Near infrared hyperspectral imaging as detection method for pre-germination in whole wheat, barley and sorghum grains

by

Paulina Engelbrecht

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Food Science at the University of Stellenbosch

Supervisor: Prof. Marena Manley
Co-supervisor: Prof. Paul Geladi

Faculty of AgriSciences
Department of Food Science

March 2011

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 sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: .....

Abstract

The use of near infrared (NIR) hyperspectral imaging for distinguishing between pre-germinated and non pre-germinated barley, wheat and sorghum kernels and, the effect of kernel shape on hyperspectral images, have been investigated.

Two sample sets were imaged. The first sample set was divided into six subsets; these subsets were treated with water and left to pre-germinate for different times (0, 6, 9, 12, 18 and 24 hrs). Subset viability was determined with the tetrazolium test. The second sample set was divided into seven subsets, treated with water and left to pre-germinate for 0, 3, 6, 9, 12, 18, 24 or 30 hrs. Individual kernel viability was determined with the tetrazolium test.

NIR hyperspectral images were acquired using two different SisuCHEMA hyperspectral imaging systems. The first system acquired images with a 150 μm spatial resolution (first sample set) and the second system acquired images with a 30 μm spatial resolution (second sample set). Principal component analysis (PCA) was performed and a distinction between pre-germinated and non pre-germinated kernels was illustrated in PCA score images. Loading line plots showed that the main compounds contributing to spectral variation were starch, water and protein. These compounds were related to starch and protein hydrolysis. The distinction between pre-germinated and non pre-germinated kernels observed in the 30 μm spatial resolution images indicated NIR hyperspectral imaging was perhaps sensing incomplete endosperm degradation. Some kernels determined as pre-germinated by the tetrazolium test had the same chemical composition according to the score image as non pre-germinated kernels in the 30 μm spatial resolution images.

A partial least squares discriminant analysis (PLS-DA) model with two classes (pre-germinated and non pre-germinated) was developed for each of the cultivars of the first sample set. The two classes were assigned in principal component (PC) 1 vs. PC 5 score plots. The model created for the barley cultivars resulted in excessive false positives and false negatives. The prediction results of wheat cultivars revealed that the model had a classification rate of 81% for the non pre-germinated class and 93% for the pre-germinated class. The sorghum prediction results revealed that the model correctly predicted 97% of the non pre-germinated class and 93% of the pre-germinated class.

Two different PLS-DA models were developed for one image of each cultivar of the 30 μm spatial resolution images. The first model was developed by assigning each kernel in the score image and the second model was developed by assigning pixels in the score plot to either the pre-germinated or non pre-germinated class. Model 1 resulted in excessive false negatives. Model 2 resulted in excessive false positives.

The differences between pre-germinated and non pre-germinated kernels were only observed in higher (PC 5 and 6) order PCs of the 150 μm spatial resolution images. The lower (PCs 1 to 4)

order PCs (of each commodity) were subsequently examined with the aid of classification gradients. Kernel shape effects were observed in these PCs.

The use of NIR hyperspectral imaging for distinguishing between pre-germinated and non pre-germinated grain kernels shows promise.

Uittreksel

Die gebruik van naby infrarooi (NIR) hiperspektrale beeld-analise is geëvalueer om onderskeid te tref tussen voor-ontkiemde en nie-voor-ontkiemde gars, koring en sorghum korrels. Die effek van korrelvorm op hiperspektrale beelde is ook geëvalueer.

Die eerste stel graan-monsters is gebruik vir 150 µm ruimtelike resolusie beelde en die tweede stel is gebruik vir 30 µm ruimtelike resolusie beelde. Die eerste kultivar stel is verdeel in ses sub-stelle en met gedistilleerde water behandel vir 0, 6, 9, 12, 18 en 24 hr. Sub-stel lewensvatbaarheid is met die tetrazolium toets vasgestel. Elke kultivar in die tweede stel is in sewe sub-stelle verdeel en is vir 0, 3, 6, 9, 12, 18, 24 of 30 hr geïnkubeer. Individuele korrel lewensvatbaarheid is met die tetrazolium toets vasgestel.

NIR hiperspektrale beelde is verkry deur gebruik te maak van twee verskillende SisuCHEMA kameras. Die verskillende kameras is gebruik om verskillende resolusie (30 en 150 µm ruimtelike resolusie) beelde te verkry. Hoofkomponent analise (HKA) is uitgevoer en ’n verskil tussen voor-ontkiemde en nie-voor-ontkiemde korrels is waargeneem in die 150 µm ruimtelike resolusie beelde. HK ladings stippe het water, stysel en proteïene uitgesonder as die verbindings wat bydrae het tot spektrale variasie. ’n Verskil tussen die voor-ontkiemde korrels en nie-voor-ontkiemde korrels is ook gesien vir die 30 µm ruimtelike resolusie beelde. Dit is egter ook waargeneem dat sommige korrels as voor-ontkiem bepaal is deur die tetrazolium toets, maar dié korrels het dieselfde chemiese samestelling volgens die punte beeld as nie-voor-ontkiemde korrels. Onvolledige endosperm hidrolise is ’n moontlike verduideliking vir die verskynsel. Die verbindings wat bygedra het tot die variasie is water, stysel en proteïene.

’n Parsiële kleinste kwadrate diskriminant analise (PKW-DA) model met twee klasse is ontwikkel vir elke kultivar van die 150 µm ruimtelike resolusie beelde. Die klasse is aangewys in the punte stip. Die model met die hoogste variasie in Y is gekies om die ander kultivars van dieselfde kommoditeit te voorspel. The PKW-DA resultate vir die gars kultivars het getoon dat die model vals positiewes en vals negatiewes opgelewer het. Die koring PKW-DA model het ’n klassifikasie koers van 81% vir die nie-voor-ontkiemde klasse en 93% vir die voor-ontkiemde klasse opgelewer. The PKW-DA resultate vir sorghum het getoon dat die model ’n klassifikasie koers van 97% vir die nie-voor-ontkiemde klasse en 93% vir die voor-ontkiemde klasse opgelewer.

Twee verskillende PKW-DA modelle is ontwikkel vir elke beeld van elke kultivar van die 30 µm ruimtelike resolusie beelde. Die eerste model is ontwikkel deur elke korrel in die punte beeld aan te wys tot een van twee klasse en die tweede model is ontwikkel deur die beeldelemente in die punte stip tot een van twee klasse toe te skryf. Model 1 het vals negatiewes opgelewer en model 2 vals positiewes.

Die verskille tussen die nie-voor-ontkiemde en voor-ontkiemde korrels is eers verduidelik in hoër orde HK van die 150 µm ruimtelike resolusie beelde. Die laer orde HK is dus ondersoek vir

hul bydrae tot spektrale variasie met die hulp van klassifikasie gradiënte. Korrel vorm effekte is waargeneem.

Die gebruik van NIR hiperspektrale beelding om onderskeid te tref tussen voor-ontkiemde en nie-voor-ontkiemde graan korrels, lyk belowend.

Acknowledgements

Firstly I would like to give praise to our heavenly Father for His guidance through these past two years.

I also recognise the following persons and institutions for their contribution to the successful completion of this thesis:

Prof. Marena Manley, my study leader, for her leadership, advice and encouragement;

Prof. Paul Geladi (Swedish University of Agricultural Sciences), my co-study leader, for his interest, support and advice during this study;

Dr David Nilsson and Oskar Jonsson (Umbio AB, Umea, Sweden) for the use of the SisuCHEMA imaging system, Evince software and assistance with imaging;

Prof. Alvaro Viljoen and Ilze Vermaak (Tshwane University of Technology, Pretoria,) for the use of the SisuCHEMA imaging system;

Tania Adams, Robert Alexander and Idelet Meijering (South African Breweries Malting, Caledon) for advice, training and barley samples;

Prof Klaus Pakendorf (Agricultural Research Council – Small Grains Institute, Stellenbosch) and Gerida du Toit (Sensako, Bethlehem) for wheat samples;

Evan Brauteseth (PANNAR, Greytown) for sorghum samples;

FoodBev SETA and Stellenbosch University for bursaries;

The Winter Cereal Trust and Sorghum Trust for project funding;

The South African-Swedish Research Partnership Programme Bilateral Agreement, NRF, (UID 60958) for funding to work at the Swedish University of Agricultural Sciences (VR. 348-2006-6715);

CST-SA for travel and subsistence grant to attend the CST-SA – ICC International Grains Symposium in Pretoria, South Africa;

FoodBev SETA for funding to attend IUFoST 2010 in Cape Town, South Africa;

Arrie Arends (Biochemistry Department, Stellenbosch University) and Jeanette Celliers (Microbiology Department, Stellenbosch University) for help with freeze drying of samples;

All the staff and postgraduate students of the Department of Food Science;

Everyone from the NIR spectroscopy laboratory, especially Paul Williams, Evette Roux, Gerida du Toit and Glen Fox;

Cushla McGoverin for her help with thesis revision and always being there to answer my questions;

Jana Visser and Kim O'Kennedy (housemates and fellow NIRds) for their support, group hugs and comic relief during the past two years;

Joske Kotze for his patience, words of encouragement and understanding; and

Last, but certainly not least, my parents and sisters for their love, support and encouragement through all my years of study. I could not have done this without you.

Table of contents

Language and style used in this thesis are in accordance with the requirements for 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.

Glossary

Chemometrics is a discipline that utilises mathematical and statistical methods to analyse chemical data.

Classification image is a score image with selected score values or areas within the sample.

Classification plot is a score plot with selected score values or clusters.

Hyperspectral imaging (or NIR chemical imaging) combines conventional imaging with NIR spectroscopy.

Hyperspectral image (or hypercube) provides the spectral information of a sample in a spatially resolved manner.

Near infrared (NIR) spectroscopy is the study of NIR radiation and its interaction with matter.

Partial least squares discriminant analysis (PLS-DA) is a pattern recognition technique that correlates variation in a dataset with class membership.

Pre-germination is the unwanted germination of mature cereal grains subjected to wet or humid conditions.

Pre-processing refers to any transformation of original data to enhance the information represented. Raw data has to be pre-processed mathematically in order to remove or reduce certain defects in spectra, such as noise, baseline shift and light scattering.

Principal component analysis (PCA) is a technique that transforms a dataset into a more compact dataset with fewer variables, without losing the main variation/information contained within the dataset.

Score image is a visual representation of the principal component (PC) score values.

Score plot is two score vectors plotted against each other, where the relationship between objects can be observed.

List of Figures

List of Tables

Chapter 1

Introduction

Chapter 1

Introduction

Cereal crops play an important role in our everyday lives, providing nutrition not only to humans but animals as well. In developing countries 70% of dietary energy is provided by cereals (Orth & Shellenberger, 1988). Cereals are therefore an important part of the agricultural industry in these countries. In southern Africa the cereal industry provides much needed employment that creates an income for otherwise impoverished families. Cereals provide essential components for a healthy diet and can be used in an array of different products, e.g. bread, biscuits, beer, whisky (Kent & Evers, 1994), tortillas and couscous (De Alencar Figueiredo *et al.*, 2006). The loss of cereal crops disrupts and has the potential to ruin agricultural, food and seed industries. Pre-germination of grains is one possible threat to the cultivation of cereals (Mohan, 2008).

Pre-germination is the untimely germination of mature cereal grains that have been subjected to wet or humid conditions prior to or after harvest (Gruwel *et al.*, 2002; Smail *et al.*, 2006). It is initiated by the uptake of water near the embryo (Fincher & Stone, 1993). Nutrients (e.g. starch) stored in the endosperm are hydrolysed by enzymes synthesized upon the initiation of germination (e.g. alpha-amylase) and delivered to the growing embryo (Fincher & Stone, 1993; Kent & Evers, 1994). These changes are beneficial when the grain is used for plantlets (seedlings), but when the grain is intended for seed or the production of food and beverages it is detrimental (Kent & Evers, 1994). Pre-germination decreases grain weight, yield and functionality (Mohan, 2008), and will result in significant economic losses (Skerritt & Heywood, 2000). Pre-germinated kernels are more easily contaminated with micro-organisms and insects (Singh *et al.*, 2009). This can lead to contamination when mixed with sound kernels (Singh *et al.*, 2009). Micro-organisms more easily infect, and are better sustained by pre-germinated grains due to kernel damage which occurs when germination begins (Singh *et al.*, 2009). Pre-germinated grains give rise to favourable conditions for mould formation (Kent & Evers, 1994; Pagano *et al.*, 1997). This is caused by the increased concentration of alpha-amylase associated with pre-germination which leads to elevated amounts of soluble sugars (Kent & Evers, 1994; Pagano *et al.*, 1997; Lijavetzky *et al.*, 2000). The formation of mould in grain will lead to a decline in grain quality and utility (Pagano *et al.*, 1997).

Pre-germination decreases the utility of cereal grains such as barley, wheat and sorghum. For example, pre-germinated barley and sorghum exhibits reduced viability for malting. Malt is a major constituent in beer and whisky manufacturing and therefore the quality of malt is of great importance (Bamforth & Barclay, 1993; Dewar *et al.*, 1997; Woonton *et al.*, 2005). Non-viable grains will not germinate during the malting process. This will lead to decreased production and economic losses, as vigorous and rapid germination of the grains is required (Carn, 1982; Ullrich *et al.*, 2009). During bread production a small amount of alpha-amylase is necessary to degrade damaged wheat starch and produce fermentable sugars for gas production (Kruger, 1994). When

an excessive amount of alpha-amylase is present, as in the case of pre-germinated wheat, the starch will be degraded to complex dextrins and simple sugars (Kruger, 1994). This extensive starch degradation reduces the water-holding capacity of the resulting bread, causes the crumb to be weakened, and the dextrins produced make the crumb sticky (Kent & Evers, 1994).

The detection of pre-germination has to be accurate as under-estimation can lead to the utilisation of unsuitable grains and over-estimation can lead to the rejection of sound grains. Conventionally grains are visually inspected by graders at grain delivery sites (Bason *et al*., 1993; Bueckert *et al*., 2007). These subjective inspections are based on the knowledge and experience of graders, which can lead to inaccurate determination of pre-germinated grains (Carn, 1982; Neethirajan *et al.*, 2007; Singh *et al.*, 2009). Pre-germination may also be detected using time-intensive and destructive tests, for example, viscosity analysis (AACC method 22-08, 2010), falling number (AACC method 56-81B, 2010) and the tetrazolium test (Rennie & Gorey, 1988; Paliwal *et al.*, 1990). A rapid, non-destructive objective test with little or no chemical waste is desired in the grain grading and utilisation industries. Near infrared (NIR) hyperspectral imaging is such a technique, and has successfully been employed in the detection of pre-germination in intact wheat kernels (Smail *et al.*, 2006; Koç *et al.*, 2008; Singh *et al.*, 2009).

The NIR region of the electromagnetic spectrum ranges from 780 to 2500 nm (Bokobza, 1998; Workman & Schenk, 2004). NIR irradiation of biological materials may provide information on the chemical constituents of the sample (Workman & Schenk, 2004). The molecular composition of a sample causes the radiation to be selectively absorbed giving rise to a sample specific spectrum (Rodriguez *et al.*, 2005). The spectral information obtained may be correlated with standard methods of analysis and ultimately used to describe the chemical composition of the sample (Osborne *et al.*, 1993; Pasquini, 2003).

NIR spectroscopy has found numerous applications in the agricultural industry, e.g. predicting chemical composition of sorghum grain (Rodriguez *et al.*, 2005), measuring protein changes during wheat maturation (Gergely & Salgó, 2007), detection of aflatoxin B1 in maize and barley (Fernandez-Ibanez *et al.*, 2009), detection of insect fragments in wheat flour (Perez-Mendoza *et al.*, 2003), predicting sweet potato starch physiochemical quality and pasting properties (Lu *et al.*, 2006), and determination of soluble solids in apples (Ventura *et al.*, 1998). The success of NIR spectroscopy lies in the ability to investigate numerous samples in a short amount of time, the ability to use one spectral measurement from a sample to measure several parameters, the limited need for sample preparation, and that the technique is non-destructive (Bokobza, 1998). One of the big disadvantages of NIR spectroscopy is the fact that only an average spectrum of the irradiated sample area is acquired and the spatial distribution of the different chemical constituents within the sample cannot be described (Wang & Paliwal, 2007).

NIR hyperspectral imaging combines conventional NIR spectroscopy with imaging (Cogdill *et al.*, 2006; Mahesh *et al.*, 2008) to obtain spatial and spectral information of a sample (Gowen *et al.*, 2007). Thereby allowing the spatial distribution of different chemical constituents within a sample to

be described (Mahesh *et al.*, 2008). Stacked images taken at different wavelengths form hyperspectral images or hypercubes (Burger & Geladi, 2005). The hypercube consists of two spatial dimensions (x and y) combined with a spectral dimension (λ) (Cogdill *et al.*, 2006). Each pixel in such a hyperspectral image includes the whole spectrum at that specific point (Gowen *et al.*, 2007; Shahin & Symons, 2008). An imaging system with the dimensions 320 (x) x 583 (y) x 239 (λ) produces 239 single channel images each with 320 x 583 pixels (Burger & Geladi, 2006). This immense amount of data poses data exploration problems and has been addressed through the use of chemometric techniques.

Chemometrics is a discipline that utilises mathematics and statistics to design and analyse chemical experiments (Massart *et al.*, 1988). Principal component analysis (PCA) is one such chemometric technique. PCA searches for sources of variability, taking into consideration that the variability explained for one PC can be associated to a pure component, a mixture of components or a physical effect, such as shape (Amigo *et al.*, 2008). The shape of the imaged sample can have a great effect on image analysis and pre-processing techniques are applied to minimise these effects (Gowen *et al.*, 2008). Pre-processing techniques remove spectral variation while retaining spectral features that enable characterisation of the sample.

NIR hyperspectral imaging has found numerous applications in the food and agricultural industries, e.g. measuring the soluble solids distribution in kiwifruit (Martinsen & Schaare, 1998), the detection of hard vitreous and starchy wheat kernels (Shahin & Symons, 2008), the prediction of oil and oleic acid concentrations in maize (Weinstock *et al.*, 2006), the classification of sound and stained wheat kernels (Berman *et al.*, 2007), the detection of processed animal protein in compound feed (Pierna *et al.*, 2010), the evaluation of maize endosperm texture (Manley *et al.*, 2009), and the identification of water matrix co-ordinates in mushrooms subjected to mechanical vibration (Gowen *et al.*, 2009).

NIR hyperspectral imaging is an emerging analytical tool and although multiple kernels are analysed simultaneously, results from single kernels are obtained. This is possible due to the spectral information located in each pixel of the image. Discriminating between pre-germinated and non pre-germinated (which differ chemically) wheat, barley and sorghum could be possible using this technique. The information gained from NIR hyperspectral imaging can be assessed by graders without subjectivity.

The aim of this study was to demonstrate the application of NIR hyperspectral imaging as an objective, non-destructive and rapid alternative to conventional pre-germination detection methods for South African wheat, barley and sorghum cultivars. The specific objectives of this thesis were to:

- distinguish between non pre-germinated and pre-germinated grains (barley, wheat and sorghum) using principal component analysis (PCA);
- discriminate between pre-germinated and non pre-germinated kernels using partial least squares discriminant analysis (PLS-DA); and

- evaluate the effect of kernel shape on hyperspectral image analysis.

References

AACC. (2010). *Approved Methods of the AACC*. Determination of Falling Number (method 56-81B). St.Paul, Minnesota: American Association of Cereal Chemists.

AACC. (2010). *Approved Methods of the AACC*. Measurement of alpha-amylase activity with the Rapid Visco Analyser (method 22-08). St Paul, Minnesota: American Association of Cereal Chemists.

Amigo, J.M., Cruz, J., Bautista, M., Maspoch, S., Coello, J. & Blanco, M. (2008). Study of pharmaceutical samples by NIR chemical image and multivariate analysis. *Trends in Analytical Chemistry*, **27**, 696-713.

Bamforth, C.W. & Barclay, A.H.P. (1993). Malting technology and the uses of malt. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 297-354. St. Paul: American Association of Cereal Chemists, Inc.

Bason, M.L., Ronalds, J.A., Wrigley, C.W. & Hubbard, L.J. (1993). Testing for sprout damage in malting barley using the rapid visco-analyser. *Cereal Chemistry*, **70**, 269-272.

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.

Bueckert, R.A., Lefol, E.B. & Harvey, B.L. (2007). Early detection of non-visible sprouting in barley seed using rapid viscosity analysis. *Canadian Journal of Plant Science*, **87**, 3-12.

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.

Carn, J.D. (1982). Alpha-amylase indicates problems with malting of stored barley. *Food Technology in Australia*, **34**, 82-83.

Cogdill, R.P., Hurburgh, C.R. & Rippke, G.R. (2006). Single kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the American Society of Agricultural Engineers*, **47**, 331-320.

De Alencar Figueiredo, L.F., Davrieux, F., Fliedel, G., Rami, J.F., Chantereau, J., Deu, M., Courtois, B. & Mestres, C. (2006). Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. *Journal of Agricultural and Food Chemistry*, **54**, 8501-8509.

Dewar, J., Taylor, J.R.N. & Berjak, P. (1997). Determination of improved steeping conditions for sorghum malting. *Journal of Cereal Science*, **26**, 129-136.

Fernandez-Ibanez, V., Soldado, A., Martinez-Fernandez, A. & de la Roza-Delgado, B. (2009). Application of near infrared spectroscopy for rapid detection of aflatoxin B1 in maize and barley as analytical assessment. *Food Chemistry*, **113**, 629-634.

Fincher, G.B. & Stone, B.A. (1993). Physiology and biochemistry of germination in barley. In: *Barley: chemistry and technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 247-294. St. Paul: American Association of Cereal Chemists, Inc.

Gergely, S. & Salgó, A. (2007). Changes in protein during wheat maturation- what is measured by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **15**, 49-58.

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 and 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 bisporus*). *Journal of Chemometrics*, **22**, 259-267.

Gowen, A.A., Tsenkova, R., Esquerre, C., Downey, G. & O'Donnell, C.P. (2009). Use of near infrared hyperspectral imaging to identify water matrix co-ordinates in mushrooms (*Agaricus bisporus*) subjected to mechanical vibration. *Journal of Near Infrared Spectroscopy*, **17**, 363-371.

Gruwel, M.L.H., Yin, X.S., Edney, M.J., Schroeder, S.W., MacGregor, A.W. & Abrams, S. (2002). Barley viability during storage: use of magnetic resonance as a potential tool to study viability loss. *Journal of Agricultural and Food Chemistry*, **50**, 667-676.

Kent, N.L. & Evers, A.D. (1994). *Kent's technology of cereals.* New York: Elsevier Science Inc.

Koç, H., Smail, V.W. & Wetzel, D.L. (2008). Reliability of InGaAs focal plane array imaging of wheat germination at early stages. *Journal of Cereal Science*, **48**, 394-400.

Kruger, J.E. (1994). Enzymes of sprouted wheat and their possible technological significance. In: *Wheat Production, Properties and Quality* (edited by W. Bushuk & V.F. Rasper). Pp. 143-151. Glasgow: Blackie Academic and Professional.

Lijavetzky, D., Martinez, M.C., Carrari, F. & Hopp, H.E. (2000). QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. *Euphytica*, **112**, 125-135.

Lu, G., Huang, H. & Zhang, D. (2006). Prediction of sweet potato starch physiochemical quality and pasting properties using near infrared reflectance spectroscopy. *Food Chemistry*, **94**, 632-639.

Mahesh, S., Manickavasagan, A., Jayas, D.S., Paliwal, J. & White, N.D.G. (2008). Feasibility of near-infrared hyperspecctral imaging to differentiate Canadian wheat classes. *Biosystems Engineering* **101**, 50-57.

Manley, M., Williams, P., Nilsson, D. & Geladi, P. (2009). Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (*Zea maize* L.) kernels. *Journal of Agricultural and Food Chemistry*, **57**, 8761-8769.

Martinsen, P. & Schaare, P. (1998). Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology*, **14**, 271-281.

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

Mohan, A. (2008). Pre-harvest sprouting in cereals: a global scenario. *Current Science*, **94**, 704-705.

Neethirajan, S., Jayas, D.S. & White, N.D.G. (2007). Detection of sprouted wheat kernels using soft x-ray image analysis. *Journal of Food Engineering*, **81**, 509-513.

Orth, R.A. & Shellenberger, J.A. (1988). Origin, production and utilization of wheat. In: *Wheat: Chemistry and Technology* (edited by Y. Pomeranz). Pp. 1-13. St. Paul: American Association of Cereal Chemists, Inc.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). In: *Practical NIR spectroscopy with applications in food and beverage analysis* (edited by D. Browning). Pp. 1-35, 49-77. Essex: Longman Scientific & Technical.

Pagano, E.A., Benech-Arnold, R.L., Wawrzkiewicz, M. & Steinbach, H.S. (1997). Alpha-amylase activity in developing sorghum caryopses from sprouting resistant and susceptible varieties. The role of ABA and GAs on its regulation. *Annals of Botany*, **79**, 13-17.

Paliwal, S., Baskin, C.C. & Delouche, J.C. (1990). The relationship between the tetrazolium test, soil cold test and standard germination test in assessing sorghum (*Sorghum bicolor* (L.) Moench) seed quality. *Journal of Seed Technology*, **14**, 56-60.

Pasquini, C. (2003). Near infrared spectroscopy: fundamentals, practical aspects and analytical application. *Journal of the Brazilian Chemical Society*, **14**, 198-219.

Perez-Mendoza, J., Throne, J.E., Dowell, F.E. & Baker, J.E. (2003). Detection of insect fragments in wheat flour by near-infrared spectroscopy. *Journal of Stored Products Research*, **39**, 305-312.

Pierna, J.A.F., Dardenne, P. & Baeten, V. (2010). In-house validation of a near infrared hyperspectral imaging method for detecting processed animal proteins in compound feed. *Journal of Near Infrared Spectroscopy*, **18**, 121-133.

Rennie, W.J. & Gorey, V. (1988). Estimation potential germination of winter wheat and winter barley seed using a tetrazolium test. *Irish Journal of Agricultural Research*, **27**, 179-186.

Rodriguez, E.R., Escalera, A.M.A. & Landin, G.M. (2005). Predicting the chemical composition of sorghum grain by near infrared reflectance spectroscopy (NIRS). *Téc Pecu Méx*, **43**, 1-11.

Shahin, M.A. & Symons, S.J. (2008). Detection of hard vitreous and starchy kernels in amber durum wheat samples using hyperspectral imaging. *NIR news*, **19**, 16-18.

Singh, C.B., Jayas, D.S., Paliwal, J. & White, N.D.G. (2009). Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry*, **86**, 256-260.

Skerritt, J.H. & Heywood, R.H. (2000). A five-minute field test for on-farm detection of pre-harvest sprouting in wheat. *Crop Science*, **40**, 742-756.

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.

Ullrich, S.E., Lee, H., Clancy, J.A., Del Blanco, I.A., Jitkov, V.A., Kleinhofs, A., Han, F., Prada, D., Romagosa, I. & Molina-Cano, J.L. (2009). Genetic relationship between pre-harvest sprouting and dormancy in barley. *Euphytica*, **168**, 331-345.

Ventura, M., de Jager, A., de Putter, H. & Roelfs, F.P.M.M. (1998). Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). *Postharvest Biology and Technology*, **14**, 21-27.

Wang, W. & Paliwal, J. (2007). Near-infrared spectroscopy and imaging in food quality and safety. *Sensing and Instrumentation for Food Quality and Safety*, **1**, 193-207.

Weinstock, B.A., Janni, J., Hagen, L. & Wright, S. (2006). Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Applied Spectroscopy*, **60**, 9-16.

Woonton, B.W., Jacobsen, J.V., Sherkat, F. & Stuart, I.M. (2005). Changes in germination and malting quality during storage of barley. *Journal of the Institute of Brewing*, **111**, 33-41.

Workman, J. & Schenk, J. (2004). Understanding and using the near-infrared spectrum as an analytical method. In: *Near-infrared Spectroscopy in Agriculture* (edited by L. Al-Amoodi). Pp. 3-10. Madison: American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.

Chapter 2

Literature review

Contents

Chapter 2

Literature Review

1. Introduction

Cereal grains are the fruits of cultivated grasses (Hoseney, 1986; Kent & Evers, 1994). Cereals grow throughout the temperate and tropical regions of the world and have adapted to a range of environmental conditions (Kent & Evers, 1994). The principal cereal crops are wheat, rice, maize, barley, oats, rye, sorghum and millets. Barley (*Hordeum vulgare* L.), wheat (*Triticum* spp.*)* and sorghum (*Sorghum bicolor*) are amongst the world's most important crops. These cereal crops contribute to the daily dietary intake of humans and animals (Orth & Shellenberger, 1988). Cereals provide carbohydrates, an important food component, as well as protein, vitamins and minerals (Desai *et al.*, 1997). The cereal production industry provides essential work for thousands of individuals in developing countries. Cereals are used for an assortment of different food products, for example bread, beer, pasta, biscuits, vinegar and chocolates (Orth & Shellenberger, 1988; Kent & Evers, 1994; Oleson, 1994). Non-food applications of cereals include varnishes, dyes, bio-fuel and bio-degradable polymers (Oleson, 1994; Lawton, 2000). Indirectly cereals contribute to the production of meat as it is given to livestock as feed (Kent & Evers, 1994; Posner, 2000).

In this literature study the uses of barley, wheat and sorghum will be reviewed to establish the importance of cereals. Pre-germination will be described and discussed in terms of the effect it has on cereal products and production. Conventional methods for pre-germination detection and proposed alternative methods will also be reviewed. Near infrared (NIR) hyperspectral imaging will be reviewed with reference to imaging instruments, image acquisition and applications. Chemometric techniques such as principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA) and pre-processing will be reviewed, as well as hyperspectral image analysis.

2. Barley, wheat and sorghum: uses and importance

2.1 *Barley*

Barley (*Hordeum vulgare* L.*)* is the fourth major cereal of the world after wheat, rice and maize (Nilan & Ullrich, 1993). Barley is one of the most tolerant cereals known. It has adapted to, and is produced over a wider range of environmental conditions than most cereals, and the seed can be stored over long periods of time (Hockett, 2000). Barley is used for animal feed (Bhatty, 1993; Kent & Evers, 1994) or may be eaten as popped barley, barley flakes and sprouts (Hockett, 2000). Barley may be used as a rice extender, for the production of soy paste and soy sauce, and roasted as a tea or coffee substitute (Bhatty, 1993). Barley flour can be used as food thickener or wheat flour additive (Bhatty, 1993; Kent & Evers, 1994). The primary use of barley, however, is the

production of malt; the main applications of malt are for beer and whisky production. The quality of malt is therefore of great importance (Bamforth & Barclay, 1993; Kent & Evers, 1994; Woonton *et al.*, 2005). Malt may also be used in the baking industry or for the production of vinegar, chocolates, gin, vodka and breakfast cereals (Bamforth & Barclay, 1993).

2.2 *Wheat*

Wheat (*Triticum* spp.), because of its nutritive value and ease of harvesting, transportation and processing, is the leading cereal produced, consumed and traded in the world (Oleson, 1994; Posner, 2000). Wheat is grown all over the world under varied environmental conditions from the borders of the arctic to near the equator (Kent & Evers, 1994; Posner, 2000). Wheat is mainly used for human consumption in a variety of foods, i.e. different types of bread, cakes, pasta, crackers, pastries, breakfast cereals, flour, and biscuits (Orth & Shellenberger, 1988; Oleson, 1994; Posner, 2000). Wheat gluten may be extracted from wheat and dried; when it is hydrated it regains nearly all of its functionality (e.g. film-forming ability for gas retention, thermosetting properties for structural rigidity and water absorption) (Ponte *et al.*, 2000). Wheat gluten is used in breads, batter mixes, roasted nut coatings and pet foods (Ponte *et al.*, 2000). Wheat also serves as animal feed and is especially good for poultry (Orth & Shellenberger, 1988; Oleson, 1994). Non-food applications include production of plastics, varnishes, paper, bio-degradable packaging, soaps, rubber and cosmetics (Oleson, 1994; Lawton, 2000). An extensive wheat loss would be devastating to the milling and wheat production industry as it is used in a diverse range of products.

2.3 *Sorghum*

Cultivated sorghums in South Africa are of the specie bicolor and are known as *Sorghum bicolor* L. Moench. Sorghum is an important grain in Asian and African countries (De Alencar Figueiredo *et al.*, 2006) as it is used for animal feed as well as human nutrition (Rodriguez *et al.*, 2005). Sorghum is relatively drought tolerant and can adapt to tropical conditions (Rooney & Serna-Saldivar, 2000). In many African countries sorghum is the major staple food, especially where agricultural and environmental conditions are unfavourable for any other cereal crops. Sorghum can be used for porridge, infant food, couscous (combined with millet and maize), rotis (Indian unfermented bread), syrup or molasses, noodles, fermented breads and tortillas (made from a sorghum-maize combination) (Asiedu *et al.*, 1993; Rooney & Serna-Saldivar, 2000; De Alencar Figueiredo *et al.*, 2006). In Africa sorghum is used for the production of malt, which is primarily utilised for the production of sorghum beer (Kent & Evers, 1994; Raschke *et al.*, 1995). Thus, good quality (viable) sorghum is needed for malting.

2.4 *Importance of barley, wheat and sorghum*

Cereal crops such as barley, wheat and sorghum are of significant value in the everyday lives of humans and animals (Orth & Shellenberger, 1988). Grains of good quality are needed for use in the malting and milling industries (Kent & Evers, 1994). Thus, the cultivation of good quality cereals is important. The loss of cereal crops disrupts and has the potential to ruin agricultural, food and seed industries. Pre-germination of grains is one possible threat to the cultivation of cereals and diminishes the quality of grains (Mohan, 2008).

3. Germination and pre-germination of cereal grains

Germination is an important part of the life cycle of a cereal grain (Kent & Evers, 1994). Germination is initiated by the uptake of water near the embryo (Duffus, 1987; Fincher & Stone, 1993). The embryonic axis produces hormones that diffuse to target cells in the aleurone layer and stimulate the synthesis and secretion of hydrolytic enzymes, e.g. alpha-amylase and endo- and exo-peptidases, into the endosperm (Fincher & Stone, 1993; Dewar *et al.*, 1997). These enzymes catalyse the hydrolysis of storage polymers such as protein and starch (Fincher & Stone, 1993; Kent & Evers, 1994; Dewar *et al.*, 1997; Singh *et al.*, 2009b). Hydrolysis is the cleaving of a molecule by reaction with water, where one part of the molecule bonds the water OH and the other to the water H (Silberberg, 2000). **Fig. 2.1** shows the hydrolysis of amylose. Storage proteins are hydrolysed to a mixture of amino acids and short peptides, and starch is hydrolysed to glucose and small oligosaccarides (Duffus, 1987; Fincher & Stone, 1993; Kent & Evers, 1994). The released amino acids and sugars nourish the growing embryo (Xing *et al.*, 2010). Ultimately coleoptiles and rootlets emerge (Fincher & Stone, 1993).

Pre-germination and pre-harvest sprouting are terms used when mature cereal grains that have been subjected to wet or humid conditions start to germinate. Grains can germinate to such an extent that they are not useful in the grain utilisation and seed industries (Pagano *et al.*, 1997; Singh *et al.*, 2001; Gruwel *et al.*, 2002; Lin *et al.*, 2008). Pre-harvest sprouting occurs before the grains are harvested (Carrari *et al.*, 2003; Li *et al.*, 2004) and pre-germination occurs before harvest or at some point between harvest and grain utilisation (Gruwel *et al.*, 2002). For the purpose of this thesis the term pre-germination will be used, as pre-germination accounts for both pre- and post-harvest germination.

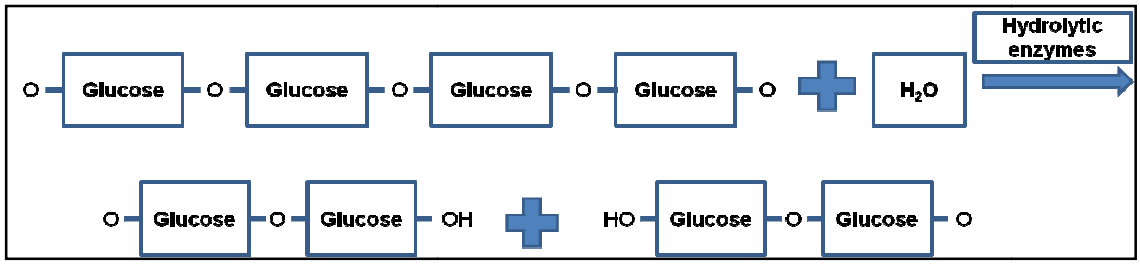


Figure 2.1 Illustration of starch hydrolysis. An amylose molecule is split into maltose molecules through the work of hydrolytic enzymes and water.

3.1 Effect of pre-germination on grain quality and utility

Pre-germination decreases grain weight, yield and functionality (Mohan, 2008). This will result in significant economic losses as the price paid to the farmer is greatly reduced (Skerritt & Heywood, 2000). Pre-germinated grains cannot be used for seed, and grain storage life is greatly reduced (Kent & Evers, 1994; Li *et al.*, 2004). Pre-germination can be interrupted by drying the kernel before there is any visible indication of germination, however, irreversible enzyme degradation has usually commenced (Carn, 1982; Lin *et al.*, 2008). Pre-germination of grains causes increased enzyme activity, and a loss of total dry matter mainly due to carbohydrate reduction (Lorenz & Kulp, 1981). The increased levels of alpha-amylase (synthesised during pre-germination) in the kernels leads to elevated amounts of soluble sugars, and consequently to favourable conditions for mould formation, resulting in a grain with decreased quality and functionality (Pagano *et al.*, 1997; Lijavetzky *et al.*, 2000). Pre-germinated grains infected with mould can lead to contamination of sound grains when mixed (Pagano *et al.*, 1997; Singh *et al.*, 2009b).

In barley and sorghum destined for malting, pre-germination can lead to the loss of viability (Carn, 1982; Pagano *et al.*, 1997; Lijavetzky *et al.*, 2000; Lin *et al.*, 2008), an important characteristic for malt quality. Grains that are not viable will not germinate vigorously during steeping (first step in the malting process) and will lead to production losses (Carn, 1982).

Wheat is used in a wide range of food products, some of which are negatively affected by pre-germination. During wheat bread production a small amount of alpha-amylase is necessary to degrade damaged starch and produce fermentable sugars for gas production by yeast (Kruger, 1994). When an excessive amount of alpha-amylase is present, as in the case of pre-germinated wheat, the starch will be degraded to complex dextrins and simple sugars. These dextrins make the crumb sticky (Kent & Evers, 1994); this may only be revealed when the bread is sliced (Kruger & Reed, 1988). Degraded starch can also have deleterious effects on the water-holding capacity of dough, as water normally bound by starch is released causing the dough to become slack. The increased proteolytic activity that usually accompanies increased amylase activity in pre-germinated wheat may also have a negative effect on bread baking quality (Halverson & Zeleny, 1988), e.g. collapsing and runny dough (Kent & Evers, 1994). In pasta products pre-germinated

wheat can lead to an increase in reducing sugars, more dissolved solids in the cooking water and a reduced dried pasta shelf-life (Kruger & Reed, 1988; Neethirajan *et al.*, 2007).

3.2 Conventional pre-germination detection methods

The detection of pre-germinated cereals is important as they are not useful in the food and beverage production industry (Kruger, 1994; Lin *et al.*, 2008; Mohan, 2008). The accurate detection of pre-germination is imperative as crop value is decreased and the farmer's compensation reduced when grains are classified as pre-germinated (Bason *et al.*, 1993; Skerritt & Heywood, 2000).

Pre-germination in cereal grains is typically detected using viscosity analysis, standard germination tests, the tetrazolium test, visual inspection or falling number. In the malting barley industry the tetrazolium test, standard germination test and visual inspection are employed. In the wheat milling and breeding industry falling number is used for the detection of pre-germination. Sorghum grains are visually inspected for pre-germination.

Graders at grain delivery sites visually inspect every batch entering the facility. The graders are trained to identify pre-germinated and damaged kernels (Bason *et al.*, 1993; Bueckert *et al.*, 2007). The graders examine a representative sample and when a certain level (established by each industry) of pre-germinated grain is detected the batch is downgraded to feed grade (Kruger, 1990). These subjective inspections are based on the knowledge and experience of the graders (Skerritt & Heywood, 2000; Neethirajan *et al.*, 2007); this can lead to the inaccurate determination of pre-germinated grains (Carn, 1982; Singh *et al.*, 2009b). Visual inspection can be unreliable as some damage to the grain occurs before pre-germination is visible to the naked eye (Carn, 1982; Skerritt & Heywood, 2000; Barnard, 2001; Lin *et al.*, 2008).

Viscosity analysis, to determine pre-germination in barley, is performed by means of a rapid visco-analyser (RVA) (AACC method 22-08, 2010). This method is based on the measurement of the starch pasting properties of a ground barley and water mixture while the sample is heated and stirred (Rasper & Walker, 2000). The stirring number (SN) after three minutes of heating and stirring are indicative of the presence or absence of pre-germinated grain (Rasper & Walker, 2000; Bueckert *et al.*, 2007). The action of alpha-amylase decreases viscosity and in turn the SN (Bueckert *et al.*, 2007). A high SN (SN>150) indicates an insignificant proportion of pre-germinated grains present in the sample, while a low SN (SN<100) indicates the presence of largely pre-germinated grains. The American Association of Brewing Chemists, however, recommended that institutions determine their own SN boundaries to accommodate genotype and environment interactions (Bason *et al.*, 1993; Bueckert *et al.*, 2007). This method can be applied to all small grains, including wheat and sorghum (Ross *et al.*, 1987; Raschke *et al.*, 1995). Viscosity analysis using the RVA takes three minutes to complete and can be classified as a rapid method (Ross *et al.*, 1987). The three minutes, however, does not include sample preparation (milling, weighing, measuring of water and mixing).

The tetrazolium test is used to determine seed viability, or potential to germinate (Rennie & Gorey, 1988; Bueckert *et al.*, 2007). A solution of tri-phenyl tetrazolium chloride is added to longitudinally cut kernels. The colourless tri-phenyl tetrazolium chloride is reduced by enzymes present in living cells to form a stable red formazan compound (Rennie & Gorey, 1988). Dead cells do not induce a colour change (Rennie & Gorey, 1988). No red staining of the embryo indicates non-viability and a red colour indicates viability (Rennie & Gorey, 1988; Paliwal *et al.*, 1990). The tetrazolium test is not a rapid method for the detection of pre-germination due to the tedious sample preparation required (cutting of each kernel).

Measurement of the falling number (AACC method 56-81B, 2010) requires a test tube filled with finely ground wheat and distilled water to be shaken and placed in a boiling water bath. The mixture is then stirred for 60 seconds after which the stirrer is dropped into the test tube from a fixed height (Kruger & Reed, 1988; Rasper & Walker, 2000). The time (seconds) required for the stirrer to reach the bottom of the test tube is the falling number. The falling number is indicative of the extent of starch degradation within a sample (Kruger & Reed, 1988; Rasper & Walker, 2000). A low falling number (< 250 s) indicates pre-germination of the grains as alpha-amylase breaks down the starch during the germination process (Halverson & Zeleny, 1988; Kruger & Reed, 1988). This method can be affected by water quality, fibre content and milling of the sample (Kruger & Reed, 1988; Skerritt & Heywood, 2000).

The standard germination test is used to determine germination energy (Bueckert *et al.*, 2007); an important characteristic for barley destined for malting. Distilled water is added to a Petri-dish containing two layers of filter paper and 100 kernels. The kernels are left to germinate for 72 hours and the percentage of sprouted kernels is recorded. The standard germination test requires no sample preparation, but is less sensitive than viscosity analysis in detecting the pre-germination of samples (Bueckert *et al.*, 2007).

Visual inspection is a subjective method for the determination of pre-germination and, unlike the previously mentioned tests, non-destructive. All conventional methods for the detection of pre-germination are time-consuming, largely due to sample preparation. There is a need for a rapid, objective and non-destructive test for pre-germination which requires minimal sample preparation and is a high throughput method (Bueckert *et al.*, 2007). Subsequently, several alternative methods have been reviewed.

3.3 *Alternative pre-germination detection methods*

In 2002 Gruwel *et al.* studied the loss of viability due to pre-germination in barley kernels by means of magnetic resonance imaging (MRI) and nuclear magnetic resonance (NMR) experiments. The barley samples were artificially pre-germinated and kernels were removed at regular intervals for non-destructive MRI and NMR measurements. Viability was established with alpha-amylase, tetrazolium and standard germination tests. With MRI it was evident that non-viable (pre-germinated) kernels showed a rapid uptake of water in the scutellum area. This could indicate

damage to this part of the kernel because of pre-germination. Nuclear magnetic resonance relaxation experiments revealed that viable and non-viable barley kernels could be distinguished from one another. Gruwel *et al.* (2002) concluded that these two methods could be used as tools for the detection of pre-germination in barley kernels. The MRI imaging and NMR experiments are time-consuming and the instruments are expensive.

The use of a soft X-ray system was evaluated for its potential to detect pre-germinated wheat kernels (Neethirajan *et al.*, 2007). Wheat was pre-germinated by subjecting kernels to distilled water for 48 hours and subsequently freeze drying to remove excess moisture. Falling number was used to determine the levels of alpha-amylase in the pre-germinated and non pre-germinated kernels. White specks or internal fissures (caused by starch degradation) were observed in the endosperm of pre-germinated kernel X-ray images, but not in the non pre-germinated kernel images. Back propagation neural network (BPNN) classifiers accurately classified 90% and 95% of pre-germinated and non pre-germinated kernels, respectively. The linear-function-parametric classifier correctly identified 87% and 92% of pre-germinated and non pre-germinated kernels, respectively. Soft X-ray images could detect pre-germinated wheat kernels non-destructively, however, X-rays pose a potential health risk to humans (Singh *et al.*, 2009b).

Near infrared (NIR) hyperspectral imaging was used to study pre-germination in wheat kernels (Smail *et al.*, 2006; Koç *et al.*, 2008). Wheat samples were artificially germinated for 3, 6, 12, 24, 36 and 48 hours and freeze dried to terminate germination processes. Kernels were imaged with an InGaAs (indium gallium arsenide) focal plane array imaging system in the germ up orientation. The system had a spectral range of 1100-1700 nm. Principal component analysis (PCA) score images revealed advanced embryo development for pre-germinated kernels (non-viable) and no embryo development in non pre-germinated kernels. Viability was confirmed with an alpha-amylase test kit (WheatRite), viscosity analysis and visual inspection (Smail *et al.*, 2006). It was concluded that chemical imaging could reveal pre-germination of grains with sensitivity greater than visual inspection, viscosity and alpha-amylase activity testing. This study was performed to measure pre-germination susceptibility and thereby assists breeding selections; pre-germination was observed in several cultivars after 24, 12 and even 6 hours of water absorption. Early pre-germination activity in the germ and aleurone layer was also observed in this study. Koç *et al*. (2008) used the method created by Smail *et al*. (2006) and examined six wheat cultivars and their susceptibilities to pre-germination. They found that some cultivars were more resistant to pre-germination than others.

The accuracy of NIR hyperspectral imaging for the detection of wheat kernel pre-germination has been assessed (Singh *et al.*, 2009b). A hyperspectral imaging system measuring in the range of 1000-1600 nm (with 60 evenly distributed wavelengths) was used to acquire images from pre-germinated and non pre-germinated kernels. PCA was employed to reduce the dimensionality of the data. From the loading of principal component (PC) one, two wavelengths of importance were identified, i.e. 1101.7 nm (second overtone of CH) and 1305.1 nm (first overtone CH combination).

The absorption at these wavelengths was attributed to starch based molecular vibrations. A linear discriminant classifier and quadratic discriminant classifier correctly (100%) classified non pre-germinated (0 hrs) and pre-germinated (48 hrs) kernels. A Mahalanobis discriminant classifier identified 100% and 98.3% of pre-germinated and non pre-germinated kernels, respectively. This study also indicated that NIR hyperspectral imaging holds promise in the detection of midge damaged kernels.

These alternative methods are promising techniques for the detection of pre-germination in cereal grains. For the purpose of this thesis only NIR hyperspectral imaging will be reviewed in more detail.

4. Near infrared spectroscopy

The discovery of the NIR region was accomplished by the experiments of William Herschel in 1800 (Osborne *et al.*, 1993; Wang & Paliwal, 2007). His experiments measured thermal energy from solar emissions beyond the red portion of the visible light spectrum (Herschel, 1800). The NIR region (780 to 2500 nm) lies between the visible and infrared (IR) regions in the electromagnetic spectrum (Miller, 2001; Workman & Schenk, 2004). The study of NIR radiation and its interaction with matter is called NIR spectroscopy (Wang & Paliwal, 2007). The NIR region contains absorption bands corresponding to overtones and combinations of CH, OH, CO, SH and NH molecular vibrations (Bokobza, 1998; Siesler, 2002; Workman & Schenk, 2004). When irradiated with NIR radiation a sample selectively absorbs energy corresponding to the vibrational frequencies (stretching or bending vibrations) of the bonds present, thus creating a sample specific spectrum (Workman & Schenk, 2004; Rodriguez *et al.*, 2005). Stretching vibrations are defined as a continuous change in the inter-atomic distance along the axis of the bond between the two atoms while a bending vibration is characterised as a change in the bond angle (Osborne *et al.*, 1993). Energy that is not absorbed is transmitted or reflected (Rodriguez *et al.*, 2005).

