

**LONG WAVELENGTH NEAR-INFRARED
HYPERSPECTRAL IMAGING FOR
CLASSIFICATION AND QUALITY
ASSESSMENT OF BULK SAMPLES OF
WHEAT FROM DIFFERENT GROWING
LOCATIONS AND CROP YEARS**

A Thesis

Submitted to the Faculty of Graduate Studies

The University of Manitoba

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

by

SIVAKUMAR MAHESH

Department of Biosystems Engineering

University of Manitoba

Winnipeg, Canada

R3T 5V6

© August 2011 by Sivakumar Mahesh

**THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES**

COPYRIGHT PERMISSION

**LONG WAVELENGTH NEAR-INFRARED HYPERSPECTRAL
IMAGING FOR CLASSIFICATION AND QUALITY ASSESSMENT
OF BULK SAMPLES OF WHEAT FROM DIFFERENT GROWING
LOCATIONS AND CROP YEARS**

By

SIVAKUMAR MAHESH

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of the University
of Manitoba in partial fulfillment of the requirement of the degree**

Of

Doctor of Philosophy

SIVAKUMAR MAHESH © 2011

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

ABSTRACT

A platform technology is identified for grain handling facilities to improve grading and determine non-destructively different quality parameters of wheat. In this study, a near-infrared (NIR) hyperspectral imaging system was used to scan four wheat classes namely, Canada Western Red Spring (CWRS), Canada Prairie Spring Red (CPSR), Canada Western Hard White Spring (CWHWS), and Canada Western Soft White Spring (CWSWS) that were collected from across various growing regions in Manitoba, Saskatchewan, and Alberta in 2007, 2008, and 2009 crop years. A database of the near-infrared (NIR) hyperspectral image cubes of bulk samples of four wheat classes at three moisture levels for each class was created. These image cubes were acquired in the wavelength region of 960-1700 nm with 10 nm intervals. Wheat classification was done using the non-parametric statistical and a four-layer back propagation neural network (BPNN) classifiers. Average classification accuracies of 93.1 and 83.9% for identifying wheat classes using the linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), respectively, were obtained for two-class identification models that included variations of moisture levels, growing locations, and crop years of samples. In the pair-wise moisture discrimination study, near-perfect classifications were achieved for wheat samples which had difference in moisture levels of about 6%. The NIR wavelengths of 1260-1380 nm had the highest factor loadings for the first principal component using the principal components analysis (PCA). A four-layer BPNN classifier was used for two-class identification of wheat classes and moisture levels. Overall average pair-wise classification accuracies of 83.7% were obtained for discriminating wheat samples based on their moisture contents. Classification accuracies of 83.2, 75.4,

73.1%, on average, were obtained for identifying wheat classes for samples with 13, 16, and 19% moisture content (m.c.), respectively. Ten-factor partial least squares regression (PLSR) and principal components regression (PCR) models were developed using a ten-fold cross validation for prediction. Prediction performances of PLSR and PCR models were assessed by calculating the estimated mean square errors of prediction (MSEP), standard error of cross-validation (SECV), and correlation coefficient (r). Overall, PLSR models demonstrated better prediction performances than the PCR models for predicting protein contents and hardness of wheat.

ACKNOWLEDGEMENTS

First and foremost, I express my sincere gratitude and owe a great many thanks to my supervisor, Dr. Digvir Singh Jayas, Vice-President (Research) and Distinguished Professor, University of Manitoba for his excellent guidance, supervision, encouragement, and support throughout the period of my doctoral study.

I convey my heartfelt thanks to my co-advisor, Dr. Jitendra Paliwal, Associate Professor, Department of Biosystems Engineering, University of Manitoba for his valuable guidance and assistance throughout my research.

I am extremely grateful to Dr. Noel D.G. White, Senior Research Scientist, Agriculture and Agri-Food Canada, for his constructive advice, providing wheat samples, and for arranging necessary analyses in his lab. I am very thankful to Mr. Colin Demianyk, Biologist, Agriculture and Agri-Food Canada for taking strenuous efforts for collecting samples from different locations and crop years.

I sincerely thank Dr. Gabriel Thomas, Associate Professor, Department of Electrical and Computer Engineering, University of Manitoba for serving on my advisory committee. I personally thank Dr. K. Alagusundaram, Director, Indian Institute of Crop Processing Technology (IICPT), Thanjavur, India, for his continuous encouragement and support for exploring opportunities in pursuing the advanced level graduate degrees in Canada.

I express my immense thanks to the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council (NSERC) of Canada, Canada

Foundation for Innovation (CFI), and the University of Manitoba Graduate Fellowship (UMGF) for providing partial financial support towards this study.

I thank administrative staff and technicians of the Department of Biosystems Engineering, University of Manitoba, for their support. I deeply acknowledge and thank my research colleagues and friends for their support. I thank Mr. Jean-Christophe Habeck for helping in data collection.

I would like to express my deepest gratitude to all my family members in India. Last but not the least; I thank my son, Sai, and my wife, Mahaa, for providing me their emotional and moral support at all times.

This thesis is dedicated to my father,

Ramarathinam Sivakumar,

who always encouraged me in all my accomplishments...

TABLE OF CONTENTS

	PAGE
ABSTRACT.....	i
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	v
LIST OF FIGURES	ix
LIST OF TABLES	xii
LIST OF ABBREVIATIONS.....	xvi
1. INTRODUCTION	1
2. BACKGROUND THEORY	7
2.1 Near-infrared hyperspectral imaging	7
2.1.1 Basic principles of hyperspectral imaging.....	7
2.1.2 Hyperspectral hardware.....	8
2.1.2.1 Hyperspectral detectors.....	9
2.1.2.2 Hyperspectral filters.....	10
2.1.2.3 Hyperspectral illumination.....	12
2.1.3 Hyperspectral imaging software.....	13
2.1.4 Hyperspectral imaging system calibration and image correction.....	14
2.1.5 Hyperspectral data analysis	16
3. REVIEW OF LITERATURE	18
3.1 Applications of near-infrared hyperspectral imaging.....	18
3.1.1 Fruits.....	19
3.1.1.1 Bruises and surface defects detection	19
3.1.1.2 Detection of bitter pits and fecal contaminations	20
3.1.1.3 Determination of quality attributes	21

3.1.2 Vegetables	24
3.1.2.1 Determination of quality attributes	24
3.1.3 Hyperspectral imaging in feeds and veterinary products	25
3.1.4 Cereals	26
3.1.4.1 Determination of quality attributes	26
3.1.5 Non agri-food applications	30
3.1.6 Limitations of hyperspectral imaging.....	30
3.2 Applications of the NIR spectroscopy	31
3.2.1 Classification of cereal grains.....	32
3.2.2 Detection of fungal, insect, and other damages in cereal grains	34
3.2.3 Determination of quality parameters in cereal grains.....	38
3.3 Summary of literature review.....	44
4. METHODS AND MATERIALS.....	52
4.1 Grain samples.....	52
4.2 Near-infrared hyperspectral imaging system	53
4.3 Spatial calibration.....	55
4.4 Illumination standardization.....	55
4.5 Image acquisition	56
4.6 Image analysis.....	56
4.6 Principal components analysis (PCA).....	57
4.6 Pattern recognizing classifiers.....	58
4.6.1 Statistical classifier.....	59
4.6.2 Artificial neural network classifier.....	59
4.7 Protein and hardness measurements.....	61
4.8 Multivariate regression algorithms.....	63

4.8.1 Partial least squares regression (PLSR)	63
4.8.2 Principal components regression (PCR)	65
5. RESULTS AND DISCUSSION	67
5.1 Segmented images and NIR reflectance spectra of wheat classes	67
5.3 Discriminant classifiers	76
5.3.1 Overall identification of wheat classes	76
5.3.2 Overall identification of moisture levels	77
5.3.3 Identification of wheat classes at 13% m.c. independent of growing locations and crop years	80
5.3.4 Identification of wheat classes at 16% m.c. independent of growing locations and crop years	82
5.3.5 Identification of wheat classes at 19% m.c. independent of growing locations and crop years	83
5.3.6 Pair-wise classification of wheat	85
5.3.6.1 Identification of wheat classes using the full data set.....	85
5.3.6.2 Identification of wheat classes at 13% m.c.....	88
5.3.6.3 Identification of wheat classes at 16% m.c.....	91
5.3.6.4 Identification of wheat classes at 19% m.c.....	93
5.3.7 Growing locations.....	95
5.3.7.1 Identification of growing locations of wheat independent of crop years...	95
5.3.8 Crop years.....	101
5.3.8.1 Identification of crop years of wheat independent of growing locations.	101
5.3.9 Identification of key wavelengths.....	104
5.4 Back propagation neural network (BPNN) classifiers for identifying moisture levels and wheat classes	108
5.4.1 Pair-wise identification of moisture contents	109

5.4.2 Classification of wheat at 13% m.c.	111
5.4.3 Classification of wheat at 16% m.c.	113
5.4.4 Classification of wheat at 19% m.c.	115
5.5 The PLSR and PCR models for protein and hardness of wheat.....	117
5.5.1 Protein content of wheat.....	118
5.5.1.1 The PLSR and PCR prediction of protein contents of wheat samples using the full data set.....	119
5.5.1.2 Class-wise PLSR and PCR prediction of protein contents of wheat samples	123
5.6.1 Hardness of wheat	126
5.6.1.1 The PLSR and PCR prediction of hardness of wheat samples using the full data set	127
5.6.1.2 Class-wise PLSR and PCR prediction of hardness of wheat samples.....	131
6. CONCLUSIONS AND RECOMMENDATIONS	135
7. REFERENCES	139
APPENDIX.....	155

LIST OF FIGURES

Number	Title	Page
Fig. 3.1.	The NIR absorption of C-H, R-OH, R-NH ₂ , ArOH, and ArCH combination bands, first, second, and third overtones (modified from Anonymous, 2011).....	32
Fig. 4.1.	The NIR hyperspectral imaging system.	54
Fig. 5.1.	Segmented NIR hyperspectral images acquired at 1330 nm for bulk samples of a) CWRS b) CPSR c) CWHWS and d) CWSWS wheat at 13% (left), 16% (center), and 19% (right) moisture content (wet basis).	68
Fig. 5.2.	Near-infrared (NIR) reflectance spectra of wheat classes that had 13% m.c. (wet basis).	69
Fig. 5.3.	Near-infrared (NIR) reflectance spectra of wheat classes that had 16% m.c. (wet basis).	69
Fig. 5.4.	Near-infrared (NIR) reflectance spectra of wheat classes that had 19% m.c. (wet basis).	70
Fig. 5.5.	The PC1 loadings plot of wheat classes that had 13% m.c. (wet basis).....	71
Fig. 5.6.	The PC1 loadings plot of wheat classes that had 16% m.c. (wet basis).....	71
Fig. 5.7.	The PC1 loadings plot of wheat classes that had 19% m.c. (wet basis).....	72
Fig. 5.8.	The PC1, PC2, and PC3 scores images of wheat samples at a moisture content of 13%.	73
Fig. 5.9.	The PC1, PC2, and PC3 scores images of wheat samples at a moisture content of 16%.	74
Fig. 5.10.	The PC1, PC2, and PC3 scores images of wheat samples at a moisture content of 19%.	75
Fig. 5.11.	Overall classification accuracies of identifying western Canadian wheat classes using the LDA and QDA classifiers.	77
Fig. 5.12.	Overall classification accuracies of identifying moisture levels of wheat using the LDA and QDA classifiers.	78
Fig. 5.13.	Pair-wise classification accuracies of identifying moisture levels of wheat using the LDA classifier.	79

Fig. 5.14. Pair-wise classification accuracies of identifying moisture levels of wheat using the QDA classifier.....	79
Fig. 5.15. Classification accuracies of identifying wheat classes that had 13% m.c. using the LDA and QDA classifiers.	82
Fig. 5.16. Classification accuracies of identifying wheat classes that had 16% m.c. using the LDA and QDA classifiers.	83
Fig. 5.17. Classification accuracies of identifying wheat classes that had 19% m.c. using the LDA and QDA classifiers.	84
Fig. 5.18. Pair-wise classification accuracies of identifying wheat classes using the LDA classifier.	86
Fig. 5.19. Pair-wise classification accuracies of identifying wheat classes using the QDA classifier.	87
Fig. 5.20. Pair-wise classification accuracies of identifying wheat classes that had 13% m.c. using the LDA classifier.	89
Fig. 5. 21. Pair-wise classification accuracies of identifying wheat classes that had 13% m.c. using the QDA classifier.	90
Fig. 5.22. Pair-wise classification accuracies of identifying wheat classes that had 16% m.c. using the LDA classifier.	91
Fig. 5.23. Pair-wise classification accuracies of identifying wheat having 16% m.c. using the QDA classifier.....	92
Fig. 5.24. Pair-wise classification accuracies of identifying wheat classes that had 19% m.c. using the LDA classifier.	94
Fig. 5.25. Pair-wise classification accuracies of identifying wheat classes that had 19% m.c. using the QDA classifier.	94
Fig. 5.26. Pair-wise classification accuracies of identifying moisture levels of wheat using a four-layer BPNN.	110
Fig. 5.27. Pair-wise classification accuracies of identifying wheat classes that had 13% m.c. using a four-layer BPNN.....	112
Fig. 5.28. Pair-wise classification accuracies of identifying wheat classes that had 16% m.c. using a four-layer BPNN.....	114

Fig. 5.29. Pair-wise classification accuracies of identifying wheat classes that had 19% m.c. using a four-layer BPNN.....	116
Fig. 5.30. Percent variance explained by the components of PLSR and PCR models for wheat.	119
Fig. 5.31. Estimated MSEPV values for the components of PLSR and PCR models for predicting protein contents of wheat.	120
Fig. 5.32. Predicted protein contents against observed protein contents of wheat using the ten-factor PLSR model (n = 7200, r = 0.68, and SECV = 1.33).....	121
Fig. 5.33. Predicted protein contents against observed protein contents of wheat using the ten-factor PCR model (n = 7200, r = 0.62, and SECV = 1.42).....	122
Fig. 5.34. Percent variance explained by the components of PLSR and PCR models for wheat.	127
Fig. 5.35. Estimated MSEPV values for the components of PLSR and PCR models for predicting hardness of wheat.	128
Fig. 5.36. Predicted hardness against observed hardness of wheat using the ten-factor PLSR model (n = 7200, r = 0.82, and SECV = 12.15).	129
Fig. 5.37. Predicted hardness against observed hardness of wheat using the ten-factor PCR model (n = 7200, r = 0.75, and SECV = 13.9).	129

LIST OF TABLES

Number	Title	Page
Table 2.1.	Detectors and their use in the hyperspectral imaging	10
Table 2.2.	Filter types, their use, and advantages of its use in hyperspectral imaging	12
Table 3.1.	Summary of hyperspectral imaging applications in fruits, vegetables, cereals, feeds, veterinary, and industrial products	47
Table 3.2.	Summary of spectroscopic applications in cereal grains	50
Table 4.1.	Growing locations (crop years) near the listed towns from where wheat samples were collected.	53
Table 4.2.	Classification of wheat based on the Perten SKCS hardness index (Source: Perten instruments, Springfield, IL)	62
Table 5.1.	The top 10 features, in descending order, based on their contribution towards classification accuracy for wheat moisture identification while using non- parametric classifiers with all features as inputs.....	80
Table 5.2.	The top 10 features, in descending order, based on their contribution towards classification accuracy for wheat class identification while using non-parametric classifiers with all features as inputs.....	85
Table 5.3.	The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for overall identification of wheat classes while using non-parametric classifiers with all features as inputs.....	88
Table 5.4.	The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for class identification for 13% m.c. wheat while using non-parametric classifiers with all features as inputs.....	90
Table 5.5.	The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for class identification for 16% m.c. wheat while using non-parametric classifiers with all features as inputs.....	93
Table 5.6.	The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for class identification for 19% m.c. wheat while using non-parametric classifiers with all features as inputs.....	95

Table 5.7. Classification accuracies of identifying growing locations near the listed towns (independent of crop years) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) for 13% moisture content wheat.	97
Table 5.8. Classification accuracies of identifying growing locations near the listed towns (independent of crop years) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) for 16% moisture content wheat.	99
Table 5.9. Classification accuracies of identifying growing locations near the listed towns (independent of crop years) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) for the 19% moisture content wheat. ...	100
Table 5.10. Classification accuracies of identifying crop years (independent of growing locations) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) models for 13% m.c. wheat.	102
Table 5.11. Classification accuracies of identifying crop years (independent of growing locations) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) models for 16% m.c. wheat.	103
Table 5.12. Classification accuracies of identifying crop years (independent of growing locations) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) models for 19% m.c. wheat.	103
Table 5.13. The top 10 features, in descending order, based on their contribution towards classification accuracy for a) identifying growing locations independent of crop years and b) identifying crop years independent of growing locations for the 13% m.c. wheat while using non-parametric classifiers with all features as inputs. ..	105
Table 5.14. The top 10 features, in descending order, based on their contribution towards classification accuracy for a) identifying growing locations independent of crop years and b) identifying crop years independent of growing locations for the 16% m.c. wheat while using non-parametric classifiers with all features as inputs. ..	106
Table 5.15. The top 10 features, in descending order, based on their contribution towards classification accuracy for a) identifying growing locations independent of crop years and b) identifying crop years independent of growing locations for the 19% m.c. wheat while using non-parametric classifiers with all features as inputs. ..	107

Table 5.16. The top 10 feature wavelengths, in descending order, based on their contribution towards classification accuracy for moisture identification while using pair-wise back propagation neural network classifiers with all features as inputs.....	111
Table 5.17. The top 10 feature wavelengths, in descending order, based on their contribution towards classification accuracy for wheat class identification at 13% m.c. while using pair-wise back propagation neural network classifiers with all features as inputs.....	113
Table 5.18. The top 10 feature wavelengths, in descending order, based on their respective contribution towards classification accuracy for wheat class identification at 16% m.c. while using pair-wise back propagation neural network classifiers with all features as inputs.....	115
Table 5.19. The top 10 feature wavelengths, in descending order, based on their contribution towards classification accuracy for wheat class identification at 19% m.c. while using pair-wise back propagation neural network classifiers with all features as inputs.....	117
Table 5.20. Summary of protein dataset statistics for wheat (n = 30 per wheat class)...	118
Table 5.21. The top 10 wavelengths, in descending order, identified for the ten-factor PLSR and PCR models for predicting protein contents of wheat samples.....	123
Table 5.22. Performance of ten-factor PLSR and PCR models for predicting protein contents using the class-specific wheat samples.....	125
Table 5.23. The top 10 wavelengths, in descending order, identified for the ten-factor PLSR and PCR models for predicting protein contents of class-specific wheat samples.....	125
Table 5.24. Summary of hardness dataset statistics for wheat (n = 300 kernels per wheat class per moisture).	126
Table 5.25. The top 10 wavelengths, in descending order, identified by the ten-factor PLSR and PCR models for predicting hardness of wheat samples.	130
Table 5.26. Performance of the ten-factor PLSR and PCR models for predicting hardness using the class-specific wheat samples	132

Table 5.27. The top 10 wavelengths, in descending order, identified for the ten-factor
PLSR and PCR models for predicting hardness of class-specific wheat samples
..... 133

LIST OF ABBREVIATIONS

A/D	Analog to digital
AACC	American Association of Cereal Chemists
AAFC	Agriculture and Agri-Food Canada
ANN	Artificial Neural Network
AOAC	American Organization of Analytical Chemists
AOTF	Acousto-Optical Tunable Filter
ASAE	American Society of Agricultural Engineers
ASCC	Average Squared Canonical Correlation
BAR	Bureau of Appeals and Review
BGYF	Bright Greenish Yellow Fluorescence
BPNN	Back Propagation Neural Network
CCD	Charge Coupled Device
CPSR	Canada Prairie Spring Red
CPSW	Canada Prairie Spring White
CWAD	Canada Western Amber Durum
CWES	Canada Western Extra Strong
CWHWS	Canada Western Hard White Spring
CWRS	Canada Western Red Spring
CWRW	Canada Western Red Winter
CWSWS	Canada Western Soft White Spring
DET	Detrending
DHV	Dark Hard Vitreous
DON	Deoxynivalenol

ETF	Electronically Tunable Filter
FLD	Fisher's Linear Discriminant
FPA	Focal Plane Array
FT	Fourier Transform
GLCM	Grey Level Co-occurrence Matrix
HRS	Hard Red Spring
HRW	Hard Red Winter
ICA	Independent Component Analysis
InGaAs	Indium-Gallium-Arsenide
LCTF	Liquid Crystal Tunable Filter
LDA	Linear Discriminant Analysis
LED	Light Emitting Diodes
LS-SVM	Least Squares Support Vector Machines
LWNIR	Long Wavelength Near-Infrared
MIA	Multivariate Image Analysis
MLR	Multiple Linear Regression
MNF	Minimum Noise Fraction
MR-MIA	Multiresolutional – Multivariate Image Analysis
MSC	Multiplicative Scatter Correction
MSEP	Mean Square Error of Prediction
MVI	Multivariate Image
NDHV	Non-Dark Hard Vitreous
NIR	Near-Infrared
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Principal Components Regression

PLS-DA	Partial Least Squares-Discriminant Analysis
PLSR	Partial Least Squares Regression
QDA	Quadratic Discriminant Analysis
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross Validation
RMSEP	Root Mean Square Error of Prediction
ROI	Region of Interest
RPD	Relative Performance Determinant
RP-HPLC	Reverse Phase – High Performance Liquid Chromatography
SE-HPLC	Size Exclusion High Performance Liquid Chromatography
SIMCA	Soft Independent Modeling of Class Analogy
SKCS	Single Kernel Characterization System
SME	Specific Mechanical Energy
SMV	Soybean Mosaic Virus
SNV	Standard Normal Variate
SRW	Soft Red Winter
SVM	Support Vector Machine
SWIR	Short Wavelength Infrared
SWNIR	Short Wavelength Near-Infrared
TSS	Total Soluble Solids
USDA-FGIS	United States Department of Agriculture-Federal Grain Inspection Service
US-FDA	United States-Food and Drug Administration
WILMA	Wavelet Interface to Linear Modeling Analysis
WT	Wavelet Transform

1. INTRODUCTION

Agricultural materials are heterogeneous substances, have inherent quality variations, and contain complex components such as proteins, carbohydrates, and fat. Wheat is the third most important cereal crop in the world after maize and rice. In Canada, wheat, which is ranked at the top in the list of agricultural crops followed by canola and maize, was produced and exported in quantities of 28.6 and 15.7 MT, respectively, in 2008 (FAOSTAT, 2011). Wheat grades of the Canadian Grain Commission (CGC) ensure a satisfactory performance in wheat quality and milling potential; and prove that grains are contaminant free (CWB, 2011). Classification of wheat is mainly based on its colour (red vs. white), hardness (soft vs. hard), and growing season (winter vs. spring). Eight major western Canadian wheat classes, which are produced throughout the prairie provinces (Manitoba, Saskatchewan, and Alberta) and exported to various countries around the world, are: Canada Western Red Spring (CWRS), Canada Prairie Spring Red (CPSR), Canada Western Extra Strong (CWES), Canada Western Red Winter (CWRW), Canada Prairie Spring White (CPSW), Canada Western Amber Durum (CWAD), Canada Western Soft White Spring (CWSWS), and Canada Western Hard White Spring (CWHWS). Production of class-based wheat, while maintaining a zero tolerance policy on insect levels during export, normally satisfies a wide range of domestic and international consumers. The price of wheat that is traded in Canada is normally fixed based on the quality and the market demand of specific wheat classes. A specific wheat class is used as a primary raw material for products such as bread, pasta, noodles, and flat bread. Traditionally, wheat classification has been carried out based on kernel morphological features such as size, shape, colour, and appearance.

A CGC report indicated that protein content of CWRS in 2009 was 0.2 and 0.9% lower than 2008 and 2007 CWRS wheat, respectively (CGC, 2011). The report also specified that the protein content of CWHWS wheat in 2009 (13.2%) was 0.2% lower than that of CWHWS wheat in 2008. Chemical compositions may vary in wheat classes because of changes in soil and climatic conditions of growing regions through crop years. Also, variations in moisture levels of procured wheat turn out to be a serious concern for grain handlers. McNeill et al. (2011) reported that wheat harvesting could be performed at higher moisture levels (above 15%) if followed by an effective drying. High moisture wheat should be dried to an optimal moisture level (12-13%) to store safely and prevent spoilage and/or sprouting prior to processing. Also, moisture levels of wheat lots are not uniform at the time they reach primary or terminal elevators or other processing facilities. Accidental mixing of wheat classes during transportation or handling reduces the value of the wheat lot considerably. Hence, inclusion of location-specific, crop year-specific, and moisture-specific wheat classes become a part of this study. Also, it helps to develop robust classification and prediction models when sample variations due to growing locations, crop years, and moisture are taken into consideration.

It is very important to know information on a wheat class and its quality in every stage of the grain handling process. Presently, visual inspection is the primary method for identifying classes and associating them with quality in grain handling facilities even though Canada has an established grain grading system for wheat. This method often needs experienced and specific skill-set personnel and is highly subjective in nature. Factors, such as tiredness, impaired vision, work pressure, poor lighting, and sometimes weather, also influence grain inspectors during wheat grading and quality assessment.

Existing laboratory-based methods include polyacrylamide gel electrophoresis (PAGE) and reversed-phase high performance liquid chromatography (RP-HPLC) for objective classification of wheat. Some of the drawbacks of using these methods are length of time per analysis, requirement of skilled operators, and higher equipment cost. To overcome the drawbacks, machine vision and spectroscopy have been researched for classifying agricultural products and evaluating their quality parameters. Paliwal et al. (1999) used a machine vision approach to classify barley, oats, rye, and two wheat classes (CWRS and CWAD). Neuman et al. (1987) used digital imaging techniques to classify wheat cultivars based on kernel type and identity. The NIR spectroscopy has been used extensively for measuring the concentration of various constituents in agricultural products. However, this technology was not helpful in resolving the spatial distribution of constituents within the sample. Imaging techniques are therefore used predominantly for evaluating the quality parameters of agricultural crops. Image analyses with hardness measurements were used to identify wheat classes and their cultivars (Zayas et al., 1996). Monochrome images of wheat bulk samples were used to classify western Canadian wheat at different moisture levels (Manickavasagan et al., 2008). Thermal imaging that generates visible images of samples from invisible radiation patterns was also found useful in classifying wheat (Manickavasagan et al., 2010).

Wheat has higher protein content; which varies between 10 and 18% of the total dry matter; than maize, paddy, or other cereal crops. Protein content of wheat is considered as one of the basic intrinsic properties and it always has an effect on functional properties of processed products. Also, the presence of specific type of protein in wheat affects the baking quality. MacRitchie (1987) described from the study to

evaluate protein fractions of wheat for bread making and dough mixing that the relative amount of globulin-type to the glutenin-type proteins seemed to be influential in baking quality. Protein content of single kernels and bulk samples of wheat were calculated from near infrared reflectance spectroscopic values of individual kernels (Delwiche, 1998; Delwiche, 2000). A near-infrared transmittance spectrophotometer was used to find out the practicability of measuring protein contents of intact kernels of wheat (Delwiche, 1995). Watson et al. (1977) developed regression models for protein content of wheat using near-infrared reflectance spectroscopy. They further reported that the near-infrared reflectance values were affected by and were based on the hardness of wheat, the key determinant of end usage. Williams (1979) utilized near infrared reflectance spectroscopy to inspect wheat for protein and hardness. Slaughter et al. (1992) specified discriminating possibilities between hard red spring and hard red winter wheat classes for protein and hardness attributes.

Spectral imaging is a modern technology that combines spectroscopy with image processing. Hyperspectral imaging, an extension of multispectral imaging, is becoming a popular research tool from which both spectral and spatial information of samples can be acquired simultaneously. Hyperspectral imaging provides a large data set, otherwise called a data cube, which facilitates a complete and reliable analysis of intrinsic properties and external characteristics of samples. It is a recognized tool that permits spectroscopic image analysis of a sample or a point within the region of interest using image processing techniques and chemical sensing methods (Headwall, 2011). Two to ten spectral images can be collected in the multispectral imaging. But, more than ten to several hundred images can be acquired by hyperspectral imaging systems (Lawrence et

al., 2003). In the laboratory, the Raman, infrared, and near infrared wavelengths were used in the hyperspectral imaging instruments. Since the late 1990s, hyperspectral imaging has been used in studies in a variety of different fields such as regional mapping of planetary surfaces (Bellucci and Formisano, 1997), environmental mapping (Clark et al., 1997), mineral mapping (Resmini et al., 1997), and precision farming (Moran et al., 1997). Hyperspectral imaging in the NIR wavelength range becomes increasingly interesting in industrial applications such as material classification (Tatzer et al., 2005). Considerable research has been reported in the last 10 years on using hyperspectral imaging for quality and safety inspection of food and agricultural products. It included detection of bruises on pickling cucumbers (Ariana et al., 2006), apples (Xing et al., 2005), starch index determination of apple (Peirs et al., 2003), detection of fecal and ingesta contaminants in poultry carcasses (Park et al., 2002), and measurement of ripeness of tomatoes (Polder et al., 2002). Image, spectra (reflectance and absorbance), and feature (distinct points in n-dimensional vector) forms are the three common methods used to quantitatively represent the hyperspectral data. Supervised classification models are developed using the feature form of a data cube. During the hyperspectral image acquisition process, high quality images with high signal-to-noise ratio have been acquired using the long wavelength near-infrared (LWNIR) region hyperspectral sensors. Singh et al. (2007) detected different fungal species in wheat using the LWNIR hyperspectral imaging in the wavelength region of 1000-1600 nm. Xing et al. (2009) evaluated the α -amylase activity in single kernels of wheat using a short wavelength infrared (SWIR) hyperspectral imaging system in the wavelength region of 1000-2500 nm. Choudhary et al. (2009) used wavelet features extracted from NIR hyperspectral

image cubes of bulk samples of wheat in the wavelength region of 960-1700 nm for identifying classes. Considering the advantages of using the LWNIR hyperspectral imaging, the overall objectives of this research were to evaluate the potential of using the LWNIR hyperspectral imaging for classifying wheat using statistical and neural network classifiers, and to develop partial least squares and principal components regression models for the prediction of protein and hardness.

The main objectives of this thesis were to:

1. examine the performance of statistical classifiers for real-time identification of location-specific and crop year-specific wheat classes and the spectral region and/or wavelengths that were most suitable for classification;
2. develop statistical classifiers to identify specified moisture levels of wheat classes and to find out the spectral region and/or wavelengths that were most suitable for the identification;
3. study the performance of neural network classifiers to identify bulk samples of wheat; and
4. investigate hyperspectral image cubes of wheat using partial least squares regression (PLSR) and principal components regression (PCR) methods for assessing quality parameters such as protein and hardness.

2. BACKGROUND THEORY

This section discusses details of the principles and components involved in the system and various analysis methods used for handling hyperspectral image data.

2.1 Near-infrared hyperspectral imaging

Electromagnetic spectrum, basics of mechanical models (harmonic and anharmonic), and different forms of energy were briefly discussed by Singh (2009).

2.1.1 Basic principles of hyperspectral imaging

Hyperspectral imaging applies conventional imaging, radiometry, and spectroscopic principles for acquiring images. Each pixel, which is otherwise a spatial resolution element, of a hyperspectral image cube has a spectrum. Hyperspectral data are generally arranged as a three dimensional (3-D) cube, otherwise called a hypercube, with two spatial dimensions and one spectral dimension. Hypercube corresponds to a stack of images of samples where images are acquired at a wavelength region of 960-1700 nm (Gat, 2000). Hyperspectral data can be transformed into radiometric quantities such as reflectance, absorbance, and transmittance. These quantities can be further related to the physical characteristics or chemical composition of samples. Extrinsic characteristics, such as size, geometry, appearance, and colour of samples, can be obtained through hyperspectral image feature extraction procedures. Chemical constituents such as water, fat, protein and other carbon-hydrogen-oxygen bonded constituents of samples can be identified using hyperspectral image analyses (Lu and Chen, 1999). Each hyperspectral image cube consists of 50 to 300 images, which are acquired at different wavelengths, with a spectral resolution of 1 to 10 nm from a specific wavelength region. Basic approaches in the spectral imaging are:

- Sequential acquisition of two dimensional (2-D) images at different wavelengths in a specific wavelength region and
- Obtaining a full spectrum for each pixel on the line by acquiring line images sequentially in a wavelength region

Use of the NIR region in hyperspectral imaging instruments gives an ideal situation for studying diverse biological materials. Reflectance and/or absorbance spectra can be derived from hyperspectral images of samples and further used for chemometric analyses. Diffuse reflectance spectra can be generated when the NIR radiation penetrates well into the samples. Production of absorbance spectra can be possible while there is no reflection in the absorbed radiation of samples. Absorbance spectra can be directly related to sample characterization and concentration determination.

2.1.2 Hyperspectral hardware

The actual configuration depends on the type of approach used for the hyperspectral imaging system development. Common parts in all types of hyperspectral imaging systems include a charge coupled device (CCD) camera, a frame grabber, a detector, a filter, a computer, and an illumination system. In general, a highly sensitive CCD camera with a high signal-to-noise ratio is required to detect a limited amount of photons entering the detector. A higher power illumination system is often required to acquire hyperspectral images effectively. The size and space occupied by hypercubes are crucial and considered important in the hyperspectral imaging. As the hyperspectral data are huge in size, they are very difficult to view, manage, and interpret. High speed computer with a massive capacity hard disk is normally required for processing hyperspectral information. Shading correction is used when illumination is not uniform

when looking at bigger areas. Use of reference standards offers effective shading corrections and true reflectance estimates.

2.1.2.1 Hyperspectral detectors

Reflectance and transmittance spectra of samples are recorded using hyperspectral detectors. Table 2.1 shows the summary of different detectors and their use in the hyperspectral imaging. Lead sulphide (PbS), silicon, and Indium-Gallium-Arsenide (InGaAs) detectors are used for the wavelength regions of 1100-2500, 360-1050, and 900-1700 nm, respectively, in the single channel detection systems. Focal plane array (FPA) detectors are used mostly in hyperspectral imaging than point detectors. The advantages of using the FPA detectors are: less scanning time, high signal-to-noise ratio, uniform background, and no image distortion. A linear array and 2-D array of detectors are used in line scan and area scan imaging systems, respectively. In the line scanning method, spatial dimension records pixels in the lines whereas spectral dimension documents the spectral information of the corresponding pixel. This approach is perfect for a conveyor belt system where the line scan hyperspectral camera is used (Polder et al., 2002). In the short wavelength near-infrared (SWNIR) hyperspectral imaging in the wavelength region of 700-1000 nm, silicon detectors are used. They are cheap but cannot be used for long wave NIR applications. Silicon detectors need coating which will reduce the quantum efficiency in long wave applications. Indium Antimonide (InSb), Platinum Silicide (PtSi), Indium Gallium Arsenide (InGaAs), Germanium (Ge), Mercury Cadmium Telluride (HgCdTe), and quantum well infrared photodetectors (QWIPs) are common commercial FPA detectors. In multispectral and/or hyperspectral imaging systems, the InSb, InGaAs, HgCdTe, and QWIP detectors are mostly used. Indium antimonide (InSb),

HgCdTe, QWIPs are also used for long wavelength IR imaging systems. Specific advantages and drawbacks of using these detectors for hyperspectral imaging are briefly discussed by Singh (2009). In hyperspectral imaging, the InGaAs detectors are normally used in wavelength region of 900-1700 nm. They are highly sensitive and used for wider wavelength regions. They produce a small amount of noise and take less response time in the NIR region but they are not highly sensitive beyond 1700 nm.

Table 2.1. Detectors and their use in the hyperspectral imaging

Detector type	Use
Lead Sulphide (PbS)	1100-2500 nm
Silicon	360-1050 nm
Indium-Gallium-Arsenide (InGaAs)	900-1700 nm, FPA detection
Indium Antimonide (InSb)	FPA detection, LWIR systems
Platinum Silicide (PtSi)	FPA detection
Germanium (Ge)	FPA detection
Mercury Cadmium Telluride (HgCdTe)	FPA detection, LWIR systems
Quantum well infrared photodetectors (QWIP)	FPA detection, LWIR systems

2.1.2.2 Hyperspectral filters

A number of optical monochromatic principles, such as prism-grating prism, tunable filters, and interferometers with sample scanning, are available for producing hyperspectral images by obtaining desired wavelength light and removing all other unnecessary wavelength radiations (Geladi et al., 2004). Hyperspectral imaging system was developed by attaching either a tunable filter or a filter wheel with a monochrome camera (Polder et al., 2003). Major limitation of using a filter wheel in the system is the limited availability of the number of bands. In some types, an imaging spectroscopy and a monochrome camera are joined together in developing a hyperspectral imaging system. Both spectral and spatial information of samples is collected by capturing hyperspectral

images. Filters are selected based on the type of hyperspectral imaging system used, i.e., grating devices are used for pushbroom hyperspectral imaging and electronically tunable filters (ETF) for area scan hyperspectral imaging. Liquid crystal tunable filters (LCTFs), one of the most common filters, is currently being used to acquire hyperspectral images (Gat, 2000; Lu and Chen, 1999). The LCTF can be used in the sequential recording of full spatial images (FSI) at each wavelength and tuned to any desired wavelength by computer. The exposure time at each wavelength should be adjusted properly in the LCTF as its transmission is mainly wavelength-dependant. Acousto-optical tunable filters (AOTF) and interferometers are other common electronically tunable filters used in spectral imaging. The AOTF and LCTF have the following advantages: large aperture size, high spectral resolution, wide spectral range, easy tuning of wavelengths, and acquisition of no-distortion images. Measurements of distinctive greyscale images with high spectral resolution can be possible using more advanced hyperspectral hardware, software, suitable wavelength filters along with broad spectral band imaging detectors. Voltage and acoustic signals are used to control spectral transmissions electronically in tunable filters, one of the core parts, in hyperspectral imaging systems. Table 2.2 shows filter types, their use, and advantages of using in hyperspectral imaging. Gat (2000) describes perfect tunable filters should have the important attributes such as minimal tunability time, minimal out-of-band transmission, minimal physical thickness, low power consumption, polarization insensitivity, selectable bandpass, insensitivity to ambient temperature and humidity, insensitivity to angle of incidence of the incoming light, infinite spectral range, top hat bandpass curve, large aperture, constant bandpass, and random access to wavelengths.

Table 2.2. Filter types, their use, and advantages of its use in hyperspectral imaging

Filter type	Use	Advantages
Liquid crystal tunable filter (LCTF)	<ul style="list-style-type: none"> • electronically tunable filters • the sequential recording of full spatial images (FSI) at each wavelength is possible • easily tunable to any desired wavelength by computer 	<ul style="list-style-type: none"> • large aperture size • high spectral resolution • wide spectral range • easy tuning of wavelengths • acquisition of no-distortion images
Acousto-optical tunable filter	<ul style="list-style-type: none"> • electronically tunable filters 	<ul style="list-style-type: none"> • large aperture size • high spectral resolution • wide spectral range • easy tuning of wavelengths • acquisition of no-distortion images
Interferometer	<ul style="list-style-type: none"> • electronically tunable filters 	<ul style="list-style-type: none"> • Spectral imaging is possible with very fine spatial resolution

2.1.2.3 Hyperspectral illumination

An ideal illumination of hyperspectral imaging system should have the following key characteristics: homogeneous illumination over a large area, perfect fiber transmission in long fibers, short pulses < 10 fs, intense polychromatic light, polarized light with known Stokes parameters, deep transmission through samples, controlled reflection from deep into the sample, and no radiation damage to the samples. In the real world, no illumination system fulfils all the properties and always some properties have to be compromised. High quality hyperspectral images without any significant noise can be acquired using proper illumination sources. In NIR instruments, tungsten halogen lamps, quartz halogen lamps, light emitting diodes (LED), tunable lasers, and heated xenon lamps are used as illumination sources. As tungsten halogen lamps are durable,

stable, and capable of emitting 400-2500 nm light, they are used as the NIR hyperspectral illumination sources.

2.1.3 Hyperspectral imaging software

Once hyperspectral images are captured, the data are further transferred to a high speed computer for storage and further analyses. Effective algorithms are developed to efficiently organize the voluminous hyperspectral data and extract important spatial and spectral features. The following steps are needed to analyze hyperspectral images: data pre-processing, enhancement of spectral images, dimension reduction of data, and material/chemical component identification or classification. Standard normal variate (SNV) and multiplicative scatter correction (MSC) are crucial pre-processing methods used in the removal of some of the large amount of variability that are generated from scattering effects of reflectance spectra. In data pre-processing, original image files are converted into 3-D hypercubes with radiometric quantities. Good quality filters and proper transformation methods are used to improve the quality of spectral images to extract the most important features effectively. The SNV and MSC often produce similar results and are commonly considered as transferable methods (Fearn et al., 2009). The SNV gave curved structures in score plots, which was generated from treated spectra, whereas the MSC produced outliers.

Principal components analysis (PCA), the minimum noise fraction (MNF) transform, and the major dimension reduction methods are involved in hyperspectral data analyses with the goal of reducing dimensions or shrinking the volume of spectral and spatial data without loss of critical information of samples. Also, spatial dimensions can be reduced using the pixel purity index (PPI) method. Vogt et al. (2005) used 3-D

wavelet transformation methods for reducing the dimensions of hyperspectral data before storing and/or performing any chemometric analyses. Five Daubechies family wavelet types were compared for their data preserving abilities. As a result, the computation time for the PCA for wavelet transformed hyperspectral data was decreased and the storage space was reduced for the transformed data. A multiresolutional multivariate image analysis (MR-MIA) performed better than the wavelet textural analysis. The MR-MIA was helpful in the data decomposition of hypercubes that were acquired from using a large number of wavelength bands in hyperspectral imaging (Liu and MacGregor, 2007).

