Milk Spectroscopy

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Engineering

at

The University of Waikato

by

Andrew Russell

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2013

Abstract

This thesis examines the capability of two different MicroNIR spectrometers to measure the composition of raw milk for an online milk sensor.

On-farm milk testing is an important tool in a developed dairy market. It is used by farmers to determine milk quality, nutrition and individual animal performance. The price farmers receive for milk is based on the fat and protein content. Currently milk composition is monitored on average three times a season, which leads to late diagnosis of poor performance. Daily monitoring of milk composition allows farmers better farm management and optimisation of the processes and feed to achieve success. The International Committee for Animal Recording (ICAR) sets standards for sensors which give farmers confidence in the data they receive. Two low cost spectrometers were assessed to see if they meet the ICAR fat and protein regulations for an online sensor.

The spectrometers covered two spectral regions in the NIR range and were both developed by JDS Uniphase Corporation (JDSU, California, USA). The MicroNIR1700 spectrometer analyses the 870 – 1660 nm region, and the MicroNIR2200 spectrometer covers the 1130 – 2150 nm region. An initial laboratory trial with the unmodified spectrometers did not meet ICAR precision limits for milk analysis. The high prediction errors were caused by insufficient signal, aged milk samples being reheated and excessive dark current.

To improve the signal to noise ratio of the spectrum the lamp voltage was increased. This increased the received signal to the spectrometers but also increased the temperature of the spectrometers. The increase in temperature caused the dark current to excessively increase. To reduce the operating temperature of the spectrometers a thermoelectric cooler was used to cool the spectrometers to 20°C.

The modified spectrometers with increased light and thermoelectric cooler were tested on three different farms to establish if the modifications done can meet ICAR fat and protein precision limits. A total of 590 milk samples

were collected across three different farms. Calibration models were tested on each individual farm and across all three farms to assess the capability of calibration models to be transferred across the different farms.

Off-line and on-farm, the modified MicroNIR spectrometers achieved ICAR fat and protein precision limits for an online sensor. For the pooled data set the MicroNIR1700 spectrometer predicted fat and protein with a root mean square error of 0.061 and 0.151 respectively. The MicroNIR2200 predicted fat and protein with a RMSE of 0.061 and 0.138. The independent calibration models could be transferred across farms with only a slight increase in prediction error but still within the ICAR tolerances for an online milk sensor

It is concluded the original MicroNIR spectrometers could not predict fat and protein to ICAR precision limits for an online sensor. To achieve ICAR accuracy with the MicroNIR spectrometers the light output needed to be increased and cooling was required to reduce the operating temperature.

This work shows that there is definite potential for low cost spectrometers to be used for the determination of milk properties.

Acknowledgements

I wish to thank the people who have in some way assisted in the research for this thesis and the Ministry of Science and Innovation for providing the necessary funding for the project. Firstly I want to acknowledge my industry supervisor Dr. Paul Johnstone who provided significant support, insight, organisation and direction throughout this project. I would also like to thank Paul for relentless help during the extensive farm trials.

Secondly I am thankful for the expertise of Associate Professor Rainer Künnemeyer who provided useful support for quantifying the sources of errors. Also I would like to thank Rainer for the significant support given in the monthly meetings throughout the project.

I would like to thank Don Friedrich and Chris Pederson for their expertise in spectroscopic techniques and data analysis. I thank Dwayne Gyde and Allan Wilson for their time and expertise in the field of electronics. I thank Rob Orchard and Vicki Farr for their insight in milk knowledge and assistance during the farm trials. I thank John McLeish, Jason Sun, and Jennie Feast for their assistance in collecting milk samples during the farm trials.

A special thanks to the farmers who allowed milk samples to be obtained for the trial must be given. George Hill must be acknowledged for his support during the initial laboratory testing of the spectrometers.

Finally I want to acknowledge my parents and friends for their endless support during my life.

Table of Contents

List of Figures

List of Tables

1 Introduction

The work presented in this thesis was carried out in collaboration with Lely Sensortec (Hamilton, NZ), JDSU (California, USA), and the School of Engineering of the University of Waikato. Funding was provided by the Ministry of Science and Innovation. Lely Sensortec is a research and development company that specialises in developing milk component-analysis sensors to enable improved herd management and farm profitability.

Measurement of milk composition is essential for the dairy industry and dairy management. Daily measurement of milk composition for each individual cow is important for animal breeding, efficient utilization of cows and for nutrition management.

Routine analytical methods for milk testing are destructive, expensive, time and labour consuming and are off line in general. The modern management of dairy farms involves automated milking systems that control the production process and collect on line data for each individual cow. Depending on the quality of the milk produced the feed type could be adjusted according to the health of the animal.

In the dairy industry, near infrared (NIR) has been widely used for analysing the major components in milk and fermented milk products. Laporte and Paquin (1999) used the Foss-NIRSystems model 6500 to measure the near infrared region of 1100 to 2200 nm. The authors were able to predict fat and protein with a standard error of 0.07% and 0.06% respectively. The spectrometer used in this trial is too large and is not a commercially feasible option.

Another study by Woo et al. (2002) used the spectral range of 700 to 1100 nm to predict fat, protein and lactose. Their study achieved a low standard error of prediction of 0.06, 0.10 and 0.10% for fat, protein and lactose, respectively. Once again this study used a laboratory analyser which is not commercially feasible.

Alternatively near infrared analysis has been used to monitor rennet-induced coagulation for cheese production for the 800 – 2500 nm spectrum by Klandar et al. (2007). NIR spectroscopy was a suitable non-destructive, non-intrusive and on-line method to evaluate the rennet induced coagulation of milk. However, the cost of the analyser limits it use in a commercial operation.

JDSU has recently developed a lightweight, robust, and portable spectrometer series which may solve these problems. The two spectrometers used in this thesis were the MicroNIR1700 and the MicroNIR2200. The MicroNIR1700 spectrometer covers 880 – 1660 nm, and the MicroNIR2200 covers the 1130 – 2160 nm spiral region.

Testing the performance of these MicroNIR spectrometers is an exciting development. Currently herd testing is done only a few times a year. The MicroNIR spectrometers may have the potential to provide farmers with daily information on the productivity and health of the individual animal. Thus giving instant feedback and improving the productivity of the farm.

The major goal of the work presented in this thesis is to determine if the MicroNIR spectrometers can meet International Committee for Animal Recording (ICAR) standards for an on-line milk analyser. Included in this thesis are six chapters and appendices. Chapter two gives background information on electromagnetic radiation, Near Infrared Spectroscopy, milk composition, and optical properties of spectroscopy. Chapter three gives details on the spectrometers used for the trial, experimental methods, reference standards used, hardware changes needed, and data processing used. Chapter four describes the results obtained for the first laboratory trial, illustrates how the hardware modification have improved the spectrum and describe the results obtained during the farm trial. Chapter five discusses the results and compares the results with respect to relevant literature. The final chapter is a summary of the significant results obtained with the MicroNIR spectrometers

2 Background information

This chapter provides an overview of near infrared spectroscopy, milk composition, and optical properties of samples. The first subsection explains electromagnetic radiation. The second subsection provides background information on near infrared spectroscopy and the history of near infrared spectroscopy. The third subsection explains milk composition. The fourth subsection explains the optical properties of the samples.

2.1 Electromagnetic radiation

All light is classified as electromagnetic radiation and consists of alternating electric and magnetic fields. The important parameters are the wavelength (λ), frequency (υ), and the wavenumbers ($\bar{\upsilon}$, number of waves per unit length) and are related by (2-1)

$$\bar{\upsilon} = \frac{\upsilon n}{c} = \frac{1}{\lambda} \qquad \textbf{(2-1)}$$

Where *c* is the speed of light and *n* is the refractive index of the medium it is passing through. Radiation is emitted from a source in discrete units called photons where the photon frequency, υ, and photon energy are related by (2-2):

$$E_p = h\upsilon \qquad \textbf{(2-2)}$$

Where *h* is Planks constant (6.62610×10^{-34} Js). Photons of specific energy may be absorbed or emitted by a molecule resulting in a transfer of energy (Kenyon 2008).

2.2 Near Infrared spectroscopy

Spectroscopic methods of analysis are based on the interaction of electromagnetic wave with matter. Spectroscopic methods are applied over a wide range of the electromagnetic spectrum; these include X-ray, ultra violet, visible, infrared, microwave, and radio-frequency (Straughan & Walker 1976).

- X-ray region (wavelengths between 0.5 and 10 nm) is involved in energy changes of electrons of the internal layers of atoms and molecules.
- Ultraviolet region (10-350 nm) is involved in electronic transitions from the excited state to the ground state are observed. In this wavelength range, luminescence (fluorescence and phosphorescence) may also be observed.
- The visible region (350 – 800 nm) is another zone where electronic transitions occur. Molecules exhibiting a large number of conjugated double bonds absorb energy in this region.
- The near infrared (NIR) region (800 – 2500 nm) is the first spectral region exhibiting absorption bands related to molecule vibrations. This region is characterized by harmonics and combination bands and is widely used for composition analyses of food products.
- The mid-infrared (MIR) region (2500 – 25000 nm) is the main region of vibrational spectroscopy. This region retains information, allowing organic molecules to be identified and the structure and conformation of molecules such as proteins and lipids to be characterized.
- In the microwave region (100 µm – 1 cm) absorbed energy is related to molecular rotation.
- The radio frequency region (1 cm – 10 m) is the region investigated by nuclear magnetic resonance.

Infrared spectroscopy is one of the most common spectroscopic methods used by chemists. Infrared spectrometry provides a means of studying pure rotational, pure vibrational and rotational-vibrational energy changes in the ground state of simple and complex molecules (Straughan & Walker 1976). Vibrational movements of molecules induce absorption in the

infrared region. Infrared radiation can also excite rotational movements of molecules, giving rotational bands. These are often superimposed on the vibrational bands.

Near infrared spectroscopy is a non-destructive method for analysing the quality and composition of agricultural products. Near-infrared spectroscopy allows rapid, accurate, and non-destructive measurements of chemical components and can provide information about structural and physical properties of materials. When light hits a sample the incident radiation may be reflected, absorbed or transmitted and the relative contribution of each phenomenon depends on the chemical and physical nature of the sample.

NIR spectroscopy analysis is based on the principle that certain types of chemical bonds absorb specific frequencies of light energy in the NIR region. If a photons energy matches the molecular energy levels difference it will be absorbed and will raise the energy of a molecule from a ground state to a specific excited state as illustrated in Figure 1.

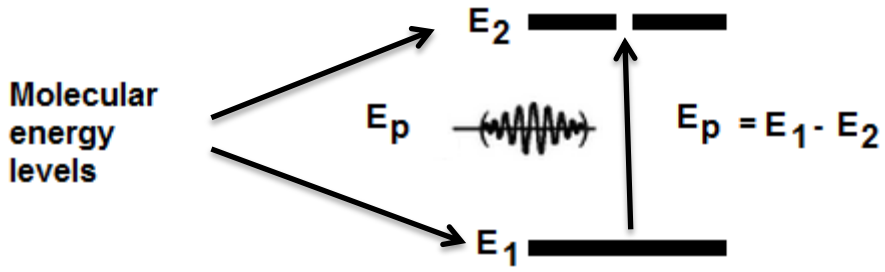

Figure 1: Absorption of electromagnetic radiation

Absorption bands in the NIR spectrum are overtones and/or combination bands of fundamental vibrations occurring in the mid infrared spectral region, which arise from rotational and vibrational movements of covalent bonds (Workman & Weyer 2008).

There are three different measurement modes NIR spectra can be obtained by: reflectance, transmittance, and transflectance mode. Transmission mode uses a light source to penetrate through the sample to the detector on the opposite side of the light source. Reflectance mode uses a light source and a detector placed in an orientation where the

reflected light from the sample can be measured. The transflectance configuration is a hybrid configuration of the transmission measurement and reflection measurement configurations. The transflectance configuration uses a thin flow cell with a diffused reflector behind the flow cell to reflect any transmitted light through the sample. The transflectance configuration increases the received signal by decreasing the optical path length which decreases the attenuation. The different measurement configurations are illustrated in Figure 2

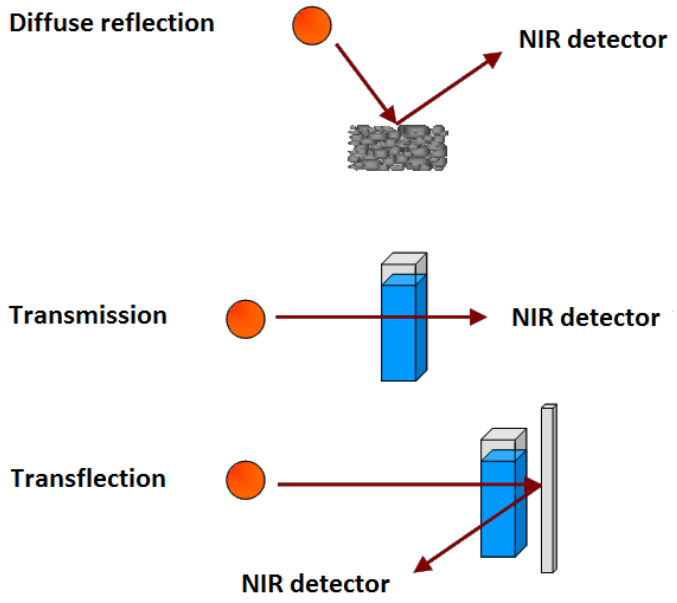

Figure 2: Different spectroscopic measurement configurations (Roos 2003).

Atoms are constantly vibrating in molecules. The oscillation of the atoms is 10^{-11} – 10^{-12} m and their vibrational frequencies are in the infrared region. Every molecule is able to absorb certain characteristic wavelengths of electromagnetic radiation. During this process electromagnetic radiation is transferred to the molecule and the intensity of the radiation decreases. The electrons in the atom become excited and are set into a natural vibrational motion. The excited electron interacts with the neighbouring atom to convert the vibrational energy into heat.

To absorb infrared light from a molecule a change in the dipole moment must occur during the normal mode of vibration. A change in dipole

moment is caused by a temporary shift in the electrical polarity of the molecule. A temporary shift in electrical polarity of a molecule is caused by stretching and bending of the molecular bonds. Stretching involves movement along the axis of the bond which causes the interatomic distance to increase. Bending causes a change in the bond angle between bonds (Marcelli et al. 2012). The two different types of stretching and four different types of bending are illustrated in Figure 3.

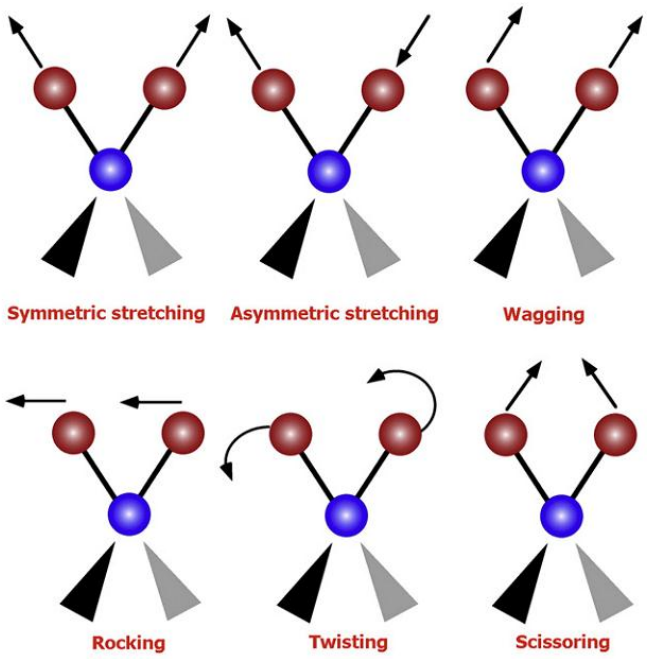

Figure 3: Diagram illustrating two different types of stretching and four different bending variations are shown (Marcelli et al. 2012).

There are two laws that govern basis of near infrared spectroscopy, “Hooke’s Law and the “Franck-Condon principle. Fundamental vibrational frequencies for diatomic molecules in the mid IR region can be calculated using Hooks law.

As a rough estimate, the vibrational movements of two atoms of a diatomic molecule can be considered to be like the compression and extension movements of a spring. The simplest model corresponds to a mass, *m*, bound to a spring with no mass. This model is defined by the strength constant, *k*, measuring the spring tightness, the displacement of the molecule, $q = r - r_0$, and the displacement of the molecule from its equilibrium position, *r.* If the spring responds to Hooke’s law, the strength,

f, applied to the particle is proportional to the molecule movement according to (2-3)

$$f = -k(r - r_0) = -kq \quad \textbf{(2-3)}$$

From equation (2-3) is possible to calculate the potential energy and using Newton's equations, the vibrational frequency, v, can be determined using equation (2-4)

$$k = 4\pi^2 v^2 m \quad \textbf{(2-4)}$$

The frequency is only dependent on the strength constant, *k*, and the mass *m*. This simple model can be improved by using quantum mechanics. Vibrational energy is quantified and can be calculated from the Schrödinger equation. The harmonic model can be used to calculate the stretching vibration observed in a spectrum (Dufour 2009).

2.2.1 History of infrared instrumentation

Infrared spectroscopy was the first spectroscopic technique used by organic chemists. Early interest in infrared technology and instrumentation started in the nineteenth century, when William Hershel first observed infrared radiation on a photographic plate. Herschel's experiments were aimed at filtering heat from a telescope but along with these developments he demonstrated that light radiation exists beyond the visible spectrum.

In the 1930's the first custom made IR instrument appeared in industrial laboratories. During World War II, the development of lead sulphide (PbS) detectors in the IR region made it possible to record IR spectra. The advance in technology revived scientific interest in IR and renewed research into IR spectroscopy.

The Becman IR-1 was the first commercial infra-red instrument and was marketed in 1942. It was used in the rubber industry and petroleum fields for quality control. The Becman IR-1 was both an infrared and visible spectrophotometer, but it was not widely sold because of governmental restrictions. The IR-1 used spherical mirrors and had long focal lengths and optical paths to affect dispersion. In 1944 Perkin-Elmer Inc. introduced

the Model 12 infrared spectrometer that used aspherical shaped mirrors, making this a more compact instrument with higher resolution due to smaller focal lengths and shorter beam paths. The Model 12 was a single beam spectrometer that was designed for point-by-point measurement and required highly trained personnel for routine operation.

In 1946 American Cyanamid Stanford laboratories worked with Perkin-Elmer Inc to begin development of the first commercially available infrared instrument. The Perkin-Elmer 21 was the first widely adopted double beam IR instrument sold in 1947. The double beam optical system was a flexible system which made it a popular analytical instrument. The spectrometer was widely adapted in almost every branch of analytical chemistry including applications in food and agricultural industry. The explosive growth of infrared spectroscopy between 1950 and 1960 was caused by the development of commercially available instruments.

One of the first applications of milk analysis with an Infrared milk analyser was done by Biggs (1967). Milk samples were homogenised and analysed using a conventional double beam spectrometer, modified for rapid wavelength changes. Precision of measurement was ± 0.03%, and standard deviation between infrared and chemical methods was ± 0.06 and ± 0.07 for fat and protein respectively.

During the 1970s and 1980s the application of near infrared spectroscopy for agricultural applications rapidly developed through government and university research programs. The first official near infrared analysis method was adopted by the Canadian grain commission for measuring the protein content of wheat.

In the 1980s simple stand-alone NIR instruments for chemical analysis appeared in the market propelled by the introduction of fibre optics. Improved monochromators and detectors made NIR a more practical technique for the applications in the agricultural and food science.

Technological advances in both hard ware and software design, has seen the development of more compact, portable, cheaper, and robust NIR and MIR spectrometer systems. The improvements in microprocessors have

decreased the time needed to analyse spectral data. Currently NIR and MIR have become so popular that they have been found practical applications in virtually all branches of agricultural and food industries (Lin et al. 2009).

2.3 Milk Composition

Milk is an extremely complex biological fluid with lots of nutrient chemicals. Milk is a particularly good source of protein, calcium, phosphorus and the vitamins A, B_1 and B_2. In addition the fat and lactose provide readily available sources of energy. Milk is produced naturally by mammals to provide nutrition and immunological protection to their young. The composition of milk is affected by many factors; these include breed, feed type, seasonal, and geographic variations.

Milk is the product of the mammary gland. The udder of a cow consist of four mammary glands, each drained by its own teat and made up by a variety of structures and tissues. The inside lining of the teats and ducts of the mammary gland are essentially modified skin. Milk originates in the secretory tissue and collects in a series of ducts, which increase in size as milk moves towards the teat as illustrated in Figure 4. The lipid phase is synthesized in the endoplasmic reticulum and collects as droplets on the cytoplasmic side of the membrane. The droplets move to the apical plasma membrane and pass into the lumen by pinocytosis. Synthesis is completed in the alveolar lumen where lactose and proteins are synthesized. When the circulating pituitary hormone, oxytocin, is bounded to the myoepithelial cells the alveolus contracts and the milk is expelled from the lumen into the duct system(Holmes & Wilson 1987).

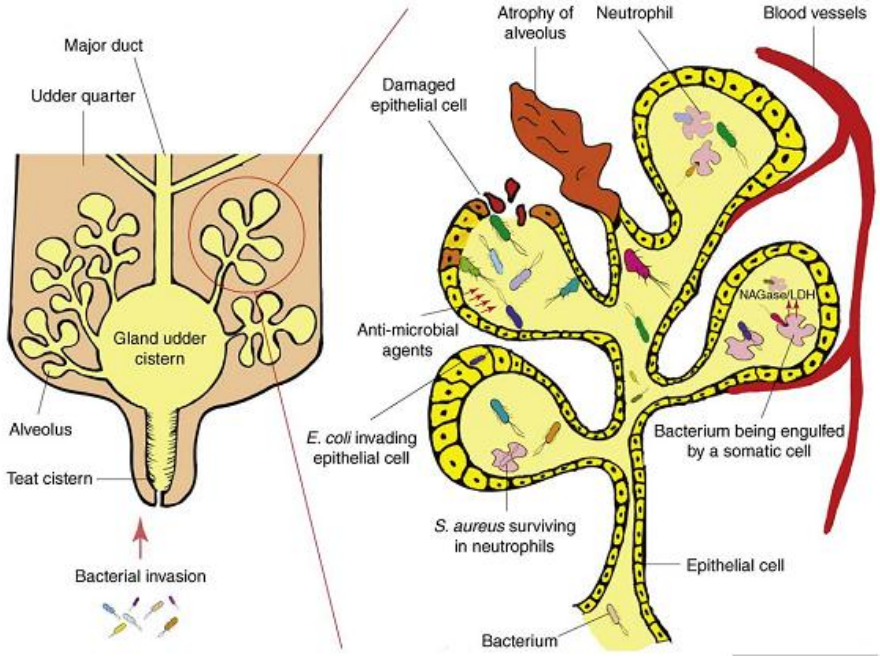

Figure 4: Structure of the udder of a cow (Viguier et al. 2009).

2.3.1 Lactose

Lactose is the principal carbohydrate of milk and primarily provides the young animal with energy and gives milk a sweetish taste. Lactose is the major solid constituent of milk. The lactose content of milk varies between 4.2% and 5.0%. The concentration of lactose is generally lower late in lactation or in cows suffering mastitis. Lactose is a disaccharide and comprises α-D-glucose and β-D-galactose molecules. Three solid forms of lactose exist, α-lactose monohydride and anhydrous α- and β-lactose. Lactose is one of the least soluble of the common sugars, having a solubility in water of only 17.8% at 25°C. The low solubility of lactose causes crystallization to occur in ice-cream which causes a sandy texture (Fox & McSweeney 1998).

Lactose is use as a food ingredient to stabilise protein. Lactose may also be used as a partial replacer for sucrose in icings and toppings to improve mouth feel without excessive sweetness. Lactose is an essential ingredient in cheese manufacture and the production of fermented milk products. It involves the conversion of lactose to constituent sugars of water soluble organic acids by lactic acid bacteria through a biochemical reaction called glycolysis. About 98% of the lactose in milk is lost in whey during cheese production, and only a relatively small amount of lactose stays with cheese (Upreti et al. 2006).

2.3.2 Lipids

Lipids are esters of fatty acids and related components that are soluble in nonpolar organic solvents and nearly insoluble in water. Alternatively, the term fat is used but fat is usually considered to consist of a mixture of triglycerides. Nearly all of the fat in milk is in globules and can be concentrated by the means of gravity creaming. Triacylglycerol's, are dominant and constitute 98% of the milk lipids, together with small amounts of di- and monoacylglycerols and free fatty acids. The main purpose of fat in milk is to provide lubrication for the suckling offspring. Fat is secreted by specialized cells in the mammary glands. The fat globules range in diameter from 0.1 µm to 20 µm with a mean diameter of 3.5 µm (Fox & McSweeney 1998).

The presence of fat is an important factor in determining the palatability of food. Dairy fats are particularly palatable due to a large number of lipids of small molecular size, short chain fatty acids and their derivatives which contribute to flavour, aroma and feel. Milk fat contains substantial quantities of the vitamin A precursor β-carotene. Milk fat is also a minor, but significant source of vitamin D for pregnant and lactating women and children who have a high dietary requirement.

The fat content of milk decreases during the first 4-6 weeks after parturition and then increases steadily throughout the remainder of lactation. Fat content is generally higher in the winter and is lowest in the summer, due partly to the effect of environmental temperature.

The quarters of a cow's udder are anatomically separate and secrete milk of markedly different composition. The fat percentage of the milk varies at different stages of milk removal. The foremilk has very low fat percentage but as more milk is removed the fat concentration rises. This is caused by the milk fat droplets clustering together and being partially trapped in the alveoli. The partially trapped fat will be expressed at the subsequent milking, and will lead to an artificially high fat content (Holmes & Wilson 1987).

2.3.3 Protein

Normal bovine milk contains about 3.5% protein. Milk proteins are of two distinct types, whey proteins (serum proteins) and caseins. Caseins constitute over 80% of the total protein of milk and whey constitutes the remaining 20%. The ratio of whey proteins to casein changes accordingly to the stage of lactation. The whey protein content of milk is highest in the first few days after calving and towards the end of lactation. The protein content of milk can be used as an indicator for the energy supply of the cow. Low milk protein content indicates deficiencies in energy intake (Varnam & Sutherland 2001).

The energy content of the diet has a major effect on the casein content of milk. Other types of protein present in milk in small quantities are albumin and globulins. These are transferred directly from blood into milk. Caseins have an essential amino acids required for the development of the muscular tissue. There are two types of whey products, β-lacto globulins and α-lactalbumins, proteose-peptones and small quantities of the blood derived proteins, serum albumin and immunoglobulins.

Caseins are highly digestible in the intestine and are a high quality source of amino acids. This high quality protein in cow milk is one of the key reasons why milk is such an important human food. Most whey proteins are relatively less digestible in the intestine, although all of them are digested to some degree.

Centrifugation of skim milk results in pelleting of the casein in whey. Whey is also called the serum phase of milk and contains water, lactose, and soluble proteins. Casein molecules can also be separated from the whey by precipitation of the casein with acid or by disrupting the micellar structure by partial hydrolysis of the protein molecules with a proteolytic enzyme. Partial hydrolysis of protein is used for the production of cheese.

The five main classes of caseins in in milk are α_{s1}-, α_{s2}-, β-, ɣ, and κ-caseins. The relative proportions of α-, β- and ɣ-caseins are subjected to genetic variation within individual herds and there can be significant differences in casein composition of milk from different cows. The casein composition of bulk milk varies very little at any single stage in lactation.

The caseins are globular proteins and have an amino acid content similar to that of other types, although cysteine is present, in small quantities, in only α_{s2}-, and κ-casein (Fox & McSweeney 1998).

Casein is extremely heat stable at temperatures up to 140°C but whey proteins are relatively unstable with extensive denaturation occurring at 80°C. Denaturing is accompanied by extensive breaking and randomization of the disulphide bonds (Varnam & Sutherland 2001).

2.3.4 Minerals

The minerals in milk consist principally of chlorides, citrates and bicarbonates of calcium, magnesium, potassium and sodium. All of the minerals are distributed between the soluble phase and the colloidal phase. The distribution of calcium, citrate, magnesium and phosphate between soluble and colloidal phases and their interaction with milk proteins have important consequences for the stability of milk. Milk is an important source of dietary calcium and is a key factor in determining healthy bone and tooth development in the young and adequate intake is essential (Varnam & Sutherland 2001).

2.3.5 Changes in milk composition

The composition of milk changes through the season and also during milking. Initially when the animal gives birth it secretes colostrum. Colostrum contains antibodies to protect the new-born against disease which ensures the offspring begins life in good physical condition. Fat and protein levels initially decrease during the first five weeks of lactation and then steadily increase over the lactation period. The concentration of lactose decreases progressively and significantly during early lactation. Then the lactose concentration steadily increases through the season. The typical changes in milk composition during lactation are illustrated in Figure 5.

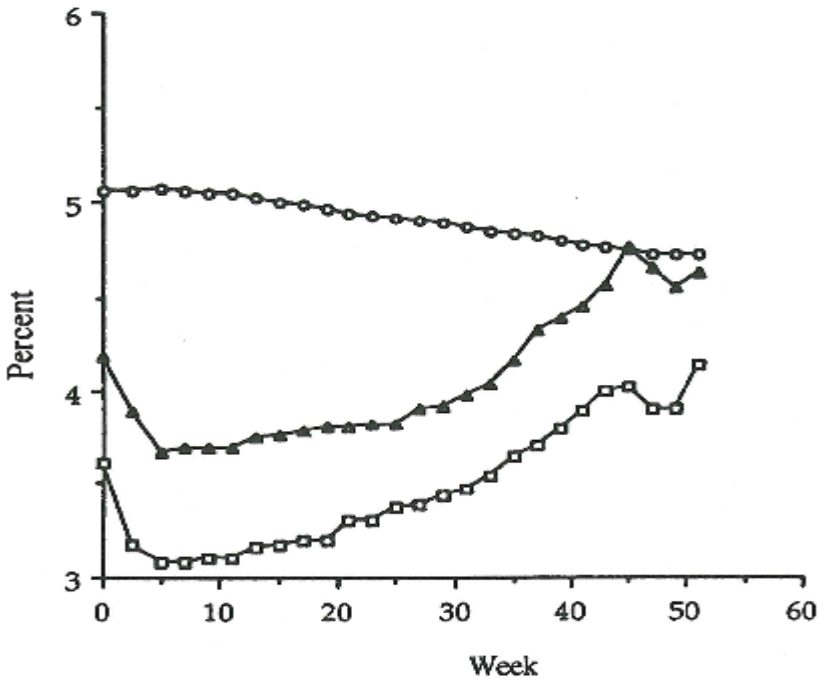

Figure 5: Changes in the concentration of fat (∆), protein (□), and lactose (O) during lactation (Fox & McSweeney 1998)

Towards the end of lactation, the number of active alveolar secretory cells slowly decline, reaching a minimum during the early dry period. The alveolar cells collapse and the space within the alveolus disappears and the udder consists of a greater proportion of connective tissue. New secretory tissue is laid down when the cow is ready for the next calving which increases the yield produced. If a cow is not dried off at all, the next lactation yield may be 20 – 30% lower.

The composition of milk also changes throughout milking. The milk fat percentage of the milk varies at different stages of milking. The foremilk has a very low fat percentage, one to two per cent, but as more milk is removed the fat concentration rises. Stripping's contain 5 to 10 per cent and residual milk 10 to 20 per cent. In bovine milk the milk fat droplets tend to aggregate together to form clusters which are large relative to the diameter of the fine ducts. These clusters tend to be held back until the final milk from the udder is removed (Holmes & Wilson 1987).

2.3.6 Effect of Mastitis

Mastitis simply means 'inflammation of the udder'. Most farmers associate mastitis with an inflamed quarter together with a change in the appearance of milk. These changes are due to the effect of the cow's inflammatory

response to infection. Bovine mastitis is a major disease that causes economic losses to dairy industry. Mastitis causes decreased milk production and reproductive disorders in dairy cows. Detection of clinical mastitis is relatively easy but subclinical mastitis is difficult to detect due to the absence of any visible sign. An increased number in somatic cell count (SCC) in the milk is the only evidence that helps in the diagnosis of subclinical mastitis. SCC measure can vary from morning milking to evening milking as well as throughout the milking. Mastitis not only results in poorer milk quality but also causes loss of productivity, increased financial cost and increased culling rates. The international market led by the European Community requires bulk milk somatic cell count to be less than 400,000 cells/ml (Madouasse et al. 2010).

Mastitis leads to changes in milk composition due to damage of the secretory cells in the mammary gland. This leads to a significant decrease in lipids, lactose and casein. This is of great significance to dairy manufacturers as it reduces the manufacturing yield from milk. Mastitis causes a decrease in the concentration of lactose which then causes an increase in sodium chloride levels to maintain the osmotic pressure of the milk. This is one of the main causes of the bitter and slightly salty taste of mastitic milk. Mastitis also produces increased levels of the enzymes lipase and plasmin which break down milk fat and casein. (Blowey & Edmondson 2010).

2.3.7 International committee for animal recording

The International Committee for Animal Recording (ICAR) is the world-wide organization for the standardization of identification, performance recording and evaluation of farm animals. ICAR establishes the accepted tolerances for monitoring the composition of milk. Online analysers allow more frequent milk analysis. Therefore a lower accuracy can be acceptable with more frequent analyses since the uncertainty can be reduced by averaging. The precision limits for an on-farm, in-line milk analytical device is a standard deviation of 0.25 g/100 g for fat and protein as illustrated in Table 1.

Table 1: Precision and accuracy limits for milk analysers (ICAR 2012).

Accuracy	Range	Standard deviation	Bias
Fat	2.0-6.0 g/100g	0.25 g/100g	0.13 g/100g
	5.0-14.0 g/100g	0.25 g/100g	0.25 g/100g
Protein	2.5-4.5 g/100g	0.25 g/100g	0.13 g/100g
	4.0-./0 g/100g	0.25 g/100g	0.25 g/100g

The accuracy limits stated in ICAR (2012) are given in standard deviation. Standard deviation is a measurement of the deviation away from the mean. The standard deviation does not include bias away from the true values. Due to the calibration being built from its own sensor there will be minimal bias. To analyse the performance of spectrometers root mean squared error was used. Root mean squared error measures the deviation from the true value. The rationale behind these measurement tolerances is the need for the analytical device to be able to identify compositional changes outside of the regular natural variation. Therefore the accuracy of the analytical device should be better than the natural day to day fluctuations (ICAR 2012).

