Investigating Aquaphotomics for Fruit Quality Assessment

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Electronics Engineering

at

The University of Waikato

by

HARPREET KAUR

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2020

Abstract

The methods of aquaphotomics were explored for internal fruit quality assessment, such as dry matter (DM) and soluble solid content (SSC) measurement using near infrared spectroscopy (NIRS). Fruits like apple and kiwifruit are more than 80% water. The NIR spectrum of whole intact fruit is dominated by water absorption peaks that shift and vary in shape with changes in quantities such as SSC, DM, and temperature. These variations can significantly reduce predictive model performance for fruit quality parameters. Because the absorption peak at 1450 nm (first overtone of the OH stretch of water) and 970 nm (second overtone of the OH stretch of water) in the samples varies with SSC, DM concentration, and temperature, an investigation was undertaken on the aquaphotomics approach to quantify changes in the water structure that are apparent in the two wavelength regions.

The key question examined was whether aquaphotomics can help in the development of better and simpler prediction models or hardware for intact fruit quality measurements beneficial to the industry. To find answers to this question, several experiments were conducted. Much work in aquaphotomics has been done in the longer wavelength region, *that is,* at the first overtone (1300–1600 nm), which requires short pathlength cells due to higher water absorption. Experiments were performed on simple

aqueous systems such as sucrose solution of various concentrations and an aqueous system such as apple and kiwifruit juice. The second overtone wavelength region (800–1100 nm) was also scrutinized for fruit juice measurements using a long pathlength cell. Eventually, experiments were conducted on whole intact fruit.

A Fourier transform near infrared (FT-NIR) spectrophotometer was used to acquire spectra of aqueous samples over the wavelength range of 900 to 1800 nm. For whole, intact fruit spectra collection, spectrophotometers in the wavelength range below 1100 nm were utilized. An in-house benchtop spectrometer based on the Zeiss MMS-1 NIR spectrometer operated in interactance mode was used along with some commercially available handheld spectrophotometers such as the F-750 produce quality spectrometer and the SCiO spectrometer. The spectra acquired with the instruments were analyzed using multivariate data analysis. The calibration models were developed using partial least square regression (PLSR) and multiple linear regression to estimate quality attributes.

For a simple aqueous system at a constant temperature, water bands were identified in the first overtone region that can explain water structure according to SSC concentration. An anharmonic oscillator model was used to calculate those wavelengths corresponding to the wavelengths in the first overtone region. It was found that for generating optimal prediction models and meaningful aquagrams (water spectral patterns) in the first and second

overtone regions, sample path length plays an important role. The PLSR model for SSC built in the second overtone region with a 10 mm path length cuvette gave a standard error of prediction (SEP) of 0.11% which was comparable to first overtone prediction results using a 1 mm cell. This confirms that longer path-lengths are essential for the short wavelength second overtone region and hence, for solid samples such as kiwifruit and apple. The PLSR model built using F-750 instrument in the 800–1000 nm region (second overtone) gave better results with a RMSECV of 0.30% compared to 0.56% in the 600–1100 nm range signifying that the important information resides in a tight window around the second overtone region.

The NIR spectrum of fruit is affected by temperature. Therefore, a calibration equation developed at one fixed temperature cannot reliably predict samples at a different temperature. Using the framework of aquaphotomics, the changes within the water structure of fruit juice and whole intact fruit occurring with changes in temperature were analysed. Water wavelengths affected by temperature variation were identified in the first and second overtone regions. Using aquagrams, it appeared that different water species were activated in the first and second overtone region. The influence of increasing temperature on the peak absorbance of the juice was a lateral shift in the first overtone region, whereas it was vertical in the second overtone region of water. Two correction pre-treatments; extended multiplicative scatter correction (EMSC) and external parameter orthogonalization (EPO) were used in conjunction with principal

component loading of water as an interferent spectrum. They were able to correct for temperature effects by at least a factor of 10 to insignificant levels.

It was concluded that the aquaphotomics paradigm offers fundamental insights into the role of the various light absorption mechanisms but only in conjunction with chemometrics tools, like EMSC and EPO, leading to an improvement in predictive model performance. The investigated methods have good potential to be used for quality assessment of fruit in industry.

Acknowledgements

First of all, I would like to dearly thank my supervisors, Dr. Rainer Künnemeyer, Dr. Andrew McGlone, and Dr. Michael Cree, for their constant guidance and support. None of this work would have been possible without their help and patience. From them, I learned to think skeptically and logically. We always followed a pragmatic approach to understand "Aquaphotomics" and did an in-depth investigation.

Besides my supervisors, I would like to thank Richard Seelye to teach me how to do destructive measurements on fruit. I would also like to thank Alona Finn for her support during the spectroscopy work. We had great discussions around to improve the performance of the models we developed. Thanks to Jason Sun for helping me in the experiments. Also, many thanks to Mark Wang for preparing the chilling injury fruit and helping me in the experimental work.

I would also like to thank the New Zealand Institute for Plant and Food Research and the University of Waikato for their financial support. My time was mainly spent at the Plant and Food Research Ruakura campus. I want to particularly thank the applied sensors-Ping Pong Rangers team for keeping me happy and motivated. I would definitely say that my table tennis skills have improved. Also, special thanks to my Toastmasters club especially Annie Muggeridge, Monica Peters, Nathan Tomer, Jason Sun to

help me to improve my three minute thesis (3MT) script and speech. Only with their guidance and support, I won the overall and people's choice award at the University of Waikato and COSMOS magazine Editor's Choice award in the Asia-Pacific Level at the University of Queensland, Australia.

I would like to give special thanks to my parents, Rajinder Singh and Jagdish Kaur, who supported me during my years of PhD study. It was really tough for me to stay away from them. They are the ones who encouraged me to reach for my goals and aim high. I certainly owe all my accomplishments to them. My mother always motivated me and also taught me cooking over the phone. Also, my brother Dr. Harmeet Singh who is a role model to me. I missed the fun times we had in India while watching History Channel, Discovery Channel, Sci-Fi, and inspirational movies together. Thanks to my sister-in-law, Geet as well for the support. Thanks to my maternal grandfather- Darji (late). I miss you so much.

I would like to thank my friends Rani, Rasshi, Sadhana, and Sandhya for their care, love, and support during my PhD studies. Also, spirituality helped me to stay calm and positive during the tough times. Sister Shivani, Manju, Niti, and Anita were my motivation. I would specifically like to thank Arvind for coming into my life. He always supported me to fulfil my dreams.

Last but not least, I would like to express my gratitude to all the teachers from my high school and college who have inspired me to pursue a career in science and engineering.

List of abbreviations

NIRS	Near Infrared Spectroscopy
PLSR	Partial Least square Regression
MLR	Multiple Linear Regression
PCA	Principal Component Analysis
WAMACS	Water Matrix Coordinates
WASP	Water Spectral Pattern
MSC	Multiple Scatter Correction
SNV	Signal Normal Variate
DM	Dry Matter
SSC	Soluble Solids Content
RMSECV	Root Mean Square Error of Cross-Validation
RMSEP	Root Mean Square Error of Prediction
SEP	Standard Error of Prediction
EPO	External Parameter Orthogonalisation
EMSC	Extended Multiplicative Scatter Correction
PC	Principal Component
CI	Chilling Injury
SNR	Signal to noise ratio

Table of Contents

Chapter 1: Introduction

This chapter outlines the introduction to the problem and a brief description of each chapter.

1.1 Introduction to the problem

Good quality fruits are always preferred for the superior taste experience and increasingly for their health benefits and their role in helping prevent many chronic diseases (Hall et al., 2009; Boeing et al., 2012). Fruit quality attributes are important considerations in terms of marketing, transportation, and storage requirements (Fu et al., 2009), especially for a country like New Zealand whose income from its horticultural sector has exceeded $9.2billion in the year 2018 ("Fresh Facts ", 2018). Therefore, fruit quality is essential for New Zealand's fruit industry, and emphasis has been placed on methods or techniques to improve quality assessment of fruit, in particular, kiwifruit. Quality components of fruit are external appearance, such as color, shape, external defects, etc., and internal parameters, such as sugar content, acid content, dry matter (DM), firmness, maturity, internal breakdowns, etc. (Nicolaï et al., 2007; Cubero et al., 2011).

Recent studies have shown that consumer acceptability could be further enhanced by improving the taste of fruit where DM is as a good indicator (McGlone & Kawano, 1998; Jordan & Seelye, 2009; Wang et al., 2015). DM remains fairly constant during ripening with only very small losses due to transpiration, starch conversion and respiration It is useful because the fruit DM is dominated by the large carbohydrate component

(around 75% of DM) that is sugar and starch at harvest which almost completely converts to sugar when the fruit ripens (McGlone & Kawano, 1998). Hence, DM indicates either the potential or actual sugar level of the fruit. The standard method to measure DM is drying slices of fruit at 65 °C for 24 h to eliminate water which is a destructive technique (Kaur et al., 2017b). Measuring soluble solids content (SSC; measured as °Brix) of kiwifruit is the authoritative method of assessing maturity at harvest time in most kiwifruit-producing countries, such as New Zealand, Italy, France, Greece, Chile, Japan and the United States (Crisosto et al., 2012). The conventional analytical techniques to measure SSC and DM are destructive, time-consuming, and labor intensive.

Several researchers have shown that a non-destructive technique, such as Near Infrared Spectroscopy (NIRS), can be useful in predicting the SSC and DM of different fruit (Kawano et al., 1992; Lu et al., 2006). Non-destructive methods based on NIRS are well established in commercial fruit grading operations for segregating fruit populations (Wang et al., 2015). By contrast, NIRS systems for in-orchard and on-tree use, where the technology might be used to make specific harvest selection of fruits or for optimizing harvest time, is still in its infancy and not in widespread commercial use. Specifically designed, commercial NIR spectrometers for in-orchard use exist. These are hand-held, portable instruments, like the NIRVANA (Integrated Spectronics, Baulkham Hills, Australia; recently discontinued), the F-750 Produce Quality Meter (Felix Instruments,

Portland, USA), the SCiO spectrometer version 1.1 (Consumer Physics, Hod HaSharon, Israel), and the NIR portable spectrometer H-100C (Sunforest, Incheon, Korea) (Fig. 1.1). The spectrophotometers used for in-orchard measurement as well as those used for commercial graders work in the visible to short wavelength (SW)-NIR region spanning 300-1100 nm (Shafie et al., 2015; Kaur et al., 2017b).

Figure 1.1. Commercially manufactured hand-held, portable instruments: (a) NIRVANA (Integrated Spectronics, Baulkham Hills, Australia; discontinued), (b) F-750 Produce Quality Meter (Felix Instruments, Portland, USA), (c) SCiO spectrometer (Consumer Physics, Hod HaSharon, Israel) and (d) H-100C NIR portable spectrometer (Sunforest, Incheon, Korea) (Kaur et al., 2017b)

Since fruit is about 80% water, this is the dominant chemical constituent of fruit. Moreover, water highly absorbs near infrared radiation. The near infrared spectrum of fruit shows a strong absorption peak around 970 nm, which corresponds to the second overtone of the OH stretch in water, Fig. 1.2 (McGlone et al., 2003; Nicolaï et al., 2007). Many multivariate models for predicting DM and SSC were developed from NIR interactance measurements using the narrow spectral range from 800 to 1100 nm around this absorption peak of water (Workman, 2000). Both DM and SSC was predicted with very good accuracy, greater than 0.39% root mean square error of prediction (RMSEP) (McGlone et al., 2003).

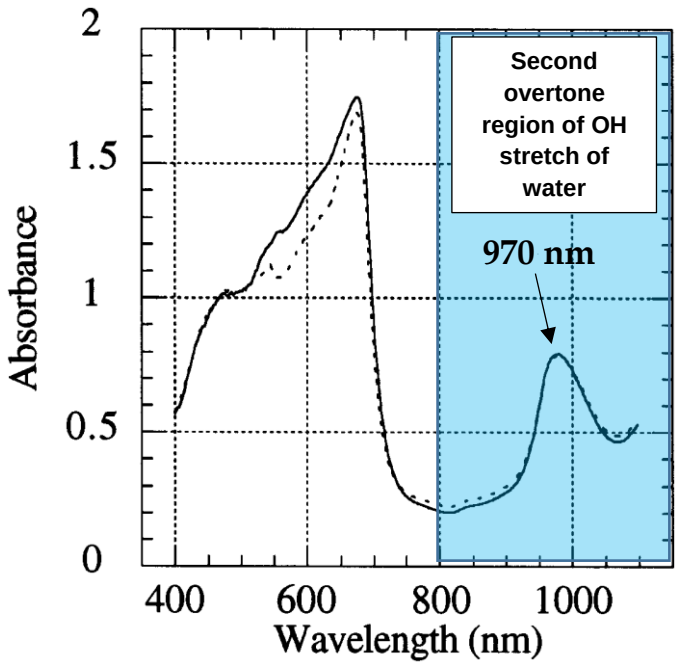


Figure 1.2. Typical absorbance spectra for kiwifruit. The solid and dashed curves are for firm and soft fruit, respectively (McGlone et al., 2003)

One research group suggests that water is a good indicator of other compounds in the matrix and stated "The overtone bands of water

1300 – 1600 nm, in a scientific discipline known as aquaphotomics have been shown to be a good discriminator and non-destructive indicator of changes in biological aqueous samples for disease diagnosis and protein conformation"(Chatani et al., 2014). This field has aided the understanding of the role of water in biological systems (Tsenkova, 2006; Tsenkova, 2007; Tsenkova, 2008a; Tsenkova, 2008b; Meilina et al., 2009; Tsenkova, 2010; Tsenkova et al., 2015; Tsenkova et al., 2018). Aquaphotomics is a scientific discipline that utilizes NIRS in which the spectral analysis focuses on absorbance patterns related to water bands and the effect of perturbations due to variation in temperature, the concentration of solutes, environment, etc. As the water absorption peak at 1450 nm (first overtone of OH stretch of water) and 970 nm (second overtone of OH stretch of water) in samples varies due to SSC or DM concentration, the aquaphotomics approach was investigated to learn more about changes in the water structure that are apparent in the 1300-1600 nm and 800-1100 nm wavelength regions.

Furthermore, this thesis examines whether aquaphotomics can help in the development of better/simpler prediction models or hardware for intact fruit quality measurement beneficial for the fruit industry. The key question examined is whether water absorptions in the short NIR range (< 1100 nm) can be used to predict fruit quality parameters. To find answers to this question, several experiments were conducted on fruit. Much work in aquaphotomics has been done in the longer wavelength region, *i.e.,* in the first overtone (1300-1600 nm), which requires short pathlength cells due to

the high absorption of water. Therefore, experiments were performed on simple aqueous systems such as sucrose solution of various concentrations and an aqueous system such as apple and kiwifruit juice.

Moreover, aquaphotomics has been investigated to find whether robust calibration models can be built in varying temperature environment. The second overtone wavelength region (800-1100 nm) was also scrutinized for fruit juice measurements using a long pathlength cell. Eventually, experiments were conducted on whole intact fruit.

In addition to the main body of work investigating aquaphotomics for fruit quality assessment, a small study was carried out to compare several commercially available spectrophotometers for fruit quality assessment. This study was published in the Journal of Near Infrared Spectroscopy; further details can be found in Appendix.

1.2 Thesis organization

This thesis is divided into nine chapters and appendices, as shown in Fig. 1.3. Chapter 2 gives an overview of the relevant literature and theory of "Aquaphotomics" and presents the techniques used for temperature correction of prediction models that are extended multiplicative signal correction (EMSC) and external parameter orthogonalisation (EPO).

Chapter 3 describes the use of aquaphotomics concepts on a simple system such as an aqueous solution of sucrose at various concentrations.

The aquaphotomics analysis was applied to both the first overtone (1300 - 1600 nm) and the second overtone (870 - 1100 nm) of OH stretch of water using an FT-NIR spectrophotometer. Anharmonic oscillator model was used to find water bands in the second overtone region. The effect of sample pathlength on the interpretation ability of aquaphotomics is discussed in this chapter. Modified aquagrams (a visualization tool for aquaphotomics) are presented.

Chapter 4 deals with the implementation of aquaphotomics for assessing the quality of apple juice in the first overtone region at different temperatures. Calibration models are developed using partial least square regression (PLSR) to see whether aquaphotomics can help in building better calibration models. Also, a temperature correction technique, EMSC, has been utilized to reduce bias when predicting a sample different in temperature to that used for the calibration model.

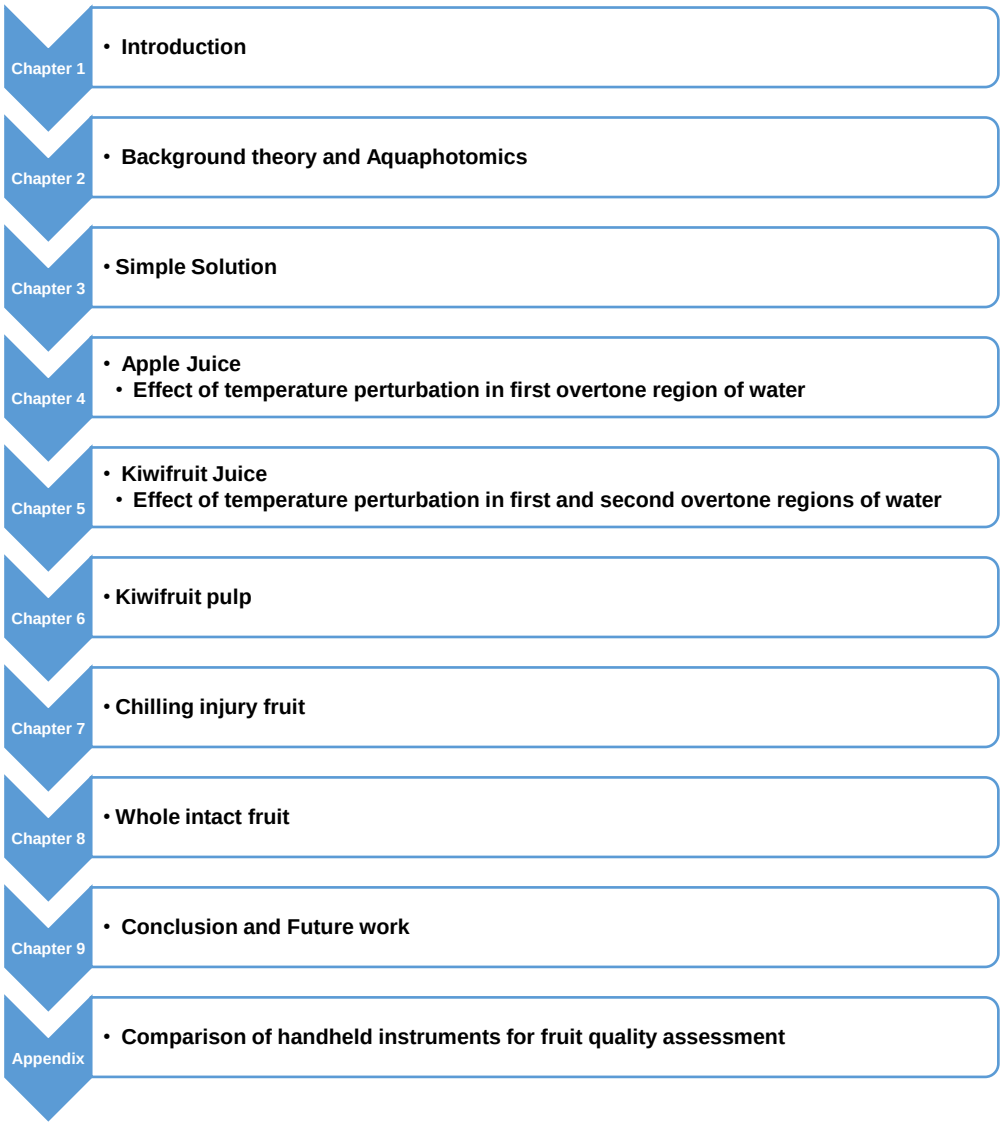


Figure 1.3 Structure of the dissertation

Chapter 5 encompasses experiments on kiwifruit juice at three different temperatures performed in the two overtone regions of the OH stretch of water. The isosbestic point was observed, and the effect of sample pathlength on the development of the calibration model for quality assessment was explored. Using EPO pre-processing, temperature correction was performed using the PC1 loading of the difference matrix

obtained after subtracting the mean of water spectra at different temperatures from the lowest temperature mean spectrum of water.

Experiments were performed on kiwifruit pulp to investigate the effect of starch on the performance ability of the calibration model which is discussed in chapter 6. Aquagrams were able to show how the water spectral pattern changes for high and low starch samples.

Chapter 7 includes work on chilling injury (CI) fruit. The methods of aquaphotomics were explored for the detection of CI in juice and whole intact fruit. Absorption of near infrared radiation by intact and CI fruit is discussed.

Chapter 8 reports the results of fruit analysis experiments, where the emphasis is placed on experiments conducted using the benchtop spectrometer and Felix instrument. The experiments were conducted at different temperatures, and water spectra were measured to correct for temperature.

Chapter 9 summarises the research work conducted and gives some proposals for future study.

1.3 Contributions

The work presented in this thesis has been carried out at the Bio-engineering Group of the New Zealand Institute for Plant and Food Research, Ruakura. The author also acknowledges the financial support of

a Ph.D. scholarship from The University of Waikato, New Zealand, and MBIE Contracts C11X1208 & C11X1601 with Plant & Food Research, New Zealand.

The main original contributions from this research work are:

- The interference term for EMSC/EPO, the PC loading generated from the pure water-temperature spectral matrix, was used to build a temperature-independent model;
- An an-harmonic oscillator model was used to establish 12 water bands in the second overtone region that corresponded to reported wavelengths in the first overtone region of water;
- The effect of sample pathlength in different overtone regions of OH stretch of water to interpret aquaphotomics ability was investigated;
- Aquaphotomics was investigated to predict the quality of apples and kiwifruit in the second overtone region of water.

The following peer-reviewed journal papers have arisen from this work:

1. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2017) *"Comparison of hand-held near infrared spectrophotometers for fruit dry matter assessment"*, Journal of Near Infrared Spectroscopy, Vol. 25, No. 4, Pages 267-277.
2. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2020) *"Investigating Aquaphotomics for Temperature Independent Prediction of*

Soluble Solids Content of Pure Apple Juice", Journal of Near Infrared Spectroscopy. Journal of Near Infrared Spectroscopy, Vol. 28, No. 2, Pages 103-112.

In addition, the following oral and poster presentations have arisen from this work:

1. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone, & Michael Cree (2016). *Performance comparison of various NIR spectrometer units for fruit quality prediction*. IEEE Instrument & Measurement Workshop 2016. Conference held at Lincoln University, New Zealand (oral presentation).
2. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2017). *Aquaphotomics for Brix Prediction of Apple Juice using FT-NIR Spectrophotometer.* The Waikato Young Research Engineers Symposium (WYRES) held at the University of Waikato, Hamilton, New Zealand (oral presentation).
3. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2017). *Investigating Aquaphotomics for Fruit Quality Assessment using Portable Handheld Spectrophotometers.* IEEE Instrumentation & Measurement Society, New Zealand Chapter Two-Day Workshop on Smart Sensors, Measurements and Instrumentation for Health, Industry, and Environment. Victoria University of Wellington, Kelburn, Wellington, New Zealand (oral presentation).

4. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2017). Assessment of Apple Juice by Aquaphotomics in a Temperature Dependent Environment. Presented at The 3rd Australia New Zealand Conference on Optics and Photonics (ANZCOP), Queenstown, New Zealand (poster presentation).
5. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2018). *Can Aquaphotomics Improve Quality Prediction of Intact Fruit?* In the 11th Annual Dodd-Walls Centre Symposium, The University of Auckland, New Zealand (oral presentation).
6. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2018). *Aquaphotomics for Fruit Quality Assessment.* In the 18th ANISG/NZNIRSS Conference at Rotorua, 9-12th April 2018 (oral presentation).
7. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2018). *Investigating Aquaphotomics for Fruit Quality Assessment.* The 3rd Aquaphotomics International Symposium - Exploring Water Molecular Systems in Nature. Awaji, Hyogo, Japan (oral presentation)**.**
8. **Harpreet Kaur,** Rainer Künnemeyer, Andrew McGlone (2019). *"Temperature independent quality prediction of kiwifruit juice using aquaphotomics and external parameter orthogonalization"* 19th International Council of Near Infrared Spectroscopy Conference 2019 held at the Gold Coast Convention & Exhibition Centre

(GCCEC), Australia, on the 15th – 20th September 2019 (oral presentation).

Recognition

1. **Three-minute thesis: 3MT** people's choice and overall winner prize at The University of Waikato, New Zealand in August 2017 for the presentation on the Ph.D. thesis work titled "*Light weight and light-based fruit quality sensing*."
2. **Best student presentation award** for the oral presentation on "*Aquaphotomics for Fruit Quality Assessment*" in the 18th ANISG/NZNIRSS Conference at Rotorua, New Zealand, 9-12th April 2018.
3. **3MT Asia Pacific semi-finalist 2017** and **Editor's choice award winner at** The University of Queensland, Australia (September 2017). The presentation was published in the COSMOS scientific magazine.

Chapter 2: Background Information and Aquaphotomics

This chapter provides an overview of fruit quality, near infrared spectroscopy, different data analysis techniques, aquaphotomics, and the effect of sample temperature on the development of a calibration model.

2.1 Fruit Quality: DM and SSC

In this thesis, the term 'fruit quality' refers to characteristics of the fruit that determine consumer acceptability in terms of the fruit's taste. In other words, it means "How good does the fruit taste to the consumer?". There are several aspects of fruit quality that together determine the overall consumer acceptance of a particular fruit. These attributes fall into two basic classes: one that includes factors that appeal to our sense of 'touch' and the other that contains the factors that appeal to our sense of 'taste' (Fillion & Kilcast, 2002). The first group includes fruit firmness, crispness, juiciness, and crunchiness and therefore, can be referred to as fruit texture. However, the second group includes traits such as fruit sweetness, acidity, and flavour, which appeal to our sense of taste.

With the rapid development of the global economy and improvement in living standards over the last couple of decades, fruit consumption has increased significantly. However, consumers now have higher expectations of obtaining good quality fruit. The research in this thesis predominantly focussed on kiwifruit and apples. Consequently, the work concentrated on the measurement of two closely related fruit quality

attributes; dry matter (DM) and soluble solid content (SSC; measured as °Brix). DM is an important quality attribute linked to taste across a broad range of fruits. For climacteric fruit that store starch during fruit growth, and are harvested mature but unripe, the DM remains fairly constant during ripening with only small changes due to transpiration, starch conversion and respiration. Fruit DM is dominated by the significant carbohydrate component (around 75% of DM for kiwifruit, for example) that is sugar and starch at harvest (McGlone & Kawano, 1998). A major characteristic of kiwifruit ripening is the increase in sugars that occurs as insoluble starch is converted into soluble solids, of which the simple sugars glucose (2–6%), fructose (1.5–8%), and sucrose (2%) provide the bulk (McGlone & Kawano, 1998). It is the soluble sugars that ultimately drive consumer preferences for sweet-tasting fruits. Hence, DM indicates either the potential or the actual sugar content, and thus a dominant part of the taste characteristics of the fruit (McGlone & Kawano, 1998).

2.2 Fruit Quality Assessment

2.2.1 Destructive Methods for DM and SSC measurement

DM and SSC measurements on the fruit are usually made by cutting the fruits, which is a destructive method of analysis. Brix readings are considered equivalent to the soluble solids in the fruit and are generally measured with a handheld refractometer, using juice expressed from two

thin slices taken from each side of the fruit. The readings are recorded and then averaged to get the mean Brix value. For DM, the measurements involve cutting two thin slices from approximately 3 mm thick slices of the fruit removed from either side of the fruit (in the case of apple) or one slice taken from the middle (for kiwifruit) and drying them in an oven for 24 hours at 65°C (McGlone & Kawano, 1998). These methods for measuring fruit quality parameters are destructive, time-consuming, and labour-intensive. Hence, there is a need to develop non-destructive techniques that will save time, money, and labour.

2.2.2 Non-destructive Methods for DM and SSC measurement

Researchers worldwide have investigated various methods for quality assessment of fruit, including acoustic techniques, spectroscopic techniques, machine vision, and electronic noses (Wang et al., 2015). Among these technologies, near infrared spectroscopy (NIRS) can be considered the most researched non-destructive technique for the assessment of internal fruit quality (Fu et al., 2009; Moghimi et al., 2010; Chen & Han, 2012; Kumar et al., 2015). NIRS analysis is non-destructive and produces no chemical waste. NIRS was first used in agricultural applications by Norris (1964) to measure moisture in grain. Since then it has been used for rapid analysis of mainly moisture, protein, and fat content of a wide variety of agricultural and food products (Davies & Grant, 1987; Gunasekaran & Irudayaraj,

2000). Early applications in horticulture focussed on DM of onions (Renfroe & Kays, 1985), SSC of apples (Bellon-Maurel, 1992) and water content of mushrooms (Roy et al., 1993), but since then many other applications have followed. As the propagation of NIR radiation in fruit and vegetable tissue is affected by their microstructure, it was soon discovered that NIRS could also be used to measure microstructure-related attributes, such as stiffness (Lammertyn et al., 1998), and internal damage (Clark et al., 2003a).

Since the spectra of the fruit contain information related to chemical composition, NIRS methods are successfully set up and commercialized for fruit grading applications, especially for SSC, DM, acidity, and firmness (Ragni et al., 2012). NIRS has been used as a rapid and non-destructive technique on a range of commodities; for example, for SSC determination of various fruits and vegetables including apples (Lammertyn et al., 2000b; McGlone et al., 2003), kiwifruit (McGlone & Kawano, 1998; Slaughter & Crisosto, 1998; McGlone et al., 2002c), peaches and nectarines (Valero et al., 2007; Kaur et al., 2017b), satsuma mandarins (Gómez et al., 2006), and tomatoes (Slaughter et al., 1996). NIR calibrations have also been developed for moisture, DM and total solids content (TSC) in some intact commodities including dates (Schmilovitch et al., 1999), kiwifruit (DC Slaughter, 1998), mangoes (Walsh et al., 2004; Subedi et al., 2007), and onions (Lohumi et al., 2014; Jantra et al., 2017).

2.3 Near infrared spectroscopy

A tribute to Karl Norris: Father of NIR spectroscopy

It is extremely sad that the founder and "the father of NIR spectroscopy": ***Karl Norris*** (Fig. 2.1) passed away on July 17, 2019, at the age of 98. Through this thesis work on NIRS and aquaphotomics, I would like to pay tribute and owe so much to Karl Norris. Karl Norris was a strong advocate for understanding the chemical and physical mechanisms that underpin the successful use of NIRS. He was no fan of the blind approach common in modern application of NIRS, where spectra are measured and then predictive models are made via automated computer algorithms that require no user input or expert knowledge. I imagine the subject of aquaphotomics, in attempting to understand fundamental mechanisms, would enjoy his favour.

Figure 2.1 Karl Norris at Beltsville with his Cary 14 in the background, photographed in 1982 (Davies, 2011)

2.3.1 Fundamentals of near infrared spectroscopy

NIRS is a vibrational spectroscopy method in the near infrared region of 780–2500 nm where overtones or combinations of the fundamental stretching bands respond (Fig. 2.2). The NIR region is divided into the short-wavelength range (SWNIR) from 700 to 1100 nm (sometimes called "the Herschel region") and the long-wavelength NIR range from 1100 nm to 2500 nm (LWNIR) (Wang et al., 2015). The SWNIR region is considered

to be the absorption band of higher-order overtones while the LWNIR belongs to the first or second overtones.

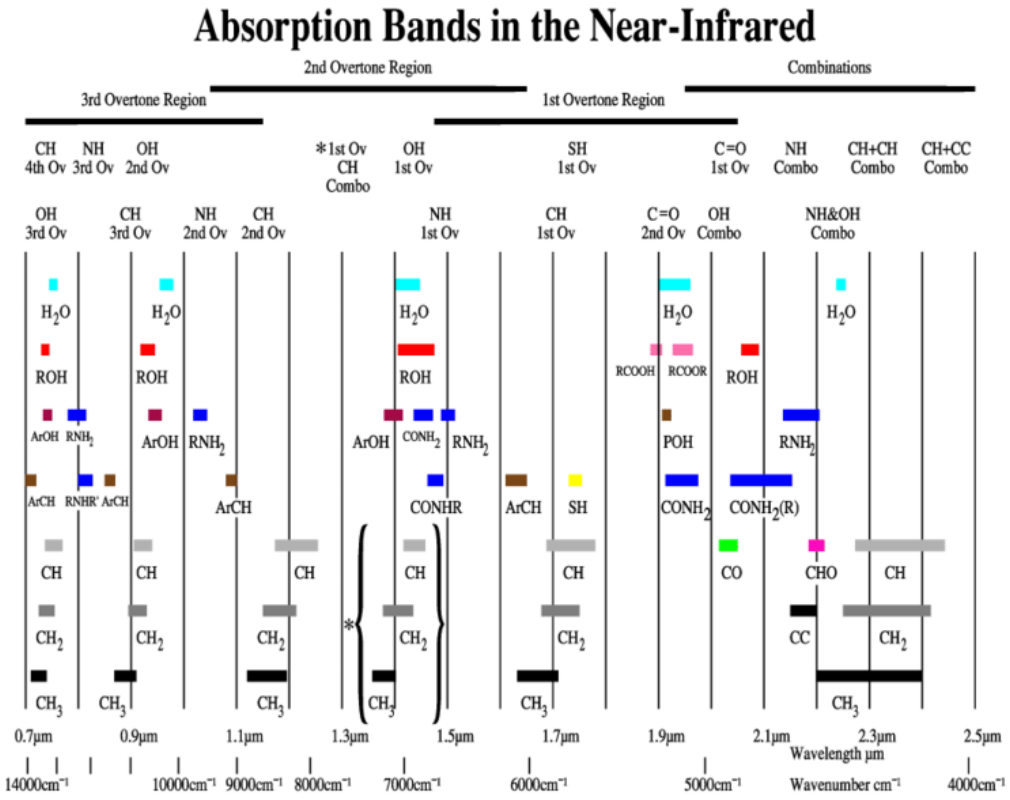


Figure 2.2 The distribution of overtone and combination bands of the main organic bonds in the electromagnetic wave region (Harris & Altaner, 2013).

The absorption intensity reduces as the overtone number increases. Molecular vibrations in the NIR region are in the form of X-H, where X is carbon, nitrogen, or oxygen and H is hydrogen. The vibration modes of X-H are due to stretching, bending, or deformation (Osborne et al., 1993). Organic molecules have specific absorption patterns in the near-infrared region that can report the chemical composition of the material being analysed.

2.3.1.1 Modes of Vibration

The absorption of electromagnetic radiation in the infrared region of the spectrum depends on the existence of an electric dipole moment across the vibrating bond. Consequently, diatomic molecules such as H_2, Cl_2, and F_2 do not produce vibrational absorption bands since they have no permanent electric dipole moments (Osborne et al., 1993). It is a triatomic molecule or AX_2, group that is of interest in NIRS, particularly in connection with this research. The vibration modes of a molecule can be either stretching or bending. Stretching modes involve a continuous change in the interatomic distance along the axis of the bond, whereas bending modes involve a change in the bond angle. For an AX_2 group, there are six possible modes of vibration; two stretching and four bending modes, known as symmetrical stretching, asymmetrical stretching, scissoring, rocking, wagging, and twisting (Fig. 2.3).

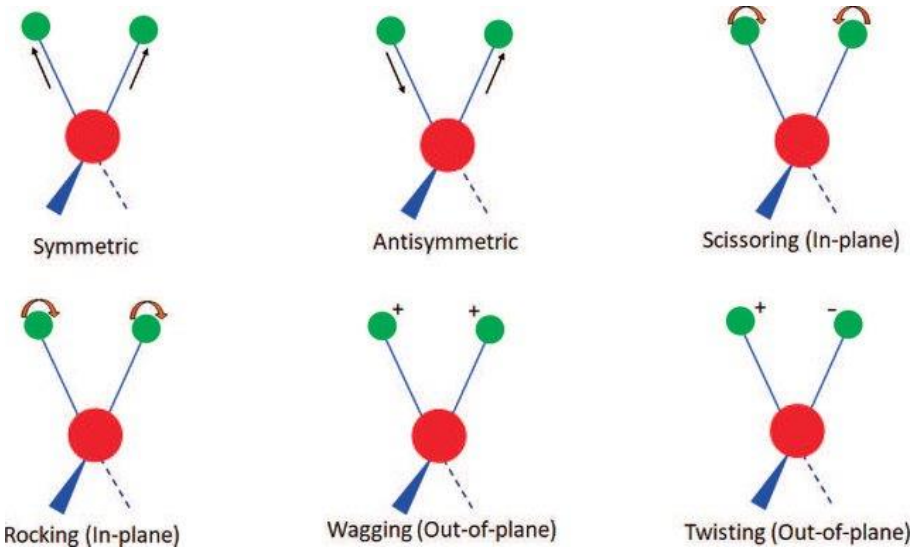


Figure 2.3 Modes of vibration of an AX_2 group (El - Azazy, 2018).

2.3.1.2 Absorption bands of important constituents

The most common absorbing species observed in fruit are listed in table 2.1. It becomes evident that O-H and C-H bonds will dominate the absorption spectra of fruit in the NIR region. Thus, using absorption peaks observed by other authors, it is possible to list the positions of the absorption peaks expected to observe in fruit.

Table 2.1 Positions of expected absorption peaks in fruit between 700 and 1100 nm (Osborne et al., 1993)

Wavelength (nm)	**Absorbing Species**	**Wavelength (nm)**	**Absorbing species**
746	CH_2	970	H_2O
760	H_2O	978	H_2O
834	H_2O	978	Sugar
838	Sugar	978	Cellulose
860	Cellulose	986	H_2O
888	Sugar	990	Starch
905	Cellulose	994	H_2O
913	CH_2	1005	Sugar
913	Sugar	1010	H_2O
920	Cellulose	1030	H_2O
938	CH_2	1030	Starch
938	H_2O	1058	Cellulose
958	H_2O	1099	H_2O

Table 2.1 shows each O-H and C-H group absorbs radiation at several wavelengths, and many of the absorption peaks of water overlap with the absorption peaks of the other molecular species present in the fruit. From this table, water is the dominant species in fruit. The overlapping of absorption peaks make the task of distinguishing between different species difficult.

2.3.2 Measurement setup

Three different measurement setups for obtaining near infrared spectra are shown in Fig. 2.4. In reflectance mode (Fig. 2.4(a)), light source and detector are mounted under a specific angle, e.g., 45°, to avoid specular reflection. In transmittance mode, the light source is positioned opposite to the detector (Fig. 2.4(b)), while in interactance mode the light source and detector are positioned parallel to each other in such a way that light due to specular reflection cannot directly enter the detector (Fig. 2.4(c)). This can be achieved by means of a bifurcated cable in which fibers leading to the source and detector are parallel to each other and in contact with the product, or by means of a special optical arrangement (Greensill & Walsh, 2000; McGlone et al., 2002b). In both reflectance and transmittance mode, integrating spheres may also be used to collect light and increase the signal to noise ratio (Nicolaï et al., 2007).

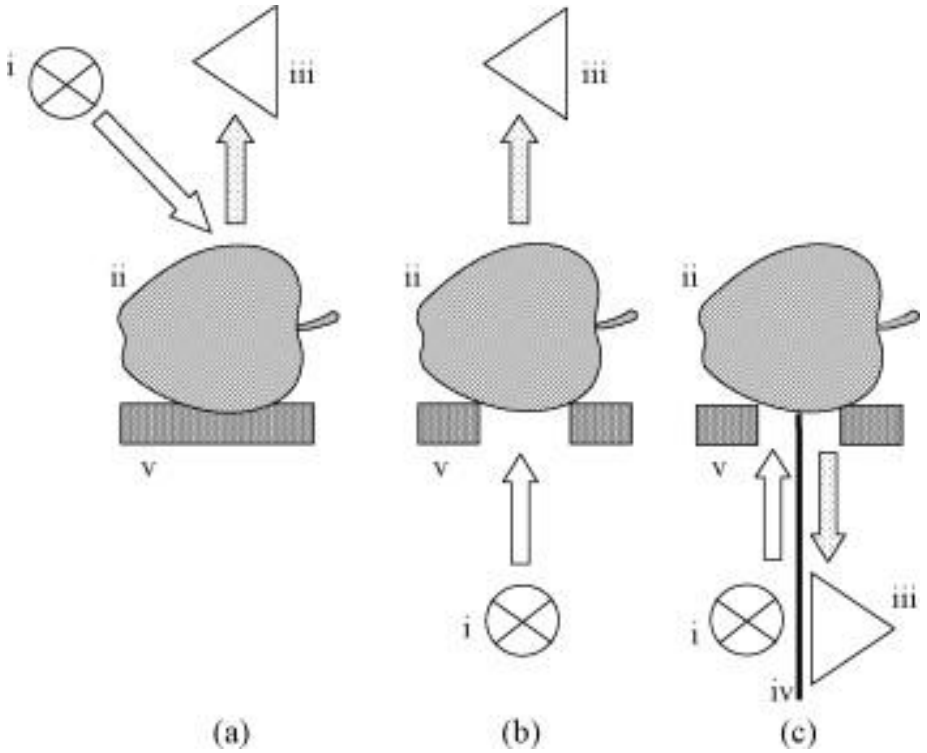


Figure 2.4 Setup for the acquisition of (a) reflectance, (b) transmittance, and (c) interactance spectra, with (i) the light source, (ii) fruit, (iii) monochromator/detector, (iv)light barrier, and (v) support. In interactance mode, light due to specular reflection is physically prevented from entering the monochromator by means of a light barrier (Nicolaï et al., 2007).

In selecting the measurement setup, it is important to know that the penetration of NIR radiation into fruit tissue decreases exponentially with the depth (Greensill & Walsh, 2000; Lammertyn et al., 2000a). The light penetration depth is defined as the distance at which the light intensity drops to a value of 1/e, or ~37%. Fraser et al. (2001) found a penetration depth of more than 25 mm in the 700–900 nm range, reaching a peak of 35 mm at 713 mm and less than 1 mm in the 1400–1600 nm range for apple. The limited penetration depth restricts the potential of reflectance or interactance measurements for detecting internal defects, and decreases the

accuracy of NIR based measurements of internal quality attributes of thick-skinned fruit, such as citrus. Transmission measurements, on the other hand, need very high light intensities which can easily burn the fruit surface and alter its spectral properties. Also, the transmitted light carries information about the skin and the core of the fruit which might or might not be relevant, depending on the application.

Schaare and Fraser (2000) compared reflectance, interactance, and transmission measurements to measure SSC, density, and internal flesh color of yellow-fleshed kiwifruit. They found that interactance mode spectra provided the most accurate estimates (Fig. 2.5). SSC was predicted with a standard error of prediction (SEP) of ± 0.80°Brix and with correlation coefficient R^2 of 0.93 (Schaare & Fraser, 2000). This study suggested that the accurate results were obtained using interactance mode, followed by transmission mode, and in general, the reflectance was least accurate. Lammertyn et al. (2000a) found only marginal differences between reflectance and interactance measurements to determine the SSC of apple.

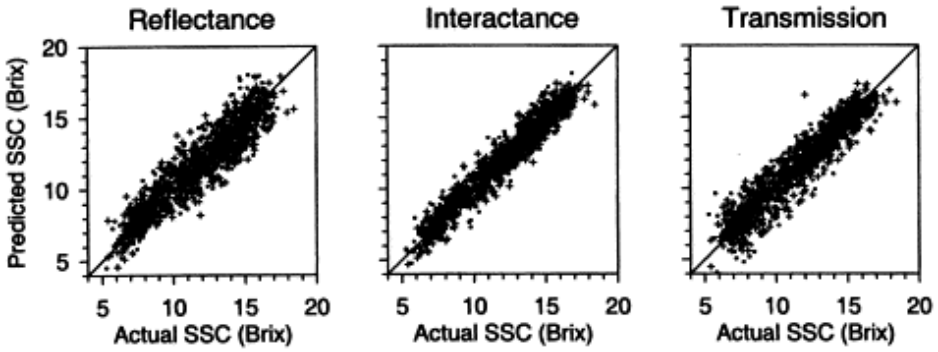


Figure. 2.5 Comparison of predicted to actual SSC content for the three different spectral modes (Schaare & Fraser, 2000).

McGlone et al. (2002c) used interactance mode for prediction of DM and SSC of kiwifruit. The SSC of kiwifruit was predicted with $R^2 = 0.92$, and RMSEP = 0.39%. Afterward, Clark et al. (2003b) used interactance mode with the same equipment and procedures of McGlone et al. (2002c) to predict storage disorders of kiwifruit (*A. chinensis*). Fu et al. (2009) compared the transmission and reflectance modes of visible (VIS) spectroscopy/NIRS using a very sensitive spectrometer for detecting brown heart in pears. With such a spectrometer, better results were obtained based on transmission spectra than those of reflectance spectra.

2.4 Data Analysis

Applications of spectroscopy to measure fruit quality attributes are usually conducted in the Vis/NIR region as spectra in this range contain a good deal of information about O-H, C-H, and N-H vibration absorptions (Pissard et al., 2013). The Vis/NIR spectrum has a low signal-to-noise ratio and high overlap of combination bands and overtones, due to the complex constitution of fruits. The result is a complexity of convolutions and collinearities in the Vis/NIR spectrum. Also there is variable light scattering phenomenon involved, confusing signal interpretation further, since fruit contains many different scattering particles (e.g. various organelles inside the fruit cells) and interfaces (e.g., rigid cell walls, skin, seeds). Therefore, chemometrics is applied for extracting information concerning specific quality attributes from the spectral data (Wang et al., 2015).

2.4.1 Chemometrics

Water is the most important chemical constituent of most fruit and vegetables. As water highly absorbs near infrared radiation, the near infrared spectrum of fruit and vegetables is dominated by water. Further, the near infrared spectrum is essentially composed of a large set of overtones and combination bands. This, in combination with the complex chemical composition of a typical fruit or vegetable causes the near infrared spectrum to be highly convoluted and confused with collinearities. Finally, the spectrum may further be complicated by wavelength dependent scattering effects, tissue heterogeneities, instrumental noise, ambient effects and other sources of variability. As a consequence, it is difficult to assign specific absorption bands to specific functional groups let alone chemical components. Multivariate statistical techniques (also called chemometrics) are therefore required to extract the information about quality attributes which is buried in the NIR spectrum ('model calibration'). Essentially this involves regression techniques coupled with spectral pre-processing (Nicolaï et al., 2007).

2.4.2 Spectral pre-processing techniques

Spectral pre-processing techniques are used to remove any irrelevant information which cannot be handled properly by the regression techniques. Several pre-processing methods have been developed for this purpose (Martens & Stark, 1991; Næs et al., 2002).

1) Multiplicative Scatter Correction

Multiplicative scatter correction (MSC) was first proposed by Ilari et al. (1988) to compensate for the effect of non-uniform scattering created by diverse particle sizes, uneven distribution and other physical effects in the spectral data. The feasibility of MSC has been confirmed by Liu et al. (2010) and Shao et al. (2009). It is done by linearizing each spectrum to an "ideal" spectrum, which corresponds to the average spectrum of the calibration set. The linear relationship between each spectrum and the average spectrum is fitted through the method of least squares. This suggests that MSC is feasible for removing the 'ideal' linear scattering when the linear relationship between absorbance and sample concentration is good.

2) Signal normal variate

Signal normal variate (SNV) is similar to MSC; the objective is to eliminate the deviations caused by particle size and scattering (Barnes et al., 1989). The method assumes that the absorbance spectrum of the sample follows some trend, such as Gaussian distribution. Therefore, based on this hypothesis, each spectrum is calibrated. The first step is to get the average value of a spectrum that is subtracted from the original spectrum and then the result is divided by the standard deviation of the spectrum. The correction capability of SNV is usually stronger than that of MSC as SNV corrects each spectrum alone (Wang et al., 2015).

3) Derivative correction

Derivative correction is a widely used method in NIRS. The first and second derivatives of the spectrum are applied to reduce baseline effects and scattering, respectively, and to amplify the presence of smaller absorbances otherwise obscured by larger absorbances. They can remove background interference, distinguish superimposed peaks, and enhance the spectral resolution. Walsh et al. (2004) utilized second derivative absorbance spectra to develop a calibration model for SSC and DM prediction for various fruits including apples, kiwifruit, and mangoes, and achieved low error values.

2.5 Why Aquaphotomics?

Fruits such as apples and kiwifruit, are more than 80% water (Osborne, 1997). A water absorption peak dominates the NIR spectrum of whole, intact fruit in the SWNIR region around 970 nm, which can shift and vary in shape in response to changes in factors such as temperature, DM, or SSC (Fig. 2.6 and Fig. 2.7).

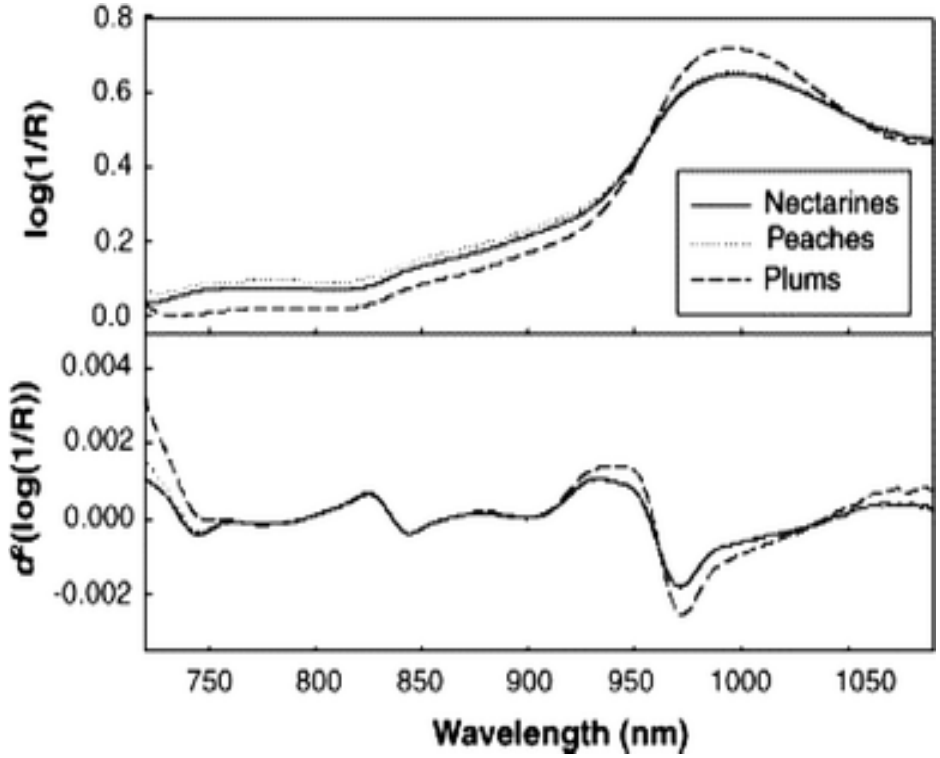


Figure 2.6. Short-wave NIR (700–1100 nm) spectra of nectarines, peaches, and plums (Golic & Walsh, 2006). The absorption peak around 970 nm is the second overtone of water (Lin & Ying, 2009).

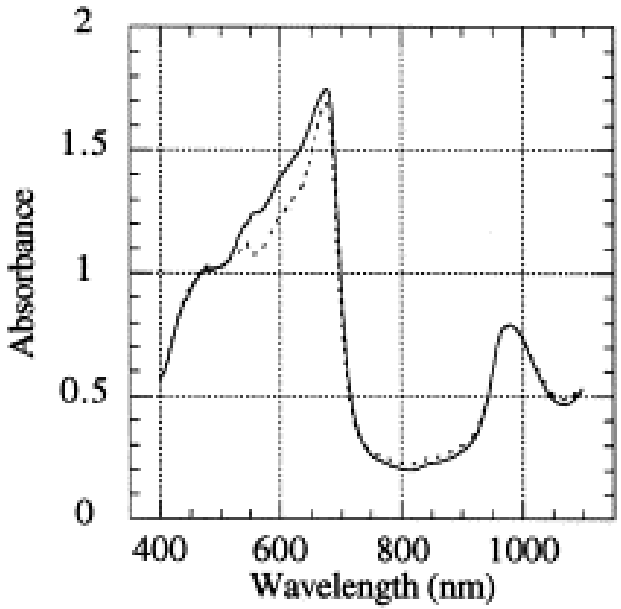


Figure 2.7. Absorbance spectra for soft (dashed line) and firm (solid line) kiwifruit (McGlone & Kawano, 1998).

Historically, the water content of fruit has been considered as interference. The shift in water bands can create a bias reducing model performance when predicting quality parameters (Kawano et al., 1995; Roger et al., 2003; Acharya et al., 2013). Therefore, some researchers have preferred to avoid the water band region when creating models to reduce bias due to temperature shift, for, e.g. using the wavelength range of 745– 825 nm and 845–900 nm (Acharya et al., 2014). Intact fruit is composed of three main constituents; water, soluble solids, and insoluble solids. It is reasonable to expect the foundation for an accurate NIR spectroscopic prediction of sugar content in fruit would be the carbohydrate peaks. However, close inspection of a fruit spectrum in the region of 780 to 1100 nm reveals that the signal is dominated by the water absorbance bands around 840 and 970 nm (Fig. 2.6, 2.7). NIR spectra have been shown to reliably predict the DM content of apples (McGlone et al., 2002a) and kiwifruit (McGlone et al., 2002c). This is not surprising given DM is equal to one minus the water content. To overcome the problem of high bias due to temperature variation and to understand how the structure of water changes with these variations, Aquaphotomics focusses on absorbance patterns related to water bands and the effect of perturbations due to variation in factors such as temperature, the concentration of solutes, and the environment as depicted in Fig. 2.8. (Tsenkova, 2006; Tsenkova, 2007;

Tsenkova, 2008b; Tsenkova, 2009; Tsenkova et al., 2015; Tsenkova et al., 2018).

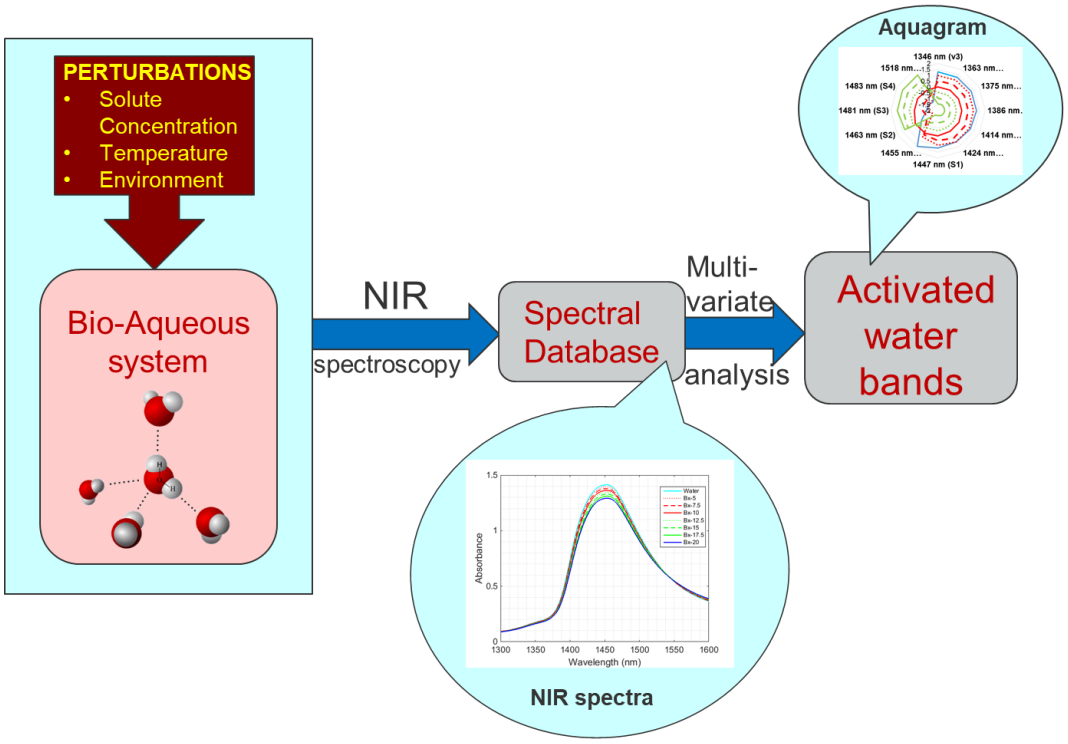


Figure 2.8. The concept of aquaphotomics as it helps in identifying perturbations (variations in temperature, solute concentration, and environment)

NIR spectroscopy, along with multivariate spectral analysis, shows that changes in the water structure reflect other solutes present in the water. The first overtone of the OH stretch of water (1300-1600 nm) has been used to measure small concentrations of solutes and for disease diagnosis (Kinoshita et al., 2012; Takemura et al., 2015). The water peaks at 1450 nm (the first overtone of the OH stretch of water) and 970 nm (the second overtone of the OH stretch of water) in fruit/juice vary due to SSC concentration, temperature, and DM. Therefore, in this research work, an aquaphotomics approach was investigated to learn more about changes in

the water structure that are apparent in the 1300-1600 nm and 800-1100 nm wavelength regions. Firstly, it is important to understand what is meant by aquaphotomics, as explained in the next sub-section.

