DISSERTATION

THE UTILITY OF NEAR-INFRARED REFLECTANCE SPECTROSCOPY FOR

WHEAT QUALITY ASSESSMENT

Submitted by

Joshua Donald Butler

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2010

COLORADO STATE UNVERSITY

May 31, 2010

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY JOSHUA DONALD BUTLER ENTITLED: THE UTILITY OF NEAR-INFRARED REFLECTANCE SPECTROSCOPY FOR WHEAT QUALITY ASSESSMENT BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

_____

Mark Brick

_____

Phillip Chapman

_____

Bradford Seabourn

_____

Advisor: Scott Haley

_____

Department Head: Gary Peterson

ABSTRACT OF DISSERTATION

THE UTILITY OF NEAR-INFRARED REFLECTANCE SPECTROSCOPY FOR
WHEAT QUALITY ASSESSMENT

End-use quality improvement is an important objective in most wheat (*Triticum aestivum* L.) breeding programs. Limited sample size, destructive parameter testing, and the short duration between harvest and planting of winter wheat are challenges for testing early-generation breeding material for end-use quality parameters. Near-infrared reflectance (NIR) spectroscopy is a rapid and non-destructive technique that could facilitate early-generation selection for end-use quality. The precision and accuracy of an NIR equation for prediction purposes is dependent on the construction of a reliable calibration. The objectives of this study were to: 1) develop and validate NIR calibration models for grain volume weight, kernel characteristics, and Farinograph parameters, and 2) evaluate the performance of NIR calibration models in a breeding context for grain volume weight and single kernel characteristics. Calibration models for prediction of grain volume weight and single kernel characteristics were developed using NIR spectra and laboratory reference values from up to 10,000 samples collected from breeding nurseries under multiple environments over four crop years. Models encompassing all years of data revealed R^2 (validation) of 0.73 for kernel diameter, 0.74 for kernel weight, 0.70 for kernel hardness, and

0.81 for grain volume weight. Of the Farinograph parameters, only absorption was effectively predicted using NIR calibration models for whole grain and flour with $R^2 \geq 0.70$. Realized heritability was estimated as a response to selection using NIR predicted values and laboratory reference values and was generally larger when using the reference values when compared to predicted values (0.17-0.77 vs. 0.05-0.77), but suggested that genetic gain was possible when using NIR models for selection. Classification errors when using the NIR models were highest in the mid-range reference values (56-66%), but could allow for divergent selection of high and low reference values. The results suggest that NIR models suitable for screening grain volume weight, SKCS kernel characteristics, and Farinograph absorption could be utilized in a breeding program and could aid in the elimination of early-generation samples with unacceptable values.

Joshua Donald Butler
Department of Soil and Crop Sciences
Colorado State University
Fort Collins, Colorado 80523
Summer 2010

ACKNOWLEDGMENTS

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LITERATURE REVIEW

Introduction

Near-infrared (NIR) light refers to a region in the electromagnetic spectrum adjacent to the visible range (between 750 and 3000 nm in wavelength). While light in the infrared region is not visible to the human eye, most organic materials have excellent transmittance and reflectance properties at those wavelengths (Williams, 2003). Diffuse reflectance spectroscopy is an extension of human vision through instrumentation (Williams and Norris, 2001). Output from NIR spectroscopy is in numeric form so mathematical treatments can be used to further analyze results.

Spectroscopists have recognized the existence of the fundamental absorption of energy in the infrared region since the early 1800's (Williams, 2003). Since then, technology has advanced, and with it, the usefulness of NIR spectroscopy. The technology was put to use in agricultural products by Karl Norris in the late 1950's, and the first commercial application of NIR spectroscopy is credited to the Canadian Grain Commission (CGC) when it applied it to testing of grain samples for protein concentration, previously tested by chemical methods (Williams and Norris, 2001). Later calibrations were developed and implemented for moisture, oil, and starch concentration of grains. The mid-1970's brought improvements in instrument design and software (Williams, 2003). In the early 1980's, development of instrumentation for testing

whole grain wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) for protein and moisture occurred. This was significant since it was no longer necessary to mill or grind the samples, thereby providing information via rapid and nondestructive means. With the advancement of computer technology and software, the state of the art is now at a stage where computer memory capacity and computing time are no longer obstacles to development of useful calibrations.

Technology and Statistics

The infrared (IR) portion of the electromagnetic spectrum is divided into three regions; far-infrared (25000-1000000 nm), mid-infrared (2500-25000 nm), and near-infrared (750-2500 nm) (Stuart, 2004). Infrared spectroscopy is commonly referred to as vibrational spectroscopy because chemical bonds have specific frequencies at which they vibrate corresponding to energy levels. Atoms commonly found in organic compounds can vibrate in six different ways: symmetrical and asymmetrical stretching, scissoring, rocking, wagging and twisting (Burns and Ciurczak, 2001). Molecules are in a continuous state of vibration at various frequencies and irradiation by an energy source, such as light, can cause the molecules to change their vibrations from one energy level to another. When this takes place, energy is absorbed, and this absorption can be measured and quantified by spectrophotometers (Williams, and Norris 2001). Different groups of atoms absorb at different wavelengths, and by taking measurements at wavelengths of a known group, chemical composition can be predicted (Williams, 2003). One complicating factor is that there is considerable

overlap due to the fact that more than one molecular group is absorbing at any one wavelength. Statistics are employed to further resolve vibrational and molecular differences.

The two most commonly used regions in infrared spectroscopy are near-infrared and mid-infrared (mid-IR). The advantage of mid-IR is that it directly monitors the fundamental vibrations corresponding to functional groups, whereas the NIR is dominated by combinations and overtones, making interpretation of results difficult (Brad Seabourn, personal communication). Therefore, mid-IR reveals IR absorptions that correspond to unique, unambiguous molecular vibrations making it more straightforward to relate specific absorbances to the chemical changes that are of interest to a given application and, as a result, mid-IR calibrations are often quite easy and fast to develop (Burns and Ciurczak, 2001). In contrast, NIR spectra are dominated by broad, overlapping overtones, and calibration models are heavily reliant on statistics (Williams, 2003).

The disadvantages of mid-IR are the expensive optical materials required for mid-IR (resulting in more expensive instruments) and the fact that the only suitable fiber optics for communication between a measurement point and a remote instrument are of limited length. While this has little bearing for research purposes, it does limit its use in process control applications. In contrast, NIR instruments have relatively inexpensive optical materials and fiber optics are routinely used in industry (Stuart, 2004).

NIR spectroscopy generally relies on diffuse reflectance, which is the reflection of light from an uneven or granular surface such that an incident ray is seemingly reflected at a number of angles. It is the complement to spectral reflection. Diffuse reflectance is also affected by particle size and shape; NIR spectroscopy can therefore be used to develop calibration models for textural properties in addition to chemical properties.

The reference method of calibration development is the primary method used in NIR spectroscopy for agricultural applications. The first step in development of a calibration using this method is to derive a sample set that encompasses as much variation as possible for the trait of interest. It is also important to include material that represents a variety of genetic and environmental variation characteristic of the samples for which the calibration will later be used. Using an NIR spectrometer, spectral data are then collected on the sample set. Mathematical pretreatments, such as derivativization and smoothing, are then applied to the data. Smoothing theoretically reduces spectral noise and derivativization compensates for baseline drift and assists in resolution of overlapping absorption bands (Williams, 2003). Multiplicative scatter correction is also routinely used and involves correction for differences between the individual spectral data for all samples at a wavelength point and the average spectral data for all samples at that wavelength (Williams, 2003). Principal component analysis is then used to evaluate the spectral data, which allows for the removal of spectral outliers and spectrally redundant samples from the data set. The other part of the process for this method involves collection of laboratory reference

data on samples that are present in the spectral dataset. When the reference data have been collected for all of the samples that remain in the spectral dataset, calibration development can begin. This can be carried out using several methods, but the two most frequently used are multiple linear regression and partial least squares regression. A prediction model is developed to predict the parameter of interest and its effectiveness is tested on an independent validation set of samples.

NIR Models for Wheat Quality

NIR spectroscopy, as a tool for prediction of wheat quality parameters, has been well documented in the literature. It not only allows for the possibility of non-destructive prediction of quality parameters, but also has the potential to quickly predict several parameters simultaneously with a very small sample that requires little or no preparation prior to analysis.

In an example of published research using the reference method, Delwiche et al. (1998a) used 30 genotypes of hard red winter wheat from eight locations throughout Nebraska and assessed the ability of NIR spectroscopy to predict concentrations of gliadin, glutenin, albumins, globulins, sodium dodecyl sulfate (SDS) sedimentation volume, and mixograph dough handling properties. Eight of the 30 genotypes contained wheat-rye chromosome translocations and the presence of secalins were determined for these genotypes. Measurement of protein fractions was done using size-exclusion high performance liquid chromatography (HPLC). Sedimentation volumes were determined using

approved method 56-61A (AACC, 2000). Mixograph analysis provided four parameters: time to peak dough development, the height of the peak at maximum resistance, the width of the curve at maximum resistance, and the width of the curve at two minutes past the point of maximum resistance. Using reference data along with spectral data collected with a Foss NIRSystems model 6500 (Foss North America, Eden Prairie, MN), prediction models were developed. Model performance was evaluated on the validation set and reported as: 1) correlation coefficient (r) determined from the linear association of the validation reference values onto model-predicted values; 2) standard deviation of validation errors (SDV), also commonly referred to as standard error of performance (SEP); 3) bias, defined as the mean difference between the modeled and reference values; 4) ratio of the standard deviation of the reference values (SD) to the SDV [residual prediction deviation (RPD), defined by Williams (1987) as RPD = SD/SDV]; and 5) ratio of the variance of the reference values with variance of the NIR model removed to the variance of the reference values with variance caused by the reference technique removed where RAP = $((SD^2 - SDV^2)/(SD^2 - laboratory\ error^2)$) (Martens and Naes, 1984). When calibrations were applied to an independent validation set, NIR models for glutenin content, gliadin content, SDS sedimentation volume, and mixograph peak resistance demonstrated relatively high values of R^2 ranging from 0.87 to 0.94.

In a variation on the traditional reference method of calibration development, Wesley et al. (1999) used a curve fitting approach to develop predictions based on theoretical spectra from nearly pure samples of gliadin and

glutenin. Internal curve fitting was based of the identification of regions in the NIR spectrum that are unique to the components of interest. In the case of flour, these components are gliadin, glutenin, starch, and water. Samples of gliadin and glutenin were isolated from wheat flour, freeze dried, and scanned with a Foss NIRSystems 6500. Spectra from the scans were used to determine regions of interest and a curve fitting approach along with a least squares unmixing methodology was used to develop a prediction model. One hundred and seventy doubled haploid lines consisting of 60 different glutenin subunit combinations were used to evaluate the prediction model. High R^2 values were seen when regressing predicted values on those obtained by size-exclusion HPLC (gliadin R^2=0.73 and glutenin R^2=0.76). While a significant bias was present in the prediction model, it was useful for ranking the samples correctly. This could prove useful in plant breeding programs where ranking is generally more important than accuracy.

Discriminant analysis is another method used in the development of NIR spectroscopy calibrations. It proves useful to identify samples by discrete classes and involves the correlation of defined classes with principal component clusters derived from spectral data. In an example of discriminant analysis, McCaig (1993) investigated the use of NIR and visible light spectroscopy for assessment of seed coat color in a wheat breeding program. Red seed coat color in wheat is controlled by up to three genes and various combinations of these genes impart various shades of red to the genotypes (Metzger and Silbaugh, 1970; Baker, 1981). Seed coat color can be determined by visual inspection in most samples,

but some genotypes can be difficult to distinguish. Environmental factors can also play a role in making red genotypes appear white, and white genotypes appear red. A total of 262 diverse wheat samples consisting of 115 red and 147 white, including common and durum wheat (*Triticum durum* L.), were used in the analysis. After a first-derivative data transformation to remove spectral variation due to differences in seed size and shape, results revealed two regions that were useful to discriminate among the three wheat groups. The visible region around 516 nm distinguished between red and white seed-coat wheat and durum could be distinguished from common white seed-coat wheat in the visible region of 450 nm. Although average spectra for each group were clearly distinguished with a bimodal distribution, the range of values for both 516 nm and 450 nm resulted in some overlap (approximately 1%).

Dowell (1998) went further to address the problem of overlap with bulk sampling methods and evaluated the accuracy of color classification of single wheat kernels using visible and near-infrared reflectance. Using a single kernel characterization system (SKCS) integrated with a DA-700 diode-array spectrometer (Perten Instruments, Reno, NV), Dowell (1998) developed calibrations using samples representing the hard red winter, soft red winter, hard red spring, soft white, and hard white wheat market classes. In addition to samples that were easily classified as red or white by grain inspectors, approximately 20% were samples that were difficult for inspectors to determine color class. A calibration based on the entire visible and near-infrared spectrum resulted in >99% correct classification for single kernels. When single kernel

classification results were averaged, 100% correct classification of bulk samples was realized. The most useful information for correct classification was found to lie within the 450-1688 nm region; however, results revealed the best classification resulted from using the entire 450-1688 nm region. The fact that the use of NIR spectra improved classification accuracy suggests that there may be some difference in the molecular structure of the red and white seed coats.

Various laboratory techniques are employed to determine bread baking quality of wheat. These include various measures of sedimentation, evaluation of dough mixing and handling properties, determination of chemical constituents of flour such as protein fractions and starch, and experimental baking. In addition to flour and bread baking quality, various other quality measures are of interest such as milling characteristics, including parameters indicative of milling performance (kernel weight, kernel diameter, grain volume weight, and hardness), and grain and flour color.

In a study to determine whether reflectance spectroscopy could be used under commercial bakery conditions to monitor functional performance in hard red spring and hard red winter flours, Delwiche and Weaver (1994) evaluated flour for baking quality. Six parameters were used in the evaluation: 1) water absorption 2) loaf height, 3) mixing time, 4) mixing tolerance or stability, 5) internal grain score, and 6) overall baking score. Spectra were collected using a Foss NIRSystems 6500. One quarter of the 317 samples were not included in the calibration set and were used for independent validation of the calibration model. Using principal component analysis (PCA) and partial least squares (PLS)

regression, calibration models were developed and evaluated with the validation set. Results revealed high R^2 for absorption (0.92), intermediate R^2 for loaf height (0.63), and low R^2 for mixing time, internal grain score, and overall bake score.

Blazek et al. (2005) explored the use of NIR spectroscopy of laboratory milled flour to predict milling characteristics in experimental and commercial wheat. The values of milling parameters are determined by the milling quality of wheat, which is affected by the physical and chemical properties of the grain endosperm. Ninety-four samples were scanned using a Foss NIRSystems 6500. Samples were milled in a Chopin CD1 Auto mill and semolina extraction rate, semolina extractability, flour extraction rate, and Mohse extraction rate were determined. Eight samples were set aside as an independent validation set. Calibrations were developed using PLS/modified PLS (mPLS) and artificial neural network (ANN) methods with both cross-validation and independent validation. Results revealed that when using cross-validation and PLS/mPLS, the highest statistical correlation value (r=0.63) in the set was achieved for semolina extractability. Similar results were observed when using the ANN method, but higher r and lower SEV values were realized. When using independent validation, the best calibration equation was developed for semolina extraction rate with a resultant r=0.92. Low correlation coefficients were found for the other two parameters studied. The potential for use of NIR spectroscopy to predict milling characteristics was examined and deemed not fully satisfactory when using independent validation. The authors suggested that better results could be realized if a larger sample set were to be used for the evaluation.

Starch level and composition in wheat have an influence on processing characteristics. Plants synthesize starch in organelles called amyloplasts, with the major starch molecules made up of amylose and amylopectin. The two starch molecules differ in their molecular structure by their degree of branching of D-glucosyl units, with amylose comprised of strait chains of α (1 → 4) linked D-glucosyl units. Amylose is believed to be synthesized by a pathway in which granule-bound starch synthase (GBSS), also known as waxy protein, is the primary enzyme. As a hexaploid, common wheat contains three unique loci that code for iso-forms of GBSS (*wx-A1, wx-B1, and wx-D1*). Generally, the greater number of genes coding for GBSS, the greater the amylose content in the grain. Lines that contain all three genes for iso-forms of GBSS are wild-type, those with one or two nulls are partial waxy, and those with three nulls are termed waxy. Delwiche and Graybosch (2002) investigated the feasibility of using NIR spectroscopy to identify and differentiate between wild-type, partial waxy, and waxy wheat samples. One-hundred and ninety-two samples (27 triple null, 126 single or double null, and 39 wild-type) were examined over two crop years. Samples were ground, and then spectra were collected with a Foss NIRSystems 6500. Reference data were generated in duplicate using iodine-binding blue complex colorimetry. SAS procedures PRINCOMP, STEPDISC, and DISCRIM were used to develop a discriminant analysis model for prediction of waxy classes. Results indicated only one of the 27 waxy samples from the first year was incorrectly classified, and in year two only two of the 27 were incorrectly classified. Only 59% of the partial waxy samples were correctly classified, but

11

only 20% were classified as wild-type and none as waxy. Approximately 54% of wild-types were correctly classified. The results suggested that while misclassifications were present, they were generally minor and classification would be effective for early generation screening for allele enrichment.

Pawlinsky and Williams (1998) evaluated the efficiency of prediction models for whole kernel wheat for quality parameters including protein concentration, wet gluten percent, Zeleny sedimentation volume, mixograph peak time, Farinograph absorption, Farinograph development time, Farinograph mixing tolerance index, extensigraph height, extensigraph area, and Canadian short process mixing time. Samples consisted of elite lines grown over two seasons (1995 and 1996) and 25% of the samples were not used in calibration development, but instead as an independent validation set. Significant results were obtained for calibration models for Farinograph absorption and mixing tolerance.

A relatively complete and comprehensive evaluation of the potential of NIR spectroscopy to measure milling, flour, dough, and bread-making quality characteristics from whole grain and flour of hard red spring (HRS) and hard red winter (HRW) wheat was conducted by Dowell et al. (2006). One-hundred and eighty-six quality parameters were evaluated using 100 HRW and 98 HRS wheat and flour samples. Using four different NIR instruments, spectral data were collected in addition to the data collected from the 186 reference methods. Mean-centered absorbance (log 1/R) and the Savitzky-Goley first-derivative of the absorbance spectrum were analyzed. Other mathematical treatments were not

tested under the assumption that improvements in prediction accuracies would be small. Ability of prediction by NIR was reported as the coefficient of determination (R^2) and the standard error of cross-validation (SECV). Results were interpreted as suggested by Williams and Norris (2001) with R^2 = 0.70-0.90 suitable for rough screening, R^2 = 0.90-0.97 suitable for screening and quality control, R^2 = 0.97-0.99 as suitable for process control, and larger values as suitable for most applications. There were no whole grain, flour, dough or baking characteristics that were consistently predicted more accurately by a specific spectrometer, with the exception of flour color, which was predicted by the Foss 6500 with R^2 values twice that of the other instruments.

When analyzing spectra from HRW whole kernels, Dowell et al. (2006) found that grain and flour protein concentration and grain moisture could be predicted with $R^2 \geq 0.97$, while average single kernel moisture, total gluten content, and mixograph absorption were predicted with $R^2 \geq 0.90$. Additionally, grain volume weight, average single kernel diameter, SDS sedimentation volume, soluble and insoluble glutenin content, soluble gliadin content, Farinograph absorption, alveograph length, alveograph swelling index, alveograph work, loaf volume and specific loaf volume were predicted with $R^2 \approx$ 0.70-0.90. When analyzing spectra from flour, NIR predictions with $R^2 \geq 0.97$ were observed for grain and flour protein concentration, and color b*. Color a*, total gluten content, and mixograph absorption were predicted with $R^2 \geq 0.90$ and SDS sedimentation volume, soluble and insoluble glutenin content, soluble

gliadin content, alveograph length, alveograph swelling index, alveograph work, loaf volume, and specific loaf volume with $R^2 \approx 0.70\text{-}0.90$.

In addition to the aforementioned papers, research on the use of NIR spectroscopy for prediction of wheat quality parameters is extensive. Delwiche and Hruschka (1998, 2000) investigated the utility of NIR spectroscopy for prediction of protein concentration of single kernels. Manley et al. (2002) and Morris et al. (2005) studied the effectiveness of predicting hardness and texture. Delwiche et al. (1991, 2002), Gergely and Salgo (2005), and Juhasz et al. (2005) examined the utility of NIR spectroscopy to predict starch and carbohydrate content including parameters measured by the rapid visco analyzer. Delwiche et al. (1994, 1998a), Wesley et al. (1999), and Nielsen et al. (2001) assessed the predictive power of NIR spectroscopy for flour characteristics, dough handling properties, and baking quality. Other studies have been conducted to determine NIR utility in determination of nitrogen uptake (Stenberg et al., 2005), *Fusarium* head blight damage (Delwiche and Hareland, 2004), insect damage (Perez-Mendoza et al., 2005), and heat damage (Wang et al., 2001).

