December 2007

Use of Near-Infrared Spectroscopy for Qualitative and Quantitative Analyses of Grains and Cereal Products

Panjama Cheewapramong
UNL, pcheewa1@bigred.unl.edu

USE OF NEAR-INFRARED SPECTROSCOPY FOR QUALITATIVE AND QUANTITATIVE ANALYSES OF GRAINS AND CEREAL PRODUCTS

by

Panjama Cheewapramong

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Doctor of Philosophy

Major: Food Science and Technology

Under the Supervision of Professor Randy L. Wehling

Lincoln, Nebraska

December, 2007

USE OF NEAR-INFRARED SPECTROSCOPY FOR QUALITATIVE AND QUANTITATIVE ANALYSES OF GRAINS AND CEREAL PRODUCTS

Panjama Cheewapramong, Ph.D.
University of Nebraska, 2007

Advisor: Randy L. Wehling

The purpose of the first part of the study was to develop a simplified near-infrared reflectance (NIR) spectroscopy method for detecting insect larvae in individual wheat kernels. Discriminant analysis, based on Mahalanobis distances calculated from log 1/R data at only four discrete wavelengths, yielded better results for classification of sound and insect infested wheat kernels than principal component analysis (PCA) using the spectral region from 1100 to 1900 nm. This simplified technique was then used to detect 3- and 4-week-old larvae of granary and maize weevils in wheat kernels. A model developed from a calibration set containing sound kernels and kernels infested with 3-week-old larvae was applied to a validation set containing sound kernels, sound air-dried kernels, kernels containing 3-week-old larvae of granary and maize weevils, kernels containing 4-week-old larvae of granary and maize weevils, and infested air-dried kernels containing dead larvae of both species. Correct classification rates of 92, 98, 77, 73, 95, 98, 96, and 94%, respectively, were achieved. Additionally, 99% of sound kernels from ten different wheat varieties were correctly classified into their respective classes. First and second derivative spectral treatments did not improve classification results for 3-week-old infested kernels.

NIR spectroscopy was also used to predict the degree of cook in products produced by HTST extrusion of corn meal. Corn meal was cooked with a Wenger TX-57 twin screw extruder using screw speeds ranging from 250 to 350 rpm, and moisture contents ranging from 13-20%, providing a wide range of pressures and shear conditions in the extruder barrel. Extruded samples were analyzed using reference methods that measure different aspects of cooking, including water absorption index (WAI), water

solubility index (WSI), viscosity profile as measured with a Rapid Viscoanalyzer (RVA), hardness and fracturability as measured by Texture Profile Analysis. Calibrations for each parameter were developed using multiple linear regression (MLR) and partial least squares (PLS) regression. Correlations with r-value>0.95 were achieved between the NIR and laboratory values. Relative predictive determinant (RPD) values ranged from 5.3 to 6.3 for the various parameters (except for hardness, and trough viscosity) indicating that the NIR measurements should be useful in quality control applications.

ACKNOWLEDGMENT

First of all, I would like to express my gratitude to my advisor, Dr. Randy L. Wehling, for his support, guidance, patience, and encouragement throughout my doctoral program for many years. I would also like to thank my committee members; Dr. Susan L. Cuppett, Dr. Curtis L. Weller, Dr. James D. Carr, and Dr. David L. Keith, for their contribution towards this dissertation.

I gratefully acknowledge the Department of Entomology at Kansas State University for providing parental insects. My sincere appreciation is extended to Mary Shipman of the Agronomy Department (Cereal Quality Lab), for providing wheat and allowing me to use her lab facilities. I am also thankful to the Department of Entomology for technical assistance and advice in growing insects. I would like express my sincere appreciation to Laurie Keeler and Steve Weier of the Food Processing Center, Food Science and Technology Department for technical assistance in extrusion process. I would also like to thank Choo Lum Ng, my lab mate, for his contributing and unforgettable friendship.

Finally, I would like to make a special acknowedgment to my loving parents, Vachara and Sirima Cheewapramong, and my younger brother, Buntoon Cheewapramong, They have been waiting for too many years. And also, my cousin, Siriporn Workman, my lovely nephew, Boston Workman, my best friend, Yachai Amornkul, and my boyfriend, Rujekorn Wittaya. This dissertation is dedicated to them for their endless love, all support, emotional, and encouragement to achieve my dream.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS.....i

TABLE OF CONTENTS.....ii

LIST OF TABLES.....v

LIST OF FIGURES.....x

OVERVIEW.....xii

LITERATURE REVIEW.....1

Near-Infrared Spectroscopy.....1

Application of NIR Spectroscopy in Qualitative Analysis.....2

Discriminant Analysis.....4

Mahalanobis Distances.....5

Principal Component Analysis Coupled with Mahalanobis Distances.....6

Application of NIR Spectroscopy in Quantitative Analysis.....8

Chemometric Methods in Quantitative Analysis.....11

Problems of Insect Infestation in Stored Grain.....14

Detection of Insect Infestation.....15

Use of NIR Spectroscopy for Detection of Insect Infestation.....18

Extrusion Cooking.....23

Analysis of Extruded Products.....24

Use of NIR Spectroscopy for Predicting Degree of Cook.....26

References.....28

RESEARCH TOPICS.....42

Chapter 1 Comparison of Selected Discrete Wavelength and Principal Component Analysis Methods for Insect Infestation Detection.....42

Abstract.....42

Introduction.....44

Materials and Methods.....44

Results and Discussion.....51

References.....69

Chapter 2 A Simplified Near-Infrared Method for Detecting Internal Insect Infestation in Wheat Kernels.....71

Abstract.....71

Introduction.....73

Materials and Methods.....74

Results and Discussion.....83

References.....103

Chapter 3 Predicting Degree of Extrusion-Cooking of Corn meal by Near-Infrared Reflectance Spectroscopy.....106

Abstract.....106

Introduction.....108

Materials and Methods.....109

Results and Discussion.....117

References.....145

CONCLUSION..149

LIST OF TABLES

Page

Chapter 1

Table 1 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra (log 1/R) from 1100 to 2498 nm using different pre-treatments 52

Table 2 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 2498 nm using derivative treatment 54

Table 3 Percentage of correct classification of wheat samples based on principal component analysis of partial NIR spectra (log 1/R, 1100-1900 nm and 1980-2498 nm) using different pre-treatments 56

Table 4 Percentage of correct classification of wheat samples based on principal component analysis of partial NIR spectra (1100-1900 nm and 1980-2498 nm) using derivative treatment 57

Table 5 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra (log 1/R) from 1100 to 1900 nm using different pre-treatments 59

Table 6 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using derivative treatment 60

Table 7 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Savitsky-Golay (SG) first derivative with different pre-treatments 62

Table 8 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Savitsky-Golay (SG) second derivative with different pre-treatments 63

Table 9 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Gap first derivative with different pre-treatments 64

Table 10 Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Gap second derivative with different pre-treatments 66

Table 11 Percentage of correct classification of wheat samples based on Mahalanobis distances of selected wavelengths of NIR spectra from 1100 To 1900 nm 68

Chapter 2

Table 1 Searching approaches for constructing calibrations using raw data (log1/R) sound and infested kernels with 4-week-old larvae performed with MultiQual software 81

Table 2 Several searching approaches for constructing calibration models using raw data (log1/R) of sound and infested kernels with 3-week-old larvae performed with MultiQual software 82

Table 3 Correct classification of wheat samples using a calibration developed from raw spectra of sound and 4-week-old larvae with different searching approaches 86

Table 4 Classification matrices from several searching approaches for constructing calibration models using raw spectra of sound and infested kernels with 4-week-old larvae 87

Table 5 Correct classification of wheat samples (calibration samples) by models using first derivative spectra of sound and 4-week-old larvae developed with different searching approaches 89

Table 6 Correct classification of wheat samples (validation samples) by models using first derivative spectra of sound and 4-week-old larvae developed with different searching approaches 90

Table 7 Correct classification of wheat samples (calibration samples) by models using second derivative spectra of sound and 4-week-old larvae developed with different searching approaches 91

Table 8 Correct classification of wheat samples (validation samples) by models using second derivative spectra of sound and 4-week-old larvae developed with different searching approaches 92

Table 9 Correct classification of wheat samples using calibrations developed from raw spectra of sound and 3-week-old larvae developed with different searching approaches 94

Table 10 Classification matrices from several searching approaches for constructing calibration models using raw spectra of sound and infested kernels with 3-week-old larvae 95

Table 11 Correct classification of wheat samples (calibration samples) by models using derivative spectra of sound and 3-week-old larvae developed with different searching approaches 96

Table 12 Correct classification of wheat samples (validation samples) by models using first derivative spectra of sound and 3-week-old larvae developed with different searching approaches 97

Table 13 Correct classification of wheat samples (calibration samples) by models using second derivative spectra of sound and 3-week-old larvae developed with different searching approaches 98

Table 14 Correct classification of wheat samples (validation samples) by models using second derivative spectra of sound and 3-week-old larvae developed with different searching approaches 99

Table 15 Correct classification of wheat samples using models developed with a modified calibration set. Models were developed from raw spectra of sound and 3-week-old larvae using different searching approaches 102

Chapter 3

Table 1 Coded samples of extrudates 111

Table 2 Summary of reference values for constituents of extruded samples in calibration and validation sets 118

Table 3 Calibration and validation statistics for prediction by each constituent's (WAI, WSI, texture analysis) best model using multiple linear regression 123

Table 4 Calibration and validation statistics for prediction of RVA parameters's best model using multiple linear regression 124

Table 5 Calibration and validation statistics for prediction by each constituent's (WAI, WSI, texture analysis) best model using partial least squares regression 125

Table 6 Calibration and validation statistics for prediction of RVA parameters (RVU)'s best model using partial least squares regression 126

LIST OF FIGURES

Chapter 2

Fig. 1 Average absorbance of sound wheat kernels, 3- and 4-week-old granary and maize weevils infested kernels 84

Chapter 3

Fig. 1 Average spectra (log 1/R) of extrudates at low, medium and high degree of processing 120

Fig. 2 Average spectra of extrudates at low, medium, and high degree of processing after treatment with first derivative 121

Fig. 3 Average spectra of extrudates at low, medium, and high degree of processing after treatment with second derivative 122

Fig. 4 Reference vs. NIR modeled values for water absorption index (WAI) of validation samples 128

Fig. 5 Reference vs. NIR modeled values for water solubility index (WSI) of validation samples 129

Fig. 6 PLS weights from second derivative spectra of extrudates from calibration model used to predict WSI 130

Fig. 7 Reference vs. NIR modeled values for hardness of validation samples 132

Fig. 8 Reference vs. NIR modeled values for fracturability of validation samples 133

Fig. 9 Reference vs. NIR modeled values for RVA cold viscosity of validation samples 135

Fig. 10 Reference vs. NIR modeled values for RVA peak viscosity of validation samples 136

Fig. 11 PLS weights from second derivative spectra of extrudates used to develop calibration models for predicting cold and peak viscosities 137

Fig. 12 Reference vs. NIR modeled values for RVA final viscosity of validation samples 138

Fig. 13 Reference vs. NIR modeled values for RVA trough viscosity of validation samples 139

Fig. 14 PLS weights from second derivative spectra of extrudates from calibration models used to predict final viscosities (excluding wavelengths of 1901-1979 nm) 140

Fig. 15 Reference vs. NIR modeled values for RVA breakdown of validation samples 142

Fig. 16 Reference vs. NIR modeled values for RVA setback of validation samples 143

Fig. 17 PLS weights from first derivative spectra of extrudates from calibration models used to predict breakdown (excluding wavelengths of 1901-1979 nm) 144

OVERVIEW

Insect infestation in stored grain can cause serious economic losses, since damage can be developed without any visible external signs. Several methods such as an acid hydrolysis test for insect fragments, cracking and flotation methods, and X-ray inspection have been accepted as official procedures; however, they still lack reliability and reproducibility. In addition, they are time-consuming, difficult to automate, and labor-intensive. They also require specialized equipment and skilled technical personnel, and the use and disposal of hazardous organic solvents. Several studies have been published indicating that the NIR technique is feasible and reliable to detect internal insect infestation in wheat kernels. This study was carried out to investigate the use of a simplified method by using a limited number of selected discrete wavelengths. This method would allow the use of much simpler and less expensive NIR instruments for detecting insect infestation than the use of full spectrum methods that have been previously reported.

NIR reflectance spectroscopy was also applied to cereal products for predicting the degree of extrusion cooking of corn meal at different extrusion parameters. Only a few studies have been reported on a rapid method, which can directly measure the dependent variables related to starch structure during extrusion processing. Multiple linear regression (MLR) and partial least squares (PLS) regression were used to develop NIR calibrations for predicting the various indicators of degree of cook.

This dissertation consists of a literature review, research topics in three chapters, and a conclusion. Chapter 1 shows a comparison of selected discrete wavelength and principal component analysis methods for insect infestation detection. Chapter 2

describes a simplified near-infrared method for detecting internal insect infestation in wheat kernels. Chapter 3 demonstrates prediction of degree of extrusion cooking of corn meal by near-infrared reflectance spectroscopy. Finally, a conclusion section summarizes the major research findings presented within this dissertation.

LITERATURE REVIEW

1. Near-Infrared Spectroscopy

Near-infrared spectroscopy (NIRS) can measure the chemical composition of biological materials by using the diffuse reflectance or transmittance of the sample at several wavelengths (Workman and Shenk, 2004). The NIR spectrum consists of a number of absorption bands that vary in intensity due to energy absorption by specific functional groups in a sample (Dunmire and Williams, 1990). NIRS is useful for the study of hydrogen bonding because it measures overtones and combinations of the molecule's vibrational modes, principally those involving hydrogen. In other words, NIRS can measure the concentration of components having different molecular structures such as protein, water, or starch (Murray and Williams 1990). The NIR spectral region, from 700 to 2500 nm, lies between the visible and mid-infrared regions of the electromagnetic spectrum. Other names for spectroscopy in this range are "far-visible spectroscopy" or "overtone vibrational spectroscopy". NIR spectra consist of overtones and combination bands of the fundamental frequencies in the mid-IR region. Low reflectivity and low absorptivity allow NIR energy to pass readily into many organic substances. Low reflectivity means that energy penetrates readily beneath the surface of most samples, including visually opaque samples. Low absorptivity means that NIR light energy passes easily through samples without rapid attenuation.

NIRS has been widely used for various foods and commodities, especially in the grain, cereal products, and oilseed processing industries (Wehling, 1998). It also is used in breeding programs for quality improvement of cereals, such as wheat flour yield, barley malting quality, durum semolina yield, rice milling yield and oat groat percentage.

In addition, it can be applied to crop management, receivable testing, and on-line process control (Osborne, 2006; Osborne, 2007). The NIR technique is fast (analysis time varies from seconds to minutes per test), reliable, non-destructive, and inexpensive in terms of cost-per-test for various uses (Osborne, 2000). Furthermore, there is no sample preparation or pretreatment, no need for dangerous reagents or solvents, and no disposal problem, either. These advantages can eliminate sampling errors caused by manual sample handling and reagent contamination. Samples may also be retained for further analysis. It can also be performed by technically unskilled personnel at-line or automatically on-line. Additionally, NIR analysis can obtain results for many constituents simultaneously by collecting the NIR spectrum of a sample over a range of wavelengths during a single scan, and analyzing the spectrum using multiple calibration equations. The calibration equations are developed through a modeling process, using chemometric methods, that employs a training set of samples to teach the computer to relate the subtle differences in spectral features to sample composition (Drennen et al, 1990). A single spectrum can be subjected to many different calibration models, to measure any number of constituents.

1.1 Application of NIR Spectroscopy in Qualitative Analysis

NIRS has been used extensively in the chemical and pharmaceutical industries for classification of raw materials (Wehling, 1998). It is currently becoming more useful in food applications. Brimmer et al (2002) reviewed the capabilities of using online NIRS to monitor and control manufacturing processes. By using pattern recognition algorithms, NIRS can provide a qualitative assessment of a process sample to ensure that its NIR spectrum is within an acceptable range of variability for what is considered a

good quality product. For example, a qualitative process NIR analyzer can be used in industrial manufacturing processes for product identification, product composition measurement, and product uniformity evaluation.

Dowell (2000) used NIRS to classify vitreous and nonvitreous single kernels of durum wheat. They found that the vitreousness of durum wheat could be quantified, perhaps because of scattering effects, or the differences in protein and starch content on NIR absorption. Studies by Wang et al (2002) showed that dark hard vitreous and nonvitreous kernels of hard red spring wheat, including kernels that are checked, cracked, sprouted, and bleached can be classified by using visible/NIR spectroscopy. They surmised that scattering is a major factor contributing to the classification of those kernels. Protein content, kernel hardness, starch content, and kernel color also are main contributors to the classification.

There are several studies focused on using NIRS in qualitative analysis based on discriminant analysis. Shah and Gemperline (1990) found that the use of principal component analysis (PCA) and Mahalanobis distances provided accurate models for classification of pharmaceutical raw materials. Delwiche and Norris (1993) also obtained the best results for classifying hard red winter and hard red spring wheats when using a discriminant analysis model based on the same technique of combining PCA and Mahalanobis distances. Subsequently, Chen et al (1995) using the same NIR spectra as Delwiche and Norris (1993), found that Neural Networks yielded better results than PCA coupled with Mahalanobis distances for classification of those wheats. However, they mentioned that Neural Networks are less interpretable than PCA techniques because of non-linear transfer data. Classification of bulk wheat (hard red spring vs. hard red

winter) was reported by Delwiche et al (1995).

In breeding programs, Delwiche and Graybosch (2002) used NIR reflectance spectroscopy to classify waxy wheat from partial-waxy or wild-type wheat using PCA according to variation of amylose content. Those results showed that linear or quadratic discriminant functions of the PC scores successfully separated waxy from non-waxy wheat. Delwiche et al (2006) recently examined the potential of NIRS of single durum wheat kernels for classification by waxy allele using linear discriminant analysis models. They successfully classified the full waxy genotype, whereas the classification accuracy of the non-waxy genotypes was very poor. Pasikatan and Dowell (2004) also created a means to rapidly segregate high- and low-protein single wheat kernels using a high-speed color sorter equipped with near-infrared optical filters at 920/1660 nm. They, however, found that the sorting was partly driven by vitreousness and color differences.

1.1.1 Discriminant Analysis

Unlike quantitative analysis, the need for actual calibration of variables is not necessary in qualitative analysis. One approach that has been used successfully for qualitative analysis is discriminant analysis, which is sometimes called pattern recognition. The purpose of this technique is to classify samples into well defined groups based on a "training set" of similar samples with limited knowledge of the composition of the group samples. Johnson and Wichern (1998) stated that the concept of using discriminant analysis is to use several variables and to see how the observations cluster together. Two methods are used in this study for developing calibration models in discriminant analysis, e.g., Mahalanobis distances, and PCA coupled with Mahalanobis distances.

1.1.2 Mahalanobis distances

Discriminant analysis based on the unit distance vector in multidimensional space, called Mahalanobis distances, has been described (Mark and Tunnell, 1985). The Mahalanobis distance can be described by an ellipsoid in multidimensional space that circumscribes the data. This method uses a matrix that describes the inverse of the matrix formed by pooling the within-group covariance matrices of all groups, which is generated by combining information from all the different materials of interest into a single matrix. Use of this matrix, therefore, defines a common metric for all groups in the data set, indeed for the multidimensional space. Dunmire and Williams (1990) described the Mahalanobis distance as the mathematical quantity that defines the position, size and shape of the ellipsoid for all clusters.

From a statistical viewpoint, the Mahalanobis distance takes the sample variability into account, whereas the Euclidian distance method does not take into account the variability of the values in all dimensions. In other words, Mahalanobis distances look at not only variation between the responses at the same wavelengths, but also at the inter-wavelength variations. Instead of treating all values equally when calculating the distance from the mean point, it weights the differences by the range of variability in the direction of the sample point. The location of each cluster in multidimensional space is described by the mean value of the absorbances (the group mean) at each wavelength. Theoretically, one Mahalanobis distance is the distance from the center of each data cluster to the edge of the ellipsoid in the direction of the measurement, and samples that have a Mahalanobis distance of 3σ (three standard deviations) or greater have a probability of 0.01 or less and can be classified as non-

members of the group. Dunmire and Williams (1990) stated that the sample can be classified unambiguously if it falls within three times the Mahalanobis distance from the respective centroid and at least six times the Mahalanobis distance from the ellipses of other groups. Mark and Tunnel (1985) briefly explained that the Mahalanobis distance is a multidimensional distance D defined by the matrix equation as follows:

$$D^2 = (x-\acute{x})'M\,(x-\acute{x})$$

Where D is the Mahalanobis distance, x is a vector consisting of optical readings at several wavelengths which describes the position in multidimensional space corresponding to the spectrum of a given sample, $\acute{x}$ is a vector describing the position of a reference point in space, and M is the pooled inverse covariance matrix describing distance measures in the multidimensional space of interest.

1.1.3 Principal Component Analysis coupled with Mahalanobis distances

PCA has been coupled with Mahalanobis distances to reduce dimensionality before carrying out the discriminant analysis (Osborne et al, 1993). PCA is a reduction technique that extracts from a large number of variables to a much smaller number of new variables, which account for most of the variability between samples and contain information from the entire spectrum. In other words, PCA decomposes the training set spectra into mathematical spectra (i.e. loading vectors, factors, principal components, etc.) which represent the most common variations to all the data. Johnson and Wichern (1998) summarized that the first principal component (PC) explains most of the variation in the original data set. It is, therefore, the most significant principal component. The second PC is less important and uncorrelated with the first one. Plots of PCs versus each other show how the variables that they account for are related. To see how the

observations cluster together, information from PCA is needed to calculate a set of scaling coefficients, called scores. Plots of PC scores versus each other can then be used to see how the observations are related.

The scores for each factor can be calculated for every spectrum in the training set. When the scores are multiplied by the loading vectors, and the results are summed, the original spectra are then constructed. Therefore, by knowing the set of loading vectors, the scores will represent the spectra as accurately as the original responses at all the wavelengths. Principle component analysis avoids the problem of overfitting by selecting too many wavelengths. This pattern recognition technique was used to measure the Mahalanobis distances that are calculated in units of standard deviations from the center (mean) of the training set cluster.

According to PLSplus/ IQ manual (Galactic Industries, Salem, NH), the use of an optimum number of factors can protect against underfitting which represents models that do not contain enough factors. Cross-validation is one approach that is used for determining the optimum number of factors. For performing this diagnostic, each sample in the calibration set is removed one by one and the remaining samples are used to construct a Mahalanobis matrix for one, two, three factors, and so on. Then using the models developed for Mahalanobis grouping, the excluded sample is predicted. The excluded sample is then returned to the calibration set, and a new sample is removed. The process is continued until all samples are removed from the calibration set and predicted. This provides an advantage of cross-validation over other methods since the predicted samples are not the same as the samples used to build the model.

The sensitivity of the analysis can be enhanced by using a spectral residual, which

is calculated by subtracting the reconstructed spectrum from the original spectrum (PLSplus/IQ manual). The reconstructed spectrum is derived from multiplying the spectrum scores by the set of primary factors and summing the results. Including the sum squared spectral residual as an additional spectral score for the Mahalanobis group can improve the sensitivity of the unknown sample classifications.

1.2 Application of NIR Spectroscopy in Quantitative Analysis

NIR instruments can be calibrated to quantitatively measure various constituents in food and agricultural commodities. Wehling (1998) described an equation that can be used to predict the amount of a constituent present in a food from its spectral measurements as follows:

$$\% \text{ constituent} = z + a \log (1/R_1) + b \log (1/R_2) + c \log (1/R_3) + ...$$

where each term represents the spectral measurement at a different wavelength multiplied by a corresponding coefficient. Each coefficient and the intercept (z) are determined by multivariate regression analysis.

There are numerous studies describing quantitative analysis by NIRS in various types of food. It provides an excellent method for the measurement of chemical composition (i.e. protein, starch, lipid, and moisture contents) in raw pork and beef (Lanza, 1983), in cheese and other dairy products (Baer et al, 1983; Rodriquez-Otero et al, 1995), and in feed for ruminant animals (Liu and Han, 2006). However, it is most widely used in the field of grains and cereal products. In some cases, such measurements are important to achieve the end-used objectives of a plant breeding program.

There are several investigations using NIRS to predict viscosity properties of rice. Delwiche et al (1996) developed calibration models on whole-grain milled rice using PLS

regression to predict viscosity properties of a flour-water paste as recorded by the Rapid ViscoAnalyzer (RVA), and other measurements that determine the cooking and processing characteristics of rice. Their results showed that none of the RVA parameters were modeled by NIR with high accuracy (r^2 ranged from 0.424 to 0.737). They reasoned that the variations in amylose-amylopectin ratio may be the primary reason why the NIR models for RVA constituents were not highly accurate.

Meadows and Barton (2002) later used NIRS to predict RVA data in rice flour. Instead of using the RVA profile obtained by following the AACC procedure as reported by Delwiche et al (1996), they used a different set of experimental conditions for RVA measurements. A PLS regression of NIR spectra (1,100-2,500 nm) vs. RVA viscosity at 0-752 sec was performed and showed that the highest correlation (r = 0.961-0.903) to NIR was at 212-228 sec, which is between the initial pasting time and peak viscosity. They mentioned that this period of time is crucial because of water absorption and rapid granular swelling. The potential of rice flours to form pastes can be distinguished because of their differences in water absorption, granule disruption, and the development of gelatinization. Furthermore, Bao et al (2001) successfully used NIRS to predict the pasting parameters of set back and break down, and gelatinization peak temperature of rice flour. Gel consistency, cool paste viscosity, gelatinization onset temperature, and textural properties were not as well predicted by NIR as those preceding parameters. Texture of cooked rice also is predicted by NIR analysis of whole grain rice (Meullenet, 2002). Five of seven sensory texture attributes can be predicted by NIR using PLS analysis of second derivative spectra to develop calibration models.

The determination of phosphate content and viscosity properties of potato starch

was studied by Thygesen et al (2001). Their results showed that NIR-based prediction of phosphate content using PLS regression was possible with a root mean square error of cross validation (RMSECV) of 0.006%. They also found that the PLS regression of NIR data was successful to predict RVA peak viscosity and breakdown. Prediction of other RVA variables was not possible.

Additionally, NIRS has been used to predict corn processing quality. Wehling et al (1993) studied the feasibility of the NIR reflectance technique to predict wet-milling starch yield from whole-kernel corn. Calibrations were developed using MLR and PLS regression. MLR of second-derivative spectra yielded the best results. However, a limiting factor in the performance of the NIR method was the lack of reproducibility of the laboratory wet-milling reference method. The NIR technique was later applied to the dry-milling process. Wehling et al (1996) found that dry-milling characteristics of corn can be reliably predicted for at least rough screening purposes. Campbell et al (1999) also predicted starch and grain amylose contents in corn by NIR transmittance spectroscopy. Calibrations were developed using PLS and artificial neural network (ANN) analyses. The results showed limited precision of this method. However, it can be used as a rough screening method for starch amylose content.

Additionally, NIRS can be used to investigate some properties of final products. Wesley et al. (1999) developed the use of dynamic NIRS with a diode array instrument to assess dough mixing time, a crucial stage in the production of bread products. In studies by Xie et al. (2003), an NIRS method was compared to a texture analysis (TA) method for measuring bread staling during storage. Their results indicated that NIR spectra (550-1700 nm) had a high correlation to firmness as measured by TA. They also found that

NIRS measures bread changes more accurately and more precisely than the TA. They explained that NIRS measurement is based on both physical and chemical changes during bread staling, whereas the TA method measures only bread firmness. Xie et al (2004) later reported that NIR spectra correlated strongly with differential scanning calorimetry (DSC) for measuring amylopectin retrogradation in bread staling. The important wavelengths were 550, 970, 1155, 1395, and 1465 nm. NIRS not only provided information about changes in bread moisture and starch structure, also differences in bread color and protein content.

In addition, different types of manufacturing processes can be controlled by quantitative information obtained from NIR measurement. Nowadays, requirements of quality control in grain milling and food processing increasingly call for on-line analyses (Osborne, 2006). Gradenecker (2003) reviewed some applications of NIR on-line measurement. They surmised that it can be used to monitor the quality of wheat in grain milling. Moreover, it can determine protein, moisture, and ash contents of wheat flour and semolina products. Changes in water absorption rate, starch damage, particle size distribution, and color measurements can also be achieved.

1.2.1 Chemometric Methods in Quantitative Analysis

Chemometric methods are means to perform calculations on measurements of chemical data. Selecting a set of calibration, or training, samples is the first and most crucial step in the chemometric application of NIR to process analytical control. Unlike usual laboratory practices for building a calibration curve, chemometrics require that a large number of samples be included in the calibration set for quantitative analysis. The best set of calibration samples should represent all of the possible variations in the

process (Cartagena et al, 1994). These include containing the constituent of interest at levels covering the range that is expected to be received, and containing a relatively uniform distribution of concentrations across that range. Once the calibration model is built using algorithms, it must be tested to ensure that the model is representative of the analyte under investigation and will produce accurate predictions for samples not used in the development of the model.

Derivatized reflectance data, either first or second derivative, has been applied to reduce particle size effects of samples (Osborne et al, 1993). It reduces the high correlations between spectral data at different wavelengths, simplifies the calibration procedure, and therefore, should lead to calibrations more robust to particle size variations. The use in NIR of derivatives beyond the second has not been successful since the signal to noise ratio decreases with each successive operation. Multiplicative scatter correction (MSC) also has been used to compensate for particle size effects (Delwiche, 1998). It rotates the spectra to remove some of this effect. MSC rotates each spectrum so that it fits as closely as possible to the mean spectrum. There are several calibration modeling methods, or algorithms, that are used to explain the relationships between a given constituent and an absorbance spectrum. Each method provides its own advantages and drawbacks, and none of them is better than others for every application. MLR is based on a multilinear equation, and is used to select the optimum wavelengths for measurement and the associated coefficients for each wavelength (Wehling, 1998). Wavelengths are selected by using a forward or reverse stepwise regression procedure, or by using a computer algorithm that tests all possible combinations of two, three, or four wavelengths to determine the combination that gives the best results. In the forward

stepwise regression, an equation is built up by adding wavelengths one at a time, each wavelength being chosen so that the resulting equation has the smallest residual sum of squares possible. Reverse stepwise regression is backward elimination, applicable only where an equation involving all the candidate wavelengths may be fitted as a first step. The wavelengths are then removed one at a time, the one removed at each step being that which causes the least increase in the residual sum of squares.