2.5.1. What is Aquaphotomics?

The aquaphotomics concept was proposed by Prof. Roumiana Tsenkova from the Biomeasurement Laboratory, Kobe University, Japan, in 2005 (Tsenkova, 2006). Aquaphotomics primarily focuses on water spectral regions using NIRS to determine water bands that change or become visible due to the change/variation. It describes the dynamic spectroscopy of biological and aqueous systems based on changes in water molecular vibrations related to other molecules present in a biological specimen, which helps in understanding the overall function of the system (Tsenkova et al., 2015; Tsenkova et al., 2018). Aquaphotomics is introduced to describe the concept of approaching water as a multi-element system that can be described by its multidimensional spectra. It uses spectroscopic techniques to follow changes in water molecular vibrations caused by other molecular vibrations present in a biological specimen, which helps in understanding the function of the whole system (Tsenkova, 2009).

2.5.1.1 Water matrix coordinates (WAMACS) and aquagrams

In the area of the first overtone of water, 12 water absorbance bands corresponding to specific water molecular species were uncovered

(Tsenkova, 2006). These 12 absorbance bands, named water matrix coordinates (WAMACS), were found to be consistently important in spectral analysis of different aqueous and biological systems, and under different perturbations. Table 2.2 provides assignments for the WAMACS of the first overtone of OH stretching vibrations (Muncan & Tsenkova, 2019). The water absorbance spectral pattern WASP is usually presented by aquagrams (Fig. 2.9). An aquagram is a radar chart that displays normalized absorbance at selected water absorbance bands. For the first overtone of water, the axes of the aquagram are usually based on previously discovered 12 WAMACS. Specifically, the absorbance at each of the selected WAMACS wavelengths was calculated as:

$$A'_{\lambda} = \frac{(A_{\lambda} - \mu_{\lambda})}{\sigma_{\lambda}} \qquad (1)$$

where A_{λ} is the SNV absorbance value for a sample at wavelength λ, and μ_{λ} and σ_{λ} are the mean and standard deviations across all SNV transformed samples at that wavelength.

The aquagrams are visually very convenient tools that enable quick and comprehensive comparison of different systems or conditions of the same system by comparison of their WASPs.

Table 2.2. Water matrix coordinates in the area of the first overtone of water in the near infrared region (1300 to 1600 nm) (Muncan & Tsenkova, 2019)

WAMACS	Range (nm)	Assignment
C1	1336–1348	$2v_3$: H_2O asymmetric stretching vibration
C2	1360–1366	$OH^-\cdot(H_2O)_{1,2,4}$: Water solvation shell
C3	1370–1376	$v_1 + v_3$: H_2O symmetrical stretching vibration and H_2O asymmetric stretching vibration
C4	1380–1388	$OH^-\cdot(H_2O)_{1,4}$: Water solvation shell $O_2^-\cdot(H_2O)_4$: Hydrated superoxide clusters $2v_1$: H_2O symmetrical stretching vibration
C5	1398–1418	Water confined in a local field of ions (trapped water) S_0: Free water Water with free OH^-
C6	1421–1430	Water hydration band H-OH bend and O-H…O
C7	1432–1444	S_1: Water molecules with 1 hydrogen bond
C8	1448–1454	$OH^-\cdot(H_2O)_{4,5}$: Water solvation shell
C9	1458–1468	S_2: Water molecules with 2 hydrogen bonds $2v_2 + v_3$: H_2O bending and asymmetrical stretching vibration
C10	1472–1482	S_3: Water molecules with 3 hydrogen bonds
C11	1482–1495	S_4: Water molecules with 4 hydrogen bonds
C12	1506–1516	v_1: H_2O symmetrical stretching vibration v_2: H_2O bending vibration Strongly bound water

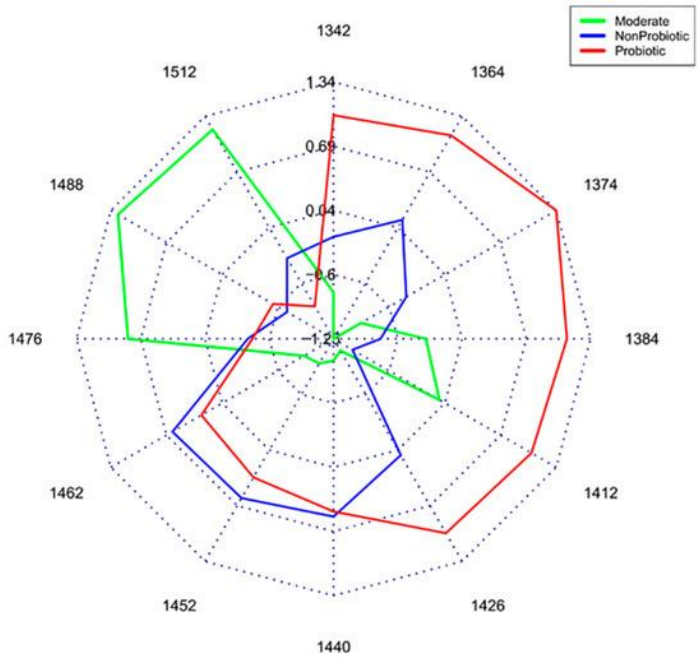


Figure 2.9. Aquagrams of culture media of groups of probiotic, moderate and non-probiotic strains. Average values of normalized absorbance values of the water matrix coordinates for each group are plotted on each wavelength axis (Slavchev et al., 2015).

2.5.2 What has been done in Aquaphotomics?

Most aquaphotomics work so far has focused on the first overtone of the OH stretch of water (1300-1600 nm) (Tsenkova, 2006; Tsenkova, 2010; Matija et al., 2012; Gowen et al., 2013; Munćan et al., 2014; Gowen et al., 2015; Tsenkova et al., 2015; Bázár et al., 2016; Cui et al., 2017). There are very few publications in which the second overtone region has been explored.

In the research carried out at the Biomeasurement Technology Laboratory at Kobe University, Japan, water was perturbed with lactose, NaCl, metals, and other solutes at various concentrations. The concept has also been applied to observe differences in milk from healthy and mastitis-infected cows (Meilina et al., 2009).

The literature review revealed that very little research has been done using aquaphotomics in the second overtone region. Jinendra et al. (Jinendra et al., 2010) used SWNIRS and aquaphotomics for the diagnosis of soybean plants infected with the soybean mosaic virus (SMV).

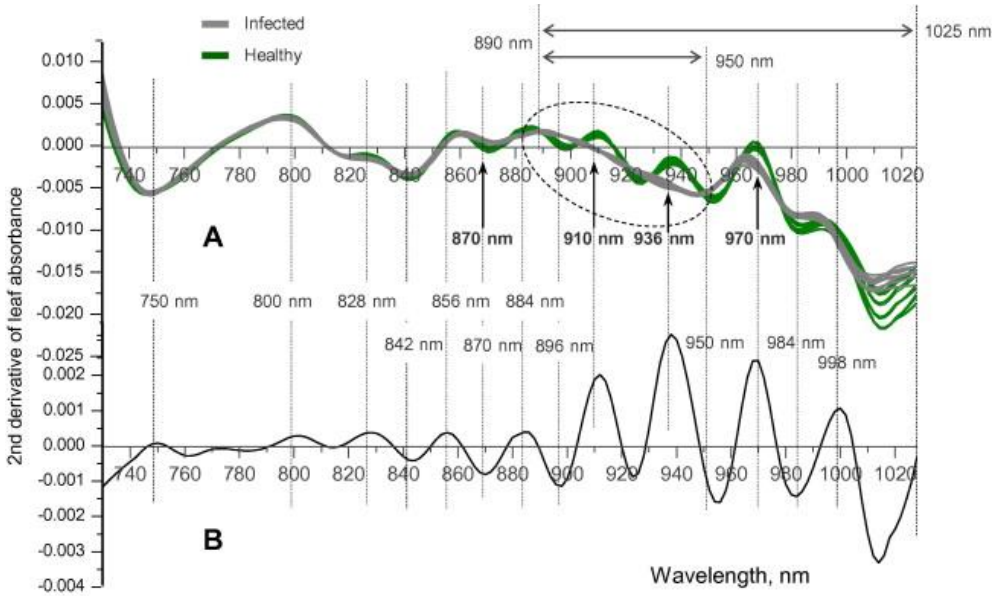


Figure 2.10. (A) The second derivative single plant average spectra of healthy and soybean mosaic virus-infected soybean plants. (B) The difference spectrum of the two sets of average spectra for healthy and soybean mosaic virus-infected plants (Jinendra et al., 2010)

Spectra of leaves were collected in the range of 730–1025 nm from healthy and infected plants using an NIR spectrometer (FQA-NIR Gun, Shizuoka Shibuya Seiki, Hamamatsu, Japan). The second derivative transformed infected plants' spectra and showed a difference in their response to the healthy plant group throughout the spectrum (Fig. 2.10). The difference was most significant in the 890–1025 nm region where the

SMV-inoculated plants' spectra showed a general downward slope and no clear absorbance peaks in the 890–950 nm region. The researchers observed spectra variations in the 890–1025 nm region resulting from different molecular vibration patterns of water molecules caused by the virus infection, which allowed successful discrimination of virus-infected soybeans in the latent stage of the disease. The peak shift observed at 970 nm towards shorter wavelengths showed the virus's influence on commonly observed leaf water stress. The wavelengths 870 nm, 910 nm, 936 nm, 950 nm showed different water absorbance patterns (WAP) for virus-induced hypersensitivity water stress, which can be used to confirm the presence of SMV infection. The classification model, based upon soft independent modelling of class analogy (SIMCA), predicted the disease with 91.6% sensitivity and 95.8% specificity when the second order derivatives of the individual plant averaged spectra were used in the water wavelength range (730-1025 nm).

In another application, aquaphotomics in the 890-1120 nm range was used to characterise 18.2 MΩ water and Nano-Harmonized Substance (NHS) as shown in Fig. 2.11(a) (Matija & Tsenkova, 2011). In the second experiment, human skin was characterised under three different conditions: non-treated, treated by the emulsion (a solution of cream base) and the cream with NHS (Fig.2.11(b)). Different water spectral patterns were obtained showing the difference between various treatments.

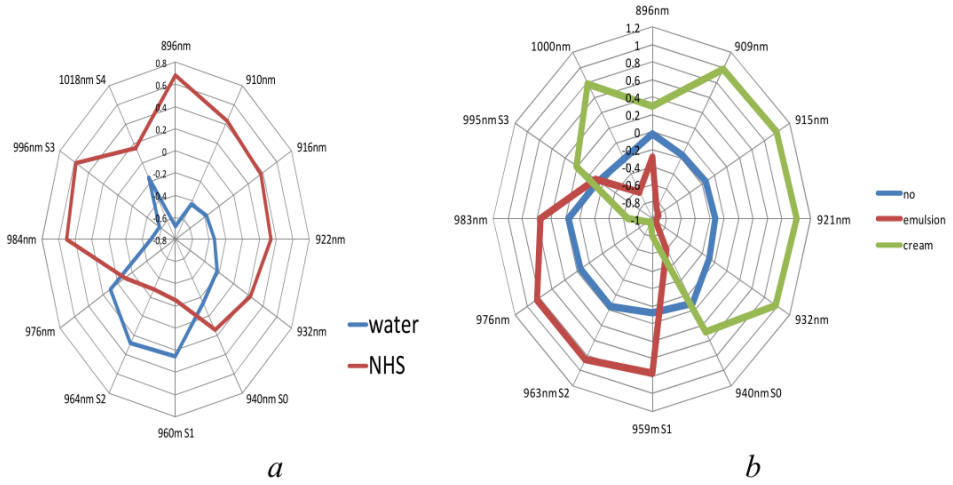


Figure 2.11. a) aquagram of water and NHS in the second overtone, b) spectrum of human skin: not treated, treated by emulsion and treated cream which contents NHS (Matija & Tsenkova, 2011).

To predict low concentrations of sugars, Bázár et al. (Bázár et al., 2015) used near infrared spectroscopy in the first overtone region of the OH stretch of water (1300-1600 nm). The accuracy of the validation models for lactose concentration showed values of R^2_{CV} =0.99 and 0.98, RMSECV=1.08 mM and 0.42 mM in the 10-100 mM and 1-10 mM concentration ranges, respectively. However, the PLSR model did not perform well in the lower 1-0.1 mM and 0.1-0.02 mM range. Converting mM concentration of sugars into °Brix, for example, for sucrose whose molecular weight is 342.2965 g/mol (Sritham & Gunasekaran, 2017), 1-100 mM sucrose concentration is converted into approximately 0.034–3.42 °Brix (Harvey, 2000). Since the SSC concentration (measured in °Brix, where 1 °Brix is a ratio of 1g of sucrose/100g of solution) of kiwifruit juice typically varies between 6 and 20 °Brix (McGlone & Kawano, 1998; Feng et al., 2011), it is

higher than the concentrations described in the above study. Therefore, the same technique was applied to observe the effect of SSC concentration on the water structure and the model performance.

To observe how water structure changes with water quality, Kovacs et al. (Kovacs et al., 2016) used NIRS along with aquaphotomics to get information on different water species. This approach provided information about qualitative changes in water molecular arrangements without taking into account the reason for the alteration in quality. Different species and concentrations of solutes in aqueous systems structure the water solvent differently. Water samples were perturbed with 1-100 mM concentrations of acetic acid, lactose, and NaCl. Measurements taken in the first overtone region of water were used to find WAMACS. Different water species are activated upon perturbing the sample with different solutes, as shown in the water absorbance pattern chart called an aquagram (Fig. 2.12). The aquagram of the aqueous solutions of acetic acid (Fig. 2.12a) show mainly higher absorbance values at shorter wavelengths, but lower absorbance values at longer wavelengths, compared to Milli- Q water, which means they contain more water molecules with free OH bonds. In Fig. 2.12b, which presents the aquagram of aqueous solutions of lactose, the water spectral pattern (WASP) of the higher concentrations implies more hydrogen bonded water structures compared to pure Milli-Q water, while the WASP of the lower concentrations represents fewer hydrogen bonded water structures.

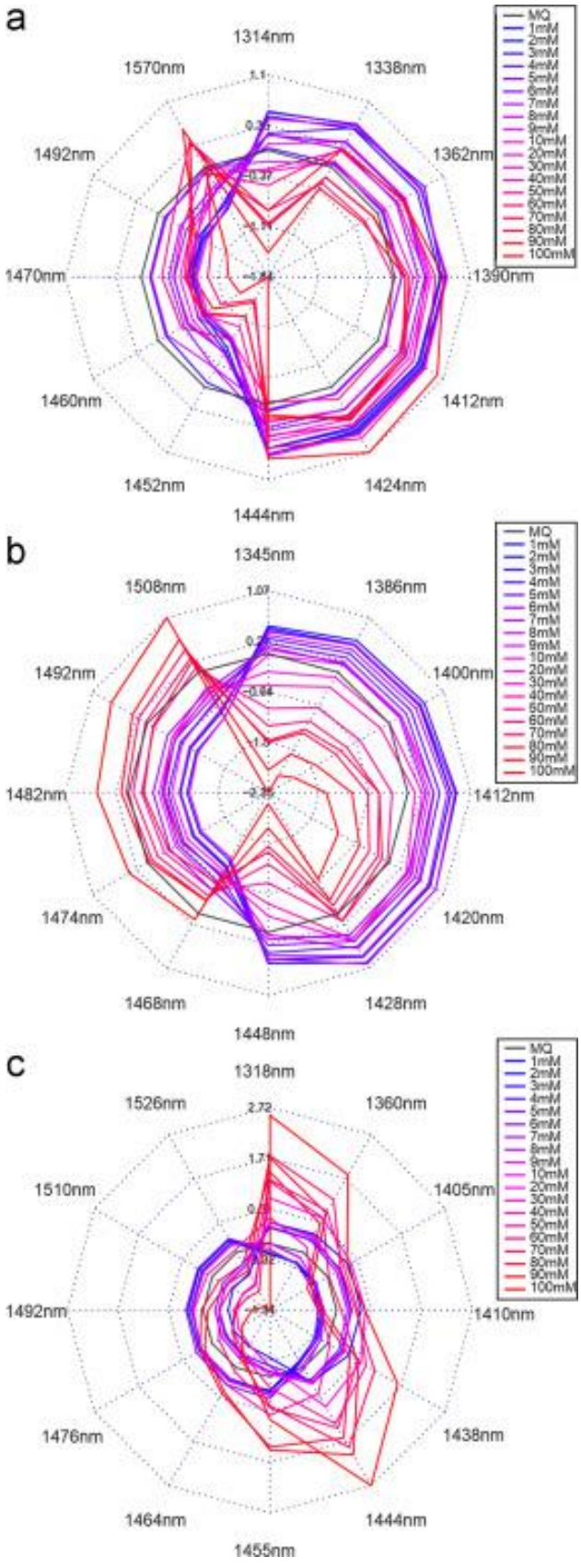


Figure 2.12. Aquagrams of aqueous solutions (a, acetic acid; b, lactose; c, sodium chloride) calculated at selected characteristic wavelengths (Kovacs et al., 2016)

The aquagram of aqueous solutions of sodium chloride (2.12(c)) in the higher concentration samples shows higher absorbance values in the range between 1318 and 1360 nm, which may refer to the first overtone of free OH stretch (OH–(H_2O)n, n=1…4), and in the range between 1438 and 1455 nm. These are known as the bands of water hydration, water molecules connected to other water molecules (S1), and the asymmetric stretching and bending of the first overtone of water (H_2O-2ν2+ ν3). However, in the range between 1492 and 1526 nm, the higher concentration samples provided lower values; these wavelengths are connected with strongly hydrogen-bonded water. These findings mean that decreasing the concentration of salt causes a shift towards longer wavelengths.

One interesting application, investigating the adulteration of honey with high fructose corn syrup (HFCS), was made by Bázár et al. (Bázár et al., 2016) in the 1300-1800 nm wavelength region. Aquaphotomics-based evaluations showed that unifloral honey contained more highly organized water than industrial sugar syrup, supposedly because of the greater variety of molecules dissolved in the multi-component honey. Adulteration with HFCS caused a gradual reduction of water molecular structures, especially water trimers, which facilitate interaction with other molecules. Levels of HFCS adulteration could be detected using leave-one-honey-out cross-validation (RMSECV = 1.48; R^2_{CV} = 0.99) and partial least squares regression using the 1300–1800 nm spectral band that contains absorption bands related to both water and carbohydrates.

2.6 Effect of sample temperature on quality prediction

The NIR spectrum of water is sensitive to temperature variations (Segtnan et al., 2001). Since fruits are more than 80% water, their NIR spectrum is also affected by temperature variations. Previous researchers have studied the effect of temperature on prediction models of fruit quality (Acharya et al., 2013), observing that a calibration equation developed at one fixed temperature could not reliably predict samples at a different temperature. The framework of aquaphotomics appears suitable for examining the temperature sensitivity of water spectra as they appear in fruit measurements.

To access model accuracy, the following parameters are used. The prediction error of a calibration model is defined as the root mean square error for cross-validation (RMSECV) when cross-validation is used or the root mean square error for prediction (RMSEP) when internal or external validation is used (Næs et al., 2002; Nicolaï et al., 2007):

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

where n_p is the number of validated objects, and $\hat{y}_i$ and y_i the predicted and measured value of the ith observation in the test set, respectively. This value gives the average uncertainty that can be expected for predictions of future samples. The number of latent variables (LV) in the calibration model is typically determined as that which minimises the RMSECV. Some

researchers also use the standard error of prediction (SEP) which is bias corrected RMSEP:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n_p}(\hat{y}_i - y_i - b)^2}{n_p}} \qquad (2)$$

with b representing the model bias.

Calibration models are called robust when the prediction accuracy is relatively insensitive to unknown changes in external factors (Wang et al., 1991). The main factor that affects model performance is temperature. The hydroxyl group of pure water causes a broad absorption band around 1449 nm due to a combination of antisymmetric and symmetric stretching modes of water (Maeda et al., 1995). This band is an overlay of various component spectra corresponding to water clusters with no, one, two, three and four hydrogen bonds. Raising the temperature decreases the average cluster size and increases the relative absorbance of the clusters with no hydrogen bonds, causing a sharpening and shift of the hydroxyl band to shorter wavelengths. In fruit and vegetable applications, temperature differences between calibration and validation sets mainly cause bias (Golic & Walsh, 2006). Roger et al. (Roger et al., 2003) found a bias of 8°Brix for a temperature variation of 20°C (range 5-25°C) for SSC prediction in apples. Several techniques are used to compensate for temperature fluctuations. Kawano et al. (1995) developed calibration equations using samples at different temperatures. Peirs et al. (2003) proposed to either develop a global robust calibration model to cover the temperature range expected in

future measurements or to develop different models for dedicated temperatures. (Roger et al., 2003) virtually removed the temperature-induced bias in SSC by applying the external parameter orthogonalisation (EPO) algorithm as a preprocessing technique to remove that part of the spectral data matrix of fruit which is most affected by temperature.

Using the framework of aquaphotomics, changes in the water structure of fruit juice and whole, intact fruit occurring with changes in temperature were analyzed. Water wavelengths affected by temperature variation were identified in the first and second overtone regions. Two correction pre-treatments; i.e. extended multiplicative scatter correction (EMSC) and EPO with principal component loading of water at different temperatures as an interferent spectrum, were used to correct for temperature variation.

2.6.1 EMSC

The concept of EMSC pre-treatment was introduced by Harald Martens (Martens & Stark, 1991; Martens et al., 2003; Martens et al., 2006). EMSC was designed to explicitly compensate for the chemical variabilities by including information about the major analyte and the interferent spectra to be removed.

Equation 3 describes the theory of EMSC:

$$X = b_o + b_1\overline{X} + b_2 I + e \qquad (3)$$

where X is the RAW observed spectra, $\overline{X}$ is the reference spectrum (the mean of all calibration spectra), I is an interferent spectrum (to be determined), b_o, b_1, and b_2 are fitting constants, and e is the residual (Gowen et al., 2011a). The corrected spectrum $\hat{X}$ is calculated by rearranging Eqn. 3 as shown below:

$$\hat{X} = \frac{X - b_o}{b_1} - \frac{b_2 I}{b_1} = \bar{X} + \frac{e}{b_1} \quad (4)$$

where the constant terms can be estimated by MLR.

2.6.2 EPO

The concept of EPO was introduced by Roger et al. (Roger et al., 2003). It is a preprocessing method that aims at removing the part most influenced by the external parameter variations from the **X** space. The method estimates this parasitic subspace by computing a PCA on a small set of spectra measured on the same objects, while the external parameter is varying. With no preprocessing Roger observed that the bias in the sugar content prediction was about 8°Brix for a temperature variation of 20°C (Calibration model at 25°C predicting samples at 5°C). After EPO preprocessing, the bias was not more than 0.3°Brix for the same temperature range.

The theory of the EPO algorithm is outlined below (Minasny et al., 2011). Let S be the m-dimensional space of the n measured spectra; the spectra can be written as

$$S = C + G + R \quad (5)$$

where

C is the chemical spectral responses;

G is the part of the matrix that is perturbed by the external parameter and is independent from C; and

R is the independent residual

In matrix form, the spectra **X** (size n x m) can be written as:

$$\mathbf{X = XP + XQ + R} \quad (6)$$

where

P is the projection matrix (size m × m) of the useful part of the spectra: $\mathbf{X^{*} = XP}$;

Q is the projection matrix (size m × m) of the not useful part (e.g., influenced by temperature) of the spectra: $\mathbf{X^{\#} = XQ;}$ and

R is the residual matrix (size n × m)

The aim of EPO is to obtain the useful spectra $\mathbf{X^{*} = X(I - Q),}$ while matrix **Q** can be written as $\mathbf{Q = GG^{T}.}$

To estimate G, the uninformative part of the spectra that is orthogonal to the useful part of the spectra, Roger et al. (Roger et al., 2003) suggested using the principal component of the difference spectra **D.**

where

D is the difference matrix generated by subtracting the average spectra for the samples at lowest temperature from the samples at all temperatures.

The spectra in both the calibration and validation sets were then multiplied by the correction factor $\mathbf{I-GG^T}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{G^T}$ is the transpose of $\mathbf{G}$.

The transformed spectra are calculated as $\mathbf{X^*=XP}$ where $\mathbf{P= I-GG^T}$.

Chapter 3: Simple sucrose solution

3.1 Introduction

To better understand the process of aquaphotomics analysis, the research work started with simple solutions of sucrose. The literature confirmed that the NIR spectrum of water is sensitive to sugar variations (Segtnan, 2001; Jensen et al., 2003). The water peak at 1450 nm (first overtone of OH stretch of water) and 970 nm (second overtone of OH stretch of water) in sucrose solutions varies with concentration.

Since kiwifruit and apple consist of more than 80% water, their spectral signature will also be affected by varying sugar concentration. The aquaphotomics discipline was investigated to learn more about changes in the water structure caused by the sucrose perturbation in the 1300-1600 nm and 870-1100 nm wavelength regions. Not only the first overtone region of OH stretch of water (1300-1600 nm) was investigated but also the SWNIR region as it is important for intact fruit quality measurement.

The effect of path length (cuvette size) was also investigated for its effect on the interpretation ability of aquaphotomics in the second overtone region of water. Aquagrams were built for varying concentrations of aqueous sucrose samples. Measurements with optimal path length improved the signal to noise (SNR) ratio (Jensen & Bak, 2002). All measurements were recorded with an FT-NIR spectrophotometer.

3.2 Materials and Methods

3.2.1 Sample preparation

Sucrose ($C_{12}H_{22}O_{11}$, M = 342.30 $g.mol^{-1}$, Sigma-Aldrich Corporation, Missouri, USA) was used to prepare aqueous solutions in the concentration range from 5 to 17.5% (w/w) with a step increase of 2.5%. Solutions were prepared using Milli-Q water. The Milli-Q water was produced by a water purification system (Millipore, Thermofisher Scientific, Australia) and had a resistivity of 18.2 MΩ·cm.

3.2.2 Reference measurement

The Brix value of all the sucrose solutions was measured at room temperature using a digital refractometer (Atago Co. Ltd, Tokyo, Japan), calibrated to zero with Milli-Q water (Vasquez & Mueller, 2019). The Brix value was recorded after placing 0.5 ml of sample into the measurement chamber of the refractometer, enough to fully cover the optical interface.

3.3.3. NIR spectra collection

The transmittance spectra of the solutions were measured at 20°C (±1 °C) with an FT-NIR spectrometer (Tango, Bruker Corporation, Germany) equipped with a temperature-controlled cuvette holder. Two measurements per sample were recorded using quartz cuvettes of 1 mm and 10 mm optical path length. The spectra were measured over the

wavelength range of 870-2500 nm. All spectra were the average of 32 successive scans with a resolution of 16 cm^{-1}. Measurements were also taken for Milli-Q water with 1 mm and 10 mm pathlength cuvettes at the beginning of the experiment.

3.3. Aquaphotomics Analysis

The standard aquaphotomics analysis investigates spectral data at 12 discrete wavebands, called the water matrix coordinates (WAMACS), which are known to be affected by perturbations of the water structure (Tsenkova et al., 2015). For the first overtone water region, these wavebands sit under the broad water absorption peak at 1450 nm, are between 5 and 20 nm in width, and typically separated by more than 5 nm (Tsenkova et al., 2018). The discrete wavebands are attributed to a mix of different water vibrational states, as well as five different H-bonded species (S0, S1, S2, S3, and S4 where the number indicates the number of hydrogen bonds per species) (Maeda et al., 1995). The standard protocol involves assigning individual wavelengths from the spectral dataset to each waveband of the WAMACS (Tsenkova et al., 2015). The exact wavelength in each case is determined from examination of the band activity with respect to:

1) the sample-water difference spectra (i.e., subtracting Raw water spectrum from Raw apple juice spectra)

2) the loading plots of principal component analysis (PCA) of Raw apple juice spectra, and

3) the second derivative spectra.

Multiple choices exist for some wavebands, but the exact value is not critical given the bandwidth of the FT-NIR system is around 3.3 nm at 1450 nm, less than the typical width and/or separation of the WAMACS bands.

Once the WAMACS wavelengths are selected, a water spectral pattern (WASP) presented on the aquagram can be drawn and examined in terms of the system perturbation. The aquagram provides a convenient graphic view for examining the cause and effect of system perturbation in terms of the underlying water structure. Aquagrams were generated from the SNV spectral data. Specifically, the absorbance at each of the selected WAMACS wavelengths was calculated as:

$$A'_{\lambda} = \frac{(A_{\lambda} - \mu_{\lambda})}{\sigma_{\lambda}} \qquad (1)$$

where A_{λ} is the SNV absorbance value for a sample at wavelength λ, and μ_{λ} and σ_{λ} are the mean and standard deviations across all SNV transformed samples at that wavelength.

3.4 Results and Discussion:

3.4.1 First overtone analysis (1300-1600 nm)

The RAW spectra of sucrose solutions at six different concentrations, plus a pure water spectrum, are shown in Fig. 3.1. The FT-NIR measurements were taken using a 1 mm pathlength cuvette since water has high absorption in this region (Jensen & Bak, 2002). There is a broad absorbance peak at 1450 nm that corresponds to O-H stretch first overtone of water (Burns & Ciurczak, 1992), decreasing in magnitude with increased concentration (%). Since water does not contain sugars, its peak intensity was highest and followed in order by lower peak intensities for the low to high sucrose concentration samples.

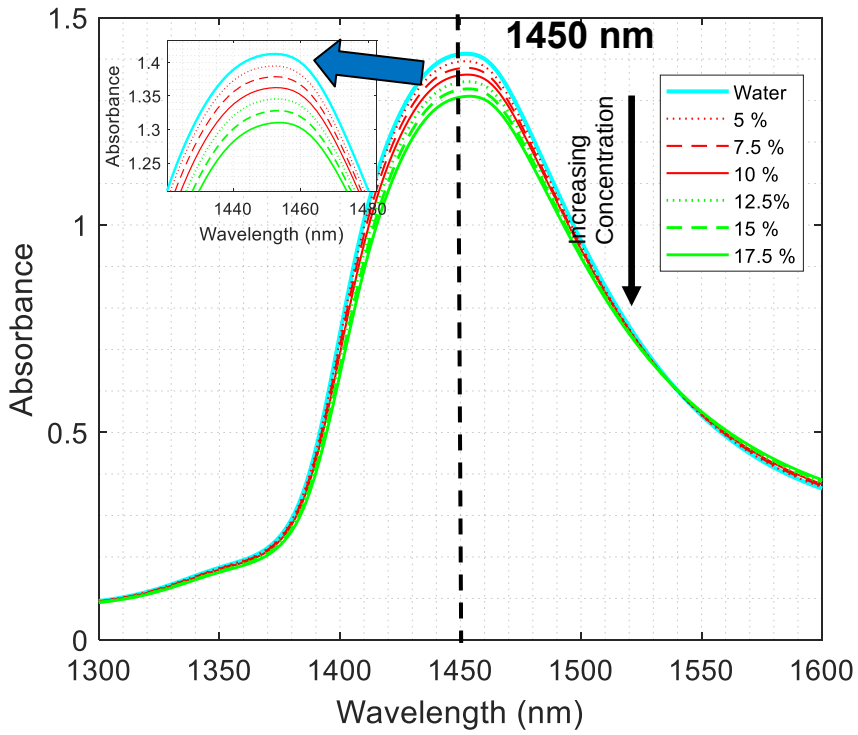


Figure 3.1. Raw absorbance spectra of sucrose solution at six different concentrations and water at 20°C in the first overtone region (1300–1600 nm) with a peak at 1450 nm.

3.4.1.1 Aquaphotomics Analysis

1) Difference absorption spectra: Raw Sucrose solution – Raw Water

With increasing concentration, the difference absorption spectra (Raw Sucrose solution– Raw Water) of the sucrose solutions decreased in the 1414 and 1455 nm regions, with respect to the zero line (Fig. 3.2). Characteristic water bands were observed at 1414 nm due to S0 (free water) and 1455 nm due to v_2+v_3, where v_2 is bending mode, and v_3 is antisymmetric stretching mode. These bands were within 2 nm to those observed in the difference spectra of aqueous sugar (Bázár et al., 2015).

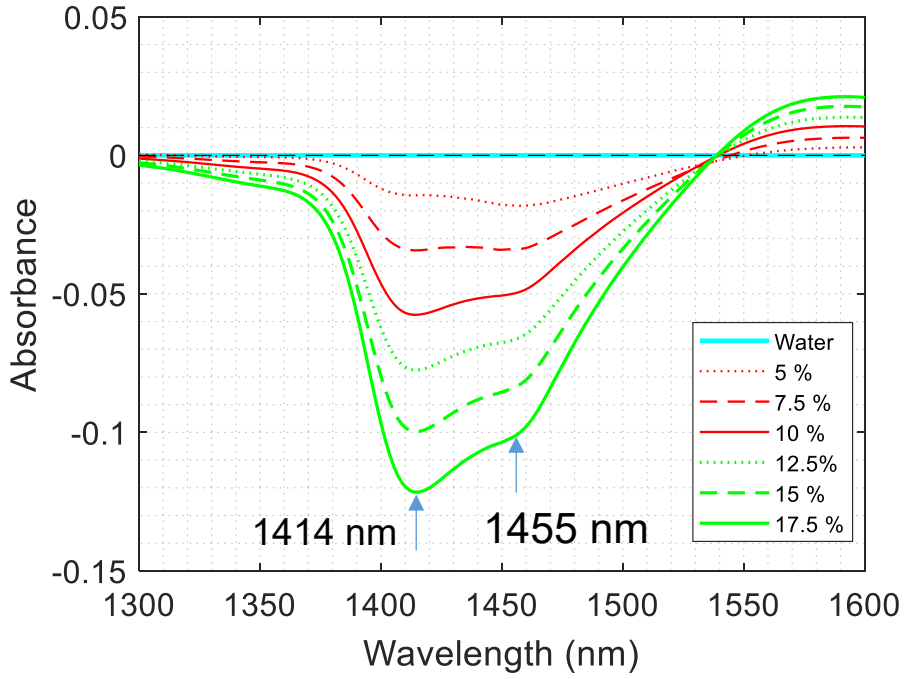


Figure 3.2. Raw absorbance difference spectra of sucrose solutions and water after subtracting raw absorbance spectrum of water in the first overtone region at 20°C.

2) PCA loadings and scores

Three principal components (PCs), which explain over 99% of the spectral variation, were extracted from the Raw absorbance data in the first overtone region (Fig. 3.3). The first PC explains 99.83% of the variation in the data and shows a peak at 1415 nm, attributed to the OH stretch of the free water molecule (S0) (Tsenkova et al., 2015). The second and third PCs explain only insignificant variation. The samples were rank-ordered in descending trend with respect to concentration. This means that with the increase in sucrose concentration, the free water molecule species, S0, decreases.

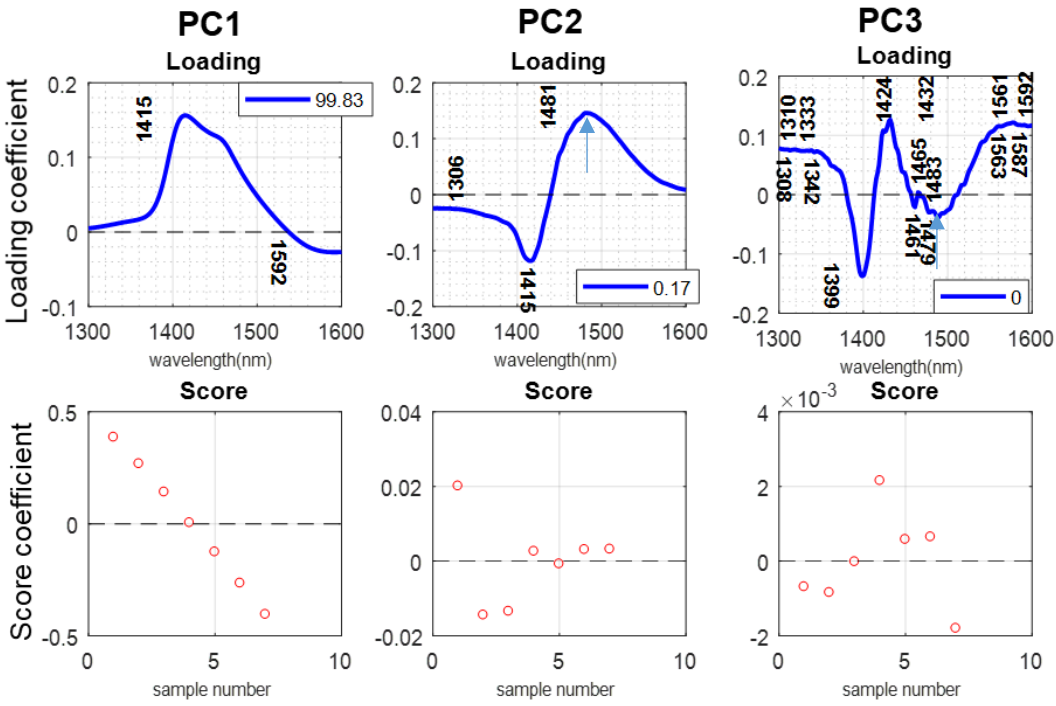


Figure **3.3.** PCA loading and score plots of raw absorbance spectra of water and sucrose solutions. Rectangular boxes signify explained variance in percentage. Labels indicate peak wavelengths in the loading plot.

3) Second derivative spectra of RAW spectra

The second derivative spectra (Fig. 3.4) were used to assist in the identification of further absorption bands, particularly small and/or overlapping absorption peaks not resolvable in the original RAW spectra. The free water, S0 band is strong at 1414 nm and bands corresponding to higher order H-bond water species, such as dimers (S2) and trimers (S3) (Segtnan, 2001; Bázár et al., 2015), are attributed to peaks/troughs in the region from 1440–1470 nm. The broad shoulder-like peak at 1518 nm is attributed to even more highly organized water structures with strong H-bonds (Tsenkova et al., 2015).

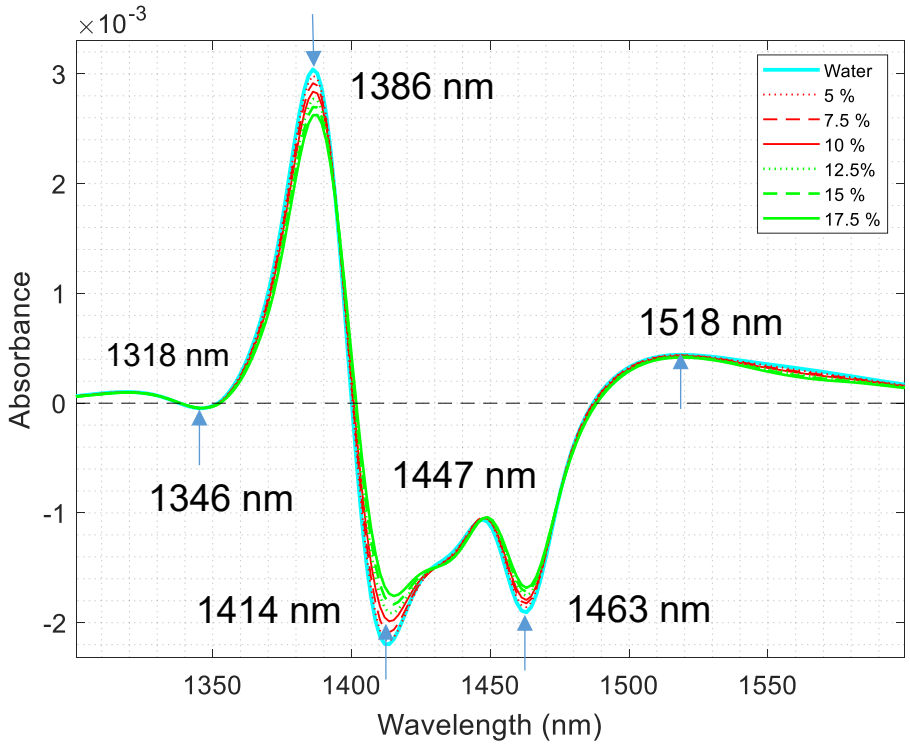


Figure **3.4.** Raw second derivative spectra of sucrose solution at six different concentrations and a spectrum of water. Labels show peak wavelengths of the water spectrum.

4) Identification of WAMACS

An aquaphotomics water matrix coordinate system (WAMACS) was created using the feature wavelengths found from the above spectral analysis that correspond closely to water band coordinates reported in the literature (Fig. 3.5) (Tsenkova et al., 2015). The final selected wavelengths for each WAMACS coordinate as represented in Table 3.1 were C1: 1346 nm from raw second derivative spectra; C2: 1363 nm (picked randomly); C3: 1375 (picked randomly); C4: 1386 nm from second derivative spectra; C5: 1414 nm from raw absorbance difference spectra; C6: 1424 nm from PC loading; C7: 1447 nm from raw second derivative spectra (close to water band); C8: 1455 nm from raw absorbance difference spectra; C9: 1463 nm, C10: 1481 nm, C11: 1483 nm all from PC loading, C12: 1518 nm from second derivative spectra (close to water band). There were no gaps or confusions evident in the selections; the wavelength coordinates all matched well with the wavelength scheme previously reported (Tsenkova et al., 2015).

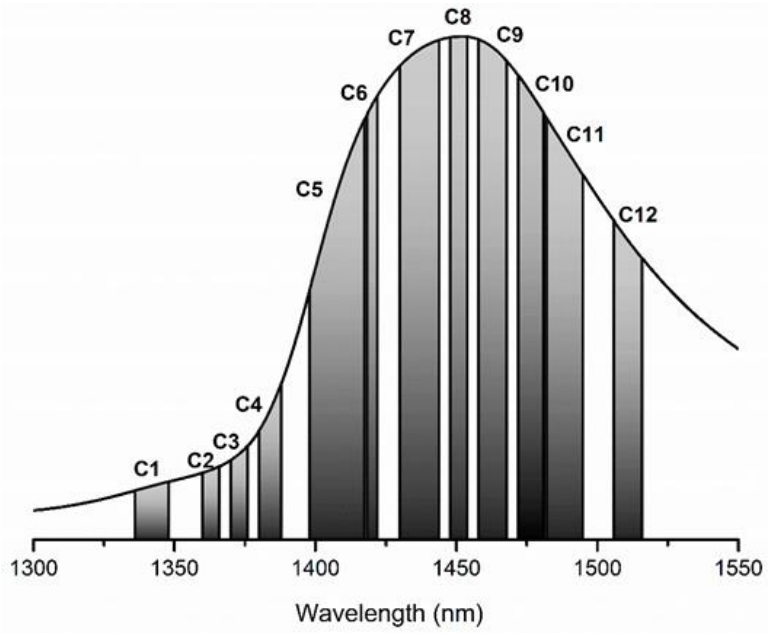


Figure 3.5 Water matrix coordinates in the area of the first overtone of water in the near infrared region (1300 to 1600 nm) (Muncan & Tsenkova, 2019).

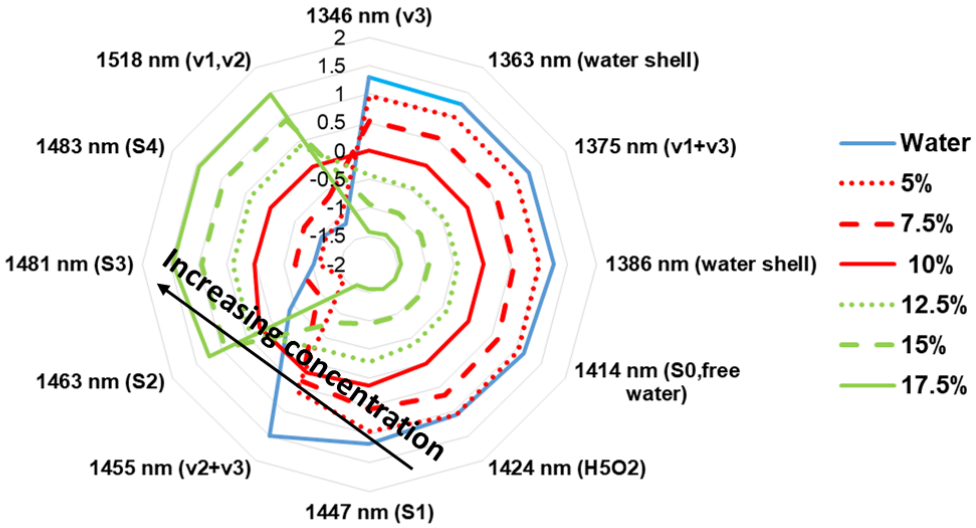
Table 3.1 Summary and assignment of the characteristic water absorbance bands for sucrose concentration

WAMACS	Assignment	Water bands, nm	Activated wavelengths, nm
		first overtone region	First overtone
C1	v_3	1336-1348	1346
C2	OH stretch (water shell)	1360- 1366	1363(picked randomly)
C3	v_1+v_3	1370-1376	1375(picked randomly)
C4	OH stretch (water shell)	1380-1388	1386
C5	S0 (free water)	1398-1418	1414
C6	Water hydration, H_5O_2	1421-1430	1424
C7	S1	1432-1444	1447(close to the band)
C8	v_2+v_3	1448-1454	1455
C9	S2	1458-1468	1463
C10	S3	1472-1482	1481
C11	S4	1482-1495	1483
C12	Strongly bonded water or v_1,v_2	1506-1516	1518 (close to the band)

5) Aquagrams

The derived WASPs presented in the aquagrams, reveal differences between pure water and sucrose solutions with varying concentrations (Fig. 3.6(a),(b)). As the concentration level rises, the number of strongly H-bonded water molecular species (S2, S3, and S4) and highly bonded water increases, with free water states (free water S0) decreasing, indicating more highly organized water structures. The aquagram (Fig. 3.6(b)), becomes more relevant when the water spectrum is subtracted from each of the spectra of the samples in Fig. 3.6(a)) that provides a baseline to the aquagram.

a)



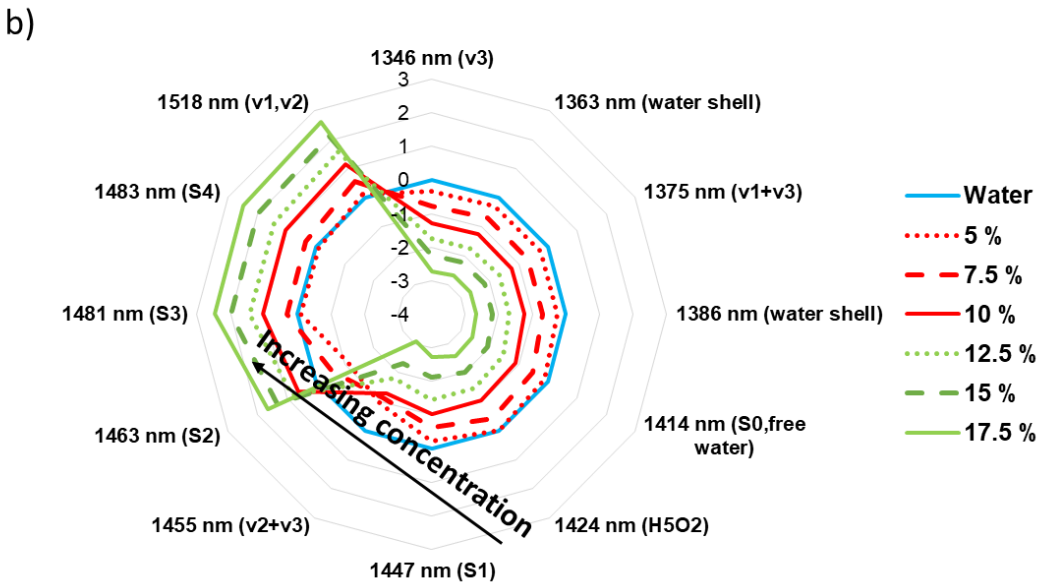


Figure 3.6 a) Aquagram for sucrose solutions at six concentrations and water at 20°C, b) Modified aquagram for the same samples by subtracting water from each sample.

3.4.2. Second overtone analysis (800-1100 nm)

In the second overtone region, spectra of the same samples were measured with 1 mm (Fig. 3.7(a)) and 10 mm (Fig. 3.7(b)) pathlength cuvettes (Jensen & Bak, 2002).

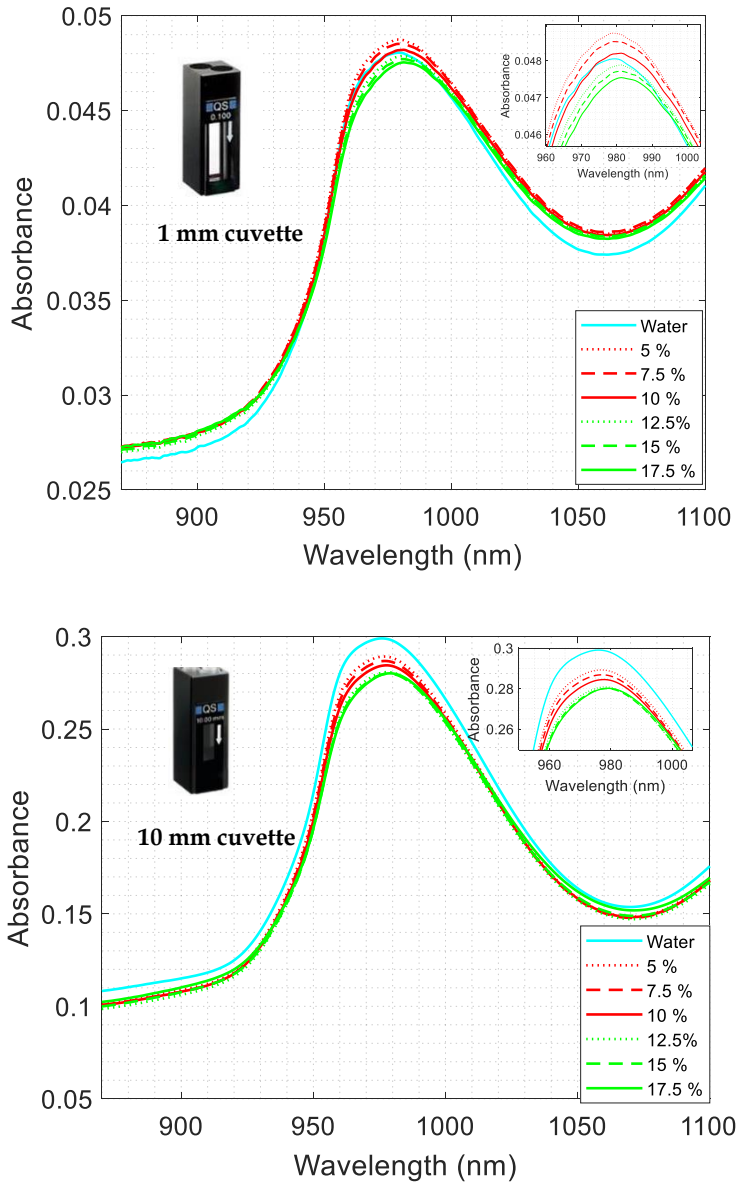


Figure 3.7 Raw absorbance spectra of sucrose solution at six different concentrations and water at 20°C in the second overtone region (870-1100 nm) with a peak at 970 nm for a a) 1 mm pathlength cuvette, b) 10 mm pathlength cuvette.

There is a broad absorbance peak at 970 nm that corresponds to the O-H stretch second overtone of water (Burns & Ciurczak, 1992), decreasing in magnitude with increased concentration (%) when using the 10 mm pathlength cell. The 1 mm cuvette measurements are much more randomly

ordered. This means that sample pathlength plays an important role to get an optimal signal to noise ratio in the first and second overtone regions (Jensen & Bak, 2002).

3.4.2.1 Aquaphotomics Analysis

1) Difference absorption spectra: Raw Sucrose solution – Raw Water

With increasing concentration levels, the difference absorption spectra of the sucrose solutions decreased in the 960 nm region with respect to the zero line (Fig. 3.8). Peaks were observed at 912 nm, 960 nm, and 1059 nm. These peak wavelengths will be later used to select 12 water wavelengths and to plot aquagram.

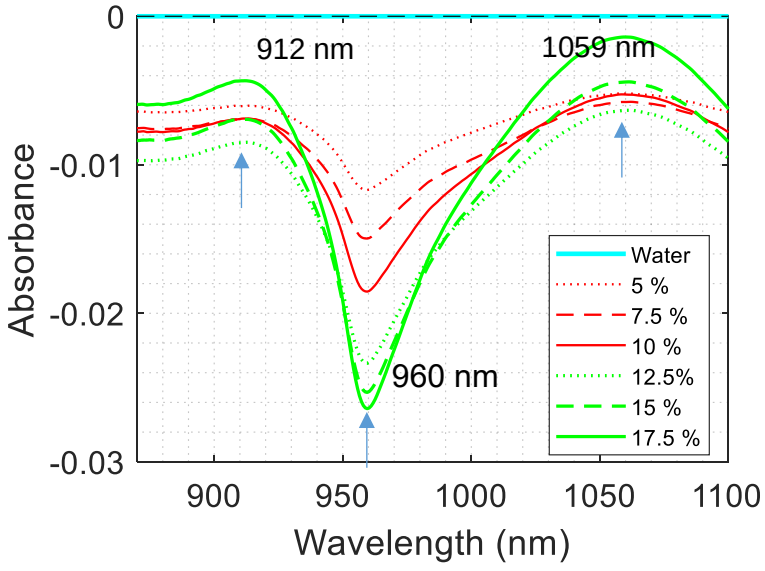


Figure 3.8 Raw absorbance difference spectra of sucrose solutions and water after subtracting the raw absorbance spectrum of water in the second overtone region at 20°C.

2) PCA loadings and scores

Two principal components (PCs), which explain over 99% of the spectral variation, were extracted from the raw absorbance data for the first overtone region (Fig. 3.9). The first PC explains 91% of the variation in the data and shows a peak at 960 nm that is close to water dimer (S1) (Sirinnapa & Sumio, 2008). The second PC explains 9% of the variation and emphasizes the same water band at 960 nm.

