Prediction of post-storage quality in canning apricots and peaches using near infrared spectroscopy (NIRS) and chemometrics

Lindie Myburgh

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

MASTER OF SCIENCE IN FOOD SCIENCE

Department of Food Science
Faculty of Agricultural and Forestry Sciences
Stellenbosch University

Study leader: Dr. Marena Manley
Co-study leader: Dr. Elizabeth Joubert

December 2003

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that it has not previously, in its entirety or in part, been submitted at any other university for a degree.

Lindie Myburgh   Date

ABSTRACT

Post-storage quality of the stone fruit, apricots and peaches, is the major factor determining their suitability for canning after cold storage in South Africa. Short harvesting periods and the limited capacity of the factory to process the large quantities of fruit within two days after delivery, necessitates cold storage until canning. Apricots develop internal breakdown, whereas peaches develop internal breakdown accompanied by loosening of the skin and adhesion of the flesh to the stone. The deterioration takes place within the fruit during a cold storage period of one to two weeks. The tendency of the fruit to develop internal defects can, to date, not be identified prior to storage and are only discovered after destoning during canning. Near infrared spectroscopy (NIRS) combined with chemometrics were investigated as a non-destructive method to predict post-storage quality in Bulida apricots and clingstone peach cultivars. Near infrared (NIR) spectra (645-1201 nm), measured on the intact fruit just after harvesting, were correlated with subjective quality evaluations performed on the cut and destoned fruit after cold storage. The cold storage periods for apricots were four weeks (2002 season) and three and two weeks for peach cultivars for the 2002 and 2003 seasons, respectively. Soft independent modelling by class analogy (SIMCA) and multivariate adaptive regression splines (MARS) were applied to the spectral and reference data to develop models for good and poor post-storage quality. The ability of these models to predict post-storage quality was evaluated in terms of recognition (sensitivity) and rejection (specificity) of the samples in independent validation sets. Total correct classification rates of 50.00% and 69.00% were obtained with Bulida apricots, using SIMCA and MARS, respectively. Classification results with apricots showed that MARS performed better than SIMCA and is thus recommended for this application. Total correct classification rates of 53.00% to 60.00% (SIMCA) and 57.65% to 65.12% (MARS) were obtained for data sets of combined peach cultivars within seasons and over both seasons. Additional aspects of fruit quality were investigated to identify possible indices of post-storage quality. Classification trees were used to find correlations between the post-storage quality and the fruit mass, diameter, firmness and soluble solids content (SSC). Among these, fruit diameter and firmness were the major indices of post-storage quality. Accurate predictions of firmness could not be achieved by near infrared spectroscopy (NIRS), making the combination of NIRS and classification trees not yet suitable for predicting post-storage quality. NIRS was further used to predict post-storage SSC within seasons in Bulida apricots and intact peach cultivars. This confirmed sufficient NIR light penetration into the intact fruit and also provided a further application of NIRS for ripeness evaluation in the canning industry. Validations on peach samples obtained

iii

correlation coefficients (r) of 0.77-0.85 and SEP-values of 1.35-1.60 °Brix using partial least squares (PLS) regression. MARS obtained r = 0.77-0.82 and SEP = 1.42-1.55 °Brix. Predictions of SSC in apricots were less accurate, with r = 0.39-0.88, SEP = 1.24-2.21 °Brix (PLS) and r = 0.51-0.82, SEP = 1.54-2.19 °Brix (MARS). It is suggested that the accuracy of SSC measurements, and the subsequent predictions, were affected by the cold storage periods as well as internal variation within the fruit. This study showed that a combination of NIRS and chemometrics can be used to predict post-storage quality in intact peaches and apricots. A small scale feasibility study showed that 4% (R117 720) (apricot industry) and 3% (R610 740) (peach industry) of production losses can be saved if this method is implemented in the South African canning industry. Although it was difficult to assign specific chemical components or quality attributes to the formulation of the storage potential models, important hidden information in the spectra could be revealed by chemometric classification methods. NIRS promises to be a useful and unique quality evaluation tool for the South African fruit canning industry. Several recommendations are made for the canning practices to reduce losses and for future research to improve the current prediction models.

UITTREKSEL

Die kwaliteit van die steenvrugte, appelkose en perskes, is die hoof bepalende faktor vir hul geskiktheid vir inmaakdoeleindes na koelopberging in Suid-Afrika. Die vrugte moet opgeberg word by lae temperature vir een tot twee weke, aangesien die oestydperk kort is en die kapasiteit van die fabriek te beperk is om die groot hoeveeheid vrugte dadelik in te maak. Tydens hierdie opbergingstydperk vind agteruitgang in die vrugte plaas. Dit word in appelkose gekenmerk deur interne verval en in perskes gekenmerk aan interne verval, tesame met 'n los skil en die vaskleef van die vrugvlees aan die pit. Tot dusver, bestaan daar geen metode om hierdie tipe agteruitgang in vrugte voor opberging te identifiseer nie. Dit word eers na opberging opgemerk wanneer die vrugte ontpit word. Naby-infrarooi spektroskopie (NIRS), gekombineerd met chemometriese metodes is gebruik om opbergingspotensiaal in Bulida appelkose en taaipitperske kultivars te bepaal. 'n Korrelasie is gemaak tussen naby-infrarooi (NIR) spektra, gemeet op die heel vrugte voor opberging en subjektiewe evaluering van kwaliteit, geïdentifiseer op die gesnyde vrugte na opberging. Die opbergingstydperke vir perskes was vir drie en twee weke vir die 2002 en die 2003 seisoene, onderskeidelik, terwyl die appelkose vir vier weke opgeberg is. Twee chemometriese metodes, "soft independent modelling by class analogy" (SIMCA) en "multivariate adaptive regression splines" (MARS) is gebruik om die spektra en ooreenstemmende subjektiewe data te kombineer en modelle is ontwikkel vir goeie en swak opbergingspotensiaal. Die vermoë van die modelle om die vrugkwaliteit na die opbergingstydperk te voorspel, is geëvalueer in terme van herkenning en verwerping van vrugtemonsters in onafhanklike toetsstelle. Totale korrekte klassifikasies van 50.00% and 69.00% is verkry vir Bulida appelkose, met SIMCA en MARS, onderskeidelik. Die klassifikasie resultate het gewys dat MARS beter gevaar het as SIMCA en word dus sterk aanbeveel vir hierdie toepassing. Totale korrekte klassifikasies van 53.00% tot 60.00% (SIMCA) and 57.65% tot 65.12% (MARS) is verkry vir gekombineerde perskekultivars tussen seisoene en oor seisoene. Verdere aspekte van vrugkwaliteit is geëvalueer om 'n moontlike indeks van opbergingspotensiaal te verkry. Klassifikasiebome is gebruik om 'n korrelasie te vind tussen kwaliteit na opberging en vrugmassa, deursnee, fermheid en totale oplosbare vastestowwe (TOV). Diameter en fermheid het die meeste gekorreleer met die kwaliteit na opberging. Voorspellings van fermheid deur die gebruik van naby infrarooi spektroskopie (NIRS) was egter nie akkuraat nie. Dus word die kombinasie van klassifikasiebome en NIRS om opbergingspotensiaal te voorspel nie tans aanbeveel nie. NIRS is verder gebruik om TOV te voorspel binne seisoene in heel Bulida appelkose en

v

perskekultivars. Dit is uitgevoer om voldoende NIR ligpenitrasie in die vrugte te bevestig en ook om 'n verdere toepassing van kwaliteitsbepaling (as indeks van soetheid en rypheid) vir die inmaakindustrie te verskaf. Validasies is op perskemonsters uitgevoer en korrelasiekoëffisiente (r) van 0.77-0.85 en voorspellingsfoute van 1.35-1.60 °Brix is verkry met "partial least squares" (PLS) regressie. MARS het r = 0.77-0.82 and voorspellingsfoute = 1.42-1.55 °Brix verkry. Die akkuraatheid van die TOV meetings en gevolglike voorspellings is waarskynlik beïnvloed deur interne variasie binne die vrugte sowel as die opbergings tydperke wat verloop het tussen metings. Hierdie studie wys dat NIRS en chemometriese metodes wel gebruik kan word om opbergingspotensiaal in heel perskes in appelkose te voorspel. 'n Kosteberekening het gewys dat besparings van 4% (R117 720) (appelkoos industrie) en 3% (R610 740) (perske industrie) moontlik is indien NIRS en MARS geïmplementeer word. Alhoewel dit moeilik was om spesifieke chemiese komponente en sekere kwaliteitsaspekte aan die ontwikkeling van die modelle te koppel, is belangrike verborge informasie in die spectra uitgebring deur chemometriese metodes. NIRS beloof om 'n bruikbare en unieke kwaliteitskontrole maatstaf te wees vir die Suid-Afrikaanse inmaakindustrie. Verskeie aanbevelings is gemaak vir die inmaakpraktyke om verliese te voorkom en ook vir toekomstige navorsing om die huidige klassifikasiemodelle te verbeter.

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 presents a compilation of manuscripts where each chapter is an individual entity and some repetition between chapters has, therefore, been unavoidable.

ACKNOWLEDGEMENTS

Chapter 1

Introduction

Approximately 45 000 tons of apricots and 130 000 tons of peaches are received for canning purposes annually in South Africa (Victor, 2002). These canned products are mainly exported, as South Africa only has a relatively small local market (Victor, 2002).

The canning industry suffers production losses to the value of R12.4 million annually in respect of canned apricots and peaches (Steyn, 2003). This loss in production is the result of fruit deterioration which occurs during cold storage rendering the fruit unsuitable for canning. Cold storage is unavoidable owing to the short harvesting period for the fruit, as well as the limited capacity of the primary canning plant in Ashton to process the large amount of fruit immediately (Victor, 2002).

The main peach and apricot producing areas in South Africa are Swellendam, Robertson, Montagu, Outshoorn, Joubertina and Ladysmith in the Western Cape Province, as well as Kimberley in the Northern Cape (Smith & Von Mollendorff, 1999). The peach cultivars most widely used in South Africa for canning purposes are Kakamas, Neethling, Sandvliet, Western Sun, Keisie, Woltemade and Goudmyn (Victor, 2002). The main apricot cultivar grown for canning purposes is Bulida (Victor, 2002).

Internal breakdown, a physiological disorder in stone fruit, results in the largest production losses for the South African peach and apricot canning industry (Victor, 2002). Internal breakdown describes changes in flesh texture of the fruit, a loss of structural integrity, discolouration (or browning) of the internal tissue and increased susceptibility to decay (Morris, 1982). Internal breakdown is also referred to as wooliness in peaches and is characterised by a loss in juiciness (Ben-Arie & Lavee, 1971). In apricots, internal breakdown is referred to as gel breakdown (Eksteen & Combrink, 1994). The defect is characterised by a gelatinous texture and a typical water-soaked appearance, accompanied by slight reddening to browning (Fernández-Trujillo et al., 1998). It is also known as chilling injury (CI), as it usually occurs among susceptible cultivars during two to four weeks of cold storage (at temperatures below 10°C) (Lill et al., 1989). The causes of internal breakdown are not clear and several theories exist to explain the phenomenon (Artés et al., 1996, Luza et al., 1992; Ben-Arie & Lavee, 1971).

No objective procedure exists at present to identify potential internal breakdown of peaches and apricots before storage, on arrival of the fruit at the factory. Internal breakdown can currently only be identified subjectively after storage, when the fruit is destoned for

1

canning. The degree of internal breakdown is determined visually by experienced workers in the peach production and canning industry (Victor, 2002).

Near infrared spectroscopy (NIRS) can provide rapid and non-destructive quantitative measurements of internal properties in peaches, such as soluble solids content (SSC) (Kawano *et al.*, 1992, 1995; Slaughter, 1995; Peiris *et al.*, 1998; Van Zyl, 2000), titratable acidity, firmness (Ortiz *et al.*, 2001), sweetness index, chlorophyll, sucrose and sorbitol content (Slaughter, 1995).

Successful predictions of SSC were also made for other fruit such as melons, pineapples (Guthrie *et al.*, 1998), apples (Moons *et al.*, 1997, Lammertyn *et al.*, 1998, Ventura *et al.*, 1998; Lu *et al.*, 2000; McGlone *et al.*, 2002a; Zude & Herold, 2001), apricots (Carlini *et al.*, 2000), kiwifruit (McGlone & Kawano, 1998; McGlone *et al.*, 2002b; Schaare & Fraser, 2000), tangerines, grapefruit (Zude & Herold, 2001), mangoes (Schmilovitch, 2000; Saranwang *et al.*, 2003) and prunes (Slaughter *et al.*, 2003).

The limits to the applications of NIRS are not yet evident. NIRS is applicable to the determination of the major organic constituents (such as protein, fat and moisture) in foods rather than trace and inorganic constituents (Osborne *et al.*, 1993). Inorganic constituents and other attributes cannot be directly measured by NIRS as these either do not absorb light in the near infrared (NIR) region or are only present in trace amounts. Information about these elements can, however, be embedded inside the spectra of NIR wavelengths (Murray, 1999). It has been possible to determine some inorganic constituents indirectly, either by way of some organic moiety to which they are bound or by the effect that they have on some absorption bands (Shenk *et al.*, 1979).

It is suggested that post-storage quality in peaches and apricots could be predicted by NIRS. Combined influences of fruit components and attributes such as SSC, pectin, Ca^{2+}, firmness and ripeness could contribute to the storage potential. By analysing the fruit absorption spectra of the entire NIR wavelength range, this information which may be buried within the spectra, can be revealed and possibly lead to a means to predict post-storage quality.

The aim of this study was to develop non-destructive models to predict the post-storage quality of intact peaches and apricots. These models would be based on the correlation between NIR spectra (obtained prior to cold storage) and subjective evaluations of internal quality (obtained after cold storage). A correlation between the storage potential and quantitative measurements of fruit firmness and SSC would also be considered. Sophisticated chemometric techniques would be used to obtain relevant information from the data collected and to develop robust prediction models.

This study is an extension of similar research on South African clingstone peaches by Van Zyl (2000). The present study was performed on a more in depth and larger scale. Two seasons (2002 and 2003), instead of one, were included to evaluate the repeatability of the results and the prediction ability when seasonal data sets were combined in the models. Larger sample sets were used for more representative results and robust prediction models. The spectral measurements also included several wavelengths from the visible range and additional chemometric methods were included. The prediction of post-storage quality of canning apricots by NIRS has not been investigated before. This study investigated this possibility and provides information and suggestions for future research.

References

Artés, F., Cano, A. & Fernandez-Trujillo, J.P. (1996). Pectolytic enzyme activity during intermittent warming storage of peaches. *Journal of Food Science*, **61**, 311-313.

Ben-Arie, R. & Lavee, S. (1971). Pectic changes occurring in Elberta peaches suffering from woolly breakdown. *Phytochemistry*, **10**, 531-538.

Carlini, P., Massantini, R. & Mencarelli, F. (2000). Vis-NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. *Journal of Agricultural and Food Chemistry, **48**, 5236-5242.

Eksteen, J. & Combrink, C. (1994). *Manual for the identification of postharvest disorders of pome and stone fruit*. P. 7. South Africa: Stellenbosch Institute for Fruit Technology, Agricultural Research Council.

Fernandez-Trujillo, J.P., Cano, A. & Artés, F. (1998). Physiological changes in peaches related to chilling injury and ripening. *Postharvest Biology and Technology*, **13**, 109-119.

Guthrie, J., Wedding, B. & Walsh, K. (1998). Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy*, **6**, 259-265.

Kawano, S., Watanabe, H. & Iwamoto, M. (1992). Determination of sugar contents in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. *Journal of the Japanese Society for Horticultural Science*, **61**, 445-451.

Kawano, S., Abe, H. & Iwamoto, M. (1995). Development of a calibration equation with temperature compensation for determining the brix value in intact peaches. *Journal of Near Infrared Spectroscopy*, **3**, 211-218.

Lammertyn, J., Nicolai, B., Ooms, K., Smedt, V. & Baerdemaeker, J. (1998). Non-destructive measurement of acidity, soluble solids and firmness of Jonagold apples using NIR spectroscopy. *Journal of Fruits, Vegetables and Nuts*, **41**, 1089-1094.

Lill, R.E., O'Donoghue, E.M. & King, G.A. (1989). Postharvest physiology of peaches and nectarines. *Horticulture Review*, **11**, 413-452.

Lu, R., Guyer, D.E. & Beaudry, M. (2000). Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *Journal of Texture Studies*, **31**, 615-628.

Luza, J.G., Van Gorsel, R., Polito, V.S. & Kader, A.A. (1992). Chilling injury in peaches: A cytochemical and ultrastructural cell wall study. *Journal of the American Society of Horticultural Science*, **117**, 114-118.

McGlone, V.A. & Kawano, S. (1998). Firmness, dry matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology*, **13**, 131-141.

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

McGlone, V.A., Jordan, R.B., Seelye, R. & Martinsen, P.J. (2002b). Comparing density and NIR methods for measurement of kiwifruit dry matter and soluble solids content. *Postharvest Biology and Technology*, **26**, 191-198.

Moons, E., Dardenne, P., Dubois, A. & Sindic, M. (1997). Nondestructive visible and NIR spectroscopy for determination of internal quality in apple. In: *Proceedings from the Sensors for Nondestructive Testing International Conference and Tour.* February, 1997. Orlando, Florida.

Morris, L.L. (1982), Chilling injury of horticultural crops: an overview. *HortScience*, **17**(2), 161-162.

Murray, I. (1999). NIR spectroscopy of food: Simple things, subtle things and spectra. *NIR news*, **10**(2), 10-12.

Oritz, C., Barreiro, P., Correa, E., Riquelme, F. & Ruiz-Altisent, M. (2001). Non-destructive identification of woolly peaches using impact response and near infrared spectroscopy. *Journal of Agricultural Engineering Research*, **78**, 281-289.

Peiris, K.H.S., Dull, G.G., Leffler, R.G. & Kays, S.J. (1998). Near infrared spectrometric method for nondestructive determination of soluble solids contents of peaches. *Journal of the American Society of Horticulture Science*, **123**, 898-905.

Saranwang, S. Sornsrivichai, J. & Kawano, S. (2003). Performance of a portable near infrared instrument for Brix determination of intact mango fruit. *Journal of Near Infrared Spectroscopy*, **11**, 175-181.

Schaare, P.N. & Fraser, D.G. (2000). Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidia chinensis*). *Postharvest Biology and Technology*, **20**, 175-184.

Schmilovitch, Z., Mizrach, A., Hoffman, A., Egozi, H. & Fuchs, Y. (2000). Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology*, **19**, 245-252.

Shenk, J.S., Wersterhaus, M.O. & Hoover, M.R. (1979). Analysis of forages by infrared reflection. *Journal of Dairy Science*, **62**, 807-812.

Slaughter, D.C. (1995). Nondestructive determination of internal quality in peaches and nectarines. *American Society of Agricultural Engineers*, **38**, 617-623.

Slaughter, D.C., Thompson, J.F. & Tan, E.S. (2003). Nondestructive determination of total and soluble solids in fresh prune using near infrared spectroscopy. *Postharvest Biology and Technology*, **28**, 437-444.

Smith C. & Von Mollendorff, L. (1999). Apricot (*Prunus armiaca*) production and breeding in South Africa. In: *Proceedings of the XIth International Symposium on the Apricot Culture*, Volume II. Pp. 119-128. May, 1997. Verra-Makedonia, Greece.

Steyn, B. (2003). Tiger Brands: Fruit unit, Ashton. Personal communication.

Van Zyl, A. (2000). The application of Fourier transform near infrared (FT-NIR) spectroscopy in the wine, fruit and dried fruit industries of South Africa. MSc in Food Science Thesis, University of Stellenbosch, South Africa.

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

Victor, W. (2002). Canned Fruit Producers' Association. Personal communication.

Zude, M. & Herold, B. (2001). Optical spectral analysis potential to determine fruit ripeness. *Erwerbs-Obstbau*, **43**, 161-166.

Chapter 2
Literature Review

A. Introduction

Internal breakdown (IB) causes the biggest losses for the South African peach and apricot canning industry (Victor, 2002). Physiological deterioration in these climacteric fruit occur during cold storage, subsequently making the fruit unsuitable for canning purposes.

Near infrared spectroscopy (NIRS) is widely used for non-destructive analysis in the food and beverage industry because of its advantages over conventional chemical methods, in terms of speed, precision and experimental simplicity (Osborne *et al.*, 1993). Near infrared (NIR) spectra of foods show broad bands which comprise overlapping absorptions corresponding to overtones and combinations involving chemical bonds, e.g. C-H, O-H or N-H.

Spectral information of materials is processed by chemometric techniques to build calibration models for future analysis of constituents. Quantitative models can be obtained by regression methods such as multiple linear regression (MLR), principal component regression (PCR), partial least squares regression (PLS) and multivariate adaptive regression splines (MARS). Qualitative (classification) methods, such as MARS and soft independent modelling by class analogy (SIMCA), can provide information about similarities that future samples have to existing models.

B. Internal fruit quality

B.1 Respiratory activity

Fruit can be divided in two groups based on their respiratory patterns, i.e. climacteric and non-climacteric fruit (Laties, 1995). Climacteric fruit includes, e.g. peaches, apricots, apples, bananas, mangos and avocados. They have a very specific respiratory pattern (Figure 1). The respiratory activity and growth are at a minimum just before ripening begins. At this stage the fruit is physiologically mature, i.e. has reached the optimum maturity stage (Ritenour *et al.*, 1997). This means that it is the right stage to harvest the fruit, as it will have the best keeping quality. During the ripening stage the respiratory activity will increase rapidly. Fruit is ripe when it reaches the climacteric peak. The longer this ripening can be delayed, the longer fruit stay fresh after harvest.
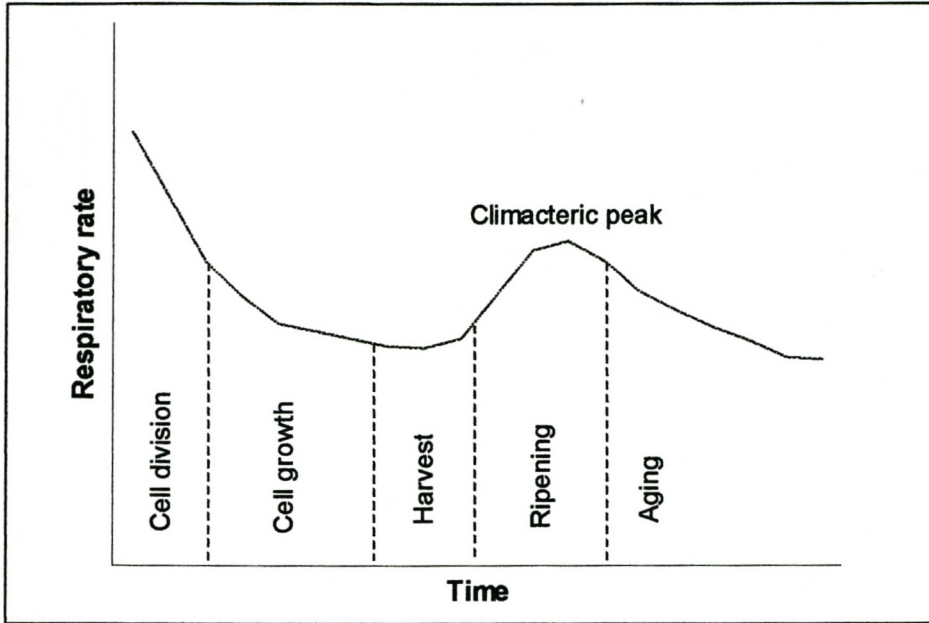
Figure 1. Respiratory activity of climacteric fruit (Fidler *et al.*, 1973).

Ethylene, a colourless, odourless gas produced by fruit, will stimulate ripening (Lill *et al.*, 1989; Ritenour *et al.*, 1997). Stone fruit (peaches, apricots and plums) produce a large amount of ethylene. The negative effects that ethylene production has on fruit include physiological changes and disorders as well as poor post-harvest quality (Kader, 1985). Low temperatures inhibit ethylene production and fruit will stay fresh for longer.

Non-climacteric fruit, such as citrus, grapes and figs, do not ripen after harvest. Their respiratory activity gradually decreases as the fruit ripens (Ritenour *et al.*, 1997), thus they are harvested when they are ready to eat.

In climacteric fruit, ripening is associated with a decrease in firmness, the concentration of metabolic intermediates and changes in skin colour, flesh colour and flavour (Lill *et al.*, 1989). Fructose and glucose decrease and sucrose increases. Malic acid, which forms the largest part of the organic acids, decreases and citric acid increases, resulting in a decline in total organic acids. The ratio of acid content to soluble solids content is closely related to quality, firmness and flesh colour.

The ageing stage (senescence) is when the respiratory activity decreases dramatically. The fruit quality deteriorates and it is also more susceptible to microorganisms, which cause the fruit to become rotten (Brady, 1987; Combrink, 1996)

B.2 Harvesting

The stage of harvesting of stone fruit depends on its uses. Fruit which is intended for fresh consumption is harvested before it reaches full flavour and colour and when it is firm enough to withstand transport and handling. Fruit intended for canning is harvested when the fruit is more mature, but still firm and fruit intended for drying is harvested at full ripeness (Manolopoulou & Mallidis, 1999). A very early harvest results in a fruit without its typical taste and aroma, while a late harvest results in the rapid deterioration of the fruit quality (Hulme, 1971). Harvest of fruit intended for canning must take place when the fruit are in the pre-climacteric stage, the production of ethylene has started and the metabolic changes for maturity take place (Pretel *et al.*, 1993). The characteristics used to measure the proper stage of harvesting include size, colour, symmetry along the suture, firmness, tenderness and sugar content (Reid, 1985).

The harvest maturity in peaches has a major influence in postharvest quality (Lill *et al.*, 1989). In a study of clingstone peaches, Kader *et al.* (1982) showed that upon harvesting, more mature fruit has a softer flesh, improved flesh colour (less green), higher ascorbic acid and soluble solids content (SSC) and lower titratable acidity (TA) than less mature fruit. Colour and flavour of canned peaches improved with more mature fruit.

8

Apricots are climacteric fruit presenting moderate respiration with a very pronounced ethylene rise (Amoros *et al.*, 1989). Harvest maturity is not uniform on the tree. It is therefore important to harvest the fruit when they reach the same stage of maturity and have uniform quality in order to have a good overall quality. The apricot fruit is very sensitive and delicate and its evolution is very fast during the last stages of ripening (Manolopoulou & Mallidis, 1999). This rapid deterioration causes a problem for apricot commercialisation (Pretel *et al.*, 1993).

B.3 Storage conditions

Generally lower temperatures result in better storage life. The optimum storage temperature for peaches and apricots is 0°C (Mitchell *et al.*, 1974; Anderson, 1979). Different methods of storage include controlled atmospheres (CA), intermittent warming (IW), low pressure storage, delayed storage and controlled humidity (Lill *et al.*, 1989).

Controlled atmosphere is storage under decreased O_2 (1%) and increased CO_2 (5%) levels. Fruit stored under normal atmospheric conditions has a significantly higher respiration rate than fruit stored in CA (Anderson, 1982). Modified atmosphere packaging (a less expensive alternative to CA) has shown to control physiological disorders such as flesh browning, but not wooliness (Zoffoli *et al.*, 2002).

Intermittent warming has been demonstrated to be a good method for alleviating chilling injury or internal breakdown during storage (Lill *et al.*, 1989; Dawson *et al.*, 1995; Artés *et al.*, 1996; Fernández-Trujillo *et al.*, 1998). Intermittent warming is achieved by interrupting the cool storage, usually at regular intervals, and warming the fruit for a short period before continuing cool storage. A possible reason for the effectiveness of IW is that the enzyme, polygalacturonase (PG), activity increases and reaches the levels attained during normal ripening (Ben-Arie & Sonego, 1980). Also, when the fruit is warmed, metabolism asserts itself and the accumulated effects of cool storage are dissipated. If the fruit is not warmed a threshold is passed where irreversible damage occurs (Ben-Arie *et al.*, 1970; Anderson, 1979; Lill *et al.*, 1989).

A combination of IW and CA reduced internal breakdown even more effectively (Anderson, 1982). It is, however, difficult to provide optimum storage conditions for stone fruit due to a variation in the storage requirements among different cultivars as well as the high costs involved in specialised storage (Lotz, 2002).

B.4 *Canning*

Fresh dessert peaches are described as freestone and melting, meaning the flesh is soft and the stone is easily removed from the flesh. In contrast, canning peaches are described as clingstone and non-melting, meaning the stone adheres to the flesh. The fruit is also firmer, allowing it to withstand the canning process (Robertson *et al.*, 1992).

The suitability of the fruit for canning depends on the shape and size of the fruit and the uniformity of maturity. The fruit should be round, not elongated on top and the one half of the fruit should not soften faster than the other half (Manolopoulou & Mallidis, 1999). The fruit should also be uniform in colour. Concerning the internal properties, the fruit should be firm and without any disintegration of the flesh or browning of the flesh. The skin should stay tight around the flesh and the flesh break away evenly from the stone during destoning (Lotz, 2002).

The canning process is as follows: The first stage is destoning, where the fruit is sized and transferred through metal channels under rotating knives to be cut longitudinally along the suture (Blaha, 1983; Mallidis *et al.*, 1993). The stones are removed mechanically and the fruit is lye peeled with NaOH. This also inactivates the fungal (*Rhizopus*) pectinolytic enzymes which contribute to softening of the fruit (Sommer *et al.*, 1984). At least 4% lye at boiling temperature is required for satisfactory peeling (Mallidis *et al.*, 1993).

The packaging materials used in the canning industry are tin cans, glass jars and plastic bags. The tin which dissolves from the can contributes to the stabilisation and brightness of the colour of the product (Andreotti *et al.*, 1979; Mallidis *et al.*, 1990).

Heat processing aims to inactivate endogenous enzymes which can impair the quality of the fruit. It also destroys microorganisms which can grow and spoil the product (Paroz *et al.*, 1973; Claypool, 1974; Dalles & Kehagias, 1980).

The main quality attributes of canned peach and apricot are colour, flavour and most importantly, texture (Katuyama *et al.*, 1971).

B.5 *Defects*

B.5.1 *Defects during storage*

Internal breakdown is a physiological defect in stone fruit. It is a general term applied to describe the change in flesh texture, loss of structural integrity, discolouration (browning) of the internal tissue, increased susceptibility to decay, loss of ability to ripen and a decrease in the overall fruit quality (Morris, 1982). Luza *et al.* (1992) described IB as a lack of juiciness. Although the fruit appears ripe, perception when consumed is not good due to texture degradation. Internal breakdown in South African clingstone peaches is identified by the

loosening of the skin, adhesion of the flesh to the stone and softening (Lotz, 2002). Internal breakdown is also known as chilling injury (CI), as it usually occurs among susceptible cultivars during storage for two to four weeks at temperatures below 10°C (Anderson, 1982; Lill et al., 1989).

It is thought that CI would damage the cell membranes, leading to loss of cell turgor and water binding to cell wall materials. Structural changes in peach cell walls become apparent during early softening stages due to dissolution of the middle lamella and disintegration of the cell wall fibrular material (Luza et al., 1992). Walsh et al. (2002) suggested that in the absence of adequate ethylene to trigger all ripening-related messages needed for softening, the fruit senesces, loses cellular water, and eventually browns.

Pectin gel formation could also be responsible for development of internal breakdown (Ben-Arie & Lavee, 1971; Dawson et al., 1993; Artés et al., 1996). Figure 2 shows the process of the development of internal changes in peaches and apricots under normal and low temperature storage. The most important enzyme activities in the development of IB during storage are those of polymethylesterase (PME) and polygalacturonase (PG) (Artés et al., 1996). PME catalyses the de-methylation of the carboxylic group in C_6 of the galacturonosyl residues of the cell wall. During normal ripening of stone fruits, PG breaks the pectin polysaccharide backbone and causes an increase of soluble pectin polysaccharides (De Haan, 1957; Bartley & Knee, 1982; Huber, 1983). During storage at low temperatures, PG is inhibited, which leads to the insufficient solubilisation of the cell wall pectin and the accumulation of insoluble high molecular weight pectin. This process produces a gel (Ben-Arie & Lavee, 1971; Buescher & Tigchelar, 1975; Buescher & Furmanski, 1978; Ben-Arie & Sonego, 1980; Dawson et al., 1992; Artés et al., 1996), particularly in the presence of Ca^{2+}, keeping the water bound and leading to loss of juiciness or wooliness (Ben-Arie & Lavee, 1971; Dawson et al., 1993).

Figure 3 shows the visible defects which develop in apricots and peaches due to internal breakdown. In the case of peaches and nectarines internal breakdown is known as wooliness or mealiness (Ben-Arie & Lavee, 1971), whereas in apricots it is referred to as gel breakdown (Eksteen & Combrink, 1994). Gel breakdown in apricots appears as a browning around the stone and a softening of the flesh (Ryall & Pentzer, 1982). This browning is of enzymatic origin and can probably be considered as alteration of the cells and a premature senescence of the tissues (Souty & Chambroy, 1994). Gel breakdown develops primarily around the stone, reaching a part of the mesocarp. The disorder is characterised by a gelatinous texture, a typical water-soaked appearance and dark olive-green colour accompanied by slight reddening to browning (Fernández-Trujillo et al., 1998). It becomes apparent after 1-2 weeks

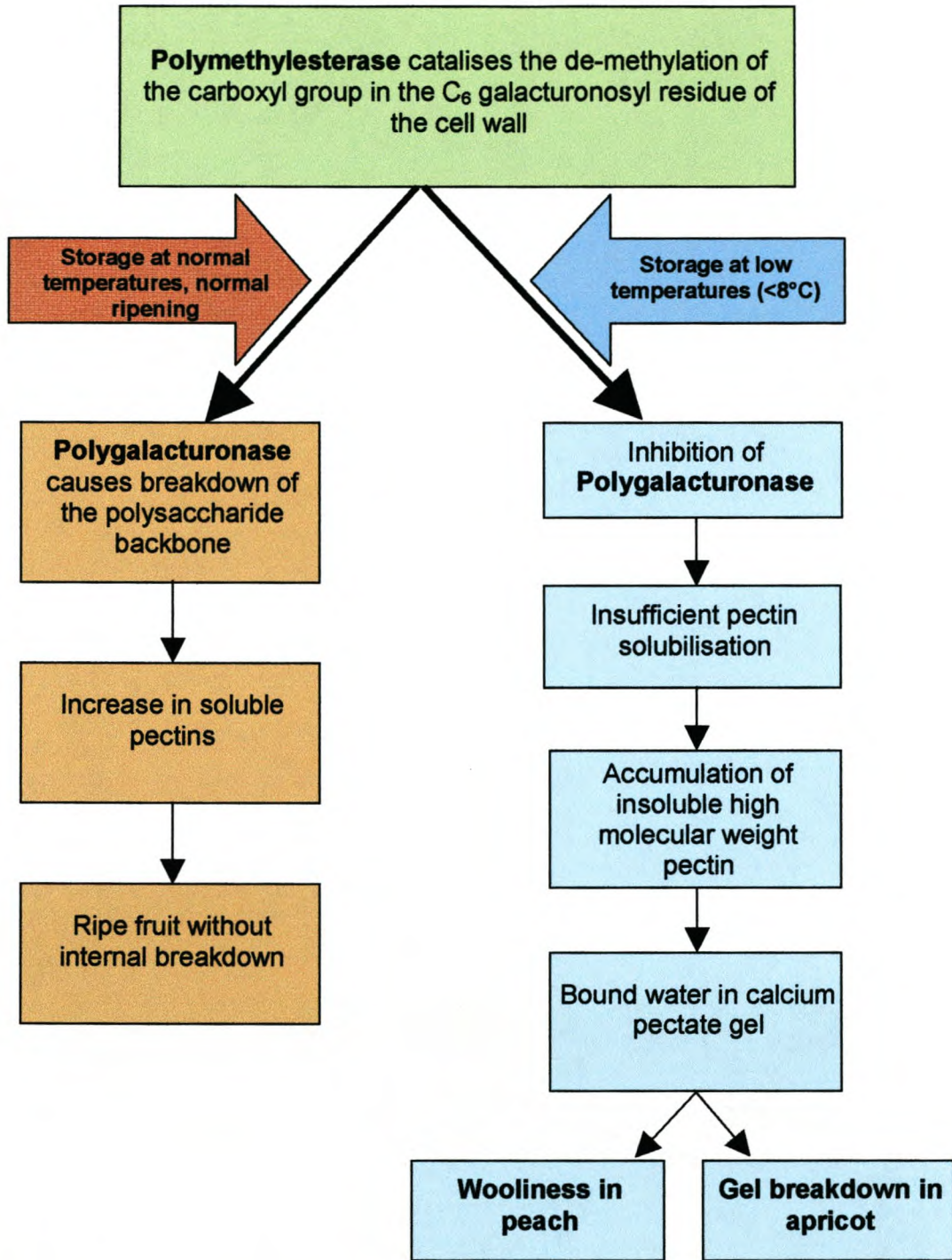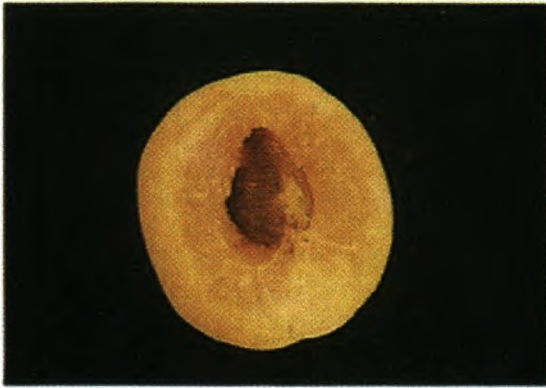**Figure 2.** Flow diagram of normal ripening versus the development of internal breakdown in peaches and apricots.

a)

b)

c)

d)

Figure 3. Internal breakdown in stone fruit. (a) Gel breakdown in an apricot, (b) wooliness (Eksteen & Combrink, 1994), (c) loosening of the skin in a peach and (d) a large cavity after removal of the stone, owing to adhesion of the flesh to the stone in a peach.

13

of cold storage, and especially in atmospheres with low O_2 levels and high CO_2 levels (Souty & Chambroy, 1994).

B.5.2 Defects during processing

Adhesion of the flesh to the stone and improper placement under the knives during the pitting stage results in an uneven and undesirable appearance. Loss of weight ranges from 5-10%, depending on the cultivar, stage of maturity and method of peeling (Blaha, 1983).

Softening (or even disintegration) after heat processing is reported to be caused by heat stable pectolytic enzymes produced in the fruit by fungi (*Rhizopus stolonifer* and *R. arrhizus*) before canning. Pretreatment with lye (NaOH) inactivates these fungal enzymes and can eliminate the softening of the canned product (Sommer *et al.*, 1984).

Enzymatic browning is responsible for the brown discolouration, involving mostly polyphenolic compounds (Luh *et al.*, 1967).

Flesh firmness is the most important determinant of fruit quality (Maness *et al.*, 1992). The average mesocarp firmness declines with the advanced stages of maturity. Internal variation in firmness for the middle and outer regions of the mesocarp occurs in stone fruit and the firmness patterns also differ highly among cultivars.

C. Near infrared spectroscopy

C.1 Background

Near infrared spectroscopy (NIRS) is an instrumental analytical technique which relies on the vibrational properties of different compounds to absorb at different frequencies. This gives rise to an absorption spectrum with different intensities of absorption and can provide complex information about molecules (Osborne *et al.*, 1993; Atkins, 1996).

The near infrared (NIR) spectrum was discovered in 1800 by the astronomer, Herschel, while experimenting with temperature increase associated with the colours of white light when they were dispersed through a glass prism. He found a maximum temperature rise just beyond the red portion (780-1100 nm) of the visible spectrum (Herschel, 1800). Only in 1881 Abney and Festing realised the importance of bonds involving hydrogen in producing NIR absorptions in the range from 700 to 1200 nm (Osborne *et al.*, 1993). In the early 1900's Coblentz recorded absorption spectra from 800 to 2800 nm of materials and showed that certain chemical groupings have characteristic absorption bands. By observing these bands, the chemical constituents can be identified (Butler, 1983). In years that followed, more assignments of absorption bands to functional groups of organic compounds were made

14

(Williams & Norris, 1987; Osborne *et al.*, 1993). Beer's law summarised the work of a number of scientists. It states that the algorithm of the absorption of light was linearly related to the concentration of the absorbing chemical bond. This provided the basis for qualitative spectrophotometric analysis (Butler, 1983).

C.2 The theory of molecular vibration

Near infrared refers to a portion of the electromagnetic (EM) spectrum near the visible portion of the EM spectrum (Figure 4). The NIR region of the EM spectrum extends from about 750 to 2600 nm and is just outside the visible range. The EM spectrum is made up of waves, and the distance between two peaks of a wave is called the wavelength (Figure 5) (Shadow, 2002). The amplitude refers to the interatomic distance covered at the extremes of a vibrating dipole and is dependent upon the amount of energy absorbed by the molecule (Workman, 1993).

When dealing with EM waves, wavelength and energy are closely related. The energy associated with a certain wavelength is given in equation 1.

$$E = \frac{hc}{\lambda}$$ 1

where

E	= energy
h	= Planck's constant (6.626 x 10^{-34} J.s)
c	= speed of light (2.998 x 10^8 m.s^{-1})
λ	= wavelength

Near infrared spectroscopy is based on the interaction of EM waves with molecules and involves the transfer of energy (Shadow, 2002). When light in the form of photons strikes an atom, the photon is absorbed by the electron and the electron will be in a higher energy orbital. Electrons that are in differing energies will absorb different amounts of energy, thus will be absorbing at different wavelengths. Molecules have different characteristics (i.e. energies) and will thus absorb different wavelengths of light. Absorptions occurring in the NIR region are considered to be vibrational absorptions.

When electrons move about atoms, these atoms will be drawn and repulsed from each other, creating a vibration motion. The disturbance of atoms and their bonds causes molecular vibration and is described by Hooke's law (Osborne *et al.*, 1993). Hooke's law

15

Figure 4. The electromagnetic spectrum indicating the near infrared region (Shadow, 2002).

λ = wavelength
A = amplitude

Figure 5. Schematic diagram of EM waves (Shadow, 2002).

states that for a two body harmonic oscillator, the frequency of a vibration (v) is equal to the reciprocal of 2π times the speed of light times the square root of the force constant ($k = 5\times10^5$ dynes.cm^{-1}) times the sum of two masses (m_1, m_2) divided by their product (Equation 2) (Barton, 2002).

$$\tilde{v} \text{ (in cm}^{-1}) = \frac{1}{2\pi c}\sqrt{\frac{k(m_1 + m_2)}{m_1 m_2}}$$

............2

Hooke's law can be used to calculate the fundamental vibrations for diatomic molecules in the mid-IR (Barton, 2002).

Vibrational energies (like all other molecular energies) can be quantised by solving quantum mechanical wave equations, as in equation 3.

$$\Delta E = (v + 0.5)hv$$

............3

where v is the frequency and υ is the vibrational quantum number which may have a value of 0, 1, 2,... Promotion to the first excited state (υ=1) thus requires absorption of radiation energy. Energy associated with transition is given in equation 4 (Osborne et al., 1993).

$$\Delta E = hv = \frac{h}{2\pi}\sqrt{\frac{k}{\mu}}$$

............4

where $\mu = (m_1 m_2)/(m_1 + m_2)$

A transition where υ is equal to 1 is known as a fundamental absorption of molecules. The fundamental absorption is the most intense absorption of energy and occurs in the mid-IR region (2500-15 000 nm) of the spectrum (Barton, 2002).

Near infrared wavelengths are associated with raising the energy within the molecules to the first and second overtones. When υ is greater than 1, the transition is known as an overtone. Quantum theory indicates that the only allowed vibrational transitions are those in which υ changes by one (+/- 1) (Osborne et al., 1993, Bokobza, 1998).