Overtone and combination bands are observable only when a dipole moment change occurs during the relevant vibration and the bond is associated with a large mechanical anharmonicity (Osborne *et al.*, 1993). NIR spectra are dominated by bands attributable to X-H stretching as these vibrations have the largest anharmonic constants and therefore the largest intensities (Siesler, 2002). Overtone and combination bands are broad and highly overlapping making direct interpretation difficult.

The NIR spectra of foods contain a vast amount of information that can only be extracted and interpreted by mathematical processing of the data, i.e. chemometrics (Osborne *et al.*, 1993). Chemometrics is a discipline that utilises mathematics and statistics to analyse chemical data (Massart *et al.*, 1988). Chemometrics involves techniques such as principal component analysis (PCA), multiple linear regression (MLR), principal component regression (PCR), partial least squares regression (PLS-R) and partial least squares discriminant analysis (PLS-DA).

Typically, a NIR spectroscopic system consists of five major components: radiation source, radiation separator, detector, sampling devices all connected to a computer (operates the spectroscopic system) (Coates, 1998; McClure, 2001; Wang & Paliwal, 2007). NIR radiation rich light sources (tungsten halogen lamps, light emitting diodes, laser diodes and lasers) illuminate samples (Osborne *et al.*, 1993; McClure, 2001) and the light separator disperses the light by using prisms, acousto-optic tunable filters (AOTF) and liquid crystal tunable filters (LCTF) (Osborne *et al.*, 1993; Pasquini, 2003; Wang & Paliwal, 2007). The surface composition and the structure of the sample will determine how deep the NIR radiation will penetrate (Workman & Schenk, 2004). The radiation that is transmitted through a sampling cell, or reflected from the sample, is detected by detectors (silicon photodiode, lead sulphide and indium gallium arsenide) and a spectrum is measured (Pasquini, 2003; Wang & Paliwal, 2007). Sampling devices are implements that aid the measurement of chemical composition of samples with NIR radiation, i.e. probes, liquid cells and other accessories (Osborne *et al.*, 1993; Wang & Paliwal, 2007).

NIR spectroscopy has become a widely popular technique for determining chemical properties of various products and foodstuffs, and has numerous applications, e.g. agriculture (Shashikumar *et al.*, 1993; Perez-Mendoza *et al.*, 2003; Saranwong *et al.*, 2004; Juhasz *et al.*, 2005; Rodriguez *et al.*, 2005; Lu *et al.*, 2006; Huang *et al.*, 2008; Woodcock *et al.*, 2008; Fernandez-Ibanez *et al.*, 2009), pharmaceutics (Patel *et al.*, 2000; Roggo *et al.*, 2005; Roggo *et al.*, 2007; Storme-Paris *et al.*, 2009; Storme-Paris *et al.*, 2010), alcoholic beverages (Cozzolino *et al.*, 2006; Zeaiter *et al.*, 2006; Kolomiets *et al.*, 2010) and medicine (Peuchant *et al.*, 1992; Mason *et al.*, 1994; Tergorg *et al.*, 2002; Olsson & Thelin, 2006; Manfredini *et al.*, 2009; Moffat *et al.*, 2010). The appeal of NIR spectroscopy over other chemical techniques lies in that it is rapid, non-destructive (Osborne *et al.*, 1993; Bokobza, 1998; Workman & Schenk, 2004) and requires little or no sample preparation; this is especially beneficial as technically unskilled personnel in a factory environment can perform NIR analyses (Osborne *et al.*, 1993). This technique has special appeal for the environmentally conscious, as chemical pre-treatments and waste material are minimised. The NIR region is also suitable for moderately concentrated samples.

NIR spectroscopy has great analytical capabilities and the applications of this technique are virtually endless, however, some limitations have been described. NIR spectroscopy still remains dependent on external chemical methods of analysis (Osborne *et al.*, 1993). The technique is not suitable for the determination of mineral content, as there is no absorption of minerals in the NIR region (Huang *et al.*, 2008). The biggest limitation of NIR spectroscopy has been the inability to source spatial information concurrent with spectral data collection (Martinsen & Schaare, 1998; Peirs *et al.*, 2003). However, with advances in technology a new form of NIR spectroscopy, where it is combined with imaging, has developed, i.e. NIR hyperspectral imaging.

5. NIR hyperspectral imaging

NIR hyperspectral imaging, also sometimes referred to as NIR chemical imaging, combines conventional imaging with NIR spectroscopy (Geladi *et al.*, 2004; Burger & Geladi, 2005; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Wang & Paliwal, 2007; Singh *et al.*, 2009b). The term hyperspectral image is given to an image that is produced by overlaying hundreds of single greyscale images, where each of these greyscale images represents the intensities of a single wavelength (Burger & Geladi, 2006). The resulting hyperspectral image provides spectral information of a sample in a spatially resolved manner (Martinsen & Schaare, 1998; Wang & Paliwal, 2007; Singh *et al.*, 2009b). Hyperspectral images can be used to identify chemical components, their concentration and distribution within a sample (Martinsen & Schaare, 1998; Burger & Geladi, 2005).

NIR hyperspectral imaging has been used successfully in the agricultural industry to detect insect damaged wheat kernels (Singh *et al.*, 2009a), predict oil and oleic acid concentrations in maize (Weinstock *et al.*, 2006), classify maize kernel hardness (Williams *et al.*, 2009), measure the soluble solids distribution in kiwifruit (Martinsen & Schaare, 1998), differentiate between wheat classes (Mahesh *et al.*, 2008) and map quality attributes of apples (Peirs *et al.*, 2003). Hyperspectral imaging has also been employed in remote sensing (Chen *et al.*, 2007; Gao *et al.*, 2009), medicine (Vo-Dinh *et al.*, 2004; Martin *et al.*, 2006) and pharmaceutical science (Roggo *et al.*, 2005; Amigo *et al.*, 2008). Review articles and books indicating the number of NIR hyperspectral imaging applications are available, and these applications will not be discussed in detail (Reich, 2005; Gowen *et al.*, 2007; Wang & Paliwal, 2007; Geladi & Manley, 2010).

Hyperspectral imaging has numerous advantageous over conventional analytical methods. It is non-destructive, quick, requires little or no sample preparation and is sensitive to minor components (Gowen *et al.*, 2007; Gowen *et al.*, 2008). The biggest advantage is the ability to describe the location of chemical and physical components (Martinsen & Schaare, 1998; Burger & Geladi, 2005; Gowen *et al.*, 2008). These physical components can include the shape of the imaged sample (Gowen *et al.*, 2008). Shape effects can cause scaling differences in the spectra; spectra from the highest part of the imaged sample can exhibit higher reflectance intensities than spectra from the lower parts of the imaged sample. This could be caused by the difference in path length from different points on the sample to the detector. It is thus important to consider the shape of the sample before image acquisition and during image analysis.

5.1 *NIR hyperspectral imaging system*

A typical NIR hyperspectral imaging system comprises a spectrograph, camera, detector, sample stage and an illumination unit, all coupled to a computer (Gowen *et al.*, 2007). The spectrograph is responsible for wavelength separation and selection; this can be achieved by liquid crystal tunable filters (LCTF), acousto-optic tunable filters (AOTF) or a prism-grating-prism (PGP) filter (Geladi *et al.*, 2007; Gowen *et al.*, 2007; Wang & Paliwal, 2007). Detectors are made of indium-gallium-arsenide (InGaAs), lead-sulfide (PbS), indium-antimonide (InSb) or mercury-cadmium-telluride

(HgCdTe) (Geladi *et al.*, 2007; Gowen *et al.*, 2007; Wang & Paliwal, 2007; Tan *et al.*, 2008). The illumination unit houses the light source which usually includes a tungsten halogen lamp, xenon plasma lamps, light emitting diodes (LED) or lasers (Geladi *et al.*, 2007; Gowen *et al.*, 2007; Wang & Paliwal, 2007). Spatial information is obtained by direct measurement through the spectrometer optics or is controlled by the positioning of the sample (Geladi *et al.*, 2007). To determine whether a hyperspectral imaging system is operating properly, a calibration step should be included after assembly and before utilisation of the instrument (ElMasry & Sun, 2010). The objectives of this procedure are to standardise the spectral axis of the hyperspectral image, validate the acceptability and credibility of the spectral data and detect instrument errors. Three different instrument configurations are employed at present, i.e. point scan, line scan and focal plane scan.

5.1.1 Point scan imaging

The point scan imaging configuration (**Fig. 2.2**) is used to measure a spectrum from a small spot of a sample after which the sample has to be repositioned before obtaining a new spectrum (Geladi *et al.*, 2007). By moving the sample in two spatial dimensions a complete hyperspectral image is acquired. This instrument configuration provides stable, high resolution spectra, but the positioning and repositioning of the sample is tedious and time-consuming (Burger, 2006; Geladi *et al.*, 2007).

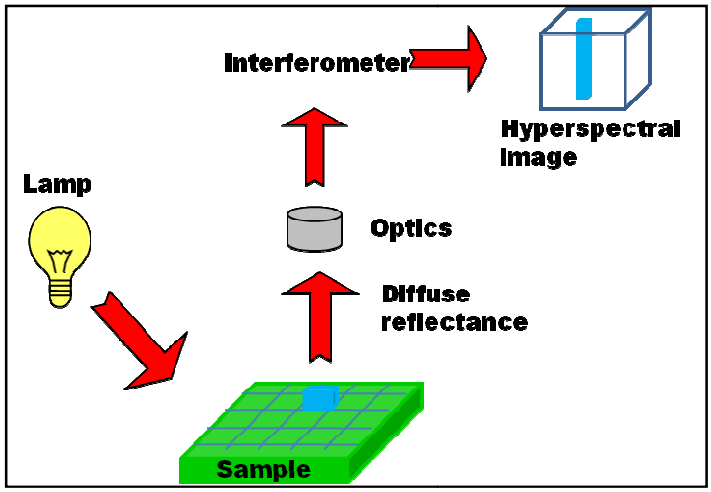


Figure 2.2 Point scan imaging configuration (adapted from Burger, 2006). A hyperspectral image is formed by measuring a spectrum at a single point on the sample after which the sample has to be moved to the next point.

5.1.2 Line scan imaging

The line scan or pushbroom configuration (**Fig. 2.3**) utilises a two dimensional detector positioned perpendicular to the sample stage (Geladi *et al.*, 2007; Gowen *et al.*, 2007). A narrow line of radiation is used to image the sample, or it is imaged through a narrow slit in the optical path leading to the detector (Geladi *et al.*, 2007). This instrument records the spectrum of each pixel in a

line simultaneously (Gowen *et al.*, 2007). By moving the sample scan line, hyperspectral images can be created by collecting sets of these matrices (Burger, 2006; Geladi *et al.*, 2007). This is a high speed imaging system configuration where the only limiting factor is camera read-out speed. This configuration is suitable for a conveyor belt and thus useful in a factory environment (Gowen *et al.*, 2007).

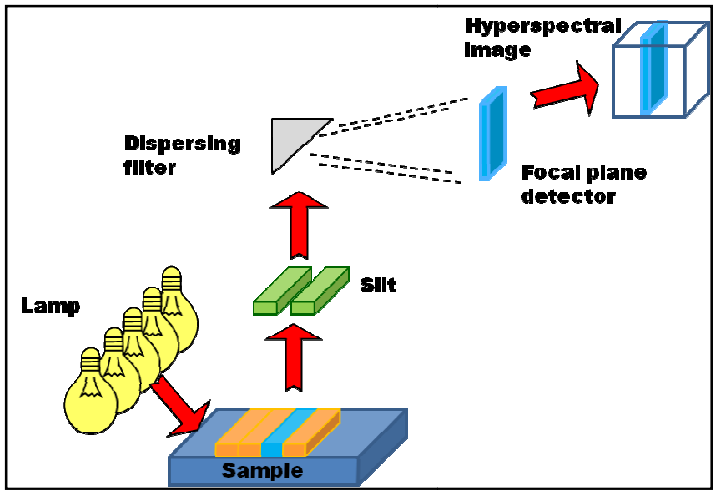


Figure 2.3 Line scan or pushbroom imaging configuration (adapted from Burger, 2006). The hyperspectral image is formed by scanning a line of the sample and then moving the sample scan line.

5.1.3 Focal plane scan imaging

The focal plane scan configuration (**Fig. 2.4**) utilises a detector in a plane parallel to the surface of the sample stage (Geladi *et al.*, 2007). The spectrometer and the sample remain stationary relative to the detector (Burger, 2006; Geladi *et al.*, 2007) while a hyperspectral image is acquired one wavelength at a time (Gowen *et al.*, 2007). This is also a very time-consuming imaging configuration (Burger, 2006; Geladi *et al.*, 2007). The lack of moving parts is advantageous as this can lead to lower background noise level.

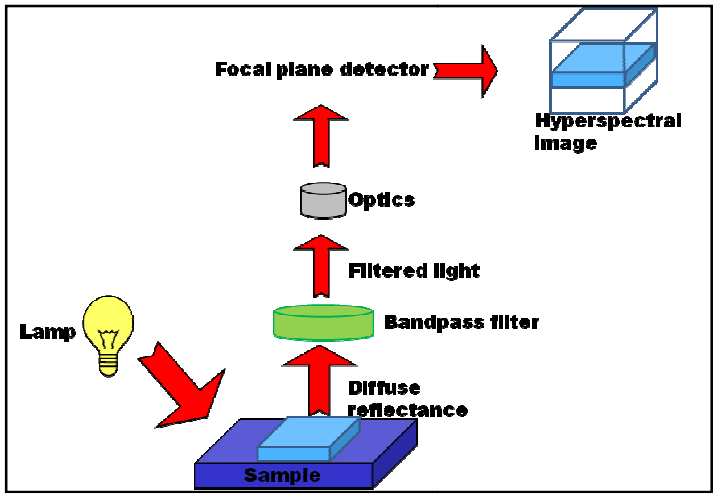


Figure 2.4 Focal plane scan imaging configuration (adapted from Burger, 2006). The hyperspectral image is acquired one wavelength at a time.

5.2 *Image acquisition*

NIR hyperspectral images are sets of images measured at different wavelengths that are progressively stacked (Geladi *et al.*, 2004; Burger & Geladi, 2005; Burger & Geladi, 2006; Gowen *et al.*, 2007). Every pixel in an NIR hyperspectral image contains the complete spectrum for that specific position (Martinsen & Schaare, 1998; Burger & Geladi, 2005; Gowen *et al.*, 2007). As such, NIR hyperspectral images, also known as hypercubes (**Fig. 2.5**), are three-dimensional datasets with two spatial (x and y) and one wavelength dimension (λ) (Cogdill *et al.*, 2004; Burger & Geladi, 2005). Hyperspectral images are often used to describe chemical composition and distribution within samples (Burger & Geladi, 2005; Shahin & Symons, 2008). An NIR hyperspectral imaging system collects large amounts of data, for example, one of the systems used in this thesis collected hypercubes with the dimensions 320(x) x 583(y) x 239(λ). This can be interpreted as 239 single images each comprising 320 x 583 pixels, alternatively, this hypercube can be seen as 186560 spectra each with 239 wavelength bands (Burger & Geladi, 2006). This enormous amount of data creates data analysis problems, however, with the use of chemometric methods relevant information can be extracted (Burger & Geladi, 2006; Wang & Paliwal, 2007; Mahesh *et al.*, 2008).

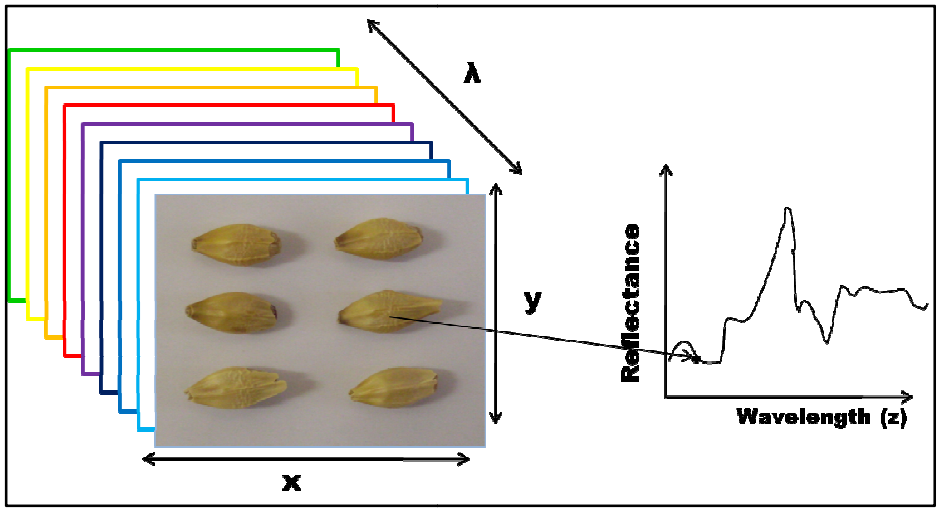


Figure 2.5 A hypercube showing the relationship between two spatial (x and y) and one wavelength (λ) dimension.

6. Chemometrics

The term chemometrics was introduced by Bruce Kowalski and Svante Wold in the early 1970's and can be defined as a discipline that utilises mathematics and statistics to design optimal procedures and experiments to provide relevant information (i.e. concentration and chemical structure) (Massart *et al.*, 1988; Burger, 2006; Wang & Paliwal, 2007; Tan *et al.*, 2008). The introduction of instrumentation, providing multivariate responses (e.g. NIR, Mid-IR and Raman spectrometers) for each sample analysed, stimulated the development of chemometrics; in many cases conventional univariate statistical methods were not adequate (Geladi, 2003). A chemometric approach is typically used to relate physical or chemical properties of examined samples to the absorption of NIR radiation (Bokobza, 1998).

As previously mentioned, NIR spectra are highly complex; NIR spectra contain a vast amount of data and interferences (Bokobza, 1998; Miller, 2001; Williams & Norris, 2001; Woodcock *et al.*, 2008). Chemometric methods are employed to reduce the acquired data into more manageable components for the purpose of extracting relevant information (Woodcock *et al.*, 2008). One such technique is principal component analysis (PCA), and is used to describe the variation within a dataset by means of fewer variables (Cowe & McNicol, 1985).

6.1 *Principal component analysis*

Principal component analysis (PCA) was first proposed in the early 1900's by Pearson, who described the analysis as "finding lines and planes of closest fit to systems of points in space" (Wold *et al.*, 1987). The goals of PCA include: finding classes of similar objects, finding relationships between objects, data dimensionality reduction, and outlier detection (Wold *et al.*, 1987). PCA is applied to transform a dataset into a more compact dataset (fewer variables) without losing the main information contained within the dataset (Bokobza, 1998; Razifar &

Bergstrom, 2007). Principal components (PCs) are calculated to describe the directions of maximum variance within the dataset (Cowe & McNicol, 1985; Bokobza, 1998; Razifar & Bergstrom, 2007). PC1 describes the linear combination of the original data with the greatest variance; PC2 is not correlated with PC1 and describes the linear combination of the greatest remaining variance (Cowe & McNicol, 1985; Wold *et al.*, 1987; Bokobza, 1998; Razifar & Bergstrom, 2007). The following components can be defined as the direction which in turn exhibits the greatest residual variation (Cowe & McNicol, 1985). Every PC is orthogonal to every other PC (Wold *et al.*, 1987; Geladi *et al.*, 1989; Bokobza, 1998; Razifar & Bergstrom, 2007). PCs, thus, explain data variance in decreasing order, for example PC1 = 90%, PC2 = 8%, PC3 = 1%, PC4 = 0.6% and PC5 = 0.4% (Wold *et al.*, 1987; Burger, 2006; Razifar & Bergstrom, 2007).

PCA is the decomposition of **X** (data matrix), usually mean-centred, into a sum of m products of score (t) and inverse loading (p') vectors, and a noise/residual matrix (**E**) (Wold *et al.*, 1987; Geladi *et al.*, 1989; Geladi, 1992):

$$\mathbf{X} = \mathbf{t}_1\mathbf{p'}_1 + \mathbf{t}_2\mathbf{p'}_2 + \mathbf{t}_3\mathbf{p'}_3 + \ldots + \mathbf{t}_m\mathbf{p'}_m + \mathbf{E}$$

where $\mathbf{t}_m$ is the score vector for the m^{th} component and $\mathbf{p}_m$ is the loading vector for the m^{th} component. This equation is schematically depicted in **Fig. 2.6**.

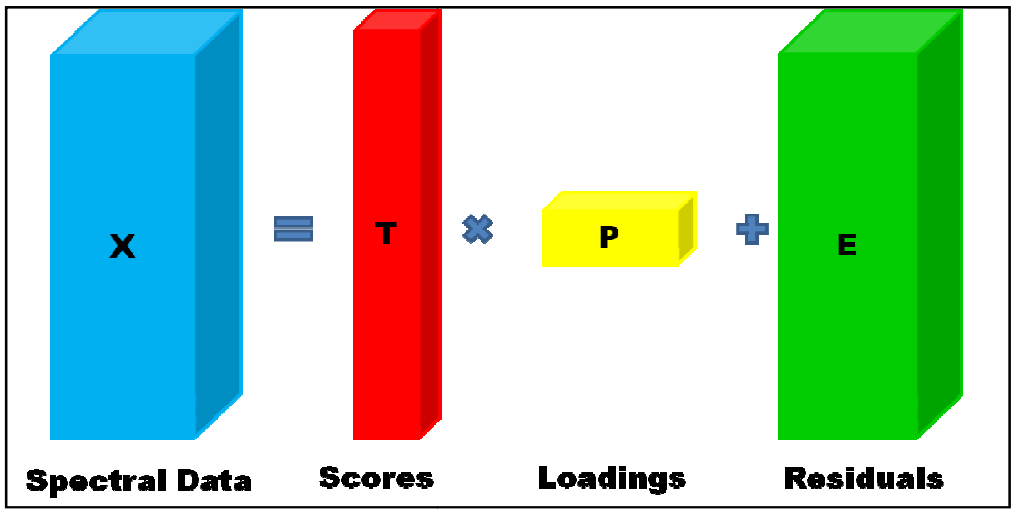


Figure 2.6 The decomposition of a data matrix into scores, loadings and residuals.

Scores and loadings resulting from the decomposition of the spectral data matrix are utilised and viewed as scatter plots and line plots respectively, thereby simplifying interpretation of the data. The scores give an indication of the relationship between the objects and the loadings show variable patterns (Wold *et al.*, 1987). PCA is very useful for the evaluation of NIR hyperspectral images as the scores (objects, samples, or pixels in the image) are observed in score plots and score images, and the most important variables (wavelengths) are observed in loading line plots.

PCA has been successfully applied in various NIR hyperspectral imaging applications: the study of pharmaceutical samples (Amigo *et al.*, 2008), the detection of insect damaged wheat kernels (Singh *et al.*, 2009a), the detection of bruise damage on mushrooms (Gowen *et al.*, 2008) and the detection *Fusarium verticillioides* in maize (Williams *et al.*, 2010).

6.2 *Partial least squares*

The partial least squares (PLS) approach was developed for the modelling of complicated datasets (Wold *et al.*, 2001a). PLS allows the relationships between two matrices to be characterised (Sharaf *et al.*, 1986). The basis of any PLS model is that a small number of latent variables drive the studied system or process; these latent variables are estimated weighted averages of the observed variables (Wold *et al.*, 2001b). Therefore, PLS can be described as a technique for the indirect examination of the latent variables (Wold *et al.*, 2001b).

Partial least squares discriminant analysis (PLS-DA) is a pattern recognition technique that correlates variation in a dataset with class membership (Karp *et al.*, 2005). A dummy matrix (**Y**) that records class membership is paired with a training set (**X**), and a PLS regression model is applied (Barker & Rayens, 2003; Perez-Enciso & Tenenhaus, 2003). The main advantage of a PLS-DA model is that it allows the visualisation and understanding of different patterns and relations in the data; the main sources of variability in the data are modelled by the latent variables, and therefore, in their associated scores and loadings (Alvarez-Guerra *et al.*, 2010). PLS-DA has been employed in fields such as proteomics (Karp *et al.*, 2005), medicine (Perez-Enciso & Tenenhaus, 2003), chemistry (Alvarez-Guerra *et al.*, 2010) and to characterise multivariate images of food and pharmaceutical samples (Chevallier *et al.*, 2006; Amigo *et al.*, 2009).

6.3 *Pre-processing*

Pre-processing refers to any transformation of the original data to enhance the information represented (Sharaf *et al.*, 1986). Before analysis, raw data sometimes has to be pre-processed mathematically in order to remove or reduce certain defects in the spectra, such as noise, baseline shift and light scattering (Bokobza, 1998; Wang & Paliwal, 2007). Pre-processing techniques include standard normal variate transformation, derivation, mean-centering, spectral smoothing algorithms and multiplicative scatter correction.

Standard normal variate (SNV) transformation removes multiplicative interferences such as particle size differences and variations in the baseline promoted by light scattering (Barnes *et al.*, 1989; Amigo *et al.*, 2008). Derivative computation is utilised for the removal of baseline offset and for the resolution of overlapping peaks (Giese & French, 1955; Hruschka, 2001). First and second derivative computation removes baseline effects without losing peak location or intensity information (Burger & Geladi, 2007). Second derivative computation is applied for the removal of sloping baseline effects, and is ideal for the identification of small features. However, this process magnifies the noise in the data (signal-to-noise ratio decreases) and is usually incorporated with a

smoothing procedure (Holler *et al.*, 1989; Bokobza, 1998; Burger & Geladi, 2007; Wang & Paliwal, 2007). Savitzky-Golay smoothing can be employed where data is very noisy and the shape of the spectral band needs to be maintained (Hruschka, 2001; Amigo *et al.*, 2008), and is often used in combination with derivative computation. Multiplicative scatter correction (MSC), as its name implies, corrects light scattering interferences from different sized particles (Geladi *et al.*, 1985). The basis of MSC is the fact that radiation scatter has wavelength dependencies different to chemically absorbed radiation.

7. Hyperspectral image analysis

Hyperspectral image analysis is a technique adapted from the multivariate image analysis strategy to explore hyperspectral images and the information contained in them. The multivariate image analysis strategy was proposed in 1989 by Esbensen and Geladi. They concluded that the multivariate image analysis strategy allows for interactive, visual exploration and classification of multivariate images, making it user-friendly (Esbensen & Geladi, 1989). This step-by-step approach is depicted in **Fig. 2.7**.

Before the analysis of hyperspectral images, the hypercube must be unfolded to produce a large two-dimensional matrix of spectra (Burger, 2006; Amigo *et al.*, 2008). The unfolding step is essential and allows the matrix to be explored using chemometrics and pre-processing techniques such as PCA, PLS-DA and SNV.

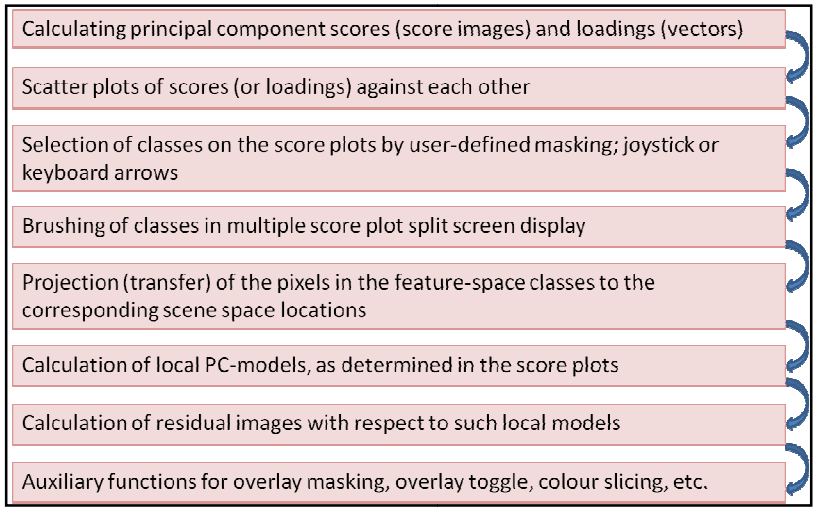


Figure 2.7 The multivariate image analysis strategy (Esbensen & Geladi, 1989).

Hyperspectral image analysis usually begins with a PCA calculation of three or more PCs. This is done with multivariate image analysis software, e.g. Evince (Umbio AB, Umeå, Sweden). PCA score plots, score images and loading line plots are generated by the software and assist in the

cleaning of the images. A score image is a visual representation of the PC score values and a score plot is two score vectors plotted against each other (Esbensen, 2002).

The interactive score plots and score images are used to identify unwanted regions (outliers, dead pixels, shading errors and background constituents) in the image (Manley *et al.*, 2009). By selecting these unwanted regions in the score plot, the unwanted regions in the score image will be highlighted facilitating the effortless cleaning of the images. Different PC combinations can be plotted against each other ensuring the effective removal of the unwanted regions in the image. The selected regions are then excluded from the PCA model and the model is recalculated. A schema of a sequence of hyperspectral image analysis may be found in Manley *et al*. (2009).

Certain pre-processing techniques can result in more prominent clustering when applied to the data. A technique is selected and applied to the data; the resulting score images and score plots are examined for clusters. The identification of the clusters is achieved by selecting a cluster in the score plot and relating that cluster to a specific area in the score image, this process is called brushing (Esbensen & Geladi, 1989). Each cluster in an image is classified accordingly and a classification plot and image is created. The analyst examining the images requires some prior knowledge of the samples, as well as knowledge about the commodity investigated, as this will assist with the identification of the clusters.

Loading line plots are studied to identify the chemical constituents responsible for the variation within the components. Large peaks within loading line plots are considered to have a large contribution to the PC model and are thus responsible for the variation within the sample (Geladi, 1992). The chemical constituents responsible for the variation can be identified by correlating the loading line peaks with group frequency tables (Osborne *et al.*, 1993).

Depending on the desired outcome of the analysis, steps may be added or excluded. Hyperspectral image analysis allows for the interactive investigation and classification of complex hyperspectral images.

8. Conclusion

Cereals are an important source of energy. A loss of cereal crops can be devastating to the agricultural industry of a country. Pre-germination is a relevant threat for the cereal production and utilisation industry, as it decreases grain weight, viability and functionality. The detection of pre-germinated cereal grains has to be accurate as undetected pre-germinated grains can lead to production and economic losses. The conventional methods used are destructive, time-consuming, subjective and produce chemical waste.

Several studies have trialled new techniques for the detection of pre-germination in cereal grains. The potential of X-ray imaging, MRI and NMR to detect pre-germination was shown, however, these techniques are not feasible as MRI and NMR are time-consuming and X-rays pose a potential health. NIR hyperspectral imaging is a more promising technique since this method is

quick, poses no health risks and images are simple to acquire. The chemical changes that occur during pre-germination, hydrolysis of starch and protein, are observable with NIR hyperspectral imaging as starch and protein contain the NIR absorbing bonds (CH, NH, OH, SH and CO). NIR hyperspectral imaging, to detect pre-germination, has been performed on wheat, but only in the wavelength range of 1000 to 1700 nm. The detection of pre-germination with NIR hyperspectral imaging has not been performed on barley or sorghum.

NIR hyperspectral imaging is a fast, effective and non-destructive method that describes sample composition in a spatial manner. The non-destructive nature of this method makes it useful in the grain breeding industry. A possible disadvantage of hyperspectral imaging is the effect of sample shape on imaging and image analysis. Nonetheless, NIR hyperspectral imaging is a promising technique for the detection of pre-germination in cereal grains.

9. References

AACC. (2010). *Approved Methods of the AACC*. Determination of falling number (method 56-81B). St. Paul, Minnesota: American Association of Cereal Chemists.

AACC. (2010). *Approved Methods of the AACC*. Measurement of alpha-amylase activity with the rapid visco analyser (method 22-08). St Paul, Minnesota: American Association of Cereal Chemists.

Alvarez-Guerra, M., Ballabio, D., Amigo, J.M., Bro, R. & Viguri, J.R. (2010). Development of models for predicting toxicity from sediment chemistry by partial least squares-discriminant analysis and counter-propagation artificial neural networks. *Environmental Pollution*, **158**, 607-614.

Amigo, J.M., Cruz, J., Bautista, M., Maspoch, S., Coello, J. & Blanco, M. (2008). Study of pharmaceutical samples by NIR chemical image and multivariate analysis. *Trends in Analytical Chemistry*, **27**, 696-713.

Amigo, J.M., Ravn, C., Gallagher, N.B. & Bro, R. (2009). A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging. *International Journal of Pharmaceutics*, **373**, 179-182.

Asiedu, M., Nilsen, R., Lie, O. & Lied, E. (1993). Effect of processing (sprouting and/ or fermentation) on sorghum and maize. I: proximate composition, minerals and fatty acids. *Food Chemistry*, **46**, 351-353.

Bamforth, C.W. & Barclay, A.H.P. (1993). Malting technology and the uses of malt. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 297-354. St. Paul: American Association of Cereal Chemists, Inc.

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

Barnard, A. (2001). Genetic diversity of South African winter wheat cultivars in relation to pre-harvest sprouting and falling number. *Euphytica*, **119**, 107-110.

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

Bason, M.L., Ronalds, J.A., Wrigley, C.W. & Hubbard, L.J. (1993). Testing for sprout damage in malting barley using the rapid visco-analyser. *Cereal Chemistry*, **70**, 269-272.

Bhatty, R.S. (1993). Non-malting uses of barley. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 355-375. St. Paul: American Association of Cereal Chemists, Inc.

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

Bueckert, R.A., Lefol, E.B. & Harvey, B.L. (2007). Early detection of non-visible sprouting in barley seed using rapid viscosity analysis. *Canadian Journal of Plant Science*, **87**, 3-12.

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

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.

Carn, J.D. (1982). Alpha-amylase indicates problems with malting of stored barley. *Food Technology in Australia*, **34**, 82-83.

Carrari, F., Benech-Arnold, R., Osuna-Fernandez, O., Hopp, E., Sanchez, R., Iusem, N. & Lijavetzky, D. (2003). Genetic mapping of the *Sorghum bicolor vp1* gene and its relationship with preharvest sprouting resistance. *Genome*, **46**, 253-258.

Chen, X., Warner, T.A. & Campagna, D.J. (2007). Integrating visible, near-infrared and short-wave infrared hyperspectral and multispectral thermal imagery for geological mapping at Cuprite, Nevada. *Remote Sensing of Environment*, **110**, 344-356.

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

Coates, J. (1998). Vibrational spectroscopy: instrumentation for infrared and raman spectroscopy. *Applied Spectroscopy Reviews*, **33**, 267-425.

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

Cowe, I.A. & McNicol, J.W. (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. (2006). Chemometrics and visible-near infrared spectroscopy monitoring of red wine fermentation in a pilot scale. *Biotechnology and Bioengineering*, **95**, 1101-1107.

De Alencar Figueiredo, L.F., Davrieux, F., Fliedel, G., Rami, J.F., Chantereau, J., Deu, M., Courtois, B. & Mestres, C. (2006). Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. *Journal of Agricultural and Food Chemistry*, **54**, 8501-8509.

Desai, B.B., Kotecha, P.M. & Salunkhe, D.K. (1997). *Seeds Handbook.* New York: Marcel Dekker, Inc.

Dewar, J., Taylor, J.R.N. & Berjak, P. (1997). Determination of improved steeping conditions for sorghum malting. *Journal of Cereal Science*, **26**, 129-136.

Duffus, C.M. (1987). Physiological aspects of enzymes during grain development and germination. In: *Enzymes: and their role in cereal technology* (edited by J.E. Kruger, D. Lineback & C.E. Stauffer). Pp. 83-111. St. Paul: American Association of Cereal Chemists.

ElMasry, G. & Sun, D.-W. (2010). Principles of hyperspectral imaging technology. In: *Hyperspectral imaging for food quality analysis and control* (edited by D.-W. Sun). Pp. 3-45. San Diego: Academic Press Elsevier.

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

Esbensen, K.H. (2002). *Multivariate data analysis - in practice. An introduction to multivariate data analysis and experimental design.* Oslo: Camo Process AS.

Fernandez-Ibanez, V., Soldado, A., Martinez-Fernandez, A. & De la Roza-Delgado, B. (2009). Application of near infrared spectroscopy for rapid detection of aflatoxin B1 in maize and barley as analytical quality assessment. *Food Chemistry*, **113**, 629-634.

Fincher, G.B. & Stone, B.A. (1993). Physiology and biochemistry of germination in barley. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 247-294. St. Paul: American Association of Cereal Chemists, Inc.

Gao, B.-C., Montes, M.J., Davis, C.O. & Goetz, A.F.H. (2009). Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sensing of Environment*, **113**, S17-S24.

Geladi, P. (1992). Some special topics in multivariate image analysis. *Chemometrics and Intelligent Laboratory Systems*, **14**, 375-390.

Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B*, **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., Isaksson, H., Lindqvist, L., Wold, S. & Esbensen, K. (1989). Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems*, **5**, 209-220.

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

Geladi, P. & Manley, M. (2010). Near-infrared hyperspectral imaging in food research. In: *Raman, infrared and near infrared chemical imaging* (edited by S. Sasic & Y. Ozaki). Pp. 243-260. New Jersey: John Wiley & Sons, Inc.

Geladi, P.L.M., Grahn, H.F. & Burger, J.E. (2007). Multivariate images, hyperspectral images: background and equipment. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P.L.M. Geladi). Pp. 1-15. Chichester: John Wiley & Sons, Ltd.

Giese, A.T. & French, C.S. (1955). The analysis of overlapping spectral absorption bands by derivative spectrophotometry. *Applied Spectroscopy*, **9**, 78-96.

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 and 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 bisporus*). *Journal of Chemometrics*, **22**, 259-267.

Gruwel, M.L.H., Yin, X.S., Edney, M.J., Schroeder, S.W., MacGregor, A.W. & Abrams, S. (2002). Barley viability during storage: use of magnetic resonance as a potential tool to study viability loss. *Journal of Agricultural and Food Chemistry*, **50**, 667-676.

Halverson, J. & Zeleny, L. (1988). Criteria of wheat quality. In: *Wheat: Chemistry and Technology* (edited by Y. Pomeranz). Pp. 15-45. St. Paul: American Association of Cereal Chemists.

Herschel, W. (1800). Experiments in the refrangibility of the invisible rays of the sun. *Philosophical Transactions of the Royal Society of London*, **90**, 284-292.

Hockett, E.A. (2000). Barley. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 81-122. New York: Marcel Dekker Inc.

Holler, F., Burns, D.H. & Callis, J.B. (1989). Direct use of second derivatives in curve-fitting procedures. *Applied Spectroscopy*, **43**, 877-882.

Hoseney, R.C. (1986). *Principles of Cereal Science and Technology.* St. Paul: American Association of Cereal Chemists, Inc.

Hruschka, W.R. (2001). Data analysis: wavelength selection methods. In: *Near-Infrared Technology in the Agricultural and Food Industries* (edited by P. Williams & K. Norris). Pp. 39-58. St. Paul: American Association of Cereal Chemists.

Huang, H., Yu, H., Xu, H. & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *Journal of Food Engineering*, **87**, 303-313.

Juhasz, R., Gergely, S., Gelencser, T. & Salgo, A. (2005). Relationship between NIR spectra and RVA parameters during wheat germination. *Cereal Chemistry*, **82**, 488-493.

Karp, N.A., Griffin, J.L. & Lilley, K.S. (2005). Application of partial least squares discriminant analysis to two-dimensional difference gel studies in expression proteomics. *Proteomics*, **5**, 81-90.

Kent, N.L. & Evers, A.D. (1994). *Kent's technology of cereals.* New York: Elsevier Science Inc.

Koç, H., Smail, V.W. & Wetzel, D.L. (2008). Reliability of InGaAs focal plane array imaging of wheat germination at early stages. *Journal of Cereal Science*, **48**, 394-400.

Kolomiets, O.A., Lachenmeier, D.W., Hoffmann, U. & Siesler, H.W. (2010). Quantitative determination of quality parameters and authentication of vodka using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **18**, 59-67.

Kruger, J.E. (1990). Instrumental assessment of sprout-damage in wheat at primary or terminal receival points. *Cereal Foods World*, **35**, 935-939.

Kruger, J.E. (1994). Enzymes of sprouted wheat and their possible technological significance. In: *Wheat Production, Properties and Quality* (edited by W. Bushuk & V.F. Rasper). Pp. 143-151. Glasgow: Blackie Academic and Professional.

Kruger, J.E. & Reed, G. (1988). Enzymes and colour. In: *Wheat: Chemistry and Technology* (edited by Y. Pomeranz). Pp. 441-487. St. Paul: American Association of Cereal Chemists.

Lawton, J.W. (2000). Non-food uses of cereals. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 725-738. New York: Marcel Dekker, Inc.

Li, C., Ni, P., Francki, M., Hunter, A., Zhang, Y., Schibeci, D., Li, H., Tarr, A., Wang, J., Cakir, M., Yu, J., Bellgard, M., Lance, R. & Appels, R. (2004). Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. *Functional Integrative Genomics*, **4**, 84-93.

Lijavetzky, D., Martinez, M.C., Carrari, F. & Hopp, H.E. (2000). QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. *Euphytica*, **112**, 125-135.

Lin, R., Horsley, R.D. & Schwarz, P.B. (2008). Associations between caryopsis dormancy, alpha-amylase activity and pre-harvest sprouting in barley. *Journal of Cereal Science*, **48**, 446-456.

Lorenz, K. & Kulp, K. (1981). Sprouting of cereal grains- effects on starch characteristics. *Starch/ Starke*, **33**, 183-187.

Lu, G., Huang, H. & Zhang, D. (2006). Prediction of sweetpotato starch physiochemical quality and pasting properties using near infrared reflectance spectroscopy. *Food Chemistry*, **94**, 632-639.

Mahesh, S., Manickavasagan, A., Jayas, D.S., Paliwal, J. & White, N.D.G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, **101**, 50-57.

Manfredini, F., Malagoni, A.M., Felisatti, M., Mandini, S., Mascoli, F., Manfredini, R., Basaglia, N. & Zamboni, P. (2009). A dynamic objective evaluation of peripheral arterial disease by near-infrared spectroscopy. *European Journal of Vascular and Endovascular Surgery*, **38**, 441-448.

Manley, M., Williams, P., Nilsson, D. & Geladi, P. (2009). Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (*Zea maize* L.) kernels. *Journal of Agricultural and Food Chemistry*, **57**, 8761-8769.

Martin, M.E., Wabuyele, M.B., Chen, K., Kasili, P., Panjehpour, M., Phan, M., Overholt, B., Cunningham, G., Wilson, D., DeNovo, R.C. & Vo-Dinh, T. (2006). Development of an advanced hyperspectral imaging (HSI) system with applications for cancer detection. *Annals of Biomedical Engineering*, **34**, 1061-1068.

Martinsen, P. & Schaare, P. (1998). Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology*, **14**, 271-281.

Mason, P.F., Dyson, E.H., Sellars, V. & Beard, J.D. (1994). The assessment of cerebral oxygenation during carotid endarterectomy utilising near infrared spectroscopy. *European Journal of Vascular Surgery*, **8**, 590-594.

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

McClure, W.F. (2001). Near-infrared instrumentation. In: *Near-Infrared Technology in the Agricultural and Food Industries* (edited by P. Williams & K. Norris). Pp. 109-127. St. Paul: American Association of Cereal Chemists, Inc.

Miller, C.E. (2001). Chemical principles of near-infrared technology. In: *Near-Infrared Technology in the Agricultural and Food Industries* (edited by P. Williams & K. Norris). Pp. 19-37. St Paul: American Association of Cereal Chemists, Inc.

Moffat, A.C., Assi, S. & Watt, R.A. (2010). Identifying counterfeit medicines using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **18**, 1-15.

Mohan, A. (2008). Pre-harvest sprouting in cereals: a global scenario. *Current Science*, **94**, 704-705.

Neethirajan, S., Jayas, D.S. & White, N.D.G. (2007). Detection of sprouted wheat kernels using soft x-ray image analysis. *Journal of Food Engineering*, **81**, 509-513.

Nilan, R.A. & Ullrich, S.E. (1993). Barley: taxonomy, origin, distribution, production, genetics and breeding. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 1-30. St. Paul: American Association of Cereal Chemists, Inc.

Oleson, B.T. (1994). World wheat production, utilization and trade. In: *Wheat: Production, Properties and Quality* (edited by W. Bushuk & V.F. Rasper). Pp. 1-10. Glasgow: Blackie Academic & Professional.

Olsson, C. & Thelin, S. (2006). Regional cerebral saturation monitoring with near-infrared spectroscopy during selective antegrade cerebral perfusion: diagnostic performance and relationship to postoperative stroke. *The Journal of Thorarcic and Cardiovascular Surgery*, **131**, 371-379.

Orth, R.A. & Shellenberger, J.A. (1988). Origin, production and utilization of wheat. In: *Wheat: Chemistry and Technology* (edited by Y. Pomeranz). Pp. 1-13. St. Paul: American Association of Cereal Chemists, Inc.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). In: *Practical NIR spectroscopy with applications in food and beverage analysis* (edited by D. Browning). Pp. 1-35, 49-77. Essex: Longman Scientific & Technical.

Pagano, E.A., Benech-Arnold, R.L., Wawrzkiewicz, M. & Steinbach, H.S. (1997). Alpha-amylase activity in developing sorghum caryopses from sprouting resistant and susceptible varieties. The role of ABA and GAs on its regulation. *Annals of Botany*, **79**, 13-17.

Paliwal, S., Baskin, C.C. & Delouche, J.C. (1990). The relationship between the tetrazolium test, soil cold test and standard germination test in assessing sorghum (*Sorghum bicolor* (L.) Moench) seed quality. *Journal of Seed Technology*, **14**, 56-60.

Pasquini, C. (2003). Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, **14**, 198-219.

Patel, A.D., Luner, P.E. & Kemper, M.S. (2000). Quantitative analysis of polymorphs in binary and multi-component powder mixtures by near-infrared reflectance spectroscopy. *International Journal of Pharmaceutics*, **206**, 63-74.

Peirs, A., Scheerlinck, N., De Baerdemaeker, J. & Nicolai, B.M. (2003). Starch index determination of apple fruit by means of a hyperspectral near infrared reflectance imaging system. *Journal of Near Infrared Spectroscopy*, **11**, 379-389.

Perez-Enciso, M. & Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, **112**, 581-592.

Perez-Mendoza, J., Thorne, J.E., Dowell, F.E. & Baker, J.E. (2003). Detection of insect fragments in wheat flour by near-infrared spectroscopy. *Journal of Stored Products Research*, **39**, 305-312.

Peuchant, E., Heches, X., Sess, D. & Clerc, M. (1992). Discriminant analysis of urinary calculi by near-infrared reflectance spectroscopy. *Clinica Chimica Acta*, **205**, 19-30.

Ponte, J.G., Dogan, I.S. & Kulp, K. (2000). Special food ingredients from cereals. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 755-775. New York: Marcel Dekker, Inc.

Posner, E.S. (2000). Wheat. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 1-27. New York: Marcel Dekker, Inc.

Raschke, A.M., Taylor, J. & Taylor, J.R.N. (1995). Use of falling number and rapid visco analyser instruments to estimate sorghum malt diastatic power. *Journal of Cereal Science*, **21**, 97-102.

Rasper, V.F. & Walker, C.E. (2000). Quality evaluation of cereals and cereal products. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 505-535. New York: Marcel Dekker, Inc.

Razifar, P. & Bergstrom, M. (2007). Application of multivariate image analysis in nuclear medicine: principal component analysis (PCA) on dynamic human brain studies with positron emission tomography (PET) for discrimination of areas of disease at high noise levels. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P. Geladi). Pp. 313-3330. Chichester: John Wiley & Sons Ltd.

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

Rennie, W.J. & Gorey, V. (1988). Estimation potential germination of winter wheat and winter barley seed using a tetrazolium test. *Irish Journal of Agricultural Research*, **27**, 179-186.

Rodriguez, E.R., Escalera, A.M.A. & Landin, G.M. (2005). Predicting the chemical composition of sorghum grain by near infrared reflectance spectroscopy (NIRS). *Téc Pecu Méx*, **43**, 1-11.

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.

Roggo, Y., Jent, N., Edmond, A., Chalus, P. & Ulmschneider, M. (2005). Characterizing process effects on pharmaceutical solid forms using near-infrared spectroscopy and infrared imaging. *European Journal of Pharmaceutics and Biopharmaceutics*, **61**, 100-110.

Rooney, L.W. & Serna-Saldivar, S.O. (2000). Sorghum. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 149-171. New York: Marcel Dekker, Inc.

Ross, A.S., Walker, C.E., Booth, R.I., Orth, R.A. & Wrigley, C.W. (1987). The rapid visco-analyser: a new technique for the estimation of sprout damage. *Cereal Foods World*, **32**, 827-829.

Saranwong, S., Sornsrivichai, J. & Kawano, S. (2004). Prediction of ripe-stage eating quality of mango fruit from its harvest quality measures non-destructively by near infrared spectroscopy. *Postharvest Biology and Technology*, **31**, 137-145.

Shahin, M.A. & Symons, S.J. (2008). Detection of hard vitreous and starchy kernels in amber durum wheat samples using hyperspectral imaging. *NIR news*, **19**, 16-18.