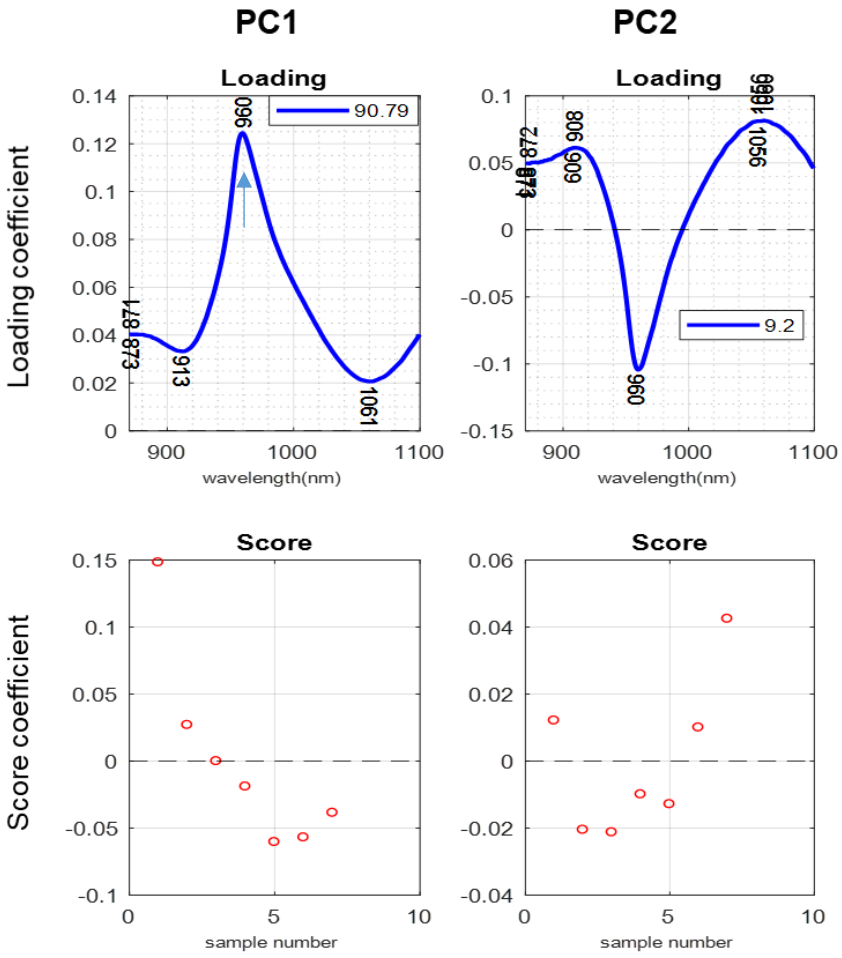


Figure **3.9** PCA loading and score plots of raw absorbance spectra of water and sucrose solutions in the second overtone region. Rectangular boxes signify explained variance in percentage. Labels indicate peak wavelengths in the loading plot.

3) Second derivative spectra of Raw spectra

The second derivative spectra (Fig. 3.10) was used to assist in the identification of further absorption bands, particularly small and/or overlapping absorption peaks not resolvable in the original Raw spectra. The water dimer S1 band is strong at 959 nm and bands corresponding to higher-order H-bond water species, such as dimers (S2) at 979 nm (Sirinnapa & Sumio, 2008). This shape is different from the first overtone region where there are two dips (Fig. 3.4). In the second overtone region there is a sharp downward peak at 959 nm.

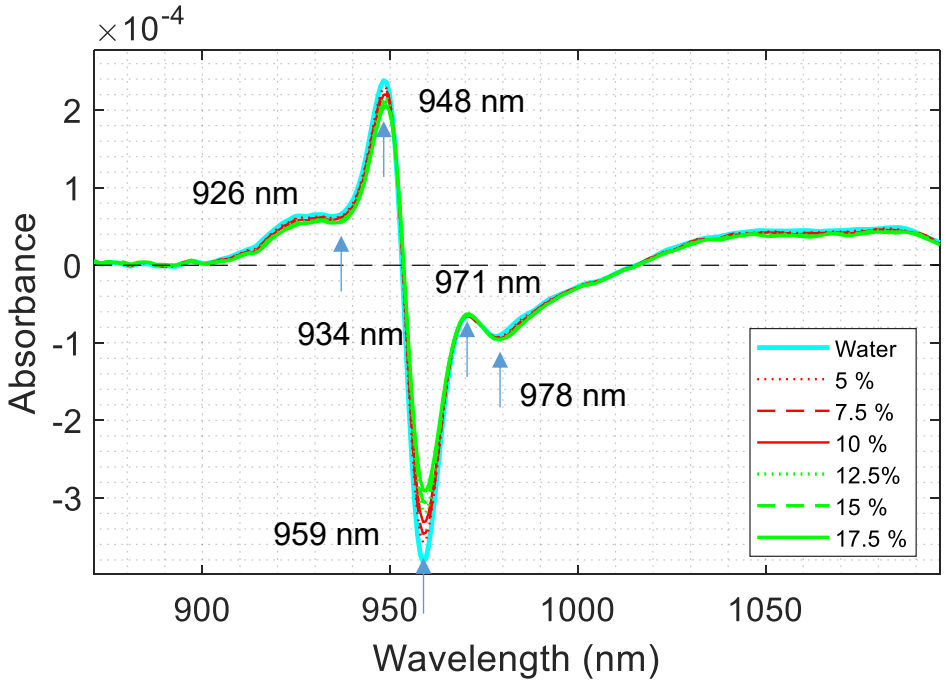


Figure **3.10** Raw second derivative spectra of sucrose solution at six different concentrations and a spectrum of water. Labels show peak wavelengths of the water spectrum.

4) Identification of WAMACS

An aquaphotomics WAMACS in the second overtone region was created using the peak wavelengths found from the peaks/dips of the above-mentioned processing techniques. However, it was found that the WAMACS in the first overtone region of water (1300–1600 nm) were well described in the literature (Tsenkova et al., 2015),very few publications are available that identify water wavelengths in the second overtone region (800–1100 nm). Matija and Tsenkova (2011) generated an aquagram using wavelengths identified by the harmonic oscillator model. In the harmonic oscillator model, the overtone frequencies are calculated from the fundamental frequency and are an integral multiple. If *f* is the fundamental frequency of vibration of a molecule, then the overtone frequency will be calculated as

$$f_{n=}(n+1)f \tag{3.1}$$

where f_n is the n^{th} overtone frequency. Values converted to the respective wavelengths (nm) are shown in column 4 of Table 2.

However, an anharmonic oscillator model should more accurately predict the overtone wavelengths, since it more realistically accounts for the potential energy asymmetry of the molecule binding state. To account for anharmonicity, the overtone frequencies were calculated using the anharmonicity constant, x, whose value is approximately 0.01 (Osborne et al., 1993). The second overtone frequency is not an integral multiple of *f* but can be calculated as

$$f_2 = 3f(1 - 4x) \quad (3.2)$$

This anharmonic oscillator model was used to establish 12 water bands in the second overtone region that correspond to reported wavelengths in the first overtone region of water (Table 3.1). The water bands are identified using the same protocol as described in section 3.3.1.1.(4).

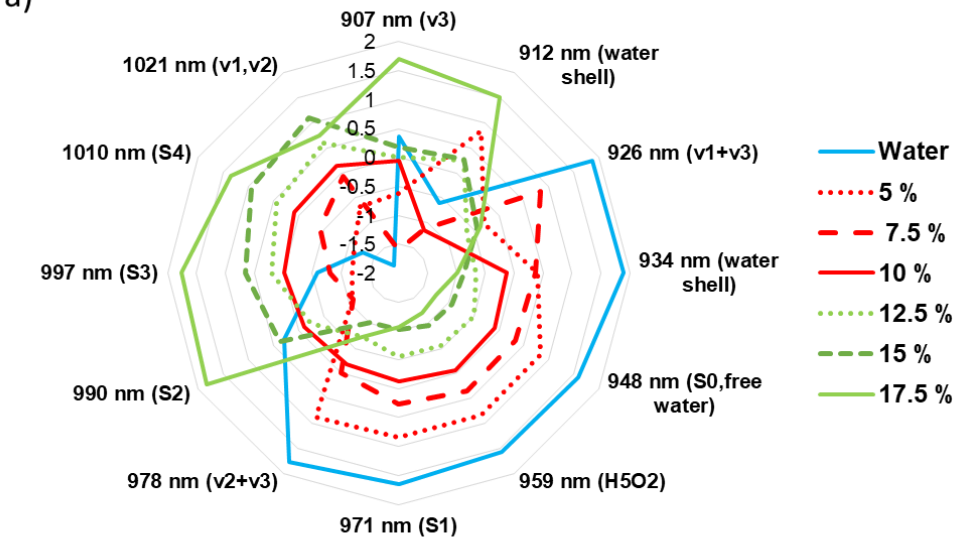
Table 3.1. Summary and assignment of the characteristic water absorbance bands for sucrose solution

WAMACS	**Assignment**	**Water band wavelengths (nm)**			**Calculated water wavelengths (nm)**	
		First overtone region (1300-1600 nm)	Second overtone region (800-1100 nm)		Second overtone region (800-1100 nm)	
			Harmonic oscillator model	Anharmonic oscillator model	Harmonic oscillator model	Anharmonic oscillator model
C1	$v3$	1336-1348	891-899	900-908	897	907
C2	OH stretch (water shell)	1360-1366	907-911	916- 920	907	912
C3	$v1$+$v3$	1370-1376	913-917	923-927	912	926
C4	OH stretch (water shell)	1380-1388	920-925	930- 935	926	934
C5	S0 (free water)	1398-1418	932-945	942-955	934	948
C6	Water hydration, H_50_2	1421-1430	947-953	957-963	948	959
C7	S1	1432-1444	955-963	965-973	959	971
C8	$v2$+$v3$	1448-1454	965-969	975-979	971	978
C9	S2	1458-1468	972-979	982-989	978	990
C10	S3	1472-1482	981-988	992-998	982	997
C11	S4	1482-1495	988-997	998-1007	990	1010
C12	Strongly bonded water or $v1$,$v2$	1506-1516	1004-1011	1014-1021	1010	1021

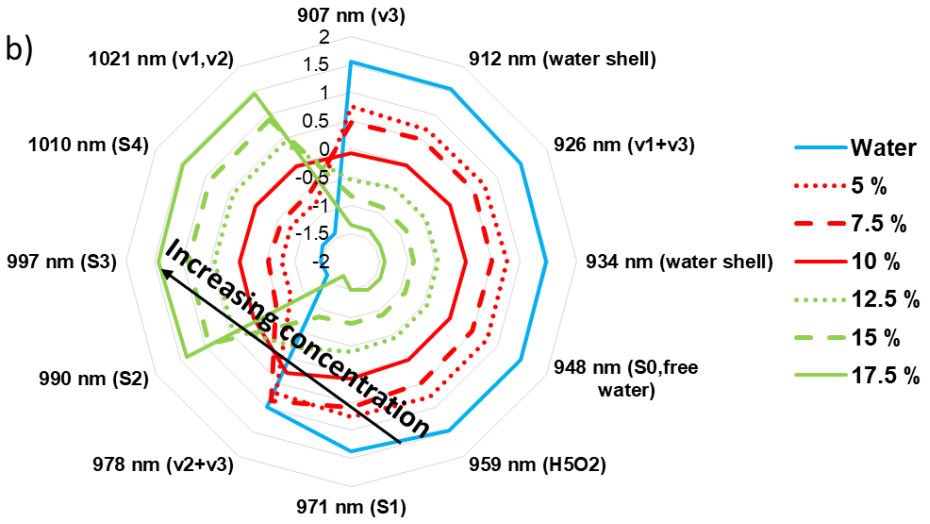
5) Aquagrams

The aquagrams of sucrose solutions in the second overtone region are shown in Fig. 3.11. The longer pathlength cell of 10 mm (Fig 3.11(b) and Fig 3.11(c)) substantially reduces the noise in the WASP. The pathlength of the sample cell is an important factor when making SWNIRS measurement since light is much less absorbed in this region. Moreover, using the anharmonic model for overtone calculation, a better aquagram can be built in the second overtone region (Fig 3.11(b),(d)) that is similar to aquagrams produced in the first overtone region (Fig 3.6(a),(b)). Bonded water species (S2, S3, S4, and (*v1,v2*)) increase with sucrose concentration in the second overtone region.

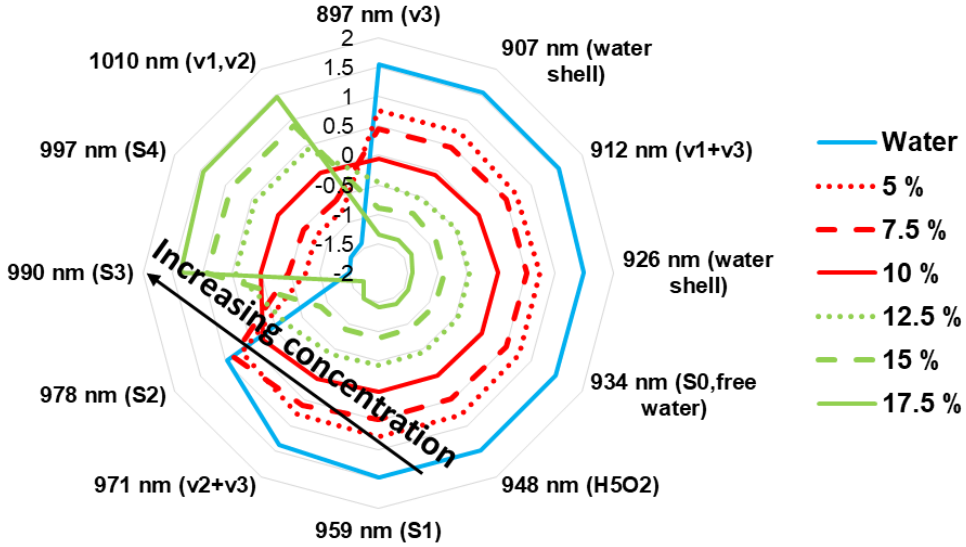
a)



b)


907 nm (v3)
912 nm (water shell)
926 nm (v1+v3)
934 nm (water shell)
948 nm (S0,free water)
959 nm (H5O2)
971 nm (S1)
978 nm (v2+v3)
990 nm (S2)
997 nm (S3)
1010 nm (S4)
1021 nm (v1,v2)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Increasing concentration
Water
5 %
7.5 %
10 %
12.5 %
15 %
17.5 %

c)


897 nm (v3)
907 nm (water shell)
912 nm (v1+v3)
926 nm (water shell)
934 nm (S0,free water)
948 nm (H5O2)
959 nm (S1)
971 nm (v2+v3)
978 nm (S2)
990 nm (S3)
997 nm (S4)
1010 nm (v1,v2)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Increasing concentration
Water
5 %
7.5 %
10 %
12.5 %
15 %
17.5 %

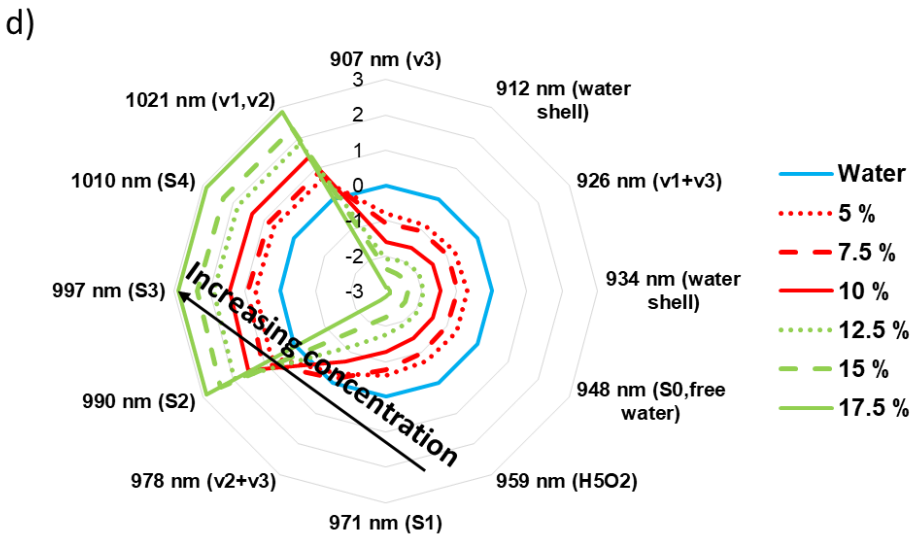


Figure 3.11 Aquagrams for sucrose solutions at six concentrations and water at 20°C with a) 1 mm cuvette and anharmonic model, b) 10 mm cuvette and anharmonic model, c) 10 mm cuvette and harmonic model, and d) 10 mm cuvette and anharmonic model after subtraction of the water spectrum from each sample.

3.5 Conclusion

Using the aquaphotomics, water bands were identified in two overtone regions that represent sucrose activity in water. The same bands were activated in the second overtone region compared with the first overtone region when using anharmonic oscillator model.

Sample path length plays an important role when generating aquagrams. Those built in the second overtone region from 10 mm path length measurements showed a clear water spectral pattern similar to that of the first overtone measurements using a 1 mm cell.

This study shows the use of NIRS and aquaphotomics to identify water bands activated by the presence of sugar in water. The same analysis method will be used in the following chapters to find water bands in fruit juice and intact whole fruit in the two different overtone regions of the OH stretch of water.

Chapter 4: Apple juice (First overtone region and temperature variation)

4.1 Introduction

Consumers have a preference for sweet-tasting apples and their juice (Aprea et al., 2017). Since the soluble solid content (SSC; measured as °Brix) is dominated by soluble sugars that determine sweet taste, so it is an important quality attribute to measure the sugar content (Magwaza & Opara, 2015). Over the last couple of decades, NIRS has been used extensively for rapid and non-destructive quality assessment of various fruits, such as apples, kiwifruit or similar (McGlone et al., 2003; Nicolaï et al., 2007; Lin & Ying, 2009; Kaur et al., 2017b). This technique has also been applied to apple juice for sugar content measurement, detection of quality deterioration during processing/storage, and prediction of adulteration. NIR transflectance spectroscopy in conjunction with chemometrics using the wavelength range 400 to 2500 nm was able to detect fructose corn syrup and added sugars adulteration in apple juice samples with an accuracy of 91–100% (León et al., 2005). Similarly, quality deterioration of 'Pinova' apple juice in different storage states was detected using NIRS in transmittance mode covering the wavelength range from 900 to 1350 nm (Zhu et al., 2011). Transmission NIRS over the range 680 to 1235 nm was used in conjunction with multiple linear regression (MLR) modelling to predict the sugar content of apple juice samples with a standard error of prediction (SEP) of 0.44% (Temma et al., 2002).

Prior studies on apple juice have not definitively elucidated the mechanism by which SSC is estimated or predicted from NIR spectra. Apple juice is 80-85% water, and so its NIR spectrum is dominated by water absorption bands, such as the major peak at 1450 nm in the first overtone region of water (1300–1600 nm) (León et al., 2005; Workman & Weyer, 2007). It is common knowledge that NIR spectra of water solutions are affected by the sugar concentration and, in the past, various water-sugar solutions, such as those involving glucose, fructose, lactose and sucrose, have been studied to learn the effect of sugars on water (Golic, 2003; Giangiacomo, 2006). Apple juice should be little different in that regard. The wavelength region from 1300–1600 nm has also been shown to be a good discriminator for applications such as the identification of honey adulteration, in monitoring the differences in probiotics growth when used for rapid bacteria selection, and in detecting contaminants in water (Gowen et al., 2015; Slavchev et al., 2015; Bázár et al., 2016). Hence, the wavelength region should also be a good discriminator for SSC concentration of apple juice.

The objective of the first part of this chapter is to investigate aquaphotomics concepts to apple juice with varying SSC (°Brix) concentration at a constant temperature (25°C). The natural variation in SSC across a range of juice samples is considered the perturbation effect, a difference from many other aquaphotomics studies involving a series of adulterants or diluents (Tsenkova et al., 2018). The key focus was on the usefulness of the aquagram, as in the water spectral pattern (WASP)

depicting the influence of the main water molecule structural features, in both understandings and in developing predictive models for the SSC of the apple juice samples.

Another part of this chapter focuses on the influence of temperature on water molecular species of apple juice. The NIR spectrum of water is sensitive to temperature variations and so will be the spectrum of fruit and fruit juice since they consist of around 85% water. This often results in bias when a calibration model developed at one temperature is applied to another temperature (Acharya et al., 2013). For example, for intact peaches, the calibration equation developed using samples at one temperature could not reliably predict sample properties at a different temperature (Kawano et al., 1995). In this chapter, the changes in the water structure of apple juice caused by variation of temperature are investigated using aquaphotomics. The work focusses on spectra in the 1300–1600 nm region surrounding the strong water absorption peak at 1450 nm. The aquaphotomics analysis has suggested the possible advantageous use of EMSC pre-processing, using an interferent derived from pure water spectra. This can help to remove the temperature sensitivity of SSC of apple juice NIR calibrations.

4.2 Materials and Methods

4.2.1 Sample preparation

A total of 110 'Braeburn' (*Malus domestica*') apples were purchased from New Zealand retail stores. The juice was expressed from about 1 cm thick slices of outer skin and flesh removed from the two opposite sides of each fruit. The juice samples were collected in Eppendorf tubes and were frozen at -10°C until the optical measurements and reference measurements were to be made. Before those measurements, the frozen samples were first thawed and then centrifuged at 13400 rpm for 3 min (MiniSpin, Eppendorf, Germany) to get clear juice solution (Fig. 4.1).

Milli-Q water was used as a reference sample for both the spectral and the refractive index measurements. The Milli-Q water was produced by a water purification system (Millipore, Thermofisher Scientific, Australia) and had a resistivity of 18.2 MΩ.cm .

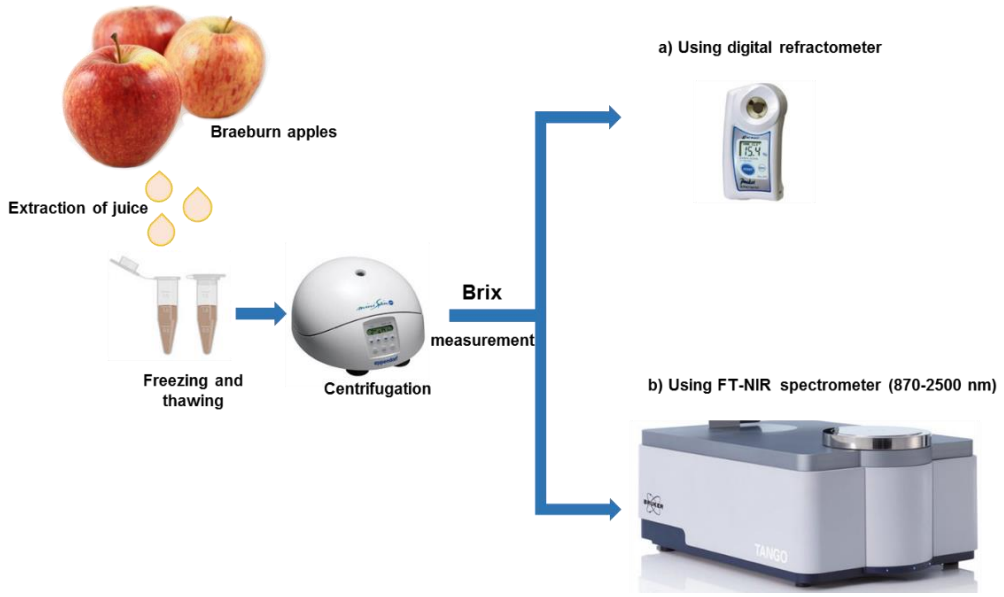


Figure 4.1. Experimental procedure for SSC (°Brix) measurement

4.2.2 Reference SSC (°Brix) measurement

The SSC value of the apple juice samples was measured at room temperature using a digital refractometer (Atago Co. Ltd, Tokyo, Japan), calibrated to zero with Milli-Q water (Vasquez & Mueller, 2019). The Brix value was recorded after placing 0.5 ml of juice into the measurement chamber of the refractometer, enough to fully cover the optical interface.

4.2.3 Spectral acquisition

The transmittance spectra from 870-2500 nm of the apple juice were measured with an FT-NIR spectrometer (Tango, Bruker Corporation, Germany), equipped with a temperature-controlled quartz cuvette (1 mm optical path length) at 20°C, 25°C, and 30°C (±1 °C). Three consecutive

spectra were recorded, each spectrum was the average of 32 spectral scans measured with a resolution of 16 cm^{-1}. The total number of apple juice spectra was 990 (110 samples x 3 consecutive scans x 3 temperatures). A reference spectrum of Milli-Q water was taken after every 11^{th} juice measurement, resulting in a total number of 33 water spectra (11 samples x 3 consecutive scans). After omission of seven samples (four clerical errors, three outliers) the final data set consisted of 103 apple juice samples (927 spectra). The wavelength range above 1800 nm was discarded as there as was no signal due to high absorption of water.

4.2.4 Aquaphotomics analysis

The water bands were calculated using the standard protocol described in Chapter 3. Once the WAMACS wavelengths are selected, WASP presented on the aquagram can be drawn and examined in terms of the system perturbation. The aquagram provides a convenient graphic view for examining the cause and effect of system perturbation in terms of the underlying water structure.

4.2.5 Multivariate analysis

The analysis involved the development of predictive models using either the original measured absorbance spectra (termed Raw) or pre-processed forms of:

- SNV+2D. This was standard normal variate transformation of the raw spectra followed by second derivative processing (Savitzky–Golay 2nd order derivative with smoothing parameters: width 15, order 2).
- EMSC. This was extended multiplicative scatter correction of the raw spectra, as described by equation 2.3 in Chapter 2.
- Combined: This was a combined dataset containing samples at all three temperatures. Predictions were done on an individual temperature dataset. SNV+2D transformation was performed on the calibration and validation datasets.

The PLSR models for SSC prediction were developed using the PLS toolbox version 8.6.2 (Eigenvector Research Inc., Wenatchee, USA) operating under MATLAB version 2018b (Math Works Inc., Natick, USA). Four-fold venetian blind cross validation was applied, to ensure there was no over fitting, with latent variables selected based on the lowest root mean square error of cross validation (RMSECV). A set of two different wavelength ranges was investigated at one temperature of 25°C for model development: 1300–1600 nm (first overtone of water) and 870–1800 nm (whole wavelength range). An SSC predictive model was also developed in the wavelength range 1300–1600 nm using the samples at 20°C for calibration and then applied to samples at 25°C and 30°C for independent SSC prediction.

4.2.6 Statistical Analysis

Four uniquely different calibration-validation sets were obtained through a 4-way leave-each-group-out process that started with splitting the data set into four unique subsets. The samples were first rank ordered by SSC value, indexed from 1 to 4 in sequence, and then split into separate sets based on the index value. Each of the four calibration-validation sets was then created by holding out each subset in turn as an independent validation dataset, leaving the remaining three subsets to be combined as the calibration dataset. Consequently, the total number of samples in each calibration set was 77 and in each validation set 26. The calibration modelling still involved a separate 4-way venetian blind cross-validation process, necessary to determine the optimal number of latent variables.

4.3 Results and Discussion

The SSC of apple juice samples varied from 7.3 to 13.7°Brix with a standard deviation of 1.19°Brix (Table 4.1). For the convenience of the aquaphotomics analysis, the samples were grouped into three classes: 'Low SSC' with 7.3– 9 °Brix, 'Med SSC' with 9–12 ° Brix, and 'High SSC' with 12–13.7 °Brix, with each group containing an equal number of samples.

Table 4.1. Distribution statistics for SSC levels on different SSC class

Sample class	SSC (°Brix)	Mean
Low SSC	7.3-9	8.6
Med SSC	9-12	10.61
High SSC	12-13.7	12.73
Total	7.3-13.7	10.63

4.3.1 The raw Spectra

The average raw absorbance spectra of apple juice samples at the three different SSC levels and three temperatures, plus a pure water spectrum, are shown in Fig. 4.2. The prominent feature is a broad absorbance peak centred around 1451 nm attributed to various water states or species related to the first overtone of the O-H stretch (Burns & Ciurczak, 1992). The peak absorbance was highest for free water and then decreased with increasing SSC. As the temperature increased in apple juice and water samples, the absorbance peak shifted to shorter wavelengths with a broadening of the peak and a decrease in intensity (Fig. 4.2 (b)). The isosbestic point of the water spectra, the wavelength where absorbance does not change with temperature, was at 1444 nm and shifted 2 nm towards the right for the apple juice samples (Workman & Weyer, 2007).

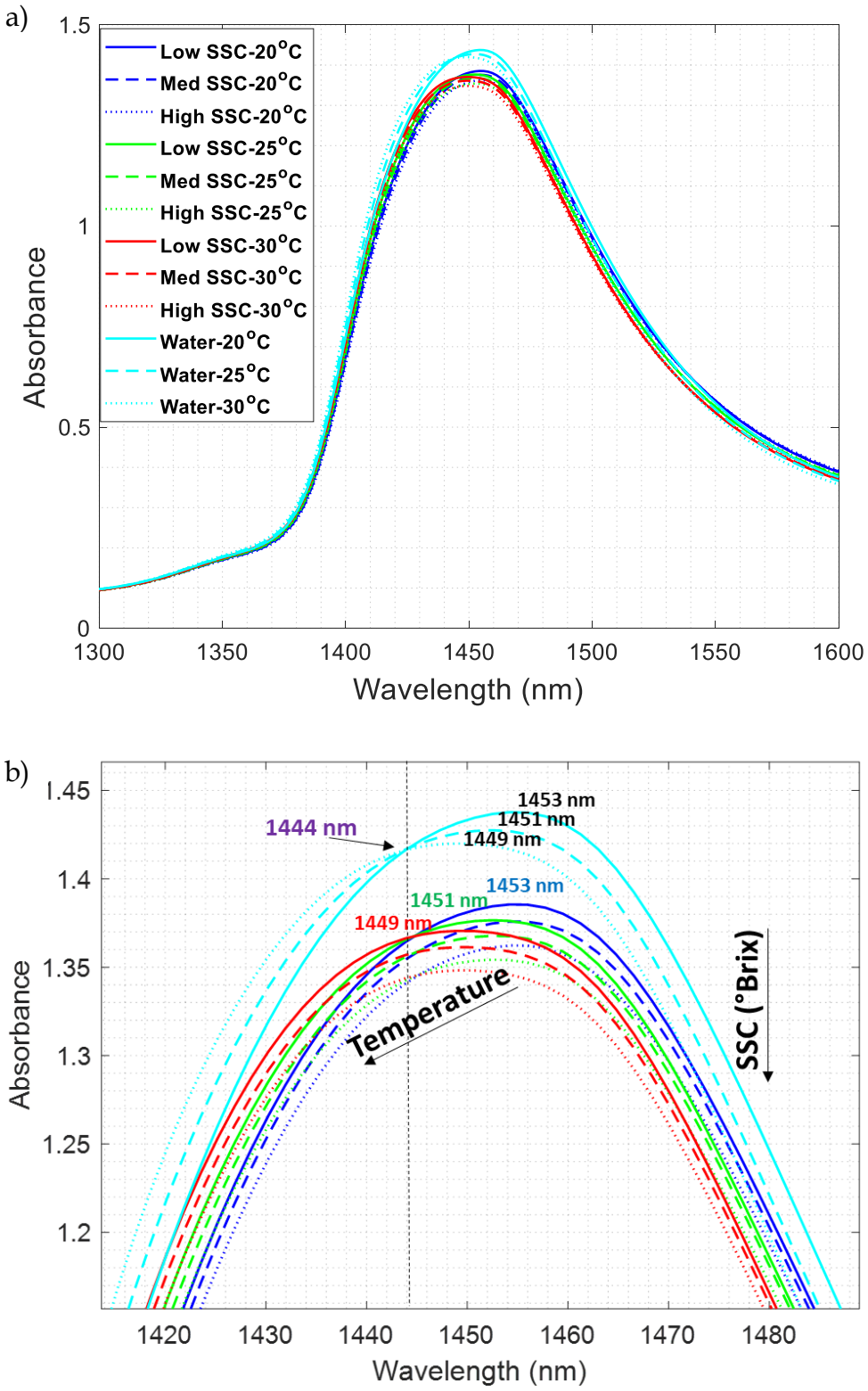


Figure 4.2. Average raw absorbance spectra of apple juice at three temperatures 20, 25, 30°C in a) the first overtone region (1300–1600 nm), b) magnified. Coloured labels indicate the wavelength of the isosbestic point of water (purple) and the peak wavelengths at three temperatures.

4.3.2 Aquaphotomics Analysis

4.3.2.1 Difference absorption spectra: Raw Apple juice – Raw Water

With increasing SSC concentration, the difference absorption spectra of the apple juice solutions decreased at 1412 and 1455 nm, while they slightly increased at 1592, 1689, and 1723 nm with respect to the zero line (Fig. 4.3). Characteristic water bands were observed at 960 nm due to $2v_1$ (symmetric stretching mode) + v_3 (antisymmetric stretching mode), 1151 nm due to v_1 +v_2 (bending mode) +v_3, 1412 nm due to v_1+v_3, and 1455 nm due to v_2+v_3. Some bands were observed at 1592 nm, 1689 nm, and 1723 nm which were within 5 nm to those observed in the difference spectra of aqueous sugar solution (Siesler et al., 2001; Tsenkova, 2002; Bázár et al., 2015).

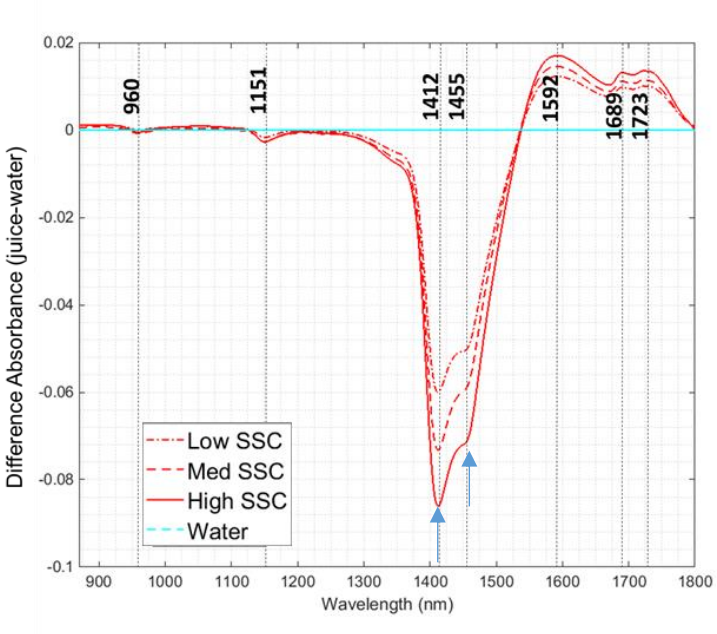


Figure 4.3 Average raw absorbance difference spectra of apple juice samples and water after subtracting average raw absorbance spectrum of water at 25°C

4.3.2.2 PCA loadings and scores

Three principal components (PCs), which explain over 99% of the spectral variation, were extracted from the raw absorbance data for the first overtone region (Fig. 4.4). The first PC (Fig. 4.4 (a)) explains 76% of the variation in the data and shows a single deep trough centred at 1414 nm, attributed to the OH stretch of the free water molecule (S0) (Tsenkova et al., 2015). That first PC demonstrates a strong negative correlation with SSC whereas the second and third PCs have little correlation, explaining 20% and 4% of variation respectively.

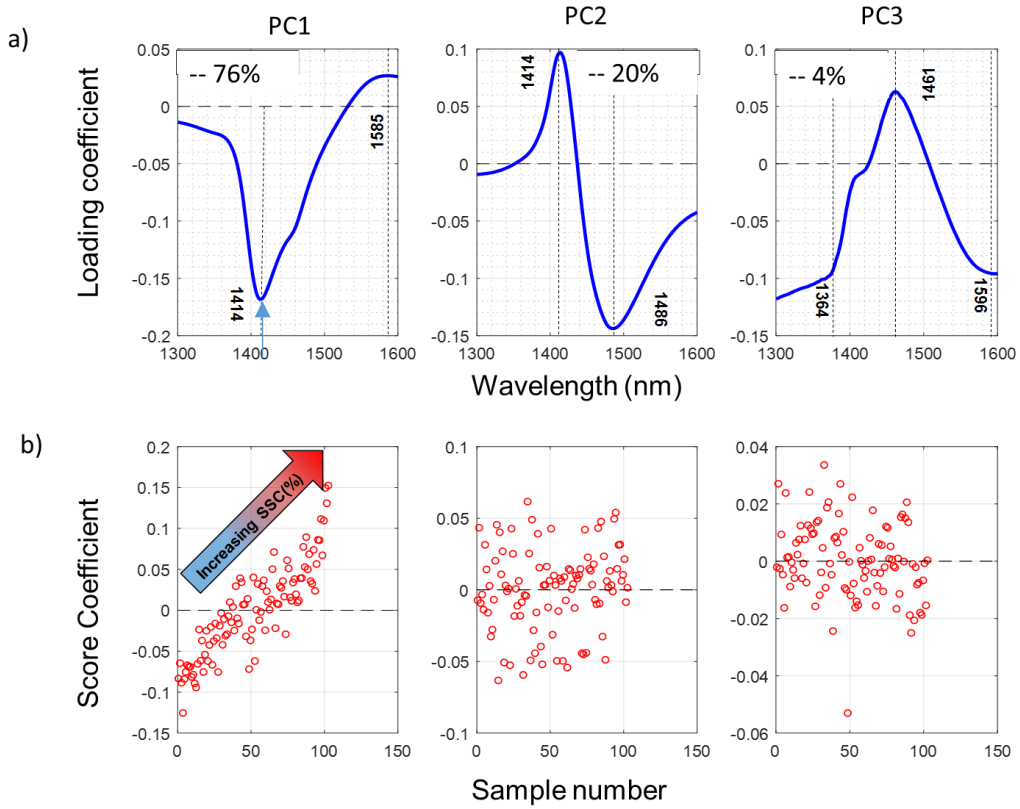


Figure 4.4. PCA a)loading, b)score plots of raw absorbance spectra of apple juice samples arranged in ascending order of SSC at 25°C in the first overtone region of water with explained variance shown at top of loading plots of each PC.

4.3.2.3 Second derivative spectra of raw spectra

The second derivative spectra (Fig. 4.5) was used to assist in the identification of further absorption bands, particularly small and/or overlapping absorption peaks not resolvable in the original raw spectra. The free water S0 band is strong at 1412 nm and bands corresponding to higher order H-bond water species, such as dimers (S2) and trimers (S3) (Segtnan, 2001; Bázár et al., 2015), are attributed to peaks/troughs in the region from 1441–1470 nm. The broad shoulder-like peak at 1518 nm is attributed to even more highly organized water structures with strong H-bonds (Tsenkova et al., 2015). It is noted that not all peaks and/or troughs in a spectrum can be attributed to unique or specific bands since 2nd derivative processing causes mathematical artefacts to appear, such as side lobe peaks around main absorption peaks.

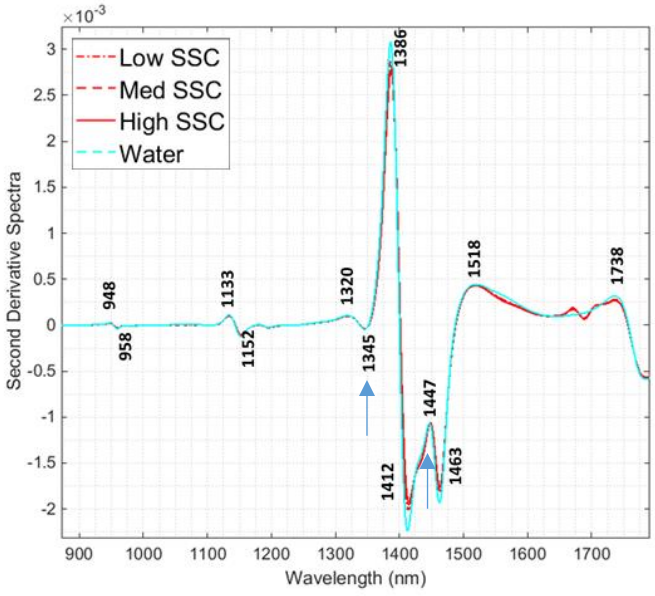


Figure 4.5 RAW second derivative spectra of apple juice at three different SSC levels and average spectrum of water at 25°C. Labels show peak wavelengths of the water spectrum.

4.3.2.4 PLS regression vector

PLS models for SSC prediction were created using the raw absorption spectra over two different spectral regions (Table 4.3). The PLS regression vectors were characterized by many small but sharply defined peaks and dips in the 1300–1600 nm region, resolved at differences of only a few nanometers (Fig 4.6). The dominant feature was a particularly strong dip at 1454 nm which corresponds closely with the peak absorption observed in the raw spectra (Fig. 4.2) (Chatani et al., 2014; Tsenkova et al., 2015). All the identified peaks, whether large or small, were considered in the assignment of the WAMACS.

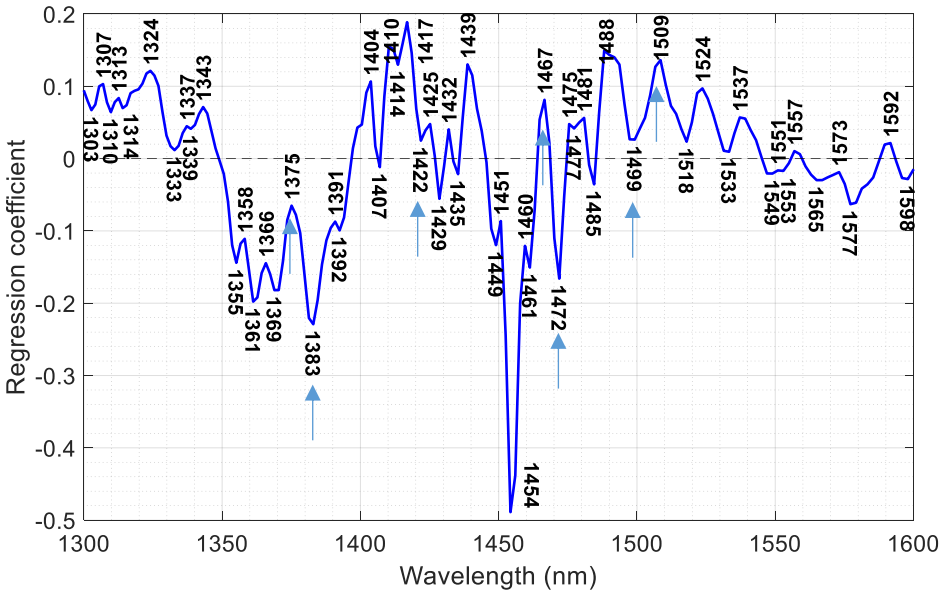


Figure 4.6. PLS regression vector coefficients for SSC (°Brix) prediction of the apple juice samples at 25°C in the first overtone region of OH stretch of water.

4.3.2.5 Pure Water Analysis

There is a clear variation with temperature in the peak wavelength of the water spectra (Fig. 4.2 (b)). When applying PCA to the water spectra, the shape of the PC1 loading (Fig. 4.7(b)) is very similar to that reported by Segtnan et al. (Segtnan, 2001) and Maeda et al. (Maeda et al., 1995), who studied the effect of temperature on the NIR spectra of water. In their work, the temperature of the water was varied between 5°C and 80°C, and PCA was applied to the resulting spectra. In our case, the shape observed in PC 1 loading (Fig. 4.7(b)) suggests a change in water structure due to a change in temperature since the corresponding PC 1 score (Fig. 4.7(a)) increases with temperature. Hence, the PC1 loading was used in the EMSC correction (equation 2.3 in Chapter 2) as an interferent spectrum to correct juice spectra against temperature variation.

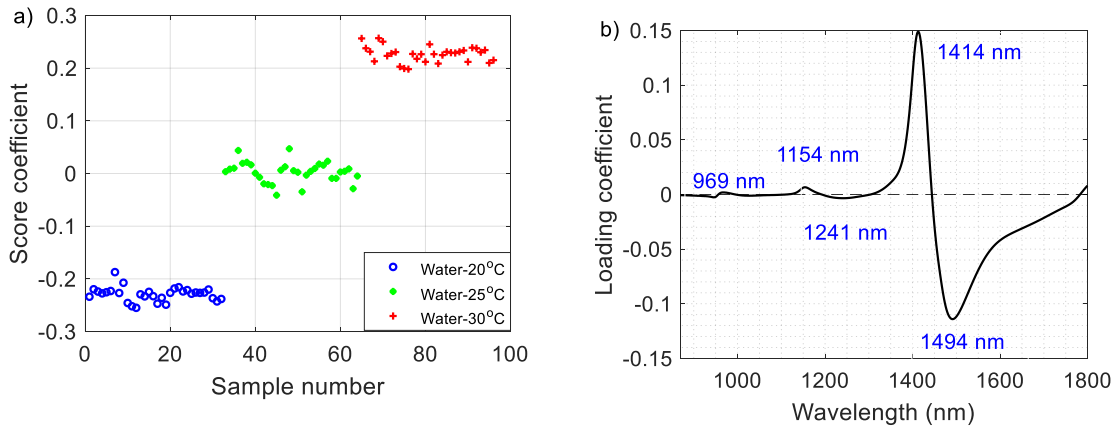


Figure **4**.7 PC1 a) score, and b) loading of water in the whole wavelength range from 870–1800 nm. Labels indicate the peak wavelengths.

4.3.2.6 Identification of WAMACS

The wavelengths picked for the display on the aquagrams were selected based on the wavelengths found to be important during previous analysis (Fig. 4.3, 4.4, 4.5, and 4.6) and they all correspond closely to water coordinates reported in the literature (Tsenkova et al., 2015). The wavelengths selected for each WAMACS coordinate were (Table 4.2) : C1: 1345 nm from the raw second derivative spectra; C2: 1364 nm from the third PC of PCA; C3: 1375 nm from the PLS regression coefficient; C4: 1383 nm from the PLS regression coefficient; C5: 1414 nm from the first PC of PCA; C6: 1422 nm from the PLS regression coefficient; C7: 1447 nm from the raw second derivative spectra; C8: 1455 nm from the raw absorbance difference spectra; C9: 1467 nm, C10: 1472 nm, C11: 1499 nm; C12: 1509 nm were all from the PLS regression coefficient. There were no gaps or confusions evident in the selections; the wavelength coordinates all matching well with the wavelength scheme previously reported (Tsenkova et al., 2015; Muncan & Tsenkova, 2019).

Table 4.2 Summary and assignment of the characteristic water absorbance bands for SSC of apple juice

WAMACS	Assignment	Water bands	Activated wavelengths, nm
		First overtone region	First overtone
C1	v_3	1336-1348	1345
C2	OH stretch (water shell)	1360-1366	1364
C3	v_1+v_3	1370-1376	1375
C4	OH stretch (water shell)	1380-1388	1383
C5	S0 (free water)	1398-1418	1414
C6	Water hydration, H_5O_2	1421-1430	1422
C7	S1	1432-1444	1447
C8	v_2+v_3	1448-1454	1455
C9	S2	1458-1468	1467
C10	S3	1472-1482	1472
C11	S4	1482-1495	1499
C12	Strongly bonded water or v_1,v_2	1506-1516	1509

4.3.2.7 Aquagrams

The derived WASPs patterns presented in the aquagrams show differences between pure water and apple juice samples with varying SSC concentrations (Fig. 4.8). As the SSC level rises, the number of strongly H- bonded water molecular species (S2, S3, and S4) increases, with free water states (free OH and S0) decreasing, indicating more highly organized water structures (Fig. 4.8(a)). Conversely, as the temperature rises, the free water (S0) increases as the water structure becomes less organized with the increased molecular motion and less stable H-bonds (Fig. 4.8(b)). The strongly H-bonded water molecular species (S2, S3, and S4) increase within each temperature group (Fig. 4.8(b)) as seen previously at 25°C (Fig. 4.8(a)).

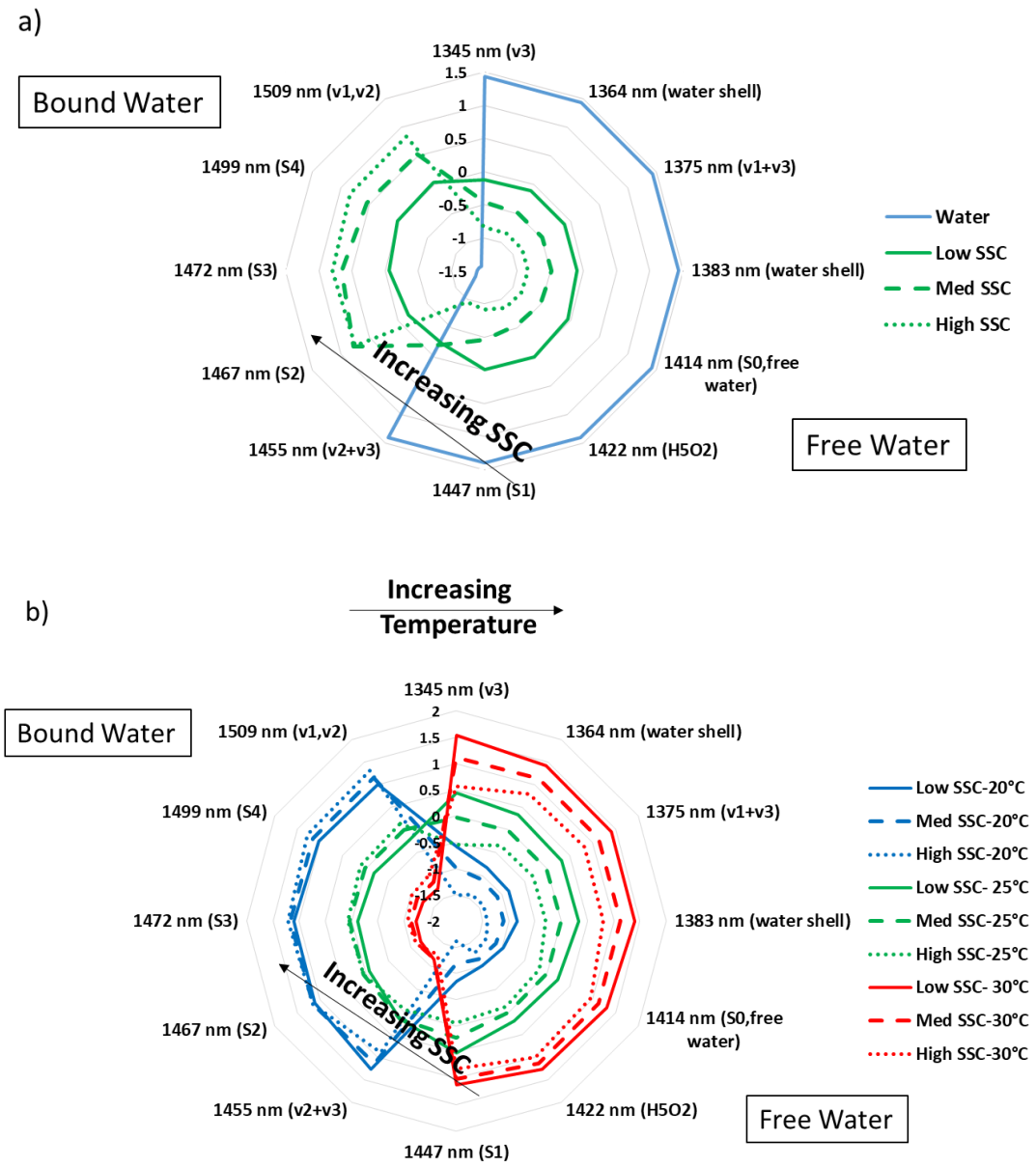


Figure 4.8. Aquagrams for a) apple juice at three SSC levels and water at 25°C b) for apple juice at three temperatures for low, medium, and high SSC levels.

4.3.3 Prediction of SSC

All PLSR models developed for SSC prediction, except those using EMSC pre-processing, proved to be more or less equivalent and yielded SEP values between 0.38 and 0.41°Brix (Table 4.3). Using only the narrow first

overtone range (1300–1600 nm) was as good as using the whole spectrum (870–1800 nm). This suggests that most of the relevant information on the SSC concentration is available in the absorption region of the first overtone of water, a finding which agrees with the literature predicting sugars (Bázár et al., 2015), salts (Gowen et al., 2015), or detection of honey adulteration (Bázár et al., 2016) and pesticides (Gowen et al., 2011b).

Table 4.3. Comparison of various pre-treatments and wavelength regions for SSC (°Brix) prediction of apple juice using PLSR at a constant temperature of 25°C

(N_{cal}=77 and N_{val}=26)								
Pre-processing	λ range (nm)	R^2_{cv}	RMSECV	LV	R^2_p	RMSEP	Bias	SEP
Raw	**1300-1600** (First overtone region of water)	0.87	0.41 (± 0.02)	4	0.88	0.41 (±0.04)	0.07 (± 0.05)	0.41 (± 0.04)
SNV+2D		0.87	0.40 (± 0.02)	3	0.89	0.39 (± 0.04)	0.06 (± 0.04)	0.38 (± 0.04)
Raw	**870-1800** (Whole region)	0.87	0.42 (± 0.03)	5	0.87	0.43 (± 0.05)	0.08 (± 0.06)	0.41 (± 0.05)
SNV+2D		0.88	0.40 (± 0.03)	3	0.88	0.40 (± 0.06)	0.07 (± 0.04)	0.39 (±0.06)

The aquaphotomics analysis, particularly the strong and consistent change of the aquagram patterns with temperature (c.f., Fig. 4.2(b)), suggested the use of EMSC pre-processing using PC1 (Fig. 4.7(b)) from the pure water-temperature change matrix as the interference component (Eqn. 2.3, Chapter 2). The spectra at the different temperatures collapsed to near-identical spectral patterns following the EMSC processing (Fig. 4.9). There was significantly improved model performance following EMSC pre-

processing. The SSC prediction bias, observed when applying the 20°C calibrated model to samples at different temperatures, was reduced to near negligible levels. Compared to SNV+2D pre-processing only, the bias dropped from 0.23 to 0.08 and 0.36 to 0.13°Brix for predicting the SSC at 25 and 30°C, respectively (Table 4.4). Similar results were achieved using a combined dataset model that had samples from all three temperatures. In addition, the model precision (SEP) was slightly improved by about 0.02°Brix in each case compared with SNV+2D method. The EMSC interference term, the PC1 generated from the pure water-temperature spectral matrix, is independent of the apple juice samples which considerably simplifies generating a temperature independent model. It does not require apple juice samples to be measured at different temperatures – measurements at one temperature will suffice. This approach may be applicable to other fruit juice or intact fruit modelling problems where robustness against temperature changes is desirable.

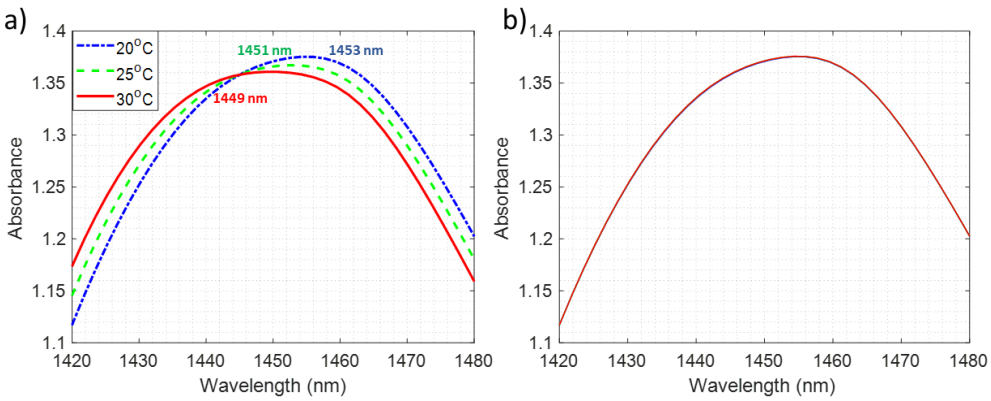


Figure 4.9. a) Uncorrected and b)EMSC corrected apple juice spectra generated by removing PC1 loading component of the water spectrum from apple juice spectra. Labels indicate the wavelengths of the peaks.