Partial least squares (PLS) regression and principle component regression (PCR) are examples of quantitative regression algorithms that are currently used for linear data. Both PLS and PCR are factor-based models. Instead of using a few selected wavelengths, PLS and PCR use information from all wavelengths in the entire NIR spectrum to predict sample composition. Wehling (1998) explained that PLS and PCR use data reduction approaches to reduce a large number of variables to a much smaller number of new variables that account for most of the variability in the samples. The amount of a constituent in samples can then be predicted by these new variables. PLS is similar to PCR but is more sensitive to variations in sample concentration. Osborne et al (1993) stated that PLS tends to produce solutions that require fewer factors than calibrations of comparable performance produced by PCR. PLS is a regression algorithm that uses concentration data during the decomposition process and includes as much information as possible into the first few loading vectors (Dowell et al, 1998). It performs the decomposition on both the spectral and concentration data simultaneously. A small number of factors are constructed as linear combinations of the original spectral data and regression on the factor scores is used to derive a prediction equation.

Some examples of nonlinear regression algorithms are artificial neural networks

(ANN), projection pursuit regression (PPR), and multivariate adaptive regression splines (MARS). Micklander et al (2006) studied the possibility of using different types of food with known fat content to develop calibration models and evaluated prediction results from different linear (PLS) and non-linear (neural networks and local regression techniques) calibration models. They found that the non-linear models gave the smallest root mean square error of prediction (RMSEP) for multi-product models.

2. Problems of Insect Infestation in Stored Grain

Insect infestation of stored grain has long been a major problem. In the US, grain loss due to insect damage is estimated to be around millions of dollars per year (Flinn et al, 2003). The U.S. standards consider wheat to be infested if 2 live insects injurious to wheat are found in a 1-kg sample (FGIS, 1997). The Food and Drug Administration (FDA) has set the defect action level (DAL) as the regulatory standard for quality control. For insect contamination, the DAL is 32 insect damaged kernels per 100 g of wheat and 75 insect fragments per 50 g of wheat flour (FDA 1998). Internal infesters, referred to as "hidden insects" (USDA, 1986), are the most serious grain infesting insects, since damage can develop without any visible external signs and cannot be completely removed by standard grain cleaning operations.

The primary infesters, granary weevil [Sitophilus granarius (L.)], maize weevil (Sitophilus zeamais), rice weevil [Sitophilus oryzae (L.)], and Lesser grain borer [Rhyzopertha dominica (F.)] are the most damaging species that develop to maturity inside grain kernels. They may also cause serious problems in the later stages of marketing channels (Vick et al, 1988) by causing products to become unsuitable for use

as human food (Russell, 1988), since they are the main source of insect fragments in flour (Harris et al, 1952). In contrast, external feeders, such as Indian meal moth [Plodia interpunctella(Hbn)], Cigarette beetle [Lasioderma serricorne (F.)], and Cadelle [Tenebroides mauritanicus (L.)] develop outside whole grain kernels. They are more easily removed by grain cleaning processes (Toews et al, 2007).

2.1 Detection of Insect Infestation

Numerous studies have focused on the development of methods for detecting internal insects. The most commonly used methods for insect detection generally rely on visual inspection of grain for insects and insect damaged kernels (Nicholson et al, 1953; USDA/FGIS, 1987). However, none of these methods can be used for detecting hidden or internally developing insects that have not emerged (Russell, 1988). Several methods have been developed to identify internal infestation including staining of egg plugs made by female weevils after laying eggs in the kernels (Goossens, 1949); flotation of hollowed kernels left by feeding insects (Apt, 1952); cracking of kernels followed by flotation to concentrate the released insect parts (Harris et al, 1952); and crushing the kernels onto paper impregnated with ninhydrin to detect amino acids corresponding to the insect (Dennis and Decker, 1962).

Other methods include measurement of respired carbon dioxide (Street and Bruce, 1976) using headspace infrared analysis. It is, however, unreliable because of the difficulty in correcting for the background level of carbon dioxide released by respiring grains. An additional complication is that the instrument must detect as little as a few ppm of CO_2 produced by insects, against a normal atmospheric background of CO_2. Hackman and Goldberg (1981) used a colorimetric procedure for measuring chitin, a

major component of insect cuticle, as an index of insect infestation in grains. This method, however, may not be adequately specific for use as an index of insect infestation in grain because a high concentration of chitin may also be found in stored grain contaminated with fungi. Chambers et al (1984) developed a technique using a nuclear magnetic resonance (NMR) spectrometer for monitoring the development of granary weevils in wheat kernels. Peaks for water and lipid were reported in wheat kernels containing third and fourth larval instars. Nevertheless, this method is not widely accepted due to the complexity of the instrument. Determination of uric acid, which is the principal end product of nitrogen metabolism of almost all terrestrial insects and accounts for more than 80% of the nitrogen of their excreta, using high-performance liquid chromatography (HPLC) has been studied (Pachla and Kissinger, 1977; Wehling and Wetzel, 1983; Ghaedian and Wehling, 1996). Fluorometry has also been used to detect uric acid (Lamkin et al, 1991). An enzyme-linked immunosorbent assay (ELISA) of kernel extracts for measuring myosin, which is a muscle protein found in all insects, has been studied (Kitto, 1991; Chen and Kitto, 1993; Rotundo et al, 2000). All these procedures can detect internal insects or larvae in kernels with varying degrees of success.

An advanced technology of acoustic detection has also been investigated (Vick et al, 1988; Shuman et al, 1993; Mankin et al, 1996; Hagstrum et al, 1996). Drzewiecki and Shuman (2001) later developed acousto-fluidic sensors to exploit the high sensitivity and signal-to-noise ratio. However, it has not been accepted as an official method, since the false positives are most often caused by electrical noise because grain is a good insulator. Pearson et al (2003) subsequently used electrical conductance, which is monitored by

measuring the voltage across the kernel. Infested kernels are classified from sound kernels based on the signal characteristics of the system and by computing the range of voltage levels in the conductance signal. Although this method is inexpensive, it is time consuming and cannot detect kernels containing dead larvae or pupae. In addition, the classification rates by this technique are very low when compared with the inspection by soft X-rays, and X-ray imaging (Karunakaran et al, 2003; Karunakaran et al, 2004; Fornal et al, 2007).

According to Schatzki and Fine (1988), an X-ray inspection technique has been developed to detect single wheat kernels infested with maize weevil, rice weevil, lesser grain borer, and angoumois grain moth at various ages of the insect. They reported that the presence of hidden insects was detected with 80% accuracy for large larvae (3^{rd} and 4^{th} instar), but misclassification increased exponentially as the size of insect decreased. False positives were also reported to be 0.8%. This was due to difficulty in distinguishing between the germ portion of the kernels and insect larvae, and damaged kernels appearing identical to 2^{nd} and 3^{rd} instar larvae. Time consumption is another concern for this method since only 13 grams of wheat could be measured per day. Keagy and Schatzki (1993) developed an image processing algorithm for machine recognition of hidden weevils in wheat radiographs; however, a low percentage of recognition (50% and 72% recognition of fourth instar larvae of maize and granary weevils, respectively), and a false positive rate of 0.5% were reported.

Although the acid hydrolysis test for insect fragments, the cracking and flotation methods, and X-ray inspection have been subsequently accepted as official procedures (AACC, 2001), they still lack reliability and reproducibility. In addition, they are time-

consuming, difficult to automate, and labor-intensive. They also require specialized equipment and skilled technical personnel, and the use and disposal of hazardous organic solvents. Pedersen (1992) and Brader et al (2002) reviewed some of these screening methods used to detect insect infestation in wheat.

2.2 Use of NIR Spectroscopy for Detection of Insect Infestation

Detection of internal insect infestation of wheat kernels using NIR reflectance spectroscopy has been reported. In the United Kingdom, Ridgway and Chambers (1996) successfully detected internal infestation by granary weevil. They investigated both bulk samples infested with larvae and pupae, and single wheat kernels infested with larvae. Their best calibration equation was obtained from PLS regression of standard normal variate-transformed spectral data (1100-2500 nm) against the infestation level. They concluded that both physical and chemical effects are responsible for differences in absorption intensity of the spectra. Chemical effects are thought to be due to moisture arising from metabolic processes of insects, insect protein and/or chitin. They also found that the criterion log1/R(1194 nm)- log1/R(1304 nm), without any form of scatter correction, can be used to classify sound kernels and kernels infested with pupae. The intensity of log 1/R spectra of infested kernels was lower at 1194 nm, which represented the starch band, leading them to conclude that the spectral response could likely be due to wheat starch lost as a result of insect feeding. Ridgway and Chambers (1998) further incorporated a simple filter-based NIR imager into a machine vision system. Differences between infested and uninfested kernels were enhanced by subtracting the image at 1300 nm from the image at 1202 nm. They concluded that the differences between those kernels are most likely due to chemical composition changes, such as loss of starch by

insect feeding, rather than increased light scattering from the insect cavity in the kernels. Ridgway et al (1999) later investigated the region from 700 nm to 1100 nm for detection of grain weevil larvae and pupae inside single wheat kernels and found that two-wavelength models based on either log 1/R (982 nm)- log 1/R (1014 nm) responding to decreasing grain starch, or log 1/R (972 nm)- log 1/R (1032 nm) responding to increasing grain moisture caused by insect activity, could correctly classify (>96% accuracy) uninfested and infested kernels by S. granarius larvae. The result was comparable to that of a full spectrum method (1100- 2500 nm).

Research in our laboratory by Ghaedian and Wehling (1997) developed calibration models using discrimant analysis based on loadings derived from PCA of full (1100–2498 nm) or partial (1100–1900 nm) NIR spectra to classify sound and granary weevil infested kernels. PCA of NIR spectra from sound kernels was used to construct calibration models by calculation of Mahalanobis distances. A five-factor PCA model from a first derivative spectral transformation in the 1100-1900 nm region provided the highest overall correct classification rates. Eliminating the region from 1900-1980 nm probably removed the differences in kernel moisture content. They found that the region 1980 to 2498 nm did not contain enough information for reliable classification. They further applied discriminant analysis based on Mahalanobis distances to log 1/R data from selected discrete wavelengths. Similar results were obtained when using 12 wavelengths, except for infested air-dried kernels. They concluded that protein, lipid (1200, 1360, 1440 and 1660 nm), and phenolic (1420 nm) compounds are the major sources of variation between sound and infested wheats.

Dowell et al (1998) showed the feasibility of using a near-infrared diode array

spectrometer integrated with an automated, single kernel characterization system to detect the presence of rice weevil, lesser grain borer, and Angoumois grain moth larvae. PLS regression in the wavelength ranges of 1000-1350 and 1500-1680 nm was used to build the calibrations. They observed that absorbance peaks of ground insect cuticle or chitin were at 1178 nm and 1500 nm. Thus, chitin present in insect tissues may explain differences between sound kernels and those containing larvae. They also studied the ability of NIRS to detect the smallest larval size in the kernels and found that 3rd and 4th instars were detected with 95% confidence, but the technique could not reliably detect smaller larvae. Classification accuracy of insect larvae was not dependent on wheat class, protein and moisture content, or insect species. They also stated that no additional useful information was obtained by including the wavelength region lower than 1100 nm. This can be explained, as absorbances from 700 nm to 1000 nm are primarily caused by weak third overtones of fundamental absorptions, and are difficult to measure (Murray and Williams, 1990).

Maghirang et al (2003) also reported the use of an automated NIRS system to detect single wheat kernels containing live or dead rice weevils. They developed calibrations using the wavelength region 950-1690 nm to detect both live and dead insects in wheat. The results showed that calibrations developed using live pupae and live large larvae correctly classified 86% to 96% of kernels having dead pupae and dead large larvae. The dead pupae and dead large larvae calibration correctly detected 92% to 93% of live pupae and live large larvae present.

In Canada, a study using NIR spectroscopy to detect rice weevil and lesser grain borer in stored wheat has also been reported by Paliwal et al (2004). They used PCA to

distinguish wheat kernels infested with pupae of those insect species at different infestation levels, and found that it was possible to differentiate those kernels with at least 25% infestation level. Easier of differentiation of insect species was obtained as infestation levels increased. They also developed calibration models for quantitative determination of infestation levels using PLS regression and found that high classification accuracies were achieved for high infestation levels but lower at low infestation levels.

Neural network analysis was subsequently applied for calibration development, and compared with PLS to detect different species of beetles commonly associated with stored grain (Dowell et al, 1999). The neural network analysis gave a higher percentage of correct classification (>99%) for primary and secondary insects, whereas PLS gave a lower classification percentage. Both calibrations classified insects by genus within primary and secondary groups with an accuracy of >95%. Baker et al (1999) used the same method and analysis procedure as Dowell et al (1998), to detect wheat kernels containing the parasitoid Anisopteromalus calandrae from rice weevils, and to separate sound kernels from those containing weevil or parasitoid larvae or pupae. They demonstrated that their model correctly classified 97.4% of sound kernels, and correctly classified as infested 90–100% of kernels with either host or parasitoid larvae or pupae, when using 13 PLS factors. When differentiating weevils from parasitoids, kernels containing weevil larvae and pupa could be classified with 100% accuracy from parasitoid pupa. The method could not reliably classify parasitoid larva, probably because of their small size.

There are additional studies on NIR detection of insect infestation. Tigabu and

Oden (2002) successfully classified sound and insect-infested tropical multipurpose trees, Cordia africana. A partial least squares model derived from orthogonal signal corrected (OSC) spectra provided the best classification results. OSC removes undesired spectral variation by taking into account the variables in its algorithm and minimizing the covariance between response variables. They stated that differences in composition of chitin, cuticular lipid in insects, and moisture content resulted in the difference spectrum and partial least squares weight.

Detection of insect fragments in wheat flour using NIRS has been studied by Perez-Mendoza et al (2003). They reported significant correlation between an NIR method with a spectral range of 400-1700 nm, and the actual number of insect fragments (less or higher than 130) from rice weevils in flour samples. However, the method was not sufficiently accurate for predicting numbers below the FDA defect action level. Further, Perez-Mendoza et al (2004) successfully graded chronological age of adults of rice weevil, lesser grain borer, and red flour beetle, three pests of stored grain. Decreasing water content and increasing cuticular lipids with increasing age were the major factors for NIRS to classify young from old beetles. Subsequently, Perez-Mendoza et al (2005) extended the NIR region to 2500 nm and detected insect fragments in flour produced from infested wheat. They found that wheat infested with a single preemergent adult of lesser grain borer provided more fragments than wheat infested with a single larva or pupa. They also surmised that the maximum level of internal infestation that can be accepted by millers to produce flour with fragment counts below the FDA defect action level ranged from 0.95 to 1.5% (380-640 infested kernels/kg of wheat) for pupae and larvae. But when the grain is internally infested with preemergent adults, the level of

maximum infestation is decreased to less than 0.05% (20 infested kernels/kg of wheat).

3. Extrusion Cooking

The extrusion technique has been used widely for processing snack foods. It has provided an approach for manufacturing new and novel products and has revolutionized many conventional snack manufacturing processes. Extrusion equipment gives many basic design advantages that result in minimizing time, energy, and cost while increasing the degree of versatility and flexibility that were not previously available (Sevatson and Huber, 2000).

Textural and nutritional properties of extruded products depend on the chemical interactions and structural changes undergone by the constituents in the extruder. Starch plays a major role in the overall quality of many processed foods. Starch consists of two glucose polymers: amylose, which is linear, and amylopectin, which is highly branched. In the starch granule, the amylose and amylopectin are intermingled but when the short linear segments of amylopectin align they become ordered into crystallites. The crystallinity arises from the extensive hydrogen bonding, both intramolecular and to water molecules, of the amylopectin molecules. The addition of thermal energy, either transferred from a heater through the barrel and/or generated by viscous dissipation, is very important in extrusion cooking (Harper, 1981). It also causes progressive disruption of the hydrogen bonds and changes in the physical characteristics of the starch granules. When starch is heated in water, the hydrogen bonds are progressively broken causing the granule to swell so that the amylose gradually diffuses out of the granule, destroying the crystalline structure and resulting in the formation of a gel. The increase in viscosity is

referred to as pasting, and occurs following gelatinization, which is the initial loss of order in the granule. A major difference between extrusion and other forms of food processing is that gelatinization occurs at much lower moisture levels (12-22%), since intensive shear force also is developed in the extruder. The extent of transformation of raw materials in extrusion cooking, referred to as the degree of cook, is crucial to final product quality. Cooking degree increases when there is an increase in depolymerization eof the starch molecules, resulting in an increase in the number of free hydroxyl (O!H) bonds and a decrease in paste viscosity.

Direct expanded snacks such as corn curls, balls, and rings are the majority of extruded snacks. Cornmeal is fed into an extruder with a feeding device at a constant rate. The meal is exposed to moisture, heat, and pressure as it is transported through the extruder toward the extruder die. Extruders for direct expanded snack products are normally short in length. Additionally, it is important that the extruder configuration consisting of screws, steamlocks, and barrel segments be properly selected to feed, knead, and cook the process material as it passes through the extruder.

3.1 Analysis of Extruded Products

Measurement of starch changes serves as an indicator of the amount of degradation during the cook, and also represents functional aspects of the product. There are several analytical methods that have been used to evaluate extruded products. The Brabender amylograph has been used previously to measure viscosity of extrusion-cooked starch (Mason and Hoseney, 1986). The Rapid ViscoAnalyzer (RVA) was later developed to permit measurement of viscosity properties similar to those measured on the Brabender Amylograph, but in one-fifth the time (Whalen et

al, 1997). RVA has been used extensively to study starch pasting characteristics for starch-based extruded products (Whalen et al, 1997; Becker et al, 2001; Ganjyal et al, 2006). Pasting parameters from RVA provide information related to starch gelatinization, disintegration, swelling, and gelling ability. In spite of this saving in time, fewer than five samples per hour can be run on the RVA instrument. The effect of extrusion on water solubility (WSI) and water absorption (WAI) indices of starch has been reported (Chinnaswamy et al, 1989; Hashimoto and Grossmann, 2003; Baik et al, 2004; Ganjyal et al, 2006). High-performance size exclusion chromatography (HPSEC) has been subsequently used to provide information on the molecular size distribution of starch polymers due to a cook (Jackson et al, 1990). Differential scanning calorimetry (DSC) has been used to study the gelatinization or melting characteristics of crystalline forms in starch (Bao et al, 2001; Becker et al, 2001; Xie et al, 2004). It can measure starch gelatinization temperature, the heat energy input required for gelatinization, and the degree of starch gelatinization. Chinnaswamy et al (1989) also reported the use of scanning electron microscopy (SEM) to study the microstructural changes, X-ray diffraction to determine the pattern of crystalline form, and SEC to discover macromolecular degradation of extruded starch. Additionally, expansion ratio, bulk density, and texture profile analysis (TPA) are generally used to determine physical properties of extruded products (Liu et al, 2000; Li et al, 2005). However, none of these methods are fast or allow on-line evaluation.

3.2 Use of NIR Spectroscopy for Predicting Degree of Cook

NIRS, as mentioned previously, is applicable to the study of chemical changes involving the O–H bond in different states of hydrogen bonding (Osborne et al., 1993). As the disruption of the hydrogen bond network within each starch granule happens, it allows water molecules to occupy new spaces within amylopectin molecules. These changes can be detected by NIRS because of differences in the spectra attributed to O–H stretching motions of bonded and free water, respectively (Osborne, 2007).

To date, a small amount of information has been published on the application of NIRS for monitoring starch degradation during the extrusion cooking process. Some applications of discriminant analysis have been recently investigated. Ben-Hdech et al (1993) applied NIRS to evaluate the intensity of extrusion cooking of pea flour, and successfully classified extrudates based on the degree of severity of extrusion cooking. Guy et al (1996) studied the feasibility of using NIRS to measure structural changes of starch in extruded wheat products. They found that the spectral data can be used to predict specific mechanical energy (SME), especially for whole meal. They also stated that the measurements are thought to be related to changes in the hydrogen bonding of the hydroxyl groups in the starch molecules. Fiber-optic probes have been further used with NIRS to measure degree of cook on-line and in-line during extrusion processing. Evans et al (1999) proved that on-line measurements using transmittance fiber optic probes can be used to track changes in degree of cook during extrusion processing of wheat flour. They showed that the second derivative of the minimum absorbance in the region of 1400 to 1450 nm of the extrudate melts can be used to follow changes in degree of cook when three process variables were changed in a factorial design. They found that the intensity of the minimum absorbance peak increased as the corresponding viscosity of

the powdered extrudates decreased. Apruzzese et al (2000) investigated in-line color and composition changes in an extruder MMduring the extrusion of yellow corn flour. They showed that differences in the peak intensities of NIR spectra of corn flour during extrusion cooking at different screw speeds could be observed at wavelengths of 2100 and 2280 nm. These wavelengths are associated with amylopectin molecules being broken down during the extrusion process. Sahni et al (2004) later used fiber-optic transmittance probes for in-line monitoring and prediction of critical parameters in emulsion-based products. They reported that NIR analysis contains information related to both the input parameters (raw materials and process variables) and the final product quality (viscosity). Furthermore, they successfully classified samples with different recipes and process variables using PCA.

References

American Association of Cereal Chemists. 2001. Method 28-21: X-ray method; Methods 28-22 and 28-51: Cracking and flotation method; Method 28-41A: Acid hydrolysis test, The Association: St. Paul, MN.

American Association of Cereal Chemists. 1995. Approved methods of the AACC, 9 th ed. Method 28-21, The Association: St. Paul, MN.

Apt, A.C. 1952. A rapid method for examining wheat samples for infestation. Northwest. Miller (Milling prod. Sect.) 247(73): 24.

Apruzzese, F., Balke, S.T., and Diosady, L.L. 2000. In-line color and composition monitoring in the extrusion cooking process. Food Res. Int. 33:621-628.

Baer, R.J., Frank, J.F., and Loewenstein, M. 1983. Compositional analysis of whey powders using near infrared reflectance spectroscopy. J. Food Sci. 48:959-961.

Baik, B-K., Powers, J., and Nguyen, L.T. 2004. Extrusion of regular and waxy barley flours for production of expanded cereals. Cereal Chem. 81:94-99.

Baker, J.E., Dowell, F.E., and Throne, J.E. 1999. Detection of parasitized rice weevils in wheat kernels with near-infrared spectroscopy. Biol. Control 16:88-90.

Bao, J.S., Cai, Y.Z., and Corke, H. 2001. Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy. J. Food Sci. 66:936-939.

Becker, A., Hill, S.E., and Mitchell, J.R. 2001. Milling-a further parameter

affecting the rapid visco analyzer (RVA) profile. Cereal Chem. 78:166-172.

Ben-Hdech, H., Gallant, D.J., Robert, P., and Gueguen, J. 1993. Use of near infrared spectroscopy to evaluate the intensity of extrusion-cooking processing of pea flour. Int. J. Food Sci. 28:1-12.

Brader, B., Lee, C.R., Plarre, R., Burkholder, W., Kitto, G.B., Kao, C., Polston, L., Dorneanu, E., Szabo, I., Mead, B., Rouse, B., Sullins, D., and Denning, R. 2002. A comparison of screening methods for insect contamination in wheat. J. Stored Prod. Res. 38:75-86.

Campbell, M.R., Mannis, S.R., Port, H.A., Zimmerman, A.M., and Glover, D.V. 1999. Prediction of starch amylose content versus total grain amylose content in corn by near-infrared transmittance spectroscopy. Cereal Chem. 76:552-557.

Cartagena, M., Mozayeni, F., and Szajer, G. 1994. Chemometric applications of NIR spectroscopy. INFORM 5:1146-1147.

Chambers, J., N.J. McKevitt, and M.R. Stubbs. 1984. Nuclear magnetic resonance spectroscopy for studying the development and detection of the grain weevil, *Sitophilus granarius*, within wheat kernels. Bull. Entomol. Res. 74:707-724.

Chen, Y.R., Delwiche, S.R., and Hruschka, W.R. 1995. Classification of hard red wheat by feed forward backpropagation neural networks. Cereal Chem. 72:217-232.

Chen, W., and Kitto, G.B. 1993. Species-specific immunoassay for Sitophilus granarius in wheat. Food Agric. Immunol. 5:165-175.

Chinnaswamy, R., Hanna, M.A., and Zobel, H.F. 1989. Microstructural,

physiochemical, and macromolecular changes in extrusion-cooked and retrograded corn starch. Cereal Foods World 34:415-422.

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., Bean, M.M., Miller, R.E., Webb, B.D., and Williams, P.C. 1995. Apparent amylose content of milled rice by near-infrared reflectance spectroscopy. Cereal Chem. 72:182-187.

Delwiche, S.R., Graybosch, R.A., Hansen, L.E., Souza, E., and Dowell, F.E. 2006. Single kernel near-infrared analysis of tetraploid (durum) wheat for classification of the waxy condition. Cereal Chem. 83:287-292.

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

Delwiche, S.R., McKenzie, K.S., and Webb, B.D. 1996. Quality characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. Cereal Chem. 73:257-263.

Delwiche, S.R., and Norris, K.H. 1993. Classification of hard red winter wheat by near-infrared diffuse reflectance spectroscopy. Cereal Chem. 70:29-33..

Dennis, N.M., and R.W. Decker. 1962. A method and machine for detecting living internal insect infestation in wheat. J. Econ. Entomol. 55: 199-203.

Dowell, F.E. 2000. Differentiating vitreous and nonvitreous durum wheat kernels by using near-infrared spectroscopy. Cereal Chem. 77:155-158.

Dowell, F.E., Throne, J.E., and Baker, J.E. 1999. Identifying stored-grain insects using near-infrared spectroscopy. J. Econ. Entomol. 92:165-169.

Dowell, F.E., Throne, J.E., and Baker, J.E. 1998. Automated Nondestructive Detection of Internal Insect Infestation of Wheat Kernels by Using Near-Infrared Reflectance Spectroscopy. J. Econ. Entomol. 91:899-904.

Drennen, J.K., Gebhart, B., Kraemer, G.E., and Lodder, R.A. 1990. Near-infrared spectrometric determination of hydrogen ion, glucose, and human serum albumin in a simulated biological matrix. Spectroscopy 6:28-32.

Drzewiecki, T.M., and Shuman D. 2001. Acousto-fluidic detection of insect larvae in grain. Acta. Hort. (ISHS) 562:233-241.

Dunmire, D.L., and Williams, R.C. 1990. Automated qualitative and quantitative NIR reflectance analyses. Cereal Foods World 35:913-918.

Evans, A.J., Huang, S., Osborne, B.G., Kotwal, Z., and Wesley, I.J. 1999. Near infrared on-line measurement of degree of cook in extrusion processing of wheat flour. J. Near Infrared Spectrosc. 7:77-84.

FDA. 1998. Defect action level handbook. Washington, D.C.: Food and Drug Administration, Center for Safety and Nutrition.

Flinn, P.W., Hagstrum, D.W., Reed, C., and Phillips, T.W. 2003. USDA-ARS Stored-grain areawide integrated pest management program. Pest Management Sci. 59:614-618.

Fornal, J., Jelinski, T., Sadowska, S. Grund as, J. Nawrot, A. Niewiada, J. R. Warchalewski, and W. Blaszczak. 2007. Detection of granary weevil Sitophilus granarius (L.) eggs and internal stages in wheat grain using soft X-ray and image analysis. J. Stored Prod Res . 43:142-148.

Federal Grain Inspection Service (FGIS). 1997. Ch.3 In Book II: Grain Grading

Procedures, Washington, D.C.:USDA-GIPSA-FGIS.

Ganjyal, G., Hanna, M.A., Supprung, P., Noomhorm, A., and Jones, D. 2006. Modeling selected properties of extruded rice flour and rice starch by neural networks and statistics. Cereal Chem. 83:223-227.

Ghaedian, A.R. and R.L. Wehling. 1997. Discrimination of sound and granary-weevil-larva-infested wheat kernels by near-infrared diffuse reflectance spectroscopy. J. Assoc. Off. Agric. Chem. 80: 997-1005.

Ghaedian, A.R. and R.L. Wehling. 1996. Distribution of uric acid in the fractions obtained from experimental milling of wheat infested with granary weevil larvae. Cereal Chem. 73: 628-631.

Goossens, H.J. 1949. A method for staining insect egg plugs in wheat. Cereal Chem. 26:419-420.

Gradenecker, F. 2003. NIR on-line testing in grain milling. Cereal Foods World. 48:18-19.

Guy, R.C.E., Osborne, B.G., and Robert, P. 1996. The application of near-infrared reflectance spectroscopy to measure the degree of processing in extrusion cooking processes. J. Food Eng. 27: 241-258.

Hackman, R.H. and M.A. Goldberg. 1981. A method for determination of microgram amounts of chitin in arthropod cuticles. Anal. Biochem. 110: 277.

Hagstrum, D.W., P.W. Flinn, and D. Shuman. 1996. Automated monitoring using acoustical sensors for insects in farm-stored wheat. J. Econ. Entomol. 89:211-217.

Harper, J.M. 1981. Extrusion of Foods, Vol. I. CRC Press, Inc., Boca Raton, FL.

Harris, K.L., J.F. Nicholson, L.K. Randolph, and J.L. Trawick. 1952. An investigation of insect and rodent contamination of wheat and wheat flour. J. Assoc. Off. Agric. Chem. 33:115-118.

Hashimoto, J.M., and Grossmann, M.V.E. 2003. Effects of extrusion conditions on quality of cassava bran/cassava starch extrudates. Int. J. Food Sci. Tech. 38:511-517.

Jackson, D.S., Gomez, M.H., Waniska, R.D., and Rooney, L.W. 1990. Effect of single-screw extrusion cooking on starch as measured by aqueous high-performance size-exclusion chromatography. Cereal Chem. 67: 529-532.

Johnson, R.A., and Wichern, D.W. 1998. Discrimination and classification. Ch.11 In: Applied Multivariate Statistical Analysis, 4th ed. A Simon and Schuster Company, Eaglewood Cliffs, NJ, 629-725 pp.

Karunakaran, C., Jayas, D.S., and White, N.D.G. 2003. Soft X-ray inspection of wheat kernels infested by Sitophilus oryzae. Trans. ASAE 46:739-745.

Karunakaran, C., Jayas, D. S., and White, N.D.G. 2004. Detection of internal wheat seed infestation by Rhyzopertha dominica using X-ray imaging. J. Stored Prod. Res. 40(5): 507-516.

Keagy, P.M. and Schatzki, T.F. 1993. Machine recognition of weevil damage in wheat radiographs. Cereal Chem. 70: 696-700.

Kitto, G.B. 1991. A new rapid biochemical technique for quantitating insect infestation in grain. Assoc. Oper. Millers Bull. 5835-5838.

Lamkin, W.M., N.C. Unruh, and Y. Pomeranz. 1991. Use of fluorometry for the

determination of uric acid in grain. Elimination of interfering fluorescence. Cereal Chem. 68: 81-86.

