NIR IMAGING AND ITS APPLICATION TO WHEAT GRADING

NIR IMAGING AND ITS APPLICATION TO WHEAT GRADING

By

ZHENG LIU, B.ENG., M. ENG.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree

Master of Applied Science

McMaster University

Master of Applied Science (2006) McMaster University

(Chemical Engineering) Hamilton, Ontario

TITLE: NIR Imaging and its Application to Wheat Grading

AUTHOR: Zheng Liu, B. Eng. (Tianjin University), M.Eng. (Tianjin University)

SUPERVISOR: Professor John F. MacGregor and Dr. Honglu Yu

NUMBER OF PAGES: xiv, 96

Abstract

Two topics related to near infrared (NIR) imaging technology are studied in this thesis. The first is on the calibration of line-scan NIR imaging systems, the second covers the feasibility of applying the NIR imaging technology for wheat grading.

In the first study, a methodology is proposed to pretreat the NIR image data acquired by the line-scan NIR imaging system used in this thesis to reduce the systematic noise introduced by the imaging system. This calibration in a standardization methodology is shown to improve the result of multivariate image analysis (MIA) based on multi-way principal component analysis (MPCA). This method represents a practical and easily used tool for calibration of line-scan NIR imaging systems in that it does not employ expensive standard reflectance material.

In the second study, two projects are accomplished. In the first project, NIR imaging is used to classify different classes of wheat kernels. Multivariate statistical algorithms, soft independent modeling of class analogy (SIMCA) and partial least square discriminant analysis (PLS-DA) are used to discriminate between different types of wheat kernels using spectral features from NIR images. A new strategy of implementing multi-class PLS-DA algorithm is proposed in this part. The results from this study show that NIR imaging provides a potentially fast and objective method for qualitatively evaluating certain characteristics of wheat samples, such as fungal infection, sprout damage and foreign types of grain, which are now graded manually in wheat industry. In the second project, NIR imaging is used to predict the "falling number" (FN) of wheat samples. Three models are built between the features extracted from NIR images of the wheat kernels and the falling number measurements made on bulk samples. One uses a regular PLS algorithm, one uses the orthogonal partial least square (O-PLS) algorithm and the other uses the PLS plus canonical correlation analysis (PLS+CCA) algorithm. The models are analyzed and the performance of the algorithms is discussed. The errors in the

prediction of the O-PLS model are investigated. The results from this study indicate that NIR imaging is a promising method for the rapid assessment of the FN of wheat samples.

Acknowledgements

For their contributions to this thesis, I would like to thank:

my supervisor, Dr. J. F. MacGregor, for his insight, encouragement and support throughout my thesis work;

my co-supervisor, Dr. Honglu Yu, for her stimulating suggestions and great help;

the Department of Chemical Engineering and Mcmaster Advanced Control Consortium for the financial support;

Dupont Canada for providing samples and cooperation;

Dr. Hatcher at Canadian Grain Commission for the useful suggestion;

my colleagues and friends in the Penthouse for sharing their expertise and for discussing those things not yet at that stage. In particular I would like to thank Kevin and David for the suggestion and great help during writing the thesis;

my parents and parents-in-law for their endless love and encouragement;

my wife, Li whose patience, love and unconditional support enabled me to complete this work.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Near infrared (NIR) spectroscopy represents a fast and non-destructive measurement technology that can predict certain properties of samples using a reference model, built in advance, between the NIR spectrum and the properties which are measured by other means. Its potential as an efficient analytical tool was recognized by Norris [1965], and NIR technology has since been one of the fastest growing analytical technologies with overwhelming application in virtually all fields of science [Tigabu, 2003]. Especially since 1990's, the availability of efficient chemometric calibration method, light-fiber optics coupled with specific probes, and miniaturization has launched NIR spectroscopy into a new era for industrial quality and process control [Siesler et al., 2002].

In recent years, NIR imaging technology has made its way from remote sensing into the laboratory and industry. NIR imaging has enabled people to obtain spatial and spectral information that characterize samples with unprecedented ease, speed and with both good spatial and spectral resolution. Its potential as an efficient sensor for process monitoring and quality control has been realized by industry and has been increasingly used in practice, for example, in the pharmaceutical industry.

Two topics related to NIR imaging technology are addressed in this thesis. One is related to the equipment itself, that is calibrating line-can NIR imaging systems as used in this thesis; the other is on the feasibility of using this technology for wheat grading.

It is important to standardize the NIR imaging instrument before using it to acquire NIR images, since for any given spectrometer equipment, accurate results from the calibration model require accurate spectra. All the literature published about

calibrating NIR imaging spectrometers use standard reflectance materials which are expensive and difficult to maintain in practice. It is therefore meaningful to develop a practical and economical method to calibrate the imaging system without the use of such standards. This is the initiative for the first study in this thesis.

The wheat industry is an important industry in the prairie provinces of Canada. Canada is known for its superior quality wheat in the global market due to its stringent wheat inspection and grading process. Canadian wheat is currently qualitatively graded manually by its visual characteristics, which is slow, labor intensive and subjective. Intelligent machine vision systems (MVS) have been developed to take over the tedious work and provide fast and objective grading. However, the current MVS are based on Red-Blue-Green (RGB) color images, which are reported as being unable to accurately segregate certain classes of wheat, the sprouted kernels for example. Some important wheat quality indices related to the quality of final wheat products, such as wheat falling number (FN), are often required by wheat customers. The FN index is currently measured using a traditional wet chemistry method, the Hagberg falling number test, which is slow and not easily implemented in grain elevators. Therefore, there exists a need to develop rapid and objective methods for quantitatively evaluating wheat quality. Most variation of wheat quality is related to chemical variation in the kernels. Based on this fact, MVS based on NIR imaging should be promising for grading wheat kernels. The feasibility of such an instrument is studied in this thesis.

The organization of the thesis is as follows:

Chapter 2 provides background knowledge on NIR spectroscopy technology. After a brief introduction to the NIR spectrum, the instrumentation, especially the NIR spectrometer used in this thesis is described. Multivariate calibration based on NIR spectral data is reviewed in the last section. Multivariate statistics based chemometric methods PLS, SIMCA, PLS-DA, which form the backbone of the methods used in the following chapters, are introduced.

Chapter 3 addresses the problem of calibrating NIR line-scan imaging systems. A simple methodology without using expensive standard reflectance materials is proposed. An example is used to show the benefit of the methodology to improve the result of multivariate image analysis (MIA) based on multi-way principal component analysis (MPCA).

Chapter 4 and Chapter 5 are related to applying NIR imaging technology for wheat grading.

In Chapter 4, NIR imaging is used to classify different classes of wheat kernels. Multivariate statistical algorithms, SIMCA and PLS-DA are used to discriminate sprouted wheat kernels from the healthy wheat. Multi-class PLS-DA is used to classify four classes of grain. A new strategy of implementing the multi-class PLS-DA algorithm, namely a *one-vs-one strategy* is proposed in this section.

In Chapter 5, the NIR images of wheat kernels are used to predict the FNs of wheat samples. Three models, regular PLS, O-PLS and PLS+CCA, are built between the features extracted from NIR images of the wheat kernels and the falling number measurements made on bulk samples. The interpretability of the O-PLS and PLS+CCA is discussed. The errors in the prediction of the models are also analyzed.

Finally, Chapter 6 summarizes the results of this thesis, draws some conclusions and highlights topics for future work.

Chapter 2

Review of NIR Technology

This chapter provides background knowledge on NIR technology. After a brief introduction of the principle of NIR spectroscopy, the instrumentation, especially the NIR spectrometer used in this thesis is described. Multivariate calibration procedures based on NIR spectral data are reviewed in the last section. Multivariate statistics based chemometric methods PLS, SIMCA, PLS-DA which form the backbone of the methods used in the following chapters are introduced.

2.1 Principle of NIR Spectrum

The NIR spectrum is commonly defined as the region of light with wavelengths from 780 nm to 2500 nm. The NIR spectrum originates from radiation energy transferred to mechanical energy associated with the motion of atoms held together by chemical bonds in a molecule. When a molecule absorbs radiation, vibrations in the bonds occur either due to stretching or bending. Stretching is vibration in which there is a continuous change in the interatomic distance along the axis of the bond between the two atoms while vibration involving a change in bond angle is referred to as bending and deformation (Figure 2.1). The molecular bonds vibrate in a manner similar to a diatomic oscillator that can be explained using the quantum-mechanical model. According to the quantum selection rules, the only allowed vibrational transitions are those in which υ (the quantum number) changes by one ($\Delta\upsilon = \pm 1$). The harmonic oscillator model, thus, explains the absorption bands observed in the infrared (IR) region (2500 nm to 5000 nm) due to fundamental modes of molecular vibration. However, real molecules do not behave exactly as predicted by the law of simple harmonic motion and real bonds do not strictly obey Hook's law due to Coulombic repulsion between the two nuclei and dissociation of bonds beyond the limit of elasticity that levels off the potential energy (Figure 2.2). Consequently, the harmonic criterion is not fulfilled at higher vibrational states, and

vibrations become rather anharmonic. Such anharmonic molecular vibrations allow energy transitions between more than one level, and thus creating overtone bands which provide the basis for the NIR spectrum (Figure 2.3). If two or three separate anharmonic vibrations absorb one part of each of the incident radiation, this type of absorption gives rise to combination bands in the NIR spectrum. So, absorption bands in the NIR spectra of chemical compounds can be observed as a consequence of overtones and combination of molecular vibrations.

The main bands typically observed in the NIR region correspond to bonds containing light atoms such as X–H, where X is carbon, nitrogen, oxygen or sulfur, and H is hydrogen that, in turn, are the major molecular moieties in virtually all organic materials. This is because the hydrogen atom is the lightest, and therefore exhibits the largest vibrations and the greatest deviations from harmonic behavior. Other important functionalities in the NIR region include C=O, C–C, and C–Cl stretching vibrations, although the bands are much weaker [Shenk et al., 2001]. Figure 2.4 illustrated a NIR absorptions chart, which indicates the major analytical bands and their relative peak positions in the NIR spectrum. Furthermore, it also illustrates the NIR absorption ranges of the above-mentioned groups.

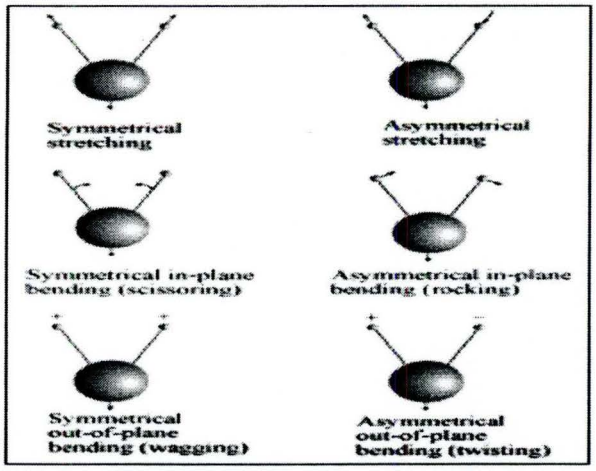


Figure 2.1. Modes of bond vibration for a hypothetical molecule AX_2

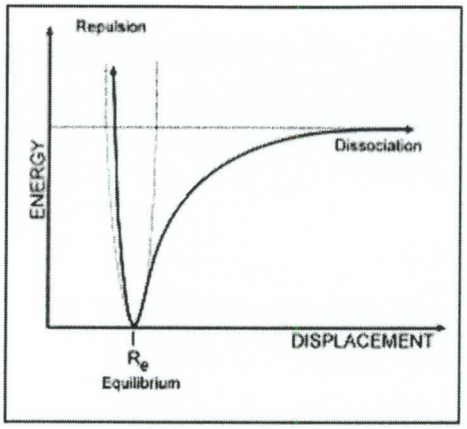


Figure 2.2. The energy of a diatomic molecule undergoing harmonic oscillation (dashed line) and anharmonic vibration (solid line) that explains absorption in the NIR region

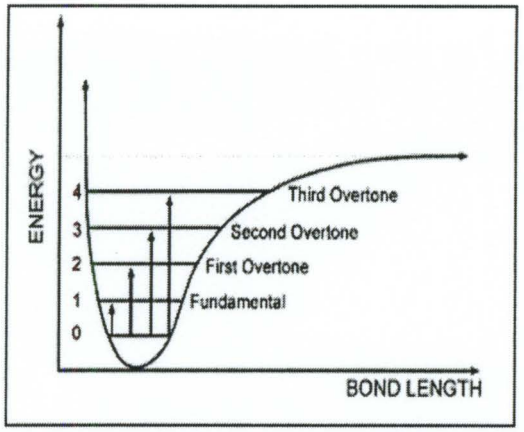


Figure 2.3. Energy transition levels creating overtone bands in the NIR region. The fundamental absorptions are the basis of IR spectroscopy

Extracting useful information about the chemical content of organic compounds from their NIR spectrum using mathematical methods was first realized by Karl Norris [Norris and Hart, 1965], who is recognized as "father" of modern NIR technology. Since the bands of overtones and combinations for many functional groups in the NIR spectrum of an organic compound overlap, they give a smooth spectrum with broad peaks, which makes the NIR spectrum more difficult to interpret as compared to its mid-IR spectrum. With the advent of digital technology NIR spectral readings are generally digitized into many hundreds of narrow wavelength bands. Since the NIR spectrum is smooth these wavelength bands are highly correlated with each other. Multivariate statistics based chemometric techniques like PCA and PLS have excelled in efficiently extracting useful information and using this information to empirically model many sample properties. The field of multivariate calibration [Martens et al.1989] develops the theory and application of such chemometric techniques in spectral data.

Compared with other analytical methods, NIR spectroscopy does have many advantages. First, there is hardly any need for sample preparation thus allowing the

technique to be applicable to a sample in any physical or chemical state. Second, NIR measurements are gathered rapidly in a non-invasive manner, thus allowing the sample to be re-used after being measured, and sent on for further analysis. The combination of these characteristics with improvements in instrumentation and the development of chemometric software has been making the NIR technology one of the fast growing analytical technologies in the world with an overwhelming application in virtually all fields of science [Tigabu, 2003].

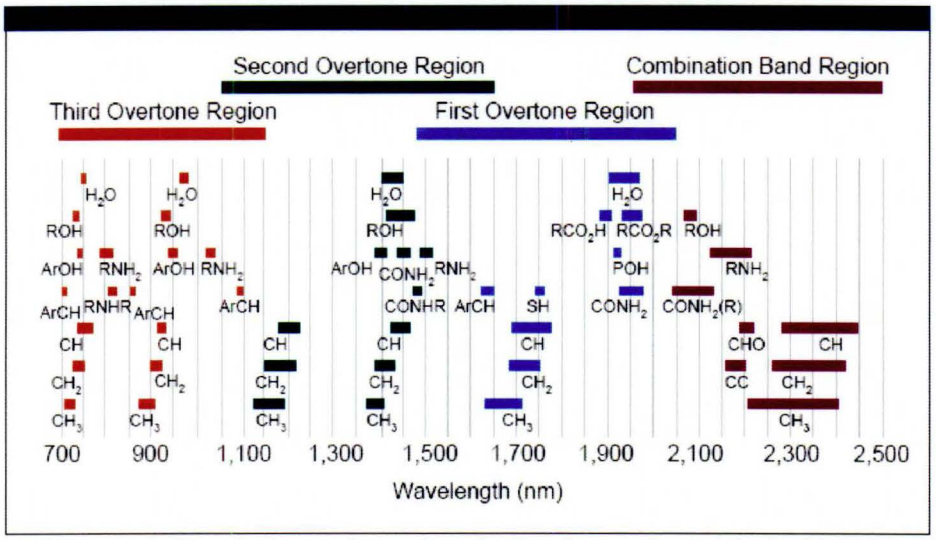


Figure 2.4. Near-Infrared absorptions chart indicating major analytical bands and their peak positions [Bharati, 2002]

2.2 Instrumentation for NIR Spectroscopy

2.2.1 General structure of NIR spectrometers

Sample information of the near-infrared region is usually collected as an absorption spectrum through transmission measurements or diffuse-reflectance measurements with a NIR spectrometer. For multi-channel spectrometers, the basic instrumental configuration includes radiation source, wavelength selector/modulator, detector and output relay. (Fig. 2.5)

Tungsten-halogen lamps with quartz envelopes are the major energy sources for NIR instruments. These lamps provide high-energy output (10–200W) over the 360-3000 nm region and last longer due to a bathing effect of the halide inside the lamp. Light emitting diodes (LED), laser diodes and lasers are non-thermal or 'cold sources', in which most of the energy consumed appears as emitted radiation over a narrow range of wavelengths. As the emitting wavelengths are predetermined, instruments based on such devices are usually dedicated for specific analysis, such as determination of moisture in samples.

Radiation emitted from a source can be spectrally separated into individual wavelengths using different optical principles, namely, dispersive, interferometric and non-thermal [Siesler et al., 2002]. The spectroscope is a dispersive system where wavelengths of light are separated spatially. Prisms were the classic dispersing elements for many years. However, prism is an inefficient arrangement with low and non-linear dispersion, and a large prism is often needed to achieve better performance. As a result, most scanning spectrometers used in laboratories and in industries today employ diffraction gratings.

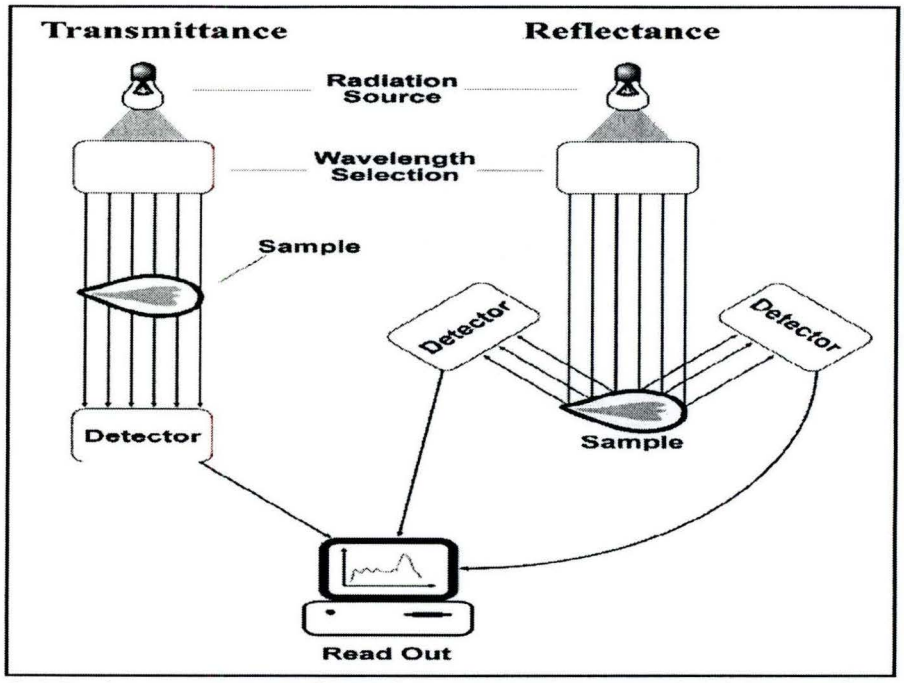


Figure 2.5 Basic components of NIR instrumentation operating in transmittance and reflectance modes

Another dispersive device incorporated into NIR spectrometers in recent years is Acousto-optically tunable filters (AOTF). AOTF choose wavelengths by using radio-frequency signals to change the refractive index of a crystal made of TeO2 (tellurium dioxide) in such a way that it transmits light of a given wavelength region or scans the whole spectral range.

The second major optical principle used for wavelength selection in NIR spectroscopy is interferometry. This method, referred to as non-dispersive, does not cause angular dispersion, but instead uses a filter, often known as a interferometer for wavelength differentiation. Among family of interferometric systems is the Michelson interferometer; the Fabry-Perot interferometer and Fourier transform NIR instruments. For more detail about interferometric systems refer to Osborne et al. [1993] and McClure [1994].

Radiation transmitted through or reflected from a sample is detected by semiconductor detectors. Lead sulfide (PbS) is the most widely used detector in the NIR over the range of 1100-2500 nm while silicon sensors are used for the 360-1000 nm range [McClure 1994]. In multi-channel system covering visible-NIR region (400-2500 nm), PbS detectors sandwiched with silicon photodiodes are often used to acquire spectral information over many wavelengths simultaneously. Another detector is a device composed of Indium gallium arsenide (InGaAs) which is sensitive in the wavelength range of 900nm to 1800 nm. This kind of detector is used in the spectrometer for this thesis. More recently, the IR camera market has seen the emergence of the uncooled microbolometer array using low-cost CMOS technologies [Lewis, 2004]. These detectors promise to set whole new performance levels for infrared focal plane arrays. Depending on the shape of the detector, the spectrometer can be categorized into traditional probe-based multi-spectral spectrometer and multi-spectral imaging spectrometer. If the wavelength selection device disperses the light spatially, for the former one, the detector is a *vector-shape* array detector and a full spectrum of the sample is obtained; for the latter one, the detector is a rectangular *matrix-shape* area array detector, where a spatial dimension is added and a multi-spectral image is formed on it.

Finally, computers are an indispensable part of NIR instrumentation for capturing spectral data as well as for process monitoring and analysis of spectral data.

2.2.2 Details of the spectrometer used in the thesis

The NIR spectrometer used in this thesis is a line-scan multi-spectral reflectance imaging spectrometer modified from a NIR digital camera. A direct imaging spectroscope [Hyvariann et al., 1998] was attached between the front optics lens and the camera back, which is an InGaAs charge-coupled devices (CCD) area array. The spectroscope consists of an entrance split, focusing lenses, and a Prism-Grating-Prism (PGP) element encased in a hollow tube. Light enters the spectroscope in a horizontal line through the entrance slit and gets vertically dispersed into its continuous spectral distribution as it goes through the lenses and PGP element. This results in an array of wavelength-specific horizontal

lines of light that are captured by the CCD area array detector in the camera back as a 2-dimensional intensity images. The horizontal axis (i.e. columns) of the captured image represents the spatial dimension, whereas the spectral dimension is represented by the vertical axis (i.e. rows). Figure 2.6 illustrates the basic operating principle of the direct-sight imaging spectroscope [ImSpector, 2003].

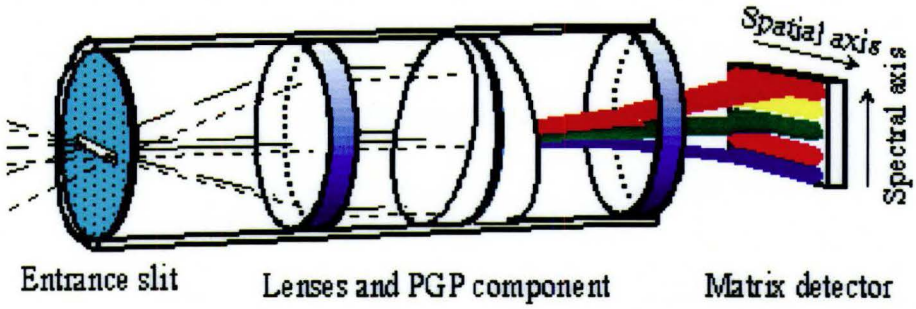


Figure 2.6 Schematic of direct-sight imaging spectroscope used to convert an area array camera into an imaging spectrometer

The pixel resolution of the CCD array in the NIR camera is 128×128 pixels. Thus the spectrally dispersed image captured by the InGaAs CCD array has dimensions of 128 rows and 128 columns. As a result, the continuous NIR reflectance spectrum (900 nm to 1700 nm) of 128 spatial pixels is vertically digitized into 128 discrete wavelength bands increasing from bottom to the top. Each band of the digitized spectrum (represented by a row of the 2-dimensional image) has a spectral resolution of approximately 6.25 nm.

Thus, for each imaged line of a moving object the system records a spatial-spectral (i.e. $x vs. \lambda$) intensity image. To capture the multi-spectral NIR image of an object, the second spatial dimension is obtained by recording multiple lines across a moving object at constant velocity in a perpendicular direction to the scan. In this thesis, this is realized with a scanner assembly. The speed of this scanner bed is controlled through a desktop computer. The $x - \lambda$ images recorded by the imaging spectrometer per line scan are joined into a 3-dimensional multi-spectral image data cube; the 3rd dimension of which represents the other spatial dimension (y). Figure 2.7 illustrate the

working principle of the imaging spectrometer to capture the *x-y* spatial dimension of a sample on a moving web.