Table 4.4 Comparison of SNV+2D and EMSC pre-processing methods in the wavelength range of 1300–1600 nm

Cal= 20°C	(N_{cal}=77 and N_{val}=26)							
Val=	**Pre-processing**	**R^2_{cv}**	**RMSECV**	**LV**	**R^2_p**	**RMSEP**	**Bias**	**SEP**
20°C	SNV+2D	0.88	0.39(± 0.02)	3	0.88	0.41(± 0.08)	0.06(± 0.05)	0.40(± 0.08)
	EMSC	0.89	0.39(± 0.01)	2	0.90	0.38(± 0.04)	0.06(± 0.05)	0.37(± 0.04)
	Combined	0.87	0.41(± 0.01)	3	0.89	0.40(± 0.05)	0.11(± 0.04)	0.38(± 0.05)
25°C	SNV+2D				0.88	0.46(± 0.10)	0.23(± 0.11)	0.40(± 0.05)
	EMSC				0.89	0.39(± 0.05)	0.08(± 0.04)	0.38(± 0.05)
	Combined				0.89	0.39(± 0.05)	0.08(± 0.03)	0.39(± 0.05)
30°C	SNV+2D				0.84	0.63(± 0.21)	0.36(± 0.32)	0.46(± 0.10)
	EMSC				0.85	0.47(± 0.10)	0.13(± 0.12)	0.44(± 0.11)
	Combined				0.85	0.48(± 0.10)	0.13(± 0.12)	0.45(± 0.10)

4.4 Conclusions

The aquaphotomics study here has delivered insight into the free and bonded water structure changes, as evident in the 1300–1600 nm spectral range (first overtone region for water absorbance), occurring with changes in SSC (7.3 to 13.7°Brix) and temperature (20, 25, and 30°C). The free water components increase, the bonded water components decrease, as temperature rises or SSC decreases. The absorbance bands of water which were activated in response to perturbation by temperature and SSC were clearly identified and they all fall within the ranges of 12 WAMACS in the first overtone of OH stretch of water. Incorporation of the EMSC pre-processing method, with PLSR modelling, significantly reduced prediction

bias when a model calibrated at one temperature (20°C) was applied to samples at different temperatures (25 and 30°C). The EMSC method used as an interference spectrum the PC1 loading vector generated from a spectral matrix for pure water measured over the same temperature range. The EMSC method with such a PC1 loading vector may well have application advantages for other fruit juice or intact fruit measurement modelling problems where robustness against temperature changes is desirable.

Chapter 5: Kiwifruit juice (first and second overtone regions with temperature variation)

5.1 Introduction

As the water peaks at 1450 nm and 970 nm in fruit juice shift due to temperature variation, aquaphotomics was investigated to learn more about changes in water structure caused by temperature (Tsenkova et al., 2015). In this chapter, calibration models were constructed for SSC (°Brix) prediction in kiwifruit juice using an FT-NIR spectrometer and PLSR with EPO correction. The water bands related to temperature variation in the first and second overtone regions for kiwifruit juice samples were identified and reported.

5.2 Methods and materials

5.2.1 Sample preparation

A total of 100 Gold Kiwifruit *"Actinidia chinensis"* were purchased from New Zealand retail stores. The juice was expressed from about 2 cm thick slices of outer skin and flesh removed from the two opposite endcaps of each fruit, collected in Eppendorf tubes and then filtered through 0.2 µm filter paper to produce a clear juice. The samples were stored in a refrigerator at 4°C. FT-NIR analysis and reference measurements were performed the next day after the samples were equilibrated to room temperature (20°C). For reference data, the SSC of the kiwifruit juice was measured using a digital refractometer (Atago Co. Ltd, Tokyo, Japan).

Milli-Q water was used as a reference sample for both the spectral and the refractive index measurements. The Milli-Q water was produced

using a water purification system (Millipore, Thermofisher Scientific, Australia) and had a resistivity of 18.2 MΩ.cm.

5.2.2 Reference SSC (°Brix) measurement

The SSC value of the kiwifruit juice samples was measured at room temperature using a digital refractometer (Atago Co. Ltd, Tokyo, Japan), calibrated with Milli-Q water (Vasquez & Mueller, 2019). The Brix value was recorded after placing 0.5 ml of juice into the measurement chamber of the refractometer; this was enough to fully cover the optical interface.

5.2.3 Spectral acquisition

Transmittance spectra of the juice samples were measured at 20, 25, and 30°C (±1°C) with an FT-NIR spectrometer (Tango, Bruker Corporation, Germany) equipped with a temperature-controlled holder. Two measurements were acquired for each juice sample, using quartz cuvettes of 1 mm and 10 mm optical path length for the 1300-1600 nm and 870-1100 nm wavelength ranges, respectively (Kaur et al., 2018). For each measurement, one spectrum was the average of 32 successive scans and was recorded with a resolution of 16 cm^{-1}. The total number of juice spectra was 600 (100 samples x 1 consecutive scan x 3 temperatures x 2 cuvettes). A reference spectrum of Milli-Q water was taken after every 11^{th} juice measurement, resulting in a total number of 11 water spectra (11 samples x 1 consecutive scans). The samples were divided into two sets; one each for the 1300-1600 nm and 870-1100 nm wavelength ranges. After removal of

five samples because of clerical errors, the final data set consisted of 95 juice samples (285 spectra for three temperatures) in each wavelength set. The spectral region above 1800 nm was discarded due to the high absorption in aqueous samples.

5.2.4 Aquaphotomics analysis

An aquaphotomics WAMACS was created using the peak wavelengths identified from PCA of the full data set in the first overtone region and in the second overtone region. An anharmonic oscillator was used to establish 12 water bands in the second overtone region that corresponded to reported wavelengths in the first overtone region of water (Osborne et al., 1993). Aquagrams displaying the resulting WASP in each wavelength region were then studied to observe temperature variation.

5.2.5 MultivariateAnalysis

Predictive models were developed using MATLAB version R2018b (Math Works Inc., Natick, USA) and the PLS toolbox version 8.6.2 (Eigenvector Research Inc., Wenatchee, USA) with four-fold Venetian blind cross-validation applied. The samples were rank-ordered according to their SSC values. The analysis involved the development of predictive models using either the original measured absorbance spectra (termed Raw) or pre-processed forms of EPO correction (Roger et al., 2003).

The main data set in each wavelength range was split into three temperature subsets for 20°C, 25°C, and 30°C, respectively. Sixteen juice

samples (covering low to high SSC) were separated for the EPO correction. This data set contained absorbance spectra of juice at three temperatures. In addition, a set of independently measured water spectra at three temperatures were taken to conduct the same EPO correction. The calibration and validation set consisted of 63 and 16 samples, respectively, selected using every fifth sample forming the validation set and the remaining samples in the calibration set. To conduct EPO correction, a difference matrix, D, was generated by subtracting the average spectrum of the 20°C juice samples from the average spectrum of samples at 20, 25, and 30°C (explained in Chapter 2, section 2.6.2). The same procedure was applied to the water temperature spectra to generate a difference matrix using independently measured water samples at the three temperatures. Then, PCA was performed on the difference matrix D generated with both juice and water samples. The first PC loading of the D matrix for both juice and water was chosen (Fig. 5.6) as it showed a strong correlation with temperature. The PC1 loading (hereafter referred to as 'G') was further used for correcting the calibration and validation spectra. A factor ($I-GG^T$) was multiplied with both the calibration and validation sets, where I is the identity matrix and G^T is the transpose of G. A calibration model was built from the samples at 30°C, which was later applied to the samples at 20°C for SSC prediction in each wavelength range (Fig. 5.8). The EPO corrected calibration and the validation sets were then used for developing models unaffected by temperature.

5.3 Results and discussion

The SSC of kiwifruit juice ranged from 11.9 to 19.2°Brix, with a standard deviation of 1.26°Brix. Fig. 5.1 shows the distribution of SSC for all fruit in the experiment. There were fewer samples in the lower range of 12-14°Brix in the overall sample population, with a higher number of samples in the 16-18°Brix region. 14-16°Brix and >18°Brix groups contained nearly equal number of samples less than 16-18°Brix region.

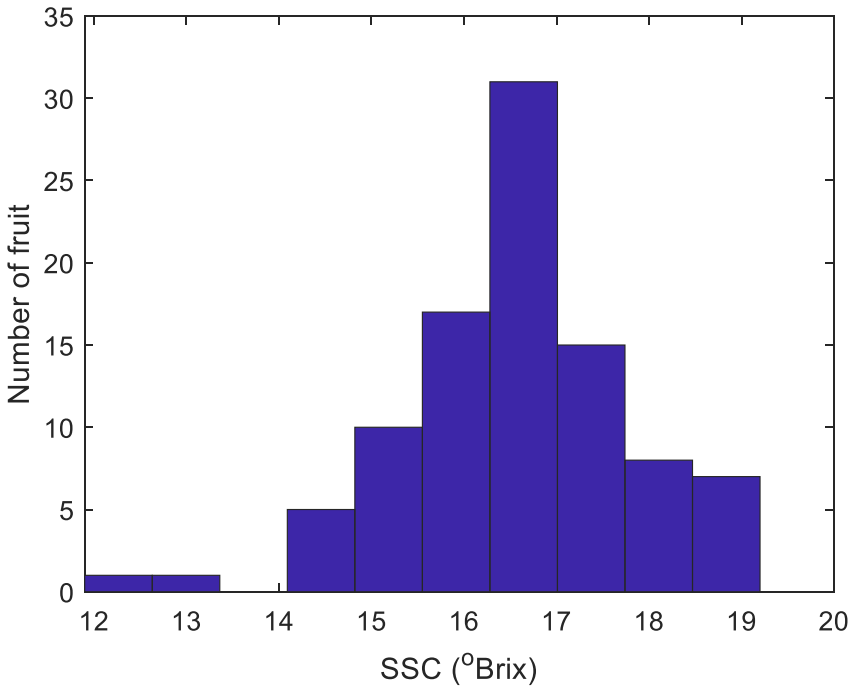


Figure 5.1: Number of fruit vs SSC (°Brix) for the kiwifruit juice samples.

5.3.1 The raw spectra

The absorbance plot in Fig. 5.2 illustrates that as the temperature increased in kiwifruit juice, the absorbance curve in the first overtone region shifted to shorter wavelengths with a broadening of the peak and a decrease in intensity (Fig. 5.2 (a)). However, in the second overtone region, there was a slight upward shift in intensity towards the shorter wavelengths with

increasing temperature (Fig. 5.2 (b)). The isosbestic points were 1444 nm (before the water peak wavelength) and 994 nm (after the water peak wavelength), respectively, approximately 4 nm away from the known water isosbestic points (Workman & Weyer, 2007).

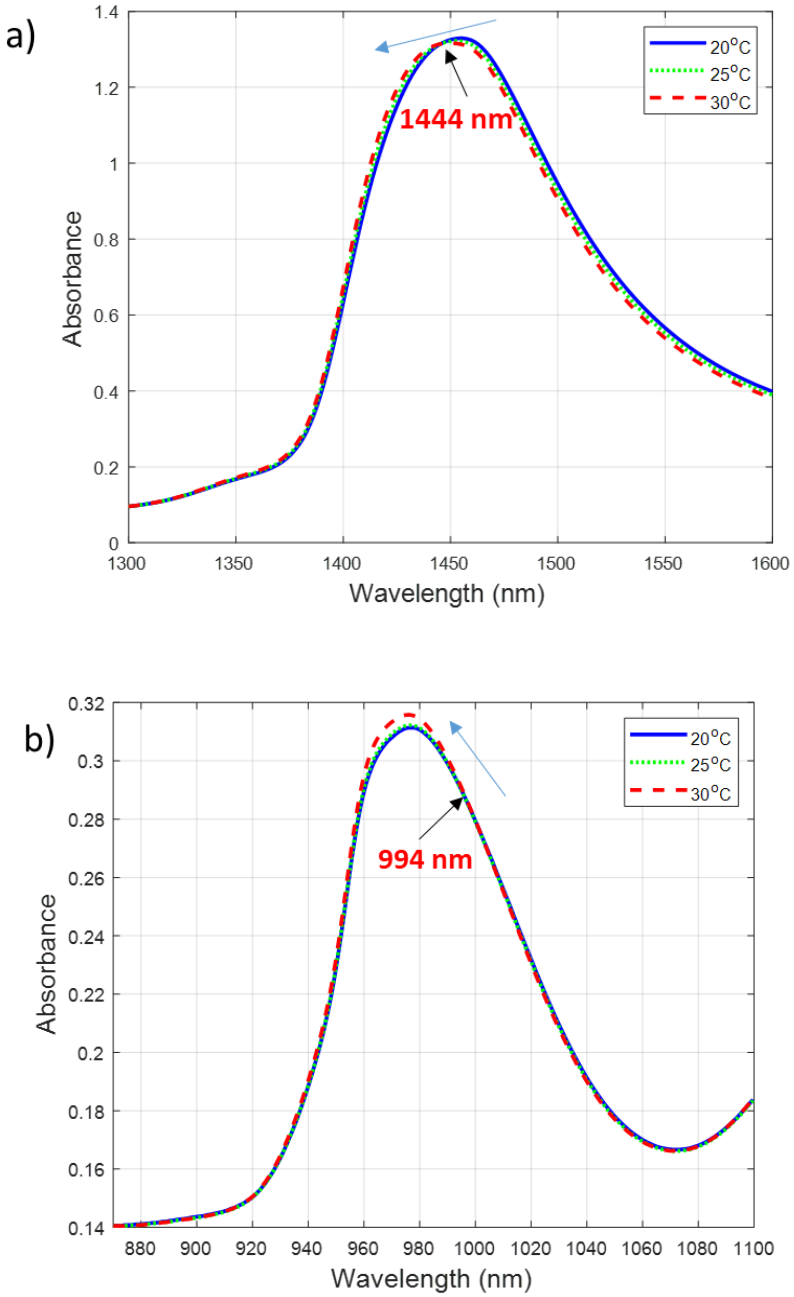


Figure 5.2: Average raw absorbance spectra of kiwifruit juice at three temperatures 20, 25, 30°C in a) the first overtone (1300-1600 nm); and b) the second overtone (870- 1100 nm) region of the OH stretch of water. Labels indicate wavelengths of the isosbestic points (the wavelengths where absorbance does not change with temperature).

5.3.2 Identification of WAMACS

The peak and trough wavelengths of the PC1 spectrum (PCA on all juice samples and temperatures) were at 1414 nm and 1494 nm (the first overtone), and 963 nm and 1027 nm (the second overtone), as shown in Figure 5.3(a) and 5.3(b). These respective wavelength pairs were chosen to be wavelengths very sensitive to temperature changes in juice (Table 5.1). As expected, the zero-crossing points for the two PC1 plots were identical to the isosbestic points observed in Figure 5.2. In the first overtone, C5: S0 (free water) and C11: S4 (species with four hydrogen bonds) were activated, whereas C6: water hydration and C12: strongly bonded water were activated in the second overtone region. Those WAMACS not identified using PCA were chosen randomly from within each water band given in Table 5.1.

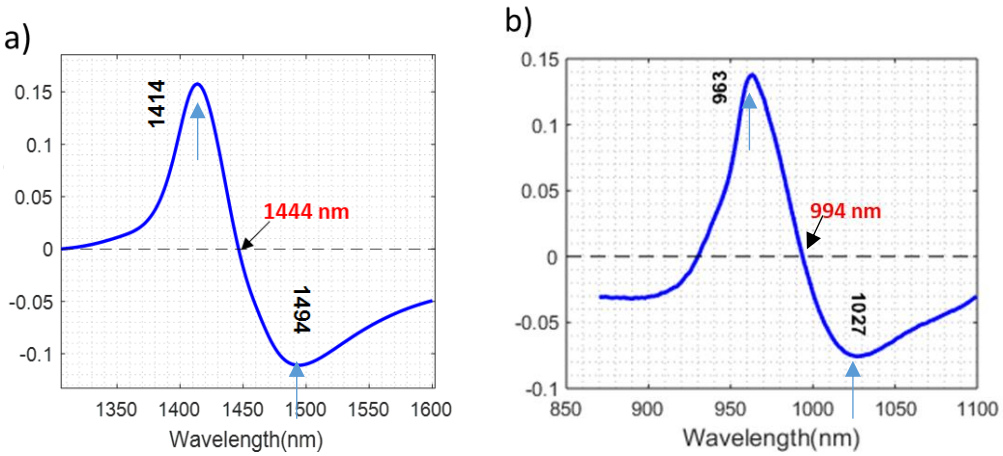


Figure 5.3: PC1 loading of kiwifruit juice in a) the first overtone (1300–1600 nm); and b) the second overtone (870–1100 nm) region of the OH stretch of water. Labels indicate peak wavelengths (black) and zero-crossing points (red).

Table 5.1. Temperature-perturbed water wavelengths of kiwifruit juice in the first ((Tsenkova et al., 2015; Muncan & Tsenkova, 2019)) and the second (Chapter 3) overtone regions of water

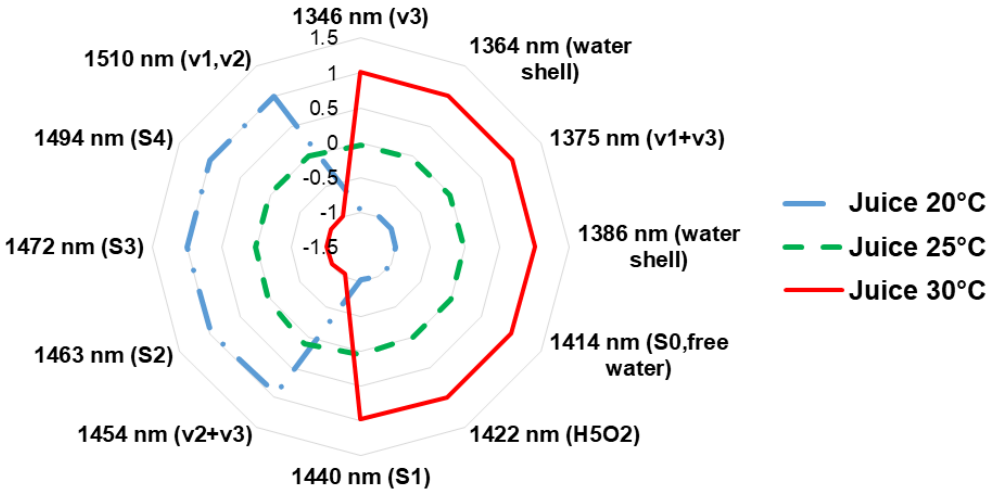
WAMACS	Assignment	Wavelengths in overtone region		Activated wavelengths, nm	
		First (1300-1600 nm)	Second (800-1100 nm)	First overtone	Second overtone
C1	$v3$	1336-1348	900-908		
C2	OH stretch (water shell)	1360- 1366	916- 920		
C3	$v1$+$v3$	1370-1376	923-927		
C4	OH stretch (water shell)	1380-1388	930- 935		
C5	S0 (free water)	1398-1418	942-955	1414	
C6	Water hydration, H_50_2	1421-1430	957-963		963
C7	S1	1432-1444	965-973		
C8	$v2$+$v3$	1448-1454	975-979		
C9	S2	1458-1468	982-989		
C10	S3	1472-1482	992-998		
C11	S4	1482-1495	998-1007	1494	
C12	Strongly bonded water or $v1$,$v2$	1506-1516	1014-1021		1027

5.3.3 Aquagrams

The aquagrams of average spectra of the juice at three temperatures are illustrated in Fig. 5.4(a) and Fig. 5.5(a) for the two overtone regions. Free water species increased with temperature in the two overtone regions. However, the asymmetric stretching and bending (v_2+v_3) can only be seen to increase with temperature for the second overtone region. As the temperature rises in juice and water, the free water (S0) increases as the water structure becomes less organized as a result of increased molecular motion and less stable H-bonds (Fig. 5.4, 5.5). Different bands were seen to change in the second overtone region compared with the first overtone region. The expectation was that the activations would be the same in both

regions; however, the findings were different. Similar aquagrams were observed for kiwifruit juice and water, with both showing the same trends (Figs. 5.4 and 5.5).

a) Kiwifruit Juice



b) Water

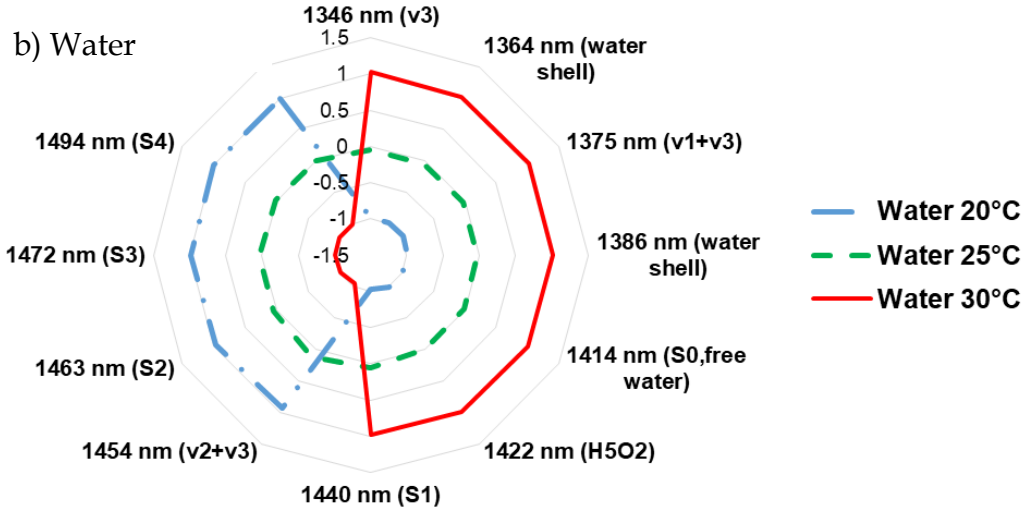
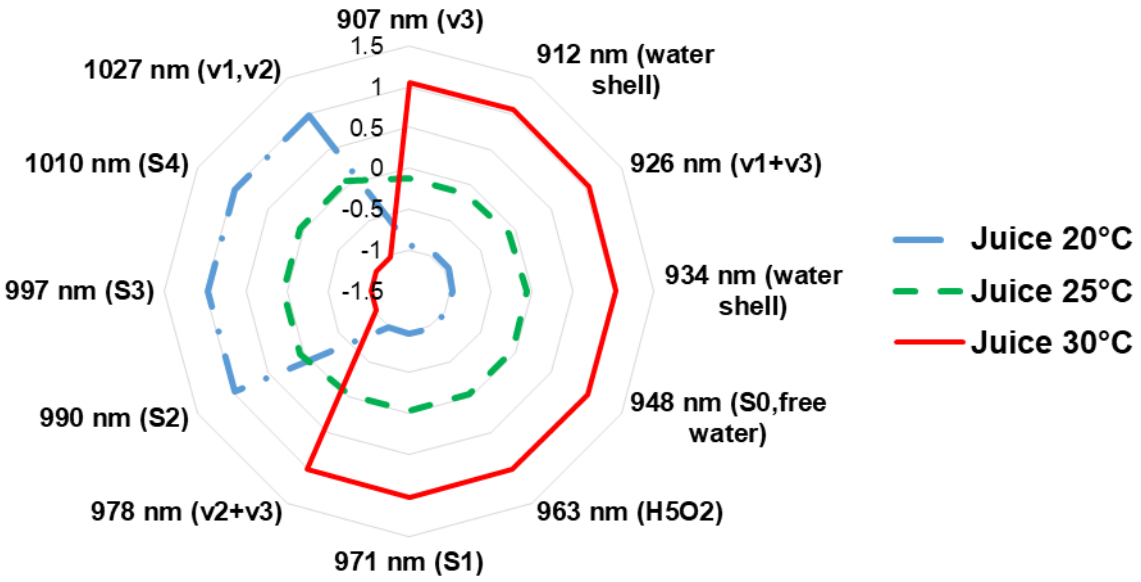


Figure 5.4: Aquagrams at three temperatures in the first overtone (1300-1600 nm) region of the OH stretch of water in a) kiwifruit juice; and b) water

a) Kiwfruit Juice



b) Water

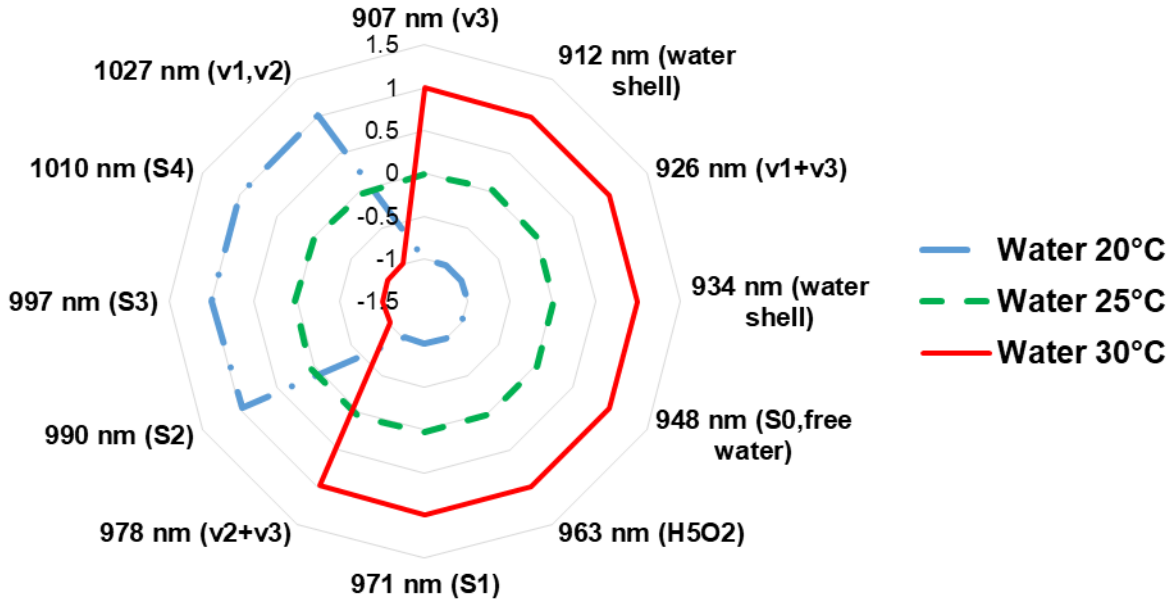


Figure 5.5: Aquagrams at three temperatures in the second overtone (870-1100 nm) region of the OH stretch of water in a) kiwifruit juice; and b) water

5.3.4 PCA of the difference matrix D for EPO correction

There was an apparent variation with temperature around the peak wavelength region of the juice spectra (Fig. 5.2). When applying PCA to the difference matrix, D, of water and juice spectra, the shape of the PC1 loading (Fig. 5.6(a),(b)) suggested a change in water structure resulting from the change in temperature. Therefore, the PC1 loading of the difference matrix D (both for juice and water) was used in the EPO correction (Fig. 5.6) as an interferent spectrum to correct juice spectra against temperature variation. The peak and trough wavelengths of the difference spectra, D were around 1414 nm and 1494 nm similar to PCA performed on the juice spectra (Fig. 5.3). However, the magnitude of water samples were higher than juice because raw absorbance of water is more than juice. The PC1 spectra of the water and the juice difference matrices overlapped (Fig. 5.7), and it was the same variation due to water temperature that occurred in the juice samples. This interference was removed using the EPO correction technique.

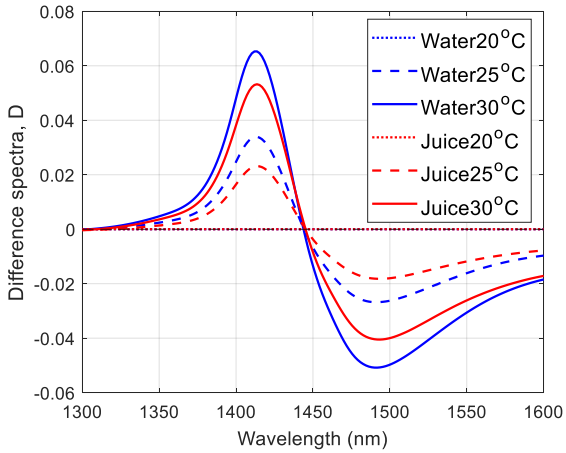


Figure **5.6** Average raw absorbance difference spectra of water and kiwifruit juice after subtracting average raw absorbance spectrum of water and juice at 20°C, respectively.

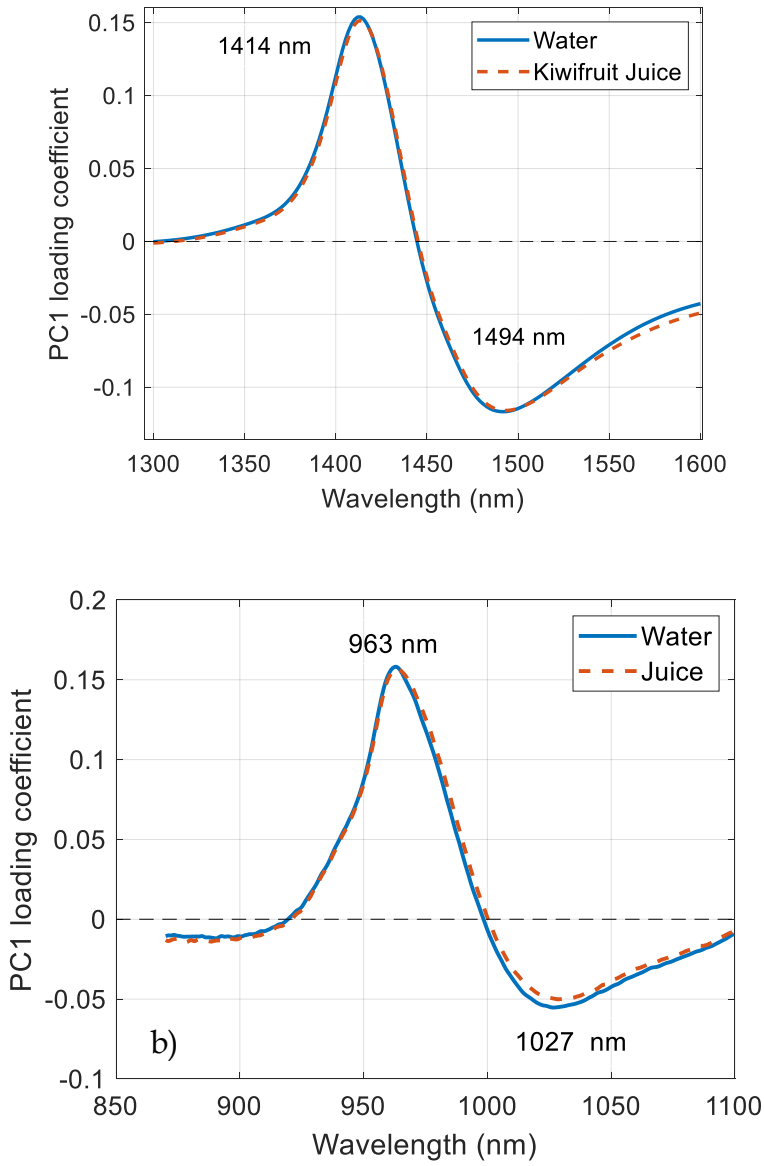


Figure **5.7** PC1 loading of water and juice difference matrix in the a) 1300-1600 nm region with a 1 mm cuvette; and b) 870–1100 nm region with a 10 mm cuvette. Labels indicate the peak wavelengths.

5.3.4 Effect of EPO correction

There were apparent spectral differences between the three juice temperature datasets, as illustrated by the PC score plots from PCA modelling of raw data in the 1300-1600 nm region (Fig. 5.8(a)). PC1

explained 87.31% and PC2 explained 12.24% of the variation in the data. The 20°C scores for the first two latent vectors did not overlap at all with the 30°C scores, showing a substantial degree of segregation. However, there was some minor overlapping between the scores at 20°C and 25°C.

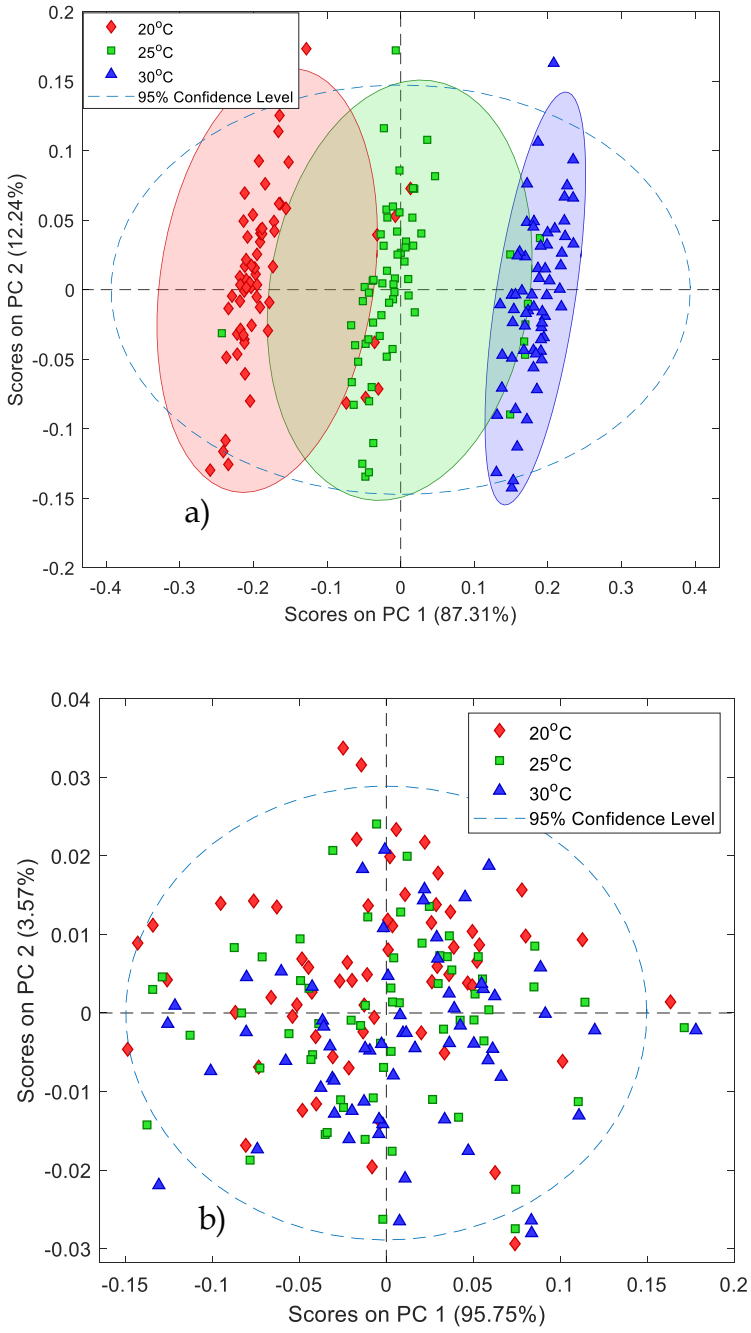


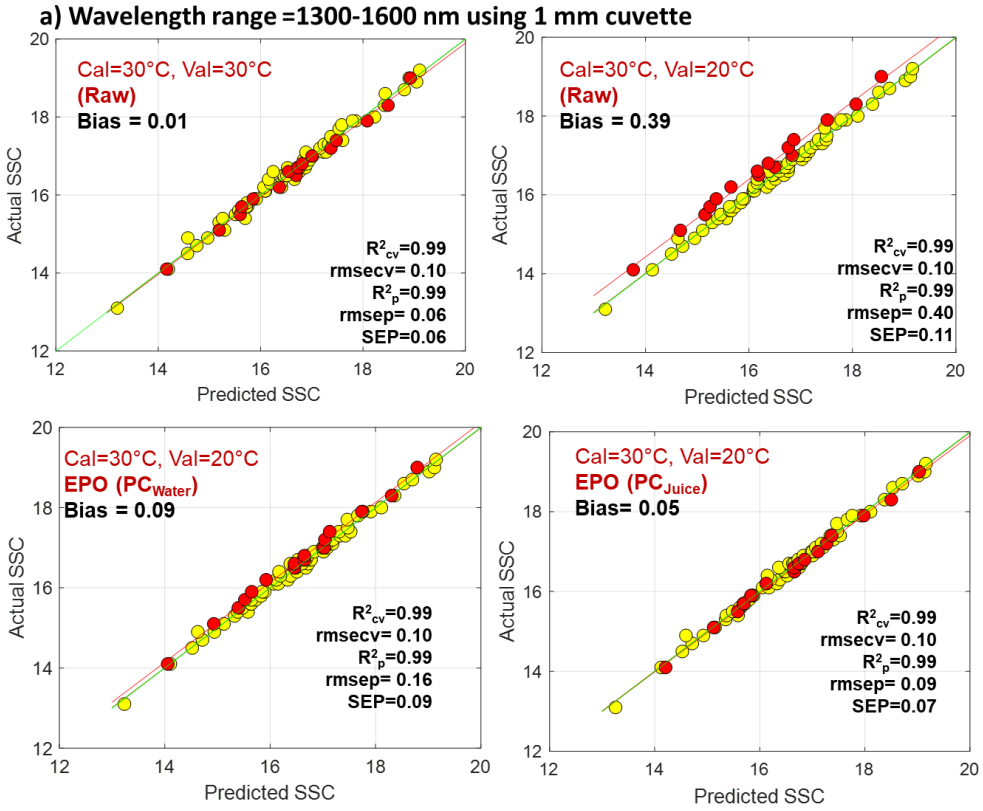
Figure 5.8. Score plot (in 1300–1600 nm) of the first two principal components (PCs) produced by the a) raw data at three temperatures; and b) EPO corrected data using PC1 of water.

To visualise the effect of EPO correction on the same dataset at three temperatures, the PC1 of the difference matrix of water spectra at three temperatures was used to correct the juice spectra against temperature variation. After correction, a PC1 vs PC2 score plot of the corrected matrix was developed (Fig. 5.8(b)). The first PC explained 95.75% and PC2 explained 3.57% of the variation in the data. The degree of segregation was reduced significantly, and there was substantial overlapping between 20°C, 25°C, and 30°C samples. This technique shows the potential for EPO correction to reduce the effects of temperature variation by using an independently measured water samples.

5.4 Prediction of SSC

The aquaphotomics analysis, particularly the strong and consistent change in the aquagram patterns with temperature (c.f., Fig. 5.4 and Fig. 5.5), suggested the use of EPO pre-processing using PC1 (Fig. 5.7) from the pure water and juice temperature change difference matrices as the interference component (Chapter 2). The PC1 vs PC2 score plot at different temperatures collapsed to strong overlapping of samples following the EPO processing (Fig. 5.8(b)). There was a significantly improved model performance with EPO pre-processing (Fig. 5.8). The SSC prediction bias, observed when applying the 30°C calibrated model to samples at lower temperature, was reduced to near negligible levels. Compared to raw data, the bias dropped from 0.39 to 0.09 using EPO with PC_{Water} and 0.07 with PC_{Juice} in the

1300– 1600 nm region (Fig. 5.9(a)). Similar results were achieved in the second overtone region of water. The EPO interference term, the PC1 generated from the pure water-temperature spectral matrix, is independent of the kiwifruit juice samples, which considerably simplifies generating a temperature-independent model. It does not require juice samples to be measured at different temperatures – measurements at one temperature will suffice. This approach may apply to other applications such as other fruit juices or intact fruit modelling problems, where robustness against temperature changes is desirable. Moreover, its performance is comparable to the results achieved using the PC1 component of the difference matrix of juice at different temperatures, with insignificant bias.



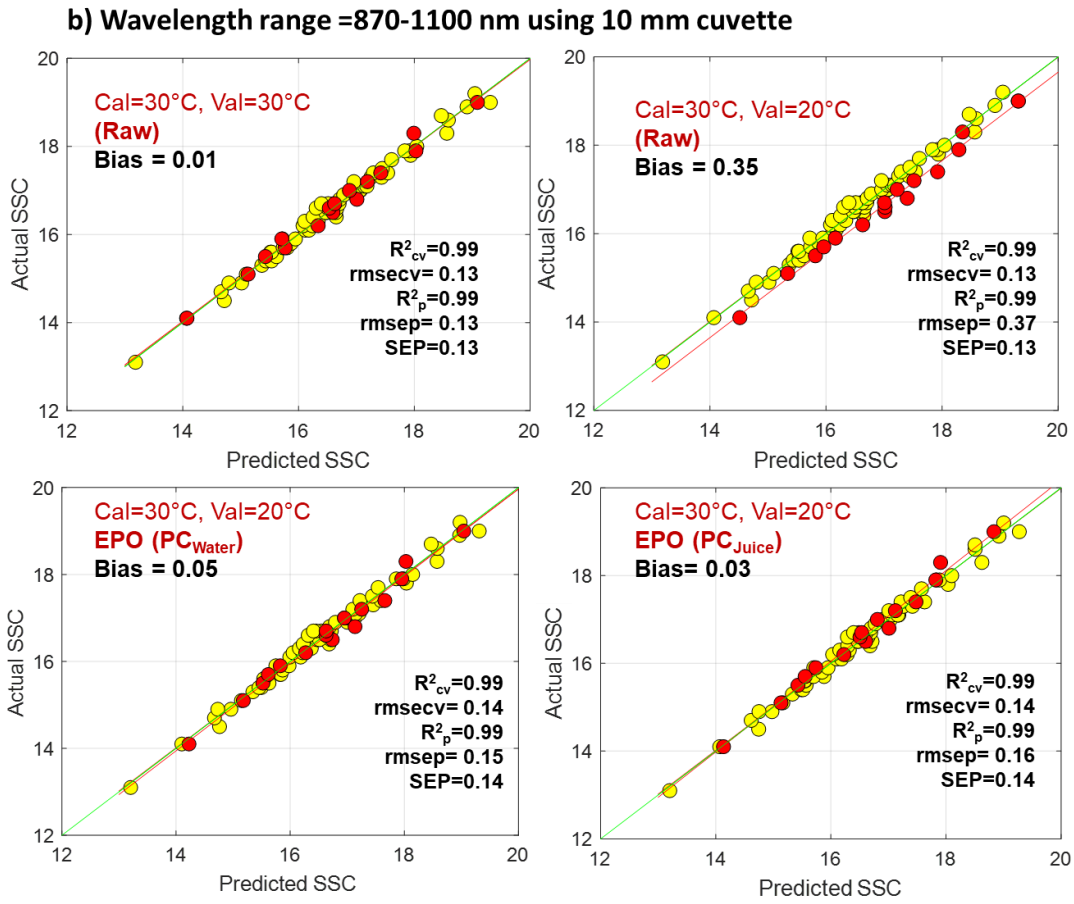


Figure 5.9. SSC prediction of kiwifruit juice in a) the first overtone (1300–1600 nm); and b) the second overtone (870–1100 nm) region of the OH stretch of water. Yellow and red circles are for calibration and validation data, respectively.

5.5 Conclusions

The influence of increasing temperature on the peak absorbance of kiwifruit juice was a lateral shift in the first overtone region, in contrast to a vertical shift in the second overtone region of water. Using aquaphotomics, the water bands were identified in the two overtone regions. Different bands were seen to change in the second overtone region compared with the first overtone region.

With the same data set, the use of EPO pre-processing was investigated to assist in building temperature-independent PLSR models to predict SSC in kiwifruit juice. The results showed that using the EPO method, a significant reduction in the prediction bias was achieved, by at least a factor of 10, when applying a model created at one temperature to measurements made at another.

Chapter 6: Kiwifruit pulp

6.1 Introduction

The DM of kiwifruit is potentially a useful taste indicator (Jordan et al., 1997). It is an internal fruit quality parameter that represents fruit flesh properties. The conventional lab method for measuring DM is to dry slices of fruit to drive off the water, which is a destructive method. Measurement vary a little in industry but generally involve cutting one or two thin central slices of approximately 2 -3 mm thickness from each kiwifruit and drying them in an oven for 24 h at 65°C. DM is then calculated as the percentage of final dry weight to the initial wet weight of the slices (McGlone & Kawano, 1998). Several studies have shown that the DM of kiwifruit can be predicted by NIR spectroscopy (McGlone & Kawano, 1998; McGlone et al., 2002c; Acharya et al., 2013; Shafie et al., 2015; Kaur et al., 2017b).

A new method of DM measurement using an FT-NIR spectrometer has been examined. The transmission measurement was performed on the pulp from kiwifruit flesh (excluding the seeds and skin). Although this method is destructive, it is faster than the conventional lab method. With the help of PLSR and various data processing techniques, the optimal wavelength range was identified to build a robust calibration model for DM prediction.

Since kiwifruit pulp is more than 80% water, its spectral signature is also affected by its DM, since DM = 100 - water content, in percentage terms. The spectral changes can be analyzed using an aquaphotomics approach in the first overtone region of the OH stretch of water (1300–1600 nm), using a

short pathlength (1mm) cuvette to optimise the signal to noise ratio (SNR) (Jensen & Bak, 2002).

In this chapter, calibration models using PLSR for DM prediction of kiwifruit pulp using an FT-NIR spectrometer are described. The WASPs are identified for different batches of Gold and Hayward kiwifruit pulp in the first overtone region.

6.2 Methods and materials

6.2.1 Sample preparation

This experiment was performed as part of a larger project being run at Plant and Food Research, titled "Delivering Perfect Fruit", that involved, in part, development of new rapid DM measurement methods. The fruit were provided by a commercial fruit testing service (Eurofins Bay of Plenty, Katikati, NZ). The fruit were obtained from many different orchards. The trial ran for seven weeks. Each week, kiwifruit were obtained on at least two days (20-24 fruit were selected randomly from 80-90 fruit each day), except for week 1 and 7 in which fruit was received only for one day in the week. A total of 240 kiwifruits were received: Gold kiwifruit "*Actinidia chinensis*" during weeks 1 to 4 and 7, and Hayward kiwifruit "*Actinidia deliciosa*" during weeks 5 and 6.

The pulp was prepared from the flesh of outer pericarp tissue (excluding seeds). After removal it was chopped into pieces and passed through a proprietary in-house pulping machine designed specifically to

pulp fruit flesh. The samples were stored separately at room temperature in several sample containers. FT-NIR analysis and reference measurements were performed on the same day after the samples had equilibrated to room temperature.

6.2.2 Reference measurement

For reference data, the DM of the kiwifruit pulp was determined using the traditional oven-drying method. These measurements involved drying approximately 2 g of the pulp in a petri dish in an oven for 24 h at 65°C. DM was then calculated as the percentage of final dry weight to the initial wet weight of the sample.

The SSC of the kiwifruit pulp was measured at room temperature using a digital refractometer (Atago Co. Ltd, Tokyo, Japan), calibrated with Milli-Q water (Vasquez & Mueller, 2019). The Brix value was recorded after placing 0.5 ml of pulp into the measurement chamber of the refractometer.

The starch content, expressed as a percentage in g/100g of FW (fruit weight) can be calculated from estimates of the insoluble solids content (I) and the residual solids content (R)

$$S = I - R \tag{6.1}$$

where I in turn can be calculated from DM and SSC values

$$I = \frac{(DM - SSC)}{100 - SSC} \times 100 \tag{6.2}$$

The residual solids content R, which represents solid components such as cell walls, seeds and skin, is typically assumed to be a fixed constant for a

cultivar, about 3% FW for the main kiwifruit cultivars (McGlone et al., 2002c), *S* is the estimated starch content, and *R* is the residual insoluble solids (2-3% FW represent seeds and cell walls).

6.2.3 Spectral acquisition

Transmittance spectra of the pulp samples were measured at 25°C (±1 °C) with an FT-NIR spectrometer (Tango, Bruker Corporation, Germany) equipped with a temperature-controlled holder and a peristaltic pump to control the flow of the pulp sample into the cuvette (Fig. 6.1). The samples were agitated using the pump in the cuvette to avoid settling of starch particles. One measurement per pulp sample was acquired using a quartz cuvette of 1 mm pathlength. For each measurement, one spectrum, which was the average of 32 successive scans, was recorded with a resolution of 16 cm^{-1}. The total number of pulp spectra was 240 (240 samples x 1 consecutive scan). The samples were divided into seven batches; one batch for each week of the trial.

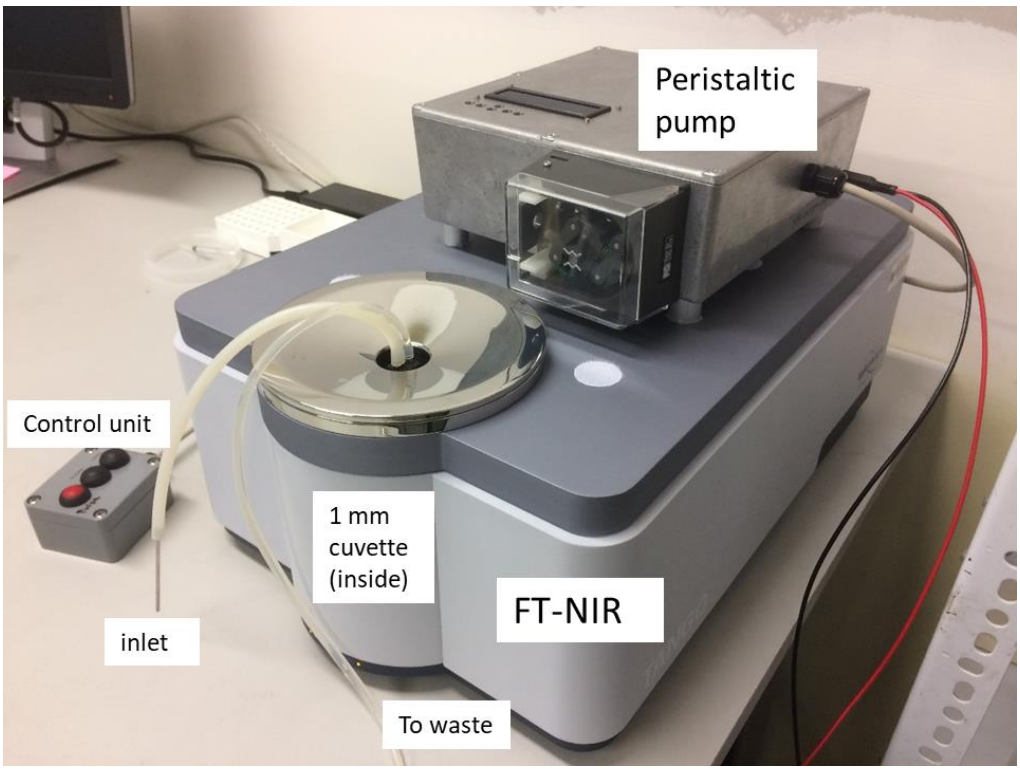


Figure 6.1 Setup for kiwifruit pulp spectral measurement.

6.2.4 Aquaphotomics analysis

An aquaphotomics WAMACS was created using the method reported in the literature (Tsenkova et al., 2015) and as described in Chapter 3. Aquagrams displaying the resulting WASP in the first overtone region were then studied for fruit variation according to the different batches.

6.2.5 Multivariate and statistical analysis

Chemometrics models were developed within Bruker's OPUS 7 software environment by aligning spectra and their matching DM values. The group of validation spectra (validation set) was set as 30% using the Kennard and Stone algorithm, with the other 70% forming the calibration or modelling

set (Kennard & Stone, 1969). Three different groups were made for DM prediction:

1) Gold-Runny: Foamy and Runny; comprising week 1, week 2, and week 3 samples

2) Gold-Gel: Gel-like; comprising week 4 and week 7 samples

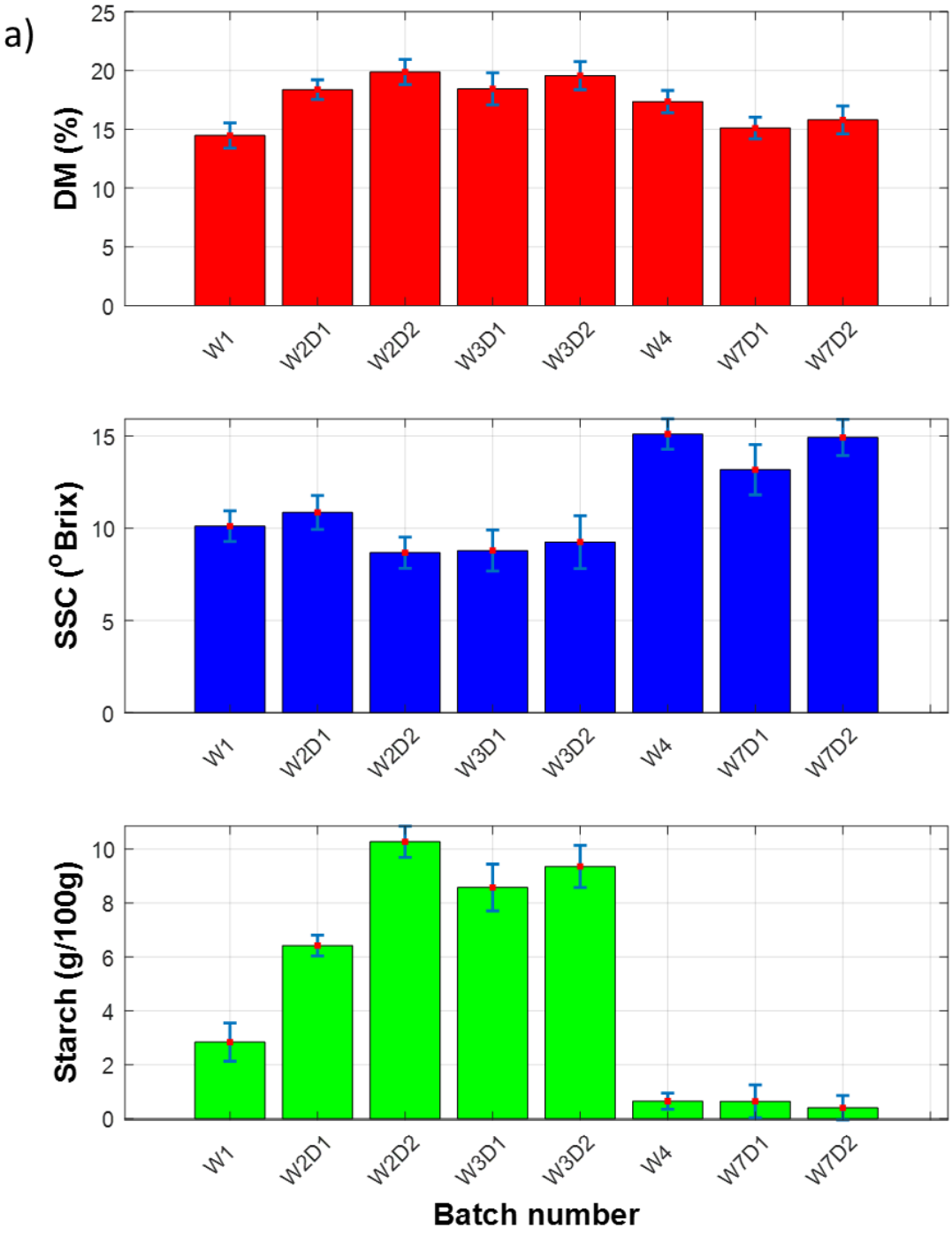
3) Hayward-Runny: Runny; comprising week 5 and week 6 samples

Four-fold venetian blind cross-validation was applied, to ensure there was no overfitting, with latent variables selected based on the lowest root mean square error of cross-validation (RMSECV). The best partial least squares (PLS) regression models were determined iteratively by an optimisation process in which different pre-processing options, e.g., SNV, MSC, first derivative (1D), second derivative(2D), 1D+SNV, and 1D+MSC, were applied to the wavelength range 870–1800 nm. This was followed by cross-validation of the calibration set using different wavelength regions. The best models, rank-ordered on their RMSECV value, were tested against the validation set to establish a robust model.