Sharaf, M.A., Illman, D.L. & Kowalski, B.R. (1986). *Chemometrics.* New York: John Wiley & Sons.

Shashikumar, K., Hazelton, J.L., Ryu, G.H. & Walker, C.E. (1993). Predicting wheat sprout damage by near-infrared reflectance analysis. *Cereal Foods World*, **38**, 364-366.

Siesler, H.W. (2002). Introduction. In: *Near-Infrared Spectroscopy: Principles, Instruments and Applications* (edited by H.W. Siesler, Y. Ozaki, S. Kawata & H.M. Hiese). Pp. 1-10. Weinheim: WILEY-VCH Verlag GmbH.

Silberberg, M.S. (2000). *Chemistry- the molecular nature of matter and change.* New York: McGraw-Hill.

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

Singh, C.B., Jayas, D.S., Paliwal, J. & White, N.D.G. (2009b). Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry*, **86**, 256-260.

Singh, H., Singh, N., Kaur, L. & Saxena, S.K. (2001). Effect of sprouting on functional and dynamic rheological properties of wheat. *Journal of Food Engineering*, **47**, 23-29.

Skerritt, J.H. & Heywood, R.H. (2000). A five-minute field test for on-farm detection of pre-harvest sprouting in wheat. *Crop Science*, **40**, 742-756.

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.

Storme-Paris, I., Clarot, I., Esposito, S., Chaumeil, J.C., Nicolas, A., Brion, F., Rieutord, A. & Chaminade, P. (2009). Near infrared spectroscopy homogeneity evaluation of complex powder blends in a small-scale

pharmaceutical preformulation process, a real-life application. *European Journal of Pharmaceutics and Biopharmaceutics*, **72**, 189-198.

Storme-Paris, I., Rebiere, H., Matoga, M., Civade, C., Bonnet, P.-A., Tissier, M.H. & Chaminade, P. (2010). Challenging near infrared spectroscopy discriminating ability for counterfeit pharmaceuticals detection. *Analytica Chimica Acta*, **658**, 163-174.

Tan, S.T., Chen, K., Ong, S. & Chew, W. (2008). Utilization of spectral vector properties in multivariate chemometrics analysis of hyperspectral infrared imaging data for cellular studies. *Analyst*, **133**, 1395-1408.

Tergorg, C., Birkner, T., Schack, B. & Witte, O.W. (2002). Acute effects of cigarette smoking on cerebral oxygenation and hemodynamics: a combined study with near-infrared spectroscopy and transcranial doppler sonography. *Journal of Neurological Sciences*, **205**, 71-75.

Vo-Dinh, T., Stokes, D.L., Wabuyele, M.B., Martin, M.E., Song, J.M., Jagannathan, R., Michaud, E., Lee, R.J. & Pan, X. (2004). A hyperspectral imaging system for in vivo optical diagnostics. *IEEE Engineering in medicine and biology magazine*, **23**, 40-49.

Wang, W. & Paliwal, J. (2007). Near infrared spectroscopy and imaging in food quality and safety. *Sensing and Instrumentation for Food Quality and Safety*, **1**, 193-207.

Weinstock, B.A., Janni, J., Hagen, L. & Wright, S. (2006). Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Applied Spectroscopy*, **60**, 9-16.

Williams, P., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta*, **653**, 121-130.

Williams, P., Manley, M., Fox, G. & Geladi, P. (2010). Indirect detection of *Fusarium verticillioides* in maize (*Zea mays* L.) kernels by near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **18**, 49-58.

Williams, P.C. & Norris, K. (2001). Variables affecting near-infrared spectroscopic analysis. In: *Near-Infrared Technology in the Agricultural and Food Industries* (edited by P. Williams & K. Norris). Pp. 171-185. St. Paul: American Association of Cereal Chemists, Inc.

Wold, S., Esbensen, K. & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**, 37-52.

Wold, S., Sjostrom, 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.

Woodcock, T., Downey, G. & O'Donnell, C.P. (2008). Better quality food and beverages: the role of near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **16**, 1-29.

Woonton, B.W., Jacobsen, J.V., Sherkat, F. & Stuart, I.M. (2005). Changes in germination and malting quality during storage of barley. *Journal of the Institute of Brewing*, **111**, 33-41.

Workman, J. & Schenk, J. (2004). Understanding and using the near infrared spectrum as an analytical method. In: *Near infrared Spectroscopy in Agriculture* (edited by L. Al-Amoodi). Pp. 3-10. Madison: American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.

Xing, J., Symons, S., Shahin, M. & Hatcher, D. (2010). Detection of sprout damage in Canada western red spring wheat with multiple wavebands using visible/ near infrared hyperspectral imaging. *Biosystems Engineering*, **106**, 188-194.

Zeaiter, M., Roger, J.M. & Bellon-Maurel, V. (2006). Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations. *Chemometrics and Intelligent Laboratory Systems*, **80**, 227-235.

Chapter 3

Detection of pre-germination in whole barley kernels with near infrared (NIR) hyperspectral imaging and image analysis

Chapter 3

Detection of pre-germination in whole barley kernels with near infrared (NIR) hyperspectral imaging

Abstract

Pre-germinated barley can not be used in the malting industry; such grains are down-graded with subsequent losses for the producer. The potential of a near infrared (NIR) hyperspectral imaging system for the detection of pre-germination in whole barley grains was evaluated. Four different barley cultivars containing pre-germinated grains were imaged using SisuCHEMA hyperspectral imaging systems (1000-2498 nm). Images with two different spatial resolutions (150 µm and 30 µm) were captured. Principal component analysis (PCA) with six components was performed on standard normal variate (SNV) transformed and mean-centered data. The acquired images were cleaned (removal of outliers, shading and background). A distinction between pre-germinated and non pre-germinated kernels was observed in the direction of principal component (PC) 5 for the 150 µm spatial resolution images and in the direction of PCs 2, 5 and 4 for the respective 30 µm spatial resolution images. The loading line plots indicated water, starch, protein and glucose as the compounds contributing to the spectral variation. During pre-germination, starch is degraded to oligosaccharides and glucose and proteins are degraded to peptides and amino acids. Water is required for the hydrolysis of starch and proteins. A partial least squares discriminant analysis (PLS-DA) model was developed from the 150 µm spatial resolution images for each of the cultivars (Puma, Erica and SSG564). The PLS-DA model for Erica had the highest explained variation in Y after six components and was used to predict the two other cultivar images. The prediction results indicated 34.57% false positives for Puma and 45.38% false negatives for SSG564. Two PLS-DA models were developed for one image of each cultivar (Clipper, Puma and SSG564) of the 30 µm spatial resolution images. The two models differed on the basis of pixel assignment; one was based on the score image and the other the score plot. Both of these models were used to predict two other independent images of the same cultivar. Overall model 1 (based on score image assignment) resulted in excessive false negatives and model 2 (based on score plot assignment) excessive false positives; neither of the models performed adequately.

Introduction

Barley (*Hordeum* vulgare L.) is primarily used for malting purposes, human food and animal feed (Nilan & Ullrich, 1993). Malt is usually made from barley; other cereals such as sorghum, wheat and rye can be malted as well, but this is done to a lesser extent (Bamforth & Barclay, 1993; Pyler & Thomas, 2000). Barley malt provides distinctive flavours and colour to a range of foodstuffs, such as biscuits, breads, breakfast cereals, cakes and sweets. Barley malt is mainly used as a source of fermentable sugars in the production of beer and whisky.

When mature barley is subjected to wet or humid conditions, pre- or post-harvest, germination may be triggered leading to pre-germination of the grain (Gruwel *et al.*, 2002; Gualano & Benech-Arnold, 2009). Grains may lose viability and functionality. The subsequent loss of viability is of great concern to the barley malting industry (Gruwel *et al.*, 2002), as the malting process requires rapid and full germination of the grain to produce good quality malt (Li *et al.*, 2004). The malting industry will reject seed lots containing pre-germinated barley (Gualano & Benech-Arnold, 2009). The rejection of barley and the subsequent down-grading to feed-grade results in economic losses for the producers as well as the barley utilisation industry (Gruwel *et al.*, 2002; Li *et al.*, 2004). Pre-germinated grains have an increased level of soluble sugars (the result of starch degradation by alpha-amylase) which favours the growth of fungi and insects (Pagano *et al.*, 1997; Singh *et al.*, 2009). This can lead to cross-contamination when pre-germinated grains are mixed with sound grains.

The detection of pre-germinated grains has to be accurate as an over-estimation can lead to the rejection of sound grains and under-estimation can lead to the utilisation of non-viable grains and subsequent production losses. Methods such as visual inspection, viscosity analysis, and the tetrazolium test are used to detect pre-germination in whole barley kernels (Matsuo *et al.*, 1982; Bueckert *et al.*, 2007). Visual inspection is a subjective method based on the knowledge and experience of the grader. Visual assessment cannot represent the actual degree of damage caused by pre-germination, as some damage occurs before it is visible (Carn, 1982; Matsuo *et al.*, 1982). Viscosity analysis, the determination of falling number and the tetrazolium test are time consuming and destructive methods (Smail *et al.*, 2006; Singh *et al.*, 2009). An alternative method for the detection of pre-germination in wheat was proposed, namely near infrared (NIR) hyperspectral imaging (Smail *et al.*, 2006; Koç *et al.*, 2008; Singh *et al.*, 2009). NIR hyperspectral imaging generates no chemical waste and requires minimal sample preparation; this may be beneficial in the barley malting industry.

A hyperspectral image (hypercube) is a three dimensional block of data with two spatial (x and y) and one wavelength (λ) dimension (Cogdill *et al.*, 2004). Such an image contains a NIR spectrum for each pixel in the image and is a very powerful tool in the evaluation of biological samples (Geladi *et al.*, 2004; Burger & Geladi, 2006). Hyperspectral images may describe the distribution of constituents within a sample (Burger & Geladi, 2005; Shahin & Symons, 2008). Chemometric techniques such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) can be used for the extraction of relevant information (e.g. sample classifications and quantitative determinations) from hyperspectral images (Geladi *et al.*, 1989; Barker & Rayens, 2003; Burger & Geladi, 2006).

NIR hyperspectral imaging has been evaluated within the agricultural and food industries. The measurement of soluble solids in kiwifruit (Martinsen & Schaare, 1998), differentiation between wheat classes (Mahesh *et al.*, 2008), starch index determination in apples (Peirs *et al.*, 2003), the evaluation of fresh mushroom shelf-life stored in different packaging films (Taghizadeh *et al.*, 2010)

and the classification of maize kernel hardness (Williams *et al.*, 2009) are a few examples of possible NIR hyperspectral imaging applications.

The aim of this research was to evaluate NIR hyperspectral imaging as a non-destructive and objective alternative to conventional pre-germination detection methods, and to discriminate between pre-germinated and non pre-germinated kernels using partial least squares discriminant analysis.

Materials and methods

Samples and sample preparation

Barley samples from commercial cultivars were kindly provided by SAB Maltings (Pty) Ltd. (Caledon, South Africa). Two sets of cultivar samples were used; the first included Puma, Erica and SSG564 and was used to collect 150 µm spatial resolution images, the second included Clipper, Puma and SSG564 and was used to collect 30 µm spatial resolution images. From the first cultivar set six subsets, comprising 25 kernels each, were randomly selected from each of the 3 cultivars. The subsets were placed in Petri-dishes lined with 2 layers of Whatman no.1 filter paper. Distilled water (3 mL) was added and the Petri-dishes were incubated at 19°C for 6, 9, 12, 18 and 24 hrs, respectively. One subset from each cultivar was not treated with water or incubated and kept as a control (0 hrs). An extra 0 and 24 hrs subset (containing 25 kernels and 3 mL distilled water) was included for each cultivar. These subsets were used for moisture analysis.

Each cultivar of the second set was divided into seven subsets, each containing 100 randomly selected kernels. The subsets were placed in Petri-dishes with 2 layers of Munktell (Grade: 3hw) filter paper and 5 mL of distilled water. The subsets were incubated for 3, 6, 9, 12, 18 or 24 hrs at 19°C. A set of control (0 hrs) samples was also included.

After incubation, the samples (including the controls) were frozen at -80°C for 24 hrs and freeze dried (Virtis, Benchtop 6.6, The Virtis Company, Gardiner, United States of America) for 72 hrs. After freeze drying all the samples were vacuum sealed.

Near infrared hyperspectral imaging system

Near infrared hyperspectral images were acquired using two different SisuCHEMA short wave infrared (SWIR) hyperspectral imaging systems (Specim, Spectral Imaging Ltd, Oulo, Finland). Both SisuCHEMA systems consisted of an imaging spectrograph coupled with a 2-D array mercury-cadmium-telluride (HgCdTe) detector. The first system was used to collect 150 µm spatial resolution images (pixels of 150 x 150 µm size). The lens had a 50 mm wide field of view and the system could image a maximum length of 100 mm. Spectra were acquired from 1000 to 2498 nm with a 6.3 nm interval, resulting in images with the maximum dimensions of 320 (x) x 583 (y) x 239 (λ). The second system was used to collect 30 µm spatial resolution images (pixels of 30 x 30 µm size). The lens had a 10 mm wide field of view and the system could image a maximum length of 200 mm. Spectra were acquired from 996 to 2498 nm with a 6.2 nm interval, resulting in images

with maximum dimensions of 320 (x) x 4302 (y) x 242 (λ). Internal dark and white reference standards were used for image calibration.

Image acquisition

The 150 µm and 30 µm spatial resolution images were acquired by positioning the samples (within a cultivar) crease down, germ to the right and in random rows (according to time) on the silicon carbide (SiC) sandpaper lined sample stage of the SisuCHEMA. This was repeated for each cultivar (**Fig. 3.1**). Digital images were acquired using a Samsung S760 camera (Samsung Opto-Electronics America, Inc., Secaucus, United States of America). A 150 µm spatial resolution image consisted of 5 columns and 12 rows (10 randomly selected kernels per subset). A 30 µm spatial resolution image consisted of 1 column and 21 rows (3 randomly selected kernels per subset) (**Fig. 3.2**). Due to mould damage, the 18 hrs subset of Clipper was not imaged.

The SisuCHEMA detector captures the spectrum of each pixel in a row simultaneously; this forms an image with one spatial and one wavelength dimension. The motorised sample stage is then moved to record the neighbouring row of pixels forming another two-dimensional image. This is done for the whole length of the image. These two-dimensional images are then stacked to form a three-dimensional hypercube with two spatial and one wavelength dimension.

Hyperspectral image analysis

Image cleaning

Within a cultivar the randomly imaged subsets were rearranged chronologically for ease of visualisation. This was done by importing an image into Evince version 2.4.0 (Umbio AB, Umeå, Sweden) multivariate image analysis software and saving each subset in the image separately. These separately saved files were imported into Evince again and merged chronologically (0 to 24 hrs) to form a multi-import image. The reorganised image was transformed from reflectance to pseudo-absorbance in Evince. Principal component analysis (PCA) with six components was used on mean-centered data to clean the images. This involved the utilisation of PCA score plots and PCA score images interactively to identify and classify unwanted pixels such as outliers, background, dead pixels, shading errors and edge effects. The unwanted pixels were removed by selecting these pixels in the score plot space and excluding them from further calculations. PCA scores and loadings were recalculated and the cleaned images were subjected to further image analysis.

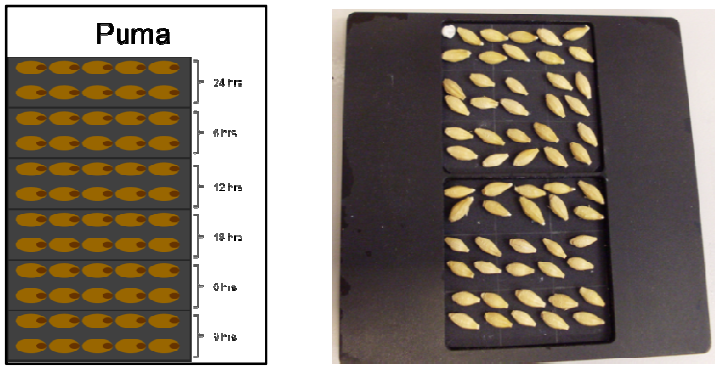
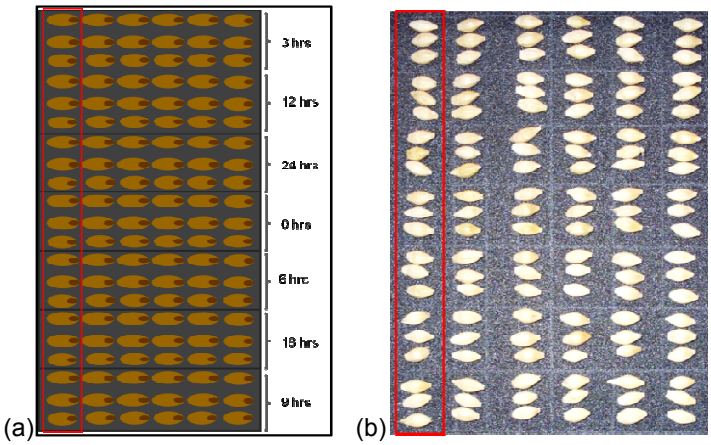


Figure 3.1 Graphical representation (left) and digital photograph (right) of the 150 µm spatial resolution images for Puma, imaged on the SisuCHEMA imaging system. The samples were positioned crease down and germ to the right in randomised rows.



(a) (b)

Figure 3.2 (a) Graphical representation of the 30 µm spatial resolution images, one image is shown with a red rectangle; and (b) a digital photograph of the samples positioned on the sandpaper lined display board, one image is highlighted in red.

Analysis of cleaned images

Standard normal variate (SNV) transformation, a mathematical transformation that removes multiplicative interferences such as scattering and particle size differences (Barnes *et al.*, 1989), was performed. Principal component (PC) score images, score plots and loading line plots were studied.

Partial least squares discriminant analysis

Partial least squares discriminant analysis (PLS-DA) was performed to discriminate between pre-germinated and non pre-germinated kernels. PLS-DA prediction results were evaluated in terms of false positives and false negatives. A false positive occurs when a sample which does not belong to a certain class is predicted as a member of that class (Contal *et al.*, 2002). A false negative arises when a sample which does belong to a given class is not classified as such.

150 µm spatial resolution images

PLS-DA was performed on the cleaned and rearranged images of each cultivar in Evince image analysis software. A PLS-DA model with two classes (pre-germinated and non pre-germinated) was developed for each cultivar (Puma, Erica and SSG564). These two classes were assigned in the score plot, i.e. the pixels associated with each class were assigned accordingly. Only the PLS-DA model with the highest explained variation in Y was used for classification. This model was validated with the images of the other two cultivars.

30 µm spatial resolution images

Two different PLS-DA models were developed for one image (cleaned and reorganised) of each cultivar (Clipper, Puma and SSG564). The first model (M1) was developed by assigning each kernel in the score image to either the pre-germinated or non pre-germinated class as determined by the tetrazolium test. The model was used to predict two other independent images of the same cultivar.

The second PLS-DA model (M2) was created in the same manner as the 150 µm hyperspectral images. The pixels in the score plot related to either the pre-germinated or non pre-germinated class were classified as such. This model was used to predict the same images as the first model.

Barley viability

For the 150 µm spatial resolution images a viability test (tetrazolium test) was performed on all 25 kernels in each of the subsets. The tetrazolium test results for the 150 µm spatial resolution images were given as an average result, i.e. the percentage of viable kernels within the total of 25 kernels. The tetrazolium test for the 30 µm spatial resolution images was performed on only the imaged kernels (18) in each subset. The position of each kernel in the image was noted and each individual kernel was tested. Therefore, each kernel in the image was classified as either viable or non-viable.

A 1% tetrazolium chloride solution was prepared by dissolving 10 g (± 0.01 g) of 2,3,5-triphenyl-tetrazolium chloride (Merck (Pty) Ltd, Germiston, South Africa) in 1000 mL distilled water (Purite Select, Purite Limited, Oxon, United Kingdom). The kernels from a subset of the 150 µm spatial resolution images were cut longitudinally. The halved kernels were placed in a test tube and it was repeated for all the subsets of the 150 µm spatial resolution images. Each kernel from the 30 µm spatial resolution images was cut independently, and both halves were placed in a marked (for identification) test tube. The kernels were covered with the tetrazolium chloride solution and kept at 40°C for 30 min. The excess solution was drained and the kernels were classified as follows: kernels in which the embryo, scutellum and aleurone layer were stained a bright pink-red colour were fully viable and kernels in which the scutellum or aleurone layer were unstained, or where the scutellum, aleurone layer and embryo were unstained were non-viable.

Moisture content

Moisture content was determined according to the moisture - modified vacuum-oven method (AACC method 44-40, 2010). Kernels from the 0 and 24 hrs subsets of the first cultivar set (150 µm spatial resolution images) were milled with a coffee bean grinder (Mellerware, aromatic coffee bean grinder, model 29105, Cape Town, South Africa) and kernels from the 0 – 24 hrs subsets included in the second cultivar set (30 µm spatial resolution images) were milled using a Cyclone sample mill (UD Corporation, Boulder, Colorado, United States of America). Two grams of ground sample from each subset were weighed and dried for five hours. The samples were left to cool in a desiccator before moisture content was determined gravimetrically (Precisa balance, model 205 SCS, Milton Keynes, United Kingdom).

Results and discussion

Image cleaning

Puma (150 µm spatial resolution image) was chosen to demonstrate the identification and removal of unwanted pixels. Principal component analysis with mean-centering was performed and the total sum of squares (SS) of six components amounted to 99.78%. Score plots and score images were created for each of the six principal components (PCs). **Fig. 3.3** illustrates the uncleaned Puma score plot of PC 1 (SS 99.1%) vs. PC 4 (SS 0.031%) with 168720 pixels and the uncleaned score image of PC 1. These score images and score plots were used interactively to investigate and identify regions of interest within the dataset. The exploration of the score images and score plots facilitated the swift identification of unwanted pixels (background, shading errors, outliers and dead pixels). These unwanted pixels was selected in the score plot and projected onto the score image, creating a classification image. The difference in chemical composition lead to the identification of three clusters, i.e. SiC sandpaper, adhesive marker and barley kernels. **Fig. 3.4** illustrates the classification plot and image of Puma, where the identified clusters are shown. The removal of unwanted pixels was simplified by the Evince software; the software allows the user to exclude unwanted pixels from further calculations with the press of a button. After the removal of the unwanted pixels, PCA scores and loadings were recalculated and the cleaned score image and score plot was used for subsequent analyses.

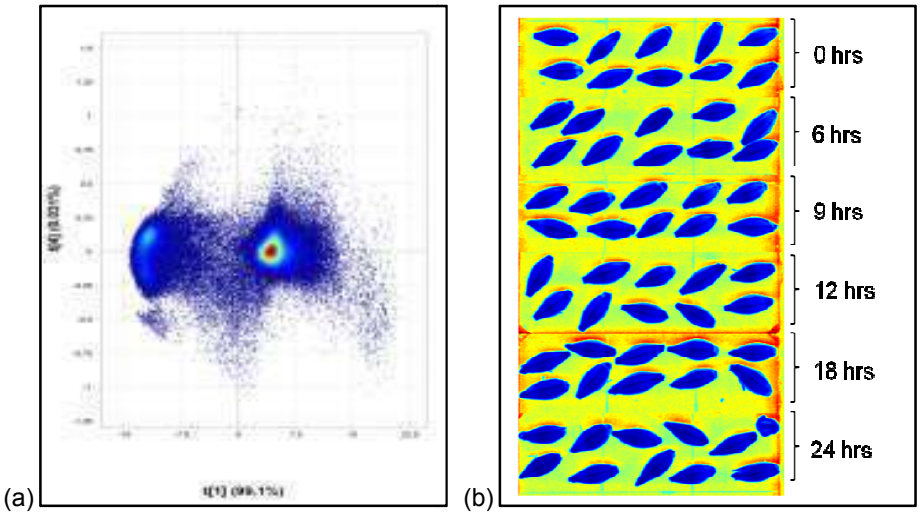


Figure 3.3 (a) Uncleaned Puma PCA score plot (PC 1 vs. PC 4) displaying three clusters; and (b) uncleaned PCA score image (PC 1).

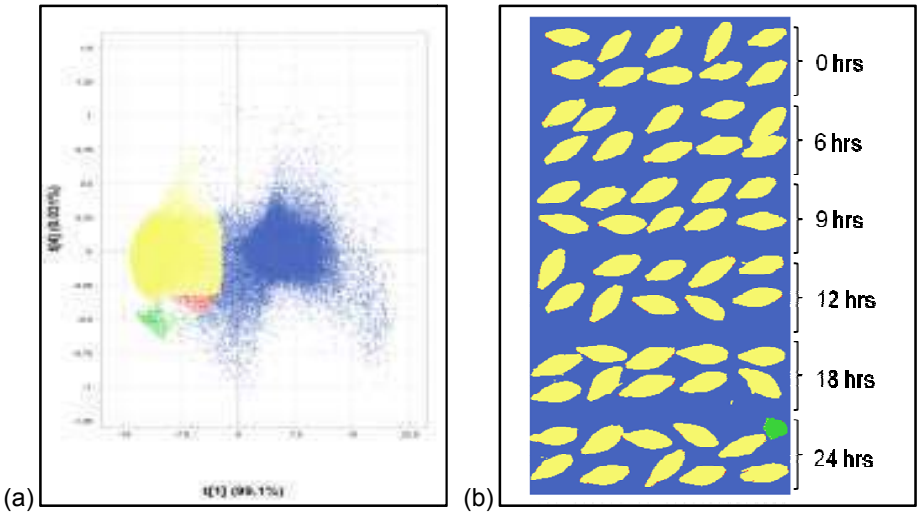


Figure 3.4 (a) Puma PCA score plot (PC 1 vs. PC 4) with selected areas projected onto the score image to create a (b) classification image (PC 1) (blue = background, yellow = barley kernels, green = adhesive marker and red = not classified).

Analysis of cleaned 150 µm spatial resolution hyperspectral images

Puma

After the removal of unwanted pixels, the PCA score images and score plots (42653 remaining pixels) were analysed for regions of interest. Different combinations of PCs (i.e. PC 1 vs. PC2; PC 1 vs. PC 3; ....; PC 5 vs. PC 6) were inspected for clustering. The combination of PC 1 and PC 5 (SS 1.53%) revealed no well defined clusters in the score plot (**Fig. 3.5a**); however, the score image (**Fig. 3.5b**) revealed a difference between the 0 hrs and 24 hrs subsets. These differences can be observed through the use of heat map colouring in the score image. Pixels with similar score values are indicated with similar colours, in this case indicating similarity in chemical composition. When correlated with the tetrazolium test (**Table 3.1** and **Fig. 3.5b**) the differences between the subsets was attributed to pre-germinated and non pre-germinated kernels; where the

0 and 6 hrs subsets include mostly non pre-germinated kernels and the 12 to 24 hrs subsets include mostly pre-germinated kernels. Kernels from the 9 hrs subset included kernels that were pre-germinated (similar in chemical composition to 24 hrs subset) and non pre-germinated (similar to 0 hrs subset). The PC 1 vs. PC 5 score plot (**Fig. 3.5a**) of Puma revealed (with interactive exploration) that positive PC 5 score values were associated with pixels of the non pre-germinated kernels. The negative score values in the score plot were associated with the pre-germinated kernels in the score image. This is better illustrated in the classification plot (**Fig. 3.6a**) and image (**Fig. 3.6b**).

The tetrazolium test results for each subset of all three cultivars were given as an average percentage (**Table 3.1**). Consequently, not only the kernels in the image but also the kernels not selected for the imaging influence the percentage viability. This could lead to a high percentage of viability for the total subset, but the image reveals pre-germinated kernels.

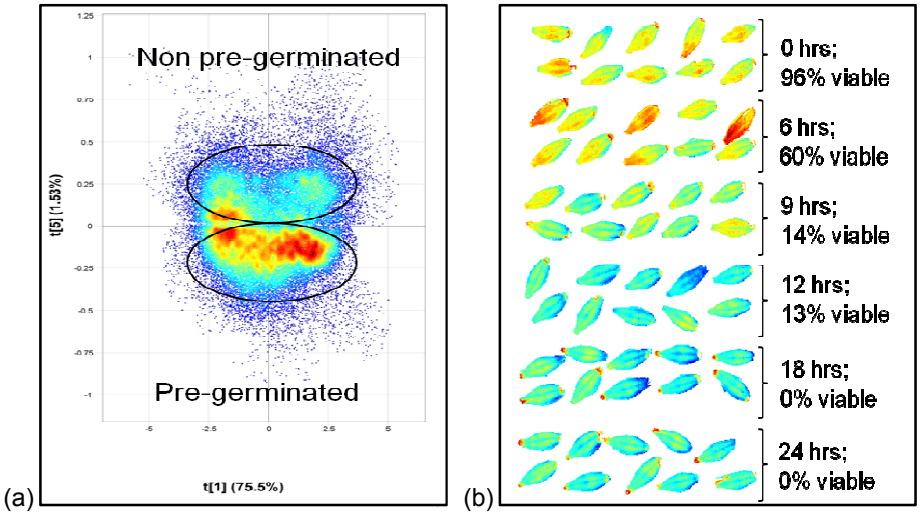
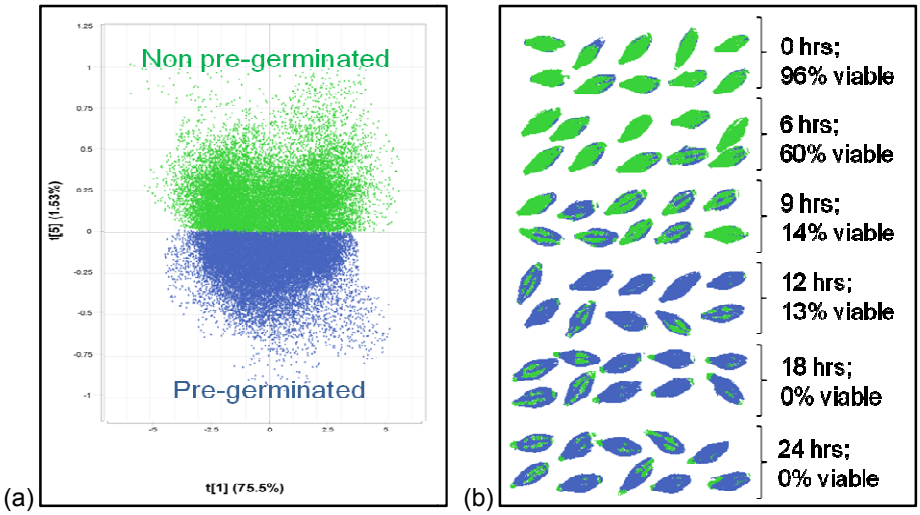


Figure 3.5 (a) Puma PCA score plot of PC 1 vs. PC 5 indicating non pre-germinated and pre-germinated pixels; and (b) PCA score image of PC 5 showing differences between the pre-germinated (9 - 24hrs) and non pre-germinated (0 and 6 hrs) subsets.



(a) (b)

Figure 3.6 (a) Puma classification plot of PC 1 vs. PC 5; and (b) classification image of PC 5 (green = non pre-germinated and blue = pre-germinated).

Table 3.1 Tetrazolium test results as percentage viability of the subsets

Cultivar	Subsets (time exposure to distilled water in hrs)	Viability (%)
Puma	0	96
Puma	6	60
Puma	9	14
Puma	12	13
Puma	18	0
Puma	24	0
Erica	0	100
Erica	6	80
Erica	9	20
Erica	12	12
Erica	18	4
Erica	24	0
SSG564	0	100
SSG564	6	74
SSG564	9	60
SSG564	12	52
SSG564	18	10
SSG564	24	8

Erica – 150 µm spatial resolution image

The PCA score images and score plots of Erica were investigated for differences between pre-germinated and non pre-germinated kernels. The PCA score plot (**Fig. 3.7a**) of PC 1 (SS 79.0%)

vs. PC 5 (SS 0.874%) revealed four clusters. The clusters in the direction of PC 1 were not attributable to the difference between pre-germinated and non pre-germinated kernels, and are therefore not discussed. The location of the pixels together with the determined viability (**Fig. 3.7b)** confirmed that pixels on the positive side of PC 5 were non pre-germinated and pixels on the negative side were pre-germinated (**Fig. 3.7a**). The kernels of the 0 hrs subset, in the PC 5 score image (**Fig. 3.7b**), were identified as non pre-germinated (100% viable) and the kernels of the 24 hrs subset were identified as pre-germinated (0% viable). The differences between these two subsets provided the means to categorize the kernels within a subset. Thus, kernels from the 6 and 9 hrs subset were identified as non pre-germinated (similarity in chemical composition to 0 hrs subset according to the score image) and the kernels from the 12 and 18 hrs subset were mostly pre-germinated. The classification plot (**Fig. 3.8a**) and image (**Fig. 3.8b**) were created to better visualise the pre-germinated and non pre-germinated kernels.

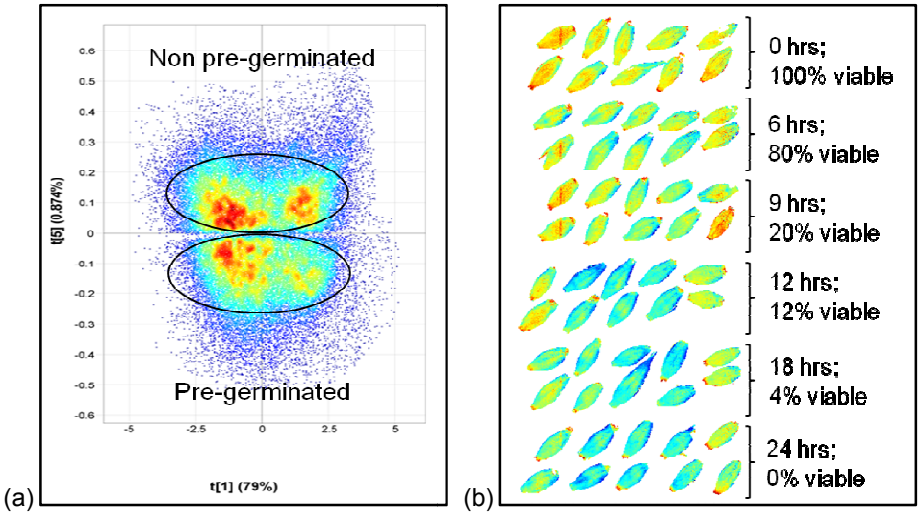
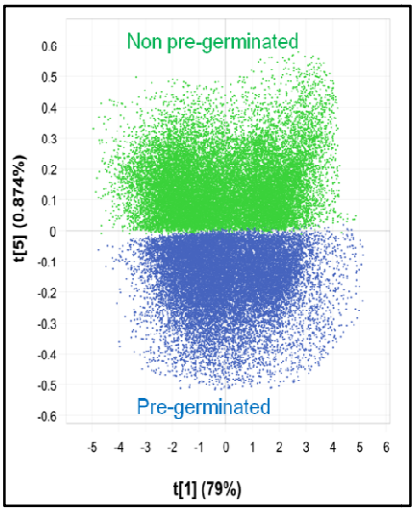


Figure 3.7 (a) Erica PCA score plot of PC 1 vs. PC 5 indicating non pre-germinated and pre-germinated pixels; and (b) PCA score image of PC 5 showing differences between the non pre-germinated (0 – 9 hrs) and pre-germinated (12 – 24 hrs) subsets.

(a) 

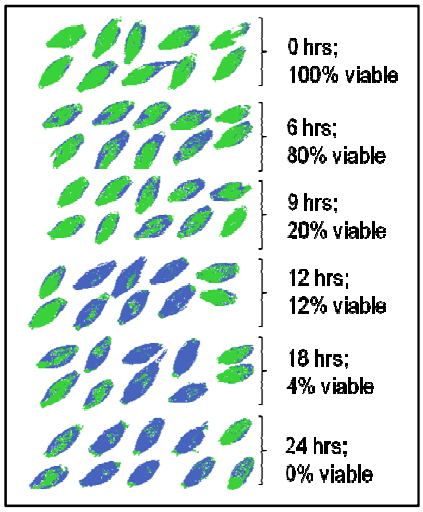
(b) 

Figure 3.8 (a) Erica classification plot of PC 1 vs. PC 5; and (b) classification image of PC 5 (blue = pre-germinated and green = non pre-germinated).

SSG564 – 150 µm spatial resolution images

Image analysis of the cultivar SSG564 produced similar results to the Puma and Erica examples. The location of pixels in the score plot (**Fig 3.9a**) and the determined viability of kernels in the score image (**Fig. 3.9b**) provided the means to identify pixels in the score plot. Pixels on the positive side of PC 5 were identified as non pre-germinated and pixels on the negative side were identified as pre-germinated. The PCA score image of PC 5 (SS 0.782%) illustrated a difference between the different subsets. The 0 hrs subset (100% viable) was identified as non pre-germinated and the 24 hrs subset (8% viable) was identified as pre-germinated. Thus, the kernels from the 0, 6 and 9 hrs subsets were non pre-germinated according to the score image and the kernels from the 12, 18 and 24 hrs subsets were pre-germinated. A classification plot (**Fig. 3.10a**) and classification image (**Fig. 3.10b**) were formed by selecting different clusters in the PCA score plot and projecting it onto the PCA score image.

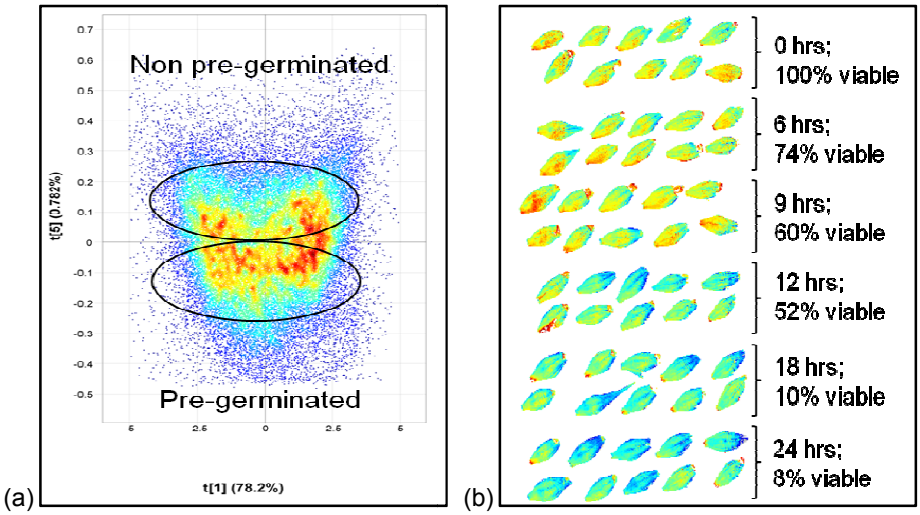


Figure 3.9 (a) SSG564 PCA score plot of PC 1 vs. PC 5 showing the pre-germinated and non pre-germinated pixels; and (b) PCA score image of PC 5 showing the different subsets (0 – 9 hrs = non pre-germinated and 12 – 24 hrs = pre-germinated).

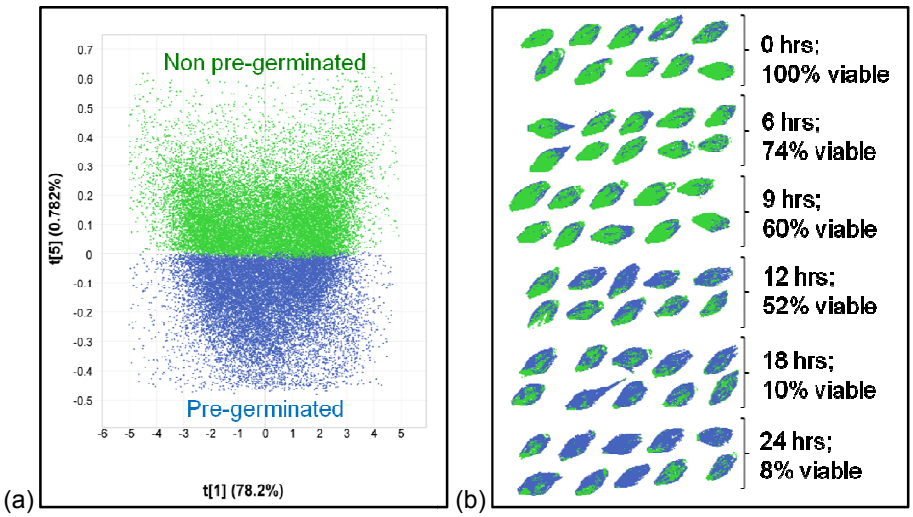


Figure 3.10 (a) SSG564 classification plot of PC 1 vs. PC 5; and (b) classification image of PC 5 (green = non pre-germinated and blue = pre-germinated).

Loading line plots – 150 µm spatial resolution images

Sample differentiation within a dataset may be better understood by examination of loading line plots. A loading line plot indicates which variables contribute most to the respective PC. The PC 5 loading line plots of all three cultivar PCAs were similar, thus only the PC 5 loading line plot of Puma is shown in **Fig. 3.11**.

The major peak observed at 1940 nm is related to water (O-H stretching and O-H deformation) (Osborne *et al.*, 1993). This peak is negatively weighted in PC 5 and can be associated with the pre-germinated kernels. The difference between pre-germinated and non pre-germinated kernels was not simple a function of moisture content as the moisture content of the 0 and 24 hrs subsets were similar (**Table 3.2**). Rather, the water environment differed between pre-

germinated and non pre-germinated kernels. In pre-germinated kernels starch and protein have been hydrolysed, this process requires the interaction of water with enzymes, starch or proteins (see Chapter 2, page 13).

Two other peaks were observed in the loading line plot of PC 5; the 1528 nm peak was related to starch [O-H stretching, first overtone (intramolecular H-bond)], and the 2150 nm peak (2x amide I and amide III) was associated with protein (Osborne *et al.*, 1993). These peaks were positively weighted within PC 5 and can be associated with the positive score values (non pre-germinated) in the score plot. During pre-germination starch and proteins are degraded, thus non pre-germinated grains will contain more intact starch and protein molecules (Fincher & Stone, 1993).

Table 3.2 Moisture content of different barley samples

	Moisture content of subsets (%)	
Cultivar	**0 hrs**	**24 hrs**
Puma	7.65	7.72
Erica	7.11	7.58
SSG564	7.50	7.97

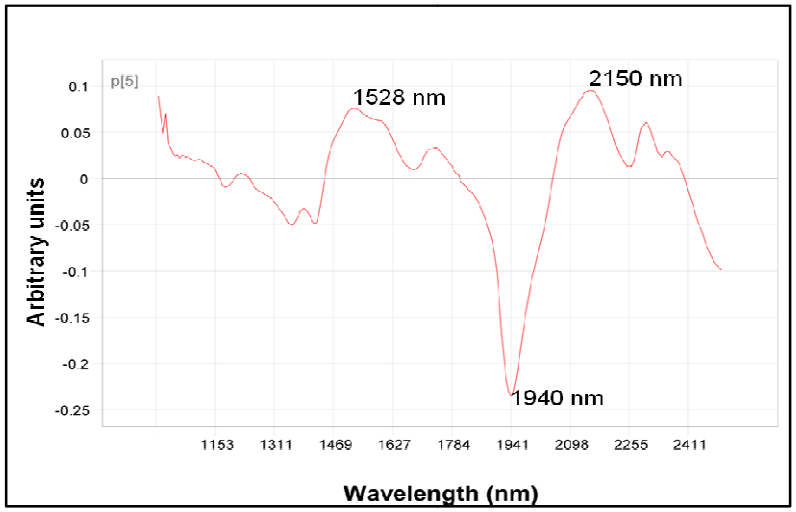


Figure 3.11 PC 5 loading line plot of Puma; indicating starch (1528 nm), water (1940 nm) and protein (2150 nm) absorption bands.

Partial least squares discriminant analysis – 150 µm spatial resolution images

The partial least squares discriminant analysis (PLS-DA) model calculated using the Erica dataset explained the most variation in Y (70% after six components), and was subsequently applied to the two other cultivar datasets. The PLS-DA model calculated using the Puma dataset only explained 68% of the variation in Y, and the SSG564 based model only 65%.

When pre-germinated kernels are predicted as non pre-germinated a false positive will arise that is detrimental for the maltster. Malting requires viable, non pre-germinated barley. Production

losses could occur if pre-germinated barley is predicted as non pre-germinated and used for the malting process. A false negative will arise if non pre-germinated kernels are not predicted as such. If non pre-germinated kernels are predicted as pre-germinated, a false positive will occur that is detrimental to the farmer. Non pre-germinated grain will be down-graded and the farmer will suffer economical losses. False negatives will arise if pre-germinated kernels are not predicted as such. The false positives arising from pre-germinated kernels predicted as non pre-germinated (detrimental for maltster) and the false negatives arising from non pre-germinated kernels not predicted as non pre-germinated will be emphasised in the prediction results.

The number of pixels assigned to each class (pre-germinated and non pre-germinated) of Puma can be seen in **Table 3.3**. The model incorrectly assigned 6701 pixels as non pre-germinated resulting in 34.57% false positives. False positives are considered to be more serious than false negatives (Contal *et al.*, 2002). Of the 19439 pixels assigned (**Table 3.4**) to the non pre-germinated class of SSG564 only 10617 pixels (54.62%) were correctly predicted. Therefore 8822 pixels (45.38% false negatives) were either predicted as pre-germinated or not classified. Thus, kernels that are non pre-germinated will be predicted as pre-germinated (or not classified) and will lead to an over estimation in the amount of pre-germinated grains in a sample. Such an error will result in the down-grading of viable barley.

The occurrence of false positives (Puma) and negatives (SSG564) could be due to the fact that the tetrazolium test was not linked to each kernel in the image. Correct identification and class assignment is needed to create a PLS-DA model that can discriminate better between pre-germinated and non pre-germinated kernels.

The prediction images (**Figs. 3.12a** and **3.13a**) illustrate the localisation of the predicted classes. Model performance can be judged by comparing the prediction and classification images (**Figs. 3.12b** and **3.13b**) to one another. If the prediction image and classification image are similar, the model performed well.

Table 3.3 PLS-DA prediction results indicating the classes, number of pixels assigned and predicted to each class for Puma

Classes	**Assigned number of pixels**	**Predicted number of pixels (% predicted)**
Non pre-germinated	19385	26086 (134.57)
Pre-germinated	23268	13589 (58.40)
Not classified	0	2978
Total	42653	42653

Table 3.4 PLS-DA prediction results indicating the classes, number of pixels assigned and predicted to each class for SSG564

Classes	Assigned number pixels	Predicted number pixels (% predicted)
Non pre-germinated	19439	10617 (54.62)
Pre-germinated	18570	25268 (136.07)
Not classified	0	2124
Total	38009	38009

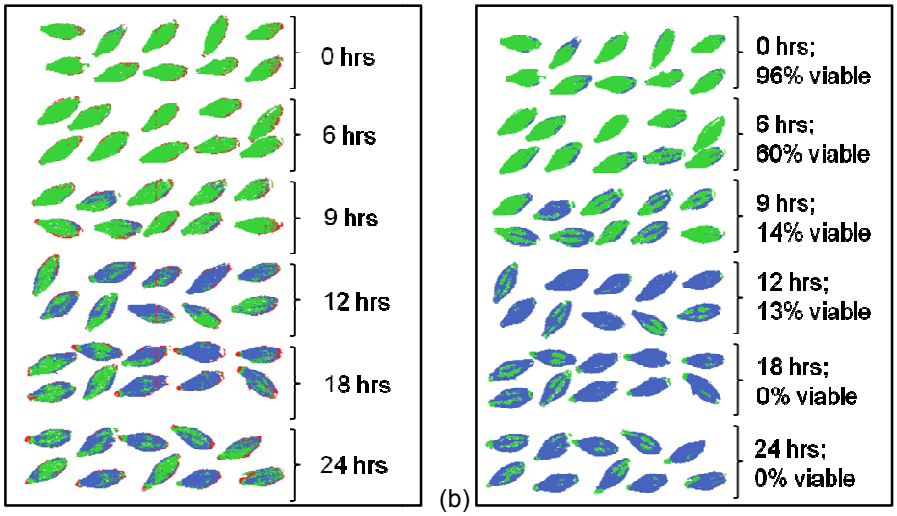


Figure 3.12 (a) Puma PLS-DA prediction image; and (b) classification image with assigned classes (blue = predicted/ assigned pre-germinated, green = predicted/ assigned non pre-germinated and red = not classified).

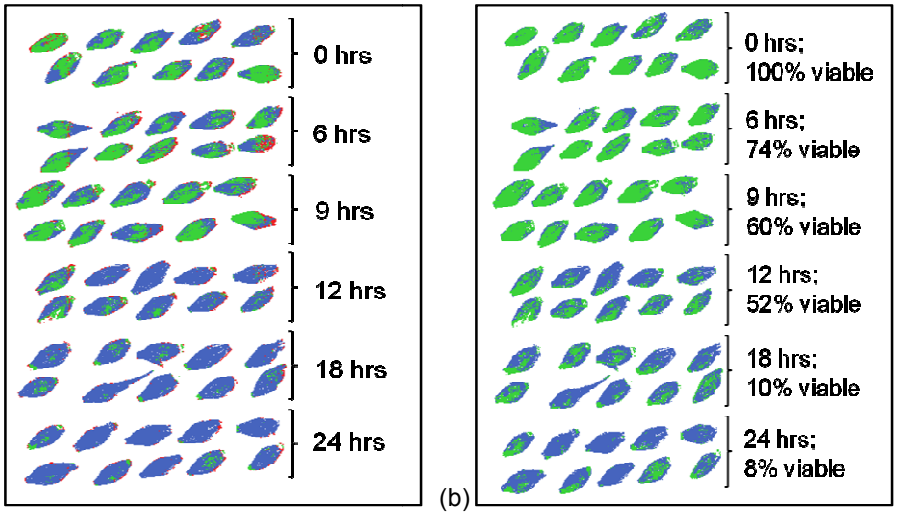


Figure 3.13 (a) SSG564 PLS-DA prediction image; and (b) classification image with assigned classes (blue = predicted/ assigned pre-germinated, green = predicted/ assigned non pre-germinated and red = not classified).