If everything was simple, NIR bands could be calculated from combinations of the mid-IR fundamentals and 2, 3, 4, ..., n times the fundamental frequency to produce overtones. The NIR region, however, is comprised of combination and overtone bands which appear at

17

imprecise multiples of the fundamentals (Barton, 2002). The quantum theory treatment of a harmonic ossilator explains the fundamental absorption bands due to molecular vibration in the IR, but it does not explain the presence of overtone bands in the NIR (Osborne *et al.*, 1993).

The Franck-Condon principle explains how atoms in a vibrating molecule spend most of their configurations in which the kinetic energy is low (Kauzmann, 1957). The photon is most likely to be absorbed when the nuclei are stationary or moving slowly. Excitation resulting from the photon absorption is not transferred to the nuclei immediately, thus the excited stage of the nuclear configuration is close to the low kinetic energy. Therefore transitions between vibrational levels take place where nuclear configurations are both in the same states and also when nuclear kinetic energies are small (Kauzmann, 1957). These small variations cause anharmonicity, which causes overtones at imprecise multiples of the fundamentals (Barton, 2002). Equation 5 gives a better idea of the spectral band position for overtones in the NIR region (Workman, 1993).

$$\bar{v} = \left(\frac{E_{vib}}{hc}\right) = \bar{v}_1 \upsilon - x_1 \bar{v}_1 \left(\upsilon + \upsilon^2\right) \qquad \ldots\ldots\ldots\ldots 5$$

where

υ = an integer number, i.e. 0, 1, 2, 3, …, n,

$x_1 \bar{v}_1$ = the unique anharmonicity constant for each bond

The more vigorous the vibration, the greater the anharmonicity (Barton, 2002). For every molecule it is possible to describe the nature and number of the vibrational modes. Vibration can be either stretching (continuous change in the interatomic distance along the bond axis) or bending (change in bond angle). Bending can be further classified into four types, scissoring (symmetrical in-plane), rocking (asymmetrical in-plane), wagging (symmetrical out-of-plane) or twisting (asymmetrical out-of-plane) (Workman, 1993). Figure 6 shows the vibrational modes indicating bending and stretching.

Each of these vibrational modes may give rise to overtones or combinations in the NIR region (Osborne *et al.*, 1993). Almost all absorptions seen in the NIR region arise from overtones of hydrogenic stretching vibrations involving AH_y functional groups or combinations involving stretching and bending modes of vibration of such groups. Thus, most NIR spectra can be explained by the assignment of the bands to overtones and combinations of fundamental vibrations involving hydrogenic stretching modes (Osborne *et al.*, 1993).

Figure 6. Molecular vibrational modes observed in the NIR region (Workman, 1993).

Absorption bands in the NIR region are broad rather than sharp peaks, due to additive effects (combinations of absorbancies) of two or more bonds at each wavelength. The intensity of the absorbance of energy depends on the degree of stretching or bending (anharmonicity) and decreases with each overtone (Osborne *et al.*, 1993; Bokobza, 1998). Hydrogen atoms are the lightest atoms, thus bonds involving hydrogen vibrate with large amplitude when undergoing stretching.

A NIR spectrum can provide useful information about hydrogen-bearing functional groups in a molecule even though it does not necessarily characterise a complete structure. As a consequence, NIRS may be used primarily for the determination of compounds containing C-H, N-H or O-H groups, such as water, alcohols, phenols, amines and unsaturated hydrocarbons, rather than for molecular structure characterisation (Giangiacomo & Nzabonimpa, 1994).

C.3 Absorption of biological samples

When light interacts with matter it may be absorbed, transmitted or reflected. According to the law of the conservation of energy, the total radiant power incident on a sample must equal the sum of radiant power absorbed, transmitted and reflected (Osborne *et al.*, 1993).

Transmission takes place when radiation passes through a medium other than a vacuum. When the radiation crosses the boundary between two media of different refractive index, a change in the direction of propagation occurs. This is called refraction.

The Beer-Lambert law states that the fraction dP/P of radiant energy P absorbed by a certain thickness of sample is proportional to the concentration of molecules in that thickness (equation 6) (Osborne *et al.*, 1993).

$$Log(P_0/P) = abc \qquad\qquad \dots\dots\dots 6$$

Where P_0 = total radiant power incident on the sample

P = radiant energy absorbed

a = absorptivity coefficient

b = thickness through which radiation passes

c = concentration of molecules in the sample

The fraction of radiation (P/P_0) transmitted by the sample is measured and this is called transmittance. Transmittance can be converted to absorbance (A), which is defined by equation 7.

$$A = \log 1/T = \log (P_0/P) \qquad \ldots\ldots\ldots 7$$

Thus, from the previous two equations,

$$A = abc \qquad \ldots\ldots\ldots 8$$

If the Beer-Lambert law holds, a plot of absorbance against concentration will give a straight line through the origin with slope ab (Osborne $et\ al.$,1993). Since thickness b is known, a may be determined easily. There are, however, deviations from the linear relationship described by the Beer-Lambert law because for example the effect of scattering. Methods for compensating for these deviations include various pre-processing methods.

When a sample is non-absorbing, the incident radiation beam, the reflected beam and the normal to the sample surface, all lie in the same plane (law of mirrors) (Osborne $et\ al.$,1993). The angle of incidence is also equal to the angle of reflection. If the sample surface is matt (Figure 7), then the boundary between the sample and the surrounding medium may be considered to be a series of small surfaces at all possible angles to the normal. If the radiation that is transmitted through the first interface undergoes absorption, the transmitted radiation will be attenuated according to the Beer-Lambert law. This radiation becomes diffused by random reflections, refractions and scatter at further interfaces inside the sample. Particles concentrated on a thick layer give rise to multiple scattering. In biological samples, which absorb radiation weakly, scattering has the effect of intensification of the absorption bands (Osborne $et\ al.$,1993).

C.4 Instrumentation

An NIR instrument normally meets certain requirements. It works over the whole wavelength range of 700-2500 nm (14 300-4000 cm^{-1}). It is capable of resolving a wavelength interval that is narrow compared with an absorption feature. It is sufficiently luminous, and efficient, to enable measurements to be made in a short time. It should also be stable and robust (Osborne $et\ al.$,1993).

The common features of an NIR instrument include the following. All instruments have a wavelength or energy producer. A tungsten-halogen or quartz-halogen light bulb serves this purpose. Detectors, to determine the amount of light transmitted or reflected, are mostly lead sulphide (PbS) detectors. Filters are used to measure certain wavelengths at a time (Shadow, 2002). Filters are installed on a rotating plate and each of them allows a specific wavelength band to pass through. This instrument is calibrated for the determination of a
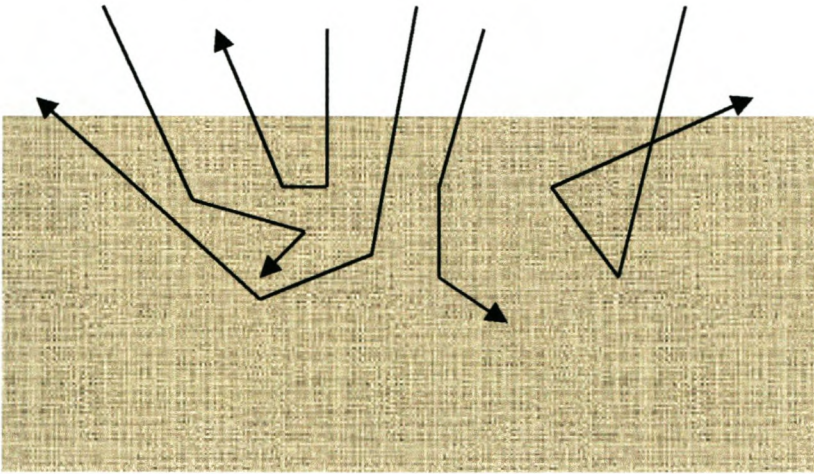
21

Figure 7. In biological samples random reflections, refractions and scatter lead to diffuse reflection (Osborne *et al.*, 1993).

certain constituent using a combination of wavelengths that correlate to the amount of the particular constituent (Giangiacomo & Nzabonimpa, 1994).

The optical operating principle is defined by the method by which the light is spectrally modulated. NIR instruments are dispersive, interferometric or non-thermal. Dispersive systems, such as the monochromators and diode array have wavelengths of light which are separated spatially. Full spectra can be taken with high photometric accuracy and repeatability. This type of instrument also has the advantage of simultaneity of measurement and large reduction of measuring time (Osborne et al.,1993, Giangiacomo & Nzabonimpa, 1994).

Interferometers, such as the Fourier Transform or Michelson interferometer, make use of the complete source spectrum rather than the individual wavelengths generated by grating and prism systems, as used in conventional IR spectroscopy. An interferometer uses a beamsplitter to divide the light from the source into two beams, reflected to a fixed and a moving mirror, respectively. The beams undergo constructive and destructive interference as they recombine due to varying path difference between the two mirrors. This results in an interferogram that contains all the spectral information related to the sample. A Fourier transform algorithm is used to convert the interferogram to a conventional IR or NIR emittance spectrum in order to obtain interpretable information (Griffiths & De Haseth, 1986). Interferometers are mostly used in the laboratory. They show simultaneity of measurement, fast response and high wavelength precision.

The NIR instruments produced in the past were the workhorse instruments for agriculture. The instrument has remained unchanged, but the sampling accessories have greatly improved. An example of such new advances is fiber optic probes, used for sampling of bulk containers and in-line processes. New instruments promise to enhance what has traditionally been done with NIRS and open new opportunities for research and industrial applications (Barton, 2002).

C.5 Advantages and disadvantages of NIRS

Traditional methods of analysis involve extraction and purification of an analyte to facilitate its determination. Analysis using NIRS has the ability to perform simultaneous analysis of several constituents of every sample at the same time and cost as a single determination (Batten, 1998).

One of the greatest uses of NIRS is for the analysis of solid samples. NIRS analysis of solids is non-destructive. Only grinding of the sample is sometimes necessary and no chemical reagents or weighing is needed. This can have significant advantages where the

23

material is either in short supply or expensive to produce. Both clear and opaque liquid materials can also be introduced to the instrument directly. No sample pretreatment is thus needed and a grinding step can be accomplished in under one minute (Day & Fearn, 1982).

NIR spectral data can be measured by transmission through intact materials by diffuse reflectance. This allows for simple, rapid and non-destructive analysis, because sample preparation is avoided (Osborne *et al.*, 1993). The increasing need for real time process control measurements has led to the demand of higher data rates (complete spectra in seconds) which NIRS can offer. The avoidance of chemical pretreatment also makes NIRS more environmentally friendly compared to toxic waste associated with many classical methods such as large scale Kjeldahl nitrogen determinations. Analysis can be performed by technically unskilled personnel at line or automatically on-line. The NIRS optical materials and low absorptivity of water make the NIR region suitable for the analysis of samples containing high proportions of water such as foods and beverages (Osborne *et al.*, 1993). Figure 8 is a summary of the flow of an experiment using NIRS.

The major limitation of NIRS is its dependence on less precise and equally empirical chemical methods of analysis (Batten, 1998). NIRS is a secondary method which relies upon calibration against other standard methods. As such its potential accuracy relates to the inherent accuracy of the technique against which it has been calibrated. As long as care is taken in the original calibration of the instruments, it should always be possible for NIRS to provide a reliable and robust technique.

Another drawback to NIRS has been in the mechanism for selecting the correct wavelengths to apply to a given analysis. Wavelength selection by computer regression and chemical theory should be studied to ensure that the wavelength set relates to the chemical composition of the material being analysed (Day & Fearn, 1982).

The acceptance of NIRS as an analytical technique in its own right has been limited by the inexperience in deriving valid calibrations (Osborne *et al.*, 1993). A large number of similar samples must be analysed for a valid and robust calibration. This fact makes the use of NIR appealing for large scale analysis such as in the dairy or milling industry (Osborne *et al.*, 1993).

The utilisation of NIRS often reflects the reason for acquiring the instrument and the potential savings relative to the costs involved in developing and maintaining calibrations (Batten, 1998).

Instead of competing with conventional analytical methods, NIRS, chemical and other instrumental methods rather complement each other. High resolution and sensitive methods
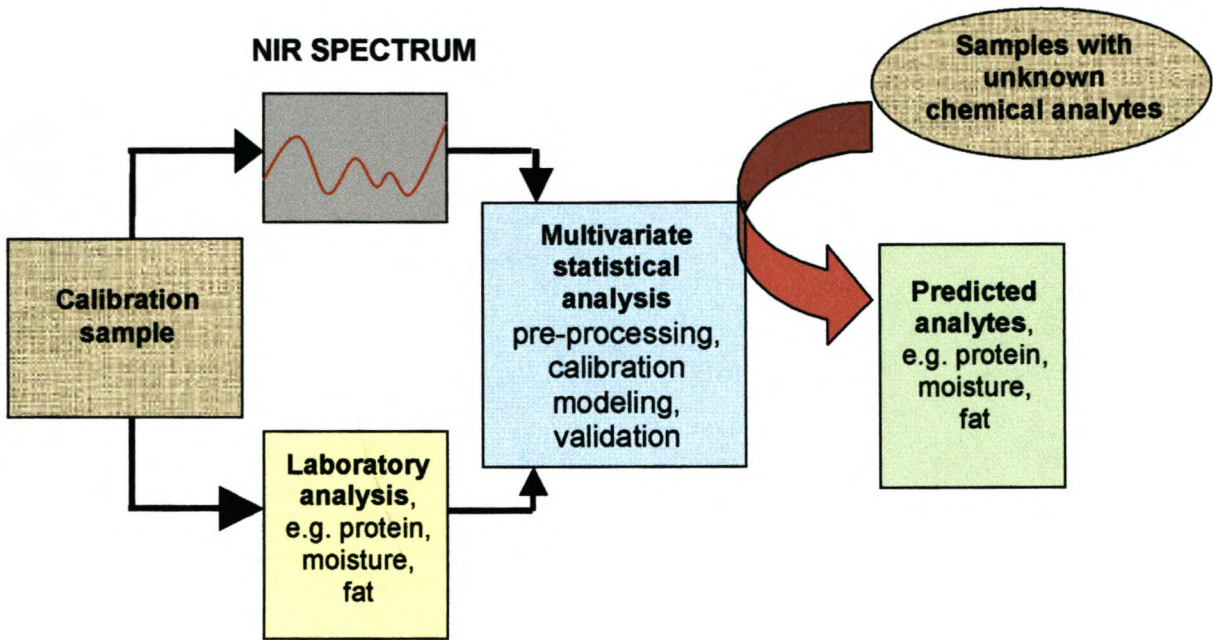
24

Figure 8. The typical flow of NIRS experiments.

are tedious and costly to perform, but can be employed to enhance the accuracy and precision of NIRS calibrations (Osborne *et al.*, 1993).

D. Chemometrics

D.1 Background

The large amount of data generated by modern analytical instruments has demanded mathematical techniques for handling multivariate data and led to the birth of a branch of science called chemometrics (Osborne *et al.*, 1993). Chemometrics is defined as information aspects of chemistry, thus how to obtain chemically relevant information from measured chemical data and how to represent and display this information (Wold, 1974).

Near infrared spectra contain information which can only be extracted by mathematical processing of the data (Osborne *et al.*, 1993). Only when data are interpreted and put to valuable use, then these data become information (Beebe & Kowalski, 1987). The utilisation of NIRS has increased since the availability, speed and capacity of computers increased to make the implementation of chemometrics possible (Osborne *et al.*, 1993).

D.2 Univariate data analysis

Traditional calibration methods are univariate, where samples with known concentration of an analyte are used in a calibration curve (amplitude vs. concentration). The calibration relationship is usually established by means of regression. By entering the amplitude measurements (the x-variable) of the unknown samples into the calibration curve, it is possible to estimate the unknown concentration (the y-variable) of the analyte (Eriksson *et al.*, 1999). Univariate analysis has shown several limitations. These include precision (such as signal noise), accuracy (such as experimental interferences) and diagnostic problems (such as non-linearities) (Eriksson *et al.*, 1999).

The statistical methods available use two shapes of data tables or matrices (Figure 9). Classical methods, such as MLR (multiple linear regression), use long and lean matrices (N>>K) (multiple observations and few variables) (Figure 9a). These have x-variables which are exact and independent and with residuals randomly distributed. MLR combines several x-variables in linear combinations, which correlate as closely as possible to the corresponding single y-vector (Esbensen, 2000). Although MLR is an excellent least squares regression tool for long and lean matrices, it cannot cope with more than 5 variables, missing data (Eriksson *et al.*, 1999) or where variables are linear combinations with each other (multicollinearity) (Beebe & Kawalski, 1987; Næs & Isaksson, 1991a; Eriksson *et al.*, 1999). This can often give poor predictions because there is a large number of irrelevant variance (information) in the

26

spectrum and this creates unstable regression equations (Beebe & Kawalski, 1987; Næs & Isaksson, 1991a). Wavelength selection methods have been used to produce stable univariate regression equations without collinearities, but this may be time-consuming (Næs & Isaksson, 1991a).

D.3 Multivariate data analysis (MVDA)

D.3.1 Data projection

Multivariate data are any data collected with multiple variables measured on multiple observations, such as in chromatographic or spectroscopic measurements. Many variables provide more information of compounds and processes than the few variables used in univariate data analysis (Eriksson *et al.*, 1999). Multivariate data analysis (MVDA) uses all the data points to derive a calibration model, which reduces the noise and is without multicollinearities (Næs & Isaksson, 1991a). It is usually the intrinsic and latent molecular properties which are responsible for the chemical behaviour of a molecule or a compound. The intrinsic molecular properties responsible for certain responses are not easily measured directly. They are measured as macroscopic molecular properties by using chemical, physical and biological model systems, such as spectroscopic (NIR, IR, NMR) and chromatographic (HPLC, GC, TLC) systems in multivariate analysis (Eriksson *et al.*, 1999).

Multivariate data must be analysed in such a manner that the desired information in the data is expressed in an understandable format, for instance a graph or a few information-rich parameters. The overview may reveal groupings of observations, trends and outliers as well as relationships among observations and variables and among the variables themselves (Eriksson *et al.*, 1999). Good multivariate calibration equations are crucial for accurate chemical determinations by NIRS. To obtain such multivariate equations, one needs adequate statistical models, efficient estimation procedures and good calibration data (Næs & Isaksson, 1994).

Multivariate data projection methods use short and fat matrices (N<<K) (multiple variables) (Figure 9b), which may have x-variables that are not independent or have errors and whose residuals may be structured. Such multivariate projection methods include principal component analysis (PCA), principal component regression (PCR) and partial least squares regression (PLS) (Eriksson *et al.*, 1999). MVDA can cope with any data matrix, is robust to noise and can tolerate moderate amounts of missing data (Eriksson *et al.*, 1999).

Multivariate data projection techniques summarise the measured variables by calculation of new variables. New variables are hidden in the measured variables and thus

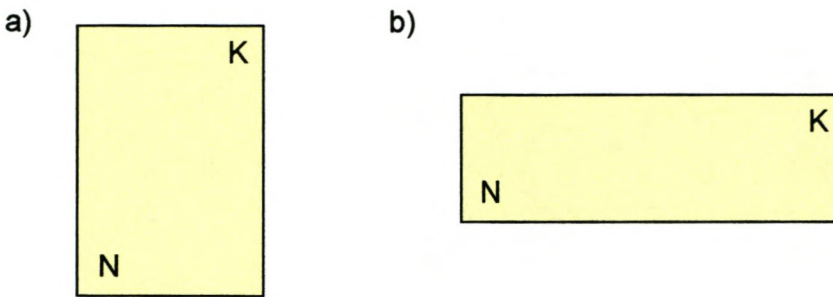**Figure 9.** Data matrices with K variables and N observations. (a) Univariate data has N>>K and (b) multivariate data has N<<K (Eriksson *et al.*, 1999).
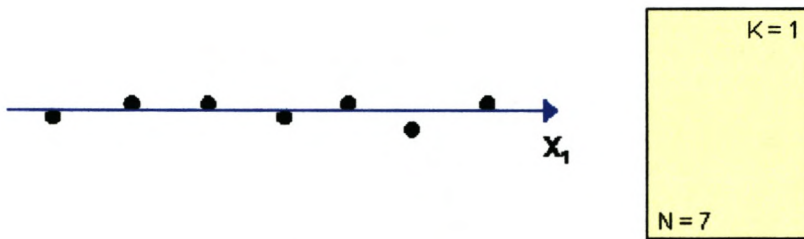

Figure 10. One dimensional line representing only one variable (Osborne *et al.*, 1993).

Figure 11. Two dimensional co-ordinate system for plotting data with two variables (Eriksson *et al.*, 1999).

called latent variables. MVDA can be explained in terms of geometry, thus spaces, lines, points, planes and hyperplanes. Multivariate data can be visualised as higher dimensional spaces with many variables. A high dimensional space is referred to as the K-space when working with x-variables (like spectra) and the M-space when working with y-variables (Eriksson *et al.*, 1999).

A data matrix with only one x-variable can be seen as a one dimensional line (Figure 10). When K = 2, there are two variables and thus two axes are needed. Each point is a sample and distances between two directions represent the spectral measurement (Figure 11).

Where three variables exist for each observation in the data table (K = 3), the three numerical values can be represented by one point in the three dimensional space (Figure 12). Should there be multiple (say twenty) variables in the x-matrix, each observation would be a point in a multi-dimensional space (twenty dimensions). To give a good overview and interpretation of the data, a plane or window is inserted. The original multidimensional space is projected down onto a two-dimensional space, the plane, and each observation is projected onto that plane (Figure 13).

The projection of observations onto a plane creates a factor. This is the main step of PCA. These new factors are also called scores or principle components (PC's). The plane is projected in such a way and angle that most of the variation in the points is captured to retain most of the information. PCA selects projections that maximise variability in the projected points (Osborne *et al.*, 1993).

A rotation of the original axes of two variables (wavelengths), K_1 and K_2, is taken which maximises the variation along a new axis, PC_1 (the first principal component). A second new axis, PC_2, which lies at right angles to the first, becomes the second PC. These new axes can be expressed as linear combinations of the original axes (equations 9-10):

$$PC_1 = C_{11}K_1 + C_{12}K_2 \qquad \qquad \text{...........9}$$
$$PC_2 = C_{21}K_1 + C_{22}K_2 \qquad \qquad \text{...........10}$$

The coefficients $C_{11}...C_{22}$ are referred to as "weights" and are always as such scaled that the sum of squared weights for any single component is one (equations 11-12)

$$C^2_{11} + C^2_{12} = 1 \qquad \qquad \text{...........11}$$
$$C^2_{21} + C^2_{22} = 1 \qquad \qquad \text{...........12}$$

Figure 12. Three dimensional space representing three variables. Each observation can be interpreted as a point in a three dimensional space (Eriksson *et al.*, 1999).

Figure 13. Each row (observation) represents a point in a twenty dimensional space. Only three dimensions are drawn here for reasons of clarity (Eriksson *et al.*, 1999).

The PC's can be used to interpret the relation between concentrations and spectra as well as the relation among the spectral variables themselves (Næs & Isaksson, 1992). The first PC will contain most of the variation, the second PC the second most, the third PC the third most, until all the variation has been explained. The PC's are thus linear combinations of the original measurements and contain information from the entire spectrum (Osborne *et al.*, 1993).

A second variate, the PC score, defines the position of each sample on each new component axis. For a sample set with 700 wavelengths, for example, it can be derived (equation 13):

$$S = C_1 E_1 + C_2 E_2 + ... + C_{700} E_{700}$$ 13

PC's are usually presented in two dimensional scatter plots (or score plots) of one component versus another (Esbensen, 2000). Score plots provide information of how observations are related to each other. These plots are used for outlier identification, identification of trends or groups and exploration of similarities. Loadings give information about the relationship between the original variables and the principal components. The loading plot provides a projection view of the inter-variable relationships and is thus used for interpreting the reasons behind the observation distribution. The corresponding score and loading plots are complementary and give valuable information about both the observations and the variables when studied together.

D.3.2. Multivariate regression

The goal of multivariate calibration is to find relationships between one or more response variables (y-variables) and a vector of predictor variables (x-variables). Calibration measurements must be made for finding the regression coefficients. An approximation called multivariate linear regression model is used (equation 14) (Geladi, 2002).

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_K x_K + f$$ 14

where b_0 = an offset

b_K (k + 1, ...K) = regression coefficients and

f = a residual.

PCR (principal component regression) and PLS (partial least squares regression) find a few relevant linear combinations of the original spectral values and use only these linear combinations in a regression equation (Næs & Isaksson, 1992). These methods can produce stable regression equations while compressing data and removing multicollinearity successfully. These data compression methods are based on the following, more compact model (Næs & Isaksson, 1992) (equation 15).

$$x = TP + E$$
$$y = Tq + f \qquad\qquad\qquad \dots\dots\dots\dots 15$$

Where x is the matrix of the mean centered NIR spectra for a set of samples and y is the corresponding mean centered vector of reference values (such as a chemical concentration) (Næs & Isaksson, 1992). T is a subset of the PC's (scores) and contain as much as possible of the relevant spectral information in x. E and f correspond to noise or irrelevant variation. P and q are loadings of x and y, respectively, and describe how x and y relate to the scores, T. The idea behind the model is to decompose the information in x and y into systematic parts TP and Tq plus noise E and f. The matrix T, which consists of linear functions of x, then serves as a carrier of information from x to y.

D.3.2.1 Principal component regression (PCR)

In PCR, T is defined as a set of principal components of x, i.e. components determined only by their variance of X (the spectra). P is the loading of x and E is the irrelevant information in x (Næs & Isaksson, 1992). PCR thus constructs factors or scores using PCA to capture as much of the variation in the x-variables as possible. It then uses MLR to estimate the relationship between the reference values and the scores, thus correcting for collinearities (Osborne *et al.*, 1993). The disadvantage of PCR is that the constructed PC's do not necessarily contain only that information which is correlated to a specific y-variable of interest. There may be other variations present in these components (Esbensen, 2000).

D.3.2.2 Partial least squares regression (PLS)

Partial least squares regression (Sjöström *et al.*, 1983; Næs & Martens, 1984) uses factors determined by using both x and y. It can also be described as a regression extension of PCA. It fits two PCA-like models, one for the x-variables and one for the y-variables, at the same time. It then connects these two blocs of variables (x and y) with each other by a linear multiple model. The end goal is to predict the y-variables from the x-variables.

32

PLS can analyse data with many noisy, collinear, incomplete variables in both x and y and the precision also improves with increasing number of relevant x-variables (Eriksson *et al.*, 1999). PLS also reduces the influence of large x-variations which do not correlate with y (Esbensen, 2000). Data analysis by PLS can also be explained with geometry in the same way as PCA. When represented graphically, each row (observations) in the data matrix corresponds to two points (in the x-space and the y-space). With many observations, two point swarms are created in these spaces. The first PC is the line in the x-space that approximates the point swarms best, while also providing a good correlation with the y-vector (Eriksson *et al.*, 1999).

If too many components are used, too much of the redundancy of the spectrum will be introduced. This leads to overfitting, which gives poor predictive ability for future samples. If we too few components are used, it is called underfitting and not enough information is used in the regression and the model then has a poor fit. The optimum number of components is determined by validation, namely cross validation (CV) and independent validation. The most used criterion is RMSEP (root mean square error of prediction). Prediction errors for different choices of number of components are computed and the one with smallest error is selected (Næs & Isaksson, 1991a).

D.3.2.3 *Multivariate adaptive regression splines (MARS)*

Multivariate adaptive regression splines, developed by Friedman (1991), is an extension of piecewise linear regression. More than one regression line is fitted to the data to account for non-linear relationships. This is achieved by the use of separate regression slopes in distinct intervals of the predictor variable space. The slope of the regression line can change from one interval to another and the point where two regression lines intersect is called a knot point. These variable intervals are defined by basis functions. As basis functions are added, the model becomes more flexible and more complex. MARS can handle more than one independent (predictor) variable as well as combinations of categorical and continuous predictors.

D.3.3 *Validation*

Validation is the measure of the prediction ability of the calibration equation for certain sample types or for a whole population of samples. Prediction is based on splitting the data set into two, one for calibration and one for validation, each with known x and y-values (Næs & Isaksson, 1991b). The two data sets should be representative of each other (Esbensen, 2000). The calibration set is used to calculate a model. The modeling error is given by comparing the y-value (analyte) from the calibration equation and the true analyte value

obtained by the reference method. The modeling error is expressed by the SEC (standard error of calibration) (equation 16). In general, the SEC will exceed the standard error of laboratory (SEL), an indication of the method repeatability (equation 17), since the laboratory plus the modeling errors are incorporated into the prediction model (Westerhaus, 1989).

$$SEC = \frac{1}{n-1-t} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad \ldots\ldots\ldots\ldots 16$$

where n = the number of samples in the calibration set

       y_i = the measured reference value for object i

       $\hat{y}_i$ = the predicted value for the same object obtained from the calibration equation

       t = the number of independent variables in the calibration equation

$$SEL = \sqrt{\frac{\sum (y_1 - y_2)^2}{2n}} \qquad \ldots\ldots\ldots\ldots 17$$

where y_1 and y_2 = the results of duplicate determinations

       n = the number of samples

This model error should be as small as possible. The calibration model is tested on the validation set or the test set to measure the predictive ability of the calibration model (Næs & Isaksson, 1991b).

The accuracy and precision of the predicted y-values can be quantified by prediction errors, which can be expected in future predictions. The overall prediction ability is best when the prediction error is at its lowest (Esbensen, 2000). Two criteria most often used for prediction performance is RMSEP (root mean square error of prediction) and SEP (standard error of prediction or performance). RMSEP is the root of the average of squared differences between predicted and measured values for the prediction objects (equation 18) (Næs & Isaksson, 1991c).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n-1}} \qquad \ldots\ldots\ldots\ldots 18$$

where n = the number of samples in the prediction set

y_i = the measured reference value for object i

$\hat{y}_i$ = the predicted value for the same object obtained from the calibration equation

SEP is defined as the standard deviation of the predicted residuals (equation 19):

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i - BIAS\right)^2}{n-1}} \qquad \ldots\ldots\ldots\ldots 19$$

Bias (average prediction error) (equation 20) is interpreted as the difference between $\hat{y}$ and y in the prediction set.

$$BIAS = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)}{n} \qquad \ldots\ldots\ldots\ldots 20$$

SEP is thus the root mean square error of prediction (RMSEP) of the bias corrected residuals. The relationship between SEP and RMSEP is simple (if the small effect of using n to determine the SEP instead of $n-1$ is ignored) (equation 21).

$$RMSEP^2 = SEP^2 + BIAS^2 \qquad \ldots\ldots\ldots\ldots 21$$

RMSEP is thus always equal to or greater than SEP. SEP could thus give an over optimistic impression of the prediction error for future samples. When the SEP is given in stead of the RMSEP, the BIAS must be reported as well (Næs & Isaksson, 1991c).

Validation is essential to ensure a quantitative model which will predict similar data sets. Independent validation and cross validation (CV) are the two methods most often used. Independent validation has been described as the best method, because it uses an independent data set (Esbensen, 2000). Cross validation is a technique based on the calibration data only, by successively deleting samples from the calibration set. First, sample no. 1 is deleted and calibration performed on the rest of the samples by comparing y with ŷ. The first sample is put back and the 2nd sample deleted and so on until all the samples have

been deleted once. The estimate of prediction ability is then measured by RMSECV (root mean square error of cross-validation) (Næs & Isaksson, 1991b). Cross validation can, however, never substitute for a proper test set validation (Esbensen, 2000).

D.3.4 Classification methods

Classification is an example of supervised pattern recognition to determine if one or more new samples belong to an already existing group of similar samples. Wold (1976) first formulated a chemometric classification method called soft independent modeling by class analogy (SIMCA). The basis for SIMCA is that objects in one class, or group, show similar rather than identical behaviour. The objects belonging to the same class show common properties or a specific class pattern, which makes all the objects in this class more similar than with respect to any other class. The goal of classification is to assign new objects to the class to which they show the largest similarity. SIMCA classification uses separate PCA models, one for each class recognised and then uses these established class models to classify to which classes new objects belong (Esbensen, 2000).

The advantages of using SIMCA as a classification method are that SIMCA is a multivariate method and is thus not restricted by the dimension of the data matrix (objects:variables ratio). Collinearities are thus also overcome. SIMCA can be used for multi-class belonging purpose, where an object can belong to more than one class. Results are also displayed graphically which makes the patterns and groupings in large data structures easy to interpret (Esbensen, 2000).

SIMCA uses statistics to quantify classification spaces and class belonging. The measures used for classification, are the object-to-model distance (Si) and the leverage or distance to model center (Hi). The object will be classified to the model to which is has the smallest distance (Si). At comparable object distances, the object is closest to the model with which it has the smallest leverage (Hi). Statistical significance levels are set to assess the risk that a new sample does not belong to a specific class. A standard F-test is used. The statistical significance levels are set by the data analyst before the classification is carried out (Esbensen, 2000).

Multivariate adaptive regression splines can also be used for classification purposes. A binary (which only have two levels) classification setting is used. The binary response variable can be coded as 0's and 1's. The MARS model is fitted and the estimated values for the response variable will lie between 0 and 1. A threshold, e.g. 0.5, can be chosen for classification purposes.

D.4 Pre-processing

The most important goal of a calibration model is to predict accurately. A regression model can be tested by plotting the predicted results (of a chosen validation set) against the measured results, as in Figure 14. The ideal situation is when the predicted values match the measured values. This means all the data points fall on a diagonal and little noise is detected (Figure 14a). A non-linear situation is shown in Figure 14b. A prediction bias is where all the points are on the one side of the diagonal (Figure 14c). Figure 14d shows a regression model with large noise. Figure 14e shows an outlier which falls out of the normal data range, but still fits the model by extrapolation. A bad outlier is one that falls in the data range, but does not fit the model (Figure 14f) (Geladi, 2002).

Pre-processing of measured data is performed to re-shape or transform the data in a way that will enhance data analysis and model interpretation (Eriksson *et al.*, 1999). The benefits of transformation are that it makes non-linear response linear, eliminates noise, stabilises the variance of the residuals and makes distributions of the residuals more normal, thus removing outliers (Eriksson *et al.*, 1999).

Pre-processing is also important because variation in x that is unrelated to y may degrade the predictive ability of a calibration model (Eriksson *et al.*, 1999). Another problem is the overlapping bands due to overtones in the NIR region that can produce collinearities. NIRS also deals with real world samples and these samples yield poor signal-to-noise (SN) ratio, baseline fluctuations and overlapping of bands due to various components. One must thus eliminate noise or transform data. Most methods focus on removing unwanted systematic variation in the data, such as baseline drifts, multiplicative scatter effects and wavelength regions with low information content (Eriksson *et al.*, 1999).

There are three basic methods of pre-processing. They are resolution enhancement, noise reduction and baseline correction (Katsumoto *et al.*, 2001). The proper use of these methods facilitates the spectral analysis, while their misuse leads to reduction in useful information or extraction of wrong information. One should also first investigate the cause of a poor regression model before applying pre-processing (Ozaki *et al.*, 2001).

D.4.1 Resolution enhancement

This is performed to separate overlapping bands and to reveal the existence of obscured bands. In NIRS, derivative methods, difference spectra, mean centering, normalisation and Fourier self-deconvulsion are used for resolution enhancement (Noda, 1993; Ozaki & Noda, 2000).
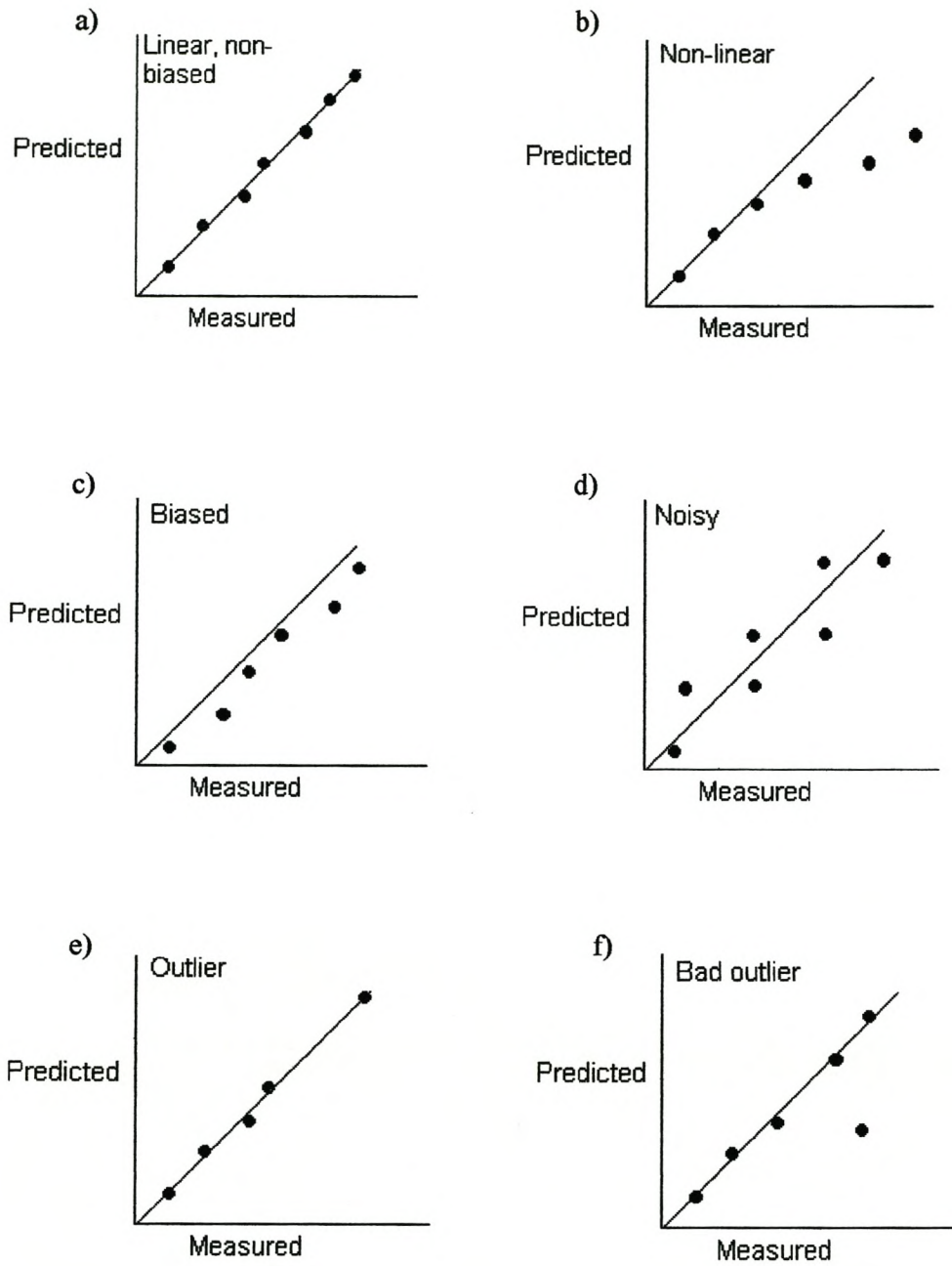
Figure 14. An ideal regression in 14a is one where the predicted values match the measured values. Pre-processing is necessary in cases of graphs of 14b-14f (Geladi, 2002).

Mean centering is powerful in resolution enhancement for FT-NIRS. A mean centered spectrum is the difference between the individual spectrum and an averaged spectrum. Then all means are equal to zero and variances are spread around zero. This method is often used in PCA as a pre-processing method (Beebe & Kowalski, 1987; Katsumoto *et al.*, 2001).

Normalisation transforms spectral data points on a unit hypersphere (it involves dividing each individual spectrum by the variance of the spectra) and all data are approximately in the same scaling (Beebe & Kowalski, 1987; Katsumoto *et al.*, 2001). One has to take great care of the use of normalisation in situations where exploratory data analysis and PCA-related procedures, such as PCA and PLS, are concerned (Katsumoto *et al.*, 2001).

D.4.2 Noise reduction

Noise is caused by the interfering of chemical and/or physical processes. High frequency noise can usually be attributed to instrument detectors and electronic circuits (Katsumoto *et al.*, 2001). Low frequency noise can be caused by instrument drift, a poor interferometer in an FT-NIR instrument or poor detector electronics (Savitzky & Golay, 1964; Eriksson *et al.*, 1999). This noise often resembles the real information in the data and is thus difficult to reduce. NIRS has high SN ratio, but there are some important bands of low molar absorptions that are more vulnerable to latent noise. High frequency noise is reduced by averaging the multiple spectra as part of collecting data. If this is still not enough, smoothing, such as the moving average method and Savitzky-Golay method, can be used to reduce high frequency noise. The moving average method is the simplest type of smoothing (Savitzky & Golay, 1964).

Wavelets is a method that removes high and low frequency noise as well as localised noise due to scattering (Katsumoto *et al.*, 2001). Wavelets transforms the spectra into the wavelet domain and returns it to the spectral domain. This is similar to applying the Fourier transform to data to reduce noise, but wavelets is more sophisticated to modeling the data than the sine/cosine of Fourier transform (Chao *et al.*, 1996; Collantes *et al.*, 1997; Stork *et al.*, 1998). The advantages of wavelets are gain in conceptual speed, lower demands for computer memory and the fact that wavelets are more generally applicable.

D.4.3 Baseline correction

Baseline fluctuations are caused by intensity changes in light scattering, changes in density and normal reflection light. These are induced by variations in average particle size,

distribution of particle shape and the density of packing of sample in a cell (Katsumoto *et al.*, 2001).

Derivative methods have long been used for baseline correction and resolution enhancement (Griffiths, 1987; Griffiths *et al.*, 1987). It was described in NIRS literature as the calculation of differences between different parts of the spectrum (Norris and Williams, 1984). Derivatives emphasise changes in the slope. The first derivative is the calculated slope of the raw spectra. The peak maximum in a spectrum then becomes zero. The second derivative is the calculated slope of the first derivative. In second derivative spectra, the peak maximum turns out clearly as separated downward peaks (Davies, 1993; Osborne *et al.*, 1993). The calculation of the second derivative enables one to investigate the shift of each band in detail. The additive and multiplicative baseline variations in the original spectrum are also removed. The disadvantage of the second derivative method is that the SN ratio deteriorates every time a spectrum is differentiated (Katsumoto *et al.*, 2001).

Multiplicative scatter correction (MSC) removes additive and multiplicative differences in a spectrum due to, for instance, differences in sample particle size. It can distinguish between light absorption and light scattering. MSC corrects spectra according to a simple linear univariate fit to a standard spectrum and is estimated by least square regression using the standard spectrum. As the standard spectrum, a spectrum of a particular sample or an average spectrum is used (Geladi *et al.*, 1985).

Detrending transformation attempts to correct baseline variations based on quadratic nonlinearity. It is usually applied to NIR diffuse reflectance spectra of powders or densely packed samples (Barnes *et al.*, 1989).

Standard normal variate (SNV) transformation processes each spectrum by subtracting the mean of the spectra from the spectral values obtained at each wavelength (Barnes *et al.*, 1989). These centered spectra are scaled by the standard deviation of the spectra. After SNV transformation the spectra have zero mean and unit variance.

Orthogonal signal correction (OSC) removes variation in the spectral data which is uncorrelated to y, such as particle size or path length variations (Sjöblom *et al.*, 1998; Wold *et al.*, 1998). OSC removes these variations before the PLS step, so that they cannot dominate in the procedure. This is achieved by seeking a small number of factors that explain variability in X, but which is not correlated with y.

Second derivatives, MSC, OSC and SNV have a similar effect. The differences between these pretreatments are small and the one best suited to use will be dictated by each specific experiment (Fearn, 1999).