2.4 Optical Properties of Samples

Electromagnetic radiation exists as waves that are propagated from a source and move in a straight line if they are not reflected or refracted. An electromagnetic wave transports energy through empty space using self-propagating electric and magnetic fields. Light traveling through matter interacts with the atoms and molecules, which results in absorption, scattering and dispersion.

2.4.1 Refractive index

The refractive index describes how light, or any other radiation propagates through that medium. When an electromagnetic wave travels through a dielectric medium the oscillating electromagnetic wave polarises the molecules of the medium at the frequency of the light. This causes the molecular dipoles to be coupled to the field. The polarisation of the molecules in the medium causes the propagation speed of the electromagnetic wave to slow down. The stronger the interaction between the field and the dipoles, the slower the propagation of the wave (Avison 1989). The ratio of the speed of light in free space to its speed in a medium is called the refractive index of the medium (n). The refractive index equation is given in (2-5).

$$n = \frac{c}{v} \qquad \textbf{(2-5)}$$

where c is the speed of light in a vacuum and
v is the speed of light in the medium.

In milk, the presence of the emulsion of fat globules generates scattering phenomena on near infrared radiation. The fat globules have a different refractive index from that of the aqueous medium that surrounds them. The difference in refractive indices causes light to be reflected and refracted (Cattaneo et al. 2009).

2.4.2 Reflection and Refraction

Refraction of light is the bending of a wave when it enters a medium where its speed is different. When light passes from a medium that has a low refractive index to a medium that has a higher refractive index it causes the light to bend towards the normal line of the difference medium.

When light hits a medium which is optically different some or all of the incident radiation is reflected. Reflection is the change in direction of a wave front between two different media when the wave front returns to the medium that it originated from (Mansfield & O'Sullivan 2010).

2.4.3 Scattering of Light

Matter is composed of discrete electric charges which are electrons and protons. If an obstacle, which could be a single electron, an atom or a molecule, a solid or liquid particle, is illuminated by an electromagnetic wave, electric charges in the obstacle are set into oscillatory motion by the electric field of the incident wave. Accelerated electric charges radiate electromagnetic energy in all directions; it is this secondary radiation that is called the radiation scattered by the obstacle (Lorenzo 2012). The type of scattering is dependent on the ratio of particle size to the wavelength of light (Lipson et al. 2010). The ease at which light is scattered increases towards the shorter wavelengths.

Rayleigh scattering occurs when the particle size is typically smaller than one-tenth of the wavelength. Rayleigh scattering causes the sky to be blue because light from the sun gets scattered on neutral atoms such as oxygen and nitrogen molecules (Kenyon 2008). Skim milk also has a blue tinge to it due to the casein micelle scattering the shorter wavelengths.

Mie scattering occurs with particles that have a diameter that is similar or larger than the wavelength of the incident light. Mie scattering is not strongly wavelength dependent and is the reason why milk appears white. During Mie scattering there is interference between light scattered from different parts of the same scatter, and the phase delays between light traveling through different thicknesses of the scatter must be taken into account (Lipson et al. 2010).

Milk contains light scattering particles in the form of fat globules and protein micelles. The main types of scattering in milk are Rayleigh scattering and Mie scattering. When light passes through milk it is scattered due to media fluctuations in density due to the fat globules which have a different refractive index compared to the suspended micelles (Cattaneo et al. 2009).

2.4.4 Water

In the infrared range the most dominant absorber in biology is water. The water molecule may vibrate in a number of ways. In the gas state, the vibrations involve combinations of symmetric stretching (v_1), asymmetric stretching (v_3) and bending (v_2) of the covalent bonds. The water molecule has a very small moment of inertia on rotation which gives rise to rich combined vibrational-rotational spectra in the vapour. In liquid the rotations tend to be restricted by the hydrogen bonds. Also the spectral lines are broader causing overlap of the absorption peaks. Absorption of light in the visible region is very low but is significant enough in the red part of the visible spectrum; this causes water to appear blue. In the MIR and NIR regions the water molecule strongly absorbs light. There are three water absorption bands that affect the near infrared spectrum of milk. The first two absorption peaks are located at 970 nm and 1440 nm relate to the first and second overtone vibrations of O-H. The water absorption band at 1440 nm is assigned to the combination of OH symmetric and antisymmetric stretching modes of water. The third absorption peak is located at 1940 nm and is a combination of the asymmetric stretch and bending of the water molecule. A plot of the near infrared water absorption bands is illustrated in Figure 6.

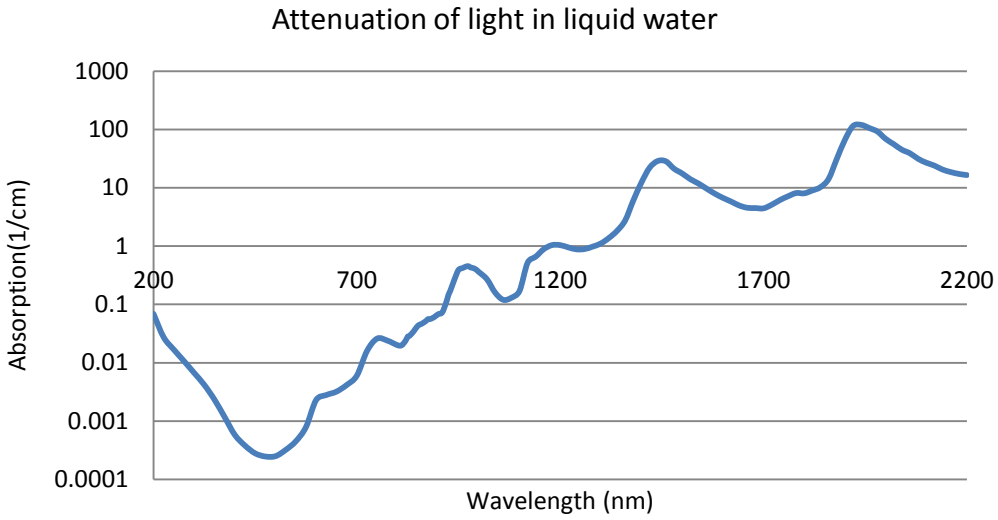

Figure 6: Visible to near infrared water absorption (Querry & Hale 1972).

In liquid water there are weak, broad combination bands near 1200 nm 1780 nm and a second set near 740 nm and 840 nm (Workman & Weyer

2008). Table 2 list the main water absorption bands for liquid water which influences the spectrum of milk.

Table 2: Significant water absorption peaks for the spectrum of milk (Yakovenko et al. 2002).

Wavelength	Assignment
1940 nm	$av_1 + v_2 + bv_3; a + b = 1$
1440 nm	$av_1 + bv_3; a + b = 2$
1200 nm	$av_1 + v_2 + bv_3; a + b = 2$
970 nm	$av_1 + bv_3; a + b = 3$
Note that a and b are integers, ≥ 0	

The water content in milk is about 87% and significantly influences the spectrum of milk. Water is also a significant source of hydration for the offspring. Without water the secretion of milk would be extremely viscous, because the milk is mostly lipids and proteins otherwise. The water content of milk is dependent on the synthesis of lactose. Water provides the aqueous medium for the suspension of organic components.

3 Hardware and methods

This chapter describes the sensing technology and methods used in this project. It covers the hardware used for the experimental work. The reference standard used for determining the composition of milk is discussed. The laboratory trial method and farm trial method are described. The final section describes how the data was processed to build a calibration model and asses the performance of the model.

3.1 Sensing technology

Two different spectrometers were used in this trial. The spectrometers were manufactured by JDSU and cover two different spectral ranges. The MicroNIR1700 spectrometer covers a wavelength range of 870 – 1660 nm. The MicroNIR2200 spectrometer covers the extended spectral range of 1130 – 2160 nm. Both spectrometers have 128 pixels with a spectral resolution of 1.25% of the centre wavelength. The MicroNIR1700 spectrometer records the spectrum at 6 nm intervals and the MicroNIR1700 spectrometer records at 8 nm intervals. Information on the specifications of the spectrometers is given in section 8.1 of the appendix.

The spectrometers use an Indium Gallium Arsenide array (InGaAs). The InGaAs detectors offer higher sensitivity with lower cooling requirements than other semiconductor detectors operating in the NIR band. Normal InGaAs arrays are grown on an InP substrate, a ratio of Indium to Gallium, 53:47 ($In_{53}Ga_{47}As$) matches the substrate structure, and resulting band gap limits the wavelength response. This is due to the low energy photons becoming transparent to the array. This is caused by the photon energy being lower than the band gap of array. The band gap of a normal InGaAs array limits the spectral response to 1700 nm.

To extend the cut off wavelength requires the band gap of the InGaAs array to be lowered. To lower the band gap of the InGaAs array the ratio of Indium: Gallium is changed to $In_{74}Ga_{26}As$. Lowering the band gap of the InGaAs array moves the materials absorption edge out to 2.2 microns. Changing the stoichiometry increases the lattice mismatch between the InGaAs and the InP substrate. This lattice mismatch combined with the

decrease in the band gap leads to an increase in the detector dark current by over a thousand times (Malchow et al. 2008).

The spectrometers use a linear variable filter as the dispersing element. The linear variable filter allows light of a specific wavelength to pass through the filter and onto the array. The spectrometers also have two integrated vacuum Tungsten lamps with a life time of 18,000 hours which are powered through the USB port. The MicroNIR1700 spectrometer is shown in Figure 7.

Figure 7: MicroNIR spectrometer.

The spectrometers are controlled and powered by a PC through a USB cable on a computer. The spectrometers are compact, lightweight, portable, and are relatively cheap for the technology.

3.2 Reference standard

Milk samples were analysed for fat, protein, lactose and somatic cell count at Testlink Laboratory, Riverlea Rd, Hamilton. To determine fat protein and lactose the MilkoScan FT6000 is used to measure the mid-infrared spectrum. Milk samples are homogenised and are placed into a cuvette where the absorption spectrum of the milk sample is measured. The MilkoScan™ performs a Fourier transform on the transmission spectrum to analyse milk composition. Specific Infra-red wavelengths have been used to correlate absorption bands and reflection bands to milk composition. By measuring absorbance of the sample it is possible to accurately measure fat, protein, and lactose of the sample. The measurement accuracy of the

MilkoScan™ is 0.8% coefficient of variation for fat, protein and lactose (Foss 2006). The MilkoScan™ is the accepted reference method for herd testing.

3.3 Laboratory trial hardware

The first objective of this project was to assess if the MicroNIR spectrometers can measure the composition of milk. To assess the spectrometer ability to predict milk composition an optical testing table was built to analyse the spectrum of milk with the MicroNIR spectrometers.

Both spectrometers that were used in the trial were in their original configuration as illustrated in Figure 8. A schematic diagram illustrating the laboratory trial hard ware is illustrated in Figure 9.

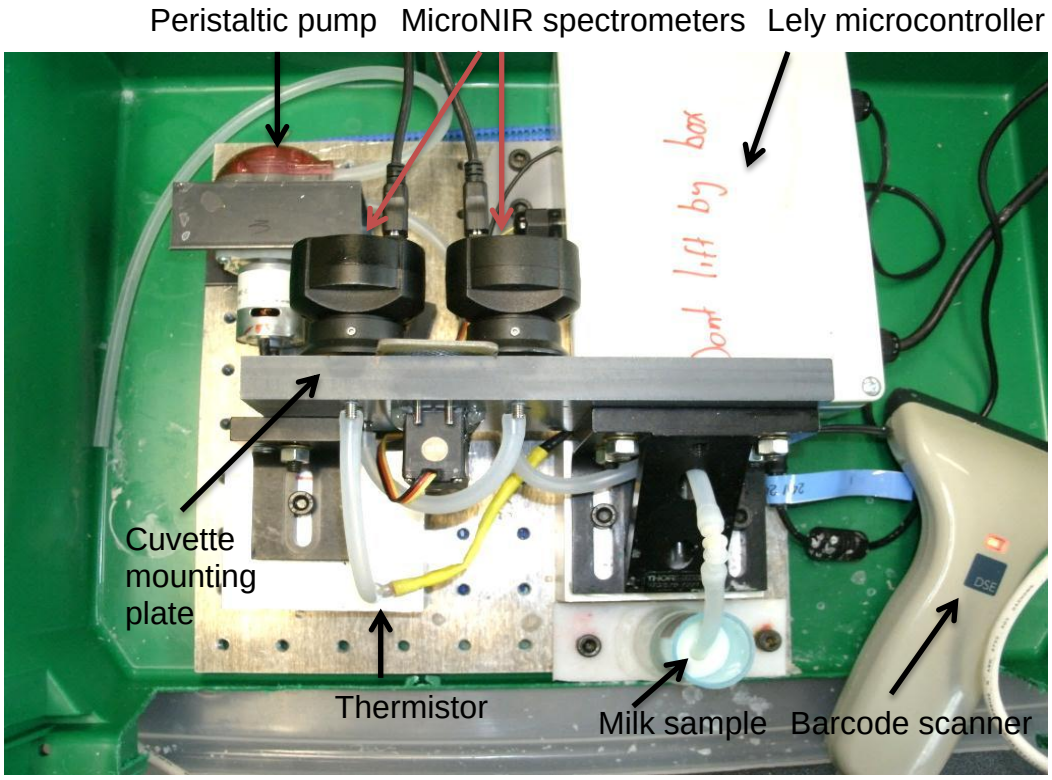

Figure 8: Laboratory trial experimental setup.

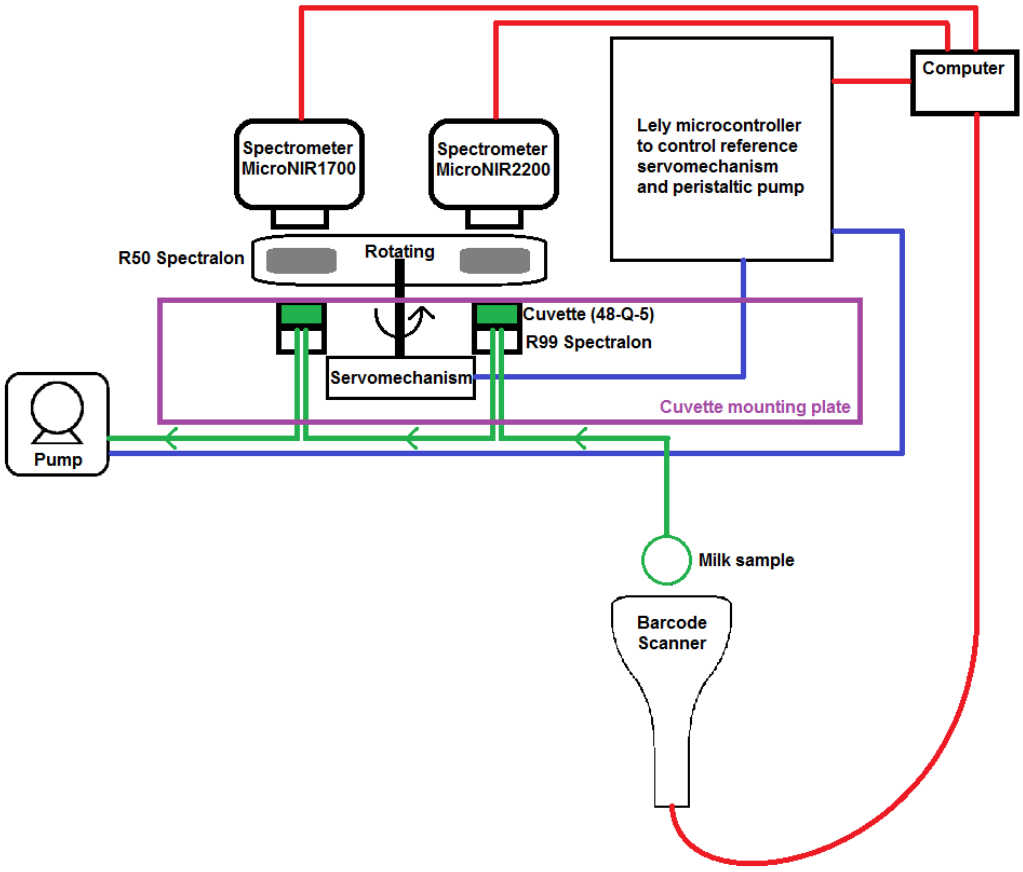

Figure 9: Schematic diagram of laboratory trial hardware.

Milk samples were analysed for fat, protein, and lactose. Two Starna 48-Q-5 (Starna Scientific Ltd, London, United Kingdom) flow cells were used to obtain reflectance spectrum of the milk samples. Samples were rapidly heated in a water bath at 45°C for 5 minutes.

Reflectance spectrum was measured with an integration time of 500 µs and was averaged over 1000 scans. Dark current and reference scans were taken between samples to calibrate out light fluctuations. The spectrum was normalised against spectralon R50% reference tiles. The barcode number of the milk samples was recorded, and the milk composition results were matched up with the spectral scans from each spectrometer.

3.4 Optimisation of hardware

The following sections describe the necessary hardware changes that were needed to meet ICAR accuracy for an online sensor. The laboratory testing identified that there was insufficient spectrum signal from the standard spectrometers. This was caused by the low integration time and the high dark current (see section 5.1 for further details). To improve the capability of the spectrometers to predict milk composition the hardware needed to be modified. This section describes; how the lamp voltage was increased, the method used to lower the operating temperature, the temperature control circuit used to maintain the spectrometer at a constant temperature, and the transflectance measurement configuration.

3.4.1 Increasing the lamp voltage

To increase the light output an independent electronically controlled power supply was built to power the two tungsten lamps. The electronically controlled power supply had to deliver a regulated supply voltage that was controlled by the spectrometer. To do this a transistor was used to amplify the lamps drive signal from the spectrometer to drive a relay as illustrated in Figure 10. The relay was used to connect the regulated power supply to the tungsten lamps. The lamps operating voltage was increased from 2.5 V to 4.8 V to increase the spectrum signal. Section 4.2.1 illustrated the increase in spectrum intensity for both spectrometers.

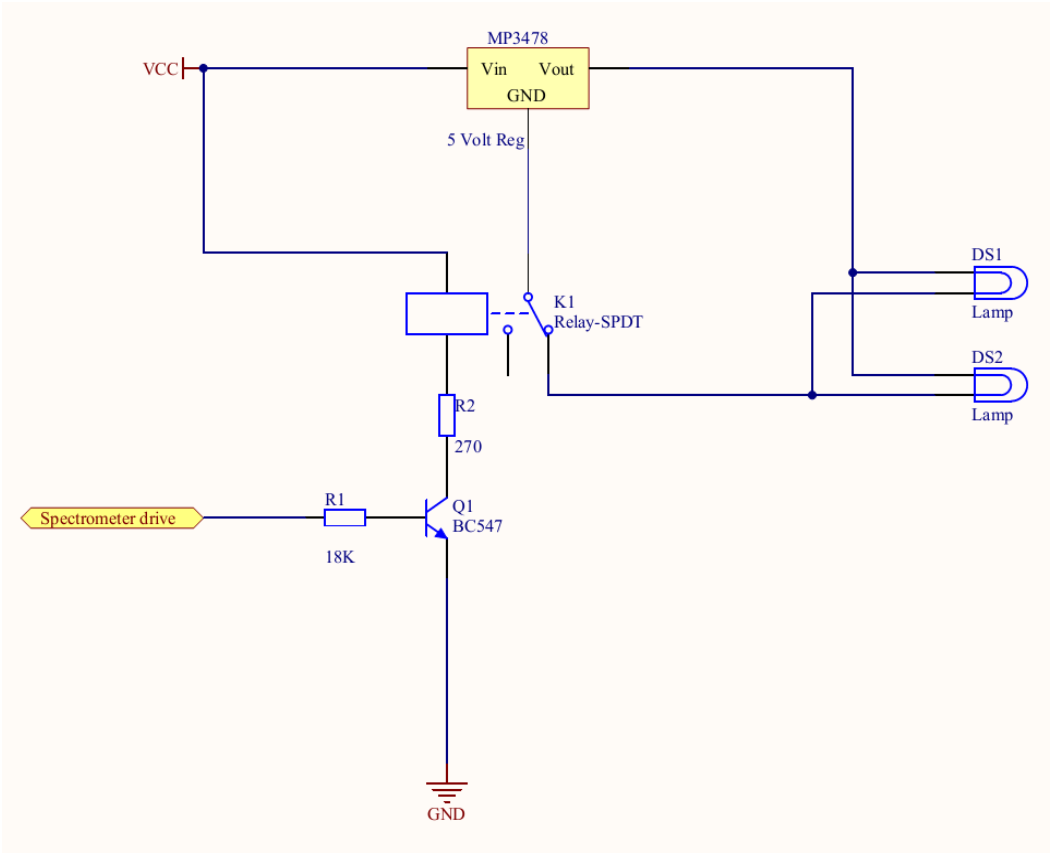

Figure 10: Spectrometer drive circuit

Increasing the lamp voltage of the spectrometers caused the operating temperature of the spectrometers to increase from 42°C to 50°C. Due to the compact nature of the spectrometer the temperature of the detector also increased. This caused the dark current to increase excessively. The high dark current caused distortion in some of the pixels which resulted in distorted spectra. The following section describes how the spectrometers temperature was decreased to reduce the spectral noise.

3.4.2 Reducing the spectral noise

The major source of noise in the spectrometers is the dark current noise. Dark current is caused by thermal excitation across the semiconductor band gap (Kenyon 2008). The rate of dark current accumulation is dependent on the temperature of the InGaAs array. Cooling the spectrometer significantly reduced the base line dark current which helps improve the detectivity of the spectrometer. Cooling the spectrometer causes the spectral response range to shift to the shorter wavelength decreasing sensitivity of the spectrometer to the longer wavelengths as illustrated in Figure 11. The decrease in sensitivity is compensated by the

large decrease in dark current as the temperature decreases thus improving the detectability of the spectrometer.

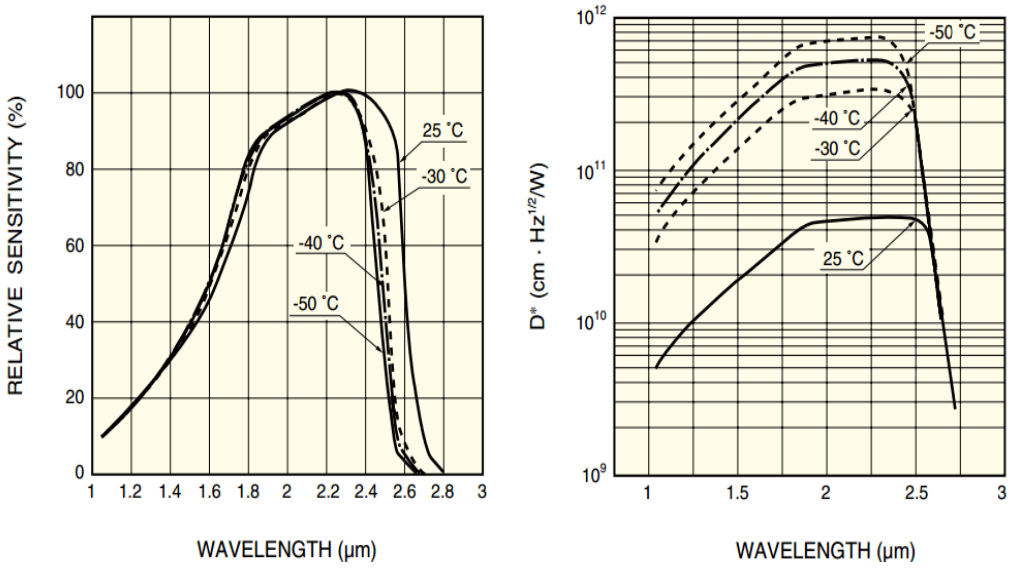

Figure 11: As the temperature of the extended spectrometer is decreased the relative sensitivity of the spectrometer decreases but the detectivity increases due to a decrease in dark current (Hamamatsu 2004).

To reduce the dark current noise and minimise the distortion of the spectra, cooling of the spectrometer was required. Cooling was achieved using a thermoelectric device called a Peltier element. Peltier cooling is a solid state technology, whereby an electric voltage creates a temperature difference in a pair of dissimilar materials. A typical Peltier element is manufactured with two thin ceramic wafers and an array of p- and n- type blocks of doped semiconductor material sandwiched between them as illustrated in Figure 12.

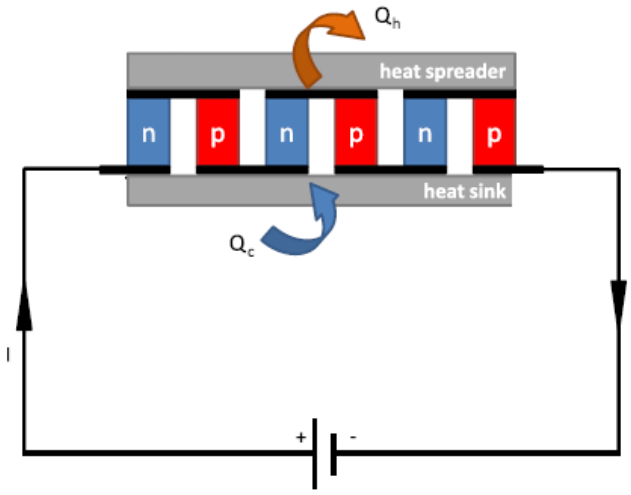

Figure 12: Schematic diagram of a thermoelectric cooler (Hermes & Barbosa Jr 2012).

The Peltier element used to cool the spectrometers was a TEC1-12708 from (Shenzhen, China) and is illustrated in Figure 13.

Figure 13: Thermoelectric cooler used to cool the spectrometers

To achieve smooth spectra it is essential to maintain the spectrometer at a constant temperature. A temperature control circuit was needed to monitor the temperature and adjust the amount of cooling done by the thermoelectric cooler. To monitor the temperature a thermistor was mounted on the case of the spectrometer with a thermally conductive epoxy as illustrated in Figure 14. This ensures an accurate temperature measurement of the array is achieved.

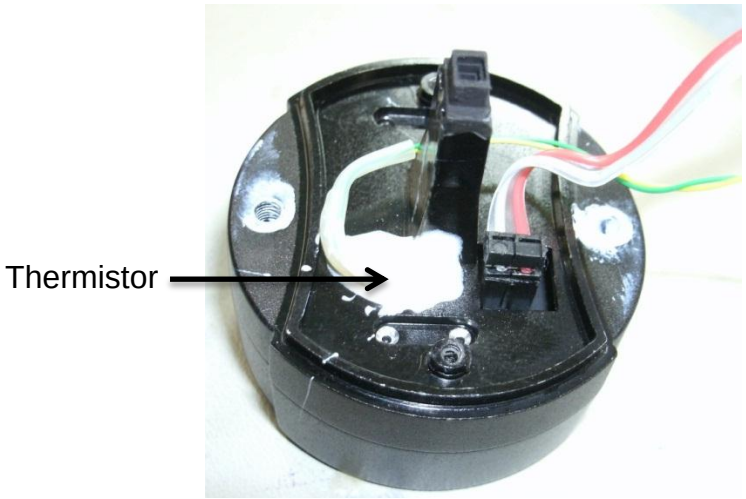

Figure 14: Thermistor mounted on the case of the spectrometer to control the temperature of the spectrometer.

The thermoelectric cooler was controlled with pulse width modulation (PWM) from the temperature controller circuit. The temperature controller circuit reads the temperature of the spectrometer and adjust the amount of cooling by varying the duty cycle of the PWM. The temperature control circuit monitors what the temperature has been in the past and compares

it with the target temperature and adjust the duty cycle accordingly to the proportional and integral gains. To control the temperature of the spectrometers a program written in C was used (See section 8.2 of the appendix). A schematic of the temperature control circuit is shown in Figure 15.

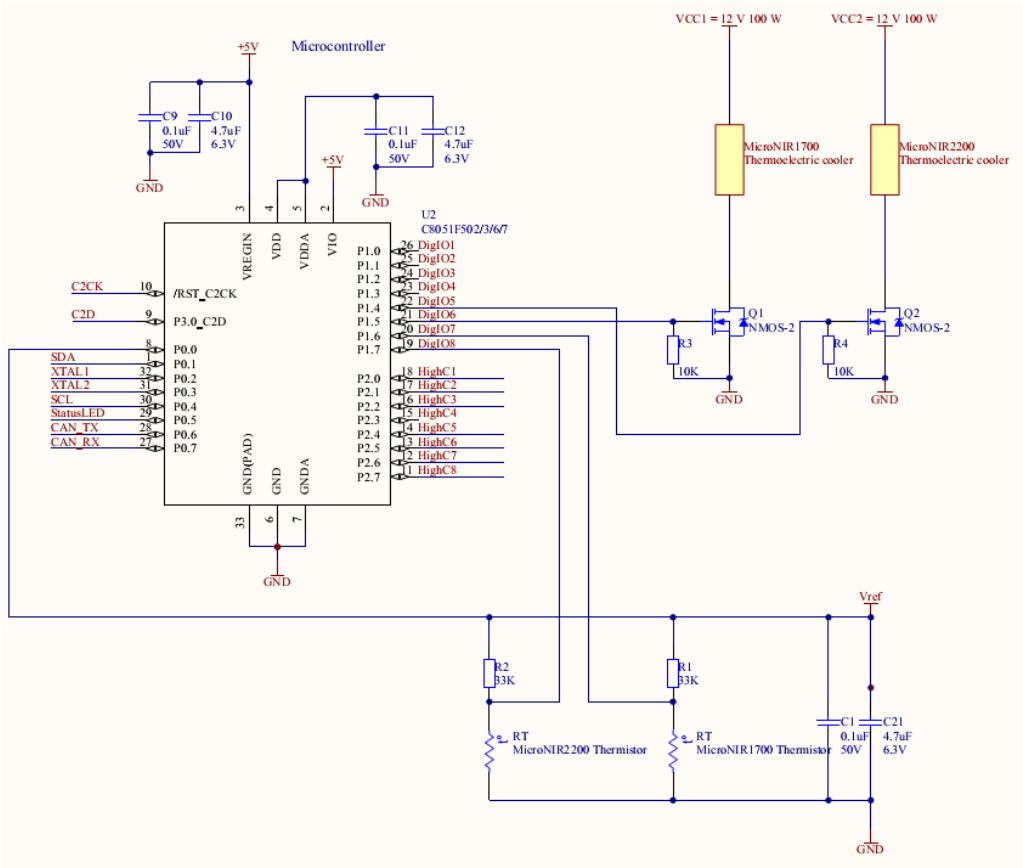

Figure 15: Temperature control circuit.

The cooling efficiencies of thermoelectric cooler are very poor and require a good heat sink with low thermal resistance to remove the thermal energy from the spectrometer. The thermoelectric cooler used was capable of transporting 71 W of heat using 131 W of electrical energy. The hot side of the Peltier element will have to dissipate 202 W. To prevent the Peltier element from overheating and to improve the efficiency a heat sink with low thermal resistance is required.

To provide adequate heat transfer, a water cooled heat sink was used to cool the thermoelectric cooler as illustrated in Figure 16. The water cooled heat sink transfers thermal energy from the thermoelectric cooler to the

water. The water transfers the thermal energy away from the heat sink increasing the efficiency of cooling. The water was cooled with ice throughout the experiments to maintain a stable water temperature.

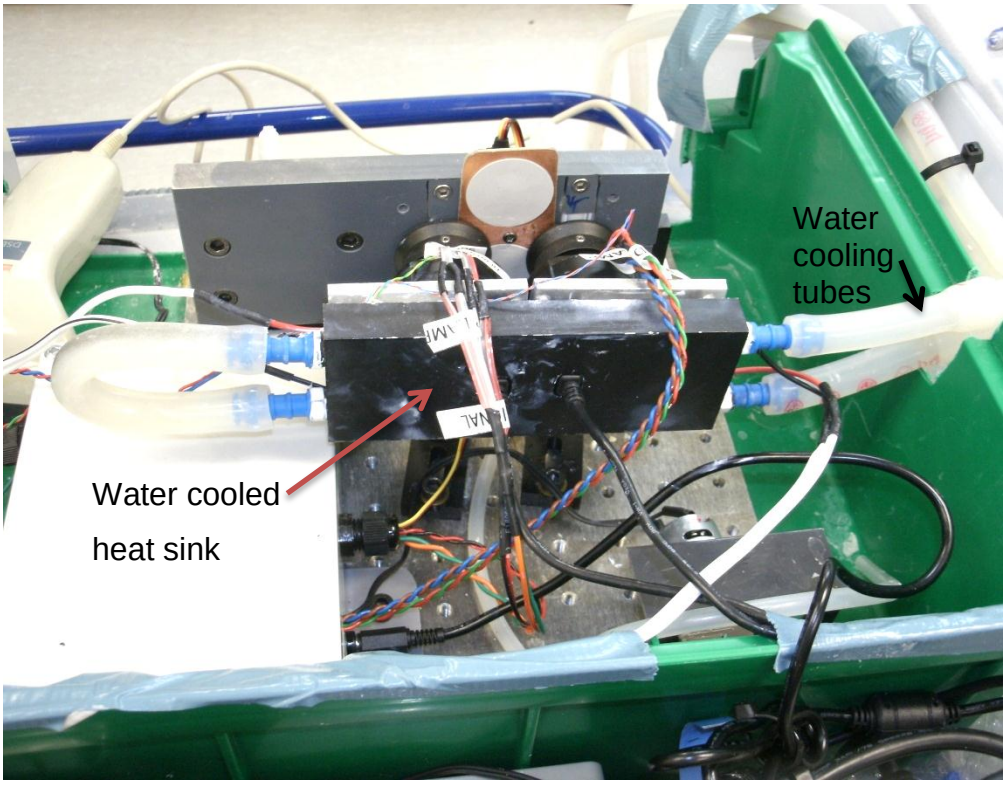

Figure 16: Water cooled heat sink mounted on the spectrometer

To increase the thermal conduction to the spectrometer a cold plate was made. The cold plate was recessed to match the outer case of the spectrometer. This maximises the contact surface area of the spectrometer which maximises the thermal conduction of the spectrometer allowing good heat transfer. The cold plate was used to connect the thermoelectric cooler to the spectrometer and to hold the Peltier element to the heat sink as illustrated in Figure 17.