Prediction of protein concentration by NIR spectroscopy has long been accepted as a useful method to assess the bread baking potential of a wheat sample. It is also important to recognize that in addition to protein quantity, protein quality is an important factor to consider. End-use quality of wheat is related to both the amount and quality of the gluten proteins and by the complex interactions of all the biochemical constituents present in flour.

14

An important issue in the development of NIR models for wheat end-use quality is the high level of correlation of many quality parameters to protein concentration. Osborne (1984) found that when evaluating a prediction model for SDS sedimentation volume, any predictive power of the model was lost when the contribution of protein concentration was removed from the model. Delwiche et al. (1998a) found that the absolute amount of gliadins and glutenins is predictable and is highly correlated with protein concentration. When the influence of protein concentration is removed from the model, R^2 values were poor. When evaluating prediction models for HRW, Dowell et al. (2006) discovered that when the contribution of protein concentration was removed, only moisture concentration predicted from whole grain spectra, and color a*and b* from the flour had $R^2 \geq 0.70$.

While several studies show that protein concentration is highly correlated with NIR-prediction of most quality parameters, there are studies to the contrary. Pawlinsky and Williams (1998) reported that in evaluation of samples from two separate growing seasons, protein concentration was largely influenced by environment. When correlating protein concentration to functionality parameters in one of the years, only weak correlations were observed while prediction models for some of the parameters were successful. From this, the authors concluded that the NIR method was apparently capable of predicting some functionality parameters, independent of the influence of protein. They went further to conclude that NIR reflectance scanning of whole grain appears to be capable of predicting wheat functionality parameters to sufficiently identify

suitable material for advancement in a breeding program. This was reiterated by Delwiche et al. (1998a) who concluded that protein quality measurements (quantity of gliadin and glutenin) by NIR may be applicable to breeding programs, but may not be attainable on commercial samples.

As stated by Pawlinsky and Williams (1998), one of the most important aspects of evaluation of the characteristics of new genetic material is the degree to which important characteristics are heritable. Protein concentration in bread wheat is largely influenced by environmental conditions and is not highly heritable. Characteristics such as gluten strength and kernel texture, on the other hand, are much more highly heritable (Pawlinsky and Williams, 1998). This suggests a genetic basis for these parameters that can be exploited in a breeding program.

In all of the papers to date, none have attempted to develop calibrations within a breeding program utilizing a broad spectrum of material in various filial generations. One benefit to this approach would be the vast genotypic, phenotypic, and environmental variability present in breeding material. In turn, this would greatly benefit the breeding program in allowing for quick, inexpensive, and non-destructive prediction of quality parameters in early-generation material and allow for improved genetic gain for heritable traits with each round of selection.

The objectives of this study were to: 1) develop and validate NIR calibration models for grain volume weight, single kernel characteristics (SKCS

kernel weight, diameter, and hardness) and Farinograph parameters (absorption, development time, stability, and mixing tolerance index), and 2) evaluate the performance of NIR calibrations in a breeding context for grain volume weight and single kernel characteristics.

CALIBRATION DEVELOPMENT AND VALIDATION

Near-Infrared Reflectance (NIR) Spectroscopy for Estimation of SKCS Parameters and Grain Volume Weight in Whole Grain Winter Wheat

Abstract

End-use quality improvement is an important objective in most wheat (*Triticum aestivum* L.) breeding programs. Limited sample size, destructive parameter testing, and the short duration between harvest and planting winter wheat are challenges for early-generation selection for end-use quality parameters. This study was conducted to develop near-infrared reflectance (NIR) spectroscopy calibration models for prediction of single kernel characteristics (SKCS; kernel diameter, weight, and hardness index) and grain volume weight, and assess the performance of calibration models for SKCS on an independent validation set and new crop samples. Calibration models were developed using NIR spectra and laboratory reference values from up to 10,000+ samples collected from breeding nurseries under multiple environments during the 2004-2007 growing seasons. Models encompassing all years of data revealed R^2_{val} (validation) in excess of 0.70 for all parameters. Increasing percentages of new crop years increased the R^2_{val} with each incremental percentage of current crop year samples added. The need to add large percentages of new crop year samples declined with each year. Additionally, the R^2_{val} approached the R^2_{cal} (calibration) for all parameters, but did not reach R^2_{cal} except for kernel hardness

index after data from all years were added to the model. Results indicate that the development of a robust calibration model for SKCS characteristics, which could be utilized in a breeding program, is feasible and could aid in the rapid and non-destructive elimination of samples with unacceptable values.

Introduction

End-use quality improvement is an important objective in most wheat (*Triticum aestivum* L.) breeding programs. In the case of winter wheat, the short duration between harvest and planting represents a major challenge to efficiently and timely conduct quality evaluations to enable selection prior to planting. Furthermore, multiple test parameters, some being destructive of the grain sample, are commonly used as predictors of overall end-use quality which complicates the selection process. The large numbers of breeding samples typically handled, even by programs of modest size, represents another limitation in the process of quality evaluation and selection in a wheat breeding program.

While actual milling and baking experiments conducted on samples would provide the best estimate of end-use quality, most early-generation lines are produced in a quantity insufficient for large-scale tests (Dowell, 2008). With samples of limited quantity, kernel characteristics and grain volume weight (test weight) are often used as indicators of milling and baking quality and flour yield.

Increasing kernel weight and size through breeding has been proposed as a method to increase flour extraction, as larger grains have a greater ratio of endosperm to bran (Finney et al., 1987; Wiersma et al., 2001). Increased kernel weight and size has the added effect of increasing the grain volume weight in some samples (Barnard et al., 2002). Additionally, kernel hardness index has

been reported to influence milling and baking properties of the resulting flour (Veha, 2007).

Ohm et al. (1998) observed strong correlation between milling characteristics and both SKCS and grain volume weight with results similar to those reported by other researchers (Finney et al., 1987; Wiersma et al., 2001; Barnard et al., 2002; and Veha, 2007). Lyford et al. (2005) developed a model using grain volume weight and SKCS parameters that accounted for 81% of the variability in mill extraction. With this model, estimates of flour extraction of a sample could be obtained without having to actually mill the sample. As a stand alone test, both SKCS and grain volume weight are relatively rapid techniques, but the large number of samples in early-generation screening, combined with other analyses lead to a substantial time commitment for data collection. In addition, SKCS requires destruction of seed, which may present a challenge in early-generation screening when seed is in limited supply.

The single kernel characterization system (SKCS) 4100 (Perten Instruments, Springfield, IL) was developed by the USDA Grain Marketing and Production Research Center in Manhattan, KS. Measurements of kernel weight, kernel diameter, hardness index, and moisture concentration are determined as an average of a user specified number of kernels. The SKCS 4100 has been accepted as a viable means of measuring these parameters and has found a place in many wheat breeding programs for early- and late-generation testing. While the test can be conducted rather quickly, the test is destructive of the seed.

Near-infrared reflectance (NIR) spectroscopy is a rapid and non-destructive technique that could facilitate early-generation selection for end-use quality. The precision and accuracy of an NIR equation for prediction purposes is dependent on the construction of a reliable calibration (Williams and Norris, 2001). Prediction models can be developed using spectral fingerprints and phenotypic reference data. In this way, a single NIR scan of a wheat grain or flour sample can provide non-destructive estimates for several different quality parameters. The first step in the application of NIR spectroscopy technology in a wheat breeding program is to develop and evaluate prediction models (Williams and Norris, 2001). Spectra collection with an NIR instrument and parameter estimates with a calibration model for kernel characteristics would offer the advantage of rapid, non-destructive sampling (approximately 50 seconds) while providing other parameter estimates such as moisture, ash, and protein concentration.

Several investigators have documented the development of NIR calibrations for wheat end-use quality assessment beyond moisture, protein concentration, and kernel hardness. Pawlinsky and Williams (1998) evaluated the efficiency of prediction models for whole grain wheat for dough mixing quality using samples of elite lines grown over two seasons and observed prediction vs. lab R^2 values ranging from 0.58-0.98. Delwiche et al. (1998a) used 30 entries of hard red winter wheat from eight locations throughout Nebraska and assessed the ability of NIR spectroscopy to predict concentrations of gluten subunits, sedimentation volume, and mixograph dough mixing properties and when

calibrations were applied to an independent validation set, the authors observed relatively high values of R^2 ranging from 0.87 to 0.94. In a variation on the traditional reference method of calibration development, Wesley et al. (1999) used a curve fitting approach to develop predictions based on theoretical spectra from nearly pure samples of gliadin and glutenin with resultant R^2 values of 0.73 for gliadin and 0.76 for glutenin. In a study to determine whether reflectance spectroscopy could be used under commercial bakery conditions to monitor functional performance in hard red spring and hard red winter wheat flours, Delwiche and Weaver (1994) evaluated flour for baking quality of 317 samples, developed calibration models, and reported low R^2 for mixing time, internal grain score, and overall bake score, but high R^2 for absorption (0.92) and intermediate R^2 for loaf height (0.63). McCaig et al. (1993) investigated the utility of using NIR and visible light spectroscopy for assessment of seed coat color and revealed two regions (516 nm and 450 nm) that were useful to discriminate among three wheat groups.. Dowell (1998) went further to evaluate the accuracy of color classification of single wheat kernels using visible and near-infrared reflectance calibration based on the entire visible and near-infrared spectrum that resulted in >99% correct classification for single kernels. Blazek et al. (2005) determined the ability of NIR to predict wheat milling characteristics in 94 experimental and commercial wheat varieties with significant prediction models for some parameters. Delwiche and Graybosch (2002) investigated the feasibility of using NIR spectroscopy to identify and differentiate wild-type, partial waxy, and waxy wheat using 193 whole grain samples and produced a discriminant analysis

model that could correctly classify waxy wheat 93-96% of the time. A relatively complete and comprehensive evaluation of the potential of NIR spectroscopy to predict whole kernel, milling, flour, dough and bread-making quality characteristics from whole grain and flour of 100 hard red spring and winter wheats was conducted by Dowell et al. (2006) where several successful calibration models were developed.

Calibration models for dough mixing and baking properties have met with limited success. Delwiche and Weaver (1994) concluded that the inability of NIR to robustly predict parameters such as dough mixing time, mixing tolerance, and overall bake score was due to the complexity of the interactions between protein, starch, and lipid. This is further complicated by the fact that the exact nature of the relationship between chemical composition and dough rheological properties is not yet fully understood (Blazek et al., 2005).

The aforementioned research has typically used samples of elite lines or cultivars, relatively small sample numbers, or a limited number of environments. To our knowledge, no previous studies have been conducted to develop calibrations within a breeding program utilizing all the inherent phenotypic (genetic and environmental) variation present in such a program. One benefit to this approach is the vast variability present in breeding material that is typically evaluated over multiple years and testing locations. In turn, this would greatly benefit the breeding program in allowing for rapid, inexpensive, and non-destructive prediction of quality parameters in early-generation material and allow for improved genetic gain for heritable traits with each round of selection.

The objectives of this research were to: 1) develop NIR calibration models for kernel characteristics (average kernel diameter, weight, and hardness index) and grain volume weight using samples from breeding nurseries grown over multiple years and testing locations; 2) validate calibrations on independent samples not used in the development of the calibrations; and 3) assess performance of SKCS calibration models on new crop year samples as an indication of utility in real-world breeding scenarios and the number of years required to develop a robust calibration.

Materials and Methods

Samples

Whole grain winter wheat samples were obtained from the Colorado State University Wheat Breeding Program for SKCS and grain volume weight evaluation. To maximize possible variation among samples, multiple test environments (years and locations) were utilized. Additionally, various levels of genetic variation including experimental lines from early-, middle-, and late-generation trials, in addition to released varieties, were utilized. Test environments included rain-fed (dryland) and irrigated nurseries. Fertilizer treatments varied by environment and were applied when necessary in a manner consistent with current farming practices in eastern Colorado. For SKCS parameters, the sample set (n=10,416) included entries from the 2004-2007 crop years over 4-13 locations per year. For grain volume weight, the sample set (n=3,437) included entries from the 2006 and 2007 crop years from a similar range of early-, mid-, and late-generation breeding nurseries. Year and location combinations, along with corresponding sample numbers, are summarized in Table 1.1.

Sample Preparation and Reference Analysis

Following harvest, samples were cleaned, and approximately one kg was sub-sampled. All samples were analyzed for single kernel characteristics using AACC method 55-31 (AACC, 2000) with a SKCS 4100 (Perten Instruments, Springfield, IL); and average kernel weight (KW), kernel diameter (KD), and

kernel hardness index (KHI) were recorded. Grain volume weight of each sample (AACC method 55-10, 2000) was determined using a measuring cup and a Seedburo test weight device (Seedburo Equipment Co., Des Plaines, IL).

<u>NIR Hardware and Spectra Collection</u>

A scanning monochromator NIRSystems 6500 (Foss NIRSystems, Inc., Eden Prairie, MN) was used to measure NIR diffuse reflectance spectra from 400 to 2500 nm at 2 nm intervals. The NIR spectra were collected on approximately 20 g whole grain samples with a ¼ cup sample cell using a standard transport module. The acquisition of NIR spectra (Log 1/R) of whole grain was facilitated by the use of ISI Scan software (Infrasoft Intl. LLC., State College, PA) and collected as an average of 25 scans for a single cell pack for each sample.

<u>Calibration and Validation</u>

The WINISI III software was used for spectral pretreatments, calibration development, and evaluation of calibration performance. Raw spectra (Log 1/R) were first evaluated visually and outliers were removed. Frequency distributions of reference values were plotted and extreme outliers removed. Distributions were then subject to a Shapiro-Wilk normality test (Shapiro and Wilk, 1965). A preliminary analysis of NIR spectra revealed that only 40% of the samples were spectrally unique. Therefore, the samples were randomly separated into two groups: a calibration set (40% of samples) used for calibration development and a validation set (60% of samples) that were not included in the calibration development but were used to evaluate the calibration model.

Table 1.1. Summary of samples and Colorado environments used in calibration development.

Parameter(s)	Year	Environments	Total Environments	*N*
SKCS	2004	AK, FC, HX, JL[†]	4	2722
SKCS	2005	AK, BL, FC, JL, SL, WL	6	1862
SKCS	2006	AK, BL, DL, FC, GN, HX, JL, SL, YM	9	2387
SKCS	2007	AK, BL, DL, FC, FR, GN, HD, JL, MK, OR, SL, WL, YM	13	3445
Grain Volume Weight	2006	AK, BL, DL, FC, GN, HX, JL, WL, YM	9	1625
Grain Volume Weight	2007	AK, DL, FC, JL,	4	1812
SKCS	Total		32	10416
Grain Volume Weight	Total		13	3437

† Environments: AK= Akron; BL= Burlington; DL= Dailey; FC= Fort Collins; FR= Fruita; GN= Genoa; HD= Hayden; HX= Haxtun; JL= Julesburg; MK= Milliken; OR= Orchard, SL= Sheridan Lake; WL= Walsh; YM= Yuma.

Principal component analysis (PCA) was used for the calculation of Mahalanobis distance (H) for the removal of spectral outliers and the removal of redundant spectra. Various mathematical treatments were applied to the absorbance spectra to maximize the accuracy of the calibration model. Treatments included multiplicative scatter correction to minimize the nonlinear effect of light scatter due to particle size differences (none, standard normal variate + detrend, standard normal variate only, and detrend only) and data transformation via derivative mathematics that reduces the intercorrelation between the data points of a spectrum (Isaksson and Naes, 1988). The treatments were applied during the calibration development and included 0,4,4,1; 1,4,4,1; 2,4,4,1; 2,6,4,1; and 3,5,5,1. The first number indicates the order of the derivative with zero (0) representing no derivative, one (1) the first derivative, and two (2) the second derivative of the log 1/R. The second number is the gap interval (in the data points) over which the derivative was calculated. The third and fourth numbers refer to the number of data points used in the first and second smoothing, respectively.

Calibration was performed using modified partial least squares (mPLS) regression available within the WINISI software. The optimum number of terms was determined by cross-validation of the calibration samples. Due to the high noise associated with the 400-1100 nm range, separate calibrations were developed using the 400 to 2500 nm and 1100 to 2500 nm wavelength ranges. Attempts were made to minimize the number of factors, and treatments were selected based on a minimum residual sum of squares. Calibrations were

evaluated using cross-validation and independent test set validation (Williams and Norris, 2001). Performance of the model was assessed with the following statistics: standard error of calibration (SEC), standard error of cross-validation (SECV), standard error of performance (SEP), and the coefficient of determination (R^2) (Williams and Norris, 2001). Also used as an overall measure of performance was the ratio of the standard deviation of the reference values (SD) to the standard error of performance (SEP) (RPD = SD/SEP; Williams and Norris, 2001). After calibrations were developed and refined, the calibration equations were examined to identify key wavelengths underlying the calibrations.

The influence of adding additional sample years on the predictive ability of the calibration models was determined by evaluating the effect of individual years and their contribution to the overall calibration model. Starting with a calibration model built with samples from the first year (2004), samples from the new crop years were added in 10% increments, calibration models were rebuilt, and validated on the remaining samples from that crop year. Resulting coefficients of determination (R^2) were plotted for the calibration set and the validation set. The analysis was conducted for SKCS parameters, but due to insufficient number of years of data, was not conducted for grain volume weight.

Results and Discussion

Summary Statistics

Descriptive statistics for all datasets in all years are summarized in Table 1.2. All datasets were subject to Shapiro-Wilk test for normality prior to division into calibration and validation sets and fit a normal distribution ($P \geq 0.05$). Differences between the calibration and validation sets were minimal in most cases, but overall ranges of the calibration set were generally broader than the validation set. In some cases, samples in the randomly selected validation set fell outside of the range of those in the calibration set. This would force the calibration model to predict samples outside of the range from which it was developed. Therefore, it was necessary to move those samples from the validation set to the calibration set. Means of the 2005 and 2007 sets (calibration and validation) were similar ($P \leq 0.05$) for KD and KHI. The KW means were similar ($P \leq 0.05$) for the 2006 and 2007 crop years. The range of data for KD was similar for all years with the exception of 2004, which had the largest range and highest overall mean. For KHI, the broadest ranges were observed in 2006 and 2007, with 2006 having the largest overall mean. The broadest range for KW was observed in 2004, which also had the largest overall mean. Both KW and KHI had wide ranges, whereas KD had a narrow range. Mean grain volume weight differed between 2006 and 2007 for both the calibration set and validation set, but the mean of the combined samples sets were similar ($P \leq 0.05$) between the

Table 1.2. Descriptive statistics of samples in calibration and validation datasets.

| Constituents | Year | | Calibration Set | | | Validation Set | |
		n	Range	Mean ± SD	N	Range	Mean ± SD
SKCS KD	2004	471	2.23-3.60	2.92 ± 0.23 a[†]	1564	2.31-3.44	2.95 ± 0.22 a[†]
(mm)	2005	286	2.16-3.20	2.68 ± 0.17 b	1038	2.27-3.20	2.67 ± 0.15 b
	2006	310	2.11-3.34	2.72 ± 0.21 c	1346	2.11-3.19	2.75 ± 0.20 c
	2007	320	1.95-3.32	2.64 ± 0.23 b	1966	2.14-3.27	2.66 ± 0.22 b
	All	915	2.11-3.44	2.76 ± 0.24	5914	2.11-3.44	2.76 ± 0.24
SKCS KW	2004	474	19.7-52.9	36.3 ± 5.5 a	1569	23.1-50.1	37.0 ± 5.3 a
(mg kernel^{-1})	2005	288	18.8-41.1	30.0 ± 3.7 b	1061	20.8-40.0	29.6 ± 3.4 b
	2006	308	19.1-42.2	30.7 ± 3.9 c	1352	19.3-40.4	30.9 ± 3.7 c
	2007	321	16.5-44.2	30.4 ± 4.6 c	1964	20.0-42.1	30.6 ± 4.5 c
	All	895	20.3-47.2	32.2 ± 5.2	5946	20.8-42.1	32.2 ± 5.3
SKCS KHI	2004	451	43.0-96.7	69.9 ± 8.9 a	1532	44.1-89.4	68.4 ± 8.8 a
(index)	2005	280	41.9-90.2	66.1 ± 8.1 b	1034	43.5-87.1	64.7 ± 7.9 b
	2006	315	43.1-94.7	68.9 ± 8.6 c	1378	43.5-89.7	69.9 ± 8.0 c
	2007	317	38.4-95.5	66.9 ± 9.5 b	1941	42.1-93.7	64.9 ± 8.6 b
	All	890	39.3-89.1	67.9 ± 8.9	5885	42.1-87.1	66.9 ± 8.7
Grain Volume	2006	496	679.0-875.1	807.7 ± 40.0 a	807	691.6-867.2	823.5 ± 27.8 a
Weight (kg m^{-3})	2007	480	446.6-802.3	721.2 ± 40.9 b	771	556.1-708.1	708.2 ± 45.7 b
	All	632	446.6-875.1	765.7 ± 59.2	1072	556.1-867.2	761.2 ± 69.2

[†] Means within a column followed by the same letter are not significantly different based on LSD test (α=0.05)

calibration and validation sets. Variation in parameter means and ranges among years are likely due to environmental influences where year-to-year effects can be greater than location or genotypic effect.