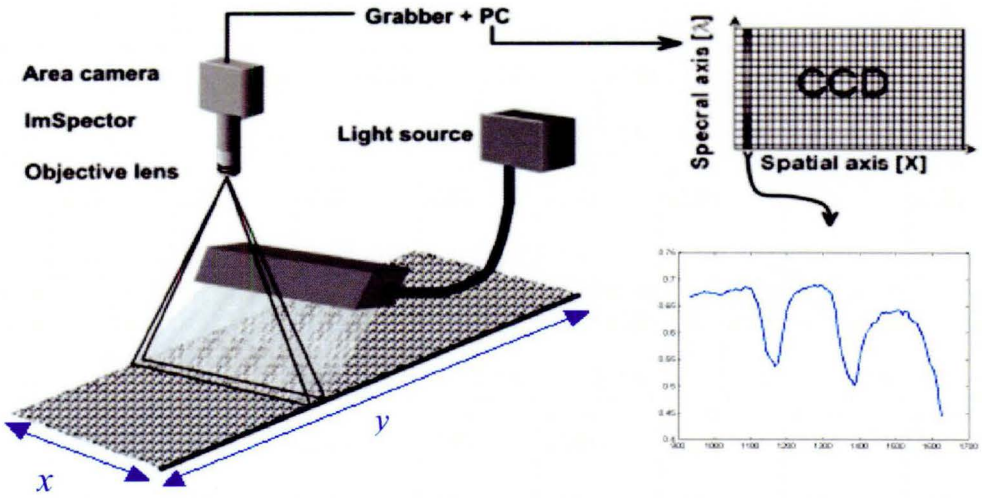


Figure 2.7 Schematic picture of the line scan NIR imaging spectrometer [ImSpector, 2003]

The resulting 3-dimensional dataset is a multi-spectral NIR reflectance image with 2 spatial dimensions and 1 specral dimension ($y \times x \times \lambda$). The number of lines scanned across the moving object can control the pixel resolution of the *y* dimension. Furthermore, the physical length of the scanned section of an object can be controlled through the number of lines scanned, whereas the distance between the object and the imaging spectrometer can control the width of the scanned section, and the resolution in this direction.

As far as the radiation source is concerned, in Chapter 3, a halogen lighting source (150W) attached with fiber-optic cables arranged in a horizontal line has been used to illuminate the object at a 45° angle with respect to the scanner bed. Typically image acquisition requires a well-illuminated object with even lighting from at least two sources at opposite 45° angles to remove any shadows cast by the object. However in Chapter 3 the objects being imaged are flat in nature, thus only a single source of lighting was deemed adequate. In chapter 4-7, since all objects imaged are seeds, two halogen bulbs (60W) are used to illuminate the seeds at opposite 45° to remove the effect of shadow.

This line-scan NIR imaging spectrometer was bought for the research of on-line monitoring the quality of the paper in pup and paper manufacturing industry [Bharati, 2002]. It is still suitable for the off-line analysis as implemented in this thesis. Another type of imaging spectrometer often used for off-line analysis in laboratory is imaging spectrometers using tunable filters. They build multi-spectral images of an object by capturing individual 2-dimensional spatial images one wavelength at a time and changing wavelengths through a grating system. Because they require a stationary object at the grating scan through the wavelength spectrum to acquire the 3^{rd} (λ) dimension of the muli-spectral image, they are not suitable for on-line use.

2.2.3 Traditional probe-based NIR spectrometers vs. NIR imaging spectrometers

NIR technology has been used in the laboratory and in industry as an analysis tool or a sensor for decades. Almost all the literature about this technology is based on the probe-based spectroscopes. The reading of these instruments is the averaged spectrum of one point or the average spectrum taken at multipoint over the local region of the sample being tested.

One of the shortcomings of probe-based NIR spectrometers is their inability to provide simultaneous multiple point readings across solid samples. On the one hand, such information could be vital to determine the homogeneity of the sample based on the spatial distribution of its chemical information across a certain area. For example, in pharmaceutical industry, it could be used to decide how the active pharmaceutical ingredients are distributed in the tablet, which affects the therapeutic performance of the medicine. On the other hand, if there were many samples need to be tested, it would take long time to obtain the spectrum of each sample sequentially thus greatly reducing the test speed.

In recent years, this issue has been addressed with the introduction of NIR imaging spectrometers. As mentioned in section 2.2.2, these instruments work on the

same principle as their older, conventional cousins except for their detector, which in this thesis is a NIR digital camera. This NIR imaging spectrometer acquires digitized NIR reflectance spectra at multiple spatial locations across the scanned section of the samples simultaneously. This is equivalent to multiple NIR probe-detectors spread across the surface of the solid sample. The acquired hyper-spectral data cube (i. e. the 3-dimensional image matrix) can be treated as a series of spatially resolved spectra (i.e. pixels) or, alternatively, as a series of spectrally resolved images (i.e. image planes or channels). Selecting a single pixel will yield the spectrum recorded that particular spatial location in the sample. Similarly, selecting a single image plane will show the intensity response of the scene at that one particular wavelength. (Figure 2.8)

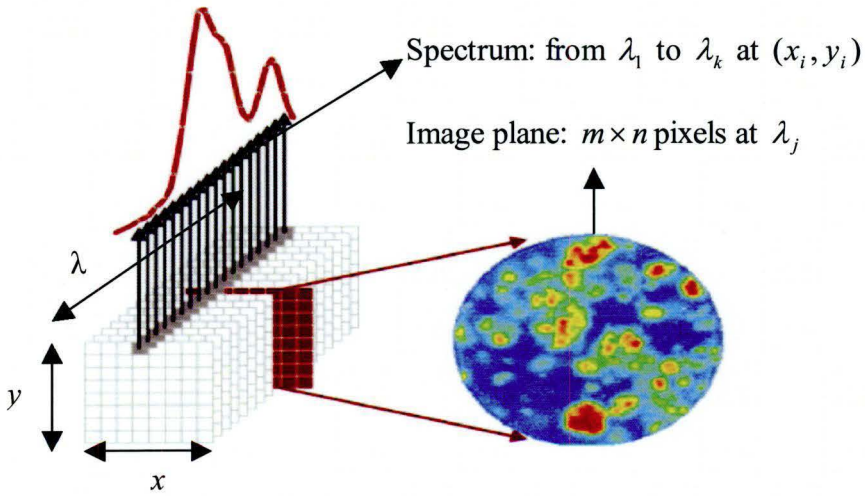


Figure 2.8 Schematic representation of a hyperspectral NIR image showing the relationship between spatial and spectral dimension

The NIR imaging spectrometer can be used as a multi-point probe-based spectrometer when many samples need to be tested. Since all samples are imaged in one picture, the average spectrum of each sample can be calculated, the multivariate calibration methods used for the probe-based spectrometers could be adopted to analyze the data. This would remarkably improve the test speed. Multivariate image analysis methods could also be used for the data cube. The spatial variation of the content within the sample or between samples could be visualized in certain feature spaces. So, NIR

imaging spectrometer retains all the advantage of probe-based spectrometers and gain several more through the addition of spatial dimensions and parallel data collection. In recent years it has become an extremely powerful adjunction to NIR spectroscopy in a number of different ways [Lewis et al., 2004].

However, there is no free lunch. Compared with the probe-based instruments, the additional complexity of the NIR imaging instruments will bring into the 3-dimension hyper-spectral image data extra systematic instrument noise which must be considered. Firstly, the area CCD array is composed of many thousands of individual infrared sensors which may be slightly different. For an ideal line-scan NIR spectrometer used in this thesis, the CCD array should produce uniform pixel response to even light if there is no spatial variation in the sample. Variations between the sensors along the spatial axis x would produce a pixel-to-pixel intensity variation across the scanned image, which results in streaks in the final image along the direction of motion of the object. Such pixel anomalies were evident in the reflectance image taken by the line-scan spectrometer used in this thesis. (Figure 3.1(a)) Secondly, lighting variations across the line of light also cause contrast variations across the imaged line, which result in contrast difference (e.g. shadow trends) across the x-y plane of the resulting image. These shadowy trends are also evident in the NIR images acquired by the imaging spectrometer used in this thesis (Figure 3.1(a)). Except for the effort for improving the hardware performance of the instruments by the spectrometer providers, some signal correction algorithm should be used to filter out the effects of such unwanted variations. This is the issue of instrument calibration and will be discussed in detail in Chapter 3.

2.3 Multivariate Statistics Based Chemometrics Techniques Using NIR spectroscopy

2.3.1 Procedure of multivariate calibration and classification

Since it is not possible to use absorbance at a single wavelength to predict the concentration of one of the absorbers due to the overlapping nature of NIR spectral peaks (the so called selectivity problem), NIR spectroscopic data are often recorded at several hundred wavelength channels, that is they are multidimensional. Because they are also highly collinear, multivariate statistics based chemometric techniques like PCA and PLS have excelled in efficiently extracting subtle differences in NIR spectra of multiple samples, and using this information to empirically model many sample properties. Multivariate chemometric methodologies fall in two categories: multivariate calibration (for quantitative application) and multivariate classification (for qualitative application). For the former one, a multivariate model based on the spectral data is built to quantify a property or a concentration whose measurement using the wet-chemistry methods is expensive or time consuming. Principal component regression (PCR) and partial least square (PLS) can be considered as standard calibration techniques for NIR spectroscopy. Multivariate classification is split into two equally important areas, cluster analysis and discriminant analysis. Cluster analysis methods can be used to find groups in the data without any predefined class structure. Cluster analysis is highly exploratory, but can sometimes, especially at the early stage of an investigation, be very useful. Discriminant analysis is a methodology which is used for building classifiers for allocating unknown samples to one of several groups. It has much in common with multivariate calibration. The difference lies in the fact that while multivariate calibration is used to predict continuous measurements, discriminant analysis is used to predict which class a sample belongs to, i.e. to predict a categorical variable. The most commonly used multivariate statistics based chemometric methods for discriminant analysis include soft independent modeling of class analogies (SIMCA) and partial least square discriminant analysis (PLS-

DA). Figure 2.9 shows the principal steps followed during the development of a quantitative/qualitative model based on NIR spectroscopy.

Sampling is the most important factor in making a robust calibration equation. The calibration sample set which is used to make a calibration equation should be representative of the population of future samples that will be predicted by NIR spectroscopy. For example, in case of agricultural products, the sample set should be sufficiently variable in respect of variety, producing area, producing year and maturity stage to meet the conditions mentioned above. The distribution of the constituent being calibrated for over the calibration samples is also important. The range of the variability should be as large as that expected in any future sample, and it is usually better to keep a more uniform spread of values over the whole range.

NIR spectra are not usually amenable for direct analysis due to unwanted systematic variation that has no correlation with the response variable. Light scattering, base line shift, instrumental drift, and path length differences are among the common sources of systematic noise in the spectra, which should be removed from the raw spectral signals [Siesler et a., 2002]. Spectral pretreatments, also called spectral filters, are mathematic functions for handling such interferences in order to reduce, eliminate or standardize the impact of the above-mentioned effects on the spectral data. Carefully designed data pretreatment algorithms can help to reduce the model complexity so that more easily interpretable methods are achieved. Often these methods are more robust against unexpected perturbations in future spectra than model based on non-pretreated spectra. However, it is worthy to be point out that most pretreatment methods also bear the potential danger of influencing a useful part of spectral information. The most commonly used data pretreatment techniques in NIR spectroscopy are the use of first or second derivatives [Savitzky and Golay, 1964], multiplicative signal correction (MSC) [Geladi et al. 1985], piece-wise multiplicative scatter correction (PMSC) [Isaksson and Kowalski, 1993], standard normal variate transformation (SNV)(Barnes *et al.* 1989) and orthogonal signal correction (OSC) [Wold et al. 1998]. In section 2.3.2, these methods

will be reviewed. One algorithm of the OSC family, Orthogonal Partial Least Square (O-PLS) will be used in Chapter 4 & 5 to eliminate the uncorrelated spectral features, and thereby simplifying models. Other approaches to handle systematic spectral variations can be found in Næs et al. [2002].

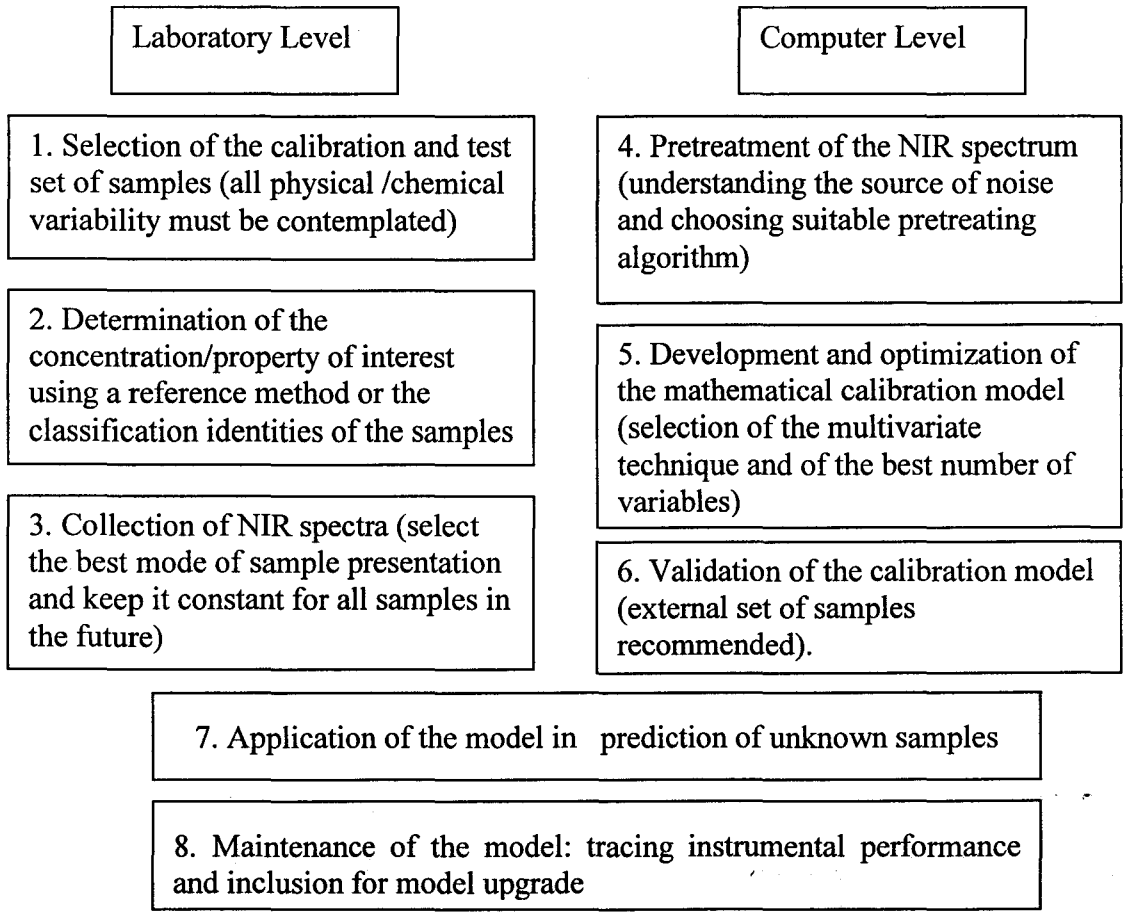


Figure 2.9 Principal steps in the development, evaluation, use and maintenance of a quantitative model based on NIR spectroscopy

PCR and PLS regression based multivariate analysis methods are generally used for NIR spectral analysis for their advantage in coping with variable nonselectivity and collinearity problems in the NIR spectral data. The structure of the calibration model, i.e. the number of principal components, is an important issue in model building and validation procedures. The criterion to evaluate underfit or overfit of the model is based

on how well the model fits observations not involved in the modeling procedure. Cross validation [Wold, S., 1978] is usually used when the dataset used for calibration and test is not very large.

The final model, employed for routine analysis, needs to be periodically checked for performance along with long term instrumental fluctuations. Intensive use may dictate the necessity of inclusion of more samples and running the construction and optimization steps again to improve the robustness of the model. In addition, efforts in making a model work with spectral data generated by an instrument other than the one used for its development has become a topic of great importance in NIR spectroscopy, which is called instrument standardization or transference of calibration [Fearn, 2001].

2.3.2 Spectral pretreatment

Among all the pretreatment methods, derivatives are used to remove or suppress constant background and to enhance the visual resolution. Background signals and global base-line variations are low-frequency phenomena, so derivatives can be interpreted as high-pass filters. The first derivative at wavelength w could be computed as:

$$x_{1der} = x_w - x_{w-1} \tag{2.1}$$

x_w is the spectral value at wavelength w in the spectrum sequence where the wavelength bands are equally spaced. The second derivative is the slope of the first derivative, and more similar to the original spectra; *i.e.*, having peaks in nearly the same locations but inverted in direction. The second derivative is computed as the difference of two adjacent first derivatives, yielding the second derivative formula:

$$x_{2der} = x_{w-1} - 2x_w + x_{w+1} \tag{2.2}$$

In spectroscopic applications, the second derivative is popular. It is a valuable tool for identifying weak peaks that are not visible in the original spectrum. The side effects of derivatives on spectroscopic data are the loss of the original shape of the spectral curve

and the reduction of the signal-to-noise ratio. To circumvent this problem, smoothing of the spectra prior to applying derivatives is essential. Savitzky and Golay (1964) described a more stringent approach based on fitting low-order polynomials.

Multiplicative signal correction (MSC) was first proposed for cases where the scatter effect is the dominating source of variability, which is very typical in many applications of diffuse NIR reflectance spectroscopy. But the idea behind this method has been extended to correct for a more general class of systematic variations in the spectral data. In MSC, it is assumed that each sample spectrum has an offset and a slope due to interference effects, one can correct for this if the variability is systematic; *i.e.*, constant over the spectral range. By plotting each spectrum, $\mathbf{x}_i$, against the reference spectrum, $\bar{\mathbf{x}}$, the offset ($\mathbf{a}_i$) and the slope ($\mathbf{b}_i$) are calculated using least squares of the equation:

$$\mathbf{x}_i = \mathbf{a}_i + \bar{\mathbf{x}}\mathbf{b}_i \tag{2.3}$$

Finally, the sample spectrum is corrected as follows:

$$\mathbf{x}_{i,corr} = (\mathbf{x}_i - \mathbf{a}_i)/\mathbf{b}_i \tag{2.4}$$

The corrected spectra give a better prediction of the response not only due to removal of irrelevant information but also due to linearization of the relationship between the predictor and the response.

This MSC method can easily be extended to a more flexible and general correction procedure in which each wavelength is corrected using individual additive and multiplicative terms. One way of doing this is by using the same idea as for MSC over a limited wavelength region, called piece-wise multiplicative scatter correction (PMSC) presented by Isaksson and Kowalski [1993]. In essence, PMSC corrects non-linear addictive and multiplicative scatter effects by fitting a linear regression in a local wavelength region. The assumption is that the scatter effects vary over the spectral range,

and hence the scatter correction should be performed piece-wise using a moving window along the wavelength range.

Orthogonal signal correction (OSC) is unique from the spectral pretreatments discussed above in one major aspect: it takes the original response variable into account in its algorithm. OSC was introduced by Wold et al. [1998] in order to avoid removal of information that is important for prediction. In OSC the signal filtering is made in such a way that the removed parts are linearly unrelated (orthogonal) to the response matrix **Y**. Different implementation of this idea has been proposed [Anderson, 1999; Fearn, 2000; Trygg and Wold, 2002]. Orthogonal partial least square (O-PLS) [Trygg and Wold, 2002] is based on the PLS NIPALS algorithm with one orthognalization step included in the algorithm. It is easy to be implemented and overcomes the overfit problem encountered in the other OSC methods. This method will be used in Chapter 4 and Chapter 5.

Basically, the OSC-treatment was developed to generate a robust prediction model for quantitative analyses through removal of interferences that have no relevance for the analyte at hand. However, in qualitative analysis where no true response variables exist, for example in PLS-DA, discrete values can be assigned to each class and used to perform OSC filtering [Wold et al. 1998]. In this thesis, this is demonstrated in Chapter 4.

2.3.3 Multivariate statistical methods for calibration

PLS is the most widely used calibration technique in NIR spectroscopy owing to its capability to handle collinearity problems, its "built in" facility for outlier detection, the possibility to analyze multiple responses, the ease for visual interpretation of the data and its ability to cope with moderate missing data. Apart from quantitative analysis, PLS can be used for pattern recognition, the so-called PLS-DA [Sjöström et al. 1986].

PLS analysis can be viewed as the regression extension of PCA. It establishes a relationship between the predictor block, **X**-matrix, and the response, **Y**, via an inner relation of their scores [Eriksson, 1999]. The **X**-scores, **T**, describe the object variation in

the predictor block (the spectral matrix in this thesis) and the corresponding variation in the response block by the Y-scores, **U**. What PLS does is to maximize the covariance between these inner variables (also called latent structures) **T** and **U**. A weight vector, **w***, is calculated for each PLS component that tells us the contribution of each **X**-variable to the explanation of **Y** in that particular component. Thus, the matrix of weights, **W***, contains the structure in **X** that maximizes the covariance between **T** and **U** over all model dimensions. Finally, the corresponding matrix of weights for the **Y**-block, **C**, and the matrix of **X**-loadings, **P**, are calculated to perform the decomposition of **X** and **Y** as follows:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \tag{2.5}$$

$$\mathbf{Y} = \mathbf{UC}' + \mathbf{F} = \mathbf{TC}' + \mathbf{G} \tag{2.6}$$

$$\mathbf{T} = \mathbf{XW}^* \tag{2.7}$$

E, **F** and **G** are residual matrices for **X**, **Y** and the inner relation, respectively left unexplained by the model.

A matrix of regression coefficients, **B**, can then be computed according to the formula:

$$\mathbf{B} = \mathbf{W}^*\mathbf{C}' \tag{2.8}$$

From the above equations, the PLS model can be expressed as

$$\mathbf{Y} = \mathbf{XW}^*\mathbf{C}' = \mathbf{XB} + \mathbf{G} \tag{2.9}$$

Each new sample is predicted either using Eq. 2.9 or by computing the scores for the new samples and multiplying with the weight from the calibration model (Eq. 2.7 and 2.6).

PLS offers many parameters and diagnostics for model interpretation, and evaluation of model performance and relevance. The scores, **T** and **U**, contain information about the observations and their similarities or dissimilarities in relation to the problem at hand. PLS score plots of the **t/t**-type are used to uncover outliers in the

descriptor matrix, **X**-space, while the **u/u**-type reveals deviation of observation in the responses matrix, **Y**-space. In addition, when PLS is used for classification/discrimination purposes, the **t/t**-type score plot for the descriptor matrix, **X**, is very useful to get an overview of the class discriminating ability of the computed PLS model. Finally, the **t/u**-type score plots are valuable tools to examine deviations from the dominating **X/Y** correlation structure as well as to identify departures from linearity between **X** and **Y**.

Similarly, the variable related information is interpreted in several ways. A plot of **X**-weights shows how the original **X**-variables are linearly combined to form the score vectors, $\mathbf{t}_a$. Using **X**-weights, it is possible to understand which original variables are summarized by the new latent variable; *i.e.* **X**-variables that are highly correlated with **Y**-variables get higher weights. In NIR spectroscopy, line plot of **X**-weights is often used, as it allows analysis of which absorption peaks are modeled by each component. Also, as a linear regression model, the coefficient plots of the PLS model are always useful for explaining the correlation relationship between the X-variables and each Y-variable.

The performance and relevance of PLS models are further evaluated by computing different statistics. The quantitative measure of *the goodness of fit* is given by the parameter $R^2(x)_{cum}$ and $R^2(y)_{cum}$, the explained variation for **X** and **Y**, respectively that can be computed as:

$$R^2(x)_{cum} = 1 - SSX[A]/SSX[0] \tag{2.10}$$

$$R^2(y)_{cum} = 1 - SSY[A]/SSY[0] \tag{2.11}$$

SSX[A] is the sum of squares of the **X**-residuals, ($\sum e_{ik}^2$), SSY[A] is the sum of squares of the **Y**-residual, ($\sum f_{im}^2$), after extracting A components; SSX[0] and SSY[0] are total sums of squares for **X** and **Y**, respectively.

The prediction ability of the computed PLS model; *the goodness of prediction*, is also quantified by a parameter called the predicted variation, $Q^2(y)$, using either cross

validation or prediction sets. The fraction of the total variation of the Y's that can be predicted by a component, $Q^2(y)$, is computed as:

$$Q^2(y) = 1 - PRESS/SS \tag{2.12}$$

PRESS is the prediction error sum of squares ($\sum(Y - \hat{Y})^2$) for the data not used to build the model and SS is the residual sum of squares of the previous dimension. This parameter is essential to determine the significance of each model dimension. The cumulative $Q^2(y)$ for all extracted components can be computed as:

$$Q^2(y)_{cum} = (1.0 - \Pi(PRESS/SS)a) \tag{2.13}$$

$\Pi(PRESS/SS)a$ is the product of PRESS/SS for each individual component, A Larger $Q^2(y)_{cum}$ value for a given response indicates that the model for that response is good. As a rule of thumb, a model with $Q^2(y)_{cum} > 0.5$ is considered as good, $Q^2(y)_{cum} > 0.75$ as very good and $Q^2(y)_{cum} > 0.9$ as excellent. The ultimate objective of developing a calibration model is to make predictions in the future. In all the studies in the thesis, the computed calibration models were applied to predict new samples in the prediction sets that were kept aside during model building. The modeling error and the prediction ability are further evaluated by computing the root mean square error of calibration (RMSEC) and the root mean square error of prediction (RMSEP), respectively; and can be computed as follows:

$$RMSEC = \sqrt{\sum(\hat{y} - y)^2 / (N - A - 1)} \tag{2.14}$$

$$RMSEP = \sqrt{\sum(\hat{y} - y)^2 / N} \tag{2.15}$$

$\hat{y}$ is the predicted value; y is the actual value; N is the number of samples in the validation sets (both for cross validation and test set) and A is the model dimension. The smaller the two values the better the calibration and prediction performance of the model.

2.3.4 Multivariate statistics methods for discriminant analysis

SIMCA and PLS-DA are two main multivariate statistics based discriminant analysis methods often used for NIR spectrum data.