2.1.4 Hyperspectral imaging system calibration and image correction

Image corrections in the hyperspectral imaging include corrections for gain, dark current offset, and the variable integration time during image processing. Use of bad sensors, sensor nonlinearity, sensor differences, heterogeneous illumination, and inadequate resolution of analog to digital (A/D) convertor influence a great deal the results of hyperspectral imaging. A number of calibrations and corrections are required to acquire good quality hyperspectral images. The raw uncorrected data can be inspected only as images as the spectral data collected from a CCD camera represent detector signal intensity counts but not actual reflectance values. The spectra become useful once the reflectance and absorbance intensities are determined. It becomes essential to calibrate and correct hyperspectral imaging instruments. Changes in the intensity of lamps can be compensated for by adjusting the integration time of the InGaAs array of the camera (Geladi et al., 2004). Many biological materials get damaged or catch fire when exposed to high intensity lamps. Saturation of the analog to digital (A/D) converter can be avoided by combining the lamp intensity with the array integration time. For the most sensitive

wavelengths, the A/D converter saturation should be set at the intensity values of 75 to 90%. Noises, such as photon, thermal, readout, and quantization, from a variety of sources can be introduced during the image acquisition process using a hyperspectral camera. Chemometric tools with image processing methods are used in hyperspectral imaging for the following reasons: to reduce the hyperspectral data dimensions, select the most significant wavelengths, extract key features, and develop classification and prediction models. In the hyperspectral imaging, a dark current (background) image, D , has to be recorded as the InGaAs array has a wavelength-dependant dark current. A reference image, W , is acquired using a reference standard of nearly 100% reflectance. Generally, the 99% reflectance standard is used and percent reflectance for samples, S , can be calculated using the following formula:

$$R = ([S - D]/[W - D]) \dots \dots \dots (2.1)$$

where R = percent reflectance of samples; S = the sample image; D = the dark current (background) image; W = the white reference image of the 99% reflectance standard.

Absorbance values of samples can also be computed as follows:

$$A = \log_{10}(1/R) = -\log_{10} \left(\frac{(S - D)}{(W - D)} \right) \dots \dots \dots (2.2)$$

Equation 2.2 is called a one-point calibration as it follows a linear trend. It is very hard to detect the nonlinear behaviour of detectors. Since the reflectance values are calculated from one single reference standard values, calibration is linear. As it is difficult to find a 100% reflectance standard, the use of nearly 100% standards may produce minor errors

in reflectance (R) and absorbance (A) values. Use of 2, 50, 75, and 99% reflectance standards produce more reference images that help detecting nonlinearities and producing superior values for reflectance intensities by averaging. Major concerns of hyperspectral imaging include the size of hypercube (usually very large) and information overlap in neighbouring image slices (Singh, 2009).

2.1.5 Hyperspectral data analysis

Once the hyperspectral data are reduced dimensionally, sample identification can then be performed using either supervised and/or unsupervised classification techniques. Common classification techniques are: minimum distance method, maximum likelihood method, and Mahalanobis distance method, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and/or the artificial neural networks (ANNs). The relevant features from hyperspectral images are acquired using mathematical and/or statistical techniques. Suitable data reduction techniques such as principal components analysis (PCA) and partial least squares (PLS) techniques are selected to reduce the dimensionality and the volume of spectral information.

In some studies, the standard error of calibration (SEC) and standard error of prediction (SEP) values, which are determined using the following formulae (ElMasry et al., (2007)), are used for evaluating PLSR and PCR models.

$$SEC = \sqrt{\frac{1}{(I_c-1)} \sum_{i=1}^{I_c} (\hat{y}_i - y_i)^2} \dots \dots \dots (4.5)$$

$$SEP = \sqrt{\frac{1}{(I_p-1)} \sum_{i=1}^{I_p} (\hat{y}_i - y_i - \text{bias})^2} \dots \dots \dots (4.6)$$

$$\text{bias} = \frac{1}{(I_p)} \sum_{i=1}^{I_p} (\hat{y}_i - y_i) \dots \dots \dots (4.7)$$

where, \hat{y}_i = the predicted value of the quality parameter in sample number i, y_i = the measured value of the quality parameter in i^{th} sample, I_c = the number of samples in the calibration set of the model, I_p = the number of samples in the validation set of the model. Attaining low SEC and SEP values and a high correlation coefficient (r) are the characteristics of a good PLSR model.

The probabilistic neural network (PNN), LDA, QDA, k-nearest neighbour (KNN), and least squares support vector machines (LS-SVM) are used for classification purposes. The NIR spectral information is reduced in dimensions using methods such as the PCA, wavelet transform (WT), and Fourier transform (FT). The PCA based multivariate image analysis (MVI) techniques are used for dimensionality reduction and necessary features extraction from hyperspectral image files. In the PCA, score values, factor loadings, and variance are used to determine the number of significant components required for explaining most of the sample variations. Some of the chemometric tools include wavelet transform and independent component analysis (ICA). In the ICA, the number of independent components has to be specified before the analysis.

3. REVIEW OF LITERATURE

In this section, major applications and limitations of using hyperspectral imaging are discussed in detail.

3.1 Applications of near-infrared hyperspectral imaging

High-resolution hyperspectral imaging provides an abundance of spectral information from which essential features are extracted for further analysis using image processing procedures. Indispensable wavelengths are selected for multispectral imaging purposes as hyperspectral images of full wavelength regions that have excessive and redundant information. Combination of spectral and data processing methods are required to select a few optimal wavelengths for detecting defects in fruits and vegetables. Consideration should be given to avoid loss of crucial information in the original hyperspectral data. In recent years, the quality and safety inspection studies of food, agricultural, and poultry products were conducted using the visible and NIR hyperspectral imaging. Hyperspectral imaging was used mainly for: bruise detection in fruits and vegetables (Ariana et al., 2006; Xing et al., 2005; Lu, 2003), detection of fecal contamination (Kim et al., 2002; Lefcourt et al., 2006) and surface defects (Mehl et al., 2004) in fruits, and measurement of bitter pit in apples (Nicolai et al., 2006). Also, it was used for inspecting cucumber chilling damage (Cheng et al., 2004) and determining moisture content, total soluble solids content, and acidity in strawberries (ElMasry et al., 2007). Gowen et al. (2007) reported that hyperspectral imaging in the visible and NIR wavelength regions (350-1700 nm) was used for assessing quality parameters of agricultural (corn, potato, and cucumber), horticultural (apple, citrus, peach, strawberry, and cherry), and meat products (poultry, pork, and codfish).

3.1.1 Fruits

3.1.1.1 Bruises and surface defects detection

If an automatic system is developed for bruise detection in fruits and vegetables, it will help provide better products for consumers, reduce potential economic losses, and increase the net profit. Lu (2003) detected both old and new bruises in Red Delicious and Golden Delicious apples using NIR hyperspectral imaging. Results showed that spectral region of 1000-1340 nm was more appropriate region in detecting bruises. Detection accuracies of 62-88% for Red Delicious and of 59-94% for Golden Delicious apples were reported using both PC and minimum noise fraction (MNF) transforms. The optimal spectral resolution for bruise detection was between 8.6 and 17.3 nm with 20-40 spectral bands.

Mehl et al. (2004) studied detection of surface defects and contaminations, such as side rots, bruises, flyspeck, scabs and moulds, fungal diseases, and soil contaminations, on the surfaces of Red Delicious, Golden Delicious, Gala, and Fuji apples using visible-NIR hyperspectral imaging. Contaminated portions of apples, independent of the apple colour and cultivar, were detected by comparing asymmetric second difference images at the wavelengths of 685, 722, and 869 nm.

The potential of using a multispectral imaging system was investigated for detecting bruises on Golden Delicious apples (Xing et al., 2005). Spectral region of 400-1000 nm was used for acquiring hyperspectral images. Four wavelengths (558, 678, 728, and 892 nm) were selected based on the PCA for the multispectral imaging. Classification accuracies were 93 and 86% for detecting healthy and bruised apples, respectively, using image processing and classification algorithms based on moments

thresholding. Lee et al. (2005) reported correlation coefficients of 0.91 and 0.79 using band ratio and band difference methods, respectively, for detecting defects on apples. Also, the wavelength region of 418-918 nm was appropriate for detection using the hyperspectral imaging.

3.1.1.2 Detection of bitter pits and fecal contaminations

Design of detection or classification system using hyperspectral imaging becomes significantly difficult than ordinary machine vision systems as the former produces large size data sets. Nicolai et al. (2006) identified bitter pit lesions on apples using the NIR hyperspectral images acquired from the wavelength region of 1000-1600 nm. The system could identify bitter pit lesions that were unseen visually and developed just after harvest. But, corky tissues could not be differentiated.

Kim et al. (2002) detected fecal contaminations on apples using the hyperspectral imaging in the wavelength region of 450-851 nm. The PCA was used on hyperspectral images of apples to identify 2-4 potential wavelengths that could be used for developing an on-line multispectral imaging system. Fecal contaminations were identified successfully using wavelengths from green and red bands in the visible region and the NIR region or two wavelengths from the ends of the NIR region. Sorting of apples based on colour was effective using three wavelengths from the visible-NIR region. Analyses with threshold detection and morphological filtering were needed for detecting diluted fecal contamination spots on apples.

Lefcourt et al. (2006) studied the multispectral detection of fecal contaminations on apples. Detection accuracies were 100 and 62.5% for Golden Delicious and Red

Delicious apples, respectively, for 1:20 dilution fecal contamination spots using optimal wavelengths identified from the NIR hyperspectral reflectance images. Concentrated fecal contamination spots were detected at a detection rate of 100%. Liu et al. (2007) explored the potential of hyperspectral images for detecting fecal contaminations on apple surfaces. Reflectance intensities were low for fecal contaminated areas on apples. In the visible-NIR region (675-950 nm), significant spectral differences were observed between uncontaminated and fecal contaminated skins. Fecal contaminated skins were identified using a dual-band ratio algorithm at 725 and 811 nm. Use of spectral features at these wavelengths reduced colour variation effects of apple cultivars.

3.1.1.3 Determination of quality attributes

Harvest time, post harvest quality, and end use are determined by the maturity of apples. Skin and flesh colour, firmness, sugar, starch pattern index, soluble solids content, and titratable acid are important maturity parameters in apples. Peirs et al. (2003) determined starch index of apples using the NIR hyperspectral imaging. The use of iodine solution, which is generally toxic in nature, is avoided in this method for the starch index determination. The PCA score images were used to classify starch and non-starch areas of apples. They suggested that the use of bandpass filters could simplify this technique and speed up the application in the future.

Noh and Lu (2007) assessed quality parameters such as skin and flesh colour, firmness, soluble solids content, and titratable acid on Golden Delicious apples using the hyperspectral laser-induced fluorescence imaging. Mean, maximum, and standard deviation spectra were found. A hybrid method, which combined the PCA with ANN, was used to predict the quality parameters at 0, 1, 2, 3, 4, and 5 min after illumination.

The apple skin hue was predicted with a correlation coefficient of 0.94. Predictions of fruit firmness, skin chroma, and flesh hue were better than the predictions of soluble solids content, titratable acid, and flesh chroma.

Non destructive determination of moisture content, total soluble solids (TSS), and acidity (pH) were done using the vis-NIR hyperspectral imaging (400-1000 nm) in strawberries (ElMasry et al., 2007). The correlation coefficients for predicting moisture, TSS, and pH were 0.90, 0.80, and 0.87, respectively, using partial least squares (PLS) analysis with the full spectral data. The β -coefficients of the PLS models were used for selecting optimal wavelengths to develop multiple linear regression (MLR) models. Correlation coefficients for predicting moisture, TSS, and pH using the MLR models were 0.87, 0.80, and 0.92, respectively. Grey-level co-occurrence matrix (GLCM) was used to perform a texture analysis on hyperspectral images for classifying strawberries based on ripeness stages. The highest classification accuracy of 89.61% was reported using the GLCM parameters at the horizontal direction ($\theta = 0^\circ$).

Nagata et al. (2004) evaluated maturity parameters of strawberries using a hyperspectral imaging system with a LCTF in the visible wavelength region of 450-650 nm. Hyperspectral images were acquired at every 2 nm interval and data were extracted to develop firmness and soluble solids content calibration models. A correlation coefficient of 0.784 with a standard error of prediction of 0.364 was observed for the firmness model with five predictors (510, 650, 644, 628, and 598 nm). Consistent predictions of soluble solids content could be obtained using maturity level analyses of individual strawberries.

Nagata et al. (2005) further evaluated internal qualities such as firmness and soluble solids content of strawberries using the vis-NIR hyperspectral imaging with the wavelength region of 650-1000 nm at every 5 nm intervals. A stepwise MLR method was used for developing firmness and soluble solids content calibration models. A three-wavelength firmness model confirmed the importance of peaks at around 675 and 980 nm (chlorophyll and water, respectively) for firmness of strawberries. A correlation coefficient of 0.786 was observed for 50% to full ripe group of strawberries. A five-wavelength soluble solids content model was developed using near-infrared wavelengths of above 800 nm where absorptions due to carbohydrate and sugar could be seen. A correlation coefficient of 0.87 for measuring soluble solids content in the 70% to full ripe group of strawberries was obtained using the five-wavelength prediction model.

Lu and Peng (2005) investigated the measurement of firmness in peaches using the hyperspectral imaging. In this study, a hyperspectral imaging system was used to acquire images of Red Haven and Coral Star peaches in a wavelength region of 500-1000 nm at 153 spectral bands. Two-parameter Lorentzian distribution function was used to fit the spectral scattering profiles for individual wavelengths with an average coefficient of determination (r^2) value of > 0.99 . The 677 nm wavelength, which was related to chlorophyll absorption, had the highest correlation among all wavelengths in MLR models for firmness. The best predictions (with $r^2 = 0.77$ and 0.58 for Red Haven and Coral Star peaches, respectively) of firmness were obtained when 10 or 11 wavelengths were combined in the MLR model. Non destructive and rapid predictions of firmness of peaches were estimated effectively using the hyperspectral scattering.

Sugar content and firmness of Empire and Red Delicious apples were predicted using the rapidly-acquired spectral data from a NIR sensing method in the wavelength range of 900-1500 nm (Lu and Ariana, 2002). Data at two sensing positions of samples, 3.5 and 5.5 mm from the illumination source, were evaluated. The PLSR method gave prediction values of ≤ 0.81 with an error of 0.5-0.7 for sugar content. Relative reflectance spectra of samples gave better predictions when comparing with the results of ratio spectra. In this method, sugar content predictions surpassed firmness predictions.

3.1.2 Vegetables

3.1.2.1 Determination of quality attributes

Gowen et al. (2009) identified freeze damaged white button mushrooms at early stages using a pushbroom line-scanning hyperspectral imaging in the wavelength region of 400-1000 nm. The SNV method was used to pre-process the hyperspectral reflectance data which were obtained by imaging mushrooms from various positions. The PCA and LDA were used for classifying freeze damaged mushrooms and healthy ones. Classification accuracies were 100 and 97.9% for whole undamaged mushrooms and freeze damaged mushrooms, respectively, for an independent test set.

Detection of bruises on pickling cucumbers was evaluated using NIR hyperspectral imaging (Ariana et al., 2006). Bruises in cucumbers were created by hidden internal damages which were caused from mechanical injuries at the time of harvesting, transporting, and handling. Band ratio, band difference methods, and Principal component analysis (PCA) were used to discriminate bruised cucumbers from healthy ones. Bruised tissues, which increased over time, had always lower reflectance intensities than healthy tissues. Spectral region of 950-1350 nm with a band width of 8.8 nm was

reported as the best region for detecting bruises using the PCA. Self-healing of bruised tissues decreased detection accuracies from 95 to 75% over a period of 6 days after injury using the PCA. Detection accuracies of 93 and 82% were reported for the best band ratio of 988/1085 nm, and of 89 and 84% for the best band difference of 1346 and 1425 nm.

Cheng et al. (2004) investigated cucumber chilling damages using a novel integrated PCA and Fisher's linear discriminant (FLD) method in hyperspectral image analyses. The integrated PCA-FLD method performed better than the PCA and FLD methods when used separately for inspecting chilling damages.

3.1.3 Hyperspectral imaging in feeds and veterinary products

Recent developments in spectroscopy have been directed to the effective monitoring of food and feed products using the spectral imaging. Pierna et al. (2006) developed an innovative method for effective screening of compound feeds using the NIR hyperspectral imaging. The SVM was used to produce discriminant equations from hyperspectral data to determine the sample composition using a classification tree method. Classification accuracies were 99-100% for more than 36 combinations of C (a penalty that was added to take into account that the samples cannot be properly separated) and σ (the width of the Gaussian function). The SVM could become a promising classification tool for determining the presence of different feed ingredients.

In broilers, detection of cecal contaminants in the visceral cavity region is essential while performing safety inspections. Cecal portions of the digestive tract were used for collecting digestive materials for studies. Cecal feces were darker in colour than duodenum or colon feces. Park et al. (2005) used a hyperspectral imaging system for

distinguishing internal cecal contaminants in the visceral cavity areas of bisected broiler carcasses (Park et al., 2005). Band ratio, thresholding, and median filtering algorithms were used to detect fecal contaminants. Detection accuracy was 92.5% to identify cecal contaminants when using a fecal threshold value of 1.05. Park et al. (2002) identified fecal and ingesta contaminations on poultry carcasses using a hyperspectral imaging system. The system had a prism-grating-prism spectrograph, fibre optic line lighting, motorized lens control, and hyperspectral image processing software. Wavelengths of 434, 517, 565, and 628 nm were identified important by the PCA for hyperspectral images acquired at a wavelength region of 400-900 nm with 512 spectral bands and further used. Images processed using the band ratio of dual wavelengths (565/517 nm) and histogram stretching were effective to detect fecal and ingesta contaminations. Detection accuracies were 97.3 and 100% for linear and non-linear histogram stretching, respectively. Fecal contaminants of duodenum, cecum, colon, and ingesta on poultry carcasses were detected effectively using the hyperspectral imaging.

3.1.4 Cereals

3.1.4.1 Determination of quality attributes

Near-infrared hyperspectral transmittance imaging in the wavelength region of 750-1090 nm was evaluated for predicting the constituent concentrations and analyzing the quality of single kernels of maize (Cogdill et al., 2004). The standardized absorbance spectra were used for developing the PLSR and PCR models to predict moisture and oil contents. Common hyperspectral data pre-processing methods such as standard normal variate (SNV), detrending (DET), multiplicative scatter correction (MSC), wavelength selection by generic algorithm, and no pre-processing were evaluated for performance.

Standard error of cross validation was 1.20% (with a relative performance determinant (RPD) of 2.74) and 1.38% (with an RPD of 1.45) for moisture and oil content models, respectively.

Mahesh et al. (2008) investigated the feasibility of using the NIR hyperspectral imaging for identifying western Canadian wheat classes. Hyperspectral images were acquired from a wavelength region of 960-1700 nm at 10 nm wavelength intervals. Spectral information of hyperspectral data were presented in relative reflectance intensities. Classification accuracies were > 94% and > 90% for the statistical and ANN classifiers, respectively.

Zhang et al. (2007) evaluated the support vector machine in the classification of wheat kernels, which were infected by storage fungi (*Aspergillus niger* van Tieghem, *Aspergillus glaucus* group, and *Penicillium* spp.), using the NIR hyperspectral images acquired from the wavelength region of 1000-1600 nm. The PCA was used for reducing dimensions of pattern vectors. Classification accuracies were 92.9, 87.2, 99.3, and 100% for identifying wheat kernels infected by *Aspergillus niger*, *Aspergillus glaucus*, *Penicillium* spp., and healthy kernels, respectively, using a multi-class support vector machine having kernel of radial basis function for classification.

Choudhary et al. (2009) extracted wavelet features for identifying wheat classes using bulk sample images taken from the NIR hyperspectral imaging camera in the wavelength region of 960-1700 nm. The LDA, QDA, and a back propagation neural network (BPNN) classifiers were used for the class identification. The highest average classification accuracies were 99.1 and 92.1% for the LDA and BPNN, respectively. The

principal component (PC) 2 features registered 80% classification accuracy using the PCA for wavelet features.

Singh et al. (2010) detected midge damaged wheat kernels using images acquired using two cameras: a short-wave NIR (700-1100 nm) and an area scan digital colour. The LDA, QDA, and Mahalanobis classifier were used for classification purposes having statistical and histogram features extracted from hyperspectral image data of significant wavelengths. The highest classification accuracies were 95.3-99.3% to discriminate healthy and midge damaged kernels using the combined NIR hyperspectral and top 10 colour image features as input.

The NIR hyperspectral imaging was used in the wavelength range of 1000-1600 nm to detect single kernels of insect-damaged wheat (Singh et al., 2009). Hyperspectral images were acquired for healthy, and *Sitophilus oryzae* (L.), *Rhizopertha dominica* (F.), *Cryptolestes ferrugineus* (Stephens) damaged kernels of wheat. The multivariate image analysis (MVI) was used for reducing dimensions of hyperspectral data. Statistical and histogram features of significant wavelengths were used as input for the LDA and QDA giving classification accuracies of 85-100% in identifying healthy and insect-damaged kernels of wheat.

Williams et al. (2009) evaluated the classification of maize kernels based on hardness levels by acquiring images using the NIR hyperspectral imaging camera in the wavelength region of 960-1662 nm and using the SWIR hyperspectral pushbroom imaging system in the wavelength range of 1000-2498 nm. Background, bad pixels, and shading effects from absorbance images were removed using an exploratory PCA

method. The PC3 of NIR hyperspectral data and the PC2 of SWIR hyperspectral data gave noticeable differences between glassy and floury endosperms of maize kernels. The root mean square errors of prediction (RMSEP) were 0.18, 0.18, and 0.29 for 12-kernel, 24-kernel NIR hyperspectral image, and SWIR image, respectively, for the partial least squares discriminant analysis (PLS-DA) models.

Prediction of the α -amylase in single kernels of CWRS and CWAD wheat was evaluated, by investigating different degrees of sprout damages, using the SWIR hyperspectral imaging with the spectral range of 1000-2500 nm (Xing et al., 2009). The PLSR was used for predicting α -amylase activity giving regression coefficient (R^2) values of 0.54 and 0.73 for CWAD and CWRS wheat, respectively. Classification accuracy of 80% was attained for discriminating high levels from low levels of α -amylase activity in CWRS wheat.

Lawrence et al. (2003) demonstrated a geometric control point correction, which is one of the modified calibration methods, to reduce smile and keystone effects from a pushbroom type hyperspectral imaging system that could be used for agricultural inspection studies. Also, wavelength, distance, and pixel-by-pixel percent reflectance calibration methods were performed on hyperspectral data. These calibration procedures were effective for the wavelength range of 430-900 nm. The stability of calibration over time had to be analyzed over time to improve calibration procedures for real-time analysis.

3.1.5 Non agri-food applications

Hyperspectral imaging is gaining popularity not only in food and agricultural sectors but also in other industrial applications. Materials classification can be achieved from hyperspectral images acquired in the NIR wavelength region of 900-1700 nm. Tatzer et al. (2005) investigated inline material sorting to classify raw and colour cardboards, newspaper, and printed papers using the NIR hyperspectral imaging. It was observed that mean classification accuracies of 93, 81, 95, and 91% were obtained for raw cardboards, coloured cardboards, newspaper, and printed papers, respectively, by combining the PCA with linear discriminant analysis (LDA).

3.1.6 Limitations of hyperspectral imaging

Hyperspectral imaging has a few limitations like any other technique. The major limitation is the production of large amount of data. This leads to a significant increase in the computational time for extracting main features from the spectral images. Hence, it becomes difficult for hyperspectral imaging to be used for on-line inspection of food, agricultural, and industrial products. Commercial hyperspectral imaging software and algorithms such as ENVI (Research Systems Inc., Boulder, CO) are developed for remote sensing purposes and are not suitable for performing quality and safety inspections of food and agricultural products. Development of reliable, fast, and efficient algorithms for hyperspectral data analyses becomes crucial. As machine vision is being combined with spectroscopy in the hyperspectral imaging, hardware implementations become more difficult than those in imaging or spectroscopic systems alone. Reliable and accurate hyperspectral data are produced using proper lighting design, system calibration, and image correction and affected by the performance of CCD camera.

3.2 Applications of the NIR spectroscopy

Quality parameters in grains, milk, meat, fruits, and beverages are determined using NIR spectroscopy. Functional, compositional, and sensory analyses are performed for verifying the adulteration and genuineness of food products. Also, the NIR spectroscopic applications for determining quality parameters of agricultural and biological materials have been discussed elsewhere (Sivakumar, 2007). Monochromators (H1034B, Jobin Yvon Inc., Edison, NJ) and diode array spectrometers (8452A, Hewlett-Packard Inc., Palo Alto, CA) are used for acquiring the NIR absorbance intensities of samples using diffuse reflectance and/or transmittance principles. The NIR absorbance/reflectance/transmittance data are pre-processed using standard normal variate and detrending (SNV-DET) and multiplicative scatter correction (MSC) methods. The NIR calibration models are developed using multiple linear regression (MLR), principal components regression (PCR), and partial least squares regression (PLSR) methods (Osborne, 2006). The NIR absorption of C-H, R-OH, R-NH₂, ArOH, and combination bands, first, second, and third overtones are shown in Fig. 3.1.

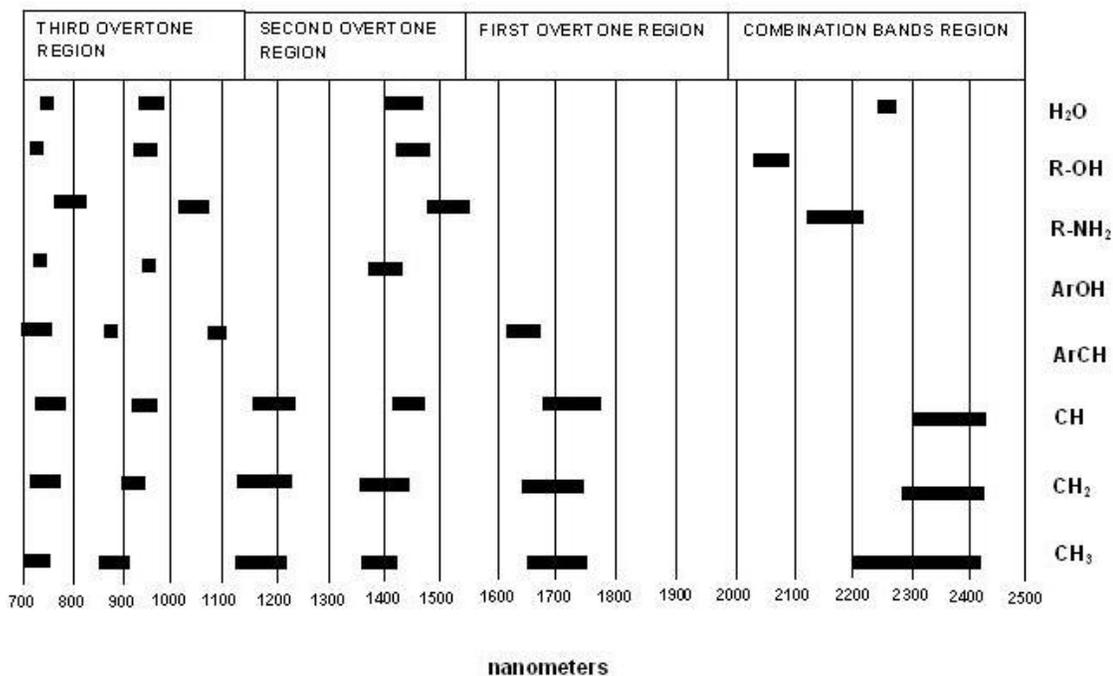


Fig. 3.1. The NIR absorption of C-H, R-OH, R-NH₂, ArOH, and ArCH combination bands, first, second, and third overtones (modified from Anonymous, 2011).

3.2.1 Classification of cereal grains

Delwiche and Massie (1996) classified single kernels of wheat using the Vis/NIR reflectance values acquired from three hard wheat (hard white (HWH), hard red spring (HRS), and hard red winter (HRW)) and two soft wheat (soft red winter (SRW) and soft white (SWH)) samples. A diode array spectrometer and a spectrophotometer were used for gathering reflectance intensities in the vis-NIR region and the NIR region, respectively. Classification accuracies were > 97% for differentiating red and white wheat using a seven-factor PLSR method. A five-factor MLR method gave classification accuracies of > 96% for identifying red and white wheat.

Dowell (2000) used the NIR spectroscopy to segregate vitreous and non vitreous durum wheat using a diode array spectrometer in the wavelength range of 400-1700 nm

at 5 nm intervals. Wheat kernels were classified in this study into two groups, i.e., easily distinguishable and easily non distinguishable. Identification of vitreous wheat kernels was improved by increasing the discriminant value from 0.5 to 0.7 in the PLSR model. The Bureau of Appeals and Review (BAR) method was used as the reference which in turn reduced the performance of the PLSR model. Classification accuracies were 80 and 70% for vitreous and non vitreous kernels, respectively, using the PLSR with a discriminant value of 0.5. The NIR spectroscopy can be used for quantifying vitreousness using the difference in the NIR absorption for protein and starch contents of durum wheat. Partially waxy and wild wheat varieties were identified using the NIR spectroscopy (Delwiche and Graybosch, 2002). The NIR reflectance spectra of ground wheat samples were collected using a spectrophotometer in the wavelength range of 1100-2498 nm at 2 nm intervals. An iodine binding blue complex colorimetric method was used for measuring apparent amylase contents of wheat samples. The PCA was used for reducing dimensions of spectral data preceded to a stepwise regression. One-out cross validation was used for determining the optimal number of discriminant functions. Perfect classification was not possible because of the overlapping effects of amylase contents of wheat classes.

Wang et al. (2002) conducted a feasibility study using the vis-NIR spectroscopy to discriminate dark hard vitreous (DHV) kernels from non dark hard vitreous (NDHV) kernels of wheat. A diode array spectrometer was used to collect reflectance intensities of wheat samples in the wavelength range of 400-1700 nm. This wavelength region was segmented into three regions, i.e., 500-750 nm (visible), 750-1700 nm (NIR), and 500-1700 nm (vis-NIR) for calibration purposes. Dorsal side kernel orientation and the

selection of specific wavelength regions produced significant improvements in detecting the DHV kernels. Lower classification accuracies (91- 97%) were observed for bleached kernels than non bleached kernels (97-100%) in the two-class PLSR model. The vis-NIR region or the NIR region itself was more suitable than the visible region for detecting the DHV kernels of wheat. Cocchi et al. (2006) measured the degree of adulteration in durum wheat flour with common bread wheat flour using the NIR spectroscopy. The PLSR and wavelet interface to linear modeling analysis (WILMA) methods were used having raw and SNV pre-treated data as input. A spectrophotometer was used in the wavelength range of 400-2498 nm at 2 nm intervals. Spectral pre-treatment reduced root mean square error for calibration (RMSEC) (= 0.2903), root mean square error of cross validation (RMSECV) (= 0.7215), and root mean square error of prediction (RMSEP) (= 0.3974) values of the eight-variable PLSR model. The WILMA-PLS model with 60 coefficients and seven latent variables had a minimum RMSEP value of 0.447.

3.2.2 Detection of fungal, insect, and other damages in cereal grains

Baker et al. (1999) differentiated kernels infested by larval and pupal stages of rice weevils (*Sitophilus oryzae* (L.)) which were parasitized by specific mites, *Anisopteromalus calandrae* (Howard), from uninfested and unparasitized kernels of wheat. Uninfested kernels, kernels infested with weevil larvae, kernels infested with weevil pupae, kernels containing parasitoid larvae, and kernels containing parasitoid pupae were used. A diode array NIR spectrometer was used to collect absorbance intensities of wheat kernels in the wavelength region of 400-1700 nm. A thirteen-factor PLSR model detected the rice weevil infestation in wheat with r of 0.90 and SECV of 0.15. Levels of the NIR absorption intensities differed due to the compositional

difference in chitin and cuticle contents. Misclassifications occurred for kernels with small parasitoids or weevils that absorbed a small amount of the NIR radiation. The NIR spectroscopy was effective to separate larval or pupal stages of insects or parasitoids in wheat.

The NIR reflectance spectroscopy was used to identify heat damaged wheat kernels from healthy kernels (Wang et al., 2001). Reference data were collected using the mixogram for identifying heat damages in wheat. A rapid viscosity analyzer was used for measuring gelatinization, pasting, and set back profiles of wheat. A diode array spectrometer was used for measuring the vis-NIR reflectance intensities in the wavelength region of 400-1700 nm. The PLSR and two-wavelength regression models were developed to separate heat damaged from undamaged kernels. They observed that heat damaged kernels were darker than undamaged kernels. The classification accuracy was 100% for the 7-factor PLSR model for 750-1700 nm for discriminating heat damaged kernels from undamaged ones. The 2-wavelength regression model for 985-1050 nm had classification accuracies of 97.5 and 96.8% for calibration and test sets, respectively. The NIR spectroscopy can be used to identify heat damaged and healthy wheat.

Wang et al. (2002) conducted a feasibility study to differentiate sound soybean kernels from damaged ones using the NIR spectroscopy. Six categories of soybean seeds, i.e., sound, weather damaged, frost damaged, sprout damaged, heat damaged, and mould damaged, were used. A diode array spectrometer was used to collect the NIR reflectance intensities in the wavelength range of 400-1700 nm at 5 nm intervals. Two and six-class PLSR models were developed using a commercial PLS software. Also, two and six-class

BPNN models were developed for differentiating sound and damaged kernels of soybean. Two-class PLSR model was developed using the values from wavelength regions of 750-1690 nm and 490-1690 nm. Accuracies were > 99.3% and > 99.5% for calibration and validation, respectively, for classifying damaged and sound kernels. The PLSR model, used spectral values from the visible wavelength region (490-750 nm), had classification accuracies of > 98.4% and > 97.8% for calibration and validation sets, respectively. Six-class PLSR model for 490-1690 nm had average classification accuracies of 75.2 and 74.5% for calibration and validation, respectively, in identifying sound and damaged soybean seeds.

Delwiche (2003) identified scab and mould damages in wheat using the NIR reflectance spectroscopy. First, mould-affected and scab-damaged kernels were visually separated from sound kernels in hard red spring wheat. A Zeiss MCS511 diode array spectrometer was used to acquire the NIR absorbance intensities in the wavelength region of 940-1700 nm. The NIR absorbance intensities of 1002-1704 nm at 6 nm intervals were used for modeling. Two kernel orientation types, i.e., crease down placement and random placement of kernels, were used in this study. The LDA with leave-one-out cross validation, soft independent modeling of class analogy (SIMCA), and the PCA were used. Cross validation accuracies of 89-98% and test set accuracies of 90.5-98.4% were obtained for the LDA for a two-way classification (sound kernels vs. damaged kernels (scab damaged + mould damaged)). Cross validation and test set accuracies were 89.3 and 86.4%, respectively, for the LDA with scores of 1st, 2nd, 7th, and 3rd principal component as input. Cross validation and test set accuracies were 85.3-86.7 and 83.6-85.8%, respectively, for 2-, 4-, and 6-factor SIMCA-PLS models. Classification

accuracies of scab-damaged, mould-affected, and sound kernels with random kernel orientations were equal to or less than the accuracies with the precise kernel orientation.

Maghirang et al. (2003) detected live or dead rice weevils at pupal and larval stages in single kernels of wheat using the NIR spectroscopy. Sound kernels, kernels infested with pupae, large larvae, medium sized larvae, and small larvae of rice weevils were selected using the x-ray imaging. A single kernel characterization system (SKCS) was used to collect data of single kernels in the wavelength region of 400-1700 nm at 5 nm intervals. The PLSR method was used to detect different stages of internal infestations using values from the wavelength region of 950-1690 nm in wheat. Ten key wavelengths were identified (990, 1135, 1210, 1250, 1370, 1395, 1425, 1510, 1610, and 1670 nm) using the PLS beta coefficients. Classification accuracies were > 90% to detect pupae or large larvae of rice weevils using the 5-7 factor PLSR models.

Perez-Mendoza et al. (2003) detected insect fragments in wheat flour using NIR spectroscopy. A diode array NIR spectrometer was used for collecting diffuse reflectance intensities in the wavelength region of 550-1700 nm. The NIR reflectance intensities were converted to absorbance values. The AOAC 972.32 floatation method was used to find out insect fragments in wheat flour. A PLSR model was developed using ten PLS factors to predict insect fragment levels in wheat flour. Six wavelengths (890, 1120, 1220, 1370, 1530, and 1630 nm) in the NIR region were identified as wavelengths responsible for the excitation of first, second, and third overtones of CH groups. Some of the main constituents of insect fragments, i.e., chitin and lipid, were responsible for the CH group absorption. The AOAC floatation method produced high accuracy in predicting insect fragments in wheat flour in 2 h. Classification accuracy was 83.3% for

classifying samples with 0, 35, and 75 insect fragments (< 130 fragment samples per 50 g flour class) and 90% for samples with 150, and 300 insect fragments (> 130 fragment samples per 50 g flour class). The NIR spectroscopy was not as sensitive as the floatation method to determine the US-Food and Drug Administration (US-FDA) levels of insect fragments (75 insect fragments per 50 g of flour) in wheat flour. They suggested that the NIR and mid-IR spectroscopy could be used to detect the US-FDA permissible levels of insect fragments in wheat flour, in future, with advancements in the field of spectroscopy.

Wang et al. (2004a) classified fungal damaged kernels from healthy kernels of soybean using NIR spectroscopy. Healthy kernels and kernels damaged by *Phomopsis*, *Cercospora kikuchii* (T. Matsumoto and Tomoy), soybean mosaic virus (SMV), and downy mildew in soybean were used. A diode array spectrometer was used to collect the NIR reflectance intensities in the wavelength region of 400-1700 nm from single kernels of soybean. The NIR reflectance intensities were interpolated to 5 nm intervals. Two (healthy vs. damaged) and five-way classifications (healthy vs. four types of damage) were developed using the PLSR and ANN methods using the vis-NIR reflectance intensities of 490-1690 nm as input. A ten-factor PLSR model had classification accuracies of > 99% for calibration and validation sets in classifying healthy kernels from damaged kernels of soybean. The highest average classification accuracies were 93.5 and 94.6% for calibration and validation sets for separating healthy and four types of fungal damage in soybean.

3.2.3 Determination of quality parameters in cereal grains

Hareland (1994) predicted the percent volume of flour particles for different wheat classes and milling methods using the NIR reflectance spectroscopy. A laser

diffraction method was used for finding out the percent volume and the PLSR model gave an accuracy of 96% for determining percent volume. Delwiche (1995) measured single kernel protein contents of wheat using the NIR transmittance spectroscopy. Nitrogen and food protein determinator and Kjeldahl methods were used for measuring protein contents. The NIR transmittance intensities (T) were acquired from the wavelength range of 740-1139 nm. The following formula was used for converting transmittance intensities into absorbance (A) values:

$$A = \log_{10} \left(\frac{1}{T} \right) \dots \dots \dots (3.1)$$

The NIR absorbance values of 850-1050 nm were used for developing a PLSR model. Three types of input data sets, i.e., no change in absorbance intensities, the MSC corrected absorbance intensities, and the MSC corrected second derivatives of absorbance intensities, were used. Model performances were improved for the second and third types of pre-processed data sets.

An NIR model was developed to measure starch structure and degree of processing of cereal products in the twin screw extrusion cooking (Guy et al., 1996). The NIR reflectance intensities were collected in the wavelength region of 1100-2500 nm at 4 nm intervals. The models were developed using forward stepwise regression to predict specific mechanical energy (SME) using principal component scores as input. Principal components were identified for the NIR reflectance intensities in the wavelength region of 1300-1800 nm. The first two principal components explained > 95% of variations of the NIR reflectance intensities. Performance of regression models was affected by damage in hydrogen bonds during extrusion cooking.

The NIR reflectance spectroscopy was used for measuring protein contents in single kernels of wheat (Delwiche, 1998). Two NIR wavelength regions, 1100-2498 nm and 1100-1798 nm, were used for scanning samples at 2 nm intervals. Protein contents were determined by the combustion method. Two hard and soft wheat classes were used and PLSR and MLR models were developed. The MSC was used to reduce the spectral variability and the NIR wavelength region of 1100-1400 nm was found effective for determining protein contents. An error of 0.411% was found from the chemometric analysis for protein. Inclusion wheat classes from more than one crop year in the calibration set might improve accuracy.

Pearson et al. (2001) detected aflatoxin in corn using the vis-NIR transmittance and reflectance spectroscopy. Corn samples with varying levels of aflatoxin were selected using the black light examination of bright greenish-yellow fluorescence (BGYF) characteristics. A silicon photo diode array fibre optic spectrometer was used to collect transmittance spectra of single kernels of corn in the wavelength region of 500-950 nm. A diode array NIR spectrometer was used for collecting reflectance intensities in the wavelength region of 550-1700 nm. The nineteen-point Savitzky-Golay second order filtering operation was used for smoothing the transmittance spectra. Mahalanobis distance method was used for grouping corn samples at three aflatoxin levels, i.e., 1, 10, and 100 ppb. A PLSR method was used to detect aflatoxin levels of corn. The USDA-FGIS (United States Department of Agriculture-Federal Grain Inspection Service) aflatest affinity chromatography was used for measuring aflatoxin levels of corn. In the discriminant analysis, better classification accuracies were obtained to identify 0, 1 – 10, > 100 ppb aflatoxin levels using transmittance and reflectance intensities as input. Error

rates were 52-56% and 65-87% for transmittance and reflectance intensities, respectively, to classify 10-100 ppb levels of aflatoxin in corn using the discriminant analysis. Germ up and germ down kernel orientation methods improved accuracies to $\geq 84\%$ to classify 10-100 ppb levels of aflatoxin in corn using the six-factor PLSR model. The vis-NIR spectroscopy could be used to detect aflatoxin levels in corn.

