Use of sensor technology to determine health status in cattle and sheep

A thesis submitted to Dublin City University for the

Degree of Doctor of Philosophy

by

Daire Thomas Byrne B.Eng. (Hons)

[1]Dublin City University, Faculty of Engineering and Computing
School of Mechanical and Manufacturing Engineering

[2]Animal and Bioscience Research Department,
Animal and Grassland Research and Innovation Centre,
Teagasc, Moorepark, Fermoy, Co. Cork.

Research Supervisors

[2]Dr. N. McHugh **[1]Assist. Prof. H. Esmonde**

July 2019

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of doctor of philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed :_____ (Candidate) ID no. : 10347231 **Date :_____**

Table of Contents

Acknowledgements

I would like to acknowledge the opportunity presented to me by Teagasc to carry out this work in the Animal and Grassland Research and Innovation Centre at Moorepark. I would also like to acknowledge the Teagasc Walsh Fellowship scheme for funding the research.

Thank you to Dr. Noirin McHugh for your unwavering support and encouragement over the years. Your patience and willingness to ensure I really understood the topic at hand was always appreciated. Thank you for teaching me what determination can achieve. To Assit. Prof Harry Esmonde, thank you for our candid chats, I did really enjoy them. Thank you for building my confidence in the early days of the PhD. I would also like to thank Prof. Donagh Berry for his guidance during my PhD. Your honest comments and diligent nature always kept me on my toes.

Thank you to all my DCU/office mates over the years Alan, Ian, Jess, Audrey, and Shauna whose advice and support were greatly appreciated. Thank you also for providing ample distraction and an enjoyable environment to work in. To the researchers, staff, summer students, and Walsh fellows who I met during the project, thank you for your help, support and for being so approachable.

To Yvonne, what can I say, you stuck by me for the whole PhD…well done. You moved to Cork during the PhD and I am so glad you did. Thank you for being there when the times were tough, it won't be forgotten. You kept me going during the PhD and helped me to "switch off" when I was away from the desk. Thanks for being there.

Finally, mam and dad, to whom this thesis is dedicated, thank you for all the support over the years, financial and otherwise. Thank you for all the sacrifices you have made for my education, now it is time to put it to use.

List of Figures

List of Tables

Glossary of terms

BRD Bovine respiratory disease

BVD Bovine viral diarrhoea

CLP Contralateral load percentage

CMT California mastitis test

DTP Descriptive temperature parameter

FMD Foot and mouth disease

GCLM grey level co-occurrence matrices

HWC Hoof weigh crate

IDD inter digital dermatitis

MIF Most influential factor

SCC Somatic cell count

SCS Somatic cell score

SOP Standard operating procedure

IRT Infrared thermography

USST Udder skin surface temperature

Author Contribution to Publications

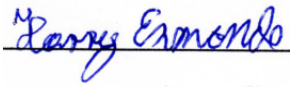
This thesis includes three manuscripts with a core theme involving the determination of health status in livestock with the use of sensor technology. The ideas, development and writing up of all manuscripts in this thesis were the principal responsibility of myself, the candidate, working within the School of Mechanical and Manufacturing Engineering in DCU under the supervision of Assist. Prof. Harold Esmonde and the Animal Bioscience department in Teagasc under the supervision of Dr. Nóirín McHugh. The inclusion of co-authors reflects the fact that part of the work came from active collaborations between researchers and acknowledgements input into team-based research. In the case of chapters 2 to 4 my contribution to the work involved the following:

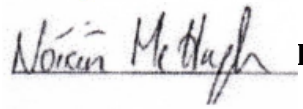
Thesis Chapter	**Publication Title**	**Publication status**	**Nature and extent of candidates contribution**
2	Temporal, spatial, inter- and intra-cow repeatability of thermal imaging	Published: Journal of animal science. 2017 Feb 1;95(2):970-9. is available online at: https://academic.oup.com/jas/article/95/2/970/4702112 doi: 10.2527/jas.2016.1005	First author, key ideas, data collection and analysis, manuscript development and writing up.

Thesis Chapter	Publication Title	Publication status	Nature and extent of candidates contribution
3	Investigation of the relationship between udder quarter somatic cell count and udder skin surface temperature of dairy cows measured by infrared thermography	Published: Journal of animal science. 2018 Jul 19;96(10):4458-70. is available online at: https://academic.oup.com/jas/article/96/10/4458/5056008 doi: 10.1093/jas/sky281	First author, key ideas, data collection and analysis, manuscript development and writing up.
4	Infrared thermography as a tool to detect hoof lesions in sheep	Published: Translational Animal Science. 2018 Dec 8. Is available online at: https://academic.oup.com/tas/advance-article/doi/10.1093/tas/txy132/5235605 doi: 10.1093/tas/txy132	First author, key ideas, data collection and analysis, manuscript development and writing up.

Thesis Chapter	Publication Title	Publication status	Nature and extent of candidates contribution
5	Sheep lameness detection from individual hoof load	Published: Computers and Electronics in Agriculture. 2019 Mar 1. Is available online at https://www.sciencedirect.com/science/article/pii/S0168169918312365 doi: 10.1016/j.compag.2019.01.048	First author, key ideas, data collection and analysis, manuscript development and writing up.

Tom Byrne (Candidate) **Signed:_____ Date 02/1/2018**

Assist. Prof. Harold Esmonde **Signed:**  **Date: 02/1/2018**

Dr. Nóirín McHugh **Signed:**  **Date: 02/1/2018**

Abstract - Use of sensor technology to determine health status in cattle and sheep

Daire Thomas Byrne

The overall objective of this thesis was to quantify the ability of new technologies to identify the health status of cattle and sheep, and to identify the factors which affect the diagnostic capabilities of these technologies. The PhD consisted of four experiments which investigated: i) the repeatability of infrared thermography in an agricultural environment, ii) the relationship between udder skin temperature (measured by infrared thermography) and the quarter somatic cell count of dairy cows, and iii) the ability of infrared thermography to detect lameness in sheep, and iv) the use of a custom hoof weigh crate to detect lameness in sheep. Approximately 8000 thermal images were manually captured throughout the PhD in an agricultural setting; these data were analysed using a variety of statistical methods including variance component analysis, mixed models, random regression models, and sensitivity and specificity analyses. Results from this thesis showed that to gain the required level of repeatability from thermal image measurements three replicate images must be captured by a trained operator. As part of this thesis, the effect of the environment on udder skin temperature was quantified but thermal imaging could not be used to estimate somatic cell count. The maximum temperature of sheep hooves proved to be a useful variable to diagnose early onset of infection in individual hooves. This thesis was the first to suggest that infrared thermography used in agriculture has the best diagnostic capabilities at colder ambient temperatures. The custom hoof weigh crate developed as part of this thesis was capable of measuring the individual hoof load of sheep, but was only a viable solution for detecting extensive infection rather than the early onset of lameness. Results from this thesis will aid in the application of infrared thermography in real world conditions, especially as a lameness detection tool for sheep.

1 General introduction

Health issues cause production and welfare to suffer on farms. Selective breeding facilitates desirable traits to be passed from parent to progeny; such techniques have been shown to be very effective at increasing production and fertility traits (Kearney et al., 2009). If animals who are genetically predisposed to be less susceptible to disease could be selected for breeding then progressive generations would be less susceptible to disease and so disease prevalence may decrease. In order for such a system to work effectively animals must be identified correctly as susceptible or resistant to disease. The difficulty here is the current lack of data that exists for health traits. Additionally, as health traits are lowly heritable, large quantities of data are required to make accurate assessments (Berry et al., 2011). Current methods for recording health traits are labour intensive and time consuming so it is currently not feasible to gather large quantities of data.

Some of the most prevalent diseases in livestock include mastitis (infection in the mammary gland) and lameness (infection on the hoof). Currently, the best method (i.e., the gold standard) for detecting infection in the udder is a laboratory analysis of the number of somatic cells in a milk sample (Harmon, 1994). The milk sample can come from one or all four of the udder quarters of a cow. For hoof infections the current gold standard involves scoring the animal as they walk (locomotion scoring) or inverting each animal and visually scoring each hoof based on infection type and severity (hoof scoring). Both of these measures are subjective as different assessors can assign different scores to the same animal.

The overarching aim of the current thesis was to assess the use of different technologies to quantify health status in livestock and to investigate the factors which may hinder the diagnostic capabilities of these technologies. Infrared thermography (IRT) was the first technology to be tested as part of the current thesis as it is a non-invasive technology which, unlike many other technologies, has shown the potential to identify a

number of different diseases in livestock including, foot and mouth disease (FMD) (Rainwater-Lovett et al., 2009), bovine viral diarrhoea (BVD) (Schaefer et al., 2004)(Schaefer et al., 2004), bovine respiratory disease (BRD) (Schaefer et al., 2012), mastitis (Colak et al., 2008; Polat et al., 2010), and lameness (Alsaaod and Büscher, 2012; Stokes et al., 2012). All thermal images in the current thesis were captured manually (Figure 1.1) when the animals (cows or sheep) were either standing in a herringbone milking parlour or temporarily restrained in a head gate. Each experiment captured images of the hooves, eye, or udder. All images were manually cropped using custom software which automatically extracted the temperature parameters, this software was built as part of the current thesis using the Thermovision LabVIEW toolkit 3.3 (FLIR Systems Inc., Stockholm, Sweden.).

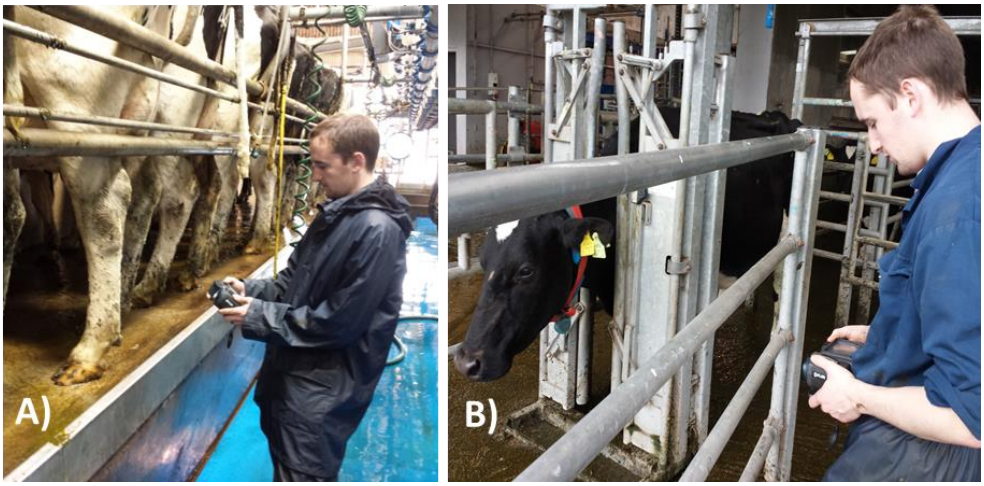


Figure 1.1: The method by which thermal images were captured from the udder (A) and the eye (B) of dairy cows is shown

As there was no standard operating procedure for capturing thermal images of cattle or sheep, the first objective of the current thesis was to quantify the repeatability of thermal image measurements captured from the eyes, udder and hooves of cows in an agricultural environment. In this first paper, the number of replicates required to obtain a certain level of precision was evaluated. Precision was defined as the 95% CI range within

which the (average of the) measured temperature(s) was expected to lie relative to the gold standard; the gold standard temperature of an entity in this study was the average of 30 temperature measurements. Furthermore the variation in temperature measurements due to different operators was quantified.

Once the number of replicate images required to gain a certain level of precision was quantified and a Standard operating procedure (SOP) was defined, the next objective was to investigate whether IRT could be used to detect the most prevalent disease in the dairy industry, mastitis. The somatic cell count (SCC) of a milk sample is known as a proxy for mastitis (Harmon, 1994). Therefore, in this, the second paper, SCC was compared to the udder skin surface temperature (USST) as measured by IRT. A total of 14 cows were milk sampled and imaged every day for two months. Additionally, 18 different temperature parameters were extracted from all thermal images. A number of different statistical methodologies were applied to this large dataset to relate USST to SCC and to predict USST on a daily basis.

As the use of IRT to detect mastitis showed little to no promise, the decision was made to move the application of technology to the leading cause of ill health in sheep, infection in the hoof (Dohoo et al., 1985). Firstly, similar to paper 1 a repeatability study was performed to investigate the precision of hoof thermal images captured from sheep. Secondly an exploratory study was performed to investigate if hoof infection could be identified using hoof temperature (measured by IRT). Additionally, the effect of ambient temperature on the diagnostic capability of IRT was also investigated herein. Finally, a validation study was performed to assess if the methods from the exploratory study would yield the same results with a new cohort of ewes.

Since the diagnostic capability of IRT was largely impacted by ambient temperature, the fourth paper focused on measuring the load a ewe places on each hoof in

order to detect lameness. A hoof weigh crate was designed, built, programmed and tested in real world conditions as part of this, the fourth paper.

1.1 Literature Cited

Alsaaod, M., and W. Büscher. 2012. Detection of hoof lesions using digital infrared thermography in dairy cows. J. Dairy Sci. 95: 735-742.

Berry, D. P., M. L. Bermingham, M. Good, and S. J. More. 2011. Genetics of animal health and disease in cattle. Ir. Vet. J. 64(1): 5.

Colak, A., B. Polat, Z. Okumus, M. Kaya, L. Yanmaz, and A. Hayirli. 2008. < I> short communication: early detection of mastitis using infrared thermography in dairy cows. J. Dairy Sci. 91: 4244-4248.

Dohoo, I. R., R. A. Curtis, and G. G. Finley. 1985. A survey of sheep diseases in canada. Can. J. Comp. Med. 49(3): 239-247.

Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell Counts1. J. Dairy Sci. 77: 2103-2112.

Kearney, F., A. Cromie, and D. P. Berry. 2009. Implementation and uptake of genomic evaluations in ireland. Interbull Bulletin. (40): 227.

Polat, B., A. Colak, M. Cengiz, L. Yanmaz, H. Oral, A. Bastan, S. Kaya, and A. Hayirli. 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. J. Dairy Sci. 93: 3525-3532.

Rainwater-Lovett, K., J. M. Pacheco, C. Packer, and L. L. Rodriguez. 2009. Detection of foot-and-mouth disease virus infected cattle using infrared thermography. Vet. J. 180: 317-324.

Schaefer, A., N. Cook, S. Tessaro, D. Deregt, G. Desroches, P. Dubeski, A. Tong, and D. Godson. 2004. Early detection and prediction of infection using infrared thermography. Can. J. Anim. Sci. 84: 73-80.

Schaefer, A., N. Cook, C. Bench, J. Chabot, J. Colyn, T. Liu, E. Okine, M. Stewart, and J. Webster. 2012. The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. Res. Vet. Sci. 93: 928-935.

Stokes, J., K. Leach, D. Main, and H. Whay. 2012. An investigation into the use of infrared thermography (IRT) as a rapid diagnostic tool for foot lesions in dairy cattle. Vet. J. 193: 674-678.

2 Literature review

2.1 Cattle and sheep production in Ireland

2.1.1 Dairy

There are 1.3 million dairy cattle (CSO, 2017a) producing 7.3 billion litres of milk (CSO, 2017b) per annum in Ireland. Approximately 7.4 % of the milk produced in Ireland is consumed as such while the rest is sold as butter, cheese, and milk powder (CSO, 2017b); as a result, farmers are financially rewarded for increased fat and protein which is key to the processing of milk. Total dairy exports were worth 4.6 billion in 2017 (Department of Agriculture, Food and the Marine, 2018). Dairy farming in Ireland is primarily pasture based and involves seasonal calving (Dillon et al., 1995). Predominantly, dairy farmers plan to produce a calf per cow between February and March to ensure peak feed demand coincides with peak grass production which occurs between May and June, this facilitates utilization of the cheapest feed available (i.e., grass) (Finneran et al., 2010). Calves and the cows are separated shortly after calving and then the cows 300 day lactation period begins where the cow will on average produce 5,397 litres of milk (i.e., 401 kg of milk solids) (Teagasc, 2017). Cows can be milked in tandem, herringbone, parallel, or rotary parlours where an operator must attach a cluster onto each udder. Alternately, cows can be trained to come into an automatic milking system where the cluster is automatically attached and large amounts of data can be gathered. Profitable dairy farming in Ireland is driven by the amount of milk solids a cow can produce and the efficient use of grass (Shalloo et al., 2014; Hanrahan et al., 2018).

2.1.2 Beef

Each year 1.7 million animals are slaughtered for beef in Ireland, while 189,000 animals are exported live; the output of the beef sector was valued at 2.4 billion per annum (Department of Agriculture, Food and the Marine, 2018). The average number of beef cows per farm is 14 (Department of Agriculture, Food and the Marine, 2018). Each

beef/suckler farmer aims to produce one calf per cow per year. After 8 to 9 months, the calf is separated from the mother and is considered a weanling. The beef farmer can choose to keep the weanling or sell it to farmers who specialise in fattening animals to a final slaughter weight. Animals are slaughtered anywhere between 16 and 30 months of age. The producer is paid on carcass weight, carcass confirmation, and fat content.

2.1.3 Sheep

There are 3.9 million sheep across 36,000 flocks in Ireland with 62,000 tonnes of sheep meat being exported annually (Department of Agriculture, Food and the Marine, 2018). The output of the sheep sector is valued at €262 million annually (Department of Agriculture, Food and the Marine, 2018). In Ireland, most sheep are grazed on the lowlands while others are grazed on hill and mountainous regions with ewes rearing 1.3 lambs per ewe joined to the ram annually. Some lambs are retained for breeding while the majority are sold for slaughter as soon as they reach a desired live weight or are sold off farm for further fattening. The key drivers of profit on an Irish sheep farm are the number of lambs weaned per hectare and the amount of grass utilised (Bohan et al., 2018).

2.2 Current gold standards for measuring health traits

The dairy, beef and sheep sectors each suffer from a slightly different set of health issues. The health issues which have a major impact on the dairy sector are mastitis, lameness, and infertility (Huxley, 2013; Geary et al., 2013; Shalloo et al., 2014). Some of the most prevalent diseases on a beef farm include lameness and BRD (Carne et al., 1964; Healy et al., 1993) with lameness, parasitic infections, and mastitis being the main health issues on sheep farms (Dohoo et al., 1985).

To improve the health status of the Irish livestock, elite animals that are genetically less susceptible to disease need to be identified and selected to become parents of the next

generation. The heritability of most health traits is low therefore large numbers of accurately recorded health records are required to accelerate genetic gain in such traits. The collection of health data is viewed as labour intensive and costly by Irish producers. Access to large quantities of accurately recorded health data is the main hindrance for the inclusion of health traits in Irish beef, sheep and dairy breeding programmes. The design of a low cost (cost limits must be determined by the type of user (farmer or geneticist) and the accuracy of the system) system to routinely record health data through use of automated and non-invasive methods will ensure that large volumes of accurately recorded data can be generated for important health traits. Two of the most common and costly (due to treatment and production loss) diseases across dairy, beef and sheep farms are mastitis (Bradley, 2002) and lameness (Huxley, 2013), therefore, research on the development of farmer friendly and robust technologies will focus on these traits.

2.2.1 Mastitis

Mastitis is the inflammation in the udder of the animal due to infection. When clinical mastitis occurs, the blood vessels dilate in the infected quarter of the udder causing an increase in blood flow to the area (Martins et al., 2013), which causes a temperature increase at the site of infection. This infection can occur naturally or due to poor management practices including: under or over milking and poor hygiene in the milking parlour. The incidence of mastitis has been reported to be as high as 41.6 cases/100 cows/ year (Bradley and Green, 2001) and has been reported to cause behavioural changes in dairy cows due to discomfort (Medrano-Galarza et al., 2012). Mastitis is estimated to cost the Irish dairy industry €37.6 million annually (Geary et al., 2013).

A common method for detecting infection in the mammary gland is the estimation of SCC in a milk sample; the udder is defined as healthy if the SCC ranges between 0 and 250,000 cells/ml, subclinical mastitis ranges from 250,000 cells/ml to 400,000 cells/ml and

clinical mastitis occurs at >400,000 cells/ml of milk (Beaudeau et al., 1998; Berry and Meaney, 2006). Subclinical mastitis is defined as the early stages of mastitis and if detected at this early stage can be treated successfully. Where laboratory analysis is unavailable the California Mastitis Test (CMT) can be performed, where a small volume of milk is taken from each of the four quarters (i.e., the four mammary glands which together are called the udder) and placed in four separate dishes. A predefined amount of a liquid test agent such as sodium hydroxide is added to each sample, when thoroughly mixed the resulting mixture begins to turn into a gel in the presence of high SCC (Whiteside, 1939). The four dishes can simply be compared as a relative test or a scoring system can be applied. The scale generally has 4 increments: 0 defined as healthy, 1+ defines when the animal is infected with clinical mastitis, 2+ is a more severe case of clinical mastitis and 3+ the most extreme case of clinical mastitis (Godden et al., 2017).

While some automated milking machines can approximate SCC on an individual udder quarter level, most milking parlours in Ireland use a cluster which combines the milk from each quarter before any milk tests can be performed. In this scenario, samples for SCC estimation and the CMT must be extracted manually before milking which can be labour intensive. While the SCC is generally estimated with laboratory equipment and is therefore empirical, the CMT can be subjective as it is based on a visual assessment of the consistency of the milk sodium hydroxide mixture.

2.2.2 Lameness

Lameness is one of the most costly diseases in both the dairy and sheep industry (Winter, 2008). The costs occur indirectly as when an animal is lame it can affect animal performance (Winter, 2008). Under wet conditions animals become increasingly susceptible to lameness as there is greater opportunity for bacteria to enter the hoof. The gold standard for the assessment of lameness is locomotion scoring. A trained observer

assesses the animal based on curvature of the back, the gait, and the weight distribution of hooves (Kaler et al., 2009a). There are numerous different scales but in one, an animal can be ranked on lameness from 0 to 6, with 0 indicating normal locomotion and 6 indicating the animal is unable to move (Kaler et al., 2009b). Although locomotion scoring is an accurate measurement of overall lameness, it does not define the cause of lameness. It does not define which hoof is infected or the severity of infection in that hoof; in some cases multiple hooves can be infected simultaneously with varying stages of infection. Since a range of contagious and non-contagious infections can cause lameness (Winter, 2004), there is no universal scale for defining the extent of infection in a hoof. Instead, each disease has its own scale, for example, the most prevalent cause of lameness in sheep is scald which can become foot rot. This scale ranges from 0 to 4, with 0 being healthy and 4 being the most extreme form of foot rot (Conington et al., 2008).

To obtain a hoof score a sheep must be manually inverted and each hoof visually assessed. This process can be labour intensive and time consuming. Additionally, the process can be subjective as it is possible for different operators to record different scores for the same hoof. As a result large scale or frequent hoof scoring is simply not done; current methods for reducing lameness prevalence include blanket treatment or individual treatment of severely infected hooves. Therefore, a new fast, non-invasive technology is required to gain a better insight into hoof health traits.

2.3 Digitization of health traits

There is a paucity of research in the area of using sensors to detect a biological change in an animal due to infection. Some specific studies have tested load sensing equipment to detect lameness (Pastell et al., 2008b; Maertens et al., 2011), others have tested various core body temperature measurements to detect pneumonia (Davis et al.,

2003b; Reuter et al., 2010), and more studies have furthered the knowledge of IRT as a disease detection tool (Berry et al., 2003; Polat et al., 2010; Gloster et al., 2011; Alsaaod and Büscher, 2012).

2.3.1 Internal body measurements

Rectal or vaginal temperature is used as a measure of core body temperature. Previous studies have portrayed a strong correlation between vaginal and rectal temperatures (George et al., 2014). In general, vaginal temperature is used to predict ovulation and rectal temperature is used for a core body temperature measurement. Rectal temperature does not vary when the animal is resting or during activity (Berry et al., 2003). It has been suggested that rectal temperature does not vary when an animal is fighting infection such as mastitis (Colak et al., 2008), however other studies disagree and have shown an increase in rectal temperature is strongly correlated to mastitis (Hovinen et al., 2008; Pezeshki et al., 2011a). Some systems are commercially available for the continuous measurement of vaginal temperature, such as VelPhone (Medria, France), while devices for continuous rectal temperature measurement have been developed for research (Reuter et al., 2010). Current methodologies for the measurement of rectal or vaginal temperatures require the insertion of a measuring device into the animal, which can be time consuming, expensive, and labour intensive.

Tympanic (inner ear) temperature can also be used as a measure of core body temperature (Davis et al., 2003a; Davis et al., 2003b) and has been used to detect disease in calves (Mahendran et al., 2017). FeverTag (Texas, USA) is a system currently on the market which measures tympanic temperature. When the FeverTag measures a tympanic temperature consistently >39.7°C for a period of six hours the animal is classified as sick (Mahendran et al., 2017). Tympanic temperature has further been used to evaluate the effect of feed intake and feeding time on body temperature (Davis et al., 2003b). An

animal's feeding regime, water cooling systems and coat colour can affect an animal's inner ear temperatures by as much as 0.5°C during heat stress conditions (Davis et al., 2003b). Additionally, a rise in ear temperature may be caused by activity and depending on climatic conditions it may take as long as 2 to 3.5 hours for the ear temperature to return to its pre-exercise value (Mader et al., 2005). The magnitude of the ear temperature rise can be dependent on the temperament of the animal (Aland and Banhazi, 2013). In addition to many factors affecting tympanic temperature, attaching ear probes to each animal can be expensive, especially considering that not all ear probes will remain in place.

2.3.2 Thermal imaging

2.3.2.1 Introduction to thermal imaging

Infrared thermography is a non-invasive technology that could be incorporated into automated systems, such as drafting gates or automatic milking machines. This technology creates a pictorial representation of the heat distribution from any object. Previously, IRT has been used to detect FMD (Rainwater-Lovett et al., 2009; Gloster et al., 2011), BVD (Schaefer et al., 2004), BRD (Schaefer et al., 2012), mastitis (Colak et al., 2008; Polat et al., 2010), and lameness (Nikkhah et al., 2005; Alsaaod and Büscher, 2012) in livestock. Infrared thermography has also been automated in an agricultural environment (Schaefer et al., 2012). Every pixel in a thermal image is a single point temperature measurement and in order to make these readings as accurate as possible, factors such as: ambient temperature, humidity, reflected temperature, distance between the camera and the object and emissivity must also be quantified (Avdelidis and Moropoulou, 2003). Emissivity is a measure of how a body emits energy via radiation and ranges from 0 to 1, where 1 is an ideal body that emits the most energy. For example, the emissivity of pig skin is 0.98 (Soerensen et al., 2014). Soerensen et al (2014) calculated the emissivity of pig skin by comparing temperature probe measurements of the pig skin to thermal images captured in tandem; any

discrepancies between probe and camera measurements were attributed to an error in emissivity. Some studies suggest an entirely new method for determining emissivity is required as the current method is difficult to apply to livestock (Church et al., 2014) but no alternative methods have been suggested in literature. Additionally, no study has investigated emissivity of the eye, which may be useful as a core body temperature measurement (Gloster et al., 2011). Studies have been found to mathematically derive the emissivity of the human eye to range from 0.97 to 1.00 (Mapstone, 1968).

2.3.2.2 Infrared thermography of the udder

Detection of mastitis using thermal imaging has been studied under different environments and previous studies have reported contrasting degrees of success (Colak et al., 2008; Hovinen et al., 2008; Polat et al., 2010; Pezeshki et al., 2011a). Previous research has shown IRT is capable of detecting subclinical mastitis (Colak et al., 2008; Polat et al., 2010). Both Polat et al. (2010) and Colak et al. (2008) used the same controlled conditions (i.e., ambient temperature between 18 and 23°C), breed of cows (i.e. Brown Swiss), thermal image view (i.e., lateral and caudal), acclimatization period (i.e., 30 minutes) and thermal camera (i.e., IR FlexCam S) and reported a temperature difference between healthy and infected udders of $\Delta T = 0.89$°C. A study by Pezeshki et al. (2011) showed that the difference between a healthy and sick animal can be as high as $\Delta T = 2$ to 3°C; this large difference may be mostly attributed to the fact that the udder was inoculated with *E.coli* whereas within other studies the infection was contracted naturally. Pezeshki et al. (2011) also differed from Polat et al. (2010) and Colak et al. (2008) by using an emissivity of 0.93 and not describing the acclimatization period before imaging or the ambient temperature during imaging. Hovinen et al. (2008) also inoculated the udder quarters but observed a difference of $\Delta T = 1$ to 1.5°C between sick and healthy udder quarters. Some of main differences between Pezeshki et al. (2011) and Hovinen et al. (2008) was the distance from the camera to the udder (1.5m vs. 0.5m), the emissivity (0.98 vs. 0.93) and the thermal

image view (caudal vs. lateral). Similar to Pezeshki et al. (2011), Hovinen et al. (2008) did not quote the acclimatization period before imaging or the ambient temperature during imaging. Another study which inoculated udders of sheep were unable to discriminate between healthy and infected udders (Castro-Costa et al., 2014). One possible reason for the differences between studies may be due to the type of bacteria which causes mastitis, Viguier et al. (2009) suggested that not all cases of mastitis cause a rise in temperature. Despite this reported variation, one system has been commercialised by Agricam (agricam, Linköping, Sweden) called CaDDi Mastitis, though the degree of success of this system is unknown.

When infection occurs, naturally or due to inoculation, blood flows to that area to fight infection; however if the infection takes hold inflammation can occur and the rate of blood flow can decrease causing a decrease in temperature (Martins et al., 2013). Therefore, the window in which a temperature increase can be detected may be limited and measurement frequency is an important factor to consider. Previous studies observed a short spike (3 hours long) in USST during the subclinical phase of infection, but as the disease progressed the temperature decreased (Hovinen et al., 2008; Pezeshki et al., 2011a). Both Hovinen et al. (2008) and Pezeshki et al. (2011) inoculated cows with *E.Coli* which leads to a more aggressive infection than would occur under normal farm conditions. Therefore, while it is not documented in literature, due to the less aggressive nature of natural infection the window for detection of naturally contracted mastitis could be much longer than inoculated mastitis; this may allow for the detection of subclinical mastitis over a longer time-frame. Additionally, repetitive measurements of the same animal will provide a baseline temperature for each animal and may aid with the identification of naturally contracted infection through abnormal temperature changes over time (Gloster et al., 2011).

2.3.2.3 Infrared thermography of the hoof

In research, the thermal view of the hoof has been successfully used to detect hoof lesions in cattle (Nikkhah et al., 2005; Alsaaod and Büscher, 2012; Stokes et al., 2012). In the past, there has been little consistency between studies on the hoof thermal image view and on the temperature variable extracted from thermal images. More recently the maximum hoof temperature extracted from the plantar aspect of the hoof has been identified as the best method for detecting lameness in cattle (Harris-Bridge et al., 2018). Two independent studies have extracted this variable from the hooves of dairy cows (breed not specified in one paper) post trimming; one study found that the average temperature difference between infected and healthy hooves from the same animal was 1.42°C (Alsaaod and Büscher, 2012) while another found that across a herd the maximum hoof temperature of infected hooves can be 7.9°C hotter than healthy hooves (Stokes et al., 2012). While the type of temperature parameter was the main difference between Alsaaod and Büscher (2012) (average temperature) and Stokes et al. (2012) (maximum temperature), a difference in the severity of infection could also have led to the disparity results, but this cannot be quantified as neither study detailed the severity of infection. Some work in lameness detection has also been conducted in horses and sheep (Eddy et al., 2001; Talukder et al., 2015), but no study has tested the ability of IRT to diagnose individual sheep hooves. Currently, hoof temperature or the difference in hoof temperatures are unable to distinguish between different forms of infection in the hoof (Stokes et al., 2012).

The capability of IRT to detect FMD has also been investigated (Rainwater-Lovett et al., 2009). Foot and mouth disease is a highly contagious disease of cloven hoofed animals (Grubman and Baxt, 2004). Research on the use of IRT in quarantined areas suggested that cow hoof temperatures (measured by IRT) can increase by approximately 4°C when infected by FMD (Rainwater-Lovett et al., 2009). However, Rainwater-Lovett et

al. (2009) acknowledge that other unrelated factors (e.g. lameness or ambient temperature) can also cause a similar increase in temperature and so more research is required to validate IRT as a FMD detection tool.

2.3.2.4 Infrared thermography of the eye

As IRT is a non-invasive technology, researchers have been attempting to identify a correlation between the temperature of an IRT measured anatomical area and rectal temperature to use it as a surrogate biomarker for disease detection (George et al., 2014). Across anatomical areas, eye temperature measured using IRT was found to have the strongest correlation with rectal temperature in comparison to other regions (i.e., nose, ear, hooves, side and back of body) (Gloster et al., 2011; Johnson et al., 2011; George et al., 2014). In addition, no temperature difference was reported between left and right eyes (George et al., 2014; Church et al., 2014). In the past, IRT measured eye temperatures have been used to identify animals infected with BVD (Schaefer et al., 2004), BRD (Schaefer et al., 2012) and FMD (Gloster et al., 2011). An eye temperature increase of 1°C was noted for animals that became infected with bovine respiratory disease (Schaefer et al., 2012) while animals who became infected with BVD showed an eye temperature increase of 2.6°C (Schaefer et al., 2004). The eye temperature difference between a healthy animal and an animal with BRD/BVD is much smaller than the hoof temperature difference between a healthy and lame hoof; this may be due to the physiology of the different anatomical areas or the nature of the respective infections but has not been investigated in the literature. No single study has identified methods for differentiating between BVD, BRD, and FMD when using eye temperature as a biomarker. Additionally, no study has used IRT in combination with other phenotypes to identify or differentiate between diseases.

2.3.2.5 Factors affecting IRT

When using a non-contact measurement system (e.g., IRT) to quantify a change in temperature of an animal due to infection, the user must be aware of the factors that can affect the accuracy of the measurement device and also the factors that can naturally change the temperature of the animal. To ensure the thermal camera measurements can account for environmental noise, a number of factors must be entered into the thermal camera software including: ambient temperature, relative humidity, emissivity, the temperature reflected off the object and the distance from camera to the object. Even after these values have been entered into the software, distance between the camera and the object distance can still affect the temperature measurement as seen in previous literature (Church et al., 2014; Montanholi et al., 2015); therefore this distance should be kept consistent between images to reduce the variability in temperature measurements. In medical research, a standard operating procedure has been developed for imaging humans (Ammer, 2008); through this a high reproducibility (intra class correlation = 0.99) and repeatability (intra class correlation = 0.99) of IRT measurements has been achieved (Fernández-Cuevas et al., 2012). No such definitive standard operating procedure has been developed for capturing thermal images of livestock (Church et al., 2014; Montanholi et al., 2015).

Many studies have mitigated factors which affect the temperature of an anatomical region by keeping ambient temperature constant (Schaefer et al., 2004; Polat et al., 2010). This is not possible in a real world scenario and since the temperature difference between sick and healthy animals can be small (Colak et al., 2008), the causes of change in the temperature of anatomical regions must be identified and quantified. Within the literature different factors have been identified to affect different regions. Udder skin surface temperature has been shown to be affected by activity, ambient temperature, the time of day (Berry et al., 2003), and the milking machine (Kunc et al., 2000). Berry et al. (2003)

suggested that using three consecutive days of USST temperature measurements and knowledge of the environmental temperature is required to accurately model healthy USST variation. A multitude of factors have been found to affect hoof temperatures, such as: stage of lactation, age of the animal, dirt, circadian rhythm (i.e., the natural 24 hour cycle of the temperature of an anatomical region), feed efficiency (some animals convert a proportion of feed into heat rather than body weight), and ambient temperature (Nikkhah et al., 2005; Montanholi et al., 2009; Gloster et al., 2011; Alsaaod and Büscher, 2012; Stokes et al., 2012). Ambient temperature has also been shown to affect the variability of hoof temperatures, as ambient temperature decreases variability increases (Gloster et al., 2011). Factors which have been determined to be correlated with eye temperature measurements include: wind speed and solar loading (Johnson et al., 2011; Church et al., 2014). Ambient temperatures may not significantly affect eye temperatures (Gloster et al., 2011). Previous work in humans has shown that the part of the brain which controls circadian rhythms (the suprachiasmatic nucleus) also has some control over core body temperature (i.e., the core body temperature of mammals has a circadian rhythm) (Buhr et al., 2010). While no study has specifically studied circadian rhythms of eye temperature in livestock (i.e., the 24 hour cycle of eye temperature), circadian rhythms have been observed in other surrogate core body temperature measurements in cattle (Davis et al., 2003b).