30 µm spatial resolution hyperspectral images

Analysis of cleaned images

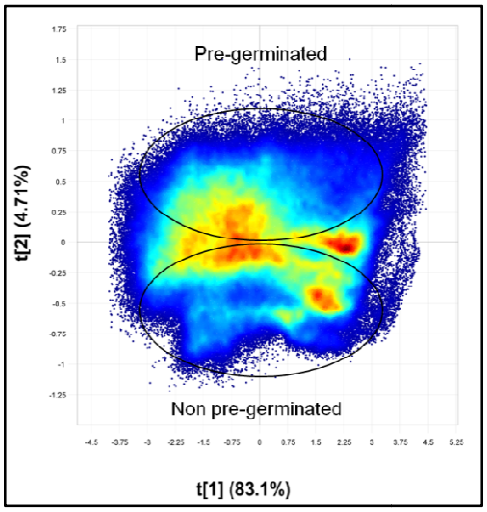
Six columns containing each subset were imaged for all three cultivars. Only one representative column from each cultivar will be discussed to minimise repetition and will be called Clipper-1, Puma-1 and SSG564-1. Two other randomly picked columns will be used to test PLS-DA models, and these columns will be called Clipper-2, Clipper-3, Puma-2, Puma-3, SSG564-2 and SSG564-3.

Clipper-1 – 30 µm spatial resolution images

The Clipper-1 PCA score plot (**Fig. 3.14a**) of PC 1 (SS 83.1%) vs. PC 2 (SS 4.71%) revealed several clusters. The clusters in the direction of PC 1 were not associated with differences between pre-germinated and non pre-germinated kernels, thus these clusters are not discussed. The clusters in the direction of PC 2 accounted for the difference between pre-germinated (positive score values) and non pre-germinated (negative score values) kernels. This was identified by interactively exploring different combinations of score plots and images.

The Clipper-1 PCA score image (**Fig. 3.14b**) of PC 2 differences between the determined pre-germinated and non pre-germinated kernels. However, kernels from the 3 and 9 hrs subsets were determined as non viable (pre-germinated), but was similar in chemical composition to the non pre-germinated kernels in the score image. The endosperm of these kernels may have undergone incomplete degradation by hydrolytic enzymes. The NIR hyperspectral imaging method is perhaps detecting degree of endosperm hydrolysis rather than kernel viability. The tetrazolium test was used to test viability; this test is based on the dehydrogenase enzymes of living cells reducing tetrazolium to formazan, a red compound. A non viable embryo does not have dehydrogenase enzymes and will thus not induce a colour change. A kernel may contain no dehydrogenase enzymes but have only been partially hydrolysed leading to a non viable result by the tetrazolium test and a viable result by the NIR hyperspectral imaging method.

(a)



(b)

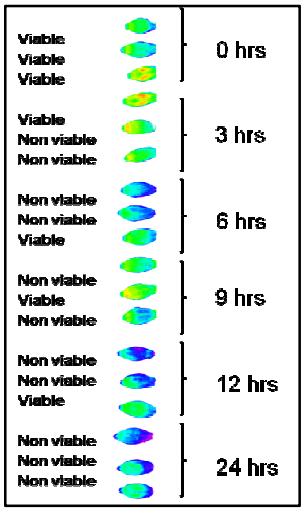


Figure 3.14 (a) Clipper-1 PC 1 vs. PC 2 PCA score plot demonstrating clusters. The clusters in the direction of PC 2 indicate differences between pre-germinated and non pre-germinated kernels. (b) PCA score image of PC 2 showing differences (chemical) between pre-germinated (non viable) and non pre-germinated (viable) kernels.

Two classification images and plots were formed for the Clipper-1 dataset. The first classification plot (**Fig. 3.15a**) and image (**Fig. 3.15b**) were created by selecting the kernels in the score image belonging to either the pre-germinated or non pre-germinated class, as identified by the tetrazolium test. **Fig. 3.15a** shows the location of pixels associated with the pre-germinated (green) or non pre-germinated (blue) kernels in the classification image. The pixels from the non pre-germinated kernels are mostly on the negative side of PC 2 and the pixels from the pre-germinated kernels are mostly on the positive side of PC 2. The second classification plot (**Fig. 3.16a**) and classification image (**Fig. 3.16b**) were created by selecting the pixels in the score plot on the negative side (non pre-germinated) and positive side (pre-germinated) of the PC and projecting it onto the score image (**Fig. 3.16b**).

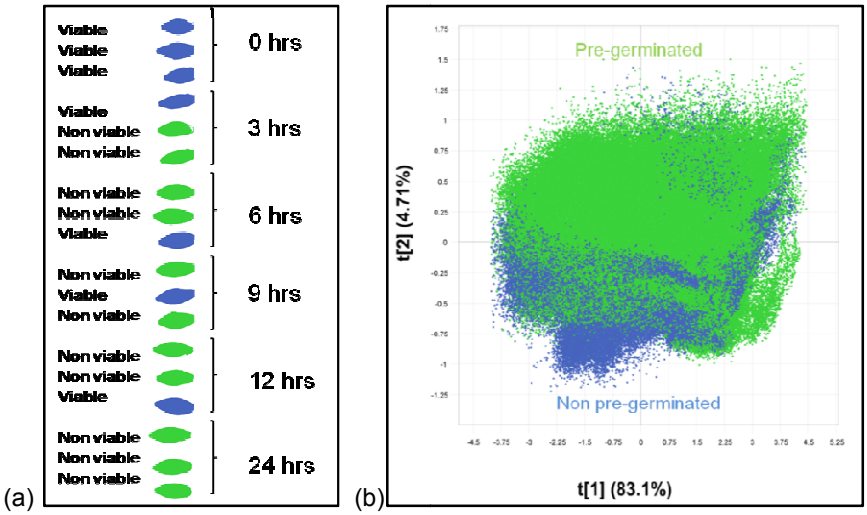


Figure 3.15 (a) Clipper-1 classification image showing the assigned pre-germinated (non viable) and non pre-germinated (viable) kernels with selection based on tetrazolium test results; and (b) classification plot with classes selected in the score image and projected onto the score plot (pre-germinated = green and non pre-germinated = blue).

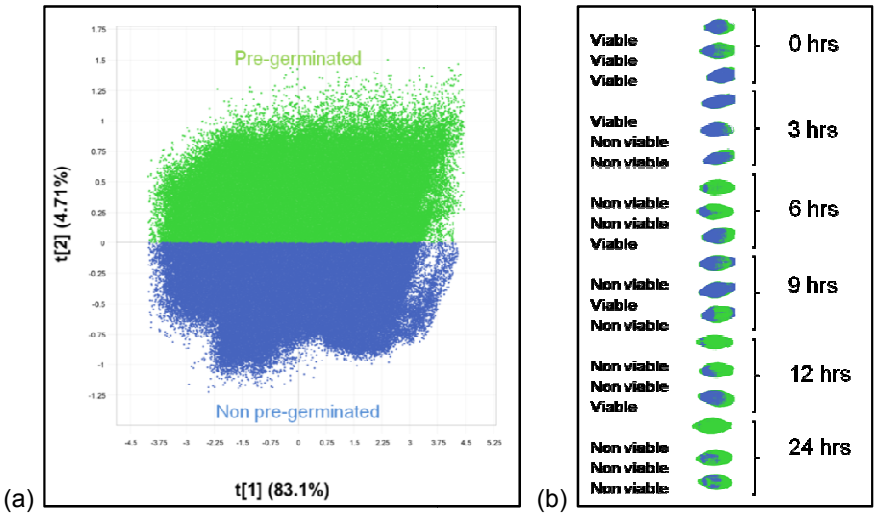


Figure 3.16 (a) Clipper-1 classification plot of PC 1 vs. 2 showing the assigned pre-germinated and non pre-germinated classes; and (b) classification image of score classes projected onto score image (pre-germinated = green and non pre-germinated = blue).

Puma-1 – 30 µm spatial resolution images

The PCA score plot (**Fig. 3.17a**) of PC 1 (SS 83.50%) vs. PC 5 (SS 0.748%) revealed two clusters in the direction of PC 1. These clusters were not investigated further, as they did not correspond to differences between pre-germinated and non pre-germinated kernels. Although no clusters were seen in the direction of PC 5 in the score plot, the pixels on the positive side of PC 5 were associated with the pre-germinated (non viable) kernels and the pixels on the negative side of PC were mostly associated with the non pre-germinated (viable) kernels. The PCA score image (**Fig. 3.17b**) of PC 5 showed differences (chemical differences) between pre-germinated and non pre-

germinated kernels. The kernels from the 18 hrs subset were all determined to be pre-germinated, however, two of these kernels had a similar chemical composition to the non pre-germinated kernels; this was possibly a result of incomplete hydrolysis of the endosperm.

Two sets of classification plots (**Figs. 3.18a** and **3.19a**) and images (**Fig. 3.18b** and **3.19b**) were formed in the manner described for Clipper-1. **Fig. 3.18a** shows that the pixels associated with the pre-germinated kernels in the score image are located on the positive side of PC 5 and the pixels associated with the non pre-germinated kernels in the score image are located on the negative side of PC 5. **Fig 3.19b** reveals the location of the pixels on the negative (blue) and positive (green) side of PC 5. The pixels from the negative side of the PC are mostly associated with the non pre-germinated kernels and the pixels from the positive side of PC 5 are associated with the pre-germinated kernels.

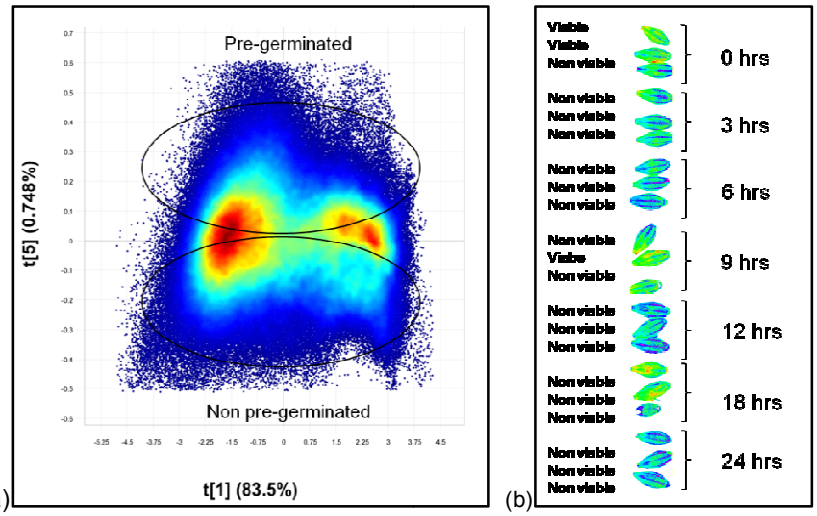


Figure 3.17 (a) Puma-1 PCA score plot (PC 1 vs. PC 5) showing score values related to pre-germinated and non pre-germinated kernels; and (b) PCA score image of PC 5 showing differences between pre-germinated (non viable) and non pre-germinated (viable) kernels.

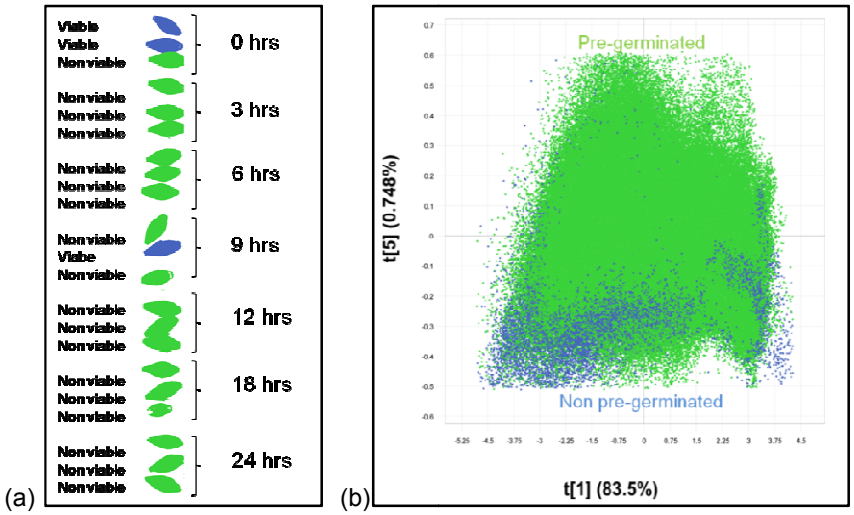


Figure 3.18 (a) Puma-1 classification image showing the assigned pre-germinated and non pre-germinated kernels with selection based on tetrazolium test results; and (b) classification plot with classes selected in the score image and projected onto the score plot (pre-germinated = green and non pre-germinated = blue).

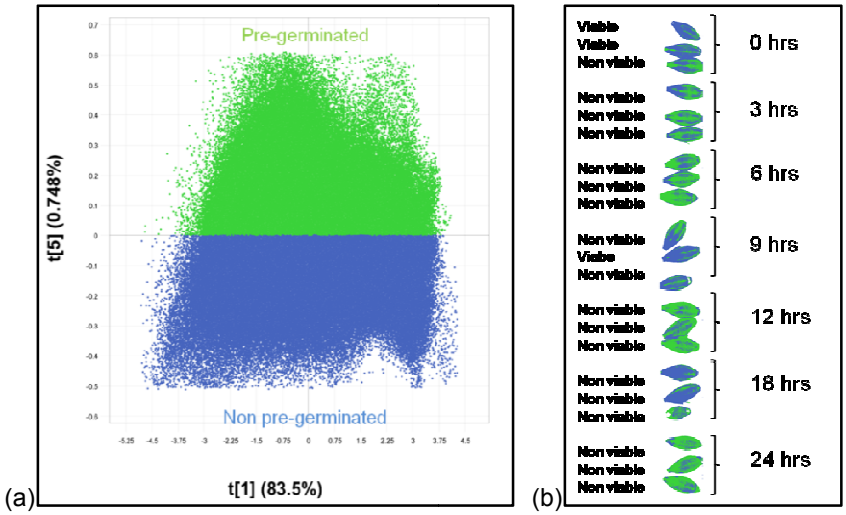


Figure 3.19 (a) Puma-1 classification plot of PC 1 vs. 5 showing the assigned pre-germinated and non pre-germinated classes; and (b) classification image of score classes projected onto score image (pre-germinated = green and non pre-germinated = blue).

SSG564-1 – 30 µm spatial resolution images

The PCA score plot (**Fig. 3.20a**) of PC 1 (SS 86.6%) vs. PC 4 (SS 1.45%) revealed clusters; the clusters in the direction of PC 4 were identified as non pre-germinated and pre-germinated. The PCA score image (**Fig. 3.20b**) of PC 4 revealed differences between the determined pre-germinated and non pre-germinated kernels. There were kernels from the 0, 6 and 9 hrs subsets that were similar to the non pre-germinated kernels, but these kernels were indicated as pre-germinated by the tetrazolium test. This might be due to incomplete endosperm hydrolysis. Two sets of classification images (**Fig. 3.21a** & **3.22a**) and plots (**Fig. 3.21b** & **3.22b**) show the pre-germinated and non pre-germinated kernels. **Fig. 3.21a** shows the pixels associated with the non

pre-germinated kernels on the positive side of PC 4 and the pixels associated with the pre-germinated kernels on the negative side of PC 4. **Fig. 3.22b** shows the pixels on the positive side of PC 4 are mostly associated with the non pre-germinated kernels and the pixels on the negative side of PC 4 are mostly associated with the pre-germinated kernels.

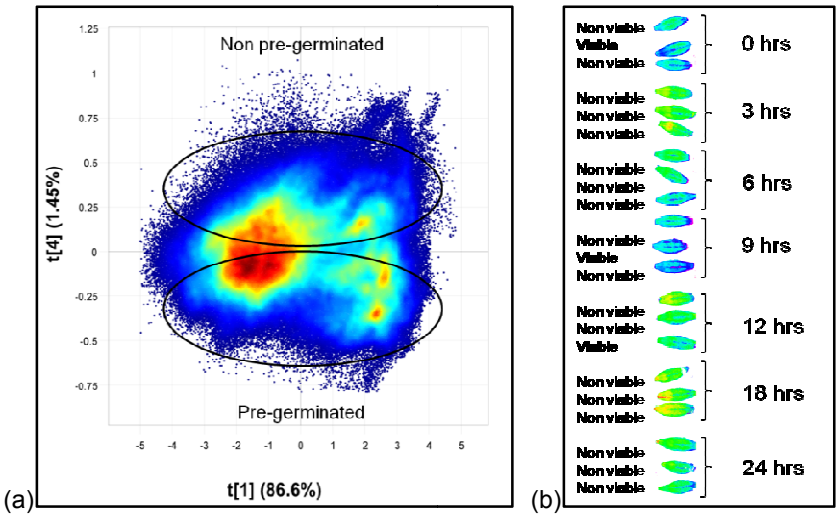


Figure 3.20 (a) SSG564-1 PCA score plot (PC 1 vs. PC 4) showing clusters (pre-germinated and non pre-germinated); and (b) PCA score image of PC 4 showing the pre-germinated (non viable) and non pre-germinated (viable) kernels.

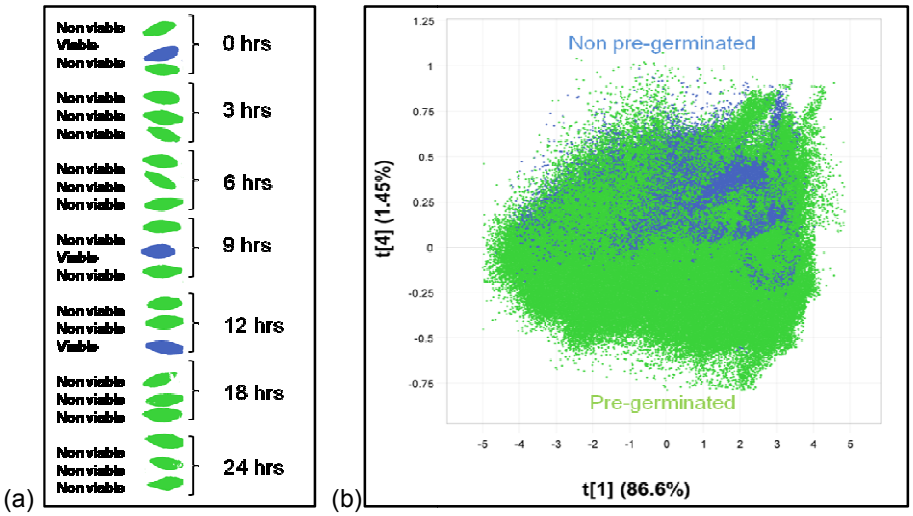


Figure 3.21 (a) SSG564-1 classification image showing the assigned pre-germinated (non viable) and non pre-germinated (viable) kernels with selection based on tetrazolium test results; and (b) classification plot with classes selected in the score image and projected onto the score plot (pre-germinated = green and non pre-germinated = blue).

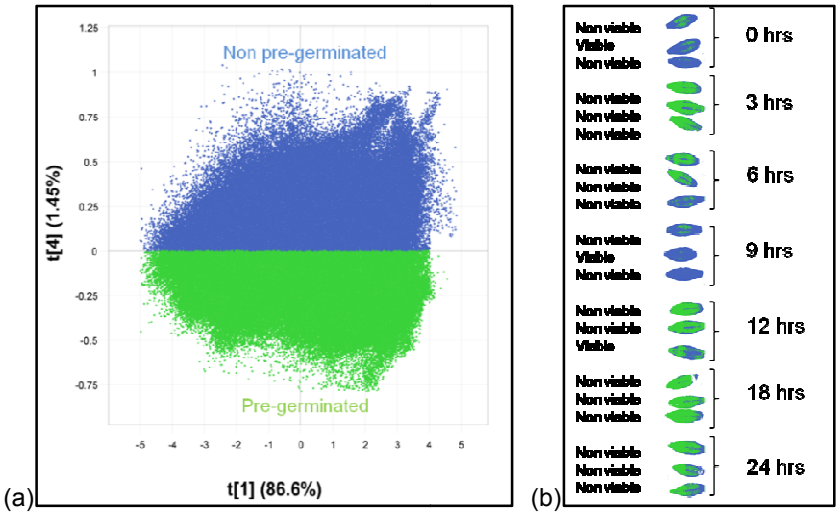


Figure 3.22 (a) SSG564-1 classification plot of PC 1 vs. 4 showing the assigned pre-germinated and non pre-germinated classes; and (b) classification image of score classes projected onto score image (pre-germinate = green and non pre-germinated = blue).

Loading line plots - 30 µm spatial resolution images

The loading line plots of Clipper, Puma and SSG564 differed slightly; however, the main compounds contributing to spectral variation were starch, water and protein.

The Clipper PC 2 loading line plot had several peaks (**Fig. 3.23**) attributable to starch: 1395 nm (2x C-H stretching and C-H deformation), 1695 nm (C-H stretching first overtone), 1725 nm (C-H stretching first overtone), a peak associated with water, 1940 nm (O-H stretching and O-H deformation), and with protein, 2294 nm (N-H stretching and C=O stretching) (Osborne *et al.*, 1993). The water peak is negatively weighted and is associated with the non pre-germinated scores in the score plot. The moisture content results (**Table 3.5**) indicated that the differences between kernels were not due to moisture content differences. The protein and starch peaks are negatively weighted and can be associated with the pre-germinated scores in the score plot. This loading line plot is different to the loading line plot of the 150 µm spatial resolution images, where the water peak was associated with the pre-germinated kernels and the protein and starch peak was associated with the non pre-germinated kernels. It is not clear why this happened, but a possible reason could be that the samples in this sample set were not pre-germinated as profoundly as the kernels from the 150 µm spatial resolution images.

The peaks observed in the Puma PC 5 loading line plot (**Fig. 3.24**) were 1440 nm (starch - O-H stretching first overtone or 2 x C-H stretching and C-H deformation) and 1940 nm (water - O-H stretching and O-H deformation), 2100 nm (starch - 2 x O-H deformation and 2 x C-O stretching), 2242 nm (protein - N-H stretching and NH_3^+ deformation) and 2294 nm (protein - N-H stretching and C=O stretching) (Osborne *et al.*, 1993). The water peak is positively weighted and can be associated with the pre-germinated kernels. The protein peaks are also positively weighted and can be associated with the pre-germinated kernels; this is also different than the 150 µm spatial

resolution images. The starch peaks are negatively weighted and may be associated with the non pre-germinated kernels (intact starch molecules).

SSG564 PC 4 loading line plot had peaks (**Fig. 3.25**) at 1940 nm (O-H stretching and O-H deformation) related to water and 2050 nm (N-H symmetric stretching and amide II or N-H asymmetric stretching and amide III) related to protein (Osborne *et al.*, 1993). The water peak could be associated with the non pre-germinated kernels and the protein peak with the pre-germinated kernels. This is not the same as the 150 µm spatial resolution images.

Smail *et al*. (2006) found that the intensity of the 1680 nm (associated with embryo development) peak allowed pre-germinated and non pre-germinated kernels to be differentiated. Singh *et al*. (2009) used the wavelengths 1101.7 and 1305.1 nm (related to starch) to discriminate between pre-geminated and non pre-germinated wheat kernels. In this study (150 and 30 µm spatial resolution images) the water peak at 1940 nm dominated pre-germinated, non pre-germinated discrimination. None of the researchers found the 1940 nm peak, because neither of the studies included a wavelength range higher than 1700 nm.

Table 3.5 Moisture analysis results for each subset of the three cultivars (Clipper, Puma and SSG564)

Cultivars	Subsets (time exposure to distilled water in hrs)	Moisture content (%)
Clipper	0	8.69
	3	8.51
	6	8.85
	9	8.83
	12	8.70
	24	8.77
Puma	0	8.17
	3	8.12
	6	8.22
	9	8.18
	12	8.15
	18	8.18
	24	8.11
SSG564	0	8.28
	3	8.30
	6	8.17
	9	8.33
	12	8.14
	18	8.29
	24	8.19

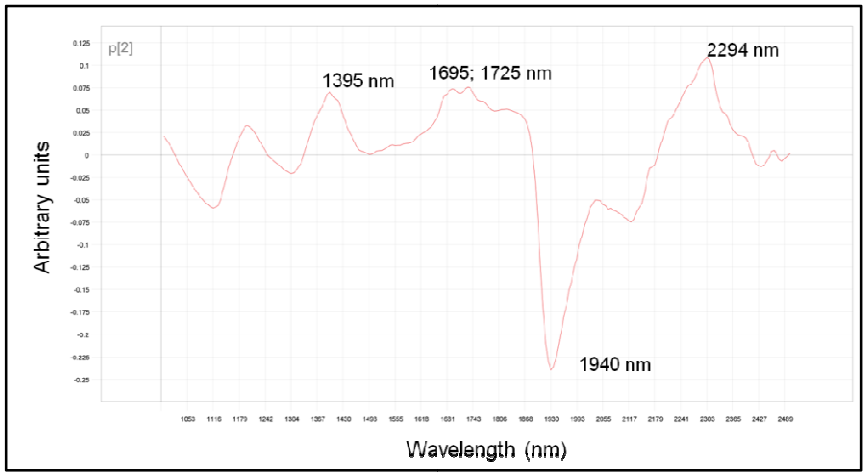


Figure 3.23 Loading line plot of Clipper PC 2, illustrating the compounds responsible for variation: starch (1395 nm, 1695 nm and 1725 nm), water (1940 nm) and protein (2294 nm).

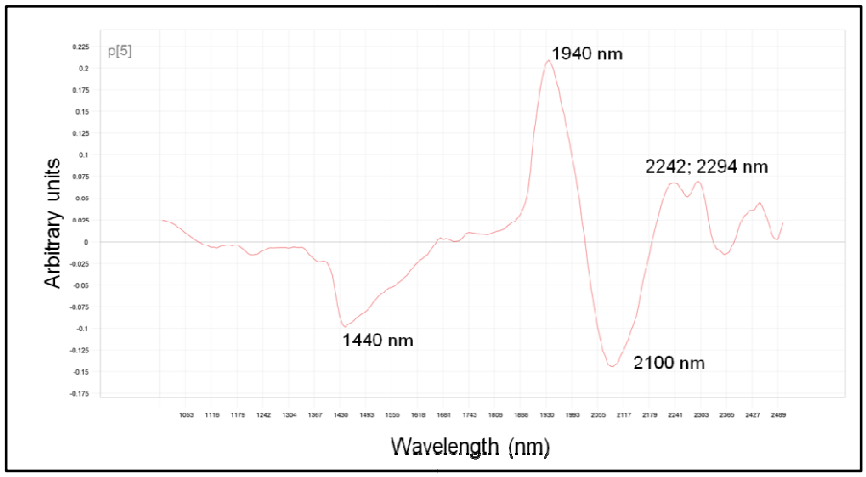


Figure 3.24 Loading line plot of Puma PC 5, showing the compounds contributing to variation; starch (1440 nm and 2100 nm), protein (2242 nm and 2294 nm) and water (1940 nm).

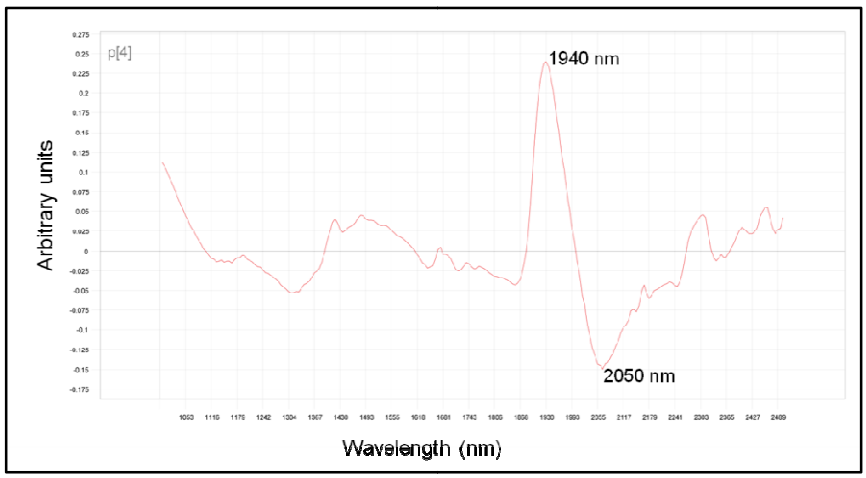


Figure 3.25 Loading line plot of SSG564 PC 4, showing peaks for water (1940 nm) and protein (2050 nm).

Partial least squares discriminant analysis – 30 µm spatial resolution images

Clipper-2 and Clipper-3

Two PLS-DA models were created for Clipper-1 and both of these models were employed to predict two other images of Clipper (2 and 3). Model one (M1) was created by classifying the pre-germinated and non pre-germinated kernels in the score image. Model two (M2) was created by classifying the pre-germinated and non pre-germinated scores in the score plot. The explained variation in Y was 35.22% and 66.20% after six components for M1 and M2 respectively.

The prediction data for Clipper-2 (**Table 3.6**) indicated that the models did not perform well. M1 which explained less variation in Y than M2, performed better; with only 4.39% false positives and a 91.22% correct prediction of the pre-germinated class. **Fig. 3.26** shows the prediction images and the classification images containing the predicted and assigned pre-germinated and non pre-germinated pixels of Clipper-2. The differences between the images of M1 reveal the model did not perform well.

The prediction results of Clipper-3 indicated that the models did not perform well either (**Table 3.6**). M1 predicted 53.46% false negatives and M2 predicted 32.3% false negatives. **Fig. 3.27** contains the prediction images and classification images of Clipper-3. Even though M2 did not perform well according to the prediction results, the images (**Fig. 3.27**) indicate that the model performed better than M1.

Better classification results were obtained in the study of Singh *et al*. (2009). Both the linear discriminant classifier and quadratic discriminant classifier calculated correctly (100%) classified non pre-germinated and pre-germinated kernels. A Mahalanobis discriminant classifier correctly identified 100% and 98.3% of pre-germinated and non pre-germinated kernels, respectively. These classification results were expected as the imaged kernels were either non pre-germinated (no treatment with water) or pre-germinated (treated with water for 48 hrs and visible sprouting recorded). The extreme nature of the two classes facilitated the discriminating abilities of these models. In this study kernels were pre-germinated for several shorter periods of time (3 – 24 hrs). Varying degrees of endosperm hydrolysis were evident in the images and may have complicated the models, leading to the models' insufficient prediction. Models developed from only pre-germinated or non pre-germinated kernels may exhibit improved prediction results.

The amount of kernels used for the development of the models may have influenced the prediction results as well. Although a huge amount of pixels are utilised in the development of the model, the reference method is done on single kernel basis. Therefore, models developed from more kernels might result in better discrimination between the classes.

Table 3.6 PLS-DA prediction results for Clipper-2 and Clipper-3 of both models M1 and M2

		Assigned number of pixels	**Predicted number of pixels (% pixels)**
Clipper-2	**M1**		
	Non pre-germinated	66125	69026 (104.39)
	Pre-germinated	319731	291653 (91.22)
	Not classified		25177
	Total	385856	385856
	M2		
	Non pre-germinated	192214	126842 (65.99)
	Pre-germinated	193642	233885 (120.78)
	Not classified		25129
	Total	385856	385856
Clipper-3	**M1**		
	Non pre-germinated	190492	88658 (46.54)
	Pre-germinated	218172	294375 (134.93)
	Not classified		25631
	Total	408664	408664
	M2		
	Non pre-germinated	207032	140163 (67.70)
	Pre-germinated	201632	239165 (118.61)
	Not classified		29336
	Total	408664	408664

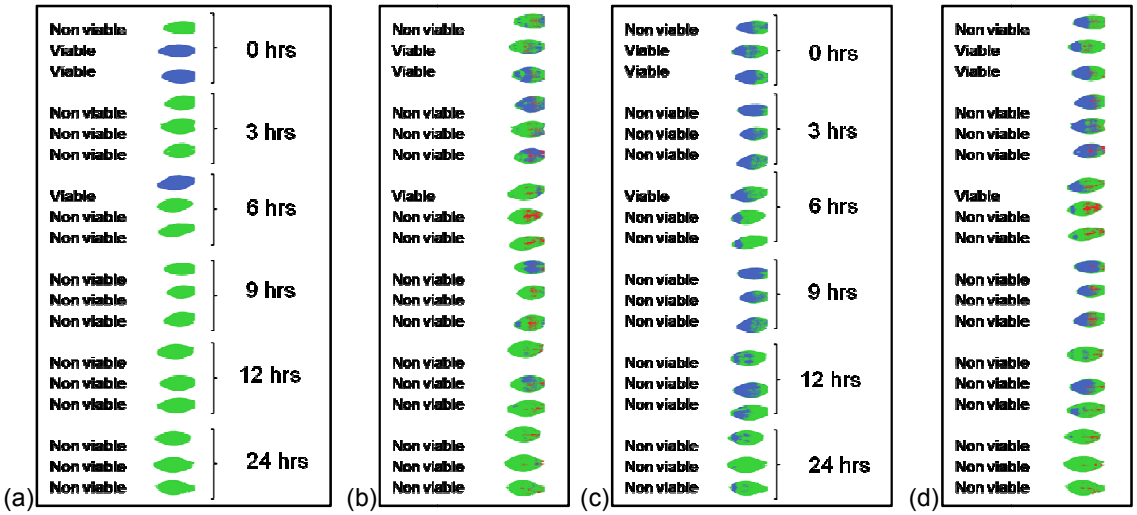


Figure 3.26 (a) Clipper-2 classification image with class assignment based on score image; and (b) prediction image of Clipper-2 with M1 (model developed according to the score image from Clipper-1). (c) Clipper-2 classification image with class assignment based on the score plot; and (d) prediction image of Clipper-2 with M2 (model developed according to the score plot from Clipper-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

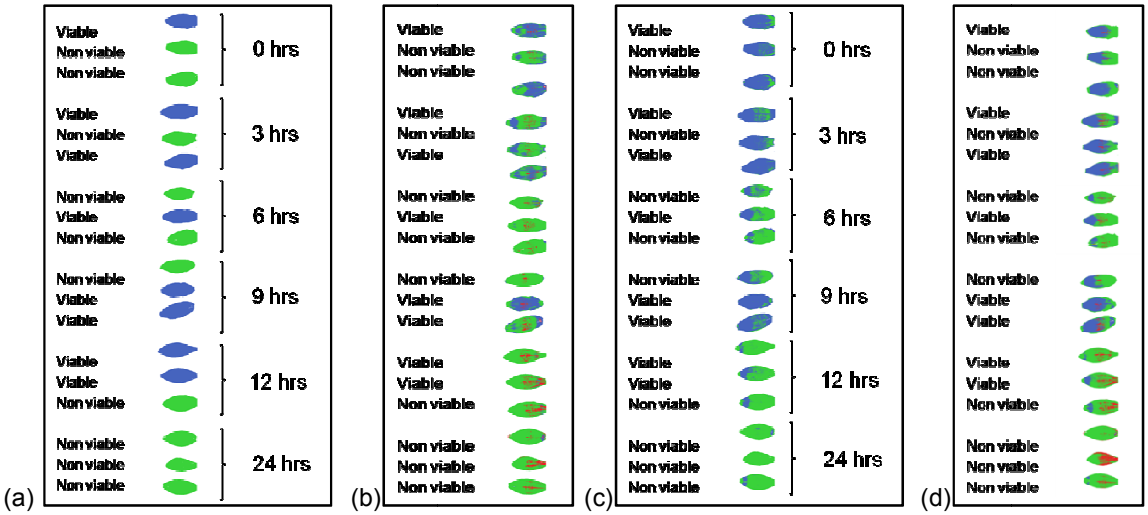


Figure 3.27 (a) Clipper-3 classification image with class assignment based on score image; and (b) prediction image of Clipper-3 with M1 (model developed according to the score image from Clipper-1). (c) Clipper-3 classification image with class assignment based on the score plot; and (d) prediction image of Clipper-3 with M2 (model developed according to the score plot from Clipper-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

Puma-2 and Puma-3

Of the two models created for Puma-1, M2 had the highest variation in Y after six components with 61.31%; M1 explained 18.16% of the variation in Y after six components. M1 predicted (**Table 3.7**) false negatives for both Puma-2 (95.47%) and Puma-3 (38.54%). M2 predicted false positives for Puma-2 (27.28%) and Puma-3 (41.69%). False positives are said to be the worse of the two

errors; however 95.47% and 38.54% false negatives are not acceptable either. Neither of the models can be used for independent classification of pre-germinated or non pre-germinated kernels. The ineffectiveness of the predictions is visualised in the **Figs. 3.28** and **3.29**.

Table 3.7 Results of PLS-DA prediction for Puma-2 and Puma-3 of both models M1 and M2

		Assigned number of pixels	**Predicted number of pixels (% pixels)**
Puma-2	**M1**		
	Non pre-germinated	133681	6052 (4.53)
	Pre-germinated	348774	458042 (131.33)
	Not classified		18361
	Total	482455	482455
	M2		
	Non pre-germinated	251400	319988 (127.28)
	Pre-germinated	231055	134994 (58.42)
	Not classified		27473
	Total	482455	482455
Puma-3	**M1**		
	Non pre-germinated	90911	55875 (61.46)
	Pre-germinated	412894	425259 (102.99)
	Not classified		22671
	Total	503805	503805
	M2		
	Non pre-germinated	243811	345460 (141.69)
	Pre-germinated	259994	77638 (29.86)
	Not classified		80707
	Total	503805	503805

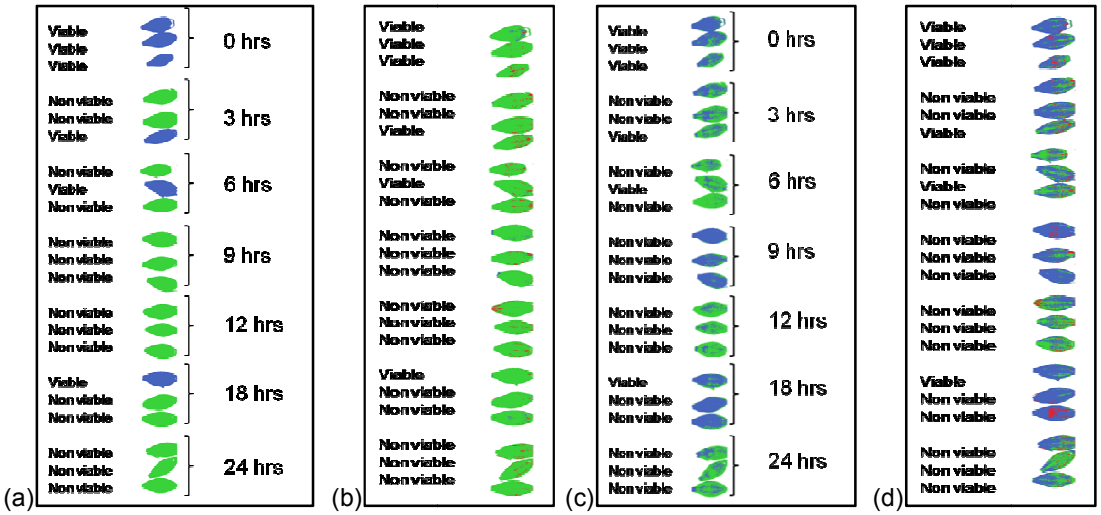


Figure 3.28 (a) Puma-2 classification image with class assignments based on score image; and (b) prediction image of Puma-2 with M1 (model developed from the score image of Puma-1). (c) Puma-2 classification image with class assignment based on the score plot; and (d) prediction image of Puma-2 with M2 (model developed from the score plot of Puma-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

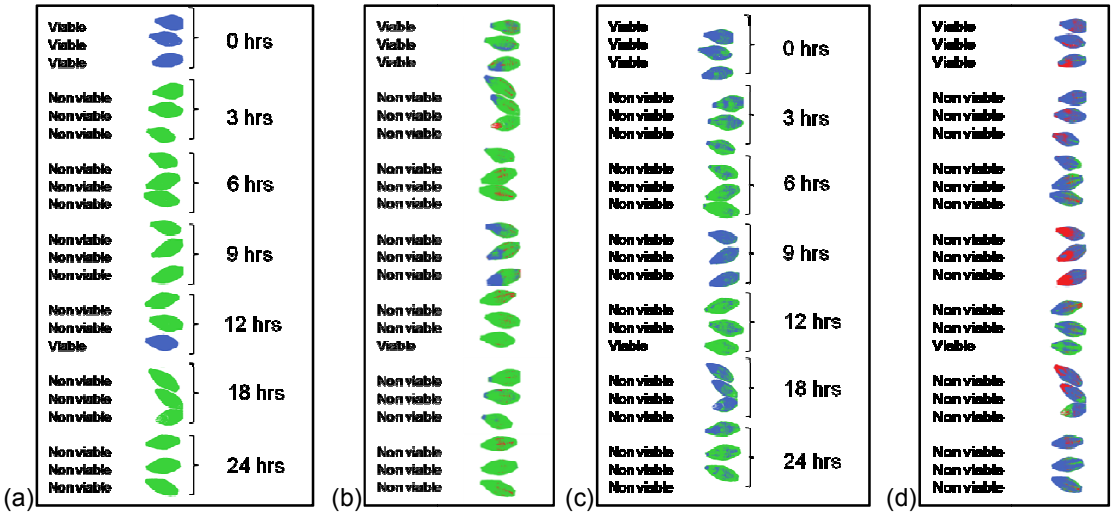


Figure 3.29 (a) Puma-3 classification image with class assignments based on score image; and (b) prediction image of Puma-3 with M1 (model developed from the score image of Puma-1). (c) Puma-3 classification image with class assignment based on the score plot; and (d) prediction image of Puma-3 with M2 (model developed from the score plot of Puma-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

SSG564-2 and SSG564-3

The explained variation in Y for M1 of SSG564-1 was 23.59%, and M2 was 66.07%, after six components. The prediction results revealed that the PLS-DA models did not perform well (**Table 3.8**). M1 predicted only 51.64% of the non pre-germinated pixels correctly for SSG564-2. The same model only predicted 25.12% non pre-germinated pixels correctly for SSG564-3. M2

predicted 35.20% (false positives) more pixels as non pre-germinated than were assigned for SSG564-2 and it predicted 21.67% (false positives) more pixels as non pre-germinated than were assigned for SSG564-3. M2 performed worse than M1, but M1 did not perform well enough to discriminate between pre-germinated and non pre-germinated kernels. The poor prediction results can be visualised in **Figs**. **3.30** and **3.31**, where the prediction images are compared with the classification images.

Table 3.8 PLS-DA prediction results for SSG564-2 and SSG564-3 of both models M1 and M2

		Assigned number of pixels	**Predicted number of pixels (% pixels)**
SSG564-2	**M1**		
	Non pre-germinated	47917	24742 (51.64)
	Pre-germinated	429501	428997 (99.88)
	Not classified		23679
	Total	477418	477418
	M2		
	Non pre-germinated	218203	295019 (135.20)
	Pre-germinated	259215	147863 (57.04)
	Not classified		34536
	Total	477418	477418
SSG564-3	**M1**		
	Non pre-germinated	103393	25973 (25.12)
	Pre-germinated	395463	456388 (115.41)
	Not classified		16495
	Total	498856	498856
	M2		
	Non pre-germinated	240345	291218 (121.67)
	Pre-germinated	258511	181567 (70.24)
	Not classified		26071
	Total	498856	498856

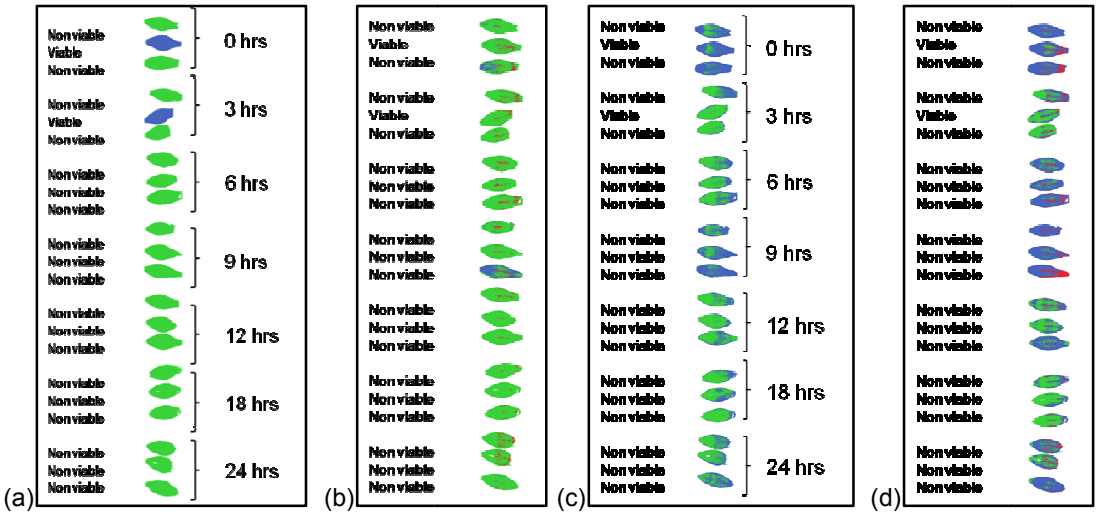


Figure 3.30 (a) SSG564-2 classification image with class assignments based on score image; and (b) prediction image of SSG564-2 with M1 (model developed from the score image of SSG564-1). (c) SSG564-2 classification image with class assignment based on the score plot; and (d) prediction image of SSG564-2 with M2 (model developed from the score plot of SSG564-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

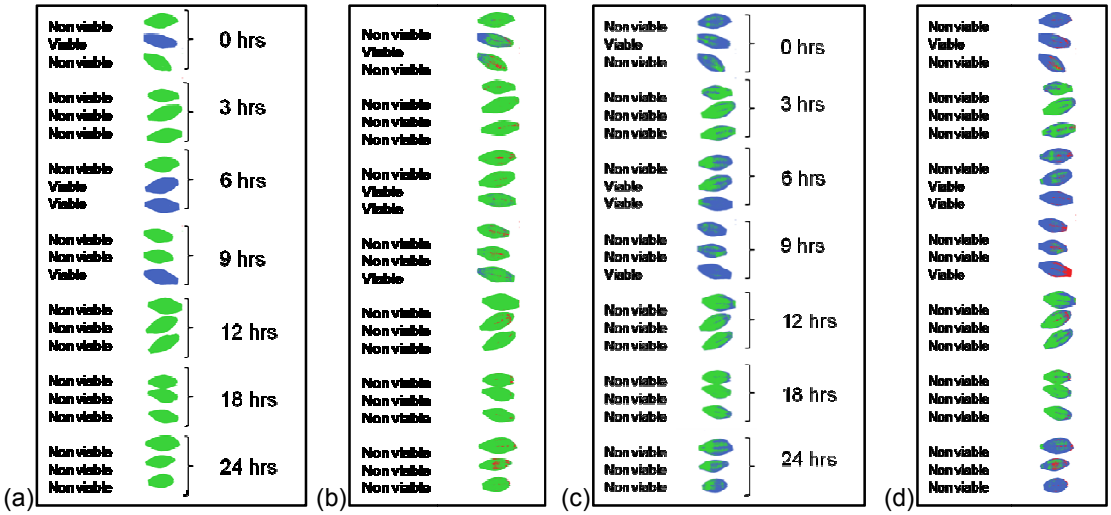


Figure 3.31 (a) SSG564-3 classification image with class assignments based on score image; and (b) prediction image of SSG564-3 with M1 (model developed from the score image of SSG564-1). (c) SSG564-3 classification image with class assignment based on the score plot; and (d) prediction image of SSG564-3 with M2 (model developed from the score plot of SSG564-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

Conclusion

Score images revealed differences between pre-germinated and non pre-germinated subsets of the 150 µm spatial resolution images. These differences, together with the tetrazolium test determined viability, lead to the classification of pre-germinated and non pre-germinated kernels

using NIR hyperspectral imaging. The analysis of the 30 µm spatial resolution images revealed more complex observations; some kernels were identified as pre-germinated by the tetrazolium test, but these kernels were similar in chemical composition to non pre-germinated kernels of the PCA score images. These observations might be attributed to incomplete endosperm hydrolysis leading to a non viable result by the tetrazolium test and a viable result by NIR hyperspectral imaging. Thus, NIR hyperspectral imaging might be employed as a method to determine endosperm hydrolysis of barley samples.

The loading line plots revealed that water was the primary compound contributing to the variation in images showing the most significant differences between pre-germinated and non pre-germinated kernels. Contradicting results were obtained with the loading line plots of the 30 µm spatial resolution images; in some cases the water peak was associated with the non pre-germinated kernels and vice versa. The kernels from the 30 µm spatial resolution images could be different from the 150 µm spatial resolution images due to the extent to which pre-germination has occurred.