Wesley et al. (2001) measured gliadin and glutenin contents by developing the NIR models for wheat. Gliadin and glutenin are related to the quality of wheat protein. The PLSR and curve fitting methods were used for predicting gliadin and glutenin contents. Size exclusion high performance liquid chromatography (SE-HPLC) was used for measuring gliadin and glutenin contents. Pre-processing of the NIR spectra was carried out using standard normal variate-detrending (SNV-DET) methods. The performance of the PLSR model was better than the curve fitting method. The PLSR model had coefficient of determination (r^2) values of 0.83 and 0.78 for glutenin and gliadin contents, respectively. The curve fitting method had r^2 values of 0.71 and 0.46 for glutenin and gliadin contents, respectively. For the PLSR model, standard errors of cross validation (SECV) were 0.38 and 0.43 for glutenin and gliadin contents, respectively. The standard errors of prediction (SEP) of a curve fitting method were 0.65 and 1.02 for glutenin and gliadin contents, respectively. The curve fitting method could rank samples qualitatively, i.e., high, medium, and low, based on amount of glutenin and gliadin contents in wheat.

Detection of fumonisin using the reflectance and transmittance spectroscopy was evaluated (Dowell et al., 2002). Two different spectrometers, i.e., fiber optic spectrometer (Model S2000, Ocean Optics, Dunedin, FL) and the NIR spectrometer (Perten

Instruments, Springfield, IL), were used for collecting transmittance (550-1050 nm) and reflectance (400-1700 nm) intensities from single kernels of corn. A fluorometer was used for finding out the total fumonisin contents. A PLSR method was used to develop models to detect the fumonisin contents (< 10 ppm and ≥ 10 ppm) of corn. The Mahalanobis distance method was used for grouping based on fumonisin levels. Classification errors were 0% and $< 7.2\%$ for identifying 1-10 ppm and > 100 ppm levels of fumonisin, respectively. Misclassifications were more for 10-100 ppm levels of fumonisin (error rate = 23-73%). Key transmittance wavelengths (650, 710, 935, and 990 nm) were identified using the PLS beta coefficients. Four reflectance wavelengths (590, 995, 1200, and 1410 nm) were also identified. The detection of fumonisin at a minimum FDA threshold levels of 2-4 ppm in corn using the NIR transmittance or reflectance spectroscopy was not possible.

Ruan et al. (2002) developed ANN models to measure mycotoxin, i.e., deoxinivalenol (DON), in barley using the NIR spectroscopy. The DON concentration of barley was measured using gas chromatography/mass spectrometry (GC/MS) method. The absorbance values from bulk barley samples were collected from the wavelength range of 400- 2500 nm at 2 nm intervals. Barley with varying DON levels were clearly separated in the NIR wavelength region of 1500-1800 nm. A three-layer BPNN model was developed using raw NIR absorbance values as input. Wavelengths of 400-700 and 700-1100 nm were crucial in predicting the DON levels in barley. The ANN models, which used values from 400-700 and 700-1100 nm, produced the best R^2 values of 0.921 and 0.912; and produced the minimum SEP values of 3.351 and 3.706. The ANN models

(400-2500 nm) at 2 nm and 4 nm intervals produced the best r^2 values of 0.933 and 0.923, respectively; and produced the minimum SEP values of 3.097 and 3.431, respectively.

Petterson and Aberg (2003) determined ergosterol and DON levels in wheat using the NIR transmittance spectroscopy. In addition, insect and mite infestations were studied using the NIR transmittance spectra of wheat. The wavelength region of 570-1100 nm was used for acquiring the NIR transmittance intensities. Gas chromatography (GC) and high performance liquid chromatography (HPLC) were used for measuring the DON levels of wheat. Eleven to thirteen PLS factors were extracted for modeling from three wavelength regions, i.e., normal (850-1100 nm), extended (570-1100 nm), and reduced (670-1100 nm). Separate PLSR models were developed for the fungal infection in wheat grown from Norway and Austria. The eleven-factor PLSR model for 670-1100 nm for Nordic samples produced r and SECV values of 0.984 and 381 μg DON per kg of wheat, respectively, for 670-1100 nm. The NIR transmittance spectroscopy could be helpful to measure the DON and ergosterol (a fungal metabolite) levels in cereal crops.

Miralbes (2004) determined quality parameters of wheat flour using the NIR models. The AACC standard methods were used for measuring quality parameters of wheat, i.e., moisture content, protein, wet gluten, dry gluten, and ash content. The NIR transmittance intensities were collected from the wavelength region of 850-1048 nm at two nm intervals. A modified PLSR model was developed after correcting spectral variations of the NIR transmittance intensities using the SNV-DET method. The r^2 was 0.99 for the PLSR validation in predicting protein and moisture contents. Minimum SEPs were 0.14 and 0.15 for predicting protein and moisture contents, respectively. The NIR

transmittance spectroscopy can control quality parameters of wheat flour during the online monitoring in milling industries.

Wang et al. (2004b) developed a linear regression model for determining moisture contents of ground wheat using the NIR spectroscopy. The NIR reflectance intensities were collected from the wavelength range 850-2000 nm at 5 nm intervals. The NIR reflectance intensities (R) were converted to absorbance values (A) using the following formula:

$$A = \log_{10} \left(\frac{1}{R} \right) \dots\dots\dots (3.2)$$

Linear regression models were developed for the averaged NIR spectra and for first derivatives of averaged NIR spectra. Pre-processing methods such as the MSC and SNV were used for correcting the NIR absorbance values. The r^2 and RMSEC values were 0.972 and 0.239, respectively, using the first derivatives of averaged NIR spectra. Baseline elimination and resolution of overlapping peaks improved the performance of models developed from first derivatives of the NIR absorbance intensities. Using the averaged NIR spectra as input, the r^2 and RMSEC values were 0.793 and 0.541, respectively.

3.3 Summary of literature review

The NIR hyperspectral imaging and the NIR spectroscopy have been used in many applications for inspecting biological and agricultural products (Sivakumar, 2007; Singh, 2009; Gowen et al., 2007). Tables 3.1 and 3.2 show the summary of hyperspectral and spectroscopic applications, respectively, in various agricultural and/or non-agricultural products. Hyperspectral imaging is a combination of both spectroscopic and

image processing techniques. Reference methods such as Kjeldahl method, nitrogen combustion method, and USDA-FGIS aflatoxin affinity chromatography, GC/MS, and SE-HPLC are time consuming, labour intensive, and skill demanding. It is apparent, after conducting an extensive review on hyperspectral imaging and spectroscopy, that hyperspectral imaging is considered more useful than other methods for performing chemometric analyses. Hyperspectral imaging can generate a full reflectance or absorbance spectrum for each pixel of an image. It is a robust technique in which biological materials are easily characterized and analysed. Hyperspectral imaging for commercial applications generally requires a high speed setup for hardware and advanced image processing software. It is an established method that has an analytical ability comparable to traditional techniques for sample separation and quality measurement. Also, it requires less analytical time than the conventional methods. Compositional distribution has been measured using PC score images and concentration maps. The type of quality parameter and the objective of the study determine the selection of proper wavelength range, pre-processing method, and type of analysis. The PCA and PLS methods are commonly used for the dimension reduction of hyperspectral data. The PCA based multivariate image analysis can effectively reduce the hyperspectral data dimensions and aid in selecting appropriate wavelengths. It is possible to consistently reproduce reflectance intensities from analog to digital (A/D) counts when known reflectance standards and values are used. Multispectral imaging can be designed using selected wavelengths where online monitoring of biological materials become easier and faster. In hyperspectral imaging, a homogeneous and intense illumination, which is not harmful to agricultural materials, is required. The LDA and QDA, which are helpful for

developing supervised classification models, have been used extensively for sample classification. The performance of statistical classifiers was better than the neural network classifiers in the LWNIR hyperspectral imaging because of the possible linear response of the InGaAs detector. The BPNN classifiers give better results if a non-linear type of training data set is used as the input.

Table 3.1. Summary of hyperspectral imaging applications in fruits, vegetables, cereals, feeds, veterinary, and industrial products

Mode	Product	Analysis	Wavelength range (nm)	Classification	Reference(s)
Fruits:					
Reflectance	Strawberry	Measurement of firmness and soluble solids content	650-1000	Stepwise MLR	Nagata et al. (2005)
Reflectance	Apple	Detection of bruises	400-1000	PCA	Xing et al. (2005)
Reflectance	Apple	Detection of bruises	900-1700	PC and MNF transforms	Lu (2003)
Reflectance	Apple	Detection of fecal contaminations	450-851	PCA	Kim et al. (2002)
Fluorescence and Reflectance	Apple	Detection of feces	452-729 and 465-900	Band ratio	Lefcourt et al. (2006)
Reflectance	Apple	Detection of surface defects and contaminations	430-900	Asymmetric second difference method	Mehl et al. (2004)
Reflectance	Apple	Measurement of bitter pits	900-1700	Discriminant PLS	Nicolai et al. (2006)
Reflectance	Strawberry	Determination of m.c., TSS, and pH	400-1000	PLS, MLR	EIMasry et al. (2007)
Reflectance	Apple	Defects	418-918	PCA, band difference, and band ratio	Lee et al. (2005)
Reflectance	Apple	Detection of fecal contaminants	447-951	PCA, band ratio, and asymmetric second difference	Liu et al. (2007)
Scattering	Peach	Measurement of firmness	500-1000	MLR	Lu and Peng (2005)
Reflectance	Apple	Measurement of sugar content and firmness	900-1500	PLSR	Lu and Ariana (2002)

Reflectance	Apple	Starch index determination	900-1700	PCA	Peirs et al. (2003)
Fluorescence	Apple	Measurement of fruit skin and flesh colour, firmness, soluble solids content, and titratable acid	500-1040	PCA with neural network	Noh and Lu (2007)
Reflectance	Strawberry	Measurement of firmness and soluble solids content	450-650	Stepwise MLR	Nagata et al. (2004)
Vegetables:					
Reflectance	White button mushroom	Freeze damage	400-1000	PCA, LDA	Gowen et al. (2009)
Reflectance	Cucumber	Detection of bruises	900-1700	PCA, band ratio, and band difference	Ariana et al. (2006)
Reflectance	Cucumber	Inspection of chilling damages	447.3-951.2	Integrated PCA-FLD	Cheng et al. (2004)
Cereals:					
Transmittance	Maize	Prediction of moisture and oil contents	750-1090	PLSR and PCR	Cogdill et al. (2004)
Reflectance	Wheat	Identification of composite wheat classes	960-1700	LDA, QDA, and ANN	Mahesh et al. (2008)
Reflectance	Wheat	Classification of fungi-infected wheat kernels	1000-1600	SVM	Zhang et al. (2007)
Reflectance	Wheat	Identification of wheat classes	960-1700	Wavelet, LDA, QDA, and ANN	Choudhary et al. (2009)
Reflectance	Wheat	Identification of midge-damaged wheat kernels	700-1100	LDA, QDA, and Mahalanobis	Singh et al. (2010)

Reflectance	Wheat	Identification wheat classes and moisture levels of composite bulk samples	960-1700	LDA and QDA	Mahesh et al. (2011)
Reflectance	Wheat	Detection of insect damages in single kernels of wheat	1000-1600	MVI, LDA, and QDA	Singh et al. (2009)
Reflectance	Maize	Classification of maize kernels based on hardness	1000-2498	PLS-DA	Williams et al. (2009)
Reflectance	Wheat	Prediction of α -amylase contents	1000-2500	PLSR	Xing et al. (2009)
Agri-related products:					
Reflectance	Feeds	Screening of compound feeds	900-1700	SVM	Pierna et al. (2006)
Industrial products:					
Reflectance	Cardboards	Sorting of raw and colour cardboards, newspaper, and printed papers	900-1700	PCA, LDA	Tatzer et al. (2005)

Table 3.2. Summary of spectroscopic applications in cereal grains

Mode	Product	Analysis	Wavelength range (nm)	Classification	Reference(s)
Reflectance	Wheat	Classification of single kernels of wheat	551-750 (colour), 1120-2476 (intrinsic properties)	PLSR, MLR	Delwiche and Massie (1996)
Absorbance	Wheat	Classification of vitreous and non vitreous kernels	400-1700	PLSR	Dowell (2000)
Reflectance	Wheat	Identification of partially waxy and wild wheat varieties	1100-2498	Stepwise regression	Delwiche and Graybosch (2002)
Reflectance	Wheat	Classification of dark hard vitreous and non dark hard vitreous kernels	400-1700	PLSR	Wang et al. (2002)
Reflectance	Wheat	Measurement of adulteration in durum wheat flour	400-2498	PLSR, WILMA	Cocchi et al. (2006)
Absorbance	Wheat	Detection of insect infestations	400-1700	PLSR	Baker et al. (1999)
Reflectance	Wheat	Detection of heat damaged kernels	400-1700	PLSR	Wang et al. (2001)
Reflectance	Soybean	Detection of damaged soybean kernels	400-1700	PLSR	Wang et al. (2002)
Reflectance	Wheat	Detection of mould and scab damages	940-1700	SIMCA, LDA, PCA	Delwiche (2003)
Reflectance	Wheat	Detection of insect infestations	400-1700	PLSR	Maghirang et al. (2003)
Reflectance	Wheat	Detection of insect fragments	550-1700	PLSR	Perez-Mendoza et al. (2003)

Reflectance	Soybean	Detection of fungal damaged seeds	490-1690	PLSR and ANN	Wang et al. (2004a)
Reflectance	Wheat	Percent volume of flour particles	400-2500	PLSR	Hareland (1994)
Absorbance	Wheat	Prediction of protein contents	740-1139	PLSR	Delwiche (1995)
Reflectance	Cereal products	Measurement of starch structure and degree of processing	1100-2500	Stepwise regression	Guy et al. (1996)
Reflectance	Wheat	Measurement of protein contents	1100-2498	PLSR, MLR	Delwiche (1998)
Transmittance and reflectance	Corn	Detection of aflatoxin levels	550-1700	PLSR	Pearson et al. (2001)
Reflectance	Wheat	Measurement of gliadin and glutenin contents	1100-2498	PLSR and curve fitting	Wesley et al. (2001)
Transmittance and reflectance	Corn	Detection of fuminosin	550-1050 (transmittance) and 400-1700 (reflectance)	Mahalanobis, PLS	Dowell (2002)
Absorbance	Barley	Measurement of deoxinivalenol (DON)	400-2500	BPNN	Ruan et al. (2002)
Transmittance	Wheat	Measurement of deoxinivalenol (DON)	570-1100	PLSR	Petterson and Aberg (2003)
Transmittance	Wheat	Measurement of quality parameters of wheat	850-1048	PLSR	Miralbes (2004)
Absorbance	Wheat	Determination of moisture content	850-2000	Linear regression	Wang et al. (2004b)

4. METHODS AND MATERIALS

This chapter focuses on discussing information on samples, sample preparation, the NIR hyperspectral imaging system, image analysis, spatial calibration, illumination standardization, image acquisition, image analysis, the PCA, pattern recognizing classifiers such as statistical and neural network classifiers, and multivariate regression algorithms such as the PLSR and the PCR.

4.1 Grain samples

Four western Canadian wheat classes, two red and two white, of varying degrees of hardness and protein grown at different locations in the prairie provinces (Manitoba, Saskatchewan, and Alberta) of western Canada and in 2007, 2008, and 2009 crop years were collected and used for this study (Table 4.1). The wheat classes used were CWRS, CPSR, CWHWS, and CWSWS. Climatic subdivisions of Canadian prairies were used for selecting sample locations (Putnam and Putnam, 1970). Sample locations were distributed over the humid prairie, the sub-boreal, the sub-humid prairie, and the semi-arid regions. Wheat samples, conditioned to three moisture contents (% wet basis), 13 (straight), 16 (tough), and 19% (damp) were used (CGC, 2004). Wheat samples for imaging were randomly selected by taking a Petri dish (90 mm in diameter and 11 mm in depth) full of grains for each location- and crop year-specific wheat from 2 kg samples. Moisture content of samples was determined by drying whole kernels at 130°C for 19 h using about 10 g samples, in triplicate (ASAE, 2003).

Table 4.1. Growing locations (crop years) near the listed towns from where wheat samples were collected.

CWRS	CPSR	CWHWS	CWSWS
Corning, SK (2008, 2009)	Edmonton, AB (2008, 2009)	Churchbridge, SK (2008, 2009)	Corning, SK (2008, 2009)
Dauphin, MB (2008, 2009)	Rosemary, AB (2008, 2009)	Kenton, MB (2008, 2009)	Jansen, SK (2008, 2009)
Domain, MB (2008, 2009)	Viking, AB (2007, 2009)	Limerick, SK (2008, 2009)	Kenton, MB (2008, 2009)
Melfort, SK (2008, 2009)	Wainwright, AB (2008, 2009)	Mather, MB (2008, 2009)	Nokomis, SK (2008, 2009)
Tisdale, SK (2008, 2009)	Corning, SK (2009)	Shaunavon, SK (2008, 2009)	Wilkie, SK (2008, 2009)
	Unity, SK (2008)		

4.2 Near-infrared hyperspectral imaging system

The near-infrared hyperspectral imaging system consisted of a near-infrared camera with two VariSpec liquid crystal tunable filters (LCTFs) (Model No. MIR06, Cambridge Research and Instrumentation Inc., Woburn, MA), a 25 mm F1.4 C-mount lens (Electrophysics Corp. Fairfield, NJ), a sample stage, and a light source controlled through a Dell Optiplex GX280 Intel(R) (Dell Inc., Round Rock, TX) computer (Figure 4.1).

An Indium Gallium Arsenide (InGaAs) camera (Model No. SU640-1.7RT-D, Sensors Unlimited Inc., Princeton, NJ), that could be operated in a room within a temperature range of 20-40°C, was used for acquiring images at different wavelengths in the NIR region of 960-1700 nm at 10 nm intervals. This system had a spatial resolution of 640×480 pixels with 27 µm pitch. Each session commenced by imaging a 10 cent Canadian coin of 17.96 mm known diameter to ensure same pixel size. For all tests, the pixel size was maintained and there was no need to adjust the distance between the camera lens and the sample throughout the study.

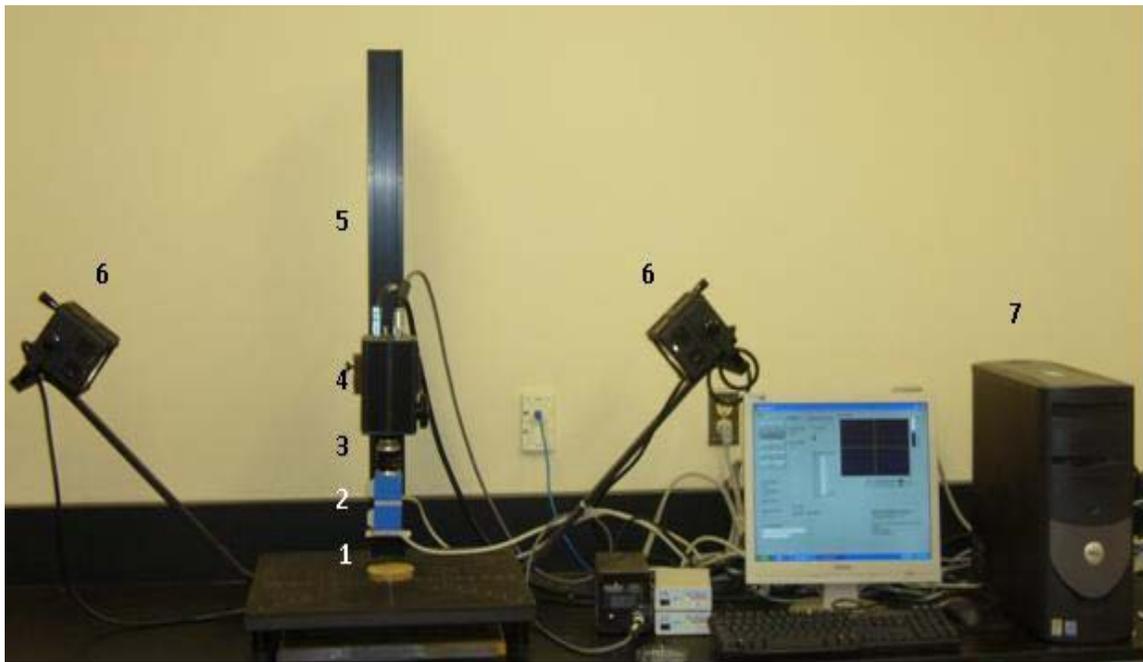


Fig. 4.1. The NIR hyperspectral imaging system.

1. Bulk wheat sample, 2. Liquid crystal tunable filter (LCTF), 3. Lens, 4. NIR camera, 5. Copy stand, 6. Illumination, 7. Data processing system.

The LCTFs had a 20 mm aperture size and a 10 mm transmission bandwidth. This high quality interference filter helped to rapidly select a wavelength in the NIR region without any vibration. This filter was attached to the camera, which ultimately produced 12-bit multispectral images. The data acquisition board (NI PCI-1422, National Instruments Corp., Austin, TX) was attuned to RS-422 signals generated from the camera system for image acquisition. The sample was illuminated by a pair of 300 W halogen lights (Ushio Lighting Inc., Cypress, CA) fitted on either side of the copy stand that supported the NIR imaging system. These halogen bulbs had the capacity to emit light in a wavelength range of 400- 2500 nm.

4.3 Spatial calibration

Martens and Naes (1992) explained necessary system calibrations that need to be done before acquiring the NIR hyperspectral images. Spatial resolution of pixels was calculated every time before starting the imaging process to make sure that the camera set up was not moved up or down from its initial position. A hyperspectral image of a Canadian 10 cent coin was acquired. Number of pixels occupied by the diameter of the coin was counted. It was confirmed that the number of pixels along the diameter of the ten cent coin were the same every time.

4.4 Illumination standardization

As the InGaAs camera had a wavelength-dependent dark current, dark current images (D) were recorded. A reference standard of 99% reflectance (Labsphere, North Sutton, NH) was used to acquire a reference image (W). Dark current and white reference spectra were collected prior to acquiring images in each session. In total, 2400 hyperspectral image cubes at each moisture level were acquired for analyses. Near-

and crop year specific samples per class \times 3 moisture levels \times 4 classes) and used for further analyses discussed in the next few sections. The NIR wavelength region was segmented into 75 slices, resulting in an NIR hypercube consisting of 75 images (an image per slice) with the first image at 960 nm. An area of 200 \times 200 pixels around the centre pixel was cropped from each image and used for further analyses. Cropping at the centre of the image was done to avoid pixels with poor reflectance intensities along the four edges of the image. The MATLAB (Mathworks Inc., Natick, MA, USA) codes were used for importing image files; and for displaying and analyzing hypercubes. Sensor defects produce insensitive pixels in images acquired from common imaging cameras. The insensitive pixels, otherwise called dead pixels, from the spatial area of hyperspectral images were removed by replacing with the median value of neighbourhood pixels. The signal-to-noise ratio was improved by co-adding the image slices at each wavelength during image acquisition. Bulk sample images were analyzed using a multivariate image analysis program written in MATLAB (Mahesh et al., 2011). The MIA was performed using the principal components analysis (Geladi and Grahn, 1996). The PCA, image segmentation, and spectral feature extraction codes were developed using MATLAB (Version 7, The Mathworks, Inc., Natick, MA). A four-layer BPNN and non-parametric statistical classifiers were used for classification.

4.6 Principal components analysis (PCA)

The PCA was used to reduce the spectral information. Here, a few correlated vectors were identified from a wide wavelength range. The top 2 or top 3 principal components could represent the most variations that existed in the raw data. The PCA

was used for conducting the multivariate image analysis (MIA) of the hyperspectral image cubes of bulk samples (Geladi and Grahn, 1996).

The MIA can be applied to bulk sample images after automatically selecting a region of interest (ROI). The ROI in this study was 200×200 pixels around the centre of the image. The hypercube data of the ROI of hyperspectral images were reshaped into a 2-D array by rearranging all the pixel intensities (reflectance) of a bulk sample image into a column at each of the 75 spectral bands. This resulted in a 40000×75 sized 2-D array in which 40000 is the total number of pixels in the cropped region of the bulk sample image and 75 is the total number of spectral bands. The PCA was then applied to the reshaped 2-D data set of hyperspectral images. First and second principal components were used for selecting the most significant wavelengths using the highest factor loadings. Loadings plot of principal components were used in identifying wheat classes. Discriminant classifiers could be developed for the data set for which dimensions were reduced using the PCA. Features such as maximum, minimum, and mean from the images at significant wavelengths were extracted for classification of fungal-infected wheat kernels from healthy kernels (Singh et al., 2007).

4.6 Pattern recognizing classifiers

Statistical and neural network classifiers were used for pattern recognition purposes. Paliwal (2002) discussed in detail about the basics of pattern recognizing, parametric and non parametric types of statistical methods, and neural network classifiers. Also, single layer and multi-layer neural network fundamentals were reviewed thoroughly in his work.

4.6.1 Statistical classifier

The class-conditional probability density function (*pdf*) of the input features is considered important in statistical classification. A non-parametric statistical classifier was used as the assumption of normal behaviour of distribution of features in the feature space was impossible. The statistical analysis software (SAS) (Version 9.1.3, SAS Institute Inc., Cary, NC) was used for developing the non-parametric statistical classifier. The linear and quadratic discriminant classifiers with a leave-one-out cross-validation method were used in this study. These were implemented using DISCRIM procedure of SAS. This SAS procedure uses the Bayes' theorem, in which the prior probability of group membership and group-specific densities are assumed, for determining the probability of an observation corresponding to a particular group. The STEPDISC procedure of SAS was used on the hypercube which was huge in size and had surplus features. Average squared canonical correlation values were used to rank the features in their order of significance. Top 10 wavelengths based on their contributions to the classification were identified. The level of contribution of each wavelength was found from the values of partial r^2 and average squared canonical correlation (ASCC). In the identification of the top 10 wavelengths, the wavelength with the highest level of contribution was identified first and subsequently removed from further analysis to find the next best wavelength. This analysis was continued until the 10th ranked wavelength was found.

4.6.2 Artificial neural network classifier

Design and implementation of neural networks were performed using Neuroshell 2 software (Ward Systems Group, Frederick, MD). The BPNN is the best suited for

classification. In all cases, the BPNN was trained and tested for 1000 epochs before using the validation data set. Over-training was arrested by saving the trained network every time once the test set got a new minimum average error for the classifier.

4.7 Protein and hardness measurements

Wheat samples were sent to the Central Testing Laboratory, Winnipeg for measuring protein contents using the American Organization of Analytical Chemists (AOAC) standard method 968.06. Dumas method was used for analyzing crude protein or total nitrogen contents of wheat samples. In this method, samples were combusted at 850°C and the resultant combustion gas was purified for nitrogen by removing CO₂ and O₂. Isolated nitrogen gas was analyzed using a thermal conductivity cell. Dumatherm Nitrogen analyzer (Gerhardt Instruments, Königswinter, Germany) and Leco FP-428 Nitrogen protein analyzer (Leco Corporation, St. Joseph, MI) were used for performing protein analysis in wheat samples.

Wheat hardness measurements were calculated at the Cereal Research Centre in Agriculture and Agri-Food Canada (AAFC), Winnipeg. A single kernel characterization system (SKCS) 4100 (Perten Instruments, Springfield, IL) was used for measuring hardness of wheat. In this method, mass of single kernels of wheat were measured. The wheat kernels were crushed individually between a rotor and a crescent gap. The force and conductivity measurements were taken as the wheat kernel was crushed. Using this information, mass, size, moisture, and hardness values were calculated for single kernels of wheat. Mean hardness index, mass, size, moisture, and their standard deviations were determined from the single kernel data obtained from a 300 kernel sample of wheat. Average SKCS hardness index values were used for classifying wheat based on hardness

(Table 4.2). The Scheffe's grouping was used for comparing protein and hardness means of wheat classes.

Table 4.2. Classification of wheat based on the Perten SKCS hardness index

(Source: Perten instruments, Springfield, IL)

Classification	SKCS hardness index	Wheat class
Extra hard	90 - 100+	Durum and other tetraploid wheat
Very Hard	80 - 89	Durum and some white wheat
Hard	65 - 79	CWES, CWRS, DNS, HRW, APH, AH, Others
Medium Hard	50 - 64	CWRW, CPSR, AH, ASW, Others
Medium Soft	40 - 49	Some CPSW, SRS, some ASW, Others
Soft	30 - 39	Some ASW, CWSWS, some club, Others
Very Soft	15 - 29	Some CWSWS, Club, SWW, SRW, AS, Others
Extra Soft	Up to 14**	SWW, SRW, Some AS, Others

HI = Hardness Index, CWES = Canada Western Extra Strong, CWRS = Canada Western

Red Spring, DNS = Dark Northern Spring (U.S.), HRW = Hard Red Winter, APH =

Australian Prime Hard, AH = Australian Hard, CPSR = Canada Prairie Spring Red,

CPSW = Canada Prairie Spring White, ASW = Australian Standard White, SRS = Soft

Red Spring, CWSWS = Canada Western Soft White Spring, SRW = Soft Red Winter,

SWW = Soft White Winter, AS = Australian Soft. ** HI data may be negative for some

extra soft wheat samples.

4.8 Multivariate regression algorithms

Two multivariate regression algorithms, the partial least squares regression (PLSR) and the principal components regression (PCR) were used for analyzing data from the full spectrum. It was hypothesized that the PLSR, one of the latent-variable regression techniques, was more advantageous than the PCR.

4.8.1 Partial least squares regression (PLSR)

The PLS analysis was used to form the prediction model between spectral responses of analyzed samples of wheat and their major quality parameters such as protein and hardness. The development PLSR algorithm, which was used in this study, was reported for determining quality parameters of strawberries (ElMasry et al., 2007). The PLS analysis was conducted between the NIR reflectance intensities and the parameter of interest (protein or hardness) using MATLAB. Maximization of the spectral data, which was formed from computing the average spectra with 75 wavelengths in wavelength range of 960-1700 nm, was done using the PLS analysis. The transfer of highly correlated data, which was also large in size and often co-linear in nature, into partial least square variables was performed. A mathematical relationship between a set of independent variables, X matrix ($N_{7200 \text{ samples}} \times K_{75 \text{ wavelengths}}$), and the dependent variable, Y vector ($N_{7200 \text{ samples}} \times 1$) was obtained for protein and hardness models using an overall data set having 7200 spectra in total. The dependent variable (Y), which was always a vector in this study, was formed using the values of one parameter (either protein or hardness) at a time. The independent variable or predictor data set (X) was formed using the reflectance values of 75 spectral bands for 7200 scanned wheat samples.

The optimal wavelengths were identified using the PLS β -coefficients. As maximum spectral information was explained, these wavelengths could be used for on-line multispectral imaging applications. Non-significant wavelengths, which had the low absolute values of β -coefficients, could completely be rejected as they had zero involvement in the prediction of intrinsic attributes of wheat. The prediction performance was estimated using the 10-fold cross validation method (Cogdill et al., 2004). In the k-fold cross-validation, which can estimate model performances accurately (Refaeilzadeh et al., 2011), the original data set was randomly divided into 10 subsample groups. Nine out of ten groups were used for calibration and the remaining one group was used for validation. The root mean squared errors of prediction (RMSEP) were estimated during the cross-validation process for the PLSR models. The probable prediction errors for the new samples were estimated using cross-validation. The cross-validation was repeated ten times by which data from each of the subsample group used exactly once for validating the model. Average of ten iterations was reported for the model performance. Estimated MSE were calculated for finding out optimal number of components for developing the PLSR model. This method was considered advantageous because all observations were used for training and validating the models. Also, each observation was used just once for validation.

4.8.2 Principal components regression (PCR)

The principal components regression (PCR) model were developed for predicting protein and hardness using the full spectral data set (960-1700 nm) of 7200 sample spectra to avoid prediction instabilities caused by co-linear nature of predictor data set (P). In this method, uncorrelated principal components, which had decreasing variances,

were generated from the NIR reflectance intensities. Regression models were developed between principal component scores and the attributes of interest (protein and hardness). A ten-fold cross-validation explained in the previous section, was used for validating the PCR models. The RMSEP values were used for finding out the optimal number of principal components for modeling. Standard error of cross-validation (SECV), estimated mean square error of prediction (MSEP), and the correlation coefficient (r) of the predictive model were used for evaluating the PCR model.

5. RESULTS AND DISCUSSION

5.1 Segmented images and NIR reflectance spectra of wheat classes

The segmented 200×200 pixel images are shown for the 13, 16, and 19% m.c. wheat in Fig. 5.1. Reflectance intensities mostly depend on the NIR wavelengths which in turn lead to dissimilar visible textures in the images. Soft and hard wheat classes have starch granules with low and high molecular proteins, respectively (Famera et al., 2004). The reflectance intensities at different NIR wavelengths in hyperspectral imaging can be attributed to the chemical composition of samples. The Figs. 5.2, 5.3, and 5.4 illustrate the NIR reflectance spectra of 13, 16, and 19% m.c. wheat classes, respectively.

Sivakumar (2007) reported that protein and oil contents of composite samples of different western Canadian wheat classes were significantly different. The NIR reflectance intensities were different for wheat samples with varying degrees of hardness and protein. The NIR absorption of water can be seen at 960 and 1420 nm; protein at 1470-1500 nm; carbohydrate at 1200-1360 nm and 1610-1700 nm; kernel hardness at 960-1060 nm, 1330-1480 nm, and 1680 nm; and oil content at 1390 nm in wheat (Wang et al., 1999; Delwiche and Massie, 1996; Murray and Williams, 1987).

In a classification study of fungal-damaged soybean, the NIR wavelength of 1330 nm was identified as one of the significant wavelengths and related to fiber and starch contents of soybean (Wang et al., 2004a). The NIR absorptions of wheat samples produced excitations for carbon-hydrogen (CH) and oxygen-hydrogen (OH) combination bonds; CH, OH, and nitrogen-hydrogen (NH) 1st overtones, 2nd overtones, and CH 3rd overtones in samples (Osborne, 2006).

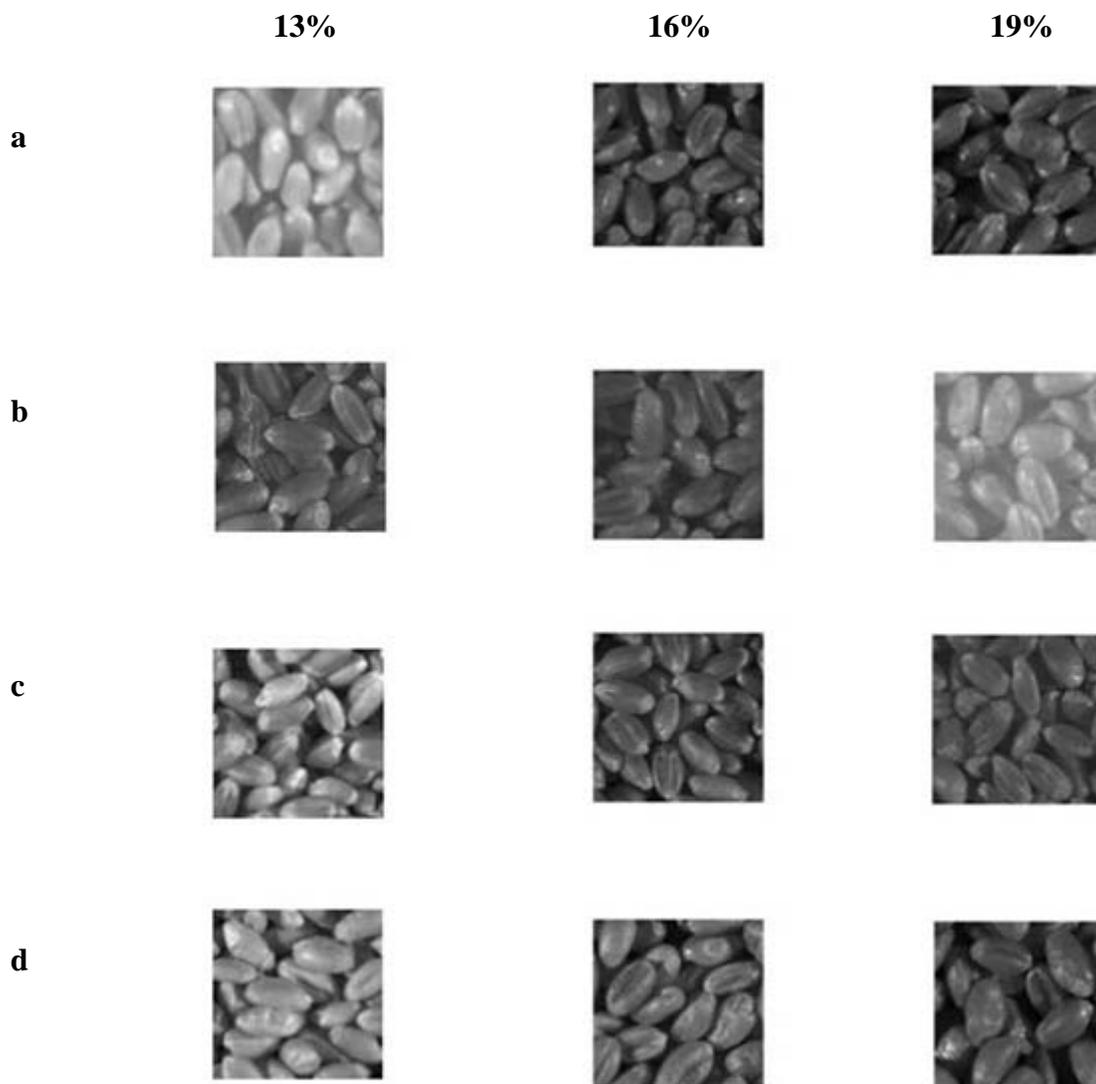


Fig. 5.1. Segmented NIR hyperspectral images acquired at 1330 nm for bulk samples of a) CWRs b) CPSR c) CWHWS and d) CWSWS wheat at 13% (left), 16% (center), and 19% (right) moisture content (wet basis).

The NIR features at the wavelengths of 1420 and 1440 nm were found significant for detecting insects in wheat (Dowell et al., 1999; Toews et al., 2007).

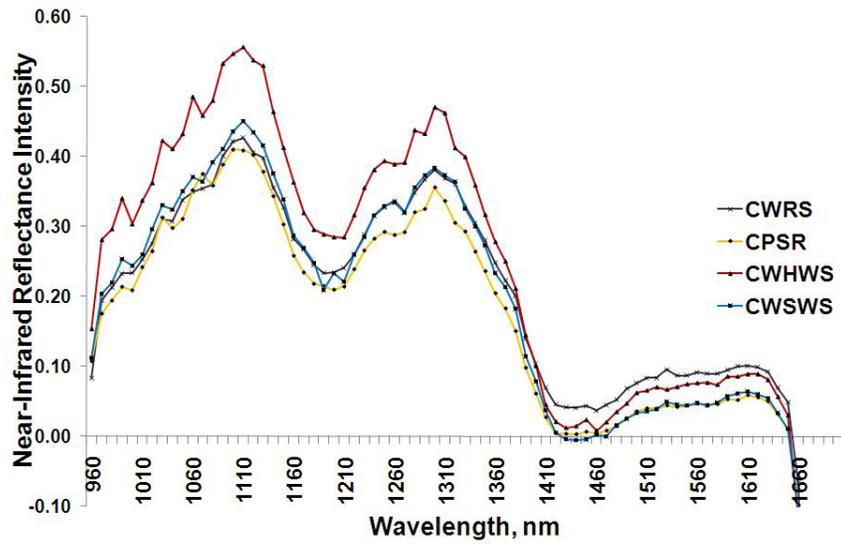


Fig. 5.2. Near-infrared (NIR) reflectance spectra of wheat classes that had 13% m.c. (wet basis).

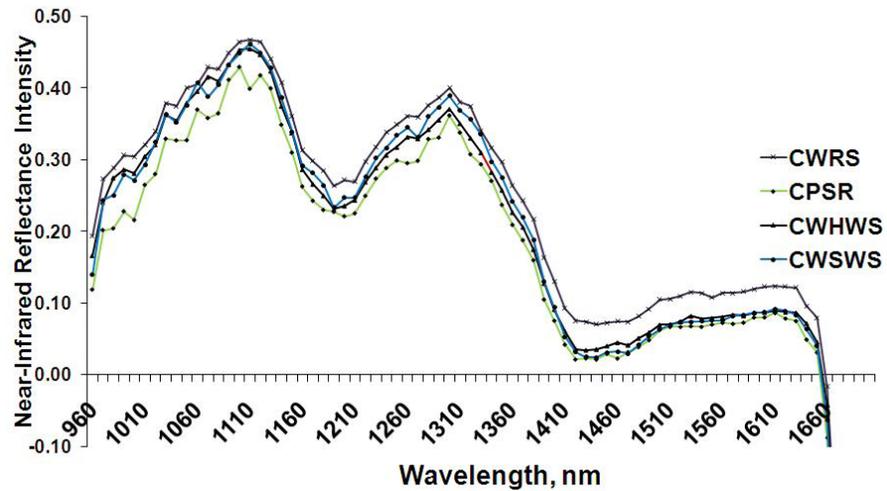


Fig. 5.3. Near-infrared (NIR) reflectance spectra of wheat classes that had 16% m.c. (wet basis).

The NIR reflectance intensities were different for wheat samples with varying degrees of hardness and protein. Maghirang et al. (2003) reported that the NIR features at 1135 and 1325 nm wavelengths were important for detecting insects in wheat.