Lanza, E. 1983. Determination of moisture, protein, fat, and calories in raw pork and beef by near-infrared spectroscopy. J. Food Sci. 48: 471-474.

Li, S.Q, Zhang, H.Q., Jin, Z.T., and Hsieh, F-H. 2005. Textural modification of soya bean/corn extrudates as affected by moisture content, screw speed and soya bean concentration. Int. J. Food Sci. Tech. 40:731-741.

Liu, Y., Hsieh, F, Heymann, H., and Huff, H.E. 2000. Effect of process conditions on the physical and sensory properties of extruded oat-corn puff. J. Food Sci. 65:1253-1259.

Liu, X., and Han, L. 2006. Prediction of chemical parameters in maize silage by near infrared reflectance spectroscopy. J. Near Infrared Spectrosc. 14:333-339.

Maghirang, E.B., Dowell, F.E., Baker, J.E., and Throne, J.E. 2003. Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy. Trans. ASAE 46:1277-1282.

Mankin, R.W., D. Shuman, and J.A. Coffelt. 1996. Noise shielding of acoustic devices for insect detection. J. Econ. Entomol. 89: 1301-1308.

Mark, H. 2001. Qualitative near-infrared analysis. Ch.13 In: Near-Infrared Technology in the Agriculture and Food Industries, 2nd ed. Williams P.C. and Norris K., Eds., American association of Cereal Chemists, Inc., MN. 233-238 pp.

Mark, H., and Tunnell, D. 1985. Qualitative near-infrared reflectance analysis

using Mahalanobis distances. Anal. Chem. 57: 1449-1456.

Mason, W.R., and Hoseney, R.C. 1986. Factors affecting the viscosity of extrusion-cooked wheat starch. Cereal Chem. 63: 436-441.

Meadows, F., and Barton, F.E., II. 2002. Determination of rapid visco analyser parameters in rice by near-infrared spectroscopy. Cereal Chem. 79:563-566.

Meullenet, J.F., Mauromoustakos, A., Horner, T.B., and Marks, B.P. 2002. Prediction of texture of cooked white rice by near-infrared reflectance analysis of whole-grain milled samples. Cereal Chem. 79:52-57.

Micklander, E., Kjeldahl, K., Egebo, M., and Norgaard, L. 2006. Multi-product calibration models of near infrared spectra of foods. J. Near Infrared Spectrosc. 14:395-402.

Murray, I., and P.C. Williams. 1990. Chemical principle of near-infrared technology. In: Near-infrared technology in the agricultural and food industries. P.C. Williams and K.H. Norris (Eds.), American Association of Cereal Chemists. St. Paul, MN, 17-34 pp.

Nicholson, J.F., O.L. Kurtz, and K.L. Harris. 1953. An evaluation of five procedures for the determination of internal insect infestation of wheat. IV. Visual examination for insect exit holes. J. Assoc. Off. Agric. Chem. 36:146-150.

Osborne, B.G. 2007. Flours and breads. Ch. 8.1. In: Near-Infrared Spectroscopy in Food Science and Technology. Ozaki Y., McClure W.F., and Christy A.A., eds., John Wiley & Sons, Inc., NJ, 281-296 pp.

Osborne, B.G. 2006. Review: Applications of near infrared spectroscopy in quality

sceening of early-generation material in cereal breeding programmes. J. Near Infrared Spectrosc. 14:93-101.

Osborne, B.G. 2000. Recent developments in NIR analysis of grains and grain products. Cereal foods World 11-15.

Osborne, B.G., Fearn, T., and Hindle, P.H. 1993. Near infrared calibration II. Ch.7 In: Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, 2nd ed. Longman Scientific & Technical, UK, 121-144 pp.

Pachla, L.A., and Kissinger, P.T. 1977. Monitoring insect infestation in cereal products-Determination of traces of uric acid by high pressure liquid chromatography. Anal. Chem, Acta 88:385.

Paliwal, J., Wang, W., Symons, S.J., and Karunakaran, C. 2004. Insect species and infestation level determination in stored wheat using near-infrared spectroscopy. Can. Bios. Eng. 46:7.17-7.24.

Pasikatan, M.C. and Dowell, F.E. 2004. High-speed NIR segregation of high- and low-protein single wheat seeds. Cereal Chem. 81:145-150.

Pearson, T.C., Brabec, D.L., and Schwartz, C.R. 2003. Automated detection of internal insect infestations in whole wheat kernels using a PERTEN SKCS 4100. Appl. Eng. Agric. 19:727-733.

Pedersen, J.R. 1992. Insects: identification, damage and detection. Storage of Cereal Grains and Their Products. American Association Cereal Chemists, St Paul, MN, 435-489 pp.

Perez-Mendoza, J., Throne, J.E., Dowell, F.E., Maghirang, E.B., and Baker, J.E. 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.

Perez-Mendoza, J., Throne, J.E., Dowell, F.E., and Baker, J.E. 2003. Detection of insect fragments in wheat flour by near-infrared spectroscopy. J. Stored Prod. Res. 39:305-312.

Ridgway, C., Chambers, J., and Cowe, I.A. 1999. Detection of grain weevils inside single wheat kernels by a very near-infrared two-wavelength model. J. Near Infrared Spectosc. 7:213-221.

Ridgway, C. and Chambers, J. 1998. Detection of insects inside wheat kernels by NIR imaging. J. Near Infrared Spectrosc. 6:115-119.

Ridgway, C. and Chambers, J. 1996. Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. J. Sci. Food Agric. 71:251-264.

Rodriquez-Otero, J.L., Hermida, M., and Cepeda, A. 1995. Determination of fat, protein, and total solids in cheese by near infrared reflectance spectroscopy. J. Assoc. Off. Agric. Chem. 78:802-806.

Rotundo, G., Germinara, G.S., and De Cristofaro, A. 2000. Immuno-osmophoretic technique for detecting Sitophilus granarius (L.) infestations in wheat. J. Stored Products Res. 36:153-160.

Russell, G.E. 1988. Evaluation of four analytical methods to detect weevils in wheat: Granary weevil, *Sitophilus granarius* (L.), in soft white wheat. J. Food Protect. 51:547-553.

Sahni, N.S., Isaksson, T., and Naes, T. 2004. In-line near infrared spectroscopy for

use in product and process monitoring in food industry. J. Near Infrared Spectrosc. 12:77-83.

Schatzki, T.F. and Fine, T.A.B. 1988. Analysis of radiograms of wheat kernels for control. Cereal Chem. 65:233-239.

Sevatson, E. and Huber, G.R. 2000. Extruders in the food industry. Ch.9 In: Extruders in Food Applications. M. Riaz (Ed.), Technomic Publishing Company, Inc., P.A. 167-204 pp.

Shadow, W. and Carrasco, A. 2000. Practical single-kernel NIR/visible analysis for small grains. Cereal Foods World 45:16-18.

Shah, N.K. and Gemperline, P.J. 1990. Combination of thee Mahalanobis distance and residual variance pattern recognition techniques for classification of near-infrared reflectance spectra. Anal. Chem. 62:465-469.

Shuman, D., J.A. Coffelt, K.W. Vick, and R.W. Mankin. 1993. Quantitative acoustical detection of larvae feeding inside kernels of grain. J. Econ. Entomol. 86:933-938.

Street, M.W. and W.A. Bruce. 1976. Carbon dioxide analyzer detects insects hidden in foods. Food Eng. 48:94-95.

Thygesen, L.G., Engelsen, S.B., Madsen, M.H., and Sorensen, O.B. 2001. NIR spectroscopy and partial least squares regression for the determination of phosphate content and viscosity behavior of potato starch. J. Near Infrared Spectrosc. 9:133-139.

Tigabu, M. And Oden, P.C. 2002. Multivariate classification of sound and insect-infested seeds of a tropical multipurpose tree, *Cordia africana*, with near

infrared reflectance spectroscopy. J. Near Infrared Spectrosc. 10:45-51.

Toews, M.D., Perez-Mendoza, J., Throne, J.E., Dowell, F.E., Maghirang, E., Arthur, F.H., Campbell, J.F. 2007. Rapid Assessment of Insect Fragments in Flour Milled from Wheat Infested with Known Densities of Immature and Adult Sitophilus oryzae (Coleoptera: Curculionidae). J. Econ. Entomol. 100:1714-1723.

U.S. Department of Agriculture, Agricultural Research Service. 1986. Stored grain insects. Agricultural Handbook No. 500, National Technical Information Service, Springfield, VA.

U.S. Department of Agriculture, Federal Grain Inspection Service. 1987. Insect infestation in grain. Federal Register 52(125):24432-24438. Food and Drug Administration, Center for Food Safety and Nutrition. The Food Defect Action Levels, Washington, DC.

Villareal, C.P., De La Cruz, N.M., and Juliano, B.O. Rice amylose analysis by near-infrared transmittance spectroscopy. Cereal Chem. 71:292-296.

Vick, K.W., J.C. Webb, B.A. Weaver, and C. Litzkow. 1988. Sound detection of stored product insects that feed inside kernels of grain. J. Econ. Entomol. 81: 1489-1493.

Wang, D., Dowell, F.E., and Dempster, R. 2002. Determining vitreous subclasses of hard red spring wheat using visible/near-infrared spectroscopy. Cereal Chem. 79:418-422.

Wehling, R.L. 1998. Infrared Spectroscopy. Ch.27 In: Food Analysis, 2nd ed. Nielson, S.S. (Ed.), Aspen Publishers, Inc., Gaithersburg, MD, 413-424 pp.

Wehling, R.L., Jackson, D.S., and Hamaker, B.R. 1996. Prediction of corn dry-milling quality by near-infrared spectroscopy. Cereal Chem. 73:543-546.

Wehling, R.L., Jackson, D.S., Hooper, D.G., and Ghaedian, A.R. 1993. Prediction of wet-milling starch yield from corn by near-infrared spectroscopy. Cereal Chem. 70:720-723.

Wehling, R.L. and Wetzel, D.L. 1983. High performance liquid chrmatography determination of low level uric acid in grains and cereal products as a measure of insect infestation. J. Chromatogr. 269:191-197.

Wesley, I.J., Uthayakumaran, S., Anderssen, R.S., Cornish, G.B., Bekes, F., Osborne, B.G., and Skerritt, J.H. 1999. A curve fitting approach to the measurement of functional proteins in wheat flour. J. Near Infrared Spectrosc. 7: 229-233.

Whalen, P.J., Bason, M.L., Booth, R.I., Walker, C.E., and Williams, P.J. 1997. Measurement of extrusion effects by viscosity profile using the rapid viscoanalyzer. Cereal Foods World 42: 469-475.

Workman, J. Jr., and Shenk, J. 2004. In: Near-Infrared Spectroscopy in Agriculture. Roberts C.A., Workman, J., Jr., and Reeves III, J.B., American Society of Agronomy, Inc., Crop Science society of America, Inc., and Soil Science Society of America, Inc., WI., 3-10 pp.

Xie, F., Dowell, F.E., and Sun, X.S. 2004. Using visible and near-infrared reflectance spectroscopy and differential scanning calorimetry to study starch, protein, and temperature effects on bread staling. Cereal Chem. 81:249-254.

Xie, F., Dowell, F.E., and Sun, X.S. 2003. Comparison of near-infrared reflectance spectroscopy and texture analyzer for measuring wheat bread changes in storage. Cereal Chem. 80:25-29.

RESEARCH TOPICS

Chapter 1

Comparison of Selected Discrete Wavelength and Principal Component Analysis Methods for Insect Infestation Detection

Abstract

Discriminant analysis using Mahalanobis distances based on selected discrete wavelengths was investigated as to whether it can provide comparable results to correctly classify sound and infested wheat kernels containing late instar granary weevil larvae, compared to using principal component analysis (PCA). Based on PCA, full or partial near-infrared (NIR) spectra from sound kernels were used to construct calibration models by the calculation of Mahalanobis distances from principal component scores. A five factor PCA model of second derivative spectra, coupled with the use of standard normal variate (SNV) over a spectral range of 1100 - 1900 nm, gave the best results overall. Correct classification rates were 100% of sound, 93% of infested, 95% of sound air-dried, 91% of infested air-dried, and 92% of sound kernels from six different wheat varieties. When discriminant analysis was applied to selected discrete wavelengths, calibrations based on the calculation of Mahalanobis distances were developed from both sound and infested wheat kernels using the NIR spectral region from 1100 to 1900 nm. The best results were achieved with a model based on four selected wavelengths at 1120, 1130, 1280 and 1860 nm. Correct classification

rates were 98% of sound, 98% of infested, 100% of sound air-dried, 96% of infested air-dried, and 100% of sound kernels from six different wheat varieties. These results indicated that the method based on selected wavelengths provided a more robust calibration than did the PCA method.

Introduction

Internally infesting insects are generally considered the most damaging of the stored-grain insects, since damage can occur with very little or no visible indication. After oviposition, females of these hidden insects secrete a gelatinous substance to glue the hole, usually flush with the seed coat (Pedersen, 1992). The hidden infestation is not easily detected, nor is it easily removed in processing. Detecting the presence of insects in grain and grain products is thus an important factor in the integrated approach to stored grain insect control.

NIR spectroscopy has shown some applications for detecting insect infestation. Previous research, including work in our laboratory (Ghaedian and Wehling, 1997), and at the USDA laboratories (Dowell et al, 1998 and 1999; Maghirang et al, 2003) in Manhattan, KS, has shown that full spectrum NIR methods can successfully detect larvae inside wheat kernels. Qualitative NIR reflectance analysis, using Mahalanobis distances based on selected discrete wavelengths, is a simple technique and requires less sophisticated instrumentation than do full spectrum techniques. The purpose of this research was to develop a protocol for detecting internal infestation that requires measurements to be taken at only a few discrete wavelengths.

Materials and Methods

Preparation of Sound and Infested Wheat Kernels

Preliminary work was done by comparing the results of discriminant functions using NIR spectra of sound, infested, air-dried, and 6 different varieties of sound wheat kernels

from a previous study in our laboratory (Ghaedian 1995). Sound and infested kernels containing 4-week-old instar larvae of granary weevil were taken from a mason jar which was maintained at 27 ∀ 3°C and 65 ∀ 5% relative humidity for four weeks. To identify infested wheat kernels, a radiograph was obtained using a General Electric X-ray Grain Inspection unit using 20 kV at 5 mA and a 2.5 min exposure time (Ghaedian 1995). Sound wheat kernels were also picked from the same radiograph. Infested and sound air-dried kernels were obtained by freezing wheat kernels for two days to kill the larvae, followed by drying at room temperature.

NIR Analysis and Data Collection

Spectral data were collected with an NIRSystems model 6500 spectrometer (NIRSystems Division of Foss Instrument, Silver Spring, MD) from 1100 to 2498 nm. Spectra were collected in the form of log (1/R) at a wavelength interval of 2 nm. Thirty two monochromator scans were averaged from each kernel. Steps for collecting NIR spectra have been previously described in more detail (Ghaedian 1995). For data collection, the Near Infrared Spectral Analysis Software (NSAS) package (version 3.16, NIRSystems) was used.

Data Analysis

Two methods were used to develop calibration models:

1) Discriminant analysis based on the loadings derived from Principal Component Analysis (PCA) of full or partial NIR spectra using PLSplus/IQ software (Galactic Industries, Salem, NH). Spectra from NSAS were imported into PLSplus/IQ software for analysis.
2) Discriminant analysis based on Mahalanobis distances of NIR reflectance data from

selected discrete wavelengths using MultiQual software (Near Infrared Research Corp., Suffern, NY). Unlike PLSplus/IQ software, spectra from NSAS were converted into ASCII (JCAMP-DX) format before analyzing data with the MultiQual software.

Development of Calibration Models Based on Principal Component Analysis

Grams/32 software (version 3.03, PLSplus/IQ, Galactic Industries, Salem, NH) was used for data analysis. This technique combines PCA and Mahalanobis distances into a single method. As mentioned earlier in the literature review section, it uses principal component scores from spectra of samples in a training set to calculate the Mahalanobis matrices, and discriminant models are then constructed. Generally, samples with Mahalanobis distances less than 3 are considered to be members of the same group as those used to develop the model. Kernels with Mahalanobis distances greater than 3 are then considered to be non-members. In this trial, a calibration set containing sound kernels was used; therefore samples having a Mahalanobis distance of less than 3 standard deviations from the training set center were classified as sound, while those with a Mahalanobis distance of 3 or greater standard deviations were classified as infested. Effectiveness of the model was evaluated by the percent correct classification of sound and infested kernels in the validation sets.

According to the PLSplus/IQ manual, several options can be applied to the data set including selection of the wavelength range to be used and the number of factors to be calculated. Fifteen factors are normally the maximum number that is recommended for use in the analysis. Models that contain excessive numbers of factors result in overfitting or over-discrimination, and may include noise vectors or vectors that are not necessary.

Additionally, a number of data pretreatments, called preprocessing algorithms, which may enhance the accuracy of the final calibration model can be applied in this software. Mean centering improves mathematical accuracy of the spectral decomposition and correlation by calculating the average spectrum of all spectra in the training set, and then subtracting the result from each spectrum. Generally speaking, mean centering enhances the subtle differences between the spectra. Variance scaling is used to give the data equal weighting, such as when analyzing low concentration constituents that have spectral bands that overlap those of higher concentration constituents. It is calculated by dividing the response at each spectral data point by the standard deviation of the response of all training set spectra at that point.

Furthermore, multiplicative scatter correction (MSC), and standard normal variate (SNV) transformation can be used for pathlength correction and removing the effects of light scattering (PLSplus/IQ manual). The use of MSC is suggested for data sets where the spectral variation between training samples is small. It corrects indeterminate pathlength effects resulting from scattering when the diffuse reflectance technique is applied, by adjusting the slope and baseline offset of each spectrum to the Aideal@ average spectrum. SNV transformation can be applied to the spectra alone or with detrending. Unlike MSC, no ideal spectrum is required for SNV. The scattering is removed by normalizing each spectrum by the standard deviation of the responses across the entire spectral range. Detrending can be applied after the SNV correction. A linear least squares regression is used to fit a second order polynomial to the SNV corrected spectrum. This curve is then subtracted from the spectrum to give the result. Effects due to changes in the physical

properties of the samples are then removed. Ridgway and Chambers (1996) stated that SNV transformation generally removes the effects of particle size and scatter; whereas, detrending corrects for variation in baseline shift and curvelinearity.

Due to a number of problems including detector drift, changing environmental conditions such as temperature, sampling accessories, etc., the baseline of a given spectrum can be changed. Therefore, derivatization (1^{st} or 2^{nd}) of the spectra is used for removing baseline effects. The first derivative is a measure of the slope of the spectral curve at every point. The second derivative is a measure of the change in the slope of the curve. They both are effective methods for removing baseline offsets because the slope of the curve is not affected by baseline shifts in the spectrum. The algorithms of Savitsky-Golay (SG) and gap methods are used for calculating derivatives. The size of the spectral segment used for calculation of the derivative can be selected.

Data Pre-treatment

The entire spectral region from 1100-2498 nm was used to develop calibration models. Mean centering was applied to all calibration spectra. Variance scaling, MSC, SNV (alone and with detrending), and derivatization were also tested to investigate if they yielded better classification rates. Using different numbers of PCA factors was also tried in an attempt to achieve more robust classification of samples.

Calibration Set

In this trial, a calibration set for developing discriminant analysis models contained only the spectra of sound kernels (n=50).

Validation Set

Once calibration models were built, they were tested with several validation sets including sound (n=25), infested (n=75), sound air-dried (n=75), infested air-dried (n=75), and six different varieties of sound wheat kernels (n=60).

Development of Calibration Models Based on Selected Wavelengths

In MultiQual software, instead of using full or partial spectra, the best possible combinations algorithm is used to select a proper number of wavelengths that prove most effective for distinguishing between the materials based on the spectroscopic data, according to the computation of Mahalanobis distances. Several approaches (e.g., forcing wavelengths, skipping wavelengths, interleaving wavelengths, etc.) are permitted in this program, allowing the incorporation of such prior information to improve the chemometric model, reduce the amount of computation required, and also to minimize the computation time. According to the MultiQual software user=s guide, the Mahalanobis distances between each pair of materials are the most important result output by the program. If two materials have a small distance between them, it means that the computer could not find wavelengths where the spectra were appreciably different and misclassification may occur. Determination of outliers is a criterion of developing a Agood@ model. This can be recognized from the classification output of the calibration set, since the computer rereads the calibration data file, and uses the model to calculate the Mahalanobis distance of each sample from each of the calibration sets. Generally, once a model is built, the calibration program then tests the model using a leave-one-out algorithm. Using the same wavelengths found from the search, each sample is successively deleted from the model. A model is then created from the

remaining data and applied to the spectral data from the deleted sample to verify the classification capabilities of the model using the selected wavelengths (MultiQual software user=s guide).

Data Processing

In this approach, log (1/R) values from NIR reflectance spectral data of sound and infested kernels were used directly in discriminant analysis. Results from the preliminary work using Galactic software show that the correct classification rate was slightly increased when using the spectral region from 1100-1900 nm. Ghaedian and Wehling (1997) also indicated that the spectral region from 1100-1900 nm contains more information for classification than does the region from 1901-2498 nm. Delwiche et al (1996) also attempted to use the lower half of the near-infrared region (1120-1800 nm). They claimed that, in this wavelength region, overtone frequencies of OH, CH, and NH occur and the reflected energy signal is relatively strong. Furthermore, they surmised that using a broader region from1120 to 2478 nm to develop models resulted in lower accuracy because of weakness in signal and nonlinear response at longer wavelengths. Therefore, in this trial, spectral data from 1901-2498 nm was removed from the calibration (training) set, leaving a wavelength range from 1100 to 1900 nm for developing models. Different combinations and numbers of wavelengths were then tested with separate validation (prediction) sets to select the best model that provides the highest correct classification rate.

Calibration Set

Unlike the PLSPlus/IQ software, the calibration set in this test was developed from both sound and infested wheat kernels. Twenty spectra from each set of sound, infested,

sound air-dried, and infested air-dried kernels were randomly selected to form a calibration set (n=80).

Validation Sets

Each model was used to predict samples in separate validation sets which are sound (n=55), infested (n=55), sound air-dried (n=55), infested air-dried (n=55), and six different varieties of sound wheat kernels (n=60).

Results and Discussion

Classification Results from Calibration Models Based on PCA

Table 1 shows the classification results using Mahalanobis distances based on PCA of the entire spectral region from 1,100 to 2,498 nm. Correct classification refers to the percentage of spectra of sound kernels in a validation set matched with spectra in a calibration set, or the percentage of spectra of infested kernels in a validation set non-matched with spectra in a calibration set. The numbers of PCA factors used in constructing the models are given and the results of using only mean centering show that a discriminant model with seven factors gave the best correct classification rate for both sound and infested kernels, including sound kernels from six different wheat varieties. The number of factors recommended by the software based only on calibration data (eight factors) gave poor classification rates for six different varieties of sound wheat kernels. This may be due to the use of excessive factors to produce models that are overfitted to the calibration set, resulting in over discrimination. Furthermore, sound air-dried kernels were not classified as sound kernels in the validation set.

Table 1: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra (log 1/R) from 1100 to 2498 nm using different pre-treatments

Pre-treatment method	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
Mean centering	7	100	83	32	97	82
	8	96	96	0	100	47
	9	96	97	0	100	42
Variance Scaling	4	100	45	91	84	93
	5	100	55	92	92	70
	6	100	79	56	97	85
	7	100	88	25	99	80
MSC [a]	5	100	88	39	97	75
MSC	6	96	95	1	100	48
MSC	7	96	95	0	100	55
MSC	8	92	96	0	100	43
SNV-detrend [b]	6	96	84	61	99	77
SNV-alone [c]	7	100	95	0	100	43
SNV-detrend	8	96	97	0	100	32

[a] Multiplicative scatter correction.

[b] Standard normal variate transformation with detrending.

[c] Standard normal variate transformation alone.

Calibrations were subsequently developed using variance scaling, MSC, and SNV with and without detrending. Different numbers of factors were also used in developing calibrations. The classification results were similar to those from calibrations using only mean centering. Variance scaling with fewer number of factors (i.e., four factor model and five factor model), however, improved the prediction rate of sound air-dried kernels. Unfortunately, it substantially lowered the prediction rate of infested kernels. Models developed from MSC provided lower correct classification rates compared with those developed from variance scaling. A fewer number of factors (six factor model vs eight factor model) could be used in the models when detrending was coupled with SNV.

Calibrations were also developed by using first and second derivative spectra. Both Savitsky-Golay (SG) and gap methods with two different gap sizes, i.e. five and nine, were used for calculation of the derivative points. The prediction results are given in Table 2. The percentage of sound air-dried kernels correctly classified was increased when first and second derivative transformations were applied to the spectra. Fewer factors were needed for use in developing calibrations. The calibration model developed by Savitsky-Golay (SG) first derivative with the size of the spectral segment of nine and four factors gave the highest percentage correct classification over all validation sets.

Table 2: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 2498 nm using derivative treatment

Derivative treatment	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
SG 1st-der [a] (5)[b]	3	100	24	100	39	82
SG 1st-der (9)	4	100	79	95	92	72
Gap 1st-der [c] (5)	6	96	93	7	100	58
Gap 1st-der (9)	7	96	95	3	100	45
SG 2nd-der [d] (5)	3	100	27	100	23	93
SG 2nd-der (9)	3	100	27	100	23	93
Gap 2nd-der [e] (5)	3	100	47	100	40	85
Gap 2nd-der (9)	5	96	89	67	95	75

[a] Savitsky-Golay first derivative.

[b] ()- Number of data point used in the derivative function.

[c] Gap first derivative.

[d] Savitsky-Golay second derivative.

[e] Gap second derivative.

Ghaedian (1995) indicated that moisture content of wheat kernels had an effect on prediction rate, since the results showed very poor classification rates for sound kernels which had been air-dried when using mean centering and MSC to develop the calibration models. In addition, Ridgway and Chambers (1996) stated that kernels with internal larvae

are typically higher in average moisture content than kernels without larvae, because of water present in the larvae itself and because insect respiration indirectly increases moisture content in those kernels. However, kernel moisture does not always specify infestation because kernels with low moisture content that contain larvae could absorb radiation similar to kernels with no larvae that have a high moisture content. To eliminate this problem, we excluded wavelengths between 1901-1979 nm when developing calibrations. Most wavelengths at which water absorbs radiation were thus excluded (Murray and Williams 1990). Classification results after removing the interference of water bands are reported in Table 3. As anticipated, the correct classification rate of air-dried sound kernels was improved by using either mean centering or variance scaling, except for the eight factor model which still gave poor performance. A seven factor model seemed to provide satisfactory prediction results for all validation sets. By using MSC and SNV, similar results were obtained with the use of fewer factors. Models developed after removing the water bands from derivatized spectra of SG second derivative and Gap first derivative gave comparable results to the full spectrum models (data not shown). Calibration models developed from SG first derivative spectra with segment size of nine and seven factors, and Gap second derivative models with gap size of nine and six factors provided high correct classification rates for all validation sets (Table 4). Calibration models with higher numbers of segments were tested for both derivative methods, however, the percentage of correct classification was not increased (data not shown).

Table 3. Percentage of correct classification of wheat samples based on principal component analysis of partial NIR spectra (log 1/R, 1100-1900 nm and 1980-2498 nm) using different pre-treatments

Pre-treatment method	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
Mean centering	4	100	43	96	57	93
	5	100	68	99	93	87
	6	100	77	92	97	87
	7	96	85	89	97	90
	8	96	96	41	99	70
Variance Scaling	4	100	43	96	53	93
	5	100	67	99	92	87
	6	100	79	92	97	87
	7	100	88	91	97	92
	8	96	96	40	99	70
MSC [a]	5	100	91	76	97	80
	6	92	96	24	99	47
	7	96	96	7	99	48
SNV-detrend [b]	6	96	84	87	96	82
	7	100	95	16	100	70
SNV-alone [c]	8	96	95	16	99	73

[a] Multiplicative scatter correction.
[b] Standard normal variate transformation with detrending.
[c] Standard normal variate transformation alone.

Table 4: Percentage of correct classification of wheat samples based on principal component analysis of partial NIR spectra (1100-1900 nm and 1980-2498 nm) using derivative treatment

Derivative treatment	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
SG 1st-der [a] (5)[b]	4	96	55	97	36	90
	5	100	52	99	33	95
	6	100	45	99	24	95
	7	100	53	97	35	95
	8	100	61	96	41	95
SG 1st-der (9)	5	100	85	88	91	85
	6	100	84	91	91	88
	7	100	84	92	91	90
	8	100	88	85	93	88
Gap 2nd-der [c] (5)	4	100	45	100	29	98
	5	100	59	99	41	97
Gap 2nd-der (9)	5	96	91	77	93	78
	6	100	88	91	92	82
	7	100	91	83	95	75

[a] Savitsky-Golay first derivative.

[b] ()-Number of data point used in the derivative function.

[c] Gap second derivative.

Ghaedian and Wehling (1997) reported the use of truncated spectra from 1100 to 1900 nm to build calibration models. The best calibration model was found when using first derivative spectra with five factors, which correctly classified 100% of sound, 93% of infested, 95% of sound air-dried, 86% of infested air-dried, and 90% of sound kernels from six different wheat varieties. In our current work, calibration models were, therefore, developed using spectra from 1100 to 1900 nm. Higher numbers of factors were used to develop calibrations. Models developed from pre-treatments provided good prediction rates for classifying infested and infested air-dried kernels (Table 5). However, prediction rates were unsatisfactory for classifying sound air-dried kernels.

Table 6 summarizes the prediction rates when using derivatization. Savitsky-Golay first derivative transformation with five segment size in a five factor model yielded an optimum prediction rate for all validation sets, and the results were similar to those reported by Ghaedian (1995). Five factor models using Savitsky-Golay first derivative with segment size of nine, and Gap first derivative with gap size of five, gave similar results and were comparable to that of Savitsky-Golay first derivative with segment size of five. Gap second derivative with segment size of nine and a six factor model gave impressive results in classifying sound air-dried, infested, and infested-air dried, but lowered the prediction rate for six different varieties of sound wheat kernels.

Table 5: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra (log 1/R) from 1100 to 1900 nm using different pre-treatments

Pre-treatment method	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
Mean centering	8	100	97	65	100	87
	9	100	99	60	100	80
	10	100	99	69	100	87
Variance Scaling	9	100	99	51	100	80
	10	100	99	69	100	87
MSC [a]	8	100	99	49	100	80
	9	96	99	53	100	73
SNV-detrend [b]	9	100	99	17	100	83
	10	100	99	32	100	78
SNV-alone [c]	9	100	99	59	100	83
	10	100	99	21	100	73

[a] Multiplicative scatter correction.

[b] Standard normal variate transformation with detrending.