The calculated PLS-DA models were not able to effectively discriminate between pre-germinated and non pre-germinated kernels. Prediction models developed from only the most extreme time point (0 and 24 hrs) may improve the prediction results. PLS-DA models developed from a bigger sample set (more kernels tested with tetrazolium test) may also improve the prediction results. Other classification methods such as support vector machines (SVM), artificial neural networks (ANN) or soft independent modelling of class analogy (SIMCA) could be investigated and their performance compared to that of PLS-DA. This, however, was outside the scope of this research. Although the PLS-DA models did not perform well, NIR hyperspectral imaging could detect a degree of endosperm hydrolysis, in contrast to the tetrazolium test that only determines viability.

References

AACC. (2010). *Approved Methods of the AACC*. Moisture – modified vacuum-oven method. St. Paul, Minnesota: American Association of Cereal Chemists.

Bamforth, C.W. & Barclay, A.H.P. (1993). Malting technology and the uses of malt. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 297-354. St. Paul: American Association of Cereal Chemists, Inc.

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

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

Bueckert, R.A., Lefol, E.B. & Harvey, B.L. (2007). Early detection of non-visible sprouting in barley seed using rapid viscosity analysis. *Canadian Journal of Plant Science*, **87**, 3-12.

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.

Carn, J.D. (1982). Alpha amylase indicates problems with malting of stored barley. *Food Technology in Australia*, **34**, 82-83.

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.

Contal, L., Leon, V. & Downey, G. (2002). Detection and quantification of apple adulteration in strawberry and raspberry purees using visible and near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **10**, 289-299.

Fincher, G.B. & Stone, B.A. (1993). Physiology and biochemistry of germination in barley. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 247-294. St. Paul: American Association of Cereal Chemists, Inc.

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

Geladi, P., Isaksson, H., Lindqvist, L., Wold, S. & Esbensen, K. (1989). Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems*, **5**, 209-220.

Gruwel, M.L.H., Yin, X.S., Edney, M.J., Schroeder, S.W., MacGregor, A.W. & Abrams, S. (2002). Barley viability during storage: use of magnetic resonance as a potential tool to study viability loss. *Journal of Agricultural and Food Chemistry*, **50**, 667-676.

Gualano, N.A. & Benech-Arnold, R.L. (2009). Predicting pre-harvest sprouting susceptibility in barley: looking for 'sensitivity windows' to temperature throughout grain filling in various commercial cultivars. *Field Crops Research*, **114**, 35-44.

Koç, H., Smail, V.W. & Wetzel, D.L. (2008). Reliability of InGaAs focal plane array imaging of wheat germination at early stages. *Journal of Cereal Science*, **48**, 394-400.

Li, C., Ni, P., Francki, M., Hunter, A., Zhang, Y., Schibeci, D., Li, H., Tarr, A., Wang, J., Cakir, M., Yu, J., Bellgard, M., Lance, R. & Appels, R. (2004). Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. *Functional Integrative Genomics*, **4**, 84-93.

Mahesh, S., Manickavasagan, A., Jayas, D.S., Paliwal, J. & White, N.D.G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, **101**, 50-57.

Martinsen, P. & Schaare, P. (1998). Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology*, **14**, 271-281.

Matsuo, R.R., Dexter, J.E. & MacGregor, A.W. (1982). Effect of sprout damage on durum wheat and spaghetti quality. *Cereal Chemistry*, **59**, 468-472.

Nilan, R.A. & Ullrich, S.E. (1993). Barley: taxonomy, origin, distribution, production, genetics and breeding. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 1-29. St. Paul: American Association of Cereal Chemists, Inc.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). In: *Practical NIR spectroscopy with applications in food and beverage analysis* (edited by D. Browning). Pp. 1-35, 49-77. Essex: Longman Scientific & Technical.

Pagano, E.A., Benech-Arnold, R.L., Wawrzkiewicz, M. & Steinbach, H.S. (1997). Alpha-amylase activity in developing sorghum caryopses from sprouting resistant and susceptible varieties. The role of ABA and GAs on its regulation. *Annals of Botany*, **79**, 13-17.

Peirs, A., Scheerlinck, N., De Baerdemaeker, J. & Nicolai, B.M. (2003). Starch index determination of apple fruit by means of a hyperspectral near infrared reflectance imaging system. *Journal of Near Infrared Spectroscopy*, **11**, 379-389.

Pyler, R.E. & Thomas, D.A. (2000). Malted cereals: their production and use. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 685-696. New York: Marcel Dekker, Inc.

Shahin, M.A. & Symons, S.J. (2008). Detection of hard vitreous and starchy kernels in amber durum wheat samples using hyperspectral imaging. *NIR News*, **19**, 16-18.

Singh, C.B., Jayas, D.S., Paliwal, J. & White, N.D.G. (2009). Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry*, **86**, 256-260.

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.

Taghizadeh, M., Gowen, A., Ward, P. & O'Donnell, C.P. (2010). Use of hyperspectral imaging for the evaluation of the shelf-life of fresh white button mushrooms (*Agaricus bisporus*) stored in different packaging films. *Innovative Food Science and Emerging Technologies*, **11**, 423-431.

Williams, P., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta*, **653**, 121-130.

Chapter 4

Detection of wheat pre-germination with near infrared (NIR) hyperspectral imaging

Chapter 4

Detection of wheat pre-germination with near infrared (NIR) hyperspectral imaging

Abstract

The partial least squares discriminant analysis (PLS-DA) results of NIR hyperspectral imaged wheat samples revealed that the PLS-DA model created from 150 µm spatial resolution images was acceptable to discriminate between pre-germinated (93%) and non pre-germinated (81%) wheat kernels. The models created from 30 µm spatial resolution images was not acceptable to discriminate between pre-germinated and non pre-germinated wheat kernels. Pre-germination is the germination (pre- or post-harvest) of wheat to such an extent that the grain is not useful in the baking industry. Wheat cultivars were each divided into subsets of which each subset was exposed to water for a different periods of time. NIR hyperspectral images were acquired with two SisuCHEMA short wave infrared (SWIR) hyperspectral imaging systems. The first system acquired images with a 50 x 100 mm field of view (150 µm spatial resolution) and the second system acquired images with a 10 x 200 mm field of view (30 µm spatial resolution).

Principal component analysis (PCA) was preformed on standard normal variate transformed and mean-centered data. The images were then effectively explored using interactive PCA score images and plots in Evince multivariate image analysis software. A difference between pre-germinated and non pre-germinated kernels was evident. Loading line plots revealed the compounds responsible for the variation were starch, moisture and protein. Proteins and starch are degraded during pre-germination and water molecules are required for hydrolysis.

Introduction

Wheat is mainly used for human food production (Orth & Shellenberger, 1988). Wheat has become a major part of the human diet because of its agronomic flexibility, ease of storage, nutritional value and the ability to produce a variety of exciting, appetizing, and satisfying foods. Among the products made from wheat is bread, noodles, cakes, biscuits, cookies, breakfast cereals, pastries and pasta (Kent & Evers, 1994; Posner, 2000). Wheat is also a popular choice for animal feed, especially in years where harvests are affected by rain (pre-germinated). Pre-germination occurs when mature wheat is subjected to rain or humidity prior to or after harvest (Smail *et al.*, 2006). Wheat is downgraded to feed grade leading to subsequent economic losses (Kruger, 1990; Neethirajan *et al.*, 2007; Singh *et al.*, 2009b). Pre-germination results in the increase of certain germinative enzymes (e.g. alpha-amylase) (Kruger, 1990). Gluten proteins (important in the bread baking process) may also be become damaged (Kruger, 1990).

A small amount of alpha-amylase is necessary to degrade damaged starch and produce fermentable sugars for gas production during by yeast bread production (Kruger, 1994). When an

excessive amount of alpha-amylase is present (as in the case of pre-germinated wheat) bread will have a wet, sticky crumb, which can create problems at the slicing stage of mechanized bread production or bakery (Lorenz & Valvano, 1981; Kruger, 1990). The higher the level of alpha-amylase in wheat intended for spaghetti will result in a softer cooked product (Matsuo *et al.*, 1982). Cakes baked from pre-germinated wheat have a coarse grain, dense and firm texture and have a dip in the centre (Lorenz & Valvano, 1981). Cookies made from pre-germinated wheat are brittle, crumbly and darker than cookies made from sound wheat (Lorenz & Valvano, 1981). Pre-germinated wheat has a greater chance of becoming infected with pathogens and becoming infested with insects as these organisms are caught in the cracks and fissure that develop during pre-germination (Singh *et al.*, 2009b). When infected and infested wheat is mixed with sound wheat, cross-contamination can occur (Singh *et al.*, 2009b).

The detection of pre-germination in wheat has to be simple, rapid, rugged, reliable and inexpensive (Kruger, 1990). Visual assessment of wheat is a simple, rapid and inexpensive method of assessing wheat quality. However, these inspections are based on the knowledge and experience of the grader, and thus are not a very objective method (Neethirajan *et al.*, 2007). Other detection methods used include falling number and viscosity analysis. These two tests are timely and destructive. An alternative detection method has been proposed, i.e. near infrared (NIR) hyperspectral imaging (Smail *et al.*, 2006; Koç *et al.*, 2008; Singh *et al.*, 2009b).

NIR hyperspectral imaging is a technique that is widely used for wheat applications; the detection of insect damaged wheat kernels (Singh *et al.*, 2009a), the determination of wheat kernel hardness and the tracking of conditioning water in single wheat kernels (Du Toit, 2009), the determination of vitreous and starchy kernels (Gorretta *et al.*, 2006; Shahin & Symons, 2008), the differentiation between Canadian wheat classes (Mahesh *et al.*, 2008) and the classification of sound and stained grains (Berman *et al.*, 2007).

The aim of this research was to evaluate NIR hyperspectral imaging for a rapid, non-destructive and objective alternative to the conventional pre-germination detection methods for whole wheat kernels; and to discriminate between pre-germinated and non pre-germinated kernels using partial least squares discriminant analysis.

Materials and methods

Samples and sample preparation

Two sets of cultivar samples were used; the first included Duzi, Kariega and Biedou and was used to collect 150 µm spatial resolution images, the second included Elands and SST 398 and was used to collect 30 µm spatial resolution images. The cultivars for the first cultivar set were kindly provided by the Small Grains Institute, Agricultural Research Council (Stellenbosch, South Africa). Six subsets, comprising 25 kernels each, were randomly selected from each of the 3 respective

cultivars. The sample preparation was performed in the same manner as described in Chapter 3 (page 42).

The cultivars used for the second cultivar set were kindly provided by Sensako (Pty) Ltd. (Bethlehem, South Africa). Each cultivar was divided into seven subsets each containing 100 randomly selected kernels. The subsets were incubated for 6, 9, 12, 18, 24 or 30 hrs at 19°C. A set of control (0 hrs) samples were also included

Near infrared hyperspectral imaging system

The two SisuCHEMA imaging systems used are described in Chapter 3 (page 42).

Image acquisition

Image acquisition was performed in the same manner as described in Chapter 3 (page 43). **Fig. 4.1** shows the graphical representation (150 µm spatial resolution images) and a digital image of the wheat cultivars on silicon carbide sandpaper. **Fig. 4.2** shows the graphical representation and digital photograph of the 30 µm spatial resolution images. The 18 hrs and 24 hrs subsets of SST 398 and the 18 hrs subset of Elands were not included for image acquisition due to mould infection.

Hyperspectral image analysis

Image cleaning and image analysis of cleaned images

The randomly imaged samples were rearranged chronologically in Evince version 2.4.0 (Umbio AB, Umeå, Sweden), as explained in Chapter 3 (page 43). Principal component analysis (PCA) with six components was performed on standard normal variate (SNV) transformed and mean-centered data to clean the images (removal of background, dead pixels and other unwanted pixels). The cleaned images were subjected to further image analysis and principal component (PC) score plots, images and loading line plots were studied.

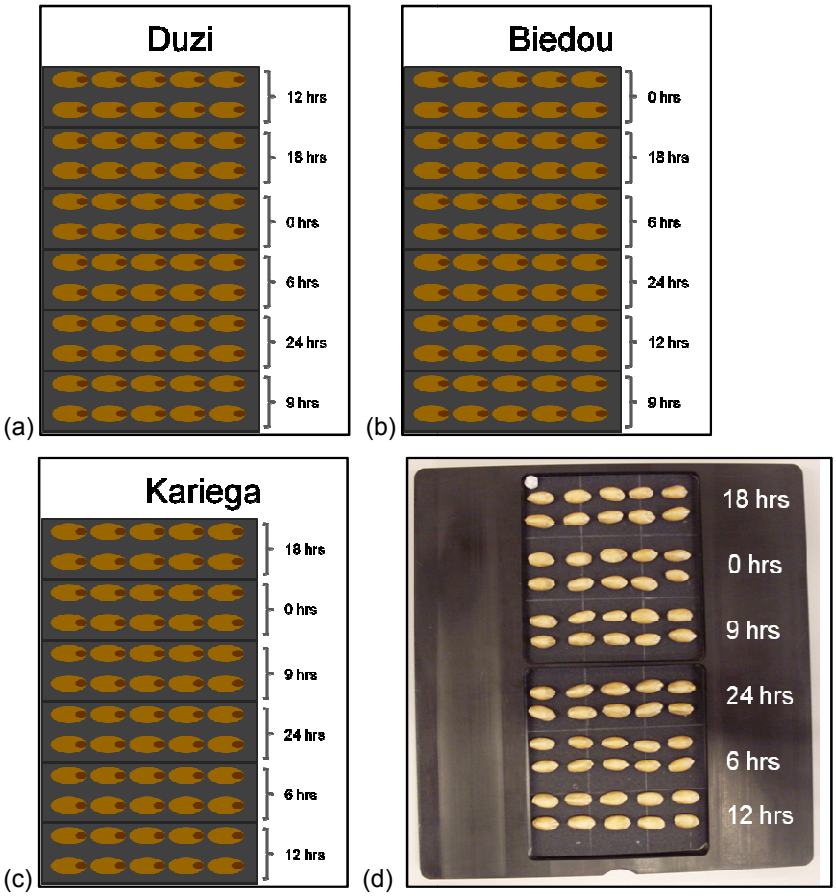


Figure 4.1 (a-c) Graphical representations of the different wheat cultivars positioned randomly according to time on the sample stage of the SisuCHEMA; and (d) digital photograph of Kariega.

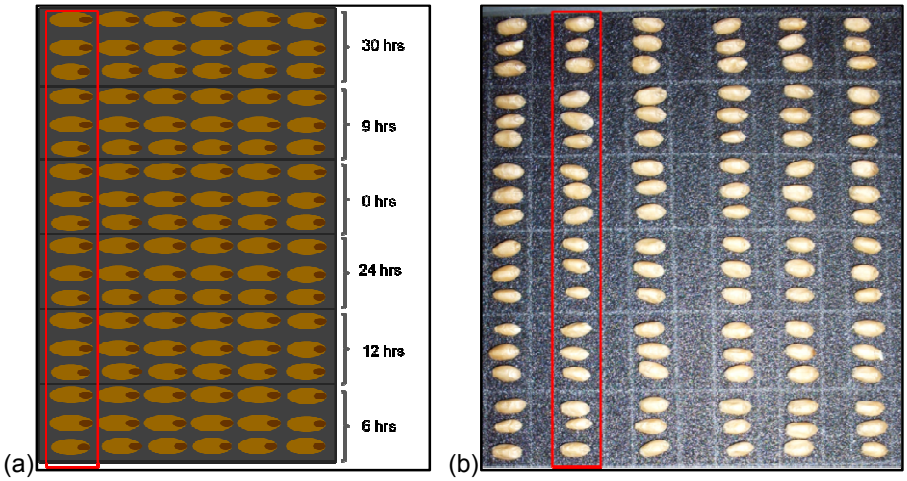


Figure 4.2 (a) Graphical representation of the 30 µm spatial resolution samples, one image is indicated with a red rectangle; and (b) a digital photograph of a representative sample.

Partial least squares discriminant analysis

Partial least squares discriminant analysis (PLS-DA) was performed to determine whether it was possible to discriminate between pre-germinated and non pre-germinated wheat kernels. The PLS-

DA prediction results will be discussed in terms of false positives and false negatives. The PLS-DA models for the 150 and 30 µm spatial resolution images were created in the same manner as described in Chapter 3 (page 45).

Wheat viability

Wheat viability was performed in the same manner as for the barley samples. The wheat kernels were exposed to a 0.5% tetrazolium chloride solution for 1 hr. The kernels were classified according to the specifications in Chapter 3 (page 45).

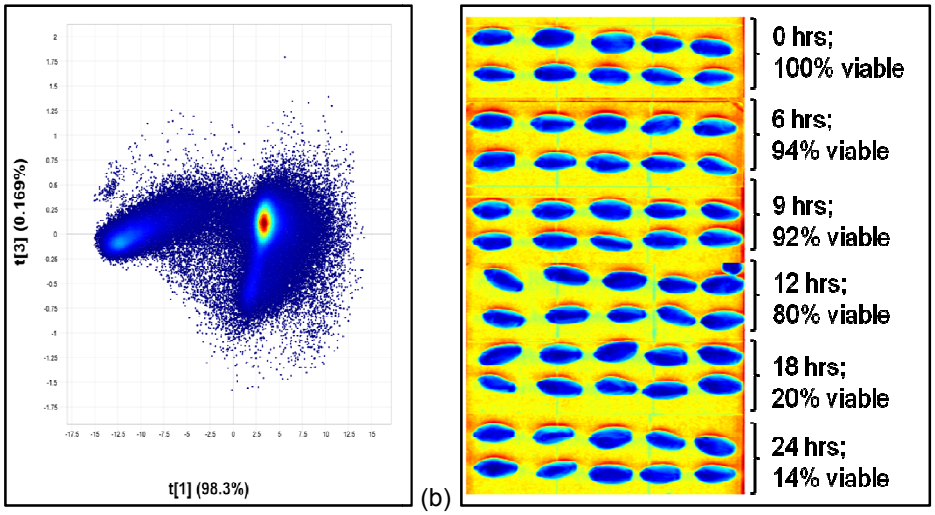
Moisture content

The moisture content of the samples was determined according to the moisture - modified vacuum-oven method (AACC method 44-40, 2010).

Results and discussion

Image cleaning

Duzi (150 µm spatial resolution image) was chosen to demonstrate the identification and removal of unwanted pixels. The PCA score plot (**Fig. 4.3a**) of PC 1 (SS 98.3%) vs. PC 3 (SS 0.169%) revealed three clusters (SiC sandpaper, adhesive marker and wheat kernels). The clusters were selected in the score plot and projected onto the score image (**Fig. 4.3b**), to create a classification plot (**Fig. 4.4a**) and image (**Fig. 4.4b**). The unwanted pixels (background, dead pixels, shading and adhesive marker) were selected and removed to clean the images.



(a) (b)

Figure 4.3 (a) Uncleaned Duzi PCA score plot (PC 1 vs. PC 3) displaying three clusters; and (b) uncleaned PCA score image (PC 1).

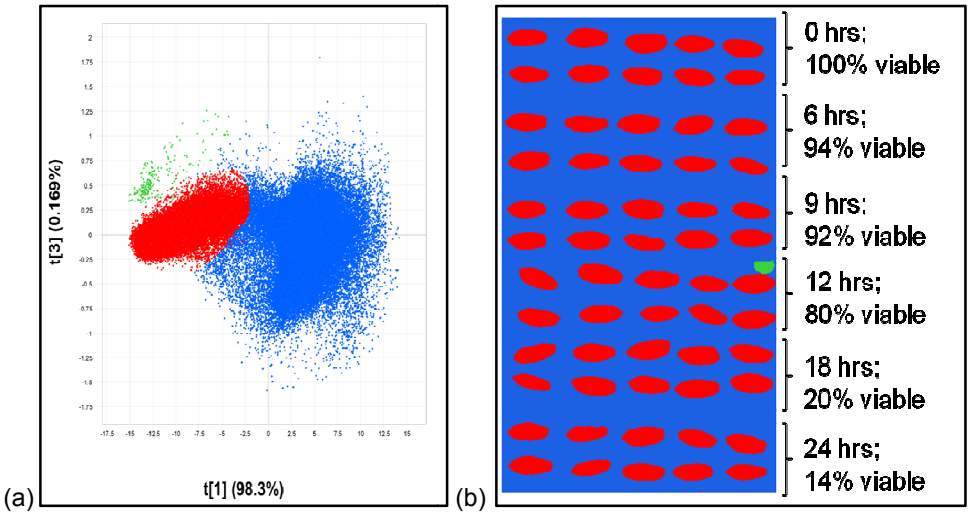


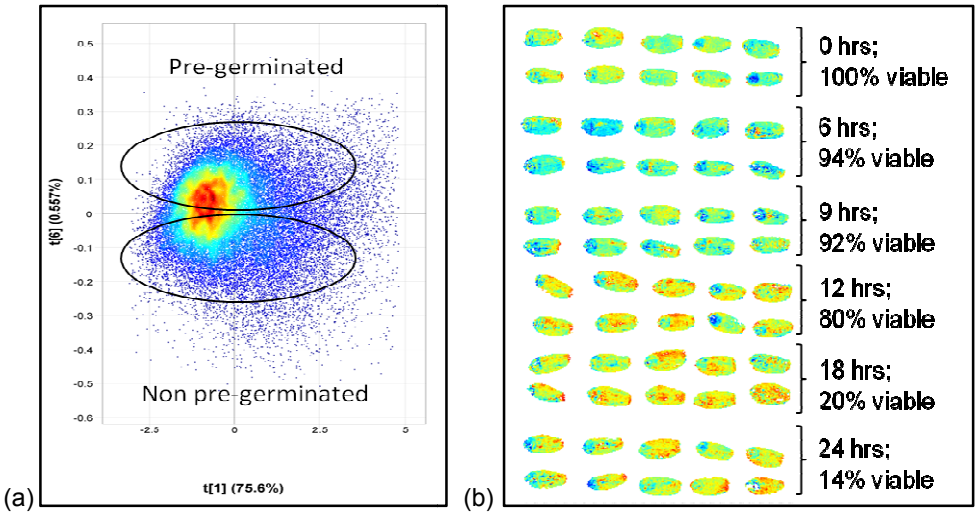
Figure 4.4 (a) Duzi PCA score plot (PC 1 vs. PC 3) with selected areas projected onto the score image to create (b) classification image (PC 1) (blue = background, red = wheat kernels and green = adhesive marker)

Analysis of cleaned 150 µm spatial resolution hyperspectral images

Duzi

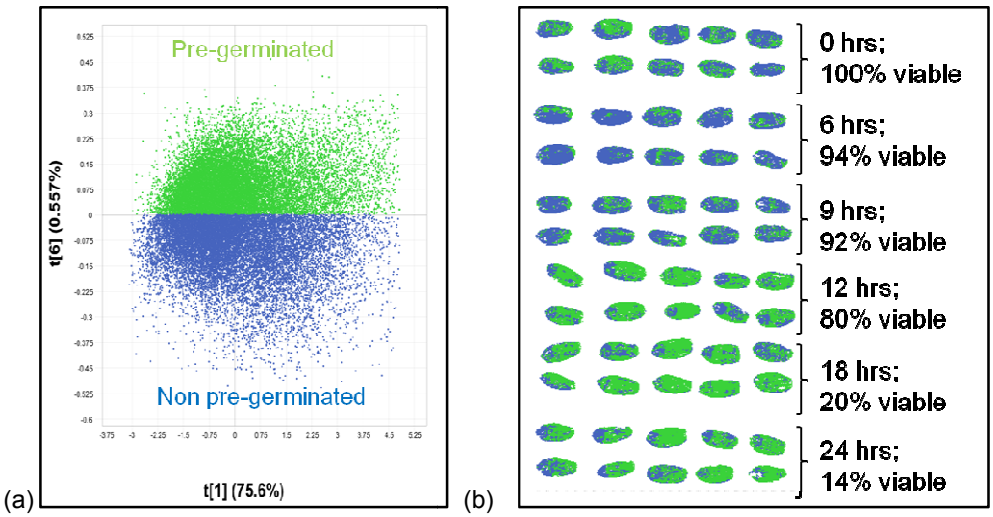
After the removal of unwanted pixels the PCA score images and score plots (37292 pixels remaining) of Duzi were analysed for regions of interest. Different combinations of PCs were plotted against each other and inspected. The PCA score plot (**Fig. 4.5a**) of PC 1 (SS 75.6%) vs. PC 6 (SS 0.557%) revealed no well defined clusters. However, the knowledge gained from a previous study (Chapter 3, page 47), aided the observation of a trend. By using the information of the location of the pixels and the determined viability (**Table 4.1, Fig. 4.5b**) of the Duzi samples, the pixels on the negative side of PC 6 were identified as non pre-germinated and the positive side were identified as pre-germinated kernels. The PCA score image (**Fig. 4.5b**) of PC 6 shows a difference between the subsets. These differences were observed through the use of heat map colouring in the score image. Pixels with similar score values are indicated with similar colours, in this case indicating similarity in chemical composition. Thus, the 0 to 9 hrs subsets in the score image are similar in chemical composition and the 12 to 24 hrs subsets are similar in chemical composition. When correlated with the tetrazolium test (**Table 4.1** and **Fig. 4.5b**), the kernels from the 0 to 9 hrs subsets were identified as non pre-germinated and the kernels from the 12 to 24 hrs subsets were identified as pre-germinated.

A classification plot (**Fig. 4.6a**) and classification image (**Fig. 4.6b**) were formed by selecting different pixels in the PCA score plot and projecting it onto the PCA score image. The classification plot and image provides a means of distinguishing between the pre-germinated and non pre-germinated kernels.



(a) (b)

Figure 4.5 (a) Duzi PCA score plot of PC 1 vs. PC 6 showing non pre-germinated and pre-germinated score values; and (b) PCA score image of PC 6 showing the pre-germinated (12- 24 hrs) and non pre-germinated (0 – 9 hrs) subsets.



(a) (b)

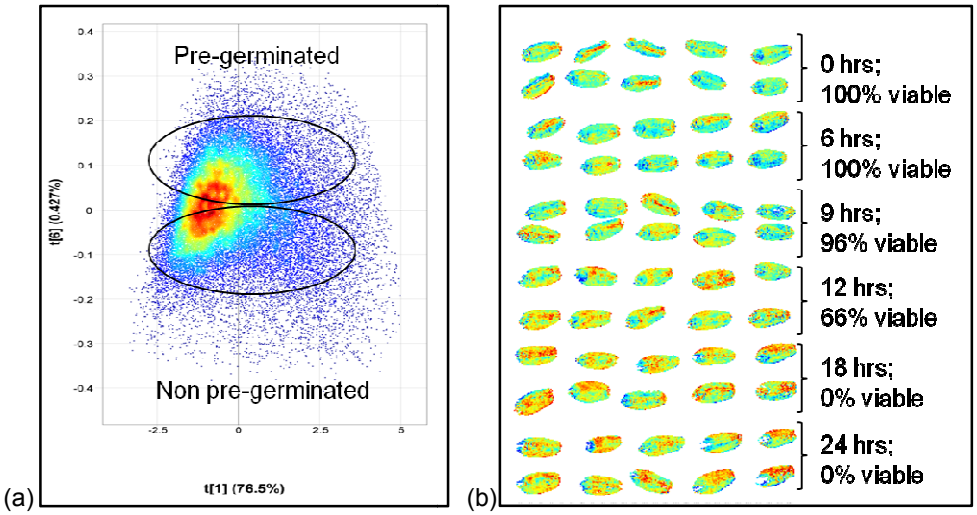
Figure 4.6 (a) Duzi classification plot of PC 1 vs. PC 6; and (b) classification image of PC 6 (green = pre-germinated and blue = non pre-germinated).

Table 4.1 Tetrazolium test results as percentage viability of subsets

Cultivar	Subsets (time exposure to distilled water in hrs)	Viability (%)
Duzi	0	100
Duzi	6	94
Duzi	9	92
Duzi	12	80
Duzi	18	20
Duzi	24	14
Biedou	0	100
Biedou	6	100
Biedou	9	96
Biedou	12	66
Biedou	18	0
Biedou	24	0
Kariega	0	100
Kariega	6	96
Kariega	9	80
Kariega	12	72
Kariega	18	70
Kariega	24	67

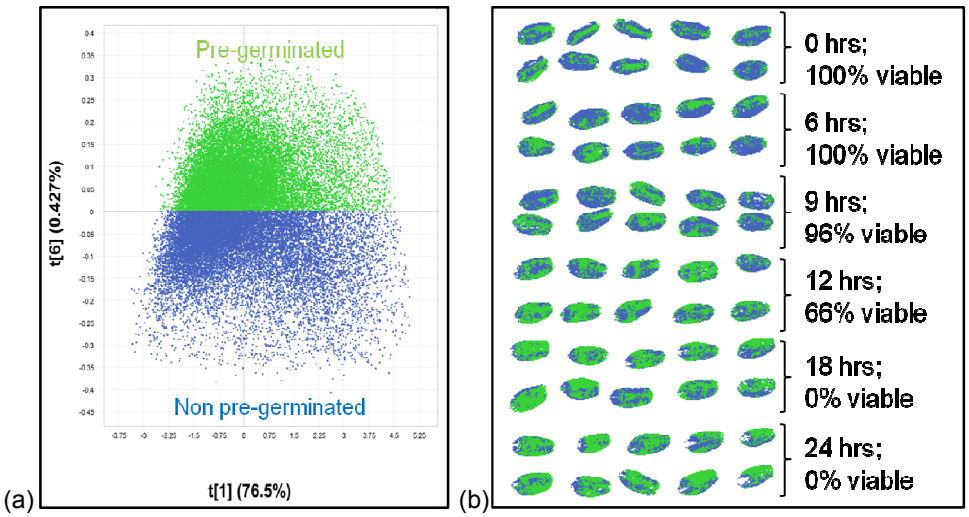
Biedou – 150 µm spatial resolution image

The PCA score plot (**Fig. 4.7a**) of PC 1 (SS 76.5%) vs. PC 6 (SS 0.427%) did not reveal any clusters, but a trend was observed in the direction of PC 6. Employing the information of the location of the pixels and the determined viability (**Table 4.1** and **Fig. 4.7b**), the pixels on the positive side of PC 6 was identified as pre-germinated and the negative side were identified as non pre-germinated kernels (**Fig. 4.7a**). The kernels in the 0 hrs subset was identified as non pre-germinated (100% viable) and the kernels in the 24 hrs subset was identified as pre-germinated (0% viable). The differences between these two subsets provided the means to categorize the kernels within a subset. Thus, the kernels from the 0 to 9 hrs subsets were identified as non pre-germinated and the kernels from the 12 to 24 hrs subsets were identified as pre-germinated. A classification plot (**Fig. 4.8a**) and classification image (**Fig. 4.8b**) were formed, to show the relationship between the pixels in the score plot and the pixels in the score image.



(a) (b)

Figure 4.7 (a) Biedou PCA score plot of PC 1 vs. PC 6 showing non pre-germinated and pre-germinated score values; and (b) PCA score image of PC 6 showing the pre-germinated (12 – 24 hrs) and non pre-germinated subsets (0 – 9 hrs).



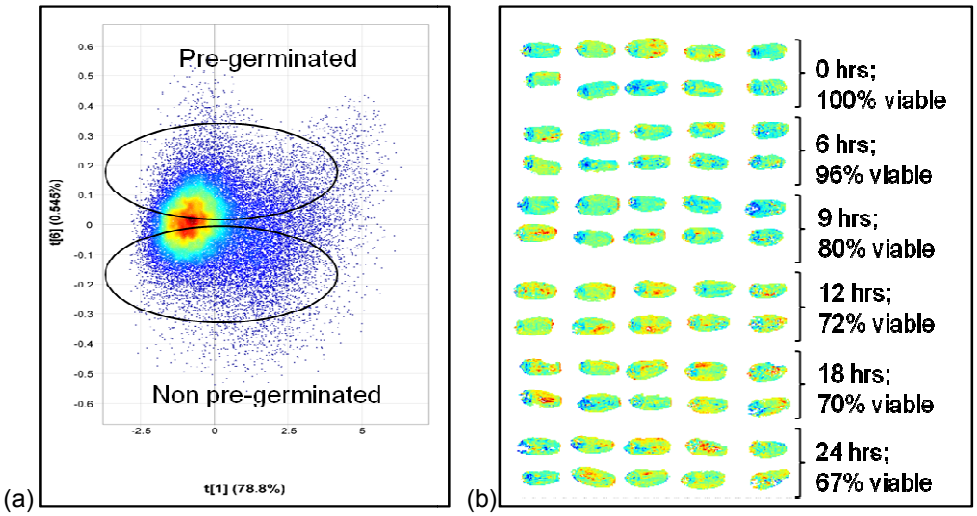
(a) (b)

Figure 4.8 (a) Biedou classification score plot of PC 1 vs. PC 6 ; and (b) classification image of PC 6 (green = pre-germinated and blue = non pre-germinated).

Kariega - 150 µm spatial resolution image

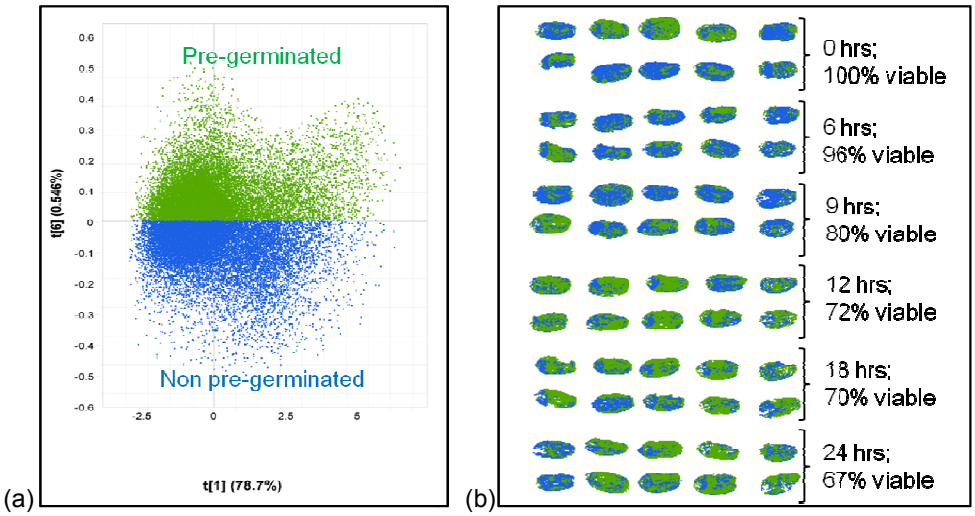
Image analysis of Kariega produced similar results to Duzi and Biedou. The location of the pixels in the score plot (**Fig. 4.9a**) and the determined viability of the kernels provided the means to identify the pixels in the score plot. The pixels on the positive side of PC 6 were identified as pre-germinated and the negative side were identified as non pre-germinated kernels. The PCA score image (**Fig. 4.9b**) of PC 6 (SS 0.545%) reveals a difference between the pre-germinated (24 hrs, 67% viable) and the non pre-germinated (0 hrs subset, 100% viable) subsets. Thus, the kernels from the 6 and 9 hrs subsets in the score image (similar in chemical composition to 0 hrs subset) are non pre-germinated and the kernels from the 12 and 18 hrs subsets are pre-germinated. The classification plot (**Fig. 4.10a**) and classification image (**Fig. 4.10b**) reveal that the pixels on the

positive side of PC 6 are associated with the pre-germinated subsets and the pixels on the negative side of PC 6 are associated with the non pre-germinated subsets.



(a) (b)

Figure 4.9 (a) Kariega PCA score plot of PC 1 vs. PC 6 showing non pre-germinated and pre-germinated score values; and (b) PCA score image of PC 6 showing the different subsets (12 – 24 hrs = pre-germinated and 0 – 9 = non pre-germinated).



(a) (b)

Figure 4.10 (a) Kariega classification score plot of PC 1 vs. PC 6; and (b) classification image of PC 6 (green = pre-germinated and blue = non pre-germinated).

Loading line plots

The loading line plots of Duzi, Kariega and Biedou were similar, thus only the loading line plot of Duzi will be discussed (**Fig. 4.11**). The peak at 1920 nm indicates the presence of water; water has large absorption bands with peak maxima around 1920 nm (combination bands) (Zhou *et al.*, 2003). This peak is positively weighted in PC 6 and can be associated with the pre-germinated kernels. The moisture analysis results revealed similar moisture content for the 0 and 24 hrs subsets, thus the differences between the subsets was not because of differences in the moisture

content (**Table 4.2**). Rather, the water environment differed between pre-germinated and non pre-germinated kernels. In pre-germinated kernels starch and protein have been hydrolysed, this process requires the interaction of water with enzymes, starch or proteins (see Chapter 2, page 13). The peaks at 1725 nm (C-H stretching first overtone) and 2294 nm (N-H stretching and C=O stretching) are related to starch and protein, respectively (Osborne *et al.*, 1993). The starch peak is negatively weighted and can be associated with the non pre-germinated score values in the score plot of PC 6. Non pre-germinated wheat kernels would include more intact starch molecules than pre-germinated kernels (Duffus, 1987; Fincher & Stone, 1993; Kent & Evers, 1994). The protein peak is positively weighted and can be associated with the pre-germinated score values in the score plot of PC 6. The protein molecules in the pre-germinated kernels would be degraded by enzymes during the pre-germination process.

Table 4.2 Moisture content of different wheat samples

Cultivar	Moisture content of subsets (%)	
	0 hrs	**24 hrs**
Duzi	6.84	7.20
Biedou	6.92	7.57
Kariega	7.12	7.58

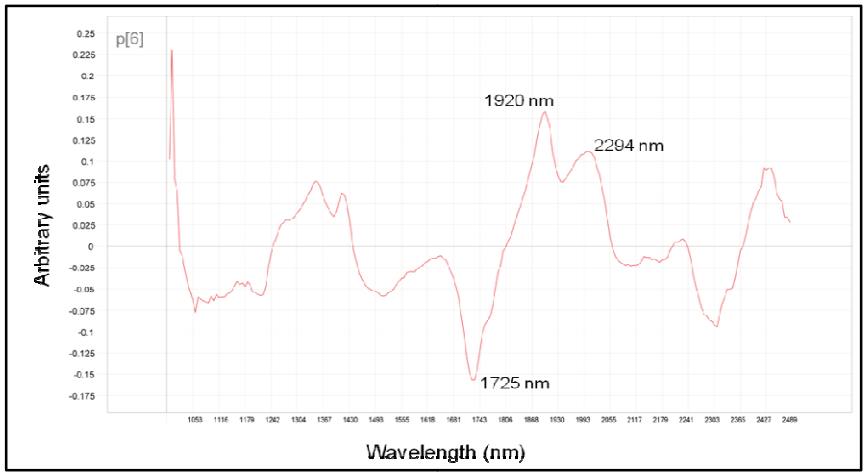


Figure 4.11 Loading line plot of Duzi (PC 6), showing the compounds contributing to variation: starch (1725 nm), water (1920 nm) and protein (2294 nm).

Partial least squares discriminant analysis – 150 µm spatial resolution images

For each of the cultivars, a PLS-DA model was created; these models were evaluated and the model with the most explained variance in Y was chosen to predict images of the other two cultivars. The PLS-DA model for Duzi presented the most explained variation in Y (63% after six components) and was applied to the other two cultivars. Biedou had 60% explained variation in Y (after six components) and Kariega 62% after six components.

Pre-germinated kernels predicted as non pre-germinated will result in false positives that is detrimental for the milling industry. Bread production requires a small amount of alpha- amylase, however the miller prefers non pre-germinated wheat to which an alpha-amylase source (in the appropriate level) is added (Kruger & Reed, 1988). Pre-germinated wheat predicted as non pre-germinated will lead to an excess amount of alpha-amylase in the flour which will have deleterious effects on product end quality. A false negative will arise if non pre-germinated kernels are not predicted as such. This type of false positives and negatives will be emphasized.

Table 4.3 specifies the amount of pixels assigned to each class (pre-germinated and non pre-germinated) of Biedou. The model correctly predicted 81.63% of the non pre-germinated class. The model also correctly predicted 100% of the pre-germinated class with a 6.82% over prediction of this class.

The model correctly predicted 92.21% of the non pre-germinated class (**Table 4.4**). Of the 17984 pixels assigned to the pre-germinated class, 16770 pixels (93.25%) were correctly predicted, showing the model's ability to correctly predict class membership. This PLS-DA model of Duzi predicted fairly well.

The prediction images (**Figs. 4.12** and **4.13a**) of Biedou and Kariega illustrates the localisation of the predicted classes. The classification image (**Figs. 4.12b** and **4.13b**) shows the assigned classes for Biedou and Kariega. Model performance can be judged by comparing the prediction and classification images to one another. The prediction images compared to the classification images confirms good predictions

Table 4.3 Prediction table for PLS-DA discrimination of Biedou

	Assigned number of pixels	**Predicted number of pixels (% predicted)**
Non pre-germinated	17649	14407 (81.63)
Pre-germinated	17649	18853 (106.82)
Not classified	0	2038
Total	35298	35298

Table 4.4 Prediction table for PLS-DA discrimination of Kariega

	Assigned number of pixels	**Predicted number of pixels (% predicted)**
Non pre-germinated	15603	14389 (92.21)
Pre-germinated	17984	16770 (93.25)
Not classified	0	2428
Total	33587	33587

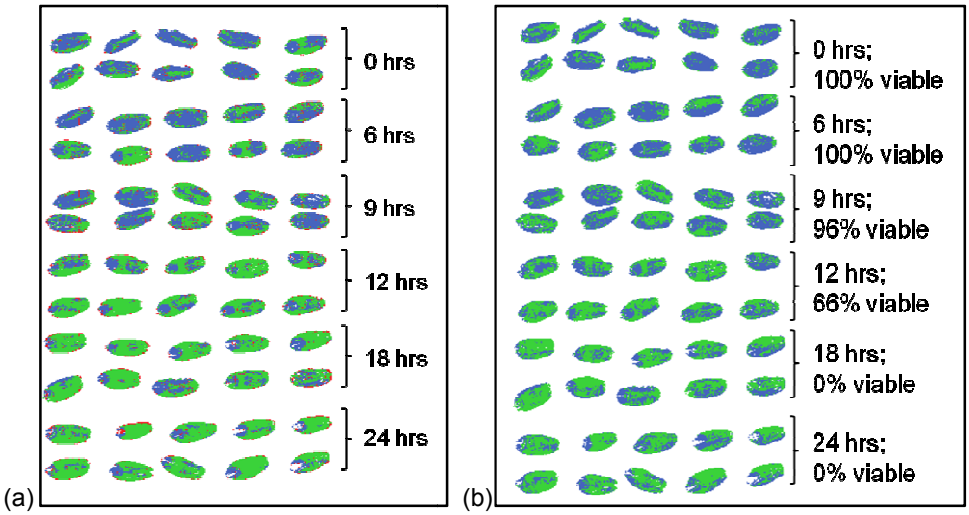


Figure 4.12 (a) Biedou PLS-DA prediction image and (b) classification image (blue = assigned/ predicted non pre-germinated, green = assigned/ predicted pre-germinated and red = not classified).

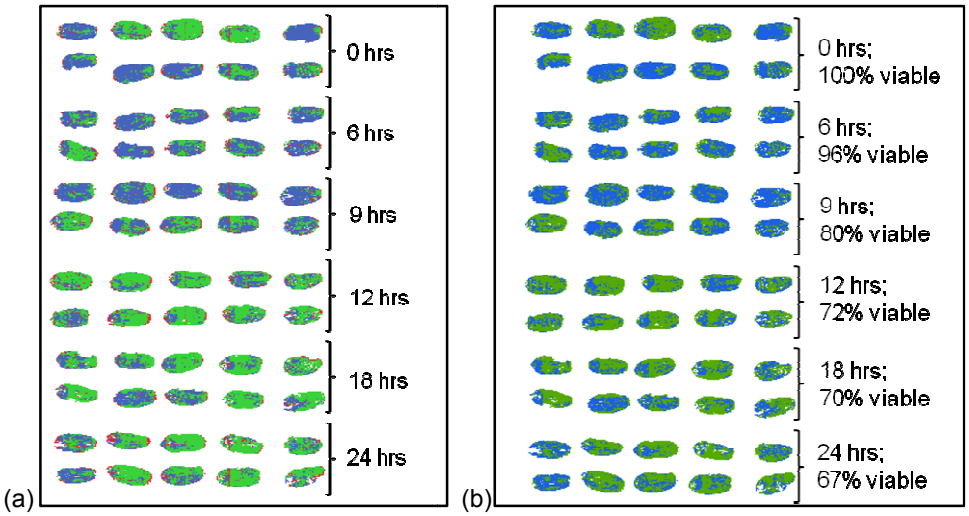


Figure 4.13 (a) Kariega PLS-DA prediction image and (b) classification image (blue = assigned/ predicted non pre-germinated, green = assigned/ predicted pre-germinated and red = not classified).

<u>30 µm spatial resolution hyperspectral images</u>

Analysis of cleaned images

Six columns containing each subset were imaged for both of the cultivars (**Fig 4.2**). Only one representative column from each cultivar will be discussed to minimise repetition and will be called Elands-1 and SST 398-1. Two other randomly picked columns will be used for PLS-DA prediction and these columns will be called Elands-2, Elands-3, SST 398-2 and SST 398-3.

Elands-1 – 30 µm spatial resolution image

The Elands-1 PCA score plot (**Fig. 4.14a**) of PC 1 (SS 85.4%) vs. PC 4 (SS 1.86%) revealed no clusters. The score values in the direction of PC 4 accounted for the difference between pre-

germinated (positive score values) and non pre-germinated (negative score values) kernels. This was identified by interactively exploring different combinations of score plots and images.

Pre-germinated and non pre-germinated kernels are distinguished from one another in the Elands-1 PCA score image (**Fig. 4.14b**). However, kernels from the 6, 12 and 24hrs subsets were determined as non viable (pre-germinated), but had a similar chemical composition according to the score image as the non pre-germinated kernels. The endosperm of these kernels may have partially been degraded by hydrolytic enzymes, but complete hydrolysis perhaps did not occur. The NIR hyperspectral imaging method is possibly detecting degree of endosperm hydrolysis rather than kernel viability. The tetrazolium test was used to test viability; a non viable embryo will not induce a colour change (tetrazolium chloride to formazan). A kernel may be non viable as described by the tetrazolium test, but have only been partially hydrolysed by hydrolytic enzymes. This will lead to a non viable result by the tetrazolium test and a viable result by the NIR hyperspectral imaging method. NIR hyperspectral imaging could, consequently, be employed as an indirect method for the detection of endosperm hydrolysis.

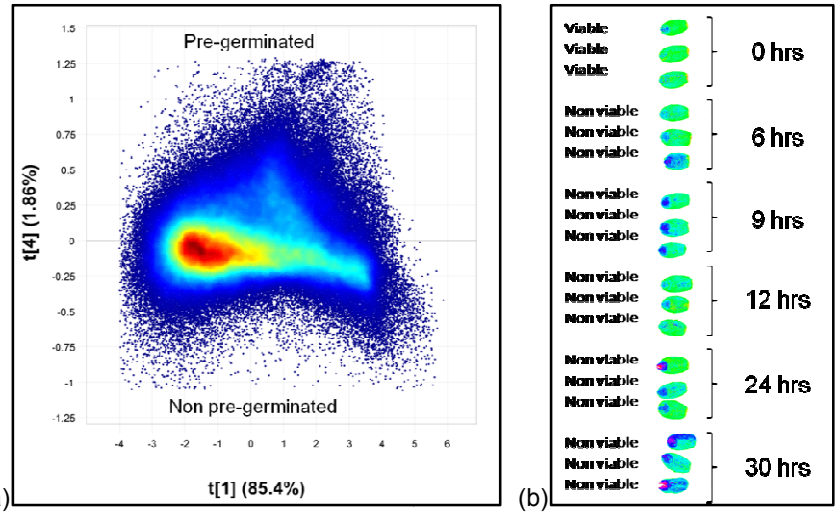
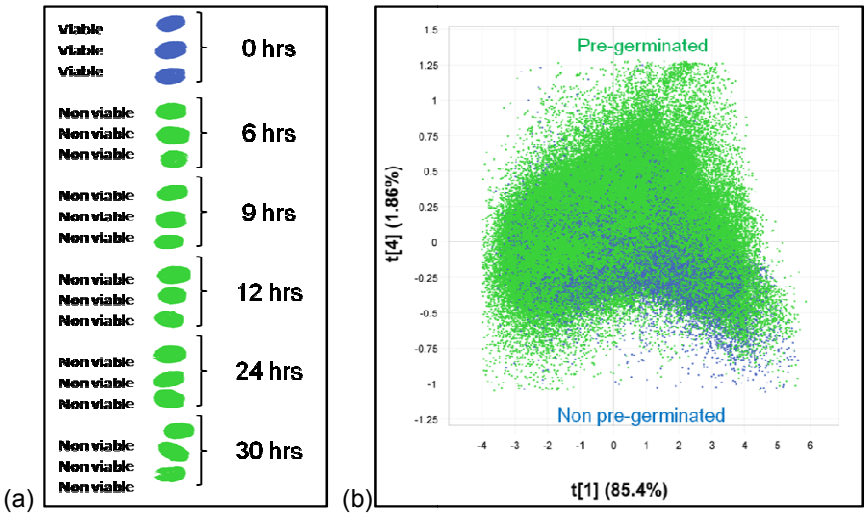


Figure 4.14 (a) Elands-1 PCA score plot of PC 1 vs. PC 4 showing pre-germinated score values on the positive side of PC 4 and non pre-germinated score values on the negative side of PC 4; and (b) Elands-1 PCA score image of PC 4 showing differences between pre-germinated and non pre-germinated kernels.