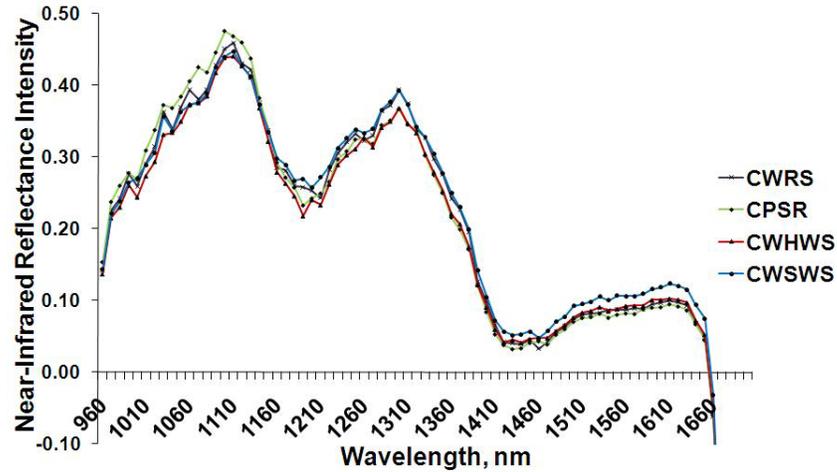


Fig. 5.4. Near-infrared (NIR) reflectance spectra of wheat classes that had 19% m.c. (wet basis).

5.2 Comparison of scores' images of wheat samples using the PCA

The PCA was performed for each image individually and the principal component scores' images were examined for determining relationships between image features and wheat classes. The first few principal components accounted for the majority of the variances of hyperspectral data (> 99%), which was explained in latter sub sections of this section. Figures 5.5, 5.6, and 5.7 show the loadings of first principal component plotted against wavelengths for wheat classes. The NIR wavelengths of 1260-1380 nm had the highest factor loadings out of the full spectral range of 960-1700 nm for all wheat samples of 13, 16, and 19% m.c. Baker et al. (1999) found that the NIR features at 1130-1200 and 1300-1400 nm wavelength regions were significant in detecting parasitized *Sitophilus oryzae* (L.) in wheat. In the wavelength regions of 1060-1160 nm and 1260-1380 nm, the NIR reflectance spectra of CWRS, CPSR, CWHWS, and CWSWS wheat showed variability.

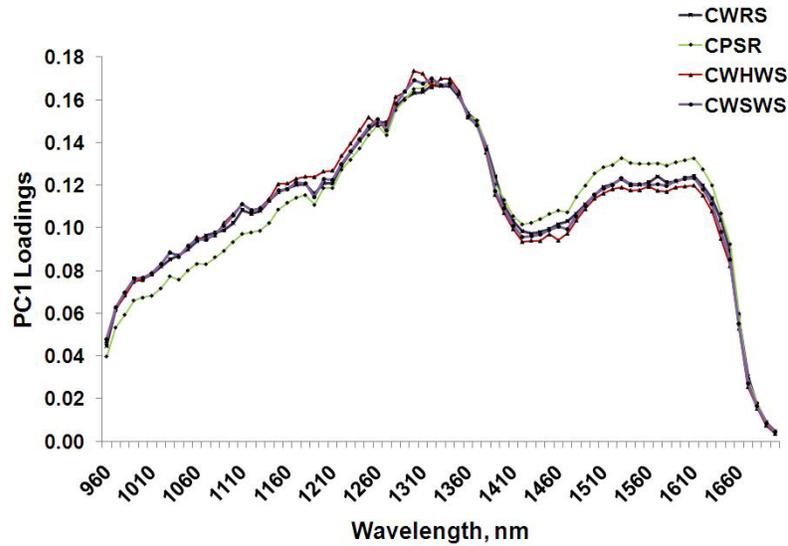


Fig. 5.5. The PC1 loadings plot of wheat classes that had 13% m.c. (wet basis)

The shapes of the reflectance spectra for all wheat classes were similar and comparable. In the highest factor loading region, peaks were also identified in the NIR reflectance spectra. Singh et al. (2007) found that wavelengths at 1284.2, 1315.8, and 1347.4 nm had the highest factor loadings for the first principal component (PC1) out of 20 wavelengths in the region of 1000-1600 nm in the fungal-detection study in wheat.

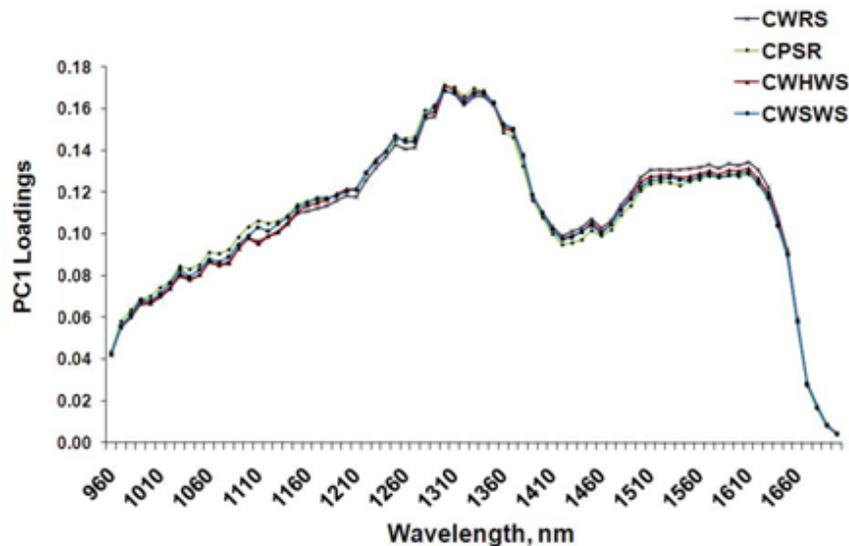


Fig. 5.6. The PC1 loadings plot of wheat classes that had 16% m.c. (wet basis)

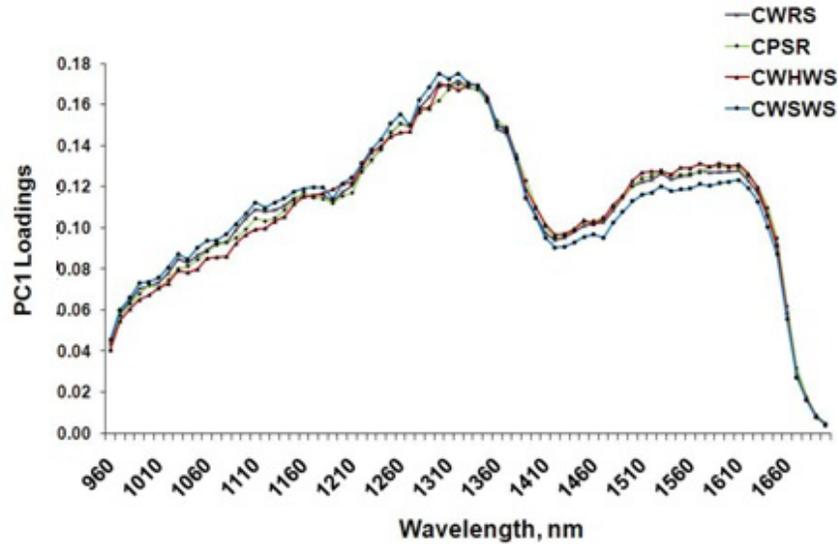


Fig. 5.7. The PC1 loadings plot of wheat classes that had 19% m.c. (wet basis)

The first three principal component scores images of wheat classes having 13, 16, and 19% m.c. are shown in Figs. 5.8, 5.9, and 5.10, respectively. The principal component 1 (PC1) and principal component 2 (PC2) scores images were similar but had an opposite display of the germs, the endosperms, and the kernel gaps in segmented images for all wheat classes. The germ portions of wheat kernels had higher PC1 scores than the rest of the kernel body images. Also, PC2 scores were lower for germ portions than the rest of kernel body for all images. The PC3 scores of all images of wheat classes accounted for a minority of the variance existing in the data. The PC3 scores images could only detect the edges of the kernels in a bulk wheat sample. The PC1 scores explained maximum variations of the data and PC2 got the maximum of the rest of the variations. As the data in the normal plane were transferred to the principal components plane in the PCA, discriminant classifiers such as LDA and QDA could be used for classification of wheat using the PCA scores as input.

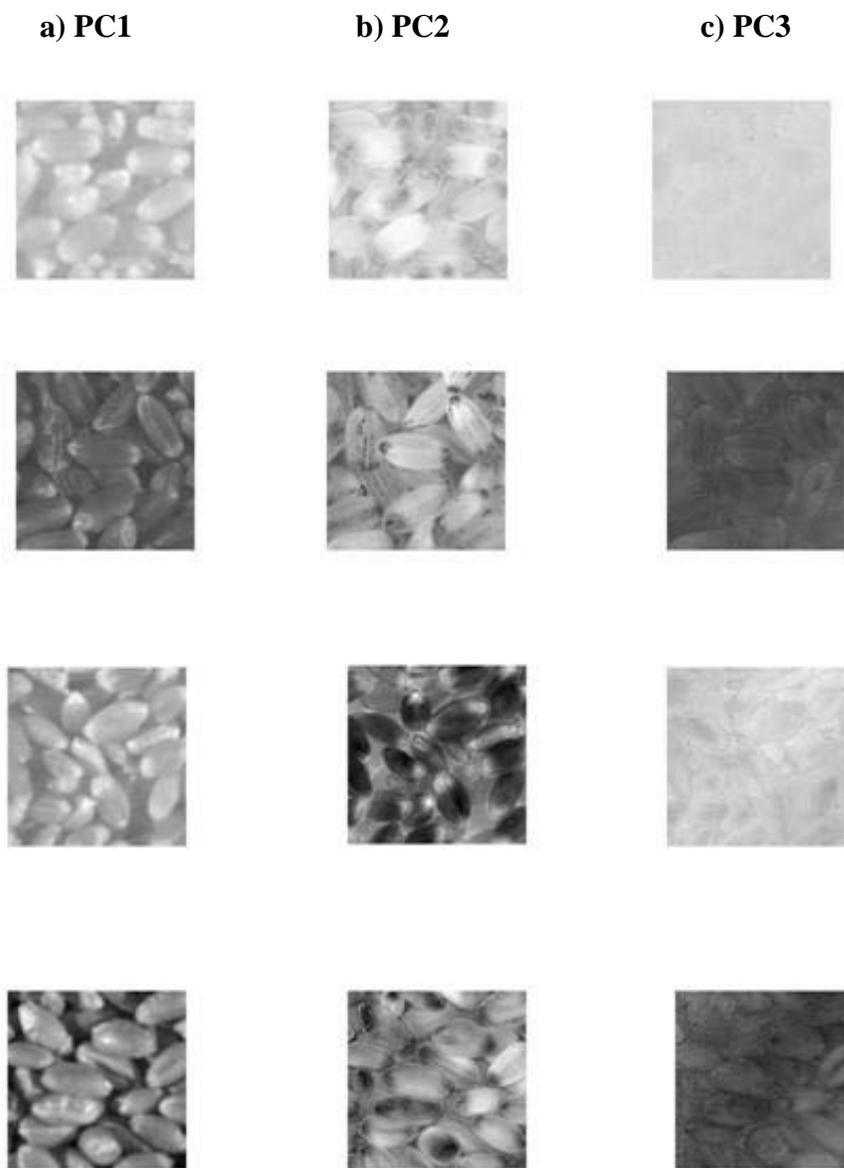


Fig. 5.8. The PC1, PC2, and PC3 scores images of wheat samples at a moisture content of 13%.

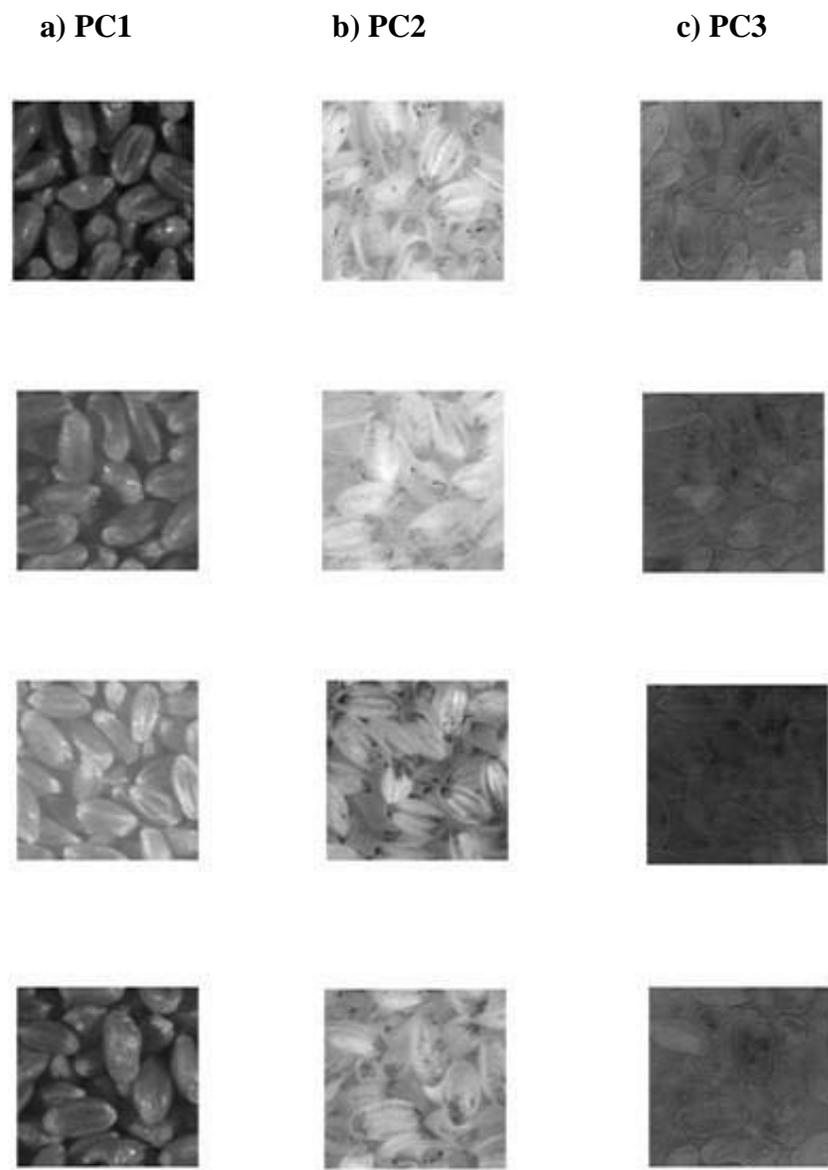


Fig. 5.9. The PC1, PC2, and PC3 scores images of wheat samples at a moisture content of 16%.

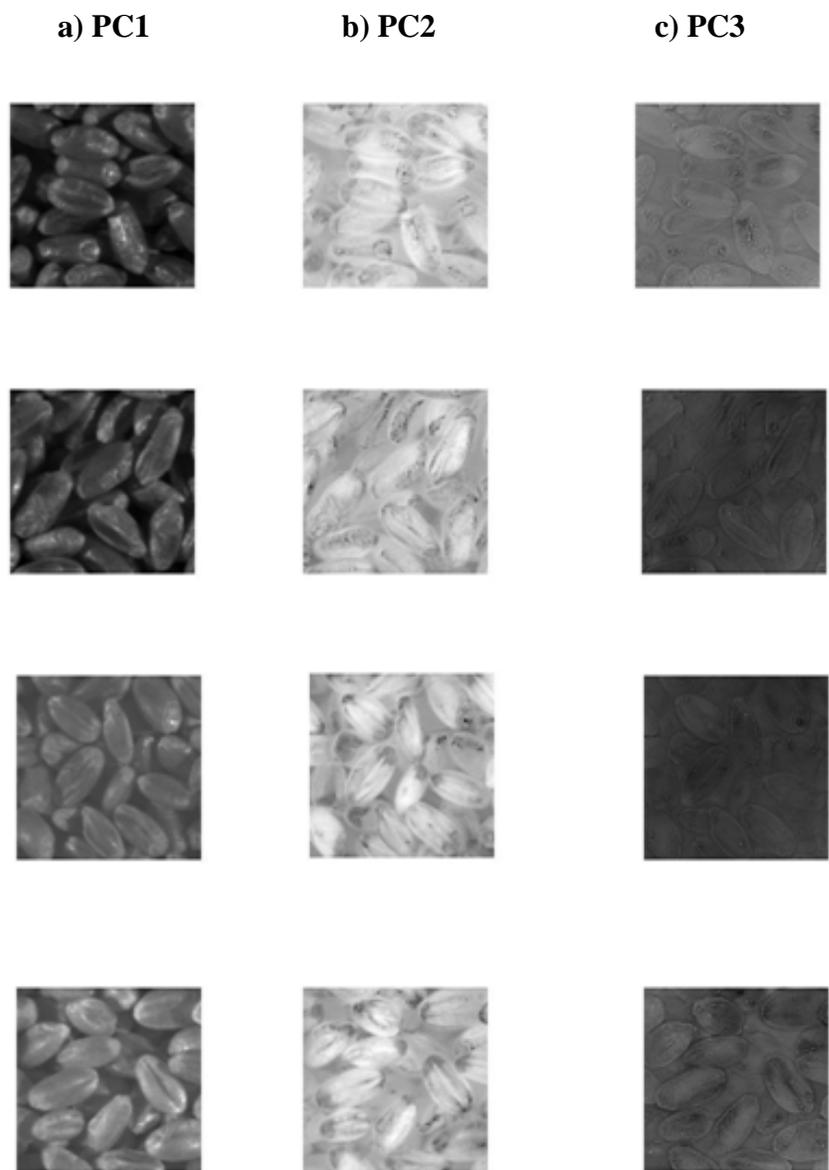


Fig. 5.10. The PC1, PC2, and PC3 scores images of wheat samples at a moisture content of 19%.

5.3 Discriminant classifiers

The LDA and QDA were used for performing discriminant analyses in this study. The LDA uses the pooled covariance and the QDA considers the individual covariance for each class for classification purposes. Equal dispersion of reflectance intensities in all wheat classes might lead to an improved performance by the LDA.

5.3.1 Overall identification of wheat classes

Wheat classes and moisture levels were identified independent of each other. At each moisture level, i.e., 13, 16, and 19%, 600 images per wheat class (containing 60 images at each growing location for each crop year) were used for developing the LDA and QDA classifiers. In total, 7200 samples were used. In overall identification of wheat classes, sample variations among wheat classes and variations within growing locations, crop years, and moisture contents were included. To verify the efficiency of classification models, a leave-one-out cross validation method was used (Mahesh et al., 2010a and 2010b). In a similar study, Mahesh et al. (2011) also identified wheat classes and their moisture levels for composite CWRS, CWES, CWRW, CWSWS, and CWHWS wheat samples, which were conditioned to 12, 14, 16, 18, and 20% moisture contents, by scanning bulk samples using the NIR hyperspectral imaging.

Average classification accuracies of 80.6 and 76.3% were obtained for the LDA and the QDA, respectively (Fig 5.11). All CWSWS wheat samples were classified with the highest classification accuracies of 88.1 and 95.2% for the LDA and QDA, respectively. Classification accuracies of 65.6% were obtained for CWHWS wheat which was the lowest in the LDA. The results of the QDA were superior to those of the LDA for all wheat classes except for CWRS. Similarly, Mahesh et al. (2011) reported

classification accuracies of 61-97% and 82-99% for the LDA and QDA, respectively, for identifying wheat classes independent of moisture contents.

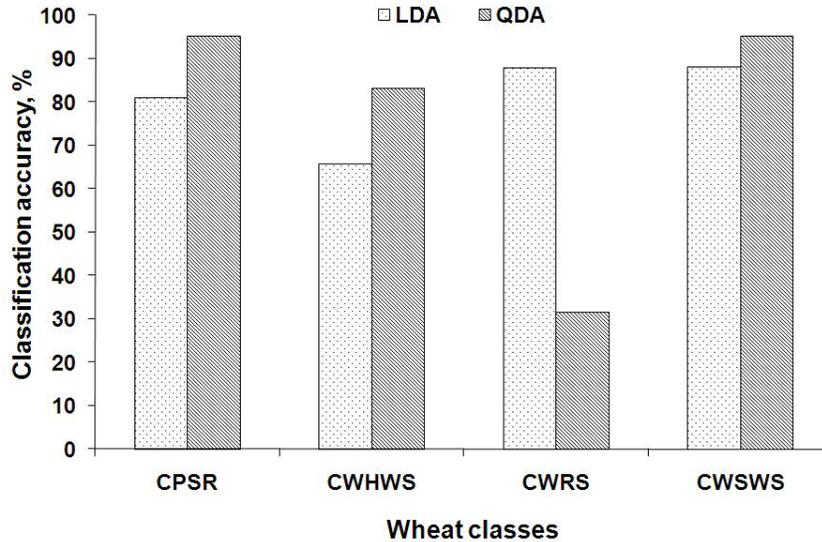


Fig. 5.11. Overall classification accuracies of identifying western Canadian wheat classes using the LDA and QDA classifiers.

5.3.2 Overall identification of moisture levels

Sample variations among moisture levels and variations within growing locations, crop years, and wheat classes were included. An input data set of 7200 images in total, that were developed by including scanned images of 600 samples for each class at each moisture level, were used. Average classification accuracies were 95.2 and 90.8% for the LDA and QDA, respectively, for overall identification of moisture levels (Fig. 5.12). Classification accuracies were 99.5 and 91.7%, which were the highest and the lowest, for the LDA for identifying wheat with the 13 and 19% m.c., respectively. For the QDA, using overall sample set, the classification accuracies were 95 and 83.3%, which were the

highest and the lowest, respectively, for identifying the 16 and 13% m.c. wheat, respectively.

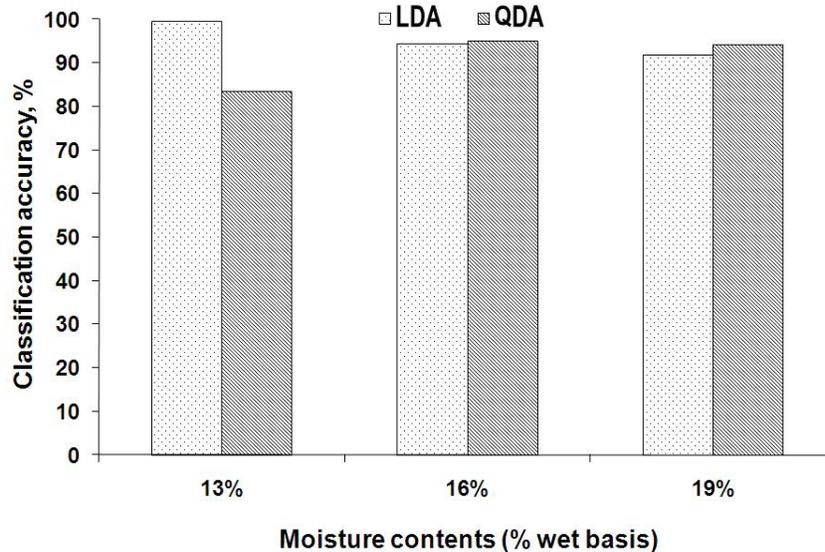


Fig. 5.12. Overall classification accuracies of identifying moisture levels of wheat using the LDA and QDA classifiers.

The pair-wise classification could classify with average classification accuracies of 98.1 and 95.2% using the LDA and QDA, respectively (Figs. 5.13 and 5.14). The thirteen vs. nineteen pair followed by the thirteen vs. sixteen pair were classified the best using the LDA. For the LDA, classification accuracies of 100% were obtained, where difference between the moisture levels were high (about 6%), for identifying low (13%) and high (19%) moisture wheat. Average classification accuracy was 97.1% for the LDA of the wheat samples that had a low difference (about 3%) in their moisture levels. For the QDA, using overall sample set for classification, the classification accuracies were 98.8 and 100%, which were the highest, for identifying the 13% m.c. and the 19% m.c. wheat, respectively. The wheat samples, that had about 3% moisture difference, were discriminated with average classification accuracies of 93.2%.

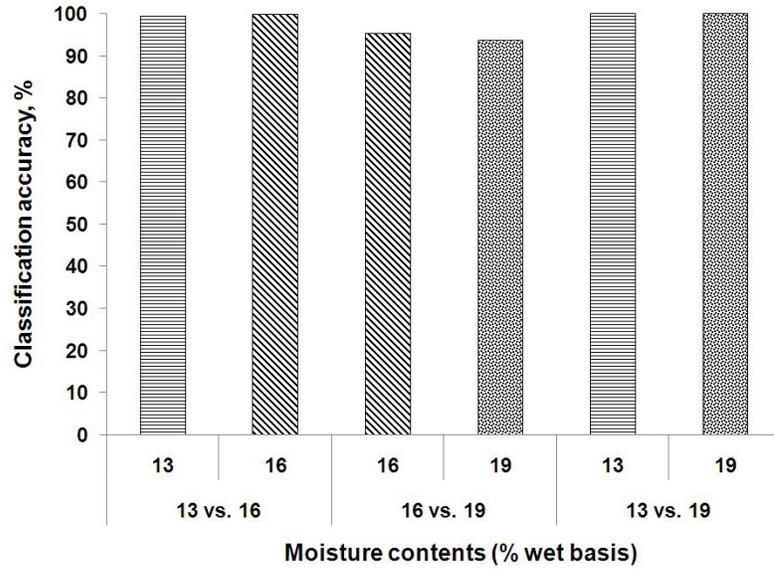


Fig. 5.13. Pair-wise classification accuracies of identifying moisture levels of wheat using the LDA classifier.

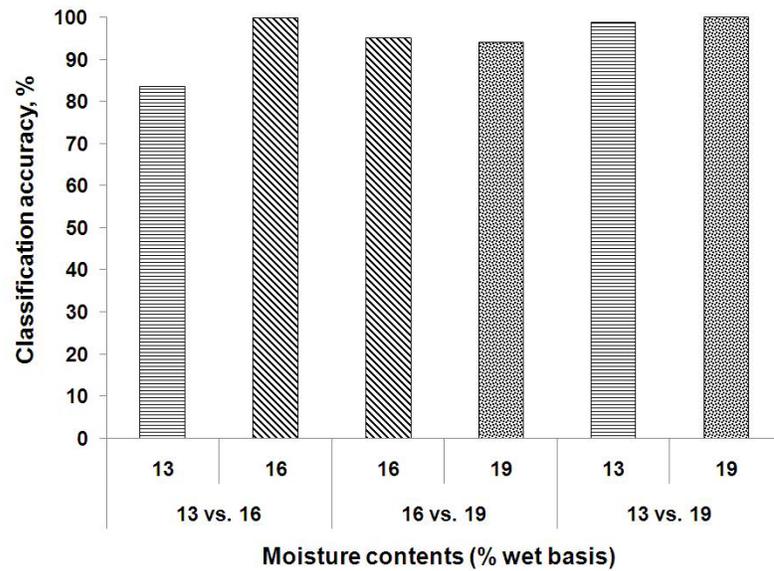


Fig. 5.14. Pair-wise classification accuracies of identifying moisture levels of wheat using the QDA classifier.

Procedure STEPDISC was used to identify the relative importance of NIR reflectance features towards classification. Table 5.1 shows the input features (related wavelengths are shown) that were ranked in descending order based on contribution to

the discriminant models. Top ten features included the NIR reflectance intensities in the wavelength regions of 1220-1230, 1320-1400, and 1630-1660 nm for overall identification of moisture contents. Rankings of the pair-wise discrimination included the NIR reflectance features from the wavelength regions of 1140-1250, 1330-1420 nm, and 1630-1660 nm. Wavelength regions identified for overall classification were within the regions recognized by the pair-wise discrimination of moisture contents. In a related study, Mahesh et al. (2011) identified that the NIR wavelengths of 1260-1360 nm were important for discriminating wheat at different moisture levels that were ranged from 12 to 20% with a 2% increase for each level.

Table 5.1. The top 10 features, in descending order, based on their contribution towards classification accuracy for wheat moisture identification while using non-parametric classifiers with all features as inputs.

Rank	Top 10 wavelengths for pair-wise discrimination (nm)			
	Overall moisture (nm)	13% vs. 16%	16% vs. 19%	13% vs. 19%
1	1660	1570	1330	1660
2	1320	1400	1250	960
3	1220	1640	1140	1400
4	1330	1190	1220	1630
5	990	1330	960	1190
6	1070	1410	1200	1650
7	1230	1630	1070	1410
8	970	1220	1340	1340
9	1400	1420	1210	1570
10	1630	1650	1150	1200

5.3.3 Identification of wheat classes at 13% m.c. independent of growing locations and crop years

An attempt was made to classify wheat classes independent of growing locations and crop years at a uniform moisture level of 13%. As the chemical compositions of

wheat classes were different, it seemed possible to classify with higher accuracies wheat classes at the same moisture content. The input data set of 600 images per class (containing 60 images at each growing location for each crop year) was used for non-parametric classifiers. Sample variations among wheat classes and variations within growing locations and crop years were included. To verify the efficiency of classifiers, a leave-one-out cross validation method was used. In this type of validation, discriminant models were developed by excluding one observation for each cycle randomly from the data set. Later, these models were validated using the excluded data set. The validation accuracy for each cycle was recorded. This cycle was performed 'n' times where 'n' was equal to the number of input samples.

For the non-parametric classifiers, using the NIR reflectance features of wheat that had 13% m.c., the average classification accuracies were 95.4 and 92.3% for the LDA and QDA, respectively. In the LDA model, classification accuracies of 86.5-99% were obtained in which CWHWS wheat had the lowest accuracy of 86.5%. This could be because of the fact that CWHWS had overlapping intrinsic qualities (e.g., protein, hardness, starch, oil content, and moisture) with other wheat classes. Average posterior probabilities of wheat classes classified into a correct class ranged from 0.97 to 0.99. Classification accuracies of 84.5-100% were achieved in the QDA model. In this model, average posterior probabilities were 0.98-1.00. The classification accuracies of wheat classes independent of growing locations and crop years using the LDA and QDA models are given in Fig. 5.15. Wheat classes were identified with accuracies of about 100% using the differences in fluorescence properties of different wheat parts, which included pericarp, aleurone layer, and endosperm, and protein and hardness (Irving et al., 1989).

Majumdar et al. (2000) reported that morphological resemblance might lead to misclassification using digital images of wheat.

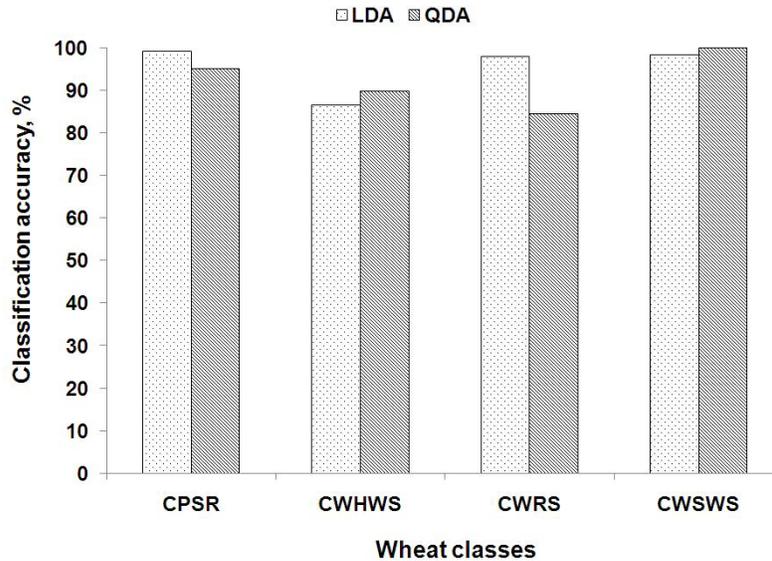


Fig. 5.15. Classification accuracies of identifying wheat classes that had 13% m.c. using the LDA and QDA classifiers.

5.3.4 Identification of wheat classes at 16% m.c. independent of growing locations and crop years

Four sets of six hundred images per class (60 images at each growing location for each crop year in each class) were used for the statistical classification to identify wheat that had 16% m.c. Sample variations among wheat classes and variations within growing locations and crop years were included for modeling. A leave-one-out cross validation method was used for validating the performance of classifiers. Classification accuracies of 91.8-95.8% and 94-98.3% were obtained for the LDA and QDA, respectively (Fig. 5.16). The mean classification accuracies with the LDA and QDA were 93.1 and 96.4%, respectively. Canada Western Red Spring wheat followed by CWHWS wheat were

classified the best using the LDA. For the QDA, higher classification accuracies of 98.3 and 96.8% were obtained for CWSWS and CWRS, respectively. The lowest classification accuracies of 91.8 and 94% were obtained, for the LDA and QDA, for CPSR and CWHWS wheat, respectively. Singh et al. (2007) investigated fungal detection in wheat using the NIR hyperspectral imaging and reported 97.8% of average classification accuracy for identifying infected kernels using the LDA.

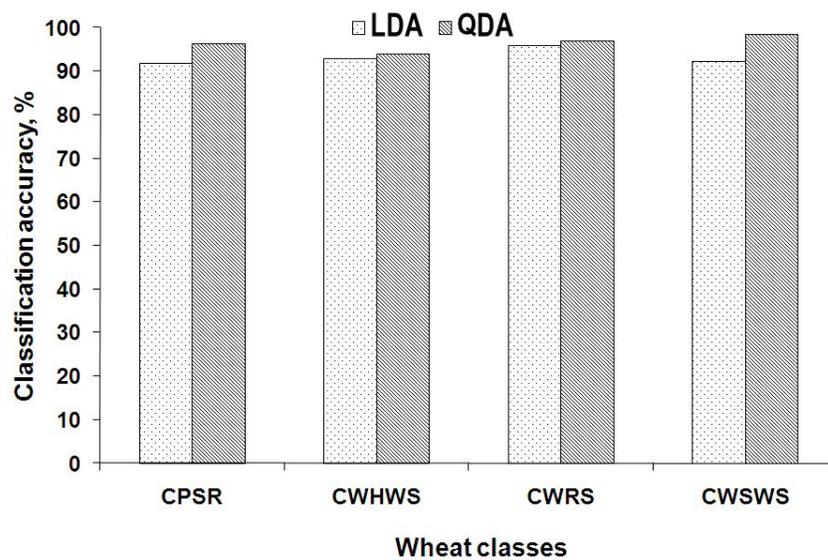


Fig. 5.16. Classification accuracies of identifying wheat classes that had 16% m.c. using the LDA and QDA classifiers.

5.3.5 Identification of wheat classes at 19% m.c. independent of growing locations and crop years

Sixty hyperspectral images acquired from each location- and crop year-specific bulk samples of wheat were used for developing the input data set. The full set, 2400 images in total, was used to develop non-parametric statistical classifiers to identify wheat that had 19% m.c. A leave-one-out cross validation method was used to confirm the performance of classifiers. Classification accuracies were 88.5-98.6% and 95.5-98.3%

for the LDA and the QDA models, respectively, for identifying wheat classes at a moisture level of 19% (Fig. 5.17). Average classification accuracies were 93.9 and 96.5% for the LDA and QDA, respectively. The results of the QDA were superior to those of the LDA for all wheat classes except for CWRS. Canada Western Red Spring and CPSR wheat were discriminated with high classification accuracies of 98.6 and 98.3%, for the LDA and QDA, respectively. The classification accuracies of QDA were superior to those of LDA for all wheat classes except CWRS.

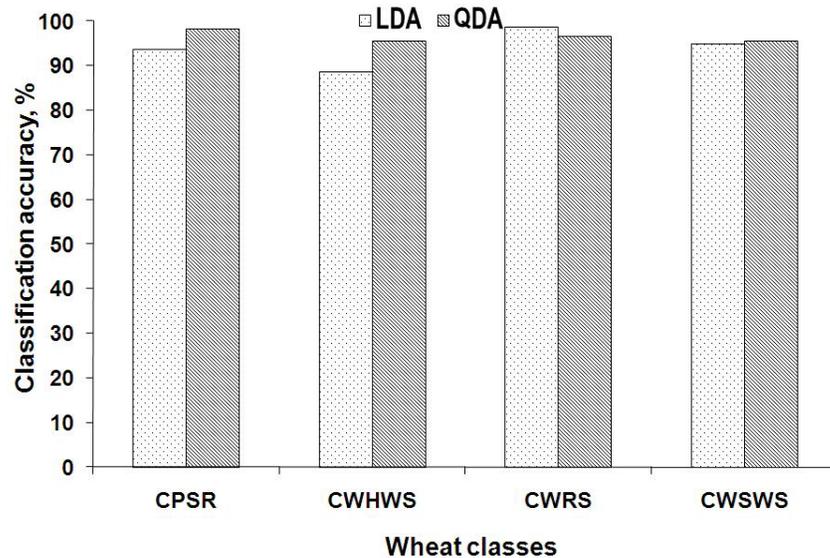


Fig. 5.17. Classification accuracies of identifying wheat classes that had 19% m.c. using the LDA and QDA classifiers.

Procedure STEPDISC was used to identify top 10 ranked features, which are presented in Table 5.2, based on their contribution towards classification for the non-parametric classifiers. Top ten features included the NIR features in the wavelength regions of 1080-1110 and 1300-1380 nm for overall identification of wheat classes. Rankings for the identification of wheat classes at each specific moisture level included the NIR reflectance features from the wavelength regions of 1110-1130, 1200-1260,

1290-1420, and 1660-1700 nm. Wavelength regions identified for overall class identification was mostly within the regions recognized by the classification of wheat at specific moisture levels.

Table 5.2. The top 10 features, in descending order, based on their contribution towards classification accuracy for wheat class identification while using non-parametric classifiers with all features as inputs.

Rank	Top 10 wavelengths identified using a stepwise discrimination method (nm)			
	Overall classes	Classes with 13% m.c.	Classes with 16% m.c.	Classes with 19% m.c.
1	1300	1030	1350	1700
2	1100	1130	1460	1210
3	1340	1260	1390	1010
4	960	1300	1700	1530
5	1490	1340	1130	1080
6	1210	1500	1080	1380
7	1380	1210	1290	1040
8	1660	1380	1210	1110
9	1080	1660	1200	1350
10	1110	1680	1420	1390

5.3.6 Pair-wise classification of wheat

Two-class LDA and QDA classifiers were tested for identifying wheat classes. Overall identification and identification of wheat classes at each moisture level were performed for the following combinations: CWRS versus CPSR, CWRS versus CWHWS, CWRS versus CWSWS, CPSR versus CWHWS, CPSR versus CWSWS, and CWHWS versus CWSWS.

5.3.6.1 Identification of wheat classes using the full data set

Pair-wise discriminant classifiers such as LDA and QDA with an independent sample set of 1800 bulk samples (600 samples in each moisture level \times 3 moisture levels)

in each wheat class were developed. A leave-one-out cross validation was used for validating model performances. Classification accuracies were 88.6-97.6% and 48.1-100% for the LDA and QDA, respectively, for identifying wheat classes (Figs. 5.18 and 5.19). Average classification accuracies of 93.1 and 83.9% were obtained for the LDA and QDA, respectively. In the LDA, the classification accuracies were high for the CWRS versus CWSWS followed by CPSR versus CWSWS. Both red wheat (CWRS vs. CPSR) and white wheat (CWHWS vs. CWSWS) classes were identified with average classification accuracies of 93.6 and 91.6%, respectively. The CPSR versus CWSWS and CWHWS versus CWSWS were superior pairs in the QDA. Wheat classes, which were compared with other wheat classes except CWRS, were discriminated with high accuracies of above 90% using the QDA. Overall, the LDA gave better results than the QDA.

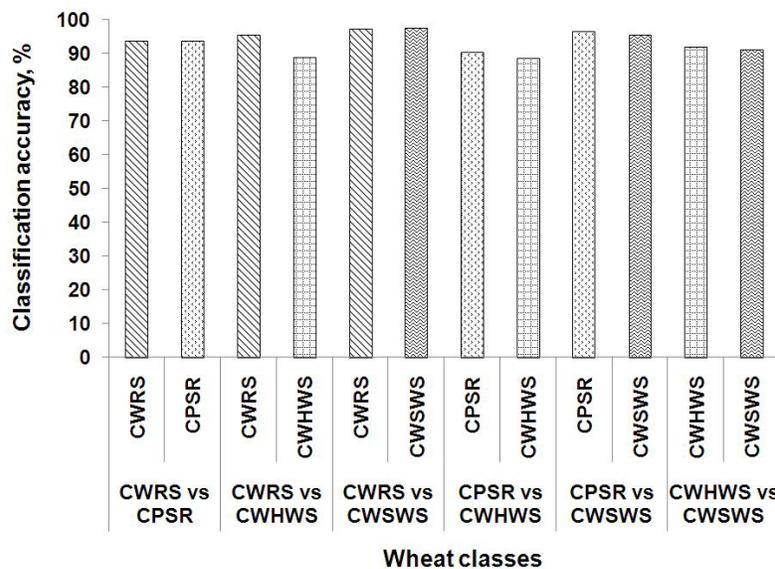


Fig. 5.18. Pair-wise classification accuracies of identifying wheat classes using the LDA classifier.

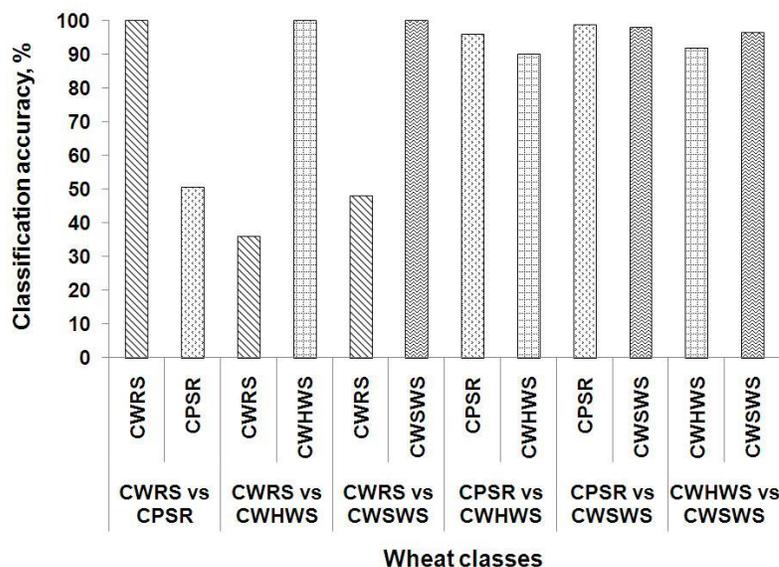


Fig. 5.19. Pair-wise classification accuracies of identifying wheat classes using the QDA classifier.

