NEAR INFRARED (NIR) HYPERSPECTRAL IMAGING FOR EVALUATION OF WHOLE MAIZE KERNELS: CHEMOMETRICS FOR EXPLORATION AND CLASSIFICATION

PAUL JAMES WILLIAMS

Thesis presented in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE IN FOOD SCIENCE

Department of Food Science
Faculty of AgriSciences
Stellenbosch University

Study leader: Dr Marena Manley
Co-study leaders: Prof Paul Geladi
Dr Glen Fox

March 2009

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 25th February 2009

Abstract

The use of near infrared (NIR) hyperspectral imaging and hyperspectral image analysis for distinguishing between whole maize kernels of varying degrees of hardness and fungal infected and non-infected kernels have been investigated.

Near infrared hyperspectral images of whole maize kernels of varying degrees of hardness were acquired using a Spectral Dimensions MatrixNIR camera with a spectral range of 960-1662 nm as well as a sisuChema SWIR (short wave infrared) hyperspectral pushbroom imaging system with a spectral range of 1000-2498 nm. Exploratory principal component analysis (PCA) on absorbance images was used to remove background, bad pixels and shading. On the cleaned images, PCA could be used effectively to find histological classes including glassy (hard) and floury (soft) endosperm. PCA illustrated a distinct difference between floury and glassy endosperm along principal component (PC) three. Interpreting the PC loading line plots important absorbance peaks responsible for the variation were 1215, 1395 and 1450 nm, associated with starch and moisture for both MatrixNIR images (12 and 24 kernels). The loading line plots for the sisuChema (24 kernels) illustrated peaks of importance at the aforementioned wavelengths as well as 1695, 1900 and 1940 nm, also associated with starch and moisture.

Partial least squares-discriminant analysis (PLS-DA) was applied as a means to predict whether the different endosperm types observed, were glassy or floury. For the MatrixNIR image (12 kernels), the PLS-DA model exhibited a classification rate of up to 99% for the discrimination of both floury and glassy endosperm. The PLS-DA model for the second MatrixNIR image (24 kernels) yielded a classification rate of 82% for the discrimination of glassy and 73% for floury endosperm. The sisuChema image (24 kernels) yielded a classification rate of 95% for the discrimination of floury and 92% for glassy endosperm.

The fungal infected and sound whole maize kernels were imaged using the same instruments. Background, bad pixels and shading were removed by applying PCA on absorbance images. On the cleaned images, PCA could be used effectively to find the infected regions, pedicle as well as non-infected regions. A distinct difference between infected and sound kernels was illustrated along PC1. Interpreting the PC loading line plots showed important absorbance peaks responsible for the variation and predominantly associated with starch and moisture: 1215, 1450, 1480, 1690, 1940 and 2136 nm for both MatrixNIR images (15 and 21 kernels). The MatrixNIR image (15 kernels) exhibited a PLS-DA classification rate of up to 96.1% for the discrimination of infected kernels and the sisuChema had a classification rate of 99% for the same region of interest. The

sisuChema image (21-kernels) had a classification rate for infected kernels of 97.6% without pre-processing, 97.7% with multiplicative scatter correction (MSC) and 97.4% with standard normal variate (SNV).

Near infrared hyperspectral imaging is a promising technique, capable of distinguishing between maize kernels of varying hardness and between fungal infected and sound kernels. While there are still limitations with hardware and software, these results provide the platform which would greatly assist with the determination of maize kernel hardness in breeding programmes without having to destroy the kernel. Further, NIR hyperspectral imaging could serve as an objective, rapid tool for identification of fungal infected kernels.

Uittreksel

Die gebruik van naby infrarooi (NIR) hiperspektrale beelding en hiperspektrale beeldanalise is ondersoek om onderskeid te tref tussen heel mieliepitte met wisselende hardheidsvlakke, asook tussen fungus geïnfekteerde en nie-geïnfekteerde mieliepitte.

Naby infrarooi hiperspektrale beelde van heel mieliepitte met wisselende hardheidsvlakke is verkry deur gebruik te maak van 'n *Spectral Dimensions MatrixNIR* kamera met 'n spektrale reikwydte van 960-1662 nm, sowel as 'n *sisuChema SWIR (kort golf infrarooi)* hiperspektrale beelding stelsel met 'n spektrale reikwydte van 1000-2498 nm. Absorbansie beelde is verken deur gebruik te maak van hoof komponent analise (HKA) om sodoende agtergrond, dooie piksels en skadu te verwyder uit die beeld. HKA is doeltreffend gebruik om histologiese klasse te vind naamlik glasagtige (hard) en meelagtige (sag) endosperm. HKA het 'n duidelike verskil tussen meelagtig en glasagtige endosperm aangedui op hoof komponent (HK). Interpretasie van die HK ladings stip het gelei tot die onderskeid van belangrike absorbansie pieke verantwoordelik vir hierdie variasie, naamlik 1215, 1395 en 1450 nm, geassosieer met stysel en vog, vir beide *MatrixNIR* beelde (12 en 24 pitte). Die ladings stip van die *sisuChema* (24 pitte) het pieke van belang by bo en behalwe die vorige pieke ook pieke aangedui by 1695, 1900 en 1940 nm wat ook met stysel en vog geassosieer word.

Vervolgens is parsiële kleinste kwadrate diskriminant analise (PKW-DA) uitgevoer om te voorspel of die verskillende tipes endosperm glas- of meelagtig is. Vir die *MatrixNIR* beeld (12 pitte) het die PKW-DA model 'n klassifikasie koers van 99% vir die diskriminasie van beide meelagtige en glasagtige endosperm aangedui. Die PKW-DA model vir die tweede *MatrixNIR* beeld (24 pitte) het 'n klassifikasie koers van 82% vir die diskriminasie van glasagtige en 73% vir dié van meelagtige endosperm aangedui. Die *sisuChema* beeld (24 pitte) het 'n klassifikasie waarde van 95% vir die diskriminasie van meelagtige en 92% vir glasagtige endosperm opgelewer.

Beelde van die geïnfekteerde en nie-geïnfekteerde heel mieliepitte is met behulp van dieselfde instrumente verkry. Agtergrond, dooie piksels en skadu is verwyder deur HKA toe te pas. HKA kon effektief op die skoon beelde toegepas word om geïnfekteerde areas van die pit, die pitsteel sowel as die nie-geïnfekteerde areas te onderskei. 'n Duidelike onderskeid tussen geïnfekteerde en nie-geïnfekteerde pitte kon getref word op HK2. Interpretasie van die HK ladings stip het belangrike pieke, hoofsaaklik geassosieer met stysel and vog, verantwoordelik vir variasie aangedui as 1215, 1450, 1480, 1690, 1940 en 2136 nm vir beide MatrixNIR beelde (15 en 21 pitte). Die *MatrixNIR* beeld (15 pitte) het 'n

PKW-DA klassifikasie koers van 96.1% aangedui vir geïnfekteerde pitte, terwyl die *sisuChema* 'n klassifikasie koers van 99% vir dieselfde area van belang aangedui het. Die *sisuChema* beeld (21 pit) het klassifikasie koerse van 97.6% vir geïnfekteerde pitte aangedui sonder vooraf korreksie; 97.7% met MSC (*multiplicative scatter correction*) en 97.4% met SNV (*standard normal variate*) aangedui.

Naby infrarooi hiperspektrale beelding is 'n belowende tegniek, wat geskik is om onderskeid te tref tussen mieliepitte van verskillende hardheidsvlakke en tussen fungus geïnfekteerde en nie-geïnfekteerde pitte. Hoewel daar steeds beperking is sover dit die harde- en sagteware aangaan, het hierdie resultate 'n platform daargestel wat 'n groot bydrae kan lewer tot die bepaling van mieliehardheid in teelprogramme sonder vernietiging van die pit. Verder kan NIR hiperspektrale beelding ook as vinnige, objektiewe metode van identifikasie van geïnfekteerde en nie-geïnfekteerde pitte.

Acknowledgements

Firstly I would like to thank our heavenly Father for allowing me to attain this for without Him it would have not been possible.

Additionally I would like to express my sincere gratitude to the following people:

Dr Marena Manley my study leader for all her help, patience, advice, support and forcing this thesis out of me, especially during the last two months;

Professor Paul Geladi (Swedish University of Agricultural Sciences, Umeå, Sweden) my co-study leader for help and advice especially with the chemometrics;

Dr Glen Fox my co-study leader from down-under, for his help, support, advice and Brazenhead moments;

Julian White (Specim, Spectral Imaging Ltd, Oulu, Finland) for his help and assistance with instrumentation and rescuing me from the ice cold winter in Haparanda (!);

Dr David Nilson and Oskar Johannson (Umbio AB, Umeå, Sweden) for the use of and constant support with the Evince software;

Dr James Burger (BurgerMetrics SIA, Jelvaga, Latvia) for showing much interest in the project and assistance with the technical advice and always being only an email away;

Prof Altus Viljoen and Ian Small (Department of Plant Pathology, Stellenbosch University) for support and preparation of maize kernels for fungal studies;

Professor Trevor Britz for help and assistance with the microbiology and use of equipment;

The Maize Trust for a bursary and funding of the project;

The National Research Foundation (NRF) for a bursary and for funding of this project (FA2006032900007);

Contents

Language and style used in this thesis are in accordance with the requirements of the *International Journal of Food Science and Technology.* This thesis represents a compilation of manuscripts where each chapter is an individual entity and some repetition between chapters has, therefore, been unavoidable.

Chapter 1

Introduction

Chapter 1

Introduction

Near infrared (NIR) spectroscopy is a spectroscopic technique utilising the NIR region of the electromagnetic spectrum from 700–2500 nm (Bokobza, 1998). A series of NIR spectra of biological samples can be used for calculating concentrations of, *e.g.* fat in pork and beef (Prevolnik *et al.*, 2005); protein in wheat (Delwiche, 1998); moisture in sheep meat (McGlone *et al.*, 2005); inorganic phosphorous in soy beans (Delwiche *et al.*, 2006); anthocyanins and total soluble solids in red grapes (Cozzolino *et al.*, 2005); and many more constituents in a variety of commodities (Osborne, 2000)*.* Typical applications include identification and quantification of pharmaceutical raw materials (Roggo *et al.*, 2007); medical diagnostics for viral infections (Sakudo *et al.*, 2006); and food and agrochemical quality control (Osborne, 2000; Nicolaï *et al.*, 2007). Ideally the whole surface of products has to be inspected; however, NIR spectroscopy relies on averaged point measurements of the sample for quantification and qualitative purposes. Recent advances in this technology enable the acquisition of NIR hyperspectral images (Burger & Geladi, 2005; Burger & Geladi, 2006; Burger & Geladi, 2007) which consist of an NIR spectrum for each pixel in the image. This enables the prediction of analyte presence and concentration at each pixel, leading to the creation of concentration images or maps which allows the entire sample to be measured and not just parts thereof.

NIR hyperspectral imaging is a form of NIR spectroscopy imaging that captures images at many wavelength bands in the NIR region (Koehler *et al.*, 2002; Cogdill *et al.*, 2004; Geladi *et al.*, 2004; Tatzer *et al.*, 2005). It is an expansion of multispectral NIR imaging, where images are captured at a much smaller number of wavelength bands, usually two or three (Koehler *et al.*, 2002; Cogdill *et al.*, 2004; Geladi *et al.*, 2004). NIR hyperspectral imaging is an exciting, new analytical advance that answers frequently posed questions such as: what chemical species are present and most importantly, where are they located? These questions can be answered simultaneously, in a single rapid measurement, through the unification of traditional NIR spectroscopy with powerful microscopic and macroscopic imaging capabilities of a hyperspectral imaging system. NIR hyperspectral imaging empowers researchers with spatial and spectral information that permits characterisation of samples with exceptional simplicity, speed and improved spatial and spectral resolution (Koehler *et al.*, 2002). This methodology is aimed at providing a comprehensive analysis of complex heterogeneous samples.

Image data collected by a NIR hyperspectral imaging system are arranged into a three-way data matrix, known as a hypercube (Cogdill *et al.*, 2004; Geladi *et al.*, 2004). The first two axes (x and y) of the matrix are the vertical and horizontal pixel coordinates while the third (z) axis is the spectral dimension. By overlaying hundreds of single channel black and white (grayscale) images, hyperspectral images (or hypercubes) are produced (Burger & Geladi, 2006). Each of these grayscale images represents a single band of spectral wavelengths. A typical commercial instrument produces hypercubes with dimensions 256×320×118. This can be construed as 118 single channel images each with 256×320 pixels. Alternatively, this hypercube may be considered as 81 920 spectra, each with 118 wavelength channels. This colossal accumulation of data poses data extraction challenges; however, in addition it creates novel possibilities.

Chemometrics is essential for the extraction of relevant information relating to the spectral content, as with conventional NIR spectroscopy, permitting sample categorisation or quantitative determinations (Burger & Geladi, 2006). Multivariate image analysis (MIA) is a methodology for analysing multivariate images, where the image coordinates are location (two- or three dimensional, usually pixel number) and variable number (Geladi & Grahn, 1996; Geladi & Grahn, 2000), *i.e.* wavelength in this instance. Multivariate images could have distinctive sizes, e.g. 1024×1024, 512×512, 256×256 or many other combinations, and have between two and many hundreds of variables. The variables can be wavelength, electron energy or particle mass. Image analysis (regarding digital images) focuses primarily on spatial relationships between pixels in a grey level image while MIA (regarding hyperspectral images) concentrates on the correlation of structure between the variables to provide additional information valuable for exploring images and classifying regions in them.

NIR hyperspectral imaging has been evaluated for quality and safety inspection of agricultural products such as poultry carcasses (Chao *et al.*, 2001; Chao *et al.*, 2002; Park *et al.*, 2002; Lawrence *et al.*, 2003; Park *et al.*, 2004; Lawrence *et al.*, 2006) as well as deficiency detection or quality determination on apples (Kim *et al.*, 2002; Mehl *et al.*, 2002; Lu, 2003; Mehl *et al.*, 2004; Liu *et al.*, 2007); cucumbers (Cheng *et al.*, 2004; Liu *et al.*, 2005); and tomatoes (Polder *et al.*, 2002). Additionally, classification of sound and stained wheat grains (Berman *et al.*, 2007) in addition to the determination of the vitreousness of durum wheat kernels (Gorretta *et al.*, 2006) have also been assessed. NIR hyperspectral imaging has also been evaluated for the detection of meat and bone meal in compound feeds (Fernández Pierna *et al.*, 2004) as well as for the screening of compound feeds (Fernández Pierna *et al.*, 2006).

Maize (*Zea mays*) is one of the three major grain crops in terms of world production (Pereira *et al.*, 2007). Additionally, it is a significant source of energy and protein in the human diet throughout the world (Rehman, 2006). Maize hardness is a quality characteristic that is important to many players within the maize industry and is a trait that has been well studied and documented (Wolf *et al.*, 1952; Wolf *et al.*, 1969; Pomeranz *et al.*, 1984; Pomeranz *et al.*, 1985; Pomeranz, 1986; Pomeranz *et al.*, 1986a; Pomeranz *et al.*, 1986b; Watson, 1987a; Watson, 1987b; Dombrink-Kurtzman & Bietz, 1993; Wehling *et al.*, 1993; Dombrink-Kurtzman, 1994; Eyherabide *et al.*, 1996; Wehling *et al.*, 1996; Dombrink-Kurtzman & Knutson, 1997; Eyherabide *et al.*, 2004; Lee *et al.*, 2005a; Lee *et al.*, 2005b; Lee *et al.*, 2006). This trait is determined using various predominantly destructive techniques which can be inefficient and unnecessary in breeding programmes. Industry seeks kernels of a particular hardness, thus responsibility lies with breeders to develop hybrids that comply. To ensure that this process is made easier, rapid, non-destructive techniques have to be employed as tools in breeding programmes. NIR spectroscopy is suited since kernels are scanned, various traits predicted using appropriate NIR calibration models, then re-installed in breeding trials. Over the years this technology has developed considerably to a stage that it now offers the ability to acquire a vast amount of data regarding a sample in one rapid, non-destructive measurement. With single-point NIR spectroscopy information regarding a sample is obtained in a series of averaged measurements at various locations on a sample. NIR hyperspectral imaging on the other hand is capable of capturing an image of the entire sample, thus allowing for measurement of the whole sample and not just part of it. Determination of single-kernel hardness could assist in separating individual kernels to ensure that kernels with appropriate hardness (and other quality traits) would be used and kernels with undesirable hardness would be removed from processing.

Another aspect of concern for the maize industry is fungal growth and mycotoxin contamination that can occur during crop growth, harvesting or storage of maize (Turner *et al.*, 1999). Temperature and moisture are critical components affecting fungal growth and mycotoxin production; however, the associated health risks are also related to the regulatory infrastructure and financial support to enforce suitable monitoring and regulation. When cereal grains and animal foods are colonised by fungi there is a significant risk of contamination with secondary metabolites or mycotoxins of these fungi (Placinta *et al.*, 1999). Fumonisins are mycotoxins produced by a variety of fungi of the *Fusarium* genus (Jackson & Jablonski, 2004). These toxins are natural contaminants of cereal grains worldwide and are mostly found in maize and products derived from maize

and were first isolated in 1988 (Gelderblom *et al.*, 1988) at the Programme on Mycotoxins and Experimental Carcinogenesis (PROMEC) of the Medical Research Council (MRC) in South Africa (Marasas, 2001). Thus, objective detection of fungal infected kernels could be of much benefit at maize silos where infected maize kernels would be removed from the storage and subsequent human and animal food chain, in so doing limiting the chance of ingestion of fumonisins.

The aim of this study was to demonstrate the application of NIR hyperspectral imaging to evaluate whole maize kernels in terms of quality and food safety purposes. Specific objectives of this study were therefore to evaluate NIR hyperspectral imaging to distinguish between:

- whole maize kernels of varying hardness using principal component analysis as well as partial least squares-discriminant analysis; and
- fungal infected non-infected whole maize kernels by means of unsupervised, supervised and pre-processing chemometrics techniques.

References

Berman, M., Connor, P.M., Whitbourn, L.B., Coward, D.A., Osborne, B.G. & Southan, M.D. (2007). Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis. *Journal of Near Infrared Spectroscopy*, **15**, 351-358.

Bokobza, L. (1998). Near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **6**, 3-17.

Burger, J. & Geladi, P. (2005). Hyperspectral NIR image regression part I: calibration and correction. *Journal of Chemometrics*, **19**, 355-363.

Burger, J. & Geladi, P. (2006). Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst*, **131**, 1152-1160.

Burger, J. & Geladi, P. (2007). Spectral pre-treatments of hyperspectral near infrared images: analysis of diffuse reflectance scattering. *Journal of Near Infrared Spectroscopy*, **15**, 29-37.

Chao, K., Chen, Y.R., Hruschka, W.R. & Park, B. (2001). Chicken heart disease characterization by multi-spectral imaging. *Applied Engineering in Agriculture*, **17**, 99-106.

Chao, K., Mehl, P.M. & Chen, Y.R. (2002). Use of hyper– and multi–spectral imaging for detection of chicken skin tumors. *Applied Engineering in Agriculture*, **18**, 113-119.

Cheng, X., Chen, Y.R., Tao, Y., Wang, C.Y., Kim, M.S. & Lefcourt, A.M. (2004). A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. *Transactions of the ASAE,* **47**, 1313-1320.

Cogdill, R.P., Hurburgh, C.R. & Rippke, G.R. (2004). Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE,* **47**, 311-320.

Cozzolino, D., Cynkar, W.U., Dambergs, R.G., Janik, L. & Gishen, M. (2005). Effect of both homogenisation and storage on the spectra of red grapes and on the measurement of total anthocyanins, total soluble solids and pH by visual near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **13**, 213-223.

Delwiche, S.R. (1998). Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. *Journal of Cereal Science*, **27**, 241-254.

Delwiche, S.R., Pordesimo, L.O., Scaboo, A.M. & Pantalone, V.R. (2006). Measurement of inorganice phosphorus in soybeans with near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, **54**, 6951-6956.

Dombrink-Kurtzman, M.A. (1994). Examination of *opaque* mutants of maize by reversed-phase high-performance liquid chromatography and scanning electron microscopy. *Journal of Cereal Science*, **19**, 57 - 64.

Dombrink-Kurtzman, M.A. & Bietz, J.A. (1993). Zein composition in hard and soft endosperm of maize. *Cereal Chemistry*, **70**, 105-108.

Dombrink-Kurtzman, M.A. & Knutson, C.A. (1997). A study of maize endosperm hardness in relation to amylose content and susceptibility to damage. *Cereal Chemistry*, **74**, 776-780.

Eyherabide, G., Robutti, J., Percibaldi, N., Presello, D. & Alvarez, M. (2004). Association between grain yield and endosperm hardness in maize cultivars. *Maydica*, **49**, 319-326.

Eyherabide, G.H., Robutti, J.L. & Borras, F.S. (1996). Effect of near-infrared transmission-based selection on maize hardness and the composition of zeins. *Cereal Chemistry*, **73**, 775-778.

Fernández Pierna, J.A., Baeten, V. & Dardenne, P. (2006). Screening of compound feeds using NIR hyperspectral data. *Chemometrics and Intelligent Laboratory Systems*, **84**, 114-118.

Fernández Pierna, J.A., Baeten, V., Michotte Renier, A., Cogdill, R.P. & Dardenne, P. (2004). Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Journal of Chemometrics*, **18**, 341-349.

Geladi, P., Burger, J. & Lestander, T. (2004). Hyperspectral imaging: calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems*, **72**, 209-217.

Geladi, P. & Grahn, H.F. (1996). *Multivariate Image Analysis.* Pp. 316. Chichester, West Sussex: John Wiley & Sons Ltd.

Geladi, P. & Grahn, H.F. (2000). Multivariate image analysis. In: *Encyclopedia of Analytical Chemistry* (edited by R.A. Meyers). **Vol. 15**. Pp. 13540-13562. Chichester: John Wiley & Sons.

Gelderblom, W.C.A., Jaskiewicz, K., Marasas, W.F.O., Thiel, P.G., Horak, R.M., Vleggaar, R. & Kriek, N.P.J. (1988). Fumonisins- novel mycotoxins with cancer-promoting activity produced by *Fusarium moniliforme*. *Applied and Environmental Microbiology*, **54**, 1806-1811.

Gorretta, N., Roger, J.M., Aubert, M., Bellon-Maurel, V., Campan, F. & Roumet, P. (2006). Determining vitreousness of durum wheat kernels using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **14**, 231-239.

Jackson, L. & Jablonski, J. (2004). Fumonisins. In: *Mycotoxins in food: detection and control* (edited by N. Magan & M. Olsen). Pp. 367-405. Cambridge, England: Woodhead Publishing, Ltd.

Kim, M.S., Lefcourt, A.M., Chao, K., Chen, Y.R., Kim, I. & Chan, D.E. (2002). Multispectral detection of fecal contamination on apples based on hyperspectral imagery. Part I. Application of visible and near-infrared reflectance imaging. *Transactions of the ASAE*, **45**, 2027-2037.

Koehler, F.W., Lee, E., Kidder, L.H. & Lewis, E.N. (2002). Near infrared spectrsocopy: the practical chemical imaging solution. *Spectoscopy Europe*, **14**, 12-19.

Lawrence, K.C., Windham, W.R., Park, B. & Buhr, R.J. (2003). A hyperspectral imaging system for identification of faecal and ingesta contamination on poultry carcasses. *Journal of Near Infrared Spectroscopy*, **11**, 269-281.

Lawrence, K.C., Windham, W.R., Park, B., Heitschmidt, G.W., Smith, D.P. & Feldner, P. (2006). Partial least squares regression of hyperspectral images for contaminant detection on poultry carcasses. *Journal of Near Infrared Spectroscopy*, **14**, 223-230.

Lee, K.M., Herrman, T.J., Lingenfelser, J. & Jackson, D.S. (2005). Classification and prediction of maize hardness-associated properties using multivariate statistical analyses. *Journal of Cereal Science*, **41**, 85-93.

Lee, K.M., Bean, S.R., Alavi, S., Herrman, T.J. & Waniska, R.D. (2006). Physical and biochemical properties of maize hardness and extrudates of selected hybrids. *Journal of Agricultural and Food Chemistry*, **54**, 4260.

Liu, Y., Chen, Y.-R., Kim, M.S., Chan, D.E. & Lefcourt, A.M. (2007). Development of simple algorithms for the detection of fecal contaminants on apples from visible/near infrared hyperspectral reflectance imaging. *Journal of Food Engineering*, **81**, 412-418.

Liu, Y.L., Chen, Y.R., Wang, C.Y., Chan, D.E. & Kim, M.S. (2005). Development of a simple algorithm for the detection of chilling injury in cucumbers from visible/near-infrared hyperspectral imaging. *Applied Spectroscopy*, **59**, 78-85.

Lu, R. (2003). Detection of bruises on apples using near-infrared hyperspectral imaging. *Transactions of the ASAE*, **46**, 523-530.

Marasas, W.F.O. (2001). Discovery and occurence of the Fumonisins: a historical perspective. *Environmental Health Perspectives*, **109**, 239-243.

Mcglone, V.A., Devine, C.E. & Wells, R.W. (2005). Detection of tenderness, post-rigor age and water status changes in sheep meat using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **13**, 277-285.

Mehl, P.M., Chao, K., Kim, M. & R., C.Y. (2002). Detection of defects on selected apple cultivars using hyperspectral and multispecteral image analysis. *Applied Engineering in Agriculture*, **18**, 219-226.

Mehl, P.M., Chen, Y.R., Kim, M.S. & Chan, D.E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering*, **61**, 67-81.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I. & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, **46**, 99-118.

Osborne, B.G. (2000). Near-infrared spectroscopy in food analysis. In: *Encyclopedia of Analytical Chemistry* (edited by R.A. Meyers). **Vol. 5**. Pp. 4069-4081. Chichester: John Wiley & Sons.

Park, B., Lawrence, K.C., Windham, W.R. & Buhr, R.J. (2002). Hyperspectral imaging for detecting fecal and ingesta contaminants on poultry carcasses. *Transactions of the ASAE*, **45**, 2017-2026.

Park, B., Lawrence, K.C., Windham, W.R. & Smith, D.P. (2004). Multispectral Imaging system for fecal and ingesta detection on poultry carcasses. *Journal of Food Process Engineering*, **27**, 311-327.

Pereira, P., Nesci, A. & Etcheverry, M. (2007). Effects of biocontrol agents on *Fusarium verticillioides* count and fumonisin content in the maize agroecosystem: Impact on rhizospheric bacterial and fungal groups. *Biological Control*, **42**, 281-287.

Placinta, C.M., D'mello, J.P.F. & Macdonald, A.M.C. (1999). A review of worldwide contamination of cereal grains and animal feed with Fusarium mycotoxins. *Animal Feed Science and Technology*, **78**, 21-37.

Polder, G., Van Der Heijden, G.W.A.M. & Young, I.T. (2002). Spectral image analysis for measurig ripeness of tomatoes. *Transactions of the ASAE*, **45**, 1155-1161.

Pomeranz, Y. (1986). Comparison of screening methods for indirect determination of sorghum hardness. *Cereal Chemistry*, **63**, 36-38.

Pomeranz, Y., Czuchajowska, Z. & Lai, F.S. (1986a). Comparison of methods for determination of hardness and breakage susceptibility of commercially dried corn. *Cereal Chemistry*, **63**, 39-43.

Pomeranz, Y., Czuchajowska, Z., Martin, C. & Lai, F. (1985). Determination of maize hardness by Stenvert hardness tester. *Cereal Chemistry*, **62**, 108-112.

Pomeranz, Y., Hall, G.E., Czuchajowska, Z. & Lai, F.S. (1986b). Test weight, hardness, and breakage susceptibility of yellow dent corn hybrids. *Cereal Chemistry*, **63**, 349-351.

Pomeranz, Y., Martin, C.R., Traylor, D.D. & Lai, F.S. (1984). Corn hardness determination. *Cereal Chemistry*, **61**, 147-150.

Prevolnik, M., Čandek-Potokar, M., Škorjanc, D., Velikonja-Bolta, Š., Škrlep, M., Žnidaršič, T. & Babnik, D. (2005). Predicting intramuscluar fat content in pork and beef by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **13**, 77-85.

Rehman, Z.U. (2006). Storage effects on nutritional quality of commonly consumed cereals. *Food Chemistry*, **95**, 53-57.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A. & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, **44**, 683-700.

Sakudo, A., Suganuma, Y., Kobayashi, T., Onodera, T. & Ikuta, K. (2006). Near-infrared spectroscopy: Promising diagnostic tool for viral infections. *Biochemical and Biophysical Research Communications*, **341**, 279-284.

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

Turner, P.C., Nikiema, P. & Wild, C.P. (1999). Fumonisin contamination of food: progress in development of biomarkers to better assess human health risks. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, **443**, 81-93.

Watson, S.A. (1987a). Measurement and maintenance of quality. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 125-183. St Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Watson, S.A. (1987b). Structure and composition. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 53-82. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Wehling, R.L., Jackson, D.S. & Hamaker, B.R. (1996). Prediction of corn dry-milling quality by near-infrared spectroscopy. *Cereal Chemistry*, **73**, 543-546.

Wehling, R.L., Jackson, D.S., Hooper, D.G. & Ghaedian, A.R. (1993). Prediction of wet-milling starch yield from corn by near-infrared spectroscopy. *Cereal Chemistry*, **70**, 720-723.

Wolf, M.J., Buzan, C.L., Macmasters, M.M. & Rist, C.E. (1952). Structure of the mature corn kernel. I. Gross anatomy and structural relationships. *Cereal Chemistry*, **29**, 321-333.

Wolf, M.J., Khoo, U. & Seckinger, H.L. (1969). Distribution and subcellular structure of endosperm protein in varieties of ordinary and high-lysine maize. *Cereal Chemistry*, **46**, 253-263.

Chapter 2

Literature Review

Chapter 2
Literature Review

Introduction

Maize is a vital source of energy and is fast becoming one of the most important dietary staple foods in the world (Schierbaum, 2007). It is used for human consumption in diverse forms, from specialised foods in developed countries, to staple food in undeveloped countries (Malvar *et al.*, 2008). In addition to being used for human consumption, maize is also used for animal feed. Thus, the quality and safety of this commodity is of growing concern to all involved in the industry from breeding programmes, processors and food manufacturers.

Maize quality for milling is based on the hardness of the kernel (Watson, 1987a; Watson, 1987b). This characteristic is measured using various predominantly destructive techniques which can be an issue for quality testing in breeding programmes. Because industry seeks kernels of a particular hardness, size and colour, responsibility lies with breeders to develop hybrids that comply with these specifications when grown across a large number of environments. From an industry perspective, maize kernels of uniform quality (hardness) allows for efficient and consistent processing, thereby minimising costs and providing quality products.

To ensure selection process is made easier, rapid, non-destructive techniques have to be employed as tools. Near infrared (NIR) spectroscopy, in particular, is suitable as kernels are scanned, measured for various attributes then re-used in breeding trials. This technology has, over the years, developed substantially to a level that it is now possible to acquire a vast multitude of data concerning a sample in one rapid, non-destructive measurement.

Fungal activity on maize also affects the quality of kernels, particularly the shape and size. More importantly, if undetected, infected kernels could enter the food chain. Maize kernels infected with *Fusarium verticillioides*, can contain a toxin that is poisonous to animals (Gelderblom *et al.*, 1996; Gelderblom *et al.*, 2001a; Gelderblom *et al.*, 2001b; Gelderblom *et al.*, 2004) as well as humans (Isaacson, 2005). Fungal infected maize cobs are identified visually; however not all infected kernels present visual symptoms. An objective method is thus required to rapidly identify infected kernels that will ensure effective sorting.

Maize hardness

Hardness is an important intrinsic property of maize kernels which affects grinding power requirements, nutritive properties, dust formation, kernel and bulk density, production of special foods, and yields of dry-milled and wet-milled products (Watson, 1987a). Maize kernel hardness is principally a genetic expression, but environment and postharvest handling (e.g. transportation, drying and storage) also have an influence on hardness properties (Pomeranz *et al.*, 1984; Pomeranz *et al.*, 1985a; Watson, 1987b)

To understand maize hardness it is necessary to be acquainted with the composition of the maize kernel (**Fig. 2.1**). The maize kernel is composed of the germ, endosperm and pericarp (Watson, 1987b; Hoseney, 1994). However, the most important component concerning hardness is the endosperm.

The endosperm constitutes 82-84% of the kernel dry weight and is 86-89% starch by weight (Wolf *et al.*, 1952; Watson, 1987b). It consists of a thin outer layer of aleurone cells, containing oil and protein, and a large inner portion of storage tissue which contains starch and protein (Wolf *et al.*, 1952). The endosperm envelops the germ, but only the aleurone layer is present over the face of the germ; the major portion of the endosperm lies to the back, sides and above the germ.

The endosperm is filled with elongated cells packed with starch granules embedded in a continuous protein matrix (Wolf *et al.*, 1952; Watson, 1987b). There are two types of starchy endosperm, floury (soft) and glassy (hard) (Bennet, 1950; Wolf *et al.*, 1952; Watson, 1987b). The floury endosperm surrounds the central fissure and is opaque to transmitted light. This is due to light refracting from minute air pockets around the starch granules, which result from tearing of the thin protein matrix as it shrinks during drying. Consequently, the protein matrix no longer completely surrounds the starch granules, which assume a round shape. In the glassy endosperm, the protein matrix is thicker and remains intact on drying. During drying, the plastic starch granules in the glassy endosperm are compressed into polyhedral shapes. Endosperm cells become progressively smaller form the central fissure to the outer endosperm.

The proportion of glassy to floury endosperm depends on the type and variety of the maize (Bennet, 1950; Wolf *et al.*, 1952; Watson, 1987b). In general, flour maize has little to no glassy endosperm; popcorn maize on the other hand, has a large proportion of glassy endosperm which forms a thin shell around a small central core of floury endosperm. Flint and dent maize kernels are intermediate between the flour and popcorn varieties. The ratio of glassy to floury endosperm averages about 2:1 in dent maize types
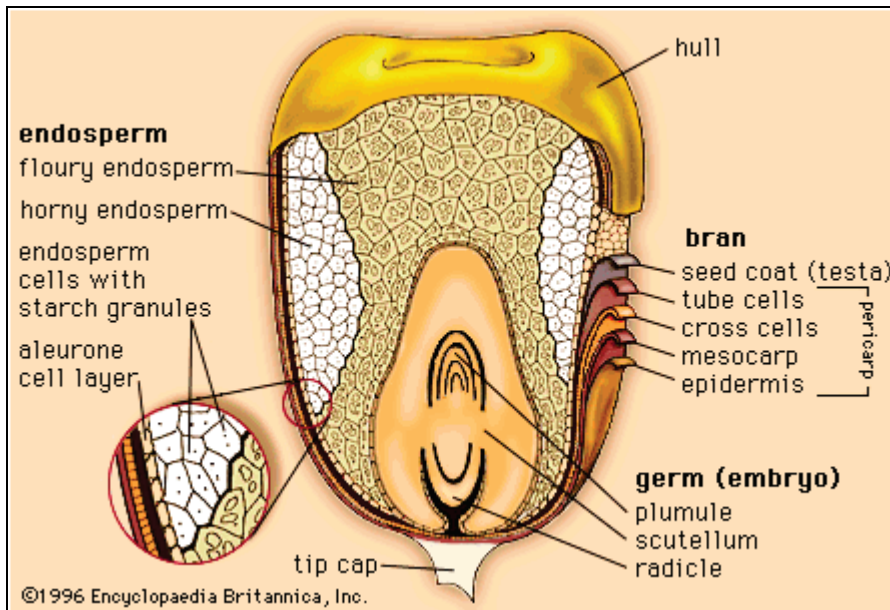
Figure 2.1 Structure of maize kernel showing location of basic constituents (Anonymous, 1996).

(Wolf *et al.*, 1952), where the glassy endosperm lies predominantly at the sides and back of the kernel. The absence of a thick glassy endosperm layer in the crown results in structural weakness in that area, thus as the kernel dries and shrinks at maturity the dent is formed by the collapse of the loose floury endosperm tissues. Neither the flint nor popcorn varieties are dented since the crown region in both is supported by a relatively thick layer of glassy endosperm.

The protein matrix plays an important role in the hardness of maize kernels and may directly influence endosperm texture and physical properties (Dombrink-Kurtzman & Bietz, 1993). Protein content has been correlated to hardness and the variation in storage protein (zein) classes have been linked to differences in hardness (Ortega & Bates, 1983; Larkins *et al.*, 1984; Dombrink-Kurtzman & Bietz, 1993; Dombrink-Kurtzman, 1994; Pratt *et al.*, 1995; Eyherabide *et al.*, 1996; Mestres & Matencio, 1996; Robutti *et al.*, 1997; Chandrashekar & Mazhar, 1999; Robutti *et al.*, 2000; Gupta *et al.*, 2003; Zhang *et al.*, 2003; Holding & Larkins, 2006; Lee *et al.*, 2006; Ioerger *et al.*, 2007; Sanchez *et al.*, 2007). While the protein content is a very low portion of the total kernel composition it would appear that it can play a strong role in influencing hardness.

A range of tests have been used and evaluated to determine maize hardness. These include resistance to grinding, grinding / sieving tests, measuring abrasion, yield of grits and starch gelatinisation properties. In addition, near infrared (NIR) spectroscopy in both reflectance and transmission modes, has also been evaluated.

Particle size index (PSI), one of the most common methods used to determine maize hardness, is based on milling the sample and then fractionating the ground material

through sieves (Abdelrahman & Hoseney, 1984; Pomeranz, 1984; Pomeranz, 1986; Wu, 1992; Haddad *et al.*, 1998). This would give rise to various layers of different degrees of hardness giving information regarding the variation of hardness within a set of kernels. Another method, the Stenvert hardness tester, timed the period for a 17 ml tube, attached to a grinder, to fill with ground material. This related hardness to resistance to grinding. A shorter period to fill the tube would consider the kernels softer (Pomeranz *et al.*, 1985b; Pomeranz *et al.*, 1986; Cavanaugh *et al.*, 1995; Armstrong *et al.*, 2007; Lee *et al.*, 2007b). This technique would give rise to the total hardness of a sample and would not, like the PSI, give an indication of the variation within the sample measured.

A different, non-destructive conceptualisation was the measurement of the density of the maize kernels (Cavanaugh *et al.*, 1995; Pratt *et al.*, 1995; Siska & Hurburgh, 1995; Wehling *et al.*, 1996; Almeida-Dominguez *et al.*, 1997; Jennings *et al.*, 2002; Kim *et al.*, 2002; Duarte *et al.*, 2005; Yang *et al.*, 2005; Cheetham *et al.*, 2006; Gadag *et al.*, 2006; Thompson & Goodman, 2006; Lee *et al.*, 2007a; Lee *et al.*, 2007b; Milasinovic *et al.*, 2007). This measurement was carried out using two different techniques. The one technique involved counting the number of maize kernels that floated in a particular solution. Solutions used included sodium nitrate, sucrose, kerosene-carbon tetrachloride and ethanol. A higher number of floating kernel indicated that the samples had a greater number of soft kernels. The other technique required the measurement of density based on the mass of kernels in a specified volume. This method used a pycnometer and kernels with higher densities were considered hard.

