

Infrared thermography as a tool to detect hoof lesions in sheep

Daire T. Byrne,^{†,‡} Donagh P. Berry,[†] Harold Esmonde,[‡] Fiona McGovern,^{||}
Philip Creighton,^{||} and Nóirín McHugh^{†,1}

[†]Animal and Grassland Research and Innovation Centre, Teagasc, Fermoy, Cork P61 P302, Ireland; [‡]School of mechanical and manufacturing engineering, Dublin City University, Glasnevin, Dublin D09 Y5N0, Ireland; and ^{||}Animal and Grassland Research and Innovation Centre, Teagasc, Mellows Campus, Athenry, Galway H65 R718, Ireland

ABSTRACT: Lameness has a major negative impact on sheep production. The objective of this study was to 1) quantify the repeatability of sheep hoof temperatures estimated using infrared thermography (IRT); 2) determine the relationship between ambient temperature, sheep hoof temperature, and sheep hoof health status; and 3) validate the use of IRT to detect infection in sheep hooves. Three experiments (a repeatability, exploratory, and validation experiment) were conducted over 10 distinct nonconsecutive 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 2 d. The maximum and average temperature of each hoof was extracted from the thermal images. Repeatability

of IRT measurements was assessed by partitioning the variance because of 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 ($P > 0.05$), whereas 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: infrared thermography, lameness, sheep

© The Author(s) 2018. Published by Oxford University Press on behalf of the American Society of Animal Science. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Transl. Anim. Sci. 2018.XX:XX–XX
doi: 10.1093/tas/txy132

¹Corresponding author: Noirin.McHugh@teagasc.ie

Received August 16, 2018.

Accepted December 6, 2018.

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., 2019). One of the most common causes of lameness in sheep is foot rot (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 foot rot requires sheep to be turned over and each hoof visually assessed, which is labor intensive and difficult to implement across large numbers of flocks.

Infrared thermography (IRT) is a noninvasive 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 (Alsaad 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 this 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 this 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 inter- and intraherd temporal benchmarking.

MATERIALS AND METHODS

A series of experiments were undertaken in Athenry Research Centre, Teagasc, Athenry, Co. Galway, Ireland (53° 17' 15.3996'' N, 8° 46' 4.224'' W). 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 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). In addition, 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 foot rot (Table 1; Conington et al. (2008)).

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

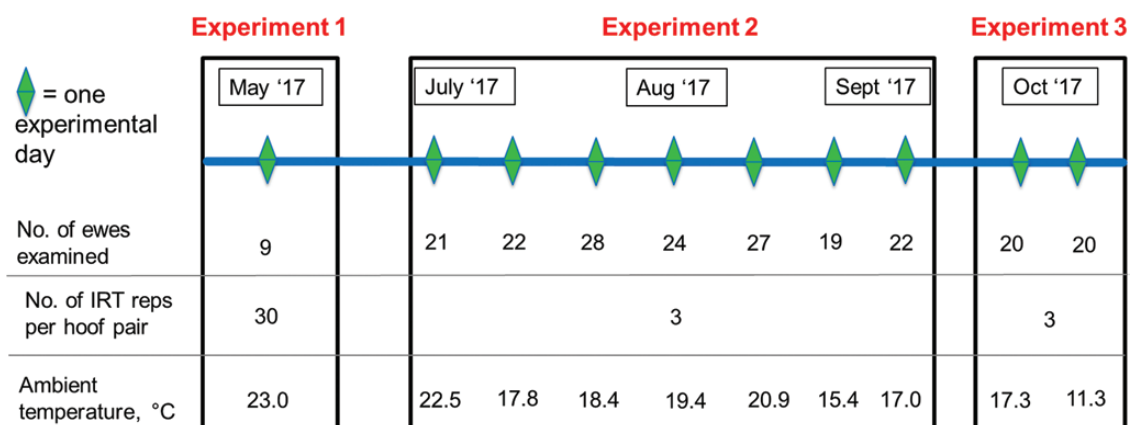


Figure 1. Timeline for all three experiments in this study is shown, with details of the ambient temperature, number of thermal image replicates captured per hoof pair (No. of IRT reps per hoof pair), and the total number of unique ewes (No. of ewes examined) used in each experiment.

