

**Automatic evaluation of infrared thermal images of
bovine udders and teats challenged with *E. coli***

von Sophie Paloma Watz

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München

**Automatic evaluation of infrared thermal images of
bovine udders and teats challenged with *E. coli***

von Sophie Paloma Watz

aus Gießen

München 2018

Aus dem Zentrum für Klinische Tiermedizin der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München
Lehrstuhl für Physiologie und Pathologie der Fortpflanzung

Arbeit angefertigt unter der Leitung von:

Univ.-Prof. Dr. Holm Zerbe

Mitbetreuung durch:

Dr. Moritz Metzner

Dr. habil. Wolfram Petzl

**Gedruckt mit der Genehmigung der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München**

Dekan: Univ.-Prof. Dr. Reinhard K. Straubinger, PhD
Berichterstatter: Univ.-Prof. Dr. Holm Zerbe
Korreferent: Univ.-Prof. Dr. Dr. Michael H. Erhard
Univ.-Prof. Dr. Mathias Ritzmann
Univ.-Prof. Dr. Dr. h.c. Erwin P. Märtlbauer
Univ.-Prof. Dr. Rolf Mansfeld

Tag der Promotion: 10.02.2018

Meiner Familie; Felix und Perle

„From my grandmother I learned that logic is relative.“

-John Irving

Content

I.	LIST OF ABBREVIATIONS.....	1
II.	INTRODUCTION	3
III.	LITERATURE SURVEY	5
1.	Physiological udder and body temperature of dairy cows	5
2.	Mastitis in dairy cows	6
2.1.	Economic impact of mastitis.....	6
2.2.	Pathogen-specific mastitis	7
3.	IRT measurements in living organisms.....	9
3.1.	Reproducibility of results provided by IRT	10
3.2.	Optimal conditions for IRT application on cattle	11
4.	Application of IRT in human medicine	12
5.	Application of IRT in buiatrics	12
5.1.	Application of IRT in detecting systemic alterations in cattle.....	13
5.2.	Application of IRT on udder and teats of dairy cows.....	13
6.	Automatic image recognition in mastitis detection.....	20
IV.	MATERIAL AND METHODS.....	22
1.	Experimental animals and thermographic material.....	22
1.1.	Conditions for experimental animals.....	22
1.2.	Housing and preparation of the animals	23
1.3.	Intramammary challenge	23
1.4.	Schedule of measurements.....	23
1.5.	Procedure of measuring	24
2.	Evaluation of thermographic material.....	27
2.1.	Evaluation by an automated image recognition software.....	27
2.2.	Manual evaluation of thermographic material.....	29
2.3.	Evaluation of the region of the teats	30
2.4.	Evaluation of thermographic material without the region of the teats	31

Content	II
<hr/>	
2.5. Documentation of obtained data	32
3. Statistical methods	33
3.1. Testing Gaussian distribution of the data	33
3.2. Precision of the different methods	33
3.3. Correlation of evaluation method and rectal temperature	35
3.4. Comparison of automatic and manual evaluation.....	36
3.5. Comparison of evaluation using the teats and evaluation using the udder	39
V. RESULTS	41
1. Testing Gaussian distribution of the data.....	41
2. Results of automatic and manual evaluation methods	41
2.1. Precision of automatic and manual evaluation	42
2.2. Correlation of evaluation method and rectal temperature	45
2.3. Comparison of automatic and manual evaluation.....	52
3. Results of evaluation method using only the teats and evaluation method using the udder surface.....	66
3.1. Precision of evaluation using the teats and evaluation using the udder..	66
3.2. Correlation of evaluation method and rectal temperature	69
3.3. Comparison of evaluation using the teats and evaluation using the udder	76
VI. DISCUSSION	92
1. Impact of measurement conditions	92
2. Impact of physiological oscillation in body temperature on IRT measurement results	94
3. Limitations of the different evaluation methods and their ROI's	94
4. Interpretation of results of automatic and manual evaluation method.....	96
4.1. Precision analysis.....	96

4.2. Results of automatic and manual evaluation and their correlation with rectal temperature.....	98
4.3. Comparison of automatic and manual evaluation.....	102
5. Interpretation of results of evaluation using the teats and evaluation using the udder and excluding the teats	109
5.1. Precision analysis.....	109
5.2. Results of evaluation using the teats and evaluation using the udder and their correlation with rectal temperature	110
5.3. Comparison of evaluation using the teats and evaluation using the udder	112
6. Conclusions	118
VII. SUMMARY.....	122
VIII. ZUSAMMENFASSUNG.....	125
IX. BIBLIOGRAPHY	129
X. APPENDIX.....	136
1. Appendix 1: Tables of data obtained by automatic and manual evaluation .	136
2. Appendix 2: Tables of data obtained by evaluation using the teats and evaluation using the udder	142
3. Appendix 3: Tables and Histograms of normality tests.....	148
4. Appendix 4: Tables of precision analysis	153
5. Appendix 5: Figures of method comparison.....	154
6. Appendix 6: Graphs of ROC analysis	160
XI. DANKSAGUNG	164

I. LIST OF ABBREVIATIONS

95% CI	95% confidence interval
a.m.	ante meridiem (before noon)
AR01	ROI selected firstly with the help of a polygon-tool
AR02	ROI selected secondly with the help of a polygon-tool
AUC	area underneath curve in ROC-analysis
‘Aut’	automatic evaluation using the hind surface of the hind udder quarters, including the teats
‘Avg’	average temperature value inside ROI
Cf.	confer (compare)
CFU	colony forming units
CMT	California Mastitis Test
<i>E. coli</i>	<i>Escherichia coli</i>
e.g.	exempli gratia (for example)
h	hour; hours
HL	left hindquarter
HR	right hindquarter
Hz	hertz
IRT	infrared thermography
IU	international units
K	Kelvin
LPS	lipopolysaccharides
LTA	lipoteichoic acid
m	meter
‘Man’	manual evaluation using the hind surface of the hind udder quarters, including the teats
‘Max’	maximum temperature value inside ROI
‘Max-Min’	range of temperature values inside ROI
‘Min’	minimum temperature value inside ROI
mL	milliliter

n	sample size
one-way ANOVA	one-way analysis of variances
p.i.	post infectionem (after infection)
p.m.	post meridiem (after noon)
q1	first quartile
q3	third quartile
r	coefficient of correlation
ROC	Receiver-Operating-Characteristics
ROI	region of interest
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
SCC	somatic cell count
SD	standard deviation
SST	skin surface temperature
‘Stdev’	standard deviation of temperature values inside ROI (given by ThermaCAM Researcher Pro 2.8 software)
‘Teats’	evaluation method using exclusively the hind surface of the hind teats, manually selected
‘Udder’	evaluation method using exclusively the hind surface of the hind udder quarters and excluding the teats, manually selected
USST	udder skin surface temperature
VC	coefficient of variation
vs.	versus
°C	Degrees Celsius
ε	emissivity
μm	micrometer
%	Percent

II. INTRODUCTION

Mastitis is one of the most frequent diseases in dairy cows, resulting in economic losses, culling and not least suffering of affected cows (DVG, 2002; BAR et al., 2008). Standard approaches in diagnosing mastitis are California Mastitis Test (CMT), somatic cell count (SCC) and bacteriological examination of milk samples. These techniques provide reliable results, especially when they are combined, but they are invasive, require a skilled examiner and can thus not be implemented several times throughout the day in large dairy cow herds.

Due to growing herd sizes in modern dairy livestock farming, there is a need for automated health monitoring. Alongside with prevention of diseases, it is of particular importance to detect health disorders at an early stage, in order to treat them effectively.

Mammals, as homoeothermic beings, keep their body temperature in a narrow range. Largest part of heat loss is due to the body surface's radiation, followed by natural convection. Therefore, changes of inner body temperature or body surface temperature lead to altered infrared radiation (JOYCE et al., 1966).

Changes in body temperature can be a reliable indicator for pathological conditions in homoeothermic individuals.

Inflammation leads to vasodilation and increased microcirculation in the affected tissue, respectively in the skin nearby. Skin surface temperature (SST) can be measured with the help of infrared thermography (IRT). An increase in SST due to inflammation can be detected in repeated measurements, for instance in udder skin of cows infected with *E. coli* (METZNER et al., 2014; METZNER et al., 2015) as well as in body surface skin of animals suffering from systemic infections (HURNIK et al., 1984), (RAINWATER-LOVETT et al., 2009), (SCHAEFER et al., 2007).

IRT can be suitable for diagnosing *E. coli* mastitis at an early stage and has the advantage of being noninvasive (METZNER et al., 2015). Nevertheless, evaluation of infrared imaging is currently done manually by a trained person and is thus very time-consuming. Although technical equipment has become more affordable over the last decades, with the recent status of evaluation, IRT is not yet operational for application in automated health monitoring.

Computer-assisted systems could be a time-saving and thus economically profitable way of supervision, and are thereby helpful in enhancing animal welfare. SCC and measurement of electrical conductivity of milk embedded in robotic milking can monitor udder health status automatically. Nevertheless, milking only takes place a restricted number of times during the day. An infrared camera in the stable area, connected to computer-assisted recognition of the udder and evaluation of surface temperature could possibly detect pathological changes hours earlier, and is thus a promising approach in automated udder health supervision.

Previous studies using IRT in dairy cows reported that in the udder region, warmest areas ('hot spots') are typically located in udder-thigh cleft and the intermammary sulcus (GLAS, 2008; HOVINEN et al., 2008; METZNER et al., 2015). In these areas, body surfaces are approaching each other in a sharp angle, thereby radiating in the infrared spectrum. Temperature of these areas probably depict inner body temperature and not udder surface temperature. Exclusion of 'hot spots' from evaluation of infrared images could possibly provide better results in mastitis detection.

Teats are the first immunological barrier of the udder and consequently show the first defense mechanisms in intramammary infections. Still, they have not yet been examined by IRT in detail when cows were infected intramammary.

In order to establish reliable systems for automated udder health monitoring in the future, this study concerns following issues:

- The current gold standard of mastitis detection using IRT is manual evaluation. Does an automated image recognition software provide comparable results?
- Is an earlier detection of clinical mastitis possible if only the surface of the teats is selected in manual interpretation?
- Can a standardized exclusion of 'hot spots' from evaluation lead to better results in detecting mastitis?

III. LITERATURE SURVEY

1. Physiological udder and body temperature of dairy cows

In order to detect pathological alterations in inner body temperature and skin surface temperature of dairy cattle, profound knowledge about physiological circumstances is essential.

The physiological udder skin surface temperature (USST) and the patterns of its variations were evaluated by BERRY et al. (2003). Ten healthy cows were observed in a period of time of eight weeks. The animals were held indoors with ambient temperatures between 11.1°C and 27.4°C and let out for two hours per day. Rectal and ambient temperatures were monitored throughout the trial. In the first part of the study, infrared images were taken before and after outdoor exercise for 31 days. Udder surface temperature, but not rectal temperature rose significantly after exercise. In the second part of the study, potential circadian rhythms of the udder surface temperature should be detected by taking infrared images of the hindquarters every two hours on four days. Mean surface temperatures of all cows were taken into consideration. The authors claim to have found notable daily variations in repeated patterns, thus circadian rhythms in udder surface temperature for mean values of all cows. Udder surface temperature was lowest in the early morning hours and rose during the day, peaking in the afternoon, differing between 2.0°C and 3.5°C per day, whereas rectal temperature showed little variation. However, it has to be said, that no correlation of udder surface temperature and ambient temperature has been calculated. The curve of ambient temperature showed distinct similarities and had a range of 2.5°C. The authors point out the possibilities of infrared thermography in mastitis detection, underlining the importance of short intervals between the measurements.

When GLAS (2008) recorded udder-surface temperature and rectal temperature of 16 lactating cows, each for 24 hours, only slight variations could be detected: rectal temperature variation ranged for $0.4^{\circ}\text{C} \pm 0.1^{\circ}\text{C}$, mean udder surface temperature ranged for $0.5^{\circ}\text{C} \pm 0.1^{\circ}\text{C}$. Still, the progress of udder surface temperature showed similar patterns as described by BERRY et al. (2003): Temperature was slightly lower than mean values in the early morning and slightly higher in the evening hours. Animals were held indoors, nevertheless environmental temperature decreased at night. A strong correlation could be found between rectal and udder surface temperature, whereas only a low but still significant dependence of

environmental temperature on udder surface temperature could be detected. FRANZE et al. (2012) also described udder surface temperature to be 0.2K to 0.4K lower in the morning hours and calculated a coefficient of correlation of 0.6 to environmental temperature.

Repetitive biphasic circadian patterns were observed in a study of BITMAN et al. (1984). Six healthy cows have been equipped with temperature sensors in the abdominal cavity and in the udder tissue. Thus, inner udder temperature and not udder surface temperature was measured. Temperature was recorded in intervals of 1.4 minutes in a period of five days, environmental temperature was held constantly at $16.7 \pm 0.3^{\circ}\text{C}$. Inner body temperature and udder temperature were almost equal. In this trial, animals were held outdoors twice per day as well. Unlike in the study of BERRY et al. (2003), body temperature declined about 1.0°C in the average when cows were outside. However, it has not clearly been stated how much space was available for the cows to exercise outdoors. Outdoor-temperature varied between 5.2 and 17.4°C . In four of six cows, biphasic circadian rhythms were detected, with variation of temperature ranging for 1.25°C . Body and udder temperature fell at 9:30 a.m. and 20:00 p.m. and temperatures were peaking from 13:00 p.m. to 20:00 p.m. and from 23:00 p.m. to 8:30 a.m. Although temperature minima were clearly associated with the exercise periods outdoors, some animals showed lower temperatures before they were let out.

Furthermore, ultradian rhythms with a time span of 90 minutes on the average and an amplitude of 0.41°C could be observed in all cows. However, the studies of BITMAN et al. (1984), BERRY et al. (2003) and GLAS (2008) support the assumption that circadian temperature patterns are not inherited by dairy cows, but rather a result of environmental factors.

2. Mastitis in dairy cows

Mastitis is the inflammatory response of the mammary gland. In dairy cows, it is mainly caused by intramammary bacterial infection. The presence or absence of evident inflammatory symptoms allows to distinguish between clinical and subclinical mastitis (SCHUKKEN et al., 2011)

2.1. Economic impact of mastitis

Clinical and subclinical mastitis generates substantial economic losses worldwide, however, the estimated costs diverge significantly, since numerous factors have to

be taken into consideration. Total losses are composed of reduced milk yield, the prohibited distribution of pathologically altered milk, increased mortality and higher replacement costs as well as costs for improving milking hygiene, veterinary consultation and medication (HAMANN & GRUNERT, 1998; DVG, 2002; BAR et al., 2008; ROLLIN et al., 2015; VAN SOEST et al., 2016)

BAR et al. (2008) estimated the average cost of clinical mastitis to be \$179 per case of clinical mastitis in five dairy farms in New York State which corresponded to \$71 per year and cow respectively. A statistical survey in 60 randomly selected dairy farms in Michigan (KANEENE & HURD, 1990) identified clinical mastitis as the most expensive disease in adult cows, causing a loss of \$35,54 per cow and year, which did not include costs of \$4,56 spent for mastitis prevention. Statistical analysis using a model of a typical US dairy farm of ROLLIN et al. (2015) exceeded formerly estimated costs by far: costs of \$444 per case of clinical mastitis in the first 30 days of lactation were calculated. Average costs of mastitis per cow and year were not given in this study, whereas VAN SOEST et al. (2016) estimated costs per cow and year to be €240 in Dutch dairy farms, prevention accounting half of the costs.

2.2. Pathogen-specific mastitis

The pathogen-specific immune response and thus the clinical symptoms of mastitis differ significantly depending on the etiological agent (PETZL et al., 2008). *Escherichia coli* mostly causes cases of clinical mastitis, which may become a severe threat to the cow. *E. coli*-mastitis is often accompanied by fever, rapid drop in milk yield, general depression, recumbence and septicemia (BURVENICH et al., 2003). In the majority of cases, a significant increase of the SCC can be observed (FROST et al., 1982). However, host responses vary considerably from case to case and there are no pathogen-specific symptoms only linked to *E. coli*-mastitis (SCHUKKEN et al., 1989). Different studies have revealed, that there is no homogenous pattern of virulence factors found in *E. coli* strains isolated from cases of clinical mastitis (SANCHEZ-CARLO et al., 1984; SUOJALA et al., 2011). ZHANG and WANG (2010) detected a higher amount of mast cells in glandular udder tissue of cows suffering from clinical mastitis compared to healthy udder tissue. The authors outline the role of mast cells in vasodilation caused by inflammation and thus increased microcirculation in clinical mastitis.

BANNERMAN et al. (2004) compared the responses of the innate immune system to induced intramammary infection with either *Staphylococcus aureus* or *E. coli*. The study revealed that higher levels of the proteins interleukin 1beta, gamma interferon, interleukin 12, soluble CD14 and lipopolysaccharide-binding-protein could be measured in the milk after inoculation with both, *E. coli* and *S. aureus*, whereas the milk levels of interleukin 8 and tumor necrosis factor alpha only significantly rose after *E. coli*-challenge. The authors assumed a correlation between the specific responses and the following clinical symptoms induced by both pathogens. The observation that the absent or reduced induction of inflammatory factors in the case of *S. aureus*-mastitis compared to *E. coli*-mastitis may lead to subclinical mastitis with chronic infections has since then been confirmed in various studies (PETZL et al., 2008; YANG et al., 2008; GILBERT et al., 2013; JENSEN et al., 2013).

CHANG et al. (2015) described the systemic inflammatory reaction during *E. coli* mastitis to increased transcription of genes encoding TLR2, TLR4 and lipopolysaccharide-binding protein, induced by *E. coli* endotoxins. GILBERT et al. (2013) compared transcriptomes of bovine mammary epithelial cells, facing either *E. coli*-lipopolysaccharides (LPS) or *S. aureus* supernatant. Both groups shared the stimulation of some genes, including genes encoding TLR2, but a significant higher number of genes was transcribed due to *E. coli*-endotoxins. Also, TLR4 was only activated in the *E. coli*-group. In contrast, PETZL et al. (2008) did not find higher transcription of TLRs in mammary tissue and lymph nodes after infection with 10.000 CFU of *S. aureus*.

A transcriptome study of MITTERHUEMER et al. (2010) provided further explanation for the systemic response: infected quarters of dairy cows showed 2154 differently expressed genes 18 hours after intracisternal inoculation with *E. coli*, compared to a control group. The neighboring quarters showed no signs of clinical mastitis, but 476 differently expressed genes could be found at that time. Transcriptome changes in bacteriologically sterile quarters next to quarters infected with *E. coli* are also reported by JENSEN et al. (2013).

The teat orifice is the key barrier between the udder and its environment. Thus previous studies stressed its importance as the first line of defense against invading pathogens (NICKERSON & PANKEY, 1983). Recently new techniques of transcriptomic profiling revealed that the initial response towards pathogenic

threats was generated in the teat (RINALDI et al., 2010). This was true during *E. coli*- and *S. aureus* induced mastitis (PETZL et al., 2016), however the pathogen-specific extent of the inflammatory response was invariantly detectable. Furthermore LIND et al. (2015) showed that the tissue response of the Fürstenberg's rosette and the teat cistern towards either LPS of *E. coli* or lipoteichoic acid (LTA) of *S. aureus* was significantly higher to LPS than to LTA.

Influence of LPS on udder tissue temperature and on inner body temperature were displayed in a study of LEFCOURT et al. (1993). Values obtained by radiotelemetry showed that udder and body temperature were almost equal throughout the trial, an observation that was supported by BITMAN et al. (1984). Intramammary challenge with *E. coli* endotoxin increased udder temperature significantly between 2.75 and 9.75 hours after challenge, peaking at 6.5 hours. Mastitis could be diagnosed in all cows at the next milking, 12 hours after injection.

3. IRT measurements in living organisms

Numerous factors influence the outcome of measurements using IRT. In order to gain reliable results, research has been done on the special requirements of examinations using IRT in living objects.

Diverse values exist for the emissivity of human and animal skin. Emissivity tables of FLIR Systems set the emissivity of human skin at 0.98. STEKETEE (1973) calculated the emissivity of human skin as 0.989 ± 0.01 and stated that it is independent from pigmentation under standardized conditions of examination. The emissivity's independence of skin pigmentation was also approved in a study of MITCHELL et al. (1967a). These authors measured the emissivity of excised human skin as 0.996 ± 0.005 and thus close to the emissivity of a blackbody (MITCHELL et al., 1967b). WATMOUGH et al. (1970) claimed that the emissivity of human skin decreases significantly, the more the angle of the detector deviates from 90° in relation to an object's surface. The authors assumed that warmer regions could possibly not be detected in areas of round objects, since the surface has an inconvenient angle towards the detector.

The special challenges of thermographic measurements of animals were described by CENA and CLARK (1973b). Unlike human skin, the surface of animals is mostly covered with coat. It has been shown that profound knowledge of the density

of the coat and also of hair diameter, structure and length are necessary to provide accurate results of temperature distributions on the skin.

In another publication, the same authors remark the influences of maculation on thermographic investigations. Under intense solar radiation, up to 9°C warmer temperatures could be measured in regions of black markings, compared to white regions of the coat. This effect concerns dairy cows as well as non-domestic animals like zebras, and underlines the importance of implementing IRT measurements in an environment without direct solar radiation (CENA & CLARK, 1973a).

Since mammals do not only differ in fur and coloring, MORTOLA (2013) tested the hypothesis that body mass has an influence on body surface temperature. As stated in Bergmann's rule, smaller individuals have more body surface in relation to body volume than larger individuals (BERGMANN, 1848). Infrared thermograms of 37 different species of mammals, ranging from mouse to African elephant, were taken under standardized conditions and indoors. Three different gradients of ambient temperature (20-22°C, 22-25°C and 25-27°C) were compared. The evaluation could not detect significant differences in average body surface temperature of the different species. However, the author did not point out how exactly the inner body temperature of each individual was taken into account. In addition, average surface temperature was measured irrespective of hair coat and structure. Thus, body surface temperature and not skin surface temperature was examined. Emissivity was set at 0.95 in this study.

3.1. Reproducibility of results provided by IRT

ZAPROUDINA et al. (2008) discussed the question how far infrared thermography is a reliable technique when examiner and time-points of measurements vary. For this study, sixteen healthy human individuals were thermographically examined by two observers on two consecutive days. 45 different regions of interest on each individual were taken into account. The two observers took their images in the same laboratory and under the same conditions with only a short time-interval of ten minutes. Intra class correlations for inter-examiner was 0.88, which depicts a good reproducibility of measurements.

Nevertheless, intra class correlation of two consecutive days was only 0.47, underlining individual variations of body surface temperature and limitations of repeated measurements. The authors admit that twenty minutes of acclimation in

the laboratory may have been too short, due to cold climate in Finland, where the examination took place. Moreover, it is not apparent from the article how examiners assured that regions of interest were repeatedly set in the exact same location.

3.2. Optimal conditions for IRT application on cattle

In order to create standardized conditions of examination for best possible results of thermal imaging in bovine medicine, OKADA et al. (2013) evaluated several influences on measuring surface temperatures. Thermograms of four lactating Holstein cows were taken in a room with constant ambient temperature of 25°C and without sun exposure. Measurements were repeated several times, but not at different days.

The factors to be tested were the distance between object and infrared camera, the camera's angle towards the object, the influence of self-heating of the camera, respectively external thermal influences on the camera, ambient air movement and humidity, the influence of fur and time of acclimation in the examination room at different temperatures.

The authors stated that raising distances between camera and object leads to measuring significantly lower surface temperatures: 26.64°C ± 0.05 at a gap of 0.5m and 26.26°C ± 0.05 at 3.0m on the average. Not an ideal distance was recommended, but choosing the same distance in repeated measurements. When distances between camera and object were named in other studies, they ranged between 0.5m (POLAT et al., 2010) and 1.8m (METZNER et al., 2014).

In comparing the camera's angle towards the object, no significant differences between an angle of 90° and 45° could be found. No lower angles were evaluated, although it would be of interest, according to WATMOUGH et al. (1970). No influence of heating of the camera after long operating time was found either, but challenging the camera with either hot or cool packs led to an increase or decrease of measured surface temperature values, respectively. This underlines the necessity to protect the camera from extreme temperature conditions.

Air movement and a humidifier in the examination room both resulted in slightly, but significantly higher measurement results. Both factors and thus the influence of cooling by evaporation were not tested in combination. The influence on the outcome of the measurement whether the fur was shaved or not was evaluated: when ambient temperature was set at 15°C, shaved skin was 5.6°K cooler than

haired skin. When cows were brought to a room with ambient temperatures of 4°C, the difference was 9.33°K when images were taken immediately and 8.6°K after 30 minutes. Acclimation time of 30 minutes did not have a significant effect on rectal temperature and on surface temperature of unshaved skin, but surface temperatures of bare skin decreased significantly.

Although standardized conditions of examination as they are common in human medical imaging (ZAPROUDINA et al., 2008) are preferable, it is hardly possible to implement them all in health monitoring in an agricultural environment. Nevertheless, for the purpose of generating reliable results of repeated measurements, conditions should be defined in the best possible way.

4. Application of IRT in human medicine

The first application of thermography as a diagnostic tool took place in 1956 when Ray Lawson detected temperature rises in tumor regions in 26 proven breast-cancer patients. In this study evaporography was used, a simple technique to depict infrared radiation into an image. Although this method only gives a vague impression of the affected regions, Lawson was able to identify carcinoma in another patient using thermography. The diagnosis was later confirmed by histopathology (LAWSON, 1956).

In the last decades, thermography has become an established tool in human medicine.

JIANG et al. (2005) and LAHIRI et al. (2012) both give detailed reviews about fields of application and prospective developments of infrared thermography in human medicine. It is currently mainly used in detecting diseases attended by vascular alterations and in oncological screening.

Reviews of NG (2009) and KENNEDY et al. (2009) outline the recent role of thermography in breast cancer screening and diagnosis. Technical advances made it possible for thermographic measurements to offer a high rate of detecting breast cancer and are thus used alongside Mammography and Clinical Breast Examination.

5. Application of IRT in buiatrics

As mentioned above, IRT is used as a helpful tool in human medicine, due to its ability of detecting changes in surface temperature, mostly resulting from

inflammation. It is thus reasonable to use this advantage in veterinary application. Inflammation, vascularization and skin surface temperature are closely linked. For instance, infrared thermography depicts increased microcirculation of the skin due to hot-iron or freeze branding induced inflammation in cattle as significantly warmer regions (SCHWARTZKOPF-GENSWEIN & STOOKEY, 1997). In this study, increased skin temperature after branding could be measured for the whole trial period of 168 hours.

5.1. Application of IRT in detecting systemic alterations in cattle

Numerous publications verify the use of infrared thermography as a tool to detect systemic diseases in cattle before clinical symptoms appear (HURNIK et al. (1984), RAINWATER-LOVETT et al. (2009), SCHAEFER et al. (2007)). SCHAEFER et al. (2004) found significant changes in facial surface temperature in calves infected with bovine viral diarrhoea virus days before other clinical parameters changed.

Currently, research is also done on the possibilities of infrared thermography to detect estrus and ovulation in cows. Although sensitivity and specificity are not sufficient yet, IRT can become a valuable part of estrus detection, if it is combined with observation of other parameters (TALUKDER et al., 2014), (TALUKDER et al., 2015).

5.2. Application of IRT on udder and teats of dairy cows

HAMANN and DUCK (1984) investigated teat surface temperature and the influence of milking using infrared thermography. Manual massage for 30 seconds reduced skin temperatures of the teat about 0.8K on the average. Milking with a conventional liner led to increased blood flow in the tip of the teat, thus, significantly higher surface temperatures were measured here. Nevertheless, this effect was compensated by reduced circulation in the teat base, detectable as 1.3°C lower temperature on the average directly after milking. These observations of influence of conventional milking liners on teat temperature were largely supported by (PAULRUD et al., 2005), underlining the suitability of infrared thermography combined with ultrasonography to evaluate teat health conditions.

5.2.1. Application of IRT in detecting mastitis in dairy cows

Surface temperature of the teats was also of interest in the study of BARTH (2000). Six cows were observed for eight days: Thermograms were taken twice per day before milking, milk samples for somatic cell count were taken simultaneously.

Surface temperature was 30.1°C on the average for the teat tip and 35.1°C for the base, respectively. Teats of quarters with cell concentration below 100.000/ml showed a significantly lower surface temperature (33.6°C) than quarters above 100.000 cells/ml (34.1°C). The author noted significant differences in values obtained from different directions (medial, caudal and lateral). A trial with shorter intervals between measurements would be of interest.

SCOTT et al. (2000) induced clinical mastitis in 20 cows by injection of *E. coli* endotoxin in the left hindquarter. Thermograms were taken every hour in the beginning of the trial, later intervals of 3 hours were chosen, followed by 12-hour-intervals. Rectal temperature, somatic cell count and bovine serum albumin in milk samples were recorded for comparison. Infrared thermography was able to detect significant increased mean udder surface temperature between 1h and 24h after infection, whereas rectal temperature did not rise until 6h p.i., and returned to normal 9h p.i. No statement was made about the exact region of the udder surface that was evaluated.

New findings about the influence of *S. aureus*-Mastitis on udder surface temperature provided the study of SCHUTZ et al. (2001). Throughout the trial of nine days, infrared thermograms were taken in an interval of five hours, no further explanation was given on the selected region. Intramammary infection in one forequarter with *S. aureus* took place on day five, as a result three of four animals showed signs of clinical mastitis. Somatic cell count and measurement of rectal temperature was implemented as well. Skin surface temperature of the udder showed a good correlation with rectal temperature ($r=0.58$). Coefficient of correlation with somatic cell count was 0.33. Although significantly higher udder surface temperatures could be detected in affected quarters compared to healthy quarters, the authors question advantages of infrared thermography, since they doubt that this technique leads to earlier diagnosis than established methods like somatic cell count.

GLAS (2008) not only tested the ability of infrared thermography to detect experimentally induced *E. coli* mastitis, but also took the existing uncertainties about the evaluation of thermograms of the udder into account. Thermograms of the hind udder quarters were taken every two hours during the trial and cows were challenged with *E. coli* in the right hindquarter after 24 hours. In order to interpret the images, polygons of the quarters were manually marked and average and

maximum temperature within these polygons was used for evaluation. Minimal temperatures seemed to depict extreme values and outliers and were considered as not useful for evaluation. In the temperature course throughout the trial, significant rising of surface temperature of both, infected and healthy quarter, was detectable, starting 7 hours after infection. A comparison with the day before infection showed that maximally higher temperatures were measured 13 hours after *E. coli* challenge for both quarters. Surface temperature stayed significantly elevated until 17 hours after infection. However, the amplitude of temperature elevation was lower in the quarter challenged with *E. coli*. For instance, mean temperature of the right side was 0.89°C cooler than for the left side 13 hours after infection. Edematous swelling in the course of mastitis was suspected as a reason. Comparison of maximum surface temperatures of the both hindquarters did not provide significant differences.

Rectal temperature had the strongest influence on udder surface temperature for both sides, before as well as after infection, followed by environmental temperature.

The author pointed out the special challenges of interpreting the regions of udder-thigh-gap and intermammary cleft: highest temperatures of the image were found here regularly, and were probably not due to local inflammation but to body surfaces approaching each other in an acute angle and thus radiating heat towards each other. The author assumed a possible falsification of the results and suggested the exclusion of these regions in further studies.

HOVINEN et al. (2008) also reported warmest pixels to be mostly situated in the udder-thigh-gap in thermograms of the udder taken from the side. In this study, *E. coli* lipopolysaccharide was injected into the left forequarter. After two hours, all cows showed signs of clinical mastitis. For five days, infrared thermograms were generated every two hours, but not during nighttime. Mean temperature of all pixels in a circle above the teat base (40x40 pixels) and maximum temperature was used for interpretation. Rectal temperature was significantly elevated between 4 hours and 8 hours after infection and surface temperature showed a similar course. Actually, Rectal temperature and surface temperature had a very strong correlation ($r=0.98$ for maximum surface temperature, $r=0.92$ for mean surface temperature), underlining the results of GLAS (2008). Nevertheless, infrared thermography could not detect changes before other mastitis parameters, e.g. somatic cell count, showed significant alterations. Since infrared thermography has the advantage of being a

noninvasive technique, the authors suggest field trials in automatic udder health monitoring.

In order to compare the ability of CMT to detect mastitis with IRT, COLAK et al. (2008) screened the udder of 94 dairy cows: first CMT was done, after that thermograms were taken from each quarter. CMT results were divided into four categories and mean quarter surface temperature was generated, but the exact region that was interpreted is not described. Quarters with high CMT scores showed significantly higher temperatures: when mean CMT score of the udder rose, udder surface temperature increased linearly, whereas rectal temperature stayed at the same level. When the results were calculated and divided by quarter and not averaged for the whole udder, coefficient of correlation for CMT and quarter surface temperature was 0.92. Rectal temperature and CMT showed lower correlation ($r=0.27$). It is to assume, that most cows suffering from mastitis did not show fever in this trial. Observations about the influence of rectal temperature on udder surface temperature are quite contradictory in literature (see below).

In the study of POLAT et al. (2010), infrared thermography was also set in relation with standard approaches in detecting subclinical mastitis: they evaluated sensitivity and specificity of somatic cell count, California Mastitis Test and udder surface temperature. Infrared images were taken twice, after milking. Forequarters were recorded from lateral, hindquarters from caudal. Milk samples for SCC and CMT were taken right after. Quarters that showed a cell number above 400.000/ml were significantly warmer (2.35°C on the average) than healthy quarters, and very strong correlations for udder surface temperature and CMT ($r=0.86$), respectively SCC ($r=0.73$) have been calculated. Rectal temperature has not been taken into account in this study. For sensitivity and specificity, similarly good results were detected (95.6% and 93.6% for IRT and 88.9% and 98.9% for CMT) when cell concentration above 400.000/ml was considered as pathological. Contrary to SCHUTZ et al. (2001), the authors consider IRT as a promising detection tool, since it offers the advantage over CMT and SCC of being noninvasive (COLAK et al., 2008; POLAT et al., 2010). However, an individual temperature course of each cow and more time points of measurements would have been interesting, as well as further grading of high SCC results and the information if cows developed signs of clinical mastitis.

A wide-ranging field study on the use of an infrared camera in the milking parlor to detect mastitis was done by FRANZE et al. (2012). The camera was installed in a Saxon dairy farm and images of the udder were taken twice a day during milking for 18 days. The images were edited manually with the help of a special software: for each image, the surface of each hindquarter and a region on the caudal thigh was selected and mean and maximum temperature was calculated. In a second dairy farm, cows were observed in the same way, but evaluation was done by software for automatic image analysis, that selected the same regions. In this study, no direct comparison is done on both interpretation methods and no finding rate for the automated software is named. Low correlation of thigh surface temperature and rectal temperature ($r=0.16$) could be explained by the fact that this region was not shaved and the surface temperature of the fur was depicted, but quite contrary to SCHUTZ et al. (2001), HOVINEN et al. (2008), GLAS (2008) and METZNER et al. (2014), rectal temperature and udder surface temperature also had a low correlation (r was between 0.15 and 0.2). It was not clearly stated in the article, but illustrations indicated that the physiologically warmest regions, e.g. udder-thigh-cleft, that are probably linked to inner body temperature (GLAS, 2008), are largely left out in image editing. COLAK et al. (2008) reported a coefficient of correlation of 0.24, however, no further explanations on selected region or exclusion of “hot spots” was made. Unlike rectal temperature, FRANZE et al. (2012) found environmental temperature to have a large influence on udder surface temperature ($r=0.6$). Hindquarters suffering from mastitis, classified by SCC and bacteriological examination, showed significantly different temperatures than healthy hindquarters when the level of significance was set at 0.10. It has to be taken into account, that those quarters neighboring mastitis-quarters were probably added to the healthy group, but still were affected by the inflammation close by, as the studies of GLAS (2008) and METZNER et al. (2015) suggest. When quarters were not only divided into two groups, but when the individual temperature course of each cow was taken into account, better mastitis detection was possible: average temperature was calculated, based on earlier measurements and compared to present udder surface temperature. In doing this, sensitivity and specificity could be raised to about 30% and 70%. Still, these values are far below those considered by POLAT et al. (2010). The long time span between the measurements could be an explanation, since other studies prove the importance of short intervals (SCOTT et al., 2000; GLAS, 2008; HOVINEN et al., 2008; METZNER et al., 2015), but POLAT et al. (2010) only

took two images per cow with an interval of twelve hours. Another factor could be, that POLAT et al. (2010) considered cell count above 400.000/ml as pathological and no microbiological examination of milk samples was done. When the threshold value was set at 200.000 cells/ml, sensitivity and specificity were 38% and 100%. Categories for mastitis were a combination of SCC above 100.000/ml and bacteriological evidence in the trial of FRANZE et al. (2012).

METZNER et al. (2014) responded to the suggestion of GLAS (2008), that better results in detecting mastitis via IRT could possibly be gained by exclusion of typically warm regions in the udder area, e.g. udder-thigh-cleft. Most suitable parameters for interpretation of infrared images in mastitis detection should be identified for future studies. METZNER et al. (2014) interpreted thermal images of hindquarters of cows before and after intramammary infection with *E. coli*, when all cows showed fever, with three different regions. They were manually selected with the help of a software: lines, that run vertically on each quarter, rectangles that aligned to the intermammary sulcus and polygons, of which outer borders were defined by udder-thigh-cleft and intermammary sulcus. Pixels in these regions were used for the evaluation of surface temperature, and minimum, average and maximum temperature was calculated. In doing this, mentioned warm areas were not involved in the region of the lines, but in rectangles and, even more, in polygons. Teats were not included in evaluation. As expected by the authors of this study and GLAS (2008), surface temperature inside the polygons showed the highest coefficient of correlation with rectal temperature (0.83 and 0.80 for the right and left hindquarter, respectively). It is thus very likely that these so-called "hot spots" depict inner body temperature. Nevertheless, evaluation using polygons and maximum surface temperature gained the best results in detecting significant differences between cows with and without fever. In ROC-analysis, these parameters yielded sensitivity and specificity of 100% and 96% of detecting inflamed udders. Moreover, maximum temperature was chosen since it detected larger differences in surface temperature of healthy udders and udders suffering from mastitis than minimum and average surface temperature. Minimum temperature showed distinct variations and lowest correlation with rectal temperature and was considered as not suitable for thermograms evaluation. Standard deviation of average and maximum temperature was significantly higher when cows suffered from mastitis, implying a more heterogenic pattern of

temperature distribution in inflamed udders. METZNER et al. (2014) suggest maximum surface temperature in polygons for interpretation in future studies, combined with a software that is able to exclude up to 20% of area originating from the outer borders, in order to cut out “hot spots” and depicting local changes more distinctly. This software should be able to detect the udder automatically, making low-cost, automated health monitoring possible.

In a second study, the authors applied the parameters detected as suitable in evaluating the course of surface temperature of inflamed udders (METZNER et al., 2015). Thermograms taken every two hours throughout a trial of 48 hours were used for this purpose. As suggested, average and maximum surface temperature in polygonal regions of the hindquarters were ascertained. A period of 24 hours, starting with intramammary injection of *E. coli*, was directly compared with the period of 24 hours before infection, when udders were healthy. Udder surface temperature was elevated between 11 and 17 hours after infection, significantly highest between 13 and 15 hours. The course of rectal temperature was similar. These peaks occur later than in the study of HOVINEN et al. (2008), who, however, used *E. coli* lipopolysaccharide and not *E. coli* suspension in the trial. Comparison of surface temperature of infected and uninfected quarter showed significant differences 13 and 17 hours after infection, but could not be approved in a mixed-model analysis. These results emphasize the similarities in USST of the infected quarter and the neighboring one. Differences of average surface temperature in right-left comparison showed a lower amplitude than those of maximum temperature. The authors assumed that the right, infected quarter was probably slightly cooler due to edematous swelling. BARTH (2000) reported neighboring quarters to be 0.3K and 1.7K cooler in one, spontaneous occurring mastitis, whereas SCOTT et al. (2000) found significant elevation of surface temperature in both hindquarters, when only the left one was challenged with *E. coli* endotoxin. When SCHUTZ et al. (2001) infected cows intracisternally with *S. aureus*, challenged quarters were significantly warmer than the others.

METZNER et al. (2015) concluded that clinical mastitis could be reliably detected with the method established earlier (METZNER et al., 2014), when intervals between measurements were kept short.

6. Automatic image recognition in mastitis detection

As mentioned above, the advantages of IRT as an early diagnosis tool cannot be used in surveillance of livestock herd health, since manual evaluation of thermograms requires a high amount of time. In order for IRT to become practically relevant, systems of automated and computer-assisted interpretation have to be developed and trained.

For this purpose, the German Federal Ministry of Education and Research founded the project VIONA (WIRTHGEN et al., 2011b; WIRTHGEN et al., 2011a; WIRTHGEN et al., 2012). Monitoring of dairy cattle by IRT and automatic evaluation of the images was tested in stable, accompanied with the challenges for the measuring occurring in this environment, such as moving animals, inconstant climate conditions and dirt in regions of interest. Two infrared cameras have been installed in milking carousel, one in the anterior region of the animals for individual identification, and the other behind the cows. The use of a reference body with known emissivity reduced uncertainty of measuring to $\pm 0.47\text{K}$ (WIRTHGEN et al., 2011b; WIRTHGEN et al., 2011a). More than two million thermograms were generated. The exact number of observed animals or the period of time is not described. It is to assume that measuring took place during milking. Regions of interest were the complete rear part of the cow, the hind udder quarters and the hind claws.

In this trial, Active Shape Approach was applied in veterinary imaging for the first time. This technique of automated recognition of silhouettes was developed for segmentation of images in human medicine (COOTES et al., 1994; COOTES et al., 1995).

The purpose of the VIONA project was to find out whether the active shape approach offers the same reliable segmentation and evaluation of veterinary thermograms as manual interpretation. Maximum and average temperature for each region of interest were calculated by manual and automated segmentation and compared by coefficient of correlation: Whereas coefficient of correlation for the rear part of the animal was 0.93 for average surface temperature, respectively 0.85 for maximum temperature, it was 0.66 (average temperature) and 0.76 (maximum temperature) for the region of the udder (WIRTHGEN et al., 2011b). However, it is not said how the manual interpretation took place and by whom.

In the studies of WIRTHGEN et al. (2011b); WIRTHGEN et al. (2011a) the technique of automated segmentation by active shape modeling under practical conditions was tested, achieving good results. Further research has to be done on the abilities of automated image processing in detecting pathological conditions in dairy cattle. Defined cut-off values for temperatures ascertained by automated infrared thermography do not exist yet.

IV. MATERIAL AND METHODS

1. Experimental animals and thermographic material

Thermographic material is used that emerged from infection studies implemented at the Clinic for Ruminants, Ludwig-Maximilians-Universität in Munich (GLAS, 2008; MITTERHUEMER, 2009). The ethics committee of the government of Upper Bavaria authorized the trial (reference: 55.2-1-54-2531-108-05).

1.1. Conditions for experimental animals

Five Friesian-Holstein cows from Bavarian dairy farms were used. Their age ranged from 25 to 30 months, daily milk yield was between 15 and 25 liters. All animals had to fulfill following conditions (MITTERHUEMER, 2009):

- primiparity
- no history of mastitis
- no pathological findings in repeated general examinations in their initial dairy farms
- no indication of bacterial infection in microbiological examination of milk samples
- Somatic cell count (SCC) of milk samples was not to exceed 50.000 cells/ml in three quarters, values up to 150.000 cells/ml were acceptable if only shown in one of four quarters and if the particular quarter was not used for intramammary challenge

Sterile milk samples were taken at least twice when animals were kept in their initial dairy farm, and three more times in a weekly interval when cows were housed at the Clinic for Ruminants. Before sample collection, 20 IU oxytocin were injected into the tail vein. Teats were cleaned using cellulose wipes. Foremilk was gained and California Mastitis Test (CMT) was implemented. Afterwards, teats were disinfected using 70% Ethanol and milk samples of each quarter were milked into sterile sample containers.