[c] Standard normal variate transformation alone.

Table 6: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using derivative treatment

Derivative treatment	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
SG 1st-der [a] (5)[b]	5	96	95	91	89	90
	6	96	96	79	97	82
	7	92	99	40	100	55
SG 1st-der (9)	5	96	93	93	88	92
	6	96	96	84	96	85
	7	96	99	47	100	63
Gap 1st-der [c] (5)	5	96	93	93	88	92
	6	100	96	83	96	85
	7	96	99	53	100	75
Gap 1st-der (9)	6	100	96	76	97	85
	7	100	95	68	97	83
SG 2nd-der [d] (5)	3	100	61	97	56	88
	5	100	75	89	68	90
SG 2nd-der (9)	3	100	61	97	56	88
	5	100	75	89	68	90
Gap 2nd-der [e] (5)	5	96	93	84	93	83
	6	96	95	89	97	80
	7	100	97	65	99	67
Gap 2nd-der (9)	6	96	95	92	93	82
	7	100	99	72	97	70

[a] Savitsky-Golay first derivative.
[b] ()-Number of data point used in the derivative function.
[c] Gap first derivative.
[d] Savitsky-Golay second derivative.
[e] Gap second derivative.

Calibration models were next developed using derivative spectra coupled with pre-treatments. Table 7 shows the classification results of the models that were created from SG first derivatives. Applying MSC to SG first derivative spectra with segment size of five and using five factors gave lower classification rates for infested kernels, and similar results for other validation sets compared with the model developed by SG first derivative with segment size of five without MSC (Table 6). When using MSC and SG first derivative with segment size of nine, an optimum prediction rate was obtained with five factors. SNV with and without detrending gave slightly lower classification rates compared with MSC. Poor prediction rates for infested and infested air-dried kernels were obtained when applying MSC to SG second derivative spectra (Table 8). The results were improved when applying SNV with and without detrending in a five factor model, however, the results were still unsatisfactory. It can be concluded that the use of SG second derivative either by itself or coupled with pathlength correction techniques did not provide good classification for infested and infested air-dried kernels. Table 9 shows prediction results when applying pre-treatments to Gap first derivative spectra. Classification results of the models using MSC coupled with Gap first derivative were comparable to those using Gap first derivative alone (Table 6), except for higher classification rates of the six varieties of sound wheat kernels. Models developed by SNV with and without detrending gave a lower percentage of correct classification compared with those developed by MSC.

Table 7: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Savitsky-Golay (SG) first derivative with different pre-treatments

Derivative treatment, pre-treatment method	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
SG 1st-der [a] (5)[b], MSC[c]	4	100	85	96	71	92
	5	100	88	93	89	90
	6	100	93	85	92	90
	7	92	96	63	99	77
SG 1st-der (5),SNV-detrend[d]	6	96	93	84	95	85
	7	92	96	79	95	85
SG 1st-der (5), SNV-alone[e]	5	100	89	96	83	90
	6	92	93	83	95	87
	7	96	95	65	99	80
SG 1st-der (9), MSC	5	96	93	88	95	90
	6	100	95	85	95	92
	7	92	97	51	99	72
	8	100	99	48	99	73
SG 1st-der (9),SNV-detrend	6	92	93	84	96	83
	7	96	96	75	97	83
SG 1st-der (9), SNV-alone	7	96	95	60	99	80
	8	96	99	21	100	60

[a] Savitsky-Golay first derivative.

[b] ()-Number of data point used in the derivative function.

[c] Multiplicative scatter correction.

[d] Standard normal variate transformation with detrending.

[e] Standard normal variate transformation alone.

Table 8: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Savitsky-Golay (SG) second derivative with different pre-treatments

Derivative treatment, pre-treatment method	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
SG 2nd-der[a] (5)[b], MSC[c]	3	100	17	88	17	80
	5	100	17	92	21	88
SG 2nd-der (5),SNV-detrend[d]	2	100	56	96	33	93
	3	100	61	99	44	93
	5	100	76	96	72	93
SG 2nd-der (5), SNV-alone[e]	2	100	56	96	33	93
	3	100	61	99	43	93
	5	100	76	96	72	93
SG 2nd-der (9), MSC	2	100	19	89	15	80
	3	100	17	88	17	80
	4	100	20	95	20	87
SG 2nd-der (9),SNV-detrend	3	100	61	97	56	88
	4	100	75	89	68	90
SG 2nd-der (9), SNV-alone	2	100	56	96	33	93
	3	100	61	99	44	93

[a] Savitsky-Golay second derivative.

[b] ()-Number of data point used in the derivative function.

[c] Multiplicative scatter correction.

[d] Standard normal variate transformation with detrending.

[e] Standard normal variate transformation alone.

Table 9: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Gap first derivative with different pre-treatments

Derivative treatment, pre-treatment method	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
Gap 1st-der[a] (5)[b], MSC[c]	6	100	95	85	93	92
	7	96	99	51	100	73
	8	100	99	43	99	77
Gap 1st-der (5),SNV-detrend[d]	6	96	93	84	95	83
	8	96	99	39	99	63
Gap 1st-der (5), SNV-alone[e]	6	92	93	77	95	83
	9	100	99	11	100	55
Gap 1st-der (9), MSC	6	96	95	84	93	90
	8	100	99	49	99	73
Gap 1st-der (9), SNV-detrend	5	100	88	96	81	90
	6	96	93	80	95	82
Gap 1st-der (9), SNV-alone	5	100	89	96	81	87
	6	92	93	69	97	78

[a] Gap first derivative.

[b] ()-Number of data point used in the derivative function.

[c] Multiplicative scatter correction.

[d] Standard normal variate transformation with detrending.

[e] Standard normal variate transformation alone.

The use of second derivative Gap spectra with gap size of five and coupled with SNV (with and without detrending) in a five factor model provided comparable results and yielded the best classification rates for all validation sets (Table 10). Classification rates were 100% for sound, 93% for infested, 95% for sound air-dried, 91% for infested air-dried, and 92% for sound kernels from six different wheat varieties, which were better results for classifying infested air-dried and six wheat varieties compared with those reported by Ghaedian (1995). This can be explained in that SNV transformation removes the effect of particle size and scatter, so effects due to changes in the physical properties of the samples are minimized. Therefore, spectral differences only arise from differences in chemical composition. Detrending did not improve classification rates. As the number of factors increased to six, overdiscrimination occurred indicating that the models were overfitted to their calibration sets.

Ghaedian (1995) explained that in a PCA model, factors one and two of both sound and infested wheat kernel spectra accounted for moisture, protein, carbohydrate and lipid which are the major constituents. Factor three of those kernels accounted for wheat carbohydrates. Factor four of sound kernels accounted for wheat protein, whereas water and Ar-OH structure were also represented in factor four of infested spectra. Lipid and moisture bands showed intensity difference in factor five for both sound and infested wheat kernels. The lipid bands of infested wheat in factors six and seven were also higher than those of sound kernels. He also reported that when the moisture band is eliminated (removing wavelengths above 1900 nm), protein, lipid, and Ar-OH structures play a major role in spectral differences between sound and infested wheat kernels.

Table 10: Percentage of correct classification of wheat samples based on principal component analysis of NIR spectra from 1100 to 1900 nm using Gap second derivative with different pre-treatments

Derivative treatment, pre-treatment method	No. of factors	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
Gap 2nd-der[a] (5)[b], MSC[c]	5	100	80	95	76	92
	6	96	93	79	96	80
Gap 2nd-der(5),SNV-detrend[d]	5	100	93	95	91	92
	6	100	93	79	96	87
Gap 2nd-der (5), SNV-alone[e]	5	100	93	95	91	92
	6	100	93	79	96	87
Gap 2nd-der (9), MSC	4	96	88	93	80	92
	5	96	95	93	92	90
	6	100	95	85	96	85
	7	92	95	75	99	80
Gap 2nd-der (9),SNV-detrend	5	100	93	97	85	95
	6	96	93	84	95	85
Gap 2nd-der (9), SNV-alone	5	100	93	97	85	95
	6	96	93	84	95	85

[a] Gap second derivative.

[b] ()-Number of data point used in the derivative function.

[c] Multiplicative scatter correction.

[d] Standard normal variate transformation with detrending.

[e] Standard normal variate transformation alone.

Classification results from calibration models using selected wavelengths

Several combinations of selected wavelengths were used to develop calibration models, which were then tested with the validation sets. A four wavelength model gave the best results, correctly classifying 98% of sound, 98% of infested, 100% of sound air-dried, 96% of infested air-dried, and 100% of six different varieties of sound wheat kernels (Table 11). This particular wavelength combination at 1120, 1130, 1280 and 1860 nm provided a greater percentage of correct classification than the results reported by Ghaedian (1995) from both techniques of PCA coupled with Mahalanobis distance, and selected wavelengths. Wavelengths at 1120 and 1130 nm are due to the second overtone of a C-H stretch found in aromatic structures (Osborne et al., 1993). The wavelength at 1280 nm is also due to the second overtone of a C-H stretch. The O-H stretch and C-O stretch at 1860 nm is probably associated with cellulose and starch in wheat. The best classification results derived from using a five PCA model based on first derivative spectra from 1100 to 1900 nm reported by Ghaedian (1995) are previously described. Also, in that study, the use of twelve discrete selected wavelengths based on underivatized, log (1/R) data from 1100 to 2498 nm gave classification results similar to the results from PCA, except correct classification of infested air-dried kernels was only 55% compared to 86% when using PCA.

Table 11: Percentage of correct classification of wheat samples based on Mahalanobis distances of selected wavelengths of NIR spectra from 1100 to 1900 nm

No. of wavelengths	Sound	Infested	Sound (air-dried)	Infested (air-dried)	6 variety
3[a]	96	91	100	96	87
4[b]	98	98	100	96	100
5[c]	100	91	100	96	93
6[d]	100	91	100	96	100
7[e]	100	84	100	96	95

[a] 3 selected wavelengths: 1150, 1260, 1360 nm.

[b] 4 selected wavelengths: 1120, 1130, 1280, 1860 nm.

[c] 5 selected wavelengths: 1140, 1270, 1360, 1380, 1650 nm

[d] 6 selected wavelengths: 1130, 1220, 1240, 1350, 1390, 1500 nm.

[e] 7 selected wavelengths: 1130, 1220, 1250, 1350, 1370, 1380, 1890 nm.

References

Delwiche, S.R., McKenzie, K.S., and Webb, B.D. 1996. Quality Characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. Cereal Chem. 73:257-263.

Dowell, F.E., Throne, J.E., and Baker, J.E. 1999. Identifying stored-grain insects using near-infrared spectroscopy. J. Econ. Entomol. 92:165-169.

Dowell, F.E., Throne, J.E., and Baker, J.E. 1998. Automated Nondestructive Detection of Internal Insect Infestation of Wheat Kernels by Using Near-Infrared Reflectance Spectroscopy. J. Econ. Entomol. 91:899-904.

Ghaedian, A.R. and R.L. Wehling. 1997. Discrimination of sound and granary-weevil-larva-infested wheat kernels by near-infrared diffuse reflectance spectroscopy. J. Assoc. Off. Agric. Chem. 80: 997-1005.

Ghaedian, A.R. 1995. Detection of insect infestation in stored products by instrumental methods of analysis. PhD Dissertation. University of Nebraska-Lincoln, Lincoln, NE.

Maghirang, E.B., Dowell, F.E., Baker, J.E., and Throne, J.E. 2003. Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy. Trans. ASAE 46:1277-1282.

Murray, I., and P.C. Williams. 1990. Chemical principle of near-infrared technology. Ch.2 In: Near-Infrared Technology in The Agricultural and Food Industries. P.C. Williams and K.H. Norris (eds.), American Association of Cereal Chemists. St. Paul, MN, 17-34 pp.

Osborne, B.G., Fearn, T., and Hindle, P.H. 1993. Theory of near infrared spectrophotometry. Ch.2 in: Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, Longman Scientific&Technological, Essex, UK, 13-35 pp.

Pedersen, J.R. 1992. Insects: identification, damage, and detection. Ch.12 In: Storage of Cereal Grains and Their Products, 4th edition. D.B. Sauer (ed.), American Association of Cereal Chemists, St. Paul, MN, 435-490 pp.

Ridgway, C. and J. Chambers. 1996. Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. J. Sci. Food Agric. 71: 251-264.

Chapter 2

A Simplified Near-Infrared Method for Detecting Internal Insect Infestation in Wheat Kernels

Abstract

A simplified near infrared (NIR) reflectance spectroscopy method for detecting internal insect infestation in wheat kernels has been developed. The method detects the presence of 3- and 4- week-old granary weevil or maize weevil larvae in individual wheat kernels. Discriminant analysis, based on Mahalanobis distances calculated from log 1/R data at only four discrete wavelengths, was able to discriminate sound kernels from those internally infested with 4-week-old larvae. When applied to a validation set, this model correctly classified 97% of sound, 100% of sound air-dried, 89% of infested kernels containing 4-week-old granary weevil larvae, 93% of kernels containing 4-week-old maize weevil larvae, 98% of air-dried kernels infested with granary weevil, and 94% of air-dried kernels infested with maize weevil. However, the model correctly identified less than 50% of kernels containing 3-week-old larvae. A model developed from a training set that contained kernels infested with 3-week-old larvae yielded a higher rate of correct classification for a validation set containing 3-week-old granary weevil larvae (77% correct), and maize weevil larvae (73% correct). This model also improved classification rates of infested kernels containing 4-week-old larvae of granary and maize weevils to 95 and 98%, respectively. The models were also tested with sound kernels from 10 varieties of wheat with varying kernel characteristics. Correct classification rates were 100 and 99% when

using calibrations developed from 4- and 3-week-old larvae, respectively. Results indicated that either live (infested kernels containing 3- or 4-week-old larvae) or dead insects (infested air-dried kernels) can be used to develop calibrations for detecting both live and dead insects in wheat.

Introduction

As mentioned in the previous chapter, the presence of internal insects in wheat is a major problem for the wheat industry. The presence of live or dead internal insects in wheat kernels can lower the quality of wheat. In addition, emergence of live insects can cause further damage to kernels and contribute to fragments in flour. NIR spectroscopy is an alternative to other methods since it is a rapid and accurate technique, and is widely used throughout the U.S. grain industry. It can also be adapted for non-destructive and automated detection.

Discriminant analysis based on Mahalanobis distances was explained by Mark and Tunnell (1985) and Mark (2001). The Mahalanobis distance, from a statistical viewpoint, takes the sample variability into account, whereas the Euclidian distance method does not take into account the variability of the values in all dimensions. In other words, Mahalanobis distances look at not only variation between the responses at the same wavelengths, but also at inter-wavelength variations. Instead of treating all values equally when calculating the distance from the mean point, it weighs the differences by the range of variability in the direction of the sample point. The location of each cluster in multidimensional space is described by the mean value of the absorbances (the group mean) at each wavelength.

Qualitative NIR reflectance analysis using Mahalanobis distances based on selected discrete wavelengths is a simple technique and requires much less expensive instrumentation than full spectrum methods using Principal Component Analysis (PCA). Studies on calibrations developed from reduced-spectrum models instead of full-spectrum models have

been reported (Ridgway et al, 1999; Dodds and Heath, 2005). The previous chapter shows that the development of discrete wavelength models can provide improved classification results compared to techniques utilizing PCA. The objectives of this study, therefore, were to determine if discriminant analysis functions could differentiate between sound and infested wheat kernels using Mahalanobis distances based on NIR reflectance spectral data at only a few discrete wavelengths, and to evaluate the ability of NIR spectroscopy to detect the earliest stages of granary weevil and maize weevil larvae in wheat kernels.

Materials and Methods

Preparation of Insect Infested and Sound Wheat Kernels

Insect colonies of granary weevil [Sitophilus granarius (L.)] and maize weevil (Sitophilus zeamis) were obtained from the stored products research laboratory of the Department of Entomology at Kansas State University, Manhattan, KS. Approximately 200 g of clean Hard Red Winter (HRW) wheat from the Cereal Quality Laboratory of the Department of Agronomy at the University of Nebraska -Lincoln was placed into a wide-mouth quart canning jar, and cultured with unsexed parent insects of a single species (~ 100 insects each). In order to protect the wheat and insects from outside contamination while providing a source of oxygen during incubation, the jars were covered with 40 mesh window screen over a circle of Whatman No.4 filter paper. A high-moisture incubator was not available to raise the insects, so in order to increase humidity in the incubator, a wide and shallow tray containing a cotton sheet made into a tent shape was placed on the bottom shelf of the incubator. Two smaller trays were also used for the same purpose and placed on the

middle shelf. In order to increase surface area of the water carrier, two triangle-shaped screens covered with cotton cloth were placed in those trays. Water was then added into all trays and maintained through the period of raising the insects. Relative humidity of the incubator could be raised only to 50 ∀ 5%, which is lower than the optimum environmental condition for hatching and development of stored-product insects (Wehling et al, 1984). Schwartz and Burkholder (1991), however, documented that immature granary weevil development in wheat can occur between 40 and 80% RH. Khan (1949) reported the number of days to larval instars (1^{st}, 2^{nd}, 3^{rd}, and 4^{th}) and pupal stage of granary weevil at 50%RH and 25^{o}C are 6, 6, 6, 11, and 12, respectively. The cultures were, therefore, raised at 50 ∀ 5 % RH and a temperature of 25 ∀ 3 ^{0}C. The female insects were allowed to deposit eggs into the wheat kernels for a period of 4 days, and the parents were then removed by sieving. Cultures were returned to the incubator, with individual culture jars removed after 2, 3, and 4 weeks as the desired level of larval development was achieved.

Identification of Sound and Infested Kernels by Radiography

A portion of each culture was collected to identify sound and infested kernels using a general Medical X-ray Unit at the conditions of 30 kV and 25mA, with a 3 sec exposure time. Radiographs were obtained by placing wheat kernels into the wells of micro-plates, arranging one kernel in each well. The radiographs were then viewed with a film illuminator. Infested kernels were visually identified, and removed using forceps from their well position on the micro-plate. In a preliminary study, the radiographs obtained by the Medical X-ray Unit did not provide sufficiently clear images to reliably identify 2-week-old insect larvae by visual inspection of the radiographs. Thus, only wheat kernels containing

3- and 4-week-old granary and maize weevil larvae were used in this study. All infested kernels chosen contained one internal insect per kernel. NIR spectra were then obtained from wheat kernels within 24 h after being X-rayed and sorted to minimize changes in the growth stage of the larvae. To insure that sound wheat kernels were treated in the same manner as infested kernels, sound kernels were also picked from the radiographed kernels.

NIR Analysis and Data Acquisition

Spectral data from all wheat kernels, expressed in the form of the absorbance (logarithm of the inverse of the reflectance, log 1/R), were collected with an NIRSystems Model 6500 spectrometer (NIRSystems Division of Foss Electric, Silver Spring, MD). Diffuse reflectance spectra were obtained from intact wheat kernels over a spectral range from 1100 to 2498 nm. Spectra were collected at 2-nm increments. Individual wheat kernels were carefully placed in a Capcell^{TM} parabolic reflector (Optical Prototypes, Mars, PA). This device allows the collection of radiation reflected from the entire surface of the kernel. The kernels were placed in a sample holder in crease down position, which was previously found to provide higher correct classification rates compared with the crease up position (Ghaedian and Wehling 1997). In addition, the orientation and position of the kernels in the sample holder were carefully reproduced. A remote reflectance probe attached to the monochromator of the instrument by a fiber optic cable was then positioned over the Capcell to measure light reflected from the kernel. Thirty-two monochromator scans were accumulated in computer memory. The scans were averaged, transformed to log (1/R), and then stored in a computer file, forming one spectrum per each kernel. The NIR spectrum of the kernel was obtained by taking the ratio of the intensity of radiation reflected from the

sample to that reflected from a ceramic reference plate. A black cloth was used to cover the remote probe and parabolic reflector during collection of each spectrum to minimize the effect of stray light.

To minimize the effect of any variation due to instrument drift, scanning was alternated between sound and infested kernels after every 10 kernels. After scanning, each kernel was numbered for further identification. Infested kernels were then frozen at -18^{o}C for 48 hr to kill the larvae, followed by air drying at room temperature on an open tray for about 14 days. These kernels contained dead larvae and simulated fumigated kernels in storage. After collection of these spectra, the kernels were sectioned with a razor blade to ensure the presence or absence of internal insect larvae.

For data collection, the spectrometer was interfaced to a personal computer running the Near Infrared Spectral Analysis Software (NSAS) package (version 3.53, FOSS NIRSystems, Silver Spring, MD). To enhance peaks and correct the baseline of the spectra, first and second derivatives of spectra were also calculated. All raw spectra (log1/R) were pretreated with first or second derivative processing with segment interval = 10 or 20 nm, and gap = 10 or 20 nm, using NSAS software. Spectra were then stored on the hard disk drive and then converted into ASCII (JCAMP-DX) format for analysis with MultiQual software (1998) from the Near Infrared Research Corp (Suffern, NY).

Discriminant Analysis

Previous results showed promise for correct classification by using discriminant analysis based on Mahalanobis distances applied to selected wavelengths. Therefore, this study was focused only on this method of analysis. The principle of this technique was

previously described in chapter 1.

Calibration Sets

Two calibration sample sets for model development were prepared. The first calibration set (n=200) was randomly selected from NIR spectra of six different groups of sound (n=100) and infested (n=100) kernels. They were:

1) Sound wheat kernels (n=50),

2) Sound air-dried what kernels (n=50),

3) Infested kernels containing 4-week-old larvae of granary weevil (n=25),

4) Infested kernels containing 4-week-old larvae of maize weevil (n=25),

5) Infested air-dried kernels containing 4-week-old larvae of granary weevil (n=25),

6) Infested air-dried kernels containing 4-week-old larvae of maize weevil (n=25).

The kernels identified as Asound air-dried@ were the sound kernels that had been air-dried at room temperature before collecting the spectra, and therefore had a different moisture content compared to the other sound kernels. The first groups of sound and sound air-dried wheat kernels were identified as Asound@ in the JCAMP-DX format in order to be analyzed with the software. The others were then identified as Ainfested@.

The second calibration set (n=200) was developed from the same NIR spectra which had been used in the first calibration set, except for the third thru sixth groups. In this second set, infested kernels containing 3-week-old larvae of granary and maize weevils were used instead of spectra from those kernels containing 4-week-old larvae of granary and maize weevils.

The preliminary results in the previous chapter showed that a calibration model of

four wavelengths gave the best classification results. Therefore, a number of four wavelength combinations in the spectral region from 1100 to 1900 nm were used to develop calibrations. Once a model was developed, a leave-one-out algorithm, or cross-validation, was used to verify classification capability of the model.

To minimize time needed for computation, skipping factors were used. Several searching strategies were tried in order to achieve the best classification results. Tables 1 and 2 show several approaches of using log 1/R data from sound and infested kernels containing 4- or 3-week-old larvae, respectively, to build calibrations. In model 1, every fifth data point was considered for building the trial model. Because of the wavelength spacing of 2 nm, this meant that wavelengths on 10 nm intervals were considered. To ensure that additional combinations were tried that may produce a better model, the third and fourth selected wavelengths were searched at every 2 nm in model 2. For faster computation, every twenty-fourth data point was used in model 3. Therefore, wavelengths on 48 nm intervals were used to build the trial model. Interleaving wavelengths, as suggested by the program author, may keep the search small and still include all of the available wavelengths in one or another of the models. In this searching approach, the number of the skip factor is equal to the number of wavelengths to be selected, and the beginning of each range starts one data point over from the previous one. Therefore, the skip factor was set to 4 and the wavelength range was incremented at a 2 nm interval as shown in model 4. An efficiency search for 4 wavelengths in model 5 was generated by the program software using an efficient proprietary algorithm to search through the multidimensional space containing the data. Searching approaches of first and second derivatized spectra for both calibration models

followed those of raw spectra. Some searching approaches were not used with derivative treatments to build calibration models, since they were not be able to be processed by the program software. For example, an approach in model 2 was removed from all derivative treatments.

Once the discriminant analysis models were developed, the models were used to predict samples from the validation sets. Samples in the validation sets were classified as Asound@ or Ainfested@. The wavelength combinations selected by the searches that provided the highest percentage of correctly classified samples were identified.

Validation Sets

Each calibration model was used to predict samples in separate validation sets. They were:

1) Sound wheat kernels (n=90),

2) Sound air-dried wheat kernels (n=85),

3) Infested kernels containing 4-week-old larvae of granary weevil (n=87),

4) Infested kernels containing 4-week-old larvae of maize weevil (n=86),

5) Infested kernels containing 3-week-old larvae of granary weevil (n=87),

6) Infested kernels containing 3-week-old larvae of maize weevil (n=75),

7) Infested air-dried kernels containing 4-week-old larvae of granary weevil (n=52),

8) Infested air-dried kernels containing 4-week-old larvae of maize weevil (n=51),

9) Sound kernels of ten different wheat varieties (Scout 66, Siouxland, Vista, Jagger, Ogallala, Windstar, Centura, Yumar, Experimental White, Nuplains White) grown in Nebraska and obtained from the Cereal Quality Laboratory of the Department of Agronomy

at the University of Nebraska -Lincoln. Ten kernels of each variety were used to construct a validation set (n=100).

Table 1: Searching approaches for constructing calibrations using raw data (log1/R) sound and infested kernels with 4-week-old larvae performed with MultiQual software.

Model	Searching Approaches			Selected Wavelengths	Mahalanobis distances
	Wavelength No.	Wavelength Range	Skip Factor		
1	1	1100-1900	5	1140, 1120, 1150, 1400	4.13151
	2	1100-1900	5		
	3	1100-1900	5		
	4	1100-1900	5		
2	1	1100-1900	5	1152, 1128, 1400, 1140	4.24386
	2	1100-1900	5		
	3	1100-1900	1		
	4	1100-1900	1		
3	1	1100-1900	24	1136, 1124, 1160, 1676	4.08155
	2	1112-1900	24		
	3	1124-1900	24		
	4	1136-1900	24		
4	1	1100-1900	4	1154, 1136, 1406, 1124	4.25396
	2	1102-1900	4		
	3	1104-1900	4		
	4	1106-1900	4		
5	Efficiency search 4 wavelengths			1116, 1136, 1158, 1550	3.85744

Table 2: Several searching approaches for constructing calibration models using raw data (log1/R) of sound and infested kernels with 3-week-old larvae performed with MultiQual software.

Model	Searching Approaches			Selected Wavelengths	Mahalanobis distances
	Wavelength No.	Wavelength Range	Skip Factor		
1	1	1100-1900	5	1130, 1140, 1160, 1390	2.33174
	2	1100-1900	5		
	3	1100-1900	5		
	4	1100-1900	5		
2	1	1100-1900	5	1128, 1138, 1390, 1160	2.48997
	2	1100-1900	5		
	3	1100-1900	1		
	4	1100-1900	1		
3	1	1100-1900	24	1136, 1124, 1160, 1388	2.40566
	2	1112-1900	24		
	3	1124-1900	24		
	4	1136-1900	24		
4	1	1100-1900	4	1138, 1128, 1390, 1156	2.47427
	2	1102-1900	4		
	3	1104-1900	4		
	4	1106-1900	4		
5	Efficiency search 4 wavelengths			1106, 1138, 1160, 1566	2.07903

Results and Discussion

Figure 1 displays the average absorbance of 50 kernels each of sound (uninfested) or internally infested wheat kernels (3-and 4-week-old granary and maize weevils). Spectra of infested air-dried wheat kernels (spectra not shown) had the same trends as those for infested wheat kernels. Absorbance (log 1/R) was generally highest for sound wheat kernels and decreased with infestation. These results agreed with results reported by Ridgway and Chambers (1996), Dowell et al (1998), and Maghirang et al (2003). This was probably due to a significant proportion of the radiation penetrating through the outer kernel to interact with the insect and surrounding cavity (Ridgway and Chambers, 1996). Interaction with both the internal insect and the surface of the insect cavity will lead to greater back reflection and thus lower log 1/R values. Therefore, changes in physical properties of wheat samples due to the presence of the internal insects probably caused spectral changes with infestation. Additionally spectral differences were most likely due to the presence of insect cuticle which contains both chitin and lipid. Several researchers have previously reported the effect of these components on the C!H absorption region around 1100-1700 nm. (Ghaedian and Wehling 1997, Dowell et al 1998 and 1999, Maghirang et al 2003, Perez-Mendoza et al 2003).

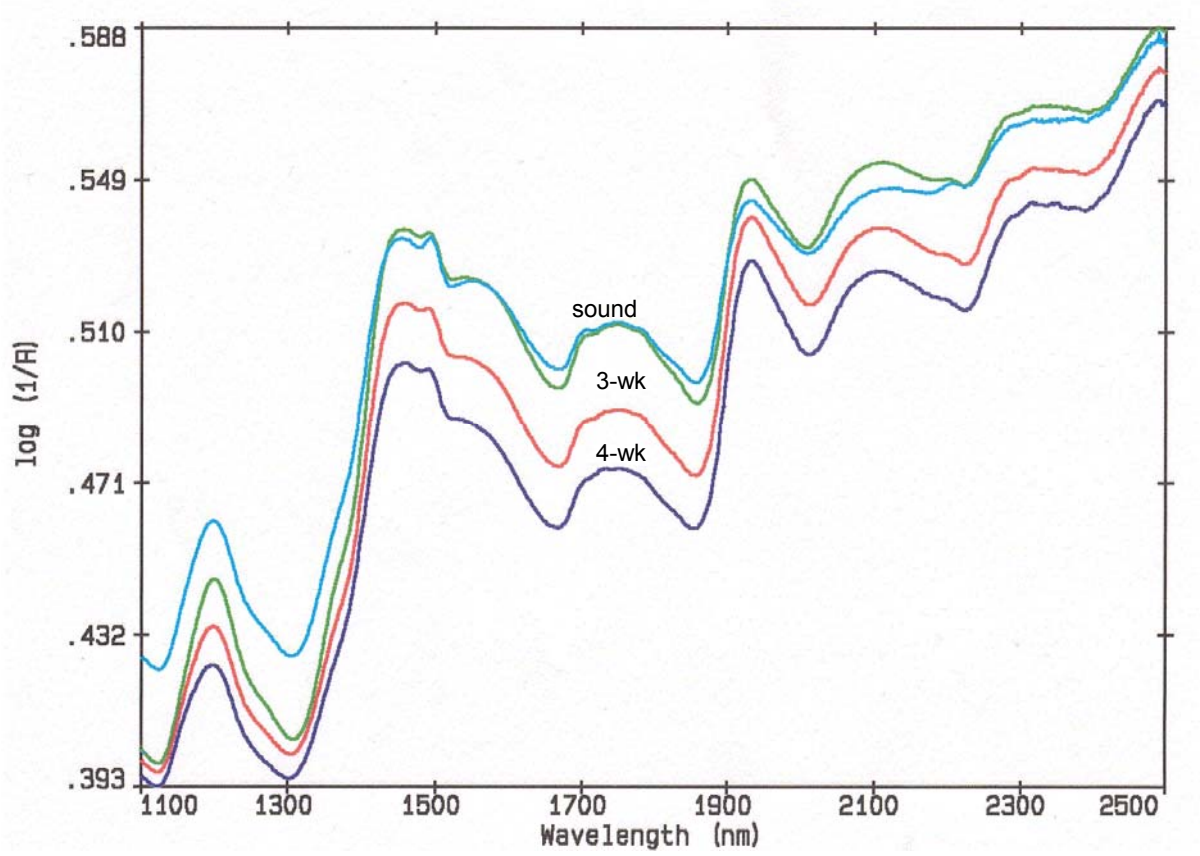

Fig. 1. Average absorbance of sound wheat kernels, 3- and 4-week-old granary and maize weevils infested kernels.

