #31	0.31	389.33	1929.65	271.64	1613.28
Sat hist bin #67	0.30	20.49	125.25	46.81	189.32
Cum DFT 1625 Hz	0.29	14.06	8.99	4.14	2.70
Cum std dev #9	0.28	0.10	0.13	0.03	0.04
First Gaussian height	0.28	9.13	8.17	3.37	3.45
Cum DFT 343 Hz	0.27	5.85	5.12	4.28	2.99
Hue hist bin #22	0.27	950.53	539.88	691.74	513.96
Kernel length	0.26	248.88	249.11	20.89	19.83
Avg saturation	0.26	44.31	46.63	2.14	3.15
Sat hist bin #20	0.25	191.60	159.04	83.71	87.67

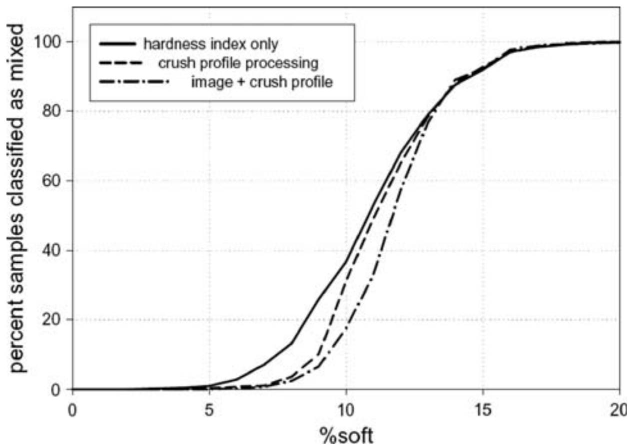

Fig. 7 Classification of mixtures containing hard white wheat blended with low levels of soft white wheat

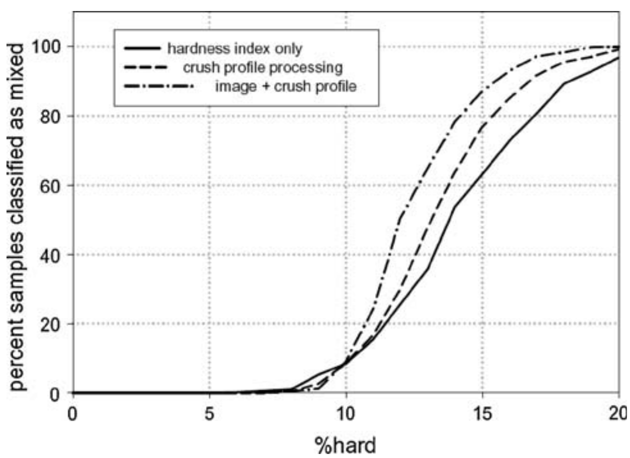

Fig. 8 Classification of mixtures containing soft white wheat blended with low levels of hard white wheat

additional processing can allow for a sharper distinction between a pure sample and mixed samples. When the blending is less than 10% of a contrasting class, the additional crush profile processing and image data reduces the number of samples classified as 'mixed', compared with hardness index alone. For example, as shown in Table 7, when the blending of HW is blended with six to 10% HW, the number of samples classified as mixed is 9% compared with 17% when only HI and histograms of HI are used. However, in the case of hard wheat blended with small amounts of SW wheat, the additional processing does not improve the number of samples that should be classified as mixed when the blending is over 10% SW. As shown in Table 7, samples of HW wheat blended with 11–15% SW have approximately the same percentages of samples correctly classified as mixed, about 75%. Samples that are predominantly SW wheat blended with greater than 10% HW wheat, the reported "mixed" classifications were improved somewhat. For samples of SW wheat having 11–15% HW blended into them, 39% are classified as mixed when using HI and HI histograms alone while 47% are classified as mixed when using additional crush force processing and 60% are classified as mixed when using additional crush force processing combined with image features.