SCC and total bacteria count was done in the laboratories of the Bavarian Association for raw milk testing (Milchprüfring Bayern e.V.) via optical fluorescent flow cytometry (Fossomatic-FC and BactoScan-FC). In addition, milk samples were streaked out and incubated on Columbia sheep blood agar, Edwards agar and Gassner agar. CMT was repeated 12 and 24 hours after intramammary challenge.

1.2. Housing and preparation of the animals

Three weeks prior to the trial, cows were housed in the stable of the Clinic for Ruminants. Flooring consisted of rubber mats and straw bedding, which was changed regularly. Walls were covered with white tiles. Windows ensured natural lighting, artificial illumination was done by fluorescent tubes. Cows were tethered with a collar. Water was supplied ad libitum by automatic drinking troughs.

Milking was done mechanically, using a portable bucket milking unit, and took place twice per day, at 6:00 a.m. and 6:00 p.m. Pre-milking was done manually. After milking, teats were disinfected by dipping (Ecolab® P3-cide special).

Cows were fed a total mixed ration and shredded grain three times per day, hay was provided ad libitum.

Estrus cycle was synchronized: a synthetic analogue of prostaglandin F2 α (Dalmazin™) was injected twice with an interval of ten days. Second injection was three days prior to the trial. All cows were in estrus during the trial (MITTERHUEMER, 2009).

1.3. Intramammary challenge

At 6:00 a.m., after injection of 20 IU of oxytocin into the tail vein and examination of pre-milk with CMT and milk sample collection as described above, cows were fully milked. 500 colony forming units (CFU) of *E. coli* strain 1303 (LEIMBACH et al., 2016), diluted in 2 milliliters (mL) of isotonic sterile saline solution, were injected into the right hindquarter via the teat canal at 6:30 a.m. 2 mL of isotonic sterile saline solution, as placebo control, were injected into the left hindquarter in the same way. Both fluids were manually massaged into the proximal udder tissue for 30 seconds (MITTERHUEMER, 2009).

1.4. Schedule of measurements

Thermographic measurements were done at defined time points (see Table 1). At each time point, three thermograms were taken in a row. Total duration of the trial was 49 hours. The first 24 hours of the trial served as reference: Between 5:30 a.m. and 3:30 a.m., thermograms were taken at 15 time points. Intervals varied between one and two hours, since additional measurements were implemented after milking (6:30 a.m., 8:30 a.m. and 6:00 p.m.) (GLAS, 2008).

Intramammary challenge took place at 6:30 a.m. of the next day. The same schedule of measurements was used for the period after challenge. Thus, each measurement at the period after challenge corresponds to a measurement of reference period. Two additional measurements were implemented at 5:30 a.m. and 6:30 a.m., respectively 23 and 24 hours after challenge.

Table 1: Schedule of measurements

	time	hours after challenge		time	hours after challenge
reference period (before challenge)	5:30 a.m.	-25	period after challenge	5:30 a.m.	-1
	6:30 a.m.	-24		6:30 a.m.	0
	7:30 a.m.	-23		7:30 a.m.	1
	8:30 a.m.	-22		8:30 a.m.	2
	9:30 a.m.	-21		9:30 a.m.	3
	11:30 a.m.	-19		11:30 a.m.	5
	1:30 p.m.	-17		1:30 p.m.	7
	3:30 p.m.	-15		3:30 p.m.	9
	5:30 p.m.	-13		5:30 p.m.	11
	6:00 p.m.	-12.5		6:00 p.m.	11.5
	7:30 p.m.	-11		7:30 p.m.	13
	9:30 p.m.	-9		9:30 p.m.	15
	11:30 p.m.	-7		11:30 p.m.	17
	1:30 a.m.	-5		1:30 a.m.	19
	3:30 a.m.	-3		3:30 a.m.	21
				5:30 a.m.	23
		6:30 a.m.	24		

1.5. Procedure of measuring

1.5.1. Hardware equipment

A thermal imaging camera of the series ThermoCAM™ B20HSV (FLIR® Systems, Wilsonville, Oregon, United States) was used for generating thermograms:

- accuracy of the measured values $\pm 2\%$
- spectral range 7.5-13 μm
- image resolution 640 x 480 pixels, full color

- image frequency 50/60 Hz, non-interlaced
- environmental temperature range for operating -15 to + 50 °C
- environmental humidity range for operating 10 to 95%, non-condensing

Focus of the camera was adjusted automatically. Emissivity correction was implemented, setting the number of emissivity manually before measurements. For udder surface, emissivity of 0.96 was selected (GLAS, 2008).

In order to detect environmental temperature and humidity at each point of measurement, a data logger was used (PCE Instruments, Meschede, Germany). Rectal temperature of the animals was also recorded throughout the trial, using a digital thermometer (Microlife AG, Widnau, Switzerland).

1.5.2. Recording of thermograms

Recording of thermograms took place in the stable, where the animals were housed. Artificial ventilation was turned off shortly before recording, to avoid falsification by airflow and thus evaporation cooling. If the udder was soiled, it was dryly cleaned with cellulose tissue. Wet cleaning was only done if it has been inevitable. In that case, cleaning was done 10 to 15 minutes before the measurements took place, to ensure that skin was dry and increased microcirculation by manipulation was normalized. If it was necessary, udder hair was clipped with the help of a trimming machine before the beginning of the trial (GLAS, 2008).



Figure 1: Photograph of the rear part of the udder, cow is fixed for recording (GLAS, 2008)

Cows were fixed in the feeding fence and their tails were tied aside. At the rear side of the udder, a patch of adhesive tape was placed at the height of the knee joints, to serve as marking (see Figure 1).

Thermograms were taken of both hindquarters. The examiner was placed with the thermal camera behind the cow. An ultrasonic distance sensor (Stellar Products, Hilden, Germany) was used to ensure the same constant distance of 1.8 m between camera and udder surface.

At each time point of measurement (see Table 1), three thermograms were taken in a row. The quality of the recorded thermograms was briefly assessed on the thermal camera's screen. If images were blurred, recording was repeated until three convenient consecutive images were taken. By triplicate recording, mean values can be generated in following evaluation.

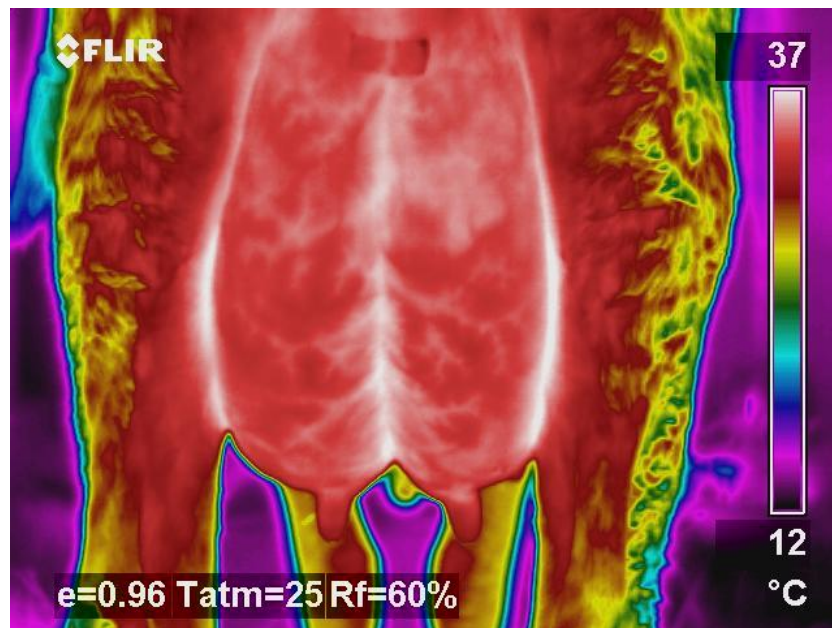


Figure 2: Thermogram of the hindquarters (GLAS, 2008)

Thermograms are displayed as images with a specified color scale. Each pixel, representing a recorded temperature, is depicted in the corresponding temperature (see Figure 2).

Immediately after each time of thermal recording, rectal temperature of the animals, environmental temperature and environmental humidity were recorded.

2. Evaluation of thermographic material

Two comparisons are done in this study:

- **Comparing automatic and manual evaluation of thermograms**
- **Comparing thermogram evaluation that uses the region of the teats with evaluation that uses the region of the udder and excludes the region of the teats**

2.1. Evaluation by an automated image recognition software

Automatic evaluation (Aut) of thermograms of the udder has to give following performances:

- Automatic recognition of the udder region
- Automatic segmentation of the udder into defined regions of interest (ROIs)
- Ascertaining the temperature values inside the ROIs

A software for automatic evaluation of thermograms based on Active Shape Model approach is developed by Fraunhofer Institute for Transportation and Infrastructure Systems IVI, Dresden, Germany (SCHRÖTER, 2015). A total of 4143 thermographic images that emerged from the trial at the Clinic for Ruminants (see ‘Material and methods; Recording of thermograms’) is provided for this purpose.

Automatic segmentation must cope with the challenge of detecting two-dimensional ROIs in images of three-dimensional objects. It is done by recognizing the contrast of the external borders of an object as lines. In the case of thermograms of the udder, the most prominent contrasts for defining external borders are found in the area of udder-thigh-clefts and intermammary sulcus. In these regions, body surfaces approach each other in a sharp angle, emitting body warmth and radiating it towards each other. Thus, these regions appear as areas with typically high temperatures (‘hot spots’) in a thermogram (see Figure 2). The contrast of relatively warm body parts towards the cooler environment is also used, as in the area of the lower udder and the teats (see Figure 2). The adhesive tape patch served as marking of the upper border of the udder. If necessary for recognition of the outer borders, adjustment of the contrast is done prior to evaluation.

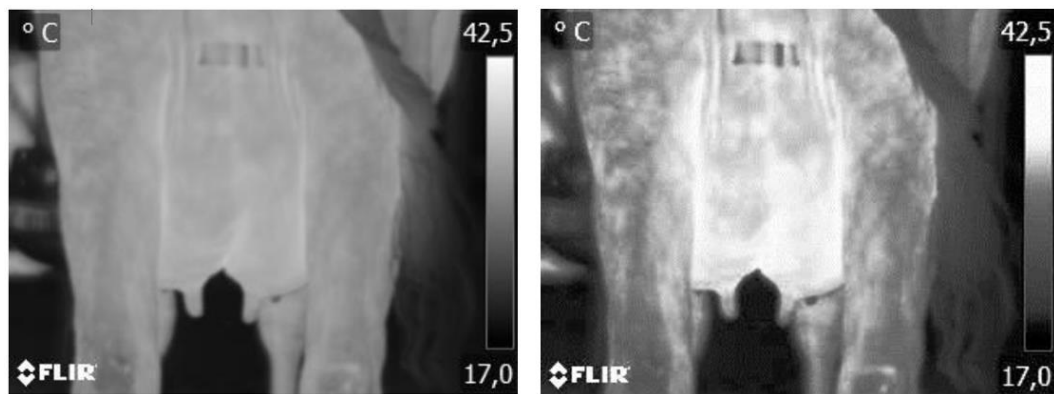


Figure 3: The same thermogram before (left) and after (right) adjustment of contrast (SCHRÖTER, 2015)

For the development of the Active-Shape-Model, a training shape, or point-distribution-model, has to be generated: the silhouette of the region of interest has to be outlined manually, each point is called a landmark. The more training shapes of the object are generated in different images, the more reliable is the detection of the algorithm. For generating the point-distribution-model of the software used in this study, a total of 70 thermograms of udders are processed manually: 35 points on the outer borders of the udder are marked clockwise in each image. These thermograms also derived from the stock of images provided from the Clinic for Ruminants. The 35 landmarks are later connected to lines that define the outer borders (see Figure 4).

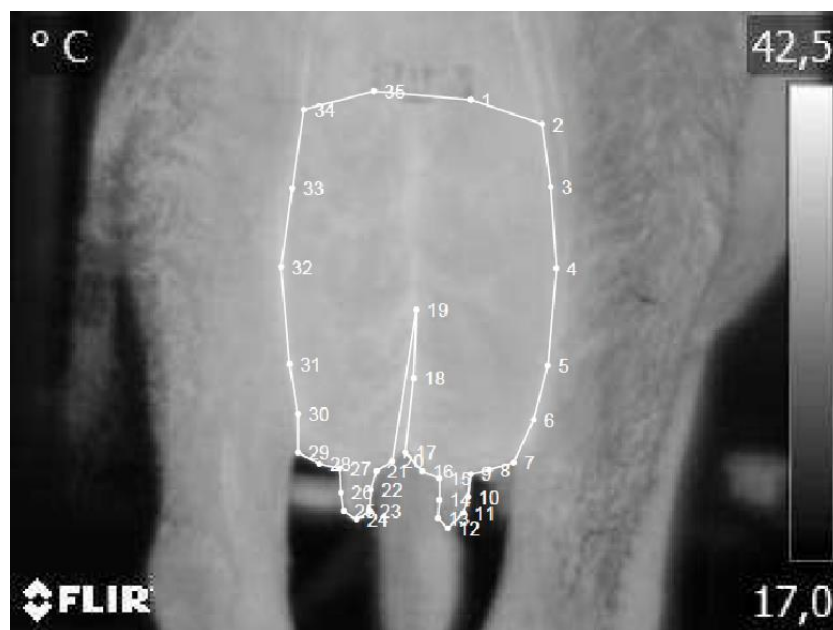


Figure 4: Landmarks manually set into the thermogram (SCHRÖTER, 2015)

Based on the point distribution model, the Active Shape model is calculated statistically, considering the average position of each landmark and its variability.

Thus, the shape of a recognized object can only differ within a limited scope, unlike in the Active Contour approach.

When the software recognizes the shape of an udder in a thermogram, it outlines the outer borders with the help of the Active Shape model algorithm. In a second step, the region of the hindquarters is separated in the line of the intermammary sulcus. Thus, two ROIs are generated: the hind surface of the left and right hindquarter, including the teats. Lastly, 5% of the pixels in the ROI, originating from the outer borders, are eliminated automatically to exclude ‘hot spots’.

For every image, minimum (‘Min’), maximum (‘Max’) and average (‘Avg’) temperature value inside each ROI, the left and right hindquarter, are generated. Automatic evaluation is performed in the context of a master thesis at the Fraunhofer Institute for Transportation and Infrastructure Systems IVI, Dresden, Germany (SCHRÖTER, 2015). The raised raw data is provided for the author in exchange of the raw data of manual evaluation.

2.2. Manual evaluation of thermographic material

Since automatic evaluation method is to compare with the gold standard of manual evaluation, the same thermograms that are interpreted automatically are assessed manually in this study.

As software for manual evaluation, ThermaCAM Researcher Pro 2.8 (FLIR® Systems, Wilsonville, Oregon, United States) is used. Values for environmental humidity, environmental temperature and emissivity, adjusted in the camera’s settings during recording, are taken into account by the software. The software offers different tools for selecting ROIs inside the thermograms. Like in automatic evaluation, rear surface of each hindquarter, including the teats, is chosen as ROI. For this purpose, the polygon-tool is used: Manually, the outer borders for the ROIs are set. These are aligned with the adhesive tape marking in the proximal udder region, the intermammary sulcus in the mediane, the udder-thigh clefts as lateral boundaries and the contrast between environment and distal udder, respectively teats. Manual selecting of ROIs is repeated for every image.

The software allows to manually change the limits of the color scale and thus adjusting the contrast, if structures cannot be recognized in the first instance due to poor contrast with the structures nearby.

For both ROIs, the surface of the left (AR01) and right hindquarter (AR02), the software calculates values for ‘Min’, ‘Max’ and ‘Avg’. Additionally, values for the temperature range (‘Max-Min’) and the standard deviation (‘Stdev’) of temperature values inside the ROI are given (see Figure 5).

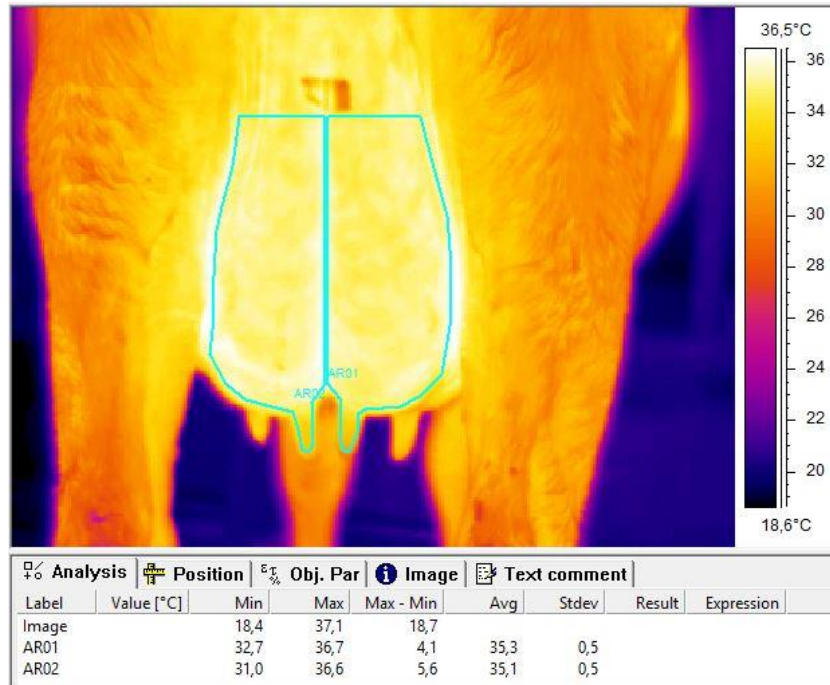


Figure 5: Screenshot of manual evaluation: left (AR01) and right hindquarter (AR02) are selected, using the polygon tool

Standardized exclusion of hot spots is not done in manual evaluation for two reasons:

- In this comparison, the influence of hot spots on the calculated course of temperature is supposed to be evaluated, so one method excludes the hot spots while the other one includes them.
- In automatic segmentation, standardized exclusion of 5% of the pixels is possible, whereas in manual evaluation exclusion would not be reliably reproducible.

2.3. Evaluation of the region of the teats

The thermograms that emerged from the study of GLAS (2008) are evaluated manually again, considering exclusively the hind teats. For segmentation of the region of the teats, ThermaCAM Researcher Pro 2.8 (FLIR® Systems, Wilsonville, Oregon, United States) is used. With the polygon tool, the outer borders of the left (AR01) and right hind teat (AR02) are circumscribed in each thermogram: The teat base is supposed to be the proximal limitation, so a line is drawn between udder and

teat surface. The other outer borders can be identified by the contrast between teat surface and environment. If necessary for recognition, the contrast is adjusted. For each ROI, values for ‘Min’, ‘Max’, ‘Avg’, ‘Max-Min’ and ‘Stdev’ are obtained.

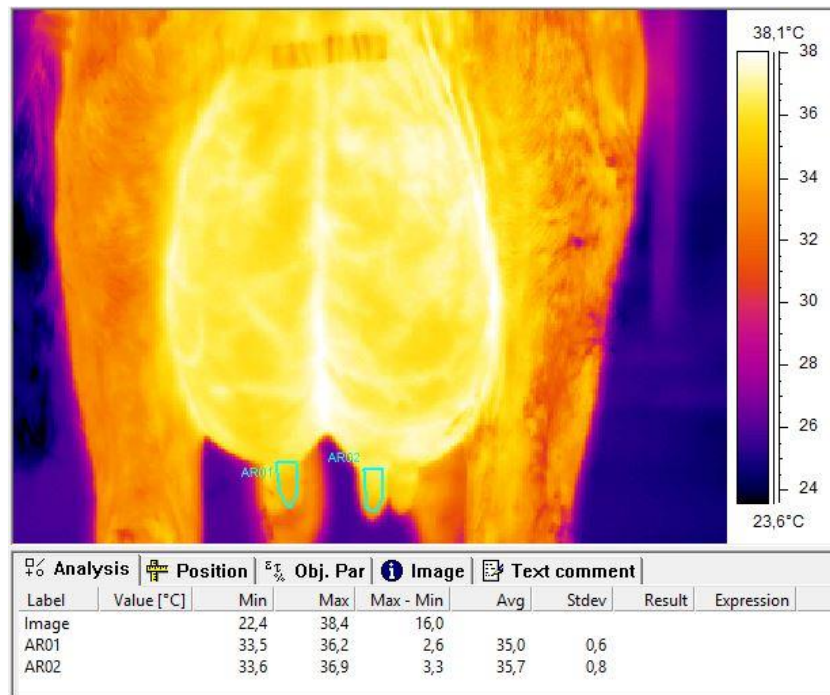


Figure 6: Screenshot of evaluation: left (AR01) and right hind teat (AR02) are selected, using the polygon tool

2.4. Evaluation of thermographic material without the region of the teats

For comparison with evaluation using the region of the teats, data are used that emerged from an earlier study, when the same thermograms were evaluated, using the region of the hind udder surface without the region of the teats (GLAS, 2008). The author of the named study gave the permit to use the data.

With the polygon-tool of ThermaCAM Researcher Pro 2.8 (FLIR® Systems, Wilsonville, Oregon, United States), the surface of the left and right hindquarter is circumscribed. The outer borders are defined by the adhesive tape patch, intermammary sulcus and udder-thigh-clefts.

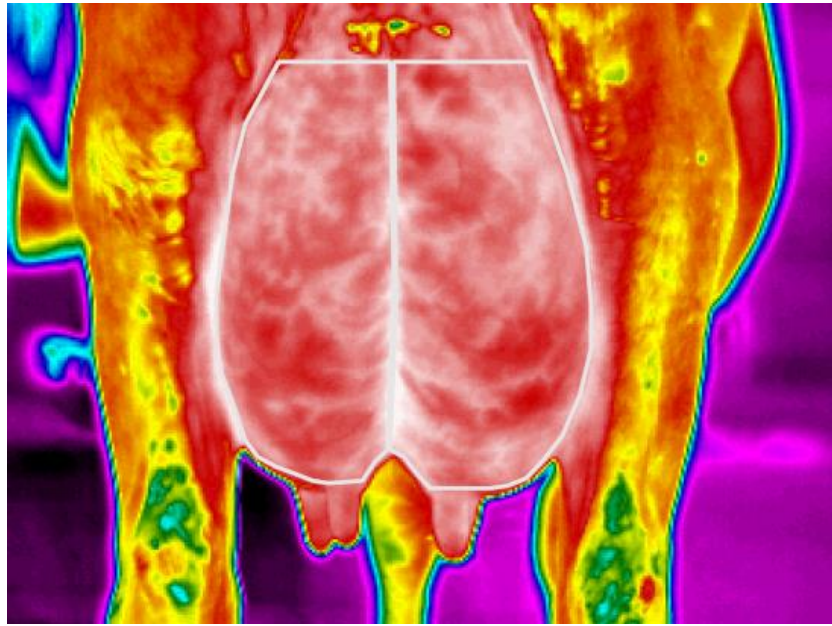


Figure 7: Left and right hindquarter circumscribed with the help of the polygon-tool, teats excluded (GLAS, 2008)

Values for 'Min', 'Max', 'Avg', 'Max-Min' and 'Stdev' are calculated for both ROIs in each thermogram.

2.5. Documentation of obtained data

Four methods are applied on the recorded thermograms:

1. automatic evaluation using the hind surface of the hind udder quarters, including the teats ('Aut')
2. manual evaluation using the hind surface of the hind udder quarters, including the teats ('Man')
3. manual evaluation using exclusively the hind surface of the hind teats ('Teats')
4. manual evaluation using the hind surface of the hind udder quarters, excluding the teats ('Udder')

For every method, obtained data from each ROI and thermogram are documented in four separate tables, using Microsoft Excel 2015. At every time-point of measurement, three thermograms have been recorded and evaluated with the four methods. For these triplets, means for 'Min', 'Max', 'Max' – min and 'Avg' are calculated. In the following calculations, the mean values are used (see 'Appendix 1: Tables of data obtained by automatic and manual evaluation' and 'Appendix 2: Tables of data obtained by evaluation using the teats and evaluation using the udder').

3. Statistical methods

GraphPad Prism[®], version 5.04 for Windows, GraphPad Software, San Diego, California, USA is used for statistical analysis.

3.1. Testing Gaussian distribution of the data

To choose the following statistical methods, it is of importance to decide whether the data emerge from a Gaussian distribution or not. For this purpose, the D'Agostino-Pearson normality test (omnibus K2) is used. This test calculates a P value that reports the probability of obtaining this distribution of values when data were randomly collected from a Gaussian distribution. Thus, a high P value indicates Gaussian distribution. This calculation is based on the differences between the observed values and the values that are expected in a perfect Gaussian distribution. Furthermore, this test provides values for skewness and kurtosis. In a Gaussian distribution, both values equal 0. Positive skewness indicates a shift of the distribution curve towards the right, negative skewness indicates a shift towards the left. A distribution curve that appears flatter than a Gaussian distribution curve has a positive kurtosis, whereas negative kurtosis indicates a curve with a pointed peak. Skewness and kurtosis provide information about the shape of the distribution curve. However, defined threshold values for assumption of Gaussian distribution do not exist.

Normality test are said to be more powerful, the larger the sample size is. A sample size of $n > 100$ is recommended.

For all four methods, the D'Agostino-Pearson test is done separately. Values of 'Min', 'Max', and 'Avg' of both hindquarters before challenge are used.

In addition, histograms of the named data are made.

3.2. Precision of the different methods

In this study, different methods and parameters are to be evaluated. The calculation of precision is done to compare the accuracy of manual and automatic evaluation, respectively evaluation using the teats and evaluation using the udder. In addition, accuracy of the different parameter 'Min', 'Max', 'Avg' and 'Max-Min' is ascertained.

For this calculation, values from the measurements before the challenge is used.

Reminder: For both hindquarters of five animals, there are triplets of thermograms for ‘Min’, ‘Max’, ‘Max-Min’, and ‘Avg’ at every time point of measurement. Data from the evaluation of four methods are available: automatic evaluation (‘Aut’), manual evaluation using udder and teats (‘Man’), manual evaluation using only the teats (‘Teats’), manual evaluation using only the udder (‘Udder’).

In this calculation, parameters with a smaller arithmetic mean (e.g.: ‘Max-Min’) are compared with parameters that have larger arithmetic means (e.g.: ‘Max’). ‘Max’ is most likely to have a larger standard deviation than ‘Max-Min’. Since the coefficient of variation (VC) is a ratio, it is independent from differences in arithmetic means. The VC calculates the relation of the standard deviation towards the arithmetic mean and thus indicates the distribution of values around the arithmetic mean. It is thus suitable to compare the precision of different measurement methods. Low values refer to a good precision.

The VC is calculated for every animal at all time points before challenge for the four different methods. The results are displayed in box-plot diagrams.

With one-way ANOVA, the VCs of ‘Min’, ‘Max’, ‘Max-Min’ and ‘Avg’ within the methods are tested on significant difference. The one-way ANOVA indicates if significant differences between the coefficients of variation of the different parameters exist, but does not calculate significance of differences among the single parameters. Thus, Tukey's multiple comparison test is performed as post-test. After deciding which of the parameters offer significantly lowest coefficients of variation and thus best precision, these parameters are separately tested on significant differences between the methods ‘Man’ and ‘Aut’, respectively ‘Teats’ and ‘Udder’. For this purpose, unmatched t-test is used.

Unmatched t-test is a nonparametric test and designed for comparing two groups in which the values variances are similar. Along with the unmatched t-test, the F-test is performed in order to identify unequal variances. An F-value exceeding 1 by far indicates different variances. The associated P-value reports the probability of finding an F-value this large in two groups with actually equal variances. Since coefficients of variation and thus ratio values are compared, similar variances are assumed. However, unmatched t-test is resistant towards unequal variances when the sample size is large and does not differ from group to group. In this analysis, n=65 in all groups.

3.3. Correlation of evaluation method and rectal temperature

The aim is to test, which of the compared evaluation methods offers the best correlation with rectal temperature throughout the trial. Data from before and after challenge are used.

Since 'Max', 'Min' and 'Avg' offer the best precision in evaluation, the course of these two parameters is set in relation to the course of rectal temperature for all evaluation methods. The calculations are done separately for left and right hindquarter.

To find out to what extent changes in one course depends on changes in the other course, Pearson correlation analysis is performed.

Pearson correlation analysis requires data that follow Gaussian distribution if small sample sizes are used for calculation. The greater the sample size (e.g.: $n \geq 100$), the less dependent this analysis is from Gaussian distribution of the data. Spearman correlation analysis is a nonparametric method, since its calculations are based on ranks. Nevertheless, it has less power, than Pearson analysis, especially when sample sizes are small.

Results of normality tests suggest that data collected before challenge follows Gaussian distribution. However, when data from before and after challenge are considered, the histogram analysis becomes asymmetrical (histograms not displayed).

In this calculation, 155 pairs ($n=155$) of evaluation parameter ('Min', 'Max' or 'Avg') and rectal temperature were available for every single analysis. Metrically scaled values are used in this correlation analysis and linear correlation is assumed. Consequently, Pearson correlation analysis is here considered as more suitable than Spearman correlation.

In addition, mean differences of rectal temperature and average surface temperature were calculated for every method. Since not the amount of the difference, but the consistence of this difference is evident for the relation between rectal temperature and measured surface temperature, standard deviations of the differences are calculated.

The standard deviations of the differences are considered before challenge, after challenge and in total. Using t-test, it is analyzed if the results of ‘Man’ and ‘Aut’, respectively ‘teats’ and ‘udder’, are significantly different.

3.4. Comparison of automatic and manual evaluation

3.4.1. Correlation analysis of automatic and manual evaluation

The temperature course of automatic evaluation is set in relation to the temperature course of manual evaluation. For this purpose, mean values of ‘Max’ and ‘Avg’ of each method are analyzed, separated by quarter. Data from all five cows before and after challenge are used, resulting in a total of n=155 pairs for each calculation. For each hindquarter and each evaluation parameter (‘Max’ and ‘Avg’), correlation analysis is done separately. Distribution of the data in scatter plots indicates a linear correlation between the parameter’s values in automatic and manual evaluation. Pearson correlation analysis is used.

For visualization, all temperature courses are displayed in line graphs.

3.4.2. Comparison of period after challenge and reference period in automatic and manual method

Due to the schedule of measurements (see Table 1), each time point of measurement at the day after challenge refers to a time point at the reference period before challenge. Thus, differences of measured USST at period after challenge and reference period can be calculated: 13 time points after challenge are compared with 13 time points before challenge:

Table 2: Time-points before and after challenge to be compared in statistical analysis. Differences between time-points are constantly 24 hours.

time points in period after challenge (hours)	compared to time points before challenge (hours)
-1*	25
0	24
1	23
2	22
3	21
5	19
7	17
9	15
11	13
11.5	12.5
13	11
15	9
17	7
19	5
21	3

* negative value indicates time before challenge

Positive values of the difference (period after challenge-reference period) indicate an increase in USST at the period after challenge.

The differences of 'Avg' and 'Max' are calculated for both, automatic and manual evaluation method. Since results of five animals are considered, n=5 for every method and parameter at each time point. The results are displayed in box-plot graphs.

After calculating the differences for both methods, it is of interest at which time point the differences become significantly higher. It is also possible that one method is able to detect significant changes earlier than the other method.

For this purpose, a multiple comparison test is performed. Since the observations are repeatedly done on the same five cows, a test for repeated measurements is chosen. D'Agostino & Pearson normality test (omnibus K2) is done for each animal, each method and each parameter ('Max' and 'Avg') separately. The results indicate Gaussian distribution of the data (results not displayed). Thus, one-way ANOVA as parametric test is chosen (n=13 groups). This test is designed to analyze the probability whether the differences of the means of three or more groups are

due to coincidence. A small P-value indicates that this probability is low. The F-ratio indicates the scattering of the analyzed data. If the means of the groups differ significantly, an F-ratio exceeding 1 is expected.

The repeated measurement one-way ANOVA is followed by a multiple comparison test (post-hoc test). In this case, Dunett's test is performed: This test allows to compare all groups with a control group. As control group, the differences of the values measured 25h before challenge and 1h before challenge are used. Differences in this period cannot be due to challenge. The Dunett's test calculates the probability (P-value) that differences between the groups are observed in randomly sampled data, although these differences do not exist. Thus, it allows to estimate which differences are significant and which are not. The mean difference between the groups and the associated P-values are calculated.

3.4.3. Determining threshold values for automatic and manual evaluation

An additional way of analyzing the accuracy of both methods is comparing sensitivity and specificity.

ROC-curves (Receiver-Operating-Characteristics-curves) are designed to test diagnostic methods by setting values observed in pathological conditions (in this case: clinical mastitis) in relation to values observed in a control group. Fever (rectal temperature exceeding 39.5°C) was observed in all cows at several time-points:

- 13h and 15h after challenge in cow 1;
- between 11.5h and 15h in cow 2;
- between 11.5h and 17h in cow 3;
- between 11.5h and 17h in cow 4;
- and between 11.5h and 15h in cow 5.

All cows were found to be healthy in clinical examination prior to the trial (see: Material and methods; 1.1 Conditions for experimental animals). Moreover, SCC implemented with milk samples from the challenged quarter 11.5h after challenge detected cell numbers exceeding 400.000 cells/ml by far in all cows. It can reasonably be assumed that the fever is due to clinical mastitis after intramammary *E. coli*-challenge. Thus, values measured in these intervals are opposed with the values measured at the remaining time-points in ROC-analysis.

Sensitivity and specificity for various threshold values are calculated. The maximum sum of sensitivity and specificity is considered as optimum and the associated threshold value is chosen.

ROC curves are displayed in Figure 30. The area underneath the curve (AUC) is equivalent with the method's ability to detect the pathological condition. Whereas a surface area of 0.5 is the poorest possible outcome of ROC-analysis, 1.0 would be an ideal surface area, representing 100% of correct differentiation between patients and control group.

For the analysis, 'Max' and 'Avg' of HL and HR of automatic and manual evaluation method are used.

3.5. Comparison of evaluation using the teats and evaluation using the udder

Statistical methods for the comparison of evaluation using the teats and evaluation using the udder follow the same principles as described in the comparison of manual and automatic method (see above).

3.5.1. Correlation analysis of evaluation using the teats and evaluation using the udder

The temperature course of evaluation using the teats is set in relation with the temperature course of evaluation using the udder. For this purpose, values of 'Max' and 'Avg' of each method are analyzed, separated by quarter. Data from all cows before and after challenge are used. As described in the comparison of automatic and manual evaluation method, values of 'Max' and 'Avg' in both methods are presented in scatter plot diagrams, indicating a linear correlation (in each case, n=155 pairs). Pearson correlation analysis is performed.

All temperature courses are displayed in line graphs.

3.5.2. Comparison of period after challenge and reference period in evaluation using the teats and evaluation using the udder

Differences of measured USST in period after challenge and reference period are calculated for 13 time points after challenge and 13 time points before challenge, see Table 2.

The differences of 'Avg' and 'Max' are calculated for evaluation using the teats and evaluation using the udder, n=5 for every method and parameter at each time point. The results are displayed in box-plot graphs.

It is determined at which time points the differences become significant: D'Agostino & Pearson normality test (omnibus K2) is done for each animal, each method and each parameter ('Max' and 'Avg') separately. The results again indicate Gaussian distribution of the data (results not displayed). and one-way ANOVA as parametric test is performed., calculating P-value and F-ratio. As post-hoc test, Dunett's test is performed: all groups are compared with a control group. As control group, the differences of the values measured 25h before challenge and 1h before challenge (-1) are used.

3.5.3. Determining threshold values for evaluation using the teats and evaluation using the udder

Rectal temperature exceeding 39.5°C is observed:

- 13h and 15h after challenge in cow 1;
- between 11.5h and 15h in cow 2;
- between 11.5h and 17h in cow 3;
- between 11.5h and 17h in cow 4;
- and between 11.5h and 15h in cow 5.

Again, values measured in these intervals are opposed with the values measured at the remaining time-points in ROC-analysis. For this purpose, 'Max' and 'Avg' of HL and HR of evaluation using the teats and evaluation using the udder are used.

The maximum sum of sensitivity and specificity is considered as optimum and the associated threshold value is chosen. ROC curves are displayed in Figure 32. Figure 32: ROC (Receiver-Operating-Characteristics) curves of evaluation parameters in evaluation using the teats ('Teats') and evaluation

V. RESULTS

1. Testing Gaussian distribution of the data

The results of D'Agostino-Pearson normality test (omnibus K2) indicate Gaussian distribution of 'Min', 'Max' and 'Avg' of automatic evaluation and 'Max' and 'Avg' of manual evaluation. They are displayed in Table 39 in Appendix 3. However, 'Min' of manual evaluation has a small P-value and high values for skewness and kurtosis. Mean and median are equal or show very low deviation

Regarding the results of D'Agostino-Pearson normality test of the data obtained by evaluation using only the region of the teats, distinct deviations from Gaussian distribution can be observed for 'Min' and 'Avg'. Negative values for skewness and positive values for kurtosis plead for an asymmetrical shape and a pointed peak of the distribution curve. However, the results of evaluation using the udder without teats indicate Gaussian distribution of 'Min', 'Max' and 'Avg'. Only slight differences of skewness and kurtosis are observed for these parameters. The results are displayed in Table 40 in Appendix 3.

Histogram analysis (see Figure 23, Figure 24, Figure 25 and Figure 26 in Appendix 3) depicts the skewness and kurtosis described in the results of D'Agostino-Pearson normality tests: Histograms appear more flattened or pointed than expected from Gaussian distribution. Some show more than one peak or an asymmetrical shape. However, D'Agostino-Pearson normality test indicates normal distribution for most parameters. Moreover, the larger the sample size, parametric tests become more robust towards discrepancies from normal distribution. Thus, parametric tests are used for statistical analysis in this study when sample sizes are large. Whenever measurements of one time-point are compared, medians are used since $n=5$. For the comparison of period after challenge and reference period, differences of the temperature values are calculated. Their distribution is analyzed separately (see '2.3.2 Comparison of period after challenge and reference period in automatic and manual method' and '3.3.2 Comparison of period after challenge and reference period in evaluation using the teats and evaluation using the udder')

2. Results of automatic and manual evaluation methods

With the manual evaluation method, it was possible to outline the ROI's in all available thermograms. The segmentation of the lower region of udder and teats was in approximately 2% of the thermograms inadequate: hind teats could not be

differentiated from fore teats or forelegs. In approximately 3% of the thermograms, the proximal border of the udder (adhesive patch) was not adequately detected. Data deriving from inadequately segmented thermograms were **not** excluded.

2.1. Precision of automatic and manual evaluation

The results of precision calculation are displayed in Figure 8. The median as well as first quartile (q1) and third quartile (q3) of coefficients of variation of the different measurement parameters in automatic and manual evaluation are displayed in Table 41 in Appendix 4. For every parameter in each evaluation method, n=65. Small coefficients of variation indicate good precision.

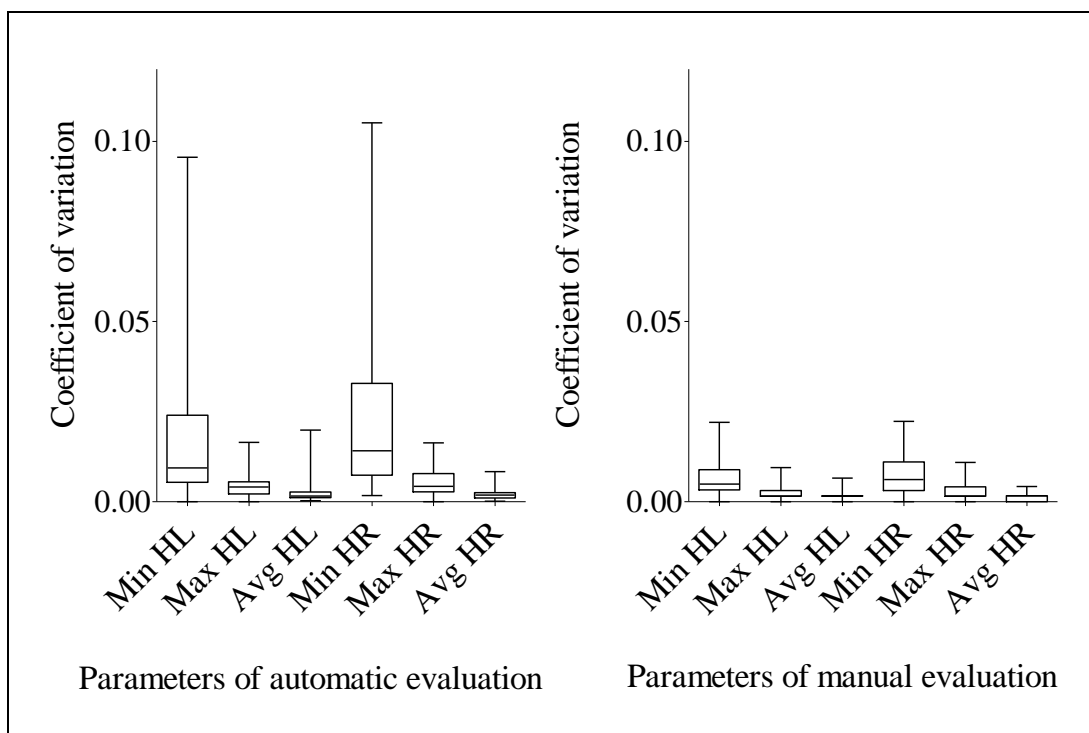


Figure 8: Coefficients of variations displayed in Box-plot diagrams, comparison of automatic and manual evaluation (coefficients of variation of 'Max-Min' not displayed). The box represents values from first to third quartile, the line marks the median. Bars depict the range of the values.

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Due to the values of coefficients of variation of 'Max-Min' exceeding those of the other parameters by far (see Table 41), they are not displayed in the box-plot-charts of Figure 8. The box-plots depicting the parameters 'Min HL' and 'Min HR' of automatic evaluation appear noticeably larger than those of the other parameters. In the chart depicting manual evaluation, the proportions differ less distinctly,

although the median coefficients of variation of ‘Min HL’ and ‘Min HR’ are also higher than those of the other parameters.

Comparing the coefficients of variation in **automatic evaluation** with One-way ANOVA, the result distinctly indicates significant differences between the eight groups of parameters (see Table 3):

Table 3: Results of One-way ANOVA and Tukey’s post-test comparing the coefficients of variation of the parameters of **automatic evaluation**.

One-way ANOVA:							
P value		< 0.0001					
Are means significantly different		Yes					
Number of groups		8					
Tukey's Multiple Comparison Test:							
Parameter A	Parameter B						
	Min HL	Max HL	Max-Min HL	Avg HL	Min HR	Max HR	Max-Min HR
Min HL							
Max HL	0.0155						
Max-Min HL	-0.1127***	-0.1282***					
Avg HL	0.0171	0.0016	0.1297 ***				
Min HR	-0.0036	-0.0191	0.1091***	-0.0206			
Max HR	0.0143	-0.0012	0.1270***	-0.0027	0.0179		
Max-Min HR	-0.1501***	-0.1656***	-0.0375	-0.1672***	-0.1466***	-0.1645***	
Avg HR	0.0177	0.0022	0.1304***	0.0007	0.0213	0.0034	0.1678***

Significant results ($P \leq 0.05$) of Tukey’s post-test are shown in bold. (*= P value 0.01 to 0.05, **= P value 0.001 to 0.01, ***= P value 0.001 to 0.0001, ****= P value <0.0001). ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values; ‘HL’ – Region of interest, left hindquarter; ‘HR’ – Region of interest, right hindquarter.

The results of Tukey’s post-test specify between which parameters the significant differences occur: The mean differences of the parameters’ VC (B-A) are displayed tabularly. A positive difference thus means that parameter B has a larger mean VC, a negative difference indicates larger values of parameter A’s coefficients of variation. The table shows distinctly that significant differences can be found between the coefficients of variation of ‘Max-Min HL’, respectively ‘Max-Min HR’ and the other parameters: As presumed, their coefficients of variation are significantly higher.