2.3.3 Hoof pressure pads

2.3.3.1 Emfit

The Emfit sensor designed by Emfit Ltd., Finland is a quasi-piezoelectric mat that is used for measuring dynamic forces. The dimensions of the sensors varies from 300mm to 50 metres long and can be either 290 or 580mm wide (Emfit L-series). The system was used to detect lameness (Pastell et al., 2008b) and compared to a system using four static load cells (Pastell et al., 2008a). As an animal steps on the mat it is compressed and with

signal conditioning an output voltage can be generated, the length and width of the mat was tailored to ensure only one hoof was on the sensor at any point in time. After calibration, time and force variables were extracted from the Emfit sensor, which were used to detect lameness. An ideal output from the system is shown in Figure 2.1 (Pastell et al., 2008b). This figure shows the waveform starting when the front hoof hits the sensor and finishing when the hind hoof leaves it. Pastell et al. (2008b) has justified the use of the Emfit sensor for detecting lameness in cattle. Other studies used load cells to measure the dynamic load an animal places on the ground while walking and also identified this variable as a useful lameness detect tool (Rajkondawar et al., 2006; Skjøth et al., 2013).

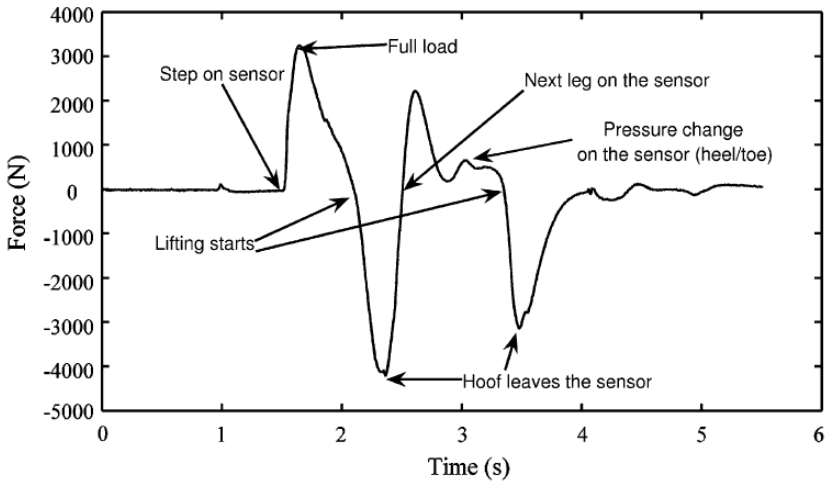


Figure 2.1: Ideal output of a healthy cow from the Emfit sensor (Force vs Time) (Pastell et al., 2008b).

2.3.3.2 Gaitwise

The GAITWISE system is a mat which uses over 18,000 pressure sensitive sensors over a surface area of 0.61m x 4.88m (Maertens et al., 2011). The system provides the x and y position co-ordinates of an animal's hooves (Figure 2.2c), the time each hoof spends in a particular position (Figure 2.2b), and the relative force exerted by each hoof (Figure

2.2a). Each hoof in Figure 2.2 is colour coded except for Figure 2.2a, where relative forces variables are colour coded instead.

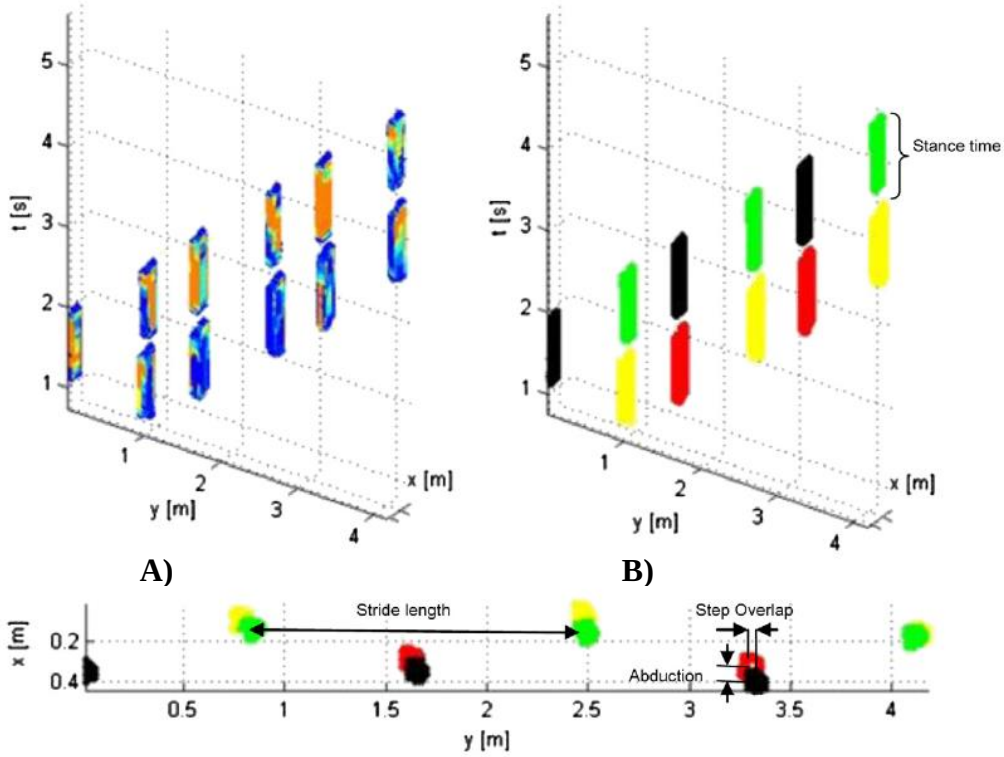


Figure 2.2: Outputs from Gaitwise system from a single animal, after data conditioning. The lower graph (C) shows the locations of each hoof imprint on the measurement area, both x and y co-ordinates. (Red = right front; Black = right hind; yellow = left front; green = left hind). In the upper right graph (B) the same data is shown with the addition of the duration each hoof spent on the system shown on the vertical axis of the 3D graph. The upper left graph (C) is similar to graph (B) but also includes detailed information on the pressure each footfall exerted on the system, the relative scale ranges from grey (lowest pressure) through cyan, yellow, magenta and red to blue (Highest pressure) (Maertens et al., 2011).

The system was compared to a visual assessment of lameness and showed that the system has the potential to be used to differentiate between healthy, mildly lame and severely lame animals when multiple gait variables are analysed (Van De Gucht et al., 2017b). Similar to

many other systems, erroneous measurements can be caused by cows stopping on the platform, running too fast across it, slipping on the mat when it is wet or multiple cows crossing the mat simultaneously (Maertens et al., 2011). To overcome these issues a roof to protect the mat from the weather and holding gates which only allowed one cow to go over the mat at any one time were designed and this decreased the percentage of unusable data from 50% to 20% (Maertens et al., 2011). The system has been reported to be highly accurate when used to detect severe lameness (sensitivity 90% and specificity 100%) (Maertens et al., 2011). As the GAITWISE system is very expensive it is not implemented on farm, to reduce cost, studies suggested the distance between sensors can be increased from 1.27cm to 5.08 and the overall length of the mat can be reduced from 4.8m to 3.2m thus reducing the total number of sensors required (Van De Gucht et al., 2017a).

2.3.3.3 Static load cells

In research, four platforms (one for each hoof) with strain gauge load cells beneath them was used to detect lameness in cattle, pigs, and horses (Hood et al., 2001; Pastell et al., 2008a; Pluym et al., 2013). When the animal becomes lame they have a tendency to remove weight from the effected hoof (Angell et al., 2015) which can be measured using static load cells (Neveux et al., 2006). The animals live weight can also be calculated from these four measurements. The minimum amount of time an animal must stand on the load cells in order to achieve an adequate measurement has not been identified and the total percentage of erroneous measurement periods can be high as high as 10% (Pastell et al., 2008a). Erroneous measurements were caused by factors such as: animals not standing on the platforms correctly, animals resting on the side of the crate, animals not settling for the duration of the measurement (restless animal), and animal individuality (Pastell et al., 2008a).

2.4 Statistical analysis

In agriculture, many gold standard measures of illness can be labour intensive and costly to gather; therefore, novel measurements are required to replace the gold standard to reduce labour and increase the frequency with which animals are measured for illness. Relating a novel measurement (e.g., IRT measured udder temperature) to an existing gold standard (e.g., Somatic cell count) is a multistage process. Firstly the repeatability (i.e., the variation between replicate measurements) of the novel measurement must be assessed to help to define the procedure with which the measurement is taken (Ammer, 2008). Secondly, how the novel measurement changes with an increase in the gold standard measurement must be assessed; does the novel measurement change at all with a unit increase in the gold standard, does it increase or decrease, if so by how much (Pluym et al., 2013)? Is there a non-linear relationship between the novel measurement and the gold standard? The relationship between the novel measurement and the gold standard can also be affected by extraneous factors (e.g., ambient temperature or differences between animals); the effect of each factor must be identified, quantified and accounted for in order for the novel measurement to be practical in a real world scenario (Nikkhah et al., 2005). As it is only possible to gather a relatively small number of measurements certain statistical inferences must be made to relate the results of the current experiments to the wider dataset. Finally, the ability of the novel measurement to correctly diagnose individual animals must be assessed (Stokes et al., 2012).

2.4.1 Statistical terminology

- The mean is the average of a set of numbers and can be calculated by summing all numbers in the dataset and dividing by the number of values in the dataset.

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

where μ is the mean, x_i represents each number in the dataset, and n is the number of elements in the data set.

- The standard deviation (SD) is a measure of variation in a data set, ±1 SD denotes where 68% of the values in a data set lie, ±1.96 SD's denotes where 95% of the values in a data set lie. The mean and standard deviation are often used to summarise a data set. The population standard deviation can be calculated as follows:

$$\sigma_n = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

Where σ is the standard deviation, x_i represents each number in the dataset, μ is the mean, and n is the number of elements in the data set.

- The standard error (SE) is the range around a sample mean which denotes where the mean of the entire data set should lie (i.e., the true mean) and is calculated as:

$$\pm se = \frac{\sigma}{\sqrt{n}}$$

Where se is the standard error, σ is the standard deviation, and n is the number of elements in the data set.

- The coefficient of variation (CV) is a standardised measure of variation, where standard deviation directly quantifies variation in a dataset once it is divided by the respective mean the calculated quantity is known as the coefficient of variation and is expressed as a percentage. The CV is calculated as:

$$CV = \frac{\sigma}{\mu}$$

Where σ is the standard deviation and μ is the mean.

- Variance is the average squared deviations from the mean and is also equal to the standard deviation squared. Variance is always greater than or equal to zero.

- Correlation (r) is a measurement of agreement between two variables. Correlation ranges from -1 to 1 and the closer a correlation is to -1 or +1 the stronger the relationship between two variables. The sign denotes how the dependant variable changes in relation to the dependant variable (e.g., when the sign is minus the dependant variable decreases as the independent variable increases).
- The coefficient of determination (R^2) is the percentage of variation in one variable that is predictable from another variable. The closer R^2 is to one the stronger the relationship between both variables.
- A fixed effect is one which is constant across individuals and generates the same variation in the experimental sample as it would in the entire population. Fixed effects are usually fitted to correct for differences in the mean of factors such as ambient temperature, gender and age, and usually have a finite number of levels (Mrode, 2005).
- A random effect is defined as a factor which has a large number of possible levels but only a random set of levels exist in the dataset. The effect itself can often consist of an infinite number of levels; an experimental unit (e.g., a cow) is generally used as a random effect (Mrode, 2005). Change in the dependent variable from one cow to the next often cannot be estimated due to the complexity of the biological mechanisms or the number of animals used in the study.
- An interaction is when the effect of an independent variable on a dependant variable is altered due to a third variable. An interaction is the result of the combined effects of a number of factors.
- Accuracy is a term to define how far a measurement is from the true value.
- Precision/ repeatability are defined as the deviation of a measurement from the average of a set of replicate measurements. Precision and accuracy are independent of each other.

- Reproducibility defines the variation of replicate measurements made by different operators.

2.4.2 Regression

Regression describes the relationship between the dependant variable (y) and an independent variable (x). When data is generated on how y varies with a changing x value, a scatter plot can be created and a line of best fit (i.e., a line which achieves the lowest possible amount of total error between estimated values and actual data) added. This simple linear regression line can be defined by the following equation:

$$y = mx + c$$

Where y is the expected output, m is the regression co-efficient and defines the slope of the regression line, x is the independent variable input and c is a constant which denotes where the regression line intersects the y-axis.

2.4.3 Linear mixed models

The influence of fixed effects and random effects on a dependant variable can be accounted for using linear mixed models. A univariate model is calculated as:

$$\boldsymbol{y} = \boldsymbol{x}b + \boldsymbol{z}u + \boldsymbol{e}$$

Where y is the vector of the dependant variables, b is the vector of the fixed effects with the design matrix X, u is the vector of the random effects with the design matrix Z and e is the vector of residuals (Mrode, 2005).

2.4.4 Random regression models

In an ideal world every animal would behave in the same way and have the same temperature as all other animals, in reality, every animal is different as there is "random" variation between animals. A lot of the variation between animals is accounted for in a linear mixed model by considering the cow as a random effect. When longitudinal data

(i.e., repeated measures over time) are recorded, any measurements which are taken of the same animal are more related than measurements taken between animals; random regression models can account for this. When random regression models are performed each animal is effectively modelled individually, and the effect of each independent factor on the dependant variable is also modelled for each animal (i.e., an individual regression co-efficient is calculated for each animal and for each independent factor).

2.4.5 Sensitivity and specificity

When a novel measurement has been shown to be dependent on infection status a cut-off point (a threshold) can be defined wherein any animal that obtained a value (e.g., a temperature) above (or below) the cut-off point is classified as infected while any animal with a value below (or above) the cut-off point is considered healthy. The proportion of animals that are correctly classified as healthy is known as the specificity and can be calculated as:

$$Specificity = \frac{TN}{TN + FP}$$

Where TN (true negative) is the number of subjects which were correctly classified as disease negative (i.e., healthy) and FP (false positive) is the number of subjects who were incorrectly classified as disease positive (i.e., infected).

The proportion of animals that are correctly classified as infected by the cut-off point is known as the sensitivity which is calculated as:

$$Sensitivity = \frac{TP}{TP + FN}$$

where TP (true positive) is the number of subjects that were correctly classified as disease positive and FN (false negative) are the number of subjects that were incorrectly classified as disease negative.

The cut-off point can be changed and a new sensitivity and specificity calculated but as sensitivity increases specificity will decrease and vice versa. The ideal trade-off between sensitivity and specificity differs from application to application. The costs (financial or otherwise) of missing an infected animal (as a result of low sensitivity) must be compared to the costs associated with screening/treating an animal which is incorrectly classified as sick (as a result of low specificity) in order for the ideal ratio between sensitivity and specificity to be decided upon (Greiner et al., 2000). In research, a sensitivity and specificity which is comparable between studies is ideal.

2.5 Gaps in Knowledge

While some research has investigated the use of IRT and load cell technology to detect disease in cattle and sheep, there are a few key areas in which this thesis aims to contribute to current knowledge and bring sensor technology one step closer to be used to detect disease on farm.. These include:

- The temperature difference between replicate thermal images of dairy cattle in an agricultural environment
- The number of replicate thermal images required for IRT to reliably quantify health status in cattle
- The use of USST as a predictor of SCC in dairy cattle in an agricultural environment
- The use of the difference between predicted and actual USST to estimate SCC
- The repeatability of thermal images captured on sheep and the number of replicates required for IRT to be used as a lameness detection tool
- The diagnostic ability of IRT to detect lameness in sheep
- The relationship between ambient temperature and sheep hoof temperature and how this relationship affects the diagnostic capabilities of IRT

- The development of a crate to measure the load a sheep places on each hoof and its ability to detect lameness

2.6 Literature Cited

Aland, A., and T. Banhazi. 2013. Livestock housing: Modern management to ensure optimal health and welfare of farm animals. Wageningen Academic Pub.

Alsaaod, M., and W. Büscher. 2012. Detection of hoof lesions using digital infrared thermography in dairy cows. J. Dairy Sci. 95: 735-742.

Ammer, K. 2008. The glamorgan protocol for recording and evaluation of thermal images of the human body. Thermol. Int. 18: 125-144.

Angell, J. W., P. J. Cripps, D. H. Grove-White, and J. S. Duncan. 2015. A practical tool for locomotion scoring in sheep: Reliability when used by veterinary surgeons and sheep farmers. Vet. Rec. 176(20): 521. doi: 10.1136/vr.102882 [doi]

Avdelidis, N., and A. Moropoulou. 2003. Emissivity considerations in building thermography. Energy Build. 35(7): 663-667.

Beaudeau, F., H. Seegers, C. Fourichon, and P. Hortet. 1998. Association between milk somatic cell counts up to 400,000 cells/ml and clinical mastitis in french holstein cows. Vet. Rec. 143(25): 685-687.

Berry, D. P., and W. J. Meaney. 2006. Interdependence and distribution of subclinical mastitis and intramammary infection among udder quarters in dairy cattle. Prev. Vet. Med. 75(1): 81-91.

Berry, R., A. Kennedy, S. Scott, B. Kyle, and A. Schaefer. 2003. Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection. Can. J. Anim. Sci. 83: 687-693.

Bohan, A., L. Shalloo, P. Creighton, E. Earle, T. Boland, and N. McHugh. 2018. Investigating the role of stocking rate and prolificacy potential on profitability of grass based sheep production systems. Livestock Science. 210: 118-124.

Bradley, A., and M. Green. 2001. A study of the incidence and significance of gram positive infections acquired during the dry period under UK field conditions. ANNUAL MEETING-NATIONAL MASTITIS COUNCIL INCORPORATED,

Bradley, A. J. 2002. Bovine mastitis: An evolving disease. The Veterinary Journal. 164(2): 116-128.

Buhr, E. D., S. H. Yoo, and J. S. Takahashi. 2010. Temperature as a universal resetting cue for mammalian circadian oscillators. Science. 330(6002): 379-385. doi: 10.1126/science.1195262

Carne, H., L. Loomis, L. Larsen, and M. Franklin. 1964. Lameness in beef cattle. Aust. Vet. J. 40(11): 382-384.

Castro-Costa, A., G. Caja, A. Salama, M. Rovai, C. Flores, and J. Aguiló. 2014. Thermographic variation of the udder of dairy ewes in early lactation and following an< i> escherichia coli endotoxin intramammary challenge in late lactation. J. Dairy Sci. 97(3): 1377-1387.

Church, J., P. Hegadoren, M. Paetkau, C. Miller, G. Regev-Shoshani, A. Schaefer, and K. Schwartzkopf-Genswein. 2014. Influence of environmental factors on infrared eye temperature measurements in cattle. Res. Vet. Sci. 96: 220-226.

Colak, A., B. Polat, Z. Okumus, M. Kaya, L. Yanmaz, and A. Hayirli. 2008. < I> short communication: early detection of mastitis using infrared thermography in dairy cows. J. Dairy Sci. 91: 4244-4248.

Conington, J., B. Hosie, G. Nieuwhof, S. Bishop, and L. Bünger. 2008. Breeding for resistance to footrot–the use of hoof lesion scoring to quantify footrot in sheep. Vet. Res. Commun. 32(8): 583-589.

CSO. 2017a. Livestock survey december. Central Statistics Office, Dublin. (https://pdf.cso.ie/www/pdf/20180309085319_Livestock_Survey_December_2017_full.pdf) Retrieved 19 September 2018.

CSO. 2017b. Milk statistics december. Central Statistics Office, Dublin. (https://pdf.cso.ie/www/pdf/20180131085949_Milk_Statistics_December_2017_full.pdf) Retrieved 19 September 2018.

Davis, J., E. Vanzant, J. Purswell, A. Green, J. Bicudo, R. Gates, L. Holloway, and W. Smith. 2003a. Methods of remote, continuous temperature detection in beef cattle. 2003 ASAE annual meeting,

Davis, M. S., T. L. Mader, S. M. Holt, and A. M. Parkhurst. 2003b. Strategies to reduce feedlot cattle heat stress: Effects on tympanic temperature. J. Anim. Sci. 81(3): 649-661.

Department of Agriculture, Food and the Marine. 2018. Annual review and outlook.(https://www.agriculture.gov.ie/media/migration/publications/2018/AnnualReviewandOutlook2018310818.pdf) Retrieved 19 September 2018.

Dillon, P., S. Crosse, G. Stakelum, and F. Flynn. 1995. The effect of calving date and stocking rate on the performance of spring- calving dairy cows. Grass Forage Sci. 50(3): 286-299.

Dohoo, I. R., R. A. Curtis, and G. G. Finley. 1985. A survey of sheep diseases in canada. Can. J. Comp. Med. 49(3): 239-247.

Eddy, A., L. Van Hoogmoed, and J. Snyder. 2001. The role of thermography in the management of equine lameness. The Veterinary Journal. 162(3): 172-181.

Fernández-Cuevas, I., J. C. Marins, P. M. Gómez Carmona, M. Á García-Concepción, J. Arnáiz Lastras, and M. Sillero Quintana. 2012. Reliability and reproducibility of skin temperature of overweight subjects by an infrared thermography software designed for human beings. Thermology International. 22: 130-137.

Finneran, E., P. Crosson, P. O'kiely, L. Shalloo, D. Forristal, and M. Wallace. 2010. Simulation modelling of the cost of producing and utilising feeds for ruminants on irish farms. Journal of Farm Management. 14(2): 95-116.

Geary, U., N. Lopez-Villalobos, B. O'Brien, D. Garrick, and L. Shalloo. 2013. Examining the impact of mastitis on the profitability of the irish dairy industry. Irish Journal of Agricultural and Food Research. : 135-149.

George, W. D., R. W. Godfrey, R. C. Ketring, M. C. Vinson, and S. T. Willard. 2014. Relationship among eye and muzzle temperatures measured using digital infrared thermal imaging and vaginal and rectal temperatures in hair sheep and cattle. J. Anim. Sci. 92(11): 4949-4955. doi: 10.2527/jas.2014-8087 [doi]

Gloster, J., K. Ebert, S. Gubbins, J. Bashiruddin, and D. J. Paton. 2011. Normal variation in thermal radiated temperature in cattle: Implications for foot-and-mouth disease detection. BMC Vet. Res. 7: 73-6148-7-73. doi: 10.1186/1746-6148-7-73 [doi]

Godden, S., E. Royster, J. Timmerman, P. Rapnicki, and H. Green. 2017. Evaluation of an automated milk leukocyte differential test and the california mastitis test for detecting intramammary infection in early-and late-lactation quarters and cows. J. Dairy Sci.

Greiner, M., D. Pfeiffer, and R. Smith. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev. Vet. Med. 45(1-2): 23-41.

Grubman, M. J., and B. Baxt. 2004. Foot-and-mouth disease. Clin. Microbiol. Rev. 17(2): 465-493.

Hanrahan, L., N. McHugh, T. Hennessy, B. Moran, R. Kearney, M. Wallace, and L. Shalloo. 2018. Factors associated with profitability in pasture-based systems of milk production. J. Dairy Sci. 101(6): 5474-5485.

Harris-Bridge, G., L. Young, I. Handel, M. Farish, C. Mason, M. Mitchell, and M. Haskell. 2018. The use of infrared thermography for detecting digital dermatitis in dairy cattle: What is the best measure of temperature and foot location to use? The Veterinary Journal. 237: 26-33.

Healy, A., M. Monaghan, H. Basset, H. Gunn, B. Markey, and J. Collins. 1993. Morbidity and mortality in a large Irish feedlot; microbiological and serological findings in cattle with acute respiratory disease. Br. Vet. J. 149(6): 549-560.

Hood, D. M., I. P. Wagner, D. D. Taylor, G. W. Brumbaugh, and M. K. Chaffin. 2001. Voluntary limb-load distribution in horses with acute and chronic laminitis. Am. J. Vet. Res. 62(9): 1393-1398.

Hovinen, M., J. Siivonen, S. Taponen, L. Hänninen, M. Pastell, A. Aisla, and S. Pyörälä. 2008. Detection of clinical mastitis with the help of a thermal camera. J. Dairy Sci. 91: 4592-4598.

Huxley, J. 2013. Impact of lameness and claw lesions in cows on health and production. Livestock Science. 156(1-3): 64-70.

Johnson, S. R., S. Rao, S. B. Hussey, P. S. Morley, and J. L. Traub-Dargatz. 2011. Thermographic eye temperature as an index to body temperature in ponies. Journal of Equine Veterinary Science. 31: 63-66.

Kaler, J., G. Wassink, and L. Green. 2009a. The inter-and intra-observer reliability of a locomotion scoring scale for sheep. The Veterinary Journal. 180(2): 189-194.

Kaler, J., G. J. Wassink, and L. E. Green. 2009b. The inter-and intra-observer reliability of a locomotion scoring scale for sheep. The Veterinary Journal. 180(2): 189-194.

Kunc, P., I. Knížková, M. Koubková, and D. O. Flusser Jan. 2000. Machine milking and its influence on temperature states of udder. abstract. Czech Journal of Animal Science. 45: 1.

Mader, T., M. Davis, and W. Kreikemeier. 2005. Case study: Tympanic temperature and behavior associated with moving feedlot cattle. Prof. Anim. Sci. 21: 339-344.

Maertens, W., J. Vangeyte, J. Baert, A. Jantuan, K. C. Mertens, S. De Campeneere, A. Pluk, G. Opsomer, S. Van Weyenberg, and A. Van Nuffel. 2011. Development of a

real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: The GAITWISE system. Biosystems Engineering. 110(1): 29-39.

Mahendran, S., R. Booth, M. Burge, and N. Bell. 2017. Randomised positive control trial of NSAID and antimicrobial treatment for calf fever. Vet. Rec.

Mapstone, R. 1968. Measurement of corneal temperature. Exp. Eye Res. 7(2): 237-242.

Martins, R. F. S., T. do Prado Paim, C. de Abreu Cardoso, B. Stéfano Lima Dallago, C. B. de Melo, H. Louvandini, and C. McManus. 2013. Mastitis detection in sheep by infrared thermography. Res. Vet. Sci. 94: 722-724.

Medrano-Galarza, C., J. Gibbons, S. Wagner, A. De Passillé, and J. Rushen. 2012. Behavioral changes in dairy cows with mastitis. J. Dairy Sci. 95(12): 6994-7002.

Montanholi, Y., K. Swanson, F. Schenkel, B. McBride, T. Caldwell, and S. Miller. 2009. On the determination of residual feed intake and associations of infrared thermography with efficiency and ultrasound traits in beef bulls. Livestock Science. 125(1): 22-30.

Montanholi, Y. R., M. Lim, A. Macdonald, B. A. Smith, C. Goldhawk, K. Schwartzkopf-Genswein, and S. P. Miller. 2015. Technological, environmental and biological factors: Referent variance values for infrared imaging of the bovine. Journal of Animal Science and Biotechnology. 6(1): 27.

Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. 2nd Edition ed. Oxfordshire,UK., CABI Publishing.

Neveux, S., D. Weary, J. Rushen, M. Von Keyserlingk, and A. de Passillé. 2006. Hoof discomfort changes how dairy cattle distribute their body weight. J. Dairy Sci. 89(7): 2503-2509.

Nikkhah, A., J. Plaizier, M. Einarson, R. Berry, S. Scott, and A. Kennedy. 2005. Short communication: infrared thermography and visual examination of hooves of dairy cows in two stages of lactation. J. Dairy Sci. 88: 2749-2753.

Pastell, M., M. Hautala, V. Poikalainen, J. Praks, I. Veermäe, M. Kujala, and J. Ahokas. 2008a. Automatic observation of cow leg health using load sensors. Comput. Electron. Agric. 62(1): 48-53.

Pastell, M., M. Kujala, A. Aisla, M. Hautala, V. Poikalainen, J. Praks, I. Veermäe, and J. Ahokas. 2008b. Detecting cow's lameness using force sensors. Comput. Electron. Agric. 64(1): 34-38.

Pezeshki, A., P. Stordeur, H. Wallemacq, F. Schynts, M. Stevens, P. Boutet, L. J. Peelman, B. De Spiegeleer, L. Duchateau, and F. Bureau. 2011. Variation of inflammatory dynamics and mediators in primiparous cows after intramammary challenge with escherichia coli. Vet. Res. 42: 15.

Pluym, L. M., D. Maes, J. Vangeyte, K. Mertens, J. Baert, S. Van Weyenberg, S. Millet, and A. Van Nuffel. 2013. Development of a system for automatic measurements of force and visual stance variables for objective lameness detection in sows: SowSIS. Biosystems Engineering. 116(1): 64-74.

Polat, B., A. Colak, M. Cengiz, L. Yanmaz, H. Oral, A. Bastan, S. Kaya, and A. Hayirli. 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. J. Dairy Sci. 93: 3525-3532.

Rainwater-Lovett, K., J. M. Pacheco, C. Packer, and L. L. Rodriguez. 2009. Detection of foot-and-mouth disease virus infected cattle using infrared thermography. Vet. J. 180: 317-324.

Rajkondawar, P., M. Liu, R. Dyer, N. Neerchal, U. Tasch, A. Lefcourt, B. Erez, and M. Varner. 2006. Comparison of models to identify lame cows based on gait and lesion scores, and limb movement variables. J. Dairy Sci. 89(11): 4267-4275.

Reuter, R. R., J. A. Carroll, L. E. Hulbert, J. W. Dailey, and M. L. Galyean. 2010. Technical note: Development of a self-contained, indwelling rectal temperature probe for cattle research. J. Anim. Sci. 88(10): 3291-3295. doi: 10.2527/jas.2010-3093 [doi]

Schaefer, A., N. Cook, S. Tessaro, D. Deregt, G. Desroches, P. Dubeski, A. Tong, and D. Godson. 2004. Early detection and prediction of infection using infrared thermography. Can. J. Anim. Sci. 84: 73-80.

Schaefer, A., N. Cook, C. Bench, J. Chabot, J. Colyn, T. Liu, E. Okine, M. Stewart, and J. Webster. 2012. The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. Res. Vet. Sci. 93: 928-935.

Shalloo, L., A. Cromie, and N. McHugh. 2014. Effect of fertility on the economics of pasture-based dairy systems. Animal. 8(s1): 222-231.

Skjøth, F., V. M. Thorup, O. F. do Nascimento, K. L. Ingvartsen, M. D. Rasmussen, and M. Voigt. 2013. Computerized identification and classification of stance phases as made by front or hind feet of walking cows based on 3-dimensional ground reaction forces. Comput. Electron. Agric. 90: 7-13.

Soerensen, D. D., S. Clausen, J. B. Mercer, and L. J. Pedersen. 2014. Determining the emissivity of pig skin for accurate infrared thermography. Comput. Electron. Agric. 109: 52-58.

Stokes, J., K. Leach, D. Main, and H. Whay. 2012. An investigation into the use of infrared thermography (IRT) as a rapid diagnostic tool for foot lesions in dairy cattle. Vet. J. 193: 674-678.

Talukder, S., G. Gabai, and P. Celi. 2015. The use of digital infrared thermography and measurement of oxidative stress biomarkers as tools to diagnose foot lesions in sheep. Small Ruminant Res. 127: 80-85.

Teagasc. 2017. Teagasc national farm survey 2017: Dairy enterprise.(https://www.teagasc.ie/media/website/publications/2018/NFS_2017_Dairy_Enterprise_Factsheet.pdf) Retrieved 19 September 2018.

Van De Gucht, T., W. Saeys, S. Van Weyenberg, L. Lauwers, K. Mertens, L. Vandaele, J. Vangeyte, and A. Van Nuffel. 2017a. Automatic cow lameness detection with a pressure mat: Effects of mat length and sensor resolution. Comput. Electron. Agric. 134: 172-180.

Van De Gucht, T., W. Saeys, S. Van Weyenberg, L. Lauwers, K. Mertens, L. Vandaele, J. Vangeyte, and A. Van Nuffel. 2017b. Automatically measured variables related to tenderness of hoof placement and weight distribution are valuable indicators for lameness in dairy cows. Appl. Anim. Behav. Sci. 189: 13-22.

Viguier, C., S. Arora, N. Gilmartin, K. Welbeck, and R. O'Kennedy. 2009. Mastitis detection: Current trends and future perspectives. Trends Biotechnol. 27(8): 486-493.

Whiteside, W. 1939. Observations on a new test for the presence of mastitis in milk. Canadian Public Health Journal. 30: 44.

Winter, A. 2008. Lameness in sheep. Small Ruminant Research. 76(1): 149-153.

Winter, A. 2004. Lameness in sheep 1. diagnosis. In Pract. 26(2): 58-63.

3 Temporal, spatial, inter- and intra- cow repeatability of thermal imaging

D.T. Byrne[1 2], D.P. Berry[1], H. Esmonde[2], and N. McHugh[1]

[1]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland, P61 P302

[2] Dublin City University, Glasnevin, Dublin 9, Co. Dublin, Ireland, D09 Y5N0

This is a pre-copyedited, author-produced version of an article accepted for publication in the journal of animal science following peer review. The version of record Byrne DT, Berry DP, Esmonde H, McHugh N. Temporal, spatial, inter-, and intra-cow repeatability of thermal imaging. Journal of animal science. 2017 Feb 1;95(2):970-9. is available online at: https://academic.oup.com/jas/article/95/2/970/4702112 (doi: 10.2527/jas.2016.1005)

3.1 Abstract

The objective of the present study was to quantify the within- and between-cow, operator and day variances of various descriptive temperature parameters from different anatomical areas captured using thermal images on Holstein-Friesian cows. Three experiments were undertaken. In Exp. 1, 30 images were captured by a single operator of each of the eye, hoof and udder from each of 45 cows. In Exp. 2, three different operators captured eye and hoof images from 12 cows and in Exp. 3, eye and hoof images were captured by a single operator on eight cows over a five day period. Maximum, minimum and average descriptive temperature parameters were manually extracted from all thermal images within the study. The repeatability of thermal imaging and the number of replicates required to obtain a certain level of precision was evaluated. Precision was defined as the 95% CI range within which the (average of the) measured temperature(s) was expected to lie relative to the gold standard; the gold standard temperature of an entity in this study was the average of 30 temperature measurements. The partitioning of the variance into error, cow, operator, and day variances was undertaken using mixed models. Results show that most repeatable anatomical area was the hoof, with the total proportion of variation attributed to the cow ranging from 91.37% to 99.28%. The descriptive temperature parameter with the lowest error variance was the maximum temperature for eye (0.11°C^2) and udder (0.03°C^2) images, whereas, the average temperature was the most precise descriptive temperature parameter for hoof (0.08°C^2) images. Additionally, no significant between-day variance was detected for maximum hoof temperatures. Results from the present study indicate that when the most precise descriptive temperature parameter is used, measurements made using infrared thermography can achieve a high level of precision in an agricultural environment if at least three replicate images of the eye, udder, or hooves of cows are captured and averaged. Additionally, when multiple operators capture thermal images in an agricultural environment, a standard operating procedure

should be put in place to minimize the variance between operators.

Key words: dairy cows, infrared thermography, repeatability

3.2 Introduction

Infrared thermography (IRT) is a quick, non-invasive procedure used to estimate the surface temperature of an object based on the radiating energy (Speakman and Ward, 1998). In research, IRT has been used in the evaluation of infrastructures (Grinzato et al., 1998; Balaras and Argiriou, 2002) and to detect breast cancer (Acharya et al., 2012; Milosevic et al., 2014). In agricultural research, IRT has also been evaluated for the detection of animal diseases such as mastitis (Polat et al., 2010), bovine respiratory diseases (Schaefer et al., 2004; Schaefer et al., 2012) and foot and mouth disease (Rainwater-Lovett et al., 2009). High repeatability (Fernández-Cuevas et al., 2012) and reproducibility (Ammer, 2008) of thermal images have been documented from medical studies undertaken under controlled conditions with calm patients. However, the same types of conditions are not attainable under practical agricultural conditions. For IRT to accurately detect disease, a precise temperature measurement is required because the difference between sick and healthy animals can be small (Colak et al., 2008). However, few studies have quantified the repeatability of thermal imaging in an agricultural environment. The objective of the present study, therefore, was to quantify the repeatability and reproducibility of thermal imaging in dairy cows under the day-to-day environmental conditions encountered on farm. The repeatability of thermal imaging and the number of replicates required to obtain a certain level of precision was evaluated. Precision was defined as the range within which the measured temperature was expected to lie relative to the gold standard; the gold standard temperature in this present study was the average of 30 temperature measurements. The impact of within- and between-operator

variability on the repeatability of the measurement system was also assessed, as well as the daily variability in individual animal temperature over a time trajectory.

