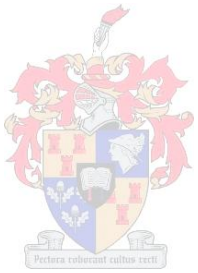
DETECTION AND QUANTIFICATION OF SPICE ADULTERATION BY NEAR INFRARED HYPERSPECTRAL IMAGING

DANWILLE JACQWIN FRANCO SEPTEMBER

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

MASTER OF SCIENCE IN FOOD SCIENCE



Department of Food Science

Faculty of AgriSciences

Stellenbosch University

Study Leader: Prof Marena Manley

Co-study Leader: Prof Paul Geladi

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.

March 2011

ABSTRACT

Near infrared hyperspectral imaging (NIR HSI) in conjunction with multivariate image analysis was evaluated for the detection of millet and buckwheat flour in ground black pepper. Additionally, mid-infrared (MIR) spectroscopy was used for the quantification of millet and buckwheat flour in ground black pepper. These techniques were applied as they allow non-destructive, invasive and rapid analysis.

Black pepper and adulterant (either millet or buckwheat flour) mixtures were made in 5% (w/w) increments spanning the range 0-100% (w/w). The mixtures were transferred to eppendorf tube holders and imaged with a sisuChema short wave infrared (SWIR) pushbroom imaging system across the spectral range of 1000–2498 nm. Principal component analysis (PCA) was applied to pseudo-absorbance images for the removal of unwanted data (e.g. background, shading effects and bad pixels). PCA was subsequently applied to the 'cleaned' data. An adulterant concentration related gradient was observed in principal component one (PC1) and a difference between black pepper adulterated with buckwheat and millet was noted in PC4. Four absorption peaks (1461, 2241, 2303 and 2347 nm) were identified in the loading line plot of PC1 that are associated with protein and oil. The loading line plot of PC4 revealed absorption peaks at 1955, 1999, 2136 and 2303 nm, that are related to protein and oil. Partial least squares discriminant analysis (PLS-DA) was applied to NIR HSI images for discrimination between black pepper adulterated with varying amounts of adulterant (millet or buckwheat). The model created with millet adulterated black pepper samples had a classification accuracy of 77%; a classification accuracy of 70% was obtained for the buckwheat adulterated black pepper samples.

An average spectrum was calculated for each sample in the NIR HSI images and the resultant spectra were used for the quantification of adulterant (millet or buckwheat) in ground black pepper. All samples were also analysed using an attenuated total reflectance (ATR) Fourier transform (FT) – infrared (IR) instrument and MIR spectra were collected between 576 and 3999 cm^{-1}. PLS regression was employed. NIR based predictions (r^2 = 0.99, *RMSEP* = 3.02% (w/w), PLS factor = 4) were more accurate than MIR based predictions (r^2 = 0.56, *RMSEP* = 19.94% (w/w), PLS factors = 7). Preprocessed NIR spectra revealed adulterant specific absorption bands (1743, 2112 and 2167 nm) whereas preprocessed MIR spectra revealed a buckwheat specific signal at 1574 cm^{-1}.

NIR HSI has great promise for both the qualitative and quantitative analysis of powdered food products. Our study signals the beginning of incorporating hyperspectral imaging in the analysis of powdered food substances and results can be improved with advances in instrumental development and better sample preparation.

UITTREKSEL

Die gebruik van naby infrarooi hiperspektrale beelding (NIR HB) tesame met veelvoudige beeldanalise is ondersoek vir die opsporing van stysel-verwante produkte (giers en bokwiet) in gemaalde swart pepper. Middel-infrarooi (MIR) spektroskopie is addisioneel gebruik vir die kwantifisering van hierdie stysel-verwante produkte in swart pepper. Albei hierdie tegnieke is toegepas aangesien dit deurdringend van aard is en dit bied nie-destruktiewe sowel as spoedige analise.

Swart pepper en vervalsingsmiddel (giers of bokwiet) mengsels is uitgevoer in 5% (m/m) inkremente tussen 0 en 100% (m/m). Eppendorfbuishouers is met die mengsels gevul en hiperspektrale beelde is verkry deur die gebruik van ‘n *sisuChema SWIR* (kortgolf infrarooi) kamera met ‘n spektrale reikwydte van 1000–2498 nm. Hoofkomponent-analise (HK) is toegepas op pseudo-absorbansie beelde vir die verwydering van ongewenste data (bv. agtergrond, skadu en dooie piksels). Hoofkomponent-analise is vervolgens toegepas op die ‘skoon’ data. Hoofkomponent (HK) een (HK1) het die aanwesigheid van ‘n vervalsingsmiddel konsentrasie verwante gradient getoon terwyl HK4 ‘n verskil getoon het tussen swart pepper vervals met giers en bokwiet. Vier absorpsiepieke (1461, 2241, 2303 en 2347 nm) was geïdentifiseer binne die HK lading stip van HK1 wat met proteïen en olie geassosieer kon word. Die HK lading stip van HK4 het absorpsipieke by 1955, 1999, 2136 en 2303 nm aangedui wat verband hou met proteïen en olie. Parsiële kleinste waarde diskriminant-analise (PKW-DA) is toegepas op die hiperspektrale beelde vir die moontlike onderskeiding tussen swart pepper vervals met verskeie hoeveelhede vervalsingsmiddel (giers of bokwiet). ‘n Klassifikasie koers van 77% is verkry vir die model ontwikkel met giers vervalsde swart pepper terwyl die model ontwikkel met bokwiet vervalsde swarte pepper ‘n klassifikasie koers van 70% bereik het.

‘n Gemiddelde spektrum is bereken vir elke monster in die hiperspektrale beelde en die resulterende spektra is gebruik vir die kwantifisering van vervalsingsmiddels (giers of bokwiet) in gemaalde swart pepper. ‘n *ATR FT-IR* instrument met spektrale reikwydte van 576-3999 cm^{-1} is additioneel gebruik vir die analise van alle monsters. Parsiële kleinste waarde regressie is gebruik vir kwantifikasie doeleindes. NIR gebasseerde voorspellings (r^2 = 0.99, *RMSEP* = 3.02% (m/m), PLS faktore = 4) was meer akkuraat as die MIR gebasseerde voorspellings (r^2 = 0.56, *RMSEP* = 19.94% (m/m), PLS faktore = 7). Vooraf behandelde NIR spektra het vervalsingsmiddel verwante absorpsiepieke (1743, 2112 en 2167 nm) aangetoon terwyl vooraf behandelde MIR spektra ‘n bokwiet verwante absorpsiepiek by 1574 cm^{-1} aangedui het.

NIR HB toon goeie potensiaal vir beide kwalitatiewe en kwantitatiewe analise van gepoeierde voedsel produkte. Ons studie kan gesien word as die begin van die inkorporasie van hiperspektrale beelding in die analise van gepoeierde voedsel material en verbeterde resulte kan verkry word deur die vordering in instrumentasie ontwikkeling en verbeterde monstervoorbereiding.

ACKNOWLEDGEMENTS

I want to thank the following people and institutions for the help they have provided throughout the duration of my postgraduate studies:

Prof Marena Manley (my supervisor) for making the invisible visible in my life, pushing me to reach my fullest scientific potential and for granting me the opportunity to go abroad;

Prof Paul Geladi (Swedish University of Agricultural Sciences, Umeå, Sweden, my co-supervisor) for all the insightful discussions on life, imaging and spectroscopy;

Dr Cushla McGoverin (NIRD head girl) for her assistance during my scientific growth, I appreciate it Cushla;

Oskar Jonsson and David Nilsson (Umbio AB, Umeå, Sweden) for allowing me to use the sisuChema imaging system, assistance during imaging acquisition and supply us with the Evince software packages (2009);

Prof Alvaro Viljoen and Ilze Vermaak (Tshwane University of Technology) for assistance during image acquisition (2010);

Paul Williams (NIRD head boy) for his inputs during my presentations (posters and seminars) and all the cold beers we shared at Brazen Head over an imaging discussion;

The following South African spice companies: Deli Spices, Freddy Hirsch, Spice Mecca and Exim Spices International for providing me with spices for this study;

National Department of Agriculture (South Africa) for my bursary and granting me the opportunity to further my academic career;

The NIRDs and Food Science staff for all the laughs, support and help;

Stephanie Eslick for her love, always being there and for always keeping me sane (as well as her family for all their support);

Finally my mom (Francina September) for her continuous support throughout my academic years, for always believing that I can do it and motivating me to only reach the best. My dad (Andreas Spetember) and sister (Gillian September) for their support and willingness to help me. The Martin

family (especially Tina Martin) for their undoubted support. Thank you all for always being there for me!

ABBREVIATIONS

AACC	- American Association of Cereal Chemists International
AOAC	- Association of Official Agricultural Chemists
ASTA	- American Spice Trade Association
ATR	- Attenuated total reflectance
CCD	- Charged coupled device
DNA	- Deoxyribonucleic acid
GC	- Gas chromatography
ELISA	- Enzyme linked immunosorbent assay
ESA	- European Spice Association
GC MS	- Gas chromatography mass spectroscopy
FPA	- Focal plane array
FT	- Fourier transform
HIA	- Hyperspectral imaging analysis
HgCdTe	- Mecury Cadmium Telluride
HPLC	- High performance liquid chromatography
InGaAs	- Indium Gallium Arsenide
IISR	- Indian Institute of Spices Research
ISI	- Indian Standards Institute
ISO	- International Organisation for Standardisation
KNN	- *K*-Nearest neigbour
LC MS	- Liquid chromatography mass spectroscopy
LCTF	- Liquid-crystal tunable filters
MIA	- Multivariate image analysis
MIR	- Mid-infrared
MSC	- Multiplicative scattering correction
NIRS	- Near infrared spectroscopy
NIR HSI	- Near infrared hyperspectral Imaging
PCA	- Principal component analysis
PCR	- Polymerase chain reaction
PC	- Principal component
PLS	- Partial least squares
PLS-DA	- Partial least squares discriminant analysis
PRESS	- prediction sum of squares
RAPD	- Rapid amplified polymorphic DNA
RMSEC	- Root mean square error of calibration
RMSECV	- Root mean square error of cross validation

RMSEP	- Root mean square error of prediction
R^2	- Coefficient of determination (calibration)
r^2	- Coefficient of determination (validation)
SCAR	- Sequence characterised amplified regions
SEC	- Standard error of calibration
SEP	- Standard error of prediction
SIMCA	- Soft independent modelling of class analogy
SNV	- Standard normal variate
SWIR	- Short wave infrared
TLC	- Thin layer chromatography

LIST OF FIGURES

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 2

Chapter 3

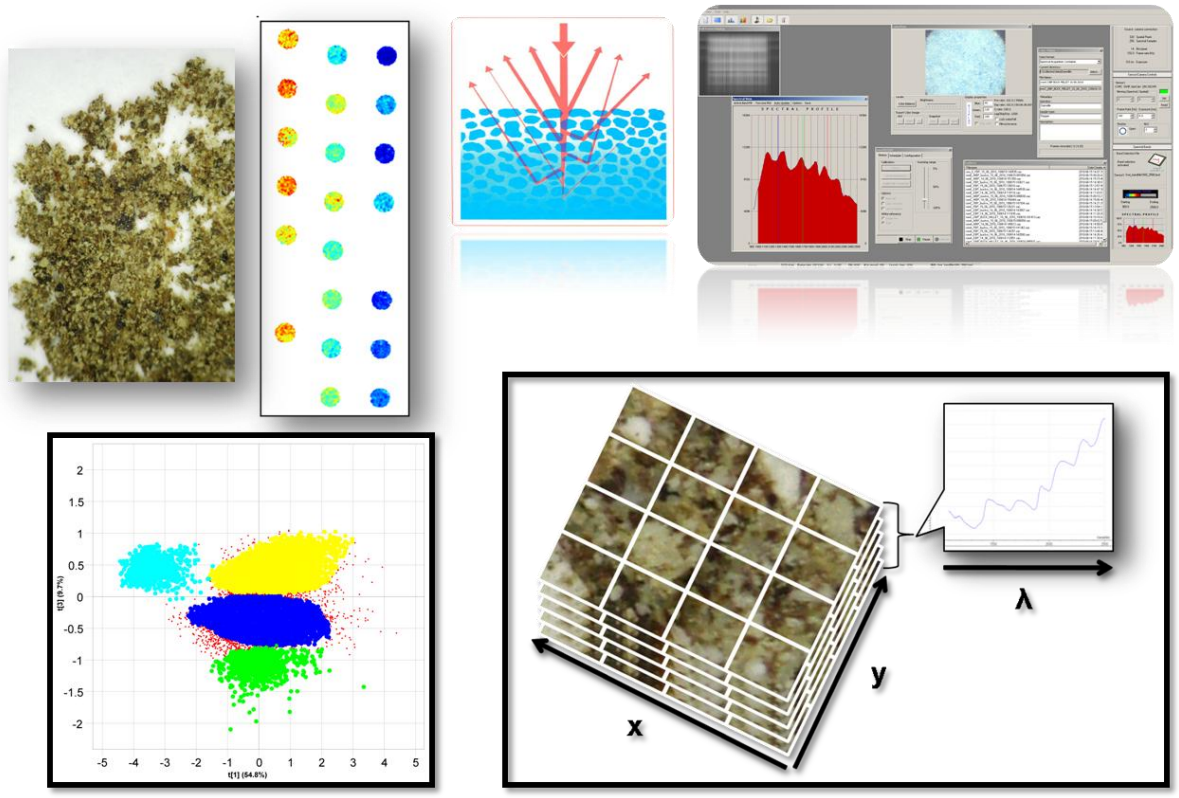
Chapter 4

TABLE OF CONTENTS

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

Chapter 1

Introduction



Introduction

The use of spices has been valued since the Pyramid Age of Egypt which began approximately 2600 B.C. (Rosengarten, 1969). Spices had great monetary value in the early ages as they were only available to the rich (Dhanya & Sasikumar, 2010). Spices have a vast range of functionalities which include flavouring of food, perfumery, cosmetics, medicinal use and preservative properties (McKee, 1995; Minakshi *et al.*, 1999; Bhattacharjee *et al.*, 2003; Sagoo *et al.*, 2008). Although spices added aroma and flavour characteristics to food, they were more valued as perfumery, cosmetic and medicinal agents in ancient times (Rosergarten, 1969; McKee, 1995). Today large amounts of spices are utilised by the commercial sector of industrialised countries, primarily in food processing (Farag Zaied *et al.*, 1996). The use of spices pervades the food industry, for example spices are used in meat, fish, bakery and vegetable products. In 1969 the Commonwealth secretariat estimated the international trade of spices to be worth R 1.2 billion; the current trade is estimated to be R 22.4 – 26.1 billion (Pruthi, 1980; Anon., 2009). The spice trade has been growing globally and with increasing consumer demand the global spice trade is expected to continue growing (Anon., 2009). The increase of both the value of spices and consumer demand has increased the demand for better quality spices from suppliers and retailers. This has encouraged illegal or fraudulent practises (e.g. adulteration), and continuous monitoring is required to identify incidences of such practices.

Food adulteration (e.g. spice adulteration) can be categorised into two separate groups namely, incidental and intentional adulteration (Anon., 2009). Incidental adulteration occurs when foreign substances are added to a food due to ignorance, negligence or improper facilities. This can occur during the harvesting of spices; examples include inclusion of pesticide residues, droppings of rodents/birds and bacterial contamination (Anon., 2009).

Intentional adulteration, better known as economic adulteration, entails the deliberate addition of inferior materials to a food to heighten appearance qualities and value for economical gain (ASTA, 2004). These inferior substances include ground material (e.g. saw dust), leave, powdered products (e.g. starches) and other spice species (Woodman, 1941; ASTA, 2004). Authentication of food has become an increasingly important factor for both the food industry and the more informed consumer (Reid *et al.*, 2006). The development of new techniques for the determination of adulteration and authentication of food is of major concern with regard to food safety and the elimination of unfair or fraudulent practices in food processing. These techniques must be sensitive, accurate and analytical, because the perpetrators involved in this type of fraud have access to adulteration methods that are difficult to detect**.**

The American Spice Trade Association (ASTA) is an institutional body ensuring the trade of good quality spices (e.g. adulteration free) in every industrial domain including the consumer retail market (ASTA, 2004). This organisation suggests different analytical methods for the identification

of spice adulteration. These include the ASTA starch method, microscopic analysis, thin layer chromatography (TLC), high performance liquid chromatography (HPLC) and spetrophotometric methods (e.g. UV/Vis spectrophotometry). Other organisational bodies such as the Association of Official Agricultural Chemists (AOAC) and International Organisation for Standardisation (ISO) also prescribe methods for the identification of spice adulteration. These include the determination of moisture content, total ash, volatile and non-volatile ash content and extraneous matter content (AOAC, 2005; ESA, 2007). All these methods are well established and are both sensitive and accurate when identifying spice adulterants (Pruthi, 1980, Dhanya & Sasikumar, 2010). However, these methods are time-consuming, costly and may require specialised personnel for system operations. Thus the opportunity for the development of rapid and more straightforward methods of analysis exists.

Near infrared (NIR) spectroscopy is a promising technique in this regard owing to its simplicity, rapidity and non-destructive nature (Osborne *et al.*, 1993). This vibrational spectroscopic technique is based on the absorption of electromagnetic radiation at wavelengths in the range 700 – 2500 nm (Osborne *et al.*, 1993; Bokobza, 1998). Observed absorption bands in the NIR region correspond to overtones and combinations of carbon-hydrogen (CH), oxygen-hydrogen (OH) and nitrogen-hydrogen (NH) vibrations. These chemical bonds are common in food constituents such as carbohydrates, protein, lipids and water. NIR spectra are a molecular fingerprint as the environment of these chemical bonds determines the energy of the overtone and combination vibrational modes. In addition, the absorption exhibits concentration dependence, hence NIR spectroscopy may be used for the quantification of components within a sample. NIR absorption bands are generally broad and overlapping resulting in complex spectra, necessitating more complex methods than univariate analyses for quantification (Wehling, 2003). A measured spectrum consists of both useful and irrelevant information, which complicates component quantification (Rodionova *et al.*, 2005).

Chemometrics is a statistical discipline, the purpose of which is to aid interpretation of chemical data. Chemometrics is usually applied to spectroscopic data in the form of multivariate data analysis techniques. These techniques include principal component analysis (PCA), partial least square discriminant analysis (PLS-DA) and PLS regression and they all reduce the dimensionality of complex data sets, i.e. spectroscopic data (Lavine & Workman, 2005). NIR spectroscopy in combination with chemometrics can be utilised to develop prediction models to identify adulterated supplies of spice in either the industrial sector or the consumer driven retail market. NIR spectroscopy with chemometrics has been applied successfully in the industrial arena these include agricultural (Wehling *et al.*, 1993), pharmaceutical (Rodionova *et al.*, 2005) and food industries (Cocchi *et al.*, 2006; Juliani *et al.*, 2006; Hernández-Hierro *et al.*, 2008; Wu *et al.*, 2009).

NIR hyperspectral imaging is a technique created by the fusion of conventional NIR spectroscopy and digital imaging (Geladi *et al.*, 2004). This technique creates hypercubes or three dimensional blocks of data, consisting of two spatial (x and y) and one spectral (λ) dimension

(Burger & Geladi, 2005). The hypercubes, also termed hyperspectral images, provide a visual presentation of the biochemical constituents in a food sample (Gowen *et al.*, 2007). With the added spatial dimensions more information is available in a hyperspectral image (Burger & Geladi, 2005) than is the case for a bulk NIR spectrum. As with bulk NIR spectroscopic analyses, multivariate techniques are applied to the analysis of hyperspectral data. These techniques have been extensively applied in agriculture (Robert *et al.*, 1992; Polder *et al.*, 2002; Beaten et al., 2005; Qin & Lu, 2006; Gowen *et al.*, 2008; Gómez-Sanchis *et al.*, 2008; Wang & ElMasry, 2010), meat (ElMasry & Sun, 2010), pharmaceuticals (Ma & Anderson, 2008) and medicine (Schultz *et al.*, 2001).

In this study the potential of near infrared (NIR) hyperspectral imaging in conjunction with multivariate image analysis was evaluated for both the identification and quantification of foreign substances (buckwheat and millet flour) in ground black pepper. More specifically:

- *PCA* was used to explore the data and to establish any possible sample relations;
- *PLS-DA* models were created for discrimination based on adulterant and black pepper content in sample mixtures; and
- *PLS* regression was applied to quantify adulterant content in ground black pepper sample mixtures.

Additionally, mid-infrared (MIR) spectroscopy in conjunction with PLS regression was employed to quantify adulterant content in ground black pepper.

References

American Spice Trade Association (2004). Spice adulteration, White paper. ASTA, New York.

Anonyms (2009). The trade in spices. [WWW document]. URL http://www.herbs-spices.net/spicetrade.html. 31 July 2009.

AOAC (2005). Adulterants in Spices (Method 916.01). In: *AOAC Official methods of Analysis.* Gaithersburg, Maryland, USA: Association of Official Agricultural Chemists International.

Baeten, V., Von Holst, C., Garrido, A., Vancutsem, J., Renier, A.M. & Dardenne, P. (2005). Detection of banned meat and bone meal feedstuffs by near-infrared microscopic analysis of the dense sediment fraction. *Analytical and Bioanalytical Chemistry*, **382**, 149 – 157.

Bhattacharjee, P., Singhal, R.S. & Achyut, S.G. (2003). Supercritical carbon dioxide extraction for identification of adulteration of black pepper with papaya seeds. *Journal of the Science of Food and Agriculture*, **83**, 783-786.

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

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

Cocchi, M., Durante, C., Foca, G., Marchetti, A., Tassi, L. & Ulrici, A. (2006). Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta*, **68**, 1505-1511.

Dhanya, K. & Sasikumar, B. (2010). Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices. *Current Trends in Biotechnology and Pharmacy*, **4**, 454-489.

ElMasry, G. & Sun, S. (2010). Meat quality assessment using a hyperspectral imaging system. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by D. Sun). Pp. 175 – 240. London, UK: Elsevier.

European Spice Association (2007). European Spice Association quality minima document, adopted at the business and technical meeting on 2nd November 2007. ESA, Germany.

Farag Zaied, S.E.A., Aziz N.H. & Ali A.M. (1996). Comparing effects of washing, thermal treatments and gamma; irradiation on quality of spices. *Nahrung*, **40**, 32-36.

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

Gómez-Sanchis, J., Gómez-Chova, L., Aleixos, N., Camps-Valls, G., Montesinos-Herrero, C., Moltó, E. & Blasco, J. (2008). Hyperspectral system for early detection of rottenness caused by Penicillium digitatum in mandarins. *Journal of Food Engineering*, **89**, 80-86.

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

Hernández-Hierro, J.M., García-Villanova, R.J. & González-Martín, I. (2008). Potential of near infrared spectroscopy for the analysis of mycotoxins applied to naturally contaminated red paprika found in the Spanish market. *Analytica Chimica Acta*, **622**, 189-194.

Juliani, H.R., Kapteyn, J., Jones, D., Koroch, A.R., Wang, M., Charles, D. & Simon, J.E. (2006). Application of near-infrared spectroscopy in quality control and determination of adulteration of african essential oils. *Phytochemical Analysis*, **17**, 121-128.

Lavine, B.K. & Workman, J.J. (2005). Chemometrics: Past, present, and future. In: *Chemometrics and chemoinformatics* (edited by B. K. Lavine). Pp. 1-13. Washington, DC: American Chemical Society.

Ma, H. & Anderson, C.A. (2008). Characterization of pharmaceutical powder blends by NIR chemical imaging. *Journal of Pharmaceutical Sciences*, **97**, 3305-3320.

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

Pruthi, J.S. (1980). *Chemical aspects – Analytical Techniques.* In: *Spices and condiments: chemistry, microbiology, technology*, 4th ed. (edited by C. O. Chichester, E. M. Mrak & G. F. Stewart). Pp 69 – 149. New York: Academic Press.

Qin, J. & Lu, R. (2006). Hyperspectral diffuse reflectance imaging for rapid, noncontact measurement of the optical properties of turbid materials. *Applied Optics*, **45**, 8366-8373.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Baverage Analysis*. P. 227. London: Longman scientific & Technical.

Reid, L.M., O'Donnell, C.P. & Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science & Technology*, **17**, 344-353.

Robert, P., Bertrand, D., Devaux, M. F., & Sire, A. (1992). Identification of chemical constituents by multivariate near-infrared spectral imaging. *Analytical Chemistry*, **64**, 664-667.

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

Rosengarten, F. *Introduction*. In: *The book of spices*. P. 10. Wynnewood: Livingston

Sagoo, S.K., Little, C.L., Greenwood, M., Mithani, V., Grant, K.A., McLauchlin, J., de Pinna, E. & Threlfall, E.J. (2009). Assessment of the microbiological safety of dried spices and herbs from production and retail premises in the United Kingdom. *Food Microbiology*, **26**, 39-43.

Schultz, R.A., Nielsen, T., Zavaleta, J.R., Ruch, R., Wyatt, R. & Garner, H.R. (2001). Hyperspectral imaging: A novel approach for microscopic analysis. *Cytometry*, **43**, 239-247.

Wang, N. & ElMasry, G. (2010). Bruise detection of apples using hyperspectral imaging. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by D. Sun). Pp. 295 – 321. London, UK: Elsevier.

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

Wehling, R.L. (2003). Near-Infrared Spectroscopy. In: *Food Analysis Laboratory Manual* (edited by S. S. Nielsen). Pp 392 – 399. New York: Kluwer Academic/Plenum Publishers

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.

Wu, D., He, Y., Shi, J. & Feng, S. (2009). Exploring near and midinfrared Spectroscopy to predict trace iron and zinc contents in powdered milk. *Journal of Agricultural and Food Chemistry*, **57**, 1697-1704.

Chapter 2

Literature review

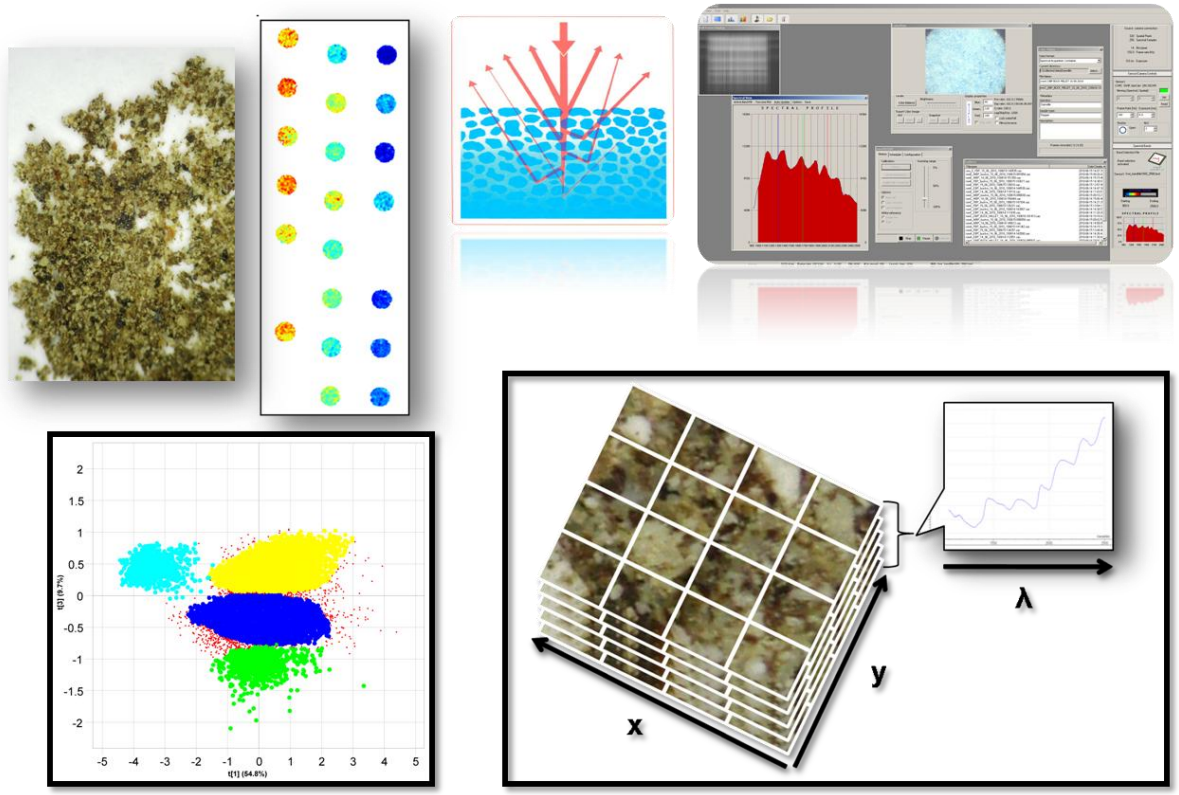


Table of contents

Literature review

1. Introduction

Value assessments of food and other agricultural commodities of plant origin rely on authenticity testing and adulterant detection (ASTA, 2004; Wilhelmsen, 2006; Dhanya & Sasikumar, 2010). Routinely analysed commodities include legumes, beverages, olive oil, fruit products, traded medicinal plant material and spices. Routine evaluation of these foodstuffs enables fraudulent practices to be identified and eradicated, thereby assuring consumer protection (Dhanya & Sasikumar, 2010). The adulteration of any foodstuff can either be intentional for economical gain or incidental due to negligence or improper facilities (ASTA, 2004; Wilhelmsen, 2006).

Powdered and whole spices have been subjected to both intentional and incidental adulteration (Krug, 1902; Woodman, 1941; Singhal & Kulkarni, 1990). Most spices are harvested in third world countries where available facilities and sanitation practises are not as advanced as those from developing and developed nations and thus allowing for incidental adulteration (Tainter & Grenis, 2001). Whole spices are often harvested directly from the jungle or hills where they grow; this in effect creates the opportunity for perpetrators to perform intentional adulteration. For example, whole black pepper is usually adulterated with papaya seeds, while starches from cheaper sources are added to ground black pepper (Krug, 1902; Woodman, 1941; Hartman, *et al.*, 1973; Tremlova, 2001; Bhattacharjee, 2003; ASTA, 2004; ISI; 2005).

Different methods for the detection of adulteration exist, including macroscopic and microscopic visual evaluation (Woodman, 1941; Pruthi & Kulkarni, 1969; AOAC, 2005), high performance liquid chromatography (HPLC) (Gonzalez *et al.*, 2003), thin layer chromatography (TLC) (Yin *et al.*, 2005), gas chromatography mass spectroscopy (GC MS) (Bhattacharjee *et al.*, 2003), liquid chromatography mass spectroscopy (LC MS) (Mazetti *et al.*, 2004) and enzyme linked immunosorbent assays (ELISA) (Stephan & Vieths, 2004). These methods are accurate and sensitive, but industrial applicability is hampered by cost and the need for specialist training.

Conventional or bulk near infrared (NIR) spectroscopy has become an important analytical tool for authenticity testing of foodstuffs and agricultural commodities (Woodcock *et al.,* 2008). This is due to the ease of application, the non-destructive nature and rapidity associated with conventional NIR spectroscopy. Advances in digital imaging and optics have lead to the development of NIR hyperspectral imaging which enables chemical mapping (Panigrahi & Gunasekaran, 2001; Reich, 2005; Burger, 2006; Gowen *et al.*, 2007; Wang & Paliwal, 2007). Both these techniques require minimal sample preparation and allow large data sets to be acquired in a minimal amount of time. Both conventional NIR spectroscopy and hyperspectral imaging are promising techniques for the identification of adulterants in spices. This review will elaborate on the theoretical background of conventional NIR spectroscopy, NIR hyperspectral imaging and multivariate image analysis. It will

also present evidence of spice adulteration, research performed on the identification of spice adulterants and adulterant related studies specifically conducted on black pepper.

2. Spice adulteration

The detection of adulterants has been investigated since 1902, and recently different associations, such as the American Spice Trade Association (ASTA) and Indian Standards Institution (ISI), have identified (**Tables 2.1 & 2.2**), and set standards for the measurement and determination of adulterants in spices (Krug, 1902; ASTA, 2004; ISI, 2005; IISR, 2007). The Indian Institute of Spices Research (IISR) actively investigates spice genetics and technology.

Table 2.1 Possible adulterants of spices as described by ASTA

Product	Adulterant
Oregano	Foreign leaves, e.g. sumac, cistus. Non-compliant herbs, e.g. savoury, thyme, marjoram.
Saffron	Floral waste, artificial colour
Cinnamon	Coffee husks
Nutmeg	Coffee husks
Ground spices	Spent spices (defatted), starch, grains, hulls, oleoresins
Capsicums	Tomato skin, dextrose, monosaccharides, disaccharides, Sudan red and related dyes
Ground black and white pepper	Buckwheat and millet seed

Table 2.2 Possible adulterants of spices as described by the ISI, New Delhi

Product	Adulterant
Cumin seed, poppy seed, black pepper	Artificially coloured foreign seeds Grass seeds coloured with charcoal
Mustard seeds	Argenome seeds
Turmeric, chilly and curry powder	Colour
Black pepper	Papaya seeds and light berries
Ground spices	Powdered bran and saw dust
Coriander powder	Dung powder and common salt
Chillies	Brick powder grit, sand, dirt and filth
Turmeric powder	Starch of maize, wheat, tapioca, rice
Turmeric	Lead chromate and metanil yellow
Asafoetida (Heeng)	Soap stone, other earthy matter and chalk

The ASTA and ISI have recommended several methods of analysis for the identification of adulterated spices. These methods may be classified as density based, microscopic probing and chemical analyses. Density based methods, as the name suggests, rely on the differing densities

of spice and adulterant. For example, when papaya seed adulterated black pepper kernels are immersed in carbon tetrachloride (CCl_4) the pure black pepper kernels will settle while the black papaya seeds will float. Adulteration of ground spices may be identified by sprinkling the suspect sample into water; saw dust and powdered bran will float on the surface while the ground spices will settle. Identification of visual differences between adulterant and spice forms the basis of microscopic probing. Pure turmeric is yellow coloured, relatively large in size and has an angular structure when viewed microscopically. Added foreign starches appear colourless and small in size compared to pure turmeric starch. The final group of methods, chemical analyses, is varied and several examples are subsequently given. Turmeric, chilli and curry powder is extracted with petroleum ether and 13N sulphuric acid (H_2SO_4) added to the extract. The presence of added dyes will be indicated by the appearance of a red colour. When distilled water is added to the mixture and the red colour disappears, it can be concluded that the sample is not adulterated. Alternatively, turmeric can be ashed, dissolved in a 1:7 dilution of H_2SO_4 and filtered. The formation of a pink colour upon the addition of one or two drops of 0.1% diphenylcarbazide to the solution is indicative of the presence of lead chromate. With the addition of a few drops of concentrated hydrochloric acid (HCl) to the sample, a violet colour will appear instantly. If the sample is pure turmeric, this colour will disappear if the sample is diluted with water. If the colour does not disappear, metanil yellow is present. Coriander powder adulterated with salt can be detected through adding a few drops of silver nitrate ($AgNO_3$) to 5 mL of the sample. The adulteration is confirmed by the occurrence of a white precipitate.