Relative importance of the NIR features towards classification was found using the procedure STEPDISC of SAS. Table 5.3 shows top 10 features that were ranked based on their contribution towards classification for non-parametric classifiers. Top ten features included the NIR features in the wavelength regions of 1060-1180, 1200-1250, 1300-1470, and 1610-1700 nm for pair-wise identification of wheat classes. The NIR reflectance features in the wavelength region of 1100-1300 nm relate to C-H 1st and 2nd overtones, and C-H combination band. In this wavelength region, starch molecules of samples absorb the NIR radiation.

Table 5.3. The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for overall identification of wheat classes while using non-parametric classifiers with all features as inputs.

Rank	Top 10 wavelengths for pair-wise discrimination (nm)					
	CWRS vs. CPSR	CWRS vs. CWHWS	CWRS vs. CWSWS	CPSR vs. CWHWS	CPSR vs. CWSWS	CWHWS vs. CWSWS
1	1690	1300	1300	1300	1340	1120
2	690	1330	1500	1110	1060	1340
3	1530	1350	1640	1700	1460	1140
4	1030	1660	1070	960	1210	1060
5	1080	1420	1210	1000	1000	1170
6	1110	1640	1470	1210	1100	1450
7	1670	1640	1650	1380	1450	1380
8	1660	1230	1110	1170	1250	1310
9	1060	980	1530	1250	1160	1200
10	1610	1160	1420	1690	1180	1070

5.3.6.2 Identification of wheat classes at 13% m.c.

Figures 5.20 and 5.21 show classification accuracies of two-class identification of wheat, which had 13% m.c., using non-parametric statistical LDA and QDA classifiers, respectively. Compared to the accuracies of identifying wheat using the full data set, classification accuracies were improved significantly for wheat classes that had uniform moisture levels of 13, 16, and 19%. This can be attributed to the elimination of moisture variations from the input data set. Average classification accuracies were 98.8 and 97.4% for the pair-wise LDA and QDA classifiers, respectively.

Classification accuracies for the QDA were improved considerably. The CWRS wheat samples were identified with perfect accuracies of 100% when they were discriminated against CPSR and CWHWS wheat classes. Overall, the accuracies of the pair-wise classification using the non-parametric discriminant classifiers were very high

(greater than 97%). In a similar classification study, the highest classification accuracies of 95.3-99.3% were reported for classifying healthy and midge-damaged wheat kernels using the NIR reflectance features combined with the top 10 colour image features (Singh et al., 2010).

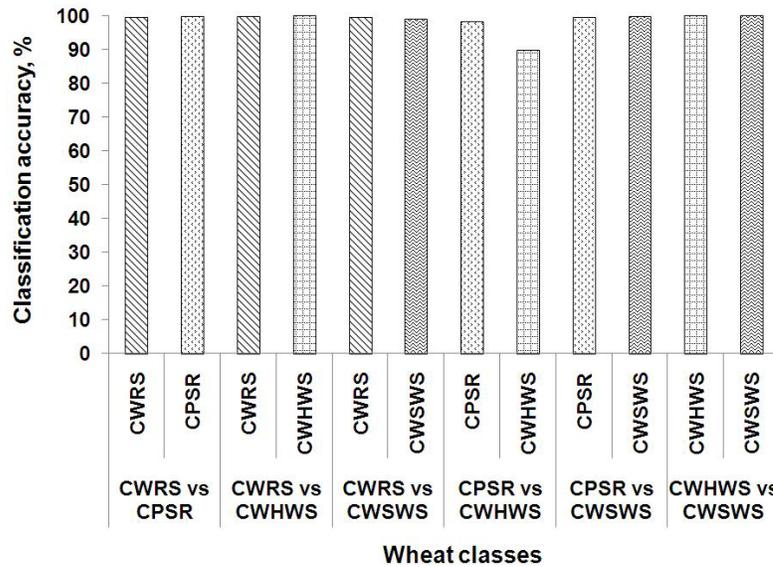


Fig. 5.20. Pair-wise classification accuracies of identifying wheat classes that had 13% m.c. using the LDA classifier.

The top 10 ranked features were identified using the procedure STEPDISC for the non-parametric classification (Table 5.4). Relative importance of input features towards classification was used to rank the features in the descending order.

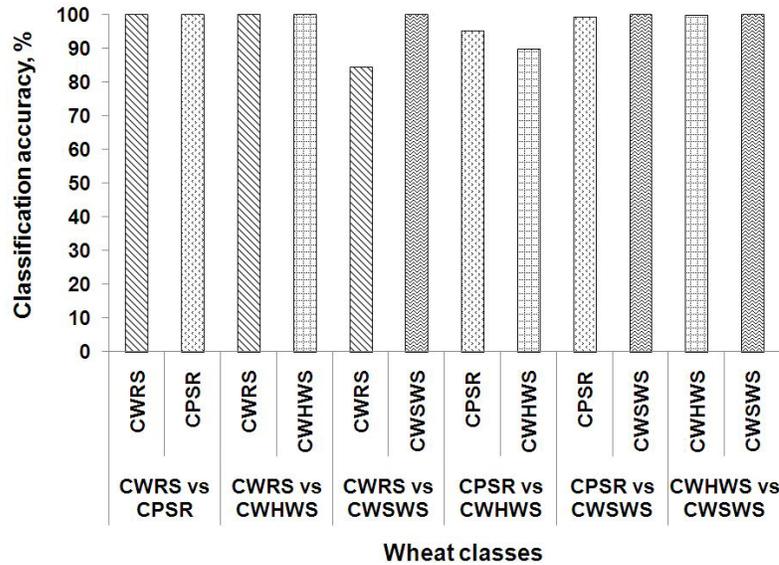


Fig. 5. 21. Pair-wise classification accuracies of identifying wheat classes that had 13% m.c. using the QDA classifier.

Top ten features included the NIR features in the wavelength regions of 960-1070, 1120-1210, 1290-1390, and 1570-1700 nm.

Table 5.4. The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for class identification for 13% m.c. wheat while using non-parametric classifiers with all features as inputs.

Rank	Top 10 wavelengths for pair-wise discrimination (nm)					
	CWRS vs. CPSR	CWRS vs. CWHWS	CWRS vs. CWSWS	CPSR vs. CWHWS	CPSR vs. CWSWS	CWHWS vs. CWSWS
1	980	1030	1320	1310	1700	1120
2	1350	1070	1270	1340	1340	1170
3	1200	1390	960	1200	1500	1500
4	1300	960	1470	1700	1300	1400
5	1330	1690	1070	1680	1030	1060
6	1290	1670	1180	1070	1210	1190
7	1690	970	1660	1640	1320	1700
8	1360	1450	1570	1170	1640	1010
9	1450	1600	1200	1160	1240	1450
10	1600	1060	1630	1500	1570	1390

5.3.6.3 Identification of wheat classes at 16% m.c.

Six hundred samples from each wheat class, which had 16% m.c., were used as input for the pair-wise discrimination using the LDA and QDA. Average classification accuracies were 98 and 98.5% for the LDA and QDA, respectively. Classification accuracies attained during the pair-wise discrimination process are shown in Figs. 5.22 and 5.23. The elimination of moisture variations from the input data set improved classification accuracies of pair-wise comparisons of wheat. The LDA showed consistent and very high classification accuracies that ranged from 94.5-99.8%.

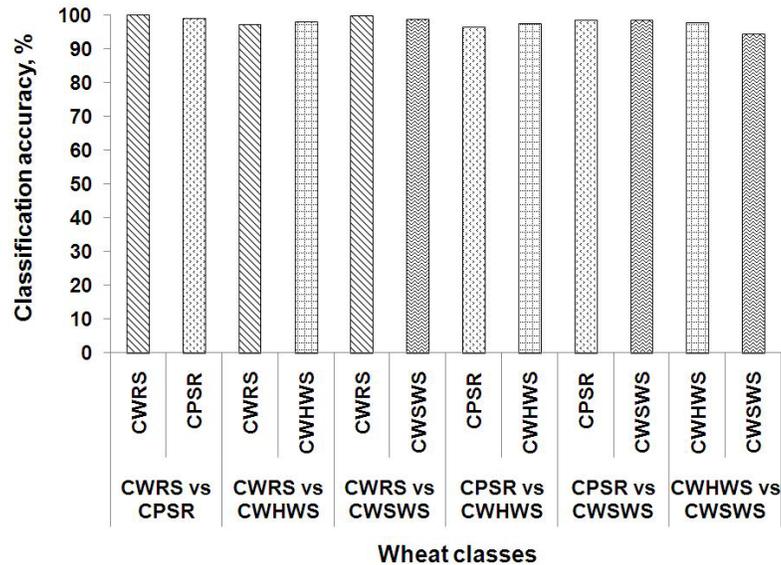


Fig. 5.22. Pair-wise classification accuracies of identifying wheat classes that had 16% m.c. using the LDA classifier.

Figure 5.23 indicates that a reduced set of features, which had no moisture effects, improved classification accuracies for the QDA. Near-perfect classification accuracies of greater than 96.5% were obtained for the QDA for all two-class combinations of wheat. The significant amount of variations in hardness and protein content of wheat leads to high identification accuracies for non-parametric classifiers.

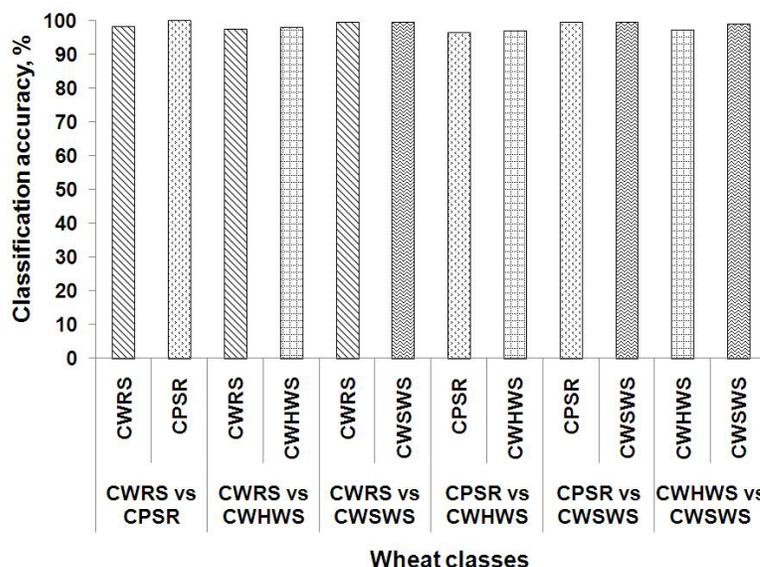


Fig. 5.23. Pair-wise classification accuracies of identifying wheat having 16% m.c. using the QDA classifier.

Procedure STEPDISC of SAS was used for finding out the relative importance of input features towards classification. Table 5.5 shows the top 10 rankings of features which were categorized based on their contributions to classification. Among the top 10 features, the NIR reflectance intensities of wavelengths in the regions of 960-1080, 1160-1210, 1270-1470, and 1640-1700 nm were present. Maghirang and Dowell (2003) reported that the NIR wavelengths of 1100, 1200, 1380, 1450, and 1670 nm were contributed mainly in predicting hardness of wheat samples. The NIR absorption of water, protein, and carbohydrate can be seen at 960 and 1420; 1470-1500; 1200-1360 and 1610-1700 nm, respectively. Wavelengths of 960-1060, 1330-1480, and 1680 nm relate to the NIR absorption of hardness. Variations in oil content of wheat samples can be seen at differences in the NIR absorption at 1390 nm. In another study, The NIR absorption of wheat in the wavelength of 1480 nm is attributed to protein content of samples (Delwiche and Massie, 1996).

Table 5.5. The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for class identification for 16% m.c. wheat while using non-parametric classifiers with all features as inputs.

Rank	Top 10 wavelengths for pair-wise discrimination (nm)					
	CWRS vs. CPSR	CWRS vs. CWHWS	CWRS vs. CWSWS	CPSR vs. CWHWS	CPSR vs. CWSWS	CWHWS vs. CWSWS
1	1540	1530	1700	1700	1330	1310
2	1130	1640	1210	960	1470	1470
3	1080	1420	1460	1210	1380	1400
4	1040	1320	1390	1200	1210	1460
5	1570	1210	1280	1330	1130	1570
6	1350	1380	1430	1040	1000	1200
7	1680	960	1400	1080	1590	1660
8	1660	1160	1270	1070	970	960
9	1160	1540	1540	1160	980	970
10	1290	1650	1640	1000	1700	1170

5.3.6.4 Identification of wheat classes at 19% m.c.

Pair-wise discriminant classifiers were developed for identifying wheat classes that had uniform moisture levels of 19%. An independent sample set of 600 reflectance spectra for each wheat class, in total 1200 samples per treatment, was used. The performance of classifiers was validated using a leave-one-out cross validation method. Classification accuracies of the LDA and QDA are shown in Figs. 5.24 and 5.25, respectively. Average classification accuracies of 97.6 and 98.6 were obtained for the LDA and QDA, respectively. Wheat classes were identified with accuracies of 92.3-99.5 and 97-100%, for the LDA and QDA, respectively. Canada Western Red Spring wheat was identified with near-perfect accuracies of 99.3 and 99.5%, when compared to CPSR and CWSWS, respectively, using the LDA.

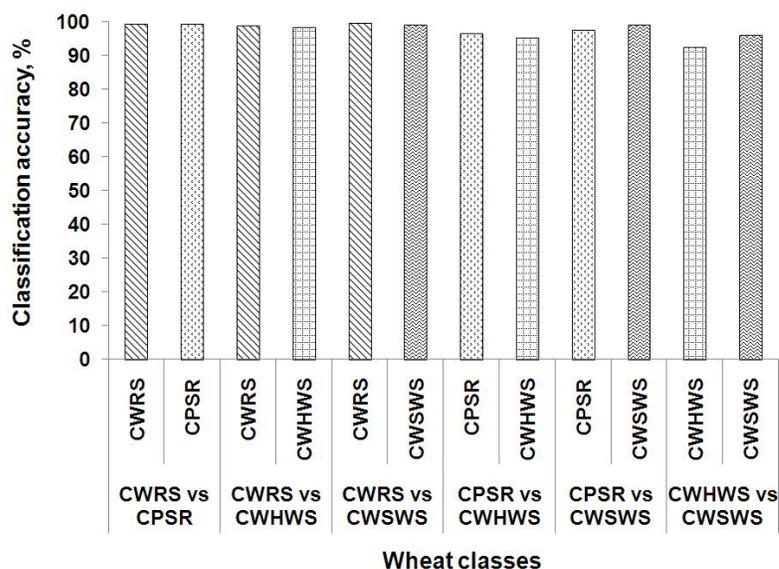


Fig. 5.24. Pair-wise classification accuracies of identifying wheat classes that had 19% m.c. using the LDA classifier.

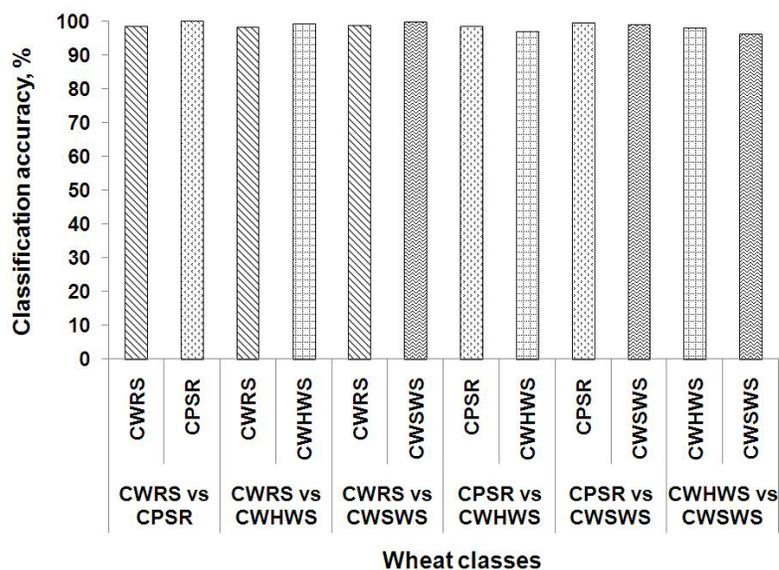


Fig. 5.25. Pair-wise classification accuracies of identifying wheat classes that had 19% m.c. using the QDA classifier.

Using the QDA, classification accuracies, which were above 97%, were more superior to those of the LDA. It is apparent that variations in intrinsic qualities such as hardness, protein, and starch content of wheat lead to very high pair-wise identification

accuracies. The procedure STEPDISC was used for finding out the relative importance of input features for classifying pair-wise the wheat classes having 19% m.c. Table 5.6 shows the top 10 features that were ranked based on their contribution to classification. The NIR reflectance intensities of wavelengths in the regions of 960-1210, 1300-1380, 1490-1570, and 1650-1700 nm were identified in the rankings.

Table 5.6. The top 10 features, in descending order, based on their contribution towards pair-wise classification accuracy for class identification for 19% m.c. wheat while using non-parametric classifiers with all features as inputs.

Rank	Top 10 wavelengths for pair-wise discrimination (nm)					
	CWRS vs. CPSR	CWRS vs. CWHWS	CWRS vs. CWSWS	CPSR vs. CWHWS	CPSR vs. CWSWS	CWHWS vs. CWSWS
1	1700	1700	1300	1700	1340	1300
2	960	960	1500	1110	1100	1680
3	1540	1540	1210	1250	1570	1110
4	1570	1210	1700	1360	1690	1040
5	980	1360	1670	1350	1250	1140
6	1490	1670	1080	1380	1360	1100
7	1650	1050	1160	1210	1350	1050
8	1010	1160	1430	1520	1330	1700
9	1080	1690	1110	1570	1140	1260
10	1040	1040	1050	1170	1040	1180

5.3.7 Growing locations

5.3.7.1 Identification of growing locations of wheat independent of crop years

A study was conducted to identify wheat classes at 13, 16, and 19% m.c. based on growing locations. It was hypothesized that sample variations of wheat classes among growing locations were significantly different. Variations among crop years within a wheat class were included in non-parametric classifiers. The data set of 600 image cubes

in wheat class, 2400 image cubes in total, was included for analyses. A leave-one-out cross validation method was used and validation accuracies were reported.

Average classification accuracies of 91.4-99.6% and 86.8-99.6% were obtained for the LDA and QDA, respectively. The LDA performed better than the QDA for identifying growing locations of wheat that had 13% m.c. (Table 5.7). The highest classification accuracies (100%) were reached when classifying CWHWS from all growing locations except from 'Limerick', CWRS from 'Melfort', and CWSWS from 'Kenton' using the LDA. The classification accuracy of 73.3% was the lowest using the LDA when identifying CWSWS from 'Jansen'. The CWHWS wheat from locations 'Limerick' and 'Mather' and CWRS wheat from 'Tisdale' had the highest classification accuracy of 100% using the QDA. Wheat (CWSWS) from 'Wilkie' had 73.3% of classification accuracy that was the lowest in the QDA. The average classification accuracies ranged from 91-99% and 86-99% using the LDA and QDA, respectively, when identifying growing locations of wheat independent of crop years (Table 5.7).

Table 5.7. Classification accuracies of identifying growing locations near the listed towns (independent of crop years) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) for 13% m.c. wheat.

Location	CPSR		Location	CWHWS		Location	CWRS		Location	CWSWS	
	LDA (%)	QDA (%)		LDA (%)	QDA (%)		LDA (%)	QDA (%)		LDA (%)	QDA (%)
Edmonton	88.3	90.8	Churchbridge	100	98.3	Corning	94.1	98.3	Corning	98.3	99.1
Rosemary	84.1	87.5	Kenton	100	99.1	Dauphin	97.5	97.5	Jansen	73.3	82.5
Viking	90.8	89.1	Limerick	98.3	100	Domain	99.1	95	Kenton	100	96.6
Wainwright	94.1	80.0	Mather	100	100	Melfort	100	99.1	Nokomis	95.8	79.1
Corning	98.3	*	Shaunavon	100	98.3	Tisdale	98.3	100	Wilkie	90	73.3
Unity	100	*									
Average	92.6	86.8		99.6	99.1		97.8	97.9		91.4	86.1

* As wheat samples from CPSR at locations 'Unity' and 'Corning' were received only from crop years 2008 and 2009, respectively, the classification accuracy of identifying these growing locations using the QDA were 0% and were not included for calculating overall classification accuracies.

Average classification accuracies were 90.5-97.1 and 85.6-96.0% for the LDA and QDA, respectively, for identifying growing locations of the 16% moisture content wheat. The highest classification accuracy (100%) was reached when classifying CPSR wheat from 'Corning' using the LDA for the 16% m.c. wheat. Classification accuracy of 80.8% was the lowest using the LDA when identifying CWRS from 'Corning'. The CWSWS from growing location 'Jansen' had the highest classification accuracy of 100% using the QDA. The CWRS from growing location 'Melfort' had 80% of classification accuracy that was the lowest using the QDA. The average classification accuracies ranged from 90.5-97.1% and 85.6-96% using the LDA and QDA, respectively, when identifying growing locations of 16% m.c. wheat independent of crop years (Table 5.8).

For identifying growing locations of the 19% m.c. wheat independent of crop years, average classification accuracies were 85.9-96.3% and 80.6-90.3% using the LDA and QDA, respectively (Table 5.9). For the 19% m.c. CPSR wheat, classification accuracies were comparatively low, which were 70.8-100 and 75-87.5%, for the LDA and QDA classifiers, respectively. Average classification accuracies of CWHWS wheat using the LDA and QDA were higher than the accuracies for other classes (LDA=96.3%; QDA=90.3%). Overall, in the non-parametric statistical classifiers, performance of the LDA was superior to that of the QDA for all moisture levels of wheat samples.

Table 5.8. Classification accuracies of identifying growing locations near the listed towns (independent of crop years) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) for 16% m.c. wheat.

Location	CPSR		Location	CWHWS		Location	CWRS		Location	CWSWS	
	LDA (%)	QDA (%)		LDA (%)	QDA (%)		LDA (%)	QDA (%)		LDA (%)	QDA (%)
Edmonton	99.1	99.1	Churchbridge	85	94.1	Corning	80.8	83.3	Corning	99.1	98.3
Rosemary	80.8	84.1	Kenton	94.1	90	Dauphin	86.6	92.5	Jansen	99.1	100
Viking	91.6	87.5	Limerick	98.3	88.3	Domain	99.1	89.1	Kenton	97.5	94.1
Wainwright	91.6	82.5	Mather	92.5	82.5	Melfort	93.3	80	Nokomis	92.5	97.5
Corning	100	*	Shaunavon	97.5	82.5	Tisdale	92.5	83.3	Wilkie	97.5	90
Unity	96.6	*									
Average	93.3	88.3		93.5	87.5		90.5	85.6		97.1	96.0

* As wheat samples from CPSR at locations 'Unity' and 'Corning' were received only from crop years 2008 and 2009, respectively, the classification accuracy of identifying these growing locations using the QDA were 0% and were not included for calculating overall classification accuracies.

Table 5.9. Classification accuracies of identifying growing locations near the listed towns (independent of crop years) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) for the 19% m.c. wheat.

Location	CPSR		Location	CWHWS		Location	CWRS		Location	CWSWS	
	LDA (%)	QDA (%)		LDA (%)	QDA (%)		LDA (%)	QDA (%)		LDA (%)	QDA (%)
Edmonton	81.6	75	Churchbridge	94.1	83.3	Corning	91.6	84.1	Corning	81.6	80
Rosemary	81.6	75.8	Kenton	96.6	90	Dauphin	95	96.6	Jansen	100	92.5
Viking	70.8	87.5	Limerick	100	90.8	Domain	86.6	79.1	Kenton	98.3	95
Wainwright	81.6	84.1	Mather	92.5	93.3	Melfort	83.3	78.3	Nokomis	88.3	80.8
Corning	100	*	Shaunavon	98.3	94.1	Tisdale	100	78.3	Wilkie	96.6	94.1
Unity	100	*									
Average	85.9	80.6		96.3	90.3		91.3	83.3		93.0	88.5

* As wheat samples from CPSR at locations 'Unity' and 'Corning' were received only from crop years 2008 and 2009, respectively, the classification accuracy of identifying these growing locations using the QDA were 0% and were not included for calculating overall classification accuracies.

5.3.8 Crop years

5.3.8.1 Identification of crop years of wheat independent of growing locations

Non-parametric classifiers were used for identifying crop years of wheat independent of growing locations. The CPSR wheat was the only class that was obtained for the crop year 2007. The rest of the sample set of CPSR wheat, which was received from 4-5 locations, was collected for the crop years 2008 and 2009, respectively. For all the other wheat classes, samples were received from five different locations for crop years 2008 and 2009. For the 13, 16, and 19% m.c. wheat, average classification accuracies using the QDA were greater than those for the LDA (Tables 5.10, 5.11, and 5.12).

In the 13% m.c. wheat, classification accuracies were 76.3-99.6% for classifying crop years using the LDA. The QDA had 84.6-100% classification accuracies for identifying crop year independent of growing locations. Near-perfect accuracies of above 98% were obtained for the CWHWS wheat for both LDA and QDA. Lower accuracies of 76.3-90% were obtained for CWSWS for non-parametric classifiers. Mahesh et al. (2010a and 2010b) reported that differing amounts of quality parameters (protein, starch, moisture, and hardness) in wheat samples can be related to higher classification accuracies for discriminating crop years using non-parametric classifiers.

Table 5.10. Classification accuracies of identifying crop years (independent of growing locations) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) models for 13% m.c. wheat.

Crop year	CPSR		CWHWS		CWRS		CWSWS	
	LDA	QDA	LDA	QDA	LDA	QDA	LDA	QDA
2007	98.3	*						
2008	92.9	94.1	98	99.6	98.6	99.3	79.3	84.6
2009	93.6	97.3	99.6	99	96.6	100	76.3	90
Average	94.9	95.7	98.8	99.3	97.6	99.7	77.8	87.3

* As CPSR wheat was the only class that was received from 2007, the classification accuracy in the QDA was 0% and it was not included for calculating average classification accuracies.

Classification accuracies were 70.5-95.5 and 90.2-97.5% for the LDA and QDA, respectively, for identifying crop years of wheat classes that had 16% m.c. High (99.1%) and low (82.6%) average classification accuracies were obtained for the CWHWS and CWSWS, respectively, for non-parametric classifiers. The performance of the QDA for identifying growing locations was better for all wheat classes except CWRS than that of the LDA. Similarly, Delwiche and Graybosch (2002) reported classification accuracies of 42-71% for identifying waxy wheat using a linear discriminant function with 1-10 principal component scores as input. The same authors also concluded that classification accuracies of 46-71% were possible to identify waxy wheat using a quadratic discriminant function with the same input criteria.

Table 5.11. Classification accuracies of identifying crop years (independent of growing locations) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) models for 16% m.c. wheat.

Crop year	CPSR		CWHWS		CWRS		CWSWS	
	LDA	QDA	LDA	QDA	LDA	QDA	LDA	QDA
2007	100	*						
2008	88.7	92.5	92.3	97.6	88.6	95.3	71.6	93
2009	91.6	91.6	98.6	97.3	93.3	94.6	69.3	87.3
Average	93.4	92.1	95.5	97.5	91.0	95.0	70.5	90.2

* As CPSR wheat was the only class that was received from 2007, the classification accuracy in the QDA was 0% and it was not included for calculating average classification accuracies.

Table 5.12. Classification accuracies of identifying crop years (independent of growing locations) using the linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) models for 19% m.c. wheat.

Crop year	CPSR		CWHWS		CWRS		CWSWS	
	LDA	QDA	LDA	QDA	LDA	QDA	LDA	QDA
2007	91.6	*						
2008	98.3	92.5	92.3	93	84	92	80.6	87
2009	87.6	96.6	93.3	97.6	79	87	86.6	88.6
Average	92.5	94.5	92.8	95.3	81.5	89.5	83.6	87.8

* As CPSR wheat was the only class that was received from 2007, the classification accuracy in the QDA was 0% and it was not included for calculating average classification accuracies.

For the 19% m.c. wheat, average classification accuracies were 81.5-92.8% and 87.8-95.3% using the LDA and QDA, respectively. Accuracies were high for CPSR and CWHWS wheat for identifying crop years. In related studies, classification accuracies of 94-100% and 86-100% were reported using the LDA and QDA, respectively, in identifying composite samples of western Canadian wheat classes with a moisture content of 11% (Mahesh et al., 2008). The average classification accuracy of the LDA using the top 90 wavelet features was the highest (99.1%) in a study conducted to identify wheat classes from the NIR hyperspectral images of bulk samples (Choudhary et al., 2009).

5.3.9 Identification of key wavelengths

Procedure STEPDISC was used to find out the relative importance of the NIR reflectance features towards classification. Table 5.13 shows the rankings of the features that were categorized based on their contributions in the descending order to the discriminatory power of the model for the 13% m.c. wheat. The stepwise discrimination method identified important wavelengths that contributed mostly to identify growing locations independent of crop years and to separate wheat classes based on crop years independent of growing locations. For the 13% m.c. wheat, features at the wavelengths in the NIR regions of 1000-1100, 1260-1380, and 1650-1700 nm were found important in identifying growing locations independent of crop years for wheat. A similar trend was observed for separating wheat samples based on crop years independent of growing locations. Similar NIR hyperspectral imaging studies showed that the NIR absorption of water, protein, and carbohydrate can be observed at 960 and 1420; 1470-1500; 1200-1360 and 1610-1700 nm, respectively.

Table 5.13. The top 10 features, in descending order, based on their contribution towards classification accuracy for a) identifying growing locations independent of crop years and b) identifying crop years independent of growing locations for the 13% m.c. wheat while using non-parametric classifiers with all features as inputs.

Rank	a) Top 10 wavelengths for identifying growing locations independent of crop years (nm)				b) Top 10 wavelengths for identifying crop years independent of growing locations (nm)			
	CWRS	CPSR	CWHWS	CWSWS	CWRS	CPSR	CWHWS	CWSWS
1	1680	1690	1700	1700	1700	1350	1100	1540
2	1700	1570	1530	1690	1670	1190	1360	1690
3	1670	1270	1270	1370	1680	1280	1070	1670
4	1690	1050	1680	1680	1690	1590	1190	1700
5	1310	1700	1190	1480	1310	1030	1210	1130
6	1080	1210	1690	1670	1240	1070	1050	1660
7	1020	1670	1310	1260	1540	1700	1240	1120
8	1580	1060	1140	1390	1640	1060	1180	1090
9	1660	1260	1100	1210	1080	1690	1610	960
10	1270	1470	1360	1240	1160	1460	1400	1290

Table 5.14 shows rankings of features (corresponding wavelengths are shown), which are crucial for finding growing locations independent of crop years and crop years independent of growing locations, in the regions of 1100-1200, 1260-1390, and 1650-1700 nm. The complexity of the nature of intrinsic components of wheat needs further investigation, which was performed in latter sections, to arrive at a chemical meaning for peak wavelengths. Variations in oil content of wheat samples can be noticed at differences in the NIR absorption at 1390 nm.

Table 5.14. The top 10 features, in descending order, based on their contribution towards classification accuracy for a) identifying growing locations independent of crop years and b) identifying crop years independent of growing locations for the 16% m.c. wheat while using non-parametric classifiers with all features as inputs.

Rank	a) Top 10 wavelengths for identifying growing locations independent of crop years (nm)				b) Top 10 wavelengths for identifying crop years independent of growing locations (nm)			
	CWRS	CPSR	CWHWS	CWSWS	CWRS	CPSR	CWHWS	CWSWS
1	1700	1290	1290	1620	1510	1690	1700	1620
2	1110	1010	1590	1700	1090	1190	1110	1140
3	1670	1700	1190	1680	1060	1450	1660	1160
4	1350	1690	1310	1090	1100	1350	1000	1020
5	1210	960	1510	1350	1400	1110	970	1570
6	1690	1560	1210	1110	1110	1140	1090	1380
7	1450	1260	1700	1100	1050	1680	1160	1240
8	1130	1210	960	1020	960	1470	1690	1230
9	1230	1310	1690	1320	1000	1030	1210	1170
10	1320	1360	1110	1080	970	1000	1010	1370

The procedure STEPDISC was used for finding out the relative importance of input features for identifying crop years and growing locations of wheat samples that had 19% m.c. Table 5.15 shows the top 10 features (related wavelengths are shown) that were identified and ranked based on their contribution in the descending order. The NIR reflectance intensities of wavelengths in the regions of 960-1210, 1300-1380, 1490-1570, and 1650-1700 nm were identified. Similar studies identified that wavelengths of 960-1060, 1330-1480, and 1680 nm associate to the NIR absorption of hardness.

Table 5.15. The top 10 features, in descending order, based on their contribution towards classification accuracy for a) identifying growing locations independent of crop years and b) identifying crop years independent of growing locations for the 19% m.c. wheat while using non-parametric classifiers with all features as inputs.

Rank	a) identifying growing locations independent of crop years (nm)				b) identifying crop years independent of growing locations (nm)			
	CWRS	CPSR	CWHWS	CWSWS	CWRS	CPSR	CWHWS	CWSWS
1	1700	1700	1700	1290	1510	1680	1700	1060
2	1380	1670	1510	1700	1260	1700	1300	1640
3	1450	1690	1290	1670	1200	1690	1580	1700
4	1110	1030	1110	1130	1210	1460	1130	1680
5	1670	1140	1130	1210	1700	1360	1060	1270
6	1210	1460	1330	1020	1690	1060	1370	990
7	1300	1210	1000	1690	1080	1140	1230	1510
8	1120	1290	1030	1140	1050	1510	1030	1670
9	1220	1190	1660	1150	1110	1670	1630	1170
10	1570	1090	1690	1080	960	1210	1160	1380

In the current study, the results showed that The LDA and QDA had very superior classification for four-class and pair-wise discrimination of wheat classes using moisture-specific samples. In a related study, Mahesh et al. (2011) used the LDA and QDA to identify wheat classes, which had varying moisture levels, with accuracies of 61-97 and 82-99%, respectively, free from the effect of moisture contents. Once wheat classes were identified, 90-100 and 72-99% of classification accuracies were attained using the LDA and QDA, respectively, to identify specific moisture levels. Reflectance features at 1060, 1090, 1340, and 1450 nm were ranked at top to identify wheat at varying levels of

moisture. In another NIR hyperspectral imaging study on wheat, classification accuracies of 85-100% were reported for identifying healthy and insect-damaged wheat kernels using the LDA and QDA (Singh et al., 2007).

5.4 Back propagation neural network (BPNN) classifiers for identifying moisture levels and wheat classes

Previous research works (reported below) showed that the BPNN classifiers are the best suited for classification of cereal grains. The classification accuracies of nearly 100% were obtained using bulk sample images of cereal grains (Majumdar and Jayas, 1999). Visen et al. (2004) reported classification accuracies of above 98% for identifying cereal grains such as wheat, barley, oats, and rye using a four-layer BPNN classifier. Paliwal et al. (2001) reported classification accuracies of over 97% for CWRS, CWAD wheat, and oats and about 88% for barley and rye while using morphological features of colour images as input for four-layer BPNN classifiers. They found that the general regression neural network architecture was not fully suitable for cereal grain identification. Performance of a four-layer BPNN and specialist probabilistic neural network (SPNN) architecture were compared towards classifying individual kernels of wheat, rye, barley, and oats (Visen et al., 2004). They concluded that BPNN was better than SPNN for identifying cereal grains. Wang et al. (1999) reported that classification accuracies of BPNN models were 98.8 and 98% for divergence feature selection method and principal components analysis (PCA), respectively, for discriminating single kernels of wheat using the reflectance spectra of 400-2000 nm. The divergence feature selection method and the PCA were helpful in reducing the number of reflectance features used during BPNN classification.

In the present study, a four-layer standard back propagation neural network classifier was used for the pair-wise discrimination of wheat classes and moisture levels. Two-way identifications were carried out for 13 versus 16%, 16 versus 19%, and 13 versus 19% moisture contents independent of wheat classes. Identification of wheat at each moisture level was performed for the following combinations: CWRS versus CPSR, CWRS versus CWHWS, CWRS versus CWSWS, CPSR versus CWHWS, CPSR versus CWSWS, and CWHWS versus CWSWS.

5.4.1 Pair-wise identification of moisture contents

The input feature set used consisted of 75 NIR reflectance intensities in the wavelength region of 960-1700 nm. A four-layer BPNN classifier was used for pair-wise identification of moisture contents. Input, two hidden and output layers consisted of 75, 34, 34, and 2 nodes, respectively. The following combinations of data sets were tested: 13 vs. 16%, 16 vs. 19%, and 13 vs. 19%. In total, 4800 samples were used for all iterations in each analysis (2400 per moisture level). The summarized results are shown in Figure 5.26. The average classification accuracies of five iterations were 88.2 and 98.7%, which were the highest, for identifying 13 and 19% m.c. respectively. It showed that the BPNN classifiers could be used for discriminating low and high moisture wheat. Overall average classification accuracy of pair-wise BPNN classifiers was 83.7%. The thirteen vs. nineteen pair followed by the thirteen vs. sixteen pair were classified the best using the BPNN classifier. Average classification accuracies of 93.5% were obtained, where difference between the moisture levels were high (about 6%), for identifying low (13%) and high (19%) moisture wheat. Average classification accuracy was 78.8% for the

BPNN for the wheat samples that had a low difference (about 3%) in their moisture levels.

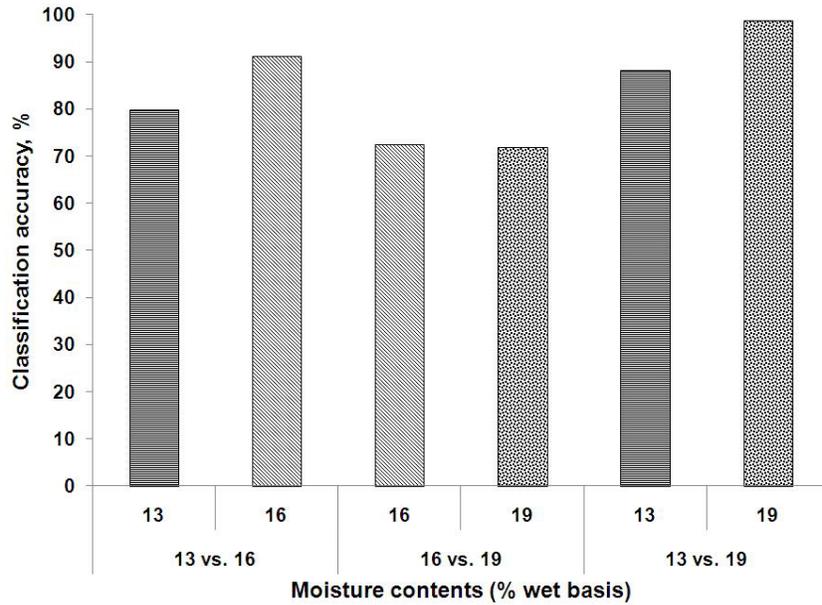


Fig. 5.26. Pair-wise classification accuracies of identifying moisture levels of wheat using a four-layer BPNN.

Contributions of input features towards classification were calculated by the BPNN network for all iterations. The top ten input feature wavelengths that had highest average contribution values are shown in Table 5.16. Contribution values show that most of the top features come from 1200-1220, 1310-1340, and 1390-1420 nm. The NIR absorption of water, protein, carbohydrate, kernel hardness, oil content, fiber, and starch content of wheat are seen in the long wavelength near-infrared (LWNIR) region of 960-1700 nm (Wang et al., 1999, Delwiche and Massie, 1996, and Murray and Williams). The NIR absorptions of wheat samples produced excitations for carbon-hydrogen (CH) and oxygen-hydrogen (OH) combination bonds; CH, OH, and nitrogen-hydrogen (NH) 1st overtones, 2nd overtones, and CH 3rd overtones in samples (Osborne, 2006). Singh et al.

(2007) reported that 1284.2, 1315.8, and 1347.4 nm were ranked at the top based on the first principal component factor loadings out of 20 LWNIR wavelengths in the region of 1000-1600 nm while detecting fungal-infected wheat using the NIR hyperspectral images of individual kernels.

Table 5.16. The top 10 feature wavelengths, in descending order, based on their contribution towards classification accuracy for moisture identification while using pair-wise back propagation neural network classifiers with all features as inputs.

Rank	Top 10 wavelengths for pair-wise discrimination (nm)		
	13 vs. 16	16 vs. 19	13 vs. 19
1	1320	1330	1330
2	1340	1340	1340
3	1400	1700	1400
4	1330	1310	1310
5	1410	1210	1410
6	1210	1400	1210
7	1310	1390	1320
8	1030	1200	1200
9	1390	1140	1630
10	1420	1150	1220

After identifying moisture contents, wheat classes were discriminated pair-wise at specific moisture levels using the BPNN classifier.

5.4.2 Classification of wheat at 13% m.c.

The network consisted of 75, 27, 27, and 2 nodes in the input, two hidden, and output layers, respectively, for each wheat class. Input data set consisted of 1200 samples in total (600 samples per wheat class) for all iterations in each treatment. The summarized results of pair-wise classification of 13% moisture content wheat are shown in Figure 5.27. The highest classification accuracies of 88.6 and 91% were obtained for identifying

CPSR and CWSWS wheat, respectively. Classification accuracies were 80.5 and 86.3% for identifying red wheat classes CWRS and CPSR, respectively. White wheat classes (CWHWS and CWSWS) were identified with classification accuracies of 69.8 and 83.3%, respectively. Average classification accuracies were 70.5 and 86.2% for identifying soft (CWSWS) wheat and hard (CWRS and CWHWS) wheat, respectively. Overall average classification accuracy of two-way BPNN classifiers was 83.2% for identifying wheat at 13% m.c.