3.3 Materials and Methods

Data were generated from a series of experiments undertaken on multiparous Holstein-Friesian cows at the Kilworth Research Farm, Teagasc Animal and Grassland Research Centre (Fermoy, Co. Cork, Ireland; 52.16832 latitude, −8.24313 longitude). In all instances, thermal images were captured using a FLIR T430sc thermal camera (FLIR Systems Inc., Stockholm, Sweden). The camera was set to auto-focus mode. The spectral range of the camera ranged between 7.5 and 13 µm. The camera resolution was 320 × 240 pixels, the thermal sensitivity was <0.03°C, and the accuracy (defined as the difference between the measured and the actual temperature) was ± 2°C.

Three experiments were undertaken to investigate the repeatability of cow temperature between October, 2015, and January, 2016. Across all experiments, thermal images of the udder were captured on the cows before milking in the milking parlour; all other images were taken in a covered shed with no direct sunlight. Milking was done twice daily; morning milking occurred between 07:00 and 09:00, while evening milking took place between 15:00 and 17:00. All cows were allocated a similar pasture herbage allowance and were subject to the same ambient temperature, sunlight exposure and walking distance to the parlour. Eye and hoof images were captured on non-lactating dairy cows. Daily ambient temperature and humidity data were obtained from a weather station located on the Teagasc research farm.

3.3.1 Exp. 1

The objective of this experiment was to quantify the repeatability of udder, eye, and hoof temperature measured by thermal imaging and to determine the number of images

required to gain a certain precision; precision was defined as the range within which the measured temperature was expected to lie relative to the average of 30 temperature measurements. A total of 15 cows were measured prior to evening milking in a 20 stand herringbone milking unit; a total of 30 udder images were taken per cow by a single operator. All animals were allowed to rest in the milking parlour for 30 minutes prior to imaging. To capture the udder images, the operator stood in the pit of the parlour directly behind the animals and captured images of the ventral faces of each udder at a distance of 0.8 m. The 30 replicate images were taken consecutively from each animal from the same approximate angle and distance. For each image, the udder was centred but the angle may have varied slightly between and within animals in order to view the ventral faces of all four quarters. Each image took approximately ten seconds to capture and care was taken to ensure all images were in focus.

Eye and hoof images were captured across 2 days after animals were given at least a 30 minute acclimatization period prior to imaging; no direct sunlight was present. A set of 30 replicates of the left eye, followed by 30 replicates of the right eye were taken on each of 15 cows by a single operator, both at a distance of 0.9 m. The operator stood just off perpendicular to the median plane to obtain an image of the eye. Any image that contained a closed eye was retaken. On the following day, a new set of 15 cows were used where 60 hoof images were captured per cow; a single operator captured 30 replicates of the front hooves followed by 30 replicates of the back hooves. To obtain a palmar view of both front hooves, the camera was placed to the left hand side of the animal pointing towards the front hooves. To obtain a plantar view of both back hooves, the camera was placed behind the animal at an angle that was just off parallel to the median plane. All hoof images were captured at a distance of 1.2 m. To mimic the capturing of real farm data, the operator focused the camera on a different object between each replicate of the udder, eye and hoof.

3.3.2 Exp. 2

The objective of Exp. 2 was to quantify the within- and between-operator variability in thermal images. Thermal images of six cows were captured after morning milking and, from a different cohort of six cows, after evening milking on the same day. Images were captured in the same location as in Exp. 1.

Images of each cow were captured by three camera operators; each operator captured each image of the required anatomical areas in the same order (i.e., image of the left eye followed by front hooves and back hooves). After taking an image of each of the three views, the camera was passed to the second operator who, upon completion of the same views, passed the camera to the third operator who repeated the procedure. The same procedure was repeated in the same order until each operator had completed four replicates of each image per animal. Each operator captured a total of twelve images per animal including the cow's eye, front hooves and back hooves. Operators varied in experience in thermal imaging, from an operator with training in thermal imaging (experienced operator), to an operator with limited experience (limited experience operator), and finally an operator with no previous experience of thermal imaging (novice operator). A standard operating procedure (SOP) was provided to each operator instructing them on how to capture each image. These methods were very similar to the procedure undertaken in Exp. 1. Each operator did not observe any other operator capture images.

3.3.3 Exp. 3

The objective of this experiment was to quantify the day-to-day repeatability of IRT on live animals. Thermal images were captured on the same eight cows over five consecutive days after evening milking. Images were taken of the left eye, front hooves and back hooves; this order of thermal imaging was repeated twice by the same operator in the same location as Exp. 1 and 2.

3.3.4 Image Analysis

Image analysis and temperature extraction was undertaken using the Thermovision LabVIEW toolkit 3.3 (FLIR Systems Inc., Stockholm, Sweden.). All image parameters (i.e., emissivity, ambient temperature, humidity, object distance, and reflected temperature) were adjusted in each image before analysis. Emissivity in all images was set to 0.98; ambient temperature and humidity data varied between images and the respective values were taken from the weather station located on the research farm. Object distance and reflected temperature changed between anatomical views.

The udder images were not cropped, the maximum, minimum and average temperature values were taken from the entire image, as only the udder was contained in the entire image. A supervised maximum value was also extracted from the udder images by manual supervision, whereby the user manually cropped each image to ensure the maximum temperature value was extracted from the same area of the udder in each replicate.

The eye and hoof images were manually cropped to ensure the same predefined areas in each image were used to calculate the descriptive temperature parameters (i.e., maximum, minimum and average) and to remove any unwanted background information. For eye images, a simple rectangle was drawn around the eye (Figure 3.1). The borders of the rectangle were defined by the outer edges of the cornea. A parallelogram shape was used to encompass the posterior face of the front and back hooves. The base of the parallelogram was drawn below the coronary band and the top of the parallelogram ceased above the dew claws (Figure 3.2). Sides were defined by the outer edges of the hoof. Manual segmentation ensured that only the hoof and no background information were contained inside the parallelogram.

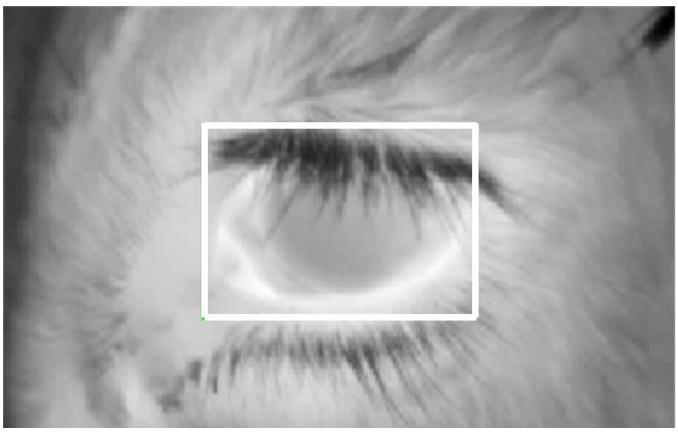


Figure 3.1 The white rectangle encompasses the region of the eye used to calculate the supervised maximum, minimum and average temperatures.

Figure 3.2 The white parallelogram encompasses the region of the hooves used to calculate the supervised maximum, minimum and average temperatures.

3.3.5 Statistical Analysis

3.3.5.1 Exp. 1.

To quantify the animal effect on temperature, between-cow and error variances were calculated for each descriptive temperature parameter (i.e., maximum, minimum, and average) and each anatomical area using a mixed model in ASReml (Gilmour et al., 2009)

with cow included as a random effect. A log likelihood ratio test was performed on nested models to determine whether a significant cow variance existed. The proportion of total variance explained by cow (H_{Cow}) was calculated as:

$$H_{Cow} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_e^2}$$

where σ_c^2 was the between-cow variance and σ_e^2 was the error variance. The CV was calculated for each anatomical area and each descriptive temperature parameter as:

$$CV = \frac{\sigma_c}{\mu}$$

where σ_C was the between-cow standard deviation and μ was the mean of the descriptive temperature parameter and anatomical area under investigation. The number of images required to gain a certain precision (P_n) with a 95% CI was calculated as:

$$P_n = 1.96 \times \sqrt{\frac{\sigma_e^2}{n_{\in(1,30)}}}$$

where n was the image count (sample size) which varied from 1 to 30. To investigate the stability of a measurement over time, temperature measurements made at each replicate number (i.e., 1 to 30 for every cow) were averaged across cows for each anatomical area and descriptive temperature parameter. The correlation between the first replicate measurement and all other replicate measurements was then calculated using PROC CORR of SAS (SAS Institute, 2010).

3.3.5.2 Exp. 2.

The between-operator and error variance in maximum and average temperature measurements for each anatomical area was estimated using a mixed model in ASReml (Gilmour et al., 2009) with both cow and operator included as random effects. A log likelihood ratio test was performed on nested models to determine whether the addition of operator as a random component improved the fit to the data. The proportion of total variation explained by operator ($H_{Operator}$) was calculated as:

$$H_{Operator} = \frac{\sigma_O^2}{\sigma_c^2 + \sigma_O^2 + \sigma_e^2}$$

where σ_O^2 was the between-operator variance, σ_c^2 was the between-cow variance and σ_e^2 was the error variance. The correlations between the temperatures of each anatomical area for each descriptive temperature parameter were calculated using PROC CORR of SAS (SAS Institute, 2010).

3.3.5.3 Exp. 3.

The partitioning of the total variance of maximum and average temperature into between-day and error variances was undertaken for each anatomical area using a mixed model in ASReml (Gilmour et al., 2009) with both cow and day included as random effects. A log likelihood ratio test was performed on nested models to determine whether the inclusion of day as a random component improved the fit of the data. The proportion of total variation explained by day (H_{Day}) was calculated from:

$$H_{Day} = \frac{\sigma_D^2}{\sigma_c^2 + \sigma_D^2 + \sigma_e^2}$$

where σ_D^2 was the between-day variance, σ_c^2 was the between-cow variance and σ_e^2 was the error variance.

3.4 Results

3.4.1 Exp. 1

Across all animals, the mean ± SD of the supervised maximum and average temperature values for the three anatomical areas was 38.22 ± 0.36°C and 32.57 ± 1.00°C for udder, 36.52 ± 0.51°C and 32.26 ± 0.84°C for eye, and 19.13 ± 4.41°C and 14.20 ± 2.88°C for hoof, respectively. The between-cow and error variance for the average temperature of the udder and both eye measurements were greater than the corresponding variances for the maximum temperature of the udder and eyes. The opposite was true for all hoof images. Greater between-cow and error variances were evident for minimum

temperature of both the udder and eye images compared with all other descriptive temperature parameters (Table 3.1). Error variances for maximum and average temperatures from the left eye were similar to the right eye. The error variance of the average temperatures was almost identical across all four hooves. The error variances of the maximum temperature for all four hooves ranged from 0.17°C^2 (right front hoof) to 0.43°C^2 (left back hoof) (Table 3.1).

Table 3.1 Mean temperature, cow variance and error variances (SE in parentheses), CV and the proportion of total variation explained by cow (H_{cow}) for each anatomical area and descriptive temperature parameter.

Body part	Variable	Mean, °C	Cow variance, °C^2	Error variance, °C^2	CV	H_{cow}, %
Udder	Maximum	38.22	0.09 (0.03)	0.05 (0.00)	0.01	64.19
	Maximum (W)[2]	38.57	0.04 (0.02)	0.03 (0.00)	0.01	56.33
	Minimum (W)[2]	19.09	3.99 (1.59)	2.00 (0.14)	0.10	66.62
	Average (W)[2]	32.57	0.79 (0.31)	0.27 (0.02)	0.03	74.55
Right eye	Maximum	36.61	0.11 (0.04)	0.11 (0.01)	0.01	49.31
	Minimum	23.07	3.21 (1.25)	2.81 (0.19)	0.08	53.32
	Average	32.52	0.3 (0.12)	0.22 (0.02)	0.02	57.25
Left eye	Maximum	36.43	0.18 (0.07)	0.11 (0.01)	0.01	62.33
	Minimum	21.95	6.54 (2.52)	3.40 (0.23)	0.12	65.83
	Average	31.99	0.57 (0.22)	0.24 (0.02)	0.02	70.20
RFH[1]	Maximum	19.90	21.22 (8.00)	0.17 (0.01)	0.23	99.17
	Minimum	10.48	3.28 (1.24)	0.14 (0.01)	0.17	91.37
	Average	15.03	11.69 (4.41)	0.08 (0.01)	0.23	99.14
RBH[1]	Maximum	18.53	15.43 (5.84)	0.33 (0.02)	0.21	99.20
	Minimum	9.76	1.59 (0.60)	0.06 (0.00)	0.13	96.00
	Average	13.80	5.42 (2.05)	0.08 (0.01)	0.17	99.28
LFH[1]	Maximum	19.05	16.31 (6.15)	0.21 (0.01)	0.21	97.30
	Minimum	9.30	1.88 (0.71)	0.09 (0.01)	0.15	96.81
	Average	13.81	7.95 (3.00)	0.07 (0.01)	0.20	97.83
LBH[1]	Maximum	19.06	27.98 (10.56)	0.43 (0.03)	0.28	97.91
	Minimum	9.74	1.95 (0.74)	0.08 (0.01)	0.14	96.32
	Average	14.16	9.01 (3.40)	0.09 (0.01)	0.21	98.56

[1] LBH = left back hoof, LFH = left front hoof, RBH= right back hoof, RFH= right front hoof.

[2] (W) = the entire image was used for analysis.

A greater proportion of the total variation was attributed to the cow for hoof images (91.37 to 99.28%) in comparison with either the eye (49.31 to 70.20%) or udder images (56.33 to 74.55%;Table 3.1). Similarly, a greater CV between cows was calculated for hoof images (0.13 to 0.28) compared with eye (0.01 to 0.12) or udder images (0.01 to 0.10). For hoof images, the greatest CV was associated with the maximum extracted temperature, whereas greater CV was associated with the minimum extracted temperature for both the eye and udder images.

The number of images required to achieve a defined level of precision, defined as the 95% CI range, where the (average of the) temperature taken different from the average 30 measurements, is shown in Table 3.2. For the hoof images, fewer replicate images were required to achieve a greater level of precision, when the average temperature was used compared with the maximum hoof temperature. When the average hoof temperature was extracted from five images, the average temperature of the five images was expected to be within ± 0.25°C of the average of 30 images, 95% of the time. Whereas when the maximum hoof temperature was extracted from the same five images, an average precision of only ± 0.46°C was achieved. For eye and udder images the maximum temperature yielded the most precise results.

Table 3.2 Standard error of one, five and thirty image replicates (°C) for each anatomical area and descriptive temperature parameter in Exp. 1.

Body part	Variable	One image, °C	Five image replicates, °C	Thirty image replicates, °C
Udder	Maximum	0.44	0.20	0.08
	Maximum (W)[2]	0.34	0.15	0.06
	Minimum (W)[2]	2.77	1.24	0.51
	Average (W)[2]	1.02	0.46	0.19
Right eye	Maximum	0.65	0.29	0.12
	Minimum	3.29	1.47	0.60
	Average	0.92	0.41	0.17
Left eye	Maximum	0.65	0.29	0.12
	Minimum	3.61	1.62	0.66
	Average	0.96	0.43	0.18
RFH[1]	Maximum	0.81	0.36	0.15
	Minimum	0.73	0.33	0.13
	Average	0.55	0.25	0.10
RBH[1]	Maximum	1.13	0.50	0.21
	Minimum	0.48	0.21	0.09
	Average	0.55	0.25	0.10
LFH[1]	Maximum	0.90	0.40	0.16
	Minimum	0.59	0.26	0.11
	Average	0.52	0.23	0.09
LBH[1]	Maximum	1.29	0.57	0.23
	Minimum	0.55	0.25	0.10
	Average	0.59	0.26	0.11

[1] LBH = left back hoof, LFH = left front hoof, RBH = right back hoof, RFH = right front hoof.

[2] (W) = the entire image was used for analysis.

Across all anatomical areas, the correlation between the first replicate temperature measurement and all other replicate temperature measurements tended to weaken as the replicate number, and therefore time interval, increased (Figure 3.3). The correlation between the first replicate temperature measurement and all other replicate temperature measurements for the maximum temperature of the left front hoof images ranged from 0.97 (replicate 1 and replicate 18) to 0.99 (replicate 1 and replicate 2). Strong to moderate

correlations existed between the first replicate measurement and all other replicate measurements for maximum temperatures of the left eye and ranged from 0.37 (replicate 1 and replicate 12) to 0.86 (replicate 1 and replicate 2) (Figure 3.3). The correlation between the maximum temperature of the first udder replicate measurement and all other udder replicate measurements ranged from 0.09 (replicate 1 and replicate 24) to 0.94 (replicate 1 and replicate 3). Correlations among average temperature values between the first replicate measurement and all other replicate measurements were similar to maximum temperature values of corresponding anatomical regions.

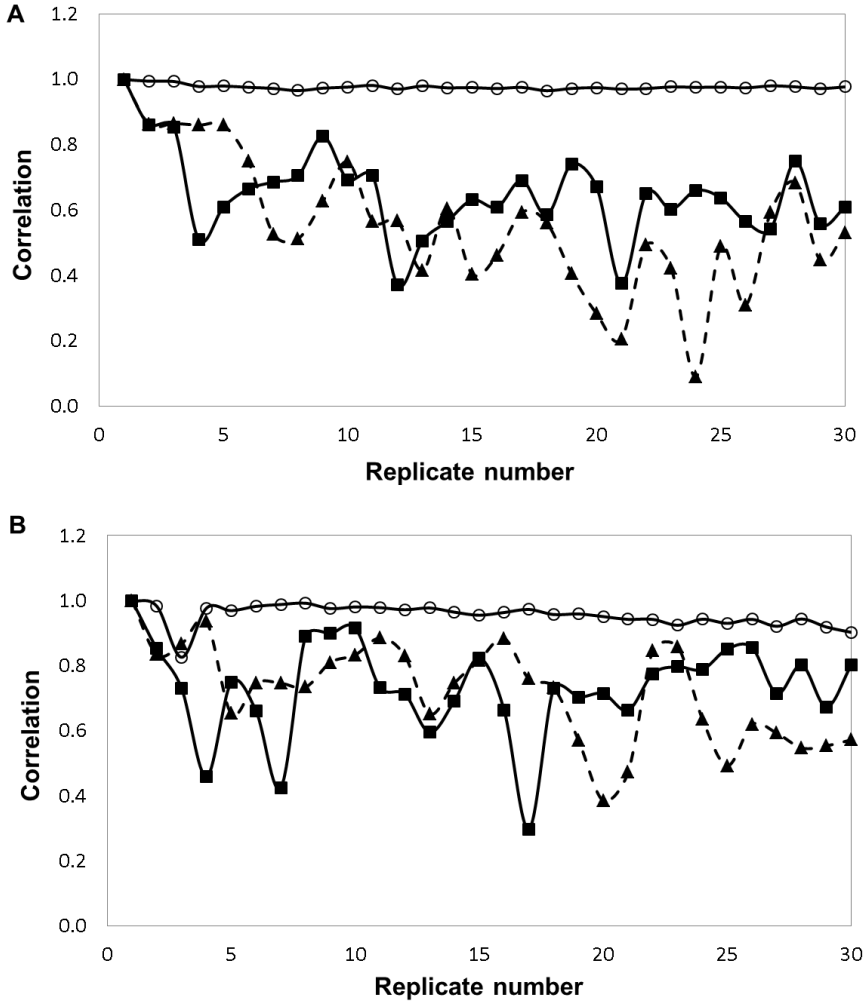


Figure 3.3 Correlations estimates between the first image and all other replicate numbers of the left eye (■), left front hoof (○) and udder (▲) for the maximum (Panel A) and the average (Panel B) extracted temperature.

3.4.2 Exp. 2

The between-cow and operator variance for the morning did not differ from the respective values for the evening and, therefore, the data were combined. The mean ± SD across all animals for each operator for maximum eye temperature was 36.31 ± 0.62°C (experienced operator), 36.52 ± 0.66°C (limited experience operator), and 37.01 ± 0.69°C (novice operator). The mean ± SD for each operator for average left front hoof temperature was 18.57 ± 3.95°C (experienced operator), 18.66 ± 3.93°C (limited experience operator), and 18.36 ± 3.99°C (novice operator). The between-operator variance was numerically greater when the maximum temperature value was extracted from each image compared to when the average temperature was extracted (Table 3.3). Error variances (i.e., within cow and operator variation) were also greater when the maximum temperature was extracted (Table 3.3). Results from the log likelihood ratio test indicated that the inclusion of operator as a random component improved the fit of the data for maximum eye temperature only. In all scenarios, the CV was $\leq$ 0.03%. The proportion of total the variation explained by operator for the maximum eye temperature was 0.22%, but in all other scenarios it was $\leq$ 0.02%. Weak correlations existed between hoof and eye temperature measurements (0.03 to 0.24), but, strong correlations existed among all four hoof temperatures.

Table 3.3 Operator and error variance (SE in parentheses), CV and the proportion of total variation explained by operator ($H_{Operator}$) for each anatomical area and descriptive temperature parameter in Exp. 2.

Body part	Variable	Operator variance, °C^2	CV	$H_{Operator}$, %	Error variance, °C^2
Eye	Maximum	0.13 (0.13)	0.01	0.22	0.18 (0.02)
	Average	0.02 (0.02)	0.00	0.02	0.42 (0.05)
RFH[1]	Maximum	0.10 (0.11)	0.02	0.01	0.31 (0.04)
	Average	0.00 (0.01)	0.00	0.00	0.17 (0.02)
RBH[1]	Maximum	0.10 (0.13)	0.02	0.01	1.31 (0.16)
	Average	0.01 (0.01)	0.00	0.00	0.37 (0.05)
LFH[1]	Maximum	0.24 (0.25)	0.03	0.01	0.54 (0.07)
	Average	0.02 (0.02)	0.01	0.00	0.20 (0.03)
LBH[1]	Maximum	0.13 (0.15)	0.02	0.01	0.79 (0.10)
	Average	0.04 (0.05)	0.01	0.00	0.41 (0.05)

[1]LBH = left back hoof, LFH = left front hoof, RBH = right back hoof, RFH = right front hoof.

3.4.3 Exp. 3

The mean ± SD maximum eye temperatures for all animals across days ranged from 37.24 ± 0.56°C to 37.97 ± 0.28°C. For the hooves, using the left back hoof as an example, the mean ± SD average temperature across all animals across days ranged from 26.13 ± 2.71°C to 29.43 ± 1.27°C. Across all anatomical areas, the variability in maximum temperature among days was numerically lower than for average temperatures. The proportion of total variation explained by day was less for maximum hoof temperatures (left front hoof, 0.44 ± 0.40°C) in comparison to average hoof temperatures (1.55 ± 1.15°C) (Table 3.4). The inclusion of day as a random effect improved the fit of the model for all anatomical areas, with the exception of the left back hoof when the average temperature was included as the dependent variable; however, for the maximum temperature, day only had a significant effect on eye temperature measurements.

Table 3.4 Day Variance (SE in parentheses), CV and the proportion of total variation explained by day (H_{Day}) for each anatomical area and descriptive temperature parameter in Exp. 3.

Body part	Variable	Day variance, °C^2	CV	H_{Day}, %
Eye	Maximum	0.08(0.06)	0.01	0.33
	Average	0.19(0.14)	0.01	0.28
RFH[1]	Maximum	0.48(0.40)	0.02	0.13
	Average	1.46(1.09)	0.04	0.29
RBH[1]	Maximum	0.41(0.33)	0.02	0.11
	Average	0.63(0.48)	0.03	0.22
LFH[1]	Maximum	0.44(0.40)	0.02	0.11
	Average	1.55(1.15)	0.04	0.33
LBH[1]	Maximum	0.14(0.16)	0.01	0.03
	Average	0.33(0.31)	0.02	0.08

[1]LBH = left back hoof, LFH = left front hoof, RBH = right back hoof, RFH = right front hoof.

3.5 Discussion

The objective of the present study was to quantify the precision achievable from IRT for measuring the temperature of various anatomical regions of cows in a day-to-day farm environment. This was achieved through quantifying the within- and between-cow, operator and day variability. Results indicate that a single IRT image does not always provide the required level of precision for every scenario. Precise measurements, however, can be achieved from multiple replicates captured following a SOP and averaging the most precise descriptive temperature parameter.

3.5.1 Temperature Measurement

There are very few published studies that have investigated the precision of various IRT-based temperature measurements of cows in an agricultural environment. Previous studies across multiple anatomical areas have, however, successfully used the maximum

temperature from thermal images to detect disease in cattle (Whay et al., 2004; Schaefer et al., 2012), indicating that the maximum temperature is indicative of the increase in blood flow that occurs at the onset of infection (Jones and Plassmann, 2002; McGavin and Zachary, 2007). Other studies have compared maximum, minimum and average udder temperatures of healthy and infected udders in dairy cows and showed that the greatest difference between sick and healthy cows existed in the maximum temperature when compared with other descriptive temperature parameters (Metzner et al., 2014). No study was found to validate minimum temperature as an indicator of disease status, as the minimum temperature may not always come from the animal's skin but rather foreign material or background information. Results from the present study corroborates the findings from previous studies in that greater precision of IRT measurements was achieved using the maximum temperature for the udder and eye images (Metzner et al., 2014). For the hoof images in the present study, however, the average hoof temperature provided the most precise temperature measurement when captured in a single day, whereas, the maximum hoof temperature was more precise across days.

3.5.2 Variances Between-Cow and across Anatomical Areas

A large between-cow variance, relative to total variance, is generally desirable because this indicates that the temperature of the cow has a greater influence on the captured temperature in comparison with all other extraneous factors that were present at the time of imaging. No study with cattle has attempted to partition the temperature variance from IRT into between and within cow variances. Hoof, eye, and udder mean temperatures from the present study are within the range previously reported for hoof temperatures (Alsaaod and Büscher, 2012; Stokes et al., 2012), eye temperatures (Schaefer et al., 2012; Church et al., 2014), and maximum and average udder temperatures (Berry et al., 2003; Metzner et al., 2014). The mean minimum udder temperature was much lower in the present study compared with other studies in dairy cows (Berry et al., 2003; Metzner et

al., 2014), although the fact that the udders of cows in the present study were not washed prior to imaging may have contributed to this apparent discrepancy.

The large between-cow variance observed for hoof images in the present study resulted in greater repeatability values associated with the hoof images compared with other anatomical areas. This therefore indicates that IRT may be a very robust tool for the detection of lameness (Nikkhah et al., 2005b; Alsaaod and Büscher, 2012; Stokes et al., 2012; Alsaaod et al., 2015), even though the mean temperature differences between animals in the present study may not be attributed to illness. In order to detect hoof lesions using IRT, previous studies employed different methods. The temperature difference between two regions of the hoof was recorded as part of some studies and, if this temperature difference was above a certain threshold, the hoof was considered to be infected (Nikkhah et al., 2005b; Alsaaod and Büscher, 2012). In other studies, a single area was measured and simply defined a threshold hoof temperature, above which a hoof was considered infected (Stokes et al., 2012). In the present study, strong correlations existed between all four hoof temperatures for each descriptive temperature parameter, a new method of lameness detection could involve comparing all four hoof temperatures to each other and any outlier with an elevated temperature could be indicative of inflammation.

Results from the present study show that the proportion of total variation attributed to between-cow variance for eye and udder images was almost equal to the proportion of total variation attributed to error variance, indicating that imaging techniques for these regions could possibly be improved to obtain a more precise temperature measurement. Improvement of imaging techniques may be achieved by a greater restriction on the camera-to-object distance or by the development of alternative descriptive temperature parameters (Milosevic et al., 2014).

3.5.3 Precision of IRT

The temperature differences between a healthy and sick animal was reported to

range from 0.89°C (Colak et al., 2008) to 7.90°C (Stokes et al., 2012) which were taken by a camera with a similar accuracy but a greater thermal sensitivity in comparison with the current study. As the difference in temperature between a sick and healthy animal can be very small, temperature measurements must, therefore, be precise to facilitate a correct diagnosis. This can be difficult in an agricultural environment where imaging conditions are not ideal because factors such as ambient temperature, previous cow activity, and the milking process are known to affect IRT capture (Mader et al., 2005; Paulrud et al., 2005; Gloster et al., 2011). One proven method to increase precision is image replication whereby a set of consecutive replicates can be captured in quick succession to provide a more precise temperature value. This can be demonstrated theoretically because, as the number of replicate records increases, the standard error reduces proportionally; assuming the variance of the data does not change. Additionally, if the precision of IRT measurements is increased, diseases could possibly be detected at earlier stages. While a larger number of replicate images are beneficial, manually capturing a large number of images on farm may be impractical. In addition, the animal's temperature can vary over time, particularly eye and udder temperature (Figure 3.3), which may result in an inaccurate temperature reading due to factors such as the animal's natural circadian rhythm (Berry et al., 2003), sunlight and strong winds (Church et al., 2014), or sudden biological or physiological changes in the animal (Valera et al., 2012). While IRT is a non-invasive technology, the presence of the thermographer within an animal's flight zone can cause an increase in the animal's temperature due to stress (Valera et al., 2012). Therefore, replicate images of a single animal must be taken in quick succession following an acclimatization period. In the past, many studies have captured a single image to detect disease (Polat et al., 2010; Alsaaod and Büscher, 2012); however, few published studies have used the average of multiple replicate thermal images to detect disease in cows. Results from the present study indicate that the difference between two udder or eye IRT replicates can

sometimes be as large as the temperature difference between a healthy and a sick animal (Colak et al., 2008; Polat et al., 2010; Schaefer et al., 2012). The temperature difference between two consecutive hoof replicates observed in the present study was less, however, than the temperature difference between a healthy and an infected hoof (Stokes et al., 2012). If five hoof replicates are captured as opposed to one, the precision of hoof temperature measurements can be improved from ± 1.03°C to ± 0.25°C when the maximum hoof temperature is used.

3.5.4 Operator Repeatability

In medical research, high operator repeatability of IRT has been reported (Fernández-Cuevas et al., 2012); however, a high degree of control can be implemented on the environment and subjects in the medical industry. The Glamorgan protocol has defined many of these controls in order to standardize medical IRT and, therefore, reduce the variance between images across the medical industry (Ammer, 2008). Hence, a thermal imaging SOP for an agricultural environment was created as part of the present study. The SOP used in Exp. 2 was sufficient to standardize IRT temperature measurements across operators because no significant operator variation was detected across most anatomical areas; however, the maximum eye temperature was affected by the variability among operators. Upon further inspection of the images, it was noted that the novice operator was closer to the animal (distance of ~0.5m) when capturing the eye images compared with the other operators (distance of 0.9m). Across all operators, the average eye temperature was extracted from the same images as the maximum eye temperature; whereas operator variation had a significant effect on maximum eye temperature, it did not have a significant effect on the average eye temperature. While few published agricultural studies have attempted to quantify between-operator repeatability, previous research has shown that the distance between the camera and the eye can affect maximum eye temperature measurements (Church et al., 2014). Strict control of the distance between the camera and

the animal during image capture should partially ensure that the variation of measurements is kept to a minimum. Future research in this area is required to include additional anatomical areas and examine temperature differences between animals of varying ages; this future research can be used to produce a more detailed SOP for use across the industry. Similar to the Glamorgan protocol, further protocols could implement the results of this study to include how each region should be segmented and the type of temperature measurement to be extracted from each region (Ammer, 2008).

3.5.5 Image Repeatability across Days

No agricultural studies have been found to compare descriptive temperature parameters across days, although previous studies have documented lower standard deviations in average udder temperatures captured in a single day compared with maximum udder temperatures (Metzner et al., 2014). In the present study, less variation across days existed for maximum temperatures in comparison with average temperatures. These results indicate that the average temperature should be extracted when images are captured in a single day, but the maximum temperature should be extracted when images are compared across a short number of days.

3.6 Conclusions

Results from the present study show that precise temperature measurements can be recorded using IRT in an agricultural setting. The capturing of multiple replicates over a short period of time can increase the precision of image capture. The present study indicates capturing three replicates of an anatomical region when using IRT for disease detection in cows. Additionally, the establishment of a detailed SOP can ensure that differences among operators are minimized and the repeatability is maximized. Maximum hoof temperatures taken across multiple days are readily comparable; however, over an

extended period, a model to predict normal and healthy temperature variations may be employed. Knowledge of these results can ensure IRT is used to its highest potential in an agricultural environment.

3.7 Literature Cited

Acharya, U. R., E. Y. Ng, J. Tan, and S. V. Sree. 2012. Thermography based breast cancer detection using texture features and support vector machine. J. Med. Syst. 36:1503-1510.

Alsaaod, M., and W. Büscher. 2012. Detection of hoof lesions using digital infrared thermography in dairy cows. J. Dairy Sci. 95:735-742.

Alsaaod, M., A. L. Schaefer, W. Büscher, and A. Steiner. 2015. The role of infrared thermography as a non-invasive tool for the detection of lameness in cattle. Sensors. 15:14513-14525.

Ammer, K. 2008. The glamorgan protocol for recording and evaluation of thermal images of the human body. Thermol. Int. 18:125-144.

Balaras, C. A., and A. A. Argiriou. 2002. Infrared thermography for building diagnostics. Energy Build. 34:171-183.

Berry, R., A. Kennedy, S. Scott, B. Kyle, and A. Schaefer. 2003. Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: potential for mastitis detection. Can. J. Anim. Sci. 83:687-693.

Church, J., P. Hegadoren, M. Paetkau, C. Miller, G. Regev-Shoshani, A. Schaefer, and K. Schwartzkopf-Genswein. 2014. Influence of environmental factors on infrared eye temperature measurements in cattle. Res. Vet. Sci. 96:220-226.

Colak, A., B. Polat, Z. Okumus, M. Kaya, L. Yanmaz, and A. Hayirli. 2008. Short communication: early detection of mastitis using infrared thermography in dairy cows. J. Dairy Sci. 91:4244-4248.

Fernández-Cuevas, I., J. C. Marins, P. M. Gómez Carmona, M. Á García-Concepción, J.

Arnáiz Lastras, and M. Sillero Quintana. 2012. Reliability and reproducibility of skin temperature of overweight subjects by an infrared thermography software designed for human beings. Thermology International 22:130-137.

Gilmour, A., B. Gogel, B. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0.VSN International Ltd, Hemel Hempstead, United Kingdom.

Gloster, J., K. Ebert, S. Gubbins, J. Bashiruddin, and D. J. Paton. 2011. Normal variation in thermal radiated temperature in cattle: Implications for foot-and-mouth disease detection. BMC Vet. Res. 7:73-6148-7-73. doi:10.1186/1746-6148-7-73

Grinzato, E., V. Vavilov, and T. Kauppinen. 1998. Quantitative infrared thermography in buildings. Energy Build. 29:1-9.

Jones, B. F., and P. Plassmann. 2002. Digital infrared thermal imaging of human skin. IEEE Engineering in Medicine and Biology Magazine 21:41-48.

Mader, T., M. Davis, and W. Kreikemeier. 2005. Case study: Tympanic temperature and behavior associated with moving feedlot cattle. Prof. Anim. Sci. 21:339-344.

McGavin, M. D., and J. F. Zachary. 2007. Pathologic basis of veterinary disease. 4th ed., Mosby Elsevier, St. Louis, MO.

Metzner, M., C. Sauter-Louis, A. Seemueller, W. Petzl, and W. Klee. 2014. Infrared thermography of the udder surface of dairy cattle: Characteristics, methods, and correlation with rectal temperature. Vet. J. 199:57-62.

Milosevic, M., D. Jankovic, and A. Peulic. 2014. Thermography based breast cancer detection using texture features and minimum variance quantization. EXCLI J. 13:1204-1215.

Nikkhah, A., J. Plaizier, M. Einarson, R. Berry, S. Scott, and A. Kennedy. 2005. Short communication: infrared thermography and visual examination of hooves of dairy cows in two stages of lactation. J. Dairy Sci. 88:2749-2753.

Paulrud, C. O., S. Clausen, P. E. Andersen, and M. D. Rasmussen. 2005. Infrared

thermography and ultrasonography to indirectly monitor the influence of liner type and overmilking on teat tissue recovery. Acta Vet. Scand. 46:137-147.

Polat, B., A. Colak, M. Cengiz, L. Yanmaz, H. Oral, A. Bastan, S. Kaya, and A. Hayirli. 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. J. Dairy Sci. 93:3525-3532.

Rainwater-Lovett, K., J. M. Pacheco, C. Packer, and L. L. Rodriguez. 2009. Detection of foot-and-mouth disease virus infected cattle using infrared thermography. Vet. J. 180:317-324.

Schaefer, A., N. Cook, S. Tessaro, D. Deregt, G. Desroches, P. Dubeski, A. Tong, and D. Godson. 2004. Early detection and prediction of infection using infrared thermography. Can. J. Anim. Sci. 84:73-80.