No significant differences can be found between the other parameters’ precision. Nevertheless, it is shown that ‘Min HR’ and ‘Min HL’ have larger mean coefficients of variation than ‘Avg’ and ‘Max’ in both hindquarters.

The comparison of coefficients of variation in **manual evaluation** shows similar results (see Table 4). Significant differences are again found between ‘Max-Min HL’, respectively ‘Max-Min HR’ and the other parameters. Also, ‘Min HL’ and ‘Min HR’ has larger mean coefficients of variation than ‘Max’ and ‘Avg’ of both hindquarters, although the differences are not significant.

Table 4: Results of One-way ANOVA and Tukey’s post-test comparing the coefficients of variation of the parameters of **manual** evaluation.

One-way ANOVA:							
P value		< 0.0001					
Are means significantly different		Yes					
Number of groups		8					
Tukey's Multiple Comparison Test:							
Parameter A	Parameter B						
	Min HL	Max HL	Max-Min HL	Avg HL	Min HR	Max HR	Max-Min HR
Min HL							
Max HL	0.0039						
Max-Min HL	-0.0414***	-0.0453***					
Avg HL	0.0046	0.0007	0.0460***				
Min HR	-0.0012	-0.0051	0.0402***	-0.0058			
Max HR	0.0037	-0.0003	0.0450***	-0.0010	0.0048		
Max-Min HR	-0.0459***	-0.0499***	-0.0045	-0.0506***	-0.0447***	-0.0496***	
Avg HR	0.0050	0.0010	0.0463***	0.0003	0.0061	0.0013	0.0509***

Significant results of Tukey’s post-test are shown in bold. (*=P value 0.01 to 0.05, **=P value 0.001 to 0.01, ***=P value 0.001 to 0.0001, ****=P value <0.0001). ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values; ‘HL’ – Region of interest, left hindquarter; ‘HR’ – Region of interest, right hindquarter.

The precision of ‘Max-Min HL’ and ‘Max-Min HR’ can be assessed as being lower than the precision of the other parameters in both, automatic and manual evaluation. ‘Max-Min HL’ and ‘Max-Min HR’ are thus excluded from further comparison of the methods’ precision. ‘Min HL’ and ‘Min HR’ show comparably large coefficients of variation in both methods. Consequently, the parameters ‘Max’ and ‘Avg’ are used to compare the precision of automatic and manual analysis.

The results of unmatched t-test comparing the coefficients of variation for ‘Max’ and ‘Avg’ of automatic and manual analysis are presented in Table 5. Regarding the mean values of the compared coefficients of variation, it becomes obvious that those of the parameters of automatic evaluation are larger than those of manual evaluation; consequently, the calculated differences are in the positive number range. This suggests better precision of the manual evaluation’s parameters. However, it has to be said that all calculated mean values of the parameters’

coefficients of variation are small, as well as the associated standard errors of the means. All differences calculated are small but significant; the differences between ‘Max’ in both evaluation methods are highly significant. These results lead to the conclusion that the precision of ‘Max’ and ‘Avg’ in automatic and manual evaluation is generally good, with the precision of manual evaluation being slightly, but significantly better.

Table 5: Mean values of the parameter’s coefficient of variation, results of unmatched t-test and F-test comparing previously selected parameters in precision analysis of automatic (‘Aut’) and manual (‘Man’) evaluation

Unmatched t-test Parameter A vs Parameter B	Aut Max HL	Aut Avg HL	Aut Max HR	Aut Avg HR
	vs Man Max HL	vs Man Avg HL	vs Man Max HR	vs Man Avg HR
Mean value of Parameter A	0.0043	0.0027	0.0054	0.0021
Mean value of Parameter B	0.0024	0.0017	0.0027	0.0014
Difference between means (A-B)	0.0019	0.0010	0.0028	0.0007
P-value	< 0.0001	0.0273	< 0.0001	0.0043
F-test to compare variances				
F	2.34	7.35	2.30	1.73
P-value	0.0008	< 0.0001	< 0.0001	0.0304

Significant differences are shown in bold. ‘Max’ – Maximum temperature values; ‘Avg’ – Average temperature values; ‘HL’ – Region of interest, left hindquarter; ‘HR’ – Region of interest, right hindquarter.

As described above (see: ‘Material and methods; 3.2 Precision of the different methods’), F-test analyzes if variances among the groups can be suggested as similar. The F-values calculated here do not provide definitive results. The corresponding P-values are small. Thus, results of F-test do not indicate similar variances among the compared groups. However, t-test is resistant towards unequal variances if the sample size is not small and does not differ too much from group to group. In all compared groups, the sample size is n=65. Results of t-test are thus accepted as valid.

2.2. Correlation of evaluation method and rectal temperature

The course of rectal temperature and the courses of the evaluation parameters ‘Min’, ‘Max’ and ‘Avg’ are displayed in Figure 9 (automatic evaluation method) and Figure 10 (manual evaluation method), for left and right hindquarter separately. The spots mark median values (n=5) of the named parameters at the defined time-points.

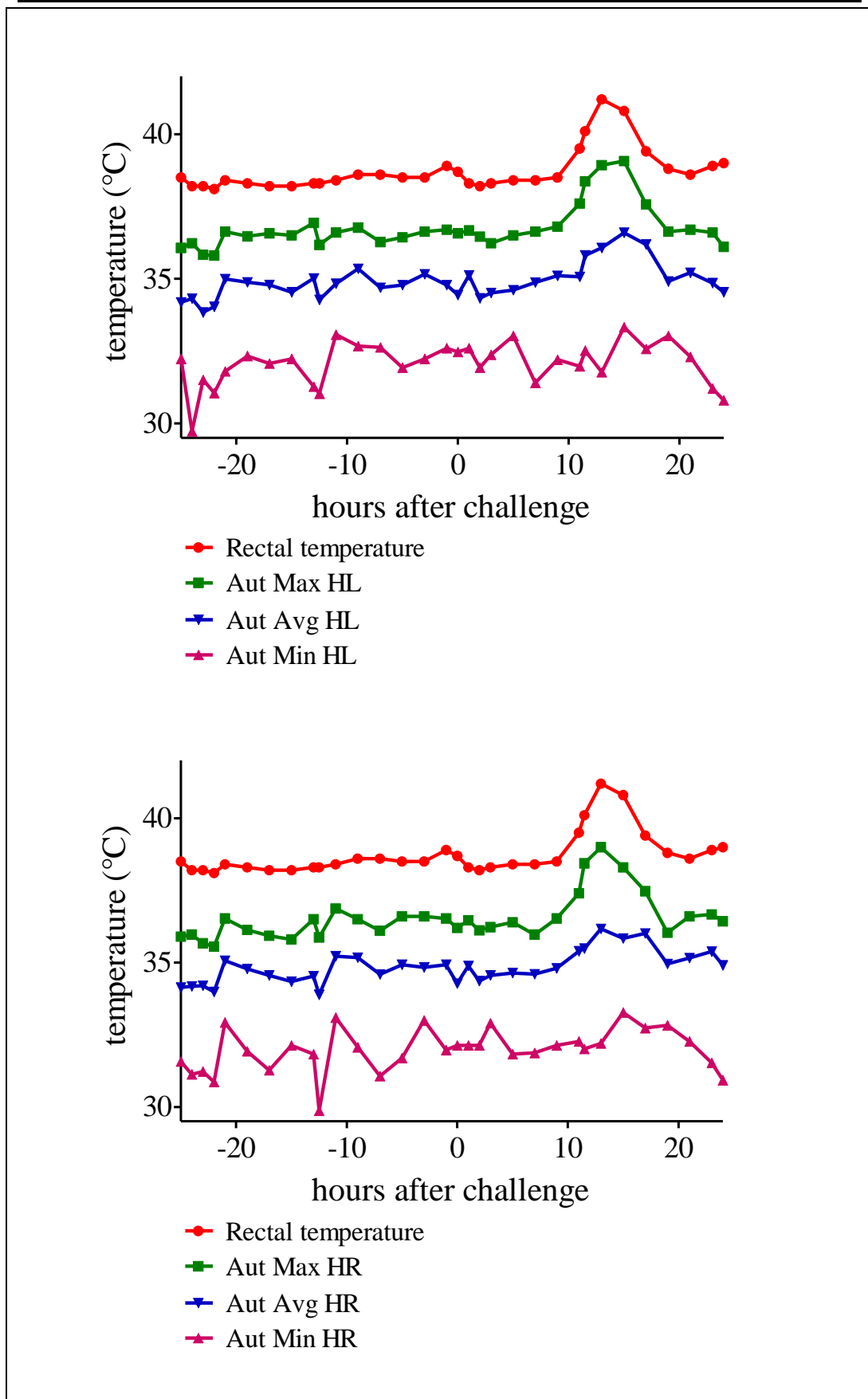


Figure 9: The course of median values of 'Max' (Maximum temperature values), 'Avg' (Average temperature values) and 'Min' (Minimum temperature values), evaluated by *automatic* method ('Aut') in the left hindquarter (HL, left) and in the right hindquarter (HR, right), set in relation with median values of rectal temperature.

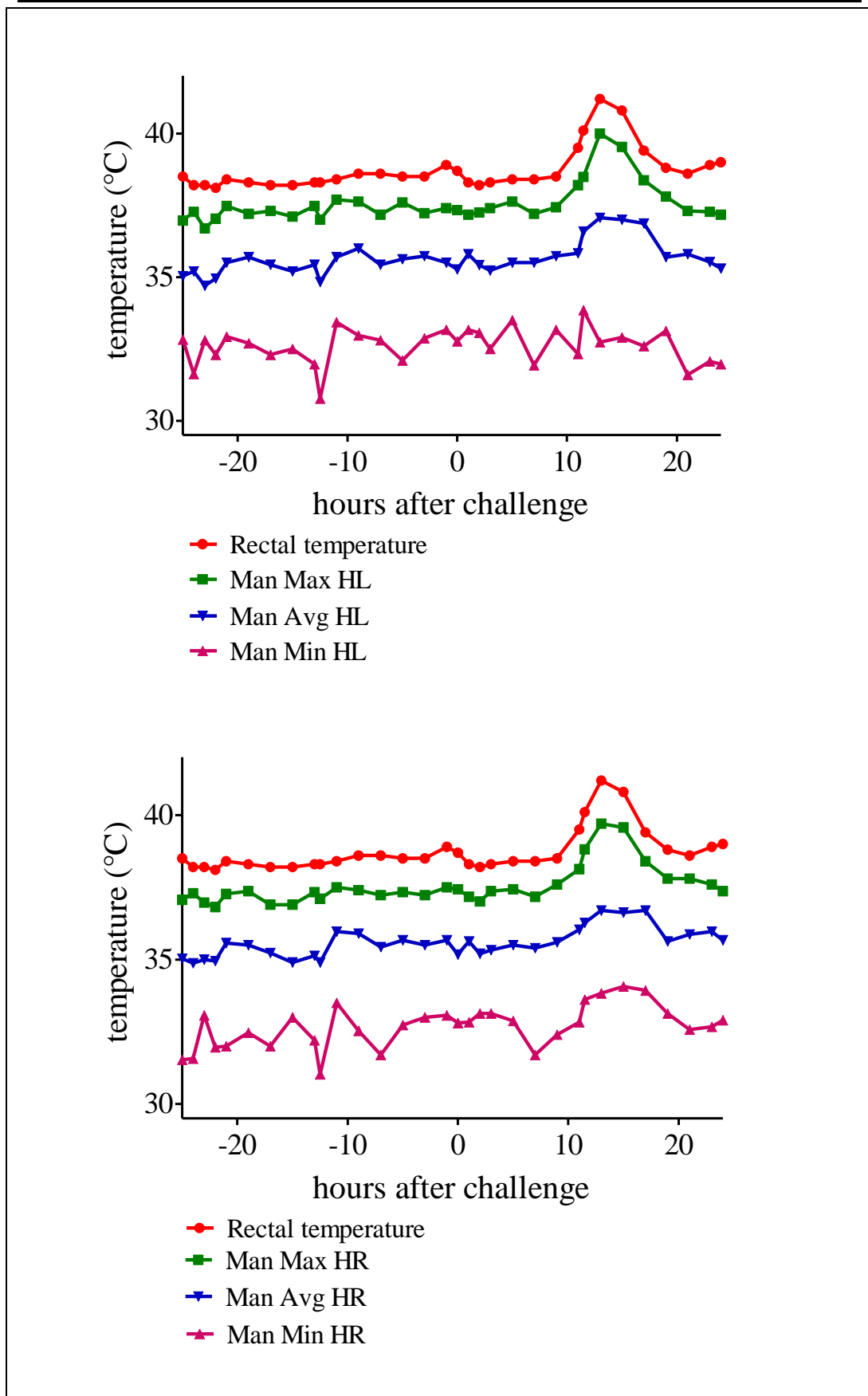


Figure 10: The course of median values of 'Max' (Maximum temperature values), 'Avg' (Average temperature values) and 'Min' (Minimum temperature values), evaluated by *manual* method ('Man') in the left hindquarter (HL, left) and in the right hindquarter (HR, right), set in relation with median values of rectal temperature.

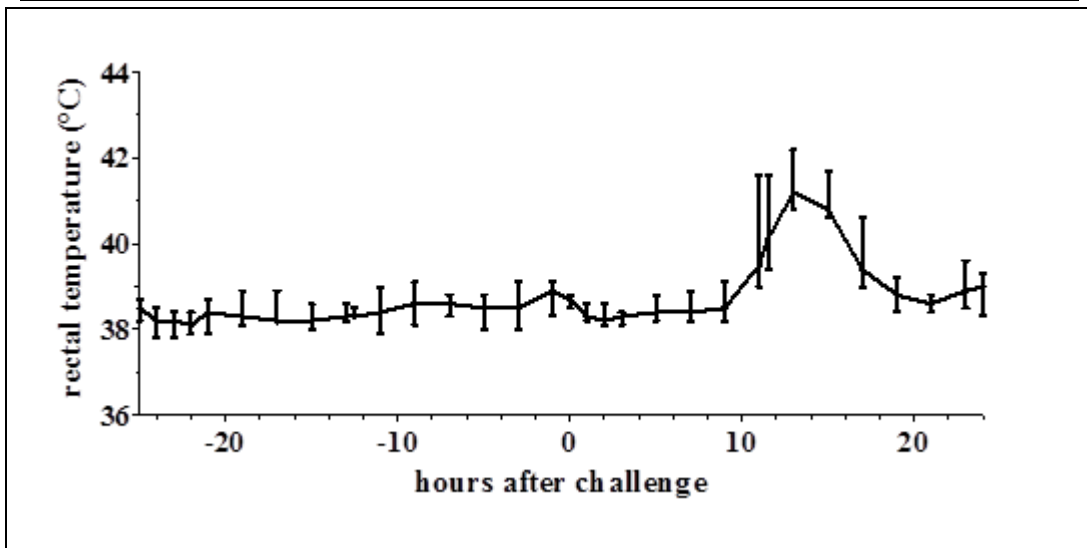


Figure 11: Course of rectal temperature of all cows throughout the trial. The line connects the median values; the bars depict the range of the values.

Figure 11 displays the course of rectal temperature in detail: The mean values of rectal temperature proceed almost horizontally: median of all rectal temperature values before challenge is 38.4°C (q1=38.2°C; q3=38.6°C). One hour before challenge, median value of rectal temperature seems to have a small peak. However, the range of the values is within the physiological limitations (37.5°C-39.5°C, VON ENGELHARDT et al. (2015)). Around the time of the challenge, there seems to be a slight drop in rectal temperature, followed by nearly horizontal temperature course up to 10 hours after challenge. At this time, a distinct increase in rectal temperature starts: median value is 40.70°C 11.5 hours after challenge, it peaks with 41.20°C at 13 hours after challenge. 15 hours after challenge, rectal temperature starts to decrease again (median value is 40.80°C), reaching the initial temperature level 19 hours after challenge. (Rectal temperatures of all cows throughout the trial are listed in Appendix 1: Tables of data obtained by automatic and manual evaluation.

Regarding Figure 9 and Figure 10, it becomes obvious that the courses of the parameter 'Max' (green line, second line from the top) follow a pattern similar to the course of rectal temperature (red line, on top). This observation concerns values from both hindquarters in both evaluation methods, although 'Man Max' seems to proceed at a higher temperature level than 'Aut Max'. The correlation coefficients of 'Max' and rectal temperature are assessed in the results of Pearson correlation analysis (see Table 6 and Table 7).

To some extent, the courses of ‘Avg’ (blue line, third line from the top) also seem to show a pattern related to rectal temperature. However, comparing the courses of ‘Avg’ in left (HL) and right hindquarter (HR), the temperature peak that is distinct in rectal temperature seems to be flattened, especially in HR, which is the challenged quarter.

Considering the curves of ‘Min’ (purple line, lowest line) in both hindquarters and in both evaluation methods, completely unsteady curves are observed.

Table 6 and Table 7 show the results of Pearson correlation analysis of rectal temperature and the named parameters, divided by evaluation method and hindquarter. Number of analyzed pairs was n=155 in each group.

Table 6: Results of Pearson correlation analysis of rectal temperature and parameters in automatic evaluation (‘Aut’)

Pearson correlation analysis of rectal temperature and evaluation parameter			
Automatic evaluation			
Evaluation parameter correlation coefficient (r)	Aut Max HL	Aut Avg HL	Aut Min HL
	0.74	0.57	0.15
P value	< 0.0001	< 0.0001	0.06
Evaluation parameter correlation coefficient (r)	Aut Max HR	Aut Avg HR	Aut Min HR
	0.79	0.49	0.24
P value	< 0.0001	< 0.0001	0.003

Significant ($P \leq 0.05$) coefficients of correlation are shown in bold.

‘Max’ – Maximum temperature values

‘Avg’ – Average temperature values

‘HL’ – Region of interest, left hindquarter

‘HR’ – Region of interest, right hindquarter

In automatic evaluation (Table 6), ‘Max HL’ and ‘Max HR’ show the best correlation coefficient (r) with rectal temperature of all parameters. They are both highly significant. ‘Avg’ and rectal temperature show moderate, but yet significant correlation coefficients in both hindquarters. They are lower as those of ‘Max’ and rectal temperature.

As suggested above, correlation coefficients of ‘Min’ and rectal temperature show poor results. Although correlation coefficient for ‘Min HR’ is significant, the values are too low to indicate a connection between the values in these parameters.

Table 7: Results of Pearson correlation analysis of rectal temperature and parameters in *manual* evaluation ('Man')

Pearson correlation analysis of rectal temperature and evaluation parameter			
Manual evaluation			
Evaluation parameter	Man Max HL	Man Avg HL	Man Min HL
correlation coefficient (r)	0.80	0.63	0.11
P value	< 0.0001	< 0.0001	0.16
Evaluation parameter	Man Max HR	Man Avg HR	Man Min HR
correlation coefficient (r)	0.85	0.54	0.30
P value	< 0.0001	< 0.0001	0.0002

Significant coefficients of correlation are shown in bold.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

The results of correlation analysis of rectal temperature and parameters of manual evaluation (Table 7) show that consistently higher correlation coefficients can be found here: All values exceed those of automatic evaluation.

'Max' shows again the best correlation coefficients with rectal temperature, followed by 'Avg'. They are all extremely significant. As expected, 'Min' in manual evaluation and rectal temperature also have small coefficients of correlation. The correlation coefficient of 'Min HR' and rectal temperature is significant, nevertheless, r is too close to zero to suggest a distinct correlation of the two parameters.

Due to the poor results of correlation analysis of rectal temperature with 'Min' in automatic and manual evaluation, 'Min' is excluded from further statistical analysis.

The differences of rectal temperature and ‘Aut Avg’, ‘respectively ‘Man Avg’ (values not displayed) are used to calculate their standard deviations. The results are shown in Table 8:

Table 8: Analysis results of differences between rectal temperature and average surface temperature

	SD of differences (mean values)		paired t-test	
	Aut	Man	difference	P value
before challenge HL	0.44	0.39	0.05	0.06
after challenge HL	0.70	0.66	0.04	0.08
total HL	0.64	0.58	0.06	0.01
before challenge HR	0.50	0.43	0.07	< 0.0001
after challenge HR	0.85	0.81	0.04	0.29
total HR	0.72	0.68	0.05	0.10

Standard deviations (SD) of differences are compared by automatic (‘Aut’) and manual (‘Man’) evaluation and tested on significant differences ($P \leq 0.05$; shown in bold). ‘HL’ – Region of interest, left hindquarter, ‘HR’ – Region of interest, right hindquarter.

Small values of standard deviation indicate a consistent difference between rectal temperature and average udder surface temperature. In both evaluation methods and both hindquarters, standard deviations of the differences are smaller before challenge than after challenge.

The mean standard deviations calculated in manual evaluation are constantly smaller than those of automatic evaluation, supporting the findings of correlation analysis that manual evaluation’s results are closer related to rectal temperature than those of automatic analysis. Nevertheless, the differences between the standard deviations are only significant between ‘total HL’ and ‘before challenge HR’.

Furthermore, it is noticeable that the standard deviations of the differences in the right hindquarter are larger than those in the left hindquarter. This applies to both evaluation methods. Apparently, average surface temperature of the right (challenged) hindquarter is less related to rectal temperature than average surface temperature of the left (unaffected) hindquarter. Correlation coefficients of rectal temperature and ‘Avg’ support this observation (see Table 6 and Table 7).

2.3. Comparison of automatic and manual evaluation

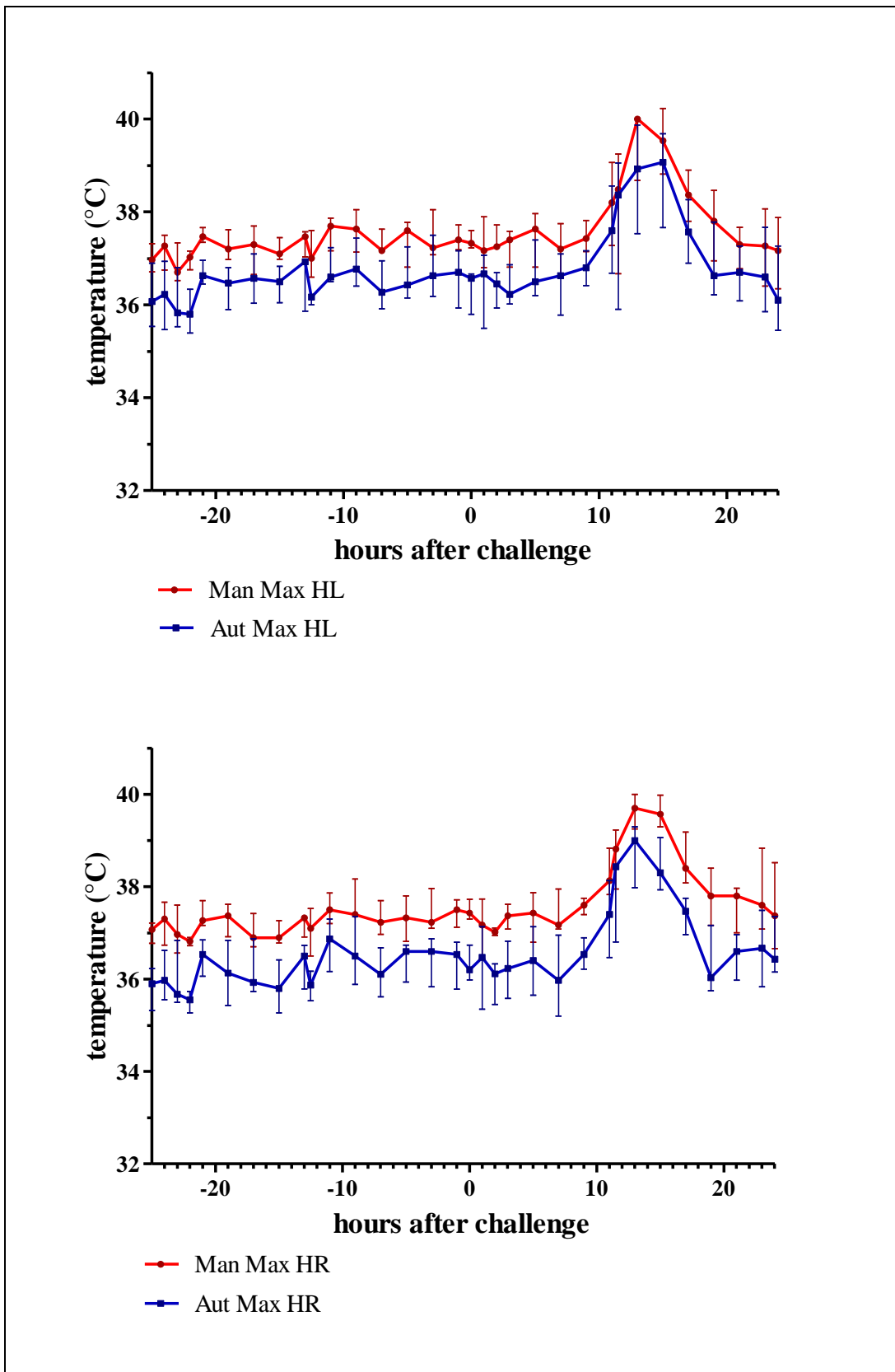


Figure 12: Temperature courses of maximum surface temperature ('Max') in automatic ('Aut', blue) and manual ('Man', red) evaluation throughout the trial, separated by left (HL) and right (HR) hindquarter. Lines connect median values, bars depict range of the values.

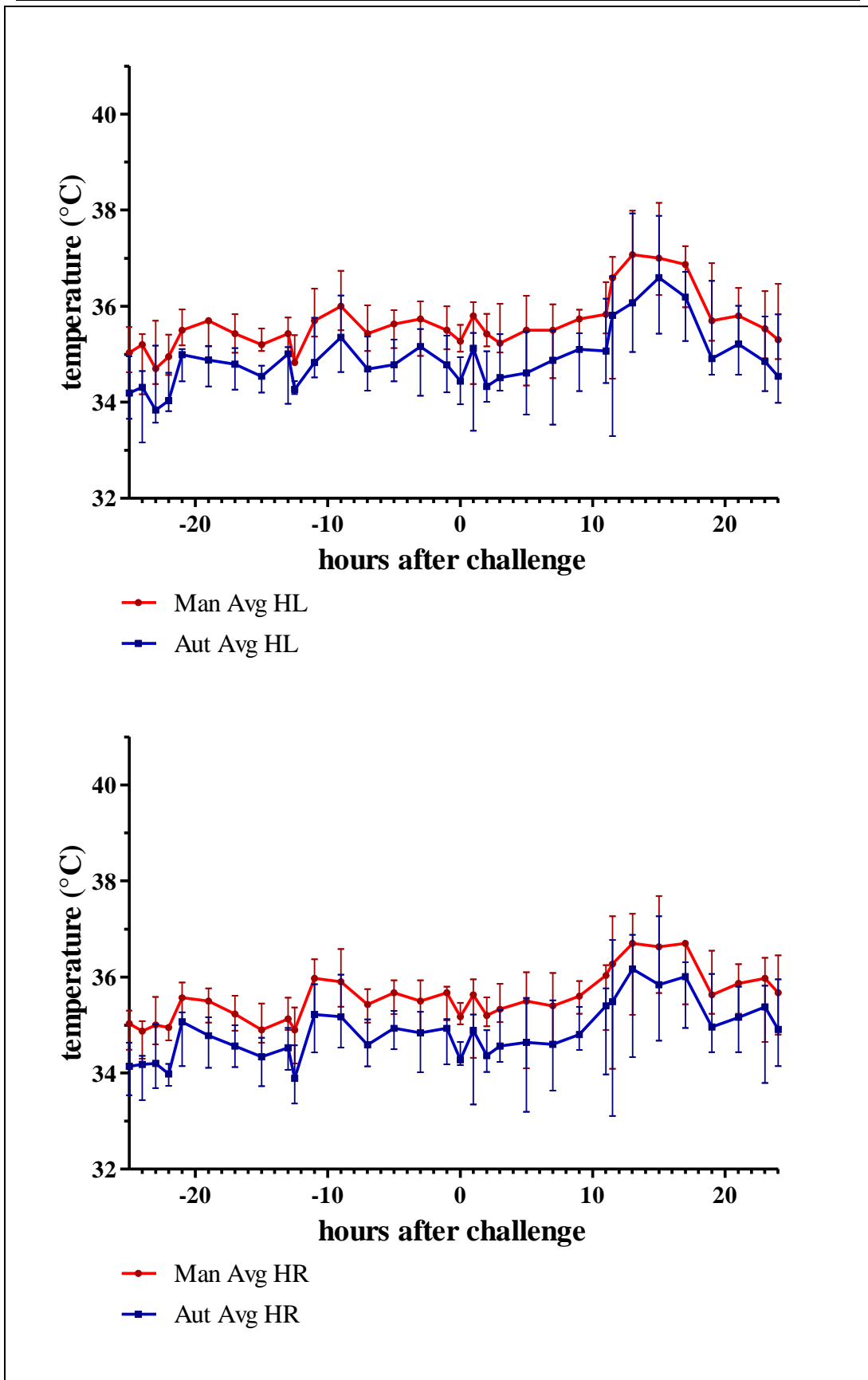


Figure 13: Temperature courses of average surface temperature ('Avg') in automatic ('Aut', blue) and manual ('Man', red) evaluation throughout the trial, separated by left (HL) and right (HR) hindquarter. Lines connect median values, bars depict range of the values.

In Figure 12 and Figure 13, temperature courses of the parameters ‘Max’ and ‘Avg’ throughout the trial are displayed, comparing automatic and manual evaluation. All temperature values obtained by automatic and manual evaluation are presented in Table 27 to Table 32 in ‘Appendix 1: Tables of data obtained by automatic and manual evaluation’.

Regarding Figure 12, values of ‘Max’ seem to proceed at a consistent level in automatic and manual evaluation before challenge. Medians of ‘Max’ and ‘Avg’ before and after challenge are shown in Table 9.

Table 9: Medians of maximum temperature values (‘Max’) and average temperature values (‘Avg’) in automatic and manual evaluation, separated by period before challenge (reference period) and period after challenge.

	Max HL	Max HR	Avg HL	Avg HR
Automatic				
reference period	36.52	36.31	34.78	34.62
period after challenge	36.74	36.55	34.97	34.97
Manual				
reference period	37.35	37.31	35.51	35.39
period after challenge	37.44	37.58	35.61	35.63

‘HL’ – Region of interest, left hindquarter

‘HR’ – Region of interest, right hindquarter

A slight drop in temperature is visible 23 hours before challenge for both methods in both hindquarters. Another oscillation in maximum temperature seems to occur between 13 hours and 10 hours before challenge.

Regarding the line of medians of maximum temperature, it is distinct that the values obtained by automatic evaluation are constantly on a lower temperature level than temperature values of manual evaluation: the median difference of ‘Man Max’ HL and ‘Aut Max’ HL is 0.81°C before challenge and 0.74°C after challenge, respectively 0.77°C throughout the whole trial. The median difference of ‘Man Max’ HR and ‘Aut Max’ HR is 1.09°C before challenge and after challenge, respectively 1.08°C throughout the whole trial.

The same observation can be made regarding Figure 13. Median values of ‘Aut Avg’ are consistently lower than median values of ‘Man Avg’ in both hindquarters: the median difference of ‘Man Avg’ HL and ‘Aut Avg’ HL is 0.74°C before

challenge and 0.66°C after challenge, respectively 0.72°C throughout the whole trial. The median difference of ‘Man Avg’ HR and ‘Aut Avg’ HR is 0.75°C challenge and 0.69°C after challenge, respectively 0.71°C throughout the whole trial.

Nevertheless, the temperature courses in Figure 12 and Figure 13 appear similar in automatic and manual evaluation, although the temperature levels differ. To what extent automatic and manual evaluation actually correlate is calculated in correlation analysis (see ‘2.3.1 Correlation analysis of automatic and manual evaluation’). Elevation in surface temperature begins around 11 hours after challenge. At which time points significant temperature changes can be detected in the different methods, the different hindquarters and the different evaluation parameters is evaluated later (see ‘2.3.2 Comparison of period after challenge and reference period in automatic and manual method’), but the appearance of the graphs in Figure 12 indicate that the courses of ‘Max’ in automatic evaluation show a faster increase than the courses of ‘Max’ in manual evaluation.

Both methods show temperature peaks at 13 and 15 hours after challenge for ‘Avg’ and ‘Max’ in both hindquarters, although the temperature peaks in ‘Avg’ are less distinct than in ‘Max’. The amplitude of the peaks seems to be similarly high in both methods. Differences between the period after challenge and reference period are analyzed later (see: ‘2.3.2 Comparison of period after challenge and reference period in automatic and manual method’).

It is noticeable that similar temperature patterns are observed in both hindquarters, although only the right hindquarter was challenged whereas the left hindquarter was treated with a placebo.

At the end of the trial, 24 hours after challenge, temperature values approach the baseline of temperature that was observed at the reference period.

2.3.1. Correlation analysis of automatic and manual evaluation

Scatter plots of the temperature values of ‘Avg’ and ‘Max’ in automatic and manual evaluation are presented in Figure 27 in Appendix 5. In each of the four graphs, $n=155$ pairs. Since temperature values obtained by manual evaluation are 0.81°C higher on the average than those of automatic evaluation (see ‘2.3 Comparison of automatic and manual evaluation’), the scatter plots are slightly shifted rightwards on the X-axis in all four graphs. Nevertheless, the scatter plots of ‘Avg HL’ and ‘Avg HR’ distinctly form in the shape of a line. The appearance of the scatter plots of ‘Max HL’ and ‘Max HR’ also distinctly indicate a strong linear correlation.

Due to the appearance of the scatter plots, Pearson correlation analysis is performed for ‘Avg’ and ‘Max’ in automatic and manual evaluation method, separately for each hindquarter. The results are presented in Table 10:

Table 10: Results of Pearson correlation analysis of evaluation parameters ‘Max’ (maximum temperature) and ‘Avg’ (average temperature) in automatic (‘Aut’) and manual (‘Man’) evaluation

Pearson correlation analysis of automatic and manual evaluation		
Evaluation parameter	Aut Max HL & Man Max HL	Aut Max HR & Man Max HR
correlation coefficient (r)	0.92	0.90
P value	< 0.0001	< 0.0001
Evaluation parameter	Aut Avg HL & Man Avg HL	Aut Avg HR & Man Avg HR
correlation coefficient (r)	0.98	0.99
P value	< 0.0001	< 0.0001

Significant ($P \leq 0.05$) coefficients of correlation are shown in bold.

It is noticeable that in all analyzed evaluation parameters, the results indicate a strong positive linear correlation between automatic and manual evaluation method. All results are highly significant.

Regarding the temperature course graphs in Figure 9 and Figure 10, strong correlation between automatic and manual evaluation can be expected.

Correlation coefficients of ‘Max HL’ and ‘Max HR’ in automatic and manual evaluation are slightly inferior to correlation coefficients of ‘Avg HL’ and ‘Avg

HR', which provide almost perfect correlation results of automatic and manual evaluation.

Again, differences in correlation of the methods between the affected and unaffected quarters are not apparent.

2.3.2. Comparison of period after challenge and reference period in automatic and manual method

Reminder: Differences of measured USST in period after challenge and reference period are calculated for 13 time points after challenge and 13 time points before challenge (see Table 2).

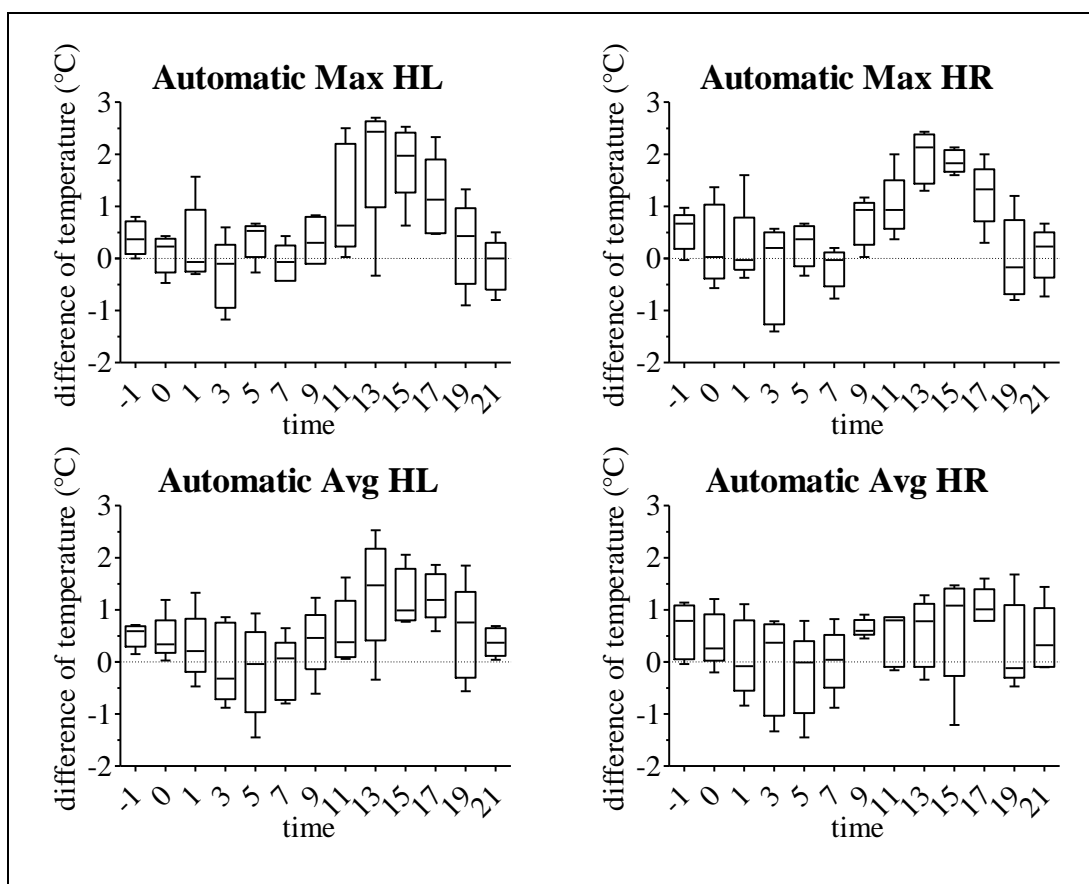


Figure 14: Box-plot diagrams of temperature differences in automatic evaluation. Lower and upper whisker display minimum and maximum values. Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

The median value of the differences of temperature values in automatic evaluation as well as the, first and third quartile are shown in Table 43 in Appendix 5. For visualization, differences are displayed in box-plot diagrams (see Figure 14). The

according box-plot diagram for manual evaluation is shown in Figure 15, the list of the median differences, q1 and q3 in Table 44 in Appendix 5.

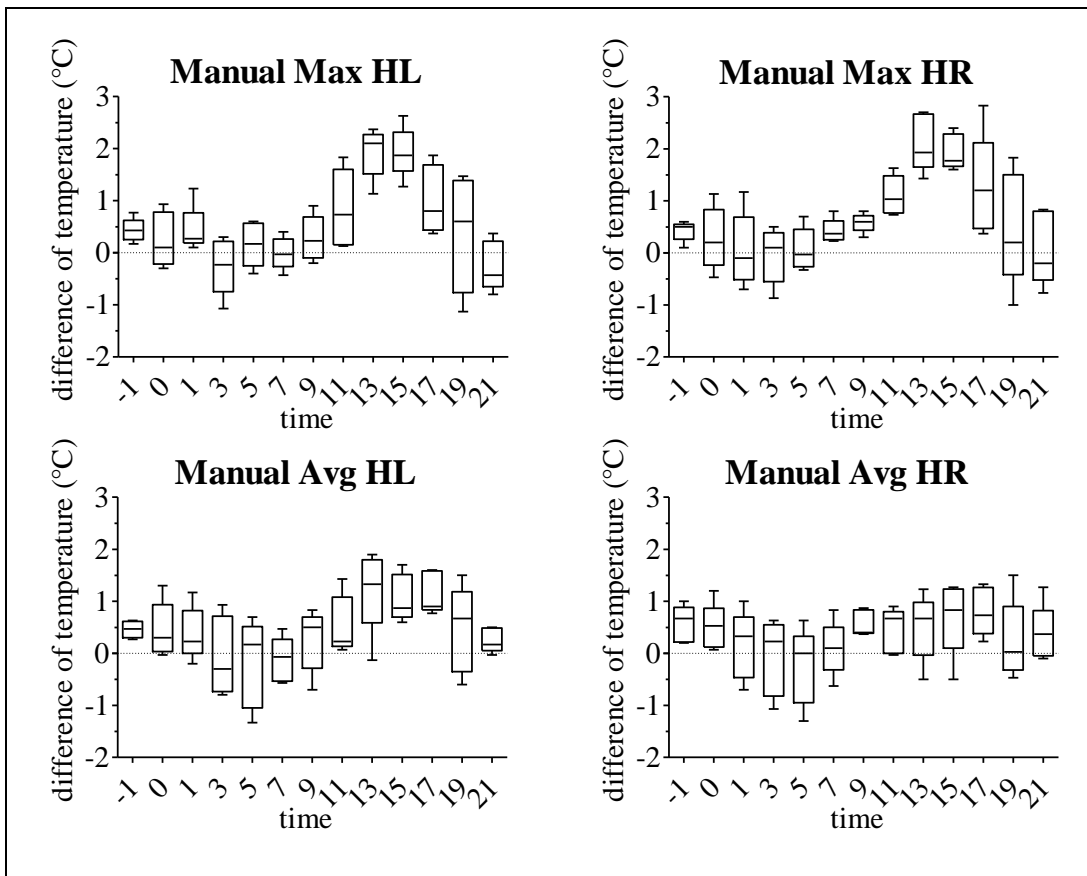


Figure 15: Box-plot diagrams of temperature differences in manual evaluation. Lower and upper whisker display minimum and maximum values. Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

In all four evaluation parameters, differences are above zero at '-1'. In manual evaluation, temperature differences are in a similar range at that time. Median temperature differences of the time of challenge versus 24 hours before ('0') approximate zero in both methods, yet there is a broad scattering.

In the following time-points ('1' to '7'), median values of temperature differences of the parameters in both methods level around the baseline of zero in both, automatic and manual evaluation. Nevertheless, the box-plot diagrams show a wider range of differences in automatic evaluation than in manual evaluation.

At '9', larger differences are observed in all evaluation parameters. In the following, temperature differences continue to increase. At '11', temperature differences are

higher in 'Max' than in 'Avg' and also higher in the challenged quarter (HR) than in the unaffected quarter (HL). This applies to both evaluation methods. It is remarkable, that in both methods, 'Avg HL' is the only evaluation parameter to show slightly decreasing temperature differences between '9' and '11'.

In both methods, temperature differences of the evaluation parameter 'Max' are distinctly largest at '13', '15' and '17'.

Concerning the evaluation parameter 'Avg', the increase in the temperature differences between '13' and '17' is not as pronounced as in 'Max', but still detectable.

It is to say that temperature differences after challenge are observed in both, challenged (HR) and unaffected (HL) hindquarter. In fact, differences at '13' and '15' are larger in the left hindquarter than in the right hindquarter, indicating a larger increase in surface temperature in the unaffected quarter than in the challenged quarter. In differences of 'Avg', this observation is even more distinct. Concerning only the evaluation parameter 'Max', differences detected in the right hindquarter exceed those of the left hindquarter at '17', when temperature differences are decreasing again. At '15', the median difference of 'Aut Avg HR' (1.08°C) slightly exceeds the median difference of 'Aut Avg HL' (0.99°C).

After the increase in temperature differences at '13', '15' and '17', differences decrease again and return to the level of the baseline at '21' in both methods.