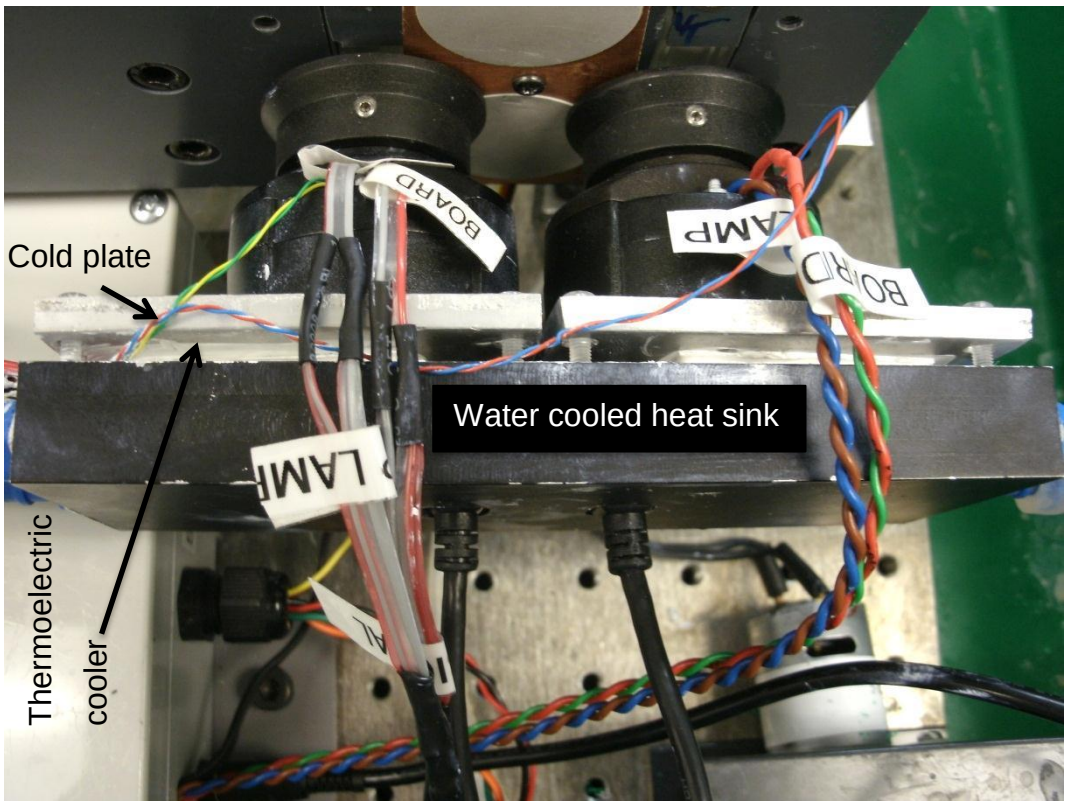

Figure 17: This illustrates the cold plate connected to the spectrometer. The thermoelectric cooler is mounted between the cold plate and the heat sink.

3.4.3 Transflectance configuration

The long path length of the reflectance configuration was attenuating the NIR light. To decrease the attenuation of the near infrared light the measurement configuration was changed to a transflectance configuration. The transflectance configuration uses a cuvette with a smaller path length. A Starna 48-Q-2 (Starna Scientific Ltd, London, United Kingdom) flow cell was used. The cuvette had a shorter path length of 2 mm compared to the 48-Q-5 flow cell which has a path length of 5 mm. A 2 mm cuvette was chosen because it allows adequate interaction of the NIR light with the milk sample and is not too thin. Figure 18 shows the difference in path length of the 2 mm cuvette (48-Q-2) and the 5 mm cuvette (48-Q-5).

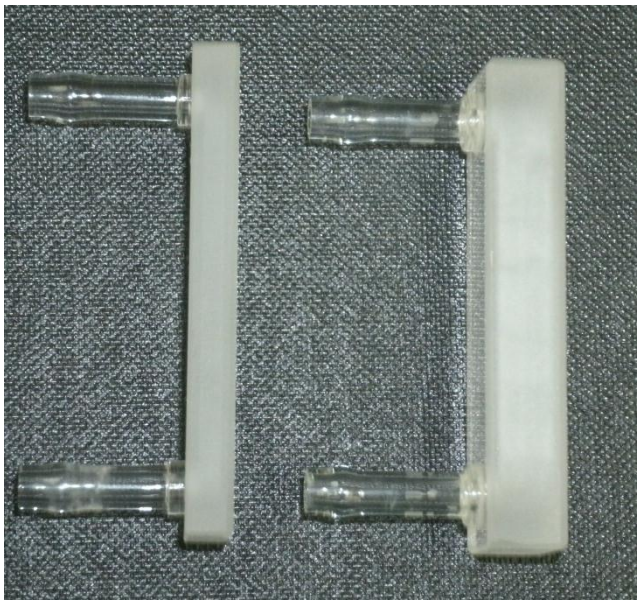

Figure 18: Comparison between 48-Q-2(left) and 48-Q-5(right) cuvettes used.

A piece of white spectralon R99 (Labsphere, North Sutton, New Hampshire, USA) was placed behind the cuvette to reflect any transmitted light. To quantify the increase in signal the transflectance configuration provides, the spectrum of milk was measured for both configurations.

A comparison of the reflectance configuration and the transflectance configuration for the MicroNIR1700 and MicroNIR2200 spectrometers is shown in Figure 19 and Figure 20.

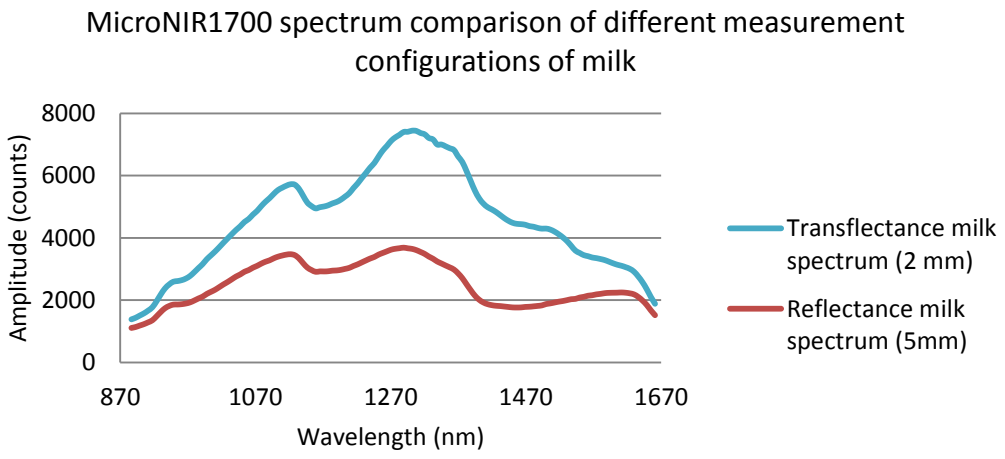

Figure 19: MicroNIR1700 spectrum comparison between the different measurement configurations.

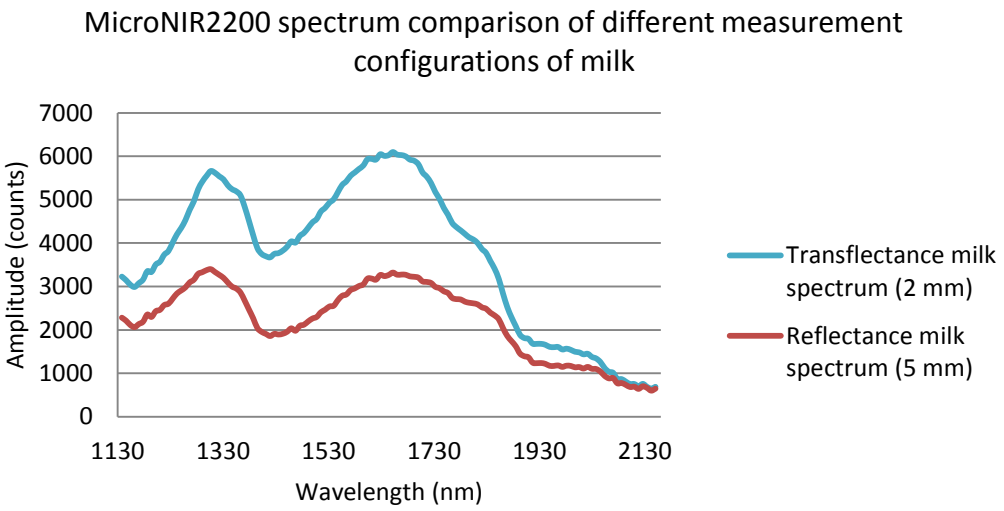

Figure 20: MicroNIR2200 spectrum comparison between the different measurement configurations.

Around the water absorption peak of 1940 nm there is only a slight increase in the received spectrum. This is because the water absorption peak at 1940 nm has a high attenuation coefficient which quickly absorbs light around 1940 nm. The reason why the spectrum converges at 2130 nm is due to the responsivity of the detector diminishing at the extended near infrared region.

The noise present on the spectrum is white noise, which is random across the spectrum. The variance in spectrum caused by the white noise can be reduced by averaging a large number of scans. During the laboratory trial 1000 scans were averaged for each sample to reduce the spectrum noise.

3.5 Spectrum noise and dark current method

To investigate the spectrum noise and dark current temperature dependences the spectrometers were mounted onto a piece of R99 spectralon as illustrated in Figure 21.

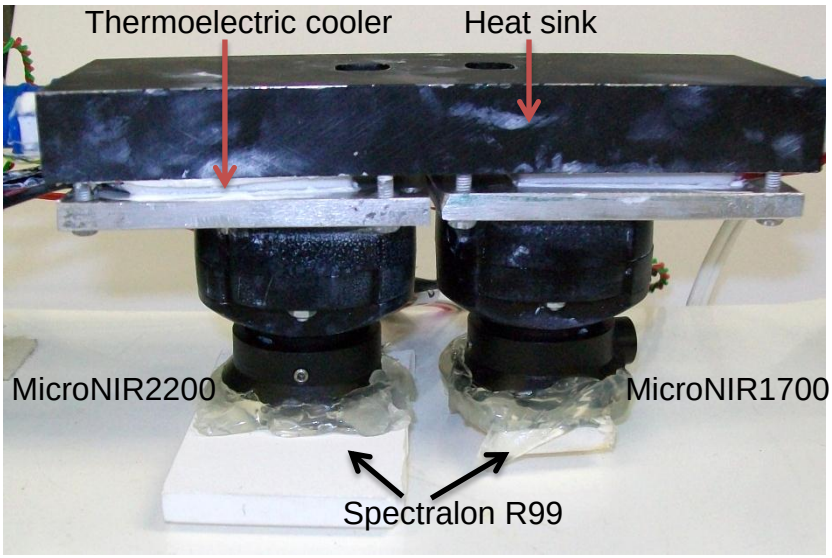

Figure 21: Experimental setup for measuring spectrum noise.

The temperature control circuit was used to maintain the spectrometers at 42°C and 20°C. For each spectral scan a dark current reading and a spectral scan of the R99 spectralon was taken. A total of 200 scans were acquired and were normalised to the first reference scan. The temperature was changed by changing the target temperature of the temperature control circuit. The temperature control circuit adjusted the PWM to the thermoelectric cooler to maintain the spectrometers at the target temperature. Dark current readings were taken at different temperatures between 35°C and -10°C. The dark current spectrum characteristics at different temperatures are illustrated in section 4.2.3. The spectrum noise of the spectrometers is illustrated in section 4.2.4.

3.6 Farm trial

Previously during the laboratory trial insufficient milk composition results were achieved with the unmodified spectrometers. For further details on the results achieved with the unmodified spectrometers see section 4.1 of the results. Due to the high prediction errors achieved in the laboratory trial hardware modifications were made to the spectrometers to improve the signal to noise ratio of the spectra. Details on the hard ware modifications made can be found in section 3.4. The hardware modifications improved the signal to noise ratio of the spectra. The MicroNIR1700 peak signal to noise ratio improved from 110 to 240 and the MicroNIR2200 spectrometers signal to noise ratio improved from 30 to 230. For further details on how the hardware changes have improved the spectra see section 4.2 of the results.

The aim of the farm trial is to see if the hardware changes made to optimise the spectrometers can predict fat and protein within ICAR precision for an online sensor. The farm trial involved collecting milk samples from three different farms and testing the capabilities of the MicroNIR spectrometers to predict fat and protein. Calibration models were built from two different farms and were tested on a third independent farm to test if the calibration models are transferable between farms.

The first farm was located at Te Pahu and had 600 cows. The herd contained a mix of Friesians and Jersey cows which allowed a diverse range of milk composition to be obtained. The second farm was located at Te Mawhai south of Te Awamutu and comprised of 750 cows. This herd consisted of Jersey cross and Holstein-Friesian cows. The third farm was located at Puketaha north of Hamilton and comprised of 450 Shorthorn cows.

A total of 590 milk samples were taken over a period of one week during November 2012. The milk samples were collected during morning milking’s from three different farms. All milk samples were taken from composite milk during the milking process. During milking, the milk is delivered to the milk line using the long delivery tube. A T sampler was inserted into the long delivery tube and was used to collect composite spot

samples during different stages of milking as illustrated in Figure 22. This varied the milk fat composition of samples collected.

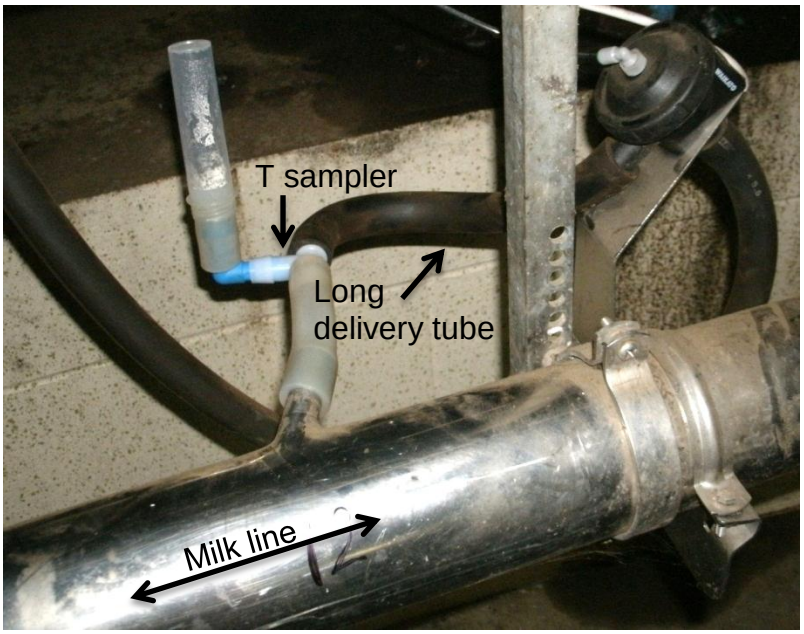

Figure 22: Composite T sampler used to acquire milk samples for spectrum analysis.

Transflectance spectra were measured through two Starna 2 mm glass flow cells (48-Q-2) with a diffused reflector behind the flow cell. This increases the received signal to the spectrometers. Each flow cell was held in place by two plastic tabs which screwed into the mounting plate. This removed the possibility of the cuvettes relative incident angle to the spectrometer changing between tests. Two reference tiles of spectralon R50 were mounted onto a servomechanism and were used to normalise the spectrum of milk as illustrated in Figure 23.

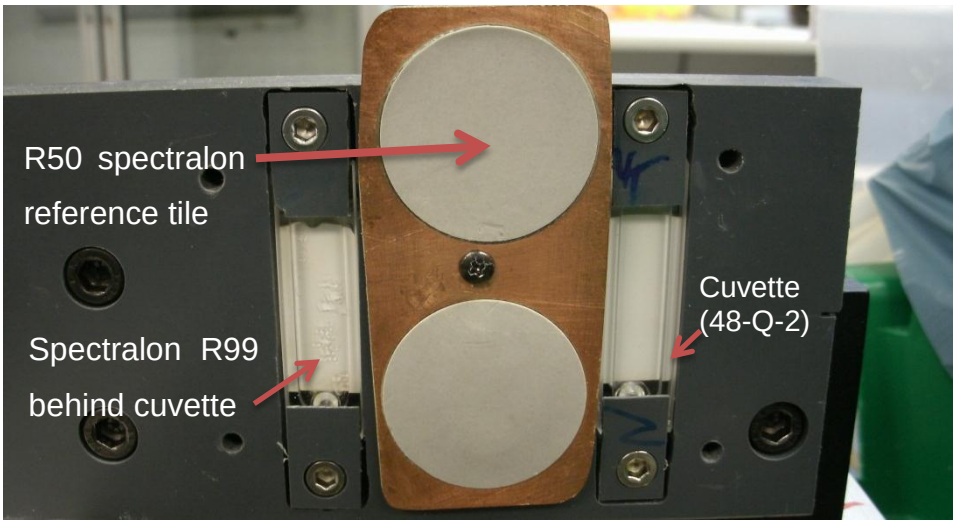

Figure 23: Cuvette mounting plate used for the experiment.

The two cuvettes were connected in series with silicon tubing. A peristaltic pump (WPX1, Welco, Japan) was used to pump the sample into the cuvettes. Figure 24 and Figure 25 illustrate the hardware configuration and a schematic diagram of the hardware used for the farm trial.

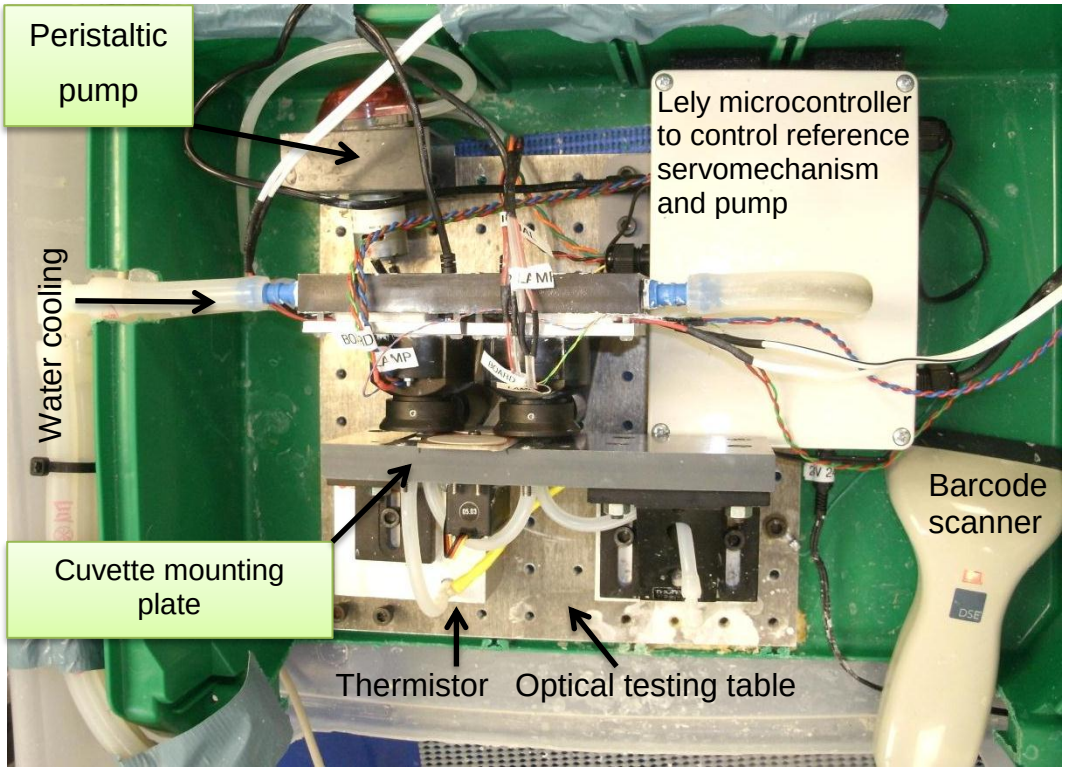

Figure 24: Farm trial optical testing table.

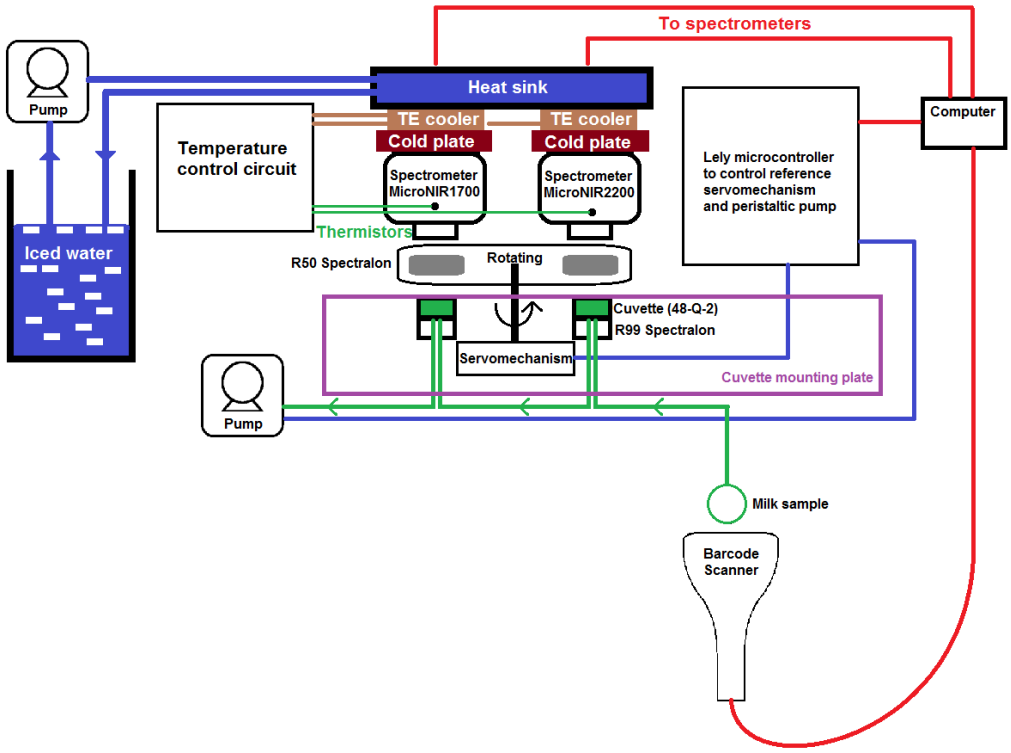

Figure 25: Schematic diagram of farm trial hardware.

The temperature control circuit maintained the spectrometers at 20°C. The spectrometer's temperature was allowed to stabilise with the lamps on for 10 minutes; this allowed the temperature of the bulbs to stabilise and helps to prevent the light output from drifting.

The temperature of the milk samples had significantly dropped once the sample was obtained. This was caused by the samples cooling down through the long delivery tube. The typical temperature of a composite milk sample was 20°C which is significantly colder than the natural temperature of milk. The samples temperature was restored using a water bath set at 48°C. Samples were placed in the water bath for 10 minutes before the spectrum was taken. This allowed the samples to be rapidly reheated to the natural temperature of 38°C. The milk samples were inverted five times to ensure homogenisation of the sample. New milk samples were inserted into the water bath during the testing process.

Samples were scanned to record the barcode number and initiate the sampling process. Excess fluid was pumped through the cuvette to reduce cross contamination from the previous samples. Fluid temperature was measured using a thermistor downstream of the cuvette. The fresh milk

samples were tested during milking using the NIR optical testing table as illustrated in Figure 26.

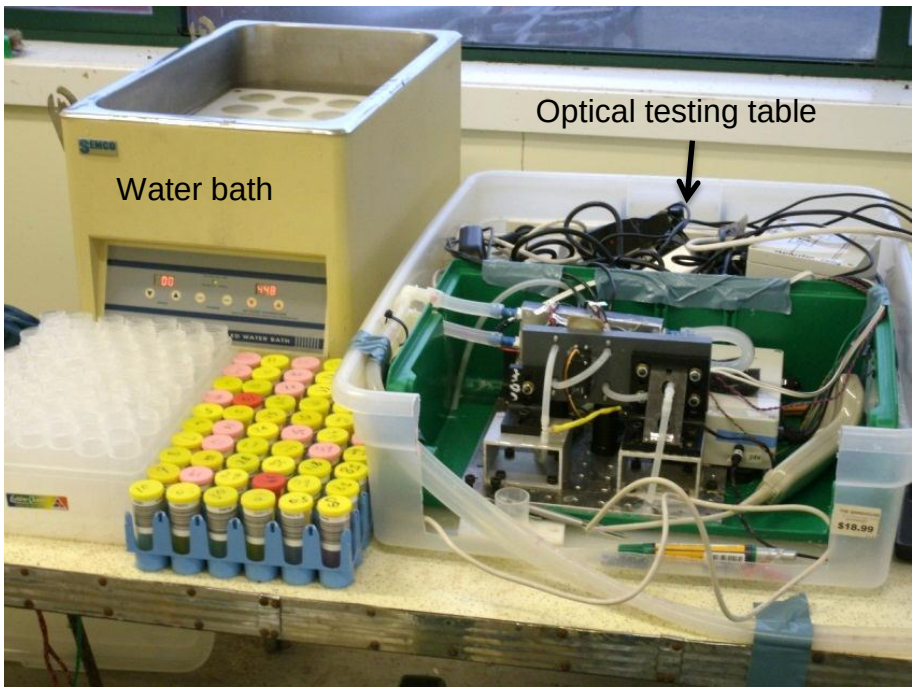

Figure 26: On farm experimental setup.

Transflectance spectra were taken in the wavelength range from 880 to 1660 nm at 6 nm intervals with the MicroNIR1700 spectrometer. Spectra taken with the MicroNIR2200 spectrometer were in the wavelength range of 1130 to 2150 nm at 8 nm intervals. The dark current was measured immediately after the milk spectrum by turning the lamps off and recording the dark current count. The reference measurement was acquired using two Spectralon R50 disks which were inserted into the optical path of the spectrometers. Spectralon R50 has a reflectance of 50%. This reference standard was used because the reflectance of the reference standard is close to the sample reflectance. Normalising the milk spectrum to a reference spectrum with a similar reflectance helps to reduce the effect of the anomalous pixels. The reference spectrum was used to calibrate out dark current and intensity variations from the spectrometers.

Normalised reflectance spectrum was measured of each sample by recording the dark current, 50% reference standard and milk scan at an

integration time of 1300 µs and the normalised reflectance value was calculated using (3-1)

$$\Gamma(\lambda) = \frac{\delta(\lambda) - D(\lambda)}{R_{ref}(\lambda) - D(\lambda)} \quad \textbf{(3-1)}$$

Where $\Gamma(\lambda)$ is the normalised reflectance,
δ is the spectrum intensity,
D is dark current, and
R_{ref} is the reference spectrum.

The normalised reflectance spectrum was averaged over 1000 scans and was recorded in a Microsoft excel worksheet. Once the spectrum data was saved and the sample was removed from the cuvettes via the peristaltic pump. No further repetitions were done with the sample after testing due to the sample being contaminated.

Data from two farms was used to build a calibration model while the third farm was used for validating the model. The data from the third farm was not included in the calibration data set making the validation set independent of the calibration set. This process was repeated on all of the different farms so each farm was independently tested with a different calibration model made from the other two farms.

To test the overall performance of the MicroNIR spectrometers the data was pooled together from all three farms. The pooled data set was ordered by fat and protein and the data set was divided up, two thirds of the data set was used for calibration and one third for a validation set. This ensured a representative sample was used for the calibration of the model which included inter farm variations. This is essential when trying to calibrate for protein which has a narrow distribution.

3.7 Data Analysis

Reference data was collected from Testlink and was stored in a Microsoft Excel file. The reference data was matched to the corresponding spectra and 11Ants model builder (11Ants Analytics, Hamilton, NZ) was used to analyse the spectrum to build chemo metric models.

11Ants model builder is a powerful yet simple to use Microsoft Excel add-in which allows easy data mining. The software emulates the best practices of a data mining expert and does the complex work for the user. The 11Ants model builder has a library of 11 machine learning algorithms including; decision tree, Gaussian processes, logistic regression, logitboost, model tree, naive Bayes, nearest neighbour, partial least squares, random forest, Ridge regression, and support vector machine.

11Ants model builder was used for calibration. Calibration is the process by which the mathematical relationship between the spectrum provided by spectrometers and those known for the measured object is established. Spectroscopic measurements are used to establish a linear relationship between the apparent absorbance and concentration via the Beer-Lambert Law.

The following sections provide an overview on the data processing methods used for calibration and testing the performance of the calibration model.

3.7.1 Spectral pre processing

Pre-processing of spectral data is often of vital importance if reasonable results are to be obtained. The main goal of pre-processing is to transform data in such a way that the signals will adhere to Beer's law (3-2). Beer's law states the absorbance and concentration are linearly correlated:

$$A = \varepsilon l c \qquad \textbf{(3-2)}$$

where ε is the molar absorptivity, l is the effective path length, and c is the concentration of the constituent of interest. The aim is to make the collective term $\varepsilon \times l$ constant for the data set, thus making the relationship between A and c linear. Many physical and chemical phenomena can

cause a deviation from this linear relationship, including scatter from particulates, molecular interactions, changes in refractive index, changes in chemical equilibrium as a function of concentration, stray light, and changes in sample size and path length (Rinnan et al. 2009)

The following spectral pre-processing methods can by 11Ants Analytics™; first derivative, second derivative, and smoothing. Performing a first derivative on the spectrum removes background effects on the spectrometer such as temperature variations. Performing a second derivative on the spectrum removes baseline variations and highlights areas of high absorbance and reflectance. The smoothing algorithm is beneficial when high frequency Gaussian noise is present on the spectrum.

The aim of calibrating is to establish a mathematical relationship between the values provided from the spectrometer and the measured concentration of the sample. Selection of the calibration samples is one of the most important steps in constructing a calibration model and involves choosing a series of samples, which ideally encompass all possible sources of physical and chemical variability in the samples to be subsequently predicted. The samples included in the calibration set should span the whole variability in both. The selected samples should be uniformly distributed throughout the calibration range in the multidimensional space defined by spectral variability (Williams et al. 2001).

The NIR spectra were divided into two random sample sets: a calibration set containing two thirds of all samples and a test set containing the remaining samples. The data is termed calibration data because it was involved in the calibration of the model. Random selection was done by sorting the samples according to the concentration of the respective milk content and assigning every third of the sorted samples to the test set. This ensured a representative sample set was used to build the calibration model. The test set is independent from the calibration set. The calibration set is used to observe patterns in the data. The structure of these patterns is then described as a model.

The model may contain multiple ensembles. Rather than using one model to build an algorithm a combination of models are used to predict values. The models are used in combination to complement each other to assist prediction.

The model was used to predict the reference value of the milk sample and the predicted value was compared for each sensor. To predict the composition of milk multivariate calibration models were used. The following sections describe the calibration models used to predict milk composition.

3.7.2 Principal component analysis

Principal component analysis (PCA) is a variable reduction method which reduces the data set of the matrix to a smaller number of (A) variables called principal components. Let the original data set be defined as $\mathbf{X}(k \times n)$, where (k) represents the number of wavelengths and (n) is the number of samples. Principal component analysis is constructed from the expression (3-3).

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \qquad \textbf{(3-3)}$$

Where $\mathbf{T}(k \times n)$ is a matrix which contains the (A) scores for the principal components. The $\mathbf{P}(k \times n)$ matrix contains the (A) loadings for the principal components and $\mathbf{E}(k \times n)$ the residuals matrix of the model. The scores are the intensities of the new variables for the samples and the loadings the new variables obtained from the original ones.

The aim of PCA is to reduce the original data matrix by deleting redundant information. The mathematical algorithm used simply calculates the eigenvectors and eigenvalues of a matrix to minimise the residual error in each step. Each PC is determined in such a way that it will account for the residual variance in the data matrix, $\mathbf{x}$, and the process is allowed to progress until the PCs equal the original variables in number and account for all of the variance in the data.

Multiplying the scores $\mathbf{t}_1$ by the loadings $\mathbf{P}_1^{\mathrm{T}}$ allows the original matrix to be converted into a new matrix $\mathbf{X}_1$ such that:

$$\mathbf{X}_1 = \mathbf{t}_1\mathbf{P}_1^{\mathrm{T}} \tag{3-4}$$

$\mathbf{X}_1$ will be different from the original matrix, $\mathbf{X}$, but constitute its best possible reproduction. A second PC can then be calculated from a loadings vector and its corresponding scores vector:

$$\mathbf{X}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{P}_1^{\mathrm{T}} = \mathbf{t}_2\mathbf{P}_2^{\mathrm{T}} \tag{3-5}$$

The process can be continued until the whole original matrix has been resolved. However the data are not compressed until the user chooses the number of PCs, which is much smaller than the original variables. In practice a compromise between the need to explain the variance in the original data and to avoid over fitting should be found (Blanco Romía & Alcalà Bernàrdez 2009).

3.7.3 Principal component regression

Principal component regression (PCR) performs multiple inverse regression of the predictor variables against the scores rather than the original data.

The first step in the process involves resolving matrix $\mathbf{X}$ into its PCs as described in (3-6)

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} = \sum_{a=1}^{A} t_a p_a{}^{\mathrm{T}} + \mathbf{E} \tag{3-6}$$

When the optimum number A of principal components describing the original matrix has been chosen, the matrix $\mathbf{X}$ can be represented by its scores matrix (3-7).

$$\mathbf{T} = \mathbf{X}\mathbf{P} \tag{3-7}$$

Where the scores matrix, $\mathbf{T}$, and the loadings matrix, $\mathbf{P}$, are both obtained from the data matrix $\mathbf{X}$. Matrix $\mathbf{Y}$ can be calculated by regressing $\mathbf{Y}$ against $\mathbf{T}$:

$$\mathbf{Y} = \mathbf{TB} + \mathbf{E} \qquad \textbf{(3-8)}$$

Where $\mathbf{B}$ is the regressor matrix and can be calculated by least squares regression provided the values of $\mathbf{Y}$ in the calibration model are known:

$$\hat{\mathbf{B}} = (\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y} \qquad \textbf{(3-9)}$$

Once an accurate model is developed, predicting the results for a set of new samples involves performing calculations similar to those used in its construction. The new spectroscopic data, $\mathbf{X}^{*}$, is auto scaled by using values calculated from the training data matrix $\mathbf{X}$. The loadings matrix obtained in the calibration process is used to calculate the scores with the new samples:

$$\mathbf{T}^{*} = \mathbf{X}^{*}\mathbf{P} \qquad \textbf{(3-10)}$$

and the regression matrix, also obtained during the calibration, is used in combination with the scores for the new samples to calculate the corresponding concentrations(3-11).

$$\mathbf{Y} = \mathbf{T}^{*}\mathbf{B} \qquad \textbf{(3-11)}$$

Over fitting of the calibration data set is the greatest problem with principal component regression. Over fitting PCR models can be avoided by limiting the number of principle components used for calibration. Also the number of principal components used may not be the optimum number of principle components to predict the sample concentration. This problem is addressed by having an intermediate step which estimates the correlation

coefficient between the sample concentrations and uses the components exhibiting significant correlation (Blanco Romía & Alcalà Bernàrdez 2009).