E. NIRS and chemometric applications for the food industry

The acceptance of NIRS as a valuable analytical tool can be attributed to two scientists. Karl Norris engineered instruments in the 1960's which could record NIR spectra and also applied multivariate treatment of the spectra to determine moisture, oil, protein and starch in biological samples (Davies, 1996). In the early 1970's Phil Williams recognised the potential of the technology to segregate wheat grain according to protein content (Williams, 1975).

The limits to the applications of NIRS are not yet evident and the technique is only limited to the range of ideas suggested by its researchers (Batten, 1998). While agricultural applications of NIRS led to its worldwide acceptance as an analytical tool, the uses and applications of NIR are now seen in many fields (food, chemical and pharmaceutical industries as well as in medicine).

Near infrared spectroscopy is applicable to the determination of the major rather than trace constituents in foods, although there are possible exceptions (Osborne et al., 1993). NIR reflectance is used successfully to determine moisture, protein, fat and carbohydrate contents of a wide variety of products in the meat, baking and dairy industries (Osborne et al., 1993). Other quantitative measurements could also be achieved with NIRS, such as organic acids, chlorophyll, defects and diseases in fruits and vegetables, as well as alcohol in the alcoholic beverage industry (Osborne et al., 1993).

Near infrared spectroscopy has been used successfully in the fruit industry to determine the internal quality of fruit. These methods were all quantitative and used multivariate regression techniques, such as PLS, PCR and MLR. Internal quality of peaches were determined by non-destructively predicting the soluble solids content (SSC) (r = 0.90, SEP = 1.0%), sucrose content (r = 0.84, SEP = 0.58%), sorbitol content (r = 0.89, SEP = 0.23%), sweetness index (r = 0.90, SEP = 0.68%) and chlorophyll content (r = 0.94, SEP = 0.23 µg/g) (Slaughter, 1995).

Previous studies have found that a strong predictable relationship exists between NIR absorbance and SSC of whole peaches (Kawano et al., 1992, 1995; Slaughter, 1995; Peiris et al., 1998; Van Zyl, 2000) and apricots (Carlini et al., 2000) (Table 1).

The application of NIRS for SSC determinations was also investigated on various other intact fruit, such as apples (Moons et al., 1997; Lammertyn et al., 1998; Ventura et al., 1998; Lu et al., 2000; McGlone et al., 2002a), kiwifruit (McGlone & Kawano, 1998; Schaare & Fraser, 2000; McGlone et al., 2002b) melons, pineapples (Guthrie et al., 1998), mangoes (Schmilovitch, 2000; Saranwang et al., 2003) and prunes (Slaughter et al., 2003) (Table 1). Good correlations between NIR estimated and actual °Brix values were obtained. Less

41

Table 1. Validation results obtained in previous studies for measuring SSC and firmness in fruit samples using NIR spectroscopy.

Sample set	NIR spectrophotometer used	Wavelength range (nm)	Constituent measured	Validation statistics			Reference
				r	SEP	RMSEP	
Peach (one cultivar)	Pacific Scientific Mode 6250 (interactance)	800-1650	SSC	0.97	0.50 °Brix		Kawano et al., 1992
Peach (one cultivar)	Pacific Scientific Mode 6250 (interactance)	800-1650	SSC	0.96	0.42 °Brix		Kawano et al., 1995
Peach (4 cultivars)	NIRSystems Model 6500, fiber optic probe (interactance)	400-1100	SSC	0.90	1.00 %		Slaughter, 1995
Peach (Individual cultivars)	Acusto-optic spectrometer (transmittance)	800-1050	SSC	0.48-0.84	0.62-1.52 %		Peiris et al., 1997
Peach (Individual cultivars)	Acusto-optic spectrometer (transmittance)	800-1050	SSC	0.20-0.94	0.49-1.63 %		Peiris et al., 1998
Peach (9 cultivars)	Spectrum Identicheck™ 2.0 FT-NIR System (Perkin Elmer)	1000-2500	SSC	0.96	0.55	0.57	Van Zyl, 2000
Apricot (one cultivar)	Vis/NIR Fiber optics (NIRSystem 6500)	400-2500	SSC	0.94	0.75		Carlini et al., 2000
Apple (5 vultivars)	Vis/NIR monochromator (NIRSystem 6500)	400-2500	SSC	0.97	0.38 °Brix		Moons et al., 1997
		400-750		0.77	0.99 °Brix		
		750-2500		0.97	0.33 °Brix		

Table 1. Continued.

Sample set	NIR spectrophotometer used	Wavelength range (nm)	Constituent measured	Validation statistics			Reference
				r	SEP	RMSEP	
Apple (one cultivar)	Vis/NIR reflectance fiber optics	380-1650	SSC	0.82	0.61 °Brix		Lammertyn *et al.*, 1998
			Firmness	0.90	2.49 N		
Apple (individual cultivars)	Diffuse reflectance	800-1700	SSC	0.93-0.97	0.37-0.58 °Brix		Lu *et al.*, 2000
			Firmness	0.38-0.58	6.6-10.1 N		
Apple (one cultivar)	Vis/NIR diode array (Zeiss MMSI-NIR)	500-1100	SSC	0.84		0.50 %	McGlone *et al.*, 2002a
			Firmness	0.74		7.50 N	
Kiwifruit (one cultivar)	Interactance fiber optics (NIRSystems 6500)	800-1100	SSC	0.95		0.39 °Brix	McGlone & Kawano, 1998
			Firmness	0.81		7.80 N	
Kiwifruit (one cultivar)	Interactance	300-1100	SSC	0.96	0.80 ° Brix		Schaare & Fraser, 2000
	Reflectance	300-1100		0.93	1.18 ° Brix		
	Transmittance	300-1100		0.94	1.01 ° Brix		

Table 1. Continued.

Sample set	NIR spectrophotometer used	Wavelength range (nm)	Constituent measured	Validation statistics			Reference
				r	SEP	RMSEP	
Kiwifruit (one cultivar)	Interactance	300-1140	SSC	0.96		0.39 ° Brix	McGlone *et al.*, 2002b
Melon (3 cultivars)	Reflectance (NIRSystems 6500)	700-1100	SSC	0.55	1.43 ° Brix		Guthrie *et al.*, 1998
Pineapple (one cultivar)	Reflectance (NIRSystems 6500)	700-1100	SSC	0.88	0.74 ° Brix		Guthrie *et al.*, 1998
Mango (one cultivar)	Reflectance fiber optics (Quantum 1200)	1200-2400	SSC	0.93	1.22 %		Schmilovitch, 2000
			Firmness	0.82	17.14 N		
Prune (one cultivar)	Diode array (Model 51000-TR Ocean Optics)	700-1100	SSC	0.98	1.20 %		Slaughter *et al.*, 2003

accurate NIR predictions of fruit firmness was reported (Lammertyn *et al.*, 1998; McGlone & Kawano, 1998; Lu *et al.*, 2000; Schmilovitch, 2000; McGlone *et al.*, 2002a). Much work remains to be done on understanding the physical or chemical basis for measuring fruit firmness using NIRS (Lu *et al.*, 2000).

Successful classifications have been made in the past on agricultural products (Sato, 1994; Wesley, 1996; Kusama *et al.*, 1997; Blennow *et al*, 2000; Woo *et al.*, 2002). NIR reflectance spectroscopy with SIMCA provided a rapid, accurate and non-destructive classification method for the determination of cultivation areas of ginseng samples despite broad and overlapping bands (Woo *et al.*, 2002). PCA models, based on complex chromatographic data were used to accurately predict crystal polymorphs and botanical origin of starch (Blennow *et al.*, 2000). Pure strawberry pulp samples could be differentiated from fruit pulp mixtures using NIRS and SIMCA (Wesley, 1996). PCA scores using NIR spectra of various vegetable oils also successfully classified nine varieties of oils (Sato, 1994). Kusama *et al.* (1997) differentiated the NIR spectra of aged and normal soybeans by PCA and discriminate analysis.

F. Conclusion

Many applications have successfully been developed in the field of NIRS. Since the commercial benefits of NIRS were realised, research and knowledge in this field have been growing. The continuous development in the field of chemometrics, i.e. new and more refined pre-processing and multivariate methods for specific applications, provides even more possibilities for the extraction of desired information from complex spectral data structures. There continues to exist many opportunities for further research in the field of NIRS and chemometrics.

A large part of the applications of NIRS were developed for analysis of raw materials and quality attributes in the food industry. Successful prediction models were developed to determine important quality attributes (such as SSC) in various fruits. Fruit is organic material, which is constantly changing as ripening, maturation and senescence occur. The internal properties of fruit, which determine its post-storage quality, are difficult to grasp and several factors may contribute to an undesirable fruit quality. NIRS and chemometrics could possibly reveal valuable information about internal quality of intact fruit that was not previously measurable non-destructively.

G. References

Amoros, A., Serrano, M., Riquelme, F. & Romojaro, F. (1989). Importancia del etileno en el desarollo y maduracion del albaricoque. *Fruits*, **44**, 171-175 (As cited by Pretel *et al.*, 1993).

Anderson, A.E. (1979). The influence of storage temperature and warming during storage on peach and nectarine fruit quality. *Journal of the American Society for Horticultural Science*, **104**, 459-461.

Anderson, R.E. (1982). Long term storage of peaches and nectarines intermittently warmed during controlled atmosphere storage. *Journal of the American Society for Horticultural Science*, **107**, 214-216.

Andreotti, R., Thomasiccio, M. & Macchiavelli, L. (1979). Packing pale fruit in syrup and peeled tomatoes in lacquered cans and in glass jars. *Industria Conserve*, **54**(2), 119-124 (As cited by Manolopoulou & Mallidis, 1999).

Artés, F., Cano, A. & Fernandez-Trujillo, J.P. (1996). Pectolytic enzyme activity during intermittent warming storage of peaches. *Journal of Food Science*, **61**, 311-313.

Atkins, P.W. (1996). Molecular spectroscopy. In: *The elements of physical chemistry*, 2nd ed. Pp. 411-427. Oxford: Oxford University Press.

Bartley, I.M. & Knee, M. (1982). The chemistry of textural changes in fruit during storage. *Food Chemistry*, **9**, 47-58.

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

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

Batten, G.D. (1998). Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Australian Journal of Experimental Agriculture*, **38**, 697-706.

Beebe, K.R. & Kowalski, B.R. (1987). An introduction to multivariate calibration and analysis. *Analytical Chemistry*, **59**, 1007-1015.

Ben-Arie, R. & Lavee, S. (1971). Pectic changes occurring in Elberta peaches suffering from woolly breakdown. *Phytochemistry*, **10**, 531-538.

Ben-Arie, R. & Sonego, L. (1980). Pectolytic enzymes activity involved in woolly breakdown of stored peaches. *Phytochemistry*, **19**, 2553-2555.

Ben-Arie, R., Lavee, S. & Guelfat-Reich, S. (1970). Control of woolly breakdown of 'Elberta' peaches in cold storage by intermittent exposure to room temperature. *Journal of the American Society for Horticultural Science,* **95**, 801-803.

Blaha, J. (1983). Results of studies on apricots from South Moravia. *Mitteilungen Klosterneuburg, Rebe und Wein Obstbau und Fruchteverwertung*, **33**, 31-32 (As cited by Manolopoulou & Mallidis, 1999).

Blennow, A., Engelsen S.B., Munck L. & Moller, B.L. (2000). Starch molecular structure and phosphorylation investigated by a combined chromatographic and chemometric approach. *Carbohydrate Polymers*, **41**, 163-174.

Bokobza, L. (1998). Origin of near-infrared absorption bands. *Journal of Near Infrared Spectroscopy*, **6**, 3-18.

Buescher, R.W. & Furmanski, R.J. (1978). Role of pectinesterase and polygalacturonase in the formation of woolliness in peaches. *Journal of Food Science*, **43**, 264-266.

Buescher, R.W. & Tigchelar, E.C. (1975). Pectinesterase, polygalacturonase, Cx-cellulaase activities and softening of the *rin* tomato mutant. *HortScience*, **10**, 624-625.

Butler, L.A. (1983). The history and background of NIR. *Cereal Foods*, **28**, 238-240.

Carlini, P., Massantini, R. & Mencarelli, F. (2000). Vis-NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. *Journal of Agricultural and Food Chemistry,* **48**, 5236-5242.

Chau, F.T., Shih, T.M., Gao, J.B. & Chan, C.K. (1996). Application of the fast wavelet transform method to compress ultraviolet-visible spectra. *Applied Spectroscopy*, **50**, 339-348.

Claypool, L.L. (1974). Apricot softening: a problem of the canned fruit. *California Agriculture*, **28**(7), 4-7.

Collantes, E.R., Duta, R., Welsh, W.J., Zielinski, W.L. & Brower, J. (1997). Preprocessing of HPLC impurity patterns by wavelet packets for pharmaceurical fingerprinting using artificial neural networks. *Analytical Chemistry*, **69**, 1392.

Dalles, T. & Kehagias, C. (1980). Effect of various factors on the texture of canned apricots of the variety "Bebecou". *Agriculture Research (Greece)*, **4**, 3345-355 (As cited by Manolopoulou & Mallidis, 1999).

Davies, A.M.C. (1993). Derivative? What do you mean derivative? *NIR news*, **4**(4), 10-11.

Davies, A.M.C. (1996). Accolades, accusations and anecdotes. *Journal of Near Infrared Spectroscopy*, **6**, 3-21 (As cited by Batten, 1998).

Davies, A.M.C. (2000). William Herschel and the discovery of near infrared energy. *NIR news*, **11**(2), 3-5.

Dawson, D.M., Melton, L.D. & Watkins, C.B. (1992). Cell wall changes in nectarines (*Prunus persica*): Solubilisation and polymerisation of the pectic and neutral polymers during ripening and in mealy fruit. *Plant Physiology*, **100**, 1203-1210.

Dawson, D.M., Watkins, C.B. & Melten, L.D. (1993). Calcium uptake and efflux, ion leakage, internalair space and cation exchange capacity in relation to mealiness in nectarine tissue. *Postharvest Biology and Technology*, **3**, 131-141.

Dawson, D.M., Watkins, C.B. & Melten, L.D. (1995). Intermittent warming affects cell wall composition of 'Fantasia' nectarines during ripening and storage. *Journal of the American Society for Horticultural Science*, **120**, 1057-1062.

Day, M.S. & Fearn, F.R.B. (1982). Near infra-red reflectance as an analytical technique. Part 1. History and development. *Laboratory Practice*, **31**, 328-330.

De Haan, I. (1957). Pectin conversion in peaches during cold storage. *The South African Industrial Chemist*, **11**, 26-34.

Eksteen, J. & Combrink, C. (1994). *Manual for the identification of postharvest disorders of pome and stone fruit*. p. 7. South Africa: Stellenbosch Institute for Fruit Technology, Agricultural Research Council.

Eriksson, L., Johansson, E., Kettaneh-Wold, N. & Wold, S. (1999). *Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS)*. Norway: Umetrics AB.

Esbensen, K.H. (2000*). Multivariate Data Analysis in practice: An introduction to multivariate data analysis and experimental design*, 4th ed. Pp. 1-352. Norway: Camo ASA.

Fearn, T. (1999). A look at some standard pre-treatments for spectra. *NIR news*, **10**(3)**,** 10-11.

Fernandez-Trujillo, J.P., Cano, A. & Artés, F. (1998). Physiological changes in peaches related to chilling injury and ripening. *Postharvest Biology and Technology*, **13**, 109-119.

Fidler, J.C., Wilkinson, B.G., Edney, K.L. & Sharples, R.O. (1973). The biology of apple and pear storage. In: *Slough: Commonwealth Agricultural Bureaux.*

Friedman, J.H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, **19**, 1-141.

Geladi, P. (2002). Some recent trends in calibration literature. *Chemometrics and Intelligent Laboratory Systems,* **60**, 211-224.

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

Giangiacomo, R. & Nzabonimpa, R. (1994). Approach to near infrared spectroscopy. *Bulletin of the International Dairy Federation*, **298**, 37-42.

Griffiths, P.R. (1987). *Laboratory Methods in Vibrational Spectroscopy* (Edited by J.H. van der Maas & R.G.J. Miller), 3rd ed. p. 121. Chichester: John Wiley & Sons (As cited by Katsumoto *et al.*, 2001).

Griffiths, P.R. & De Haseth, J.A. (1986). *Fourier Transform Infrared Spectroscopy.* New York: John Wiley & Sons (As cited by Van de Voort, 1992).

Griffiths, P.R., Pierce, J.A. & Hongjin, G. (1987). *Computer-enhanced Analytical Spectroscopy* (Edited by H.L.C. Meuzelaar & T.L. Isenhour), 3rd ed. p. 29. New York: Plenum Press (As cited by Katsumoto *et al.*, 2001).

Guthrie, J., Wedding, B. & Walsh, K. (1998). Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy*, **6**, 259-265.

Herschel, W. (1800). *Philosophical Transactions of the Royal Society* (London), **284** (As cited by Davies, 2000).

Huber, D.J. (1983). The role of cell wall hydrolases in fruit softening. *Horticultural Reviews*, **5**, 169-219.

Hulme, A.C. (1971). The biochemistry of fruit and their products, Volume II. Pp. 411-431. London and New York: Academic Press.

Kader, A.A., Heintz, C.M. & Chordas, A. (1982). Postharvest quality of fresh and canned clingstone peaches as influenced by genotype and maturity at harvest. *Journal of the American Society for Horticultural Science*, **107**, 947-951.

Kader, A.A. (1985). Ethylene-induced senescence and physiological disorders in harvested horticultural crops. *HortScience*, **20**, 54-57.

Katsumoto, Y., Jiang, J., Berry, R.J. & Ozaki, Y. (2001). Modern pretreatment methods in NIR spectroscopy. *Near Infrared Analysis*, **2**, 29-36.

Katuyama, T., Nakayama, T.O.M., Lee, T.H. & Chichester, C.O. (1971). Carotenoid transformations in ripening apricots and peaches. *Journal of Food Science*, **47**, 429-431.

Kauzmann, W. (1957). *Quantum Chemistry.* Pp. 667. New York: Academic Press (As cited by Barton, 2002).

Kawano, S., Abe, H. & Iwamoto, M. (1995). Development of a calibration equation with temperature compensation for determining the brix value in intact peaches. *Journal of Near Infrared Spectroscopy*, **3**, 211-218.

Kawano, S., Watanabe, H. & Iwamoto, M. (1992). Determination of sugar contents in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. *Journal of the Japanese Society for Horticultural Science*, **61**, 445-451.

Kusama, T., Abe, H., Kawano, S. & Iwamoto, M. (1997). Classification of normal and aged soybean seeds by discriminant analysis using principal component scores of near infrared spectra. *Journal of Japanese Society of Food Science and Technology*, **44**, 569-578.

Lammertyn, J., Nicolai, B., Ooms, K., Smedt, V. & Baerdemaeker, J. (1998). Non-destructive measurement of acidity, soluble solids and firmness of Jonagold apples using NIR spectroscopy. *Journal of Fruits, Vegetables and Nuts*, **41**, 1089-1094.

Laties, G.G. (1995). Review: Franklin Kidd, Charles West and F.F. Blackman: The start of modern postharvest physiology. *Postharvest Biology and Technology*, **5**, 1-10.

Lill, R.E., O'Donoghue, E.M. & King, G.A. (1989). Postharvest physiology of peaches and nectarines. *Horticulture Review*, **11**, 413-452.

Lotz, E. (2002). Agricultural Research Council, Infruitec-Nietvoorbij, Stellenbosch. Personnal communication.

Lotz, E. & Combrink, J.C. (2001). *Voorkoming van Naoesverliese, die Sleutel tot Gesonde Vrugte wat Lank Hou*. Pp. 1-20. Stellenbosch: Infruitec-Nietvoorbij, Agricultural Research Council

Lu, R., Guyer, D.E. & Beaudry, M. (2000). Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *Journal of Texture Studies*, **31**, 615-628.

Luh, B.S., Usu, E.T. & Stachowicz, K. (1967). Polyphenolic compounds in canned cling peaches. *Journal of Food Science*, **32**, 251-258.

Luza, J.G., Van Gorsel, R., Polito, V.S. & Kader, A.A. (1992). Chilling injury in peaches: A cytochemical and ultrastructural cell wall study. *Journal of the American Society of Horticultural Science*, **117**, 114-118.

Mallidis, C. & Katsaboxis, C. (1999). Suitability of apricot cultivars for canning. In: *Proceedings of the XIth International Symposium on the Apricot Culture*, Volume II. Pp. 593-597. May, 1997. Verra-Makedonia, Greece.

Mallidis, C., Stylianidis, D., Katsaboxis, K., Karrayiannis, I. & Papanicolaou, D. (1993). Evaluation of apricot cultivars for canning In: Agriculture. *Deuxieme recontre sur l' abricotier, Avignon (Fance)*. EUR 15009 FR, 219-226 (As cited by Manolopoulou & Mallidis, 1999).

Mallidis, C., Stylianidis, D., Katsaboxis, K. & Papanicolaou, D. (1990). Evaluation of physiochemical and technological characteristics of clingstone peaches. In: *3rd National Greek Conference on Food Science and Technology.* Athens, Greece (As cited by Manolopoulou & Mallidis, 1999).

Maness, N.O., Brusewitz, G.H. & McCollum, T.G. (1992). Internal variation in peach fruit firmness. *HortScience*, **27**, 903-905.

Manolopoulou, H. & Mallidis, C. (1999). Storage and processing of apricots. In: *Proceedings of the XIth International Symposium on the Apricot Culture*, Volume II. Pp. 567-573. May, 1997. Verra-Makedonia, Greece.

McGlone, V.A. & Kawano, S. (1998). Firmness, dry matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology*, **13**, 131-141.

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

McGlone, V.A., Jordan, R.B., Seelye, R. & Martinsen, P.J. (2002b). Comparing density and NIR methods for measurement of kiwifruit dry matter and soluble solids content. *Postharvest Biology and Technology*, **26**, 191-198.

Mitchell, F.G., Mayer, G., Maxie, E.C. & Coates, W.W. (1974). Cold storage effects on fresh market peaches, nectarines and plums: In: *I. Estimating freezing points, II. Using low temperatures to delay internal breakdown. California Agriculture*, **28**, 12-14. (As cited by Luza *et al.*, 1992).

Morris, L.L. (1982). Chilling injury of horticultural crops: An overview. *HortScience*, **17**, 161-162.

Moons, E., Dardenne, P., Dubois, A. & Sindic, M. (1997). Nondestructive visible and NIR spectroscopy for determination of internal quality in apple. In: *Proceedings from the Sensors for Nondestructive Testing International Conference and Tour*. Pp. 122-132. February, 1997. Orlando, Florida.

Næs, T. & Isaksson, T. (1991a). Multicollinearity and the need for data reduction. *NIR news*, **2**(6), 10-11.

Næs, T. & Isaksson, T. (1991b). Fitting, prediction testing, cross validation or leaverage correction? *NIR news*, **2**(5), 10-11.

Næs, T. & Issaksson, T. (1991c). SEP or RMSEP, which is best? *NIR news*, **2**(4), 16.

Næs, T. & Issaksson, T. (1992). Interpreting principle components in NIR spectroscopy. *NIR news*, **3**(3), 7-9.

Næs, T. & Issaksson, T. (1994). Selection of samples for calibration. *NIR news*, **5**(6), 16.

Næs, T. & Martens, H. (1984). Multivariate calibration. II. Chemometric methods. *Trends in Analytical Chemistry*, **3**(10), 266-271.

Noda, I. (1993). Generalised two-dimensional correlation method applicable to infrared raman and other types of spectroscopy. *Applied Spectroscopy*, **47**, 1329-1336.

Norris, K.H. & Williams, P.C. (1984). Optimisation of mathematical treatments of raw near infrared signal in the measurement of protein in hard red spring wheat. 1. Influence of particle size. *Cereal Chemistry*, **61**, 158.

Ortiz, C., Barreiro, P., Correa, E., Riquelme, F & Ruiz-Altisent, M. (2001) Non-destructive identification of woolly peaches using impact response and near-infrared spectroscopy. *Journal of Agricultural Engineering Research*, **78**, 281-289.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis* (Edited by D. Browning), 2nd ed. Pp. 1-220. Harlow: Longman Scientific and Technical.

Ozaki, Y. & Noda, I. (2000). Traditional correlation spectroscopy. *AIP Conference series*, Vol. 503. New York: American Institute of Physics.

Ozaki, Y. Šašic, S. & Jiang, J. H. (2001). Review: How can we unravel complicated near infrared spectra? Recent progress in spectral analysis methods for resolution enhancement methods and band assignments in the near infrared region. *Journal of Near Infrared Spectroscopy*, **9**, 63-95.

Paroz, P., Seale, P.E. & Harger, K.A. (1973). Textural breakdown in canned apricots. *Food Technology Australia.*, **25**, 130-135 (As cited by Manolopoulou & Mallidis, 1999).

Peiris, K.H.S., Dull, G.G., Leffler, R.G. & Kays, S.J. (1998). Near infrared spectrometric method for nondestructive determination of soluble solids contents of peaches. *Journal of American Society of Horticulture Science*, **123**, 898-905.

Pretel, M.T., Serrano, M., Martinez, G., Riquelme, F. & Romojaro, F. (1993). Influence of films of different permeability on ethylene synthesis and ripening of MA-packaged apricots. *Lebensmittel-Wissenschaft & Technologie*, **26**, 8-13.

Reid, S.M. (1985). *Product maturation and maturity indices in post harvest technology of horticultural crops.* Pp. 8-11. University of California, Division of Agriculture and Natural Resources. (As cited by Manolopoulou & Mallidis, 1999).

Ritenour, M.A., Mangrich, M.E., Beaulieu, J.C., Rab, A. & Salveit, M.E. (1997). Ethanol effects on the ripening of climacteric fruit. *Postharvest Biology and Technology*, **12**, 35-42.

Robertson, J.A., Meredith, F.I., Lyon, B.G., Chapman, G.W. & Sherman, W.B. (1992). Ripening and cold storage changes in the quality characteristics of nonmelting clingstone peaches (FLA 1-20C). *Journal of Food Science*, **57**, 462-465.

Ryall, M.S. & Pentzer, M.S. (1982). *Handling, transportation and storage of fruits and vegetables*, Volume II. Pp. 495-498. Connecticut: AVI Publishing Company, Inc.

Saranwang, S, Somsrivichai, J. & Kawano, S. (2003). Performance of a portable near infrared instrument for Brix determination of intact mango fruit. *Journal of Near Infrared Spectroscopy*, **11**, 175-181.

Sato, T. (1994). Application of principal component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *Journal of the American Oil Chemists' Society*, **71**, 293-298.

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

Schaare, P.N. & Fraser, D.G. (2000). Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidia chinensis*). *Postharvest Biology and Technology*, **20**, 175-184.

Schmilovitch, Z., Mizrach, A., Hoffman, A., Egozi, H. & Fuchs, Y. (2000). Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology*, **19**, 245-252.

Shadow, W. NIR theory – The basics, http://www.charpa.co.th/bulletin/nir_theory.html, 2002/10/01.

Sjöblom, J., Svensson, O., Josefson, M., Kullberg, H. & Wold, S. (1998). An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, **44**, 229-244.

Sjöström, M., Wold, S., Lindberg, W., Persson, J. & Martens, H. (1983). A multivariate calibration problem in analytical chemistry solved by partial least squares models in latent variables. *Analytica Chimica Acta*, **150**, 61-70.

Slaughter, D.C. (1995). Nondestructive determination of internal quality in peaches and nectarines. *American Society of Agricultural Engineers*, **38**, 617-623.

Slaughter, D.C., Thompson, J.F. & Tan, E.S. (2003). Nondestructive determination of total and soluble solids in fresh prune using near infrared spectroscopy. *Postharvest Biology and Technology*, **28**, 437-444.

Sommer, N.F., Buchanan, J.R. & Fortlage, R.J. (1984). Inactivation of pectolytic fungal enzymes in raw apricot to prevent softening of canned fruits. *Lebensmittel-Wissenschaft & Technologie*, **17**(3)167-170.

Souty, M. & Cambroy, Y. (1994). C.R. Contrat CEE, CT 900016 (As cited by Manolopoulou & Mallidis, 1999).

Stork, C.L., Veltkamp, D.J. & Kowalski, B.R. (1998). Detecting and identifying spectral anomalies using wavelet processing. *Applied Spectroscopy*, **52**, 1348-1352.

Van de Voort, F.R. (1992). Fourier transform infrared spectroscopy applied to food analysis. *Food Research International*, **25**, 397-403.

Van Zyl, A. (2000). The application of Fourier transform near infrared (FT-NIR) spectroscopy in the wine, fruit and dried fruit industries of South Africa. MSc in Food Science Thesis, University of Stellenbosch, South Africa.

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

Victor, W. (2002). Canned Fruit Producers' Association, Personal communication.

Visagie, T.R. & Eksteen, G.J. (1981). Picking maturity of stone fruits. *The Deciduous Fruit Grower*, **8**, 312-317.

Walsh, C.S., Daberkow, P.N., Hoffman, N., Follin, K., Lane, H. & McDowell, E.F. (2002). The effects of chilling temperatures on juiciness and ethylene evolution in peach fruit. In: *Proceedings of the 5th International Peach Symposium*, Volume 1. Pp. 629-633. July, 2001. Davis, CA.

Wesley, I. (1996). The application of near infrared spectroscopy to the identification of fruit pulps. *Fruit Processing*, **6**, 263-266

Westerhaus, W.O. (1989). Calibration: Interpretation of regression statistics. In: *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality* (Edited by G.C. Marten, J.S. Shenk & F.E. Barton II). Pp. 39-40. United States Department of Agriculture.

Williams, P.C. (1975). Application of near infrared reflectance spectroscopy to the analysis of cereal grains and oilseeds. *Cereal Chemistry*, **52**, 561-576.

Wold, S. (1974). Kemometri-kemi och tillämpad matematik. In: *Yearbook of Swedish Natural Science Research Council, Stockholm*. Pp. 200-206 (As cited by Wold & Sjöström, 1998).

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognition*, **8**, 127-139.

Wold, S. & Sjöström, M. (1998). Chemometrics, present and future success. *Chemometrics and Intelligent Laboratory Systems*, **44**, 3-14.

Wold, S., Antii, H., Lindgren, F. & Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, **44**, 175-185.

Woo, Y., Cho, C., Kim, H., Yang, J. & Seong, K. (2002). Classification of cultivation area of ginseng by near infrared spectroscopy and ICP-AES. *Microchemical Journal*, **73**, 299-306.

Workman, J. (1993). A brief review of the near infrared measurement technique. *NIR news*, **4**(6), 8-16.

Zoffoli, J.P, Balbontin, S. & Rodriguez, J. (2002). Effect of modified atmosphere packaging and maturity on susceptibility to mealiness and flesh browning of peach cultivars. In: *Proceedings of the 5th International Peach Symposium*, Volume 1. Pp. 573-579. July, 2001. Davis, CA.

Chapter 3

Prediction of soluble solids content and post-storage internal quality of Bulida apricots using near infrared spectroscopy (NIRS) with PLS, MARS and SIMCA as chemometric regression and classification techniques

Abstract

The development of internal breakdown during cold storage of South African canning apricots (Bulida) necessitates canning after harvesting or short term cold storage of the fruit. The logistics of the peak harvesting period and the limited capacity of the factory do not allow immediate canning of all the fruit harvested. Production losses occur when the fruit becomes unsuitable for canning during cold storage. The prediction of post-storage quality using Fourier transform near infrared spectroscopy (FT-NIRS) and chemometric classification techniques was investigated. The prediction models were based on near infrared (NIR) spectra, measured on the intact fruit prior to storage, and subjective quality evaluations performed after four weeks of cold storage. The chemometric regression method, multivariate adaptive regression splines (MARS), obtained a correct classification rate of 69%, compared to the 50% obtained by soft independent modelling by class analogy (SIMCA). A preliminary feasibility study showed that 4% (amounting to an income of R117 720) of the production losses can be saved annually if near infrared spectroscopy (NIRS) and MARS are used to predict post-storage quality. NIRS was also used to non-destructively measure soluble solids content (SSC) just after harvesting and after four weeks of cold storage. Soluble solids content was measured on the intact fruit as well as on the fruit flesh after part of the skin was removed, to confirm sufficient NIR light penetration into the fruit. Measurement of the SSC also provided an index of pre- and post-storage ripeness and possibly also of the quality of the apricots. Partial least squares (PLS) regression and MARS were used as regression methods. When one spectral and SSC measurement per fruit was used in regression, validations resulted in correlation coefficients (r) of 0.39-0.72 (PLS) and 0.51-0.82 (MARS) and standard error of prediction (SEP) of 1.59-2.21 °Brix (PLS) and 1.54-2.19 °Brix (MARS), depending on the data set used. Better prediction results were obtained when two spectral measurements per fruit (one on each half of the fruit), allocating each the same SSC value (one SSC value measured per fruit), were used in PLS regression. PLS regression prediction results were r = 0.65-0.88 and SEP = 1.24-1.46 °Brix. More spectral variability for each fruit resulted in a more robust regression model with better prediction results. NIRS

measurements made on the intact fruit did not result in poorer validation results than NIRS measurements made on the fruit flesh. This confirms sufficient NIR light penetration into the intact fruit, allowing future measurements to be made non-destructively. This preliminary study showed promising results for further, more in-depth and larger scale research on the prediction of post-storage quality of Bulida apricots.

Introduction

Bulida apricots are mainly used for canning purposes in the Western Cape Province of South Africa (Victor, 2002). Annual production losses of these apricots can amount to 900 tons (approximately R1.4 million) of the total 45 000 tons of the fruit received for canning each year. The fruit is ideally canned upon arrival at the factory but, as the harvesting season commences its peak period, approximately 30% of the fruit requires cold storage for one to two weeks owing to the limited canning capacity of the factory (Steyn, 2003). The production losses are the result of internal breakdown within the fruit during cold storage, which causes the fruit to be unsuitable for canning purposes (Victor, 2002).

Internal breakdown in apricots is referred to as gel breakdown (Eksteen & Combrink, 1994). This disorder appears as a browning around the stone, a softening of the flesh (Ryall & Pentzer, 1982) and a gelatinous texture (Fernández-Trujillo et al., 1998). Gel formation in stone fruit is attributable to the accumulation of insoluble high molecular weight pectins during cold storage (Ben-Arie & Lavee, 1971).

South African apricots are graded upon arrival at the factory and farmers are compensated for their produce according to the quality of a small sample taken for grading. The quality for canning is determined by the fruit size, colour, ripeness and bruising. The fruit diameter should be at least 36 mm for canning grade 1 (best quality) and 34 mm for canning grade 2. The colour of the fruit is compared to a standard colour chart. Ripeness is not measured but is evaluated by a subjective evaluation of the stage of fruit development. The fruit should be at the right stage of development during grading to ensure a firm and ripe fruit after canning. Light bruising is acceptable for canning grade 1 and medium bruising acceptable for canning grade 2 (Steyn, 2003). The potential of apricots to develop internal breakdown cannot be measured before or after storage, but can only be identified subjectively after the fruit had been destoned (Victor, 2002).

Other possible measures for fruit quality include soluble solids content (SSC), firmness and titratable acidity (Byrne, 2002), although these measures are not used in the South African industry. These measurements are not necessarily related to the occurrence of

internal breakdown or to the post-storage internal quality. Non-destructive measurements need to be done on an individual fruit basis to ensure consistently high quality fruit (Byrne, 2002). This testing should also include a measurement to indicate post-storage fruit quality (Victor, 2002).

Soluble solids content comprises sugar (as the major component), organic and amino acids and soluble pectins. Sugar content can be measured directly via chemical analysis, however, since sugars are usually the major component of soluble solids, it is much easier to measure the total SSC in extracted juice with a refractometer (Wills *et al.*, 1989) Both these methods are, however, destructive.

Previous studies (Carlini *et al.*, 2000) showed that visible and near infrared (Vis/NIR) spectroscopy could be applied to measure SSC in intact apricots at harvest. Using individual wavelength selection, a regression coefficient (r) of 0.94 and SEP of 0.75 °Brix were obtained.

Near infrared spectroscopy (NIRS) was also used to measure post harvest SSC in various other intact fruits such as tomatoes (Slaughter, 1996), pineapples (Guthrie *et al.*, 1998), apples (Ventura *et al.*, 1998; Zude & Herold, 2001; McGlone *et al.*, 2002), kiwifruit (McGlone & Kawano, 1998; Schaare & Fraser, 2000), mangos (Schmilovitch *et al.*, 2000) and peaches (Kawano *et al.*, 1992, 1995; Slaughter, 1995; Peiris *et al.*, 1998; Ortiz *et al.*, 2001).

Considerable SSC variability exists within individual fruit (peach, apple, cantaloupe, grapefruit, melon, orange, peach, tomato and pineapple) and this variability may affect the accuracy of calibration equations and their prediction capability (Peiris *et al.*, 1999).

Fresh fruit is characterised by a high moisture content. The spectral region from 700-1100 nm is relevant because water absorption peaks are less strong and broad. As a result there exists less chance of concealing spectral information correlated to low concentrations of constituents. Light from this region can also penetrate much further in fruit of different species (Carlini *et al.*, 2000). The visible part (400-700 nm) of the electromagnetic spectrum is important as chlorophyll and anthocyanin (indices of ripeness) absorption bands belong to this region (McGlone & Kawano, 1998).

In addition to quantitative measurements, NIRS could also be used for qualitative measurements. Successful qualitative classifications were made using NIRS and chemometrics, such as the determination of cultivation areas of ginseng samples (Woo *et al.*, 2002), classification of varieties of vegetable oils (Sato, 1994) and differentiation between aged and normal soybeans (Kusama *et al.*, 1997).

By analysing the spectra of the entire near infrared (NIR) wavelength range measured on an intact sample, it may be possible to obtain more information about the sample (Murray, 1999). Information about internal changes in fruit texture and changes in constituent

concentrations (sugars, organic acids and pectin) which occur within apricots during ripening and storage may lie buried within the NIR spectra. This information can possibly be revealed by chemometric methods and lead to a means to predict post-storage quality of intact apricots.

Objectives

The objectives of this study were to develop NIRS calibrations for determining pre- and post-storage SSC as an index of fruit ripeness. Validation results would also confirm sufficient NIR light penetration into intact Bulida apricots. The possibility of predicting the post-storage internal quality of intact apricots was also investigated. These predictions would be based on NIR spectra, measured just after harvesting, and subjective evaluations of post-storage quality. The final objective was to perform a feasibility study to determine whether the South African apricot canning industry will benefit from this classification method.

Materials and methods

Samples

Bulida apricots from different production areas in the Western Cape Province of South Africa were harvested for analyses during December 2001. The fruit was collected and packed onto moulded cardboard trays (to separate the fruit) in cardboard cartons (Figure 1) and transported to the ARC Infruitec-Nietvoorbij premises in Stellenbosch. Four cartons containing 40 apricots each (total of 160 fruit) were received every day over a three-day period (12-14 December, n = 480). The apricots were numbered individually (1,2,3, …, n) and on both halves (a or b) with a soft permanent marker pen for identification after storage. The fruit were marked on the areas that would not be used for any of the measurements. The fruit was not washed or treated with chemicals before analysis. The sequence of the procedures followed for the experiment are shown in Figure 2. The NIR spectra of the fruit were measured on arrival and at room temperature (21°C). Each day, after NIR measurements, fruit of one carton was cut on both halves to expose the fruit tissue and the NIR spectra measured again. Cuts were made manually on the fruit surface with a sharp fruit knife. These cuts were approximately 15 mm in diameter and were made just deep enough to remove the skin of the area to be measured with NIRS. The SSC was subsequently measured on each fruit. The whole fruit of the remaining three cartons (cartons 1-3) was cold stored at -0.5°C for a period of four weeks.

Figure 1. Carton of Bulida apricots as they arrived at the laboratory for NIRS measurements.
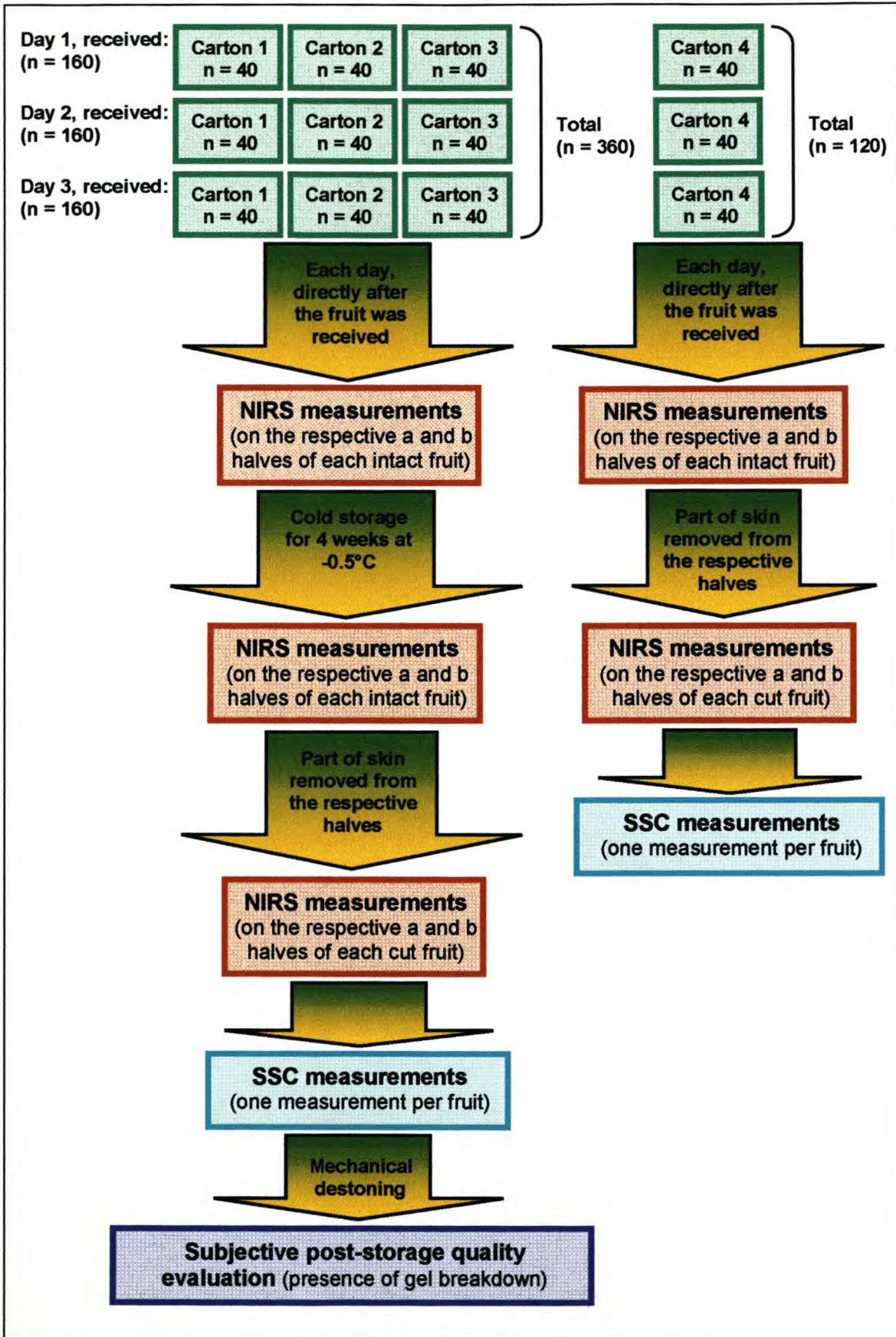
Figure 2. Sequence of procedures followed during the experiment.

NIR spectral measurements were performed on intact fruit after the four-week cold storage period (8-10 January 2002). The fruit was cut (15 mm in diameter) on both halves to expose the fruit tissue and the NIR spectra measured again. SSC measurements were subsequently performed on each fruit. The fruit was destoned mechanically and subjective evaluations of post-storage quality were made.