Table 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)

Type of hoof lesion scale	Score	Definition*	Exploratory exp.		Validation exp.	
			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/foot rot	2	0.33	0	0.00
	4	Severe foot rot	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

Both the categorical and binary hoof lesion scales were derived from the hoof lesion scale from [Conington et al. \(2008\)](#).

*IDD = Interdigital dermatitis.

removed prior to analysis. Hooves that received a healthy hoof score (i.e., hoof lesion score = 0 from all operators) that 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. As a small number of category 3 ($n = 2$) and no category 4 hoof lesions were observed ([Table 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 2. 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 ([Alsaad 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 0 from one operator and 1 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 1 h after scoring. Following this, they were moved into the shed for a 30-min 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 320×240 , the thermal sensitivity was <0.03 $^{\circ}\text{C}$, and the accuracy was ± 2 $^{\circ}\text{C}$. All images were captured by the same operator in the same shed that 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 toward 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 later). 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 an EL-USB-2 data logger (Lascar electronics, Whiteparish, England).

Image analysis and temperature extraction were undertaken using the Thermovision LabVIEW Toolkit 3.3 (FLIR Systems Inc.) 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 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 dewclaws ([Figure 2](#)). The maximum and average temperature of each hoof was recorded. The 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.

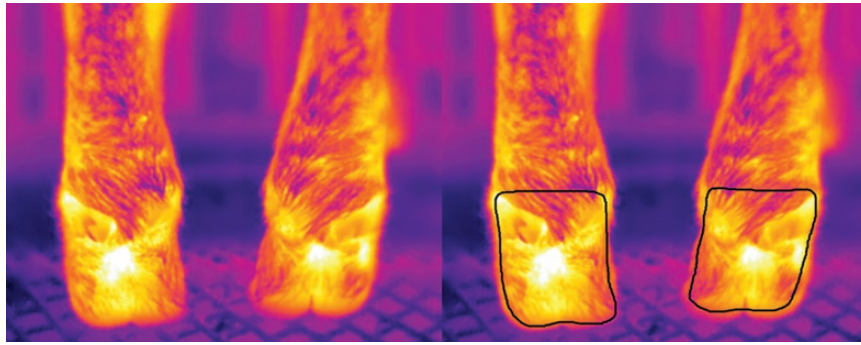


Figure 2. A thermal image of a pair of hooves before (left) and after (right) a freehand border was applied for temperature extraction is shown. The freehand border encompassed the posterior face of each hoof from below the coronary band to above the dewclaws.

A series of experiments were undertaken with the objectives of 1) quantifying the repeatability of IRT-estimated sheep hoof temperatures; 2) investigating the interrelationship between ambient temperature, hoof temperature, and hoof health status; and 3) validating the use of IRT as a tool to detect hoof infection.

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 10 s to capture.

Statistical analyses. The between-ewe and error variances for both maximum and average temperature of each of the four hooves were 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 fitted 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 estimated 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).

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, 7 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 1 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 2). Data from the control group and the lame group were combined for analyses.

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 or 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 2 ($n = 12$; Table 1). A multiple regression analysis was performed in SAS using PROC MIXED (SAS Institute, 2010), whereby either maximum or average hoof temperature was the dependent variable and binary hoof lesion score (0 = healthy hoof, and 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: 1) average hoof temperature, 2) maximum hoof temperature, 3) 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 4) 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 this 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).

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 fifth and sixth 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 2).

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 to 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.

RESULTS

Across the entire experiment period, ambient temperature ranged from 11.3 to 23.0 °C, whereas 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), whereas 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 three 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 three 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.

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 2](#)). The greatest between-ewe variance was observed in the right front hoof whereas 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 2](#)).

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 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 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.