2.1. Research into spice adulteration

Beyond these recommended methods of adulterant detection several studies have been conducted to survey the extent of the spice adulteration problem (Tripathi *et al.*, 2007; Dhanya *et al.*, 2007) and to assess the efficacy of new detection methods. J. S. Pruthi (1980) summarised the spice adulteration studies published between 1919 and 1979. The spices discussed include: asafoetida, caraway, coriander, dill, ajowan, fennel, cinnamon, cassia, cloves, chillies, paprika, black pepper, white pepper, saffron, turmeric and vanilla (**Table 2.3**). This review made apparent the need for simple, rapid, reliable and inexpensive analytical techniques for the detection of different adulterants in spices. In a more recent paper, Dhanya and Sasikumar (2010) reviewed spice adulteration studies completed prior to 2009; a selection of which will be briefly discussed.

The detection of spice poppy seeds adulteration (*Papaver somniferum*) with rajgeera (*Amaranthus paniculatas*) seeds has been investigated (Singhal *et al.*, 1990). In this study fat and squalene concentrations were evaluated to determine the presence and content of rajgeera seeds. Fat content decreases and squalene content increases with increasing rajgeera addition to poppy seeds. When rajgeera is puffed (at 185 – 195ºC) it produces a product with flavour attributes different to that of puffed poppy seeds, allowing rajgeera seeds to be distinguished from poppy seeds.

Table 2.3 A selection of spice adulteration studies

Spice	Adulterant	Technique used for analysis	Confirmation of adulterant presence	Reference
Vanilla extract	Foreign plant material	Paper chromatography	Foreign spots in chromatogram	Fitelson (1961)
Capsicum spices	Dehydrated red beets	Paper chromatography	Red to pinkish-red chromatogram	Schwien & Miller (1967)
		UV-Vis spectrophotometry	Difference in absorbance	
		Microscopic examination	Red particles (parenchyma)	
Poppy seeds (*Papaver somniferum*)	Rajeera (*Amaranthus paniculatas*) seeds	Analysis of fat and squalene contents (AOAC method)	Decrease in fat content Increase in squalene content	Singhal & Kulkarni (1990)
		Evaluation of puffing	Flavoured product	
Chilli and curry powder	Synthetic colours/dyes (metanil yellow, sudan I, sudan III and orange II)	Paper chromatography, UV – light (365 nm)	Fluorescent spots	Tripathi *et al.* (2007)

Table 2.3 ... **(*continue*)**

Study	Adulterant	Technique	Confirmation of adulterant presence	Reference
Turmeric powder	*Curcuma zedoaria* powder	RAPD*	Adulterant specific bands	Sasikumar *et al* (2005)
Powdered spices	Plant based adulterants	SCAR#	Reference to specific SCAR marker	Dhanya (2009)
Whole black pepper	Light berries	Starch determination	Variation in starch content	Mitra *et al.* (1966)
Whole black pepper	Papaya seeds	Flotation test	Floating papaya seeds (density difference)	Pruthi & Kulkarni (1969)
		Microscopic examination	Visual confirmation of floaters	
Whole black pepper	Papaya seed	GC	Presence of benzyl glucosinolate	Curl & Fenwick (1983)
Black pepper powder	Ground papaya seed	TLC	Fluorescent bands observed at 366 nm at R_f 0.172 and R_f 0.943	Paradkar *et al.* (2001)
Black pepper powder	Ground papaya seeds	SCAR#	Reference to specific SCAR marker	Dhanya *et al.* (2009)

*Random amplified polymorphic DNA

#Sequence characterised amplified regions

The use of synthetic colours in foodstuffs has been surveyed (Tripathi *et al*., 2007). The non-permitted colours, associated with the adulteration of chilli and turmeric powder, were identified as a mixture of Sudan I and Sudan III, and orange II and metanil yellow respectively. This study served as evidence that adulteration with non-permitted colours occurred in some loose/non-branded products in both rural and urban market sectors even though regulatory surveillance was in place (Tripathi *et al*., 2005).

Di Anibal *et al.* (2009) proposed UV-visible spectroscopy with the inclusion of multivariate techniques as a method for the identification of Sudan dyes (I, II, III and IV) in commercial spices. These commercial spices include turmeric, curry, mild paprika and hot paprika. They experimented with three different classification techniques, namely *K*-nearest neighbour (KNN), soft independent modelling of class analogy (SIMCA) and partial least squares discriminant analysis (PLS-DA). The classification results were 99.3%, 96.3% and 90.4% correct for PLS-DA, KNN and SIMCA, respectively.

The adulteration of chilli powder (*Capsicum annum*) with plant based adulterants (dried red beet pulp, almond shell dust and powdered *Ziziphus nummularia* fruit) has been investigated using a random amplified polymorphic DNA (RAPD) technique (Dhanya *et al.*, 2007). One sample out of the six market samples analysed amplified the *Ziziphus nummularia* specific band, indicating the occurrence of adulteration in market samples. All tested market samples were free from almond shell dust and dried red beet pulp adulteration.

One of the most common adulterants of whole black pepper is dried papaya seeds (*Carica papaya*), since they resemble dried whole black pepper kernels in size colour and shape (Dhanya & Sasikumar, 2010). Whole black pepper kernels are also substituted with wild Piper berries (*P. attenuatum, P. galeatum*), dried fruits of West Indian Lantana (*Lantana camara*) and False Black Pepper (*Embelia ribes*). Alternatively, whole black pepper can be adulterated with exhausted pepper, pinheads (dried under developed black pepper berries), stem and chaff of black pepper. The adulteration of ground black pepper with coloured starches from cheaper sources has been reported.

3. NIR spectroscopy

Vibrational spectroscopic techniques, including NIR spectroscopy, are based on interactions between electromagnetic radiation and vibrational modes of covalently bound molecules (Herschel, 1800; Osborne *et al.*, 1993; Pasquini, 2003; Reich, 2005; Walsh & Kawano, 2009). The spectral range 700 to 2500 nm (14 300 to 4000 cm^{-1}) is examined when working with NIR spectroscopy. In this region overtone and combination vibrational modes impart chemical and physical information (Osborne *et al*., 1993; Walsh & Kawano, 2009). Overtones occur when a molecule is vibrationally excited from the ground state to the

second or higher vibrational energy level. A combination is created when two or more vibrational modes are excited simultaneously. The vibrations observed in the NIR region involve primarily C-H, O-H, N-H and S-H bonds (Pasquini, 2003). The observed peaks in a NIR spectrum are usually broad and overlapping.

When matter is exposed to NIR radiation the radiant energy can be transmitted, absorbed and reflected, or reflected from the surface of the material (Osborne *et al.*, 1993). The latter phenomenon is termed specular reflectance and offers no or little information of the analysed sample. The vibrational modes of molecules within a sample will absorb specific frequencies of radiant energy so that the profile of light incident on a sample is different to that of absorbed and reflected, or transmitted light (Pasquini, 2003).

3.1. Instrumentation

Various instrumental configurations for NIR measurements exist: these include wavelength selection methods (liquid crystal tunable filters and acousto optical tunable filters), dispersive optics-based instruments (grating monochromators), diode array systems and interferometric systems. A Fourier transform system, i.e. Fourier transform NIR (FT-NIR) (**Fig. 2.1**) was employed for all bulk NIR spectroscopic measurements presented in this thesis. The system utilises interferometry when obtaining data from a sample (Gunasekaran & Irudayaraj, 2001). Interferometry is based on the interference of two or more radiation beams after passing through different optical paths (Osborne *et al.*, 1993; Gunasekaran & Irudayaraj, 2001). The Michelson interferometer forms the basis of the FT-NIR design. Fourier transform instruments provide increased sensitivity, permit a high energy throughput and improve the rapidity of spectral acquisition**.**

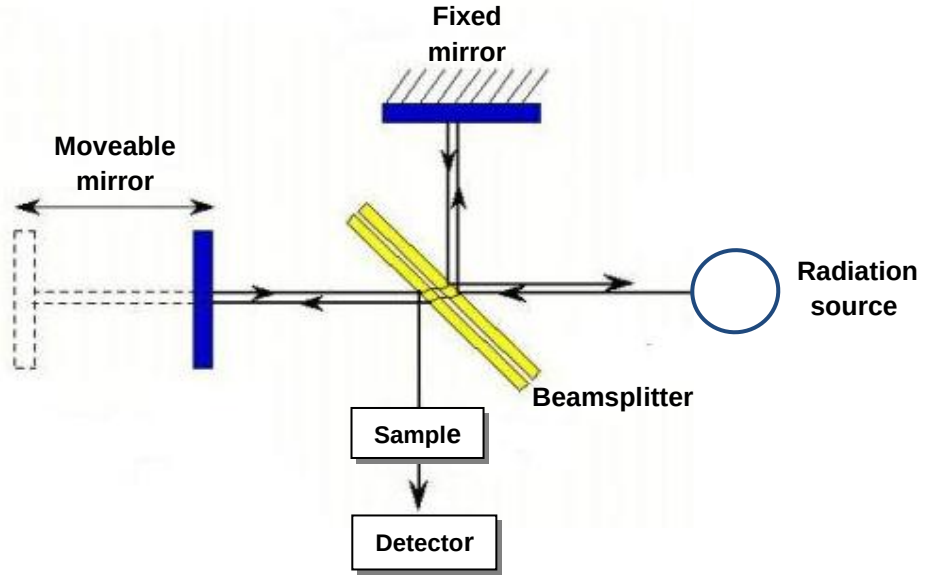


Figure 2.1 Schematic illustration of the Michelson interferometer action (Gunasekaran & Irudayaraj, 2001)

3.2. Applications

3.2.1. General applications

Applications of NIR spectroscopy can be found as early as the 1950s (Benson, 1993). In a review by McClure (2003) it was mentioned that application of this technique appeared dormant until the 1950s and only 91 publications are available on the CNIRS database for the period 1800 to 1950. It was only in 1965 when Karl Norris and David Massie made a breakthrough with their work on the spectral reflectance and transmittance properties of grain, that NIR spectroscopy was seen as a valid analytical tool.

NIR spectroscopy has been applied to foods for qualitative and quantitative analyses. The American Association of Cereal Chemists (AACC) International has adopted this technique for measuring protein in barley, oats, rye, triticale and wheat (AACC, 2009a). Wheat flour hardness estimation (AACC, 2009b) and measurements from wheat whole kernels for protein determination (AACC, 2009c) have been approved as official methods of analysis. NIR spectroscopy has also been successfully applied to poultry, fish, red and processed meats for the measurement of moisture, protein and fats (Wehling, 2003). On–line application is another area where NIR spectroscopy has proven its capabilities (Benson, 1993). On–line moisture measurement of cereals, flour, milk powders, starch powders and cocoa powders has been performed using NIR spectroscopy.

NIR spectroscopy has been used as a method to confirm authenticity for several food commodities (Reid *et al.*, 2006). These food products included fruit purées and juices (Rodriguez-Saona *et al.*, 2001; Contal *et al.*, 2002), maple syrup (Paradkar *et al.*, 2002), honey (Downey *et al.*, 2004), echinacea root (Laasonen *et al.*, 2002), milk powder (Maraboli *et al.*, 2002) and fishmeal (Murray *et al.*, 2002). In the differentiation of wines on the basis of grape variety, Cozzolino *et al.* (2003) found that NIR spectroscopy yielded correct classification levels of up to 100%.

3.2.2. Applications specific to spices

NIR spectroscopy has been employed for the rapid evaluation and quantitative analysis of thyme, oregano and chamomile essential oils (carvacrol, thymol and α-bisabolol) (Schulz *et al.*, 2003). The experiment was performed in reflectance mode and using the spectral range between 1100 and 2500 nm; all volatile oil contents could be successfully predicted (R^2 > 0.97). Standard errors for main essential oil components were within the same range as for the applied GC method.

The characterisation of peppercorn, pepper oil and pepper oleoresin, using vibrational spectroscopy methods, has been investigated (Schulz *et al.*, 2005). NIR, attenuated total reflectance (ATR) – infrared (IR) and NIR-FT-Raman spectroscopies were used to identify piperine (a pungent alkaloid within pepper) and pepper oleoresins. Piperine Raman signals

were well resolved in the Raman spectra of pepper and related oleoresins. This allowed Raman mapping of the whole green berry and dried peppercorn to determine piperine distribution. NIR measurements were calibrated by GC measurements to create a prediction model for essential oil concentrations in ground black and white pepper.

NIR-FT-Raman micro-spectroscopic mapping was employed to investigate the identification of secondary metabolites in medicinal and spice plants (Baranska *et al*., 2004). The potential of NIR-FT-Raman micro-spectroscopic mapping to identify the microstructure and chemical composition of fennel fruits, curcuma roots and chamomile inflourescence was demonstrated. Through studying microscopic Raman maps the authors could successfully identify anethole, the main essential oil component of fennel fruits. Anethole is distributed through the whole mericarp with the highest concentration at the fruit's top. Spiroethers were identified to be accumulated in the middle part of the flower after studying chamomile inflourescence Raman images. Curcumin distribution could be clearly identified inside the curcuma root whereas the core root contained the highest curcumin concentration.

The potential of NIR spectroscopy for analysing red paprika powder naturally contaminated with mycotoxins (aflatoxin B_1, ochratoxin A, total aflatoxin) was investigated (Hernández-Hierro *et al.*, 2008). Aflatoxin B_1 was the mycotoxin predicted with the greatest accuracy. The occurrence of aflatoxin B_1 is indicative of general mycotoxin contamination as it was present in all cases of mycotoxin contamination. NIR spectroscopy has the potential as an alternative to conventional methods for mycotoxin detection.

With advances in optics and digital imaging a more advanced technique that incorporates both NIR spectroscopy and imaging has been developed. This technique is called NIR hyperscpectral imaging and has shown its potential use in the pharmaceutical industry, defense forces and agricultural industry. In addition mid-infrared (MIR) spectroscopy, focusing on the MIR region (4000 – 400 cm^{-1}) of the electromagnetic spectrum, also poses great advantages for the analysis of adulterated spices. This technique is based on the same vibrational principles then NIR spectroscopy, but is focused on the fundamental vibrations occurring in the MIR region and therefore well resolved peaks can be obtained (Osborne *et al.*, 1993). MIR spectroscopy has been used in authentication studies of oils (Yang & Irudayaraj, 2001), powdered coffee (Downey *et al.*, 1997) and recently for the monitoring of red wine fermentation (Di Egidio *et al.*, 2010). The technique has also been found useful in the determination of structural characteristics of lignocellulosic biomass (Adapa *et al.*, 2009). Comprehensive work on the characterisation of peppercorn, pepper oil and pepper oleoresin has also been performed using MIR spectroscopy in conjunction with other vibrational spectroscopy techniques (Schulz *et al.,* 2005). The identification of various plant substances in the MIR region have also been confirmed (Schulz & Baranska, 2007).

4. NIR hyperspectral imaging

NIR hyperspectral imaging (NIR HSI) is applied in various scientific fields and is known by names including chemical or spectroscopic imaging, NIR imaging and spectral imaging (McClure, 2003; Garini *et al.*, 2006; Gowen *et al.*, 2007). NIR hyperspectral imaging integrates conventional imaging and NIR spectroscopy to attain spatially resolved spectral information from an object (Reich, 2005; Burger, 2006; Gowen *et al.*, 2007; Wang & Paliwal, 2007). Hyperspectral imaging was originally developed for remote sensing applications and has found its way to the spectroscopist laboratory environment through fusion with NIR spectroscopy (McClure, 2003; Gowen *et al.*, 2007).

NIR HSI poses various advantages when compared to conventional analytical methods. These include no physical contact and it enables non-destructive measurements from a sample (McClure, 2003; Burger, 2006; Gowen *et al*., 2007). NIR hyperspectral imaging has been applied to wheat, maize, corn and cucumber whereas hyperspectral imaging in the UV/Vis/NIR region has been applied to foodstuffs including apples, cucumbers, citrus fruit, poultry, cherries and maize. Selections of these applications are detailed in **Table 2.4**. NIR HSI has also been applied to animal feed and feed ingredients with great success (Fernández Pierna *et al.*, 2004; Von Holst, 2008). This technique also shares a great deal of advantages and disadvantages with NIR spectroscopy (**Table 2.5**).

4.1. Instrumentation

A NIR hyperspectral system is equipped with an illumination source, an imaging optic, a spectral encoder (for wavelength selection) and a focal plane array (FPA) detector. The sample (adulterated spice in this thesis) is illuminated with NIR radiation by a radiation source and a diffuse reflected image is collected with an imaging optic. The configuration of the imaging optic is dependent on sample size and type.

The FPA detector (e.g. InGaAs or HgCdTe) records a series of images in the NIR region at each wavelength selected by a spectral encoder. Liquid crystal tunable filters (LCTF) and interferometers are typical spectral encoders utilised during NIR hyperspectral analysis (Burger & Geladi, 2005). The data collected (sets of images at different wavelengths) is a three dimensional data set, known as a hypercube, comprising of two spatial dimensions (*x* and *y*) and one spectral dimension (*λ*) (Burger & Geladi, 2005; Reich, 2005). Each pixel contains the spectroscopic signature for the selected region (**Fig. 2.2**). When a sample is imaged with a hyperspectral imaging system a large amount of data is attained slice by slice creating a three dimensional block of data or hypercube (**Fig. 2.2**). Hypercube dimensions can vary from 320 x 256 pixels with 118 wavebands (a total of 9 666 560 data points) to 320 x 992 pixels with 242 wavebands (a total of 76 820 480 data points), depending on the type

of hyperspectral system used. This amount of data complicates data handling and requires specific approaches (i.e. hyperspectral image analysis) to extract relevant information.

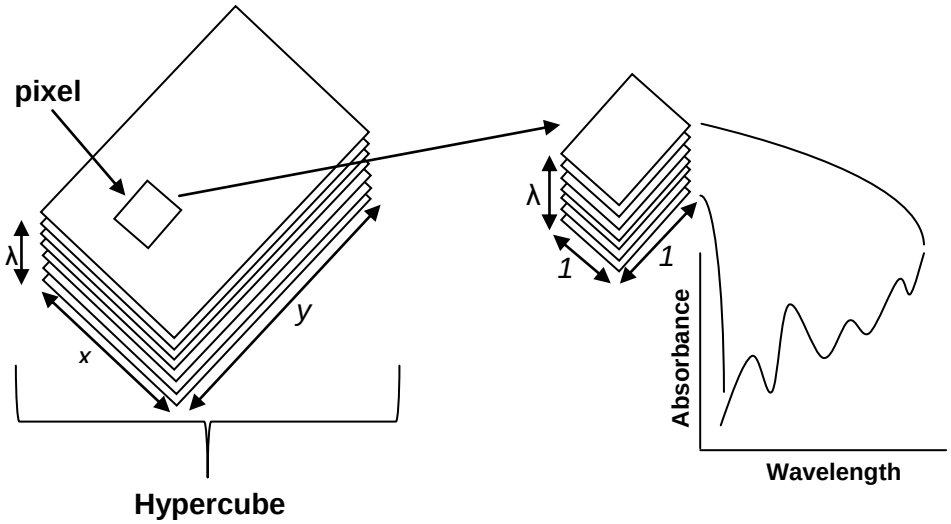


Figure 2.2 Digital acquired hypercube with pixel selected showing spectroscopic signature (Gowen *et al.*, 2007).

Table 2.4 Summary of NIR hyperspectral imaging applications of foodstuffs

Mode	Product	Wavelength (nm)	Study at hand	Reference
Reflectance	Apple	500 - 950	Bruise detection	Xing *et al.* (2007)
	Corn	950 - 1700	Prediction of oil and oleic acid concentration in individual kernels	Weinstock *et al.* (2006)
	Cucumber	900 - 1700	Bruise detection	Ariana *et al.* (2006)
	Pork	430 - 1000	Pork quality and marbling level assessment	Qiao *et al.* (2007)
	Poultry	430 – 850	Contamination detection on carcasses	Lawrence *et al.* (2006)
	Strawberry	650 - 1000	Firmness testing	Tallada *et al.* (2006)
	Animal feed	1100 – 2500	Detection of meat and bone meal	Von Holst *et al.* (2008)
Transmittance	Cherries	450 - 1000	Pit detection	Qin & Lu (2005)
	Cucumber	450 – 950	Internal damage	Ariana & Lu (2006)

Table 2.5 Comparison of NIR conventional spectroscopy and NIR hyperspectral imaging based on analytical abilities and limitations

Remark	NIR conventional spectroscopy	NIR hyperspectral imaging
Rapid and accurate analysis	✓	✓
Non-destructive analysis	✓	✓
Determination of multiple components from single sample spectrum	✓	✓
Avoidance of chemical reagents	✓	✓
Enable mapping of constituents (spatial distribution)	×	✓
Sensitivity to minor constituents	×	✓
Availability of large amounts (ca. 78 000 000 spectra) of data for analysis	×	✓
Large up-front cost	✓	✓
Instrument calibration	✓	✓
Capability of creating sufficient quantitative models	✓	✓
Illumination complications	×	✓

4.2. Image acquisition

Images can be acquired in one of three ways: a point-to-point spectral scan, a line-by-line spatial scan or pushbroom method, and wavelength scanning or staring imager method (Burger, 2006; Wang & Paliwal, 2007). The latter two imaging methods have been applied in food quality and safety inspection (Wang & Paliwal, 2007). Relative movement between camera and sample is a requirement when applying the pushbroom method (**Fig. 2.3**) for imaging (Reich, 2005). Spatial information is captured line-wise, while the spectral information for each pixel is projected along the perpendicular axis of the two-dimensional detector plane. Spectral information is attained by dispersive optics, linear variable filters or a combination of a digital micro-mirror array with a grating. This system configuration was employed for all NIR HSI measurements presented in this thesis.

With the staring imager method (**Fig. 2.4**) single images are recorded for each wavelength (selected by various filters) while the sample and camera are kept stationary (Reich, 2005). In a point-to-point scan a diffuse reflectance measurement is taken at a single spot on the sample and the complete spectrum is obtained (Burger, 2006). A new spectrum is obtained by repositioning the sample. Tuneable filters or an imaging Fourier transform spectrophotometer provides the spectral information of the two-dimensional image plane (Reich, 2005). Sample repositioning is time consuming and influences measurement repeatability since high demands are placed on repositioning hardware (Burger, 2006). With the staring imaging method, the analysed sample is exposed to heavy heat load from the illuminating source, since the sample remains stationary.

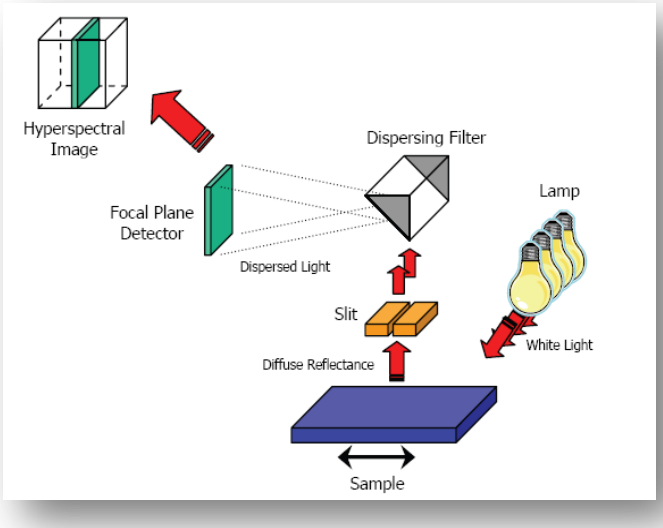


Figure 2.3 Schematic representation of the line-scan/pushbroom configuration (Burger, 2006).

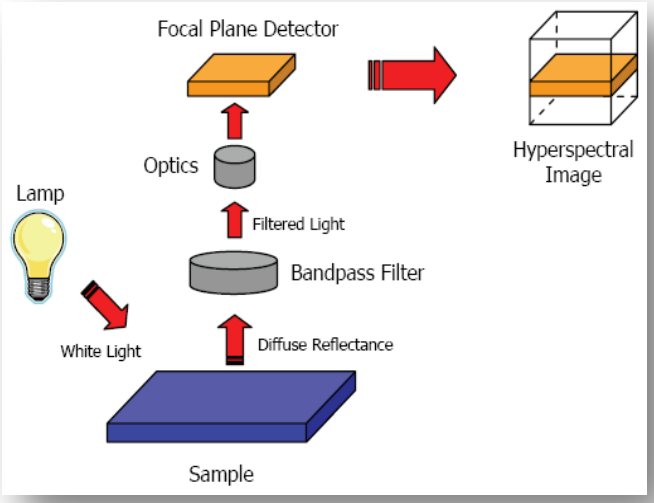


Figure 2.4 Schematic representation of the focal plane scan/staring imaging configuration (Burger, 2006).

When conventional NIR spectroscopy or NIR hyperspectral imaging is considered as analytical tool for experimental analysis, it poses both limitations and advantages to the user. With the addition of imaging, NIR hyperspectral imaging creates better opportunity in understanding sample behaviour. Although still costly, this technique is very promising and with further developments can reveal many phenomena unknown to the researcher. **Table 2.5** provides the reader with a brief summary of existing limitations and advantages of both these analytical techniques.

5. Chemometrics

Chemometrics is used to mathematically reduce the dimensionality of spectral data obtained from both conventional NIR spectroscopic and NIR hyperspectral imaging (NIR HSI) experiments (Tatzer *et al*., 2005). This statistical discipline allows for better understanding of chemical constituent behaviour through exploration and classification of mathematically reduced data (Beebe *et al.*, 1998). Sample spectra collected in the NIR region contain both physical and chemical information (Ozaki *et al.*, 2007). Extraction of this information from NIR spectra is complicated by multicollinearity and multiple sources of spectral noise originating from light scattering and instrument defects. Multicollinearity is when different variables (e.g. wavelengths) correlate within a specific data set (Næs *et al.*, 2002). Extraction of chemical and physical information from spectral data sets is possible through the application of spectral chemometric and preprocessing techniques.

When spice samples whether adulterated or not, are analysed by means of either conventional NIR spectroscopy or NIR HSI a large amount of relevant information is encapsulated inside the obtained data sets. This information includes hidden patterns, outliers or trends amongst variables (Cozzolino *et al.*, 2009). The recognition of these data structures becomes difficult once a matrix

exceeds three variables, but become more apparent when multivariate chemometric techniques such as principal component analysis (PCA), partial least squares (PLS) and PLS discriminant analysis (PLS-DA) are applied to the data matrix. These multivariate data analysis methods will be discussed fully in the remaining paragraphs.

5.1. Principal component analysis (PCA)

Principal component analysis (PCA) is an unsupervised clustering technique used for data visualisation and pattern recognition (Wold *et al.*, 1987; Geladi *et al.*, 2004; Roggo *et al.*, 2007; Cozzolino *et al.*, 2009). PCA reduces the dimensionality of a multivariate data set. Newly formed variables (principal components, PCs) are linear functions of the original variables (absorbance at each wavelength) and each consecutive PC encapsulates a different part of the total variation, since each is perpendicular to the other. Each PC consists of scores and a loading counterpart, where the scores indicate association between samples and the loadings illustrate how variables relate. Loadings of each PC are presented as a line plot and the scores, as a scatter plot. The simultaneous evaluation of these plots ensures the successful interpretation of measured spectroscopic data.

When highly collinear variables exist in a data set often using an alternate coordinate system allows better visualisation of the available information (Geladi *et al.*, 2004; Lavine & Workman, 2004). In PCA this new coordinate system is based on variance, where the first PC is along the direction of maximum variance in the original variables (**Fig. 2.5**) and each consecutive PC is along the direction of maximum remaining variance. Mean-centering can also be applied to the data matrix, to ensure optimal description of the analysed samples by new PCs (Geladi, 2003). With mean-centering the average of the data matrix is subtracted from each individual sample. The use PCA in analytical chemistry has been reviewed by Brereton (2000).

PCA is a bilinear modelling technique, thus before PCA can be applied to multivariate images (e.g. 3D block of data) the hypercube must be unfolded into a two dimensional data set or table (Geladi *et al.*, 1992; Geladi & Grahn, 2007). A multivariate image consists of three dimensions, where the first two are spatial (*x*,*y*) and the third is a spectral dimension (*λ*). A graphical representation of unfolding is presented in **Fig. 2.6.** The unfolding is in such a way that the samples or observations are the product of pixel coordinates (*x***y*). The number of variables for each observation corresponds to the number of recorded wavelengths (***λ***). The created two dimensional data table (observations x corresponding variables) can be subjected to PCA calculations.

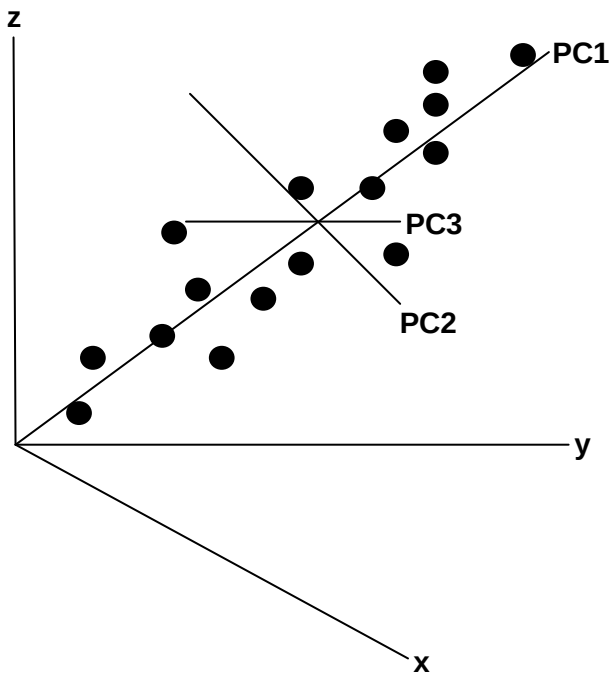


Figure 2.5 In the alternate coordinate system generated by PCA, samples may be defined by PC1 and PC2 rather than x, y and z (the original variables). PC3 here only encompasses noise in the data (Wold *et al.*, 1987).

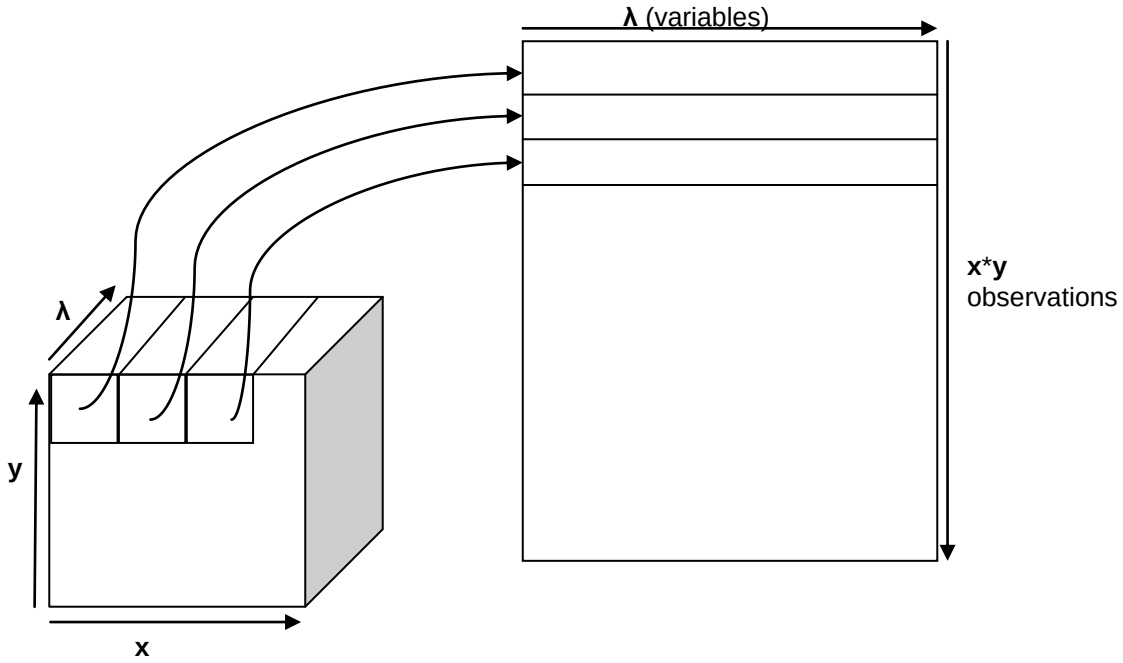


Figure 2.6 Unfolding of multivariate image (***x*y*λ***) into 2D block of data (Grahn & Geladi, 2007).

5.2. Partial least squares (PLS) regression

PLS utilises information from all wavelengths in the entire NIR spectrum to predict sample composition and is therefore a full spectrum method (Wold, 1982; Geladi & Kowalski, 1986; Brereton, 2000; Martens, 2001; Wold *et al.*, 2001; Wehling, 2003). PLS reduces the dimensionality of a large set of variables, by extracting a smaller amount of new variables explaining the maximum covariance or a linear link between spectral data **X** and reference data **Y** (Næs *et al.*, 2002). Both matrices (**X** and **Y**) are modelled to identify variables in the **X** space that will best describe the **Y** space. A regression equation can be developed using the new variables to predict sample constituents of a food.