The tangential abrasion dehulling device (TADD) is a process where kernels are abraded for a 10 minute period (Wehling *et al.*, 1996; Almeida-Dominguez *et al.*, 1997; Lee *et al.*, 2007b). The amount of material removed from the kernels is calculated, with higher values of material removed indicating soft kernels. The Rapid Visco Analyser (RVA) is a technique that uses ground material, mixed with water that is constantly stirred while the sample is heated to $100^{o}C$. The technique provides information on starch properties, including paste viscosity, gelatinisation temperature and time. The data has been related to maize hardness (Almeida-Dominguez *et al.*, 1997; Lee *et al.*, 2005; Narvaez-Gonzalez *et al.*, 2006; Narvaez-Gonzalez *et al.*, 2007) as the RVA measures variation in amylose content and starch granule characteristics which have been shown to contribute to maize hardness (Dombrink-Kurtzman & Knutson, 1997). Similar to the PSI method and Stenvert hardness tester, the RVA requires a grinding step.

NIR technology has been applied to substantiate efficiencies in breeding selection or to determine the hardness of samples prior to milling (Robutti, 1995; Siska & Hurburgh, 1995;

Eyherabide *et al.*, 1996; Wehling *et al.*, 1996; Campbell *et al.*, 1997; Campbell *et al.*, 1999; Campbell *et al.*, 2000; Fox *et al.*, 2002; Lee *et al.*, 2005; Armstrong *et al.*, 2007; Cen & He, 2007; Fox *et al.*, 2007). Calibrations developed using NIR are extremely sensitive to the accuracy and precision of the reference technique (Williams, 2007) and are thus dependant on the accurate execution of these techniques. Calibrations for maize kernel hardness have been developed using the TADD (Wehling *et al.*, 1996; Lee *et al.*, 2005), Stenvert mill (Lee *et al.*, 2005; Armstrong *et al.*, 2007), density method (Eyherabide *et al.*, 1996), kernel density (Siska & Hurburgh, 1995) and coarse-fine ratio (Robutti, 1995). A number of these calibrations have used samples corrected to a fixed moisture content (Campbell *et al.*, 2000; Lee *et al.*, 2005; Armstrong *et al.*, 2007) while the rest used maize samples on an as is basis (Robutti, 1995; Siska & Hurburgh, 1995; Eyherabide *et al.*, 1996; Wehling *et al.*, 1996). The use of a single wavelength 860 or 1680 nm (Robutti, 1995; Lee *et al.*, 2005) as well as the maximum absorbance/reflectance between 620 and 680 nm have also been related to hardness (Robutti, 1995).

Fusarium verticillioides

Fusarium verticillioides (synonym, *Fusarium moniliforme* Sheldon; teleomorph, *Gibberella moniliformis* [synonym, *Gibberella fujikuroi* mating population A]) is the most commonly reported fungal species infecting maize worldwide (Gelderblom *et al.*, 2001b; Schulthess *et al.*, 2002; Oren *et al.*, 2003; Yates *et al.*, 2003; Bush *et al.*, 2004; Yates & Sparks, 2008). They are ubiquitous in soils and are associated with infection at all stages of plant development and infect all plant parts (Bacon *et al.*, 1992; Munkvold & Desjardins, 1997; Munkvold *et al.*, 1997; Bacon *et al.*, 2001; Pascale *et al.*, 2002; Gelderblom *et al.*, 2004; Desjardins, 2006).

On potato dextrose agar (PDA) medium, *F. verticillioides* produces white mycelium (**Fig. 2.2a**) that may become violet as cultures age (Desjardins, 2006). The colour of the colony under-surface may vary from white to greyish orange (**Fig. 2.2b**), dark violet or almost black in some strains. Macroconidia have three to five septa and are slender and straight while chlamydospores are absent (Samson, 2000; Desjardins, 2006). *F. verticillioides* produces club-shaped microconidia in false heads and in chains on monophialedes.

Fusarium verticillioides is an endophyte since its hyphae occur systemically in the leaves, stems, roots and cobs of the maize plant (Bacon *et al.*, 1992; Bacon & Hinton, 1996; Munkvold & Desjardins, 1997; Schulthess *et al.*, 2002; Fandohan *et al.*, 2003). The word endophyte means "in plant" and mycologists employ the term "fungal endophyte" for

fungi that inhabit plants without visible disease symptoms (Schulz & Boyle, 2005). *F. verticillioides*, however, is capable of causing asymptomatic infection as well as symptomatic infections (Bacon *et al.*, 1992; Bacon & Hinton, 1996; Munkvold & Desjardins, 1997; Fandohan *et al.*, 2003).

Symptomatic infections include severe damage to kernels, seed quality loss and deformation of kernel shape and size and is most likely a result of a complex interaction of virulence of an isolate, maize variety and environment (Headrick *et al.*, 1990). In addition, the symptomatic infections manifest themselves as ear and kernel rot that are characterised by the presence of white, pink or salmon-coloured mould and brown, tan or white-streaked kernels.

Asymptomatic infection, on the other hand, can exist throughout the plant in leaves, stems, roots, grains, and the presence of the fungus is in many cases ignored because it does not cause visible damage to the plant (Munkvold & Desjardins, 1997; Fandohan *et al.*, 2003; Jackson & Jablonski, 2004). The endophytic habit suggests that the fungus arrives at its internal location in kernels from growth through the stalk, into the cob, rachilla and pedicle (Foley, 1962). Furthermore using scanning electron microscopy (SEM), it was found that in asymptomatic kernels, the fungus was always found in the pedicle or tip cap end of the kernel (Bacon *et al.*, 1992). Thus, the fungus enters the kernel at the pedicle and invades the embryo and enters the endosperm, sporulating internally under high moisture conditions, resulting in the production of mycotoxins.
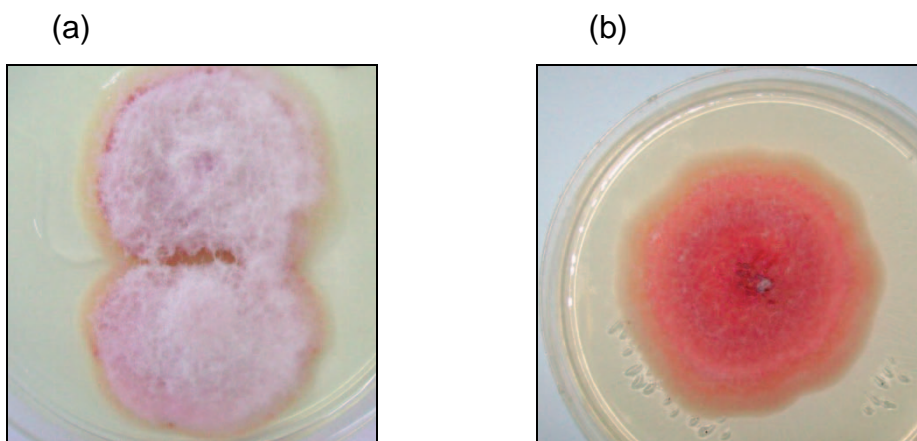
(a)   (b)

Figure 2.2 (a) White mycelium of *Fusarium verticillioides* on potato dextrose agar (PDA) medium; and (b) under-surface showing greyish-orange colour.

Fusarium verticillioides produces a group of mycotoxins known as fumonisins (Gelderblom *et al.*, 1984; Bezuidenhout *et al.*, 1988; Gelderblom *et al.*, 1988; Marasas, 1995; Gelderblom *et al.*, 1996; Musser & Plattner, 1997; Gelderblom *et al.*, 2001a;

Gelderblom *et al.*, 2001b; Marasas, 2001; Gelderblom *et al.*, 2004; Desjardins *et al.*, 2005; Desjardins, 2006). These toxins are natural contaminants of cereal grains worldwide and are mostly found in maize and products derived from maize. Fumonisins were first isolated from cultures of *F. verticillioides* strain MRC 826 (Gelderblom *et al.*, 1988) and later structurally elucidated as the di-ester of propane-1,2,3-tricarboxylic acid (Bezuidenhout *et al.*, 1988). They are polar compounds that are soluble in water and aqueous solutions of methanol and acetonitrile, but are not soluble in non-polar compounds (Jackson & Jablonski, 2004). Fumonisins have cancer-promoting activity (Gelderblom *et al.*, 1988) and have been associated with equine leukoencephalomalacia (ELEM) in horses and oesophageal cancer (EC) in humans in the Transkei region of South Africa (Marasas, 2001). Furthermore culture material on maize of *F. verticillioides* strain MRC 826, which was isolated from mouldy maize in Transkei, was shown to cause ELEM in horses, porcine pulmonary edema (PPE) syndrome in pigs, and liver cancer in rats (Gelderblom *et al.*, 1988; Gelderblom *et al.*, 2001a; Gelderblom *et al.*, 2001b; Marasas, 2001; Gelderblom *et al.*, 2004).

Fumonisins have been isolated from symptomatic as well as asymptomatic kernels (Bacon *et al.*, 1992; Bacon & Hinton, 1996; Munkvold & Desjardins, 1997; Munkvold *et al.*, 1997; Fandohan *et al.*, 2003) and the degree of infection is not correlated with elevated concentrations of fumonisins. It is thus important to be able to detect kernels with asymptomatic infection since they also pose a risk to human and animal health. In addition, these are more likely to go unnoticed because of the lack of visible symptoms. An objective, rapid method of detection is thus required that will ensure that asymptomatic kernels do not enter the food chain. Rapid, objective techniques have been evaluated for the detection and determination of fungi as well as the respective mycotoxins produced. A rapid technique for the detection of aflatoxin had been proposed (Hesseltine & Shotwell, 1973) that involves a greenish-gold fluorescence produced under ultraviolet light (365 nm) by maize kernels containing the mycotoxin.

NIR has been evaluated for the detection of maize kernels infected by a variety of fungi with a classification rate of up to 96% (Pearson & Wicklow, 2006). The method showed to be sensitive at spectral bands at 715 and 965 nm and could correctly identify 98.1% of asymptomatic kernels and 96.6% of kernels showing extensive discolouration. In another study, NIR, in the region of 500 – 1700 nm, was used to separate kernels contaminated with mycotoxins (Pearson *et al.*, 2004). It was found that absorbances at 750 and 1200 nm could accurately identify more than 99% of the kernels as contaminated with aflatoxin. In a similar study, transmittance spectra (500 – 950 nm) and reflectance spectra (550 –

1700 nm) were analysed to determine if aflatoxin contamination in single whole maize kernels could be distinguished (Pearson *et al.*, 2001). More than 95% of the kernels were correctly classified as containing either high (>100 ppb) or low (<10 ppb) levels of aflatoxin; however, classification accuracy for kernels between 10 and 100 ppb was only 25%. The rapid detection of kernel rots and mycotoxins in maize have also been evaluated using NIR spectroscopy (Berardo *et al.*, 2005). The best predictive ability for the percentage of global fungal infection and *F. verticillioides* was obtained using a calibration model utilising maize kernels (r^2 = 0.75; standard error of cross validation (SECV) = 7.43) and maize meals (r^2 = 0.79; SECV = 10.95), respectively.

Recently, NIR hyperspectral imaging has been evaluated for the fungal detection in wheat with classification rates of up to 97.8% for the two-class discriminant classification and 95% for the four class discriminant classification (Singh *et al.*, 2007). In a similar study, classification rates of up to 100% have been achieved for the detection of fungal infected wheat kernels (Zhang *et al.*, 2007). Using hyperspectral imaging at shorter wavelengths (458, 541 and 743 nm), a classification rate of 97.7% was achieved for the differentiation of toxigenic fungi (Yao *et al.*, 2008).

Near infrared (NIR) spectroscopy

The discovery of NIR energy is ascribed to Herschel who had recorded the first NIR spectrum while measuring heat energy of solar emission beyond the red portion of the visible spectrum (Herschel, 1800). The NIR region spans the range of the electromagnetic spectrum between 780 – 2500 nm and consists of absorption bands corresponding to overtones and combination bands of fundamental C-H, O-H, C-O and N-H vibrations because of the large anharmonicity of those vibrations involving the light hydrogen atom (Osborne *et al.*, 1993b; Bokobza, 1998; Barton, 2002; Bokobza, 2002; Reich, 2005; Sakudo *et al.*, 2006; Cen & He, 2007; Nicolaï *et al.*, 2007; Roggo *et al.*, 2007; Williams, 2007; Siesler, 2008). These bonds (C-H, O-H, C-O and N-H) are subject to vibrational energy changes when irradiated by NIR frequencies (Bokobza, 1998; Barton, 2002; Bokobza, 2002; Cen & He, 2007a). Two vibration patterns exist in these bonds: stretching vibration and bending vibration. Stretching is characterised as a continuous change in the inter-atomic distance along the axis between two atoms and bending is defined as a change in the bond angle between two atoms (Workman, 1993). NIR absorption occurs when the vibrations at a given frequency coincides with those of a molecular bond in the material being scanned. Thus, NIR spectroscopy includes the measurement of the absorption of NIR radiation from molecules within a given sample.

Due to the broad band nature of NIR spectra that consists of overlapping combination and overtone bands, individual chemical species are not always well resolved (Workman, 1993). In addition, many compounds absorb NIR energy throughout the entire wavelength region making it difficult, if not impracticable, to clearly resolve a usable standard for quantitative methods. Thus, chemometrics and chemometric techniques are essential for the decomposition and interpretation of NIR spectroscopic data. Techniques such as: multiple linear regression (MLR), principal component regression (PCR), partial least squares (PLS), partial least squares-discriminant analysis (PLS-DA) (Barker & Rayens, 2003) and linear discriminant analysis (LDA) were the basic techniques necessary for successful interpretation, decomposition, cluster analysis and quantification of sample data. However, as the technology and instruments became more sophisticated, newer techniques were required for the ever-growing data sets produced.

The emergence and acceptance of NIR spectroscopy into the analytical world as a recognised technique commenced with the work of Karl Norris of the US Department of Agriculture, Agricultural Research Service in the early 1960s (Norris, 1962; Norris, 1964). Since then the applications of NIR spectroscopy has spread throughout the scientific world into various disciplines. It has been evaluated in food and agriculture (Szabo *et al.*, 1991; Siska & Hurburgh, 1994; Lammertyn *et al.*, 2000; McGlone *et al.*, 2002; Lovett *et al.*, 2004; Lovett *et al.*, 2005; Gómez *et al.*, 2006; Han *et al.*, 2006; Siuda *et al.*, 2006; Uddin *et al.*, 2006; Zhao & Yan, 2006; Juhász *et al.*, 2007; Masoero *et al.*, 2007; Nicolaï *et al.*, 2007; Ortiz-Somovilla *et al.*, 2007; Yang *et al.*, 2008); medicine (Attas *et al.*, 2002; Sakudo *et al.*, 2006; Tisdall *et al.*, 2007); pharmaceutical industry (Reich, 2005; Rodionova *et al.*, 2005; Roggo *et al.*, 2007); and animal science (Cozzolino *et al.*, 2006b; Landau *et al.*, 2006).

Despite the fact that NIR spectroscopy has been and is currently being successfully evaluated in various subject areas, it has one inadequacy that has of late become of paramount significance: the lack of spatial information. With the advance of technology and the ever growing concern of quality and safety, it has become important to trace molecules, chemicals or components within samples. This could, for example, lead to superior quality control and efficient monitoring of processes within in a factory. Through the combination of digital imaging with NIR spectroscopy, a new form of imaging spectroscopy was born, i.e. NIR hyperspectral imaging.

Principles of NIR hyperspectral imaging

What is NIR hyperspectral imaging?

NIR hyperspectral imaging also referred to as NIR chemical imaging, NIR spectroscopic imaging or hyperspectral NIR imaging, is an emerging, cutting-edge analytical technique that integrates conventional digital imaging and NIR spectroscopy to attain both spatial and spectral information from an object (Koehler *et al.*, 2002; Cogdill *et al.*, 2004; Burger, 2006; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Gowen *et al.*, 2008a). In contrast to digital imaging that captures a complete set of information regarding an object in only three layers, red, green and blue (RGB), NIR hyperspectral imaging acquires data pertaining the object in many layers, varying from 118 to 229, depending on the instrument. It is not only capable of identifying the chemical species and determining the concentration present in a sample, but it is also able to indicate location.

NIR hyperspectral instrument configurations and image acquisition

A typical NIR hyperspectral instrument is composed of a camera, a spectrograph and detector and an illuminator all coupled to a computer (Tatzer *et al.,* 2005; Gowen *et al.,* 2008a). The camera unit in most cases posses an InGaAs (Indium Gallium Arsenide) detector sensitive in the NIR region, a PbS (lead sulphide) detector or a HgCdTe (mercury-cadmium-telluride) detector. The spectrograph which either comprises a liquid crystal tunable filter (LCTF), acousto-optic tunable filter or a prism-grating-prism (PGP) is responsible for wavelength selection, separation and measurement. The illuminator is simply composed of a light source, normally tungsten halogen or xenon gas plasma lamps (Geladi *et al.,* 2007). Various instrument configurations exist allowing for different image acquisition, consequently, resulting in the same NIR hyperspectral image. At present three instrument configurations exist: point scan imaging configuration, line scan or pushbroom configuration and focal plane scan or staring imager configuration.

Point scan imaging configuration

In this configuration, as the name implies, a complete NIR spectrum is obtained from the diffuse reflectance at a point on the sample (Burger, 2006; Geladi *et al.*, 2007). The sample is then repositioned prior to acquiring a current spectrum. Through repositioning the sample proportionately in two spatial dimensions (x and y) a comprehensive hyperspectral image can be obtained (Burger, 2006; Geladi *et al.*, 2007). Instrument calibration for reflectance and wavelength is only required once preceding commencement of image acquisition. Very stable high resolution spectra are obtained from this system;

however, image acquisition is very time consuming due to sample repositioning. In addition, the field of view is limited. This configuration is commonly used for microscopes.

Focal plane scan/staring imaging configuration

Within the focal plane scan configuration (**Fig. 2.3a**) the sample and all spectrometer components remain stationary while an image of the entire field of view is acquired (Burger, 2006; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Gowen *et al.*, 2008a). An image of the sample at each wavelength (depending on the capability of the detector and spectrograph) is recorded resulting in images slices *i.e.* there is one image per wavelength stacked sequentially (Koehler *et al.,* 2002).

Line scan/pushbroom configuration

The line scan configuration (**Fig. 2.3b**) uses a two dimensional detector perpendicular to the surface of the sample (Burger, 2006; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Gowen *et al.*, 2008a). The sample is imaged by a narrow line of radiation falling on the sample or by a narrow slit in the optical path leading to the detector (Burger, 2006; Geladi *et al.*, 2007). Thus, the spectrum of each pixel in a line of sample is captured which is then concurrently recorded by the array detector (Gowen *et al.,* 2007; Gowen *et al.,* 2008a). Since no filter changes are required, image acquisition time is considerably faster than the point scan configuration and the focal plane scan imaging configuration. This configuration is especially well-suited to online scanning on conveyor belts.
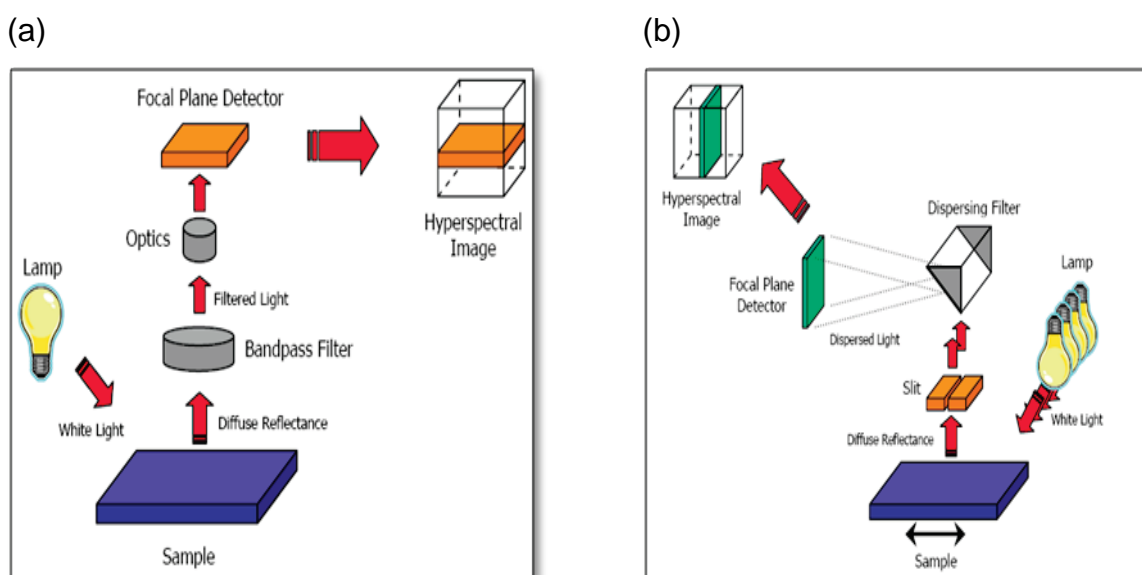
(a)   (b)

Figure 2.3 (a) Focal plane scan/staring imaging configuration, and (b) line-scan configuration showing how images of samples are acquired and hypercube constructed (Burger, 2006).

NIR hyperspectral images

NIR hyperspectral images are composed of hundreds of neighbouring wavebands for each spatial position of an object studied (Martinsen *et al.*, 1999; Koehler *et al.*, 2002; Cogdill *et al.*, 2004; Geladi *et al.*, 2004; Dubois *et al.*, 2005; Ariana *et al.*, 2006; Burger, 2006; Burger & Geladi, 2006; Dubois *et al.*, 2007; Gowen *et al.*, 2007; ElMasry *et al.*, 2008; Gowen *et al.*, 2008a).   They are created when hundreds of single black and white or grayscale images are layered; where each grayscale image represents a single band of spectral wavelengths (Geladi *et al.*, 2004; Burger, 2006; Burger & Geladi, 2006).   Each pixel in an NIR hyperspectral image contains the spectrum of that exact location (Burger, 2006; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Gowen *et al.*, 2008a).   The composition of that particular pixel can be characterised by the ensuing spectrum that acts as identity verification.   NIR hyperspectral images, known as hypercubes, are three-dimensional cubes of data, comprised of two spatial (x and y) and one wavelength dimension (z) (**Fig. 2.4**).   Since regions of a sample with similar spectral properties have similar chemical composition, separated into particular areas of the image, the hypercube permits the visualisation of the biochemical components within the sample (Cogdill *et al.*, 2004; Burger, 2006; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Gowen *et al.*, 2008a).   NIR hyperspectral images may be perceived as chemical maps (Koehler *et al.,* 2002).   The intensity of a pixel in the hypercube plotted as a function of the wavelength dimension represents a standard NIR spectrum and the intensity for all pixels at a single wavelength represents an image of absorption.   By isolating different chemical species to different spectral regions of the chemical map, the wavelength dimension visualises chemical specificity (Koehler *et al.*, 2002; Burger, 2006; Gowen *et al.*, 2007; Gowen *et al.*, 2008a; Gowen *et al.*, 2008b).

   As with single-point NIR spectroscopy, chemometrics must be applied to extract relevant information relating to the spectral content, allowing sample classifications or quantitative determinations (Burger & Geladi, 2006).
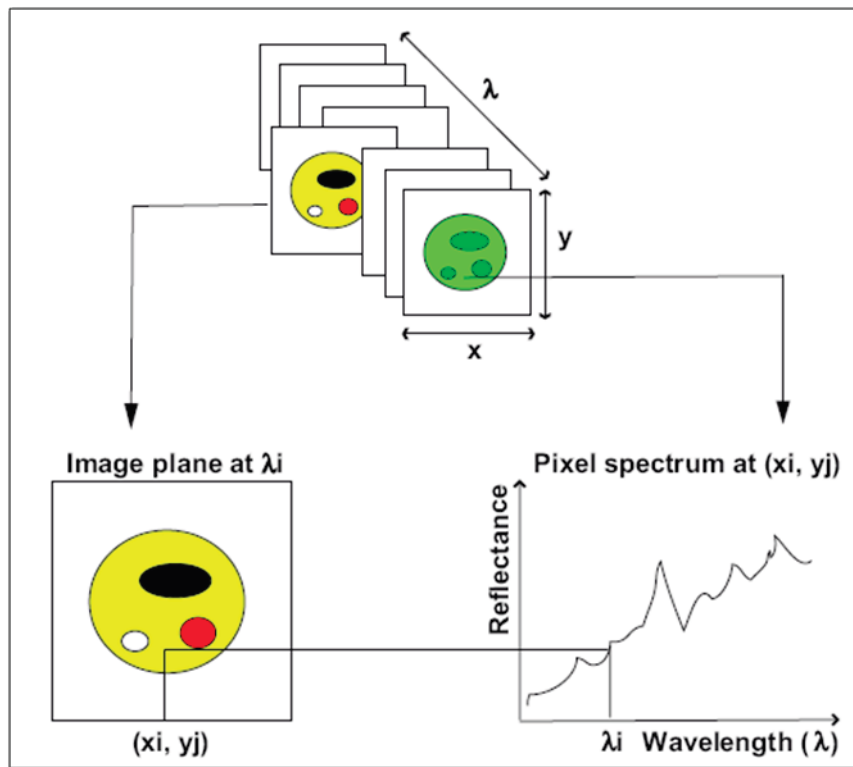
Figure 2.4 Schematic representation of a NIR hyperspectral image or hypercube showing the relationship between spectral and spatial dimensions (Gowen *et al.*, 2007).

Chemometrics and multivariate data analysis (MDA)

Chemometrics can be defined as the chemical discipline concerned with the use of mathematical, statistical and other methods to design or select optimal measurement procedures and experiments and to provide maximum relevant chemical information by analysing chemical data (Massart *et al.*, 1988; Brown, 1995; Bokobza, 1998; Næs *et al.*, 2002). Another term most frequently used is multivariate data analysis (MDA) that is synonymous with chemometrics, however is often confused. Chemometrics is the science, while MDA is the process applying the chemometric techniques. The chemometric approach is frequently used to relate physical or chemical properties of investigated samples to the absorption of energy in the NIR wavelength range (Bokobza, 1998).

To improve interpretation of the following section, a few important points are listed:

- samples are referred to as *objects*;
- measurement results (e.g. concentration, pH, absorbances) are referred to as *variables*; and
- a data table consisting of K variables determined for I objects is referred to as a *data matrix* of *size $I * K$* (**Fig. 2.5**).
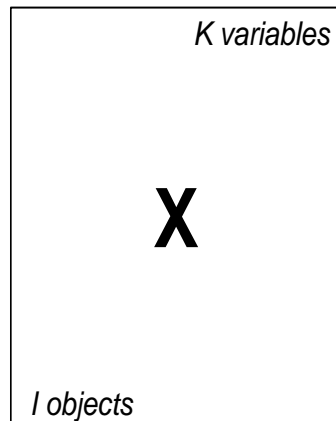
Figure 2.5 A data matrix of size *I* * *K;* where *K* can be anything from three to≥ 50 000 and the *I* ≥ 100 000.

A great amount of the work in chemometrics involves extracting meaningful information about the objects and the variables from data matrices as depicted in **Fig. 2.5** (Geladi, 2003). The observation that these data matrices often contain superfluous information and can be attenuated considerably is the concept of chemometrics. Thus chemometrics seeks to decompose these often colossal matrices into fewer, more manageable components. The reduced terms are easier to interpret, have more stability and the residual contains noise or less useful information (Geladi, 2003). The aforementioned is the basics of a powerful chemometric technique, principal component analysis (PCA,) that is used most frequently (Cowe & McNicol, 1985). However, prior to delving into the various chemometric techniques a few definitions and notations are in order. A *matrix* is a rectangular array of numbers or functions (Gorsuch, 1974). A capital letter is used to signify the entire array. The following nomenclature is applicable:

- x   Single scalar value
- **x**   Vector of scalar values
- **X**   Two-dimensional matrix of scalar values
- **X̱**   Three-dimensional array of scalar values
- **X**^{T} The transpose of the matrix **X**

MDA is about the extraction of information from a data matrix or three-way array. Different goals, methods of data generation and nature of the sample determine which chemometric

technique is most applicable. The first step in MDA is data exploration. This basically involves getting to know the data by surveying it for intriguing phenomena without prior expectations. This allows for the observation and detection of clusters and clustering of objects, outliers and gradients between clusters. An excellent means of data exploration is PCA that projects the large data space in the $I * K$ matrix into a smaller space that is easier to analyse (Jolliffe, 1986). It is also known as an unsupervised means of classification.

Principal component analysis (PCA)

PCA constitutes the most basic work horse of all of MDA (Esbensen, 2006). The central idea of PCA is to reduce the dimensionality of a data set which consists of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set (Cowe & McNicol, 1985; Jolliffe, 1986; Massart *et al.*, 1988; Bokobza, 1998; Næs *et al.*, 2002; Geladi, 2003; Benito & Peña, 2005; Reich, 2005; Esbensen, 2006; Roggo *et al.*, 2007). This is achieved by transforming to a new set of variables, the principal components (PC's), which are uncorrelated and which are ordered so that the first few retain most of the variation present in all of the original variables (Jolliffe, 1986; Massart *et al.*, 1988; Bokobza, 1998; Næs *et al.*, 2002; Geladi, 2003; Benito & Peña, 2005; Reich, 2005; Esbensen, 2006; Roggo *et al.*, 2007):

PCA: $\mathbf{X} = \mathbf{t}_1\mathbf{p}'_1 + \mathbf{t}_2\mathbf{p}'_2 + ...+ \mathbf{t}_A\mathbf{p}'_A + \mathbf{E}$ 　　　　　　…equation 1

where $\mathbf{X}$ is an ($I * K$) matrix usually mean-centred or scaled in some way, the $\mathbf{t}_A$ are the *score* values for the ath component, $\mathbf{p}'_A$ are the *loading* values for the ath component and $\mathbf{E}$ is the residual ($I * K$) matrix. A more general name for the components is *latent variable* or more commonly *principal component*. Equation one can be viewed geometrically as illustrated in **Fig. 2.6**.

Once the data has been decomposed, the resultant scores and loadings can be viewed in scatter plots and line plots that allow for efficient interpretation of the whole data space. Studying these plots is fundamental to data exploration. The scores portray the correlation between the objects while the loadings express the variation within the variables. They are used simultaneously to understand the nature of the data such as the cause of the variation and which component drives it. In spectroscopy this is especially useful as it (the loading line plot) indicates which variable(s) (wavelength) is responsible for variation. Thus, the score plot indicates whether variation is present while the loading line plot shows the variable accountable.
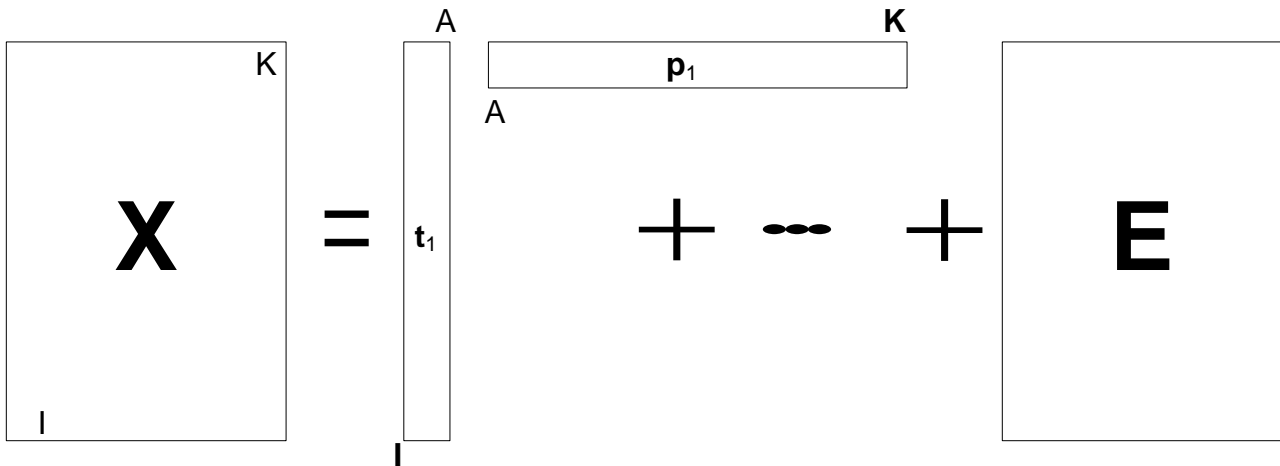
Figure 2.6 A data matrix of size ($I * K$) is reduced to smaller matrices of size (I * A) and (A * K) that are easier to interpret and understand in addition to containing all the relevant information. Noise and other disturbances are left in the residual matrix of size I * K. Adapted from (Geladi, 2003; Esbensen, 2006).

PCA is the first step in MDA; all other calculations follow on the new variable space. Because of its success in dimensionality reduction, it is an excellent tool for image analysis especially for the analysis of NIR hyperspectral images where it is integral. PCA has been successfully evaluated in various applications, i.e.

- to detect bruise damage on white mushrooms (Gowen *et al.,* 2008b) and pickling cucumbers (Ariana *et al.,* 2006);
- to analyse multispectral images of works of art for classification of painted zones by pigments with different chemical composition or physical properties (Baronti *et al.,* 1997);
- as a fast dimensionality reduction of image data (Benito & Peña, 2005);
- for the investigation of tomato authenticity (Arvanitoyannis & Vaitsi, 2007);
- image texture analysis (Bharati *et al.,* 2004) and image compression using principal component neural networks (Costa & Fiori, 2001);
- spectroscopic monitoring of red wine fermentation (Cozzolino *et al.*, 2006a) and
- crop-weed discrimination by line-image spectroscopy (Borregaard *et al.,* 2000).

Although very successful, PCA has the disadvantage that the loadings are arbitrary values and thus can be positive or negative. This makes interpretation of loading line plots rather tricky as uncertainty often arises when deciding whether to analyse peaks or valleys as

important absorption areas. Nevertheless, when classifying it is often necessary to discriminate between classes using models constructed from unsupervised classification. This then leads to supervised classification, whereby data are classified according to known information. An excellent means of discriminating between classes is the use of partial least squares-discriminant analysis (PLS-DA).

Partial least squares-discriminant analysis (PLS-DA)

Partial least squares (PLS) was not initially designed as a tool for statistical discrimination (Barker & Rayens, 2003), in fact it was intended for the modelling of complicated data sets in terms of chains of matrices (Wold *et al.*, 2001a). The fact that the studied system or process is driven by a small number of latent variables, and that these latent variables are weighted averages of the observed variables, is the fundamental theory of any PLS model (Wold *et al.*, 2001b). Therefore, PLS is a method for the indirect observation of the latent variables.

The way in which PLS facilitates the classification varies according to the user. The idea is that effective group membership is predicted. For this reason a dummy matrix (**Y**) that records the membership with ones and zeros is paired with a training set (**X**) and PLS is implemented in the normal way (**Fig. 2.7**) (Barker & Rayens, 2003; Chevallier *et al.*, 2006).

Pre-processing

Spectroscopic data and image data often contain disturbances caused by scattering due to surface inhomogeneities. Pre-processing is usually performed to remove non-chemical biases from the spectral information and prepare the data for further processing. A number of techniques exist; however, for the purpose of this thesis only two techniques, i.e. multiplicative scatter correction (MSC) and standard normal variate (SNV) will be discussed.

The basis of the MSC is the fact that physical light scatter has wavelength dependencies different from that of chemically absorbed light and separates chemically absorbed light from physical light scatter (Geladi *et al.*, 1985; Isaksson & Næs, 1998). With the aid of data from many wavelengths, it is possible to distinguish between absorption and scatter. The scatter for each sample is estimated relative to that of an ideal sample and each sample's spectrum is then corrected so that all samples appear to have the same scatter level as the ideal.
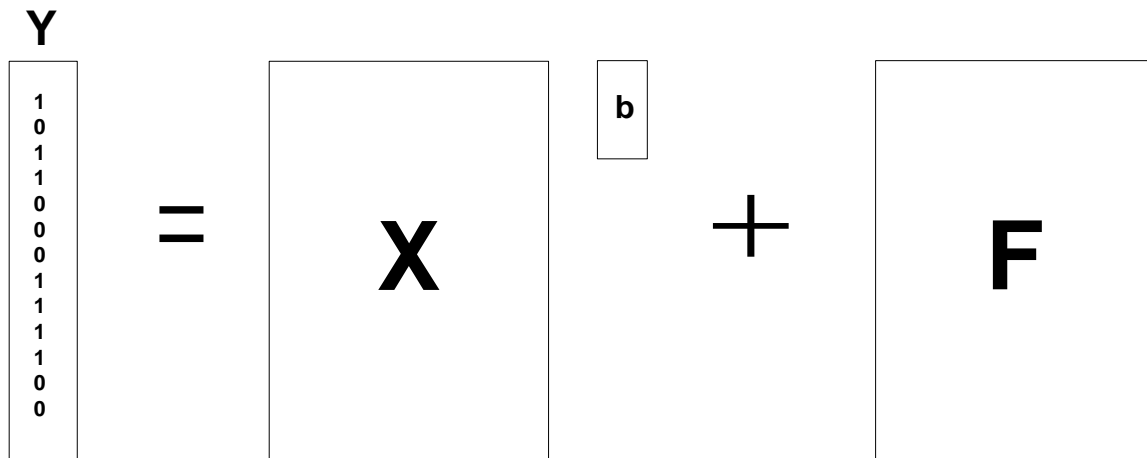
Figure 2.7 Diagrammatic representation of the implementation of PLS-DA. **Y**, a dummy matrix with ones and zeros, is paired with **X** to obtain **b**, regression coefficients, and **F** the residual.

The correction is carried out based on the supposition that all the samples have equivalent scatter coefficients at all the NIR wavelengths. In MSC, the mean spectrum is calculated from all the spectra in a defined data set (Pizarro *et al.*, 2004). Afterwards, a least squares linear regression is performed on absorbance values of the sample spectrum versus those at corresponding wavelengths in the mean spectrum. This function provides a linear equation (equation 2) with a definite intercept and slope. Subsequently, the value of the intercept is subtracted from each data point in the spectrum. Thus MSC rotates each spectrum so that it fit as closely as possible to the mean spectrum (Osborne *et al.*, 1993a). Ultimately, each absorbance value in the resultant spectrum is divided by the value of the slope as illustrated in equation 3.

$$x_i = a_i + b_i \bar{x}_j + 0e_i \qquad\qquad \text{...equation 2}$$

where x_i is an individual spectrum, i, $\bar{x}_j$ the mean spectrum of the data set and e_i the residual spectrum, which ideally represents the chemical information in the data. The fitted constants a_i (intercept) and b_i are used to correct each value of the spectrum i.

$$x_{i,MSC} = \frac{(x_i - a_i)}{b_i}$$ ...equation 3

MSC attempts to remove multiplicative and additive effects of the scatter in NIR measurements, minimising spectral variations that are not due to the analyte concentration.

Standard normal variate (SNV) (Barnes *et al.*, 1989) centres and scales individual spectra, having an effect similar to that of MSC (Osborne *et al.*, 1993a; Næs *et al.*, 2002). The main practical difference is that SNV standardises each spectrum using only the data from that spectrum (Næs *et al.*, 2002), in other words it does not use the mean spectrum of any set. Arithmetic conversion of the spectra by computation of the standard normal variate at each wavelength removes slope deviation on an individual sample basis by means of equation 4:

$$SNV = \left(\frac{x_i - \bar{x}_i}{s_i} \right)$$ …equation 4

where x_i is the spectral measurement from the whole spectrum, $\bar{x}_i$ the mean of each spectrum and s_i the standard deviation of each spectrum.