Experiment 2: Exploratory Experiment

The frequency of each hoof lesion score is given in [Table 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 interdigital dermatitis (33.81 °C; $P < 0.05$) (i.e., categorical hoof lesion score = 1), but

Table 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 hooves) as well as the precision achieved with three images

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

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.

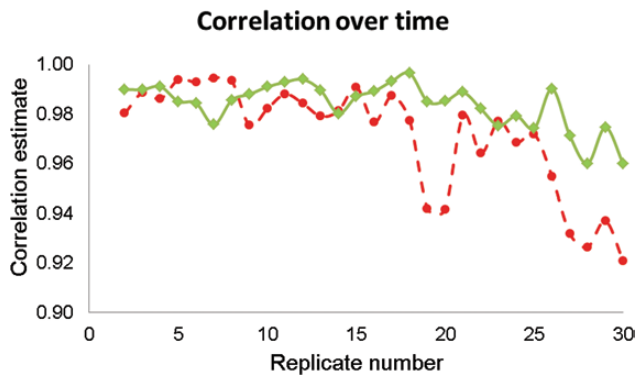


Figure 3. Correlation estimates between the right front hoof temperatures (average (in-filled diamond) and maximum (in-filled circle) temperature) recorded from the first thermal images of nine ewes and each of the subsequent replicate images (i.e., 2 to 30) are shown.

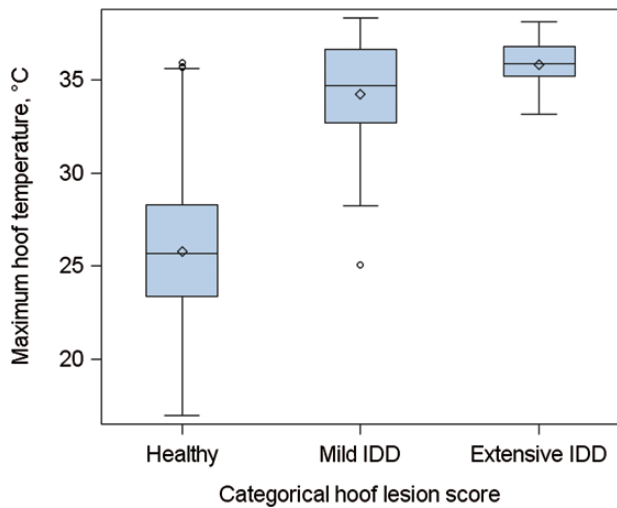


Figure 4. A box and whisker plot of the relationship between the maximum hoof temperature and the stages of interdigital dermatitis (IDD) (i.e., categorical hoof lesion score), where the mean (diamond outline), median (line within a blue box), interquartile range (IQR) (blue box), greatest (upper cap on error bars) and lowest (lower cap on error bars) values within $1.5 \times$ IQR of the IQR, and outliers (circle outline) of the maximum hoof temperature are shown.

this value did not differ when infection increased from mild to extensive interdigital 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 4. The mean \pm SE average hoof temperature for hooves with no lesions (i.e., score = 0), mild interdigital dermatitis (i.e., score = 1), and extensive interdigital 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 coefficient 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.

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 \pm 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 \pm 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 ($P < 0.001$). Although the maximum temperature of healthy hooves increased with ambient temperature (regression coefficient = 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. The least square mean \pm 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 ($P > 0.05$).

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 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.

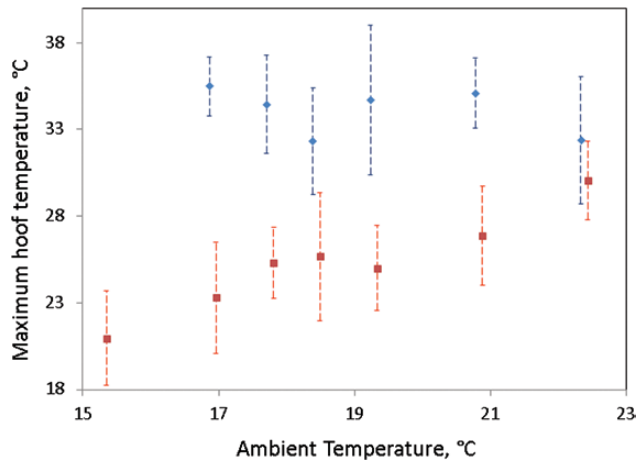


Figure 5. Mean and standard deviation of maximum temperatures for infected (in-filled blue diamond) and healthy (in-filled red square) hooves at various levels of ambient temperature.