3.7.4 Method of Least squares

The following section gives an overview on method of least squares for further details see Rawlings (1998).

The simplest linear model involves only one independent variable and states that the true mean of the dependent variable changes at a constant rate as the value of the independent variable increases or decreases. Thus, the functional relationship between the true mean of, Y_i, denoted by, $\mathcal{E}(Y_i)$, and X_i is the equation of a straight line (3-12).

$$\mathcal{E}(Y_i) = \beta_0 + \beta_1 X_i \qquad \textbf{(3-12)}$$

where β_0 is the intercept, the value of $\mathcal{E}(Y_i)$ when $X = 0$, and β_1 is the slope of the line, the rate of change in $\mathcal{E}(Y_i)$ per unit change in X.

Any measurement will to some degree contain noise, variations, or errors which are likely to occur at random. The deviation of an observation, Y_i, from its population mean, $\mathcal{E}(Y_i)$, is taken into account by adding a random error, ε_i, to give the equation (3-13)

$$\mathcal{E}(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \textbf{(3-13)}$$

The subscript i indicates the particular observational unit, $i = 1,2,...,n$. The $\mathbf{X}_i$ are the n observations on the independent variable and are assumed to be measured without error. The random errors have zero mean and are assumed to have common variance, σ^2, and to be pairwise independent.

The least squares estimation procedure uses the criterion that the solution must give the smallest possible sum of squared deviations of the observed Y_i from the estimates of their true means provided by the solution. Let $\hat{\beta}_0$

and $\hat{\beta}_1$ be numerical estimates of the parameters β_0 and β_1 , respectively, and let

$$\hat{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad \textbf{(3-14)}$$

be the estimated mean of Y for each $X_i, i = 1,...,n$. The least squares principle chooses β_0 and β_1 that minimises the sum of squares of the residuals, $SS(\text{Res})$.

$$\begin{aligned} SS(\text{Res}) &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \sum e_i^2 \end{aligned} \qquad \textbf{(3-15)}$$

Where $e_i = (Y_i - \hat{Y}_i)$ is the observed residual for the ith observation. The summation is over all observations of the data set. The estimators for β_0 and β_1 are obtained by using calculus to find the values that minimise $SS(\text{Res})$. The derivatives of $SS(\text{Res})$. with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ in turn are set to zero. This gives two equations in two unknowns called the normal equations (3-16).

$$\begin{aligned} n(\hat{\beta}_0) + (\sum X_i)\hat{\beta}_1 &= \sum Y_i \\ (\sum X_i)\hat{\beta}_0 + (\sum X_i^2)\hat{\beta}_1 &= \sum X_i Y_i \end{aligned} \qquad \textbf{(3-16)}$$

Solving the normal equations simultaneously for $\hat{\beta}_0$ and $\hat{\beta}_1$ gives the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ (3-17)

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}. \end{aligned} \qquad \textbf{(3-17)}$$

These estimates of the parameters give the regression equation (3-18)

$$\hat{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad \textbf{(3-18)}$$

3.7.5 Partial least squares

The following section is an overview on partial least squares, for further detail see Malinowski (2002). Partial least squares (PLS) is a factor analytical technique that is useful when the target matrix does not contain the full model representation, that is, when there are more factors in the data matrix than in the target matrix. The method was invented by Herman Wold who began developing it in the early 1970s and completed it in 1977. The first chemical application of PLS appeared in 1979 in a study by Gerlach, Kowalski and Herman Wold.

In PLS the combination step is amalgamated with the decomposition step so that the eigenvectors of the data matrix are extracted in a sequence congruent with the eigenvectors of the target matrix.

The PLS method involves regressions between the scores of two matrices, X and Y. For spectroscopic analysis of mixtures, X and Y may represent, respectively, spectra and concentration matrices of the mixtures. PLS seeks a calibration model such that

$$\underset{k\times p}{Y} = \underset{k\times m}{X}\ \underset{m\times p}{B_{PLS}} \qquad \textbf{(3-19)}$$

Where B_{PLS} is the best set of calibration constants for the system. Partial least squares require a training set consisting of matrices X and Y, which are used to determine B_{PLS} It also requires an independent test set of the measurable variables and properties that can be used to test the accuracy and validity of the developed PLS model.

Normally multiple regression analysis minimises the sum squares of errors in Y, as described by (3-20)

$$SS(\mathrm{Res}) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad \textbf{(3-20)}$$

The mean squared error of prediction (MSEP) is simply (3-21)

$$\text{MSEP} = \frac{SS(\text{Res})}{df} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{(p-k)} \quad \textbf{(3-21)}$$

In this case the number of degrees of freedom, *df*, is equal to the number of random independent variables *p* minus the number of parameters, *k*, employed in the model. There is no simple way for determining the degrees of freedom.

For nonlinear models additional factors are required to account for the nonlinearities. Determining the appropriate number of latent factors represents is important for a successful PLS. Using too few factors leads to under fitting and poor predictability. Using too many factors may result in over fitting of the training data set. When the calibration model is tested on the independent test set the calibration model may fail to predict the new data well.

There are a variety of different methods to determine the optimum number of latent factors. The normal method of determining the number of latent factors involves determining the MSEP at each factor level and extracting the number of latent variables until the MSEP falls below a threshold. Other methods use cross validation to estimate the optimum number of latent factors.

3.7.6 Gaussian process

The following chapter provides an overview of Gaussian processes obtained from Rasmussen and Williams (2006). For full proof see Gaussian Processes for Machine Learning.

Gaussian process models are constructed from classical statistical models by replacing latent functions of parametric form by random processes with Gaussian prior. The basic idea in Gaussian process is to model a given dataset as a realisation of a stochastic process.

A Gaussian process is completely specified by its mean function and covariance function. We define the mean function $m(\mathbf{x})$ and the covariance function $k(\mathbf{x},\mathbf{x}')$ of a real process $f(\boldsymbol{x})$ as (3-22)

$$m(\mathbf{x}) = \mathrm{E}\left[f(\mathbf{x})\right]$$
$$k(\mathbf{x},\mathbf{x}') = \mathrm{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\right] \quad \textbf{(3-22)}$$

And the Gaussian process is written as (3-23)

$$f(\mathbf{x}) \sim \mathrm{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')) \quad \textbf{(3-23)}$$

Random variables represent the value of the function $f(\boldsymbol{x})$ at location **x**. A Gaussian process is defined as a collection of random variables, any finite number of which has a joint Gaussian distribution. This implies there is a consistency requirement, also known as the marginalization property. This simply means that if the Gaussian process specifies $(y_1, y_2) \sim \mathcal{N}(\mu, \sum)$, then it must also specify $y_1 \sim \mathcal{N}(\mu_1, \sum_{11})$ where $\sum_{11}$ is the relevant sub matrix of $\sum$

A simple example of a Gaussian process can be obtained from the Bayesian linear regression model $f(x) = \varphi(\mathbf{x})^{\mathrm{T}}\mathbf{w}$ with prior $\mathbf{w} \sim \mathcal{N}(0, \sum_p)$.

We have the mean and covariance (3-24).

$$\mathrm{E}\left[f(\mathbf{x})\right] = \varphi(\mathbf{x})^{\mathrm{T}}\mathrm{E}\left[\mathbf{w}\right] = 0,$$
$$\mathrm{E}\left[f(\mathbf{x})f(\mathbf{x}')\right] = \varphi(\mathbf{x})^{\mathrm{T}}\mathrm{E}\left[\mathbf{w}\mathbf{w}^{\mathrm{T}}\right]\varphi(\mathbf{x}') = \varphi(\mathbf{x})^{\mathrm{T}}\sum_p \varphi(\mathbf{x}') \quad \textbf{(3-24)}$$

Thus $f(\mathbf{x})$ and $f(\mathbf{x}')$ are jointly Gaussian with zero means and covariance given by $\phi(\mathbf{x})^{\mathrm{T}}\sum_p \phi(\mathbf{x}')$. Indeed the function values $f(\mathbf{x}_1)\ldots f(\mathbf{x}_n)$ corresponding to any number of input points *n* are jointly Gaussian, although if $N < n$ then this Gaussian is singular as the joint covariance matrix will be of rank *N.*

A graphical model for a Gaussian process for a regression is illustrated in Figure 27. The squares represent observed variables and circles represent unknowns. The thick horizontal bar represents a set of fully connected nodes. Note that an observation y_1 is conditionally independent of all other nodes given the corresponding latent variable, f_1. Because of the marginalization property of Gaussian processes addition of further

inputs, $\mathbf{x}$, latent variables, *f*, unobserved targets, y_*, does not change the distribution of any other variables (Rasmussen & Williams 2006).

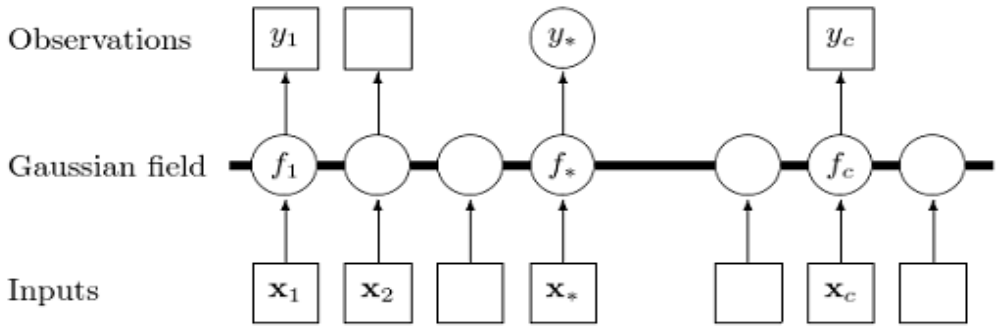

Figure 27: Gaussian process for a regression(Rasmussen & Williams 2006)

3.7.7 Analysing the performance of the model

Once a model is created, its outcome must be tested on an independent data set. The quality of calibration models can be evaluated via statistical parameters, of which those allowing the mean error of the whole population are preferred. The data was plotted on a prediction scatter graph comparing the predicted value with the known value. Statistical analysis of the data was performed to quantify the accuracy of the model. The statistics typically used to assess the quality of calibration models calculate the error of prediction as the summation of the squares of the residuals (3-25), which is usually designated as the predicted residual error sum of squares (PRESS).

$$\text{PRESS} = \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2 \qquad \textbf{(3-25)}$$

Where $\hat{y}_i$ is the true value and

Y_i is the predicted value

The mean prediction error value is determined by dividing equation (3-25) by the number of samples *(n)* to obtain the mean square error (3-26).

$$\text{MSE} = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n} \qquad \textbf{(3-26)}$$

Another measure of the models quality is the root mean square error (RMSE), which is defined in equation (3-28)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n}} \qquad \textbf{(3-27)}$$

To measure how strongly the calibration model is correlated the Pearson's sample correlation coefficient (r) is used (3-28). It is used to make a quantitative assessment of the strength of the relationship between the predicted value and the known value.

$$r = \frac{n(\sum_{i=1}^{n} x_i y_i) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^{\,2} - \sum_{i=1}^{n}(y_i)^2\right]}} \qquad \textbf{(3-28)}$$

The Pearson's sample correlation coefficient gives an indication of how strongly correlated the predicted values are against the known values. The Pearson's correlation coefficient can be between +1 and -1. When r = +1 it indicates the predicted values are fully correlated with the known values. If r = 0 it means there is no correlation and if r = -1 it means there is a negative correlation.

Another measure of the calibration models performance is the coefficient of determination. The coefficient of determination, (r^2), is a measure of the proportion of variability in the predicted value against the known value. The coefficient of determination is determined using (3-29):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \qquad \textbf{(3-29)}$$

The coefficient of determination can be between 1 and 0. When r^2 = 1 it indicates the regression line fits the data perfectly and when r^2 = 0 it means there is no correlation between the data(Devore & Peck 2001).

4 Results

This chapter presents the results from the laboratory trial, optimisation of the hardware, and the results obtained during the farm trial.

Chapter 4.1 presents the results from the laboratory trial where the original spectrometers were used to predict the composition of milk. Chapter 4.2 illustrates how the hardware changes improve the spectrum quality. Chapter 4.3 reports on the ability of the modified MicroNIR spectrometers to predict milk composition and test the capability of the calibration models transferability between farms.

4.1 Laboratory testing of the JDSU spectrometers

A total of 1675 herd test milk samples, from a total of three different farms were used to assess the capability of the MicroNIR spectrometers to predict milk composition. The milk samples originated from Holstein-Friesian and Jersey cows and were tested immediately after Foss analysis at Testlink. Thus the samples had been collected and stored under refrigeration for at least 24 hours.

4.1.1 Laboratory spectra

Two different sets of spectra were acquired using the two MicroNIR spectrometers. The MicroNIR1700 spectrometer covered the spectrum between 880 – 1660 nm and the MicroNIR2200 spectrometer covered the 1130 – 2160 nm region. The spectrum for the laboratory testing of the MicroNIR1700 and MicroNIR2200 spectrometers are illustrated in Figure 28 and Figure 29 respectively.

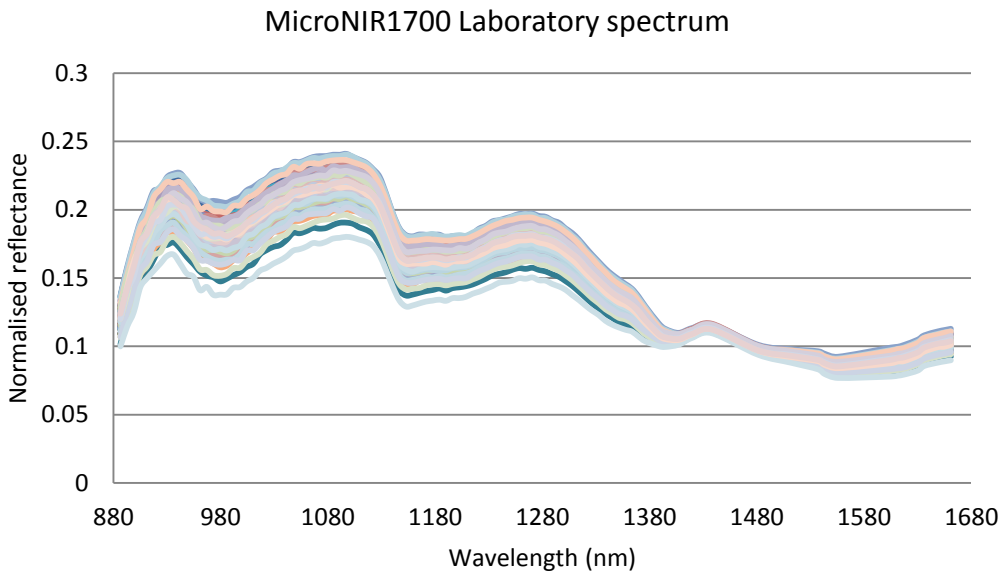

Figure 28: Spectra of milk samples acquired during the laboratory experiment with the MicroNIR1700 spectrometer

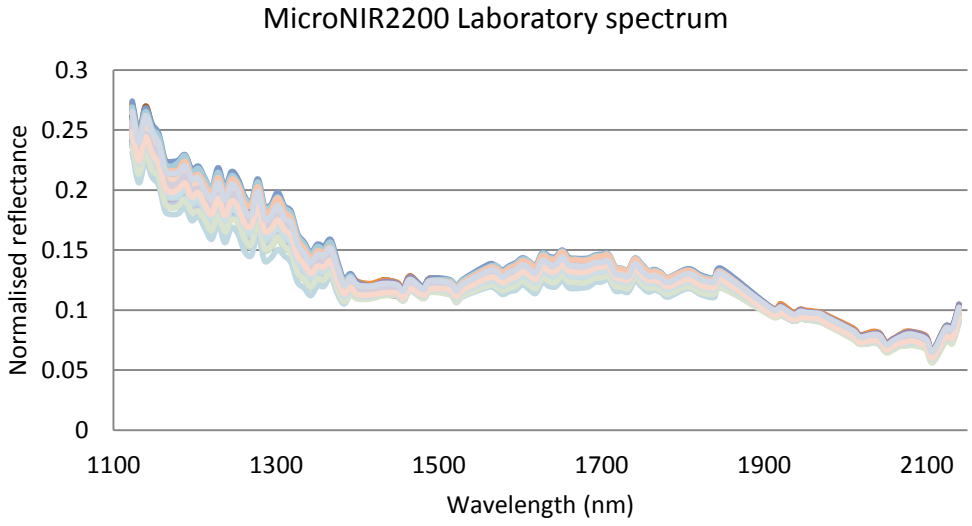

Figure 29: Spectra of milk samples acquired during the laboratory experiment with the MicroNIR2200 spectrometer

4.1.2 Laboratory results - Lipids

The range of fat content was from 1.97% to 11.26%, the average being 5.88% with a standard deviation of 1.18.

Figure 30: Laboratory trial fat histogram

Scatter plots of predicted against known fat concentration for the MicroNIR1700 and MicroNIR2200 are shown in Figure 31 and Figure 32 respectively. The MicroNIR1700 results are obtained using 126 pixels between 886 nm to 1661 nm. The first two pixels of the MicroNIR1700 detection array were erroneous and therefore discarded. Two PLS models were used to predict the fat content. One of the PLS models uses 12 latent variables and the other has 7 latent variables as determined by cross validation as the optimal number. The MicroNIR1700 spectrometer predicted the fat content with a RMSE of 0.272. This error of prediction was too high and does not meet the ICAR precision tolerance of 0.25%.

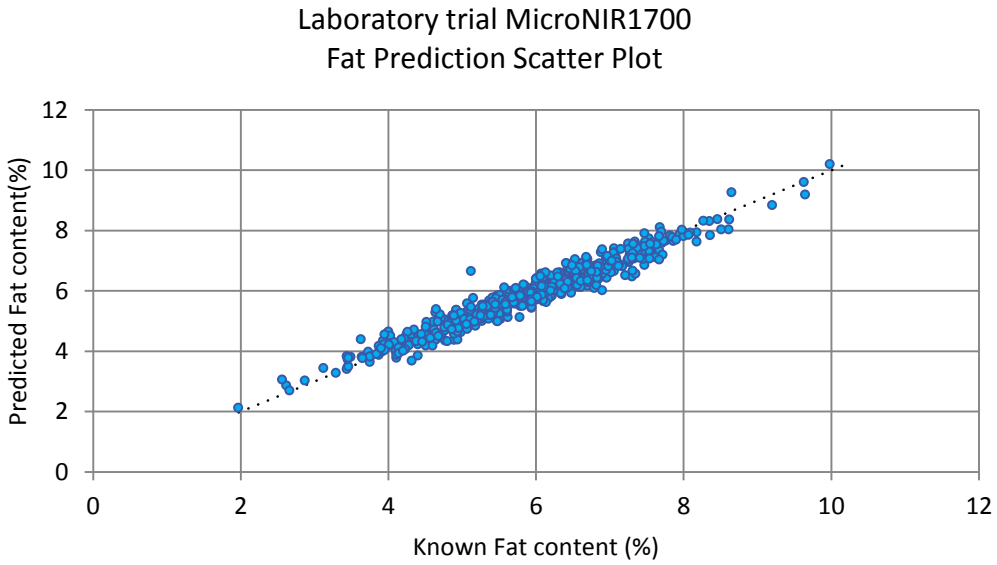

Figure 31: MicroNIR1700 fat prediction scatter plot of the laboratory data (RMSE = 0.272, r^2 = 0.946, Bias = -0.031)

The MicroNIR2200 results are obtained using 126 pixels between 1123 nm to 2140 nm. Once again the first two pixels were discarded due to erroneous signals. A PLS model with 6 latent variables was used to predict the fat content. The MicroNIR2200 spectrometer had a similar prediction error to the MicroNIR1700 spectrometer; a RMSE of 0.277 was achieved. This level of prediction error does not meet the ICAR precision limits.

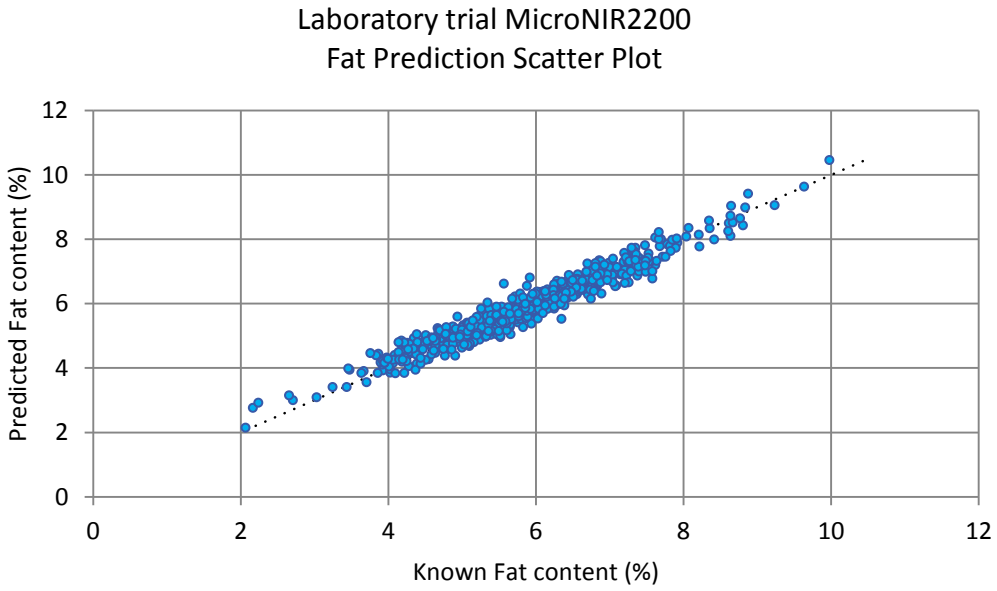

Figure 32: MicroNIR2200 fat prediction scatter plot of the laboratory data (RMSE = 0.277, r^2 = 0.946, Bias = 0.020)

4.1.3 Laboratory results - Protein

The range of protein content was from 2.81% to 6.51%, the average being 4.41% with a standard deviation of 0.47. A histogram illustrating the range of protein tested is shown in Figure 33.

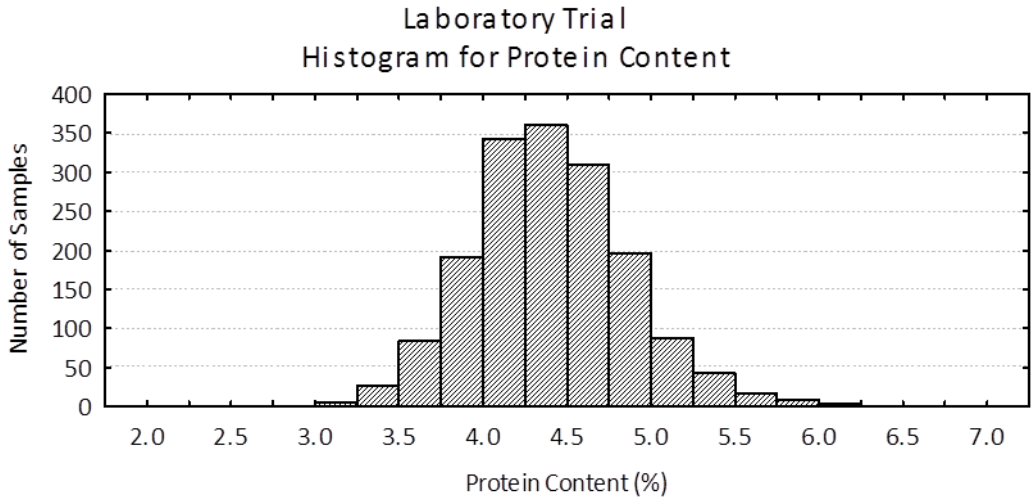

Figure 33: Laboratory trial protein composition

Scatter plots of predicted against known protein concentration for the MicroNIR1700 are shown in Figure 34. The MicroNIR1700 results are obtained using 126 wavelengths between 886 nm to 1661 nm. The MicroNIR1700 spectrometer predicted the protein composition with a RMSE of 0.364.

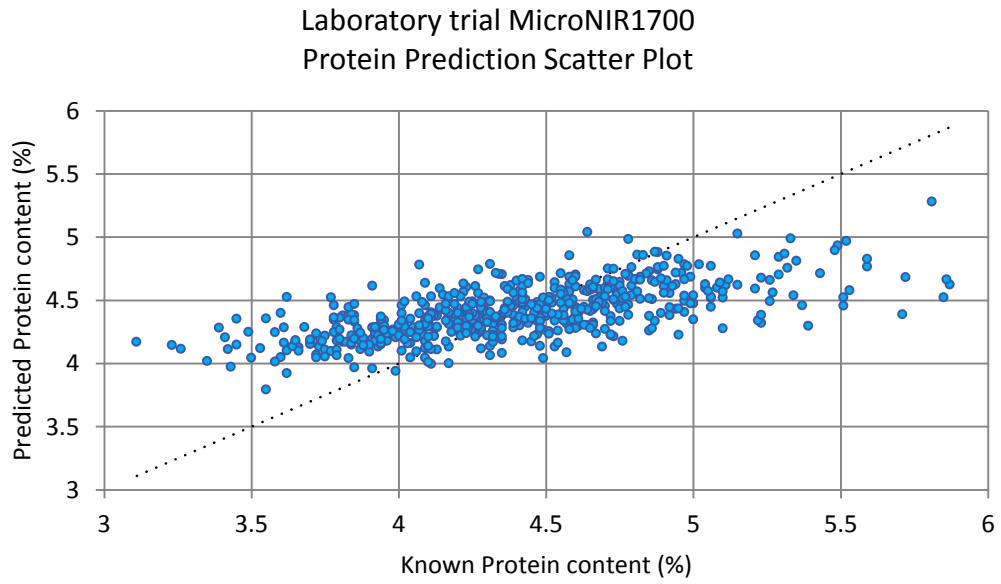

Figure 34: MicroNIR1700 protein prediction scatter plot of the laboratory data (RMSE = 0.364, r^2 = 0.437, Bias = 0.008)

The MicroNIR2200 results are obtained using 126 wavelengths between 1123 nm to 2140 nm. The MicroNIR2200 spectrometer had a higher prediction error compared to the MicroNIR1700. The MicroNIR2200 predicted the protein composition with a RMSE of 0.379. A protein prediction scatter plot for the MicroNIR2200 spectrometer is illustrated in Figure 35

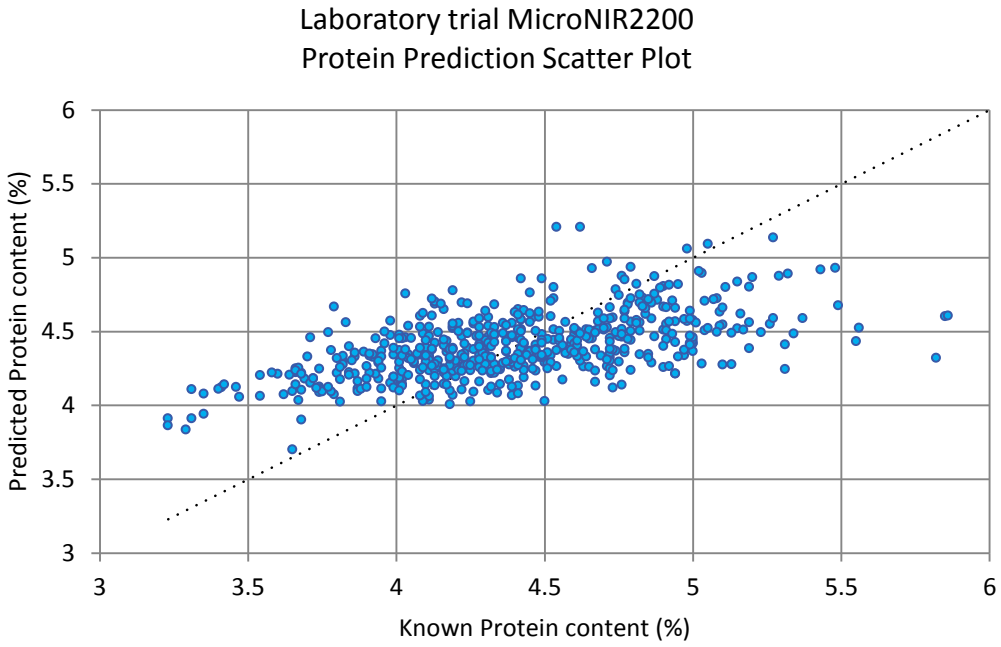

Figure 35: MicroNIR2200 protein prediction scatter plot of the laboratory data (RMSE = 0.379, r^2 = 0.338, Bias = -0.005)

The protein prediction scatter plots with the MicroNIR spectrometers are not well correlated with actual protein. Therefore the unmodified MicroNIR spectrometers are not capable of achieving ICAR accuracy for an online milk sensor. For further discussions see section 5.1 of the discussion section. To meet ICAR accuracy for an on milk sensor the hardware of the MicroNIR spectrometers needed to be modified. Details on the hardware modifications can be found in section 3.4. Section 4.2 shows how the hardware modifications have improved the quality of the spectrum.

4.2 Optimisation results

Due to high prediction errors and low correlations achieved during the laboratory trial modifications were needed to reduce the prediction errors. The following sections illustrate how hardware changes have improved the quality of the spectra used to predict milk composition.

4.2.1 Lamp intensity

To reduce the milk composition prediction errors the amplitude of the milk spectrum needed to be increased to improve its signal to noise ratio. The MicroNIR spectrometers were modified by powering the lamps with an external power supply to increase the light output. The lamp voltage was increased from 2.5 V to 4.8 V by using an external power supply (see section 3.4.1 for a schematic). Increasing the lamp voltage increased the light output by an average of 300% for the MicroNIR2200 as illustrated in Figure 36.

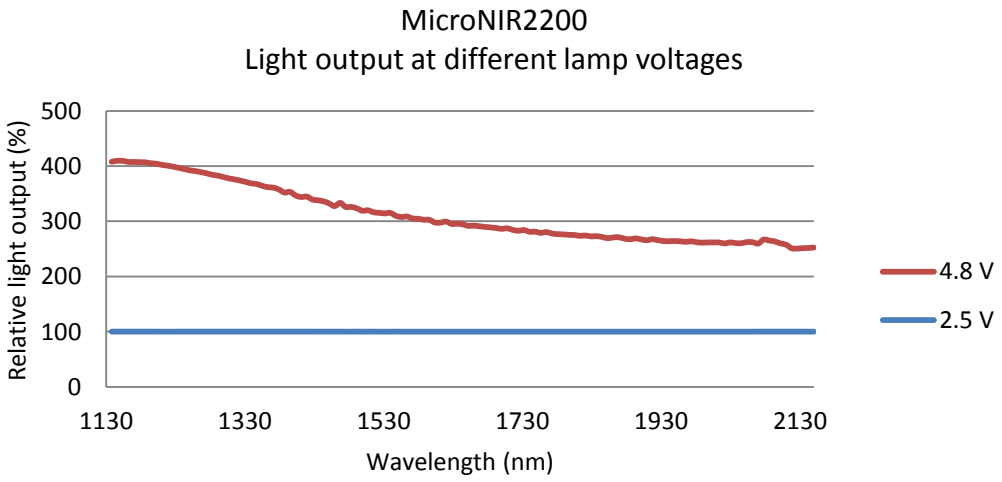

Figure 36: MicroNIR2200 Normalised spectrum of white reference at the original lamp voltage of 2.5 V and the modified 4.8 V.

The MicroNIR1700 spectrum intensity increased more than the MicroNIR2200. The relative light output increased by a maximum of 480% for the MicroNIR1700 spectrometer as illustrated in Figure 37.

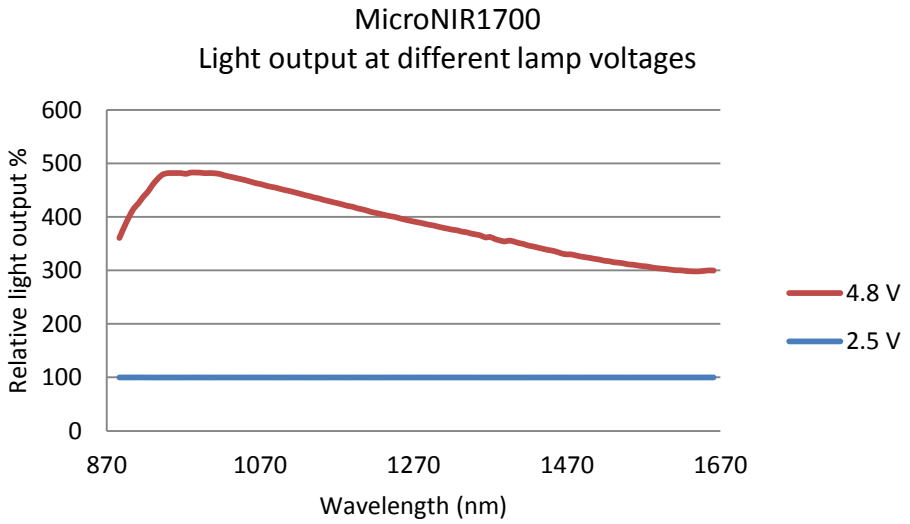

Figure 37: MicroNIR1700 Normalised spectrum of white reference at the original lamp voltage of 2.5 V and the modified 4.8 V.

The total intensity over all wavelengths increases as the temperature of the bulbs is increased. As the lamp output is increased there is a slight shift in the spectrum output. This shift causes the peak intensity of the spectrum to decrease in wavelength as the lamps are operated at a higher temperature. The reason why the location of the maximum light output shifts as the lamp voltage is varied is described by Wien's displacement law (4-1). The location of the peak intensity of a black body spectrum at temperature T is located at λ_{peak} .

$$\lambda_{peak} = 2.898 \times 10^{-3} / T \qquad \textbf{(4-1)}$$

It was discovered experimentally in 1893 by Wien and is known Wein's Law (Kenyon 2008).