The first classification plot (**Fig. 4.15a**) and image (**Fig. 4.15b**) was created by selecting the kernels in the score image belonging to either the pre-germinated or non pre-germinated class, as identified by the tetrazolium test. **Fig. 4.15a** shows that the few pixels associated with the viable kernels (**Fig. 4.15b**) are located mostly on the negative side of PC 4.

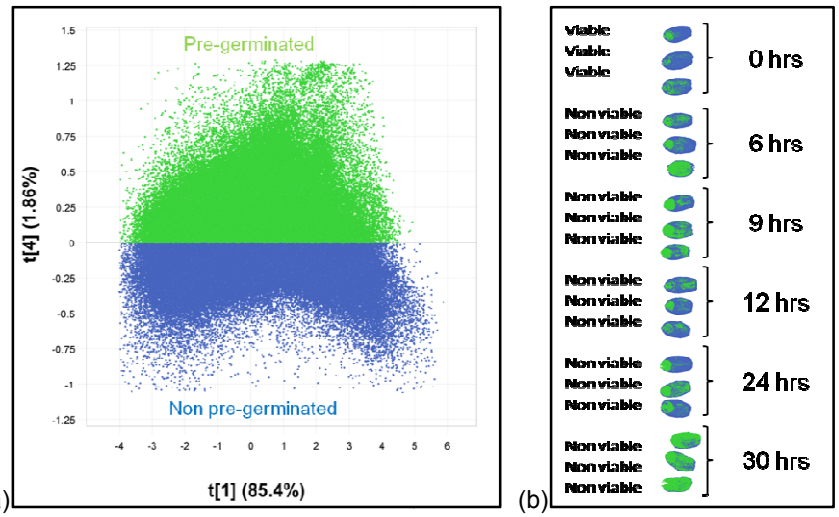
The second classification plot (**Fig. 4.16a**) and classification image (**Fig. 4.16b**) were created by selecting the pixels in the score plot on the negative side (non pre-germinated) and positive side (pre-germinated) of the PC and projecting it onto the score image. The classification image (**Fig.**

4.16b) shows that most of the kernels were classified as non pre-germinated (blue), except the 30 hrs subset.



(a) (b)

Figure 4.15 (a) Elands-1 classification image showing the assigned pre-germinated (non viable) and non pre-germinated (viable) kernels with selection based on tetrazolium test results; and (b) classification plot with classes selected in the score image and projected onto the score plot (pre-germinated = green and non pre-germinated = blue).



(a) (b)

Figure 4.16 (a) Elands-1 classification plot of PC 1 vs. 4 showing the assigned pre-germinated and non pre-germinated classes; and (b) classification image of score classes projected onto score image (pre-germinate = green and non pre-germinated = blue).

SST 398-1 - 30 µm spatial resolution images

The PCA score plot (**Fig. 4.17a**) of PC 1 (SS 86.9%) vs. PC 3 (SS 2.00%) showed only slight grouping of score values. However, the PCA score image (**Fig. 4.17b**) of PC 3 revealed differences between the determined pre-germinated and non pre-germinated kernels. There were kernels from the 6 and 9 hrs subsets that had a similar composition according to the score image

as the non pre-germinated kernels, but these kernels were indicated as pre-germinated by the tetrazolium test. This was possibly a result of incomplete hydrolysis of the endosperm. Two sets of classification images (**Fig. 4.18a** & **4.19a**) and plots (**Fig. 4.18b** & **4.19b**) was created as described above. **Fig. 4.18a** shows that the few pixels associated with the non pre-germinated kernels (**Fig. 4.18b**) are mostly located on the negative side of PC 3; the pixels associated with the pre-germinated kernels are much more because only two kernels were determined as viable by the tetrazolium test. **Fig 4.19b** reveals that the pixels on the negative side of PC 3 are mostly associated with the 0, 6 and 9 hrs subsets and the pixels on the positive side of PC 3 are mostly associated with the 12 and 30 hrs subsets in the classification image.

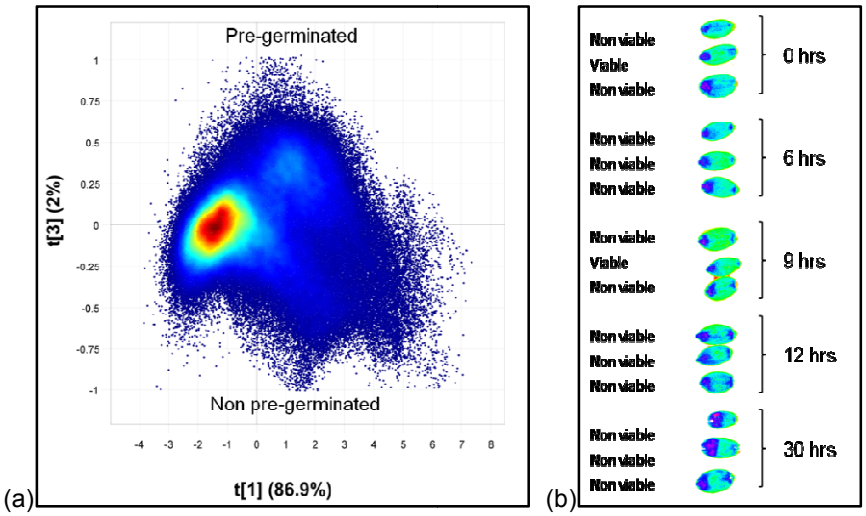


Figure 4.17 (a) SST 398-1 PCA score plot of PC 1 vs. PC 3 showing pre-germinated (positive) and non pre-germinated (negative) score values; and (b) SST 398-1 PCA score image of PC 3 showing differences between pre-germinated and non pre-germinated kernels.

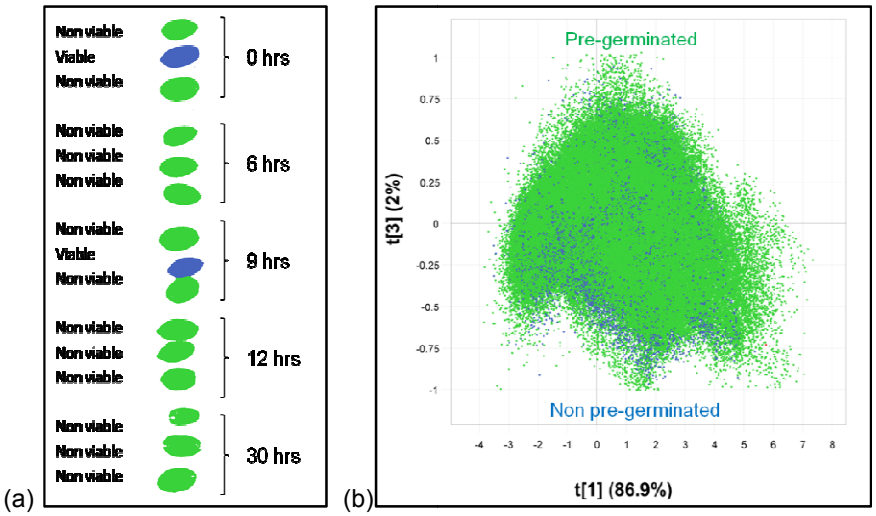


Figure 4.18 (a) SST 398-1 classification image showing the assigned pre-germinated (non viable) and non pre-germinated (viable) kernels with selection based on tetrazolium test results; and (b) classification plot with classes selected in the score image and projected onto the score plot (pre-germinated = green and non pre-germinated = blue).

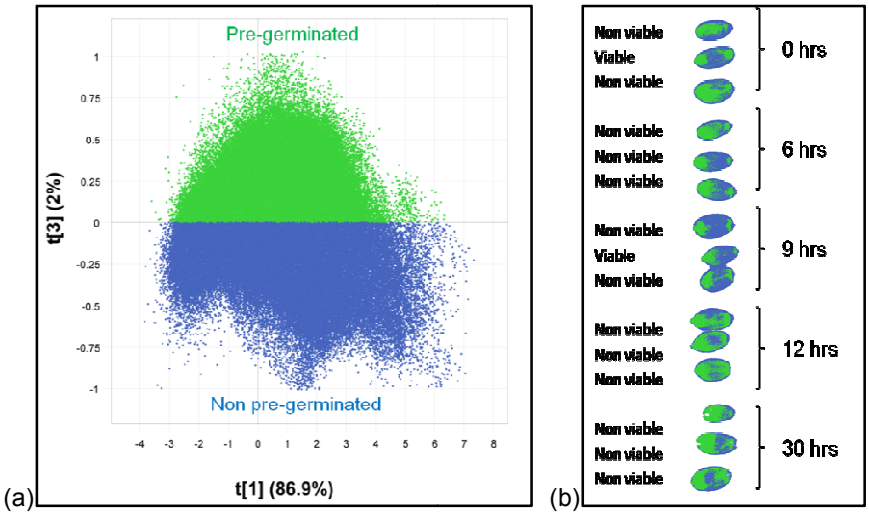


Figure 4.19 (a) SST 398-1 classification plot of PC 1 vs. 3 showing the assigned pre-germinated and non pre-germinated classes; and (b) classification image of score classes projected onto score image (pre-germinate = green and non pre-germinated = blue).

Loading line plots – 30 µm spatial resolution images

The loading line plots of Elands and SST 398 differed slightly, but the main compounds contributing to the variation was starch, water and protein. These compounds were the main compounds contributing to the variation in the 150 µm spatial resolution images.

For Elands peaks (**Fig. 4.20**) were observed at 1471 nm (protein - N-H stretching first overtone), 1725 nm (starch - C-H stretching first overtone), 1940 nm (water - O-H stretching and O-H deformation) and 2252 nm (starch - O-H stretching and O-H deformation) (Osborne *et al.*, 1993). The water peak is negatively weighted and can be associated with the non pre-germinated scores in the score plot. The moisture content results (**Table 4.5**) indicated that the differences between kernels were not due to moisture content differences. The protein peak is negatively weighted and can be associated with the negative (pre-germinated) scores in the score plot. The starch peaks are positively weighted and may be associated with the pre-germinated score in the score plot (positive). This loading line plot is different from the 150 µm spatial resolution images, where the water and protein peaks were associated with the pre-germinated kernels and the starch peak was associated with the non pre-germinated kernels. It is not clear why this happened, a possible reason could be the kernels from the 30 µm spatial resolution images were not pre-germinated to the same extent as the 150 µm spatial resolution images.

The peaks observed for SST 398 (**Fig. 4.21**) was at 1415 nm (starch - 2x C-H stretching and C-H deformation), 1940 nm (water - O-H stretching and O-H deformation) and 2252 nm (starch - O-H stretching and O-H deformation) (Osborne *et al.*, 1993). The water peak is positively weighted, thus associated with the pre-germinated scores in the score plot of PC 3. The water environment of pre-germinated kernels differs from non pre-germinated kernels. The variation due to water could be because of the different bonding environments of water induced during pre-germination. The

starch peaks are negatively weighted and can be associated with the non pre-germinated scores in the score plot of PC 3. Non pre-germinated kernels would contain more intact starch molecules than pre-germinated kernels.

According to other NIR hyperspectral imaging research done on wheat the wavelengths used to discriminated between pre-germinated and non pre-germinated kernels were 1680, 1101.7 and 1305.1 nm (Smail *et al.*, 2006; Singh *et al.*, 2009b). According to this study the water peak at 1940 nm dominated pre-germinated, non pre-germinated discrimination. None of the other researchers observed the 1940 nm water peak, because their studies only included a NIR wavelength range below 1700 nm.

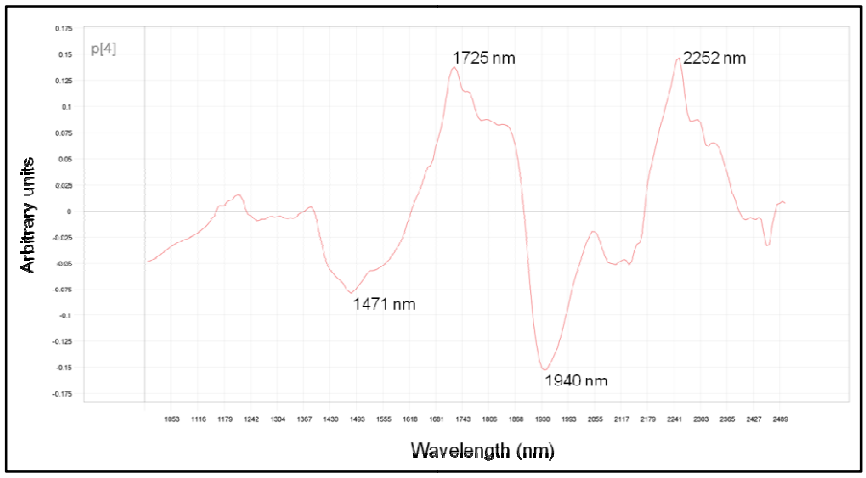


Figure 4.20 Loading line plot of Elands PC 4, illustrating the compounds responsible for variation; starch (1725 nm and 2252 nm), water (1940 nm) and protein (1471 nm).

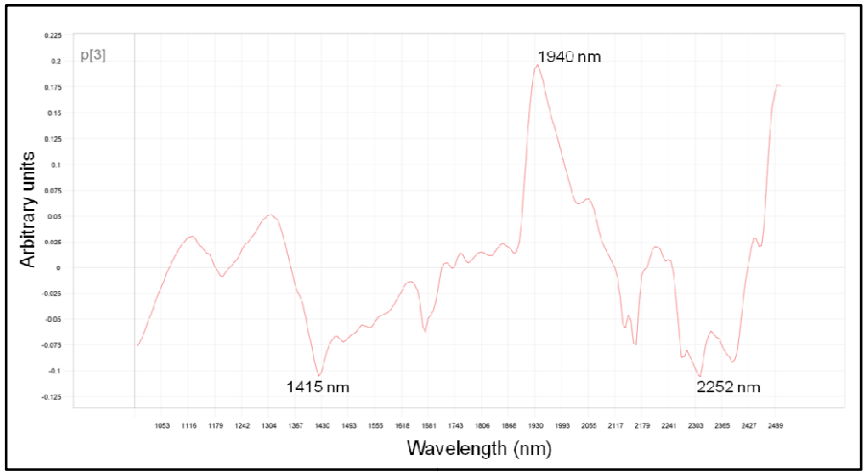


Figure 4.21 Loading line plot of SST 398 PC 3, illustrating the compounds responsible for variation; starch (1415 nm and 2252 nm) and water (1940 nm).

Table 4.5 Moisture analysis results for each subset of the three wheat cultivars

Cultivars	Subsets (time exposure to distilled water in hrs)	Moisture content (%)
Elands	0	8.69
	6	8.70
	9	8.50
	12	8.35
	24	8.52
	30	8.74
SST 398	0	8.87
	6	8.93
	9	8.72
	12	8.92
	30	8.73

Partial least squares discriminant analysis – 30 µm spatial resolution images

Elands-2 and Elands-3

Two PLS-DA models were created for Elands-1 and both of these models were employed to predict two other images of Elands (2 and 3). The explained variation in Y for model one (M1) was 23.05% and model two (M2) was 60.01% after six components. **Table 4.6** contains the amount of pixels assigned to both models one and two of Elands 2 and 3.

The prediction data for Elands-2 and Elands-3 (**Table 4.6**) revealed that the models did not perform well. M1 resulted in 88.63% (Elands-2) and 80.09% (Elands-3) false negatives and M2 resulted in 39.80% (Elands-2) and 16.91% (Elands-3) false positives. Even though false negatives are the less serious of the two errors (Contal *et al.*, 2002), 88.63% and 80.09% false negatives are not acceptable. **Figs. 4.22** and **4.23** show the prediction images and the classification images containing the predicted and assigned pre-germinated and non pre-germinated pixels of Elands-2 and Elands-3. The differences between the prediction and classification images corroborates that the models did not perform well.

Better classification results were obtained in the study of Singh *et al*. (2009). Both the linear discriminant classifier and quadratic discriminant classifier calculated correctly (100%) classified non pre-germinated and pre-germinated kernels. A Mahalanobis discriminant classifier correctly identified 100% and 98.3% of pre-germinated and non pre-germinated kernels, respectively. The researchers expected good prediction results because the kernels were either non pre-germinated (no treatment with water) or pre-germinated (48 hrs pre-germination and visible sprouting recorded). The discriminating abilities of the models were facilitated by the extreme differences between the pre-germinated and non pre-germinated kernels. In this study kernels were pre-

germinated for several shorter periods of time (6 – 30 hrs). Different degrees of endosperm hydrolysis were apparent in the images and may have diminished the model's discriminating abilities. Models developed from only pre-germinated or non pre-germinated kernels may exhibit improved prediction results.

Table 4.6 PLS-DA prediction results for Elands-2 and Elands-3 of both models M1 and M2

		Assigned number of pixels	Predicted number of pixels (% pixels)
Elands-2	**M1**		
	Non pre-germinated	34548	3929 (11.37)
	Pre-germinated	267472	279091 (104.34)
	Not classified		19000
	Total	302020	302020
	M2		
	Non pre-germinated	153520	214622 (139.80)
	Pre-germinated	148500	62774 (42.27)
	Not classified		24624
	Total	302020	302020
Elands-3	**M1**		
	Non pre-germinated	55954	11141 (19.91)
	Pre-germinated	283984	315728 (111.17)
	Not classified		13069
	Total	339938	339938
	M2		
	Non pre-germinated	189474	221511 (116.91)
	Pre-germinated	150464	103617 (68.86)
	Not classified		14810
	Total	339938	339938

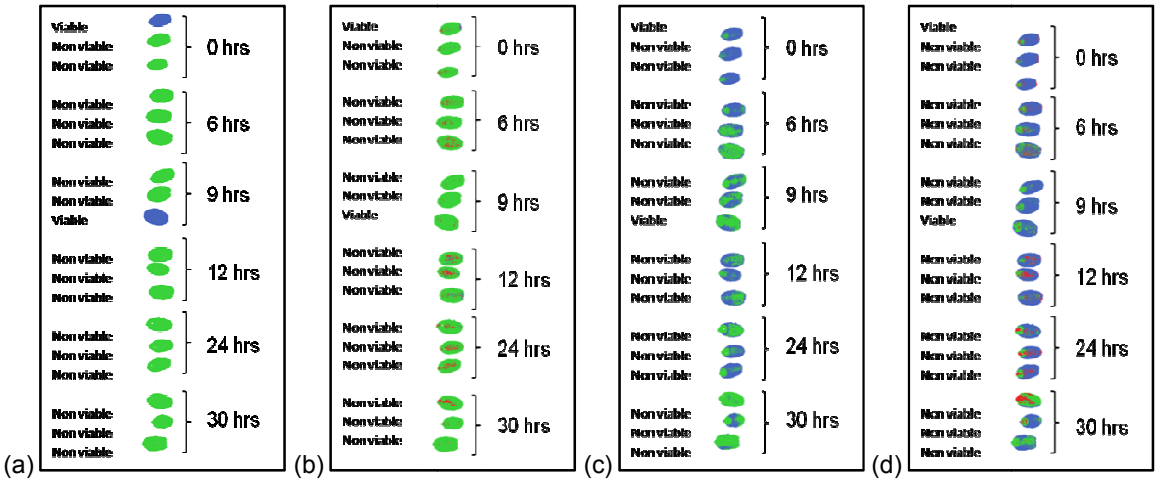


Figure 4.22 (a) Elands-2 classification image with class assignments based on score image; and (b) prediction image of Elands-2 with M1 (model developed according to the score image from Elands-1). (c) Elands-2 classification image with class assignment based on the score plot; and (d) prediction image of Elands-2 with M2 (model developed according to the score plot from Elands-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

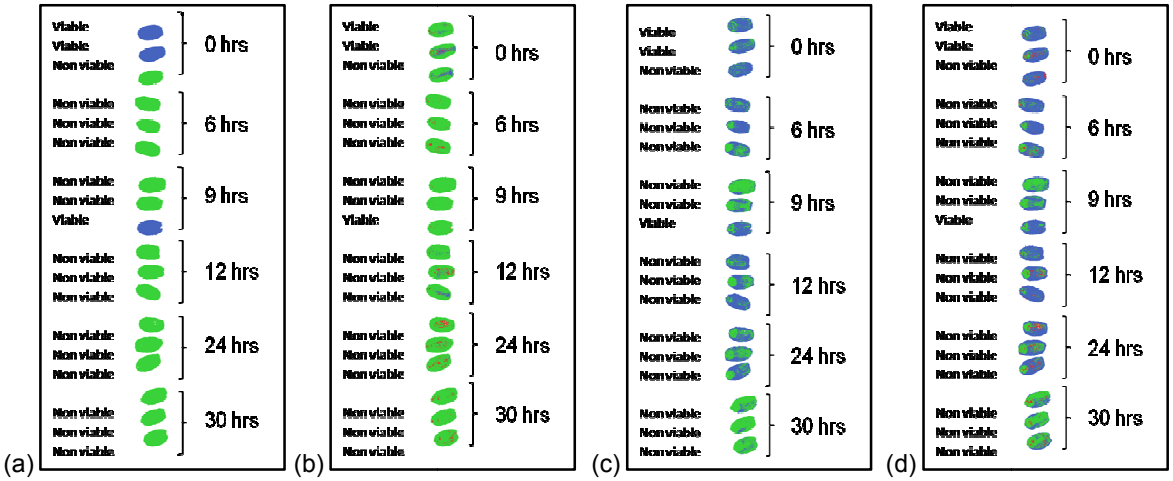


Figure 4.23 (a) Elands-3 classification image with class assignments based on score image; and (b) prediction image of Elands-3 with M1 (model developed according to the score image from Elands-1). (c) Elands-3 classification image with class assignment based on the score plot; and (d) prediction image of Elands-3 with M2 (model developed according to the score plot from Elands-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

SST 398-2 and SST 398-3

The explained variation in Y for M1 of SST 398-1 was 8.82% after six components and M2 was 60.01% after six components. The prediction results revealed that the PLS-DA models did not perform very well (**Table 4.7**). M1 predicted 99.96% and 99.71% false negatives for SST 398-2 and SST 398-3 respectively. M2 correctly predicted 86.10% of the non pre-germinated pixels for SSG564-2 and 25.18% (false positives) more pixels was predicted as non pre-germinated than

was assigned for SSG564-3. Neither of the models can be used to discriminate between pre-germinated and non pre-germinated wheat kernels. **Figs. 4.24** and **4.25** show the predicted and assigned pixels of each class for SST 398-2 and SST 398-3.

Only 15 (SST 398) and 18 (Elands) kernels were used to create the respective PLS-DA models. Although an enormous amount of pixels are used during model development, the tetrazolium test is only performed on 15 kernels. It would thus be better to develop models based on more kernels.

Table 4.7 PLS-DA prediction results for SST 398-2 and SST 398-3 of both models M1 and M2

		Assigned number of pixels	**Predicted number of pixels (% pixels)**
SST 398-2	**M1**		
	Non pre-germinated	35783	16 (0.04)
	Pre-germinated	247484	270991 (109.50)
	Not classified		12260
	Total	283267	283267
	M2		
	Non pre-germinated	144605	124511 (86.10)
	Pre-germinated	138662	143018 (103.14)
	Not classified		15738
	Total	283267	283267
SST 398-3	**M1**		
	Non pre-germinated	50161	146 (0.29)
	Pre-germinated	234679	276251(117.71)
	Not classified		8443
	Total	284840	284840
	M2		
	Non pre-germinated	135865	170076 (125.18)
	Pre-germinated	148975	103283 (69.33)
	Not classified		11481
	Total	284840	284840

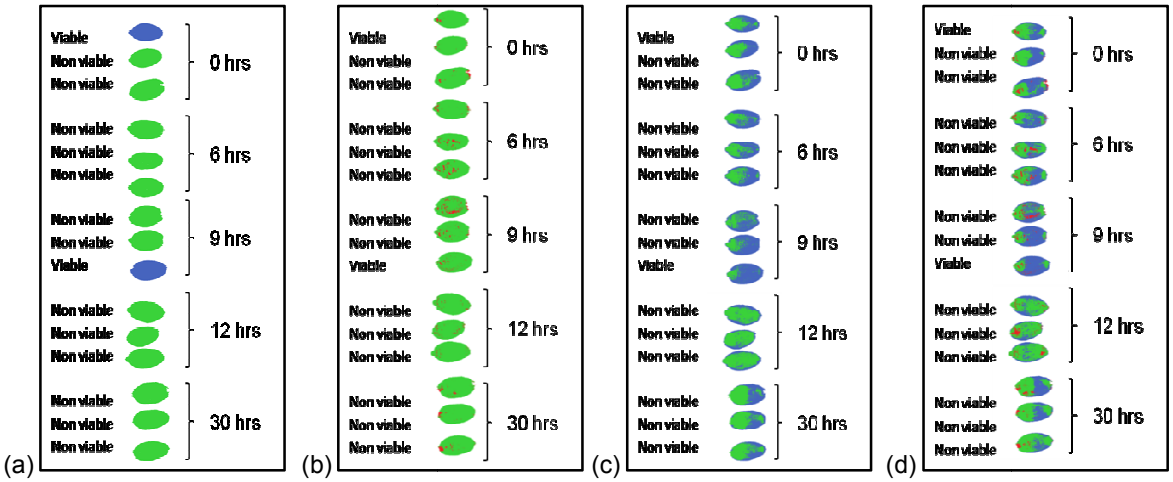


Figure 4.24 (a) SST 398-2 classification image with class assignments based on score image; and (b) prediction image of SST 398-2 with M1 (model developed from the score image of SST 398-1). (c) SST 398-2 classification image with class assignment based on the score plot; and (d) prediction image of SST 398-2 with M2 (model developed from the score plot of SST 398-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

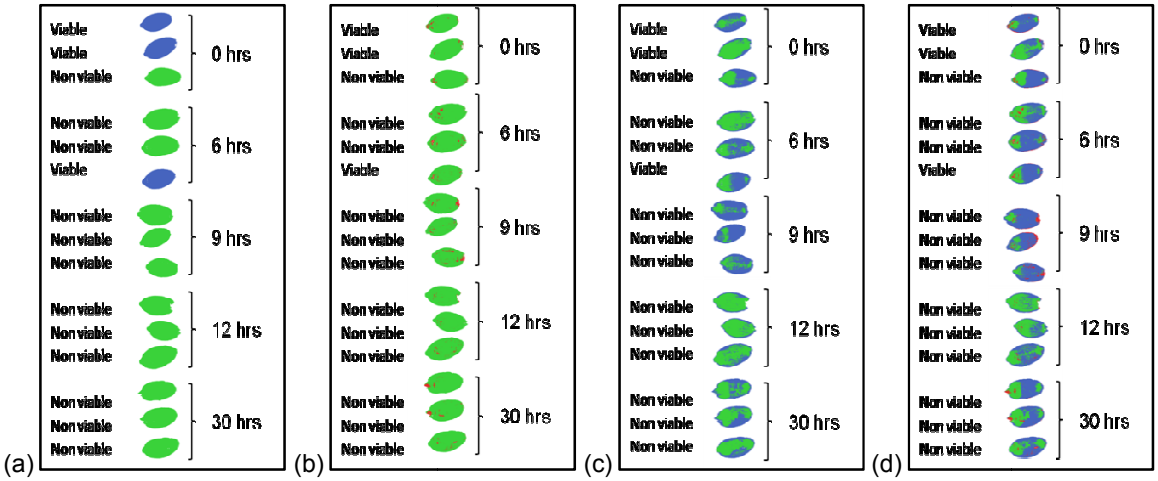


Figure 4.25 (a) SST 398-3 classification image with class assignments based on score image; and (b) prediction image of SST 398-3 with M1 (model developed from the score image of SST 398-1). (c) SST 398-3 classification image with class assignment based on the score plot; and (d) prediction image of SST 398-3 with M2 (model developed from the score plot of SST 398-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

Conclusion

The pre-germinated and non pre-germinated subsets of the 150 µm spatial resolution images showed differences between the 0 and 24 hrs subsets, this lead to the classification of pre-germinated and non pre-germinated kernels using NIR hyperspectral imaging. The 30 µm spatial resolution score images revealed differences between pre-germinated and non pre-germinated

kernels, however, some kernels were identified as pre-germinated by the tetrazolium test, but these kernels were similar in chemical composition according to the score images as the non pre-germinated kernels. NIR hyperspectral imaging could detect differences between the kernels due to chemical differences (e.g. starch and protein hydrolysis), and this method could be employed as an indirect method to determine endosperm hydrolysis.

The loading line plots revealed that water was a compound contributing to the variation in the images. Contradicting results were obtained with the loading line plots of the 30 µm spatial resolution images; the water and protein peaks were associated with the non pre-germinated kernels and the starch peak was associated with the pre-germinated kernels. The kernels from the 30 µm spatial resolution images could be different from the 150 µm spatial resolution images due to the extent to which pre-germination has occurred.

The PLS-DA models for the 150 µm spatial resolution images could distinguish between pre-germinated and non pre-germinated kernels of Biedou, with an accuracy of 81.63% for the non pre-germinated class and 100% (with a 6.82% over-prediction) of the pre-germinated class. The model predicted 92.21% (non pre-germinated) and 93.25% (pre-germinated) of the kernels correctly for Kariega. The PLS-DA models for the 30 µm spatial resolution images did not perform as well. Model 1 predicted more pixels as pre-germinated than was assigned to the image (false negatives) and model 2 predicted more pixels as non pre-germinated than was assigned (false positives). PLS-DA models developed from a bigger sample set and models developed only from kernels in the 0 and 30 hrs subsets may improve the prediction results. The models developed from the second sample set (30 µm spatial resolution images) did not result in accurate predictions, but the models created from the first sample set shows promise for the prediction of pre-germinated and non pre-germinated wheat kernels.

Future studies should include field pre-germinated kernels, since these kernels may differ from laboratory sprouted kernels (Kruger, 1994). To completely validate the PLS-DA models, they too should be tested by predicting images of field pre-germinated kernels.

References

AACC. (2010). *Approved Methods of the AACC*. Moisture – modified vacuum-oven method. St. Paul, Minnesota: American Association of Cereal Chemists.

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.

Contal, L., Leon, V. & Downey, G. (2002). Detection and quantification of apple adulteration in strawberry and raspberry purees using visible and near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **10**, 289-299.

Du Toit, G. (2009). Near infrared hyperspectral imaging and chemometrics for exploration and classification of whole wheat kernels. MSc Thesis Food Science, Stellenbosch University.

Duffus, C.M. (1987). Physiological aspects of enzymes during grain development and germination. In: *Enzymes: and their role in cereal technology* (edited by J.E. Kruger, D. Lineback & C.E. Stauffer). Pp. 83-111. St. Paul: American Association of Cereal Chemists.

Fincher, G.B. & Stone, B.A. (1993). Physiology and biochemistry of germination in barley. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 247-294. St. Paul: American Association of Cereal Chemists, Inc.

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.

Kent, N.L. & Evers, A.D. (1994). *Kent's technology of cereals.* New York: Elsevier Science Inc.

Koç, H., Smail, V.W. & Wetzel, D.L. (2008). Reliability of InGaAs focal plane array imaging of wheat germination at early stages. *Journal of Cereal Science*, **48**, 394-400.

Kruger, J.E. (1990). Instrumental assessment of sprout-damage in wheat at primary or terminal receival points. *Cereal Foods World*, **35**, 935-939.

Kruger, J.E. (1994). Enzymes of sprouted wheat and their possible technological significance. In: *Wheat Production, Properties and Quality* (edited by W. Bushuk & V.F. Rasper). Pp. 143-151. Glasgow: Blackie Academic and Professional.

Kruger, J.E. & Reed, G. (1988). Enzymes and colour. In: *Wheat: Chemistry and Technology* (edited by Y. Pomeranz). Pp. 441-487. St. Paul: American Association of Cereal Chemists.

Lorenz, K. & Valvano, R. (1981). Functional characteristics of sprout-damaged soft white wheat flours. *Journal of Food Science*, **46**, 1018-1020.

Mahesh, S., Manickavasagan, A., Jayas, D.S., Paliwal, J. & White, N.D.G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, **101**, 50-57.

Matsuo, R.R., Dexter, J.E. & MacGregor, A.W. (1982). Effect of sprout damage on durum wheat and spaghetti quality. *Cereal Chemistry*, **59**, 468-472.

Neethirajan, S., Jayas, D.S. & White, N.D.G. (2007). Detection of sprouted wheat kernels using soft x-ray image analysis. *Journal of Food Engineering*, **81**, 509-513.

Orth, R.A. & Shellenberger, J.A. (1988). Origin, production and utilisation of wheat. In: *Wheat: Chemistry and Technology* (edited by Y. Pomeranz). Pp. 1-13. St. Paul: American Association of Cereal Chemists.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). In: *Practical NIR spectroscopy with applications in food and beverage analysis* (edited by D. Browning). Pp. 1-35, 49-77. Essex: Longman Scientific & Technical.

Posner, E.S. (2000). Wheat. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 1-27. New York: Marcel Dekker, Inc.

Shahin, M.A. & Symons, S.J. (2008). Detection of hard vitreous and starchy kernels in amber durum wheat samples using hyperspectral imaging. *NIR news*, **19**, 16-18.

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

Singh, C.B., Jayas, D.S., Paliwal, J. & White, N.D.G. (2009b). Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry*, **86**, 256-260.

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.

Zhou, G.X., Ge, Z., Dorwart, J., Izzo, B., Kukura, J., Bicker, G. & Wyvratt, J. (2003). Determination and differentiation of surface and bound water in drug substances by near infrared spectroscopy. *Journal of Pharmaceutical Sciences*, **92**, 1058-1065.

Chapter 5

Near infrared (NIR) hyperspectral imaging for the detection of pre-germination in sorghum grains

Chapter 5

Near infrared (NIR) hyperspectral imaging for the detection of pre-germination in sorghum grains

Abstract

Pre-germination is the non visible germination of sorghum to such an extent that it is not useful in the sorghum industry. Near infrared (NIR) hyperspectral imaging was evaluated for its potential to distinguish between pre-germinated and non pre-germinated sorghum kernels. Sorghum cultivars, South African landrace 4442, PAN 8816 (150 µm spatial resolution images), PAN 8446 and PAN 8816 (30 µm spatial resolution images) were each divided into subsets, where each subset was subjected to distilled water for different time periods. NIR hyperspectral images of each cultivar were acquired using a SisuCHEMA hyperspectral imaging system (1000-2498 nm). Two spatial resolutions were used; 150 µm and 30 µm.

Principal component analysis (PCA) was performed on standard normal variate transformed and mean-centered data. PCA was effective in identifying pre-germinated and non pre-germinated individual kernels. The loading line plots revealed the compounds responsible for the variation corresponded to starch, water and protein.

Partial least squares discriminant analysis (PLS-DA) was performed. Two classes (pre-germinated and non pre-germinated) were modelled. The model for SA landrace 4442 (highest explained variation in Y) was used to predict the image of PAN 8816 for the 150 µm spatial resolution images. The PLS-DA model predicted the amount of pre-germinated class with 93.93% accuracy and the non pre-germinated class with 97.13% accuracy. Model development for the 30 µm spatial resolution images was not successful as the model did not perform satisfactory. Models 1 and 2 of PAN 8446 (PAN 8446-2 and PAN 8446-3) over-predicted the amount of pre-germinated pixels. Model 1 of PAN 8816-3 and model 2 of PAN 8816-2 over-predicted the amount of pre-germinated pixels. PAN 8816-2 had a 92.33% and 93.94% prediction for the non pre-germinated and pre-germinated classes, respectively. PAN 8816-3 had a 98.56% (non pre-germinated) and 81.56% (pre-germinated) correct prediction. Compared to the prediction images, the location of the predicted pixels was not completely correct

Introduction

Near infrared hyperspectral imaging is an emerging technique that captures images together with multiple NIR wavelength bands (Cogdill *et al.*, 2004; Gowen *et al.*, 2007). NIR hyperspectral images contain a NIR spectrum for each pixel in the image and have the potential of describing the distribution of constituents within a sample (Geladi *et al.*, 2004; Burger & Geladi, 2005; Burger & Geladi, 2006; Shahin & Symons, 2008). Hyperspectral imaging is a technique that is widely used in

agriculture (Martinsen & Schaare, 1998; Mahesh *et al.*, 2008; Singh *et al.*, 2009a; Williams *et al.*, 2009; Taghizadeh *et al.*, 2010), medicine (Vo-Dinh *et al.*, 2004; Martin *et al.*, 2006) and pharmaceutics (Roggo *et al.*, 2005; Roggo *et al.*, 2007; Amigo *et al.*, 2008). NIR hyperspectral imaging was proposed to detect pre-germination in wheat (Smail *et al.*, 2006; Koç *et al.*, 2008; Singh *et al.*, 2009b). NIR hyperspectral imaging is a rapid and non destructive method that could aid the sorghum grain industry in the detection of non visible pre-germination.

Sorghum is a major food crop in African and Asian countries (De Alencar Figueiredo *et al.*, 2006). It is used for a variety of products such as bread, porridge, couscous, tortillas, molasses and malt to produce beer (Asiedu *et al.*, 1993; Rooney & Serna-Saldivar, 2000; De Alencar Figueiredo *et al.*, 2006). Grain quality is thus of great importance to farmers, consumers and breeders (De Alencar Figueiredo *et al.*, 2006). Pre-germination can lead to the reduction in seed viability and promotes the hydrolysis of starch and protein in the endosperm of the grain (Lijavetzky *et al.*, 2000). This will lead to reduced quality of the grain. Pre-germination occurs when grain maturity coincides with a period of high humidity and or rain fall (Pagano *et al.*, 1997). Pre-germination is triggered by water absorption of the grain (Singh *et al.*, 2009b). During pre-germination the embryonic axis grows; this growth is accompanied by the hydrolysis of stored nutrients by means of hydrolytic enzymes (Kent & Evers, 1994). This hydrolysis of the stored nutrients decreases the grain protein and starch yield (Kent & Evers, 1994; Xing *et al.*, 2010). The hydrolysis of starch increases the soluble sugar content in the endosperm and favours the conditions for fungal attack (Pagano *et al.*, 1997). In grains destined for malting, pre-germination can lead to the loss of viability and germination energy, (Lin *et al.*, 2008) and subsequent decrease in malt quality and usability.

The detection of pre-germination in sorghum grains can be determined using visual inspection, viscosity analysis, falling number or tetrazolium test (Paliwal *et al.*, 1990; Raschke *et al.*, 1995). Visual assessment can not determine the degree of damage (starch hydrolysis) done to the kernel as some damage occur before pre-germination is visible (Matsuo *et al.*, 1982). Methods like tetrazolium test, falling number and viscosity analysis are time-consuming and tedious (De Alencar Figueiredo *et al.*, 2006). NIR hyperspectral imaging will provide a means to determine pre-germination quickly, effectively and objectively.

The aim of this research was to evaluate NIR hyperspectral imaging for objective alternative to the conventional pre-germination detection methods; and to discriminate between pre-germinated and non pre-germinated kernels using partial least squares discriminant analysis.

Materials and methods

Samples and sample preparation

Sorghum samples from commercial cultivars were kindly provided by PANNAR (Pty) Ltd. (Greytown, South Africa). Two sets of cultivar samples were used; the first included South African

landrace 4442 and PAN 8816 and was used to collect 150 µm spatial resolution images, the second included PAN 8446 and PAN 8816 and was used to collect 30 µm spatial resolution images. Six subsets, comprising 25 kernels each, were randomly selected from each of the 3 respective cultivars. The sample preparation was performed in the same manner as described in Chapter 3 (page 42).

Each cultivar was divided into seven subsets each containing 100 randomly selected kernels. The samples were prepared in the same manner as described for the 30 µm spatial resolution images of the wheat (Chapter 4, page 77).

Near infrared hyperspectral imaging system

The two SisuCHEMA imaging systems used are described in Chapter 3 (page 42).

Image acquisition

Image acquisition was performed as described in Chapter 3 (page 43). **Fig. 5.1** shows a digital photograph of the 30 µm spatial resolution images.

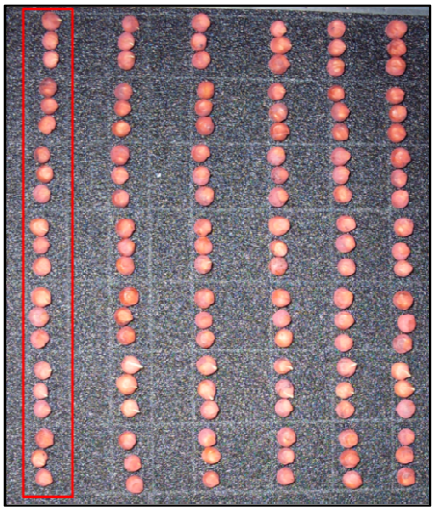


Figure 5.1 A digital photograph of the sorghum samples of the 30 µm spatial resolution, positioned on the sandpaper lined display board, one image is shown with a red rectangle.

Hyperspectral image analysis

Image cleaning and image analysis of cleaned images

The randomly imaged samples were rearranged chronologically in Evince version 2.4.0 (Umbio AB, Umeå, Sweden), as explained in Chapter 3 (page 43).

Partial least squares discriminant analysis

Partial least squares discriminant analysis (PLS-DA) was performed to determine whether it was possible to discriminate between pre-germinated and non pre-germinated wheat kernels. The PLS-

DA prediction results will be discussed in terms of false positives and false negatives. The PLS-DA models for the 150 and 30 µm spatial resolution images were created similarly as described in Chapter 3 (page 45).

Sorghum viability

The sorghum viability test was performed in the same manner as for the barley samples. The sorghum kernels were exposed to a 0.5% tetrazolium chloride solution for 1 hr. The kernels were classified according to the specifications in Chapter 3 (page 45).

Moisture content

The moisture content of the samples was determined according to the moisture - modified vacuum-oven method (AACC method 44-40, 2010).

Results and discussion

Image cleaning

Principal component analysis (PCA) with mean-centering was carried out and six components were calculated with a total of 99.80% of the total sum of squares (SS). SA landrace 4442 (150 µm spatial resolution image) was selected to demonstrate the identification and removal of unwanted pixels. The PCA score plot of SA landrace 4442 (**Fig. 5.2a**) comprised of 164496 pixels before cleaning. The score image (**Fig. 5.2b**) was used interactively to identify the clusters in the score plot. The selected clusters in the score plot were projected onto the score image, creating a classification plot (**Fig. 5.3a**) and image (**Fig. 5.3b**). The respective clusters were identified as; sandpaper, adhesive marker and sorghum kernels. The background and adhesive marker constituents were removed and PCA were recalculated. The cleaned score image and plot were used in further analysis of the dataset.

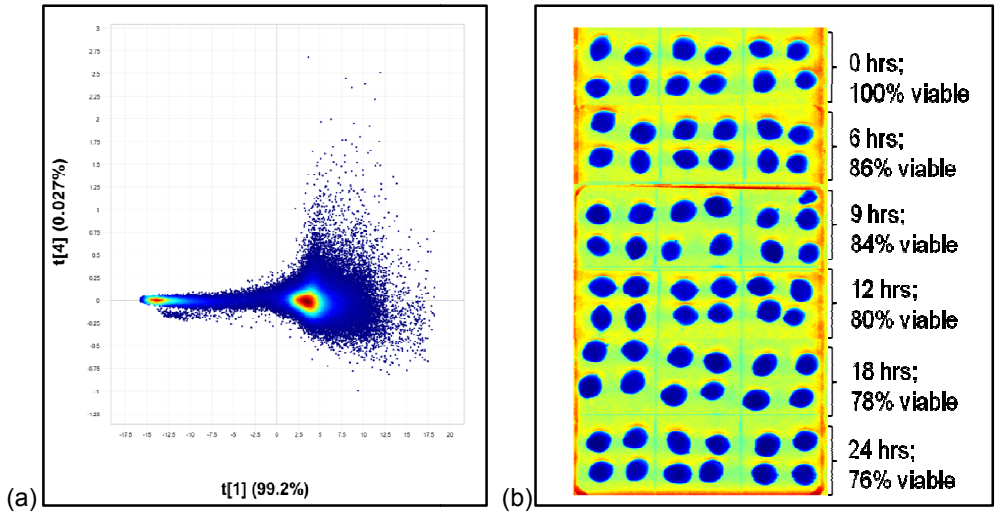


Figure 5.2 Uncleaned SA landrace 4442 (a) PCA score plot (PC 1 vs. PC 4) with 164496 pixels, displaying three clusters; and (b) uncleaned PCA score image (PC 1).

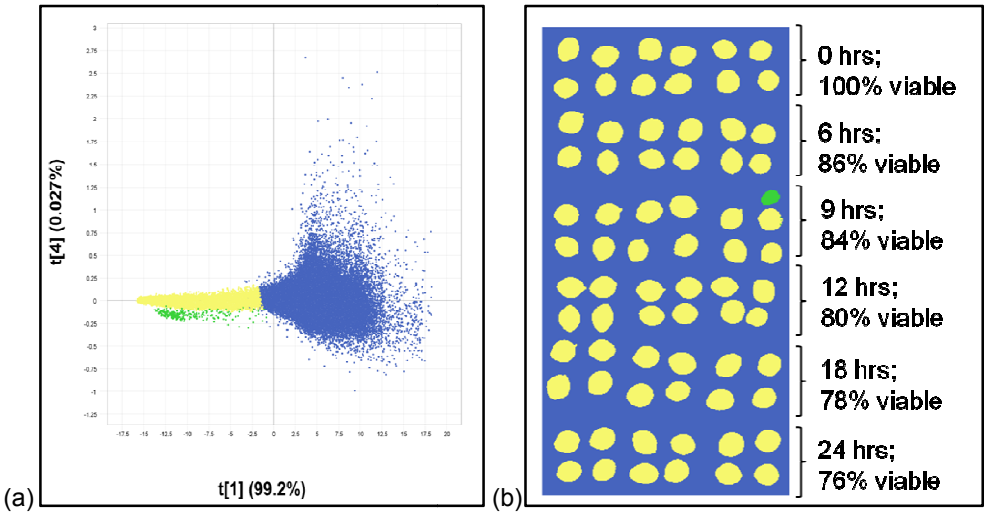


Figure 5.3 SA landrace 4442 (a) PCA score plot (PC 1 vs. PC 4) with selected areas projected onto the score image to create (b) classification image (PC 1) (blue = background, yellow = wheat kernels and green = adhesive marker).

Analysis of cleaned 150 µm spatial resolution hyperspectral images

SA landrace 4442

After the removal of unwanted clusters the PCA score images and score plots of SA landrace 4442 were analysed for regions of interest (34609 remaining pixels). Different combinations of PCs were plotted against each other (i.e. PC 1 vs. PC2; PC 1 vs. PC 3; ....; PC 5 vs. PC 6) in the PCA score plot space and inspected.

The combination of PC 1 and PC 5 (SS 1.75%) revealed no well defined clusters in the score plot (**Fig. 5.4a**), however the score image (**Fig. 5.4b**) revealed a difference between the 0 hrs and 24 hrs subsets. Pixels with similar score values are indicated with similar colours, in this case, indicating similarity in chemical composition (heat map colouring). It was anticipated that the 0 and 24 hrs subsets differ from one another, because the 0 hrs subset was not treated with water or incubated. The PCA score image of PC 5 shows a change in chemical composition from the 0 hrs subset to the 24 hrs subset. Thus, the 0 - 9 hrs subsets are similar in chemical composition and the 12 hrs to 24 hrs subsets are similar in chemical composition. When correlated with the tetrazolium test (**Table 5.1**, **Fig. 5.4b**), the difference was attributed to pre-germinated and non pre-germinated kernels; the 0 to 9 hrs subsets are non pre-germinated and the 12 to 24 hr subsets are pre-germinated.

The PCA score plot (**Fig. 5.4a**) of SA landrace 4442 with PC 1 vs. PC 5 revealed a trend in the direction of PC 5 could be observed when selecting the score values in the positive and negative regions of PC 5. The PC 1 vs. PC 5 score plot (**Fig. 3.6a**) of SA landrace 4442 revealed that positive PC 5 score values were associated with pixels of the pre-germinated kernels. The negative score values were associated with the pixels of the non pre-germinated kernels. This is better illustrated in the classification plot (**Fig. 5.5a**) and image (**Fig. 5.5b**).