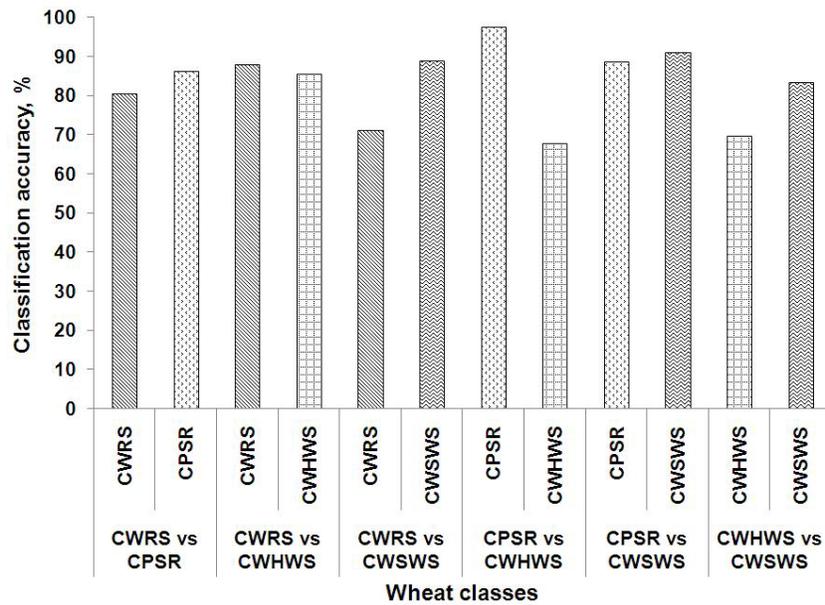


Fig. 5.27. Pair-wise classification accuracies of identifying wheat classes that had 13% m.c. using a four-layer BPNN.

The top 10 NIR feature wavelengths that were ranked based on their highest contribution values towards classification are shown in Table 5.17. Most of the features that contribute significantly towards classification are derived from wavelengths in the regions of 1070-1130, 1210-1270, 1300-1350, and 1660-1700 nm. When comparing to non-parametric statistical classifiers, Mahesh et al. (2010a) reported that the linear and

quadratic discriminant classifiers had average classification accuracies of 95.4 and 92.3%, respectively, for identifying the 13% moisture content wheat that included sample variations due to growing locations and crop years.

Table 5.17. The top 10 feature wavelengths, in descending order, based on their contribution towards classification accuracy for wheat class identification at 13% m.c. while using pair-wise back propagation neural network classifiers with all features as inputs.

Rank	Top 10 wavelength for pair-wise discrimination (nm)					
	CWRS vs. CPSR	CWRS vs. CWHWS	CWRS vs. CWSWS	CPSR vs. CWHWS	CPSR vs. CWSWS	CWHWS vs. CWSWS
1	1700	960	1680	1700	1700	1700
2	960	1300	960	1680	1690	1100
3	1300	1700	1700	1690	1320	1690
4	1690	1210	1320	1670	1680	1400
5	1270	1680	1690	1310	1260	1300
6	1030	1130	1070	1130	1270	1130
7	1670	1450	1300	1320	960	1330
8	1450	1690	1670	1070	1350	1110
9	1210	1310	1310	1660	1210	1210
10	1680	1670	1660	1300	1330	1230

5.4.3 Classification of wheat at 16% m.c.

In the class identification part of the 16% moisture content wheat, a pair-wise four-layer BPNN classifier was used. Input, two hidden, and output layers of neural network classifiers had 75, 27, 27, and 2 nodes, respectively, for each wheat class. Input feature set consisted of 600 samples in each wheat class and 1200 samples for each analysis. The pair-wise classification results are shown in Figure 5.28. In the present study, the highest classification accuracies were 91.1 and 78.2% for identifying CWRS and CWSWS wheat, respectively. Classification accuracies of 84.9 and 80.6% were

obtained for identifying CWRS and CPSR (red wheat classes), respectively. The CWHWS and CWSWS (white wheat classes) were identified with classification accuracies of 71.7 and 69.6%, respectively. Overall average classification accuracy of 75.4% was obtained for two-way BPNN classifiers for identifying the 16% m.c. wheat. Song et al. (1995) developed BPNN models for classifying wheat using the NIR transmittance spectra that were developed in the wavelength region of 850-1049 nm. They reported that two-class BPNN models had average classification accuracies of 97-100% which were higher than that of six-class models.

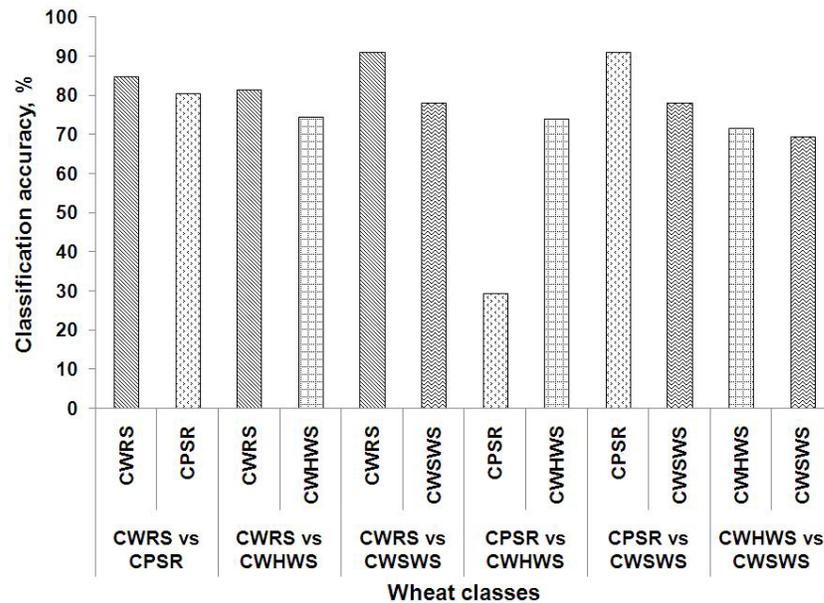


Fig. 5.28. Pair-wise classification accuracies of identifying wheat classes that had 16% m.c. using a four-layer BPNN.

The contribution rankings in descending order for the NIR features (corresponding wavelengths are shown) to the classification process are given in Table 5.18. Features from NIR wavelengths in the regions of 1070-1130, 1300-1410, and 1690-1700 nm contributed mostly to the classification process. Mahesh et al. (2010b) reported

that features at NIR wavelength regions of 1000-1200 and 1260-1390 nm were ranked at the top for identifying the 16% m.c. wheat.

Table 5.18. The top 10 feature wavelengths, in descending order, based on their respective contribution towards classification accuracy for wheat class identification at 16% m.c. while using pair-wise back propagation neural network classifiers with all features as inputs.

Rank	Pair-wise discrimination (nm)					
	CWRS vs. CPSR	CWRS vs. CWHWS	CWRS vs. CWSWS	CPSR vs. CWHWS	CPSR vs. CWSWS	CWHWS vs. CWSWS
1	1700	1320	1210	1700	1700	1700
2	1690	1010	1700	1400	1320	1690
3	1320	1700	1070	1330	1340	1320
4	1010	1330	1120	1130	1350	1310
5	1330	1350	1190	1690	1370	1360
6	1070	1690	1380	1010	1070	1380
7	1350	1260	1570	1550	1360	1300
8	1100	1070	1390	1100	1300	1350
9	1540	1110	1640	1600	1380	1340
10	1400	1410	1400	1310	1690	1200

5.4.4 Classification of wheat at 19% m.c.

In the present study, a two-way four-layer BPNN classifier was used for identifying classes of 19% m.c. wheat. The classifier had 75 input nodes, 27 nodes in each of two hidden layers, and 2 output nodes for each wheat class. Input feature set consisted of 600 samples in each wheat class and 1200 samples for each analysis. The pair-wise classification results are shown in Figure 5.29. The highest classification accuracies were 82.3 and 87.6% for identifying CPSR and CWSWS wheat, respectively. In the present study, average classification accuracies of 71.4 and 77.4% were obtained

for identifying red (CWRS vs. CPSR) and white (CWSWS vs. CWHWS) wheat classes, respectively. Overall average classification accuracy was 73.1% for two-way BPNN classifiers for discriminating the 19% m.c. wheat classes.

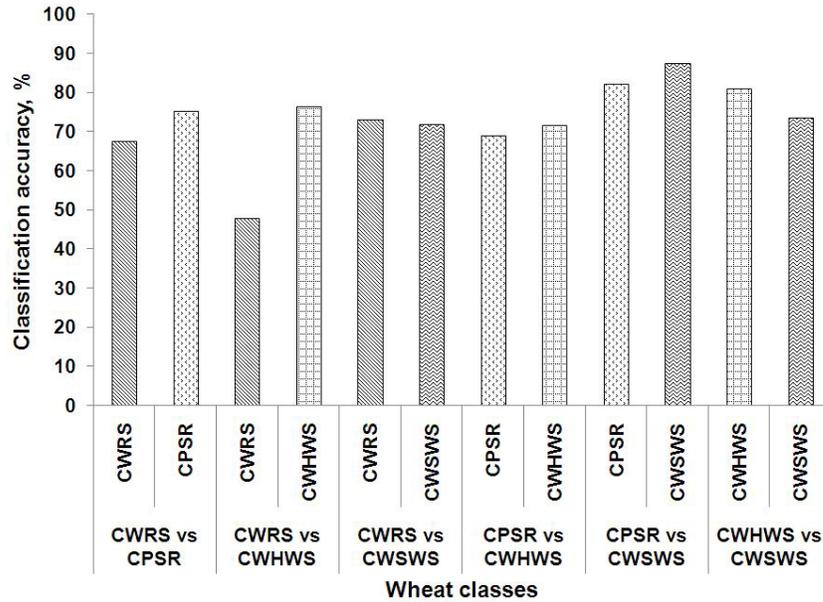


Fig. 5.29. Pair-wise classification accuracies of identifying wheat classes that had 19% m.c. using a four-layer BPNN.

Table 5.19 shows the rankings of the NIR features, where related wavelengths are shown, based on their contribution towards classification. Features in the NIR regions of 1040-1130, 1290-1350, and 1660-1700 nm are crucial to the BPNN classification. Wavelengths of 960 and 1420; 1470-1500; 1200-1360 and 1610-1700 nm were responsible for the NIR absorption of water, protein, and carbohydrate contents of samples, respectively. The NIR absorptions of wheat in the wavelength regions of 960-1060, 1330-1480 nm, and at 1680 nm are likely associated to protein, starch, and hardness of samples (Maghirang and Dowell, 2003; Murray and Williams, 1987). These wavelengths can be added to a multispectral imaging system that will be used for

conducting on-line class identification in wheat in the future. The 960-1060, 1330-1480 nm wavelength regions, and 1680 nm wavelength relate to the NIR absorption of hardness of samples. Maghirang and Dowell (2003) reported that the NIR wavelengths of 1100, 1200, 1380, 1450, and 1670 nm were contributed mainly in predicting hardness of wheat. Variations in oil content of wheat samples can be seen by differences in the NIR absorption at 1390 nm.

Table 5.19. The top 10 feature wavelengths, in descending order, based on their contribution towards classification accuracy for wheat class identification at 19% m.c. while using pair-wise back propagation neural network classifiers with all features as inputs.

Rank	Top 10 wavelengths for pair-wise discrimination (nm)					
	CWRS vs. CPSR	CWRS vs. CWHWS	CWRS vs. CWSWS	CPSR vs. CWHWS	CPSR vs. CWSWS	CWHWS vs. CWSWS
1	1700	1700	1700	1700	1700	1700
2	1690	1010	1110	1570	1350	1690
3	1110	1690	1690	1350	1060	1680
4	1660	1230	1310	1690	1340	1320
5	1670	960	1210	1060	1100	1130
6	1400	1400	1470	1100	1320	1110
7	1680	1330	1330	1400	1210	960
8	1540	1130	1320	1680	1040	1290
9	1070	1490	1130	980	1070	1670
10	1080	1200	1300	1110	1310	1350

5.5 The PLSR and PCR models for protein and hardness of wheat

Wheat milling gives various types of flour such as whole wheat flour, all purpose flour, and semolina for human consumption and considerable quantities of by-products such as bran for animal feeds. Famera et al. (2004) stated that cake and biscuit flour

require low protein and soft wheat classes that would have minimum damages of starch during milling. But, bakery flour required high protein and hard wheat classes. Durum wheat is used for producing pasta because its protein, hardness and gluten strength values are high.

In the present study, hyperspectral image cubes in the LWNIR wavelength region of 960-1700 nm were examined for non-destructive determination of protein content and hardness in wheat. The NIR reflectance spectra were analyzed using multivariate calibration techniques such as the partial least squares regression (PLSR) and the principal components regression (PCR).

5.5.1 Protein content of wheat

Four western Canadian wheat classes, CWRS, CPSR, CWHWS, and CWSWS, were used in this study. Table 5.20 shows the summary of protein dataset statistics and results of grouping that was performed using the Scheffe's test. The American Organization of Analytical Chemists (AOAC) standard method 968.06 was used for determining protein contents of wheat samples. Dumas method (Schmitter and Rihs, 1989), a modification of Leco version 2.2, was used for analyzing crude protein or total nitrogen contents of wheat samples.

Table 5.20. Summary of protein dataset statistics for wheat (n = 30 per wheat class)

Wheat class	Protein content			
	Average (%)*	Maximum (%)	Minimum (%)	Standard deviation (%)
CWRS	14.62 ^a	16.19	11.88	1.28
CPSR	13.26 ^b	15.39	11.38	1.13
CWHWS	14.38 ^a	16.45	12.12	1.44
CWSWS	11.25 ^c	12.93	9.51	1.06

*Values with same letters in the column are not significantly different by Scheffe's test for comparing means.

5.5.1.1 The PLSR and PCR prediction of protein contents of wheat samples using the full data set

The PLSR and PCR models were developed using the average spectra from 7200 samples utilizing the full spectral range of 960-1700 nm (75 wavebands). A ten-fold cross-validation method was used for both PLSR and PCR methods for choosing the optimal number of components that were involved in model development processes. It is a more statistically reliable method for choosing components in the PLSR or PCR. Figure 5.30 shows the percent variations, which was represented by the components of PLS and PCR, of the input data.

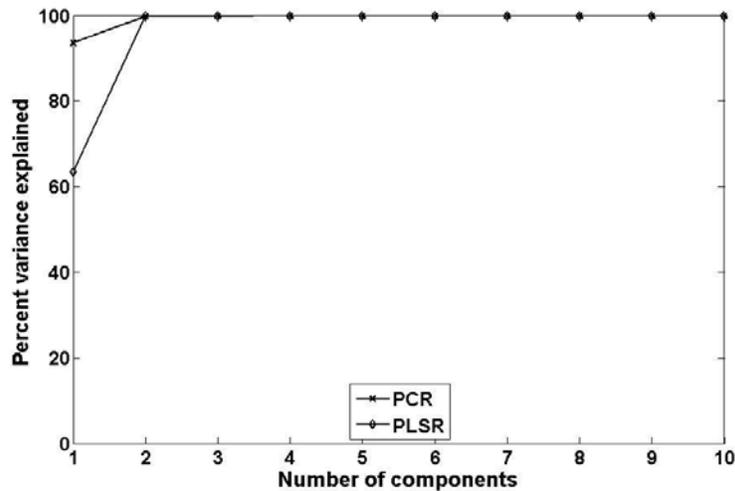


Fig. 5.30. Percent variance explained by the components of PLSR and PCR models for wheat.

Over-fitting of data was restricted in both model fitting and prediction error estimation. The first 3 components of the PLSR and PCR accounted for 99.9% of

variations in the input data set. For both PLSR and PCR, optimal model components and prediction performance were estimated using a ten-fold cross-validation method. In ten-fold cross-validation method, the full data set was randomly divided into ten groups (i.e., by having 720 samples in each group). The PLSR or PCR predictions for each of the ten groups, which had 720 samples, were calculated using the calibration equations derived from the rest of the nine groups (i.e., by using $720 \times 9 = 6,480$ samples) of NIR spectra of wheat samples. Ten iterations were performed in the cross-validation and mean square errors of prediction (MSEP) were estimated for explaining model performances. The optimal number of components for the PLSR and PCR was selected by considering the component that had the minimum MSEP values. Figure 5.31 shows the estimated MSEP values for components used in the PLSR and PCR models for predicting protein contents of wheat samples. The minimum MSEP values were 2.02 and 1.76 for the PCR and PLSR with ten components, respectively.

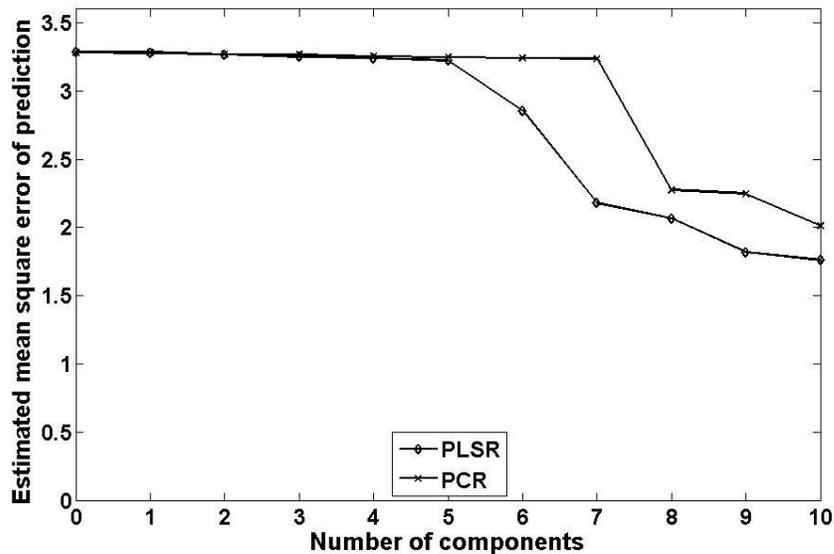


Fig. 5.31. Estimated MSEP values for the components of PLSR and PCR models for predicting protein contents of wheat.

For the protein content measurement part, the accurate results were obtained using the PLSR on the NIR reflectance spectra. With 10 factors included, the PLSR calibration achieved an estimated mean square error of prediction (MSEP) of 1.76 with a standard error of prediction (SECV) of 1.33 and a correlation coefficient (r) of 0.68. The PCR, which had first 10 components in the model, attained 2.02, 1.42, and 0.62 for the estimated MSEP, SECV, and r , respectively. Cogdill et al. (2004) reported an SECV of 1.2% and an r^2 of 0.87 for predicting moisture concentrations in individual kernels of maize using the NIR hyperspectral imaging. The predicted protein contents against the observed protein contents of wheat samples for the ten-factor PLSR and ten-factor PCR models are shown in Figs. 5.32 and 5.33, respectively.

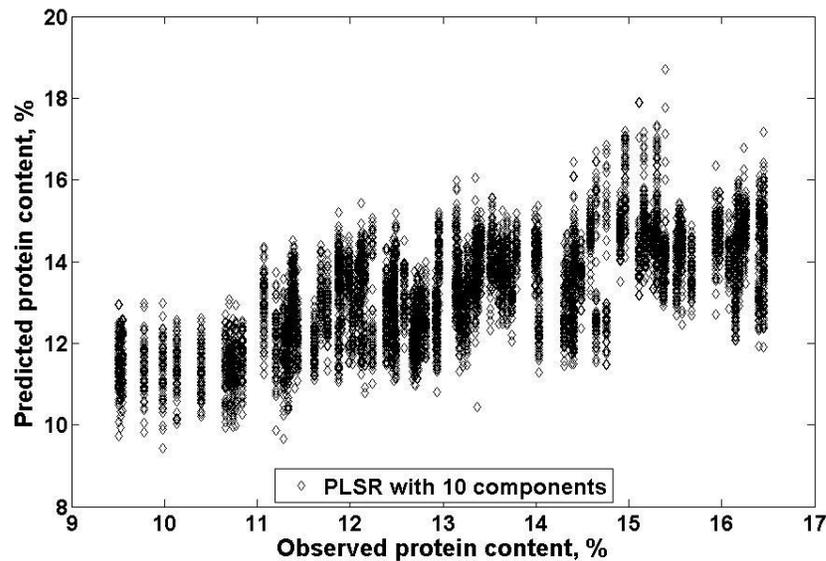


Fig. 5.32. Predicted protein contents against observed protein contents of wheat using the ten-factor PLSR model ($n = 7200$, $r = 0.68$, and $SECV = 1.33$).

In a similar study, ElMasry et al. (2007) observed that the PLSR method had very high correlation for predicting moisture, total soluble solids content, and pH for strawberries using the Vis-NIR hyperspectral images acquired from the wavelength

region of 400-1000 nm. Also, in the same study, the moisture contents of strawberries were precisely predicted using the PLSR with r value 0.96 for the validation set.

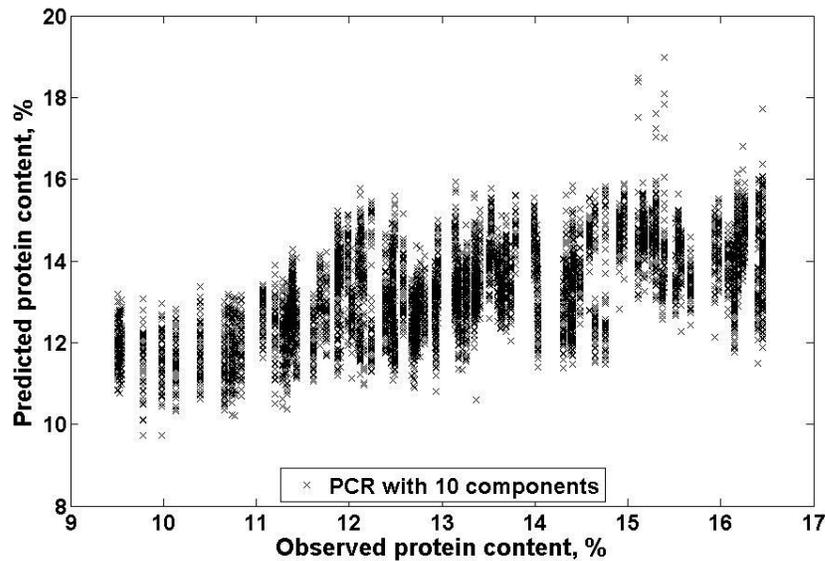


Fig. 5.33. Predicted protein contents against observed protein contents of wheat using the ten-factor PCR model ($n = 7200$, $r = 0.62$, and $SECV = 1.42$).

Based on the values of β -coefficients and the first PC loadings, the optimal wavelengths for predicting protein contents of wheat were identified for the PLSR and PCR methods, respectively (Table 5.21). As optimal wavelengths explained maximum information of the spectral data, they could be used in future for on-line applications that were performed by multispectral imaging. The NIR reflectance features from the wavelength regions of 1180-1200 and 1460-1500 nm were found important by considering β -coefficient values of the PLSR method. The NIR absorptions of protein contents of samples were seen at the wavelengths of 1470-1500 nm. Wavelengths from the regions of 960-1030 and 1670-1700 nm identified vital for the first PC loading values in the PCR method. These wavelengths could be added for developing a multispectral

imaging system that will be used in future for conducting on-line quality assessments in wheat.

Table 5.21. The top 10 wavelengths, in descending order, identified for the ten-factor PLSR and PCR models for predicting protein contents of wheat samples

Rank	Top 10 wavelengths for PLSR (nm)	Top 10 wavelengths for PCR (nm)
1	1190	1700
2	1210	1690
3	1180	1670
4	1090	1680
5	1670	960
6	1220	970
7	1130	980
8	1490	1030
9	1500	1020
10	1460	990

5.5.1.2 Class-wise PLSR and PCR prediction of protein contents of wheat samples

In this part of the study, the multivariate regression models, which included PLSR and PCR, were developed using the class-specific NIR reflectance intensities and reference protein contents of wheat samples as independent and dependent variables, respectively. The average NIR reflectance spectra from the LWNIR wavelength region of 960-1700 nm of class-specific wheat samples, which included 1800 samples per class from CWRS, CPSR, CWHWS, and CWSWS, were used. In the Appendix, the percent variations, which were explained by the components of PLSR and PCR for CWRS, CPSR, CWHWS, and CWSWS, are shown in Figs. D1, D2, D3, and D4, respectively. The first 3 components of the PLSR and PCR accounted for > 99.6% of variations of the input data set for all wheat classes.

A ten-fold cross validation method was used for PLSR and PCR for estimating the model performance and finding out an optimal number of components for model development. In this cross-validation method, the data set was randomly split into ten groups (i.e., by having 180 samples in each group). The PLSR or PCR predictions of protein contents of samples for each of the ten groups, which consisted of 180 samples each, were calculated using the calibration equations developed from the NIR spectra of wheat samples of the rest of the nine groups (i.e., by using $180 \times 9 = 1,620$ samples). The MSE values were calculated from ten iterations that were performed using the cross-validation. The optimal number of components for the PLSR and PCR was selected. Figures D5, D6, D7, and D8 in the Appendix show the estimated MSE values for the components of the PLSR and PCR for ten-fold cross-validation for predicting protein contents of class-specific wheat samples of CWRS, CPSR, CWHWS, and CWSWS, respectively. Table 5.22 shows the estimated MSE, SECV, and r values for the ten-factor PCR and PLSR, respectively, for predicting protein contents of class-specific wheat samples. The ten-factor PLSR models for all class-specific wheat samples performed better than the ten-factor PCR models. For all wheat classes, the correlation coefficients of PLSR were higher than those of PCR. The more accurate protein content predictions were expected from the CWHWS and CWSWS wheat samples that had correlation coefficient values of 0.71 and 0.77, respectively. The correlation coefficient values of 0.57-0.61 were achieved for the ten-factor PCR models for predicting protein contents. Predicted protein contents against observed protein contents for class-specific wheat samples for the ten-factor PLSR and the ten-factor PCR models are shown in Figs. D9-D12 and Figs. D13-D16, respectively.

Table 5.22. Performance of ten-factor PLSR and PCR models for predicting protein contents using the class-specific wheat samples.

Wheat class	Ten-factor PLSR model			Ten-factor PCR model		
	Estimated MSEP	SECV	r	Estimated MSEP	SECV	r
CWRS	0.961	0.980	0.65	1.071	1.035	0.59
CPSR	0.712	0.844	0.68	0.849	0.921	0.57
CWHWS	1.044	1.022	0.71	1.310	1.145	0.59
CWSWS	0.467	0.683	0.77	0.683	0.826	0.61

Using β -coefficients and the first PC loading values, the optimal wavelengths for predicting protein contents of class-specific wheat samples were identified (Table 5.23). The NIR reflectance features from the wavelength regions of 960-1220, 1280-1360, and 1610-1670 nm were important using the 10-factor PLSR method.

Table 5.23. The top 10 wavelengths, in descending order, identified for the ten-factor PLSR and PCR models for predicting protein contents of class-specific wheat samples.

Rank	Top 10 wavelengths for PLSR (nm)				Top 10 wavelengths for PCR (nm)			
	CWRS	CPSR	CWHWS	CWSWS	CWRS	CPSR	CWHWS	CWSWS
1	1010	1150	1330	1090	1700	1580	1570	1700
2	1110	1340	1130	1210	1670	1590	1590	1690
3	1120	1210	1260	1220	1690	1570	1610	1680
4	1190	1190	980	1100	1680	1600	1600	1670
5	1280	1110	1630	1080	960	1610	1560	960
6	960	1360	1640	1450	970	1520	1580	970
7	1000	1100	970	1670	1030	1560	1510	980
8	1150	1120	1180	1190	980	1510	1520	1660
9	1060	1040	1610	1330	1020	1550	1550	990
10	1220	960	1620	1110	990	1530	1530	1010

From the ten-factor PCR method, the NIR reflectance features from the wavelength regions of 960-1030 and 1510-1700 nm were ranked at the top. These

wavelengths could be used on-line quality assessments in class-specific wheat samples. From similar studies, it was reported that the NIR absorptions of wheat in the wavelength regions of 960-1060, 1330-1480 nm, and at 1680 nm were likely associated to protein, starch, and hardness of samples (Maghirang and Dowell, 2003; Murray and Williams, 1987).

5.6.1 Hardness of wheat

In this part of the study, four western Canadian wheat classes, CWRS, CPSR, CWHWS, and CWSWS, were used. Table 5.24 shows the summary of hardness dataset statistics and grouping results that was performed by comparing means using the Scheffe's test. A single kernel characterization system (SKCS) 4100 (Perten Instruments, Springfield, IL) was used for measuring hardness of wheat. The force and conductivity measurements were taken when wheat kernels were crushed.

Table 5.24. Summary of hardness dataset statistics for wheat (n = 300 kernels per wheat class per moisture).

Wheat class	Hardness			
	Average*	Maximum	Minimum	Standard deviation
CWRS	70.8 ^{ab}	87	57	8.73
CPSR	66.1 ^b	78.8	54.5	7.21
CWHWS	78.9 ^a	86.6	66.2	7.87
CWSWS	27.9 ^c	37.9	12.8	8.67

* Values with same letters in the column are not significantly different by the Scheffe's test for comparing means.

Mean hardness index, weight, size, moisture, and their standard deviations were determined from the single kernel data obtained from a sample of 300 kernels of wheat.

Delwiche and Norris (1993) reported that the NIR hardness values were significantly different for most of the hard red winter (HRW) and hard red spring (HRS) wheat classes.

5.6.1.1 The PLSR and PCR prediction of hardness of wheat samples using the full data set

Average NIR reflectance spectra from 7200 samples utilizing the full spectral range of 960-1700 nm were utilized for predicting hardness using PLSR and PCR models. A ten-fold cross-validation method was used for both PLSR and PCR for selecting the optimal number of components that were further used in the prediction studies. Figure 5.34 shows the percent variations, which were represented by the components of PLS and PCR, of the input data. The first 3 components of the PLSR and PCR accounted for 99.9% of variations in the input data set.

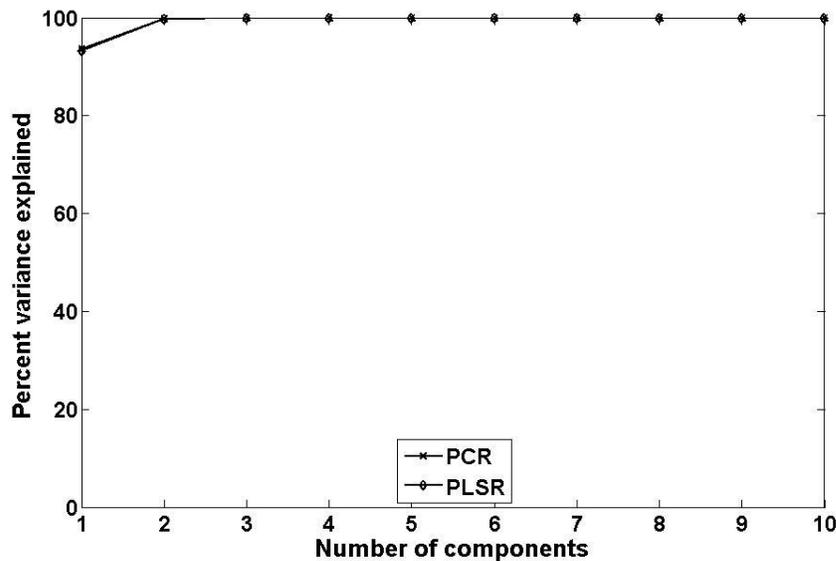


Fig. 5.34. Percent variance explained by the components of PLSR and PCR models for wheat.

For both PLSR and PCR methods, optimal model components and prediction performances were calculated using a ten-fold cross-validation method. The procedure for performing the ten-fold cross-validation method was already discussed in the section 5.5.1. Ten replications were performed in the cross-validation process and mean square errors of prediction (MSEP) were estimated. The optimal number of components for the PLSR and PCR was selected by considering the component that had the minimum MSEP value. Figure 5.35 shows the estimated MSEP values for components of the PLSR and PCR models for predicting hardness of wheat samples. The minimum MSEP values were 193.1 and 147.7 for the tenth component of PCR and PLSR models, respectively.

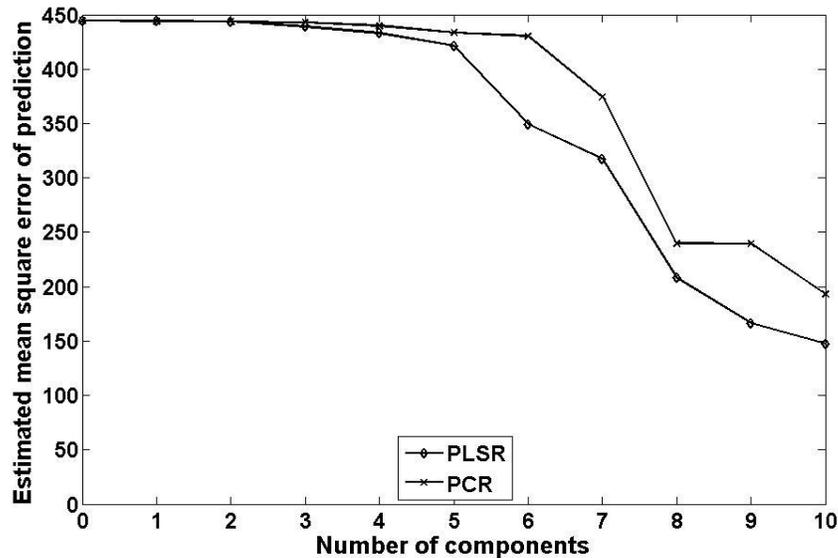


Fig. 5.35. Estimated MSEP values for the components of PLSR and PCR models for predicting hardness of wheat.

The predicted hardness against the observed hardness of wheat samples for the ten-factor PLSR and ten-factor PCR models are shown in Figs. 5.36 and 5.37, respectively.

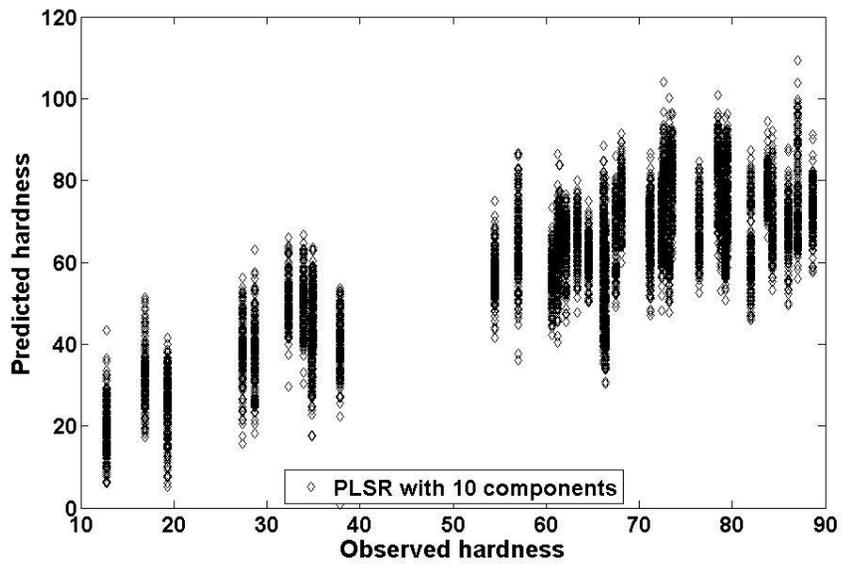


Fig. 5.36. Predicted hardness against observed hardness of wheat using the ten-factor PLSR model ($n = 7200$, $r = 0.82$, and $SECV = 12.15$).

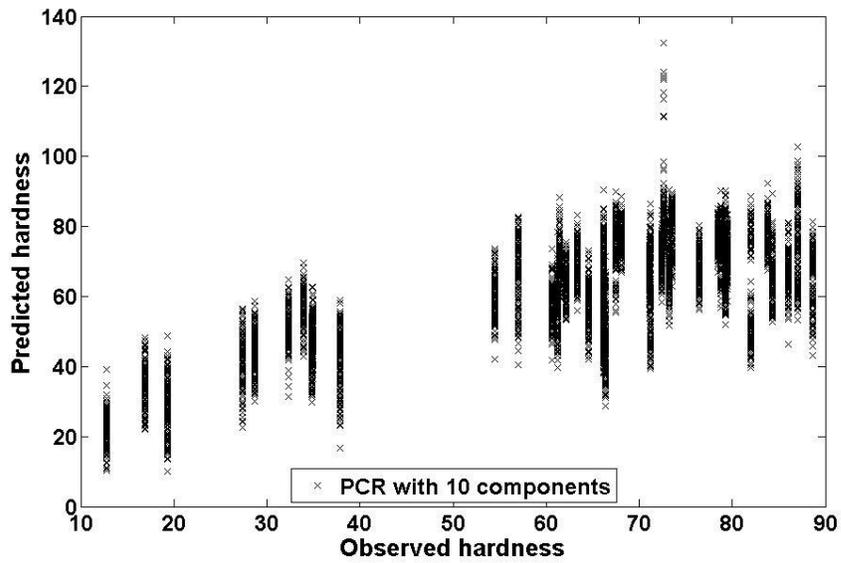


Fig. 5.37. Predicted hardness against observed hardness of wheat using the ten-factor PCR model ($n = 7200$, $r = 0.75$, and $SECV = 13.9$).

In a related study, Pomeranz et al. (1988) stated that spring wheat classes had more hardness values than winter wheat classes. For the hardness prediction, the ten-factor PLSR method gave relatively better hardness predictions than the ten-factor PCR method for wheat. With the first 10 components included, the PLSR calibration achieved an estimated MSEP of 147.7 with a SECV of 12.15 and an r of 0.82. In a similar prediction study, Lu (2001) reported that the PLSR models had r values of 0.80 and 0.65 for predicting firmness of Hedelfinger and Sam cherries, respectively. The PCR, which had the first 10 components in the model, attained 193.1, 13.9, and 0.75 for the estimated MSEP, SECV, and r, respectively.

The optimal wavelengths for predicting hardness of wheat were identified based on the values of β -coefficients and the first PC loadings using PLSR and PCR, respectively (Table 5.25).

Table 5.25. The top 10 wavelengths, in descending order, identified by the ten-factor PLSR and PCR models for predicting hardness of wheat samples.

Rank	Top 10 wavelengths for PLSR (nm)	Top 10 wavelengths for PCR (nm)
1	1390	1700
2	1220	1690
3	1180	1670
4	1320	1680
5	1480	960
6	1190	970
7	1210	980
8	1460	1030
9	1490	1020
10	1400	990

The NIR reflectance features from the wavelength regions of 1180-1220, 1320-1400, and 1460-1490 nm were critical for predicting hardness by considering β -coefficient values of the ten-factor PLSR model. Wavelengths from the regions of 960-1030 and 1670-1700 nm were identified important from the loading values of the first PC of the ten-factor PCR model. In a similar study, Maghirang and Dowell (2003) reported that the NIR wavelengths of 1100, 1200, 1380, 1450, and 1670 nm mainly contributed predicting hardness of wheat samples using the PLSR. The NIR absorptions of wheat in the wavelength regions of 960-1060, 1330-1480 nm, and at 1680 nm are likely associated to protein, starch, and hardness of samples (Maghirang and Dowell, 2003; Murray and Williams, 1987). These wavelengths can be added to a multispectral imaging system that will be used in the future for conducting on-line hardness assessments in wheat.

5.6.1.2 Class-wise PLSR and PCR prediction of hardness of wheat samples

The PLSR and PCR were used for predicting hardness of class-specific wheat samples by having the NIR reflectance intensities and reference hardness values of wheat samples as independent and dependent variables, respectively. The average NIR reflectance spectra from the LWNIR wavelength region of 960-1700 nm of class-specific wheat samples, which included 1800 samples from each of CWRS, CPSR, CWHWS, and CWSWS wheat classes, were used. The percent variations, which were explained by the components of PLSR and PCR for CWRS, CPSR, CWHWS, and CWSWS, are shown in Figs. D17, D18, D19, and D20, respectively, in the Appendix. Above 99.5% variations of input data were explained by the first three components of the PLSR and PCR for all class-specific wheat samples.

A ten-fold cross validation method was used for both PLSR and PCR methods for estimating the model performance and finding out an optimal number of components that were further used for prediction studies. In this cross-validation method, the data set was randomly split into ten groups (i.e., by having 180 samples in each group). The PLSR or PCR predictions of hardness of samples for each of the ten groups were found. The MSEP values were calculated from ten iterations that were performed in the cross-validation. The optimal number of components for the PLSR and PCR was selected. The estimated MSEP values for the PLSR and PCR components are shown in Figures D21, D22, D23, and D24 for ten-fold cross-validation for predicting hardness of CWRS, CPSR, CWHWS, and CWSWS wheat, respectively. Table 5.26 shows the estimated MSEP, SECV, and r values for the ten-factor PCR and PLSR, respectively. The ten-factor PLSR models for all class-specific samples performed better than the ten-factor PCR models. For all wheat classes, the correlation coefficients of PLSR were higher than those of PCR. The more accurate hardness predictions were expected from the CWSWS, CPSR wheat classes followed by CWHWS. High correlations with correlation coefficient values of 0.88 and 0.81 were achieved for the PLSR and PCR methods for predicting hardness of CWSWS wheat.