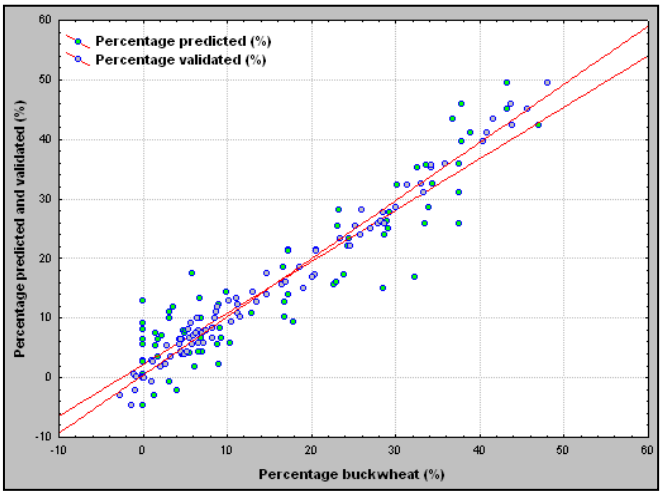


Figure 2.7 Regression model illustrating the predicted buckwheat content against the known buckwheat content. All bulk NIR measurements were made with a Perkin Elmer Spectrum IdentiCheck™ FT-NIR system (Wellesley, MA, USA).

5.3. Partial least squares discriminant analysis (PLS-DA)

This supervised pattern recognition technique is a variant of the previously described PLS regression, with the purpose of making qualitative assignation instead of predicting a qualitative parameter (Kasemsumran *et al.*, 2005; Liu *et al.*, 2008; Amigo *et al.*, 2009). With this technique the identified variation in the data set are correlated with class membership and therefore it is employed to discriminate between different classes (Liu *et al.*, 2008; Görlitz *et al.*, 2009). The creation of the PLS-DA model consists of two basic steps. Firstly, a conventional PLS model is built on group indicator variables after which the resultant observations are classified based on the group indicator variables (Liu *et al.*, 2008). For class membership identification a dummy matrix (**Y**), consisting of ones and zeros for classes, are paired with **X** spectral data (**Table 2.6**). The developed PLS-DA model can then be employed to classify new samples. This is done by predicting the X spectral data of each new sample and identifying how good the predicted data belong to each class Y (Kasemsumran *et al.*, 2005; Liu *et al.*, 2008; Görlitz *et al.*, 2009).

Table 2.6 Class membership identification based on black pepper and presence of adulterant

Classes		X Block	Y Block	
1	**Black pepper presence**	NIR spectral data	1	0
2	**Adulterant presence**		0	1

When NIR measurements are collected, different instrumental and light scattering effects can cause erroneous interpretation of the resulting data. For this reason it is helpful to mathematically pretreat or preprocess the data prior to PCA and PLS calculations.

5.4. Spectral preprocessing

Mathematical pretreatments, also known as preprocessing methods, are applied to spectral data to account for irregularities (scatter effects, instrumental drift, and uneven illumination) that give rise to unwanted variance in the data (Beebe *et al.*, 1998; Roggo *et al*., 2007). Preprocessing is used to reduce noise, enhance spectral resolution, reduce baseline variation and to normalize the data (Beebe *et al.*, 1998; Ozaki *et al.*, 2007). Ozaki *et al.* (2007) categorises preprocessing methods into four classes: noise reduction methods (e.g. smoothing), baseline correction methods (e.g. derivative methods), centering and normalisation, resolution enhancement (e.g. mean-centering). Noise and baseline correction methods will be discussed subsequently. Mean-centering was discussed previously in the PCA section whereas normalisation will be discussed with reference to standard normal variate (SNV) (Barnes *et al.*, 1989).

5.4.1. Noise reduction methods

High frequency noise results from instrument detector and electronic circuit interferences, whereas instrument drift during scanning causes low frequency noise (Ozaki *et al.*, 2007). Light scattering can also add to the noise found in captured data. Both high and low frequency noise contains little relevant information and can result in biased modelling. Low frequency noise usually resembles the real information in the data and is therefore difficult to reduce. The structure of typical noise in a NIR spectrum is illustrated in **Fig. 2.8**. Smoothing methods reduce random variation due to high frequency noise and therefore improve the signal-to-noise (S/N) ratio within a spectrum. Smoothing methods are window specific; the window is the spectral region used to define the smoothing function. The window width affects the resulting smooth, since each point in the window is utilised to determine the value at the center of the window (Beebe *et al.*, 1998; Ozaki *et al.*, 2007).

Various smoothing methods exist, these include moving average, running median, mean smoother and the most commonly applied Savitzky-Golay method (Beebe *et al.*, 1998; Ozaki *et al.*, 2007). Savitzky and Golay (1964) suggested that in the vicinity of a measurement point a

spectrum can be fitted by low-degree polynomials. Polynomials are fitted to each successive curve segment and therefore original values are replaced with values with more regular variation. This method is sensitive to polynomial order and with an increase of polynomial order band shape distortion will occur (Ozaki *et al.*, 2007). With this resulting distortion the spectral resolution is decreased, care must be taken to avoid distortion when applying smoothing methods.

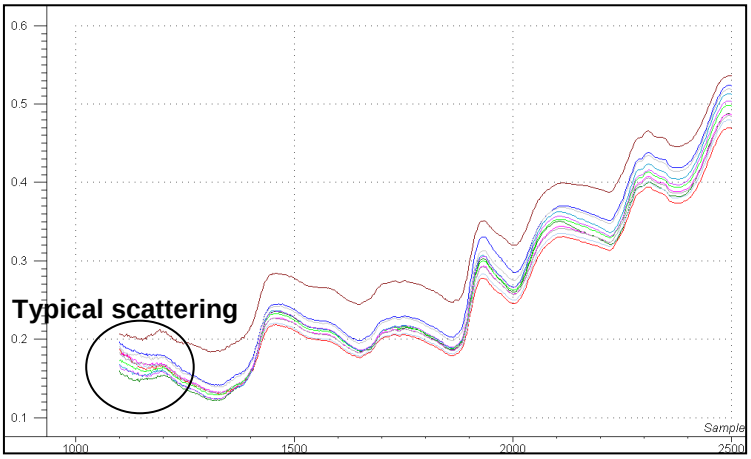


Figure 2.8 NIR spectra of samples before preprocessing showing the presence of noise.

5.4.2. Baseline correction methods

Baseline changes in the NIR spectrum can be due to a range of physical effects. Some of these factors include light scattering (e.g. illumination sources), particle differences (e.g. size) and influences of fibre optical cables (Ozaki *et al.*, 2007). There are different suggested methods for the reduction and elimination of these factors. Derivative methods and multiplicative scatter correction (MSC) are two of the methods commonly used for baseline correction.

First and second derivatives are the two most common forms of derivative preprocessing. The first derivative method is used for baseline correction whereas the second derivative is applied to account for additive and multiplicative baseline variation in the original spectrum. The MSC method regresses original spectral values of a sample against that of a reference spectrum (most often the mean spectrum) to identify both additive and multiplicative baseline variation. The wavelength dependence of light scattering is different from that of chemically based light absorbance; this forms the foundation of MSC (Geladi *et al.*, 1985). When raw spectroscopic data is preprocessed using MSC the resulting spectra resemble the original spectra simplifying data interpretation (Beebe *et al.*, 1998). Although MSC is an excellent technique for linearity in NIR spectroscopy, care should be taken when applying the technique, since a portion of the relevant information may be removed with the noise fraction (Yukihiro *et al.*, 2007). Similar to MSC treatment is a technique called standard normal variate (SNV). Through SNV treatment individual spectra are normalised without the use of the mean spectrum (reference spectrum) of any data set. Each spectrum is effectively centered on zero roughly varying between -2 and +2 on the vertical scale (Næs *et al.*, 2002).

Preprocessing methods are user specific and can be applied in any order followed by mean-centering, prior to PCA and PLS calculations. Calibration development and the validation thereof are of major concern when applying PLS regression, thus a comprehensive description follows.

5.5. Calibration development

Quantitative models strive to predict compositional properties of unknown samples using their NIR spectra. When calculating quantitative models accurate reference values for the samples are necessary, as the NIR spectral response is correlated with these values in a regression (Roggo *et al.*, 2007). Reference values are generally obtained via wet chemistry methods (e.g. HPLC, GC and spectrophotometric methods). Different sample sets are used in the development and validation of a quantitative model. The model is computed using the calibration set and the model's ability to predict is evaluated by the validation set. To ensure robustness of the model these two sets must be independent and must contain samples from different batches.

Cross-validation may be applied in cases where a separate sample set for validation is not available due to a limited number of samples. Leave-one-out cross-validation refers to the removal of one sample at a time (Næs *et al.*, 2002). The removed sample is then predicted by the model and the predicted value compared with the actual value. This procedure is applied to the whole data set until all the samples have been removed once. The prediction sum of squares (PRESS) is than calculated by adding the squared differences between predicted and observed values. The predictive capability of the model is measured by PRESS. These measures are employed to assess the robustness of the created models and are only two of a few other measures for model assessment.

The root mean square error of prediction (RMSEP, **eq. 2.1**) is an indication of the goodness of the prediction. Similarly, the root mean square error of cross-validation (RMSECV) is an estimate based on cross-validation and is calculated in the same manner as RMSEP ($\hat{y}_i$ is substituted with $\hat{y}_{CV,i}$.; $\hat{y}_{CV,i}$ is the estimate for the reference value, y_i, with sample *i* deleted). Other statistical measures such as a root mean square error of calibration (RMSEC), coefficient of determination (R^2), and bias are all indicators of how good the fit of a prediction model is (Næs *et al.*, 2002; Williams, 2007). The RMSEC is based on the validation of a calibration set whereas the RMSEP is specifically based on prediction testing. Care should be taken when considering the RMSEC as this only estimates model error and not the prediction error. Bias is defined as the mean difference between the measured (reference value, e.g. HPLC data) and predicted (based on NIR spectrum) values. Changes in source of raw materials, in processing conditions and in ambient temperature are a few causes of bias in a prediction model.

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \qquad \text{… 2.1}$$

Where $\hat{y}_i$ = predicted value

y_i = reference value

n = total number of samples

5.6. From multivariate image analysis to hyperspectral analysis

Esbensen and Geladi (1986) first introduced the strategy of multivariate image analysis (MIA) (Esbensen & Geladi, 1989; Grahn & Geladi, 2007), from this the hyperspectral imaging analysis (HIA) approach originated. The MIA strategy is based on the analysis of multivariate images which they describe as an array of pixels where each pixel is associated with a variable (waveband) (Esbensen & Geladi, 1989). The operations involved in the MIA classical approach are as follows:

1. The calculation of principal component scores (score images) and loadings (vectors).
2. Scatter plots of scores or loadings against each other.
3. Class selection on score plots.
4. Brushing of classes on multiple score plot.
5. Projection (transfer) of the pixels in the feature-space classes to the corresponding scene space location.
6. Calculation of local PC-models, as determined in the score plots.
7. Calculation of residual images with respect to such local models.
8. Auxiliary functions for overlay masking, overlay toggle and colour slicing.

The HIA approach is derived from the MIA classical approach, but has been simplified to make the image analysis more user-friendly. Evince multivariate image analysis software (Umbio AB, Umeå, Sweden) is design especially for HIA. This software is employed to perform HIA on corrected data. The following operations are involved in the HIA process:

1. *Image correction*

   Dark and white reference standards are employed for image correction and pseudo absorbance values are formed after the conversion of detector counts (**Fig. 2.9a**).

2. *Image cleaning*

   To start the HIA process a three PC PCA model is calculated by default, thereby generating score images, score plots and loading line plots of the hypercube. Created score images and score plots are of an interactive nature and are utilised to detect and discard unwanted information (background, shading errors, dead pixels, detector errors, detector saturation) (Geladi *et al.*, 2004; Grahn & Geladi, 2007). Unwanted pixels (highlighted in green, **Fig. 2.9b**) are identified by selecting certain regions inside the score plot after which the related regions

are projected onto the score image due to their interactive nature (Grahn & Geladi, 2007). Various PC combinations are probed to successfully identify and discard all unwanted pixels. Once all the unwanted pixels are removed a new PCA model is calculated revealing information previously obscured due to the inclusion of unwanted data.

3. *Addition of supplementary PCs*

   To further the search of unwanted data supplementary PCs are added to the PCA model, as a rule-of-thumb no more than six PCs per PCA model (**Fig. 2.9c**). Higher order PCs tend to describe excessive amounts of noise, thus complicating the extraction of relevant information. Apart from being noisy, higher components can also provide relevant information, thus the user is advised to probe all possible PC combinations.

4. *Application of preprocessing techniques if necessary*

   Preprocessing techniques may be applied to the data to improve overall data visualisation (**Figs. 2.9d-f**). When MSC was applied in this example to the cleaned image more unwanted pixels (highlighted in black, **Fig. 2.9d**) were identified and removed. After the application of an appropriate preprocessing method better cluster identification may be achieved (**Fig. 2.9d-e**).

5. *Probe PCs for possible pixel clustering*

   The scores of different PCs are plotted against each other for the identification of possible clustering of data points (**Fig. 2.9e**). Through this procedure the components contributing to the cluster formation can be established and with *a priori* knowledge of the analysed samples one can establish which samples caused the identified clustering. Biological samples usually consist of similar chemical compounds, thus overlapping of clusters may occur. The duality of score plot and image is once again used to identify region inside the score image (**Fig. 2.9g**). With the selection of identified clusters in the score plot corresponding regions in the score image will be highlighted (**Fig. 2.9g**).

6. *Identify and determine clusters using score plots and images interactively*

   Clusters can be labelled since *a priori* knowledge of the imaged sample exists (**Fig. 2.9h**).

7. *Classification of identified clusters in score plots and score images*

   Clusters in the score plot can be assigned to different classes (**Fig. 2.9h**) based on both similarities and differences in sample components, since the score plot and image are interactively linked, the score image will be changed into a classification image (**Fig. 2.9h**). As a result this will enable the mapping and distribution of the chemical compounds in the sample.

8. *Interpretation of loading line plots to establish correlations*

Loading line plots of each component are also formed when a PCA model is calculated. Once clusters are identified loading line plots can be used to identify crucial absorption peaks which relate to the chemical compounds responsible for spectral variation. Subsequently, absorption tables (Osborne *et al.*, 1993) can be employed to assign identified absorption peaks in the loading line plot to specific chemical compounds.

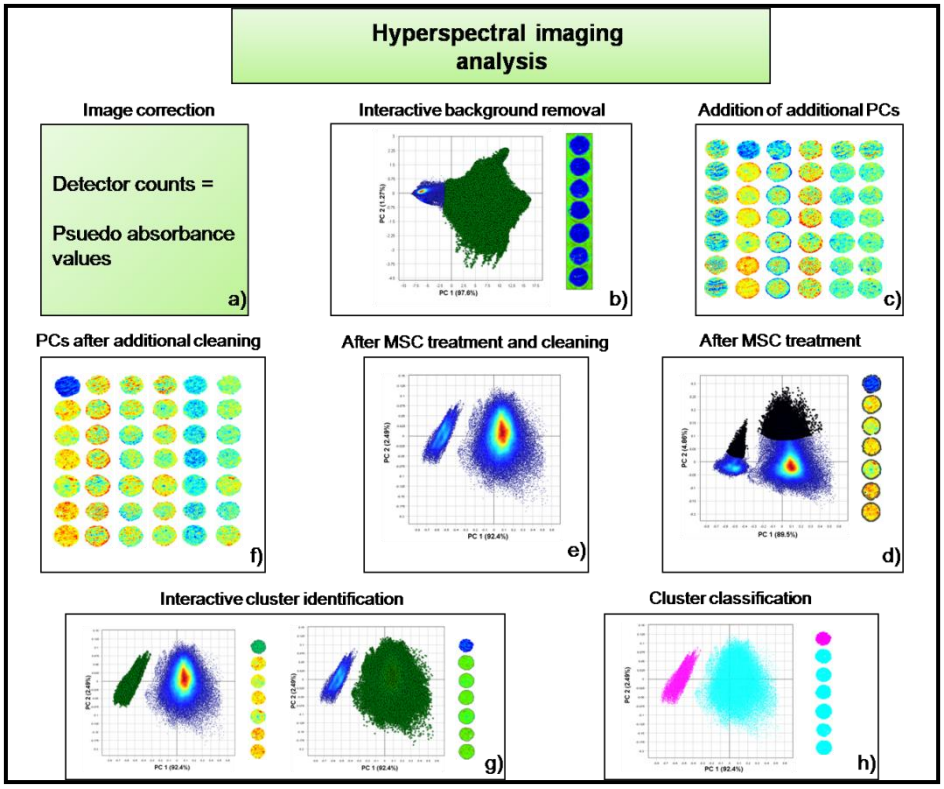


Figure 2.9 Diagrammatic representation of the typical steps involved during hyperspectral imaging analysis, **a)** image correction step, **b)** elimination of background and unwanted information, **c)** analysis of additional PCs, here presented as score images, **d)** identification of unwanted data (e.g. edge effects sample holders) after MSC treatment, **e)** clear visibility of separate clusters, **f)** PC score image after MSC treatment, **g)** selection of clusters in score plot and corresponding projection onto score image, **h)** classification plot and score image of identified clusters (pink cluster = pure millet flour; turquoise cluster = millet adulterated and unadulterated black pepper).

6. Conclusion

Established organisational bodies such as ASTA and ISI have recommended various methods for the analysis of spices. These methods are used in quality assurance laboratories in the commercial arena and on traditional scale (e.g. rural market). The methods are specifically developed to maintain good quality spices free from fraudulent admixtures. Through the analysis of spices various specifications/attributes were established for the purpose of authenticity

classification. These specifications ensure that quality and authenticity of spices are maintained. Fraudsters attempt to surpass these specifications utilising adulterants (e.g. starches, husks, ground material, leaves, spice species) to create spice blends with the same visual characteristics as the authentic spices, for economical gain. For this reason sensitive analytical techniques are crucial for spice analysis. Current spice analytical methods are sensitive and based on wet chemistry methodologies (e.g. HPLC, TLC) and microscopic analyses; these methods are laborious, expensive to maintain and require specialised personnel. NIR spectroscopy has shown great potential in quality and authenticity studies of food related products. NIR hyperspectral imaging has also been applied to food related products with great success.

7. References

AACC (2009a). AACC International Approved Methods of Analysis, 11th Ed. Method 39-10.01. Near-Infrared Reflectance Method for Protein Determination in Small Grains. Approved November 3, 1999. AACC International. St. Paul, MN, USA. Doi: 10.1094/AACCIntMethod-39-10.01.

AACC (2009b). AACC International Approved Methods of Analysis, 11th Ed. Method 55-31.01. Single Kernel Characterisation System for Wheat Kernel Texture. Approved November 3, 1999. AACC International. St. Paul, MN, USA. Doi: 10.1094/AACCIntMethod-55-31.01.

AACC (2009c). AACC International Approved Methods of Analysis, 11th Ed. Method 39-25.01. Near-Infrared Reflectance Method for Protein Content in Whole-Grain Wheat. Approved November 3, 1999. AACC International. St. Paul, MN, USA. Doi: 10.1094/AACCIntMethod-39-25.01.

Alonso, G.L., Salinas, M.R. & Garijo, J. (1998). Method to determine the authenticity of aroma of Saffron (*Crocus sativus* L.). *Journal of Food Protection*, **61**, 1525-1528.

Adapa, P., Karunakaran, C., Tabil, L. & Schoenau, G. (2009). Potential applications of Infrared and Raman spectromicroscopy for agricultural biomass*. Agricultural Engineering International: the CIGR Ejournal*, **11**, Manuscript 1081.

Amigo, J., 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.

Ariana, D. & Lu, R. (2006). Visible/near-infrared hyperspectral transmittance imaging for detection of internal mechanical injury in pickling cucumbers. In: *ASABE Annual International Meeting*, Paper No. 063039, July 2006.

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

AOAC (2005). Adulterants in spices (Method 916.01). In: *AOAC Official methods of Analysis.* Gaithersburg, Maryland, USA: Association of Official Agricultural Chemists International.

ASTA (2004) Spice adulteration, White paper. New York: American Spice Trade Association.

Baranska, M., Schulz, H., Rösch, P., Strehle, M. A. & Popp, J. (2004). Identification of secondary metabolites in medicinal and spice plants by NIR-FT-Raman microspectroscopic mapping, *Analyst*, **129**, 926.

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

Beebe, K.R., Pell, R.J. & Seasholtz, M.B. (1998). *Chemometrics: A Practical Guide.* Pp. 1-8, 26-55. New York, USA: John Wiley & Sons, Inc.

Benson, I.B. (1993). Compositional analysis using near infrared absorption spectroscopy. In: *Instrumentation and sensors for the food Industry* (edited by E. K. Rogers). Pp.121-166. London: Butterworth-Heinemann Ltd.

Bhattacharjee, P., Singhal, R.S. & Achyut, S.G. (2003). Supercritical carbon dioxide extraction for identification of adulteration of black pepper with papaya seeds. *Journal of the Science of Food and Agriculture*, **83**, 783-786.

Brereton, R.G. (2001). Introduction to multivariate calibration in analytical chemistry. *The Analyst*, **125**, 2125 – 2154.

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

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

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.

Cozzolino, D., Cynkar, W.U., Shah, N., Dambergs, R.G. & Smith, P.A. (2009). A brief introduction to multivariate methods in grape and wine analysis. *International Journal of Wine Research*, **1**, 123–130.

Cozzolino, D., Smyth, H.E. & Gishen, M. (2003). Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *Journal of Agricultural and Food Chemistry*, **51**, 7703-7708.

Curl, C.L. & Fenwick, G.R. (1983). On the determination of papaya seed adulteration of black pepper. *Food Chemistry*, **12**, 241-247.

Dhanya, K. & Sasikumar, B. (2010). Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices. *Current Trends in Biotechnology and Pharmacy*, **4**, 454-489.

Dhanya, K. (2009). Detection of probable plant based adulterants in selected powdered market samples of spices using molecular techniques. Ph.D thesis, Mangalore University, Mangalore, India. p. 251.

Dhanya, K., Syamkumar, S. & Sasikumar, B. (2009). Development and application of SCAR marker for the detection of papaya seed adulteration in traded black pepper powder. *Food Biotechnology*, **23**, 97-106

Dhanya, K., Kizhakkayil, J., Syamkumar, S. & Sasikumar, B. (2007). Isolation and amplification of genomic DNA from recalcitrant dried berries of black pepper (*Piper nigrum* L.) a medicinal spice. *Molecular Biotechnology*, **37**,165-168.

Di Anibal, C.V., Odena, M., Ruisánchez, I. & Callao, M.P. (2009). Determining the adulteration of spices with Sudan I-II-II-IV dyes by UV-visible spectroscopy and multivariate classification techniques. *Talanta*, **79**, 887-892.

Di Egidio V., Sinelli, N., Giovanelli, G., Moles, A. & Casiraghi, E. (2010). NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation, *European Food Research and Technology*, **230**, 947 – 955.

Downey, G., Fouratier, V. & Kelly, J. (2004). Detection of honey adulteration by addition of fructose and glucose using near-infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **11**, 447-456.

Downey, G., Briandet, R., Wilson, R.H. & Kemsley, E.K. (1997). Near- and Mid-Infrared Spectroscopies in food authentication: Coffee varietal identification. *Journal of Agricultural and Food Chemistry*, **45**, 4357-4361.

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

Fernández Pierna, J.A., Baeten, V. & Michotte Renier, A, Cogdill, R.P. & Dardenne, P. (2004). Combination of SVM and NIR imaging spectroscopy for the detection of MBM in compound feeds. *Journal of Chemmometrics*, **18**, 341 – 349.

Fitelson, J. (1961). The detection of foreign plant material in Vanilla extract, *Journal of the Association of Official Agricultural Chemists,* **44**, 531 – 534.

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

Geladi, P. & Kowalski, B.R. (1986). An example of 2-block predictive partial least-squares regression with simulated data. *Analytica Chimica Acta*, **185**, 19 -32.

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 imaging: Background and equipment. In: *Techiques and Applications of Hyperspectral Image Analysis* (edited by P.L.M. Geladi & H.F. Grahn). Pp. 1 - 16. West Sussex, England.: John Wiley & Sons.

González, M., Gloria Lobo, M., Méndez, J. & Carnero, A. (2005). Detection of colour adulteration in cochineals by spectrophotometric determination of yellow and red pigment groups. *Food Control*, **16**, 105-112.

Görlitz, L., Menze, B.H., Kelm, B.M., & Hamprecht, F.A. (2009). Processing spectral data. *Surface and Interface Analysis*, **41**, 636-644.

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

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

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

Grahn, H.F. & Geladi, P. (2007). *Techniques and applications of hyperspectral image analysis.* Pp. 1-15, 313-334. West Sussex, England: John Wiley & Sons Ltd.

Greule, M., Hänsel, C., Bauermann, U. & Mosandl, A. (2008). Feed additives: authenticity assessment using multicomponent-/multielement-isotope ratio mass spectrometry. *European Food Research and Technology,* **227**, 767-776.

Gunasekaran, S. & Irudayaraj, J. (2001). Optical methods: Visible, NIR, and FTIR spectroscopy. In: *Nondestructive Food Evaluation Techniques to Analyze Properties and Quality* (edited by S. Gunasekaran). Pp. 1 – 37. New York: Marcel Dekker, Inc.

Hartman, C.P., Divakar, N.G. & Rao, V.N.N. (1973). A study of identification of papaya seed in black pepper. *Journal of Food Science and Technology*, **10**, 43.

Hernández-Hierro, J.M., García-Villanova, R.J. & González-Martín, I. (2008). Potential of near infrared spectroscopy for the analysis of mycotoxins applied to naturally contaminated red paprika found in the Spanish market. *Analytica Chimica Acta*, **622**, 189-194.

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

IISR (2007). Spices News (edited by R. Dinesh, P. Rajeev & C. K. Sushamadevi). *IISR*, **18**, 1 – 12. Kerala, India: Indian Institute of Spice Research.

ISI (2005). Common adulterants/contaminants in food and simple screening tests for their detection (edited by P.K. Jaiswal). Nagpur, India: Indian Standards Institution.

Kasemsumran, S., Kang, N., Christy, A. & Ozaki, Y. (2005). Partial least squares processing of near-infrared spectra for discrimination and quantification of adulterated olive oils. *Spectroscopy Letters: An International Journal for Rapid Communication*, **38**, 839 - 851.

Krug, H. (1902). The analytical methods for carbohydrates as applied to foods and feeding stuffs. *Journal of the Franklin Institute*, **154,** 418 – 420.

Laasonen, M., Harmia-Pulkkinen, T., Simard, C.L., Michiels, E., Rasanen, M. & Vuorela, H. (2002). Fast identification of *Echinacea purpurea* dried roots using near-infrared spectroscopy. *Analytical Chemistry*, **74**, 2493-2499.

Lavine, B.K. & Workman, J.J. (2005). Chemometrics: Past, present, and future. In: *Chemometrics and chemoinformatics* (edited by B. K. Lavine). Pp. 1-13. Washington, DC: American Chemical Society.

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

Lewis, E.N., Dubois, J. & Kidder, L.H. (2007). NIR Imaging and its applications to agricultural and food engineering. In: *Near- Infrared Spectroscopy in food science and technology* (edited by Ozaki, Y, McClure, W.F. & Christy, A.A.). Pp. 121 – 131. New Jersey: John Wiley & Sons, Inc.

Liu, F., He, Y. & Wang, L. (2008). Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Analytica Chimica Acta*, **615**, 10-17.

Maraboli, A., Cattaneo, T.M.P. & Giangiacomo, R. (2002). Detection of vegetable proteins from soy, pea and wheat isolates in milk powder by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **10**, 63-69.

Marieschi, M., Torelli, A., Poli, F., Sacchetti, G. & Bruni, R. (2009). RAPD-based method for the quality control of mediterranean oregano and its contribution to pharmacognostic techniques. *Journal of Agricultural and Food Chemistry*, **57**, 1835- 1840.

Martens, H. (2001). Reliable and relevant modelling of real world data: a personal account of the development of PLS regression. *Chemometrics and Intelligent Laboratory Systems*, **58**:85–95.

McClure, W.F. (2003). Review: 204 years of near infrared technology: 1800 – 2003, *Journal of Near Infrared Spectroscopy*, **11**, 487 – 518.

Mitra, S.N., Roy, B.R. & Roy, A.K. (1966). Note on the importance of starch in the analysis of black pepper. *Journal and Proceedings of the Institute of Chemistry*, **38**, 215.

Murray, I., Aucott, A.S. & Pike, I. (2001). Use of discriminant analysis on visible and near infrared reflectance spectra to detect adulteration of fishmeal with meat and bone meal. *Journal of Near Infrared Spectroscopy*, **9**, 297-311.

Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). *A User Friendly Guide to Multivariate Calibration and Classification.* Pp. 39-54. West Sussex, UK: NIR Publications.

Noh, H. K. & Lu, R. (2007). Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. *Postharvest Biology and Technology*, **43**, 193-201.

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

Ozaki, Y., Morita, S. & Du, Y. (2007). Spectral analysis. In: *Near- Infrared Spectroscopy in Food Science and Technology* (edited by Ozaki, Y, McClure, W.F. & Christy, A.A.). Pp. 47 - 72. New Jersey: John Wiley & Sons, Inc.

Panigrahi, S. & Gunasekaran, S. (2001). Computer vision. In: *Nondestructive Food Evaluation Techniques to Analyze Properties and Quality* (edited by S. Gunasekaran). Pp. 39 – 98. New York: Marcel Dekker, Inc.

Paradkar, M.M. & Irudayaraj, J. (2002). Discrimination and classification of beet and cane inverts in honey by FT-Raman spectroscopy. *Food Chemistry*, **76**, 231-239.

Paradkar, M.M., Singhal, R.S., & Kulkarni, P.R. (2001). A new TLC method to detect the presence of ground papaya seed in ground black pepper. *Journal of the Science of Food and Agriculture*, **81**, 1322-1325

Paramita, B., Singhal, R.S. & Achyut, S.G. (2003). Supercritical carbon dioxide extraction for identification of adulteration of black pepper with papaya seeds. *Journal of the Science of Food and Agriculture*, **83**, 783-786.

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

Pruthi, J.S. (1980). Detection of adulteration. In: *Spices and Condiments: Chemistry, Microbiology, Technology* (edited by C. O. Chichester, E. M. Mrak & G. F. Stewart). Pp. 139 – 149. New York: Academic Press

Pruthi, J.S. & Kulkarni, B.M. (1969). A simple technique for the rapid and easy detection of papaya seeds in black pepper berries. *Indian Food Packer*, **23**, 51-52.

Qiao, J., Ngadi, M.O., Wang, N., Gariépy, C. & Prasher, S.O. (2007). Pork quality and marbling level assessment using a hyperspectral imaging system. *Journal of Food Engineering*, **83**, 10-16.

Qin, J., & Lu, R. (2005). Detection of pits in tart cherries by hyperspectral transmission imaging. *Transactions of the ASAE*, **48**, 1963-1970.

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

Reid, L.M., O'Donnell, C.P. & Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science & Technology* **17**, 344-353.

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

Rodriguez-Saona, L.E., Fry, F.S., McLaughlin, M.A. & Calvey, E.M. (2001). Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. *Carbohydrate Research*, **336**, 63-74.

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.

Sasikumar, B., Syamkumar, S., Remya, R. & John Zachariah, T. (2005). PCR based detection of adulteration in the market samples of turmeric powder. *Food Biotechnology*, **18**, 299-306.

Savitzky, A. & Golay, M. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, **36**, 1627-1639.

Schwien, W.G. & Miller, B.J. (1967). Detection and identification of dehydrated red beets in capsicum spices, *Journal of the Association of Official Agricultural Chemists,* **50**, 523 – 525.

Schulz, H. & Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy*, **43**, 13 - 25.

Schulz, H., Baranska, M., Quilitzsch, R., Schutze, W. & Losing, G. (2005). Characterization of peppercorn, pepper oil, and pepper oleoresin by vibrational spectroscopy methods. *Journal of Agricultural and Food Chemistry*, **53**, 3358-3363.

Schulz, H., Quilitzsch, R. & Krüger, H. (2003). Rapid evaluation and quantitative analysis of thyme, origano and chamomile essential oils by ATR-IR and NIR spectroscopy. *Journal of Molecular Structure*, **661-662**, 299-306.

Singhal, R. & Kulkarni, P.R. (1990). Detection of adulteration of the spice poppy seeds (*Papaver somniferum*) with *Amaranthus paniculatas* (Rajgeera) seeds, *Journal of Food Quality*, **13**, 375 – 381.

Smith, E.R., Samuel, A. and Mitchell, L.C. (1926). Detection of added papper-shells in pepper. *Journal of the Association of Official Agricultural Chemists*, **9**, 233.

Stephan, O. & Vieths, S. (2004). Development of a real-time PCR and a sandwich ELISA for detection of potentially allergenic trace amounts of peanut (*Arachis hypogaea*) in processed foods. *Journal of Agricultural Food Chemistry*, **52**, 3754-3760.

Tainter, D.R. & Grenis A.T. (2001). U.S. regulations as they apply to spices. In: *Spices and Seasoning: A Food Technology Handbook*, 2nd ed. p. 1. New York: John Wiley & Sons.

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

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

Tremlova, B. (2001). Evidence of spice black pepper adulteration. *Czech Journal of Food Scienc*es, **19**, 235-239.

Tripathi, M., Khanna, S.K. & Das, M. (2007). Surveillance on use of synthetic colours in eatables vis a vis Prevention of Food Adulteration Act of India. *Food Control*, **18**, 211-219.

Von Holst, C., Baeten, V., Boix, A., Slowikowski, B., Fernández Pierna, J.A., Tirendi, S. & Dardenne, P. (2008). Transferability study of a near-infrared microscopic method for the detection of banned meat and bone meal in feedstuffs. *Analytical Bioanalytical Chemistry*, **392**, 313 – 317.