Hyperspectral image analysis

A typical commercial instrument produces hypercubes with dimensions 256×320×118 (Burger & Geladi, 2006). This may be construed as 118 single channel images each with 256×320 pixels. Alternatively, this hypercube may be considered as 81 920 spectra, each with 118 wavelength channels resulting in 9 666 560 data points. This colossal accumulation of data poses data extraction challenges; however, it creates novel possibilities.

Hyperspectral image analysis (HIA) is a relatively new term (Grahn & Geladi, 2007) that is an extension of multivariate image analysis (MIA). As the name suggests, the amount of variables involved is far more than that dealt with in MIA. MIA is a methodology for analysing multivariate images, where the image coordinates are location (two- or three dimensional, usually pixel number) and variable number (Esbensen & Geladi, 1989; Geladi & Grahn, 1996). The methodology was based on a specific, strategic, step-by-step approach toward analysing multivariate images (Esbensen & Geladi, 1989). The approach contained the following operations:

1. Calculating principal component scores (score images) and loadings (vectors)

2. Scatter plots of scores or loadings against each other

3. Selection of classes on the score plots

4. Brushing of classes in multiple score plot split-screen display

5. Projection of the pixels in the feature-space classes to the corresponding scene-space locations

6. Calculation of local PC-models, as determined in the score plots

7. Calculation of residual images with respect to such local models

8. Auxiliary functions for overlay masking, overlay toggle and colour slicing

HIA can be approached similarly, however, vast improvements in technology in the past 19 years, since this method was published, has made this approach more simplistic and elementary. In other words, HIA involves similar steps; nevertheless the process is easier and faster. Thus the MIA approach was modified and adapted to create an efficient, effective and fast method of analysing hyperspectral images. The HIA approach contains the succeeding operations:

1. Image correction (correct image for dark and white references and convert to absorbance *i.e.*

2. Image cleaning – calculate PCA (score images and score plots) and study both plots; use various combinations of components plotted against each other in conjunction with score image to identify and remove background, shading, dead pixels and detector errors

3. Add more components to model

4. Examine various combinations of components plotted against each other in score plots and survey for clusters or clustering of pixels

5. Identify and determine (anatomically or histologically) what the clusters are using score plot, score image and knowledge of sample

6. Assign clusters to classes

7. Study loading line plots and identify important absorption peaks and wavebands

8. If necessary apply appropriate pre-processing technique

Dependant on the desired outcome of an analysis, more steps may be included or some may be removed. These steps are the general outline of the HIA approach and are usually applied first before any quantitative analyses can be done.

For the purpose of uniformity and clarity throughout this chapter, data analysis and exploration will be discussed with specific reference to the Evince version 2.020 (UmBio AB, Umeå, Sweden) hyperspectral image analysis software package. The HIA approach from step two will be illustrated with the analysis of NIR hyperspectral images of freeze dried devil's claw (*Harpagophytum procumbens*) root slices with the active component harpagoside and three commercial products manufactured from devil's claw. Devil's claw (*Harpagophytum procumbens*) is an herbaceous, medicinal plant indigenous to the dry, arid plains of southern Africa and is found particularly in the Kalahari Desert (Watt & Breyer-Brandwijk, 1962; Baranska *et al.*, 2005; Stewart & Cole, 2005; Catelan *et al.*, 2006; Ludwig-Müller *et al.*, 2008). It has been used by the inhabitants of southern Africa for treating a large number of human ailments including fever, diabetes, diarrhoea and blood diseases (Stewart & Cole, 2005).

Similar to the MIA approach (Esbensen & Geladi, 1989), the starting point of HIA is a traditional PCA of the image to be studied, generating score images (contour plot in Evince), score plots (scatter plot) and loading line plots. Thereafter, utilising the duality and interactivity of the score plot and score image, the data analyst becomes acquainted with the image data set. In addition the initial PCA, which has as default three components in Evince, is used for image cleaning purposes whereby background effects, shading effects, dead pixels (if any and instrument dependant), improper illumination, detector errors and detector saturation can be identified and removed. To avoid repetition of the aforementioned list, they will henceforth collectively be cited as bad pixels. Firstly these bad pixels are identified and their location determined through trail-and-error. Pixels in the score plot (**Fig. 2.7a**) are selected and because of the duality of the score plot and score image, these areas are simultaneously highlighted in the score image (**Fig 2.7b**) indicating their location within the image. When satisfied with the selection of these pixels, they can be excluded, and by simply "applying changes" the highlighted pixels are removed from the image and no longer form part of further calculations. Once the bad pixels have been removed additional principal components are added to the model with a consequent vast improvement in the clarity of the score image (**Fig. 2.8a**). In addition the score plot (**Fig. 2.8b)** becomes more dispersed and clusters that were once hidden are more apparent.

Once the image has been cleaned and additional components have been added, the plots of the various components plotted against each other in score plots are studied to determine which component(s) contribute most to distinctly separating the clusters. For this illustration it was established that when plotting t2 versus t6 (**Fig. 2.9a**) the data was

clearly separated into three separate clusters and one high-density cluster; however, when plotting t4 versus t6 (**Fig. 2.9b**) it was found that the data was even more uncluttered and five definite clusters were visible. One expected to see five clusters, since it is known that that there were objects in the image. If the objects are very similar chemically then overlapping of clusters would occur. This is likely to be the case in the high-density cluster found in both plots (**Fig. 2.9a** and **Fig. 2.9b**). Consequently these score plots were compared with the corresponding score image to relate the clusters to the root and tablets in the score image. Normally the clusters would be selected in the score plot and interactively projected onto the score image for localisation of classes. However, due to the distinct objects being imaged, for the purpose of this illustration, each object in the score image was assigned to a class and colour coded (**Fig. 2.10**). This was visualised interactively in the score plot (**Figs. 2.11a & b**). Distinct classes, related to the freeze dried devil's claw root, the three different tablets and the pure harpagoside were observed in the score plot. As this approach only serves as an illustration to be followed when evaluating hyperspectral images, this example will not be discussed in further detail.

Subsequently the loading line plots for each principal component is studied to establish which wavelengths are critical and to determine what, if any, absorption peaks are present. When peaks are correlated with a table of known absorptions in the NIR region(Osborne *et al.*, 1993b), the chemical constituents responsible for the variation within the component can be identified. Depending on the nature of the sample and the slope of the loading line plots, appropriate pre-processing techniques may be applied. In addition if prior knowledge of the sample matter is known, supervised chemometric techniques such PLS-DA (Barker & Rayens, 2003) may be performed.
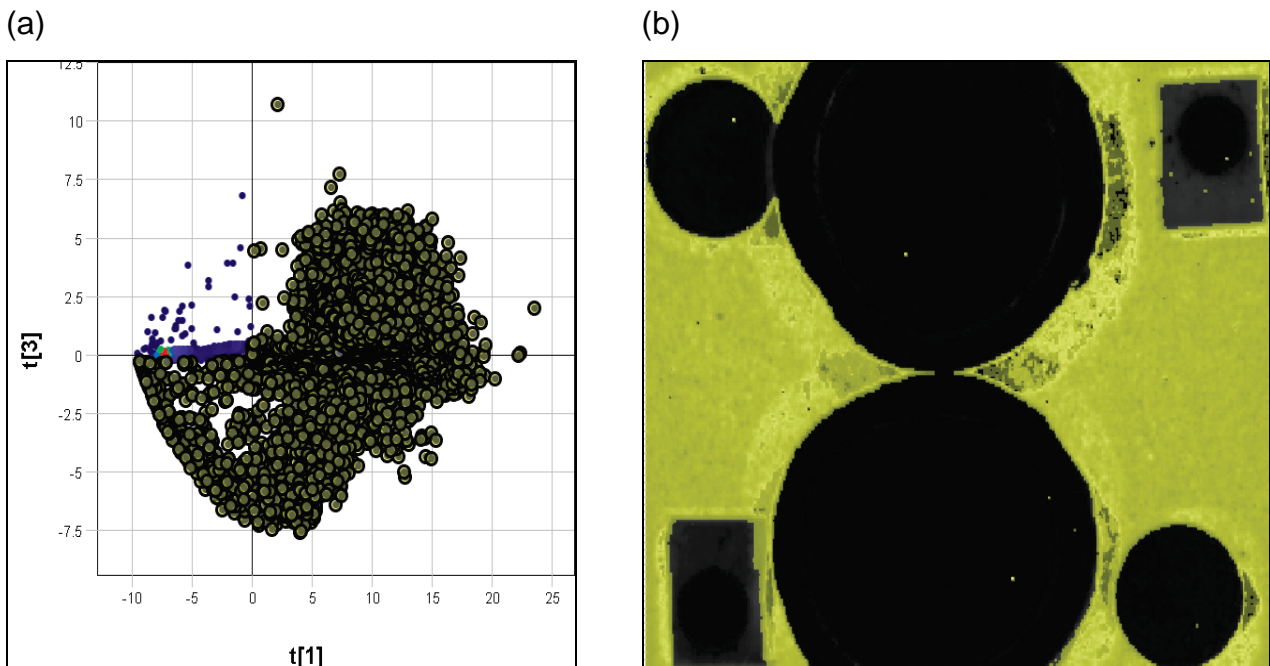
(a)

(b)

Figure 2.7 (a) PCA score plot of t1 versus t2 illustrating manually selected pixels; and (b) highlighted area (yellow) in score image corresponding with selected pixels of (a).
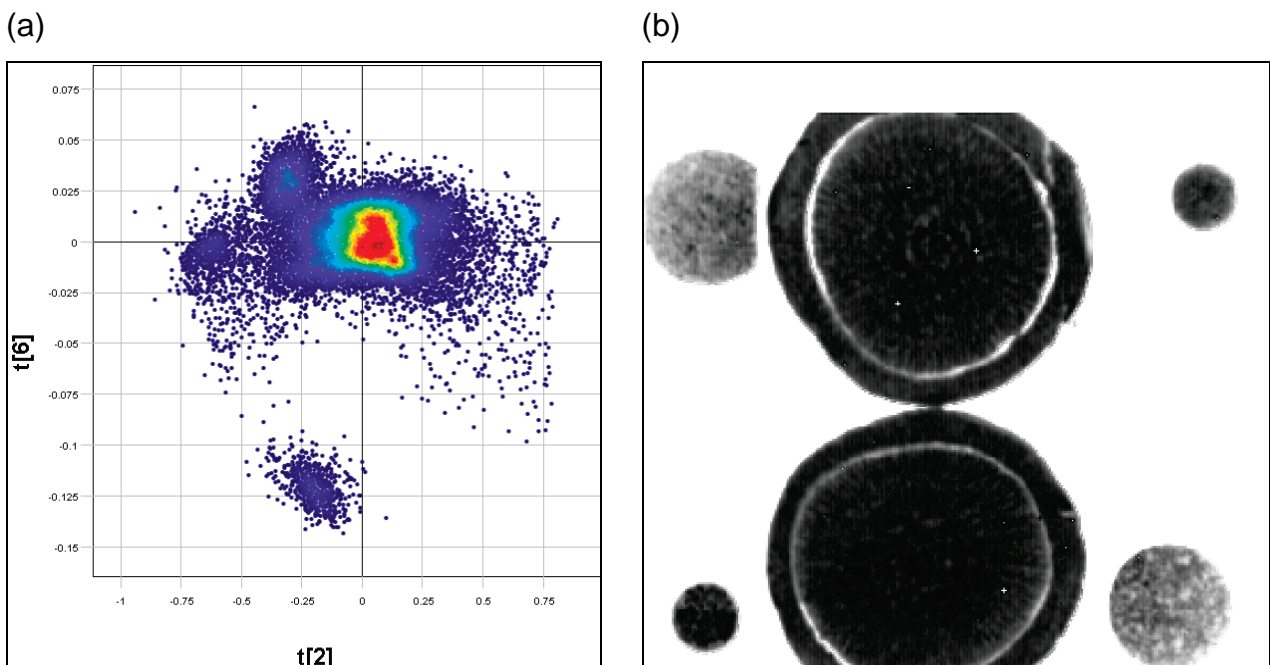
(a)

(b)

Figure 2.8 (a) Score plot of t2 versus t6 subsequent to bad pixels being removed and supplementary components added. (b) PCA score image after bad pixels have been removed and additional components added.
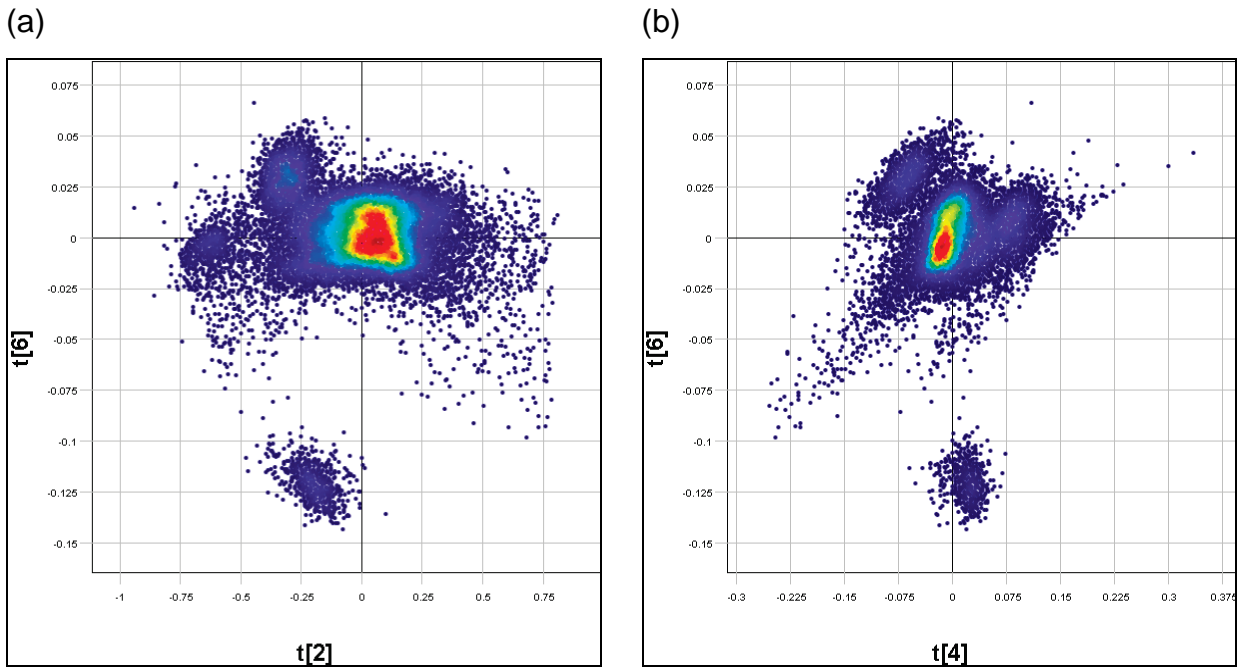
Figure 2.9 (a) Score plot of t2 versus t6 illustrating three distinct clusters and one high-density displaying overlapping of clusters indicating chemical or biochemical similarities. (b) Score plot of t4 versus t6 displaying more distinct separation between clusters and slightly less overlapping.
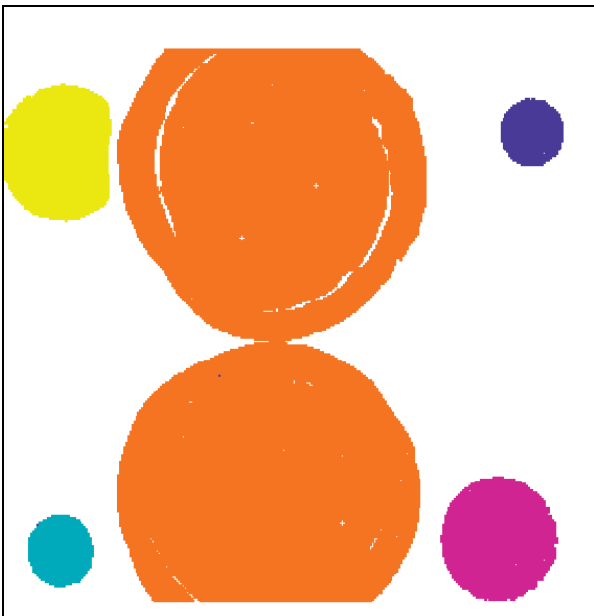

Figure 2.10 Classification image generated assigning each object in the score image to a class.
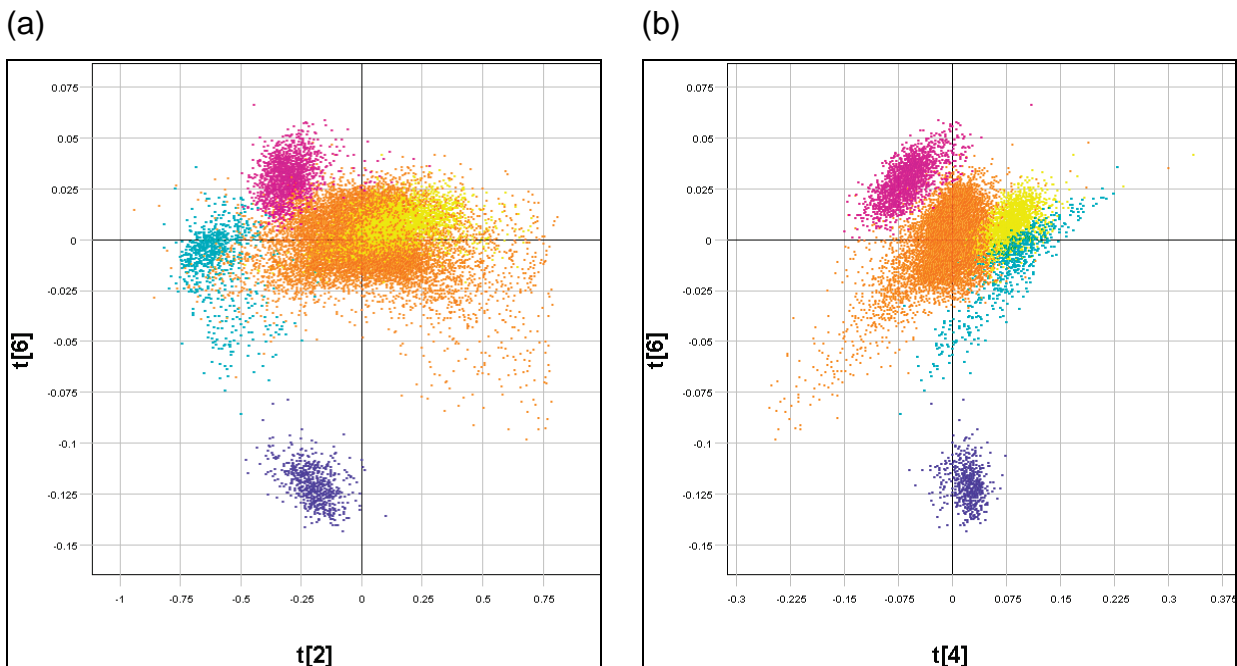
(a)

(b)

Figure 2.11 (a) score plot of t2 versus t6 indicating the clusters now assigned to classes. (b) plot of t4 against t6. For both plots the classes were assigned as follows (blue: pure harpagoside, cyan: table1, orange: devil's root slices, magenta: tablet2 and yellow: tablet3.

Conclusion

Rapid analytical methods are continually required to perform routine analyses that ensure accurate, reliable data. This is especially required in the food industry where routine analyses are performed without respite. NIR, a technique frequently used in the food industry, offers speed, accuracy, reliability and precision. However, in a world where consumers are more educated, there is a need to be cognisant about many aspects of products and processes. With the ability to identify and indicate location of possible contaminants, NIR hyperspectral imaging is an excellent tool for analyses.

Maize hardness analyses require revision since all the techniques are time consuming, destructive and require bulk samples for measurement. NIR hyperspectral imaging could prove to be beneficial since single kernel measurements are possible that will greatly assist in breeding trials. Identification of fungal infected kernels are very important, thus objective techniques are required for discrimination. Here too, NIR hyperspectral imaging would offer great potential as an objective means of discriminating between sound and fungal infected kernels.

Although much research is still required with regards to instrumentation, software and chemometrics, NIR hyperspectral imaging is a promising tool for the food industry of the future.

References

Abdelrahman, A.A. & Hoseney, R.C. (1984). Basis for hardness in pearl millet, grain sorghum and corn. *Cereal Chemistry*, **61**, 232-235.

Almeida-Dominguez, H.D., Suhendro, E.L. & Rooney, L.W. (1997). Factors affecting rapid visco analyser curves for the determination of maize kernel hardness. *Journal of Cereal Science*, **25**, 93-102.

Anonymous. (1996). Corn: layers and structures of corn kernel. [WWW document]. http://www.britannica.com/EBchecked/topic-art/103350/162/The-outer-layers-and-internal-structures-of-a-kernel-of. 17 December 2008

Ariana, D.P., Lu, R.F. & Guyer, D.E. (2006). Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture*, **53**, 60-70.

Armstrong, P.R., Lingenfelser, J.E. & McKinney, L. (2007). The effect of moisture content on determining corn hardness from grinding time, grinding energy, and near-infrared spectroscopy. *Applied Engineering in Agriculture*, **23**, 793-799.

Arvanitoyannis, I.S. & Vaitsi, O.B. (2007). A review on tomato authenticity: quality control methods in conjunction with multivariate analysis (chemometrics). *Critical Reviews in Food Science and Nutrition*, **47**, 675 - 699.

Attas, M., Posthumus, T., Schattka, B., Sowa, M., Mantsch, H. & Zhang, S. (2002). Long-wavelength near-infrared spectroscopic imaging for in-vivo skin hydration measurements. *Vibrational Spectroscopy*, **28**, 37-43.

Bacon, C.W., Bennett, R.M., Hinton, D.M. & Voss, K.A. (1992). Scanning electron microscopy of *Fusarium moniliforme* within asymptomatic corn kernels and kernels associated with leukoencephalomalacia. *Plant Disease*, **76**, 144-148.

Bacon, C.W. & Hinton, D.M. (1996). Symptomless endophytic colonization of maize by *Fusarium moniliforme*. *Canadian Journal of Botany*, **74**, 1195-1202.

Bacon, C.W., Yates, I.E., Hinton, D.M. & Meredith, F. (2001). Biological control of *Fusarium moniliforme* in maize. *Environmental Health Perspectives Supplements*, **109**, 325-332.

Baranska, M., Schulz, H., Siuda, R., Strehle, M.A., Rösch, P., Popp, J., Joubert, E. & Manley, M. (2005). Quality control of *Harpagophytum procumbens* and its related phytopharmaceutical products by means of NIR-FT-Raman spectroscopy. *Biopolymers*, **77**, 1-8.

Barker, M. & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, **17**, 166-173.

Barnes, R.J., Dhanoa, M.A. & Lister, S.J. (1989). Standard normal variate transformatoin and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, **43**, 772-777.

Baronti, S., Casini, A., Lotti, F. & Porcinai, S. (1997). Principal component analysis of visible and near-infrared multispectral images of works of art. *Chemometrics and Intelligent Laboratory Systems*, **39**, 103-114.

Barton, F.E. (2002). Theory and principles of near infrared spectroscopy. *Spectoscopy Europe*, **14**, 12-18.

Benito, M. & Peña, D. (2005). A fast approach for dimensionality reduction with image data. *Pattern Recognition*, **38**, 2400-2408.

Bennet, E.H. (1950). Kernel hardness in corn. II. A microscopic examination of hard and soft types of dent corn. *Cereal Chemistry*, **27**, 232-238.

Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A. & Marocco, A. (2005). Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry*, **53**, 8128-8134.

Bezuidenhout, S.C., Gelderblom, W.C.A., Gorst-Allman, C.P., Horak, R.M., Marasas, W.F.O., Spiteller, G. & Vleggaar, R. (1988). Structure elucidation of the fumonisins, mycotoxins from *Fusarium moniliforme*. *Journal of the Chemical Society, Chemical Communications*, **11**, 743-745.

Bharati, M.H., Liu, J.J. & MacGregor, J.F. (2004). Image texture analysis: methods and comparisons. *Chemometrics and Intelligent Laboratory Systems*, **72**, 57-71.

Bokobza, L. (1998). Near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **6**, 3-17.

Bokobza, L. (2002). Origin of near-infrared absorption bands. In: *Near-Infrared Spectroscopy: Principles, Instruments, Applications* (edited by H.W. Siesler, Y. Ozaki & H.M. Heise). Pp. 11-39. Weinheim: Wiley-VCH Verlag.

Borregaard, T., Nielsen, H., Nørgaard, L. & Have, H. (2000). Crop-weed discrimination by line imaging spectroscopy. *Journal of Agricultural Engineering Research*, **75**, 389-400.

Brown, S.D. (1995). Has the chemometrics revolution ended? Some views on the past, present and future of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **30**, 49-58.

Burger, J. (2006). Hyperspectral NIR image analysis: data exploration, correction and regression. PhD Thesis Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, Umeå, Sweden

Burger, J. & Geladi, P. (2006). Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst*, **131**, 1152-1160.

Bush, B.J., Carson, M.L., Cubeta, M.A., Hagler, W.M. & Payne, G.A. (2004). Infection and Fumonisin roduction by *Fusarium verticillioides* in developing maize kernels. *Phytopathology*, **94**, 88-93.

Campbell, M.R., Brumm, T.J. & Glover, D.V. (1997). Whole grain amylose analysis in maize using near-infrared transmittance spectroscopy. *Cereal Chemistry*, **74**, 300-303.

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

Campbell, M.R., Sykes, J. & Glover, D.V. (2000). Classification of single- and double-mutant corn endosperm genotypes by near-infrared transmittance spectroscopy. *Cereal Chemistry*, **77**, 774-778.

Catelan, S.C., Belentani, R.M., Marques, L.C., Silva, E.R., Silva, M.A., Caparroz-Assef, S.M., Cumen, R.K.N. & Bersani-Amado, C.A. (2006). The role of adrenal corticosteroids in the anti-inflammatory effect of the whole extract of *Harpagophytum procumbens* in rats. *Phytomedicine*, **13**, 446-551.

Cavanaugh, K.J., Zehr, B.E., Nyquist, W.E., Hamaker, B.R. & Crane, P.L. (1995). Responses to Selection for Endosperm Hardness and Associated Changes in Agronomic Traits after 4 Cycles of Recurrent Selection in Maize. *Crop Science*, **35**, 745-748.

Cen, H.Y. & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, **18**, 72-83.

Chandrashekar, A. & Mazhar, H. (1999). The biochemical basis and implications of grain strength in sorghum and maize. *Journal of Cereal Science*, **30**, 193-207.

Cheetham, H., Millner, J. & Hardacre, A. (2006). The effect of nitrogen fertilisation on maize grain quality and yield. *Agronomy New Zealand*, **36**.

Chevallier, S., Bertrand, D., Kohler, A. & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, **20**, 221-229.

Cogdill, R.P., Hurburgh, C.R. & Rippke, G.R. (2004). Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE*, **47**, 311-320.

Costa, S. & Fiori, S. (2001). Image compression using principal component neural networks. *Image and Vision Computing*, **19**, 649-668.

Cowe, I. & McNicol, J. (1985). The use of principal components in the analysis of near-infrared spectra. *Applied Spectroscopy*, **39**, 257-266.

Cozzolino, D., Parker, M., Dambergs, R.G., Herderich, M. & Gishen, M. (2006a). Chemometrics and visible-near infrared spectroscopic monitoring of red wine fermentation in a pilot scale. *Biotechnology and Bioengineering*, **95**, 1101-1107.

Cozzolino, D., Vadell, A., Ballesteros, F., Galietta, G. & Barlocco, N. (2006b). Combining visible and near-infrared spectroscopy with chemometrics to trace muscles from an autochthonous breed of pig produced in Uruguay: a feasibility study. *Analytical and Bioanalytical Chemistry*, **385**, 931-936.

Desjardins, A.E. (2006). Selected Mycotoxigenic *Fusarium* Species. In: *Fusarium Mycotoxins: Chemistry, Genetics and Biology*. Pp. 192-194. St. Paul, Minnesota U.S.A: The American Phytopathological Society.

Desjardins, A.E., Plattner, R.D., Stessman, R.J., McCormick, S.P. & Millard, M.J. (2005). Identification and heritability of fumonisin insensitivity in *Zea mays*. *Phytochemistry*, **66**, 2474-2480.

Dombrink-Kurtzman, M.A. (1994). Examination of *opaque* mutants of maize by reversed-phase high-performance liquid chromatography and scanning electron microscopy. *Journal of Cereal Science*, **19**, 57 - 64.

Dombrink-Kurtzman, M.A. & Bietz, J.A. (1993). Zein composition in hard and soft endosperm of maize. *Cereal Chemistry*, **70**, 105-108.

Dombrink-Kurtzman, M.A. & Knutson, C.A. (1997). A study of maize endosperm hardness in relation to amylose content and susceptibility to damage. *Cereal Chemistry*, **74**, 776-780.

Duarte, A.P., Mason, S.C., Jackson, D.S. & Kiehl, J.D. (2005). Grain quality of Brazilian maize genotypes as influenced by nitrogen level. *Crop Science*, **45**, 1958-1964.

Dubois, J., Neil Lewis, E., Fry, J., Frederick S. & Calvey, E.M. (2005). Bacterial identification by near-infrared chemical imaging of food-specific cards. *Food Microbiology*, **22**, 577-583.

Dubois, J., Wolff, J.C., Warrack, J.K., Schoppelrei, J. & Lewis, E.N. (2007). NIR chemical imaging for counterfeit pharmaceutical products analysis. *Spectroscopy*, **22**, 40-50.

ElMasry, G., Wang, N., Vigneault, C., Qiao, J. & ElSayed, A. (2008). Early detection of apple bruises on different background colors using hyperspectral imaging. *LWT - Food Science and Technology*, **41**, 337-345.

Esbensen, K.H. (2006). *Multivariate Data Analysis in Practice.* Pp. 598. Oslo: CAMO Software AS.

Esbensen, K.H. & Geladi, P. (1989). Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems*, **7**, 67-86.

Eyherabide, G.H., Robutti, J.L. & Borras, F.S. (1996). Effect of near-infrared transmission-based selection on maize hardness and the composition of zeins. *Cereal Chemistry*, **73**, 775-778.

Fandohan, P., Hell, K., Marasas, W. & Wingfield, M. (2003). Infection of maize by Fusarium species and contamination with fumonisin in Africa. *African Journal of Biotechnology*, **2**, 570-579.

Foley, D.C. (1962). Systemic infection of corn by *Fusarium moniliforme. Phytopathology*, **52**, 870-872.

Fox, G.P., Onley-Watson, K. & Osman, A. (2002). Multiple linear regression calibrations for barley and malt protein based on the spectra of hordein. *Journal of the Institute of Brewing*, **108**, 155-159.

Fox, G.P., Osborne, B., Bowman, J., Kelly, A., Cakir, M., Poulsen, D., Inkerman, A. & Henry, R. (2007). Measurement of genetic and environmental variation in barley (Hordeum vulgare) grain hardness. *Journal of Cereal Science*, **46**, 82-92.

Gadag, R.N., Jha, S.K. & Singh, A. (2006). Physical characteristics of different types of maize kernels. *Maize Genetics Cooperation Newsletter*.

Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B: Atomic Spectroscopy*, **58**, 767-782.

Geladi, P., Burger, J. & Lestander, T. (2004). Hyperspectral imaging: calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems*, **72**, 209-217.

Geladi, P. & Grahn, H.F. (1996). *Multivariate Image Analysis.* Pp. 316. Chichester, West Sussex: John Wiley & Sons Ltd.

Geladi, P., Grahn, H.F. & Burger, J. (2007). Multivariate images, hyperspectral imaging: background and equipment. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P. Geladi). Pp. 1-14. Chichester, West Sussex: John Wiley & Sons Ltd.

Geladi, P., MacDougall, D. & Martens, H. (1985). Linearizaton and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*, **39**, 491-500.

Gelderblom, W.C.A., Abel, S., Smuts, C.M., Marnewick, J., Marasas, W.F.O., Lemmer, E.R. & Ramljak, D. (2001a). Fumonisin-Induced hepatocarcinogenesis: mechanisms related to cancer initiation and promotion. *Environmental Health Perspectives Supplements*, **109**, 291.

Gelderblom, W.C.A., Jaskiewicz, K., Marasas, W.F.O., Thiel, P.G., Horak, R.M., Vleggaar, R. & Kriek, N.P.J. (1988). Fumonisins- novel mycotoxins with cancer-promoting activity produced by *Fusarium moniliforme. Applied and Environmental Microbiology*, **54**, 1806-1811.

Gelderblom, W.C.A., Marasas, W.F.O., Steyn, P.S., Thiel, P.G., van der Merwe, K.J., van Rooyen, P.H., Vleggaar, R. & Wessels, P.L. (1984). Structure elucidation of Fusarin C, a mutagen produced by *Fusarium moniliforme. Journal of the Chemical Society, Chemical Communications*, **2**, 122-124.

Gelderblom, W.C.A., Rheeder, J.P., Leggott, N., Stockenstrom, S., Humphreys, J., Shephard, G.S. & Marasas, W.F.O. (2004). Fumonisin contamination of a corn sample associated with the induction of hepatocarcinogenesis in rats-role of dietary deficiencies. *Food and Chemical Toxicology*, **42**, 471-479.

Gelderblom, W.C.A., Seier, J.V., Snijman, P.W., Van Schalkwyk, D.J., Shephard, G.S. & Marasas, W.F.O. (2001b). Toxicity of culture material of *Fusarium verticillioides* strain MRC 826 to nonhuman primates. *Environmental Health Perspectives Supplements*, **109**, 267.

Gelderblom, W.C.A., Smuts, C.M., Abel, S., Snyman, S.D., Cawood, M.E., Van Der Westhuizen, L. & Swanevelder, S. (1996). Effect of fumonisin B1 on protein and lipid synthesis in primary rat hepatocytes. *Food and Chemical Toxicology*, **34**, 361-369.

Gómez, A.H., He, Y. & Pereira, A.G. (2006). Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *Journal of Food Engineering*, **77**, 313-319.

Gorsuch, R.L. (1974). *Factor Analysis.* Pp. 370. Philadelphia: W.B. Saunders Company.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J. & Bell, S.E.J. (2008a). Recent applications of chemical imaging to pharmaceutical process monitoring and quality control. *European Journal of Pharmaceutics and Biopharmaceutics*, **69**, 10-22.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, **18**, 590-598.

Gowen, A.A., O'Donnell, C.P., Taghizadeh, M., Cullen, P.J., Frias, J.M. & Downey, G. (2008b). Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (*Agaricus bisporu*s). *Journal of Chemometrics*, **22**, 259-267.

Grahn, H.F. & Geladi, P. (2007). Techniques and Applications of Hyperspectral Image Analysis. Pp. 368. John Wiley & Sons, Ltd.

Gupta, H.O., Sachdev, A., Johari, R.P. & Singh, R.P. (2003). Storage protein and its impact on nutritional quality in quality protein maize. *Indian Journal of Agricultural Biochemistry*, **16**.

Haddad, Y., Benet, J.C. & Abecassis, J. (1998). A rapid general method for appraising the rheological properties of the starchy endosperm of cereal grains. *Cereal Chemistry*, **75**, 673-676.

Han, D., Tu, R., Lu, C., Liu, X. & Wen, Z. (2006). Nondestructive detection of brown core in the Chinese pear (Yali) by transmission visible-NIR spectroscopy. *Food Control*, **17**, 604-608.

Headrick, J.M., Pataky, J.K. & Juvik, J.A. (1990). Relationships among carbohydrate content of kernels, condition of silks after pollination, and the response of sweet corn inbred lines to infection of kernels by *Fusarium moniliforme*. *Phytopathology*, **80**, 487-494.

Herschel, W. (1800). Investigation of the powers of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers. *Philosophical Transactions of the Royal Society of London*, **90**, 255-283.

Hesseltine, C.W. & Shotwell, O. (1973). New methods for rapid detection of aflatoxin. *Pure and Applied Chemistry*, **35**, 259-266.

Holding, D.R. & Larkins, B.A. (2006). The development and importance of zein protein bodies in maize endosperm. *Maydica*, **51**, 243-254.

Hoseney, R.C. (1994). *Principles of Cereal Acience and Technology.* Pp. 378. St. Paul, Minnesota: American Association of Cereal Chemists, Inc.

Ioerger, B., Bean, S.R., Tuinstra, A.R., Pedersen, J.F., Erpelding, J., Lee, K.A. & Herrman, T.J. (2007). Characterization of polymeric proteins from vitreous and floury sorghum endosperm. *Journal of Agricultural and Food Chemistry*, **55**, 10232-10239.

Isaacson, C. (2005). The change of the staple diet of black South Africans from sorghum to maize (corn) is the cause of the epidemic of squamous carcinoma of the oesophagus. *Medical Hypotheses*, **64**, 658-660.

Isaksson, T. & Næs, T. (1998). The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Applied Spectroscopy*, **42**, 1273-1284

Jackson, L. & Jablonski, J. (2004). Fumonisins. In: *Mycotoxins in Food: Detection and Control* (edited by N. Magan & M. Olsen). Pp. 367-405. Cambridge, England: Woodhead Publishing, Ltd.

Jennings, S.D., Myers, D.J., Johnson, L.A. & Pollak, L.M. (2002). Effects of maturity on grain quality and wet-milling properties of two selected corn hybrids. *Cereal Chemistry*, **79**, 697-702.

Jolliffe, I.T. (1986). *Principal Component Analysis.* Pp. 271. New York: Springer-Verlag.

Juhász, R., Gergely, S., Szabóki, Á. & Salgó, A. (2007). Correlation between NIR spectra and RVA parameters during germination of maize. *Cereal Chemistry*, **84**, 97.

Kim, T.H., Hampton, J.G., Opara, L.U., Hardacre, A.K. & Mackay, B.R. (2002). Effects of maize grain size, shape and hardness on drying rate and the occurrence of stress cracks. *Journal of the Science of Food and Agriculture*, **82**, 1232-1239.

Koehler, F.W., Lee, E., Kidder, L.H. & Lewis, E.N. (2002). Near infrared spectrsocopy: the practical chemical imaging solution. *Spectoscopy Europe*, **14**, 12-19.

Lammertyn, J., Peirs, A., De Baerdemaeker, J. & Nicolaï, B. (2000). Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biology and Technology*, **18**, 121-132.

Landau, S., Glasser, T. & Dvash, L. (2006). Monitoring nutrition in small ruminants with the aid of near infrared reflectance spectroscopy (NIRS) technology: A review. *Small Ruminant Research*, **61**, 1-11.

Larkins, B.A., Pedersen, K., Marks, M.D. & Wilson, D.R. (1984). The zein proteins of maize endosperm. *Trends in Biochemical Science*, **9**, 306-.

Lee, K.M., Bean, S.R., Alavi, S., Herrman, T.J. & Waniska, R.D. (2006). Physical and biochemical properties of maize hardness and extrudates of selected hybrids. *Journal of Agricultural and Food Chemistry*, **54**, 4260-4269.

Lee, K.M., Herrman, T.J., Bean, S.R., Jackson, D.S. & Lingenfelser, J. (2007a). Classification of dry-milled maize grit yield groups using quadratic discriminant analysis and decision tree algorithm. *Cereal Chemistry*, **84**, 152-161.

Lee, K.M., Herrman, T.J., Lingenfelser, J. & Jackson, D.S. (2005). Classification and prediction of maize hardness-associated properties using multivariate statistical analyses. *Journal of Cereal Science*, **41**, 85-93.

Lee, K.M., Herrman, T.J., Rooney, L., Jackson, D.S., Lingenfelser, J., Rausch, K.D., McKinney, J., Iiams, C., Byrum, L., Hurburgh, C.R., Johnson, L.A. & Fox, S.R. (2007b). Corroborative study on maize quality, dry-milling and wet-milling properties of selected maize hybrids. *Journal of Agricultural and Food Chemistry*, **55**, 10751-10763.