6.3 Results and discussion

Fig. 6.2((a), (b)) shows variation in DM, SSC, and starch content across different batches of fruit. For Gold kiwifruit (Fig. 6.2(a)), the mean DM ranged from 14–20%DM. Weeks 1-3 batches represent early season fruit while weeks 4-7 batches represent late-season fruit. The %DM was higher

for W2D1, W2D2, W3D1, and W3D2 compared to W1, W4, W7D1, and W7D2 batches where 'W' and 'D' represents the week and the day, respectively. The SSC concentration decreased for the same batches, while it was higher for late-season fruit, i.e., W4, W7D1, and W7D2. Higher SSC concentrations indicate greater fruit maturity. Eventually, the starch concentration was lower for the same batches of late-season fruit, which was less than 1 g / 100 g (Fig. 6.2(a)).


a)
DM (%)
SSC (°Brix)
Starch (g/100g)
W1
W2D1
W2D2
W3D1
W3D2
W4
W7D1
W7D2
Batch number

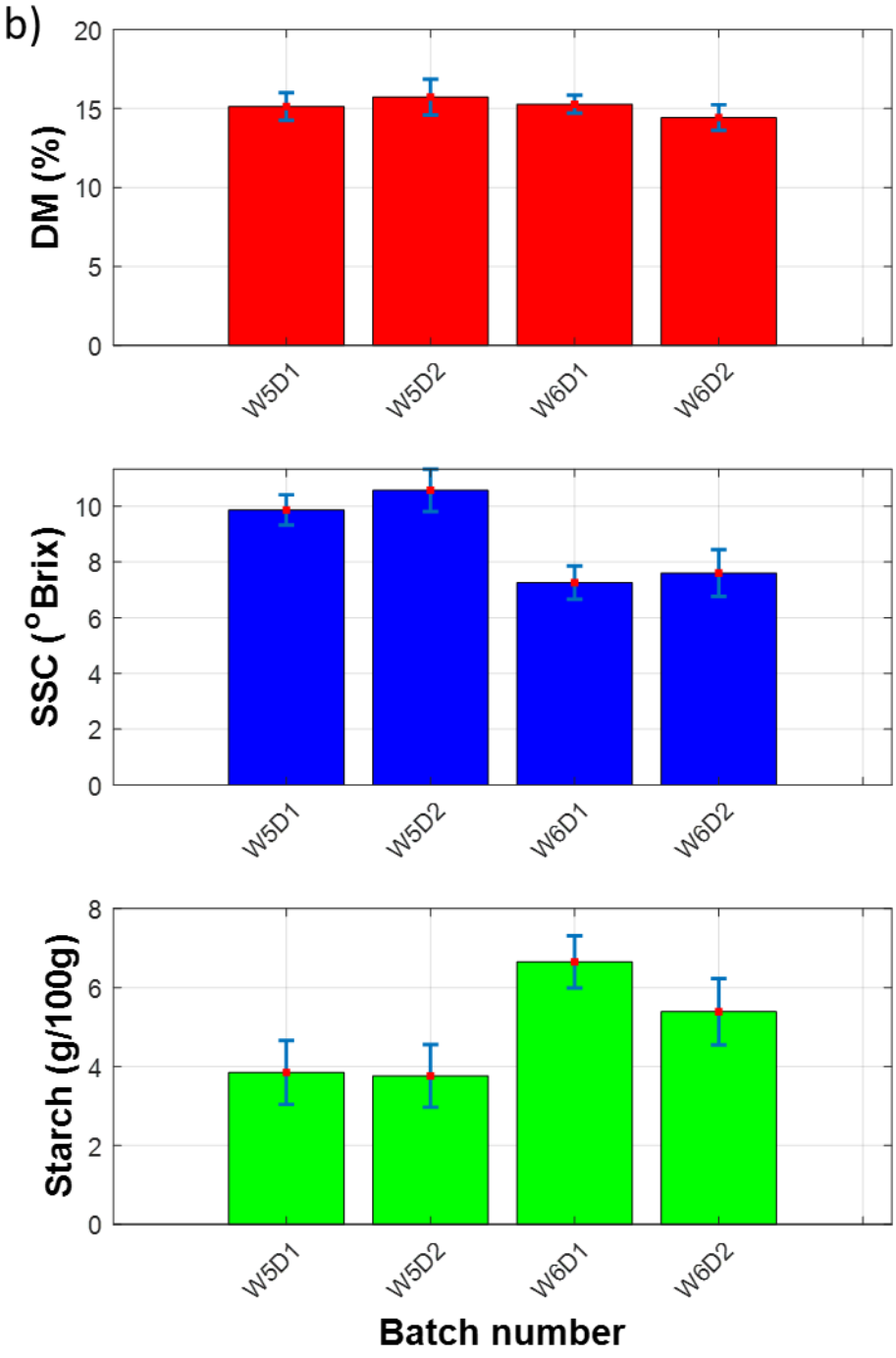


Figure 6.2 Mean DM, SSC, and starch in a) Gold; and b) Hayward kiwifruit vs. the batch number. 'W' represents the week number, and 'D' represents the day. The error bar is the standard deviation of the same batch of the fruit population.

For Hayward kiwifruit, the %DM remained consistent around 15%DM for weeks 5 and 6 (Fig. 6.2(b)). However, there was a decrease in SSC concentration for week 6. The decrease in SSC concentration was due to higher starch concentrations for W6D1 and W6D2 batch fruit.

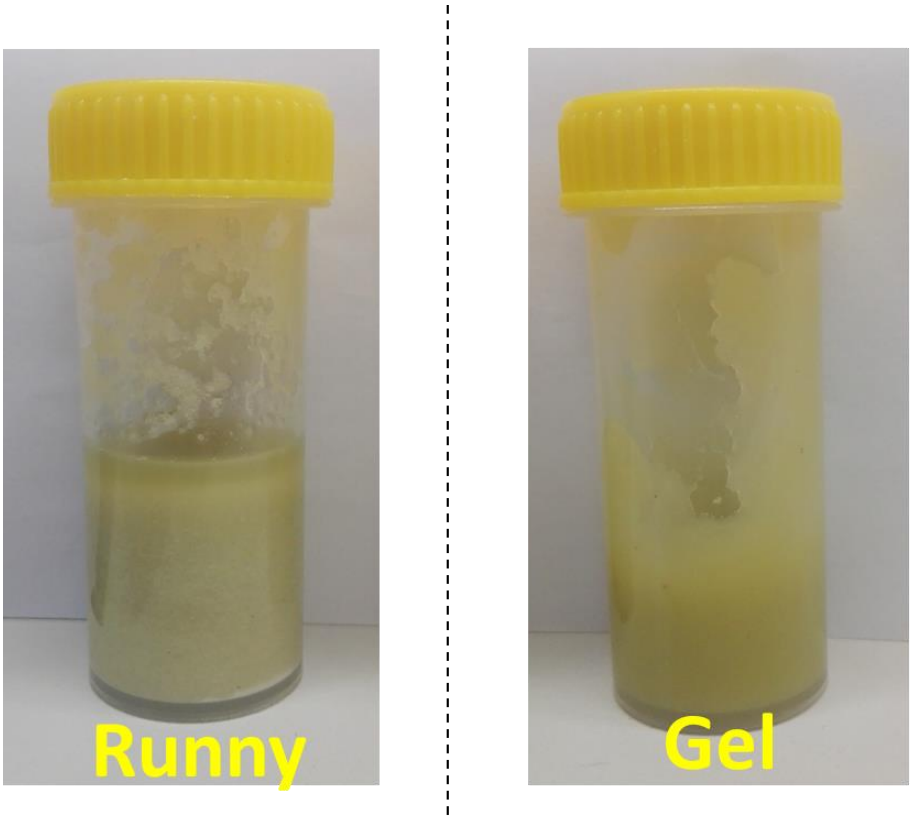


Figure 6.3 Runny and Gel-like (Gold) kiwifruit pulp samples

Fig. 6.3 shows the variation in the early and late season fruit pulp. The variation in starch concentration resulted in different pulp consistencies and textures. For W2 and W3 batch fruit, the consistency was heterogenous 'Runny': with a separation in solid and liquid phase whereas, for W4 and W7, it was homogeneous 'Gel-Like'. W1 was foamy, and the pulp was neither runny nor gel-like. The Hayward pulp samples were runny, with variations in starch concentration.

6.3.1 The raw spectra

The absorbance plot in Fig. 6.4 illustrates that as the batch number of the kiwifruit increased, the absorbance curve for the pulp in the first overtone region decreased in intensity, because of the decrease in starch concentration. Higher starch concentrations caused scattering of NIR light and hence induced offset in the NIR absorbance spectrum. Runny batches had higher absorbance intensities compared to gel-like batches of fruit. In Hayward fruit (Fig. 6.4(b)), the Week 6 fruit had higher starch concentrations and showed higher intensity offsets.

The prominent feature in both Gold and Hayward pulp (Fig. 6.4(a),(b)) was a broad absorbance peak centred around 1451 nm, attributed to various water states or species related to the first overtone of the O-H stretch (Burns & Ciurczak, 1992).

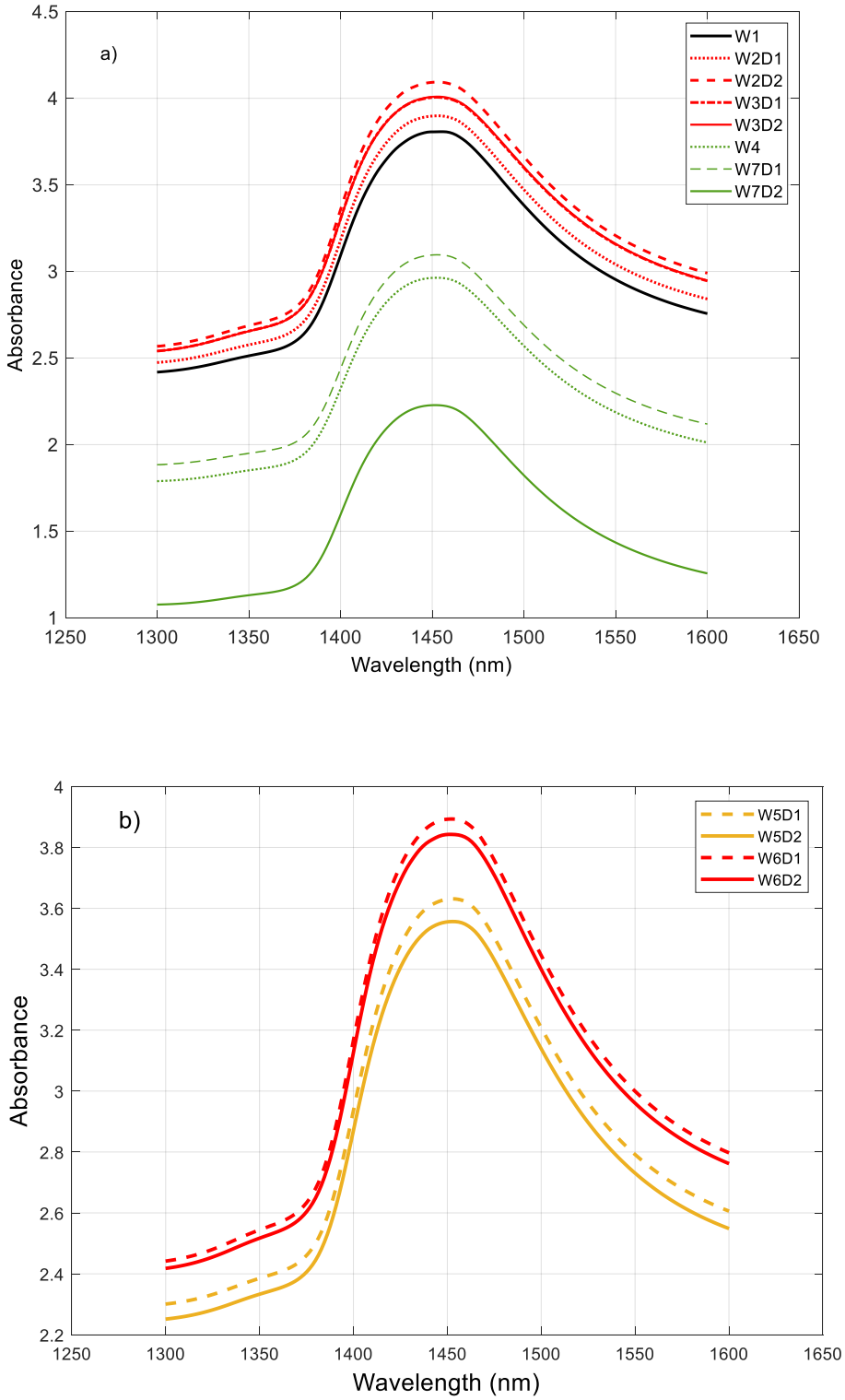


Figure 6.4 Average raw absorbance spectra of each batch of fruit for a) Gold; and b) Hayward kiwifruit pulp at 25°C in the first overtone (1300–1600 nm).

6.3.2 Aquagrams

The derived WASPs presented in the aquagrams showed differences between runny and gel-like kiwifruit pulp samples (Fig. 6.5(a)). As the starch concentration rose for W2 and W3 pulp samples, the number of strongly H-bonded water molecule species (S4 and ($v1$,$v2$)) increased, along with an increase in free water S0 and water shell. Conversely, with fruit maturation, and thus a decrease in starch concentration for W4, W7D1, and W7D2 batch fruit, the consistency became gel-like, and the WASP (green- colored) became diagonal, increasing left to right and irregular broad band. The water species related to solvation shell and bonded water structures (S1, S2, and S3). H_5O_2, $v3$, and ($v1$+$v3$) increased. In contrast the WASPs for runny samples (red-coloured) was a narrow band, titled down from left to right. The foamy fruit in the W1 batch covered all the water species (black-coloured).

Some similar trends were observed for Hayward fruit with high and low starch concentrations. As the starch rose and SSC decreased, the free water species increased (Fig. 6.5(b)). The strongly H-bonded water molecular species (S1, S2, and S3) increased with SSC concentration, as seen previously for Gold kiwifruit pulp (Fig. 6.5(a)).

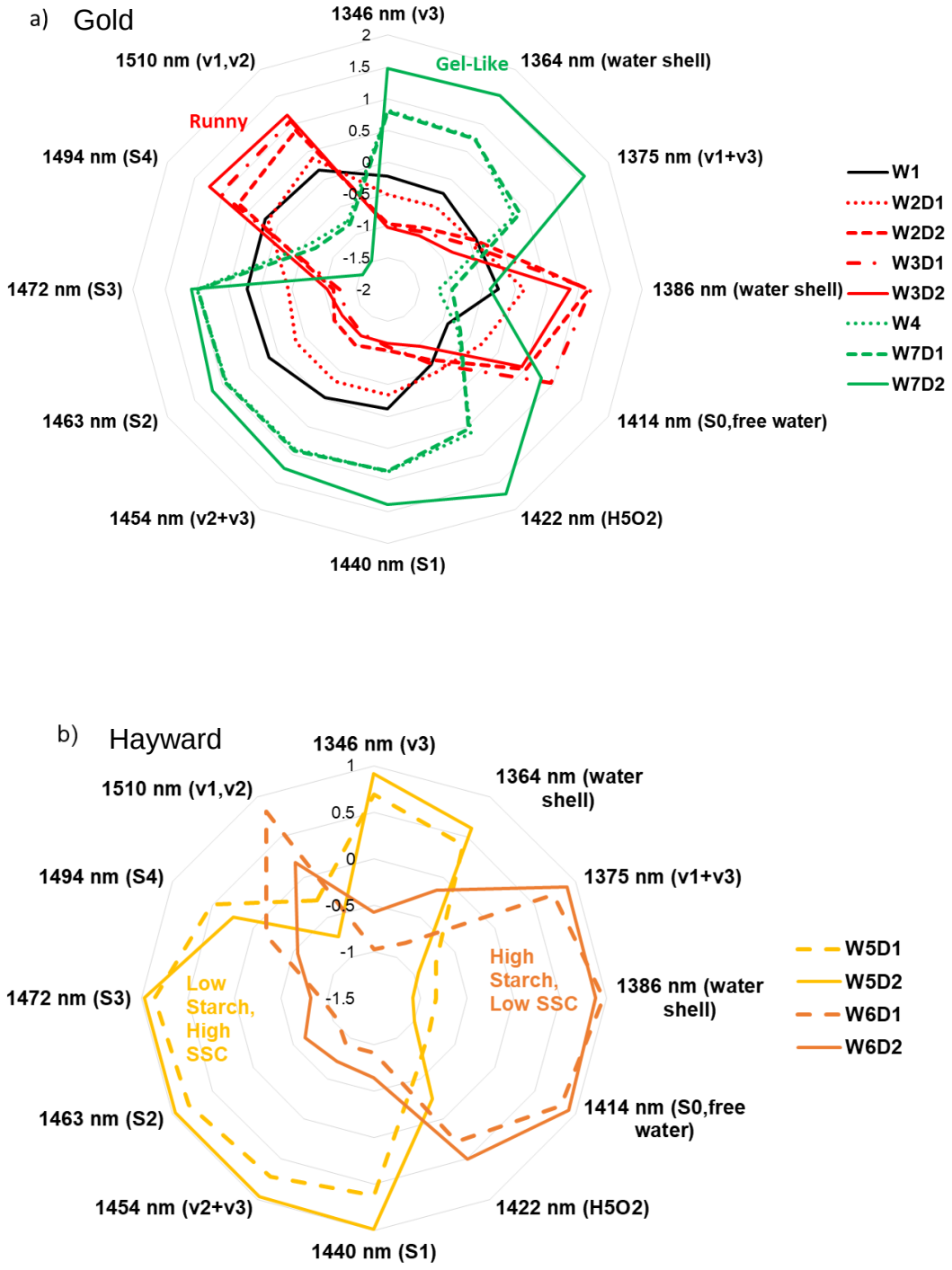


Figure 6.5 Aquagrams at 25°C in the first overtone (1300–1600 nm) region of the OH stretch of water for the pulp of a) Gold; and b) Hayward kiwifruit

6.4 Prediction of DM (%)

PLSR models developed for DM prediction for different groups of samples proved to be better for gel-like gold samples (Fig. 6.7) compared to foamy and runny group samples (Fig. 6.6) and yielded an RMSEP value of 0.20%DM. Gold runny samples gave the best model with 1D+SNV pre-processing in the 1063–1800 nm region. The RMSEP was 0.27%DM which was higher than the error value for gel-like Gold samples. However, for Hayward fruit, the RMSEP was similar to the error obtained with Gold 'runny' samples and was 0.29%DM (Fig. 6.8). Prediction models for Gold, Foamy and Runny samples (Fig. 6.7) gave high RMSECV value of 0.30%DM due to the high starch concentrations in the sample, which caused scattering.

For generating robust calibration models for three groups of samples, the water absorbances in the vicinity of the first overtone water peak appear to play an important role as a whole. The wavelengths regions were: Gold, Foamy and Runny = 1063-1800 nm; Gold, Gel-like = 1333- 1740 nm, and Hayward, Runny = 1063-1800 nm. All these wavelength regions well covered the vicinity of the water peak at 1450 nm.

6.4.1 Gold: Foamy and Runny samples:- Week 1, week 2, and week 3

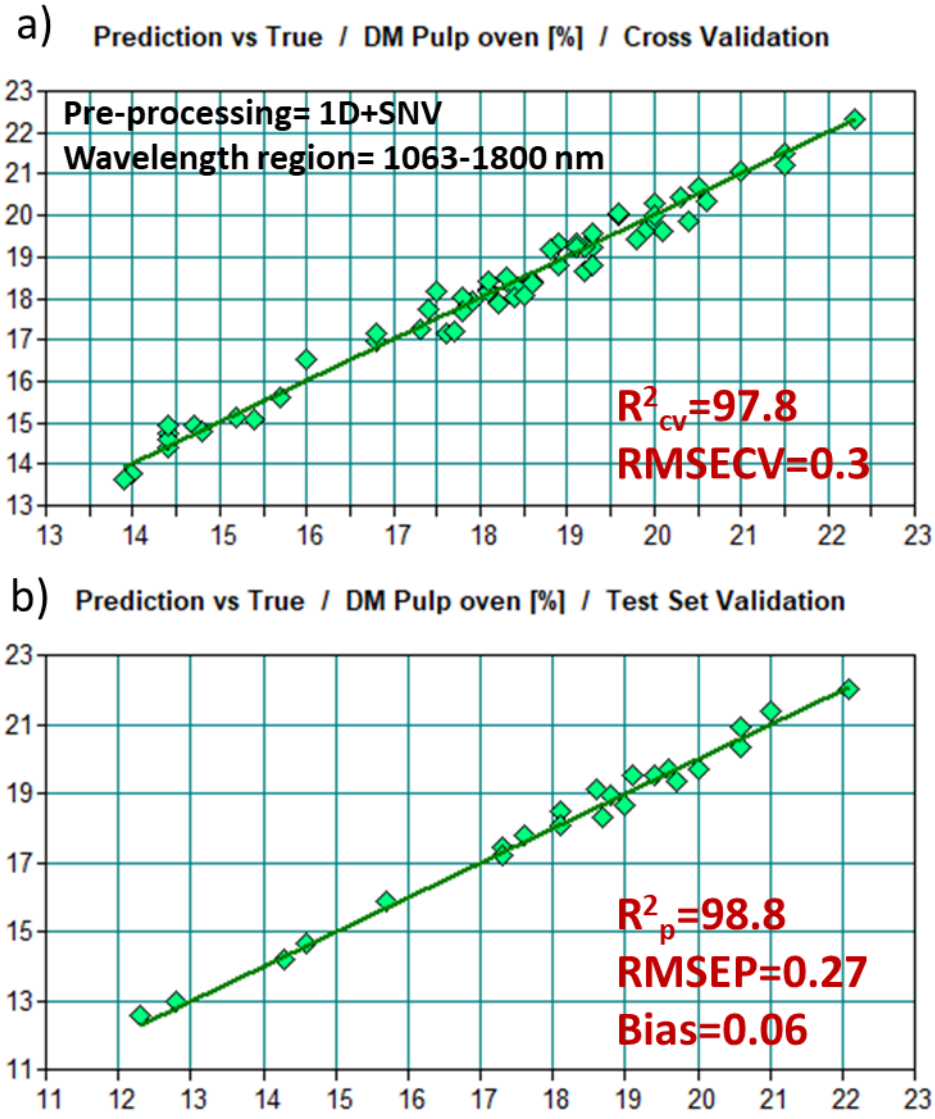


Figure 6.6 DM prediction of Gold: foamy and runny kiwifruit pulp in a) Cross-validation; and b) Prediction. R^2cv/R^2p: the coefficient of determination for calibration/prediction; RMSECV/RMSEP: root mean square error of calibration/prediction.

6.4.2 Gold: Gel-like samples:- Week 4 and week 7

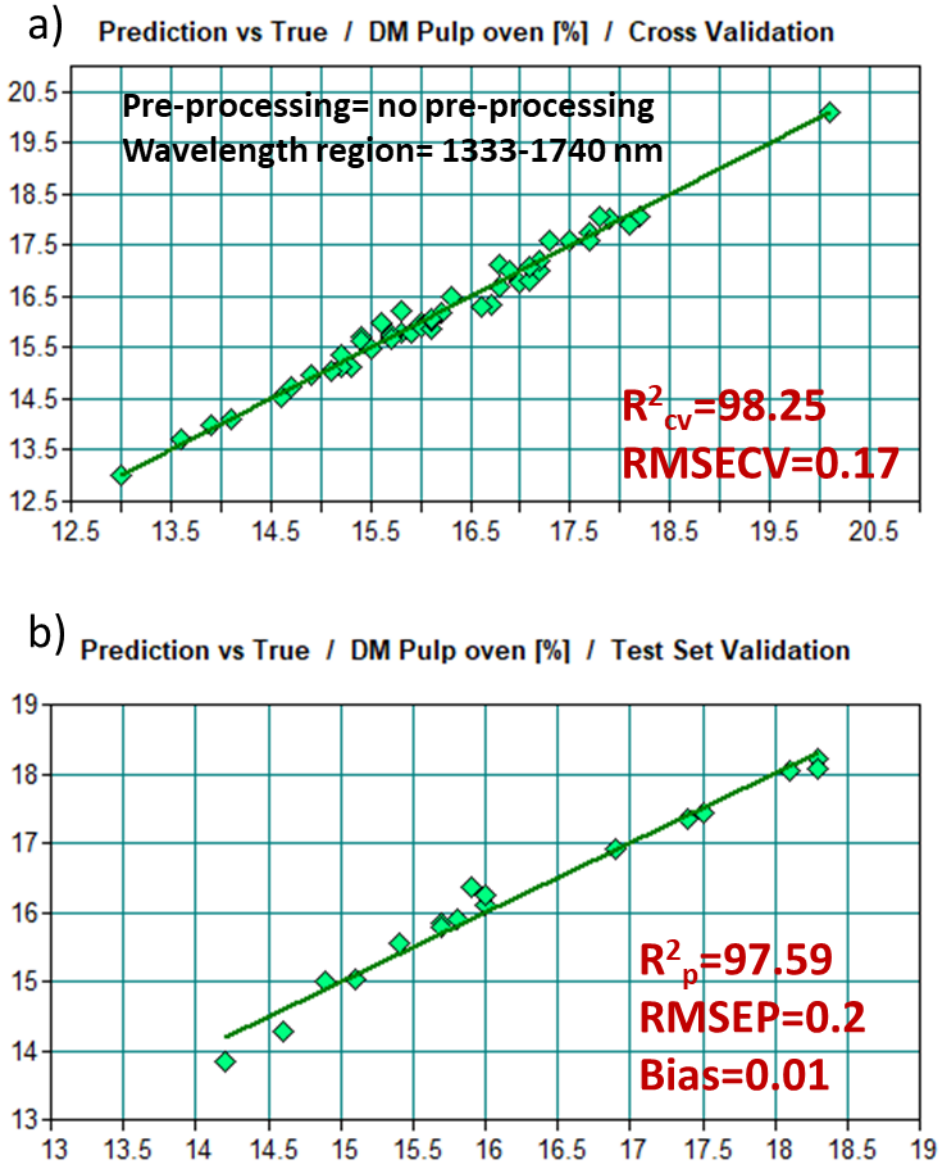


Figure 6.7 DM prediction of Gold: gel-like kiwifruit pulp in a) Cross-validation; and b) Prediction. R^2cv/R^2p: the coefficient of determination for cross-validation/prediction; RMSECV/RMSEP: root mean square error of cross-validation/prediction.

6.4.3 Hayward: Runny samples:- Week 5 and week 6 analysis

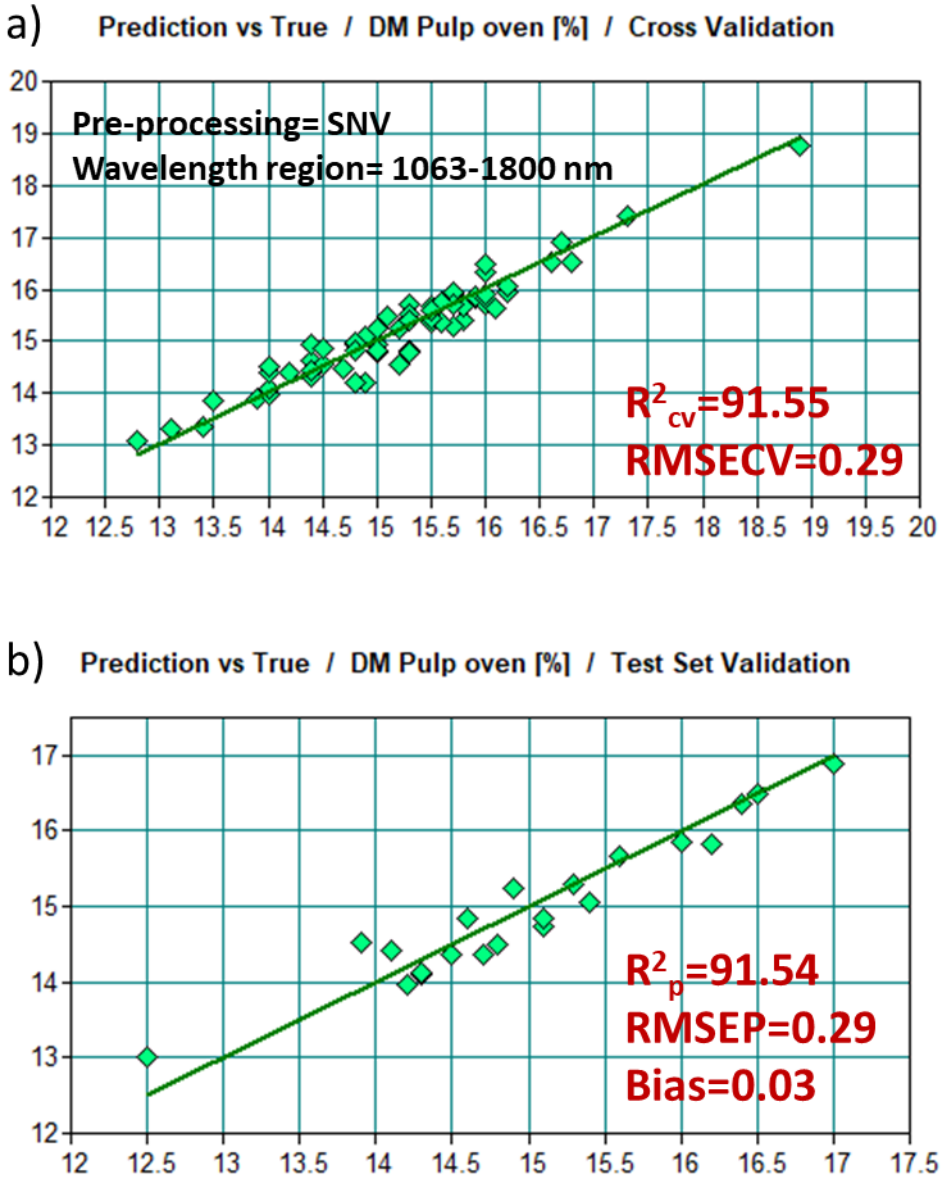


Figure 6.8 DM prediction of Hayward: runny kiwifruit pulp in a) Cross-validation; and b) Prediction. R^2cv/R^2p: the coefficient of determination for cross-validation /prediction; RMSECV/RMSEP: root mean square error of cross-validation /prediction.

6.5 Conclusions

Using aquaphotomics, the water spectral pattern was identified for different consistencies of kiwifruit pulp samples in the first overtone region of water. Different trends in WASPs were seen in the first overtone region for 'Runny' and 'Gel-like' pulp samples. Gel-like samples were oriented to downward left to right diagonal pattern with an increase in the water species related to solvation shell and bonded water structures (S1, S2, and S3) and H_5O_2, $v3$ and ($v1$+$v3$). In contrast the WASPs for runny samples oriented themselves to upward left to right diagonal pattern. The foamy fruit pulp in the W1 batch covered all the water species.

A similar trend was seen for Hayward fruit with high and low starch concentrations. As the starch rose and SSC decreased, the free water species increased. The strongly H-bonded water molecular species (S1, S2, and S3) increased with SSC concentration.

The results demonstrated that it was possible to produce excellent PLS regression models for the determination of DM in kiwifruit pulp from each of these three groups using FT-NIR spectroscopy in the NIR between 870 and 1800 nm. The models had very good prediction errors of 0.27, 0.20, and 0.29% DM for Gold: foamy and runny, Gold: gel-like, and Hayward: runny samples, respectively. Gel-like samples delivered the best prediction model because of negligible starch content, which caused scattering. The PLSR models built confirm that the important information resides in the water window of the first overtone region.

Chapter 7: Chilling Injury Kiwifruit

7.1 Introduction

For most fruits, cool-storage is the primary means by which quality is maintained after harvest. However, prolonged low-temperature storage can cause the breakdown of tissues, which is known, for kiwifruit, as low-temperature breakdown (LTB) or chilling injury (CI). The resulting damage reduces the fruit quality and shelf life, leading to economic losses (Lallu, 1997). Symptoms of kiwifruit CI include grainy tissue in the outer pericarp, water-soaked areas in the outer and inner pericarp, and outer skin discoloration (Lallu, 1997; Gwanpua et al., 2018), as shown in Fig. 7.1. The development of CI is dependent on the storage duration and temperature of chilling as well as the fruit maturity stage.

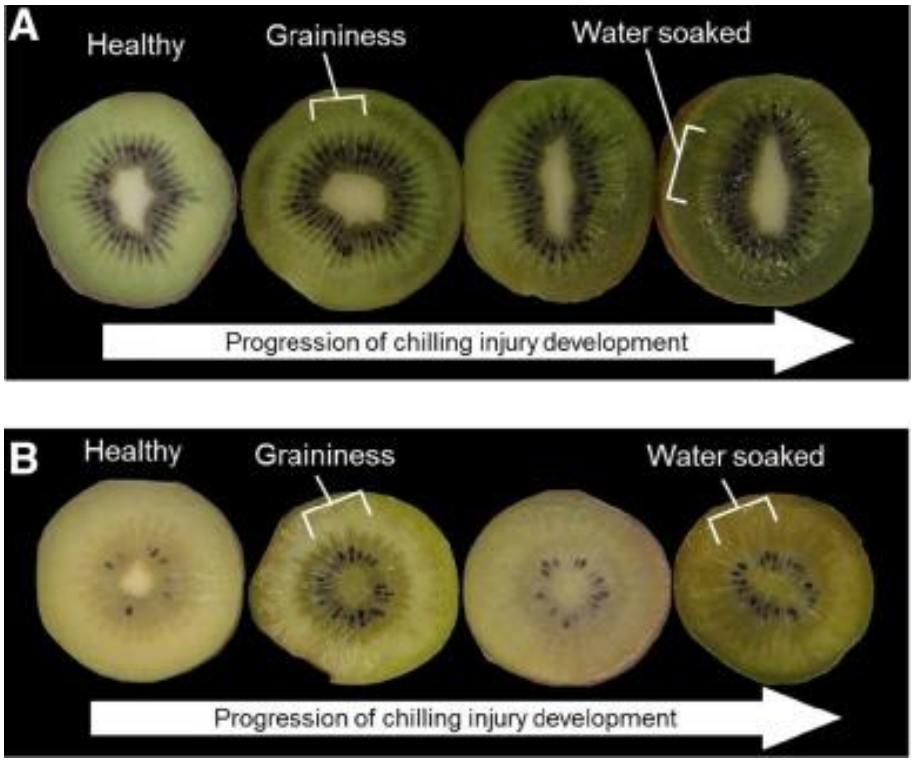


Fig. 7.1 Progression of chilling injury in 'Hayward' (A), and 'Zesy002' SunGold (B) kiwifruit (Gwanpua et al., 2018).

Currently, the presence of CI in kiwifruit can only be detected when the fruit is cut open. A non-destructive method of detecting kiwifruit CI would be beneficial. Near-infrared spectroscopy (NIRS) is a popular and efficient non-destructive method for internal quality evaluation of fruits and vegetables, including aspects such as SSC, DM, acidity, and firmness (Cen et al., 2016). A typical NIRS set-up uses a broadband light source to illuminate the sample, and the reflected or transmitted light is measured using a spectrometer (Wang et al., 2019).

As the water absorption peaks at 1450 nm and 970 nm in the kiwifruit juice change with temperature and SSC (Kaur et al., 2018), the differences between control and CI kiwifruit can be traced by observing changes in the water spectral patterns presented on an aquagram. Therefore, in this chapter, aquaphotomics was investigated to identify changes in the water structure caused by CI. The water bands were identified for CI kiwifruit juice in the first overtone region and for whole intact fruit in the second overtone region of the OH stretch of water.

7.2 Methods and materials

7.2.1 Sample preparation

A total of 132 Gold kiwifruit (*Actinidia chinensis*) were used for the experiment. Non-destructive NIR measurements were made on intact fruit; then the juice was extracted from the whole fruit. The juice was prepared using a juicer (BJE410, Breville, Sydney, Australia) and then centrifuged at

10,000 rpm to produce a clear light-coloured juice as a supernatant. The clear juice was filtered with 0.22 μm paper to get clear juice before testing.

7.2.2 Reference SSC (°Brix) measurement

The SSC of the kiwifruit juice samples was measured at room temperature using a digital refractometer (Atago Co. Ltd, Tokyo, Japan), calibrated with Milli-Q water (Vasquez & Mueller, 2019). The Brix value was recorded after placing 0.5 ml of juice into the measurement chamber of the refractometer; this was enough to fully cover the optical interface.

7.2.3. CI measurement

The CI was identified by cutting each fruit in half and visually assessing the fruit flesh for granular appearance or water soaking.

7.2.4 Spectra acquisition

7.2.4.1 NIR intact fruit measurement

To investigate aquaphotomics in the second overtone of the OH stretch of water (800– 1100 nm), interactance measurements were obtained non-destructively for each kiwifruit in the wavelength range 400–1100 nm, using a custom-built Vis-NIR spectroscopy system (McGlone et al., 2002b). It contained a broadband light source (50 W quartz halogen, RJL 5012 FL, Radium, Germany) and a diode array spectrometer (MMS1-NIR, Zeiss, Jena, Germany) with a sampling interval of 3.3 nm and an optical resolution

varying between 8 and 13 nm. Two Vis/NIR interactance measurements were recorded for opposite locations at the blossom end of each fruit (Wang et al., 2019). The spectra, T, were corrected for dark current and normalised by a Teflon white reference (McGlone et al., 2002b). Then, the transmittance was transformed into absorbance taking $\log_{10}(1/T)$.

7.2.4.2 FT- NIR Transmittance measurement of the kiwifruit juice

SSC variation in kiwifruit can affect the water spectral pattern and thus cause absorbance bands to change (Kaur et al., 2017a; Kaur et al., 2018). Therefore, to scrutinize the effect of CI, with no contribution from SSC, only kiwifruit with similar SSC values were selected. A total of 15 kiwifruits were chosen, with SSC values of 14-15°Brix. Five of the 15 fruit were selected for their moderate-high CI classification based upon skin discoloration at the blossom end (external) and severe granular and water soaking areas (internal). The remaining ten fruit were used as controls. Transmittance spectra of the juice were measured at 22°C (±1°C) with an FT-NIR spectrometer (Tango, Bruker Corporation, Germany) equipped with a temperature-controlled holder. One measurement for each juice sample was acquired using a quartz cuvette of 1 mm optical path length. Each measurement was the average of 32 consecutive scans and was recorded with a resolution of 16 cm^{-1}. The total number of juice spectra was 15 (15 samples x 1 consecutive scan).

7.2.5 Aquaphotomics analysis

The water bands were calculated using the standard protocol described in Chapter 3. Once the WAMACS wavelengths are selected, a WASP on the aquagram can be drawn and examined in terms of the system perturbation. The aquagram provides a convenient image for examining the cause and effect of system perturbation in terms of the underlying water structure.

7.3 Results and discussion

7.3.1 The raw spectra (1300-1600 nm) of the kiwifruit juice

The plots in Fig. 7.2 show the absorbance curve of the juice in the first overtone region. There is a slight shift to longer wavelengths with a minor shortening of the peak and a decrease in intensity for the CI juice (Fig. 7.2 (b),(c)). The prominent feature is a broad absorbance peak centered around 1451 nm (Fig. 7.2 (a),(b),(c)) attributed to various water states or species related to the first overtone of the O-H stretch (Burns & Ciurczak, 1992).

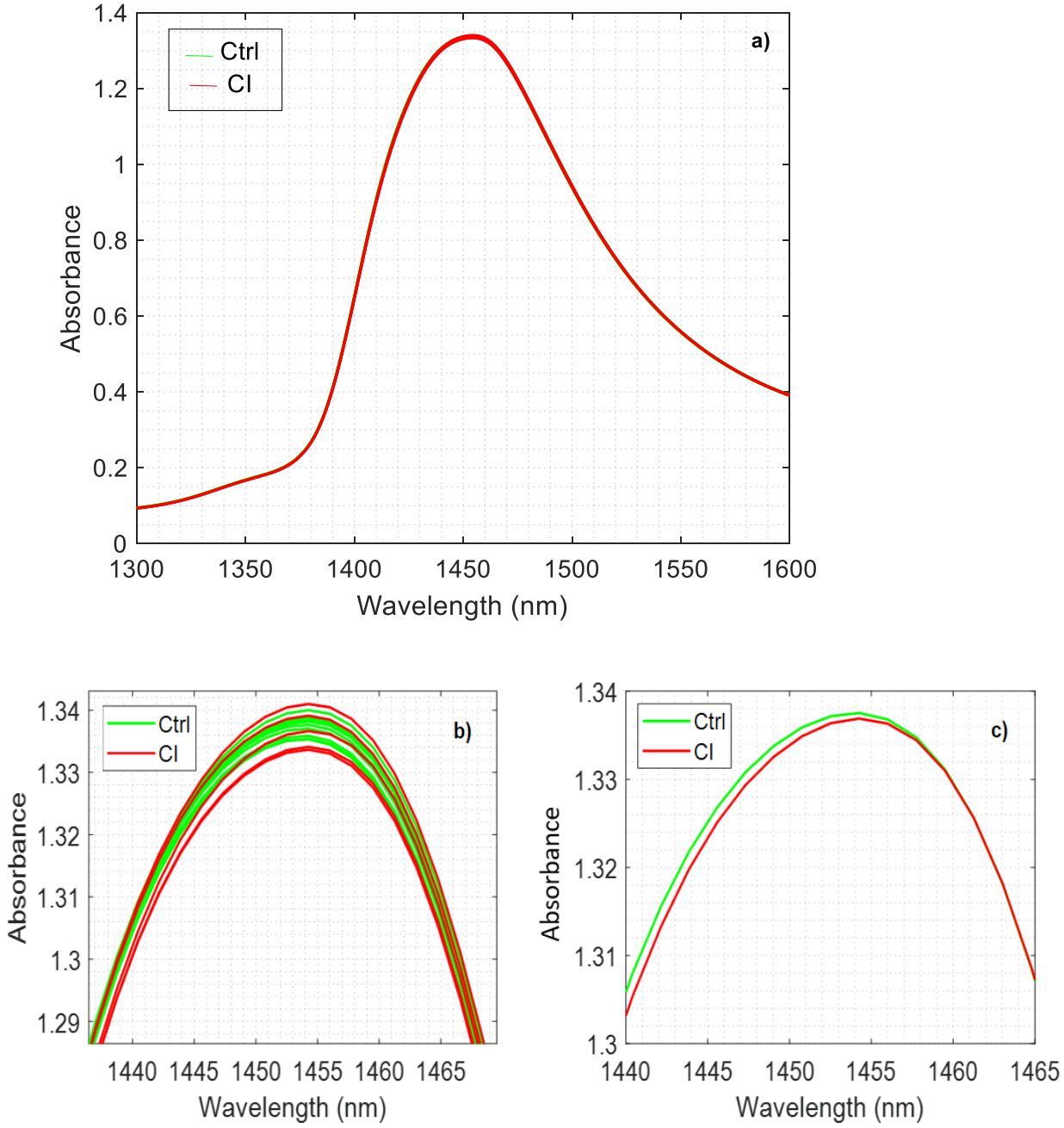


Figure 7.2 Absorbance spectra of kiwifruit juice in the first overtone region of the OH stretch of water: a) raw; b) magnified; and c) averaged and magnified. CI represents chilling injury, and Ctrl represents control fruit juice.

7.3.2 Aquaphotomics Analysis (First overtone region)

7.3.2.1 Second derivative absorption spectra (mean): Raw CI and Raw Control

The second derivative spectrum (Fig. 7.3) were used to assist in the identification of further absorption bands, particularly small and/or overlapping absorption peaks not resolvable in the original raw spectra. The free water S0 band is strong at 1415 nm, and bands corresponding to higher-order H-bond water species, such as dimers (S2) and trimers (S3) (Segtnan, 2001; Bázár et al., 2015), are attributed to peaks/troughs in the region from 1441–1500 nm. The broad shoulder-like peak at 1520 nm is attributed to even more highly organized water structures with strong H-bonds (Tsenkova et al., 2015). Comparing the identified bands with those of sucrose solution (Chapter 3, Fig. 3.4), the features looked similar but a slight 2 nm shift was observed in juice samples. This might be due to other components such as acids, scattering materials, etc. present in juice.

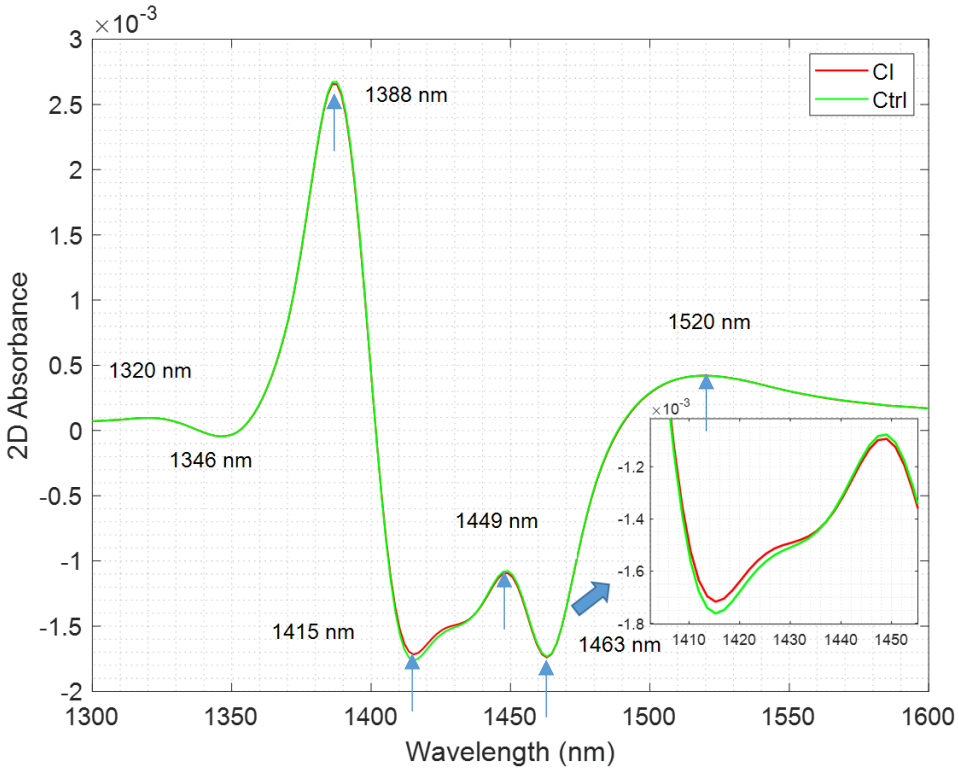


Figure 7.3 Average raw second derivative spectra of CI and Ctrl kiwifruit juice. Labels show peak wavelengths of the CI fruit.

7.3.2.2 Difference absorption spectrum (mean): Raw CI – Raw Control

The average difference absorption spectrum of the kiwifruit juice decreased at 1415 nm, while it slightly increased at 1496 nm with respect to the zero axis (Fig. 7.4). Characteristic water bands were observed at 1415 nm due to S0 and at 1496 nm due to S4 (Siesler et al., 2001; Tsenkova, 2002; Bázár et al., 2015).

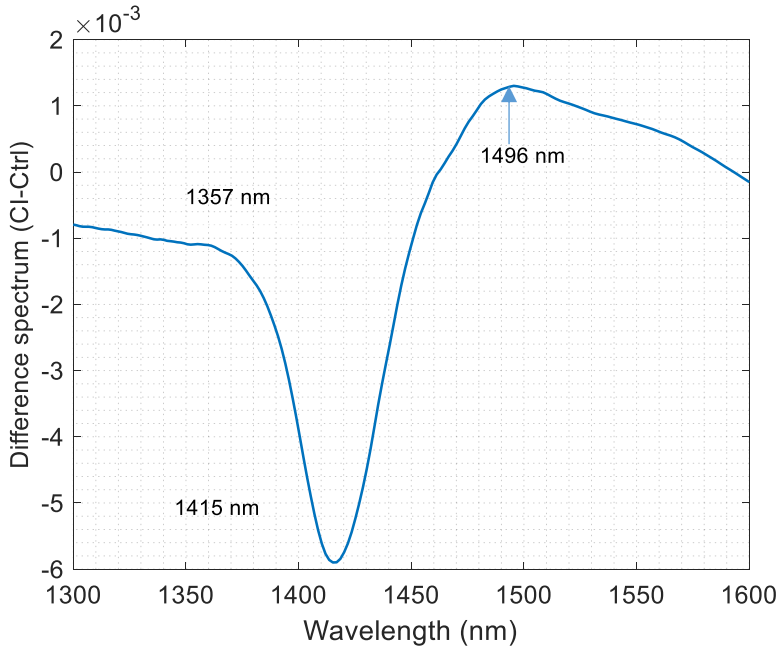


Figure 7.4 Average raw absorbance difference spectrum of CI kiwifruit juice after subtracting the spectrum of control kiwifruit juice at 22°C.

7.3.2.3 Second derivative difference absorption spectrum (mean): Raw CI – Raw Control

The second derivative of the difference absorption spectrum of the kiwifruit juice revealed many other hidden bands (Fig. 7.5) such as those at 1349 nm and 1470 nm, which are related to characteristic water bands of v_3 and S3 (Siesler et al., 2001; Tsenkova, 2002; Bázár et al., 2015).

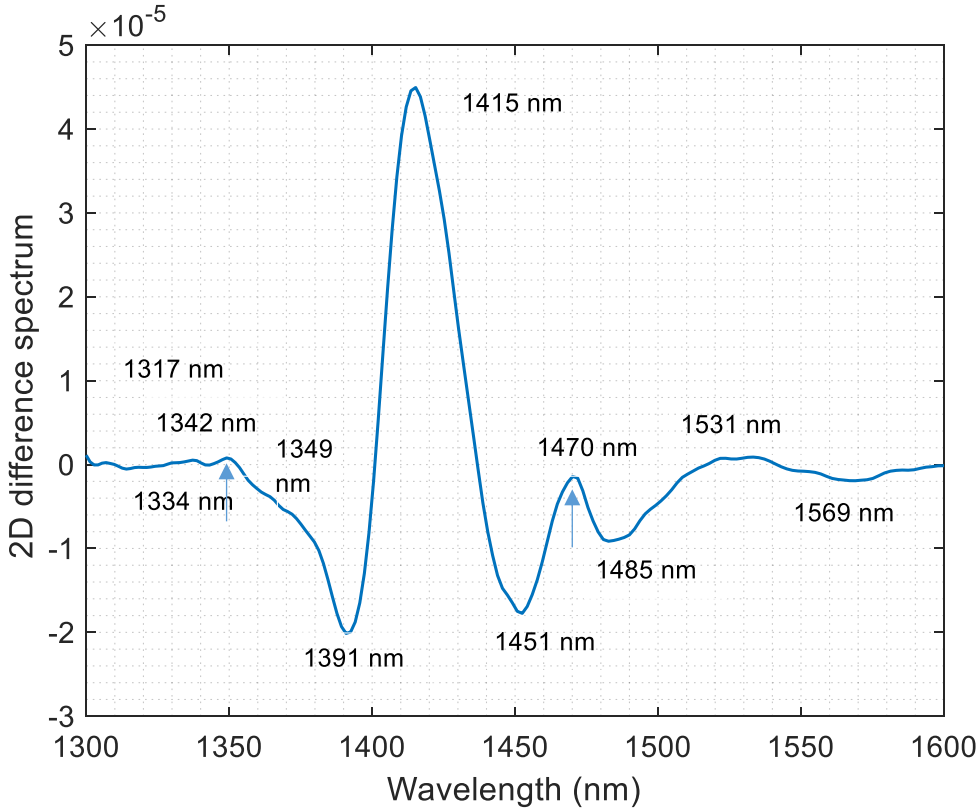


Figure 7.5 Second derivative of the average raw absorbance difference spectrum of CI kiwifruit juice after subtracting the average raw absorbance spectrum of control kiwifruit juice at 22°C.

7.3.2.4 PCA loadings of Raw CI and Raw Control

Three principal components (PCs), which explained over 99% of the spectral variation, were extracted from the raw absorbance data for the first overtone region (Fig. 7.6). The first PC (Fig. 7.6 (b)) explained 57% of the variation in the data and showed a single deep peak centered at 1415 nm, attributed to the OH stretch of the free water molecule (S0) (Tsenkova et al., 2015). That first PC demonstrated a strong negative correlation with CI. The first PC score plot (Fig. 7.6 (a)) shows that with a decrease in CI, free water at 1415 nm increases. The second and third PCs explained 40% and 2% of the variation, respectively (Fig. 7.6 (b)).

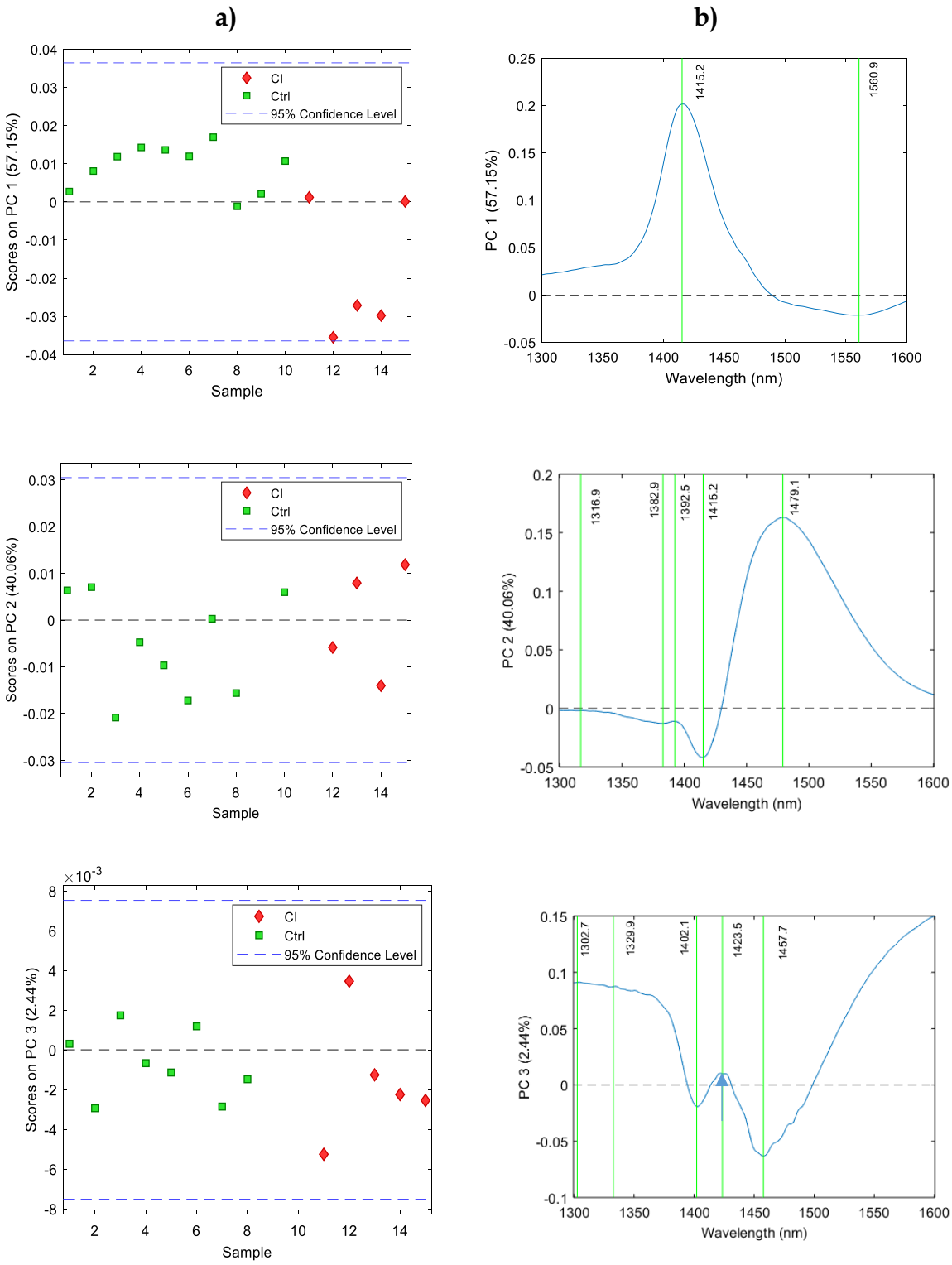


Figure 7.6. PCA: a) score; and b) loading plots of raw absorbance spectra of kiwifruit juice in the first overtone region of water with explained variance shown with each y-axis label for each PC.