Walsh, K.B. & Kawano, S. (2009). Near-Infrared Spectroscopy. In: *Optical Monitoring of Fresh and Processed Agricultural Crops* (edited by M. Zude). p. 193. New York: CRC Press

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.

Wehling, R.L. (2003). Near-infrared spectroscopy. In: *Food Analysis Laboratory Manual* (edited by S. S. Nielsen). Pp 392 – 399. New York: Kluwer Academic/Plenum Publishers

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.

Wilhelmsen, E.C. (2006). Adulteration determination. In: *Meyers, R. A. (Ed.). Encyclopedia of analytical chemistry: applications, theory and instrumentation,* (edited by R. A. Meyers). p. 14344New York: John Wiley and Sons, Inc

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

Wold, H. (1982). Soft modeling. The basic design and some extensions. In: *Systems Under Indirect Observation*. Pp. 1-53. Amsterdam: North-Holland

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

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

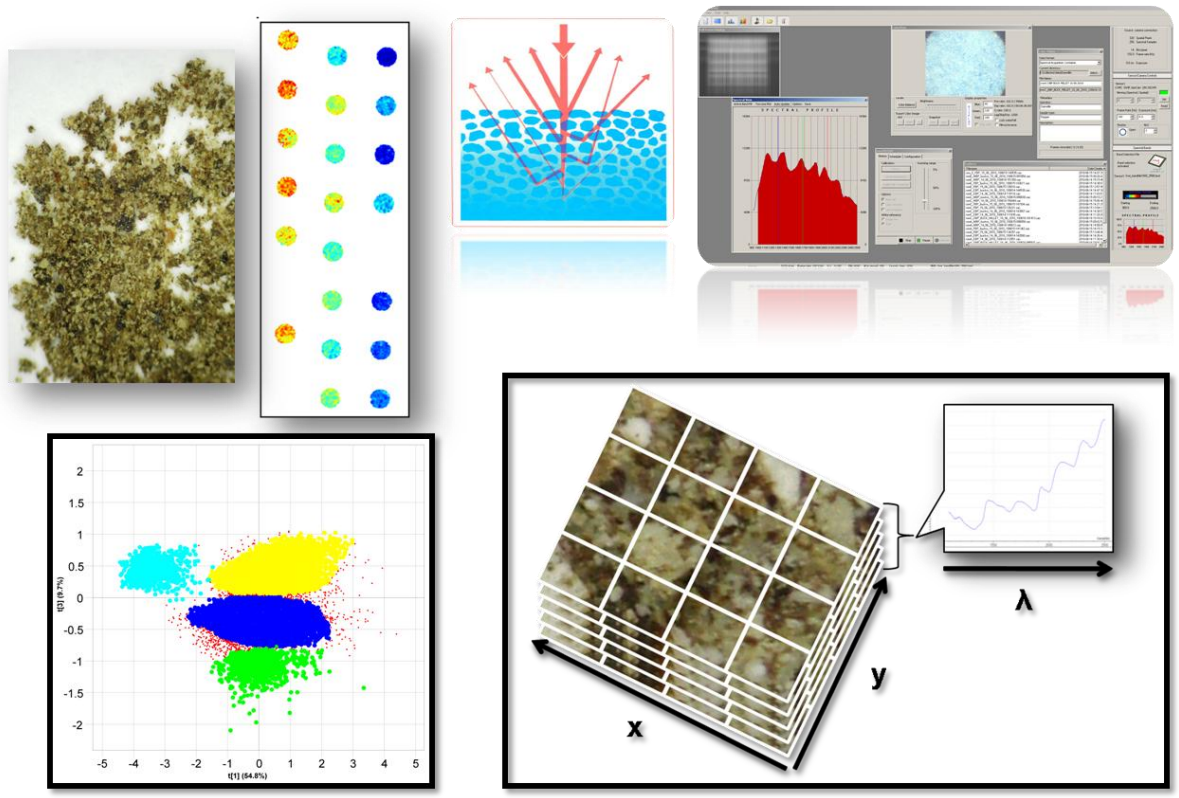
Woodcock, T., Downey, G. & O'Donnell, C.P. (2008). Better quality food and beverages: the role of near infrared spectroscopy. *Journal of Near Infrared Spectroscopy,* **16**, 1-29.

Woodman, A.G. (1941). Pepper. In: *Food Analysis: Typical Methods and the Interpretation of Results*, 4th ed. p. 396. New York: McGraw-Hill Book Company, Inc.

Xing, J., Saeys, W. & De Baerdemaeker, J. (2007). Combination of chemometric tools and image processing for bruise detection on apples. *Computers and Electronics in Agriculture*, **56**, 1-13.

Chapter 3

Investigating the presence of buckwheat and millet adulterant in ground black pepper by near infrared (NIR) hyperspectral imaging



Investigating the presence of buckwheat and millet adulterant in ground black pepper by near infrared (NIR) hyperspectral imaging

Abstract

Black pepper is known for its diverse uses and importance in the global spice trade. This creates reason for the intentional adulteration of black pepper by fraudsters for economical gain. The potential of near (NIR) hyperspectral imaging as a rapid method for detecting the presence of adulterants (buckwheat and millet flour) in ground black pepper was evaluated. A sisuChema short wave infrared (SWIR) pushbroom imaging system with spectral range of 1000–2498 nm was used for hyperspectral image measurements. Ground black pepper was adulterated with either buckwheat or millet flour in 5% (w/w) increments from 0–100%. Exploratory principal component analysis (PCA) was performed on absorbance images for removal of non–essential information (background, bad pixels and shading errors). Multiplicative scattering correction (MSC) was applied to the cleaned image. Consequently, mean-centering was applied and the PCA model was recalculated. An adulterant dependent gradient was recognised along PC1 this is possibly due to a difference in protein and oil presence of the adulterant and black pepper. Analysis of images (mosaics of buckwheat and millet adulterated black pepper) revealed that buckwheat adulterated samples were separated from millet adulterated samples along PC4 and the loading line plot of PC4 suggests that this separation was due to a protein difference. Subsequently, partial least squares-discriminant analysis (PLS-DA) was performed on both buckwheat and millet adulterated black pepper. PLS-DA models calculated for buckwheat and millet adulterated black pepper demonstrated accuracies of 70% and 77%, respectively.

Introduction

Black pepper (*Piper nigrum* L.) has been used in human diets, perfumery and for medicinal purposes and is considered to be the most widely used spice (Bhatacharjee *et al.*, 2003; Dhanya & Sasikumar, 2010). It is also crowned as the King of Spices, since it fetches the highest return amongst all spices (Bhatacharjee *et al.*, 2003). This confirms black pepper's important role in the global spice trade and creates a reason for the possible adulteration with inferior materials. Utilisation of adulterants in food is usually intentional to maximize revenues (ASTA, 2004; Dhanya & Sasikumar, 2010). Incidental adulteration may also occur due to ignorance, negligence or lack of proper facilities (ASTA, 2004).

Adulteration of black pepper with inexpensive or inferior material (e.g. pepper shells, nutshells, buckwheat and other cereals) has been reported (Woodman, 1941; Dhanya *et al.*, 2007). Microscopic examination and spectrophotometric techniques have been used for adulterant detection in black pepper (Woodman, 1941; Tremlova, 2001). Recently a polymerase chain reaction (PCR) based method has been developed for detection of papaya seeds in ground black

pepper (Dhanya, 2009). Although the mentioned techniques have been applied with success in identifying black pepper adulteration, each requires great expertise and is time-consuming.

Near infrared (NIR) reflectance spectroscopy complimented with chemometrics, has proven to be a successful analytical technique in various fields, including agriculture and food industries (Workman, 2001; Cen & Hie, 2007). This technique is based on attaining information from a substance by observing its absorption of NIR radiation and has been used in a number of food authentication studies (Siesler, 2002; Pasquini, 2003; Woodcock *et al.*, 2008). The rapidity, ease of use and non-destructive nature of NIR spectroscopy are the key characteristics making it an appropriate technique for studying biological material. Once spectral data are captured, chemometric methods such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) can be applied for dimensional reduction. These techniques reveal hidden patterns and sample associations (Wold *et al.*, 1987; Beebe *et al.*, 1998; Næs *et al.*, 2002; Geladi, 2003; Chevallier *et al.*, 2006; Cozzolino, 2009).

NIR hyperspectral imaging (HSI) is the fusion of NIR spectroscopy and digital imaging to attain both spatial (*x* and *y*) and spectral (*λ*) information from an object (Geladi *et al.*, 2004; Gowen *et al.*, 2007; Grahn & Geladi, 2007). With this technique a large amount of data (295 040 spectra) can be obtained from the imaged object. Apart from this the spatial feature also permits studying the location of chemical constituents (Garini, *et al.*, 2006; Grahn & Geladi, 2007; Gowen *et al.*, 2008). NIR HSI applications are just as diverse in food quality analysis as NIR spectroscopy, and it is increasingly applied to investigate food products (Wang & Paliwal, 2007; Elmasry & Sun, 2010; Ariana & Lu, 2010; Kamruzzaman, *et al.*, 2010; Wang & Elmasry, 2010).

In this study, the efficacy of NIR HSI for the identification of buckwheat or millet flour adulterated black pepper was assesed. PCA was employed to identify specific grouping of samples (adulterated black pepper, unadulterated black pepper and adulterants) and PLS-DA was applied for discrimination between established groups based on adulteration concentration.

Material and methods

Sample material

Four batches (1, 2, 3 and 4) of whole black pepper, kindly supplied by four manufacturers, millet and buckwheat kernels were milled to approximately 500 µm particle size using a Retsch mill (Retsch model ZM1: sieve with 500 µm hole width, Haan, Germany). All samples were milled to reduce particle size differences, which has been identified as a factor influencing NIR spectra. Ground black pepper was then adulterated with buckwheat or millet flour in evenly spread intervals between 0 and 100% (increments of 5% w/w) forming 19 adulteration levels, one unadulterated black pepper and one pure flour (millet or buckwheat). All ground samples were dried at 74°C for 2 hrs in a vacuum oven (Heraeus model RVT 360, Hanau, Germany) to reduce any variation caused by moisture content, cooled in a desiccator and transferred to clear sepcap vials (Kimble Glass Incorporated, New Jersey, USA: 15 x 45 mm). The given temperature was chosen to dry the

samples effectively without interfering with the chemical integrity of the samples. For the NIR HSI, the sepcap vials containing the adulterated black pepper samples, adulterant (buckwheat or millet flour) as well as unadulterated ground black pepper were packed into four eppendorf tube holders (Eppendorf AG, Hamburg, Germany: 6.5 cm x 21 cm; 80 holes). The vial content was transferred to 24 holes of the holder and imaged with a 100 mm lens providing a pixel size of 300 x 300 µm.

Hyperspectral imaging system

Images were acquired with a sisuChema SWIR pushbroom imaging system (Specim, Oulu, Finland) from 1000-2498 nm with 6.2 nm intervals and field of view of 6.5 cm x 21 cm producing images of 320(*x*) x 922(*y*) x 242(*λ*). Data acquisition, system control and data management were made possible by ChemaDAQ software (Specim, Oulu, Finland). The sisuChema consists of an imaging spectrograph coupled to a 2-D array Mercury-Cadmium-Telluride (HgCdTe) detector.

Eppendorf tube holders, containing the samples, were placed on top of the imaging conveyer belt (covered with black rubber pad) of the sisuChema imaging system. It was illuminated line-by-line, diffuse reflected light was focused onto the entrance slit, where after wavelengths were selected by the imaging spectrograph. Lines of the target were imaged onto the rows of the 2D array (spatial axis) and the spectrograph generated a spectrum for each pixel in the line along the spectral axis (**Fig. 3.1**). Theoretically, image correction transformation is performed according the following equation (**eq. 3.1**):

$$I_{\lambda,n} = -\log[(\frac{S_{\lambda,n} - B_{\lambda,n}}{W_{\lambda,n} - B_{\lambda,n}}) \times x \qquad \text{... } \mathbf{3.1}$$

Where:

- n = pixel index variable (n = 1...N)
- $I_{\lambda,n}$ = standardised absorbance intensity, pixel n, at wavelength λ
- $S_{\lambda,n}$ = sample image, pixel n, at wavelength λ
- $B_{\lambda,n}$ = dark reference image , pixel n, at wavelength λ
- $W_{\lambda,n}$ = white reference image , pixel n, at wavelength λ
- x = total reflectance

Through this process the obtained images are transformed from instrumental acquired reflectance counts to absorbance. This transformation is performed for correction of dark counts with a dark reference image ($B_{\lambda,n}$) subtracted from the raw image ($S_{\lambda,n}$). The latter are then divided by a total reflectance spectrum (x) of a white reference image ($W_{\lambda,n}$) subtracted from a dark reference image ($B_{\lambda,n}$). The explained process was automatically performed in Evince hyperspectral image analysis software version 2.4.0 (Umbio, Umeå, Sweden).

Multivariate image analysis

Prior to data preprocessing and image analysis, two single images (320 x 922 x 242) were combined at a time using Evince image analysis software to form mosaics (320 x 1844 x 242).

Principal component analysis (PCA)

Principal component analysis was applied to reduce the dimensionality of the obtained mosaics (590 080 spectra). Before any data interpretation, the detection and removal of bad pixels and background was first performed on the mosaics (Geladi *et al.*, 2004). This was done by delineating clusters in the PCA score plot with 3 principal components (PCs) and using the brushing technique between the score plot and corresponding score image.

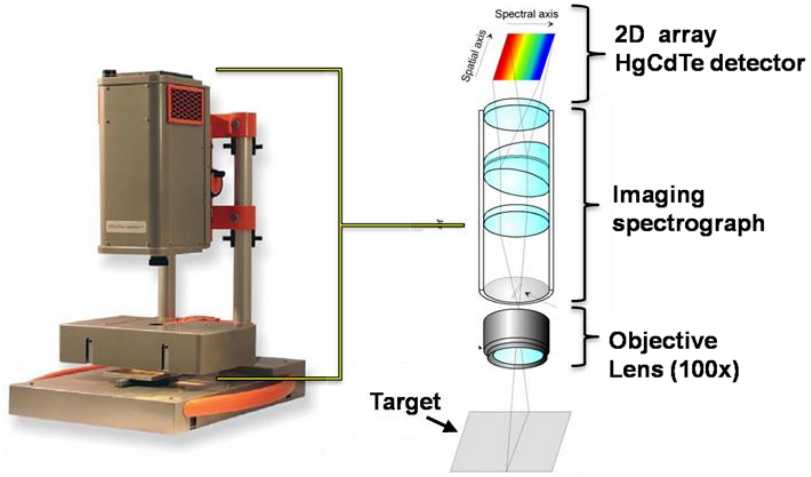


Figure 3.1 The sisuChema hyperspectral imaging instrument with the crucial components depicted.

The cleaned image was then subjected to a PCA with 6 PCs. Prior to this PCA calculation, multiplicative scatter correction (MSC) was applied to remove any remaining variation due to light scattering effects where after mean-centering was applied to center all data points for ease of interpretation (Geladi *et al.*, 1985; Beebe *et al.*, 1998; Næs *et al.*, 2002). Both MSC and standard normal variate (SNV) was separately applied to the cleaned mosaics. All interpretations were made on preprocessed data calculated with Evince imaging software.

Partial least squares discriminant analysis (PLS-DA)

Various PLS-DA models were developed to evaluate the membership classification based on percentage adulterant (buckwheat or millet) added. Prior to PLS-DA model development, MSC was applied. The following paragraphs explain the process involved in the development of PLS-DA models for buckwheat and millet adulterated black pepper.

Buckwheat adulterated black pepper

Different single images containing black pepper adulterated with buckwheat flour were combined and subsequently investigated. A mosaic contained 42 samples and was divided into a training and test set where each set contained 21 samples (1 = unadulterated black pepper; 1 = millet flour; 19 = adulterated black pepper). The samples in both sets were categorised into three different

classes based on the level of adulteration (A = 0 – 25%; B = 30 – 60%; and C = 65 – 100%. Each class was assigned a dummy variable, class A (1.0.0), class B (0.1.0) and class C (0.0.1), for discrimination. After PLS-DA model development the test set was predicted and the amount of pixels correctly classified in the test set was evaluated, to assess the model validity. The coefficient of determination (R^2) was also used to determine calibration model validity.

Millet adulterated black pepper

The same procedure was followed with the investigation of mosaics of black pepper adulterated with millet flour. A single mosaic, also consisting of 42 samples, was divided similarly into a training and test set. For the PLS-DA calculation each class was again assigned a dummy variable with 1.0.0, 0.1.0 and 0.0.1 for classes A, B and C respectively (A = 0 – 25%; B = 30 – 65%; C = 70 – 100%. The calibration model validity was assessed by predicting the test set, evaluating the amount of pixels correctly classified in the test set and also observing the resultant R^2. The amount of unclassified pixels also gives a good indication of how well the calibration model performed in predicting the test set.

Results and discussion

Principal component analysis

Even though PCA was applied on all mosaics (four) formed, only the results from one mosaic of buckwheat and millet adulterated black pepper will be discussed in depth. Only the results obtained after MSC treatment will be discussed, since this preprocessing technique performed better than SNV in identifying outlying data points. After MSC application and subsequent cleaning of the mosaic from one batch of black pepper separately adulterated with buckwheat and millet, PC score images (1 and 4) revealed pertinent sample associations. The remaining PC score images (2, 3, 5 and 6) comprised irrelevant (systematic error, scattering effects) or unexplained information and were therefore not included in the interpretations. PCs 1 and 4 explained 92.1% and 0.85% respectively of the total variance. Both PC1 and PC4 was combined with PC2 (2.68%) to improve visualisation of datal distribution in the PC score plots.

It was expected to identify clusters based on amount of adulterant present in ground black pepper, since black pepper was mixed with varying amounts of adulterant (buckwheat and millet). This, however, was not apperent after investigating the score plot of PC1 vs. PC2 (**Fig. 3.2a**), but the score image of PC1 indicated the presence of chemical variation amongst the samples (**Fig. 3.2b**). The score image illustrates the location of score values relative to the original image whereas the score plot demonstrates the position of score values in the multidimensional space.

a) b)

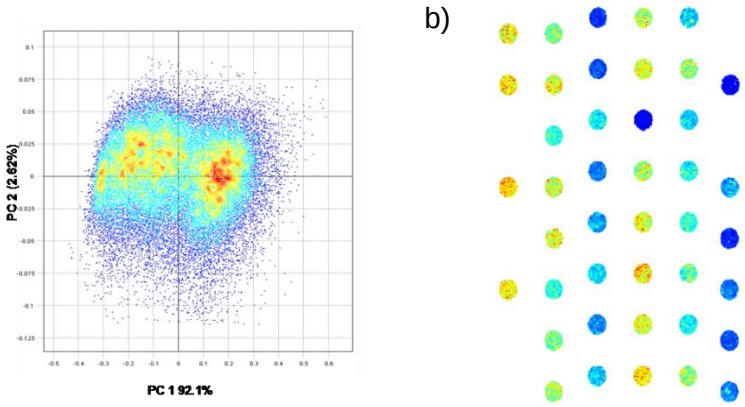


Figure 3.2 a) PC score plot (PC1 vs. PC2) and **b)** score image PC1 after MSC treatment indicating presence of differing chemical variation amongst adulterant present in black pepper (orange to light green = black pepper presence; light to dark blue = adulterant presence).

The different ground black pepper samples in the score images contain varying amounts of adulterant between 0 and 100%. When the score image of PC1 were studied the different samples portrayed differing colour intensities. The differing colour intensities in the score image (**Fig. 3.2b**) denotes the concentration of adulterant (light to dark blue) or black pepper (orange to light green) present in the samples. The score plot of PC1 vs. PC2 illustrated no clear clustering based on the specific amount of adulterant present , but it is well known that score plots can also be employed to study gradients (Geladi *et al.*, 1992). Four different areas were categorised along PC1 forming a classification plot (**Fig. 3.3a**) and with subsequent projection onto the score image, a classification image (**Fig. 3.3b**) was created. The interactive examination of the classification plot and image suggests that there is an adulterant concentration gradient present along PC1, starting from unadulterated black pepper and ending at the pure adulterants.

a) b)

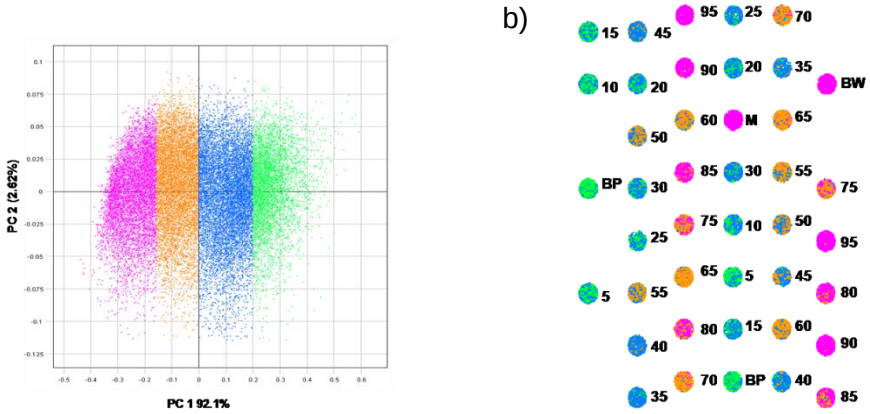


Figure 3.3 a) Classification plot (PC1 vs. PC2) with four different classes (green = 0 - 30% adulteration; blue = 20 – 45% adulteration; orange = 50 – 75% adulteration and pink = 80 -100% adulteration) and **b)** classification image with percentage adulterant present (BP = unadulterated black pepper; M = millet flour and BW = buckwheat flour).

The loading line plot of PC1 revealed three prominent positively loaded (2241, 2303 and 2347 nm) and only one (1461 nm) negatively loaded, wavelengths. The wavelength or absorption peak at 2241 nm (N-H stretching and NH^{+}_{3} deformation) can be associated with an amino acid, thus relates to protein, whereas the absorption peaks at 2303 nm (CH stretching and CH deformation) and 2347 nm (CH_2 symmetric stretching and $=CH_2$ deformation) are both associated with oil (Fassio *et al*., 2007; Williams, 2007; Osborne *et al.*, 1993). Protein and oil can be seen as the chemical descriptors of all the score values situated on the positive side of PC1 (**Fig. 3.4a**), thus relating to unadulterated black pepper and the samples containing a low adulterant content.

One negatively loaded wavelength, 1461 nm (N-H stretching first overtone), associating with a $-CONH_2$ structure (peptide) and relating to protein (**Fig. 3.4b**) was identified (Osborne *et al.*, 1993). Protein thus relates to all score values observed on the negative side of PC1, therefore relating to the pure adulterants and the samples containing higher amounts of adulterant. The loading line plot of PC1 indicated that the observed variation along PC1 (**Fig. 3.4a**) might be due to a protein structural difference between samples. In addition, oil presence also contributed to the observed variation (Davies & Grant, 1987; Bruun *et al*., 2007; Williams, 2007). Earlier it was mentioned that two clusters are present along PC1 and through further interactive investigation it was established that the noted clusters seen along PC1 are due to sample grouping based on the amount of adulterant (either high or low) present in black pepper.

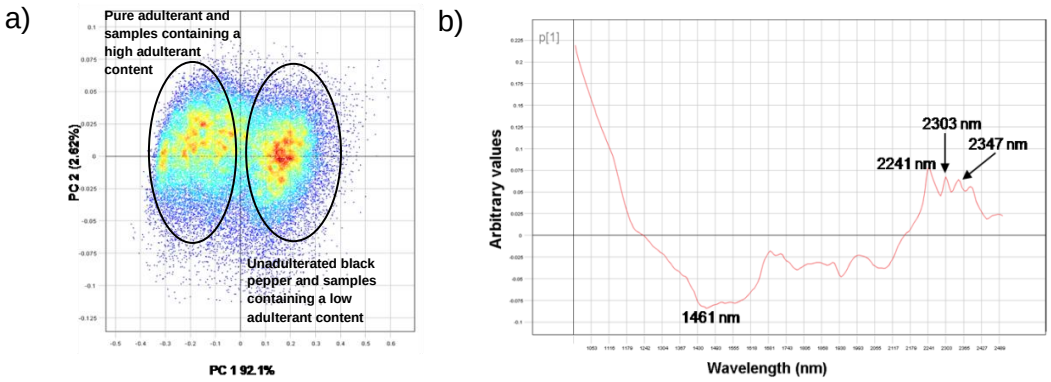


Figure 3.4 a) PC score plot of PC1 vs. PC2 with ellipses showing the type of samples (left = pure adulterant and samples containing a high adulterant content; right = unadulterated black pepper and samples containing a low adulterant content) predominantly found in that region along PC1 and **b)** loading line plot PC1 illustrating the prominent wavelengths.

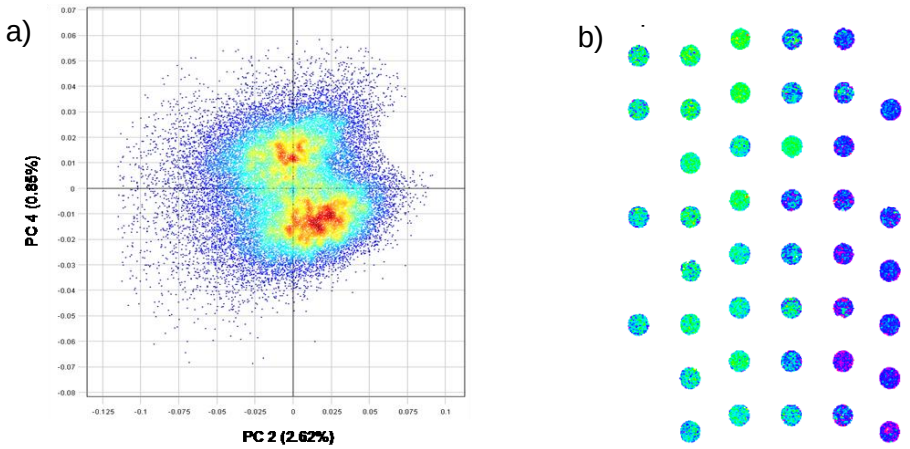


Figure 3.5 a) PC score plot of PC2 vs. PC4 showing two distinct clusters along PC4 and **b)** score image PC4 indicating sample differences (predominantly green = millet adulterated black pepper; predominantly blue = buckwheat adulterated black pepper).

When PC1 and PC4 were plotted against each other (not shown) no clear clusters were observed, but when PC2 was plotted against PC4 two distinct clusters were observed along PC4 (**Fig. 3.5a**). Upon investigation of the score image (**Fig. 3.5b**) a clear difference amongst the samples was visible. A classification plot (**Fig. 3.6a**) and image (**Fig. 3.6b**) were created to define the observed clusters (**Fig. 3.5a**) and differences (**Fig. 3.5b**). Having *a priori* knowledge of the samples it was possible to associate samples appearing predominantly red to millet adulterated samples and those appearing predominantly purple to buckwheat adulterated samples. The latter thus explains the observed differences present in the score image (**Fig. 3.5b**). Upon investigation of the classification image it became apparent that the unadulterated black pepper samples also appeared red. This suggests that there is an existing similarity between millet flour and ground black pepper. This might be due to the presence of similar chemical structures or physical particle size of both millet flour and ground black pepper.

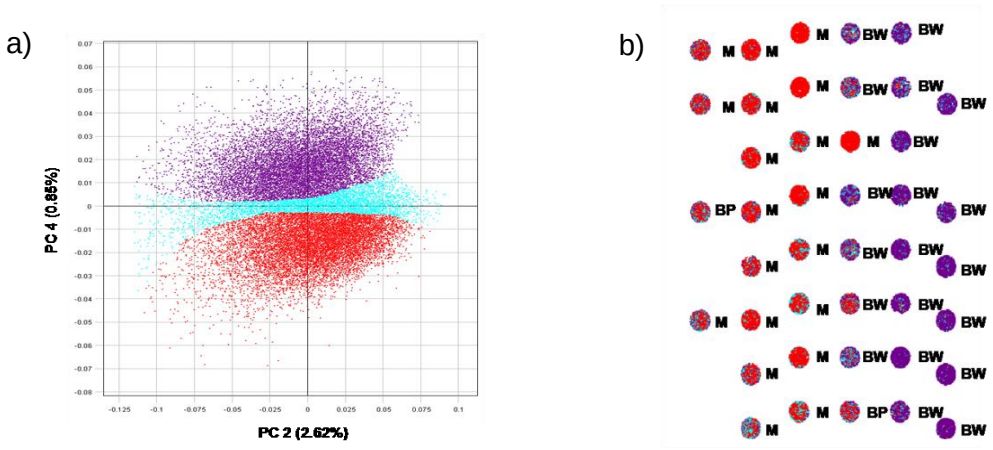


Figure 3.6 a) Classification plot illustrating two clusters (red = millet adulterated; purple = buckwheat adulterated) **b)** classification image depicting the location of pixels classified in the classification plot (BP = unadulterated black pepper; M = millet adulterated samples; BW = buckwheat adulterated samples).

The two clusters along PC4 can thus be ascribed to either millet adulterated samples or buckwheat adulterated samples and the loading line plot of PC4 (**Fig. 3.7b**) was studied to identify the chemical bonds responsible for the observed variation along PC4. Two prominent wavelengths (1955 nm: N-H asymmetric stretching and amide II and 1999 nm: N-H symmetric stretching and amide II) related to positive score values, were identified on the positive side of loading line plot PC4. Both these absorption peaks indicates the presence of a protein related structure, thus indicating the presence of protein. Other groups have also elucidated the presence of protein related structures at similar absorption peaks (Davies & Grant, 1987; Osborne *et al.*, 1993; Bruun *et al.*, 2007; Williams, 2007; Wu *et al.*, 2009).

Protein (structural/content) can then collectively serve as the chemical descriptor of all score values forming the cluster on the positive side along PC4 (**Fig. 3.7a**) or the buckwheat adulterated samples (**Fig. 3.7a**). Conversely, 2303 nm (CH stretching and CH deformation) was identified as the most prominent absorption peak on the negative side of the loading line plot. This peak previously assigned to oil in loading line plot PC1 gives us an indication of black pepper (Fassio *et al*., 2007; Williams, 2007).

The above mentioned gave us reason to believe that the identified clusters along PC4 are due to a difference in oil and protein presence. To substantiate the latter, an absorption peak at 2136 nm (N-H stretching and C=O stretching) was observed directly opposite 1999 nm (N-H symmetric stretching and amide II). The absorption peak at 2136 nm can be assigned to an amino acid and glutamine was previously assigned to an absorption peak at 2184 nm (Bruun *et al.*, 2007).

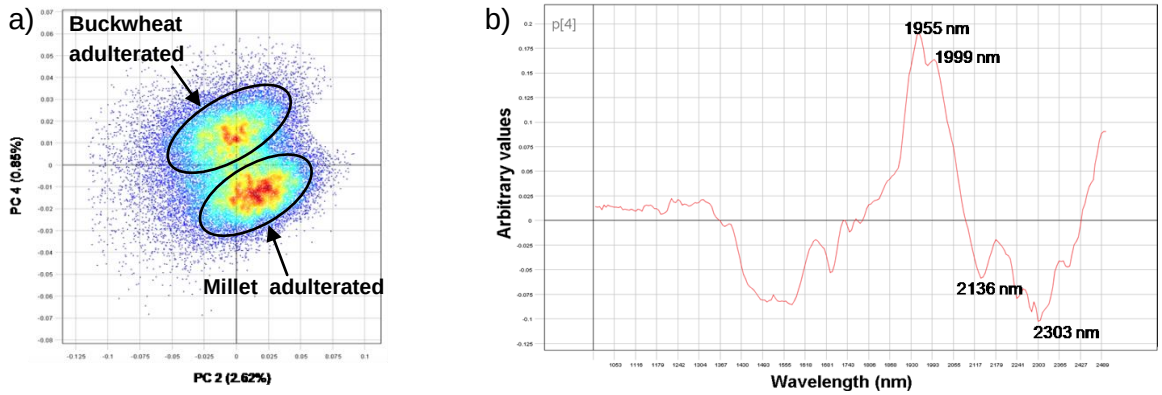


Figure 3.7 a) PC score plot (PC2 vs. PC4) with ellipses depicting buckwheat and millet adulterated black pepper along PC4 and **b)** their corresponding wavelengths in loading line plot PC4.

The separation illustrated in the score plot of PC2 and PC4 can either be due to protein structure differences or the amount of protein present in buckwheat and millet flour. A great deal of research has been conducted on millet and buckwheat flour protein structure and content (Prakash *et al.*, 1987; Bejosano & Corke, 1999; Kasaoka *et al.*, 1999). Millet flour is particularly low in lysine (amino acid) and the protein fractions present are usually albumins, globulins, prolamin and

glutelin. Conversely, buckwheat flour proteins are lysine rich and with globulin as its major protein fraction (Campbell, 1997). In addition, buckwheat protein contains no glutelin fractions whereas albumin and prolamins are present to a lesser extent. The biochemical characterization of rice glutelin has been studied extensively and results show that glutamine was one of the most abundant amino acids found in glutelin (Wen & Luthe, 1985). This gives us strong reason to believe that there are indeed known protein structural differences between millet and buckwheat flour. The protein content of buckwheat typically ranges between 9.7 – 15% (Prakash *et al.*, 1987; Bejosano & Corke, 1999) whereas the protein content of millet varies between 9.3 – 12.7% (Kasaoka *et al*., 1999). The above mentioned suggests that there are prominent differences between buckwheat and millet flour when referring to both protein content and protein structure. With application of PCA it was possible to identify data clustering, concentration gradients and chemical structures (protein and oil) responsible for the observed variation. The identification of chemical structures indicates that physical effects were successfully minimised.