These findings suggest that no distinct temperature differences can be detected 21 hours after challenge. Furthermore, it can be expected that significant temperature differences between period after challenge and reference period can be found at '13', '15' and '17'. Regarding the results, it is noticeable that evaluation parameters of automatic evaluation mostly show larger temperature differences than those of manual evaluation. The following one-way analysis of variances is performed to clarify if the differences occurring between period after challenge and reference period are significant, and if they are, it is of interest at what time they are significant. In addition, it is clarified whether one evaluation method is able to detect more or earlier significant temperature differences than the other evaluation method.

Table 11 shows the results of one-way ANOVA and Dunett's multiple comparison post-test in automatic evaluation, Table 12 shows the according results in manual evaluation.

*Table 11: Results of One-way ANOVA and Dunett's post-test comparing the temperature differences of reference period and period after challenge in **automatic** evaluation*

Automatic evaluation				
One-way ANOVA:	Max HL	Max HR	Avg HL	Avg HR
P value	< 0.0001	< 0.0001	0.0049	0.2261
significant different group means	Yes	Yes	Yes	No
Number of groups	13	13	13	13
Dunett's Multiple Comparison Test:	Max HL	Max HR	Avg HL	Avg HR
mean difference of '-1' and				
0	0.30	0.28	0.05	0.19
1	0.13	0.32	0.21	0.53
3	0.69	0.81	0.56	0.66
5	0.03	0.28	0.67	0.85
7	0.48	0.72	0.64	0.60
9	0.05	-0.18	0.11	-0.04
11	-0.70	-0.47	-0.07	0.15
13	-1.54 (*)	-1.41 (**)	-0.82	0.05
15	-1.47 (*)	-1.32 (**)	-0.72	-0.06
17	-0.79	-0.70	-0.74	-0.46
19	0.12	0.56	-0.06	0.32
21	0.51	0.44	0.13	0.17

Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Significant results (P value ≤ 0.05) are shown in bold. (=P value 0.01 to 0.05, **=P value 0.001 to 0.01, ***=P value 0.001 to 0.0001, ****=P value < 0.0001).*

In the one-way analysis of variances of temperature differences in automatic evaluation, highly significant differences are detected among the groups in the evaluation parameter 'Max' in the challenged as well as in the placebo-treated quarter. Among the temperature differences of 'Aut Avg HL', one-way ANOVA detects highly significant differences, whereas no significant differences are detected among the temperature differences of 'Aut Avg HR'.

In Dunett's multiple comparison test, the temperature differences of the different parameters in all groups ('0' to '21') are compared with the temperature differences of the evaluation parameters at '-1'. The table shows the mean difference of the differences in the two compared groups (difference at '-1' – difference at 'x'). Large negative values thus indicate large differences. Significant mean differences are shown in bold. The parameters 'Aut Max HL' and 'Aut Max HR' provide significant mean differences between '-1' and '13' and between '-1' and '15'. However, between '-1' and '17', no significant differences can be detected. The same applies to the comparison of '-1' to all other groups.

Although one-way ANOVA calculated very significant differences among the groups of differences of 'Aut Avg HL', no significant differences between the differences at '-1' and the differences of the other groups can be found in Dunett's multiple comparison test. However, mean differences among the groups are large at '13', at '15' and at '17'.

The comparison of temperature differences of 'Aut Avg HR' provides no significant differences between the differences at '-1' and all other groups. Mean differences between '-1' and '13', respectively between '-1' and '15', where 'Aut Max HL' and 'Aut Max HR' detected significant differences are small. The largest mean difference of 'Aut Avg HR' is found between '-1' and '17', yet not significant.

The results of one-way analysis of variances and Dunett's post-test of temperature differences in manual evaluation method (see Table 12) are in large parts consistent with those of automatic evaluation method. One-way ANOVA detects highly significant differences among the temperature differences of 'Man Max HL' and 'Man Max HR'.

In Dunett's multiple comparison post-test comparing the temperature differences of '-1' with all other temperature differences, significant differences can be found for 'Man Max HL' and 'Man Max HR' between '-1' and '13' and between '-1' and '15'. Between the temperature differences of '-1' and '17' no significant differences can be found, and mean differences decrease again.

Regarding the results of comparing the differences of average surface temperature in manual evaluation, no significant differences can be found between the differences of '-1' and all other differences, although one-way ANOVA indicated

significant differences among the groups for at least ‘Man Avg HL’. For ‘Man Avg HL’, mean differences are largest for the differences between ‘-1’ and ‘13’, between ‘-1’ and ‘15’ and between ‘-1’ and ‘17’. The according differences calculated with ‘Man Avg HR’, the challenged quarter, are small between ‘-1’ and ‘13’, between ‘-1’ and ‘15’ and between ‘-1’ and ‘17’.

Table 12: Results of One-way ANOVA and Dunett’s post-test comparing the temperature differences of reference period and period after challenge in *manual* evaluation

Manual evaluation				
One-way ANOVA:	Max HL	Max HR	Avg HL	Avg HR
P value	< 0.0001	< 0.0001	0.0023	0.19
significant different group means	Yes	Yes	Yes	No
Number of groups	13	13	13	13
Dunett's Multiple Comparison Test:	Max HL	Max HR	Avg HL	Avg HR
mean difference of '-1' and				
0	0.19	0.15	0.01	0.07
1	0.00	0.38	0.09	0.41
3	0.69	0.47	0.53	0.64
5	0.27	0.36	0.64	0.82
7	0.44	0.01	0.58	0.48
9	0.15	-0.15	0.19	0.01
11	-0.41	-0.68	-0.07	0.12
13	-1.50 (**)	-1.68 (***)	-0.76	0.06
15	-1.49 (**)	-1.51 (**)	-0.60	-0.13
17	-0.57	-0.85	-0.69	-0.23
19	0.07	-0.05	-0.01	0.34
21	0.69	0.35	0.21	0.19

Differences are calculated for values measured after challenge (‘-1’ to ‘21’ hours, time (h)) and values measured 24 hours earlier.

‘Max’ – Maximum temperature values

‘Avg’ – Average temperature values

‘HL’ – Region of interest, left hindquarter

‘HR’ – Region of interest, right hindquarter

Significant results ($P \text{ value} \leq 0.05$) are shown in bold. (*= $P \text{ value } 0.01 \text{ to } 0.05$, **= $P \text{ value } 0.001 \text{ to } 0.01$, ***= $P \text{ value } 0.001 \text{ to } 0.0001$, ****= $P \text{ value } < 0.0001$).

Comparing the results of one-way ANOVA and Dunett’s multiple comparison test of the temperature differences in automatic and manual evaluation, both methods show similar results:

For both methods, the temperature differences of the evaluation parameters ‘Max HL’ and ‘Max HR’ are detected as significantly different in one-way ANOVA and significant differences could be found between the differences of ‘-1’ and ‘13’ as well as between the differences of ‘-1’ and ‘15’.

Regarding the analysis of temperature differences of average surface temperature, the results are also consistent in both methods: differences of ‘Avg HL’ are significantly different in one-way ANOVA, but no significant differences can be found in Dunett’s multiple comparison test. ‘Avg HR’ provided the poorest results in detecting temperature differences between surface temperature of the udder before and after challenge, although it concerns the challenged quarter.

2.3.3. Determining threshold values for automatic and manual evaluation

Box-plot graphs of USST values detected when cows showed rectal temperatures below 39.5°C versus USST values detected when cows had rectal temperatures above 39.5°C are displayed in Figure 29 in Appendix 6, as well as the associated ROC curves (Figure 30 in Appendix 6).

Table 13: Results of ROC (Receiver-Operating-Characteristics) analysis: Area underneath curve (AUC) and P-Value of evaluation parameters in automatic (‘Aut’) and manual (‘Man’) evaluation discriminating between healthy udder and udder suffering from clinical mastitis

ROC Analysis		
	AUC	P-Value
Automatic		
Aut Max HL	0.87	< 0.0001
Aut Max HR	0.96	< 0.0001
Aut Avg HL	0.83	< 0.0001
Aut Avg HR	0.75	0.0010
Manual		
Man Max HL	0.92	< 0.0001
Man Max HR	0.98	< 0.0001
Man Avg HL	0.86	< 0.0001
Man Avg HR	0.78	0.0002

‘Max’ – Maximum temperature values

‘Avg’ – Average temperature values

‘HL’ – Region of interest, left hindquarter

‘HR’ – Region of interest, right hindquarter

In Table 13, the AUCs of the evaluation parameters ‘Avg’ and ‘Max’ in the left and right hindquarter in automatic and manual evaluation method are displayed.

Since all P-values are small, results of ROC analysis of all evaluation parameters are considered as significant. In both methods, ‘Max HR’ obtains the largest area underneath the ROC curve, followed by ‘Max HL’. ‘Avg HL’ still shows moderately good results in ROC analysis, whereas the results of ‘Avg’ in the challenged quarter (HR) are poorer. It is noticeable that AUCs of evaluation parameters in manual evaluation are constantly larger than those of automatic evaluation, although the differences are small (differences between the AUCs range from 0.02 to 0.05).

Table 14 shows the threshold values of the evaluation parameters in automatic and manual evaluation gaining the maximum sum of sensitivity and specificity (Youden’s Index).

Table 14: Results of ROC (Receiver-Operating-Characteristics) analysis: Threshold value, sensitivity, specificity and sum of sensitivity and specificity of evaluation parameters in automatic (‘Aut’) and manual (‘Man’) evaluation

Sensitivity and Specificity				
	Threshold (°C)	Sensitivity (%)	Specificity (%)	Sum
Automatic				
Aut Max HL	> 37.50	81.25	92.09	173.34
Aut Max HR	> 37.42	93.75	94.96	188.71
Aut Avg HL	> 35.83	75.00	92.09	167.09
Aut Avg HR	> 35.80	62.50	92.81	155.31
Manual				
Man Max HL	> 38.07	93.75	92.09	185.84
Man Max HR	> 38.65	93.75	96.40	190.15
Man Avg HL	> 36.65	75.00	95.68	170.68
Man Avg HR	> 36.55	62.50	96.40	158.90

‘Max’ – Maximum temperature values

‘Avg’ – Average temperature values

‘HL’ – Region of interest, left hindquarter

‘HR’ – Region of interest, right hindquarter

‘Max HR’ gains the best results of sensitivity and specificity in both evaluation methods ‘Max HL’ gains slightly lower, but still good results.

Concerning the results of 'Avg', sensitivity and specificity of detecting clinical mastitis are again poorer in the challenged quarter (HR) than in the placebo-treated quarter (HL).

As already indicated by the AUCs, sensitivity and specificity of evaluation parameters in manual evaluation are equal or slightly higher than sensitivity and specificity of evaluation parameters in automatic evaluation.

3. Results of evaluation method using only the teats and evaluation method using the udder surface

3.1. Precision of evaluation using the teats and evaluation using the udder

Precision analysis results are presented as box-plots in Figure 16. Sample size of VCs for every parameter in each method is $n=65$. Small VCs indicate good precision.

The median as well as first quartile (q_1) and third quartile (q_3) of VCs of the different measurement parameters are displayed tabularly in Table 42 in Appendix 4.

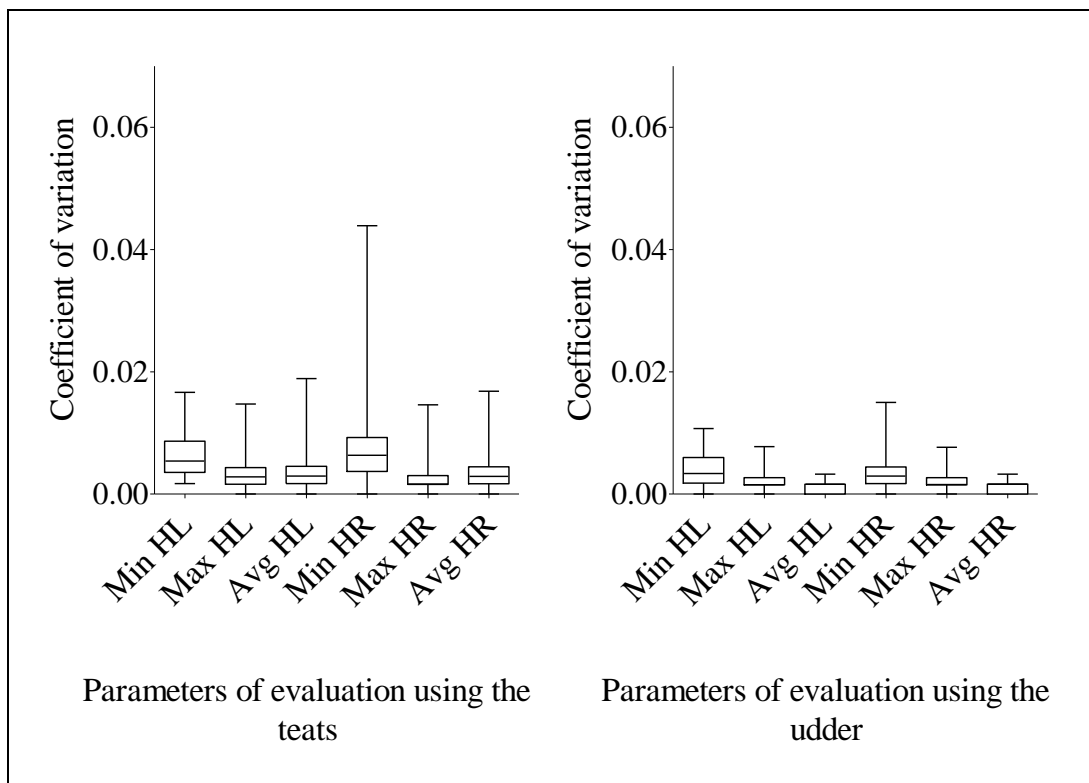


Figure 16: Coefficients of variations displayed in Box-plot diagrams, comparison of evaluation using the teats and evaluation using the udder (coefficients of variation of ‘Max-Min’ not displayed). The box represents values from first to third quartile, the line marks the median. Bars depict the range of the values.

‘Min’ – Minimum temperature values

‘Max’ – Maximum temperature values

‘Avg’ – Average temperature values

‘HL’ – Region of interest, left hindquarter, respectively left hind teat

‘HR’ – Region of interest, right hindquarter, respectively right hind teat

Similarly to automatic and manual evaluation, VCs of ‘Min HL’ and ‘Min HR’ are larger than those of the other parameters in Figure 16. Also, results in Table 42 indicate that values of VCs of ‘Max-Min’ again exceed those of the other parameters by far. Hence, they are not displayed in the box-plot-charts of Figure 16. Moreover, evaluation using the udder seems to have lower VCs in most of the

parameters than evaluation using the teats. Especially the parameters ‘Avg HL’ and ‘Avg HR’ seem to differ from method to method. In following analysis, it is tested if significant differences can be assumed. One-way ANOVA tests if the group’s mean VCs differ generally, whereas Tukey’s post-test directly compares each parameter’s VCs with another (parameter A vs. parameter B).

Table 15: Results of One-way ANOVA and Tukey’s post-test comparing the coefficients of variation of the parameters of evaluation *using the teats*

One-way ANOVA:							
P value	< 0.0001						
Are means significantly different	Yes						
Number of groups	8						
Tukey's Multiple Comparison Test:							
Parameter A	Parameter B						
	Min HL	Max HL	Max-Min HL	Avg HL	Min HR	Max HR	Max-Min HR
Min HL							
Max HL	0.0031						
Max-Min HL	-0.0762***	-0.0792***					
Avg HL	0.003	-0.0001	0.0792***				
Min HR	-0.0007	-0.0038	0.0755***	-0.0037			
Max HR	0.0044	0.0013	0.0805***	0.0014	0.0050		
Max-Min HR	-0.0815***	-0.0846***	-0.0053	-0.0845***	-0.0808***	-0.0859***	
Avg HR	0.0033	0.0002	0.0795***	0.0003	0.0040	-0.0010	0.0848***

Significant results of Tukey’s post-test are shown in bold. (*=P value 0.01 to 0.05, **=P value 0.001 to 0.01, ***=P value 0.001 to 0.0001, ****=P value <0.0001). ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values; ‘HL’ – Region of interest, left hind teat; ‘HR’ – Region of interest, right hind teat.

Results of One-way ANOVA and Tukey’s post-test comparing the parameters’ VCs within both methods are displayed in Table 15 and Table 16. In both evaluation methods, VC of the parameters ‘Max-Min’ are constantly significantly larger than those of all other parameters. This distinctly underlines the inferior precision of ‘Max-Min’. However, between ‘Max-Min HL’ and ‘Max-Min HR’, no significant differences are detectable.

Table 16: Results of One-way ANOVA and Tukey's post-test comparing the coefficients of variation of the parameters of evaluation **using the udder**

One-way ANOVA:							
P value		< 0.0001					
Are means significantly different		Yes					
Number of groups		8					
Tukey's Multiple Comparison Test:							
Parameter A	Parameter B						
	Min HL	Max HL	Max-Min HL	Avg HL	Min HR	Max HR	Max-Min HR
Min HL							
Max HL	0.0022						
Max-Min HL	-0.0385***	-0.0408***					
Avg HL	0.0029	0.0007	0.0415***				
Min HR	0.0007	-0.0015	0.0392 ***	-0.0022			
Max HR	0.0019	-0.0003	0.0404***	-0.0010	0.001		
Max-Min HR	-0.0334***	-0.0356***	0.0051	-0.0363***	-0.0341***	-0.0353***	
Avg HR	0.0028	0.0006	0.0413***	-0.0001	0.002	0.0009	0.0362***

Significant results of Tukey's post-test are shown in bold. (*=P value 0.01 to 0.05, **=P value 0.001 to 0.01, ***=P value 0.001 to 0.0001, ****=P value <0.0001). 'Min' – Minimum temperature values; 'Max' – Maximum temperature values; 'Max-Min' – Range of temperature values; 'Avg' – Average temperature values; 'HL' – Region of interest, left hindquarter; 'HR' – Region of interest, right hindquarter.

Comparing VCs of 'Min' with those of 'Max' and 'Avg', results of Tukey's post-test of both methods indicate that 'Min' constantly shows larger VCs and thus a lower precision. However, differences are not significant.

For further comparison of the methods' precision, VCs of the parameters 'Max' and 'Avg' are used. The results of unmatched t-test and associated F-test are shown in Table 17:

Table 17: Mean values of the parameter's coefficient of variation, results of unmatched t-test and F-test comparing previously selected parameters in precision analysis of evaluation using the teats ('Teats') and evaluation using the udder ('Udder')

Unmatched t-test Parameter A vs Parameter B	Teats Max HL	Teats Avg HL	Teats Max HR	Teats Avg HR
	vs Udder Max HL	vs Udder Avg HL	vs Udder Max HR	vs Udder Avg HR
Mean value of Parameter A	0.0035	0.0036	0.0023	0.0033
Mean value of Parameter B	0.0019	0.0012	0.0023	0.0014
Difference between means (A-B)	0.0016	0.0024	-0.00001	0.0019
P value	0.0002	< 0.0001	0.9689	< 0.0001
F-test to compare variances				
F	4.15	10.00	2.13	6.92
P value	< 0.0001	< 0.0001	0.0018	< 0.0001

Significant differences are shown in bold. Parameters: 'Max' – Maximum temperature values; 'Avg' – Average temperature values; 'HL' – Region of interest, left hindquarter, respectively left hind teat; 'HR' – Region of interest, right hindquarter, respectively right hind teat.

The mean values of the VC of 'Max HL', 'Avg HL' and 'Avg HR' are larger in evaluation using only the teats than those in evaluation using the region of the udder. Their differences are all highly significant ($P < .001$) and in the positive number range. VCs of 'Max HR' are nearly equal in both methods. Consequently, no significant difference can be found here.

The results of t-test indicate a significantly higher precision of evaluation method using the udder. However, it has to be taken into consideration that in one of four compared parameters, no differences could be found.

The F-values calculated in F-test appear widespread. The corresponding P values are small. Thus, results of F-test do not suggest equal variances among the compared groups. However, t-test is resistant towards unequal variances if the sample size is not small and does not differ too much from group to group (as described above, see 'Material and Methods: Precision of the different methods'). In all compared groups, sample size is $n=65$. Results of t-test are thus accepted as valid.

3.2. Correlation of evaluation method and rectal temperature

In Figure 17 and Figure 18, courses of median values of rectal temperature and 'Min', 'Max' and 'Avg' in evaluation using the teats and evaluation using the udder are shown, divided by left and right hindquarter.

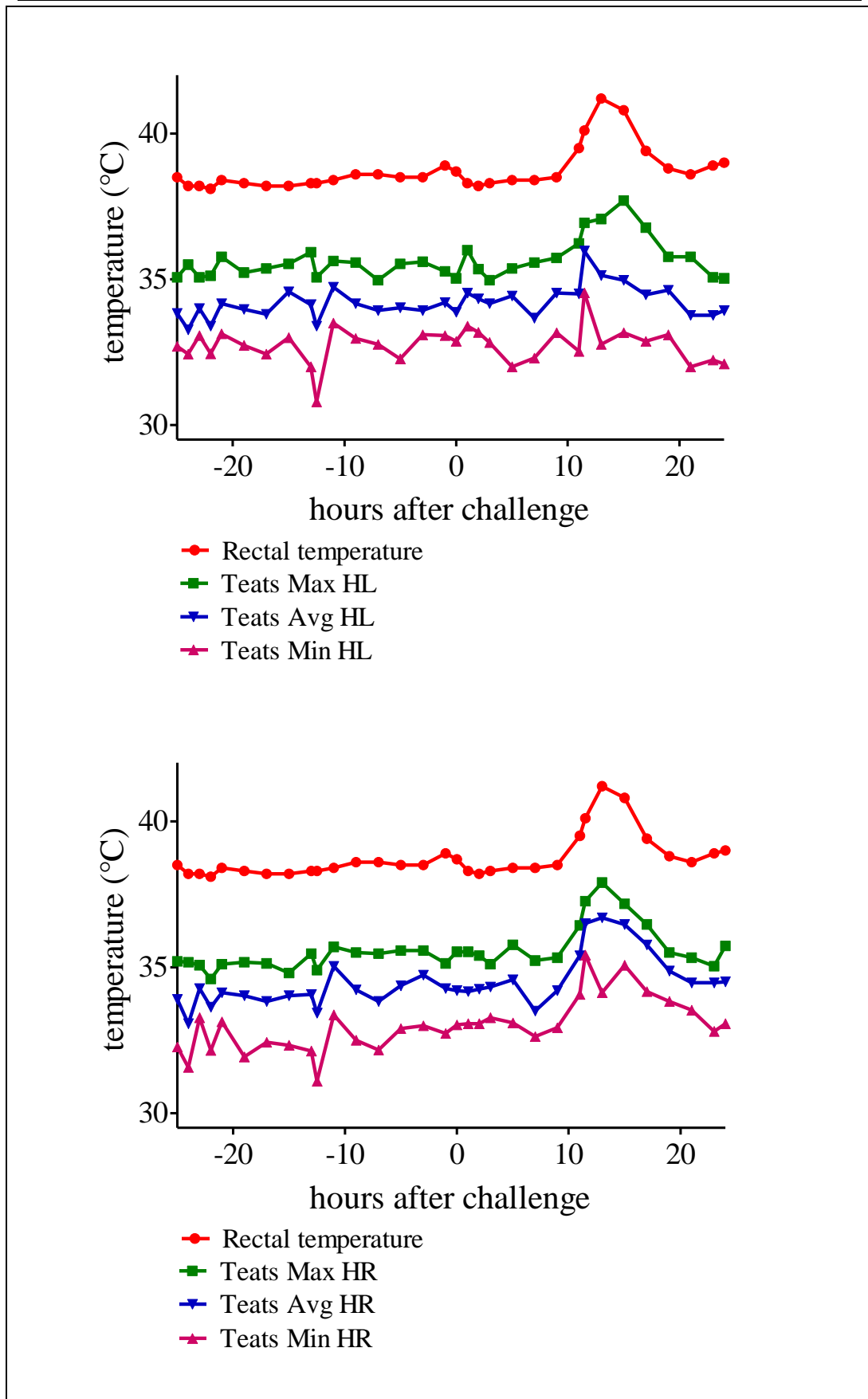


Figure 17: The course of median values of 'Max' (Maximum temperature values), 'Avg' (Average temperature values) and 'Min' (Minimum temperature values), evaluated by **evaluation using the teats** ('Teats') in the left hind teat (HL, left) and in the right hind teat (HR, right), set in relation with median values of rectal temperature.

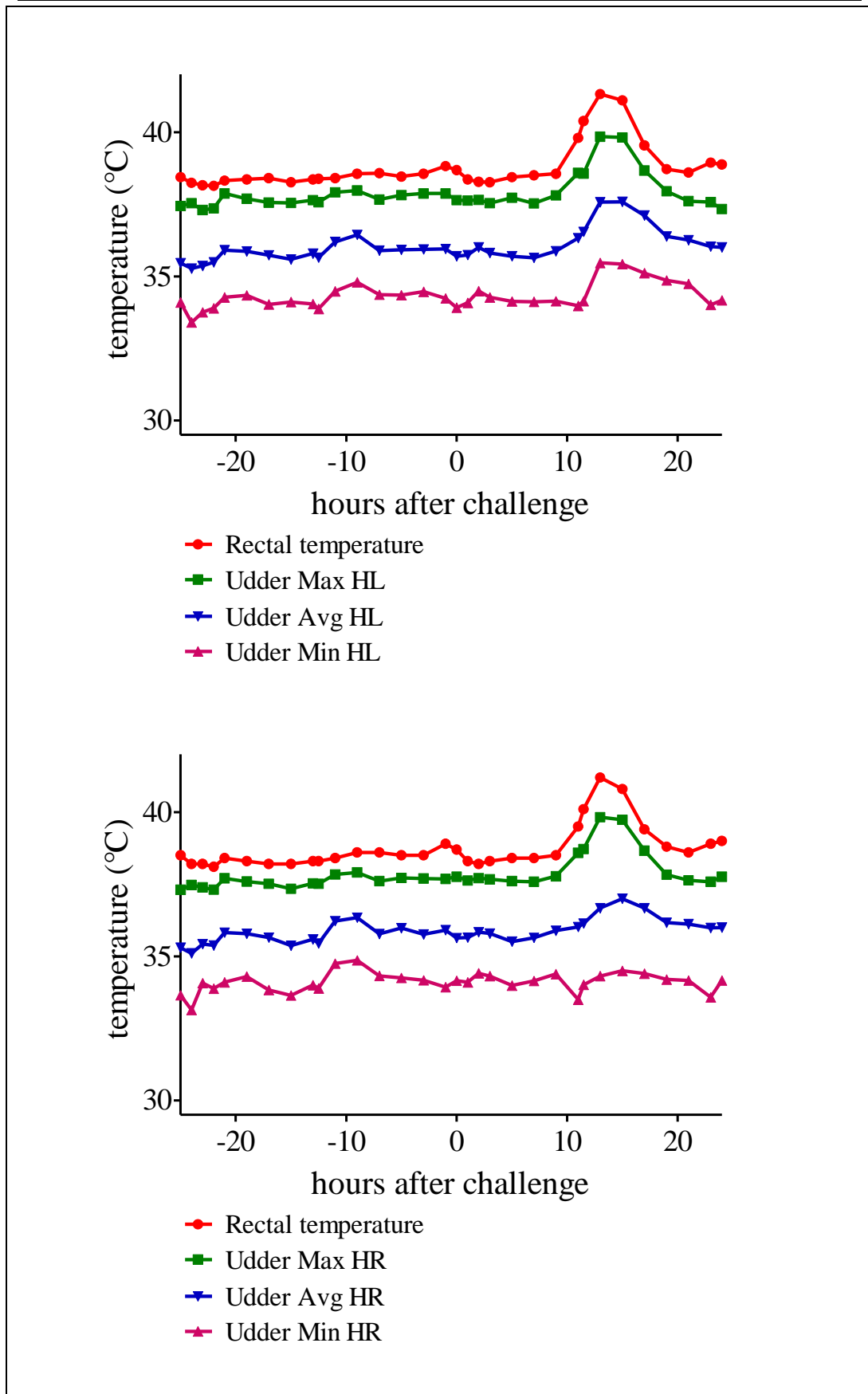


Figure 18: The course of median values of 'Max' (Maximum temperature values), 'Avg' (Average temperature values) and 'Min' (Minimum temperature values), evaluated by **evaluation using the udder** ('Udder') in the left hindquarter (HL, left) and in the right hindquarter (HR, right), set in relation with median values of rectal temperature.

The course of rectal temperature is described in detail in ‘3.3 Correlation of evaluation method and rectal temperature’ (see Figure 11).

Comparing Figure 17 with Figure 18, it becomes obvious that temperature courses of the parameters in evaluation using only the teats do not resemble the course of rectal temperature, the way the courses of parameters in evaluation using the udder do.

The course of ‘Teats Max’ also shows a distinct temperature peak 13 and 15 hours after challenge, but appears to run irregularly, compared to the course of ‘Udder Max’. The parameter ‘Teats Avg’ shows even more irregularities in the course throughout the trial. In the right hindquarter (HR), the temperature peak at 13 and 15 hours after challenge is visible, whereas the curve of the left hindquarter (HL) shows two smaller peaks in this period of time.

The course of the parameter ‘Min’ seems to run in an arbitrary pattern in evaluations using only the teats. In evaluations using the udder, the course of ‘Min’ seems to be more related to rectal temperature, but the peak is less distinct, compared to ‘Avg’ and ‘Max’ (HL), respectively no peak is visible (HR).

Table 18: Results of Pearson correlation analysis of rectal temperature and parameters in evaluation using the teats (‘Teats’)

Pearson correlation analysis of rectal temperature and evaluation parameter				
Evaluation using the teats	Evaluation parameter correlation coefficient (r)	Teats Max HL	Teats Avg HL	Teats Min HL
	P value	0.57	0.31	0.12
		< 0.0001	< 0.0001	0.12
	Evaluation parameter correlation coefficient (r)	Teats Max HR	Teats Avg HR	Teats Min HR
	P value	0.67	0.56	0.36
		< 0.0001	< 0.0001	< 0.0001

Significant ($P \leq 0.05$) coefficients of correlation are shown in bold.

‘Max’ – Maximum temperature values

‘Avg’ – Average temperature values

‘HL’ – Region of interest, left hind teat

‘HR’ – Region of interest, right hind teat

Table 19: Results of Pearson correlation analysis of rectal temperature and parameters in evaluation **using the udder** ('Udder')

Pearson correlation analysis of rectal temperature and evaluation parameter				
Evaluation using the udder	Evaluation parameter	Udder Max HL	Udder Avg HL	Udder Min HL
	correlation coefficient (r)	0.80	0.66	0.35
	P value	< 0.0001	< 0.0001	< 0.0001
	Evaluation parameter	Udder Max HR	Udder Avg HR	Udder Min HR
	correlation coefficient (r)	0.83	0.52	0.18
	P value	< 0.0001	< 0.0001	0.02

Significant ($P \leq 0.05$) coefficients of correlation are shown in bold.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Regarding the results of Pearson correlation analysis (Table 18 and Table 19), the parameter 'Max' yields the best correlations with rectal temperature in both evaluation methods and in both hindquarters. Their correlation coefficient results are all extremely significant. However, the correlation coefficients of 'Max' and rectal temperature in evaluation **using the udder** are higher than those in evaluation **using only the teats**.

A highly significant positive correlation was also detected for 'Avg' and rectal temperature in both hindquarters and evaluation methods, although it has to be said that correlation coefficients only indicate poor to moderate correlation. Again, results of evaluation using the udder exceed those of evaluation using the teats.

'Teats Min HR' still has a significant correlation coefficient with rectal temperature, nevertheless the correlation coefficient is low. In the left hind teat, no correlation with 'Min' and rectal temperature can be detected.

In the evaluation using the udder, the parameter 'Min' yields the poorest results of correlation with rectal temperature of the three parameters. However, correlation coefficient results are significant.

It is notable that, in the evaluation using only the teats (see Table 18), all parameters applied at the right (challenged) teat yielded higher results in correlation with rectal temperature than the parameters applied at the left (unaffected) teat. In evaluation

using the udder (see Table 19), this cannot be observed: correlation coefficient of rectal temperature with ‘Max HR’ is slightly higher than with ‘Max HL’, but regarding the correlation of ‘Avg’ and ‘Min’ with rectal temperature, the left hindquarter exceeds the right hindquarter.

Since the parameter ‘Min’ shows inconstant and, in evaluation using the teats, multiply peaked temperature courses, it is most likely that this parameter is prone to falsifications. ‘Min’ is thus excluded from further statistical analysis.

The differences of rectal temperature and ‘Teats Avg’, respectively ‘Udder Avg’ (values not displayed) are used to calculate their standard deviations. The results are shown in Table 20. Small values of standard deviation indicate a consistent difference between rectal temperature and average surface temperature. In evaluation method using the udder, standard deviations of the differences are smaller before challenge than after challenge.

Table 20: Analysis results of differences between rectal temperature and average surface temperature

	SD of differences (mean values)		paired t-test	
	Teats	Udder	difference	P value
before challenge HL	0.80	0.39	0.41	0.002
after challenge HL	1.13	0.68	0.45	0.07
total HL	0.82	0.58	0.25	0.02
before challenge HR	0.94	0.44	0.50	0.002
after challenge HR	0.81	0.88	-0.06	0.67
total HR	0.91	0.71	0.20	0.02

Standard deviations (SD) of differences are compared by evaluation using the teats (‘Teats’) and evaluation using the udder (‘Udder’) and tested on significant differences ($P \leq 0.05$; shown in bold). ‘HL’ – Region of interest, left hindquarter, respectively left hind teat; ‘HR’ – Region of interest, right hindquarter, respectively right hind teat.

In evaluation using only the region of the teats, smaller standard deviations are also seen before challenge than after challenge in the left hindquarter, but at the right hindquarter the mean standard deviation is slightly lower after challenge than before challenge. As seen in the comparison of automatic and manual evaluation, temperature values seem to differ to a larger extent after challenge in the udder surface, but this effect is not clearly observed in the region of the teats.

In the right hindquarter after challenge the mean standard deviation is lower in evaluation using the teats than in evaluation using the udder. The difference is not

significant. In all other findings, the mean standard deviations calculated in evaluation using the udder are constantly smaller than those of evaluation using the teats, indicating that the surface temperature of the udder is closer related to rectal temperature than surface temperature of the teats. These findings support the results of correlation analysis (see Table 18 and Table 19). The differences between the mean SD of evaluation using and using the udder are significant in both hindquarters before challenge and in total. However, the differences between the methods are larger before challenge than after challenge and in total.

As already observed in comparison of automatic and manual analysis, the differences' standard deviations of the right hindquarter are also larger than those of the left hindquarter. As suggested above, average surface temperature of the right (challenged) hindquarter is less related to rectal temperature than average surface temperature of the left (unaffected) hindquarter, and this assumption can apparently be extended to the region of the teats. However, correlation coefficients of rectal temperature and 'Avg' support this observation only for the region of the udder, whereas for the region of the teats correlation coefficients oppose these findings (see Table 18 and Table 19).

3.3. Comparison of evaluation using the teats and evaluation using the udder

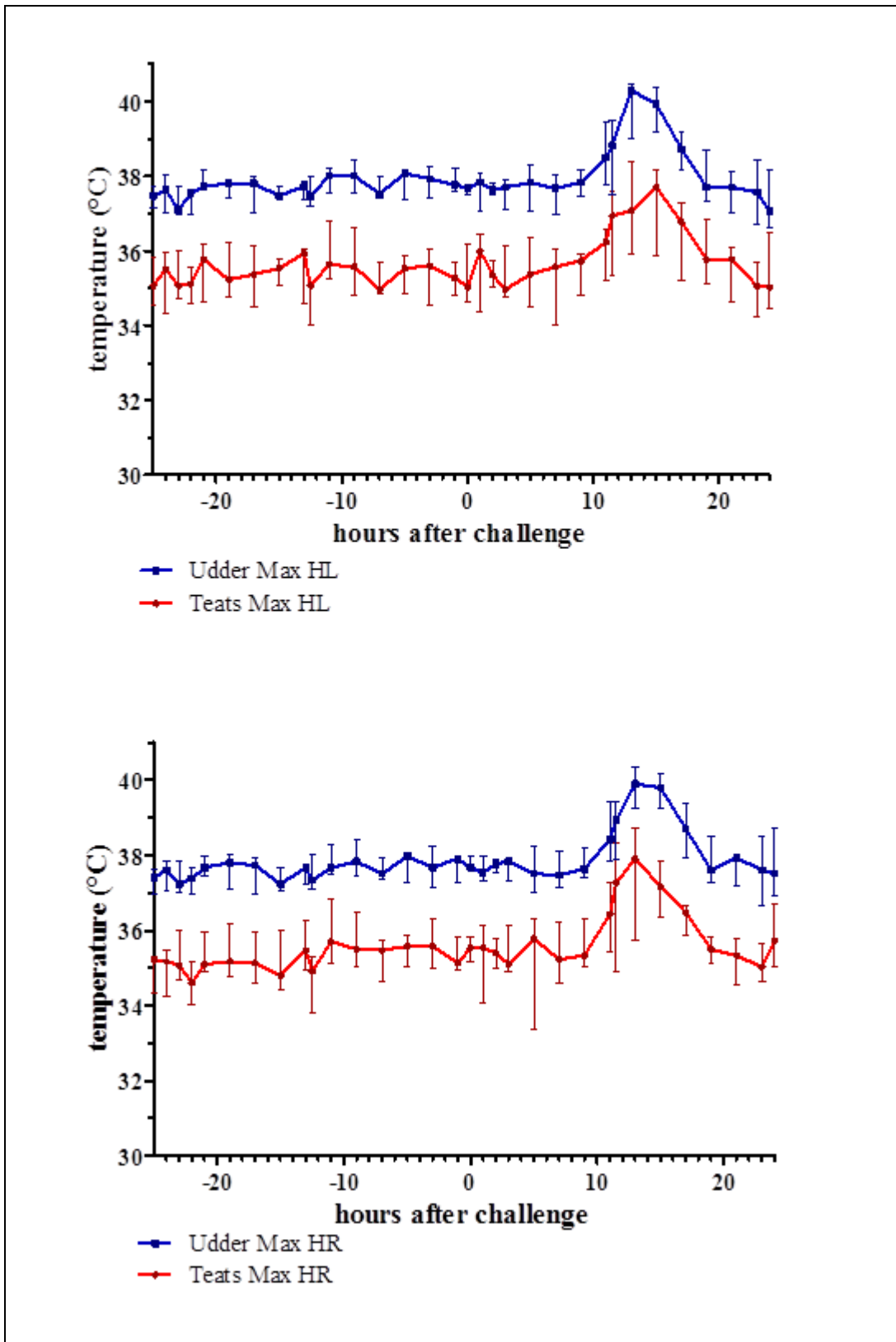


Figure 19: Temperature courses of maximum surface temperature ('Max') in evaluation using the teats ('Teats', red) and evaluation using the udder ('Udder', blue) evaluation throughout the trial, separated by left (HL) and right (HR) hindquarter, respectively teats. Lines connect median values, bars depict range of the values.

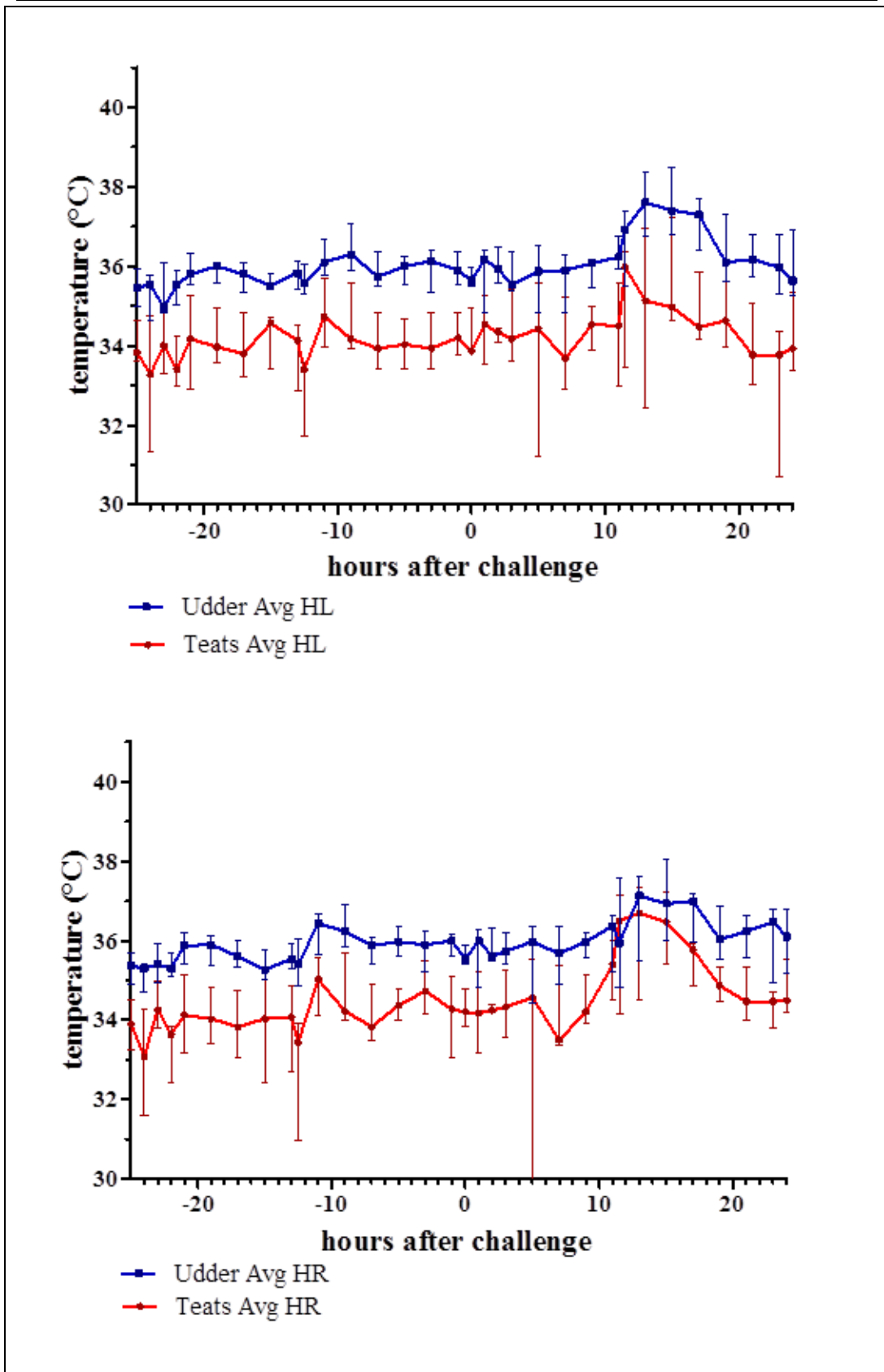


Figure 20: Temperature courses of average surface temperature ('Avg') in evaluation using the teats ('Teats', red) and evaluation using the udder ('Udder', blue) evaluation throughout the trial, separated by left (HL) and right (HR) hindquarter, respectively teats. Lines connect median values, bars depict range of the values.

Figure 19 and Figure 20 present the temperature courses of the parameters ‘Max’ and ‘Avg’ throughout the trial. In all graphs, the courses of ‘Teats Avg’ and ‘Teats Max’ seem to run in a less consistent pattern than the courses of ‘Udder Avg’ and ‘Udder Max’. Nevertheless, the graph’s patterns indicate that values before challenge proceed around a baseline in both methods. Medians of ‘Max’ and ‘Avg’ before and after challenge are shown in Table 21:

Table 21: Medians of maximum temperature values (‘Max’) and average temperature values (‘Avg’) in evaluation using the teats (‘Teats’) and evaluation using the udder (‘Udder’), separated by period before challenge (reference period) and period after challenge

	Max HL	Max HR	Avg HL	Avg HR
Teats				
reference period	35.37	35.36	34.10	34.11
period after challenge	35.45	35.47	34.43	34.62
Udder				
reference period	37.67	37.61	35.87	35.76
period after challenge	37.73	37.76	36.03	36.00

‘HL’ – Region of interest, left hindquarter, respectively left hind teat

‘HR’ – Region of interest, right hindquarter, respectively right hind teat

The courses of ‘Avg’ at reference period show more oscillation than the courses of ‘Max’ at reference period, especially the oscillation between 15 hours and 11 hours before challenge that is already described in the comparison of automatic and manual evaluation is also distinct here.