SIMCA is based on a series of PCA models. For each group in the training set a PCA model is built. A new unknown item is assessed against each of the groups in turn by evaluating a criterion taking into account of the following factors:

(1) The Euclidean distance of the new item to the principal component model for that group and

(2) Where the new item lies relative to the training samples within the principal component model

The nearest class of a sample is defined as the class model which results in a minimum "reduced" distance of sample i to model j, d_{ij}

$$d_{ij} = \sqrt{(E_r)^2 + (T_r^2)^2} \tag{2.16}$$

where $E_r = E_i / E_{0.95}$ among which $E_i = \sqrt{\sum_k e_{ik}^2}$ (e_{ik} is the component of the residual vector $\mathbf{e}_i$') is the Euclidean distance of the sample i from the plane defined by the PCA model and $E_{0.95}$ is the 95% confidence limit of the residual of the samples in the training set for the jth PCA model; $T_r^2 = T_i^2 / T_{0.95}^2$ among which $T_i^2 = \sum_{a=1}^{A} \frac{t_{ia}^2}{s_{ta}^2}$ (s_{ta}^2 is the variance of $\mathbf{t}_a$ according to the PCA model) is the Hotelling statistic for sample i and $T_{0.95}^2$ is the 95% confidence limit of the Hotelling T^2 of the training set. The distance measure d_{ij} gives equal weighting to distance in the model space (T^2) and residual space (E). A graphic illustration of SIMCA and the Euclidean distance and the Hotelling statistics is shown in figure 2.10.

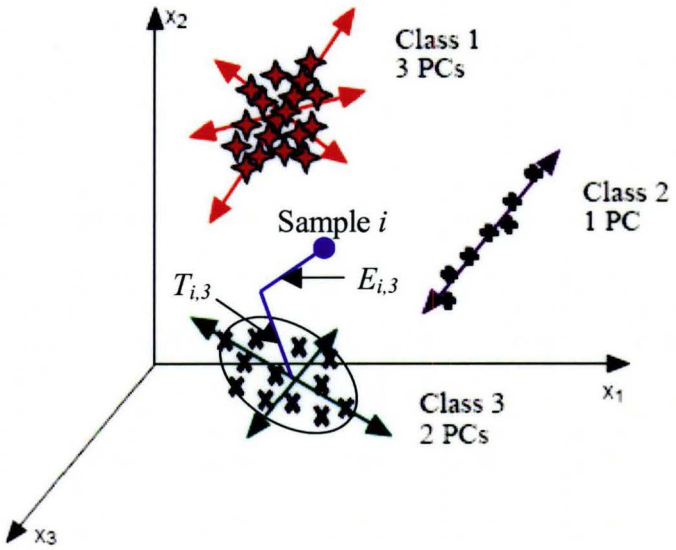


Figure 2.10 Illustration of SIMCA and the two distances

While SIMCA is often a very useful classification tool, it does have drawbacks. The main one of these is that the PCA sub-models in SIMCA are computed with the goal of capturing the variation within each class. PCA finds the directions in the multivariate space that represent the largest source of variation. However, it is not necessary the case that these maximum variations coincide with the maximum separation directions among the classes. Rather, it may be that other directions are more pertinent for discriminating among classes of observations. Therefore, the directions of the principal components in each model do not guarantee differentiating the classes. This issue is addressed by PLS-DA, which rotates the direction of the projection to give latent variables that focus on class separation. This is realized by fabricating a dummy matrix which describes the class membership of each observation and using it as the **Y**-data for PLS. A dummy variable is an artificial variable that assumes a discrete numerical value in the class description. The dummy matrix **Y** has G columns (for G classes) with ones and zero, such that the entry of the *g*th columns is one and the entries in other columns are zeros for observations of class *g*.

PLS-DA is a linear discriminant analysis method, however, it does not works well in all situations where LDA works. Since only the values of one class are set to one and all the others are set to zeros in each column of **Y**, this indicates that all the observations corresponding to these zero values belong to one general class. Therefore, PLS-DA is equivalent to deciding a discriminant plane for each class to separate it from all the other classes and then the class boundaries defined by all the discrminant planes would be considered by the PLS2-DA model when performing classification. Therefore, PLS-DA works well only when such discriminant plane exists for each class. This issue will be addressed in detail in Chapter 4.

Since PLS–DA is a special version of PLS implementing discrimination by regression, it enjoys all the advantages of PLS and can use all the parameters and diagnostics of PLS for model interpretation and evaluation as mentioned in section 2.3.2. When PLS-DA is used for analyzing NIR spectrum of different samples, it would not only give the classification result but also the complementary information such as the contribution of each wavelength bands to one score variable, which would be a good indication of some quality information of the sample. In view of this advantage, PLS-DA will be employed as a main classification method for discriminating different kinds of wheat kernels in this thesis.

Chapter 3

Calibration of Line-scan NIR Imaging Systems

The NIR imaging instrumentation and its use are not free from problems [Geladi et al., 2004]. In this chapter we propose a method to pretreat NIR image data to reduce the systematic noise introduced by the line-scan NIR imaging system described in section 2.2. It is observed that some detailed information blurred by the systematic noise can be visualized after the image is pretreated using the proposed method. An example is used to show the benefit of the methodology to improve the result of multivariate image analysis (MIA) based on multi-way principal component analysis (MPCA). This method represents a practical and easily used tool for the calibration of line-scan NIR imaging systems since it does not employ expensive standard reflectance materials.

3.1 Introduction

3.1.1 Standardization of NIR instruments

As mentioned in section 2.3, data pretreatment is usually needed to reduce the noise level before the NIR spectral data is used for sample characterization or concentration determination. There are two sources of noise affecting the signal-to-noise ratio in the spectral data: one is from the instrumentation; the other is from the physical properties of the sample, such as light scattering effects caused by the particle size distribution of the sample. In this chapter, we will focus on data pretreatment methods to reduce the influence of the noise from the NIR spectroscopic instrument, namely the calibration (or standardization or correction) of the instrument.

It is important to standardize the NIR spectroscopic instrument before using it to measure spectra on unknown samples. For any given spectrometer equipment, accurate results of the calibration model require accurate spectra. Any spectral changes, such as

wavelength shifts and baseline offsets resulting from inherent instabilities or aging parts, will lead to biased predictions from the multivariate calibration model. For multiple instruments, differences between instruments introduce the calibration model transfer problem, i. e. a calibration model developed on one instrument can not be used for other instruments of the same type [Wang et al., 1991].

Much work has been published regarding the standardization of probe-based NIR spectroscopic instruments and the transfer of multivariate calibration models [Wang et al., 1991, 1992, Fearn, 2001, Feudale et al., 2002]. The general idea of standardization is to model the instrumental differences. The spectral response of a subset of samples measured on the primary instrument is regressed against the same subset measured on the secondary instrument. Thus, changes in the response variables between the two instruments can be corrected and the original model can be used for prediction on the secondary instrument without having to compute new regression coefficients [Feudale et al., 2002].

Standardization of NIR imaging spectrometers is subject to all the error contributions of conventional one-dimensional probe-based spectroscopy (noise, drift, non-linear response of detectors, wavelength-dependent errors) as well as the two-dimensional or spatial error components associated with camera devices and illumination (readout errors, in-consistent detector responses, quantization errors, and non-uniform lighting) [Burger and Geladi, 2005]. This requires that the standardization of the imaging spectrometer must be done for each spatial or pixel position and also for each wavelength if there are wavelength-dependent errors.

The standardization of the NIR imaging spectrometer has received attention recently as NIR imaging is making its way into laboratory practice. Geladi et al. [2004] published one paper addressing the standardization of a Spectral Dimensions™ NIR imaging instrument using a liquid crystal tunable filter (LCTF) in combination with an InGaAs diode array detector. The results are based on calibration against known reference standards. Standard NIR reflectance materials, the calibration surfaces made of

Spectrolon [Pro-lite technology, 2005] with different levels of reflectance (99%, 75%, 50% and 2% reflectance) and known reflectance spectra in units of reflectance percentage, were employed in his method. Each calibration surface was rotated to different positions by a rotating bearing, imaged several times and averaged so that the influence from the non-uniformity on the surface were eliminated and thus the noise in the image data (in units of signal intensity counts read out from the A/D converter of the spectrometer) is only from the imaging instrument and illumination. Then a linear or a quadratic regression model was fitted between the true reflectance spectral value of the standard materials at each wavelength band and the measurement value at specific spatial positions at the same wavelength in the hyperspectral images taken by the NIR imaging spectrometer. Thus a linear or quadratic calibration model cube with the same dimension as the hyperspectral image was obtained and used to correct the readout from the spectrometer to the reflectance image in units of reflectance percentage. This method is a direct extension of the traditional standardization method for the probe-based NIR instruments to the imaging instrument by taking into account of the spatial dimensions of the image. Therefore, it is able to compensate for both the sensitivity difference of the InGaAs detector at different wavelengths and the illumination unevenness and detector inhomogeneities in the spatial dimensions. Recently, Burger and Geladi [2005] published another paper addressing further calibration of this instrument. External standards (i. e. the standard reflectance materials that are imaged separately from the images to be corrected) are used to correct pixel-to-pixel variances due to camera inconsistencies and variation in sample illumination, and internal standards (i. e. a mosaic of different standard reflectance materials imaged together with the objects of which the images need to be corrected) are used to compensate for signal drift over time due to changes in power or temperature effects.

The shortcoming of the methods in Geladi [2004] and Burger and Geladi [2005] is that standard reflectance materials must be used. The ideal standard material for their methods is a kind of material with spatial and spectral uniformity. However, such standard materials are not easily found for the NIR region [Burger and Geladi, 2005]. The

Spectralon materials with different reflectance rates used in their papers are created by adding different amount of carbon black to a white Teflon-based material and appear inhomogeneous and textured at high resolution [Burger and Geladi, 2005]. Although the material was imaged several times and the images were averaged, it is not easy to guarantee the uniformity of the material. Maintenance of such material is another problem because their physical properties may deteriorate with time due to scratches on the surface and shape change affecting precision. Any errors in the reference standards will ultimately compromise the standardization result. In addition, the standard materials are usually very expensive.

3.1.2 Error sources in line-scan NIR imaging spectrometers

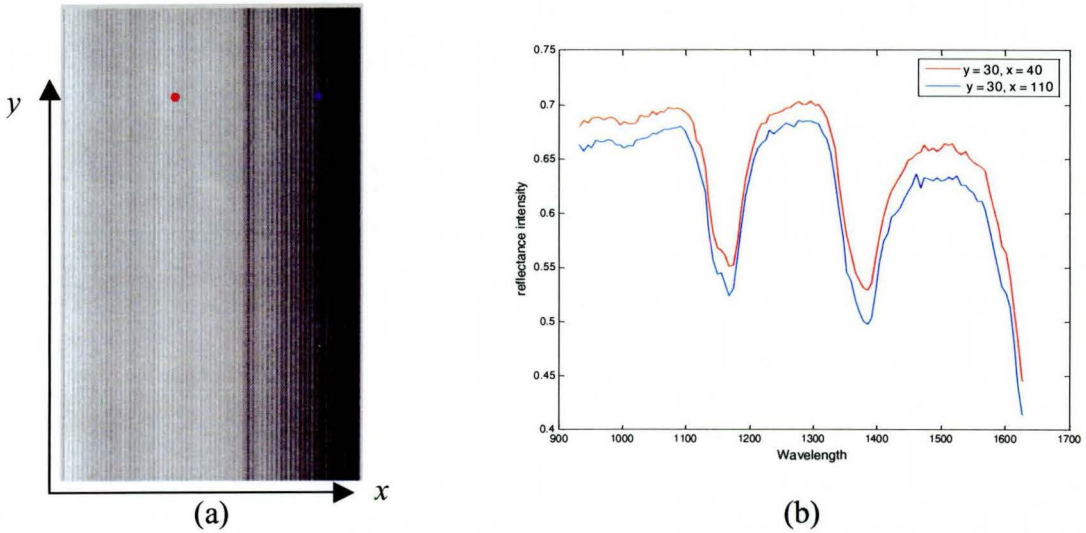
The line-scan NIR imaging spectrometer used in this thesis is converted from a monochrome area NIR camera by adding an *ImSpector* imaging spectrograph [ImSpector, 2003] between the front optics lens and the back InGaAs area CCD array of the camera (section 2.2). For each scanned line across the sample, the reflected light is vertically dispersed into its continuous spectral distribution by the *ImSpector* spectrogragh and is captured by the area CCD array detector as a spatial-spectral ($x \times \lambda$) intensity image. By moving the sample at a constant velocity in a perpendicular direction to the scan, multiple lines are recorded by the CCD array and a hyperspectral image ($y \times x \times \lambda$) of the sample is obtained. A graphic representation of the imaging system is shown in figure 2.7.

At a given wavelength band along the spectral axis, any variations among the sensors along the spatial axis x will result in streak lines along the direction of motion (y) of the object in the monochromatic image at this wavelength band. Figure 3.1 shows the mono-spectral NIR image of a red plastic shim with uniform surface and thickness at the wavelength band around 1200 nm. The evident streak lines indicate the existence of non-uniformity of the CCD array along the x axis in the line-scan NIR imaging spectrometer.

Lighting variations across the lightline also compromise the image quality by resulting in contrast difference (e.g. shadowy trends) across the x-y plane of the resulting

image. These shadow trends are also evident on the right side of the image in figure 3.1(a). The baseline difference between the spectra (figure 3.1(b)) of the two pixels highlighted in figure 3.1(a) is caused by a combination of the illumination difference between the two positions, and the sensor differences between the CCD array elements at those locations.

Furthermore, for a given pixel position in the x-y plane, if the corresponding sensors along the λ axis, giving the multi-band spectrum of this pixel, had different sensitivity to the light at different wavelength bands, its reflectance spectrum measured by the spectrometer would be different from its true reflectance spectrum. Standard illuminating sources with peaks at precisely known wavelengths are usually used to correct the spectrum [ImSpector, 2003].



(a) (b)

Figure3.1 (a) The monochromatic image of a red plastic shim with uniform surface and thickness at the wavelength around 1200 nm (b) the spectra of the two pixels in the image at the locations as marked in figure 3.1(a)

The spectral data collected from the A/D converter of the imaging spectrometer represents the signal intensity counts not actual reflectance values. The raw spectral data are mainly influenced by the light intensity of the lamp. As the lamp is used, the values decrease due to decreasing light intensity. On the other hand, the raw data do not reflect

the true intensity of the reflected light because the CCD detector generates charges even though there is no light exposure on the detector. These temperature generated charges cause a small signal, called dark current, typically varying from pixel to pixel. In precise measurements, this offset must be measured and deducted from the A/D converter counts.

In practice, the raw spectral data are transformed into reflectance or absorbance units by comparing with spectra of standard materials. The usual transformation to reflectance values is obtained by correcting sample spectra for dark current and dividing by a similarly corrected total reflectance spectrum. This is also the inherent correction mechanism integrated into the data acquiring software of the NIR spectrometer, and completed at the start of an imaging run. The procedure is described as follows: first, a spatial-spectral ($x \times \lambda$) image for a scanned line of the dark current is recorded with the lens cap in place to block light from entering the spectrometer; second, a spatial-spectral ($x \times \lambda$) image for a scanned line of a white total reflectance standard is recorded. An optical diffuse material (OP.DI.MA.) 15/10 white diffuse plastic with 98% reflectance from Gigahertz-Optik (Germany) is used for this purpose [Gigahertz-Optik, 2005]. The sample NIR reflectance image $\underline{\mathbf{R}}$ captured by the spectrometer is separated from the system response by taking, pixel by pixel, the ratio of each sample to the white image using the following equation [Hyvarinen et al., 1998]:

$$r_{yx\lambda} = \frac{s_{yx\lambda} - d_{x\lambda}}{w_{x\lambda} - d_{x\lambda}} \tag{3.1}$$

where $r_{yx\lambda}$ is an element of the hyperspectral reflectance image cube $\underline{\mathbf{R}}$ in the units of reflectance percentage, $w_{x\lambda}$ is an element of the raw spatial-spectral image **W** of a scanned line of the total reflectance standard and $d_{x\lambda}$ is an element of the raw spatial-spectral image **D** of a scanned line of the dark current imaged by blocking the lens. The above equation inherently compensates for both lighting spatial non-uniformity across the scene line, and light source color drift with aging.

Equation (3.1) is a linear calibration where the coefficient ($\frac{1}{w_{x\lambda} - d_{x\lambda}}$) is found from one standard reference value only and therefore is often termed one-point calibration. Geladi et al. [2004] found that improvements were obtained by using four references including 2%, 50%, 75% and 99% reflectance standards.

It is worthy to be pointed out that the methods of Geladi et al. [2004] and Burger and Geladi [2005] can not be used for line-scan imaging spectrometers because of the different image capturing mechanism between the filter based and the line scan imaging spectrometers (Section 2.2.2). Line scan imaging spectrometers take the image of a moving object. It is difficult to *precisely* image the *same* sample several times and then calculate the average image to eliminate the influence from the nonuniformity of the surface of the standard material. In industry, line-scan NIR imaging spectrometers are being used for on-line process monitoring and quality control. Thereby, a practical calibration method for line scan NIR imaging spectrometers is needed in practice.

In this chapter, we develop a simple method for calibrating the line-scan imaging spectrometer to reduce the systematic errors along the spatial axis x and the spectral axis λ without using the expensive uniform reflectance standard materials with known spectra.

3.2 Methodology

Color-coded plastic shims were used for the calibration. Each shim looks uniform and has even thickness at different positions. For the NIR image of the plastic shim taken by the line scan imaging spectrometer, each scanned line is measured by the same line of sensors in the InGaAs CCD array of the camera. Due to the uniformity of the plastic shim, it is reasonable to assume that the variation between the scanned lines (in the y direction) is random noise. Calculating the average along the dimension of y in an image, we obtain the average spatial-spectral intensity image of the scanned lines (called *average line image* for the purpose of this thesis, with the spatial-spectral resolution of

126×110 pixels.). Based on the "uniformity" assumption, the noise from the physical variation in the plastic shim is almost eliminated by this averaging and thereby the noise in the *average line image* is only from the difference between the sensors on the InGaAs CCD array and the unevenness of the illumination. Then by calculating the average along the dimension of x in the *average line image*, we obtain the average multi-band spectrum (called *average spectrum* for the purpose of this thesis, with the spectral resolution 110×1) of the 126 spectra in the *average line image*. This *average spectrum* is the *average spectrum* of all the pixels in the NIR image. The influence of the variation between the sensors on the CCD array along the spatial axis of x is reduced by the second averaging, and the influence from unevenness of the illumination which is not eliminated by equation 3.1 is also reduced. This *average spectrum* is almost free from the influence of variations in the imaging system and will be used as the reference spectrum for the 126 spectra in the spatial-spectral *average line image*. Our objective is to get a correction factor for each element in the spatial-spectral *average line image*, which also means getting a correction factor for each sensor of the InGaAs CCD array. Each scanned line forming the hyperspectral image will be corrected by the factors and thereby each element in the hyperspectral image cube is calibrated.

Four different plastic shims with the colors of coral, pink, white and yellow were imaged. Six images were taken, one at each of six different locations on each shim. 24 images were taken totally. Each image had 200 scanned lines in the y direction. The NIR hyperspectral image had a $y \times x \times \lambda$ dimension of $200 \times 126 \times 110$. The images were in the units of % reflectance ratio to the white reference OP.DI.MA. calculated with equation (3.1). The spatial-spectral *average line image* and the *average spectrum* of each image are calculated. Figure 3.2 illustrates the *average line image* using a false color image, figure 3.3 shows the plot of two spectra on the *average line image* and figure 3.4 shows the *average spectrum* of the image. It is observed that the spectrum in figure 3.4 is less noisy than the spectra in figure 3.3. The *average line image*s and the average spectra of the 24 images are used to estimate the correction factor for each sensor in the InGaAs CCD array.

Figure 3.5 (a) shows the relationship between the elements of the 24 *average line images* at the spatial-spectral position $x = 30$ and $\lambda = 90$ and their reference, the average spectral values at $\lambda = 90$. Figure 3.5 (b) shows the relationship between the elements of the 24 *average line images* at the position $x = 100$ and $\lambda = 70$ and their reference values. Both plots indicate that there is a linear relationship between them as denoted by the straight line in the each figure. The same relationship is also observed between the elements of the *average line images* at other positions and their reference values. That relationship can be expressed as

$$s_\lambda = \alpha_{x\lambda} + \beta_{x\lambda} l_{x\lambda} \tag{3.2}$$

Where s_λ is the average spectral value at the wavelength band λ, $l_{x\lambda}$ is the value of the *average line image* at the spatial-spectral coordinate position x and λ, $\alpha_{x\lambda}$ is the intercept coefficient and $\beta_{x\lambda}$ is the slope coefficient. The coefficients for all the elements in the *average line image* are obtained by fitting a line regression model by least squares between the *average line images* and the average spectra of the 24 images. Figure 3.6 and figure 3.7 illustrate the slope matrix $\boldsymbol{\beta}$ and the intercept matrix $\boldsymbol{\alpha}$ by visualizing them as false color images. It can be observed that the streak lines also appear in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ which indicates that the pixels along the axis x have different correction factors. The slope matrix $\boldsymbol{\beta}$ and the intercept matrix $\boldsymbol{\alpha}$ then will be used to filter each scanned line in the NIR image $\underline{\mathbf{R}}$ to counteract systematic errors from the imaging system using the following equation:

$$r_{yx\lambda,corr} = \alpha_{x\lambda} + \beta_{x\lambda} r_{yx\lambda} \tag{3.3}$$

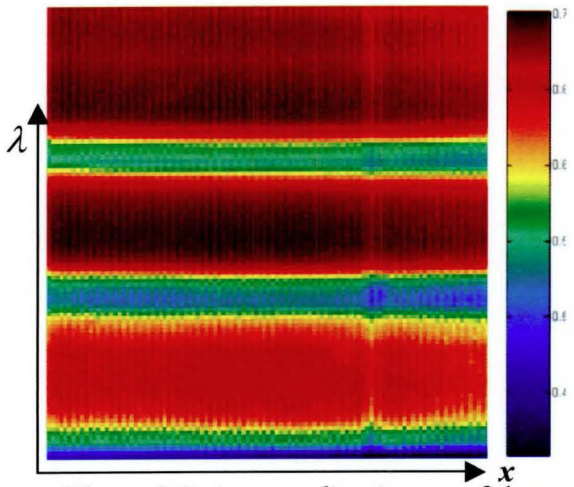


Figure 3.2 *Average line image* of the red plastic shim

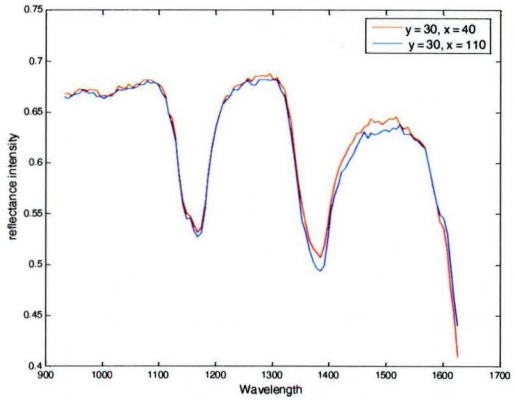


Figure 3.3 Two spectra from the *average line image* of the red plastic shim at $x = 59$ and $x = 60$

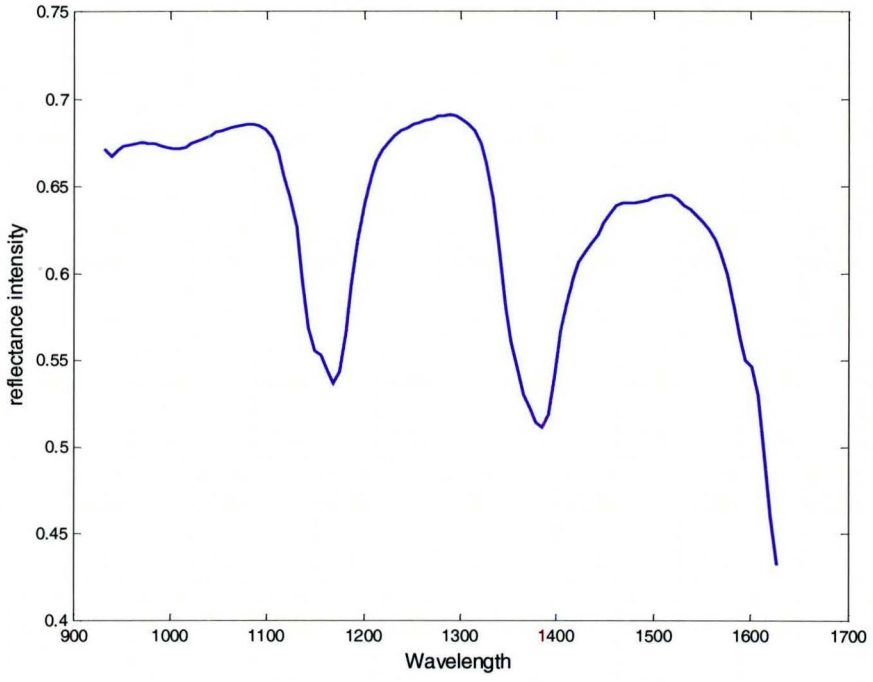
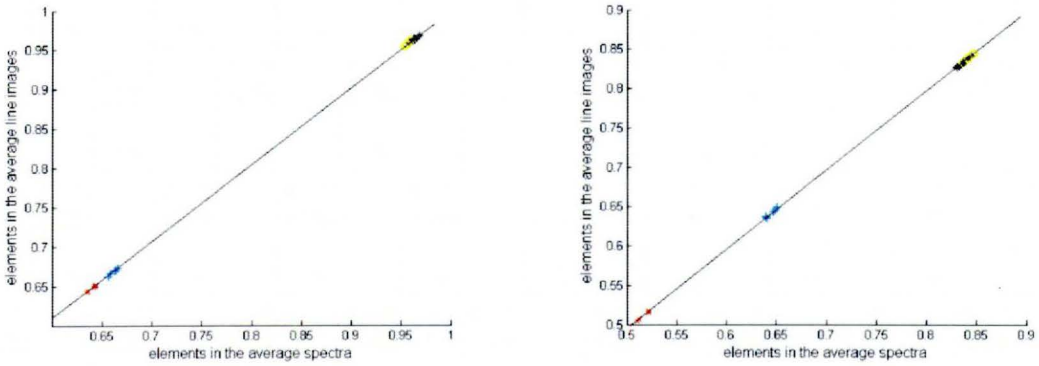


Figure 3.4 *Average spectrum* of all the pixels in the NIR image of the red plastic shim



(a) (b)

Figure 3.5 The elements in the *average line image*s vs. their reference values (a) The elements in the 24 *average line image*s at the spatial location x = 30 and $\lambda = 90$ vs. their reference, the average spectral values at $\lambda = 90$. (b) The elements in the 24 *average line image*s at the spatial location x = 100 and $\lambda = 70$ vs. their reference, the average spectral values at $\lambda = 70$.