NIR Calibration and Validation

For SKCS parameters, spectral outliers (Mahalanobis distance H≥3.0) identified in the principal component analysis (PCA) varied among years (from 0-7 samples). Spectrally redundant samples (Mahalanobis distance H≤0.6) were also identified and removed (44.1% for 2004, 39.2% for 2005, 33.6% for 2006, 23.7% for 2007, and 22.6% for the combined analysis). For grain volume weight, similar results were observed with a 26.3% reduction in 2006 and a 25.0% reduction in 2007 due spectrally redundant samples. A greater reduction of samples (50.4%) was observed in the combined analysis for grain volume weight perhaps due to decreased year effect compared to the other parameters.

Overall, all SKCS parameters produced models that would be acceptable for screening (Table 1.3). The combined analysis, including data from all years and locations, produced R^2_{val}≥0.70 for each trait. The SECV values tended to be high, resulting in RPD values ≤2.0, the exception being grain volume weight that had a combined analysis RPD value of 2.62. Since the SKCS testing required sample destruction, a random bulk sample was used for the NIR analysis. This could have contributed to higher variability, thus higher SECV values. Additionally, a large number of principal components (11-16) were necessary to account for the spectral differences among samples for each trait. This was not

Table 1.3. Results of modified partial least squares (mPLS) model development (wavelength 400 to 2500 nm). Models are optimum results obtained through spectral pretreatments including derivative mathematics and scatter correction.

| Parameter | Year | Model | | | | | |
| | | Calibration | | | Validation | | |
		SEC	SECV	R^2	SEP(C)	R^2	RPD
SKCS	2004	0.12	0.13	0.73	0.11	0.72	1.77
Kernel	2005	0.08	0.10	0.77	0.09	0.61	1.70
Diameter	2006	0.09	0.12	0.77	0.11	0.70	1.75
(mm)	2007	0.09	0.11	0.81	0.10	0.78	2.09
	Combined	0.11	0.12	0.76	0.12	0.73	2.00
SKCS	2004	2.4	2.8	0.81	2.5	0.77	1.96
Kernel	2005	1.8	2.1	0.75	2.1	0.64	1.76
Weight	2006	1.8	2.2	0.76	2.2	0.66	1.77
(mg kernel^{-1})	2007	2.2	2.4	0.75	2.3	0.73	1.92
	Combined	2.2	2.6	0.78	2.6	0.74	2.00
SKCS	2004	4.5	4.8	0.73	4.6	0.72	1.85
Hardness	2005	3.7	4.6	0.75	4.3	0.71	1.76
(index)	2006	4.3	4.8	0.75	4.6	0.68	1.79
	2007	3.3	4.4	0.88	4.1	0.78	2.16
	Combined	4.5	4.8	0.74	4.7	0.70	1.85
Grain Volume Weight (kg m^{-3})	Combined	17.9	20.4	0.83	26.5	0.81	2.62

the case for grain volume weight where the NIR analysis sample was taken directly from the reference sample and the number of principal components was 14. Details of calibration development and validation are broken down by individual parameter to allow for comparisons of environmental influences and to analyze results for each parameter independently.

Kernel Diameter

Statistical terms related to the calibration development and analysis on calibration and validation datasets are summarized in Table 1.3. The KD was best modeled by the first-derivative transformation (1,4,4,1) with no multiplicative scatter correction (none) in the combined analysis with the exception of 2006 where a second-derivative transformation (2,6,4,1) and a standard-normal variate + detrend scatter correction resulted in a model with the best theoretical accuracy. The optimum number of terms (as determined by cross-validation) varied among crop years from 9 to 14. The highest R^2 of calibration (R^2_{cal}) was observed in 2007 (0.81) which also showed the highest R^2 of validation (R^2_{val}=0.78).

Regressions of the NIR predicted values vs. reference values for average KD are shown in Figure 1.1. All of the intercepts were greater than zero resulting from NIR predicted values being less than the corresponding reference value. Additionally, the slopes ($\beta_0 < 1$) suggest that the degree of underestimation would increase with increased reference values. The R^2_{val} varied among years with the

Figure 1.1. Average single kernel diameter reference values (y-axis) vs. near-infrared kernel diameter modeled values (x-axis) for individual years and all years combined, evaluated on validation samples (n = number of validation samples used in analysis). Regression line and corresponding equation are best fit linear models.

2004
$y = 0.74x + 0.75$
$R^2 = 0.72$
$n = 1564$

2005
$y = 0.6569x + 0.9132$
$R^2 = 0.609$
$n = 1038$

2006
$y = 0.70x + 0.83$
$R^2 = 0.70$
$n = 1346$

2007
$y = 0.77x + 0.59$
$R^2 = 0.78$
$n = 1966$

Combined
$y = 0.75x + 0.67$
$R^2 = 0.73$
$n = 5914$
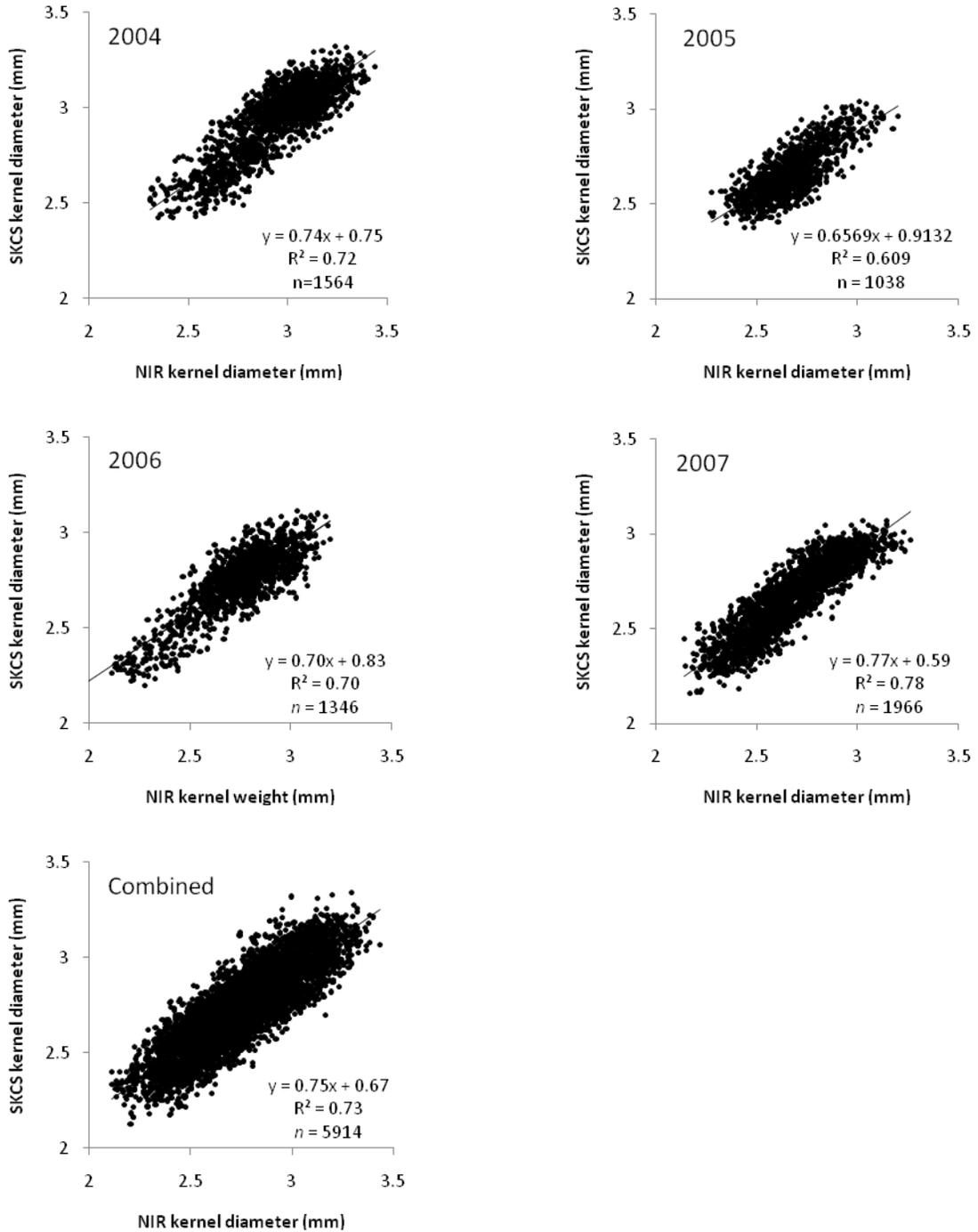
largest (0.78) in 2007 and the smallest in 2005 (0.61). The combined analysis model had R^2_{val}=0.73, meaning that 73% of the variability among the grain samples was explained by the model. In the combined analysis, the relative performance determinant (RPD, the ratio of the standard deviation (SD) to the SECV, Williams and Norris, 2001) was 2.00. While this is lower than the recommended value for screening purposes proposed by Williams and Norris (2001), the regression reveals an obvious association between predicted and reference values. This would indicate utility in a breeding program where the objective is generally to eliminate samples with the lowest values.

Kernel Weight

As observed for KD, KW was best modeled by the first-derivative transformation (1,4,4,1) with no multiplicative scatter correction (none) in all analyses resulted in a model with the best theoretical accuracy. The number of terms varied by year from 8 to 15. Models for SKCS KW produced the highest R^2_{cal} (0.81) and R^2_{val} (0.77) in 2004 (Table 1.3). The 2005 model had the lowest SEC (1.8), SECV (2.1), and SEP (2.1) but also had the lowest R^2_{cal} (0.75) and R^2_{val} (0.64). The combined analysis model resulted in R^2_{val}=0.74, similar to the combined models for SKCS KD (R^2_{val}=0.73) and KHI (R^2_{val}=0.72). The RPD of the combined analysis model (2.00) was lower than recommended for screening (Williams, 2001), but higher than the RPD values for KHI and equal to KD RPD. Results of the regression of NIR predicted values vs. reference values for SKCS KW are shown in Figure 1.2. Slopes of the best-fit linear model for 2006 (β_0=0.71), 2007 (β_0=0.71), and combined analysis models (β_0=0.75) were

Figure 1.2. Average single kernel weight reference values (y-axis) vs. near-infrared modeled values (x-axis) for individual years and all years combined, evaluated on validation samples (n = number of validation samples used in analysis). Regression line and corresponding equation are best fit linear models.
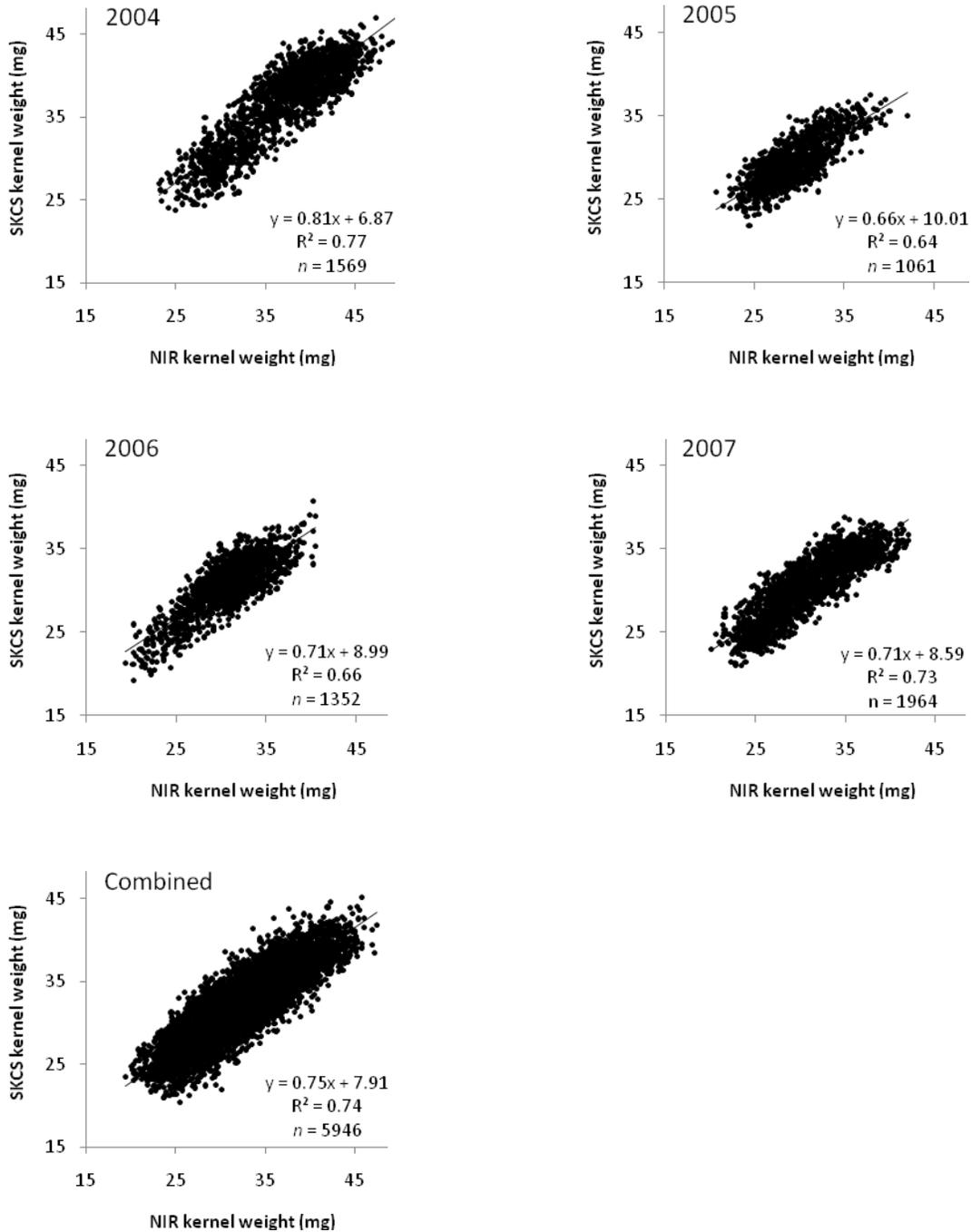
relatively consistent, while 2004 had the greatest slope (β_0=0.81) and 2005 the smallest (β_0=0.66). The small slope (β_0=0.66) and the large intercept (10.01) in the 2005 model may be the result of the limited range of values for KW in the calibration dataset.

Kernel Hardness Index

Results of statistical analysis and calibration development and validation are summarized in Table 1.3. The KHI was also best modeled by the first-derivative transformation (1,4,4,1) with no scatter correction (none), except in 2006 where again a second-derivative transformation (2,6,4,1) with standard-normal variate + detrend provided the best model. Optimum number of terms, as determined by cross-validation, varied by dataset from 5 to 14. The calibration model created with the 2007 data resulted in the highest R^2_{cal} (0.88) and R^2_{val} (0.78) and the lowest SEC (3.3), SECV (4.4), and SEP (4.1). The combined analysis model had the lowest R^2_{cal} and R^2_{val} (0.74 and 0.70) and the highest SEC (4.5), SECV (4.8), and SEP (4.7) and had a RPD value of 1.85. As with KD and KW, although the RPD $\leq$ 2.0 the results of the regression analysis (Figure 1.3) of NIR predicted values vs. reference values indicate predictive ability that could be useful in a breeding program. The PCA analysis and corresponding Mahalanobis (H) distance calculations for elimination of spectral outliers and redundant spectra were similar to those for SKCS KD and KW analysis. Slopes of the best-fit linear models revealed similarities in the 2004, 2005, and combined analysis, while the 2006 model had the smallest slope (0.65) and 2007 the largest (0.86).

Figure 1.3. Average single kernel hardness index reference values (y-axis) vs. near-infrared modeled values (x-axis) for individual years and all years combined, evaluated on validation samples (n = number of validation samples used in analysis). Regression line and corresponding equation are best fit linear models.
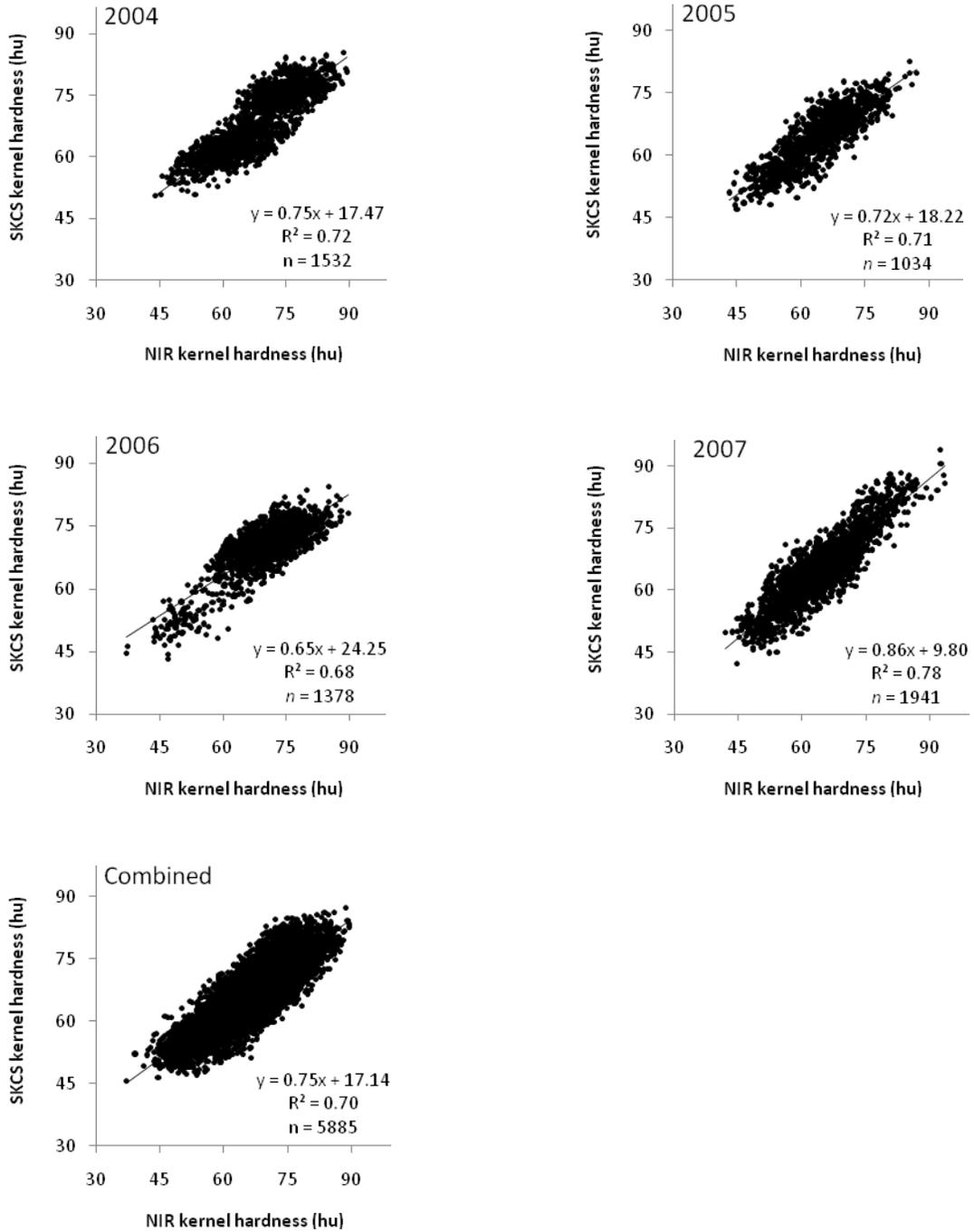
Slopes for all models were again less than one ($\beta_0 < 1$) and would indicate that the model would underestimate the parameter.