Nevertheless, the interquartile range of the medians of ‘Avg’ in both evaluation methods is narrow, indicating no wide scattering around the baseline (interquartile range of ‘Avg’ here not depicted). All temperature values obtained by evaluation using the teats and evaluation using the udder are presented in Table 33 to Table 38 in Appendix 2: Tables of data obtained by evaluation using the teats and evaluation using the udder.

In Figure 19 and Figure 20 it becomes distinct that the values of average and maximum surface temperature of the teats are constantly on a lower temperature level than those of the udder. Calculating differences of both method’s results approve this assumption: the difference of ‘Udder Max’ HL and ‘Teats Max’ HL is 2.40°C before challenge and 2.10°C after challenge, respectively 2.30°C

throughout the whole trial. The difference of 'Udder Max' HR and 'Teats Max' HR is 2.33°C before challenge and 2.23°C after challenge, respectively 2.23°C throughout the whole trial.

Regarding the differences of the median values of 'Avg' in evaluation using the udder and evaluation using the teats, similar observations are made: the difference of 'Udder Avg' HL and 'Teats Avg' HL is 1.77°C before challenge and 1.73°C after challenge, respectively 1.75°C throughout the whole trial. The difference of 'Udder Avg' HR and 'Teats Avg' HR is 1.60°C challenge and 1.40°C after challenge, respectively 1.53°C throughout the whole trial.

The graphs of the courses of 'Max' in both methods and both hindquarters seem to proceed in a similar pattern, although the temperature levels clearly differ (see Figure 19). The courses of 'Avg' seem to be less related (see Figure 20) and the differences between the results are less consistent than in the evaluation parameter 'Max': the interquartile range of the differences of 'Max' in the methods are 0.27 (HL) and 0.30 (HR), whereas the interquartile range of the differences of 'Avg' are 0.87 (HL) and 0.60 (HR) after challenge.

To what extent the results of both evaluation parameters in evaluation using the teats and using the udder are associated is calculated in correlation analysis (see '3.3.1 Correlation analysis of evaluation using the teats and evaluation using the udder').

The temperature peaks at 13 and 15 hours after challenge were mentioned in the comparison of automatic and manual evaluation method, and they can also be distinctly observed in the parameter 'Max' of evaluation using the teats and evaluation using the udder.

Regarding the results of the parameter 'Avg' (see Figure 20), it is noticeable that in the left (unaffected) hindquarter, the mentioned temperature peak is apparently distinct in evaluation using the udder, but less pronounced in evaluation using the teats.

In the right (challenged) hindquarter, 'Avg' in evaluation using the teats seems to show an earlier and more pronounced temperature elevation than 'Avg' in evaluation using the udder, and 11.5 hours after challenge, median value of 'Teats Avg' HR (36.5°C) exceeds the median value of 'Udder Avg' HR (35.93°C).

Differences between the temperature values in the period after challenge and the reference period are calculated later (see ‘3.3.2 Comparison of period after challenge and reference period in evaluation using the teats and evaluation using the udder’).

Similar to the observations made in the comparison of automatic and manual evaluation, temperature values also quickly decrease again after the temperature peak. This is applicable to ‘Avg’ and ‘Max’ in both evaluation methods. 24 hours after challenge, temperature values once more approach the baseline of temperature that was observed at reference period.

3.3.1. Correlation analysis of evaluation using the teats and evaluation using the udder

The appearance of the temperature courses of ‘Avg’ and ‘Max’ in evaluation using the teats and evaluation using the udder (see Figure 17 and Figure 18) indicate a good correlation between the methods’ evaluation results, especially for results of the evaluation parameter ‘Max’.

In the scatter plots in Figure 28 in Appendix 5, temperature values of ‘Avg’ and ‘Max’ in evaluation using the teats (X-axis) are plotted against the temperature values of ‘Avg’ and ‘Max’ in evaluation using the udder (Y-axis), separated by hindquarter. In each of the four graphs, n=155 pairs. Since temperature values obtained by evaluation using the teats are 1.95°C lower on the average than those of evaluation using the udder (see ‘3.3 Comparison of evaluation using the teats and evaluation using the udder’), the scatter plots are shifted upwards on the Y-axis in all four graphs. The appearance of the scatter plots of ‘Max HL’ and ‘Max HR’ imply a linear distribution. This linear distribution is also visible in the scatter plots of ‘Avg HL’ and ‘Avg HR’, although they appear more widespread.

Due to the appearance of the scatter plots, Pearson correlation analysis is performed for ‘Avg’ and ‘Max’ in evaluation method using the teats and evaluation method using the udder, separated by hindquarter. The results are presented in Table 22:

Table 22: Results of Pearson correlation analysis of evaluation parameters ‘Max’ (maximum temperature) and ‘Avg’ (average temperature) in evaluation using the teats (‘Teats’) and evaluation using the udder (‘Udder’)

Pearson correlation analysis of evaluation using the teats and evaluation using the udder		
Evaluation parameter	Teats Max HL & Udder Max HL	Teats Max HR & Udder Max HR
correlation coefficient (r)	0.78	0.82
P value	< 0.0001	< 0.0001
Evaluation parameter	Teats Avg HL & Udder Avg HL	Teats Avg HR & Udder Avg HR
correlation coefficient (r)	0.64	0.73
P value	< 0.0001	< 0.0001

Significant ($P \leq 0.05$) coefficients of correlation are shown in bold. ‘HL’ – Region of interest, left hindquarter, respectively left hind teat; ‘HR’ – Region of interest, right hindquarter, respectively right hind teat

As expected regarding the temperature courses, the evaluation parameter ‘Max’ yields better results in correlation analysis than the evaluation parameter ‘Avg’. The results thus indicate a good positive linear correlation between evaluation method using the teats and evaluation method using the udder when evaluation parameter ‘Max’ is used, and a moderate to good correlation when evaluation parameter ‘Avg’ is used. All results are highly significant.

In this correlation analysis, results of the parameters ascertained at the challenged quarter (HR) show better correlation of the methods than those of the unaffected quarter (HL).

3.3.2. Comparison of period after challenge and reference period in evaluation using the teats and evaluation using the udder

Reminder: Differences of measured surface temperature in period after challenge and in reference period are calculated for 13 time points after challenge and 13 time points before challenge (see Table 2)

The median values of the differences of temperature values in evaluation using the teats as well as the first and third quartiles are listed in Table 45 in Appendix 5. For visualization, differences are displayed in box-plot diagrams (see Figure 21). Lower and upper whisker display minimum and maximum values. The according table and diagrams for evaluation using the udder are shown in Table 46 in Appendix 5 and in Figure 22.

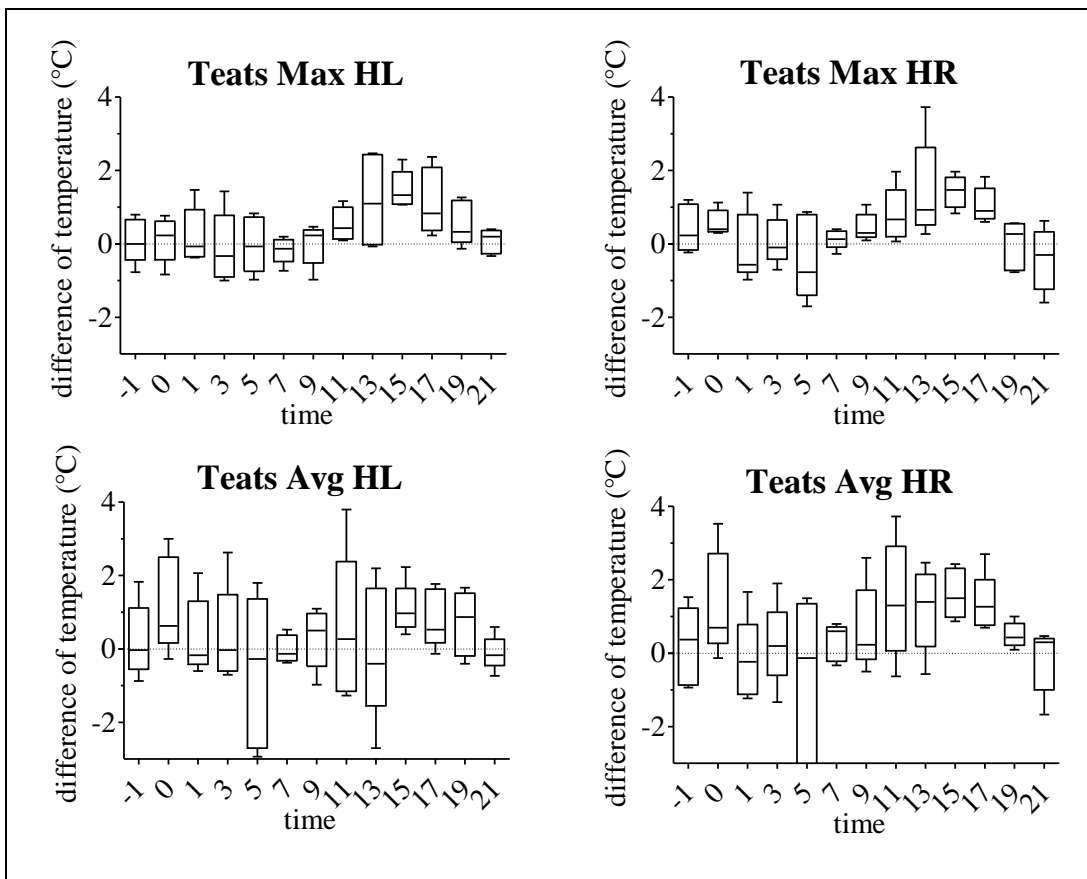


Figure 21: Box-plot diagrams of temperature differences in evaluation using the teats. Lower and upper whisker display minimum and maximum values. Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hind teat

'HR' – Region of interest, right hind teat

Outliers of 'Teats Avg HR' at '5': Minimum value=-4.10°C; first quartile=-3.37°C

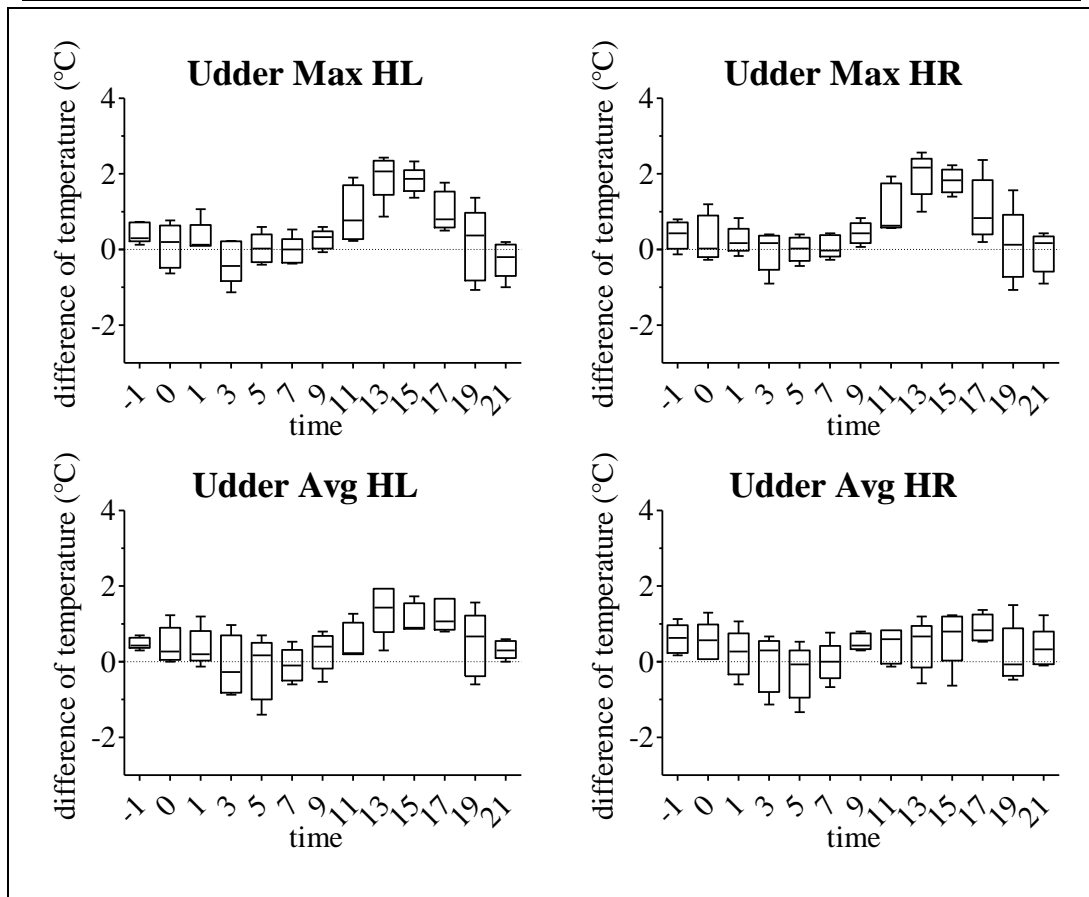


Figure 22: Box-plot diagrams of temperature differences in evaluation using the udder. Lower and upper whisker display minimum and maximum values. Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

As explained above, '-1' describes the differences between temperature values in the evaluation parameters one hour before challenge and 24 hours earlier. At that time, differences range around zero in evaluation using the teats. In evaluation using the udder, temperature differences are slightly larger at that time and range above zero. Temperature differences of the time of challenge versus 24 hours before ('0') are on a similar level slightly above zero in both methods.

The median values of temperature differences from '1' to '7' range around the baseline of zero in both, evaluation using and evaluation using the udder. Solely the parameter 'Teats Max HR' shows a median difference of -0.77°C at '5'. A large negative difference cannot be observed in 'Udder Max HR' at that time.

Regarding the box-plot diagrams in Figure 14 and Figure 15, it is shown that despite consistent median values in both methods, temperature differences in evaluation

using the teats show a distinctly wider range than in evaluation using the udder. Moreover, the box-plot diagrams of differences in evaluation using the teats indicate that the differences of the evaluation parameters 'Max HL' and 'Max HR' vary to a lesser extent than those of 'Avg HL' and 'Avg HR'. In evaluation using the udder, variation of the parameters' differences is not that pronounced.

At '9', median differences slightly increase in both methods when compared to '7', but still show a similar level to differences detected at '-1'. Solely 'Teats Avg HR' shows a decreasing median difference from '7' to '9'.

In the following, median temperature differences continue to increase. At '11', 'Teats Avg HL' is the only evaluation parameter to show decreasing differences compared to '9', whereas 'Teats Avg HR' now shows the most pronounced elevation in temperature difference. In evaluation using the udder, all evaluation parameters but 'Udder Avg HL' show an increasing temperature difference compared to '9'.

In both methods, median temperature differences of the evaluation parameter 'Max' are distinctly largest at '13'. Comparing both methods, it is obvious that median differences of evaluation parameters in evaluation using the teats are distinctly smaller than in evaluation using the udder. Comparing the challenged and unaffected hindquarter, no distinct differences are found. At '17', median differences are still elevated but decreasing again.

Concerning the evaluation parameter 'Avg', disparities between the methods are observed: At '13', median difference of 'Teats Avg HL' is -0.40°C and median difference of 'Teats Avg HR' is 1.40°C ; at '15', median difference of 'Teats Avg HL' is 0.97°C and median difference of 'Teats Avg HR' is 1.50°C . Elevation in differences of average surface temperature is thus earlier and more pronounced in the challenged teat than in the placebo-treated teat.

In contrast to these findings, the average udder surface temperature of the challenged quarter shows a lower difference between period after challenge and reference period than the placebo-treated quarter in evaluation using the udder: median difference of 'Udder Avg HL' is higher than median difference of 'Udder Avg HR' at '13' and '15', in evaluation using the teats it is vice versa.

Similar to median differences of the parameter ‘Max’, median differences of ‘Avg’ are still elevated at ‘17’ but decreasing again in evaluation using the teats, whereas in evaluation using the udder, median difference of ‘Udder Avg HL’ shows a slight increase.

After the increase in differences of average surface temperature at ‘13’, ‘15’ and ‘17’, median differences of most parameters decrease again and return to the level of the baseline at ‘21’ in both methods. Solely differences of ‘Teats Avg HL’ show another increase at ‘19’.

It can be expected that significant temperature differences between period after challenge and reference period can be found for the parameter ‘Max’ at ‘13’ and ‘15’ in both methods. Regarding the parameter ‘Avg’, discrepancies exist between the methods. Differences are most distinct in ‘Teats Avg HR’ from ‘11’ to ‘17’ and in ‘Udder Avg HL’ at ‘13’ and ‘15’. However, it is questionable if these differences are more distinct than the differences in ‘Max’ in both methods. Since temperature differences of ‘Max’ in evaluation using the teats mostly are larger than those of evaluation using the udder, it is also of interest if it provides more significant differences. One-way analysis of variances is used to clarify if the differences occurring between period after challenge and reference period are significant, and if they are, it is of interest at what time they are significant. In addition, it is clarified whether one evaluation method is able to detect more or earlier significant temperature differences than the other evaluation method.

Table 23 shows the results of one-way ANOVA and Dunett’s multiple comparison post-test in evaluation using the teats, Table 24 shows the according results in evaluation using the udder.

Table 23: Results of One-way ANOVA and Dunett's post-test comparing the temperature differences of reference period and period after challenge in evaluation **using the teats**

Evaluation using the teats				
One-way ANOVA:	Max HL	Max HR	Avg HL	Avg HR
P value	0.0029	0.0009	0.64	0.046
significant different group means	Yes	Yes	No	Yes
Number of groups	13	13	13	13
Dunett's Multiple Comparison Test:	Max HL	Max HR	Avg HL	Avg HR
mean difference of '-1' and				
0	-0.03	-0.17	-0.97	-1.11
1	-0.13	0.52	-0.10	0.40
3	0.21	0.34	-0.13	-0.03
5	0.11	0.81	0.81	1.20
7	0.26	0.28	0.22	-0.10
9	0.10	-0.04	-0.08	-0.45
11	-0.45	-0.39	-0.33	-1.23
13	-1.09	-1.03	0.26	-0.99
15	-1.39 (*)	-1.01	-0.87	-1.40
17	-1.05	-0.65	-0.61	-1.14
19	-0.47	0.43	-0.49	-0.28
21	0.01	0.84	0.33	0.40

Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hind teat

'HR' – Region of interest, right hind teat

Significant results ($P \text{ value} \leq 0.05$) are shown in bold. (*= $P \text{ value } 0.01 \text{ to } 0.05$, **= $P \text{ value } 0.001 \text{ to } 0.01$, ***= $P \text{ value } 0.001 \text{ to } 0.0001$, ****= $P \text{ value } < 0.0001$).

In the one-way analysis of variances of temperature differences in evaluation using the teats, significant differences are detected among the groups in the evaluation parameters 'Teats Max HL', 'Teats Max HR' and 'Teats Avg HR'. Among the temperature differences of 'Teats Avg HL', one-way ANOVA detects no significant differences.

In Dunett's multiple comparison test, the temperature differences of the different parameters in all groups ('0' to '21') are compared with the temperature differences of the evaluation parameters at '-1'. The table shows the mean difference of the differences in the two compared groups (difference at '-1' – difference at 'x'). Large

negative values thus indicate large differences. Significant mean differences are shown in bold.

The parameter 'Teats Max HL' is the only evaluation parameter in evaluation using the teats to show a significant mean difference between the difference at '-1' and the difference at '15'. Mean differences of the differences of '-1' and '13' and of '-1' and '17' are also prominent, yet they are not significant.

All other parameters in evaluation using the teats do not show significant differences between the differences of '-1' and the other temperature differences, although one-way ANOVA calculates very significant differences among the groups of differences of 'Teats Max HR' and 'Teats Avg HR'. Largest negative differences of temperature differences in 'Teats Max HR' are found between '-1' and '13' and between '-1' and '15'. The parameter 'Teats Avg HL' shows small mean differences between '-1' and '13' and between '-1' and '15'. Temperature differences of 'Teats Avg HR' show largest negative values of mean differences between '-1' and '11', between '-1' and '13', between '-1' and '15' and between '-1' and '17'.

One-way analysis of variances and Dunett's multiple comparison post-test of temperature differences in evaluation method using the udder (see Table 24) provide results different from those of evaluation using the teats. One-way ANOVA also detects significant differences among the temperature differences of 'Udder Max HL' and 'Udder Max HR'. However, significant mean differences are found among the differences of 'Udder Avg HL' but not among the differences of 'Udder Avg HR'.

Table 24: Results of One-way ANOVA and Dunett's post-test comparing the temperature differences of reference period and period after challenge in evaluation **using the udder**

Evaluation using the udder				
One-way ANOVA:	Max HL	Max HR	Avg HL	Avg HR
P value	< 0.0001	< 0.0001	0.0003	0.1422
significant different group means	Yes	Yes	Yes	No
Number of groups	13	13	13	13
Dunett's Multiple Comparison Test:	Max HL	Max HR	Avg HL	Avg HR
mean difference of '-1' and				
0	0.33	0.09	0.05	0.07
1	0.11	0.14	0.11	0.39
3	0.76	0.41	0.59	0.65
5	0.40	0.37	0.65	0.88
7	0.46	0.31	0.58	0.61
9	0.16	-0.05	0.20	0.09
11	-0.51	-0.68	-0.05	0.17
13	-1.50 (***)	-1.60 (***)	-0.89	0.15
15	-1.40 (**)	-1.44 (**)	-0.66	-0.05
17	-0.58	-0.68	-0.74	-0.29
19	0.30	0.27	0.02	0.41
21	0.70	0.44	0.17	0.25

Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Significant results ($P \text{ value} \leq 0.05$) are shown in bold. (*= $P \text{ value} 0.01 \text{ to } 0.05$, **= $P \text{ value } 0.001 \text{ to } 0.01$, ***= $P \text{ value } 0.001 \text{ to } 0.0001$, ****= $P \text{ value } < 0.0001$).

Regarding the results of Dunett's multiple comparison post-test comparing the temperature differences of '-1' with all other temperature differences, significant differences can be found for 'Udder Max HL' and 'Udder Max HR' between '-1' and '13' and between '-1' and '15'. Between the temperature differences of '-1' and '17' no more significant differences can be found, and differences decrease again.

Comparing the differences of average surface temperature in evaluation using the udder, no significant differences can be found between the differences of '-1' and all other differences, although one-way ANOVA indicated significant differences

among the groups for at least ‘Udder Avg HL’. In ‘Udder Avg HL’, mean differences are largest for the differences between ‘-1’ and ‘13’, between ‘-1’ and ‘15’ and between ‘-1’ and ‘17’. In the challenged quarter, mean differences of ‘Udder Avg’ are smaller between the temperature differences of ‘-1’ and ‘13’, between ‘-1’ and ‘15’ and between ‘-1’ and ‘17’ than in the placebo-treated quarter.

Comparing the results of one-way ANOVA, the temperature differences of the evaluation parameters ‘Max HL’ and ‘Max HR’ are detected as significantly different in both, evaluation method using the teats and using the udder. However, in Dunnett’s multiple comparison test significant differences can be found between the differences of ‘-1’ and ‘15’ in ‘Teats Max HL’, whereas significant mean differences can be found for both, ‘Udder Max HL’ and ‘Udder Max HR’ between the differences of ‘-1’ and ‘13’ as well as between the differences of ‘-1’ and ‘15’.

Regarding the analysis of temperature differences of average surface temperature, the results also vary: whereas in evaluation method using only the teats, differences of ‘Avg HR’ are larger and more pronounced than differences of ‘Avg HL’ between ‘11’ and ‘17’; in evaluation using the udder ‘Avg HL’ shows more distinct temperature differences. However, no significant differences can be found in Dunnett’s multiple comparison test.

3.3.3. Determining threshold values for evaluation using the teats and evaluation using the udder

Udder surface, respectively teat surface temperature values detected when cows showed rectal temperatures below 39.5°C versus values detected when cows had rectal temperatures above 39.5°C, are displayed in box-plot graphs in Figure 31 in Appendix 6, as well as the associated ROC curves (Figure 32 in Appendix 6).

The AUCs of the evaluation parameters ‘Avg’ and ‘Max’ in the left and right hindquarter, respectively in the left and right teat, are displayed in Table 25.

Table 25: Results of ROC (Receiver-Operating-Characteristics) analysis: Area underneath curve (AUC) and P-Value of evaluation parameters in evaluation using the teats ('Teats') and evaluation using the udder ('Udder'), discriminating between healthy udder and udder suffering from clinical mastitis

ROC Analysis		
	AUC	P-Value
Teats		
Teats Max HL	0.84	< 0.0001
Teats Max HR	0.87	< 0.0001
Teats Avg HL	0.71	0.0050
Teats Avg HR	0.86	< 0.0001
Udder		
Udder Max HL	0.93	< 0.0001
Udder Max HR	0.94	< 0.0001
Udder Avg HL	0.89	< 0.0001
Udder Avg HR	0.76	0.0004

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter, respectively left hind teat

'HR' – Region of interest, right hindquarter, respectively right hind teat

Since all P-values are small, results of ROC analysis of all evaluation parameters are considered as significant. In evaluation using the teats, 'Max HR' and 'Avg HR' obtain almost equally large AUCs, whereas in evaluation using the udder, the AUC of 'Avg HR' is distinctly smaller than the AUC of 'Avg HL'.

The AUCs of 'Max HL' are in both methods larger than those of 'Avg HL'. It is thus noticeable that evaluation parameters concerning the right teat show better results in ROC analysis than those concerning the left teat. However, results of ROC analysis of evaluation using the teats are inferior to those of evaluation using the udder.

Table 26 shows the threshold values of the evaluation parameters in evaluation using the teats and evaluation using the udder gaining the maximum sum of sensitivity and specificity (Youden's Index).

Table 26: Results of ROC (Receiver-Operating-Characteristics) analysis: Threshold value, sensitivity, specificity and sum of sensitivity and specificity of evaluation parameters in evaluation using the teats (Teats') and evaluation using the udder ('Udder')

Sensitivity and Specificity				
	Threshold (°C)	Sensitivity (%)	Specificity (%)	Sum
Teats				
Teats Max HL	> 36.68	68.75	94.24	162.99
Teats Max HR	> 36.52	68.75	94.24	162.99
Teats Avg HL	> 34.92	62.50	79.86	142.36
Teats Avg HR	> 35.05	87.50	76.26	163.76
Udder				
Udder Max HL	> 38.37	94.12	92.31	186.43
Udder Max HR	> 38.57	94.12	95.10	189.22
Udder Avg HL	> 36.45	88.24	83.92	172.16
Udder Avg HR	> 36.85	58.82	96.50	155.32

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter, respectively left hind teat

'HR' – Region of interest, right hindquarter, respectively right hind teat

In evaluation using the udder, 'Max HR' gains slightly better results of sensitivity and specificity than 'Max HL', whereas in evaluation using the teats, 'Max HL' and 'Max HR' gain the same sensitivity and specificity at different threshold values

In evaluation using the teats, the sensitivity of 'Avg HR' is larger than the sensitivity of 'Avg HL', whereas the specificity of 'Avg HL' exceeds the specificity of 'Avg HR'.

Concerning the results of 'Avg' in evaluation using the udder, the findings are the exact opposite: sensitivity of 'Avg HR' is distinctly lower than sensitivity of 'Avg HL', whereas better specificity is gained by 'Avg HR'. However, the sum of sensitivity and specificity of 'Avg HL' exceeds the sum of 'Avg HR' by far.

'Avg HR' is the only evaluation parameter to show a larger sum of sensitivity and specificity in evaluation using the teats than in evaluation using the udder. In all other evaluation parameters, the results of evaluation using the udder are superior to those of evaluation using the teats.

VI. DISCUSSION

The comparison of automatic and manual evaluation of thermograms of the udder serves the aim of this study to analyze if both methods gain comparable results, bearing the objective in mind to investigate whether automatic evaluation is a suitable tool in automated udder health monitoring in large dairy herds.

As outlined in the introduction, further objectives of the study are if evaluation of the region of the teats, as well as exclusion of typical ‘hot spots’, enable an earlier detection of clinical mastitis.

1. Impact of measurement conditions

The importance of consistent measurement conditions for reliable results are described by GLAS (2008), METZNER et al. (2014) and OKADA et al. (2013) (see: ‘III.3.2 Optimal conditions for IRT application on cattle’). Thermographic material used in this study was recorded with constant distance of 1.8 m between camera and udder as recommended by OKADA et al. (2013). Measurements took place indoors where the animals were housed and were thus unaffected from direct solar radiation. The camera was not exposed to extremely hot or cold temperature effects, which could have influenced the measurement results (OKADA et al., 2013). As suggested by METZNER et al. (2014) and BERRY et al. (2003), intervals between the measurements were kept short in order to not miss temperature peaks.

Values of environmental temperature and air humidity were recorded throughout the trial. Their influences on the USST are investigated in detail in the study of GLAS (2008) and are thus not part of the correlation analysis in this study. GLAS (2008) finds the air humidity to have no influence on USST, whereas environmental temperature has a small but detectable influence on USST. It is, however, not as pronounced as the influence of rectal temperature.

Since cows were kept in an enclosed room, environmental temperature only varied to a small extent. Nevertheless, alterations in environmental temperature are a factor that has to be taken into consideration in future studies, especially when IRT is applied in open stables.

FRANZE et al. (2012) report a correlation coefficient of $r=0.6$ between USST and environmental temperature in their study in the stable area, but do not find any

correlation between air humidity and USST, respectively between air pressure and USST.

The preparation of the udder for the thermographic material used in this study was extensive: depending on necessity hair was clipped and the udder was dry-cleaned. In some cases of gross contamination of the udder, wet-cleaning was necessary and measurements consequently had to be postponed for several minutes. Bearing the idea in mind to use IRT as an automatic health supervision tool, these complex preparations are unrealistic to implement in large livestock farming.

For further studies concerning the use of IRT in automatic udder health monitoring, it is of particular interest to evaluate the resistance of IRT measurements towards falsifications by contaminations in the ROI's. Results of precision analysis in this study indicate that evaluation parameter 'Max-Min' has the poorest precision in all methods, followed by evaluation parameter 'Min'. This is most probably due to contaminations, since 'Min' represents the pixel with the lowest temperature in the ROI. Evaluation parameters 'Avg' and 'Max' both show good precision: 'Avg' takes the temperatures of all pixels in the ROI into account, thereby reducing the impact of outliers. 'Max' stands for the pixel with the highest temperature in the ROI, which most likely originates from the udder surface if no external heat sources are nearby. It is thus conceivable that evaluation parameters 'Avg' and 'Max' are robust to a certain level of contamination and usable for the evaluation of thermograms taken without extensive preparations. However, this assumption has to be validated by further research. GLAS (2008) also concludes that minimum values of USST are prone to falsifications and should not be used for evaluation.

Cows were tethered with a collar which reduced their mobility, thereby measurements were simplified and could consistently be recorded from the same angle with the same distance. The automatic image recognition software was able to correctly detect the hindquarters in all thermograms. Installation of an IRT-camera in stables where cows can move freely can involve further challenges: The Active Shape Approach algorithm would not only have to be trained for naturally occurring variations in the shape of the udder, but also in recognizing the udder from different angles and from different distances.

The authors of a large field study in the stable area (FRANZE et al., 2012) state that IRT is applicable for automatic udder health supervision. IRT-cameras were

installed behind the milking carousel and, inter alia, evaluated automatically. However, no rates of correct detection are named. Fixing cows in a milking carousel has the advantage of relatively consistent measurement conditions. Nevertheless, milking is only done a limited number of times throughout the day and intervals between the measurements could become too long to detect temperature peaks of USST (cf. METZNER et al. (2014)).

2. Impact of physiological oscillation in body temperature on IRT measurement results

It is not definitively clarified if circadian rhythms of body temperature in cattle exist. BITMAN et al. (1984); BERRY et al. (2003) and GLAS (2008) find repetitive circadian patterns in rectal temperature in cows. Although it is not proven in their studies, it is most probable that these patterns are due to influence of environmental temperature, since rectal temperature is in all cases described to be lowest in the early morning and to increase throughout the day, peaking in the late afternoon or early evening. GLAS (2008) describes rectal temperature to range for $0.4^{\circ}\text{C} \pm 0.1^{\circ}\text{C}$ in healthy cows throughout the day, whereas BITMAN et al. (1984) find oscillations of rectal temperature of 1.25°C . Considering the good correlation of rectal temperature and USST, these variations in healthy individuals are noticeable. In this study, intramammary challenge of all cows consistently took place at the same daytime, thereby eliminating influences of circadian patterns in body temperature. However, these influences cannot be excluded when IRT is applied in automated health monitoring in the field.

3. Limitations of the different evaluation methods and their ROI's

The software ThermaCAM Researcher Pro 2.8 (FLIR[®] Systems, Wilsonville, Oregon, United States) is used for the manual evaluation ('Man') of hind udder **including** teat surface that serves as comparison with automatic evaluation, but also for the evaluation of **exclusively** the teat surface ('Teats') and the udder surface **excluding** the teats ('Udder'), which are both manual methods as well. Due to the lack of alternatives, manual evaluation of thermograms is the current gold-standard in veterinary medicine. However, this software is specialized for the application in engineering research and construction. Selecting polygons as ROIs with this software was proved to be superior to the selection of lines or rectangles in the

udder in the study of METZNER et al. (2014). It provides reliable results for the examiner, but evaluating a large number of thermograms with the polygon-tool is extremely time-consuming since the ROIs have to be repeatedly selected in each thermogram and results have to be manually transferred into a table. In order to select the same region in each thermogram consistently, strict orientation towards the defined borders is necessary, and standardized exclusion of ‘hot spots’, as suggested by METZNER et al. (2014), is thus not possible when the udder is selected.

Selecting only the region of the teats automatically excludes the ‘hot spots’ of udder-thigh-gap and intermammary sulcus. Nevertheless, it has to be considered that the teats have a large surface in relation to their volume and they are in a position exposed to environmental influences (e.g. extreme temperatures and air movement). It can be assumed that they are more vulnerable to temperature falsifications by external factors than the udder.

In the comparison of automatic and manual evaluation, the hind surface of the udder **including** the teats was chosen as ROI, since the teats in contrast to environment serve as important landmarks in the point-distribution-model. Automatic segmentation was inadequate in 2% of the thermograms, when hind teats or distal udder could not be differentiated from fore teats or forelegs), respectively in approximately 3% of the thermograms when the proximal border of the udder was not correctly detected (SCHRÖTER, 2015). For a valid comparison of automatic and manual evaluation method, the data of inadequate segmented thermograms must not be excluded from further analysis.

It is to assume that inadequate segmentation of the lower udder and teats has more impact on the results of evaluation than inadequate segmentation of the proximal udder: when teats were falsely segmented, ROIs included cooler regions (e.g. parts of the forelegs). When the adhesive patch was not recognized, the proximal border of the udder was falsely set lower, consequently, the ROI still was located on the udder surface.

However, finding rates of the Active-Shape Model in this study have to be assessed as very good. WIRTHGEN et al. (2011b) evaluated over two million thermograms and describe correct finding rates of 80% with the Active-Shape Model. However, substantial differences between these studies have to be considered: WIRTHGEN

et al. (2011b) recorded the complete rear view of cows in a milking carousel, resulting in a more complex point-distribution model. Unlike in this study, the tail was not held away and covered parts of the udder. Moreover, seven ROIs were evaluated in each thermogram, including small regions like the claws. It is not clearly stated, but a thermogram was probably assessed as falsely segmented when one of seven ROIs was not correctly found. In summary, it can be said that the study of WIRTHGEN et al. (2011b) applies more to field conditions than this study does.

Unlike in manual evaluation, automatic evaluation was able to exclude 5% of the pixels originating from the outer borders of the ROIs, thereby reducing the impact of the ‘hot spots’.

The right hindquarter was challenged with *E. coli*, whereas the left hindquarter was treated with a placebo, thus it is evident to evaluate the measurements in both hindquarters separately.

4. Interpretation of results of automatic and manual evaluation method

It is to presume that automatic evaluation has time-saving advantages over manual evaluation. Averaged from the evaluation time of five thermograms, manual evaluation of a single thermogram with two ROI's by a practiced examiner and the following transfer of the data into a table took 117 seconds. However, the exact automatic evaluation time was not recorded. It was reported that once the automatic image recognition program was trained, evaluation of the whole set of thermograms took a few seconds. Since this is not a reliable measurement, no exact statement about the saving of time can be done.

4.1. Precision analysis

Coefficients of variation (VCs) of the evaluation parameters in thermogram-triplets before challenge are used to determine their precision. They allow to draw a conclusion to what extent values of the different evaluation parameters differ in three consecutive recordings at the same time-point.

Regarding the median values of the VCs in Table 41 in Appendix 4, it becomes obvious that those of ‘Max-Min’ are noticeably larger than those of the other parameters, and even between twenty and sixty times higher than those of ‘Avg’. This observation is not too surprising: ‘Max-Min’ depicts the temperature range

between the warmest pixel and the coolest pixel in each ROI. Consequently, smallest changes in the borders of the ROI, or dirt particles that are temporarily inside the ROI can lead to an altered ‘Max-Min’ value. The fact that actually surprises is that median VCs of ‘Max-Min’ in automatic evaluation are larger than those of manual evaluation. It is unknown if variations of the coolest or the warmest pixel are responsible for the large VC’s of ‘Max-Min’, but since automatic evaluation excludes hot spots more than manual evaluation, less variation concerning the warmest pixels is expected. It is thus conceivable that variations derive from cool pixels belonging to the environment, falsely selected into the ROI, and that automatic evaluation selects the outer borders less precise than manual evaluation. However, VCs of ‘Max-Min’ exceeding those of the other parameters by far indicates insufficient precision of this evaluation parameter in both methods.

The box-plot graphs of Figure 8 give a first visual impression of the VC’s distribution: The box-plots of VCs of ‘Min’ appear noticeably larger than those of the other parameters. METZNER et al. (2014) suggest that minimum temperature is prone to falsifications since it, as mentioned above, depicts the coolest pixel inside the ROI, which can be a dirt particle or belong to the structures neighboring the ROI. Comparing automatic and manual evaluation, VCs of ‘Min’ are distinctly larger in automatic evaluation. This is probably due the automatic detection of the ROIs: even if the outer borders are detected correctly, it is possible that either parts of the adhesive patch or the neighboring structures of the body or surroundings are included. One pixel is enough to alter the outcome.

Regarding the box-plots, ‘Avg’ seems to have the best precision, followed by ‘Max’. The appearance of the box-plots does not allow to distinguish which method is more precise. It was expectable that ‘Avg’ has the best precision results since it is a mean value and thus less prone to outliers. The variations of the parameter ‘Max’ largely depend on the selection of the ROI: warmest pixels are in most cases located in the ‘hot spots’ (udder-thigh-cleft and intermammary sulcus) which define the outer borders of the ROI. In Table 41, medians of VCs of ‘Max’ are larger in automatic evaluation than in manual evaluation. Inclusion of the ‘hot spots’ in manual evaluation may thus lead to more consistent findings of warmest pixels in these areas.

One-way analysis of variances and following Tukey’s post-hoc test (Table 3 and Table 4) have similar results for both method: significant differences are only found

between the VCs of ‘Max-Min’ and the VCs of all other parameters, although it was expected that ‘Min’ is also significantly less precise. However, ‘Avg’ and ‘Max’ consistently have the smallest VCs, so these two parameters are chosen to compare the precision of automatic and manual evaluation directly:

The results of unmatched t-test comparing the precision of ‘Max’ and ‘Avg’ in both methods (see Table 5) lead to the conclusion that evaluating the thermograms manually is slightly, but significantly more precise. This result is not too surprising, taking into account that each ROI was carefully selected in manual evaluation, whereas in automatic evaluation false findings occurred to a low percentage. Nevertheless, the mean differences between the parameters VCs are small, ranging from 0.0007 to 0.0028. It can thus be assumed that automatic evaluation also provides valid results, even if they are slightly less precise than those of manual evaluation.

The evaluation parameter ‘Avg’ yields in both methods and both hindquarters the lowest VCs, but ‘Max’ is almost as precise. It should be considered that measurements took place under optimized conditions in a trial. Although ‘Avg’ is a robust evaluation parameter since it is a mean value, it may be vulnerable to gross contamination or wetness of the udder when IRT is applied in the stable area. ‘Max’ depicting the warmest pixel in the ROI may be more consistent in these cases. ‘Max’ should thus not be rejected due to slightly lower precision.

4.2. Results of automatic and manual evaluation and their correlation with rectal temperature

One of the main questions in this study is if exclusion of the ‘hot spots’ leads to better detection of local inflammation processes in the udder. Assuming that surface temperature in the ‘hot spots’ is closely related to rectal temperature (METZNER et al., 2014), results of manual evaluation would be more correlated with rectal temperature than results of automatic evaluation.

Figure 9 and Figure 10 show the course of median values in the left and right hindquarter in automatic and manual evaluation. The course of ‘Min’ in both methods confirms what was suggested in the discussion of Precision analysis: median values proceed completely irregularly, the graphs are multiply peaked and the peak around 15 hours after challenge is not apparent, unlike in the medians of the other parameters. The pattern of the course of rectal temperature is not

recognizable in the courses of 'Min'. Thus, poor correlation of 'Min' and rectal temperature is expectable. Due to the poor precision and the irregular course of 'Min', this parameter is excluded from further statistical analysis.

Comparing the course of 'Max' and rectal temperature, the graphs appear similar in automatic and manual evaluation: median temperatures proceed at a consistent level, they are increased between 11 and 17 hours after challenge, peaking 13 and 15 hours after challenge. 'Aut Max' seems to be related with rectal temperature to the same extent that 'Man Max' is: median temperatures proceed in the same pattern, but at a distinctly lower level. This is probably due to the exclusion of 'hot spots' in automatic evaluation, maximum temperatures are thus lower than in manual evaluation. However, the fact that the pattern of 'Max' is still related to the pattern of rectal temperature, contradicts the assumption that exclusion of 'hot spots' leads to a better depiction of local inflammatory processes.

Rectal temperature and USST peaking 13 and 15 hours after challenge is consistent with the findings of METZNER et al. (2015), but when HOVINEN et al. (2008) performed intramammary challenge with *E. coli* lipopolysaccharide, they found temperatures to be peaking much earlier, between 4 and 8 hours after challenge.

It is remarkable that maximum surface temperature has a similar course in the left, placebo-treated hindquarter as in the right, challenged hindquarter: temperature peaks are just as distinct and occur at the same time. This applies to both methods. This observation leads to positive and negative conclusions: On the one hand, it suggests that screening the surface of the hindquarters in automatic health supervision would probably be sufficient to detect clinical mastitis in all quarters, even in the forequarters. Additional recordings from the side would not be required. More generally spoken, systemic diseases accompanied by fever could possibly also be detected by screening the surface of the hindquarters.

On the other hand, the fact that the neighboring quarter shows the same distinct USST elevation as the challenged quarter leads to the conclusion that a simple comparison of left and right hindquarter is not sufficient for mastitis detection. Much rather, the temperature course of each cow must be observed in short intervals, as suggested in numerous studies (SCOTT et al., 2000; GLAS, 2008; HOVINEN et al., 2008; METZNER et al., 2015). METZNER et al. (2015) calculated differences of maximum and average surface temperature between left

and right hindquarter and tested them for statistical significance: the differences of maximum surface temperature were not significant at any time. The differences of average surface temperature were significant at two time points before challenge and two time points after challenge. However, these significant differences could not be approved in multivariate analysis. SCOTT et al. (2000) also detected significant temperature elevation in both hindquarters when the left hindquarter was challenged with *E. coli* endotoxin, whereas FRANZE et al. (2012) found different temperatures in the challenged quarter ($P < 0.1$). However, the authors were able to improve the mastitis detection rate when they compared two consecutive measurements in one cow, instead of comparing challenged and unaffected quarter. SCHUTZ et al. (2001) report quarters challenged with *S. aureus* to be significantly warmer than the other quarters.