Lovett, D.K., Deaville, E.R., Givens, D.I., Finlay, M. & Owen, E. (2005). Near infrared reflectance spectroscopy (NIRS) to predict biological parameters of maize silage: effects of particle comminution, oven drying temperature and the presence of residual moisture. *Animal Feed Science and Technology*, **120**, 323-332.

Lovett, D.K., Deaville, E.R., Mould, F., Givens, D.I. & Owen, E. (2004). Using near infrared reflectance spectroscopy (NIRS) to predict the biological parameters of maize silage. *Animal Feed Science and Technology*, **115**, 179-187.

Ludwig-Müller, J., Georgiev, M. & Bley, T. (2008). Metabolite and hormonal status of hairy root cultures of Devil's claw (*Harpagophytum procumbens*) in flasks and in a bubble column bioreactor. *Process Biochemistry*, **43**, 15-23.

Malvar, R.A., Revilla, P., Moreno-Gonzalez, J., Butron, A., Sotelo, J. & Ordas, A. (2008). White maize: genetics of quality and agronomic performance. *Crop Science*, **48**, 1373-1381.

Marasas, W.F.O. (1995). Fumonisins: Their implications for human and animal health. *Natural Toxins*, **3**, 193-198.

Marasas, W.F.O. (2001). Discovery and occurence of the Fumonisins: a historical perspective. *Environmental Health Perspectives*, **109**, 239-243.

Martinsen, P., Schaare, P. & Andrews, M. (1999). A versatile near infrared imaging spectrometer. *Journal of Near Infrared Spectroscopy*, **7**, 17-25.

Masoero, G., Sala, G. & Peiretti, P.G. (2007). Development of near infrared (NIR) spectroscopy and electronic nose (EN) techniques to analyse the conservation quality of farm silages. *Journal of Food Agriculture & Environment*, **5**, 172-177.

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. & Kaufman, L. (1988). *Chemometrics: A Textbook.* Amsterdam: Elsevier Science Publishers.

McGlone, V.A., Jordan, R.B. & Martinsen, P.J. (2002). Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biology and Technology*, **25**, 135-144.

Mestres, C. & Matencio, F. (1996). Biochemical basis of kernel milling characteristics and endosperm vitreousness of maize. *Journal of Cereal Science*, **24**, 283-290.

Milasinovic, M., Radosavljevic, M., Dokic, L. & Jakovljevic, J. (2007). Wet-milling properties of ZP maize hybrids. *Maydica*, **52**, 289-292.

Munkvold, G.P. & Desjardins, A.E. (1997). Fumonisins in maize: can we reduce their occurrence? *Plant Disease*, **81**, 556-565.

Munkvold, G.P., Hellmich, R.L. & Showers, W.B. (1997). Reduced fusarium ear rot and symptomless infection in kernels of maize genetically engineered for european corn borer resistance. *Phytopathology*, **87**, 1071-1077.

Musser, S.M. & Plattner, R.D. (1997). Fumonisin composition in cultures of *Fusarium moniliforme*, *Fusarium proliferatum*, and *Fusarium nygami. Journal of Agricultural and Food Chemistry*, **45**, 1169-1173.

Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). *A User-friendly Guide to Multivariate Calibration and Classification.* Pp. 114-125. Chichester, UK: NIR Publications.

Narvaez-Gonzalez, E.D., Cardenas, J.D.F., Taba, S., Tostado, E.C. & Peniche, R.A.M. (2007). Effect of starch granule size on the thermal and pasting properties of maize. *Revista Fitotecnia Mexicana*, **30**, 269-277.

Narvaez-Gonzalez, E.D., Figueroa-Cardenas, J.D., Taba, S., Tostado, E.C., Peniche, R.A.M. & Sanchez, F.R. (2006). Relationships between the microstructure, physical features, and

chemical composition of different maize accessions from Latin America. *Cereal Chemistry*, **83**, 595-604.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I. & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology*, **46**, 99-118.

Norris, K.H. (1962). Instrumetation of infrared radiation. *Transaction of the ASAE*, **5**, 17-20.

Norris, K.H. (1964). Simple spectroradiometer for 0.4 to 1.2-micron region. *Transactions of the ASAE*, **7**, 240-242.

Oren, L., Ezrati, S., Cohen, D. & Sharon, A. (2003). Early events in the *Fusarium verticillioides*-maize interaction characterized by using a green fluorescent protein-expressing transgenic isolate. *Applied and Environmental Microbiology*, **69**, 1695-1701.

Ortega, E.I. & Bates, L.S. (1983). Biochemical and agronomic studies of two modified hard-endosperm *opaque*-2 maize (*Zea mays* L.) populations. *Cereal Chemistry*, **60**, 107-111.

Ortiz-Somovilla, V., España-España, F., Gaitán-Jurado, A.J., Pérez-Aparicio, J. & De Pedro-Sanz, E.J. (2007). Proximate analysis of homogenized and minced mass of pork sausages by NIRS. *Food Chemistry*, **101**, 1031-1040.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993a). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis.* Pp. 114-117. Essex, England: Longman Scientific & Technical.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993b). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis.* Pp. 29-33. Essex, England: Longman Scientific & Technical.

Pascale, M., Visconti, A. & Chelkowski, J. (2002). Ear rot susceptibility and mycotoxin contamination of maize hybrids inoculated with Fusarium species under field conditions. *European Journal of Plant Pathology*, **108**, 645-651.

Pearson, T.C. & Wicklow, D.T. (2006). Detection of corn kernels infected by fungi. *Transactions of the ASABE*, **49**, 1235-1245.

Pearson, T.C., Wicklow, D.T., Maghirang, E.B., Xie, F. & Dowell, F.E. (2001). Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. *Transactions of the ASAE*, **44**, 1247-1254.

Pearson, T.C., Wicklow, D.T. & Pasikatan, M.C. (2004). Reduction of aflatoxin and fumonisin contamination in yellow corn by high-speed dual-wavelength sorting. *Cereal Chemistry*, **81**, 490-498.

Pizarro, C., Esteban-Díez, I., Nistal, A.N. & González-Sáiz, J.M. (2004). Influence of data pre-processing and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, **509**, 217-227.

Pomeranz, Y. (1986). Comparison of screening methods for indirect determination of sorghum hardness. *Cereal Chemistry*, **63**, 36-38.

Pomeranz, Y., Czuchajowska, Z. & Lai, F.S. (1986). Comparison of methods for determination of hardness and breakage susceptibility of commercially dried corn. *Cereal Chemistry*, **63**, 39-43.
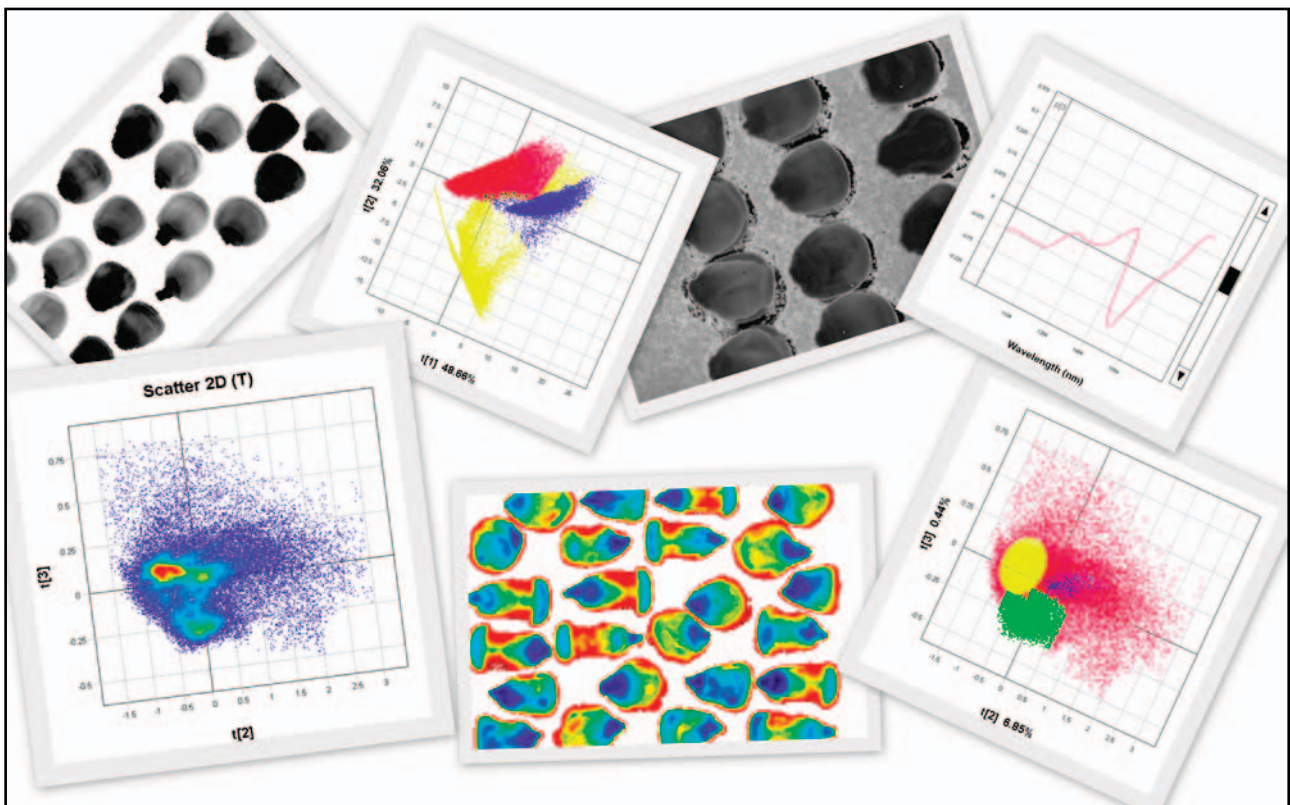
Pomeranz, Y., Czuchajowska, Z., Martin, C. & Lai, F. (1985a). Determination of maize hardness by Stenvert hardness tester. *Cereal Chemistry*, **62**, 108-112.

Pomeranz, Y., Czuchajowska, Z., Martin, C.R. & Lai, F.S. (1985b). Determination of corn hardness by the Stenvert hardness tester. *Cereal Chemistry*, **62**, 108-112.

Pomeranz, Y., Martin, C.R., Traylor, D.D. & Lai, F.S. (1984). Corn hardness determination. *Cereal Chemistry*, **61**, 147-150.

Pratt, R.C., Paulis, J.W., Miller, K., Nelsen, T. & Bietz, J.A. (1995). Association of zein classes with maize kernel hardness. *Cereal Chemistry*, **72**, 162-167.

Reich, G. (2005). Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews Non-Invasive Spectroscopic and Imaging Techniques in Drug Delivery*, **57**, 1109-1143.

Robutti, J.L. (1995). Maize kernel hardness estimation in breeding by Near-infrared transmission analysis. *Cereal Chemistry*, **72**, 632-636.

Robutti, J.L., Borras, F.S. & Eyherabide, G.H. (1997). Zein compositions of mechanically separated coarse and fine portions of maize kernels. *Cereal Chemistry*, **74**, 75-78.

Robutti, J.L., Borras, F.S., Ferrer, M.E. & Bietz, J.A. (2000). Grouping and identification of Argentine maize races by principal component analysis of zein reversed-phase HPLC data. *Cereal Chemistry*, **77**, 91-95.

Rodionova, O.Y., Houmøller, L.P., Pomerantsev, A.L., Geladi, P., Burger, J., Dorofeyev, V.L. & Arzamastsev, A.P. (2005). NIR spectrometry for counterfeit drug detection: a feasibility study. *Analytica Chimica Acta*, **549**, 151-158.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A. & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, **44**, 683-700.

Sakudo, A., Suganuma, Y., Kobayashi, T., Onodera, T. & Ikuta, K. (2006). Near-infrared spectroscopy: promising diagnostic tool for viral infections. *Biochemical and Biophysical Research Communications*, **341**, 279-284.

Samson, R.A. (2000). *Introduction to food- and airborne fungi.* Utrecht: Centraalbureau voor Schimmelcultures.

Sanchez, F.C., Salinas, M.Y., Vazquez, C.M.G. & Aguilar, G.N. (2007). Effect of the prolamins in maize (Zea mays L.) grain on tortilla texture. *Archivos Latinoamericanos De Nutricion*, **57**, 295-301.

Schierbaum, F. (2007). World corn production, uses and ending stocks: recorded figures for fiscal years 2004/05 and 2005/06. *Starch*, **59**, 192-195.

Schulthess, F., Cardwell, K. & Gounou, S. (2002). The effect of endophytic *Fusarium verticillioides* on infestation of two maize varieties by Lepidopterous Stemborers and Coleopteran grain feeders. *Phytopathology*, **92**, 120-128.

Schulz, B. & Boyle, C. (2005). The endophytic continuum. *Mycological Research*, **109**, 661-686.

Siesler, H.W. (2008). Basic principles of near-infrared spectroscopy. In: *Handbook of Near-infrared Analysis* (edited by D.A. Burns & E.W. Ciurczak). Pp. 7-19. Boca Raton: CRC Press, Talyor & Francis Group.

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

Siska, J. & Hurburgh, C.R. (1994). Prediction of wisconsin tester breakage susceptibility of corn from bulk density and NIRS measurments of composition. *Transactions of the ASAE*, **37**, 1577-1582.

Siska, J. & Hurburgh, C.R. (1995). Corn density measurement by near-infrared transmittance. *Transactions of the ASAE*, **38**, 1821-1824.

Siuda, R., Balcerowska, G. & Sadowski, C. (2006). Comparison of the usability of different spectral ranges within the near ultraviolet, visible and near infrared ranges (UV-VIS-NIR) region for the determination of the content of scab-damaged component in blended samples of ground wheat. *Food Additives and Contaminants*, **23**, 1201-1207.

Stewart, K.M. & Cole, D. (2005). The commercial harves of devil's claw (*Harpagophytum* spp.) in southern Africa: The devil's in the details. *Journal of Ethnopharmacology*, **100**, 225-236.

Szabo, L., Pinter, L. & Alfoldi, Z. (1991). The NIR technique as a fast and cheap tool for improving silage maize breeding and evaluation. *Novenytermeles*, **40**, 135-140.

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

Thompson, D.L. & Goodman, M.M. (2006). Increasing kernel density for two inbred lines of maize. *Crop Science*, **46**, 2179-2182.

Tisdall, M.M., Tachtsidis, I., Leung, T.S., Elwell, C.E. & Smith, M. (2007). Near-infrared spectroscopic quantification of changes in the concentration of oxidized cytochrome c oxidase in the healthy human brain during hypoxemia. *J Biomed Opt*, **12**, 024002.

Uddin, M., Okazaki, E., Ahmad, M.U., Fukuda, Y. & Tanaka, M. (2006). NIR spectroscopy: A non-destructive fast technique to verify heat treatment of fish-meat gel. *Food Control*, **17**, 660-664.

Watson, S.A. (1987a). Measurement and maintenance of quality. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 125-183. St Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Watson, S.A. (1987b). Structure and composition. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 53-82. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Watt, J.M. & Breyer-Brandwijk, M.G. (1962). *The Medicinal and Poisonous Plants of Southern and Eastern Africa.* Pp. 830. Edinburgh and London: E. & S. Livingstone LTD.

Wehling, R.L., Jackson, D.S. & Hamaker, B.R. (1996). Prediction of corn dry-milling quality by near-infrared spectroscopy. *Cereal Chemistry*, **73**, 543-546.

Williams, P. (2007). Near-infrared Technology - getting the best out of light. In: A short course in the practical implementation of near-infrared spectroscopy for the user. Nanaimo, Canada: PDK Projects, Inc.

Wold, S., Sjöström, M. & Eriksson, L. (2001a). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **58**, 109-130.

Wold, S., Trygg, J., Berglund, A. & Antti, H. (2001b). Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems*, **58**, 131-150.

Wolf, M.J., Buzan, C.L., MacMasters, M.M. & Rist, C.E. (1952). Structure of the mature corn kernel. I. Gross anatomy and structural relationships. *Cereal Chemistry*, **29**, 321-333.

Workman, J. (1993). A brief review of the near infrared measurement technique. *NIR News*, **4**, 8-16.

Wu, Y.V. (1992). Corn hardness as related to yield and particle size of fractions from a micro hammer-cutter mill. *Cereal Chemistry*, **69**, 343-347.

Yang, Y., Li, X. & Xie, E. (2005). Correlation between ultra-microstructure and nutrition quality in opaque-2 corn kernels with different hardness endosperm. *Scientia Agricultura Sinica*, **38**.

Yang, Z., Ren, H.Q. & Jiang, Z.H. (2008). Discrimination of wood biological decay by NIR and partial least squares discriminant analysis (PLS-DA). *Spectroscopy and Spectral Analysis*, **28**, 793-796.

Yao, H., Hruska, Z., Kincaid, R., Brown, R. & Cleveland, T. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215-224.

Yates, I., Arnold, J., Hinton, D., Basinger, W. & Walcott, R. (2003). *Fusarium verticillioides* induction of maize seed rot and its control. *Canadian Journal of Botany*, **81**, 422-428.

Yates, I.E. & Sparks, D. (2008). *Fusarium verticillioides* dissemination among maize ears of field-grown plants. *Crop Protection*, **27**, 606-613.

Zhang, H., Paliwal, J., Jayas, D.S. & White, N.D.G. (2007). Classification of fungal infected wheat kernels using near-infrared reflectance hyperspectral imaging and support vector machine. *Transactions of the ASABE*, **50**, 1779-1785.

Zhang, Y., Darlington, H., Jones, H.D., Halford, N.G., Napier, J.A., Davey, M.R., Lazzeri, P.A. & Shewry, P.R. (2003). Expression of the gamma-zein protein of maize in seeds of transgenic barley: effects on grain composition and properties. *Theoretical and Applied Genetics*, **106**, 1139-1146.

Zhao, H.H. & Yan, Y.L. (2006). The effects of noise on NIR analysis and related mathematic pretreatments and models. *Spectroscopy and Spectral Analysis*, **26**, 842-845.

Chapter 3

Determination of maize kernel hardness by near infrared (NIR) hyperspectral imaging and hyperspectral image analysis

Chapter 3

Determination of maize kernel hardness by near infrared (NIR) hyperspectral imaging and hyperspectral image analysis

Abstract

The use of near infrared (NIR) hyperspectral imaging and hyperspectral image analysis for distinguishing between hard, intermediate and soft maize kernels from inbred lines was evaluated. Near infrared hyperspectral images of whole maize kernels were acquired using a Spectral Dimensions MatrixNIR camera with a spectral range of 960-1662 nm as well as a sisuChema SWIR (short wave infrared) hyperspectral pushbroom imaging system with a spectral range of 1000-2498 nm. Exploratory principal component analysis (PCA) on absorbance images was used to remove background, bad pixels and shading. On the cleaned images, PCA could be used effectively to find histological classes including glassy and floury endosperm. PCA illustrated a distinct difference between floury (soft) and glassy (hard) endosperm along principal component (PC) three with two distinguishable clusters. Interpreting the PC loading line plots important absorbance peaks responsible for the variation were 1215, 1395 and 1450 nm for both MatrixNIR images (12 and 24 kernels). The loading line plots for the sisuChema illustrated peaks of importance at 1695, 1900 and 1940 nm. Subsequently partial least squares-discriminant analysis (PLS-DA) was applied. The MatrixNIR-12 kernel image with a PLS-DA model exhibited a classification rate of up to 99% for the discrimination of both floury and glassy endosperm. This was repeated on the MatrixNIR-24 kernel image which yielded a classification rate of 82% for the discrimination of glassy endosperm and 73% for floury endosperm. The sisuChema image yielded a classification rate of 95% for the discrimination of floury endosperm and 92% for glassy endosperm.

Introduction

Maize kernel hardness is principally a genetic expression, but environment and postharvest handling (e.g. transportation, drying and storage) also have an influence on hardness properties (Watson, 1987a). Hardness is related to kernel density, bulk density, and attack by storage insects, breakage susceptibility caused by drying, storage, handling, and processing. Maize is different from wheat in that both glassy and floury endosperm are found within a single kernel in a particular ratio (Watson, 1987a; Watson, 1987b; Hoseney, 1994). It is this ratio that determines whether the kernel is hard, soft or intermediate (Watson, 1987b). Hard kernels have predominantly glassy endosperm; soft

kernels consist principally of floury endosperm while intermediate kernels possess approximately equal quantities of both. The glassy endosperm is tightly compacted with few or no air spaces. The starch granules are held together by a protein matrix and protein bodies (zein) are found on the starch granules (Hoseney, 1994; Lee *et al.*, 2006). The floury endosperm on the other hand comprises spherical starch granules that are covered with a protein matrix without zein bodies.

Maize hardness is important to producers and processors in the grain trade (Pomeranz *et al.*, 1984; Watson, 1987a; Lee *et al.*, 2005) since it greatly influences end-use processing performance, including dry milling yield and power requirements; dust formation during processing; and processing of maize grits into certain foods.

Maize hardness has been assessed to date using near infrared (NIR) spectroscopy (Pomeranz *et al.*, 1984; Eyherabide *et al.*, 1996; Wehling *et al.*, 1996); density (Pomeranz *et al.*, 1984; Almeida-Dominguez *et al.*, 1997); particle size index (PSI) or particle size analysis (PSA) (Pomeranz *et al.,* 1984); the Stenvert hardness tester (Pomeranz *et al.,* 1985); the tangential abrasive dehulling device (TADD) (Wehling *et al.*, 1996; Almeida-Dominguez *et al.*, 1997); and the rapid visco analyser (RVA) (Almeida-Dominguez *et al.,* 1997). Most of these methods require destruction of the sample.

Using some of the aforementioned techniques together with NIR spectroscopy and multivariate data analyses, predictions of maize dry-milling quality indicated that NIR spectroscopy can predict dry-milling characteristics with a reliability suitable for rough screening (Wehling *et al.,* 1996). Additionally, wet-milling starch yield with a bias-corrected standard error of prediction of 1.41% was found using NIR spectroscopy (Wehling *et al.,* 1993). In another study the classification and prediction of maize hardness properties into hardness clusters using discriminant analysis showed a classification rate of 87% (Lee *et al.,* 2005). The problem with selecting the most appropriate method to determine maize hardness, either as a reference method for NIR calibration development or as a standalone method, remains unresolved. The selected method should satisfy both maize breeders and industry.

NIR hyperspectral imaging is a powerful spectroscopic technique that is capable of capturing images at many wavelengths in the NIR region (Peirs *et al.*, 2003; Cogdill *et al.*, 2004; Geladi *et al.*, 2004; Tatzer *et al.*, 2005; Burger & Geladi, 2006; Tallada *et al.*, 2006) and has been used in a number of applications in wheat. However, the technique has not yet been evaluated as a technique for determining maize hardness.

A NIR hyperspectral imaging system collects image data by arranging it into a three-way data matrix, known as a hypercube (Cogdill *et al.*, 2004; Burger & Geladi, 2006). The

first two axes (x and y) of the matrix are the vertical and horizontal pixel coordinates while the third (z) axis depicts the spectral dimension. Hundreds of single channel black and white (grayscale) images are stacked on top of each other to produce hyperspectral images (or hypercubes) (Burger & Geladi, 2006). Each of these grayscale images represents a single band of spectral wavelength. Typically, a commercial NIR imaging instrument produces hypercubes with dimensions 256×320×118, *i.e.* 118 single channel images each with 256×320 pixels.

Despite posing data extraction challenges, this immense accumulation of data creates novel possibilities. Near infrared hyperspectral imaging has been evaluated for the classification of sound and stained wheat grains with an overall accuracy of 95% over the 420-2500 nm range, as well as for the reduced ranges of 420–1000 nm and 420–700 nm (Berman *et al.*, 2007). In another study it was shown that NIR hyperspectral imaging reveals sprouting for simultaneous single wheat kernels with sensitivity greater than that of the human eye (Smail *et al.*, 2006). The discrimination between vitreous and non-vitreous wheat kernels with a classification rate of 94% has been achieved using NIR hyperspectral imaging with a wavelength range of 650–1100 nm (Gorretta *et al.*, 2006) and single kernel maize analysis by hyperspectral transmittance in the range of 750–1090 nm was done for the development of predictive calibrations for moisture (SECV 1.20% and RPD 2.74) and oil (SECV 1.38% and RPD 1.45) content (Cogdill *et al.*, 2004).

Multivariate image analysis (Geladi & Grahn, 1996) or, lately, hyperspectral image analysis (Geladi *et al.*, 2007) is helpful for image data exploration, classification and quantification. As with single-point NIR spectroscopy, chemometric techniques are applied to the data (image data set in this case) to decompose the data, pre-process the data and perform regression or classification analyses. Principal component analysis (PCA) was evaluated for the investigation of visible and NIR multispectral images of works of art (Baronti *et al.*, 1997). In a recent study NIR hyperspectral imaging combined with (PCA) has been evaluated for the analysis for bruise damage detection on white mushrooms (Gowen *et al.*, 2008). Partial least squares-discriminant analysis (PLS-DA) has been assessed in multivariate image analysis to discriminate food products of different natures with classification rates of 75%, 83%, 98% and 89% for maize, pea, soya bean meal and wheat respectively (Chevallier *et al.*, 2006).

The aim of this study was to determine whether NIR hyperspectral imaging could distinguish between whole maize kernels from inbred lines of varying hardness by evaluating:

- the use of PCA and the interpretation of principal component (PC) loading line plots to explain the chemical variation within and between the maize kernels of different hardness; and
- partial least squares-discriminant analysis (PLS-DA) as a possible chemometric classification technique.

Materials and methods

Samples

Maize kernels of varying degrees of hardness were kindly provided by Pioneer Hi-Bred Research RSA (P.O. Box 699 Delmas, 2210 South Africa). These whole kernels were selected randomly from yellow maize inbred lines used in breeding trials. The yellow maize consisted of three categories of hardness, i.e. hard, intermediate and soft. These descriptions of the hardness categories were provided by experienced maize breeders.

Near infrared hyperspectral imaging systems

Near infrared hyperspectral images were acquired using a Spectral Dimensions MatrixNIR camera (Malvern Instruments Ltd, Malvern, Worcestershire, UK) as well as a sisuChema SWIR (short wave infrared) hyperspectral imaging system (Specim, Spectral Imaging Ltd, Oulu, Finland).

The Spectral Dimensions MatrixNIR camera comprised of an Indium Gallium Arsenide (InGaAs) diode array detector with a liquid crystal tunable filter (LCTF) bandpass filter enabling wavelength selection. Four current controlled quartz-halogen lamps provided illumination (Gilway 12V 75W). The image field of view was 50 mm x 62 mm. The spectral range recorded was 960-1662 nm at increments of 6 nm, producing hyperspectral images with 256 × 320 pixels at 118 wavelength channels.

The sisuChema comprised an imaging spectrograph coupled to a 2-D array Mercury-cadmium-telluride (HgCdTe) detector. Individual images were acquired with a spectral range of 1000-2498 nm with 6-7 nm resolution. When stacked, the individual images created a 301× 220 × 239 hypercube.
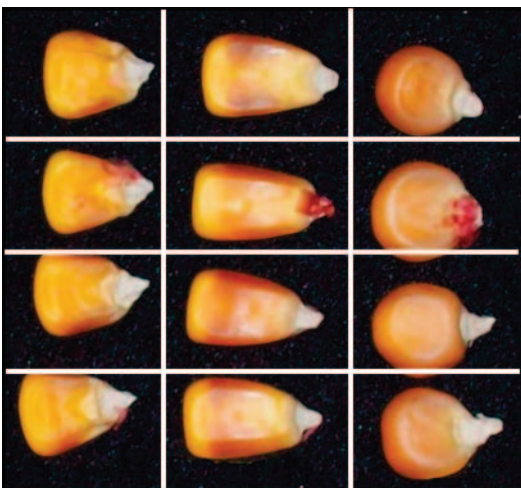
Image acquisition

Twelve yellow maize kernels, four kernels randomly selected from each of the three hardness categories (hard, intermediate and soft) were positioned on a 70 mm x 70 mm silicon carbide (SiC) sandpaper as illustrated in **Fig. 3.1**. Digital images were acquired with a Kodak EasyShare, model C713 (Eastman Kodak Company, Rochester, NY 14650).

These samples were presented on the sample holder of the Spectral Dimensions MatrixNIR camera for image acquisition. In addition twenty four yellow maize kernels, eight kernels randomly selected from each hardness category were positioned on a 70 mm x 70 mm SiC sandpaper according to a Latin Square design as depicted in **Fig. 3.2**. Duplicate kernels of each category were within each block within the grid. These samples were presented on the sample holders of both the Spectral Dimensions MatrixNIR camera and the sisuChema for image acquisition. Digital images were acquired with a Leica Digilux 3 (Leica Camera AG, Solms, Germany).

The image acquisition time on the Spectral Dimensions MatrixNIR camera was *ca*. 8 minutes 128 ms integration time. This includes the acquiring of the images of reflectance standards (Labsphere, USA) necessary for image calibration and to correct for pixel-to-pixel variances due to camera inconsistencies and variation in sample illumination (Burger & Geladi, 2005). Each channel was scanned 16 times and the samples were kept cool with a small box fan positioned just beyond the field of view. Standards of 2% (dark reference) and 75% (white reference) reflectance were used. The image acquisition time on the sisuChema was *ca*. 8 seconds with an exposure time of 1.5 ms and a frame rate of 100 Hz. Both internal dark and white reference standards were imaged immediately before image acquisition of the sample. Digital images were acquired with a Leica Digilux 3 (Leica Camera AG, Solms, Germany).

(a)   (b)

Figure 3.1 (a) Digital image illustrating sample presentation of the 12 yellow maize kernels for image acquisition with the Spectral Dimensions MatrixNIR camera and (b) grid illustrating the hardness class of each maize kernel (H = hard; I = intermediate; S = soft).
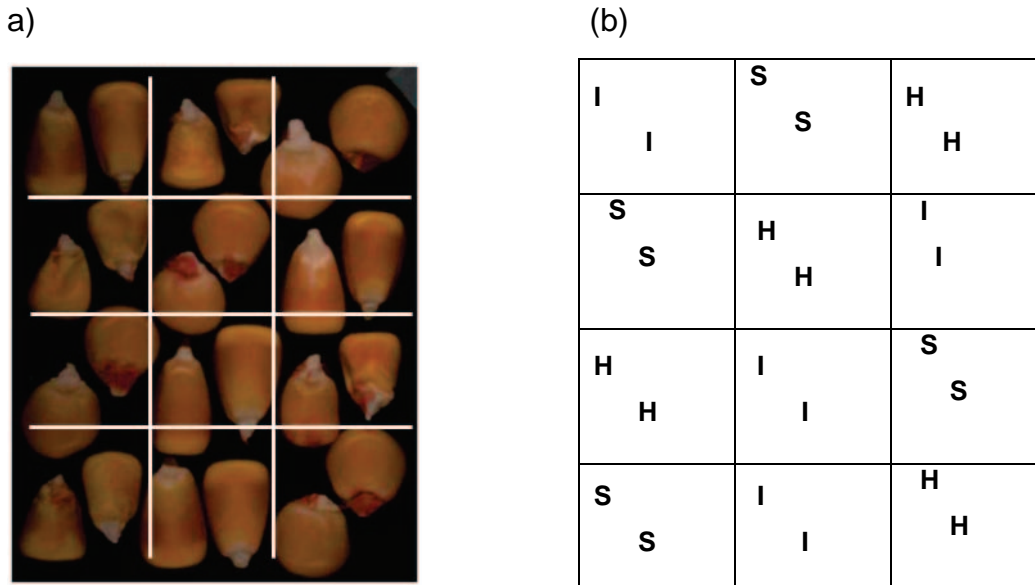
a)   (b)

I I	S S	H H
S S	H H	I I
H H	I I	S S
S S	I I	H H

Figure 3.2 (a) Digital image illustrating sample presentation of the 24 yellow maize kernels for image acquisition with the Spectral Dimensions MatrixNIR camera as well as the sisuChema SWIR imaging system and (b) grid illustrating the hardness class of each maize kernel (H = hard; I = intermediate; S = soft).

Hyperspectral image analysis (HIA)

Image correction

Sample images (I_0) acquired with the Spectral Dimensions MatrixNIR camera were converted from instrument measurement values, i.e. A/D (analogue to digital) counts to absorbance values according to Equation 1. The transformation involved correcting spectra for detector dark current (dark reference subtracted from sample spectra) and dividing by a similarly corrected total reflectance spectrum; 75% in this case (dark reference subtracted from white reference). Image acquisition was realised using MatrixAcquire software (Malvern Instruments Ltd, Malvern, Worcestershire, UK) while transformation to absorbance was done using ISys 4.0 (Malvern Instruments Ltd, Malvern, Worcestershire, UK). In addition, absorbance images were converted to Matlab files and exported to Evince 2.020 (UmBio AB, Umeå, Sweden) for further image analysis.

$$I_{\lambda,n} = -\log[(\frac{S_{\lambda,n} - B_{\lambda,n}}{W_{\lambda,n} - B_{\lambda,n}}) \times 0.75]$$

… (Equation 1)

Where

    n = pixel index variable ($n = 1...N$)

    $I_{\lambda,n}$ = standardised absorbance intensity, pixel n, at wavelength λ

    $S_{\lambda,n}$ = sample image, pixel n, at wavelength λ

$B_{\lambda,n}$ = dark reference image, pixel *n*, at wavelength λ

$W_{\lambda,n}$ = white reference image, pixel *n*, at wavelength λ

0.75   = total reflectance

The images acquired with the sisuChema were automatically converted to pseudo-absorbance in the Evince 2.020 hyperspectral image analysis software package (UmBio AB, Umeå, Sweden) using the same principles as above.

<u>Image cleaning</u>

Absorbance images were subjected to PCA. PCA score plots as well as the PCA score images were used to identify and classify outliers, bad pixels, illumination errors and shading errors, edge effects and background. After the unwanted classes were removed, PCA was recalculated with additional components obtaining a cleaned image. The cleaned image was used in subsequent analysis.

<u>Image analysis of cleaned image</u>

PCA score plots and the accompanied score images were used to identify and classify clusters to the various hardness classes (glassy and floury). These were then projected onto PCA score images.

<u>Partial least squares-discriminant analysis (PLS-DA)</u>

Each object in the two classes (glassy and floury endosperm) was assigned to a dummy variable, zero or one, *i.e.* floury endosperm or glassy endosperm. The images of the second sample set were divided into a training set and a test set by assigning two thirds of the image to the training set and the remainder to the test set. The pedicle of the maize kernels was excluded before performing PLS-DA.

Results and discussion

Hyperspectral image analysis – MatrixNIR (12 kernels)

<u>Image correction</u>

Figure 3.3a shows the raw image of the 12 maize kernels acquired on the MatrixNIR at 1310 nm. The bright zones are areas where there was high signal (instrument count or reflectance). Thus at this wavelength there was a very high signal near the pedicle of the kernels. The other parts of the kernel have a lower (darker) signal, and therefore less reflectance (greater absorbance). In addition, stripes of contrasting colour across the

entire image were found which also negatively affected the spectra as illustrated in **Fig. 3.3b**. Correcting for these inconsistencies by means of imaging system calibration was therefore imperative.

Calibration of the system compensates for differences in wavelength intensity values in the different physical components of the camera system (Geladi *et al.*, 2004). Every camera system behaves differently for different wavelengths of light (Geladi *et al.*, 2004) and there are numerous factors that impact on this. The camera sensors are more sensitive to some wavelengths of light than others; the lenses and other optical components act as filters and 'block' or absorb some of the light with different effects at different wavelengths; and the light sources do not illuminate the sample with the same amount of light at different wavelengths.

There were also spatial variances which must be measured and this too was the result of the different physical system components: the lamps may not be aimed accurately, or even under best conditions they do not illuminate with the same intensity at all positions on the sample; the throughput of light through the lenses and other optical components may not be the same at all spatial locations in the image; and the individual pixel sensors may not behave exactly the same across the camera sensor. **Figure 3.4a** shows the absorbance-corrected image without the striations and bright zones on the kernels, illustrating the efficacy of the dark/white reference calibration and **Fig. 3.4b** depicts the more consistent absorbance spectra.
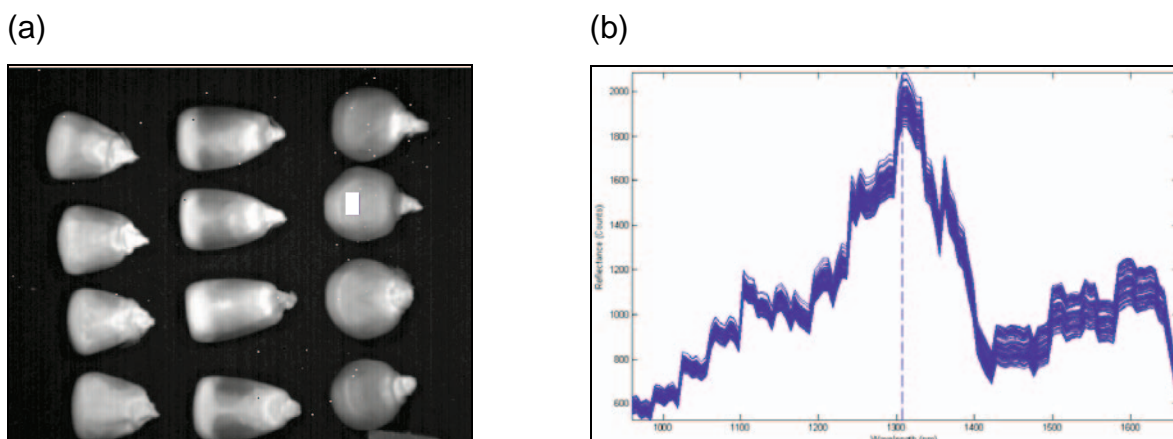
(a)   (b)

Figure 3.3 (a) Raw NIR hyperspectral image at 1310 nm with striations and bright zones indicative of an image in detector counts and (b) spectra of the selected region, indicated by white square in (a), illustrating high detector counts and wavelength intensity inconsistencies.
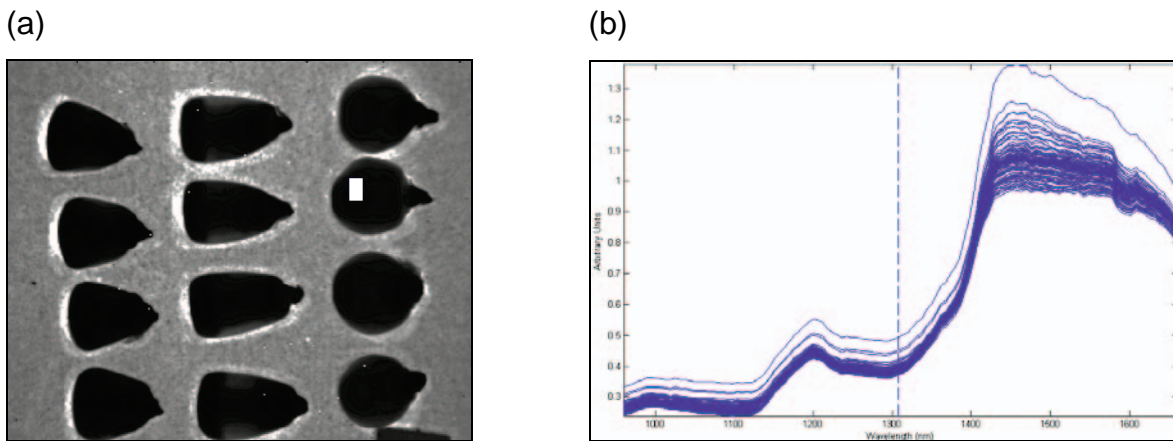
(a)

(b)

Figure 3.4 (a) Absorbance corrected image at 1310 nm without striations and bright zones and (b) consistent absorbance spectra of selected region, indicated by white square in (a).