Grain Volume Weight

Grain volume weight was best modeled by the first-derivative transformation (1,4,4,1) with standard normal variate and no detrend. Results of the regression on NIR predicted values vs. reference values for grain volume weight are shown in Figure 1.4. Based on the combined analyses, grain volume weight had the largest R^2_{cal} (0.83) and R^2_{val} (0.81) when compared to KD, KW, and KHI, despite the reduced number of samples and years (Table 1.3). The optimum number of terms, determined by cross-validation, was 14 and the RPD of the combined analysis model was 2.62, the largest of calibration models for any of the parameters and higher than the recommended level for screening (Williams, 2001). The slope of the best fit linear model for grain volume weight ($\beta_0 = 0.99$) was the closest to $\beta_0 = 1$ of any of the parameter models. As it is still less than one, this would indicate that model would only slightly underestimate the parameter. The graph reveals a tapering of data points as values increase, suggesting more accurate prediction is possible as values increase. The results of the prediction model for grain volume weight suggest that the prediction model may have utility in a breeding program for screening.

Loading plots from the mPLS regression equations for KD, KW, KHI, and grain volume weight for the combined model do not reveal an obvious chemical basis for any of the parameters (Appendices 1-4). Loadings in the visible

Figure 1.4. Grain volume weight values (y-axis) vs. near-infrared modeled values (x-axis) for and all years combined, evaluated on validation samples (n = number of validation samples used in analysis). Regression line and corresponding equation are best fit linear models.
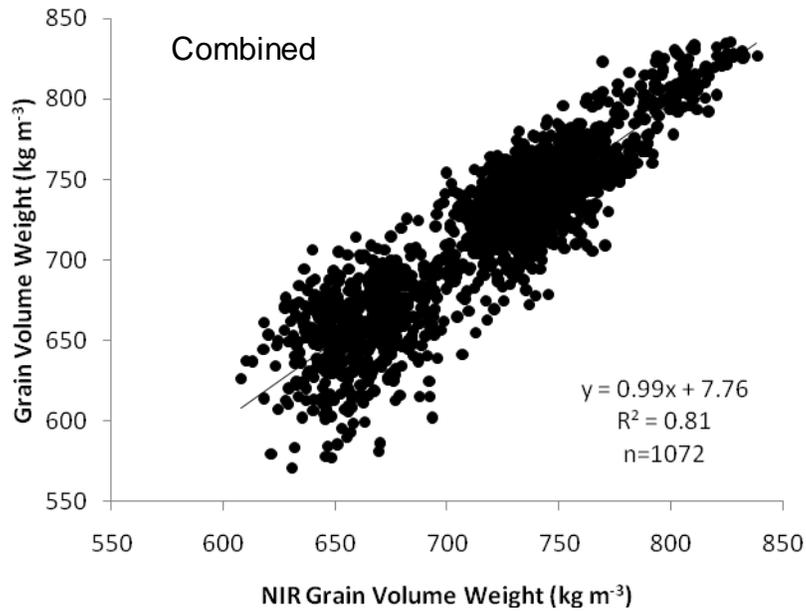
wavelength (400-1100) range have much less influence when compared to wavelengths in the near-infrared range (1100-2500 nm). However, attempts made to develop models with wavelength ranges excluding the visible range resulted in inferior models with lower R^2_{val} and higher SEC, SECV and SEP(C). The influence of visible wavelengths, albeit small, allows for a more theoretically accurate prediction model. The lack of large, isolated, wavelength ranges in the loadings would indicate that the calibration model is acting as an 'optical sieve' and is related more to light scattering effects and particle size than chemical constituents (B. Seabourn, personal communication). This would be further enforced by the results of calibration development, which in all cases except 2006, minimal derivative transformation and no scatter correction was necessary to produce the calibration models.

New Crop Year Analysis

To assess performance of calibration models on new crop year samples and the number of years required to develop robust calibrations, samples were added on a percentage-wise basis for each year and the model's predictive ability on the remaining samples evaluated. For KD, the influence of the addition of new crop year samples on the calibration and the corresponding ability to predict the remainder of that crop year's samples are illustrated in Figure 1.5. Predictive ability of the calibration model including 2005 and 10% of 2006 samples is very low ($R^2_{val}<0.20$) for KD, but increases rapidly as an increasing percentage of 2006 samples are added. Regardless, even when 100% of 2006

Figure 1.5. Effect of adding percentages (%) of new crop year samples on prediction of kernel diameter. Validation was done on remaining samples of the new crop year.
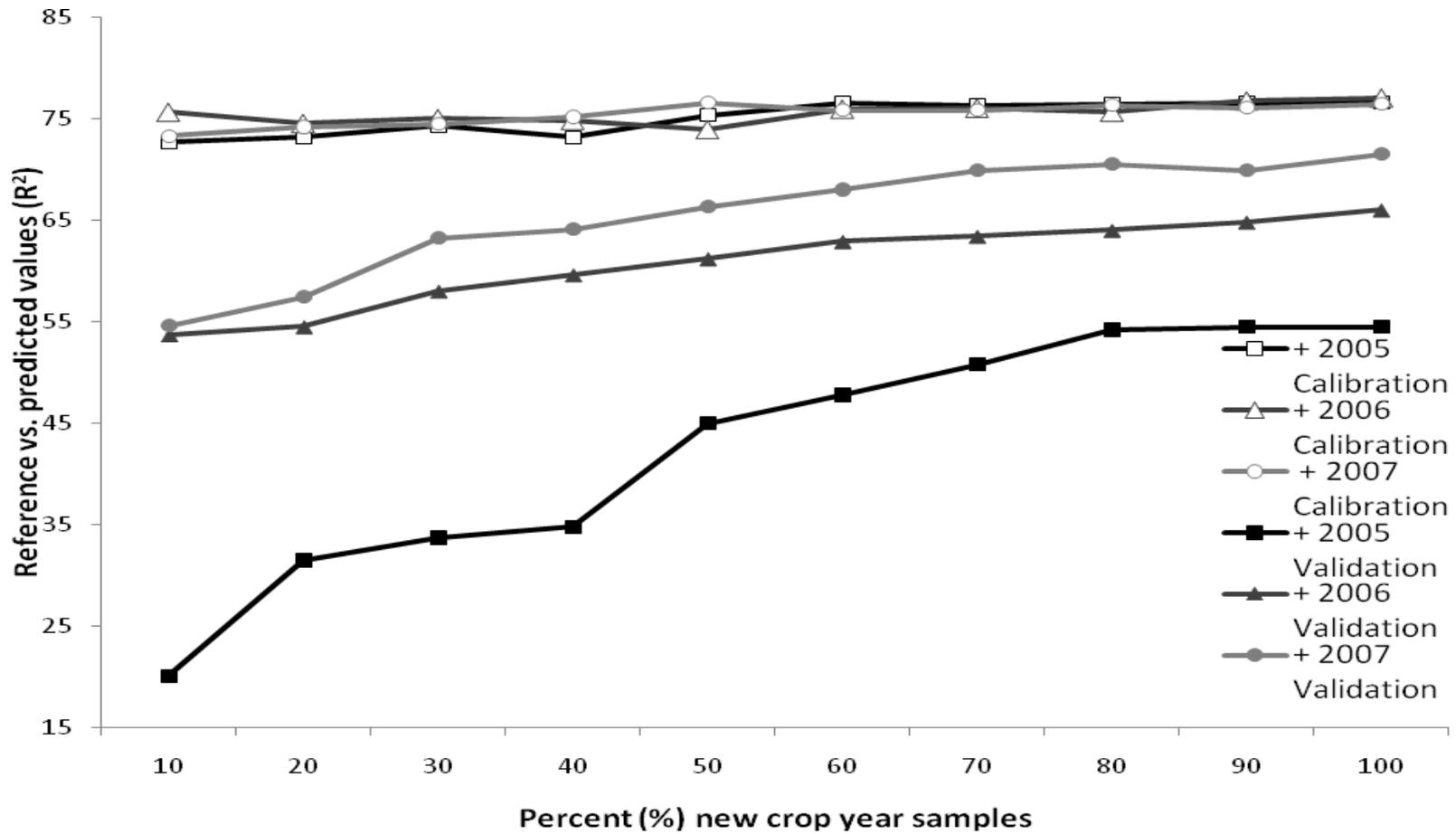
Figure 1.6. Effect of adding percentages (%) of new crop year samples on prediction of kernel weight. Validation was done on remaining samples of the new crop year.
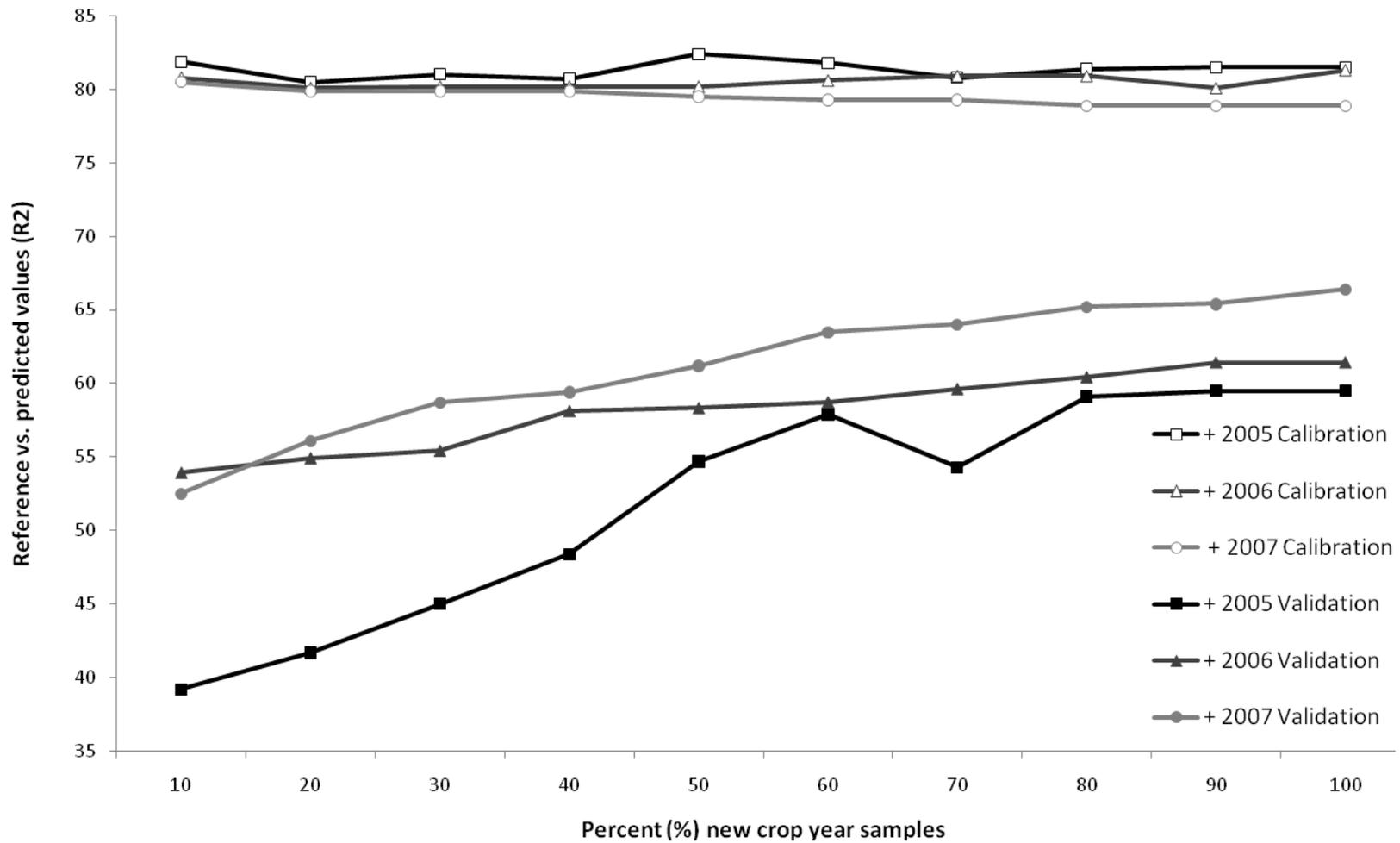
Figure 1.7. Effect of adding percentages (%) of new crop year samples on prediction of kernel hardness index. Validation was done on remaining samples of the new crop year.
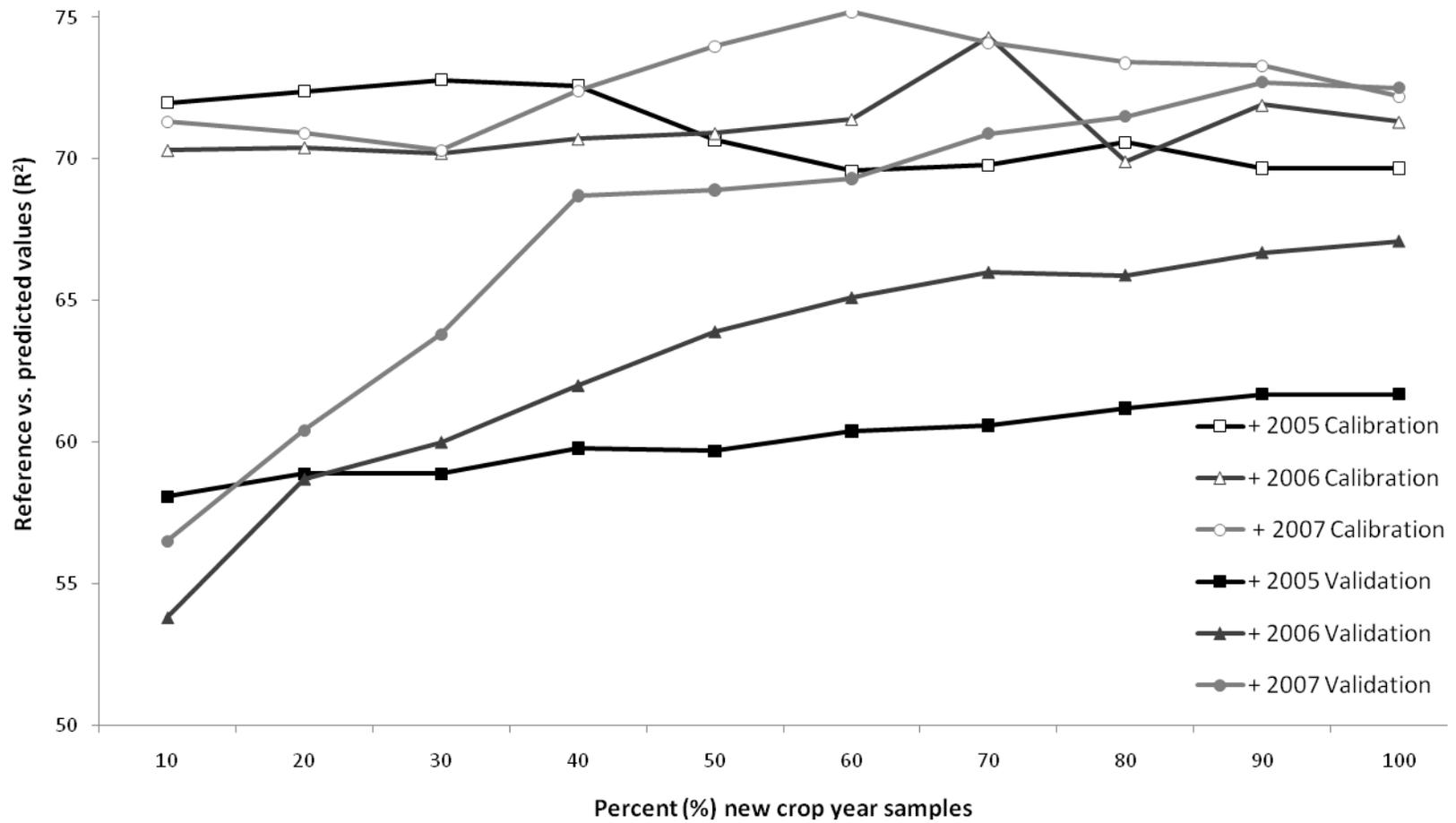
samples are added the R^2_{val} is still less than 0.55. As more crop years are added to the model, predictive ability increases, but still does not reach the level of predictive ability indicated by cross-validation where all models had $R^2_{val}<0.70$. This trend is consistent with KW (Figure 1.6).

Several published studies use a small number of samples encompassing only one or a few years of data and report the predictive ability of a NIR calibration model based on cross-validation results or on a small validation set. Calibrations for KD were consistent with those found by Dowell et al. (2006) where R^2 for calibration models on whole wheat kernels were 0.71, compared to 0.73 found in this study. Dowell et al. (2006) reported a much lower R^2 for KHI of 0.46, whereas in this study a R^2 value of 0.70 was observed. Dowell et al. (2006) reported an R^2_{cal} of 0.74 for NIR predicted grain volume weight with a whole grain winter wheat model, while the results in this study were slightly higher with an R^2_{cal} of 0.83. The addition of many samples over multiple years clearly has an influence on the predictive ability of these models. Even with samples grown under multiple environmental conditions over four years, the predictive ability, in terms of R^2_{val}, does not reach that estimated by the by R^2_{cal} for KD and KW. The model for KHI was an exception, and reveals that after adding all spectrally unique samples over four years, the R^2_{val} approaches that of R^2_{cal} (Figure 1.7). The results may differ for other parameters, but reveal that more sample and crop year combinations are needed for prediction of new crop year samples. Even within the same crop year, the calibration models for KD and KW are not yet robust enough to approach the results indicated by R^2_{cal}.

The trends of increasing predictive ability of calibration models as new crop year samples are added are likely to continue as more information is added to the model, eventually resulting in a calibration model that only requires regular maintenance and the addition of minimal amounts of new crop year samples.

Conclusions

Development of NIR calibration models was met with limited success. Reference data from the SKCS parameters (average kernel diameter, weight, and hardness index) resulted in moderately predictive modified partial least squares models, with models for grain volume weight slightly better. Clustering effects detected in the principal component analysis, due to environmental and genotypic influences, may be one of the reasons for the high number of principal components used and why models failed to provide greater predictive ability. The number of genotypes and environments in our study, while greater than many published studies, has not adequately provided the range of reference data necessary for accurate prediction of new crop year samples. The potential predictive ability observed in R^2_{cal} values indicate potential for improvement of models. This is consistent with the thought that it may take 7-10 years of data, under multiple environments, to develop a robust prediction model (P. Williams, personal communication).

Published research on the development of calibration models for wheat end-use quality parameters has met with mixed results. Parameters such as protein concentration [R^2=0.99 (Delwiche et al., 1998b)], starch damage [R^2=0.92

(Morgan and Williams, 1995)], and particle size [Approved Method 39-70A, NIR method for hardness determination in wheat (AACC, 2000)] have been highly successful. Other parameters such as Alveograph measures [$R^2 \leq 0.25$ (Hruskova et al., 2004)] and baking absorption [$R^2 = 0.65$ (Delwiche and Weaver, 1994)] have failed to result in accurate calibration models. Our resulting R^2 were consistent with Dowell et al. (2006) for SKCS kernel diameter (0.73 vs. 0.71) but R^2 were higher for SKCS hardness index (0.70 vs. 0.46). The objective of Dowell et al. (2006) was not to develop calibration models, but to examine the possibilities of NIR prediction for the grain industry and only limited spectral pretreatments were used. Additionally, this research went further to include breeding material grown under multiple environments.

The predictive models indicate that, while not sufficient for quality control, they may be useful in a breeding program for the elimination of early-generation samples with low parameter values. Improvements and modifications could be made if the analysis were repeated including: 1) only scan the 300 kernels that are to pass through the SKCS, and 2) increase the number of repacks and the overall number of scans.