The storage period of four weeks was chosen to identify clear differences between fruit with good and poor post-storage quality and also to ensure that enough fruit with poor quality were available for the development of a model for poor storage potential.

Spectral acquisition

The optical absorption spectrum was measured on each half of the intact and cut fruit, respectively, using a Perkin Elmer Spectrum Identicheck™ 2.0 FT-NIR system in reflectance mode. Each fruit was positioned on the Identicheck Reflectance Accessory (ICRA), where light interacted with the intact fruit. Every absorption spectrum was the average of 16 individual optical scans in the wavelength range of 645 to 1201 nm at 2 nm intervals and a resolution of 64 cm^{-1}.

Sample evaluation

Immediately after the last NIRS measurement was performed on the cut fruit, the SSC was measured on each fruit. The SSC, expressed in °Brix, was determined with an Atago N1 hand refractometer. Juice from the fruit was dropped directly on the refractometer after the flesh was punctured. The sampling area of the refractometer was cleaned with tissue paper after each measurement. Only one SSC measurement was made on each fruit due to the small size of Bulida apricots. Following SSC measurements, each fruit (that was stored) was destoned and subsequently evaluated subjectively for gel breakdown (internal breakdown). The presence of gel breakdown was recorded for each half of the fruit. A fruit half which had gel breakdown was evaluated as having poor post-storage quality and allocated a value of one. A fruit half without gel breakdown was evaluated as having good post-storage quality and allocated a value of zero.

Data processing

Spectral data sets were merged with the respective SSC and post-storage quality data. Figure 3 shows which data sets were merged for the regression and classification models. The spectral data sets were subjected to pre-processing and spectral evaluation.
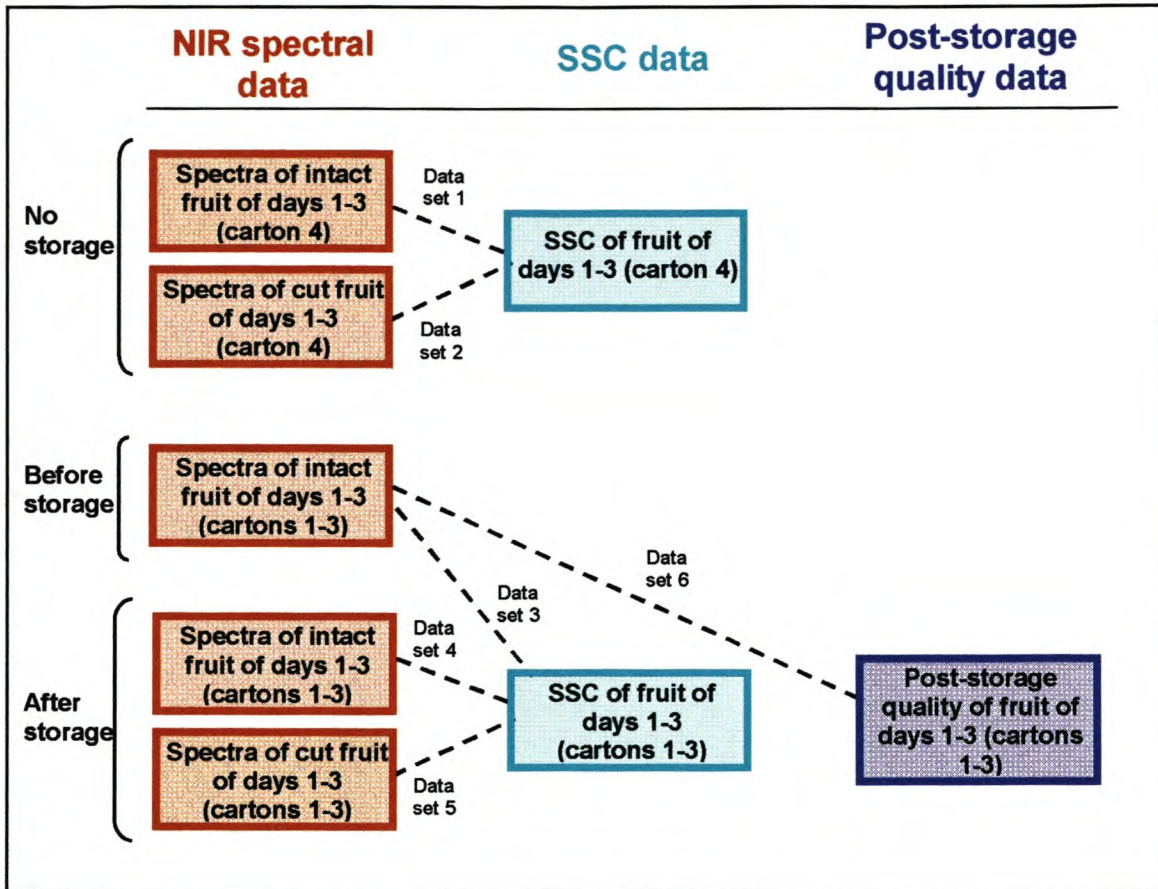
62

Figure 3. The respective spectral and reference data sets merged for SSC regression models and classification models of post-storage quality. The merged data sets were numbered as data sets 1-6)

Pre-processing of spectral data

The spectral data sets were subjected to mathematical pre-processing in order to resolve overlapping bands, to investigate absorption peaks and to identify differences in the spectra of the fruit just after harvesting and after storage. Multiplicative scatter correction (MSC) and first derivatives (Savitzky-Golay) were applied to the spectra. MSC compensates for additive and multiplicative effects in the spectra due to physical phenomena, such as differences in particle size. Derivation is relevant for spectral variables which are themselves functions of some underlying variable. The goal of derivation is baseline correction and resolution enhancement. Pre-processing was also performed when it resulted in better regression results.

Spectral evaluation

The spectra of low (9 °Brix) and high (17 °Brix) SSC of intact and cut apricot fruit were evaluated. For each SSC value, different spectra (intact fruit before storage, intact fruit after storage and cut fruit after storage) of the same sample were evaluated.

Development of regression models

Partial least squares (PLS) regression

Partial least squares (PLS) regression (Sjöström *et al.*, 1983; Næs & Martens, 1984) finds a few relevant linear combinations (principal components) of the original spectral values as well as the reference measurements and use only these linear combinations in a regression equation. These methods can produce stable regression equations while compressing data and removing multicollinearity successfully.

Spectral data were merged with the respective SSC data using The Unscrambler™ program, version 6.11 (CAMO ASA, Norway). A master data file was made for each data set (data sets 1-5, Figure 3).

The calibrations were performed on approximately two-thirds of each data set. The calibration samples were chosen to cover the entire range of the reference data and included the highest and lowest reference value. Independent validation was performed on each data set. The validation set included the other third of the data set (thus samples not used in calibration) and also covered the entire range of the reference data.

Initially, spectra of only one half of each fruit were used in regression, as only one SSC measurement was made on each fruit. A further regression was performed on each data set, using spectra of both halves of each fruit. The two spectra of each fruit were allocated the same SSC value. The inclusion of both spectra provided more spectral variability.

64

Multivariate adaptive regression splines (MARS)

Multivariate adaptive regression splines (MARS), developed by Friedman (1991), is an extension of piecewise linear regression, where more than one regression line is fitted to the data to account for non-linear relationships. This is achieved by the use of separate regression slopes in distinct intervals of the predictor variable space. The slope of the regression line can change from one interval to another and the point where two regression lines cross, is called a knot point. These variable intervals are defined by basis functions. As basis functions are added, the model becomes more flexible and more complex. MARS can handle more than one independent (predictor) variable as well as combinations of categorical and continuous predictors.

The MARS™ program, version 2 (Salford Systems, California, USA) was used to develop MARS regressions. Spectra from only one half of each fruit were used, as there was only one SSC measurement made on each fruit. A MARS model was subsequently fitted for each data set. The calibration (training) set consisted of 80% of the total samples in each season. The model was then tested on the validation (test) set, which consisted of 20% (samples not used for calibration) of the total samples in each season. The samples for calibration and validation were randomly selected.

The regression results calculated for each PLS and MARS regression included the standard error of calibration (SEC), root mean square error of prediction (RMSEP), standard error of prediction (SEP), and bias (Equations 1-4) (Næs & Isaksson, 1991).

$$SEC = \frac{1}{n-1-t} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad \ldots\ldots\ldots\ldots 1$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n-1}} \qquad \ldots\ldots\ldots\ldots 2$$

The SEP is calculated similar to the RMSEP, but corrected for bias.

$$SEP = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i - BIAS\right)^2}{n-1}} \qquad \ldots\ldots\ldots\ldots 3$$

65

Bias (average prediction error) is interpreted as the systematic difference between the predicted and measured values.

$$BIAS = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)}{n} \qquad \dots\dots\dots\dots 4$$

where n =   the number of samples in the calibration (equation 2) or validation (equation 3) set

y_i =   the measured reference value for object i

$\hat{y}_i$ =   the predicted value for the same object obtained from the calibration equation

t =   the number of independent variables in the calibration equation

Classification of post-storage quality

Soft independent modeling by class analogy (SIMCA)

Soft independent modeling by class analogy (SIMCA) (Wold, 1976) is a multivariate classification method that is not restricted by the dimension of the data matrix, thus often used for NIRS data. The basis for SIMCA is that objects in one class, or group show common similarities, properties or a specific class pattern, which makes all the objects in this class more similar than with respect to any other class. The goal of SIMCA classification is to assign new objects to the class to which they show the largest similarity (Esbensen, 2000).

SIMCA was performed using The Unscrambler™ program, version 6.11 (CAMO ASA, Norway). Before classification, the spectral variation of the data sets were analysed by principal component analysis (PCA). PCA was performed on mean centered spectra (subtraction of an average spectrum from each individual spectrum for resolution enhancement) to obtain an overview of the spectral data. Outlier spectra were eliminated and possible clustering or classes were investigated. SIMCA was performed on the data set which contained spectra measured on intact fruit before storage and evaluations measured after storage (data set 6, Figure 3). A calibration set, containing approximately two-thirds of the sample set (n = 236) and a validation set, containing the other third of the sample set (n = 117), were created for the data set. The calibration set was used to calculate two separate PCA models, a model indicating good post-storage quality and a model indicating poor post-storage quality. These models were based on results of the subjective evaluations. For each PCA model, the optimum number of principal components (PC's) were determined. Finally,

the validation samples were evaluated by these models according to their analogy to each model. The measurements used for classification are the object-to-model distance (Si) and the leverage or distance to model center (Hi). Each object in the test set was classified to the model to which is had the smallest distance to (Si). At comparable object distances, the object was closest to the model with which it had the smallest leverage (Hi) (Esbensen, 2000).

Multivariate adaptive regression splines (MARS)

Multivariate adaptive regression splines (MARS) (Friedman, 1991) was applied as an alternative method used for classification of post-storage quality. MARS can be used in a binary (which only have two levels) classification setting. The binary response variable can be coded as 0 or 1. When the MARS model is fitted, estimated values for the response variable will lie between 0 and 1. A threshold (e.g. 0.5) can be chosen for classification purposes.

The MARS™ program, version 2 (Salford Systems, California, USA) was used for MARS classification. The subjective evaluations after storage were used as response variables and spectral measurements made on one side of the intact fruit before storage were used as the predictor variables. The subjective response variables were coded as zero (good post-storage quality) and one (poor post-storage quality). The calibration set consisted of 75% (n = 266) and the validation set the other 25% (n = 87) of the data set. The MARS model was fitted on the validation set and a threshold value of 0.36 was chosen for classification purposes, as this value gave the best classification results.

Outliers

The final regression and classification models were obtained after eliminating the outliers for each model. Outlier samples are likely to have a high influence on the model and could be harmful to the model. Defective or outlier data were data points that were abnormal compared to the major part of the data. They were identified by investigating the respective PCA, leverage and influence plots, which confirmed that they were outliers. The leverage can be interpreted as the distance between a projected point and the model center. The influence plot is a plot of the squared residuals versus the leverage.

Feasibility study

The feasibility of implementing NIRS to predict post-storage quality in Bulida canning apricots was evaluated. The classification result obtained using the MARS method was applied. The losses due to poor post-storage quality were calculated for the two scenarios, e.g.:

1. No prior prediction of post-storage quality made (current situation) and
2. Prediction (based on NIR spectra and MARS) of post-storage quality made prior to storage (proposed method)

The feasibility of the classification method was calculated if 30% (the current situation in the industry) of the apricots had to be stored upon arrival at the factory.

Results and discussion

Descriptive statistics of SSC measurements

Overall descriptive statistics of SSC measurements, including mean, minimum, maximum and standard deviations are given in Table 1. The standard error of laboratory (SEL) was not calculated as measurements were only performed on one side of the fruit.

NIR spectra of the apricot

The original spectra (no pre-processing) from 645-1185 nm of some of the apricots are shown in Figure 4. The peaks in the spectra represent frequencies at which light is absorbed by the fruit. Peaks are visible at 645-740 nm and at 979 nm. The peak at 979 nm is due to OH overtones which could indicate moisture (Osborne *et al.*, 1993) or carbohydrates (Williams & Norris, 1987; Osborne *et al.*, 1993). Figure 5 shows the same spectra treated with MSC and first derivative. MSC corrected for all the differences in the spectra due to unwanted physical elements, while first derivatives enhanced the significant absorption peaks. The major absorption peaks for apricot fruit were in the visible range (683 nm) as well as at 721, 955 and 1143 nm. McGlone & Kawano (1998) identified peaks for kiwifruit at 680 nm. This is the red end of the visible spectrum. High absorbance in this area is indicative of red absorbing pigments, particularly chlorophyll (McGlone & Kawano, 1998). Chlorophyll is related to fruit colour (Moons *et al.*, 1998). Williams & Norris (1987) assigned moisture to 958 nm. The peak at 721 nm could be due to OH overtones, also indicating moisture (Osborne *et al.*, 1993). The peak at 1143 nm correlates with results of Kawano *et al.* (1995), who identified peaks for SSC at 1141 nm.

68

Table 1. Descriptive statistics for SSC (°Brix) measured on all the samples (cartons 1-4).

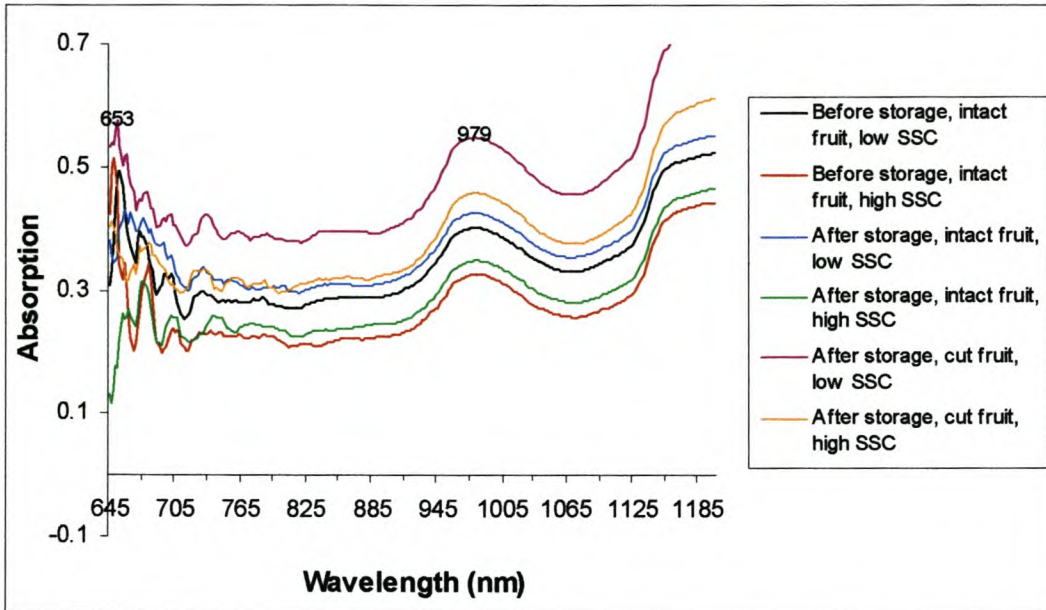| Data set	n	Mean	Min.	Max.	Std. Dev.
No storage, days 1-3 (carton 4)	120	14.35	7.8	20.0	2.88
After storage (cartons 1-3)	360	15.56	8.2	20.2	1.81

Figure 4. NIR spectra (no pre-processing) of samples from the various data sets. The spectra of samples with low (11 °Brix) and high (17 °Brix) SSC values are shown for each data set.
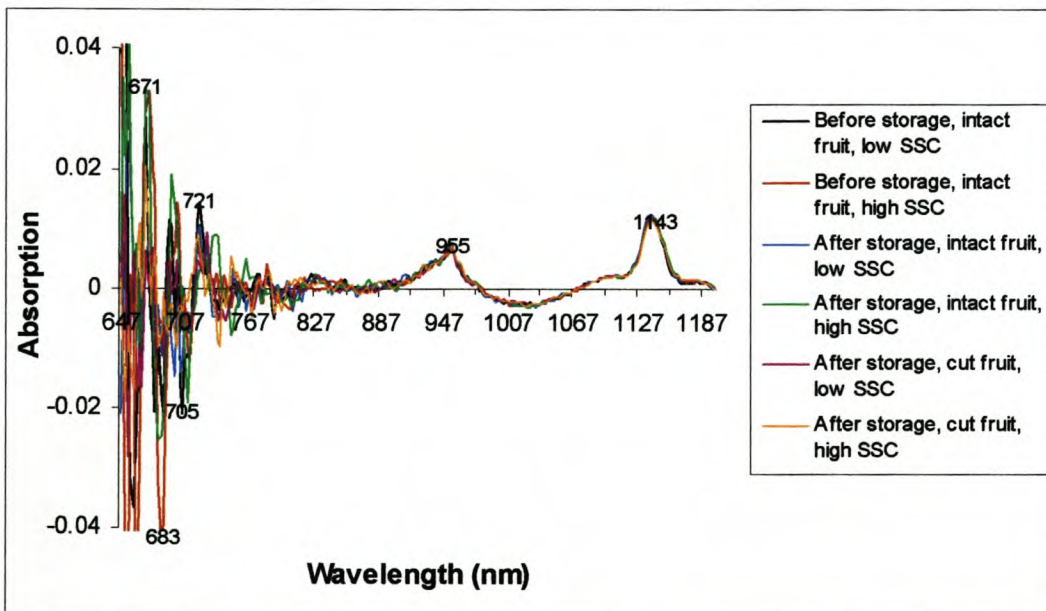

Figure 5. NIR spectra treated with first derivatives and MSC. The spectra of samples with low (11 °Brix) and high (17 °Brix) SSC values are shown for each data set.

69

Figure 6 shows the spectra measured on the same sample before and after storage. The apricot spectra measured before storage had a larger absorption peak at 679 nm. This indicates that the chlorophyll decreased during the storage period. This was expected, as the chlorophyll concentration in apricots decreases as the fruit matures. Carotenoids (yellow pigments) increase during apricot maturation (Chahine *et al.*, 1999), but peaks indicating carotenoids are not visible in the wavelength range measured, as these peaks are at 400-490 nm (Atkins, 1996).

Soluble solids content

Regressions were preformed on the various data sets, using PLS and MARS. Regressions were initially performed using spectra from one half (a-half) of each fruit. Tables 2 & 3 show the regression results obtained using the PLS and MARS methods, respectively. Correlation coefficients (r) of 0.39-0.72 (PLS) and 0.51-0.82 (MARS) and SEP's of 1.59-2.21 °Brix (PLS) and 1.54-2.19 °Brix (MARS) were obtained. Although the samples selected for the calibration and validation sets for the two methods were not identical, the two methods produced similar results. As the regression was linear, the ability of MARS to fit more than one regression line to compensate for nonlinear relationships did not improve on the PLS model.

PLS regressions were also performed using spectra measured on both halves of each fruit, instead of only one half. The two spectra of each fruit were allocated the same SSC value as reference value. The regression results and prediction plots are shown in Table 4 and Figures 7-11, respectively. These regression results were better than those obtained using only one spectral measurement per fruit. Correlation coefficients (r) of 0.65-0.88 and SEP's of 1.24-1.45 °Brix were obtained. A possible reason for the improved results is that more variability was included in the spectra of each fruit, making the model more robust. The results were, however, poorer than the results (r = 0.94 and SEP 0.75 °Brix) obtained by Carlini *et al.* (2000), who used Vis/NIRS, with wavelength selection and PLS regression.

The best correlation coefficients (0.72-0.88) were obtained where there was no storage time spectral and SSC measurements. The best SEP values (1.24-1.59 °Brix) were obtained where there was a four-week storage period between spectral measurements and SSC measurements. Validation results obtained on intact fruit were similar to results obtained on the fruit flesh after removal of the skin. This confirms sufficient NIR light penetration into the intact fruit. Non-destructive measurements on apricot fruit are thus possible.

The optimum number of PC's for the PLS models ranged from 2-17 PC's. These numbers were determined by monitoring the residual matrix (E-matrix) for the accumulating PC model fit. Figures 12-16 show the residual validation variance plots correlating to each.
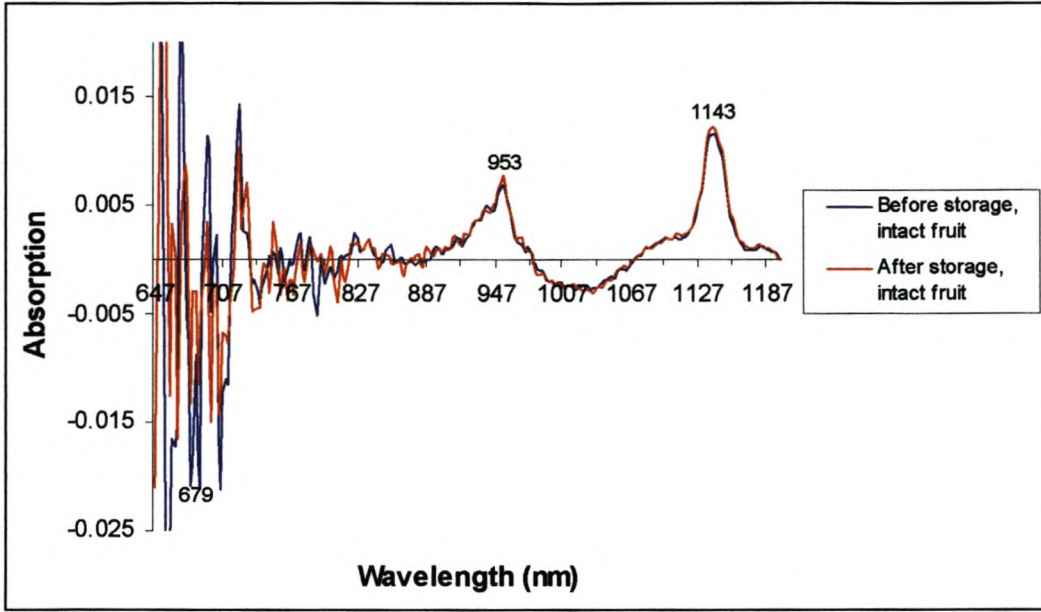
Figure 6. NIR spectra treated with first derivatives and MSC. The spectra were measured on the same sample before and after storage.

Table 2. PLS regression statistics for apricot SSC using spectra measured on one half (a-half) of each fruit.

Statistic	Data set used for regression				
	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5
Calibration					
No. of PC's	2	5	5	4	4
n	75	77	230	237	232
r	0.65	0.80	0.45	0.30	0.43
SEC (°Brix)	2.22	1.76	1.59	1.76	1.60
RMSEC (°Brix)	2.23	1.77	1.59	1.76	1.61
Bias (°Brix)	0.002	0.002	0.002	0.002	0.001
Validation					
n	38	38	116	118	117
r	0.59	0.72	0.42	0.39	0.43
SEP (°Brix)	2.21	1.92	1.59	1.63	1.59
RMSEP (°Brix)	2.23	1.89	1.58	1.62	1.58
Bias (°Brix)	0.23	-0.13	0.09	-0.003	0.04

Table 3. MARS regression statistics for apricot SSC using spectra measured on one half (a-half) of each fruit.

Statistic	Data set used for regression				
	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5
Calibration					
No. of basis functions	15	15	15	30	20
n	90	90	284	284	279
r	0.92	0.79	0.68	0.77	0.80
SEC (°Brix)	1.13	1.82	1.32	1.16	1.09
RMSEC (°Brix)	1.13	1.82	1.32	1.16	1.09
Validation					
n	23	23	71	71	70
r	0.82	0.57	0.53	0.51	0.58
SEP (°Brix)	1.56	2.19	1.65	1.56	1.54
RMSEP (°Brix)	1.56	2.19	1.65	1.56	1.54

Table 4. PLS regression statistics for apricot SSC using spectra measured on both halves of each fruit.

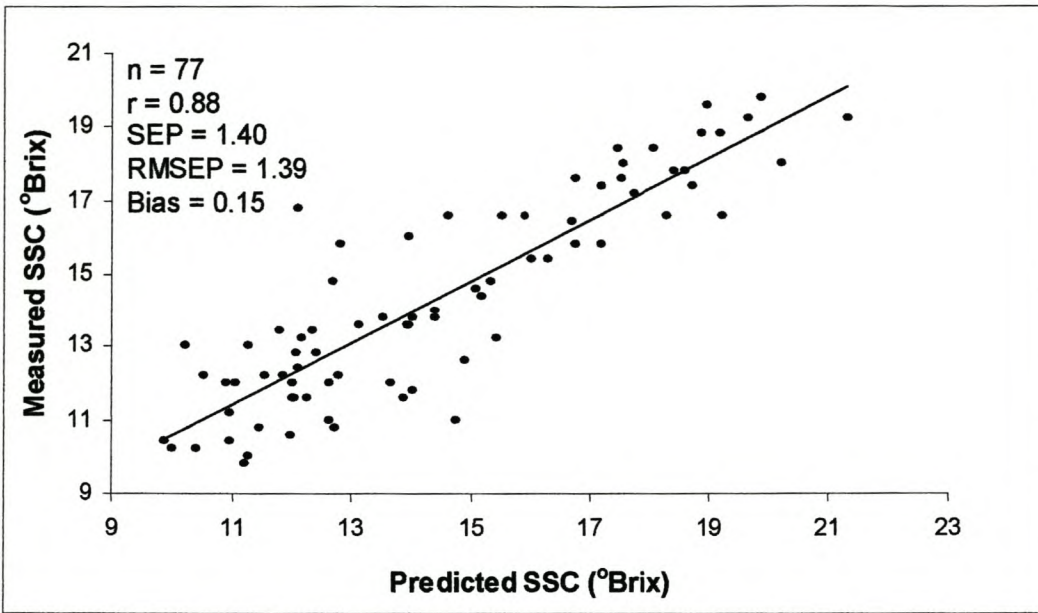| Statistic	Data set used for regression				
	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5
Calibration					
No. of PC's	17	9	16	13	13
n	155	154	464	470	462
r	0.94	0.88	0.74	0.73	0.76
SEC (°Brix)	1.00	1.38	1.67	1.24	1.20
RMSEC (°Brix)	0.99	1.38	1.67	1.24	1.20
Bias (°Brix)	-0.01	0.002	0.01	-0.01	-0.001
Validation					
n	77	77	230	235	234
r	0.88	0.86	0.65	0.67	0.72
SEP (°Brix)	1.40	1.45	1.30	1.32	1.24
RMSEP (°Brix)	1.39	1.46	1.29	1.32	1.24
Bias (°Brix)	0.15	-0.03	-0.03	-0.01	-0.01

Figure 7. Validation plot of measured SSC (°Brix) vs. predicted SSC of intact fruit, using spectra and SSC measured before storage (data set 1) (no pre-processing; 17 PC's).
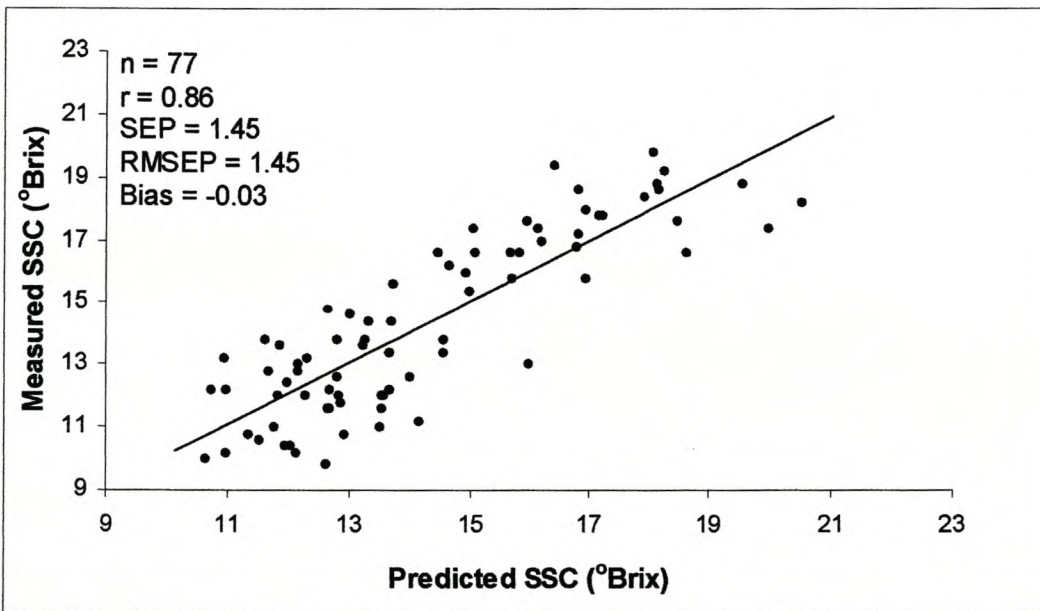

Figure 8. Validation plot of measured SSC (°Brix) vs. predicted SSC of cut fruit, using spectra and SSC measured before storage (data set 2) (no pre-processing; 9 PC's).
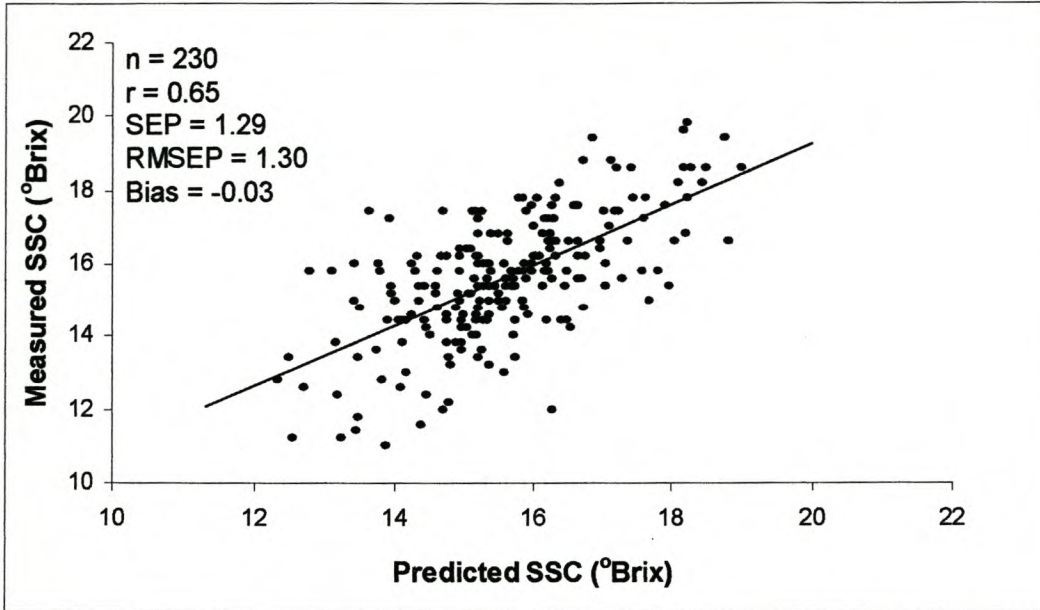
73

Figure 9. Validation plot of measured SSC (°Brix) vs. predicted SSC of intact fruit, using spectra measured before storage and SSC measured after storage (data set 3) (MSC corrected spectra; 16 PC's).
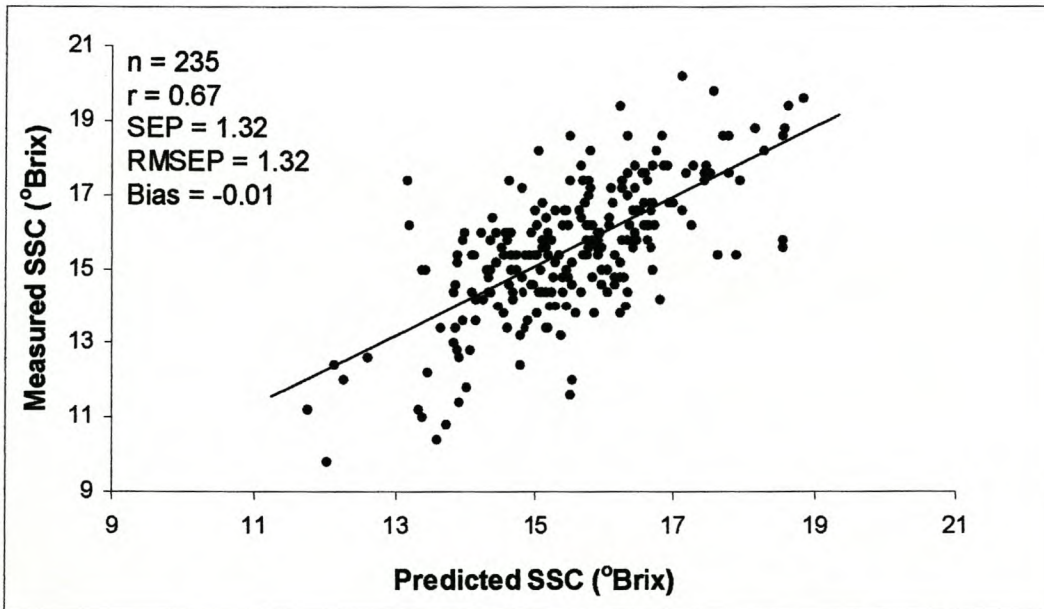

Figure 10. Validation plot of measured SSC (°Brix) vs. predicted SSC of intact fruit, using spectra and SSC measured after storage (data set 4) (MSC corrected spectra; 9 PC's).
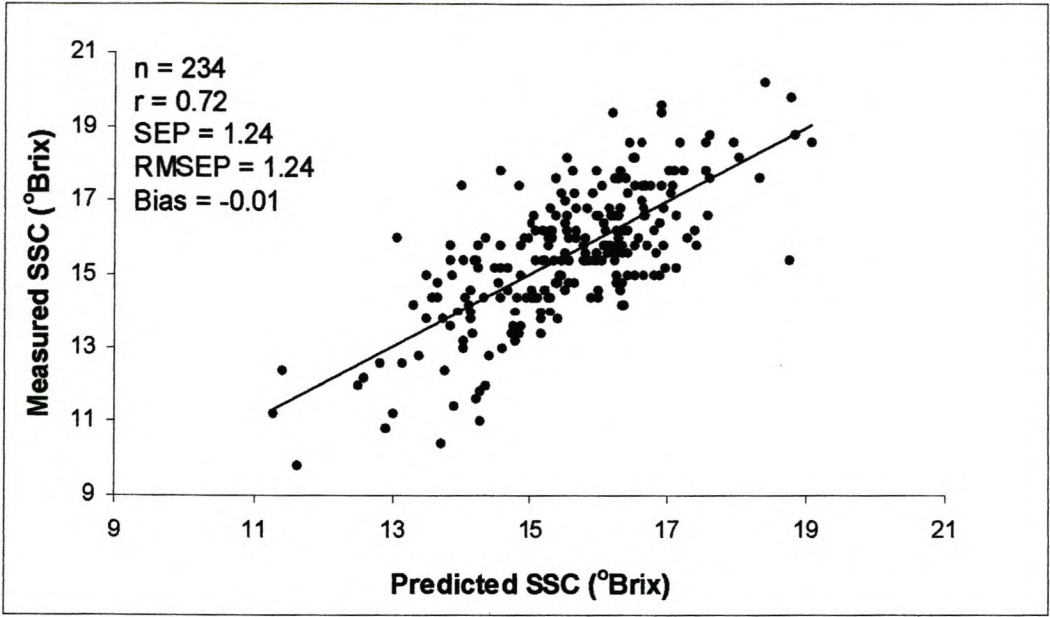
74

Figure 11. Validation plot of measured SSC (°Brix) vs. predicted SSC of cut fruit, using spectra and SSC measured after storage (data set 5) (MSC corrected spectra; 13 PC's).

Figure 12. Residual validation variance plot for the SSC regression model of intact fruit, using spectra and SSC measured before storage (data set 1).

Figure 13. Residual validation variance plot for the SSC regression model of cut fruit, using spectra and SSC measured before storage (data set 2),

Figure 14. Residual validation variance plot for the SSC regression model of intact fruit, using spectra measured before storage and SSC measured after storage (data set 3).
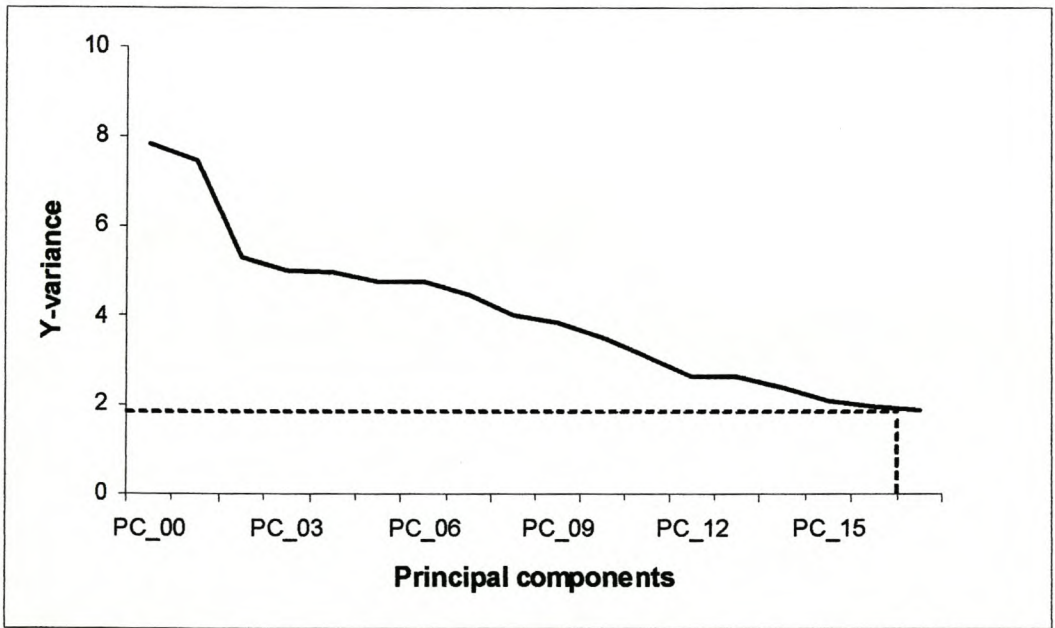
Figure 15. Residual validation variance plot for the SSC regression model of intact fruit, using spectra and SSC measured after storage (data set 4).
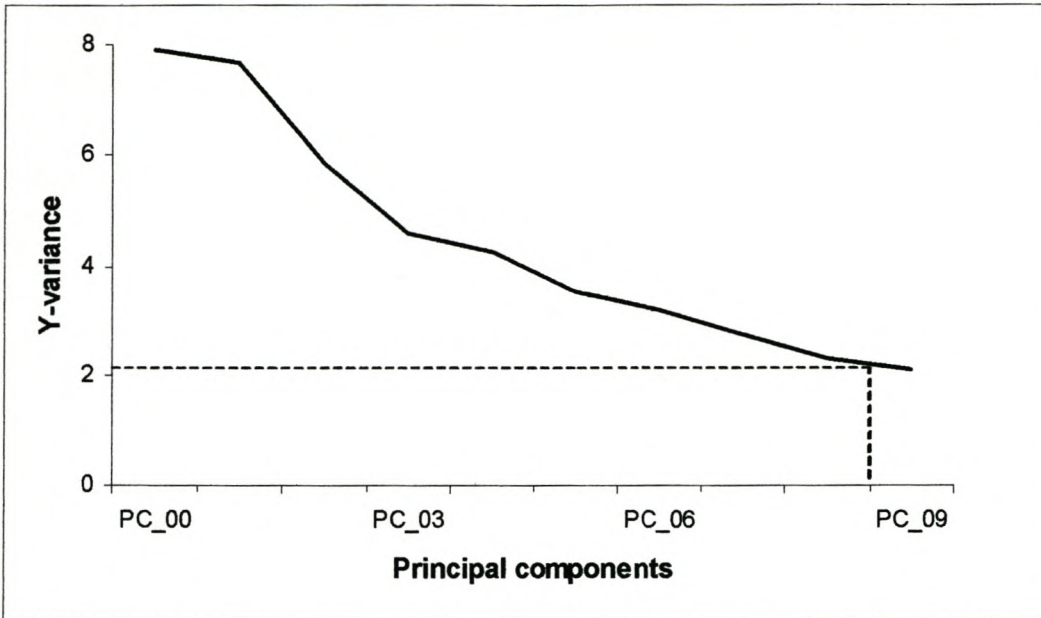

Figure 16. Residual validation variance plot for the SSC regression model of cut fruit, using spectra and SSC measured after storage (data set 5).
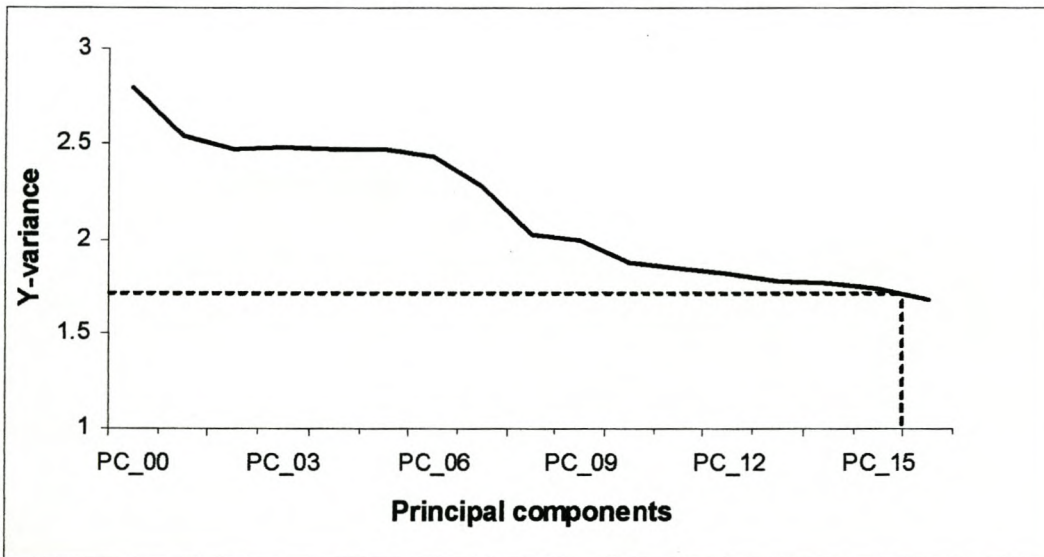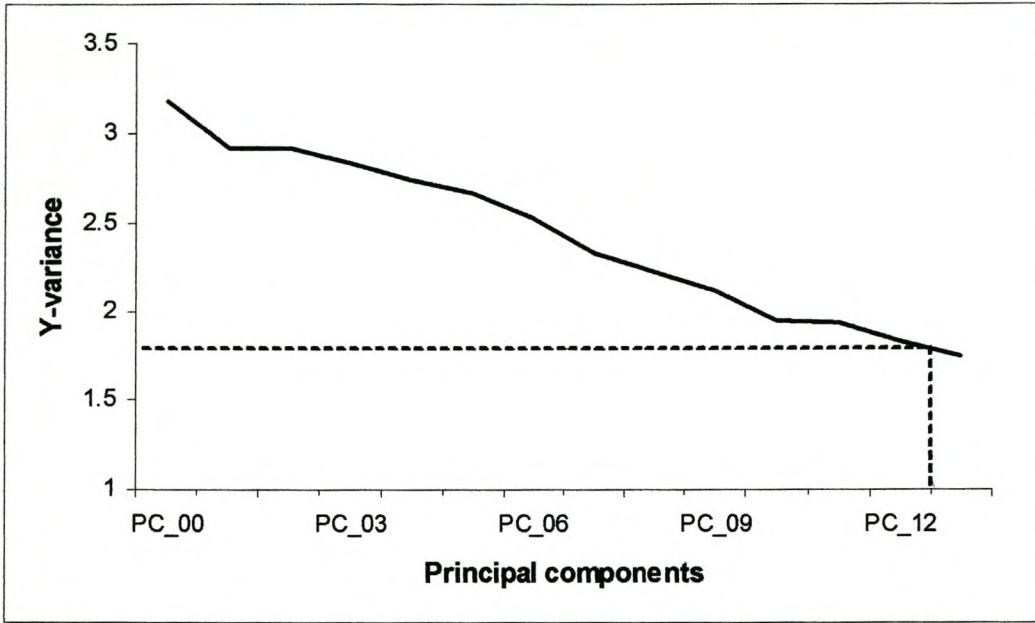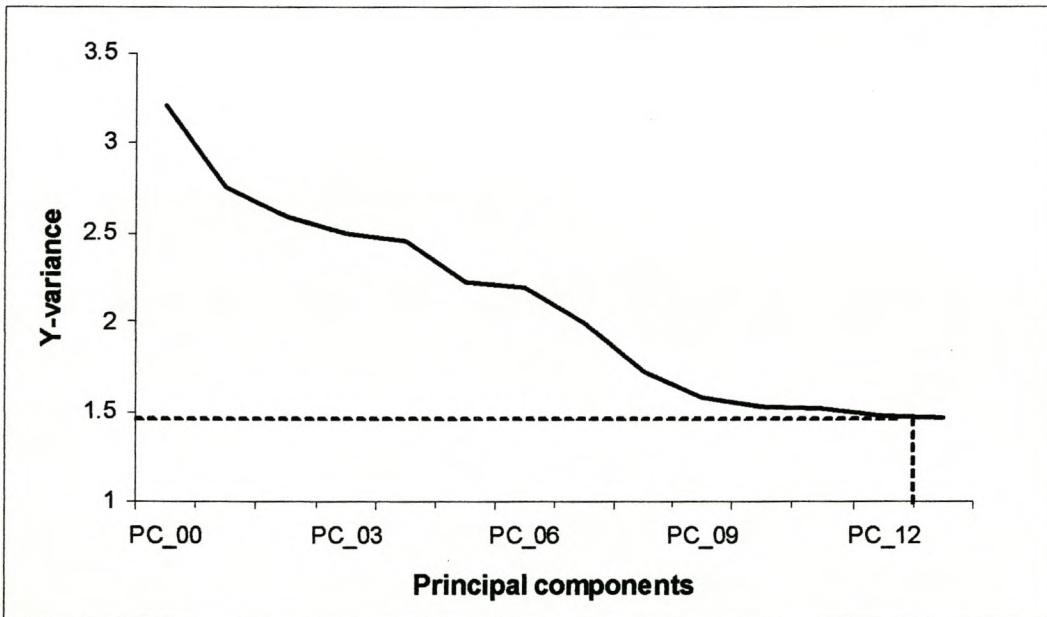
PLS regression shown in Figures 7-11. There was a gradual decrease and then a flattening in the residual variance in each curve. The PC which had the minimum Y-variance was chosen as the last PC to include in the model. The chosen number of PC's also gave the lowest SEP values in the regression results. These plots were also useful as an indication of how well the model was able to describe the relationships between X and Y. A decrease in the residual variance indicated a "good" model, i.e. that no extreme outliers were present and that there were systematic relationships between X and Y.