Image cleaning

Principal component analysis with mean-centering was carried out and three components were calculated for a total of 91.74% of the total sum of squares (SS). The PCA score plot (**Fig. 3.5a**) of the MatrixNIR acquired image of the 12 maize kernels comprised of 81 920 pixels. This plot was used for exploration and identification of any special features (clusters or groupings) within the data set. Used in conjunction with the score image (**Fig. 3.5b**), it was possible to identify various clusters within the score plot that were associated with different components in the image. The score plot with selected clusters (**Fig. 3.5c**) was subsequently projected on the score image (**Fig. 3.5b**) which resulted in a classification image (**Fig. 3.5d**). This allowed for the identification of the respective clusters such as SiC sandpaper, bad pixels and edge effects, shading, maize kernels and plastic marker. Since these clusters differed either chemically, physically or by degree of intensity, they were separated to different locations within the hypercube. The hypercube thus allowed for the localisation of the chemical and biochemical constituents of the entire image field of view including the background (Gowen *et al.*, 2007). This made the selection, classification and removal of clusters relatively uncomplicated. The removal of the unwanted clusters, as indicated in the score plot (**Fig. 3.5c**) and classification image (**Fig. 3.5d**) resulted in a cleaned score image (**Fig. 3.6a**) that was used in subsequent analyses.
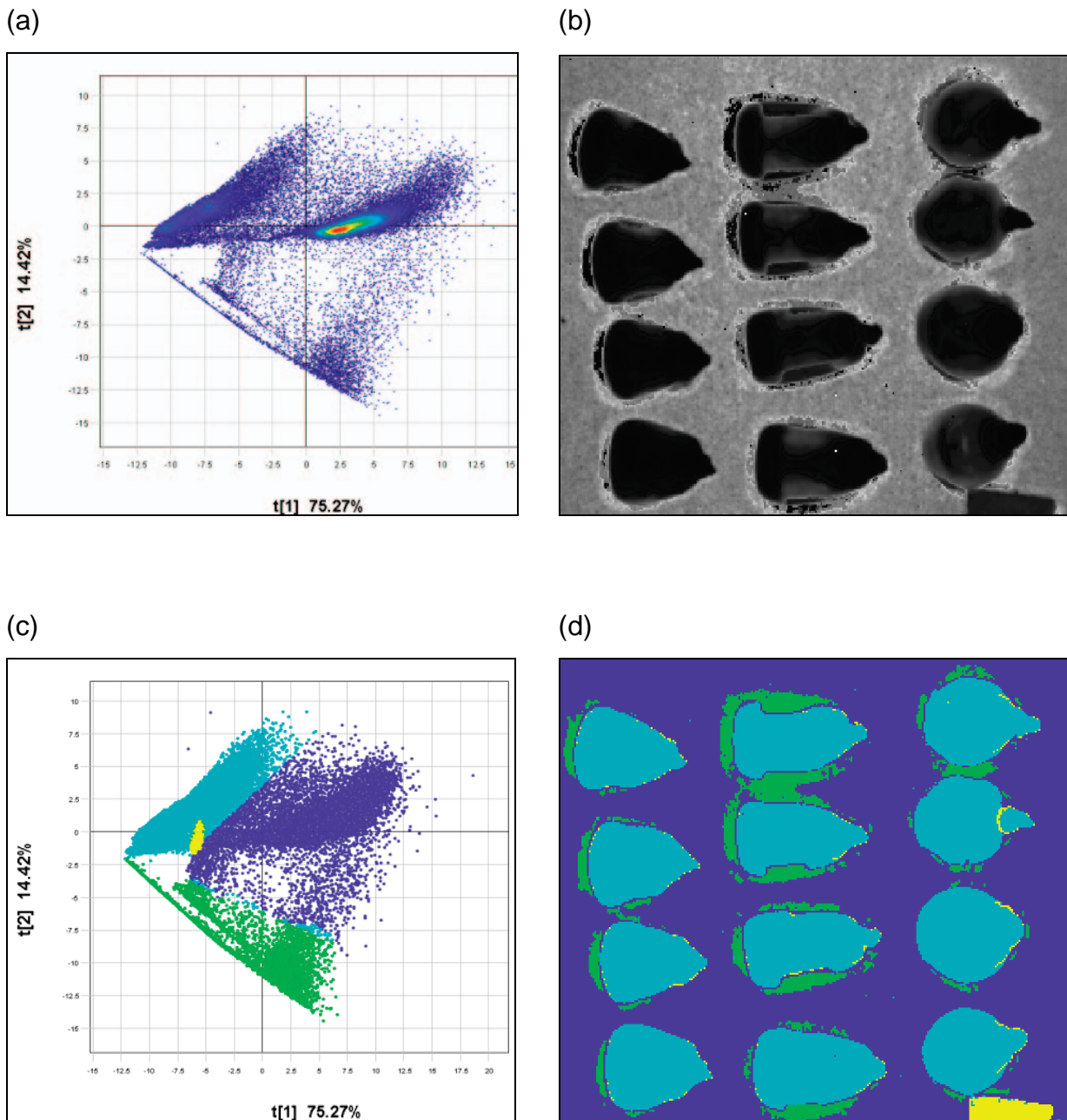
(a)

(b)

(c)

(d)

Figure 3.5 (a) PCA score plot of raw absorbance image (PC1 versus PC2) obtained with the MatrixNIR (81 920 pixels); (b) score image of PC1 used to identify and locate unwanted pixels; (c) PCA score plot with selected clusters projected onto (b) to obtain (d) classification image (blue = sandpaper, bad pixels and edge effects; cyan = maize kernels; green = shading; yellow = plastic marker).

Image analysis of cleaned images

After removal of the irrelevant clusters the cleaned score image (**Fig. 3.6a**) was more interpretable. PC1 (91.53%) was not useful in separating the data swarm effectively when plotted against PC2 (7.58%) or any of the other components as seen the score plots (**Figs. 3.6b-d**). It is known that if mean centred data are used PC1 typically contains no chemical information and usually include all physical variation (reflection, intensity or contrast and background) within an image; however this is case dependant (Esbensen & Lied, 2007). This is illustrated in the score plots (**Figs. 3.6b-d**) where PC1, although depicting clusters, was more likely explaining variation due to different light scattering properties than chemical differences. Combinations of higher order principal components will therefore be of more value to interpret the chemical differences between the maize kernels. Because similar clustering along PC1 was observed for the 24 kernel images acquired with both the MatrixNIR and the sisuChema imaging systems, the detailed results are not shown.
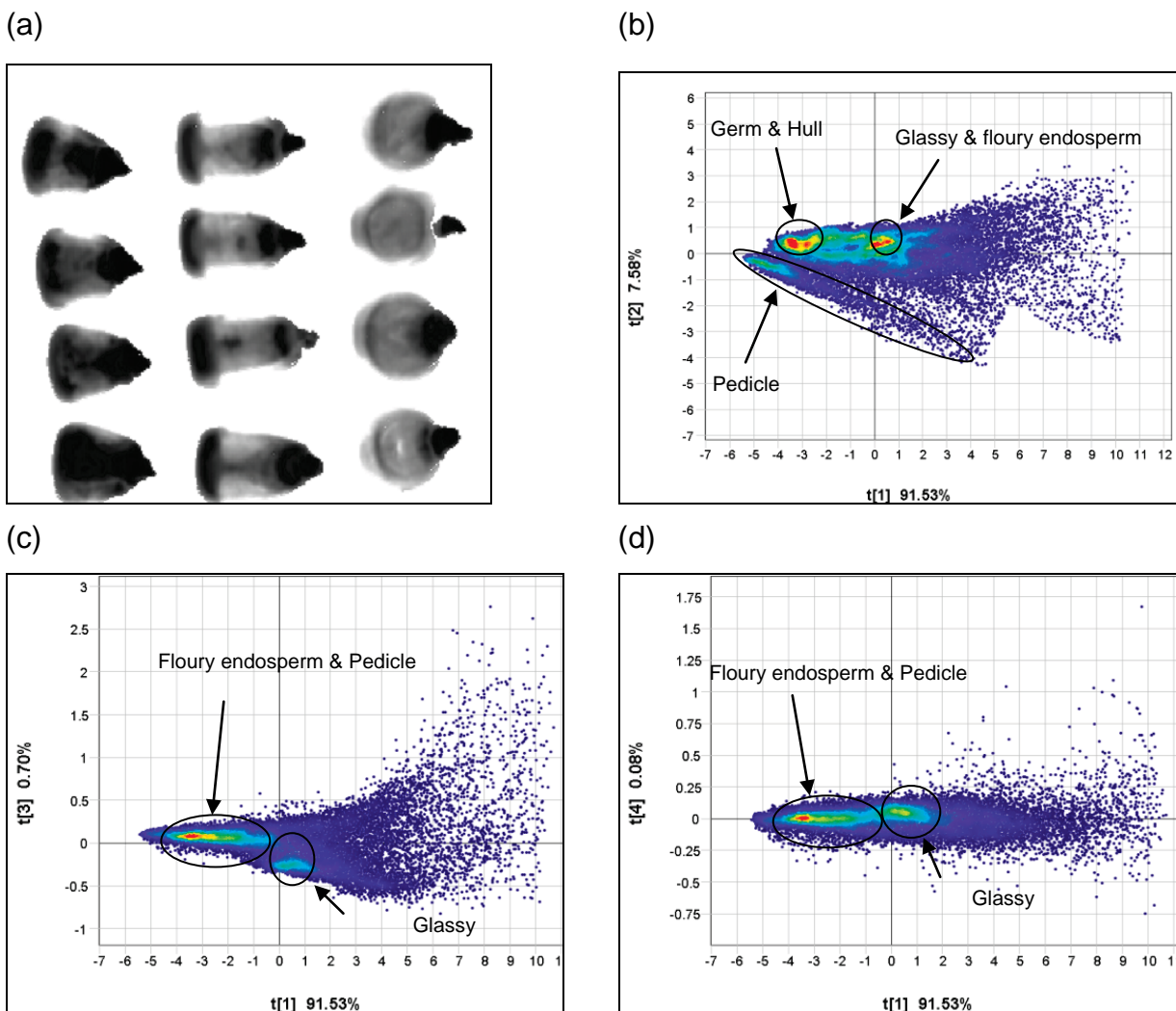
(a)

(b)

(c)

(d)

Figure 3.6 (a) Cleaned PCA score image (PC1) for MatrixNIR (23 798 pixels) and PCA score plots of (b) PC1 versus PC2; (c) PC1 versus PC3; and (d) PC1 versus PC4 illustrating clusters.

The score plot of PC2 versus PC3 (**Fig. 3.7a**) illustrated three clusters. Due to the interactive nature of the Evince software package the respective clusters in the score plot could be associated with specific parts of the maize kernels in the score image. This was accomplished by selecting a cluster in the score plot that simultaneously activated the score image wherein the location of the highlighted cluster was visualised. Employing this information in addition to histological knowledge of a maize kernel (Hoseney, 1994), the clusters could be assigned to classes with the resulting classification plot (**Fig. 3.7b**). The latter was projected onto the score image (**Fig. 3.7c**) resulting in a classification image (**Fig. 3.7d**) comprising three classes, i.e. glassy endosperm (blue), floury endosperm (green) and the pedicle and hull (yellow). Other combinations of principal components could not improve on the classification obtained from the score plot of PC2 versus PC3.

From these results it was observed that using the MatrixNIR with wavelength range 960-1660 nm it was possible to distinguish between the different endosperm types, i.e. glassy and floury, but not individual kernels from each hardness categories. This was due to the fact that each kernel, albeit hard, intermediate or soft, contains both glassy and floury endosperm in different ratios (Watson, 1987b). Due to the different ratios it was possible to distinguish between hard and soft kernels. Clear interpretation of loading line plots was not possible since averaged loading spectra of all kernels were obtained that explained the variation of the fundamental components of maize kernels, i.e. starch, moisture and protein.

It is well-known that differences between floury and glassy endosperm exist and these differences are both physical and chemical (Wolf *et al.*, 1952; Watson, 1987a; Watson, 1987b). Therefore it was decided to investigate the glassy and floury endosperm regions individually to determine what could be detected.
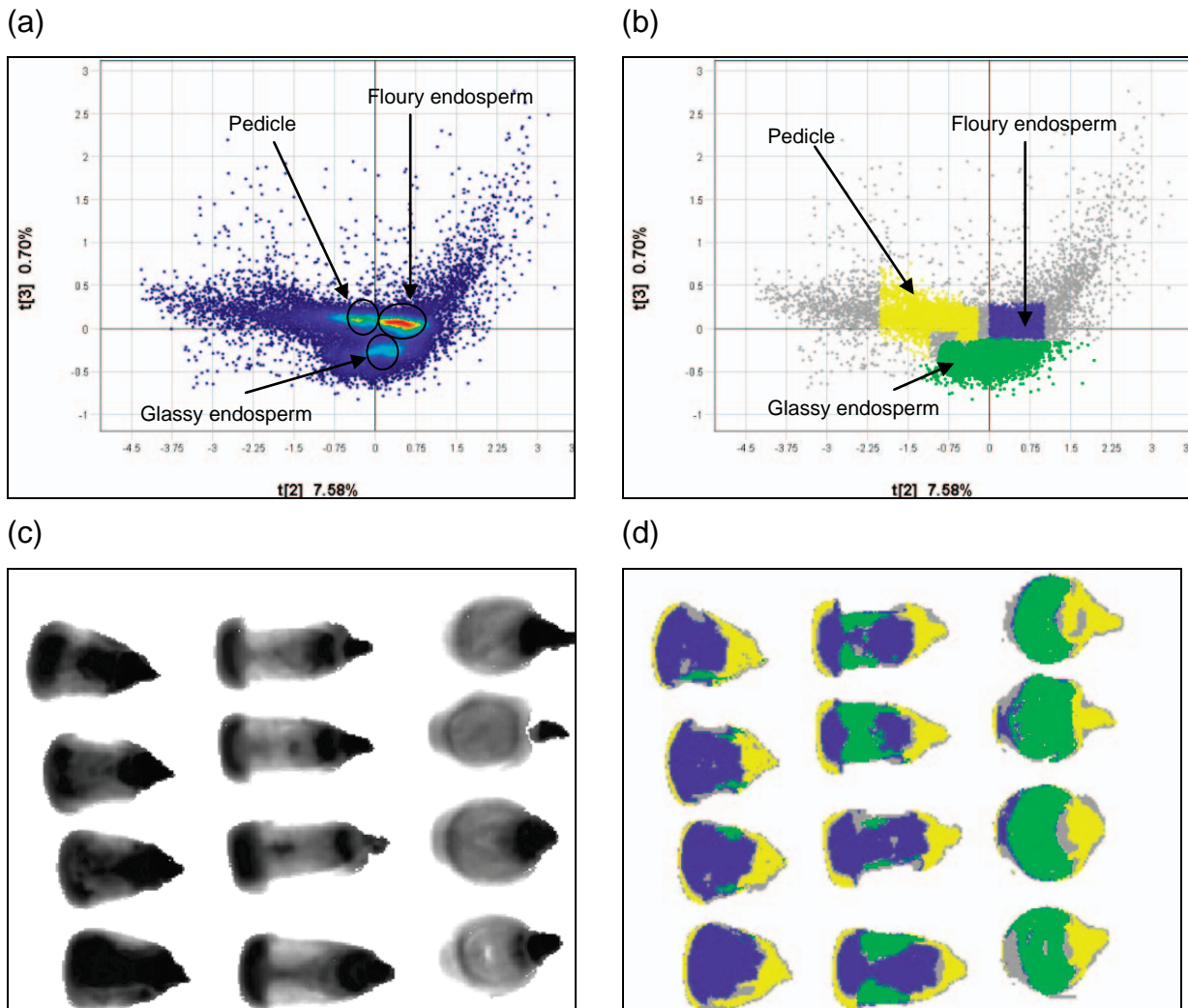
Figure 3.7 PCA score plots of PC2 (7.58%) versus PC3 (0.70%) depicting (a) three distinct density clusters and (b) with the three clusters assigned to classes which was then projected onto (c) the score image (PC1) resulting in (d) the classification image (green = glassy endosperm; blue = floury endosperm; yellow = pedicle).

To explain the chemical variation between the floury and glassy endosperm these two clusters or regions of interest (ROI's) were subjected to PCA individually. The respective loading line plots were interpreted based on chemical bonds present. The pedicle and hull were also excluded since floury and glassy endosperm are known to be absent in these regions (Watson, 1987a; Watson, 1987b; Hoseney, 1994). As expected, the loading line plot for PC1 (**Fig. 3.8a**) for both the floury and glassy endosperm portrayed no valuable chemical variation and depicted loading plots similar to an average spectrum of a maize kernel image (Esbensen & Lied, 2007).

It was ascertained that PC2 of both floury and glassy endosperm loadings depicted an absorption peak at 1215 nm (C-H stretching vibrations, second overtone) associated with CH_2 which in turn is associated with starch. Since the predominant component of the maize kernel is starch it was not unexpected to find that it could be a source of variation. In addition, given that every maize kernel is composed of both floury and glassy endosperm in different degrees, it was anticipated to find more variation due to starch in addition to finding multiple starch-related absorption peaks portraying the different endosperms.

The loading line plots of PC2 (**Fig. 3.8b**) for both floury (1.34% of total variance SS) and glassy endosperm (5.04% of total variance SS) illustrated variation primarily ascribed to differences in starch and moisture. This could be observed at 1450 nm which is a well-known region of moisture absorption in the NIR. However, starch also exhibits absorption in this region due to an O-H stretch first overtone. Although both loadings were similar, the loading line plot of the glassy endosperm portrayed an absorption peak at 1570 nm (N-H stretching vibrations, first overtone from a CH structure) which is associated with a peptide bond, thus associated to protein. This concurs with the literature (Cox *et al.*, 1944; Wolf *et al.*, 1952; Watson, 1987a; Watson, 1987b), protein and the starch forms a highly structured matrix in glassy endosperm on which zein bodies are found indented into the starch granules. Additionally, in the glassy endosperm, the protein matrix is thicker and remains intact on drying (Watson, 1987b). Floury endosperm, on the other hand, has a thin protein matrix that tears during drying resulting in air pockets.

Loading line plots of PC3 (**Fig. 3.8c**) for both floury and glassy endosperm exhibited variation due to starch and starch-moisture contrast. The variation in starch was expected since PC3 exhibited distinct separation of the hard and soft endosperm. Peaks at 1215 nm (C-H stretching vibrations, second overtone) associated with CH_2 and 1450 nm (O-H stretch, first overtone) with H_2O, confirmed this. For the floury endosperm, PC3 accounted for 0.12% of total SS while the glassy endosperm contributed 0.14%.

The variation in PC4 (**Fig. 3.8d**) for both floury and glassy endosperm was attributed to either an O-H stretch first overtone from starch at 1440 nm, or 2 x C-H stretch and C-H deformation due to a CH structure also at 1440 nm. This PC explained 0.07% of the variation in the floury endosperm and 0.14% of variation in the glassy endosperm. Although this component could be used to identify the endosperm, it could not be used to distinguish between the glassy and floury endosperm. Additionally, the loading indicated a degree of noise at the higher wavelengths. Principal components 5 and 6 (results not shown) displayed a considerable amount of noise and detector artefacts and were thus not useful in the elucidation of this model.
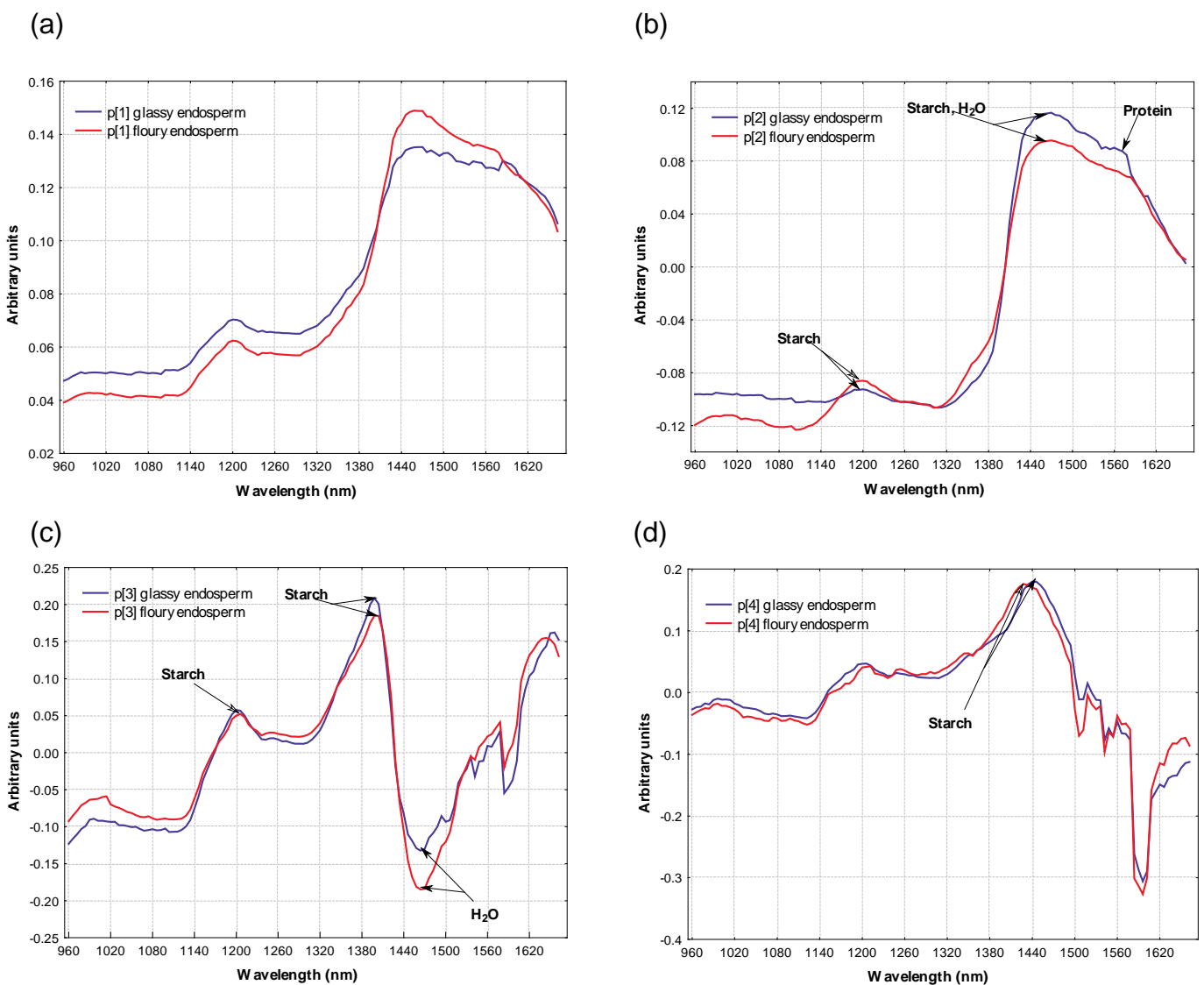
(a)

(b)

(c)

(d)

Figure 3.8 Loading line plots of (a) PC1, (b) PC2, (c) PC3 and (d) PC4 for both floury and glassy endosperm with PC1 portraying typical average spectra, PC2 absorption peaks at 1215 (starch), 1450 (moisture) and 1570 (protein) nm; PC3 starch-moisture contrast (1395 and 1450 nm); and PC4 a starch peak at 1440 nm and detector and/or monochromator anomalies at higher wavelengths.

Hyperspectral image analysis – MatrixNIR (24 kernels)

<u>Image correction and image cleaning</u>

The raw images of the 24 maize kernels acquired with the MatrixNIR camera were corrected and cleaned following the same approach as described earlier. Since the observations for these images were similar, the detailed results of the image cleaning process are not shown.

<u>Image analysis of cleaned images</u>

After removal of the irrelevant clusters, the PC score plot of PC2 (6.85%) versus PC3 (0.44%) showed similar clustering of pixels (**Fig. 3.9a)** as the previous MatrixNIR score plot (**Fig. 3.7a**). Already in the score image of PC2 (**Fig. 3.9c**) differences between glassy and floury endosperm regions were evident and with the knowledge gained from the previous analyses, it was clear which kernels were predominantly hard, intermediate and soft. As previously discussed, the score plot (**Fig. 3.9a**) was used to identify clusters and generated a classified score plot (**Fig. 3.9b**) that was further used together with the score image (**Fig. 3.9c**) to create a classification image (**Fig. 3.9d**). In spite of more kernels being imaged, the results were similar since the chemical constituents remained the same. This confirms the possibility to distinguish between different endosperm types using NIR hyperspectral imaging.

Hyperspectral image analysis – sisuChema (24 kernels)

<u>Image correction and image cleaning</u>

The raw image of the 24 yellow maize kernels acquired with the sisuChema imaging systems was corrected and cleaned following the same approach as described earlier. Since the observations for this image were similar, the detailed results of the image cleaning process are not shown.

<u>Image analysis of cleaned images</u>

A score plot of PC2 (4.60%) versus PC3 (0.66%), was obtained subsequent to the removal of the irrelevant pixels (**Fig. 3.10a**). This plot was unlike the previous score plots in that the separation of clusters was not as distinct. Even though the clusters were not that apparent, it was clear that three clusters existed, i.e. pedicle, floury endosperm and glassy endosperm. These clusters were identified, assigned to classes in the score plot (**Fig. 3.10b**) and projected onto the score image (**Fig. 3.10c**) to generate the classification image (**Fig. 3.10d**). The score image of PC2 (**Fig 3.10c**) illustrates the difference between

the MatrixNIR and the sisuChema. It appeared that the score image of PC2 contained more detail than the previous score images and to be of a higher resolution. However, similar to the previous score image, the dark regions represent areas of low absorption and are situated where the hard endosperm is expected to be.
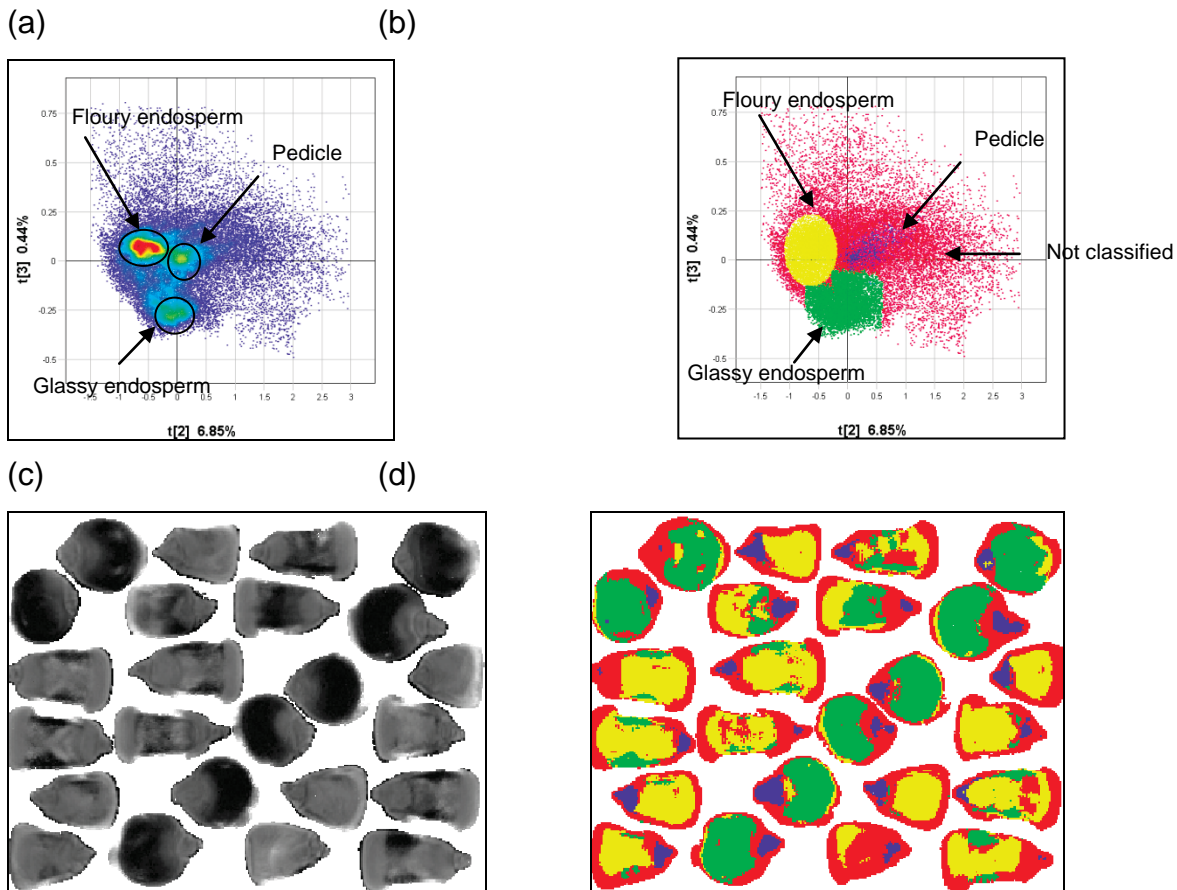

Figure 3.9 PCA score plots of PC2 (6.85%) versus PC3 (0.44%) for MatrixNIR 24-kernel image (42 672 pixels) depicting (a) three distinct density clusters and (b) with the three clusters assigned to classes which was then projected onto (c) the score image (PC2) resulting in (d) the classification image (green = glassy endosperm; blue = pedicle; yellow = floury endosperm; red = not classified).
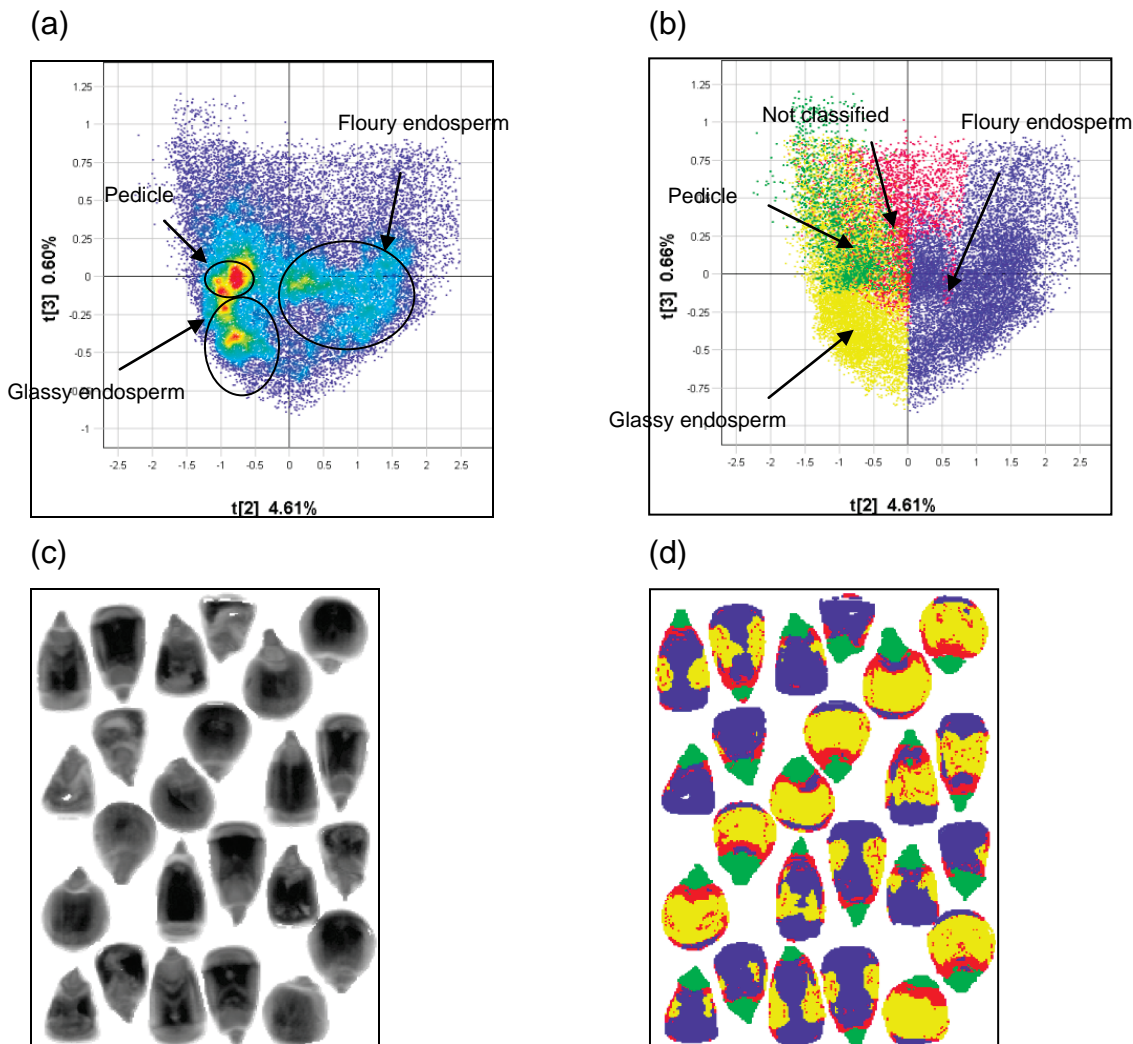
(a)

(b)

(c)

(d)

Figure 3.10 PCA score plots of PC2 (4.61%) versus PC3 (0.87%) for sisuChema (32 410 pixels) depicting (a) three distinct density clusters and (b) with the three clusters assigned to classes which was then projected onto (c) the score image (PC2) resulting in (d) the classification image (yellow = glassy endosperm; blue = floury endosperm; green = pedicle; red = not classified).

The loading line plot for the MatrixNIR image for the 24 yellow maize kernels, **Fig. 3.11a**, was similar to the loading line plots previously discussed except it was smoother than the previous loading line plots, especially in the higher components. The previous loading line plots were constructed from data extracted from portions of the kernels that were believed to be floury and glassy endosperm. This led to minimal variation implying most, if not all, of the variation was modelled within the first few components. The resultant was the loading line plots of higher components being noisy. Nevertheless, similar absorption peaks assigned to starch were evident at 1215 nm for both PC2 and PC3 and 1395 nm for both PC3 and PC4.

The loading line plot for the sisuChema image for the 24 yellow maize kernels, **Fig. 3.11b**, differed vastly from MatrixNIR loading line plot analyses since there were more

wavelengths. However, since the sample remained the same, similar absorption peaks were anticipated but at higher wavelength bands. These were found at 1695 nm (C-H stretch, first overtone) and 1900 nm (O-H stretch + 2 × C-O stretch) both associated with starch for PC3 and PC4.

The regions of interest differed primarily on PC2 where the glassy endosperm exhibited a protein peak at 1570 nm while the floury endosperm did not. This would imply the variation in the glassy endosperm was due to protein which is consistent with the literature since it has a more defined protein-starch matrix. However, PC3 was responsible for differentiating between glassy endosperm and floury endosperm since these clusters were clearly separated along this component. The loading line plots PC's 4, 5 and 6 for both floury and glassy endosperm were relatively similar and could not be used to indentify chemical components. The signal-to-noise ratio for the sisuChema seemed to be better than the MatrixNIR most likely because the monochromator of the MatrixNIR was not always functioning properly. This was apparent in the loading line plots that showed detector anomalies in the higher components of the MatrixNIR.
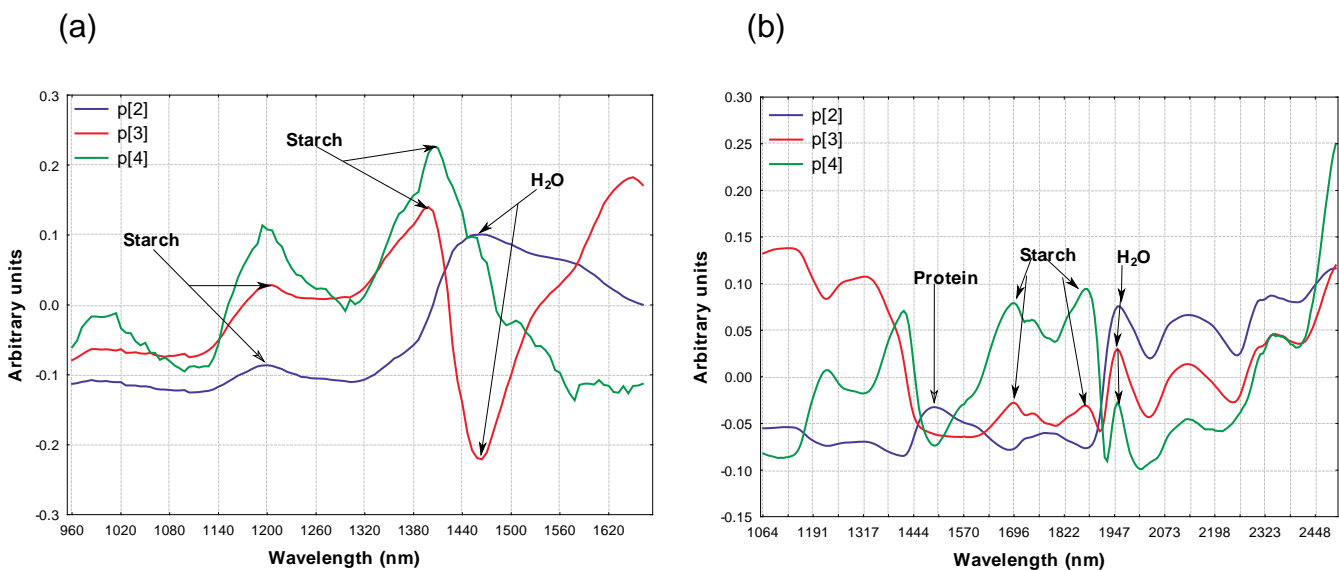
(a)

(b)

Figure 3.11 Loading line plots of PC2, PC3 and PC4 for (a) MatrixNIR image and (b) sisuChema for the 24 yellow maize kernels. Absorption peaks attributed to starch were apparent at 1215 nm for both PC2 and PC3 and 1395 nm for both PC3 and PC4 for the MatrixNIR and at 1695 nm and 1900 nm for both PC3 and PC4 for the sisuChema SWIR.

Partial least squares-discriminant analysis – MatrixNIR (12 kernels)

Figure 3.12 depicts the PLS-DA model overview illustrating the variation explained of the model up to 10 PLS components (88% of variation in Y explained). Using 3 to 6 components would be adequate as it would include 84-86% **Y** (floury and glassy endosperm classes) variation. PLS-DA was used to 1) determine whether it was possible to discriminate between the classes and 2) to build a model that could be used to predict future images.

Table 3.1 specifies how much of the test set observations were predicted as either floury or glassy endosperm. The first column indicates the classes assigned, the "not classified" class comprised all the pixels not assigned to either hard or soft endosperm. The second column shows the amount of pixels these classes were comprised of. Columns three, four and five indicate how much of these were accurately predicted for each class.