Near-infrared Reflectance (NIR) Calibration Development for Farinograph Parameter Prediction of Wheat Whole Grain and Flour

Abstract

The Farinograph is used for evaluation of wheat (*Triticum aestivum* L.) flour and provides data predictive of dough rheology and baking quality. Farinograph testing, however, is very time and labor intensive and it requires milling of the sample. While success in the development of flour NIR models for Farinograph parameters has been reported, previous methods have included small sample numbers and few environments. This research, using 500 sample collected from breeding nurseries under multiple Eastern Colorado environments from 2004-2007, was conducted to develop whole grain and flour near-infrared reflectance (NIR) spectroscopy calibration models for prediction of Farinograph parameters (absorption, development time, stability, and mixing tolerance index) and to validate the performance of the models on an independent validation set. Calibration models were developed using NIR spectra (every 2 nm between 400 and 2500) and laboratory reference values. All parameters were significantly correlated with protein concentration for flour (r=-0.31 to 0.58) and whole grain (r=-0.28 to 0.63) with the largest correlations observed with water absorption in both cases. Only Farinograph absorption was effectively predicted using NIR calibration models for whole grain and flour with $R^2 \geq 0.70$. While whole grain NIR models for Farinograph absorption ($R^2_{val}=0.88$) were reasonably high, prediction models for Farinograph development time ($R^2_{val}=0.46$), stability ($R^2_{val}=0.37$), and mixing tolerance index ($R^2_{val}=0.37$) were not promising for selection in a breeding

50

program. Prediction models developed with flour followed the same trends with relatively high predictability for Farinograph absorption (R^2_{val}=0.89) but low predictability for development time (R^2_{val}=0.45), stability (R^2_{val}=0.41), and mixing tolerance index (R^2_{val}=0.38). The results suggest that a whole grain NIR model for Farinograph absorption may be useful as a non-destructive selection tool in wheat breeding programs.

Introduction

With increasing market demands, breeding for end-use quality in winter wheat (*Triticum aestivum* L.) has taken higher priority among the objectives of many breeding programs. Full-scale bake testing of flour obtained from breeding samples provides the most accurate assessment of bread baking quality, but requires milling of the samples, specialized equipment and training, and considerable time. This presents a challenge in winter wheat breeding programs due to the short duration between harvest and planting. Additionally, when dealing with early-generation material, the large numbers of samples and limited seed supply make comprehensive evaluations of quality difficult or prohibitive.

With limited sample supply, one alternative is to use various quantitative and qualitative tests that are predictive of end-use quality (Dowell et al., 2006). Numerous methods have been developed for routine quality assessment (AACC, 2000). A significant amount of research has been conducted to establish the relationships among the values from these tests and the functionality of the flour produced from grain, including evaluation of kernel characteristics, grain volume weight (test weight), dough-handling properties, and various measures of dough strength. The relationship between protein concentration, and several quality measures including water absorption, mixing strength, and loaf volume, have been well established (Graybosch et al., 1993; Preston et al., 1995; Evers and Millar, 2002; Fowler and Kovacs, 2004). Other research has concluded that analysis based on protein concentration alone results in false assumptions about end-use quality. Weegels et al. (1996) concluded that the gluten content in flour

was more important than the protein concentration in determining loaf volume. Environmental effects have also been found to influence protein quality in addition to protein concentration, which in turn has a substantial influence on end-use quality. As protein concentration alone is commonly used for selection of early-generation breeding material, the need for a rapid, non-destructive measure of protein quality is apparent. Significant differences in dough strength can be detected as early as the F_4 filial generation, but at this stage there is limited grain and stocks are valuable for breeding purposes (Pawlinksy and Williams, 1998).

The Farinograph (Brabender GmbH & Co., Duisburg, Germany) is used for evaluation of flour in the baking industry as a predictor of bake water absorption, dough strength, and stability. Farinograph analysis is an approved method of AACC (2000) and Farinograph water absorption is highly correlated with loaf volume (Dowell et al., 2008). Analysis of flour samples with the Farinograph requires milling of the grain and requires a significant amount of time as the test may need to be repeated multiple times to optimize water absorption. The analysis also requires a relatively large sample, although recently the development of 50 g and 10 g Farinograph bowls has reduced the amount required.

Near-infrared reflectance (NIR) spectroscopy is a rapid and non-destructive technique that could facilitate early-generation selection for end-use quality. Prediction models can be developed using spectral fingerprints and phenotypic reference data. In this way, a single NIR scan of a wheat grain or

flour sample can provide simultaneous estimates for several different quality parameters. Spectra collection with an NIR instrument, followed by parameter estimation with a calibration model for Farinograph parameters, would offer the advantage of rapid, non-destructive sampling (approximately 50 seconds) while providing other parameter estimates such as moisture, ash concentration, and protein concentration. An additional benefit would be realized with the development of whole grain calibrations as destruction of the seed would not be necessary for parameter estimation.

Several studies have reported inconsistent results in the development of NIR models for Farinograph parameters. Williams et al. (1988) described the use of NIR models to predict Farinograph parameters in a dataset of 92 hard red spring wheat samples grown in multiple environments with protein range of samples limited to 3%. Williams et al. (1988) reported R^2 values of 0.73 for absorption and 0.86 for stability, and identified key wavelengths used in the calibration model related to protein, oil, and water. Pawlinsky and Williams (1998) later reported R^2 values of 0.91 for absorption, 0.71 for development time, 0.81 for stability, and 0.92 for mixing tolerance index from a NIR model developed with a set of 50 flour samples from two production years. Pawlinsky and Williams (1998) concluded that the efficiency of predictions is affected by growing season, and a significant amount of overall improvement could be achieved by combining samples from more than one season. Dowell et al. (2006) evaluated the potential of NIR to measure multiple quality parameters, including those from the Farinograph, from whole grain wheat and flour on 100 hard red winter wheats

grown over two years under multiple environments. In this study, the authors reported R^2 values of 0.76 for Farinograph water absorption, while all other Farinograph parameters had observed R^2 values < 0.27.

The objective of this study was to develop NIR calibration models for Farinograph parameters (water absorption, development time, stability, and mixing tolerance index) and validate the models on an independent validation dataset. The present study has expanded on the previous studies by incorporating a larger number of unique samples (*n*=500). Furthermore, the samples were collected from breeding nurseries grown under a diverse set of environments over multiple years, some of which are likely to have had an influence on protein quality in addition to protein concentration.

Materials and Methods

Sample Origin, Reference and NIR Analysis

A total of 500 whole grain winter wheat samples were obtained from field trials throughout eastern Colorado. To maximize genotypic and environmental sample variation, samples were obtained from multiple test environments (years and locations) from breeding nurseries grown between 2004 and 2007. Test environments included rain-fed (dryland) and irrigated nurseries. Fertilizer treatments varied by environment and were applied when necessary in a manner consistent with current farming practices in eastern Colorado. Whole grain samples included released cultivars and mid- to late-generation experimental lines. Samples were milled with a modified Brabender Quadrumat Sr. (Brabender GmbH & Co., Duisburg, Germany) mill, and the break and reduction fractions were combined, well mixed, and used for further analysis.

A scanning monochromator NIRSystem 6500 (Foss NIRSystems, Inc., Eden Prairie, MN) was used to measure NIR diffuse reflectance spectra from 400 to 2500 nm at 2 nm intervals for both whole grain and flour. Analyses were performed on 50 ml volumetric measure samples with a ¼ cup sample cell using a standard transport module. The acquisition of NIR spectra of whole grain and flour was facilitated by the use of WINISI III (Infrasoft International LLC., State College, PA) and collected as an average of 25 scans for each sample (one cell-pack), recorded as log(1/R).

Flour samples were evaluated using a Farinograph-E (Brabender GmbH & Co., Duisburg, Germany) in accordance with Approved method 54-21 (AACC, 2000) with a 50 g bowl and sample weight adjusted to 140 g kg^{-1} (14%) moisture basis. Samples were rerun when necessary to optimize absorption to peak development at 500 ± 20 Brabender units (BU). Parameters measured by the Farinograph included absorption, development time, stability, and mixing tolerance index. Farinograph water absorption is a measure of the water-carrying capacity of the flour; development time is the time required for the flour to reach maximum strength; and stability is a measure of tolerance to over-mixing. Mixing tolerance index is the difference, in Brabender or Farinograph units, between the top of the curve at the peak and the top of the curve 5 min after the peak is reached.

Statistical Analysis and Calibration Development

Statistical analyses were conducted using SAS/STAT software, version 9.1.3 (SAS Institute Inc., 2004). The UNIVARIATE procedure was used for the calculation of summary statistics and the CORR procedure was used to calculate Pearson correlation coefficients between whole grain and flour protein concentration and the Farinograph parameters.

Commercial spectral analysis software (WINISI III, Infrasoft International LLC., State College, PA) was used for spectral pretreatments, calibration development, and evaluation of calibration performance. Spectra were first evaluated visually and outliers were removed. A preliminary analysis of the NIR

spectral data indicated that 60% of the samples were spectrally unique, therefore samples were then randomly separated into two groups: a calibration set (60% of samples) used for calibration development and a validation set (40% of samples) that was not included in calibration development but was used to evaluate the calibration model.

Principal component analysis (PCA) was used for the calculation of Mahalanobis distance (H) for the removal of additional outliers based on spectral data and the removal of spectrally redundant samples. Various mathematical treatments were applied to the absorbance spectra to maximize the accuracy of the calibration model. Treatments included multiplicative scatter correction to minimize the nonlinear effect of light scatter due to particle size differences (none, standard normal variate + detrend, standard normal variate only, and detrend only) and data transformation via derivative mathematics that reduces the intercorrelation between the data points of a spectrum (Isaksson and Naes, 1988). The treatments were applied during the calibration development and included 0,4,4,1; 1,4,4,1; 2,4,4,1; 2,6,4,1; and 3,5,5,1. The first number indicates the order of the derivative with zero (0) representing no derivative, one (1) the first derivative, and two (2) the second derivative of the log 1/R. The second number is the gap interval (in data points) over which the derivative is calculated. The third and fourth numbers refer to the number of data points used in the first and second smoothing, respectively.

Calibration was performed using modified partial least square (mPLS) regression available within the WINISI software. The optimum number of terms was

58

determined by cross-validation of the calibration samples. Calibrations were developed using the 400 to 2500 nm wavelength range. Attempts were made to minimize the number of factors, and treatments were selected based on a minimum residual sum of squares. Calibrations were evaluated using cross-validation and independent validation (Williams and Norris, 2001). Performance of the model was assessed with the following statistics: standard error of calibration (SEC), standard error of cross-validation (SECV), standard error of performance (SEP), and the coefficient of determination (R^2) (Williams and Norris, 2001). To measure overall performance of the calibrations, the residual prediction deviation (RPD) was also used, defined by Williams and Norris (2001) as the ratio of the standard deviation of the reference values (SD) to the SEP (RPD = SD/SEP).

Results and Discussion

<u>Summary Statistics</u>

Descriptive statistics for the calibration and validation datasets for Farinograph absorption, development time, stability, and mixing tolerance index are presented in Table 2.1. The range, mean and standard deviation for all parameters are relatively consistent between the calibration and validation sets. The difference of range of Farinograph development time between the calibration and validation datasets is the only difference of note. Regardless of this difference, the mean and standard deviation of both sets are consistent, indicating that the data used in calibration development are consistent with the data used for validation of the NIR models. In some cases, the values of entries in the validation set fell outside the range of entries in the calibration set. In these cases, the samples were removed from the validation set and added to the calibration set. The results of the summary statistics indicate that a representative data set was used in both calibration and validation.

Pearson correlations between both whole grain and flour protein concentration and Farinograph parameters are summarized in Table 2.2. All parameters were significantly correlated ($P \leq 0.001$) with protein concentration of both whole grain and flour, and all but Farinograph mixing tolerance index showed positive correlation. Results were consistent among flour and whole grain with regard to direction and strength of the correlation. Correlations were intermediate for Farinograph absorption and development time (r=0.50 to 0.63),

Table 2.1. Summary statistics for Farinograph water absorption, development time, stability, and mixing tolerance index for samples used in NIR calibration development and independent validation.

Parameter	Calibration				Validation			
	n	Range	Mean	SD	n	Range	Mean	SD
Water Absorption (g kg^{-1})	268	506-712	629	32	198	561-710	630	33
Development time (min)	281	0.3-20.0	9.0	5.8	198	0.3-20.0	9.2	5.9
Stability (min)	281	1.5-19.9	11.6	5.3	200	1.5-19.3	11.5	5.3
Mixing tolerance index (BU)	203	4.0-142.0	29.7	21.7	198	4.0-122.0	30.1	21.8

Table 2.2. Pearson correlation coefficients between whole grain and flour protein concentration and Farinograph parameters (absorption, development time, stability, and mixing tolerance index).

	Whole grain protein concentration	Flour protein concentration
Farinograph water absorption	0.63[†]	0.58
Farinograph development time	0.50	0.52
Farinograph stability	0.21	0.22
Farinograph mixing tolerance index	-0.28	-0.32

[†] All coefficients were significant at $P \leq 0.001$.

and low for stability (r=0.21 to 0.22). A negative correlation was observed between protein concentration and Farinograph mixing tolerance index in both flour and whole grain. Maghirang et al. (2006) reported that for hard red wheat, the only observed Farinograph parameter that had high correlation with protein concentration ($r \geq 0.70$) was absorption. Pawlinsky and Williams (1998) reported correlation for Farinograph parameters and flour protein concentration of r=0.48 to 0.68 for absorption, r=0.11 to 0.44 for development time, r=0.08 to 0.51 for stability, and r=0.34 for mixing tolerance index. Results from Pawlinsky and Williams (1998) varied greatly by year for all parameters, except Farinograph mixing tolerance index. These results indicate that protein concentration is correlated with all Farinograph parameters, but the parameters cannot be explained solely by protein concentration alone.

Calibration Models

Results of the whole grain NIR calibration and validation are shown in Table 2.3. All whole grain NIR calibrations were best modeled with a second-derivative transformation and standard-normal variate and detrend scatter correction. The number of factors varied by parameter with 11 factors for Farinograph absorption and stability, nine factors for development time, and 10 factors for mixing tolerance index. Only one of the Farinograph parameters (absorption) was predicted with an $R^2_{val} \geq 0.70$, a value suggested by Williams and Norris (2001) as suitable for rough screening in a breeding program. Dowell et al. (2006) reported R^2_{val} values of whole grain NIR prediction of 0.63 for

63

Table 2.3. Summary statistics for whole grain calibration development, cross-validation, and independent validation for Farinograph absorption, development time, stability, and mixing tolerance index.

Parameter[†]	Calibration				Validation			
	n	SEC	SEC(V)	R^2	N	SEP	R^2	RPD
Absorption (g kg^{-1})	268	11	19	0.87	198	12	0.88	1.74
Development time (min)	281	4.2	4.4	0.47	198	4.6	0.46	1.32
Stability (min)	281	3.8	4.3	0.45	200	4.2	0.37	1.23
Mixing tolerance index (BU)	203	10.9	12.3	0.50	198	18.1	0.37	1.76

[†] SEC is standard error of calibration; SEC(V) is standard error of cross-validation, and SEP the standard error of prediction.

Farinograph absorption, 0.29 for development time, 0.06 for stability, and 0.14 for mixing tolerance index. These results were more promising in all cases with R^2_{val} values of 0.88 for Farinograph absorption, 0.46 for development time, 0.37 for stability, and 0.37 for mixing tolerance index. It is important to note that the objective of the study by Dowell et al. (2006) was to determine the feasibility of NIR modeling for multiple parameters and only a first-derivative spectral pretreatment was used, likely explaining the difference in results from this study.

The regression of Farinograph water absorption on whole grain NIR predicted absorption is presented in Figure 2.1. The y-intercept of the best fit linear regression model was 82.4, suggesting that the NIR prediction whole-wheat model would underestimate the reference absorption values. The slope of the regression line was less than one, indicating that the underestimation would decrease with increased Farinograph absorption reference values. Overall, there were no obvious outliers in the validation set. The validation set results were indicative of a typical breeding scenario where outliers are not known or removed from the analysis.

The flour NIR calibration models for Farinograph water absorption and stability were best modeled with a second-derivative transformation and no scatter correction, and best modeled for development time and mixing tolerance index with a first-derivative transformation with detrend. Farinograph absorption was again the only model with an $R^2_{val} \geq 0.70$ (Table 2.4). Williams et al. (1988) reported results of NIR prediction using a small sample size hard red spring wheat flour in terms of coefficients of correlation (*r*) between the NIR and

65

Figure 2.1. Farinograph water absorption (ABS) reference values (y-axis) vs. whole grain near-infrared modeled values (x-axis), evaluated on validation samples (n = 198). Regression line and corresponding equation are best fit linear models.

Plot: Reference Farinograph ABS (g kg^{-1}) on y-axis (550–750) vs. NIR Predicted Farinograph ABS (g kg^{-1}) on x-axis (550–750).

$y = 0.869x + 82.48$
$R^2 = 0.8782$

Table 2.4. Summary statistics for flour calibration development, cross-validation, and independent validation for Farinograph water absorption, development time, stability, and mixing tolerance index.

Parameter[†]	Calibration				Validation		
	n	SEC	SEC(V)	R^2	N	SEP	R^2
Absorption (g kg^{-1})	268	12	15	0.88	198	14	0.89
Development time (min)	281	4.1	4.5	0.51	198	4.5	0.45
Stability (min)	281	3.8	4.2	0.48	200	4.2	0.41
Mixing tolerance index (min)	203	13.5	14.4	0.53	198	14.4	0.38

[†] SEC is standard error of calibration, SEC(V) is standard error of cross-validation, and SEP(C) the standard error of prediction.

reference analysis, and observed values of 0.73 for Farinograph absorption and 0.86 for stability. Williams et al. (1988) reported more promising results for Farinograph stability than those reported by Dowell et al. (2006), where flour NIR models for hard red winter wheat had R^2_{val} values of 0.63 for Farinograph absorption, 0.29 for development time, 0.06 for stability, and 0.06 for mixing tolerance index. Pawlinsky and Williams (1998) observed NIR prediction results of 0.91 for Farinograph absorption, 0.71 for development time, 0.80 for stability, and 0.92 for mixing tolerance index. The results of Dowell et al. (2006) are more consistent with this study where NIR models for wheat flour resulted in R^2_{val} values of 0.89 for Farinograph absorption, 0.45 for development time, 0.41 for stability, and 0.38 for mixing tolerance index. Pawlinsky and Williams (1998) concluded that the efficiency of NIR predictions is affected by the growing season and speculated that an overall improvement in prediction efficiency could be achieved by combining samples from more than season in calibration development. This conclusion is not supported by the results of this study where multiple growing seasons and environments were included, and the NIR prediction efficiencies were not greatly improved. This may be due to varied environments used in this study.

The regression of Farinograph water absorption on flour NIR predicted absorption is presented in Figure 2.2. The results are similar to those for the whole grain model as the y-intercept of the best fit linear regression model is greater than zero (55.5), suggesting that the whole grain NIR prediction model would underestimate the reference absorption values. The underestimation

Figure 2.2. Farinograph water absorption (ABS) reference values (y-axis) vs. flour near-infrared modeled values (x-axis), evaluated on validation samples (n = 198). Regression line and corresponding equation are best fit linear models.
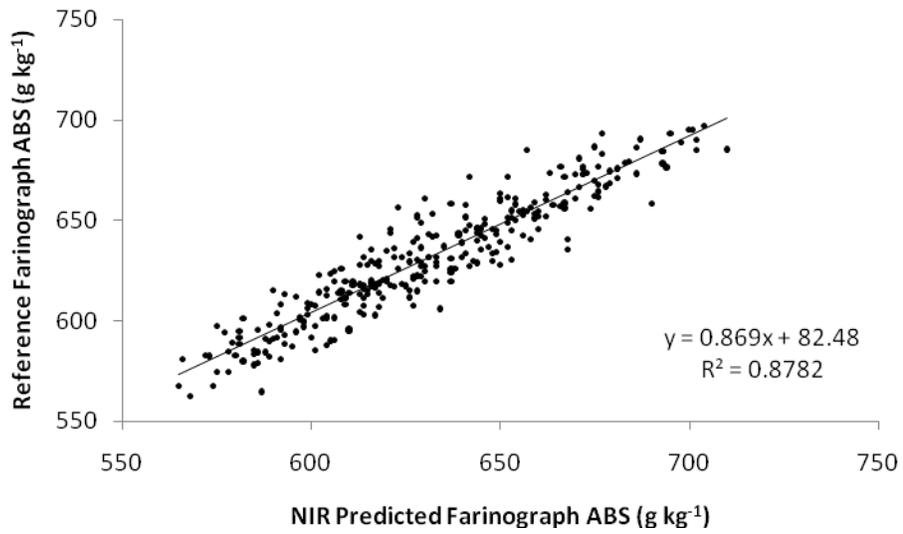

$y = 0.912x + 55.52$
$R^2 = 0.8933$
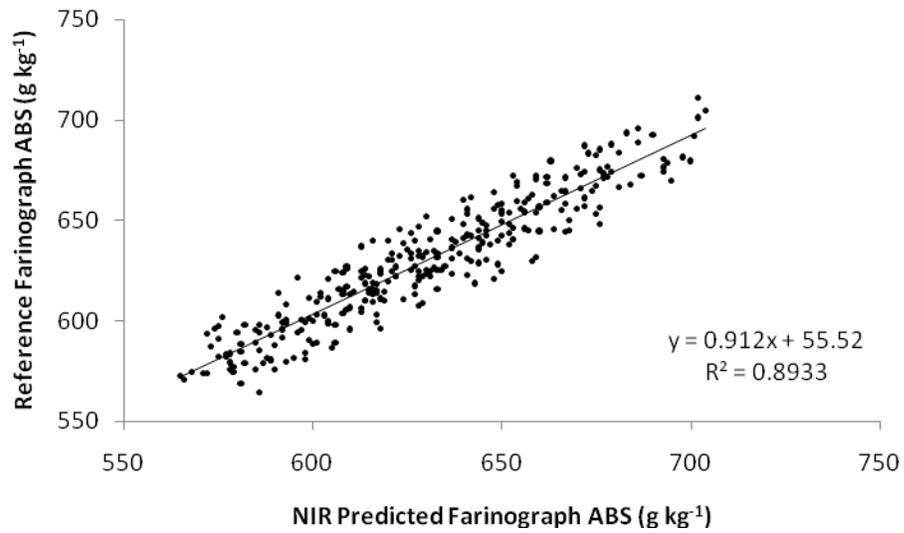
would be less than observed in the whole grain model. Again, the slope of the regression line is less than one, indicating that the underestimation would decrease with increased Farinograph absorption reference values. As reflected in the R^2 values, the flour NIR prediction model also reveals less deviation from the regression than observed in the whole grain model.