The importance of various X-variables (wavelengths) for predicting Y (SSC) was evaluated by plotting the PLS regression coefficients (B-coefficients) for each principal component (PC) of the model. Large absolute values for the variables indicate large significance, whereas small values indicate unimportant or noisy variables. Figure 17 shows the regression coefficients where spectra (of intact fruit) and SSC, measured before storage, were used for regression. The significant variables (wavelengths) used were 685 and 699 nm. Chlorophyll was assigned to these wavelengths (McGlone & Kawano, 1998). Figure 18 shows the regression coefficients where spectra measured (on intact fruit) before storage and SSC measured after storage were used for regression. The significant variables were those from the visible range as well as around 787, 913 and 1081 nm. Chlorophyll decreases as the fruit matures (Chahine et al., 1999). SSC increases slightly during storage (Pretel et al., 1993). The fruit maturity thus had a large influence in predicting SSC. Dull et al. (1989) used 913 nm to measure SSC in cantaloupe. The significant bands were also compared with chemical assignments summarised by Osborne et al. (1993). The C-H overtones can be assigned to 787, 913 and 1081 nm. C-H can be organic acids or carbohydrates (mono-, di- and oligosaccharides), which could be sugars and pectins (the constituents of SSC). Fruit ripening is known to be associated with an increase in sucrose and soluble pectin and a decrease in organic acids (Bartley & Knee, 1982; Lill et al., 1989). The wavelengths with the largest significance used in the measurement of SSC in apricots were assigned to constituents which all change in concentration during ripening and maturation of the fruit.

Similar to the regression coefficients for the PLS method, the most significant wavelengths used in MARS for pre-storage regression were in the visible range (653, 669 and 675 nm), once again indicating chlorophyll. The longer wavelengths (951, 1017, 1053, 1067, 1117, 1153 and 1191 nm) were more important in the post-storage SSC regressions. Kawano et al (1995) identified 951 nm in SSC regressions of peaches and Williams & Norris (1987) reported 1053 and 1153 nm to be absorptions of carbohydrates and water, respectively.
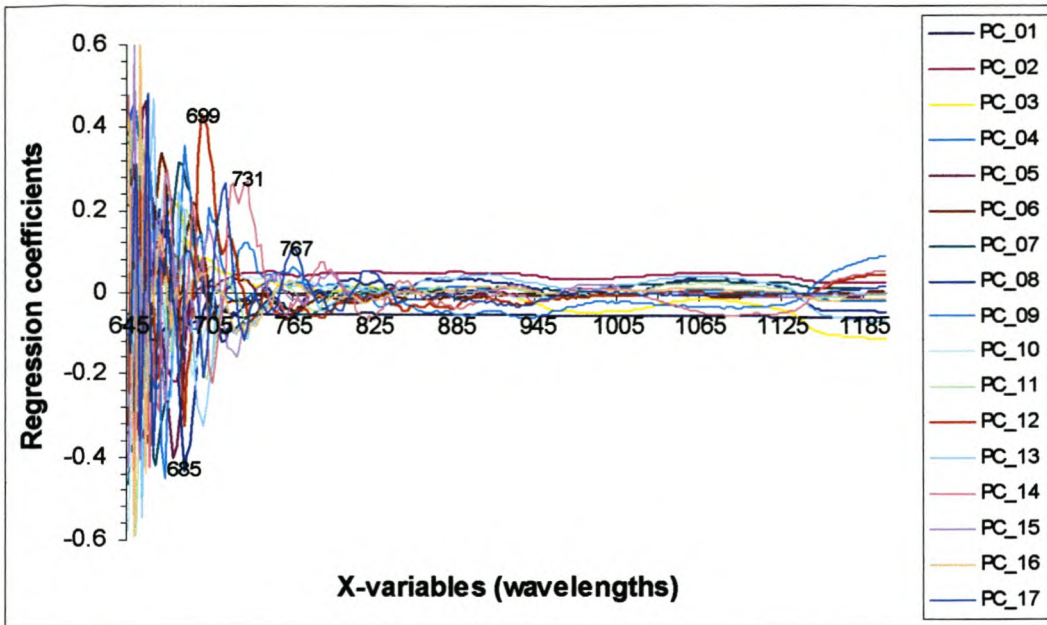
78

Figure 17. Regression coefficient plot of the 17 PC's calculated for the SSC regression model of whole fruit using spectral and SSC data before storage.
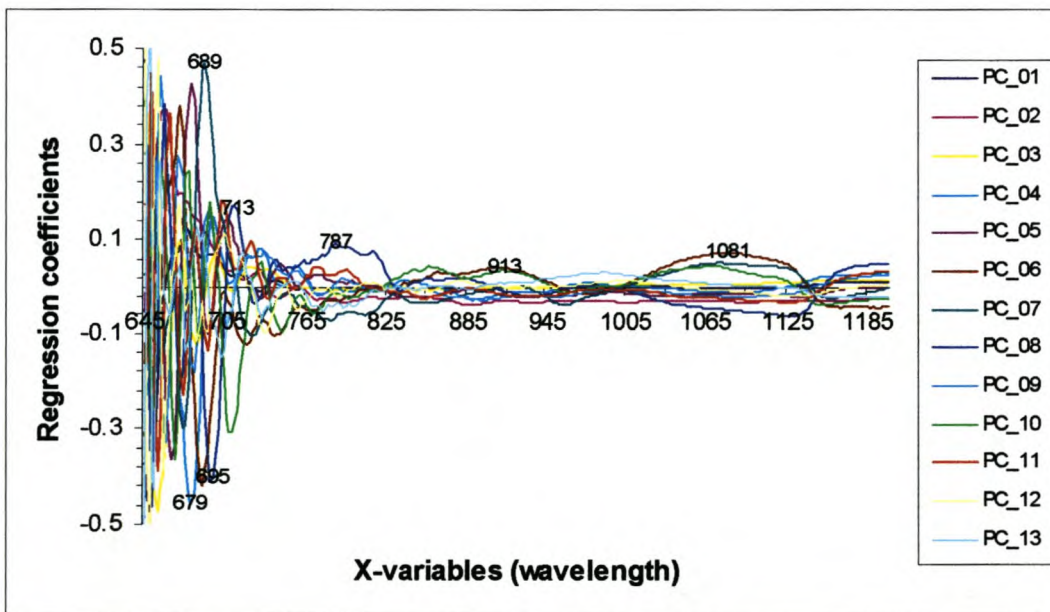

Figure 18. Regression coefficient plot of the 13 PC's calculated for the SSC regression model for whole fruit using spectral and SSC data after storage.

Classification of storage potential

Principal component analysis (PCA) was performed to inspect the spectral variation in the spectral data set. All possible combinations of the eight PC's calculated were plotted against each other (two- and three dimensionally) in order to find some clustering or pattern on which to base further storage potential classifications. The PCA plot of PC1 versus PC2 (Figure 19) showed that the majority of the samples with poor post-storage quality grouped together, but they could not be separated from those samples with good post-storage quality.

The sample set was subsequently subjected to the two classification techniques, SIMCA and MARS. In each case, a validation set was assigned to the post-storage quality models, which were created using the corresponding calibration set. The classification results are shown in Table 5. Classification results are given with respect to the model created for "good" post-storage quality. The recognition rate, or sensitivity, is the number of samples correctly classified as belonging to the model (classified to have good post-storage quality) out of the total number that truly belonged to the model (the total number subjectively evaluated as good post-storage quality). The rejection rate, or specificity, is the percentage of samples not assigned to the model (classified as poor post-storage quality) out of the total number of samples that truly did not belong to the model (the total number subjectively evaluated as poor post-storage quality).

Applying SIMCA resulted in a recognition rate of 47.87% and a rejection rate of 60.87%, with a total correct classification rate of 50.43%. MARS classifications obtained better results than SIMCA classifications. A recognition rate of 71.23% and a rejection rate of 57.14% resulted in a total correct classification rate of 69.00%. MARS can account for non-linear relationships in data (Friedman, 1991). A possible reason for the better classification rate obtained by MARS is that a non-linear relationship exists between the spectral data and the post-storage quality data.

The wavelengths that played a major role in the MARS and SIMCA classifications were those in the visible range (645-709 nm), indicating chlorophyll, as well as 841, 959 and 969-1005 nm. Williams & Norris (1987) assigned water to 841, 958, 978, 986 and 994 nm. These moisture peaks could relate to post-storage quality, as gel breakdown occurs when water is bound in a calcium pectate gel (Ben-Arie & Lavee, 1971). The wavelength at 1005 nm can be assigned to carbohydrates (C-H). These bands possibly explain the increase in pectin (mostly insoluble) which occurs during gel breakdown (Ben-Arie & Lavee, 1971). The wavelength which had the largest influence in the MARS classification was 817 nm. Osborne *et al.* (1993) assigned N-H third overtones to 815 nm. An increase in the amino acid, alanine, was reported when peaches were stored at 0°C and alanine accumulation has been

Figure 19. PCA score plot (PC 1 versus PC 2) of apricot spectra, indicating no obvious separation between good and poor post-storage quality.

Table 5. SIMCA and MARS classification results on each test set. There was a four-week cold storage period between spectral acquisition and subjective evaluations.

Method	n	Recognition rate		Rejection rate		Total classification rate	
		n	%	n	%	n	%
SIMCA	117	45/94	47.87	14/23	60.87	59/117	50.43
MARS	87	52/73	71.23	8/14	57.14	60/87	69.00

associated with chilling injury in other crops (Lill *et al.*, 1989). It is possible that this, or some other amino acid also had an influence on the classification of the post-storage quality in apricots.

Feasibility study

The results of the MARS model (Table 6) developed on the calibration set, were used for the subsequent calculations to determine the feasibility of implementing NIRS in the apricot canning industry.

By using the proposed classification method (NIRS and MARS), 36 tons (4%) of apricots can be saved from the total 900 tons of apricots lost annually. From the total annual losses of R1.35 million, an amount of R54 000 can thus be saved. If 36 tons of fruit is saved, the potential profit that will be gained from this saving amount to R63 720 (36 tons at a profit of R1770 per ton). The savings and potential profit add up to a total income of R117 720. Very few samples were used in the feasibility study, because the sample set for the study was small (n = 360) and only 71 (20%) fruit out of the total had poor quality after storage. This study on a larger scale would provide more representative results.

Conclusion

PLS and MARS regressions are both satisfactory methods for measuring SSC in Bulida apricots. This confirms that non-destructive measurements can be made on apricots using NIRS. Results were improved by including two spectral values per fruit for each SSC value. More spectral variation for each reference (SSC) value resulted in more robust calibration models with better prediction ability.

Acceptable classification results from this study showed that the accurate prediction of post-storage quality (storage potential) in apricots is possible. It is not, however, clear what the basis for such classifications should be. So far, ripeness, moisture and polysaccharides (pectin) are considered to have an effect on the post-storage internal quality. The implementation of NIRS in the factory, with MARS as the preferred classification method, could be of valuable benefit for the South African apricot canning industry. This research should, however, be continued on a larger scale and include a larger sample size for more representative results. A more in-depth study should continue to investigate the possible link between the storage potential and the internal constituent concentrations and interactions of the apricot fruit. This would assist in the understanding and improvement of the current classification models.

Table 6. Results of classification of storage potential on the calibration set using MARS.

	Good post-storage quality (predicted) (n)	Poor post-storage quality (predicted) (n)	Total (n)
Good post-storage quality (measured) (n)	52	21	73
Poor post-storage quality (measured) (n)	6	8	14
Total (n)	58	29	87

Number of fruit to be stored (if 30% of the fruit have to be stored) $= 0.3 \times 87$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = 26$

1. If no prior prediction of post-storage quality is made:

    Number of fruit with good post-storage quality $= (73/87)26$

    $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = 22\ (84\%)$

    Number of fruit with poor post-storage quality $= (14/87)26$

    $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = 4\ (16\%)$

2. If prediction (based on NIRS and MARS) of post-storage quality is made prior to storage:

    All fruit predicted as good + rest of the fruit predicted as poor = amount stored

    Thus,$\qquad\qquad\qquad$ 26 $\qquad\qquad$ + $\qquad$ 0 $\qquad\qquad$ = 26

    Number of fruit which will have good and poor post-storage quality:

    $[(52/58)26]_{good}$ $\qquad$ + $[(6/58)26]_{poor}$ $\qquad$ + $\qquad$ 0 $\qquad\qquad$ = 26

    23_{good} $\qquad\qquad\quad$ + 3_{poor} $\qquad\qquad\qquad$ + $\qquad$ 0 $\qquad\qquad$ = 26

    Amount with good post-storage quality = 23 (88%)

    Amount with poor post-storage quality = 3 (12%)

Thus, the amount of fruit lost due to poor post-storage quality:

Without the classification method: $\qquad\qquad$ 4 fruit (16%)

Using the MARS classification method: $\qquad$ 3 fruit (12%)

Total fruit saved using MARS classification: $\quad$ 1 fruit (4%)

References

Atkins, P.W. (1996). Molecular spectroscopy. In: *The elements of physical chemistry*, 2nd ed. Pp. 411-427. Oxford: Oxford University Press.

Bartley, I.M. & Knee, M. (1982). The chemistry of textural changes in fruit during storage. *Food Chemistry*, **9**, 47-58.

Ben-Arie, R. & Lavee, S. (1971). Pectic changes occurring in Elberta peaches suffering from woolly breakdown. *Phytochemistry*, **10**, 531-538.

Byrne, D.H. (2002). Peach breeding trends: A world wide perspective. In: *Proceedings of the 5th International Peach Symposium*, Volume 1. Pp. 49-59. July, 2001. Davis, CA.

Carlini, P., Massantini, R. & Mencarelli, F. (2000). Vis/NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. *Journal of Agricultural Food Chemistry*, **48**, 5236-5242.

Chahine, H., Gouble, B., Audergon, J.M., Souty, M. & Albagnac, G. (1999). Effect of ethylene on certain quality parameters of apricot fruit (*Prunus armeniaca*, L.) during maturation and postharvest evolution. In: *Proceedings of the XIth International Symposium on the Apricot Culture*, Volume II. Pp. 577-584. May, 1997. Verra-Makedonia, Greece.

Dull, G.G., Birth, G.S., Smittle, D.A. & Leffler, R.G. (1989). Near infrared analysis of soluble solids in intact cantaloupe. *Journal of Food Science*, **54**, 393-395.

Eksteen, J. & Combrink, C. (1994). *Manual for the identification of postharvest disorders of pome and stone fruit*. p. 7. South Africa: Stellenbosch Institute for Fruit Technology, Agricultural Research Council.

Esbensen, K.H. (2000). Multivariate Data Analysis in Practice: An Introduction to Multivariate Data Analysis and Experimental Design, 4th ed. Pp. 1-352. Norway: Camo ASA.

Fernandez-Trujillo, J.P., Cano, A. & Artés, F. (1998). Physiological changes in peaches related to chilling injury and ripening. *Postharvest Biology and Technology*, **13**, 109-119.

Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, **19**, 1-141.

Guthrie, J., Wedding, B. & Walsh, K. (1998). Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy*, **6**, 259-265.

Kawano,S., Watanabe, H. & Iwamoto, M. (1992). Determination of sugar contents in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. *Journal of the Japanese Society for Horticultural Science*, **61**, 445-451.

Kawano, S., Abe, H. & Iwamoto, M. (1995). Development of a calibration equation with temperature compensation for determining the brix value in intact peaches. *Journal of Near Infrared Spectroscopy*, **3**, 211-218.

Kusama, T., Abe, H., Kawano, S. & Iwamoto, M. (1997). Classification of normal and aged soyabean seeds by discriminant analysis using principal component scores of near infrared spectra. *Journal of Japanese Society of Food Science and Technology*, **44**, 569-578.

Lill, R.E., O'Donoghue, E.M. & King, G.A. (1989). Postharvest physiology of peaches and nectarines. *Horticulture Review*, **11**, 413-452.

McGlone, V.A. & Kawano, S. (1998). Firmness, dry matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology*, **13**, 131-141.

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

Moons, E., Dardenne, P., Dubois, A. & Sindic, M. (1997). Nondestructive visible and NIR spectroscopy for determination of internal quality in apple. In: *Proceedings from the Sensors for Nondestructive Testing International Conference and Tour*. Pp. 122-132. February, 1997. Orlando, Florida.

Murray, I. (1999). NIR spectroscopy of food: Simple things, subtle things and spectra. *NIR news*, 10(2), 10-12.

Næs, T. & Issaksson, T. (1991). SEP or RMSEP, which is best? *NIR news*, 2(4), 16.

Næs, T. & Martens, H. (1984). Multivariate calibration. II. Chemometric methods. *Trends in Analytical Chemistry*, **3**, 266-271.

Oritz, C., Barreiro, P., Correa, E., Riquelme, F. & Ruiz-Altisent, M. (2001). Non-destructive identification of woolly peaches using impact response and near infrared spectroscopy. *Journal of Agricultural Engineering Research*, **78**, 281-289.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis* (Edited by D. Browning), 2nd ed. Pp. 1-220. Harlow: Longman Scientific and Technical.

Peiris, K.H.S., Dull, G.G., Leffler, R.G. & Kays, S.J. (1999). Spatial variability of soluble solids or dry-matter content within individual fruits, bulbs, or tubers: implications for the development and use of NIR spectrometric techniques. *HortScience*, **34**, 114-118.

Peiris, K.H.S., Dull, G.G., Leffler, R.G. & Kays, S.J. (1998). Near infrared spectrometric method for nondestructive determination of soluble solids contents of peaches. *Journal of American Society of Horticulture Science*, **123**, 898-905.

Pretel, M.T., Serrano, M., Martinez, G. Riquelme, F. & Romojaro, F. (1993). Influence of films of different permeability on ethylene synthesis and ripening of MA-packaged apricots. *Lebensmittel-Wissenscaft & Technologie*, **26**, 8-13.

Ryall, M.S. & Pentzer, M.S. (1982). Handling, transportation and storage of fruits and vegetables. Pp. 495-498. AVI Publishing Company, Inc.

Sato, T. (1994). Application of principal component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *Journal of the American Oil Chemists' Society,* **71**, 293-298.

Schaare, P.N. & Fraser, D.G. (2000). Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidia chinensis*). *Postharvest Biology and Technology*, **20**, 175-184.

Schmilovitch, Z., Mizrach, A., Hoffman, A., Egozi, H. & Fuchs, Y. (2000). Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology*, **19**, 245-252.

Sjöström, M., Wold, S., Lindberg, W., Persson, J. & Martens, H. (1983). A multivariate calibration problem in analytical chemistry solved by partial least squares models in latent variables. *Analytica Chimica Acta*, **150**, 61-70.

Slaughter, D.C., Barrett, D. & Boersig, M. (1996). Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. *Journal of Food Science,* **61**, 695-697.

Slaughter, D.C. (1995). Nondestructive determination of internal quality in peaches and nectarines. *American Society of Agricultural Engineers*, **38**, 617-623.

Steyn, B. (2003). Tiger Brands: Fruit unit, Ashton. Personal communication.

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

Victor, W. (2002). Canned Fruit Producers' Association. Personal communication.

Williams P.C. & Norris, K.H. (1987). Qualitative applications of near-infrared reflectance spectroscopy. In: *Near Infrared Technology in the Agricultural and Food Industries* (Edited by P.C. Williams & K.H. Norris) American Association of Cereal Chemistry Inc.: St. Paul, MN, USA.

Wills, R.B.H., Mc Glasson, W.B., Graham, D., Lee, T.H. & Hall, E.G. (1989). Physiology and biochemistry of fruit and vegetables. In: *Postharvest: An Introduction to the Physiology and Handling of Fruits and Vegetables*. Pp. 34-37. Oxford: BSP Professional Books.

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognition*, **8**, 127-139.

Woo, Y., Cho, C., Kim, H., Yang, J. & Seong, K. (2002). Classification of cultivation area of ginseng by near infrared spectroscopy and ICP-AES. *Microchemical Journal*, **73**, 299-306.

Zude, M. & Herold, B. (2001). Optical spectral analysis potential to determine fruit ripeness. *Erwerbs-Obstbau*, **43**, 161-166.

Chapter 4

Prediction of post-storage soluble solids content and internal quality of intact clingstone peaches using near infrared spectroscopy (NIRS) with PLS, SIMCA and MARS as chemometric regression and classification techniques

Abstract

The South African peach canning industry suffers production losses owing to poor storage potential of clingstone peaches. The fruit needs to be cold stored owing to the short harvesting period and the limited capacity of the factory to process the large amount of fruit during the peak harvesting period. The possibility and feasibility of implementing Fourier transform near infrared spectroscopy (FT-NIRS) to predict storage potential of whole peach samples were investigated. Soluble solids content (SSC) and firmness are good indices of ripeness and possibly also of fruit quality. For this study, nine cultivars of clingstone peaches were collected from different peach-producing areas in the Western Cape Province of South Africa over 8-week periods during two consecutive seasons. Near infrared spectroscopy (NIRS) was used to develop quantitative calibration models for SSC and firmness of intact peaches. The spectral measurements were made on intact fruit after harvesting and the SSC and firmness measurements were made after three- (2002) and two-week (2003) periods of cold storage. Correlation coefficients (r) between actual and predicted values were 0.77 (2002), 0.80 (2003) and 0.76 (combined seasons) and SEP-values were 1.60 °Brix (2002), 1.60 °Brix (2003) and 1.68 °Brix (combined seasons), using partial least squares (PLS) regression. An alternative regression method, multivariate adaptive regression splines (MARS), resulted in r = 0.82 (2002) and 0.77 (2003) and SEP = 1.42 °Brix (2002) and 1.55 °Brix (2003). PLS validations within cultivars obtained r-values ranging from 0.68 to 0.83 and SEP-values ranging from 0.99 to 1.59 °Brix. Firmness predictions were less successful. PLS obtained r-values ranging from 0.40 to 0.52 and SEP-values from 1.41 kg (13.83 N) to 1.53 kg (15.00 N) for the 2002, 2003 and combined seasons, respectively. MARS obtained validation results of r = 0.66 and SEP = 1.24 kg (12.16 N) for 2002 and r = 0.52 and SEP = 1.33 kg (13.05 N) for 2003. A Diode Array spectrophotometer gave slightly better results for measuring SSC (r = 0.85 and SEP = 1.35 °Brix) and firmness (r = 0.52 and SEP = 1.41 kg) than a Fourier Transform spectrophotometer in a comparative study over the 2002 season. The regression results confirmed sufficient near infrared (NIR) light penetration into the intact

fruit and provided information for further studies on the possibility of NIRS as means to predict post-storage quality of intact clingstone peaches. Poor storage quality was subjectively identified by softening, loosening of the skin and adhesion of the flesh to the stone just after destoning, after a three (2002 season) and two-week (2003 season) cold storage period. Two approaches to predict storage potential were investigated. An indirect approach involved the correlation (using regression) between near infrared spectra and objective measurements (firmness and soluble solids content), combined with the correlation (using classification trees) between the objective measurements (mass, size, soluble solids content, firmness and cultivar) and subjective quality evaluations of post-storage quality. Fruit was classified into two post-storage quality classes according to fruit size and firmness. This indirect method could be used for classification of post-storage quality in intact peaches, provided a more accurate NIRS prediction model for firmness can be obtained. A more direct approach was to classify peach samples into two storage potential models, created using NIR spectra and subjective post-storage quality evaluations. For this purpose, two different chemometric methods, soft independent modeling by class analogy (SIMCA) and MARS were applied. SIMCA obtained overall correct classification rates of 53%, 60% and 53% for the 2002, 2003 and combined seasons, respectively. Classification results within cultivars were poorer than classification results within seasons. MARS obtained classification rates of 57% to 65% within each of the two seasons. Results from a MARS model was used to determine the feasibility of the implementation of NIRS in the peach canning industry. A total saving of 3%, amounting to R610 740 was calculated. In order to make NIRS more feasible for the South African peach canning industry, more research is required to improve this classification technique.

Introduction

Approximately 130 000 tons of clingstone peaches are received from producers for canning annually in South Africa (Victor, 2002). A large portion (67%) of the fruit is cold stored for one to two weeks owing to a short harvesting period and the limited capacity of the factory to process all the fruit immediately. The industry, however, suffers production losses of 6500 tons of peaches (R11,05 million) annually owing to internal breakdown during cold storage (Steyn, 2003).

Upon arrival at the factory, the peaches are graded and each producer compensated for his fruit according to the quality of a small sample taken for grading (Steyn, 2003). The quality for canning is determined by the fruit size, colour, ripeness and bruising. The fruit

diameter should be at least 57 mm for canning grade 1 (best quality) and 55 mm for canning grade 2. The colour of the fruit is compared to a standard colour chart and should be light yellow to deep yellow. Ripeness is not measured, but is evaluated by a subjective evaluation of the stage of fruit development. The fruit should be at the right stage of development upon arrival at the factory to ensure a firm and ripe fruit after canning. Light bruising is acceptable for canning grade 1 and medium bruising acceptable for canning grade 2. Fruit firmness is not measured during grading, but noticeably soft and disintegrated fruit is rejected for canning.

Internal breakdown in peaches is often referred to as wooliness or mealiness (Morris, 1982). Breakdown is associated with an increase in insoluble, high molecular weight pectin, which binds water in a pectate gel (Ben-Arie & Lavee, 1971). The retention properties of Ca^{2+} possibly enhance this condition (Dawson et al., 1993). Indices of internal breakdown include softening, loosening of the skin and adhesion of the flesh to the stone (leaving a large stone cavity). The identification of internal breakdown is based on subjective evaluations by experienced workers after the fruit is destoned. The canning industry requires an objective and non-destructive method to identify whether or not peaches will develop internal breakdown during cold storage. Decisions, whether peaches can be cold stored or should be canned immediately, will be possible and will subsequently lead to decreased production losses (Victor, 2002).

Near infrared spectroscopy (NIRS) can provide rapid and non-destructive quantitative measurements of internal properties in peaches, such as soluble solids content (SSC) (Kawano et al., 1992, 1995; Slaughter, 1995; Peiris et al., 1998; Van Zyl, 2000), titratable acidity, firmness (Ortiz et al., 2001), sweetness index, chlorophyll, sucrose, and sorbitol content (Slaughter, 1995). Table 1 shows the conditions and validation results obtained from previous studies for predicting SSC in peach samples. Soluble solids content within an individual peach can vary up to 4%. Thus a measurement on a single spot on the fruit may vary substantially from the average measurement of the entire fruit, regardless of how accurate the method is to determine SSC (Slaughter, 1995).

Season and cultivar variability in SSC was responsible for a major amount of the spectral variability in apple cultivars (Peirs et al., 2003). The accuracy of the NIRS measurements increased when greater variability in the calibration set was included and overfitting was avoided. However, it was suggested that cultivar specific models be used if possible.

A credible theory does not exist on which to base NIRS models for predicting firmness in fruit. The normal hand-held penetrometer is also renowned for being imprecise, thus limiting

Table 1. Validation results obtained in previous studies for measuring SSC in peach samples using NIR spectroscopy.

NIR instrument used	Wavelength range (nm)	Peach sample set	n (validation)	Validation statistics		Reference
				r	SEP	
Spectrum Identicheck™ 2.0 FT-NIR System (Perkin Elmer)	1000-2800	Nine cultivars	23	0.96	0.55 °Brix	Van Zyl, 2000
NIRSystems Model 6500, fiber optic probe (interactance)	400-1100	Four cultivars	41	0.90	1.00%	Slaughter, 1995
Acusto-optic spectrometer (transmittance)	800-1050	Four individual cultivars	50, 50, 50, 200	0.51-0.94	0.49-1.40%	Peiris et al., 1998
Acusto-optic spectrometer (transmittance)	800-1050	One cultivar	50	0.80-0.84	0.66-1.94%	Peiris et al., 1997
Pacific Scientific Mode 6250 (interactance)	800-1650	One cultivar	100	0.96	0.42 °Brix	Kawano et al., 1995
Pacific Scientific Mode 6250 (interactance)	800-1650	One cultivar	114	0.97	0.50 °Brix	Kawano et al., 1992

the accuracy of any prediction model based on visible and near infrared (Vis/NIR) spectra (McGlone *et al.*, 2002). NIRS predictions of firmness in fruit were reported for kiwifruit (r = 0.66 and RMSEP = 7.8 N (McGlone & Kawano, 1998), apples (r = 0.74 and RMSEP = 7.5 N) (McGlone *et al.*, 2002), (r = 0.51-0.60 and SEP = 7.1-9.5 N) (Lu *et al.*, 2000) and mango (r = 0.82 and SEP = 17.4 N) (Schmilovitch *et al.*, 2000). Much work remains to be done on understanding the physical or chemical basis for measuring fruit firmness using NIRS (Lu *et al.*, 2000).

Near infrared (NIR) spectra contain information which can only be extracted by mathematical processing of the multivariate data. Various sophisticated chemometric techniques have been developed which take into account both the variability in the spectra and spectral patterns (Osborne *et al.*, 1993).

Oritz *et al.* (2001) used two non-destructive methods to identify woolly peaches, with both methods obtaining 80% correct classification. NIRS was applied to classify the fruit into categories of juiciness, while impact response classified the fruit according to texture categories.

Van Zyl (2000) showed the potential of NIRS to classify South African canning peaches according to post-storage quality. SIMCA was applied to FT-NIR spectra (1000-2500 nm) to classify peach samples (six cultivars combined), with subjective post-storage quality evaluations as reference values. Recognition rates of 72-100% and rejection rates of 4-39% were obtained on the validation sets (n = 23-49) over six cultivars and within eight weeks of harvest of the 2000-season.

This study is an extension of the research by Van Zyl (2000). A larger sample set of South African clingstone peaches from two consecutive seasons were used for SSC and firmness regressions. Classification according to the storage potential of canning peaches was investigated more in-depth. The performance of the classification models was investigated when cultivar, weekly and seasonal variation were included. Additional chemometric methods of regression and classification were applied to obtain the best possible regression and classification models. It was also attempted to identify a correlation between the post-storage quality and an objectively measurable fruit characteristic. The FT-NIR absorption spectra were measured at 645-1201 nm to include some wavelengths from the visible range as well.

Objectives

The first objective of this study was to confirm sufficient NIR light penetration into intact peaches by NIRS measurements of SSC in intact clingstone peaches. The main objective of this study was to determine the post-storage quality of fresh clingstone peaches based on their NIR spectra and subjective post-storage quality evaluations (presence of internal breakdown, loosening of the skin and adhesion of the flesh to the stone). The first approach was to find an indirect correlation by including several other factors, such as cultivar, mass, size, SSC and firmness, to predict post-storage quality. This approach would involve a combination of a regression based on NIR spectra and the reference values and classification trees for predicting post-storage quality.

The second approach was to find a direct correlation between NIR spectra and post-storage quality. Two chemometric classification methods, soft independent modeling by class analogy (SIMCA) and multivariate adaptive regression splines (MARS) were investigated for this purpose. A final objective was to suggest the most appropriate method to predict the post-storage quality in peaches and to determine the feasibility of implementing an NIRS instrument as a means to predict post-storage quality in the peach canning industry.

Materials and methods

Samples

Peaches from different production areas (Ceres, Worcester, Robertson, Ashton) in the Western Cape Province of South Africa were collected for analyses during the 2002 and 2003 harvesting seasons. The fruit was packed onto moulded cardboard trays (to separate the fruit) in cartons and transported to the ARC Infruitec-Nietvoorbij premises in Stellenbosch. For each season, peaches were collected over an eight week period (09 January - 27 February 2002 and 15 January – 5 March 2003). During the 2002 and 2003 seasons, 12 and 10 cartons, respectively, containing 20 peaches each were received. The cultivars included Goudmyn, Keisie, Neethling, Professor Black, Sandvliet, Western Sun, Woltemade and Kakamas. Table 2 summarises the total number of samples, including cultivar range, received during the two seasons. The fruit was numbered individually (1, 2, 3, ..., n) and on both halves (a or b) with a soft permanent marker pen for identification after storage. The fruit were marked on the areas that would not be used for any of the measurements. NIR spectral measurements were performed on arrival at room temperature (21°C). The peaches were not washed or treated with chemicals prior to analyses. They were also not treated with antimicrobials, since this application would increase the risk of damage to the fruit. After NIR

93

Table 2. Summary of peach samples received during 2002 and 2003 seasons

Week of harvest	Cultivar	Number for 2002 season		Number for 2003 season	
		Cartons	Samples	Cartons	Samples
1	Goudmyn	2	40	0	0
	Keisie	1	20	4	80
	Neethling	3	60	2	40
	Sandvliet	4	80	4	80
	Western Sun	2	40	0	0
2	Keisie	6	120	2	40
	Neethling	5	100	8	160
	Western Sun	1	20	0	0
3	Goudmyn	1	20	0	0
	Keisie	2	40	3	60
	Neethling	6	120	5	100
	Professor Black	1	20	1	20
	Woltemade	2	40	1	20
4	Keisie	1	20	1	20
	Neethling	5	100	5	100
	Sandvliet	0	0	1	20
	Woltemade	6	120	3	60
5	Neethling	1	20	0	0
	Sandvliet	1	20	0	0
	Woltemade	5	100	2	40
	Kakamas	5	100	8	160
6	Woltemade	0	0	1	20
	Kakamas	12	240	9	180
7	Kakamas	12	240	10	200
8	Kakamas	12	240	10	200
Total	Goudmyn	3	60	0	0
Total	Keisie	10	200	10	200
Total	Neethling	20	400	20	400
Total	Sandvliet	5	100	5	100
Total	Western Sun	3	60	0	0
Total	Professor Black	1	20	1	20
Total	Woltemade	13	260	7	140
Total	Kakamas	41	820	37	740
Total	All cultivars	96	1920	80	1600

94

spectral measurements the fruit was subjected to cold storage at -0.5ºC. In South Africa, the cold storage period is approximately one to two weeks, depending on the quantity of fruit harvested for canning. For this study it was decided to store the fruit for a maximum period of time. The reason was to ensure that there was enough fruit with good and poor quality after storage to develop representative models for storage potential and to identify clear differences between fruit with good and poor post-storage quality. The samples of the 2002 season were cold stored for a period of three weeks. Too many samples had poor quality after three weeks and therefore the samples of the 2003 season were cold stored for a period of two weeks.

Spectral acquisition

The optical absorption spectrum was measured on each half (180° apart) of each intact fruit using a Perkin Elmer Spectrum Identicheck[TM] 2.0 FT-NIR system in reflectance mode. Each fruit was positioned on the Identicheck Reflectance Accessory (ICRA), where light interacted with the fruit tissue (Figure 1). Every absorption spectrum was the average of 16 individual optical scans in the wavelength range of 685 to 1191 nm at 2 nm intervals and a resolution of 64 cm[-1]. NIR spectra of the samples from the 2002 season were also measured with a Perten Diode Array 7000 Flexi-mode NIR/Visible Spectrophotometer. For the DA 7000 flexi-mode system a wavelength range of 425 to 1690 nm was used and the samples were illuminated from the top (down-view mode).

Sample evaluation

Each week, after the peaches were cold stored for the required number of weeks at -0.5ºC, they were analysed for the quantitative determinations of fruit diameter, mass, firmness and SSC. The fruit diameter was measured with a Digimatic calliper. A Mettler PE 3600 scale was used to measure the mass of each fruit. Fruit firmness, expressed as a pressure reading in kg, was measured with a Penetrometer Mod FT 444 fitted with a 11 mm tip and mounted on a modified drill stand. The SSC, expressed in ºBrix, was determined with an Atago N1 hand refractometer. A drop of juice that was left on the penetrometer tip after pressure measurements was placed directly on the refractometer. The tip of the penetrometer and the sampling area of the refractometer were cleaned with tissue paper after each measurement. Firmness and SSC readings were taken on both halves of each fruit (at the same place where each fruit was scanned optically). SSC and firmness measurements were performed by the same operator throughout this study. The fruit was destoned using a Filper pitter and the mass determined again for each fruit.

95

Figure 1. Peach sample placed on the ICRA of the Perkin Elmer Spectrum Identicheck™ 2.0 FT-NIR system for spectral acquisition.

Each fruit half was evaluated subjectively for post-storage quality. These evaluations were performed by the same researcher, specialising in post harvest defects of fruit. The following quality characteristics were identified: adhesion of flesh to the stone after destoning (leaving a large stone cavity); loosening of the skin and softening. A score was allocated to each fruit half according to all three characteristics. A fruit half which had any one of these defects, was reported as having poor storage potential and allocated a value of one. Where the fruit half had none of the defects, it was reported as having good storage potential and allocated a value of zero.

Data processing

The spectra were subjected to pre-processing and spectral evaluation. A master data file was made for each season by merging all spectral and reference data (quantitative measurements and subjective storage potential data) collected throughout each season. Regressions and classifications were performed on the data within seasons. Subsets of each master file were made for the data of each cultivar in order to perform regressions and classifications for each cultivar separately. Due to internal variability within individual peach samples, it was decided to also perform each regression and classification by only using data from only one half of the fruit (either the a- or b-half). Finally, all the data sets of the two seasons (2002 and 2003) were combined to evaluate the performance of the regression and classification models when seasonal variability is included.

Pre-processing of spectral data

The spectral data sets were subjected to mathematical pre-processing in order to resolve overlapping bands and to investigate absorption peaks and the spectral differences between low and high SSC. Multiplicative scatter correction (MSC) and first derivatives (Savitzky-Golay) (described in Chapter 3) were applied to the spectra.

Spectral evaluation

The absorption peaks in the NIR spectra of intact peaches were evaluated. The spectral patterns between peaches with low (9 °Brix) and high (16 °Brix) SSC contents were compared.

Development of regression models

Partial least squares (PLS) regression

Partial least squares (PLS) regression (Sjöström *et al.*, 1983; Næs & Martens, 1984) (described in Chapter 3) was performed to develop SSC and firmness regressions. Spectral data were merged with the respective SSC and firmness data using The Unscrambler™ program, version 6.11 (CAMO ASA, Norway).

Calibration was performed on approximately two-thirds of each sample set. The calibration samples were chosen to cover the entire range of the reference data and included the highest and lowest reference value. Independent validation was performed on each data set. The validation set included the other third of the data set (thus samples not used in calibration) and also covered the entire range of the reference data.

Multivariate adaptive regression splines (MARS)

Multivariate adaptive regression splines (MARS) (Friedman, 1991) (described in Chapter 3) was performed as an alternative method to develop SSC and firmness regressions. The MARS™ program, version 2 (Salford Systems, California, USA) was used to develop MARS regressions. A MARS model was developed for each season using a calibration set which consisted of 80% of the total samples in each season. The model was then tested on the validation set, which consisted of 20% (the samples not used in calibration) of the total samples in each season. The samples used for calibration and validation were randomly chosen. Regression results, in terms of r, SEC, RMSEP, SEP and bias, were calculated for each PLS and MARS regression as described in Chapter 3.