Schaefer, A., N. Cook, C. Bench, J. Chabot, J. Colyn, T. Liu, E. Okine, M. Stewart, and J. Webster. 2012. The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. Res. Vet. Sci. 93:928-935.

Speakman, J., and S. Ward. 1998. Infrared thermography: Principles and applications. Zoology: Analysis of Complex Systems. 101:224-232.

Stokes, J., K. Leach, D. Main, and H. Whay. 2012. An investigation into the use of infrared thermography (IRT) as a rapid diagnostic tool for foot lesions in dairy cattle. Vet. J. 193:674-678.

Valera, M., E. Bartolomé, M. J. Sánchez, A. Molina, N. Cook, and A. Schaefer. 2012. Changes in eye temperature and stress assessment in horses during show jumping competitions. J. Equine Vet. Sci. 32:827-830.

Whay, H. R., M. J. Bell, and D. C. J. Main. 2004. Validation of lame limb identification through thermal imaging. 13th international symposium on lameness in ruminants, Maribor, Slovenia.

4 Investigation of the relationship between udder quarter somatic cell count and udder skin surface temperature of dairy cows measured by infrared thermography

D.T. Byrne[1 2], D.P. Berry[1], H. Esmonde[2], and N. McHugh[1]

[1]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland, P61 P302.

[2]Dublin City University, Glasnevin, Dublin 9, Co. Dublin, Ireland, D09 Y5N0.

This is a pre-copyedited, author-produced version of an article accepted for publication in the journal of animal science following peer review. The version of record Byrne DT, Berry DP, Esmonde H, McHugh N. Investigation of the relationship between udder quarter somatic cell count and udder skin surface temperature of dairy cows measured by infrared thermography. Journal of animal science. 2018 Jul 19;96(10):4458-70. is available online at: https://academic.oup.com/jas/article/96/10/4458/5056008 (doi: 10.1093/jas/sky281)

4.1 Abstract

The objective of the current study was to quantify the relationship between udder skin surface temperature (USST) and somatic cell count (SCC) in lactating dairy cows. Data were recorded on the same 14 Holstein-Friesian cows, at evening (15:00 to 16:00) milking every day over a two-month period. Surface temperature measurements of all udders were extracted from thermal images. Post imaging, milk was extracted from each quarter and analysed for SCC. Environmental and cow related factors (i.e., ambient temperature, humidity, rainfall, wind speed, distance walked to the parlour, number of days since the udder was shaved, parity, and stage of lactation) were recorded on each day of the experiment. A large array of descriptive temperature parameters (DTP) were extracted from every udder image including temperature- (e.g., maximum, average and minimum USST), pixel count- and textural-based DTPs. Several different analytical methods were tested in an attempt to relate any given DTP to SCC; this included investigating the relationship between USST and the log transform of SCC (i.e. somatic cell score; SCS). The temperature range within each udder was also compared to the natural log of the range in SCC of the respective quarters. In a separate analysis, the temperature difference between each DTP and its respective daily baseline (i.e., average of the five lowest values of that DTP across the herd) was compared to SCS. Finally, the association between environmental and cow related factors with each DTP was investigated to create prediction models for each DTP, the residuals of which were compared to SCC. Results from the present study indicate that the correlation between any DTP and SCS was weak (range of -0.16 to 0.19) and so could not be used to identify quarters with high SCC. While some alternative measures had a significant relationship with SCS, again, the correlation was too weak for practical use on its own. Maximum and average USST could be predicted with a root mean square error of 0.23°C and 0.35°C, respectively, although the residuals from the prediction model could not be used to identify animals with high SCC. This suggests that

infrared thermography alone could not be used as a real time automated tool to detect high SCC for dairy cows in a pasture based system.

Key words: dairy cows, infrared thermography, somatic cell count

4.2 Introduction

An elevation in somatic cell count (SCC) in the milk of a cow is a common method of detecting infection in the mammary gland (Harmon, 1994). Mammary gland infections and high SCC can reduce farm profitability due to loss in milk sales, greater veterinary treatment, and greater involuntary culling (Geary et al., 2011). Previous studies have shown the prevalence of mastitis to vary from 31 to 48% (Wilson et al., 1997; Pitkälä et al., 2004) with culling rates in Ireland due to mastitis ranging from 3 to 13% (Geary et al., 2013b). Currently, only parlours which use separate milk lines for each teat cup (e.g., automatic milking systems) have the capability to automatically measure the SCC of each quarter on a daily basis. When a single milk line is used for all four teat cups, obtaining a measure of SCC for each quarter can be subjective and labour intensive; therefore alternative, non-invasive methods are required.

Infrared thermography (IRT) is a non-invasive technology which can estimate the temperature of an object based on the radiating energy (Speakman and Ward, 1998). Previous research has indicated that IRT has the ability to approximate the SCC of lactating cows by measuring the udder skin surface temperature (USST) (Colak et al., 2008; Polat et al., 2010), but both of these studies were conducted on a single day and in a controlled environment. In Ireland, the majority of milk is produced from cows grazing in-situ; therefore, environmental and cow related factors must be considered (Berry et al., 2003).

The objective of the present study was to investigate the feasibility of using udder IRT to predict SCC in dairy cows on a daily basis. This study aims to investigate the relationship between SCC and a plethora of descriptive temperature parameters (DTPs) and ultimately create a prediction model for USST, the residuals of which may be used to predict SCC. Results from this study could aid in the implementation of a real time automated tool to detect high SCC.

4.3 Materials and Methods

The study was conducted over a two-month period on the Moorepark Research Farm, Teagasc, Fermoy, Co. Cork, Ireland (52.16175 latitude, -8.25344 longitude) commencing in September 2016. All procedures were conducted under approval from the Teagasc Animal Ethics Committee on experimental animal use (TAEC127-2016) in accordance with the Cruelty to Animals Act 1876 and the European Communities Regulations, 1994.

To quantify the relationship between USST and SCC over time, data were recorded on the same 14 Holstein-Friesian cows at evening (15:00 to 16:00) milking every day over a two-month period. Cows enrolled in the study were chosen to have a highly variable SCC in the months preceding the experiment. All udders were shaven before (21st September 2016) and half way through (1st November 2016) the two-month experimental period. Animals were allocated to a new grass paddock each day and received a small quantity of concentrate during each morning and evening milking. The animals walked from the paddock to the parlour at their own pace each day and were allowed a 10-minute acclimatization period in the parlour prior to imaging. Three replicate thermal images were taken of each udder each day prior to the evening milking as per Byrne et al. (2017). All thermal images were captured using a calibrated FLIR T430sc thermal camera (FLIR SYSTEMS Inc., Stockholm, Sweden). The spectral range of the camera was between 7.5

and 13 µm. The camera resolution was 320x240, the thermal sensitivity was <0.03°C, and the accuracy was ± 2°C. All images were captured by the same operator each day in a 30-stand herringbone unit. The operator stood directly behind the animal in the pit of the milking parlour to capture the ventral face of each udder at a distance of 0.8 m.

After the thermal images were captured and the first 10 ml of milk from each quarter discarded, milk samples, of 35ml each (1% of total yield), were then taken from each quarter to be analysed individually for SCC (Somacount 300, Bentley Instruments, Inc, Minnesota, USA). A California Mastitis Test was also performed on all quarter samples by the same operator (Godden et al., 2017).

Additional information was also recorded during the experimental period including the distance the animals walked to the parlour every day using Google maps (87 to 967 m) (Google Inc., California, USA), whether each animal was laying or standing in the paddock before evening milking, and the number of days since the animal's udder was shaved. Ambient temperature and humidity of the parlour was recorded every minute using a Lascar EL-USB-2 data logger (Lascar Electronics, Whiteparish, UK). External environment related variables were recorded on an hourly basis from a weather station located on the Teagasc research farm and these included ambient temperature, relative humidity, rainfall and wind speed. The data for each external environment related factor were retained one and 2 hours before imaging each day, and the maximum and average values for each variable were used in the analysis. Parity and stage of lactation (i.e., days in milk) were also available for each cow.

4.3.1 Image Analysis

Image analysis and temperature extraction were undertaken using the Thermovision LabVIEW toolkit 3.3 (FLIR Systems Inc., Stockholm, Sweden) based on the procedures outlined previously (Byrne et al., 2017). All image parameters (i.e., emissivity, ambient

temperature, humidity, object distance, and reflected temperature) were adjusted in each image before analysis. Emissivity in all images was set to 0.98; ambient temperature and humidity data varied between images and the respective values were taken from the Lascar EL-USB-2 data logger (Lascar Electronics).

For every udder image, a border was drawn by freehand around the ventral face of each quarter, and all calculations were made using only the enclosed pixels (Figure 4.1). Teats and any background information (e.g., legs, underbelly) were not included within the borders.

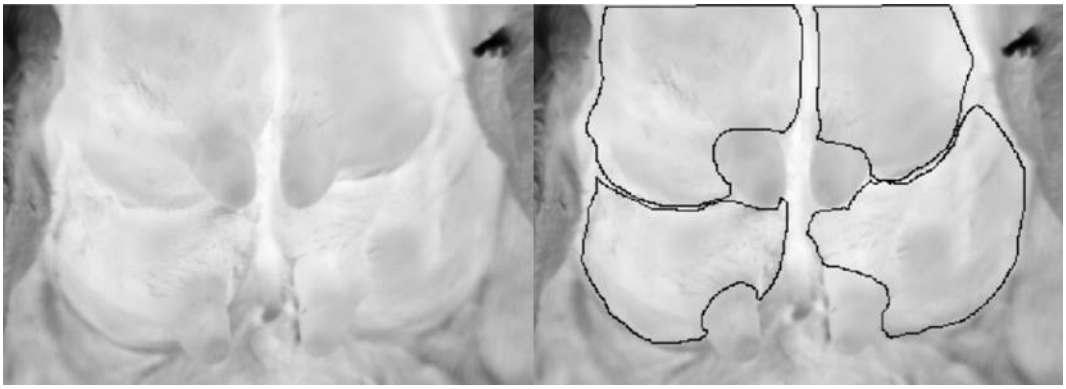


Figure 4.1 A thermal image of an udder is shown on the left, the image on the right shows how this image was cropped. A freehand border (shown in black) was drawn around the ventral face of each quarter and all descriptive temperature parameters were extracted from these regions.

4.3.2 Descriptive Temperature Parameters

Initial DTPs that were extracted from each image consisted of the maximum, minimum and average temperature. As the Thermovision LabVIEW toolkit 3.3 allows manipulation of every pixel in an image, more complex DTPs were also created (Table 4.1). Because the extracted maximum temperature is derived from a single pixel in the image it can be prone to noise; therefore a DTP called Mxavg was created whereby the

average temperature of the 3x3 pixel region with its centre lying on the hottest pixel in an udder quarter was calculated. Additionally, to remove the influence of dirt or hair on the udder, a threshold of 32°C was applied to each quarter whereby any pixel with a temperature less than 32°C was removed from the analysis; the average temperature was then calculated using the remaining pixels thus creating a threshold average above 32°C (Tavg32). A threshold average above 35°C (Tavg35) was also created using the same approach. For two additional DTPs, a threshold relative to the maximum temperature of each quarter was chosen; the weighted maximum at 0.5°C (Wmax0.5) was created by retaining only the pixels within 0.5°C of the maximum temperature of each udder quarter and calculating an average temperature using the retained pixels. The weighted maximum at 2°C (Wmax2.0) was also calculated in the same way, but the threshold was set to within 2°C of the maximum temperature of each udder quarter.

Table 4.1 The abbreviation and description of all descriptive temperature parameters (DTPs) calculated in the present study.

DTP abbreviation	DTP description
Avg	The average temperature of an udder quarter
AASM[1]	$Average\ angular\ second\ moment\ = \sum_i \sum_j \{P(i,j)\}^2$
ACT[1]	$Average\ Contrast = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \right\},\ \lvert i-j \rvert = n$
AC[1]	$Average\ correlation = \frac{\sum_i \sum_j (i,j) P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
ASOS[1]	$Average\ sum\ of\ squares = \sum_i \sum_j (i-\mu)^2 P(i,j)$
AIDM[1]	$Average\ inverse\ difference\ moment = \sum_i \sum_j \frac{P(i,j)}{1+(i-j)^2}$
AE[1]	$Average\ Entropy = \sum_i \sum_j P(i,j) \log\big(P(i,j)\big)$
Max	The maximum temperature of an udder quarter
Min	The minimum temperature of an udder quarter
Mxavg	The average of the 3x3 pixel region who's centre lies on the hottest pixel in an udder quarter
PA32	The percentage of pixels above 32°C in an udder quarter
PA35	The percentage of pixels above 35°C in an udder quarter
PAM-2.0	The percentage of pixels above 2.0°C less than the maximum temperature
PAM-0.5	The percentage of pixels above 0.5°C less than the maximum temperature
Tavg32	The average temperature of all pixels above 32°C
Tavg35	The average temperature of all pixels above 35°C
Wmax2.0	The average temperature of all pixels above 2°C less than the maximum temperature
Wmax0.5	The average temperature of all pixels above 0.5°C less than the maximum temperature

[1] Where *P(i,j)* is the (i,j)th entry in a matrix of temperature measurements, N_g is the number of distinct temperature measurements in the quantized matrix, $\sum_i x_i = \sum_{i=0}^{N_g} x_i$ was the sum of all rows , $\sum_j x_j = \sum_{j=0}^{N_g} x_j$ was the sum of all columns, $\mu_x = \sum_i \sum_j i \times P(i,j)$ was the mean of P_x , $\mu_y = \sum_i \sum_j j \times P(i,j)$ was the mean of P_y, $\mu = \sum_{ij} i \times P(i,j)$ is the mean value of P, $\sigma_x = \sum_i \sum_j (i-\mu_x)^2 \times P(i,j)$ was the standard deviation of P_x and $\sigma_y = \sum_i \sum_j \big(i-\mu_y\big)^2 \times P(i,j)$ was the standard deviation of P_y.

A series of new DTPs were also created based on the pixel count percentages. A

threshold was set at 32°C and the number of pixels above this threshold was divided by the total number of pixels in the udder quarter to calculate the percentage of pixels above 32°C (PA32). A similar method was applied but using a threshold at 35°C to calculate the percentage of pixels above 35°C (PA35). The percentage of pixels within 0.5°C of the maximum temperature (PAM-0.5) was calculated as the number of pixels within 0.5°C of the maximum temperature divided by the total number of pixels in the udder quarter. A similar procedure was used to calculate the percentage of pixels within 2°C of the maximum temperature of each udder quarter (PAM-2.0).

Textural patterns on the surface of the udder were also investigated using grey level co-occurrence matrices (GCLMs) (Haralick and Shanmugam, 1973) which were adapted for thermal imaging for use in the present study. While the aforementioned DTPs quantify temperature and pixel count based attributes of the udder, the GCLMs allow the thermal pattern on the udder to be compared to infection status, similar to Milosevic et al. (2014) who used GCLMs of thermal images to detect breast cancer in humans. Each cell (i, j) in the GCLM denotes the number of times two pixels with grey levels i and j lie a distance D apart at a given angle Ɵ in the image of interest, irrespective of location. Both D and Ɵ were defined before the GCLMs were constructed. Generally, the dimensions of the square GCLM is defined by the number of grey levels in an image, but in the present study the dimensions of the GCLM was defined as the number of distinct temperature levels in a given image. To facilitate the definition of the dimensions of the GCLM, every temperature in each thermal image was rounded to one decimal place. For the current study, a distance of one, as well as four different angles 0°, 45°, 90° and 135° were considered for all GCLMs. Therefore, four different GCLMs were constructed for each udder quarter and six different textural values (Milosevic et al., 2014) were calculated for each GCLM. The mean value for each textural feature was taken across all four angles for each udder quarter; all the textural features were unit less. The calculated textural features

were as follows:

Eq. 1

$$Average\ angular\ second\ moment\ = \sum_i \sum_j \{P(i,j)\}^2$$

Eq. 2

$$Average\ contrast\ = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \right\}, |i-j| = n$$

Eq. 3

$$Average\ correlation\ = \frac{\sum_i \sum_j (i,j) P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Eq. 4

$$Average\ sum\ of\ squares\ = \sum_i \sum_j (i-\mu)^2 P(i,j)$$

Eq. 5

$$Average\ inverse\ difference\ moment\ = \sum_i \sum_j \frac{P(i,j)}{1+(i-j)^2}$$

Eq. 6

$$Average\ Entropy\ = -\sum_i \sum_j P(i,j) \log\big(P(i,j)\big)$$

where *P(i,j)* is the (i,j)th entry in a matrix of temperature measurements, N_g is the number of distinct temperature measurements in the quantized matrix, $\sum_i x_i = \sum_{i=0}^{N_g} x_i$ was the sum of all rows , $\sum_j x_j = \sum_{j=0}^{N_g} x_j$ was the sum of all columns, $\mu_x = \sum_i \sum_j i \times P(i,j)$ was the mean of P_x , $\mu_y = \sum_i \sum_j j \times P(i,j)$ was the mean of P_y, $\mu = \sum_{ij} i \times P(i,j)$ is the mean value of P, $\sigma_x = \sum_i \sum_j (i-\mu_x)^2 \times P(i,j)$ was the standard deviation of P_x and $\sigma_y = \sum_i \sum_j \big(i-\mu_y\big)^2 \times P(i,j)$ was the standard deviation of P_y.

4.3.3 Statistical analyses

To investigate the association between the DTPs (described in Table 4.1) and SCC, as well as the association between each environmental factor and DTP, a series of analyses was undertaken as shown below:

4.3.3.1 Linear regression.

By examining 62 cows on a single day Polat et al. (2010) demonstrated a linear relationship between the natural log of SCC (also known as somatic cell score) and USST; to test the validity of this, the Pearson correlation co-efficient between every DTP in the present study and SCS across the entire dataset was calculated (PROC CORR; SAS Institute, 2010).

4.3.3.2 Association of the range of a DTP and SCC within udder.

To investigate whether an animal can be used as its own control to facilitate the prediction of milk SCC from a DTP, a selection of additional DTP and SCC variables were created. The range in each DTP within each udder was calculated as the difference in temperature between the quarter with the highest DTP value and the quarter with the lowest DTP value. The natural logarithm of the difference in SCC between the respective quarters was also calculated. A linear regression was performed using PROC REG (SAS Institute, 2010) for each of the 18 DTPs individually where the dependent variable was the log of the difference in SCC and the calculated DTP range was the independent variable. Fit statistics, the regression coefficient and the Pearson correlation co-efficient (PROC CORR; SAS Institute, 2010) were calculated.

4.3.3.3 Association of SCS and the divergence of a DTP from a daily baseline.

The difference in a DTP from a quarter of a given cow relative to the average of the five coldest quarters on that day (i.e., baseline value), which could have come from between two and five different cows each day, was calculated to quantify its usefulness as

a predictor of milk SCS (natural log of SCC). A linear regression was performed using PROC REG (SAS Institute, 2010) for each of the 18 DTPs individually where SCS was the dependent variable and the difference between a DTP value and the respective baseline DTP value was the independent variable. Fit statistics, the regression coefficient and the Pearson correlation co-efficient (PROC CORR; SAS Institute, 2010) were calculated.

4.3.3.4 Association between environmental factors and DTPs.

To create prediction models for each DTP, the residuals of which could be compared to SCC, the association between each DTP and a range of environmental-level (i.e., external ambient temperature, external relative humidity, ambient temperature at the site of imaging, relative humidity at the site of imaging, rainfall, and wind speed) and cow-level factors (i.e., parity (1, 2, 3, 4, and ≥5) and stage of lactation (classified in 20 day intervals)) was first quantified. A stepwise selection using a multiple regression model in PROC REG (SAS Institute, 2010) was performed whereby each DTP was included separately as the dependent variable and all environmental and cow related factors were considered as covariates. The significance threshold for entry and exit of the terms into/from the model was set at 0.1% before testing for co-linearity between the factors. Normality checks were performed on all model residuals.

4.3.3.5 Prediction of each DTP using a single environmental factor.

To investigate the usefulness of a single environmental factor in predicting a DTP on a per quarter basis, the most influential environmental factor (MIF) (calculated from previous analysis (section 3.3.3.4: Association between environmental factors and DTPs)) for each DTP was used to predict a DTP on a single day (day 0) for each cow. A pre-defined number of training days (day -2, day -5, and day -10) prior to each prediction day (day 0) were used while all other recorded DTPs were temporarily discarded. Each DTP was then estimated for every day using a linear mixed model with a random intercept term in PROC

MIXED (SAS Institute, 2010), where quarter nested within cow was used as a random effect (Eq. 7):

Eq. 7 $$\boldsymbol{DTP} = \left(\mu_{\mathbf{p}} + \boldsymbol{a_j}(\mu_{\mathbf{i}})\right) + \left(\mathbf{b_p} * \mathbf{MIF}\right)$$

where μ_p is the population intercept, a_j is the individual cow intercept, μ_i is the individual quarter intercept, and b_p is the population regression co-efficient of the MIF. To measure the fit of the model, the estimated DTP and actual recorded DTP were used to calculate root mean square error (RMSE) and R^2. The 5^{th} and 95^{th} percentiles of the residual values were also calculated. To investigate if the residuals from a linear mixed model with a random intercept could be used to predict health status of an animal, the residuals (calculated from Eq. 7) and actual DTPs were compared to animal health status. The udder quarter health status of the animals was determined based on SCC; a healthy quarter was defined as a quarter with a SCC <400,000 cells/ml and if the SCC was between 400,000 and 1,000,000 cells/ml but did not increase by 300,000 cells/ml from the previous day it was also considered healthy; all other quarters were classified as non-healthy. To investigate the association between the USST of healthy and non-healthy animals, and the residuals from the mixed model analysis, a linear model in PROC GLM (SAS Institute, 2010) was used where the mixed model residuals were the dependent variable and animal health status (categorized by SCC) was the independent variable.

4.4 Results

Ambient temperature during the experimental period ranged from 2.6 to 16.2°C, while the relative humidity ranged from 58 to 95%; the average daily USST ranged from 31.1 to 35.5°C. Somatic cell count ranged from 1,000 to 10,000,000 cells/ml. The mean (SD in parenthesis) days in milk of all cows at the start of the experiment was 218 (72) d.

Two cows were treated for mastitis during the experiment, but, only a temporary reduction in SCC was observed in both cases.

4.4.1 Linear regression

The Pearson correlation co-efficient between recorded maximum USST and SCS was -0.01. Similarly, the Pearson correlation co-efficient between recorded average USST and SCS was 0.01. The average USST recorded in this current study and USST calculated by Polat et al. (2010) ($\widehat{\mathrm{USST}} = 32.12 + (0.49 \times \ln(\frac{\mathrm{SCC}}{1000}))$) are shown in Figure 4.2, where a clear disparity between the results of both studies can be seen. All textural-based DTPs had a significant relationship with SCS with the Pearson correlation co-efficient ranging from -0.16 (average entropy (Eq. 6)) to 0.19 (average inverse difference moment (Eq. 5)). The only other DTP that had a significant relationship ($P<0.05$) with SCS was PAM-0.5, where a Pearson correlation co-efficient of 0.07 was observed.

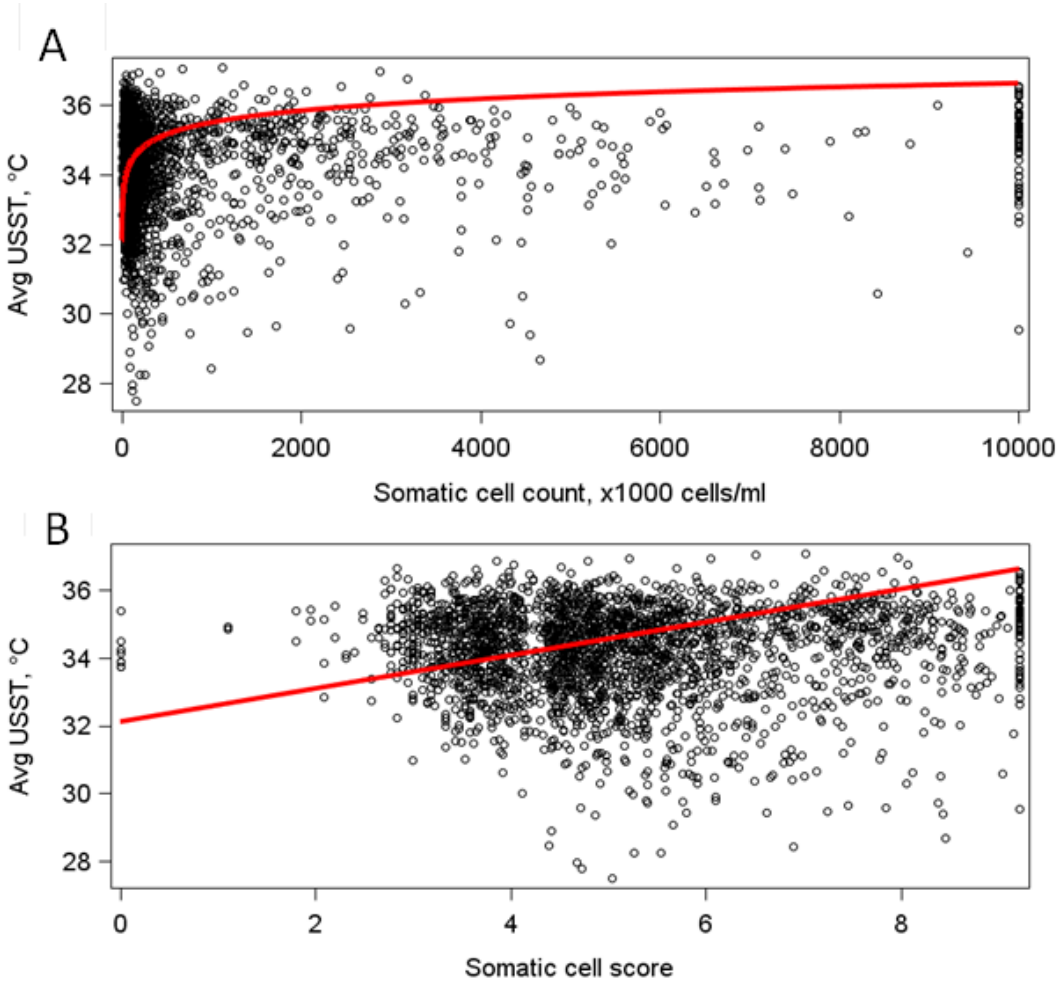


Figure 4.2 The recorded (○) and expected (—) average udder skin surface temperature (Avg USST) (Polat et al., 2010) for: a) the expected logarithmic relationship between SCC and USST, and b) the expected linear relationship between somatic cell score (SCS) and USST.

4.4.2 Association of the range of a DTP and SCC within udder

The range of PAM-2.0 (2 to 66%) within each udder was the only DTP to be associated ($P<0.05$) with the log of the difference in SCC. The R^2 for PAM-2.0 was 0.02 and the regression coefficient (SE in parenthesis) was 18 (5).

4.4.3 Association of SCS and the divergence of a DTP from a daily baseline

When the difference between a DTP value and its respective daily baseline was regressed onto SCS all DTPs were associated ($P<0.05$) with SCS, with the exception of PAM-2.0. The regression coefficients of all DTPs were positive, with the exception of average entropy (Eq. 6), average contrast (Eq. 2), and PAM-0.5. The regression coefficients for temperature-based DTPs ranged from 0.07 (minimum USST) to 0.44 (Tavg35). The regression coefficients (SE in parenthesis) for average entropy (Eq. 6), average contrast (Eq. 2) and PAM-0.5 were -1.59 (0.13), -0.31 (0.03), and -8.63 (2.92), respectively. The proportion of variation (i.e., R^2) in SCS which was explained by the DTPs ranged from 0.00 (PAM-0.5) to 0.05 (average entropy (Eq. 6)) (Table 4.2).

Table 4.2 The r-squared (R^2), regression coefficient (b) and the corresponding standard error (SE) for the association between somatic cell score and the difference between a DTP value and the baseline (i.e., average of the five lowest values of that DTP) for the respective day.

DTP[1]	R^2	b	SE
AE	0.05	-1.59	0.13
AIDM	0.05	4.88	0.41
ACT	0.04	-0.31	0.03
Avg	0.03	0.30	0.03
Max	0.02	0.36	0.05
Wmax0.5	0.02	0.36	0.05
PA32	0.02	1.42	0.18
AASM	0.02	91.77	12.18
ASOS	0.02	0.00	0.00
Wmax2.0	0.02	0.32	0.04
Min	0.02	0.07	0.01
Tavg32	0.02	0.30	0.05
Tavg35	0.01	0.44	0.08
AC	0.01	0.00	0.00
Mxavg	0.01	0.15	0.03
PA35	0.01	0.49	0.12
PAM-0.5	0.00	-8.63	2.92

[1]where Avg = The average temperature of an udder quarter, AASM = Average angular second moment, ACT = Average contrast AC = Average correlation, ASOS = Average sum of squares, AIDM = Average inverse difference moment, AE = Average entropy,Max = The maximum temperature of an udder quarter, Min = The minimum temperature of an udder quarter, Mxavg = The average of the 3x3 pixel region who's centre lies on the hottest pixel in an udder quarter, PA32 = The percentage of pixels above 32°C in an udder quarter, PA35 = The percentage of pixels above 35°C in an udder quarter, PAM-0.5 = The percentage of pixels within 0.5°C of the maximum temperature, Tavg32 = The average temperature of all pixels above 32°C, Tavg35 = The average temperature of all pixels above 35°C, Wmax2.0 = The average temperature of all pixels within 2°C of the maximum temperature, Wmax0.5 = The average temperature of all pixels within 0.5°C of the maximum temperature.

4.4.4 Association between environmental factors and DTPs

When the association between the environmental factors and DTPs were investigated, the environmental factors accounted for between 6 (average correlation (Eq. 3)) and 53% (average USST) of the variation in the investigated DTPs. Irrespective of

DTP, the number of environmental factors that were associated ($P<0.01$) with each DTP ranged from 4 (average correlation (Eq. 3)) to 7 (average USST), albeit much of the variation was accounted for by a single factor (R^2 = 3 to 36%), which differed by the DTP under investigation (Table 4.3). The number of days since the udder was shaved accounted for between 10 and 23% of the variation in most textural-based DTPs (i.e., average angular second moment, average contrast, average inverse difference moment and average entropy). For the average correlation (Eq. 3), ambient temperature at the time of imaging accounted for 3% of the variation, while the maximum ambient temperature observed 2 hours before imaging accounted for 13% of the variation in the average sum of squares DTP (Eq. 1). For all temperature and pixel count-based DTPs, except PAM-0.5, the maximum ambient temperature observed 2 hours before imaging accounted for the largest proportion of variation (R^2 = 9 to 36%). Other influential factors which accounted for some of the variation in the DTPs included parity (R^2 = 2 to 12%), stage of lactation (R^2 = 0 to 4%) and the total amount of precipitate that fell 2 hours before imaging (R^2 = 0 to 4%). Humidity and wind related factors accounted for a small proportion of the variation in any DTPs (0 to 3%). The distance animals walked to the parlour was not associated with any of the DTPs investigated in the current study. The DTPs with the greatest proportion of total variation accounted for by multiple environmental factors were average USST (53%), PA32 (51%), and maximum USST (50%) (Table 4.3). The DTPs with the greatest proportion of total variation accounted for by a single environmental factor, which in all cases was the maximum ambient temperature observed 2 hours before to imaging, were average USST (36%), PA32 (35%), and Wmax0.5 (31%) (Table 4.3).

Table 4.3 The number of significant environmental factors (No. of factors), associated R-squared (R^2), the most influential environmental factor (MIF) and the r squared when only the MIF is included in the model (R^2 with MIF) for each descriptive temperature parameters (DTP).

DTP[1]	R^2	No. of factors	MIF[2]	R^2 with MIF
Avg	0.53	7	Amb2hmax	0.36
PA32	0.51	6	Amb2hmax	0.35
Max	0.50	5	Amb2hmax	0.30
Wmax0.5	0.49	5	Amb2hmax	0.31
Tavg32	0.48	7	Amb2hmax	0.29
Wmax2.0	0.48	5	Amb2hmax	0.30
AIDM	0.45	7	days unshaven	0.23
PA35	0.41	7	Amb2hmax	0.23
Tavg35	0.40	5	Amb2hmax	0.25
ACT	0.39	6	days unshaven	0.18
AE	0.37	6	days unshaven	0.12
Min	0.36	7	Amb2hmax	0.23
Mxavg	0.34	5	Amb2hmax	0.21
AASM	0.29	7	days unshaven	0.10
ASOS	0.27	5	Amb2hmax	0.11
PAM-2.0	0.23	6	Amb2hmax	0.09
PAM-0.5	0.07	5	Amb1h	0.03
AC	0.06	4	Amb1h	0.03

[1]where Avg = The average temperature of an udder quarter, AASM = Average angular second moment, ACT = Average contrast AC = Average correlation, ASOS = Average sum of squares, AIDM = Average inverse difference moment, AE = Average entropy,Max = The maximum temperature of an udder quarter, Min = The minimum temperature of an udder quarter, Mxavg = The average of the 3x3 pixel region who's centre lies on the hottest pixel in an udder quarter, PA32 = The percentage of pixels above 32°C in an udder quarter, PA35 = The percentage of pixels above 35°C in an udder quarter, PAM-2.0 = The percentage of pixels within 2.0°C of the maximum temperature,PAM-0.5 = The percentage of pixels within 0.5°C of the maximum temperature, Tavg32 = The average temperature of all pixels above 32°C, Tavg35 = The average temperature of all pixels above 35°C, Wmax2.0 = The average temperature of all pixels within 2°C of the maximum temperature, Wmax0.5 = The average temperature of all pixels within 0.5°C of the maximum temperature.

[2]where amb2hmax = the maximum ambient temperature observed 2 hours before imaging, days unshaven = the number of days since the udder was shaven, Amb1h = the ambient temperature at the time of imaging.

4.4.5 Prediction of each DTP

The R^2 for the prediction of maximum USST on a single day (day 0) for each cow, increased from 67 to 76% when the number of training days increased from two to ten; for average USST, the R^2 increased from 65 to 74% for the same increase in training days. The RMSE decreased as the number of training days increased, decreasing from 0.27 to 0.23°C for maximum USST and from 0.38 to 0.35°C for average USST when two and ten training days were used, respectively. The difference between the 5^{th} and 95^{th} percentile of the maximum USST model residuals (discrepancy between predicted and actual maximum USST) decreased from 4.78°C to 1.84°C when the number of training days was increased from two to ten; a similar trend was observed for all other DTPs.

When a linear mixed model with a random intercept was used, a single environmental factor accounted for between 48 (PAM-0.5) and 79% (average entropy (Eq. 3)) of the variation in the DTPs using a 10 day training period (Table 4.4). The temperature-based DTP with the greatest R^2 (77%) was Wmax2.0; a RMSE of 0.23°C was achieved and the 5^{th} and 95^{th} percentiles of the model residuals were -0.99°C and 0.92°C, respectively (Table 4.4

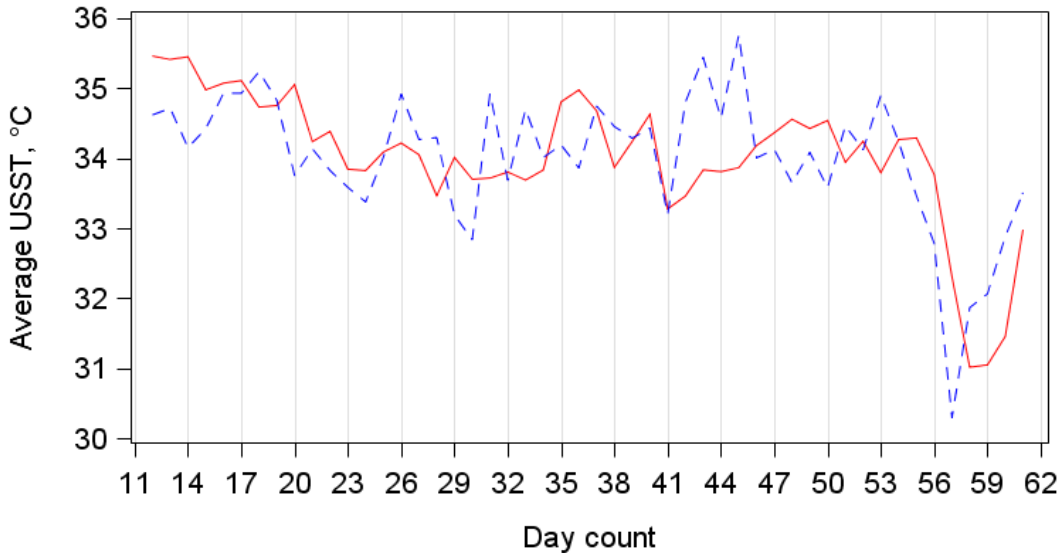


Figure 4.3 The recorded (– –) and estimated (—) maximum udder skin surface temperature across the experimental period of the right hind quarter of a single cow using linear mixed model with a random intercept and 10 days of training data.