Conclusions

Moderate correlation coefficients (r) for both flour and whole grain were observed with protein concentration and Farinograph absorption, but low-intermediate correlations between protein concentration with other Farinograph parameters were observed. These results suggest that protein concentration is likely contributing to the calibration models, especially Farinograph absorption, but the parameters cannot be explained by protein concentration alone. Only Farinograph absorption was effectively predicted using NIR calibration models for whole grain and flour with $R^2 \geq 0.70$. Williams and Norris (2001) suggested $R^2 = 0.70$ to 0.90 is suitable for rough screening, $R^2 = 0.90$ to 0.97 is suitable for screening or quality control, $R^2 = 0.97$ to 0.99 is suitable for process control, and larger values are suitable for most applications. The results of NIR prediction of Farinograph parameters in this study suggest that whole grain and flour models for Farinograph absorption may have utility for rough screening. Results of R^2 values for the remaining parameters were not as large as reported in some previous literature yet larger than found in other studies, but are unlikely to have utility in breeding for end-use quality.

Utility of Near-infrared Spectroscopy Based Selection for Wheat

Single Kernel Characteristics and Grain Volume Weight

Abstract

End-use quality improvement is an important objective in most wheat (*Triticum aestivum* L.) breeding programs. Grain volume weight (test weight) and single kernel characterization system (SKCS) parameters are commonly used for selection of early-generation breeding material. While effective methods, both are performed independently and the SKCS requires destruction of seed which prevents the sample from being replanted. An alternative would be the development and implementation of NIR prediction models where estimates from multiple parameters are returned in a single, rapid, and non-destructive scan. This study was conducted to determine the selection efficiency and accuracy of selection based on realized heritability (h^2_r) and classification errors when using NIR spectroscopy as a selection tool for grain volume weight and SKCS parameters (kernel diameter, weight, and hardness index). One hundred $F_{3:4}$ lines from each of three single-cross populations were randomly selected in 2006, and planted in replicated field trials in 2007 in four eastern Colorado environments. Using NIR calibration models for SKCS parameters grain volume weight, realized heritability was estimated as a response to selection using NIR predicted values and laboratory reference values. Heritability estimates from the

reference method varied by environment and among populations, but in general were intermediate-high (0.40 to 0.75) for grain volume weight, kernel weight, and hardness index, but were low-intermediate (0.16 to 0.45) for kernel diameter. The heritability estimates were generally higher when using the reference values, but suggested that genetic gain was possible when using NIR models for selection. Classification errors were observed when using the NIR models, but in general were not extreme misclassifications. The results suggest that while not as accurate as the reference selection method, the speed and efficiency of NIR-based selection may be effective for early-generation selection.

Introduction

There are standard and recommended measurement methods to determine end-use quality in winter wheat (*Triticum aestivum* L.). However, the methods are generally costly and time consuming, and many cannot be used to rapidly measure quality characteristics and functionality (Dowell et al., 2006). In addition to the more time-consuming flour and bread-baking quality characteristics, various other milling-related parameters are also of interest such as grain volume weight and kernel diameter, weight and hardness from the single kernel characterization system (SKCS; SKCS 4100, Perten Instruments, Springfield, IL). Kernel characteristics derived from the SKCS and grain volume weight have been accepted as viable means for evaluation of early-generation samples and are indicative of milling extraction.

Ohm et al. (1998) observed strong correlation between milling characteristics and both SKCS and grain volume weight with results similar to those reported by other researchers (Finney et al., 1987; Wiersma et al., 2001; Barnard et al., 2002; and Veha, 2007). Lyford et al. (2005) developed a model using grain volume weight and SKCS parameters that accounted for 81% of the variability in mill extraction. With this model, estimates of flour extraction of a sample could be obtained without having to mill the sample. As a stand alone test, both SKCS and grain volume weight are relatively rapid techniques, but the large number of samples in early-generation screening, combined with other analyses lead to a substantial time commitment for data collection. In addition, SKCS requires

destruction of seed, which may be limited for replanting for further field evaluation.

Near-infrared reflectance (NIR) spectroscopy is a rapid and non-destructive technique that could facilitate early-generation screening in breeding programs (Williams and Norris, 2001). Recent advances that facilitate calibration development and permit whole grain analysis make NIR technology a promising tool for prediction of wheat end-use functionality. The development of whole grain NIR calibration models offers an advantage to winter wheat breeders for early-generation screening, when seed supplies are limited and a quick turnaround is required. It also eliminates the need for the time consuming step of milling the samples for analysis of the flour and mill extraction. An additional benefit is that a single NIR scan could return predicted estimates for multiple parameters.

Several investigators have documented the development of NIR calibrations for wheat end-use quality assessment beyond simple moisture, protein concentration, and grain hardness. Models have been developed for prediction of parameters including those obtained by the SKCS and grain volume weight (Chapter 1 of this dissertation; Dowell et al., 2006); degree of starch damage (Morgan and Williams, 1995); the presence of 1AL.1RS and 1BL.1RS wheat-rye chromosomal translocations (Delwiche et al., 1999); and dough-handling properties (Delwiche et al., 1998b).

One of the most important aspects of evaluation of the quality characteristics of genetic material is the degree to which they are heritable (Pawlinsky and

Williams, 1998). High heritability estimates suggest that there is genetic influence that can be exploited in a breeding program. In early-generation selection, the main focus of a breeding program is to eliminate lines that perform poorly in environments representative of the target growing area. Using NIR spectroscopy for early-generation selection could allow rapid elimination of lines with undesirable quality characteristics. While NIR calibration models may not explain all of the variation for the reference values, it may be useful for screening early-generation breeding materials. One risk in using a less than perfect prediction model is incorrect classification and elimination of lines with superior trait values or selection of lines with inferior trait values. By incorrectly classifying lines, the heritability and selection efficiency would be reduced. The risk of using a less than perfect calibration model can be examined by characterizing classification errors.

The reliability of selection in a breeding program using NIR-based estimates of grain quality parameters has not been reported in the literature. For a breeding program, heritability of the evaluation method for any trait is a primary criterion of interest as this has a direct bearing on genetic gain over time. By comparing heritability of selection using the reference method with that of the NIR prediction, the usefulness of NIR as an early-generation selection tool may be determined. While several reports suggest the promise of NIR in wheat quality assessment, no whole grain NIR prediction models have been evaluated in the context of a breeding program for selection of early-generation samples with desirable characteristics. The objectives of this study were therefore to assess the utility of

NIR-based selection for grain volume weight and SKCS parameters by: 1) comparing the realized heritability of selection using NIR-based and conventional methods, and 2) determining classification errors resulting from NIR prediction of SKCS parameters and grain volume weight.

Materials and Methods

Population Development and Reference Analysis

Populations were developed using a modified bulk breeding procedure. All early-generation population and line development was done in the greenhouse or an irrigated field-testing location at Fort Collins, CO. The crosses between parents CO980376/TX97V2838, CO99W329/'Overley', and TX97V2838/ CO99W188 were made in the greenhouse and F_1 seed was harvested and planted in a field nursery. The parents CO980376, CO99W329, and CO99W188 are unreleased experimental lines from Colorado State University. The TX97V2838 is an experimental line from Texas A&M University and Overley is a released variety from Kansas State University. The populations were selected based on contrasting performance of the parents for grain volume weight and SKCS weight and diameter. Seed from the F_1 plants was harvested in bulk and planted in an unreplicated F_2 bulk nursery. The F_2 population was harvested in bulk with a small-plot combine and was planted in an unreplicated F_3 bulk nursery. Populations were advanced by random sampling of approximately 300 spikes harvested at maturity from the F_3 bulks. Selected spikes were threshed individually and planted in an irrigated $F_{3:4}$ headrow nursery in a paired-row arrangement 1 m long with 23 cm spacing between rows. One-hundred lines were randomly selected from $F_{3:4}$ headrow populations of each of the three single-cross populations in the 2005-2006 crop year.

The F$_{3:4}$ populations were planted in fall 2006 in short paired-rows (1 m long, 23 cm spacing between rows) at four locations in eastern Colorado (Akron, Dailey, Fort Collins, and Julesburg). Akron, Dailey, and Julesburg were rain-fed (dryland), and Fort Collins was sprinkler irrigated. Experimental units were replicated twice in a latinized row-column arrangement to allow for analysis of spatial variation. A planting rate of 2.47 million seeds ha^{-1} was used for each entry based on kernel weight counts.

Plots were harvested with a small-plot combine, cleaned, and dockage-free samples were immediately subjected to grain volume weight, SKCS, and NIR analysis. Single kernel characteristics were determined by AACC method 55-31 (AACC, 2000) using a SKCS 4100 and average kernel weight (SKCS weight), average kernel diameter (SKCS diameter), and average kernel hardness index (SKCS hardness index) on approximately 300 kernels were recorded. Grain volume weight of each sample (AACC method 55-10, 2000) was determined using a measuring cup and a Seedburo test weight device (Seedburo Equipment Co., Des Plaines, IL).

Calibration Development and Evaluation

A scanning monochromator NIRSystem 6500 (Foss NIRSystems, Inc., MN) was used to measure NIR diffuse reflectance spectra from 400 to 2500 nm at 2 nm intervals. Analyses were performed on approximately 20 g whole grain samples with a ¼ cup sample cell using a standard transport module. The acquisition of NIR spectra was accomplished using ISI Scan (Infrasoft

International LLC., State College, PA) and collected as an average of 25 scans for each sample (one cell-pack), recorded as log(1/R).

Commercial spectral analysis software (WINISI III, Infrasoft International LLC., State College, PA) was used to develop the calibration equations and evaluate the calibration performance. Calibrations for parameters developed by Butler (Chapter 1) for grain volume weight and SKCS kernel diameter, weight, and hardness index were used for NIR predicted estimates. All NIR calibration models used in this study were developed with samples from breeding nurseries grown over several years in multiple environments. The samples used in this study were not included in the development of the calibrations, but were used for validation of the models.

<u>Statistical Analysis</u>

Statistical analyses were conducted using SAS/STAT software, version 8.2 (SAS Institute Inc., 1999). To account for spatial variability, a SAS PROC MIXED anisotropic model was used to adjust trait values on the basis of spatial covariance analysis of adjacent plots (Butler et al., 2005). Rows and columns were considered as random effects and entries as fixed effects. Least square means (LS means) were estimated separately within each environment for each trait within each population. Separate LS means were calculated based on reference values and NIR predicted values. The LS means from both methods were used for all further analyses. Least significant difference (LSD) was estimated among populations within environments (α=0.05).

Heritability

Narrow sense heritability (h^2_N) as a function of parent-offspring correlation was estimated for each trait by calculating the correlation coefficient r between the values observed on the $F_{3:4}$ and the value observed in the $F_{3:5}$ (Frey and Horner, 1957). Adjustments were made to account for inbreeding according to Smith and Kinman (1965). Narrow sense heritability was calculated independently using the NIR values and the reference values. The standard error of h^2_N was calculated as:

$$S_{h^2_N} = \sqrt{\frac{s^2_{xy}}{\sum (x - \bar{x})^2}}$$

where s denotes the standard error (Motzo et al., 2004).

Realized heritability (h^2_r) estimates for SKCS characteristics and grain volume weight were calculated based on reference and NIR-predicted data from each environment for grain from $F_{3:4}$ and $F_{3:5}$ lines. The heritability of a trait can be estimated by the amount of genetic improvement that is realized by selection within a population (Falconer, 1981). Using the formula $h^2_r = R/S$, where R is the response realized by selection and S is the selection differential, genetic gain can be determined (Fehr, 1993). The selection differential is the difference between the mean of the individuals selected from a population and the overall mean of the population from which they were selected resulting in the following formula:

$$h_r^2 = \frac{(\bar{X}_{Selected_{F_{3:5}}} - \bar{X}_{population_{F_{3:5}}})}{(\bar{X}_{Selected_{F_{3:4}}} - \bar{X}_{population_{F_{3:4}}})}$$

Selection intensity in this study was 0.30, where the top 30% of lines were selected. By comparing realized heritability from the reference and NIR-based analyses, the genetic gain realized by each selection method was estimated.

Classification errors

The $F_{3:4}$ lines were ranked based on reference and NIR predicted parameter values and assigned into three separate, equally-sized classification groups (TOP, MID, BOTTOM). Samples were classified independently using data obtained from the reference analysis and NIR-based analysis. Results were entered into a jackknife classification table to analyze the percentage of samples that were incorrectly classified when using the NIR-based analysis. From this, the severity of misclassification (e.g., TOP reference samples incorrectly classified as BOTTOM by NIR) when using NIR-based analysis was assessed.

Results and Discussion

<u>Summary statistics</u>

Analysis of variance showed highly significant variation ($P \leq 0.001$) among the 100 $F_{3:5}$ lines in each population, at each of the four locations for all parameters. Ranking of the population means varied by environment and revealed some trends (Table 3.1). Greater ($P \leq 0.001$) grain volume weight values (overall average 756.3 kg m^{-3}) were observed at Fort Collins than any of the other environments, likely due to positive effects of irrigation on grain volume weight maintenance. Drought stress was evident at Julesburg and was reflected in an overall grain volume weight of 659.8 kg m^{-3} which was lower ($P \leq 0.001$) than the other locations. Greater ($P \leq 0.001$) grain volume weight values for population CO99W329/Overley were observed at Fort Collins and Dailey, whereas populations CO980376/TX97V2838 and TX97V2838/CO99W188 had greater ($P \leq 0.001$) grain volume weight values at Akron and Julesburg. Mean SKCS kernel weight at Julesburg (26.2 mg kernel^{-1}) and Dailey (26.9 mg kernel^{-1}) were lower ($P \leq 0.001$) than the other two environments, likely due to drought stress. The greatest ($P \leq 0.001$) SKCS kernel weight was observed at Fort Collins, the only irrigated environment, with a combined mean of 35.3 mg kernel^{-1}.

Population means for SKCS kernel diameter were not different ($P \leq 0.001$) within the two environments with the greatest values (Akron and Fort Collins). Fort Collins had greater ($P \leq 0.001$) observed SKCS kernel diameter

Table 3.1. Mean values (± SD) for grain volume weight and kernel characteristics of $F_{3:4}$ and $F_{3:5}$ lines of three populations grown at four locations in eastern Colorado.

Parameter	Population	$F_{3:4}$	$F_{3:5}$			
		Fort Collins	Akron	Dailey	Fort Collins	Julesburg
Grain volume weight (kg m^{-3})	CO980376/TX97V2838	749.2 ± 14.3	743.7 ± 15.3 b[†]	724.3 ± 17.0 b	756.1 ± 12.4 b	656.1 ± 24.6 b
	CO99W329/'Overley'	744.9 ± 16.0	750.2 ± 10.8 a	729.1 ± 17.9 b	756.8 ± 9.0 b	669.0 ± 15.2 a
	TX97V2838/CO99W188	754.8 ± 14.9	748.8 ± 15.1 b	736.8 ± 18.7 a	765.0 ± 10.3 a	655.1 ± 25.4 b
	Average	748.8 ± 16.4	747.1 ± 14.2 B[‡]	730.1 ± 18.5 C	756.3 ± 11.3 A	659.8 ± 23.0 D
SKCS Weight (mg kernel^{-1})	CO980376/TX97V2838	33.1 ± 2.6	29.3 ± 1.7 a	27.2 ± 1.7 a	33.1 ± 2.6 a	25.5 ± 1.6 b
	CO99W329/'Overley'	31.3 ± 2.8	28.6 ± 2.0 b	26.9 ± 1.8 a	34.8 ± 1.7 a	26.4 ± 1.7 a
	TX97V2838/CO99W188	31.3 ± 2.5	28.1 ± 2.3 b	26.7 ± 1.9 a	35.0 ± 2.3 a	25.7 ± 1.8 b
	Average	31.9 ± 2.7	28.7 ± 2.1 B	26.9 ± 1.8 C	35.3 ± 2.1 A	26.2 ± 1.7 C
SKCS Diameter (mm)	CO980376/TX97V2838	2.84 ± 0.12	2.55 ± 0.09 a	2.45 ± 0.09 b	2.89 ± 0.09 a	2.42 ± 0.09 b
	CO99W329/'Overley'	2.78 ± 0.13	2.54 ± 0.10 a	2.46 ± 0.09 b	2.87 ± 0.08 a	2.45 ± 0.08 a
	TX97V2838/CO99W188	2.80 ± 0.11	2.55 ± 0.11 a	2.49 ± 0.10 a	2.88 ± 0.10 a	2.43 ± 0.09 b
	Average	2.81 ± 0.13	2.55 ± 0.97 B	2.47 ± 0.09 C	2.88 ± 0.09 A	2.43 ± 0.09 D
SKCS Hardness (index)	CO980376/TX97V2838	70.5 ± 4.8	75.2 ± 3.3 b	76.1 ± 3.6 b	65.9 ± 4.0 b	54.5 ± 3.5 b
	CO99W329/'Overley'	69.8 ± 4.7	75.2 ± 2.9 b	76.8 ± 2.9 b	65.5 ± 3.4 b	54.6 ± 4.5 b
	TX97V2838/CO99W188	75.8 ± 4.7	78.9 ± 3.1 a	81.3 ± 3.0 a	70.3 ± 4.3 a	56.9 ± 3.7 a
	Average	72.1 ± 5.4	76.5 ± 3.5 B	78.1 ± 3.9 A	67.3 ± 4.5 C	55.3 ± 4.1 D

[†] Means within a column followed by the same lower case letter are not significantly different based on LSD test (α=0.05)
[‡] Means within a row followed by the same upper case letter are not significantly different based on LSD test (α=0.05)

values than the other three environments. Additionally, there was no significant difference among the populations in the two environments with the smallest combined means: Dailey (KD=2.47 mm) and Julesburg (KD=2.43 mm).

The mean SKCS hardness index values at Akron (76.5 hardness units) and Dailey (78.1 hardness units) were greater ($P \leq 0.001$) than values observed at Fort Collins (67.3 hardness units) and Julesburg (55.3 hardness units). Differences among populations, within environments were evident in all environments with the TX97V2838/CO99W188 population having greater ($P \leq 0.001$) SKCS hardness index than the CO980376/TX97V2838 and CO99W329/Overley populations. Observed SKCS hardness index means for Julesburg were lower ($P \leq 0.001$) than the other three environments.

Results (Table 3.1) demonstrated overall differences among the environments and among populations within environments. The rank of the average for all parameters, except kernel hardness index, was Fort Collins with the greatest values, followed by Akron, Dailey, and Julesburg with the lowest values. Fort Collins was irrigated and had the largest ($P \leq 0.001$) observed values for all of the parameters, with the exception of SKCS hardness index. Population by environment interactions were observed as performance of the populations varied among locations for most parameters and rankings changed among environments. As an exception, the population rank for SKCS hardness index was consistent across all environments, and the TX97V2838/CO99W188 population had larger values in all environments for SKCS kernel hardness index. This was likely due to the increased kernel hardness index values for parents

used in the formation of this population. The varied results under multiple environments suggest that the data derived from this study encompass a range of values that are consistent with what is observed in a typical hard winter wheat breeding program.