Table 5.26. Performance of the ten-factor PLSR and PCR models for predicting hardness using the class-specific wheat samples

Wheat class	Ten-factor PLSR model			Ten-factor PCR model		
	Estimated MSEP	SECV	r	Estimated MSEP	SECV	r
CWRS	36.4	6.03	0.70	46.8	6.83	0.57
CPSR	17.0	4.12	0.81	22.6	4.75	0.72
CWHWS	25.1	5.01	0.76	40.2	6.34	0.76
CWSWS	16.2	4.03	0.88	24.0	4.90	0.81

The PLSR method performed better than the PCR method for predicting hardness. The ten-factor PLSR method had r values of 0.7-0.88 for all class-specific samples. The correlation coefficient values of 0.57-0.81 were achieved for the ten-factor PCR method for predicting hardness. Predicted hardness against observed hardness for class-specific wheat samples for the ten-factor PLSR and the ten-factor PCR methods are shown in Figs. D25-D28 and Figs. D29-D32, respectively.

Based on the β -coefficients of PLSR and the first PC loading values of PCR, the optimal wavelengths were identified for predicting hardness of class-specific wheat samples (Table 5.27). The NIR reflectance features from the wavelength regions of 1030-1250, 1300-1480, and 1630-1670 nm were identified important using the 10-factor PLSR method. From the ten-factor PCR method, the NIR reflectance features from the wavelength regions of 960-1030 and 1510-1700 nm were ranked at the top.

Table 5.27. The top 10 wavelengths, in descending order, identified for the ten-factor PLSR and PCR models for predicting hardness of class-specific wheat samples

Rank	Top 10 wavelengths for PLSR (nm)				Top 10 wavelengths for PCR (nm)			
	CWRS	CPSR	CWHWS	CWSWS	CWRS	CPSR	CWHWS	CWSWS
1	1070	1470	1370	1090	1700	1580	1570	1700
2	1040	1360	1080	1370	1670	1590	1590	1690
3	1360	1280	1150	1390	1690	1570	1610	1680
4	1140	1440	1060	1650	1680	1600	1600	1670
5	1170	1480	1630	1380	960	1610	1560	960
6	1080	1460	1240	1400	970	1520	1580	970
7	1300	1330	1340	1230	1030	1560	1510	980
8	1640	1450	1250	1640	980	1510	1520	1660
9	1200	1130	1470	1170	1020	1550	1550	990
10	1160	1030	1430	1670	990	1530	1530	1010

Maghirang and Dowell (2003) reported that the NIR wavelengths of 1100, 1200, 1380, 1450, and 1670 nm contributed mainly in predicting hardness of wheat samples using the PLSR. The results of the protein and hardness prediction studies implied the greater potential likelihood of developing a non-destructive technique using the NIR hyperspectral imaging for measuring the intrinsic quality parameters such as protein and hardness for location- and crop year-specific wheat samples.

6. CONCLUSIONS AND RECOMMENDATIONS

This thesis research investigated the potential of using the LWNIR (960-1700 nm) hyperspectral imaging system for identifying wheat classes, both moisture-specific and moisture-non specific, using non-parametric statistical (LDA and QDA) and a four-layer BPNN classifiers for location- and crop year-specific wheat samples. Stepwise discriminant and BPNN contribution factor values were used to identify the relative importance of NIR reflectance features to the classification process for non-parametric statistical and four-layer BPNN classifiers, respectively. The PCA was used for identifying relationships between image features and wheat classes. Further research was extended to identify variations among growing locations and crop-years of wheat samples. Also, protein contents and hardness of wheat were predicted using the ten-factor PLSR and PCR methods.

A hyperspectral image database was developed for four wheat classes, which were collected from five-six different locations for two-three crop years from the three prairie provinces of Canada, each at three different moisture levels. Computer codes were written in MATLAB to extract the NIR reflectance intensities of wheat samples. Non-parametric statistical classifiers were used to identify classes using moisture non-specific wheat samples. The LDA and QDA were tested for four-class and two-class identification of wheat classes. Average classification accuracies of 80.6 and 76.3% were obtained for the LDA and QDA, respectively, for identifying moisture non-specific wheat classes. Classification accuracies were 86.5-99, 91.8-95.8, and 88.5-98.6% for identifying moisture-specific wheat classes that had 13, 16, and 19% m.c., respectively, using the LDA. In the same study, using the QDA, 84.5-100, 94-98.3, and 95.5-98.3% of

classification accuracies were achieved for discriminating wheat classes. Non-parametric classifiers, which were used for moisture-specific wheat samples, gave better results for identifying wheat classes than neural network classifiers. Hence, moisture-specific non-parametric classifiers can be used for accurate identification of wheat samples that included variations of growing locations and crop years. Using the PCA, the NIR wavelengths of 1260-1380 nm had the highest factor loading values for the first PC.

Non-parametric statistical classifiers were used for identifying moisture levels of wheat samples. Average classification accuracies of 95.2 and 90.8% were obtained for the LDA and QDA, respectively, for discriminating wheat samples into 13, 16, and 19% moisture groups. In the pair-wise moisture discrimination study, near-perfect classification (around 100%) was achieved for wheat samples which had a difference in moisture levels of about 6%. Non-parametric high-low m.c. classifier can be used for accurately discriminating wheat samples that had highly varying moisture contents.

For the pair-wise identification of wheat classes, average classification accuracies were 93.1 and 83.9% for identifying wheat classes, which included variations of moisture levels, growing locations, and crop years, for the LDA and QDA, respectively.

Classification accuracies were considerably improved for two-class identification of wheat samples that had only 13, 16, or 19% m.c. (wet basis). Non-parametric pair-wise moisture-specific classifiers can be used for accurately discriminating wheat classes that had very similar visual features. Classification accuracies were ≥ 85 and ≥ 80 % for identifying growing locations of wheat independent of crop years using the LDA and QDA, respectively. In the crop year identification study the non-parametric QDA classifier that had accuracies of 84.6% or above, performed better than the LDA. The

wavelengths, which were identified important towards classification using the STEPDISC procedure, were also listed for all classification studies.

For the neural network classification, four-layer BPNN architecture was used for pair-wise identification of wheat classes and moisture levels. Overall average classification accuracies of 83.7% were obtained for the pair-wise discrimination of wheat samples based on their moisture contents. When moisture-specific samples of 13, 16, or 19% m.c. were used, a two-class four-layer BPNN classifier had average classification accuracies of 83.2, 75.4, or 73.1%, respectively, for identifying wheat classes. Key wavelengths were identified and further listed based on contribution values of input features using the BPNN.

In the protein content prediction study, using the full data set, the ten-factor PLSR, which had 1.76%, 1.33%, and 0.68 for the estimated MSE_P, SECV, and *r*, respectively, produced better results than the ten-factor PCR model. The *r* values were 0.65, 0.68, 0.71, and 0.77, for CWRS, CPSR, CWHWS, and CWSWS wheat, respectively, for predicting protein contents using the 10-factor PLSR method for class-specific wheat samples. Based on the β -coefficient values of the ten-factor PLSR method, 960-1220, 1280-1360, and 1610-1670 nm were found important.

In the hardness prediction part, the estimated MSE_P, SECV, and *r* values were 147.7, 12.15, and 0.82, respectively, for the ten-factor PLSR model that was developed using the full data set of samples. The 10-factor PLSR models had *r* values of 0.70, 0.81, 0.76, and 0.88 for CWRS, CPSR, CWHWS, and CWSWS for predicting hardness of class-specific wheat samples. The NIR reflectance features from the wavelength regions

of 1030-1250, 1300-1480, and 1630-1670 nm were identified important using the β -coefficient values of the ten-factor PLSR models. Overall, PLSR models demonstrated better prediction performances than the PCR models for predicting protein contents and hardness of wheat. The PLSR models are thus recommended for predicting protein and hardness of western Canadian wheat classes.

In the present study, the LWNIR region of 960-1700 nm was used for acquiring hyperspectral images of wheat classes. In future, if the wavelength region is extended to 2500 nm, the complete NIR absorption of different quality parameters can be measured and further added to the models to improve classification and prediction accuracies. Also, hyperspectral imaging can be used for other major crops such as rye, barley, and canola for identification and quality-prediction purposes. Multispectral imaging systems, which use wavelengths identified by hyperspectral imaging, can be developed and evaluated for on-line quality assessments for all major crops grown in Canada.

7. REFERENCES

- Anonymous. 2011. NIR Instrumentation. Available at: http://www.glatt.com/times/times23site/tms23_fundamentals1.htm. Accessed 14 April 2011.
- Ariana, D.P., R. Lu, and D.E. Guyer. 2006. Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture* 53: 60-70.
- ASAE. 2003. Moisture measurement-unground grain and seeds. In: *ASAE Standards*. St. Joseph, MI: ASAE.
- Baker, J.E., F.E. Dowell, and J.E. Throne. 1999. Detection of parasitized rice weevils in wheat kernels with near infrared spectroscopy. *Biological Control* 16: 88-90.
- Bellucci, G. and V. Formisano. 1997. Regional mapping of planetary surfaces with imaging spectroscopy. *Journal of Optical Society of America* 12(9): 1884-1901.
- CGC. 2004. Canadian Grain Commission, Winnipeg, MB. Grain grading guide. Available at: <http://www.omicnet.com/omiccan/form/ggg2004.pdf>. Accessed 14 April 2011.
- CGC. 2011. Canadian Grain Commission, Winnipeg, MB. Quality of western Canadian wheat 2009. Available at: <http://www.grainscanada.gc.ca/wheat-ble/harvest-recolte/2009/hqww09-qrbo09-2-eng.htm>. Accessed 13 April 2011.
- Cheng, X., Y.R. Chen, Y. Tao, C.Y. Wang, M.S. Kim, and A.M. Lefcourt. 2004. A novel integrated PCA and FLD method on hyperspectral image feature extraction for

- cucumber chilling damage inspection. *Transactions of the ASAE* 47(4): 1313-1320.
- Choudhary, R., S. Mahesh, J. Paliwal, and D.S. Jayas. 2009. Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosystems Engineering* 102(2): 115-127.
- Clark, C.D., H.T. Ripley, E.P. Green, A.J. Edwards, and P.J. Mumby. 1997. Mapping and measurement of tropical coastal environment with hyperspectral and high spatial resolution data. *International Journal of Remote Sensing* 18(2): 237-242.
- Cocchi, M., C. Durante, G. Foca, A. Marchetti, L. Tassi, and A. Ulrici. 2006. Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta* 68: 1505-1511.
- Cogdill, R.P., C.R. Hurburgh, Jr., and G.R. Rippke. 2004. Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE* 47(1): 311-320.
- CWB. 2011. Canadian Wheat Board, Winnipeg, MB. Available at: http://www.cwb.ca/public/fr/library/research/popups/wheat_Quality_standards.jsp. Accessed 14 April 2011.
- Delwiche, S.R. 1995. Single wheat kernel analysis by near infrared transmittance: protein content. *Cereal Chemistry* 72(1): 11-16.
- Delwiche, S.R. 1998. Protein content of single kernels of wheat by near infrared reflectance spectroscopy. *Journal of Cereal Science* 27: 241-254.

- Delwiche, S.R. 2000. Protein content of bulk wheat from near-infrared reflectance of individual kernels. *Cereal Chemistry* 77(1): 86-88.
- Delwiche, S.R. 2003. Classification of scab and other mould damaged wheat kernels by near-infrared reflectance spectroscopy. *Transactions of ASAE* 46(3): 731-738.
- Delwiche, S.R. and D.R. Massie. 1996. Classification of wheat by visible and near-infrared reflectance from single kernels. *Cereal Chemistry* 73(3): 399-405.
- Delwiche, S.R. and K.H. Norris. 1993. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chemistry* 70(1): 29-35.
- Delwiche, S.R. and R.A. Graybosch. 2002. Identification of waxy wheat by near-infrared reflectance spectroscopy. *Journal of Cereal Science* 35: 29-38.
- Dowell, F.E. 2000. Differentiating vitreous and nonvitreous durum wheat kernels by using near-infrared spectroscopy. *Cereal Chemistry* 77(2): 155-158.
- Dowell, F.E., J.E. Throne, D. Wang, and J.E. Baker. 1999. Identifying stored-grain insects using near-infrared spectroscopy. *Journal of Economic Entomology* 92: 165-169.
- Dowell, F.E., T.C. Pearson, E.B. Maghirang, F. Xie, and D.T. Wicklow. 2002. Reflectance and transmittance spectroscopy applied to detecting fumonisin in single corn kernels infected with *Fusarium verticillioides*. *Cereal Chemistry* 79(2): 222-226.

- ElMasry, G., N. Wang, A. ElSayed, and M. Ngadi. 2007. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering* 81: 98-107.
- Famera, O., M. Hruskova, and D. Novotna. 2004. Evaluation of methods for wheat grain hardness determination. *Plant, Soil and Environment* 50(11): 489-493.
- FAOSTAT. 2011. Food and Agriculture Organization of the United Nations, Rome. Available at: <http://faostat.fao.org>. Accessed 14 April 2011.
- Fearn, T., C. Riccioli, A. Garrido-Varo, and J.E. Guerrero-Ginel. 2009. On the geometry of SNV and MSC. *Chemometrics and Intelligence Laboratory Systems* 96(1): 22-26.
- Gat, N. 2000. Imaging spectroscopy using tunable filters: A review. *Proceedings of the SPIE* 4056: 50-64.
- Geladi, P. and H. Grahn. 1996. *Multivariate Image Analysis*. Chichester, UK: John Wiley & Sons Limited.
- Geladi, P., J. Burger, and T. Lestander. 2004. Hyperspectral imaging: calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems* 72: 209-217.
- Gowen, A.A., C.P. O'Donnell, P.J. Cullen, G. Downey, and J.M. Frias. 2007. Hyperspectral imaging: an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology* 18: 590-598.

- Gowen, A.A., M. Taghizadeh, and C.P. O'Donnell. 2009. Identification of mushrooms subjected to freeze damage using hyperspectral imaging. *Journal of Food Engineering* 93: 7-12.
- Guy, R.C.E., B.G. Osborne, and P. Robert. 1996. The application of near infrared reflectance spectroscopy to measure the degree of processing in extrusion cooking processes. *Journal of Food Engineering* 27: 241-258.
- Hareland, G.A. 1994. Evaluation of flour particle size distribution by laser diffraction, sieve analysis and near infrared reflectance spectroscopy. *Journal of Cereal Science* 21: 183-190.
- Headwall. 2011. Spectral imaging capabilities of hyperspec™ imaging technology in pharmaceutical operations. Available at: http://www.headwallphotonics.com/downloads/hw_hyperspectral-in-pharma.pdf. Accessed 13 April 2011.
- Irving, D.W., R.G. Fulcher, M.M. Bean, and R.M. Saunders. 1989. Differentiation of wheat based on fluorescence, hardness, and protein. *Cereal Chemistry* 66:471-477.
- Jayas, D.S., J. Paliwal, and N.S. Visen. 2000. Multi-layer neural networks for image analysis of agricultural products. *Journal of Agricultural Engineering Research* 77(2): 119-128.
- Kim, M.S., A.M. Lefcourt, K. Chao, Y.R. Chen, I. Kim, and D.E. Chan. 2002. Multispectral detection of fecal contamination on apples based on hyperspectral imagery: Part I. Application of visible and near-infrared reflectance imaging. *Transactions of the ASAE* 45(6): 2027-2037.

- Lawrence, K.C., B. Park, W.R. Windham, and C. Mao. 2003. Calibration of pushbroom hyperspectral imaging system for agricultural inspection. *Transactions of the ASAE* 46(2): 513-521.
- Lee, K.J., S. Kang, M.S. Kim and S.H. Noh. 2005. Hyperspectral imaging for detecting defect on apples. Paper No. 053075. St. Joseph, MI: ASABE.
- Lefcourt, A.M., M.S. Kim, Y.R. Chen, and S. Kang. 2006. Systematic approach for using hyperspectral imaging data to develop multispectral imaging systems: Detection of feces on apples. *Computers and Electronics in Agriculture* 54: 22-35.
- Liu, J.J. and J.F. MacGregor. 2007. On the extraction of spectral and spatial information from images. *Chemometrics and Intelligence Laboratory Systems* 85(1): 119-130.
- Liu, Y., Y.R. Chen, M.S. Kim, D.E. Chan, and A.M. Lefcourt. 2007. Development of simple algorithms for the detection of fecal contaminants on apples from visible/near infrared hyperspectral reflectance imaging. *Journal of Food Engineering* 81: 412-418.
- Lu, R. 2001. Predicting firmness and sugar content of sweet cherries using near-infrared diffuse reflectance spectroscopy. *Transactions of the ASAE* 44(5): 1265-1271.
- Lu, R. 2003. Detection of bruises on apples using near-infrared hyperspectral imaging. *Transactions of the ASAE* 46(2): 523-530.
- Lu, R. and D. Ariana. 2002. A near-infrared sensing technique for measuring internal quality of apple fruit. *Applied Engineering in Agriculture* 18(5): 585-590.

- Lu, R. and Y. Peng. 2005. Hyperspectral scattering for assessing peach fruit firmness. *Biosystems Engineering* 93(2): 161-171.
- Lu, R. and Y.R. Chen. 1999. Hyperspectral imaging for safety inspection of food and agricultural products. *Proceedings of the SPIE* 3544: 121-133.
- MacRitchie, F. 1987. Evaluation of contributions from wheat protein fractions to dough mixing and bread making. *Journal of Cereal Science* 6(3): 259-268.
- Maghirang, E.B. and F.E. Dowell. 2003. Hardness measurement of bulk wheat by single-kernel visible and near-infrared reflectance spectroscopy. *Cereal Chemistry* 80(3): 316-322.
- Maghirang, E.B., F.E. Dowell, J.E. Baker, and J.E. Throne. 2003. Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy. *Transactions of ASAE* 46(4): 1277-1282.
- Mahesh, S., A. Manickavasagan, D.S. Jayas, J. Paliwal, and N.D.G. White. 2008. Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering* 101: 50-57.
- Mahesh, S., D.S. Jayas, J. Paliwal, and N.D.G. White. 2010a. Near-infrared (NIR) hyperspectral imaging – An emerging analytical tool for classification of western Canadian wheat classes from different locations and crop years. Paper No. MB-SK 10-302. St. Joseph, MI: ASABE.
- Mahesh, S., D.S. Jayas, J. Paliwal, and N.D.G. White. 2010b. Identification of classes and moisture contents for location-specific and crop year-specific wheat samples

using the Near-infrared (NIR) hyperspectral imaging. Paper No. MB-SK 10-404.
St. Joseph, MI: ASABE.

Mahesh, S., D.S. Jayas, J. Paliwal, and N.D.G. White. 2011. Identification of wheat classes at different moisture levels using near-infrared hyperspectral images of bulk samples. *Sensing and Instrumentation for Food Quality and Safety* 5(1): 1-9.

Majumdar, S. and D.S. Jayas. 1999. Classification of bulk samples of cereal grains using machine vision. *Journal of Agricultural Engineering Research* 73(1): 35-47.

Majumdar, S. and D.S. Jayas. 2000. Classification of cereal grains using machine vision. I. Morphology models. *Transactions of the ASAE* 43(6): 1669-1675.

Manickavasagan, A., D.S. Jayas, N.D.G. White, and J. Paliwal. 2010. Wheat class identification using thermal imaging. *Food and Bioprocess Technology* 3: 450-460.

Manickavasagan, A., G. Sathya, D.S. Jayas, and N.D.G. White. 2008. Wheat class identification using monochrome images. *Journal of Cereal Science* 47:518-527.

Martens, H. and T. Naes. 1992. *Multivariate Calibration*. Chichester, UK: John Wiley & Sons Limited.

McNeill, S., D. Overhults, and M. Montross. 2011. Harvesting, Drying and Storing Wheat. Available at: <http://www.ca.uky.edu/agc/pubs/id/id125/10.pdf>. Accessed 13 April 2011.

- Mehl, P.M., Y.R. Chen, M.S. Kim, and D.E. Chan. 2004. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering* 61: 67-81.
- Miralbes, C. 2004. Quality control in the milling industry using near infrared transmittance spectroscopy. *Food Chemistry* 88: 621-628.
- Moran, M.S., Y. Inoue, and E.M. Barnes. 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment* 61(3): 319-346.
- Murray, I. and P.C. Williams. 1987. Chemical principles of near-infrared technology. In *Near-infrared Technology in the Agricultural and Food Industries*, 17-34. St. Paul, MN: AACC.
- Nagata, M., J.G. Tallada, T. Kobayashi, and H. Toyada. 2005. NIR hyperspectral imaging for measurement of internal quality in strawberries. Paper No. 053131. St. Joseph, MI: ASABE.
- Nagata, M., J.G. Tallada, T.K. Kobayashi, Y. Cui, and Y. Gejima. 2004. Predicting maturity quality parameters of strawberries using hyperspectral imaging. Paper No. 043033. St. Joseph, MI: ASABE.
- Neuman, M., H.D. Sapirstein, E. Shwedyk, and W. Bushuk. 1987. Discrimination of wheat class and variety by digital image analysis of whole grain samples. *Journal of Cereal Science* 6(2): 125-132.

- Nicolai, B.M., E. Lotze, A. Peirs, N. Scheerlinck, and K.I. Theron. 2006. Non-destructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging. *Postharvest Biology and Technology* 40:1-6.
- Noh, H.K. and R. Lu. 2007. Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. *Postharvest Biology and Technology* 43: 193-201.
- Osborne, B.G. 2006. NIR spectroscopy in food analysis. In: *Encyclopedia of Analytical Chemistry*, 1-14. Chichester, U.K.: John Wiley & Sons Limited.
- Paliwal, J. 2002. Digital image analysis of grain samples for potential use in grain cleaning. *Unpublished Ph.D. Thesis*. Department of Biosystems Engineering, Winnipeg, MB: University of Manitoba.
- Paliwal, J., N.S. Sashidhar, and D.S. Jayas. 1999. Grain kernel identification using kernel signature. *Transactions of the ASAE* 42: 1921-1924.
- Paliwal, J., N.S. Visen, and D.S. Jayas. 2001. Evaluation of neural network architectures for cereal grain classification using morphological features. *Journal of Agricultural Engineering Research* 79(4): 361-370.
- Park, B., K.C. Lawrence, W.R. Windham and D.P. Smith. 2005. Detection of cecal contaminants in visceral cavity of broiler carcasses using hyperspectral imaging. *Applied Engineering in Agriculture* 21(4): 627-635.
- Park, B., K.C. Lawrence, W.R. Windham, and R.J. Buhr. 2002. Hyperspectral imaging for detecting fecal and ingesta contaminants on poultry carcasses. *Transactions of the ASAE* 45(6): 2017-2026.

- Pearson, T.C., D.T. Wicklow, E.B. Maghirang, F. Xie, and F.E. Dowell. 2001. Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. *Transactions of ASAE* 44(5): 1247-1254.
- Peirs, A., N. Scheerlinck, J.D. Baerdemaeker, and B.M. Nicolai. 2003. Starch index determination of apple fruit by means of a hyperspectral near infrared reflectance imaging system. *Journal of Near Infrared Spectroscopy* 11: 379-389.
- Perez-Mendoza, J., J.E. Throne, F.E. Dowell, and J.E. Baker. 2003. Detection of insect fragments in wheat flour by near-infrared spectroscopy. *Journal of Stored Products Research* 39: 305-312.
- Petterson, H. and L. Aberg. 2003. Near infrared spectroscopy for determination of mycotoxins in cereals. *Food Control* 14: 229-232.
- Pierna, J.A.F., V. Baeten, and P. Dardenne. 2006. Screening of compound feeds using NIR hyperspectral data. *Chemometrics and Intelligent Laboratory Systems* 84: 114-118.
- Polder, G., G.W.A.M. van der Heijden, and I.T. Young. 2002. Spectral image analysis for measuring ripeness of tomatoes. *Transactions of the ASAE* 45(4): 1155-1161.
- Polder, G., G.W.A.M. van der Heijden, L.C.P. Keizer, and I.T. Young. 2003. Calibration and characterization of imaging spectrographs. *Journal of Near Infrared Spectroscopy* 11: 193-210.

- Pomeranz, Y., Z. Czuchajowska, M.D. Shogren, G.L. Rubenthaler, L.C. Bolte, H.C. Jeffers, and P.J. Mattern. 1988. Hardness and functional (bread and cookie-making) properties of U.S. wheats. *Cereal Foods World* 33: 297-304.
- Putnam, D.F. and R.G. Putnam. 1970. *Canada: A Regional Analysis*. Toronto, ON: J.M. Dent and Sons, Inc.
- Refaeilzadeh, P., L. Tang, and H. Liu. 2011. Cross-validation. Available at: <http://www.public.asu.edu/~ltang9/papers/ency-cross-validation.pdf>. Accessed 24 April 2011.
- Resmini, R.G., M.E. Kappus, W.S. Aldrich, J.C. Harsanyi, and M. Anderson. 1997. Mineral mapping with hyperspectral digital imagery collection experiment (HYDICE) sensor data at Cuprite, Nevada, U.S.A. *International Journal of Remote Sensing* 18(7): 1553-1570.
- Ruan, R., Y. Li, X. Lin, and P. Chen. 2002. Non-destructive determination of deoxynivalenol levels in barley using near-infrared spectroscopy. *Applied Engineering in Agriculture* 18(5): 549-553.
- Schmitter, B.M. and T. Rihs. 1989. Evaluation of macrocombustion method for total nitrogen determination in feedstuffs. *Journal of Agricultural and Food Chemistry* 37(4): 992-994.
- Singh, C.B. 2009. Detection of insect and fungal damage and incidence of sprouting in stored wheat using near-infrared hyperspectral and digital colour imaging.

Unpublished Ph.D. Thesis. Department of Biosystems Engineering, Winnipeg, MB: University of Manitoba.

Singh, C.B., D.S. Jayas, J. Paliwal, and N.D.G. White. 2007. Fungal detection in wheat using near-infrared hyperspectral imaging. *Transactions of the ASABE* 50(6): 2171-2176.

Singh, C.B., D.S. Jayas, J. Paliwal, and N.D.G. White. 2009. Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. *Journal of Stored Products Research* 45: 151-158.

Singh, C.B., D.S. Jayas, J. Paliwal, and N.D.G. White. 2010. Detection of midge-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. *Biosystems Engineering* 105: 380-387.

Sivakumar, M. 2007. Determination of main constituents in wheat using near-infrared hyperspectral imaging. *Unpublished M.Sc. Thesis.* Department of Biosystems Engineering, Winnipeg, MB: University of Manitoba.

Slaughter, D.C, K.H. Norris, and W.R. Hruschka. 1992. Quality and classification of hard red wheat. *Cereal Chemistry* 69(4): 428-432.

Song, H., S.R. Delwiche, and Y.R. Chen. 1995. Neural network classification of wheat using single kernel near-infrared transmittance spectra. *Optical Engineering* 34(10): 2927-2934.

Tatzer, P., M. Wolf, and T. Panner. 2005. Industrial application for inline material sorting using hyperspectral imaging in the NIR range. *Real-Time Imaging* 11: 99-107.

- Toews, M.D., J. Perez-Mendoza, J.E. Throne, F.E. Dowell, E. Maghirang, F.H. Arthur, and J.F. Cambell. 2007. Rapid assessment of insect fragments in flour milled from wheat infested with known densities of immature and adult *Sitophilus oryzae* (Coleoptera: Curculionidae). *Journal of Economic Entomology* 100: 1704-1723.
- Visen, N.S., J. Paliwal, D.S. Jayas, and N.D.G. White. 2004. Image analysis of bulk grain samples using neural networks. *Canadian Biosystems Engineering* 46: 7.11-7.15.
- Vogt, F., S. Banerji, and K. Booksh. 2005. Utilizing three-dimensional wavelet transforms for accelerated evaluation of hyperspectral image cubes. *Journal of Chemometrics* 18(7-8): 350-362.
- Wang, D., F.E. Dowell, and D.S. Chung. 2001. Assessment of heat-damaged wheat kernels using near-infrared spectroscopy. Paper No. 01-6006. St. Joseph, MI: ASABE.
- Wang, D., F.E. Dowell, and R. Dempster. 2002. Determining vitreous subclasses of hard red spring wheat using visible/near-infrared spectroscopy. *Cereal Chemistry* 79(3): 418-422.
- Wang, D., F.E. Dowell, and R.E. Lacey. 1999. Single wheat kernel colour classification using neural networks. *Transactions of the ASAE* 42: 233-240.
- Wang, D., F.E. Dowell, M.S. Ram, and W.T. Schapaugh. 2004. Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. *International Journal of Food Properties* 7: 75-82.

- Wang, D., F.E. Dowell, M.S. Ram, and W.T. Schapaugh. 2004a. Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. *International Journal of Food Properties* 7: 75-82.
- Wang, W., J. Paliwal, and D.S. Jayas. 2004b. Determination of moisture content of ground wheat using near-infrared spectroscopy. Paper No. MB04-200. St. Joseph, MI: ASABE.
- Wang, D., M.S. Ram and F.E. Dowell. 2002. Classification of damaged soybean seeds using near-infrared spectroscopy. *Transactions of ASAE* 45(6): 1943-1948.
- Watson, C.A., W.C. Shuey, O.J. Banasik, and J.W. Dick. 1977. Effect of wheat class on near infrared reflectance. *Cereal Chemistry* 54(6): 1264-1269.
- Wesley, I.J., O. Larroque, B.G. Osborne, N. Azudin, H. Allen, and J.H. Skerritt. 2001. Measurement of Gliadin and Glutenin content of flour by NIR spectroscopy. *Journal of Cereal Science* 31: 125-133.
- Williams, P., P. Geladi, G. Fox, and M. Manley. 2009. Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta* 653: 121-130.
- Williams, P.C. 1979. Screening wheat for protein and hardness by near infrared reflectance spectroscopy. *Cereal Chemistry* 56(3): 169-172.
- Xing, J., C. Bravo, P.T. Jancsok, H. Ramon, and J.D. Baerdemaeker. 2005. Detecting bruises on 'Golden Delicious' apples using hyperspectral imaging with multiple wavebands. *Biosystems Engineering* 90(1): 27-36.

- Xing, J., P.V. Hung, S. Symons, M. Shahin, and D. Hatcher. 2009. Using a short wavelength infrared (SWIR) hyperspectral imaging system to predict alpha amylase activity in individual Canadian western wheat kernels. *Sensing and Instrumentation for Food Quality and Safety* 3: 211-218.
- Zayas, I.Y., C.R. Martin, J.L. Steele, and A. Katsevich. 1996. Wheat classification using image analysis and crush-force parameters. *Transactions of the ASAE* 39(6): 2199-2204.
- Zhang, H., J. Paliwal, D.S. Jayas, and N.D.G. White. 2007. Classification of fungal infected wheat kernels using near-infrared reflectance hyperspectral imaging and support vector machine. *Transactions of the ASABE* 50(5): 1779-1785.

APPENDIX

Graphs of protein content and hardness prediction studies in wheat

Protein content prediction graphs in wheat

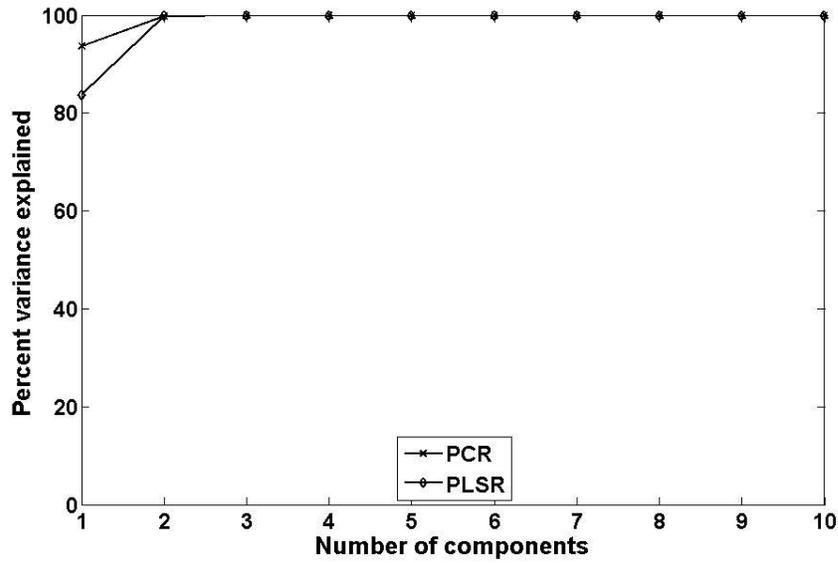


Fig. D.1. Percent variance explained by the components of PLSR and PCR for CWRS wheat.

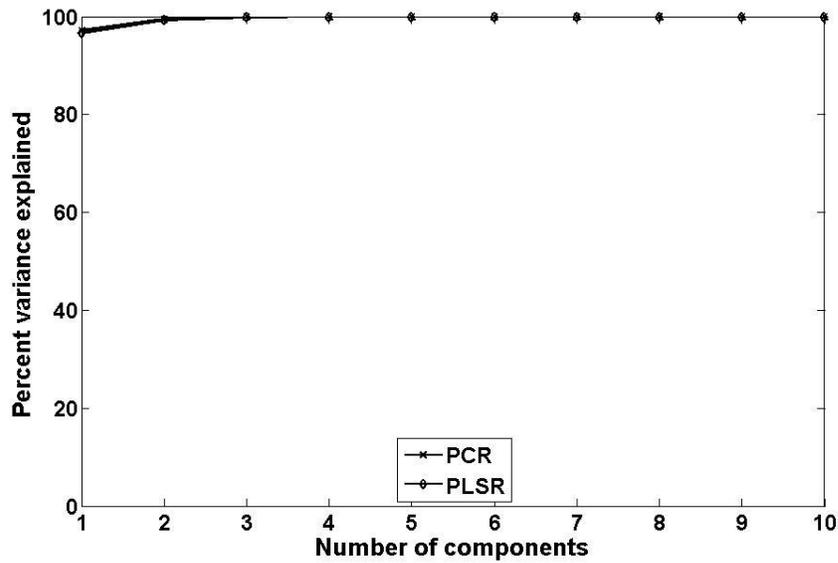


Fig. D.2. Percent variance explained by the components of PLSR and PCR for CPSR wheat.

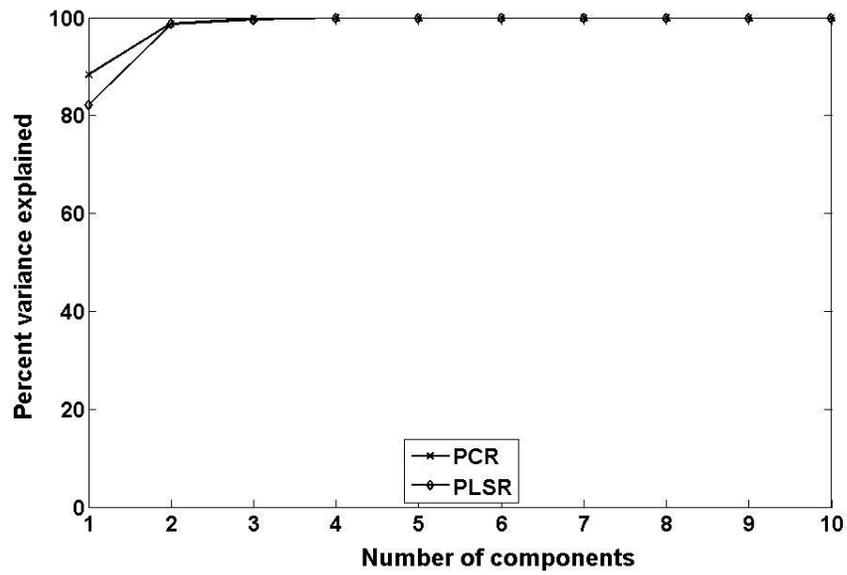


Fig. D.3. Percent variance explained by the components of PLSR and PCR for CWHWS wheat.

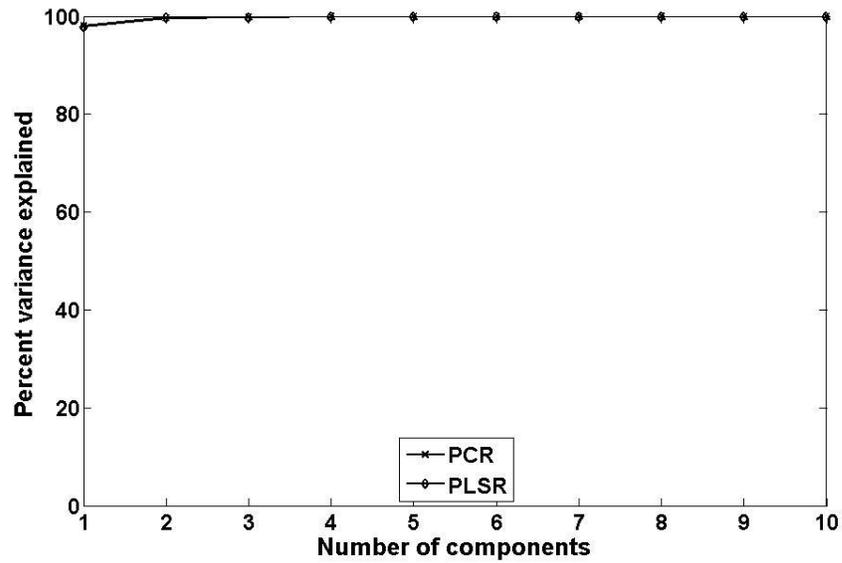


Fig. D.4. Percent variance explained by the components of PLSR and PCR for CWSWS wheat.

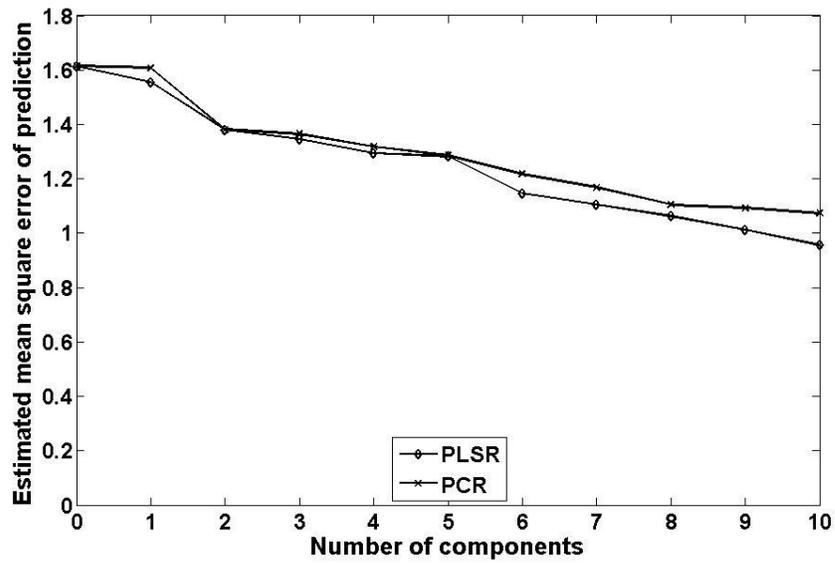


Fig. D.5. Estimated MSEP values for the components of PLSR and PCR models for predicting protein contents of CWRS wheat.

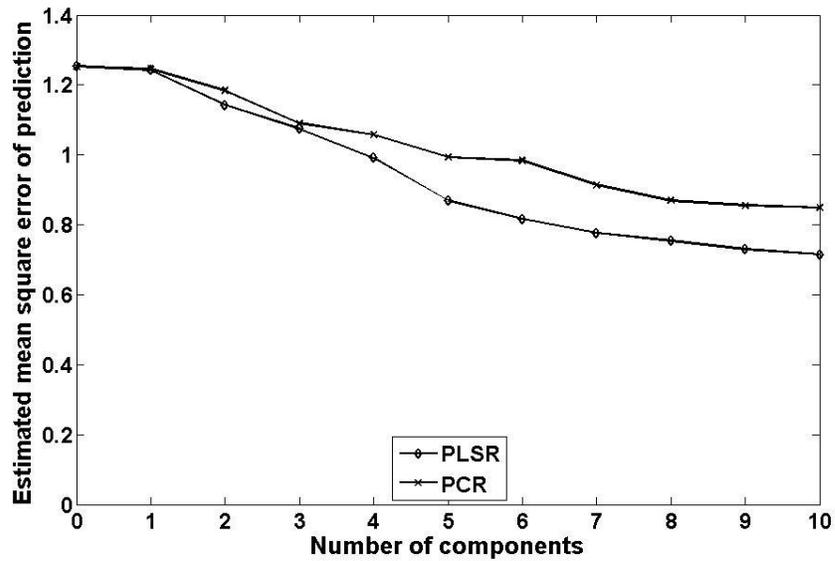


Fig. D.6. Estimated MSEP values for the components of PLSR and PCR models for predicting protein contents of CPSR wheat.

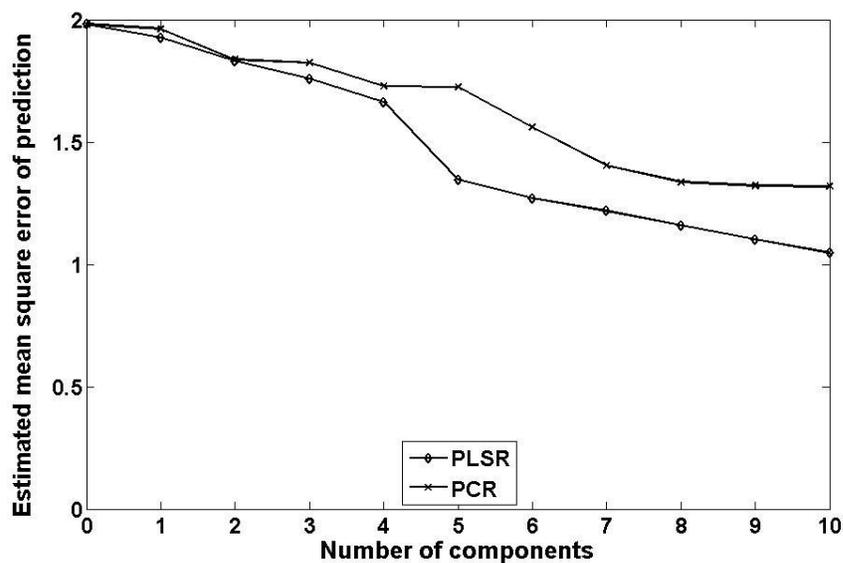


Fig. D.7. Estimated MSE values for the components of PLSR and PCR models for predicting protein contents of CWHWS wheat.