Another observation surprises when the courses of 'Avg' in the two hindquarters are compared: in the right (challenged) hindquarter, the amplitude of the temperature peak appears noticeably lower than in the left (placebo-treated) hindquarter. Consequently, average surface temperature in the challenged quarter rose less than in the placebo-treated quarter. This is quite contrary to what would be expected. In the study of BARTH (2000), a naturally occurring mastitis also led to the affected quarter being slightly cooler than the others. The causal pathogen was not stated. It is assumed that this observation was due to edema accompanying the clinical mastitis. Edema of the udder can occur due to non-inflammatory causes, for instance in the periparturient period, or due to inflammatory causes, especially alongside with clinical mastitis (GRUNERT, 1996). Endotoxins influence the vascular permeability and fluid accumulates in the interstitial tissue. However, ultrasonographic investigations revealed that the structures mostly affected are the distal subcutis and the suspensory apparatus of the udder (STOCKER & RÜSCH, 1997) Consequently, it is of interest if this effect is also detectable in evaluation using only the teats, where subcutaneous edema does not occur.

For automated udder health supervision, it is important to bear in mind that simple comparison of udder quarters does not provide reliable results, since the affected quarter may not be warmer or even cooler than the others. Much rather it should be recognized if USST exceeds a certain level and the according cow should undergo clinical examination. However, the present data are only applicable for mastitis due

to *E. coli*. USST in mastitis caused by other pathogens may have a completely different course.

The results of Pearson correlation analysis prove what has already been assumed: In both methods and both hindquarters, maximum surface temperature and rectal temperature are significantly positively correlated (see Table 6 and Table 7). The good correlation can be either due to a rise in rectal temperature influencing the udder surface temperature, or due to other variables (for instance local and systemic reaction towards the challenge) influencing both, rectal and udder surface temperature. It is important to bear in mind that all cows in this trial showed signs of clinical mastitis and had fever after challenge. These results do not allow conclusions about the course of USST and its correlation with rectal temperature in cases of subclinical mastitis.

‘Avg’ and rectal temperature are in both methods indeed less correlated in the challenged quarter than in the placebo-treated quarter, and although correlation coefficients are significant, they are only moderately large. This was expectable, since ‘Avg’ takes all pixels in the ROI into account, and the influence of ‘hot spots’ is thus relativized. This is not automatically a disadvantage, since ‘Avg’ may be able to depict the local udder inflammation better than ‘Max’.

Automatic evaluation excludes 5% of the outer pixels in order to minimize the impact of ‘hot spots’, the results of automatic evaluation should thus be less related to rectal temperature than those of manual evaluation. The results of correlation analysis of rectal temperature and parameters of manual evaluation (Table 7) show that consistently higher correlation coefficients can be found here: All values exceed those of automatic evaluation. This does not consequently mean that automatic evaluation is inferior towards manual evaluation. Much rather, it indicates that automatic evaluation actually depicts less of inner body temperature than manual evaluation by the exclusion of ‘hot spots’, as described in the introduction.

The course of ‘Min’ and the course of rectal temperature have an expectably low correlation, and although the results of Pearson correlation analysis are at least significant for the values of minimum temperature in the right hindquarter and rectal temperature, correlation coefficients are too low to indicate a distinct relation.

For further understanding of the relation of rectal temperature and both evaluation methods, the standard deviations of the differences between each method's 'Max', respectively 'Avg' and rectal temperature were analyzed and compared. Small values of standard deviation indicate a consistent difference between rectal temperature and average udder surface temperature. In both evaluation methods and both hindquarters, standard deviations of the differences are smaller before challenge than after challenge. It is thus likely that temperature values differ to a larger extent after challenge and that local temperature changes in the udder surface have a higher impact after challenge. The standard deviation of the differences between manually determined values and rectal temperature are consistently lower than those of the differences between automatically determined values and rectal temperature. However, results of t-test to detect significant differences between the methods' SDs are not straightforward.

It can be concluded that by the exclusion of 'hot spots' in automatic evaluation, impact of inner body temperature on the measurements' outcome could be reduced, but they are still distinctly related.

4.3. Comparison of automatic and manual evaluation

The graphs in Figure 12 and Figure 13 once more emphasize that the temperature courses evaluated automatically proceed on a lower level than the manually evaluated courses. It is self-evident that excluding the 'hot spots' leads to ascertaining lower temperatures, however, it is surprising that the course itself remains similar. It would have been expectable, that by excluding the regions that show temperatures mostly related to inner body temperature, a course would have been ascertained that resembles less the course of rectal temperature. Nevertheless, the similar courses of 'Max' and 'Avg' suggest that both methods provide equally precise results and that both detect temperature elevations equally quick: The amplitude of the temperature peak 13 and 15 hours after challenge seems to be comparably high in both methods. The differences between temperature values in the period after challenge and the reference period are discussed in '4.3.2 Comparison of period after challenge and reference period before challenge in automatic and manual evaluation'.

Exclusion of the 'hot spots' is still recommendable, also for practical reasons: it reduces the possibility to include pixels belonging to environmental structures neighboring the ROIs' borders. However, for future automatic evaluation of

thermograms of the udder it is essential to bear in mind that if exclusion of the ‘hot spots’ is performed, it has consistently to be done to the same percentage, since it leads to temperature results on a distinctly different level. Otherwise, if exclusion is done inconsistently, false temperature drops or increases could be detected.

Medians of evaluation parameters at the reference period and period after challenge in automatic and manual evaluation (see Table 9) once more emphasize that calculating medians or mean values for long time-spans is not sufficient for mastitis detection: differences of the medians of the 24-hours periods before and after challenge range between 0.1°C and 0.35°C. These small temperature differences would hardly be recognized as pathological and can of course also occur by chance.

4.3.1. Correlation analysis of automatic and manual evaluation

Bearing the questioning of this study in mind, if an automated image recognition software can gain results comparable to those of the gold standard (manual evaluation), correlation analysis of the two methods is of particular importance.

The scatter plots in Figure 27 obviously depict a linear correlation between the parameters ‘Max’, respectively ‘Avg’ in automatic and manual evaluation, Pearson correlation analysis is thus a valid way to calculate the correlation coefficients. The rightwards shifting of the scatter plots is due to manual evaluation providing higher temperature values than automatic evaluation (see ‘4.3 Comparison of automatic and manual evaluation’).

All results of correlation analysis between evaluation parameters ‘Avg’ and ‘Max’ in automatic and manual evaluation indicate a very good correlation. This is a major advantage for automatic evaluation since it proves it to be comparably adequate as the gold standard of manual evaluation. The results of the evaluation parameter ‘Avg’ are slightly superior to the results of ‘Max’, gaining almost perfect correlation coefficients. This makes ‘Avg’ favorable for automatic evaluation, but as already mentioned, ‘Avg’ might be affected by gross contamination or covering in some cases when recording conditions are less standardized. ‘Max’, always depicting the warmest pixel and gaining promising results in correlation analysis as well, should also be considered. It is conceivable that ‘Max’ and ‘Avg’, being more reliable in combination, are both used in automated udder health supervision.

The results of WIRTHGEN et al. (2011b) support this assumption. In their study, IRT was applied to cows in the milking carousel, evaluating a total of seven ROIs

per cow automatically and manually, amongst them the surface of the hindquarters. Correlation coefficients of average surface temperature in automatic and manual evaluation is $r=0.66$, respectively $r=0.76$ for maximum surface temperature. These results are noticeably lower than in this study. Obviously, this study is performed under stable conditions which may influence the outcome of the correlation analysis. Unfortunately, no detailed description of the manual evaluation process was done, which may have had impact on the results of correlation analysis.

4.3.2. Comparison of period after challenge and reference period before challenge in automatic and manual evaluation

All differences are calculated for the value detected by an evaluation parameter at a certain time point and the value detected by the same evaluation parameter 24 hours earlier. It would have been possible to just calculate differences between the values after challenge and a median or mean value measured before challenge. However, calculating the differences between healthy and challenged udder consistently between measurements that are 24 hours apart has the advantage of neutralizing the possible influence of circadian temperature rhythms that were observed in numerous studies (BITMAN et al., 1984; BERRY et al., 2003; GLAS, 2008).

The presentation of the differences in box-plot graphs simplifies the analysis of their courses (see Figure 14 (automatic evaluation) and Figure 15 (manual evaluation)). Differences above zero imply that the maximum, respectively average surface temperature is higher at this number of hours after challenge (see 'time' on x-axis) than 24 hours before.

The appearance of the differences of 'Max' in both methods prove what was already expected: distinct differences before and after challenge are found in certain intervals but not throughout the whole time: differences increase at '11', peak at '13' and '15' and quickly decrease again, approaching the baseline of zero again at '21'. The appearance of the box-plot graphs is very similar in both hindquarters and both methods, solely the differences of 'Aut Max HL' appear more widespread than the others. In the tables of median differences, q1 and q3 (see Table 43 and Table 44), none of the methods seems to be distinctly superior to the other in detecting temperature differences. This observation suggests that the automatic method is able to detect pathological temperature elevations equally quick as the gold

standard. However, this has to be approved by one-way analysis of variances (see below).

Regarding the parameter ‘Avg’, temperature differences increase between ‘13’ and ‘17’, but much less pronounced as the differences of ‘Max’. Moreover, the course of the differences of average surface temperature is quite different in the left (placebo-treated) and in the right (challenged) hindquarter. As already observed in the temperature courses (see: ‘4.2 Results of automatic and manual evaluation and their correlation with rectal temperature’), this depicts the right hindquarter showing a lesser temperature increase which is most likely due to subcutaneous edema. It is unlikely that the challenged udder quarter itself is cooler than the other quarter, since clinical mastitis was diagnosed in all cows and increased temperature is a cardinal symptom of inflammation. Much more, it is conceivable that this edema functions as a buffer between increasingly warm udder tissue and skin surface.

It is thus not too surprising that one-way ANOVA was able to detect significant different groups means between the differences of ‘Max HL’, ‘Max HR’ and ‘Avg HL’, but not between the differences of ‘Avg HR’. This applies to both methods and underlines the importance of taking more than one hind udder quarter into account. Furthermore, it suggests that if the whole hind udder surface was conceived as a unit in automatic evaluation, changes in average surface temperature might be neutralized by the edema of the affected udder quarter. The separation of the hind udder surface into two ROIs, corresponding to left and right hindquarter, is valid and important.

Since one-way ANOVA only analyzes **if** significant differences between the groups’ means exist, but not between **which** groups, the results of Dunett’s multiple comparison test are essential. Dunett’s post-hoc tests compares all groups with one single group. The differences that are marked with ‘-1’ are the differences of the value one hour before challenge minus the value 25 hours before challenge. Consequently, average or maximum udder surface temperature could not be affected by the intramammary challenge at both time points since they are both in the reference period. Therefore, differences of ‘-1’ are used as ‘control difference’ in Dunett’s post-hoc test. Alternatively, the value zero could have been used as control, since differences varying around zero are also expected for all parameters at ‘-1’, but using existing differences from before challenge was regarded as more

reliable since it takes into account that temperature differences can also occur naturally or by chance within 24 hours, even if the udder is not manipulated.

In both methods, significant differences are detected for the temperatures in the evaluation parameters 'Max HL' and 'Max HR' between the differences of '-1' and '13', respectively '-1' and '15'. This means, that a significant increase in maximum temperature is detectable by both methods between 13 and 15 hours after challenge. Consequently, both methods detect temperature differences equally quick via evaluation parameter 'Max' in both hindquarters. In addition, no method detects significant differences for a longer period of time than the other method. However, a time-span of two hours of significant temperature elevation is relatively short and underlines once more that measurement intervals must not exceed two hours. Much more it can be stated that measurement intervals cannot be too short: for instance, it is not known how average and maximum temperature proceeded between the measurements in this trial. It is possible that a more distinct peak between 13 and 15 hours after challenge was missed, or that temperature elevation actually started 9.5 hours after challenge, but the next recording took place 11 hours after challenge.

This raises the question where the IRT-camera for automatic health supervision should be positioned in the stable area. A position where the animals can be recorded multiple times throughout the day- and nighttime is recommendable, for instance places that are highly frequented in the free-range stable. However, this suggestion has disadvantages: firstly, it is expectable that diseased cows move less and would thus be less often recorded by the cameras. Nevertheless, it can be countered that in the first place, automated health supervision should be a tool to detect pathological alterations before clinical symptoms appear. A cow that shows increased time-spans of lying down or reduced mobility should be noticed by a trained supervising person. Furthermore, promising results have been achieved in trials using accelerometers to analyze mobility and behavior patterns in cattle (ROBERT et al., 2009; VÁZQUEZ DIOSDADO et al., 2015). A second limitation of this suggestion is more concerning: highly frequented places in the stable area are usually the most attractive places, for instance feeding stations or cow brush arrangements. Lower-ranking cows are expected to have less access to these areas and could evade the automated supervision. Consequently, field studies analyzing different positioning of the IRT cameras and involving a larger number of cows are of interest.

However, installing the camera in the milking parlor like in the studies of FRANZE et al. (2012) and WIRTHGEN et al. (2011b) is probably not sufficient for mastitis detection, since intervals between the measurements are up to twelve hours and temperature peaks of two or three hours are most likely to be missed.

4.3.3. Limitations and results of ROC-analysis of automatic and manual evaluation

One of the main objectives in this study is the comparison of two methods and its ability to detect clinical mastitis. ROC-analysis is a valuable way to analyze diagnostic methods. However, the raised data have to be split into two groups: a group of data raised when a scientifically proven method detected pathological findings, and a control group. In this study, this was not straightforward. SCC detected cell concentration exceeding 400.000 cells/ml by far in milk samples of the challenged quarters of all cows 11.5 hours after challenge. A cell concentration exceeding this threshold is classified as pathological and alongside with clinical symptoms (in this case: fever) it can be considered as clinical mastitis. SCC was not implemented at each time point of measurement, but with an interval of twelve hours. Although it can be assumed that cell concentration was not only increased 11.5 hours after challenge, it is not proven. Moreover, it is unknown when the increase has begun. Rectal temperature was recorded consistently throughout the whole trial, so the time points when cows had fever are definite, at least for intervals of one to two hours (see 'IV.3.4.3 Determining threshold values for automatic and manual evaluation'). All cows were found to be healthy in clinical examinations prior to the trial, and fever occurred in a similar period of time after challenge in all cows, ranging from between 13 and 15 hours after challenge to between 11.5 and 17 hours after challenge. Moreover, macroscopic examination of the milk of the challenged quarter obtained 11.5 hours revealed distinct pathological alteration. MITTERHUEMER et al. (2010), whose study refers to the same trial as this study does, reported that all cows developed symptoms of acute mastitis and *E. coli* was ascertained from all milk samples of the challenged quarters, taken 12 hours after challenge.

Consequently, it can almost certainly be stated that all cows developed clinical mastitis, and that fever was a symptom of clinical mastitis. Therefore, data raised at the time points when cows had rectal temperature exceeding 39.5°C are opposed with data raised at the other time points.

The hypothesis, that exclusion of ‘hot spots’ leads to results not related with rectal temperature and depicting only the local inflammatory responses is disproved: temperature courses obtained by automatic and manual evaluation are similar and correlation coefficients of rectal temperature and the results of automatic evaluation are only slightly smaller than those of rectal temperature and manually obtained results. Influence of rectal temperature is reduced, but not eliminated (see ‘4.2 Results of automatic and manual evaluation and their correlation with rectal temperature’). ROC-analysis comparing both methods in detecting clinical mastitis on the basis of fever is thus applicable. Nevertheless, it is important to bear in mind that this analysis in the first place calculates the methods’ sensitivity and the specificity to detect fever, although the fever was due to clinical mastitis in this case. If infrared thermography detects increasing surface temperatures due to increasing body temperatures, or if udder surface temperature rises equally quick but autonomously cannot be definitely resolved in this study.

The overall results of the evaluation parameters ‘Avg’ and ‘Max’ in detecting clinical mastitis are good to very good in both methods. The AUCs (see Table 13) and sums of sensitivity and specificity (see Table 14) indicate that automatic and manual method have a comparably good ability of mastitis detection. The results of manual evaluation are consistently better, but the differences are quite small. For instance, the parameter ‘Max HR’ yields the largest AUCs in both methods: 0.96 in automatic evaluation and 0.98 in manual evaluation, which are both excellent results. The ROC-curves in Figure 30 also appear in a similar pattern in both methods. These findings once more suggest that, in this study, automatic evaluation is able to detect clinical mastitis almost as accurately as manual evaluation.

The results of the parameter ‘Avg HR’ in ROC-analysis are in both methods distinctly inferior to those of the other parameters. This was expectable since, as already discussed, average temperature of the challenged quarter does not increase as much as average temperature in the neighboring quarter. This effect is not apparent in maximum surface temperature, which is an advantage in health supervision when the affected quarter is, of course, unknown.

The threshold values of the parameters in automatic evaluation are smaller than the threshold values in manual evaluation, which was expectable due to the exclusion of ‘hot spots’. When these threshold values are applied in further measurements it

is important to choose a threshold value depending on the selected ROI: if ‘hot spots’ are excluded or not.

5. Interpretation of results of evaluation using the teats and evaluation using the udder and excluding the teats

5.1. Precision analysis

In this precision analysis, parameters evaluating relatively small ROIs (‘Teats’) are opposed with parameters evaluating the larger ROIs of the hind udder surface, excluding the teats (‘Udder’). This relation suggests better precision of evaluation using the udder, especially for ‘Avg’, depicting mean values of all pixels in the ROI. However, this precision analysis is also done to define the most accurate evaluation parameters within the methods.

Regarding the box-plot graphs of Figure 16, median values of the parameters’ VCs seem to be at a similar level in both methods, whereas the ranges of the VCs in evaluation using the teats are noticeably larger than those of evaluation using the udder. VCs of the parameters in evaluation using the teats are broadly scattered. Similarly to the findings in precision analysis of automatic and manual method, the VCs of the parameter ‘Min’ stand out: they are larger than those of the other parameters. This was expectable, as the vulnerability of ‘Min’ towards falsifications is discussed in ‘VI.4.1 Precision analysis’.

The median values of the evaluation parameters’ VCs in Table 42 confirm the findings of precision analysis of automatic and manual evaluation: VCs of ‘Max-Min’ exceed the others by far. Results of one-way analysis of variance (Table 15 and Table 16) prove that ‘Max-Min’ has a significantly lower precision than the other parameters. This applies to both methods. Differences between the VCs of ‘Min’ and those of the other parameters are not significant. However, ‘Avg’ has the lowest median VCs, followed by the VCs of ‘Max’. Consequently, these parameters are picked for further analysis.

What is of interest here is the direct comparison of evaluation using the teats and evaluation using the udder: median values in Table 42 and box-plots in Figure 16 already indicate that the overall precision of evaluation using the udder is superior to evaluation using the teats. This may be due to the ROIs of the teats being a small area, as mentioned above: ‘Avg’ is calculated from less pixels and the manually selected borders may be more likely to include pixels of neighboring structures,

when ROIs are small (see Figure 6: Screenshot of evaluation: left (AR01) and right hind teat (AR02) are selected, using the polygon tool).

Unmatched t-test (see Table 17) comparing the precision of ‘Max’ and ‘Avg’ in evaluation using the teats and evaluation using the udder, separated by hindquarter, shows quite interesting results: As expected, ‘Max HL’, ‘Avg HL’ and ‘Avg HR’ are significantly more precise in evaluation using the udder than in evaluation using the teats, whereas the difference between the mean VCs of ‘Max HR’ in the two methods approaches zero (-0.00001). This result is surprising, since only data from before challenge were analyzed. Consequently, there is no explanation why such a distinct difference between the precision in the left and right teat exists. In Tukey’s post-hoc test, no significant difference between the precision of ‘Teats Max HR’ and ‘Teats Max HL’ was detectable, so the only finding that can be stated is that the precision of ‘Max’ and ‘Avg’ is inferior in evaluation using the teats, except for the precision of ‘Max HR’, which is equal in both methods. This result may be due to chance, but the P-value of 0.97 indicates that these results are observed in two groups between which actually no difference exists. However, concluding that only the maximum temperature of the right teat should be evaluated since it has the same accuracy as the gold standard would be exaggerated. It should not be forgotten that VCs of ‘Max’ and ‘Avg’ are also relatively small in evaluation using the teats, even though they were significantly higher than in evaluation using the udder. ‘Max’ and ‘Avg’ are valid parameters for both, evaluation using the teats and evaluation using the udder.

5.2. Results of evaluation using the teats and evaluation using the udder and their correlation with rectal temperature

Comparing Figure 17 with Figure 18, it becomes obvious that temperature courses of the parameters in evaluation using only the teats do not resemble the course of rectal temperature, the way the courses of parameters in evaluation using the udder do. Consequently, a lower correlation with rectal temperature is expected for evaluation using the teats. Regarding the aim of this study, it was of interest to depict local temperature changes with evaluation using the teats. Therefore, evaluation using the teats is not consequently inferior to evaluation using the udder.

In total, courses of median temperature values in evaluation using the teats proceed at a distinctly lower level than in evaluation using the udder. This is not too surprising, since teats have a small volume related to their surface, and they are in

an exposed position, distant from the center of the body. Logically, their surface temperature is constantly lower than the surface temperature of the udder. However, the appearance of the courses gives rise to the hypothesis that surface temperature of the teats is more vulnerable towards environmental influences, for instance airflow, than udder surface temperature. Especially the course of 'Min' is completely irregular in evaluation using the teats, compared to evaluation using the udder. This was expectable, since cooler pixels of the environment, falsely selected into the ROI, are much more probable for the ROI of the teat, being in an exposed position, than for the ROI of the udder without teats.

The course of 'Avg' also seems to proceed with some irregularities: As mentioned above, since 'Avg' is a mean value, outliers have a larger impact on it in a small ROI (like a teat) compared to larger ROI (like the hind udder surface). However, the average temperature is peaking distinctly between 13 and 15 hours after challenge in evaluation using the teats, especially in the right, challenged teat: an increase is already visible 9 hours after challenge, and not 11 hours after challenge, like in evaluation using the udder. In the left, placebo-treated teat, the temperature increase is split into two peaks and has a lower amplitude than in the challenged teat. It is quite interesting that the challenged teat shows a more distinct temperature elevation than the placebo-treated teat. However, in the course of maximum temperature in evaluation using the teats, differences between left and right hindquarter are not apparent. The direct comparison of evaluation using the teats and evaluation using the udder is discussed below (see '5.3 Comparison of evaluation using the teats and evaluation using the udder').

As it was expected, evaluation using the teats gains in Pearson correlation analysis with rectal temperature much lower correlation coefficients than evaluation using the udder (see Table 18 and Table 19). The parameter 'Max' still yields moderately good results in evaluation using the teats, but they are distinctly inferior to those of evaluation using the udder. This finding proves that evaluating only the teat and thereby excluding the 'hot spots' reduces the impact of inner body temperature on the measurement results, but up to now, no statements can be made if this improves the mastitis detection or not.

Concerning the parameter 'Avg', it should be noticed that average surface temperature of the challenged teat correlates better ($r=0.56$) with rectal temperature than average surface temperature of the placebo-treated teat ($r=0.33$). In evaluation

using the udder it is the exact opposite. This may be due to the teat lacking subcutaneous tissue and thus not suffering from the subcutaneous edema of the challenged udder quarter described in earlier studies (BARTH, 2000; METZNER et al., 2015).

It is not surprising that 'Min' yields poor results in correlation analysis with rectal temperature. Although some of the correlation coefficients are statistically significant, they are too small to indicate a distinct relation. This applies to both methods and both hindquarters.

The standard deviations of the differences (see Table 20) between 'Avg' and rectal temperature being constantly smaller in evaluation using the udder than in evaluation using the teats support the findings of correlation analysis: small standard deviations indicate a consistent difference and thus a consistent relation between average surface temperature and rectal temperature. T-test can only detect significant differences between both methods' standard deviations before challenge but not after challenge. Especially between the right (challenged) teat and udder quarter, differences of the standard deviations are small after challenge. It can be assumed that the rectal temperature and teat surface temperature become closer related when the cow is suffering from clinical mastitis accompanied by fever.

5.3. Comparison of evaluation using the teats and evaluation using the udder

The courses of the median temperature values in Figure 19 and Figure 20 illustrate the different patterns obtained by evaluation using the teats and evaluation using the udder. In the first place, it is apparent that the teats show much lower median surface temperatures than the udder. Differences of maximum temperatures between both methods are noticeable (2.30°C (HL), respectively 2.23°C (HR)) and underline that different temperature levels should be applied for mastitis detection using the teats and mastitis detection using the udder. However, this temperature differences are expectable, as well as the unsteady pattern of teat surface temperature, since the teats are located in a distant position from the body center and exposed to environmental influences (see '5.2 Results of evaluation using the teats and evaluation using the udder and their correlation with rectal temperature').

Regarding the maximum surface temperature (see Figure 19), the peaks seem to occur in the same period of time. However, maximum surface temperature seems

to rise less in the placebo-treated teat compared to the placebo-treated quarter, whereas the maximum surface temperature of the challenged teat appears to show a sharper increase than in the challenged quarter. Nevertheless, it is questionable if these slight differences would have an impact on mastitis detection under practical conditions.

More distinct differences between the methods are found when the evaluation parameter ‘Avg’ is considered (see Figure 20). In the left, placebo-treated quarter, evaluation using the udder detects an obviously more pronounced temperature increase than evaluation using the teat, whereas in the right, challenged quarter, it is the exact opposite: average surface temperature of the teat increases earlier than average surface temperature of the udder. Moreover, the peak has a larger amplitude and it even exceeds the median value of average surface temperature of the udder 11.5 hours after challenge. This observation is important for two reasons:

Firstly, it approves the hypothesis of BARTH (2000); GLAS (2008) and METZNER et al. (2015), that average temperature of the challenged quarter rises less than average temperature of the unaffected quarter, which is probably due to subcutaneous edema. This is again observed in ‘Avg’ of evaluation using the udder, but not in evaluation using the teats. However, when DRUMMER (2009) evaluated the ultrasonographic assessment of peripartal udder edema, the teat basis was chosen as the most suitable localization. As stated by STOCKER and RÜSCH (1997), largest fluid accumulation of udder edema occurs in the distal udder subcutis. This may be the reason why DRUMMER (2009) detected udder edema most distinctly at the teat base. The teat itself lacks subcutaneous tissue, and although it is not mentioned by STOCKER and RÜSCH (1997), it is most probably that in this case fluid accumulates in the interstitium of the layer of muscles and connective tissue. Obviously, due to the different expression of the edema in the teat, less buffer is situated between inflamed teat tissue and skin surface than between udder tissue and skin surface.

Secondly, it concerns the question raised in the introduction: “Is an earlier detection of clinical mastitis possible if only the surface of the teats is selected in manual interpretation?”. Evaluating only the teats eliminates the influences of ‘hot spots’ in the udder-thigh gap and intermammary sulcus. Furthermore, recent studies underline the role of the teat in first defense mechanisms towards pathogens (RINALDI et al., 2010; LIND et al., 2015; PETZL et al., 2016). The challenged teat

indeed shows a more distinct peak in average temperature than the challenged quarter. If this peak is statistically significant and if it is actually earlier than in evaluation using the udder is discussed later (see: ‘5.3.2 Comparison of period after challenge and reference period in evaluation using the teats and evaluation using the udder’).

5.3.1. Correlation analysis of evaluation using the teats and evaluation using the udder

Prior to correlation analysis, scatter plots of the values ‘Avg’ and ‘Max’ in evaluation using the teats and evaluation using the udder are assessed to choose an appropriate test. Since the scatter plots in Figure 28 indicate linear correlation, Pearson correlation analysis is regarded as suitable.

The overall correlation coefficients are moderately good, and it is noticeable that maximum and average surface temperature of evaluation using the teats and using the udder are better correlated in the challenged quarter than in the placebo-treated quarter. Temperature of the challenged teat is also more related to rectal temperature than temperature of the placebo-treated teat, which may explain this observation (see ‘5.2 Results of evaluation using the teats and evaluation using the udder and their correlation with rectal temperature’).

Nevertheless, it has to be said that for the studies’ questioning if evaluation method using only the teats is more suitable for mastitis detection than evaluation method using the udder, correlation between the methods may not be decisive. Following analysis’ results, which method is able to detect temperature changes earlier (see: ‘3.3.2 Comparison of period after challenge and reference period in evaluation using the teats and evaluation using the udder’) have to be taken into account.

High correlation coefficients are preferable when one method’s results are compared with the results of the gold standard, like in the comparison of automatic and manual method (see: ‘4.3.1 Correlation analysis of automatic and manual evaluation’). When, like in this case, an evaluation method shall be enhanced, for instance by different ROIs, perfect correlation coefficients indicate no improvement. The results of correlation analysis observed here suggest that the results of both methods are related to some extent. Nevertheless, it is possible that one method is superior to the other in mastitis detection.

5.3.2. Comparison of period after challenge and reference period in evaluation using the teats and evaluation using the udder

The calculation of the differences between reference period and period after challenge, as well as One-way ANOVA and Dunnett's post-hoc test, are similar to those of automatic and manual evaluation. Their design is thus discussed in '4.3.2 Comparison of period after challenge and reference period before challenge in automatic and manual evaluation'.

The differences of the temperature values at the time points in reference period and period after challenge are visualized in box-plots in Figure 21 (evaluation using the teats), respectively Figure 22 (evaluation using the udder).

In evaluation using the teats, differences of maximum temperature increase at '11' and peak at '13', '15' and '17'. Differences between left and right hindquarter are not apparent. This is similar to what is observed in automatic evaluation and (manual) evaluation using teats and udder as a union. Maximum temperature differences in evaluation using the udder proceed in the same pattern, but appear larger and less scattered than in evaluation using the teats between '13' and '15'. These findings do not support the assumption that evaluation using the teats detects clinical mastitis earlier than evaluation using the udder.

Regarding the differences of average surface temperature, they appear extremely widespread in evaluation using the teats. Although the medians and mean values range around zero, several differences are noticeably large from '-1' to '9'. Starting at '13', temperature differences in evaluation seem to scatter less, as if the warming of the teats due to mastitis prevents external temperature alterations. Even in the right (challenged) teat, where course of average surface temperature indicates more distinct peaks than in average surface temperature of the udder, differences between reference period and period after challenge become less pronounced due to wide scattering of the differences before.

In evaluation using the udder, the differences between reference period and period after challenge proceed in a steadier pattern and only show little scattering compared to evaluation using the teats. However, it becomes obvious that average temperature of the challenged quarter is hardly different before and after challenge. Nevertheless, the distinct differences of maximum temperature can compensate for this.

The results of One-way ANOVA and Dunnett's post-hoc test (see Table 23 and Table 24) are surprising: Although One-way ANOVA detects significant differences among the differences of 'Max HL', 'Max HR' and 'Avg HR', Dunnett's post-hoc test finds that evaluation using the teats is only able to detect significant differences 15 hours after challenge with the evaluation parameter 'Max HL'. Consequently, evaluation using the teats is inferior to evaluation using the udder in this study: Using the udder as ROI, significant differences can be detected 13 and 15 hours after challenge with 'Max HL' and 'Max HR'. Consequently, evaluation method using the udder detects significant temperature differences with the parameters 'Max HL' and 'Max HR' earlier than evaluation using only the teats and in both hindquarters.

It is also surprising, that the differences of 'Avg HR' in the reference period and period after challenge are not conclusive. In Table 45 and Table 46 it is shown that they are noticeably larger than those of evaluation using the udder, yet they are not significant in Dunnett's post-hoc test. A possible explanation is that the differences at the other time points are too widespread to let the differences of 'Avg HR' between '11' and '17' stand out. It is also possible that in a trial with a larger number of cows less scattering would occur and that this effect would be more pronounced.

Evaluation using the teats is less precise and not as suitable for mastitis detection as evaluation using the udder in this study. However, bearing in mind that average temperature of the challenged teat proceeds completely different than average temperature of the placebo-treated teat, an observation that is not made in the other methods, and that average temperature of the challenged teat seems to rise earlier than average temperature of the challenged quarter; evaluation using the teats should not be rejected from the outset, but considered as an additional ROI in large-scale studies.

5.3.3. Results of ROC-analysis of evaluation using the teats and evaluation using the udder

ROC-analysis of evaluation using the teats and evaluation using the udder is performed with the same criteria as ROC-analysis of automatic and manual evaluation. Consequently, its limitations are discussed in: '4.3.3 Limitations and results of ROC-analysis of automatic and manual evaluation'.

It is already known that evaluation using the teats is not able to detect surface temperature changes earlier than evaluation using the udder. Much more, it is disadvantaged in detecting significant differences in surface temperature compared to evaluation using the udder. However, ROC-analysis evaluates both methods' ability to correctly identify pathological and physiological conditions and calculates an according threshold-value.

The AUCs (see Table 25) of the parameters in evaluation using the teats are smaller than the AUCs of the parameters in evaluation using the udder, except for 'Avg HR' (0.86 versus 0.76). 'Avg HR' of evaluation using the teats has also a larger sensitivity, but a lower specificity than 'Avg HR' of evaluation using the udder. Apart from 'Avg HR' in evaluation using the teats, evaluation parameter 'Max' gains better results of sensitivity and specificity than 'Avg' in both methods. In total, evaluation using the udder has clearly better results in ROC-analysis than evaluation using the teats. However, it has to be considered that criteria for this analysis are based on rectal temperature, and surface temperature of the udder quarters was proved to be more related to rectal temperature than surface temperature of the teats, except average surface temperature of the challenged teat (see: 'V.3.2 Correlation of evaluation method and rectal temperature'). If surface temperature of the teats showed earlier temperature peaks or a temperature course completely independent from rectal temperature, this would have falsified the outcome of this ROC-analysis. However, apart from 'Avg HR' seeming to increase sharper, evaluation using the teats was not able to detect significant earlier temperature changes than evaluation using the udder.

6. Conclusions

Interpretation of the results lead to the following conclusions concerning the study's questioning:

The current gold standard of mastitis detection using IRT is manual evaluation. Does an automated image recognition software provide comparable results?

Compared to the current gold standard of manual evaluation, automatic evaluation of the same set of thermograms provides a good detection rate. Evaluation parameters of automatic evaluation have a good precision; however, it is slightly lower than in manual evaluation. The detected course of average and maximum surface temperature is similar in both methods. Due to slightly different ROIs the courses of automatic evaluation proceed on a lower temperature level. Both methods detect distinct temperature peaks between 13 and 15 hours after challenge with a similar amplitude. They are significant in evaluation parameter 'Max' in both hindquarters. Both methods are closely correlated in their results of average and maximum surface temperature. Results of manual evaluation are slightly closer related to rectal temperature than results of automatic evaluation. However, rectal temperature has a large impact on the results of both methods. Automatic and manual evaluation both have very good sensitivity and specificity in detecting acute *E. coli* mastitis accompanied by fever with the parameters 'Avg' and 'Man'. The according threshold values differ.

The results of maximum temperature are similar for the challenged and the placebo treated quarter in both methods. Average surface temperature increased less in the challenged quarter than in the placebo-treated quarter. This was equally detected by both methods.

In total, automatic evaluation provides comparably good results as manual evaluation.

Is an earlier detection of clinical mastitis possible if only the surface of the teats is selected in manual interpretation?

The teats are situated at a position distant from the body center and exposed to environmental temperature influences. Results of evaluation using the teats are less precise and the courses less steady and on a lower temperature level than those of

evaluation using the hind surface of the udder. The results of both methods are moderately correlated. Teat surface temperature is less related to rectal temperature than udder surface temperature. Evaluation using the udder detects temperature peaks earlier and for a longer period of time. Solely the average surface temperature of the challenged teat shows a sharper and probably earlier increase. This is probably due to differently located inflammatory edema, that functions as a temperature buffer in the subcutis of the challenged quarter but not in the challenged teat. However, the differences are not significant.

Evaluation using the teats is inferior in detecting *E. coli* mastitis to evaluation using the udder in this study. Nevertheless, the different courses of average surface temperature of challenged and placebo treated teat are noticeable. Although the teats should not be used as only ROI in mastitis detection, their use as an additional ROI in large-scale studies would be of interest.

Can a standardized exclusion of ‘hot spots’ from evaluation lead to better results in detecting mastitis?

Automatic evaluation excludes the ‘hot spots’ by 5% of the pixels, originating from the outer borders, and thereby reduces impact of rectal temperature on evaluation results. However, results of automatic evaluation are equally good, but not better, compared to manual evaluation in which ‘hot spots’ are not excluded: temperature courses proceed on a lower level in automatic evaluation but in a similar pattern as in manual evaluation.

Evaluation using only the teats obviously excludes the ‘hot spots’. However, the attained results in detecting significant temperature differences are inferior to those of evaluation using the udder, which includes ‘hot spots’.

Exclusion of ‘hot spots’ is thus not necessary for detection of acute *E. coli* mastitis. Nevertheless, it may help to reduce falsifications in automatic evaluation: If the outer borders of the ROI are the outer borders of the udder surface, pixels belonging to the neighboring structures might falsely be included into the ROI.

If exclusion of ‘hot spots’ is applied, it has to be consistently done to the same percentage to avoid falsifications. Moreover, different threshold values for mastitis detection are indicated in that case.

Further conclusions

Polygons of the hindquarters' surface, excluding a small percentage of pixels originating from the outer borders, are suitable ROIs for automatic evaluation of thermograms of the udder. The silhouette of the teats plays an important role in automatic image recognition using the Active Shape Model, thus, they are included into the ROIs.

The evaluation parameters 'Max' and 'Avg' are proven to be most useful and most precise in evaluation of udder thermograms. 'Max' is most suitable for detection of significant temperature differences and has better results of sensitivity and specificity than 'Avg'. Since it is unknown how vulnerable 'Avg' is towards gross contamination or covering in the field, whereas 'Max' always depicts the warmest pixel; 'Max' is considered as most valuable in detection of acute *E. coli* mastitis. If 'Avg' is used, left and right hindquarter have to be separated into two ROIs since affected and unaffected quarters show different average surface temperature. 'Min' and 'Max-Min' are not suitable for evaluation.

A simple comparison of temperatures of left and right hindquarter is not sufficient for mastitis detection. Average temperature even rises less in the challenged quarter, which is most probably due to edema. Challenged and placebo-treated quarter show a similar course in maximum temperature. Fortunately, this also means that in all probability, clinical *E. coli* mastitis of the forequarters can be detected by screening the maximum temperature of the hindquarters, which would be a major benefit for automatic evaluation.

Material for this study was recorded under optimized conditions. Further research has to be done how automatic evaluation works under the conditions of a stable area.

Results of this trial once more emphasize the importance of short intervals between measurements. In this case, temperature elevation was distinct in a period of approximately four hours and significant in a period of two hours. A definite recommendation for the localization of the thermal camera in the stable to ensure these short intervals does not exist yet.

In this study, significant temperature elevation is detected starting from 13 hours after challenge. This is a promising short time, aiming the successful treatment of

clinical mastitis. Of course, time of challenge with a pathogen is unknown, and various reasons for temperature elevation have to be considered. It is important to bear in mind that this method probably also depicts changes in inner body temperature, respectively fever. However, IRT is to be used as automatic health supervision tool and shall not replace examination of individuals. The fact that it also reports fever should not be regarded as disadvantage. Animals that have abnormal USST are identified quickly and can be assessed in detail. It is conceivable that either the individual temperature course of each animal, or the surpassing of a threshold value is monitored in short intervals. However, this method's ability to detect subclinical mastitis, clinical mastitis caused by other pathogens or systemic diseases without fever is not evaluated.

VII. SUMMARY

Automatic evaluation of infrared thermal images of bovine udders and teats challenged with *E. coli*

Mastitis is one of the most frequent diseases in dairy cows and causes substantial economic losses and suffering of the affected animals. Infrared thermography is a noninvasive tool to detect clinical mastitis early. The current gold-standard of manual evaluation of the thermograms is however time-consuming and requires a skilled examiner. Due to growing herd sizes in dairy cow farming, there is a need for automated health supervision. This study concerns the question if evaluation of thermograms of the bovine udder by an automatic image recognition software provides results comparable to those of manual evaluation in detecting clinical *E. coli* mastitis. Moreover, it is questioned if the exclusion of typical ‘hot spots’ (udder-thigh cleft and intermammary sulcus), which show temperatures closely related to inner body temperature, leads to a better depiction of local inflammation processes in the udder. Since the teat is the first immunological barrier to react towards invading pathogens, it is evaluated, whether evaluating only the region of the teats leads to an earlier mastitis detection.

For this purpose, thermographic material is used that emerged from an experimental infection study (GLAS, 2008): Five healthy Holstein-Friesian dairy cows were challenged with *E. coli* intracisternally into the right hindquarter. As a result, all cows developed signs of clinical mastitis. In a period of 24 hours before challenge and 24 hours after challenge, thermograms of the hind udder surface were taken in intervals of two hours with the help of an infrared camera.

The same thermograms are repeatedly evaluated with different methods (I-III): an automatic image recognition software (**‘Aut’, I**), based on the Active Shape Model, detects the silhouette of the udder and creates two regions of interest (ROIs) in each thermogram: the hind surface of the udder including the teats, divided by left (HL) and right hindquarter (HR). In a second step, 5% of the pixels inside each ROI, originating from the outer borders, are automatically excluded. For comparison with automatic evaluation, thermograms are evaluated manually with a polygon-tool (**‘Man’, II**). The same ROIs are selected, but no exclusion of ‘hot spots’ is done.

For the second part of the study, the ROIs of left and right hind teat are manually evaluated with the polygon tool ('Teats', III) and compared with data emerging from a former study in which the same thermograms were manually evaluated with the polygon tool, using the surface of the hindquarters without the teats as ROIs ('Udder').

Results of automatic ('Aut') and manual ('Man') evaluation:

Automatic evaluation has a low rate of falsely detected ROIs (2-3%). Results of automatic evaluation are less correlated with rectal temperature than results of manual evaluation, but correlation coefficients are still moderately large. All cows showed fever ($>39.5^{\circ}\text{C}$) after intramammary *E. coli* challenge, peaking 13-15 hours after challenge. Peaks of average ('Avg') and maximum surface temperature ('Max') of challenged **and** placebo-treated quarter occur in a similar time-span. Peaks of maximum temperature are equally high in both quarters, average temperature peaks less in the challenged quarter. The course of automatically ascertained temperatures is similar to the course of manually ascertained temperatures, but proceeds on a lower temperature level. The results of both methods are highly correlated: $r=0.98$ ('Avg HL'), respectively $r=0.99$ ('Avg HR'). Significant temperature differences between period after challenge and reference period are detected in both hindquarters 13 and 15 hours after challenge, using the parameter 'Max'. This applies to both methods. In ROC-analysis, both methods provide good results for sensitivity and specificity at different threshold values: 'Aut Max HR': threshold $\geq 37.42^{\circ}\text{C}$, sensitivity=93.75%, specificity=94.96%; respectively: 'Man Max HR': threshold $\geq 38.65^{\circ}\text{C}$, sensitivity=93.75%, specificity=96.40%.

Results of evaluation using the teats ('Teats') and evaluation using the udder ('Udder'):

Average surface temperature of the challenged teat peaks more than average surface temperature of the placebo-treated teat. In evaluation using the udder it is the other way round: average surface temperature of the challenged quarter rises less. Temperature peaks occur around 13 and 15 hours after challenge in both methods. Results of evaluation using the teats are distinctly less correlated with rectal temperature than results of evaluation using the udder. Maximum temperature results of both methods are highly correlated whereas results of average surface

temperature are only moderately correlated. Evaluation using the teats detects significant temperature differences only in maximum temperature of the left teat 15 hours after challenge, whereas evaluation using the udder detects significant differences of maximum temperature of both hindquarters 13 to 15 hours after challenge. Differences of 'Avg' are not significant. In ROC-analysis, evaluation using the teats is distinctly inferior to evaluation using the udder. Solely the parameter 'Teats Avg HR' provides moderately good results.