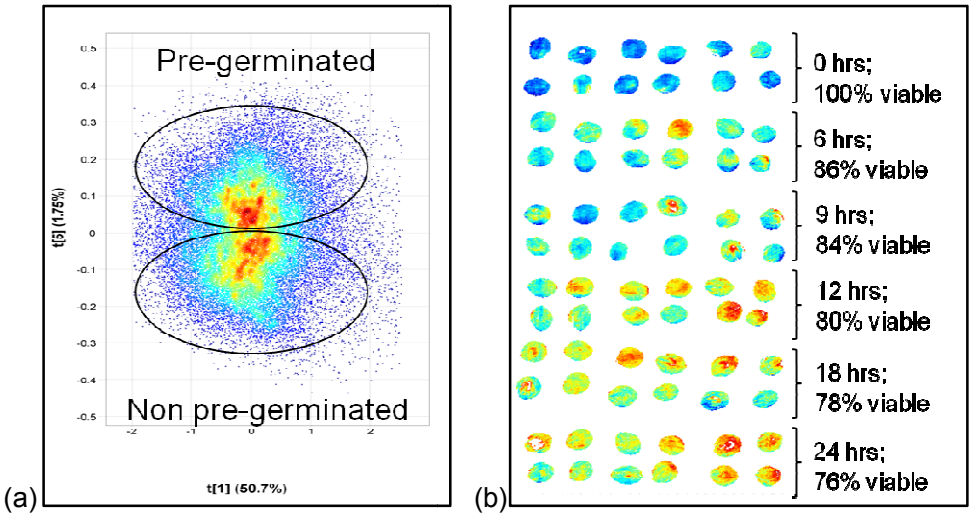


Figure 5.4 SA landrace 4442 (a) PCA score plot of PC 1 vs. PC 5 revealing two classes, i.e. non pre-germinated and pre-germinated. (b) PCA score image of PC 5 showing the different subsets; 0 – 9 hrs = non pre-germinated and 12 – 24 hrs = pre-germinated.

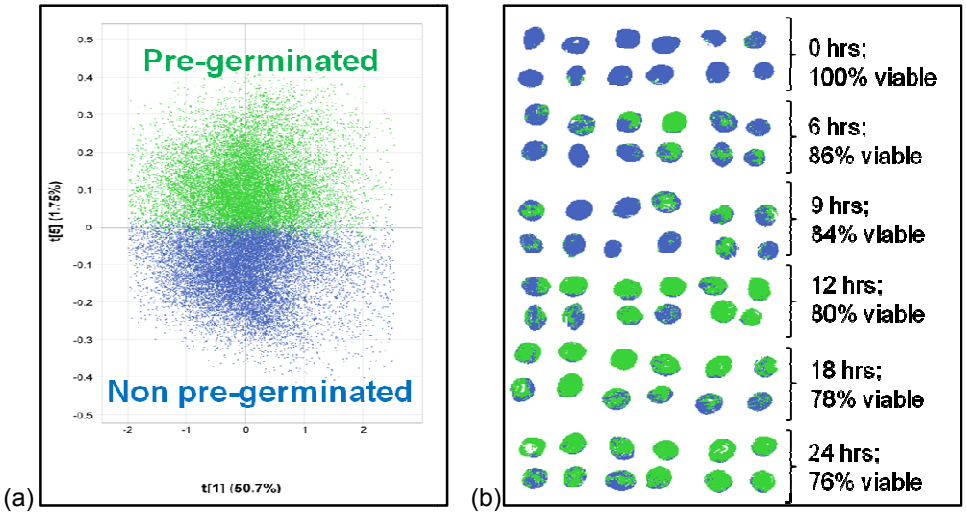


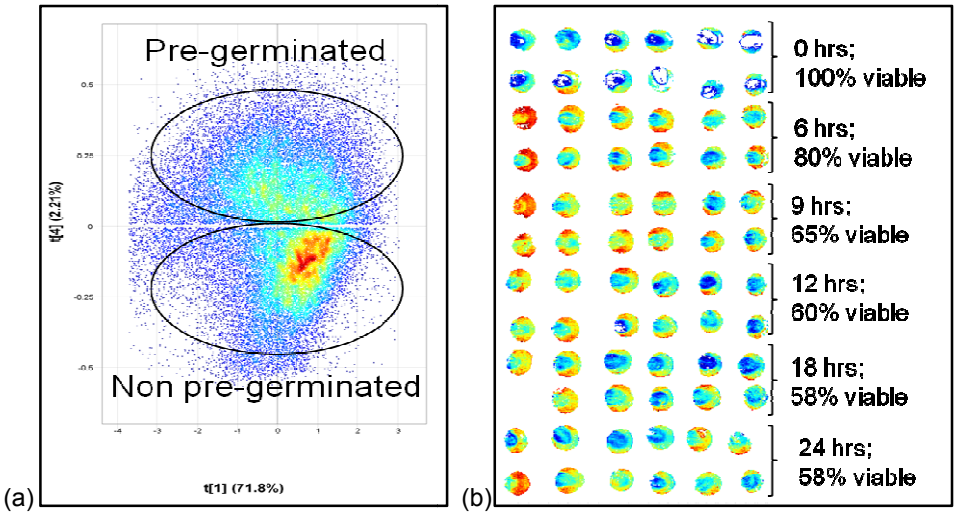
Figure 5.5 SA landrace 4442 (a) classification plot of PC 1 vs. PC 5 and (b) classification image of PC 5 (blue = non pre-germinated and green = pre-germinated).

Table 5.1 Tetrazolium test results given as percentage viability of the 25 kernels investigated

Cultivar	Subsets (time exposure to water in hrs)	Viability (%)
SA landrace 4442	0	100
SA landrace 4442	6	86
SA landrace 4442	9	84
SA landrace 4442	12	80
SA landrace 4442	18	78
SA landrace 4442	24	76
PAN 8816	0	100
PAN 8816	6	80
PAN 8816	9	65
PAN 8816	12	60
PAN 8816	18	58
PAN 8816	24	58

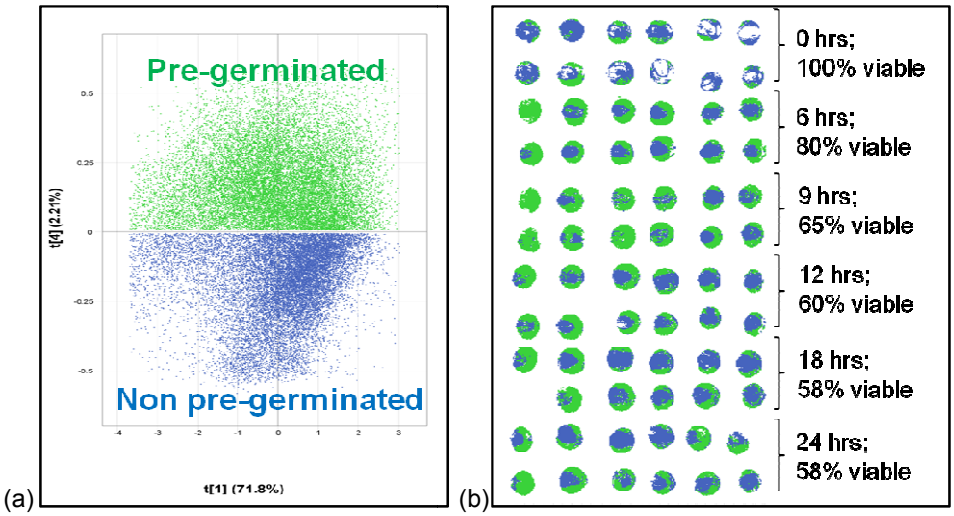
PAN 8816 – 150 µm spatial resolution image

PCA facilitated the removal of unwanted pixels such as background, shading and adhesive marker. After the removal of these pixels, 33077 pixels remained for further analysis. From previous experience and with the aid of the interactive nature of the software, the same trend was observed during the image analysis of PAN 8816. The location of the pixels in the score plot (**Fig 5.6a**) and the determined viability of the kernels provided the means to identify the pixels in the score plot. The score values on the positive side of PC 4 were identified as pre-germinated and the negative side were identified as non pre-germinated kernels. The PCA score image of PC 4 (SS 2.21%) revealed a difference between the different subsets. The 0 hrs subset and kernel from the 12, 18 and 24 hrs subsets were similar in chemical composition (non pre-germinated) and the 6 and 9 hrs subsets were pre-germinated. The tetrazolium test results (**Table 5.1**) are given as an average value for the whole subset (25 kernels). This could lead to low percentage viability for the total subset, but the image reveals non pre-germinated kernels. A classification plot (**Fig. 5.7a**) and classification image (**Fig. 5.7b**) were formed by selecting different clusters in the PCA score plot and projecting it onto the PCA score image.



(a) (b)

Figure 5.6 (a) PAN 8816 PCA score plot of PC 1 vs. PC 4 revealing two classes, i.e. non pre-germinated and pre-germinated. (b) PCA score image of PC 4 showing the viability of the different subsets.



(a) (b)

Figure 5.7 PAN 8816 (a) classification plot of PC 1 vs. PC 4 and (b) classification image of PC 4 (blue = non pre-germinated and green = pre-germinated).

Loading line plots – 150 µm spatial resolution images

The loading line plots of PAN 8816 and SA landrace 4442 were similar, thus only the loading line plot of SA landrace 4442 will be discussed. The loading line plot of PC 5 of SA landrace 4442 (**Fig. 5.8**) reveals major peaks at 1360 nm, 1940 nm and 2150 nm. The peak at 1360 nm (2 x C-H stretching and C-H deformation) indicates the presence of a CH_3 and is related to starch. This peak was positively weighted within PC 5 and can be associated with the positive score values (pre-germinated) in the score plot. During pre-germination starch is degraded (Fincher & Stone, 1993). The variation due to starch could be the result of the hydrolysis of starch during pre-germination. The major peak observed at 1940 nm is related to water (O-H stretching and O-H deformation) (Osborne *et al.*, 1993). This peak is positively weighted in PC 5 and can be

associated with the pre-germinated kernels. The difference between pre-germinated and non pre-germinated kernels was not simple a function of moisture content as the moisture content of the 0 and 24 hrs subsets were similar (**Table 5.2**). Rather, the water environment differed between pre-germinated and non pre-germinated kernels. In pre-germinated kernels starch and protein have been hydrolysed, this process requires the interaction of water with enzymes, starch or proteins (see Chapter 2, page 13). The peak at 2150 nm (2 x amide I and amide III) was associated with protein. This peak was negatively weighted within PC 5 and can be associated with the negative score values (non pre-germinated) in the score plot. Protein remains intact if kernels are not pre-germinated.

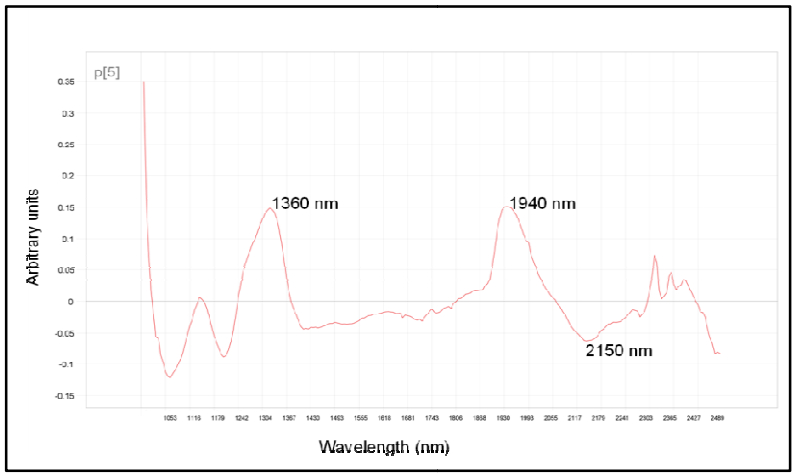


Figure 5.8 Loading line plot of SA landrace 4442 illustrating the compounds contributing to variation; starch (1360 nm), water (1940 nm) and protein (2150 nm).

Table 5.2 Moisture content results of different sorghum samples

	Moisture content of subsets (%)	
Cultivars	**0 hrs**	**24 hrs**
SA landrace 4442	7.79	8.09
PAN 8816	6.86	7.47

Partial least squares discriminant analysis – 150 µm spatial resolution images

For each of the cultivars a PLS-DA model was created; these models were evaluated and the model with the highest explained variation in Y was applied to the image of the other cultivar. The PLS-DA model for SA landrace 4442 presented the most explained variation in Y (62% after six components) and was used to predict PAN 8816 (61% explained variation in Y after six components).

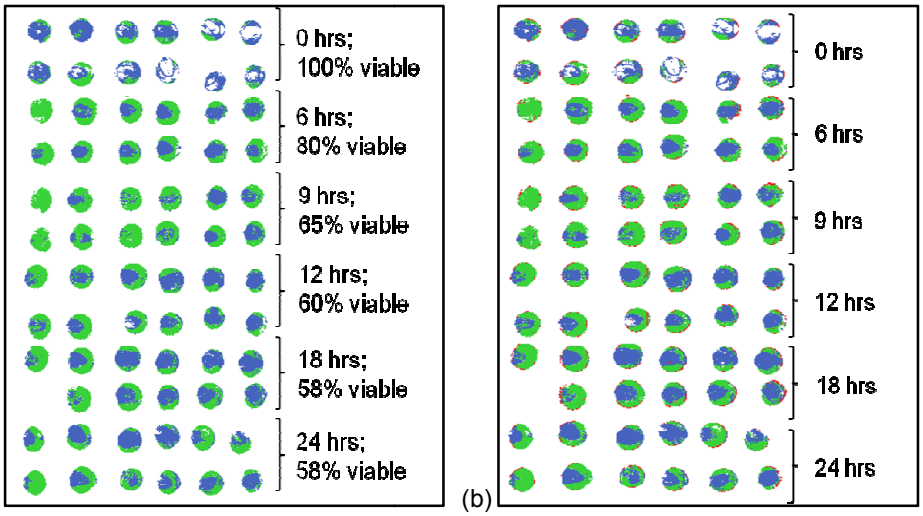
When pre-germinated kernels are predicted as non pre-germinated a false positive will arise that is detrimental for the maltster. Malting requires viable, non pre-germinated sorghum. Production losses could occur if pre-germinated sorghum is predicted as non pre-germinated and used for malting purposes. A false negative will arise if non pre-germinated kernels are not

predicted as such. If non pre-germinated kernels are predicted as pre-germinated, a false positive will occur that is detrimental to the farmer.

The model correctly predicted 97.13% of the non pre-germinated class and 93.93% of the pre-germinated class (**Table 5.3**). The classification image (**Fig. 5.9a**) of PAN 8816 illustrates the localisation of the assigned pixels. The prediction image (**Fig. 5.9b**) provides a visual interpretation of the model's performance. When the classification image is compared to the prediction image it is possible to visualise that the model predicted very good.

Table 5.3 PLS-DA prediction results indicating the classes, number of pixels assigned and predicted to each class for PAN 8816

	Assigned pixels	**Predicted**
Non pre-germinated	16367	15897 (97.13%)
Pre-germinated	16710	15696 (93.93%)
Not classified		1484
Total	33077	33077



(a) (b)

Figure 5.9 (a) PAN 8816 classification image and (b) PAN 8816 PLS-DA prediction image (blue = predicted non pre-germinated, green = predicted pre-germinated and red = predicted not classified).

Analysis of cleaned 30 µm spatial resolution hyperspectral images

Six columns containing each subset were imaged for both of the cultivars. Only one representative column from each cultivar will be discussed to minimise repetition and will be called PAN 8446-1 and PAN 8816-1. Two other randomly picked columns will be used for PLS-DA prediction and these columns will be called PAN 8446-2, PAN 8846-3, PAN 8816-2 and PAN 8816-3.

PAN 8446-1 – 30 µm spatial resolution images

The PAN 8446-1 PCA score plot (**Fig. 5.10a**) of PC 1 (SS 73.9%) vs. PC 4 (SS 2.88%) revealed two clusters. After exploration it was found that the clusters in the direction of PC 4 accounts for the difference between pre-germinated (positive score values) and non pre-germinated (negative score values) kernels.

The PAN 8446-1 PCA score image (**Fig. 5.10b**) of PC 4 revealed differences (chemical differences) between the determined pre-germinated and non pre-germinated kernels. However, kernels from the 0 and 6 hrs subsets were determined as non viable (pre-germinated), but had the same chemical composition as non pre-germinated kernels in the score image. The endosperm of these kernels may have undergone incomplete degradation by hydrolytic enzymes. The NIR hyperspectral imaging method is perhaps detecting degree of endosperm hydrolysis rather than kernel viability. A kernel with a partially hydrolysed endosperm might lead to a non viable result by the tetrazolium test and a viable result with the NIR hyperspectral imaging method.

The first classification plot (**Fig. 5.11a**) and image (**Fig. 5.11b**) was created by selecting the kernels in the score image belonging to either the pre-germinated or non pre-germinated class, as identified by the tetrazolium test. **Fig. 5.11** shows that the pixels associated with the non pre-germinated kernels are situated mostly on the negative side of PC 4. The second classification plot (**Fig. 5.12a**) and classification image (**Fig. 5.12b**) was created by selecting the pixels belonging to either the pre-germinated or non pre-germinated clusters in the score plot. The non pre-germinated kernels in the classification image are mostly associated with the non pre-germinated pixels in the classification plot, and the pre-germinated kernels in the classification image are mostly associated with the pre-germinated pixels from the classification plot (**Fig. 5.12**).

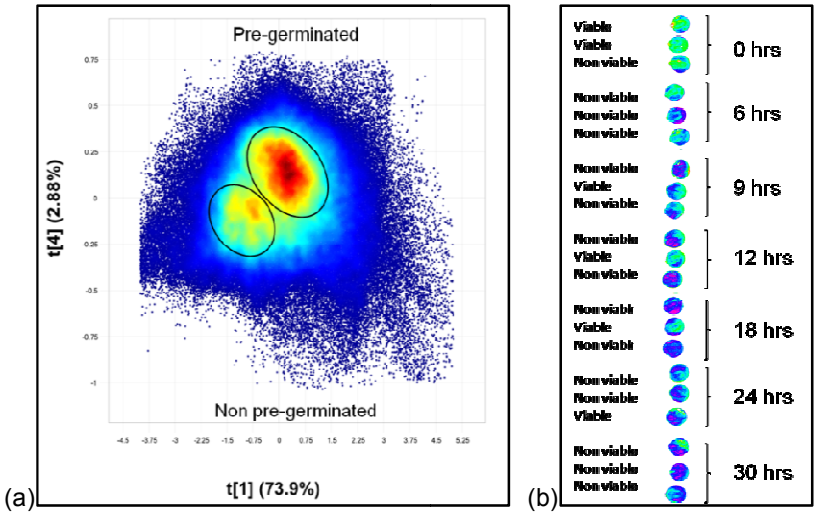


Figure 5.10 (a) PAN 8446-1 PCA score plot of PC 1 vs. PC 4 showing pre-germinated and non pre-germinated clusters; and (b) PAN 8446-1 PCA score image of PC 4 showing differences between pre-germinated and non pre-germinated kernels.

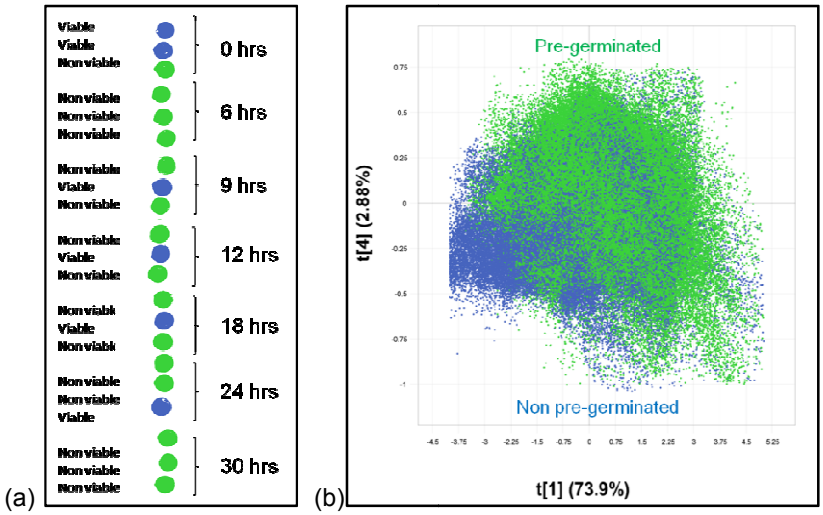


Figure 5.11 (a) PAN 8446-1 classification image showing the assigned pre-germinated (non viable) and non pre-germinated (viable) kernels with selection based on tetrazolium test results; and (b) classification plot with classes selected in the score image and projected onto the score plot (pre-germinated = green and non pre-germinated = blue).

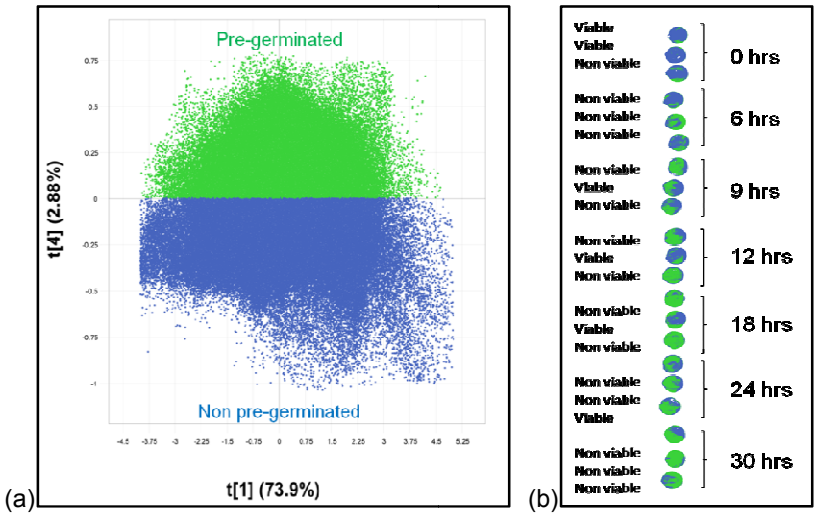


Figure 5.12 (a) PAN 8446-1 classification plot of PC 1 vs. 4 showing the assigned pre-germinated and non pre-germinated classes; and (b) classification image of score classes projected onto score image (pre-germinated = green and non pre-germinated = blue).

PAN 8816-1 – 30 µm spatial resolution image

The PCA score plot (**Fig. 5.13a**) of PC 1 (SS 72.4%) vs. PC 3 (SS 3.42%) revealed no clusters, but with exploration it was found that the scores on the positive side of PC 3 are associated with the pre-germinated kernels in the score image (**Fig. 5.13b**); and the negative scores are associated with the non pre-germinated kernels in the score image. The PCA score image of PC 3 revealed differences between the determined pre-germinated and non pre-germinated kernels. There were kernels from the 12, 18 and 24 hrs subsets that were similar to the non pre-germinated kernels in the score image, but these kernels were indicated as pre-germinated by the tetrazolium test. This

might be because of incomplete endosperm hydrolysis. Two of classification images and plots were formed in the same manner as described for PAN 8446-1 (**Figs. 5.14** and **5.15**). **Fig. 5.14** shows that the pixels associated with the non pre-germinated kernels are situated mostly on the negative side of PC 3. **Fig. 5.15** shows that the pixels on the negative side of PC 3 are mostly associated with the non pre-germinated kernels and the pixels on the positive side of PC 3 are mostly associated with the pre-germinated kernels in the classification image.

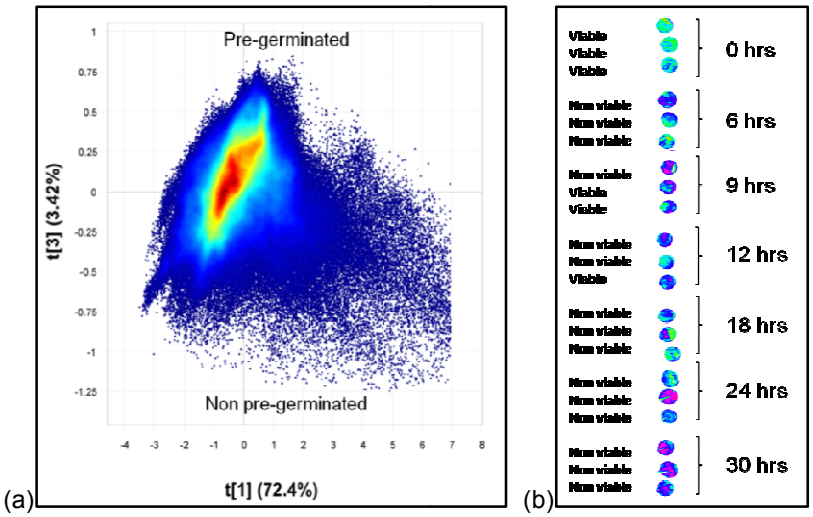


Figure 5.13 (a) PAN 8816-1 PCA score plot of PC 1 vs. PC 3 showing pre-germinated and non pre-germinated score values. (b) PAN 8816 1 PCA score image of PC 3 showing differences between pre-germinated and non pre-germinated kernels.

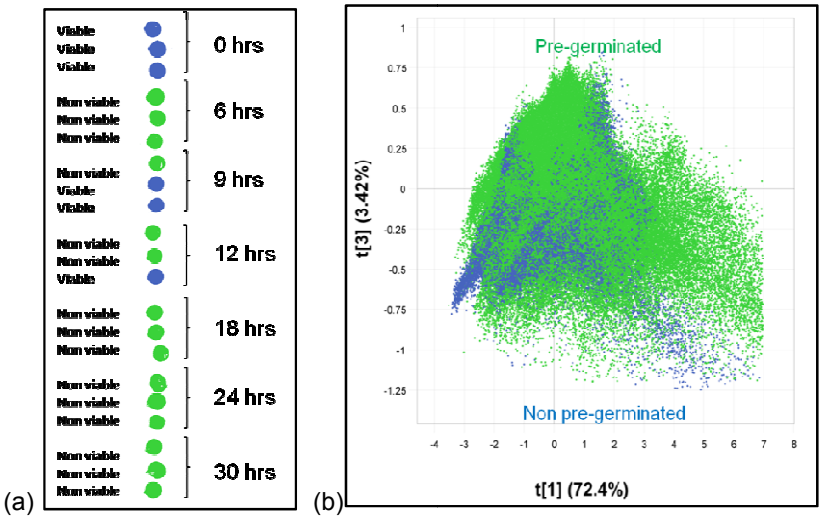
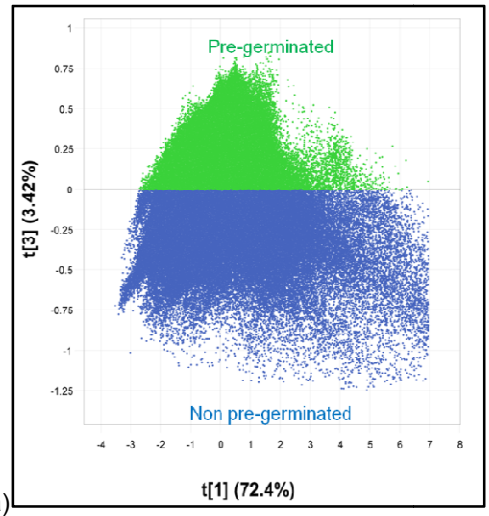


Figure 5.14 (a) PAN 8816-1 classification image showing the assigned pre-germinated (non viable) and non pre-germinated (viable) kernels with selection based on tetrazolium test results; and (b) classification plot with classes selected in the score image and projected onto the score plot (pre-germinated = green and non pre-germinated = blue).

(a) 

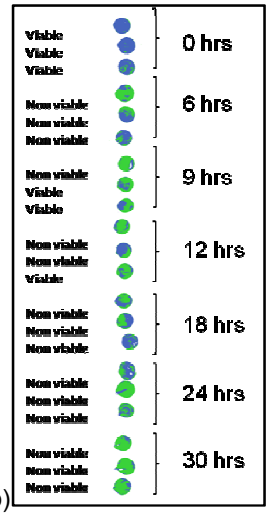
(b) 

Figure 5.15 (a) PAN 8816-1 classification plot of PC 1 vs. 2 showing the assigned pre-germinated and non pre-germinated classes; and (b) classification image of score classes projected onto score image (pre-germinated = green and non pre-germinated = blue).

Loading line plots – 30 µm spatial resolution images

The loading line plots of PAN 8446 and PAN 8816 differed slightly, but the main compounds contributing to the variation was starch, water and protein. These compounds were the main compounds contributing to the variation in the 150 µm spatial resolution images.

For PAN 8446 peaks (**Fig. 5.16**) were observed at 1395 nm (starch - 2x C-H stretching and C-H deformation), 1940 nm (water - O-H stretching and O-H deformation) and 2132 nm (protein - N-H stretching and C=O stretching) (Osborne *et al.*, 1993). The starch peak is positively weighted and can be associated with the pre-germinated score values in the score plot of PC 4. During pre-germination starch is hydrolysed to glucose and oligosaccharides. The water peak is negatively weighted and is associated with the non pre-germinated scores in the score plot. This peak was associated with the pre-germinated kernels in the 150 µm spatial resolution images. It is unclear why the water peak changed, but a possible reason could be that the kernels from the 150 µm spatial resolution images was not pre-germinated to the same extent as the kernels from the 30 µm spatial resolution images. The moisture content results (**Table 5.4**) indicated that the differences between kernels were not due to moisture content differences. The protein peak was negatively weighted and can be associated with the negative (pre-germinated) scores in the score plot. Non pre-germinated kernels would contain more intact protein molecules than pre-germinated kernels.

PAN 8816 had peaks (**Fig. 5.17**) at 1360 nm (starch - 2 x C-H stretching and C-H deformation), 1395 nm (starch - 2x C-H stretching and C-H deformation) and 1940 nm (water - O-H stretching and O-H deformation) (Osborne *et al.*, 1993). The starch peaks are positively weighted and can be associated with the pre-germinated score values in the score plot of PC 3. During pre-germination starch is hydrolysed. The water peak is negatively weighted and can be associated

with the non pre-germinated score values of PC 3. This is also different than the 150 µm spatial resolution images

The water peak at 1940 nm was the dominating compound contributing to spectral variation. None of the published studies of NIR hyperspectral imaging to detect pre-germination in wheat detected this water peak (Smail *et al.*, 2006; Singh *et al.*, 2009b). There research did not include an extended wavelength range (higher than 1700 nm)

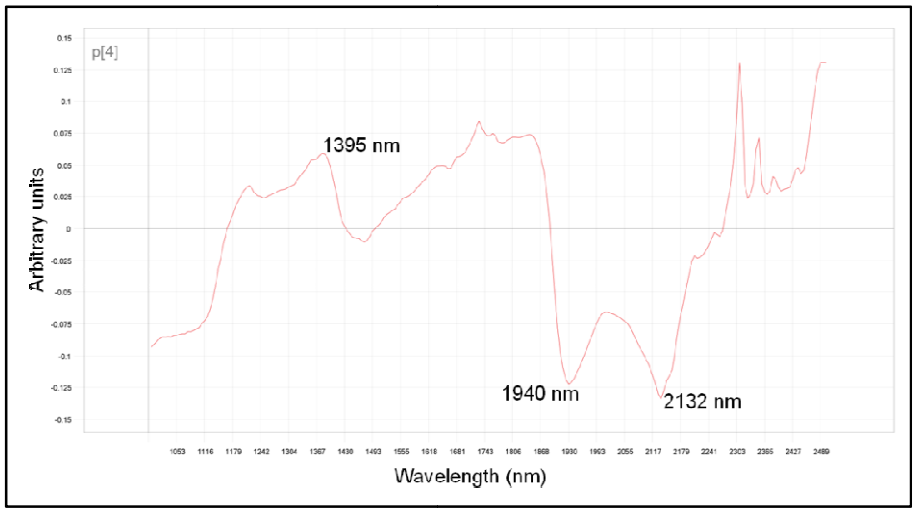


Figure 5.16 Loading line plot of PAN 8446 PC 4, illustrating the compounds responsible for variation; starch (1395 nm), water (1940 nm) and protein (2132 nm).

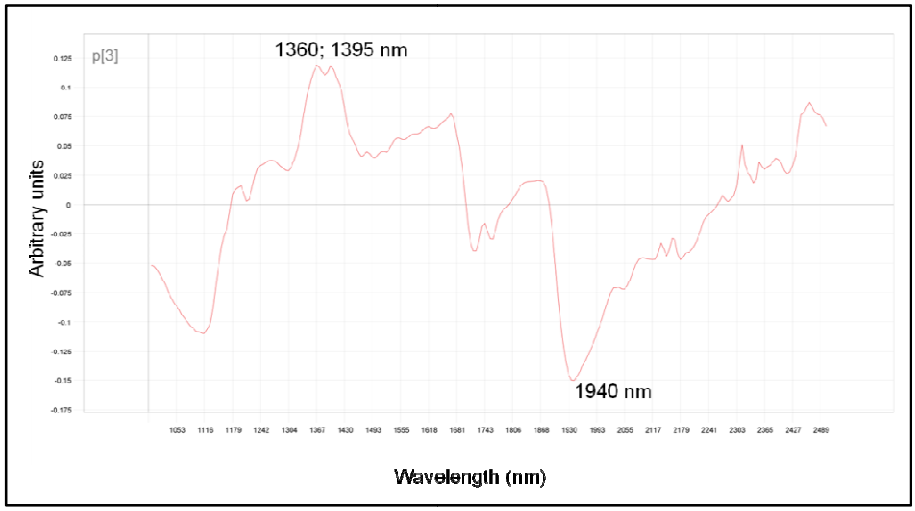


Figure 5.17 Loading line plot of PAN 8816 PC 3, illustrating the compounds responsible for variation; starch (1360 nm and 1395 nm) and water (1940 nm).

Table 5.4 Moisture analysis results for each subset of the three cultivars

Cultivars	Subsets (time exposure to distilled water in hrs)	Moisture content (%)
PAN 8446	0	9.89
	6	9.40
	9	9.79
	12	9.39
	18	9.79
	24	9.71
	30	9.46
PAN 8816	0	9.99
	3	9.91
	6	9.72
	9	9.72
	12	9.59
	18	9.43
	24	9.47

Partial least squares discriminant analysis – 30 µm spatial resolution

PAN 8446-2 and PAN 8446-3

Two PLS-DA models were created for PAN 8446-1 and both of these models were employed to predict two other images of PAN 8446 (2 and 3). Model one (M1) was created by classifying the pre-germinated and non pre-germinated kernels in the score image. Model two (M2) was created by classifying the pre-germinated and non pre-germinated scores in the score plot. The explained variation in Y for M1 was 24.47% and M2 was 66.42% after six components. **Table 5.5** contains the amount of pixels assigned to each of the classes in both models one and two of PAN 8446-2 and PAN 8446-3.

The prediction data for PAN 8446-2 (**Table 5.5**) revealed that the models did not perform very well. M1 and M2 only predicted 3.10% and 39.36% of the non pre-germinated class correctly, respectively. This resulted in false negatives, 96.9% and 60.64% for M1 and M2, respectively. **Fig. 5.18** shows that the model predicted more pixels as pre-germinated than was assigned for PAN 8446-2.

The prediction results for PAN 8446-3 revealed that M1 and M2 predicted fewer pixels (83.45% and 27.75%, respectively) to the non pre-germinated class than was assigned (false negatives) (**Table 5.5**). **Fig. 5.19** shows the prediction images and classification images of PAN 8446-3 for models 1 and 2. These images also reveal that the models predicted more pixels as pre-germinated than was assigned.

Singh et al. (2009) obtained better classification results. The developed models were based on sound kernels (not treated with water) and fully pre-germinated kernels (incubated for 48hrs and sprouting occurred). Models developed from such kernels will exhibit better classification results. In this study kernels were pre-germinated for a shorter amount of time (6-30 hrs). Varying degrees of endosperm hydrolysis was evident in the images and may have lead to the model's insufficient prediction. Models developed from only pre-germinated or non pre-germinated kernels may improve the prediction results.

The prediction results could be improved by developing the models from a bigger sample set. Even though the models were developed from thousands of pixels the reference method used was only performed on 21 kernels.

Table 5.5 PLS-DA prediction results for PAN 8446-2 and PAN 8446-3 of both models M1 and M2

		Assigned number pixels	**Predicted number pixels (% pixels)**
PAN 8446-2	**M1**		
	Non pre-germinated	131068	4066 (3.10)
	Pre-germinated	170634	286734 (168.04)
	Not classified		10902
	Total	301702	301702
	M2		
	Non pre-germinated	157094	61845 (39.36)
	Pre-germinated	144608	228278 (157.86)
	Not classified		11579
	Total	301702	301702
PAN 8446-3	**M1**		
	Non pre-germinated	116771	19324 (16.55)
	Pre-germinated	186997	255738 (136.76)
	Not classified		28706
	Total	303768	303768
	M2		
	Non pre-germinated	142881	103227 (72.25)
	Pre-germinated	160887	175320 (109.02)
	Not classified		25221
	Total	303768	303768

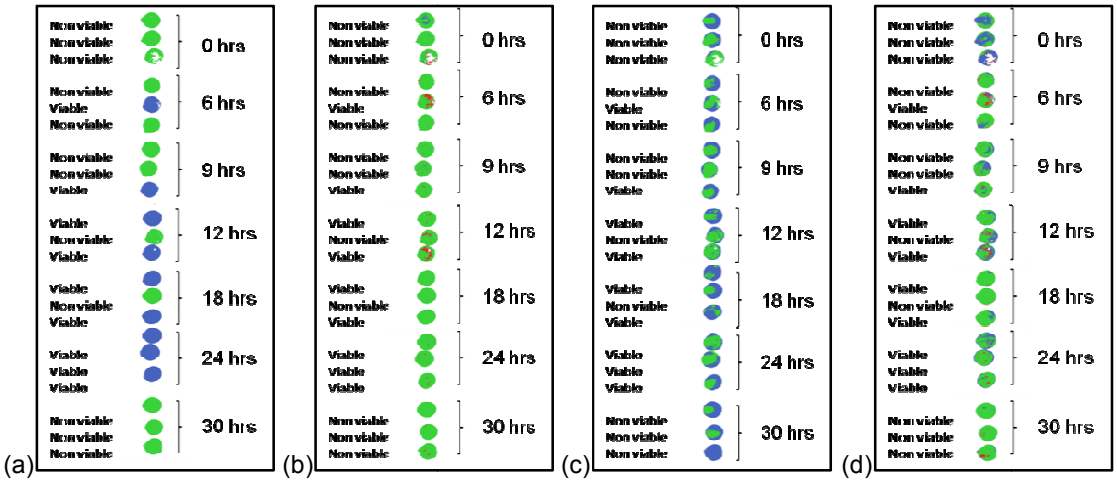


Figure 5.18 (a) PAN 8446-2 classification image with class assignments based on score image; and (b) prediction image of PAN 8446-2 with M1 (model developed from the score image of PAN 8446-1). (c) PAN 8446-2 classification image with class assignment based on the score plot; and (d) prediction image of PAN 8446-2 with M2 (model developed from the score plot of PAN 8446-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

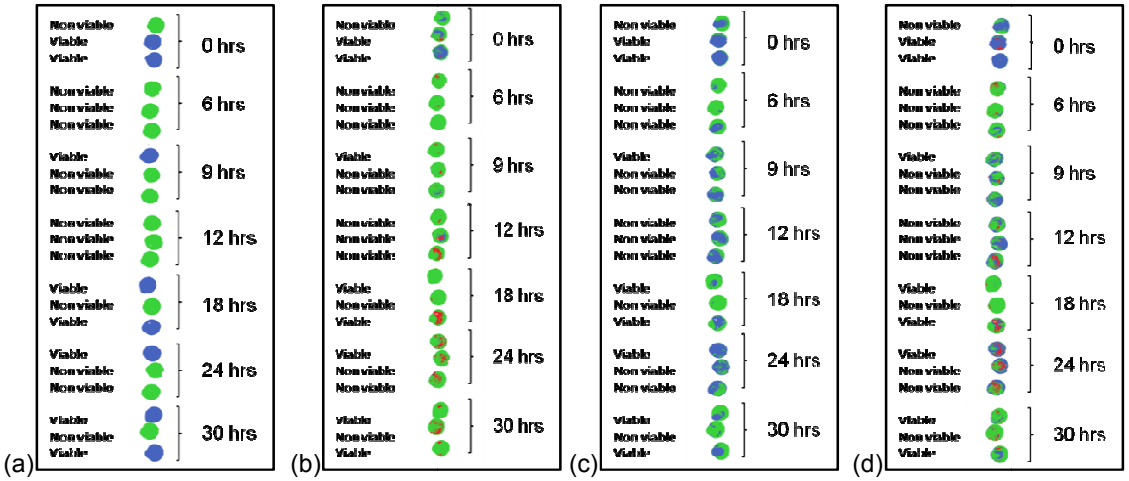


Figure 5.19 (a) PAN 8446-3 classification image with class assignments based on score image; and (b) prediction image of PAN 8446-3 with M1 (model developed from the score image of PAN 8446-1). (c) PAN 8446-3 classification image with class assignment based on the score plot; and (d) prediction image of PAN 8446-3 with M2 (model developed from the score plot of PAN 8446-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

PAN 8816-2 and PAN 8816-3

The explained variation in Y for M1 of PAN 8816-1 was 26.05% and M2 was 65.08% after six components. The prediction results revealed that the PLS-DA models performed fairly well (**Table 5.6**). M1 predicted correctly predicted 92.33% of the non pre-germinated pixels and 93.94% of the pre-germinated pixels of PAN 8816-2 (**Fig. 5.20**). The same model only predicted 28.08% non pre-

germinated pixels correctly for PAN 8816-3, thus, 71.92% false negatives (**Fig. 5.21**). M2 predicted 30.21% (false positives) more pixels as non pre-germinated than was assigned for PAN 8816-2 (**Fig. 5.20**) and correctly predicted 98.56% pixels as non pre-germinated for PAN 8816-3 (**Fig. 5.21**).

Table 5.6 PLS-DA prediction results for PAN 8816-2 and PAN 8816-3 for both models M1 and M2

		Assigned number of pixels	**Predicted number of pixels (% pixels)**
PAN 8816-2	**M1**		
	Non pre-germinated	142825	131866 (92.33)
	Pre-germinated	191816	180198 (93.94)
	Not classified		22577
	Total	334641	334641
	M2		
	Non pre-germinated	153020	199249 (130.21)
	Pre-germinated	181621	102808 (53.60)
	Not classified		32584
	Total	334641	334641
PAN 8816-3	**M1**		
	Non pre-germinated	141914	39843 (28.08)
	Pre-germinated	185798	260834 (140.39)
	Not classified		27035
	Total	327712	327712
	M2		
	Non pre-germinated	144450	142366 (98.56)
	Pre-germinated	183262	149460 (81.56)
	Not classified		35886
	Total	327712	327712

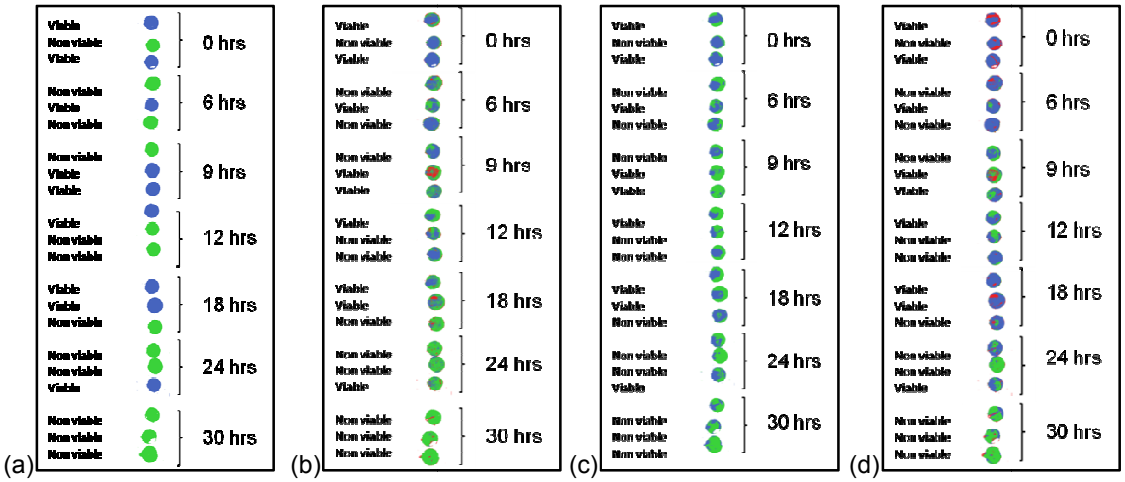


Figure 5.20 (a) PAN 8816-2 classification image with class assignments based on score image; and (b) prediction image of PAN 8816-2 with M1 (model developed from the score image of PAN 8816-1). (c) PAN 8816-2 classification image with class assignment based on the score plot; and (d) prediction image of PAN 8816-2 with M2 (model developed from the score plot of PAN 8816-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

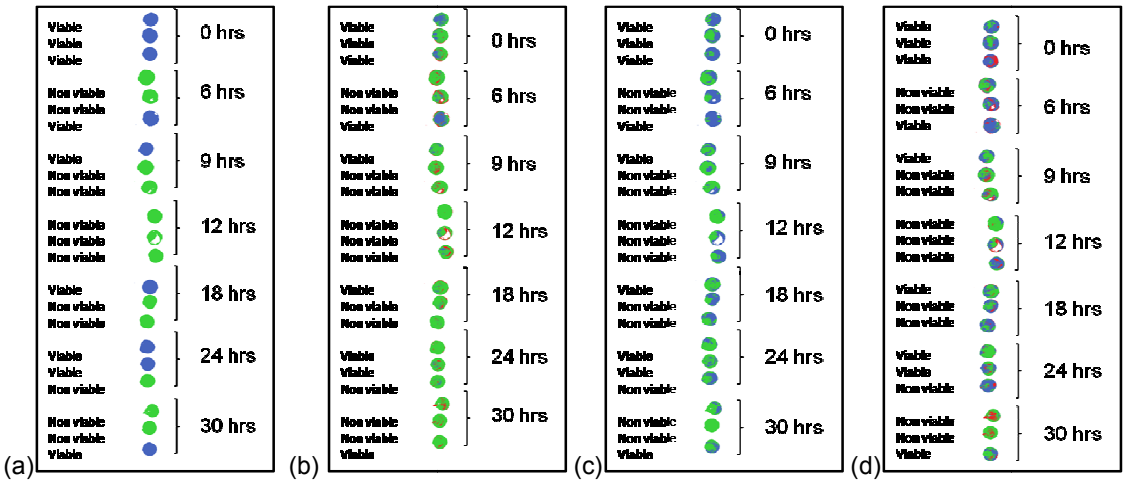


Figure 5.21 (a) PAN 8816-3 classification image with class assignments based on score image; and (b) prediction image of PAN 8816-3with M1 (model developed from the score image of PAN 8816-1). (c) PAN 8816-3 classification image with class assignment based on the score plot; and (d) prediction image of PAN 8816-3 with M2 (model developed from the score plot of PAN 8816-1) (green = assigned/predicted pre-germinated, blue = assigned/predicted non pre-germinated and red = not classified).

Conclusion

The score images of the 150 µm spatial resolution images revealed a difference between pre-germinated and non pre-germinated subsets. This led to the classification of these kernels using NIR hyperspectral imaging. The score images of the 30 µm spatial resolution images revealed some differences. Kernels that was determined to be pre-germinated by the tetrazolium test, was similar in chemical composition to the other non pre-germinated kernels in the score image. NIR

hyperspectral imaging could be detecting differences between the kernels due to chemical differences (e.g. starch and protein hydrolysis), and this method might be employed as an indirect method to determine kernel viability.

Although the loading line plots of the 150 and 30 µm spatial resolution images differed, the loading line plots revealed that water (1940 nm) was a major compound contributing to the variation in the images. This was not mentioned in any of the other studies done on the detection of pre-germination with NIR hyperspectral imaging, since their studies only included a wavelength range from 900 – 1700 nm. Other molecules such as starch and protein also had an influence on the spectral variation. The difference between the loading line plots of the 30 and 150 µm spatial resolution images could be due to the extent to which pre-germination has occurred, the kernels from the 150 µm spatial resolution images may have pre-germinated more extensively.

The PLS-DA models of the 150 µm spatial resolution images were able to discriminate between pre-germinated (93.93% correct prediction) and non pre-germinated (97.13% correct prediction) classes. The PLS-DA models of the 30 µm spatial resolution images were not able to effectively distinguish between pre-germinated and non pre-germinated classes. Model 1 could, however, correctly predict the pre-germinated and non pre-germinated classes of PAN 8816-2 and M2 the classes of PAN 8816-3 only. Model development based on more samples and samples that are either pre-germinated or not, might improve model prediction. The PLS-DA results of the 150 µm spatial resolution images illustrate promise in the prediction of pre-germinated and non pre-germinated sorghum kernels. NIR hyperspectral imaging results indicated the possibility of this method to evaluate the degree of endosperm hydrolysis during pre-germination processes.

References

AACC. (2010). *Approved Methods of the AACC*. Moisture – modified vacuum-oven method. St. Paul, Minnesota: American Association of Cereal Chemists.

Amigo, J.M., Cruz, J., Bautista, M., Maspoch, S., Coello, J. & Blanco, M. (2008). Study of pharmaceutical samples by NIR chemical image and multivariate analysis. *Trends in Analytical Chemistry*, **27**, 696-713.

Asiedu, M., Nilsen, R., Lie, O. & Lied, E. (1993). Effect of processing (sprouting and/ or fermentation) on sorghum and maize. I: proximate composition, minerals and fatty acids. *Food Chemistry*, **46**, 351-353.

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.

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.

De Alencar Figueiredo, L.F., Davrieux, F., Fliedel, G., Rami, J.F., Chantereau, J., Deu, M., Courtois, B. & Mestres, C. (2006). Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. *Journal of Agricultural and Food Chemistry*, **54**, 8501-8509.