Significant gains in amplitude were achieved across the spectrum for both spectrometers. The spectrum amplitude increased by approximately 200% over the entire spectrum. A comparison of the milk spectrum taken at 2.5 V and 4.8 V is illustrated in Figure 38 and Figure 39 for the MicroNIR2200 and MicroNIR1700 respectively.

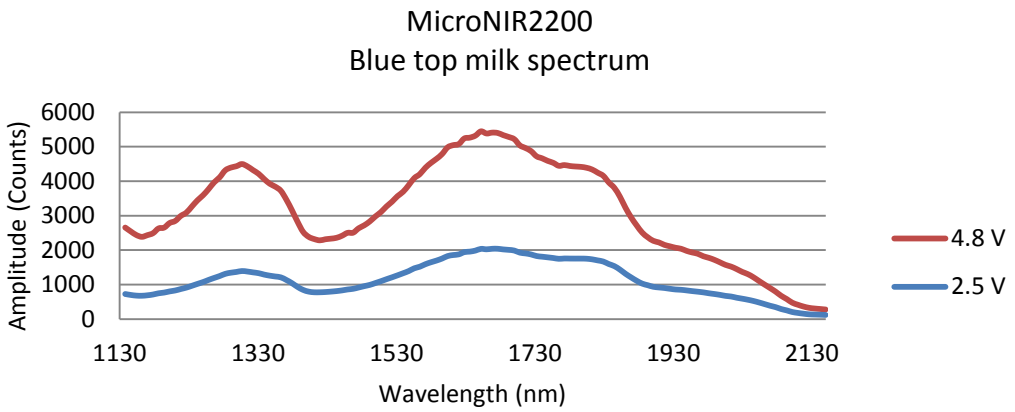

Figure 38: MicroNIR2200 comparison of milk spectrum intensity at different lamp voltages

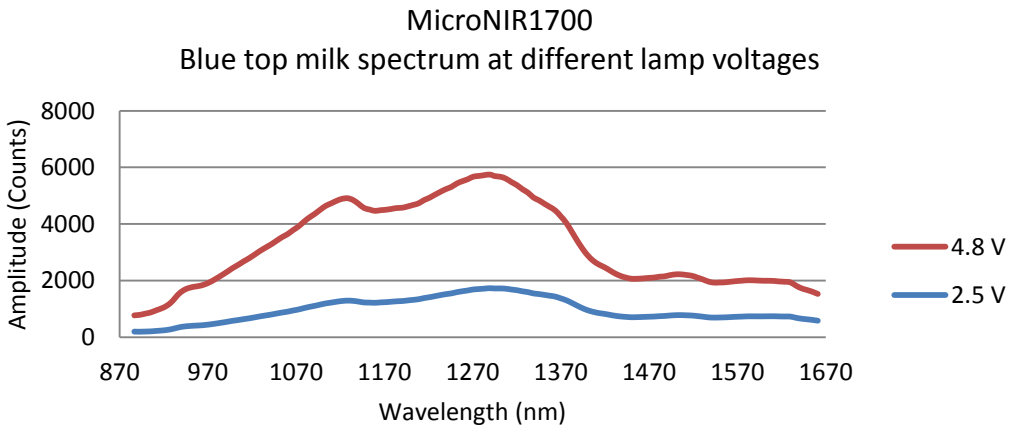

Figure 39: MicroNIR1700 comparison of milk spectrum intensity at different lamp voltages

A side-effect of increasing the lamp voltage to the spectrometers is it causes the temperature of the spectrometers to increase to 50°C. The high temperature causes the dark current to excessively rise, adding noise to the spectrum. Section 4.2.2 illustrates what effect temperature has on the spectrum

4.2.2 Spectrum at different temperatures

The high operating temperature causes the inter pixel wobble to increase. Figure 40 illustrates the water spectrum and dark current at two temperatures for the MircoNIR2200 spectrometer. When the temperature is at 35°C there is a noticeable wobble present on the water spectrum and dark current. When the spectrometer is cooled down to 12°C the wobble is noticeably reduced.

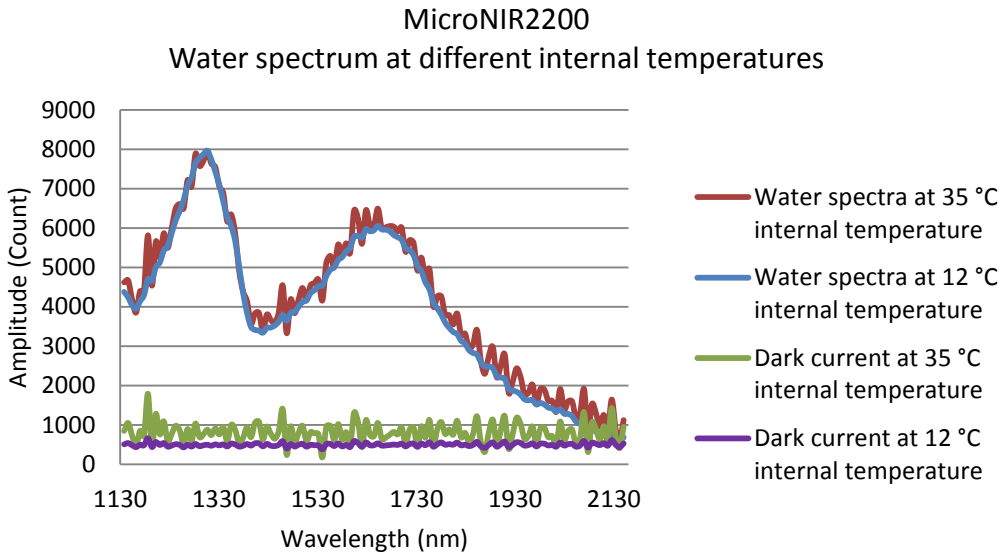

Figure 40: MicroNIR2200 water spectrum measured at different internal temperatures

The wobble present on the spectrum is caused by variations in the inter-pixel photometric response. Cooling the MicroNIR2200 spectrometer reduces the inter-pixel wobble which helps to produce a smooth spectrum.

Cooling the MicroNIR1700 spectrometer does not make a noticeable difference to the water spectrum at two different temperatures as illustrated in Figure 41. This is caused by the variation in pixel to pixel photometric response remaining constant for different temperatures. The MicroNIR1700 spectrometer does not have the same temperature dependency as the MicroNIR2200 spectrometer.

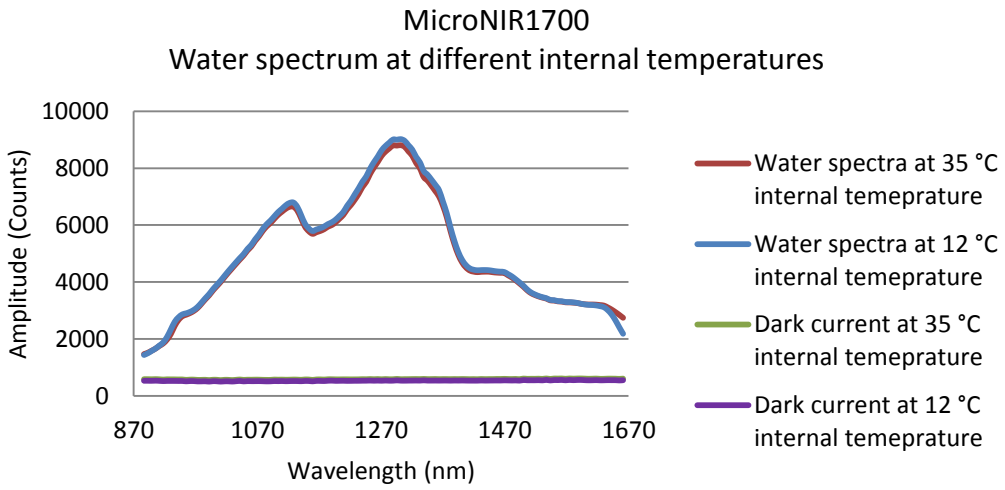

Figure 41: MicroNIR1700 water spectrum measured at different internal temperatures

4.2.3 Dark current

To quantify the behaviour of the photo detector arrays an experiment was done measuring the dark current at different temperatures. An internal temperature range from -10°C to 35°C was investigated to identify how the dark current was affected by temperature.

The extended MicroNIR2200 spectrometers dark current is not very temperature stable above room temperature. The dark current exponentially increases as temperature increases. The mean dark current for the MicroNIR spectrometers is plotted in Figure 42.

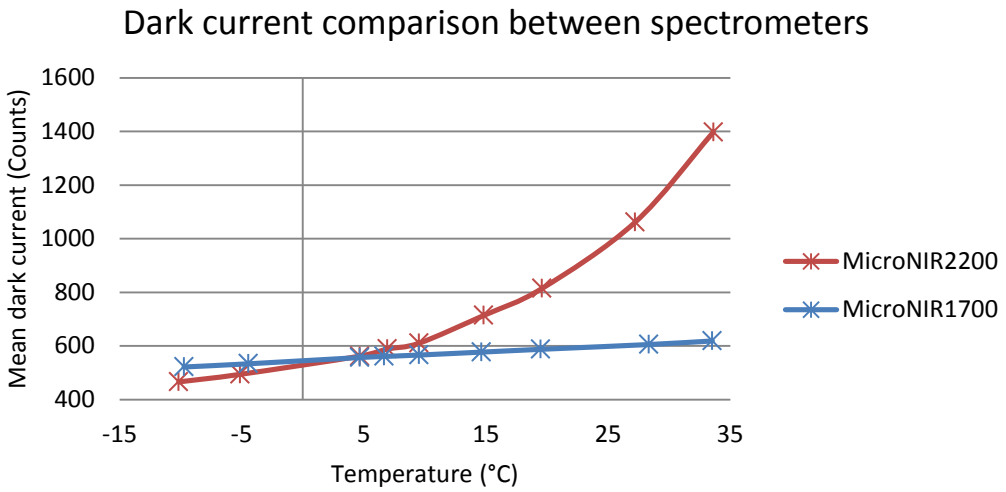

Figure 42: Mean dark current at different internal temperatures

Increasing the internal temperature for the MicroNIR2200 spectrometer increases the variation of the pixel-to-pixel photometric response. As the temperature is increased the mean dark current for each pixel increases at a different rate compared with the neighbouring pixels. This causes the pixel-to-pixel variation to appear unstable as illustrated in Figure 43. This means there is variation in the inter pixel dark current temperature dependencies on the extended InGaAs array of the MicroNIR2200 spectrometer.

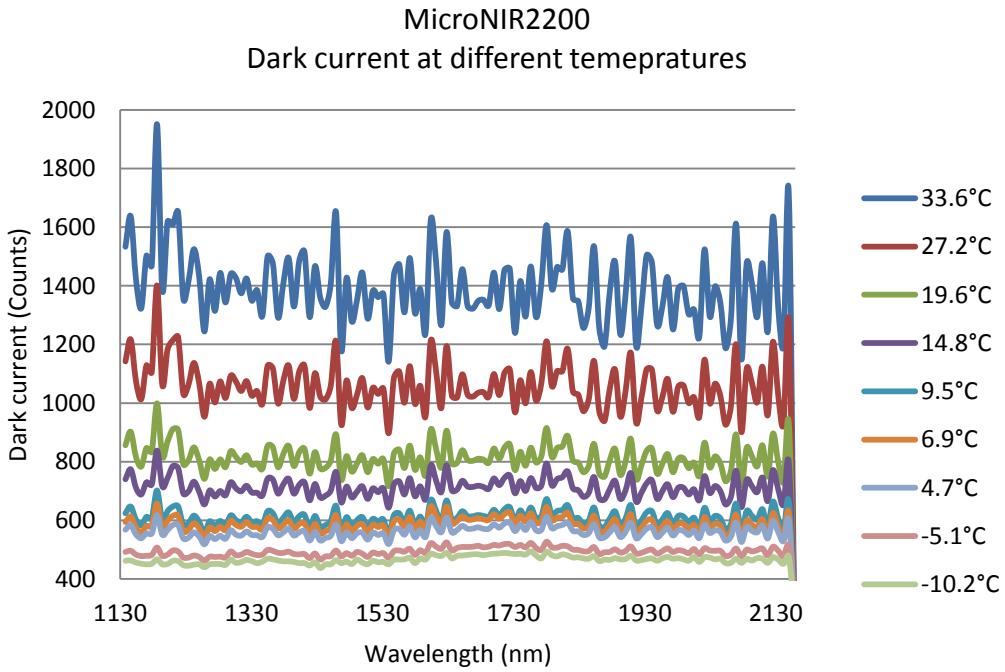

Figure 43: This is a plot of the MicroNIR2200 dark current at various temperatures. As the temperature is increased the baseline dark current increases proportionally to temperature squared. Also the inter pixel variation increases proportionally to the temperature squared.

The MicroNIR1700 dark current is fairly stable with the mean dark current linearly rising with increasing temperature. Also the variation in pixel-to-pixel photometric response remains constant across the array for different temperatures as illustrated in Figure 44. This is the ideal response to variations in temperature and makes the spectrometer more robust to temperature fluctuations.

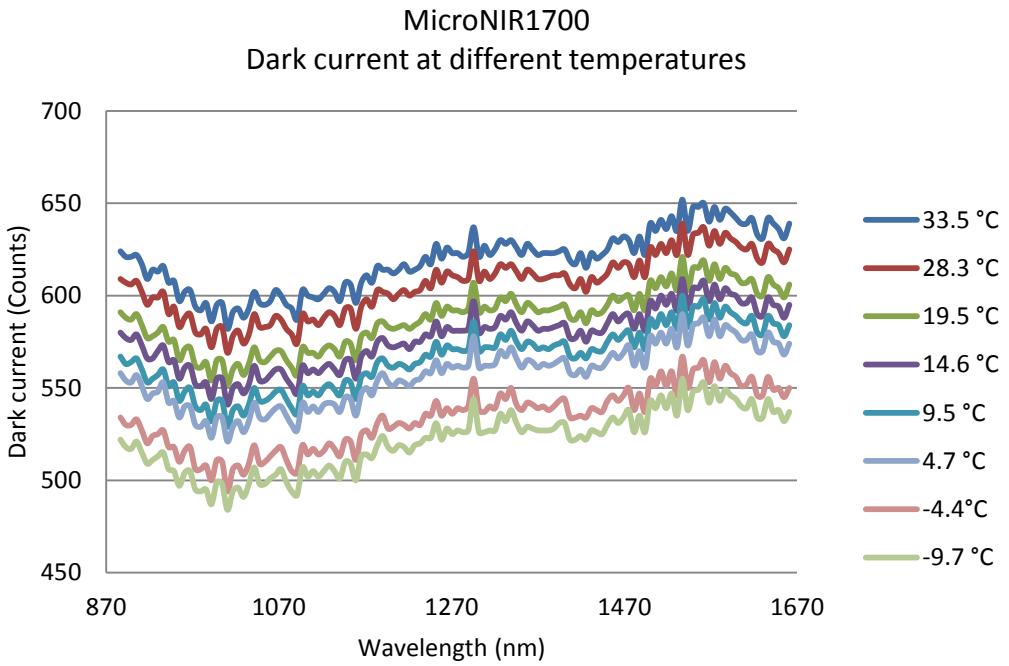

Figure 44: This is a plot of the MicroNIR1700 dark current at various temperatures. As the temperature increases the baseline dark current linearly increases. The inter pixel variation remains constant for all temperatures.

A sensor on a milking robot is typically exposed to a temperature between -5°C to 45°C. The MicroNIR2200 dark current starts to rise excessively above 30°C adding noise to the spectrum. The stability of the MicroNIR1700 dark current to temperature fluctuations is a significant advantage for the spectrometer.

4.2.4 Spectrum noise results

Both spectrometers have a high level of noise at the outer edges of their detection range. This is caused by the reduced responsivity at the outer wavelengths of the InGaAs array. The reduced responsivity causes the amplitude of the received signal to decrease, which decreases the signal to noise ratio.

The MicroNIR1700 spectrometer has a lot of spectrum noise from the 870 – 1000 nm region as illustrated in Figure 45. The majority of the MicroNIR2200 spectrum noise occurs in the 1130 – 1400 nm and the 2100 – 2150 nm regions as illustrated in Figure 46.

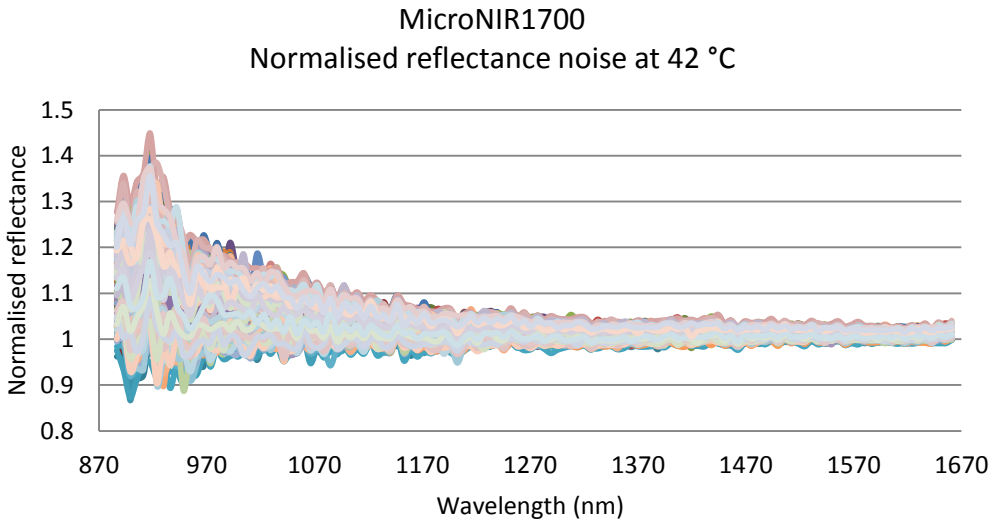

Figure 45: Normalised reflectance spectrum of 200 white references using the MicroNIR1700 spectrometer at the normal operating temperature

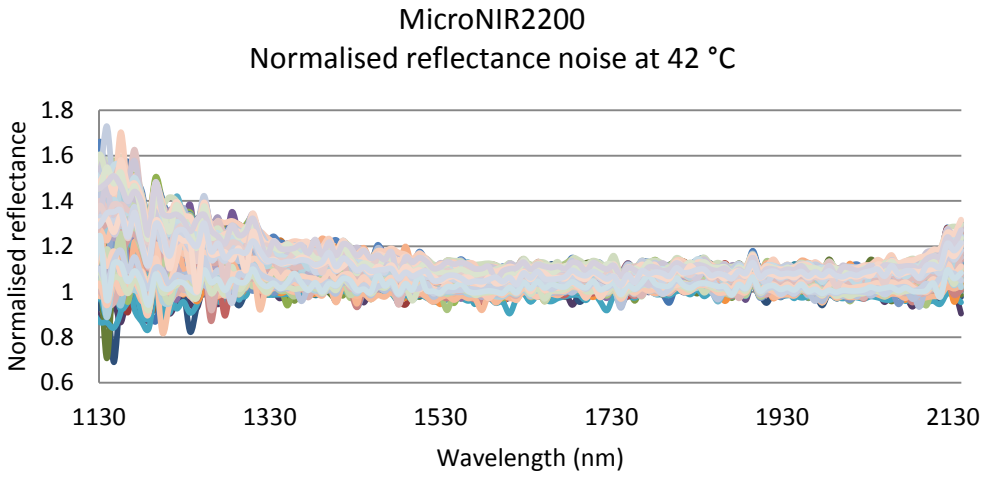

Figure 46: Normalised reflectance spectrum of 200 white references using the MicroNIR2200 spectrometer at the normal operating temperature

Due to the high variance that was present in the normalised reflectance spectrum, cooling was required to reduce the spectrum noise and the operating temperature of the spectrometers. The spectrum noise at 20°C of the MicroNIR1700 and MicroNIR2200 spectrometers is plotted in Figure 47 and Figure 48 respectively. Cooling the spectrometers to 20°C significantly reduces the dark current variation for both spectrometers at the outer wavelengths of the array.

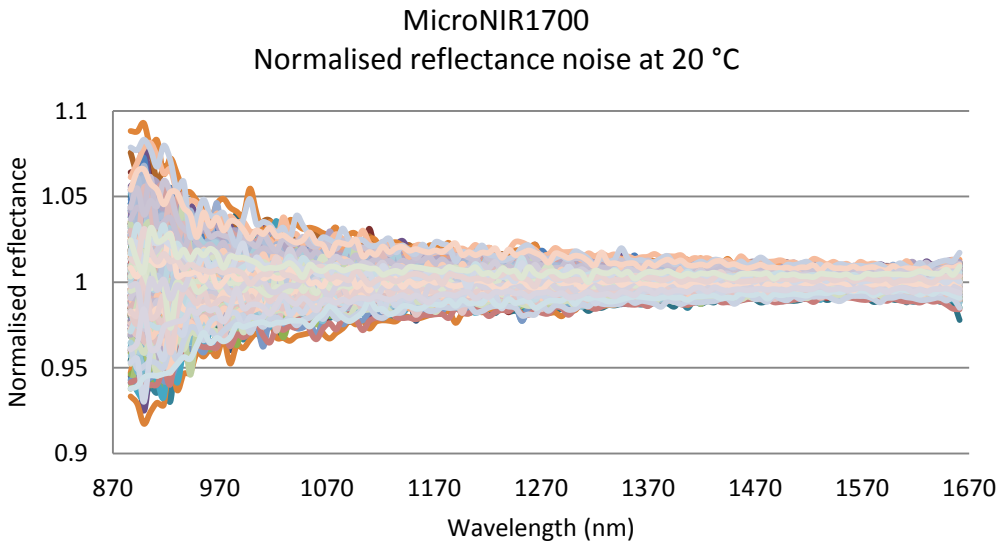

Figure 47: Normalised reflectance spectrum of 200 white references using the MicroNIR1700 spectrometer at 20°C

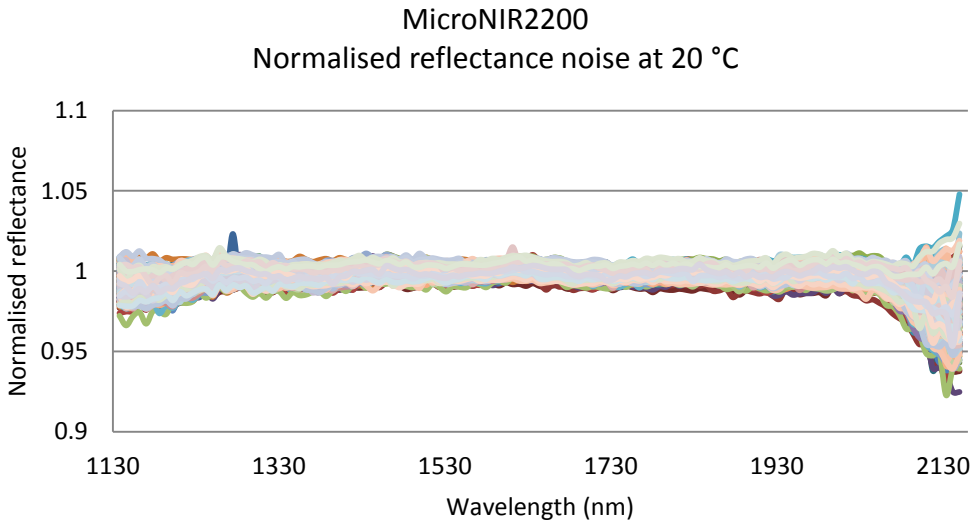

Figure 48: Normalised reflectance spectrum of 200 white references using the MicroNIR2200 spectrometer at 20°C

The biggest improvement in spectrum quality was achieved with the MicroNIR2200 spectrometer with the peak signal to noise ratio improving from 31 to 228 when the temperature is reduced from 42°C to 20°C. The average signal to noise ratio for the spectrum improved from 21 to 169 for the MicroNIR2200. A significant improvement occurs at the shorter wavelength range of 1130 - 1400 nm for the MicroNIR2200 spectrometer when the internal temperature is cooled to 20°C. Cooling the MicroNIR1700 improved the maximum signal to noise ratio from 111 to

246. The average signal to noise ratio for the spectrum improved from 57 to 146. The signal to noise ratio for the MicroNIR2200 and MicroNIR1700 spectrometers is plotted in Figure 49 and Figure 50.

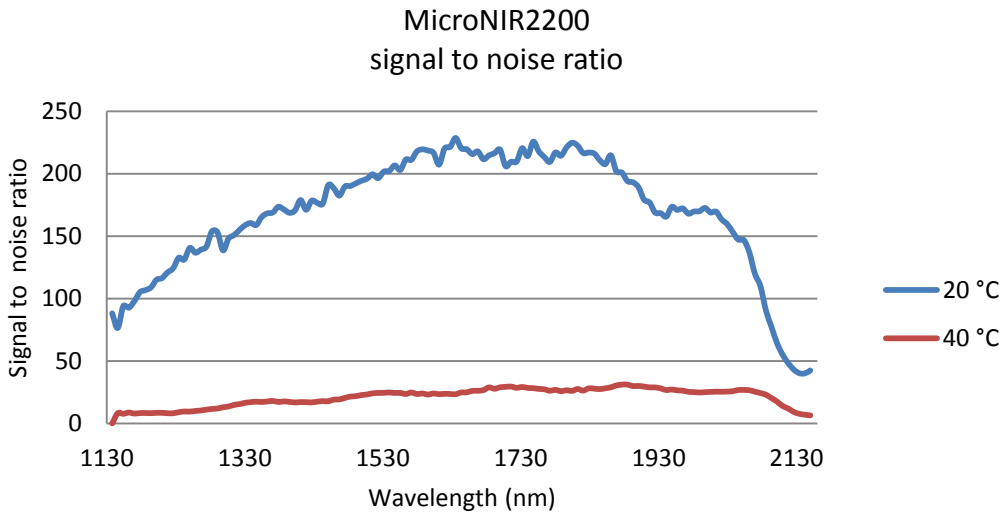

Figure 49: MicroNIR2200 signal to noise ratio for 200 individual scans at two different temperatures

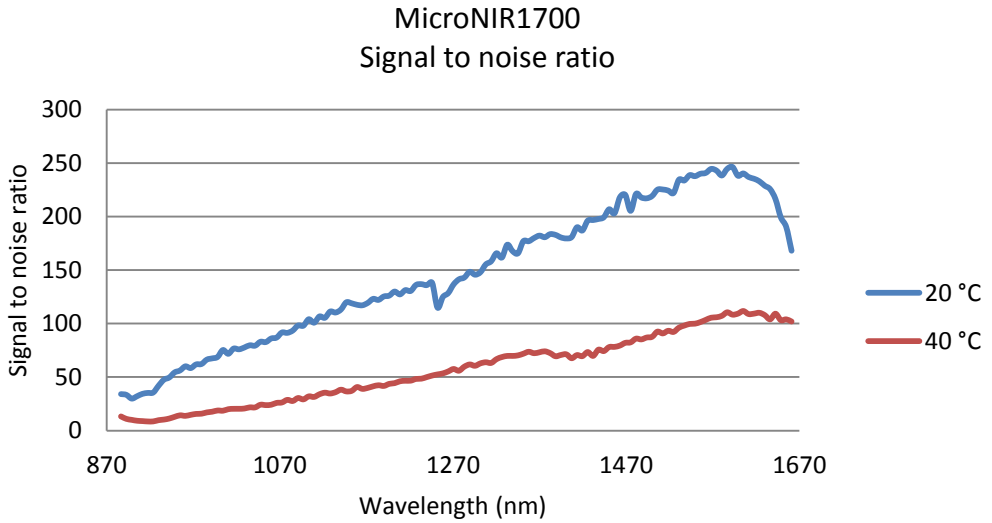

Figure 50: MicroNIR1700 signal to noise ratio for 200 individual scans at two different temperatures

4.3 Farm trial results

This chapter presents the results from 590 milk samples collected across three different farms during November. Calibration models were built from two farms and were validated on the third farm. This process was repeated for all three farms so each farm was independently tested with a different calibration model. Prediction models were built for fat and protein for each farm.

4.3.1 Farm trial spectra

The normalised reflectance spectra of the respective milk samples acquired are shown in Figure 51 (870 – 1660 nm region) and Figure 52 (1130 - 2160 nm region). These are the wavelength regions from the MicroNIR1700 and MicroNIR2200 spectrometers respectively.

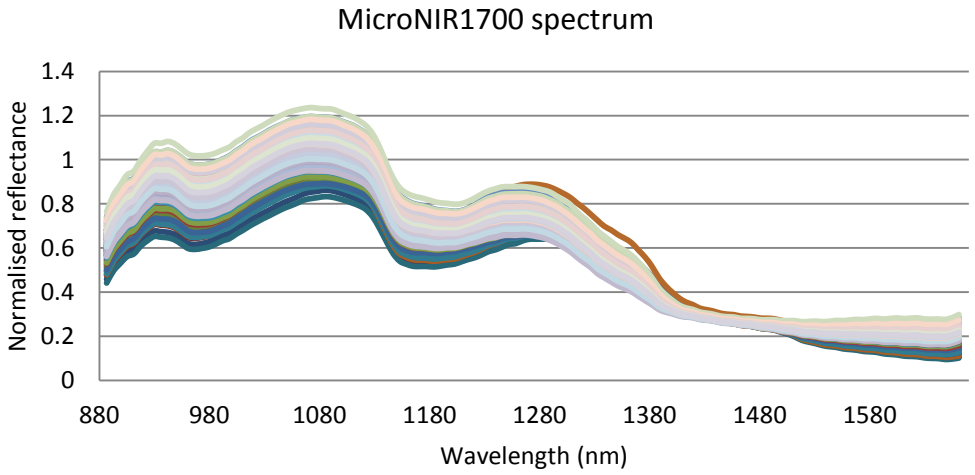

Figure 51: Spectra of milk samples acquired during the experiment using the MicroNIR1700 spectrometer.

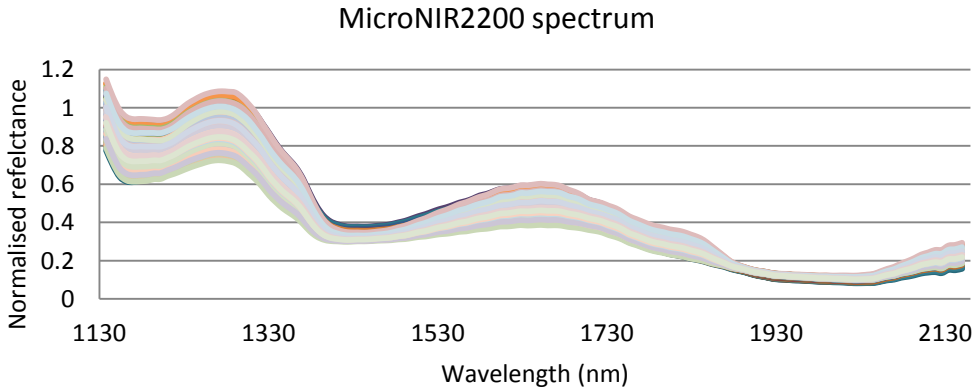

Figure 52: Spectra of milk samples acquired during the experiment using the MicroNIR2200 spectrometer.

4.3.2 Farm trial results - Lipids

Collected milk samples varied widely in their fat content; Figure 53, Figure 54 and Figure 55 show histograms of the samples collected. The range of milk fat content in the samples was from 0.61% to 13.81%, the average value being 5.09% with a standard deviation of 2.47.