Partial least squares discriminant analysis

Buckwheat adulterated black pepper

Black pepper from batches 1 and 2 adulterated with buckwheat flour were used in this investigation. Batch one adulterated with buckwheat demonstrated better separation between samples and was consequently used in the calibration process. The three remaining batches were then used as test sets. After studying the score image of PC1 (**Fig. 3.8a**) sample differences were recognised. The calibration image consisting of 21 samples was then divided into three classes based on the amount of buckwheat present in the black pepper samples (**Fig. 3.8c**). The classes (**Fig. 3.8b**) were created with reference to the score image, forming a classification image which was projected onto the score plot. After the classes was assigned and projected onto the score plot a classification plot (**Fig. 3.8b**) was obtained. Also note that there was overlapping present between the three classes (**Fig. 3.8b**). This will ultimately influence the prediction quality of the created PLS-DA model.

Subsequently the PLS-DA model was created explaining 70% of the Y variation after six PLS components. No further variation increase was observed after six PLS components, thus making the later adequate to explained the variation in the model (**Fig. 3.9**). In previous research performed on sample mixtures, eight PLS components were used to explain the Y variation, thus validating the obtained amount of PLS components in our research (Chevallier *et al.*, 2006). The created calibration model performance was validated by predicting a test set (**Figs. 3.10 a & b**).

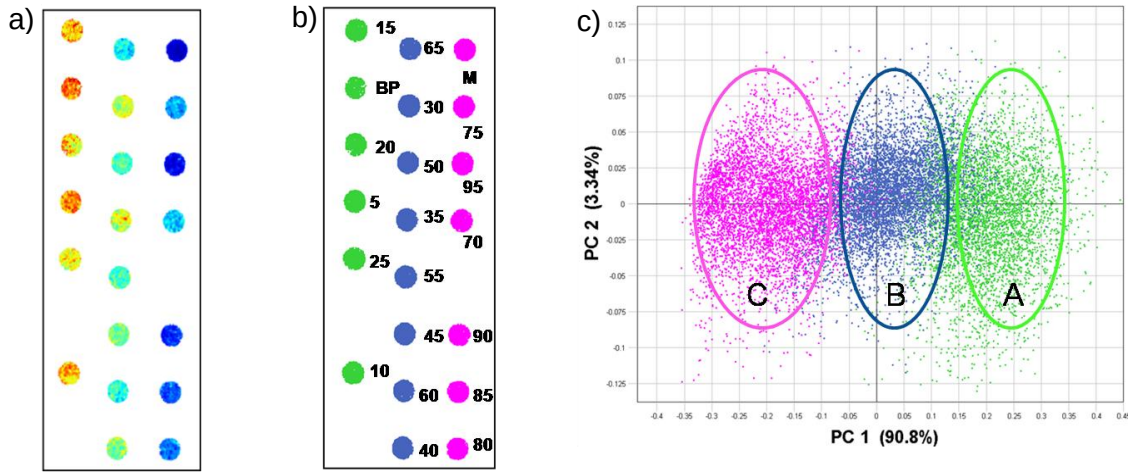


Figure 3.8 a) Score image PC1 depicting sample differences, **b)** classification image (calibration set) of PC1 illustrating three classes A (green), B (blue), C (pink) with their respective adulterant content in bold, and **c)** classification plot depicting the position of selected classes in the data swarm along PC1 (projection of classification image).

Table 3.1 depicts the amount of pixels initially assigned in the test set image to each class (A, B, C) as well as the amount of pixels correctly and incorrectly predicted for each class. The test set consisted of 17 850 pixels where 5011, 6023 and 6815 pixels were assigned to classes A, B and C respectively. Only 16.70% of the pixels initially assigned as A was predicted as A of which 97.85% were correctly predicted as A. Of the 6023 pixels assigned to class B, 92.33% were predicted as B, but only 38.01% were correctly predicted as B. Class C initially consisted of 6815 pixels and 81.44% were predicted as C of which 78.28% were correctly predicted.

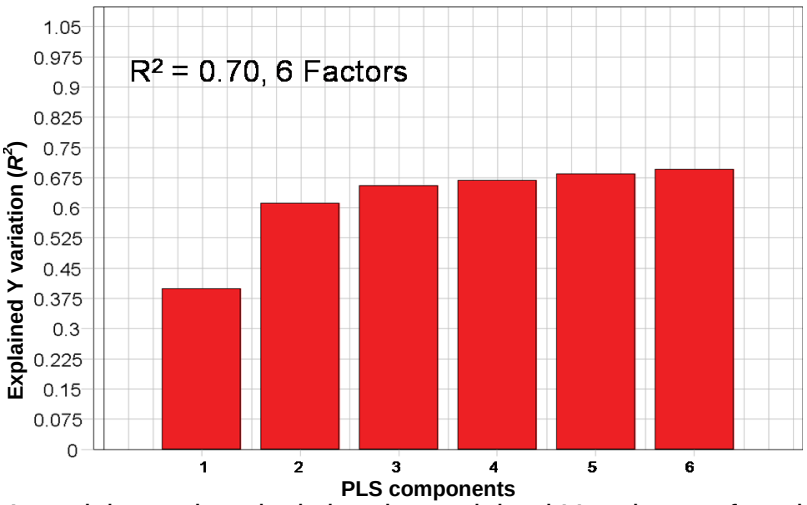


Figure 3.9 PLS-DA model overview depicting the explained Y variance after six PLS components (calibration).

Table 3.1 Classes A, B and C with their assigned pixels and classes predicted as A, B, C and not classified, with allocated pixels for test set image

Actual classes and assigned pixels		**Predicted classes with pixel allocation**			
Classes	Assigned pixels	A (0 – 25% adulteration)	B (30 – 60% adulteration)	C (65 – 100% adulteration)	Not Classified
A (0 – 25% adulteration)	5011	819 (97.85%)*	1470	17	2705
B (30 – 60% adulteration)	6023	17	2114 (38.01%)*	1188	2704
C (65 – 100% adulteration)	6815	1	1977	4345 (78.28%)*	493
Total	17850	837 (4.70%)#	5561 (31.15%)#	5550 (31.09%)#	5902 (33.06%)

#*Percentage predicted as class*
**Percentage correctly predicted as class*

Upon investigation of the prediction image (**Fig. 3.10b**), it became clear that numerous black pepper related pixels were classified in the unadulterated black pepper sample thus giving us an indication that NIR HSI has the ability of predicting unadulterated black pepper better than the other adulterated samples in class A. Earlier it was mentioned that class B overlaps into both class A and C, this overlapping is clearly evident when viewing the model's ability of predicting class B. Of the 5561 pixels predicted as class B, 61.99% was incorrectly predicted as class B. The prediction of class C was by far the best with only 21.71% of the pixels incorrectly predicted as class C. An enormous amount of pixels were unclassified (**Table 3.1**), indicating the model's inability of identifying the previously assigned classes.

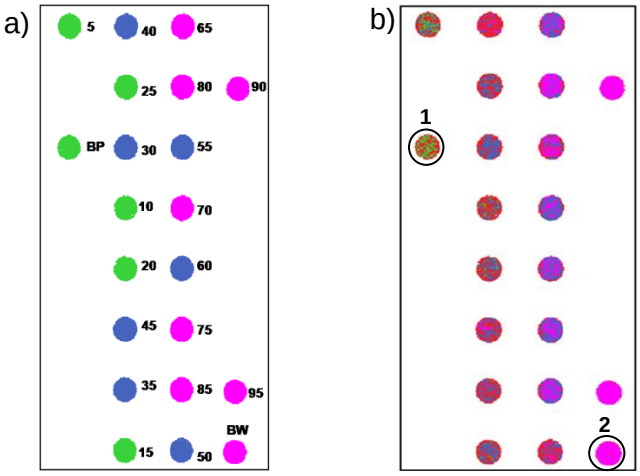


Figure 3.10 a) Classification image (test set) with the same classes (A = green, B = blue and C = pink) as the calibration set and **b)** predicted image illustrating the prediction of classes (1 = unadulterated black pepper and 2 = pure buckwheat).

Table 3.2 Discrimination results of two additional test sets (batches 3 and 4)

Batches predicted	Pixel total	Predicted classes with pixel allocation			
		A (0 – 25% adulteration)	B (30 – 60% adulteration)	C (65 – 100% adulteration)	Not classified
3	16 007	949 (5.93%)	10 200 (63.72%)	1424 (8.90%)	3434 (21.45%)
4	15 987	741 (4.64%)	7015 (43.88%)	4336 (27.12%)	3895 (24.36%)

The discrimination results of two additional test sets are illustrated in **Table 3.2**. The results obtained for the prediction of class A remained similar to those found in **Table 3.1**, but varied for class B and class C. The unclassified pixels also decreased for batches 3 and 4 when compared to **Table 3.1**. No prominent improvement of predictions was observed when the obtained results for batches 3 and 4 were studied.

Millet adulterated black pepper

Black pepper batches 2 and 4 adulterated with millet were investigated. Batch 2 was used in the calibration process, whereas the remaining batches were used as test sets. The score image of PC1 (**Fig. 3.11a**) was divided into similar classes as with the buckwheat adulterated samples and class overlap was present once the classification image (**Fig. 3.11b**) was projected onto the classification plot (**Fig. 3.11c**). A PLS-DA model was created using the calibration set (**Fig. 3.11b**) explaining 77% of the Y variation after 6 PLS components. No further variation increase was observed after six PLS component and therefore only six PLS components were used to explain the variation in the model (**Fig. 3.12**).

The amount of pixels initially assigned to each class (A, B, C) and the amount of pixel correctly and incorrectly predicted for each class is illustrated in **Table 3.3**. The calibration set presented by classes A, B and C collectively consisted of 16051 pixels. Classes A, B and C contained 4586, 6343 and 5122 pixels respectively. For class A, 40.06% of the pixels were predicted as A and 98.37% thereof were correctly predicted. A small proportion (30 pixels) of the 1837 pixels, were predicted as class B. The PLS-DA model predicted 8002 pixels as class B, but only 6343 pixels were initially assigned to class B (**Table 3.3**).

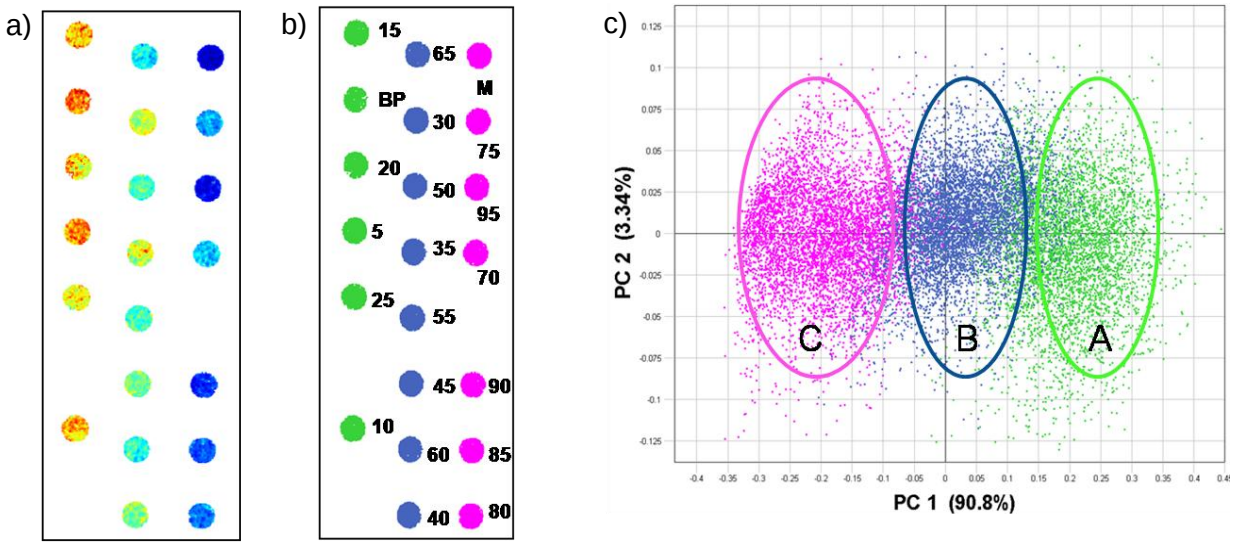


Figure 3.11 a) PC score image of PC1 illustrating sample differences **b)** Classification image (calibration set) with three classes A (green = 0 – 25% adulteration), B (blue = 30 – 65% adulteration), C (pink = 70 -100% adulteration) and **c)** projection of selected classes onto classification plot.

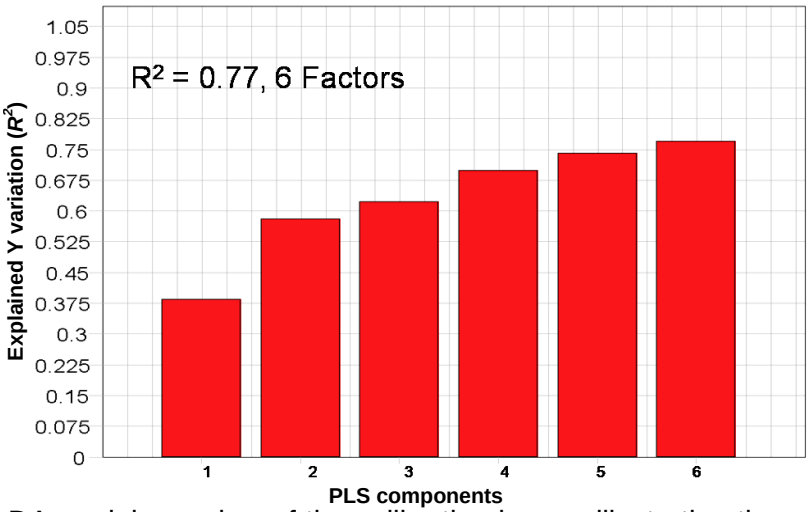


Figure 3.12 PLS-DA model overview of the calibration image, illustrating the explained Y variance after six PLS components.

Table 3.3 Classes A, B and C with their assigned pixels and classes predicted as A, B, C and not classified, with allocated pixels

Actual classes and assigned pixels		**Predicted classes with pixel allocation**			
Classes	Assigned pixels	A (0 – 25% adulteration)	B (30 – 65% adulteration)	C (70 – 100% adulteration)	Not Classified
A (0 – 25% adulteration)	4586	1807 (98.37%)*	1357	22	1400
B (30 – 65% adulteration)	6343	30	3471 (43.38%)*	1532	1310
C (70 – 100% adulteration)	5122	0	3174	1921 (55.28%)*	27
Total	16051	1837 (11.44%)#	8002 (49.85%)#	3475 (21.65%)#	2737 (17.05%)

#*Percentage predicted as class*
**Percentage correctly predicted as class*

This indicates an over-estimation and a reason for this might be due to the fact that overlapping into class A and C was present upon class selection. The pixels correctly predicted as class B amounted to 43.38%, the remaining pixels were either predicted as class A (1357 pixels) or class C (3174 pixels). More pixels were predicted as class C and thus suggest that the overlap was more into class C then class A, which is evident when looking at the classification plot (**Fig. 3.11c**).

Of the initial 5122 pixels assigned to class C, 67.84% were predicted as class C. Of the latter 55.28% were correctly predicted as class C representing 37.50% of pixels initially assigned as class C. The remaining 1554 pixels were either incorrectly predicted as class A (22 pixels) or class B (1532 pixels). The model was not able to classify 2737 pixels into any of the three classes, indicating the model inability to predict specific classes successfully. Upon investigation of the prediction image (**Fig. 3.13b**), the unadulterated black pepper sample again appeared the best predicted in class A. This then shows that NIR HSI has the ability to classify unadulterated black pepper better than the other adulterated samples in class A.

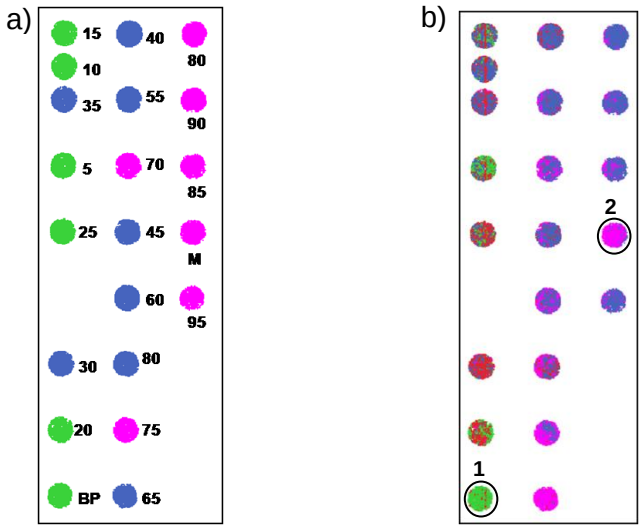


Figure 3.13 a) Classification image of test set with the same classes found in calibration set and **b)** the predicted image illustrating the prediction of the three selected classes (1 = unadulterated black pepper and 2 = pure buckwheat).

Even though some difficulty was experienced with the PLS models predicting class B, the models performed well in correctly predicting the unadulterated black pepper samples and pure adulterants. This difficulty is partly due to the fact that the particle size (≤ 500 µm, in diameter) of all powdered samples may be bigger or smaller than the pixel size (300 x 300 µm) of the obtained images. Taking this into account it is possible that a pixel in the image may consist of more black pepper or adulterant particles than measured. Therefore confusion is visible when performing a prediction based on percentage adulterant present. A possible recommendation is to rather create the PLS-DA model based on black pepper and adulterant presence than percentage adulterant present. This approach (even though not shown) was attempted, but with not that much success. It was possible to identify the presence of adulterant in samples adulterated at 30%. Another contributing factor to the PLS-DA model development and predicting performance is the surface area of the samples. The surface area will rarely be representative of the exact amount of black pepper or adulterant present in the samples, thus the chances of over- or under-estimation of adulterant or black pepper presence during calibration and prediction, are great.

The results obtained from two additional validation sets are summarised in **Table 3.4**. These results are similar to those found in **Table 3.3**, but with lesser unclassified pixels. It was also observed that class A was better predicted when compared to results found in **Table 3.3**. This might be due the fact that the samples were better mixed inside the eppendorf tube holes. The speed of data collection and the non-destructiveness of the technique are two good advantages that the food industry can utilise in identifying the adulteration of foods. The application of NIR HSI also allows the visualisation of chemical compounds within samples which is yet another advantage to the disposal of the food industry.

Table 3.4 Discrimination results of two additional test sets (batches 1 and 3)

Batches predicted	Pixel total	Predicted classes with pixel allocation			
		A (0 – 25% adulteration)	B (30 – 65% adulteration)	C (70 – 100% adulteration)	Not classified
1	17 188	5548 (32.28%)	960 (5.59%)	7942 (46.21%)	2738 (15.93%)
3	16 262	3335 (20.51%)	11 326 (69.65%)	755 (4.64%)	842 (5.18%)

Powdered black pepper are heterogeneous of nature, thus complicating data analysis, but with minimisation of physical effects (sample size, light exposure) more emphasis can be placed on chemical constituent identification. In this study black pepper, millet flour and buckwheat flour were used and all three of them are biological material, thus containing similar chemical structures, therefore it can possibly influence both the model creation and model prediction quality.

Conclusion

The use of PCA in data exploration was very effective, since gradients and sample separation (clusters) were observed. The observed gradients along PC1 were driven by the adulterant concentration present in the samples. This could be established, since *a priori* knowledge of the samples was available. Previous studies performed on buckwheat and millet flour strongly suggest both protein structural and content differences. The investigation of loading line plot PC4 suggests a protein related difference between buckwheat and millet adulterated black pepper, but further investigation into both adulterants needs to be performed to establish if the observed differences are due to protein structure or content.

The influence of particle size vs. pixel size on PLS-DA model creation pose great challenges in correct classification of assigned classes, but this problem can be addressed with advances in spatial data analysis. Sample presentation (e.g. surface area) and penetration depth of near infrared radiation are other factors that also might have influenced the prediction results. The success of this study was greatly influenced by the aforementioned factors, but we speculate to find better success in using the obtained hyperspectral data for regression purposes. The current study signals the beginning of research specifically based on powdered food material and the use of NIR HSI combined with PLS-DA in online screening procedures. NIR HSI provides more information (590 080 spectra) about the studied material to the analyst compared to conventional NIR spectroscopy (168 spectra). Therefore NIR HSI permits a better understanding of the studied material. Our study proves NIR HSI is a promising technique for the identification of adulterants in ground black pepper and the developed approach can also be utilised in identifying adulterants in other food material of a powdered nature. Using the current application new insights into ground

black pepper, buckwheat and millet flour were gained. It is also important to note that with the more informed consumer and increase in food demand, quality parameters of food gets more astringent, thus developing a need for faster and non-destructive analytical methods, i.e. NIR HSI.

References

Ariana, D.P. & Lu, R. (2010) Hyperspectral waveband selection for internal defect detection of pickling cucumbers and whole pickles. *Computers and Electronics in Agriculture*, **137**, 137 – 144.

Campbell, C.G. (1997). Buckwheat *Fagopyrum esculentum* Moench. Promoting the conservation and use of underutilized and neglected crops. P. 20. Rome, Italy: Institute of Plant Genetics and Crop Plant Research.

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

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

Cozzolino, D., Cynkar, W. U., Shah, N., Dambergs, R.G. & Smith, P.A. (2009). A brief introduction to multivariate methods in grape and wine analysis. *International Journal of Wine Research*, **1**, 123–130.

Beebe, K.R., Pell, R.J. & Seasholtz, M.B. (1998). *Chemometrics: A practical guide.* Pp. 1-8, 26-55. New York, USA: John Wiley & Sons, Inc.

Bhattacharjee, P., Singhal, R.S. & Achyut, S.G. (2003). Supercritical carbon dioxide extraction for identification of adulteration of black pepper with papaya seeds. *Journal of the Science of Food and Agriculture*, **83**, 783-786.

Bejosano, F.P. & Corke H. (1999). Properties of protein concentrates and hydrolysates from *Amaranthus* and Buckweat. *Industrial Crops and Products*, **10**, 175-183.

Bruun, S.W., Søndergaard, I. & Jacobsen, S. (2007). Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 1. Gluten powder. *Journal of Agricultural and Food Chemistry*, **55**, 7234-7243.

Davies, A.M.C. & Grant, A. (1987). Review: Near infra-red analysis of food. *Interantional Journal of Food Science and Technology*, **22**, 191-207.

Dhanya, K. & Sasikumar, B. (2010). Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices. *Current Trends in Biotechnology and Pharmacy*, **4**, 454-489.

Dhanya, K., Syamkumar, S. & Sasikumar, B. (2009). Development and application of SCAR marker for the detection of papaya seed adulteration in traded black pepper powder. *Food Biotechnology*, **23**, 97-106.

Dhanya, K., Kizhakkayil, J., Syamkumar, S. & Sasikumar, B. (2007). Isolation and amplification of genomic DNA from recalcitrant dried berries of black pepper (*Piper nigrum* L.) a medicinal spice. *Molecular Biotechnology*, **37**,165-168.

ElMasry, G. & Sun, S. (2010). Meat quality assessment using a hyperspectral imaging system. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by D. Sun). Pp. 175 – 240. London, UK: Elsevier.

Fassio, A., Gimenz, A., Fernandez, E., Vaz Martins, D. & Cozzolino, D. (2007). Prediction of chemical composition in sunflower and silage (*Helianthus annus* L.) by near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, **15**, 201-207.

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

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

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

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

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

Grahn, H.F. & Geladi, P. (2007). *Techniques and Applications of Hyperspectral Image Analysis.* Pp. 1-15. West Sussex, England: John Wiley & Sons Ltd.

Kamruzzaman, M., ElMasry, G., Sun, D. & Allen, P. (2010). Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering* (Article in press: doi: 10.1016/j.jfoodeng.2010.12.024).

Kasaoka, S., Oh-hashi, A., Morita, T. & Kiriyama, S. (1999). Nutritional characterization of millet protein concentrates produced by a heat-stable α -amylase digestion. *Nutrition Research*, **19**, 899-910.

Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). *A User Friendly Guide to Multivariate Calibration and Classification.* Pp. 39-54. West Sussex, UK: NIR Publications.

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

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

Prakash, D., Prakash N. & Misra, P.S. (1987). Protein and amino acid composition of *Fagopyrum* (buckwheat). *Plant Foods for Human Nutrition*, **36**, 341-344.

Schulz, H., Quilitzsch, R. & Krüger, H. (2003). Rapid evaluation and quantitative analysis of thyme, origano and chamomile essential oils by ATR-IR and NIR spectroscopy. *Journal of Molecular Structure*, **661-662**, 299-306.

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

Tremlova, B. (2001). Evidence of spice black pepper adulteration. *Czech Journal of Food Scienc*es, **19**, 235-239.

Wang, N. & ElMasry, G. (2010). Bruise detection of apples using hyperspectral imaging. In: *Hyperspectral Imaging for Food Quality Analysis and Control* (edited by D. Sun). Pp. 295 – 321. London, UK: Elsevier.

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.

Wen, T. & Luthe, D. (1985). Biochemical characterization of rice glutelin, *Plant Physiology*, **78**, 172-177.

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

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

Workman, J.J. (2001). Review of chemometrics applied to spectroscopy: Quantitative and qualitative Analysis. In *The Handbook of organic compounds*, pp. 301-326. Burlington: Academic Press.

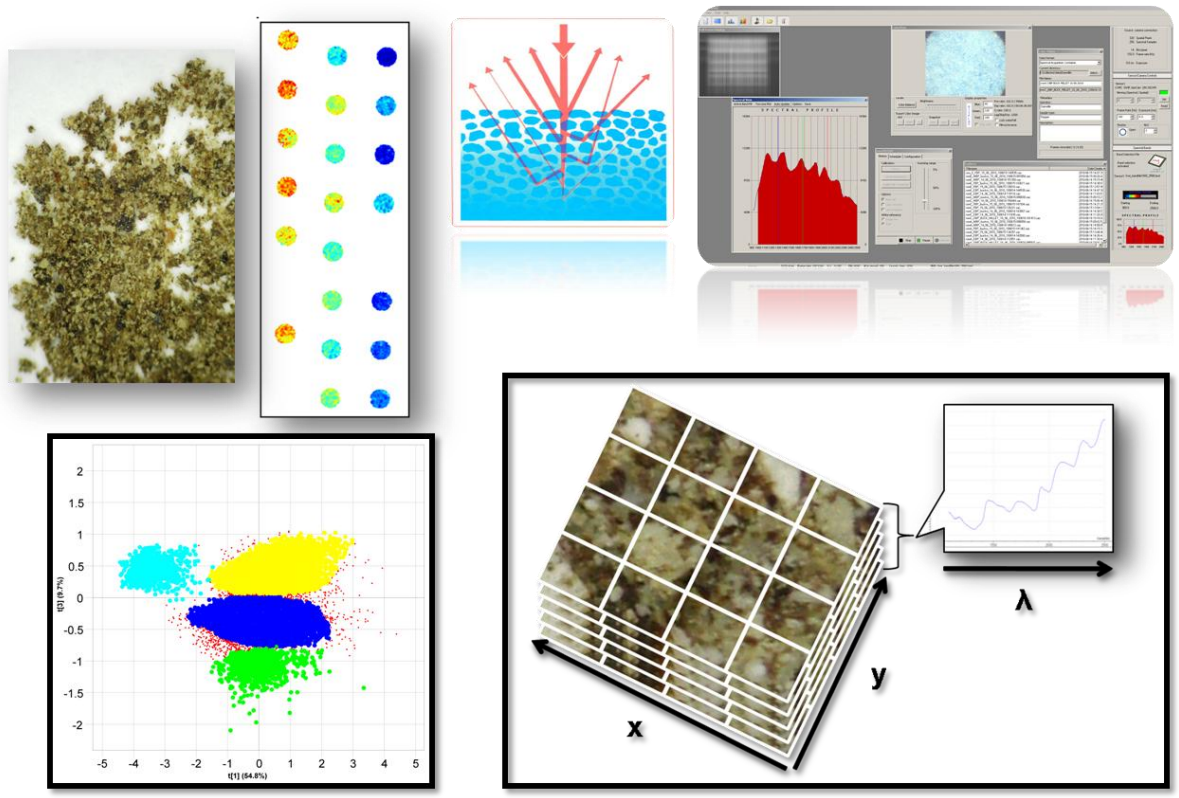
Woodcock, T., Downey, G. & O'Donnell, C. P. (2008). Better quality food and beverages: the role of near infrared spectroscopy. *Journal of Near Infrared Spectroscopy,* **16**, 1-29.

Woodman, A.G. (1941). Pepper. In: *Food Analysis: Typical methods and the interpretation of results*, 4th ed. p. 396. New York: McGraw-Hill Book Company, Inc.

Wu, D., He, Y., Shi, J. & Feng, S. (2009). Exploring near and midinfrared spectroscopy to predict trace and zinc contents in powdered milk. *Journal of Agricultural and Food Chemistry*, **57**, 1697-1704.

Chapter 4

Near infrared (NIR) and mid-infrared (MIR) spectroscopy for the quantification of adulterants in ground black pepper



Near infrared (NIR) and mid-infrared (MIR) spectroscopy for the quantification of adulterants in ground black pepper

Abstract

The use of near infrared (NIR) and mid-infrared (MIR) spectroscopy is well known in the food, pharmaceutical and agricultural industries. NIR and MIR (in an attenuated reflectance (ATR) configuration) spectroscopies allow the non-destructive analysis of powdered mixtures. Partial least squares regression is usually applied to NIR and MIR data to obtain quantitative information. In this study the feasibility of NIR and MIR spectroscopy in conjunction with PLS regression to quantitatively predict the presence of millet and buckwheat in ground black pepper, was evaluated. Whole black pepper kernels received from four different manufactures were milled and subsequently adulterated with buckwheat or millet flour in proportions ranging between 5 and 95% w/w in increments of 5% w/w. All samples were dried prior to mixing; analysed with NIR hyperspectral imaging (1000 – 2498 nm) and ATR Fourier transform (FT) – infra red (IR) instrumentation (576 – 3999 cm^{-1}). An average spectrum was calculated for each sample in the hyperspectral images. Raw NIR and MIR data were preprocessed using Savitzky Golay 2nd derivative and multiplicative scatter correction (MSC) respectively. Millet levels (NIR based) were more accurately predicted (r^2 = 0.99, *RMSEP* = 3.02% (w/w), PLS factors = 4) than buckwheat (r^2 = 0.83, *RMSEP* = 12.90% (w/w), PLS factors = 2) concentration in black pepper. The PLS model for the prediction of adulterant, irrespective of adulterant type performed well (r^2 = 0.99, *RMSEP* = 3.32% (w/w), PLS factors = 4). Poor predictions were obtained for MIR data (r^2 = 0.56, *RMSEP* = 19.94% (w/w), PLS factors = 7) when compared to NIR data.

Introduction

Near infrared (NIR) and mid-infrared (MIR) spectroscopy, in conjunction with partial least squares regression (PLS-R), have been extensively employed for quantification of food, agricultural and pharmaceutical products (Yang & Irudayaraj, 2002; Shultz *et al.*, 2003; Cozzolino & Moron, 2004; Wang *et al.*, 2006; Cen & He, 2007; González *et al.*, 2007). These vibrational spectroscopy techniques provide an abundance of chemical information to the user, making them suitable for quantification studies (Workman, 2001; Paradkar, 2002). NIR and MIR spectroscopy both require a dipole moment change for a molecular vibration to be observable; MIR spectroscopy utilises the fundamental forms of these vibrational modes and NIR spectroscopy the overtones and combinations (McKelvy *et al.*, 1998). In addition, NIR observable molecular vibrations must be associated with a large anharmonicity. MIR concentrates on samples absorption of radiation in the MIR region (4000– 400 cm^{-1}) of the electromagnetic spectrum. NIR spectroscopy is focused on a higher energy region of the electromagnetic spectrum: 714 - 2500 nm or 14 000 to 4000 cm^{-1}

(Downey, 1994; Wu & Siesler, 1999; Hein *et al.*, 2010). These differences and their related consequences (e.g. sample preparation), influence the precision of developed PLS-R models.

PLS-R reduces the dimensionality of any given multivariate data set and maximizes the covariance between direct (reference values) and indirect (e.g. MIR and NIR spectra) responses (Geladi & Kowalski, 1986). Spectral information obtained using both MIR and NIR spectroscopy has been employed in soil carbon measurements (McCarty *et al.*, 2002); in this specific study MIR out performed NIR in the prediction of soil composition. Various research groups have investigated NIR and MIR spectroscopy to determine essential oil authenticity (Lai *et al.*, 1995; Guillen *et al.*, 1997; Küpper *et al.*, 2001; Kasemsumran *et al.*, 2005; Sandasi *et al.*, 2010). The use of MIR, NIR and Raman spectroscopy for the assessment of Buchu oil has also been investigated (Sandasi *et al.*, 2010). In this study the application of MIR together with PLS-R generated the most accurate regression model. Other investigative studies where MIR and NIR were used simultaneously include coffee varietal identification (Downey *et al.*, 1997), prediction of trace iron and zinc content in powdered milk (Wu *et al.*, 2009), the examination of biodegraded spruce wood (Schwanninger *et al.*, 2004) and the characterisation of pepper oil, peppercorn and pepper oleoresin (Shultz *et al.*, 2005). In a more recent study the feasibility of using NIR and MIR in the monitoring of red wine fermentation was tested (Di Egidio *et al.*, 2010). This research group used the spectral data (MIR and NIR) in the development of regression models and found that NIR data performed better than MIR data in predicting the main chemical parameters (e.g. glucose, ethanol, total phenolics, total flavanoids) involved in the fermentation process. Even though many feasibility studies have been conducted on pharmaceutical, agricultural and food products using NIR and MIR, to our knowledge, it has never been attempted to quantitatively predict the adulteration of ground black pepper. Current methods, i.e. microscopic examination, chromatographic techniques and deoxyribonucleic acid (DNA) based techniques, are all focused on the identification of foreign material in black pepper (Woodman, 1941; Tremlova, 2001; Dhanya *et al.*, 2009; Dhanya & Sasikumar, 2010).

In this study the use of NIR and MIR spectroscopy to determine the amount of adulterant (millet and buckwheat) present in ground black pepper was evaluated. NIR data were obtained using NIR hyperspectral imaging. PLS-R models were developed on

- Ground black pepper separately adulterated with millet and buckwheat flour - two separate data sets (NIR and MIR data)
- Ground black pepper separately adulterated with millet and buckwheat flour as a complete set - global model (NIR and MIR data).