Automatic evaluation of thermograms of bovine udders challenged intramammary with *E. coli* provides good results in clinical mastitis detection and is comparably valid as the current gold standard of manual evaluation, alongside with a good detection rate.

Evaluation using the teats is inferior in detecting mastitis to evaluation using the udder in this study. Nevertheless, the different courses of average surface temperature of challenged and placebo treated teat are noticeable.

Exclusion of 'hot spots' does not lead to better detection of acute *E. coli* mastitis. Nevertheless, it may help to reduce falsifications in automatic evaluation.

The evaluation parameters 'Max' and 'Avg' are proven to be most useful and most precise in evaluation of udder thermograms. 'Max' is more suitable for detection of significant temperature differences and has better results of sensitivity and specificity than 'Avg'.

A simple comparison of temperatures of left and right hindquarter is not sufficient for detection of acute *E. coli* mastitis. Average temperature even rises less in the challenged quarter, which is most probably due to edema. *E. coli* mastitis of the forequarters can probably be detected by screening the maximum temperature of the hindquarters. For reliable udder health monitoring, recording of udder thermograms should be implemented in short intervals not exceeding two hours.

VIII. ZUSAMMENFASSUNG

Automatisierte Auswertung von Thermogrammen des Euters und der Zitzen von Kühen mit induzierter *E. coli*-Mastitis

Mastitis ist eine der häufigsten Erkrankungen der Milchkuh und verursacht erhebliche wirtschaftliche Verluste sowie Leiden der betroffenen Tiere. Die Infrarot-Thermografie ist eine nichtinvasive Methode zur frühzeitigen Erkennung von klinischen Mastitiden. Der aktuelle Gold-Standard der manuellen Auswertung von Thermogrammen ist jedoch sehr zeitaufwändig. Dieser Studie liegt die Fragestellung zugrunde, ob eine automatisierte Auswertung von Thermogrammen des Euters mithilfe einer Bilderkennungssoftware vergleichbar gute Ergebnisse zur Erkennung von klinischen *E. coli*-Mastitiden liefert wie die manuelle Auswertung. Des Weiteren wird untersucht, ob ein Ausschluss sogenannter „Hot Spots“ (Euter-Schenkel-Spalt und Sulcus intermammarius) zu einer besseren Darstellung lokaler Entzündungsprozesse im Rahmen einer klinischen Mastitis führt. Da die Zitze die erste immunologische Hürde darstellt, die auf in das Euter eindringende Pathogene reagiert, wird außerdem untersucht ob eine Auswertung der Thermogramme, die sich ausschließlich auf die Oberfläche der Zitzen beschränkt, eine frühere Erkennung von *E. coli*-Mastitiden bietet.

Zu diesem Zweck wird thermografisches Material verwendet, das aus Infektionsexperimenten einer Vorläuferstudie entstammt (GLAS, 2010): fünf gesunde Milchkuhe der Rasse Holstein-Friesian wurden auf dem rechten, hinteren Euterviertel mit *E. coli* (1303) infiziert, woraufhin alle Kühe eine klinische Mastitis entwickelten. In einem Zeitraum von 24 Stunden vor und 24 Stunden nach der Infektion wurden in einem Intervall von zwei Stunden Thermogramme der hinteren Euteroberfläche mithilfe einer Infrarot-Kamera angefertigt.

Diese Thermogramme werden mit drei verschiedenen Methoden (I-III) ausgewertet: eine dafür entwickelte automatisierte Bilderkennungssoftware (SCHRÖTER, 2015), basierend auf dem „Active Shape Model“, erkennt die Eutersilhouette (**Aut; I**) und erstellt zwei Auswahlflächen: die Oberfläche des linken (HL) und rechten (HR) hinteren Euterviertels inklusive der Zitze. In einem zweiten Schritt werden 5% der Pixel innerhalb der Auswahlflächen, ausgehend von den Außengrenzen, automatisiert entfernt um die sogenannten „Hot Spots“ auszuschließen. Dieselben Thermogramme werden mithilfe eines Polygon-Tools

manuell ausgewertet (**Man‘; II**): die gleichen Auswahlflächen werden verwendet, ein Ausschluss der „Hot Spots“ findet bei dieser Methode nicht statt. Für den zweiten Teil der Studie werden die hinteren Zitzen als Auswahlfläche mit dem Polygon-Tool manuell ausgewertet (**Teats‘; III**) und mit Daten einer früheren Studie verglichen, in dem dieselben Thermogramme mit den Auswahlflächen des rechten und linken Hinterviertels ohne Zitzen manuell ausgewertet wurden (**Udder‘**).

Ergebnisse von automatisierter (,Aut‘) und manueller (,Man‘) Auswertung:

Die automatisierte Bilderkennungssoftware hat eine geringe Fehlerrate, mit der die Auswahlflächen nicht korrekt erkannt werden (2-3%). Die Ergebnisse der automatisierten Auswertung sind mit der Rektaltemperatur deutlich korreliert, jedoch weniger als die Ergebnisse der manuellen Auswertung. Alle Kühe zeigten nach der intrazisternalen *E. coli*-Infektion Fieber ($>39.5^{\circ}\text{C}$); die höchsten Werte der Rektaltemperatur wurden 13-15 Stunden p.i. gemessen. Die Höchstwerte der maximalen (,Max‘) und durchschnittlichen (,Avg‘) Oberflächentemperatur des infizierten Viertels **und** des mit einem Placebo behandelten Viertels werden im gleichen Zeitraum gemessen. Die Höchstwerte von ,Max‘ sind in beiden Vierteln gleich hoch, die Höchstwerte von ,Avg‘ steigen im infizierten Viertel weniger an als im Placebo-behandelten Viertel. Die automatisch ermittelten Temperaturen folgen dem Muster der manuell ermittelten Temperaturen, verlaufen aber auf einem deutlich niedrigeren Temperaturniveau. Die Ergebnisse der automatischen und manuellen Methode sind stark positiv korreliert: $r=0.98$ (,Avg HL‘), beziehungsweise $r=0.99$ (,Avg HR‘). Signifikante Unterschiede von ,Max‘ zwischen den Referenzmessungen vor der Infektion und den Messungen nach der Infektion ermitteln beide Methoden 13 und 15 Stunden p.i. in beiden Hintervierteln. Die ROC-Analyse ermittelt für beide Methoden gute Ergebnisse für Sensitivität und Spezifität, jedoch bei unterschiedlichen Schwellenwerten: ‘Aut Max HR’: Schwellenwert $\geq 37.42^{\circ}\text{C}$, Sensitivität=93.75%, Spezifität=94.96%; beziehungsweise: ‘Man Max HR’: Schwellenwert $\geq 38.65^{\circ}\text{C}$, Sensitivität=93.75%, Spezifität=96.40%.

Ergebnisse der Auswertung die die Zitzen nutzt (,Teats‘) und der Auswertung die den Euterspiegel nutzt (,Udder‘):

„Avg“ der infizierten Zitze steigt stärker an als die der mit einem Placebo behandelten Zitze. Bei der Auswertung „Udder“ wird das Gegenteil beobachtet: „Avg“ des infizierten Viertels steigt weniger an. Bei beiden Methoden werden Temperatur-Höchstwerte 13 und 15 Stunden nach der Infektion aufgezeichnet. Die Ergebnisse der Auswertungsmethode „Teats“ sind deutlich weniger mit der Rektaltemperatur korreliert als die der Methode „Udder“. Signifikante Unterschiede zwischen den Messungen nach der Infektion und den Referenzmessungen 24 Stunden zuvor sind bei der Auswertungsmethode „Teats“, nur für die Messung 15 Stunden p.i. und „Max HL“ nachweisbar, während bei der Methode „Udder“ für die Messungen 13 und 15 Stunden p.i. signifikante Unterschiede von „Max“ nachweisbar sind. Die Differenzen von „Avg“ sind bei beiden Methoden nicht signifikant. In der ROC-Analyse ist die Auswertungsmethode „Teats“ der Auswertungsmethode „Udder“ deutlich unterlegen, lediglich „Teats Avg HR“ erzielt mäßig gute Ergebnisse.

Die automatisierte Auswertung der Thermogramme bietet eine gute Erkennungsrate der Eutersilhouette und eine ähnlich gute Erkennung von klinischen *E. coli*-Mastitiden wie der aktuelle Gold-Standard der manuellen Auswertung.

Die Auswertungsmethode, die lediglich die Auswahlfläche der Zitzen nutzt, ist der Auswertungsmethode, die den Euterspiegel nutzt, in der Erkennung von *E. coli*-Mastitiden unterlegen. Dennoch sollte beachtet werden, dass die „Avg“ der infizierten Zitze stärker ansteigt, als die der nicht-infizierten Zitze, während es sich am Euterspiegel gegensätzlich verhält.

Ein Ausschluss der „Hot Spots“ führt nicht zu einer verbesserten Erkennung von *E. coli*-Mastitiden. Dennoch könnte er dazu beitragen, Verfälschungen der Auswahlfläche zu vermeiden.

Die Messparameter „Max“ und „Avg“ erweisen sich als präzise und zweckdienlich in der Auswertung von Thermogrammen des Euters. „Max“ erzielt bessere Ergebnisse in der Erkennung von signifikanten Temperaturunterschieden sowie der Sensitivität und Spezifität in der Erkennung klinischer *E. coli*-Mastitiden als „Avg“.

Ein einfacher Vergleich von infiziertem und nicht infiziertem Viertel ist nicht ausreichend für die Erkennung einer klinischen *E. coli*-Mastitis, da ‚Max‘ beidseits in gleichem Maße ansteigt und ‚Avg‘ des infizierten Viertels, wahrscheinlich bedingt durch das subkutane Ödem, sogar weniger ansteigt. Es ist daher wahrscheinlich, dass eine klinische *E. coli*-Mastitis eines vorderen Euterviertels auch durch die Auswertung der Oberflächentemperatur der Hinterviertel feststellbar wäre. Für eine verlässliche Überwachung der Eutergesundheit mittels Infrarotthermografie sollte der zeitliche Abstand zwischen den Messungen zwei Stunden nicht überschreiten.

IX. BIBLIOGRAPHY

Bannerman, D. D., Paape, M. J., Lee, J. W., Zhao, X., Hope, J. C., Rainard, P. (2004):

Escherichia coli and *Staphylococcus aureus* elicit differential innate immune responses following intramammary infection. *Clinical and Diagnostic Laboratory Immunology*; 11: 463-472.

Bar, D., Tauer, L. W., Bennett, G., Gonzalez, R. N., Hertl, J. A., Schukken, Y. H., Schulte, H. F., Welcome, F. L., Grohn, Y. T. (2008):

The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. *Journal of Dairy Science*; 91: 2205-2219.

Barth, K. (2000):

Basic investigations to evaluate a highly sensitive infrared-thermograph-technique to detect udder inflammation in cows. *Milchwissenschaft*; 55: 607-609.

Bergmann, C. (1848):

Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse, abgedruckt aus den Göttinger Studien, Vandenhoeck und Ruprecht, Göttingen: 89-96

Berry, R. J., Kennedy, A. D., Scott, S. L., Kyle, B. L., Schaefer, A. L. (2003):

Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection. *Canadian Journal of Animal Science*; 83: 687-693.

Bitman, J., Lefcourt, A., Wood, D. L., Stroud, B. (1984):

Circadian and ultradian temperature rhythms of lactating dairy cows. *Journal of Dairy Science*; 67: 1014-1023.

Burvenich, C., Van Merris, V., Mehrzad, J., Diez-Fraile, A., Duchateau, L. (2003):

Severity of *E. coli* mastitis is mainly determined by cow factors. *Veterinary Research*; 34: 521-564.

Cena, K., Clark, J. A. (1973a):

Thermographic Measurements of the Surface Temperatures of Animals. *Journal of Mammalogy*; 54: 1003-1007.

Cena, K., Clark, J. (1973b):

Thermal radiation from animal coats: coat structure and measurements of radiative temperature. *Physics in medicine and biology*; 18: 432.

Chang, G., Petzl, W., Vanselow, J., Gunther, J., Shen, X., Seyfert, H. M. (2015):

Epigenetic mechanisms contribute to enhanced expression of immune response genes in the liver of cows after experimentally induced *Escherichia coli* mastitis. *Veterinary Journal*; 203: 339-341.

Colak, A., Polat, B., Okumus, Z., Kaya, M., Yanmaz, L. E., Hayirli, A. (2008):

Short communication: early detection of mastitis using infrared thermography in dairy cows. *Journal of Dairy Science*; 91: 4244-4252.

Cootes, T., Hill, A., Taylor, C., Haslam, J. (1994):

Use of active shape models for locating structures in medical images. *Image and Vision Computing*; 12: 355-365.

-
- Cootes, T. F., Taylor, C. J., Cooper, D. H., Graham, J. (1995):**
Active Shape Models -Their Training and Application. Computer Vision and Image Understanding; 61: 38-59.
- Drummer, C. (2009):**
Ultrasonographische Untersuchungen zum Verlauf von Euterödemen beim Rind und Kontrolle des Behandlungserfolges mit einem Diuretikum. Dissertation, Tierärztliche Hochschule Hannover; http://elib.tiho-hannover.de/dissertations/drummerc_ws09.pdf
- Deutsche Veterinärmedizinische Gesellschaft (2002):**
Leitlinien zur Bekämpfung der Mastitis des Rindes als Herdenproblem. In: Sachverständigenausschuss "Subklinische Mastitis". DVG, Gießen.
- Franze, U., Geidel, S., Heyde, U., Schroth, A., Wirthgen, T., Zipser, S. (2012):**
Möglichkeiten des Einsatzes der Infrarot-Thermographie zur automatischen Gesundheitsüberwachung bei Milchkühen. Züchtungskunde; 84: 158-170.
- Frost, A. J., Hill, A. W., Brooker, B. E. (1982):**
Pathogenesis of experimental bovine mastitis following a small inoculum of *Escherichia coli*. Research in veterinary science; 33: 105-112.
- Gilbert, F. B., Cunha, P., Jensen, K., Glass, E. J., Foucras, G., Robert-Granie, C., Rupp, R., Rainard, P. (2013):**
Differential response of bovine mammary epithelial cells to *Staphylococcus aureus* or *Escherichia coli* agonists of the innate immune system. Veterinary research; 44: 40.
- Glas, A. (2008) :**
Vergleichende Untersuchung klinisch gesunder und mit *Escherichia coli* infizierter Euterviertel von Kühen mittels Infrarotthermographie. Dissertation, Ludwig-Maximilians-Universität München; https://edoc.ub.uni-muenchen.de/9081/1/Glas_Andrea.pdf
- Grunert, E. (1996)**
Euterödeme. In: Buiatrik Band I: Euterkrankheiten, Geburtshilfe und Gynäkologie, Andrologie und Besamung, 5., überarbeitete und erweiterte Auflage: 24-25; Verlag M. & H. Schaper, Alfeld-Hannover, ISBN: 9783794401819.
- Hamann, J., Duck, M. (1984):**
First results from measuring the temperature of teats using IR thermography. Milchpraxis; 22: 148-152.
- Hamann, J., Grunert, E. (1998):**
Ökonomie der Mastitisbekämpfung - eine diagnostische Frage? Praktischer Tierarzt, coll vet; 28: 62-70.
- Hovinen, M., Siivonen, J., Taponen, S., Hanninen, L., Pastell, M., Aisla, A. M., Pyörala, S. (2008):**
Detection of clinical mastitis with the help of a thermal camera. Journal of Dairy Science; 91: 4592-4598.
- Hurnik, J. F., Boer, S. D., Webster, A. B. (1984):**
Detection of health disorders in dairy cattle utilizing a thermal infrared scanning technique. Canadian Journal of Animal Science; 64: 1071-1073.

Jensen, K., Gunther, J., Talbot, R., Petzl, W., Zerbe, H., Schuberth, H. J., Seyfert, H. M., Glass, E. J. (2013):

Escherichia coli- and *Staphylococcus aureus*- induced mastitis differentially modulate transcriptional responses in neighbouring uninfected bovine mammary gland quarters. BMC Genomics; 14: 36.

Jiang, L. J., Ng, E. Y. K., Yeo, A. C. B., Wu, S., Pan, F., Yau, W. Y., Chen, J. H., Yang, Y. (2005):

A perspective on medical infrared imaging. Journal of Medical Engineering and Technology; 29: 257-267.

Joyce, J., Blaxter, K., Park, C. (1966):

The effect of natural outdoor environments on the energy requirements of sheep. Research in veterinary science; 7: 342-359.

Kaneene, J. B., Hurd, H. S. (1990):

The National Animal Health Monitoring-System in Michigan. 3. Cost Estimates of Selected Dairy-Cattle Diseases. Preventive Veterinary Medicine; 8: 127-140.

Kennedy, D. A., Lee, T., Seely, D. (2009):

A Comparative Review of Thermography as a Breast Cancer Screening Technique. Integrative Cancer Therapies; 8: 9-16.

Lahiri, B., Bagavathiappan, S., Jayakumar, T., Philip, J. (2012):

Medical applications of infrared thermography: a review. Infrared Physics & Technology; 55: 221-235.

Lawson, R. (1956):

Implications of Surface Temperatures in the Diagnosis of Breast Cancer. Canadian Medical Association Journal; 75: 309-310.

Lefcourt, A. M., Bitman, J., Wood, D. L., Stroud, B., Schultze, W. D. (1993):

Radiotelemetry temperature responses of mammary gland and body to intramammary injection of *Escherichia coli* endotoxin or *Streptococcus agalactiae* in lactating dairy cows. American journal of veterinary research; 54: 798-804.

Leimbach, A., Poehlein, A., Witten, A., Wellnitz, O., Shpigel, N., Petzl, W., Zerbe, H., Daniel, R., Dobrindt, U. (2016):

Whole-Genome Draft Sequences of Six Commensal Fecal and Six Mastitis-Associated *Escherichia coli* Strains of Bovine Origin. Genome Announcements; 4: e00753-00716.

Lind, M., Sipka, A. S., Schuberth, H. J., Blutke, A., Wanke, R., Sauter-Louis, C., Duda, K. A., Holst, O., Rainard, P., Germon, P., Zerbe, H., Petzl, W. (2015):

Location-specific expression of chemokines, TNF-alpha and S100 proteins in a teat explant model. Innate Immunity; 21: 322-331.

Metzner, M., Sauter-Louis, C., Seemueller, A., Petzl, W., Klee, W. (2014):

Infrared thermography of the udder surface of dairy cattle: Characteristics, methods, and correlation with rectal temperature. The Veterinary Journal; 199: 57-62.

- Metzner, M., Sauter-Louis, C., Seemueller, A., Petzl, W., Zerbe, H. (2015):**
Infrared thermography of the udder after experimentally induced *Escherichia coli* mastitis in cows. *The Veterinary Journal*; 204: 360-362.
- Mitchell, D., Wyndham, C. H., Hodgson, T. (1967a):**
Emissivity and transmittance of excised human skin in its thermal emission wave band. *Journal of applied physiology*; 23: 390-394.
- Mitchell, D., Wyndham, C. H., Hodgson, T., Nabarro, F. R. N. (1967b):**
Measurement of the Total Normal Emissivity of Skin Without The Need For Measuring Skin Temperature. *Physics in Medicine and Biology*; 12: 359.
- Mitterhuemer, S. (2009):**
Analyse des Transkriptoms von bovinem Milchdrüsengewebe nach experimenteller Infektion mit *E. coli* 1303. Dissertation, Ludwig-Maximilians-Universität München; <https://edoc.ub.uni-muenchen.de/10468/>.
- Mitterhuemer, S., Petzl, W., Krebs, S., Mehne, D., Klanner, A., Wolf, E., Zerbe, H., Blum, H. (2010):**
Escherichia coli infection induces distinct local and systemic transcriptome responses in the mammary gland. *BMC Genomics*; 11: 138.
- Mortola, J. P. (2013):**
Thermographic Analysis of Body Surface Temperature of Mammals. *Zoological Science*; 30: 118-124.
- Ng, E. Y. K. (2009):**
A review of thermography as promising non-invasive detection modality for breast tumor. *International Journal of Thermal Sciences*; 48: 849-859.
- Nickerson, S., Pankey, J. (1983):**
Cytologic observations of the bovine teat end. *American journal of veterinary research*; 44: 1433-1441.
- Okada, K., Takemura, K., Sato, S. (2013):**
Investigation of Various Essential Factors for Optimum Infrared Thermography. *The Journal of Veterinary Medical Science*; 75: 1349-1353.
- Paulrud, C. O., Clausen, S., Andersen, P. E., Rasmussen, M. D. (2005):**
Infrared thermography and ultrasonography to indirectly monitor the influence of liner type and overmilking on teat tissue recovery. *Acta veterinaria Scandinavica*; 46: 137-147.
- Petzl, W., Zerbe, H., Gunther, J., Yang, W., Seyfert, H. M., Nurnberg, G., Schuberth, H. J. (2008):**
Escherichia coli, but not *Staphylococcus aureus* triggers an early increased expression of factors contributing to the innate immune defense in the udder of the cow. *Veterinary Research*; 39: 18.
- Petzl, W., Günther, J., Mühlbauer, K., Seyfert, H.-M., Schuberth, H.-J., Hussen, J., Sauter-Louis, C., Hafner-Marx, A., Zerbe, H. (2016):**
Early transcriptional events in the udder and teat after intra-mammary *Escherichia coli* and *Staphylococcus aureus* challenge. *Innate immunity*; 22: 294-304.

Polat, B., Colak, A., Cengiz, M., Yanmaz, L., Oral, H., Bastan, A., Kaya, S., A., H. (2010):

Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. *Journal of Dairy Science*; 92: 3525-3532.

Rainwater-Lovett, K., Pacheco, J. M., Packer, C., Rodriguez, L. L. (2009):

Detection of foot-and-mouth disease virus infected cattle using infrared thermography. *The Veterinary Journal*; 180: 317-324.

Rinaldi, M., Li, R. W., Bannerman, D. D., Daniels, K. M., Evoke-Clover, C., Silva, M. V., Paape, M. J., Van Ryssen, B., Burvenich, C., Capuco, A. V. (2010):

A sentinel function for teat tissues in dairy cows: dominant innate immune response elements define early response to *E. coli* mastitis. *Functional & integrative genomics*; 10: 21-38.

Robert, B., White, B. J., Renter, D. G., Larson, R. L. (2009):

Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Computers and Electronics in Agriculture*; 67: 80-84.

Rollin, E., Dhuyvetter, K. C., Overton, M. W. (2015):

The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. *Preventive Veterinary Medicine*; 122: 257-264.

Sanchez-Carlo, V., McDonald, J. S., Packer, R. A. (1984):

Virulence factors of *Escherichia coli* isolated from cows with acute mastitis. *American journal of veterinary research*; 45: 1775-1777.

Schaefer, A. L., Cook, N., Tessaro, S. V., Deregt, D., Desroches, G., Dubeski, P. L., Tong, A. K. W., Godson, D. L. (2004):

Early detection and prediction of infection using infrared thermography. *Canadian Journal of Animal Science*; 84: 73-80.

Schaefer, A. L., Cook, N., Church, J. S., Basarab, J., Perry, B., Miller, C., Tong, A. K. (2007):

The use of infrared thermography as an early indicator of bovine respiratory disease complex in calves. *Research in Veterinary Science*; 83: 376-384.

Schröter, W. (2015):

Automatisierte Analyse von veterinären Wärmebildern. Master Thesis, Fakultät Elektrotechnik und Informatik, Hochschule Zittau/Görlitz.

Schukken, Y. H., Kremer, W. D., Lohuis, J. A. (1989):

Escherichia coli mastitis in cattle. I. Clinical diagnosis and epidemiological aspects. *Tijdschrift voor diergeneeskunde*; 114: 829-838 (Abstract).

Schukken, Y. H., Günther, J., Fitzpatrick, J., Fontaine, M., Goetze, L., Holst, O., Leigh, J., Petzl, W., Schuberth, H.-J., Sipka, A. (2011):

Host-response patterns of intramammary infections in dairy cows. *Veterinary immunology and immunopathology*; 144: 270-289.

Schutz, M. M., Eichser, S. D., Townsend, J. M., Shaw, G., Kocak, D. M. (2001):

Evaluation of early detection of induced *Staphylococcus aureus* mastitis using infrared thermography. *Journal of Dairy Science*; 84 (Supplement 1): 150-151.

Schwartzkopf-Genswein, K., Stookey, J. (1997):

The use of infrared thermography to assess inflammation associated with hot-iron and freeze branding in cattle. *Canadian Journal of Animal Science*; 77: 577-584.

Scott, S. L., Schaefer, A. L., Tong, A. K. W., Lacasse, P. (2000):

Use of infrared thermography for early detection of mastitis in dairy cows. *Canadian Journal of Animal Science*; 80: 764-765.

Steketee, J. (1973):

Spectral emissivity of skin and pericardium. *Physics in medicine and biology*; 18: 686.

Stocker, H., Rüschi, P. (1997):

Euter und Zitzen. In: *Textbook of Ultrasonography in Cattle*. Berlin, Parey: 163-175.

Suojala, L., Pohjanvirta, T., Simojoki, H., Myllyniemi, A.-L., Pitkälä, A., Pelkonen, S., Pyörälä, S. (2011):

Phylogeny, virulence factors and antimicrobial susceptibility of *Escherichia coli* isolated in clinical bovine mastitis. *Veterinary Microbiology*; 147: 383-388.

Talukder, S., Kerrisk, K. L., Ingenhoff, L., Thomson, P. C., Garcia, S. C., Celi, P. (2014):

Infrared technology for estrus detection and as a predictor of time of ovulation in dairy cows in a pasture-based system. *Theriogenology*; 81: 925-935.

Talukder, S., Thomson, P. C., Kerrisk, K. L., Clark, C. E. F., Celi, P. (2015):

Evaluation of infrared thermography body temperature and collar-mounted accelerometer and acoustic technology for predicting time of ovulation of cows in a pasture-based system. *Theriogenology*; 83: 739-748.

Van Soest, F. J., Santman-Berends, I. M., Lam, T. J., Hogeveen, H. (2016):

Failure and preventive costs of mastitis on Dutch dairy farms. *Journal of Dairy Science*; 99: 8365-8374.

Vázquez Diosdado, J. A., Barker, Z. E., Hodges, H. R., Amory, J. R., Croft, D. P., Bell, N. J., Codling, E. A. (2015):

Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*; 3: 15.

Von Engelhardt, W., Breves, G., Diener, M., Gäbel, G. (2015)

Wärmebilanz und Temperaturregulation. In: *Physiologie der Haustiere*, 2. Auflage, 448; Enke, Stuttgart; ISBN: 9783830412700.

Watmough, D. J., Fowler, P. W., Oliver, R. (1970):

The thermal scanning of a curved isothermal surface: Implications for clinical thermography. *Physics in Medicine and Biology*, 15: 1-8.

Wirthgen, T., Zipser, S., Franze, U., Geidel, S., Lempe, G. (2011a):

Automatic infrared based temperature measuring system for health monitoring in veterinary applications. In: *Conference proceedings; Sensors*, 2011 IEEE, 28-31 Oct. 2011: 1800-1803.

-
- Wirthgen, T., Zipser, S., Franze, U., Geidel, S., Dietel, F., Alary, T. (2011b):**
Automatic Segmentation of Veterinary Infrared Images with the Active Shape Approach. In: Image Analysis; Springer Berlin Heidelberg: 435-446. ISBN: 978-3-642-21226-0.
- Wirthgen, T., Lempe, G., Zipser, S., Grünhaupt, U. (2012):**
Level-Set Based Infrared Image Segmentation for Automatic Veterinary Health Monitoring. In: Computer Vision and Graphics, Springer Berlin Heidelberg: 685-693. ISBN: 978-3-642-33563-1.
- Yang, W., Zerbe, H., Petzl, W., Brunner, R. M., Günther, J., Draing, C., Von Aulock, S., Schuberth, H.-J., Seyfert, H.-M. (2008):**
Bovine TLR2 and TLR4 properly transduce signals from *Staphylococcus aureus* and *E. coli*, but *S. aureus* fails to both activate NF- κ B in mammary epithelial cells and to quickly induce TNF α and interleukin-8 (CXCL8) expression in the udder. *Molecular immunology*; 45: 1385-1397.
- Zaproudina, N., Varmavuo, V., Airaksinen, O., Narhi, M. (2008):**
Reproducibility of infrared thermography measurements in healthy individuals. *Physiological measurement*; 29: 515-524.
- Zhang, Z.-M., Wang, R.-L. (2010):**
Histopathological Observation on Cows Suffering from Clinical Mastitis. *Animal Husbandry and Feed Science* 2010: 2 (Abstract).

X. APPENDIX

1. Appendix 1: Tables of data obtained by automatic and manual evaluation

Table 27: Cow 1; mean temperature values (°C) throughout the trial, obtained by automatic and manual evaluation method. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

hours after challenge	rectal temperature (°C)	AUTOMATIC EVALUATION						MANUAL EVALUATION									
		left hindquarter			right hindquarter			left hindquarter			right hindquarter						
		Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg
-25	38.2	33.80	37.17	3.37	35.15	33.43	36.33	2.90	34.98	34.10	37.33	3.23	35.60	34.33	37.33	3.00	35.47
-24	38.2	32.53	36.70	4.17	34.31	31.83	36.77	4.93	34.47	33.33	37.37	4.03	35.20	33.23	37.43	4.20	34.87
-23	38.1	33.67	37.23	3.57	35.54	30.77	37.17	6.40	35.55	34.50	37.60	3.10	35.83	34.47	37.70	3.23	35.87
-22	38.1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
-21	38.1	33.43	36.63	3.20	34.99	33.43	36.53	3.10	35.07	34.00	37.47	3.47	35.50	33.97	37.27	3.30	35.57
-19	38.1	33.87	36.80	2.93	35.19	33.73	36.97	3.23	35.35	34.00	37.47	3.47	35.70	33.90	37.77	3.87	35.80
-17	38.2	33.20	37.33	4.13	35.17	31.67	37.07	5.40	34.81	33.80	37.60	3.80	35.80	33.67	37.37	3.70	35.33
-15	38.1	31.73	37.13	5.40	34.35	28.27	36.50	8.23	34.54	34.03	37.40	3.37	35.20	33.00	36.70	3.70	34.90
-13	38.2	29.37	36.93	7.57	35.05	29.70	36.83	7.13	34.70	28.97	37.57	8.60	35.43	28.57	37.37	8.80	35.13
-12.5	38.3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
-11	38.4	34.27	37.30	3.03	35.85	34.03	36.87	2.83	35.86	35.07	37.70	2.63	36.40	34.37	37.33	2.97	36.27
-9	39.1	35.67	38.00	2.33	36.84	35.67	38.00	2.33	36.84	35.77	38.43	2.67	37.27	35.73	38.60	2.87	37.23
-7	38.8	34.23	37.47	3.23	35.70	33.83	37.23	3.40	35.42	34.63	38.00	3.37	36.27	34.47	38.03	3.57	36.00
-5	38.8	34.23	37.37	3.13	35.64	33.07	36.87	3.80	35.51	33.93	37.73	3.80	36.20	34.17	37.77	3.60	36.07
-3	38.8	33.63	37.80	4.17	35.88	33.60	36.77	3.17	35.48	33.67	38.27	4.60	36.37	33.77	38.00	4.23	36.07
-1	39.1	33.37	37.33	3.97	35.30	31.27	37.00	5.73	34.93	33.17	37.67	4.50	35.87	33.43	37.83	4.43	35.67
0	38.6	32.97	36.63	3.67	34.65	30.77	36.20	5.43	34.26	33.40	37.47	4.07	35.30	32.80	37.43	4.63	35.03
1	38.2	34.17	37.03	2.87	35.75	34.07	37.10	3.03	35.47	34.33	37.87	3.53	36.30	34.43	37.90	3.47	36.20
2	38.2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
3	38.3	34.37	37.23	2.87	35.85	33.90	36.97	3.07	35.74	34.97	37.60	2.63	36.43	34.87	37.77	2.90	36.20
5	38.3	33.27	37.47	4.20	35.14	31.97	37.00	5.03	35.37	34.60	37.63	3.03	36.03	33.67	37.43	3.77	35.83
7	38.7	34.20	37.27	3.07	35.82	32.83	37.03	4.20	35.63	34.03	37.73	3.70	36.27	34.23	38.17	3.93	36.17
9	38.2	33.47	37.03	3.57	35.58	33.60	36.53	2.93	35.24	34.10	37.40	3.30	36.03	34.07	37.33	3.27	35.77
11	39.3	34.23	38.83	4.60	36.67	34.00	37.77	3.77	35.56	33.57	38.93	5.37	36.87	34.57	38.70	4.13	36.03
11.5	40.7	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
13	41.2	36.83	39.87	3.03	38.38	35.70	39.00	3.30	37.14	36.43	40.07	3.63	38.30	36.17	40.03	3.87	37.50
15	40.8	36.90	39.97	3.07	38.90	35.83	39.83	4.00	38.32	36.10	40.43	4.33	38.97	36.20	40.20	4.00	38.50
17	39	34.70	37.93	3.23	36.89	34.30	37.53	3.23	36.43	34.67	38.37	3.70	37.17	34.40	38.40	4.00	36.73
19	38.4	35.10	37.80	2.70	36.48	34.03	37.13	3.10	36.01	35.40	38.33	2.93	36.87	33.87	37.97	4.10	36.37
21	38.6	34.57	37.40	2.83	36.25	32.27	37.10	4.83	36.11	34.97	37.77	2.80	36.53	33.47	37.80	4.33	36.43
23	39.6	32.70	38.27	5.57	36.21	33.10	37.97	4.87	36.24	32.80	38.77	5.97	36.60	32.97	39.07	6.10	36.67
24	39.2	33.80	38.00	4.20	36.67	34.40	38.00	3.60	36.85	35.37	38.47	3.10	37.03	33.50	38.77	5.27	37.13

reference period

period after challenge

Table 28: Cow 2; mean temperature values (°C) throughout the trial, obtained by automatic and manual evaluation method. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

hours after challenge	rectal temperature (°C)	AUTOMATIC EVALUATION					MANUAL EVALUATION												
		left hindquarter		right hindquarter		Avg	left hindquarter		right hindquarter		Avg								
		Min	Max	Min	Max		Max-Min	Min	Max	Min		Max	Max-Min						
-25	38.7	30.27	36.07	5.80	34.19	34.19	31.57	36.13	4.57	34.14	28.60	36.97	8.37	35.03	30.37	37.07	6.70	35.03	
-24	38.5	28.73	36.23	7.50	34.40	34.40	30.13	35.83	5.70	34.18	28.67	37.27	8.60	35.20	28.47	37.30	8.83	35.10	
-23	38.3	30.07	35.53	5.47	33.79	33.79	31.27	35.57	4.30	33.78	31.33	36.70	5.37	34.70	31.17	36.40	5.23	34.63	
-22	38	31.13	35.30	4.17	33.83	33.83	30.63	35.63	5.00	33.95	32.20	36.70	4.50	34.77	31.70	36.83	5.13	34.93	
-21	38.5	30.27	36.60	6.33	34.33	34.33	29.60	36.10	6.50	34.11	31.07	37.27	6.20	35.17	30.13	37.10	6.97	35.07	
-19	38.9	31.43	36.80	5.37	34.88	34.88	32.20	36.70	4.50	34.98	32.47	37.77	5.30	35.70	31.97	37.47	5.50	35.73	
-17	38.9	31.60	36.87	5.27	35.09	35.09	32.07	36.67	4.60	35.18	31.20	37.80	6.60	35.87	32.00	37.47	5.47	35.90	
-15	38.6	32.37	36.53	4.17	34.97	34.97	32.90	36.33	3.43	34.93	33.57	37.50	3.93	35.70	33.23	37.53	4.30	35.70	
-13	38.4	33.37	36.97	3.60	35.25	35.25	32.97	36.63	3.67	35.17	33.40	37.47	4.07	35.90	32.77	37.33	4.57	35.77	
-12.5	38.3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
-11	39	33.40	37.17	3.77	35.67	35.67	33.10	37.50	4.40	35.84	33.97	37.90	3.93	36.33	32.97	38.03	5.07	36.47	
-9	38.6	32.80	36.77	3.97	35.35	35.35	31.87	36.70	4.83	35.17	33.53	37.63	4.10	36.00	31.90	37.73	5.83	35.93	
-7	38.8	32.63	36.43	3.80	34.69	34.69	31.27	36.10	4.83	34.50	32.80	37.13	4.33	35.43	31.70	37.17	5.47	35.33	
-5	38.6	30.60	37.13	6.53	34.97	34.97	31.70	36.60	4.90	34.93	32.10	37.83	5.73	35.63	31.87	37.83	5.97	35.67	
-3	39.1	32.97	37.20	4.23	35.17	35.17	33.30	36.97	3.67	35.07	33.20	37.83	4.63	35.83	32.60	37.93	5.33	35.80	
-1	39.1	31.90	36.70	4.80	34.78	34.78	32.40	36.53	4.13	34.93	34.27	37.40	3.13	35.50	33.93	37.50	3.57	35.70	
0	38.8	31.47	36.57	5.10	34.44	34.44	32.87	37.20	4.33	34.43	32.77	37.13	4.37	35.17	33.50	37.83	4.33	35.17	
1	38.6	32.60	37.10	4.50	35.12	35.12	32.80	37.17	4.37	34.89	33.63	37.93	4.30	35.87	33.93	37.57	3.63	35.63	
2	38.6	31.23	36.43	5.20	34.48	34.48	32.10	36.10	4.00	34.49	32.83	37.20	4.37	35.27	32.20	37.03	4.83	35.20	
3	38.4	32.37	36.50	4.13	34.98	34.98	33.20	36.67	3.47	34.89	34.13	37.57	3.43	35.67	33.63	37.37	3.73	35.53	
5	38.8	34.53	37.33	2.80	35.81	35.81	34.47	37.27	2.80	35.76	33.63	38.30	4.67	36.40	33.97	38.17	4.20	36.37	
7	38.9	33.10	36.93	3.83	35.18	35.18	33.03	36.87	3.83	35.40	33.90	37.77	3.87	35.80	33.37	37.73	4.37	36.00	
9	38.7	32.10	37.30	5.20	35.30	35.30	33.63	36.83	3.20	35.52	33.67	37.73	4.07	35.83	33.47	37.83	4.37	36.07	
11	39.6	32.13	37.60	5.47	35.64	35.64	33.47	37.40	3.93	35.97	33.60	38.20	4.60	36.13	33.60	38.13	4.53	36.43	
11.5	40.1	33.70	38.70	5.00	36.73	36.73	34.97	38.40	3.43	36.91	35.33	38.80	3.47	37.07	35.57	38.70	3.13	37.33	
13	41.5	32.93	39.87	6.93	37.48	37.48	34.30	39.07	4.77	36.62	32.73	40.00	7.27	37.67	33.83	39.97	6.13	37.13	
15	40.8	35.33	39.07	3.73	36.87	36.87	33.27	38.30	5.03	35.84	35.07	39.50	4.43	37.33	34.07	39.50	5.43	36.63	
17	39.3	33.60	37.57	3.97	35.81	35.81	33.93	37.23	3.30	35.29	34.80	37.93	3.13	36.33	33.50	38.37	4.87	35.57	
19	38.4	32.50	36.23	3.73	34.41	34.41	32.83	35.80	2.97	34.46	31.40	36.70	5.30	35.03	32.30	36.83	4.53	35.20	
21	38.4	33.30	36.40	3.10	35.21	35.21	33.90	36.23	2.33	34.98	31.60	37.03	5.43	35.80	33.60	37.17	3.57	35.70	
23	38.5	32.10	35.83	3.73	34.56	34.56	31.37	36.07	4.70	34.21	32.07	36.43	4.37	35.30	32.67	37.00	4.33	35.03	
24	38.6	30.80	35.47	4.67	34.30	34.30	29.43	35.97	6.53	34.11	30.90	36.03	5.13	35.10	30.73	36.53	5.80	34.77	

reference period

period after challenge

Table 30: Cow 4; mean temperature values (°C) throughout the trial, obtained by automatic and manual evaluation method. 'Min' – Minimum temperature values; 'Max' – Maximum temperature values; 'Max-Min' – Range of temperature values; 'Avg' – Average temperature values.

hours after temperature challenge	rectal temperature (°C)	AUTOMATIC EVALUATION						MANUAL EVALUATION									
		left hindquarter			right hindquarter			left hindquarter			right hindquarter						
		Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg
-25	38.5	31.53	36.63	5.10	34.76	30.10	35.90	5.80	34.28	32.83	37.30	4.47	35.53	31.47	37.10	5.63	35.13
-24	37.8	31.57	37.17	5.60	34.90	31.73	36.47	4.73	34.25	33.70	37.63	3.93	35.63	32.77	37.90	5.13	35.07
-23	37.8	32.07	36.37	4.30	34.81	31.90	36.50	4.60	34.48	33.20	37.07	3.87	35.57	33.07	37.50	4.43	35.30
-22	37.9	29.90	36.47	6.57	34.74	31.07	35.77	4.70	34.25	30.63	37.13	6.50	35.50	28.87	36.93	8.07	35.00
-21	37.9	30.50	37.23	6.73	35.06	32.93	36.83	3.90	35.09	31.80	37.70	5.90	35.97	30.77	37.70	6.93	35.60
-19	38.1	32.33	36.47	4.13	35.15	31.40	36.13	4.73	34.78	32.70	37.20	4.50	35.80	32.47	37.07	4.60	35.50
-17	38.2	32.10	36.20	4.10	34.79	30.97	35.93	4.97	34.56	32.33	36.77	4.43	35.43	31.53	36.80	5.27	35.23
-15	38	30.93	36.50	5.57	34.54	31.50	35.57	4.07	33.89	30.57	36.97	6.40	35.17	26.57	37.00	10.43	34.67
-13	38.3	31.27	36.97	5.70	35.01	31.37	36.27	4.90	34.28	31.23	37.57	6.33	35.63	30.83	37.40	6.57	35.00
-12.5	38.5	31.03	36.17	5.13	34.17	29.87	35.87	6.00	33.37	30.77	36.60	5.83	34.77	31.03	36.50	5.47	34.20
-11	37.9	32.00	36.60	4.60	34.61	29.97	36.20	6.23	34.25	32.97	37.23	4.27	35.37	32.60	37.07	4.47	35.00
-9	38.1	32.57	36.87	4.30	35.60	32.73	36.50	3.77	35.26	32.87	37.67	4.80	36.20	32.43	37.40	4.97	35.90
-7	38.3	30.90	36.27	5.37	35.04	30.83	35.97	5.13	34.81	31.50	37.17	5.67	35.77	30.93	36.77	5.83	35.50
-5	38	30.50	36.43	5.93	34.72	29.73	36.00	6.27	34.43	29.60	37.13	7.53	35.43	33.30	37.00	3.70	35.23
-3	38	32.07	36.63	4.57	35.16	33.00	36.60	3.60	34.84	29.43	37.20	7.77	35.73	33.87	37.10	3.23	35.50
-1	38.3	31.47	37.00	5.53	35.48	33.33	36.60	3.27	35.31	32.80	37.77	4.97	36.13	28.40	37.60	9.20	35.90
0	38.5	33.50	36.70	3.20	35.22	33.63	36.27	2.63	34.87	32.90	37.33	4.43	35.93	33.50	37.43	3.93	35.60
1	38.3	32.73	36.67	3.93	35.14	32.13	36.47	4.33	34.97	33.17	37.17	4.00	35.80	32.83	37.17	4.33	35.70
2	38.1	32.63	36.77	4.13	35.26	32.17	36.40	4.23	35.03	33.33	37.23	3.90	35.93	33.03	37.13	4.10	35.70
3	38.1	32.13	36.07	3.93	34.51	31.73	35.70	3.97	34.35	32.50	36.63	4.13	35.17	31.30	36.83	5.53	35.03
5	38.5	28.87	36.20	7.33	33.70	29.43	35.80	6.37	33.33	26.43	36.80	10.37	34.47	26.47	37.03	10.57	34.20
7	38.4	31.40	36.63	5.23	34.87	31.87	35.97	4.10	34.60	31.47	37.17	5.70	35.50	31.70	37.17	5.47	35.40
9	39.1	32.20	36.80	4.60	35.10	32.13	36.53	4.40	34.80	32.30	37.43	5.13	35.73	30.70	37.60	6.90	35.47
11	41.6	30.87	37.40	6.53	35.07	28.87	37.27	8.40	34.12	29.37	37.73	8.37	35.83	31.40	38.13	6.73	35.03
11.5	41.6	31.33	38.03	6.70	35.31	29.53	38.47	8.93	34.61	33.57	38.17	4.60	36.27	32.10	38.93	6.83	35.47
13	42.2	28.43	38.90	10.47	36.07	30.67	38.53	7.87	34.40	30.40	39.13	8.73	37.07	32.70	39.70	7.00	35.43
15	41.6	33.23	39.40	6.17	36.59	30.73	38.23	7.50	34.05	32.90	39.53	6.63	37.00	32.97	39.57	6.60	35.40
17	39.4	28.40	38.60	10.20	36.55	32.73	37.97	5.23	36.01	32.10	39.03	6.93	37.33	35.23	39.60	4.37	36.70
19	38.8	33.03	37.77	4.73	36.57	33.97	37.20	3.23	36.12	32.63	38.60	5.97	36.93	34.60	38.83	4.23	36.73
21	38.6	29.83	37.13	7.30	35.77	32.87	36.60	3.73	35.16	28.70	37.57	8.87	36.23	32.47	37.87	5.40	35.87
23	38.9	29.20	37.07	7.87	35.36	32.60	37.00	4.40	35.38	23.27	37.37	14.10	36.03	33.50	38.60	5.10	36.13
24	39	30.33	36.53	6.20	34.99	29.93	36.70	6.77	34.91	33.03	37.17	4.13	35.90	33.43	38.27	4.83	35.77

period after challenge

reference period

Table 31: Cow 5; mean temperature values (°C) throughout the trial, obtained by automatic and manual evaluation method. 'Min' – Minimum temperature values; 'Max' – Maximum temperature values; 'Max-Min' – Range of temperature values; 'Avg' – Average temperature values.