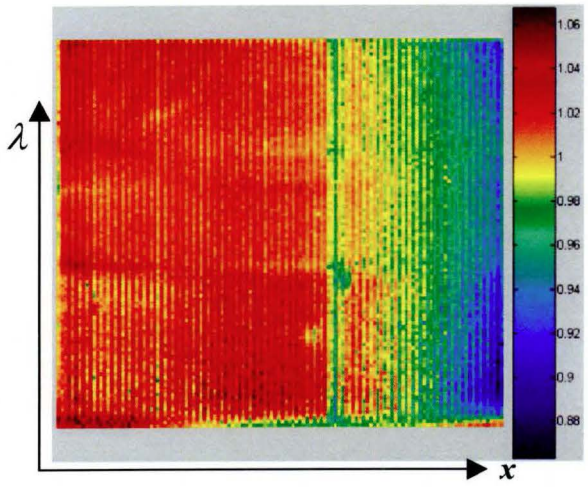


Figure 3.6 Image presentation of the slope matrix β'

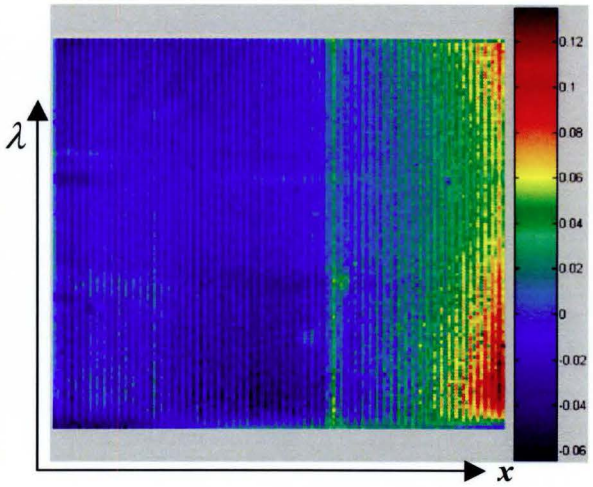
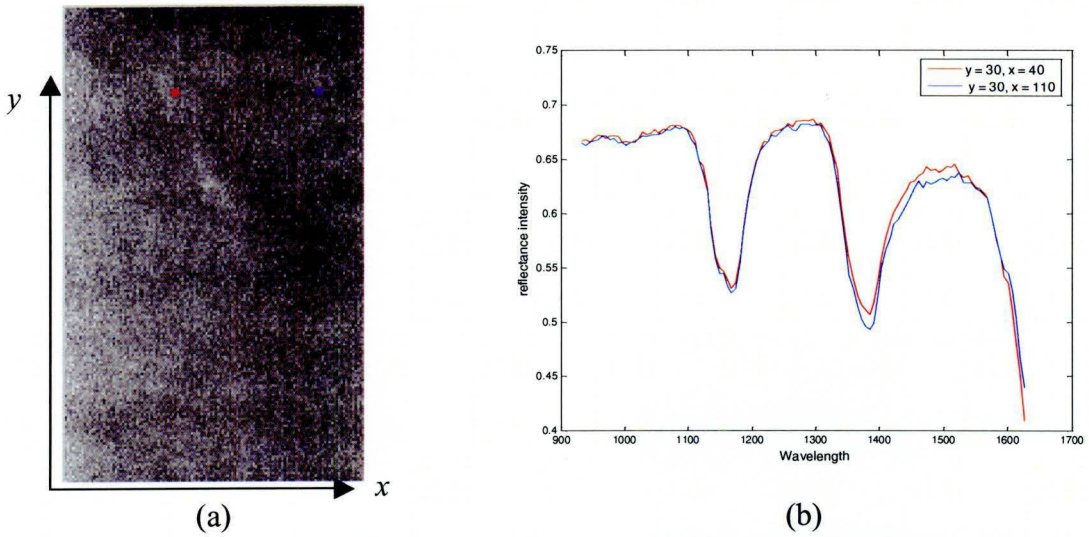


Figure 3.7 Image presentation of the intercept matrix α'

3.3 Results

Figure 3.8(a) shows the corrected result of the image in figure 3.1. It is observed that both the streaks caused by the non-uniformity of the CCD array and the shadow caused by the unevenness of illumination along the spatial axis x are reduced remarkably. The small thickness variation in the sample, which is submerged by the systematic noise

in the original images, is shown more clearly after the correction. Figure 3.8(b) shows the corrected results of the spectra of the two pixels in figure 3.1. Compared with the plots in figure 3.1(b), the baseline shift is remarkably reduced and the two spectra look more consistent with each other, which illustrate the influence of the uneven illumination along the axis x and the sensitivity difference between the sensors along the spectral axis λ are reduced.



(a) (b)

Figure3.8 Correction result (a) Corrected monochromatic image of the red plastic shim at wavelength band 1200 nm (b) The corrected spectra of the two pixels marked in figure 3.1 (a)

The calibration model is also tested on another image not used in building the calibration model. A yellow plastic shim with a fingerprint and some glue residual from a removed piece of adhesive tape at the center was imaged. Since the spectral channels are highly correlated, multivariate image analysis (MIA) technique using multi-way principal component analysis (MPCA) decomposition [Geladi et al., 1996] was used to extract the variations in the hyperspectral image. Two score images which explained 99.99% variation were adopted. Figure 3.9(a) shows the composite false color image by combining the first two score images. It can be observed that the streak lines blur the fingerprint and the glue residual and make them less visible. This indicates that the spatial variation in the sensors of each pixel location is much greater than the signal arising from

the fingerprint and the glue residual. The corresponding t_1-t_2 score plot for this uncorrected image is shown below in figure 3.9 (b). It is observed that the pixels of the background scatter over a wide area in the t_1-t_2 score space. To delineate the background from the fingerprint and the glue residual, the masking/highlighting strategy is used by manually creating a mask in the score space scattering plot and highlighting the corresponding pixels in the composite false color score image space [Geladi et al., 1996] (figure 3.9 (c) and figure 3.9 (d)). Although the masking procedure has extracted the essence of the fingerprint and glue residual from the removed tape because of the influence of streak lines, the boundaries between the fingerprint and background and between the fingerprint and residual glue region are not clear.

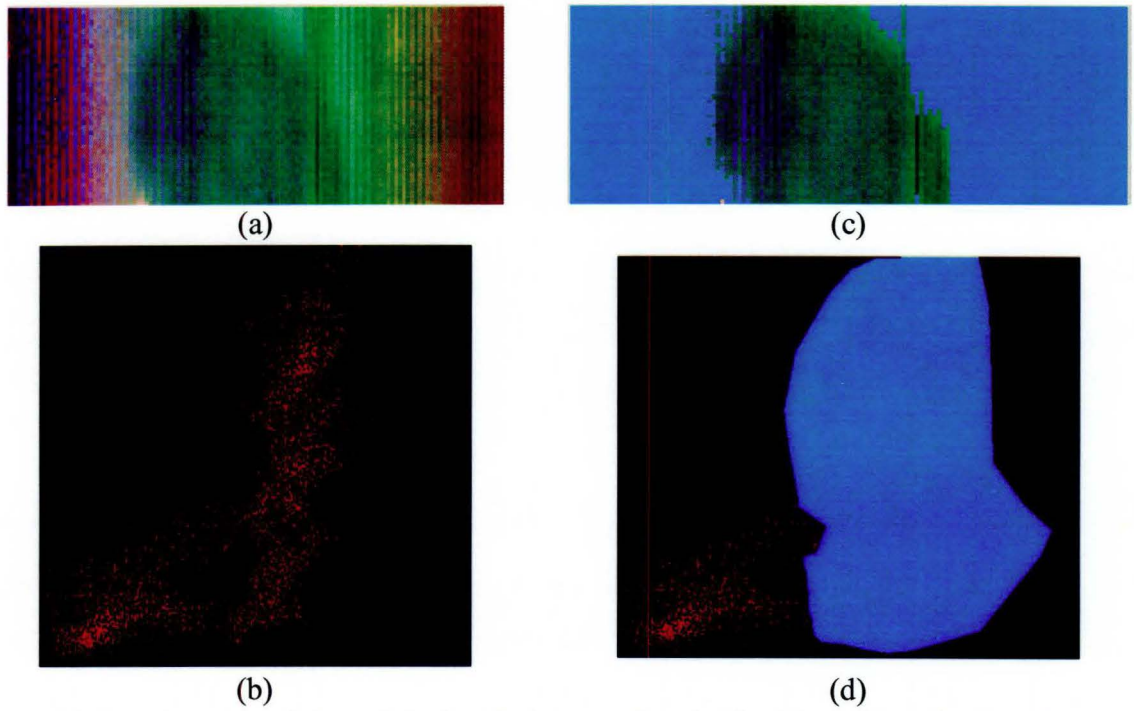
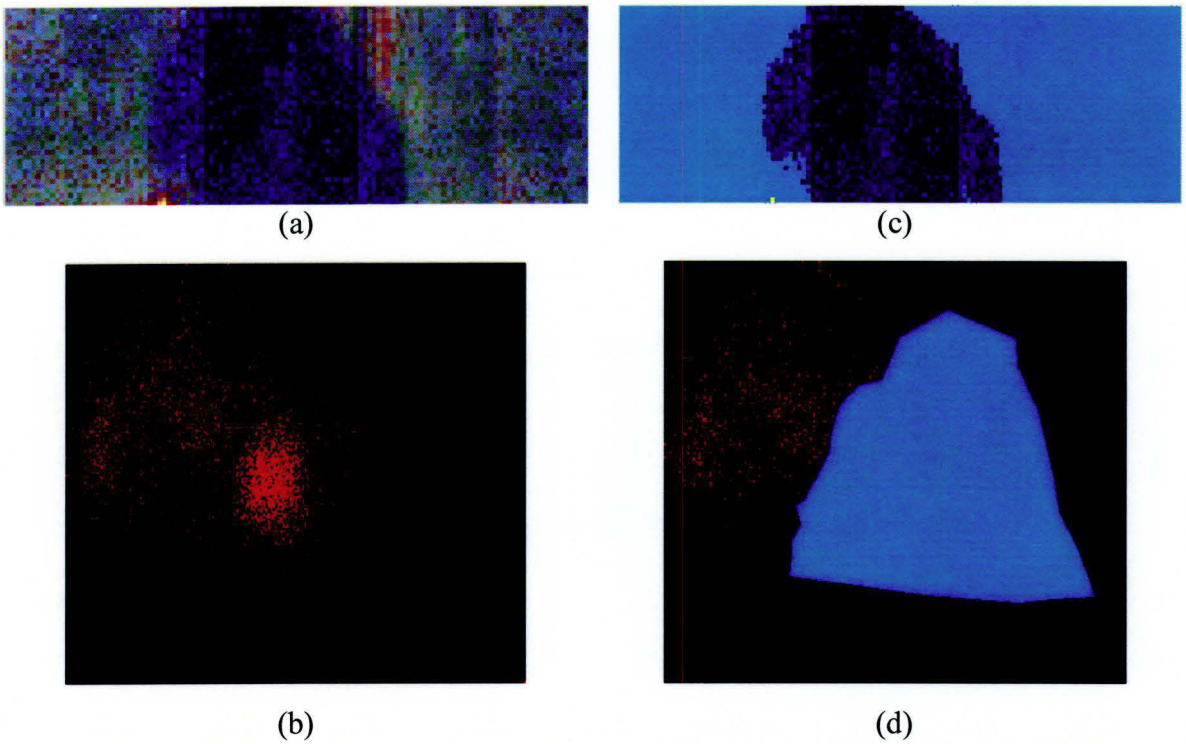


Figure 3.9 MIA result of the original NIR image of a plastic shim with a finger print and some glue residual at the center (a) Combined t1+t2 false color score image (b) t1-t2 scattering plot (c) Combined t1+t2 false color score image with the background marked (d) t1-t2 scattering plot with the background pixels marked

Figure 3.10 shows the result of the same MPCA-based MIA procedure after pretreating the NIR image by the calibration model, equation 3.3. It is shown that the fingerprint and

the glue residual from the strip of adhesive tape are more clearly distinguished from the background and from each other in the composite false color image and the background pixels cluster much more tightly in the scattered t_1-t_2 score plot. Consequently, the masking/highlighting result is improved.



(a) (c)

(b) (d)

Figure 3.10 MIA result of the corrected NIR image of a plastic shim with a finger print and some glue residual at the center (a) Combined t1+t2 false color score image (b) t1-t2 scattering plot (c) Combined t1+t2 false color score image with the background marked (d) t1-t2 scattering plot with the background pixels marked

3.4 Conclusions and Discussion

From the visual difference between the images before and after the correction, it can be seen that the instrument calibration method proposed is an effective way to reduce the systematic errors from the line-scan imaging system and to improve the accuracy of the NIR image. Both of the inconsistencies along the axis x and the axis λ in the imaging system are corrected. The benefit to the result of subsequent image analysis is also

demonstrated with an example. The method provides a practical and inexpensive approach since it only employs homogeneous objects with even thickness instead of using reflectance standard materials which are very expensive and difficult to maintain.

Chapter 4

NIR Imaging for Classification of Wheat Kernels

This chapter is a preliminary feasibility study of applying NIR imaging for the classification of different wheat kernels. SIMCA and PLS-DA methods are employed for the classification. The implementation of these methods is also discussed.

4.1 Introduction

Canada is among the top exporters of grain in an increasingly sophisticated and competitive international market [CGC, 2005]. Canada has a stringent grain inspection and grading process and hence is known for its superior quality grain in the global market. The main gain is wheat in Canada [CGC, 2005]. Total wheat production was 25.86 million metric tons in 2004 [USDA, 2004]. Based on the information of cultivars and the growing region, the wheat in Canada is designated to different classes, such as Canada Western Red Spring (CWRS), Canada Western Hard White Spring (CWHWS), etc. For each class, grades are carefully established by Canadian Grain Commision (CGC) to describe the processing qualities of the grain. Grading factors are associated with adverse growing conditions in Canada and affect the edibility and end-use performance of common wheat. The frequently encountered grading factors include fungal infections such as fusarium, ergot and mildew, insect infections like midge and the influences from growing and storing conditions, pre-harvest sprouting for example. CGC is responsible for providing the definition or standard of each grading factor and the extent or values of the factors in different grades. Generally, each class of Canadian wheat is graded in four levels. The lower the grade the better the quality.

Wheat is graded by its visual characteristics. At present, wheat inspection and grading are carried out manually in Canada. When wheat is unloaded at the elevators, CGC staff grade the wheat samples by comparing them to the definition of different grades and to the standards that represent, as close as possible, the minimum level of

quality expected for a particular grade. The sample in review must be better than or equal to these resources, otherwise it is assigned to the next lower grade. This grading process is, however, subjective and is limited by experience and expertise of the individual. The decision making capabilities of a grain inspector can be seriously affected by his/her physical condition such as fatigue and eyesight, mental state caused by biases and work pressure and working conditions such as proper lighting, climate, etc. Human involvement induces problems like inconsistency, high labor cost, and fatigue. This can lead to economic losses due to poor grade determinations. As a result of the increased number of cultivars and amount of grain handled in recent times, it is difficult to train grain inspectors every year to grade all incoming grain objectively [Delwiche and Norris 1993]. In addition, the manual process is usually time-consuming.

These problems can be eliminated by the use of an automated system based on the principles of machine vision. Machine vision systems (MVS) and pattern recognition are used to determine external features and internal characteristics of products giving objective results rapidly. The feasibility of using MVS for identification and classification of seeds has been reported widely [Barker et al. 1992; Sapirstein and Bushuk, 1989; Symons and Fulcher, 1988; Zayas et al. 1986]. MVS based automated grain analysis could offer objective and rapid analysis of grain and reduce the subjectivity inherent in the grain quality assessment process. Majumdar et al. have published a series of papers [Majumdar et al., 1996, 1999, 2000a, 2000b, 2000c, 2000d] on classifying individual kernels of different species of Canadian grains using the color, textural and morphological features from the color images. Many references on color machine imaging in the grain industry can be obtained from these papers.

Recently, automated grain analysis instruments based on MVS have been developed and appeared in the market. One such product is the Dupont ™ Acurum™ system. It can instantly analyze samples of over 3000 seeds at a time, measuring physical characteristics such as texture, length, width, as well as color characteristics. This information is then processed into judgments, using artificial intelligence and neural

network modeling. This system is capable of accessing most of the visual grading factors defined by the CGC.

However, it is reported that RGB based MVSs do not work well in evaluating sprout damage, one of the most important grading factors affecting the processing performance of wheat. Sprout damage is due to pre-harvest germination. Under conditions of prolonged dampness or rain, wheat kernels may start to germinate while the wheat crop is lying in the swath. This may also occur in lodged stands or, under very warm and wet conditions, when the mature crop is still standing. Germination begins when mature kernels absorb water and generate enzymes that break down stored starch and protein in the endosperm. The enzymes release sugars from starch and amino acids from proteins which nourish the growing embryo. The most important of these enzymes is called α-amylase. Sprout damage is detrimental to bread quality because of the action of the starch degrading enzyme α-amylase which is present in very high levels in sprouted wheat [Dexter, 1998]. The α-amylase degrades starch during mixing and fermentation reducing the water holding capacity of starch. Baking absorption must then be reduced, lowering the number of loaves of bread obtained from a given weight of flour, an important economic consideration to bakers. Loaf volume is often not affected by sprout damage, and can actually increase due to more rapid gas production during fermentation [Dexter, 1998]. In addition, sprout damage leads to sticky dough and gummy crumb which causes handling problems. Gummy crumb causes build-up on slicer blades and interferes with effective bread slicing (Fig. 4.1). All of the effects of α-amylase are exaggerated for baking processes with long fermentation times because α-amylase continues to degrade starch during the fermentation stage. Loafs of bread made from different levels of sprouted wheat are shown in figure 4.1. The influence of sprout damage of wheat kernels to the end-product quality can be easily observed.

Figure 4.1 The loaf made from sprouted wheat is sticky. When it is sliced, it shreds. The problem is exacerbated with the loaf made from severely sprouted wheat, Courtesy CGC

Severely sprouted kernels Sprouted kernels Sound kernel

Figure 4.2 Samples of wheat kernels with different grades of sprout damage, courtesy CGC

It is not surprising that sprout damage is a critical grading factor in Canada's grading system. All classes of Western Canadian wheat are assessed for sprouted and severely sprouted kernels. All classes of eastern Canadian wheat are assessed for sprouted kernels [CGC, 2006]. For each grade, CGC has specified the tolerance of (severely) sprouted kernels in term of the percentage of the weight of (severely) sprouted seeds to the whole sample. Currently sprout damage is assessed visually. Figure 4.2 demonstrates the definition sample of sprouted and severely sprouted wheat kernels from the website of CGC. It can be observed that:

1) There is obvious wrinkle on the surface of the severely sprouted seeds.

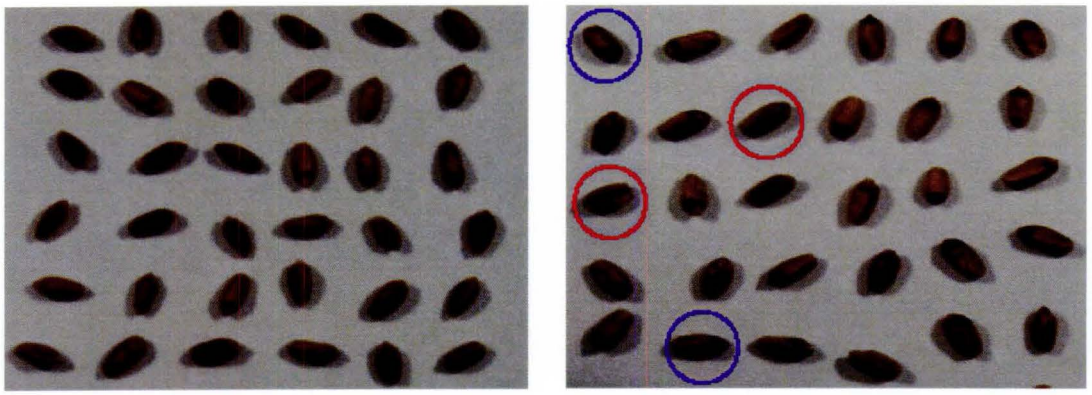
2) The color changes after sprouting. There is grayish discoloration on the surface of the sprouted seeds.

3) The germ can be observed at one end of the sprouted seed and is more obvious in the severely sprouted seed).

However, it can be observed that there is large variation in the visual characteristics between different sprouting levels of wheat kernels. Detection of these features also depends on the orientation of the kernels when they are imaged. For example, it is easy to detect the germ and discriminate between the kernels if the dorsal side is on top, however if the crease side on top, it is hard to see the difference (illustrated in Figure 4.3). It seems that "visual" features are not a "robust" characteristic that can be used to distinguish the sprouted from the healthy wheat kernels. This is probably the reason why MVSs based on RGB images are not "smart" enough to discriminate the sprouted kernels from the healthy kernels.

NIR images are a type of chemical image in which each pixel represents the NIR spectrum of the location on the object. The NIR spectrum has been widely used in industry to measure the chemical variation of organisms. It is noticed that most grading factors for wheat quality evaluation are involved with the chemical variations in or on the surface of the kernels. From this perspective, MVS based on NIR imaging should be a promising means for grading wheat kernels. Most literature and all the MVS for grain grading in industry focus on RGB imaging technology and no work has been reported on using NIR imaging technology. The objective of this chapter is to perform a preliminary feasible study using NIR imaging technology for the segregation of different kinds of kernels related to the wheat grading factors. In section 4.2, NIR imaging is used to separate the sprouted kernels from the healthy kernels. SIMCA and PLS-DA algorithms are employed to do the classification. In section 4.3, NIR imaging is used for separating four classes of grain kernels including healthy, sprouted, fusarium infected wheat kernels and barley kernels. Two implementation strategies of the PLS-DA algorithm are used for

this case. In section 4.4, the conclusions from the results are summarized. Some comments on the classification methods are given.



(a) (b)

Figure 4.3 RGB images of wheat kernels: (a) sound kernels (b) sprouted kernels The kernels marked with red circles are sprouted kernels with the crease side on top when imaged; the kernels marked with blue circles are sprouted kernels with the dorsal side on top when imaged

4.2 Discriminating Sprouted Wheat Kernels from Healthy Kernels

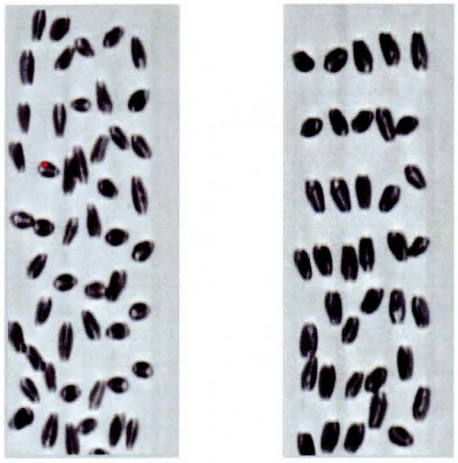
The samples used in this study include 164 healthy Canada West Red Spring (CWRS) wheat kernels, 132 sprouted CWRS kernels, 93 fusarium infected CWRS kernels and 86 barley kernels. Six images are taken of them using the NIR imaging spectrometer described in section 2.2. The information about the images is listed in table 4.1. Figure 4.4 shows the monochromatic NIR images of some healthy kernels and some sprouted kernels at the wavelength 1450 nm. Figure 4.5 shows the spectra of two pixels, one from a healthy kernel and the other from a sprouted kernel. Since we want to discriminate between the ***kernels*** in the images, the features extracted from the image should be able to represent the character of each kernel. In this case study, the average spectrum of all the pixels of the kernel is used as its feature.