NIR calibration development and validation

Calibration models for grain volume weight, SKCS kernel weight, kernel hardness index, and kernel diameter used in this study were developed by Butler (2010, dissertation Chapter 1). Details and statistics related to the cross-validations and independent validations are presented in Table 3.2. Cross-validation was performed on the samples used in the development of the calibration and validation was performed using the samples from this study. The R^2 variable was used to account for the variability contained in the grain samples that was explained by both the validation and calibration model. In regard to R^2 of the validation set (R^2_{val}), results for validation were consistent with those found when the calibration was developed for grain volume weight (0.798 vs. 0.810), kernel weight (0.703 vs. 0.740), kernel hardness index (0.740 vs. 0.700) and kernel diameter (0.728 vs. 0.730).

For the grain volume weight calibration model, the R^2 of the calibration set (R^2_{cal}) was greater (82.5) than the R^2_{cal} for any of the SKCS parameters. These results are consistent with other documented calibration models. Dowell et al. (2006) reported an R^2_{cal} of 0.74 for NIR predicted grain volume weight with a whole grain winter wheat model using a Foss NIR 6500.

Table 3.2. Summary statistics for calibration development, cross-validation, and independent validation for grain volume weight and SKCS kernel weight diameter and hardness index.

Parameters	Calibration				Independent validation		
	n	SEC	SEC(V)	R^2	n	SEP(C)	R^2
Grain volume weight (kg m^{-3})	632	4.21	5.73	0.825	2691	7.31	0.798
Kernel weight (mg kernel^{-1})	895	2.25	2.64	0.776	2683	3.12	0.703
Kernel diameter (mm)	915	0.114	0.124	0.762	2669	0.161	0.728
Kernel hardness (hardness units)	890	4.89	4.81	0.744	2674	6.10	0.740

†SEC is standard error of calibration, SECV is standard error of cross-validation, and SEP(C) the standard error of prediction.

An R^2_{cal} of 0.80 was reported by Sissons et al. (2006) for a whole grain NIR calibration model for grain volume weight in durum wheat (*Triticum turgidum*). The R^2_{val} was also higher for grain volume weight (79.8) compared to kernel weight (70.3), hardness index (74.0), and diameter (72.8).

The regression plots for NIR predicted values vs. reference values for each trait are shown in Fig. 3.1. The y-intercept for each trait is greater than zero, suggesting that NIR predicted values underestimate the reference values. A slight clustering of two distinct groups is evident for kernel hardness index and diameter, but is very pronounced for grain volume weight. This may be falsely inflating the R^2_{val} values, since a best fit linear regression line drawn through two clusters may not be representative of a linear regression line drawn through each cluster independently. This clustering is the result of low values in the Julesburg environment. Regardless, the trend for grain volume weight is apparent and suggests that whole grain NIR has some predictive ability for grain volume weight.

Figure 3.1. Regression of kernel characteristics and grain volume weight determined by prediction model and by reference method in all populations combined.
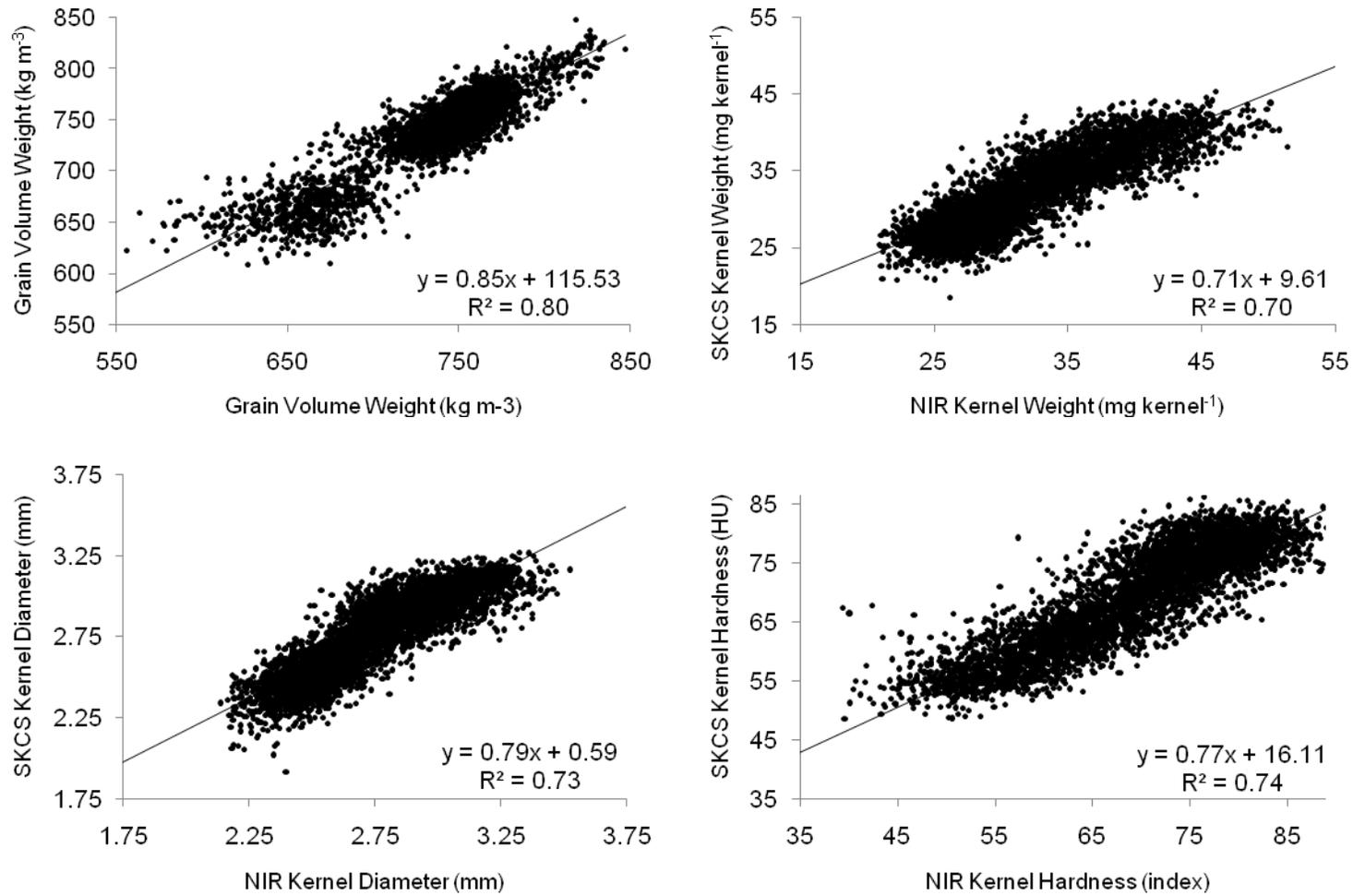
<u>Heritability</u>

Narrow-sense heritability was estimated for each location-population-parameter combination by parent-offspring regression using both reference data and NIR predicted data (Table 3.3). Estimates differed widely among population and environment for each trait, but when averaged across all data, estimates were 49.1% larger when calculated using the reference data when compared to the NIR prediction method. The difference is in some part due to several low estimates when using the NIR method that pulled the average down. Six (12.5%) of the estimates based on NIR predicted values were equal to or less than 0.10, whereas only one (2.1%) of the values based on reference data were less than 0.10. Of the six low NIR heritability estimates, four were from the Julesburg environment. While the estimates from the reference data at Julesburg were not below 0.10, the average heritability estimates from Julesburg across populations were among the bottom half. This is likely due to drought stress at the Julesburg environment which reduced the range of values making among-line comparisons difficult.

Heritability estimates for grain volume weight are in general agreement with those reported by Barnard et al. (2002, $h^2 = 0.31$, narrow-sense, based on variance components) with average $h^2_{NIR}=0.25$ and average $h^2_{Ref}=0.33$. Average estimates of heritability for kernel weight based on reference values (0.44) and NIR values (0.21) flanked narrow-sense heritability estimates reported by Dere and Yildirim (2006) (h^2 of 0.31), but were lower than those reported by Wiersma et al. (2001, h^2 of 0.59) and Barnard et al. (2002, h^2 of 0.71) for both methods

Realized heritability was estimated separately for each parameter by population in each environment (Table 3.4). Estimates were calculated for both selection with the reference method and selection using the NIR predicted values with a selection intensity of 0.30. The selection intensity of 0.30 indicates that the top 30% of lines were selected in each population within environments.

Grain volume weight

Realized heritability estimates for grain volume weight with the reference method was greater than for the NIR method in all cases (Table 3.4), with the exception of the CO980376/TX97V2838 population in Fort Collins (h^2_r NIR=0.64 vs. h^2_r reference=0.55) and TX97V2838/CO99W188 in Julesburg (h^2_r NIR of 0.77 vs. h^2_r reference of 0.42). Overall, realized heritability estimates were intermediate-high (0.45 to 0.77) in all populations in all locations, with the exception of the CO99W329/Overley population in Akron (h^2_r reference of 0.20) and Dailey (reference of 0.17). The heritability results using the reference method are consistent with broad-sense heritability estimates based on variance components reported by Wiersma et al. (2001) in spring wheat (mean H^2=0.55, following 8 cycles of recurrent selection), but higher than values reported by Barnard et al. (2002; h^2 of 0.31, narrow-sense, based on variance components). Estimates of h^2_r NIR when compared to h^2_r Ref are not consistent with the high R^2_{val} values for the NIR grain volume weight model, but are low to intermediate in general in all locations except Dailey.

Table 3.3. Narrow-sense heritability estimates (± standard error) calculated with parent-offspring correlation using NIR (h^2_rNIR) and reference (h^2_rRef) data for grain volume weight and SKCS kernel weight, diameter, and hardness index.

Trait	Population	Akron		Dailey		Fort Collins		Julesburg	
		h^2NIR	h^2Ref	h^2NIR	h^2Ref	h^2NIR	h^2Ref	h^2NIR	h^2Ref
Grain Volume Weight	CO980376/TX97V2838	0.28 ±0.03	0.45 ±0.05	0.30 ±0.03	0.32 ±0.04	0.59 ±0.06	0.52 ±0.05	0.32 ±0.04	0.42 ±0.08
	CO99W329/'Overley'	0.15 ±0.02	0.08 ±0.01	0.09 ±0.01	0.11 ±0.02	0.18 ±0.01	0.27 ±0.01	0.09 ±0.01	0.26 ±0.03
	TX97V2838/CO99W188	0.21 ±0.02	0.42 ±0.05	0.31 ±0.01	0.44 ±0.06	0.14 ±0.02	0.42 ±0.03	0.34 ±0.04	0.29 ±0.06
SKCS Weight	CO980376/TX97V2838	0.10 ±0.01	0.47 ±0.08	0.38 ±0.06	0.42 ±0.07	0.27 ±0.03	0.44 ±0.09	0.07 ±0.01	0.50 ±0.08
	CO99W329/'Overley'	0.26 ±0.04	0.49 ±0.09	0.19 ±0.03	0.40 ±0.07	0.43 ±0.05	0.36 ±0.06	0.13 ±0.02	0.29 ±0.05
	TX97V2838/CO99W188	0.17 ±0.03	0.53 ±0.12	0.13 ±0.02	0.51 ±0.10	0.21 ±0.02	0.49 ±0.11	0.17 ±0.02	0.37 ±0.07
SKCS Diameter	CO980376/TX97V2838	0.15 ±0.01	0.49 ±0.01	0.34 ±0.01	0.43 ±0.01	0.13 ±0.01	0.34 ±0.01	0.18 ±0.01	0.54 ±0.01
	CO99W329/'Overley'	0.32 ±0.01	0.36 ±0.01	0.27 ±0.01	0.24 ±0.01	0.39 ±0.02	0.22 ±0.02	0.07 ±0.01	0.19 ±0.01
	TX97V2838/CO99W188	0.29 ±0.01	0.42 ±0.01	0.32 ±0.01	0.44 ±0.01	0.28 ±0.01	0.28 ±0.01	0.10 ±0.01	0.28 ±0.01
SKCS Hardness	CO980376/TX97V2838	0.19 ±0.04	0.66 ±0.21	0.41 ±0.08	0.51 ±0.17	0.36 ±0.11	0.59 ±0.22	0.43 ±0.07	0.55 ±0.19
	CO99W329/'Overley'	0.40 ±0.08	0.43 ±0.12	0.20 ±0.03	0.39 ±0.11	0.49 ±0.13	0.42 ±0.14	0.51 ±0.11	0.48 ±0.20
	TX97V2838/CO99W188	0.40 ±0.09	0.54 ±0.16	0.27 ±0.05	0.34 ±0.10	0.39 ±0.14	0.53 ±0.22	0.46 ±0.09	0.42 ±0.15

Table 3.4. Realized heritability estimates for selection using NIR (h^2_r NIR) and reference (h^2_r Ref) data for grain volume weight and SKCS kernel weight, diameter, and hardness index.

Trait	Population	Akron		Dailey		Fort Collins		Julesburg	
		h^2_rNIR	h^2_rRef	h^2_rNIR	h^2_rRef	h^2_rNIR	h^2_rRef	h^2_rNIR	h^2_rRef
Grain volume weight	CO980376/TX97V2838	0.32	0.50	0.09	0.46	0.53	0.61	0.22	0.45
	CO99W329/'Overley'	0.16	0.20	0.09	0.17	0.64	0.55	0.14	0.42
	TX97V2838/CO99W188	0.33	0.77	0.16	0.69	0.36	0.53	0.77	0.42
SKCS Weight	CO980376/TX97V2838	0.42	0.48	0.44	0.50	0.18	0.53	0.45	0.54
	CO99W329/'Overley'	0.65	0.58	0.51	0.44	0.63	0.42	0.34	0.24
	TX97V2838/CO99W188	0.43	0.47	0.37	0.53	0.69	0.52	0.05	0.23
SKCS Diameter	CO980376/TX97V2838	0.45	0.31	0.53	0.44	0.51	0.42	0.82	0.18
	CO99W329/'Overley'	0.42	0.41	0.36	0.16	0.40	0.26	0.02	0.20
	TX97V2838/CO99W188	0.86	0.36	0.15	0.42	0.91	0.31	0.04	0.63
SKCS Hardness	CO980376/TX97V2838	0.75	0.74	0.41	0.46	0.60	0.63	0.70	0.66
	CO99W329/'Overley'	0.37	0.35	0.57	0.30	0.50	0.46	0.70	0.32
	TX97V2838/CO99W188	0.13	0.47	0.09	0.29	0.50	0.50	0.71	0.48

Kernel Weight

Estimates of h^2_r NIR for kernel weight differ from those calculated from the reference data (Table 3.3). In several comparisons between h^2_r NIR and h^2_r Ref, the NIR-based analysis resulted in larger values. Overall, h^2_r Ref estimates are intermediate-high (0.40 to 0.60) for all populations in all environments, with the exception of Julesburg where the CO99W329/'Overley' and TX97V2838/ CO99W188 populations had low-intermediate h^2_r Ref estimates. This is again likely due to drought stress which commonly compresses distributions and reduces the range of parameter values. Estimates of realized heritability based on reference values for kernel weight in this study were consistent with narrow-sense heritability estimates reported by Dere and Yildirim (2006) (h^2 of 0.31), but lower than those reported by Wiersma et al. (2001, h^2 of 0.59) and Barnard et al. (2002, h^2 of 0.71).

Kernel Hardness Index

Heritability estimates for SKCS hardness index varied among populations within environments (Table 3.4), but were intermediate-high (0.29 to 0.74) when using the reference method for selection. Estimates of h^2_r derived from selection with NIR predictions were also intermediate-high (0.37 to 0.75), with the exception of the TX97V2838/CO99W188 populations at Akron (h^2_r Ref of 0.13) and Dailey (h^2_r Ref of 0.09). Overall, estimates using the reference method and the NIR-based method were more consistent for kernel hardness index than with any other parameter. This is likely due to the parents of all populations having

relatively large SKCS hardness index values that differ less between parents than the other parameters.

Kernel Diameter

When comparing the h^2_r estimates calculated for the reference method and the NIR-based analysis, the estimates for kernel diameter were the most variable (Table 3.4). The TX97V2838/CO99W188 population at Akron (h^2_r NIR of 0.86 vs. h^2_r Ref of 0.31) and Fort Collins (h^2_r NIR of 0.91 vs. h^2_r Ref of 0.31) had estimates that were much higher for the NIR selection method, but at Dailey and Julesburg the opposite was true. Overall, estimates of h^2_r Ref were low-intermediate (0.16 to 0.44) with the exception of TX97V2838/CO99W188 in Julesburg (h^2_r Ref of 0.63).

Overall, the greatest amount of consistency between the heritability estimates based on SKCS reference values and NIR-based analysis were observed in Fort Collins. Fort Collins was the only irrigated environment and had the largest mean grain volume weight, SKCS weight and diameter. The availability of moisture would allow lines to reach their full genetic potential and would allow for maximum observable differences among lines.

Based on the estimates for all environments, results indicate genetic gains can be obtained through selection with NIR predicted parameters. While in general for NIR-based analysis heritability estimates were lower than estimates calculated by the reference method, the NIR-based analysis suggest than genetic gains could be made using NIR predicted values. The speed and

efficiency of the NIR-based analysis may provide more benefits than risks in early-generation selection when the number of samples is great and time between harvest and planting is limited.

<u>Classification Errors</u>

Samples from the three populations were combined and based on parameter values obtained from reference testing and NIR-based analysis, all $F_{3:4}$ lines were ranked in descending order. Based on these ranks, samples were assigned into three separate classification groups (TOP, MID, BOTTOM). The TOP class included samples ranked in the top 1/3 of all samples, the MID class samples ranked in the middle 1/3, and BOTTOM class samples ranked in the bottom 1/3. Samples were classified independently using data obtained from the reference analysis and NIR-based analysis. Results were entered into a jack-knife classification table (Table 3.5) to determine the percentage of samples that were incorrectly classified when using the NIR-based analysis.

The results indicate that when using NIR-based analysis, the percentage of correct classification when selecting for the best one-third (TOP) of samples ranges from 45% to 64%. Similar results are observed when making NIR-based selection for the bottom one-third (BOTTOM) of samples with a range of correct classification of 51% to 67%. In the context of breeding, when selecting the best samples for advancement, or the worst for elimination, approximately one-half to one-third of the selected samples would actually fall into the desired categories. At first glance, this would suggest that the NIR-based analysis would be

Table 3.5. Jackknife classification table for all parameters. Columns represent classification by reference method and rows represent classification by NIR prediction.

		Reference classification rate (%)			
		Top	Middle	Bottom	Correct
NIR Grain volume weight	Top (n=95)	56	33	12	56
	Middle (n=95)	27	43	29	43
	Bottom (n=95)	17	24	59	59
				Average	53
NIR Kernel Weight	Top (n=100)	52	32	16	52
	Middle (n=100)	26	43	31	43
	Bottom (n=100)	22	25	53	53
				Average	49
NIR Kernel Diameter	Top (n=100)	45	39	16	45
	Middle (n=100)	33	34	33	34
	Bottom (n=100)	22	27	51	51
				Average	43
NIR Kernel Hardness	Top (n=95)	64	31	5	64
	Middle (n=95)	28	44	27	44
	Bottom (n=95)	7	25	67	67
				Average	58

ineffective in a breeding program. A closer look at the degree of misclassification may suggest otherwise. In all parameters, when selecting the top one-third of samples based on NIR predicted values, between 84 and 94% fall into the top two-thirds (TOP and MID) of all samples, and only between 5 and 16% would be in the bottom one-third (BOTTOM). In the context of breeding, between 5 and 16% of samples selected for advancement, based on NIR predicted values, would be inferior lines. A similar result is observed when selecting the bottom one-third (BOTTOM) of sample for elimination where between 7% and 22% of the lines would actually be superior lines.

Conclusions

The objective of this study was to determine the selection efficiency and accuracy of selection based on realized heritability and classification errors when using NIR spectroscopy as a selection tool by comparison to traditional methods. Both genetic and environmental effects on kernel characteristics were observed under several environments, thereby reinforcing the approach of collecting samples from multiple environments when developing a NIR calibration model for kernel characteristics. Heritability estimates suggest that genetic gains can be made when using NIR predicted values for selection. Analysis of classification errors indicate that while classification errors do occur when using NIR predicted values, they are in general not extreme misclassifications. Results suggest that a small percentage of desirable lines (7 to 22%) would be discarded when using NIR spectroscopy and calibration models for discarding inferior lines in early-generation material. Severe misclassifications are slightly reduced when selecting superior lines where 5 to 16% of the advanced lines would be inferior. While not yet perfected, NIR prediction models for grain volume weight, SKCS kernel weight, diameter, and hardness index may be useful in a breeding program for early-generation selection. The degree of selection errors using this approach are relatively low, but it would be up to the individual breeders to decide if this is acceptable and if the benefits of using NIR-based analysis justify the associated selection error.