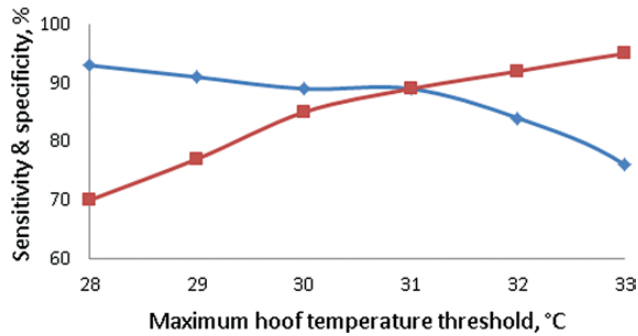


Figure 6. The percentage of hooves that were correctly classified as healthy (i.e., specificity (in-filled red square)) or infected (i.e., sensitivity (in-filled blue diamond)) when the threshold for maximum hoof temperature varied from 28 to 33 °C. Hooves with a maximum temperature above the threshold were considered infected, whereas all others were considered healthy.

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 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.

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 because of the labor requirement. The objective of this 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 that 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.

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 has 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; Alsaod and Büscher, 2012; Talukder et al., 2015) and so did not investigate the repeatability of IRT measurements of hooves. Results from this 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 this 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 this 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 $\pm 2\%$, indicating that replicate measurements did not improve the overall diagnostic ability, at least in this study.

Factors Associated With Hoof Temperatures

Hoof health status and ambient temperature were two major factors that were associated with hoof temperature. Although 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 ([Alsaad 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 interdigital 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 interdigital dermatitis (i.e., a hoof score of 1) and extensive interdigital 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 in hoof temperature averaged across the flock examined by [Talukder et al. \(2015\)](#) and this study was observed, despite differences because of 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 this 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 this study and that of [Talukder et al. \(2015\)](#); this study captured the palmar or planter face of each hoof as the animal was standing ([Figure 2](#)), whereas [Talukder et al. \(2015\)](#) restrained animals in a standing position, lifted each hoof, manually separated the toes, and captured images of the interdigital space. The interdigital 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. In addition, 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% ([Alsaad 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 this study, the maximum hoof temperature was derived from the same region where the median artery splits in two (below the dewclaws; [Figure 2](#)). Results from this 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 2](#)).

Diagnostic Capabilities of IRT

Although some studies have used hoof temperatures to identify hoof lesions in cattle (Alsaod and Büscher, 2012; Stokes et al., 2012), no study used IRT to diagnose foot rot in the individual hooves of sheep. In addition, no study has conducted a true validation on the diagnostic capabilities of IRT in any species; instead, many studies conduct a single exploratory experiment (Alsaod 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, whereas 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), whereas 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 (Alsaod and Büscher, 2012; Stokes et al., 2012; Alsaod et al., 2015). When the same threshold was applied to a separate cohort of ewes (i.e., 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 that 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 (Alsaod and Büscher, 2012; Stokes et al., 2012; Alsaod 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 coefficient 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 foot rot 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 that 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.

This 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). 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 whereas healthy hooves should cool down. This technique is known as dynamic thermography (Ohashi and Uchida, 2000).

CONCLUSION

The present study suggests that IRT has the potential to detect infection in sheep hooves, and as 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 that relate the temperature of an anatomical region to disease should be conducted at multiple levels of ambient temperature. In addition, 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.

ACKNOWLEDGMENTS

Funding from the Department of Agriculture, Food, and Marine Research Stimulus Fund (RSF 11/S/133) is gratefully acknowledged.

Conflict of interest statement. None declared.