Classification of post-storage quality

Two approaches were followed to classify samples according to storage potential. Figure 2 describes the approaches proposed for classification. The first approach was to find a correlation between all objective measurements (mass, diameter, SSC, firmness and cultivar) and the concluding post-storage quality data (subjective quality evaluations). A non-destructive classification for post-storage quality would then be possible by a correlation between NIR spectra and the objective measurements. The second approach was to simply find a correlation between the NIR spectra and the concluding subjective post-storage quality data.
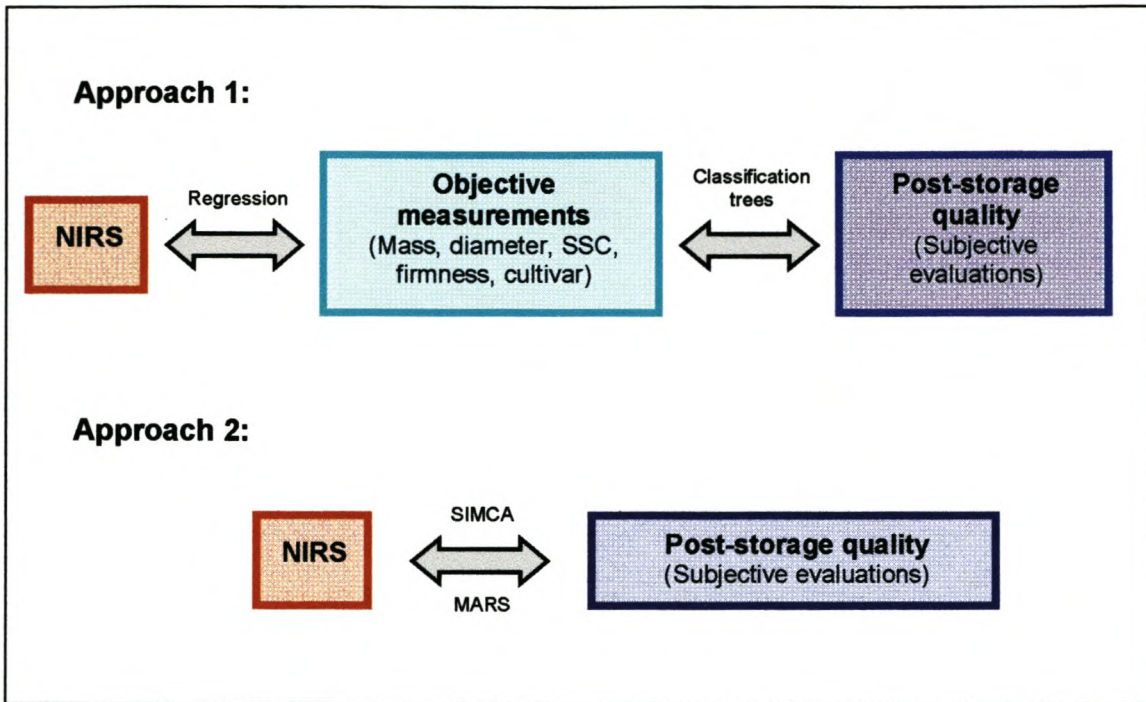
98

Figure 2. The two approaches followed for classification of storage potential of peach samples.

Approach 1: Classification trees and regression

Classification trees

Classification trees were used to find a correlation between all objective measurements and the post-storage quality. In the case of classification trees the dependent (response) variable y is a discrete variable consisting of 2 or more classes (e.g. yes/no, present/not present, good/poor post-storage quality). The concept of entropy (chaos) is used as basis for constructing classification trees. To explain entropy, consider a response variable with 2 classes namely yes/no. If the samples in a data set consist of 50% yes and 50% no responses, then the entropy of that data set is a maximum because the chance of correctly predicting the class of the response variable is 50%. As the proportion of one of the classes approaches 100%, the lower the entropy becomes, and it reaches a minimum when a data set consists of 100% of one class. In this case there is a 100% chance of correctly predicting the class of the response variable. The aim of a classification tree is to divide the data set into subsets so that the subsets have a lower entropy than the full data set. Thus it strives to group the classes together into the best possible subsets based on the independent or predictor variables. This is achieved as follows:

The method selects a point x_p between the minimum and maximum of x, e.g. SSC data, that splits the data into two sets (or nodes in a tree). All instances for which $x \leq x_p$ goes to the left node and all instances where $x > x_p$ goes to the right node. The point where the split is made is the point that decreases the entropy from the parent node to the child nodes the most. The procedure is then repeated for each of the two nodes.

Thus, a binary split is made on each node using the criteria mentioned above. Stopping rules are used to decide when the splitting process should stop. For example a minimum number of cases per node can be specified, and if that minimum number is reached, the node will split no further.

A sample subset of the peach data from the 2003 season was made to provide a calibration set for classification tree analysis. This calibration set consisted of 80% of the total number of samples. The validation set comprised of the other 20% (the samples not used in calibration) of the samples. The samples used for calibration and validation were randomly chosen. Only data from one half (b-half) of each fruit were used in order to avoid variability within each fruit. The predictor variables included the objective measurements of mass, diameter, SSC, firmness as well as cultivar, while the subjective storage potential data was used as response variables. The CART™ program (Salford Systems, California, USA) was used to obtain

classification trees, which grouped the response data as best possible based on the predictor variables.

Approach 2: SIMCA and MARS
Principal component analysis (PCA)

Before classification, the spectral variation of the data sets were analysed by PCA (described in Chapter 3). PCA was performed on the spectral data of each cultivar measured within one week of harvest; on the spectra of each cultivar measured within seasons; and on the spectra of all cultivars measured within seasons. PCA was performed to obtain an overview of the spectral data and to seek possible groupings of the data relating to the storage potential.

Soft independent modelling by class analogy (SIMCA)

Soft independent modeling by class analogy (SIMCA) (Wold, 1976) (described in Chapter 3) was performed using The Unscrambler™ program, version 6.11 (CAMO ASA, Norway). SIMCA was performed on the various data sets by creating a calibration set (containing approximately two-thirds of each sample set) and a validation set (containing the other third of each sample set) for each data set. The calibration set was used to calculate two separate PCA models, a model indicating good post-storage quality and a model indicating poor post-storage quality. These models were based on results of the subjective evaluations. For each PCA model, the optimum number of principal components (PC's) was determined. Finally, the samples from the validation set were evaluated by these models according to their analogy to each model.

Multivariate adaptive regression splines (MARS)

Multivariate adaptive regression splines (MARS) (Friedman, 1991) (described in Chapter 3) was performed on the data as an alternative classification method. The MARS™ program, version 2 (Salford Systems, California, USA) was used for MARS classifications. MARS was applied on the samples within each season. Separate models were made for each half of the fruit. The calibration set consisted of 80% and the validation set the other 20% of each data set. The samples used for the calibration and validation sets were randomly chosen. The subjective response variables were coded as 0's (good post-storage quality) and 1's (poor post-storage quality). The MARS model was fitted and estimated values for post-storage quality lay between zero and one. A threshold of 0.5 was chosen for classification purposes.

Outliers

The final regression and classification models were obtained after eliminating the outlier data points for each regression and classification model. Outliers were eliminated as described in Chapter 3.

Feasibility study

The feasibility of implementing NIRS to predict post-storage quality in canning peaches was evaluated. The classification result obtained using the MARS method was applied. The losses due to poor post-storage quality were calculated for the two scenarios:

1. No prior prediction of post-storage quality made (current method)
2. Prediction (based on NIR spectra and MARS) of post-storage quality made prior to storage (proposed method)

The feasibility of the classification method was calculated if 67% (the current situation in the industry) of the peaches have to be stored upon arrival at the depot.

Results and discussion

Descriptive statistics of SSC and firmness measurements

Overall descriptive statistics of SSC and firmness measurements, including mean, minimum, maximum and standard deviations are given in Table 3. The difference between measurements made on the a- and b-half of each fruit was calculated to determine the variation within the fruit. Due to the large variations within the fruit, the two measurements made on each fruit were not averaged, but kept as separate observations for further statistical analysis. The standard error of laboratory (SEL) was therefore not calculated. Box plots (Figures 3-6) show the distribution of measurements for all the reference data.

NIR spectra of the intact peach

The peaks in the spectra represent frequencies at which light is absorbed by the fruit. Figure 7 illustrates the original spectra (645-1185 nm). The spectra did not show many distinct wavelengths that could be used for quantitative measurements. High absorption peaks fall in the visible region (657 & 679 nm). The spectrum of a fruit half with a low SSC had higher peaks at 657 and 679 nm. These peaks could be due to higher chlorophyll content.

Table 3. Descriptive statistics for SSC (°Brix) and firmness (kg) measurements for the 2002 and 2003 seasons.

Season	Variable	n	Mean	Min.	Max.	Std. Dev.
2002	SSC	3664	13.14	2.6	22.4	2.56
	SSC difference[1]	1832	0.88	0.0	12.2	0.95
	Firmness	3238	5.28	2.0	18.0	1.75
	Firmness difference	1607	1.00	0.0	13.4	1.03
2003	SSC	3010	14.02	1.8	24.0	2.61
	SSC difference	1506	0.82	0.0	8.6	0.80
	Firmness	3010	5.82	2.0	14.0	1.60
	Firmness difference	1506	0.90	0.0	6.6	0.78

[1]The difference between the measurements of the two halves of the fruit
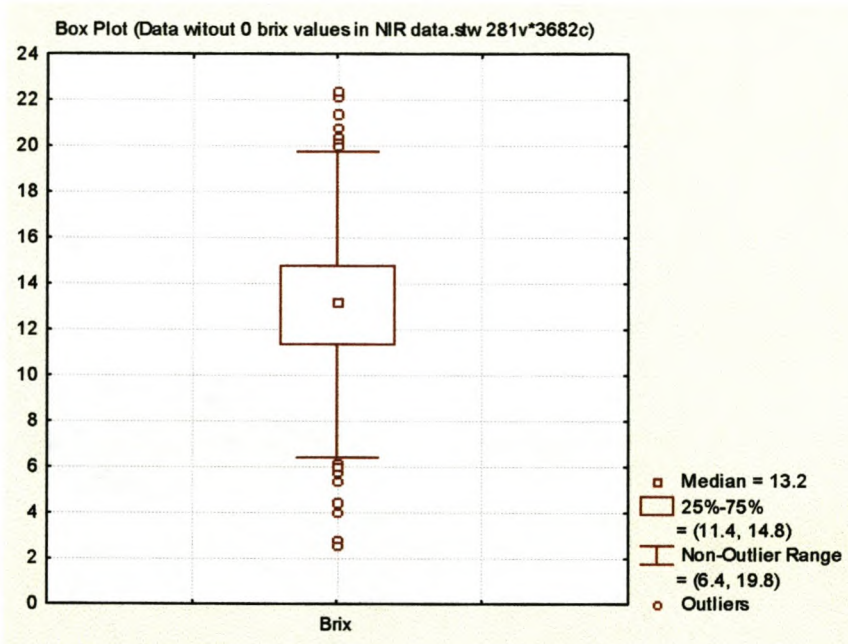

Figure 3. Box plot showing the distribution of SSC (°Brix) measurements of the 2002 season.
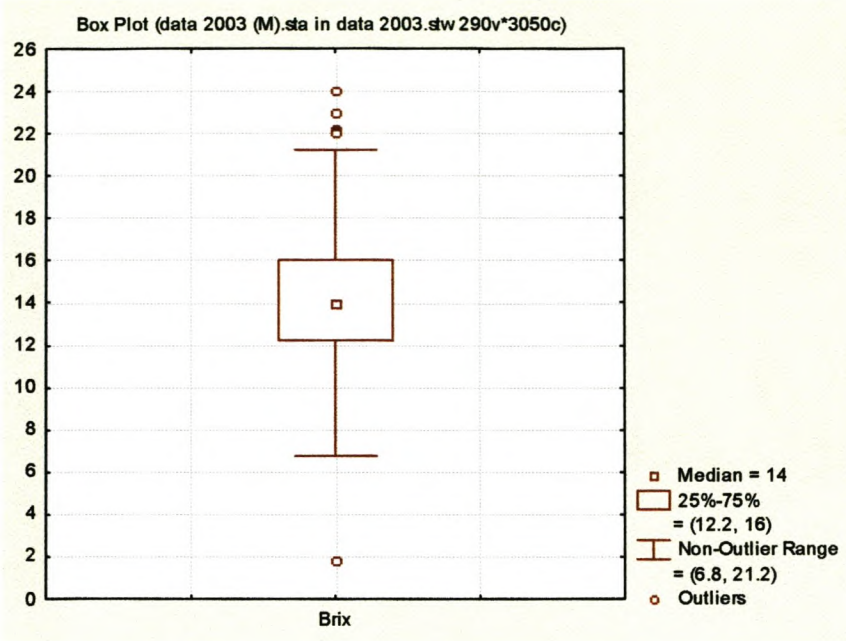
Figure 4. Box plot showing the distribution of SSC (°Brix) measurements of the 2003 season.
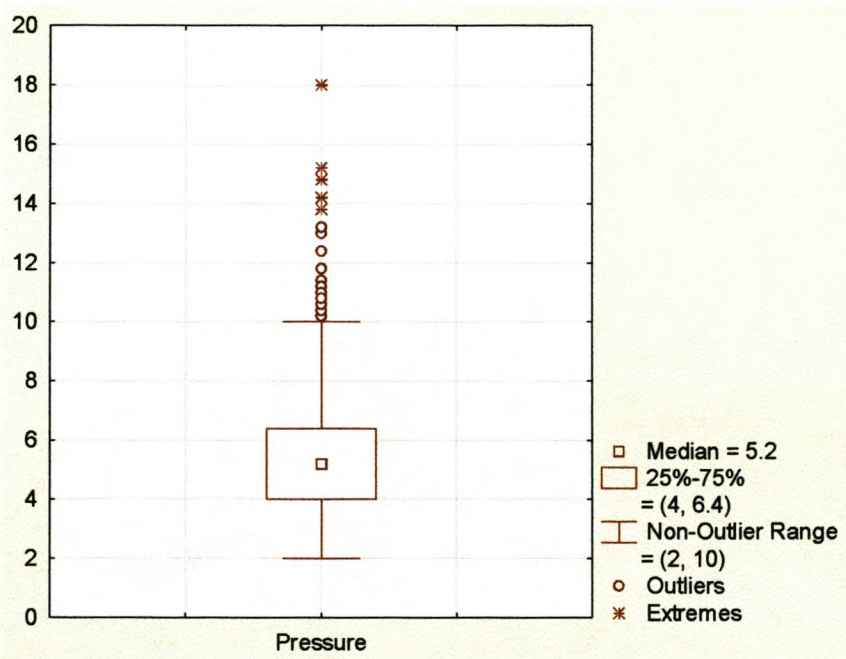

Figure 5. Box plot showing the distribution of firmness (kg) measurements of the 2002 season.
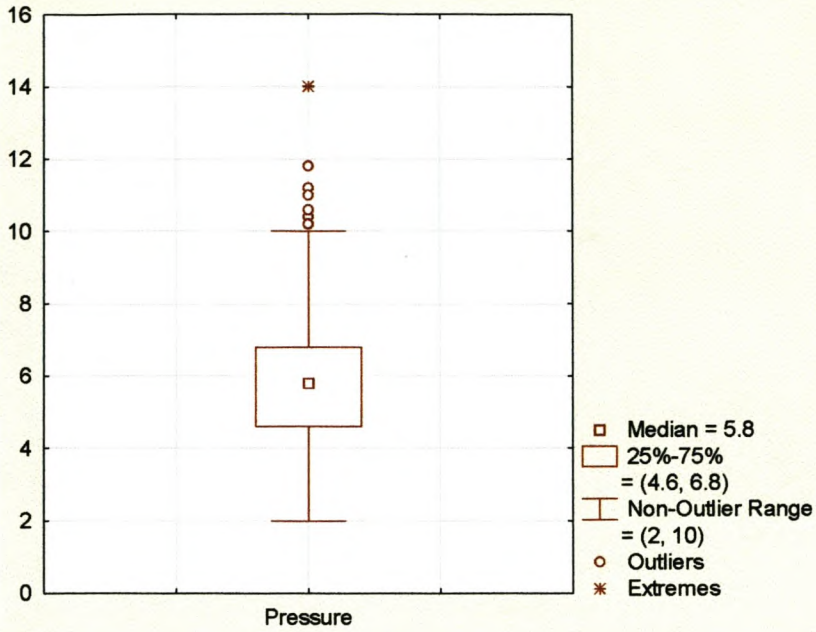
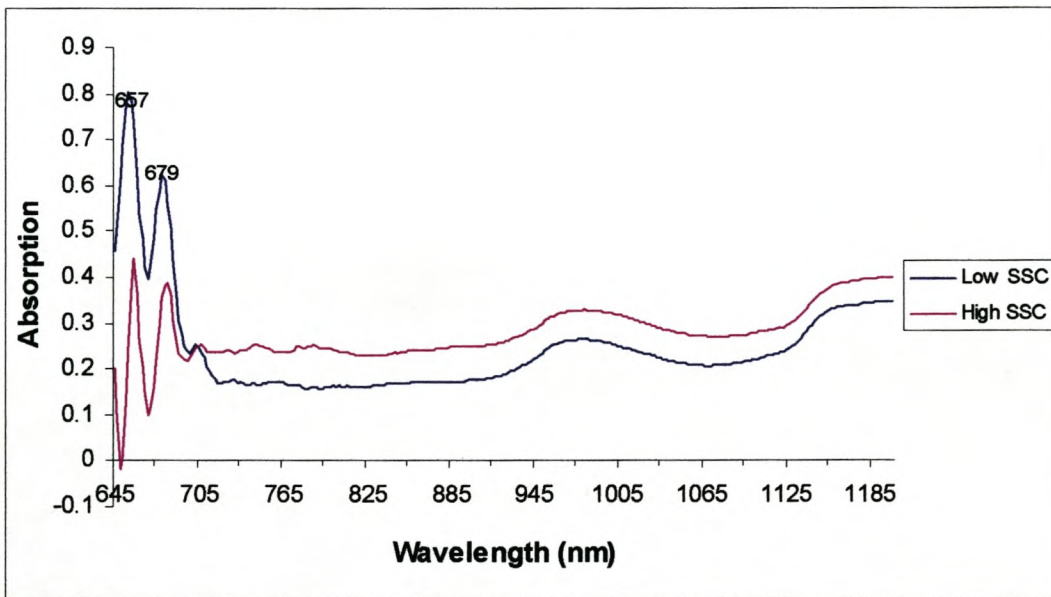**Figure 6**. Box plot showing the distribution of firmness (kg) measurements of the 2003 season.

Figure 7. NIR spectra (no pre-processing) of peach samples with a high and low SSC, respectively.

Chlorophyll was assigned to 680 nm (McGlone & Kawano, 1998). An immature fruit has more chlorophyll. As fruit mature, a decrease in chlorophyll and an increase in carotenoids (red and yellow pigments) take place (Wills *et al.*, 1989). The differences in absorption peaks show a correlation between chlorophyll and SSC, respectively, which also relates to fruit maturity. Moons *et al.* (1998) suggested that some of the visible wavelengths should be included in the calibration, as the SSC and firmness may be related to the fruit colour.

Spectra treated with MSC and first derivative (Figure 8) showed clear peaks at 741, 773, 841, 953 and 1139 nm. Peaks at 841 nm, 953 nm and 1139 nm correlate with the results of Kawano *et al.* (1995) who identified peaks for water at 841 nm, 966 nm and 942 nm and SSC at 904 nm, 951 nm and 1141 nm. The peaks at 741 and 773 nm could be CH_2 and OH overtones (Osborne *et al.*, 1993), indicating sugars (SSC) and moisture, respectively. The two spectra showed the same peaks, but separated the low and high SSC spectra more clearly.

Regression

Soluble solids content

Partial least squares regression results of the SSC calibrations and validations within seasons and for combined seasons are shown in Table 4. Figures 9-11 show the validation plots obtained within seasons. Correlation coefficients of 0.77 for 2002 (MSC-corrected spectra), 0.80 for 2003 (no pre-processing of spectra) and 0.76 for the combined seasons (no pre-processing of spectra) and SEP of 1.60 °Brix for 2002, 1.51 °Brix for 2003 and 1.68 °Brix for the combined seasons were obtained using the spectra from the Fourier Transform spectrophotometer. Regression results over the combined seasons gave slightly poorer results than the results obtained within seasons. The storage time between the spectral and reference measurements of the seasons differed (two and three weeks for the 2003 and 2002 seasons, respectively), possibly leading to a less accurate regression model.

MARS models were fitted to uncorrected spectra in conjunction with the first derivative spectral data. The MARS regression results (Table 5) for these data sets were similar to the PLS results, obtaining r-values of 0.82 and 0.77 and SEP-values of 1.42 °Brix and 1.55 °Brix for the 2002 and 2003 seasons, respectively. MARS can deal with non-linear relationships in data (Friedman, 1991). The similar regression results are probably owed to the linear relationship between the spectra and the SSC values. PLS regression performed on spectral data (MSC-corrected) from the Diode Array spectrophotometer gave a r-value of 0.85 and SEP of 1.35 °Brix.
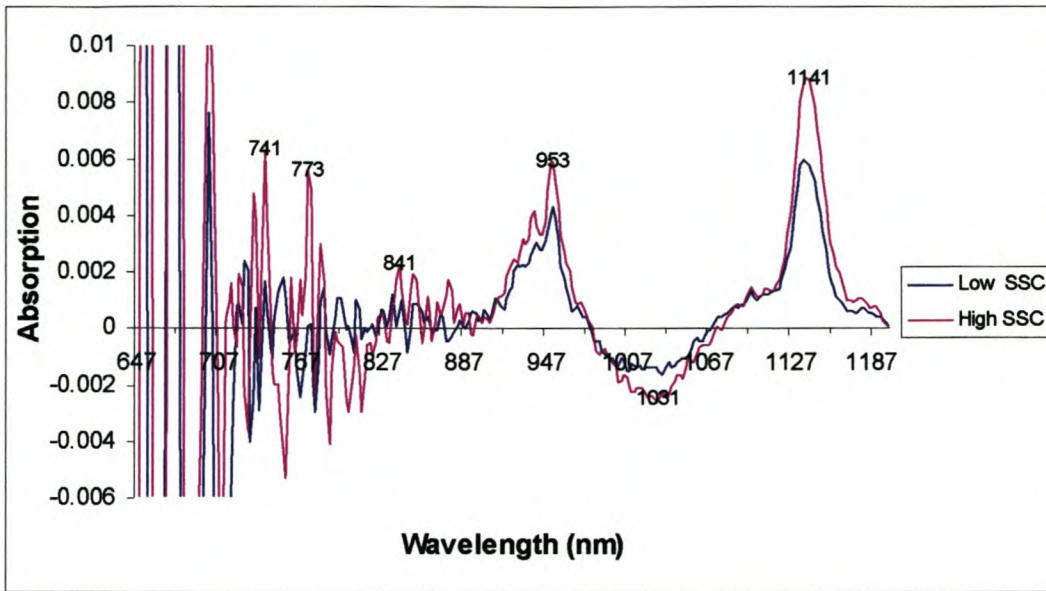
106

Figure 8. Pre-processed NIR spectra of peach samples with high and low SSC. MSC and first derivatives were applied to the spectra.

Table 4. PLS regression statistics for peach SSC obtained for the data within seasons and over the combined seasons.

Statistic	2002[1]	2003[1]	Combined seasons[1]	2002[2]
Calibration				
No. of PC's	17	15	23	25
n	2371	2089	4517	2583
r	0.78	0.81	0.74	0.85
SEC (°Brix)	1.59	1.52	1.73	1.34
RMSEC (°Brix)	1.59	1.52	1.73	1.34
Bias (°Brix)	0.002	-0.004	0.01	-0.004
Validation				
n	1292	1034	2279	1095
r	0.77	0.80	0.76	0.85
SEP (°Brix)	1.60	1.51	1.68	1.35
RMSEP (°Brix)	1.60	1.51	1.68	1.35
Bias (°Brix)	-0.01	0.01	-0.04	-0.02

[1]Fourier Transform spectrophotometer

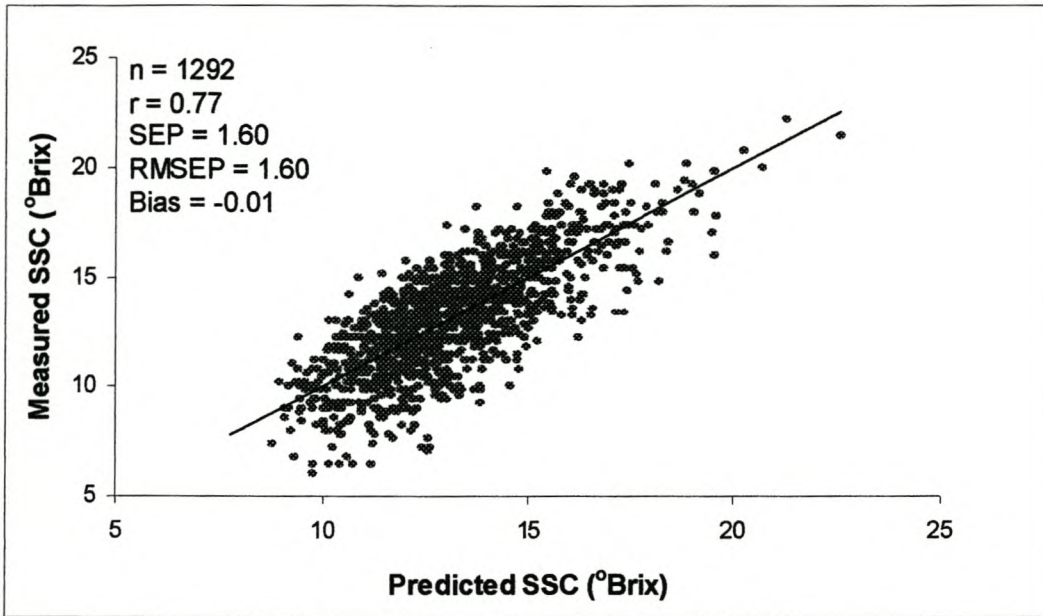[2]Diode Array spectrophotometer

107

Figure 9. Validation plot of measured SSC (°Brix) vs. predicted SSC for samples of the 2002 season (MSC corrected spectra, 17 PC's) using FT-NIRS.
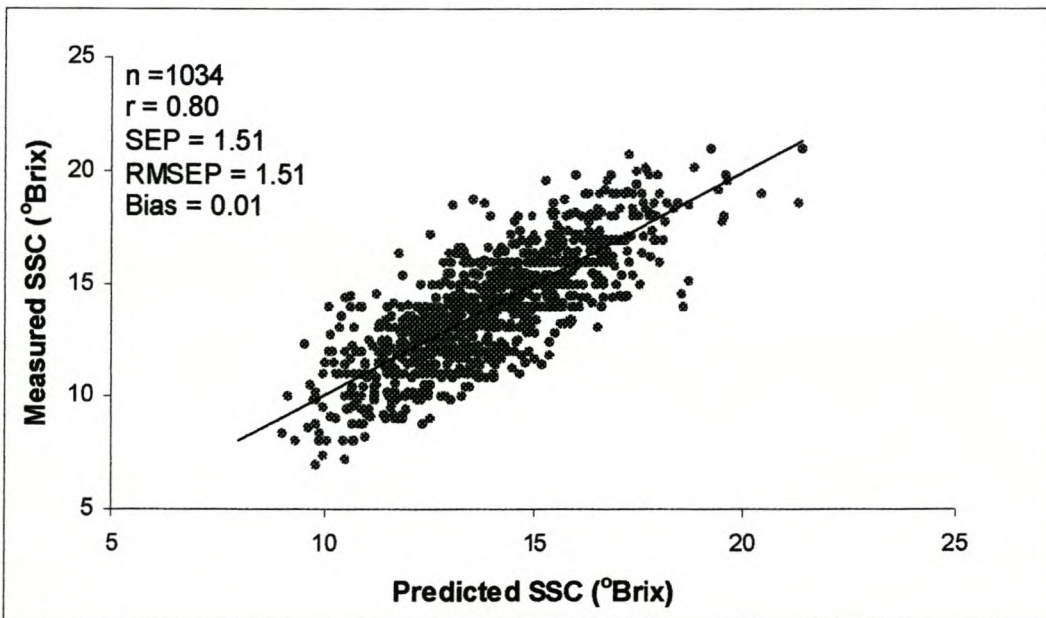

Figure 10. Validation plot of measured SSC (°Brix) vs. predicted SSC for samples of the 2003 season (no pre-processing, 17 PC's) using FT-NIRS.
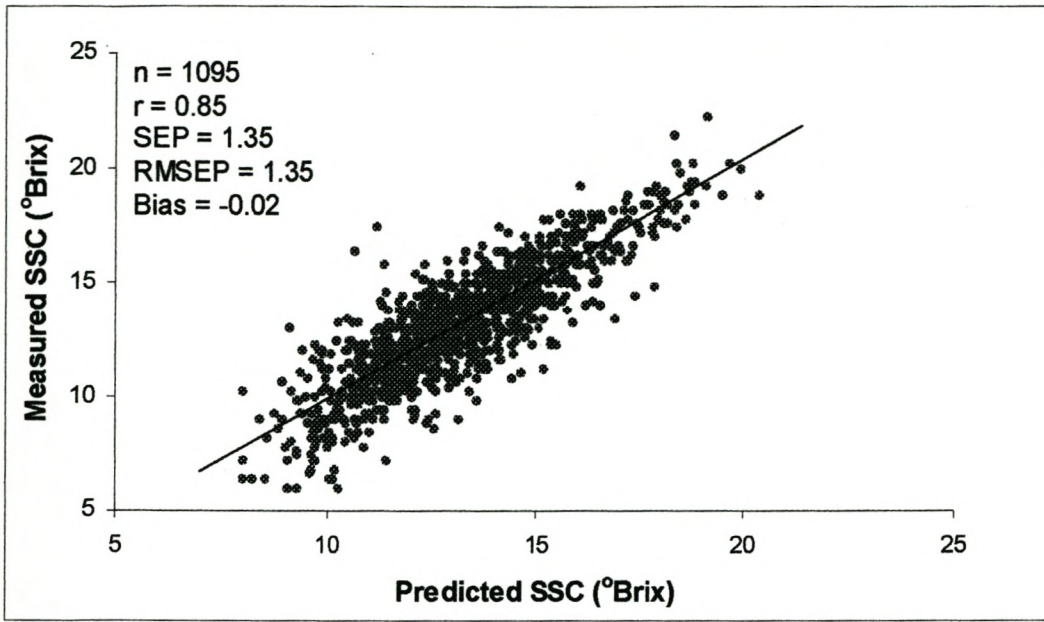
Figure 11. Validation plot of measured SSC (°Brix) vs. predicted SSC for samples of the 2002 season (MSC corrected spectra, 25 PC's) using the Diode Array spectrophotometer.

Table 5. MARS regression statistics for peach SSC obtained for the data within seasons (FT-NIRS used).

Statistic	2002	2003
Calibration		
No. of basis functions	30	30
n	2274	1954
r	0.84	0.0.81
SEC (°Brix)	1.29	1.47
RMSEC (°Brix)	1.29	1.47
Bias (°Brix)	0.001	0.001
Validation		
n	550	491
r	0.82	0.77
SEP (°Brix)	1.42	1.55
RMSEP (°Brix)	1.42	1.55
Bias (°Brix)	0.01	0.06

The Diode Array system gave slightly better results than the Fourier Transform instrument. Overall, the results are poorer than those reported by Slaughter (1995), who obtained r = 0.90 and SEP = 1.0% over four cultivars within a season. The reason for this could be that the SSC measurements of the present study were not made immediately after NIRS measurements, but two (2003) and three (2002) weeks later.

Regression was also performed on the data of separate cultivars to determine if variation between cultivars influenced the spectra and subsequent SSC regressions. Tables 6 & 7 summarise the PLS regression results for the 2002 and 2003 seasons, respectively. Correlation coefficients ranged from 0.68 to 0.83 for 2002 and 0.61 to 0.79 for 2003. SEP ranged from 1.00 to 1.59 °Brix for 2002 and 1.25 to 1.57 °Brix for 2003. Previous SSC calibrations performed on single peach cultivars within a harvesting season reported r-values of 0.96 (Kawano *et al.*, 1995) and 0.97 (Kawano *et al.*, 1992) and SEP-values of 0.42°Brix (Kawano *et al.*, 1995) and 0.50°Brix (Kawano *et al.*, 1992).

The optimum number of principal components (PC's) for the PLS models was 17 PC's for 2002 and 15 PC's for 2003. The numbers of PC's used were determined by monitoring the residual matrix (E-matrix) for the accumulating PC-model fit. Figures 12 & 13 show the residual validation variance for 2002 and 2003, respectively. There was a gradual decrease and then a flattening in residual variance in each curve. The PC which had the minimum Y-variance was chosen as the last PC to include in the model. The chosen number of PC's also gave the lowest SEP value in the regression results. These plots were also useful as an indication of how well the model described the relationships between X and Y. A decrease in the residual variance indicated a "good" model, i.e. that no extreme outliers were present and that there were systematic relationships between X and Y.

The importance of various X-variables (wavelengths) for predicting Y (SSC) was studied by plotting the PLS regression coefficients (B-coefficients) for the model. Figure 14 shows absolute values for regression coefficients for each variable (wavelength) of each PC for the 2002 model. Large absolute values for the variables indicate large significance, whereas small values indicate unimportant or noisy variables. The wavelengths 701, 705, 749, 753, 785, 795, 843, 883, 907, 953, 981, 1025, 1069 and 1143 nm were the most significant in the regression model for 2002. The regression coefficients were also plotted for the 2003 season (Figure 15). The wavelengths 723, 745, 773, 787, 815, 843, 915, 955, 1000-1061 and 1145 all had high regression coefficients. These variables are similar to those identified for the 2002 results. Peaks at 841 nm, 907, 955 nm and 1143 nm correlate well with results of Kawano *et al.* (1995) who identified peaks for water at 841 nm, 966 nm and 942 nm and SSC at 904 nm, 951 nm and 1141 nm.

111

Table 6. PLS regression statistics for peach SSC within cultivars of the 2002 season (FT-NIRS used).

Statistic	Cultivar					
	Goudmyn	Kakamas	Keisie	Neethling	Sandvliet	Western Sun
Calibration						
No. of PC's	5	17	17	15	20	15
n	77	1069	264	471	114	79
r	0.72	0.82	0.84	0.75	0.95	0.92
SEC (°Brix)	1.39	1.45	1.13	1.47	0.47	0.71
RMSEC (°Brix)	1.38	1.45	1.13	1.47	0.47	0.71
Bias (°Brix)	0.003	0.004	0.006	-0.005	0.001	-0.008
Validation						
n	39	529	118	227	57	38
r	0.68	0.83	0.71	0.71	0.75	0.72
SEP (°Brix)	1.44	1.38	1.40	1.59	0.99	1.23
RMSEP (°Brix)	1.46	1.38	1.41	1.59	0.98	1.23
Bias (°Brix)	0.32	-0.06	-0.22	0.03	0.02	-0.15

Table 7. PLS regression statistics for peach SSC within cultivars of the 2003 season (FT-NIRS used).

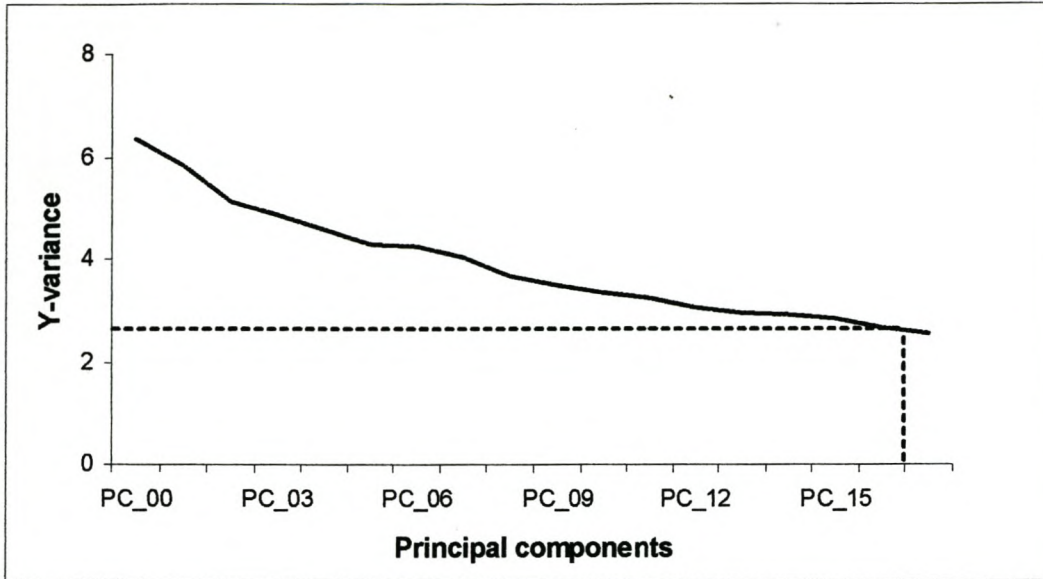| Statistic	Cultivar				
	Kakamas	Keisie	Neethling	Sandvliet	Woltemade
Calibration					
Nr. of PC's	19	15	15	14	10
n	957	247	518	124	181
r	0.81	0.85	0.82	0.90	0.64
SEC (°Brix)	1.43	1.25	1.32	0.93	1.60
RMSEC (°Brix)	1.43	1.25	1.32	0.93	1.60
Bias (°Brix)	0.00	0.015	-0.001	0.008	0.004
Validation					
n	489	134	257	67	84
r	0.79	0.73	0.79	0.78	0.61
SEP (°Brix)	1.57	1.53	1.52	1.25	1.56
RMSEP (°Brix)	1.57	1.52	1.52	1.24	1.55
Bias (°Brix)	0.14	0.02	-0.14	-0.01	-0.08

Figure 12. Residual validation variance plot for the SSC regression model of the 2002 (FT-NIRS used, MSC-corrected spectra) data.
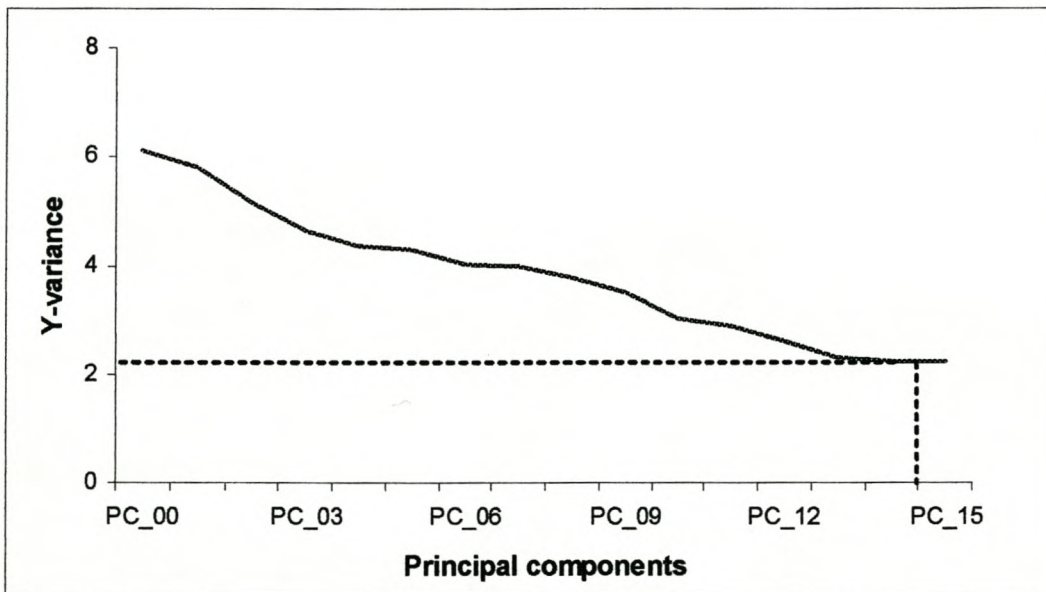

Figure 13. Residual validation variance plot for the SSC regression model of the 2003 data (FT-NIRS used, no pre-processing).
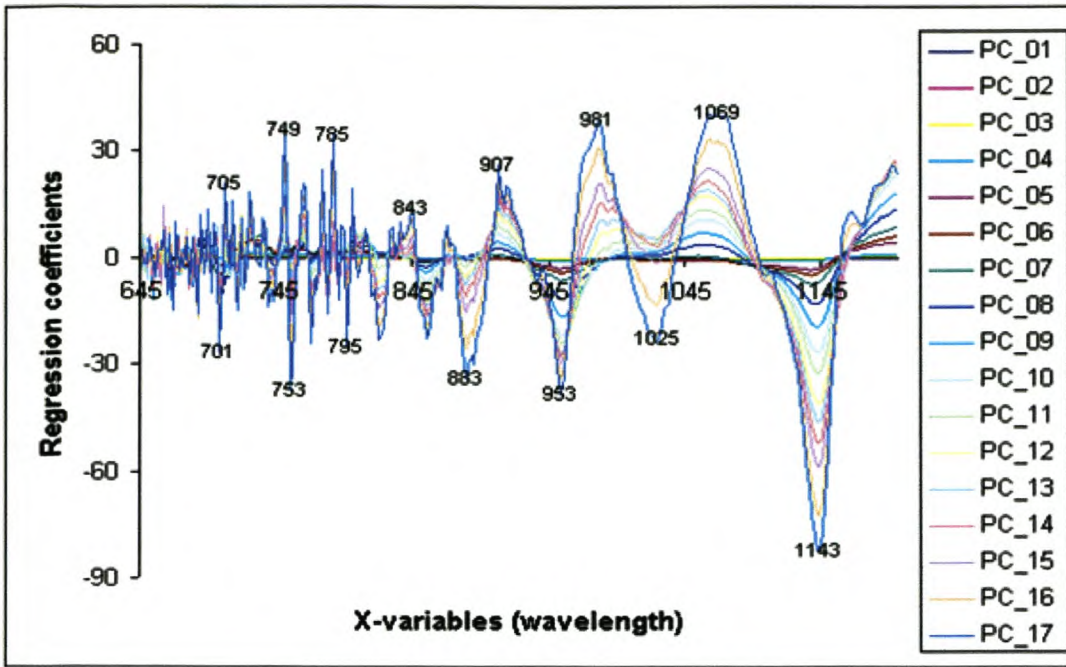
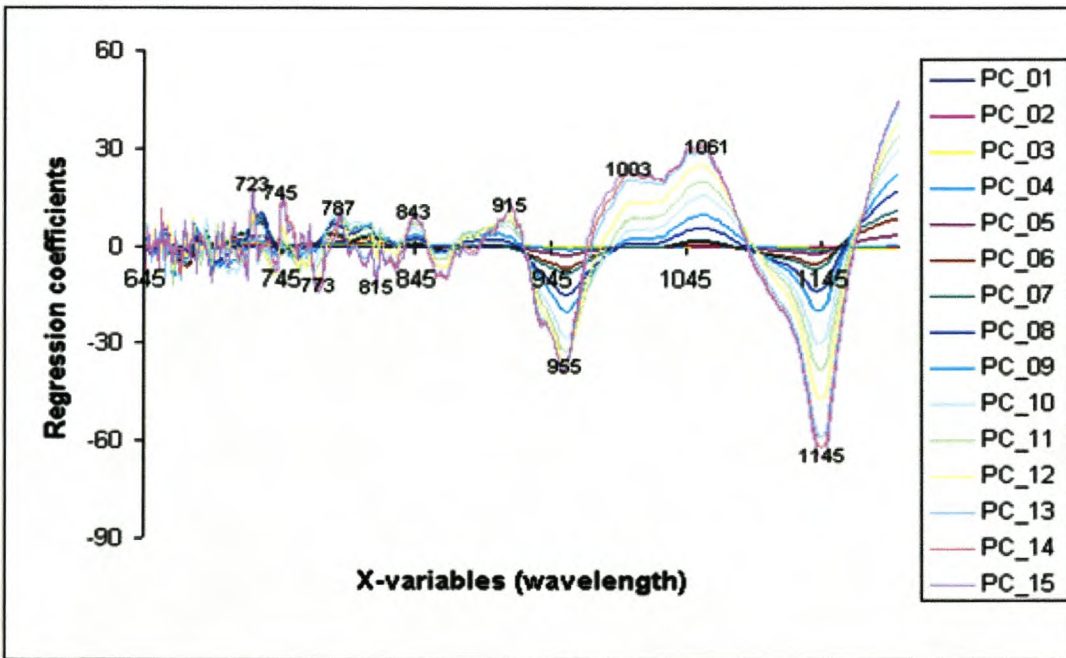**Figure 14**. Regression coefficient plot of the 17 PC's calculated for the SSC regression model of the 2002 data.