Table 7 Average number of mixed samples reported for various HW–SW blends

Hard white (%)	Soft white (%)	Average percentage of mixed samples reported		
		HI only	Crush profile features	Image + crush profile features
0–5	100–95	0	0	0
6–10	90–94	3	2	2
11–15	85–89	39	47	60
100–95	0–5	0	0	0
90–94	6–10	17	9	5
85–89	11–15	76	75	70

Conclusion

New data processing and imaging hardware additions to the SKCS 4100 system were proposed and tested. While the imaging hardware appears to aid in distinguishing HW from SW kernels, it does add some complexity to the system, and the robustness of calibrations over time and across different instruments needs to be further investigated. It does appear that additional data processing of the raw crush profile acquired while kernels are crushed can improve discrimination between SW and HW as well. This method has the advantage that only software changes to the SKCS are required for its implementation.

When SKCS data was combined with image features, errors in distinguishing SW from HW were reduced to less than 5%, from 17.1% using HI alone. However, adding image data decreased hardness classification accuracy for SR and added complexity to the system. When additional data processing is applied to the low-level data currently produced by the SKCS 4100, classification errors between SW and HR are reduced by over 50% with respect to HI alone (from 17.1% for HI to 7.7% with additional data processing). Similar improvements to classification accuracies were achieved with 300-kernel samples when the individual kernel classification was used to modify the histogram of HI values from the sample. These results indicate that improved classification for HW and SW can be achieved through additional data processing software.

References

1. USDA-GIPSA, U.S. Standards for Grain>Wheat (5/1/06). USDA, Grain Inspection, Packers and Stockyard Administration (2006), Available at: www.gipsa.udsa.gov
2. D. Goter, Kansas wheat farmers may be raising near-useless wheat. Salina Journal. **332**, 1–11, November 28 (1985)
3. D. Steimel, Plan could cost wheat growers millions. Wichita Eagle Beacon. August 3, 7B (1986)
4. C.R. Martin, R. Rousser, D.L. Brabec, Development of a single-kernel wheat characterization system. Trans. ASAE **36**(5), 1399–1404 (1993)
5. NIST, *Reference Material 8441—wheat hardness* (National Institute of Standards and Technology, Gaithersburg, MD, 1997)
6. K. Blocker (2005) Hard white wheat the wave of the future. Wheat Life. 2005(October), 46–47
7. D.A. Engle, C.F. Morris, *Genotype and Environment Study, 8-Year Summary, 1997–2004 Crop Years* (USDA-ARS, Western Wheat Quality Laboratory, Pullman, Washington, 2005)
8. T.C. Pearson, J. Wilson, J. Gwirtz, E. Maghirang, F. Dowell, P. McCluskey, S. Bean, The relationship between single wheat kernel particle size distribution and the Perten SKCS 4100 hardness index. Cereal Chem. **84**(6), 567–575 (2007)
9. I. Zayas, F.S. Lai, Y. Pomeranz, Discrimination between wheat classes and varieties by image analysis. Cereal Chem. **63**(1), 52–56 (1986)
10. M. Neuman, H.D. Sapirstein, E. Shwedyk, W. Bushuk, Wheat grain color analysis by digital image processing: II. Wheat class discrimination. J. Cereal Sci. **10**(30), 183–188 (1989)
11. S. Majumdar, D.S. Jayas, Classification of cereal grains using machine vision: II. Color models. Trans. ASAE **43**(6), 1677–1680 (2000)
12. T.C. Pearson, D.L. Brabec, Camera attachment for automatic measurement of single-wheat kernel size on a Perten SKCS 4100. Appl. Eng. Agric. **22**(6), 927–933 (2006)
13. T.C. Pearson, A.E. Cetin, A.H. Tewfik, R.P. Haff, Detection of damaged wheat kernels by impact-acoustic emissions. Digital Signal Process. J. **17**(2007), 617–633 (2007)
14. B. Gold, N. Nelson, *Speech and Audio Signal Processing: Processing and Perception of Speech andMusic* (John Wiley and Sons, New York, NY, 2000)
15. J. Hintze, *NCSS and PASS Number Cruncher Statistical Systems* (Kaysville, Utah, 2004)
16. Perten Instruments, *Single Kernel Classification System 4100 Operation Manual* (Springfield, IL, 1999)