Classification Results using Calibrations Developed from 4-Week-Old Larvae

A summary of classification results when infested kernels from 4-week-old larvae were used to develop calibrations from raw spectra is given in Table 3. Overall, the misclassification rate of the calibration set was 4.2% and the correct classification rate was 95.8%. The calibration models were then used to test samples in the validation (prediction)

set, which had not been used in the training (calibration) set. By looking at overall correct classification, the best results were obtained using the searching approach in model 5. The results revealed that 97% of sound, 100% of sound air-dried, 89% of infested kernels containing 4-week-old larvae of granary weevil, 93% of kernels containing 4-week-old maize weevil larvae, 98% of air-dried kernels infested with granary weevil, 94% of air-dried kernels with maize weevil, and 100% of sound kernels from ten different wheat varieties were correctly classified into their respective classes. Table 4 shows the distribution of sound and infested wheat kernels in both misclassification and correct classification. With model 5, for example, three of two hundred seventy five sound wheat kernels, and one hundred fourteen of four hundred thirty eight infested kernels, were actually misclassified in the validation set. These impressive results demonstrate that NIR spectroscopy and discriminant analysis techniques can be used for identifying sound and infested wheat kernels, regardless of differences in moisture content and wheat varieties. However, infested kernels containing 3-week-old larvae of maize and granary weevils were poorly classified with correct percentages of 37 and 46, respectively.

Table 3: Correct classification of wheat samples using a calibration developed from raw spectra of sound and 4-week-old larvae with different searching approaches.

	Model	Percentage of correct classification								
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)		10 variety
						4-wk-old	3-wk-old	4-wk-old	3-wk-old	
Calibration	1	100	100	100	100	92	-	92	-	-
samples	2	100	100	100	100	92	-	92	-	-
	3	100	100	100	100	88	-	96	-	-
	4	100	100	100	100	88	-	96	-	-
	5	100	100	100	96	92	-	96	-	-
Validation	1	100	100	98	96	86	35	91	32	99
samples	2	100	100	98	96	86	32	91	29	100
	3	98	100	98	94	87	40	92	31	100
	4	99	100	98	96	86	36	91	33	100
	5	97	100	98	94	89	46	93	37	100

Table 4: Classification matrices from several searching approaches for constructing calibration models using raw spectra of sound and infested kernels with 4-week-old larvae.

Model	Class found ↓	Calibration		Validation	
		Actual class →		Actual class →	
		Sound (N=100)	Infested (N=100)	Sound (N=275)	Infested (N=438)
1	Sound	99	2	274	131
	Infested	1	98	1	307
2	Sound	100	3	275	135
	Infested	0	97	0	303
3	Sound	100	4	273	126
	Infested	0	96	2	312
4	Sound	100	4	274	129
	Infested	0	96	1	309
5	Sound	100	4	272	114
	Infested	0	96	3	324

The chosen wavelengths were 1116, 1136, 1158 and 1550 nm. The first three wavelengths are associated primarily with the second overtone of a C-H stretch found in aromatic structures, CH_3, and CH_2. These absorbance regions agreed with those reported by Dowell et al (1998) and Maghirang et al (2003). The wavelength at 1550 nm is due to an N-H first overtone (Osborne et al, 1993), indicating that the ability to detect infested wheat

kernels may partly be due to the change in protein content of the wheat kernels. CH_3 and CH_2 groups are common chemical moieties in components that make up the cuticular lipids in insects. Generally, insect cuticular lipids are composed of fatty acids, alcohols, esters, glycerides, sterols, aldehydes, ketones and hydrocarbons (Jackson et al, 1974; Blomquist and Dillwith, 1985). Chitin and protein are also predominant constituents in insect larvae (Singh and Sinha, 1977).

Table 5 provides calibration results for models developed by applying a first derivative treatment to spectra of sound and infested wheat kernels and using a leave-one-out algorithm to classify samples in the calibration set. When applied to validation samples (Table 6), the calibration using a segment interval= 5 and gap= 5 in model 4 improved classification results for sound and infested wheat kernels containing 4-week-old granary and maize weevils to 99, 92, and 94%, respectively. The wavelengths chosen were 1304, 1330, 1358, and 1500 nm. The first three wavelengths correspond to C-H first and second overtones, and C-H combination vibrations, i.e., stretching modes and various deformation modes, which are found in CH_2 and CH_3 (Osborne et al, 1993). The band at 1500 nm is also associated with protein. A segment interval= 10 and gap= 10 gave lower percentages of correct classification for wheat kernels infested with live larvae. Calibrations developed from second derivative processing gave lower correct classification rates for infested kernels containing both 3- and 4-week-old larvae compared to those built from raw spectra (Table 7 for calibration samples and Table 8 for prediction samples).

Table 5: Correct classification of wheat samples (calibration samples) by models using first derivative spectra of sound and 4-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment =5, gap =5	1	98	100	100	100	96	-	96	-
	2	-	-	-	-	-	-	-	-
	3	98	100	100	96	96	-	92	-
	4	98	100	100	100	96	-	96	-
	5	96	98	100	100	92	-	88	-
segment=10, gap=10	1	96	100	100	100	96	-	96	-
	2	-	-	-	-	-	-	-	-
	3	98	100	100	96	96	-	88	-
	4	96	100	100	96	96	-	96	-
	5	98	100	100	96	92	-	88	-

Table 6: Correct classification of wheat samples (validation samples) by models using first derivative spectra of sound and 4-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification for validation samples							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment =5, gap =5	1	99	100	96	92	91	20	93	31
	2	-	-	-	-	-	-	-	-
	3	99	100	94	92	89	31	92	41
	4	99	100	96	92	92	24	94	31
	5	100	100	98	92	82	22	90	43
segment=10, gap=10	1	100	100	98	96	85	31	88	27
	2	-	-	-	-	-	-	-	-
	3	100	100	98	90	82	23	87	25
	4	100	100	100	96	85	30	86	25
	5	99	99	98	94	85	29	89	40

Table 7: Correct classification of wheat samples (calibration samples) by models using second derivative spectra of sound and 4-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment =5, gap =5	1	98	100	100	100	100	-	92	-
	2	-	-	-	-	-	-	-	-
	3	96	98	100	96	80	-	84	-
	4	98	100	100	96	96	-	92	-
	5	96	98	100	96	88	-	84	-
segment=10, gap=10	1	98	100	100	100	92	-	88	-
	2	-	-	-	-	-	-	-	-
	3	94	96	100	96	88	-	100	-
	4	98	100	100	100	96	-	88	-
	5	96	100	100	100	88	-	92	-

Table 8: Correct classification of wheat samples (validation samples) by models using second derivative spectra of sound and 4-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification for validation samples							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment =5, gap =5	1	98	99	98	96	87	30	91	27
	2	-	-	-	-	-	-	-	-
	3	98	99	100	94	78	26	86	27
	4	99	100	98	94	86	33	90	25
	5	97	98	98	92	86	28	87	33
segment=10, gap=10	1	100	100	98	92	84	34	93	31
	2	-	-	-	-	-	-	-	-
	3	99	98	100	94	77	31	85	31
	4	99	100	98	94	86	34	92	28
	5	98	100	98	96	86	30	90	32

Classification Results using Calibration Models Developed from 3-Week-Old Larvae

Models developed from raw spectra of 3-week-old larvae gave improved prediction results for infested kernels containing 3-week-old maize and granary weevil larvae (Table 9). Overall, the searching approach in model 2 provided the best results as shown in Table 10. Nine sound wheat kernels, and fifty-one infested kernels were misclassified. Table 9 reveals that infested kernels containing 3-week-old larvae of granary weevil and maize weevil were correctly classified with percentages of 77, and 73, respectively. Moreover, classification rates of infested kernels containing 4-week-old larvae of maize and granary weevils were also increased to 98 and 95%, respectively. Nevertheless, classification results of sound and granary weevil air-dried kernels were slightly decreased compared to the results from the model developed from 4-week-old larvae. It may be that when 3-week-old larvae are used to develop calibrations, the Mahalanobis distances becomes smaller, resulting in increased difficulty of classification between sound and infested kernels. Wavelengths used for the best classification were 1128, 1138, 1160, and 1390 nm. The first three wavelengths are associated with a second overtone of a C-H stretch found in aromatic structures and CH_3. The wavelength at 1390 nm is associated with C-H combination vibrations (Shenk et al, 1992). Table 11 and 12 show that correct classification rates for calibration and validation samples were decreased when a first derivative treatment was applied to raw spectra. When spectra were treated with a second derivative (Table 13 and 14), results comparable to raw spectra for the validation set were found when using segment=10 and gap=10 in model 3. However, the correct classification rate of sound wheat kernels was slightly lower compared to the result from raw spectra. Using the same treatment, 98% of

Table 9: Correct classification of wheat samples using calibrations developed from raw spectra of sound and 3-week-old larvae developed with different searching approaches.

	Model	Percentage of correct classification								
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)		10 variety
						4-wk-old	3-wk-old	4-wk-old	3-wk-old	
Calibration	1	82	96	100	100	-	72	-	72	-
samples	2	90	98	100	100	-	68	-	72	-
	3	88	98	100	100	-	72	-	72	-
	4	90	98	100	100	-	64	-	68	-
	5	84	90	96	92	-	76	-	72	-
Validation	1	89	96	96	94	95	79	97	72	99
samples	2	92	98	96	94	95	77	98	73	99
	3	87	99	98	94	94	79	97	69	96
	4	92	98	98	94	97	75	98	69	95
	5	79	93	96	96	95	82	95	73	98

sound wheat kernels were correctly classified in model 1 and 5. However, the percentage of correct classification of infested wheat kernels containing 3-week-old larvae was decreased.

Table 10: Classification matrices from several searching approaches for constructing calibration models using raw spectra of sound and infested kernels with 3-week-old larvae.

Model	Class found ↓	Calibration		Validation	
		Actual class →		Actual class →	
		Sound (N=100)	Infested (N=100)	Sound (N=275)	Infested (N=438)
1	Sound	89	14	261	51
	Infested	11	86	14	387
2	Sound	94	15	266	51
	Infested	6	85	9	387
3	Sound	93	14	258	53
	Infested	7	86	17	385
4	Sound	94	17	261	54
	Infested	6	83	14	384
5	Sound	87	16	248	48
	Infested	13	84	27	390

Table 11: Correct classification of wheat samples (calibration samples) by models using derivative spectra of sound and 3-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment =5, gap =5	1	90	96	96	96	-	68	-	80
	2	-	-	-	-	-	-	-	-
	3	90	94	100	96	-	68	-	72
	4	90	96	96	96	-	72	-	76
	5	84	90	96	96	-	60	-	68
segment =10, gap =10	1	92	98	100	100	-	68	-	68
	2	-	-	-	-	-	-	-	-
	3	88	92	100	96	-	76	-	64
	4	88	94	100	100	-	72	-	72
	5	90	94	100	96	-	68	-	64

Table 12: Correct classification of wheat samples (validation samples) by models using first derivative spectra of sound and 3-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification for validation samples							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment =5, gap =5	1	89	93	90	94	95	56	99	68
	2	-	-	-	-	-	-	-	-
	3	83	95	94	94	95	60	97	77
	4	86	94	90	94	95	56	99	69
	5	83	88	87	94	92	49	95	68
segment =10, gap =10	1	98	99	98	92	95	63	98	73
	2	-	-	-	-	-	-	-	-
	3	89	96	92	92	95	77	95	65
	4	89	96	98	92	95	64	97	64
	5	89	95	92	88	97	75	95	65

Table 13: Correct classification of wheat samples (calibration samples) by models using second derivative spectra of sound and 3-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment = 5, gap = 5	1	90	90	100	92	-	64	-	68
	2	-	-	-	-	-	-	-	-
	3	84	92	92	96	-	68	-	68
	4	-	-	-	-	-	-	-	-
	5	84	94	92	100	-	68	-	84
segment=10, gap=10	1	96	98	100	100	-	56	-	60
	2	-	-	-	-	-	-	-	-
	3	88	94	100	100	-	72	-	80
	4	-	-	-	-	-	-	-	-
	5	92	100	100	96	-	60	-	72

Table 14: Correct classification of wheat samples (validation samples) by models using second derivative spectra of sound and 3-week-old larvae developed with different searching approaches.

Derivative treatment	Model	Percent correct classification for validation samples							
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)	
						4-wk-old	3-wk-old	4-wk-old	3-wk-old
segment = 5, gap = 5	1	87	92	94	90	92	69	95	69
	2	-	-	-	-	-	-	-	-
	3	77	92	94	90	97	63	97	77
	4	-	-	-	-	-	-	-	-
	5	80	95	86	86	95	69	97	75
segment=10, gap=10	1	98	100	98	94	94	68	97	60
	2	-	-	-	-	-	-	-	-
	3	89	95	98	92	99	77	97	72
	4	-	-	-	-	-	-	-	-
	5	98	98	94	94	94	68	97	76

An attempt to improve classification results for infested kernels containing 3-week-old larvae was later tried by developing another calibration set. The trial calibration set, therefore, contains raw spectra from samples used previously to build calibrations, including sound wheat kernels (n=50), sound air-dried wheat kernels (n=50), infested wheat kernels containing 3-week-old larvae of granary weevil (n=25), and maize weevil (n=25). In addition, another 25 each of infested kernels containing 3-week-old larvae of granary and maize weevils were randomly selected from the validation set, and added into the calibration set. Therefore, this calibration set (n=200) contained wheat samples of sound (n=100) and infested (n=100) kernels containing 3-week-old larvae of both insect species. Infested air-dried kernels were not included in this trial. The calibration developed was then applied to wheat samples in the validation set used previously, except the numbers of infested kernels containing 3-week-old larvae of granary and maize weevils were decreased to 62 and 50, respectively. Overall, the searching approach in model 4 provided the best results.

Table 15 shows the classification results achieved, with 85% of sound, 98% of sound air-dried, 92% of infested air-dried kernels containing granary weevil, 90% of infested air-dried kernels containing maize weevil, 99% of kernels infested with 4-week-old larvae of granary weevil, 99% of kernels containing 4-week-old larvae of maize weevil, 87% of kernels infested with 3-week-old larvae of granary weevil, 84% of kernels infested with 3-week-old larvae of maize weevil, and 96% of sound kernels from ten different wheat varieties correctly classified. The selected wavelengths were 1128, 1146, 1156, and 1390 nm, which are similar to those selected in previous

calibrations. This model, however, yielded lower correct classification rate for sound kernels than the previous model that includes infested air-dried for developing calibration (85% compared to 92%).

Table 15: Correct classification of wheat samples using models developed with a modified calibration set. Models were developed from raw spectra of sound and 3-week-old larvae using different searching approaches.

	Model	Percentage of correct classification								
		Sound	Sound (air-dried)	Granary weevil (air-dried)	Maize weevil (air-dried)	Granary weevil (live)		Maize weevil (live)		10 Variety
						4-wk-old	3-wk-old	4-wk-old	3-wk-old	
Calibration	1	89	90	-	-	-	88	-	83	-
samples	2	85	98	-	-	-	77	-	83	-
	3	83	98	-	-	-	79	-	79	-
	4	87	98	-	-	-	81	-	83	-
	5	78	96	-	-	-	79	-	77	-
Validation	1	79	94	73	82	98	87	100	80	95
samples	2	83	98	81	80	98	85	98	84	96
	3	76	95	96	94	95	81	96	78	92
	4	85	98	92	90	99	87	99	84	96
	5	73	90	77	86	96	81	96	84	90

References

Blomquist, G.J., and Dillwith, J.W. 1985. Cuticular lipids. In: Comprehensive Insect Physiology, Biochemistry, and Pharmacology, Vol. III. Kerkut, G.A., and Gilbert, L.I. (eds), Pergamon Press, New York, 117-154 pp.

Dodds, S.A., and Health, W.P. 2005. Construction of an online reduced-spectrum NIR calibration model from full-spectrum data. Chemom. Intell. Lab. Syst. 76:37-43.

Dowell, F.E., Throne, J.E., and Baker, J.E. 1999. Identifying stored-grain insects using near-infrared spectroscopy. J. Econ. Entomol. 92:165-169.

Dowell, F.E., Throne, J.E., and Baker, J.E. 1998. Automated Nondestructive Detection of Internal Insect Infestation of Wheat Kernels by Using Near-Infrared Reflectance Spectroscopy. J. Econ. Entomol. 91:899-904.

Ghaedian, A.R. and R.L. Wehling. 1997. Discrimination of sound and granary-weevil-larva-infested wheat kernels by near-infrared diffuse reflectance spectroscopy. J. Assoc. Off. Agric. Chem. 80: 997-1005.

Jackson, L.L., Armold, M.T., and Regnier, F.E. 1974. Cuticular lipids of adult fleshflies, *Sarcophaga bullata*. Insect Biochem. 4:369-379.

Khan, M.Q. 1949. A contribution to a further knowledge of the structure and biology of the weevils *Sitophilus oryzae* (Linn.) and S. *granarius* (Linn.) with special reference to the effects of temperature and humidity on the rate of their development. Indian J Entomol. 11:143-201.

Maghirang, E.B., Dowell, F.E., Baker, J.E., and Throne, J.E. 2003. Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance

spectroscopy. Trans. ASAE 46:1277-1282.

Mark, H.L. 2001. Qualitative near-infrared analysis. Ch.13. In: Near-Infrared Technology in the Agriculture and Food Industries, 2nd ed. Williams P.C. and Norris K., eds., American association of Cereal Chemists, Inc., MN. 233-238 pp.

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

Osborne, B.G., Fearn, T., and Hindle, P.H. 1993. Theory of near infrared spectrometry. Ch.2. In: Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, 2nd ed. Longman Scientific & Technological, Essex, UK, 13-35 pp.

Perez-Mendoza, J., Throne, J.E., Dowell, F.E., and Baker, J.E. 2003. Detection of insect fragments in wheat flour by near-infrared spectroscopy. J. Stored Prod. Res. 39:305-312.

Ridgway, C., Chambers, J., and Cowe, I.A. 1999. Detection of grain weevils inside single wheat kernels by a very near-infrared two-wavelength model. J. Near Infrared Spectosc. 7: 213-221.

Ridgway, C. and J. Chambers. 1996. Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. J. Sci. Food Agric. 71: 251-264.

Schwartz, B.E., and Burkholder, W.E. 1991. Development of the granary weevil (Coleoptera: curculionidae) on barley, corn, oats, rice, and wheat. J. Econ. Ent. 84: 1047-1052.

Shenk, J.S., Workman Jr., J.J., and Westerhaus, M.O. 1992. Application of NIR spectroscopy to agricultural products. In: Handbook of Near-Infrared Analysis.

Burns, D.A. and Ciurczak, E.M. (eds.), Marcel Dekker, Inc., New York, 383-431 pp.

Singh, N.B., and Sinha, R.N. 1977. Carbohydrate, lipid and protein in the developmental stages of *Sitophilus oryzae* and *Sitophilus granarius* (Coleoptera: Curculionidae). Ann. Entomol. Soc. Am. 70(1): 107-111.

Wehling, R.L., Wetzel, D.L., and Pedersen J.R. 1984. Stored wheat insect infestation related to uric acid as determined by liquid chromatography. J. Assoc. Off. Agric. Chem. 67: 644-647.

CHAPTER 3

Prediction Degree of Extrusion-Cooking of Corn Meal by Near-Infrared Reflectance Spectroscopy

Abstract

Near-Infrared (NIR) spectroscopy has been used to predict degree of cook in products produced by HTST extrusion of corn meal. Corn meal was cooked with a Wenger TX-57 twin screw extruder using screw speeds ranging from 250 to 350 rpm, and moisture contents ranging from 13-20%, providing a wide range of pressures and shear conditions in the extruder barrel. Extruded samples were analyzed for various physical properties that relate to degree of cooking, including water absorption index (WAI), water solubility index (WSI), viscosity properties as measured with a Rapid ViscoAnalyzer (RVA), hardness and fracturability as measured by Texture Profile Analysis (TPA). Samples were ground and their NIR reflectance spectra obtained with a Foss/NIRSystems 6500 spectrometer over a range of 1100-2500 nm. After separating samples into calibration and validation sets, multiple linear regression (MLR) and partial least squares (PLS) regression were used to develop NIR calibrations for measuring the various indicators of degree of cook. Calibrations developed were then applied to samples in the validation set. In general, correlations with r-value> 0.95 were achieved between the NIR and laboratory values. RPD values, which compare the standard error of prediction to the standard deviation of the reference data, ranged from 5.3 to 6.3 for the various parameters (except for hardness, and

trough viscosity), indicating that the NIR measurement should be useful in quality control applications.

Introduction

Extrusion is widely used in manufacturing many cereal-based products ranging from snacks to pet food. Corn is widely used to produce extruded snack foods (e.g., corn curls). The physical and chemical properties of the extrudate change when the material leaves the extruder, caused by the difference in temperature and pressure between the extruder and its ambient surroundings (Stauffer, 1993). Thermal and mechanical energy from extrusion process causes cleavage of hydrogen bonds between starch molecules and between starch and bound water, which can be measured by NIR (Osborne, 2007; Lee, 2007). The extent of physical changes to the native starch is a major factor influencing the rheology of the developing fluids and is also related to structure creation in the products. It is, therefore, an important variable to control during extrusion in order to obtain the desired physical characteristics in the products. There is no practical method for directly measuring degree of cook. Indirect methods of measuring functional properties related to degree of cook have been studied as previously indicated in the literature review.

Only a limited number of studies have been reported on rapid methods which can measure the dependent variables related to starch structure that result from extrusion cooking. NIR was found to be the most practical technique for the development of an on-line system to monitor starch-based extrusion processes (Scotter and Millar, 2004). The energy input to the material in the extruder as a result of high pressure and temperature indicates the degree of transformation in the starch fraction. Researchers have reported some applications of NIR for rapid quality control and process monitoring during extrusion processing in food (Ben-Hdech et al, 1993; Guy et al, 1996; Evans et al, 1999; Apruzzese et

al, 2000; Sahni et al, 2004; Dodds and Health, 2005) and other industries (Rohe et al, 1998; Tumuluri et al, 2004; Swarbrick et al, 2005).

In this study, the application of NIR spectroscopy for predicting starch changes during extrusion cooking was investigated. Objectives of this research project included (1) studying the feasibility of using NIR spectroscopy to predict the degree of cook of corn meal extruded under different extrusion conditions, (2) investigating the relationships between NIR spectra and physical properties of extrudates, and (3) developing NIR calibrations for predicting physicochemical parameters using MLR and PLS regression algorithms.

Materials and Methods

Raw Material

Fine corn meal was obtained from Trujillo & Sons Inc. (Miami, FL) and used for sample preparation.

Extrusion Process

High-temperature short-time extrusion cooking was conducted with a corotating twin-screw extruder (model TX-57, Wenger Manufacturing, Sabetha, KS). The length-to-diameter ratio was 8:1, and die opening was 3.96 mm. Extrusion feed rate was 90 kg/hr. A preconditioner was used at a speed of 150 rpm to convey samples to the extruder barrel. Temperature was set to be constant at 35^{o}C at the first head and 120^{o}C at the second head. The head pressure was approximately 500 PSI. The operating conditions were varied for moisture content and screw speed. Extruder water flow rates were 0.06, 0.09, 0.11, 0.15, and 0.19 kg/min as determined by a flow meter, resulting in feed moisture content at levels of 13,

15, 16, 18, and 20% (wb), respectively. Screw speeds were adjusted to 250, 275, 300, 325, and 350 rpm for each moisture level. The extrusion process was started from low to high screw speed with each moisture level, i.e., 250 rpm from high moisture (20%) to low moisture (18, 16, 15, and 13%, respectively). The screw speed was then changed to 275 rpm and from low moisture (13%) to high moisture (15, 16, 18, and 20%), and vice versa. All other extrusion parameters were held constant throughout the experiment. The experiment was run in three replicates (total= 5H5H3= 75 samples). Extrusion was allowed to reach a steady state. Extrudates were then collected and cooled on trays at ambient temperature (25^{o}C) for 1 hr. The extruder was allowed at least 5 min to achieve equilibrium between each run. Seventy five extrusion runs were completed in one day. Extrudates from each run were held in plastic bags and coded as shown in Table 1, and then air-dried overnight. Extrudates were ground in a laboratory impact grinder (A-10 analytical mill, Tekmar Co., Cincinnati, OH) to pass a U.S. No. 60 screen and stored in Ziploc7 bags at -18^{o}C for further analyses. The moisture contents of the ground extrudates were determined in duplicate using an AOAC air-oven procedure (1984).

Table 1: Coded samples of extrudates.

Run No.	Sample ID	Screw speed (rpm)	Moisture (%wb)	Rep	Run No.	Sample ID	Screw speed (rpm)	Moisture (%wb)	Rep
1	AE1	250	20	1	39	CB2	300	15	2
2	AD1	250	18	1	40	CA2	300	13	2
3	AC1	250	16	1	41	DA2	325	13	2
4	AB1	250	15	1	42	DB2	325	15	2
5	AA1	250	13	1	43	DC2	325	16	2
6	BA1	275	13	1	44	DD2	325	18	2
7	BB1	275	15	1	45	DE2	325	20	2
8	BC1	275	16	1	46	EE2	350	20	2
9	BD1	275	18	1	47	ED2	350	18	2
10	BE1	275	20	1	48	EC2	350	16	2
11	CE1	300	20	1	49	EB2	350	15	2
12	CD1	300	18	1	50	EA2	350	13	2
13	CC1	300	16	1	51	AE3	250	20	3
14	CB1	300	15	1	52	AD3	250	18	3
15	CA1	300	13	1	53	AC3	250	16	3
16	DA1	325	13	1	54	AB3	250	15	3
17	DB1	325	15	1	55	AA3	250	13	3
18	DC1	325	16	1	56	BA3	275	13	3
19	DD1	325	18	1	57	BB3	275	15	3
20	DE1	325	20	1	58	BC3	275	16	3
21	EE1	350	20	1	59	BD3	275	18	3
22	ED1	350	18	1	60	BE3	275	20	3
23	EC1	350	16	1	61	CE3	300	20	3
24	EB1	350	15	1	62	CD3	300	18	3
25	EA1	350	13	1	63	CC3	300	16	3
26	AE2	250	20	2	64	CB3	300	15	3
27	AD2	250	18	2	65	CA3	300	13	3
28	AC2	250	16	2	66	DA3	325	13	3
29	AB2	250	15	2	67	DB3	325	15	3
30	AA2	250	13	2	68	DC3	325	16	3
31	BA2	275	13	2	69	DD3	325	18	3
32	BB2	275	15	2	70	DE3	325	20	3
33	BC2	275	16	2	71	EE3	350	20	3
34	BD2	275	18	2	72	ED3	350	18	3
35	BE2	275	20	2	73	EC3	350	16	3
36	CE2	300	20	2	74	EB3	350	15	3
37	CD2	300	18	2	75	EA3	350	13	3
38	CC2	300	16	2					

Reference Analyses

Water Absorption (WAI) and Water Solubility (WSI) Indices

Water absorption index is the weight of the gel (g) obtained per gram of dry sample. It was determined by following the procedure of Anderson et al. (1969). A 2.5 g ground sample was placed into a 50 ml round centrifuge tube. Water was added (30 ml), a stopper was placed in the tube, and then shaken vigorously for 30 sec to suspend the sample before incubating in a water bath (30^{o}C) for 30 min. The suspension was stirred intermittently over a 30-min period, and then centrifuged at 3000 g for 10 min. The supernatant was carefully poured into a tared evaporating dish, and the remaining gel was weighed.

$$\text{WAI} = \frac{(\text{Weight of gel + tube}) - (\text{Weight of tube})}{\text{Sample dry weight}}$$

Water solubility index is the amount of solids recovered by evaporating the supernatant from the water absorption test, expressed as percentage of dry solids in the sample. The supernatant was dried in a air oven at 105^{o}C for 8 hr and then weighed. WSI was calculated as follows:

$$\text{WSI} = \frac{(\text{Weight of container + dried supernatant}) - (\text{Weight of container})}{\text{Sample dry weight}}$$

For WAI and WSI, each extruded sample was analyzed in triplicate.

Viscosity Properties

RVA measurements were performed using a Rapid ViscoAnalyzer (RVA-Series 4, Newport Scientific Pty. Ltd., Warriewood, Australia) along with the accompanying software Thermocline for Windows. A suspension was prepared by weighing approximately 4 g (14% moisture basis) of ground extrudate into an RVA canister, and adding water

approximately 25 g to yield 14% dry basis. The mixture was then manually mixed by capping with a No.12 stopper, and shaking vigorously for 15 sec to ensure homogeneity of the sample mixture. The sample was then inserted into the RVA tower and analyzed using the following heating profile; 25°C for 2 min, heated to 95°C at 14°C/min, held at 95°C for 3 min, cooled to 25°C at 14°C/min, and held at 25°C for 5 min. Each experiment was initiated by a 10 sec, 960 rpm mixing period, followed by a 160 rpm paddle speed for the remainder of data collection. Each extruded sample was run in duplicate. Values for viscosity were reported in units termed Arapid visco units@ (RVU). RVA parameters were recorded as follows; cold viscosity (peak viscosity in 0.2-2 min), peak (maximum viscosity recorded during ramp to 95°C), trough (lowest viscosity at 95°C), final viscosity (viscosity at the finish of the test or cool paste viscosity at 25°C), breakdown (peak-trough; an indication of the breakdown in viscosity of the paste during the 95°C holding period), and setback (final viscosity-trough; a gauge of the texture of the starch paste). Generally, product that is less cooked or less sheared has a higher final viscosity value, and this can be a straightforward measure of the cook (Whalen, 1999).