LITERATURE CITED

- Alsaad, M., and W. Büscher. 2012. Detection of hoof lesions using digital infrared thermography in dairy cows. *J. Dairy Sci.* 95:735–742. doi:10.3168/jds.2011-4762
- Alsaad, 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 (Basel)*. 15:14513–14525. doi:10.3390/s150614513
- 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.* 176:521. doi:10.1136/vr.102882
- 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. doi:10.4141/A03-012.
- Bohan, A., L. Shalloo, P. Creighton, T. M. Boland, and N. McHugh. 2019. Deriving economic values for the Irish national sheep breeding objectives using a bio-economic model. In press: *Livest Sci.*
- Byrne, D.T., D.P. Berry, H. Esmonde, and N. McHugh. 2017. Temporal, spatial, inter-, and intra-cow repeatability of thermal imaging. *J. Anim. Sci.* 95:970–979. doi:10.2527/jas.2016.1005
- Church, J.S., P.R. Hegadoren, M.J. Paetkau, C.C. Miller, G. Regev-Shoshani, A.L. Schaefer, and K.S. Schwartzkopf-Genswein. 2014. Influence of environmental factors on infrared eye temperature measurements in cattle. *Res. Vet. Sci.* 96:220–226. doi:10.1016/j.rvsc.2013.11.006
- Conington, J., B. Hosie, G.J. Nieuwhof, S.C. Bishop, and L. Bünger. 2008. Breeding for resistance to footrot—the use of hoof lesion scoring to quantify footrot in sheep. *Vet. Res. Commun.* 32:583–589. doi:10.1007/s11259-008-9062-x
- 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. Anim. Biosci.* 1:340. doi:10.1017/S2040470010004838
- Dohoo, I.R., R.A. Curtis, and G.G. Finley. 1985. A survey of sheep diseases in Canada. *Can. J. Comp. Med.* 49:239–247.
- Fitzpatrick, J., M. Scott, and A. Nolan. 2006. Assessment of pain and welfare in sheep. *Small Ruminant Res.* 62:55–61. doi:10.1016/j.smallrumres.2005.07.028
- 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.* 64:123–130. doi:10.12681/jhvms.15485
- Gilmour, A., B. Gogel, B. Cullis, and R. Thompson. 2009. *ASReml user guide release 3.0*. Hemel Hempstead (UK): VSN International Ltd.
- 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. doi:10.1186/1746-6148-7-73
- Greiner, M., D. Pfeiffer, and R.D. Smith. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev. Vet. Med.* 45:23–41. doi:10.1016/S0167-5877(00)00115-X
- 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. In: *Proceedings from the New Zealand Society of Animal Production*; Christchurch, New Zealand. p. 117–122.
- Kaler, J., and L.E. Green. 2008. Recognition of lameness and decisions to catch for inspection among sheep farmers and specialists in GB. *BMC Vet. Res.* 4:41. doi:10.1186/1746-6148-4-41.
- 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., 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. doi:10.1016/j.rvsc.2012.10.021.
- 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. Mag.* 19:42–51. doi:10.1109/51.844379
- 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. doi:10.1016/j.tvjl.2008.01.003
- SAS Institute. 2010. SAS/STAT software, release 9.3. Cary (NC): SAS Institute Inc.
- Schaefer, A.L., N.J. Cook, C. Bench, J.B. Chabot, J. Colyn, T. Liu, E.K. Okine, M. Stewart, and J.R. 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. doi:10.1016/j.rvsc.2011.09.021

- Stewart, M., K.J. Stafford, S.K. Dowling, A.L. Schaefer, and J.R. Webster. 2008. Eye temperature and heart rate variability of calves disbudded with or without local anaesthetic. *Physiol. Behav.* 93:789–797. doi:10.1016/j.physbeh.2007.11.044
- Stokes, J.E., K.A. Leach, D.C. Main, and H.R. 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. doi:10.1016/j.tvjl.2012.06.052
- 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. doi:10.1016/j.smallrumres.2015.04.006
- 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. doi:10.1016/j.jevs.2012.03.005