). Results for maximum USST were very similar to Wmax2.0; an illustration of the relationship between recorded and predicted maximum USST is shown in Figure 4.3. The pixel count-based DTP with the greatest R^2 (71%) was PA35; a RMSE of 10% was achieved and the 5^{th} and 95^{th} percentiles of the model residuals were -39% and 40%, respectively (Table 4.4).

Table 4.4 The r-squared (R^2), the root mean squared error (RMSE), and the 5th (5th Perc) and 95th (95th Perc) percentiles of the model residuals when linear mixed model with a random intercept was used to estimate each descriptive temperature parameter (DTP) with 10 days of training data.

DTP	R^2	RMSE	5th Perc	95th Perc
AE	0.79	0.09	-0.34	0.35
AIDM	0.78	7.81	-0.11	0.11
Wmax2.0	0.77	0.23	-0.99	0.92
Max	0.77	0.23	-0.96	0.88
Wmax0.5	0.76	0.24	-0.97	0.88
AASM	0.75	0.04	0.00	0.00
ACT	0.75	0.37	-1.26	1.61
ASOS	0.75	6366.35	-151.47	179.28
Avg	0.74	0.35	-1.58	1.44
Tavg35	0.74	0.14	-0.56	0.54
PA35	0.71	0.10	-0.39	0.40
Tavg32	0.71	0.23	-1.15	1.10
PA32	0.67	0.08	-0.21	0.16
AC	0.66	152951.53	-422726.25	457744.04
PAM2.0	0.64	0.10	-0.27	0.31
Mxavg	0.59	0.61	-1.06	1.05
Min	0.54	1.68	-5.35	4.27
PAM0.5	0.48	0.01	-0.01	0.02

[1] where Avg = The average temperature of an udder quarter, AASM = Average angular second moment, ACT = Average contrast AC = Average correlation, ASOS = Average sum of squares, AIDM = Average inverse difference moment, AE = Average entropy,Max = The maximum temperature of an udder quarter, Min = The minimum temperature of an udder quarter, Mxavg = The average of the 3x3 pixel region who's centre lies on the hottest pixel in an udder quarter, PA32 = The percentage of pixels above 32°C in an udder quarter, PA35 = The percentage of pixels above 35°C in an udder quarter, PAM-2.0 = The percentage of pixels within 2.0°C of the maximum temperature,PAM-0.5 = The percentage of pixels within 0.5°C of the maximum temperature, Tavg32 = The average temperature of all pixels above 32°C, Tavg35 = The average temperature of all pixels above 35°C, Wmax2.0 = The average temperature of all pixels within 2°C of the maximum temperature, Wmax0.5 = The average temperature of all pixels within 0.5°C of the maximum temperature.

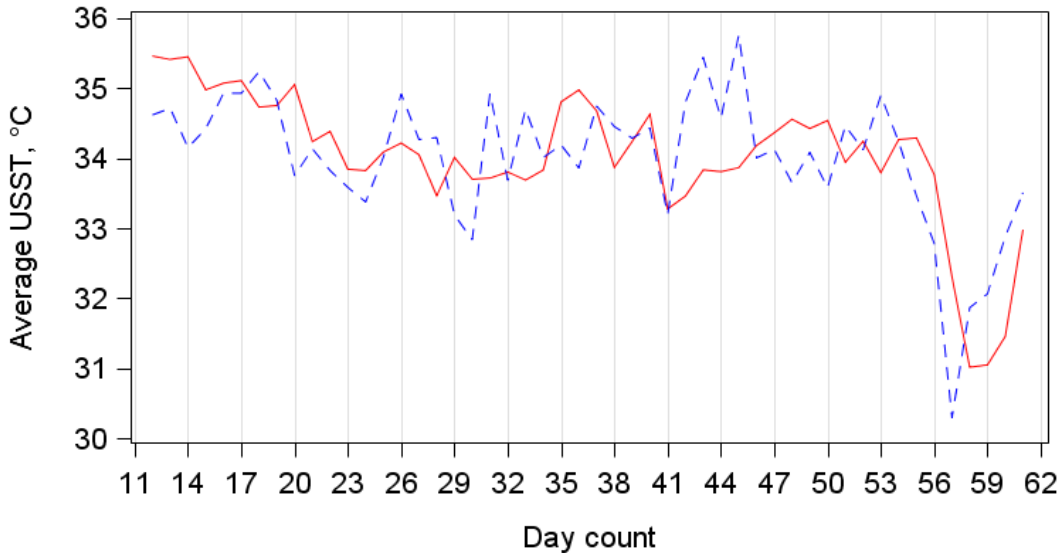


Figure 4.3 The recorded (– –) and estimated (—) maximum udder skin surface temperature across the experimental period of the right hind quarter of a single cow using linear mixed model with a random intercept and 10 days of training data.

A total of 459 records (18%) were categorized as non-healthy using the observed SCC. Based on the residuals calculated from the mixed model, four DTPs (PA35, Tavg32, PAM-2.0, and average correlation (Eq. 3)) had the ability to differentiate healthy from non-healthy animals (Table 4.5). When Tavg32 was considered, the SD of residuals for healthy and non-healthy animals was 0.74°C and 0.73°C, respectively, while the mean (SE in parenthesis) of the residuals for healthy and non-healthy animals was 0.01 (0.01) and 0.10 (0.03)°C, respectively. The residual mean of PA35 and PAM-2.0 for healthy and non-healthy animals was similar across both DTPs, but the standard deviation of the model residuals for PA35 (25%) was larger in comparison to PAM-2.0 (19%). The mean residual (SE) for healthy and non-healthy animals was 35,449.61 (11,949) and 24,314.25 (25,859.34) (unit less), respectively, implying that healthy and non-healthy animals were indistinguishable.

Table 4.5 The mean (SE in parenthesis) and SD of the residuals from the significant (P<0.05) descriptive temperature parameter (DTP) models for healthy and non-healthy animals.

DTP	Healthy mean	Healthy SD	Non-healthy mean	Non-healthy SD
PA35	0.00(0.01)	0.25	0.03(0.01)	0.25
Tavg32	0.01(0.01)	0.74	0.10(0.03)	0.73
PAM-2.0	0.00(0.00)	0.19	0.02(0.01)	0.19
AC	35449.61(11949)	591881.81	-24314.25 (25859.34)	572590.55

[1] where AC = Average correlation, ASOS = Average sum of squares, PA35 = The percentage of pixels above 35°C in an udder quarter, PAM-2.0 = The percentage of pixels within 2.0°C of the maximum temperature, Tavg32 = The average temperature of all pixels above 32°C.

4.5 Discussion

Elevated SCC can be due to an infection in the mammary gland and, as such, can negatively impact farm profitability. Most milking parlours cannot easily measure SCC at an individual quarter level; therefore, a new approach is required to generate such data. The objective of the present study was to investigate the feasibility of using IRT of the udder to predict the SCC of grazing Holstein-Friesian cows on farm. The relationship between SCC and a multitude of different DTPs was tested using various analytical techniques. Results indicate that SCC could not be readily predicted using IRT. While the maximum and average USST could be predicted with a RMSE of 0.23°C and 0.35°, respectively, deviations in recorded USST from an expectation could not differentiate healthy and non-healthy animals.

4.5.1 Integrity in experimental procedure

To ensure reliability of the results in any experiment, the data must be collected and analysed thoroughly; the two main sources of data in this experiment were USST measured using infrared thermography and SCC of milk, both of which require certain procedures to reduce erroneous measurements. A temperature measurement taken from an animal using IRT can be affected by a range of factors, including the environment, stress on the animal, activity, circadian rhythm, hair on the region of interest, the precision of the camera, and the consistency of the operator (Berry et al., 2003; Church et al., 2014; Byrne et al., 2017). To ensure all of these factors do not cause erroneous measurements, each must be recorded or mitigated through certain procedures. Firstly, environmental factors were recorded throughout the experiment both on farm and at the site of imaging so the impact of each factor could be quantified through analysis. In the present study, ambient temperature accounted for most of the explainable variation in temperature- and pixel count-based DTPs (range of 9 to 36%). Secondly, stress experienced by the animal was minimized as all animals were allowed to walk from the field into milking parlor at their own pace in the same manner as they would for milking. The distance animals walked to the parlor was recorded every day as an estimate of exercise. This distance did not have a significant relationship with any of the DTPs, indicating that the ten minute acclimatization period was sufficient to reduce the effect of exercise on USST. A change in USST due to the circadian rhythm was mitigated against as animals were imaged at the same time each day; 15:00 to 16:00 was chosen because variation of USST due to the circadian rhythm is minimized during this time period (Berry et al., 2003). The number of days since the udder was shaved was recorded as an estimate for the amount of hair on the udder, but only had a significant association with textural-based DTPs and average USST which indicates that some temperature-based DTPs, such as maximum USST, were robust to variations in udder hair. As the animals were allowed out to pasture each day, none had large amounts

of dirt on the udder. Byrne et al. (2017) recommended that three thermal image replicates should be taken of every udder, and the same operator should capture every image within a study to increase the precision of IRT; the present study followed this protocol. Finally, to further increase the integrity of the experimental procedure, a calibrated camera was used and no udder was handled prior to imaging, as this can cause a change in USST (Paulrud et al., 2005).

When extracting milk samples for the measurement of SCC, similar methods were used to those described by Polat et al. (2010), whereby the first 10ml of milk was discarded before 35ml of milk was collected from each quarter on a daily basis. Alternative methods of milk collection involve taking samples of milk throughout the entire milking. Nielsen et al. (2005) demonstrated that while SCC can increase slightly during milking, the difference in SCC between healthy and non-healthy quarters remains consistent throughout milking. Thermal images and milk samples in the present study were taken at evening milking (15:00 to 16:00). Consequently, animals in the present study only had a 7 hour period since last milking; this may have resulted in slightly higher SCC, though this does not offset the difference in SCC between healthy and unhealthy animals (Nielsen et al., 2005). Both SCS and the California mastitis test results were assessed as part of this study and found to have a high concordance ($r = 0.85$).

4.5.2 Experimental outputs

Polat et al. (2010) demonstrated a linear relationship between SCS and USST with an R^2 of 72%. Results from the present study showed a large discrepancy between expected (Polat et al., 2010) and recorded USST (Figure 4.2). One of the main factors which possibly contributed to the differences between both studies was the ambient temperature, which may affect an animal's thermoregulatory response to infection; Polat et al. (2010) kept the animals in a room which was at an animal's thermal neutral zone (18 to 23°C)

before imaging whereas the present study allowed animals to stand in a parlor with temperatures ranging from 3.5 to 16.2°C, which is typical of grazing based systems in Ireland. Dissimilarities were not resolved when data from the single hottest day in the current study was compared to the results from Polat et al. (2010). Additional distinctions between the methodology of the present study and that of Polat et al. (2010) include animal breed (Holstein-Friesian (current study) vs Brown Swiss (Polat et al. 2010)), acclimatization period (30 minutes vs 10 minutes), number of non-healthy records (459 vs 154), and udder thermal image view (lateral and caudal view vs ventral view).

Previous studies have traditionally used the average temperature of the udder (Berry et al., 2003; Hovinen et al., 2008; Polat et al., 2010; Pampariene et al., 2016) to determine the effect of exercise, ambient temperature and infection status on USST. While Metzner et al. (2014) also compared maximum and minimum USST to infection status, no agricultural study has investigated how alternative DTPs from udder thermal images relate to SCS. In the present study, the correlation observed between any of the DTPs and SCS (-0.16 to 0.19) was too weak for practical use as any predictions would contain a large number of false positives and false negatives. Therefore, alternative methodologies were tested. Previous studies have successfully related the temperature difference between two regions on the hoof to hoof lesions (Nikkhah et al., 2005; Alsaaod and Büscher, 2012), in effect using the animal as her own control. Paulrud et al. (2004) also suggested detecting mastitis using contralateral teat temperature differences. Similarly, in the current study, the temperature range within each udder was compared to the respective natural log of the range in SCC, essentially also using the animal as her own control in an attempt to mitigate factors such as ambient temperature and the amount of hair on the udder. This technique would not be useful on farm as many of these measures did not have a significant relationship with the log of the range in SCC. In a separate analysis, the temperature difference between a DTP and its daily baseline (i.e., average of the five lowest values of

that DTP) was compared to SCS, thereby using the daily baseline as a control in an attempt to mitigate environmental factors which the entire herd would be subjected to (e.g., ambient temperature). The proportion of variation in SCS (0 to 5%) which was accounted for by the difference between a DTP and its daily baseline was too weak for practical use.

While using the animal as its own control did not help to predict SCS, the next step was to quantify the association between various environmental and cow related factors with each DTP. Berry et al. (2003) showed that ambient temperature, exercise and circadian rhythm was associated with average USST but no study has investigated how alternative udder DTPs are associated with various environmental or cow factors. Temperature-based DTPs (maximum USST, average USST, minimum USST, Wmax0.5, Wmax2.0, Tavg32, Tavg35, Mxavg) had the greatest association with ambient temperature (R^2 ranged from 21 (Mxavg) to 36% (average USST)), which is consistent with previous literature for the average USST of dairy cattle (Berry et al., 2003). Parity had the second greatest association with temperature-based DTPs (R^2 ranged from 2 (Tavg35) to 12% (maximum USST)), with all temperature-based DTPs decreasing as parity increased; Nikkhah et al. (2005) showed a similar association between parity and the hoof temperature of cattle but not with USST. Other factors such as days since the udder was shaved had a minor association with average USST (R^2 = 3%), yet it was not related to maximum USST. As maximum USST is derived from the single hottest pixel it was taken from the skin rather than the hair, unlike average USST which is calculated using every element of the udder. While some studies have used shaven (Metzner et al., 2014) or unshaven udders (Berry et al., 2003), no study has investigated the association between hair on the udder of a dairy cow and various DTPs. Only a small proportion of the total variation (6 to 45%) in textural-based DTPs could be accounted for by environmental factors, indicating that the textural-based DTPs may be more prone to noise in comparison to other DTPs.

When linear mixed models were used to predict each DTP, a single factor could account for between 3 and 36% of the variation in the DTPs. When linear mixed models with a random incept term were used, a single factor could account for between 48 and 79% of the variation because each udder quarter was modelled individually. A linear mixed model with a random regression coefficient was also tested but RMSE and R^2 values were not improved. Berry et al. (2003) used a linear regression model with lag to predict average USST and has achieved slightly better RMSE and R^2 in comparison to the present study. Gloster et al. (2011) suggested that colder conditions cause a greater variability in the temperatures of some anatomical regions, which may account for the difference in results between the present study and the study by Berry et al. (2003).

Ultimately, the present study aimed to differentiate healthy and non-healthy animals (defined by SCC) using USST. Gloster et al. (2011) suggested that mathematical models could be used to predict the temperature of an anatomical area and any deviations from the prediction (the model residuals) could be related to infection. Therefore, a prediction model for USST was developed. The natural deviation due to infection must be greater than the distribution of the prediction model residuals as any deviation due to infection less than the distribution of the residuals can only be attributed to model inaccuracy rather than infection. The 5^{th} and 95^{th} percentiles of maximum USST with a 10 day training period (-0.96 and 0.88°C Table 4.4) were less than the difference between healthy and non-healthy animals reported by previous studies which ranged from 0.9 to 2.4°C (Colak et al., 2008; Polat et al., 2010). However, in the present study the magnitude of the maximum USST residuals did not differ between healthy and non-healthy animals. While the residuals of Tavg32 differed (0.9°C; $P<0.05$) between healthy and non-healthy animals, a large number of animals from both data sets could not be differentiated. The expected difference between actual and predicted USST is due to a change in blood flow to the site of infection due to the inflammatory response (Martins et al., 2013). Hovinen et al. (2008) and

Pezeshki et al. (2011) inoculated udders with *E.coli* and suggested that this change in blood flow may only last a number of hours and hence images taken daily may not be sufficient to detect infection in the udder. While these studies (Hovinen et al., 2008; Pezeshki et al., 2011b) have investigated the change in USST due to *E.coli*, no study exists that investigated the USST differences between various udder pathogens.

4.6 Conclusion

The current study attempted to relate USST and SCC by using simple linear regression, a cow as her own control, a daily base line, and a linear mixed model with a random intercept; none of these proved to be sufficiently accurate to differentiate healthy and non-healthy animals. Future studies may attempt to repeat this work in warmer/ more controlled environments and to gather more frequent measurements (i.e. greater than once a day) in conjunction with an automatic milking machine or water trough; the effect of the circadian rhythm on the animals USST must be accounted for in these scenarios. Finally, the analytical methods and additional DTPs developed as part of this study could be tested on various anatomical regions for measuring alternative phenotypes, for example, feed efficiency or infection in the hoof.

4.7 Literature Cited

Alsaaod, M., and W. Büscher. 2012. Detection of hoof lesions using digital infrared thermography in dairy cows. J. Dairy Sci. 95: 735-742.

Berry, R., A. Kennedy, S. Scott, B. Kyle, and A. Schaefer. 2003. Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection. Can. J. Anim. Sci. 83: 687-693.

Byrne, D., D. Berry, H. Esmonde, and N. McHugh. 2017. Temporal, spatial, inter-, and intra-cow repeatability of thermal imaging. J. Anim. Sci. 95: 970-979.

Church, J., P. Hegadoren, M. Paetkau, C. Miller, G. Regev-Shoshani, A. Schaefer, and K. Schwartzkopf-Genswein. 2014. Influence of environmental factors on infrared eye temperature measurements in cattle. Res. Vet. Sci. 96: 220-226.

Colak, A., B. Polat, Z. Okumus, M. Kaya, L. Yanmaz, and A. Hayirli. 2008. < I> short communication: early detection of mastitis using infrared thermography in dairy cows. J. Dairy Sci. 91: 4244-4248.

Geary, U., N. Begley, F. McCoy, B. O'Brien, L. O'Grady, and L. Shalloo. 2011. Estimating the Impact of Mastitis on the Profitability of Irish Dairy Farms. Pages 221-228 in Udder Health and Communication. Springer,

Geary, U., N. Lopez-Villalobos, B. O'Brien, D. Garrick, and L. Shalloo. 2013. Examining the impact of mastitis on the profitability of the irish dairy industry. Irish. J. Agr. Food Res. : 135-149.

Gloster, J., K. Ebert, S. Gubbins, J. Bashiruddin, and D. J. Paton. 2011. Normal variation in thermal radiated temperature in cattle: Implications for foot-and-mouth disease detection. BMC Vet. Res. 7: 73-6148-7-73. doi: 10.1186/1746-6148-7-73 [doi]

Godden, S., E. Royster, J. Timmerman, P. Rapnicki, and H. Green. 2017. Evaluation of an automated milk leukocyte differential test and the california mastitis test for detecting intramammary infection in early-and late-lactation quarters and cows. J. Dairy Sci.

Haralick, R. M., and K. Shanmugam. 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. : 610-621.

Harmon, R. J. 1994. Physiology of mastitis and factors affecting somatic cell Counts1. J. Dairy Sci. 77: 2103-2112.

Hovinen, M., J. Siivonen, S. Taponen, L. Hänninen, M. Pastell, A. Aisla, and S. Pyörälä. 2008. Detection of clinical mastitis with the help of a thermal camera. J. Dairy Sci. 91: 4592-4598.

Martins, R. F. S., T. do Prado Paim, C. de Abreu Cardoso, B. Stéfano Lima Dallago, C. B. de Melo, H. Louvandini, and C. McManus. 2013. Mastitis detection in sheep by infrared thermography. Res. Vet. Sci. 94: 722-724.

Metzner, M., C. Sauter-Louis, A. Seemueller, W. Petzl, and W. Klee. 2014. Infrared thermography of the udder surface of dairy cattle: Characteristics, methods, and correlation with rectal temperature. Vet. J. 199: 57-62.

Milosevic, M., D. Jankovic, and A. Peulic. 2014. Thermography based breast cancer detection using texture features and minimum variance quantization. EXCLI J. 13: 1204-1215.

Nielsen, N. I., T. Larsen, M. Bjerring, and K. L. Ingvartsen. 2005. Quarter health, milking interval, and sampling time during milking affect the concentration of milk constituents. J. Dairy Sci. 88: 3186-3200.

Nikkhah, A., J. Plaizier, M. Einarson, R. Berry, S. Scott, and A. Kennedy. 2005. Infrared thermography and visual examination of hooves of dairy cows in two stages of lactation. J. Dairy Sci. 88: 2749-2753.

Pampariene, I., V. Veikutis, V. Oberauskas, J. Zymantiene, R. Zelvyte, A. Stankevicius, D. Marciulionyte, and P. Palevicius. 2016. 1908. thermography based inflammation monitoring of udder state in dairy cows: Sensitivity and diagnostic priorities comparing with routine california mastitis test. Journal of Vibroengineering. 18(1)

Paulrud, C.O. & Rasmussen, M.D., 2004. Teat thermography, hot or not. Proc. NMC 43rd Annual Meeting, Charlotte, NC, 159-168

Paulrud, C. O., S. Clausen, P. E. Andersen, and M. D. Rasmussen. 2005. Infrared thermography and ultrasonography to indirectly monitor the influence of liner type and overmilking on teat tissue recovery. Acta Vet. Scand. 46: 137-147.

Pezeshki, A., P. Stordeur, H. Wallemacq, F. Schynts, M. Stevens, P. Boutet, L. J. Peelman, B. De Spiegeleer, L. Duchateau, and F. Bureau. 2011. Variation of inflammatory dynamics and mediators in primiparous cows after intramammary challenge with escherichia coli. Vet. Res. 42: 15.

Pitkälä, A., M. Haveri, S. Pyörälä, V. Myllys, and T. Honkanen-Buzalski. 2004. Bovine mastitis in finland 2001—prevalence, distribution of bacteria, and antimicrobial resistance. J. Dairy Sci. 87: 2433-2441.

Polat, B., A. Colak, M. Cengiz, L. Yanmaz, H. Oral, A. Bastan, S. Kaya, and A. Hayirli. 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. J. Dairy Sci. 93: 3525-3532.

SAS Institute. 2010. SAS/STAT Software, Release 9.3, SAS Institute Inc. Cary, NC.

Speakman, J., and S. Ward. 1998. Infrared thermography: Principles and applications. Zool-Anal. Complex Sy. 101: 224-232.

Wilson, D. J., R. N. Gonzalez, and H. H. Das. 1997. Bovine mastitis pathogens in new york and pennsylvania: Prevalence and effects on somatic cell count and milk production. J. Dairy Sci. 80: 2592-259

5 Infrared thermography as a tool to detect hoof lesions in sheep

D.T. Byrne[1 3], D.P. Berry[1], H. Esmonde[3], F. McGovern[2], and N. McHugh[1]

[1]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland, P61 P302.

[2]Animal & Grassland Research and Innovation Centre, Teagasc, Mellows Campus, Athenry, Co. Galway, Ireland, H65 R718;

[3]Dublin City University, Glasnevin, Dublin 9, Co. Dublin, Ireland, D09 Y5N0

This is a pre-copyedited, author-produced version of an article accepted for publication in the journal of translational animal science following peer review. The version of record Byrne DT, Berry DP, Esmonde H, McGovern F, Creighton P, McHugh N. Infrared thermography as a tool to detect hoof lesions in sheep. Translational Animal Science. 2018 Dec 8. is available online at: https://academic.oup.com/tas/advance-article/doi/10.1093/tas/txy132/5235605 (doi: 10.1093/tas/txy132)

5.1 Abstract

Lameness has a major negative impact on sheep production. The objective of the current study was to: i) quantify the repeatability of sheep hoof temperatures estimated using infrared thermography (IRT), ii) determine the relationship between ambient temperature, sheep hoof temperature, and sheep hoof health status, and iii) validate the use of IRT to detect infection in sheep hooves. Three experiments (a repeatability, exploratory and validation experiment) were conducted over 10 distinct non-consecutive days. In the repeatability experiment, 30 replicate thermal images were captured from each of the front and back hooves of nine ewes on a single day. In the exploratory experiment, hoof lesion scores, locomotion scores, and hoof thermal images were recorded every day from the same cohort of 18 healthy ewes in addition to a group of lame ewes, which ranged from one to nine ewes on each day. Hoof lesion and locomotion scores were blindly recorded by three independent operators. In the validation experiment, all of the same procedures from the exploratory experiment were applied to a new cohort of 40 ewes across two days. The maximum and average temperature of each hoof was extracted from the thermal images. Repeatability of IRT measurements was assessed by partitioning the variance due to ewe and error using mixed models. The relationship between ambient temperature, hoof temperature, and hoof health status was quantified using mixed models. The percentage of hooves correctly classified as healthy (i.e., specificity) and infected (i.e., sensitivity) was calculated for a range of temperature thresholds. Results showed that a small to moderate proportion of the IRT-estimated temperature variability in a given hoof was due to error (1.6 to 20.7%). A large temperature difference (8.5 °C) between healthy and infected hooves was also detected. The maximum temperature of infected hooves was unaffected by ambient temperature ($P > 0.05$), while the temperature of healthy hooves was associated with ambient temperature. The best sensitivity (92%) and specificity (91%) results in the exploratory experiment were observed when infected hooves were defined as having a

maximum hoof temperature ≥ 9 °C above the average of the five coldest hooves in the flock on that day. When the same threshold was applied to the validation dataset a sensitivity of 77% and specificity of 78% was achieved, indicating that IRT could have the potential to detect infection in sheep hooves.

Key words: sheep, infrared thermography, lameness

5.2 Introduction

Lameness has a major impact on the welfare and profitability of sheep production (Hickford et al., 2005; Fitzpatrick et al., 2006), with every 10% increase in prevalence costing an additional €2.40 per ewe in treatment costs alone (Bohan et al., 2018a). One of the most common causes of lameness in sheep is footrot (Conington et al., 2010); the average prevalence ranges from 0.4 to 23.3% across sheep production systems (Conington et al., 2010; Gelasakis et al., 2013). Currently, the gold standard for recording footrot requires sheep to be turned over and each hoof visually assessed which is labour intensive and difficult to implement across large numbers of flocks.

Infrared thermography (IRT) is a non-invasive technology that can estimate the temperature of an object based on the radiating energy (Luzi et al., 2013). Previous research successfully used IRT to detect lameness in cattle (Alsaaod et al., 2015), respiratory disease in calves (Schaefer et al., 2012), and breast cancer in humans (Milosevic et al., 2014). In sheep, Talukder et al. (2015) used 15 rams to demonstrate an association between hoof temperature and hoof lesions, but did not test the ability of IRT to diagnose individual hooves. Other studies have shown how ambient temperature can affect the temperature of an anatomical region, and if left unaccounted for, the diagnostic

ability of IRT (Berry et al., 2003; Church et al., 2014). No study however, has investigated the association between environmental factors and hoof temperature in sheep.

The objective of the present study, therefore, was to investigate and validate the feasibility of using IRT to detect lameness in sheep while taking cognizance of prevailing environmental factors. Results from the present study could aid in the development of an automated lameness detection tool for sheep and would facilitate large quantities of lameness data to be gathered for accurate genetic evaluations and the inter- and intra-herd temporal benchmark.

5.3 Materials and Methods

A series of experiments were undertaken in Athenry Research Centre, Teagasc, Athenry, Co. Galway, Ireland (53.287611 latitude, -8.767840 longitude). All procedures were conducted under approval from the Teagasc Animal Ethics Committee on experimental animal use (TAEC141-2017) in accordance with the Cruelty to Animals Act 1876 and the European Communities Regulations, 1994.

To investigate the relationship between hoof health status and hoof temperature, thermal images, individual hoof lesion scores, and locomotion scores were recorded from 103 purebred Texel, Suffolk, Belclare, and crossbred ewes for 10 unique days between May and October 2017 (Figure 5.1). On the morning of each experimental day, all sheep were blindly locomotion scored by three independent operators on a scale ranging from 0 to 3, where 0 = sound and 3 = severely lame (Angell et al., 2015). Additionally, each sheep was turned and the same three operators scored each hoof for lesions on a scale ranging from 0 to 4, where 0 = healthy and 4 = severe footrot (Table 5.1; Conington et al. (2008)).

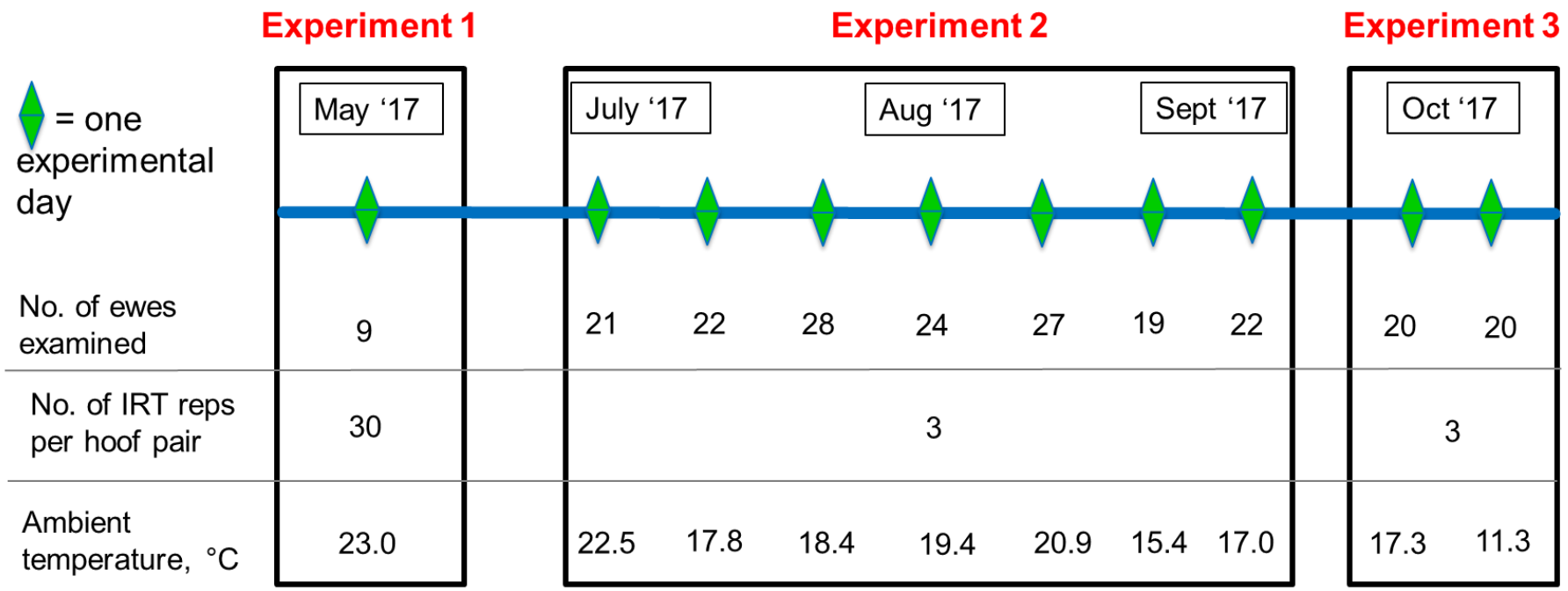


Figure 5.1 Timeline for all three experiments in the present study is shown, with details of the ambient temperature, number of thermal image replicates captured per hoof pair (No. of IRT reps per hoof pair), and the total number of unique ewes (No. of ewes examined) used in each experiment.

Table 5.1 Number of hooves (No. of hooves) and their percentage of the total dataset (% of total) from the exploratory (Exploratory exp.) and validation (Validation exp.) experiments, categorized by type of hoof lesion scoring scale and hoof score (Score). Both the categorical and binary hoof lesion scales were derived from the hoof lesion scale from Conington et al. (2008).

			Exploratory exp.		Validation exp.	
Type of hoof lesion scale	Score	Definition*	No. of hooves	% of total	No. of hooves	% of total
Hoof lesion scale by Conington et al. (2008)	0	Healthy hoof	577	94.44	119	94.44
	1	Mild IDD	26	4.26	6	4.76
	2	Extensive IDD	6	0.98	1	0.79
	3	Severe IDD/footrot	2	0.33	0	0.00
	4	Severe footrot	0	0.00	0	0.00
Categorical hoof lesion scale	0	Healthy hoof	577	93.82	119	93.70
	1	Mild IDD	26	4.23	6	4.72
	2	Extensive IDD or worse	12	1.95	2	1.57
Binary hoof lesion scale	0	Healthy hoof	577	92.03	119	89.47
	1	Mild IDD or worse	50	7.97	14	10.53

*IDD = Inter digital dermatitis

The locomotion scores for each ewe and the hoof lesion scores for each hoof were averaged across all three operators. A number of edits were imposed on the recorded hoof lesion scores to ensure that erroneous hoof measurements were removed prior to analysis. Hooves that received a healthy hoof score (i.e., hoof lesion score = 0 from all operators) which was preceded or proceeded by a hoof lesion score ≥ 1 on any other experimental day were removed from the dataset; data from six hooves were removed. Since a small number of category 3 (n = 2) and no category 4 hoof lesions were observed (Table 5.1), a categorical hoof lesion score was created whereby, any hoof with a hoof score averaged across the three operators of >2 was set equal to two. Any remaining records where all three independent operators did not universally agree on a hoof lesion score of 0, 1, or 2 were deleted; data from 135 hooves were discarded. Similar to previous research in cattle (Alsaaod and Büscher, 2012; Stokes et al., 2012), a binary hoof score was created whereby hooves were classified as either healthy (universal agreement between all three operators on a hoof score of 0) or infected (hoof score averaged across all three operators was ≥1). Any hoof that received a hoof lesion score of zero from one operator and one from another operator was deleted; data from 125 hooves were discarded.

All animals participating in the study were allowed to rest in a paddock close to the shed for one hour after scoring. Following this, they were moved into the shed for a 30 minute acclimatization period before imaging. All thermal images were captured using a FLIR T430sc thermal camera (FLIR Systems Inc., Stockholm, Sweden). The spectral range of the camera was between 7.5 and 13 µm. The camera resolution was 320x240, the thermal sensitivity was <0.03°C, and the accuracy was ± 2°C. All images were captured by the same operator in the same shed which did not receive direct sunlight. To obtain a palmar view of both front hooves, the camera was placed to the right hand side of the animal pointing towards the front hooves. To obtain a plantar view of both back hooves, the camera was placed behind the animal at an angle that was just off parallel to the

median plane. All hoof images were captured at a distance of 0.7 m. The number of images captured per hoof pair (i.e., front hooves or back hooves) varied from 3 to 30 between experiments (described below). Care was taken to ensure all images were in focus. Ambient temperature and humidity of the shed were recorded every minute during each experimental day using a Lascar EL-USB-2 data logger (Lascar Electronics, Whiteparish, UK).

Image analysis and temperature extraction was undertaken using the Thermovision LabVIEW toolkit 3.3 (FLIR Systems Inc., Stockholm, Sweden) using the procedures previously outlined by Byrne et al. (2017). All image parameters (i.e., emissivity, ambient temperature, humidity, object distance, and reflected temperature) were adjusted in each image before analysis. Emissivity in all images was set to 0.98. Ambient temperature and humidity data varied between images and the respective values were taken from the Lascar EL-USB-2 data logger (Lascar Electronics). A freehand border was drawn around all hooves to extract the required pixels and remove any background information. The freehand border encompassed the posterior face of each hoof from below the coronary band to above the dew claws (Figure 5.2). The maximum and average temperature of each hoof was recorded. A minimum of three thermal image replicates was captured on every hoof and respective temperature values from the replicate images were averaged as recommended by Byrne et al. (2017) in their study about cattle.