7.3.3 The raw spectra (800-1100 nm) of the whole intact fruit

The absorbance plot in the second overtone region of the OH stretch of water in Fig. 7.7 illustrates that the average absorbance curve of CI fruit was slightly shifted to shorter wavelengths at 976 and 1053 nm (Fig. 7.7 (b)). The prominent feature is a broad absorbance peak centered around 976 nm (Fig. 7.7 (a),(b)) attributed to various water states related to the second overtone of the OH stretch (Burns & Ciurczak, 1992).

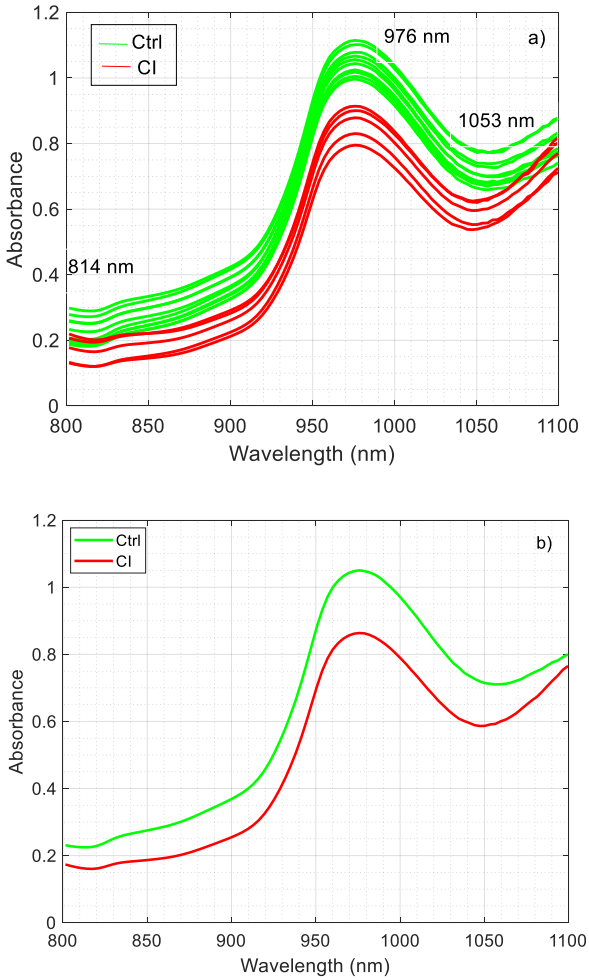


Figure 7.7 Raw absorbance spectra of whole, intact kiwifruit in the second overtone region of OH stretch of water: a) raw; and b) average. CI represents chilling injury fruit, and Ctrl represents control fruit.

7.3.4 Aquaphotomics Analysis of Intact fruit (Second overtone region)

7.3.4.1 Second derivative absorption spectra (mean): Raw CI and Raw Control

The second derivative spectrum (Fig. 7.8) were used to assist in the identification of further absorption bands, particularly small and/or overlapping absorption peaks not resolvable in the original raw spectra. The dip was strong at 963 nm, which is related to water bands with strong H-bonding (Sirinnapa & Sumio, 2008).

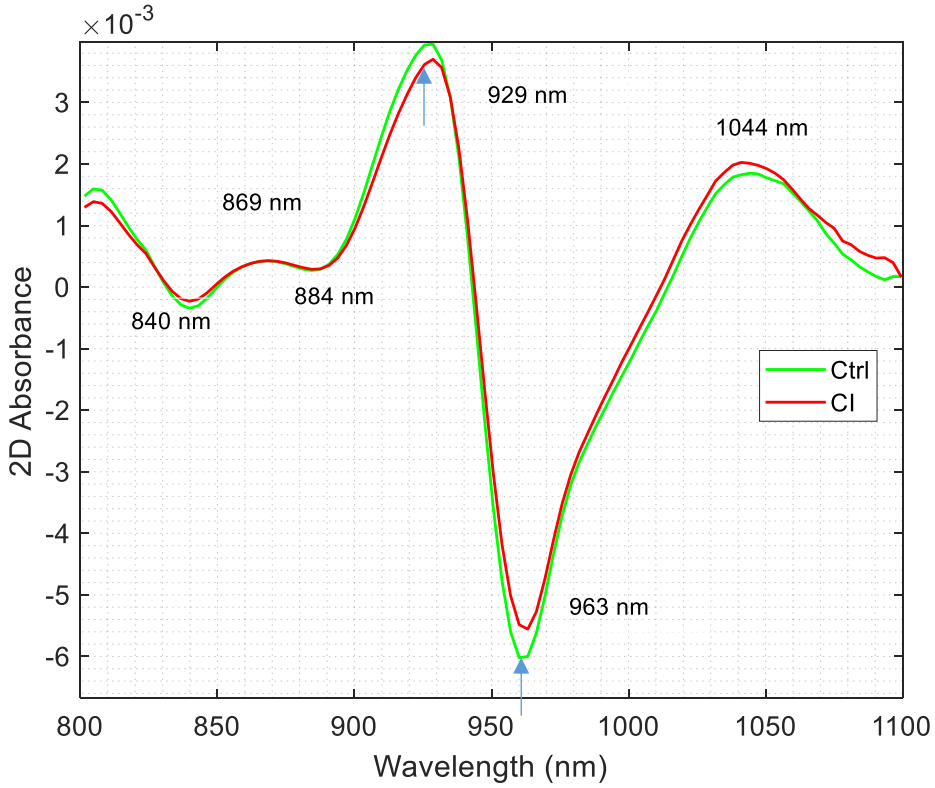


Figure 7.8 Average raw second derivative spectra of CI and Ctrl intact kiwifruit in the second overtone region. Labels show peak wavelengths for the CI fruit.

7.3.4.2 Difference absorption spectrum (mean): Raw CI – Raw Control

The difference absorption spectrum of the intact kiwifruit decreased at 966 nm with respect to the zero axis (Fig. 7.9). The dip was broad and strong at 966 nm.

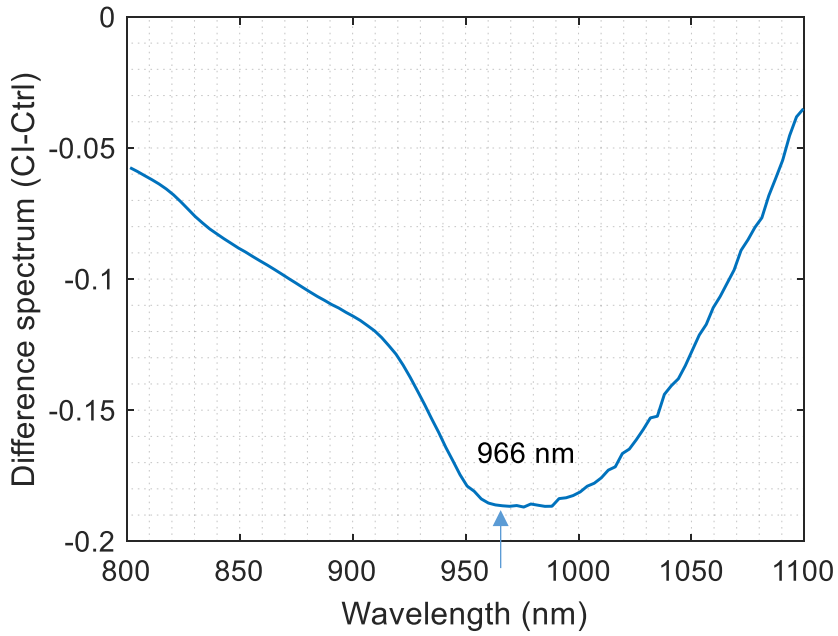


Figure 7.9 Average raw absorbance difference spectrum of CI intact kiwifruit after subtracting the spectrum of Ctrl kiwifruit in the second overtone region at 22°C.

7.3.4.3 Second derivative difference absorption spectrum (mean): Raw CI – Raw Control

The second derivative of the difference absorption spectrum of the intact kiwifruit spectra revealed many other hidden bands (Fig. 7.10) such as those at 864 nm, 919 nm, 954 nm, 992 nm, and 1016 nm. The water band at 992 nm and 1016 nm were due to S3 and S4 water species (Sirinnapa & Sumio, 2008).

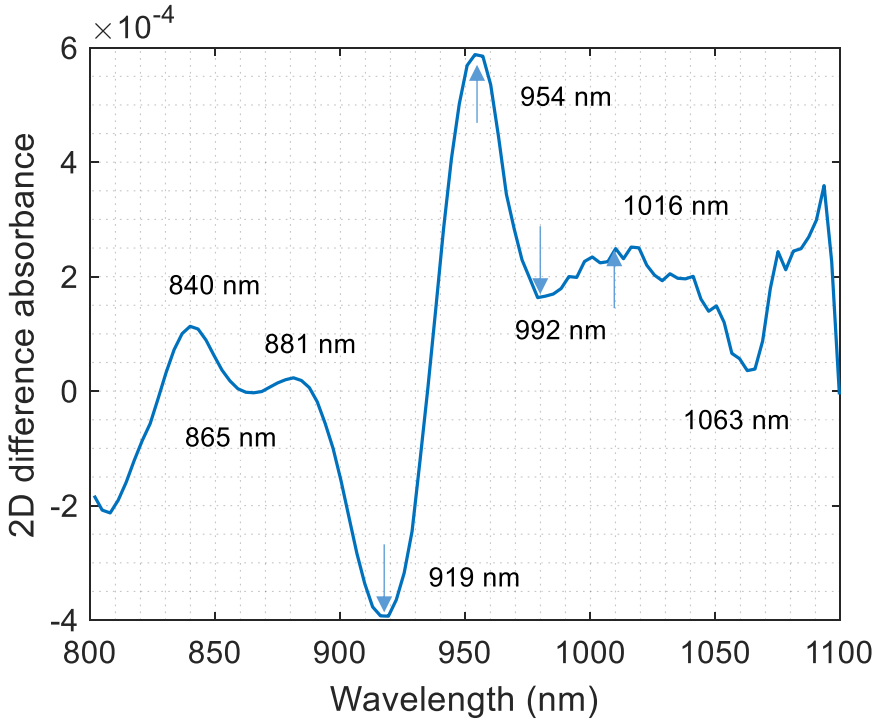


Figure 7.10 Second derivative of the average raw absorbance difference spectrum of CI intact kiwifruit after subtracting the spectrum of Ctrl kiwifruit at 22°C.

7.3.4.4 PCA loadings of Raw CI and Raw Control

Three principal components (PCs), which explained over 99% of the spectral variation, were extracted from the raw absorbance data for the first overtone region (Fig. 7.11). The first PC (Fig. 7.11(b)) explained 97% of the variation in the data and showed a single deep peak centered at 979 nm, attributed to vibrations involving a higher number of H-bonds. The first PC demonstrates a strong negative correlation with CI. The first PC score plot (Fig. 7.11 (a)) shows that with a decrease in CI, the intensity at 979 nm increases. The second and third PCs explained 2% and 0.1% of the variation, respectively (Fig. 7.6 (b)).

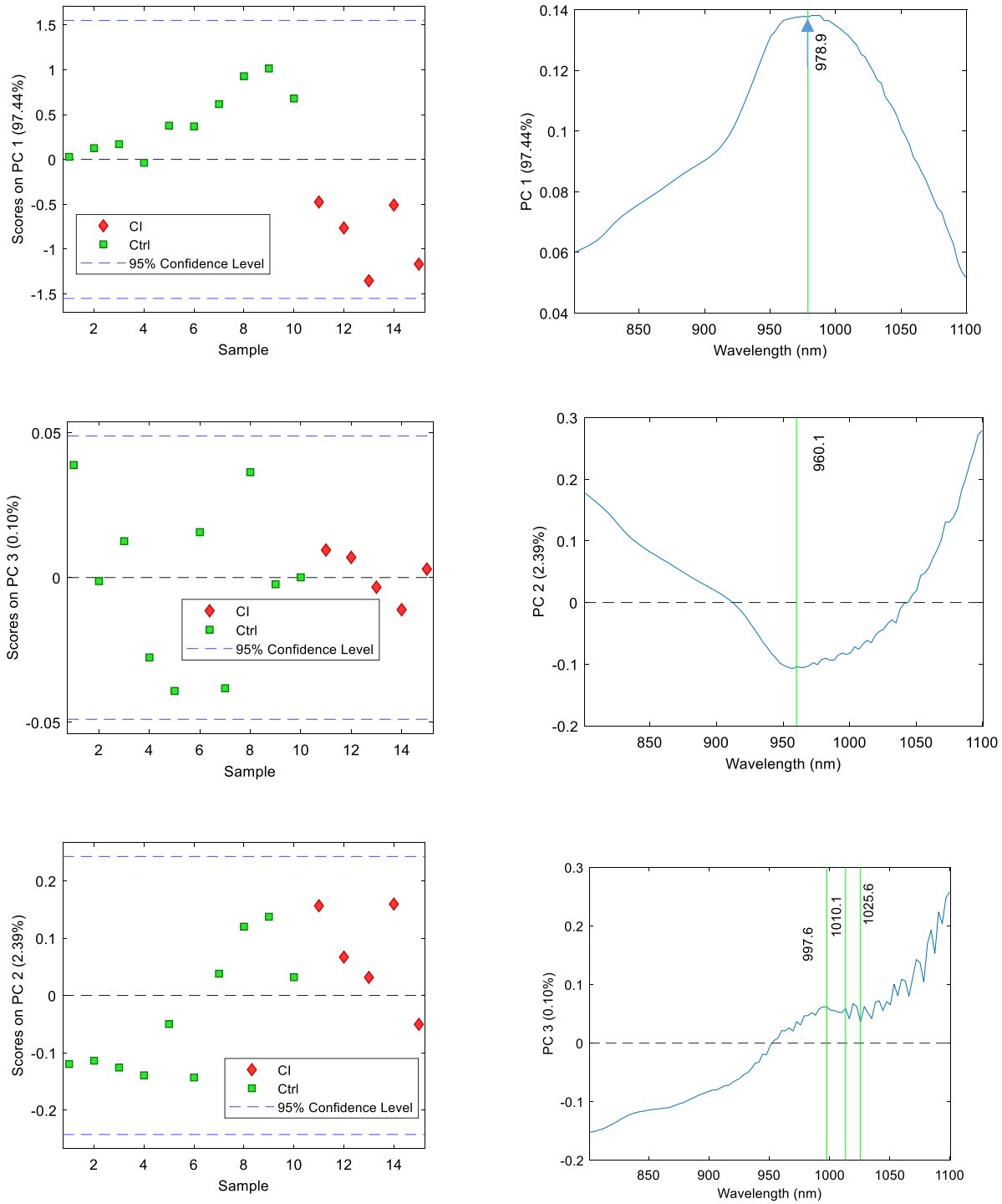


Figure 7.11. PCA: a) score; and b) loading plots of raw absorbance spectra of intact kiwifruit spectra in the second overtone region of water with explained variance shown with y-axis label of each PC.

7.3.4 Identification of WAMACS

WAMACS were created using the feature wavelengths found using different pre-processing techniques as described in section 7.3.2 and 7.3.3. The wavelengths selected for each WAMACS are provided in Table 7.1.

Table 7.1 Summary and assignment of the characteristic water absorbance bands for CI kiwifruit

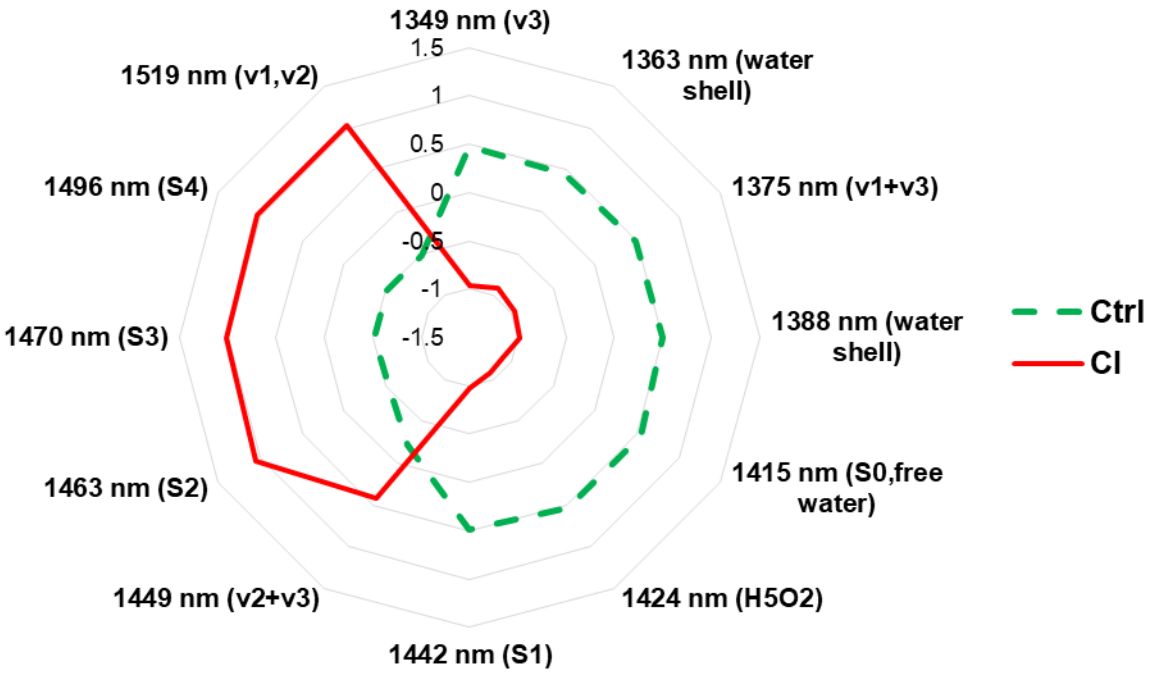
WAMACS	Assignment	Wavelengths in overtone region		Activated wavelengths, nm	
		First (1300-1600 nm)	Second (800-1100 nm)	First overtone	Second overtone
C1	$v3$	1336-1348	900-908	1349	901
C2	OH stretch (water shell)	1360- 1366	916- 920	1363	919
C3	$v1+v3$	1370-1376	923-927	1375	924
C4	OH stretch (water shell)	1380-1388	930- 935	1388	929
C5	S0 (free water)	1398-1418	942-955	1415	954
C6	Water hydration, H_50_2	1421-1430	957-963	1424	963
C7	S1	1432-1444	965-973	1442	966
C8	$v2+v3$	1448-1454	975-979	1449	979
C9	S2	1458-1468	982-989	1463	985
C10	S3	1472-1482	992-998	1470	992
C11	S4	1482-1495	998-1007	1496	1010
C12	Strongly bonded water or $v1,v2$	1506-1516	1014-1021	1519	1016

7.3.5 Aquagrams

The aquagrams of average juice spectra (Fig, 7.12(a)) and whole, intact kiwifruit (Fig. 7.12(b)) are shown for the two overtone regions. All the bonded water species increased with CI in the first overtone region.

However, the water hydration (H_5O_2) only increased with CI for the second overtone region, with a decrease in S4 and strongly bonded water. As the CI increases in the juice and whole intact fruit, bonded water increases; the water structure becomes more organized, with more stable H-bonds (Fig. 7.12 (a),(b)). In the second overtone region compared with the first overtone region, different water species showed changes. The expectation was that the WASP would be the same in both regions; however, the findings were different.

a) The first overtone



b) The second overtone

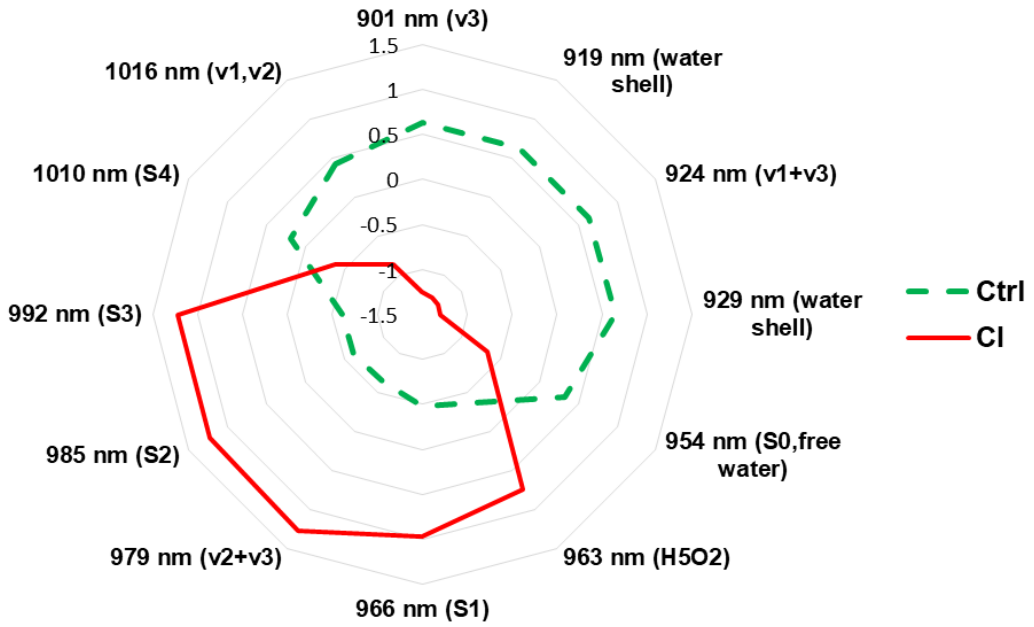


Figure 7.12 Aquagrams of Ctrl and CI: a) kiwifruit juice in the first overtone region of OH stretch of water; and b) intact kiwifruit in the second overtone of water.

7.4 Conclusions

Aquaphotomics helped in the identification of the water wavelengths in the two overtone regions that showed changes due to the effect of CI in the fruit. Different water species in the second overtone region show the CI perturbation compared with the first overtone region. The expectation was that the WASP would be the same in both regions. However, the findings were different. The S4 water species decreased in the second overtone region. For both juice and intact fruit measurements, CI increases H-bonded water structures despite the fruit having the same SSC concentrations.

Chapter 8: Whole intact fruit (the second overtone region)

8.1 Introduction

The main purpose of this chapter was to examine whether aquaphotomics can help in the development of better and simpler prediction models and/or hardware for intact fruit quality measurement beneficial to the fruit industry. In previous chapters, experiments were performed on simple aqueous systems such as sucrose solutions of various concentrations and aqueous systems such as apple and kiwifruit juice. The second overtone wavelength region was scrutinized for fruit juice measurements using a long pathlength cell. This chapter focuses on intact fruit measurement, which is different from juice sample measurement. For NIR measurement of juice samples, a transmission setup with a fixed pathlength cuvette is used (see Chapters 4 and 5). In contrast, for fruit spectral measurements, interactance-based instruments are normally used (Fig. 8.1 and Fig. 8.2). Here the penetration depth of light inside the fruit is not fixed, since the light is scattered inside the fruit. This phenomenon changes the pathlength of light between the source and the detector and thus also affects the absorption. Applying the aquaphotomics concept to intact fruit is challenging. Due to the high absorption of water at 1300–1600 nm, intact fruit cannot be scanned with the interactance-based systems because the signal to noise ratio of the returned signal is poor in the first overtone region. Therefore, fruit measurements are usually done in the wavelength region below 1100 nm (McGlone et al., 2003; Shafie et al., 2015).

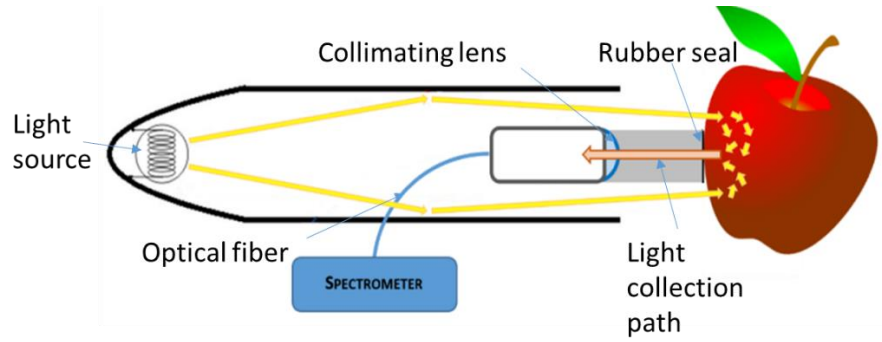


Figure 8.1 Interactance optical design; light rays illuminate an annulus on the sample. The light then interacts with the sample by internally scattering through the tissue. The light that undergoes remission normal to the collimating lens is collected and focused onto the fibre ("F-750 Instruction Manual," 2019).

The objective of this chapter was to investigate the application of aquaphotomics concepts to intact fruit with varying DM concentrations at a constant temperature (22°C). The natural variation in quality attributes across a range of samples is considered the perturbation (Tsenkova et al., 2015; Tsenkova et al., 2018). The key focus was on the usefulness of the aquagram – the water spectral pattern (WASP) depicting the influence of the main water molecular species – in both understanding and developing predictive models for the DM of intact fruit, using two different spectrophotometers.

A goal was also to determine a small number of discrete wavelengths to observe whether WAMACS are optimal or other better wavelengths are available that could be used to build calibration models. By undertaking an exhaustive wavelength search in the 800–1100 nm region using the "branch and bound" algorithm (Shafie et al., 2015), and from 12 aquaphotomics water wavelengths, the optimal wavelengths were identified.

Another section of this chapter focuses on the influence of temperature on intact apples. In this chapter, changes in the water structure of apple juice caused by variations in temperature are investigated using aquaphotomics. The work focuses on spectra in the 800-1000 nm region, surrounding the strong water absorption peak at 970 nm. Aquaphotomics analysis has suggested the possible advantageous use of EMSC and EPO pre-processing, using an interferent derived from pure water spectra. This may help to remove the temperature sensitivity of DM in apple NIR calibrations.

8.2 Methods and materials

8.2.1 Sample preparation

Lot I: Kiwifruit at one temperature

A total of 100 Gold Kiwifruit *"Actinidia chinensis"* were purchased from New Zealand retail stores. The fruit were kept at room temperature (20±2°C) in the laboratory and measured within two days of purchase. Reference DM measurements were made within 2 to 3 h of the spectral measurements.

Lot II: Apples at three temperatures (10°C, 20°C, 26°C)

A total of 45 *"Royal Gala"* apples were purchased from New Zealand retail stores. The fruit were stored in a refrigerator at 4°C for 6 h. Prior to spectral measurement, the fruit were taken out of the refrigerator, and the near-

surface fruit temperature was allowed to increase to approximately 10°C. The room temperature was set to 20°C. The temperature of the fruit was measured using a non-contact infrared thermometer (QM222, Digitech). The first spectral measurements were taken at 10°C. After this, the fruit was left for 3 hours to allow the fruit temperature to equilibrate to 20°C. Another set of spectral measurements were then taken at 20°C. Eventually, to increase the temperature of fruit to 26°C, the fruit were stored in an incubator for 3 hours at 26°C, and measurements were taken at the higher temperature.

Spectral measurements of Milli-Q water were also acquired at three temperatures; 10°C, 20°C, and 26°C to enable temperature corrections. The temperature of the Milli-Q water was controlled using a water bath (Grant Instruments, Cambridge, England).

8.2.2 DM measurement

For Lot I samples, kiwifruit DM was measured by cutting one equatorial slice of 3 mm thickness from the middle of each fruit, and drying these slices at 65°C for 24 h. The fruit DM was calculated as a percentage final dry weight of the initial wet weight of the slices (McGlone & Kawano, 1998).

For Lot II samples, apple DM was measured by cutting two equatorial slices, of approximately 3 mm thickness, and drying them at 65 °C for 24 h. The fruit DM was calculated as a percentage final dry weight of the initial wet weight of the slices. The two DMs values were then averaged to get one mean DM per fruit.

8.2.3 Spectral acquisition

For Lot I kiwifruit, non-destructive NIR interactance measurements were acquired in the 300–1100 nm range using two hand-held NIR instruments (Fig. 8.2); an F-750 Produce Quality Meter (Felix Instruments, Portland, USA) and a customised in-house benchtop spectrophotometer (McGlone et al., 2002c; Kaur et al., 2017b) at a room temperature of 22°C. Fruit spectra were recorded, taking two separate measurements on opposite sides in the equatorial plane of each fruit. The total number of fruit spectra per instrument was 200 (100 samples × 2 consecutive scans). The two consecutive scans were averaged to get one mean spectrum per fruit.

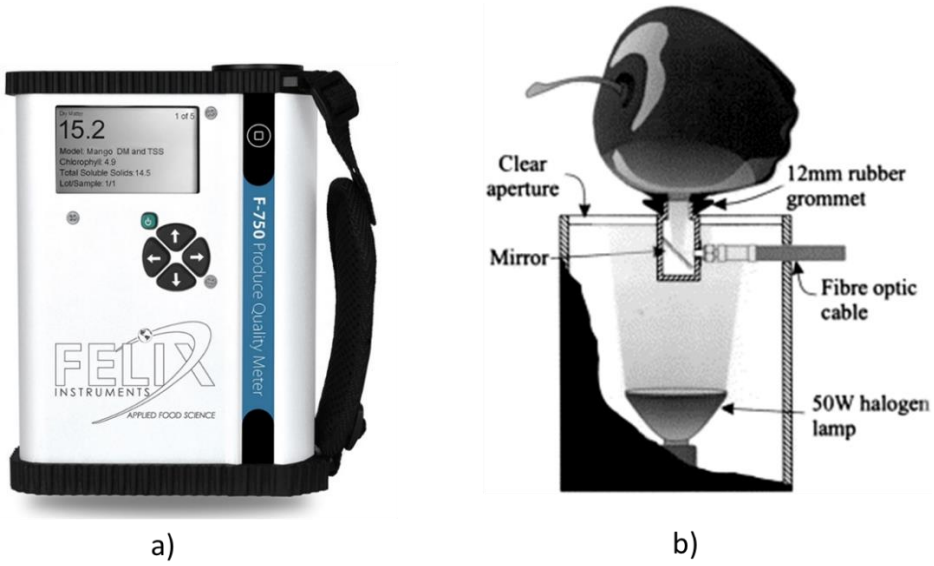


Figure 8.2 NIR portable spectrophotometers used for the experiment a) F-750 Produce Quality Meter (Felix Instruments, Portland, USA); and b) benchtop spectrophotometer (McGlone et al., 2002a).

For Lot II apples, non-destructive NIR interactance measurements were acquired in the 300–1100 nm range using the benchtop spectrophotometer (Fig. 8.2(b)). The Milli-Q water was also measured with

the benchtop spectrophotometer. Measurements on water were performed by placing a glass petri dish (5 cm diameter and 1.5 cm height) filled with approximately 10ml of water onto the rubber grommet (Fig. 8.2 (b)). The top of the petri dish was covered with a white Teflon block to reflect the light into the detector. A total of 10 samples were each measured twice at three different temperatures. Thus, in total, 60 water spectra (10 samples x 2 consecutive scans x 3 temperatures) were collected.

8.2.4 Aquaphotomics analysis

The water bands were calculated using the standard protocol described in Chapter 3. Once the WAMACS wavelengths are selected, WASPs presented on the aquagram can be drawn and examined in terms of the system perturbation. The aquagram provides a convenient graphic view for examining the cause and effect of system perturbation in terms of the underlying water structure.

8.2.5 Multivariate and statistical analysis

The analysis involved the development of predictive models using the pre-processed forms of:

- 2D: This was standard second derivative processing (Savitzky–Golay 2nd order derivative with smoothing parameters: width 15, order 2, polynomial 2) of the raw spectra for Lot I kiwifruit samples.
- EMSC. This was extended multiplicative scatter correction of the spectra, as described in Chapter 2 for Lot II apple samples.

- EPO. This was external parameter orthogonalisation of the spectra, as described in Chapter 2 for Lot II apple samples.

The PLSR models for DM prediction were developed using the PLS toolbox version 8.6.2 (Eigenvector Research Inc., Wenatchee, USA) operating under MATLAB version 2018b (Math Works Inc., Natick, USA). The samples were rank ordered according to their DM value. Four-fold venetian blind cross validation was applied, to ensure there was no over fitting, with latent variables selected based on the lowest root mean square error of cross validation (RMSECV). For Lot I samples, a set of two different wavelength ranges was investigated at one temperature of 22°C for model development: 800–1000 nm and 600–1100 nm (whole wavelength range). The validation set (25 samples) was selected by taking every fourth sample from the main data set, and the remaining 75 samples formed the calibration set

For Lot II samples, the samples were rank ordered according to their DM value. The data set in each wavelength range was split into three temperature subsets; 10, 20, and 26°C. The validation set (15 samples) was selected by taking every third sample from the main data set, and the remaining 30 samples formed the calibration set. Calibration models built at one temperature were then applied to samples at the other temperatures.

8.3 Results and discussion

The DM of Lot I kiwifruit samples varied from 13.8 to 20.7% with a standard deviation of 1.29% (Table 8.1). For convenience in the aquaphotomics analysis, the samples were grouped into three classes: 'Low DM' (13.8– 16%), 'Med DM' (16.5–18%), and 'High DM' (19–20.7%), with each group containing equal class interval.

Table 8.1. Distribution statistics for DM levels on different DM class

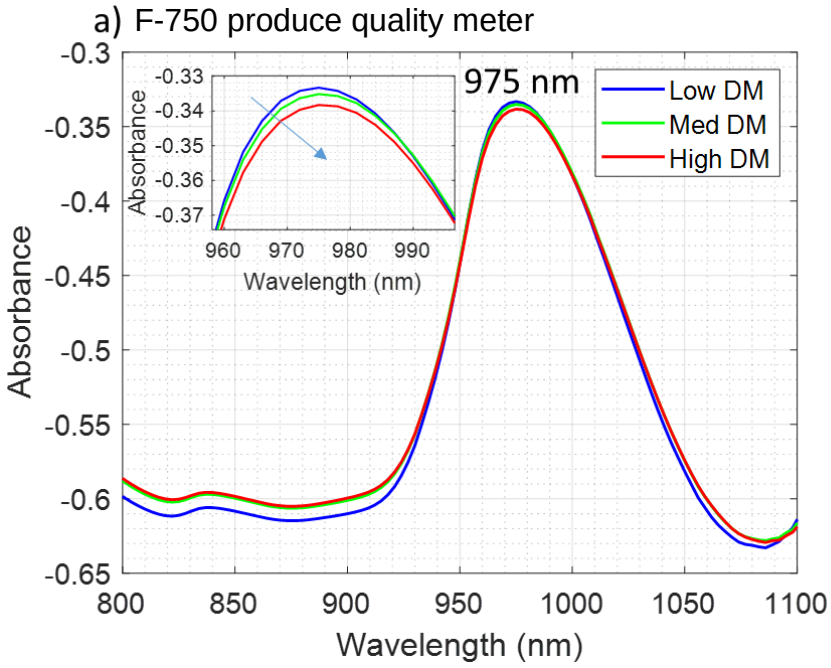
Lot I: Kiwifruit	**Sample class**	**DM (%)**	**Mean**
	Low DM	13.8-16	15.30 (SD=0.95)
	Med DM	16.5-18	17.98 (SD=0.74)
	High DM	19-20.7	19.75 (SD=0.61)
	Total	13.8-20.7	18.40 (SD=1.29)
Lot II: Apple	**Sample class**	**DM (%)**	**Mean**
	Low DM	11.12-13.34	12.47 (SD=0.79)
	Med DM	13.37-15.56	14.39(SD=0.70)
	High DM	15.68-17.79	16.45 (SD=0.84)
	Total	11.12-17.79	14.20 (SD=1.47)

The DM of Lot II apple samples varied from 11.12 to 17.79% with a standard deviation of 1.47% (Table 8.1). The samples were grouped into three classes: 'Low DM' (11.12–13.34%), 'Med DM' (13.37–15.56%), and 'High DM' (15.68–17.79%), with each group containing equal class interval.

8.3.1 The raw spectra of Lot I Kiwifruit

The absorbance plots in the second overtone region for the two spectrophotometer units are shown in Fig. 8.3. The absorbance curves obtained using the F- 750 unit were slightly shifted to shorter wavelengths with no baseline offset (Fig. 8.3(a)) as compared to the absorbance spectra

obtained using the benchtop instrument (Fig. 8.3(b)). The prominent feature in both of these plots is a broad absorbance peak centred around 975 nm and 980 nm (Fig. 8.3 (a),(b)) attributed to various water states related to the second overtone of the OH stretch of water (Burns & Ciurczak, 1992). The position of the water peak varied between instruments. The peak absorbance was highest for low DM and then decreased with increasing DM content for the F-750 produce quality meter. The rather flat baseline of the raw spectra produced by the F-750 instrument suggests an unknown pre-processing method has probably been used to remove any baseline offset and slope. In contrast, the raw absorbance spectra of the benchtop spectrophotometer had a strong baseline slope, angling up towards the longer wavelength end (Fig. 8.3(b)).



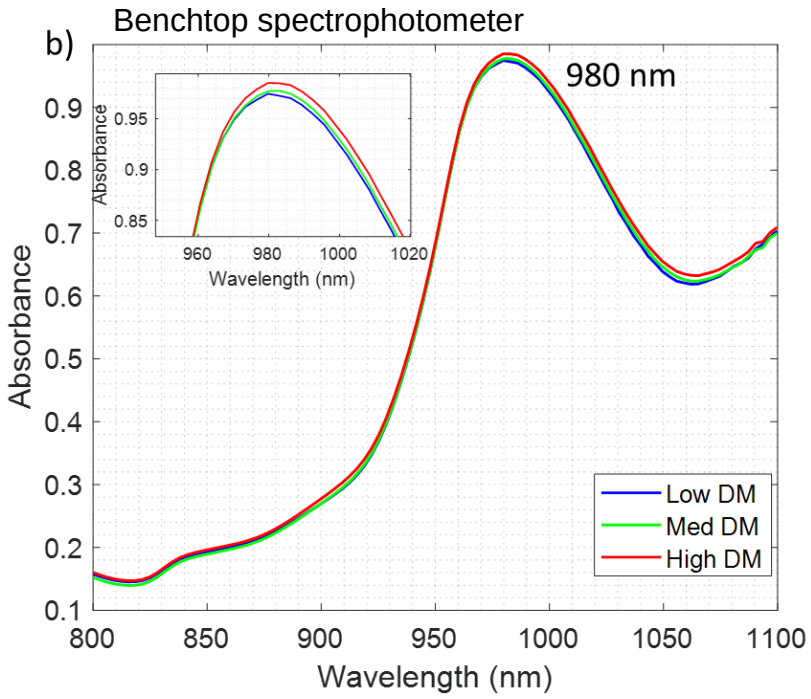


Figure 8.3 Average raw absorbance spectra of intact kiwifruit at three DM levels in the second overtone region collected by a) F-750 produce quality meter; and b) benchtop spectrophotometer.

8.3.2 Aquaphotomics Analysis

8.3.2.1 Second derivative absorption spectra (mean): Raw Kiwifruit

The second derivative spectra (Fig. 8.3) were used to assist in the identification of further absorption bands, particularly small and/or overlapping absorption peaks not resolvable in the original raw spectra. The dip was strong at 966 nm for the F-750 and at 970 nm for the benchtop instrument (Fig. 8.4), which is related to water bands with strong H-bonding (Sirinnapa & Sumio, 2008).

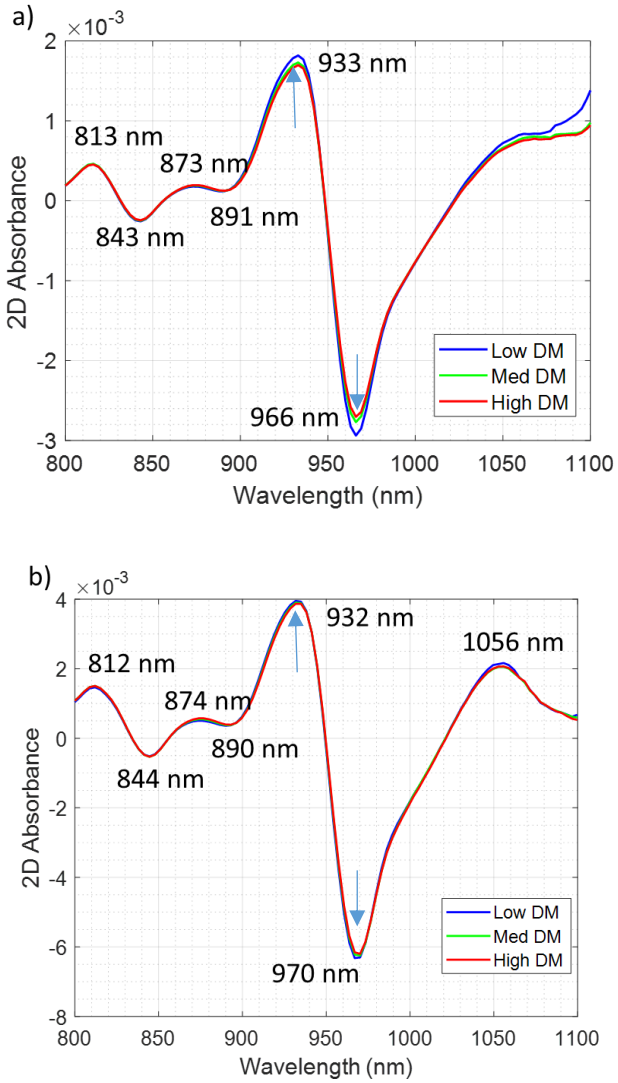


Figure 8.4. Average raw second derivative absorbance spectra of kiwifruit at three DM levels in the second overtone region for a) F-750 produce quality meter; and b) benchtop spectrophotometer.

8.3.2.2 Difference absorption spectra (mean): Raw fruit – Raw lowest DM

The difference absorption spectra of the intact kiwifruit decreased at 966 nm with respect to the zero axis in measurements taken using the F-750 meter (Fig. 8.5). The dip was broad and strong at 966 nm, which is related to water bands with strong H-bonding, S1 (Sirinnapa & Sumio, 2008). The benchtop unit showed a different trend with a dip at 960 nm and a peak at 916 nm, with different baselines.

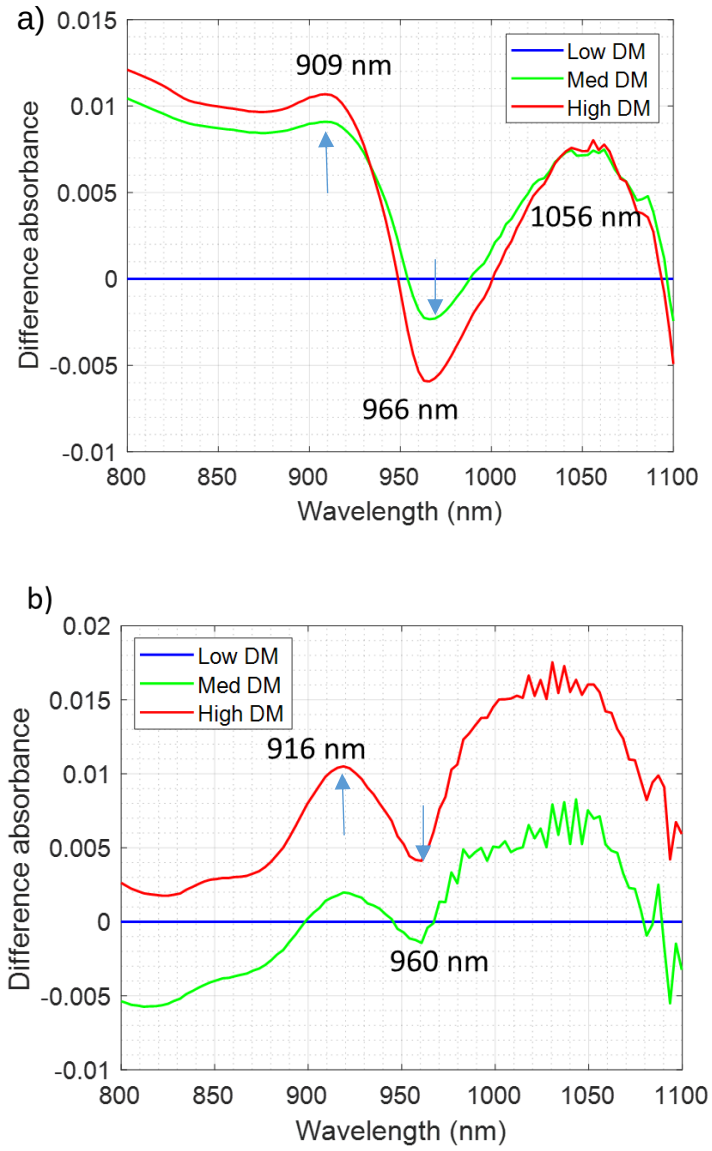


Figure 8.5. Average raw difference absorbance spectra of kiwifruit at three DM levels in the second overtone region a) F-750 produce quality meter; and b) benchtop spectrophotometer.

8.3.2.3 Second derivative difference absorption spectra

The second derivative of the difference absorption spectra of the intact kiwifruit revealed many other hidden bands for the measurements obtained using the F-750 meter (Fig. 8.6(a)), including those at 833 nm, 882 nm, and 930 nm. The benchtop instrument also showed many peaks/dips, particularly at 880 nm, 919 nm, 986 nm, and 1015 nm. Some of the

wavelengths were common in the previous pre-processing techniques (8.3.2.2) such as 966 nm for the F-750 produce quality meter and 960 nm for the benchtop instrument.

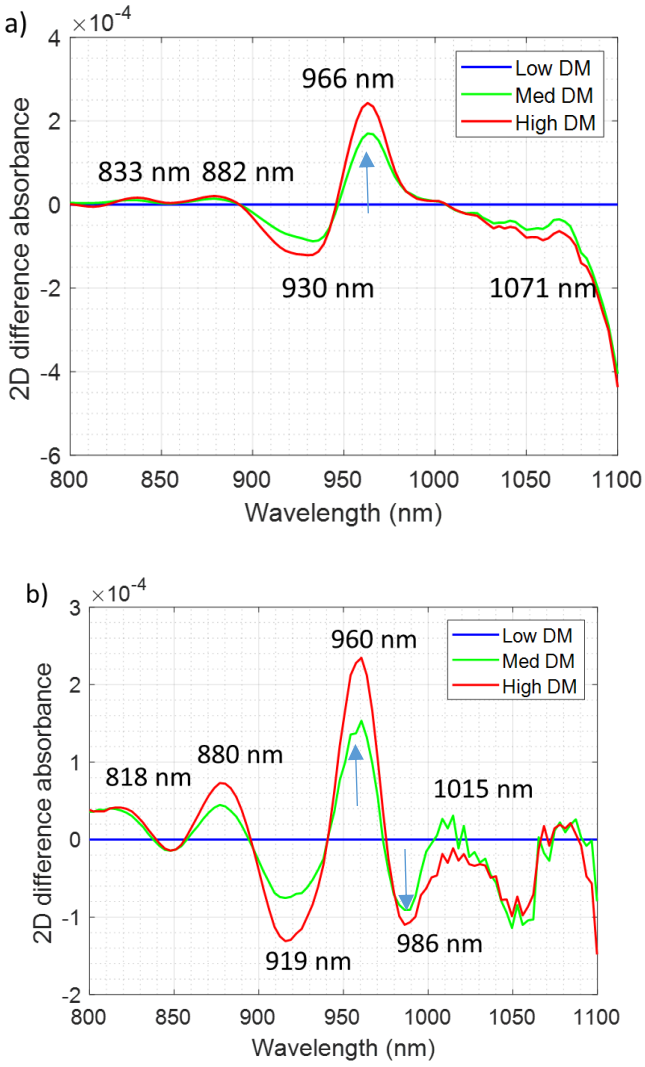
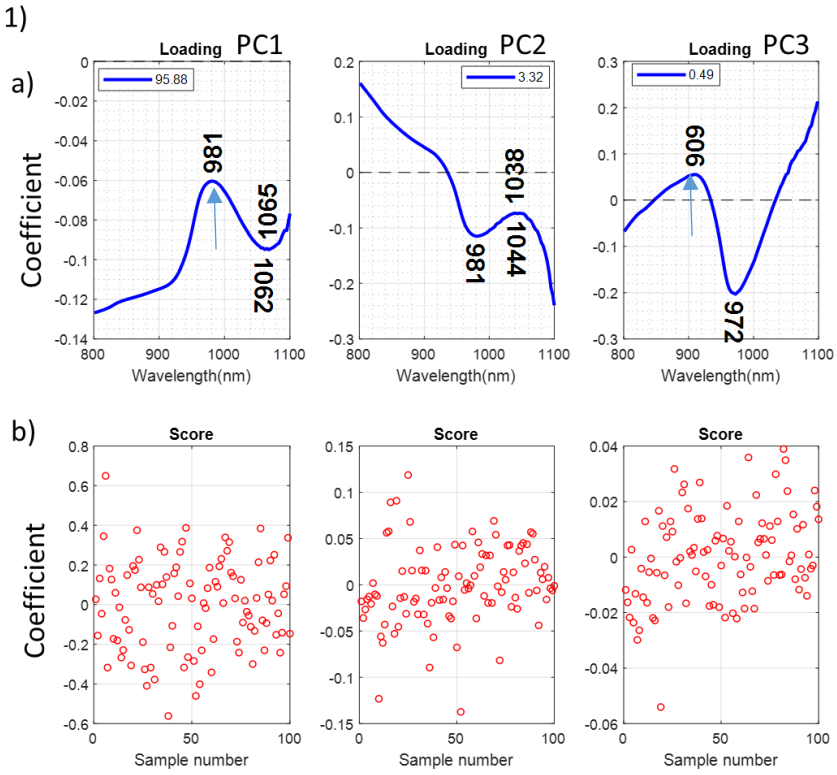


Figure 8.6. Average raw second derivative difference absorbance spectra of kiwifruit at three DM levels in the second overtone region a) F-750 produce quality meter; and b) Benchtop spectrophotometer.

8.3.2.4 PCA loadings of raw fruit

Three principal components (PCs), which explained over 99% of the spectral variation, were extracted from the raw absorbance data for the second overtone region (Fig. 8.6(a)) for both the F-750 produce quality

meter (Fig. 8.7(1)) and the benchtop spectrophotometer (Fig. 8.7(2)). For the F-750 measurements, the first PC (Fig. 8.7(1(a))) explained 96% of the variation in the data. The peak for the first PC was at 981 nm. The third PC showed a single dip centered at 972 nm, attributed to higher H-bonding vibration. The second and third PCs explained 3% and 0.5% of the variation, respectively (Fig. 8.7(1(a))). Peaks were also found using the benchtop spectrometer. The first PC (Fig. 8.7(2(a))) explained 97% of the variation in the data. The dip at the first PC was at 986 nm. The third PC showed a single deep peak centered at 977 nm, attributed to higher H-bonding vibration. The second and third PCs explained 2% and 0.5% of the variation, respectively (Fig. 8.7(2(a))).



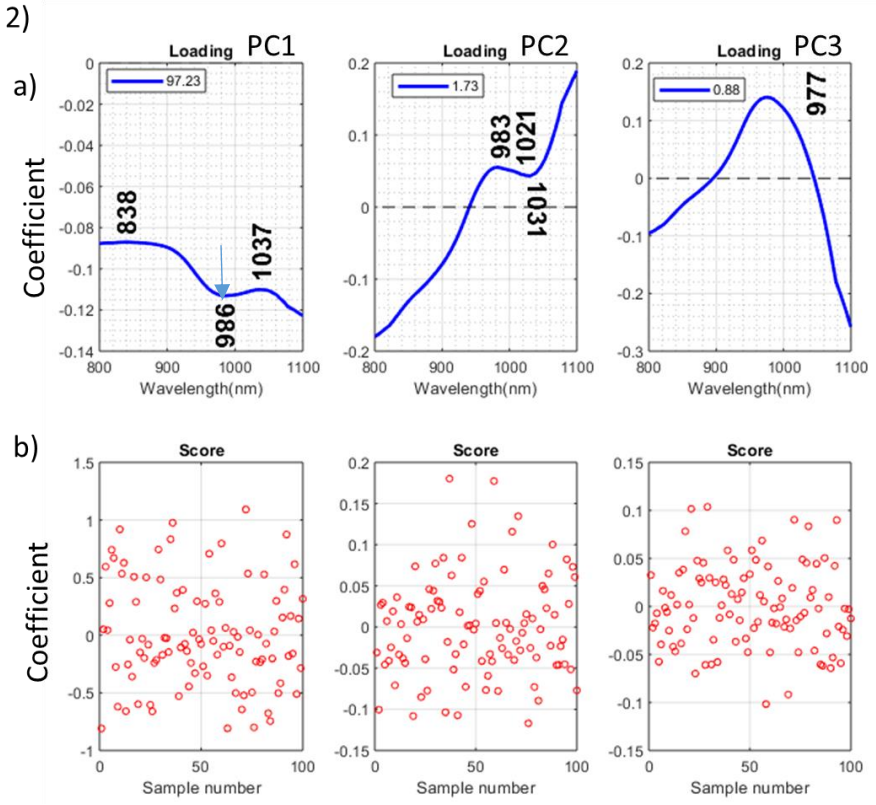


Figure 8.7 PCA a) loading; and b) score plots of raw absorbance spectra of kiwifruit samples arranged in ascending order of SSC in the second overtone region of water with explained variance shown at the top of the loading plots of each PC for 1) F-750 produce quality meter; and 2) benchtop spectrophotometer.

8.3.2.5 PLS regression of raw spectra

PLS models for DM prediction were created using the raw absorption spectra over the 800–1100 nm range. The PLS regression vectors were characterized by many small but sharply defined peaks and dips in the region, resolved at differences of only a few nanometers (Fig 8.8). All the identified peaks, whether large or small, were considered in the assignment of the WAMACS.

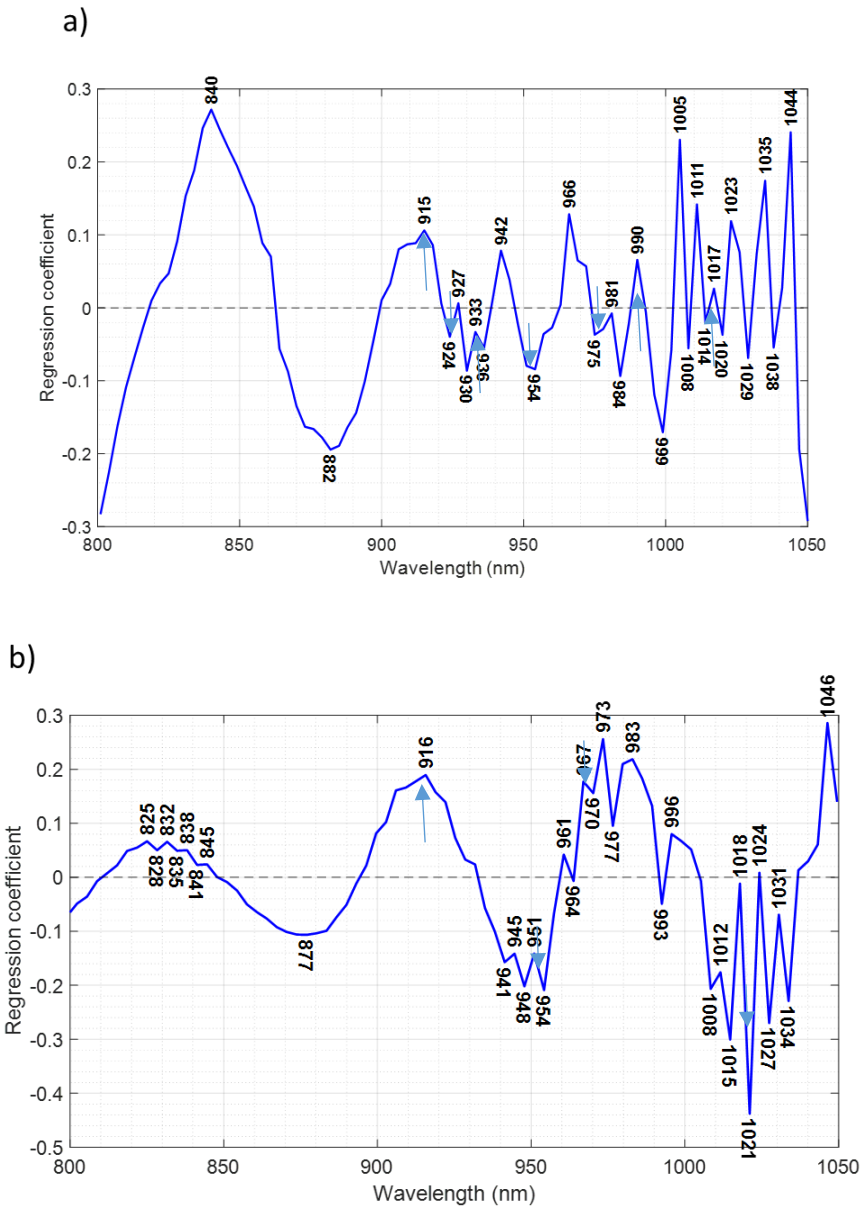


Figure 8.8 PLS regression vector coefficients for DM prediction of whole kiwifruit at 22°C in the second overtone region a) F-750 produce quality meter; and b) benchtop spectrophotometer.