Texture Profile Analysis (TPA)

Hardness and fracturability of extrudates were measured on a Texture Analyzer (TA-XT2i, Texture Technologies Corp., Scarsdale, N.Y.) with a cylinder probe (TA-42), and equipped with a Texture Expert software program (Version 5.16). The parameters used for operation were set as follows: Pretest = 2 mm/s, test speed = 1 mm/s, post test = 2 mm/s, distance = 70% strain, force = 0.98 N. The higher the value of the maximum peak force required, meaning that a higher force was required to breakdown the extrudates, the higher

the hardness of the extrudates. The fracturability of the extrudates was the first peak force measured, and is associated with the crispness of those extrudates (Charles et al, 2001 and Li et al, 2005). The lower the force measured at first crackdown, the higher the fracturability or crispness of the extrudates. Extrudates from each run were measured ten times.

Thermal Properties of Extrudates

DSC thermograms were performed with a Pyris 1 Scanning Calorimeter (Perkin-Elmer Corp., Norwalk, CT) equipped with an Intracooler II System and Pyris thermal analysis software (Perkin-Elmer). Starch and water mixtures (1:3, w/w) were sealed in aluminum pans and equilibrated at room temperature for 2 hr before analysis. An empty pan was used as a reference. The samples were heated at 5^{o}C/min over a temperature range of 15-125^{o}C to obtain the endotherms. The specific heat of gelatinization (ΔH, J/g) was indicated from the area under the curve. Onset temperature (To), peak temperature (Tp), and conclusion temperature (Tc) were provided from the software.

NIR Analysis

Ground extrudates were scanned using a Foss/NIRSystems Model 6500 spectrometer (Silver Spring, MD) equipped with a rotating drawer attachment. For instrument control and calibration development, the spectrometer was interfaced to a personal computer running the Near Infrared Spectral Analysis Software (NSAS) package (version 3.53) provided by NIRSystems. About 4 g of samples were placed in a small ring cup. Diffuse reflectance readings were referenced to those from a ceramic disk. The spectra were collected over a wavelength range of 400 to 2500 nm. Each spectrum represented the average of 32 scans, and was recorded as log (1/R) at 2 nm increments. Samples were

randomly selected by the NSAS software program for a calibration set (n= 45) for the purpose of developing quantitative models, and the rest (n=30) were used as an independent data set for validating the performance of the models. Log 1/R spectra were transformed using first and second derivatives with different combinations of wavelength segments and gaps (10, 20, 30, and 40 nm) to enhance absorption peaks and remove baseline difference.

Multiple linear regression (MLR)equations were developed to relate log 1/R, first, and second derivative spectra to laboratory reference values using forward stepwise and best possible regression algorithms. To avoid the interference of water bands with the spectral bands of other constituents, wavelengths of 1901 to 1979 nm were removed from the spectra (Ghadian and Wehling, 1997). Therefore, in the present study, spectral regions from 1100 to 1900 and 1980 to 2500 nm were used in developing calibration models. The optimum number of wavelengths to include in the calibration equations was determined by comparing regression results for multiple correlation coefficient (r), which should approach 1, and standard error of calibration (SEC), which should be low. Each model was tested for its ability to predict degree of cook expressed by each reference property using separate validation samples. Performance of models was compared by evaluating 1) correlation coefficient (r), which indicates the closeness of fit between the NIR and reference data over the range of composition, (which should be close to 1), 2) root mean square of the differences (RMSD) between NIR and reference values (Delwiche and Reeves, 2004) which should be low, 3) slope of the relationship between the values of the reference data and the values predicted by the NIR model (Liu and Han, 2006) which should be close to 1, and 4)

relative predictive determinant (RPD) or standardized performance, which is defined as the ratio of the standard deviation of the constituent=s reference values to the standard error of prediction (SEP) of the validation samples (Williams, 2001). In this study, RMSD was used instead of SEP. RMSD is not corrected for bias and is an appropriate estimator for the random error assuming that there is no systematic error of the measurement (Hruschka, 2001). The higher the RPD value, the better is the performance of the model. A value of 1.0 represents a lack of modeling power since the variability attributable to the NIR model has the same level as the variance of the naturally occurring constituent. Values ∃2.5 indicate that the NIR model may be suitable for screening and breeding programs. Values above 5.0 are potentially useful in quality control.

The second algorithm used to develop calibration equations was partial least squares (PLS) regression, which is a data compression technique that provides good relationships in cases of collinearity (Sahni et al, 2004). The number of PLS factors reported was the minimum required to give the best classification results. The use of PLS algorithms allows information from all wavelengths in the entire spectrum to be included in the calibration, rather than information from only a few wavelengths (Wehling, 1998). The optimum number of PLS terms (factors) was selected based on the lowest standard errors of cross validation (SECV) with the fewest terms, along with the R value obtained from the regressions. Each model was verified for its ability to predict degree of starch cooking using independent validation samples. The predictive abilities of the models were compared as previously described for MLR.

Calibrations were developed from raw and derivatized spectral data over the regions

1100 to 1900 and 1980 to 2500 nm. Calibrations were also built with truncated wavelength ranges of 1100-1800 nm and 1100-1900 nm, corresponding to the lower half of the near-infrared region, where the reflected energy signal is relatively strong (Delwiche et al, 1996).

Results and Discussion

Properties of Extrudates Determined by Reference Methods

The minimum, maximum, and mean values and standard deviation (SD) of each constituent are listed in Table 2. These statistics were calculated separately for the calibration and validation sets. For all constituents except WSI and set back, the range of values for the validation set fell within the calibration set range. In addition, the wide ranges in each set for each constituent, and the small differences in means, ranges, and SD between the calibration set and the validation set, revealed that both sets could be considered representative of the same overall diversity.

Differential Scanning Calorimetry (DSC) Analysis

Upon extrusion cooking, the typical native starch gelatinization endotherm disappeared (results not shown). This indicates no ungelatinized starch or crystalline forms were present after extrusion.

Table 2: Summary of reference values for constituents of extruded samples in calibration and validation sets.

Constituent, Units[a]	Minimum	Maximum	Mean	Standard Deviation
WAI, g/g				
Calibration	2.32	7.47	5.27	1.64
Validation	2.32	7.45	5.11	1.57
WSI, %				
Calibration	20.81	72.64	43.91	15.87
Validation	21.24	72.97	45.82	14.89
Viscosity, RVU				
Cold				
Calibration	13.29	123.46	55.83	31.02
Validation	18.71	120.53	52.09	29.89
Peak				
Calibration	8.88	124.09	50.51	32.73
Validation	13.21	122.00	46.92	31.54
Trough				
Calibration	1.80	21.58	9.60	4.80
Validation	2.80	20.89	9.27	4.43
Final				
Calibration	7.54	62.97	32.56	15.96
Validation	10.92	60.83	30.50	14.61
Breakdown				
Calibration	5.71	102.34	40.97	28.28
Validation	8.81	101.12	37.71	27.52
Setback				
Calibration	4.37	41.50	23.00	11.55
Validation	7.00	41.58	21.23	10.59
Hardness, N				
Calibration	34.56	115.68	59.74	25.09
Validation	34.63	115.01	55.63	25.60
Fracturability, N				
Calibration	18.41	114.99	48.74	24.92
Validation	20.54	113.33	46.14	25.91

[a] n = 45 samples in the calibration set, 30 samples in the validation set. RVU = Rapid Visco Units.

Spectral Characteristics of Extrudates

Figure 1 illustrates some examples of log 1/R spectra of ground extrudates from different extrusion parameters. The upper spectrum is the average from the parameters of lowest screw speed and highest feed moisture content (lowest degree of processing), the intermediate spectra are the averages from lower feed moisture contents at the lowest screw speed, and the lowest spectrum is the average from the highest screw speed at the highest feed moisture content. When derivative transformations were applied, the absorption bands were sharpened and distinct. Fig. 2 and 3 show some differences in absorption band intensities of first and second derivative spectra, respectively. Wavelengths at 1152 and 1414 nm are due to second overtones of C-H stretch and deformations, which are associated with CH_3 and CH_2 . Wavelengths at 1432 and 1922 nm, are most likely associated with moisture content variation; however, other differences in spectral intensity can be observed at 1902, 2012, 2238, 2256, 2284 and 2430 nm. These wavelengths are attributable to an O-H stretch first overtone, C-O stretch, and C-H stretch deformation, which are associated with starch (Osborne et al, 1993). Osborne (1996) also reported absorptions at 1410 and 1430 nm are changed in wheat extrusion due to an O-H stretch first overtone in starch or from water bound to starch. The branched molecules of amylopectin at wavelength around 2280 nm were reported to break down through shearing forces (Lee, 2007).

Calibration and Prediction Results

Calibration models were developed from log 1/R, 1st derivative and 2nd derivative spectra using both MLR and PLS regression. Statistical results of calibration and validation

samples for WAI, WSI, textural properties (hardness and fracturability), and RVA pasting properties using MLR are summarized in Tables 3 and 4, respectively. In MLR equations, a maximum of four wavelengths was used, to prevent overfitting. Table 5 and 6 show the results of PLS modeling of those ten constituents of extrudates. For both algorithms, calibrations with low SEC values did not always provide low RMSD=s for the validation set.

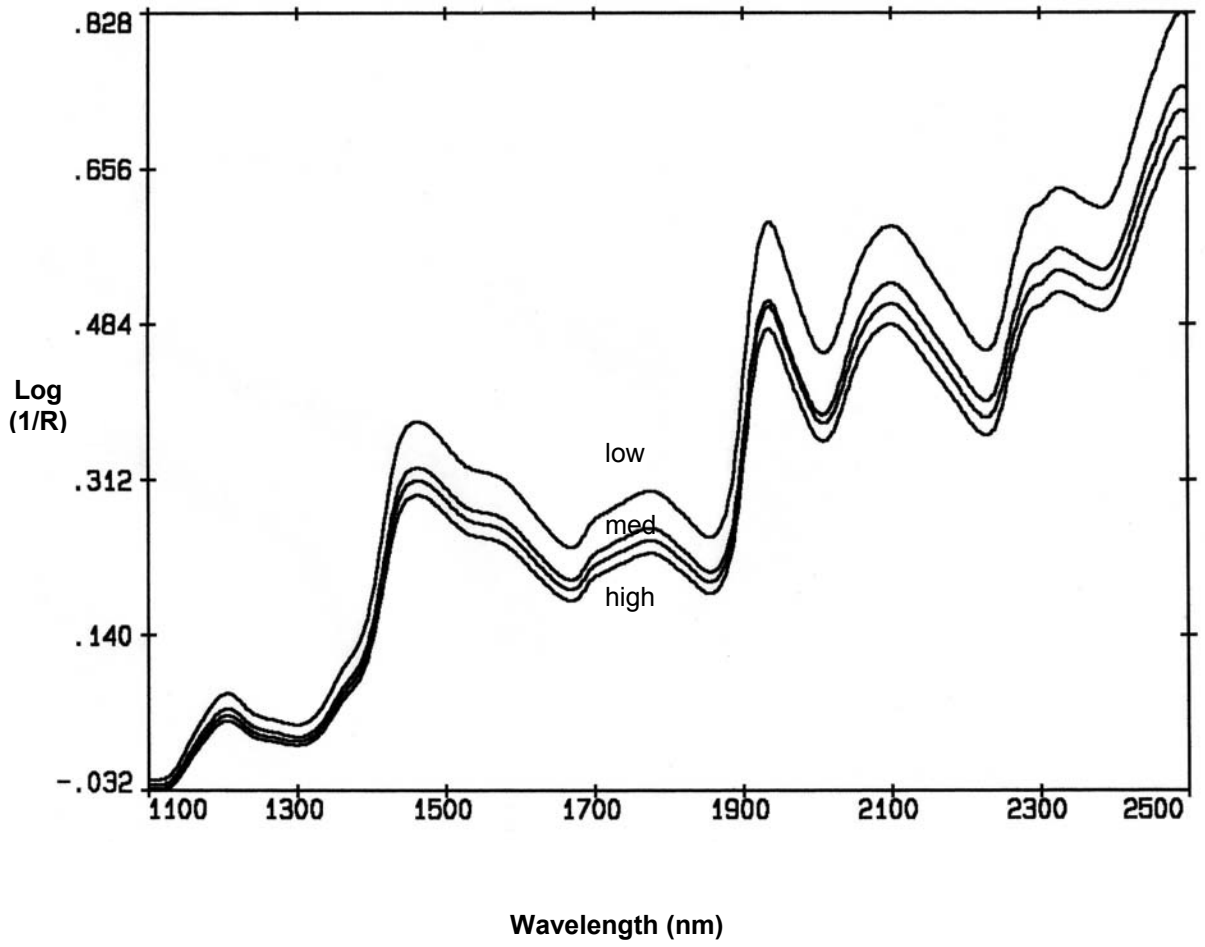

Fig. 1. Average spectra (log 1/R) of extrudates at low, medium and high degree of processing.

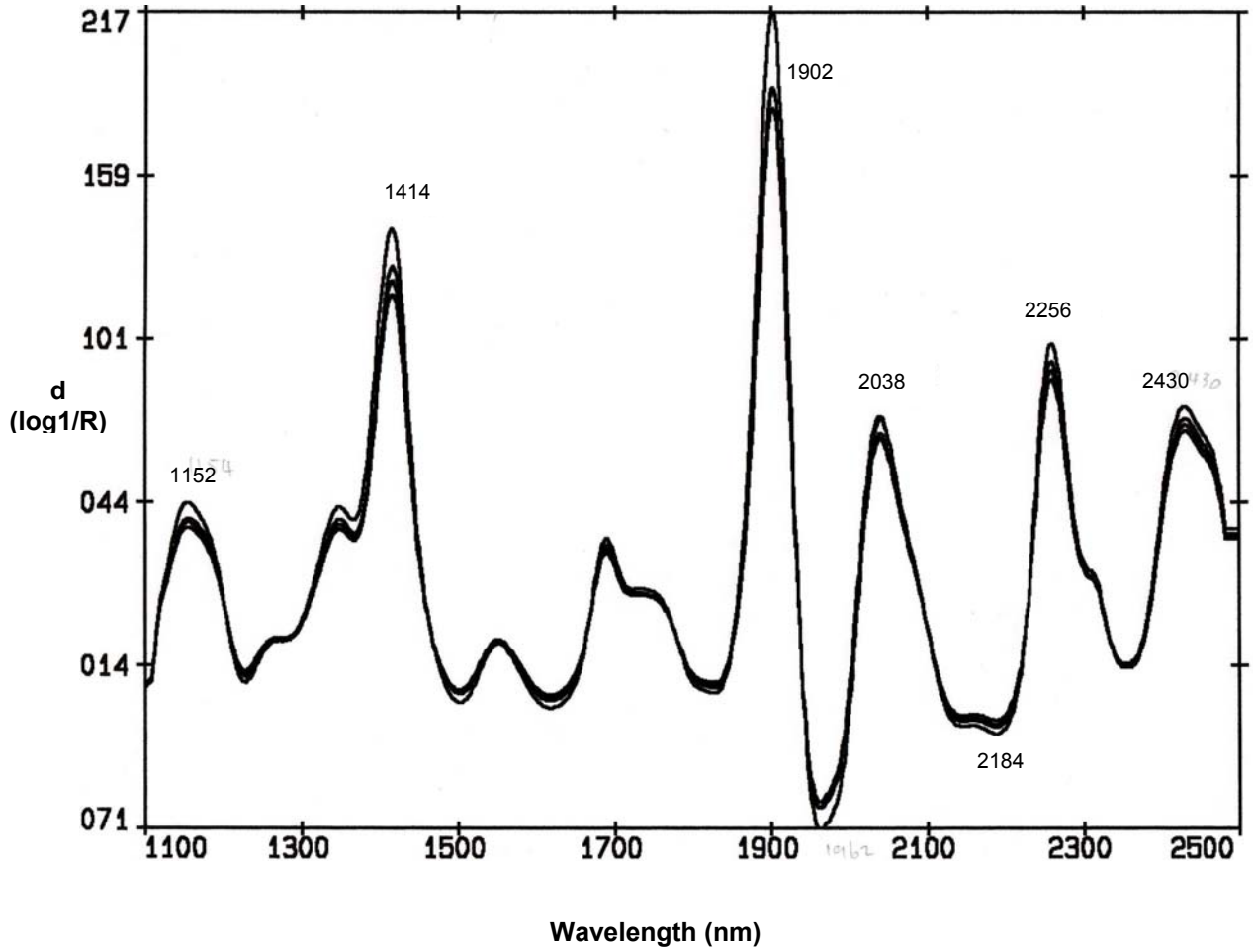

Fig. 2. Average spectra of extrudates at low, medium, and high degree of processing after treatment with first derivative.

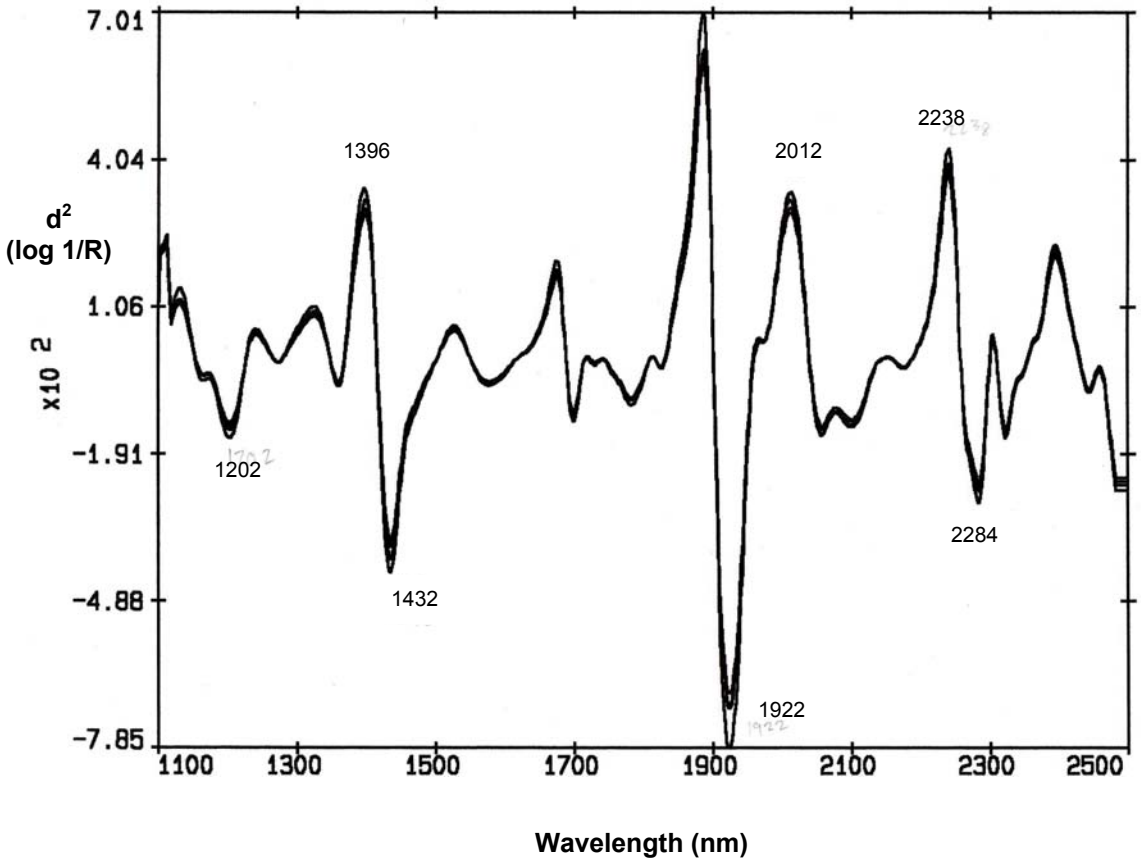

Fig. 3. Average spectra of extrudates at low, medium, and high degree of processing after treatment with second derivative.

Table 3: Calibration and validation statistics for prediction by each constituent=s (WAI, WSI, texture analysis) best model using multiple linear regression.

Constituents	Spectral treatment	Calibration		Validation			
		r	SEC	r	RMSD	Slope	RPD
WAI (g/g)	**log 1/R** (1388, 1468, 2040, 2168)[a]	0.979	0.35	0.963	0.44	0.972	3.57
	1st derivative[b] (1122, 2106, 2222, 2300)	0.984	0.31	0.981	0.32	1.045	4.91
	2nd derivative[c] (1752, 1788, 2140, 2436)	0.985	0.30	0.984	0.28	1.031	5.61
WSI (%)	**log 1/R** (1218, 1470, 2038, 2416)	0.981	3.20	0.972	3.88	0.947	3.84
	1st derivative[d] (1118, 1480, 2220, 2300)	0.977	3.54	0.983	2.80	0.998	5.32
	2nd derivative[c] (1124, 1570, 2406, 2482)	0.980	3.35	0.979	3.43	0.918	4.34
Hardness (N)	**log 1/R** (1218, 1408, 2000, 2220)	0.927	9.86	0.928	9.75	1.056	2.63
	1st derivative[d] (1110, 1204, 1852, 2330)	0.950	8.19	0.963	7.24	1.010	3.54
	2nd derivative[e] (1268, 1366, 2310, 2424)	0.949	8.27	0.965	6.97	1.025	3.67
Fracturability (N)	**log 1/R** (1218, 1408, 2222, 2398)	0.955	7.75	0.974	6.84	1.145	3.79
	1st derivative[f] (1520, 1580, 2138, 2300)	0.976	5.68	0.987	4.34	1.043	5.97
	2nd derivative[d] (1236, 1566, 1814, 2190)	0.967	6.65	0.986	4.68	1.029	5.54

[a] Wavelengths (nm).
[b] Segment = 40 nm, gap = 20 nm.
[c] Segment = 10 nm, gap = 20 nm.
[d] Segment = 30 nm, gap = 40 nm.
[e] Segment = 20 nm, gap = 10 nm.
[f] Segment = 10 nm, gap = 10 nm.

Table 4: Calibration and validation statistics for prediction of RVA parameters=s best model using multiple linear regression (MLR).

Constituents (RVU)	Spectral treatment	Calibration		Validation			
		r	SEC	r	RMSD	Slope	RPD
Cold viscosity	**log 1/R** (1138, 1332, 2298, 2348)[a]	0.965	8.59	0.976	6.64	0.987	4.50
	1st derivative[b] (1210, 1550, 2000, 2104)	0.970	7.86	0.963	9.16	0.899	3.26
	2nd derivative[c] (1234, 1692, 2032, 2428)	0.969	8.05	0.977	6.80	0.941	4.40
Peak viscosity	**log 1/R** (1144, 1332, 2298, 2370)	0.969	8.51	0.982	6.09	0.976	5.18
	1st derivative[c] (1210, 1666, 2226, 2310)	0.975	7.57	0.982	6.74	0.926	4.68
	2nd derivative[d] (1232, 1692, 2034, 2360)	0.977	7.27	0.988	5.55	0.922	5.68
Trough	**log 1/R** (1110, 1894, 2054, 2494)	0.900	2.19	0.950	1.39	0.987	3.19
	1st derivative[d] (1110, 1666, 2098, 2226)	0.933	1.81	0.921	1.88	0.853	2.36
	2nd derivative[e] (1236, 1696, 2178, 2356)	0.920	1.97	0.917	1.82	0.900	2.43
Final viscosity	**log 1/R** (1138, 1408, 2244, 2324)	0.975	3.71	0.983	2.76	0.985	5.29
	1st derivative[f] (1112, 1664, 2400, 2478)	0.985	2.93	0.978	3.18	0.940	4.59
	2nd derivative[d] (1234, 1482, 2258, 2426)	0.987	2.68	0.976	3.38	0.946	4.32
Breakdown	**log 1/R** (1138, 1332, 2298, 2390)	0.968	7.47	0.979	5.66	0.986	4.86
	1st derivative[d] (1210, 1668, 2228, 2314)	0.976	6.46	0.975	6.81	0.928	4.04
	2nd derivative[c] (1234, 1692, 2032, 2166)	0.975	6.59	0.975	6.49	0.938	4.24
Setback	**log 1/R** (1118, 1410, 2242, 2300)	0.985	2.11	0.980	2.19	0.978	4.84
	1st derivative[f] (1562, 1880, 2104, 2226)	0.990	1.69	0.978	2.38	0.945	4.45
	2nd derivative[d] (1166, 1268, 2194, 2430)	0.992	1.54	0.981	2.13	0.951	4.97

[a] Selected wavelengths (nm).
[b] Segment = 20 nm, gap = 20 nm.
[c] Segment = 40 nm, gap = 10 nm.
[d] Segment = 30 nm, gap = 10 nm.
[e] Segment = 30 nm, gap = 40 nm.
[f] Segment = 20 nm, gap = 40 nm.

Table 5: Calibration and validation statistics for prediction by each constituent=s (WAI, WSI, texture analysis) best model using partial least squares (PLS) regression.

Constituents	Spectral treatment	Calibration		Validation			
		r	SECV	r	RMSD	Slope	RPD
WAI (g/g)	**log 1/R** (WL=1100-1900, 1980-2500, F=5)[a]	0.984	0.31	0.976	0.35	1.012	4.49
	1st derivative[b] (WL=1100-1800, F=4)	0.984	0.31	0.982	0.30	1.020	5.23
	2nd derivative[c] (WL=1100-1900, F=4)	0.987	0.28	0.982	0.29	1.017	5.41
WSI (%)	**log 1/R** (WL=1100-1900, F=5)	0.978	3.48	0.979	3.12	0.992	4.77
	1st derivative[d] (WL=1100-1900, F=5)	0.979	3.45	0.983	2.82	1.024	5.28
	2nd derivative[c] (WL=1100-1900, F=5)	0.985	2.87	0.985	2.74	0.941	5.43
Hardness (N)	**log 1/R** (WL=1100-1900, F=5)	0.939	9.18	0.957	7.86	1.095	3.26
	1st derivative[e] (WL=1100-1900, 1980-2500, F=9)	0.989	3.99	0.979	6.66	1.031	3.84
	2nd derivative[e] (WL=1100-1900, 1980-2500, F=7)	0.987	4.45	0.973	6.85	1.048	3.74
Fracturability (N)	**log 1/R** (WL=1100-1900, F=6)	0.958	7.66	0.987	4.86	1.103	5.33
	1st derivative[f] (WL=1100-1900, 1980-2500, F=3)	0.960	7.27	0.988	4.74	1.099	5.47
	2nd derivative[g] (WL=1100-1900, 1980-2500, F=4)	0.961	7.20	0.987	4.67	1.079	5.55

[a] Spectral range (nm) and number of factors used.
[b] Segment = 40 nm, gap = 20 nm.
[c] Segment = 10 nm, gap = 20 nm.
[d] Segment = 20 nm, gap = 10 nm.
[e] Segment = 10 nm, gap = 10 nm.
[f] Segment = 10 nm, gap = 40 nm.
[g] Segment = 20 nm, gap = 20 nm.

Table 6: Calibration and validation statistics for prediction of RVA parameters (RVU)=s best model using partial least squares regression (PLSR).

Constituents	Spectral treatment	Calibration		Validation			
		r	SEC	r	RMSD	Slope	RPD
Cold viscosity	**log 1/R** (WL=1100-1800, F=3)[a]	0.963	8.66	0.979	6.20	0.983	4.82
	1st derivative[b] (WL=1100-1900, F=6)	0.968	8.38	0.981	6.20	0.959	4.82
	2nd derivative[c] (WL=1100-1800, F=4)	0.976	7.11	0.985	5.65	0.939	5.29
Peak viscosity	**log 1/R** (WL=1100-1800, F=3)	0.966	8.76	0.983	5.97	0.988	5.28
	1st derivative[d] (WL=1100-1900, F=6)	0.976	7.68	0.984	6.37	0.927	4.95
	2nd derivative[c] (WL=1100-1800, F=4)	0.981	6.67	0.989	5.35	0.925	5.90
Trough	**log 1/R** (WL=1100-1900, 1980-2500, F=4)	0.901	2.18	0.950	1.39	0.963	3.19
	1st derivative[e] (WL=1100-1900, 1980-2500, F=3)	0.907	2.09	0.935	1.62	0.910	2.73
	2nd derivative[f] (WL=1100-1900, F=4)	0.924	1.93	0.943	1.53	0.908	2.90
Final viscosity	**log 1/R** (WL=1100-1900, F=3)	0.976	3.58	0.985	2.57	0.974	5.68
	1st derivative[g] (WL=1100-1800, F=3)	0.976	3.53	0.993	2.38	0.987	6.14
	2nd derivative[d] (WL=1100-1900, 1980-2500, F=3)	0.979	3.32	0.987	2.34	0.983	6.24
Breakdown	**log 1/R** (WL=1100-1900, F=3)	0.968	7.36	0.983	5.21	0.980	5.28
	1st derivative[f] (WL=1100-1900, 1980-2500, F=3)	0.972	6.93	0.984	5.02	1.026	5.48
	2nd derivative[h] (WL=1100-1900, F=3)	0.970	7.15	0.981	5.48	0.971	5.02
Setback	**log 1/R** (WL=1100-1900, F=6)	0.989	1.82	0.986	1.87	0.966	5.66
	1st derivative[e] (WL=1100-1800, F=6)	0.991	1.69	0.987	1.76	0.977	6.02
	2nd derivative[g] (WL=1100-1900, F=6)	0.993	1.50	0.987	1.78	0.972	5.95

[a] Spectral range (nm) and number of factors used.
[b] Segment = 40 nm, gap = 30 nm.
[c] Segment = 10 nm, gap = 10 nm.
[d] Segment = 40 nm, gap = 10 nm.
[e] Segment = 20 nm, gap = 20 nm.
[f] Segment = 10 nm, gap = 20 nm.
[g] Segment = 40 nm, gap = 40 nm.
[h] Segment = 30 nm, gap = 20 nm.

Water Absorption and Water Solubility Indices

For WAI, optimal model conditions occurred when using a linear combination of second-derivative terms (segment= 10 nm and gap= 20 nm), which produced a prediction error of 0.28 g/g, correlation coefficient of 0.98, and RPD= 5.61. Prediction results were not improved when PLS regression was used. From observation of the MLR equations, the significant wavelengths were 1752, 1788, 2140, and 2436 nm. Wavelengths at 1752 and 1788 nm are due to the first overtone C-H stretch associated with CH_2. The wavelength at 2140 nm lies on the side of the starch combination band, and also responds to a combination of C-H stretch and C-C stretch corresponding to HC=CH structure associated with lipid. The wavelength at 2436 nm is associated with starch. A scatter plot of the modeled and reference values for WAI is shown in Fig. 4.