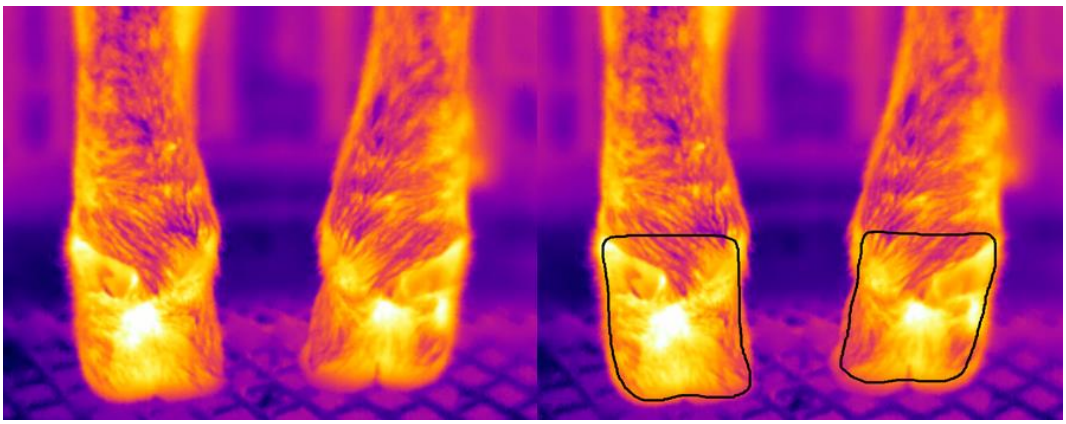


Figure 5.2 A thermal image of a pair of hooves before (left) and after (right) a freehand border was applied for temperature extraction is shown. The freehand border encompassed the posterior face of each hoof from below the coronary band to above the dew claws.

A series of experiments were undertaken with the objectives of i) quantifying the repeatability of IRT-estimated sheep hoof temperatures, ii) investigating the inter-relationship between ambient temperature, hoof temperature and hoof health status, and iii) validating the use of IRT as a tool to detect hoof infection.

5.3.1 Experiment 1 - repeatability experiment

The objective of this experiment was to quantify the repeatability of sheep hoof temperature estimated by IRT and to assess the number of replicates required to achieve a certain precision of hoof temperature using IRT. Precision was defined as the largest expected difference (95% of the time) between (the average of) the measured temperature(s) and the average of 30 replicates of the measurement. To measure the repeatability of sheep hoof IRT, a set of 30 consecutive thermal image replicates of the palmar view of the front hooves, followed by 30 replicates of the plantar view of the back hooves, was taken of each ewe by a single operator. Each image took approximately ten seconds to capture.

5.3.1.1 Statistical analyses.

The between-ewe and error variances for both maximum and average temperature of each of the four hooves was estimated separately using mixed models in ASReml (Gilmour et al., 2009) with ewe included as a random effect. A log likelihood ratio test was performed to test if the model fit the data better with or without the inclusion of the random ewe effect. The proportion of total variance explained by ewe (H_{ewe}) was calculated as the between-ewe variance divided by the sum of the error and between-ewe variance. The coefficient of variation (CV) was calculated for the maximum and average hoof temperature of each anatomical area separately as the respective SD divided by the mean. The number of images required to gain a certain precision (P_n) with a 95% CI was calculated as follows:

$$P_n = 1.96 \times \sqrt{\frac{\sigma_e^2}{n_{\in(1,30)}}}$$

where *n* was the image count (sample size) which varied from 1 to 30 and σ_e^2 was the error variance. The stability of the temperature measurement from 30 images across time was investigated; the correlation between the first replicate measurement and all other replicate measurements (i.e., 2 to 30) was calculated for the maximum and average temperature separately for each of the four hooves using PROC CORR of SAS (SAS Institute, 2010).

5.3.2 Experiment 2 - exploratory experiment

The objective of the exploratory experiment was to investigate the relationship between ambient temperature, hoof temperature, and hoof health status. Prior to the commencement of the experiment, a cohort of 18 purebred Texel, Suffolk, and Belclare ewes was randomly selected as a control group and was locomotion scored on each experimental day. Between July and September 2017, seven experimental days were conducted. In addition to the control group, a further group of 120 ewes was locomotion scored on each experimental day (Angell et al., 2015) and any ewe with a locomotion score

of one or greater was assigned to a lame group; the size of the lame group varied by experimental day from one to nine ewes. Subsequently, on each experimental day, both the control group (n = 18) and a lame group (n = 1 to 9) were scored for hoof lesions (Conington et al., 2008) by all three operators as per the experimental procedure outlined previously. A set of three thermal image replicates of the palmar face of the front hooves, followed by three thermal image replicates of the plantar view of the back hooves, was captured of every animal within both the control and lame groups (Figure 5.2). Data from the control group and the lame group were combined for analyses.

5.3.2.1 Statistical analyses.

To investigate if temperature differences between hooves with different levels of infection existed, a linear mixed model was performed in SAS using PROC MIXED (SAS Institute, 2010), with either maximum and average hoof temperature as the dependent variable and categorical hoof lesion score (score 0 to 2) as the independent variable. Ewe was included as a random effect. Anatomical area (i.e., left front hoof, right front hoof, left hind hoof and right hind hoof) was nested within the combination of ewe and date, which was included as a repeated effect. The association between categorical hoof lesion score and hoof temperature was also assessed using the Spearman rank correlation coefficient which was calculated in SAS using PROC CORR.

To quantify the association between hoof health status, ambient temperature and hoof temperature, the binary hoof score scale (i.e., 0 = healthy and 1 = infected) was used as a small number of records received a categorical hoof lesion score of two (n = 12;

Table 5.1). A multiple regression analysis was performed in SAS using PROC MIXED (SAS Institute, 2010), whereby either maximum and average hoof temperature was the dependent variable and binary hoof lesion score (0 = healthy hoof, 1 = infected hoof), ambient temperature, and the interaction between ambient temperature and binary hoof lesion score were all included as fixed effects. Anatomical area (i.e., left front hoof, right front hoof, left hind hoof and right hind hoof) was nested within the combination of ewe and date, which was included as a repeated effect. Fit statistics as well as the regression coefficients were calculated.

To investigate whether hoof temperatures estimated by IRT could be useful to differentiate infected (i.e., hoof score averaged across operators of ≥1 (Conington et al., 2008), which is equivalent to a binary hoof score = 1) from healthy (i.e., binary hoof score of 0) hooves, four distinct hoof temperature variables were considered: i) average hoof temperature, ii) maximum hoof temperature, iii) the difference between the average temperature of the hoof in question and the average of the five coldest average hoof temperatures on that day (i.e., average daily baseline), and iv) the difference between the maximum temperature of the hoof in question and the average of the five coldest maximum hoof temperatures on that day (i.e., maximum daily baseline). Numerous temperature thresholds were applied to each of these four temperature metrics (e.g., thresholds were tested at 1 °C intervals from 28 to 35 °C for maximum hoof temperatures), which facilitated the classification of each hoof as either infected or healthy by IRT. If a hoof temperature was above the given threshold then the hoof was diagnosed as infected, otherwise the hoof was considered healthy. This classification of hooves was compared to the actual presence or absence of hoof lesions, so the sensitivity and specificity could be calculated for each temperature threshold. Throughout the current study, the ideal threshold was considered to be the one that achieved a balanced sensitivity and specificity, as results can then be readily compared with past and future studies (Greiner et al., 2000).

5.3.3 Experiment 3 - validation experiment

The objective of this experiment was to investigate whether the thresholds defined in the exploratory experiment could be used to accurately diagnose hoof health in a separate cohort of ewes. A cohort of 40 crossbred ewes (i.e., the validation dataset) was scored for locomotion and hoof lesions on the 5th and 6th of October from the same location and using the same experimental procedure as in the exploratory experiment. A set of three thermal image replicates of the palmar face of the front hooves, followed by three thermal image replicates of the plantar view of the back hooves, was captured from all animals (Figure 5.2).

5.3.3.1 Statistical analyses.

The association between categorical hoof lesion score (score 0 to 2) and hoof temperature was also assessed using the Spearman rank correlation coefficient which was calculated in SAS using PROC CORR. To validate the ability of IRT to diagnose healthy from infected hooves (i.e., score 0 from 1), the optimum temperature thresholds as defined in the exploratory experiment for each temperature variable (i.e., maximum hoof temperature, average hoof temperature, the difference between the maximum hoof temperature and the maximum daily baseline, and the difference between the average hoof temperature and the average daily baseline) were applied to the validation dataset and the resulting sensitivity and specificity were calculated for each. Similar to the exploratory experiment, any hoof with a temperature above a given threshold was diagnosed as infected, whereas a hoof temperature below the given threshold was diagnosed as healthy. The classification of hooves as either healthy or infected using IRT data was tested against the actual presence or absence of lesions which facilitated the sensitivity and specificity to be calculated.

5.4 Results

Across the entire experiment period, ambient temperature ranged from 11.3 to 23.0°C, while relative humidity ranged from 53.0 to 88.9%. Individual maximum hoof temperatures ranged from 17.0 to 38.4°C, which was similar to (Gloster et al., 2011) who examined the hoof temperature of cattle at a wide range of ambient temperatures. The average temperature of each hoof averaged across the flock ranged from 23.13 (left hind hoof) to 24.18°C (left front hoof), while the maximum temperature of each hoof averaged across the flock ranged from 29.61 (left hind hoof) to 30.40°C (left front hoof). The number of infected hooves (i.e., all 3 operators agree upon a score of ≥1) on each day of the experiment ranged from 1 to 15. The percentage of lame ewes (i.e., had one hoof where all 3 operators agreed upon a score of ≥1 on the hoof lesion scale by Conington et al. (2008)) recorded each day ranged from 4.8 to 47.4%. Of the 18 ewes selected for the control group in the exploratory experiment, 15 remained free from hoof lesions throughout the entire experiment.

5.4.1 Experiment 1 - repeatability experiment

The between-ewe and error variances were greater for the maximum temperature in comparison to average temperature across the four anatomical regions assessed (Table 5.2). The greatest between-ewe variance was observed in the right front hoof while the lowest between-ewe variance was observed in the right hind hoof for both average and maximum hoof temperatures. A lower error variance was noted for the front hooves in comparison to the back hooves (Table 5.2).

Table 5.2 The ewe variance, error variance, percentage of total variance due to the ewe (H_{ewe}), for the maximum and average temperatures of all four hooves (Right front, left front, right hind and left hind hoof) as well as the precision achieved with three images. Precision was defined as the largest expected difference (95% of the time) between (the average of) the measured temperature(s) and the gold standard, where the gold standard was the average of 30 measurements.

Variable	Quantity	Right front	Left front	Right Hind	Left Hind
Maximum	Ewe variance (SE) , $°C^2$	19.92(9.94)	14.29(7.13)	3.71(1.86)	15.79(7.90)
	Error variance (SE) , $°C^2$	0.49(0.04)	0.40(0.04)	0.97(0.08)	0.96(0.08)
	H_{ewe} , %	97.60	97.26	79.31	94.26
	Precision with 3 images, °C	0.79	0.72	1.11	1.11
Average	Ewe variance (SE) , $°C^2$	8.00(3.99)	5.24(2.61)	2.05(1.02)	4.85(2.43)
	Error variance (SE) , $°C^2$	0.13(0.01)	0.13(0.01)	0.15(0.01)	0.24(0.02)
	H_{ewe} , %	98.36	97.62	93.21	95.20
	Precision with 3 images, °C	0.41	0.41	0.44	0.55

A large proportion of the variation in both the average and maximum hoof temperature was due to the ewe (79.31 to 98.36%; Table 5.2). The CV tended to be greater for maximum hoof temperatures (7.60 to 15.15%) in comparison to the respective average hoof temperatures (6.38 to 11.13%). Maximum temperature averaged across three replicate images was expected to lie within ± 0.79 °C of the average of 30 replicates (i.e., the precision was 0.79 °C). The precision achieved when an average temperature was extracted from the same three images was ± 0.41 °C.

Across both maximum and average hoof temperatures, the correlation between the first replicate temperature measurement and all other replicate temperature measurements tended to weaken as the interval between the replicates compared lengthened (Figure 5.3). Strong to moderate correlations existed between the first replicate and all other replicates of maximum hoof temperature, which ranged from 0.76 (replicate 1 and replicate 24; right

hind hoof) to 0.99 (replicate 1 and replicate 13; left front hoof) across all hooves. Similarly, the correlation between the first replicate and all other replicates of average hoof temperature was strong and ranged from 0.86 (replicate 1 and replicate 29; left hind hoof) to 0.99 (replicate 1 and replicate 22; left front hoof) across all hooves.

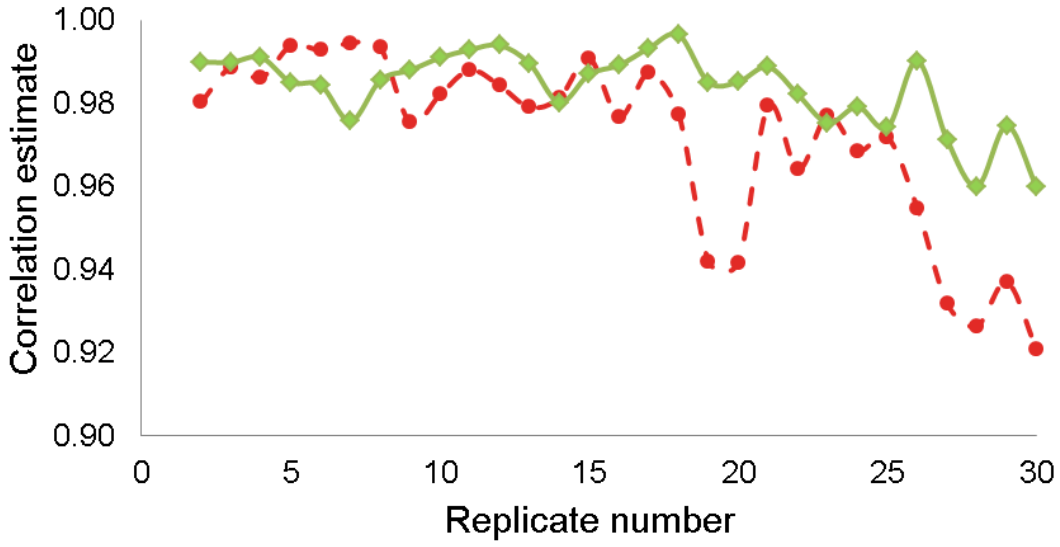


Figure 5.3 Correlation estimates between the right front hoof temperatures (average (♦) and maximum (●) temperature) recorded from the first thermal images of nine ewes and each of the subsequent replicate images (i.e., 2 to 30) are shown.

5.4.2 Experiment 2 - exploratory experiment

The frequency of each hoof lesion score is given in Table 5.1. The maximum hoof temperature averaged across healthy hooves (i.e., score = 0) was lower (26.28 °C) in comparison to that of hooves with mild inter-digital dermatitis (33.81 °C; $P<0.05$) (i.e., categorical hoof lesion score = 1), but this value did not differ when infection increased from mild to extensive inter-digital dermatitis (i.e., categorical hoof lesion score = 2) ($P > 0.05$). A box and whisker plot of the relationship between categorical hoof lesion score and maximum hoof temperature is shown in Figure 5.4. The mean ± SE average hoof temperature for hooves with no lesions (i.e., score = 0), mild inter-digital dermatitis (i.e.,

score = 1), and extensive inter-digital dermatitis (i.e., score = 2) was 21.20 ± 0.20, 25.98 ± 0.51, and 27.34 ± 0.76 °C, respectively. The average hoof temperature differed between all hoof lesion categories (P<0.05). The Spearman rank correlation co-efficient between the categorical hoof lesion score (score 0 to 2) and hoof temperature was 0.39 (P<0.001) when the maximum hoof temperature was examined and 0.38 (P<0.0001) when the average hoof temperature was used.

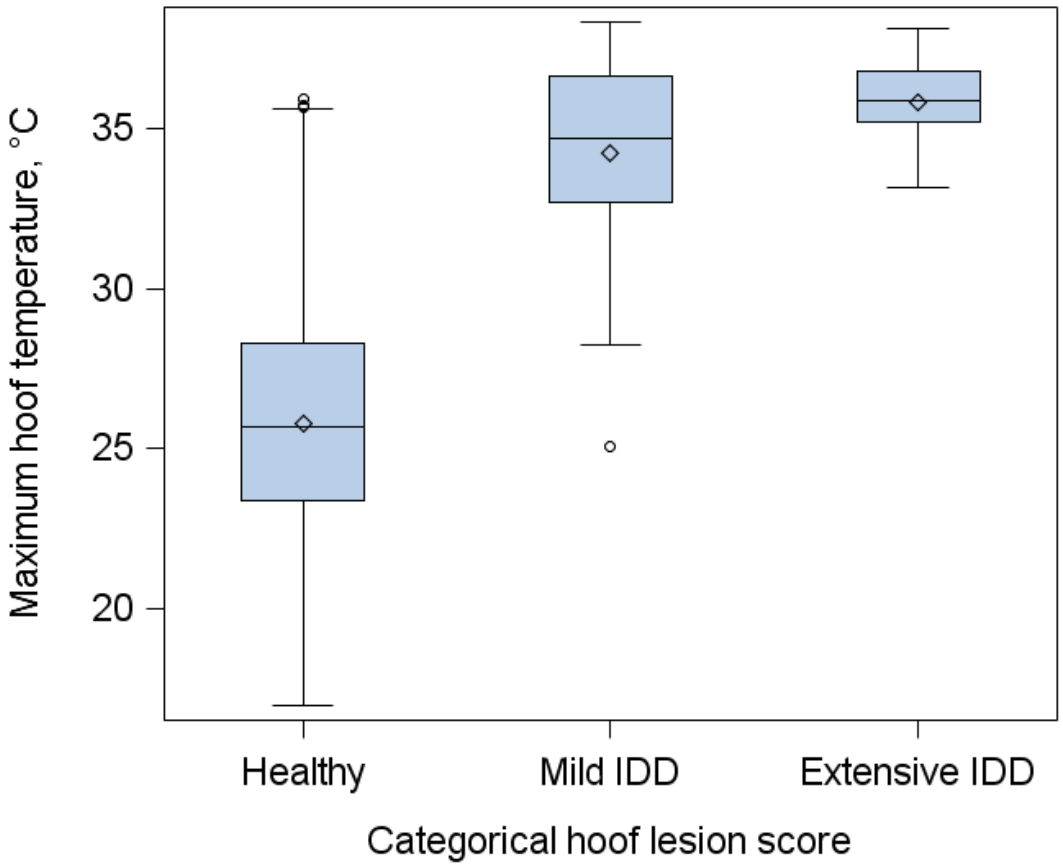


Figure 5.4 A box and whisker plot of the relationship between the maximum hoof temperature and the stages of inter-digital dermatitis (IDD) (i.e., categorical hoof lesion score), where the mean (◊), median (—), interquartile range (IQR) (■), greatest (┬) and lowest (┴) values within 1.5*IQR of the IQR, and outliers (○) of the maximum hoof temperature are shown.

When a multiple regression analysis was performed with average hoof temperature as the dependent variable and ambient temperature, binary hoof score (i.e., 0 = healthy hooves and 1 = infected hooves), and their interaction as the independent variables, all three independent variables were observed to be significant ($P < 0.001$). The regression coefficient ± SE associated with ambient temperature was greater for healthy hooves (i.e., binary hoof score of 0; 0.80 ± 0.03 °C) in comparison to infected hooves (i.e., binary hoof score of 1; 0.43 ± 0.15 °C). The least square mean ± SE average temperature differed between healthy (20.98 ± 0.07 °C) and infected (26.35 ± 0.27 °C) hooves ($P < 0.05$). When a multiple regression analysis was performed with maximum hoof temperature as the dependent variable, ambient temperature, binary hoof score and their interaction were observed to be significant fixed effects ($P < 0.001$). While the maximum temperature of healthy hooves increased with ambient temperature (regression co-efficient = 0.94 ± 0.05 °C; $P < 0.01$), the maximum temperature of infected hooves did not differ with ambient temperature ($P>0.05$). An illustration of the relationship between ambient temperature, maximum hoof temperature and binary hoof score is shown in Figure 5.5. The least square mean ± SE maximum temperature of healthy (i.e., binary hoof score of 0) and infected (i.e., binary hoof score of 1) hooves was significantly different ($P < 0.05$) at 25.87 ± 0.12 and 34.36 ± 0.46 °C, respectively. The maximum or average hoof temperature did not differ by relative humidity ($P > 0.05$).

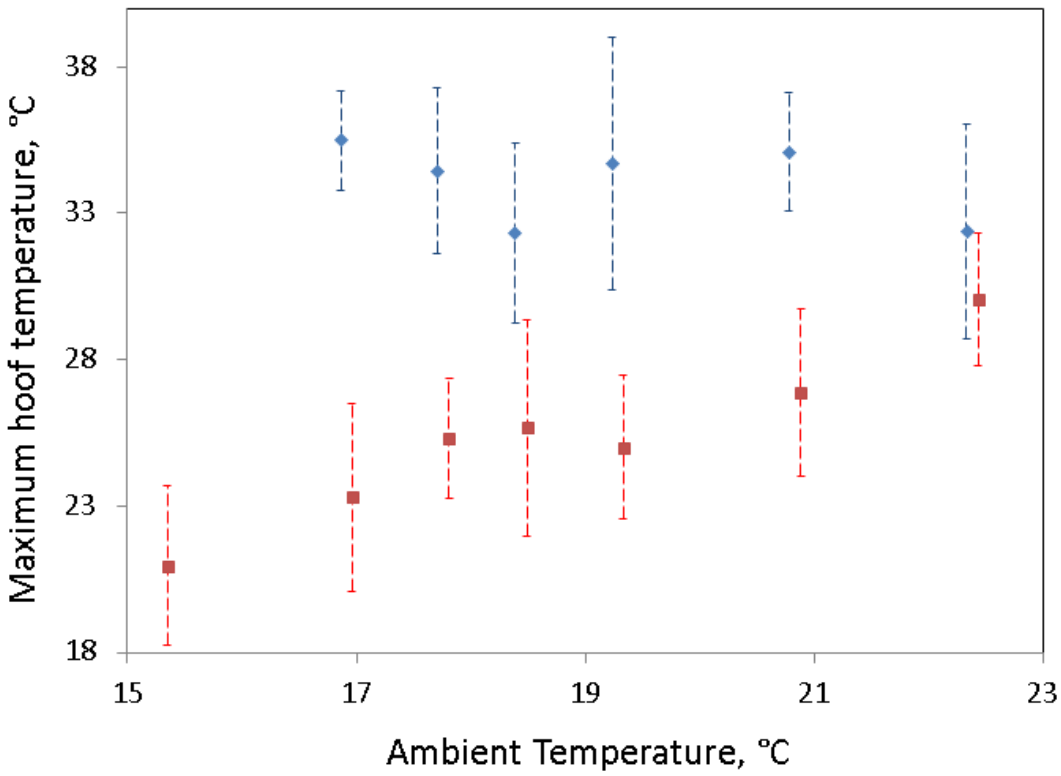


Figure 5.5 Mean and standard deviation of maximum temperatures for infected (♦) and healthy (■) hooves at various levels of ambient temperature.

When temperature thresholds ranging from 28 to 35 °C in 1 °C intervals were applied to the raw maximum hoof temperature, the optimum threshold (i.e., the threshold at which sensitivity and specificity were balanced) was observed at 31 °C, with a sensitivity of 92% and a specificity of 91% (Figure 5.6). The optimum threshold for average hoof temperatures was 24 °C, achieving a sensitivity of 86% and specificity of 88%. When the difference between maximum hoof temperatures and the respective daily baseline was investigated, a threshold of 9 °C above the daily baseline achieved the optimum sensitivity (92%) and specificity (91%). A threshold of 5 °C above the daily baseline of average hoof temperatures achieved the optimum results (i.e., sensitivity of 90% and specificity of 89%) for the difference between average hoof temperatures and the respective daily baseline.

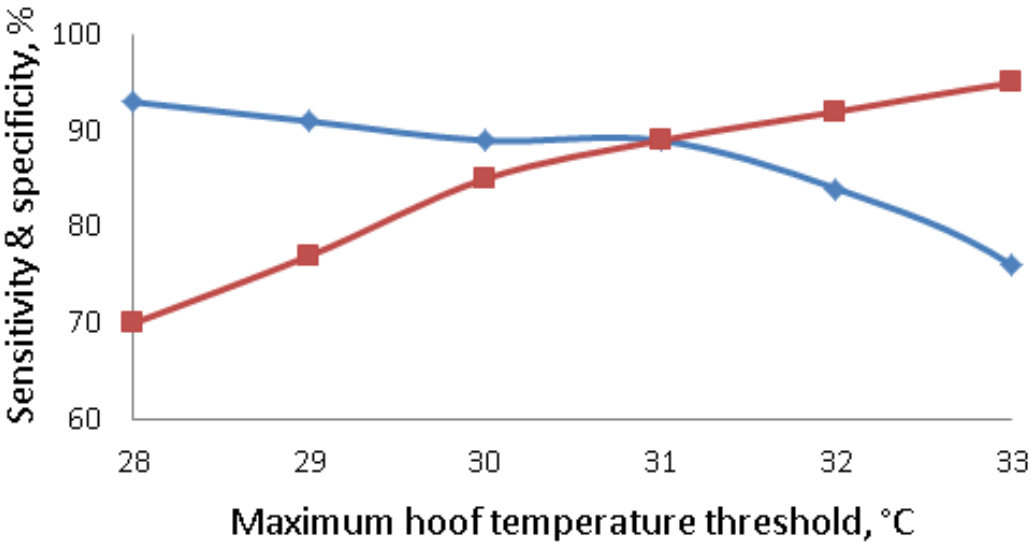


Figure 5.6 The percentage of hooves which were correctly classified as healthy (i.e., specificity (■)) or infected (i.e., sensitivity (♦)) when the threshold for maximum hoof temperature varied from 28 to 33°C. Hooves with a maximum temperature above the threshold were considered infected, while all others were considered healthy.

5.4.3 Experiment 3 - validation experiment

As in the exploratory experiment, a positive Spearman rank correlation coefficient was observed between categorical hoof lesion score and hoof temperatures ($P > 0.001$) ($\rho = 0.36$ for the maximum hoof temperature and $\rho = 0.30$ for the average hoof temperature). The percentage of lame hooves within the validation experiment is shown in Table 5.1. When the optimum threshold for the maximum hoof temperature from the exploratory experiment (i.e., 31 °C) was applied to the validation dataset, a sensitivity of 46% and specificity of 96% was achieved. A similar deterioration of predictive ability was observed when the optimum average hoof temperature threshold (i.e., 24 °C) from the exploratory experiment was applied to the current dataset, where a sensitivity of 15% and specificity of 99% was achieved. The optimum threshold for the difference between maximum hoof temperatures and the respective daily baseline (i.e., 9 °C) achieved a balanced but lower sensitivity (77%) and specificity (78%) compared to the exploratory experiment, indicating

that IRT could be a potential solution to detect infection in sheep hooves. When a threshold of 5°C was applied to the difference of average hoof temperatures and the respective daily baseline, a sensitivity of 69% and a specificity of 80% was achieved.

5.5 Discussion

Lameness is one of the leading causes of morbidity in sheep (Dohoo et al., 1985). The best method for reducing lameness prevalence involves early detection and subsequent treatment (Kaler and Green, 2008). The current gold standard for recording lameness (i.e., hoof lesion scoring) is difficult to implement across large flocks due to the labour requirement. The objective of the present study was not only to investigate whether IRT could be used as a tool to detect lameness in sheep but also to determine the factors which must be accounted for to improve the diagnostic capabilities of IRT. Results indicate that when a single image of each hoof is captured under optimal conditions, IRT could indeed be a valuable hoof infection detection tool.

5.5.1 Repeatability of sheep hoof IRT

Using experiments on a population of 15 lactating dairy cows, Byrne et al. (2017) reported that the temperature difference between replicate thermal images of the udder and eye can be larger than the temperature difference between healthy and infected in these anatomical areas; no study investigated the repeatability of IRT measurements from sheep hooves. Many studies in cattle and sheep have relied upon on a single temperature measurement (Rainwater-Lovett et al., 2009; Alsaaod and Büscher, 2012; Talukder et al., 2015), and so did not investigate the repeatability of IRT measurements of hooves. Results from the present study suggest that sheep hoof temperature measurements made using IRT are repeatable as depicted by the small to moderate percentage (1.6 to 20.7%) of the variability in temperature being due to unknown factors encapsulated within the error

variation. A slightly greater variation between replicate images (error variance) was detected in the present study relative to the cattle based study of Byrne et al. (2017). A larger error variance generally means that more replicate images are required, but a single image of each hoof may actually suffice to detect disease if infection causes a large shift in temperature. In the present study, the average difference between the maximum temperature of healthy and infected hooves was much larger (8.5°C) than the mean temperature difference between two thermal hoof image replicates (ranging from 1.2 to 1.9°C across anatomical areas). Nonetheless, consistent with the recommendations of Byrne et al (2017), the present study used the average of three replicate measures for analysis to minimize the error variation. To evaluate alternative methodologies, the maximum temperature of the three replicate images was also taken which resulted in an increased sensitivity but at a cost of eroding the specificity; taking the minimum of the three thermal image replicates had the opposite effect (i.e., decreased sensitivity and increased specificity). If the aim was to increase sensitivity to the detriment of specificity, or vice versa, then altering the temperature threshold between healthy and infected hooves would be a more reliable option to achieve the same goal. When a single image was randomly chosen for each hoof and used for analysis, the sensitivity and specificity results varied by ± 2%, indicating that replicate measurements did not improve the overall diagnostic ability, at least in the present study.

5.5.2 Factors associated with hoof temperatures

Hoof health status and ambient temperature were two major factors which were associated with hoof temperature. While some studies have documented a temperature difference between healthy and infected hooves of 1.4°C in sheep and between 0.4 and 7.9°C in cattle (Alsaaod and Büscher, 2012; Stokes et al., 2012; Talukder et al., 2015), no study in ruminants investigated the change in hoof temperature with increasing severity of hoof lesion score. Results herein revealed that when the severity of infection increases (i.e.,

from mild to extensive inter-digital dermatitis), the maximum temperature of the hoof does not increase but instead thermal energy spreads through the hoof thereby increasing the average hoof temperature. In sheep, it may be important for IRT to differentiate between mild inter-digital dermatitis (i.e., a hoof score of 1) and extensive inter-digital dermatitis (i.e., a hoof score of 2) as the appropriate treatment for each score is different. Alternatively, if IRT is used only as a preliminary screening tool or as a means of collecting phenotypes for genetic evaluations, then IRT may only need to discern between healthy (i.e., score = 0) and infected hooves (i.e., score ≥ 1).

A large difference of hoof temperature averaged across the flock examined by Talukder et al. (2015) and the present study was observed, despite differences due to ambient temperature being mitigated against. Talukder et al. (2015) extracted the maximum temperature from each healthy hoof of nine rams and calculated the average to be 35.7°C (ambient temperature = 14.3°C); in the present study, the equivalent value from a flock of healthy hooves subject to a similar ambient temperature (i.e., 15.4°C) was 21.0°C. The thermal image view was a key difference between the present study and that of Talukder et al. (2015); the present study captured the palmar/ planter face of each hoof as the animal was standing (Figure 5.2), while Talukder et al. (2015) restrained animals in a standing position, lifted each hoof, manually separated the toes, and captured images of the inter-digital space. The inter-digital space is more enclosed and would be subject to friction between the toes, therefore, it should be warmer than the posterior face of the hoof. Additionally, lifting each hoof may cause stress to the animal which can cause a change in temperature (Stewart et al., 2008; Valera et al., 2012). Therefore, future studies should ideally capture the posterior face of each hoof while the animal is standing to reduce labor and stress to the animal.

Previous studies have noted the coefficient of determination between ambient temperature and the hoof temperature in cattle to range from 10 to 92% (Alsaaod and Büscher, 2012; Stokes et al., 2012). Results from the present study suggest that, unlike cattle hoof temperatures, the relationship between sheep hoof and ambient temperature is actually dependent on hoof health status, indicating that studies using IRT to detect disease should be conducted at multiple levels of ambient temperature. Martins et al. (2013) postulated that skin temperature is derived from internal blood flow which increases during infection; in the present study the maximum hoof temperature was derived from the same region as where the median artery splits in two (below the dew claws; Figure 5.2). Results from the present study suggest that when infection occurs, the median artery blood flow has a larger impact on the maximum hoof temperature than ambient temperature as the hoof is generating more thermal energy than can be absorbed from the environment. On the other hand, both infected and healthy average hoof temperatures were associated with ambient temperature which suggests that blood flow does not influence the entire cropped region of the hoof (i.e., the posterior face; Figure 5.2).

5.5.3 Diagnostic capabilities of IRT

While some studies have used hoof temperatures to identify hoof lesions in cattle (Alsaaod and Büscher, 2012; Stokes et al., 2012), no study used IRT to diagnose footrot in the individual hooves of sheep. Additionally, no study has conducted a true validation on the diagnostic capabilities of IRT in any species; instead, many studies conduct a single exploratory experiment (Alsaaod and Büscher, 2012; Stokes et al., 2012; Talukder et al., 2015). In the present study, some temperature metrics were not capable of differentiating healthy from infected hooves, while other temperature metrics showed that IRT has the potential to be a useful infection detection tool. The optimum threshold for the maximum hoof temperature in the exploratory experiment was 31 °C, where 92% of the truly infected hooves were identified as such by IRT data (i.e., sensitivity), while 91% of the truly

healthy hooves were also classified correctly (i.e., specificity). The sensitivity (92%) and specificity (91%) achieved with the maximum hoof temperature data were superior to most other comparable studies in cattle (Alsaaod and Büscher, 2012; Stokes et al., 2012; Alsaaod et al., 2015). When the same threshold was applied to a separate cohort of ewes (the validation dataset), the sensitivity deteriorated to 46%; this implies that the maximum hoof temperature cannot be used as an infection detection tool as random selection would achieve a sensitivity of 50%. This deterioration in accuracy clearly demonstrates the necessity of conducting a proper validation experiment, as without it, the maximum hoof temperature would appear to be a very viable solution to lameness detection. An increase in the temperature of healthy hooves (mean increase of 1.7 °C) and a reduction in the temperature of infected hooves (mean decrease of 1.6 °C) in the validation experiment contributed to the observed deterioration in sensitivity. The ideal temperature threshold to differentiate between healthy and infected hooves is one which can be applied to the temperature of any hoof in any scenario and correctly classify the health status of the hoof. Stokes et al. (2012) showed that the optimum threshold for differentiating healthy from infected hooves in cattle changes depending on the orientation and cleanliness of the hoof. As sheep hoof temperature can differ with ambient temperature, an optimal threshold which was defined when ambient temperature was 12 °C, may underperform when the ambient temperature rises to 20 °C. Therefore, to mitigate the influence of ambient temperature on a diagnosis, the temperature difference between a hoof and the respective daily baseline (average of the five coldest hooves from the flock on that day) was calculated. When hooves with a maximum temperature of 9 °C above the daily baseline were considered infected, the best results for differentiation ability were achieved (sensitivity of 92% and specificity of 91%), which again, were superior to most other comparable studies in cattle (Alsaaod and Büscher, 2012; Stokes et al., 2012; Alsaaod et al., 2015). As hoof temperatures were associated with ambient temperature, the daily

baseline for the maximum or average hoof temperature could also be derived from ambient temperature; the regression co-efficient and intercept to calculate the maximum hoof temperature daily baseline were 0.98 and 2.43 °C, respectively. The coefficient of determination between the maximum hoof temperature daily baseline derived from the coldest hooves in the flock and ambient temperature was 93%. When the threshold of 9 °C above the maximum daily baseline was applied to the validation dataset, a sensitivity of 77% and specificity of 78% was achieved. It may be possible that some of the incidents of lameness in the validation study were caused by metabolic or mechanical issues, which could have impacted the sensitivity and specificity. If a sensitivity of 77% and specificity of 78% were applied to a flock of 100 ewes with a 10% prevalence of footrot and assuming every lame ewe only has one lame hoof, then 3 of 10 lame hooves would be misclassified as healthy, whereas 86 of 390 healthy hooves would be misclassified as lame. Thresholds could be altered to reduce the number of misclassified healthy hooves (false positives) at the cost of increasing the number of misclassified lame hooves (false negatives), but this decision should be made based on actual prevalence and the costs associated with treatment or further screening of the hooves which are classified as lame (Greiner et al., 2000). To improve upon the predictive ability of IRT, temperature changes due to ancillary information (e.g., parity, breed, or feed efficiency) could also be investigated with a larger dataset of animals.

The present study suggests that colder ambient temperatures (<17°C) are best for detecting lameness in sheep as a greater temperature difference between healthy and infected hooves exists (Figure 5.5). Future work could compare the temperature of a hoof before and after the application of a cold stimulus; in theory infected hooves should remain hot while healthy hooves should cool down. This technique is known as dynamic thermography (Ohashi and Uchida, 2000).