Figure 15. Regression coefficient plot of the 17 PC's calculated for the SSC regression model of the 2003 data.

114

The significant bands were also compared with chemical assignments summarised by Osborne *et al.* (1993). The most significant variable for these data sets was around 1143 nm. A C-H (aromatic structure) second overtone vibration can be assigned to this absorption band. The C-H second overtone can also be assigned to 915 nm. C-H can be any carbohydrates. Fruit ripening is known to be associated with an increase in SSC (sucrose and soluble pectin) and a decrease in organic acids (Bartley & Knee, 1982; Lill *et al.*, 1989). The significant bands at 905 and 1143 nm are probably owed to these constituents. The absorption bands 970 and 990 nm are vibrations of O-H second overtones, which can be ROH and H_2O. The large coefficients in this area (981 nm) are probably related to moisture and SSC. The large coefficients from about 1000 to 1069 could be O-H, C-H or N-H. Table 8 is a summary of the most important variables (wavelengths used for SSC regressions using PLS and MARS. They are compared to band assignments made by previous researchers. Overlapping absorptions in the NIR region made chemical assignments to specific absorption bands a difficult task.

Firmness

Partial least squares regression results of the firmness calibration and validation are shown in Table 9. Regression was preformed on all the samples within each harvesting season (Figures 16 & 17) and for the combined seasons, as was performed for the SSC regressions. Using the Fourier Transform spectrophotometer, correlation coefficients of 0.45, 0.46 and 0.40 were obtained for 2002 (MSC-corrected), 2003 (no pre-processing) and combined seasons (no pre-processing), respectively. SEP-values of 1.47 kg (14.42 N) for 2002, 1.43 kg (14.03 N) for 2003 and 1.53 kg (15.00 N) for the combined seasons were obtained, respectively.

MARS models were fitted to uncorrected spectra in conjunction with the first derivative spectral data. The MARS regression results (Table 10) for these data sets were similar, obtaining r-values of 0.66 and 0.52 and SEP-values of 1.24 kg (12.16 N) and 1.33 kg (13.05 N) for the 2002 and 2003 seasons, respectively. PLS regression performed on spectral data (MSC-corrected) from the Diode Array spectrophotometer gave a r-value of 0.52 and SEP-value of 1.41 kg (13.83 N) for the 2002 season. Previous studies predicted firmness of kiwifruit (r = 0.66 and RMSEP = 7.8 N) (McGlone & Kawano, 1998) and apples (r = 0.74 and RMSEP = 7.5 N) (McGlone *et al.*, 2002).

Various factors could contribute to poor firmness predictions models based on NIR spectra. Spectra measured before storage was correlated with firmness values measured after weeks of storage. During this time structural changes such as softening and

115

Table 8. Important variables used in the regression equations for SSC and firmness with the respective possible chemical and constituent assignments.

Wavelength region (nm)				Previously assigned (specific wavelength)	Reference	Chemical assignment Osborne et al., 1993	Possible constituent
SSC		Firmness					
PLS	MARS	PLS	MARS				
	1179^2, 1167^2		1177^2	Water (1168)	Kawano et al., 1995	HC=CH	Sugars (SSC)
	1151^2, 1153^2, 1149^2			Water (1153)	Williams & Norris, 1987	C-H 2nd overtone (CH_3)	Carbohydrates e.g. sugars, pectins
1141^1, 1143^1, 1145^1	1139^2, 1133^2	1145^1, 1135^1	1145^1, 1133^1	SSC (1141)	Kawano et al., 1995	C-H (aromatic) 2nd overtone	Carbohydrates e.g. sugars (SSC)
1003^1-1061^1	1097^2	997^1-1035^1	1005^1	Sugar (1005)	Williams & Norris, 1987	CH – cyclopropane structure	Sugars (SSC)
843^1				Water (841)	Kawano et al., 1995	CH – benzene structure	Sugars (SSC)
	1083^2, 1081^2	861^1	1081^2, 867^2	Starch (860)	Williams & Norris, 1987		
				Firmness (864)	McGlone et al., 1997		
	1061^2		1065^2			RNH_2	Nitrogen
	1053^2			Starch (1053)	Williams & Norris, 1987	CH_2	Sugars (SSC), starch
	1011^2			Water (1010)	Williams & Norris, 1987	CH_3	Sugars (SSC), moisture, pectins
981^1, 955^1, 953^1	993^2, 969^2, 961^2, 959^2		985^2, 979^2,	Water (994, 986, 978, 958)	Williams & Norris, 1987	O-H second overtones (ROH, H_2O, starch)	Sugars (SSC), moisture, starch, cellulose
				Starch (979), Cellulose (978)	Williams & Norris, 1987		

[1] Spectra without pre-processing

[2] Spectra treated with first derivatives

116

Table 8. Continued.

Wavelength region (nm)				Previously assigned (specific wavelength)	Reference	Chemical assignment Osborne et al., 1993	Possible constituent
SSC		Firmness					
PLS	MARS	PLS	MARS				
				Water (968), SSC (951)	Kawano et al., 1995	O-H second overtones (ROH, H_2O, starch)	Sugars (SSC), moisture, starch, cellulose
	935[2]	937[1],				CH_2 (third overtone)	Sugars (SSC)
915[1]		915[1]		Cellulose (920, 905)	Williams & Norris, 1987		Sugars (SSC), pectins
				Starch (918, 901)	Williams & Norris, 1987		
				SSC (910)	Slaughter, 1995		
				Carbohydrates (906)	Kawano et al., 1992 Kawano et al., 1995		
				SSC (904)	Williams & Norris, 1987		
				Sucrose (913, 888, 838)			
907[1], 883[1]	905[2], 897[2]	889[1]	891[2]			CH_3 (third overtone)	Sugars (SSC), pectins
785[1]	767[2]	785[1], 769[1]	755[2]			CH_2 (fourth overtone)	Sugars (SSC) Moisture
753[1], 749[1], 745[1]		751[1]				O-H (third overtone) (H_2O)	
	697[2],	697[1],		Chlorophyll (680)	McGlone & Kawano, 1998	Chlorophyll	Colour, ripeness (SSC, firmness)
	689[2], 673[1], 653[1]	671[1]	699[2], 689[2]				Colour, ripeness (SSC, firmness)

Table 9. PLS regression statistics for peach firmness (kg) within seasons and over the combined seasons.

Statistic	2002[1]	2003[1]	Combined seasons	2002[2]
Calibration				
No. of PC's	9	14	15	16
n	2597	2216	4479	2453
r	0.46	0.46	0.40	0.52
SEC (kg)	1.48 kg	1.42	1.53	1.41
RMSEC (kg)	1.48 kg	1.42	1.53	1.41
Bias (kg)	-0.001	-0.003	0.002	-0.002
Validation				
n	1042	920	2296	1164
r	0.45	0.47	0.40	0.52
SEP (kg)	1.47	1.43	1.53	1.41
RMSEP (kg)	1.47	1.43	1.53	1.41
Bias (kg)	0.03	-0.03	0.004	0.01

[1]Fourier Transform spectrophotometer
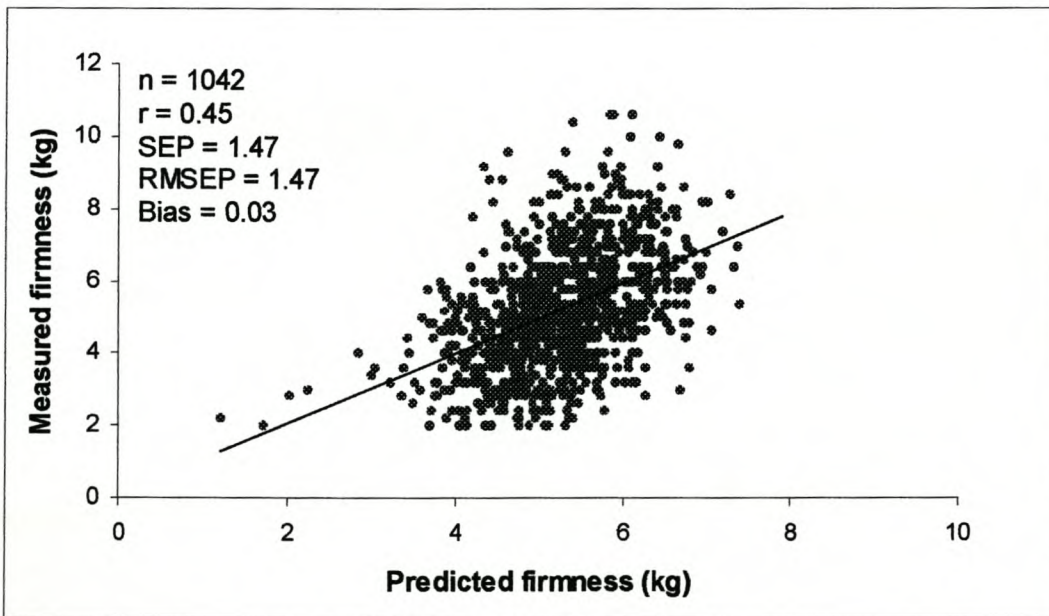
[2]Diode Array spectrophotometer

Figure 16. Validation plot of measured firmness (kg) versus predicted firmness for samples of the 2002 season (MSC-corrected spectra, 9 PC's) using FT-NIRS.
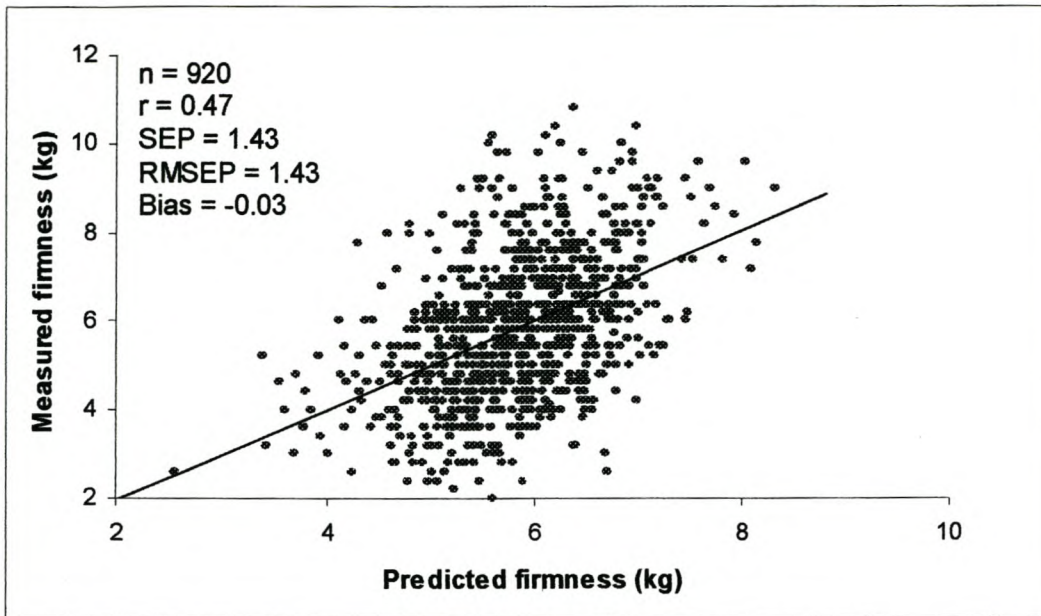
Figure 17. Validation plot of measured firmness (kg) versus predicted firmness for samples of the 2003 season (no pre-processing, 14 PC's) using FT-NIRS.

Table 10. MARS regression statistics for peach firmness (kg) of data from the 2002 and 2003 season (FT-NIRS used).

Statistic	2002	2003
Calibration		
No. of basis functions	30	30
n	1492	2140
r	0.72	0.60
SEC (kg)	1.11	1.24
RMSEC (kg)	1.11	1.24
Bias (kg)	0.001	0.002
Validation		
n	353	528
r	0.66	0.52
SEP (kg)	1.24	1.33
RMSEP (kg)	1.24	1.33
Bias (kg)	0.001	0.03

119

disintegration took place in some of the fruit. The hand-held penetrometer is also not a very accurate reference method (McGlone *et al.*, 2002). Both these factors could lead to poor relationships between spectral and reference measurements. No specific chemical assignments for firmness could be made as was possible for SSC.

The residual matrix (E-matrix) for the accumulating PC-model fit was studied to determine optimum number of PC's for each PLS regression of firmness. Figures 18 & 19 show that the residual variance came to a minimum at PC 9 for the 2002 regression and at PC 14 for the 2003 regression. The very slight decrease in residual variance in both cases is an indication of a systematic relationship between X and Y. However, the model could not be evaluated as a "good" model, because there was very little decrease in the variance of each accumulating PC. All the PC's had almost the same low variance.

The importance of various X-variables (wavelengths) for predicting Y (firmness) was evaluated by plotting the PLS regression coefficients (B-coefficients) for the model. Figures 20 & 21 show the regression coefficients for each variable (wavelength) of each PC for the 2002 and 2003 models, respectively. The significant bands were compared to chemical assignments summarised (Table 8). Significant variables appear in the visible range (around 680 nm), which indicates that the firmness could be related to the colour (chlorophyll content) of the fruit (Moons *et al.*, 1998, McGlone & Kawano, 1998). A decrease in chlorophyll is an indication of fruit ripening (Wills *et al.* 1989) and ripening is also associated with a decrease in firmness (Lill *et al.*, 1989).

The most significant variable in the NIR range was at 785 nm. The C-H third overtone can be assigned to this wavelength, possibly indicating sugars. Another significant variable is in the region of 1135 and 1143 nm. This was also observed in the SSC regression model. A C-H (aromatic structure) second overtone vibration can be assigned to this absorption band. The high peaks in the region of approximately 977 to 1045 nm could be vibration bands of O-H second overtones (ROH and H_2O), C-H (CH_3) and N-H second overtones. These bands were also significant in calculating the SSC regression model. C-H third overtones can also be assigned to the high peaks at 915, 937 and 889 nm (Figure 21). These overtones include mono-, di- and oligosaccharides, which could include sugars and pectins (the constituents of SSC). Fruit ripening is known to be associated with an increase in sucrose and soluble pectin (Bartley & Knee, 1982; Lill *et al.*, 1989). The measurement of firmness can thus be used as an index of fruit ripening.

The 861 nm region corresponds with the region used by McGlone *et al.* (1997) who used 864 nm for firmness determinations. Pectins, which play a major role in fruit ripening and senescence, can probably be assigned to this wavelength. It seems that the most important
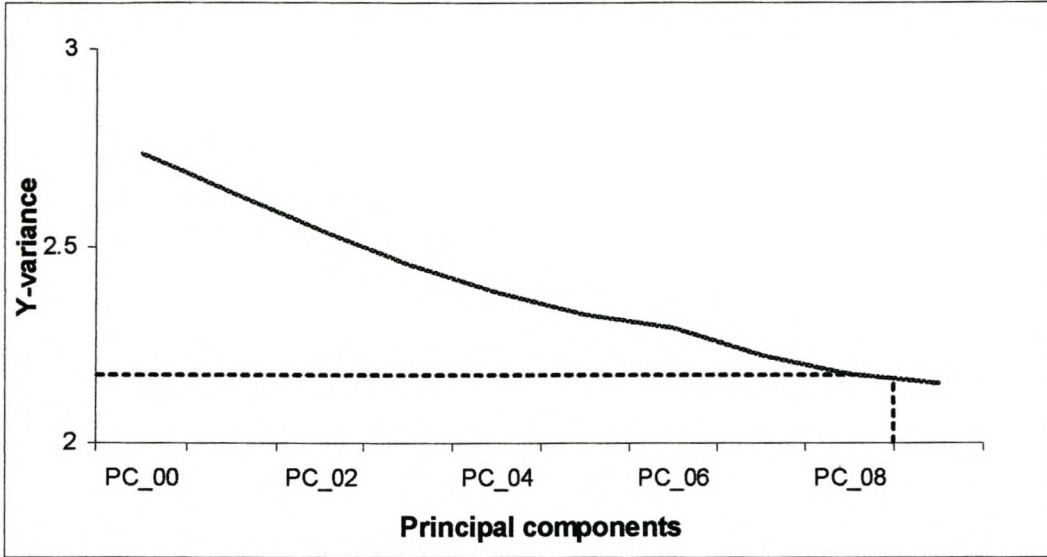
120

Figure 18. Residual validation variance plot indicating the PC-model fit for the firmness regression model of the 2002 season (FT-NIRS used).
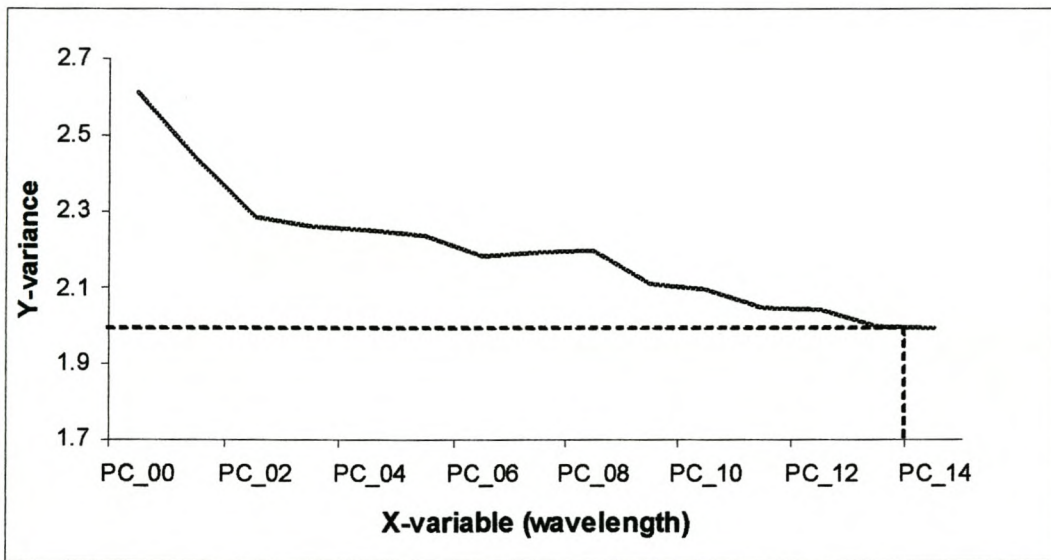

Figure 19. Residual validation variance plot indicating the PC-model fit for the firmness regression model of the 2003 season (FT-NIRS used).
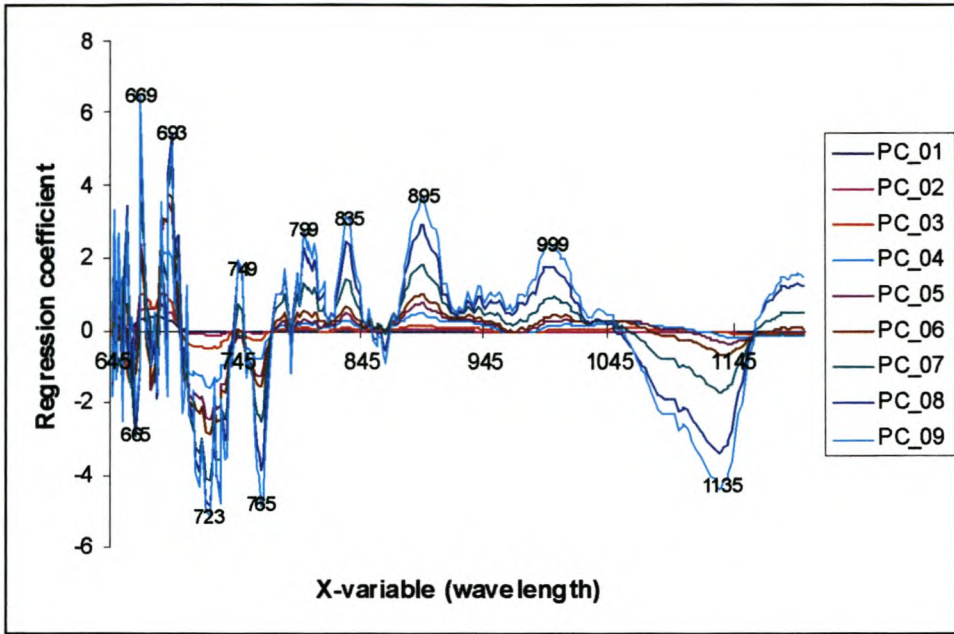
Figure 20. Plot of regression coefficients calculated for X-variables (wavelengths) for the PC's of the 2002 firmness regression model.
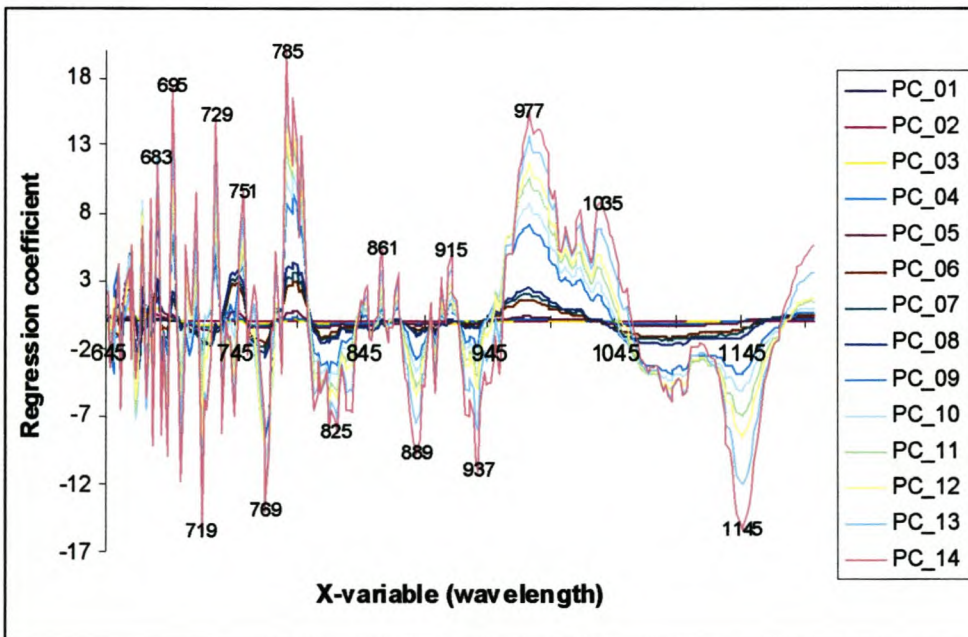

Figure 21. Plot of regression coefficients calculated for X-variables (wavelengths) for the PC's of the 2003 firmness regression model.

122

variables used for calculating a firmness regression model were those of C-H vibrations (due to sugars and pectins) in the NIR region (720-1145 nm) as well as chlorophyll in the visible range (650-700 nm).

Subjective evaluations

Several fruit developed microbial spoilage during the storage period, because the fruit could not be treated with antimicrobials. Fruit which developed spoilage were not included in the sample set for evaluations. Tables 11 & 12 show the results of the subjective evaluations made on each half of the fruit. These results are compared in Figure 22. The overall quality of the samples after the two and three week storage periods was poor, as over 50% of the fruit in each season were defective and would have been unsuitable for canning. In the industry, however, the fruit are usually stored for a shorter period and would thus have less deterioration than the results obtained in this study. Only 37% and 42% of the fruit were evaluated as having good post-storage quality for the 2002 and 2003 seasons, respectively. The cultivar Sandvliet had the best keeping quality (62% for 2002 and 73% for 2003), followed by Neethling and Keisie, then by Kakamas, Western Sun, Goudmyn, Woltemade and lastly Professor Black. Too few samples (n = 40) of Professor Black were, however, present to make any significant conclusions about this cultivar's keeping quality. Recommendations can be made about which cultivars are more suitable for cold storage. Cultivars such as Woltemade should be processed upon arrival at the factory, while cultivars such as Sandvliet can be cold stored before canning.

Classification

Approach 1: Classification trees

The first approach was to find a correlation between all objective measurements (mass, diameter, SSC, firmness and cultivar) made after storage and the concluding post-storage quality data (subjective evaluations). A non-destructive classification for storage potential would then be possible by a correlation between NIR spectra and the objective measurements.

A classification tree (Figure 23) was obtained for data from the 2003 season. The split of the data was made by the classification tree according to firmness (pressure) measurements, followed by the size (diameter) measurements, as these measurements gave best splits based on entropy reduction. The fruit cultivar, mass and SSC measurements were rejected by the classification tree and were thus not sufficient for predicting post-storage quality.

Table 11. Results of subjective evaluations after 3 weeks storage during the 2002 season.

Data set	Total[1] (n)	Good post-storage quality[2] (n)	Good post-storage quality (%)
Kakamas	1628	596	36.61
Neethling	800	370	46.25
Keisie	392	163	41.58
Sandvliet	200	123	61.50
Woltemade	520	93	17.88
Goudmyn	120	37	30.83
Western Sun	120	43	35.83
Professor Black	40	5	12.50
Total	3820	1430	37.43

[1]Total number of samples used for the data set

[2]Total number of samples with no defects after cold storage

Table 12. Results of subjective evaluations after 2 weeks storage during the 2003 season.

Data set	Total[1] (n)	Good post-storage quality[2] (n)	Good post-storage quality (%)
Kakamas	1453	524	36.06
Neethling	787	350	44.47
Keisie	394	195	49.49
Sandvliet	196	144	73.47
Woltemade	279	95	34.05
Professor Black	40	1	2.50
Total	3149	1309	41.57

[1]Total number of samples used for the data set

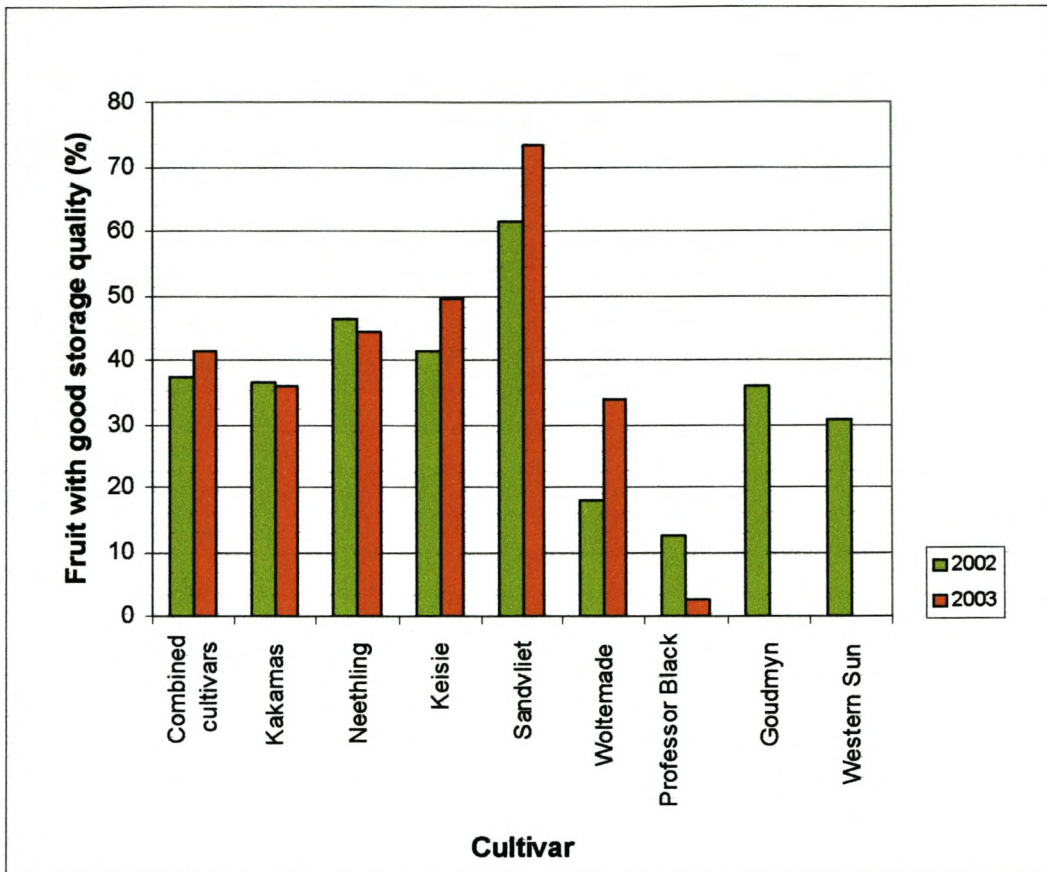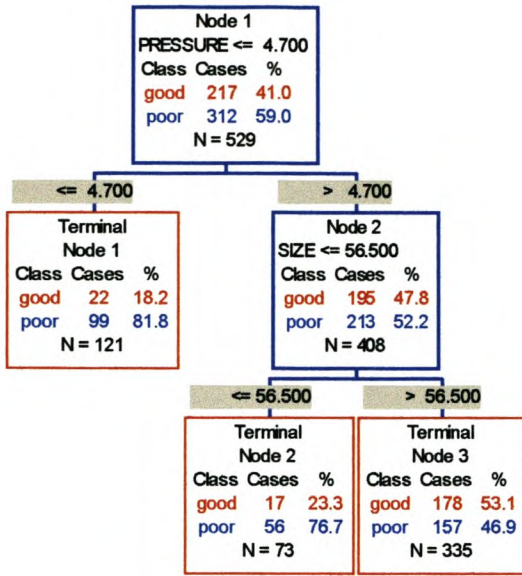[2]Total number of samples with no defects after cold storage

Figure 22. Percentage of fruit which had good post-storage quality.

(a)

```
                          Node 1
                   PRESSURE <= 4.700
                   Class  Cases   %
                   good    217   41.0
                   poor    312   59.0
                          N = 529
```

| <= 4.700 | | > 4.700 |

```
        Terminal                     Node 2
        Node 1                  SIZE <= 56.500
   Class  Cases   %             Class  Cases   %
   good    22    18.2           good    195   47.8
   poor    99    81.8           poor    213   52.2
        N = 121                       N = 408
```

| <= 56.500 | | > 56.500 |

```
        Terminal                   Terminal
        Node 2                     Node 3
   Class  Cases   %           Class  Cases   %
   good    17    23.3         good    178   53.1
   poor    56    76.7         poor    157   46.9
        N = 73                     N = 335
```
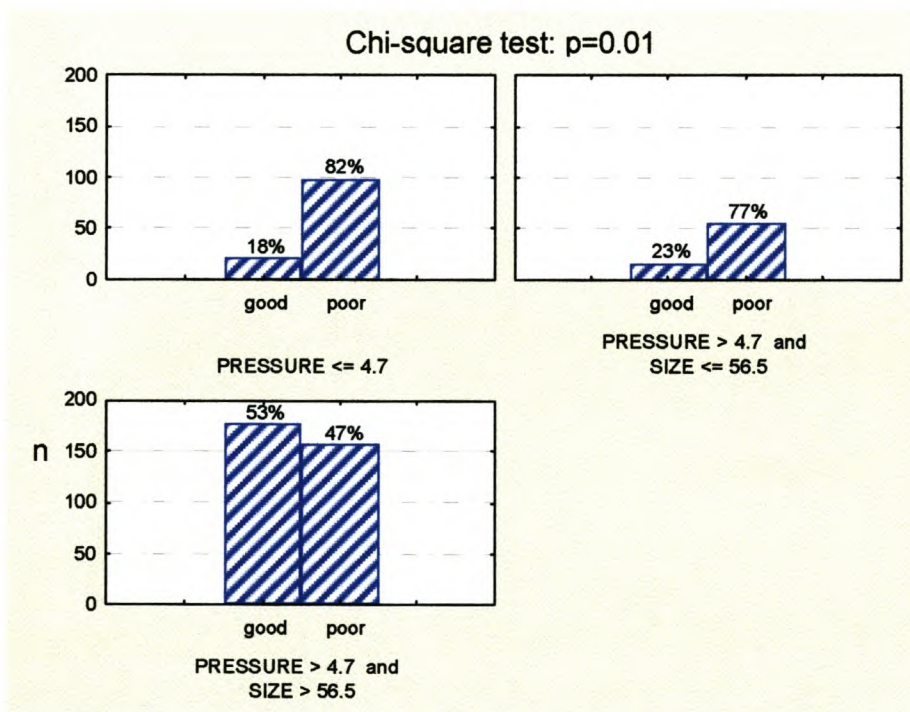
(b)

Figure 23. Classification trees splitting the data into post-storage quality by firmness and then by fruit diameter, with (a) a diagram showing the split of the classification tree and (b) a graphical summary.

A calibration set of 529 observations were firstly divided according to firmness, and the best split was found at 4.7 kg. According to this value, the calibration set was split into two nodes consisting of 121 and 408 samples, respectively. When the firmness reading was less than or equal to 4.7 kg, 81.8% of the 121 samples had poor post-storage quality and 18.2% good post-storage quality. When the firmness reading was more than 4.7 kg, 52.2% of the 408 samples had poor storage quality and 47.8% good post-storage quality. A further split was made on this node using the size (diameter) of the fruit, with a threshold value of 56.5 mm. When the diameter of the fruit was less than or equal to 56.5 mm, 76.7% of the 73 samples had poor post-storage quality and 23.3% good post-storage quality. When the diameter of the fruit was more than 56.5 mm, 46.9% of the 335 samples had poor post-storage quality and 53.1% good post-storage quality.

From this classification tree, one can conclude that less firm fruit (pressure readings ≤4.7 kg) after storage had a high percentage of fruit with poor post-storage quality. Firm and small fruit (pressure readings >4.7 kg and diameters ≤56.5 mm) after storage also showed a high percentage of fruit with poor storage quality. However, firm and large fruit (pressure readings >4.7 kg and diameters >56.5 mm) gave a split of 47% good and 53% poor post-storage quality.

This threshold value for fruit diameter correlates well with the current South African grading system, where fruit with a minimum of 57 mm and 55 mm are graded as canning grade one (best quality) and canning grade two, respectively. Firmness is not measured in the South African peach grading system. The classification tree can be used to some extent to predict storage potential based on firmness, once the firmness values of the fruit are known. The industry can process small fruit with pressure readings of less than 4.7 kg immediately, as these fruit will not have a good storage potential. NIRS and regression statistics provide a possible non-destructive method to determine such values. It is suggested that NIR spectra of peach samples can be used in a regression equation to obtain the predicted firmness values. However, results from firmness regressions earlier in this chapter (Tables 9 & 10) showed correlation coefficients (r) of 0.45-0.66 and standard errors of predictions (SEP) of 1.24-1.47 kg within seasons. Firmness predictions should be improved or an alternative method of classification should be developed.

Approach 2: SIMCA and MARS

Spectral analysis

Principal component analysis (PCA) performed within cultivars and within weeks of harvest showed some separation of the spectra of the fruit samples according to their storage quality.

Very little separation was visible when PCA was performed within cultivars over all the weeks of harvest in each season (Figures 24-28) No separation of spectral data of the samples according to their storage quality was visible in the PCA of all the samples (all cultivars and all weeks of harvest) within each season (Figure 29). All possible combinations of PC's were plotted against each other and viewed in two and three dimensional plots. The best separation was obtained each time by plotting the first three PC's against each other in three dimensional plots. For each of the cultivar-specific PCA, 6 PC's were calculated. For each seasonal data set 10 PC's were calculated. These numbers were determined by monitoring the residual matrix (E-matrix) for the accumulating PC-model fit of each PCA. The explained variance plot in Figure 30 shows the total variance explained by the accumulating PC's of a seasonal data set. The PC which had the maximum X-variance was chosen as the last PC to include in the model. The same strategy for determining the optimum number of PC's was used for all the other data sets.

The loading plot shows how much each variable (wavelength) contributes to each PC. Variables with a high degree of variation show large loadings. Figure 31 is a loading plot indicating which wavelengths were important in the calculation of each of the 10 PC's. Variables showing the most variation are those wavelengths below 745 nm (visible range). Chlorophyll absorbs in the visible range at about 680 nm (McGlone & Kawano, 1998). Thus, most of the variation in these data sets can be attributed to the chlorophyll content, or the fruit colour, which relates to the fruit maturity (Wills *et al.*, 1989). This observation supports the hypothesis that breakdown in peaches is associated with fruit maturity (Fernández-Trujillo *et al.*, 1998). It was decided to use the entire wavelength range measured (645-1201 nm), as this resulted in better classification results.

Classification of peach samples

SIMCA

Tables 13 & 14 show classification results within seasons and within cultivars for the 2002 and 2003 seasons, respectively. Classification results are given with respect to the model created for good post-storage quality. The recognition rate, or sensitivity, is the number of samples correctly predicted as belonging to the model out of the total amount that truly belonged (subjectively evaluated as good post-storage quality) to the model. The rejection rate, or specificity, is the percentage of samples not assigned to the model out of the total amount of samples that truly did not belong (subjectively evaluated as poor post-storage quality) to the model.
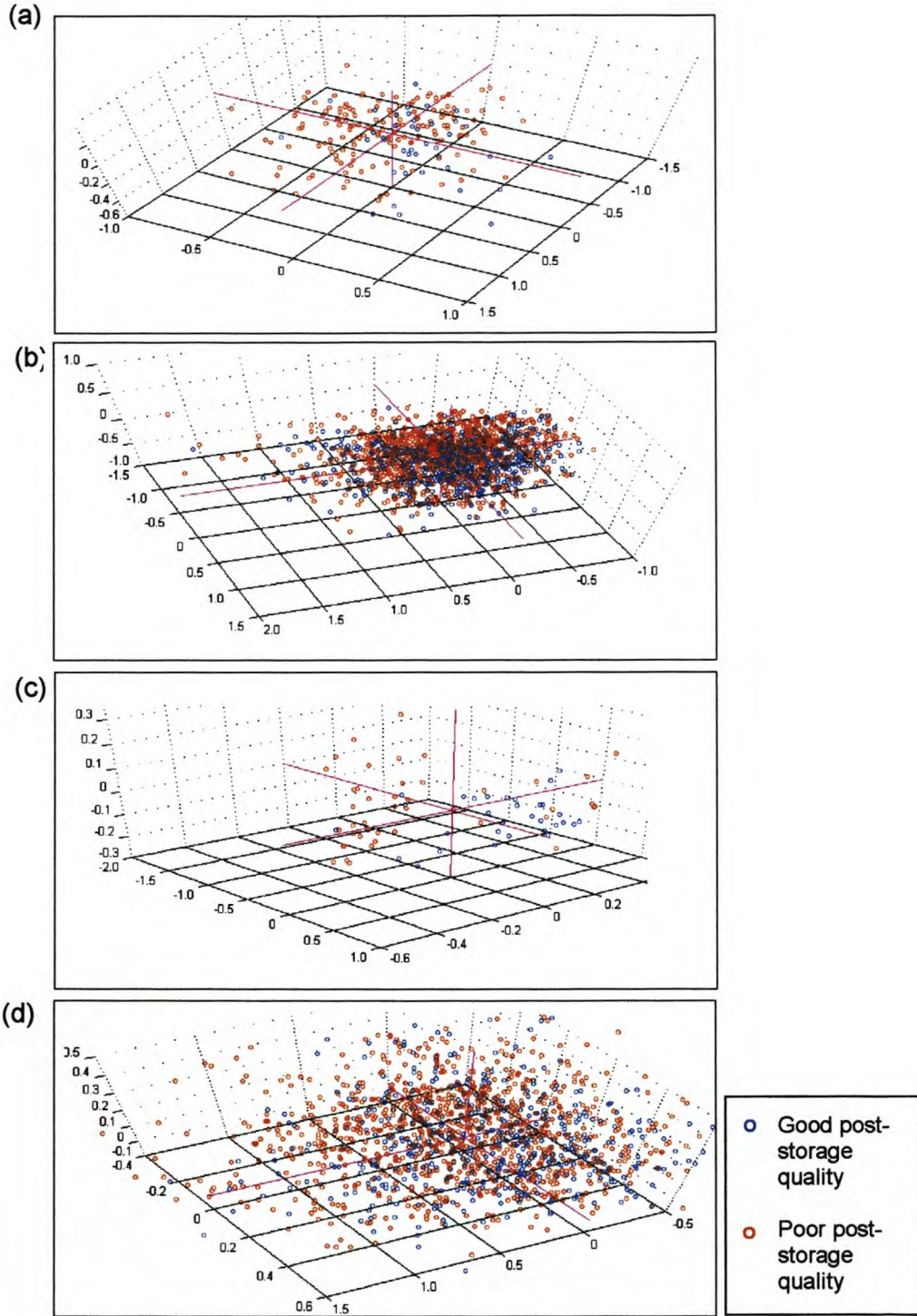
128

Figure 24. PCA (PC 1 vs. PC 2 vs. PC 3) of the spectra of Kakamas peaches, indicating some separation between fruit with good and poor post-storage quality. (a) Week 5, 2002, (b) all weeks, 2002, (c) week 7, 2003 and (d) all weeks, 2003.
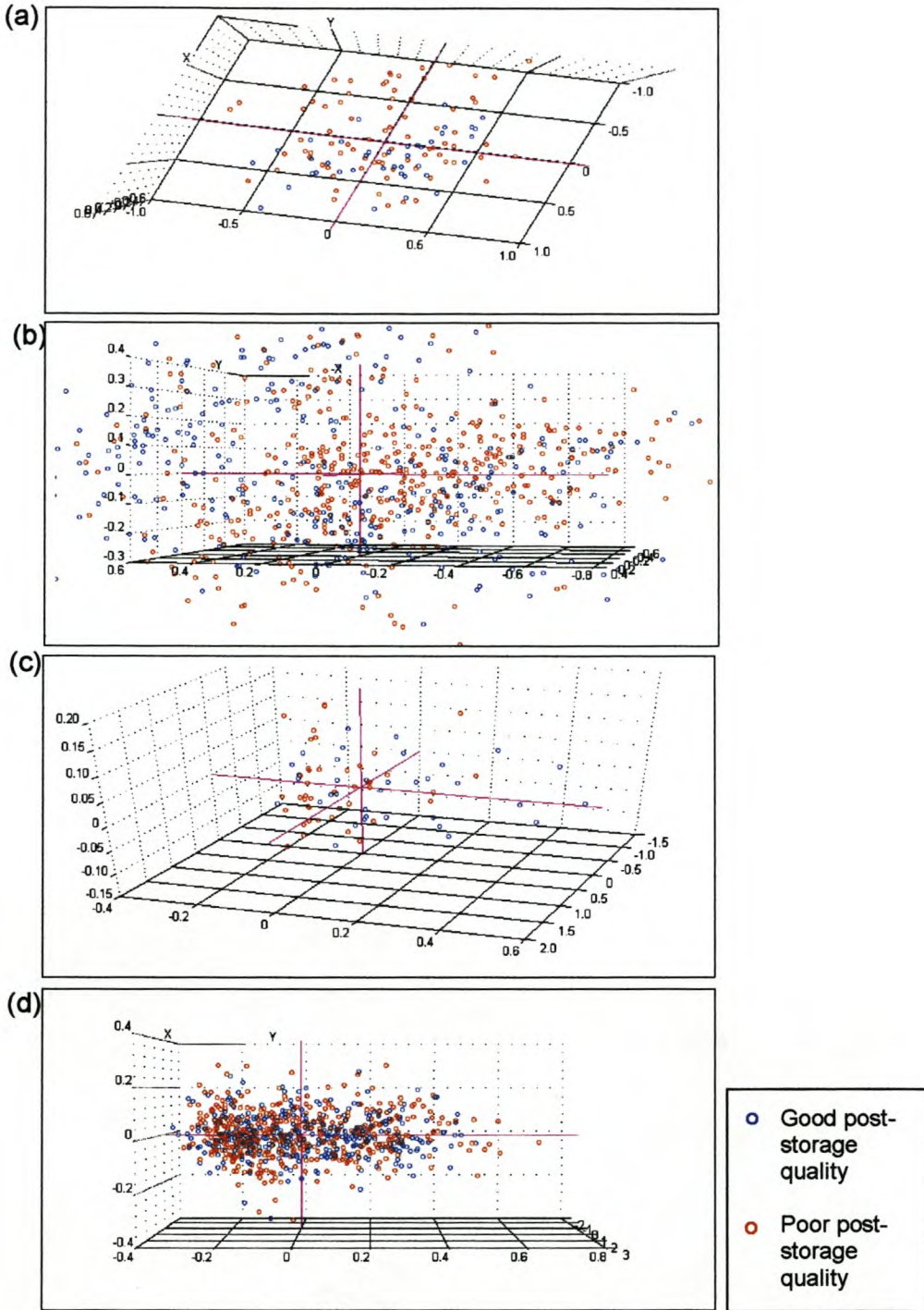
Figure 25. PCA (PC 1 vs. PC 2 vs. PC 3) of the spectra of Neethling peaches, indicating some separation between fruit with good and poor post-storage quality. (a) Week 1, 2002, (b) all weeks, 2002, (c) week 1, 2003 and (d) all weeks, 2003.
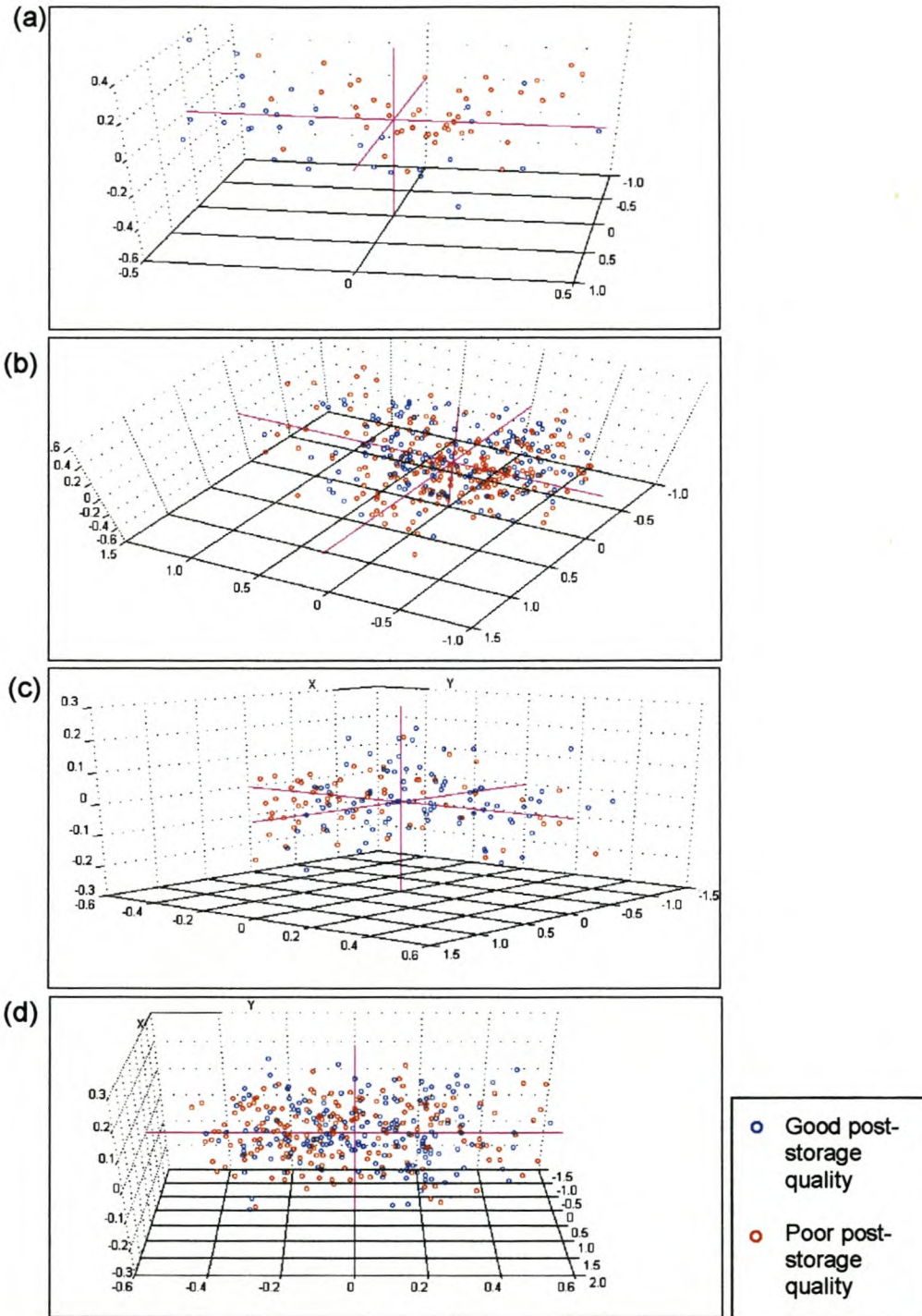
Figure 26. PCA (PC 1 vs. PC 2 vs. PC 3) of the spectra of Keisie peaches, indicating some separation between fruit with good and poor post-storage quality. (a) Week 3, 2002, (b) all weeks, 2002, (c) week 1, 2003 and (d) all weeks, 2003.