A second derivative coupled with the PLS algorithm provided the best prediction result for WSI. Surprisingly, the truncated wavelengths of 1100-1900 nm yielded better results than using wavelengths of 1100-1900 and 1980-2500 nm, with 5 PLS factors (RMSD= 2.74%, r-value = 0.99, and RPD= 5.43). Fig. 5 illustrates a linear relationship between predicted and reference WSI values. Weights, which indicate the degree that the variance has been used in the computation of the factors across the wavelength range, are more useful than loadings to interpret the process used to develop the calibration (Williams, 2001). Fig. 6 presents PLS weights from second derivative spectra at wavelengths from 1100 to 1900 nm. The absorption bands at 1342 and 1360 nm are due to C-H stretch and deformation combinations which are associated with CH_3 groups. The strong absorption bands at around 1620, 1742, 1806 and 1820 nm are assigned to C-H stretch, O-H stretch and C-O stretch which are associated with carbohydrates, including starch and cellulose

(Osborne et al, 1993 and Williams, 2001).

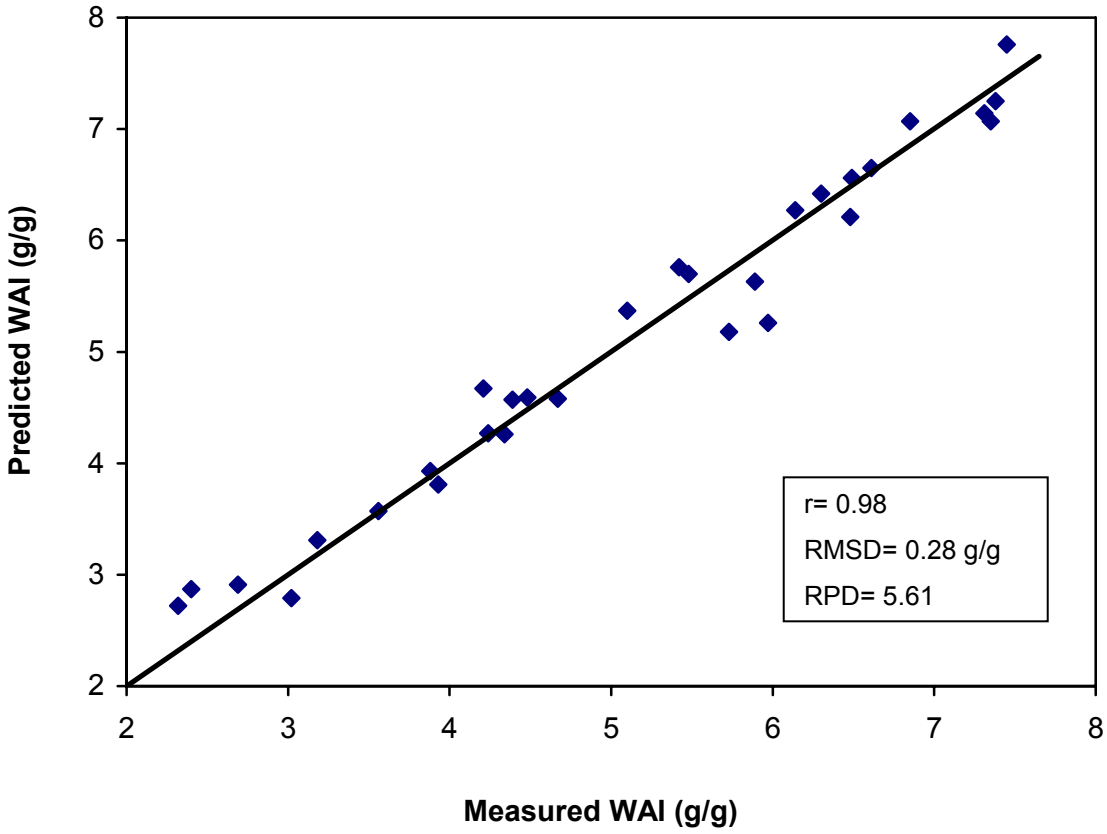

Fig. 4. Reference vs. NIR modeled values for water absorption index (WAI) of validation samples.

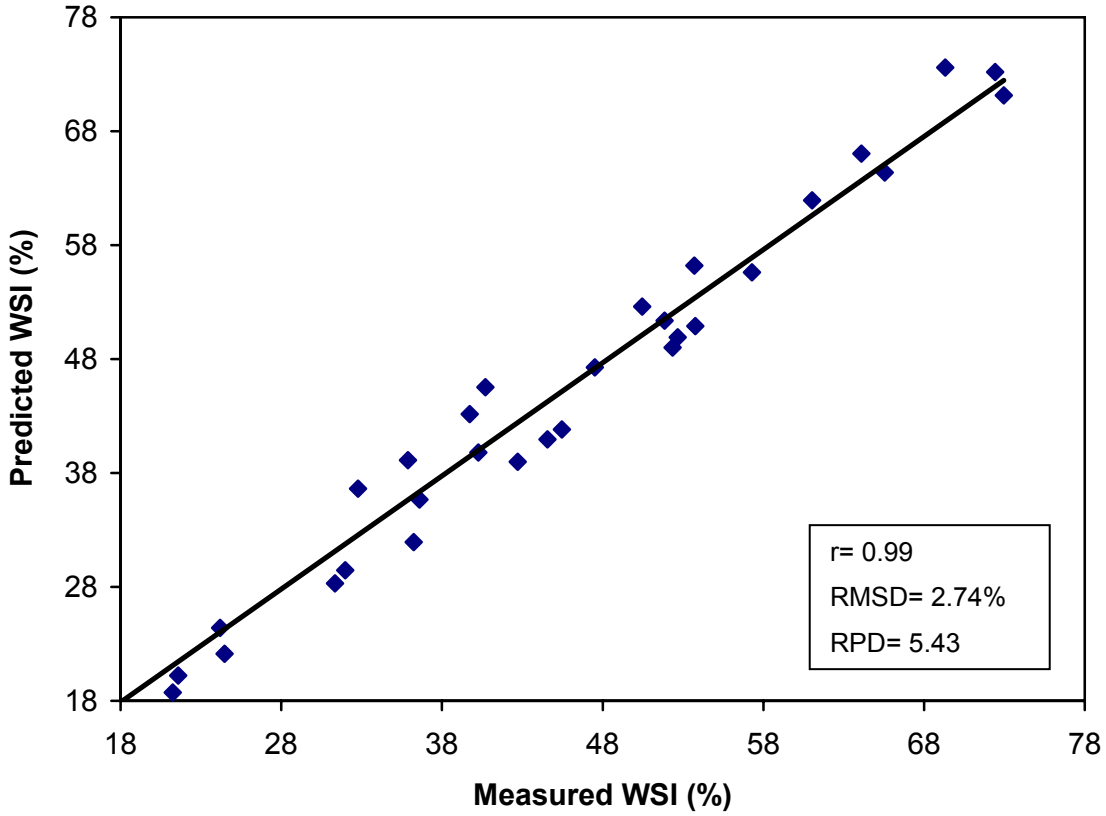

Fig.5. Reference vs. NIR modeled values for water solubility index (WSI) of validation samples.

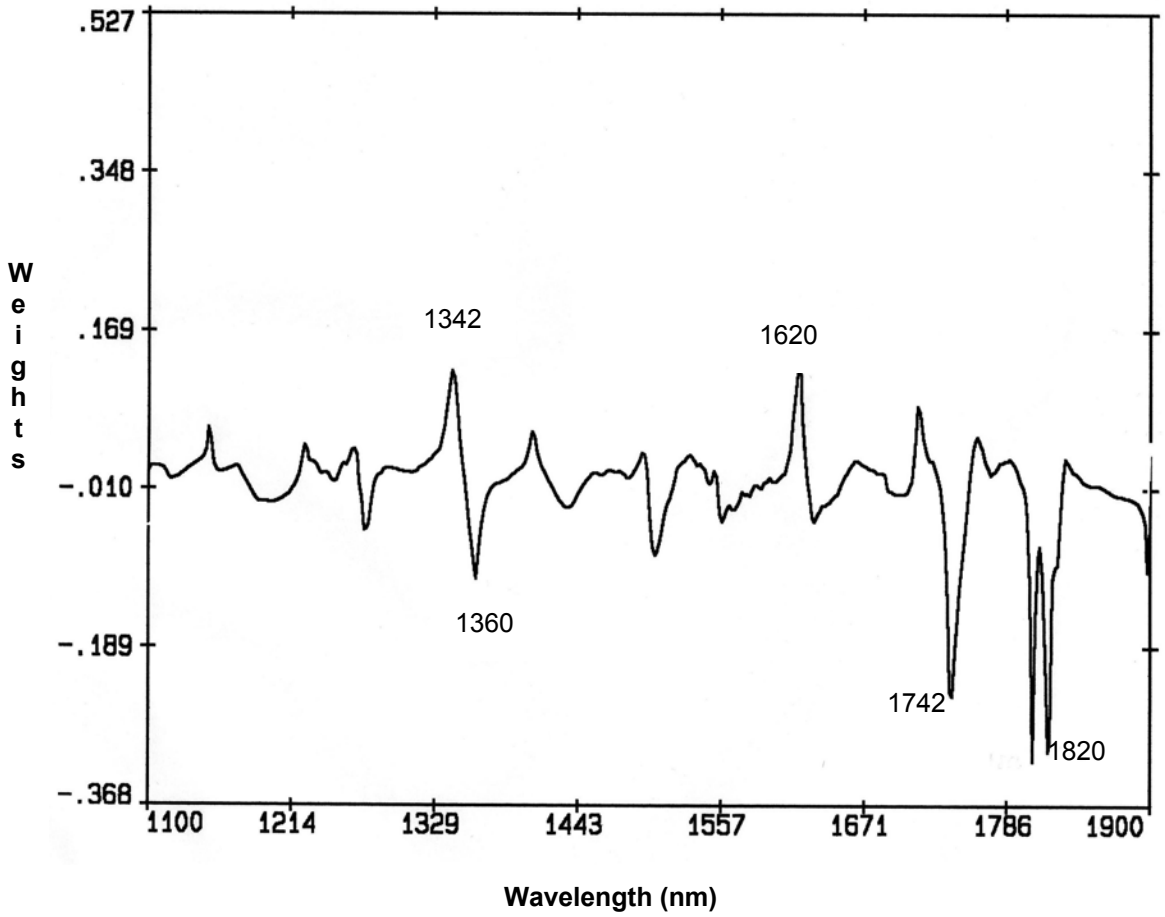

Fig. 6. PLS weights from second derivative spectra of extrudates from calibration model used to predict WSI.

Textural Properties

PLS regression of first derivative with the wavelength range of 1100-1900 nm and 1980-2500 nm with 9 PLS factors (segment= 10 nm and gap= 10 nm) yielded the best result for hardness between predicted values and measured values (r-value= 0.98 and RMSD= 6.66 N). When the reliability of the NIR determination was assessed with RPD, the value was relatively low (RPD= 3.84). However, it is still acceptable. Williams (2007) indicated that minimum acceptable values for the RPD is 3.0. Fig. 7 shows a scatter plot between predicted and NIR values for hardness. PLS weights from first derivative spectra showed peaks at around 1360, 1400 and 2280 nm (figure not shown) which are due to C-H stretch and deformation which correspond with CH_3 groups and CH_2 of starch and cellulose, and O-H stretch first overtone of starch (Osborne et al, 1993 and Williams, 2001). When using MLR regression of first derivative, NIR prediction for fracturability showed better results than for hardness (r-value= 0.99, RMSD= 4.34 N, and RPD= 5.97). The wavelengths chosen were 1520, 1580, 2138, and 2300 nm, which are attributed to O-H stretch first overtone, C-H and C-C stretches corresponding to CH_2 groups, respectively, which are associated with starch. A linear relationship between modeled and reference values for fracturability is illustrated in Fig. 8.

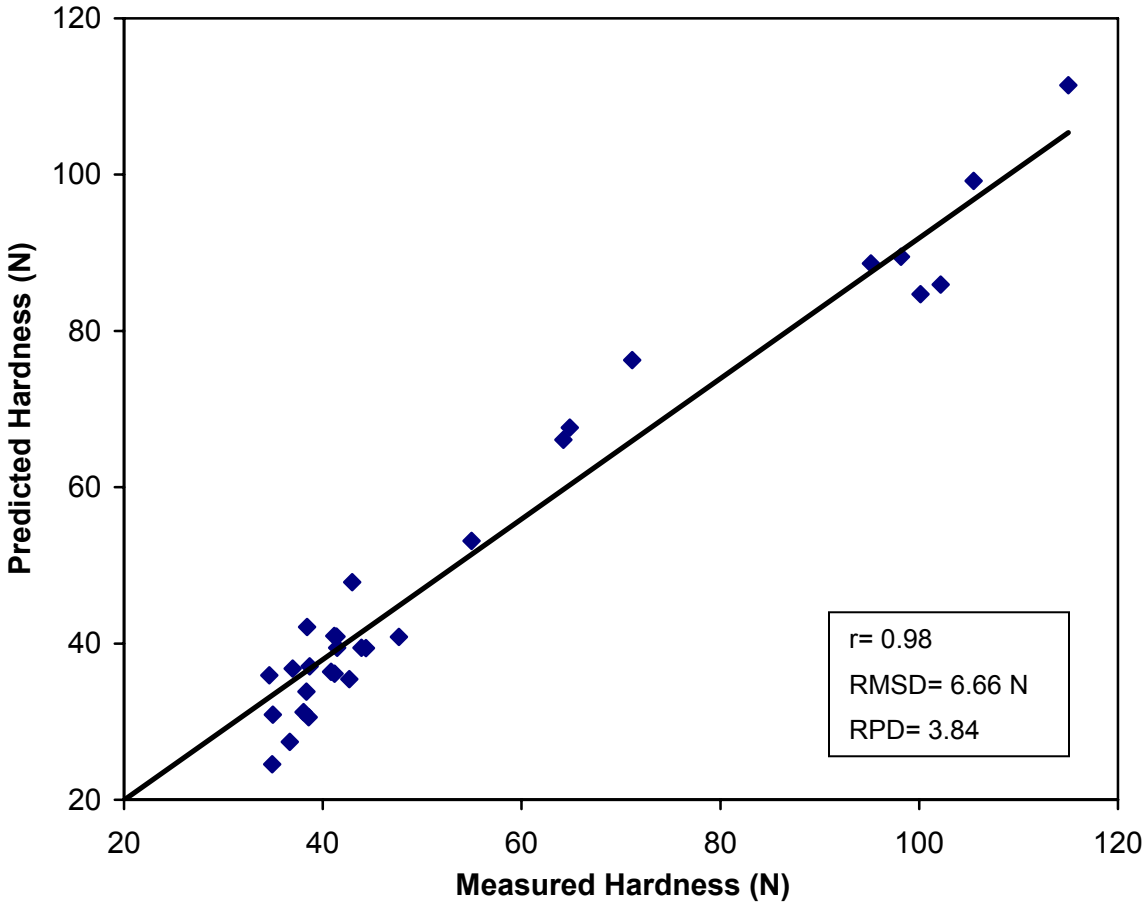

Fig. 7. Reference vs. NIR modeled values for hardness of validation samples.

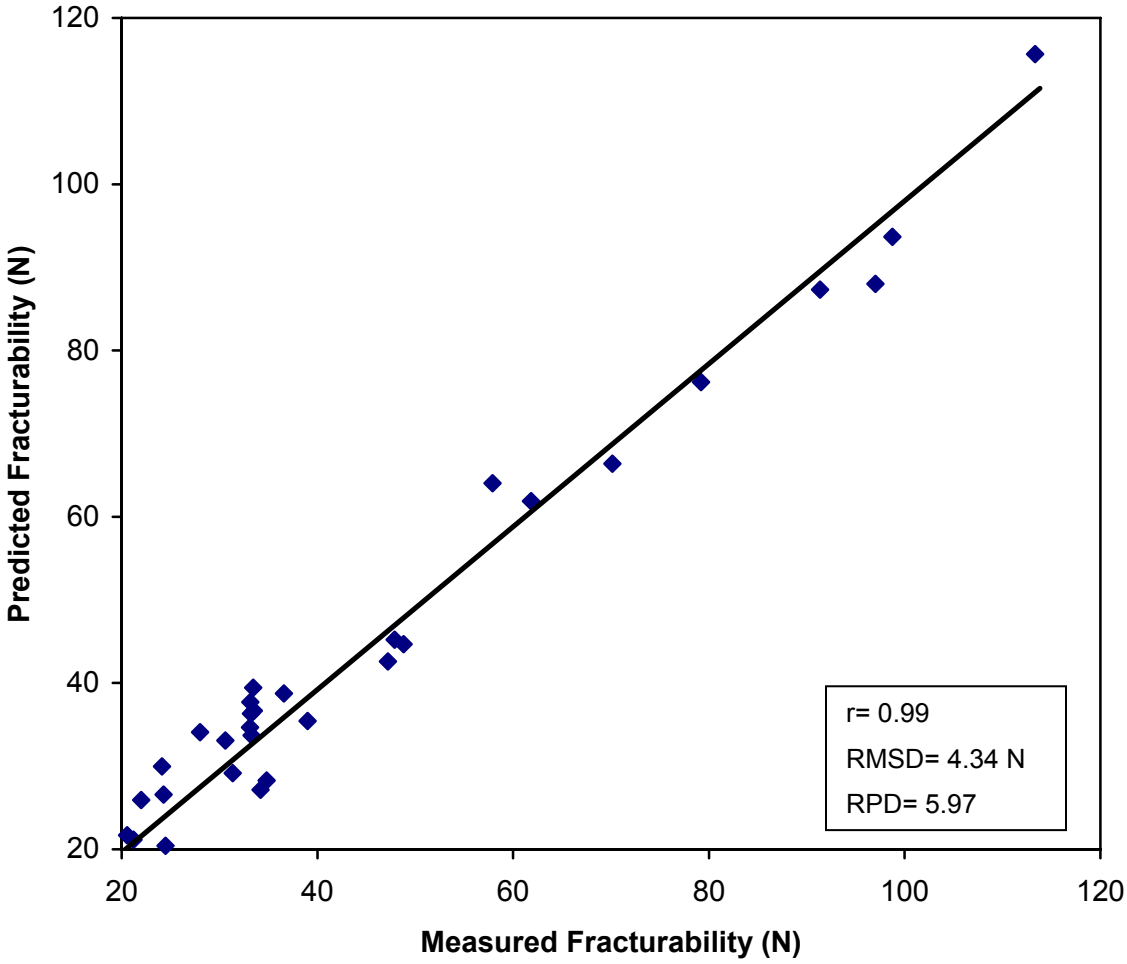

Fig. 8. Reference vs. NIR modeled values for fracturability of validation samples.

Viscosity Properties

For RVA constituents, decreasing moisture content resulted in a lowering of the pasting and a shift to cold viscosity with a decrease in the trough and setback (data not shown). These results are similar to those reported by Whalen et al (1999). NIR analysis gave better prediction, as expressed by low RMSD and high r-value, for cold, peak and final viscosities, compared to trough. For cold viscosity, the best prediction result was derived from PLS regression of second derivative using wavelengths of 1100-1800 nm with 4 factors (r-value= 0.99, RMSD= 5.65 RVU, and RPD= 5.29). Using the same treatment also provided the best prediction result for peak viscosity (r-value= 0.99, RMSD= 5.35 RVU, and RPD= 5.90). Scatter plots between NIR modeled and reference values of cold and peak viscosities are shown in Figs. 9 and 10, respectively. PLS weights for prediction of cold and peak viscosities are similar and are shown in Fig. 11. Strong absorption bands can be seen at wavelengths around 1128, 1200, 1390, 1426, and 1700 nm, which are assigned to C-H overtone, O-H overtone, C-H stretch and deformation (Osborne et al, 1993). These are associated with starch and cellulose (Williams, 2001). Unlike the results for cold and peak viscosities, a 3-factor PLS model developed from second derivative using the entire wavelength range of 1100-2500, excluding wavelengths from 1901-1979 nm, gave the best results for predicting final viscosity of extruded samples (r-value= 0.99, RMSD= 2.34 RVU, and RPD= 6.24). PLS and MLR analyses yielded similar results for predicting trough with values of r= 0.95, RMSD= 1.39 RVU, and RPD = 3.19. Figs. 12 and 13 present relationships between predicted and lab values of final and trough viscosities, respectively. Fig. 14 shows strong absorption bands (1624, 1696, 2188, 2260, and 2430 nm) by plotting PLS weights for final viscosity prediction. These are also due to C-H first overtone,

associated with O-H stretch, and C-H stretch, which correspond to starch (Williams, 2001).

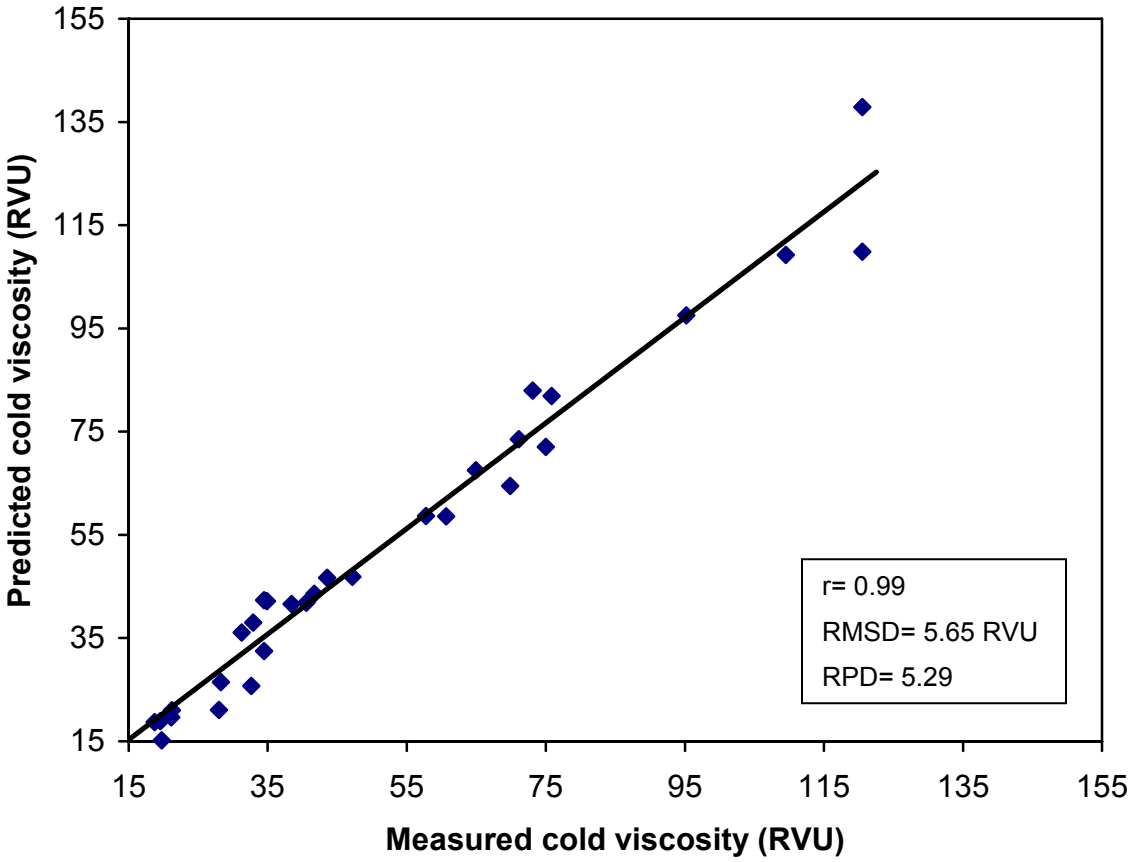

Fig. 9. Reference vs. NIR modeled values for RVA cold viscosity of validation samples.

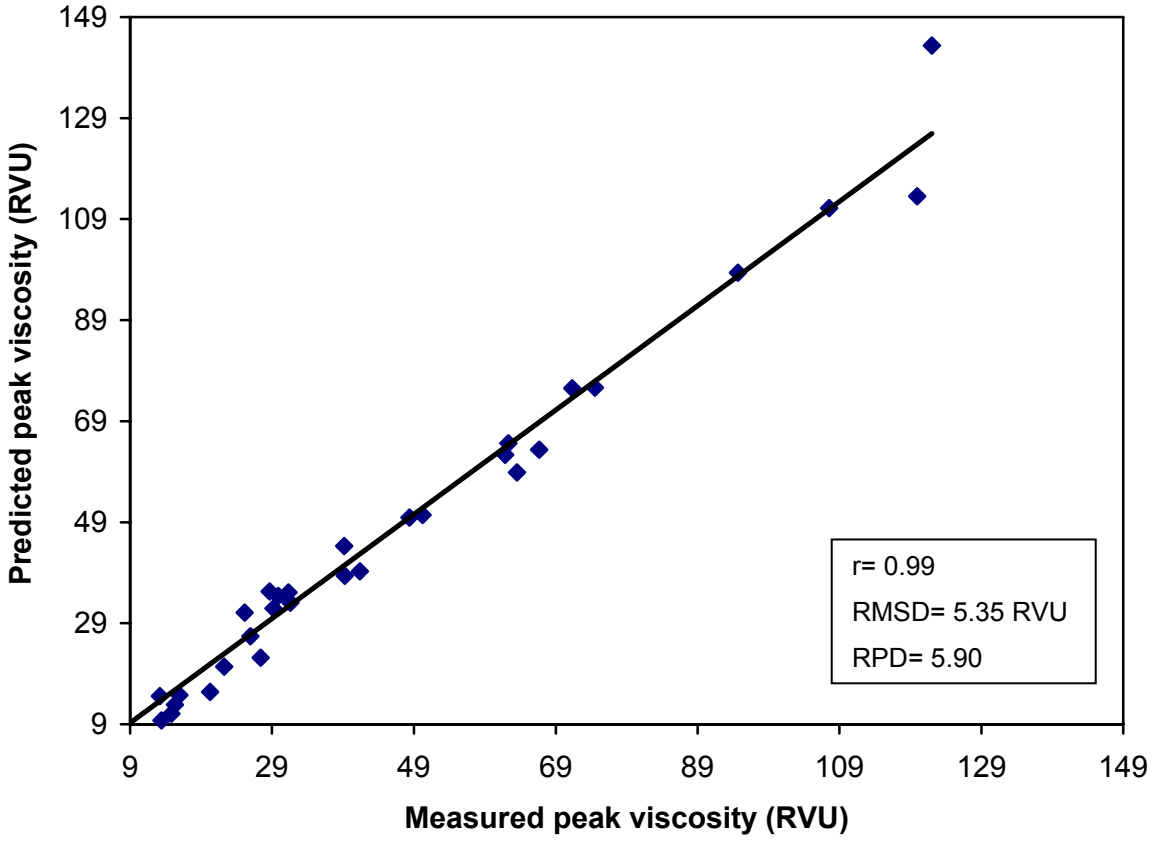

Fig. 10. Reference vs. NIR modeled values for RVA peak viscosity of validation samples.

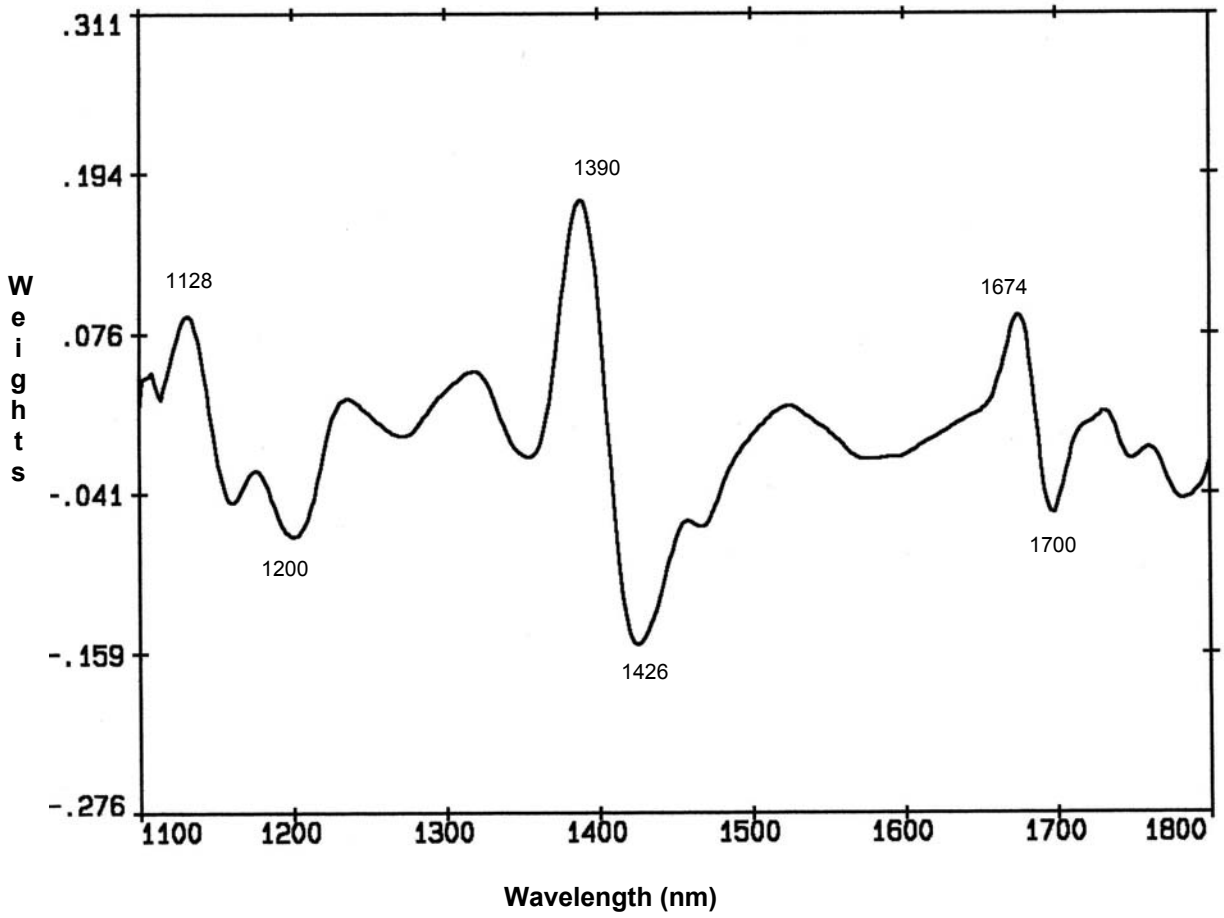

Fig. 11. PLS weights from second derivative spectra of extrudates used to develop calibration models for predicting cold and peak viscosities.