5.6 Conclusion

The current study suggests that IRT has the potential to detect infection in sheep hooves, and since a large temperature difference between healthy and infected hooves was observed, a single image of each hoof may suffice to detect disease. The relationship between ambient temperature and hoof temperature was dependent on hoof health status and therefore, future studies which relate the temperature of an anatomical region to disease should be conducted at multiple levels of ambient temperature. Additionally, future work may consider using dynamic thermography techniques, as the temperature difference between sick and healthy hooves is greater when hooves are subjected to colder conditions.

5.7 Literature Cited

Alsaaod, M., and W. Büscher. 2012. Detection of hoof lesions using digital infrared thermography in dairy cows. J. Dairy Sci. 95: 735-742.

Alsaaod, M., A. L. Schaefer, W. Büscher, and A. Steiner. 2015. The role of infrared thermography as a non-invasive tool for the detection of lameness in cattle. Sensors. 15: 14513-14525.

Angell, J. W., P. J. Cripps, D. H. Grove-White, and J. S. Duncan. 2015. A practical tool for locomotion scoring in sheep: Reliability when used by veterinary surgeons and sheep farmers. Vet. Rec. 176(20): 521. doi: 10.1136/vr.102882 [doi]

Berry, R., A. Kennedy, S. Scott, B. Kyle, and A. Schaefer. 2003. Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection. Can. J. Anim. Sci. 83: 687-693.

Bohan, A., L. Shalloo, P. Creighton, T. M. Boland, and N. McHugh. 2018. Deriving economic values for the Irish national sheep breeding objectives using a bio-economic model. Submitted: Livest Sci.

Byrne, D., D. Berry, H. Esmonde, and N. McHugh. 2017. Temporal, spatial, inter-, and intra-cow repeatability of thermal imaging. J. Anim. Sci. 95: 970-979.

Church, J., P. Hegadoren, M. Paetkau, C. Miller, G. Regev-Shoshani, A. Schaefer, and K. Schwartzkopf-Genswein. 2014. Influence of environmental factors on infrared eye temperature measurements in cattle. Res. Vet. Sci. 96: 220-226.

Conington, J., B. Hosie, G. Nieuwhof, S. Bishop, and L. Bünger. 2008. Breeding for resistance to footrot–the use of hoof lesion scoring to quantify footrot in sheep. Vet. Res. Commun. 32(8): 583-589.

Conington, J., M. Speijers, A. Carson, S. Johnston, and S. Hanrahan. 2010. Foot health in sheep-prevalence of hoof lesions in UK and irish sheep. Advances in Animal Biosciences. 1(1): 340.

Dohoo, I. R., R. A. Curtis, and G. G. Finley. 1985. A survey of sheep diseases in canada. Can. J. Comp. Med. 49(3): 239-247.

Fitzpatrick, J., M. Scott, and A. Nolan. 2006. Assessment of pain and welfare in sheep. Small Ruminant Res. 62(1): 55-61.

Gelasakis, A., G. Oikonomou, R. Bicalho, G. Valergakis, G. FTHENAKIS, and G. Arsenos. 2013. Clinical characteristics of lameness and potential risk factors in intensive and semi-intensive dairy sheep flocks in greece. J. Hell Vet. Med. Soc. 64(2): 123-130.

Gilmour, A., B. Gogel, B. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0.VSN International Ltd, Hemel Hempstead, HP1 1ES, United Kingdom.

Gloster, J., K. Ebert, S. Gubbins, J. Bashiruddin, and D. J. Paton. 2011. Normal variation in thermal radiated temperature in cattle: Implications for foot-and-mouth disease detection. BMC Vet. Res. 7: 73-6148-7-73. doi: 10.1186/1746-6148-7-73

Greiner, M., D. Pfeiffer, and R. Smith. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev. Vet. Med. 45(1-2): 23-41.

Hickford, J., S. Davies, H. Zhou, and B. Gudex. 2005. A survey of the control and financial impact of footrot in the new zealand merino industry. PROCEEDINGS-NEW ZEALAND SOCIETY OF ANIMAL PRODUCTION,

Kaler, J., and L. Green. 2008. Recognition of lameness and decisions to catch for inspection among sheep farmers and specialists in GB. BMC Vet. Res. 4(1): 41.

Luzi, F., M. Mitchell, C. Nanni, and V. Redaelli. 2013. Thermography: Current status and advances in livestock animals and in veterinary medicine. . Brescia, Italy, Fondazione Iniziative Zooprofilattiche e Zootecniche.

Martins, R. F. S., T. do Prado Paim, C. de Abreu Cardoso, B. Stéfano Lima Dallago, C. B. de Melo, H. Louvandini, and C. McManus. 2013. Mastitis detection in sheep by infrared thermography. Res. Vet. Sci. 94: 722-724.

Milosevic, M., D. Jankovic, and A. Peulic. 2014. Thermography based breast cancer detection using texture features and minimum variance quantization. EXCLI J. 13: 1204-1215.

Ohashi, Y., and I. Uchida. 2000. Applying dynamic thermography in the diagnosis of breast cancer. IEEE Eng. Med. Biol. 19(3): 42-51.

Rainwater-Lovett, K., J. M. Pacheco, C. Packer, and L. L. Rodriguez. 2009. Detection of foot-and-mouth disease virus infected cattle using infrared thermography. Vet. J. 180: 317-324.

SAS Institute. 2010. SAS/STAT Software, Release 9.3, SAS Institute Inc. Cary, NC.

Schaefer, A., N. Cook, C. Bench, J. Chabot, J. Colyn, T. Liu, E. Okine, M. Stewart, and J. Webster. 2012. The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. Res. Vet. Sci. 93: 928-935.

Stewart, M., K. Stafford, S. Dowling, A. Schaefer, and J. Webster. 2008. Eye temperature and heart rate variability of calves disbudded with or without local anaesthetic. Physiol. Behav. 93(4-5): 789-797.

Stokes, J., K. Leach, D. Main, and H. Whay. 2012. An investigation into the use of infrared thermography (IRT) as a rapid diagnostic tool for foot lesions in dairy cattle. Vet. J. 193: 674-678.

Talukder, S., G. Gabai, and P. Celi. 2015. The use of digital infrared thermography and measurement of oxidative stress biomarkers as tools to diagnose foot lesions in sheep. Small Ruminant Res. 127: 80-85.

Valera, M., E. Bartolomé, M. J. Sánchez, A. Molina, N. Cook, and A. Schaefer. 2012. Changes in eye temperature and stress assessment in horses during show jumping competitions. J. Equine Vet. Sci. 32: 827-830.

6 Sheep lameness detection from individual hoof load

D.T. Byrne[1 3], D.P. Berry[1], H. Esmonde[3], F. McGovern[2], Creighton P[2], and N. McHugh[1]

[1]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland, P61 P302.

[2]Animal & Grassland Research and Innovation Centre, Teagasc, Mellows Campus, Athenry, Co. Galway, Ireland, H65 R718.

[3]Dublin City University, Glasnevin, Dublin 9, Co. Dublin, Ireland, D09 Y5N0.

This is a pre-copyedited, author-produced version of an article accepted for publication in the journal of Computers and Electronics in Agriculture following peer review. The version of record Byrne DT, Berry DP, Esmonde H, McGovern F, Creighton P, McHugh N. Sheep lameness detection from individual hoof load Computers and Electronics in Agriculture. 2019 Mar 1 is available online at: https://www.sciencedirect.com/science/article/pii/S0168169918312365?via%3Dihub (doi: 10.1016/j.compag.2019.01.048)

6.1 Abstract

Lameness is a leading cause of morbidity in sheep but routine inspection of sheep for lameness is labor intensive. The objectives of the current study were to i) build and test a custom hoof weigh crate (HWC) to measure the individual hoof load of sheep, ii) quantify the relationship between hoof health status and the load a sheep distributes to each hoof, and iii) evaluate the ability of the HWC to differentiate healthy from infected hooves. The overall footprint of the HWC was 950mm x 450mm, wherein two strain-gauge load cells were placed underneath each of four individual hoof platforms. An experiment was conducted over nine non-consecutive days between July and October 2017. On each experimental day, a total of 20 ewes (consisting of lame and healthy ewes) were placed on the HWC for five-minutes each. Each sheep hoof was visually assessed for lesions by three independent operators and a hoof lesion score assigned (scale from 0 -healthy to 4 -severe footrot). In addition to individual hoof load, the load placed on each hoof was divided by the sum of the load of the respective contralateral pair (front or back hooves), and multiplied by 100 to express the contralateral load percentage. A linear mixed model was invoked for each of the two load parameters as the dependent variable while hoof lesion score, contralateral pair, and their interaction were included as fixed effects. Each hoof was classified into a hoof lesion score category based on its load parameter values, and the numbers of correct and incorrect classifications were quantified. Healthy front hooves naturally carried more weight (60% of total weight) in comparison to healthy back hooves (40% of total weight), but when front or back hooves were infected to the same extent, they carried the same load. Results from the linear mixed model showed a small mean difference (4.5kg) in hoof load between healthy front hooves and those with a mild infection (i.e., score = 1), but there was no hoof load difference between healthy back hooves and those with a mild infection ($P<0.05$). The lowest proportion of misclassified hooves (sensitivity of 100% and specificity of 95%) was observed when the contralateral

load percentage was used to differentiate between healthy hooves and those with extensive inter-digital dermatitis. Results herein indicate that the HWC could be used to automatically detect extensive infection in sheep hooves.

Key words: weight distribution, static load cells, precision livestock farming.

6.2 Introduction

Lameness is a serious welfare concern and leading causes of morbidity in sheep (Dohoo et al., 1985; Winter, 2004). Footrot, the most prevalent cause of lameness (Winter, 2004; Conington et al., 2010), has a reported international prevalence of between 0.4 and 23.3% (Conington et al., 2010; Gelasakis et al., 2013). Since footrot can be contagious, the best strategy to pre-emptively reduce prevalence is early detection and subsequent treatment (Kaler and Green, 2008). Routine detection of hoof lesions, however, is labour intensive as the gold standard involves physically overturning each animal and visually assessing each hoof.

Previous research has attempted to automate the visual assessment of cattle, chickens, pigs, and horses for lameness using machine vision (Aydin et al., 2010; Poursaberi et al., 2011; Kashiha et al., 2014; Abdul Jabbar et al., 2017). Automatically extracting useable animal features from images can be problematic in real world conditions (Kashiha et al., 2014) and sometimes requires manipulation of the imaging environment (Nääs et al., 2018). Thermal imaging has also been proven to be a promising technology to detect lameness in cattle, sheep, and horses (Eddy et al., 2001; Stokes et al., 2012; Byrne et al., 2018) but results can depend on the ambient temperature (Byrne et al., 2018). Alternatively, as accelerometer data are less influenced by environmental conditions, wearable accelerometers are commercially available to detect lameness in dairy cattle (IceRobotics, 2017). For a flock of sheep, a wearable accelerometer device on every

animal may not be economically viable due to larger flock sizes and smaller profits per animal when compared to cattle.

When severe lameness occurs, an animal tends to elevate the affected hoof, removing or reducing any downward pressure (Angell et al., 2015); therefore, previous research (Maertens et al., 2011; Pluym et al., 2013; Byrne et al., 2018) attempted to quantify load distribution in order to detect lameness in horses, cattle, and pigs. A quasi-piezoelectric mat (i.e., the Emfit sensor; Pastell et al. (2008b)) and a mat with pressure sensitive sensors placed in a 2D array (i.e., the Gaitwise system; Maertens et al. (2011)) have been developed to detect lameness in cattle; under the right conditions, the Gaitwise system achieved a sensitivity and specificity of 90 and 100%, respectively (Maertens et al., 2011). Studies have also used strain gauge based load cells to detect lameness in horses, cattle, and pigs (Hood et al., 2001; Pastell et al., 2008a; Pluym et al., 2013). None of the aforementioned weighing systems have been tested for their ability to detect lameness in sheep. Without certain design features to reduce erroneous measurements, these technologies many underperform; some studies have reported discarding the entire measurement period for up to 10% of all animals (Pastell et al., 2008a; Pastell et al., 2008b).

The objective of the current study was to quantify the relationship between sheep hoof health status and the load a sheep distributes to each hoof. The ability of hoof load to differentiate healthy and infected hooves was also evaluated. Results from the present study could aid in the development of an automated lameness detection tool for sheep.

6.3 Materials and Methods

6.3.1 Development of the hoof weigh crate system

A hoof weigh crate (HWC) was designed to fit within a pre-existing 550mm wide sheep raceway; such raceway dimensions are consistent with what is generally used internationally. The hoof weigh crate *in situ* is illustrated in Figure 6.1a, while Figure 6.1b is a 3D model of the crate.

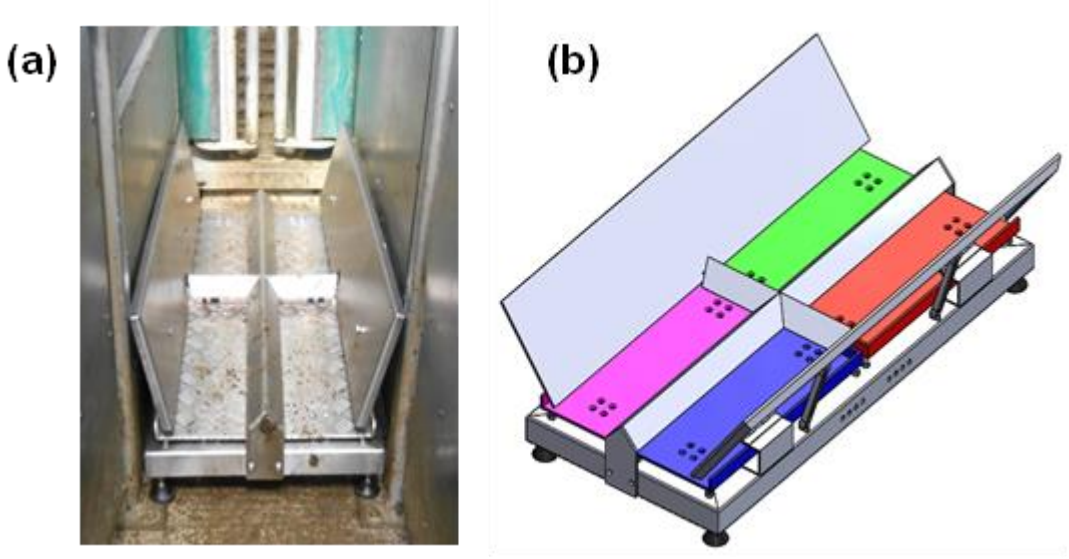


Figure 6.1 (a) An image of the hoof weigh crate in the sheep raceway, and (b) an isometric view of the 3D model of the hoof weigh crate with each load platform a different color.

In order to determine the dimensions for each of the four hoof platforms, physical measurements on 20 sheep of various breeds were used to determine the distance 1) between all four hooves, 2) from the sheep's head to the front hooves, and 3) the overall body length. Based on the observed range of these distances, the final size of each of the four load platforms was 455mm x 185mm, and so the overall footprint of the HWC was 950mm x 450mm. In addition, the total weight of each of the 20 sheep was measured to determine the load requirements of the HWC. The weight of each sheep ranged from 60 to 100kg, and therefore every single load cell in the system had a rated capacity of 100kg as it

was possible for a sheep to place all of its weight on a single hoof, albeit for a short period of time. While four single point load cells could have been used per platform, two strain-gauge cantilever load cells (load cell 1242, Sensor techniques Limited, United Kingdom) were placed underneath each load platform, thereby reducing the amount of equipment used (from a total of 16 to 8 load cells). Due to the type of load cells used and the size of each hoof platform, errors from eccentric loading were expected to be minimal; these errors could increase if a single load cell was used per platform.

The HWC was manufactured from stainless steel with aluminium checker plate for each of the four hoof platforms, thus ensuring no corrosion occurred and that hooves did not slip. Hoof placement was also a factor considered in the design of the HWC; large steep adjustable side panels ensured that the sheep could not stand to the side of the platform while pieces of angle iron covered the intersections between each platform as illustrated in Figure 6.1.

The analogue signal output from each load cell was connected to a laptop using two NI9237 simultaneous bridge modules and a NI cDAQ – 9174 (National instruments corporation, Texas, USA) compact DAQ module so measurements could be processed in LabVIEW. The digitized output of the loads cells was sampled at 1kHz and converted into kilograms within a LabVIEW program using the linear response characteristics that were established during the calibration of each load platform. The LabVIEW program facilitated for all HWC data to be recorded to an excel spread sheet.

Before any animal was placed into the HWC, each load cell was calibrated individually using a 0.1kg and an 80kg calibrated weight. Following this, the output of each hoof platform (i.e., two load cells joined by a single plate) was assessed, whereby, loads of 1, 5, 20, 40, 60 and 80kg were incrementally placed (using a set of assorted calibrated weights) on each hoof platform five times and the readings were recorded; this

facilitated a response characteristic to be developed and applied to the raw outputs of each hoof platform. In order to investigate whether position of the load had an effect on the recorded load, an eccentricity test was performed whereby a 20kg calibrated weight was placed at each of the four extremities of each hoof platform (Figure 6.2); this was also repeated five times per hoof platform.

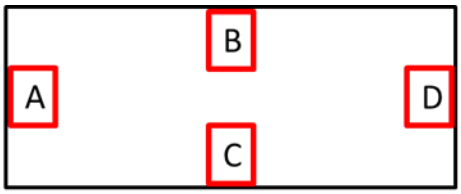


Figure 6.2 Load placement positions (A, B, C, and D) for the hoof platform eccentricity test

6.3.2 Experimental Procedure

The objectives of the present study were to: i) evaluate the repeatability of hoof load measurements made using the custom HWC, ii) quantify the relationship between measured hoof load and observed hoof health status, and iii) investigate whether the HWC could be used to differentiate between healthy and infected hooves. An experiment was undertaken in Athenry research centre, Teagasc, Athenry, Co. Galway, Ireland (53.287611 latitude, -8.767840 longitude). All procedures were conducted under approval from the Teagasc Animal Ethics Committee on experimental animal use (TAEC141-2017) in accordance with the Cruelty to Animals Act 1876 and the European Communities Regulations, 1994.

The experiment was conducted over nine non-consecutive days between July and October 2017, using a group of Texel, Suffolk, Belclare, and crossbred ewes. Before the experiment began, a cohort of 18 ewes was randomly selected as a control group. An additional flock of 120 ewes were assessed for locomotion (Angell et al., 2015), and any

ewe that walked abnormally (i.e., locomotion score averaged across three independent operators of ≥ 1 on the locomotion scoring scale from Angell et al. (2015)) on the morning of each experimental day was included in the "lame group" for that day. On each experimental day, the entire lame group was chosen to participate in the experiment alongside a number of randomly chosen ewes from the control group; the total number of ewes examined each day was limited to 20. A total of 18 ewes were observed at least twice during the experimental period as part of the control group. The number of ewes from the control group observed on each day ranged from 7 to 15. On the morning of each experimental day, all four hooves of each overturned sheep were blindly scored for lesions by three independent operators on a scale from 0 to 4 (Table 6.1) where, 0 = healthy hoof and 4 = severe inter-digital dermatitis (Conington et al., 2008). Following hoof lesion scoring, all animals were allowed to rest in a grass paddock close to the shed for one hour and were then returned to the shed and individually placed into the HWC for a five-minute period. Immediately following this, the entire weight of each animal was measured using a Prattley weigh crate (Prattley industries Ltd., Temuka, New Zealand.).

Table 6.1 The hoof score definition (Definition), the number of front hooves (No. of front hooves), and number of back hooves (No. of back hooves) categorized by type of hoof lesion scoring scale and hoof score (Score) in the present study.

Hoof lesion scoring scale	Score	Definition*	No. of front hooves	No. of back hooves
Hoof lesion score by Conington et al. (2008)	0	Healthy hoof	247	271
	1	Mild IDD	8	3
	2	Extensive IDD	3	1
	3	Severe IDD/footrot	0	0
	4	Severe footrot	0	0
Edited hoof lesion score	0	Healthy hoof	245	268
	1	Mild IDD	7	3
	2	Extensive IDD or worse	7	5

*IDD=Inter-digital dermatitis

6.3.2.1 Hoof score edits

Hoof lesion scores per hoof were averaged across all three operators to obtain an average hoof lesion score per hoof. To ensure any erroneous hoof measurements were removed from the final dataset, a number of edits were imposed on the recorded hoof measures which were similar to those described by Byrne et al. (2018), who used thermal imaging to detect lameness in sheep. The hoof load records of hooves with a healthy hoof score (i.e., hoof lesion score = 0 from all operators) which was preceded or proceeded by a hoof lesion score ≥ 1 on any other experimental day were removed from the dataset as changes in hoof load may exist before or after clinical infection; data from six healthy hooves on three animals were removed. Since no category three or four hoof lesions were observed (Table 6.1), any hoof with a hoof score averaged across the three operators of >2 were set equal to two. Any remaining records where all three independent operators did not universally agree on a hoof lesion score of 0, 1, or 2 were deleted; data from 125 hooves

were discarded. If any ewe received a score averaged across operators of ≥1 on two or more hooves on a single day then all hoof load data from that ewe were removed from the dataset on that day; data from six ewes were deleted.

6.3.2.2 Load cell data filtering

To reduce erroneous measurements due to major ewe movement on the HWC over the five-minute weighing period, any load of > 40kg or < 0kg from a single load platform was identified and all load platform records captured at that time were deleted; 9% of the data was deleted. Load measurements for each hoof were then averaged across one second intervals throughout each five-minute measurement period.

Following the deletion of data likely due to major ewe movement, and the averaging over one second intervals, further data quality control measures were invoked to ensure that the steady-state load of each hoof was measured. A first derivative deletion of data was undertaken whereby, if the load placed on a hoof increased or decreased by 1.5kg from one second to the next, then all hoof load data for that ewe on the 2nd second was deleted; 34% of data points were deleted. Some erroneous measurements were also generated when a ewe temporarily placed her weight off the HWC by leaning on the side panels of the raceway. Therefore, after first derivative deletion, a process herein known as off-weight deletion was applied where data were deleted if the total weight of a ewe at a single second differed by 1.5kg from the median total weight of the entire measurement period; 3% of data points were deleted. The 1.5kg threshold imposed was chosen based on the distribution of the dataset. The number of seconds remaining per ewe after data deletion was approximately normally distributed with an average ± standard deviation (SD) of 195 ± 48 seconds.

6.3.2.3 Load parameters

The initial parameter recorded from each of the four load platforms was the hoof load, and this was dependent on both the total weight of the ewe but also the distribution of weight to each of the four hooves. Based on previous research in cattle and pigs (Pastell et al., 2010; Pluym et al., 2013), an additional load parameter was generated; the load placed on each hoof was divided by the sum of the respective contralateral pair (i.e., front hooves or back hooves), and multiplied by 100 to express the contralateral load percentage, which was calculated for every second of the measurement period. For example, to calculate the contralateral load percentage of the right front hoof, the right front hoof load was divided by the sum of the load placed on both of the front hooves and expressed as a percentage.

6.3.3 Statistical analyses

6.3.3.1 Repeatability of the HWC before and after data edits

To quantify the repeatability of the load parameters (i.e., the individual hoof load and the contralateral load percentage) for each hoof and the total load, the variation in each parameter due to the ewe, date, ewe-date, time in the HWC (measured in one second intervals), and error was estimated using mixed models in ASreml (Gilmour et al., 2009). Ewe, date, time, and the concatenation of ewe and date were included as random effects. Only the 18 healthy ewes (i.e., universal agreement of a hoof score = 0 on all four hooves) that were placed into the HWC two or more times during the entire experimental period were included in the analysis. A log likelihood ratio test was performed to test if the inclusion of each random term in the statistical model improved the fit to the data. The proportion of total variation due to each random effect was also calculated as the variation due to the random effect divided by the total variation. The co-efficient of variation for each random effect was calculated as the SD due to the random effect divided by the respective mean.

6.3.3.2 Association between hoof load and hoof health status

To quantify the load differences between hooves with various hoof lesion scores (recoded to a scale from 0 to 2 as previously described), linear mixed models were performed in SAS using PROC MIXED (SAS Institute, 2010). The hoof load and contralateral load percentage post-editing were individually averaged across the entire five-minute measurement period for each ewe of a given day. Following this, a linear mixed model was invoked for each of the load parameters, where the dependent variable was the load parameter in question, and the fixed effects were the hoof lesion score (i.e., score 0 to 2), hoof location (i.e., front or back), and their interaction. Ewe was included as a random effect. Both load parameters were tested for normality. When the contralateral load percentage was the dependent variable, any hoof that had a contralateral counterpart with a contralateral load percentage of <30% was deleted [data from 19 hooves were deleted, all of which were healthy (i.e., score = 0)]; these 19 healthy hooves were deemed to be outliers as they were not subject to the same conditions of most other healthy hooves (i.e., they had a contralateral counterpart holding an unusually low amount of weight, possibly due to infection).

6.3.3.3 Diagnostic capability of the HWC

To quantify the proportion of hooves which the HWC could correctly identify as having a hoof lesion score of 0, 1, or 2, several load thresholds were applied to both load parameters (i.e., the hoof load and the contralateral load percentage). The diagnosis of each hoof predicted from the HWC data was compared to the actual hoof lesion scores and the resulting sensitivity and specificity calculated. In the current study, the optimum load threshold was one which achieved a balanced sensitivity and specificity; as this facilitates comparison between past and future studies (Greiner et al., 2000). When load thresholds were used to differentiate between two hoof lesion scores (e.g., 0 vs. 1), then all hooves which received the other hoof lesion score (e.g., a score of 2) were temporarily removed.

6.4 Results

When the accuracy of each of the four load platforms was tested with calibrated weights, the mean ± SD of the difference between the actual and recorded load (n = 120) was -0.001 ± 0.008kg. After a calibrated 20kg weight was repeatedly placed at the extremities of each load platform, the mean ± SD of the difference between the actual and recorded load (n = 80) was 0.007 ± 0.028kg. The percentage of lame ewes (i.e., had one hoof where all 3 operators agreed upon a score of ≥1) recorded daily during the experimental period ranged from 10 to 33%. When all records from all the ewes with an infected (hoof score ≥1 average across all three operators) hoof were temporarily removed from the dataset, the mean ± standard error (SE) hoof load and contralateral load percentage for each of the four hooves was calculated and is shown in Table 6.2. The hoof load distribution (before first derivative or off-weight deletion) from an example ewe with four healthy hooves and an example ewe with extensive inter-digital dermatitis (i.e., score = 2) in the left front hoof is shown in Figure 6.3. As seen in Figure 6.3, the healthy ewe distributed her weight evenly between the left and right side of her body, but the ewe with an affected hoof removed almost all of the weight from the infected hoof (i.e., left front hoof) and placed additional weight on the contralateral counterpart (i.e., right front hoof).

Table 6.2 The mean (standard error in parenthesis) healthy (hoof score = 0 from all three operators) hoof load and contralateral load percentage (CLP) for each of the four hooves. Hoof load or CLP values denoted with different letters (i.e., a, b, or c) were significantly ($P<0.05$) different from each other.

Load Variable	Left front hoof	Left back hoof	Right front hoof	Right back hoof
Hoof load (kg)	20.66 (0.50)^{a}	15.14 (0.50)^{b}	21.34 (0.50)^{a}	13.95 (0.50)^{b}
CLP (%)	48.62 (0.01)^{ab}	52.14 (0.01)^{c}	50.06 (0.01)^{a}	47.60 (0.01)^{b}

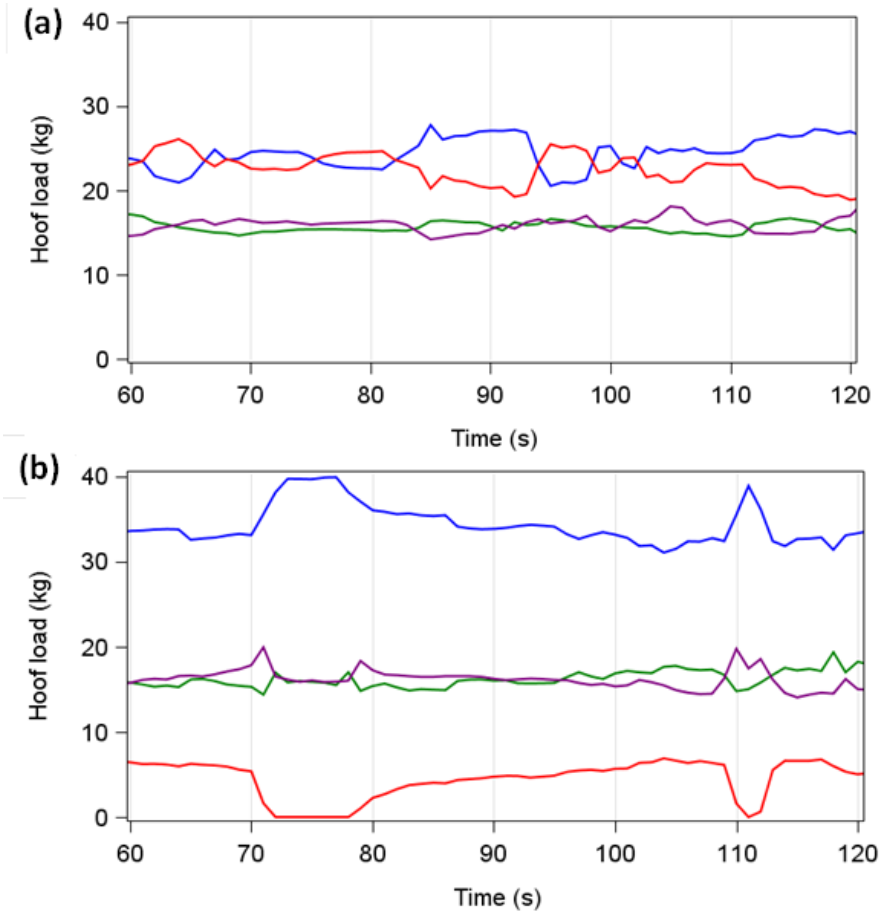


Figure 6.3 The hoof load distribution [right front hoof (—), left front hoof (—), right back hoof (—) and left back hoof (—)] of a) a ewe with four healthy hooves and b) a ewe with extensive inter-digital dermatitis (i.e., score = 2) in the left front hoof.

6.4.1 Repeatability of the HWC before and after data edits

The inclusion of all random terms in the mixed model improved the fit to the pre and post edited data (i.e., before and after first derivative and off weight deletion), with the exception of the term for each second of time spent in the HWC (i.e., 1 to 300). An example of how first derivative deletion worked for an example ewe is shown in Figure 6.4. After first derivative data deletion and off weight deletion, the error SD across all load parameters and anatomical areas decreased by 0.81kg to 1.02kg for hoof load and decreased by 1.95% to 2.03% for the contralateral load percentage. The change in SD due to all other significant factors (ewe, date, ewe-date) was smaller and inconsistent (-0.32kg to 0.17kg for hoof load, -0.63% to 0.31% for contralateral load percentage). The SD and the co-efficient of variation due to each random effect after first derivative data deletion and off weight deletion are shown in Table 6.3. The ewe accounted for between 36.84 (left back hoof) and 44.22% (right front hoof) of the total variation in hoof load after edits, whereas the proportion of total variation due to error ranged from 32.78 (right front hoof) to 38.90% (left back hoof). Error accounted for the greatest proportion of total variation in the edited contralateral load percentage across all four hooves (50.57% for the left front hoof to 51.61% for the right back hoof). The ewe contributed between 13.76 (left front hoof) and 19.49% (left front hoof) of the total variability in the edited contralateral load percentage. While data edits did not have a major impact on the SD of total weight due to the ewe (13.80kg; increase of 0.04kg), date (1.23kg; decrease of 0.59kg), or ewe-date (0.92kg; decrease of 0.11kg), the error variation of total weight decreased from a SD of 4.68kg before edits to 0.33kg after data edits.

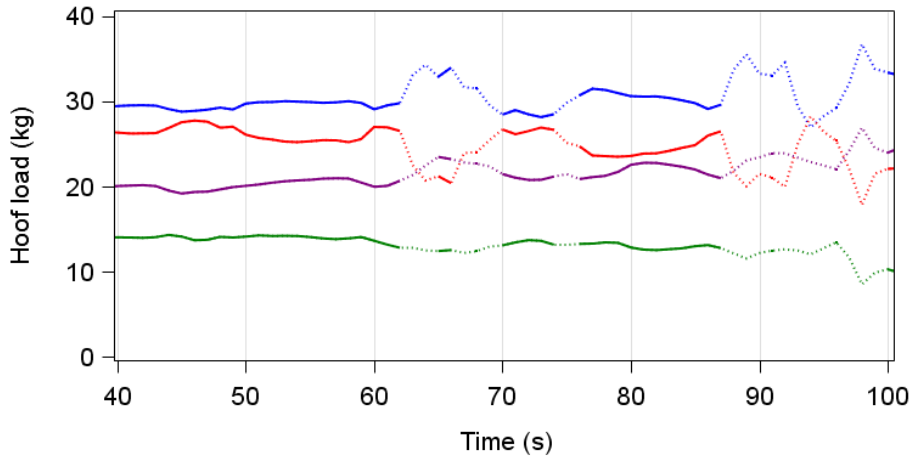


Figure 6.4 The hoof load distribution [right front hoof (—), left front hoof (—), right back hoof (—) and left back hoof (—)] of a healthy ewe where the dotted lines represent data that was deleted when one of the four hooves changed by 1.5kg or more within one second (first derivative deletion). All solid lines represent the remaining data.

Table 6.3 The standard deviation and co-efficient of variation (CV) due to the ewe, date, ewe-date (the combination of ewe and date) and error are shown for both load parameters [hoof load and contralateral load percentage (CLP)] from each anatomical area (Ano. area).

Ano. area	Variable	Hoof load (kg)	CV of hoof load (%)	CLP (%)	CV of CLP (%)
Right front hoof	Ewe	4.52 (2.74)	17.56	2.72 (1.99)	5.11
	Date	0.76 (0.71)	2.95	0.88 (1.06)	1.64
	Ewe-Date	1.59 (0.74)	6.16	3.31 (1.53)	6.20
	Error	3.35 (0.37)	13.01	7.07 (0.77)	13.26
Left front hoof	Ewe	4.19 (2.54)	18.43	2.72 (1.99)	5.83
	Date	0.50 (0.55)	2.21	0.88 (1.06)	1.87
	Ewe-Date	1.52 (0.7)	6.67	3.31 (1.53)	7.08
	Error	3.45 (0.38)	15.17	7.07 (0.77)	15.14
Left back hoof	Ewe	3.04 (1.86)	17.34	2.44 (2.17)	4.70
	Date	0.61 (0.59)	3.46	1.33 (1.58)	2.57
	Ewe-Date	1.39 (0.64)	7.90	4.81 (2.23)	9.27
	Error	3.21 (0.35)	18.34	9.15 (1.00)	17.63
Right back hoof	Ewe	3.45 (2.11)	21.08	2.44 (2.17)	5.07
	Date	0.56 (0.58)	3.43	1.33 (1.58)	2.77
	Ewe-Date	1.47 (0.68)	8.96	4.81 (2.23)	10.00
	Error	3.03 (0.33)	18.49	9.15 (1.00)	19.03

6.4.2 Association between hoof load and hoof health status

The number of hooves with lesion scores of 0, 1, and 2 are shown in Table 6.3. Hoof lesion score (i.e., 0, 1, or 2), hoof location (i.e., front or back), and their interaction were all found to be significantly associated with hoof load ($P<0.05$), when a linear mixed model was invoked. A box-and-whisker plot of the relationship between hoof load and hoof lesion score for the front and back hooves is shown in Figure 6.5. The mean ± SE hoof load for front hooves with lesion scores 0, 1, and 2 were 21.6 ± 0.4, 17.1 ± 1.2 and 6.4 ± 1.3kg, respectively. For the back hooves, the mean ± SE hoof load at hoof lesion scores of 0, 1, and 2 were 14.7 ± 0.4, 14.9 ± 2.0 and 6.3 ± 1.4kg, respectively. Hoof load of the front hooves reduced with lesion severity, but the same was not true for the back hooves. There was no difference between the hoof load of healthy back hooves (i.e., score = 0) and those with mild inter-digital dermatitis (i.e., score = 1) ($P>0.05$). Hoof load was lower in back hooves with extensive inter-digital dermatitis (i.e., score = 2) ($P<0.01$) in comparison to back hooves with all other levels of infection (i.e., score of 0 or 1). Healthy front hooves carried more weight than healthy back hooves, but when front or back hooves were infected to the same extent, there was no difference in the amount of weight carried by the front or back hooves ($P<0.05$).