The kernels in Image 1 and Image 2 in table 4.1 are used as the training set to build the calibration model to separate the healthy kernels from the sprouted kernels and

the kernels in the other images are used as the test set for validating the model. SIMCA and PLS-DA are employed to classify these two classes of wheat seeds.

Image No.	Class	No. of kernels	Utility
1	Healthy	83	Training data
2	Sprouted	64	Training data
3	Healthy	81	Test data
4	Sprouted	68	Test data
5	Barley	86	Test data
6	Fusarium infected	93	Test data

Table 4.1 Information on the wheat kernels used in this chapter



(a) (b)

Figure 4.4 Monochromatic images at wavelength 1450 nm (a) healthy kernels (b) sprouted kernels

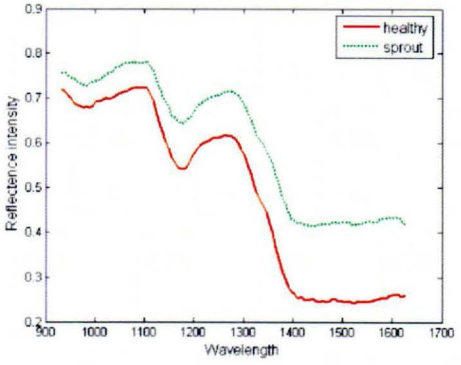


Figure 4.5 Reflectance spectra of two pixels in the left images

4.2.1 Classification using the SIMCA algorithm

A PCA model with three components, based on cross-validation, is built for the features from image 1 and denoted as "healthy model'. Another PCA model also with three components is built for the features from Image 2 and denoted as "spout model". The information of the two models is summarized in table 4.2. It is observed that more than 99% of variation in each class is explained.

Healthy model		Sprout model	
Component	R^2_{cum}	Component	R^2_{cum}
1	0.9108	1	0.9137
2	0.9893	2	0.9893
3	0.9943	3	0.9944

Table 4.2 Information of the two sub PCA models in the SIMCA

To decide the class assignment of an unknown kernel, its feature is projected to each of the two models. The classification decision is made by comparing the distances of the feature to the two PCA models (DMoDX). The Euclidean distance is usually used. It describes the vertical distance from the sample to the hyper plane formed by the PCs of the PCA model and represents the variation not explained by the PCA model. It is calculated as the root of the sum of squares of the residual of the sample projected to the hyper plane. A small Euclidean distance indicates that the sample has characteristics that are similar to the model. In SIMCA, the classification decision for one sample is made following the rules as followed:

1. If the Euclidean distances of the sample to both models are larger than the 95% confidence limits of the distances (the statistics calculated using the samples in the calibration data set), the sample is recognized as not belonging to the two classes in the SIMCA model.

2. If the Euclidean distance to one model is below the 95% confidence limit but the distance to another model is larger than the 95% confidence limit, the sample is assigned to the class described by the former model.

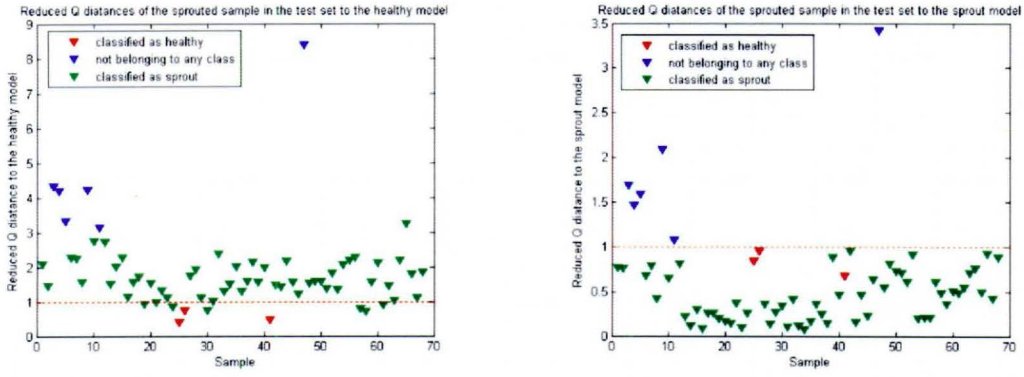
3. If the Euclidean distances to both of the two models are below the 95% confidence limits, the sample is assigned to the class with the smaller "reduced" Euclidean distance which is the ratio of the Euclidian distance to the 95% confidence limit.

Figure 4.6 shows the "reduced" Euclidean distances of the sprouted kernels in the test set to the models. It can be observed that the distances of most kernels to the sprouted

models are below and to the healthy model are above the 95% confidence limit. According to the decision rules, most kernels are correctly classified (marked with green triangles). It also shows that 6 kernels (marked with blue triangles) are far from both models and not recognized by the models as belonging to either. Three sprouted kernels are close to both models but "relatively" closer to the healthy model and therefore are misclassified as the healthy kernels (marked with red triangles). The classification results of all the kernels both in the training set and in the test set are illustrated with false color images in figure 4.7. The results are summarized in table 4.3. However, it is observed that the classification result for samples not belonging to either of the two classes is not good. There are 18 barley and 14 fusarium-infected kernels inaccurately recognized as sprouted seeds. It is interesting to observe what they look like after being projected to the hyper plane formed by the PCs of the sprout model. Figure 4.8(b) illustrates the "reduced" Euclidean distances of the barley kernels to the sprout model and figure 4.8(a) shows their "reduced" Hotelling T^2 (which is the ratio of the Hotelling T^2 to the 95% confidence limit) in the hyper plane formed by the PCs of the sprout model. It is observed that all of the Hotelling T^2s are outside the 95% confidence limits meaning that all the barley kernels are far from the center of the sprout model and are actually outliers. However, this information can not be detected by only using the Euclidean distance. Therefore, it is reasonable to take the Hotelling T^2 statistic into account when judging the class belonging of the sample. A second usually used distance is called "combined distance" in the SIMCA algorithm. It is a combination of the Euclidean distance and the Hotelling's T^2 statistic. In the *PLS_Toolbox 3.5 for Matlab* [Eigenvector, 2004], these two distances are given the same weight. The "reduced" combined distance is calculated as [Eigenvector, 2004]:

$$d_{ij} = \sqrt{\left(\frac{E_{ij}}{E_j}\right)^2 + \left(\frac{T_{ij}^2}{T_j^2}\right)^2} \tag{4.1}$$

where E_{ij} is the Euclidian distance of sample i to the PCA model j. T_{ij}^2 is the Hotelling T^2 of sample i in the PCA model j. E_j is the 95% confidence limit of the Euclidean distance calculated using the samples for building the PCA model j. T_j^2 is the 95% confidence limit of the Hotelling T^2 statistic of the PCA model j. From the geometric point of view, this distance represents the distance of the sample to the center of the hyper plane formed by the PCs of the PCA model. It is a distance standard which can be used to detect the outliers in two directions, one is vertical to the hyper plane and the other in the hyper plane. Figure 4.9 shows the combined distances of the barley kernels to the two PCA models. It is observed that all the distances are outside the 95% confidence limit. According to the decision rules, they are correctly assigned to neither of the two classes. The classification result based on the "combined" distance is illustrated in Figure 4.10. The results are summarized in table 4.4. Compared with the result of using the Euclidean distance to the model, it is observed that the accuracy rate is greatly improved for the barley samples in the testing dataset. However, the fusarium-infected samples still are not well separated from the healthy and sprouted classes.



(a) (b)

Figure 4.6 (a) Reduced Euclidean distances of the sprouted samples in the test set to the healthy model. (b) Reduced Euclidean distances of the sprouted samples in the test set to the sprout model

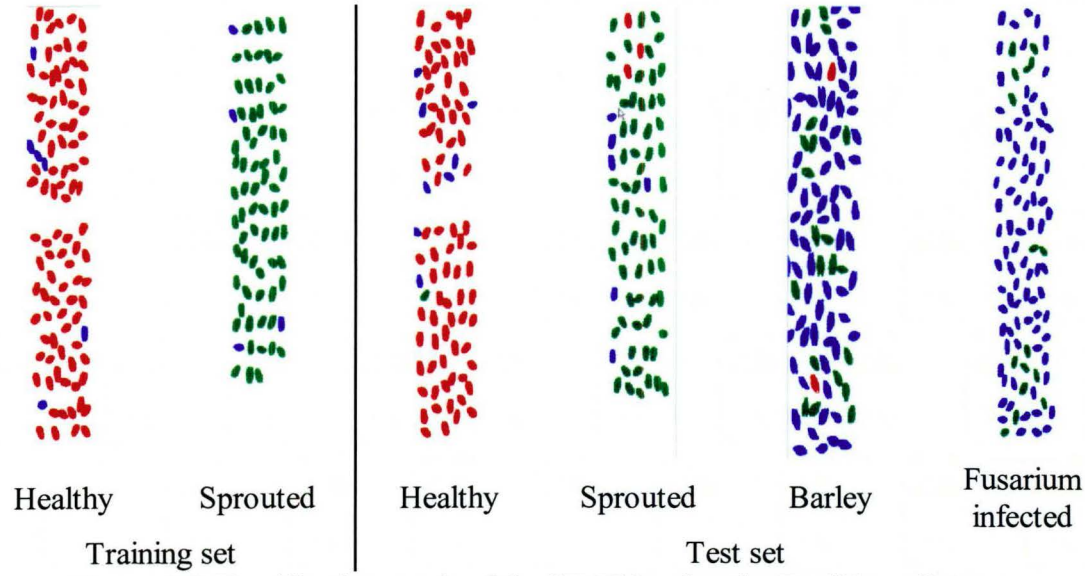
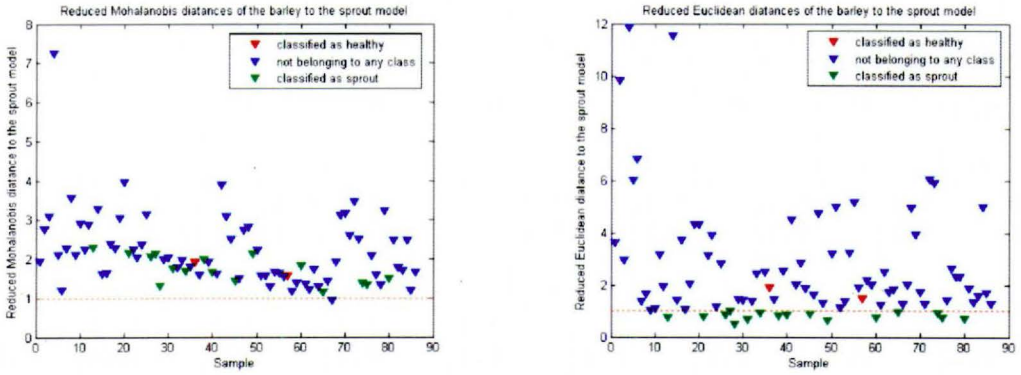


Figure 4.7 Classification result of the SIMCA using the Euclidean distance

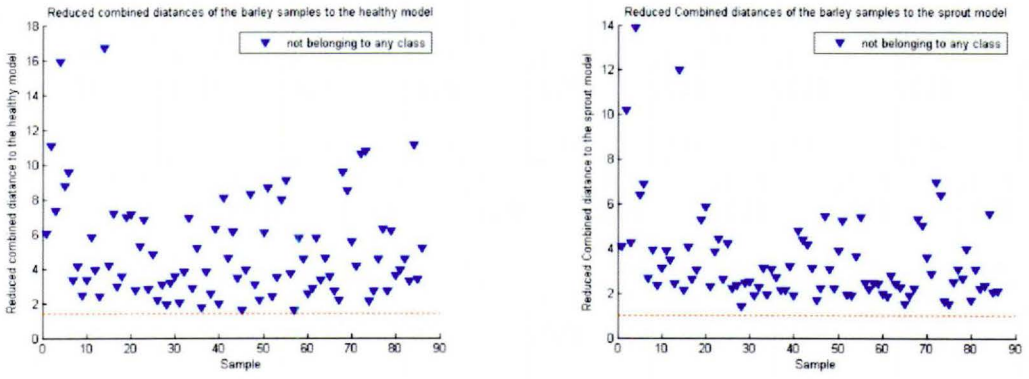
	Training set		Test set			
Classes	Healthy	Sprouted	Samples belonging to the two classes		Samples not belonging to the two classes	
			Healthy	Sprouted	Barley	Fusarium infected
No. of kernels	83	64	81	68	86	93
Classified as not belonging to either class (blue color)	4	4	8	6	68	79
Misclassified	0	0	1	3	18	14
Accuracy rate	94.6%		87.9%		82.1%	

Table 4.3 Statistics of the classification result using the Euclidean distance in SIMCA



(a) (b)

Figure 4.8 (a) Reduced Mahalanobis distance (Hotelling T^2) of the barley samples to the healthy model (b) Reduced Euclidean distances of the barley samples to the sprout model



(a) (b)

Figure 4.9 (a) Combined distances of the barley samples to the healthy model (b) Combined distances of the barley samples to the sprout model

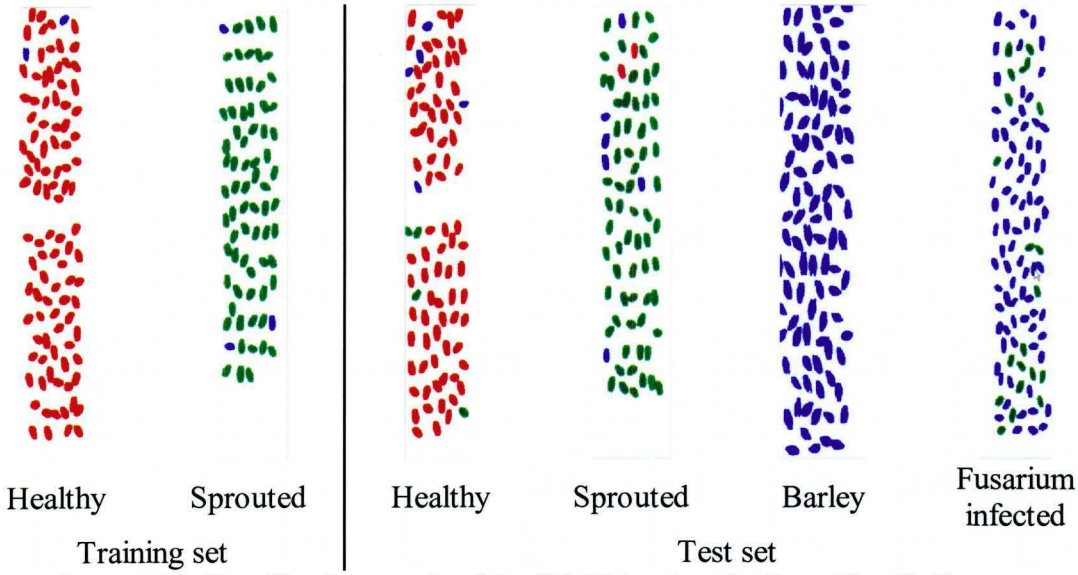


Figure 4.10 Classification result of the SIMCA using the "combined" distance

	Training set		Test set			
Classes	Healthy	Sprouted	Samples belonging to the two classes		Samples not belonging to the two classes	
			Healthy	Sprouted	Barley	Fusarium infected
No. of kernels	83	64	81	68	86	93
Classified as not belonging to either class (blue color)	2	3	5	6	86	74
Misclassified	0	0	4	2	0	19
Accuracy rate	96.6%		88.6%		89.4%	

Table 4.4 Statistics of the classification result using the "combined" distance

4.2.2 Classification using the PLS-DA algorithm

The PCA sub models in SIMCA are computed with the goal of computing the nature of the variation within each class. It is not easy to obtain the information about the classification such as which wavelength range is important for discriminating the classes. This information is usually important in practice. For example it will be instructive for choosing new suitable instrumentation. PLS-DA is more suitable than SIMCA to achieve this objective. PLS-DA is a special version of the PLS regression algorithm with the

objective of classification by setting the dummy y variables to integer values of 0 and 1 (Section 2.4). Therefore, PLS-DA also enjoys the advantages of ordinary PLS for continuous values of **Y** such as noise reduction and statistics that assist with variable selection.

Based on cross-validation, a PLS-DA model is fitted to the training dataset in table 4.1. A model summary is shown in table 4.5. It is observed that the model is very complicated having 5 components. This indicates that there is much variation in the feature space that is not directly related to the classification. It is also noticed that the fraction of the variance explained in **y**, $R^2(y)_{cum}$ and the cross-validated $Q^2(y)_{cum}$ are quite low in the first two components compared with the fraction of the variance explained in the features, $R^2(x)_{cum}$. This means that the first PLS component is mainly used to explain the **Y**-uncorrelated variation in **X**. The regression coefficients of the PLS model for the dummy variable y_1 is plotted and shown in figure 4.11. It is shown that the important wavelength variables that contribute to the separation of the two classes are the wavelengths near 950 nm, 1000nm, 1100 nm, 1200 nm and 1400 nm.

Com.	$R^2(x)_{cum}$	$R^2(y)_{cum}$	$Q^2(y)_{cum}$	Misclassified in the training set	Misclassified in the test set
1	0.904	0.152	0.14	51	61
2	0.988	0.258	0.241	45	62
3	0.993	0.811	0.797	5	8
4	0.997	0.85	0.836	0	5
5	0.998	0.869	0.852	0	3

Table 4.5 Summary of the PLS-DA model for classification of healthy and sprouted kernels

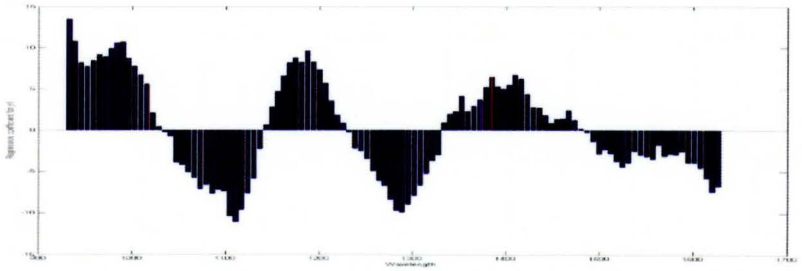


Figure 4.11 Regression coefficients of the original PLS-DA model for y_1

To simplify the PLS model, the OSC method [Wold et al. 1998] is employed to prefilter the feature data before fitting a simpler PLS-DA model. The O-PLS algorithm [Trygg and Wold, 2002] is used in this case due to its advantage over the other OSC algorithms in solving the overfitting problem and also due to it using a similar framework with the NIPALS algorithm used in the PLS modeling. In O-PLS, the number of the orthogonal components is decided by the ratio $\|\mathbf{w}_{ortho}\|/\|\mathbf{p}\|$ which represents the percentage of the y-uncorrelated variation of the feature data with respect to the variation in the covariance direction. It converges to zero if the "proper" number of orthogonal components has been calculated. The ratio $\|\mathbf{w}_{ortho}\|/\|\mathbf{p}\|$ of this study is shown in Figure 4.12, which indicates three or four orthogonal components are reasonable. Four orthogonal components are used and 86.8% of variation in the **X** is eliminated(Table 4.6).

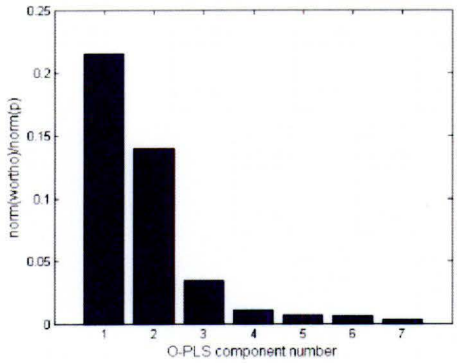


Figure 4.12 Ratio of $\|\mathbf{w}_{ortho}\|/\|\mathbf{p}\|$ for each O-PLS component

The O-PLS-DA model			
Comp.	$R^2(x)_{cum}$	$R^2(y)_{cum}$	$Q^2(y)_{cum}$
Ortho. Comp. 1	0.469		
Ortho. Comp. 2	0.837		
Ortho. Comp. 3	0.848		
Ortho. Comp. 4	0.868		
PLS Comp. 1	0.998	0.828	0.826

Table 4.6 Summary of the O-PLS-DA model for classification of healthy and sprouted kernels

A new PLS-DA model is built between the 13.2% variation left in the feature space and the dummy matrix **Y**. One PLS component is used for this PLS-DA model (Table 4.6). The coefficient plot for variable y_1 is shown in figure 4.13. It is observed that the important wavelengths for separating these two classes are at the high wavelength range. It needs to be noted that the prediction variables in the simple PLS-DA model of O-PLS-DA are not the original wavelength variables, but the filtered wavelength variables. This issue is discussed in detail in Chapter 5.

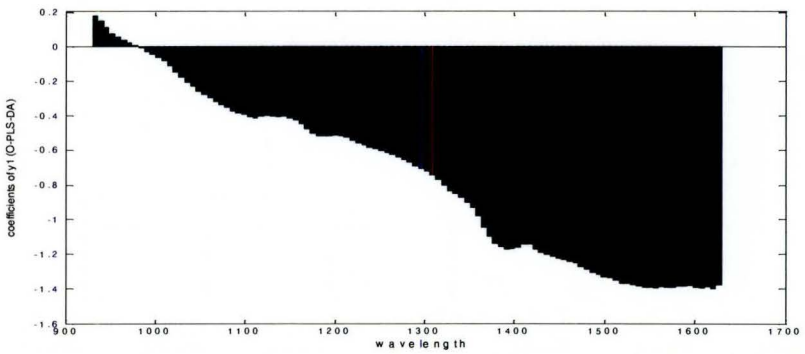


Figure 4.13 Coefficient plot of the O-PLS-DA model for y_1

The class assignment of a new unknown sample is decided based on its predicted y values from the PLS-DA model. A threshold value for separating the two classes is calculated for each y variable using the predicted y-values from the training data.

Figure 4.14 illustrates how the threshold is calculated for y_1 in this case. The red bars are a histogram of the predicted values of y_1 for the healthy samples in the training data set and the green bars are a histogram of the predicted values of y_1 for the sprout samples in the training data set. If a normal distribution is fitted to each of those histograms, they would cross at y_1-pred = 0.45. This means the probability of measuring a sample with the predicted y_1 value 0.45 for the healthy class is equal to the probability of measuring it for the sprout class. So, 0.45 is set as the threshold for separating the two classes based on the predicted values of y_1.

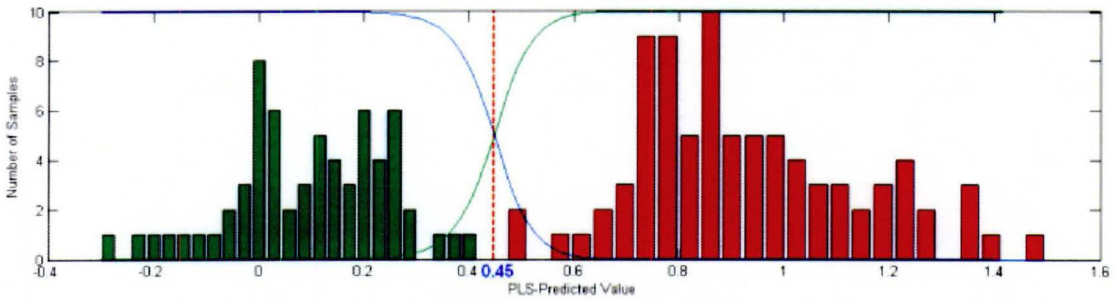


Figure 4.14 Histogram of the predicted values of *y1* for the samples in the training set. The green plot is the normal distribution fitted to the predictions for the healthy class and the blue plot is the normal distribution fitted to the predictions for the sprout class

For an unknown sample, the normal distribution marked with the green line in figure 4.14 is used to calculate its prior probability of being classified as the healthy, $P(y_1 \mid H)$ and the normal distribution marked with the blue line in figure 4.14 is used to calculate its prior probability of being classed as sprout $P(y_1 \mid S)$. Assuming that sample definitely belongs to one of the two samples, the posterior probability of this unknown sample being classified as the healthy is calculated as

$$P(H \mid y_1) = \frac{P(y_1 \mid H)}{P(y_1 \mid H) + P(y_1 \mid S)} \tag{4.2}$$

where y_1 is the y_1-value predicted from the PLS-DA model for the sample in question. The probability of the sample being classed as the sprout class is calculated as

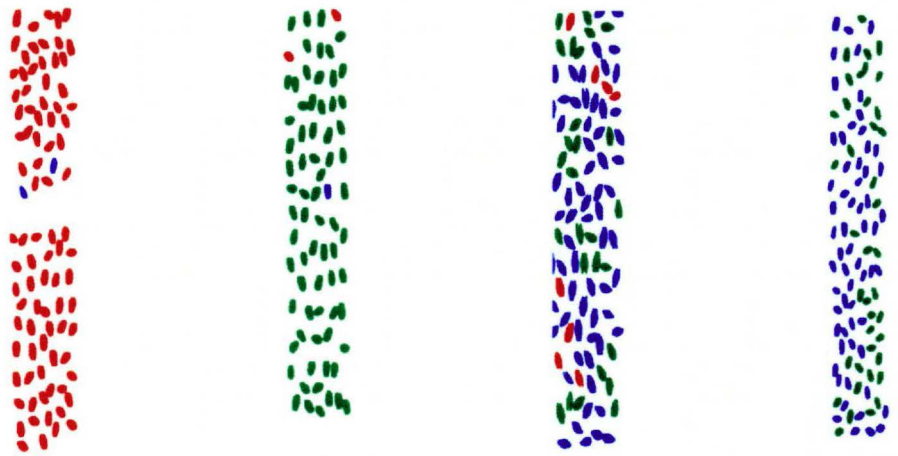
$$P(S \mid y_1) = \frac{P(y_1 \mid S)}{P(y_1 \mid H) + P(y_1 \mid S)} \tag{4.3}$$

The sample is assigned to the class with higher posterior probability.