8.3.2.6 Identification of WAMACS

WAMACS were created using the feature wavelengths found using different pre-processing techniques as described in section 8.3.2. The wavelengths selected for each WAMACS are provided in Table 8.1. Some of the wavelengths that were not identified using these pre-processing

techniques were chosen randomly from within each water band to complete the 12 WAMACS for making the aquagram and are highlighted in red.

Table 8.2 Summary and assignment of the characteristic water absorbance bands for DM of intact kiwifruit

WAMACS	**Assignment**	**Wavelengths in overtone region**		**Felix wavelengths, nm**	**Benchtop wavelengths, nm**
		First (1300-1600 nm)	Second (800-1100 nm)	Second overtone	Second overtone
C1	$v3$	1336-1348	900-908	909	909
C2	OH stretch (water shell)	1360-1366	916-920	915	916
C3	$v1+v3$	1370-1376	923-927	924	924
C4	OH stretch (water shell)	1380-1388	930- 935	933	932
C5	S_0 (free water)	1398-1418	945-955	954	954
C6	Water hydration, H_5O_2	1421-1428	957-962	960	960
C7	S_1	1436-1444	967-973	966	970
C8	$v2+v3$	1448-1454	975-979	975	980
C9	S_2	1462-1468	984-989	981	986
C10	S_3	1472-1482	991-998	990	993
C11	S_4	1482-1495	998-1007	999	1008
C12	Strongly bonded water or $v2$	1506-1516	1014-1021	1017	1021

8.3.2.7 Aquagrams

The derived WASP patterns presented in the aquagrams show differences between intact kiwifruit with varying DM levels (Fig. 8.9). As the DM level rises, the number of strongly H-bonded water molecular species (S2, S3, S4, ($v1$,$v2$)) increases along with minor increases in $v3$, water shell and $v1+v3$, with free water states (S0, H_5O_2, and S1) decreasing, indicating more highly organized water structures (Fig. 8.8(a)). This trend can be seen in the WASPs

derived from both the F-750 (Fig. 8.8(a)) and the benchtop spectrophotometer (Fig. 8.8(b)), only the S3 water species not increasing with DM.

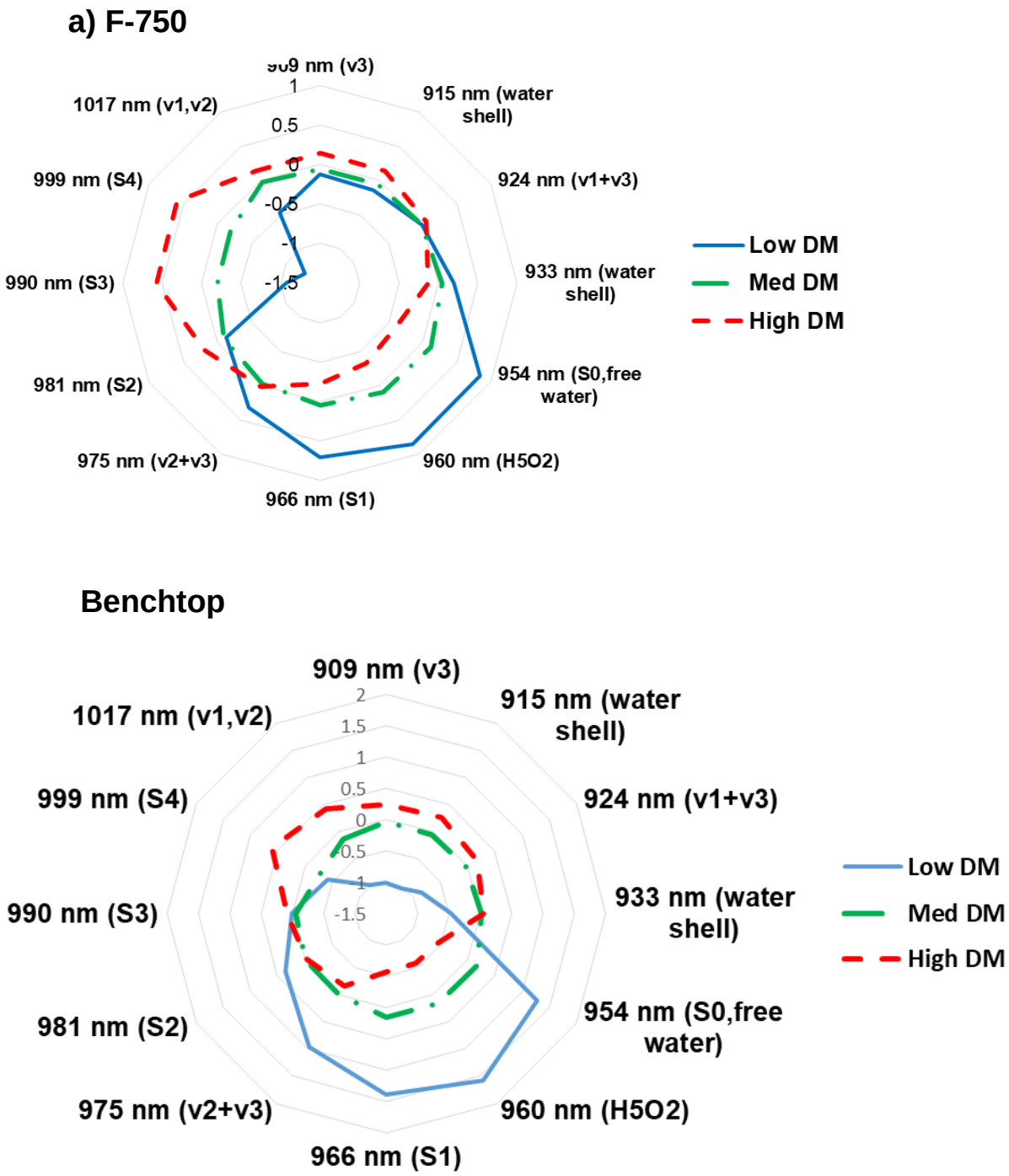


Figure 8.9. Aquagrams for intact kiwifruit at three DM levels at 22°C for a) F-750 produce quality meter; b) benchtop spectrophotometer.

8.3.3 Absorbance spectra of Lot II intact apples and water at three temperatures

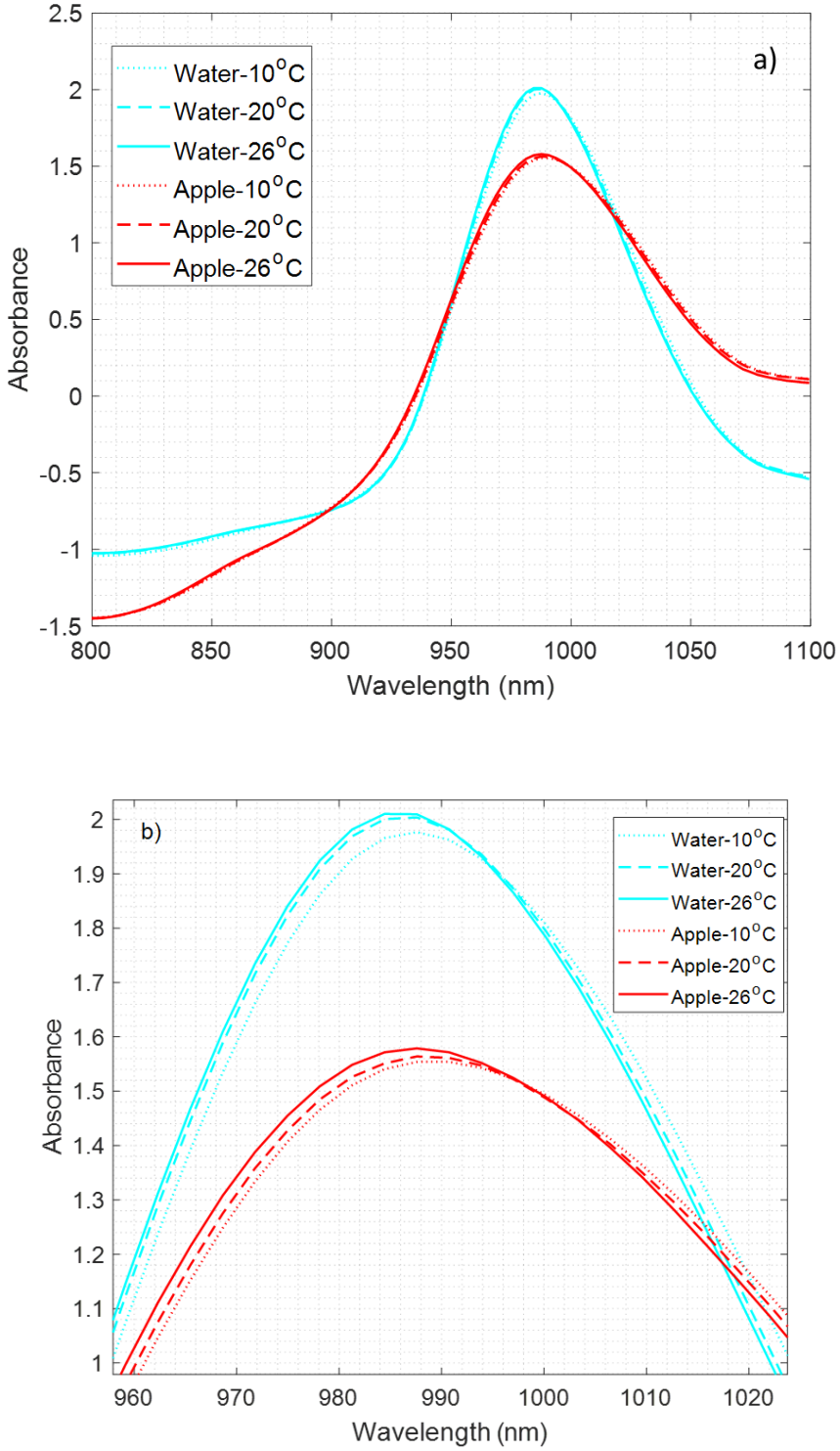


Figure 8.9. Absorbance spectra of intact apple and Milli-Q water at three temperatures; 10, 20, and 26°C in the second overtone region of the OH stretch of water a) Smoothed+SNV transformed; b) magnified.

The absorbance plots after smoothing (Savitzky–Golay with smoothing parameters: width 15, order 0) followed by SNV in Fig. 8.9(a) and 8.9(b) illustrate that as the temperature increased in intact apples and water in the second overtone region, the absorbance curve was shifted upward to shorter wavelengths (Fig. 8.9(b)) for both apples and water. The isosbestic points were 996 nm (for apple) and 992 nm (for water), respectively.

8.3.4 Pure Water analysis

There was a clear variation with temperature in the peak wavelength of the water (Fig. 8.9 (b)). When PCA was applied to the water spectra, the shape of the PC1 loading (Fig. 8.10) was very similar to that reported by Segtnan et al. (Segtnan, 2001) and Maeda et al. (Maeda et al., 1995), who studied the effect of temperature on the NIR spectra of water. In their work, the temperature of the water varied between 5°C and 80°C, and PCA was applied to the resulting spectra. In our case, the shape observed in PC 1 loading (Fig. 8.10) suggests a change in water structure due to a change in temperature. The PC1 loading was used in the EMSC correction (equation 2.3 in Chapter 2) as an interferent spectrum to correct apple spectra for temperature variation. The wavelengths 965 nm and 1025 nm represent the free and H-bonded water species that were similar to the PC1 loading of water in the second overtone region. The peaks in Fig. 8.10 were within 2 nm of those observed in the PC1 of water at different temperatures measured using an FT-NIR spectrophotometer in the 870-1100 nm region.

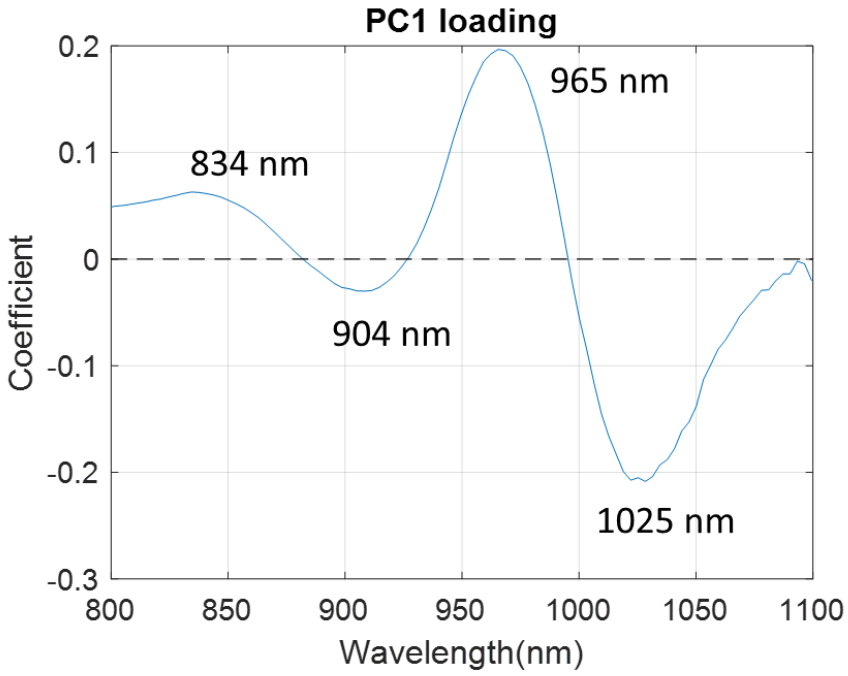


Figure 8.10. PC1 loading of water in the wavelength range from 800–1100 nm. Labels indicate the peak wavelengths.

8.3.5 Prediction of DM

8.3.5.1 PLSR on Lot I samples

All PLSR models developed for DM prediction yielded SEP values of 0.25 and 0.41%DM for the F-750 and benchtop spectrophotometer, respectively (Fig. 8.11). Using only the narrow second overtone range (800–1000 nm) was as good as using the wider spectrum (600–1100 nm). This suggests that most of the relevant information about DM concentration is available in the absorption region of the second overtone of water, a finding which agrees with the literature predicting DM (McGlone & Kawano, 1998; McGlone et al., 2003). The results from the F-750 handheld unit were better than those from the benchtop spectrophotometer in terms of the low RMSECV value of 0.30%DM (800-1000 nm) in Fig. 8.11(1).

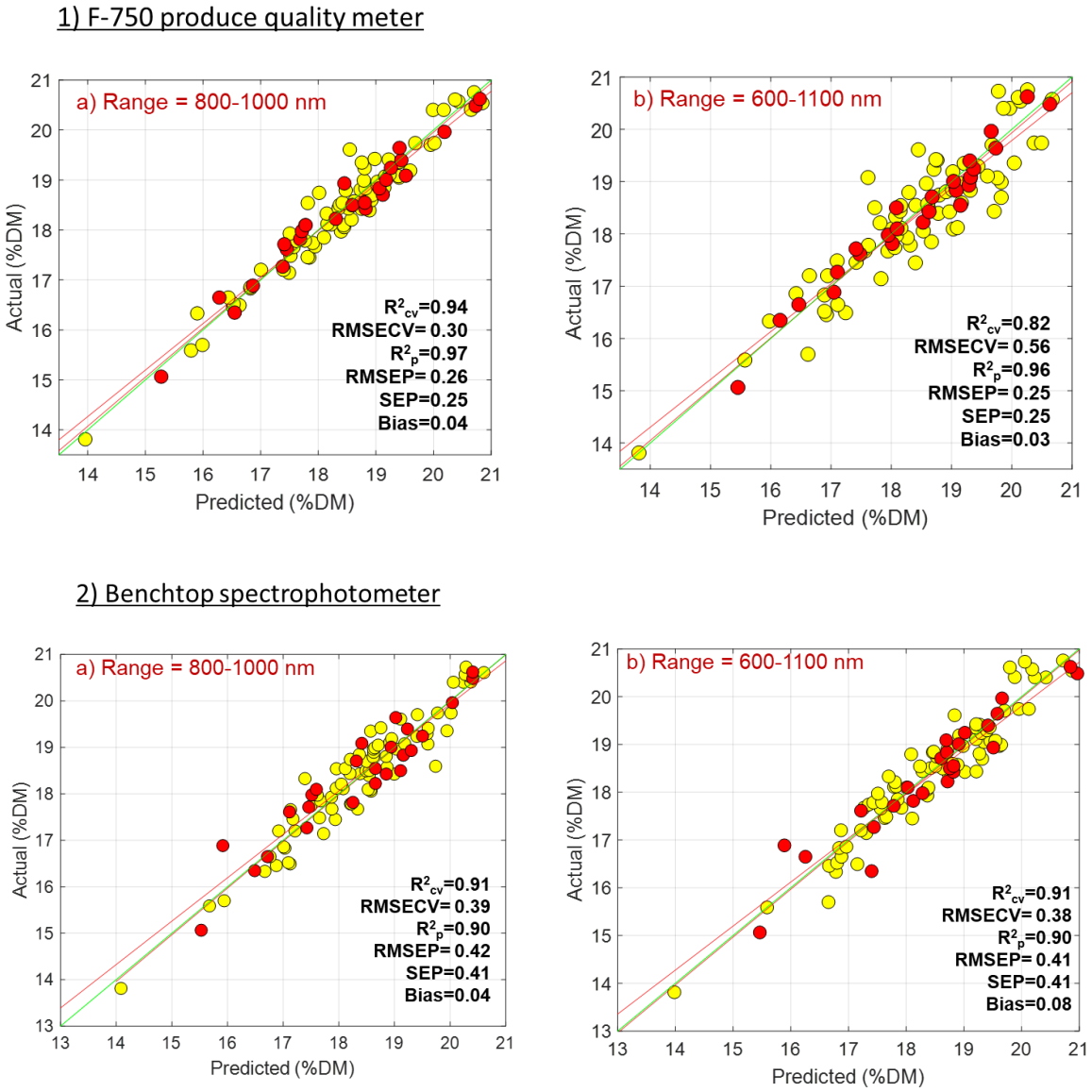


Figure 8.11. DM prediction of intact kiwifruit using 1) F-750 produce quality meter; and 2) Benchtop spectrophotometer in a) 800–1000 nm, and b) 600–1100 nm wavelength range. Yellow and red circles are for calibration and validation data, respectively. R^2_{cv}/R^2_p: the coefficient of determination for calibration/prediction; RMSECV/RMSEP: root mean square error of calibration/prediction; SEP: standard error of prediction (bias-corrected RMSEP).

8.3.5.1 MLR on Lot I samples

Multiple linear regression (MLR) was performed on two sets of wavelengths from the F-750 handheld unit absorbance data (Table 8.3). The first set consisted of discrete water wavelengths obtained after aquaphotomics analysis. The second set consisted of all the wavelengths between 800-1100 nm, with 3 nm pixel spacing. The wavelength search involved a systematic examination of every linear wavelength combination of up to six wavelengths using an MLR model for DM prediction. The spectral bandwidth of the data was reduced to 800–1100 nm as this region consisted of water and carbohydrates. No spectrum pre-processing was performed to keep the application of the data general, since any particular chosen method may not be appropriate to the irregular non-contiguous spectra likely to result from a discrete LED-based system.

Table 8.3. Comparison of various wavelength combinations for DM (%) prediction of intact kiwifruit at 22°C using MLR on the raw absorbance data from the F-750 produce quality meter

F-750 produce quality meter (N_{cal}=75 and N_{val}=25)							
Selection of wavelength	Wavelength used (nm)	R^2_{cv}	RMSECV	R^2_p	RMSEP	Bias	SEP
909 915 924 933 954 960 966 975 981 990 999 1017 (Aquaphotomics wavelengths)	966	0.01	1.3	0.03	1.24	0.06	1.24
	960 1017	0.52	0.90	0.74	0.67	0.15	0.65
	933 960 981	0.71	0.70	0.85	0.55	0.21	0.52
	954 960 981 999	0.73	0.66	0.86	0.51	0.16	0.49
	909 933 954 990 1017	0.84	0.51	0.89	0.47	0.21	0.42
	909 933 954 960 990 1017	0.84	0.50	0.91	0.44	0.24	0.37
800–1100 nm (branch-and-bound)	966	0.01	1.3	0.03	1.24	0.06	1.24
	960 1020	0.52	0.89	0.72	0.69	0.17	0.66
	945 960 978	0.73	0.67	0.86	0.52	0.19	0.49
	861 888 915 924	0.89	0.41	0.92	0.36	0.03	0.36
	855 873 906 936 954	0.94	0.32	0.96	0.30	0.08	0.29
	861 879 915 933 987 1041	0.95	0.28	0.96	0.28	0.05	0.27

The data set was divided into two independent sets: a calibration set that consisted of 75 samples and a validation set of 25 samples. MLR models were generated for DM, and the predictive ability of the models was confirmed on the validation set. Separate MLR models were generated with the calibration samples for subsets of one wavelength (n = 1) to six wavelengths (n = 6). The wavelength selections were made using a "branch-and-bound" exhaustive search algorithm as implemented in the "leaps" software package used with the R programming language (Shafie et al., 2015). The "leaps" software was considerably faster than the simple brute force algorithm that was initially used, which calculated and ranked all possible MLR models. Apart from the wavelength search algorithm "leaps" that was implemented in R version 3.5.3, all other data analysis and processing were performed using MATLAB software and the PLS Toolbox. To compare the performance of the wavelength searches, the analysis of the spectral data was performed on the model generated from the six best aquaphotomics wavelengths.

The optimum wavelength sets found in the various searches are listed in Table 8.3. The full-range PLS model delivered the best DM result with an SEP of 0.25% and an R^2_p of 0.97 (Fig. 8.11(1)). The steady decrease in prediction error with increasing wavelength selections is shown in Table 8.3. The six-wavelength MLR model performed similarly to the full range PLS model, which gave an SEP = 0.27% and R^2_p = 0.96. Therefore, the six-wavelength MLR model was considered the optimal model in the sense of balancing the need for accuracy against the desire for low numbers of

wavelengths. However, the best six wavelengths chosen from the 12 aquaphotomics wavelengths did not perform as well, with an SEP of 0.37% and R^2_p of 0.91. Therefore, Table 8.3 indicates that while the aquaphotomics wavelengths were useful for investigating the water species in a sample they may not be the best selections for a predictive model. Other factors may be at play, such as variable sample light scattering effects, that diminish the practical usefulness of individual WAMACS wavelengths for model building. The alternative exhaustive search techniques, like "branch and bound", were simple to use and productive at providing precise MLR based predictive models known to be equivalent in performance to PLSR models.

8.3.5.2 PLSR on Lot II apple samples using benchtop spectrophotometer

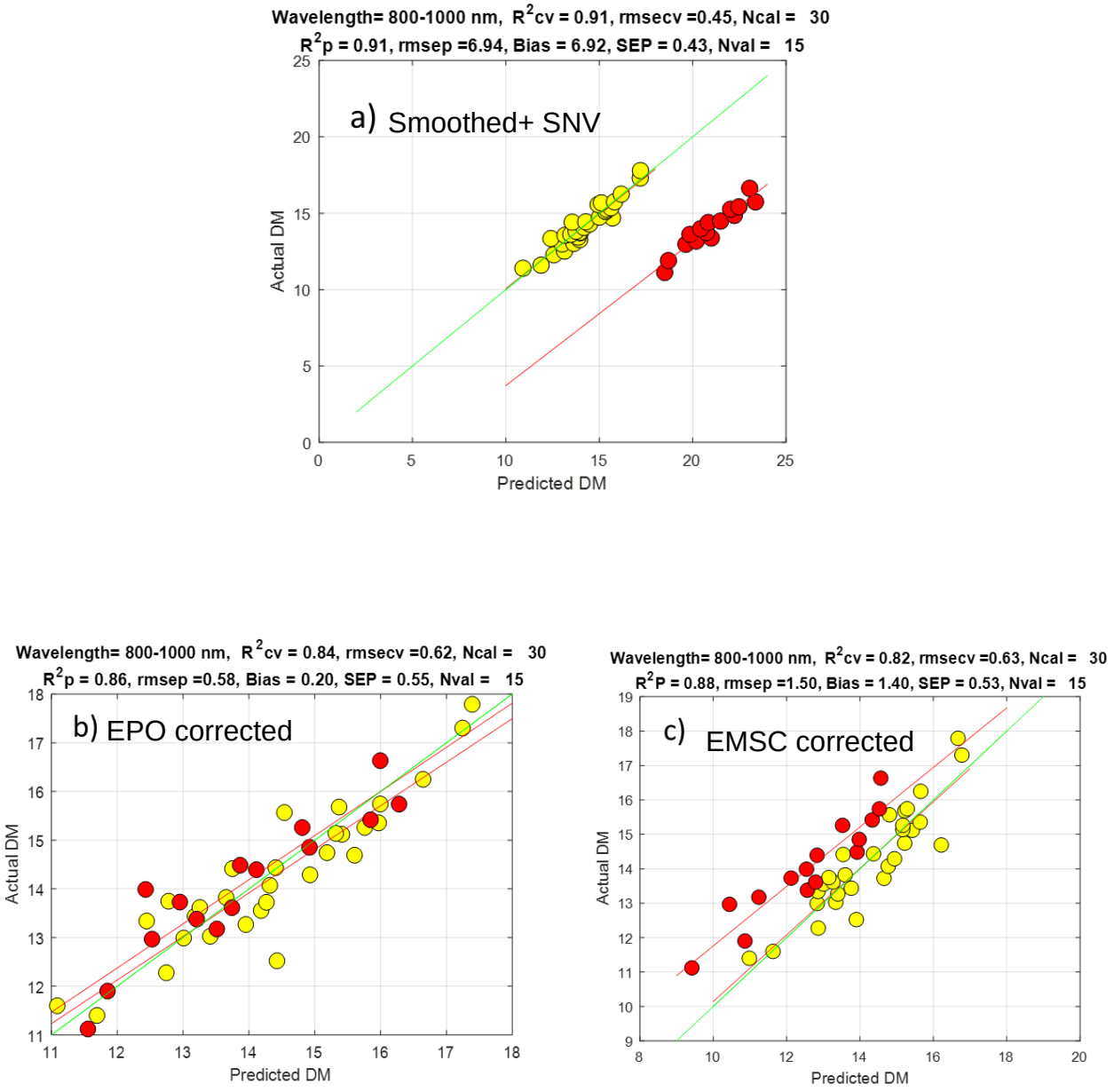


Figure 8.12. DM prediction for intact apples in the second overtone of the OH stretch of water on a) smoothed+ SNV corrected spectra; b) EPO corrected spectra; and c) EMSC corrected spectra. Yellow and red circles are for calibration and validation data, respectively.

Model performance improved following EPO and EMSC pre-processing (Fig. 8.12(b),(c)). The DM prediction bias, observed when applying the 26°C calibrated model to samples at 10°C, was reduced to a near-negligible level

of 0.2 for the EPO method (Fig. 8.12(b)). Compared to smoothed+SNV pre-processing only, the bias dropped from 6.92 to 0.2% and 1.4% for EPO and EMSC pre-processing techniques, respectively. However, the model precision (SEP) was increased from 0.43 to 0.55%DM compared with the smoothed+SNV method. The EMSC and EPO interference term, the PC1 generated from the pure water-temperature spectral matrix, was independent of the apple samples, which considerably simplified generating a temperature-independent model. This model does not require apple samples to be measured at different temperatures – measurements at one temperature will work. The EPO technique was found to be better than the EMSC technique. However while both of these techniques reduced the bias they did so with increased noise and, for EMSC, sloped offset. Obviously the problem is not entirely solved then, using EPO or EMSC techniques, and more research is warranted to understand the poorer performance, especially compared to earlier results on juice samples (Chapter 4 & 5), and so learn to overcome the limitations on whole fruit samples.

8.4 Conclusions

The aquaphotomics study undertaken here has delivered insights into the free and bonded water structure changes, as evident in the second overtone region for water absorbance, occurring with changes in DM in intact kiwifruit (13.8 to 20.7%DM) at a constant temperature of 22°C. The free water components decreased and the bonded water components increased

as DM decreased. The absorbance bands of water, which were activated in response to perturbation by DM, were clearly identified, and they all fell within the ranges of 12 WAMACS in the second overtone of the OH stretch of water. The PLSR model built with the F-750 instrument in the 800-1000 nm region (second overtone) gave a better result, with an RMSECV of 0.30% compared to 0.56% in the 600-1100 nm range, signifying that the important information resides in the tight window of the second overtone region. The MLR model built using the six best wavelengths with the branch and bound algorithm gave the same accuracy as that achieved with the PLS model. However, the MLR model built with the six best aquaphotomics water wavelengths was not as good as the one created using the six best wavelengths found using the branch and bound algorithm. This suggests that to identify the best wavelengths for developing a robust MLR model, aquaphotomics might not be useful. However, the best PLSR model was constructed in the full second overtone wavelength range of water.

Use of the EPO and EMSC pre-processing method with Lot II apples at different temperatures, with PLSR modelling, significantly reduced prediction bias when a model calibrated at one temperature (26°C) was applied to samples at different temperatures (10°C). This application is useful in creating robust models to deal with temperature variations but much more research is required to make it fully successful.

Chapter 9: Conclusions and future scope

The objective of this thesis was to investigate the feasibility of using aquaphotomics for fruit internal quality assessment. This study examined whether aquaphotomics can help in the development of better/simpler prediction models and/or hardware for intact fruit quality measurement that may be beneficial to the fruit industry. Much work in aquaphotomics has been done in the longer wavelength region, i.e., in the first overtone (1300-1600 nm), which requires short pathlength cells due to the high absorption of water. Therefore, experiments were performed on simple aqueous systems in the first overtone such as sucrose solution of various concentrations and an aqueous system such as apple and kiwifruit juice.

The experiments on aqueous samples were performed using an FT- NIR spectrophotometer. While kiwifruit and apples consist of more than 85% water, their spectral signatures will also be affected by varying sugar concentrations. The aquaphotomics approach was used to learn more about changes in water structure caused by sucrose perturbations in the 1300-1600 nm and 870-1100 nm wavelength regions. Both, the first overtone region of OH stretch of water (1300-1600 nm) and the SWNIR region were investigated as the latter is important for intact fruit quality measurement. Water bands were identified in two overtone regions that represented sucrose activity in water. The same bands were activated in the second overtone region compared with the first overtone region when using an anharmonic oscillator model. Sample path lengths played an important role when generating aquagrams. Those constructed in the second overtone

region from 10 mm path length measurements showed a clear water spectral pattern similar to that of the first overtone measurements using a 1 mm cell, and demonstrated the usefulness of NIRS and aquaphotomics in identifying water bands activated by the presence of sugar in water.

Aquaphotomics concepts were applied to apple juice samples with varying SSC (°Brix) concentrations at a constant temperature (25°C), using the first overtone of the OH stretch of water. The key focus was to observe the usefulness of the water spectral patterns (WASPs) presented on the aquagrams to observe water molecule structural changes. Then, the approach of aquaphotomics was used in the development of predictive models for the SSC of apple juice samples. Changes in the water structure of apple juice caused by variations in temperature were investigated using aquaphotomics. The work focused on spectra in the 1300–1600 nm region, surrounding the strong water absorption peak at 1450 nm. The aquaphotomics analysis indicated that EMSC pre-processing might be useful, using an interferent derived from pure water spectra. This may help to remove the temperature sensitivity of SSC of apple juice in NIR calibrations. The aquaphotomics study helped to get better understanding in terms of free and bonded water structures for apple juice samples whose SSC varied from 7.3 to 13.7°Brix at three temperatures 20, 25, and 30°C. The free water components increased, and the H-bonded water components decreased, as temperature rose or SSC decreased. The absorbance bands of water that were activated in response to perturbations by temperature and

SSC were identified, and they all fell within the tolerances of the 12 WAMACS of the first overtone of the OH stretch of water. Incorporation of the EMSC pre-processing method with PLSR modelling significantly reduced prediction bias when a model calibrated at one temperature (20°C) was applied to samples at different temperatures (25 and 30°C). The EMSC method used, as an interference spectrum, the PC1 loading vector generated from a spectral matrix for pure water measured over the same temperature range. The EMSC method with such a PC1 loading vector may well have applications in other fruit juice or intact fruit measurement modelling problems where robustness against temperature changes is desirable.

As the water peaks at 1450 nm and 970 nm in fruit juice shift with variations in temperature, aquaphotomics was also investigated as a way to learn more about changes in water structure caused by temperature in both these regions. The calibration models were constructed for SSC (°Brix) prediction in kiwifruit juice using an FT-NIR spectrometer and PLSR with EPO correction. The influence of increasing temperature on the peak absorbance of kiwifruit juice was a lateral shift in the first overtone region, in contrast to a vertical shift in the second overtone region of water. The water bands related to temperature variation in the first and second overtone regions for kiwifruit juice samples were identified. Different bands were seen to change in the second overtone region compared with the first overtone region. With the same data set, the use of EPO pre-

processing with independently measured water samples was investigated to help in building temperature-independent PLSR models to predict SSC in kiwifruit juice. The results showed that, using the EPO method, a significant reduction in the prediction bias was achieved by at least a factor of 10, when applying a model created at one temperature to measurements made at another.

Another study explored the potential of NIRS and aquaphotomics in DM prediction of kiwifruit pulp using an FT-NIR spectrophotometer. The water spectral pattern was identified for different consistencies of kiwifruit pulp samples in the first overtone region of water. Different trends in WASPs were seen in the first overtone region for 'Runny' and 'Gel-like' pulp samples. An increase in the water species related to solvation shell and bonded water structures (S1, S2, and S3) and H_5O_2, $v3$ and ($v1$+$v3$) were found in Gel-like samples. Whereas, the runny samples had S4, ($v1$, $v2$), S0, and water shells species. The foamy fruit pulp in the W1 batch covered all the water species. A similar trend was seen for Hayward fruit with high and low starch concentrations. As the starch rose and SSC decreased, the free water species increased. The strongly H-bonded water molecular species (S1, S2, and S3) increased with SSC concentration.

Aquaphotomics was used to examine the differences between control and chilling injury (CI) kiwifruit by observing the changes in the water spectral patterns presented on aquagrams. The water bands were identified for CI kiwifruit juice in the first overtone region and for intact

fruit in the second overtone region of the OH stretch of water. It was found that for both juice and intact fruit measurements, CI increased H-bonded water structures despite the fruit having the same SSC concentrations.

In the last chapter, the use of aquaphotomics for intact fruit quality measurement was investigated. The technique delivered insight into the free and bound water structure changes, as evident in the second overtone region for water absorbance, occurring with changes in DM of gold kiwifruit (13.8 to 20.7%DM) at a constant temperature of 22°C. The free water components decreased and the bonded water components increased as DM increased. The PLSR model built using the F-750 instrument in the 800-1000 nm region (second overtone) gave better results, with an RMSECV of 0.30% compared to 0.56% in the 600-1100 nm range, signifying that the important information resides in the tight window of the second overtone region. The MLR model built using the six best wavelengths with a branch and bound algorithm gave the same accuracy as that achieved with the PLS model. However, the MLR model built with the six best aquaphotomics water wavelengths was not as good as that built with the six best wavelengths found using the branch and bound algorithm. This suggests that to identify the best wavelengths for developing a robust MLR model, aquaphotomics may not be useful. However, the best PLSR model was built in the full second overtone wavelength range of water. The application of EPO and EMSC pre-processing to intact apples at different temperatures, with PLSR modelling, reduced prediction bias when a model calibrated at

one temperature was applied to samples at different temperatures. This application is useful in creating robust models to deal with temperature variations but much more research is required to make it fully successful as seen with juice samples.

In future work, the aquaphotomics approach could be applied to monitor fruit maturity/ripeness/variety through changes in water spectral patterns. Development of hardware in the narrow wavelength region of 800-1100 nm may be helpful in creating a portable spectrometer since the majority of the information about the quality attributes of fruit resides in the region of the water bands. In conjunction with the EPO and EMSC techniques, using an interferent component derived independently from water spectra may be beneficial for generating robust calibration models.

References

Acharya, U. K., Walsh, K. B., & Subedi, P. (2013). *Effect of temperature on SWNIRS based models of fruit DM and colour.* Paper presented at the NIR 2013 - 16th International Conference on Near Infrared Spectroscopy, La Grande-Motte, France.

Acharya, U. K., Walsh, K. B., & Subedi, P. (2014). Robustness of partial least-squares models to change in sample temperature. II, Application to fruit attributes.

Aprea, E., Charles, M., Endrizzi, I., Laura Corollaro, M., Betta, E., Biasioli, F., & Gasperi, F. (2017). Sweet taste in apple: the role of sorbitol, individual sugars, organic acids and volatile compounds. *Scientific Reports, 7*, 44950.

Barnes, R., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied spectroscopy, 43*(5), 772-777.

Bázár, G., Kovacs, Z., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y., Sugiyama, H., & Tsenkova, R. (2015). Water revealed as molecular mirror when measuring low concentrations of sugar with near infrared light. *Analytica Chimica Acta, 896*, 52-62.

Bázár, G., Romvári, R., Szabó, A., Somogyi, T., Éles, V., & Tsenkova, R. (2016). NIR detection of honey adulteration reveals differences in water spectral pattern. *Food Chemistry, 194*, 873-880.

Bellon-Maurel, V. (1992). *Application de la spectroscopie proche infrarouge au contrôle en ligne de la qualité des fruits et legumes.* Toulouse, INPT.

Boeing, H., Bechthold, A., Bub, A., Ellinger, S., Haller, D., Kroke, A., Leschik-Bonnet, E., Müller, M. J., Oberritter, H., Schulze, M., Stehle, P., & Watzl, B. (2012). Critical review: vegetables and fruit in the prevention of chronic diseases. *European journal of nutrition, 51*(6), 637-663.

Burns, D. A., & Ciurczak, E. W. (1992). *Handbook of near-infrared analysis.* . New York, NY: Marcel Dekker.

Cen, H., Lu, R., Zhu, Q., & Mendoza, F. (2016). Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification. *Postharvest Biology and Technology, 111*, 352-361.

Chatani, E., Tsuchisaka, Y., Masuda, Y., & Tsenkova, R. (2014). Water molecular system dynamics associated with amyloidogenic nucleation as revealed by real time near infrared spectroscopy and aquaphotomics. *PloS one, 9*(7), e101997-e101997.

Chen, X., & Han, W. (2012). Spectroscopic determination of soluble solids content of 'Qinmei' kiwifruit using partial least squares. *African Journal of Biotechnology, 11*(10), 2528-2536.

Clark, C., McGlone, V., & Jordan, R. (2003a). Detection of Brownheart in 'Braeburn'apple by transmission NIR spectroscopy. *Postharvest Biology and Technology, 28*(1), 87-96.

Clark, C. J., McGlone, V. A., Requejo, C., White, A., & Woolf, A. B. (2003b). Dry matter determination in 'Hass' avocado by NIR spectroscopy. *Postharvest Biology and Technology, 29*(3), 301-308.

Crisosto, G., Hasey, J. K., Zegbe, J. A., & Crisosto, C. H. (2012). New quality index based on dry matter and acidity proposed for Hayward kiwifruit. *California Agriculture, 66*(2), 70-75.

Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables. *Food and Bioprocess Technology, 4*(4), 487-504.

Cui, X., Liu, X., Yu, X., Cai, W., & Shao, X. (2017). Water can be a probe for sensing glucose in aqueous solutions by temperature dependent near infrared spectra. *Analytica Chimica Acta, 957*, 47-54.

Davies, A., & Grant, A. (1987). Near infra - red analysis of food. *International Journal of Food Science & Technology, 22*(3), 191-207.

Davies, T. (2011). Happy 90th Birthday to Karl Norris, Father of NIR Technology. *NIR news, 22*(4), 3-16.

DC Slaughter, C. C. (1998). Nondestructive internal quality assessment of kiwifruit using near-infrared spectroscopy *Seminars in Food Analysis, 3*, 13 11-140

El - Azazy, M. (2018). Introductory Chapter: Infrared Spectroscopy - A Synopsis of the Fundamentals and Applications.

F-750 Instruction Manual. (2019). Retrieved from https://felixinstruments.com/file-collection/f-750-manuals/

Feng, J., McGlone, A. V., Currie, M., Clark, C. J., & Jordan, B. R. (2011). Assessment of yellow-fleshed kiwifruit (actinidia chinensis

'hort16a') quality in pre-and post-harvest conditions using a portable near-infrared spectrometer. *HortScience, 46*(1), 57-63.

Fillion, L., & Kilcast, D. (2002). Consumer perception of crispness and crunchiness in fruits and vegetables. *Food Quality and Preference, 13*(1), 23-29.

Fraser, D. G., McGlone, V., Jordan, R., & Künnemeyer, R. (2001). Near Infrared (NIR) light penetration into an Apple. *PostHarvest Biology and Technology 22*, 191-195.

Fresh Facts (2018). Mt Albert, Auckland: The New Zealand Institute for Plant & Food Research.

Fu, X.-p., Li, J.-p., Zhou, Y., Ying, Y.-b., Xie, L.-j., Niu, X.-y., Yan, Z.-k., & Yu, H.-y. (2009). Determination of soluble solid content and acidity of loquats based on FT-NIR spectroscopy. *Journal of Zhejiang University. Science. B, 10*(2), 120-125.

Giangiacomo, R. (2006). Study of water–sugar interactions at increasing sugar concentration by NIR spectroscopy. *Food Chemistry, 96*(3), 371-379.

Golic, M., & Walsh, K. B. (2006). Robustness of calibration models based on near infrared spectroscopy for the in-line grading of stonefruit for total soluble solids content. *Analytica chimica acta, 555*(2), 286-291.

Golic, M. W., Kerry; Lawson, Peter. (2003). Short-Wavelength Near-Infrared Spectra of Sucrose, Glucose, and Fructose with Respect to Sugar Concentration and Temperature. *Applied Spectroscopy, 57*(2), 139-145.

Gómez, A. H., He, Y., & Pereira, A. G. (2006). Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *Journal of Food Engineering, 77*(2), 313-319.

Gowen, A., Stark, E., Tsuchisaka, T., & Tsenkova, R. (2011a). Extended multiplicative signal correction as a tool for aquaphotomics. *NIR news, 22*(7), 9-13.

Gowen, A., Tsuchisaka, Y., O'Donnell, C., & Tsenkova, R. (2011b). Investigation of the Potential of Near Infrared Spectroscopy for the Detection and Quantification of Pesticides in Aqueous Solution. *American Journal of Analytical Chemistry, 2*(8).

Gowen, A. A., Amigo, J. M., & Tsenkova, R. (2013). Characterisation of hydrogen bond perturbations in aqueous systems using aquaphotomics and multivariate curve resolution-alternating least squares. *Analytica Chimica Acta, 759*, 8-20.

Gowen, A. A., Marini, F., Tsuchisaka, Y., De Luca, S., Bevilacqua, M., O'Donnell, C., Downey, G., & Tsenkova, R. (2015). On the feasibility of near infrared spectroscopy to detect contaminants in water using single salt solutions as model systems. *Talanta, 131*, 609-618.

Greensill, C. V., & Walsh, K. B. (2000). A remote acceptance probe and illumination configuration for spectral assessment of internal attributes of intact fruit. *Measurement Science and Technology, 11*(12), 1674.

Gunasekaran, S., & Irudayaraj, J. (2000). Optical methods—Visible, NIR and FTIR spectroscopy *Nondestructive Food Evaluation—Techniques to Analyze Properties and Quality* (pp. 1-38): CRC Press CA.

Gwanpua, S. G., Jabbar, A., Zhao, M., Heyes, J. A., & East, A. R. (2018). Investigating the potential of dual temperature storage as a postharvest management practice to mitigate chilling injury in kiwifruit. *International Journal of Refrigeration, 86*, 62-72.

Hall, J. N., Moore, S., Harper, S. B., & Lynch, J. W. (2009). Global Variability in Fruit and Vegetable Consumption. *American Journal of Preventive Medicine, 36*(5), 402-409.

Harris, P., & Altaner, C. (2013). *Workshop on commercial application of IR spectroscopies to solid wood*.

Harvey, D. (2000). *Modern analytical chemistry*: McGraw-Hill.

Ilari, J. L., Martens, H., & Isaksson, T. (1988). Determination of Particle Size in Powders by Scatter Correction in Diffuse Near-Infrared Reflectance. *Applied Spectroscopy, 42*(5), 722-728.

Jantra, C., Slaughter, D. C., Liang, P.-S., & Pathaveerat, S. (2017). Nondestructive determination of dry matter and soluble solids content in dehydrator onions and garlic using a handheld visible and near infrared instrument. *Postharvest Biology and Technology, 133*, 98-103.

Jensen, P. S., & Bak, J. (2002). Near-Infrared Transmission Spectroscopy of Aqueous Solutions: Influence of Optical Pathlength on Signal-To-Noise Ratio. *Applied Spectroscopy, 56*(12), 1600-1606.

Jensen, P. S., Bak, J., & Andersson-Engels, S. (2003). Influence of Temperature on Water and Aqueous Glucose Absorption Spectra in the Near- and Mid-Infrared Regions at Physiologically Relevant Temperatures. *Applied Spectroscopy, 57*(1), 28-36.

Jinendra, B., Tamaki, K., Kuroki, S., Vassileva, M., Yoshida, S., & Tsenkova, R. (2010). Near infrared spectroscopy and aquaphotomics: Novel

approach for rapid in vivo diagnosis of virus infected soybean. *Biochemical and Biophysical Research Communications, 397*(4), 685-690.

Jordan, R., Osborne, S., Kunnemeyer, R., & Seelye, R. (1997). Harvest time prediction of eating time properties of kiwifruit using NIR transmission. *Proceedings of Sensors for Nondestructive Testing: Measuring the Quality of Fresh Fruits and Vegetables NRAES-97, Orlando, FL. Published by Natural Resource, Agriculture, and Engineering Service, Ithaca, NY*.

Jordan, R. B., & Seelye, R. J. (2009). Relationship between taste perception, density and soluble solids concentration in kiwifruit (Actinidia deliciosa). *New Zealand Journal of Crop and Horticultural Science, 37*(4), 303-317.

Kaur, H., Künnemeyer, R., & McGlone, A. (2017a). *Assessment of Apple Juice by Aquaphotomics in a Temperature Dependent Environment.* Paper presented at the The 3rd Australia New Zealand Conference on Optics and Photonics (ANZCOP), Queenstown, New Zealand.

Kaur, H., Künnemeyer, R., & McGlone, A. (2017b). Comparison of hand-held near infrared spectrophotometers for fruit dry matter assessment. *Journal of Near Infrared Spectroscopy, 25*(4), 267-277.

Kaur, H., Künnemeyer, R., & McGlone, A. (2018). *Investigating Aquaphotomics for Fruit Quality Assessment.* Paper presented at the 3rd Aquaphotomics International Symposium Exploring Water Molecular Systems in Nature, Awaji, Hyogo, Japan.

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*(4), 211-218.

Kawano, S., Watanabe, H., & Iwamoto, M. (1992). Determination of Sugar Content in Intact Peaches by Near Infrared Spectroscopy with Fiber Optics in Interactance Mode. *Journal of the Japanese Society for Horticultural Science, 61*(2), 445-451.

Kennard, R. W., & Stone, L. A. (1969). Computer Aided Design of Experiments. *Technometrics, 11*(1), 137-148.

Kinoshita, K., Miyazaki, M., Morita, H., Vassileva, M., Tang, C., Li, D., Ishikawa, O., Kusunoki, H., & Tsenkova, R. (2012). Spectral pattern of urinary water as a biomarker of estrus in the giant panda. [Article]. *Scientific Reports, 2*, 856.

Kovacs, Z., Bázár, G., Oshima, M., Shigeoka, S., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y., & Tsenkova, R. (2016). Water

spectral pattern as holistic marker for water quality monitoring. *Talanta, 147*, 598-608.

Kumar, S., McGlone, A., Whitworth, C., & Volz, R. (2015). Postharvest performance of apple phenotypes predicted by near-infrared (NIR) spectral analysis. *Postharvest Biology and Technology, 100*, 16-22.

Lallu, N. (1997). Low Temperature Breakdown in Kiwifruit *III International Symposium on Kiwifruit* (444 ed., pp. 579-586): Acta Horticulture.

Lammertyn, J., Nicolaï, B., Ooms, K., De Smedt, V., & De Baerdemaeker, J. (1998). Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Transactions of the ASAE, 41*(4), 1089.

Lammertyn, J., Peirs, A., De Baerdemaeker, J., & Nicolaı, B. (2000a). Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biology and Technology, 18*(2), 121-132.

Lammertyn, J., Peirs, A., De Baerdemaeker, J., & Nicolaï, B. (2000b). Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biology and Technology, 18*(2), 121-132.

León, L., Kelly, J. D., & Downey, G. (2005). Detection of Apple Juice Adulteration Using Near-Infrared Transflectance Spectroscopy. *Applied Spectroscopy, 59*(5), 593-599.

Lin, H., & Ying, Y. (2009). Theory and application of near infrared spectroscopy in assessment of fruit quality: a review. *Sensing and Instrumentation for Food Quality and Safety, 3*(2), 130-141.

Liu, Y., Sun, X., & Ouyang, A. (2010). Nondestructive measurement of soluble solid content of navel orange fruit by visible–NIR spectrometric technique with PLSR and PCA-BPNN. *LWT-Food Science and Technology, 43*(4), 602-607.

Lohumi, S., Lee, S., Lee, W.-H., Kim, M. S., Mo, C., Bae, H., & Cho, B.-K. (2014). Detection of Starch Adulteration in Onion Powder by FT-NIR and FT-IR Spectroscopy. *Journal of Agricultural and Food Chemistry, 62*(38), 9246-9251.

Lu, H.-s., Xu, H.-r., Ying, Y.-b., Fu, X.-p., Yu, H.-y., & Tian, H.-q. (2006). Application Fourier transform near infrared spectrometer in rapid estimation of soluble solids content of intact citrus fruits. *Journal of Zhejiang University. Science. B, 7*(10), 794-799.

Maeda, H., Ozaki, Y., Tanaka, M., Hayashi, N., & Kojima, T. (1995). Near Infrared Spectroscopy and Chemometrics Studies of Temperature-Dependent Spectral Variations of Water: Relationship between Spectral Changes and Hydrogen Bonds. *Journal of Near Infrared Spectroscopy, 3*(4), 191-201.

Magwaza, L. S., & Opara, U. L. (2015). Analytical methods for determination of sugars and sweetness of horticultural products—A review. *Scientia Horticulturae, 184*, 179-192.

Martens, H., Bruun, S. W., Adt, I., Sockalingum, G. D., & Kohler, A. (2006). Pre-processing in biochemometrics: correction for path-length and temperature effects of water in FTIR bio-spectroscopy by EMSC. *Journal of Chemometrics, 20*(8-10), 402-417.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. *Analytical Chemistry, 75*(3), 394-404.

Martens, H., & Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis, 9*(8), 625-635.

Matija, L., & Tsenkova, R. (2011). *Aquaphotomics of hydrogenated fullerenes.* Paper presented at the The Second Scientific International Conference on Water and Nanomedicine, Banja Luka.

Matija, L. R., Tsenkova, R. N., Miyazaki, M., Bamba, K., & Muncan, J. S. (2012). Aquagrams: Water spectral pattern as characterization of hydrogenated nanomaterial. *FME Transactions, 40*, 51-56.

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*(2), 135-144.

McGlone, V. A., Jordan, R. B., & Martinsen, P. J. (2002b). Vis/NIR estimation at harvest of pre-and post-storage quality indices for 'Royal Gala'apple. *Postharvest Biology and Technology, 25*(2), 135-144.

McGlone, V. A., Jordan, R. B., Seelye, R., & Clark, C. J. (2003). Dry-matter—a better predictor of the post-storage soluble solids in apples? *Postharvest Biology and Technology, 28*(3), 431-435.

McGlone, V. A., Jordan, R. B., Seelye, R., & Martinsen, P. J. (2002c). Comparing density and NIR methods for measurement of Kiwifruit dry matter and soluble solids content. *Postharvest Biology and Technology, 26*(2), 191-198.

McGlone, V. A., & Kawano, S. (1998). Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology, 13*(2), 131-141.

Meilina, H., Kuroki, S., Jinendra, B. M., Ikuta, K., & Tsenkova, R. (2009). Double threshold method for mastitis diagnosis based on NIR spectra of raw milk and chemometrics. *Biosystems Engineering, 104*(2), 243-249.

Minasny, B., McBratney, A. B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., & Joalland, S. (2011). Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma, 167-168*, 118-124.

Moghimi, A., Aghkhani, M. H., Sazgarnia, A., & Sarmad, M. (2010). Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosystems Engineering, 106*(3), 295-302.

Muncan, J., & Tsenkova, R. (2019). Aquaphotomics—From Innovative Knowledge to Integrative Platform in Science and Technology. *Molecules, 24*(15), E2742.

Munćan, J. S., Matija, L., Simić-Krstić, J. B., Nijemčević, S. S., & Koruga, D. L. (2014). Discrimination of mineral waters using near-infrared spectroscopy and aquaphotomics. *Hem. Ind, 68*(2), 257-264.

Næs, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user friendly guide to multivariate calibration and classification*: NIR publications.

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*(2), 99-118.

Norris, K. H. (1964). Design and development of a new moisture meter. *Agric. Eng, 45*(7), 370-372.

Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis* Harlow, UK: Longman Scientific & Technical

Osborne, S. D. (1997). *Non-Destructive Measurement of Internal Fruit Quality.* Unpublished PhD thesis, The University of Waikato.

Peirs, A., Scheerlinck, N., & Nicolaï, B. M. (2003). Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. *Postharvest Biology and Technology, 30*(3), 233-248.

Pissard, A., Fernández Pierna, J. A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., Dupont, P., Rondia, A., & Lateur, M. (2013). Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture, 93*(2), 238-244.

Ragni, L., Cevoli, C., Berardinelli, A., & Silaghi, F. A. (2012). Non-destructive internal quality assessment of "Hayward" kiwifruit by waveguide spectroscopy. *Journal of Food Engineering, 109*(1), 32-37.

Renfroe, W., & Kays, S. (1985). Nondestructive spectrophotometric determination of dry matter in onions. *J. Amer. Soc. Ilort.. Sci, 110*(2), 297-303.

Roger, J. M., Chauchard, F., & Bellon-Maurel, V. (2003). EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometrics and Intelligent Laboratory Systems, 66*(2), 191-204.