Of the 2364 pixels assigned to the floury endosperm class 2349 (99%) were correctly predicted, 14 (0%) were incorrectly predicted as glassy endosperm and 1 (0%) wrongly predicted as not classified. The glassy endosperm class comprised of 1215 pixels. Of the 1215 pixels, 1207 (99%) were correctly predicted, 7 (0%) incorrectly predicted as floury endosperm and 1 (0%) as not classified. The not classified class comprised of only 4 pixels that were probably associated with the pedicle area. This indicated the classes were accurately assigned and the removal of the pedicle was necessary to accurately predict the classes. The calibration and prediction images (**Figs. 3.13a & b**) illustrate the localisation of the modelled and predicted classes, i.e. floury and glassy endosperm.
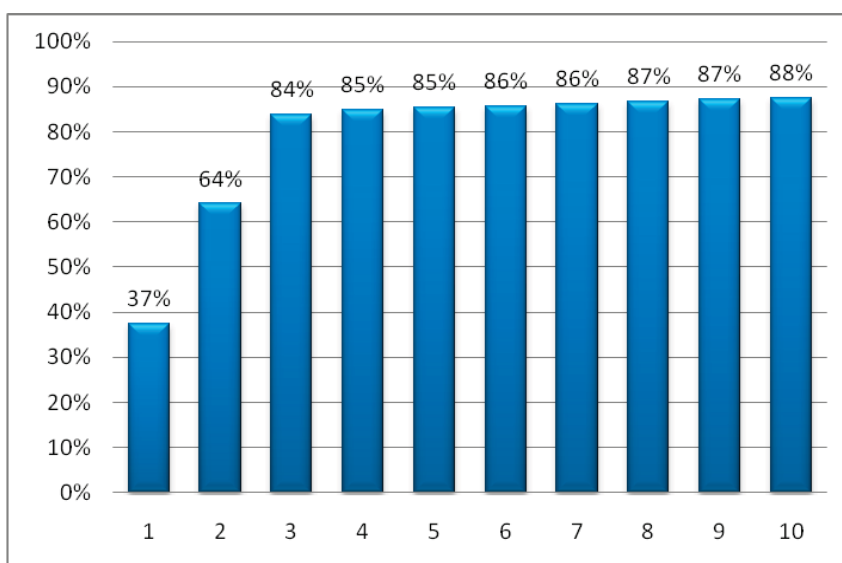

Figure 3.12 PLS-DA model overview illustrating the **Y** variation explained up to 10 PLS components for the MatrixNIR 12-kernel image.

Table 3.1 PLS prediction performance indicating the classes, number of pixels each class comprises and the prediction performance for floury endosperm, glassy endosperm and that of the classes not classified (MatrixNIR 12 kernels)

Classes	Number of Pixels (% pixels)	Floury	Glassy	Not classified
Floury	2364 (65%)	2349 (99%)	14 (0%)	1 (0%)
Glassy	1215 (33%)	7 (0%)	1207 (99%)	1 (0%)
Not classified	4 (0%)	2 (50%)	2 (50%)	-
Total	3583 (100%)	2358 (65%)	1223 (34%)	2 (0%)

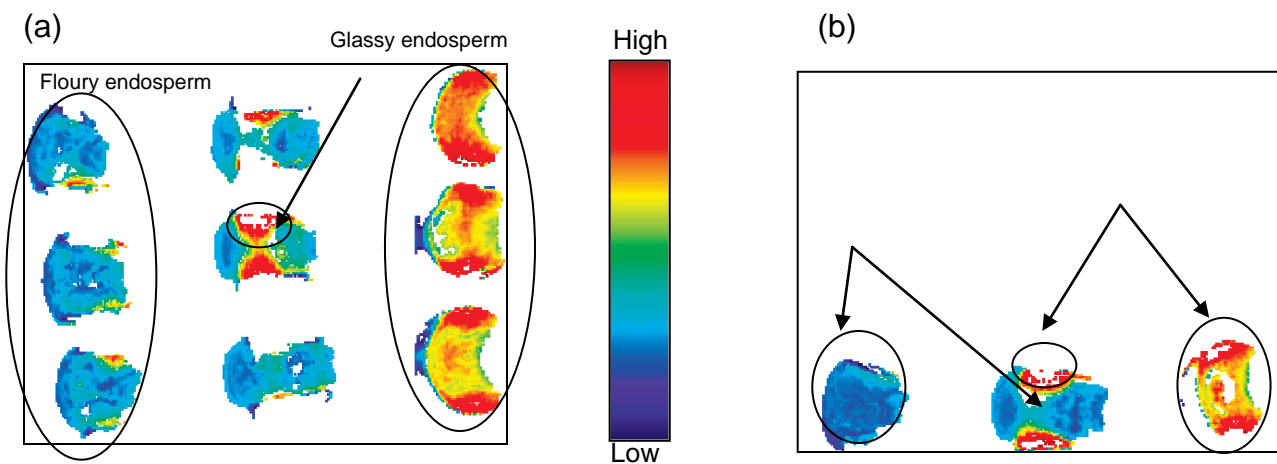Floury = floury endosperm; Glassy = glassy endosperm

Figure 3.13 Calibration image (a) and prediction image (b) for glassy endosperm on MatrixNIR 12 kernels after 4 PLS components (red = glassiness modelled well, blue = glassiness not modelled well).

Partial least squares-discriminant analysis – MatrixNIR (24 kernels)

After three PLS-factors 84.9% of the variation in **Y** was accounted for by the model. This was an immediate indication that this model was good and could perform well (**Table 3.2)**. **Table 3.2** specified how much of the test set observations were predicted as floury or glassy as previously described. Of the 2671 pixels allotted to the glassy endosperm class 2669 (99.9%) was correctly predicted as such and 2 (0.1%) were incorrectly predicted as floury endosperm. The floury endosperm class consisted of 5696 pixels. Of the 5696 pixels 95 (1.7%) were wrongly predicted as glassy endosperm and 5601 (98.3%) correctly predicted as floury endosperm.

The training set **(Fig. 3.14a)** illustrates the ability of the model to represent glassy endosperm in all the kernels. It is apparent from this image which kernels were hard, intermediate or soft. The kernels that are almost entirely red are those with a predominantly glassy endosperm, the rest are intermediate and soft. The intensity of the colour indicates how well the characteristic, glassy endosperm, was modelled. The warmer the colour the better it was modelled or predicted. **Figure 3.14b** gives an indication of the model's ability to predict glassy endosperm.

Table 3.2 PLS prediction performance indicating the classes, number of pixels each class comprised and the prediction performance for glassy endosperm soft endosperm (MatrixNIR 24 kernels)

Classes	Number of Pixels (%pixels)	Floury	Glassy
Floury	5696 (68.1%)	5601 (98.3%)	95 (1.7%)
Glassy	2671 (31.9%)	2 (0.1%)	2669 (99.9%)
Total	8367 (100%)	5603 (67%)	2764 (33%)

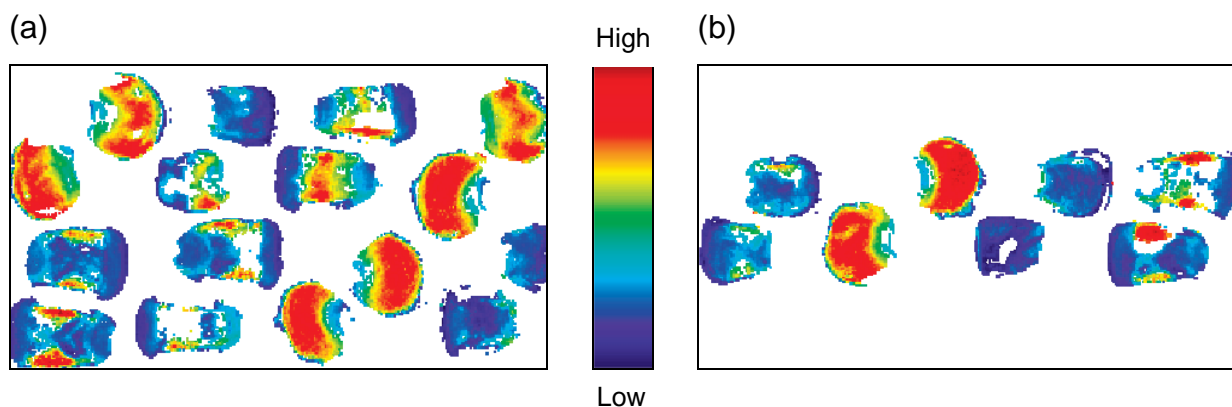Floury = floury endosperm; Glassy = glassy endosperm

Figure 3.14 Calibration image (a) and prediction image (b) for glassy endosperm on the MatrixNIR 24-kernel image after 3 PLS components (red = glassiness modelled well, blue = glassiness not modelled well).

Partial least squares-discriminant analysis – sisuChema (24 kernels)

The cumulative explained variance of **Y** increased dramatically after one PLS-factor (52%) to 76.3% after three PLS-factors. After an additional three PLS-factors a total of 77.5% of the SS in **Y** was explained; no discernable improvement of the model.

The prediction performance of the PLS-DA model is shown in Table 3.4. The image to which the model was applied consisted of 6430 pixels of where 3600 (56%) were assigned to the floury endosperm class and 2830 (44%) to the glassy endosperm. Of the 3600 pixels assigned to the soft class 3579 (99.4%) were correctly predicted while 21 (0.6%) were predicted as glassy endosperm. Of the 2830 pixels ascribed to the glassy endosperm, 96 (3.4%) were wrongly predicted as floury endosperm and 2734 (96.6%) correctly predicted as glassy endosperm.

The calibration image (**Fig. 3.15a**) represents the variation explained after 3 PLS components for glassy endosperm. Each column in **Table 3.3**, when selected would highlight the corresponding region of interest in the test set (**Fig. 3.15b**) allowing localisation of the predicted area.

The PLS-DA model for the MatrixNIR 12-kernel image portrayed the ability to discriminate between floury and glassy endosperm with a classification rate of 99% for both glassy and floury endosperm. The Matrix 24-kernel image had a classification rate of 82% for glassy endosperm and 73% for floury endosperm. The PLS-DA model for the sisuChema 24-kernel image had a classification rate of 95% for floury endosperm and 92% for glassy endosperm. Thus, it was successful in discriminating between glassy and floury endosperm using both the MatrixNIR and the sisuChema instruments that could ultimately lead to classification of single kernels of a certain hardness category.

Table 3.3 PLS prediction performance indicating the classes, number of pixels each class comprised and the prediction performance for glassy endosperm and for soft endosperm (sisuChema 24 kernels)

Classes	Number of Pixels (% pixels)	Floury	Glassy
Floury	3600 (56%)	3579 (99.4%)	21 (0.6%)
Glassy	2830 (44%)	96 (3.4%)	2734 (96.6%)
Total	6430 (100%)	3675 (57.2%)	2755 (42.8%)

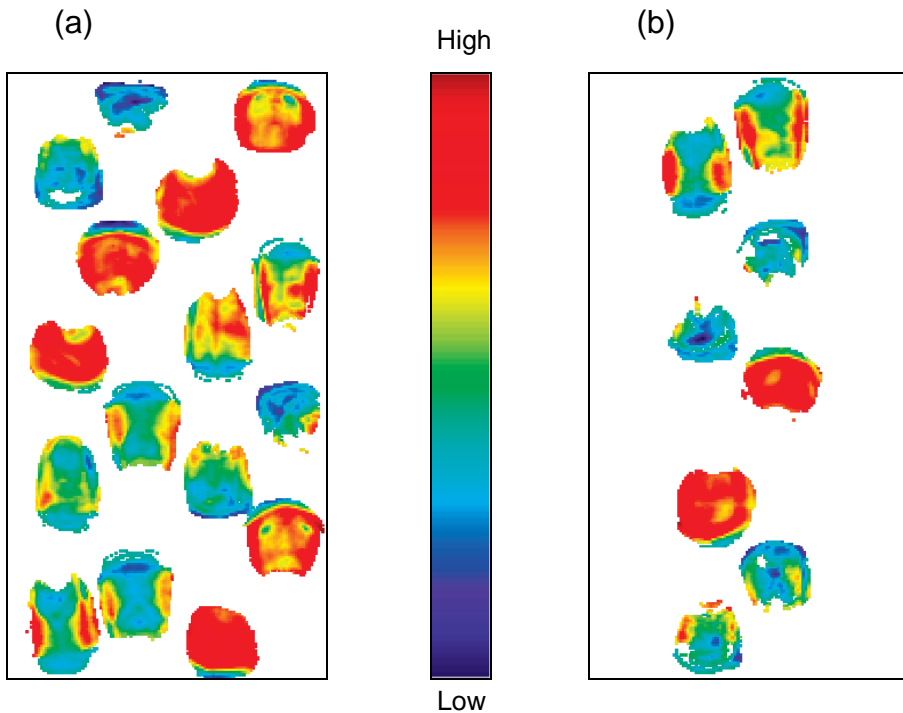Floury = floury endosperm; Glass = glassy endosperm

Figure 3.15 Calibration image (a) and prediction image (b) for glassy endosperm on sisuChema 24-kernel image after 3 PLS components (red = glassiness modelled well, blue = glassiness not modelled well).

Conclusion

The use of longer wavelength instruments could be more beneficial than shorter wavelength instruments; however this is application dependant. For research purposes it would be more useful to obtain a full spectrum of a sample to identify important wavelengths and components. Furthermore, very little detector error was found with the longer wavelength instrument and it was possible to identify more peaks associated with moisture, fat, protein and starch. This could then be used to develop a calibration which could be implemented in an instrument with only the wavelengths of importance to be used in industry allowing for quicker throughput. It is important to have prior knowledge concerning the application, sample and chemical constituents of importance.

It would, however, be more meaningful if it were possible to test this model on a new set of samples to determine its accuracy as this was not possible with the software package used during this study. A recommendation would be to validate the data by predicting a new image from the model obtained. In addition, images could be taken of kernels that have been dissected to verify the information that had been determined with the intact kernels, since more penetration would be possible with the dissected kernels. This would also give light to the interaction of floury and glassy endosperm.

The determination of maize hardness using NIR hyperspectral imaging and hyperspectral image analysis was investigated. It was found that the method of PLS-DA was adequate for the discrimination of floury and glassy endosperm, eventually leading to classification of kernels based on the floury to glassy endosperm ratio.
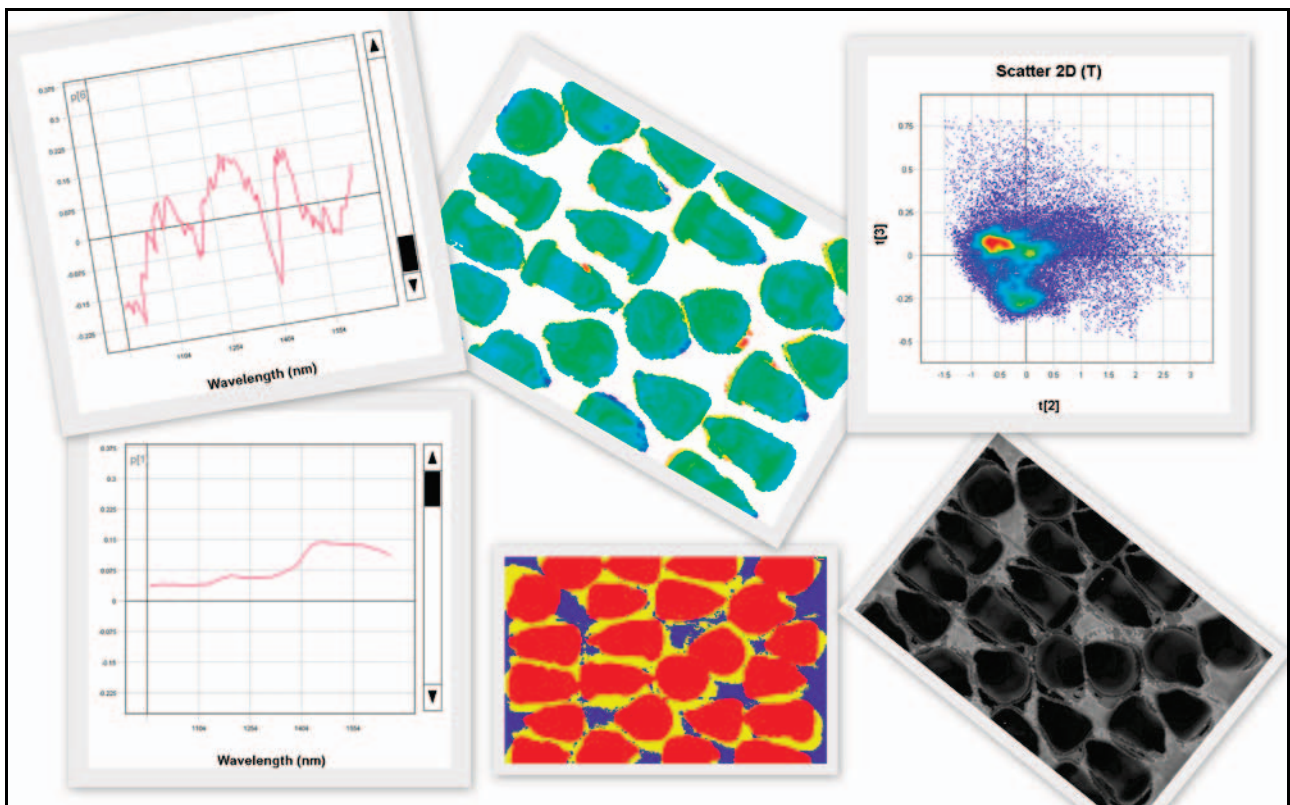
References

Almeida-Dominguez, H.D., Suhendro, E.L. & Rooney, L.W. (1997). Factors affecting rapid visco analyser curves for the determination of maize kernel hardness. *Journal of Cereal Science*, **25**, 93-102.

Baronti, S., Casini, A., Lotti, F. & Porcinai, S. (1997). Principal component analysis of visible and near-infrared multispectral images of works of art. *Chemometrics and Intelligent Laboratory Systems*, **39**, 103-114.

Berman, M., Connor, P.M., Whitbourn, L.B., Coward, D.A., Osborne, B.G. & Southan, M.D. (2007). Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis. *Journal of Near Infrared Spectroscopy*, **15**, 351-358.

Burger, J. & Geladi, P. (2005). Hyperspectral NIR image regression part I: calibration and correction. *Journal of Chemometrics*, **19**, 355-363.

Burger, J. & Geladi, P. (2006). Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst*, **131**, 1152-1160.

Chevallier, S., Bertrand, D., Kohler, A. & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, **20**, 221-229.

Cogdill, R.P., Hurburgh, C.R. & Rippke, G.R. (2004). Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE*, **47**, 311-320.

Cox, M.J., MacMasters, M.M. & Hilbert, G.E. (1944). Effect of the sulfurous acid steep in corn wet milling. *Cereal Chemistry*, **21**, 447-465.

Esbensen, K.H. & Lied, T.T. (2007). Principles of multivariate image analysis (MIA) in remote sensing, technology and industry. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P. Geladi). Pp. 17-41. Chichester, West Sussex: John Wiley & Sons, Ltd.

Eyherabide, G., Robutti, J. & Borras, F. (1996). Effect of near infrared transmission-based selection on maize hardness and the composition of zeins. *Cereal Chemistry*, **73**, 775-778.

Geladi, P., Burger, J. & Lestander, T. (2004). Hyperspectral imaging: calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems*, **72**, 209-217.

Geladi, P. & Grahn, H.F. (1996). *Multivariate Image Analysis.* Pp. 316. Chichester, West Sussex: John Wiley & Sons Ltd.

Geladi, P., Grahn, H.F. & Burger, J. (2007). Multivariate images, hyperspectral imaging: background and equipment. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P. Geladi). Pp. 1-14. Chichester, West Sussex: John Wiley & Sons Ltd.

Gorretta, N., Roger, J.M., Aubert, M., Bellon-Maurel, V., Campan, F. & Roumet, P. (2006). Determining vitreousness of durum wheat kernels using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **14**, 231-239.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, **18**, 590-598.

Gowen, A.A., O'Donnell, C.P., Taghizadeh, M., Cullen, P.J., Frias, J.M. & Downey, G. (2008). Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (*Agaricus bisporu*s). *Journal of Chemometrics*, **22**, 259-267.

Hoseney, R.C. (1994). *Principles of Cereal Acience and Technology.* Pp. 378. St. Paul, Minnesota: American Association of Cereal Chemists, Inc.

Lee, K.M., Herrman, T.J., Lingenfelser, J. & Jackson, D.S. (2005). Classification and prediction of maize hardness-associated properties using multivariate statistical analyses. *Journal of Cereal Science*, **41**, 85-93.

Lee, K.M., Bean, S.R., Alavi, S., Herrman, T.J. & Waniska, R.D. (2006). Physical and biochemical properties of maize hardness and extrudates of selected hybrids. *Journal of Agricultural and Food Chemistry*, **54**, 4260.

Peirs, A., Scheerlinck, N., Nicolai, B.M. & De Baerdemaeker, J. (2003). Starch degradation analysis of apple fruits measured with a hyperspectral (NIR) imaging system. *Proceedings of the International Conference Postharvest Unlimited*, 315-321.

Pomeranz, Y., Czuchajowska, Z., Martin, C. & Lai, F. (1985). Determination of maize hardness by Stenvert hardness tester. *Cereal Chemistry*, **62**, 108-112.

Pomeranz, Y., Martin, C.R., Traylor, D.D. & Lai, F.S. (1984). Corn hardness determination. *Cereal Chemistry*, **61**, 147-150.

Smail, V.W., Fritz, A.K. & Wetzel, D.L. (2006). Chemical imaging of intact seeds with NIR focal plane array assists plant breeding. *Vibrational Spectroscopy*, **42**, 215-221.

Tallada, J.G., Nagata, M. & Kobayashi, T. (2006). Non-destructive estimation of firmness of strawberries (Fragaria x ananassa Duch.) using NIR hyperspectral imaging. *Environment Control in Biology*, **44**, 245-255.

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

Watson, S.A. (1987a). Measurement and maintenance of quality. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 125-183. St Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Watson, S.A. (1987b). Structure and composition. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 53-82. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Wehling, R.L., Jackson, D.S. & Hamaker, B.R. (1996). Prediction of corn dry-milling quality by near-infrared spectroscopy. *Cereal Chemistry*, **73**, 543-546.

Wehling, R.L., Jackson, D.S., Hooper, D.G. & Ghaedian, A.R. (1993). Prediction of wet-milling starch yield from corn by near-infrared spectroscopy. *Cereal Chemistry*, **70**, 720-723.

Wolf, M.J., Buzan, C.L., MacMasters, M.M. & Rist, C.E. (1952). Structure of the mature corn kernel. I. Gross anatomy and structural relationships. *Cereal Chemistry*, **29**, 321-333.

Chapter 4

Indirect detection of *Fusarium verticillioides* in maize kernels by NIR hyperspectral imaging

Chapter 4

Indirect detection of *Fusarium verticillioides* in maize kernels by NIR hyperspectral imaging

Abstract

Near infrared (NIR) hyperspectral imaging and hyperspectral image analysis was evaluated to distinguishing between fungal infected and sound kernels. Using a Spectral Dimensions MatrixNIR camera with a spectral range of 960-1662 nm as well as a sisuChema SWIR (short wave infrared) hyperspectral pushbroom imaging system with a spectral range of 1000-2498 nm, NIR hyperspectral images of infected and sound whole maize kernels were acquired. Background, bad pixels and shading were removed using exploratory principal component analysis (PCA) on absorbance images. PCA could be used effectively on the cleaned images to find classes including infected regions and the pedicle as well as non-infected regions. A distinct difference between infected and sound kernels along principal component (PC) one with two distinguishable clusters was found. Interpretation of the PC loading line plots showed important absorbance peaks responsible for the variation at 1215, 1450, 1480, 1690, 1940 and 2136 nm for both MatrixNIR images (15 and 21 kernels). Partial least squares-discriminant analysis (PLS-DA) was applied afterwards. Classifications rate of up to 96.1% and 99% for the discrimination of infected kernels were obtained on The MatrixNIR-15 kernel image as well as the sisuChema-15 kernel image. The sisuChema 21-kernel image had a classification rate for infected kernels of 97.6% without pre-processing, 97.7% with multiplicative scatter correction (MSC) and 97.4% with standard normal variate (SNV).

Introduction

Fusarium verticillioides (Saccardo) Nirenberg is commonly associated with maize kernels (Bacon *et al.*, 1992; Munkvold & Desjardins, 1997; Marin *et al.*, 1999; Fandohan *et al.*, 2003; Oren *et al.*, 2003; Bush *et al.*, 2004; Desjardins, 2006; Yates & Sparks, 2008). It is associated with diseases at all stages of maize plant development; infecting the roots, stalk, and kernels (Munkvold & Desjardins, 1997). This fungus is not only the most common pathogen of maize, but is also among the most common fungi found colonising symptomless maize plants. The fungus is an endophyte suggesting it arrives at its internal location in kernels from growth through the stalk, into the cob and pedicle (Bacon *et al.*, 1992). *Fusarium verticillioides* is an almost constant companion of maize plants and seed (Munkvold & Desjardins, 1997; Marasas, 2001; Fandohan *et al.*, 2003; Gelderblom *et al.*, 2004; Fandohan *et al.*, 2005; Desjardins, 2006). In many cases, its presence is ignored

because it does not cause visible damage. Symptomless infection can exist throughout the plant, and seed-transmitted strains of the fungus can develop systemically to infect the kernels.

NIR hyperspectral imaging data are arranged into a three-way matrix known as a hypercube (Cogdill *et al.*, 2004; Burger & Geladi, 2006; Gowen *et al.*, 2007; Gowen *et al.*, 2008). It comprises two spatial axes (x & y), horizontal and vertical pixel coordinates, and one wavelength dimension (z) that includes the spectral data. The hypercube is, depending on the instrument, composed of hundreds of single channel, grey scale images that represents a single band of spectral wavelength (Burger & Geladi, 2006). This combination of spatial and spectral data allows for identification and localisation of chemical constituents in samples.

NIR hyperspectral imaging and support vector machine has been evaluated for the classification of fungal infected wheat kernels with classification rates of 92.9% for *Aspergillus niger*, 87.2% *for Aspergillus glaucus* and 99.3% and 100% for two Penicillium species (Zhang *et al.*, 2007). In two similar studies, a classification rate of up to 95% was achieved for the classification of sound and stained wheat kernels (Berman *et al.*, 2007) and 95% for the fungal detection in wheat (Singh *et al.*, 2007). More recently a classification rate of up to 97.7% for the differentiation of toxigenic fungi has been obtained (Yao *et al.*, 2008).

The aim of this study was to determine whether NIR hyperspectral imaging could distinguish and localise infected whole maize kernels by evaluating:

- the use of PCA and the interpretation of principal component (PC) loading line plots to explain the chemical variation within and infected and non-infected regions; and
- partial least squares-discriminant analysis (PLS-DA) as a possible chemometric classification technique.

Materials and methods

Samples

Maize kernels of varying degrees of fungal infection were kindly provided by the Department of Plant Pathology (Stellenbosch University, Stellenbosch, South Africa). These consisted of kernels infected with *Fusarium verticillioides* (MRC 0826) and were divided into three groups: infected with visual symptomatic effects, infected with asymptomatic effects and a control group that was treated with distilled H_2O.

The kernels were obtained from maize plants that were inoculated with the MRC 0826 strain of *Fusarium verticillioides*. *Fusarium verticillioides* (MRC 0826) was cultured on half

strength potato dextrose agar (19.5 g.L^{-1}). After 14 days, autoclaved water with Tween 20 (3 drops.L^{-1}) was used to wash the spores from the agar. The spore suspension was poured through two layers of sterile cheesecloth to remove mycelium thereafter the suspension was adjusted to 10^6 spores.mL^{-1} using a haemocytometer. The maize plants were inoculated at the mid-silk stage with 2 mL of spore suspension by injecting the spore suspension into the silk channel of the maize ear using a modified 1 mL pipette tip, fitted to a disposable 10 mL syringe. For the control maize plants 2mL of autoclaved water was injected instead of the spore suspension. Maize cobs were harvested 3 months after inoculation.

Near infrared hyperspectral imaging systems

Near infrared hyperspectral images were acquired using a Spectral Dimensions MatrixNIR focal plane array camera (Malvern Instruments Ltd, Malvern, Worcestershire, UK) as well as a sisuChema SWIR (short wave infrared) hyperspectral, pushbroom imaging system (Specim, Spectral Imaging Ltd, Oulu, Finland). The Spectral Dimensions MatrixNIR camera comprised of an Indium Gallium Arsenide (InGaAs) diode array detector with a liquid crystal tunable filter (LCTF) while the sisuChema comprised an imaging spectrograph coupled to a 2-D array Mercury-cadmium-telluride (HgCdTe) detector as described in Chapter 3.

Image acquisition

Fifteen white maize kernels, five kernels randomly selected from each infection category were positioned on a 70 mm x 70 mm silicon carbide (SiC) sandpaper according to a Latin Square design as depicted in **Fig. 4.1**. These samples were presented for image acquisition on both the MatrixNIR and the sisuChema. In addition 21 white maize kernels, eight non-infected kernels, eight infected kernels and five asymptomatic kernels were randomly selected from each category and positioned on 70 mm x 70 mm SiC sandpaper according to a Latin Square design as depicted in **Fig. 4.2**. These samples were presented on the sample holder of the sisuChema for image acquisition. Digital images were acquired with a Leica Digilux 3 camera (Leica Camera AG, Solms, Germany).

The image acquisition time on the Spectral Dimensions MatrixNIR camera was *ca.* 8 min at 128 ms integration time. This included the acquiring of the images of reflectance standards (Labsphere, USA) necessary for image calibration and to correct for pixel-to-pixel variances due to camera inconsistencies and variation in sample illumination (Burger & Geladi, 2005). Each channel was scanned 16 times and the samples were kept cool

with a small box fan positioned just beyond the field of view. Standards of 2% (dark reference) and 25% (white reference) reflectance were used. The image acquisition time on the sisuChema was *ca.* 8 seconds with an exposure time of 1.5 ms and a frame rate of 100 Hz. Both internal dark and white reference standards were imaged immediately before image acquisition of the sample.
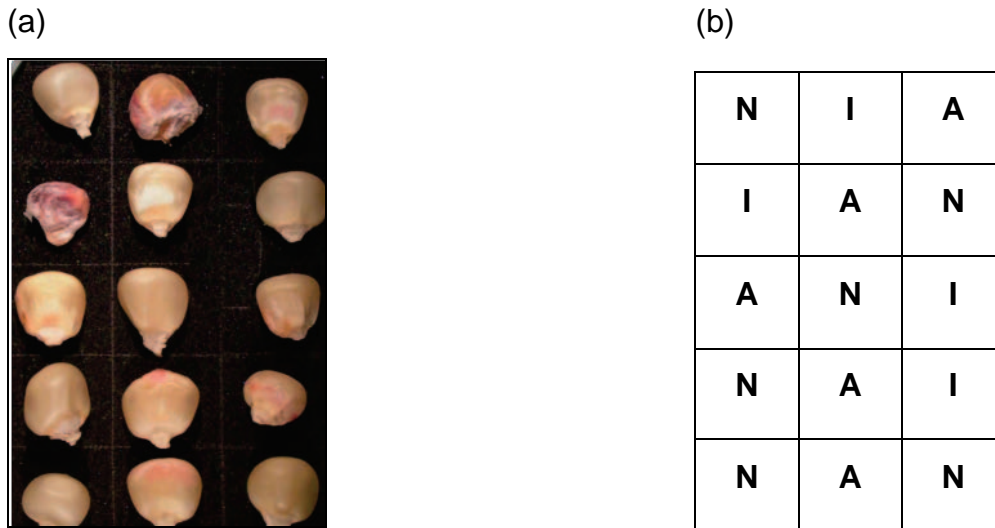
(a)

(b)

N	I	A
I	A	N
A	N	I
N	A	I
N	A	N

Figure 4.1. (a) Digital image illustrating sample presentation of the 15 white maize kernels for image acquisition with the Spectral Dimensions MatrixNIR camera as well as the sisuChema SWIR imaging system and (b) grid illustrating the infection class of each maize kernel (N = not infected; I = infected; A = infected but asymptomatic)
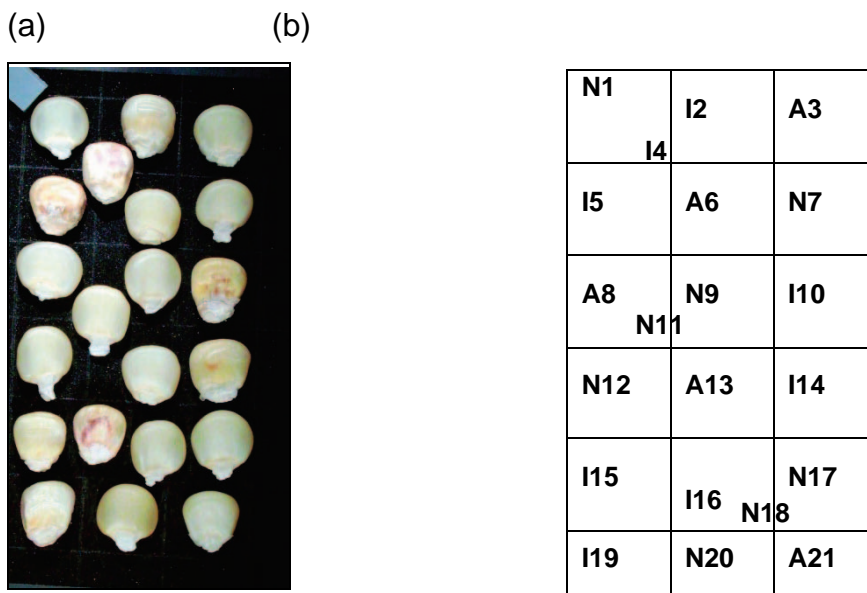
(a)   (b)

N1 I4	I2	A3
I5	A6	N7
A8 N11	N9	I10
N12	A13	I14
I15	I16 N18	N17
I19	N20	A21

Figure 4.2 (a) Digital image illustrating sample presentation of the 21 white maize kernels for image acquisition with the sisuChema SWIR imaging system and (b) the layout grid illustrating the infection class of each maize kernel (N = not infected; I = infected; A = infected but asymptomatic) and the number assigned.

Microbial analysis

The 21 kernels imaged with the sisuChema were dissected; the germ (pedicle included) and endosperm regions were scraped and plated on potato dextrose agar (PDA) then incubated at 30 $^{\circ}$C for 72 hrs. Petri dishes were examined for growth and where growth was present the fungus was isolated, stained with Lactophenol blue solution (E. Merck, D-61, Darmstadt, 13741) and identified microscopically (Samson, 2000). Kernels and plates were numbered.

Hyperspectral image analysis (HIA)

<u>Image correction</u>

Sample images (I_0) acquired with the Spectral Dimensions MatrixNIR camera were converted from instrument measurement values, i.e., A/D (analogue to digital) counts to absorbance values as described in Chapter 3.

The images acquired with the sisuChema were automatically converted to pseudo-absorbance in the Evince 2.2 hyperspectral image analysis software package (UmBio AB, Umeå, Sweden).

<u>Image cleaning</u>

Absorbance images were subjected to PCA. PCA score plots as well as the PCA score images were used to identify and classify outliers, bad pixels, illumination errors and shading errors, edge effects and background. PCA was recalculated with additional components after the irrelevant classes were removed obtaining a cleaned image. The cleaned image was used in subsequent analysis.

<u>Image analysis of cleaned image</u>

PCA score plots and the accompanied score images were used to identify and classify clusters to the various classes (infected, asymptomatic and not-infected). These were then projected onto PCA score images.

<u>Partial least squares-discriminant analysis (PLS-DA)</u>

Each object in the two classes (infected and not-infected) was assigned to a dummy variable. The images were divided into a training set and a test set by assigning one third of each image on each instrument to the test set and the remainder to the training set. For the sisuChema 21-kernel image, two pre processing techniques multiplicative scatter correction (MSC) and standard normal variate (SNV) were applied individually.

Results and discussion

Hyperspectral image analysis – MatrixNIR (15 kernels)

<u>Image cleaning</u>

Principal component analysis with mean-centering was carried out and three components were calculated for a total of 97.46% of the total sum of squares (SS). Similar to the PCA score plot of the MatrixNIR 12-kernel image in Chapter 3, the PCA score plot (**Fig. 4.3a**) and score image (**Fig. 4.3b**) was composed of 81 920 pixels. As described in Chapter 3, the interactive nature of the plots made it possible not only to identify irrelevant pixels but also to assign them to classes. Since the entire image field of view was associated with different organic bonds in different combinations it was possible to differentiate between each element in the image by way of the hypercube (Burger, 2006; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Gowen *et al.*, 2008). This made it possible to assign the various clusters to classes to obtain a score plot with classes (**Fig. 4.3c**) that was projected onto the score image (**Fig. 4.3b**) to produce a classification image (**Fig. 4.3d**).