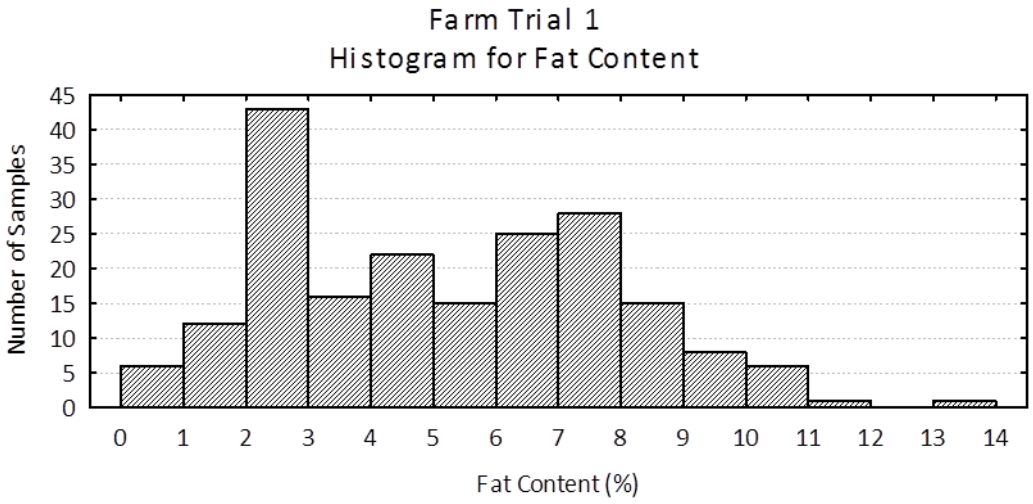

Figure 53: Fat histogram of samples collected during the first farm trial

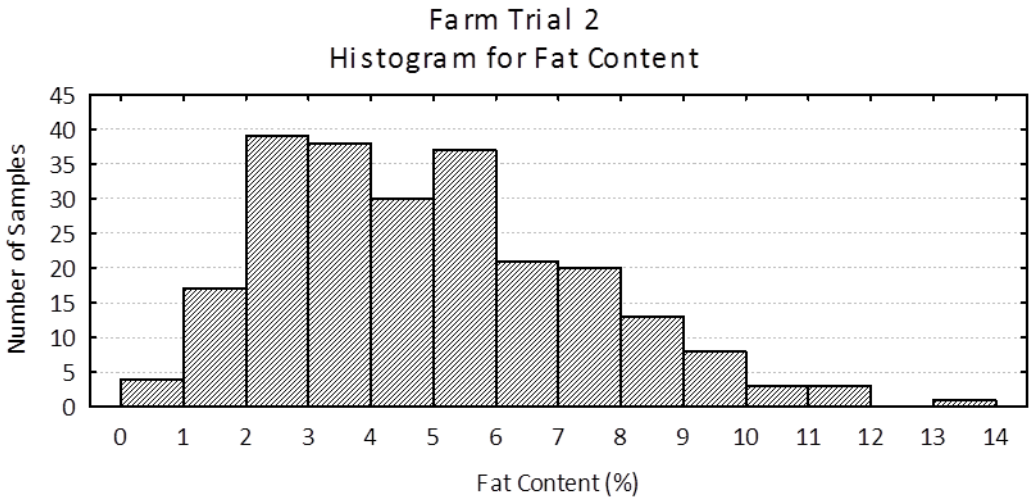

Figure 54: Fat histogram of samples collected during the second farm trial

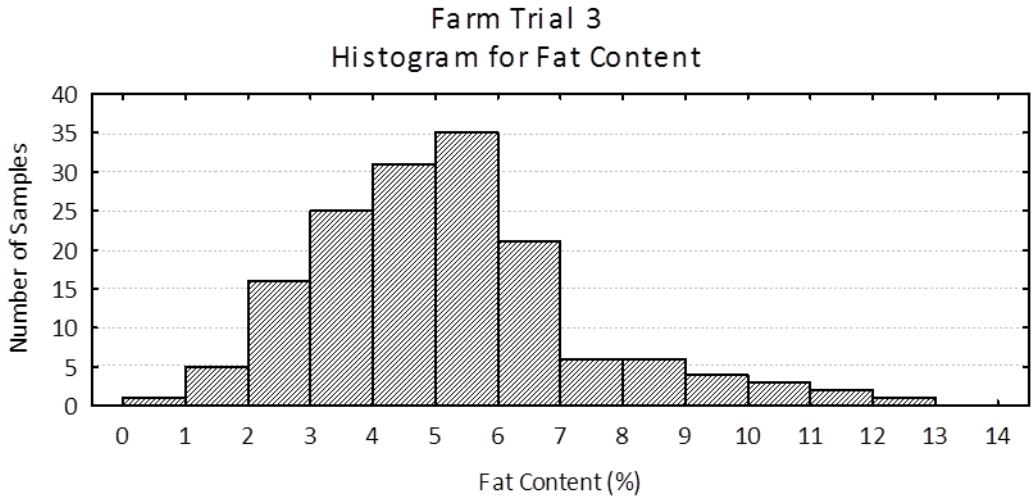

Figure 55: Fat histogram of samples collected during the third farm trial

First calibration models were built for the pooled data set. The pooled data set contained data from all three farms. For the pooled data set the fat prediction results for the MicroNIR1700 spectrometer are plotted in Figure 56. The results are plotted on a scatter plot with the known fat measured by Testlink. This fat prediction is outstanding with all the data points situated on the one to one line. The RMSE for the validation data set is 0.061 and the r^2 = 0.999 which indicates the fit is superb.

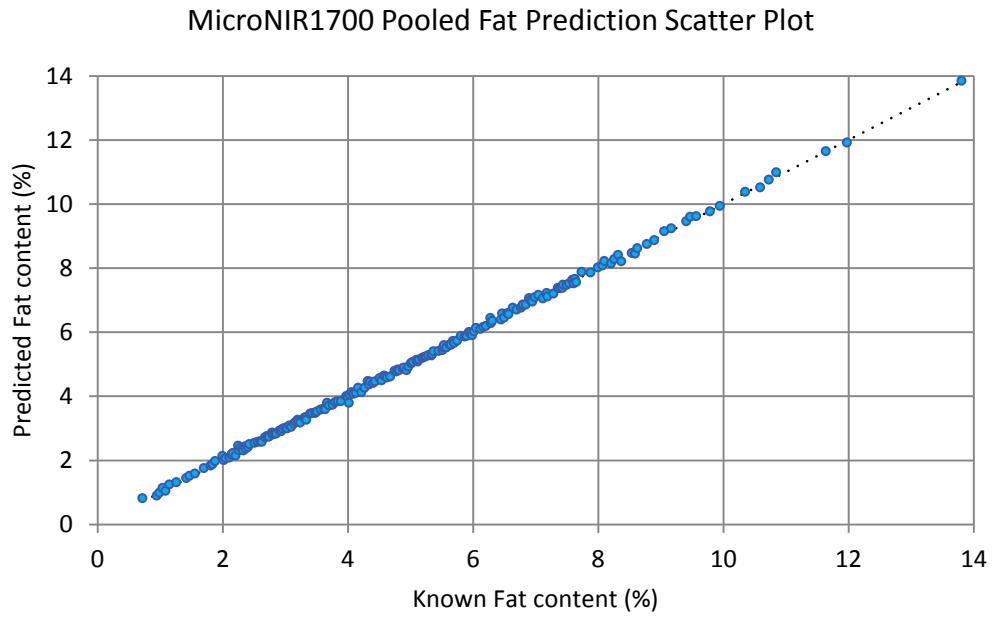

Figure 56: MicroNIR1700 fat prediction scatter plot of the combined data from all three farms (RMSE = 0.061, r^2 = 0.9994, Bias = 0.005)

Variations in fat content cause the spectrum to shift in intensity. Samples with high fat content reflect more light across the spectrum. Figure 57 and Figure 58 illustrates how milk samples with higher fat content have a higher normalised reflectance spectrum for both spectrometers.

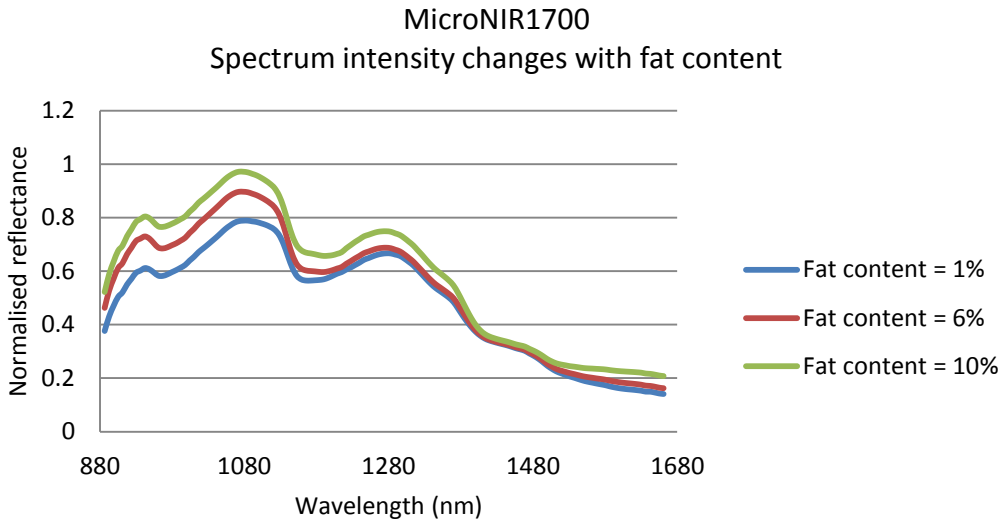

Figure 57: MicroNIR1700 spectrum intensity changes with changing fat content. Milk sample with high fat content have a high normalised reflectance spectrum than samples with low fat content

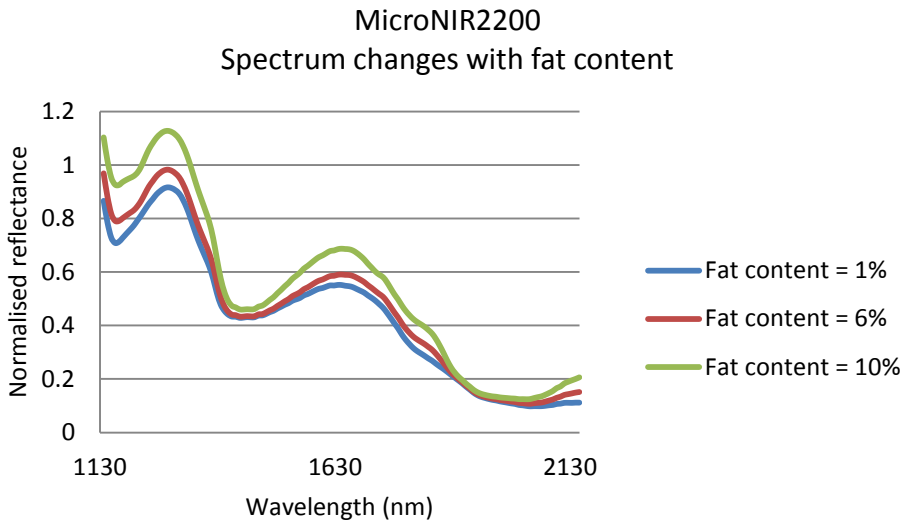

Figure 58: MicroNIR2200 spectrum intensity changes with changing fat content. Milk sample with high fat content have a high normalised reflectance spectrum than samples with low fat content

The 11Ants analytic software determined Partial Least Squares to give the optimum calibration fit. A total of 13 latent variables were used to predict fat with the MicroNIR1700 spectrum. The most significant wavelength bands for milk fat determination are the 892 – 1041 nm and 1140 –

1295 nm bands. The relative wavelength influence for fat prediction is illustrated in Figure 59.

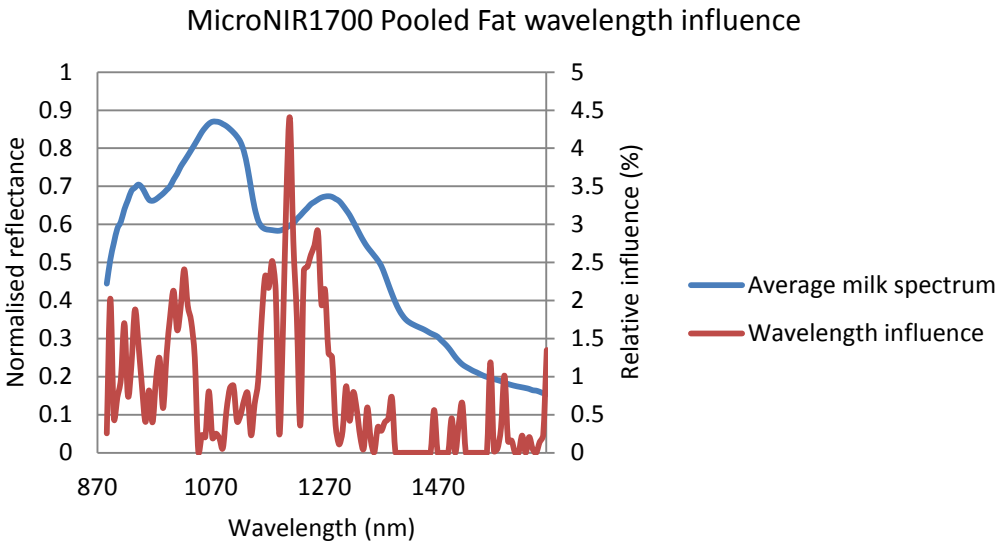

Figure 59: MicroNIR1700 wavelength influence for fat prediction using the combined data from all three farms

The MicroNIR2200 predicted fat with similar prediction errors that were obtained with the MicroNIR1700 spectrometer. The MicroNIR2200 spectrometer predicted the fat composition with a RMSE of 0.071 and an r^2 of 0.999 as illustrated in Figure 60. This prediction error is slightly higher than the MicroNIR1700 which has a RMSE of 0.061.

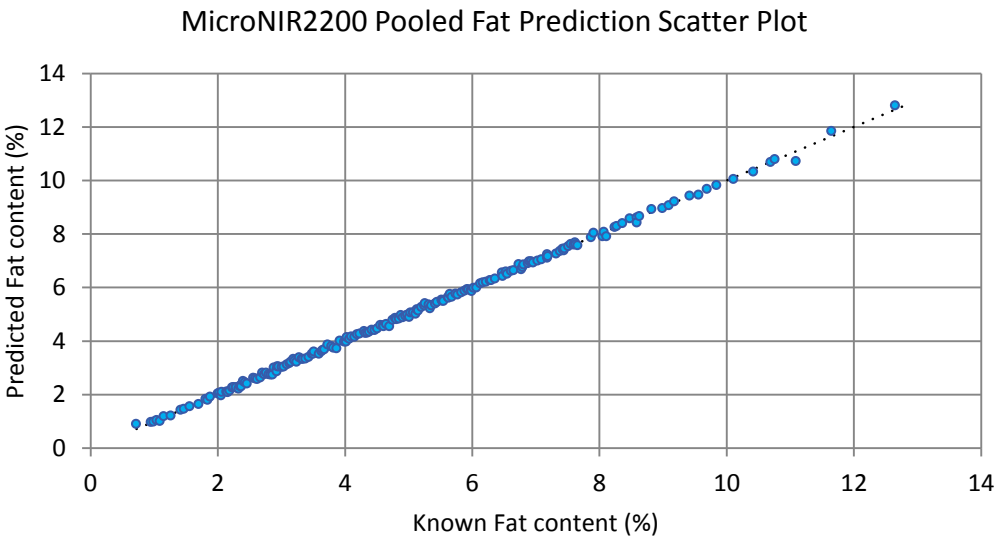

Figure 60: MicroNIR2200 fat prediction scatter plot of the combined data from all three farms (RMSE = 0.071, r^2 = 0.999, Bias = 0.002)

The 11Ants analytics software predicted the fat composition with the MicroNIR2200 spectrometer using a Gaussian process. There are two distinct bands used to predict the fat composition with the MicroNIR2200. The wavelength bands are located at 1162 – 1395 nm and 1571 – 1876 nm. Figure 61 illustrates the relative wavelength influence for fat prediction. The first wavelength band located at 1162 – 1395 nm coincides with the wavelength band used for fat prediction for the MicroNIR1700 which suggest is a very important band for prediction of fat. Wavelengths located at the two water absorption peaks at 1450 nm and 1935 nm do not significantly contribute to the prediction of fat. This is caused by the water absorbing light in the near infrared region. The strong water absorption peaks causes the normalised reflectance values to converge at 1450 nm and 1935 nm. This reduces the variations in normalised reflectance for different fat compositions which minimises the influence these wavelengths have for the prediction of fat.

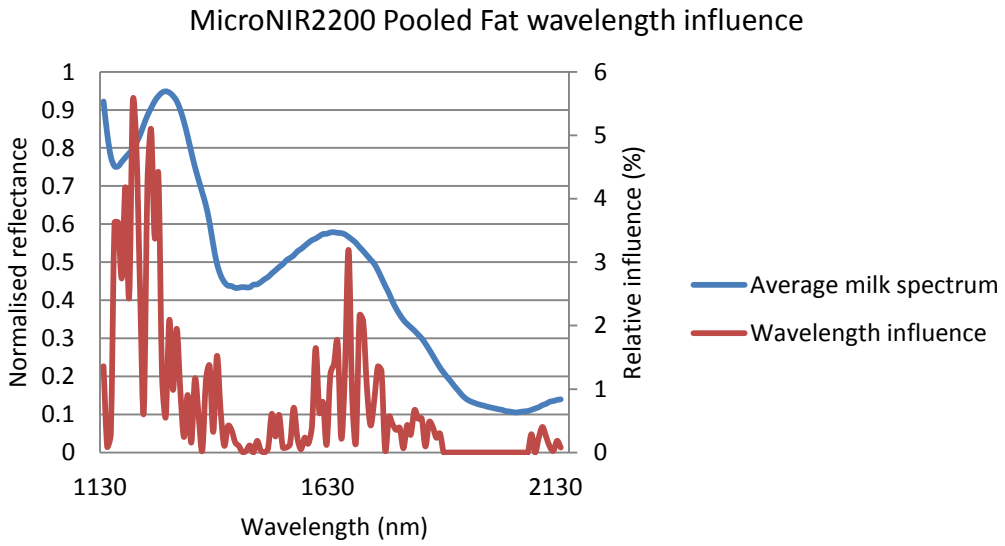

Figure 61: MicroNIR2200 wavelength influence for fat prediction using the combined data from all three farms.

The fat predictions obtained from both spectrometers using the pooled data were outstanding. The MicroNIR1700 predicted the milk fat composition better than the MicroNIR2200. Both spectrometers can predict the milk fat composition within ICAR for an online sensor.

4.3.2.1 Inter farm lipid transferability

Next an independent test was done to test the ability of the calibration model to be transferred across different farms. New calibration models were built on two of the three farms tested. The calibration model created from the two farms was tested on the third farm for independent validation. A tabulated summary of the milk fat composition collected from the different farms is summarised in Table 3. The results of the calibration models tested on an independent farm are summarised in Table 4.

The calibration models between farms are transferable with very low RMSE for both spectrometers. The calibration models have a RMSE of 0.064% to 0.087% for the different farm trials. The MicroNIR1700 predicted fat composition better than the MicroNIR2200 across all farm trials. The prediction errors for the independent farm calibrations are slightly higher than the pooled data set. This is expected because the calibration model does not include regional variations. All of the calibration models have coefficient of determination of 0.999 which is outstanding. Also there is no apparent bias present in the prediction. This suggests the fat calibration models are transferrable between farms.

Table 3: Descriptive statistics off the fat composition of samples obtained during the farm trials (F1 represents the first farm)

Data set	F2 + F3	F1 + F3	F1 + F2	F1	F2	F3
Mean	5.02	5.19	5.08	5.20	4.95	5.08
Median	4.88	5.02	4.81	4.99	4.785	5
Standard Deviation	2.35	2.47	2.58	2.72	2.47	2.17
Sample Variance	5.51	6.13	6.66	7.39	6.09	4.71
Range	12.91	12.81	13.2	12.81	12.87	11.75
Minimum	0.9	0.61	0.61	0.61	0.94	0.9
Maximum	13.81	13.42	13.81	13.42	13.81	12.65

Table 4: This is a summary of the milk fat prediction RMSE obtained during this trial. The calibration & validation rows signify what data set was used for calibration and validation respectively.

Sensor	MicroNIR1700	MicroNIR1700	MicroNIR1700	MicroNIR2200	MicroNIR2200	MicroNIR2200
Wavelength (nm)	880 - 1661	880 - 1661	880-1661	1137-2150	1137-2150	1137-2150
Calibration	F2 & F3	F1 & F3	F1 & F2	F2 & F3	F1 & F3	F1 & F2
Validation	F1	F2	F3	F1	F2	F3
Calibration set Count	389	351	431	389	351	431
Validation set Count	196	233	155	196	233	155
Root Mean Squared Error (RMSE)	0.064	0.070	0.067	0.085	0.077	0.087
Mean Absolute Error (MAE)	0.049	0.056	0.054	0.062	0.058	0.049
Mean Error (Bias)	-0.004	-0.022	-0.009	0.013	0.029	0.025
Pearson Correlation Coefficient (r)	0.999	0.999	0.999	0.999	0.999	0.999
r^2	0.999	0.999	0.999	0.999	0.999	0.999

4.3.3 Farm trial results - Protein

The collected milk samples varied widely in their protein content. The range in protein content in the samples was from 2.15% to 4.87%; the average being 3.68% with a standard deviation of 0.35. Histograms of the protein content for the three farm trials are illustrated in Figure 62, Figure 63 and Figure 64. These are typical protein distributions for herd testing.

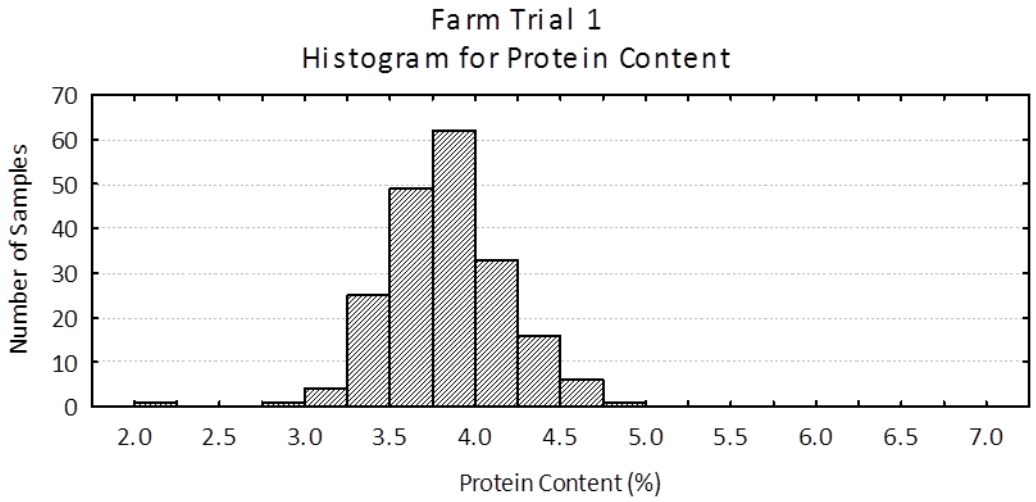

Figure 62: First farm protein distribution

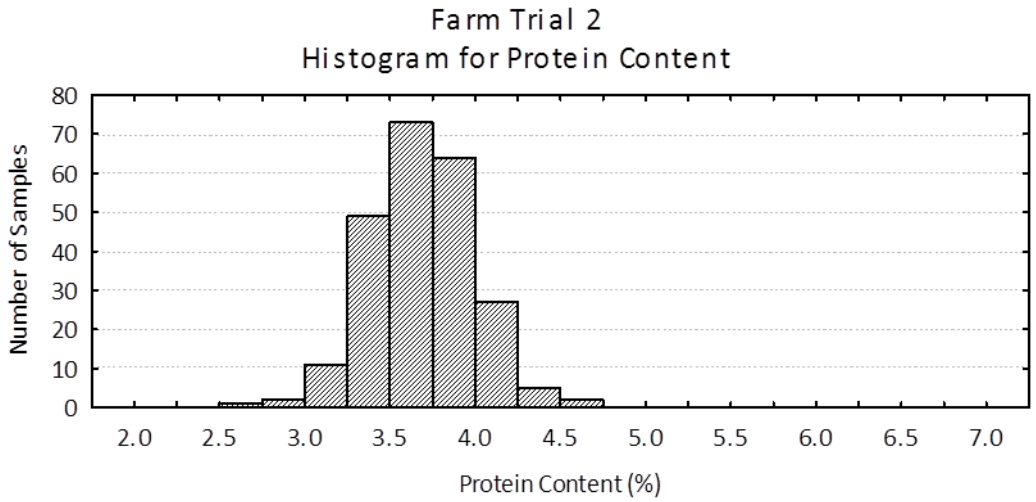

Figure 63: Second farm protein distribution

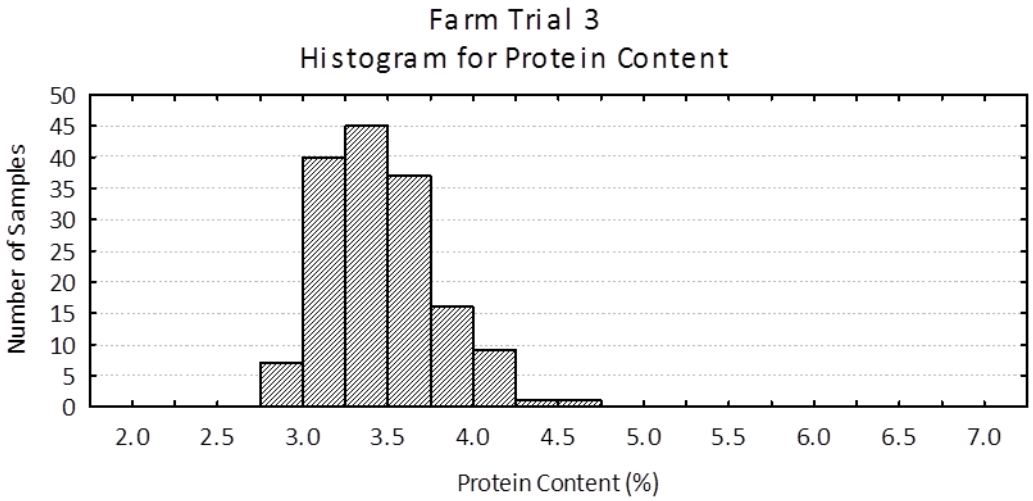

Figure 64: Third farm protein distribution

The MicroNIR1700 spectrometer was used to predict the protein composition. The spectrum of milk from 880 nm to 1660 nm at 6 nm intervals was used to predict the protein composition. First a calibration data set was built by listing the samples from lowest to highest and using the first two samples for calibration. Next the third sample was removed from the calibration data set and was used for the validation data set.

A protein prediction scatter plot using the MicroNIR1700 spectrometer is illustrated in Figure 65. The MicroNIR1700 spectrometer was able to predict protein with a RMSE of 0.151 for a total of 195 samples. This prediction error meets ICAR precision tolerances for protein.

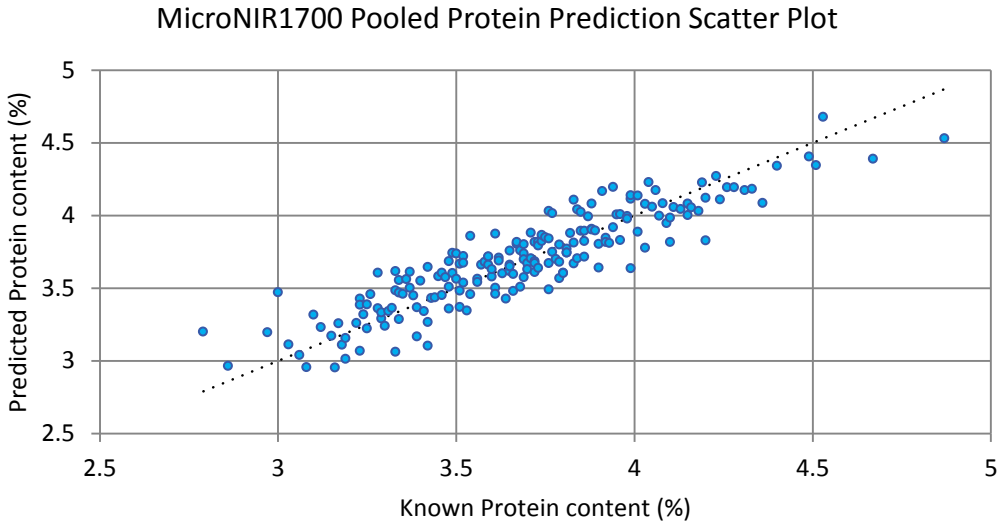

Figure 65: MicroNIR1700 protein prediction scatter plot of the combined data from all three farms (RMSE = 0.151, r^2 = 0.817, Bias = 0.008)

To predict the protein composition a partial lease squares model was used. The model used 15 components to predict protein. The most significant wavelength used to predict protein was the 1177 nm. Other significant bands used to predict protein were 950 – 990 nm, 1035 – 1135 nm, 1260 – 1370 nm, 1300 – 1360 nm and 1550 – 1630 nm. Figure 66 illustrates the relative wavelength influences for the prediction of protein using the MicroNIR1700 spectrometer.

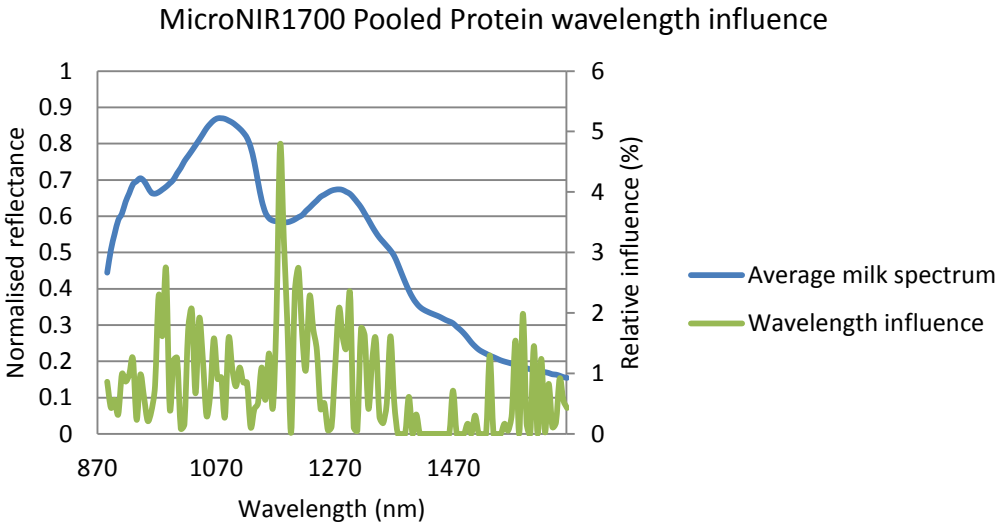

Figure 66: MicroNIR1700 wavelength influence for protein prediction using the pooled farm data.

Next the MicroNIR2200 capability to predict protein was tested. The same calibration data set used for the MicroNIR1700 pooled data set was used to build a calibration model with the MicroNIR2200. The spectrum of milk from 1137 nm to 2150 nm at 8 nm intervals was used to predict the protein composition. Scatter plots of predicted protein against actual protein concentration are shown in Figure 67 for the MicroNIR2200 spectrometer. The MicroNIR2200 could predict protein composition slightly better than the MicroNIR1700 spectrometer with a RMSE of 0.138 and r^2 of 0.842.

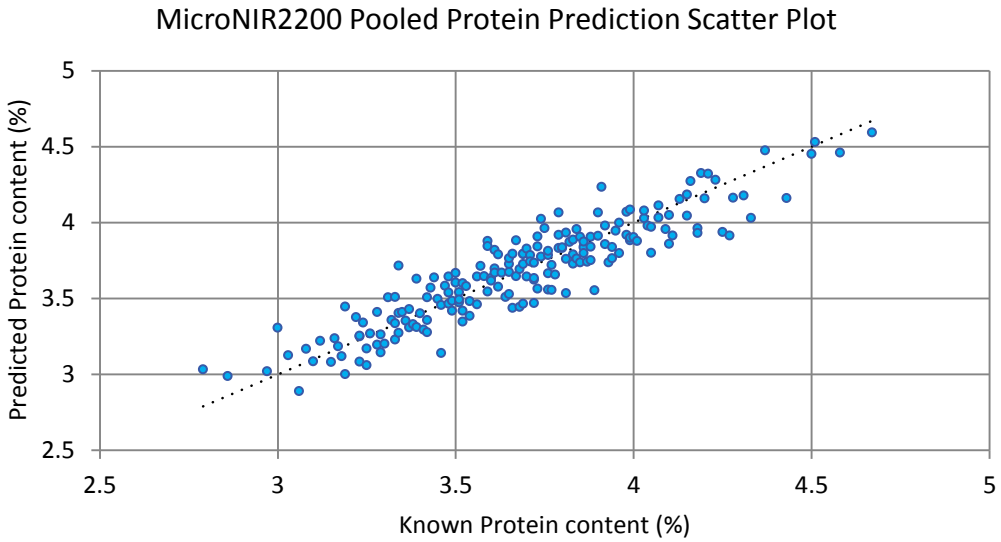

Figure 67: MicroNIR2200 protein prediction scatter plot of the combined data from all three farms (RMSE = 0.138, r^2 = 0.842, Bias = -0.006)

The significant wavelength bands that were used to predict protein are located between 1130 – 1400 nm and 1570 – 1860 nm. Figure 68 illustrates the relative wavelength influences for the prediction of protein. The 1130 – 1400 nm wavelength band used to predict protein with the MicroNIR2200 coincides with the wavelength influences for the MicroNIR1700. This suggests this is a significant region for the prediction of protein.

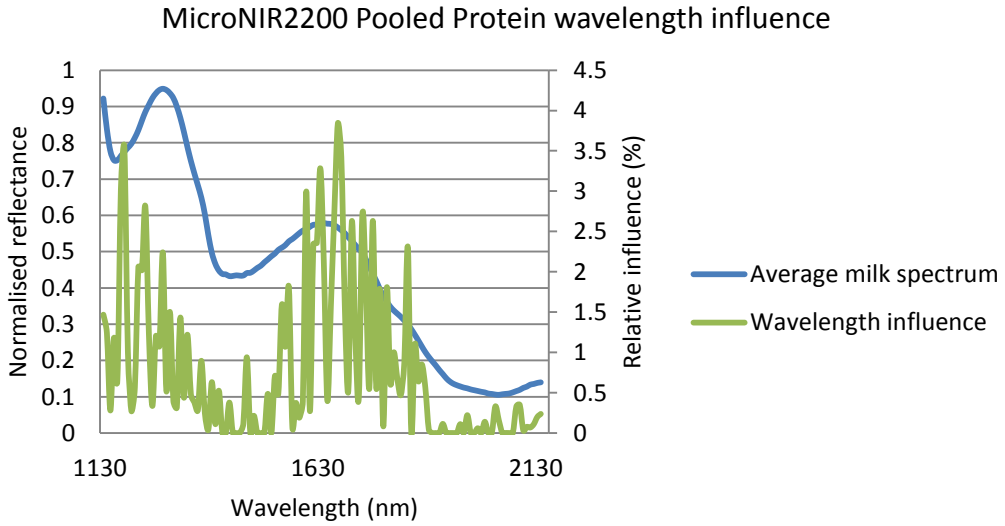

Figure 68: MicroNIR2200 wavelength influence for protein prediction using the pooled farm data.

4.3.3.1 Inter farm protein transferability.

It is important to test the capability of the calibration models ability to be transferred across farms. It is preferable to use a single calibration model to predict milk composition because it reduces operating cost.

To test the calibration models ability to be transferred across farms a calibration data set using spectrum from two farms was used to build a model. The calibration model was then tested on the third farm. This provides an independent test because no samples from the third farm were used for the calibration data set.

The calibration models are able to be transferred across different farms with a similar prediction error compared with the pooled data set. A protein prediction scatter plot of the MicroNIR1700 and MicroNIR2200 is illustrated in Figure 69 and Figure 70.