Figure 27. PCA (PC 1 vs. PC 2 vs. PC 3) of the spectra of Sandvliet peaches, indicating some separation between fruit with good and poor post-storage quality. (a) Week 1, 2002, (b) all weeks, 2002, (c) week 1, 2003 and (d) all weeks, 2003.
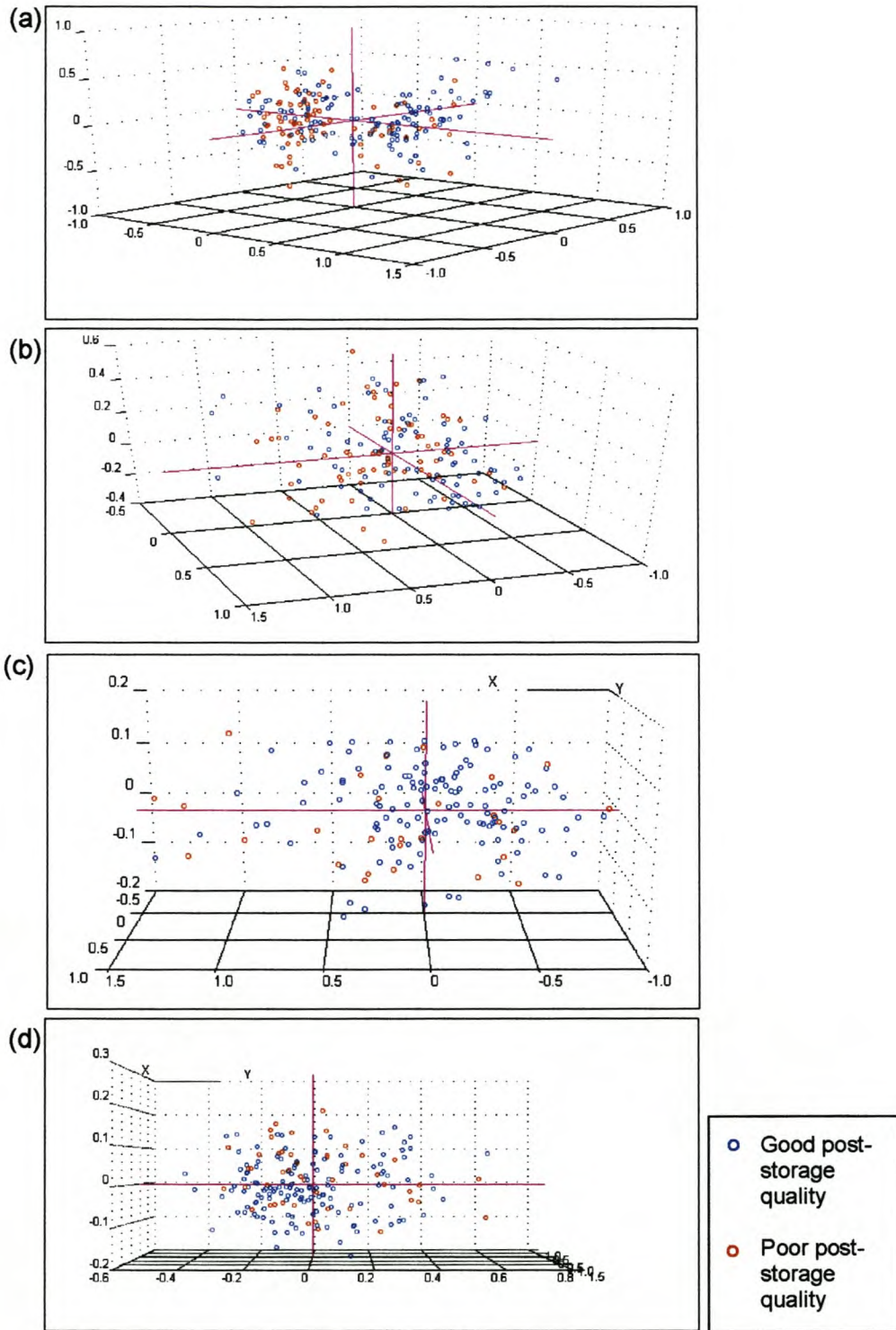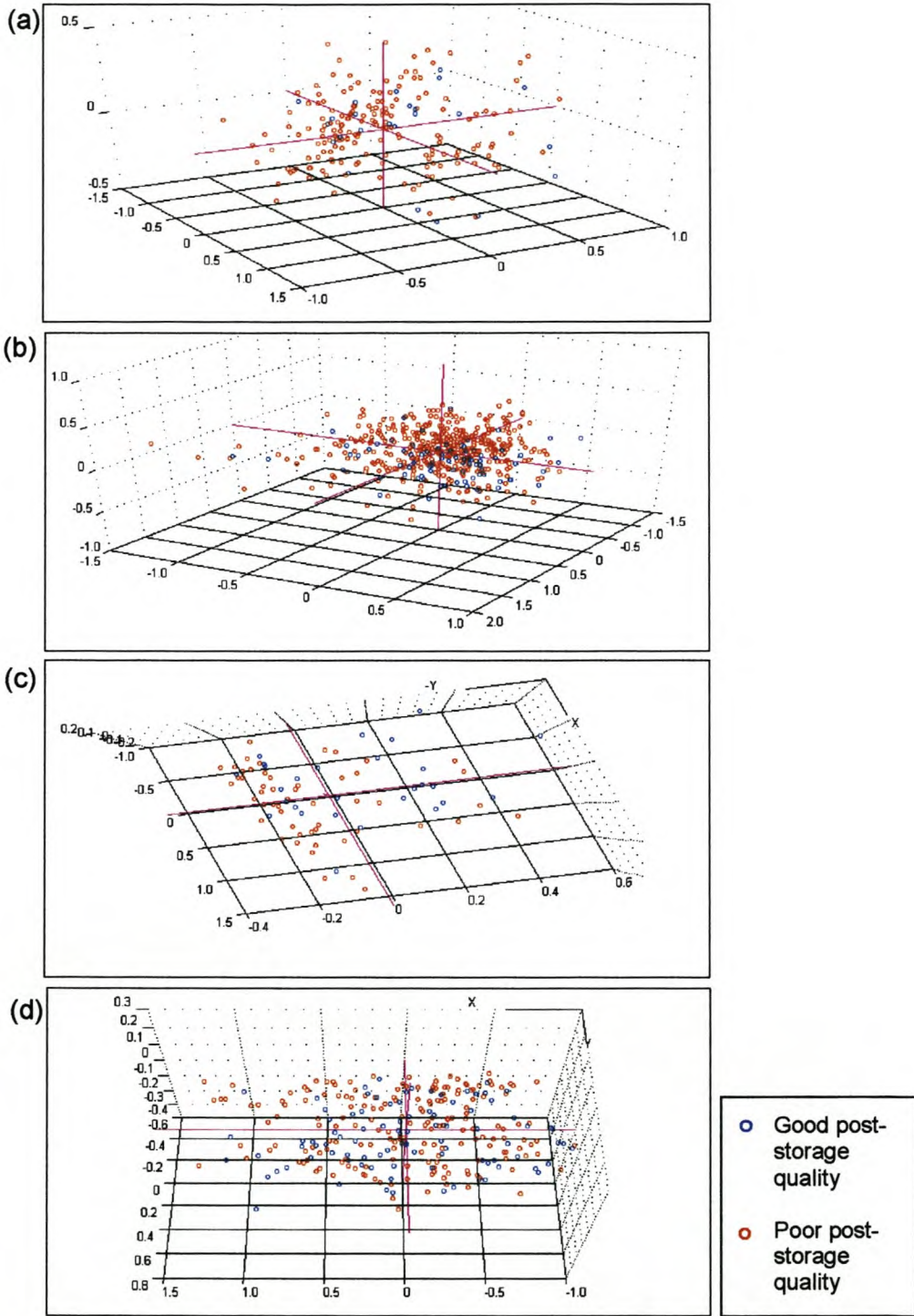
Figure 28. PCA (PC 1 vs. PC 2 vs. PC 3) of the spectra of Woltemade peaches, indicating the spectra of fruit with good and poor post-storage quality. (a) Week 4, 2002, (b) all weeks, 2002, (c) week 5, 2003 and (d) all weeks, 2003.
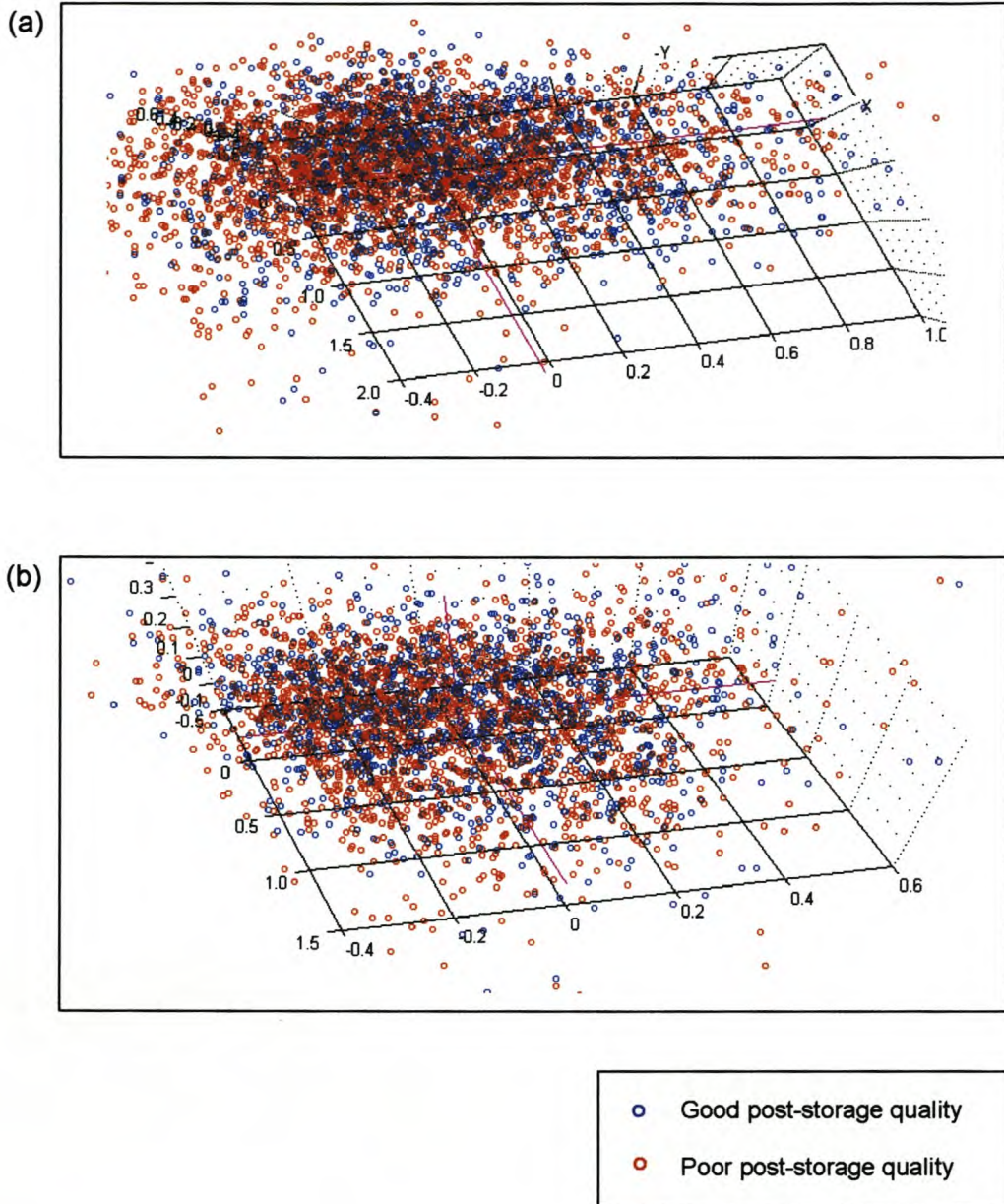
Figure 29. PCA (PC 1 vs. PC 2 vs. PC 3) of the spectra of peaches within the entire (a) 2002 and (b) 2003 season, indicating the spectra of fruit with good and poor post-storage quality. No separation of spectra with good and poor post-storage quality is visible.
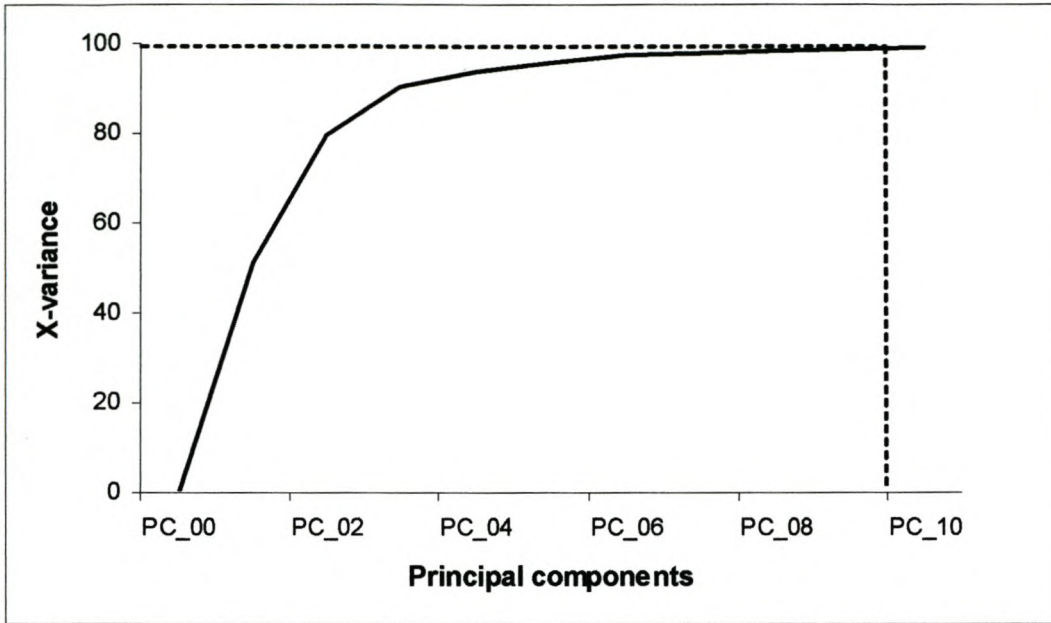
Figure 30. Explained variance plot of the spectral data from the 2002 season, indicating the total variance explained by the accumulating PC's.
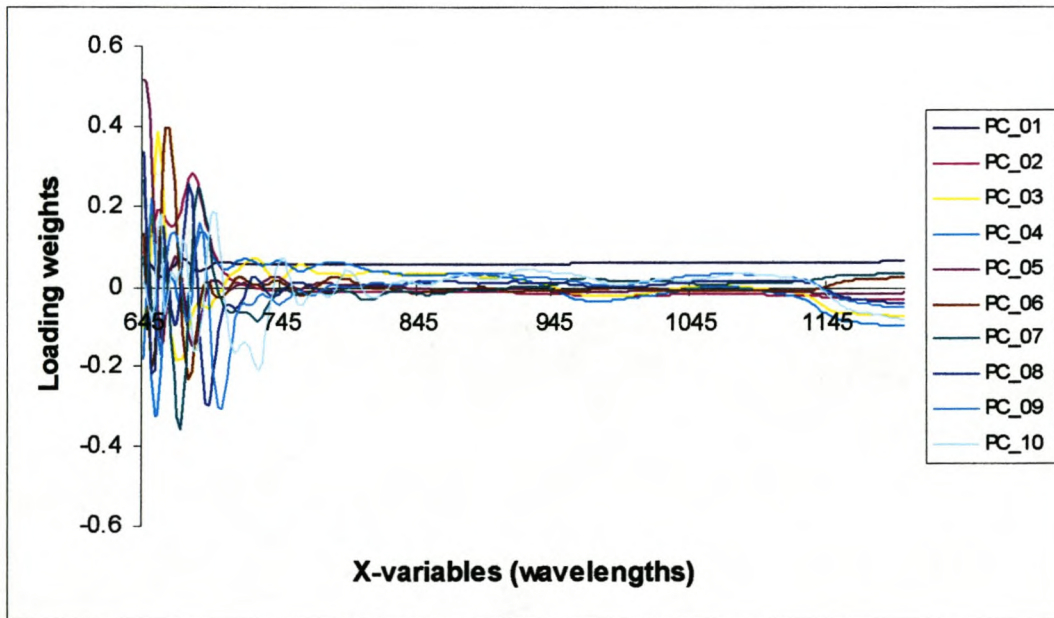

Figure 31. The loadings plot indicates the most important variables (wavelengths) used for the calculation of each principal component for the spectral data from the 2002 season.

Table 13. SIMCA classification results performed on data within cultivars and within seasons.

Data set	Recognition rate		Rejection rate		Total classification rate	
	n	%	n	%	n	%
2002						
Kakamas	154/198	77.78	86/293	29.35	240/491	49.00
Neethling	59/110	53.64	91/143	63.64	150/253	59.00
Keisie	14/54	25.93	64/76	84.21	78/130	60.00
Sandvliet	10/41	24.39	19/25	76.00	29/66	44.00
Woltemade	12/29	41.38	80/130	61.54	92/159	58.00
2003						
Kakamas	95/174	54.60	150/310	48.39	245/484	51.00
Neethling	61/116	56.90	80/146	54.79	141/262	54.00
Keisie	47/65	72.31	23/66	34.85	70/131	53.00
Sandvliet	37/47	78.72	3/17	17.65	40/64	63.00
Woltemade	13/31	41.94	39/62	62.90	52/93	56.00

Table 14. SIMCA Classification results performed on all data within seasons.

Data set	Recognition rate		Rejection rate		Total classification rate	
	n	%	n	%	n	%
2002 a-half	131/228	57.00	224/403	56.00	355/631	56.00
2002 b-half	124/246	50.00	229/393	58.00	353/639	55.00
2002 both halves	265/475	55.79	453/795	56.98	718/1270	57.00
2003 a-half	42/213	20.00	248/285	87.00	290/498	58.00
2003 b-half	115/223	52.00	201/303	66.00	316/526	60.00
2003 both halves	233/438	53.20	328/611	53.68	561/1049	53.00
Seasons combined (both halves)	568/913	62.12	682/1407	48.47	1250/2320	53.88

Classifications within cultivars (Table 13) obtained recognition rates ranging from 24.39% to 78.72% and rejection rates ranging from 17.65% to 63.64. Unfortunately, the high recognition rates were accompanied by low rejection rates. The best classification results obtained within cultivars were those of Sandvliet, with an overall classification rate of 63%.

Classifications within seasons (Table 14) obtained recognition rates of 55.79% and 53.20% and rejection rates of 56.98% and 53.68% for the 2002 and 2003 seasons, respectively. The overall classification rates were thus 57.00% and 53.00%. Separate classifications for each half (a- and b-half) of the fruit obtained overall correct classification rates of 56% and 55% for the 2002 season and 58% and 60% for the 2003 season. Classification performed over both seasons resulted in a recognition rate of 62% and a rejection rate of 48%, with an overall correct classification rate of 54%. Figure 32 is a summary of SIMCA classifications made within seasons and within cultivars.

MARS

MARS classifications are shown in Table 15. The best results were obtained for the 2002 season, with correct classification rates of 64% (a-half) and 65% (b-half). For the 2003 season, total correct classification rates of 58% (a-half) and 60% (b-half) were obtained. Each MARS classification model used 30 basis functions, as this number gave the optimal model. These basis functions represent all the information contained in the spectral variables,

The results of the two classification methods were similar. MARS can account for non-linear relationships (Friedman, 1991). A possible reason for similar results is that a linear relationship existed between the spectral and reference data. Figure 33 compares the results of SIMCA and MARS. Throughout the seasons classification results ranged between 52% and 65%. Classifications on separate cultivars and halves of the fruit, in an effort to eliminate some of the variability, did not seem to improve on the results obtained over all the data.

The most important wavelengths used in the MARS classifications were 1077, 1045, 1019, 855, 685 and 769 nm. C-H overtones (carbohydrates) can be assigned to 1077 and 1045 nm (Osborne et al., 1993). These carbohydrates may relate to SSC, but could also be due to pectin, which will support the hypothesis that insoluble pectin accumulation plays a major role in fruit deterioration (Ben-Arie & Lavee, 1971). Chlorophyll can be assigned to 685 nm, indicating that the maturity of the fruit had a role in classifying the post-storage quality.

Feasibility study

Table 16 shows the results of the MARS model developed on the calibration set. These values were used for the subsequent calculations to determine the feasibility of implementing
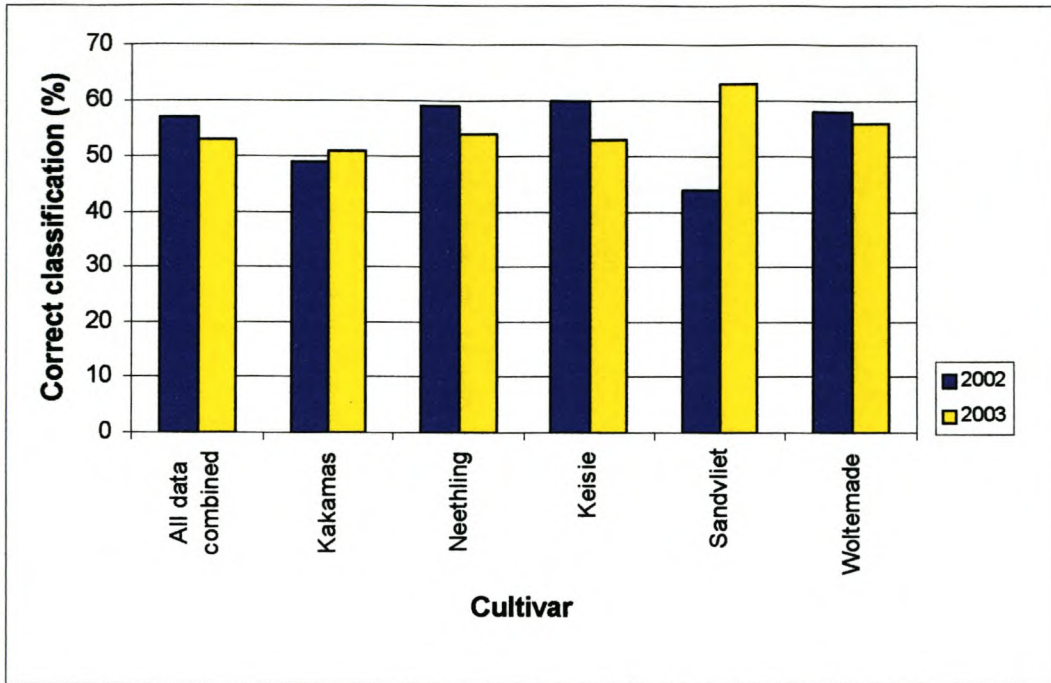
137

Figure 32. Classification results obtained for the various data sets (within seasons and cultivars) using SIMCA as classification method.

Table 15. MARS Classification results performed on data within seasons.

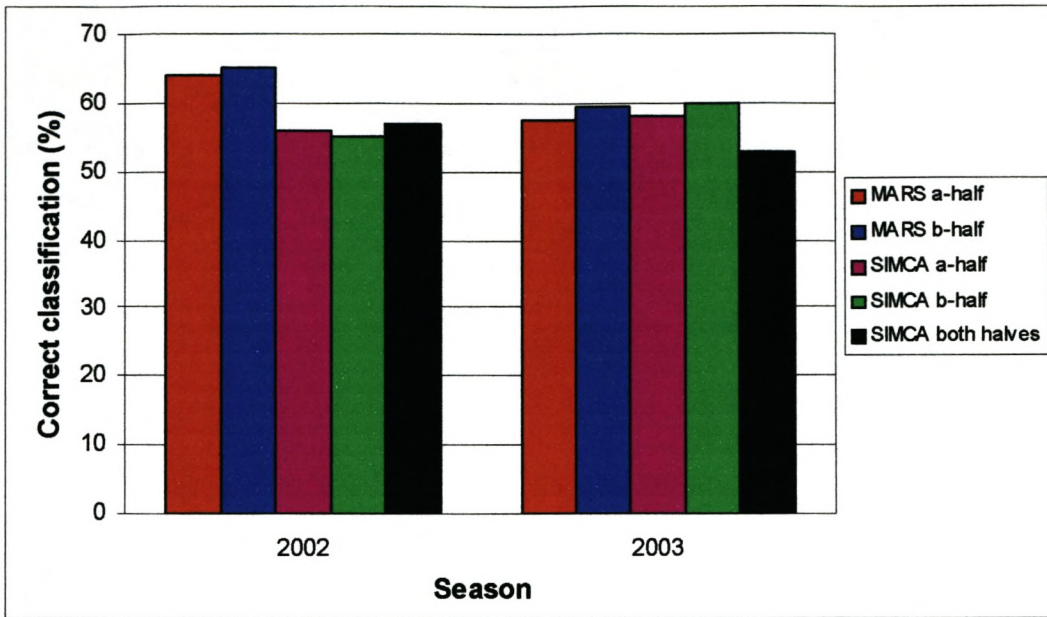| Data set	Recognition rate		Rejection rate		Total classification rate	
	n	%	n	%	n	%
2002 a-side	55/140	39.29	184/233	78.97	239/373	64.08
2002 b-side	57/142	40.14	195/245	79.59	252/387	65.12
2003 a-side	46/85	54.12	67/111	60.36	113/196	57.65
2003 b-side	26/74	35.14	77/99	77.78	103/173	59.54

Figure 33. Differences in the ability of MARS and SIMCA to predict post-storage quality correctly.

Table 16. Results of classification of storage potential on a calibration set (2003 data) using MARS.

	Good post-storage quality (predicted) (n)	Poor post-storage quality (predicted) (n)	Total (n)
Good post-storage quality (measured) (n)	46	39	85
Poor post-storage quality (measured) (n)	44	67	111
Total (n)	90	106	196

The feasibility of the classification method was calculated if 67% (current situation in the industry) of the peaches have to be stored upon arrival at the depot.

Number of fruit to be stored = 0.67 x 196
= 131 fruit

1. If no prior prediction of post-storage quality is made:

   Number of fruit with good post-storage quality = (85/196)131
   = 57 fruit (43%)
   Number of fruit with poor post-storage quality = (111/196)131
   = 74 fruit (57%)

2. If prediction (based on NIRS and MARS) of post-storage quality is made prior to storage:

   Store all peaches predicted as having good post-storage quality as well as a few peaches predicted as having poor post-storage quality to make up 67%. Thus,
   All fruit predicted as good + rest of the fruit predicted as poor = Number of fruit stored
   Thus, 90 + 41 = 131
   Number of fruit which will have good and poor post-storage quality:

   $46_{good} + 44_{poor}$ + (39/106)41 + (67/106)41 = 131

140

$$46_{good} + 44_{poor} \quad + \quad 15_{good} + 26_{poor} \quad = 131$$
$$61_{good} \quad + \quad 70_{poor} \quad = 131$$

Number of fruit with good post-storage quality = 61 fruit (47%)

Number of fruit with poor post-storage quality = 70 fruit (53%)

Thus, the number of fruit lost due to poor post-storage quality:

Without the classification method:	74 fruit	(43%)
Using the NIRS and MARS classification method:	70 fruit	(53%)
Total fruit saved using NIRS and MARS classification:	4 fruit	(3%)

MARS for the peach canning industry. Using the proposed classification method (NIRS and MARS), 195 tons (3%) of peaches can be saved from the total 6500 tons of peaches lost annually. From the total annual losses of R11,05 million, an amount of R331 500 can thus be saved. If 195 tons of peaches is saved, the potential profit that will be gained, amount to R279 240 (195 tons at a profit of R1432 per ton). The savings and potential profit add up to a total income of R610 740.

Conclusion

Near infrared spectroscopy, combined with MARS or SIMCA as chemometric classification techniques, can predict post-storage quality of clingstone peach samples. Both classification methods were suitable for this application. The method can only decrease losses by 3% for the canning industry, but could be feasible on a very large scale. Several recommendations can be made to minimise losses and to obtain the most functionality and value out of this classification method:

- Cultivars with poor post-storage quality, such as Woltemade, should be processed upon arrival at the factory. Cultivars like Sandvliet, Neethling and Keisie show better post-storage quality and can be cold stored before canning, if necessary.

- More research is essential to obtain an understanding of the physical or chemical basis for measuring post-storage quality. Small fruit and fruit with a low firmness were associated with poor post-storage quality. A link was also made between the maturation or ripening stage (change in chlorophyll content) of the fruit and the wavelengths that were most relevant for the classification models. More research on the role of pectin and Ca^{2+} in fruit deterioration, as well as the possibility of NIRS as a means to evaluate these conditions is recommended. A better understanding of the cause of internal fruit deterioration and it's influence on spectral patterns could assist in decision making for future research.

References

Bartley, I.M. & Knee, M. (1982). The chemistry of textural changes in fruit during storage. *Food Chemistry*, **9**, 47-58.

Ben-Arie, R. & Lavee, S. (1971). Pectic changes occurring in Elberta peaches suffering from woolly breakdown. *Phytochemistry*, **10**, 531-538.

Dawson, D.M., Watkins, C.B. & Melten, L.D. (1993). Calcium uptake and efflux, ion leakage, internalair space and cation exchange capacity in relation to mealiness in nectarine tissue. *Postharvest Biology and Technology*, **3**, 131-141.

Friedman, J.H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, **19**, 1-141.

Kawano, S., Watanabe, H. & Iwamoto, M. (1992). Determination of sugar contents in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. *Journal of the Japanese Society for Horticultural Science*, **61**, 445-451.

Kawano, S., Abe, H. & Iwamoto, M. (1995). Development of a calibration equation with temperature compensation for determining the brix value in intact peaches. *Journal of Near Infrared Spectroscopy*, **3**, 211-218.

Lill, R.E., O'Donoghue, E.M. & King, G.A. (1989). Postharvest physiology of peaches and nectarines. *Horticulture Review*, **11**, 413-452.

Lu, R., Guyer, D.E. & Beaudry, M. (2000). Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *Journal of Texture Studies*, **31**, 615-628.

McGlone, V.A. & Kawano, S. (1998). Firmness, dry matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology*, **13**, 131-141.

McGlone, V.A., Abe, H. & Kawano, S. (1997). Kiwifruit firmness by near infrared light scattering. *Journal of Near Infrared Spectroscopy*, **5**, 83-89.

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

Moons, E., Dardenne, P., Dubois, A. & Sindic, M. (1997). Nondestructive visible and NIR spectroscopy for determination of internal quality in apple. In: *Proceedings from the Sensors for Nondestructive Testing International Conference and Tour*. Pp. 122-132. February, 1997. Orlando, Florida.

Morris, L.L. (1982), Chilling injury of horticultural crops: An overview. *HortScience*, **17**, 161-162.

Murray, I. (1999). NIR spectroscopy of food: Simple things, subtle things and spectra. *NIR news*, 10(2), 10-12.

Næs, T. & Martens, H. (1984). Multivariate calibration. II. Chemometric methods. *Trends in Analytical Chemistry*, **3**, 266-271.

Ortiz, C., Barreiro, P., Correa, E., Riquelme, F. & Ruiz-Altisent, M. (2001) Non-destructive identification of woolly peaches using impact response and near-infrared spectroscopy. *Journal of Agricultural Engineering Research*, **78**, 281-289.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis* (Edited by D. Browning), 2nd ed. Pp. 1-220. Harlow: Longman Scientific and Technical.

Peiris, K.H.S., Dull, G.G., Leffler, R.G. & Kays, S.J. (1997). Nondestructive determination of soluble solids content of peach by near infrared spectroscopy. In: *Proceedings from the Sensors for Non-destructive Testing International Conference and Tour*. February, 1997. Orlando, Florida.

Peiris, K.H.S., Dull, G.G., Leffler, R.G. & Kays, S.J. (1998). Near infrared spectrometric method for nondestructive determination of soluble solids contents of peaches. *Journal of American Society of Horticulture Science*, **123**, 898-905.

Peirs, A., Tirry, J., Verlinden, B., Darius, P. & Nicolai, B.M. (2003). Effect of biological variability of NIR models for soluble solids content of apples. *Postharvest Biology and Technology*, **28**, 269-280.

Sjöström, M., Wold, S., Lindberg, W., Persson, J. & Martens, H. (1983). A multivariate calibration problem in analytical chemistry solved by partial least squares models in latent variables. *Analytica Chimica Acta*, **150**, 61-70.

Slaughter, D.C. (1995). Nondestructive determination of internal quality in peaches and nectarines. *American Society of Agricultural Engineers*, **38**, 617-623.

Schmilovitch, Z., Mizrach, A., Hoffman, A., Egozi, H. & Fuchs, Y. (2000). Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology*, **19**, 245-252.

Steyn, B. (2003). Tiger Brands: Fruit unit, Ashton. Personal communication.

Van Zyl, A. (2000). The application of Fourier transform near infrared (FT-NIR) spectroscopy in the wine, fruit and dried fruit industries of South Africa. MSc in Food Science Thesis, University of Stellenbosch, South Africa.

Victor, W. (2002). Canned Fruit Producers' Association. Personal communication.

Williams P.C. & Norris, K.H. (1987). Qualitative applications of near-infrared reflectance spectroscopy. In: *Near Infrared Technology in the Agricultural and Food Industries* (Edited by P.C. Williams & K.H. Norris) American Association of Cereal Chemistry Inc.: St. Paul, MN, USA.

Wills, R.B.H., Mc Glasson, W.B., Graham, D., Lee, T.H. & Hall, E.G. (1989). Physiology and biochemistry of fruit and vegetables. In: *Postharvest: An Introduction to the Physiology and Handling of Fruits and Vegetables*. Pp. 34-37. Oxford: BSP Professional Books.

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognition*, **8**, 127-139.

Chapter 5

General discussion and conclusions

This study investigated the possibility of non-destructively predicting the post-storage quality of intact Bulida apricots and clingstone peach cultivars for the South African canning industry. The short harvesting period and limited capacity of the factory to process the large amount of fruit directly after harvesting, necessitates cold storage until the fruit can be canned (Victor, 2002). Fruit deterioration during cold storage causes a large portion of the fruit to be unsuitable for canning and leads to economic losses. The basis for measurement of storage potential is unclear, as this attribute cannot be measured quantitatively. It can only be identified subjectively after storage, when the fruit is destoned.

Near infrared spectroscopy (NIRS) allows for simple, rapid and non-destructive analysis (Osborne *et al.*, 1993). Furthermore, NIRS allows for simultaneous analysis of several constituents of each sample for the same time and cost as a single determination (Batten, 1998). The advantages that NIRS offers over conventional chemical methods led to the idea of implementing NIRS as a non-destructive tool to predict post-storage quality of the fruit for the canning industry.

Near infrared spectroscopic calibrations to predict soluble solids content (SSC) in intact Bulida apricots and clingstone peaches were successful, confirming sufficient light penetration into the intact fruit. Accurate predictions were made for peach samples in spite of the two and three week periods between spectral acquisition and the SSC measurements. Less accurate predictions of SSC were made for apricots. Firmness calibrations in peach samples were less successful than SSC calibrations and more research is required to investigate the basis for firmness measurements. The accuracy of the calibrations was dependent on the accuracy of the reference values, which are by their nature not as precise as would be desirable. Variation within the fruit, as well as the occurrence of soft and disintegrated fruit after storage, made the reference measurements less accurate.

Several factors (fruit firmness, SSC, diameter, cultivar and weight) contributing to fruit quality were investigated by means of a classification tree to find a possible explanation or link to the deterioration of fruit quality during cold storage. Large ($\geq$ 57mm) and firm (pressure readings $\geq$ 4.7kg) fruit displayed good post-storage quality whilst small and soft fruit displayed poor post-storage quality. The non-destructive measurement of fruit firmness by NIRS was, however, not accurate enough to make these predictions of post-storage quality.

Near infrared spectroscopy (NIRS) was used in combination with chemometric techniques to classify samples according to their post-storage quality. Classification models were developed using soft independent modelling by class analogy (SIMCA) and multivariate adaptive regression splines (MARS). The models were developed by combining pre-storage NIR spectral data and the corresponding post-storage subjective quality evaluations. This classification method provides an objective classification method and improves on the current situation in the fruit processing industry, where no classification method exists and samples are stored randomly. Savings of 3% and 4% was calculated for the apricot and peach canning industries, respectively. Whether the savings will justify the costs involved in implementing and maintaining such a classification system in the long run, is not yet evident.

The ability of these models to predict storage potential to a certain extent indicates a further application of NIRS in the food industry. By including all the wavelengths of the range measured on the sample and investigating the entire spectral matrix, additional and previously hidden information about the samples could be obtained. Multivariate processing of the complex data provided a way to extract, express and interpret useful information from the sample set.

Recommendations

Fruit should, ideally, be processed as soon as possible – as this is reflected in industry practice. However, a recommendation to the industry that could reduce product losses would be to prioritise the canning of cultivars with poor storage quality, such as Woltemade, upon arrival at the factory. Cultivars like Sandvliet, Neethling and Keisie show better storage potential and can be cold stored before canning, if necessary.

Potential for decreased deterioration and improved storage quality exist if storage conditions can be improved significantly. Additional methods to improve storage conditions and limit fruit deterioration should be investigated.

A wider wavelength range could be included for improvement of the models developed. A wavelength range of 645-1201 nm was measured in this study. Along with the wavelengths in the NIR range, those in the visible range were also significant in the development of regression and classification models. It is suggested that spectral measurements from 400-2600 nm may reveal more hidden information relating to fruit quality.

More research is required to understand the physical or chemical basis for measuring post-storage quality. A better understanding of the internal changes associated with fruit deterioration during storage and its cause could assist in decision making to improve prediction models.

147

More factors could be included in future prediction models. In addition to the NIR spectra, one could also include external factors, such as rainfall, area of cultivation, soil conditions, the use of nitrogen, pesticides and herbicides and general farming practices. These factors have a large influence on the overall quality of the fruit (Victor, 2002) and could also affect the storage potential. Although a complex model should be avoided if possible, the addition of these factors could lead to a better understanding of the contributing factors to post-storage quality. This knowledge could assist in decision making and subsequently lead to improved prediction models.

The large or industrial scale cost effectiveness of implementing the classification system described above should be investigated, as the potential exists for reduced product losses through successful prediction of post-storage quality using the methodology developed in this study.

References

Batten, G.D. (1998). Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Australian Journal of Experimental Agriculture*, **38**, 697-706.

Osborne, B.G., Fearn, T. & Hindle, P.H. (1993). *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis* (Edited by D. Browning), 2nd Ed. Pp. 1-220. Harlow: Longman Scientific and Technical.

Victor, W. (2002). Canned Fruit Producers' Association. Personal communication.