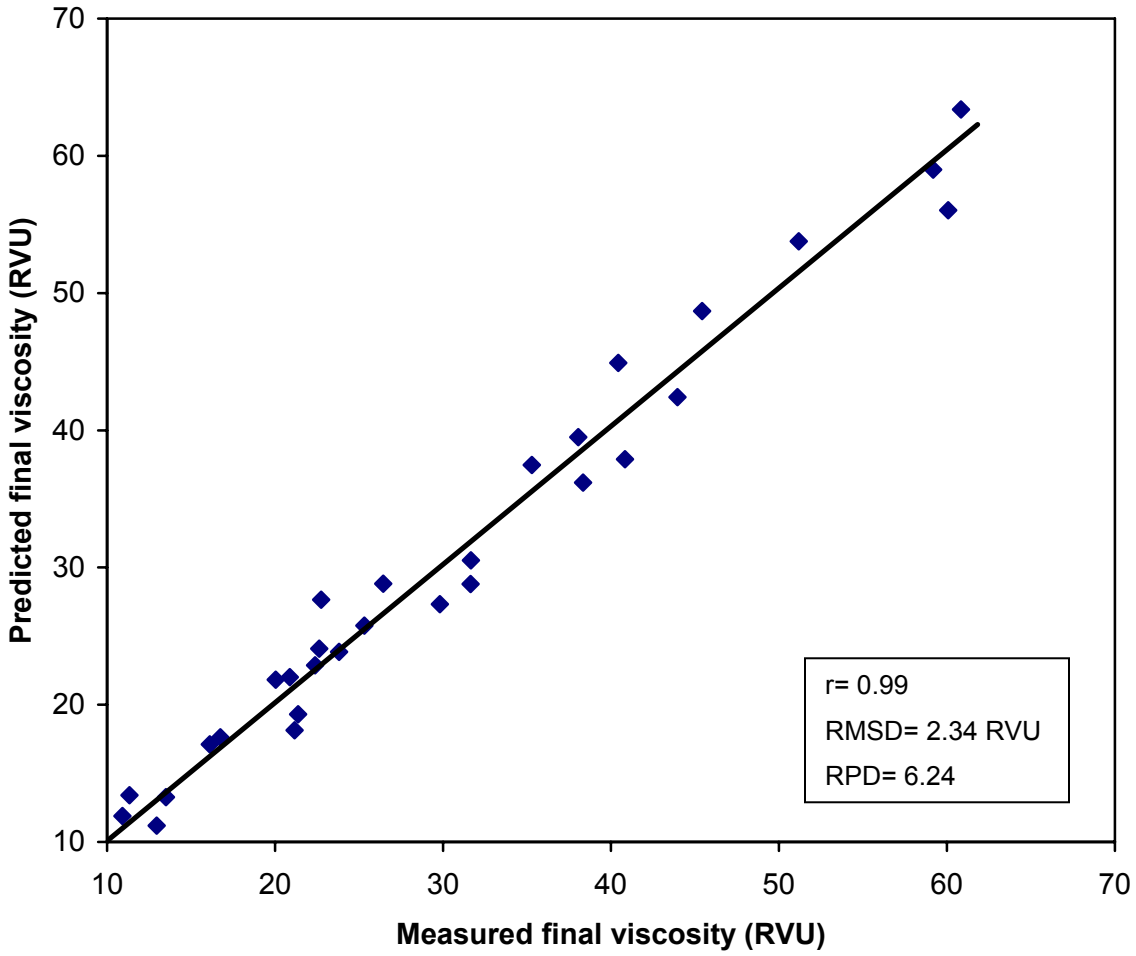

Fig. 12. Reference vs. NIR modeled values for RVA final viscosity of validation samples.

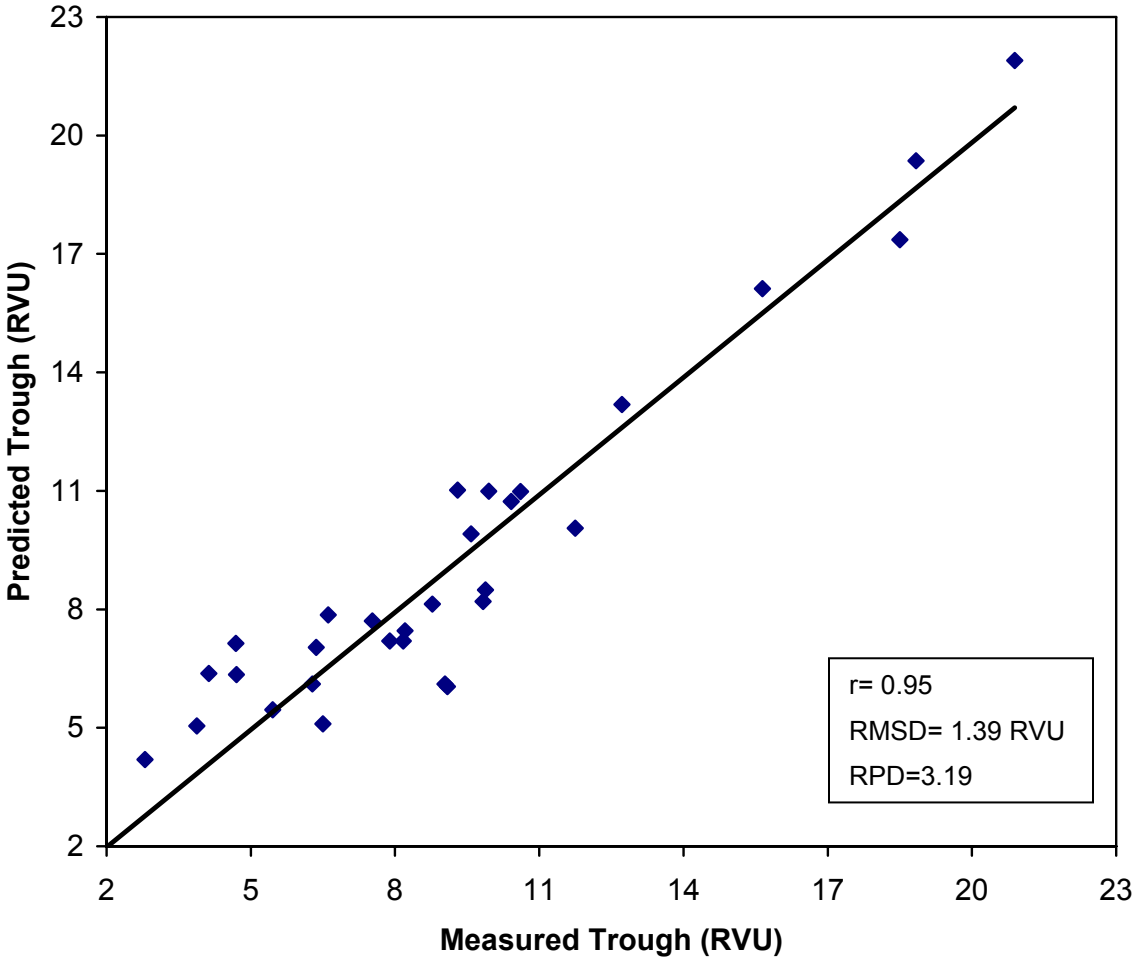

Fig. 13. Reference vs. NIR modeled values for RVA trough viscosity of validation samples.

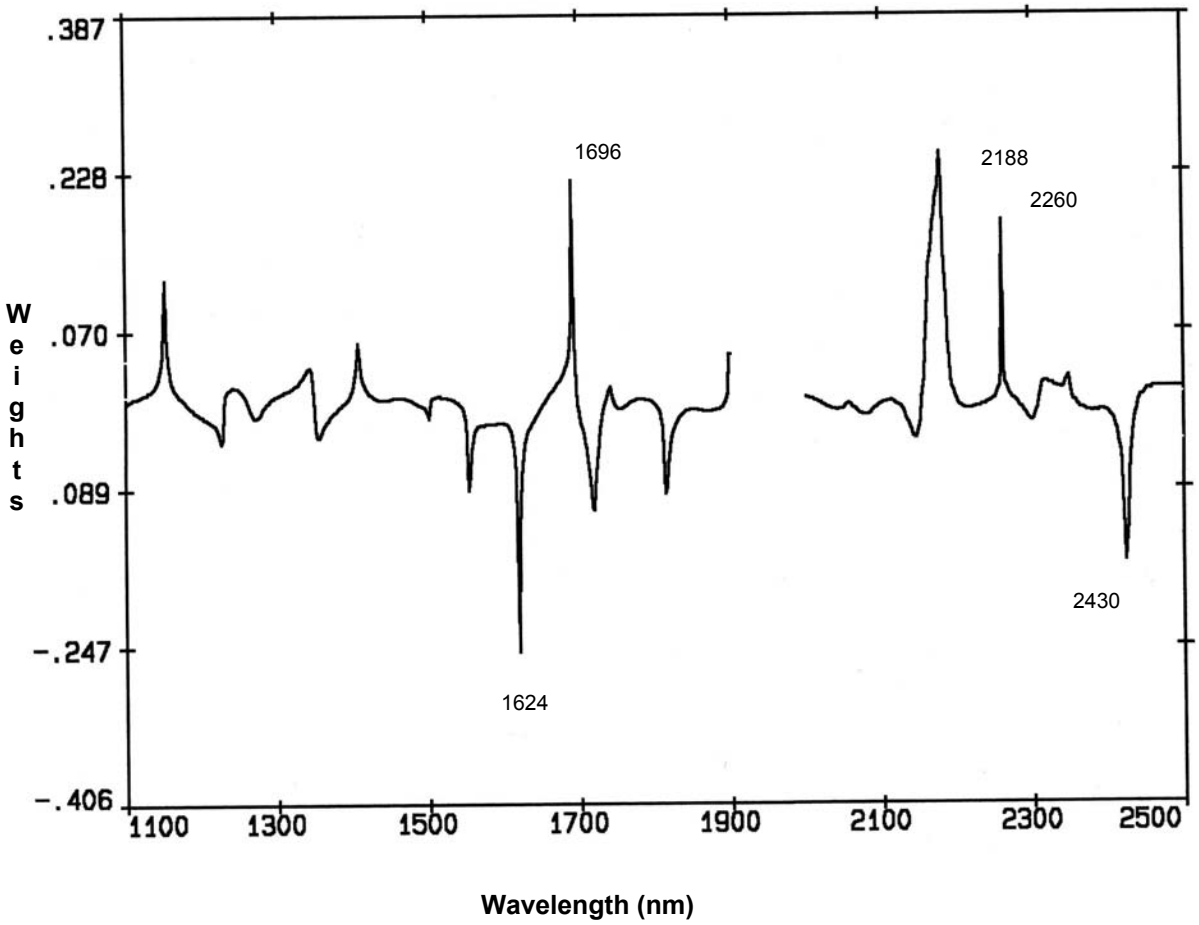

Fig. 14. PLS weights from second derivative spectra of extrudates from calibration modeled used to predict final viscosities (excluding wavelengths of 1901-1979 nm).

For RVA parameters of breakdown and setback, first derivative with PLS calibration yielded the best model. The best prediction result for breakdown, as shown by r-value= 0.98, RMSD= 5.02 RVU, and RPD= 5.48, were derived from the model with the wavelength ranged from 1100 to 2500 nm, leaving the wavelength from 1901 to 1979 nm, with 3 factors. For setback, the r, RMSD, and RPD values obtained (0.99, 1.76 RVU, and 6.02, respectively) indicate the best calibration model using the wavelengths ranging from 1100 to 1800 nm with 6 factors. Comparisons between modeled and reference values for breakdown and setback of validation samples are plotted in Fig. 15 and 16, respectively. Fig. 17 presents PLS weights from first derivative spectra of the breakdown. The absorption band at 1216 nm is due to a C-H second overtone associated with CH_2 group. Wavelengths at 1396, 2038, 2284, and 2440 nm are assigned to C-H stretch and deformation, and C-C stretch, which are related to CH_3 group, starch, and cellulose (Osborne et al, 1993 and Williams, 2001). PLS weights for setback are also associated with CH_2, CH_3, starch, and cellulose.

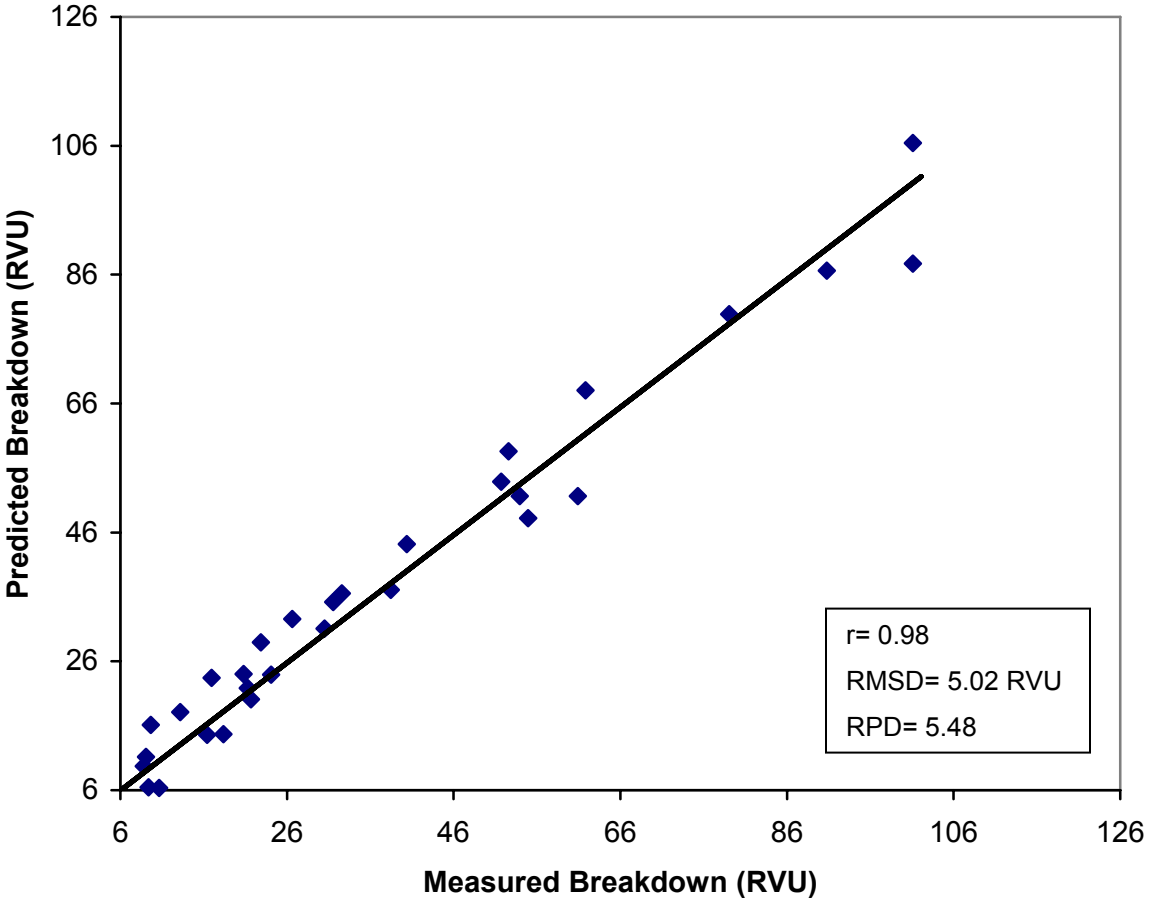

Fig. 15. Reference vs. NIR modeled values for RVA breakdown of validation samples.

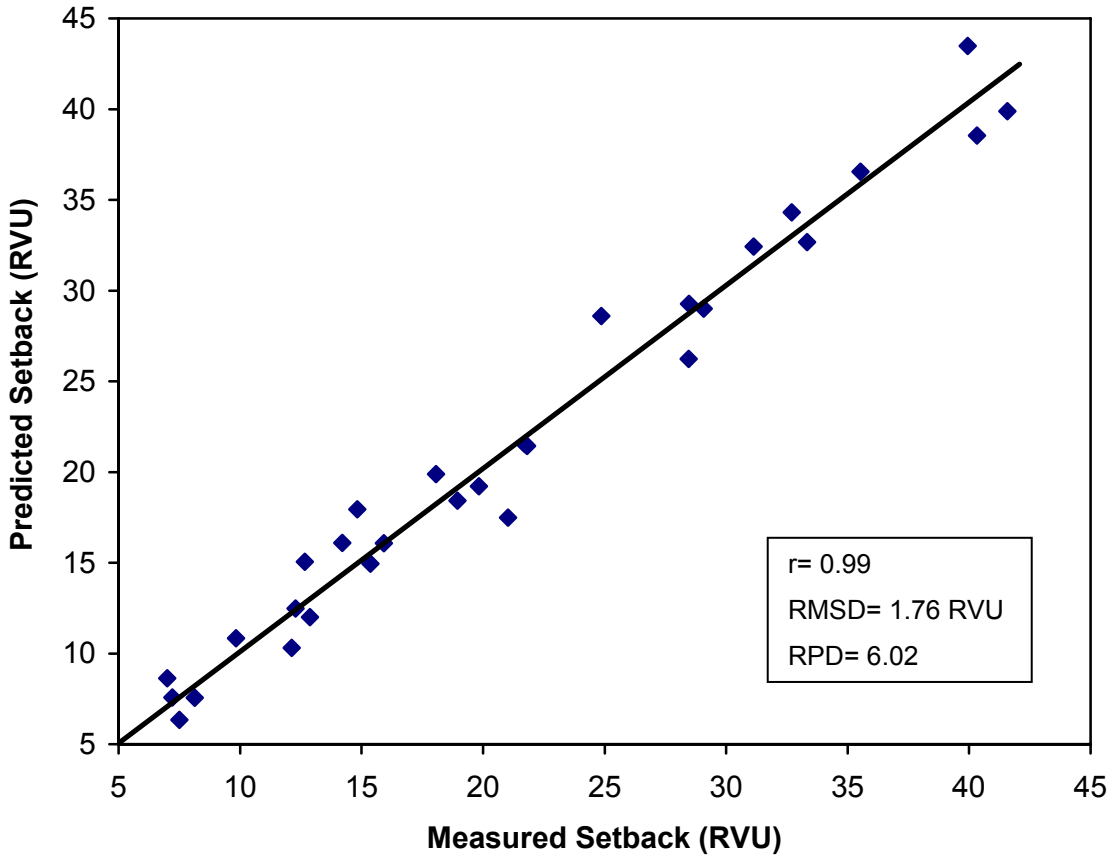

Fig. 16. Reference vs. NIR modeled values for RVA setback of validation samples.

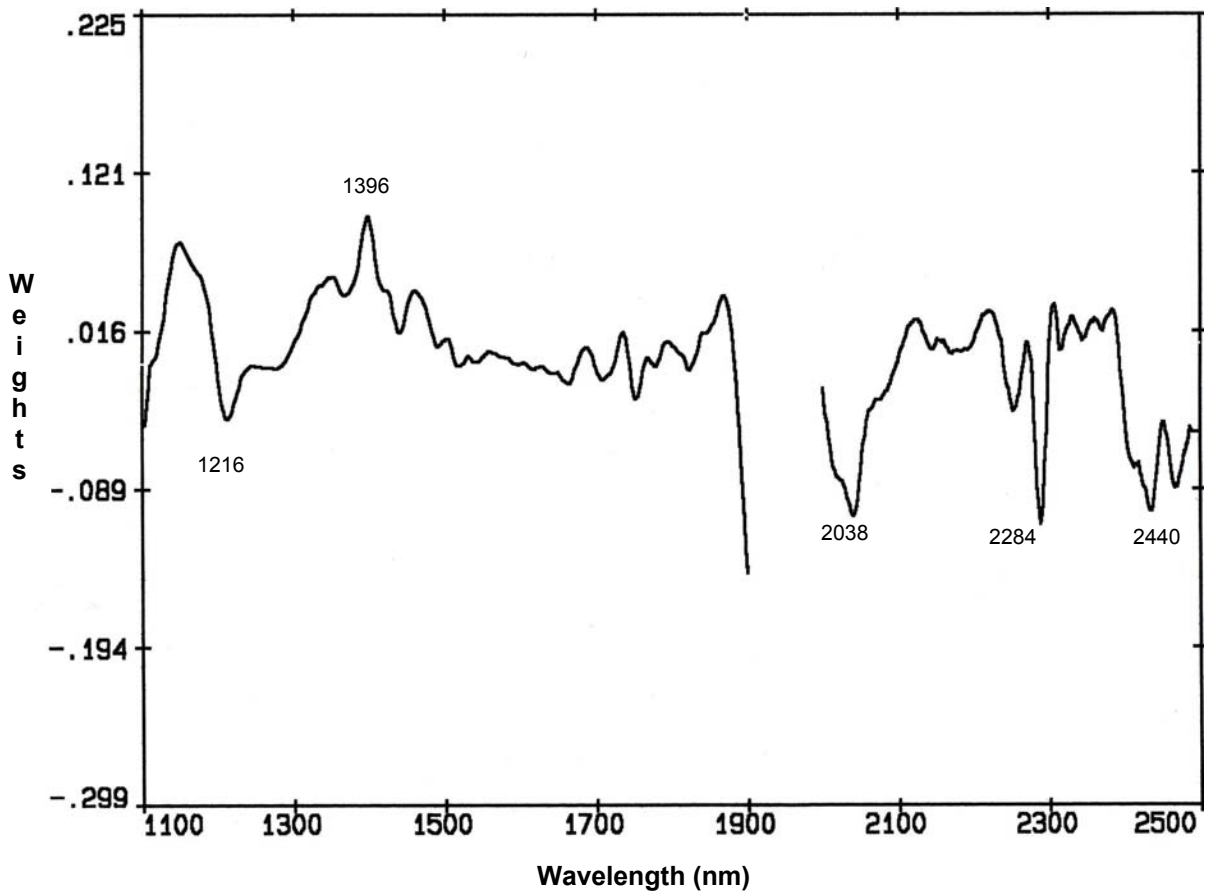

Fig. 17. PLS weights from first derivative spectra of extrudates from calibration modeled used to predict breakdown (excluding wavelengths of 1901-1979 nm).

References

Anderson, R.A., Conway, H.F., Pfeifer, V.F., and Griffin, E.L. 1969. Gelatinization of corn grits by roll- and extrusion-cooking. Cereal Sci. Today 14:4-12.

Apruzzese, F., Balke, S.T., and Diosady, L.L. 2000. In-line color and composition monitoring in the extrusion cooking process. Food Res. Int. 33:621-628.

Ben-Hdech, H., Gallant, D.J., Robert, P., and Gueguen, J. 1993. Use of near infrared spectroscopy to evaluate the intensity of extrusion-cooking processing of pea flour. Int. J. Food Sci. Tech. 28:1-12.

Charles, A.L., Ho, C.T., and Huang, T.C. 2001. Role of lipids in Taiwanese flaky snack. J. Food Lipids 8:115-130.

Delwiche, S.R. and Reeves, J.B.,III. 2004. The effect of spectral pre-tretments on the partial least squares modeling of agricultural products. J. Near Infrared Spectrosc. 12:177-182.

Dodds, S.A., and Health, W.P. 2005. Construction of an online reduced-spectrum NIR calibration model from full-spectrum data. Chemom. Intell. Lab. Syst. 76:37-43.

Evans, A.J., Huang, S., Osborne, B.G., Kotwal, Z., and Wesley, I.J. 1999. Near infrared on line measurement of degree of cook in extrusion processing of wheat flour. J. Near Infrared Spectrosc. 7:77-84.

Ghaedian, A.R. and R.L. Wehling. 1997. Discrimination of sound and granary-weevil-larva-infested wheat kernels by near-infrared diffuse reflectance spectroscopy. J. Assoc. Off. Agric. Chem. 80: 997-1005.

Guy, R.C.E., Osborne, B.G., and Robert, P. 1996. The application of near-infrared

reflectance spectroscopy to measure the degree of processing in extrusion cooking processes. J. Food Eng. 27: 241-258.

Hruschka, W.R. 2001. Data Analysis: Wavelength selection methods. Ch.3 In: Near-Infrared Technology in the Agriculture and Food Industries, 2nd ed. Williams P.C. and Norris K., Eds., American association of Cereal Chemists, Inc., MN. 39-58 pp.

Lee, K.A. 2007. On-line analysis in food engineering. Ch. 9.2. In: Near-Infrared Spectroscopy in Food Science and Technology. Ozaki Y., McClure W.F., and Christy A.A., eds., John Wiley & Sons, Inc., NJ, 361-378 pp.

Li, S.Q, Zhang, H.Q., Jin, Z.T., and Hsieh, F-H. 2005. Textural modification of soya bean/corn extrudates as affected by moisture content, screw speed and soya bean concentration. Int. J. Food Sci. Tech. 40:731-741.

Liu, X. And Han, L. 2006. Prediction of chemical parameters in maize silage by near infrared reflectance spectroscopy. J. Near Infrared Spectrosc. 14:333-339.

Osborne, B.G. 2007. Flours and breads. Ch. 8.1. In: Near-Infrared Spectroscopy in Food Science and Technology. Ozaki Y., McClure W.F., and Christy A.A., eds., John Wiley & Sons, Inc., NJ, 281-296 pp.

Osborne, B.G. 1996. Near infrared spectroscopic studies of starch and water in some processed cereal foods. J. Near Infrared Spectrosc. 4:195-200.

Osborne, B.G., Fearn, T., and Hindle, P.H. 1993. Theory of near infrared spectrophotometry. Ch.2. In: Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, 2nd ed. Longman, Essex, UK. 14-35 pp.

Rohe, Th., Becker, W., Krey, A., Nagele, H., Kolle, S., and Eisenreich, N. 1998. In-line monitoring of polymer extrusion processes by NIR spectroscopy. J. Near Infrared

Spectrosc. 6:325-332.

Sahni, N.S., Isaksson, T., and Naes, T. 2004. In-line near infrared spectroscopy for use in product and process monitoring in food industry. J. Near Infrared Spectrosc. 12:77-83.

Scotter, C-N.G. and Millar, S. 2004. Analysis of baking products. Ch.7. In: Near-Infrared Spectroscopy in Agriculture. Roberts C.A., Workman, J. Jr., and Reeves III J.B., American Society of Agronomy, Inc., Crop Science society of America, Inc., and Soil Science Society of America, Inc., WI. 439-463 pp.

Stauffer, C.E. 1993. Extrusion basics. Bak. Snacks 18-23.

Swarbrick, B., Grout, B., and Noss, J. 2005. The rapid, at-line determination of starch content in sucrose-starch blends using near-infrared reflectance spectroscopy; a process analytical technology initiative. J. Near Infrared Spectrosc. 13:1-8.

Tumuluri, S.V., Prodduturi, S., Crowley, M.M., Stodghill, S.P., McGinity, J.W., Repka, M.A., and Avery, B.A. 2004. The use of near-infrared spectroscopy for the quantitation of a drug in hot-melt extruded films. Drug Dev. Ind. Pharm. 30:505-511.

Wehling, R.L. 1998. Infrared spectroscopy. In: Food Analysis, 2nd ed. Nielsen S.S., ed. Aspen Publishers, Inc., MA, 413-424 pp.

Williams, P.C. 2007. Grains and seeds. Ch.7.1. In: Near-Infrared Spectroscopy in Food Science and Technology. Ozaki Y., McClure W.F., and Christy A.A., eds., John Wiley & Sons, Inc., NJ, 165-218 pp.

Williams, P.C. 2001. Implementation of near-infrared technology. Ch.8 In: Near-Infrared Technology in the Agriculture and Food Industries, 2nd ed. Williams P.C. and Norris

K., eds., American association of Cereal Chemists, Inc., MN. 145-169 pp.

Whalen, P.J. 1999. Measuring process effects in ready-to-eat breakfast cereals. Cereal Foods World 44:407-412.

CONCLUSION

Chapter 1

These preliminary results show the comparison between discriminant analysis models based on principal component analysis, and Mahalanobis distances of discrete wavelengths, and found slightly greater correct classification rates when using four selected discrete wavelengths than when using PCA models. Classification results based on PCA showed that a five factor calibration model constructed by mean centering and Gap second derivative transformation with the gap size of five coupled with the use of SNV, provided the highest overall correct classification rate when partial spectra from 1100 to 1900 nm were used. This calibration model correctly classified 100% of sound, 93% of infested, 95% of sound air-dried, 91% of infested air-dried, and 92% of sound kernels from six different wheat varieties. Results of calibrations based on discrete wavelengths showed that the highest overall correct classification was achieved when using 4 wavelengths. This combination of wavelengths gave a model that correctly classified 98% of sound, 98% of infested, 100% of sound air-dried, 96% of infested air-dried, and 100% of sound kernels from six different wheat varieties. Therefore, detection of internal insect

infestation based on selected discrete wavelengths was promising, and was used to analyze data from sound and infested kernels containing 3- and 4-week old granary and maize weevil larvae.

Chapter 2

NIR spectroscopy using discriminant analysis with Mahalanobis distances based on selected discrete wavelengths can yield reliable results to classify infested kernels containing either live or dead larvae of both granary and maize weevils. The calibration developed from sound and infested kernels containing 3-week-old larvae of granary and maize weevils provided correct classification of sound, sound air-dried, and infested kernels containing 3-week-old larvae of granary and maize weevils, 4-week-old larvae of granary and maize weevils, and infested air-dried kernels containing dead larvae of granary and maize weevils of 92, 98, 77, 73, 95, 98, 96, and 94% respectively. Additionally, 99% of sound kernels from ten different wheat varieties were correctly classified into their respective classes. When applying first derivative treatment to the calibration developed from 4-week-old larvae using a segment interval= 5 and gap= 5, improved classification results were achieved for sound and infested wheat kernels containing 4-week-old granary and maize weevils to 99, 92, and 94%,

respectively. Second derivative treatment did not improve classification results for calibration developed from either 4-week-old or 3-week-old larvae.

The need for only four wavelengths allows the use of simpler and less expensive instrumentation than is required for the full spectrum methods previously used for this purpose. Additionally, automated single-kernel presentation systems need to be developed to make this technique practical for routine use of large numbers of samples. A high humidity incubator to raise insects and high resolution X-ray equipment are needed for further studies in order to achieve improved classification of smaller size larvae. Kernels infested by larvae of other internal infesters should also be included. Calibration models using fewer discrete wavelengths should also be evaluated for their effectiveness.

Chapter 3

The use of NIR spectroscopy for predicting degree of starch cooking seems to be feasible. The changes in the spectra at wavelengths associated with C-H first and second overtones, C-H stretch and deformation, O-H first overtone, etc. caused by changing of starch molecules can be measured. The data obtained from this study are the only information currently available.

Online NIR spectroscopy, which will be able to scan the extrudates automatically through fiber optics, needs to be further studied by examining the absorbance spectrum of the extruder melt as it approaches or passes through the die of the extruder. Monitoring of the melt fluid can provide information for a feedback control system. Additionally, the next phase of research can be investigated the technique with more complex formulations containing added salt, sugar, and multiple sources of starch.