Roy, S., Anantheswaran, R. C., Shenk, J. S., Westerhaus, M. O., & Beelman, R. B. (1993). Determination of moisture content of mushrooms by Vis—NIR spectroscopy. *Journal of the Science of Food and Agriculture, 63*(3), 355-360.

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*(2), 175-184.

Schmilovitch, Z. e., Hoffman, A., Egozi, H., Ben-Zvi, R., Bernstein, Z., & Alchanatis, V. (1999). Maturity determination of fresh dates by near infrared spectrometry. *Journal of the Science of Food and Agriculture, 79*(1), 86-90.

Segtnan, V. H. (2001). Studies on the Structure of Water Using Two-Dimensional Near-Infrared Correlation Spectroscopy and Principal Component Analysis. *Analytical Chemistry, 73*(13), 3153-3161.

Segtnan, V. H., Šašić, Š., Isaksson, T., & Ozaki, Y. (2001). Studies on the Structure of Water Using Two-Dimensional Near-Infrared Correlation Spectroscopy and Principal Component Analysis. *Analytical Chemistry, 73*(13), 3153-3161.

Shafie, K. A., Künnemeyer, R., McGlone, A., Talele, S., & Vetrova, V. (2015). An Optimised Six-Wavelength Model for Predicting Kiwifruit Dry Matter. *Journal of Near Infrared Spectroscopy, 23*(2), 103-109.

Shao, Y., He, Y., Bao, Y., & Mao, J. (2009). Near-infrared spectroscopy for classification of oranges and prediction of the sugar content. *International Journal of Food Properties, 12*(3), 644-658.

Siesler, H., Ozaki, Y., Kawata, S., & Heise, H. (2001). Near - Infrared Spectroscopy: Principles, Instruments, Applications *Near - Infrared Spectroscopy*. Weinheim, Germany: Wiley-VCH.

Sirinnapa, S., & Sumio, K. (2008). Interpretation of near Infrared Calibration Structure for Determining the Total Aerobic Bacteria Count in Raw Milk: Interaction between Bacterial Metabolites and Water Absorptions. *Journal of Near Infrared Spectroscopy, 16*(6), 497-504.

Slaughter, D., Barrett, D., & Boersig, M. (1996). Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. *Journal of food science, 61*(4), 695-697.

Slaughter, D. C., & Crisosto, C. H. (1998). *Nondestructive internal quality assessment of kiwifruit using near-infrared spectroscopy.* Paper presented at the Seminars in food analysis.

Slavchev, A., Kovacs, Z., Koshiba, H., Nagai, A., Bázár, G., Krastanov, A., Kubota, Y., & Tsenkova, R. (2015). Monitoring of water spectral pattern reveals differences in probiotics growth when used for rapid bacteria selection. *PloS one, 10*(7), e0130698.

Sritham, E., & Gunasekaran, S. (2017). FTIR spectroscopic evaluation of sucrose-maltodextrin-sodium citrate bioglass. *Food hydrocolloids, 70*, 371-382.

Subedi, P. P., Walsh, K. B., & Owens, G. (2007). Prediction of mango eating quality at harvest using short-wave near infrared spectrometry. *Postharvest Biology and Technology, 43*(3), 326-334.

Takemura, G., Bázár, G., Ikuta, K., Yamaguchi, E., Ishikawa, S., Furukawa, A., Kubota, Y., Kovács, Z., & Tsenkova, R. (2015). Aquagrams of Raw Milk for Oestrus Detection in Dairy Cows. *Reproduction in Domestic Animals, 50*(3), 522-525.

Temma, T., Hanamatsu, K., & Shinoki, F. (2002). Measuring the Sugar Content of Apples and Apple Juice by Near Infrared Spectroscopy. *Optical Review, 9*(2), 40-44.

Tsenkova, R. (2002). Near Infrared Spectroscopy for Nondestructive Study of Prion Protein Isoforms. In D. R. Brown (Ed.), *Prion Diseases and Copper Metabolism: BSE, scrapie and CJD research* (pp. 79-95). Sawston, UK: Woodhead Publishing.

Tsenkova, R. (2006). AquaPhotomics: water absorbance pattern as a biological marker. *NIR news, 17*(7), 13-10.

Tsenkova, R. (2007). AquaPhotomics: water absorbance pattern as a biological marker for disease diagnosis and disease understanding. *NIR news, 18*(2), 14-16.

Tsenkova, R. (2008a). Aquaphotomics: acquiring spectra of various biological fluids of the same organism reveals the importance of water matrix absorbance coordinates and the aquaphotome for understanding biological phenomena. *NIR news, 19*(1), 13-15.

Tsenkova, R. (2008b). Aquaphotomics: the extended water mirror effect explains why small concentrations of protein in solution can be measured with near infrared light. *NIR news, 19*(4), 13-14.

Tsenkova, R. (2009). Introduction: Aquaphotomics: dynamic spectroscopy of aqueous and biological systems describes peculiarities of water. *Journal of Near Infrared Spectroscopy, 17*(6), 303-313.

Tsenkova, R. (2010). Aquaphotomics: Water in the biological and aqueous world scrutinised with invisible light. *Spectroscopy Europe, 22,* 6-10.

Tsenkova, R., Kovacs, Z., & Kubota, Y. (2015). Aquaphotomics: Near Infrared Spectroscopy and Water States in Biological Systems. In E. A. Disalvo (Ed.), *Membrane Hydration: The Role of Water in the Structure and Function of Biological Membranes* (pp. 189-211). Cham: Springer International Publishing.

Tsenkova, R., Munćan, J., Pollner, B., & Kovacs, Z. (2018). Essentials of Aquaphotomics and Its Chemometrics Approaches. *Frontiers in Chemistry, 6*, 363.

Valero, C., Crisosto, C. H., & Slaughter, D. (2007). Relationship between nondestructive firmness measurements and commercially important ripening fruit stages for peaches, nectarines and plums. *Postharvest Biology and Technology, 44*(3), 248-253.

Vasquez, S., & Mueller, S. (2019). *Refractometer Calibration, Use and Maintenance.* University of California Cooperative Extension Fresno County. Retrieved from http://cemerced.ucanr.edu/files/40574.pdf

Walsh, K. B., Golic, M., & Greensill, C. V. (2004). Sorting of Fruit Using near Infrared Spectroscopy: Application to a Range of Fruit and Vegetables for Soluble Solids and Dry Matter Content. *Journal of Near Infrared Spectroscopy, 12*(3), 141-148.

Wang, H., Peng, J., Xie, C., Bao, Y., & He, Y. (2015). Fruit quality evaluation using spectroscopy technology: a review. *Sensors (Basel, Switzerland), 15*(5), 11889-11927.

Wang, Y., Veltkamp, D. J., & Kowalski, B. R. (1991). Multivariate instrument standardization. *Analytical Chemistry, 63*(23), 2750-2756.

Wang, Z., Kunnemeyer, R., & McGlone, A. (2019). *Near-Infrared Spectroscopy for Kiwifruit Water-soaked Tissue Detection*. Paper presented at the Dodd-Walls Centre 12th Annual Symposium 2019, Dunedin, Otago.

Workman, J. (2000). *The Handbook of Organic Compounds, Three-Volume Set: NIR, IR, R, and UV-Vis Spectra Featuring Polymers and Surfactants*: Elsevier Science.

Workman, J. J., & Weyer, L. (2007). *Practical Guide to Interpretive Near-Infrared Spectroscopy*. Boca Raton, FL: CRC Press.

Zhu, D., Ji, B., Qing, Z., Wang, C., & Zude, M. (2011). The Detection of Quality Deterioration of Apple Juice by Near Infrared and Fluorescence Spectroscopy. In D. Li, Y. Liu & Y. Chen (Eds.), *Computer and Computing Technologies in Agriculture* (pp. 84-91). Berlin, Heidelberg: Springer Berlin Heidelberg.

Appendix: Comparison of hand-held near infrared spectrophotometers for fruit dry matter assessment

---

A journal paper

by

Harpreet Kaur, Rainer Künnemeyer, Andrew McGlone

Published in the

Journal of Near Infrared Spectroscopy

Original article

JNIRS

Journal of Near Infrared Spectroscopy
2017, Vol. 25(4) 267–277

DOI: 10.1177/0967033517725530
journals.sagepub.com/home/jns
SAGE

Comparison of hand-held near infrared spectrophotometers for fruit dry matter assessment

Harpreet Kaur[1,2,3], Rainer Künnemeyer[1,3] and Andrew McGlone[2]

Abstract

Comparisons are reported for developing predictive models for dry matter across a wide variety of fruits with near infrared spectroscopy instrumentation, using a number of commercially available hand-held portable instruments (NIRVANA by Integrated Spectronics, F-750 by Felix Instruments, H-100C by Sunforest and SCiO by Consumer Physics) and an in-house laboratory based instrument (Benchtop). Three intrinsic (same fruit type) and combined (all fruit types) data sets were created from two separate batches of fruit populations. The first batch (Lot I) consisted of 205 ripe fruits from three different main fruit types (apples, kiwifruit and summerfruit) and 12 distinct fruit sub-categories. The second batch (Lot II) consisted of 91 ripe fruits from two different fruit types (apples and kiwifruit) and seven distinct fruit sub-categories. The laboratory based Benchtop instrument performed the best overall with typically higher prediction r^2 values (>0.92). The hand-held instruments delivered moderate to high r^2 values between 0.8 and 0.95. Results obtained with the intrinsic data sets revealed typically lower root mean square errors of prediction for apples and kiwifruit (0.32% to 0.73%) and larger prediction errors for summerfruit (0.53% to 0.82%). Some large performance variations between instruments of the same type were observed suggesting caution in evaluating the relative performance of different instrument types or formats on the basis of data generated with just a single instrument and/or data set. However, performance differences between the different hand-held portable instruments, on the same data sets, were often not statistically significant ($p < 0.05$). Instrument choice for any particular application will likely come down to matters not considered here, such as, for example, ease and accuracy during in-field operation and overall reliability.

Keywords

Near infrared spectroscopy, dry matter, multivariate analysis, field portable, prediction

Received 17 March 2017; accepted 15 July 2017

Introduction

Good quality fruits are always preferred for the superior taste experience and increasingly for their health benefits and their role in helping prevent many chronic diseases.[1,2] Fruit quality attributes are also important considerations in terms of marketing, transportation and storage requirements.[3] Fruit dry matter (DM) is an important quality attribute linked to taste across a broad range of fruits.[4,5] For climacteric fruit that store starch during fruit growth, it remains fairly constant during the subsequent ripening stages, with only small changes due to respiration. Fruit DM is typically dominated by the large carbohydrate component (around 75% of DM for kiwifruit, for example) that is sugar and starch at harvest. The starch completely converts to soluble sugars when it ripens, and it is the soluble sugars that ultimately drive consumer preferences for sweet tasting fruits. Hence, the DM indicates either the potential or the actual sugar content, and thus a dominant part of the taste character of the fruit.[6]

The standard method to measure fruit DM is drying and weighing slices of fruit to eliminate water, a method which is obviously destructive. Non-destructive methods based on near infrared (NIR) spectroscopy are well established in commercial fruit grading operations for segregating fruit populations.[7] The first fruit grading system with NIR spectroscopy measurement

[1]School of Engineering, The University of Waikato, Hamilton, New Zealand
[2]The New Zealand Institute for Plant & Food Research Limited, Ruakura Research Centre, Hamilton, New Zealand
[3]The Dodd Walls Centre for Photonic and Quantum Technologies, Dunedin, New Zealand

Corresponding author:
Harpreet Kaur, The University of Waikato, Hamilton 3240, New Zealand.
Email: harpreet.kaur@plantandfood.co.nz

capability was introduced in 1989 by Mitsui Mining Co. Ltd. (Tokyo, Japan) for sorting peaches.[8] Today, NIR spectroscopy for fruit sorting is a mainstay of horticulture packing line operations, and many commercial grader systems offer this feature, e.g. Aweta (Nootdorp, The Netherlands), Sacmi (Bologna, Italy), Greefa (Tricht, The Netherlands), Maf Roda Agrobotic (Montauban, France) and Compac Sorting Equipment Ltd. (Auckland, New Zealand). By contrast, NIR spectroscopy systems for in-orchard and on-tree use, where the technology might be used to make specific harvest selection of fruits or for optimising harvest time, is still in its infancy and not in wide-spread commercial use. However, specifically designed commercial NIR spectrometer instruments for in-orchard use exist. These are hand-held, portable instruments, like the NIRVANA (Integrated Spectronics, Baulkham Hills, Australia; recently discontinued), the F-750 Produce Quality Meter (Felix Instruments, Portland, USA), the SCiO spectrometer version 1.1 (Consumer Physics, Hod HaSharon, Israel), and the NIR portable spectrometer H-100C (Sunforest, Incheon, Korea).

The objective of this research was to examine the performance of these instruments when predicting DM. Three of the hand-held instruments are relatively new (SCiO, H-100C and F-750) and not well covered in the literature, whilst the fourth (NIRVANA) is older and well established.[9,10] As a benchmark, an in-house laboratory NIR spectrometer instrument (Benchtop) of known and reported performance levels was also included in the study.[11] Fruit data sets were specifically created that involved up to three basic fruit types (apples, kiwifruit, summerfruit), each including subcategories of each fruit type depending on what was readily available from a local retail market. This paper discusses the performance of the various spectrophotometer instruments in terms of developing predictive models on each of three 'intrinsic' (same fruit type) data sets and also with a 'combined' (mixed fruit types) data set.

Materials and methods

NIR instruments

The NIRVANA and F-750 instruments are rectangular box-shaped spectrophotometers, of similar measurement geometry, weighing around 1 kg (Figure 1(a) and (b)). The F-750 was developed as a new instrument

Figure 1. Commercially manufactured hand-held, portable instruments: (a) NIRVANA (Integrated Spectronics, Baulkham Hills, Australia; discontinued), (b) F-750 Produce Quality Meter (Felix Instruments, Portland, USA), (c) SCiO spectrometer (Consumer Physics, Hod HaSharon, Israel) and (d) H-100C NIR portable spectrometer (Sunforest, Incheon, Korea).

based on the functional principles of the NIRVANA, which is no longer commercially available. Both models are built around a diode array spectrometer module (MMS1, Zeiss, Jena, Germany).[12] The measurement principle is based on non-contact interactance measurements where the source and detector optics are on the same side of the sample. The main optical components are a small distance (~1 cm) off the target surface. A halogen lamp in the base of the instruments illuminates the fruit surface through a sampling window of 30 mm diameter. A fibre-optic probe in front of the lamp, and close to the sampling window, casts a circular shadow of about 8 mm diameter onto a central zone of the illuminated fruit surface when the fruit is held in position against the sampling aperture of the instrument. When light from the lamp penetrates the fruit, some light is scattered and transmitted through the fruit flesh and is emitted back through the shadowed region into the fibre-optic probe.[9] The probe directs the light to the MMS1 spectrometer. The spectral range of the spectrometer is 310–1100 nm with a sampling interval of 3.3 nm and an optical resolution varying between 8 and13 nm across the wavelength range. Both of these instruments use gold-coated foil as their reference standard, usually set to be measured before or after every sample measurement. The total time to record a spectrum, including reference standard measurement, is approximately 5 s.[9] Two instruments of the NIRVANA (labelled N010 and N011) and two of the F-750 type (F750-09 and F750-10) were used in our experiments.

The SCiO (Figure 1(c)) is a small 68 × 40 mm palm sized spectrophotometer that weighs around 36 g. It can acquire a 740–1070 nm wavelength range spectrum with a sampling interval of 1 nm. The light source is LED based, emitting light broadly in the visible and NIR ranges. The device operates in interactance mode, measuring light emitted back out of the sample from the same side the light enters. The light scattered and emitted from the sample passes first through an optical filter and is then dispersed by a fast Fourier transform focussing element onto an image sensor. The sensor data are automatically interrogated to provide a spectrum that is wirelessly transferred to a smartphone for display, storage or further analysis. The low power consumption and zero warm up time makes the instrument highly responsive and efficient to use, with sufficient sensitivity and resolution (typically $< 10\,cm^{-1}$) to obtain useful spectra.[13] Hundreds of scans can be performed from a small rechargeable battery. There is a default option that transfers the spectra to a cloud-based service for further analysis and review. For version 1.0 of the SCiO, the use of an optical shade was recommended to maintain a distance of about 0.5 in. (~1.25 cm) between the sample and the optical interface of the sensor. For later versions (1.1) that also incorporate an additional colour filter to maximise the spectral response for fruit measurement, the recommendation is that the sensor optics makes direct contact with the fruit.

The Sunforest H-100C instrument is a portable gun-shaped spectrophotometer which is advertised as applicable for mandarins and kiwifruits. It weighs about 420 g in normal operation (without head cap) and measures a spectrum at 2 nm intervals between 650 nm and 950 nm. The head of the H-100C contains apertures for the light source (a small halogen lamp) and sensor (enhanced CMOS spectrometer), which are arranged so that an interactance measurement is made on one side the fruit. The optical interfaces are in contact or close to contact with the fruit. With a fully charged battery, the H-100C can make 5000 measurements when operating in the standalone mode. The on-board memory is able to store 1000 measurements, and a user program is provided to allow transfer to a PC for display, storage and further analysis.

The Benchtop instrument is a specially developed laboratory system designed to work in full interactance mode without any risk of stray light from the source getting to the sensor.[11] It consists of a wideband light source (50 W quartz halogen, RJL 5012 FL, Radium, Germany), a fruit holder/light collection fixture and a non-scanning polychromatic/diode array spectrometer (MMS1-NIR, Zeiss, Jena, Germany). Fruit is usually placed on the fruit holder and illuminated from below by the light source. Light from the source penetrates the fruit through the exposed regions around the holder and is scattered and transmitted through the kiwifruit flesh to exit from the fruit on the inside of the holder. A 12 mm diameter rubber grommet acts as both a light seal against surface reflections and as a flexible support to accommodate differently shaped fruit. The inside of the holder contains a mirror and fibre optic cable for directing the exiting light into the spectrometer, which had a spectral range of 300–1140 nm with a 3.3 nm sampling interval. A 15-bit analogue to digital converter (FEE-001, tec5, Oberursel, Germany) is used to amplify and digitise the spectral signal. Data acquisition and spectra storage are achieved with a PC running software developed in-house (NIRFruit, HortResearch, 1999).[11]

Basic specifications for each instrument, compiled from manufacturer's data, are summarised in Table 1. Figure 2 illustrates the different approaches to sample presentation.

Data sets

Two data sets were created, Lot I and Lot II. Lot I consisted of spectral data and DM measurements made on fully ripe and ready-to-eat fruits purchased from New Zealand retail stores in the month of February. The set contained three different fruit types: apples (a mix of Granny Smith (GS), Royal Gala (RG), Sweet Tango (ST), Braeburn (BB), and New Zealand Rose (NR)), kiwifruit (a mix of Hayward (HW), Zespri Gold (HA), and Zespri SunGold (SG)), and summerfruit (a mix of apricots (AP), nectarines (NC), peaches (PC), and plums

Table 1. Specifications of the various spectrophotometers used in our experiment.

Instrument	λ range (nm)	Resolution	Sampling interval	Light source	Spectrometer	Weight	Measurement time
Benchtop (Plant and Food Research, Hamilton, New Zealand)	310–1100	8–13 nm	3.3 nm	Quartz halogen	Carl Zeiss MMS-1	15 kg	100 ms
F750 (Felix Instruments, Portland, USA)	310–1100	8–13 nm	3.3 nm	Xenon tungsten lamp	Carl Zeiss MMS-1	1.05 kg	5 s
N011 (Integrated Spectronics, Baulkham Hills, Australia)	310–1100	8–13 nm	3.3 nm	Halogen lamp	Carl Zeiss MMS-1	1 kg	5 s
H-100C (Sunforest, Incheon, Korea)	650–950	<20 nm[a]	2 nm	Halogen lamp	Enhanced CMOS	420 g	2 s
SCiO (Consumer Physics, Hod HaSharon, Israel)	740–1070	<10 cm^{-1} (1 nm)	1 nm	LED	Image sensor	36 g	1.5 s

[a]Estimated from full width half maximum (FWHM) signal of narrowband diode laser.

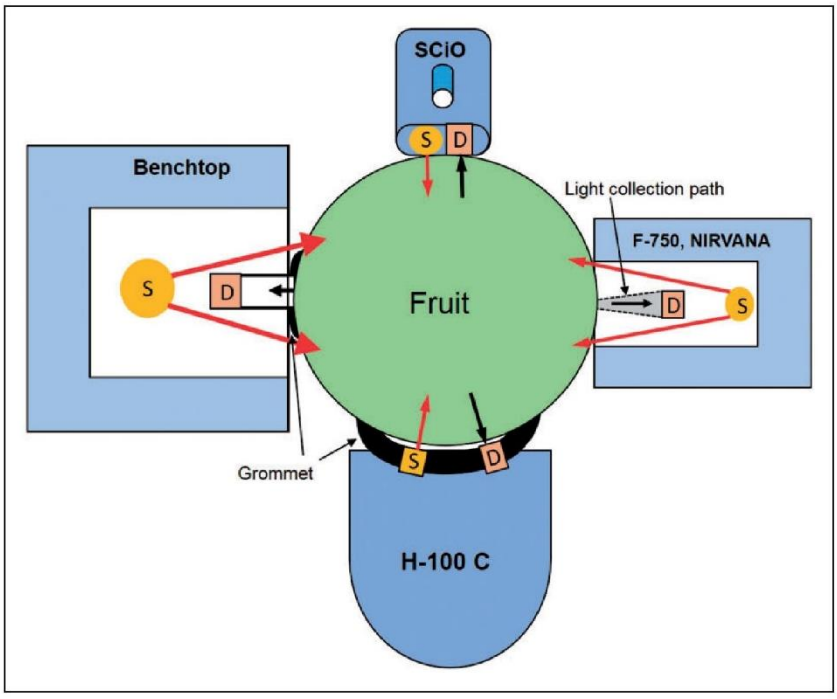


Figure 2. Sample presentation of the various spectrophotometers. S: source, D: detector/spectrometer, Red arrow: incident light, and Black arrow: returned light.

(PL)). The fruit were kept at room temperature (20 ± 2°C) in the laboratory and measured within a one or two days from purchase. Fruit spectra were recorded with each of the different NIR spectroscopy instruments taking two separate measurements on opposite sides in the equatorial plane of each fruit.

Reference DM measurements were made within 2 to 3 h of the spectral measurements.[4] For apples and summerfruits, DM was measured on the side of the fruit from which spectral readings were taken. Two thin slices of approximately 3 mm thickness were taken from the sides of each fruit and dried in an oven at 65°C for 24 h. DM was calculated as the percentage of final dry weight to initial wet weight of the slices. Each DM value was then associated with the respective spectrum. Only one DM measurement was taken from the equatorial plane of kiwifruits which was then used as the calibration value for both spectral measurements on the fruit.

After omission of four fruits because of clerical errors, the final Lot I data set consisted of 201 fruit (402 spectra) of which 47 were apples, 74 were kiwifruit

and 80 were summerfruit. The SCiO instrument was eventually excluded from the data set as the early version (1.0) of the instrument proved to be inadequate for fruit measurement. Another data set, Lot II, was collected when a new and improved version of the SCiO (version 1.1) became available. The Lot II data set consisted of fully ripe and ready-to-eat apples and kiwifruit, which were readily available during the month of August. Spectral measurements were made with the SCiO, Benchtop and H-100C instruments. In total there were 46 apples from a mix of Granny Smith, Royal Gala, Braeburn, Fuji and Pink lady and 45 kiwifruits from a mix of Hayward and Zespri SG fruit.

Multivariate analysis

Predictive models were developed using MATLAB version R2014b (MathWorks Inc., Natick, MA) and the PLS toolbox version 8.1.1 (Eigenvector Research Inc., Wenatchee, WA) with four-fold venetian blind cross-validation applied. Consumer Physics, the manufacturers of the SCiO device, provides a cloud-based facility for predictive model development for their device, but that was not used here as it produced poorer results than we could otherwise obtain. To ensure no over fitting of the model, latent variables were selected based upon lowest root mean square error of cross validation (RMSECV).

The spectral region around 734–931 nm has been used successfully in the literature to make calibration models for DM of fruit.[5] However, the investigated instruments have quite different operating ranges (Table 1). To ensure that the performance of an instrument was not restricted by choice of wavelength range, a set of seven different wavelength ranges was investigated: 800–950 nm, 800–1000 nm, 740–1070 nm, 650–950 nm, 740–950 nm, 740–1000 nm, and 600–1000 nm. If an instrument covered one or more of these ranges, calibration models were built on various data sets. The wavelength ranges, shown in Table 3 and Table 5, were selected as those where the RMSECV was lowest.

The H-100C spectra were already in second derivative form and were used directly, while the spectra from the other instruments were transformed into a similar form. The spectral data of the F750, NIRVANA and Benchtop instruments were all pre-processed using a combination of a standard normal variate (SNV) transformation followed by a second derivative transformation (Savitzky-Golay with filter width of 15 pixels and 2nd order polynomial smoothing). The combination of SNV followed by second derivative improved results compared to using each separately, confirming literature claims that it is a useful pre-processing combination.[14] Pre-processing was the same for the SCiO, except that a 49 pixel filter width was used for the Savitzky-Golay transformation. This resulted in a second derivative spectral resolution that was similar to the other devices. The resolution of the SCiO is stated to be about one-third of the other devices, but averaging to the coarser resolution gave better calibration results.

Predictive models were developed and analysed for two different scenarios:

1. *Intrinsic*: Here each of the main fruit types were analysed separately (apples, kiwifruit or summerfruit for Lot I; apples or kiwifruit for Lot II).
2. *Combined*: All fruit types were considered together in Lot I or Lot II. Predictive models were developed for the total population of each lot.

Statistical analysis

Calibration modelling and validation subset splits were arranged on each data set through a venetian blind selection process. The intrinsic data sets were split into four, the combined Lot I into 10, and the combined Lot II into seven subset splits. The selection process was applied to data that had first been DM rank ordered within each fruit sub-category. That is, for example, the five different apples varieties in data set Apple Lot I were each individually rank ordered in terms of DM and then combined to form the full Apple Lot I data set upon which the venetian blind selection process was conducted. This meant all subset splits had very similar DM profiles. The data sets were split 4, 10 or 7 ways to ensure each independent validation subset had at least 20 fruits which was considered sufficient for statistical analysis purposes. Calibration modelling and validation analysis were undertaken by holding out each subset split in turn from the calibration modelling and then using the generated model on the held-out subset split for independent validation of the model. In that way performance statistics were generated from multiple models, and ANOVA analysis of the differences between instruments was made possible. One-way ANOVA was employed, using the Matlab Statistics Toolbox 11.1 (MathWorks Inc), with the data set splits included as a blocking effect. Post hoc pair-wise comparisons between different instruments were made using the Bonferroni method which adjusts for the circumstance of multiple comparisons.

Results and discussion

Across the full selection of fruit the mean DM varied from 11.80% to 18.21% in Lot I and from 14.06% to 18.05% in Lot II (Table 2). Fruit varieties with the highest and lowest mean DM on Lot I were Hayward kiwifruit (DM 18.21%) and peaches (DM 11.8%), respectively. For the Lot II data the varieties with the highest and lowest mean DM were SG kiwifruit (DM 18.05%) and Granny Smith (DM 14.06%). Summerfruit exhibited the largest within variety variations, particularly peaches and apricots for which the standard deviations were $> 2\%$. For apples and

Table 2. Distributional statistics for DM % on different fruit types making up the Lot I and II data sets.

	Lot I				Lot II			
Data sets	Fruit types	*N*	Mean	SD	Fruit types	*N*	Mean	SD
Apple	Royal Gala (RG)	20	13.74	0.90	RG	20	15.25	1.01
	Braebrun (BB)	18	15.35	0.98	BB	18	15.83	1.09
	Granny Smith (GS)	20	13.80	0.83	GS	18	14.06	1.96
	NZ Rose (NR)	18	15.15	0.73				
	Sweet Tango (ST)	20	15.78	1.13				
					Fuji (FJ)	18	14.76	1.01
					Pink Lady (PL)	18	14.94	0.91
	All	94	14.74	1.24	All	92	14.97	1.13
Kiwifruit	Hayward (HW)	48	18.21	0.95	HW	40	17.80	1.35
	SunGold (SG)	50	15.92	0.89	SG	50	18.05	1.67
	Hort16A (HA)	50	17.47	0.69				
	All	148	17.19	1.27	All	90	17.9	1.56
Summerfruit	Plum (PL)	48	13.39	1.50				
	Peach (PC)	50	11.80	2.18				
	Apricot (AP)	20	16.54	2.03				
	Nectarine (NC)	42	12.07	1.38				
	All	160	12.94	2.32				

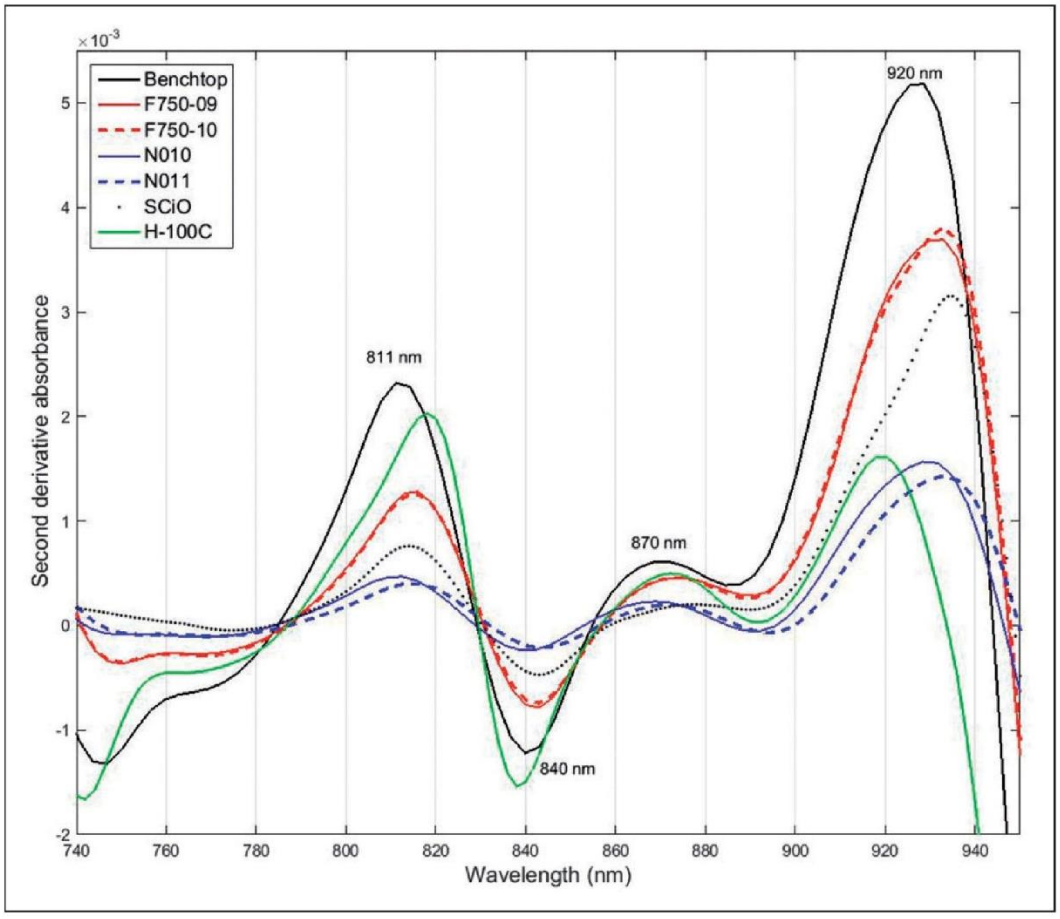


Figure 3. Average second derivative spectra of apples measured by different spectrometers. Spectra for SCiO are plotted at 10 times scale, Felix F750-09 and F750-10 at two times scale. Labels indicate the wavelengths of the peaks measured with the Benchtop instrument. N010 and N011 are two different NIRVANA instruments; F750-09 and F750-10 are two different F750 devices.

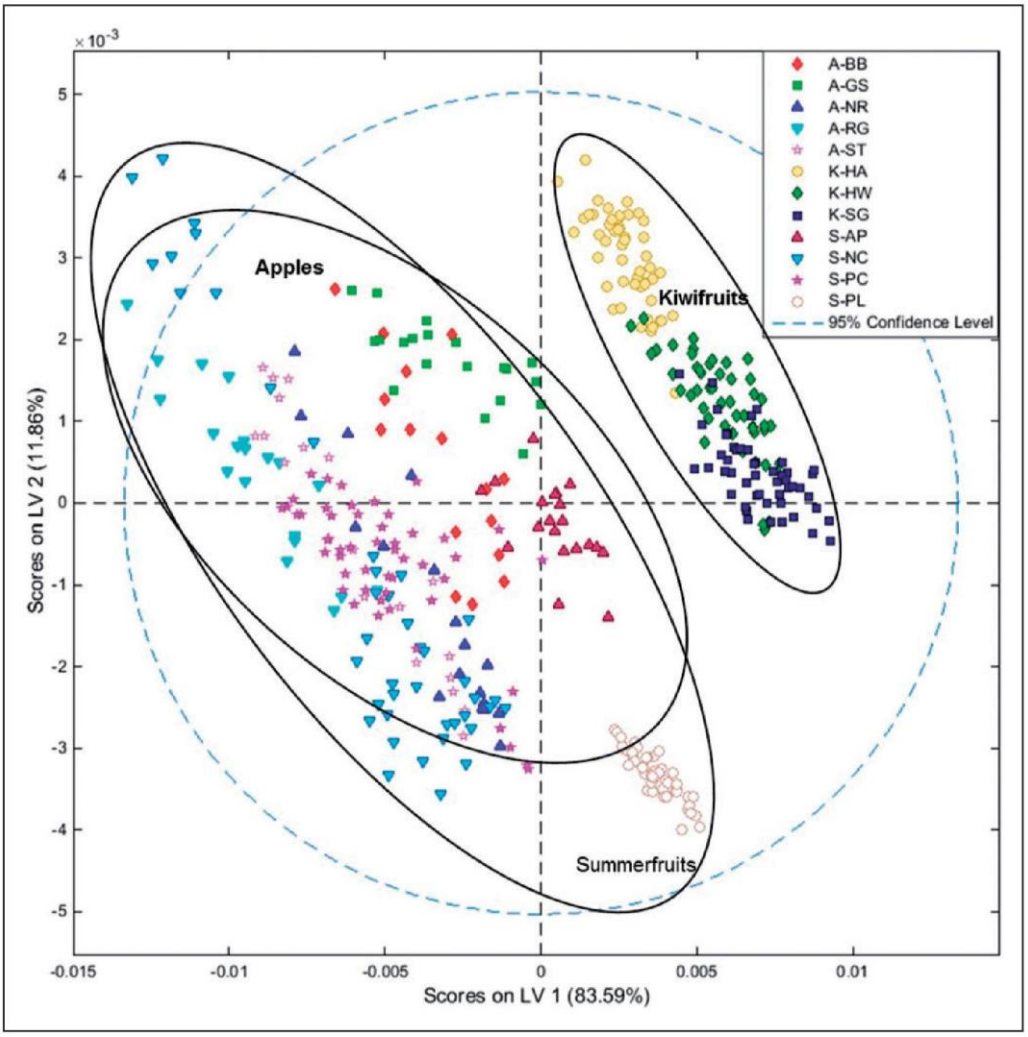


Figure 4. Score plot of the first two latent vectors (LVs) produced by the combined data set of Lot I with the Benchtop instrument. Circles group the fruits by main types with symbols separating sub-categories (A: apples; K: kiwifruit; S: summerfruit).

kiwifruit the standard deviations were typically much lower, < 1% for Lot I and between 1% and 1.7% for Lot II fruit.

The average second derivative absorbance spectra were quite similar for the seven spectrophotometers, but showed slight differences in the main modelling region of 740–950 nm (Figure 3). The expected spectral features were observed, particularly troughs (with attendant side lobe peaks) corresponding to water absorbance bands around 760 nm, 840 nm and 960 nm. However, precise peak or dip positions varied between instruments. At 928 nm the H-100C had the largest deviation from the other instruments of −6 to −10 nm. The two NIRVANA instruments had shifted wavelength scales, for instrument N010 the scale was 4 nm lower than that of N011. The SCiO instrument exhibited some subtle differences in shape to the other instruments, such as an inflexion point around 920 nm and a broader/flatter peak around 870 nm.

There were clear spectral differences between the main three fruit types and the various sub-categories. This is illustrated by the latent vector scoreplots from PLS modelling of Lot I data from the Benchtop instrument (Figure 4). The kiwifruit scores for the first two latent vectors did not overlap at all with the other fruit types, and the three individual kiwifruit sub-categories also showed a strong degree of segregation. The kiwifruit scores for the Lot II data set were similarly isolated and different to that of the apples in that data set. The Lot I scores for apples and summerfruit were moderately overlapping, with the only strongly distinct and isolated group being the plums (S-PL in Figure 4). It is not surprising that such distinct groupings were observed in the first two latent variables. The obvious physical differences (both morphological and physiological) that result in the fruit visibly looking (size, shape, colour or similar) and feeling (stiffness, texture) quite different will result in spectral differences. A DM

Table 3. Performance of various instruments in predicting DM on the three intrinsic fruit data sets.[a]

Apples	Lot I ($N_{cal}=70$ and $N_{val}=24$)					
Instrument	λ range	R^2_{cv}	RMSECV	LV	r^2_p	RMSEP
Benchtop	740–950	0.93	0.33 (±0.03)	6	0.95	0.32(±0.08)[b]
F750-09	600–1000	0.78	0.59 (±0.03)	13	0.84	0.54 (±0.05)[c,d]
F750-10	600–1000	0.80	0.56 (±0.02)	12	0.78	0.63 (±0.06)[d,e]
N010	650–950	0.76	0.62 (±0.03)	10	0.77	0.64 (±0.11)[d,e]
N011	650–950	0.85	0.49 (±0.04)	14	0.88	0.46 (±0.08)[b,c]
H-100C	740–950	0.72	0.67 (±0.04)	8	0.72	0.73 (±0.11)[e]
	Lot II ($N_{cal}=69$ and $N_{val}=23$)					
Benchtop	650–950	0.85	0.44 (±0.05)	11	0.88	0.41 (±0.09)[b]
H-100C	740–950	0.56	0.77 (±0.08)	7	0.61	0.72 (±0.13)[c]
SCiO	740–950	0.82	0.48 (±0.02)	6	0.86	0.45 (±0.04)[b]
Kiwifruit	Lot I ($N_{cal}=111$ and $N_{val}=37$)					
Instrument	λ range	R^2_{cv}	RMSECV	LV	r^2_p	RMSEP
Benchtop	740–950	0.92	0.35 (±0.02)	13	0.93	0.34 (±0.02)[b,f,g]
F750-09	600–1000	0.81	0.56 (±0.01)	13	0.81	0.58 (±0.08)[c,d]
F750-10	600–1000	0.84	0.52 (±0.03)	14	0.86	0.49 (±0.07)[c,d,e,f,g]
N010	650–950	0.87	0.46 (±0.01)	12	0.88	0.46 (±0.04)[d,e,f,g]
N011	650–950	0.90	0.41 (±0.01)	14	0.91	0.38 (±0.02)[b,d,e,f,g]
H-100C	800–950	0.85	0.49 (±0.02)	7	0.89	0.44 (±0.07)[b,d,e,f,g]
	Lot II ($N_{cal}=68$ and $N_{val}=22$)					
Benchtop	740–950	0.91	0.46 (±0.03)	8	0.90	0.48 (±0.08)[b]
H-100C	800–950	0.87	0.55 (±0.04)	6	0.91	0.50 (±0.04)[b]
SCiO	740–1000	0.85	0.59 (±0.03)	9	0.84	0.66 (±0.02)[c]
Summerfruit	Lot I ($N_{cal}=120$ and $N_{val}=40$)					
Instrument	λ range	R^2_{cv}	RMSECV	LV	r^2_p	RMSEP
Benchtop	650–950	0.93	0.64 (±0.09)	15	0.93	0.61 (±0.12)[b]
F750-10	800–1000	0.90	0.73 (±0.04)	7	0.90	0.77 (±0.11)[b]
N010	650–950	0.90	0.72 (±0.04)	12	0.92	0.69 (±0.05)[b]
N011	650–950	0.95	0.53 (±0.07)	13	0.95	0.53 (±0.13)[b]
H-100C	700–950	0.88	0.80 (±0.05)	11	0.89	0.82 (±0.17)[b]

LV: latent variable; RMSECV: root mean square error of cross validation; RMSEP: root mean square error of prediction.
[a]The wavelength range (nm) was that producing the lowest RMSECV for each data set/instrument combination. The RMSECV and RMSEP values are DM %; the bracketed value is the standard error calculated from the 4 data splits. The same superscript letter alongside the RMSEP value indicates there is no statistically significant difference between these results.

model must take account of these by either amplifying or reducing their spectral effect, depending on whether the effects are related directly to the fruit DM or not.

Results for PLS modelling of the intrinsic data sets are reported in Table 3. The root mean square error of prediction was calculated as

$$RMSEP = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{N}} \quad (1)$$

where N denotes the size of the prediction set, and $\hat{y}_i$ and y_i are the prediction and reference values for sample i, respectively. Differences between RMSECV and root mean square error of prediction (RMSEP) values were small and variable in each case, equally positive and negative across the data sets, and so statistically consistent with optimised fitting of the models. Significant differences in RMSEP between instruments for the same data sets were revealed through ANOVA ($p < 0.05$).

Post hoc pair-wise comparisons showed that for the kiwifruit Lot I data set the Benchtop, N011 and H-100C instruments were the best performers with RMSEP values of 0.34%, 0.39% and 0.44%, respectively. This was significantly better than the other three instruments. With the Apples Lot I data set, the Benchtop and N011 instruments were again significantly better than the other instruments with RMSEP values of DM 0.32% and 0.46%, respectively.

The H-100C instrument was the worst performer on the Apple data set which is not unexpected given it is not sold as suitable for apples. For the Summerfruit Lot I data set the N011 instrument was the best recording a low RMSEP value of DM 0.53%, although there was no significant difference between the results from all five instruments.

The SCiO instrument, only tested on the Lot II data sets, performed relatively well on the Apple data set with an RMSEP value of DM 0.45%. This was not statistically different to the Benchtop result (DM 0.41%) and is clearly comparable to the results reported with the other hand-held portable instruments on the Apple Lot I data set (RMSEP values ranging from 0.46% to 0.73%). It was surprising that the SCiO instrument did not perform well on the kiwifruit Lot II data set. We do not have an explanation for that.

There was some significant variation in model performance between instruments of the same type. The pairs of F-750 and NIRVANA instruments on the Lot I data sets delivered RMSEP values different by between DM 0.08% and 0.18%. These seem quite large differences at first glance, around 20% of an average RMSEP value at DM 0.5%, but only the largest

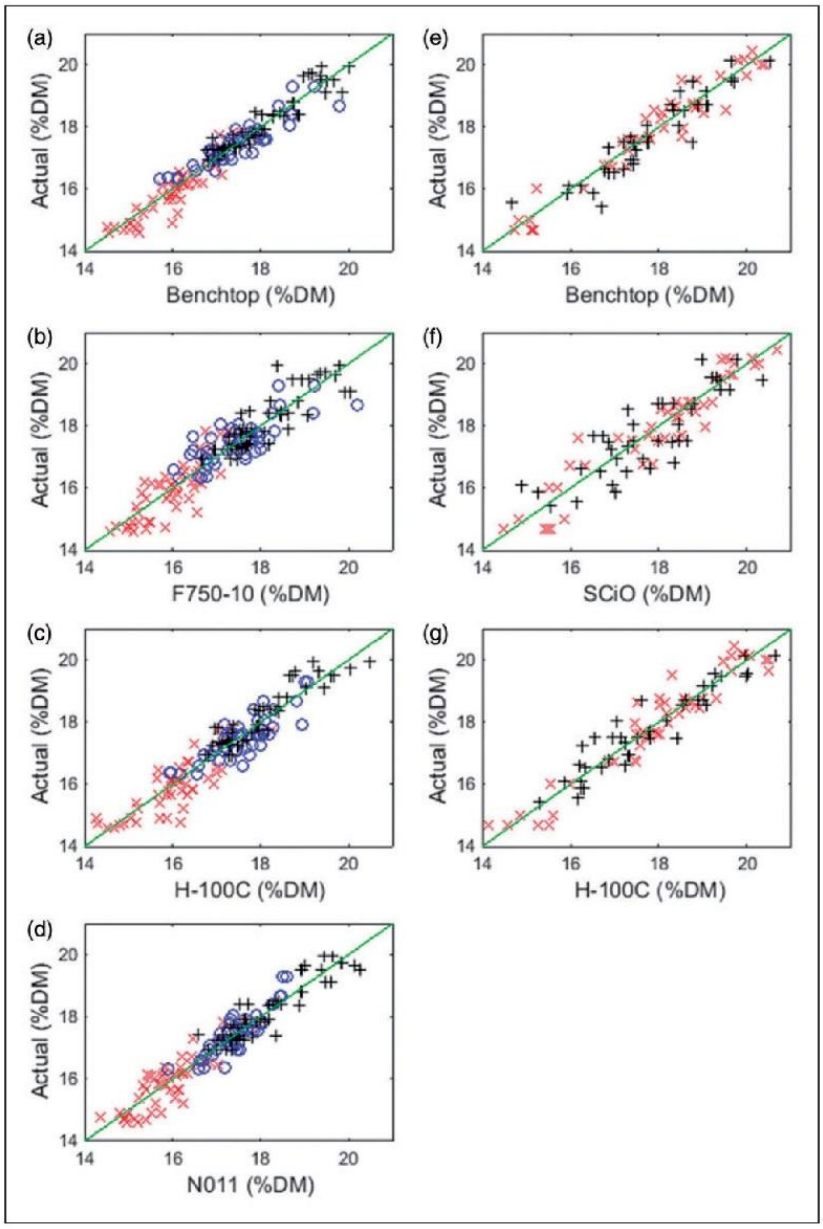


Figure 5. Actual versus predicted DM values for all four validation data set splits of the intrinsic kiwifruit data sets (x: SunGold; o: Zespri Gold; +: Hayward). (a), (b), (c) and (d) represent Lot I predictions, and (e), (f) and (g) represent Lot II predictions.
DM: dry matter.

difference of DM 0.18% between N011 and N010 on the Apple Lot I data set proved to be statistically significant ($p < 0.05$). It is possible that the differences were caused by the wavelength calibrations which are markedly different (about 4 nm) between the two NIRVANA instruments (Figure 3).

The optimised wavelength ranges were very similar for most of the instruments on the intrinsic data sets. The upper edge of the wavelength window is always either 950 nm or 1000 nm, indicating the importance of incorporating information related to the broad water absorbance peak that is centred at 970 nm. Its spectral influence is obvious at 950 nm and below (Figure 3). The lower edge of the window is 600 nm to 800 nm, varying by instrument within each data set and suggesting it is related to the optical geometry of the instruments. For instance, the non-contact interactance systems (F-750, NIRVANA) favoured a lower edge at 600 or 650 nm whereas the other interactance systems, involving more direct sample contact (Benchtop, H-100C, and SCiO), favoured 740 or 800 nm. It is probable that the non-contact interactance systems are more sensitive to surface information and that a higher fraction of directly reflected surface or near-surface light would likely be recorded. Hence visible light wavelengths related to skin pigments at the red-end of the absorbance spectrum, such as chlorophyll at 670 nm, might then be usefully incorporated into a predictive model.

Scatterplots of actual versus predicted values for some of the different instruments are shown in Figure 5 for the kiwifruit data sets. The predictive models were linearly well behaved with reasonably uniform scatter across the range. This was also observed for the other fruit data sets. The kiwifruit scatterplots in Figure 5 also show that for the Lot I data set the DM predictions were moderately segregated with respect to the fruit sub-categories. The SG variety has lower average DM than the other two kiwifruit types. This might suggest that the sub-category type was helping forge the DM model. However, no such sub-category segregation was observed with the Lot II kiwifruit predictions, implying the Lot I observation of kiwifruit sub-category segregation in the predictions may only be coincidental for forging the Lot I DM model.

Models generated with the Benchtop and the H-100C instruments, on the Lot I Intrinsic data sets for apples and kiwifruit, were applied to make predictions on the Lot II Intrinsic data set for the same instruments (Table 4). The Benchtop model from Lot I worked well on the Lot II data with high prediction r_p^2 values of 0.89 and 0.92, low standard error of prediction (SEP; RMSEP corrected for bias) of DM 0.37% and 0.44%, and moderate bias values of DM 0.44% and 0.43% for the apples and kiwifruit data sets, respectively. The H-100C DM model for Lot I kiwifruit was successfully applied to Lot II kiwifruit as well, with an r_p^2 value of 0.83, SEP of 0.64% and a bias of −0.4. This is still significantly different and poorer, at the 95% confidence level, from the Benchtop result using the standard error comparison method by Fearn.[15]

Table 4. Performance statistics for the Benchtop and H-100C instruments when applying Lot I generated models to Lot II data.

Data set	Instrument	N_{val}	r_p^2	RMSEP	SEP	Bias
Apple	Benchtop	92	0.89	0.57	0.37	0.44
	H-100C	92	0.52	1.05	0.79	0.7
Kiwifruit	Benchtop	90	0.92	0.62	0.44	0.43
	H-100C	90	0.83	0.76	0.64	−0.41

N_{val}: the number of samples in validation set; r_p^2: the coefficient of determination for prediction; RMSEP: root mean square error of prediction; SEP: standard error of prediction (bias corrected RMSEP).

Table 5. Prediction of dry matter using combined fruit data sets for Lot I (apples, kiwifruit, summerfruit) and Lot II (apples and kiwifruit).[a]

	Lot I ($N_{cal} = 370$ and $N_{val} = 32$)					
Instrument	λ range	R_{cv}^2	RMSECV	LV	r_p^2	RMSEP
Benchtop	650–950	0.95	0.59 (±0.01)	15	0.96	0.54 (±0.10)[b]
F750-10	650–950	0.91	0.75 (±0.01)	16	0.93	0.70 (±0.11)[c]
N010	600–1000	0.89	0.86 (±0.01)	18	0.93	0.69 (±0.06)[c]
N011	650–950	0.95	0.59 (±0.01)	17	0.95	0.56 (±0.08)[b]
H-100C	740–950	0.91	0.78 (±0.01)	15	0.93	0.68 (±0.09)[c]
	Lot II ($N_{cal} = 158$ and $N_{val} = 24$)					
Benchtop	740–950	0.92	0.58 (±0.02)	12	0.93	0.55 (±0.10)[b]
H-100C	800–950	0.89	0.66 (±0.03)	7	0.89	0.66 (±0.13)[c]
SCiO	800–1000	0.88	0.70 (±0.02)	12	0.87	0.70 (±0.07)[c]

LV: latent variable; RMSECV: root mean square error of cross validation;RMSEP: root mean square error of prediction.
[a]The RMSECV and RMSEP values are DM %; the bracketed value is the standard error calculated from the 10 individual data splits for Lot I and 7 for Lot II. The same uppercase letter alongside the RMSEP value indicates there is no statistically significant difference between results with the same letter.

The H-100C DM model for Lot I apples was unsuccessfully applied to Lot II apples, which was expected as the H-100C instrument were not specifically designed for apples measurements.

Table 5 shows the performance statistics of DM models developed for the combined fruit data sets of Lot I and Lot II. Overall, the results were very good for all instruments with high R^2 and r^2 values > 0.87, whether in cross-validation or on the independent prediction sets. The Benchtop and N011 instruments performed significantly better than the rest with very high R^2 and r^2 values > 0.95 and low prediction errors of DM $< 0.6\%$. Results with the other field portable instruments were more moderate with R^2 values between 0.88 and 0.93, and predictive errors of between DM 0.66% and 0.86%. The SCiO instrument, only tested on Lot II data sets, performed equivalently to the H-100C with $r^2_p \sim 0.87$ and RMSEP of 0.70%. Interestingly the H-100C and SCiO both used a relatively high lower edge of 800 nm for the wavelength range on Lot II, which only involved apple and kiwifruit samples, compared to lower edges from 600 to 740 nm for most of the other instruments on Lot I. The Lot II data set contained no summerfruit, and it is quite possible that the red-end visible wavelengths then carry no useful information for spectra collected with those two instruments.

Conclusion

Each instrument in this study delivered models with high linear correlations, r^2, ranging from around 0.80 to 0.95, between actual DM and NIR predicted values. Single fruit type or cultivar data sets delivered better performing models, lower RMSEP values, than models created from large combined fruit type data sets. The older and/or more established instruments (Benchtop and NIRVANA) were typically better performing than the more recently developed ones (F-750, H-100C, and SCiO). One of the NIRVANA instruments delivered the best results amongst the hand-held field portable units. However, the other NIRVANA instrument examined demonstrated some large and poorer differences in performance on the same data sets. This suggests caution is required when evaluating the relative performance of different instrument types or formats on the basis of data generated with just a single instrument and/or data set. Choice of instrument then may likely come down to matters other than those of instrument precision and accuracy, not addressed here, such as reliability and robustness in field operation under conditions of varying ambient light and/or temperature, etc.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: PhD scholarship from The University of Waikato, New Zealand and MBIE Contracts C11X1208 & C11X1601 with Plant & Food Research, New Zealand.

References

1. Boeing H, Bechthold A, Bub A, et al. Critical review: vegetables and fruit in the prevention of chronic diseases. *Eur J Nutr* 2012; 51: 637–663.
2. Hall JN, Moore S, Harper SB, et al. Global variability in fruit and vegetable consumption. *Am J Prev Med* 2009; 36: 402–409.e5.
3. Ahmad MS and Siddiqui MW. Factors affecting postharvest quality of fresh fruits. In: *Postharvest quality assurance of fruits: practical approaches for developing countries*. Cham: Springer International Publishing, 2015, pp.7–32.
4. McGlone VA and Kawano S. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biol Technol* 1998; 13: 131–141.
5. Walsh K, Golic M and Greensill C. Sorting of fruit using near infrared spectroscopy: application to a range of fruit and vegetables for soluble solids and dry matter content. *J Near Infrared Spectrosc* 2004; 12: 141–148.
6. Gordon Mitchell F, Mayer G and Biasi W. Effect of harvest maturity on storage performance of "hayward" kiwifruit. *Acta Hortic* 1992; 297: 617–626.
7. Kawano S. Present condition of nondestructive quality evaluation of fruits and vegetables in Japan. *Jpn Agric Res Q* 1994; 28: 212–216.
8. Kawano S. New application of nondestructive methods for quality evaluation of fruits and vegetables in Japan (quality and its evaluation of horticultural products, for further development of horticulture in East Asia). *J Jpn Soc Hortic Sci* 1998; 67: 1176–1179.
9. Shafie K, Künnemeyer R, McGlone A, et al. An optimised six-wavelength model for predicting kiwifruit dry matter. *J Near Infrared Spectrosc* 2015; 23: 103–109.
10. Kumar S, McGlone A, Whitworth C, et al. Postharvest performance of apple phenotypes predicted by near-infrared (NIR) spectral analysis. *Postharvest Biol Technol* 2015; 100: 16–22.
11. McGlone VA, Jordan RB, Seelye R, et al. Comparing density and NIR methods for measurement of kiwifruit dry matter and soluble solids content. *Postharvest Biol Technol* 2002; 26: 191–198.
12. Greensill CV and Walsh KB. Optimization of instrumentation precision and wavelength resolution for the performance of NIR calibrations of sucrose in a water–cellulose matrix. *Appl Spectrosc* 2001; 54: 426–430.
13. Goldring D and Sharon D. *Low-cost spectrometry system for end-user food analysis*. Patent 2014/0320858 A1, USA, 2013.
14. Rinnan Å, Berg Fvd and Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Anal Chem* 2009; 28: 1201–1222.
15. Fearn T. Comparing standard deviations. *NIR News* 1996; 7: 5.