Before using the above equations to calculate the probability for an unknown sample, its combined distance as expressed in equation 4.1 to the PLS-DA model is calculated. If its reduced combined distance to the model is larger than the 95% confidence limit, it is classified to neither of the classes. If below the 95% confidence limit, it is assigned to the class with higher posterior probability. This is the implementation of PLS-DA in the *PLS_Toolbox 3.5 for Matlab* [Eigenvector, 2004].

The classification result for the samples in the test data set is shown in figure 4.15. Compared with the results from SIMCA (Figure 4.10), the results for the healthy and sprouted kernels are better, however the results for the barley and fusarium infected kernels are worse. 64 barley and fusarium infected kernels are incorrectly classified as healthy or sprouted kernels. This is because the combined distance is not an effective measure for PLS-DA to identify the sample from a third class. Unlike SIMCA, in which each sub-PCA model explains the nature of the variation of each class, the PLS-DA

model focuses on finding the score space that maximally separates between the two classes. Therefore a small combined distance to the PLS-DA model does not necessarily mean that the sample belongs to these two classes. Figure 4.16 illustrates the combined distances of the barley and fusarium infected wheat kernels and shows that the distances of some kernels are below the 95% confidence limit. When calculating the probability of the class assignment, these kernels are assumed definitely belonging to one of the classes in the calibration set. This is the reason why so many barley and fusarium infected kernels are misclassified. This result indicates that all classes that will appear in the future should be included in the PLS-DA model so that the model can recognize them and provide accurate classification results.



Healthy Sprouted Barley Fusarium infected

Figure 4.15 Classification result of the O-PLS-DA model for the samples in the test set

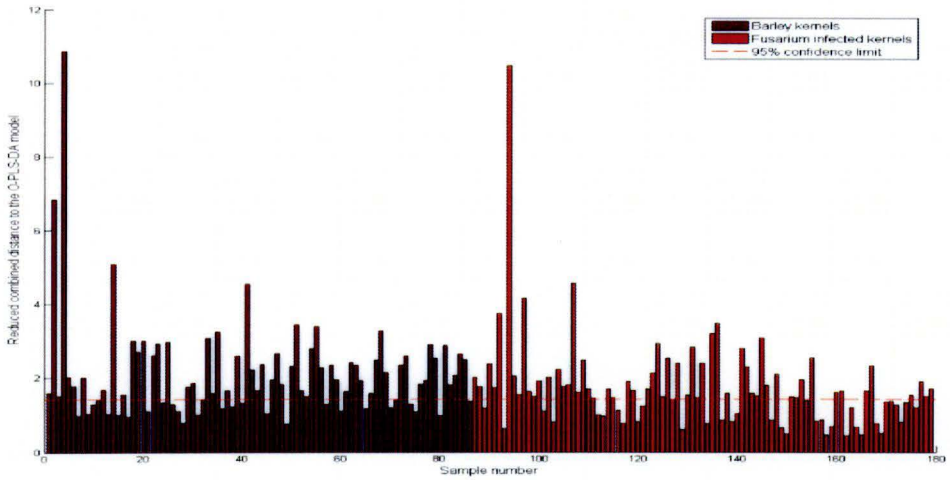


Figure 4.16 "Reduced" combined distances of the barley and fusarium infected kernels in the test set to the O-PLS-DA model

4.3 Multi-category Grain Classification Using PLS-DA

Most of the factors related to wheat grading are involved with chemical changes inside or on the surface of wheat kernels. Therefore, it is expected that NIR imaging technology should work well in discriminating between them. In this section, we further evaluate the feasibility of NIR imaging to classify four classes of grain: healthy wheat, sprouted wheat, fusarium infected wheat and barley kernels. They will be classified using the PLS-DA algorithm based on the features from their NIR images. At the same time, this will be used as a case to study the performance of the PLD-DA algorithm for multi-category classification.

One image of 89 barley kernels and one image of 97 fusarium infected wheat kernels are used to build a multi-class PLS-DA model, together with the healthy and sprouted images used for the calibration data in the previous section. The test data used for the previous section is also used for validating the model. The average spectrum of the kernel is used again as its features. Based on cross-validation, a PLS-DA model with 8 components is fitted for the training data. The classification result of applying the model

to the test data set is illustrated in figure 4.18(a) with false color images. It is observed that 14 kernels are misclassified. The accuracy rate is 95.75%. From this result, it can be said that NIR imaging technology is feasible for separating these four classes of grain.

A single PLS-DA model for discrimination between multiple classes can be considered as a classifier combining K *one-vs-rest* sub-classifiers if the $N \times K$ dummy matrix **Y** is set up in the form where a one indicates that an observation belongs to a class and a zero that it does not (where K is the number of classes and also the number of the columns of the dummy matrix **Y** and N is the number of observations). For each column of the dummy matrix Y, at least one PLS component is needed to separate the samples with y-value one in this column from the other samples. The PLS components of the final PLS-DA model are the combination of the components of all the sub-classifiers. Figure 4.17 illustrates the prediction from the final PLS-DA model for each column of the Y matrix and the threshold for separating the samples of one class from the other samples. It is observed that the sprouted kernels are not well separated from the other samples (figure 4.17(b)). This indicates that when PLS-DA is used for multi-category classification, in some situations, it is not easy to find a discriminant line or plane in the score space to accurately separate ***each*** class from ***all the other*** classes.

In the field of pattern recognition, most classifiers were initially designed for binary classification, such as support vector machine (SVM) [Vapnik, 1998] and multi-layer perceptrons [Minsky, 1969]. Two strategies are usually used when they are used for the multi-class situation. One is the so called *one-vs-rest strategy* in which K binary classifiers that separate each class from all the others are built, the other one is the so called *one-vs-one strategy* in which $K(K-1)/2$ binary classifiers are built between every two classes to separate between them. The class assignment of a sample is decided by combining the results of the K (or the $K(K-1)/2$) classifiers. The *Max-wins voting strategy* is usually used to make the decision, i. e. the result of each binary classifier votes the sample to one class and the class with the highest number of votes is set as the sample's class assignment. It has been illustrated that the *one-vs-one strategy* usually gives better

results than the *one-vs-rest strategy* [Statnikov *et al.,* 2004]. The philosophy behind the advantage of the *one-vs-one strategy* is that "even if the entire multi-category problem is non-separable, while some of the binary sub-problems are separable, then the *one-vs-one* can lead to improvement of classification compared to *one-vs-rest*" [Statnikov *et al.,* 2004].

This strategy is applied here to PLS-DA. Instead of building a single PLS-DA model for all the classes, we build *K(K-1)/2* PLS-DA models between every two classes and then combine the results of all the models to make a classification decision. It should be easier to find a discriminating line or plane in the score space to separate one class from another class than to separate one class from all other classes. The *max-win voting strategy* is used to combine the results of the sub PLS-DA models and make the decision to assign the sample to one of the *K* classes.

Six PLS-DA models are built between every two classes of grain in the training data. The number of PLS components for each model is decided by cross validation. The classification result of applying the models to the test data is shown in figure 4.18(b). It is observed that 9 kernels are misclassified and the result is much better than the result of the single PLD-DA model shown in figure 4.18(a). The classification result of each sub PLS-DA model is shown in figure 4.19. It shows that each model can separate one class from another class perfectly. It is these accurate results from each sub PLS-DA model that allows an accurate final classification result, and therefore the *one-vs-one strategy* is a good choice for implementing multi-category PLS-DA.

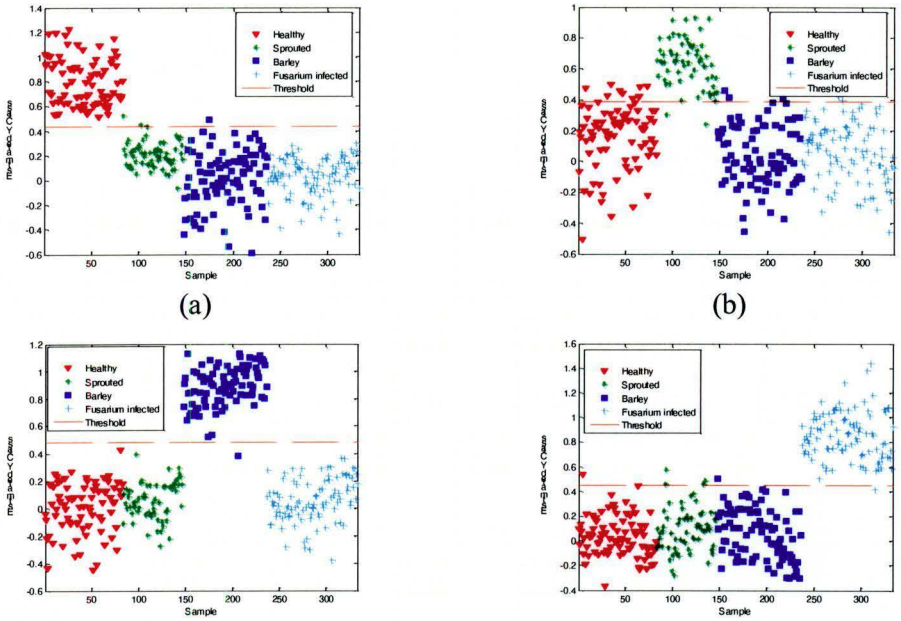


Figure 4.17 Estimated *y* values from the single PLS-DA model for each column of **Y** and the thresholds (for the training data)

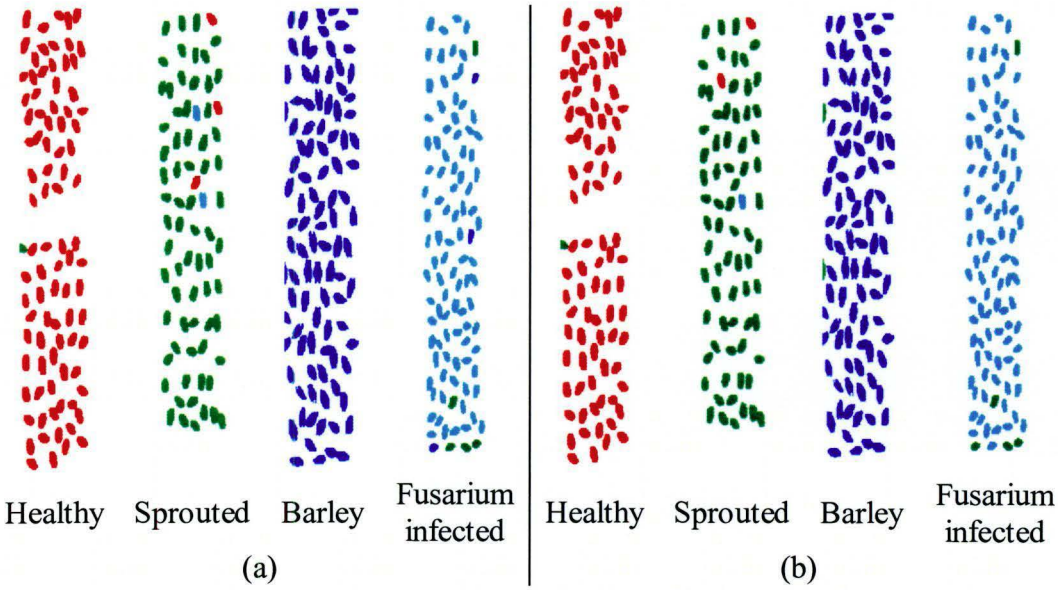


Figure 4.18 Classification result of multi-class PLS-DA for the test data
(a) result of *one-vs-rest* strategy, **14** kernels misclassified (b) result of *one-vs-one* , **9** kernels misclassified.

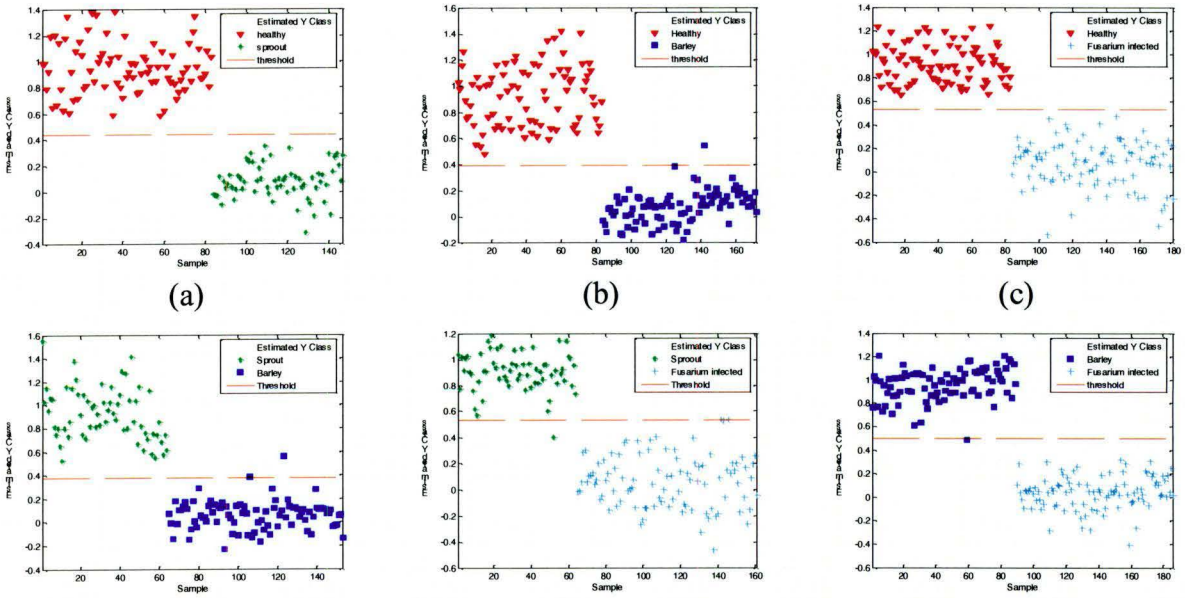


Figure 4.19 Illustration of the classification result of the six sub PLS-DA models in the *one-vs-one strategy* (for the training data)

4.4 Conclusions and Discussion

The feasibility of using NIR imaging for the grading of wheat kernels was studied in this chapter. Multivariate statistical algorithms, SIMCA and PLS-DA were used to discriminate between different types of wheat kernels using spectral features from NIR images. The results show that NIR imaging technology is feasible for wheat grading. In view of the fact that RGB image based MVS is not robust in identifying some classes of wheat kernels, the sprouted kernels for example, NIR imaging can provide an improved separation of some classes.

The results of different chemometric methods were compared. A summary of these results follows:

1. In SIMCA, the "combined" distance should be used to decide the class assignment of an unknown sample. i. e. the distance of the sample to the center of the model should be used.

2. SIMCA performs better than PLS-DA in identifying an unknown sample not belonging to the modeled classes.

3. The *one-vs-one strategy* is a better implementation of PLS-DA for multi-class classification than traditional PLS-DA.

The results from this feasibility study show that NIR imaging provides a potentially powerful new method for separating certain classes of abnormal fungal infections, types of seeds, and undesirable features such as sprouted seeds.

Chapter 5

Predicting the Falling Number Index of Wheat Flour Using NIR Imaging Technology on Samples of Wheat Kernels

The work reported in this chapter is intended to be a feasibility study of predicting the "falling number" of wheat samples using NIR imaging technology on samples of wheat kernels. Three models are built between the features extracted from NIR images of the wheat kernels and the falling number measurements made on bulk samples. Three multivariate regression methods are used: a regular PLS algorithm, the O-PLS algorithm and a PLS plus canonical correlation analysis algorithm (PLS+CCA [Yu and MacGregor, 2004]). Some useful information is obtained by analyzing the coefficients and loadings of the models. The performance of the algorithms is compared. The errors in the prediction of the PLS models are also analyzed.

5.1 Introduction

In section 4.1, we discussed the impact of sprout damage on wheat quality. However, sprout damage in wheat is difficult to assess. One simple assessment of sprout damage is to count the percentage of sprouted and severely sprouted kernels. CGC has established sprout damaged kernel tolerance for each class of Canadian wheat. The embryo and endosperm of wheat seeds produce the enzyme alpha-amylase at an accelerating rate when germination begins. A severely sprout-damaged kernel contains many thousands of times the amount of alpha-amylase than is present in the early stages of germination [CGC, 2002]. For the wheat quality assessor, it is not easy to accurately evaluate the extent of sprouting for each kernel by inspection. Therefore, it is easy for a cargo that appears to be high quality in terms of percentage of sprout damage to exhibit significant alpha-amylase activity which is what finally affects the quality of the wheat

flour. As a result, the percentage assessment of sprout damage is no longer the best indicator the milling and baking quality of a wheat sample [Canadian Wheat Board, 2002]. A more accurate and complex method is to measure the wheat falling number (FN), an index correlated with the content of alpha-amylase. The bakers' and other customers now prefer buying on the basis of FN rather than the percentage of sprout damage [Canadian Wheat Board, 2002].

The Hagberg FN test is the internationally accepted measurement of sprout damage in wheat due to the amount of alpha-amylase [Canadian Wheat Board, 2002]. The FN is the time in seconds for a stirrer to fall through the hot slurry of milled wheat. The theory behind the test is that as the enzyme acts on the slurry, the slurry will become thinner. If the slurry is thinner, it will not be able to hold the weight of the plunger. Therefore, the more enzymes, the faster the plunger will fall. A high FN (or the longer it takes the stirrer to fall) indicates that the wheat is sound and satisfactory for most baking processes. A No.1 Canada West Red Spring (CWRS) normally has a FN greater than 350 seconds. The steps of Hagberg FN test are described below [Perten Instruments, 2005]:

1. *Sample Preparation*

   A 300 gram sample of wheat kernels is ground in a laboratory miller equipped with a 0.8 mm sieve.

2. *Weighing*

   7.0 ± 0.05g of flour, based on a 14% moisture basis, is weighed and put into a viscometer tube.

3. *Dispensing*

   25 ± 0.2ml of distilled water is added to the tube.

4. *Shaking*

   Sample and water are mixed by vigorously shaking the tube to obtain a homogeneous suspension.

5. *Stirring*

The viscometer tube with the stirrer inserted is put into the boiling water bath and the instrument is started. After 5 seconds the stirring begins automatically.

6. *Measuring*

The stirrer is automatically released in its top position after 60 (5 + 55) seconds and allowed to fall under its own weight.

7. *The Falling Number*

The total time in seconds from the start of the instrument until the stirrer has fallen a measured distance (including the 60 second stirring time) is registered by the instrument. This is the FN.

Since 2001 the CGC has been developing another method, the Rapid Viscosity Analysis (RVA) [CGC, 2003], which is based on the same principle as the Hagberg FN test. In the RVA method, steps in preparing wheat samples are similar to sample preparation in the Hagberg FN method. It uses a smaller sub-sample, 4 grams of flour. Instead of measuring the time of falling, a paddle is attached to a rotating calibrated motor which measures the force required to spin the paddle in the slurry. This produces a measurement of the viscosity of the slurry in scientific units, i.e., centipoises [3]. Compared with the FN test, the RVA test is faster. It can be completed in three minutes, independent of the degree of sprout damage. The Hagberg test, however, can take over seven minutes to complete. The grain research laboratory of the CGC has evaluated RVA under laboratory conditions using CWRS wheat. The Laboratory has generated a strong calibration formula between viscosity at three minutes and the Hagberg FN value. At present, this method is being evaluated under operating conditions in some country elevators and terminals in Canada [CGC, 2003].

However, both the Hagberg FN test and RVA are wet chemistry methods. They are labor intensive and need a lot of money and time in training the technical staff. Although both methods have been standardized and are performed according to the description in each step, the involvement of people introduces operator variability at each

step of the sample preparation. In addition, both methods are implemented in laboratories and are not easy to implement at elevator and terminal facilities at the ports where space for specialized laboratory equipment is limited and rapid turnaround is key. As a result, it is desirable in the market to develop a rapid and objective method which would ideally be able to test FN in a more rapid, less labor intensive, and more precise manner.

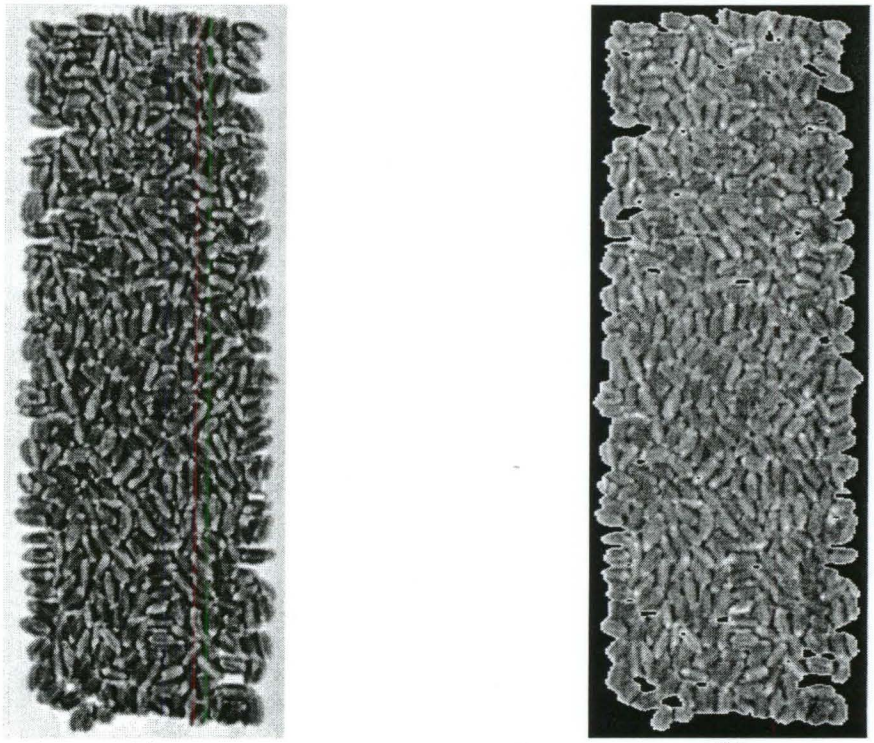
The project in this chapter is designed in response to this desire of the wheat industry. It considers the feasibility of using NIR imaging of the wheat kernel samples to predict the FN using the method of Multivariate Image Regression (MIR). The organization of this chapter is as follows: In section 5.2, after a short introduction of the available wheat samples and the imaging procedure, features that are representative of the sample characteristics are extracted from the image. In section 5.3, the regular PLS algorithm, an O-PLS algorithm and a PLS+CCA algorithm are employed respectively to build the reference model between the extracted features and the FN measurements. Some useful information is obtained by analyzing the related coefficients and loadings of the models. The sources of the prediction error are analyzed and the performance of the algorithms is compared. In Section 5.4, the main result is summarized and the conclusion is given. Possible research topics for the future are also discussed.

5.2 NIR Imaging of the Samples of Wheat Kernels and Feature Extraction

The wheat samples used in this study include 27 packages of CWRS wheat kernels with measured FNs tested at *Intertek*, a wheat quality control company in Winnipeg. The FN range is between 190 seconds and 400 seconds. For each sample, four tablespoons of kernels (around 12 grams in one tablespoon) are imaged separately to acquire 4 images. Totally, 108 images are taken for the 27 samples. During imaging, each tablespoon of wheat was evenly distributed in a rectangular area of 136mm (L) x 50mm (W) on a white tile (as shown in figure 5.1(a)) and imaged under the NIR imaging spectrometer with the scanner bed moving at a speed of 7mm/sec. NIR reflectance images

were captured with the spatial resolution of 650 pixels (y) x 126 pixels (x) and spectral resolution of 110 unique wavelengths spanning the 933nm – 1663 nm range.

The shadow influence and background are removed by selecting a threshold at the wavelength of 1450nm. The spectral values of the pixels with an intensity larger than 0.5 at this wavelength were set as zero on the whole wavelength range from 933nm to 1663 nm, and these pixels are excluded from the subsequent analysis. This background removal is illustrated in Figure 5.1.



(a) (b)

Figure 5.1 (a) Monochromatic intensity image of a sample at wavelength around 1200 nm (b) Monochromatic intensity image of a sample at wavelength around 1200 nm after removing the shadow influence and background

Extracting relevant feature information from the NIR image is the crucial step of the overall MIR modeling scheme. The features extracted from the image should be representative of the overall characteristics of the wheat seed sample in the image as well as predictors of chemical information affecting the wheat property.

Two features are extracted from the images in this study. Both of the two features extracted are based on the assumptions that: (1) a weighted average NIR reflectance spectrum over the scanned area of the sample is an adequate feature vector representing overall chemical information, and (2) the chemical information captured by the feature vector is indicative of the FN.

The first feature is the average spectrum of all the pixels of the kernels in the image. Because the variation of the spectrum is related to absorbance signatures of various functional groups, this feature is a good indicator of the variation in the organic content of the seeds, which is the chemical basis of wheat FN variation.

The second feature is the 1st MPCA loading vector of the image. No mean centering of the image is performed in MIA. Therefore, the 1st PC explains the mean variability throughout the spectral data. Thus the 1st loading vector **p**1 represents a normalized mean spectrum of all pixels throughout the un-scaled multivariate image.