In addition, the predictive quality of PLS-R models developed on both NIR and MIR data were compared.

Materials and methods

Samples

Four batches of whole black pepper were supplied by four different manufactures. Buckwheat and millet were purchased as whole kernels. Black pepper, buckwheat and millet were all milled to less than 500 µm (in diameter) with a Retsch mill (Retsch model ZM1: sieve with 500 µm hole width, Haan, Germany). Two series of samples were made: Buckwheat adulterated and millet adulterated. In each series the level of adulterant varied in 5% w/w increments from 0-100% w/w adulterant, this resulted into an adulteration series containing 19 adulterated samples, one unadulterated black pepper sample and one pure adulterant (millet or buckwheat) (n = 21 per batch of black pepper).

All samples were oven (Heraeus model RVT 360, Hanau, Germany) dried at 74°C for 2 hrs prior to sample mixing to reduce moisture variation. Sample mixing was performed in clear sepcap vials (Kimble Glass Incorporated, New Jersey, USA: 15 x 45 mm) before presentation to the MIR instrument and NIR hyperspectral imaging system. The mixing process was carried as follow:

- Sepcap vials were weighed and zeroed on a balance (Ohaus model AS200S, Ohaus Corporation, USA)
- All sepcap vials were filled with 2 g of dried unadulterated ground black pepper, pure adulterant (millet or buckwheat) or adulterated black pepper (mixtures)
- After filling the sepcap vials with the ground material, it was carefully mixed with a cleaned spatula
- Sepcap vials were then closed and placed onto a Vortex-Genie® 2 (Scientific Industries, Inc., model G560E, USA), to further mix the content
- All samples were then transferred to eppendorf tube holders for imaging purposes and prior to transferral, the vial content was shaken
- A portion (50 mg) of each sample were used for MIR analysis

NIR hyperspectral imaging instrumentation

NIR hyperspectral data were acquired using the pushbroom sisuChema hyperspectral imaging system (Specim, Oulo, Sweden), which was comprised of an imaging spectrograph coupled to a 2-D array Mercury-Cadmium-Telluride (HgCdTe) detector. Samples were presented to the imaging system in eppendorf tube holders (Eppendorf AG, Hamburg, Germany: 6.5 cm x 21 cm; 80 holes). Images were acquired from 1000 to 2498 nm in 6.2 nm intervals (242 waveband channels) and a spatial resolution of 300 µm x 300 µm was obtained using a 100 mm (field of view width) lens. This produced single images of the following size: 320(x) x 922(y) x 242(λ). Acquired reflectance counts were converted to pseudo-absorbance values before any data processing or analysis was performed.

Seven single images were obtained and subsequently imported into Evince hyperspectral image analysis software version 2.4.0 (Umbio, Umeå, Sweden). A PCA model was calculated and

different combinations of principal components (PCs) were investigated for the identification and removal of outlying data points and background (i.e. eppendorf tube holder). After this procedure the spectra from each well/sample were averaged resulting into a NIR data set containing 168 samples with one spectrum per sample.

MIR instrumentation

All MIR measurements were performed using the single reflection configuration of an alpha-P ATR FT-IR spectrometer (Model A220/D-01, Bruker OPTIK GmbH, Ettlingen, Germany) fitted with a diamond crystal (**Fig. 4.1**). Spectra were obtained using OPUS version 6.5. Samples were placed on the flat sampling plate and compressed with the one-finger clamp mechanism. Spectral measurements were recorded in the range between 576 and 3999 cm^{-1} with a spectral resolution of 8 cm^{-1}. The measured spectrum was the average of 32 scans. Background measurements were performed every 20 samples. Duplicate spectra from two different aliquots were collected from each sample. The obtained spectra were imported into The Unscrambler® X (CAMO Software AS, Norway) where duplicate spectra were averaged.

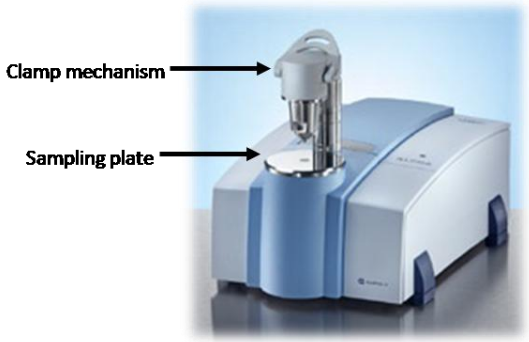


Figure 4.1 The alpha-P ATR FT-IR spectrometer illustrating the clamp mechanism and sample plate on which sample portions were placed (Bruker OPTIK, Ettlingen, Germany).

Partial least squares regression

The appropriate chemometric techniques for regression purposes were executed in The Unscrambler statistical software program. Savitzky Golay 2nd derivative (nine point averaging, 3rd polynomial order) was applied to NIR data whereas multiplicative scattering correction (MSC) was applied to MIR data, as preprocessing methods.

For PLS calibration development the collected NIR and MIR data were evaluated in the following manner: 1) development of PLS models based on the type of adulterant (millet or buckwheat) present in the ground black pepper; and 2) development of a global PLS model which incorporated all the samples irrespective of the adulterant present. Two sample sets (millet or buckwheat adulterated black pepper) containing 84 samples each, were used for PLS model development. The sample sets were divided into a training (n = 63) and test (n = 21) set, with the test set containing a single black pepper batch. Extra unadulterated black pepper samples were added to the training set (n = 66) and test set (n = 22). A calibration model was calculated using

the samples in the training set and subsequently used to predict the samples in the test set. This procedure was carried out until each adulterated batch of black pepper had been predicted once; this resulted into four models for each adulterant (buckwheat and millet).

Global PLS models were created using all samples plus 8 additional unadulterated black pepper samples (n = 176). The training set consisted of 154 samples whereas the test set contained 22 samples: 2 unadulterated black pepper samples, 19 adulterated black pepper samples and 1 pure adulterant (millet or buckwheat). The same procedure was followed as described earlier, where the test set is predicted using a calibration model. This resulted in eight PLS models, one model for the prediction of one adulterated black pepper batch (2 adulterants x 4 black pepper batches).

Full cross-validation was applied to determine the appropriate number of PLS factors needed for calibration models. Calibration accuracy was explained by the root mean square error of cross validation (*RMSECV*), root mean square error of calibration (*RMSEC*), coefficient of determination (R^2) and the bias. Whereas the predictive quality of the created models was described by the root mean square error of prediction (*RMSEP*), r^2 and bias.

Results and discussion

NIR spectra

The averaged raw and second derivative spectra of four different unadulterated black pepper batches are shown in **Fig. 4.2**. Grouping of batches was observed in the raw spectra. This may have been due to either 1) particle size or 2) batch quality similarities. Second derivative spectra of unadulterated black pepper emphasised an absorption peak at 2378 nm (not present in millet or buckwheat). This peak arises from the O-H deformation second overtone of ROH groups. ROH groups are abundant in carbohydrates (e.g. cellulose and starches) and the epicarp of black pepper kernels primarily comprises cellulose (Woodman, 1941).

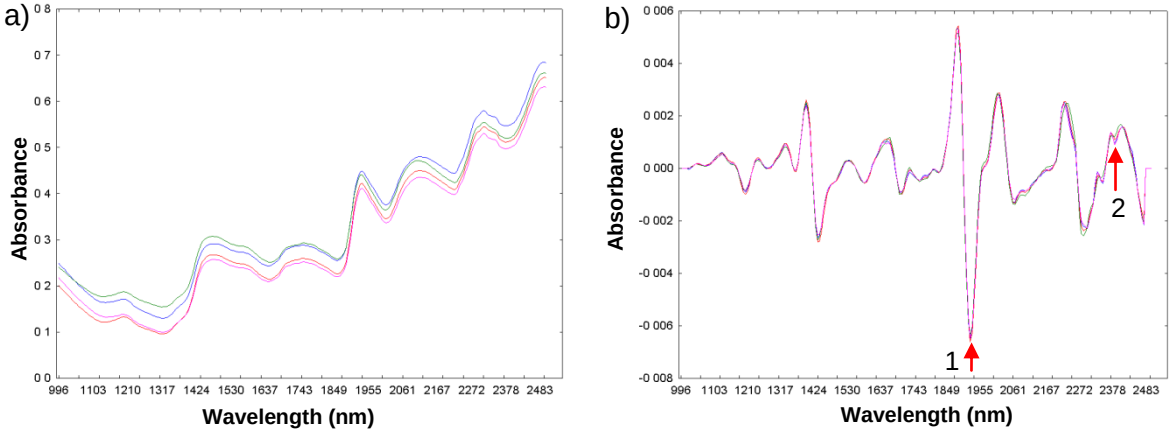


Figure 4.2 a) Averaged raw NIR spectra of four different unadulterated black pepper batches and their (with grouping of batches) **b)** second derivative spectra. Red arrows in b) indicate the presence of peaks at 1) 1955 nm and 2) 2378 nm.

A strong peak at 1955 nm was apparent in the second derivative spectra. The peak at 1955 nm is due to combined N-H asymmetrical stretching and amide II bond vibrations of the $CONH_2$ group. The peak at 1955 nm is an indication of protein (see Chapter 3, P. 50).

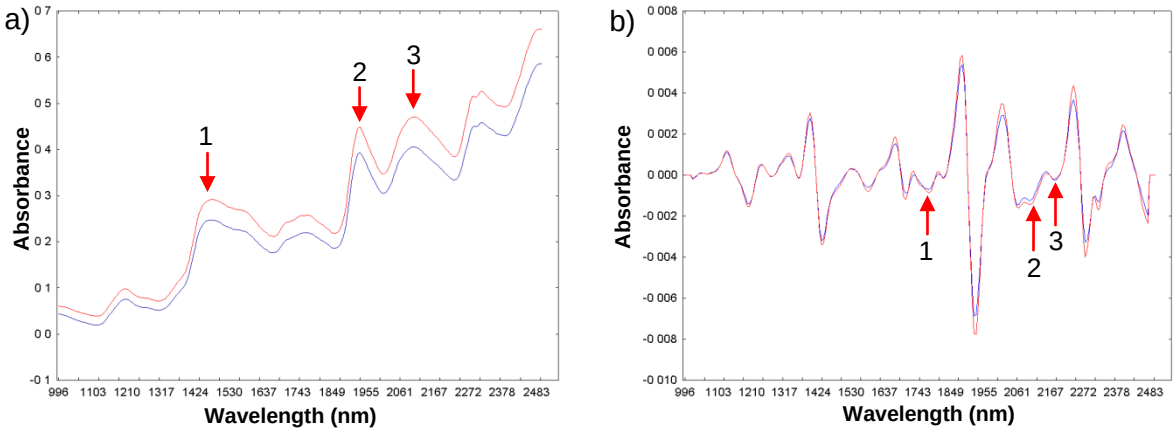


Figure 4.3 a) Averaged raw NIR spectra of pure millet (red) and buckwheat (blue) flour and **b)** the resultant second derivative spectra. Red arrows in a) indicate spectral differences at 1) 1424 nm, 2) 1955 nm and 3) 2114 nm. Red arrows in b) indicate peaks specific to millet and buckwheat 1) 1743 nm, 2) 2112 nm and 3) 2167 nm.

The averaged raw and second derivative spectra of pure buckwheat and millet flour are illustrated in **Fig. 4.3**. The second derivative spectra depict the presence of absorption bands at 1743 nm, 1955 nm (protein presence), 2112 nm and 2167 nm. The absorption peak at 1743 nm (not present in adulterated black pepper) typically arises from C-H first overtones of CH bonds and cellulose. The peak at 2167 nm arises from bond vibrations of protein related structures. The peaks at 2167 nm, 1743 nm and 2112 nm are specific for millet and buckwheat. They can therefore be used to determine whether buckwheat or millet is present in ground black pepper. The broad peak found at 1424 nm in the raw spectra of millet and buckwheat is also present in the raw spectra of the unadulterated black pepper samples, but with lower absorbance values.

Buckwheat does not contain glutelin and is particularly rich in lysine whereas millet cereals are low in lysine and glutelin is one of the protein fractions present (Campbell, 1997). Starch is a major component in millet and buckwheat, but the chemical composition of millet starch differs from that found in buckwheat. The spectral differences observed between buckwheat and millet agree with these reported compositional differences. Pertinent spectral variation was observed in the raw spectra of millet and buckwheat adulterated samples and is typically due to physical differences leading to non-systematic noise; the second derivative spectra clearly illustrate the successful removal of this variation.

Averaged raw and second derivative spectra of buckwheat and millet adulterated black pepper are illustrated in **Figs. 4.4** and **4.5**, respectively. Higher absorption values were observed at 2114 nm for millet adulterated black pepper (**Fig. 4.5a**) compared to buckwheat adulterated black pepper (**Fig. 4.4a**)

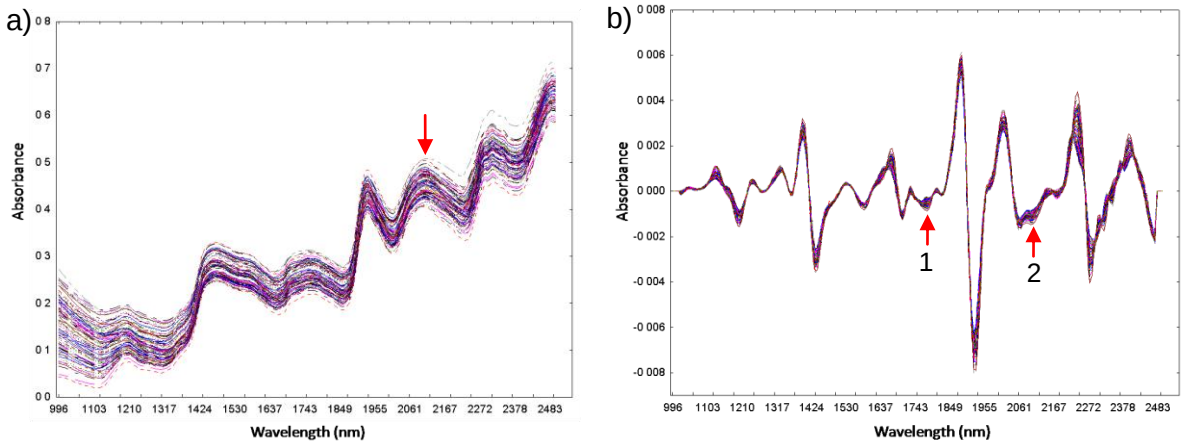


Figure 4.4 a) Raw spectra obtained after averaging each sample in the hyperspectral images of the buckwheat adulterated black pepper and **b)** the resultant spectra after applying Savitzky Golay 2nd derivative. Red arrow in a) indicates the absorption peak at 2114 nm. Red arrows in b) indicate peaks at 1) 1743 nm and 2) 2112 nm.

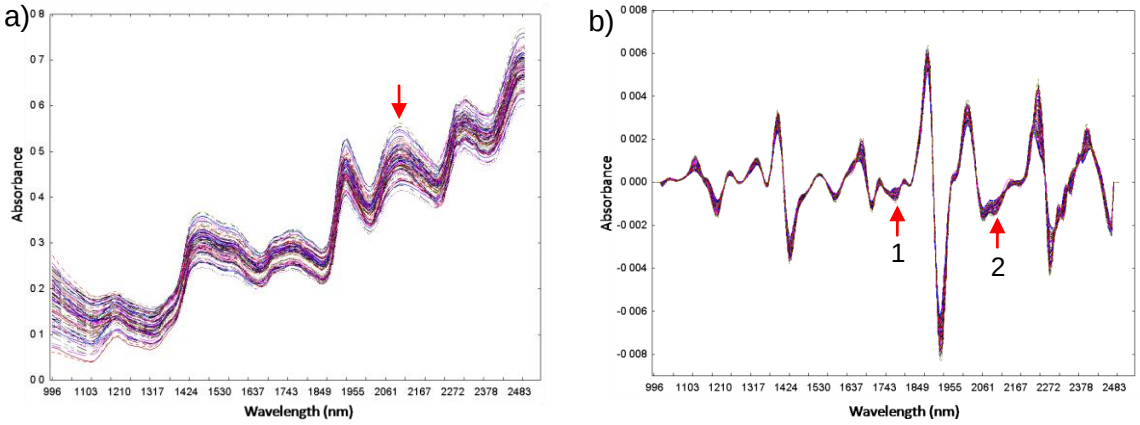


Figure 4.5 a) Raw spectra of averaged NIR hyperspectral data of the millet adulterated black pepper samples and **b)** the resultant spectra after apply Savitzky Golay 2nd derivative. Red arrow in a) indicates the absorption peak at 2114 nm. Red arrows indicate peaks at 1) 1743 nm and 2) 2112 nm.

PLS calibration models for buckwheat adulterated black pepper

A summary of the regression statistics for both calibration and test set validation is given in **Table 4.1**. Good calibration and validation results were achieved with a low number of PLS factors. It is noted that weaker calibration models performed better in test set validation with low *RMSEP* (4.41 and 4.61% w/w) and higher r^2 (0.98) values. In a previous study researchers examined the dissolution of tablets and found high values for *RMSEP* and lower values for r^2 compared to calibration models (Tatavarti *et al.*, 2004).

Table 4.1 PLS regression statistics for 4 batches of ground black pepper adulterated with buckwheat flour

PLS Model	Preprocessing	Training set (Calibration)					Test set (Validation)			
		n	R^2	*RMSEC (% w/w)*	*Bias*	*PLS factors*	*n*	r^2	*RMSEP (% w/w)*	*Bias*
1	Sav.Gol_2nd derv._9point av._3nd poly. order*	66	0.95	7.24	0	2	22	0.98	4.61	-3.72
2	Sav.Gol_2nd derv._9point av._3nd poly. order	66	0.94	7.06	0	2	22	0.98	4.41	2.52
3	Sav.Gol_2nd derv._9point av._3nd poly. order	66	0.98	4.41	0	2	22	0.91	9.16	-7.73
4	Sav.Gol_2nd derv._9point av._3nd poly. order	66	0.98	4.61	0	2	22	0.83	12.89	8.93

*Savitzky Golay 2nd derivative, 9 point averaging, 3rd polynomial order

The differences observed between the models indicated physical differences (e.g. particle size) influenced the spectra despite preprocessing. It should be noted that preprocessing only mathematically reduces the obtained spectra and does not account for all physical effects imposed on the samples during preparation and analysis. The full spectrum (1000 – 2498 nm) was used during the calibration development process and thus some spectral regions containing higher noise levels are included in the regression calculations which negatively influenced the obtained results.

Full cross-validation was only applied to model 4 to determine the optimal number of PLS factors required for the calibration model. Subsequently, a test set was used to validate the predictive quality of the model. The model was then recalculated without the use of cross-validation and again validated using test set validation. After applying full cross-validation, 98% of the Y-variance was explained within 2 PLS factors (**Fig. 4.6**). This resulted in a *RMSEC* and *RMSECV* of 4.61% w/w and 5.09% w/w, respectively (**Fig. 4.7 & 4.8**).

In a recent study, researchers orientated their study on investigating the influence of sample preparation on estimating specific chemical properties of *Eucalyptus urophylla* S. T. Blake wood using NIR data (Schwanninger *et al.*, 2004). Different sample particle sizes were used in this study and the best result was obtained using milled wood with 0.5 mm particle size. Better *RMSECV* results for the content of Klason lignin (0.53), acid soluble lignin (0.10) and syringyl-to-guaiacyl ratios (0.13) were obtained when compared to our results (5.09% w/w), but lower R^2 (0.88, 0.88 and 0.92) were obtained for their calibration models within a higher number of PLS factors (6, 6 and 7). Results we obtained for calibration development, even though not as satisfactory, are

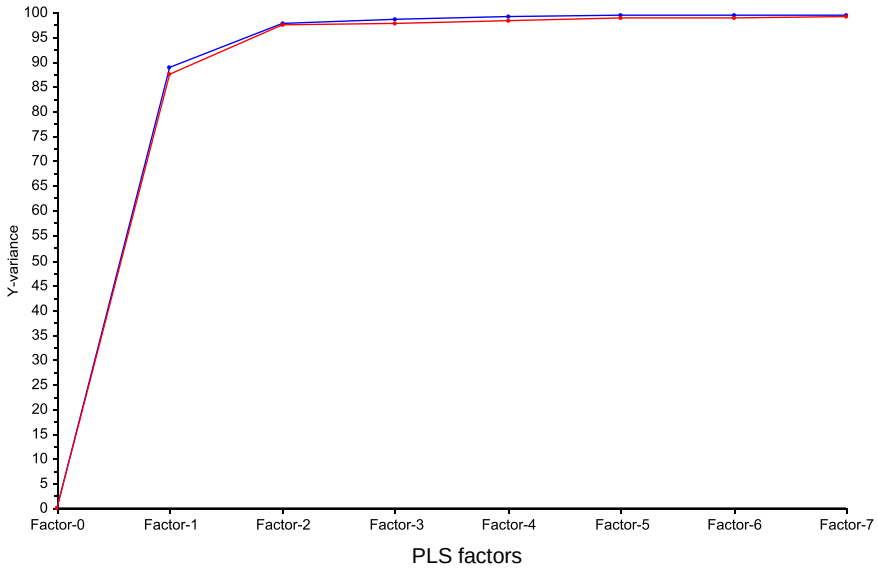


Figure 4.6 Residual validation (red) and calibration (blue) variance plots for the buckwheat adulterated black pepper samples. The explained variance are plotted against the number of PLS factors.

concurrent with results obtained with other studies where products of powdered nature were examined. It should also be noted that black pepper and buckwheat flour are heterogeneous

biological material which contain similar chemical constituents. This should therefore be considered as a contributing factor that influences the calibration development and predictive quality of the created PLS models.

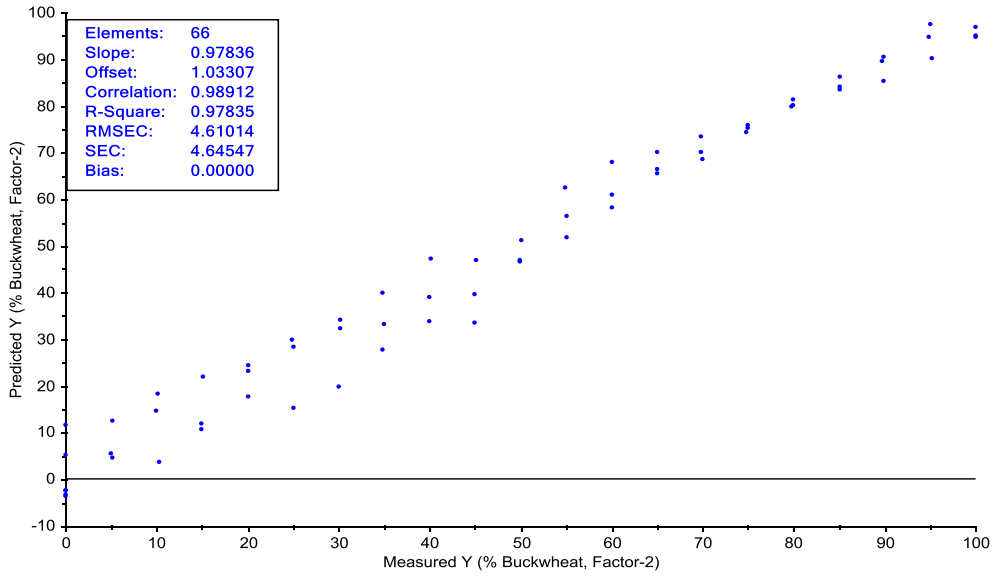


Figure 4.7 Scatter plot of calibration (*n* = 66) samples for the NIR data of the buckwheat adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the modelled percentages (NIR data).

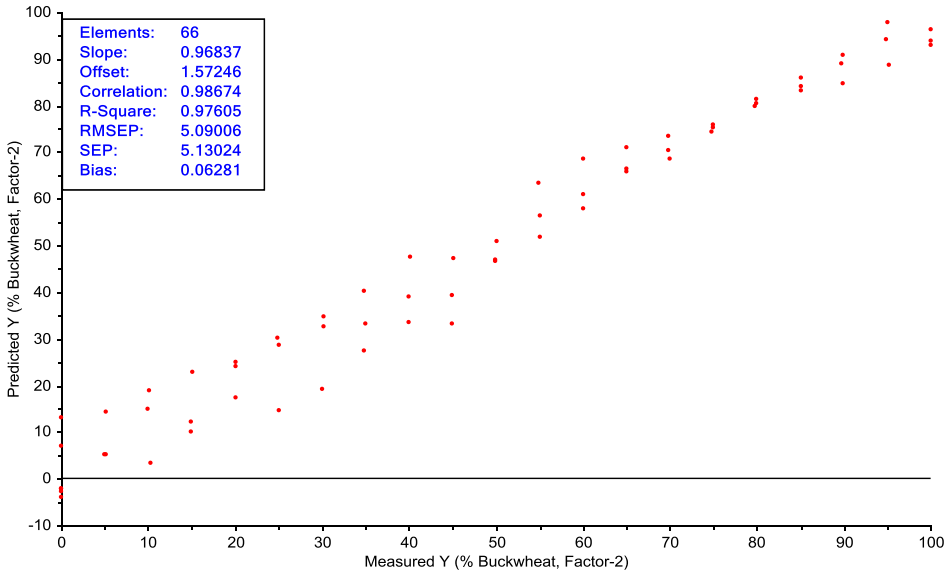


Figure 4.8 Scatter plot of validation (*n* = 66) samples for the NIR data of the buckwheat adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the modelled percentages (NIR data). RMSEP = RMSE*CV*.

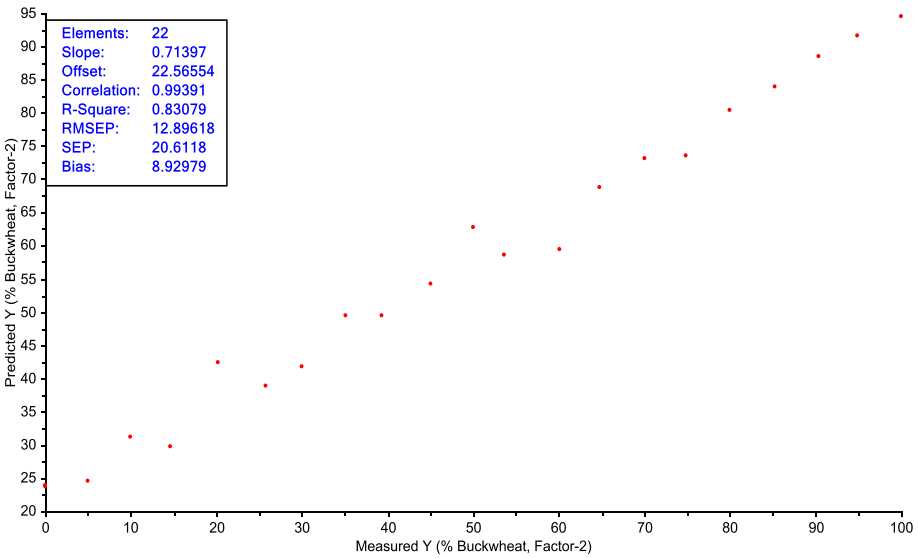


Figure 4.9 Regression graph of independent set (*n* = 22) of samples. Modelled percentages (NIR data) are plotted against actual percentage buckwheat present in the ground black pepper.

After using the developed calibration (**Fig. 4.7**) model to predict an independent set of samples, unsatisfactory results were obtained (**Fig. 4.9**). Over and under estimated (± 20% w/w) results were obtained for unadulterated black pepper and pure buckwheat flour, respectively. Using the same data set, a calibration model was created without full cross-validation (**Figs. 4.10 & 4.11**). No improvement in the predictive quality of the created calibration model (**Fig. 4.10**) was observed when compared with the calibration model created with full cross-validation (**Fig. 4.7**). The use of cross-validation did aid in estimating the correct number of PLS factors required for calibration purposes (**Figs. 4.6 & 4.10**). Even though poor predictive quality was observed during the application of cross-validation and test set validation, **Figs. 4.9** and **4.12** indicate the presence of an increasing trend relative to the amount of buckwheat present in the ground black pepper.

An irregular curve was observed after studying the explained Y-variance for the test set (**Fig. 4.10**). This might be indicative of a batch difference existing between the black pepper batch used as the test set and those used for the training set. The whole black pepper batches were initially obtained from four different manufactures and can be of varying quality. Subsequently, the milled black pepper material can thus contain varying amounts of physical black pepper related material (e.g. pericarp and parenchyma). This introduces more variability, which should be considered upon interpretation of the obtained results (suggestion – sieve material after milling). Similarly, whole buckwheat kernels were also milled and the resulting buckwheat flour can contain mixed buckwheat related material (e.g. husk, fibers), which adds to the already existing variability.

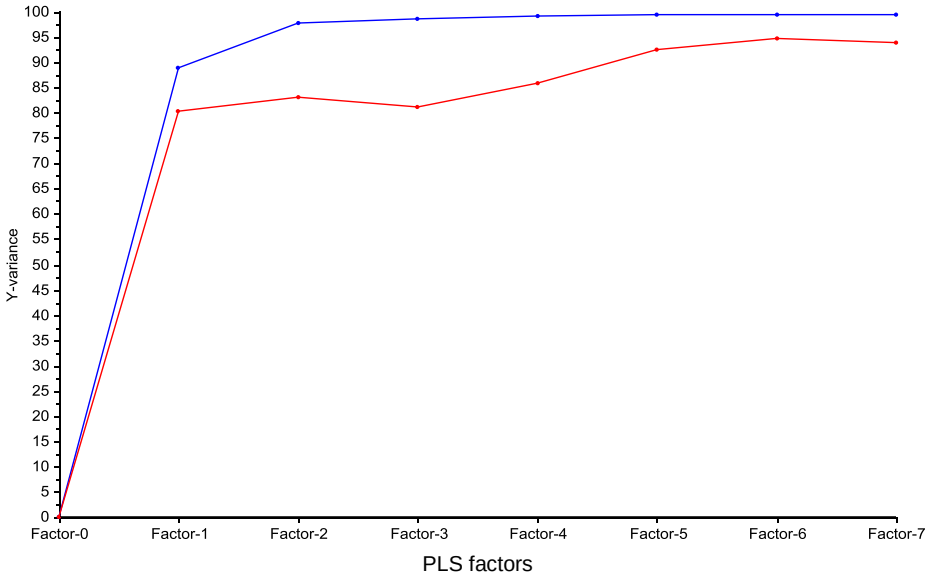


Figure 4.10 Residual test set validation (red) and calibration (blue) variance plots for the buckwheat adulterated black pepper samples. The explained variance is plotted against the number of PLS factors.

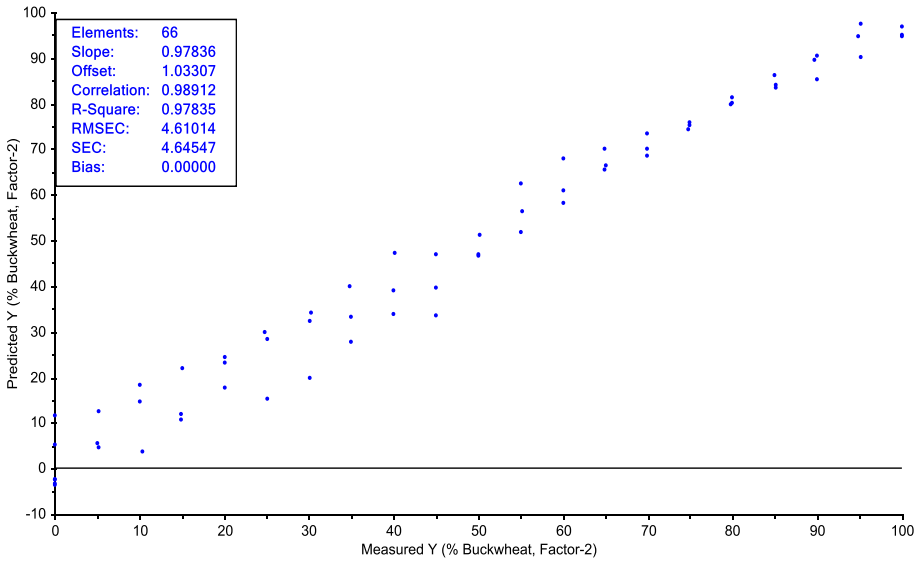


Figure 4.11 Scatter plot of calibration (*n* = 66) samples for the NIR data of the buckwheat adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the modelled percentages (NIR data).

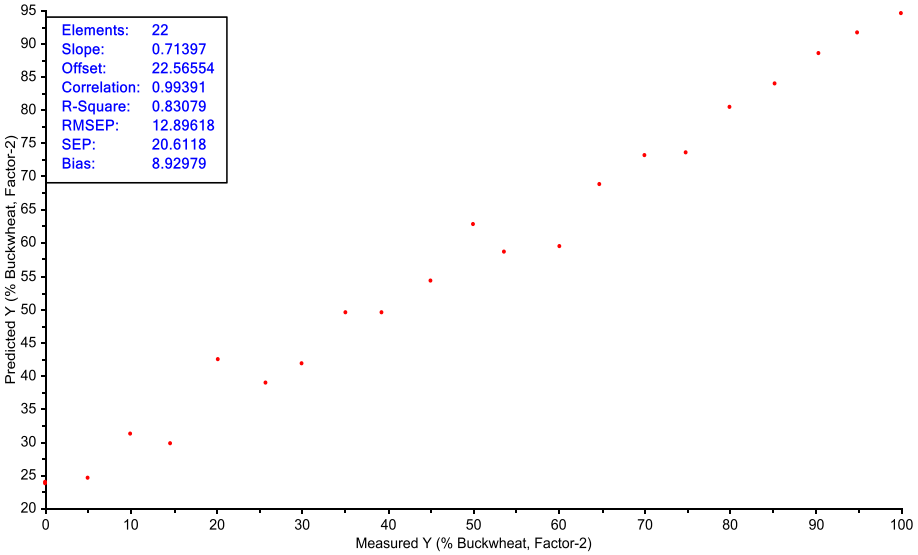


Figure 4.12 Scatter plot of test set validation (*n* = 22) samples for the NIR data of the buckwheat adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the modelled percentages (NIR data).