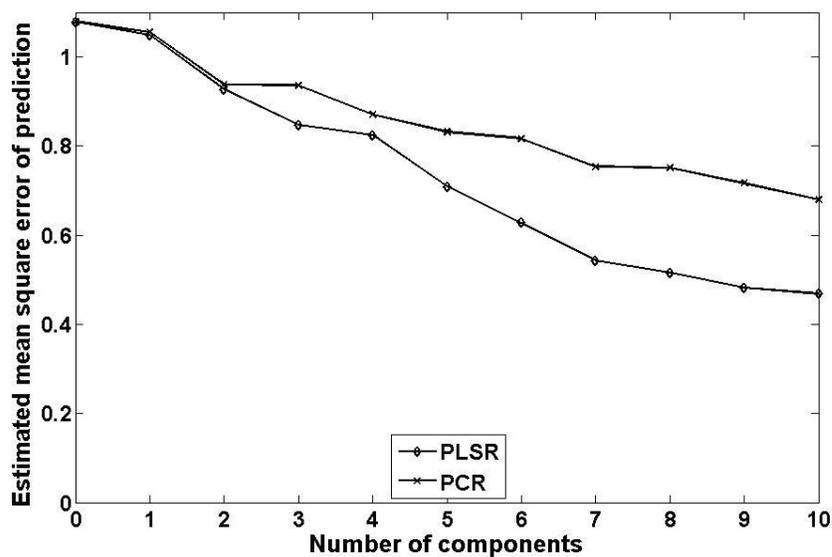


Fig. D.8. Estimated MSE values for the components of PLSR and PCR models for predicting protein contents of CWSWS wheat.

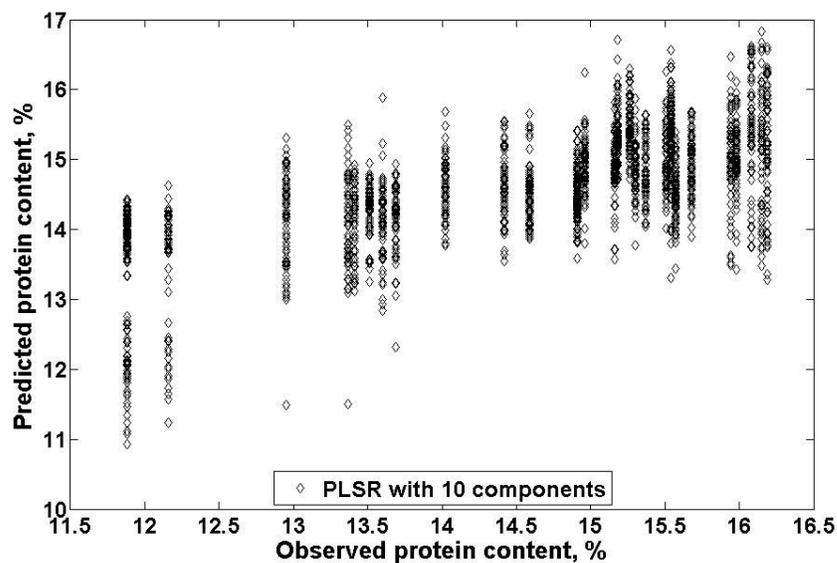


Fig. D.9. Predicted protein contents against observed protein contents of CWRS wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.65$, and $SECV = 0.98$).

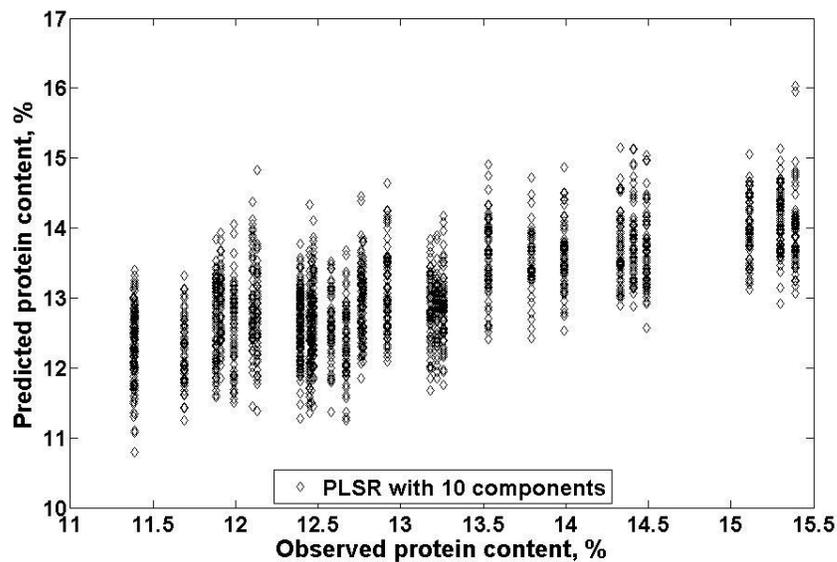


Fig. D.10. Predicted protein contents against observed protein contents of CPSR wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.68$, and $SECV = 0.84$).

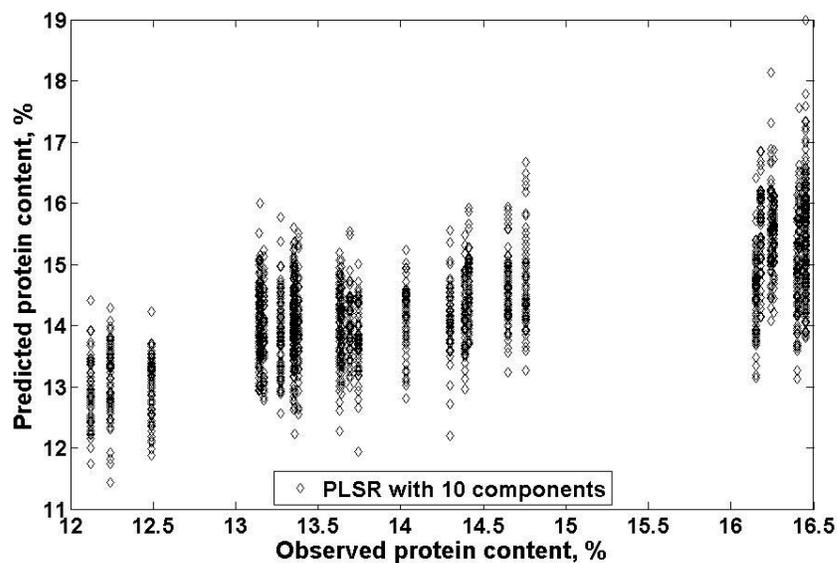


Fig. D.11. Predicted protein contents against observed protein contents of CWHWS wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.71$, and $SECV = 1.02$).

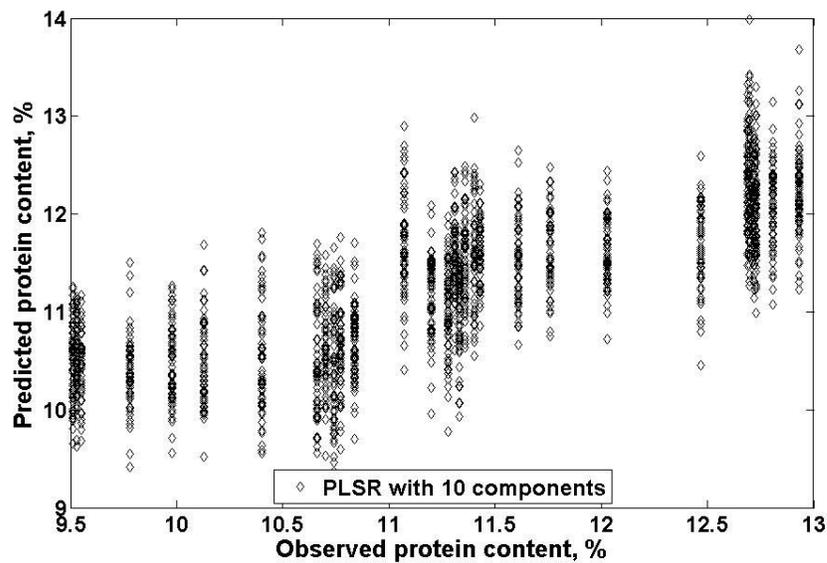


Fig. D.12. Predicted protein contents against observed protein contents of CWSWS wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.71$, and $SECV = 1.02$).

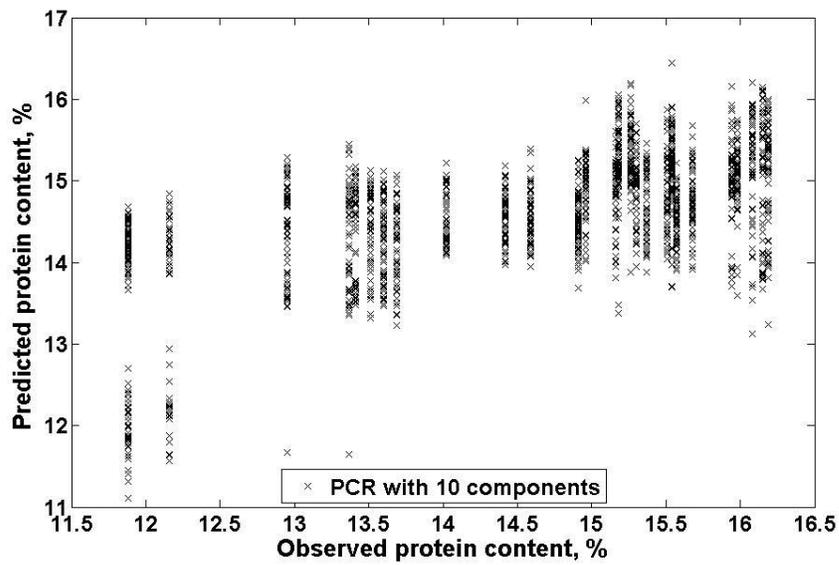


Fig. D.13. Predicted protein contents against observed protein contents of CWRS wheat using the ten-factor PCR model ($n = 1800$, $r = 0.59$, and $SECV = 1.04$).

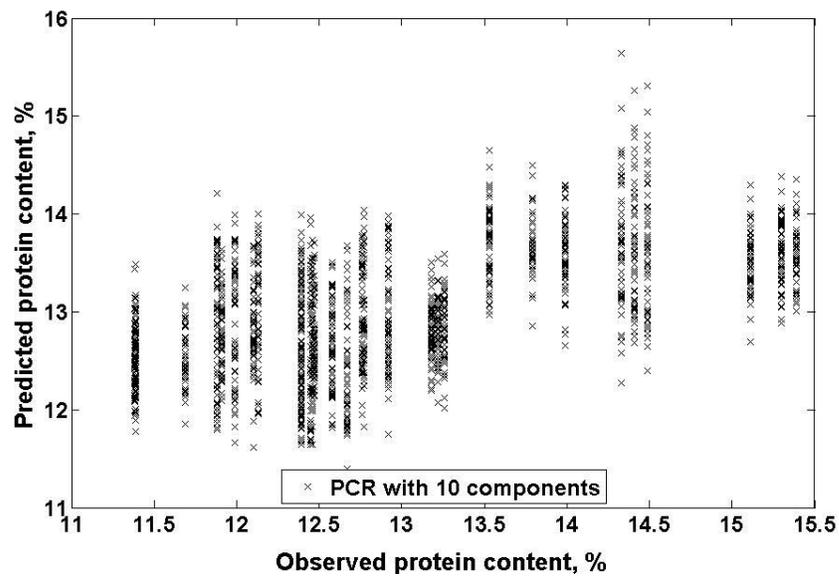


Fig. D.14. Predicted protein contents against observed protein contents of CPSR wheat using the ten-factor PCR model ($n = 1800$, $r = 0.57$, and $SECV = 0.921$).

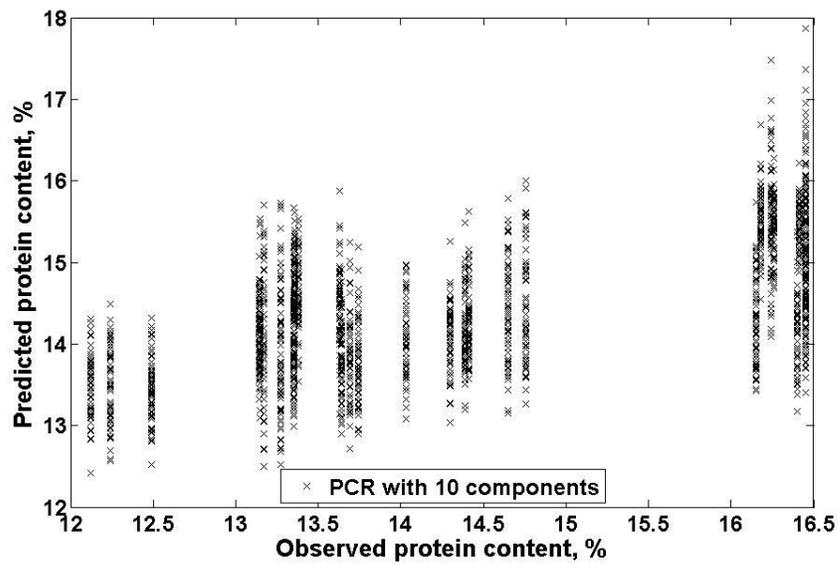


Fig. D.15. Predicted protein contents against observed protein contents of CWHWS wheat using the ten-factor PCR model ($n = 1800$, $r = 0.59$, and $SECV = 1.15$).

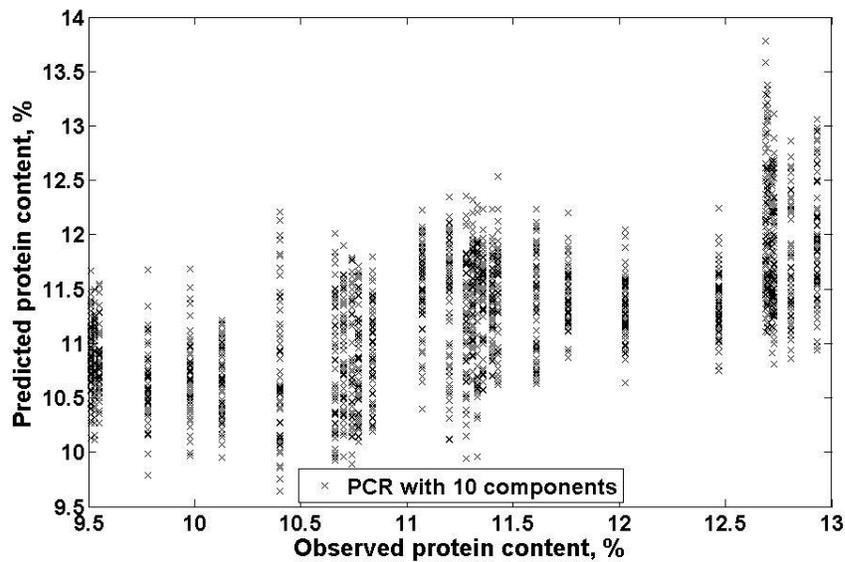


Fig. D.16. Predicted protein contents against observed protein contents of CWSWS wheat using the ten-factor PCR model ($n = 1800$, $r = 0.61$, and $SECV = 0.83$).

Hardness prediction graphs for wheat

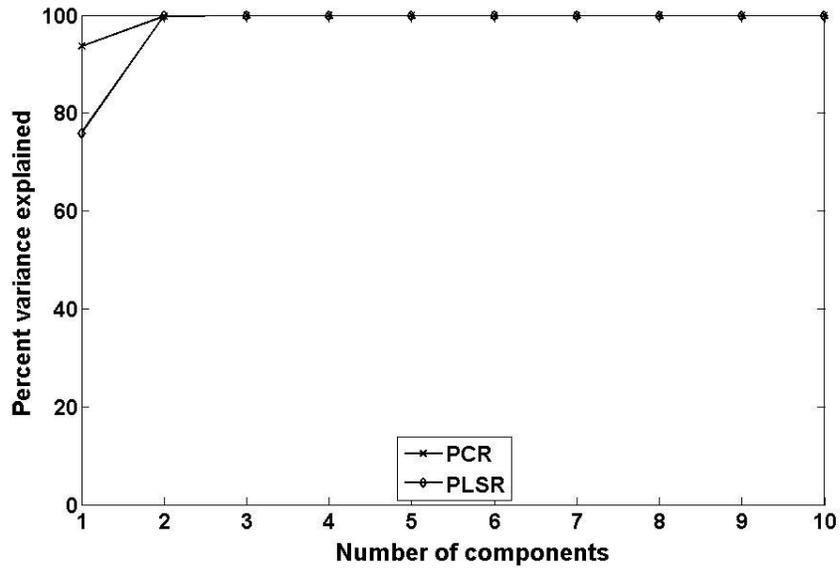


Fig. D.17. Percent variance explained by the components of PLSR and PCR for CWRS wheat.

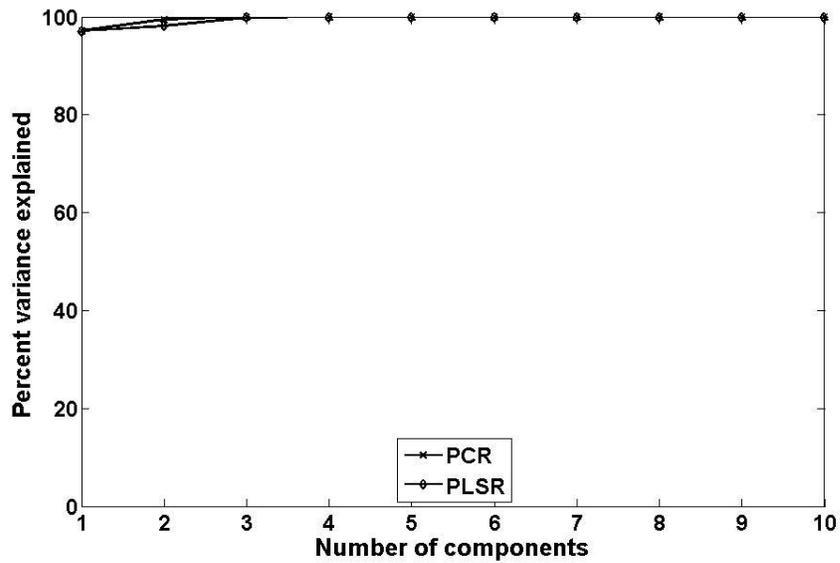


Fig. D.18. Percent variance explained by the components of PLSR and PCR for CPSR wheat.

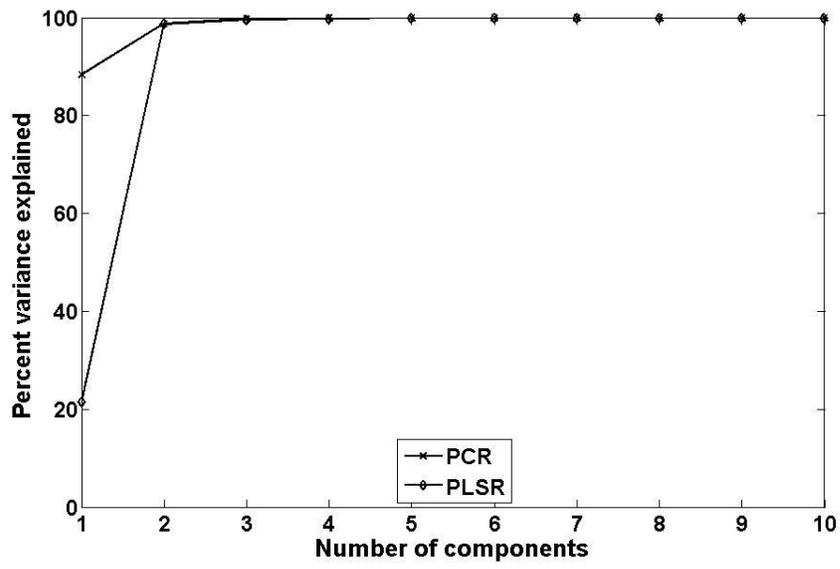


Fig. D.19. Percent variance explained by the components of PLSR and PCR for CWHWS wheat.

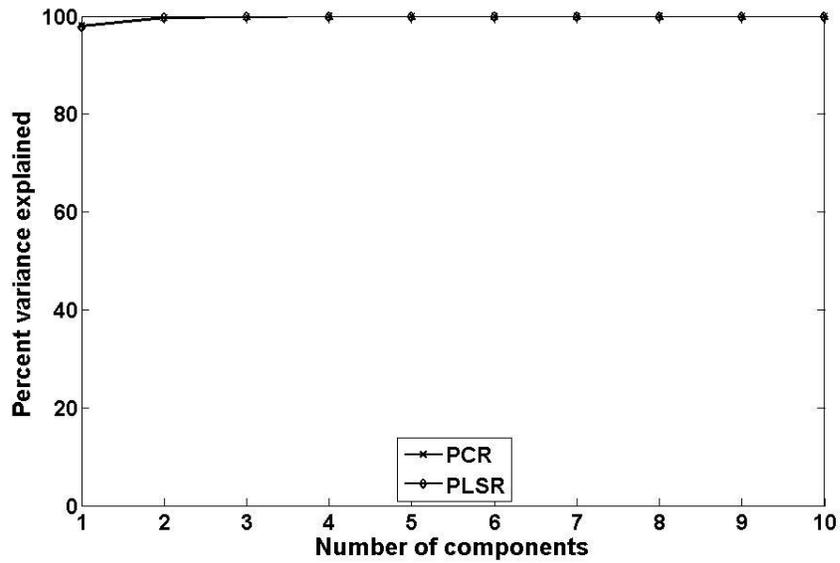


Fig. D.20. Percent variance explained by the components of PLSR and PCR for CWSWS wheat.

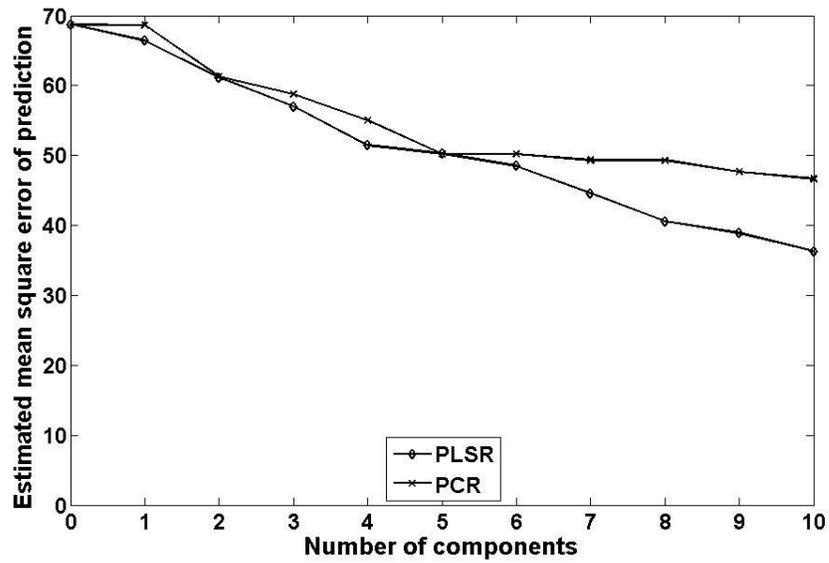


Fig. D.21. Estimated MSE values for the components of PLSR and PCR models for predicting hardness of CWRs wheat.

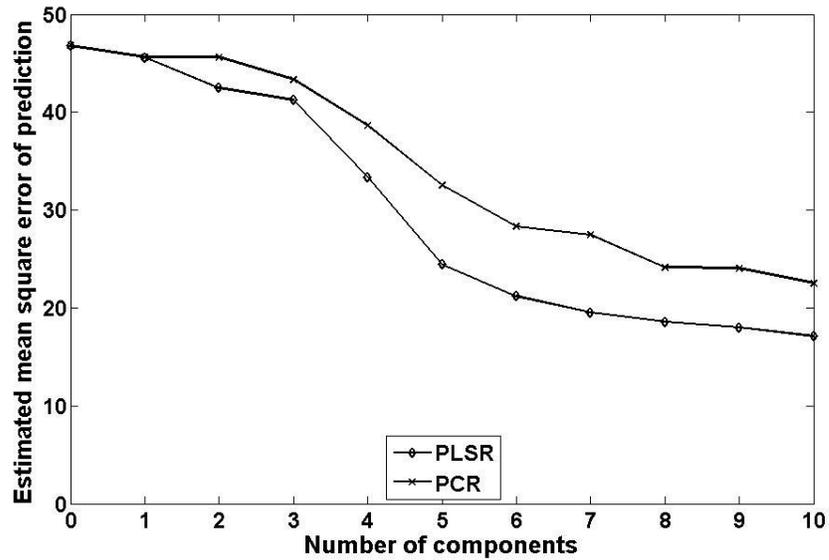


Fig. D.22. Estimated MSE values for the components of PLSR and PCR models for predicting hardness of CPSR wheat.

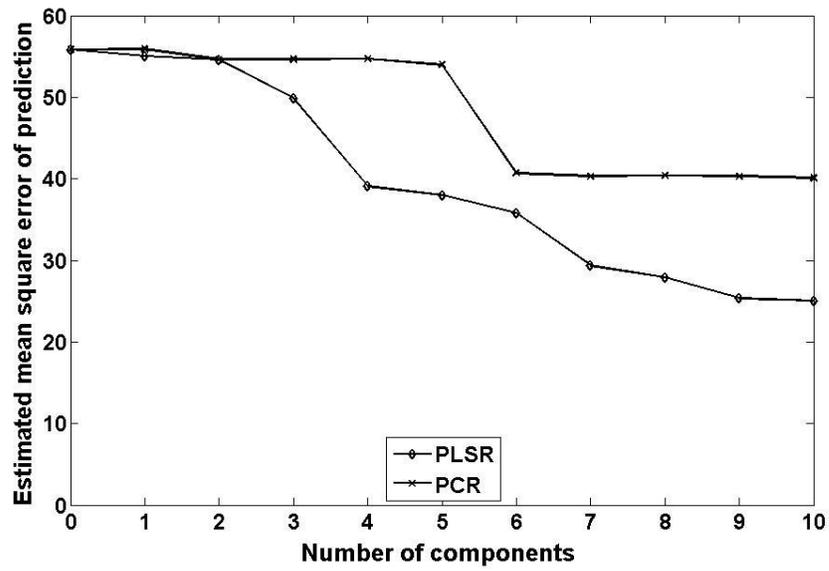


Fig. D.23. Estimated MSE values for the components of PLSR and PCR models for predicting hardness of CWHWS wheat.

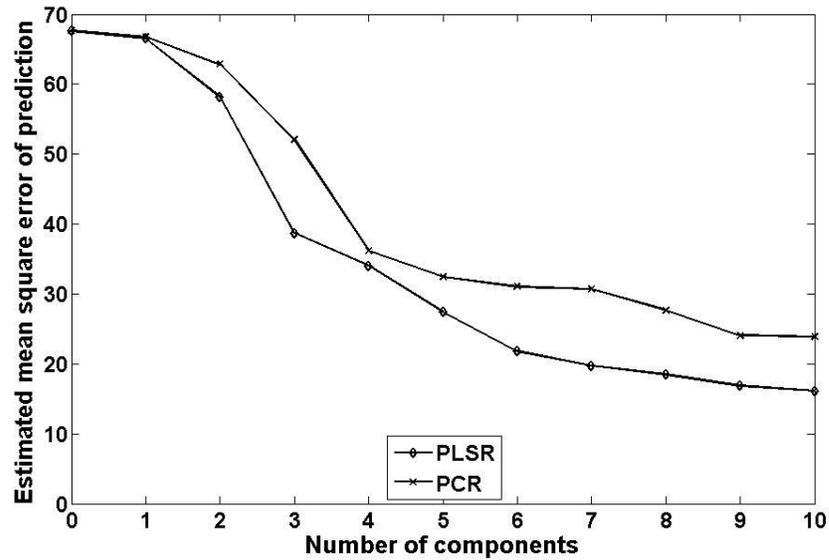


Fig. D.24. Estimated MSE values for the components of PLSR and PCR models for predicting hardness of CWSWS wheat.

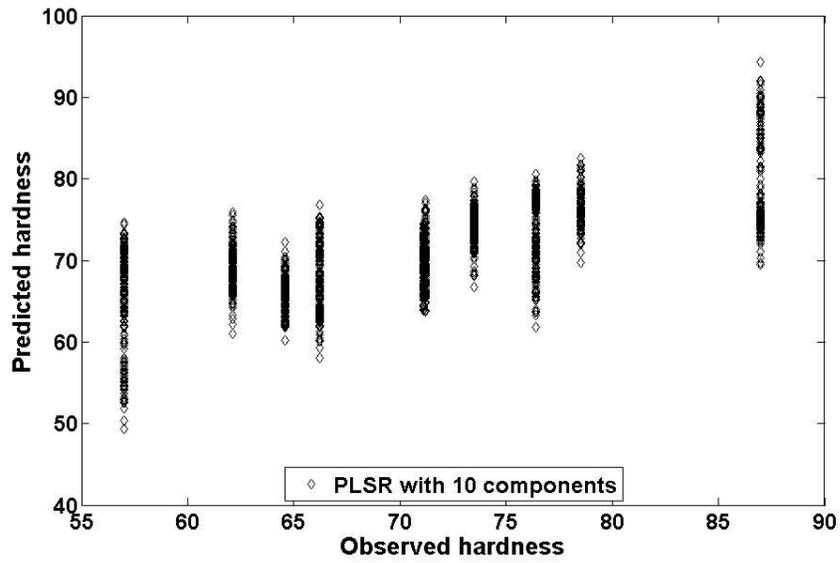


Fig. D.25. Predicted hardness against observed hardness of CWRS wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.70$, and $SECV = 6.03$).

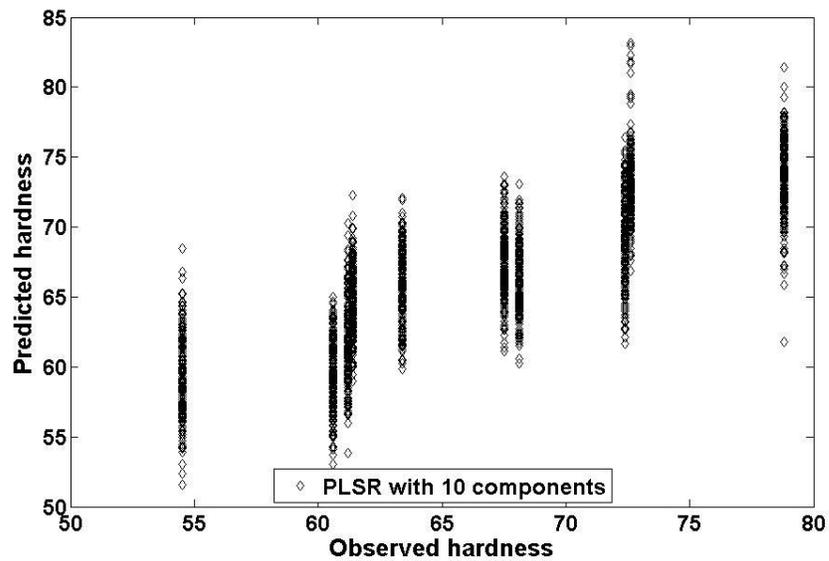


Fig. D.26. Predicted hardness against observed hardness of CPSR wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.81$, and $SECV = 4.12$).

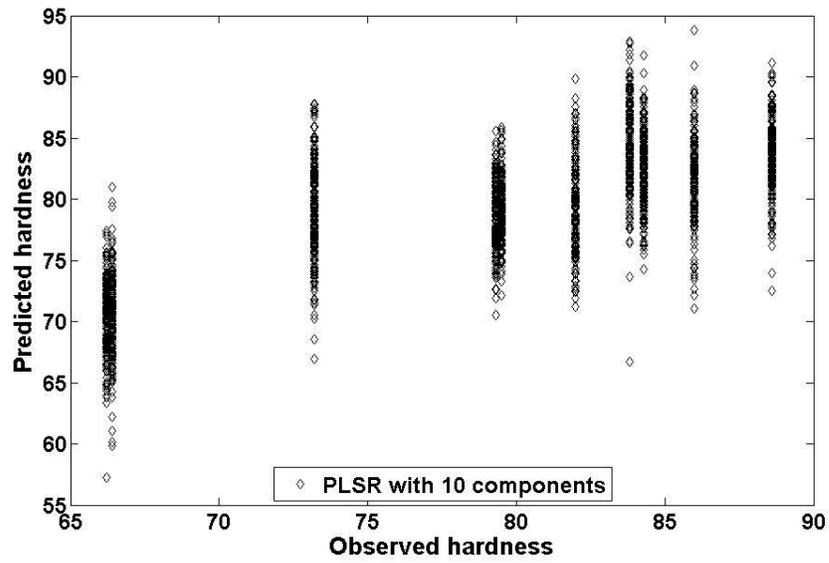


Fig. D.27. Predicted hardness against observed hardness of CWHWS wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.76$, and $SECV = 5.01$).

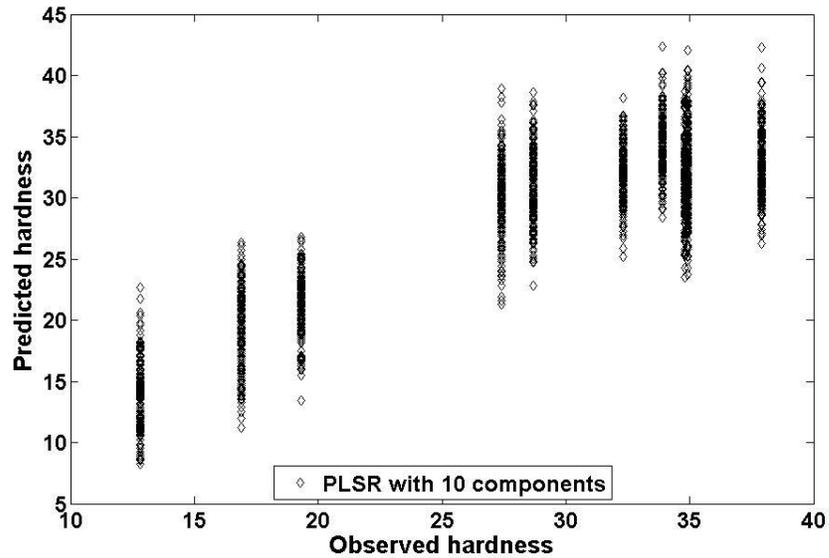


Fig. D.28. Predicted hardness against observed hardness of CWSWS wheat using the ten-factor PLSR model ($n = 1800$, $r = 0.88$, and $SECV = 4.03$).

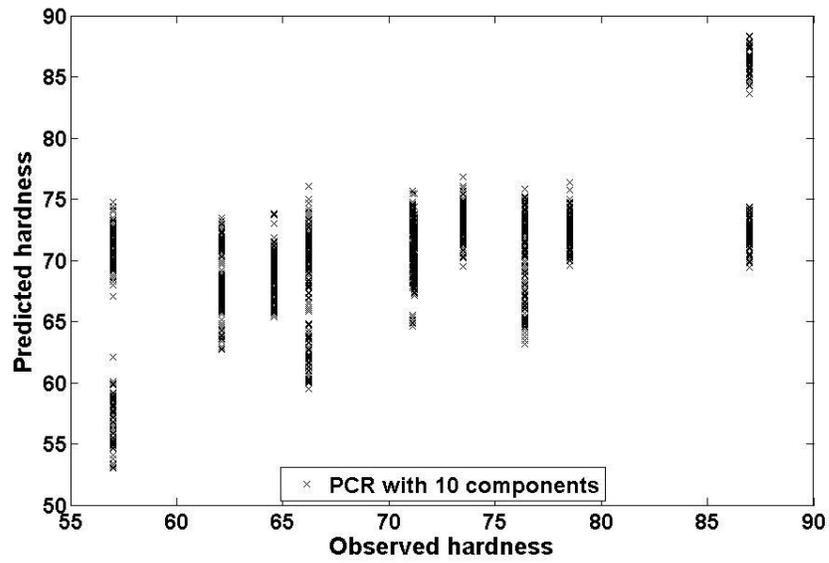


Fig. D.29. Predicted hardness against observed hardness of CWRS wheat using the ten-factor PCR model ($n = 1800$, $r = 0.57$, and $SECV = 6.83$).

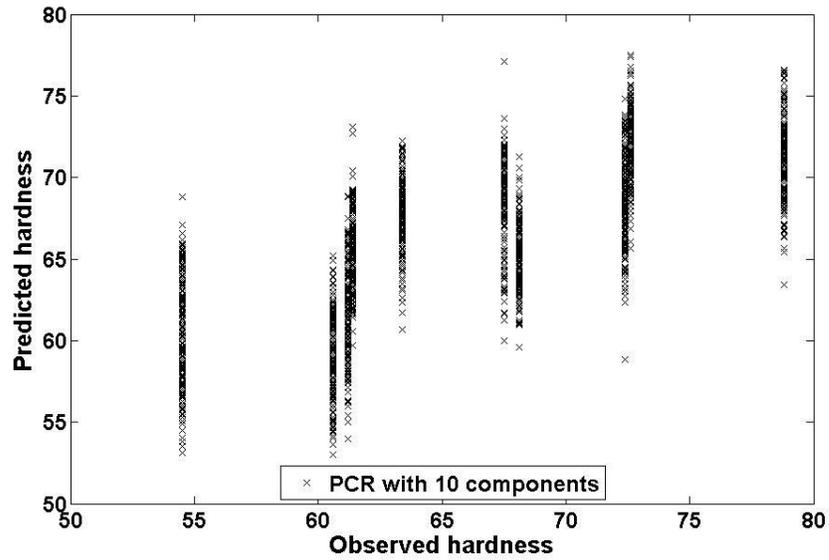


Fig. D.30. Predicted hardness against observed hardness of CPSR wheat using the ten-factor PCR model ($n = 1800$, $r = 0.72$, and $SECV = 4.75$).

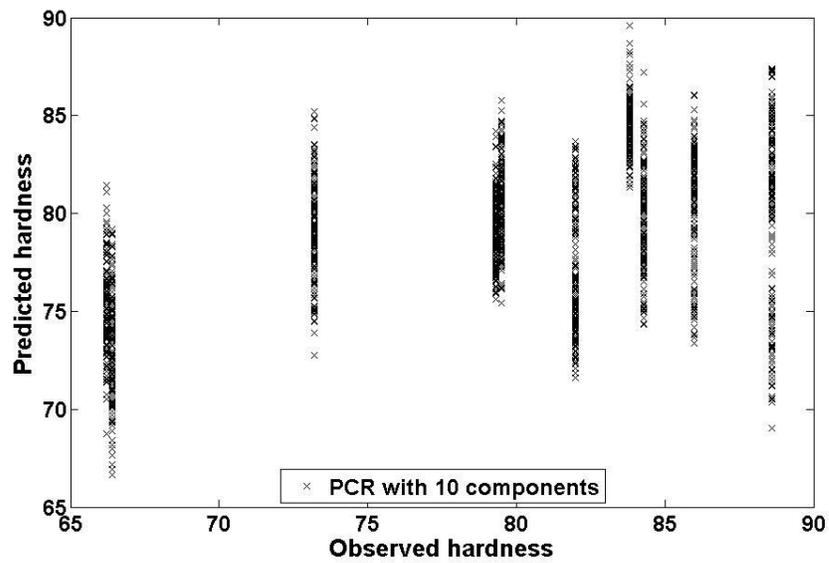


Fig. D.31. Predicted hardness against observed hardness of CWHWS wheat using the ten-factor PCR model ($n = 1800$, $r = 0.76$, and $SECV = 6.34$).

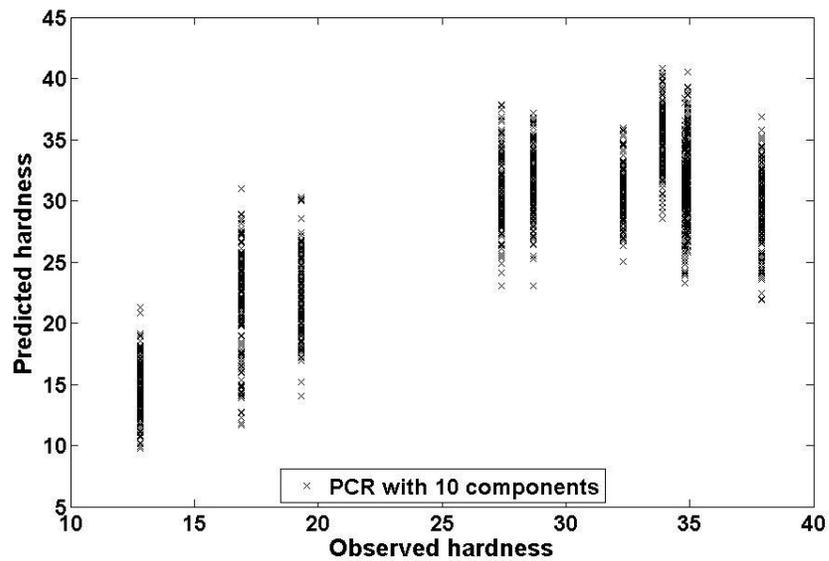


Fig. D.32. Predicted hardness against observed hardness of CWSWS wheat using the ten-factor PCR model ($n = 1800$, $r = 0.81$, and $SECV = 4.90$).