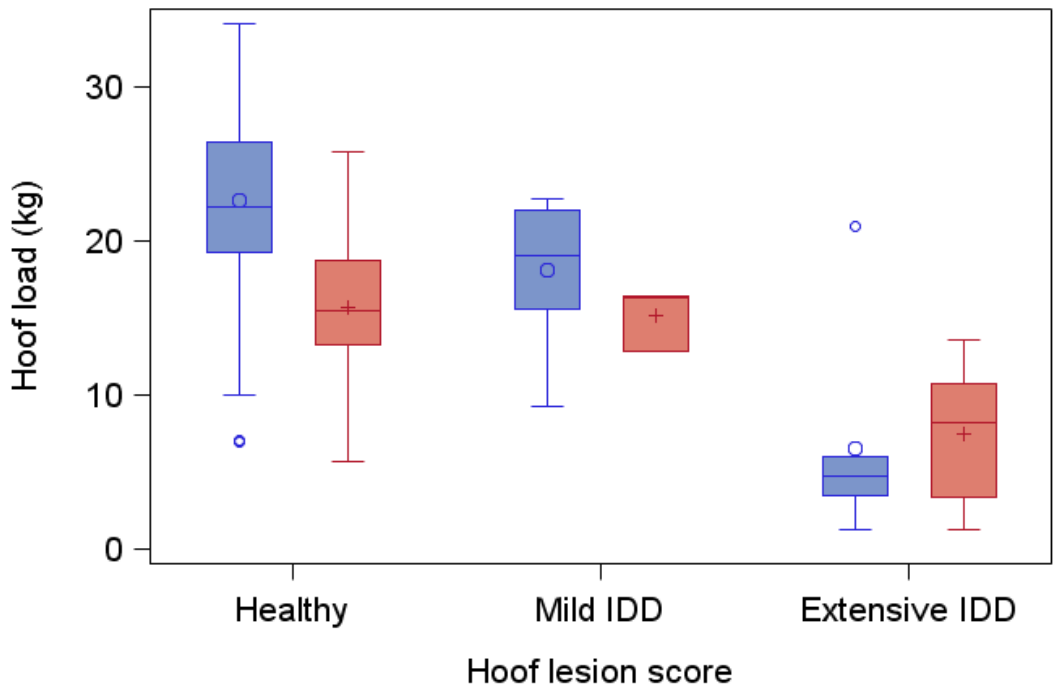


Figure 6.5 A box-and-whisker plot of the relationship between hoof load and hoof lesion score of front (■) and back (■) hooves; the mean (O or +), median (—), interquartile range (IQR) (■ or ■), greatest (┬) and lowest (┴)values within 1.5*IQR of the IQR, and outliers (○) are shown. Inter-digital dermatitis was abbreviated to IDD.

Hoof lesion score (i.e., 0, 1, or 2), hoof location (i.e., front or back), and their interaction were all associated ($P<0.05$) with contralateral load percentage. A box-and-whisker plot of the relationship between contralateral load percentage and hoof lesion scores for the front and back hooves is in Figure 6.6. The mean ± SE contralateral load percentage for front hooves with lesion scores of 0, 1, and 2 were 49.8 ± 0.4, 41.2 ± 2.5, 16.0 ± 2.5%, respectively. For the back hooves, the mean ± SE contralateral load percentage associated with hoof lesion scores of 0, 1, and 2 were 50.0 ± 0.4, 51.0 ± 3.9, and 24.5 ± 3.0%, respectively. No difference was observed between the contralateral load percentage of healthy front hooves, healthy (i.e., score = 0) back hooves, or back hooves with mild inter-digital dermatitis (i.e., score = 1). The contralateral load percentage of front hooves differed between each category of hoof lesion score (i.e., 0, 1. and 2). Front hooves

with extensive inter-digital dermatitis (i.e., score = 2) had a lower contralateral load percentage in comparison to back hooves with the same infection.

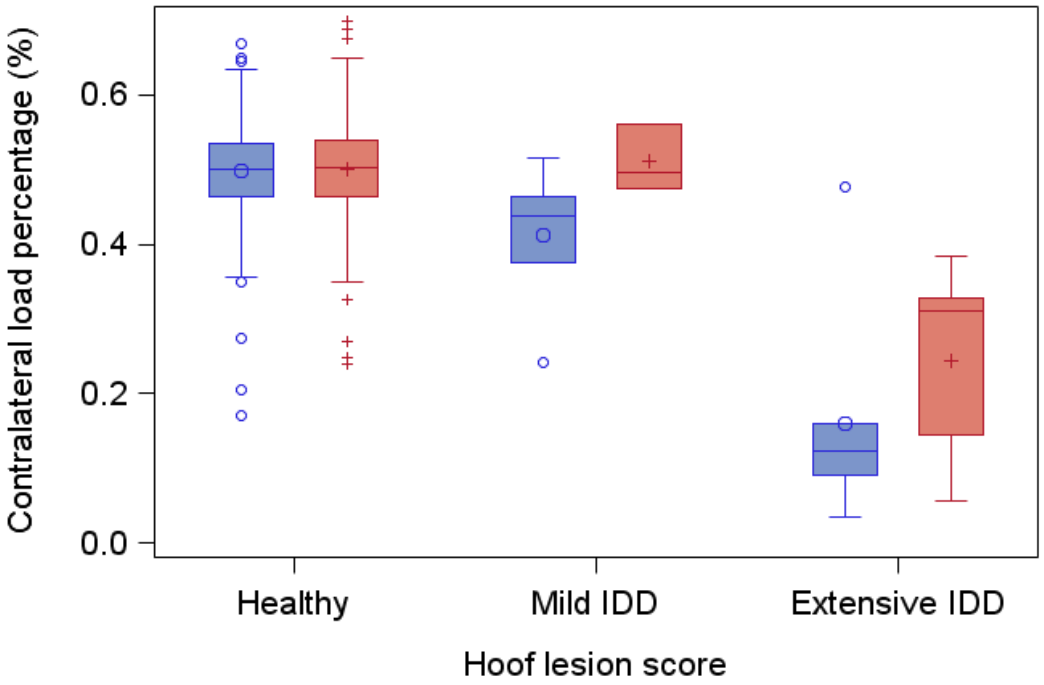


Figure 6.6 A box-and-whisker plot of the relationship between contralateral load percentage and hoof lesion score of front (■) and back (■) hooves; the mean (O or +), median (—), interquartile range (IQR) (■ or ■), greatest (┬) and lowest (┴)values within 1.5*IQR of the IQR, and outliers (○ or +) are shown. Inter-digital dermatitis was abbreviated to IDD.

6.4.3 Diagnostic capability of the HWC

The optimum load parameter thresholds (i.e., threshold which achieved a balanced sensitivity and specificity) to differentiate between hooves of two hoof lesion scores (i.e., 0 vs. 1, or 1 vs. 2) and the resulting sensitivity and specificity are shown in Table 6.4. When front hooves with a hoof load less than 20kg were considered to have mild inter-digital dermatitis (i.e., score = 1) and all other hooves were considered healthy (i.e., score = 0) the sensitivity and specificity was balanced yet moderate (sensitivity of 71.43% and specificity

of 70.20%); similar results were attained when a load threshold of 46% was applied to the contralateral load percentage to differentiate the same hooves. No load parameter was capable of differentiating back hooves with a hoof score of 0 and 1 to level at which both sensitivity and specificity were greater than 60% (Table 6.4). Contralateral load percentage was the best load parameter to differentiate between healthy hooves (front or back) and those with extensive inter-digital dermatitis (i.e., score = 2), where sensitivity and specificity ranged from 85.71% to 100.00%. The load thresholds which were identified to optimally differentiate hooves with lesion scores of 0 and 2 and the load thresholds identified to differentiate hooves with lesion scores of 1 and 2 were identical for respective load parameters and hoof location. For example, 13kg was the optimum hoof load threshold to differentiate back hooves which had hoof lesion scores of 0 and 2, or hoof lesion scores of 1 and 2.

Table 6.4 The optimum hoof load and contralateral load percentage (CLP) thresholds with the resulting sensitivity and specificity for differentiating front and back hooves of various hoof lesion score categories (0 = healthy, 1 = mild inter-digital dermatitis and 2 = extensive inter-digital dermatitis).

Load variable	Hoof score categories	Anatomical area	Threshold	Sensitivity (%)	Specificity (%)
CLP (%)	0 vs. 1	Front	46	71.43	76.73
		Back	50	66.67	51.49
	0 vs. 2	Front	20	85.71	99.59
		Back	40	100.00	94.78
	1 vs. 2	Front	20	85.71	100.00
		Back	40	100.00	100.00
Hoof load (kg)	0 vs. 1	Front	20	71.43	70.20
		Back	16	33.33	45.15
	0 vs. 2	Front	6	85.71	100.00
		Back	13	80.00	77.99
	1 vs. 2	Front	6	85.71	100.00
		Back	13	80.00	66.67

6.5 Discussion

Overall the HWC proved to be fit for purpose, as it was capable of recording the hoof load distribution of sheep in the present study. Hoof load measurements proved to be repeatable, and large load differences were observed between healthy hooves (i.e., score = 0) and those with extensive inter-digital dermatitis (i.e., score = 2).

6.5.1 The HWC design

The main differences between the mechanical design of the HWC in the present study and that of other hoof weigh crates constructed for cattle, pigs, and horses (Hood et al., 2001; Pastell et al., 2008a; Pluym et al., 2013), was the dimensions, number of load cells per hoof platform, and the design features intended to ensure correct hoof placement. The load variation due to eccentricity in the present study was less than that of other studies in pigs (Pluym et al., 2013), which may be due to the greater number of load cells per hoof platform in the current study. While accurate load platforms are a requirement to measure individual hoof load, the HWC must also be designed in such a way to ensure the animals consistently place each hoof on the correct platform. The side panels and angle irons separating each hoof platform (Figure 6.1) did help to ensure correct hoof placement in the present study, but higher partitions (e.g., 100 to 200mm) between each hoof platform may have been beneficial; higher partitions were used by Sun et al. (2011) who developed a similar hoof weigh crate to measure the load a sow placed on each hoof. Since errors due to hoof placement were minimized in the present study, the next objective was to obtain stable load measurements. A stable measurement may be difficult to attain when an animal is held in an unfamiliar environment, since they can become nervous and shift their weight from one hoof to another. To reduce stress (and thereby weight shifting) in the sheep on the HWC in the present study, another ewe was held in a Prattley weigh crate (for the entire five-minute measurement period) which was directly in front of the HWC. When the ewe on the HWC could see another ewe, a reduction in agitation was visually obvious.

Measuring hoof load while the animal drinks or eats may also reduce agitation, but this may cause extraneous variations in load distribution (Sun et al., 2011). Placing the HWC at a drinking/ feeding trough would facilitate frequent hoof load measurements to be gathered for lameness detection or growth monitoring. If growth monitoring was the sole function of a weigh crate, only the load placed on the front hooves would need to be measured as, in the current study, a correlation of 0.98 was observed between the sum of the load on the front hooves and the total weight of healthy ewes.

6.5.2 Repeatability of HWC measurements before and after data edits

As a live animal had some limited freedom to move around on the HWC, the entire duration of the measurement period was not a true representation of their normal load distribution, and may have instead been caused by excessive movement or an improper stance on the HWC. Previous studies described simple data-driven thresholds to define what data needs to be deleted from a measurement period (Pastell et al., 2008a; Sun et al., 2011; Pluym et al., 2013). Pastell et al. (2008a) examined the hoof load of cattle and removed any data points which were less than 30kg from the maximum hoof weight observed during an entire measurement period; a similar method was adopted by Pluym et al. (2013) who measured the hoof load of sows. Some studies in horses and cattle did not remove any data points at all (Neveux et al., 2006; Hood et al., 2001). The current study achieved the intended goal of reducing the error variation in all load parameters by using first derivative deletion and off-weight deletion (6.3.2.2 Load cell data filtering). Hoof load measurements edited as previously described, proved to be repeatable as only a moderate proportion of the variation was due to error (32.78% for the right front hoof to 38.90% for the left back hoof). After data deletion, the variation in hoof load due to error was equivalent to the observed variation in sow hoof load due error from Pluym et al. (2013), and as a ewe typically weighs less than a pig, the co-efficient of variation due to error was higher in the present study [5.4 to 7.2% in Pluym et al. (2013) and 13.1 to 18.5% in the

present study]. Error contributed a large proportion of the total variation (50.57% for the left front hoof to 51.61% for the right back hoof) to the contralateral load percentage which may suggest the contralateral load percentage is not suitable as a repeatable measure. However, it can still be used as a disease detection tool since the contralateral load percentage difference between hooves with extensive inter-digital dermatitis and healthy hooves (between 25.5% and 33.8%) was greater than the total variation in the contralateral load percentage (between 13.9% and 17.7%). Unlike other studies, the current study quantified the variability in the load parameters due to date, ewe-date, and time on the HWC. Date had a significant effect on the variability of all load parameters, which makes sense as the weight of a ewe may change from day to day. Time spent on the HWC did not have a significant effect on the variance of any load parameters, which indicates that the load distribution of a ewe did not change much before or after data editing over the five-minute measurement period.

6.5.3 Hoof load distribution and severity of infection

The natural healthy load distribution of a species must be understood before the change in load distribution due to infection can be determined. Previous research showed that the front hooves of cattle, horses, and pigs (Hood et al., 2001; Pastell et al., 2006; Pluym et al., 2013) support more weight than the back hooves, while weight is evenly distributed between the left and right side of the body; the current study shows that sheep also conform to these trends (Table 6.2). Many studies in cattle, pigs, and horses (Neveux et al., 2006; Pluym et al., 2013; Hood et al., 2001) showed that when a hoof becomes lame, the animal removes weight from the affected hoof and primarily places weight on the contralateral counterpart. The current study corroborates this statement as the mean ± SE percentage of total load placed on each hoof from the ewes (n = 5) who had extensive inter-digital dermatitis (i.e., score = 2) in the left front hoof was 4.9 ± 2.1 (left front hoof), 28.6 ± 4.6 (left back hoof), 49.5 ± 1.6 (right front hoof) and 16.9 ± 4.0% (right back hoof).

While the ewes did place some additional weight onto the left back hoof (i.e., the ipsilateral hoof) most of the load which was removed from the infected hoof was placed onto the right front hoof (i.e., the contralateral hoof). No study has quantified the load normally placed on an infected front hoof and compared it to how much load a ewe would normally place on a back hoof which was infected to the same extent. The current study showed that when using hoof load to identify mild infection, front and back hooves should be assessed separately as they can both react differently to a mild infection.

As neither the hoof load nor the contralateral load percentage differed between healthy back hooves and those with mild inter-digital dermatitis, the optimum threshold yielded a poor sensitivity and specificity. No other study has attempted to use the standing load distribution of a species to detect mild hoof infection, but instead classifies mildly lame animals as not lame (Van Nuffel et al., 2015). The hoof load and contralateral load percentage in the current study was somewhat able to differentiate between front hooves which were healthy (i.e., score = 0) and those with mild inter-digital dermatitis (i.e., score = 1). In an attempt to find a better method to differentiate between healthy and afflicted hooves, alternative load parameters were tested as part of the present study. The percentage of total load placed on a hoof was evaluated but did not out-perform any of the aforementioned load parameters (i.e., hoof load or contralateral load percentage). Many studies in cattle and horses have suggested using the SD of hoof load over time to identify infection as a lame animal will tend to move more due to discomfort (Hood et al., 2001; Pastell et al., 2006; Neveux et al., 2006). As part of the present study, the SD of hoof load over time, the SD of the contralateral percentage over time, and the SD of the percentage of total load placed on a hoof over time (before data edits) were tested for their ability to differentiate between healthy and mildly afflicted hooves (i.e., score = 1); they did not out-perform any of the aforementioned load parameters (i.e., hoof load or contralateral load percentage).

While mild inter-digital dermatitis was difficult to detect using various load parameters, the HWC was capable of detecting the load differences between healthy hooves and those with extensive inter-digital dermatitis; this is because extensive inter-digital dermatitis was associated with a large reduction in hoof load (8.4 to 15.2kg) and contralateral load percentage (25.5 to 33.8%). Sensitivity and specificity results seen herein were comparable to those of previous work in cattle (Pastell and Kujala, 2007; Pastell et al., 2010a; Maertens et al., 2011). The optimum hoof load and contralateral load percentage thresholds for differentiating between hoof lesion scores of 0 and 2 were the same as those defined to differentiate between hoof lesion scores of 1 and 2 because there was minimal hoof load or contralateral load differences between healthy (i.e., score = 0) and mildly afflicted hooves (i.e., score = 1). The research herein is a preliminary look into how the load distribution of sheep changes when a single hoof is lame. Future work may investigate how the load distribution of sheep changes with multiple lame hooves or seek to validate the results herein. While the current study defined an optimal load threshold to be one which the sensitivity and specificity were balanced, in practice the threshold would be defined by the cost of identifying false positive and false negative results (Greiner et al., 2000).

6.6 Conclusion

The HWC developed as part of the present study was suitable for measuring the hoof load distribution of sheep; the mechanical design and post-processing techniques ensured that erroneous data were minimized. Measurements made using the HWC were found to be repeatable. Healthy front hooves naturally carried more weight (60% of total weight) in comparison to healthy back hooves (40% of total weight), but when front or back hooves were infected to the same extent, they carried the same load. The present study suggests

that a HWC could be used to reliably differentiate healthy hooves and those with extensive inter-digital dermatitis. Lesser forms of infection (e.g., mild inter-digital dermatitis) were more difficult to detect as there was little (front hooves; 4.5kg) to no difference (back hooves) between the hoof load of healthy and mildly afflicted hooves. Future studies may seek to validate the results herein or attempt to automate a hoof weigh crate for lameness detection in sheep. If future studies attempt to use the load distribution of any species to detect a mild form of hoof infection, front and back hooves should be assessed separately.

6.7 Conflict of interest statement

The authors herein declare there is no conflict of interest.

6.8 Literature Cited

Angell, J. W., P. J. Cripps, D. H. Grove-White, and J. S. Duncan. 2015. A practical tool for locomotion scoring in sheep: Reliability when used by veterinary surgeons and sheep farmers. Vet. Rec. 176(20): 521. doi: 10.1136/vr.102882

Aydin, A., O. Cangar, S. E. Ozcan, C. Bahr, and D. Berckmans. 2010. Application of a fully automatic analysis tool to assess the activity of broiler chickens with different gait scores. Comput. Electron. Agric. 73(2): 194-199.

Byrne, D. T., D. P. Berry, H. Esmonde, F. McGovern, P. Creighton, and N. McHugh. 2018. Infrared thermography as a tool to detect hoof lesions in sheep. Translational Animal Science.

Conington, J., B. Hosie, G. Nieuwhof, S. Bishop, and L. Bünger. 2008. Breeding for resistance to footrot–the use of hoof lesion scoring to quantify footrot in sheep. Vet. Res. Commun. 32(8): 583-589.

Conington, J., M. Speijers, A. Carson, S. Johnston, and S. Hanrahan. 2010. Foot health in sheep-prevalence of hoof lesions in UK and irish sheep. Adv, in Ani. Biosci. 1(1): 340.

Dohoo, I. R., R. A. Curtis, and G. G. Finley. 1985. A survey of sheep diseases in canada. Can. J. Comp. Med. 49(3): 239-247.

Eddy, A., L. Van Hoogmoed, and J. Snyder. 2001. The role of thermography in the management of equine lameness. The Veterinary Journal. 162(3): 172-181.

Gelasakis, A., G. Oikonomou, R. Bicalho, G. Valergakis, G. FTHENAKIS, and G. Arsenos. 2013. Clinical characteristics of lameness and potential risk factors in intensive and semi-intensive dairy sheep flocks in greece. J. Hell Vet. Med. Soc. 64(2): 123-130.

Gilmour, A., B. Gogel, B. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0.VSN International Ltd, Hemel Hempstead, HP1 1ES, United Kingdom.

Greiner, M., D. Pfeiffer, and R. Smith. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev. Vet. Med. 45(1-2): 23-41.

Hood, D. M., I. P. Wagner, D. D. Taylor, G. W. Brumbaugh, and M. K. Chaffin. 2001. Voluntary limb-load distribution in horses with acute and chronic laminitis. Am. J. Vet. Res. 62(9): 1393-1398.

IceRobotics 2017. COWALERT Lameness Detection Highly Commended. www.icerobotics.com. Retrieved on 14 January 2019, from http://www.icerobotics.com/news/cowalert-lameness-detection-highly-commended/.

Jabbar, K. Abdul, M. F. Hansen, M. L. Smith, and L. N. Smith. "Quadruped locomotion analysis using three-dimensional video." In Students on Applied Engineering (ICSAE), International Conference for, pp. 398-401. IEEE, 2016.

Kaler, J., and L. Green. 2008. Recognition of lameness and decisions to catch for inspection among sheep farmers and specialists in GB. BMC Vet. Res. 4(1): 41.

Kashiha, M. A., C. Bahr, S. Ott, C. P. Moons, T. A. Niewold, F. Tuyttens, and D. Berckmans. 2014. Automatic monitoring of pig locomotion using image analysis. Livestock Science. 159: 141-148.

Maertens, W., J. Vangeyte, J. Baert, A. Jantuan, K. C. Mertens, S. De Campeneere, A. Pluk, G. Opsomer, S. Van Weyenberg, and A. Van Nuffel. 2011. Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: The GAITWISE system. Biosyst. Eng. 110(1): 29-39.

Neveux, S., D. Weary, J. Rushen, M. Von Keyserlingk, and A. de Passillé. 2006. Hoof discomfort changes how dairy cattle distribute their body weight. J. Dairy Sci. 89(7): 2503-2509.

Nääs, I. d. A., L. C. M. Lozano, S. A. Mehdizadeh, R. G. Garcia, and J. M. Abe. 2018. Paraconsistent logic used for estimating the gait score of broiler chickens. Biosystems Engineering. 173: 115-123.

Pastell, M., H. Takko, H. Gröhn, M. Hautala, V. Poikalainen, J. Praks, I. Veermäe, M. Kujala, and J. Ahokas. 2006. Assessing cows' welfare: Weighing the cow in a milking robot. Biosyst. Eng. 93(1): 81-87.

Pastell, M., and M. Kujala. 2007. A probabilistic neural network model for lameness detection. J. Dairy Sci. 90(5): 2283-2292.

Pastell, M., M. Hautala, V. Poikalainen, J. Praks, I. Veermäe, M. Kujala, and J. Ahokas. 2008a. Automatic observation of cow leg health using load sensors. Comput. Electron. Agric. 62(1): 48-53.

Pastell, M., M. Kujala, A. Aisla, M. Hautala, V. Poikalainen, J. Praks, I. Veermäe, and J. Ahokas. 2008b. Detecting cow's lameness using force sensors. Comput. Electron. Agric. 64(1): 34-38.

Pastell, M., L. Hänninen, A. De Passillé, and J. Rushen. 2010a. Measures of weight distribution of dairy cows to detect lameness and the presence of hoof lesions. J. Dairy Sci. 93(3): 954-960.

Pastell, M., L. Hanninen, A. M. de Passille, and J. Rushen. 2010b. Measures of weight distribution of dairy cows to detect lameness and the presence of hoof lesions. J. Dairy Sci. 93(3): 954-960. doi: 10.3168/jds.2009-2385 [doi]

Pluym, L. M., D. Maes, J. Vangeyte, K. Mertens, J. Baert, S. Van Weyenberg, S. Millet, and A. Van Nuffel. 2013. Development of a system for automatic measurements of force and visual stance variables for objective lameness detection in sows: SowSIS. Biosyst. Eng. 116(1): 64-74.

Poursaberi, A., C. Bahr, A. Pluk, D. Berckmans, I. Veermäe, E. Kokin, and V. Pokalainen. 2011. Online lameness detection in dairy cattle using body movement pattern (BMP).

Intelligent systems design and applications (ISDA), 2011 11th international conference on,

Stokes, J., K. Leach, D. Main, and H. Whay. 2012. An investigation into the use of infrared thermography (IRT) as a rapid diagnostic tool for foot lesions in dairy cattle. Vet. J. 193: 674-678.

Sun, G., R. F. Fitzgerald, K. J. Stalder, L. A. Karriker, A. K. Johnson, and S. J. Hoff. 2011. Development of an embedded microcomputer-based force plate system for measuring sow weight distribution and detection of lameness. Appl. Eng. Agric. 27(3): 475-482.

Van Nuffel, A., I. Zwertvaegher, S. Van Weyenberg, M. Pastell, V. M. Thorup, C. Bahr, B. Sonck, and W. Saeys. 2015. Lameness detection in dairy cows: Part 2. use of sensors to automatically register changes in locomotion or behavior. Animals. 5(3): 861-885.

Winter, A. 2004. Lameness in sheep 1. diagnosis. In Pract. 26(2): 58-63.

7 Summary

The overall objective of this thesis was to investigate whether certain technologies could be used to quantify the health status of cattle or sheep. According to the literature thermal imaging could be used to detect various diseases but much of this work was undertaken in a controlled environment and most studies did not investigate the factors which affect the ability of IRT to correctly quantify the health status of an animal. The current thesis showed that the temperature difference between thermal image replicates can be larger than the temperature difference between sick and healthy animals. Therefore, a minimum of three replicates should be captured when thermal imaging is used to quantify the health status of livestock

One of the most prevalent health issues in dairy cattle is infection of the udder. Previous literature has shown that SCC (the current gold standard for measuring udder infection) of dairy cattle is related to USST (measured by IRT). The current thesis disagreed with previous studies by noting that in real world conditions SCC and USST could not be related despite the use of a large dataset as well as extracting a large number of variables from thermal images of the udder.

On the other hand, the current thesis showed that thermal imaging could be used to detect early stages of hoof infection in sheep. Moreover, the current thesis was the first body of work to demonstrate that the relationship between ambient temperature and hoof temperature was dependent on the health status of the hoof. This phenomenon meant that IRT had the optimal diagnostic capabilities at colder ambient temperatures.

An alternative technology with detection capabilities that were not dependant on ambient temperature was a device which measured the load a sheep placed on each hoof. The device, also known as the hoof weigh crate, was the first such system developed to detect lameness in sheep. The hoof weigh crate was shown to be capable of capturing

repeatable measurements from live sheep, although it was only capable of detecting severe lameness rather than mild lameness.

In summary, the procedures set out in the current thesis may be used to assess future technologies, thermal imaging has the potential to be used to quantify the health status of sheep hooves and the hoof load of a sheep may be a useful tool to detect severe lameness.

7.1 Paper 1: Temporal, spatial, inter-, and intra-cow repeatability of thermal imaging

Objective: To investigate the repeatability temperature measurements made using thermal images of dairy cattle in an agricultural environment, with a view to inform the imaging process.

- Little work has been published on the difference between thermal image replicates
- High repeatability and reproducibility seen in the medical industry with thermal imaging, but an agricultural environment is not as controllable
- Three experiments aimed to:
    - Investigate the repeatability of replicate images of the udder, eyes and hooves
    - Test the temperature variation between operators
    - Quantify the temperature variance between days
- Variation due to each random effect (i.e., cow, operator, and day) was quantified
- Difference between replicate images (e.g., 1.3°C for the eye) was in some cases greater than the difference between sick and healthy animals reported in the literature (e.g., 0.9°C for the eye)
- Hoof images were the most repeatable with 91.4 to 99.3% of variation attributed to the cow rather than error
- Operator variance showed that operators must be properly trained to achieve a low operator variance
- Day variance had a minimal effect
- A recommendation of three thermal images was put forth for using IRT to quantify the health status of animals

7.2 Paper 2: Investigation of the relationship between udder quarter somatic cell count and udder skin surface temperature of dairy cows measured by infrared thermography

Objective: To quantify the relationship between the somatic cell count of dairy cows and their udder skin surface temperature measured by IRT.

- Previous work conducted in controlled conditions has shown a logarithmic relationship between USST and SCC
- Three thermal images were captured each from 14 cows every day for a 2 month period, milk samples were also taken from each quarter every day post imaging
- A total of 28 different temperature variables were extracted from each quarter in the udder thermal images
- A range of methods were tested to relate USST and SCC including, simple correlations, using the animal as her own control, and using the five coldest quarters in the herd as a control
- Linear mixed models identified the factors which were attributed to the greatest proportion of variation in USST (ambient temperature and the amount of hair on the udder)
- Once random regression models were used to predict USST, the difference between actual and predicted USST was compared to SCC
- While maximum and average USST could be predicted to within 0.23 and 0.35°C respectively, USST could not be used to predict SCC

7.3 Paper 3: Infrared thermography as a tool to detect hoof lesions in sheep

Objective: To test the ability of IRT to detect hoof infection in sheep and to investigate the factors which hinder the diagnostic capability of IRT.

- Some work has shown that a temperature difference exists between healthy sheep hooves and those with infection
- No work was found to have tested the hoof infection detection capabilities of IRT or how this is affected by an agricultural environment
- Three experiments were conducted
    - The repeatability experiment: Aimed to test the repeatability of thermal images captured from sheep hooves
    - The exploratory experiment: Investigated the relationship between hoof temperature and hoof infection and tested the diagnostic capability of IRT
    - Validation experiment: Aimed to validate IRT as a lameness detection tool
- Repeatability experiment showed that hoof IRT measurements were highly repeatable
- Large hoof temperature differences existed between healthy and infected hooves
- A new variable was created (the daily baseline) whereby the average value of the five coldest hooves in the flock on a day was used to inform the hoof temperature cut-off point for that day
- IRT was found to be highly effective as an infection detection tool in the exploratory experiment (sensitivity of 92% and a specificity of 91%)
- Validation results showed that maximum hoof temperature may not be as useful as the daily baseline threshold. When any hoof with a temperature of 9°C above the daily baseline was considered infected, a sensitivity and specificity of 77% and 78% were achieved.

7.4 Paper 4: Sheep lameness detection from individual hoof load

Objective: Develop and test a crate to measure the load a sheep places on each hoof to detect lameness in sheep.

- Systems to detect lameness using the hoof load of cattle, pigs and horses have been developed although no study has developed a crate to detect lameness in sheep
- A custom hoof weigh crate was developed and used to investigate the:
    - Repeatability of hoof load measurements
    - Relationship between hoof load and hoof infection status.
    - The ability of the hoof weigh crate to detect lameness.
- Hoof load and the contralateral load percentage (i.e., the load of a single hoof divided by sum of the load from the respective contralateral pair and expressed as a percentage) tested as lameness detection variables.
- The percentage of total load a sheep placed on front hooves (60%) was greater than back hooves (40%)
- Large load difference observed between lame hooves and those with severe infection (15.2 kg for front hooves and 13.3kg for back hooves)
- When sheep removed weight from a single hoof they had a tendency to place more weight on the contralateral counterpart of the lame hoof (i.e., if the right front hoof was lame sheep tended to place more weight on the left front)
- Less weight was placed onto front hooves with mild infection in comparison to healthy front hooves, although there was no difference between the weigh placed on healthy back hooves or back hooves with a mild infection
- The best diagnostic capability results were observed when the contralateral load percentage was used to differentiate healthy back hooves and those with severe infection (sensitivity 100% and specificity 95%)

7.5 Overall conclusions and implications

There is great scope for investigating the use of various sensor technologies to quantify health status in livestock. Standardized methods for testing new technologies in real world conditions may speed up the development, demonstration, and implementation of such sensors. Many of the experimental methods and analytical techniques seen in this thesis can be applied to testing the diagnostic ability of future technologies. Before future researchers begin to use IRT, they should conduct a trial similar to paper 1 of the current thesis in order to test the variation present in their imaging procedures. These procedures and statistical techniques can also be applied to test the repeatability of alternative technologies as seen in paper 4. Much of the methods seen in paper 2 can be applied when IRT is used to detect different diseases from different anatomical regions in different species e.g., using thermal images of the eye to detect BRD in calves. Few studies have investigated the difference between predicted and actual temperature of anatomical regions as a tool to quantify health status but it has been suggested a number of times throughout the literature. The application of this methodology is clearly demonstrated in the current thesis. Paper 3 in the current thesis demonstrates the research that must be conducted before a product can be implemented on farm; therein factors which effected the diagnostic capability of IRT were tested which lead to a validation study of the diagnostic capabilities of IRT. Most papers with positive results do not include a validation study which is required for a robust assessment of the application of the technology.

As the current thesis was the first to show that the relationship between ambient temperature and hoof temperature can be affected by health status, this may have implications for other regions of the body on a range of species. For example, it may be possible that the temperature of cow hooves or the eyes of pigs behave in a similar way. This phenomenon also meant that IRT was best suited to detect disease at colder ambient temperatures, which has implications for all research using IRT to detect disease, as

previously the ambient temperature at which the animal was in their thermal neutral zone was widely considered to be the best temperature to detect disease. Due to the rigorous nature of the current research, results from this thesis could also be used to build an automated lameness detection unit for sheep. As IRT has been shown to be capable of detecting mild infection, an automated disease detection unit could lead to the infection being treated at an early stage thus preventing severe infection from occurring. With large scale and long term monitoring of hoof infection, it would be possible to identify animals that are genetically less susceptible to infection and they in turn could be used to breed the next generation of disease resistant animals.

The current thesis showed that it is possible to capture repeatable hoof load measurements from sheep. The designs of the hoof weigh crate in the current thesis can be used in future studies to capture automated measurements of sheep hoof load at drinking troughs or in a raceway. Furthermore, as the current thesis showed that a ewe's weight can be accurately estimated by only weighing their front hooves, a simplified weigh crate (a single load platform) could be set up at a drinking trough to monitor the growth of lambs or the body weight of ewes. The current thesis also demonstrated how hooves can react differently to mild infection depending on the body location, i.e., when front hooves contracted mild infection the ewe tended to remove a small amount of weight from those hooves, yet when back hooves contracted a mild infection ewes did not remove any weight from those hooves. This phenomenon has implication for research in many other species, currently, hoof weigh crates developed for cattle, pigs and horses only investigate how hoof load changes with severe infection rather than mild infection.

7.6 Future work

Future work may apply the techniques shown in paper 1 and 2 when using IRT to investigate other health issues or novel animal characteristics. The methodology used in paper 1 for testing the repeatability of thermal imaging in an agricultural environment should be used by future researchers to inform their imaging procedures. Additionally, when future researchers attempt to relate surface temperature of an animal to the gold standard of other infections, the variables extracted and statistical methods used in paper 2 of the current thesis could be applied.

For future work it is suggested to develop a system which could automatically capture thermal images of sheep hooves while animals drink from a water trough or walk through a raceway. Measurements captured while the animal is moving may vary by a greater amount and therefore, the repeatability of these measurements would need to be tested. This can be conducted with similar procedures as seen in paper 1 and paper 3 of the current thesis. As the current thesis has shown that IRT has the best diagnostic capabilities when hooves are subject to colder conditions, future work may consider using dynamic thermography techniques whereby a thermal image is taken of a sheep hoof before and after the application of a cold stimulus. The difference in temperature between the two thermal images can then be used as an indicator of infection. In theory, the infected hooves should stay warm while the healthy hooves should cool down.

Future work may also implement an automated hoof weigh crate system to detect severe lameness in sheep; such a system could be placed at a water trough or in a raceway. As the current thesis was the first to show that the hoof load of back hooves does not change in the same way as front hooves when a mild infection sets into the hoof, there is scope for other researchers to test this phenomenon in other species.

8 Appendix

- In section 3.6 the sentence “The present study indicates capturing three replicates of an anatomical region when using IRT for disease detection in cows.” should be rephrased to “The present study suggests that when IRT is being used to detect disease in cattle, the average of three thermal image replicates should be taken of the region of interest, rather than a single image.”
- In section 3.5.2 it states that “ the large between-cow variance observed for hoof images in the present study resulted in greater repeatability values associated with the hoof images compared with other anatomical areas” should be reworded to “the large between-cow variance observed for hoof images in the present study resulted in a higher proportion of the variance being attributed to the cow rather than error; this is favourable as temperature variations due to cow can be accounted for when using IRT as a diagnostic tool but variations due to error cannot”
- In section 5.5 and 6.1 the sentence “Lameness is one of the leading causes of morbidity in sheep” should be rephrased to “Hoof infection is the leading cause of ill health in sheep”
- In section 5.5.3 the sentence “When the same threshold was applied to a separate cohort of ewes (the validation dataset), the sensitivity deteriorated to 46%; this implies that the maximum hoof temperature cannot be used as an infection detection tool as random selection would achieve a sensitivity of 50%.” should instead say “When the same threshold was applied to a separate cohort of ewes (the validation dataset), the sensitivity deteriorated to 46%”
- In section 6.2 the sentence “Lameness is a serious welfare concern and leading causes of morbidity” should be “Hoof infection is a serious welfare concern and one of the leading causes of ill health”

- Throughout the document the “hoof temperature” may be rephrased to the “pastern temperature”