PLS models for millet adulterated black pepper

Full cross-validation was applied to the preprocessed NIR data of one batch of black pepper adulterated with millet flour, to determine the optimal number of factors needed to explain the Y variance (**Fig. 4.13**). Four PLS factors were found to be adequate to explain 99% of the variance (**Fig. 4.13**). One outlying sample was identified in the scatter plot of cross-validated data (not shown) and subsequently removed and a PLS model was recalculated. Excellent calibration results were achieved for model three (R^2*= 0.99, RMSEC= 2.87 % w/w, RMSECV= 3.78% w/w)* (**Fig. 4.14 & 4.15**). The predictive quality of the PLS model was subsequently tested using a test set containing 22 samples (2 = unadulterated black pepper; 1 = pure millet flour; 19 = adulteration levels). Over and under estimated results were achieved for all samples in the test set, better results were obtained compared to buckwheat adulterated black pepper samples (**Fig. 4.16**). A better trend of increasing adulterant presence was also observed compared to buckwheat adulterated black pepper (**Fig. 4.16**).

The regression statistics achieved on preprocessed data for calibration and test validation of millet adulterated black pepper are summarised in **Table 4.2**. Better validation and calibration results were obtained for all created PLS models compared to PLS models created for buckwheat adulterated black pepper. The bias of the validation set varied between -5.01 and 8.00 with good coefficients of determination ranging between 0.92 and 0.98. The predictive quality of the created calibration models was exceptionally good compared to buckwheat adulterated black pepper.

After a visual inspection of the millet flour we found that the millet particles were more homogenous than the buckwheat flour particles. In our study we used Savitzky Golay 2nd derivative as the only preprocessing method to minimise spectral variation. Relatively good results

Table 4.2 PLS regression statistics of the created calibration models for millet adulterated black pepper and test set validations

PLS model	Preprocessing	Training set (Calibration)					Test set (Validation)			
		n	R^2	*RMSEC (%w/w)*	*Bias*	*PLS factors*	*n*	r^2	*RMSEP (%w/w)*	*Bias*
1	Sav.Gol_2nd derv._9point av._3nd poly. order	66	0.99	2.23	0	5	22	0.98	4.80	1.76
2	Sav.Gol_2nd derv._9point av._3nd poly. order	66	0.99	3.09	0	4	22	0.99	3.02	-1.00
3	Sav.Gol_2nd derv._9point av._3nd poly. order	66	0.94	7.63	0	1	22	0.92	9.09	8.00
4	Sav.Gol_2nd derv._9point av._3nd poly. order	66	0.99	3.17	0	4	22	0.94	7.47	-5.01

* Savitzky Golay 2nd derivative, 9 point averaging, 3rd polynomial order

were obtained and therefore Savitzky Golay 2nd derivative was accepted as the relevant preprocessing method. Alternatively, MSC was applied to the data, but the obtained calibration and validation results (not shown) were not better than those obtained with Savitzky Golay 2nd derivative.

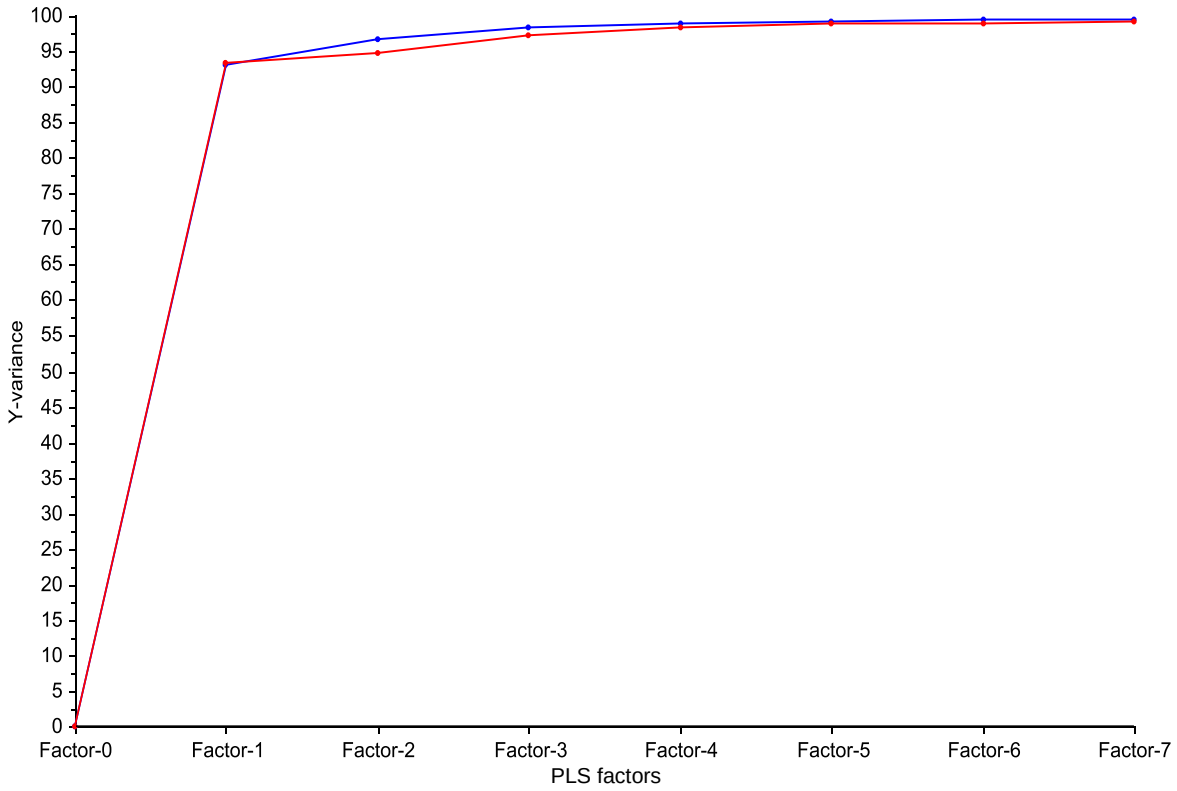


Figure 4.13 Residual validation (red) and calibration (blue) variance plots for the millet adulterated black pepper samples. The explained variance are plotted against the number of PLS factors.

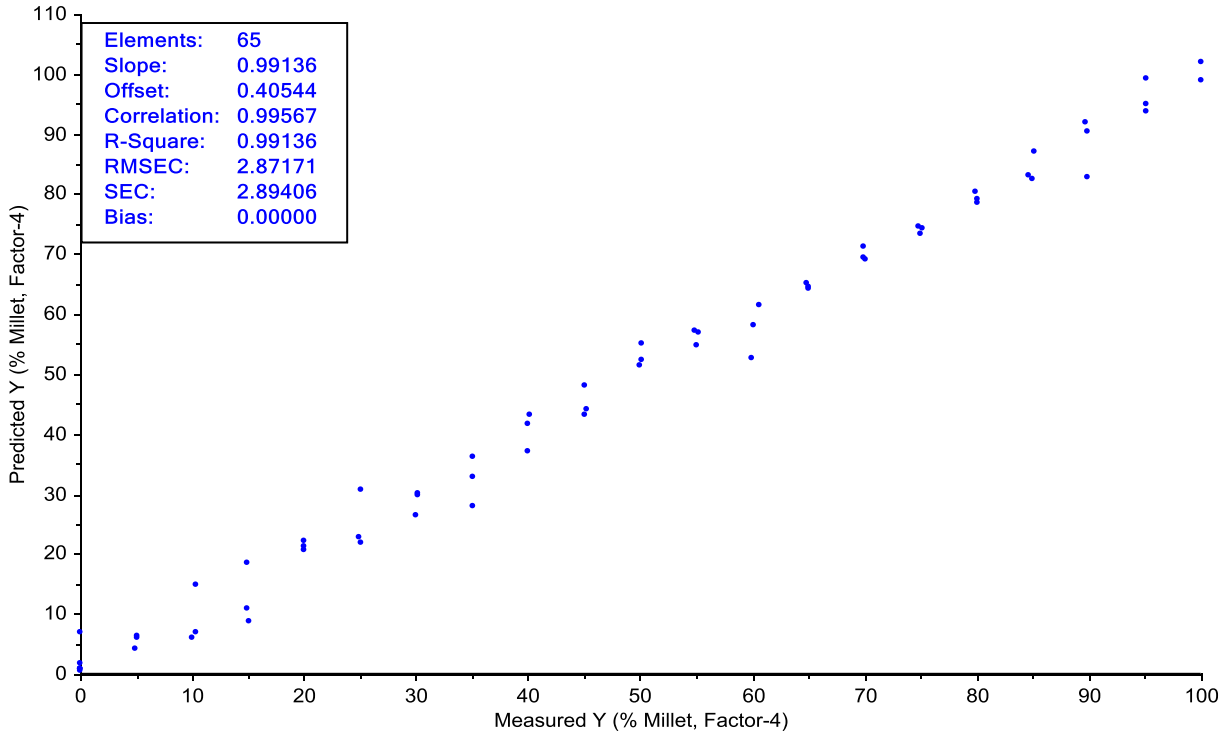


Figure 4.14 Scatter plot of calibration (*n* = 65) samples for the NIR data of the millet adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the modelled percentages (NIR data).

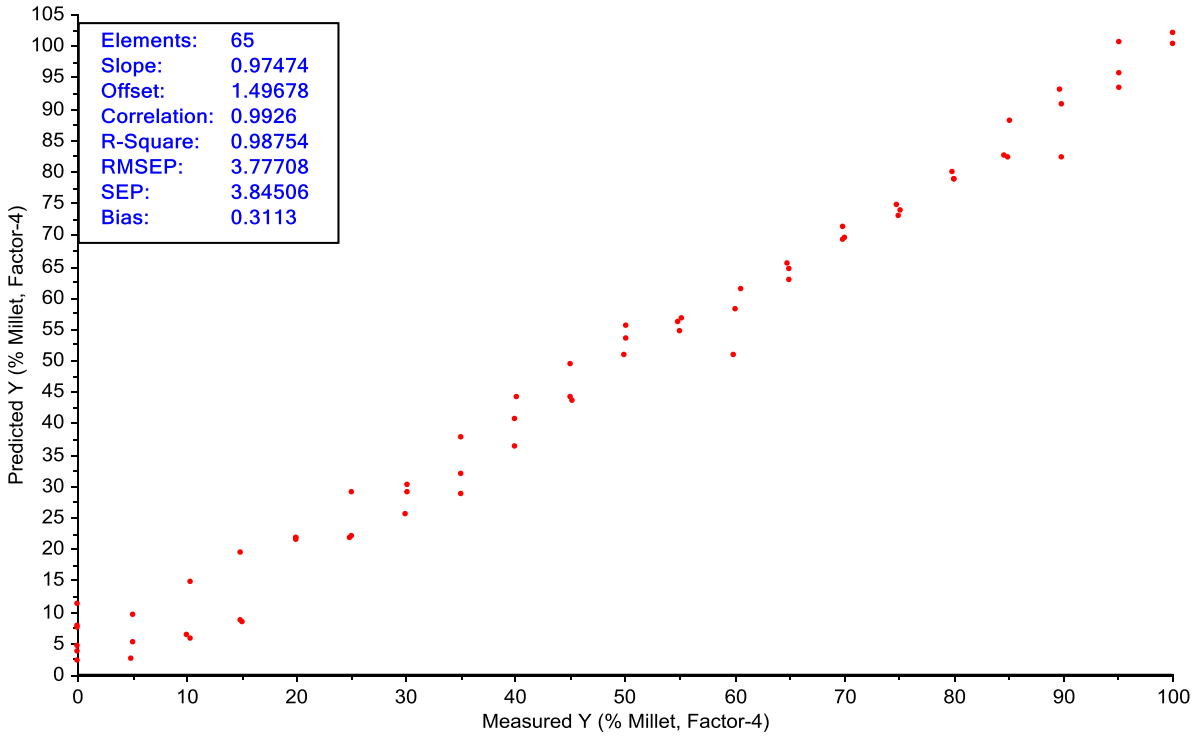


Figure 4.15 Scatter plot of validation (*n* = 65) samples for the NIR data of the millet adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the modelled percentages (NIR data). RMSEP = RMSE*CV*.

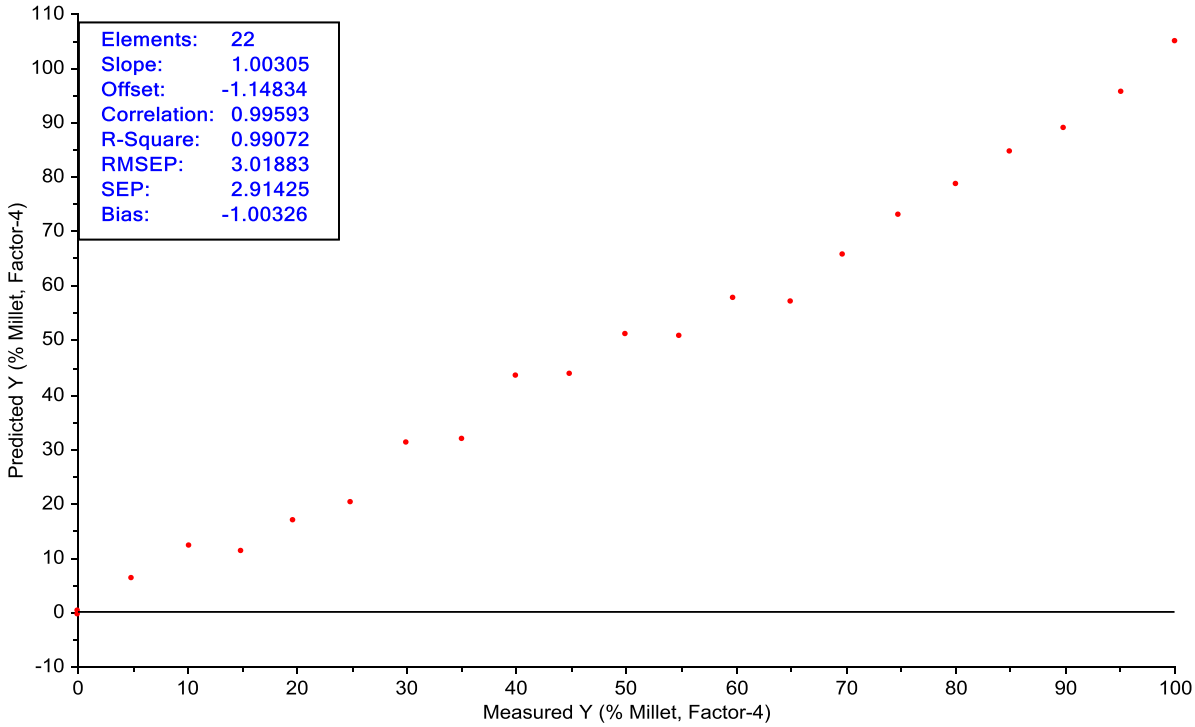


Figure 4.16 Regression graph of independent set (*n* = 22) of samples. Modelled percentages (NIR data) are plotted against actual percentage millet present in the ground black pepper.

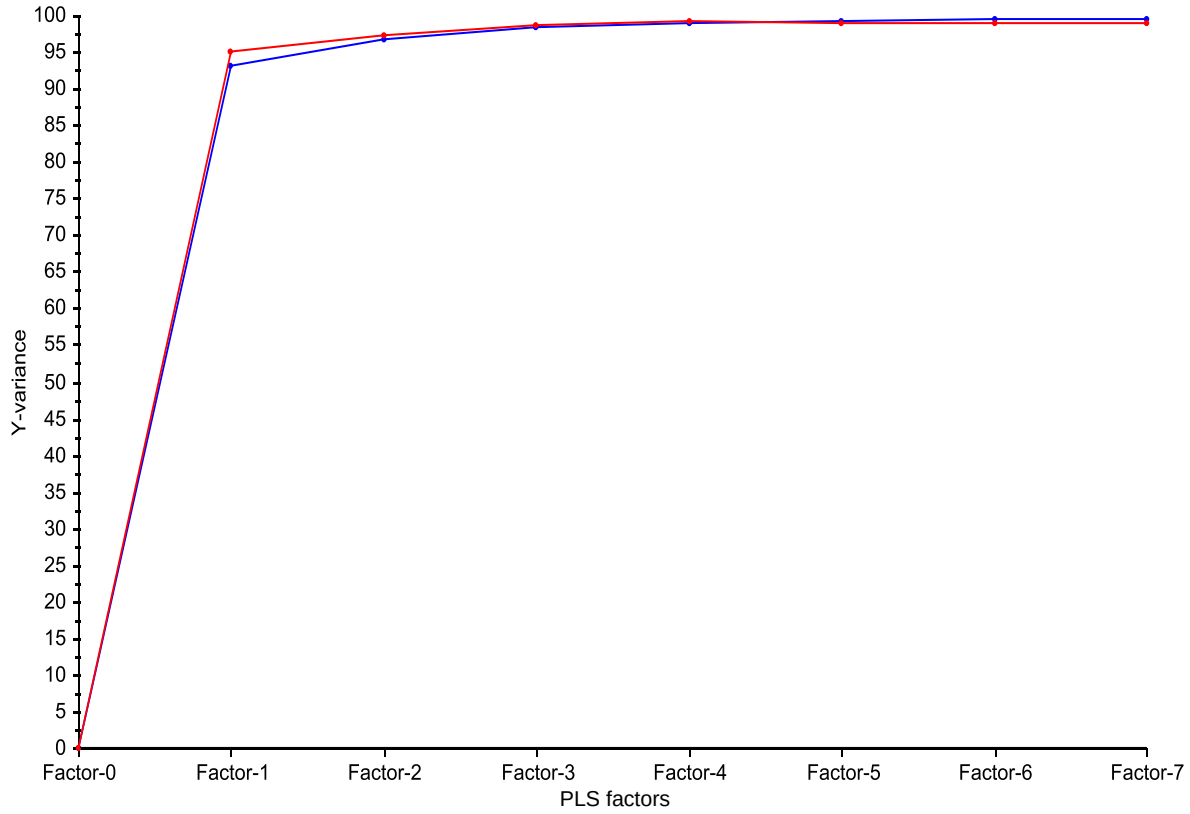


Figure 4.17 Residual test set validation (red) and calibration (blue) variance plots for the millet adulterated black pepper samples. The explained variance are plotted against the number of PLS factors.

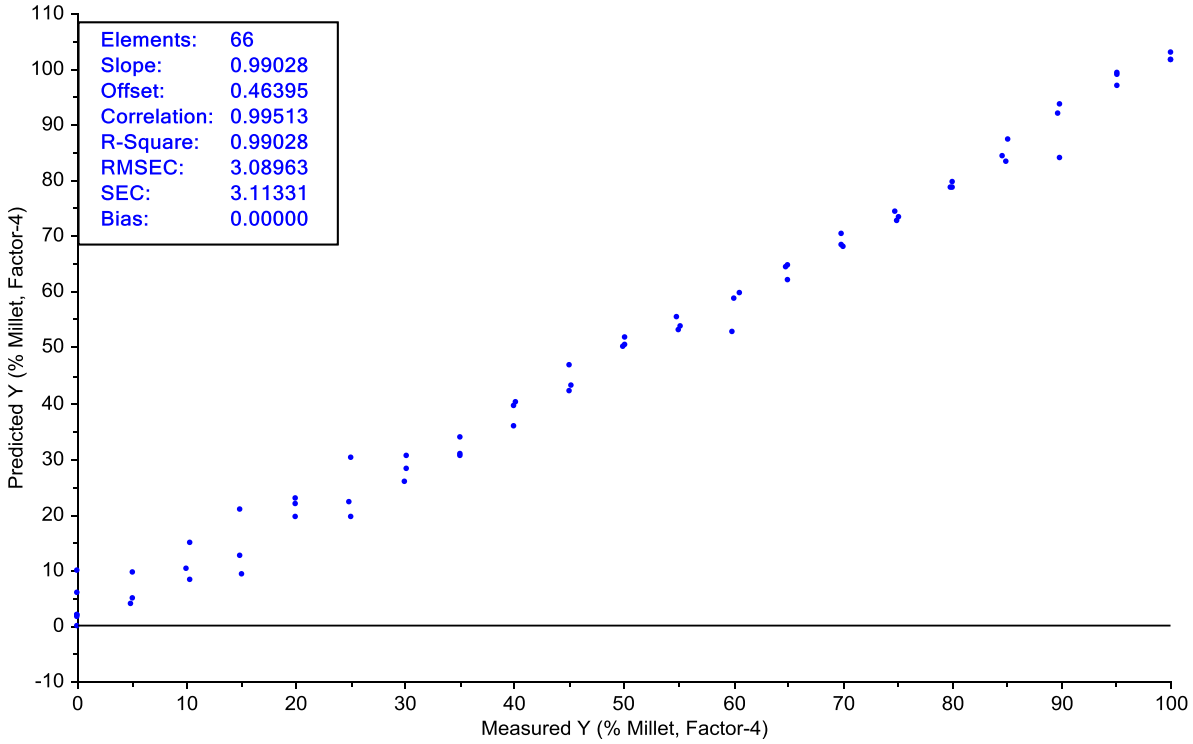


Figure 4.18 Scatter plot of calibration (*n* = 66) samples for the NIR data of the millet adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the modelled percentages (NIR data).

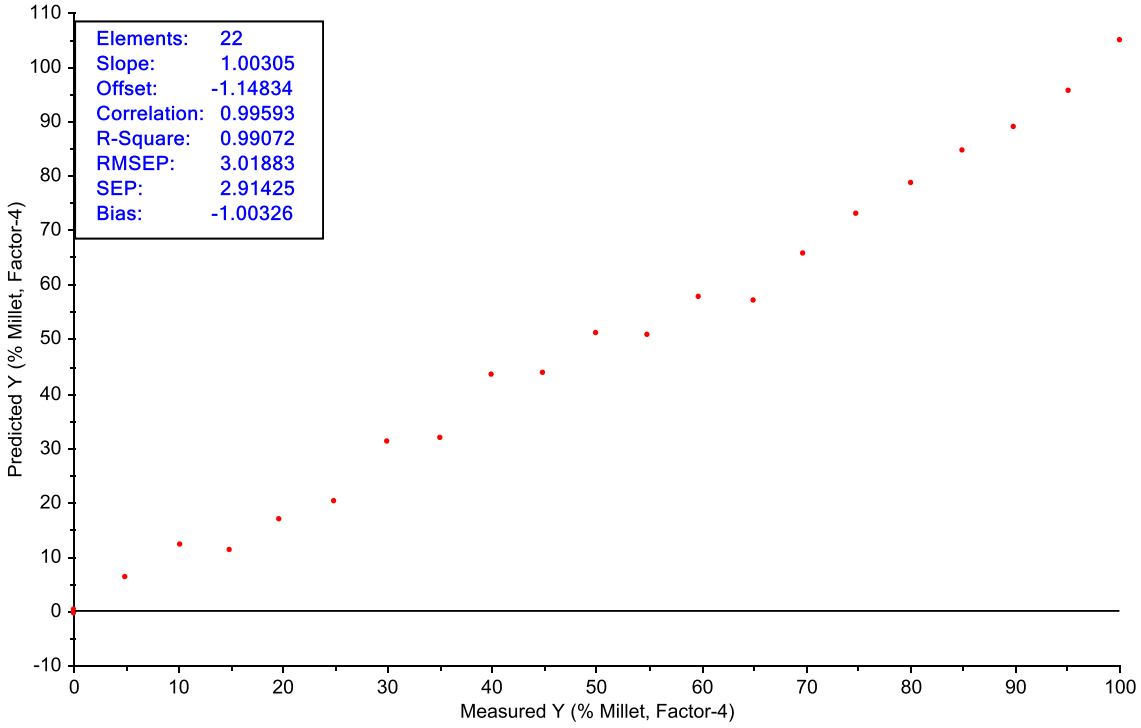


Figure 4.19 Scatter plot of test set validation (*n* = 22) samples for the NIR data of the millet adulterated black pepper. The measured adulterant percentages (w/w) are plotted against the predicted percentages (NIR data).

We suspected that millet flour had a more homogenous particle size distribution than buckwheat flour and therefore the prediction performance for the created models was better. The data used for the development of PLS model 2, were then used for the development of an additional model without applying full cross-validation (**Fig. 4.17**). Test set validation of model 2 produced similar results as when full cross-validation was applied. Most of the Y-variance (99%) was explained within four PLS factors and a relatively low *RMSEC* (3.09% w/w) was obtained (**Fig. 4.18**). The created PLS model also performed well in predicting the test set (r^2= 0.99, *RMSEP*= 3.02) (**Fig. 4.19**).

Global PLS models

PLS models were created to quantify the presence of adulterant irrespective of the type (buckwheat or millet). The regression statistics are summarised in **Table 4.3**. The obtained regression results for the 8 models varied; the best model was obtained with 4 PLS factors (Model 8). An *RMSEC* of 3.71% w/w was obtained and 99% of the Y-variance was explained with 4 PLS factors. The resultant *RMSEP* of the test set was 3.32% w/w and the obtained r^2 value was 0.99. All the created PLS models had high R^2 values ranging between 0.99 and 0.96 for calibration and test set validation. The *RMSEC* values ranged between 3.70 and 6.40% w/w for calibration and the *RMSEP* values ranged between 3.32 and 5.96% w/w for test set validation. The number of PLS factors did not exceed 4 PLS factors, thus our results are relatively good when compared to other research where a higher number of PLS factors were used to explain the Y-variance (Hein *et al.*, 2010).

Regression results obtained for millet adulterated black pepper samples were better compared to that of buckwheat adulterated samples. We suspected that the predictive performance of the PLS models based on the buckwheat adulterated black pepper samples were influenced by physical particle size differences existing between the samples. The mill used had a sieve size of 500 µm and therefore the ground samples had a particle size of 500 µm or smaller. Compositional differences (e.g. cellulose content) between buckwheat and millet can influence the interaction of NIR radiation. In addition, ground black pepper also contains a variety of black pepper material (e.g. epicarp) that can prevent the interaction of NIR radiation with other related black pepper components (e.g. piperine). This adds to the particle difference effect, that influences the quality of the PLS models. Ground millet had a more homogenous particle size distribution and this allowed better calibration development. Similar problems were experienced when powdered wood was analysed using NIR spectroscopy (Hein *et al.*, 2010). The NIR spectra used for calibration development were averaged spectra of sample hyperspectral images. The arrangement of the particles inside each sample in the hyperspectral images varies and therefore causes over or under estimation of the adulterant content present in each sample. The application of Savitzky Golay 2nd derivative as a preprocessing method did aid in reducing the spectral variation, which allowed better spectral interpretation.

Table 4.3 PLS regression results obtained for global data, irrespective of adulterant present (millet or buckwheat)

PLS Model	Preprocessing	Training set (Calibration)					Test set (Validation)			
		n	R^2	*RMSEC (% w/w)*	*Bias*	*PLS factor*	*n*	r^2	*RMSEP (% w/w)*	*Bias*
Model 1[a]	Sav.Gol_2nd derv._9point av._3nd poly. order[*]	154	0.98	4.35	0	3	22	0.99	3.76	2.29
Model 2[a]	Sav.Gol_2nd derv._9point av._3nd poly. order	154	0.99	3.76	0	4	22	0.99	3.69	-1.45
Model 3[a]	Sav.Gol_2nd derv._9point av._3nd poly. order	154	0.96	6.40	0	2	22	0.98	4.66	0.56
Model 4[a]	Sav.Gol_2nd derv._9point av._3nd poly. order	154	0.96	6.36	0	2	22	0.98	4.79	-3.55
Model 5[b]	Sav.Gol_2nd derv._9point av._3nd poly. order	154	0.98	4.72	0	3	22	0.99	3.03	0.62
Model 6[b]	Sav.Gol_2nd derv._9point av._3nd poly. order	154	0.98	3.88	0	4	22	0.96	5.96	0.52
Model 7[b]	Sav.Gol_2nd derv._9point av._3nd poly. order	154	0.96	6.34	0	2	22	0.98	4.73	3.12
Model 8[b]	Sav.Gol_2nd derv._9point av._3nd poly. order	154	0.99	3.70	0	4	22	0.99	3.32	-1.59

[*] Savitzky Golay 2nd derivative, 9 point averaging, 3rd polynomial order

[a] Millet flour

[b] Buckwheat flour

MIR spectra

The obtained MIR raw and multiplicative scatter corrected (MSC) spectra are illustrated in **Fig. 20**. Spectra were collected between 576 and 3999 cm^{-1}, but 1699 – 2389 cm^{-1}, 2393 – 2705 cm^{-1} and 3594 – 3999 cm^{-1} were excluded from the spectral region, since it did not include any sample specific information. Particle size differences caused a large amount of spectral variation and MSC effectively removed it.

Broad water peaks were observed in the region between 3462 and 3195 cm^{-1}. The MIR region is very sensitive to moisture and even though the samples were dried prior to MIR analysis, it is possible that the ground black pepper absorbed some moisture during the analysis; powdered substances are very susceptible to moisture uptake. The MSC spectra illustrate sharp peaks in the regions between of 2929 – 2806 cm^{-1}, 1574 -1307 cm^{-1} and the sharpest peaks were identified at 1240 cm^{-1} and 975 cm^{-1}. The region between 2929 and 2806 cm^{-1} arises from C-H stretching of CH_3 and CH_2 (Di Egidio *et al.*, 2010). Cellulose and protein related groups usually absorb in the region between 1574 and 1307 cm^{-1} due to deformations of CH_2 and CH as well as amide bond vibrations. The peak at 1240 cm^{-1} arises from twisting and rocking vibrations of $–CH_2$ of piperine (Schulz *et al.*, 2005). The sharpest peak at 975 cm^{-1} arises from $–OCH_3$ bond vibrations and is related to cellulose. Researchers have reported the presence of an anti-symmetric out plane ring stretch of amorphous cellulose at 900 cm^{-1} (C-O) when they studied wheat straw (Michell, 1990). When soft wood cellulose was investigated, a weak band at 971 cm^{-1} was identified arising from C-C and C-O stretching (Wiley & Atalla, 1987). The characteristic black colour of black pepper is due to the dark brown polygonal cells of the epicarp and comprises cellulose (Woodman, 1941).

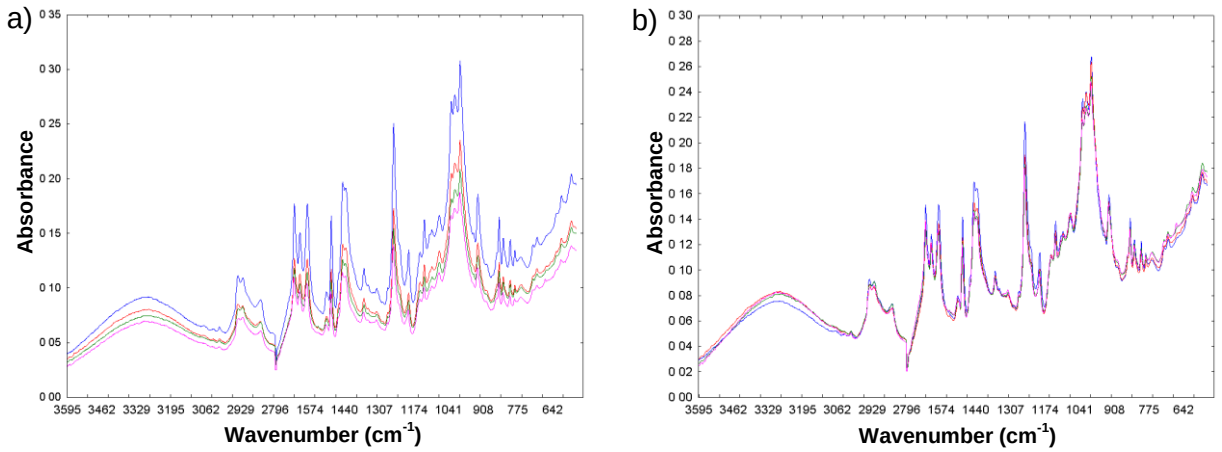


Figure 4.20 a) MIR raw data of four different black pepper samples and **b)** the resultant MSC corrected spectra.

Fig. 21 depicts the raw and MSC spectra of millet and buckwheat flour. Spectral variation can be observed in the raw spectra, but it is reduced after applying MSC (**Fig. 21b**). Broad water

peaks are observed in the raw and MSC spectra in the region between 3462 and 3195 cm^{-1}; this is due to moisture uptake of the samples during MIR analysis.

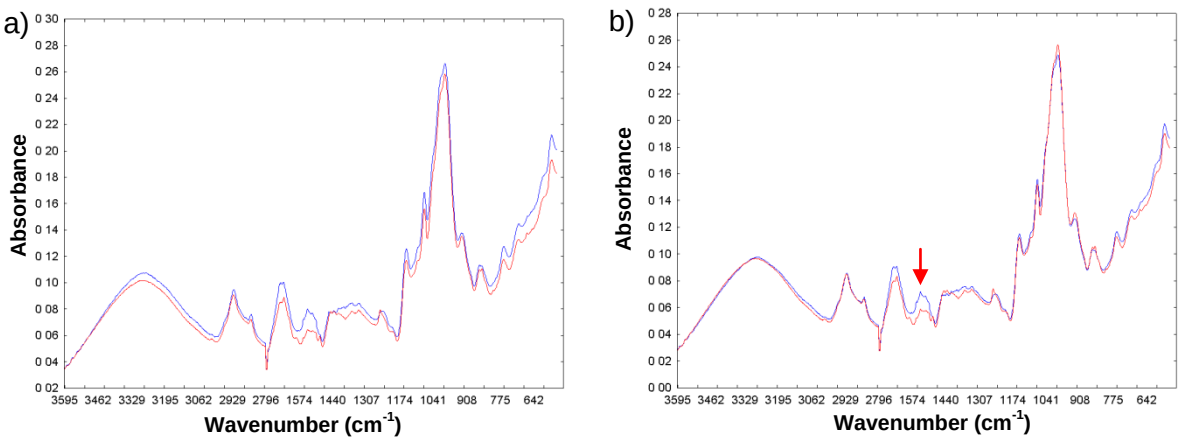


Figure 4.21 a) Raw MIR data of pure millet (red) and buckwheat (blue) flour and **b)** MSC corrected spectra. Red arrow indicates spectral difference at 1574 cm^{-1}.