<u>Image analysis of cleaned images</u>

Subsequent to the removal of the irrelevant pixels the score image (**Fig. 4.4a**) was more comprehensible and the six components encompassed 99.11% of the total SS. The image now comprised 33 506 pixels that only contained variation pertaining to the kernels. PC1 (94.51%) plotted against PC2 (3.31%) was not very effective in separating the data swarm in the score plot (**Fig. 4.4b**). Since clusters were not that apparent in the score plot, PC1 versus PC2 was not used to determine if differences between the infected and non-infected kernels could be detected. Furthermore, the lack of separation between the clusters could lead to misclassification when assigning the clusters to classes that would eventually lead to poor modelling when applying PLS-DA. Although PC1 is generally accepted to show an overall albledo effect (Esbensen & Lied, 2007) it was found that when plotted against any combination of other components (**Figs. 4.4c-f**), an improved separation of clusters was apparent. To determine what the clusters represented, they were selected in the score plot (**Figs. 4.5a & b**) and the associated location was simultaneously highlighted in the maize kernels in the score image (**Figs. 4.5c & d**). This was only done for PC1 versus PC3 as the other combinations of components with PC1 yielded similar results, i.e. similar separation of clusters, and are thus not shown. Using this method and known information concerning the kernels, it was determined that the cluster in the right half of the score plot (**Fig. 4.5a**) was associated with the pedicle and the

infected regions of the maize kernel while the left cluster in the score plot (**Fig. 4.5b**) was associated with the non-infected regions of the kernel.
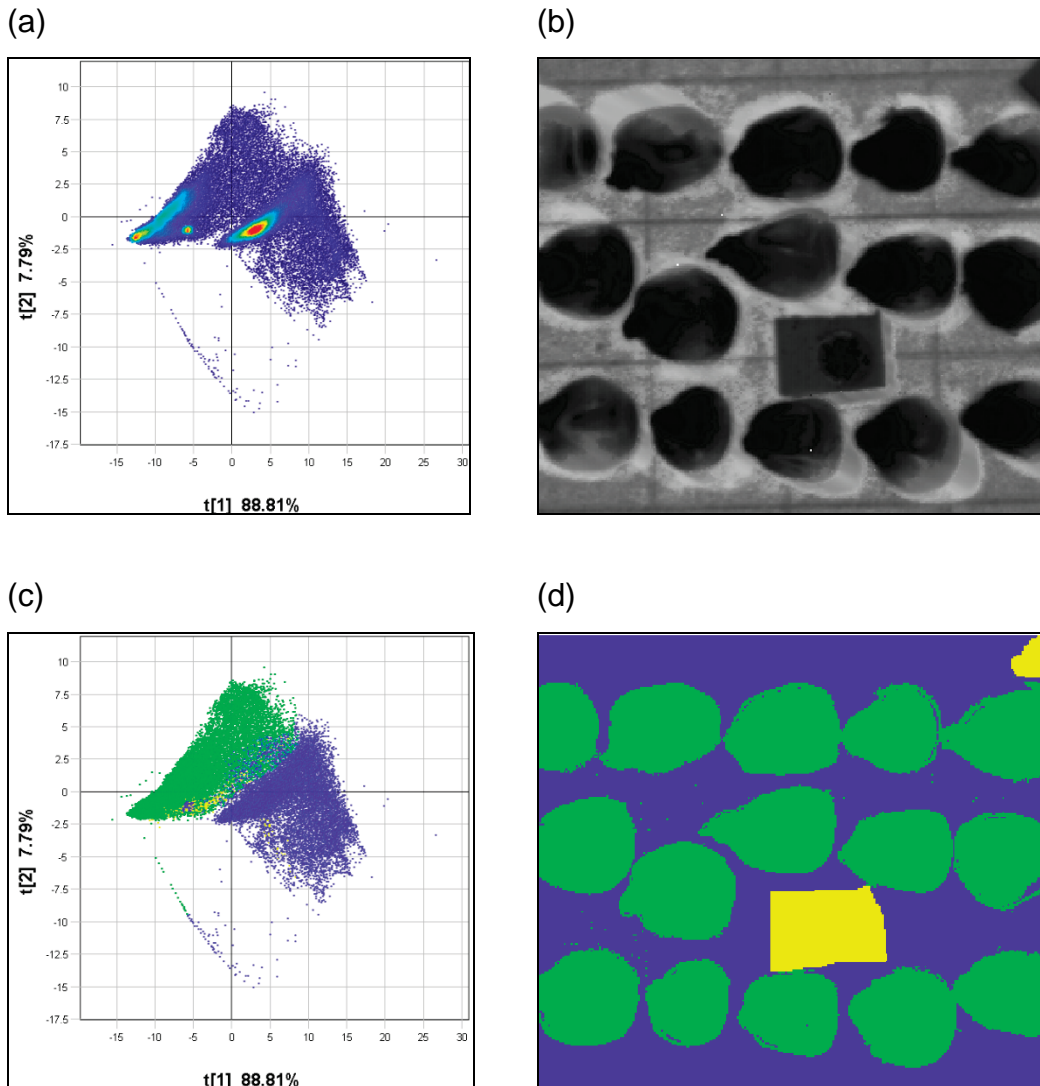
(a)

(b)

(c)

(d)

Figure 4.3 (a) PCA score plot of uncleaned image (PC1 versus PC2) obtained with the MatrixNIR (15 kernels) (81 920 pixels); (b) score image of PC1 used to identify and locate unwanted pixels; (c) PCA score plot with selected clusters projected onto (b) to obtain (d) classification image (blue = sandpaper, bad pixels and edge effects; green = maize kernels; yellow = plastic marker).
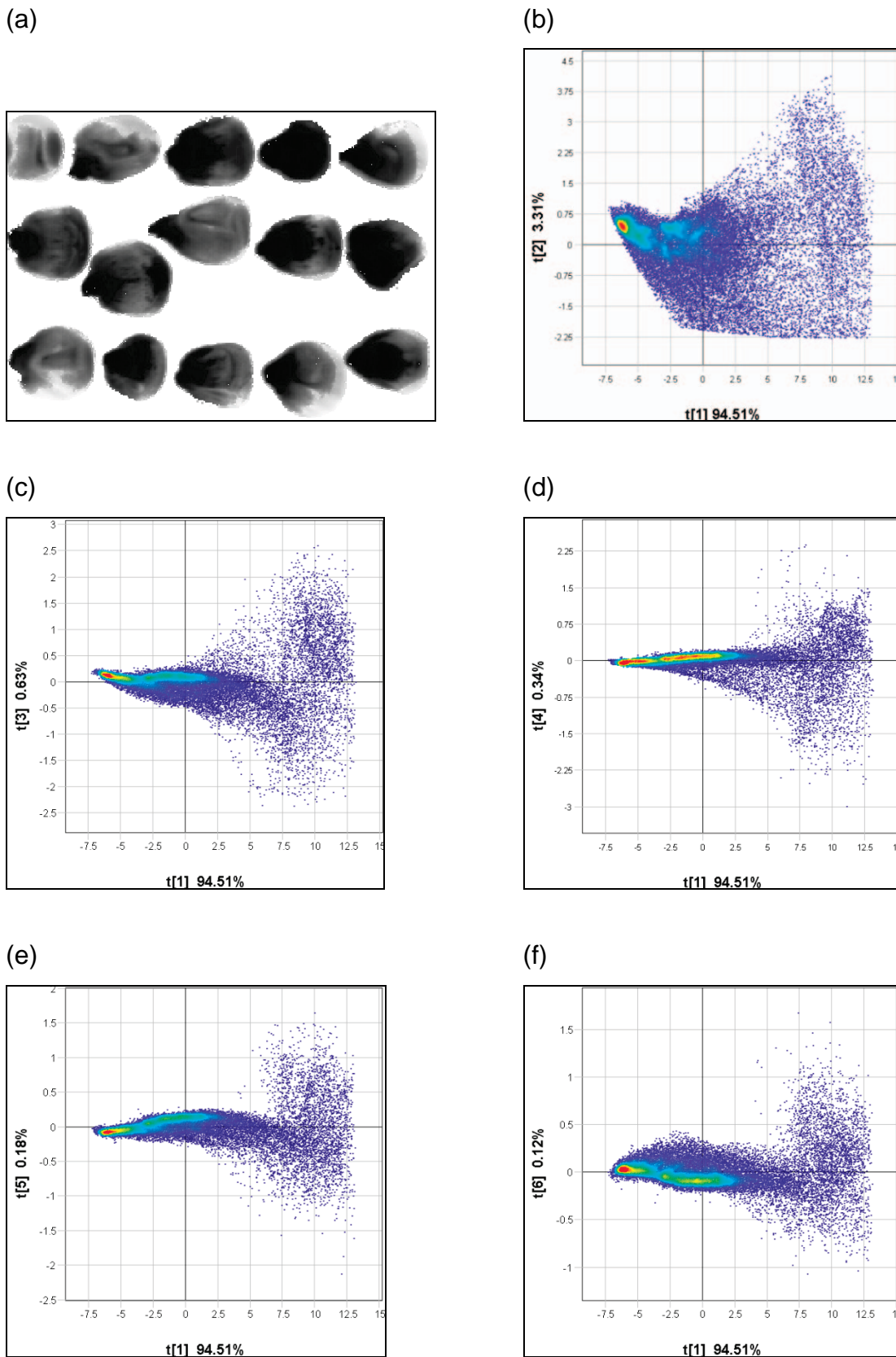
(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.4 (a) Cleaned PCA score image (PC1) for MatrixNIR (15 kernels) (33 506 pixels) and PCA score plots of (b) PC1 versus PC2; (c) PC1 versus PC3; (d) PC1 versus PC4, (e) PC1 versus PC5; and (f) PC1 versus PC6 illustrating clusters.

These findings were in agreement with background knowledge. However, whenever the infected region was selected it was found that the pedicle always appeared highlighted in all the kernels (infected or non-infected) in the score image, suggesting that it too was infected even in the non-infected kernels. This could be ascribed to the fact that the pedicle is enclosed in dead tissue in addition to containing less than *ca*. 3% starch, proteins and oil found in the whole kernel (Watson, 1987). Since it is known that the fungus has α-amylase activity and that this is the primary starch utilisation enzyme of filamentous fungi (Bluhm & Woloshuk, 2005), the phenomena in the score plot (**Figs. 4.5a & b**) and score image (**Figs.4.5c & d**) could be attributed to the hydrolysis of starch. In the more heavily infected kernels it could be seen that more starch had been hydrolysed while the non-infected kernels displayed the presence of starch. Thus the fact the pedicle of the non-infected kernels occur in the infected cluster is probably due to fact that there is not much starch present in this region of the kernel.
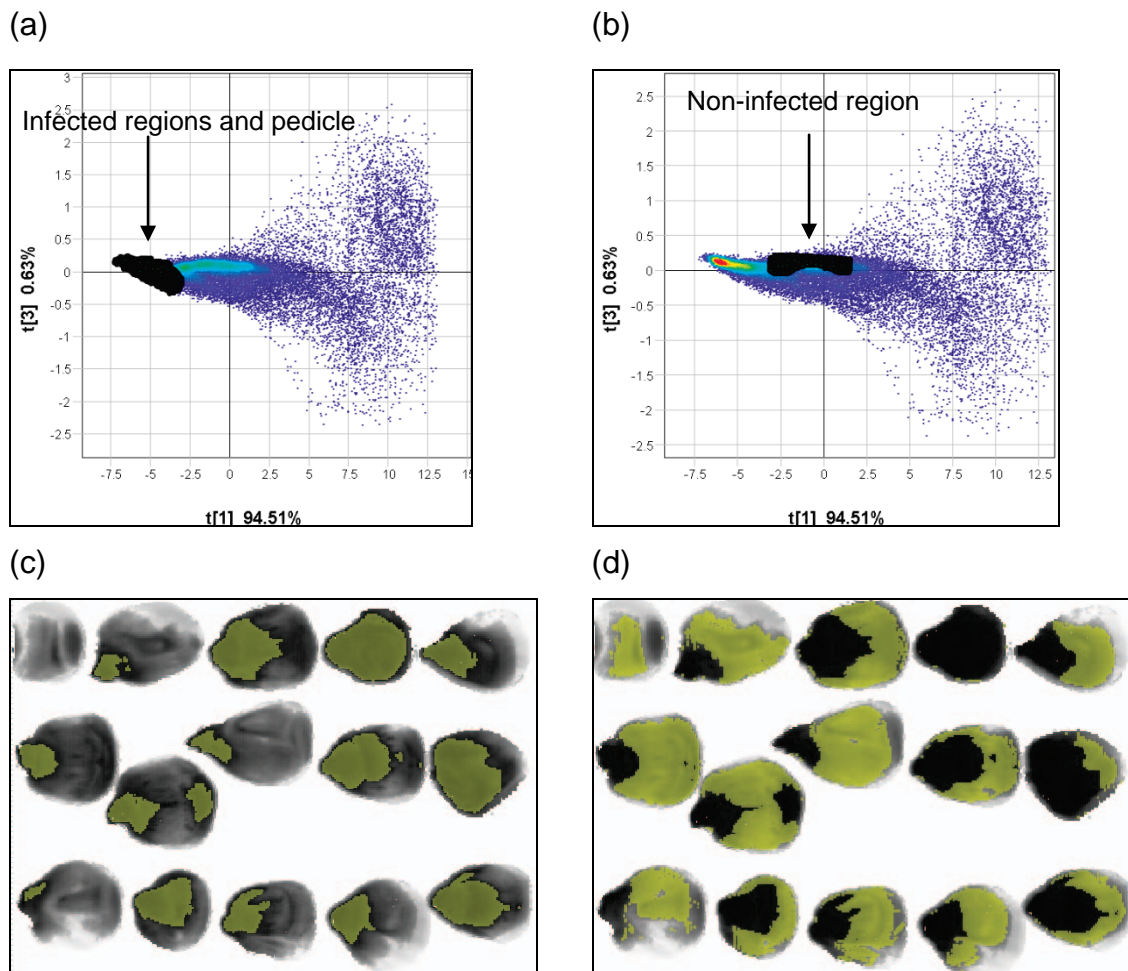
(a)   (b)

(c)   (d)

Figure 4.5 (a) PCA score plot of PC1 versus PC3 (MatrixNIR 15 kernels) with cluster associated with infected regions selected; and (b) with cluster associated with non-infected regions selected; (c) corresponding score image of (a) and (d), the corresponding score image of (b) showing the infected and non-infected regions, respectively (yellow = infected and non-infected regions).

To establish if this variation was due to the absence or presence of starch it was necessary to study the loading line plots. For that reason each class was subjected to PCA individually and the respective loading line plots interpreted based on chemical bonds present. PC1 for both infected (90.62% of SS) and non-infected (87.22% of SS) regions showed a typical grain spectrum (**Fig. 4.6a**). Although it might be an averaged spectrum it was this component that was responsible for the separation of the two clusters and was thus considered as being imperative in the analysis of loading line plots. Absorption peaks at 1215 nm (C-H stretching vibrations, second overtone) associated with CH_2 which in turn is related to starch and 1450 nm (O-H stretch, first overtone) associated with starch and (or) H_2O (Osborne *et al.*, 1993). This illustrates that the variation in PC1 is due to starch and moisture. In the infected kernels it would be expected to have a degree of moisture since it is needed for fungal proliferation (Bacon *et al.*, 1992; Bacon *et al.*, 2001).

The loading line plot for PC2 (**Fig. 4.6b**) for both infected and non-infected was similar to that of PC1; however in addition to the starch and moisture peaks it also exhibited absorption peaks at 1500 nm (N-H stretch, first overtone) for the infected region and 1580/1584 nm for the non-infected region (O-H stretch, first overtone, intermolecular hydrogen bond). The N-H stretch is associated with an NH structure present in proteins while the O-H stretch is associated with carbohydrates, i.e. starch (Juhász *et al.*, 2007). These components are the basic constituents of maize (Watson, 1987) and variation owing to these components was expected. Nevertheless, it was evident that starch in addition to protein played an important role as the contributing factor to the variation within the component. PC 2 for the infected region contributed 8.72% to the SS while the non-infected part contributed 11.46%.

The variation in PC3, (**Fig. 4.6c**) for the non-infected region (1.08% of SS) is predominantly due to a starch-moisture contrast. Absorption peaks at 1215 nm (C-H stretching vibrations, second overtone) associated with CH_2 which in turn is related to starch, 1395 nm (2xC-H stretch + C-H deformation) also associated with CH_2 and 1450 nm (O-H stretch, first overtone) associated with starch and/or water (Osborne *et al.*, 1993) were found. The infected region contributed 0.37% to the total SS and absorption peaks at 1215 (C-H stretching vibrations, second overtone) associated with CH_2 most likely attribute to starch and 1450 nm (O-H stretch, first overtone) associated with starch and (or) H_2O were found. The two loadings, although similar, differed by the absorption peak at 1450 nm. The loading for the infected region showed a lower absorption than the non-infected region. This could be ascribed to the fact that fungal-infected kernels would scatter more light than non-infected kernels, since the invasion of the fungus can cause

the kernel endosperm to become porous (Hesseltine & Shotwell, 1973). This scattering would cause less NIR radiation to be absorbed in reflectance mode (Pearson & Wicklow, 2006).
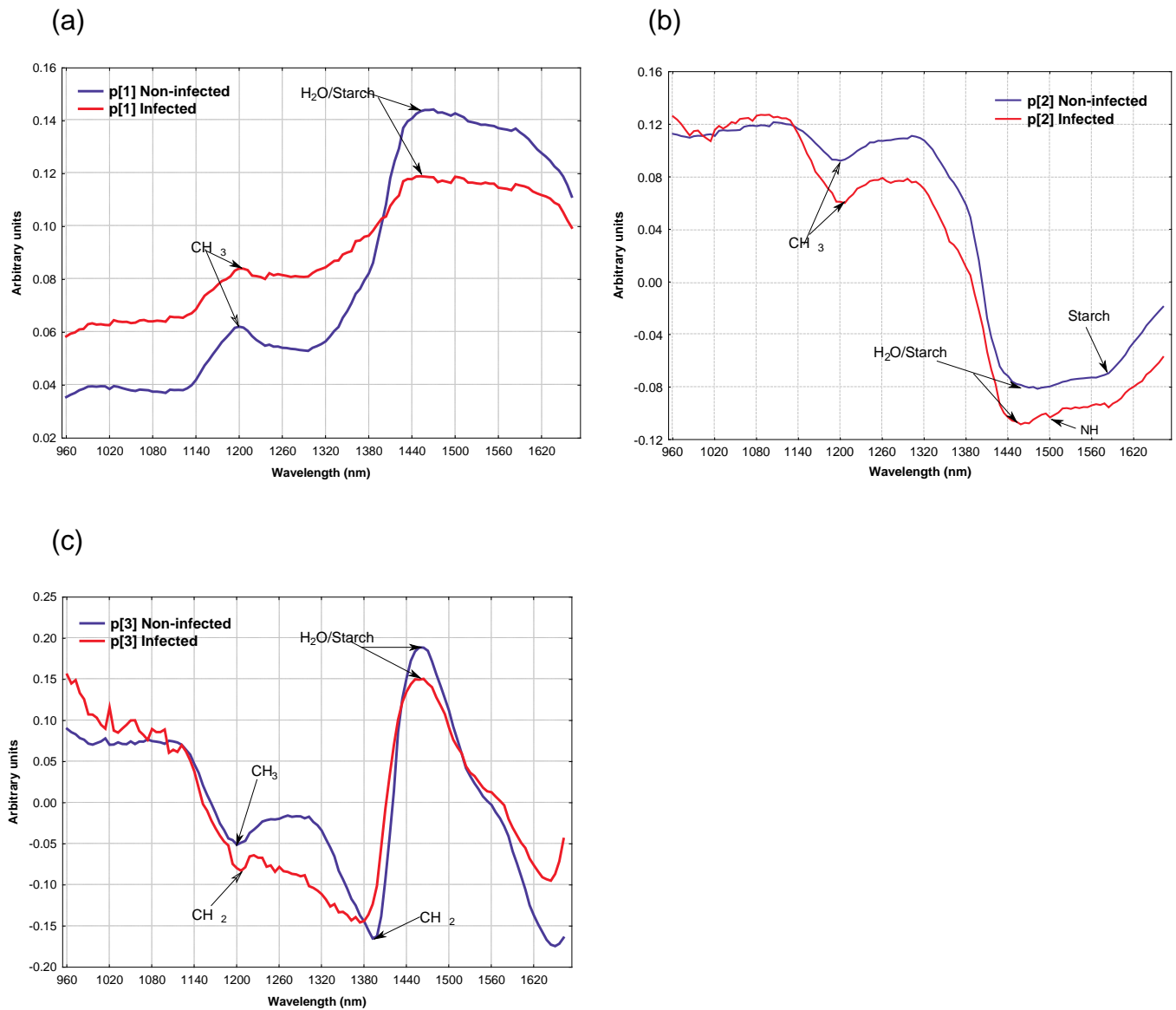
(a)

(b)

(c)

Figure 4.6 Loading line plots of MatrixNIR for (a) PC1, (b) PC2, (c) PC3 for both infected and non-infected regions with PC1 portraying typical grain spectra with absorption peaks at 1215 (starch) and 1450 (moisture/starch), PC2 absorption peaks at 1215 (starch), 1450 (moisture/starch) nm for both infected and non-infected and 1580 (starch) nm for non-infected and 1500 (NH) nm; PC3 a starch-moisture contrast (1395 and 1450 nm) for non-infected and a starch peak for infected at 1215 nm.

Hyperspectral image analysis – sisuChema (15 kernels)

Image cleaning

As with the MatrixNIR image, PCA with mean-centering was carried out and three components were calculated for a total of 99.30% of the total sum of squares (SS). The score plot and score image (**Figs. 4.7a & b**) was composed of 165 440 pixels. As described in Chapter 3, the interactive nature of the score plot and the score image made it possible not only to identify irrelevant pixels but also assign them to classes. This was depicted in the classified score plot and classification image (**Figs. 4.7c & d**). It was evident in these plots that there were differences between the data obtained from the MatrixNIR and the sisuChema. The score plot (**Fig. 4.7a**), although plotted against a different component, illustrated differences in the separation of clusters. The clusters seemed more apparent with the sisuChema generated image, and the selection and removal of classes was much easier.

Image analysis of cleaned images

Identification of relevant clusters and pixels were possible, once the irrelevant clusters had been removed. The cleaned score plot and score image (**Figs. 4.8a & b**) now comprised 23 959 pixels and after three additional principal components had been calculated, 99.96% of the total SS was accounted for by the model. The score plot (**Fig. 4.8a**) of PC1 versus PC3 showed a separation of the data swarm into two clusters. When the cluster in the left half of the score plot (**Fig. 4.8c**) was selected, the infected regions and pedicle was highlighted in the score image (**Fig. 4.8d**). The same was true for the cluster in the right half of the score plot; however, this time when selected in the score plot (**Fig. 4.8e**), the non-infected kernels and regions were highlighted in the corresponding score image (**Fig. 4.8f**). These findings were in agreement with the results obtained from the MatrixNIR, thus confirming that it was possible to distinguish between infected and non-infected kernels. In addition it was possible to determine which kernels were partially infected or totally infected. Even though asymptomatic kernels presented no visual confirmation of infection (Munkvold & Desjardins, 1997; Fandohan *et al.*, 2003), they were detected as infected in the score image (**Fig. 4.8d**).
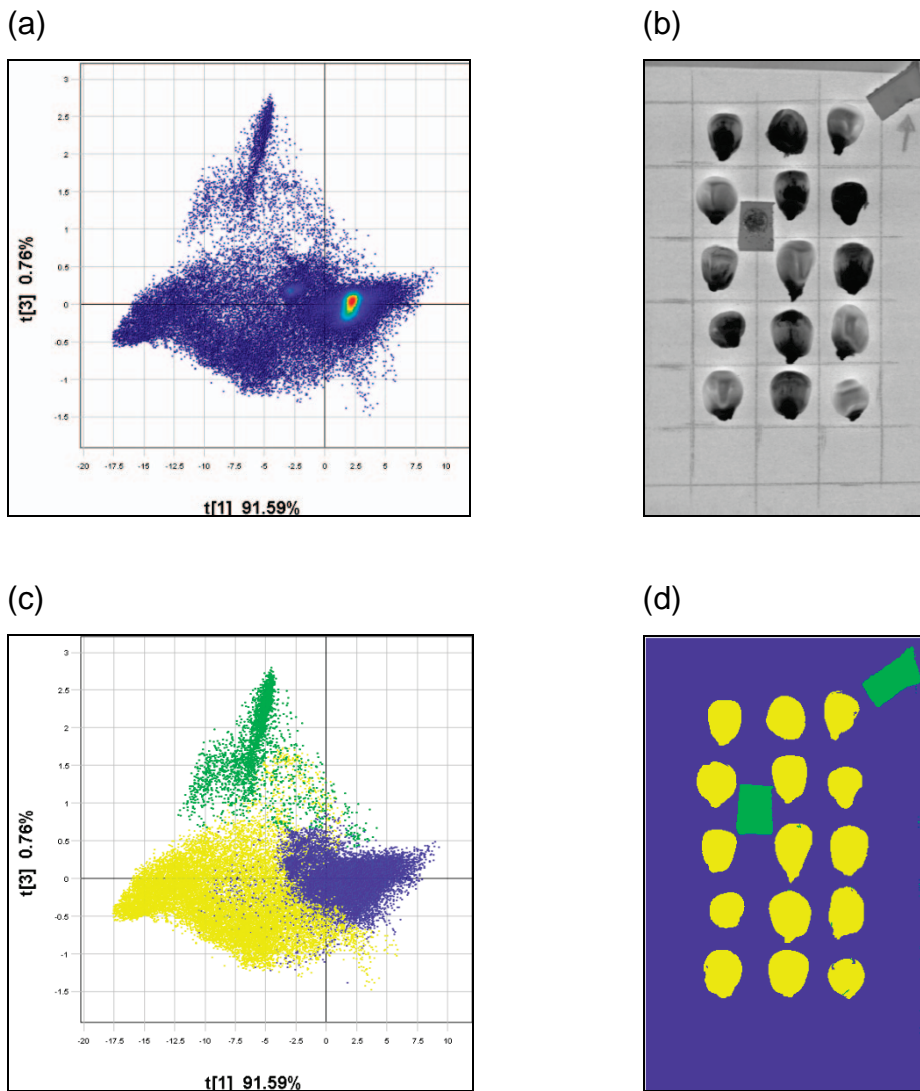
(a)

(b)

(c)

(d)

Figure 4.7 (a) PCA score plot of uncleaned image (PC1 versus PC3) obtained with the sisuChema (15 kernels) (165 440 pixels); (b) score image of raw absorbance image (PC1) used to identify and locate unwanted pixels; (c) PCA score plot with selected clusters projected onto (b) to obtain (d) classification image (blue = sandpaper, bad pixels and edge effects; yellow = maize kernels; green = plastic marker).
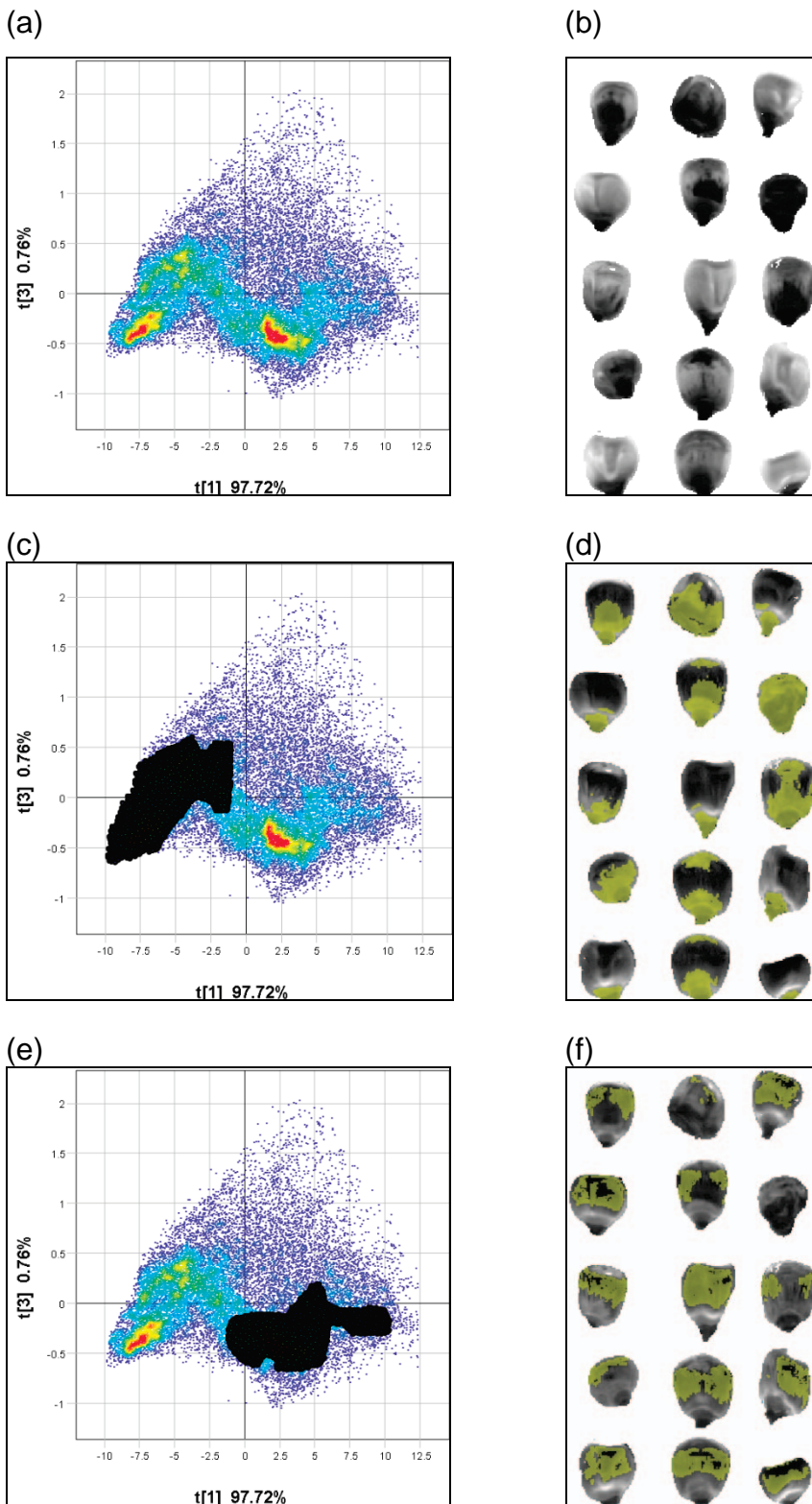
Figure 4.8(a) Cleaned PCA score plot of PC1 versus PC3; and (b) cleaned score image (PC1) for sisuChema (15 kernels) (23 959 pixels); (c) PCA score plot with cluster associated with infected regions selected; and (d) the corresponding score image of (c) with cluster associated with infected regions and pedicle highlighted; (e) PCA score plot with cluster associated with non-infected regions selected; and (f) corresponding score image of (e) with non-infected region highlighted.

As with the analyses of the MatrixNIR 15-kernel image, the two clusters were extracted from the image, subjected to PCA individually and the loading line plots (**Fig. 4.9a-c**) scrutinised. PC1 for both infected (99.89% of the total SS) and non-infected (89.84% of the total SS) regions showed a typical grain spectrum, similar to that of the MatrixNIR 15-kernel image. However, because of the longer wavelength capability of the instrument additional absorption peaks were present, thus more information was present that would otherwise not be detected with the MatrixNIR. Like the previous PC loading line analyses, PC1 was responsible for the separation of the clusters and was considered as crucial. Absorption peaks at 1215 nm (C-H stretching vibrations, second overtone) associated with CH_2 most likely attribute to starch, 1480 nm (O-H stretch first overtone, intramolecular hydrogen-bond) associated with glucose and 1940 nm (O-H stretch + O-H deformation) associated with H_2O (Osborne *et al.*, 1993) were present. These findings suggest that the variation in PC1 for both the infected and non-infected regions is due to starch and H_2O. Glucose is also responsible for the variation for the infected region. The fact that glucose contributes to the variation for the infected region could signal the hydrolysis of starch by the fungus to simpler compounds.

Absorption peaks at 1690 nm (C-H stretch first overtone) associated with starch (Pearson & Wicklow, 2006) and 1940 nm (O-H stretch + O-H deformation) associated with H_2O (Osborne *et al.*, 1993) were responsible for the variation in PC3 for the infected region. For the non-infected region the variation was due to an absorption peak at 1940 nm (O-H stretch + O-H deformation) associated with water (Osborne *et al.*, 1993).

From these observations it was lucid the variation in both, infected and non-infected regions was primarily due to starch and moisture, the basic constituents of maize kernels. The presence of these components are also necessary for the proliferation of the fungus (Bacon *et al.*, 1992; Munkvold & Desjardins, 1997) and could be used as a means of separating infected from non-infected kernels.

The analyses of the images from the both the MatrixNIR and the sisuChema made it possible, firsthand, to observe the differences between these instruments. The data obtained from the sisuChema had less detector error than that of the MatrixNIR most likely because it had a more stable detector. Furthermore, the sisuChema had more wavelengths which led to the detection of additional peaks associated with moisture, oil, protein and starch.
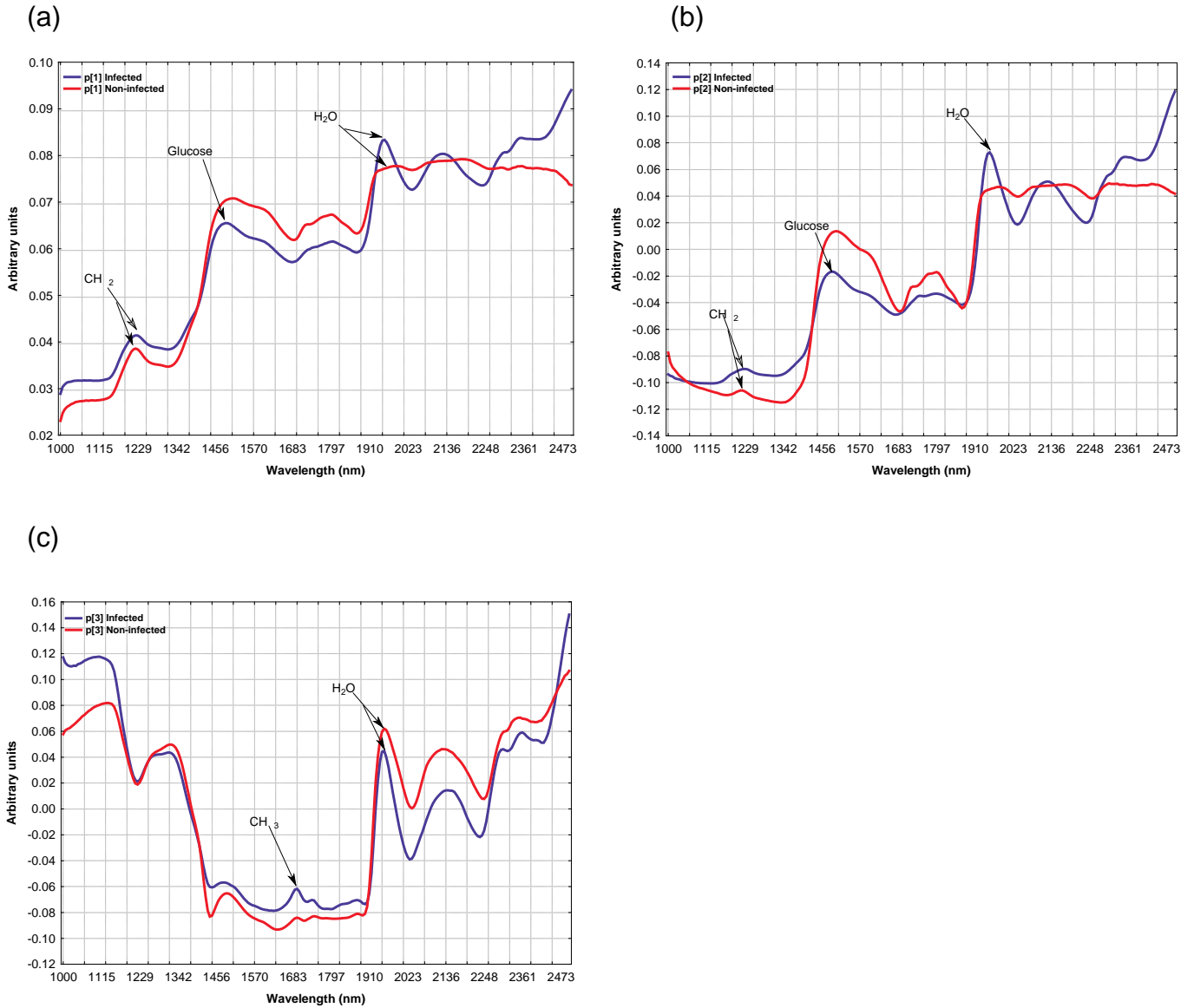
(a)

(b)

(c)

Figure 4.9 Loading line plots of sisuChema (15-kernels) for (a) PC1, (b) PC2, and (c) PC3 for both infected and non-infected regions with PC1 portraying typical grain spectra with absorption peaks at 1215 nm (starch) and 1480 nm (glucose) for infected only, PC2 absorption peaks at 1215 (starch), 1480 (glucose) nm for both infected and non-infected and 1940 (H_2O) nm for non-infected only; PC3 absorption peaks at 1690 nm (starch) for infected and 1940 nm (H_2O) for both non-infected and infected.

Hyperspectral image analysis – sisuChema (21 kernels)

<u>Image cleaning</u>

Image cleaning was performed as the sisuChema image. The score plot was similar to the previous sisuChema generated image and is not shown. The image comprised 165 120 pixels and three components contributed 99.63% to the total SS.

<u>Image analysis of cleaned images</u>

Image analysis was consistent with previous analyses and similar results were obtained. The PCA score plot (**Fig. 4.10a**) of PC1 (97.21%) versus PC2 (1.93%) showed a clear separation of the data swarm. Two big clusters comprising two smaller clusters each were apparent. When selected in the score plot (**Figs. 4.10c & e**), the two clusters were associated with non-infected region and infected region and pedicle, in the score image (**Figs. 4.10d & f**), respectively. The smaller clusters, when selected, showed some of the same information as the big clusters. It was thus concluded that these clusters had to be selected as a whole for each half of the plot.

Figure 4.10 (a) Cleaned PCA score plot of sisuChema (21 kernels) for PC1 versus PC2; and (b) cleaned score image of PC1 (45 550 pixels); (c) PCA score plot with cluster associated with non-infected regions selected; and (d) the corresponding score image of (c) with cluster associated with non-infected regions; (e) PCA score plot with cluster associated with infected regions selected; and (f) corresponding score image of (e) with corresponding region highlighted

The data obtained from the analysis of the sisuChema 21-kernel image was compared with the microbial analysis as well as background information concerning the kernels (**Table 4.1**). Typical growth was indicated with the presence of peach, pale cream, whitish to grey and brown colonies (**Fig. 4.11a**) and was identified microscopically by the presence of conidiophores, micro- and macroconidia (**Figs. 4.11b & c**). **Figure 4.11d** represents an agar plate with no growth. The microbiological and NIR hyperspectral imaging analysis of kernels 1, 2, 4, 5, 7, 10, 11, 12, 14, 15, 16, and 19 were in agreement with the data supplied from plant pathology. These kernels were either infected or non-infected and were thus easy to discriminate because the fungus was either present or not. Kernels 3, 6, 8, 13 and 21 were identified as asymptomatic kernels by the plant pathologists. These kernels were infected with the fungus but present no visible sign of infection (Bacon *et al.*, 1992; Munkvold & Desjardins, 1997; Oren *et al.*, 2003). According to the microbiological analysis, kernels 3, 6 and 21 were not infected and no growth was detected on the petri dishes, thus the kernels were either misclassified prior to the experiment or, the fungus was present but in very low numbers and were not detected on the plates. These kernels were also interpreted as not infected from the NIR hyperspectral imaging data, thus confirming that this technique can be used to distinguish between infected and non-infected kernels. Kernels 9, 17, 18 and 20 were identified as not infected by the plant pathologists but had growth present on the agar plates. Kernel 9 exhibited growth on the agar plate due to contamination since the growth was not from the centre of the plate where the initial inoculation was made. The remaining kernels were either misclassified or growth was present due to contamination. These kernels were thus incorrectly identified by NIR hyperspectral imaging.

Table 4.1 Comparison of data supplied by plant pathologists with data obtained from microbial analysis and NIR hyperspectral imaging for each kernel of the sisuChema 21-kernel image.