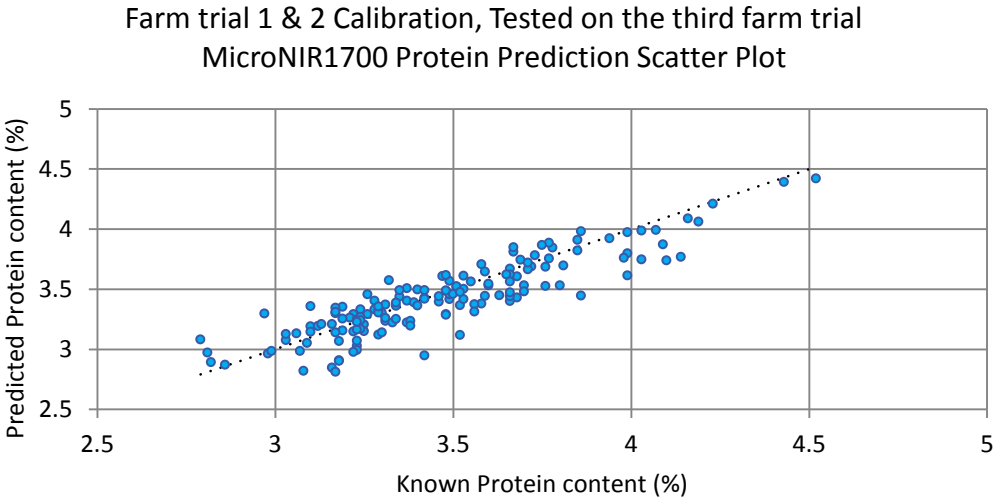

Figure 69: MicroNIR1700 protein prediction scatter plot using the first and second farm trial data as a calibration set and validating the data on the third trial at a different farm (RMSE = 0.154, r^2 = 0.795, Bias = -0.044)

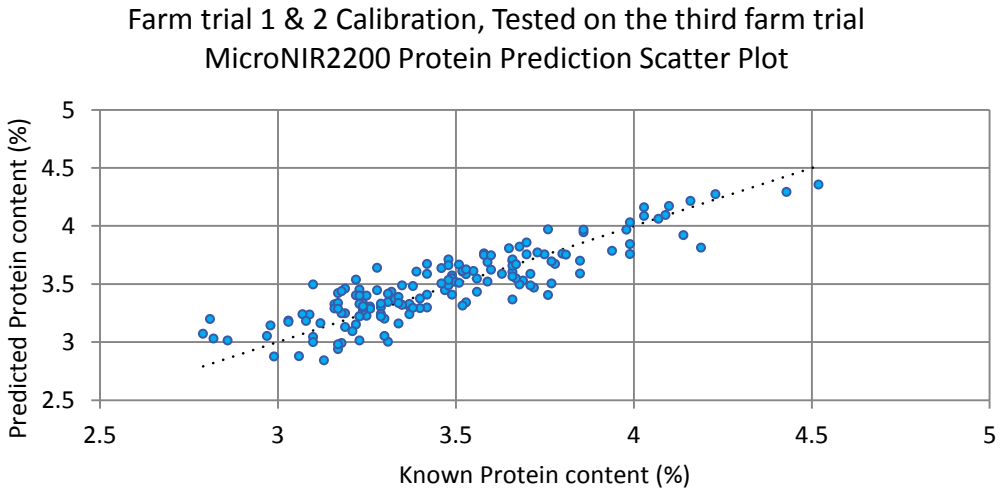

Figure 70: MicroNIR2200 protein prediction scatter plot using the first and second farm trial data as a calibration set and validating the data on the third trial at a different farm (RMSE = 0.153, r^2 = 0.781, Bias = 0.014)

Calibration models were independently tested on all three farms used in the trial. A tabulated summary of the milk fat composition collected from the different farms is summarised in Table 5. A summary of the prediction statistics used for testing the independent calibration models ability to be transferred across the different farms is illustrated in Table 6

The prediction errors obtained for the independent farm validation test are slightly higher than the prediction errors acquired in the pooled data set. This is caused by the calibration data set not including regional variations between the different farms.

Table 5: Descriptive statistics off the protein composition of samples obtained during the farm trials.

Data set	F2 + F3	F1 + F3	F1 + F2	F1	F2	F3
Mean	3.60	3.68	3.76	3.84	3.69	3.45
Median	3.615	3.66	3.76	3.83	3.71	3.42
Standard Deviation	0.33	0.38	0.33	0.35	0.30	0.36
Sample Variance	0.11	0.14	0.11	0.12	0.09	0.13
Range	2.03	2.08	2.28	2.72	2.03	3.08
Minimum	2.59	2.15	2.15	2.15	2.59	2.79
Maximum	4.62	4.87	4.87	4.87	4.62	4.52

Table 6: This is a summary of the protein prediction RMSE obtained during this trial. The calibration & validation rows signify what data set was used for calibration and validation respectively.

Sensor	MicroNIR1700	MicroNIR1700	MicroNIR1700	MicroNIR2200	MicroNIR2200	MicroNIR2200
Wavelength (nm)	880-1661	880-1661	880-1661	1137-2150	1137-2150	1137-2150
Calibration	F2 & F3	F1 & F3	F1 & F2	F2 & F3	F1 & F3	F1 & F2
Validation	F1	F2	F3	F1	F2	F3
Calibration set Count	389	351	431	389	351	431
Validation set Count	196	233	155	196	233	155
Root Mean Squared Error (RMSE)	0.178	0.178	0.154	0.176	0.218	0.153
Mean Absolute Error (MAE)	0.140	0.141	0.117	0.140	0.174	0.124
Mean Error (Bias)	-0.052	0.043	-0.044	0.044	0.050	0.014
Pearson Correlation Coefficient(r)	0.860	0.825	0.892	0.858	0.726	0.884
r^2	0.739	0.681	0.795	0.736	0.526	0.781

5 Discussion

The results presented in the previous chapter will now be discussed in detail. Section 5.1 explains why the original spectrometers could not predict the composition of milk. Section 5.2 explains how the hardware modifications improve the spectrometers ability to predict milk composition. Section 5.3 discusses the results achieved during the farm trial with the modified spectrometers. Section 5.4 relates the results obtained during the farm trial to relevant published material. Section 5.5 describes improvements which could be made in the future to improve the spectrometers ability to predict milk composition.

5.1 Laboratory trial discussion

The purpose of this study was to determine if a low cost portable NIR spectrometer can determine fat and protein composition within ICAR tolerances.

The Testlink trial tested the MicroNIR spectrometers in an unmodified configuration. Both spectrometers could not predict fat and protein within ICAR precision. The best fat and protein prediction RMSE was achieved with the MicroNIR1700 spectrometer. The MicroNIR1700 could predict fat with a RMSE of 0.272 and protein with a RMSE of 0.364. The MicroNIR2200 spectrometer predicted the composition of milk with a higher RMSE than the MicroNIR1700 spectrometer. The fat and protein prediction RMSE with the MicroNIR2200 spectrometer were 0.277 and 0.379 respectively. These prediction errors are greater than ICAR precision tolerances for fat and protein of 0.25 g/100 g.

The reason why the MicroNIR spectrometers could not meet ICAR tolerances for fat and protein was caused by the low signal to noise ratio of the milk spectrum. The low signal to noise ratio was caused by the low integration time the spectrometers were operating at. The low integration time was recommended by the manufacturer to reduce the effect of the anomalous pixels for the extended MicroNIR2200 spectrometer. The anomalous pixels response to intensity and integration time is different from the other pixels. The difference in responsivity of the anomalous

pixels causes spikes and wobbles to be present in the normalised reflectance which was hindering our ability to predict milk composition.

The protein prediction scatter plots do not follow the one to one line. The samples below the population mean have been overestimated and the samples higher than the mean have been underestimated. A possible cause for this is the choice of samples used for the calibration data set. The typical protein content of milk has a very narrow Gaussian distribution with relatively few samples at the outer extremes as illustrated in Figure 33. Because the calibration data set was built with a Gaussian distribution it caused the predictions to regress towards the mean which causes the noticeable slope away from the one to one line. This is often called the Dunne effect. Ideally when constructing a calibration model a uniform distribution of composition should be used (Williams et al. 2001).

The age of the milk samples was a source of variability. The milk samples were not fresh when analysed. Milk samples were refrigerated for at least 24 hours before testing. Then the milk samples were reheated twice; the first was for Foss analysis of milk composition, then the sample was reheated for a second time for spectral analysis with the MicroNIR spectrometers. Reheating the milk sample caused flocculation to occur which decreases the homogeneity of the sample

Another possible reason why the milk composition errors were high is due to regional effects. Cows from different areas could be fed different feed types by the farmer as a supplement for milk production. This could cause slight changes in the NIR spectrum due to the different feed types. The different feed types could be changing the colour of the milk in the near infrared region.

The main reason why the spectrometers had a high milk composition prediction error was because of the low signal to noise ratio. The low signal to noise ratio was caused by the low integration time for the MicroNIR spectrometers. The integration time was 500 µs which was too low. This was the manufacturers recommended settings for the MicroNIR2200 to minimise the variation in inter pixel photometric response. The MicroNIR2200 spectrometer uses an uncooled, extended

wavelength range InGaAs array detector. This detector has increased dark current and dark noise, greater thermal drift, and more variation of pixel-to-pixel photometric response.

The variation in pixel-to-pixel photometric response causes anomalies in the normalised reflectance readings. When the reference reflectance is higher than the sample reflectance, the anomalous pixels show up as anomalously low reflectance values as illustrated in Figure 71. The anomalies in the spectrum increase the difficulty in predicting the composition of milk.

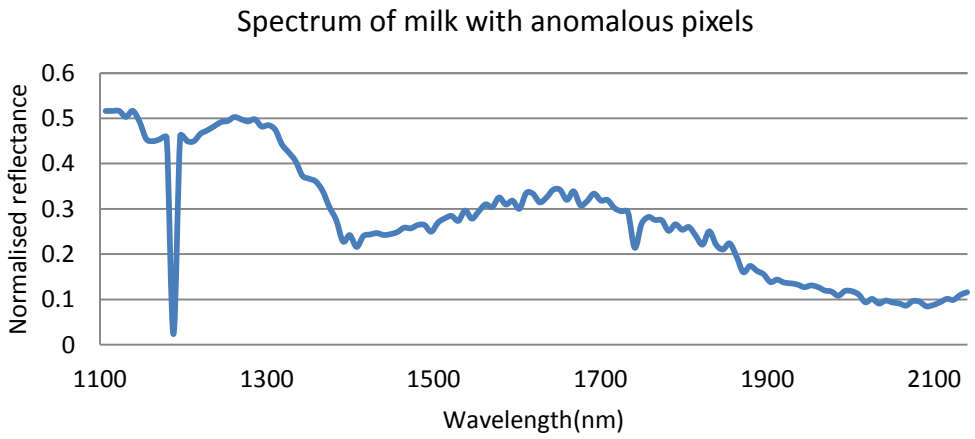

Figure 71: This is the normalised reflectance spectrum of milk at a high integration time. The pixel located at 1188 nm is a pixel with a significantly different photometric response compared with the neighbouring pixels causing the normalised reflectance to be significantly lower than the surrounding pixels.

When the integration time is at the manufacturers recommended setting the dark current is excessively high as illustrated in Figure 72.

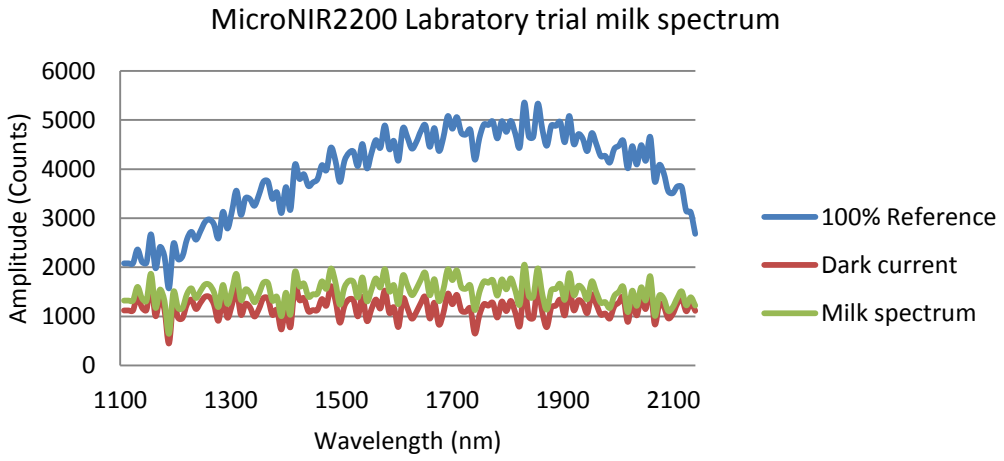

Figure 72: MicroNIR2200 milk spectrum for the laboratory trial

The low integration time causes the dark current to be the major source of noise because the milk spectrum signal is low. This causes the normalised reflectance spectrum to be significantly influenced by the dark current. The low amplitude of the milk spectrum causes wobbles to be present on the normalised reflectance spectrum. Setting the integration time at the manufacturers recommended setting did reduce the effect of the anomalous pixels. However the prediction errors obtained during the laboratory trial were too high and did not meet ICAR standard for fat and protein for an online sensor.

The excessive prediction errors were caused by the samples being old, low signal amplitude, high dark current and inter-pixel wobble. Also the sample distribution was very narrow which was causing the Dunne effect to be present in the predictions. This caused samples below the mean protein composition to be overestimated and samples above the mean to be under estimated. The spectrometers hardware needed to be changed to reduce the prediction error. The following section discusses how the hardware changes improved the quality of the spectrum.

5.2 Optimisation of the hardware

To improve the ability of the MicroNIR spectrometers to predict milk composition the signal to noise ratio of the spectrum had to be improved. This required the amplitude of the milk spectrum to be increased. To increase the amplitude of the milk spectrum the lamp voltage of the spectrometers was increased from 2.5 V to 4.8 V which increased the peak intensity of the spectrum by 300% for the MicroNIR2200 and 380% for the MicroNIR1700 spectrometers. Increasing the lamp voltage increases the received milk spectrum which improves the signal to noise ratio.

Due to the compact nature of the spectrometers, increasing the lamp voltage caused the internal temperature to increase to 50°C. The increase in temperature significantly affects the dark current of the MicroNIR2200 spectrometer. The dark current and the inter-pixel variation increases proportionally to the temperature squared. The MicroNIR1700 dark current increases linearly as the temperature is increased and the inter pixel variation remains constant. This has the effect of changing the responsivity between each pixel which causes changes in the normalised reflectance spectrum. Differences between inter pixel responsivity hindered our ability to predict milk composition. To minimise the effect of inter pixel variance; cooling was required to decrease the operating temperature.

To cool the spectrometers a thermoelectric cooler was mounted onto the case of the spectrometers. A Peltier element was chosen due to its small size and also due to it having no moving parts. The thermoelectric cooler had two main tasks; the first was to remove excess heat from the spectrometers, and the second was to maintain the spectrometers at a constant temperature.

A temperature control circuit was used to maintain the spectrometers at a constant temperature. A thermistor was mounted in the spectrometer to monitor the temperature. The temperature was controlled by proportional and integral control to the thermoelectric cooler.

The dark current of the spectrometers is directly related to the temperature of the array. If the spectrometers are kept at a constant temperature the dark current remain constant during the experiment. This helps to improve the spectrum during the experiment because it stops the dark current from drifting between samples.

Cooling the spectrometers from 42°C to 20°C significantly decreased the spectrum noise for both spectrometers which improves the signal to noise ratio. The average signal to noise ratio across the spectrum improved from 21 to 169 for the MicroNIR2200 spectrometer. There were similar improvements for the MicroNIR1700 with the average spectrum signal to noise ratio improving from 57 to 146.

The transflectance configuration increased the received signal from the milk samples. The transflectance configuration used a 2 mm path length compared with the 5 mm path length used for the reflectance configuration. Reducing the path length of the sample reduced the attenuation of the NIR light caused by the water absorption bands. A piece of spectralon R99% was placed behind the cuvette to reflect any transmitted light.

This investigation on how the spectrometers hardware could be optimised has highlighted the spectrometers did not have a good signal to noise ratio for the laboratory testing. This was caused by the low integration time which caused a low level of signal and the high dark current to be present on the spectrum. The integration time was set low to minimise the effect of the anomalous pixels in the MicroNIR2200. The anomalous pixels had a different photometric responsivity to the neighbouring pixels which distorted the normalised reflectance spectrum. To increase the signal the lamp voltage was increased to 4.8 V which improved the signal to noise ratio. The increased lamp voltage caused the spectrometer to overheat causing an excessive dark current. To cool the spectrometer a Peltier cooler was used to cool and maintain the spectrometers at a constant temperature. This reduced the dark current and also reduced dark current noise. Also the measurement configuration was changed from reflectance to a transflectance configuration. The transflectance configuration has a shorter path length which reduces the attenuation of the NIR light caused

by the water absorption bands. These were the necessary hardware changes that needed to be done to the spectrometers to improve the spectrum.

5.3 Farm trial discussion

The aim of the farm trial was to see if the modifications done to the MicroNIR spectrometers allowed the sensor to predict fat and protein at ICAR precision. The second goal was to test the ability of the calibration model to be transferred between farms. The modified spectrometers with the increased lamp voltage and temperature control circuit were tested on three different farms during November 2012. A total of 590 milk samples were collected from three different farms.

The modifications done to the MicroNIR spectrometers significantly improved the milk composition prediction errors. The MicroNIR1700 was capable of predicting fat and protein with a RMSE of 0.061 and 0.151 for the pooled data set. The MicroNIR2200 predicted protein with a RMSE of 0.138 which is better than the MicroNIR1700. The MicroNIR2200 predicted the composition of fat to a similar prediction RMSE of 0.071. These prediction errors are within the allowed ICAR precision limits for fat and protein.

The calibration models were able to be transferred across the three different farms with only a slight increase in fat prediction errors. The MicroNIR1700 predicted milk fat with a RMSE between 0.064 and 0.070. These prediction errors are slightly higher than the prediction errors obtained for the pooled data set where a RMSE of 0.061 was obtained.

When independent farm calibrations were used the milk fat prediction errors increased for the MicroNIR2200. A milk fat prediction RMSE of 0.077 to 0.087 was achieved for the independent calibration models. These prediction errors are slightly higher than the pooled data set where a RMSE of 0.071 was obtained. The summary statistics for the fat prediction errors is illustrated in Table 4

The prediction errors are higher for the independent farm calibration models because the calibration does not include subtle inter farm differences for the test farm.

The overall performance of the calibration models for fat was excellent. The reason for the high performance of the calibration model for fat was that the milk spectra had much information of fat content from scattering

light by fat globules and absorption by C-H strings and C-C strings of triacylglycerols. The fat content dominates the NIR spectra for both spectrometers. Variations in milk fat content cause the normalised reflectance spectra to be displaced. Samples with high milk fat content have high normalised reflectance values and samples with low milk fat have low normalised reflectance values. Fat globules are responsible for back scattering a significant part of the light in the direction of the light source, resulting in a higher reflectance and lower transmittance for fatty milk samples.

Milk fat is easy to predict because a wide sample distribution can be easily obtained. The concentration in fat varies during the milking process with the fore milk containing a low concentration of fat and the striping's containing the highest milk fat (Holmes & Wilson 1987). It is possible to adjust the milk fat composition by sampling at different points during milking to obtain samples at the extreme ends of the sample population. This helps to achieve a uniform distribution of composition across the anticipated range which improves the performance of the calibration model.

The prediction results for protein were worse than those obtained for fat. Protein is difficult to predict because the changes in fat concentration causes variations in the amount of light scattered from sample to sample. The main errors in milk analysis are associated with changes in fat concentration from one sample to another and variations in the amount of scattering of infrared energy by fat globules (O'Sullivan et al. 1999). Spectral pre-processing such as multiplicative scatter correction can be used to correct for additive scatter effects (Rinnan et al. 2009).

Protein is difficult to predict because the concentration of protein does not vary widely between samples. The protein content for the farm trial varied from 2.59% to 4.87% which is a narrow range compared with the range of fat. The fat content of the milk samples varied from 0.61% to 13.81% which improves the calibration model. Also the protein sample population is characterised by a Gaussian distribution with a narrow width. This means there are relatively fewer samples at the extremes.

One of the main difficulties when studying milk is the high absorption of water in the NIR spectrum. The NIR bands for determining protein are typically weak in nature and are around the water absorption bands. The assignment of the peptide bands in the NIR spectrum are broad bands between 1100 and 2500 nm (Dufour 2009). The broad nature of the protein bands are clearly illustrated in Figure 66 and Figure 68 where there are numerous wavelengths contributing to predict the protein composition.

Both spectrometers predicted the protein composition within the acceptable ICAR tolerances for an online sensor for the inter farm transferability test. The MicroNIR1700 spectrometer predicted the protein composition on independent farms with a RMSE of 0.154 to 0.178. These prediction errors are slightly higher than the results obtained with the pooled data set. The best independent protein prediction errors were obtained on the third farm with a RMSE of 0.154. A possible reason why the prediction errors were lower for the third farm is the calibration data set had a greater population to calibrate against. This created a calibration data set that was more representative of the independent test farm compared to the other tests. The other two independent farm calibration tests had the same prediction errors of 0.178. An additional reason why better protein composition results were obtained for the third farm is the milk fat variance is significantly lower than the other test farms. The milk fat variance for the third farm was 4.71% compared with 7.39% and 6.09% for the first and second farms respectively. The high variance in milk fat composition ensures the calibration data set covers the anticipated variety of samples to be tested.

The MicroNIR2200 spectrometer predicted the protein composition with a RMSE between 0.153 and 0.215. The lowest prediction RMSE obtained was 0.153 for the third farm validation test. This is a similar prediction error acquired with the MicroNIR1700 which had a RMSE of 0.154 for the third farm. A significantly higher prediction RMSE of 0.218 was obtained for the second farm validation set. A possible reason why the prediction errors were higher is due to the smaller calibration data set which may exclude some sample variability in the calibration. The reason why the prediction

errors are higher for the independent validation data set is due to the exclusion of regional variations in the calibration data set. The performance of the calibration model is extremely dependent on the training set chosen. Choosing every third sample in order of composition ensures a representative data set is used for calibration.

5.4 Comparison to other scientific studies

The analysis of raw milk with NIR spectroscopy is not new and has been done by other researchers in the past. The ability to predict the composition of milk with NIR has been proven possible in the past.

A trial investigating online milk spectroscopy was done by Spitzer (2005). Transmission spectra through a 0.50 mm polysulphone flow cell. Spectra were measured in the range from 380 to 1100 nm at 0.20 nm intervals and 900 to 1700 nm at 1 nm intervals. The flow cell was installed in the milk line of a Merlin milking system. A total of 368 samples were collected from individual quarters. The flow cell was rotated across all four quarters. For fat a RMSE of 0.70% was obtained using partial least squares regression in the NIR region (900 – 1700 nm) with a correlation coefficient of 0.93. These prediction errors and correlation coefficients are of poorer quality than the results obtained with the MicroNIR spectrometers for fat. High protein prediction errors were obtained for the NIR region. Partial least square models with fifteen latent variables were used to predict protein composition in the NIR region. A RMSE of prediction of 0.65% and correlation coefficient of 0.50 for protein was achieved. These prediction errors are significantly higher than the prediction errors obtained during the farm trial.

A possible reason why high milk composition errors were obtained during this trial may be due to the sampling method used. To collect milk samples a solenoid valve was placed directly downstream from the flow cells. This sampling method may be prone to contamination due to milk constantly flowing through the milk line. This may cause error in the calibration due to the sample not having the same milk composition compared to the sample used for spectral analysis. Ideally when testing the capability of a spectrometers ability to predict milk composition spectral analysis should be done on a subsample to avoid cross contamination.

Aernouts et al. (2011) investigated transmission and transflection measurement configurations at two wavelength ranges. The trial done by Aernouts consisted of 300 cows from 30 different farms. A Zeiss Corona VISNIR1.7 diode array spectrometer was used to measure the visible to

NIR spectrum. The visible range was covered using a silicon array from 306.5 nm to 1135.5 nm with a spectral resolution of 3.2 nm. The NIR spectrum was covered with an InGaAs array with 6 nm resolution in the 944.5 to 1710.9 nm range. Aernouts achieved a milk fat prediction RMSE of 0.047 for the reflectance configuration. Similar results were obtained in the farm trial with a fat prediction RMSE of 0.061 using the 880 – 1661 nm spectrum. In the transmission configuration the fat prediction errors were slightly better with a RMSE of 0.046. The protein composition was able to be predicted with a RMSE of 0.099 using the reflectance configuration. The transmittance configuration had a slightly higher RMSE of 0.133 for protein. The MicroNIR1700 transflectance configuration for the farm trial obtained a slightly higher RMSE of 0.133 for protein. The short wavelength range (400 – 1100 nm) contained less information for the fat prediction. When the 400 – 1100 nm spectra was used to predict fat significantly worse prediction errors resulted. The Zeiss Corona VISNIR1.7 is a big and heavy spectrometer with dimensions of 280 x 230 x 110 mm and weighs 6 kg. The spectrometer used has a slightly higher spectral resolution in the NIR range of 10 nm compared with 12.5 nm for the MicroNIR which accounts for the milk composition prediction errors.

A study done by Melfsen et al. (2012a) achieved low milk composition errors. A total of 1610 milk samples were collected from 38 Holstein cows (8 – 49 samples per cow). Diffuse reflectance spectra were collected in the 851 - 1649 nm range. Milk was stirred with nominal rotation speed of 102 rpm while the spectra were acquired. A 30 mm static measurement cell was used to take spectra. A total of 1073 samples were used for the calibration data set and 537 for the validation set. Good fat and protein prediction RMSE were obtained with 0.03 for fat and 0.07 for protein. High correlation coefficients were achieved in this trial with a $R^2 = 0.998$ for fat and $R^2 = 0.94$ for protein. The spectrometer used for this trial was the PSS-1720, Polytec GmbH which had 256 pixels and a spectral resolution of 7.5 nm. The main reason why better prediction errors were obtained from this trial is the spectrometer has a higher resolution compared to the MicroNIR spectrometers. Although higher correlations are obtained with this spectrometer it is unnecessary to have this level of accuracy for an

online situation because many readings can be averaged during the milking. Also the measurement method used in this trial is not practical for an online situation. The results obtained from this trial meet the ICAR specifications but the spectrometer is over engineered for an online sensor. This results in a more expensive product which is undesirable for a competitive market.

A second trial done by Melfsen et al. (2012b) investigated building calibration models and applying the models to farms that were not part of the calibration set. Three different farms in Germany were visited twelve times. A total of 3119 samples were collected. A PSS-1720 spectrometer was used to analyse the milk spectra. The spectrometer was installed in the long milk tube. NIR diffused reflection spectra were acquired in the wavelength region of 851 – 1649 nm. Spectra were acquired during the milking process every 500 ms. For each 2 kg of milking, spectra were averaged and subsamples of raw milk in the bypass system were collected for reference analysis. The milk fat composition prediction errors achieved in this trial had a RMSE of 0.16. An excellent protein prediction RMSE of 0.06 was achieved for the pooled data set. When an external calibration set and validation set was used the prediction errors increased. The milk fat composition was predicted with a RMSE of 0.17 to 0.2. A protein prediction RMSE of 0.16 to 0.18 was achieved using the external data set.

When a dependent calibration set was built with two thirds of the data from all three farms the prediction errors improved. For fat the RMSE was 0.17 to 0.19 and for protein the RMSE was 0.10 to 0.12. This trial found the prediction improved when external spectra from other farms was included in the calibration set. This is due to the superior amount of sample variation in terms of feeding diets and different cow breeds in external calibration sets. Slightly higher prediction errors were achieved when independent farms were used for calibration and validation. The increase in prediction errors observed by Melfsen et al. (2012b) agrees with the results gathered from the farm trial. When an independent calibration data set was used there was a slight increase in prediction errors. The fat prediction errors obtained in this trial are higher than the results obtained

for the farm trial where a RMSE of 0.061 to 0.070 was achieved for the wavelength range of 880 – 1661 nm. This trial had a greater sample size which may have increased the prediction errors.

Kawasaki et al. (2006) analysed 216 milk samples on-line using a diffused transmission configuration. The system consisted of an NIR spectrometer, a milk flow meter, and a milk sampler. The spectrometer used was a diffraction grating spectrometer with a linear array CCD and a Peltier cooling system. The wavelength resolution was 5 nm. The system was installed in a Lely Astronaut robotic milking system. VIS-NIR spectrum was analysed from 600 nm to 1050 nm. For validation set of 72 milk samples a standard error of prediction of 0.25 and 0.15 was achieved for fat and protein respectively. A higher fat prediction error was obtained for this trial which may be caused by the small sample size used to build the calibration set. A similar protein prediction error was obtained from the farm trial of 0.15. This trial used a small validation set which can influence the robustness of the model; in addition the samples are based on a small number of cows. This method leads to favourable results due to the small data set.

Sasic and Ozaki (2001) analysed 100 raw milk samples and predicted fat and protein using partial least squares. Transmission spectra in the 800 – 1100 nm region were recorded at 2 nm intervals by a NIR systems 6500 spectrometer. The milk samples were homogenised prior to the NIR measurement. When 6 partial least square factors were used a fat prediction RMSE of 0.083 results. The RMSE of proteins in the independent data sets varied from 0.121 to 0.180. Once again the spectrometer used in this trial has a higher resolution than the MicroNIR spectrometers. Also homogenising milk samples is not practical in an online situation because it requires time and can be quite bulky. For these reasons this measurement configuration is undesirable. The potential of near infrared spectroscopy in the 1100 – 2400 nm range was studied by Tsenkova et al. (2000). A total of 260 individual milk samples were collected from three Holstein cows over a six month period. Transmittance spectra of 1 mm thick milk samples were obtained using a FOSS NIRSystems at 2 nm intervals. A standard error of cross-validation of

0.125 and 0.138 was obtained for fat and protein when all the cows were included in the sample set. For fat content determination, the correlation coefficients obtained from calibration and cross-validation procedures were greater than 0.997. A Relatively low correlation coefficient for protein of 0.760 was obtained. This was due to the small total variation of protein content in milk. The farm trial results obtained better milk fat RMSE results using the MicroNIR2200 spectrometer of 0.071 for a similar spectral region of 1137 – 2150 nm. The reason why smaller prediction errors for fat were obtained was the farm trial had a larger data set with more variation in sample composition and cow variety, this minimises the error of prediction. Slightly better protein prediction errors were obtained in this trial compared with the farm trial. A possible reason why a better fat prediction was obtained was due to the small variety of cows used in the trial. Also the validation test was not tested on independent data because the cows were used for the calibration and validation data sets. The spectral features at 1132 nm, 1460 nm to 1520 nm, 1980 nm to 2070 nm, and 2170 nm to 2180 nm, respectively were connected with the absorption of N-H groups in proteins. These areas of significant wavelength influences are similar to the wavelength influences used for predicting protein with the MicroNIR2200 spectrometer as illustrated in Figure 68.

Saranwong and Kawano (2008) compared a compact NIR systems called MilkSpec-2. A total of 103 non-homogenised raw milk samples were collected from 26 individual milking cows for two days. 1.2 mm NIR transmittance spectra from 400 nm to 1100 nm at 2 nm intervals were used to predict the composition of milk. Excellent fat prediction results were obtained in this trial with a standard error of prediction of 0.03. A good standard error of 0.07 for protein was achieved. The calibration and validation samples are not independent because the samples used were from the same cows. This is a significant advantage because the validation model was tested on cows used for calibration. This does not test if the calibration model can be transferred across breed types. The spectrometer used in this trial has higher resolution than the MicroNIR spectrometers which helps to achieve lower milk composition errors. Transmission spectra were used to measure the milk composition. This is a different measurement configuration used on the farm trial.

The main reason why the transmission configuration was neglected for measuring milk composition was due to the substantial attenuation of NIR light in milk caused by water absorption. To receive enough signal to the spectrometers a 1 mm flow cell must be used to receive enough signal. This path length is not practical for an inline sensor where debris may be jammed in the cuvette. A summary of the scientific studies and the results achieved during the farm trial is illustrated in Table 7.

Table 7: Summary of other scientific studies

Trial	Sensor	Sensing configuration	Resolution	Spectrum	Fat RMSEP	Protein RMSEP
Spitzer (2005)	SNLC681	Transmittance	0.20 nm	380 – 1100 nm	0.91	0.47
	SNIR 349		1.0 nm	900 – 1700 nm	0.70	0.65
Aernouts (2011)	Zeiss Corona 45 VISNIR 1.7 with InGaAs	Reflection	3 nm 6 nm	307 - 1136 nm 944 - 1711 nm	0.047	0.099
		Transmittance			0.046	0.133
Melfsen et al. (2012a)	PSS-1720, Polytec GmbH, Wald-bronn, Germany	Reflection	6.4nm	851 - 1649 nm	0.03	0.07
Melfsen et al. (2012b)	PSS-1720, Polytec GmbH, Wald-bronn, Germany	Reflection	6.4nm	851 - 1649 nm	0.19	0.11
Kawasaki et al. (2008)	-	Transmittance	5 nm	600–1050 nm	0.25	0.15
Sasic & Ozaki (2001)	NIRSystem 6500 spectrophotometer	Transmittance	2 nm	800 - 1100 nm	0.083	0.121
Tsenkova et al. (2000)	NIRSystem 6500 spectrophotometer	Transmittance	2 nm	700 - 1100 nm	0.125	0.138
Saranwong & Kawano (2008)	MilkSpec-3	Transmittance	2 nm	700-1000 nm	0.03	0.07
Farm trial (2012)	MicroNIR1700	Transflectance	6 nm	950 - 1650 nm	0.061	0.151
Farm trial (2012)	MicroNIR2200	Transflectance	8 nm	1130- 2150 nm	0.071	0.138

Overall the results obtained with the MicroNIR spectrometers are comparable with the other scientific studies. Lower milk composition prediction errors can be obtained by using spectrometers with higher resolution. The spectrometers used in the previous trials are suited for laboratory analysis of milk samples and are not practical for an online situation on a milking robot. The MicroNIR spectrometers can meet ICAR accuracy for an online milk sensor using a low-priced spectrometer. This is a significant advantage and provides a competitive advantage over the other spectrometers.

5.5 Improvements

During the farm trial there was noise present on the reference spectrum. During the three farm experiments there was a gradual decline of the MicroNIR1700 reference spectrum during the initial stages of sampling. Also there were random spikes or sudden shifts in the amplitude of the reference spectrum for both spectrometers as illustrated in Figure 73 and Figure 74.