Various peaks are observed in the MSC spectra of millet and buckwheat. These include 576 cm^{-1}, 974 cm^{-1}, 1174 cm^{-1}, 1243 cm^{-1}, 1574 cm^{-1}, 1640 cm^{-1} and 2896 cm^{-1} and the possible assignment of these bands are summarised in **Table 4.4**. Differences between buckwheat and millet flour at 1574 cm^{-1} and 1640 cm^{-1} were protein related. These protein differences were also observed in the NIR region (previously discussed). Most of the key bands identified are related to polysaccharides, this is in accordance with literature.

Carbohydrates are the most abundant chemical component of both millet and buckwheat. Spectral variation was observed in the millet and buckwheat adulterated black pepper samples (**Figs. 22a & 23a**). This deviation was removed after applying MSC to the raw spectra (**Figs. 22b & 23b**). The broad water peak indentified in the MIR spectra of pure millet flour, buckwheat flour and unadulterated black pepper spectra was also identified in the buckwheat and millet adulterated black pepper spectra (**Figs. 22b & 23b**). Sharp spectral peaks were observed in MSC spectra (**Figs. 22b & 23b**). This is due to black pepper, since all the samples except the pure flours, contained black pepper. A discrete difference was noted when buckwheat adulterated MSC spectra were compared to millet adulterated spectra. In the buckwheat adulterated black pepper spectra the previously identified peak at 1574 cm^{-1} was present (**Fig. 22b**) and absent in the millet adulterated black pepper spectra (**Fig. 23b**). This band is indicative of buckwheat presence and therefore it can be used as a descriptor of black pepper adulterated with buckwheat flour.

Table 4.4 Summary of possible MIR band assignments for the identified bands in MIR MSC spectra of millet and buckwheat flour

Wavenumber (cm^{-1})	Assignment	Chemical groups	Researchers
576	555 cm^{-1}, 591 cm^{-1}: skeletal deformation of aromatic rings, substituent group and side chains	Aromatic aminoacids – tyrosine and phenylalanine	Agarwal (2008)
974	969 cm^{-1}: CCH and –HC=CH- deformation	Lignin	Agarwal (2008)
	972 cm^{-1}: -OCH	Pectin	Schulz & Baranska (2006)
1174	1162 cm^{-1}: C-O-C stretching, ring	Cellulose	Schulz & Baranska (2006)
1243	1246 cm^{-1}: cellulosic compounds	Cellulose	Yu (2005)
1574	1543 – 1480 cm^{-1}: amide II (N-H deformation + C-N stretching)	Protein	Schulz & Baranska (2006)
1640	1655 cm^{-1}: amide I (C=O stretching + N-H wagging)	Protein	Schulz & Baranska (2006)
2896	2895 cm^{-1} – very strong C-H stretch	Hemicellulose	Himmelsbach & Akin (1998)

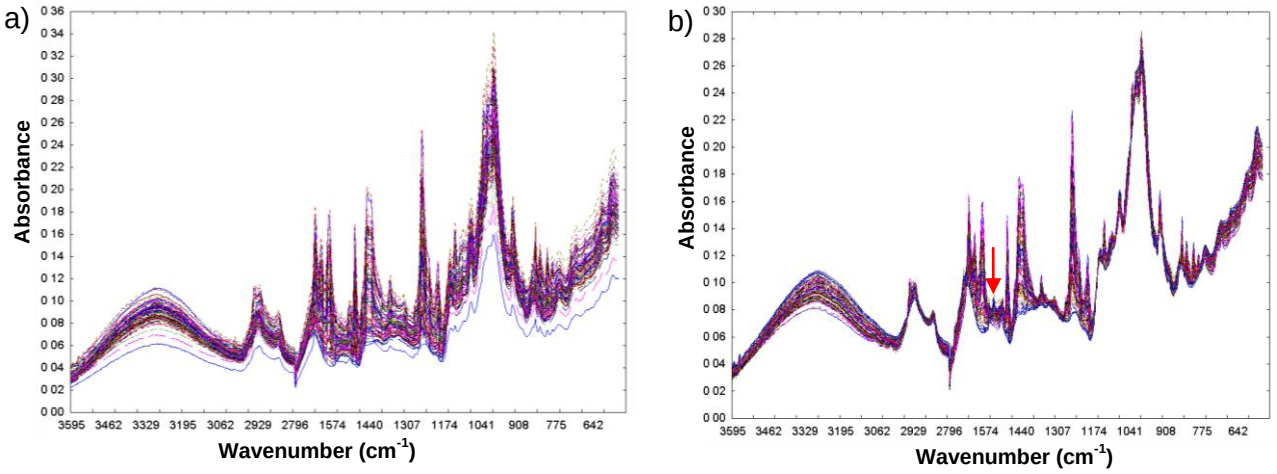


Figure 4.22 a) Buckwheat adulterated black pepper raw MIR spectra and **b)** the resultant MSC corrected spectra. Red arrow indicates the 1574 cm^{-1} peak.

PLS models created for buckwheat and millet adulterated black pepper samples (MIR data)

The regression results obtained for calibration and test validation are summarised in **Tables 4.5**, **4.6** and **4.7**. Outliers in the scatter plots were detected during the calibration development and subsequently removed (not showen). Removing the outliers and reducing the spectral variation induced by particle size differences, did not improve the prediction quality of the created PLS

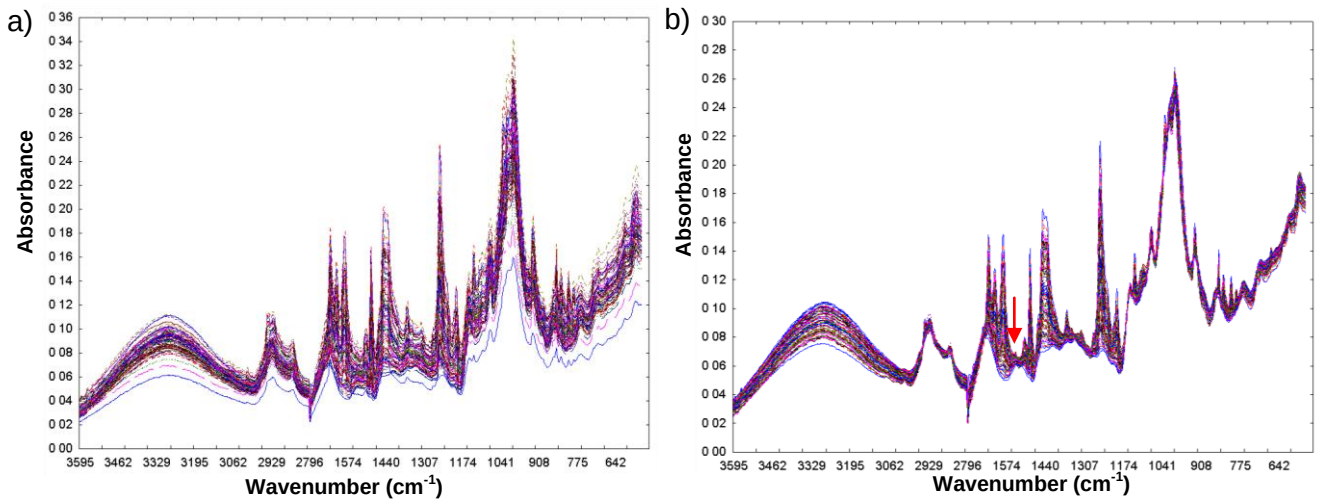


Figure 4.23 a) Millet adulterated black pepper raw MIR spectra and **b)** the resultant MSC corrected spectra. The red arrow indicates the absence of the 1574 cm^{-1} peak.

models. Poor calibration and validation results were obtained compared to PLS models created from NIR data. The R^2 of PLS models created for buckwheat adulterated black pepper ranged between 0.63 and 0.98 whereas better R^2 values (0.94 – 0.98) were obtained for PLS models of millet adulterated black pepper. Results obtained for global PLS models were not better than those obtained for buckwheat and millet adulterated black pepper. *RMSEP values*, ranging between 6.44 and 20.67% w/w was obtained for all test set validations. The number of PLS factors used to explain the Y-variance of created PLS models ranged between 1 and 7.

The obtained regression results for PLS models created from MIR data were not as accurate as those obtained from NIR data. The success rate of any created PLS model based on MIR and NIR data is dependent on a few factors including 1) type of material studied, 2) particle size of the material, 3) spectral variation and 4) penetration depth of NIR and MIR radiation (McKelvy *et al.*, 1998). In this study buckwheat, millet and black pepper were used as our test materials and have various compositional similarities and differences.

In addition, a mixture of these materials was created resulting in a heterogeneous mixture adding to the overall variation. In pharmaceutical studies powdered mixtures are also studied, but pure materials are used (e.g. lactose and salicylic acid) which can reduce the variation caused by sample constituent heterogeneity (Berntsson *et al.*, 2000). The particle size of the studied powdered material is of major concern for both NIR and MIR calibration development (McKelvy *et al.*, 1998; Hein *et al.*, 2010). Varying particle size induces light scattering and therefore spectral variation, but this can be accounted for by using preprocessing techniques (e.g. MSC) (Geladi *et al.*, 1985).

Table 4.5 PLS regression statistics obtained from multiplicative scatter corrected (MSC) MIR data of buckwheat adulterated black pepper

PLS Model	Preprocessing	Training set (Calibration)					Test set (Validation)			
		n	R^2	*RMSEC (% w/w)*	*Bias*	*PLS factor*	*n*	r^2	*RMSEP (% w/w)*	*Bias*
Model 1	MSC	62	0.98	5.44	0	6	21	0.71	16.21	9.37
Model 2	MSC	62	0.86	11.37	0	2	21	0.93	7.80	0.74
Model 3	MSC	63	0.63	18.21	0	1	20	0.92	8.50	-1.07
Model 4	MSC	62	0.96	5.80	0	7	21	0.74	15.54	-9.50

Table 4.6 PLS regression statistics obtained from multiplicative scatter corrected (MSC) MIR data of millet adulterated black pepper

PLS Model	Preprocessing	Training set (Calibration)					Test set (Validation)			
		n	R^2	*RMSEC (% w/w)*	*Bias*	*PLS factor*	*n*	r^2	*RMSEP (% w/w)*	*Bias*
Model 1	MSC	59	0.96	5.44	0	2	21	0.84	12.02	0.98
Model 2	MSC	60	0.94	7.39	0	1	21	0.95	6.66	-2.93
Model 3	MSC	60	0.94	6.93	0	1	21	0.85	11.70	-0.54
Model 4	MSC	60	0.97	4.63	0	3	21	0.54	20.67	-5.53

Table 4.7 PLS regression results obtained for global data after multiplicative scattering correction (MSC)

PLS Model	Preprocessing	Training set (Calibration)					Test set (Validation)			
		n	R^2	*RMSEC (% w/w)*	*Bias*	*PLS factor*	*n*	r^2	*RMSEP (% w/w)*	*Bias*
Model 1 (millet)	MSC	146	0.75	15.12	0	1	21	0.63	18.27	-0.37
Model 2 (millet)	MSC	146	0.74	15.46	0	2	21	0.94	6.98	-0.67
Model 3 (millet)	MSC	146	0.75	15.03	0	2	21	0.84	12.01	-0.54
Model 4 (millet)	MSC	146	0.76	14.83	0	2	21	0.80	13.45	2.80
Model 5 (buckwheat)	MSC	147	0.74	15.45	0	2	20	0.98	6.44	-0.02
Model 6 (buckwheat)	MSC	146	0.89	10.18	0	7	21	0.56	19.94	-15.99
Model 7 (buckwheat)	MSC	146	0.81	13.40	0	3	21	0.94	7.24	4.36
Model 8 (buckwheat)	MSC	146	0.73	15.82	0	1	21	0.78	14.24	7.05

Particle size complicates the calibration process, since NIR and MIR radiation interacts with the irregular particles of the powders and therefore unwanted information are captured inside the resultant spectra (Hein *et al.*, 2010). This factor can be dealt with by employing better sample preparation. Spectral variation is typically caused by 1) inhomogeneous illumination, 2) instrumental non-linearities and 3) specular reflection (Naes *et al.*, 2002). The first two points are of a technical nature and can be improved by technological advances whereas point 3 is based on sample preparation and can be dealt with practically. The NIR region of the electromagnetic spectrum are of a longer wavelength range (1000 – 2498 nm) then the MIR region (576 – 3999 cm^{-1}) and therefore the penetration depth in the NIR region are deeper into the material then MIR region (Geladi, 2008). The reported penetration dept in the MIR region is in the range of few 100 µm at 3500 nm and higher values have been obtained for polymers (0 - 2.5 mm) and exicipients (0.025 – 0.18 mm) in the NIR region (Van den Broek, 1997; Geladi, 2008).

A low sample volume was used during MIR analysis of the adulterated black pepper powder. This influenced the MIR based PLS regression results, since the percentage adulterant (millet or buckwheat) sensed by the instrument is not representative of the actual amount present in the ground black pepper. The varying moisture in the laboratory environment can also be a contributing factor to the obtained results for MIR based PLS models.

Conclusion

The results obtained from PLS models created with NIR data were superior to those obtained for MIR data. MIR radiation only interacts with a limited amount of material due to penetration depth restrictions; therefore PLS modelling was greatly influenced. However, well resolved spectral signals were identified in MIR spectra, simplifying constituent identification. It is recommended that the ATR based technique should rather be employed in qualitative studies, since minimal sample preparation is needed. The preprocessing methods (Savitzky Golay 2^{nd} derivative for NIR data; MSC for MIR data) did manage to reduce the spectral variation, but particle size differences remains a problem. In addition, second derivative and MSC spectra aided in the process of identifying specific chemical components related to buckwheat, millet and black pepper. This enabled us to identify the presence of millet and buckwheat in ground black pepper. The expected particle size of ground buckwheat, millet and black pepper is ≤ 500 µm (in diameter); we recommend that an extra sieving step should be introduced to the experimental layout for the creation of more homogeneous particle distribution. Various factors influenced the calibration development, but it can be addressed with better sample preparation and advances in instrumental technologies. In this study the NIR data was obtained from hyperspectral images. This introduced the multivariate image regression concept to our study. The results obtained from

multivariate image regression were more promising than to those obtained for PLS-DA modelling (see Chapter 3) illustrating the success of applying multivariate image regression to powdered material. This adds yet another method for analysing complex hyperspectral data. A better understanding of ground millet, buckwheat and black pepper was gained from this study and the highlighted consideration can also be implemented in other studies involving powdered food material. It should be noted that NIR hyperspectral imaging poses great promise in food security, since a large amount of information is obtained from a single sample within a matter of seconds. This can aid in rapid quality analysis of powdered food samples, securing better quality for the consumer.

References

Agarwal, U.P. (2008). Chapter 9: An overview of Raman spectroscopy as applied to lignocellulosic materials. In: *Advances in Lignocellulosic Characterization* (edited by T. Hu). Pp 201 – 225. USA: Blackwell Publishing.

Campbell, C.G. (1997). Buckwheat *Fagopyrum esculentum* Moench. Promoting the conservation and use of underutilized and neglected crops. P. 20. Rome, Italy: Institute of Plant Genetics and Crop Plant Research.

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

Cozzolino, D. & Moron, A. (2004). Exploring the use of near infrared reflectance spectroscopy (NIRS) to predict trace minerals in legumes. *Animal Feed Science and Technology*, **111**, 161-173.

Di Egidio V., Sinelli, N., Giovanelli, G., Moles, A. & Casiraghi, E. (2010). NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation, *European Food Research and Technology*, **230**, 947 – 955.

Dhanya, K. & Sasikumar, B. (2010). Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices. *Current Trends in Biotechnology and Pharmacy*, **4**, 454-489.

Dhanya, K., Syamkumar, S. & Sasikumar, B. (2009). Development and application of SCAR marker for the detection of papaya seed adulteration in traded black pepper powder. *Food Biotechnology*, **23**, 97-106

Downey, G., Briandet, R. Wilson, R.H. & Kemsley, E.K. (1997). Near- and mid-Infrared spectroscopies in food aunthentication: Coffee varietal identification. *Journal of Agricultural and Food Chemistry*, **45**, 4357 – 4361.

Downey, G. (1994). Tutorial review: Quantitative analysis in the near-infrared region. *Analyst*, **119**, 2367 – 2375.

Geladi, P. (2008). Are pixels sample cells? Hyperspectral diffuse near infrared imaging experiments with pinholes. *Journal of Near Infrared Spectroscopy*, **16**, 357 – 363.

Geladi, P. & Kowalski, B.R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, **185**, 1-17.

González-Martín, I., Hernández-Hierro, J. & González-Cabrera, J. (2007). Use of NIRS technology with a remote reflectance fibre-optic probe for predicting mineral composition (Ca, K, P, Fe, Mn, Na, Zn), protein and moisture in alfalfa. *Analytical and Bioanalytical Chemistry*, **387**, 2199-2205.

Guillen, M.D. & Cabo, N. (1997). Infrared spectroscopy in the study of edible oils and fats. *Journal of the Science of Food and Agriculture*, **75**, 1 -11.

Hein, P.R.G., Limu, J.T. & Chaix, G. (2010). Effects of sample preparation on NIR spectroscopic estimation of chemical properties of *Eucalyptus urophylla* S.T. Blake wood. *Holzforschung*, **64**, 45 – 54.

Himmelsbach, D.S., Khalili, S. & Akin, D.E. (1998). FT-IR Microspectroscopic imaging of Flax (*Linum usitatissimum* L.) stems. *Cellular and Molecular Biology*, **44**, 99 - 108.

Kasemsumran, S., Kang, N., Christy, A. & Ozaki, Y. (2005). Partial least squares processing of near-Infrared spectra for discrimination and quantification of adulterated olive oils. *Spectrocopy Letters*, **38**, 839 - 851.

Küpper, L., Heise, H.M., Lampen, P., Davies, A.N. & McIntyre, P. (2001). Authentication and quantification of extra virgin olive oils by attenuated total reflectance infrared spectroscopy using silver halide fiber probes and partial least squares calibration. *Applied Spectroscopy*, **55**, 563 – 570.

Lai, Y.W., Kelmsley, E.K. & Wilson, R.H. (1995). Quantitative analysis of potential adulterants of extra virgin olive oil using infrared spectroscopy. *Food Chemistry*, **53**, 95 – 98.

Landgrebe, D., Haake, C., Hpfner, T., Beutel, S., Hitzmann, B., Scheper, T., Rniel, M. & Reardon, K. (2010). On-line infrared spectroscopy for bioprocess monitoring. *Applied Microbiology and Biotechnology*, **88**, 11 - 22.

McCarty, G.W., Reeves III, J.B., Follett, R.F. & Kimble, J.M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal*, **66**, 640 – 646.

McKelvy, M.L., Britt, T.R., Davis, B.L., Gillie, J.K., Graves, F.B. & Lentz, L.A. (1998). Infrared spectroscopy. *Analytical Chemistry*, **70**, 119 – 177.

Michell, A.J. (1990). Second-derivative FT-IR spectra of native cellulose. *Carbohydrate Research*, **197**, 53 - 60.

Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). *A User Friendly Guide to Multivariate Calibration and Classification.* Pp. 39-54. West Sussex, UK: NIR Publications.

Schulz, H. & Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy*, **43**, 13 - 25.

Sandasi, M., Kamatou, G.P.P., Baranska, M. & Viljoen, A.M. (2010). Application of vibrational spectroscopy in the quality assessment of Buchu oil obtained from two commercially important *Agathosma* species (Rutaceae). *South African Journal of Botany*, **76**, 692 – 700.

Schwanninger, M., Hintertoisser, B., Gradinger, C., Messner, K. & Fackler, K. (2004). Examination of spruce wood biodegraded by *Ceriporiopsis subvermispora* using near and mid infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, **12**, 397 – 409.

Schulz, H., Quilitzsch, R. & Krüger, H. (2003). Rapid evaluation and quantitative analysis of thyme, origano and chamomile essential oils by ATR-IR and NIR spectroscopy. *Journal of Molecular Structure*, **661-662**, 299-306.

Tatavarti, A.S., Fahmy, R., Wu, H., Hussain, A.S., Marnane, W., Bensley, G.H. & Hoag, S.W. (2005). Assessment of NIR spectroscopy for non-destructive analysis of physical and chemical attributes of sulfamethazine bolus dosage forms. *AAPS PharmSciTech*, **6**, E91 – E99

Tremlova, B. (2001). Evidence of spice black pepper adulteration. *Czech Journal of Food Scienc*es, **19**, 235-239.

Van den Broek, W.H.A.M. (1997). Chemometrics in spectroscopic near infrared imaging for plastic material recognition. PhD Thesis, Katholieke Universiteit Nijmegen.

Wang, L., Lee, F.S.C., Wang, X. & He, Y. (2006). Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. *Food Chemistry*, **95**, 529-536.

Wiley, J. & Atalla, R.H. 1987. Bands assignments in the Raman spectra of celluloses. *Carbohydrate Research*, **160**, 113 - 129.

Woodman, A.G. (1941). Pepper. In: *Food analysis: Typical methods and the interpretation of results*, 4th ed. p. 396. New York: McGraw-Hill Book Company, Inc.

Workman, J.J. (2001). Review of chemometrics applied to spectroscopy: Quantitative and qualitative analysis. *The Handbook of Organic Compounds.* Pp. 301-326. Burlington: Academic Press.

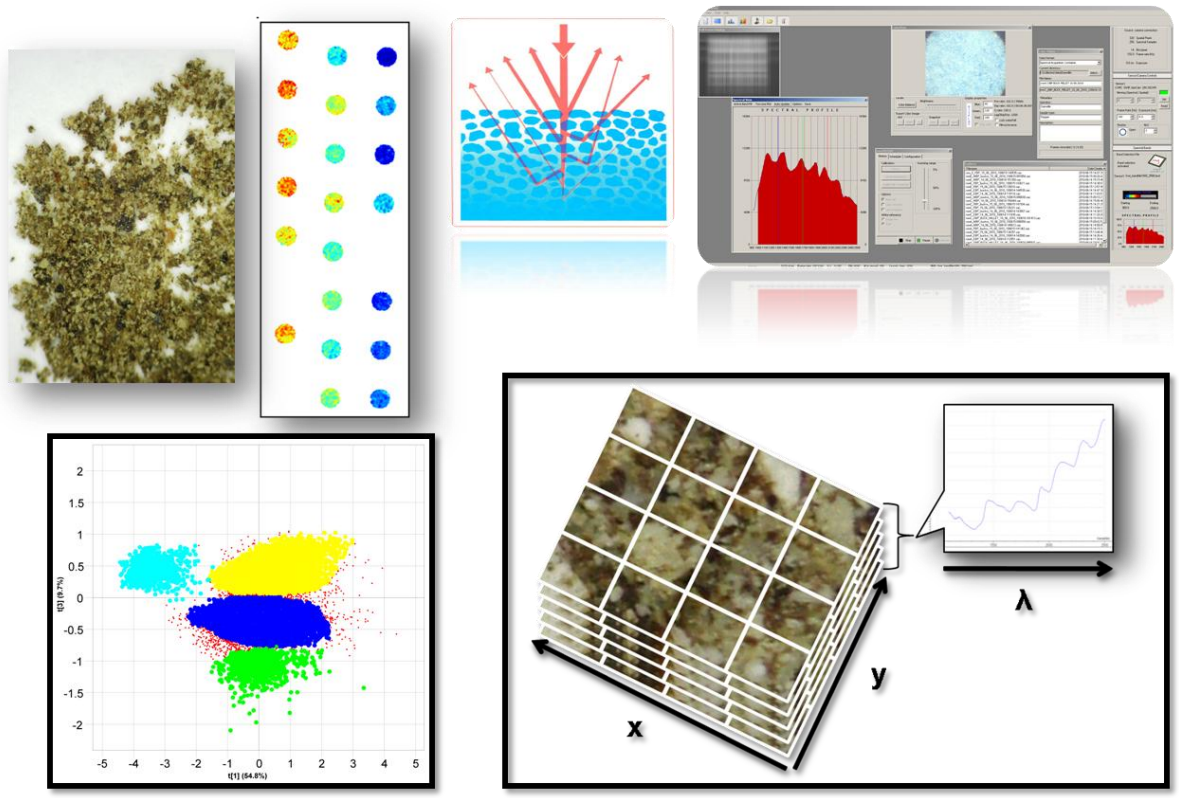
Wu, D., He, Y., Shi, J. & Feng, S. (2009). Exploring near and mid-infrared spectroscopy to predict trace Iron and Zinc contents in powdered milk. *Journal of Agricultural and Food Chemistry*, **57**, 1697 – 1704.

Wu, P. & Siesler, H.W. (1999). The assignment of overtone and combination bands in the near infrared spectrum of polyamide 11. *Journal of Near Infrared Spectroscopy*, **7**, 65 – 76.

Yang, H. & Irudayaraj, J. (2002). Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques. *Journal of Pharmacy and Pharmacology*, **54**, 1247-1255.

Chapter 5

General discussion and conclusions



General discussion and conclusions

The efficacy of near infrared hyperspectral imaging (NIR HSI) and mid-infrared (MIR) spectroscopy in the detection of adulteration of black pepper was assessed. Ground black pepper was adulterated with either buckwheat or millet flour in 5% (w/w) increments from 0–100% and imaged using a sisuChema short wave infrared (SWIR) pushbroom imaging system with spectral range of 1000–2498 nm. The use of PCA allowed unsupervised classification and aided in a better understanding of the studied ground material. An adulterant (millet or buckwheat) dependent gradient was recognised along PC1 and the loading line plot of PC1 revealed four absorption bands (1461, 2241, 2303 and 2347 nm). These absorption bands were related to protein and oil, which in turn relate to adulterant and black pepper respectively. The images of millet and buckwheat adulterated black pepper were combined (mosaic) and analysed together. A distinct separation between buckwheat and millet adulterated black pepper was observed along PC4. The observed separation was caused by a protein difference, but further chemical analysis is needed to confirm whether the observed difference is due to protein content or protein structure. It has been shown that buckwheat flour is lysine rich and contains no glutelin, whereas millet flour is particularly low in lysine and glutelin forms part of the protein fractions (Prakash *et al.*, 1987; Kasaoka *et al.*, 1999; Campbell, 1997; Bejosano & Corke, 1999). Differences in protein content have also been reported (Prakash *et al.*, 1987; Bejosano & Corke, 1999; Kasaoka *et al*., 1999).

Partial least squares discriminant analysis (PLS-DA) was performed on buckwheat and millet adulterated black pepper individually. PLS-DA models calculated for buckwheat and millet adulterated black pepper demonstrated a coefficient of determination (R^2) of 70% and 77%, respectively. The predictive quality of these models was poor. Powdered samples are composed of a range of particles sizes (in this case ≤ 500 µm) and these particles may be bigger or smaller than the pixel size (300 x 300 µm) of the obtained images. A single pixel may therefore contain both black pepper and adulterant complicating pixel classification. The classes created for the PLS-DA models were based on percentage adulterant present in the ground black pepper and therefore an alternative approach was tested. In this alternative approach two classes (1 = black pepper presence; 2 = adulterant presence) were created for the PLS–DA model development. Limited success was achieved with this as it was only able to identify adulterants in samples containing ≥ 30% adulterant. Additionally, the surface should also be considered as a factor influencing the PLS-DA model development, since the sample surface may not be representative of the sample bulk. Black pepper, millet flour and buckwheat flour are all of biological origin, and thus contain similar chemical structures, adding to the difficulty experienced during the calibration and prediction

process. With the successful application of PLS-DA to powdered food material, it can be implemented in online screening procedures.

The feasibility of NIR and MIR spectroscopy in conjunction with PLS regression to quantitatively predict the presence of millet and buckwheat in ground black pepper was evaluated. All samples were analysed with NIR hyperspectral imaging (1000 – 2498 nm) and ATR FT-IR instrumentation (576 – 3999 cm^{-1}). Both averaged raw and second derivative spectra of ground black pepper, millet and buckwheat were evaluated. Black pepper batch similarities were discovered after studying the raw spectra; this gave an indication of particle size differences, which were successfully removed after the application of Savitzky Golay 2^{nd} derivative. Second derivative spectra of black pepper showed the presence of an absorption peak at 2378 nm that was not present in millet and buckwheat spectra. This peak arises from bond vibrations related to ROH groups that are particularly abundant in cellulose and starches. The characteristic black colour of black pepper is due to the epicarp, which primarily comprises cellulose (Woodman, 1941). Millet and buckwheat showed spectral differences at 1424 nm, 1955 nm and 2114 nm. All these peaks typically arise from bond vibrations of starch and protein related structural groups. Second derivative spectra revealed the presence of buckwheat and millet specific absorption bands at 1743 nm, 2112 nm and 2167 nm arising from bond vibrations of cellulose and protein related structures. These peaks can be used as identifiers of the presence of millet and buckwheat since they were not present in ground black pepper.

Raw MIR spectra of unadulterated black pepper, millet and buckwheat contained spectral variation caused by light scattering; multiplicative scatter corrected (MSC) spectra depicted the successful removal of these spectral variation. Broad water peaks were observed in all raw and MSC MIR spectra of unadulterated black pepper, millet and buckwheat. All samples were dried prior to analysis, but the influence of atmospheric moisture should be considered. The raw material was of powdered form and is very susceptible to moisture absorption. MSC spectra of unadulterated black pepper had various peaks at 2929 – 2806 cm^{-1}, 1574 -1307 cm^{-1}, 1240 cm^{-1} and 975 cm^{-1} that arise from bond vibration of cellulose and protein related structures. The highest peak was observed at 975 cm^{-1} that arises from bond vibrations of cellulose, indicating the strong influence of cellulose in unadulterated black pepper. MSC spectra of millet and buckwheat revealed the presence of characteristic strong signals at 576 cm^{-1}, 974 cm^{-1}, 1174 cm^{-1}, 1243 cm^{-1}, 1574 cm^{-1}, 1640 cm^{-1} and 2896 cm^{-1}, but specific differences in peak intensities between millet and buckwheat were observed at 1574 cm^{-1} and 1640 cm^{-1}. These peaks arise from bond vibrations of protein related structures and again place emphasis on protein related differences between millet and buckwheat that was also noted in the NIR region. Moreover, the peak at 1574 cm^{-1} was identified in buckwheat adulterated black pepper MSC MIR spectra which was absent in the MSC MIR spectra of

millet adulterated black pepper. It can therefore be used as an indicator of buckwheat presence in ground black pepper.

NIR based PLS results obtained for millet adulterated black pepper (r^2 = 0.99, *RMSEP* = 3.02% (w/w), PLS factors = 4) were better than buckwheat adulterated black pepper (r^2 = 0.83, *RMSEP* = 12.90% (w/w), PLS factors = 2). Different factors influenced the predictive quality of the created PLS models, but we suspect that particle size played a major role. The particle size differences of buckwheat flour led to increased scattering effects which was manifested as less accurate predictions. Millet flour had a more homogeneous particle distribution which positively influenced the obtained regression results. We suggest that an extra sieving step should be introduced to the experimental layout, which will secure particle homogeneity.

A PLS model for the prediction of adulterant content, irrespective of adulterant type, performed well (Global model: r^2= 0.99, *RMSEP*= 3.32% (w/w), PLS factors= 4). More promising results were obtained using PLS regression when compared to PLS-DA results. It should be noted that an averaged spectrum was calculated of each sample in the NIR hyperspectral images for PLS regression; therefore the resultant data are more representative of the true values making it more objective. In PLS-DA, the model is dependent on the classes assigned, making the predictions more complex, since class overlap is very prominent. The MIR prediction results were poor (r^2 = 0.56, *RMSEP* = 19.94% (w/w), PLS factors = 7) compared to NIR data. It should be noted that only a limited amount of the sample are penetrated with MIR radiation during MIR analysis and therefore complicating the prediction process. Moreover, the penetrated areas are not representative of the percentage adulterant present in the ground black pepper.

With our study new insight was gained and our results can be improved with advances in hyperspectral technology (e.g. spatial resolution). The identified problems can be addressed with improved sample preparation and the introduction of references methods (e.g. scanning electron microscopy (SEM)). Strong protein related differences were identified in the MIR and NIR spectra of millet and buckwheat, but protein content and structural studies are needed to gain better understanding of these materials. NIR and MIR spectra did retain valuable information relating to black pepper, millet and buckwheat; indicating the possible implementation of these spectral regions for qualitative studies. The studied techniques are non-destructive and invasive, allowing the food industry to identify suspect samples and remove them without interrupting the processing of these food materials. It should also be noted that this technique has not been employed in the study of ground millet, buckwheat and black pepper; therefore this study signals the beginning of implementing this technique in routine quality analysis of processed powdered food products. Food authenticity is a very important aspect in the food industry, since it secures consumer buying power and brand

loyalty. NIR HSI poses great potential in the field of authenticity testing of ground food material and adds another dimension to securing food quality.

References

Bejosano, F.P. & Corke H. (1999). Properties of protein concentrates and hydrolysates from *Amaranthus* and buckweat. *Industrial Crops and Products*, **10**, 175-183.

Campbell, C.G. (1997). Buckwheat *Fagopyrum esculentum* Moench. Promoting the conservation and use of underutilized and neglected crops. P. 20. Rome, Italy: Institute of Plant Genetics and Crop Plant Research.

Kasaoka, S., Oh-hashi, A., Morita, T. & Kiriyama, S. (1999). Nutritional characterization of millet protein concentrates produced by a heat-stable α -amylase digestion. *Nutrition Research*, **19**, 899-910.

Prakash, D., Prakash N. & Misra P.S. (1987). Protein and amino acid composition of *Fagopyrum* (buckwheat). *Plant Foods for Human Nutrition*, **36**, 341-344.

Woodman, A.G. (1941). Pepper. In: *Food Analysis: Typical methods and the interpretation of results*, 4th ed. p. 396. New York: McGraw-Hill Book Company, Inc.