LITERATURE CITED

AACC International. 2000. Approved Methods of the American Association of Cereal Chemists, 10th Ed. The Association: St. Paul, MN.

Barnard, A.D., M.T. Labuschagne, and H.A. van Niekerk. 2002. Heritability estimates of bread wheat quality traits in the Western Cape province of South Africa. Euphytica 127: 115-122.

Baker, R.J. 1981. Inheritance of seed coat color in 8 spring wheat cultivars. Can. J. Plant Sci. 61: 719-721.

Blazek, J., O. Jirsa, and M. Hruskova. 2005. Prediction of wheat milling characteristics by near-infrared reflectance spectroscopy. Czech J. Food Sci. 23: 145-151.

Butler, J.D., S.D. Haley, and B.W. Seabourn. 2005. Efficiency of selection for wheat kernel characteristics using near-infrared reflectance spectroscopy. American Society of Agronomy Annual Meeting, Salt Lake City, UT.

Burns, D.A., and E.W. Ciurczak (eds.) 2001. Handbook of near-infrared analysis. Marcel Dekker, New York, New York.

Delwiche, S.R. 1998. Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. J. Cereal Sci. 27: 241-254.

Delwiche, S.R., and G. Weaver. 1994. Bread quality of wheat-flour by near-infrared spectrophotometry - feasibility of modeling. J. Food Sci. 59: 410-415.

Delwiche, S.R., and W.R. Hruschka. 2000. Protein content of bulk wheat from near-infrared reflectance of individual kernels. Cereal Chem. 77: 86-88.

Delwiche, S.R., and R.A. Graybosch. 2002. Identification of waxy wheat by near-infrared reflectance spectroscopy. J. Cereal Sci. 35: 29-38.

Delwiche, S.R., and G.A. Hareland. 2004. Detection of scab-damaged hard red spring wheat kernels by near-infrared reflectance. Cereal Chem. 81: 643-649.

Delwiche, S.R., R.E. Pitt, and K.H. Norris. 1991. Examination of starch-water and cellulose-water interactions with near-infrared (NIR) diffuse reflectance spectroscopy. Starch-Starke 43: 415-422.

Delwiche, S.R., R.A. Graybosch, and C.J. Peterson. 1999. Identification of wheat lines possessing the 1AL.1RS or 1BL.1RS wheat-rye translocation by near-infrared reflectance spectroscopy. Cereal Chem. 76: 255-260.

Delwiche, S.R., R.A. Graybosch, and C.J. Peterson. 1998a. Predicting protein composition, biochemical properties, and dough-handling properties of hard red winter wheat flour by near-infrared reflectance. Cereal Chem. 75: 412-416.

Delwiche, S.R., R.O. Pierce, O.K. Chung, and B.W. Seabourn. 1998b. Protein content of wheat by near-infrared spectroscopy of whole grain: Collaborative study. J. AOAC Int. 81: 587- 603.

Dere, S, and M.B. Yildirim. 2006. Inheritance of plant height, tiller number per plant, spike height and 1000-kernel weight in a 8x8 diallel cross population of bread wheat. Cereal Res. Commun. 34: 965-972.

Dowell, F.E. 1998. Automated color classification of single wheat kernels using visible and near-infrared reflectance. Cereal Chem. 75: 142-144.

Dowell, F.E., E.B. Maghirang, F. Xie, G.L. Lookhart, R.O. Pierce, B.W. Seabourn, S.R. Bean, J.D. Wilson, and O.K. Chung. 2006. Predicting wheat quality characteristics and functionality using near-infrared spectroscopy. Cereal Chem. 83: 529-536.

Dowell, F.E., E.B. Maghirang, R.O. Pierce, G.L. Lookhart, S.R. Bean, F. Xie, M.S. Caley, J.D. Wilson, B.W. Seabourn, M.S. Ram, S.H. Park, and O.K. Chung. 2008. Relationship of bread quality to kernel, flour, and dough properties. Cereal Chem. 85: 82-91.

Evers, T., and S. Millar. 2002. Cereal grain structure and development: Some implications for quality. Journal Cereal Sci. 36: 261-284.

Falconer, D. S. 1981. Introduction to quantitative genetics. Longman, New York.

Fehr, W. R. 1993. Principles of cultivar development. Vol. 1. Macmillan Publishing Company, New York.

Finney, P.L., C.S. Gaines, and L.C. Andrews. 1987. Wheat quality – a quality assessor's view. Cereal Food World 32: 313.

Fowler, D.B., and M.I.P. Kovacs. 2004. Influence of protein concentration on Farinograph absorption, mixing requirements and mixing tolerance. Can. J. Plant Sci. 84: 765-772.

Frey, K.J., and Horner, T. 1957. Heritability in standard units. Agron. J. 49:59–62.

Gergely, S., and A. Salgo. 2005. Changes in carbohydrate content during wheat maturation - what is measured by near-infrared spectroscopy? J. Near Infrared Spec. 13: 9-17.

Graybosch, R., C.J. Peterson, K.J. Moore, M. Stearns, and D.L. Grant. 1993. Comparative effects of wheat-flour protein, lipid, and pentosan composition in relation to baking and milling quality. Cereal Chem. 70: 95-101.

Hruskova, M., M. Bednarova, and P. Smejda. 2004. Prediction of rheological parameters of dough by NIR spectral analysis of wheat flour. Chem. Listy 98: 423-431.

Isaksson, T., and T. Naes. 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in the NIR spectroscopy. Appl. Spectrosc. 42: 1273-1284.

Juhasz, R., S. Gergely, T. Gelencser, and A. Salgo. 2005. Relationship between NIR spectra and RVA parameters during wheat germination. Cereal Chem. 82: 488-493.

Lyford, C.P., W. Kidd, P. Rayas-Duarte, and C. Deyoe. 2005. Prediction of flour extraction rate in hard red winter wheat using the single kernel characterization. J. Food Quality 28: 279-288.

Maghirang, E.B, G.L. Lookhart, S.R. Bean, R.O. Pierce, F. Xie, M.S. Caley, J.D. Wilson, B.W. Seabourn, M.S. Ram, S.H. Park, O.K. Chung, and F.E. Dowell. 2006. Comparison of quality characteristics and breadmaking functionality of hard red winter and hard red spring wheat. Cereal Chem. 83: 520-528.

Manley, M., L. Van Zyl, and B.G. Osborne. 2002. Using Fourier transform near-infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour. J. Near Infrared Spec. 10: 71-76.

Martens, H., and T. Naes. 1984. Multivariate calibration .1. concepts and distinctions. Trac-Trend Anal. Chem. 3: 204-210.

McCaig, T.N., R.M. Depauw, and P.C. Williams. 1993. Assessing Seed-Coat Color In A Wheat Breeding Program With A NIR Vis Instrument. Can. J. Plant Sci. 73: 535-539.

Metzger, R.J., and B.A. Silbaugh. 1970. Location of genes for seed coat color in hexiploid wheat, Triticum aestivum L. Crop Sci. 10: 495-497.

Morgan, J.E., and P.C. Williams. 1995. Starch damage in wheat flours – a comparison of enzymatic, iodometric, and near-infrared reflectance techniques. Cereal Chem. 72: 209-212.

Morris, C.F., K.G. Campbell, and G.E. King. 2005. Kernel texture differences among US soft wheat cultivars. J. Sci. Food Agr. 85: 1959-1965.

Motzo, R., F. Giunta, and M. Deidda. 2004. Expression of a tiller inhibitor gene in the progenies of interspecific crosses Triticum aestivum L. x T. turgidum subsp durum. Field Crops Res. 85: 15-20.

Nielsen, J.P., D. Bertrand, E. Micklander, P. Courcoux, and L. Munck. 2001. Study of NIR spectra, particle size distributions and chemical parameters of wheat flours: a multi-way approach. J. Near Infrared Spec. 9: 275-285.

Ohm, J.B, O.K. Chung, and C.W. Deyoe. 1998. Single-kernel characteristics of hard winter wheats in relation to milling and baking quality. Cereal Chem. 75: 156-161.

Osborne, B.G. 1984. Investigations into the use of near-infrared reflectance spectroscopy for the quality assessment of wheat with respect to its potential for bread baking. J. Sci. Food Agr. 35: 106-110.

Pawlinsky, T., and P. Williams. 1998. Prediction of bread-baking functionality in whole kernels, using near-infrared reflectance spectroscopy. J. Near Infrared Spec. 6: 121-127.

Perez-Mendoza, J., J.E. Throne, E.B. Maghirang, F.E. Dowell, and J.E. Baker. 2005. Insect fragments in flour: Relationship to lesser grain borer (Coleoptera: Bostrichidae) infestation level in wheat and rapid detection using near-infrared spectroscopy. J. Econ. Entomol. 98: 2282-2291.

Peterson, C.J., R.A. Graybosch, D.R. Shelton, and P.S. Baenziger. 1998. Baking quality of hard winter wheat: Response of cultivars to environment in the great plains (Reprinted from Wheat: Prospects for global improvement, 1998). Euphytica 100: 157-162.

Preston, K.R., B.C. Morgan, and J.E. Dexter. 1995. Influence of protein segregation on the quality characteristics of Biggar and Genesis Canada Prairie spring wheat. Can. J. Plant Sci. 75: 599-604.

SAS Institute Inc., 2004. SAS 9.1.3 Help and Documentation, SAS Institute Inc., Cary, NC.

Shapiro, S.S., M.B. Wilk, and H.J. Chen. 1968. A comparative study of various tests for normality. J. Am. Stat. Assoc. 63:1343-1372.

Sissons, M., B. Osborne, and S. Sissons. 2006. Application of near infrared reflectance spectroscopy to a durum wheat breeding programme. J. Near Infrared Spec. 14: 17-25.

Smith, J.D., and M.L. Kinman. 1965. The use of parent-offspring regression as an estimator of heritability. Crop Sci 5: 595-596.

Stenberg, B., A. Jonsson, and T. Borjesson. 2005. Use of near-infrared reflectance spectroscopy to predict nitrogen uptake by winter wheat within fields with high variability in organic matter. Plant Soil 269: 251-258.

Stuart, B.H. 2004. Infrared spectroscopy: fundamentals and application. John Wiley and sons. West Sussex, England.

Veha, A. 2007. Correlation between the kernel structure and the quality parameters on some Hungarian winter wheat varieties. Cereal Res. Commun. 35: 1289-1292.

Wang, D., F.E. Dowell, and D.S. Chung. 2001. Assessment of heat-damaged wheat kernels using near-infrared spectroscopy. Cereal Chem. 78: 625-628.

Weegels, P.L., R.J. Hamer, and J.D. Schofield. 1996. Functional properties of wheat glutenin. J. Cereal Sci. 23: 1-18.

Wesley, I.J., S. Uthayakumaran, R.S. Anderssen, G.B. Cornish, F. Bekes, B.G. Osborne, and J.H. Skerritt. 1999. A curve-fitting approach to the near-infrared reflectance measurement of wheat flour proteins which influence dough quality. J. Near Infrared Spec. 7: 229-240.

Wiersma, J.J., R.H. Busch, G.G. Fulcher, and G.A. Hareland. 2001. Recurrent selection for kernel weight in spring wheat. Crop Sci. 41: 999-1005.

Williams, P. 2003. Near-infrared Technology - Getting the Best out of Light. 2.0.2 ed. PDK Projects, Inc., Nanaimo, British Columbia, Canada.

Williams, P., and K. Norris, (eds.) 2001. Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, St. Paul, MN.

Williams, P.C., F.J. Elharamein, G. Ortizfereira, and J.P. Srivastava. 1988. Preliminary-observations on the determination of wheat strength by near-infrared reflectance. Cereal Chem. 65: 109-114.
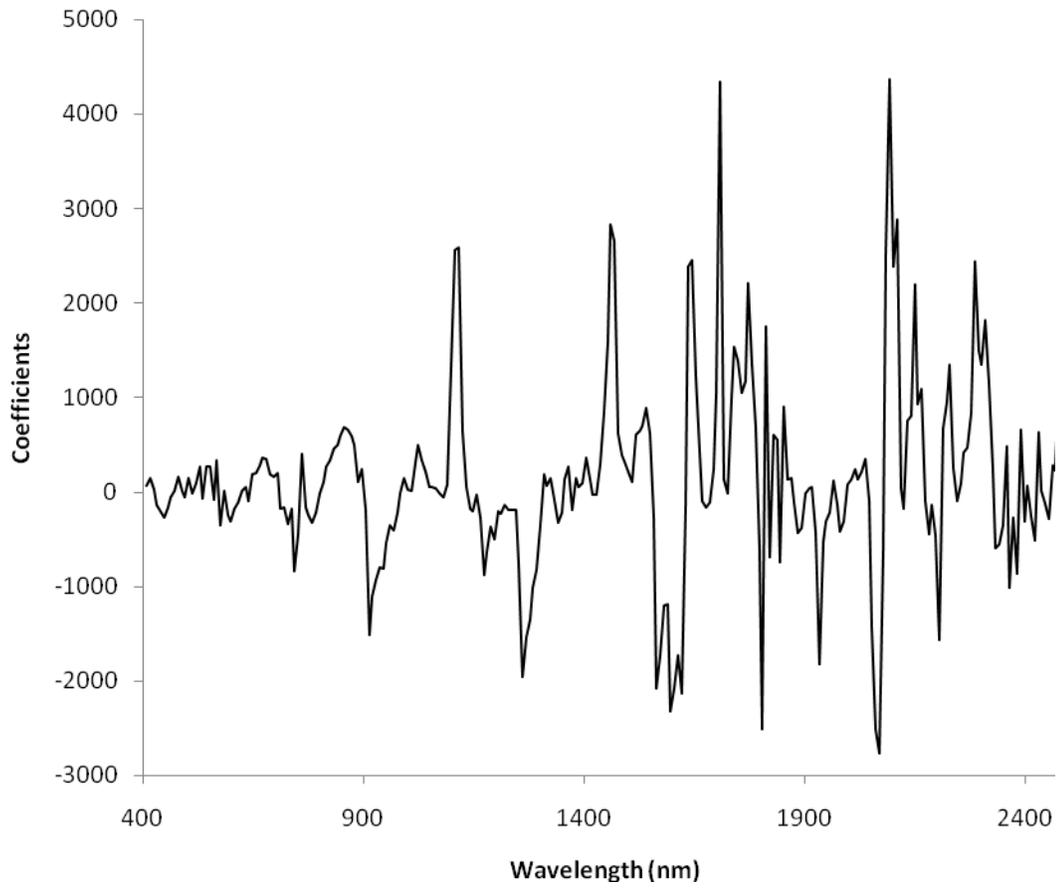
APPENDICES

Appendix 1. Calibration loadings for kernel diameter.
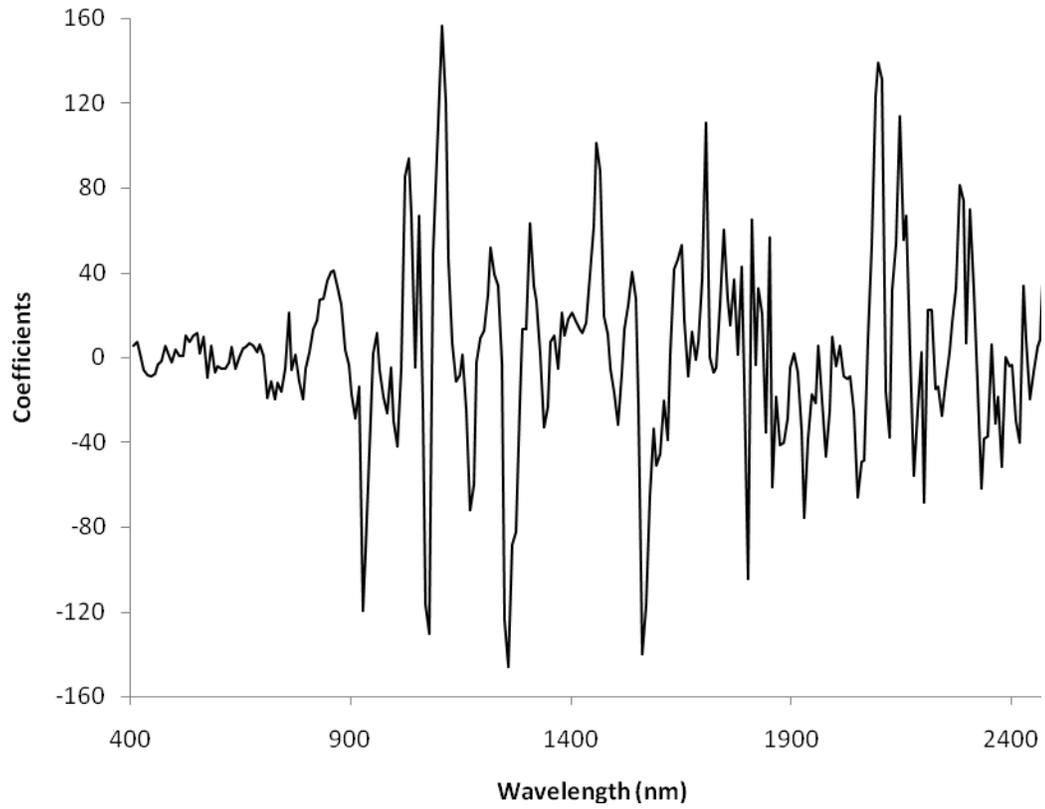
Appendix 2. Calibration loadings for kernel hardness.
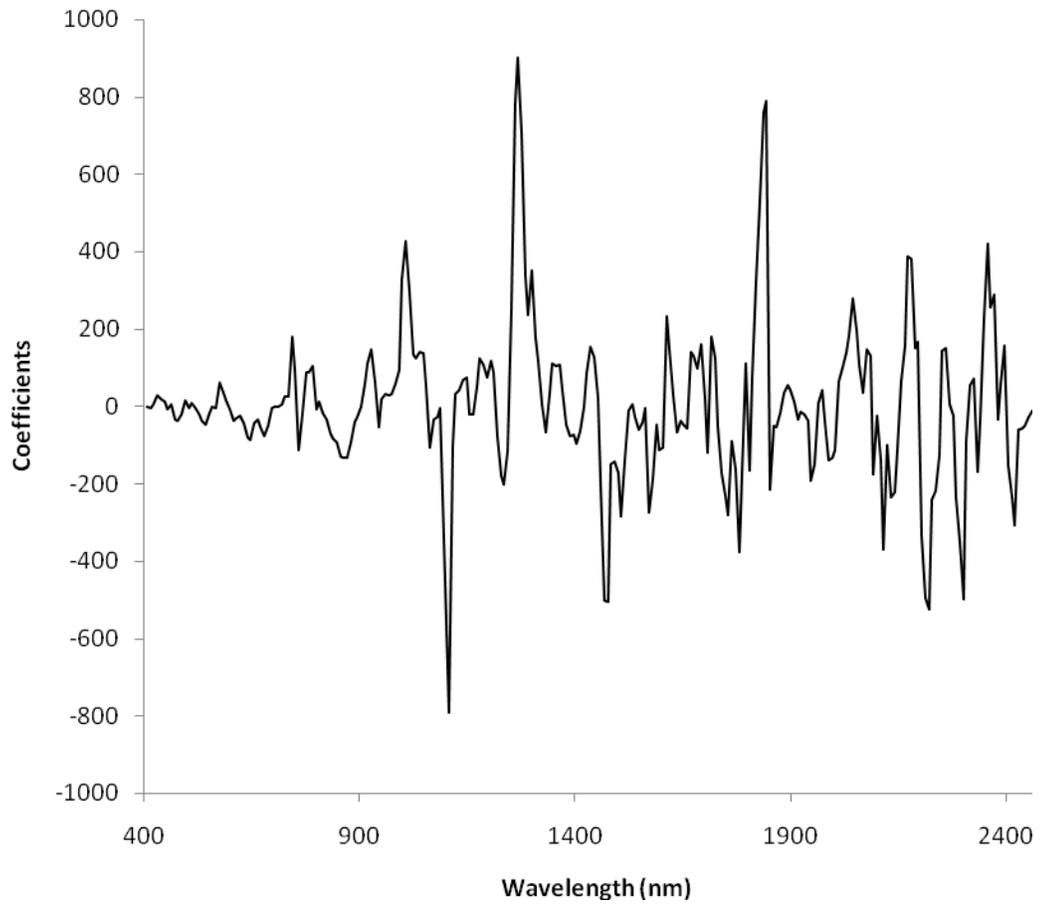
Appendix 3. Calibration loadings for kernel weight.

Appendix 4. Calibration loadings for grain volume weight.