The MIR models based on these two features gave almost the same results. In the following sections, only the results using the first feature are shown and discussed.

5.3 PLS Regression Modeling of FN Using Features from NIR Wheat Kernel Images

After extracting the features, a PLS inferential model is built by regressing these features against the FN measurements of the samples. 14 samples (labeled as 1, 3, ... , 27) are used for the training set, the other 13 samples (labeled as 2,4,...26) for the test set. The procedure of the MIR scheme for this study is illustrated in figure 5.2.

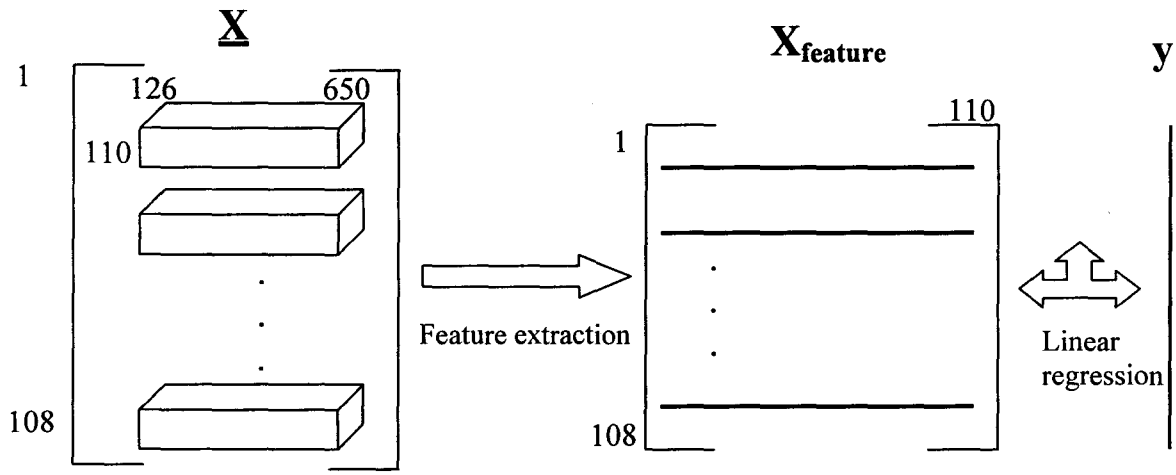


Figure 5.2 Schematic of proposed MIR strategy for predicting FN (**y**) from multi-spectral NIR wheat seed images (**X**)

To reduce non-linearity and improve PLS model fit and prediction, the features are transformed by taking the negative logarithm ($-\log_{10}\mathbf{X}_{feature}$), which means that the reflectance average spectra are transformed to the absorbance average spectra.

Prior to application of PLS regression modeling, the NIR spectral data in $\mathbf{X}_{feature}$ is mean-centered with respect to the 110 wavelengths (columns). The data of **y**, i. e. the measurements of the FNs, are mean-centered and auto-scaled to unit variance.

5.3.1 Regular PLS model

Table 5.1 shows the summary of the regular PLS1 model. This model is good in that the amount of variation explained ($R^2(x)_{cum}$) is 86.8% and the amount predicted (by cross-validation, $Q^2(y)_{cum}$) is 71.2%. The information from the model indicates that there exists in $\mathbf{X}_{feature}$ a lot of unrelated variation with regard to **y**. Firstly, it can be seen that ***five*** PLS components are employed in the model. Only ***one*** component should be enough for the PLS1 model if there were no y-unrelated variation in $\mathbf{X}_{feature}$. However, in this case, extra PLS components are needed to explain the high uncorrelated variance and lead to a more complex model. Secondly, the variance explained in **y**, $R^2(y)_{cum}$ and the cross-validated $Q^2(y)_{cum}$ are quite low in the first two components compared with the variance explained in $\mathbf{X}_{feature}$, $R^2(x)_{cum}$. This means that the first two PLS components are

employed mainly to explain the **y**-unrelated variation in $\mathbf{X}_{\mathbf{feature}}$. It is also noticed that the 95% confidence intervals of most PLS1 regression coefficients (Figure 5.3) are very wide and many include zero, which makes the interpretation of the original PLS model ambiguous. Hence, $\mathbf{X}_{\mathbf{feature}}$ is preprocessed to eliminate the **y**-uncorrelated variation to simplify the PLS1 model and to improve its interpretability.

Original PLS1 model			
PLS comp.	$R^2(x)_{cum}$	$R^2(y)_{cum}$	$Q^2(y)_{cum}$
1	0.907	0.159	0.0351
2	0.998	0.231	0.058
3	0.999	0.586	0.433
4	0.999	0.83	0.698
5	0.999	0.868	0.712

Table 5.1 Summary of the original PLS1 model

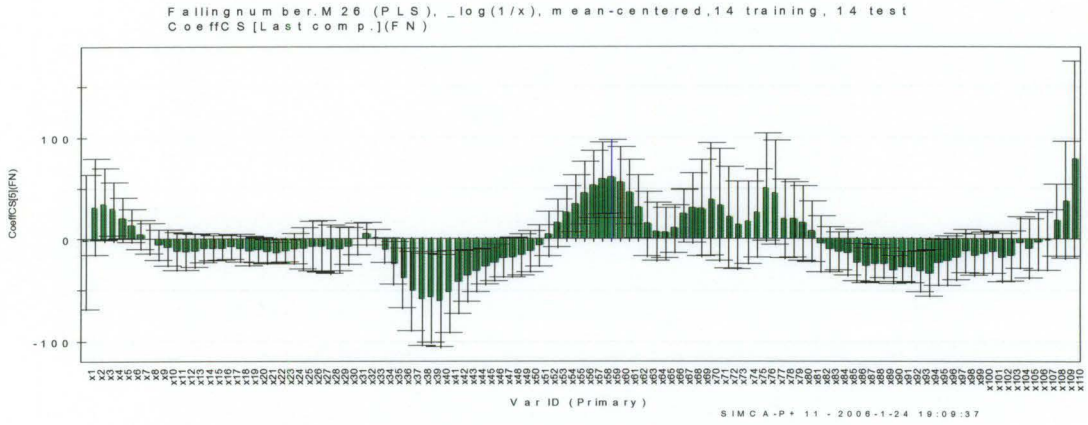


Figure 5.3 Coefficients of the regular PLS1 model

O-PLS and PLS+CCA are employed to solve this problem. These two methods are chosen because both of them overcome the over-fitting problem encountered in the OSC+PLS algorithm proposed by Wold [1998] where cross-validation and eigenvalue

criteria cannot be easily implemented to decide the appropriate number of orthogonal components.

5.3.2 O-PLS model

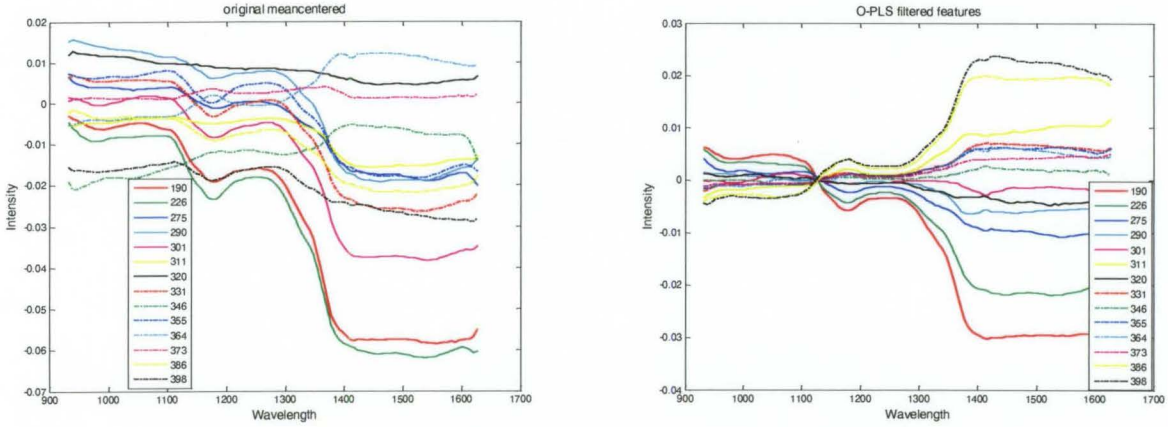
The O-PLS model is summarized in table 5.2. It shows that 84.05% of the total variation in $\mathbf{X}_{\mathbf{feature}}$ is removed by the four O-PLS components orthogonal to **y**. Figure 5.4 shows the original and the O-PLS preprocessed feature spectra of fourteen images. It is easily observed from comparing figure 5.4(a) and (b) that the irrelevant baseline variations have been greatly reduced. It can also be seen that the difference between the features are more consistent with the FN variation between the samples. One can see the consistent trend of FN change from the plots in Figure 5.4(b).

The number of the orthogonal components is decided by the ratio $\|\mathbf{w}_{ortho}\|/\|\mathbf{p}\|$, which represents the percentage of the y-uncorrelated variation with respect to the y-covariation in the feature space. It converges to zero if a "proper" number of orthogonal components has been calculated. The ratio $\|\mathbf{w}_{ortho}\|/\|\mathbf{p}\|$ of this study is shown in Figure 5.5, which indicates 4 orthogonal components are reasonable.

A new PLS1 model is built between the 15.95% variation left in $\mathbf{X}_{\mathbf{feature}}$ and the measurements of the FN. This PLS1 model has only one PLS component, which indicates that almost all the uncorrelated variation in $\mathbf{X}_{\mathbf{feature}}$ has been removed. Note that, as expected, the percent of the some of squares explained by the fit $(R^2(x)_{cum})$ is exactly the same as for the regular five component PLS1 model (Table 5.1). It is observed that $Q^2(y)_{cum}$ of the O-PLS model is a little larger than one in the original PLS1 model. In fact, O-PLS does not contribute to improving the predictability of the model. The larger $Q^2(y)_{cum}$ here is because it is the $Q^2(y)_{cum}$ only calculated for the PLS1 step of the O-PLS. If the cross-validation is also used for the orthogonal component extraction step in the O-PLS, the $Q^2(y)_{cum}$ will be almost the same as in the original regular PLS1 model.

O-PLS model			
Comp.	$R^2(x)_{cum}$	$R^2(y)_{cum}$	$Q^2(y)_{cum}$
Ortho. Comp. 1	0.3988		
Ortho. Comp. 2	0.76362		
Ortho. Comp. 3	0.8331		
Ortho. Comp. 4	0.8405		
PLS Comp. 1	0.1595	0.868	0.861

Table 5.2 Summary of the O-PLS1 model



(a) (b)

Figure 5.4 (a) Plots of 14 mean-centered original features, (b) Plots of 14 mean-centered features filtered by O-PLS

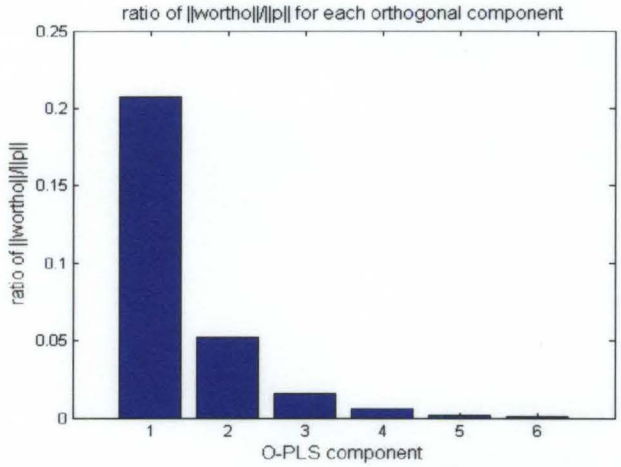


Figure 5.5 Ratio of $||\mathbf{w}_{ortho}||/||\mathbf{p}||$ for each O-PLS component

The coefficients of the O-PLS model are plotted in figure 5.6. These coefficients correspond to the O-PLS filtered data (the data illustrated in figure 5.4(b) for example) and not the raw feature data (the data illustrated in figure 5.4(a) for example). It can be observed that the coefficients are not so noisy as in the original PLS1 model. Since most of the variation in the feature that is orthogonal to y has been removed, the O-PLS model built on the filtered data is focused only on the subspace that is highly correlated with y. It shows that all the filtered variables on the whole wavelength range from 933 nm to 1663 nm are correlated with the FN. The fist 31 filtered variables in the low wavelength region (around 933 nm – 1133nm) are negatively and the other filtered variables are positively correlated with the FN. The filtered variables over the high wavelength range contribute more to the FN prediction than the other variables.

It is important to note that the O-PLS is between the **filtered** x variables and the original y variable. Thereby, any information observed by analyzing the coefficients or the loadings of the parsimonious model is about **the variables in the filtered (correlated) subspace, not** the original variables. If the information related to the original variables is needed, for example what are the interesting wavelengths for prediction of the response variable y, the coefficients or the loadings of the PLS model for the original variable space have to be analyzed. The only advantage of O-PLS is to allow for interpretation of the orthogonal and correlated subspace.

From table 5.2, it is observed that most of the y-uncorrelated information in $\mathbf{X}_{\text{feature}}$ is explained by the first two orthogonal components. The loading vectors of these two components are shown in figure 5.7 and 5.8 respectively. Figure 5.7 indicates that the first orthogonal component corresponds to the average baseline shift over the whole wavelength range among the features. Figure 5.8 indicates that the second orthogonal component explains a contrast between on the low wavelength range and on the high wavelength range, or a baseline trend that is not correlated with y.

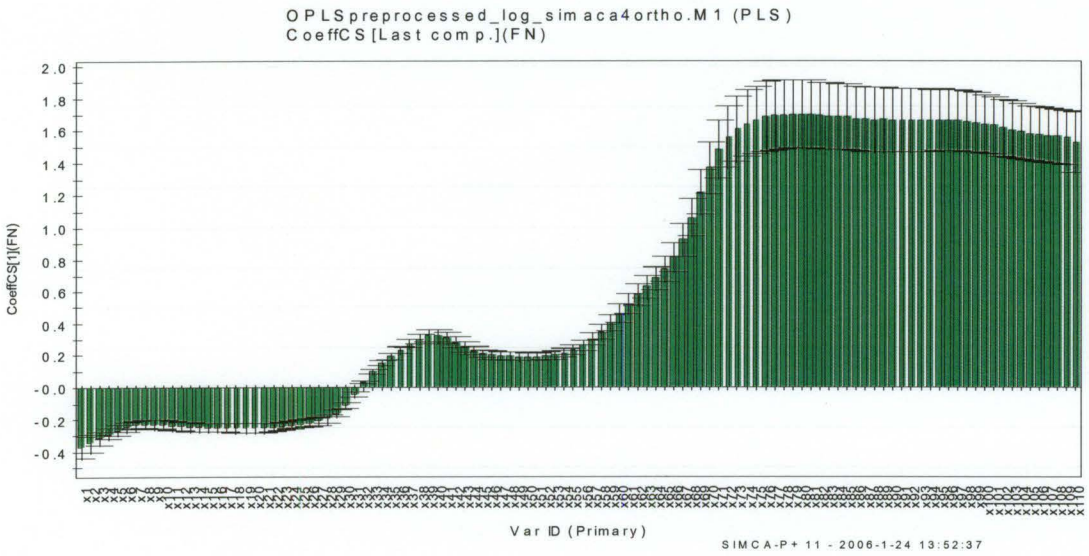


Figure 5.6 Coefficients of the PLS model built using the filtered data

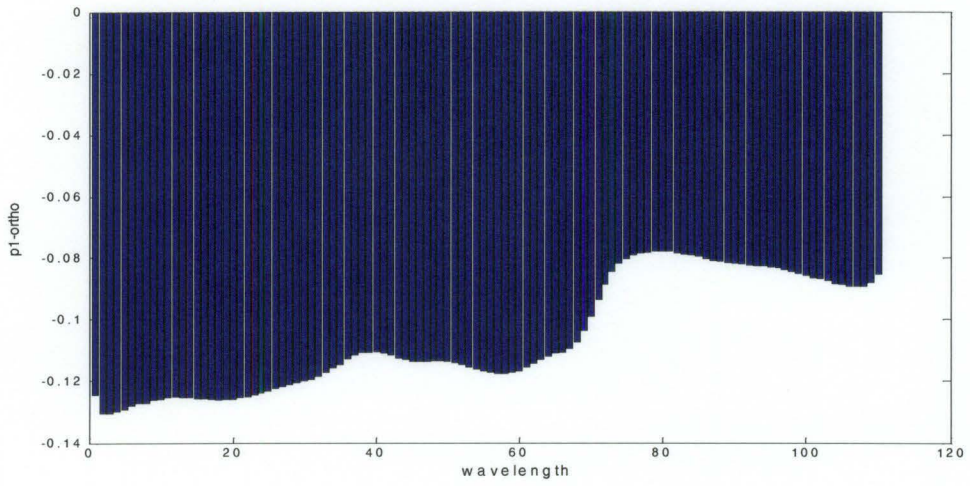


Figure 5.7 Loading plot of the first orthogonal component of the O-PLS model

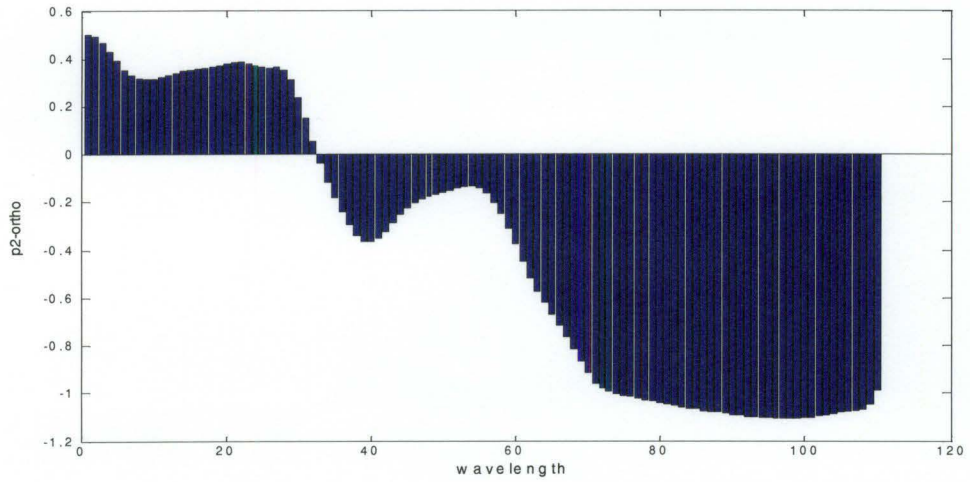


Figure 5.8 Loading plot of the second orthogonal component of the O-PLS model

The O-PLS model is tested by the test set including 52 images of 13 samples (four images for each sample) with different FN measurements. The predictions of the O-PLS model vs. the measurement of the FNs are shown in figure 5.9. Every four markers at the same horizontal level in the figure denote four images of each sample. They have the same FN measurements. One can see that both the fit of the training set and the prediction of the test set are not bad. The root mean square error of prediction (RMSEP) of the test set is 26.56 seconds. It is worthy to be pointed out that the RMSEP represents the error originated from three sources:

1. The error from the sample-to-sample variation,
2. The error from the reference model,
3. Large errors in the FN measurements themselves.

The standard deviation of the four predictions for each sample in the test set is calculated and plotted in figure 5.10. The average of the standard deviations is 13.4 seconds. It indicates that approximately half of the prediction error (described as RMSEP) is from the sample-to-sample variability. This is easy to understand because of severe

nonlinearity between sprout kernels and FN. Due to the widely varying amount of alpha amylase in a single severely sprouted vs. a sound kernel, the activity ranges from 5000 units to literally 5 units [Hatcher, 2006], the difference between two samples say just one single severely sprouted kernel would introduce a large error to the prediction of the O-PLS model which is based on the average spectrum of the pixels.

The average of the four predictions for each sample in the test set is calculated and plotted vs. the measurement in Figure 5.11. It can be observed that the average of the predictions follows the variation trend of the FN measurements reasonably well. The standard deviation of the average predictions from the measurement is 17.7 seconds. This value can be approximately looked upon as the prediction error with little sample-to-sample variation. In the wheat industry, the acceptable deviation of the FN measurements without the sample-to-sample variation (That means the error is only from the error from step 2 –step 6 in the Hagburg FN test) is 20 seconds [Hatcher, 2006]. Compared with this standard, 17.7 seconds is acceptable implying that the NIR prediction precision is close to its lowest achievable bound, namely the standard deviation of the FN measurement itself.

In view of the fact that there is severe nonlinearity of the alpha amylase activity among the kernels with different degree of sprouting, the average spectrum of the pixels will not be the best feature to describe the chemical variation of the sample. To count for this nonlinearity one option is to predict the FN by extracting feature from separate kernels and then recombining them in a nonlinear manner. A problem is that the FN measurement is not evaluable for each kernel, only for a sample taken from the batch. Therefore, a nonlinear model relating the single kernel features to this overall sample FN will have to be built. This idea will be followed up in the future.

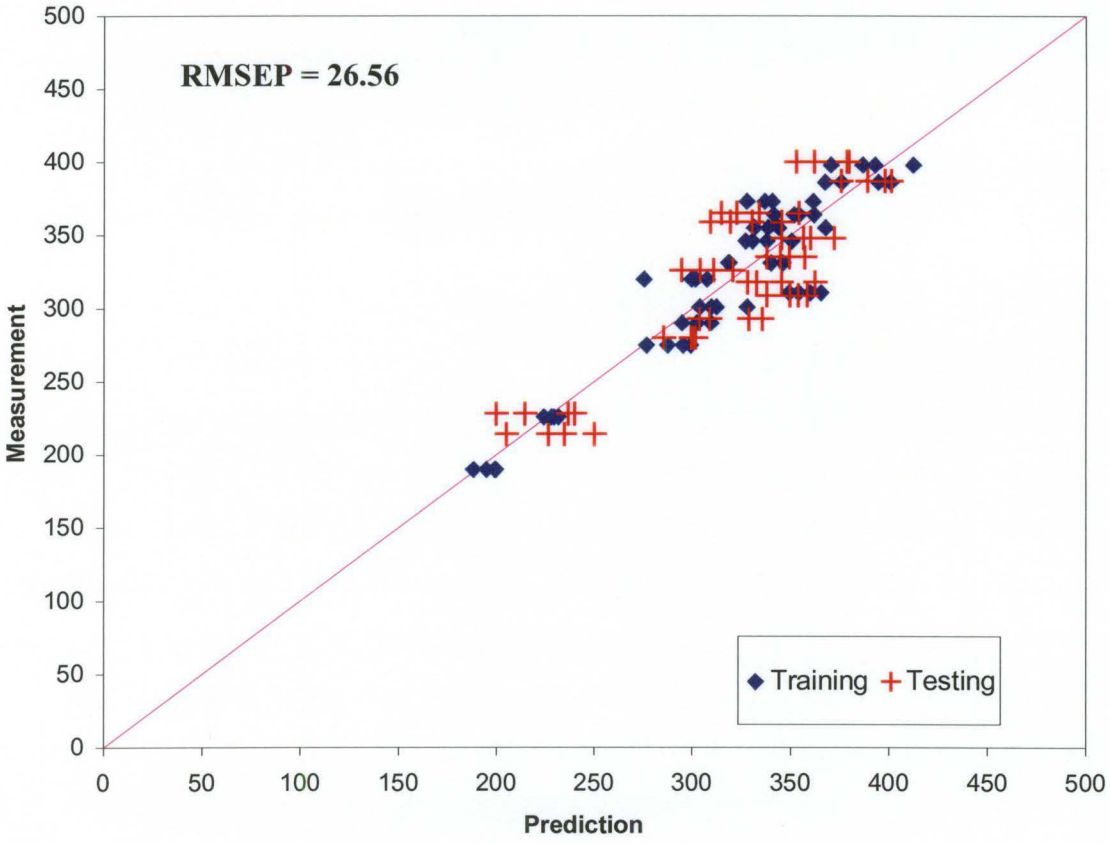


Figure 5.9 O-PLS model prediction vs. measurement

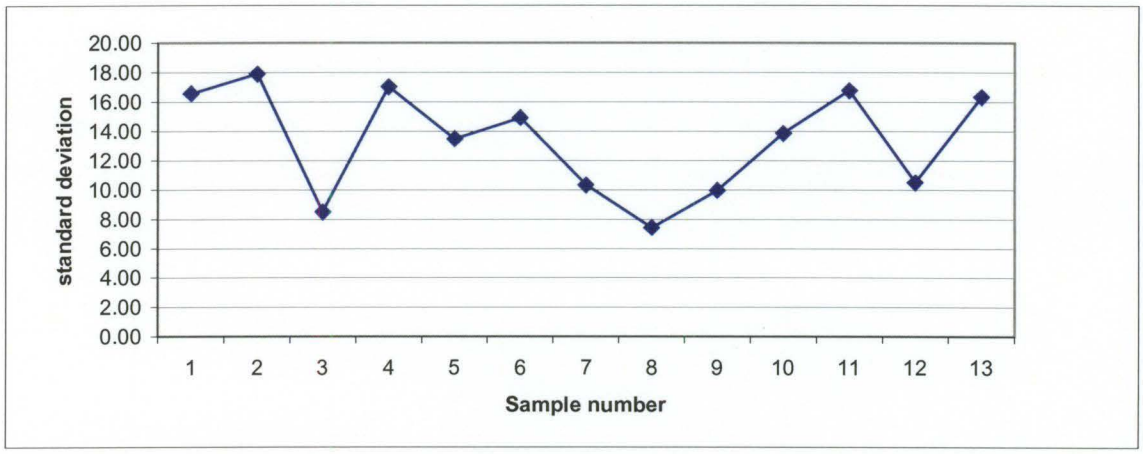


Figure 5.10 The standard deviation of the four predictions for each sample in the test set (approximate sample-to-sample variability)

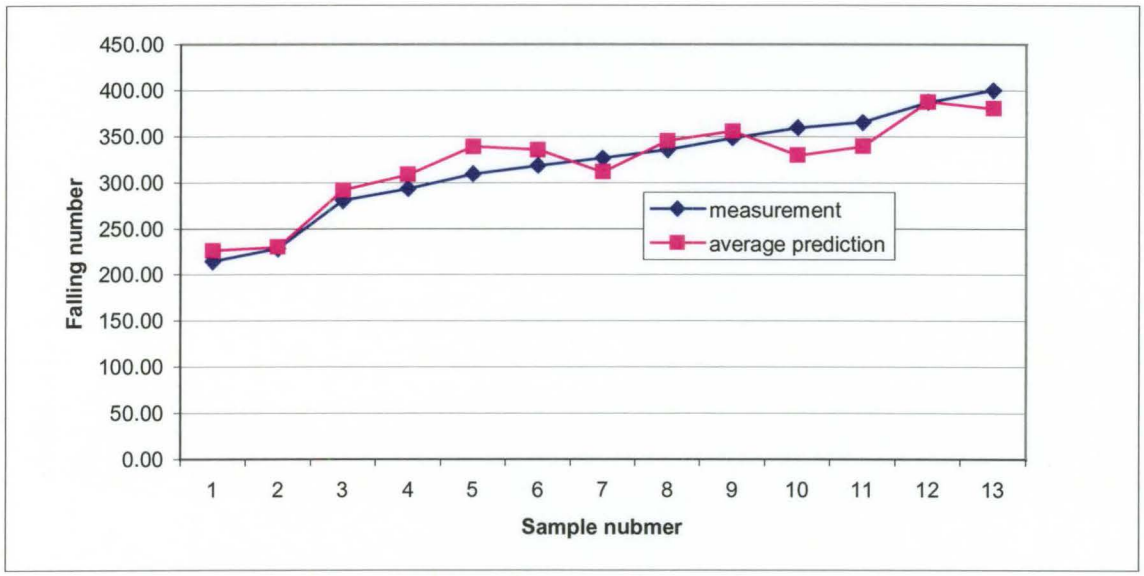


Figure 5.11 The average of the four predictions for each sample in the test set vs. the measurements

5.3.3 PLS +CCA

The post processing method PLS+CCA was proposed by Yu and MacGregor [2004] as an alternative method to OSC + PLS. The incentive of this algorithm is the fact that all preprocessing methods based on OSC have the risk of overfitting except for O-PLS and Fearn's OSC [Fearn, 2000] in the case of one single response variable. Instead of using OSC as a preprocessing step to remove the orthogonal information from the predictor matrix **X**, A CCA is performed between the response matrix **Y** and the score matrix **T** obtained from an initial PLS model between **X** and **Y**. The initial PLS step is employed to avoid ill condition problem when using CCA directly.