Fincher, G.B. & Stone, B.A. (1993). Physiology and biochemistry of germination in barley. In: *Barley: Chemistry and Technology* (edited by A.W. MacGregor & R.S. Bhatty). Pp. 247-294. St. Paul: American Association of Cereal Chemists, Inc.

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

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 and Technology*, **18**, 590-598.

Kent, N.L. & Evers, A.D. (1994). Botanical Aspects of Cereals. In: *Technology of Cereals*. Pp. 29-36. New York: Elsevier Science Ltd.

Koç, H., Smail, V.W. & Wetzel, D.L. (2008). Reliability of InGaAs focal plane array imaging of wheat germination at early stages. *Journal of Cereal Science*, **48**, 394-400.

Lijavetzky, D., Martinez, M.C., Carrari, F. & Hopp, H.E. (2000). QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. *Euphytica*, **112**, 125-135.

Lin, R., Horsley, D. & Schwarz, P.B. (2008). Associations between caryopsis dormancy, alpha-amylase activity and pre-harvest sprouting in barley. *Cereal Science*, **48**, 446-456.

Mahesh, S., Manickavasagan, A., Jayas, D.S., Paliwal, J. & White, N.D.G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, **101**, 50-57.

Martin, M.E., Wabuyele, M.B., Chen, K., Kasili, P., Panjehpour, M., Phan, M., Overholt, B., Cunningham, G., Wilson, D., DeNovo, R.C. & Vo-Dinh, T. (2006). Development of an advanced hyperspectral imaging (HSI) system with applications for cancer detection. *Annals of Biomedical Engineering*, **34**, 1061-1068.

Martinsen, P. & Schaare, P. (1998). Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology*, **14**, 271-281.

Matsuo, R.R., Dexter, J.E. & MacGregor, A.W. (1982). Effect of sprout damage on durum wheat and spaghetti quality. *Cereal Chemistry*, **59**, 468-472.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). In: *Practical NIR spectroscopy with applications in food and beverage analysis* (edited by D. Browning). Pp. 1-35, 49-77. Essex: Longman Scientific & Technical.

Pagano, E.A., Benech-Arnold, R.L., Wawrzkiewicz, M. & Steinbach, H.S. (1997). Alpha-amylase activity in developing sorghum caryopses from sprouting resistant and susceptible varieties. The role of ABA and GAs on its regulation. *Annals of Botany*, **79**, 13-17.

Paliwal, S., Baskin, C.C. & Delouche, J.C. (1990). The relationship between the tetrazolium test, soil cold test and standard germination test in assessing sorghum (*Sorghum bicolor* (L.) Moench) seed quality. *Journal of Seed Technology*, **14**, 56-60.

Raschke, A.M., Taylor, J. & Taylor, J.R.N. (1995). Use of falling number and rapid visco analyser instruments to estimate sorghum malt diastatic power. *Journal of Cereal Science*, **21**, 97-102.

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.

Roggo, Y., Jent, N., Edmond, A., Chalus, P. & Ulmschneider, M. (2005). Characterizing process effects on pharmaceutical solid forms using near-infrared spectroscopy and infrared imaging. *European Journal of Pharmaceutics and Biopharmaceutics*, **61**, 100-110.

Rooney, L.W. & Serna-Saldivar, S.O. (2000). Sorghum. In: *Handbook of Cereal Science and Technology* (edited by K. Kulp & J.G. Ponte). Pp. 149-171. New York: Marcel Dekker, Inc.

Shahin, M.A. & Symons, S.J. (2008). Detection of hard vitreous and starchy kernels in amber durum wheat samples using hyperspectral imaging. *NIR News*, **19**, 16-18.

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

Singh, C.B., Jayas, D.S., Paliwal, J. & White, N.D.G. (2009b). Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry*, **86**, 256-260.

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.

Taghizadeh, M., Gowen, A., Ward, P. & O'Donnell, C.P. (2010). Use of hyperspectral imaging for the evaluation of the shelf-life of fresh white button mushrooms (*Agaricus bisporus*) stored in different packaging films. *Innovative Food Science and Emerging Technologies*, **11**, 423-431.

Vo-Dinh, T., Stokes, D.L., Wabuyele, M.B., Martin, M.E., Song, J.M., Jagannathan, R., Michaud, E., Lee, R.J. & Pan, X. (2004). A hyperspectral imaging system for in vivo optical diagnostics. *IEEE Engineering in medicine and biology magazine*, **23**, 40-49.

Williams, P., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta*, **653**, 121-130.

Xing, J., Symons, S., Shahin, M. & Hatcher, D. (2010). Detection of sprout damage in Canada western red spring wheat with multiple wavebands using visible/ near infrared hyperspectral imaging. *Biosystems Engineering*, **106**, 188-194.

Chapter 6

Effect of kernel shape on near infrared (NIR) hyperspectral imaging and image analysis

Chapter 6

Effect of kernel shape on near infrared (NIR) hyperspectral imaging and image analysis

Abstract

The identification of intact pre-germinated barley, wheat and sorghum kernels are advantageous for the grain utilisation industry. Previous studies (Chapters 3, 4 and 5) revealed that differences between pre-germinated and non pre-germinated kernels used for the 150 µm spatial resolution images were only observed in principal component (PC) 5 (barley and sorghum) and PC 6 (wheat). In this study the subsequent lower order PCs (PC 1 – 4) were evaluated to understand the major sources of spectral variation. Images acquired in the previous studies were inspected with classification gradients. Classification gradients were made in the score plot in the direction of the respective PCs from negative to positive scores. A different colour was given to each gradient in the score plot to simplify the localisation of those scores in the score image. Classification gradients facilitated the identification of certain shape attributes in the classification gradient images. The shape of the grain kernels were observed in PCs one to three for the barley and sorghum kernels, and PCs one to four for the wheat kernels. The shape of the grain kernels influenced hyperspectral image analyses. It is therefore important to consider the physical characteristics of a sample before hyperspectral imaging and image analysis.

Introduction

Near infrared (NIR) hyperspectral imaging is a technique used for characterising and analysing biological and organic samples (Geladi *et al.*, 2004; Gowen *et al.*, 2008). It combines conventional imaging with NIR spectroscopy, resulting in an image with a related NIR spectrum for each pixel in the image.

NIR hyperspectral imaging has several advantages. No sample preparation is required and the technique does not require the use of chemicals and potentially dangerous reagents (ElMasry & Sun, 2010). High spectral resolution of hyperspectral images provides qualitative and quantitative measurements. The imaged samples are not destroyed during imaging and can be used for other purposes or analyses. It is possible to examine several constituents simultaneously in the same sample. The biggest advantage of hyperspectral imaging is the ability to identify the location of chemical and physical constituents within a sample (Burger & Geladi, 2005; Gowen *et al.*, 2008). This is possible because regions of a sample with similar chemical compositions have similar spectral properties (Gaston *et al.*, 2010).

The combination of imaging and NIR spectroscopy makes NIR hyperspectral imaging suitable for numerous applications, however, the technique has some disadvantages which must be considered before implementation in the food and agricultural industry (ElMasry & Sun, 2010). NIR hyperspectral images create data exploration problems (large blocks of data) (Burger & Geladi, 2005; Burger & Geladi, 2006; ElMasry & Sun, 2010). Hyperspectral imaging is not suitable for liquids or homogeneous samples because the value of imaging lies in the ability to resolve the spatial distribution of constituents in a sample (ElMasry & Sun, 2010). NIR hyperspectral imaging is subjected to the combined error contributions of conventional spectroscopy (e.g. light scattering) and imaging (e.g. sample shape contributions) (Burger & Geladi, 2005; Burger & Geladi, 2007; Gowen *et al.*, 2008). When whole samples are imaged, as in the case of hyperspectral imaging, the shape of the sample could have a great effect on the spectra.

The effect of sample shape on spectra of hyperspectral images was observed by Gowen *et al*. (2008). Hyperspectral imaging was combined with principal component analysis (PCA) for the detection of bruises on mushrooms. The shape of the mushroom surface caused a scaling difference in the spectra. The spectra from the top of the mushroom exhibited higher reflectance intensity than the spectra from the side of the mushroom. This was caused, partially, by the difference in path length from different points of the curved mushroom to the detector. Points on the mushroom, which were nearer to the detector, resulted in higher AD (analogue to digital) counts than points further away, such as those on the edge. Pre-processing techniques were employed to minimise the spectral variability caused by sample shape.

The overall aim of this thesis was to demonstrate the potential of NIR hyperspectral imaging as an alternative to conventional pre-germination detection method for South African wheat, barley and sorghum cultivars (Chapters 3, 4 and 5). It was observed during exploration of the 150 µm spatial resolution images that the desired information (difference between pre-germinated and non pre-germinated) is only explained in principal component (PC) 5 (barley and sorghum) and higher (wheat). Thus, the aim of this research was to examine the lower order principal components (PC 1 to PC 4 or 5) to understand the major sources of spectral variation, given that the lower order PCs explains the most variation in the dataset.

Materials and methods

The 150 µm spatial resolution images acquired (after PCA and cleaning) in previous studies (Chapter 3, 4 and 5) for Puma (barley), Duzi (wheat) and South African landrace 4442 (sorghum) were inspected with classification gradients. Classification gradients were selected in the score plot of each PC and projected onto the score image. Classification gradients are groups of score values selected in almost equal sections in the score plot. Each classification gradient in the classification gradient plot is given a different colour which eases the identification of those pixels in the accompanying classification gradient image. These gradients were selected in the direction of the

PC from negative to positive score values. The interactive nature of the software allows the visualisation of the selected gradients. Identification of the source of variance explained by the PC was based on prior knowledge of the cereals (morphology) and subset differences.

Results and discussion

Image analysis of cleaned barley samples

The Puma score plots and images acquired in a previous study (Chapter 3, page 47) were inspected to assess the type of spectral variation each explained in the lower order PCs. Principal component (PC) 5 of Puma was selected to explain the creation of classification gradient plots and images and their functionality in image analysis (**Fig. 6.1**). The classification gradients formed in the plot were observed in the classification gradient image. In the previous study (Chapter 3, page 48) a classification plot was formed with two classes, in this study a classification gradient plot was formed with four gradients. The four gradients provide a different means to identify the variation in the direction of PC 5. Pixels on the positive side of PC 5 (green and turquoise) were associated with the non pre-germinated subsets (Chapter 3, page 47) and the pixels on the negative side of PC 5 (red and blue) were associated with the pre-germinated subsets. Thus, this classification plot (**Fig. 6.1**) with gradients was also able to identify the variation (pre-germinated and non pre-germinated kernels) in the direction of PC 5.

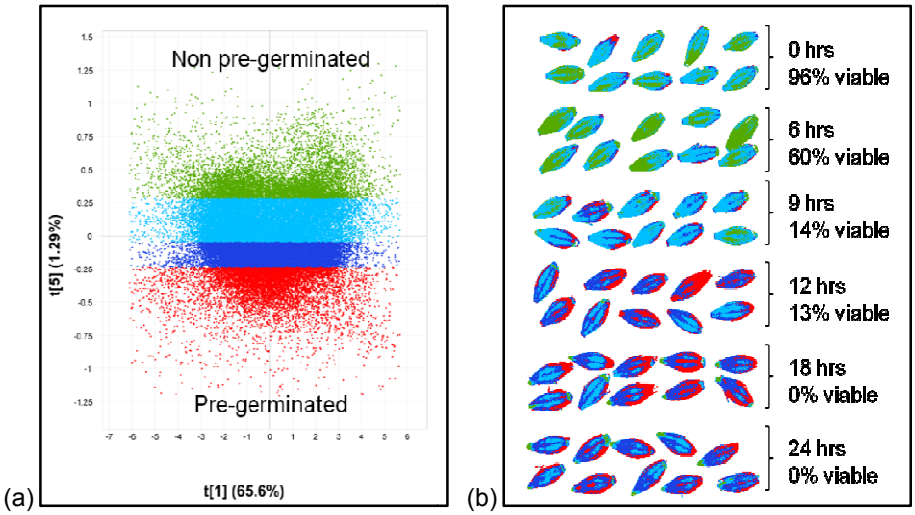


Figure 6.1 (a) Puma classification plot showing gradients from negative (red) to positive (green) score values of PC 5; and (b) classification gradient image showing pre-germinated subsets in red and blue and non pre-germinated subsets in green and turquoise.

During the image analysis of Puma (Chapter 3, page 47) the PCA score image of PC 1 (**Fig. 6.2**) did not show differences between known pre-germinated and non pre-germinated subsets. Therefore, PC 1 did not account for the variation between pre-germinated and non pre-germinated

kernels. A classification plot with gradients was created to investigate the variation in the direction of PC 1 (**Fig. 6.3a**). The pixels in the classification gradient plot with negative PC 1 score values were associated with the top parts of the barley kernels in the classification gradient image (**Fig. 6.3b**) and the pixels with negative PC 1 score values (**Fig. 6.3a**) were associated with the lowest parts of the kernels (**Fig. 6.3b**). Consequently, PC 1 accounted for the shape of the kernels from the highest (red) to the lowest (orange) parts.

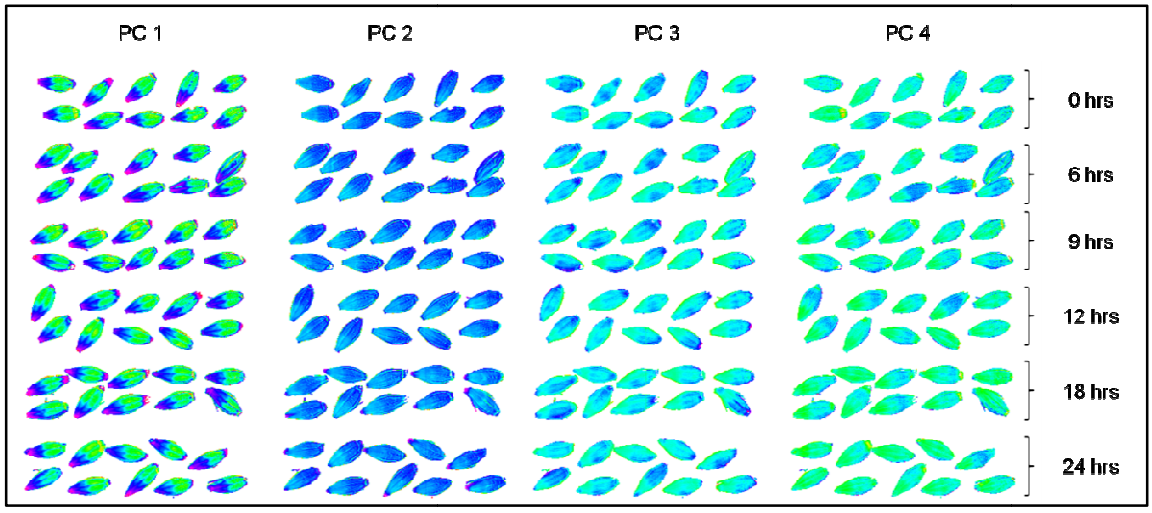


Figure 6.2 Puma score images of PCs 1 – 4 showing no distinct differences between the 0 (non pre-germinated) and 24 hrs (pre-germinated) subsets.

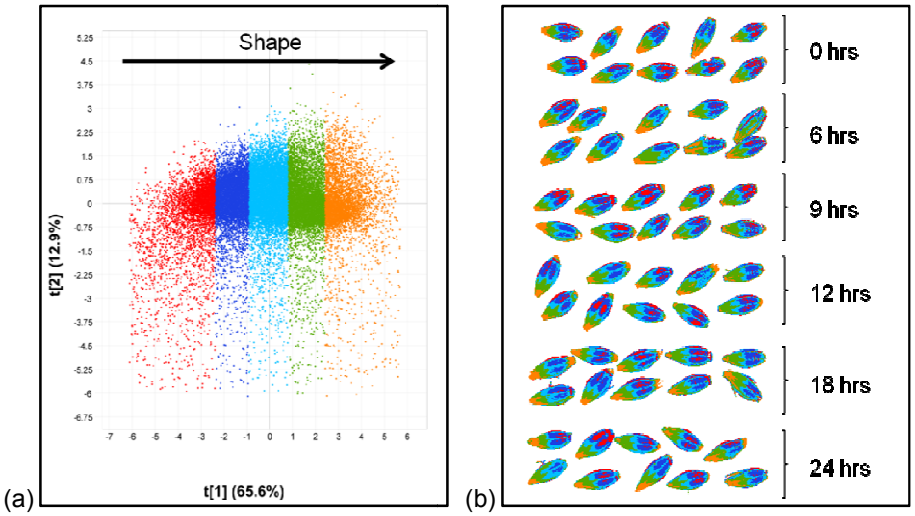


Figure 6.3 (a) Puma classification plot showing gradients from negative PC 1 (red) score values to positive (orange); and (b) classification gradient image of PC 1 showing kernel shape effects, the red areas are the highest and orange the lowest.

There were no distinct differences between kernels in the score image of PC 2 (**Fig. 6.2**). The variation was not due to differences between pre-germinated and non pre-germinated kernels. In the PC 1 vs. PC 2 classification gradient plot (**Fig. 6.4a**) the outermost positioned pixels (red and

green) on the negative and positive side of PC 2 were associated with the edges of the barley kernels in the classification gradient image (**Fig. 6.4b**). The pixels grouped around zero were associated with the embryo half (blue), and the endosperm and non embryo end (turquoise) of the kernels. The embryo half of the kernel (**Fig. 6.5**) is positioned lower than the endosperm and non embryo end (distal) when kernels are positioned crease down on a flat surface and this had an effect on the images. With the aid of classification gradient plots and images it was possible to explain the variation in the direction of PC 2; the shape of the kernel from the embryo to the endosperm-distal end of the kernel with some edge effects.

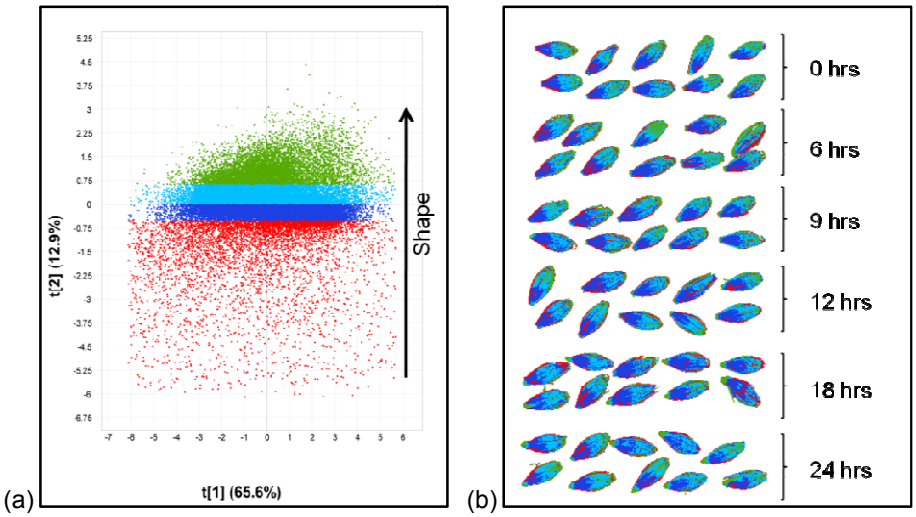


Figure 6.4 (a) Puma classification plot showing gradients from negative score values of PC 2 (red) to positive (green); and (b) classification gradient image of PC 2 showing kernel shape effects, the red and green are edges and the blue and turquoise are the lower positioned embryo and higher positioned endosperm halves respectively.

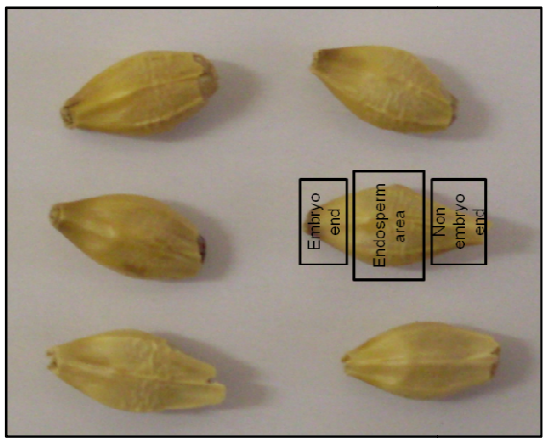


Figure 6.5 Barley kernels with shape areas marked.

The spectral variation explained by PC 3 was examined with the creation of a classification gradient plot and image. The two groups created in the classification gradient plot with negative PC

3 score values (**Fig. 6.6a**) were associated with the edge of the kernel (red) and the embryo end (blue) in the classification gradient image (**Fig. 6.6b**). The two groups with positive PC 3 score values were associated with the endosperm (turquoise) and distal end (green) of the kernels. PC 3 accounted for shape effects from the embryo to the distal end as well as some edge effects.

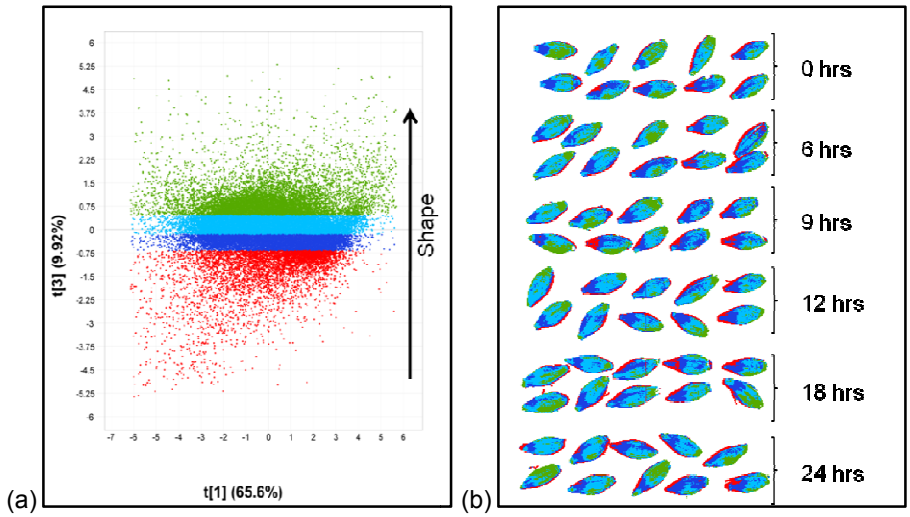


Figure 6.6 (a) Puma classification plot showing gradients from negative (red) to positive (green) PC 3 score values; and (b) classification gradient image of PC 3 showing kernel shape effects, the red areas are edge effects, the blue, turquoise and green areas are the embryo end, endosperm area and non embryo end respectively.

The image analysis of Puma PC 4 showed slight differences between pre-germinated and non pre-germinated kernels. These differences were more prominent in PC 5, thus PC 5 was used for further image analysis in the previous study (Chapter 3). A classification gradient plot and image was created for PC 4 to examine the variation in the direction of this PC (**Fig. 6.7**). The pixels in the classification gradient plot (**Fig. 6.7a**) with negative PC 4 score values were mostly associated with pixels from the pre-germinated subsets in the classification gradient image (**Fig. 6.7b**). The pixels with positive PC 4 score values were mostly associated with pixels from the non pre-germinated subsets. Another trend was observed; the most negative pixels (red) in the classification gradient plot (**Fig. 6.7a**) were associated with the lowest parts of the kernels in the classification gradient image. The pixels on the negative side of PC 4 (**Fig. 6.7a**) grouped closer to zero (blue) were associated with the embryo end of the kernels in the classification gradient image. The pixels on the positive side of PC 4 (**Fig. 6.7a**) grouped closer to zero were associated with the endosperm areas (turquoise) of the kernels and the most positive pixels were associated with the distal end (green) of the kernel (**Fig. 6.7b**). PC 4 accounted for shape effects and slight differences between pre-germinated and non pre-germinated kernels.

Gowen *et al*. (2008) indicated that the inherent shape of the mushrooms caused a scaling difference in the spectra. The shape of the barley kernel could have had the same effect as the

shape of the mushrooms. The difference in path length from the top of the kernel to the detector could have resulted in higher spectral reflectance intensities than the spectra from the lowest part of the kernel. Gowen *et al*. (2008) investigated four spectral pre-treatments (multiplicative scatter correction, median normalization, mean normalization and maximum normalization) and mean normalization was found to be the best pre-treatment. With mean normalization, PC 2 showed contrast between damaged and undamaged mushrooms, and this PC was used for classification. The shape of the barley kernel had a great effect on image analysis; the difference between pre-germinated and non pre-germinated barley kernels was only fully visible in PC 5 after SNV transformation.

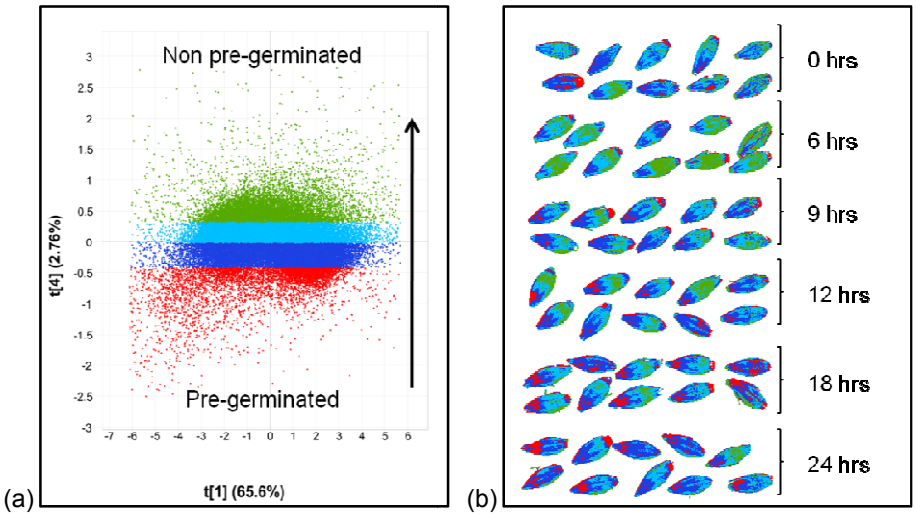


Figure 6.7 (a) Puma classification plot (PC 1 vs. PC 4) showing gradients from the negative (red) to positive score values (green); and (b) classification gradient image of PC 4 showing kernel shape effects and differences between the pre-germinated and non pre-germinated subsets.

Fig. 6.8 is an illustration of the effects seen from PC 1 to PC 5 for Puma. The top row is the same kernel from a non pre-germinated (viable) subset taken from the classification gradient image, represented over all 5 PCs. The bottom row is another kernel from a pre-germinated (non viable) subset taken from the classification gradient image, represented over all 5 PCs. From this illustration it was possible to observe that the first three PCs described kernel shape effects, and the last two PCs described the differences between pre-germinated and non pre-germinated kernels.

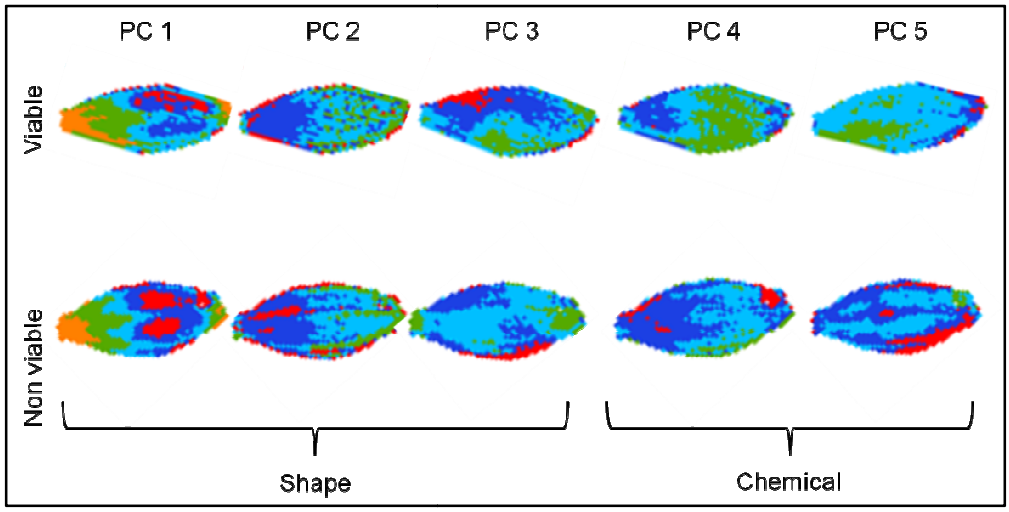


Figure 6.8 Puma (barley) kernels taken from the classification gradient images from PC 1 to 5. The top row is the same kernel from a non pre-germinated subset shown over all 5 PCs and the bottom row is a kernel from a pre-germinated subset shown over all 5 PCs.

Image analysis of cleaned wheat samples

During the image analysis of Duzi (Chapter 4, page 81) the PCA score images of PC 1 to PC 4 (not shown here) did not show differences between known pre-germinated and non pre-germinated subsets. Principal component 5 showed only a slight difference between pre-germinated and non pre-germinated subsets, but PC 6 showed definite differences and was used for further image analysis. Therefore, PC 1 to 5 was investigated to determine the variation explained in them. Classification gradient plots and images were created in the same manner as for Puma. **Fig. 6.8** only shows two kernels taken from the classification gradient images, represented over all 6 PCs. **Fig. 6.9** was sufficient to describe the variation in the PCs, thus the classification gradient plots and images are not shown.

The pixels with negative PC 1 score values were associated with the sides of the wheat kernels (**Fig. 6.9**) and the pixels with positive PC 1 score values were associated with the embryo and non embryo ends of the kernels. The pixels grouped around zero of PC 1 were associated with the highest part of the kernel. Consequently, PC 1 accounts for shape effects; the sides of the kernel in red and blue, the top of the kernel in turquoise, the area around the embryo and non embryo end in green to the embryo and non embryo end in yellow (**Fig. 6.9**).

The variation in the direction of PC 2 was investigated. The pixels with negative PC score values were associated with the lowest part of the kernel (including some shading) and the pixels with positive side PC score values were associated with the highest parts of the kernels. This is evident in **Fig. 6.9** where the lowest part of the kernels is coloured red and the highest part of the kernel in yellow or orange. Therefore, PC 2 accounted for the variation due to shape, from the lowest part of the kernel in red, to the highest in orange. PC 3 is the reverse of PC 2, where the negative pixels were associated with the top of the kernel and the positive pixels were associated

with the bottom of the kernel (**Fig. 6.9**). PC 4 revealed the same trend as PC 2, it accounted for the shape of the kernel from the lowest part (red) to the highest part (yellow).

PC 5 showed slight differences between the pre-germinated and non pre-germinated kernels. The variation was inspected and it was found that the non pre-germinated subsets included more pixels with negative (red) PC 5 score values and the pre-germinated subsets included more pixels with positive (yellow) PC 5 score values (**Fig. 6.9**). PC 5 also revealed that the negative score were associated with the bottom of the kernel (red) and the positive pixels were associated with the top of the kernel (yellow). PC 5 accounted for shape effects and slight differences between pre-germinated and non pre-germinated subsets.

Distinct differences between pre-germinated and non pre-germinated subsets were seen in PC 6 of Duzi. The non pre-germinated subsets in PC 6 included more pixels with negative (red and blue) PC 6 score values and the pre-germinated subsets included more pixels with positive (turquoise and green) PC 6 score values. The shape of the wheat kernels had a greater effect than the barley kernels, as the differences between the pre-germinated and non pre-germinated subsets were only fully observed in PC 6 after SNV transformation.

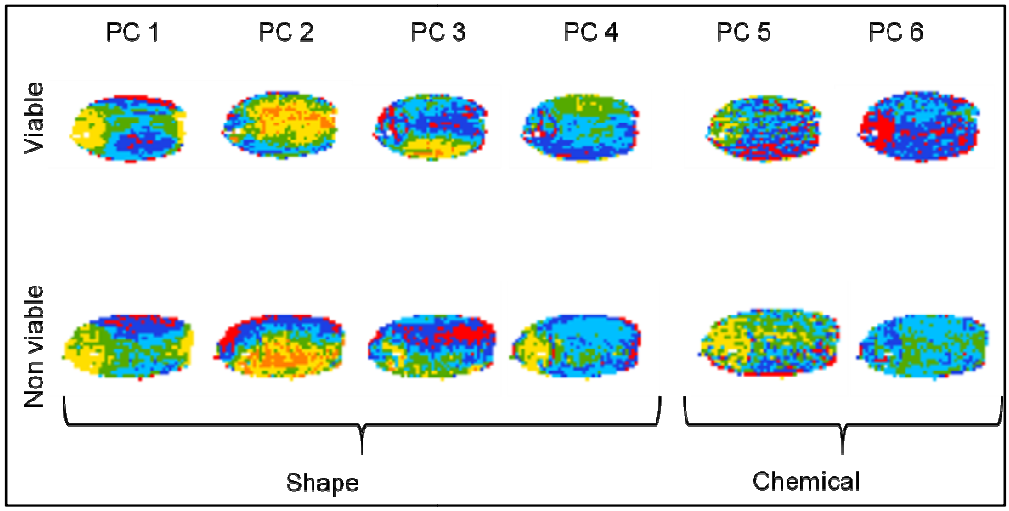


Figure 6.9 Duzi (wheat) kernels taken from the classification gradient images from PC 1 to 6. The top row is the same kernel from a non pre-germinated subset shown over all 6 PCs and the bottom row is a different kernel from a pre-germinated subset shown over all 6 PCs.

Image analysis of cleaned sorghum samples

The image analysis of SA landrace 4442 (Chapter 5, page 107) revealed no difference between pre-germinated and non pre-germinated subsets from PC 1 to PC 3. PC 4 (as for Puma) showed slight differences between pre-germinated and non pre-germinate subsets, but PC 5 revealed distinct differences. Classification gradient plots and images were created similarly to Puma and Duzi. **Fig. 6.10** was sufficient to describe the variation in the PCs.

The pixels in the on the negative side of PC 1 were associated with the lowest parts of the sorghum kernels (including some shading errors) and the pixels on the positive side of PC 1 were associated with the highest parts of the kernels (**Fig. 6.10**). PC 1 accounted for shape effects; the lowest part of the kernels is red and blue and the top of the kernel are turquoise and green.

The variation in the direction of PC 2 was explained with classification gradient images and plots. The pixels with negative PC score values were associated with the lowest part of the kernel on the embryo end (including some shading) and the pixels with positive PC score values were associated with the lowest part of the kernel on the non embryo end. Therefore, PC 2 accounted for the variation due to shape, from the lowest part of the kernel on the embryo end in red, to the lowest part on the non embryo end in orange. PC 3 is the reverse of PC 2, where the negative pixels were associated with the lowest part of the kernel on the non embryo end and the positive pixels were associated with the lowest part of the kernel on the embryo end (**Fig. 6.10**).

PC 4 showed slight differences between the pre-germinated and non pre-germinated subsets. The non pre-germinated subsets included more pixels with positive (turquoise-green) PC 5 score values and the pre-germinated subsets included more pixels with negative (red-blue) PC 4 score values (**Fig. 6.10**). PC 4 also revealed that the negative scores were associated with the top of the kernel (red) and the positive pixels were associated with the bottom of the kernel (green). PC 4 accounted for shape effects and slight differences between pre-germinated and non pre-germinated kernels.

SA landrace 4442 revealed distinct differences between pre-germinated and non pre-germinated subsets in the direction of PC 5. The non pre-germinated subsets in PC 5 included more pixels with negative (red and blue) PC 5 score values and the pre-germinated subsets included more pixels with positive (turquoise and green) PC 5 score values (**Fig. 6.10**).

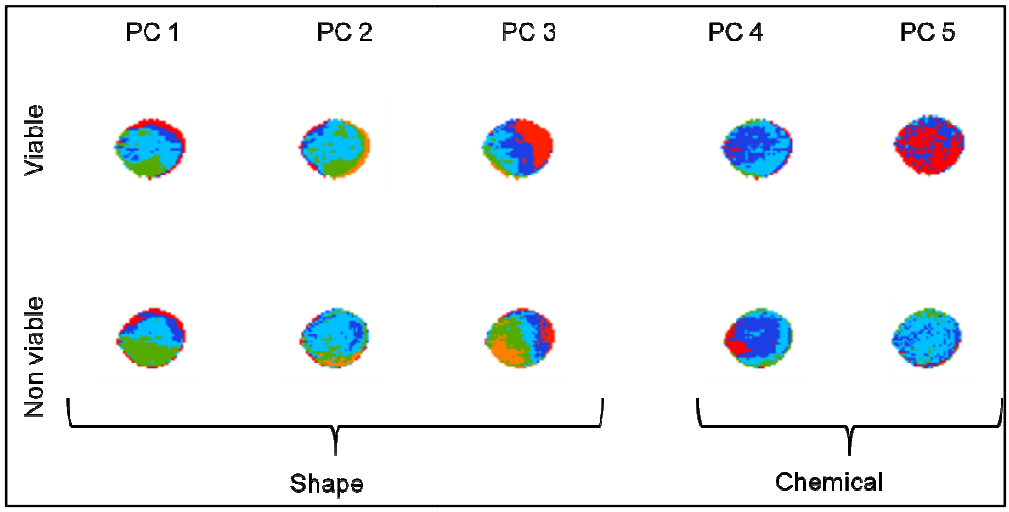


Figure 6.10 SA landrace 4442 (sorghum) kernels taken from the classification gradient images of PC 1 to 5. The top row is the same kernel from a non pre-germinated subset shown over all 5 PCs and the bottom row is a kernel from a pre-germinated subset shown over all 5 PCs.

Conclusion

Classification gradient images and plots were interpreted to determine the nature of spectral variation explained in the lower order PCs. The variation in the lower order PCs was due to kernel shape effects. The variation caused by kernel shape was not explained in only one PC, even after pre-processing was applied. Therefore, the shape of the cereal kernels has a great influence on hyperspectral image analysis. It is important to consider the physical characteristics of a sample before hyperspectral imaging and image analysis; kernels should be arranged in such a way to minimize the effects of shape. Different pre-processing techniques and combinations of pre-processing techniques should be investigated and the best technique, or combination, should be employed to reduce the effects of shape. Image analysis allows the effective identification of unwanted data and the subsequent exclusion of such data.

References

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.

ElMasry, G. & Sun, D.-W. (2010). Principles of hyperspectral imaging technology. In: *Hyperspectral imaging for food quality analysis and control* (edited by D.-W. Sun). Pp. 3-45. San Diego: Academic Press Elsevier.

Gaston, E., Frias, J.M., Cullen, P.J., O'Donnell, C.P. & Gowen, A.A. (2010). Prediction of polyphenol oxidase activity using visible near infrared hyperspectral imaging on mushrooms (*Agaricus bisporus*) caps. *Journal of Agricultural and Food Chemistry*, **58**, 6226-6233.

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

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 bisporus*). *Journal of Chemometrics*, **22**, 259-267.

Chapter 7

General discussion and conclusion

Chapter 7

General discussion and conclusion

Near infrared (NIR) hyperspectral imaging, also known as NIR chemical imaging, combines conventional imaging with NIR spectroscopy (Geladi *et al.*, 2004; Burger & Geladi, 2005; Geladi *et al.*, 2007; Gowen *et al.*, 2007; Wang & Paliwal, 2007; Singh *et al.*, 2009). Hyperspectral images not only have great potential for describing concentration and identifying chemical components, but also describing chemical component distribution within a sample (Martinsen & Schaare, 1998; Burger & Geladi, 2005). In this research, NIR hyperspectral imaging was evaluated as a non destructive method for the the detection of pre-germination in whole barley, wheat and sorghum grains. In addition, the effect of kernel shape on hyperspectral images was evaluated.

Pre-germination is the unwanted germination of mature cereal grains subjected to wet or humid conditions. Pre-germination decreases grain weight, yield and functionality (Mohan, 2008). This will result in significant economic losses as the price paid to the farmer is greatly reduced (Skerritt & Heywood, 2000). Pre-germination can be interrupted by drying of the kernel before any visible indication, however, irreversible enzyme degradation has usually already commenced (Carn, 1982; Lin *et al.*, 2008). Viable, non pre-germinated barley and sorghum is required for malting purposes (Lijavetzky *et al.*, 2000; Lin *et al.*, 2008). An excess amount of alpha-amylase (found in pre-germinated wheat) has damaging effects on bread production (Kruger & Reed, 1988; Kent & Evers, 1994; Kruger, 1994). The detection methods used in the industry are either subjective (visual inspection) or destructive and time-consuming (viscosity analysis, falling number and tetrazolium test). Therefore a need has emerged for a rapid, objective and non-destructive test for pre-germination.

Two spatial resolutions were trialled over two sample sets. For the first sample set no specific individual kernel viability information was available, just an average result. For the second sample set specific individual kernel viability information was collected. The first sample set was investigated using 150 µm spatial resolution images and the second sample set was investigated using 30 µm spatial resolution images.

The analysis of the principal component (PC) 5 (barley and sorghum) and PC 6 (wheat) loading line plots of the 150 µm spatial resolution images showed that the main compounds contributing to the spectral variation were starch, water and protein. It was anticipated that starch and protein could have an influence on the images because starch and proteins are hydrolysed by hydrolytic enzymes during pre-germination. It was however not anticipated that water (1940 nm) would have an influence on the images. The difference between pre-germinated and non pre-germinated kernels was not simply a function of moisture content as the moisture content of the 0 and 24 hrs subsets were similar. Rather, the water environment differed between pre-germinated and non pre-germinated kernels. In pre-germinated kernels starch and protein have been

hydrolysed, this process requires the interaction of water with enzymes, starch and proteins. NIR hyperspectral imaging allows the user to obtain results from single kernels, even though multiple kernels were imaged. This technique could allow the grader to evaluate single kernels on the basis of their chemical differences.

The barley partial least squares discriminant analysis (PLS-DA) model were developed from Erica. The prediction results of Puma revealed that the model predicted 34.57% false positives and 36.07% false negatives for SSG564. The PLS-DA model created for the barley cultivars of the 150 µm spatial resolution images cannot be employed in the barley malting industry, the non pre-germinated and pre-germinated barley will not be correctly predicted. The prediction results of wheat cultivar Biedou revealed that the model developed from Duzi correctly predicted 81.63% of the non pre-germinated pixels and 100% (6.82% over prediction) of the pre-germinated pixels. This model performed better than the barley model. The same model correctly classified 92.21% of the non pre-germinated pixels of Kariega and 93.25% of the pre-germinated pixels. This model may be used to discriminate between intact pre-germinated and non pre-germinated wheat kernels. The prediction results of PAN 8816 revealed that the model developed from SA landrace 4442, correctly predicted 97.13% and 93.93% of the non pre-germinated and pre-germinated kernels, respectively. This model predicted better than the barley and wheat models as no false positive or negatives were recorded. Therefore this model may be employed to distinguish between whole pre-germinated and non pre-germinated sorghum kernels. To completely validate the models, they should be applied to images of field pre-germinated kernels. Laboratory pre-germinated kernels may be quite different from field pre-germinated kernels (Kruger, 1994).

A distinction between pre-germinated and non pre-germinated kernels was observed in the 30 µm spatial resolution images. However, some kernels determined to be non viable (pre-germinated) by the tetrazolium test were similar to non pre-germinated kernels in the score image. The endosperm of these kernels may have undergone incomplete degradation by hydrolytic enzymes. The NIR hyperspectral imaging method is perhaps detecting degree of endosperm hydrolysis rather than kernel viability. NIR hyperspectral imaging may be employed as an indirect method for endosperm hydrolysis rather than a tool for viability determinations. NIR hyperspectral imaging and viscosity analysis or falling number results should be compared to determine the feasibility of this proposal.

The loading line plots of the 30 µm spatial resolution images revealed that the main compounds contributing to the spectral variation were water, starch and protein. These compounds did however differ from the 150 µm spatial resolution images with regards to their association with pre-germinated and non pre-germinated kernels. In some cases the water, starch and protein peaks were associated with pre-germinated kernels and in other cases with the non pre-germinated kernels. It is not clear what the reason for this might be, but a possible reason could be that the kernels used for the 150 µm spatial resolution images underwent profound endosperm hydrolysis. In contrast to the 150 µm spatial resolution images the 30 µm spatial resolution images

did not reveal the spectral differences in one consistent PC, i.e. the differences between pre-germinated and non pre-germinated kernels was observed in PC 2 for one cultivar and PC 5 for another. Contributions not attributable to pre-germination might have an influence on the loading line plots of the samples where pre-germination differences were captured in the lower order PCs. The water peak at 1940 nm dominated the loading line plots. It would be interesting to see how field pre-germinated kernels compare to laboratory pre-germinated kernels with respect to this water peak. Future studies should include images of field pre-germinated kernels.

The developed PLS-DA models of the 30 µm spatial resolution images did not perform satisfactory. Model 1 resulted in excessive false negatives, non pre-germinated pixels was not predicted as such. This model will lead to the down grading of viable kernels, and will lead to economic losses for the farmers. Model 2 resulted in excessive false positives, pre-germinated pixels was predicted as non pre-germinated. Such a model will result in production losses if pre-germinated kernels are used. The PLS-DA models were developed from a small amount of kernels (15 -21 kernels), and the models were developed from kernels with varying degrees of endosperm hydrolysis. Future research should include PLS-DA models developed from more samples and include only kernels from the most extreme time points.

NIR hyperspectral imaging has several advantages over conventional analytical methods, the primary being the ability to assess the distribution of chemical constituents within a sample (Burger & Geladi, 2005; Gowen *et al.*, 2008). However, NIR hyperspectral imaging suffers the collective errors of conventional spectroscopy and imaging (e.g. shape contributions) (Burger & Geladi, 2005; Burger & Geladi, 2007; Gowen *et al.*, 2008). The shape of whole grain samples has an effect on the spectral dataset. It was observed during the exploration of the 150 µm spatial resolution images that the differences between pre-germinated and non pre-germinated kernels were only explained in PC 5 (barley and sorghum) and higher (wheat). The lower order PCs (PC 1 to PC 4 or 5) were subsequently examined using classification gradients. Classification gradient images and plots served to be an effective and easy-to-use method to determine the variation explained in the lower order PCs. It was observed that the variation was attributable to kernel shape effects. The classification gradient plots of the different PCs revealed that pixels associated with a certain part of the kernel lay close together, resulting in the effective employment of classification gradients. The shape of the kernel from the highest to lowest points, edge effects and different halves (embryo and endosperm) were evident in the classification images. The variation caused by kernel shape was not explained in only one PC, even after pre-processing was applied. The physical characteristics of a sample should be considered in future research and kernels should be arranged to minimise shape effects. The reduction of spectral variability caused by shape effects might be minimised by pre-processing or combinations of pre-processing techniques.

The detection of pre-germination in whole barley, wheat and sorghum kernels was illustrated with the use of non-destructive NIR hyperspectral imaging. Differences between pre-germinated and non pre-germinated kernels were observed and lead to the classification of these images in

the score plots. The possible incomplete endosperm hydrolysis was observed when individual kernel viability was determined. NIR hyperspectral imaging might be better employed as an indirect method to determine endosperm hydrolysis as an alternative to destructive viscosity analysis and falling number. NIR hyperspectral imaging shows promise for the detection of pre-germination in grains, however, further research is needed.

References

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

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.

Carn, J.D. (1982). Alpha-amylase indicates problems with malting of stored barley. *Food Technology in Australia*, **34**, 82-83.

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

Geladi, P.L.M., Grahn, H.F. & Burger, J.E. (2007). Multivariate images, hyperspectral images: background and equipment. In: *Techniques and Applications of Hyperspectral Image Analysis* (edited by H.F. Grahn & P.L.M. Geladi). Pp. 1-15. Chichester: John Wiley & Sons, Ltd.

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 and 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 bisporus*). *Journal of Chemometrics*, **22**, 259-267.

Kent, N.L. & Evers, A.D. (1994). *Kent's technology of cereals.* New York: Elsevier Science Inc.

Kruger, J.E. (1994). Enzymes of sprouted wheat and their possible technological significance. In: *Wheat Production, Properties and Quality* (edited by W. Bushuk & V.F. Rasper). Pp. 143-151. Glasgow: Blackie Academic and Professional.

Kruger, J.E. & Reed, G. (1988). Enzymes and colour. In: *Wheat: Chemistry and Technology* (edited by Y. Pomeranz). Pp. 441-487. St. Paul: American Association of Cereal Chemists.

Lijavetzky, D., Martinez, M.C., Carrari, F. & Hopp, H.E. (2000). QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. *Euphytica*, **112**, 125-135.

Lin, R., Horsley, R.D. & Schwarz, P.B. (2008). Associations between caryopsis dormancy, alpha-amylase activity and pre-harvest sprouting in barley. *Journal of Cereal Science*, **48**, 446-456.

Martinsen, P. & Schaare, P. (1998). Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology*, **14**, 271-281.

Mohan, A. (2008). Pre-harvest sprouting in cereals: a global scenario. *Current Science*, **94**, 704-705.

Singh, C.B., Jayas, D.S., Paliwal, J. & White, N.D.G. (2009). Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry*, **86**, 256-260.

Skerritt, J.H. & Heywood, R.H. (2000). A five-minute field test for on-farm detection of pre-harvest sprouting in wheat. *Crop Science*, **40**, 742-756.

Wang, W. & Paliwal, J. (2007). Near infrared spectroscopy and imaging in food quality and safety. *Sensing and Instrumentation for Food Quality and Safety*, **1**, 193-207.