Kernel	Plant Pathology	Micro-analysis	NIR Hyperspectral imaging
1	N	N	N
2	I	I	I
3	A	N	N
4	I	I	I
5	I	I	I
6	A	N	N
7	N	N	N
8	A	I	N
9	N	I	N
10	I	I	I
11	N	N	N
12	N	N	N
13	A	I	I
14	I	I	I
15	I	I	I
16	I	I	I
17	N	I	N
18	N	I	N
19	I	I	I
20	N	I	N
21	A	N	N

N = Non-infected
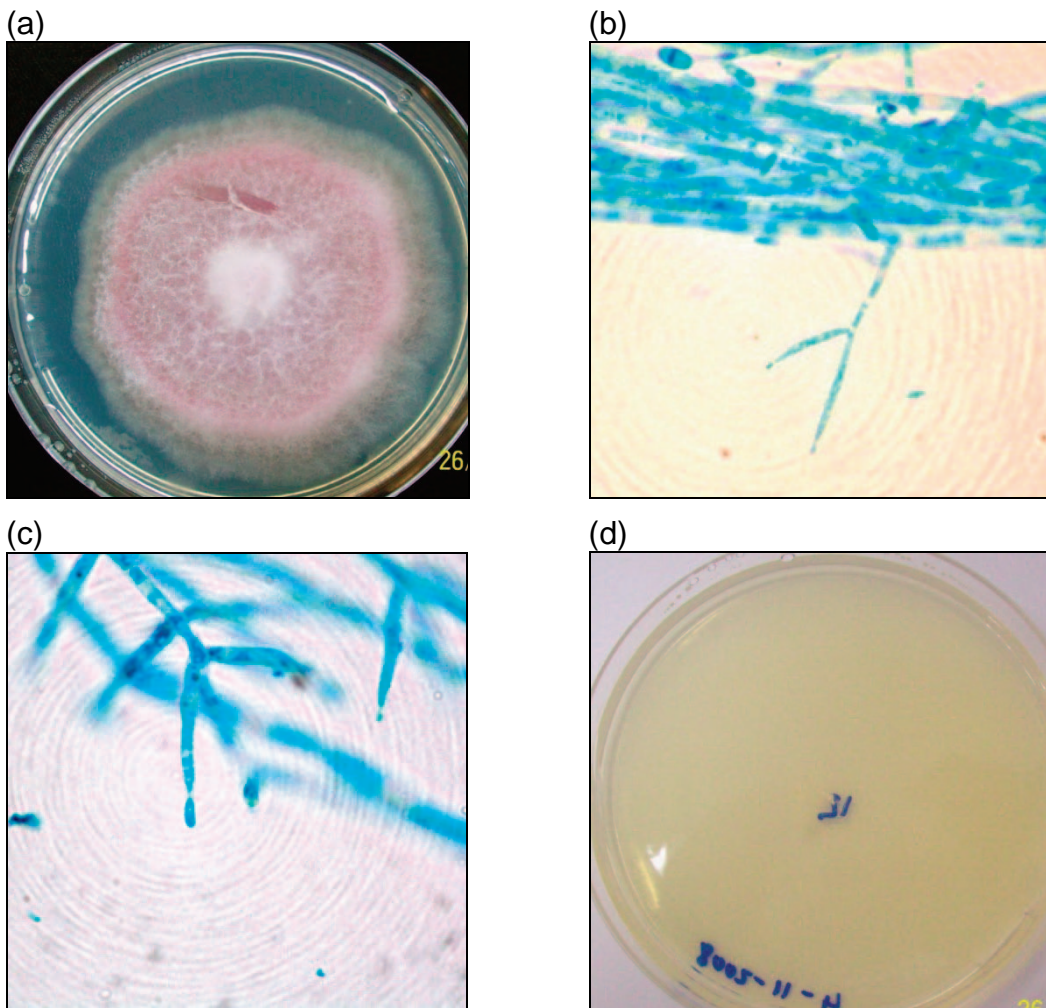A = Asymptomatic
I = Infected

Figure 4.10 (a) Petri dish showing typical growth of *Fusarium verticillioides;* and (b) microscope image of conidiophores x100; and (c) microscope image of conidiophores x1000; and (d) petri dish showing no growth;.

PCA loading line plots were also evaluated (**Fig. 4.12a-c**). Unlike the previous analyses of loading line plots, the two classes, i.e. infected and non-infected, from the sisuChema 21-kernel image were not extracted and subjected to individual PCA. Similar results were obtained as with the previous analyses. PC1 accounted for much of the variation. The source of the variation was, as expected, the basic components of maize kernels, i.e. starch, moisture, and protein. Absorption peaks at 1215 nm (C-H stretching vibrations, second overtone) probably corresponding to starch, 1500 nm (N-H stretch, first overtone) corresponding to protein, 1790 nm (C-H stretch, first overtone) corresponding to cellulose and 1940 nm (O-H stretch, first overtone and O-H deformation) corresponding to H_2O were found. The variation in PC2 was due to the same absorption peaks as PC1; this was expected as these bonds are predominant in the maize kernel and would thus account for most of the variation. The loading line plots for PC3 and PC4 were similar in that both accounted for variation mainly due to the presence of starch and to a lesser extent protein.

Absorption peaks at 1215 nm (C-H stretching vibrations, second overtone) and 2336 nm (C-H stretch and C-H deformation) were associated with starch and 1500 nm (N-H stretch, first overtone) and 2136 nm (N-H stretch and C=O stretch) corresponded with protein. This showed that, in terms of loadings, with 15 or 21 kernels the same data was generated with the sisuChema. However, when extracting the classes and applying PCA to the smaller data sets noisy loadings were often observed in the higher components, especially with the MatrixNIR. This could be attributed to the instability of the monochromator as pointed out in Chapter 3. There did not seem to be differences between the loadings of the extracted regions versus the loadings of all the kernels on the sisuChema. Nevertheless, it was apparent that differences existed between the loadings of the MatrixNIR and the sisuChema. Most prominent was the additional absorption peaks found in the sisuChema loading line plots due to the longer wavelength capability. This was useful for identifying chemical components that would not be detected in the region of the MatrixNIR, however it added to the complexity of interpreting the loading line plots.
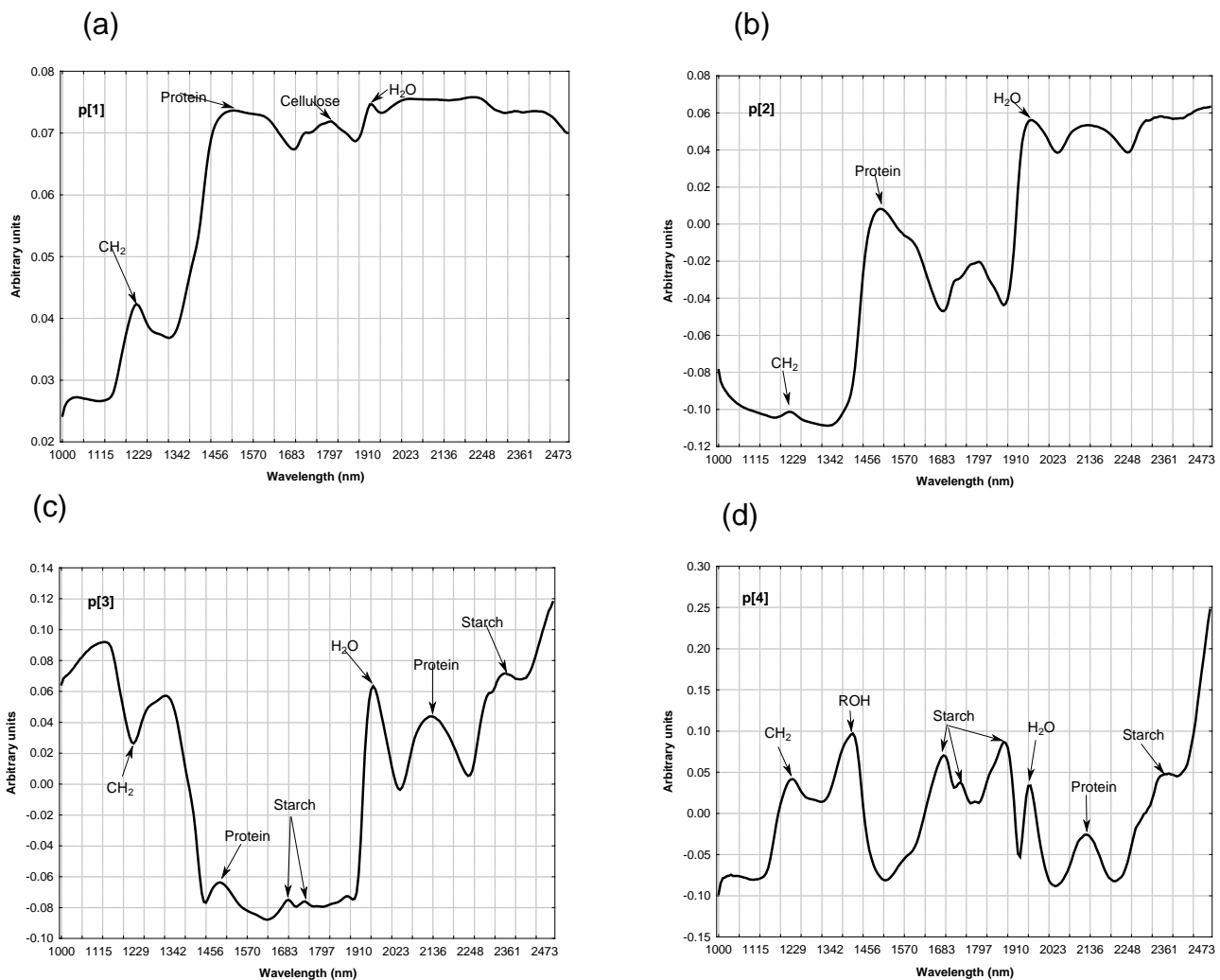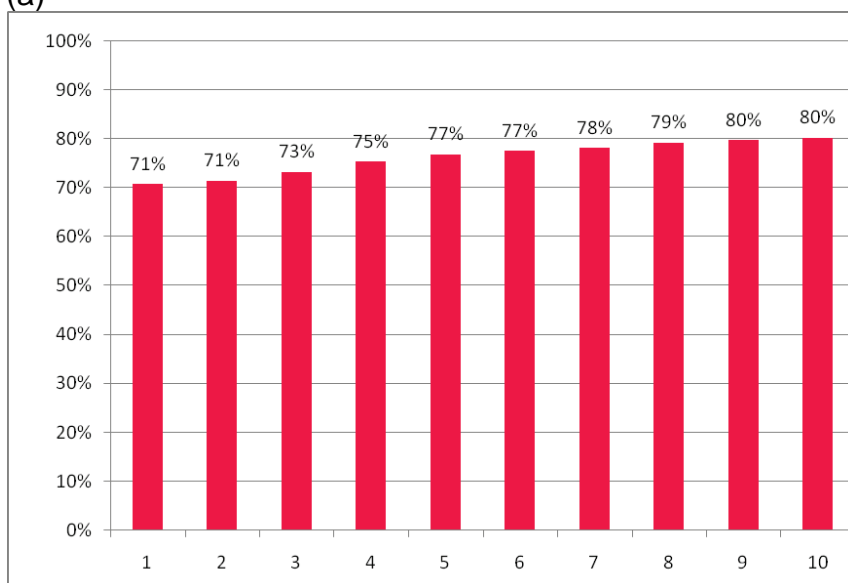
(a)

(b)

(c)

(d)

Figure 4.12 Loading line plots of (a) PC1, (b) PC2, (c) PC3 and (d) PC4 of whole image for sisuChema (21 kernels).

Partial least squares-discriminant analysis - MatrixNIR and sisuChema (15 kernels)

PLS-DA models of both the MatrixNIR and sisuChema generated images yielded satisfactory results. Both models (**Figs. 4.13a & b**) were capable of modelling more than 75% of the **Y** variation, an indication that the model would perform well. After 10 PLS components the MatrixNIR image modelled 80% of the **Y** variation and the sisuChema model account for 88% **Y** variation. Already after two PLS components, the sisuChema model explained 85% of the total **Y** variation, there was thus not a substantial increase with the addition of eight components and could thus lead to over fitting and possible inclusion of noise. The calibration and prediction images after 3 PLS components for the both the MatrixNIR and the sisuChema (**Figs. 4.14a-d**) illustrate the localisation of the modelled and predicted classes, i.e. infected regions and pedicle and non-infected regions. The predictive ability of the models (**Tables 4.2 & 4.3**) for both instruments yielded classification rates of 96.1% for infected regions and 99.9% for non-infected regions on the MatrixNIR and 99.9% for infected regions and 95.5% for non-infected regions on the sisuChema. These results are more than satisfactory and compare favourably to those reported earlier (Berman *et al.*, 2007; Zhang *et al.*, 2007).
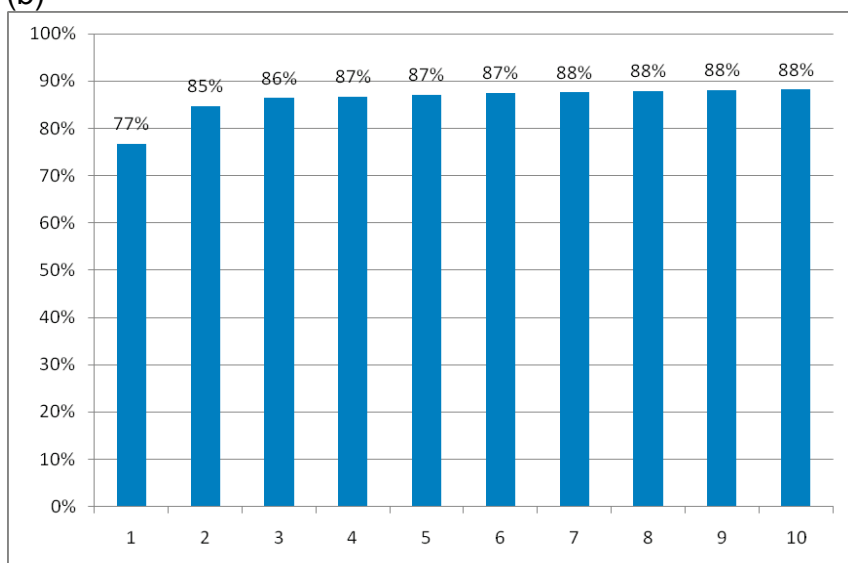
(a)

(b)

Figure 4.13 PLS-DA model overview for the 15 kernel image of both the MatrixNIR (a) and sisuChema (b) acquired images.

(a)

(b)

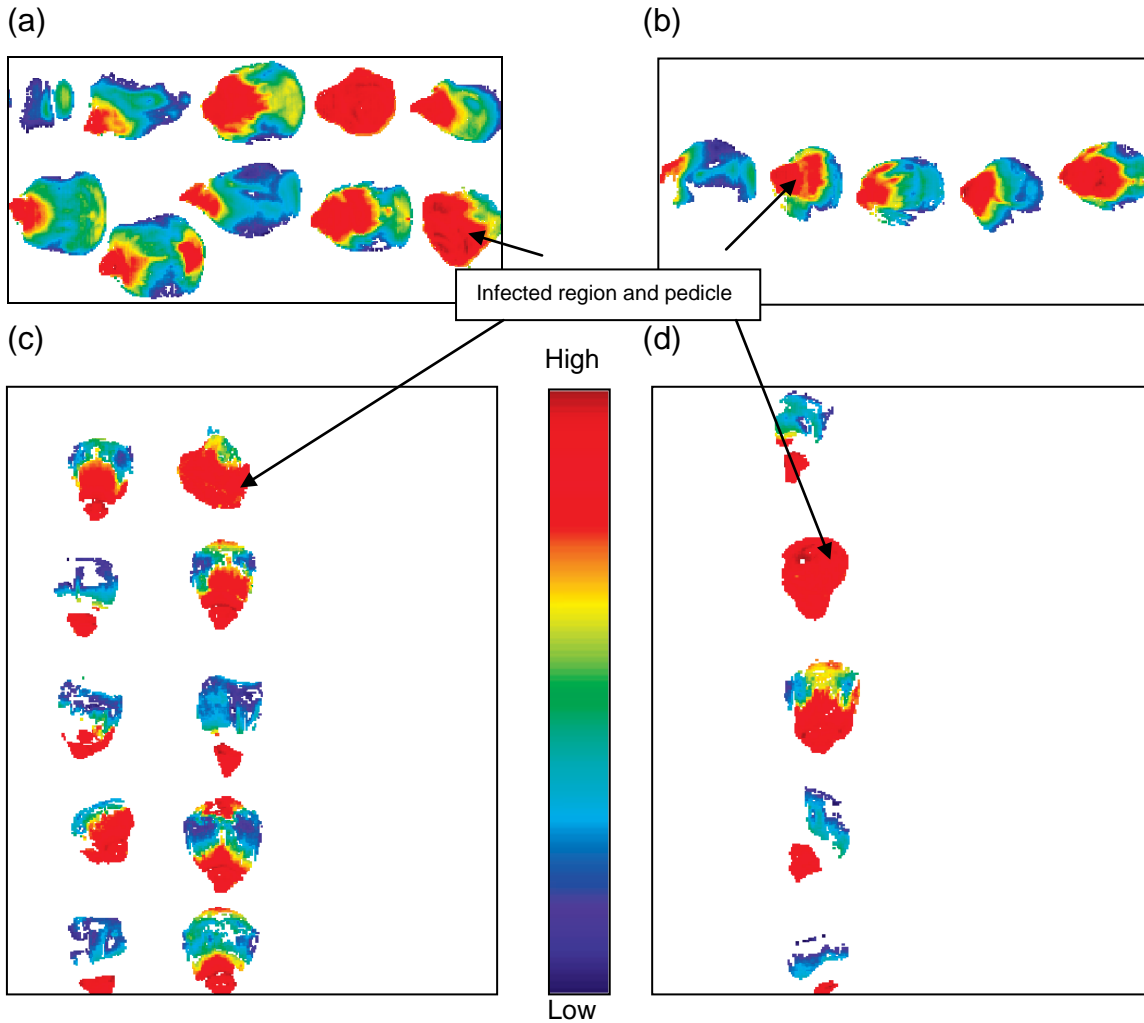Infected region and pedicle

(c)

High

(d)

Low

Figure 4.14 (a) Calibration image (a) and prediction (b) of MatrixNIR (15 kernels) for the infected region and pedicle after 3 PLS components; and calibration image (c) and prediction image (d) of sisuChema (15 kernels) after 3 PLS components for infected region. (red = infected region modelled well; blue = infected region not modelled well).

Table 4.2 PLS prediction performance indicating the classes, number of pixels predicted as each class and the prediction performance for infected and non-infected regions for the MatrixNIR

Classes	# Predicted	Infected	Non-infected
Infected	2779 (39.1%)	2672 (96.1%)	107 (3.9%)
Non-infected	4321 (60.9%)	3 (0.1%)	4318 (99.9%)
Total	7100(100%)	2675 (37.7%)	4425 (62.3%)

Infected = infected regions and pedicle; Non-infected = non-infected regions

Table 4.3 PLS prediction performance indicating the classes, number of pixels predicted as each class and the prediction performance for infected and non-infected regions for the sisuChema

Classes	# Predicted	Infected	Non-infected
Infected	2805 (66.2%)	2802 (99.9%)	3 (0.1%)
Non-infected	1429 (33.8%)	65 (4.5%)	1364 (95.5%)
Total	4234 (100%)	2675 (37.7%)	1367 (32.3%)

Infected = infected regions and pedicle; Non-infected = non-infected regions

Partial least squares-discriminant analysis - sisuChema (21 kernel)

The PLS-DA models for the various forms of pre-processing gave similar results. The model overview (**Fig. 4.15**) shows that there was no substantial difference between, no pre-processing, MSC and SNV. After 10 PLS components 84.5%, 84.6% and 85% of **Y** was modelled for each one respectively. Already after three PLS components more than 75% of **Y** was accounted for. To further illustrate that pre-processing made no substantial improvement in the model, the classification rates are shown (**Table 4.4**). Without any pre-processing, a classification rate of up to 97.6% for infected was achieved, 97.7% for MSC and 97.4% for SNV. The reason for no considerable differences between the pre-processing techniques is most likely ascribable to efficient and even illumination in the sisuChema system since only a single line of the sample is illuminated. These attributes, when negatively impacting imaging, could result in images that would necessitate pre-processing prior to analyses. This was not the case for these images. In addition, removal of irrelevant pixels was most probably done in such a way that most of these effects, i.e. shading had been accounted for.
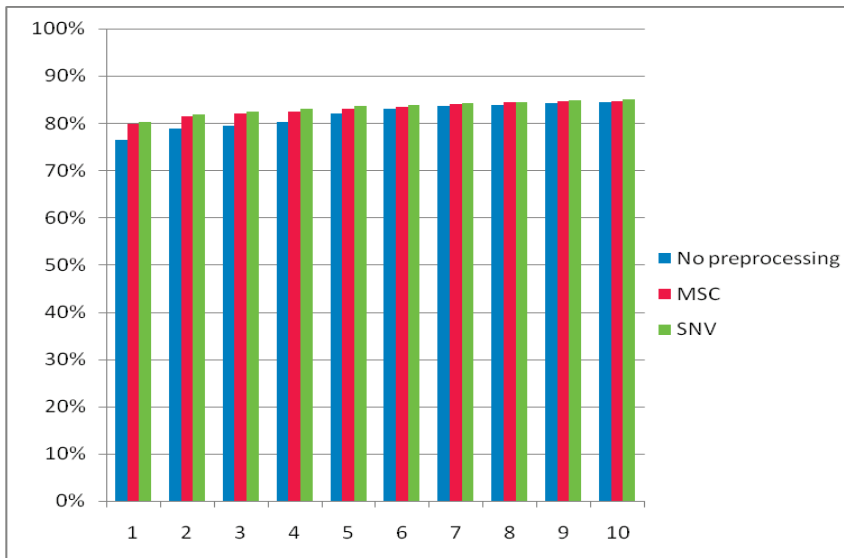
Figure 4.15 PLS-DA model overview for sisuChema (21 kernels) illustrating the **Y** percentage SS explained after 10 PLS components with no pre-processing, MSC and SNV applied.

Table 4.4 An overview of the classification rates of the PLS-DA models with various pre-processing techniques applied

Classes	Pre-processing	Classification rate (%)
Infected	No pre-processing	97.6
	MSC	97.7
	SNV	97.4
Non-infected	No pre-processing	99.2
	MSC	94.0
	SNV	94.2

MSC = multiplicative scatter correction
SNV = standard normal variate

Conclusion

Although the loading line plot of PC1 is generally considered to be an average spectrum of the image being analysed, it cannot be ignored or overlooked as is often the case. It does represent an average spectrum when compared with the rest of the loading line plots; however, the variation it accounts for is substantial. This is, nevertheless, dependant on the image being analysed. The infection of maize kernels with fungi is often signalled by physical as well as chemical changes. The chemical changes may be subtle but severe physical changes are manifested in extreme cases of infection with visible signs of infection.

NIR hyperspectral imaging can be used as a means of detecting infected kernels with a classification rate of up to 97%. PC1 for both the MatrixNIR and sisuChema 15-kernel

image as well as the sisuChema 21-kernel image showed considerable separation of the data into two distinct clusters that were eventually identified as infected and non-infected regions. PC loading line plots showed that most of the variation can be ascribed to starch. This was expected since the fungus uses the starch as a carbon source.

Although this work may still be in the preliminary stages, it shows that it is possible to distinguish between infected and sound kernels. Further studies should include the infection of kernels and acquiring NIR hyperspectral images each day after incubation. This will allow for the study of the progression of the fungus within in the kernel. The reason for the pedicle always being classified as infected should also be investigated.

References

Bacon, C.W., Bennett, R.M., Hinton, D.M. & Voss, K.A. (1992). Scanning electron microscopy of *Fusarium moniliforme* within asymptomatic corn kernels and kernels associated with leukoencephalomalacia. *Plant Disease*, **76**, 144-148.

Bacon, C.W., Yates, I.E., Hinton, D.M. & Meredith, F. (2001). Biological control of *Fusarium moniliforme* in maize. *Environmental Health Perspectives Supplements*, **109**, 325-332.

Berman, M., Connor, P.M., Whitbourn, L.B., Coward, D.A., Osborne, B.G. & Southan, M.D. (2007). Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis. *Journal of Near Infrared Spectroscopy*, **15**, 351-358.

Bluhm, B.H. & Woloshuk, C.P. (2005). Amylopectin induces Fumonisin B1 production by *Fusarium verticillioides* during colonization of maize kernels. *Molecular Plant-Microbe Interactions*, **18**, 1333-1339.

Burger, J. (2006). Hyperspectral NIR image analysis: data exploration, correction and regression. PhD Thesis Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, Umeå, Sweden

Burger, J. & Geladi, P. (2005). Hyperspectral NIR image regression part I: calibration and correction. *Journal of Chemometrics*, **19**, 355-363.

Burger, J. & Geladi, P. (2006). Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst*, **131**, 1152-1160.

Bush, B.J., Carson, M.L., Cubeta, M.A., Hagler, W.M. & Payne, G.A. (2004). Infection and Fumonisin roduction by *Fusarium verticillioides* in developing maize kernels. *Phytopathology*, **94**, 88-93.

Cogdill, R.P., Hurburgh, C.R. & Rippke, G.R. (2004). Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE*, **47**, 311-320.

Desjardins, A.E. (2006). Selected Mycotoxigenic *Fusarium* Species. In: *Fusarium Mycotoxins: Chemistry, Genetics and Biology*. Pp. 192-194. St. Paul, Minnesota U.S.A: The American Phytopathological Society.

Esbensen, K.H. & Lied, T.T. (2007). Principles of multivariate image analysis (MIA) in remote sensing, technology and industry. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P. Geladi). Pp. 17-41. Chichester, West Sussex: John Wiley & Sons, Ltd.

Fandohan, P., Gnonlonfin, B., Hell, K., Marasas, W.F.O. & Wingfield, M.J. (2005). Natural occurrence of Fusarium and subsequent fumonisin contamination in preharvest and stored maize in Benin, West Africa. *International Journal of Food Microbiology*, **99**, 173-183.

Fandohan, P., Hell, K., Marasas, W. & Wingfield, M. (2003). Infection of maize by Fusarium species and contamination with fumonisin in Africa. *African Journal of Biotechnology*, **2**, 570-579.

Geladi, P., Grahn, H.F. & Burger, J. (2007). Multivariate images, hyperspectral imaging: background and equipment. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P. Geladi). Pp. 1-14. Chichester, West Sussex: John Wiley & Sons Ltd.

Gelderblom, W.C.A., Rheeder, J.P., Leggott, N., Stockenstrom, S., Humphreys, J., Shephard, G.S. & Marasas, W.F.O. (2004). Fumonisin contamination of a corn sample associated with the induction of hepatocarcinogenesis in rats-role of dietary deficiencies. *Food and Chemical Toxicology*, **42**, 471-479.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J. & Bell, S.E.J. (2008). Recent applications of chemical iImaging to pharmaceutical process monitoring and quality control. *European Journal of Pharmaceutics and Biopharmaceutics*, **69**, 10-22.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G. & Frias, J.M. (2007). Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, **18**, 590-598.

Hesseltine, C.W. & Shotwell, O. (1973). New methods for rapid detection of aflatoxin. *Pure and Applied Chemistry*, **35**, 259-266.

Juhász, R., Gergely, S., Szabóki, Á. & Salgó, A. (2007). Correlation between NIR spectra and RVA parameters during germination of maize. *Cereal Chemistry*, **84**, 97.

Marasas, W.F.O. (2001). Discovery and occurence of the Fumonisins: a historical perspective. *Environmental Health Perspectives*, **109**, 239-243.

Marin, S., Magan, N., Serra, J., Ramos, A., Canela, R. & Sanchis, V. (1999). Fumonisin B1 production and growth of *Fusarium moniliforme* and *Fusarium proliferatum* on maize, wheat, and barley grain. *Journal of Food Science*, **64**, 921-924.

Munkvold, G.P. & Desjardins, A.E. (1997). Fumonisins in maize: can we reduce their occurrence? *Plant Disease*, **81**, 556-565.

Oren, L., Ezrati, S., Cohen, D. & Sharon, A. (2003). Early events in the *Fusarium verticillioides*-maize interaction characterized by using a green fluorescent protein-expressing transgenic isolate. *Applied and Environmental Microbiology*, **69**, 1695-1701.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis.* Pp. 29-33. Essex, England: Longman Scientific & Technical.

Pearson, T.C. & Wicklow, D.T. (2006). Detection of corn kernels infected by fungi. *Transactions of the ASABE*, **49**, 1235-1245.

Samson, R.A. (2000). *Introduction to food- and airborne fungi.* Utrecht: Centraalbureau voor Schimmelcultures.

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

Watson, S.A. (1987). Structure and composition. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 53-82. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Yao, H., Hruska, Z., Kincaid, R., Brown, R. & Cleveland, T. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215-224.

Yates, I.E. & Sparks, D. (2008). *Fusarium verticillioides* dissemination among maize ears of field-grown plants. *Crop Protection*, **27**, 606-613.

Zhang, H., Paliwal, J., Jayas, D.S. & White, N.D.G. (2007). Classification of fungal infected wheat kernels using near-infrared reflectance hyperspectral imaging and support vector machine. *Transactions of the ASABE*, **50**, 1779-1785.

Chapter 5

General discussion and conclusion

Chapter 5

General discussion and conclusion

Near infrared (NIR) hyperspectral imaging is a rapid, non-destructive technique that is capable of capturing data regarding a sample in the form of images. Data are arranged in three-way matrices with two spatial pixel coordinates and one wavelength dimension. This allows for the identification and localisation of chemical components within samples that can be used to determine presence and concentration. In this research, NIR hyperspectral imaging has been explored for two applications in maize: firstly to assess maize hardness and variation within and between maize kernels, and secondly to assess the detection of a fungal disease with serious impacts on maize quality and food safety.

Maize hardness is a quality attribute very important to many players in the industry, especially millers. It is well-known that, more often than not, hard (glassy) endosperm maize is more desirable than soft (floury) endosperm for industry to produce products for human consumption (Alonso Ferro *et al.*, 2008). Thus, breeders need to develop hybrids with this trait. This is a lengthy process and non-destructive techniques to evaluate traits are beneficial. Vitreousness has been evaluated with NIR hyperspectral imaging in durum wheat (Gorretta *et al.*, 2006) and the determination of desirable and undesirable traits to assist plant breeding has been assessed (Smail *et al.*, 2006). The technique has to date, not yet been assessed on maize kernels for the determination of kernel hardness.

Maize kernels of varying degrees of hardness were imaged using a MatrixNIR (960-1662 nm) and sisuChema (1000-2498 nm) imaging systems. In this study principal component analysis (PCA) has been effectively used to distinguish between glassy and floury endosperm. Exploratory PCA is essential for image analysis, particularly hyperspectral image analysis as a dimension reduction method and as a means of image exploration. The interactivity of the score plot and score image in the Evince software illustrated this. This proves that unsupervised classification can be used to understand much concerning samples and can be used to effectively discriminate between classes.

The distinction between the glassy and floury endosperm was found primarily along PC3 for both the MatrixNIR and the sisuChema images. Analysis of the loading line plot for PC3 showed absorbance peaks at 1215, 1395 and 1450 nm associated with starch and moisture for both MatrixNIR (12 and 24 kernel) acquired images. The loading plot line for the sisuChema acquired image also had absorbance peaks at these wavelengths in addition to 1695, 1900 and 1940 nm. The first two peaks were associated with starch and

the latter with moisture. This was anticipated since maize kernels hardness is primarily due to differences in starch which in turn is interpreted as differences in starch-protein associations and ratios of glassy to floury endosperm present (Wolf *et al.*, 1952; Wolf *et al.*, 1969; Watson, 1987; Lee *et al.*, 2006). Although the variation along PC3 was attributed to starch, it is not known what exactly was contributing to these differences as illustrated using NIR hyperspectral imaging. This would have to be evaluated using a reference technique; however, there are no non-destructive reference techniques available for single-kernel hardness determination.

At present hardness measurements are carried out on bulk samples that give an average estimation of hardness even though, as is known, hardness varies within a single kernel and between kernels. NIR hyperspectral imaging has shown and reasserted that the current reference techniques have many shortcomings and requires revision. To determine precisely what NIR hyperspectral imaging is detecting would most likely require a combination of the reference techniques with imaging. This will involve acquiring images of kernels with a range of hardness, starch content and protein content, particularly with known zein content in addition to using scanning electron microscopy (SEM), then performing various reference techniques. The results could be correlated to determine exactly what is being detected using NIR hyperspectral imaging. Nevertheless with the current results obtained, if the method was to be used in industry, it would be recommended to install a system focused on 1215, 1395 and 1690 nm. This would greatly benefit industry as the routine measurements would not be so time-consuming, results could be obtained immediately and samples remain intact.

Even though it was possible to distinguish between hard and soft kernels, in order to use this information in industry, the results had to be statistically evaluated. For that reason, partial least squares-discriminant analysis (PLS-DA) was applied to determine whether good discrimination between glassy and floury endosperm could be estimated.

Partial least squares-discriminant analysis showed the potential of predicting future samples in terms of type of endosperm present. Classification rates for glassy endosperm up to 99% for the MatrixNIR (12 kernel image), 82% for the MatrixNIR (24 kernel image) and 92% for the sisuChema 24 (kernel image) were obtained. This demonstrates the ability of the method to be used as a means of predicting unknown samples; however could not have been validated independently, i.e. on a separate image due to current software limitations in the Evince software package. For this study prediction was not done on a new image, instead the original absorbance image was divided into a training

and test set. To validate the model completely independently, it has to be tested on an entire new image. This would give a true indication of the predictive ability of the model.

Fungal infected kernels are typically identified by the presence of visible fungal growth. Symptomatic infections include severe damage to kernels, seed quality loss and deformation of kernel shape and size and is most likely a result of a complex interaction of virulence of an isolate, maize variety and environment (Headrick *et al.*, 1990). In addition, the symptomatic infections manifest themselves as ear and kernel rot that are characterised by the presence of white, pink or salmon-coloured mould and brown, tan or white-streaked kernels. Asymptomatic infection, on the other hand, can exist throughout the plant in leaves, stems, roots, grains, and the presence of the fungus is in many cases ignored because it does not cause visible damage to the plant (Munkvold & Desjardins, 1997; Fandohan *et al.*, 2003; Jackson & Jablonski, 2004). This leaves room for much operator error as these means of identification could be subjective and would certainly differ between operators. Thus, an objective technique for the detection of fungal infected kernels is necessary.

With regards to the fungal infected maize kernels, NIR hyperspectral imaging has shown to be an excellent objective means of discriminating between these and sound kernels. Once again PCA has shown to be an excellent exploratory tool for hyperspectral image analysis, allowing for identification and location of infected as well as non-infected regions in the maize kernels. Even though infected kernels had regions that were identified as non-infected, this would still signal that the kernel is tainted and not suitable for human or animal consumption. This most likely indicated the region of the kernel that indeed was infected or the possible location of the fungal material, i.e. hypha. Additionally this could be indicating the location of the origin of the infection.

PLS-DA showed excellent classification rates, 96.1% and 99% for the MatrixNIR (15 kernel) and sisuChema (15 kernel) image and the sisuChema (21 kernel) image had a classification rate for infected kernels of 97.6% without pre-processing, 97.7% with multiplicative scatter correction (MSC) and 97.4% with standard normal variate (SNV). This was equal to and better than those previously reported on wheat and agar plates (Zhang *et al.*, 2007; Yao *et al.*, 2008). This illustrates the potential of NIR hyperspectral imaging together with PCA and PLS-DA to be used in the maize industry for early detection of fungal infected kernels; however, this requires supplementary work.

Future research should include acquiring NIR hyperspectral images of the fungus, the sound maize kernels as well as infected kernels. Additionally, three dimensional (3-D) imaging could be applied whereby infected kernels are imaged thereafter a section would

be sliced away and another image taken. This would be repeated till the entire kernel had been sliced and imaged. With this a 3-D image could be assembled that could assist in understanding the mechanisms of fungal development and the chemical changes that accompany it.

NIR hyperspectral imaging is an important technique to be considered as a routine analytical tool. To date this was the first work done concerning maize hardness whereby endosperm hardness was assessed using NIR hyperspectral imaging. This gave information concerning distribution of the types of endosperm without complicated techniques such as scanning electron microscopy (SEM). However, further research should be combined with SEM to determine exact location and distribution based on starch granule shape and size. This could have a great impact for breeders to not just rapidly determine a range of traits but to be able to have this information localised, allowing these traits to be studied with specific reference to location within the sample. Moreover, single kernel analysis will now be possible which will allow for single kernels to be selected based on specific traits or a combination thereof. The most important advantage though would be that after analyses these kernels would continue their journey in the breeding trial. In addition, as the technology and software improves, commercial on-line systems could assist in separating individual kernels to ensure those kernels with appropriate hardness (and other quality traits) would be used and kernels with undesirable hardness would be removed from processing. This would improve production through increased efficiency in processing for specific end-products.

With regards to detection of fungal infected kernels, much like the detection of defects in other produce (Fernández Pierna *et al.*, 2004; Mehl *et al.*, 2004) it is important that these techniques are carried out objectively. Much too often human errors are made that affect the lives of others. This technique could greatly reduce this likelihood. Much research has been done with regards to the identification of fungal contaminated foodstuffs, however much is still required. The results of this research indicated that NIR hyperspectral imaging together with PCA and PLS-DA is ideal for this objective detection and could be of much benefit at maize silos where infected maize kernels would be removed from the storage and subsequent human and animal food chain.

NIR hyperspectral imaging has revolutionised the way one looks at a sample and emphasises the necessity to obtain a holistic assessment of the sample and not just one dimension.

References

Alonso Ferro, R.C., Malvar, R.A., Revilla, P., Ordás, A., Castro, P. & Moreno-González, J. (2008). Genetics of quality and agronomic traits in hard endosperm maize. *Journal of Agricultural Science*, **146**, 551-560.

Fandohan, P., Hell, K., Marasas, W. & Wingfield, M. (2003). Infection of maize by Fusarium species and contamination with fumonisin in Africa. *African Journal of Biotechnology*, **2**, 570-579.

Fernández Pierna, J.A., Baeten, V., Michotte Renier, A., Cogdill, R.P. & Dardenne, P. (2004). Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Journal of Chemometrics*, **18**, 341-349.

Gorretta, N., Roger, J.M., Aubert, M., Bellon-Maurel, V., Campan, F. & Roumet, P. (2006). Determining vitreousness of durum wheat kernels using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **14**, 231-239.

Headrick, J.M., Pataky, J.K. & Juvik, J.A. (1990). Relationships among carbohydrate content of kernels, condition of silks after pollination, and the response of sweet corn inbred lines to infection of kernels by *Fusarium moniliforme*. *Phytopathology*, **80**, 487-494.

Jackson, L. & Jablonski, J. (2004). Fumonisins. In: *Mycotoxins in food: detection and control* (edited by N. Magan & M. Olsen). Pp. 367-405. Cambridge, England: Woodhead Publishing, Ltd.

Lee, K.M., Bean, S.R., Alavi, S., Herrman, T.J. & Waniska, R.D. (2006). Physical and biochemical properties of maize hardness and extrudates of selected hybrids. *Journal of Agricultural and Food Chemistry*, **54**, 4260.

Mehl, P.M., Chen, Y.R., Kim, M.S. & Chan, D.E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering*, **61**, 67-81.

Munkvold, G.P. & Desjardins, A.E. (1997). Fumonisins in maize: can we reduce their occurrence? *Plant Disease*, **81**, 556-565.

Smail, V.W., Fritz, A.K. & Wetzel, D.L. (2006). Chemical imaging of intact seeds with NIR focal plane array assists plant breeding. *Vibrational Spectroscopy*, **42**, 215-221.

Watson, S.A. (1987). Structure and composition. In: *Corn: Chemistry and Technology* (edited by S.A. Watson & P.E. Ramstad). Pp. 53-82. St. Paul, Minnesota, USA: American Association of Cereal Chemists, Inc.

Wolf, M.J., Buzan, C.L., Macmasters, M.M. & Rist, C.E. (1952). Structure of the mature corn kernel. I. Gross anatomy and structural relationships. *Cereal Chemistry*, **29**, 321-333.

Wolf, M.J., Khoo, U. & Seckinger, H.L. (1969). Distribution and subcellular structure of endosperm protein in varieties of ordinary and high-lysine maize. *Cereal Chemistry*, **46**, 253-263.

Yao, H., Hruska, Z., Kincaid, R., Brown, R. & Cleveland, T. (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing and Instrumentation for Food Quality and Safety*, **2**, 215-224.

Zhang, H., Paliwal, J., Jayas, D.S. & White, N.D.G. (2007). Classification of fungal infected wheat kernels using near-infrared reflectance hyperspectral imaging and support vector machine. *Transactions of the ASABE*, **50**, 1779-1785.