In fact, O-PLS and PLS+CCA are two opposite operations. The OSC step in O-PLS remove the uncorrelated variation from the PLS components; on the contrary, the CCA step in PLS+CCA extracts only the correlated variation from the significant PLS components of the initial PLS model.

A canonical correlation regression model is built between the five score matrix of the PLS model summarized in table 5.1 and the FN measurements. Since there is only one y variable, the CCA model has one canonical covatiate.

The y-related part of the feature matrix can be estimated as

$$\mathbf{X}_{corr} = (\mathbf{t}_{cca}\mathbf{p}^{T}_{cca})\mathbf{P}^{T}_{pls}$$

Where $\mathbf{t}_{cca}$ is the score vector and $\mathbf{p}_{cca}$ is the loading vector of the CCA analysis, and $\mathbf{P}_{pls}$ is the loading matrix of the initial PLS model. The correlated part of the 14 features is shown in figure 5.12. It can be observed that these plots resemble the plots in figure 5.4(b), but a little less noisy. This is because the estimation from the above formula is based on the significant components of the initial PLS model, which is not as noisy as the original features.

The coefficient of the CCA model is shown in Figure 5.13. It shows that the last three PLS components are more correlated with the FN than the first two PLS components. This is consistent with the analysis of the PLS model in section 5.3.1. The first two PLS components mainly explain the y-unrelated variation in $\mathbf{X}_{feature}$.

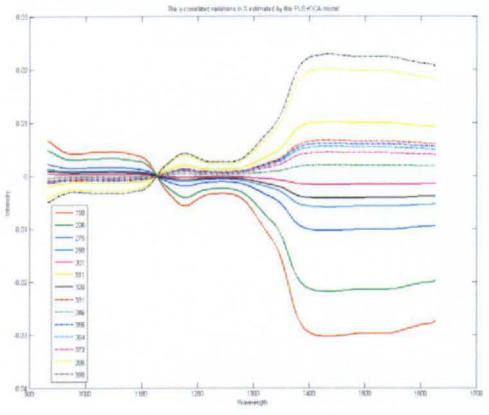


Figure 5.12 the **y**-correlated variation in $\mathbf{X}_{feature}$ estimated by the PLS+CCA model

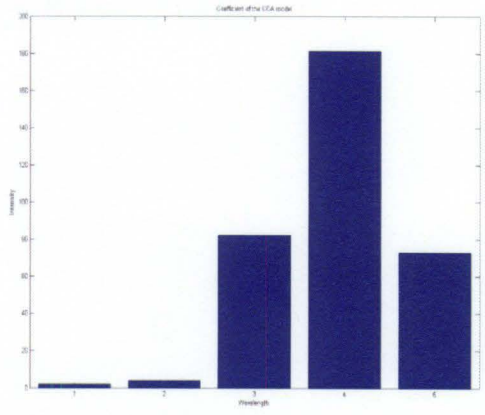


Figure 5.13 The coefficients of the CCA sub-model in PLS+CCA

The PLS+CCA model is also tested by the same test set as in the O-PLS model. The result is shown in Figure 5.14, The RMSEP is 27.29, a little larger than in the O-PLS model.

In this study, there is not much difference between the results of O-PLS and PLS+CCA. However, it is worthy to be pointed out that PLS+CCA shows evident advantage over O-PLS when there are more than one y variable [Yu and MacGregor, 2004]. In this situation, there are no established rules to follow to decide how many OSC components should be removed in O-PLS.

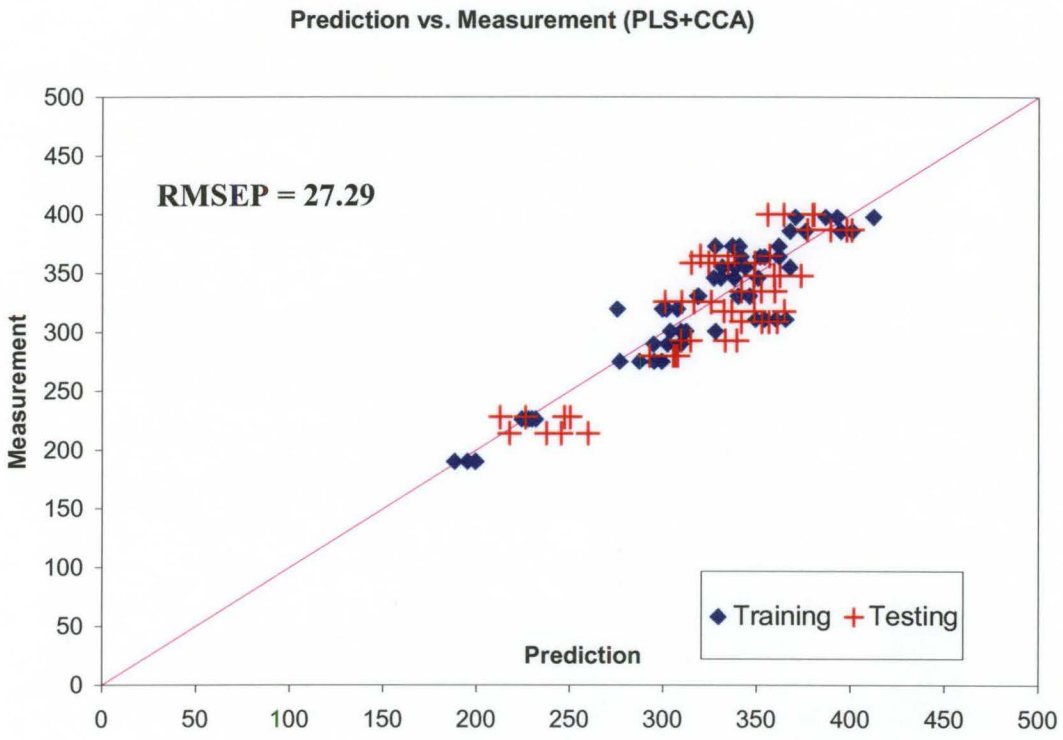


Figure 5.14 PLS+CCA model prediction and measurement

5.4 Conclusions and Discussion

In this chapter, multivariate image regression methodology has been employed to develop models to predict the FN of wheat samples. This is the first attempt to measure this wheat property using NIR imaging technology based on a kernel sample. The average spectrum of the sample kernels in the NIR image is calculated as the feature representing the chemical character of the sample. PLS algorithms are employed to develop the reference models. The sources of the prediction error are analyzed. It is found approximately half of the prediction error originates from the sample-to-sample variation. Except for this part of error, the prediction error variance of the O-PLS models comparable to the variance of the standard Hagberg FN test in the wheat industry. In conclusion, NIR imaging technology is a feasible and promising means to measure the wheat falling number. However, the method still requires more development to increase its robustness and accuracy.

O-PLS and PLS+CCA algorithms are also tried and compared with the PLS algorithm. All three algorithms had essentially the same performance in this case study. Both O-PLS and PLS+CCA were initially proposed to simplify the original PLS model with many components and increase the interpretability of the model. However, it is worthy to be pointed out that the parsimonious models obtained from both these methods are in the filtered subspace, not in the original variable space. Hence, the information obtained by analyzing the parsimonious model is about the subspace. However, the information about the original variable space is usually needed in practice. In this situation, one has to go back to the original PLS model or combines the two steps in either of the two methods to mine such information.

To reduce the impact of nonlinearity effect on the prediction of the reference model, one promising method is to evaluate the FN on the level of separate kernels and then combine them. This idea will be implemented in the future. A NIR imaging spectrometer with higher resolution is necessary to examine separate kernels. In addition,

it is observed that there is texture variation on the kernel surfaces having sprouts. This feature can be extracted from images with higher resolution and used as part of the feature vector in modeling.

Chapter 6

Summary and Conclusions

6.1 Summary and Conclusions

Two topics related to the NIR imaging technology are studied in this thesis. One is on the calibration of line-scan NIR imaging systems, the other covers the feasibility of applying the NIR imaging technology for wheat grading.

In the first study, a simple methodology, that is not based on expensive standard reflectance materials is proposed to calibrate line-scan NIR imaging systems. The main idea of this method is to obtain a calibration model for each sensor on the InGaAs CCD array. The models are able to correct almost all the errors caused by the inconsistency between the sensors along the spatial axis x and the spectral axis λ, and the variation in the illumination along the spatial axis x. The calibration results of the image are shown. Some detailed information blurred by the systematic noise from the imaging system can be clearly visualized after calibrating the image. The benefit of this method to the multivariate image analysis is illustrated with an example. The method can be used to calibrate any line-scan imaging system. The advantage of this method is that it does not employ the expensive standard reflectance materials which are also difficult to maintain.

In the second study, two projects are accomplished:

In chapter 4, NIR imaging is used to classify different classes of wheat kernels. The average spectrum of the kernel is used as its feature. In section 4.2, SIMCA and PLS-DA is used respectively to separate the sprouted kernels from the healthy ones. The two distances used in SIMCA is compared. It is shown that the "combined distance" is better than the simple Euclidean distance to the model. The results of SIMCA and PLS-DA are

compared. PLS-DA works a little better than SIMCA in classifying the kernels within these two classes. However, SIMCA has the advantage over PLS-DA in discriminating the kernels from other class. In section 4.2, multi-class PLS-DA is used to classify four classes of grain. A new strategy of implementing multi-class PLS-DA algorithm, the *one-vs-one strategy* is proposed in this part. It is shown that this strategy gives better result than the traditional multi-class PLS-DA, which is virtually a *one-vs-rest strategy* implementation. The good results from this study show that NIR imaging provides a potentially fast and objective method for qualitatively evaluating certain characteristics of wheat samples, such as fungal infection, sprout damage and foreign types of grain, which are now graded manually in the wheat industry.

In chapter 5, the NIR images of wheat kernels are used to predict the FNs of the bulk wheat samples. Three models, regular PLS, O-PLS and PLS+CCA, are built between the features extracted from NIR images of the wheat kernels and the falling number measurements made on bulk samples. The interpretability of the O-PLS is discussed. It is pointed out that the only advantage of O-PLS is to allow for interpretation of the orthogonal and correlated subspace. The errors in the RMSEP of the O-PLS model are analyzed. It is shown that half of the prediction error is from the sampling errors. Except for that, NIR prediction precision is close to the standard deviation of the FN measurement itself. The results from this study indicate that NIR imaging is a promising alternative tool other than the current Hagberg FN test for fast, objectively and non-destructively measuring the FN index of wheat samples.

6.2 Future work

In this thesis, only spectral features from the NIR images are used for classification or prediction. It is noticed that there are texture features on certain classes of wheat kernels, the wrinkle on the severely sprouted kernel surface for example. The texture features are also related to the chemical variation in the kernels and should be helpful for the classification or prediction. Because of the low resolution of the spectrometer used in this thesis, this feature cannot be extracted and used in this thesis. In

the future, this idea can be implemented by using a NIR imaging spectrometer with higher resolution.

To reduce the impact of nonlinearity of the alpha-amylase variation among the sprouted kernels on the prediction of the FN, one promising method is to evaluate the FN on the level of separate kernels and then combine them. A NIR imaging spectrometer with higher resolution is necessary to examine separate kernels.

References

Anderssson, C. A., "Direct orthogonalization", Chemometrics and Intelligent Laboratory Systems, **47**, pp. 51-63, 1999

Barker, D.A., T.A. Vouri, M.R. Hegedus and D.G. Myers, "The Use of Ray Parameters for the Discrimination of Australian Wheat Varieties", Plant Varieties and Seed, **5(1)**, pp. 35-45, 1992

Barnes, B.J., Dhanoa, M.S. and Lister, S.J., "Standard Normal Variate Transformation and De-trending of Near Infrared Diffuse Reflectance Spectra", Applied Spectroscopy, **43**, pp. 772-777, 1989

Bharati, M. H., "Multivariate Image Analysis and Regression for Industrial Process Monitoring and Product Quality Control", Doctoral Thesis, Mcmaster University, Hamilton, 2002

Burger, James and Geladi, Paul, "Hyperspectral Image Regression Part I: Calibration and Correction", Journal of Chemometrics, **19**, pp. 355-363, 2005

Canadian Wheat Board, "Sprout Damage Causes Milling Concerns", http://cwb.ca/en/publications/farmers/nov-dec-2002/11-12-02-04.jsp, 2002

CGC, "CGC reminds producers: Don't Blend Sprout-damaged Wheat with Sound Wheat", http://www.grainscanada.gc.ca/newsroom/news_releases/2002/2002-09-27-e.htm, September, 2002

CGC, "Evaluation of a Rapid Test for Sprout Damage Underway" http://www.grainscanada.gc.ca/qualit_matter/rva-e.htm, 2003

CGC, "Organization and Operations of Canadian Grain Commission", http://www.grainscanada.gc.ca/pubs/OrgOps/orgops-e.pdf, 2005

CGC, "Sprouted and Severely Sprouted Kernels"
http://www.grainscanada.gc.ca/FAQ/sproutdamage-e.htm, Last updated: March, 6, 2006

Delwiche, S.R. and Norris, K.H., "Classification of Hard Red Wheat by Near-infrared Diffuse Reflectance Spectroscopy", Cereal Chemistry, **70(1)**, pp. 29-35, 1993

Dexter, J. E. and Edwards, N. M., "Implications of Frequently Encountered Grading Factors on the Processing Quality of Common Wheat", http://www.grainscanada.gc.ca/PUBS/confpaper/Dexter/Grading212/dgrading2-3c.htm, Last updated: Feb. 6,1998

Eigenvector Research Inc. "PLS_Toolbox 3.5 for Use with Matlab", 2004

Eriksson, L., Johansson, E., Wold, N.K. and Wold, S., "Introduction to Multi and Megavariate Analysis Using Projection Methods (PCA and PLS)", Umetrics AB, Umeå, Sweden, pp. 490, 1999

Fearn, T., "On Orthogonal Signal Correction", Chemometrics and Intelligent Laboratory Systems, **50**, pp. 47-52, 2000

Fearn, T., "Standardization and Calibration Transfer for Near Infrared Instruments: a Review", Journal of Near Infrared Spectroscopy, **9**, pp. 229-244, 2001

Feudale R, Woody N, Tan H, Myles A, Brown S and Ferre, J, "Transfer of Multivariate Calibration Models: a Review", Chemometrics and Intelligent Laboratory Systems, **64**, pp. 181-192, 2002

Gegahertz-Optic, "Calibration Standards and Uniform Light Source"
http://www.gigahertz-optik.de/pdf/catalogue/Calibration_Standards.pdf, 2005

Geladi, P., MacDougall, D. and Martens, H., "Linearization and Scattercorrection for Near Infrared Reflectance Spectra of Meat", Applied Spectroscopy, **39**, pp. 491-500, 1985

Geladi, P., Burger, J. and Lestander, T., "Hypespectral Imaging: Calibration Problems and Solutions", Chemometrics and Intelligent Laboratory Systems, **72,** pp. 209-217, 2004

Geladi., P. and Grahm, H., Multivariate Image Analysis, John Willey & Sons,1996

Hatcher, D., Personal E-mail Communication with Dr. Hatcher at Canadian Grain Commission, 2006

Hyvarinen, T., Herrala, E. and Dall'Ava, A., "Direct Sight Imaging Spectrograph: A Unique Add-on Component Brings Spectral Imaging to Industrial Applications Proceedings of SPIE – Digital Solid State Cameras: Designs and Applications", San Jose, CA, January 25-30,1998

ImSpector, ImSpector User Manual
http://www.specim.fi/pdf/User_Manual%2BApp_2_21.pdf, 2003

Isaksson, T. and Kowalski, B., "Piece-wise Multiplicative Scatter Correction Applied to Near-infrared Diffuse Transmittance Data from Meat Products", Applied Spectroscopy, **47**, pp. 702-709, 1993

Lewis, E. N., Schoppelrei, J., Lee, E., "Near-Infrared Chemical Imaging and the PAT Initiative", Spectroscopy, **19(4)**, pp. 26-36, 2004

Majumdar, S., D.S. Jayas, J.L. Hehn and N.R.Bulley, "Classification of Various Grains Using Optical Properties", Canadian Agricultural Engineering, **38(2)**, pp. 139-145, 1996

Majumdar, S. and D.S. Jayas., "Classification of Bulk Samples of Cereal Grains Using Machine Vision", Journal of Agricultural Engineering Research, **73**, pp. 35-47, 1999

Majumdar, S. and D.S. Jayas., "Classification of Cereal Grains Using Machine Vision: I. Morphology models", Transactions of the ASAE, **43(6)**, pp. 1669-1675, 2000a

Majumdar, S. and D.S. Jayas, "Classification of Cereal Grains Using Machine Vision: II. Color Models", Transactions of the ASAE, **43(6)**, pp. 1677-1680, 2000b

Majumdar, S. and D.S. Jayas. "Classification of Cereal Grains Using Machine Vision: III. Texture Models", Transactions of the ASAE, **43(6),** pp. 1681-1687, 2000c

Majumdar, S. and D.S. Jayas. "Classification of Cereal Grains Using Machine Vision: IV. Morphology, Color, and Texture Models", Transactions of the ASAE, **43(6)**, pp. 1689-1694, 2000d

Martens, H. and Næs, T., Multivariate Calibration, John Wiley and Sons Ltd, Guildford, Great Britain, pp. 419, 1989

McClure, W.F., "Near-Infrared Spectroscopy: the Giant Is Running Strong", Analytical Chemistry, **66**, pp. 43A-53A, 1994

Minsky, M. and Papert, S., "Perceptrons: An Introduction to Computational Geometry", MIT Press, 1969

Næs, T., Isaksson, T., Fearn, T. and Davies, T., A User Friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK., pp. 344, 2002

Norris, K.H. and Hart, J.R., "Direct Spectrophotometric Determination of Moisture Content of Grain and Seeds", in Principles and Methods of Measuring Moisture Content in Liquids and Solids (ed. A. Waxler), Reinhold Publishing Corporation, New York, Vol. IV, pp. 19-25, 1965

Osborne, B.G., Fearn, T. and Hindle, P.H., Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, Longman Scientific and Technical, Harlow, UK., 2nd edition, pp. 227, 1993

Pasquini, C., "Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications", J. Braz, Chem. Soc., Vol. 14, No.2, pp.198-219, 2003

Perten Instruments, "The Falling Number Method" http://www.perten.com/pages/ProductPage____368.aspx, 2005

Pro-lite Technology, "Spectralon Diffuse Reflectance Targets", http://www.pro-lite.uk.com/Light/Labsphere/lab_mats.html, 2005

Sapirstein, H.D. and W. Bushuk., "Quantitative Determination of Foreign Material and Vitreosity in Wheat by Digital Images", In Proceedings ICC 89 Symposium: Wheat End-use Properties, Helsinki, Finland: University of Helsinki, ed. H. Salovaara, pp. 453-474, 1989

Savitzky, A. and Golay, M.J.E., 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures", Analytical Chemistry, **36**, pp. 1627-1639, 1964

Shenk J.S., Workman J.J. and Westerhaus M.O., "Application of NIR Spectroscopy to Agricultural Products", in Handbook of Near-Infrared Spectroscopy (eds. D.A. Burns and E.W. Ciurczak), Marcel Dekker Inc, New York, 2nd ed., pp 419-474, 2001

Siesler, H. W., Ozaki, Y., Kawata, S. and Heise,H. M., "Near_Infrared Spectroscopy -- Principles, Instruments, Applications", Wiley-vchVerlag Gmbh, 2002

Sjöström, M., Wold, S. and Söderström, B., "PLS Discriminant Plots", in Pattern Recognition in Practice II, Elsevier Science Publisher B.V., Holland, pp. 461-470, 1986

Statnikov, A., Aliferis, C. F., Tsamardinos, I. Hardin, D. and Levy S., "A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis", Bioinformation Advanced Acess, Sept. 16, 2004

Symons, S. J. and Fulcher, R. G., "Determination of Wheat Kernel Morphological Variation by Digital Image Analysis: II. Variation in Cultivars of Soft White Winter Wheat", Journal of Cereal Science, **8**, pp. 219-229, 1988b

Symons, S.J. and R.G. Fulcher., "Determination of Wheat Kernel Morphological Variation by Digital Image Analysis: Variation in Eastern Canadian Milling Quality Wheats", Journal of Cereal Science, **8**, pp. 211-218, 1988

Tigabu, M, "Characterization of Forest Tree Seed Quality with Near Infrared Spectroscopy and Multivariate Analysis", Doctoral Thesis, Swedish University of Agricultural Science, Umea, 2003

Trygg, J. and Wold, S., "Orthogonal Projections to Latent Structures (O-PLS)", Journal of Chemometrics, **16**, pp. 119-128, 2002

USDA, "Canada Agricultural Situation: This Week in Canadian Agriculture", USDA Foreign Agricultural Service: Grain Report, **43**, 2004

Vapnik, V., Statistical Learning theory, Wiley-Interscience, 1998

Wang, Y., Veltkamp, D. and Kowalski, B., "Multivariate Instrument Standardization", Analytical Chemistry, **63**, pp. 2750-2756, 1991

Wang, Y. and Kowalski, B., "Improvement of Multivariate Calibration through Instrument Standardization", Analytical Chemistry, **64**, pp. 562-564, 1992

Wold, S., "Cross-validatory Estimation of the Number of Components in Factor and Principal Components Models", Technometrics, **20**, pp. 397-405, 1978

Wold, S., Antii, H., Lindgren, F. and Öhman, J., "Orthogonal Signal Correction of Near-infrared Spectra", Chemometrics and Intelligent Laboratory Systems, **44**, pp. 175-185, 1998

Wold, S., Sjöström, M. and Eriksson, L., "PLS-regression: a Basic Tool of Chemometrics", Chemometrics and Intelligent Laboratory System, **58**, pp. 109-130, 2001

Yu, H. L. and MacGregor, J. F., "Post Processing Methods (PLS-CCA): Simple Alternatives to Preprocessing Methods (OSC+PLS)", Chemometrics and Intelligent Laboratory Systems, **73**, pp. 199-205, 2004

Zayas, I., F.S. Lai and Y. Pomeranz., "Discrimination Between Wheat Classes and Varieties by Image Analysis", Cereal Chemistry, **63(1)**, pp. 52-56, 1986