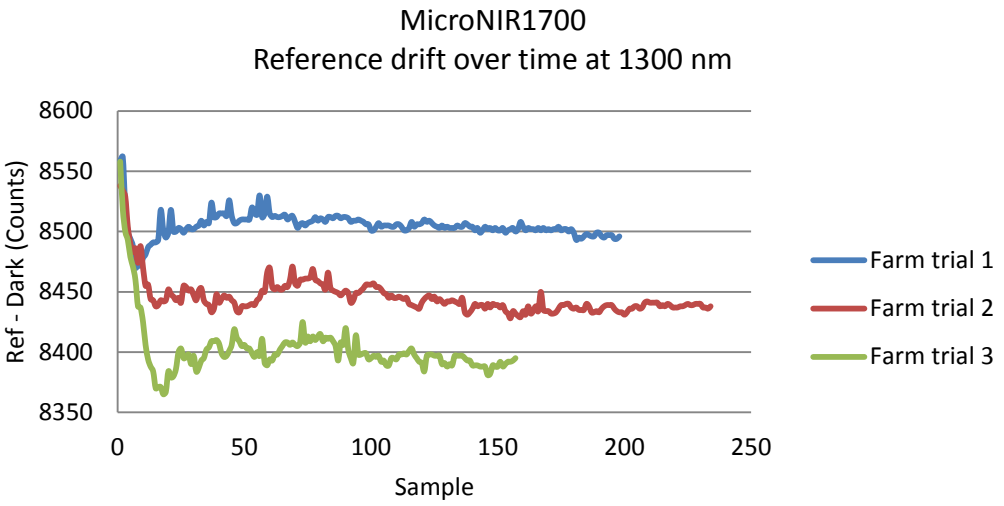

Figure 73: Reference variation during the three different farm trials for the MicroNIR1700 at 1300 nm

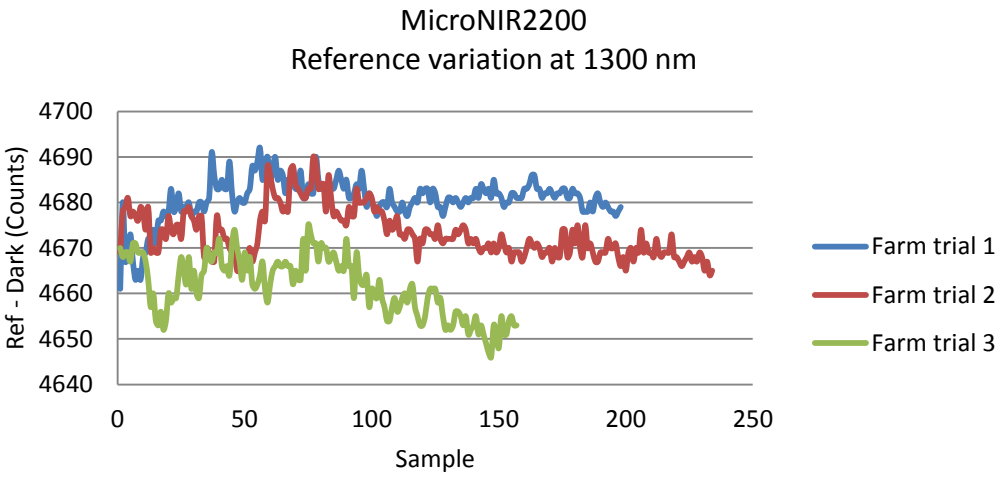

Figure 74: Reference variation during the three different farm trials for the MicroNIR2200 at 1300 nm

It is unclear what causes the gradual decline in amplitude for the MicroNIR1700 spectrometer. A possible reason why the amplitude of the reference spectrum decreases may be caused by the lamps warming up

during start up. It is unusual to not see this effect on both spectrometers at the same wavelength.

The sudden shifts in amplitude could be caused by a change in lamp voltage during the experiment. This would cause the amplitude of the spectrum to shift during the experiment. To eliminate voltage fluctuations it is recommended to use a voltage regulator with low noise and high accuracy to power the lamps of the spectrometer.

The shift in reference amplitude may be caused by the reference tiles not being placed at the same position. Displacement variations would cause changes in the amplitude of the reference spectrum. Also the servomechanism may not be rotating the spectralon reference tile to the same rotational position. To reduce the noise on the reference spectrum it is recommended to remove the moving reference on the servomechanism. This will eliminate mechanical noise on the spectrum. The reference spectrum should be obtained using the spectrum of water in the flow cell or use an empty flow cell with a piece of white spectralon R99 reference tile behind the cuvette.

Possible problems with using the spectrum of water for a reference spectrum is water will attenuate the NIR spectrum causing the amplitude to decrease at the water absorption bands. The low amplitude of the water reference spectrum will increase the noise present on the normalised reflectance spectrum. Another problem with using the water spectrum is the absorption of water changes with temperature. An increase in water temperature causes a hypochromic shift of the maximum absorption band and causes the absorption band to narrow. It was found that changing the sample temperature generally increased the number of partial least square factors by one (Blanco & Valdes 2004). This leads to an unstable reference spectrum which is dependent on temperature.

A possible problem of using an empty flow cell for a reference spectrum is there may be a residual layer of moisture on the cuvette which will attenuate the NIR light. The variations in residual moisture on the cuvette will cause variations in the reference spectrum. To remove the moisture it

may be possible to dry the cuvette with compressed air between milk samples but this adds complexity to the design.

Another possible problem in the future is maintaining a clean cuvette. If the cuvette is not cleaned properly between test a residual fat layer may build up inside the cuvette. This may cause a gradual drift in the spectra and cause a bias to emerge over time. The bias could be reduced if the sensor is frequently recalibrated but this is not practical for a commercial farmer. Therefore it is important proper cleaning of the cuvette is done periodically to prevent drift.

The calibration data set used in this trial was not uniform across the milk composition range. The sample population has a typical Gaussian or “normal” distribution about the mean. Ideally, calibration sets should be assembled with uniform distribution of composition across the anticipated range. A Gaussian shaped calibration data set was used because of the limited number of samples acquired during the trial. Calibrations developed with the sample set having a Gaussian distribution of composition with respect to a particular constituent may cause the future analysis to regress towards the mean. This is also known as the Dunne effect and it causes the predictions at higher concentrations to appear lower, while samples at the low concentrations to appear higher than the actual reference result (Williams et al. 2001).

It may be possible to obtain better milk composition predictions if the outer edges of the spectrum are excluded. It has been proven the outer edges of the measured spectrum have a lower signal to noise ratio. If the spectrum with the low signal to noise ratio is removed an improvement in milk composition prediction errors may result.

The signal to noise ratio of the spectrum can be improved by increasing the number of scans taken for each measurement. Increasing the number of scans used will reduce the spectrum noise. The standard deviation of the averaged samples is reduced by the number of samples taken for each measurement:

$$\sigma_{ave} = \sigma\sqrt{\frac{1}{N}} \quad \textbf{(5-1)}$$

Where σ is the standard deviation and
N is the number of samples.

Unfortunately there must be a trade-off between the degree of certainty in the spectral measurement and the time taken to acquire the scans. Due to time restrictions during the experiments the number of scans taken was limited to 1000.

The calibration models built during this trial are only for one spectrometer. This trial has not tested if the calibration models can be transferred between spectrometer. Different MicroNIR spectrometers will have slightly different wavelength positions for each individual pixel. Calibration models would need to include instrumental variations for a commercial product. Also the number of farms used in the trial is limited. Calibration models should ideally be built across several farms to achieve robustness in commercial farm applications.

Further milk composition improvements could be made with a higher integration time. Increasing the integration time on the spectrometers reduces the spectrum noise. Using a low integration time does not fully utilise the 16 bits on the analogue to digital converter. Currently the peak milk spectrum amplitude is approximately 6000 counts which only utilises 13 bits.

$$\frac{\log_{10} 6000}{\log_{10} 2} = 12.6 \rightarrow 13\text{Bits} \quad \textbf{(5-2)}$$

It is recommended to increase the integration time to increase the signal received. Increasing the signal increases the number of bits used to read the spectrum which increases the resolution.

6 Conclusions

The main objective of this study was to investigate the feasibility of using low cost MicroNIR spectrometers to measure lipid and protein content of milk. 11Ants model builder was used to build calibration models with the spectra of the raw milk samples.

The laboratory testing of the JDSU spectrometers used a 5 mm path length flow cell with the unmodified MicroNIR spectrometers in a reflectance configuration. A total of 1675 reflectance spectra were collected with the MicroNIR1700 and MicroNIR2200 spectrometers. The MicroNIR1700 measured the 880 – 1661 nm spectrum and the MicroNIR2200 measured the 1130 – 2160 nm spectrum. Inadequate lipid and protein composition prediction errors were obtained from the laboratory trial. This implies the original MicroNIR spectrometers are inadequate for predicting fat and protein to ICAR standard.

Following the laboratory trial the hardware needed to be optimised to meet ICAR accuracy tolerances. To optimise the spectrometers the signal to noise ratio needed to be increased. To increase the received signal the lamp voltage was increased from 2.5 V to 4.8 V. This increased the spectrum intensity by at least 200% over all wavelengths. Increasing the lamp voltage caused the spectrometers to overheat. To reduce the operating temperature a Peltier element was used to cool the spectrometers to 20°C. Cooling the spectrometers significantly improved the dark current spectrum. Maintaining the spectrometers at a constant temperature reduced dark current fluctuations and prevents the dark current from drifting. Cooling the spectrometers also minimised the inter pixel non-linear photometric response which helps to produce a smooth spectrum. The measurement configuration was changed from reflectance to transflectance configuration by changing the path length of the cuvette from 5 mm to 2 mm. The transflectance configuration doubled the received signal because the shorter path length reduced the attenuation of the NIR light.

To test the capability of the optimised spectrometers to predict the composition of milk a farm trial was done on farm. 590 milk samples were collected from three different farms. The spectrometers were cooled to 20°C and the lamp voltage was increased, providing a 300% increase in received signal. Transflectance spectra were obtained using a 2 mm flow cell with R99 spectralon reflecting the transmitted light back to the spectrometers.

Calibration models were built across all three farms. Pooled milk composition prediction errors are summarised below.

- A total of 590 composite milk samples were collected
- The MicroNIR1700 obtained the following RMSE and correlation coefficients for the pooled data set during the farm trial
    - 0.061%, 0.999 for milk fat
    - 0.151%, 0.904 for protein
- The MicroNIR2200 obtained the following RMSE and correlation coefficients for the pooled data set during the farm trial
    - 0.071%, 0.999 for milk fat
    - 0.138%, 0.918 for protein

To test if calibration models could be transferred across different farms and breed types independent calibration models were built. Milk spectra from two farms were used to build a calibration model. The model was independently validated on the third farm

Independent validation milk composition prediction RMSE and correlation coefficients are summarised below

- The MicroNIR1700 achieved
    - 0.064 – 0.070%, 0.999 for milk fat
    - 0.154 – 0.178%, 0.825 – 0.892 for protein
- The MicroNIR2200 achieved
    - 0.077 – 0.087%, 0.999 for milk fat
    - 0.153 – 0.218%, 0.858 – 0.726 for protein

The prediction errors for the independent validation test were slightly higher than the pooled data. This was due to the calibration model not including regional variation from the independent farm.

The results achieved show the modified spectrometers can meet the ICAR accuracy restrictions for an online milk sensor. Also the calibration models can be used to predict fat and protein on independent farms with minimal bias. This proves the modified MicroNIR spectrometers can predict the composition of milk within ICAR accuracy limits for an online milk sensor. Further work needs to be done on reducing the noise on the reference spectrum which should improve our ability to predict milk composition.

It appears that both spectrometers are evenly matched in the current configuration. Both spectrometers have excellent fat prediction errors and can predict protein within ICAR limits. It appears the MicroNIR2200 spectrometer has a slightly better protein prediction RMSE of 0.138 compared with 0.151 for the MicroNIR1700 when using the pooled data. This indicates that the MicroNIR2200 is slightly better at predicting protein from a pooled data set but there are stability issues with this spectrometer. The MicroNIR2200 spectrometer has inter-pixel wobble problems which get significantly worse as the internal temperature approaches 30°C. The dark current noise exponentially rises as temperature increases for the MicroNIR2200 spectrometer. This is a significant disadvantage of the MicroNIR2200 spectrometer.

The MicroNIR1700 spectrometer does not have the inter-pixel variation problem which the extended spectrometer has. The dark current linearly rises with temperature and the inter-pixel variation remains constant with increasing temperature. This makes the MicroNIR1700 spectrometer more robust to temperature fluctuations. The MicroNIR1700 spectrometer is also US$2,000 cheaper than the extended MicroNIR2200 spectrometer and has similar protein prediction errors.

This makes the MicroNIR1700 spectrometer an attractive option for predicting the composition of milk with.

7 References

Aernouts, B., Polshin, E., Lammertyn, J. & Saeys, W. 2011, 'Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: Reflectance or transmittance?', *Journal of dairy science*, vol. 94, no. 11, pp. 5315-29.

Avison, J. 1989, *The World of Physics*, Nelson.

Biggs, D.A. 1967, 'Milk Analysis with the Infrared Milk Analyzer', *Journal of dairy science*, vol. 50, no. 5, pp. 799-803.

Blanco, M. & Valdes, D. 2004, 'Influence of temperature on the predictive ability of near infrared spectroscopy models', *Journal of near Infrared Spectroscopy*, vol. 12, no. 2, pp. 121-6.

Blanco Romía, M. & Alcalà Bernàrdez, M. 2009, 'Chapter 3 - Multivariate Calibration for Quantitative Analysis', in *Infrared Spectroscopy for Food Quality Analysis and Control*, Academic Press, San Diego, pp. 51-82.

Blowey, R. & Edmondson, P. 2010, *Mastitis Control in Dairy Herds*, CABI.

Cattaneo, T.M.P., Cabassi, G., Profaizer, M. & Giangiacomo, R. 2009, 'Contribution of light scattering to near infrared absorption in milk', *Journal of near Infrared Spectroscopy*, vol. 17, no. 6, pp. 337-43.

Devore, J.L. & Peck, R. 2001, *Statistics: The Exploration and Analysis of Data*, Brooks/Cole.

Dufour, É. 2009, 'Chapter 1 - Principles of Infrared Spectroscopy', in *Infrared Spectroscopy for Food Quality Analysis and Control*, Academic Press, San Diego, pp. 1-27.

Foss 2006, *MilkoScan™ FT2 advance milk analysis for profit in dairy production*, viewed 26 June 2012, <http://www.foss.dk/-/media/Files/Documents/IndustrySolution/BrochuresandDatasheet/MilkoScanFT2/MilkoScan%20FT2%20datasheet_GB.ashx>.

Fox, P.F. & McSweeney, P.L.H. 1998, *Dairy Chemistry and Biochemistry*, Springer.

Hamamatsu 2004, *Characteristics and use of infrared detectors*, viewed 22 July 2012, <http://sales.hamamatsu.com/assets/applications/SSD/Characteristics_and_use_of_infrared_detectors.pdf>.

Hermes, C.J.L. & Barbosa Jr, J.R. 2012, 'Thermodynamic comparison of Peltier, Stirling, and vapor compression portable coolers', *Applied Energy*, vol. 91, no. 1, pp. 51-8.

Holmes, C.W. & Wilson, G.F. 1987, *Milk Production from Pasture*, Elsevier Science & Technology Books.

ICAR 2012, *International Agreement of Recording Practices*, viewed 15 November 2012, <http://www.icar.org/Documents/Rules%20and%20regulations/Guidelines/Guidelines_2012.pdf>.

Kawasaki, M., Kawamura, S., Morita, S., Komiya, M. & Natsuga, M. 2006, 'Near-infrared spectroscopic sensing system for online milk quality assessment in a milking robot for dairy precision farming', in *Preprints. 3rd IFAC International Workshop on Bio-Robotics, Information Technology and Intelligent Control for Bioproduction Systems, 9-10 Sept. 2006*, Japan, Japan, pp. 60-4.

Kenyon, I.R. 2008, *The Light Fantastic: A Modern Introduction to Classical and Quantum Optics*, Oxford University Press.

Klandar, A.H., Lagaude, A. & Chevalier-Lucia, D. 2007, 'Assessment of the rennet coagulation of skim milk: A comparison of methods', *International Dairy Journal*, vol. 17, no. 10, pp. 1151-60.

Laporte, M.F. & Paquin, P. 1999, 'Near-infrared analysis of fat, protein, and casein in cow's milk', *Journal of Agricultural and Food Chemistry*, vol. 47, no. 7, pp. 2600-5.

Lin, M., A Rasco, B., G Cavinato, A. & Al-Holy, M. 2009, 'Chapter 6 - Infrared (IR) Spectroscopy—Near-Infrared Spectroscopy and Mid-Infrared Spectroscopy', in *Infrared Spectroscopy for Food Quality Analysis and Control*, Academic Press, San Diego, pp. 119-43.

Lipson, A., Lipson, S.G. & Lipson, H. 2010, *Optical Physics*, Cambridge University Press.

Lorenzo, J.R. 2012, *Principles of Diffuse Light Propagation: Light Propagation in Tissues With Applications in Biology and Medicine*, World Scientific Publishing Company Incorporated.

Madouasse, A., Huxley, J.N., Browne, W.J., Bradley, A.J. & Green, M.J. 2010, 'Somatic cell count dynamics in a large sample of dairy herds in England and Wales', *Preventive Veterinary Medicine*, vol. 96, no. 1–2, pp. 56-64.

Malchow, D.S., Brubaker, R.M. & Hansen, M.P. 2008, 'Development of linear array ROIC for InGaAs detector arrays with wavelength response to 2.5 microns for NIR spectroscopy and machine vision', *Proceedings of SPIE*, vol. 6940, p. 69402v.

Malinowski, E.R. 2002, *Factor analysis in chemistry*, Wiley.

Mansfield, M. & O'Sullivan, C. 2010, *Understanding Physics*, John Wiley & Sons.

Marcelli, A., Cricenti, A., Kwiatek, W.M. & Petibois, C. 2012, 'Biological applications of synchrotron radiation infrared spectromicroscopy', *Biotechnology Advances*, vol. 30, no. 6, pp. 1390-404.

Melfsen, A., Hartung, E. & Haeussermann, A. 2012a, 'Accuracy of milk composition analysis with near infrared spectroscopy in diffuse reflection mode', *Biosystems Engineering*, vol. 112, no. 3, pp. 210-7.

—— 2012b, 'Robustness of near-infrared calibration models for the prediction of milk constituents during the milking process', *Journal of Dairy Research*, vol. FirstView, pp. 1-10.

O'Sullivan, A., O'Connor, B., Kelly, A. & McGrath, M.J. 1999, 'The use of chemical and infrared methods for analysis of milk and dairy products', *International Journal of Dairy Technology*, vol. 52, no. 4, pp. 139-48.

Querry, M.R. & Hale, G.M. 1972, 'Optical constants of water in the 200-nm to 0.2-mm wavelength region', in *1972 Annual Meeting of the Optical Society of America. Abstracts only, 17-20 Oct. 1972*, Washington, DC, USA, p. 54.

Rasmussen, C.E. & Williams, C.K.I. 2006, *Gaussian Processes for Machine Learning*, Mit Press.

Rawlings, J.O., Pantula, Sastry G. Dickey, David A. 1998, *Applied regression analysis: A research tool.*, 2nd edn, Springer texts in statistics, Springer, New York.

Rinnan, Å., Nørgaard, L., Berg, F.v.d., Thygesen, J., Bro, R. & Engelsen, S.B. 2009, 'Chapter 2 - Data Pre-processing', in *Infrared Spectroscopy for Food Quality Analysis and Control*, Academic Press, San Diego, pp. 29-50.

Roos, W.S. 2003, *Near Infrared Spectroscopy Introduction into the method*, Society for Medical Plant and Natural Product Research, viewed 20 June 2012, <http://www.ga-online.org/files/kiel2003/WS_Roos.pdf>.

Saranwong, S. & Kawano, S. 2008, 'System design for non-destructive near infrared analyses of chemical components and total aerobic bacteria count of raw milk', *Journal of near Infrared Spectroscopy*, vol. 16, no. 4, pp. 389-98.

Sasic, S. & Ozaki, Y. 2001, 'Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment', *Analytical Chemistry*, vol. 73, no. 1, pp. 64-71.

Spitzer, K. 2005, 'Online milk spectroscopy', MSc(Tech) thesis, University of Waikato.

Straughan, B.P. & Walker, S.D. 1976, *Spectroscopy*, Chapman and Hall.

Tsenkova, R., Atanassova, S., Itoh, K., Ozaki, Y. & Toyoda, K. 2000, 'Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a spectral region from 1,100 to 2,400 nanometers', *Journal of Animal Science*, vol. 78, no. 3, pp. 515-22.

Upreti, P., McKay, L.L. & Metzger, L.E. 2006, 'Influence of Calcium and Phosphorus, Lactose, and Salt-to-Moisture Ratio on Cheddar Cheese Quality: Changes in Residual Sugars and Water-Soluble Organic Acids During Ripening', *Journal of dairy science*, vol. 89, no. 2, pp. 429-43.

Varnam, A.H. & Sutherland, J.P. 2001, *Milk and Milk Products: Technology, Chemistry and Microbiology*, Springer.

Viguier, C., Arora, S., Gilmartin, N., Welbeck, K. & O'Kennedy, R. 2009, 'Mastitis detection: current trends and future perspectives', *Trends in Biotechnology*, vol. 27, no. 8, pp. 486-93.

Williams, P., Norris, K.H. & Chemists, A.A.o.C. 2001, *Near-Infrared Technology: In the Agricultural and Food Industries*, American Association of Cereal Chemists.

Woo, Y.A., Terazawa, Y., Chen, J.Y., Iyo, C., Terada, F. & Kawano, S. 2002, 'Development of a new measurement unit (MilkSpec-1) for rapid determination of fat, lactose, and

protein in raw milk using near-infrared transmittance spectroscopy', *Applied Spectroscopy*, vol. 56, no. 5, pp. 599-604.

Workman, J. & Weyer, L. 2008, *Practical Guide to Interpretive Near-Infrared Spectroscopy*, CRC Press.

Yakovenko, A.A., Yashin, V.A., Kovalev, A.E. & Fesenko, E.E. 2002, 'Structure of the vibrational absorption spectra of water in the visible region', *Biophysics*, vol. 47, no. 6, pp. 891-5.

8 Appendix

8.1 MicroNIR data sheet

The table below is the data sheet supplied with the spectrometer.

Table 8: MicroNIR spectrometer data sheet

Parameter	Specification
Illumination source	Two integrated vacuum tungsten lamps; 18,000 hrs lifetime
Illumination geometry	Flood illumination/0° observer
Input aperture dimensions	0.5 x 2.0 mm
Active sampling area	2 x 4 mm at window face
Dispersing element	Linear variable filter
Detector type	128-pixel uncooled InGaAs photodiode array
Pixel size/pitch	30 x 250 µm/50 µm
Optical wavelength range	950 to 1650 nm or custom
Spectral bandwidth (FWHM); optical resolution	<1.25% of center wavelength, or custom. For example, at 1000 nm, the resolution is 12.5 nm.
Spectral in-band LVF blocking	>4 OD average
Analog-to-digital converter	16 bit
Dynamic range (max.)	1000:1
Measurement time (typical)	0.25 s
Sampling integration time	Minimum 100 µs; maximum limited by dark signal only
Host interface	USB 2.0, high speed (480 Mb/s)
Dimensions (diameter x height)	45 x 42 mm
Weight	<60 g (2.1 oz)
Operating environment	–20 to 40°C, noncondensing
Storage environment	–40 to 70°C, noncondensing
Power requirement	USB powered (<500 mA @ 5 V)

8.2 Temperature controller code

The following code was used to control the temperature of the spectrometers using proportional and integral control.

```
//////////////////////////////////////
//  Generated Initialization File  //
//////////////////////////////////////

// This code is used to control the temperature control for the spectrometers using a Peltier

#include "compiler_defs.h"
#include "C8051F500_defs.h"
#include "math.h"
void temp_error (void);
// Peripheral specific initialization functions,
// Called from the Init_Device() function
void PCA_Init()                                          // Initiates the PWM drive to the Peltier
{
                  PCA0CN    = 0x40;
   PCA0MD   &= ~0x40;
   PCA0MD    = 0x02;
   PCA0CPM0  = 0xC2;
   PCA0CPM1  = 0xC2;
   PCA0CPL0  = 0xFF;          //Low byte of 1700 nm PWM
   PCA0CPH0  = 0xFF;          //High byte of 1700 nm PWM
   PCA0CPL1  = 0xFF;          //Low byte of 2200 nm PWM
   PCA0CPH1  = 0xFF;          //High byte of 2200 nm PWM
}

void Timer_Init()
{
   TCON     = 0x40;
   TMOD     = 0x20;
   CKCON    = 0x02;
   TH1      = 0x06;
}

void ADC_Init()
{
   ADC0CN    = 0x80;
                ADC0MX    = 0x0F;
}

void Voltage_Reference_Init()
{
   REF0CN    = 0x13;
}

void Port_IO_Init()                              // Initiate the ports
{
   // P0.0  -  Skipped,    Open-Drain, Analog
   // P0.1  -  Skipped,    Open-Drain, Digital
   // P0.2  -  Skipped,    Open-Drain, Digital
   // P0.3  -  Skipped,    Open-Drain, Digital
   // P0.4  -  Skipped,    Open-Drain, Digital
   // P0.5  -  Skipped,    Open-Drain, Digital
   // P0.6  -  Skipped,    Open-Drain, Digital
   // P0.7  -  Skipped,    Open-Drain, Digital

   // P1.0  -  Skipped,    Open-Drain, Digital
   // P1.1  -  Skipped,    Open-Drain, Digital
   // P1.2  -  Skipped,    Open-Drain, Digital
   // P1.3  -  Skipped,    Open-Drain, Digital
   // P1.4  -  Skipped,    Open-Drain, Digital
   // P1.5  -  Skipped,    Open-Drain, Digital
   // P1.6  -  CEX0  (PCA), Open-Drain, Digital
   // P1.7  -  Unassigned,  Open-Drain, Analog
   // P2.0  -  Unassigned,  Open-Drain, Digital
   // P2.1  -  Unassigned,  Open-Drain, Digital
   // P2.2  -  Unassigned,  Open-Drain, Digital
   // P2.3  -  Unassigned,  Open-Drain, Digital
   // P2.4  -  Unassigned,  Open-Drain, Digital
   // P2.5  -  Unassigned,  Open-Drain, Digital
   // P2.6  -  Unassigned,  Open-Drain, Digital
   // P2.7  -  Unassigned,  Open-Drain, Digital
   // P3.0  -  Unassigned,  Open-Drain, Digital

   SFRPAGE   = CONFIG_PAGE;
   P0MDIN    = 0xFE;
   P1MDIN    = 0x5F;
   P1MDOUT   = 0x50;                             // configures PCA for used in PWM for push pull configuration
   P0SKIP    = 0xFF;
   P1SKIP    = 0x2F;
   XBR1      = 0x08;
   XBR2      = 0x40;
   SFRPAGE   = ACTIVE_PAGE;
}

void Oscillator_Init()
{
   int i = 0;
   SFRPAGE   = CONFIG_PAGE;
   CLKMUL    = 0x80;
   for (i = 0; i < 20; i++);    // Wait 5us for clock multiplier initialization
   CLKMUL    |= 0xC0;
   while ((CLKMUL & 0x20) == 0);
```

```
   CLKSEL   = 0x02;
   OSCICN   = 0xC7;
   SFRPAGE  = ACTIVE_PAGE;
}

void Interrupts_Init()
{
   IE       = 0x88;
}

// Initialization function for device,

void Init_Device(void)
{
   PCA_Init();
   Timer_Init();
   ADC_Init();
   Voltage_Reference_Init();
   Port_IO_Init();
   Oscillator_Init();
   Interrupts_Init();
}


// Variables used for the 2200 nm temperature controler
                double V_supply = 2.208;                                        // Supply voltage for the thermistors used
in temperature calculations.
                unsigned int ADC_read2200;                                      // ADC value read from temperature
circuit. (unsigned int shifts the value so it can only be +ve
                double voltage_2200;                                            // Voltage from thermistor circuit
                double temperature_2200 = 0.0;                                  // First temperature calculation
                double temperature_2200_2 = 0.0;                                // Last temperature calculation
                double average_temperature_2200;                                // Average temperature reading
                double target_2200 = 20.0;                                      // Target temperature
                double T_error_2200 = 0.0;                                      // Temperature error
                double P_error_2200 = 0.0;                                      // Previous error used for control
                double resistance_2200;                                         // Thermistor resistance
                double kp_2200 = 0.31*65536.0;                                  // Proportional gain
                double ki_2200 = 0.079*65536.0;                                 // Integral gain
                double cooling_2200 = 3.0;                                      // Cooling effort used
                float           effort_2200;                                    // Convert cooling effort into a float
                unsigned int value_2200;
                int count;                                                      // Used for triggering
                bit timer1_on = 0;

// Variables used for the 1700 nm temperature controller

                unsigned int ADC_read1700;                                      // ADC value read from temperature
circuit. (unsigned int shifts the value so it can only be +ve)
                double voltage_1700;                                            // Voltage from thermistor circuit
                double temperature_1700 = 0.0;                                  // First temperature calculation
                double temperature_1700_2 = 0.0;                                // Last temperature calculation
                double average_temperature_1700;                                // Average temperature reading
                double target_1700 = 20.0;                                      // Target temperature
                double T_error_1700 = 0.0;                                      // Temperature error
                double P_error_1700 = 0.0;                                      // Previous error used for control
                double resistance_1700;                                         // Thermistor resistance
                double kp_1700 = 0.31*65536.0;                                  // Proportional gain
                double ki_1700 = 0.07*65536.0;                                  // Integral gain
                double cooling_1700 = 0.0;                                      // Cooling effort used
                float           effort_1700;                                    // Convert cooling effort into a float
                unsigned int value_1700;

void tmr1 (void) interrupt INTERRUPT_TIMER1
{

                        timer1_on = 1;

}

int ReadADC(unsigned char ADCONVal)             // General read ADC function - pass ADC MUX value in
{
 SFRPAGE = ACTIVE_PAGE;                 // Switch to the required SFR page for the ADC
 ADC0MX = ADCONVal;                     // Set the multiplexer to the specified input
 AD0BUSY = 1;                   // Start conversion
 while(AD0BUSY);                        // Wait for conversion
 return (ADC0);                 // Return result
}



void temp_error (void)
{
                resistance_2200 = voltage_2200/((V_supply-voltage_2200)/33000);                 // calculate the resistance of the thermistor
                temperature_2200 = 1.6575*(log(resistance_2200))*(log(resistance_2200))-58.939*log(resistance_2200)+468.14;//calculate the
temperature.
                average_temperature_2200 = (temperature_2200+29*temperature_2200_2)/30;// Averages the temperature
                T_error_2200 = average_temperature_2200 - target_2200;                  // Reads the temperature error
                cooling_2200 = kp_2200*T_error_2200 + ki_2200*T_error_2200;// Calculates the amount of cooling needed to reach target temperature
                temperature_2200_2 = average_temperature_2200;                          // Assigns the calculated temperature to temperature2
used later for averaging

                if(cooling_2200 > 65535)                                                // Temperature is colder than target temperature
switch off PWM
                        {
                                effort_2200 = 65535;                                    // Sets PWM to minimum value
                                PCA0CPL1 = 0x01;
                                PCA0CPH1 = 0x00;
                        }

                else if(cooling_2200 < 1)                                               // Temperature is hotter than target temperature
switch PWM to maximum output
```

```
                {
                        effort_2200 = 1;                        // 16 bit PWM subtracts from maximum value to get duty cycle
                        PCA0CPL1 = 0xff;
                        PCA0CPH1 = 0xff;
                }

                else
                {
                        effort_2200 = cooling_2200; // If the effort is not a maximum value it will be used as the PWM value

                        SFRPAGE = ACTIVE_PAGE;
                        value_2200 = 65535 - effort_2200;               // Assigns the control effort to the PWM drive
                        PCA0CPL1 = (value_2200 & 0xff);                 // Writes to the lower 8 bit register by AND function
                        PCA0CPH1 = (value_2200 & 0xff00)>>8;            // Captures the higher 8 bit register using AND function
then shifts down 8 bits
                }
                P_error_2200 = T_error_2200 + T_error_2200;             // Integral part of the control loop
                resistance_1700 = voltage_1700/((V_supply-voltage_1700)/33000);
                // calculate the resistance of the thermistor
                temperature_1700 = 1.6575*(log(resistance_1700))*(log(resistance_1700))-58.939*log(resistance_1700)+468.14;
        //calculate the temperature.
                average_temperature_1700 = (temperature_1700+29*temperature_1700_2)/30;// Averages the temperature
                T_error_1700 = average_temperature_1700 - target_1700;          // Reads the temperature error
                cooling_1700 = kp_1700*T_error_1700 + ki_1700*T_error_1700;     // Calculates the amount of cooling
needed to reach target temperature
                temperature_1700_2 = average_temperature_1700;  // Assigns the calculated temperature to temperature2 used later for
averaging

        if(cooling_1700 > 65535)                                // Temperature is colder than target temperature switch off PWM
                {
                        effort_1700 = 65535;                            // Sets PWM to minimum value
                        PCA0CPL0 = 0x01;
                        PCA0CPH0 = 0x00;
                }

        else if(cooling_1700 < 1)                       // Temperature is hotter than target temperature switch PWM to maximum output
                {
                        effort_1700 = 1;                // 16 bit PWM subtracts from maximum value to get duty cycle
                        PCA0CPL0 = 0xff;
                        PCA0CPH0 = 0xff;
                }

                else
                {
                        effort_1700 = cooling_1700; // If the effort is not a maximum value it will be used as the PWM value
                        SFRPAGE = ACTIVE_PAGE;
                        value_1700 = 65535 - effort_1700;               // Assigns the control effort to the PWM drive
                        PCA0CPL0 = (value_1700 & 0xff);                 // Writes to the lower 8 bit register by AND function
                        PCA0CPH0 = (value_1700 & 0xff00)>>8;    // Captures the higher 8 bit register using AND function then shifts
down 8 bits
                }

        P_error_1700 = T_error_1700 + T_error_1700;                     // Integral part of the control loop


}

void main (void)
{

        Init_Device();
        while(1)
        {
                if (timer1_on)
                {
                        timer1_on = 0;
                        ADC_read2200 = ReadADC(0x0F);
                        voltage_2200 = ((float)ADC_read2200/4096.0)*V_supply;// Reads the ADC value and converts to a voltage
                        ADC_read1700 = ReadADC(0x0D);
                        voltage_1700 = ((float)ADC_read1700/4096.0)*V_supply;// Reads the ADC value and converts to a voltage
                        temp_error();
                }

        }

}
```