hours after challenge	rectal temperature (°C)	AUTOMATIC EVALUATION						MANUAL EVALUATION									
		left hindquarter			right hindquarter			left hindquarter			right hindquarter						
		Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg				
-25	38.3	32.70	35.27	2.57	33.86	30.77	34.97	4.20	33.31	32.83	36.63	3.80	34.67	31.53	37.07	5.53	34.30
-24	38.2	29.73	36.03	6.30	33.67	31.13	35.97	4.83	34.02	31.63	37.10	5.47	34.70	31.57	36.97	5.40	34.80
-23	38.2	31.50	35.83	4.33	33.84	30.97	35.67	4.70	34.20	32.17	36.47	4.30	34.53	32.77	36.97	4.20	35.00
-22	38.3	30.97	35.93	4.97	33.81	30.67	35.20	4.53	33.66	32.40	36.93	4.53	34.50	32.63	36.80	4.17	34.60
-21	38.4	31.80	36.70	4.90	34.54	31.73	36.03	4.30	34.19	32.93	37.43	4.50	35.20	32.00	37.23	5.23	35.10
-19	38.3	32.60	35.87	3.27	34.27	31.93	34.83	2.90	33.57	32.83	36.93	4.10	35.00	32.57	36.77	4.20	34.60
-17	38.2	32.07	35.87	3.80	34.13	31.27	35.83	4.57	34.20	32.30	36.53	4.23	34.73	32.43	36.90	4.47	34.87
-15	38.2	32.87	36.40	3.53	34.56	32.13	34.97	2.83	33.57	32.30	37.10	4.80	35.37	32.10	36.90	4.80	34.60
-13	38.3	27.80	35.93	8.13	34.05	31.83	35.30	3.47	33.86	32.97	36.70	3.73	35.13	32.97	36.50	3.53	34.80
-12.5	38.3	29.33	36.17	6.83	34.27	28.90	35.53	6.63	33.89	29.73	37.00	7.27	34.83	29.60	37.10	7.50	34.90
-11	38.4	28.73	36.50	7.77	34.43	32.80	36.13	3.33	34.61	32.37	37.10	4.73	35.37	33.50	37.50	4.00	35.50
-9	38.4	28.40	36.17	7.77	34.18	31.33	35.50	4.17	34.22	32.97	36.87	3.90	35.17	32.53	37.37	4.83	35.10
-7	38.4	31.80	35.77	3.97	34.15	30.33	35.27	4.93	33.79	33.03	37.17	4.13	34.87	30.37	37.23	6.87	34.77
-5	38.4	31.93	36.03	4.10	34.15	32.30	35.87	3.57	34.56	32.53	36.50	3.97	34.83	32.73	36.63	3.90	35.23
-3	38.4	29.33	35.77	6.43	34.00	32.17	35.50	3.33	33.99	32.87	36.97	4.10	34.80	32.20	37.10	4.90	34.77
-1	38.9	32.63	36.07	3.43	34.52	31.97	35.93	3.97	34.44	33.20	37.40	4.20	35.30	32.07	37.17	5.10	35.30
0	38.8	31.23	36.27	5.03	34.08	31.93	36.00	4.07	34.28	32.27	37.73	5.47	35.27	31.57	37.17	5.60	35.33
1	38.3	30.77	35.53	4.77	33.37	29.87	35.63	5.77	33.36	31.70	36.73	5.03	34.33	31.83	36.27	4.43	34.30
2	38.2	29.93	36.47	6.53	33.95	29.87	36.13	6.27	34.24	33.03	37.87	4.83	35.57	33.23	37.00	3.77	35.20
3	38.2	31.67	35.97	4.30	34.21	32.90	36.23	3.33	34.56	32.20	37.00	4.80	34.90	33.13	37.33	4.20	35.33
5	38.2	31.17	36.20	5.03	33.79	31.07	35.50	4.43	33.06	24.93	36.83	11.90	34.23	23.63	36.57	12.93	34.00
7	38.2	30.10	35.43	5.33	33.33	31.43	35.07	3.63	33.32	31.93	36.10	4.17	34.23	31.57	37.13	5.57	34.23
9	38.3	27.53	36.30	8.77	33.95	31.93	35.90	3.97	34.17	32.13	36.90	4.77	34.67	32.40	37.47	5.07	35.00
11	39	29.97	35.97	6.00	34.18	31.43	35.67	4.23	33.83	30.07	36.83	6.77	35.20	31.70	37.53	5.83	34.77
11.5	39.4	28.80	35.20	6.40	32.62	28.43	36.27	7.83	32.61	29.53	36.17	6.63	33.90	29.93	37.70	7.77	33.63
13	40.8	31.77	36.17	4.40	34.09	32.20	37.43	5.23	34.27	26.13	38.23	12.10	35.23	30.70	38.93	8.23	35.00
15	40.6	31.73	36.80	5.07	35.01	32.03	37.63	5.60	35.30	32.57	38.13	5.57	35.77	33.00	39.10	6.10	35.93
17	39.4	32.17	36.27	4.10	34.74	31.23	36.70	5.47	34.58	32.10	37.67	5.57	35.63	31.70	37.80	6.10	35.30
19	38.8	32.17	36.63	4.47	34.91	31.63	35.70	4.07	34.41	33.13	37.80	4.67	35.70	33.13	37.80	4.63	35.27
21	38.6	32.30	35.77	3.47	34.18	29.77	35.73	5.97	33.89	32.83	36.53	3.70	34.93	32.57	36.83	4.27	34.77
23	38.5	31.07	35.87	4.80	33.90	31.53	35.60	4.07	33.38	30.07	36.37	6.30	34.53	30.53	37.17	6.63	34.27
24	38.3	28.13	35.43	7.30	33.68	30.93	36.43	5.50	34.19	31.50	37.30	5.80	34.70	32.90	37.37	4.47	34.83

reference period

period after challenge

Table 32: Median temperature values of all five cows (°C) throughout the trial, obtained by automatic and manual evaluation method. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

recal hours after temperature challenge (°C)	AUTOMATIC EVALUATION						MANUAL EVALUATION					
	left hindquarter			right hindquarter			left hindquarter			right hindquarter		
	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg
-25	32.11	36.19	4.08	34.28	31.56	35.80	4.24	34.10	32.24	37.01	4.77	35.08
-24	30.34	36.21	5.87	33.99	31.07	36.06	4.99	33.96	30.93	37.15	6.22	34.87
-23	31.53	36.10	4.57	34.27	31.23	36.07	4.84	34.32	32.80	36.88	4.08	34.97
-22	*	*	*	*	*	*	*	*	*	*	*	*
-21	31.91	36.69	4.79	34.81	32.30	36.47	4.17	34.78	32.68	37.50	4.82	35.55
-19	32.43	36.37	3.95	34.78	32.12	36.13	4.01	34.67	32.82	37.28	4.46	35.51
-17	32.08	36.57	4.49	34.71	31.13	36.23	5.10	34.56	32.07	37.20	5.14	35.43
-15	32.03	36.45	4.43	34.49	31.45	35.83	4.39	34.25	32.59	37.19	4.60	35.28
-13	30.79	36.52	5.73	34.65	31.59	36.31	4.71	34.51	31.71	37.33	5.63	35.41
-12.5	*	*	*	*	*	*	*	*	*	*	*	*
-11	32.29	36.81	4.52	35.08	32.60	36.76	4.16	35.16	33.56	37.55	3.99	35.83
-9	32.42	36.89	4.47	35.41	32.73	36.59	3.86	35.27	33.37	37.60	4.23	36.09
-7	32.47	36.40	3.93	34.78	31.47	36.14	4.67	34.62	32.88	37.35	4.47	35.52
-5	31.89	36.65	4.76	34.85	31.62	36.39	4.77	34.90	31.97	37.36	5.39	35.55
-3	32.05	36.80	4.75	34.89	32.83	36.40	3.57	34.68	32.34	37.50	5.16	35.57
-1	32.39	36.58	4.19	34.80	31.99	36.34	4.35	34.71	33.15	37.44	4.29	35.54
0	32.33	36.30	3.97	34.45	32.27	36.33	4.06	34.38	32.75	37.40	4.66	35.32
1	32.48	36.36	3.88	34.57	32.14	36.29	4.15	34.40	33.09	37.31	4.22	35.35
2	*	*	*	*	*	*	*	*	*	*	*	*
3	32.67	36.40	3.73	34.76	32.67	36.21	3.53	34.73	33.23	37.24	4.01	35.48
5	32.17	36.74	4.57	34.61	31.75	36.39	4.64	34.43	30.62	37.44	6.82	35.33
7	32.04	36.48	4.44	34.59	32.08	36.05	3.97	34.58	32.41	37.19	4.78	35.31
9	31.63	36.79	5.16	34.89	32.61	36.55	3.94	34.90	33.07	37.47	4.40	35.55
11	31.83	37.62	5.79	35.24	32.01	37.32	5.31	34.97	31.79	38.18	6.39	35.94
11.5	*	*	*	*	*	*	*	*	*	*	*	*
13	31.96	38.75	6.79	36.40	32.60	38.71	6.11	35.72	31.93	39.49	7.56	37.05
15	34.11	38.75	4.65	36.64	33.19	38.46	5.27	35.94	33.91	39.53	5.62	37.15
17	32.29	37.58	5.29	36.04	32.83	37.38	4.55	35.70	33.25	38.35	5.10	36.67
19	33.21	36.93	3.71	35.42	32.75	36.37	3.62	35.19	33.31	37.73	4.41	36.01
21	31.97	36.68	4.71	35.27	32.15	36.50	4.35	35.12	31.66	37.24	5.58	35.82
23	31.25	36.73	5.47	34.97	31.99	36.66	4.67	34.92	30.07	37.24	7.16	35.60
24	30.83	36.31	5.47	34.84	31.57	36.69	5.12	35.02	32.55	37.13	4.57	35.61

reference period

period after challenge

2. Appendix 2: Tables of data obtained by evaluation using the teats and evaluation using the udder

Table 33: Cow 1, mean temperature values (°C) throughout the trial, obtained by evaluation using the teats and evaluation using the udder. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

hours after challenge	rectal temperature (°C)	EVALUATION USING THE TEATS						EVALUATION USING THE UDDER									
		left hindquarter			right hindquarter			left hindquarter			right hindquarter						
		Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg
-25	38.2	34.50	36.03	1.53	35.30	33.90	35.20	1.33	34.90	34.90	37.73	2.83	35.93	34.53	37.77	3.27	35.87
-24	38.2	33.30	35.50	2.20	34.80	33.53	35.17	1.63	34.33	33.93	37.73	3.83	35.53	33.80	37.83	4.03	35.30
-23	38.1	34.57	36.43	1.87	35.63	34.30	36.63	2.33	35.40	34.90	37.77	2.83	36.20	35.07	37.97	2.90	36.23
-22	38.1	*	*	*	*	*	*	*	*	34.50	37.70	3.20	36.03	34.57	37.70	3.13	36.03
-21	38.1	34.60	36.47	1.87	35.63	34.30	36.20	1.90	35.13	34.30	37.73	3.43	35.80	34.27	37.57	3.30	35.90
-19	38.1	34.07	36.50	2.43	35.67	34.27	36.53	2.27	35.27	34.77	37.80	3.07	36.00	34.33	37.93	3.53	36.13
-17	38.2	34.27	36.07	1.80	35.20	33.63	35.93	2.30	34.67	34.10	37.93	3.87	36.03	34.27	37.77	3.53	35.70
-15	38.1	33.83	35.53	1.73	34.80	33.20	34.80	1.60	33.83	34.30	37.43	3.13	35.47	34.00	37.07	3.07	35.27
-13	38.2	29.10	35.93	6.83	32.20	28.70	36.10	7.40	32.07	34.63	37.70	3.13	35.83	34.23	37.63	3.33	35.53
-12.5	38.3	*	*	*	*	*	*	*	*	34.33	37.47	3.13	35.57	33.97	37.33	3.37	35.23
-11	38.4	35.20	36.50	1.30	35.87	34.67	36.37	1.70	35.33	35.57	38.00	2.43	36.70	35.60	37.67	2.10	36.60
-9	39.1	35.97	37.17	1.20	36.70	35.87	36.97	1.10	36.47	36.33	38.77	2.40	37.57	36.30	38.87	2.60	37.53
-7	38.8	34.83	36.07	1.23	35.53	34.80	35.90	1.10	35.27	35.57	38.23	2.67	36.53	35.40	38.10	2.70	36.30
-5	38.8	34.03	35.80	1.77	35.10	34.07	35.57	1.53	34.87	35.43	38.07	2.63	36.50	35.43	38.00	2.57	36.40
-3	38.8	33.63	36.00	2.33	34.97	33.63	35.57	1.97	34.73	35.77	38.33	2.57	36.67	35.50	38.20	2.77	36.37
-1	39.1	33.07	35.27	2.20	34.43	31.83	34.97	3.13	34.10	34.87	38.03	3.13	36.23	34.63	37.93	3.33	36.03
0	38.6	33.13	35.97	2.83	34.53	33.03	35.53	2.50	34.20	34.57	37.40	2.83	35.63	34.23	37.57	3.33	35.37
1	38.2	34.13	36.37	2.23	35.47	34.30	36.07	1.77	35.17	35.40	38.00	2.57	36.63	35.47	38.07	2.60	36.50
2	38.2	*	*	*	*	*	*	*	*	35.57	37.93	2.37	36.67	35.53	38.00	2.47	36.53
3	38.3	35.23	36.60	1.37	35.97	34.87	36.10	1.20	35.47	35.70	37.93	2.30	36.77	35.57	37.97	2.37	36.57
5	38.3	32.00	35.97	3.97	35.40	33.93	35.77	1.83	35.13	35.40	37.83	2.40	36.30	35.30	37.50	2.23	36.07
7	38.7	34.00	36.10	2.10	35.43	34.53	36.07	1.53	35.27	35.63	37.93	2.30	36.57	35.47	38.20	2.77	36.47
9	38.2	34.20	36.00	1.80	35.30	34.17	35.33	1.17	34.67	35.27	37.57	2.30	36.27	34.83	37.63	2.73	35.97
11	39.3	33.87	36.77	2.90	36.00	34.87	36.43	1.57	35.80	35.33	39.20	3.87	37.10	34.80	39.20	4.43	36.37
11.5	40.7	*	*	*	*	*	*	*	*	34.97	38.83	3.87	36.93	34.53	38.93	4.33	35.93
13	41.2	36.67	38.97	2.30	38.07	36.43	37.90	1.47	37.17	37.07	40.43	3.33	38.63	36.47	39.83	3.37	37.80
15	40.8	36.47	38.50	2.03	37.77	36.33	37.80	1.47	37.33	38.17	40.63	2.47	39.30	36.60	40.27	3.63	38.77
17	39	34.77	36.90	2.13	36.00	34.87	36.50	1.63	35.97	36.53	38.73	2.23	37.60	34.80	38.70	3.87	37.13
19	38.4	35.67	37.07	1.40	36.47	34.40	35.83	1.43	35.30	35.87	38.43	2.60	37.17	34.97	38.13	3.13	36.67
21	38.6	35.03	36.20	1.17	35.57	33.70	35.60	1.90	35.03	35.53	38.13	2.63	36.87	35.00	37.93	2.93	36.73
23	39.6	33.00	36.33	3.33	34.57	33.37	36.03	2.67	34.90	35.47	38.97	3.50	37.00	35.10	38.80	3.70	37.03
24	39.2	35.00	37.63	2.63	36.67	33.07	37.53	4.47	35.97	35.70	38.83	3.13	37.43	35.53	39.00	3.47	37.47

period after challenge

Table 34: Cow 2, mean temperature values (°C) throughout the trial, obtained by evaluation using the teats and evaluation using the udder. 'Min' – Minimum temperature values; 'Max' – Maximum temperature values; 'Max-Min' – Range of temperature values; 'Avg' – Average temperature values.

hours after challenge	rectal temperature (°C)	EVALUATION USING THE TEATS										EVALUATION USING THE UDDER									
		left hindquarter					right hindquarter					left hindquarter					right hindquarter				
		Min	Max	Max-Min	Avg		Min	Max	Max-Min	Avg		Min	Max	Max-Min	Avg		Min	Max	Max-Min	Avg	
-25	38.5	28.77	35.07	6.30	33.40	30.37	35.40	5.03	34.13		34.20	37.47	3.23	35.47		33.87	37.47	3.60	35.37		
-24	38.3	28.23	35.50	7.27	31.87	28.00	35.27	7.27	32.77		34.27	37.63	3.37	35.57		33.67	37.60	3.87	35.40		
-23	38	31.10	35.07	3.97	33.00	31.03	34.77	3.73	33.60		33.80	37.07	3.23	34.97		33.47	37.07	3.50	34.93		
-22	38.5	32.17	34.80	2.63	33.10	31.83	34.90	3.07	33.80		34.00	37.03	3.00	35.10		34.00	37.17	3.13	35.20		
-21	38.9	31.00	34.23	3.23	32.27	29.87	35.10	5.23	33.17		34.10	37.60	3.53	35.53		33.93	37.67	3.73	35.40		
-19	38.9	32.53	35.90	3.37	33.97	31.93	35.80	3.87	34.40		34.70	38.03	3.33	36.00		34.77	38.07	3.30	36.10		
-17	38.6	32.10	36.17	4.07	34.47	32.43	36.00	3.57	34.83		35.00	38.07	3.07	36.13		35.07	38.10	3.03	36.27		
-15	38.4	33.37	35.90	2.53	34.63	33.47	36.17	2.70	35.33		35.07	37.80	2.73	35.90		34.00	37.83	3.80	35.97		
-13	38.3	33.57	36.07	2.50	34.87	32.97	36.40	3.43	35.40		34.93	37.73	2.77	36.20		35.03	37.80	2.70	36.10		
-12.5	39	*	*	*	*	*	*	*	*		35.07	37.97	2.90	36.30		35.13	37.97	2.83	36.23		
-11	38.6	33.97	37.13	3.17	35.53	33.37	37.27	3.90	35.77		35.57	38.23	2.63	36.63		35.67	38.33	2.67	36.77		
-9	38.8	33.13	36.07	2.93	34.47	32.50	35.97	3.47	34.90		35.27	38.00	2.73	36.30		35.10	37.97	2.87	36.23		
-7	38.6	32.77	34.97	2.20	33.93	32.60	35.57	2.97	34.50		34.70	37.50	2.80	35.73		34.43	37.50	3.07	35.63		
-5	39.1	32.27	35.90	3.63	34.23	31.93	36.17	4.23	34.73		34.80	38.20	3.40	36.00		34.60	38.13	3.53	35.97		
-3	39.1	33.40	36.10	2.70	34.70	32.63	36.27	3.63	35.17		34.70	38.23	3.53	36.17		34.80	38.27	3.50	36.10		
-1	38.8	34.33	35.87	1.53	35.23	34.67	36.37	1.70	35.67		34.57	37.77	3.20	35.90		34.30	37.90	3.67	36.00		
0	38.6	32.83	34.67	1.83	33.87	33.57	35.57	2.00	34.67		34.33	37.83	3.50	35.57		33.93	37.63	3.70	35.47		
1	38.6	34.17	36.53	2.37	35.07	33.87	36.17	2.30	35.27		35.00	38.13	3.13	36.17		34.80	37.90	3.13	36.00		
2	38.4	33.13	35.60	2.47	34.00	32.17	35.73	3.57	34.23		34.30	37.53	3.20	35.57		33.97	37.50	3.53	35.60		
3	38.8	34.00	35.67	1.67	34.90	33.63	36.17	2.53	35.07		34.80	37.83	3.03	35.97		34.33	37.83	3.53	35.83		
5	38.9	34.57	36.73	2.17	35.77	34.40	36.67	2.27	35.90		35.73	38.63	2.87	36.70		35.43	38.47	3.00	36.63		
7	38.7	33.97	36.03	2.07	35.00	33.53	36.40	2.87	35.47		35.03	38.10	3.07	36.03		34.90	38.00	3.10	36.27		
9	39.6	33.53	35.83	2.30	34.67	33.77	36.43	2.67	35.57		35.17	38.13	2.93	36.07		35.37	38.10	2.73	36.33		
11	40.1	33.53	36.23	2.70	35.13	34.07	37.07	3.00	36.17		35.33	38.50	3.20	36.43		35.60	38.43	2.87	36.70		
11.5	41.5	35.23	37.77	2.53	36.40	35.70	38.00	2.30	36.90		36.13	39.10	2.97	37.37		36.40	39.00	2.60	37.60		
13	40.8	32.77	37.07	4.30	35.13	33.93	38.03	4.10	36.70		36.63	40.30	3.67	38.07		35.27	40.27	5.00	37.43		
15	39.3	35.40	37.70	2.30	36.70	35.47	37.93	2.47	37.10		36.40	39.73	3.37	37.67		34.30	39.80	5.50	36.93		
17	38.4	34.90	36.77	1.87	35.70	34.13	36.47	2.33	35.77		35.47	38.30	2.80	36.63		34.90	38.33	3.43	36.23		
19	38.4	31.40	35.77	4.37	33.83	32.87	35.50	2.63	34.83		34.00	37.13	3.10	35.40		34.33	37.07	2.73	35.50		
21	38.5	32.00	35.77	3.77	34.53	34.50	35.97	1.47	35.63		35.10	37.23	2.20	36.17		34.87	37.37	2.50	36.00		
23	38.6	32.63	35.10	2.47	34.13	32.80	35.03	2.23	34.50		34.43	36.73	2.33	35.67		33.03	36.93	3.93	35.33		
24	39.2	31.77	35.03	3.27	33.33	31.73	35.13	3.40	34.10		34.43	36.40	1.93	35.43		33.40	36.63	3.20	35.10		

reference period

period after challenge

Table 35: Cow 3, mean temperature values (°C) throughout the trial, obtained by evaluation using the teats and evaluation using the udder. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

rectal temperature (°C)	EVALUATION USING THE TEATS										EVALUATION USING THE UDDER										
	left hindquarter					right hindquarter					left hindquarter					right hindquarter					
hours after challenge	Min	Max	Max-Min	Avg		Min	Max	Max-Min	Avg		Min	Max	Max-Min	Avg		Min	Max	Max-Min	Avg		
-25	32.70	34.70	2.00	33.83		32.53	35.23	2.70	33.90		33.43	37.33	3.93	34.90		33.57	37.33	3.80	35.10		
-24	37.8	27.47	34.37	30.83		26.47	34.33	7.87	30.43		32.90	36.83	3.90	34.17		32.43	37.00	4.60	34.23		
-23	37.9	33.83	34.80	0.97	34.30		34.17	35.07	0.90	34.57		33.17	37.10	3.93	34.70		33.80	37.23	3.43	34.93	
-22	37.9	33.20	35.60	2.40	34.40		32.47	35.23	2.77	33.87		33.87	37.53	3.67	35.53		34.20	37.60	3.43	35.37	
-21	38.1	33.60	35.77	2.17	34.87		34.10	35.70	1.60	35.13		34.70	38.13	3.40	36.33		34.80	38.07	3.30	36.53	
-19	38.2	31.97	34.73	2.77	33.50		31.80	35.17	3.37	33.37		33.93	37.80	3.87	35.70		34.43	37.80	3.33	35.90	
-17	38	30.87	34.73	3.87	33.03		28.70	35.13	6.43	32.67		33.47	37.80	4.30	35.60		33.87	37.73	3.83	35.40	
-15	38.3	32.57	34.70	2.13	33.70		32.27	35.87	3.70	34.70		33.47	37.60	4.10	35.37		34.23	37.50	3.23	35.60	
-13	38.5	32.00	34.47	2.47	33.53		32.13	35.47	3.33	34.07		33.37	37.77	4.43	35.30		33.77	37.70	3.93	35.73	
-12.5	37.9	32.13	35.10	2.97	33.57		32.10	35.30	3.20	33.93		34.17	38.03	3.90	35.80		34.27	38.07	3.80	35.87	
-11	38.1	33.50	35.63	2.13	34.73		34.00	35.70	1.70	35.03		34.23	38.23	4.00	36.10		34.67	38.20	3.57	36.43	
-9	38.3	32.23	35.03	2.80	34.17		32.50	35.43	2.93	34.23		34.13	37.83	3.67	36.23		34.53	37.83	3.33	36.13	
-7	38	32.37	34.90	2.53	33.83		31.90	35.47	3.57	33.83		33.87	37.73	3.87	35.63		34.30	37.73	3.37	35.90	
-5	38	32.13	34.87	2.73	33.87		31.53	35.60	4.07	33.77		34.13	38.07	3.90	36.00		34.17	37.97	3.80	36.30	
-3	38.3	32.57	34.33	1.77	33.57		33.00	35.30	2.30	34.17		33.93	37.63	3.70	35.50		33.83	37.50	3.67	35.27	
-1	38.5	32.63	34.70	2.07	33.80		33.57	35.13	1.57	34.27		33.83	37.47	3.63	35.33		34.00	37.20	3.17	35.40	
0	38.3	32.87	34.60	1.73	33.83		32.37	35.47	3.10	33.97		34.03	37.60	3.57	35.40		34.17	38.20	4.03	35.53	
1	38.1	32.53	34.47	1.93	33.70		32.13	34.50	2.37	33.33		33.47	37.20	3.73	34.90		33.30	37.50	4.17	34.87	
2	38.1	33.23	35.00	1.77	34.40		33.33	35.07	1.73	34.20		34.07	37.47	3.40	35.57		33.97	37.77	3.77	35.37	
3	38.5	32.83	34.97	2.13	34.17		32.30	35.00	2.70	33.80		33.50	37.70	4.23	35.47		33.53	37.90	4.33	35.40	
5	38.4	33.53	35.37	1.83	34.43		33.10	35.90	2.80	34.57		34.63	38.00	3.37	35.87		34.57	38.03	3.50	35.97	
7	39.1	30.80	34.00	3.20	32.67		30.37	35.23	4.87	33.47		33.33	37.43	4.10	35.00		33.90	37.47	3.57	35.20	
9	41.6	33.17	34.93	1.77	34.53		31.97	36.17	4.20	34.20		34.00	38.20	4.20	35.77		34.93	38.33	3.37	36.03	
11	41.6	32.53	35.63	3.10	34.50		33.00	37.43	4.43	35.37		33.20	39.67	6.43	36.10		34.30	39.63	5.30	36.57	
11.5	42.2	34.70	36.90	2.20	36.17		36.00	38.40	2.40	37.23		35.80	39.90	4.10	37.37		35.57	39.87	4.33	37.53	
13	41.6	34.33	36.73	2.40	35.83		34.70	39.43	4.73	37.50		35.90	40.50	4.57	37.37		34.57	40.43	5.87	37.13	
15	39.4	33.17	36.10	2.93	34.97		34.27	36.90	2.63	35.73		35.03	40.17	5.13	37.10		34.80	40.07	5.23	37.30	
17	38.8	32.73	35.40	2.67	34.37		34.17	36.23	2.07	35.13		34.80	39.03	4.07	37.30		34.80	39.03	4.23	37.27	
19	38.6	33.97	35.20	1.23	34.73		32.90	34.83	1.93	34.10		34.33	37.50	3.17	35.83		33.83	37.60	3.80	36.03	
21	38.9	31.23	34.67	3.43	33.50		32.43	34.43	2.00	33.83		34.37	37.70	3.33	36.00		33.83	37.93	4.13	36.50	
23	39	32.23	34.40	2.17	33.77		32.60	34.50	1.90	33.93		33.97	37.57	3.63	35.97		32.90	37.60	4.70	36.47	
24	39.2	31.97	34.63	2.67	34.00		32.07	35.90	3.83	34.50		33.97	37.07	3.07	35.63		34.17	37.23	3.10	36.10	

Table 36: Cow 4, mean temperature values (°C) throughout the trial, obtained by evaluation using the teats and evaluation using the udder. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

hours after challenge	rectal temperature (°C)	teats					udder										
		Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg								
-25	37.8	32.70	34.70	2.00	33.83	32.53	35.23	2.70	33.90	33.43	37.33	3.93	34.90	33.57	37.33	3.80	35.10
-24	37.8	27.47	34.37	6.90	30.83	26.47	34.33	7.87	30.43	32.90	36.83	3.90	34.17	32.43	37.00	4.60	34.23
-23	37.9	33.83	34.80	0.97	34.30	34.17	35.07	0.90	34.57	33.17	37.10	3.93	34.70	33.80	37.23	3.43	34.93
-22	37.9	33.20	35.60	2.40	34.40	32.47	35.23	2.77	33.87	33.87	37.53	3.67	35.53	34.20	37.60	3.43	35.37
-21	38.1	33.60	35.77	2.17	34.87	34.10	35.70	1.60	35.13	34.70	38.13	3.40	36.33	34.80	38.07	3.30	36.53
-19	38.2	31.97	34.73	2.77	33.50	31.80	35.17	3.37	33.37	33.93	37.80	3.87	35.70	34.43	37.80	3.33	35.90
-17	38	30.87	34.73	3.87	33.03	28.70	35.13	6.43	32.67	33.47	37.80	4.30	35.60	33.87	37.73	3.83	35.40
-15	38.3	32.57	34.70	2.13	33.70	32.27	35.87	3.70	34.70	33.47	37.60	4.10	35.37	34.23	37.50	3.23	35.60
-13	38.5	32.00	34.47	2.47	33.53	32.13	35.47	3.33	34.07	33.37	37.77	4.43	35.30	33.77	37.70	3.93	35.73
-12.5	37.9	32.13	35.10	2.97	33.57	32.10	35.30	3.20	33.93	34.17	38.03	3.90	35.80	34.27	38.07	3.80	35.87
-11	38.1	33.50	35.63	2.13	34.73	34.00	35.70	1.70	35.03	34.23	38.23	4.00	36.10	34.67	38.20	3.57	36.43
-9	38.3	32.23	35.03	2.80	34.17	32.50	35.43	2.93	34.23	34.13	37.83	3.67	36.23	34.53	37.83	3.33	36.13
-7	38	32.37	34.90	2.53	33.83	31.90	35.47	3.57	33.83	33.87	37.73	3.87	35.63	34.30	37.73	3.37	35.90
-5	38	32.13	34.87	2.73	33.87	31.53	35.60	4.07	33.77	34.13	38.07	3.90	36.00	34.17	37.97	3.80	36.30
-3	38.3	32.57	34.33	1.77	33.57	33.00	35.30	2.30	34.17	33.93	37.63	3.70	35.50	33.83	37.50	3.67	35.27
-1	38.5	32.63	34.70	2.07	33.80	33.57	35.13	1.57	34.27	33.83	37.47	3.63	35.33	34.00	37.20	3.17	35.40
0	38.3	32.87	34.60	1.73	33.83	32.37	35.47	3.10	33.97	34.03	37.60	3.57	35.40	34.17	38.20	4.03	35.53
1	38.1	32.53	34.47	1.93	33.70	32.13	34.50	2.37	33.33	33.47	37.20	3.73	34.90	33.30	37.50	4.17	34.87
2	38.1	33.23	35.00	1.77	34.40	33.33	35.07	1.73	34.20	34.07	37.47	3.40	35.57	33.97	37.77	3.77	35.37
3	38.5	32.83	34.97	2.13	34.17	32.30	35.00	2.70	33.80	33.50	37.70	4.23	35.47	33.53	37.90	4.33	35.40
5	38.4	33.53	35.37	1.83	34.43	33.10	35.90	2.80	34.57	34.63	38.00	3.37	35.87	34.57	38.03	3.50	35.97
7	39.1	30.80	34.00	3.20	32.67	30.37	35.23	4.87	33.47	33.33	37.43	4.10	35.00	33.90	37.47	3.57	35.20
9	41.6	33.17	34.93	1.77	34.53	31.97	36.17	4.20	34.20	34.00	38.20	4.20	35.77	34.93	38.33	3.37	36.03
11	41.6	32.53	35.63	3.10	34.50	33.00	37.43	4.43	35.37	33.20	39.67	6.43	36.10	34.30	39.63	5.30	36.57
11.5	42.2	34.70	36.90	2.20	36.17	36.00	38.40	2.40	37.23	35.80	39.90	4.10	37.37	35.57	39.87	4.33	37.53
13	41.6	34.33	36.73	2.40	35.83	34.70	39.43	4.73	37.50	35.90	40.50	4.57	37.37	34.57	40.43	5.87	37.13
15	39.4	33.17	36.10	2.93	34.97	34.27	36.90	2.63	35.73	35.03	40.17	5.13	37.10	34.80	40.07	5.23	37.30
17	38.8	32.73	35.40	2.67	34.37	34.17	36.23	2.07	35.13	34.80	39.03	4.07	37.30	34.80	39.03	4.23	37.27
19	38.6	33.97	35.20	1.23	34.73	32.90	34.83	1.93	34.10	34.33	37.50	3.17	35.83	33.83	37.60	3.80	36.03
21	38.9	31.23	34.67	3.43	33.50	32.43	34.43	2.00	33.83	34.37	37.70	3.33	36.00	33.83	37.93	4.13	36.50
23	39	32.23	34.40	2.17	33.77	32.60	34.50	1.90	33.93	33.97	37.57	3.63	35.97	32.90	37.60	4.70	36.47
24	39.2	31.97	34.63	2.67	34.00	32.07	35.90	3.83	34.50	33.97	37.07	3.07	35.63	34.17	37.23	3.10	36.10

reference period

period after challenge

Table 37: Cow 5, mean temperature values (°C) throughout the trial, obtained by evaluation using the teats and evaluation using the udder. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

hours after challenge	rectal temperature (°C)	EVALUATION USING THE TEATS						EVALUATION USING THE UDDER									
		left hindquarter			right hindquarter			left hindquarter			right hindquarter						
		Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg
-25	38.2	32.70	34.70	2.00	33.83	32.53	35.23	2.70	33.90	33.43	37.33	3.93	34.90	33.57	37.33	3.80	35.10
-24	38.2	27.47	34.37	6.90	30.83	26.47	34.33	7.87	30.43	32.90	36.83	3.90	34.17	32.43	37.00	4.60	34.23
-23	38.3	33.83	34.80	0.97	34.30	34.17	35.07	0.90	34.57	33.17	37.10	3.93	34.70	33.80	37.23	3.43	34.93
-22	38.4	33.20	35.60	2.40	34.40	32.47	35.23	2.77	33.87	33.87	37.53	3.67	35.53	34.20	37.60	3.43	35.37
-21	38.3	33.60	35.77	2.17	34.87	34.10	35.70	1.60	35.13	34.70	38.13	3.40	36.33	34.80	38.07	3.30	36.53
-19	38.2	31.97	34.73	2.77	33.50	31.80	35.17	3.37	33.37	33.93	37.80	3.87	35.70	34.43	37.80	3.33	35.90
-17	38.2	30.87	34.73	3.87	33.03	28.70	35.13	6.43	32.67	33.47	37.80	4.30	35.60	33.87	37.73	3.83	35.40
-15	38.3	32.57	34.70	2.13	33.70	32.27	35.87	3.70	34.70	33.47	37.60	4.10	35.37	34.23	37.50	3.23	35.60
-13	38.3	32.00	34.47	2.47	33.53	32.13	35.47	3.33	34.07	33.37	37.77	4.43	35.30	33.77	37.70	3.93	35.73
-12.5	38.4	32.13	35.10	2.97	33.57	32.10	35.30	3.20	33.93	34.17	38.03	3.90	35.80	34.27	38.07	3.80	35.87
-11	38.4	33.50	35.63	2.13	34.73	34.00	35.70	1.70	35.03	34.23	38.23	4.00	36.10	34.67	38.20	3.57	36.43
-9	38.4	32.23	35.03	2.80	34.17	32.50	35.43	2.93	34.23	34.13	37.83	3.67	36.23	34.53	37.83	3.33	36.13
-7	38.4	32.37	34.90	2.53	33.83	31.90	35.47	3.57	33.83	33.87	37.73	3.87	35.63	34.30	37.73	3.37	35.90
-5	38.4	32.13	34.87	2.73	33.87	31.53	35.60	4.07	33.77	34.13	38.07	3.90	36.00	34.17	37.97	3.80	36.30
-3	38.9	32.57	34.33	1.77	33.57	33.00	35.30	2.30	34.17	33.93	37.63	3.70	35.50	33.83	37.50	3.67	35.27
-1	38.8	32.63	34.70	2.07	33.80	33.57	35.13	1.57	34.27	33.83	37.47	3.63	35.33	34.00	37.20	3.17	35.40
0	38.3	32.87	34.60	1.73	33.83	32.37	35.47	3.10	33.97	34.03	37.60	3.57	35.40	34.17	38.20	4.03	35.53
1	38.2	32.53	34.47	1.93	33.70	32.13	34.50	2.37	33.33	33.47	37.20	3.73	34.90	33.30	37.50	4.17	34.87
2	38.2	33.23	35.00	1.77	34.40	33.33	35.07	1.73	34.20	34.07	37.47	3.40	35.57	33.97	37.77	3.77	35.37
3	38.2	32.83	34.97	2.13	34.17	32.30	35.00	2.70	33.80	33.50	37.70	4.23	35.47	33.53	37.90	4.33	35.40
5	38.2	33.53	35.37	1.83	34.43	33.10	35.90	2.80	34.57	34.63	38.00	3.37	35.87	34.57	38.03	3.50	35.97
7	38.3	30.80	34.00	3.20	32.67	30.37	35.23	4.87	33.47	33.33	37.43	4.10	35.00	33.90	37.47	3.57	35.20
9	39	33.17	34.93	1.77	34.53	31.97	36.17	4.20	34.20	34.00	38.20	4.20	35.77	34.93	38.33	3.37	36.03
11	39.4	32.53	35.63	3.10	34.50	33.00	37.43	4.43	35.37	33.20	39.67	6.43	36.10	34.30	39.63	5.30	36.57
11.5	40.8	34.70	36.90	2.20	36.17	36.00	38.40	2.40	37.23	35.80	39.90	4.10	37.37	35.57	39.87	4.33	37.53
13	40.6	34.33	36.73	2.40	35.83	34.70	39.43	4.73	37.50	35.90	40.50	4.57	37.37	34.57	40.43	5.87	37.13
15	39.4	33.17	36.10	2.93	34.97	34.27	36.90	2.63	35.73	35.03	40.17	5.13	37.10	34.80	40.07	5.23	37.30
17	38.8	32.73	35.40	2.67	34.37	34.17	36.23	2.07	35.13	34.80	39.03	4.07	37.30	34.80	39.03	4.23	37.27
19	38.6	33.97	35.20	1.23	34.73	32.90	34.83	1.93	34.10	34.33	37.50	3.17	35.83	33.83	37.60	3.80	36.03
21	38.5	31.23	34.67	3.43	33.50	32.43	34.43	2.00	33.83	34.37	37.70	3.33	36.00	33.83	37.93	4.13	36.50
23	38.3	32.23	34.40	2.17	33.77	32.60	34.50	1.90	33.93	33.97	37.57	3.63	35.97	32.90	37.60	4.70	36.47
24	39.2	31.97	34.63	2.67	34.00	32.07	35.90	3.83	34.50	33.97	37.07	3.07	35.63	34.17	37.23	3.10	36.10

period after challenge

reference period

Table 38: Median temperature values of all five cows (°C) throughout the trial, obtained by evaluation using the teats and evaluation using the udder. ‘Min’ – Minimum temperature values; ‘Max’ – Maximum temperature values; ‘Max-Min’ – Range of temperature values; ‘Avg’ – Average temperature values.

hours after challenge	rectal temperature (°C)	EVALUATION USING THE TEATS						EVALUATION USING THE UDDER									
		left hindquarter			right hindquarter			left hindquarter			right hindquarter						
		Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg	Min	Max	Max-Min	Avg
-25	38.24	32.35	35.15	2.81	34.07	32.11	34.91	2.80	33.88	34.10	37.44	3.33	35.46	33.65	37.31	3.65	35.30
-24	38.16	31.00	35.21	4.21	33.09	30.39	34.92	4.53	32.96	33.41	37.54	4.13	35.27	33.14	37.47	4.33	35.10
-23	38.14	33.05	35.30	2.25	34.11	33.20	35.27	2.07	34.37	33.75	37.30	3.53	35.36	34.07	37.39	3.31	35.43
-22	38.32	*	*	*	*	*	*	*	*	33.89	37.35	3.49	35.49	33.88	37.31	3.43	35.37
-21	38.36	32.80	35.47	2.67	34.10	32.49	35.36	2.87	34.15	34.27	37.87	3.64	35.91	34.10	37.70	3.60	35.83
-19	38.4	32.96	35.43	2.47	34.20	32.59	35.40	2.81	34.10	34.34	37.69	3.36	35.87	34.30	37.59	3.29	35.78
-17	38.26	32.43	35.32	2.89	33.99	31.85	35.25	3.39	33.88	34.03	37.56	3.53	35.73	33.83	37.51	3.68	35.65
-15	38.36	32.69	35.44	2.76	34.16	31.61	35.14	3.55	33.78	34.11	37.54	3.41	35.59	33.64	37.34	3.68	35.37
-13	38.38	31.91	35.44	3.53	33.78	31.56	35.57	4.01	33.83	34.04	37.64	3.62	35.79	34.00	37.53	3.49	35.59
-12.5	38.4	*	*	*	*	*	*	*	*	33.87	37.57	3.69	35.65	33.89	37.51	3.62	35.45
-11	38.56	33.73	35.96	2.23	34.81	33.59	35.92	2.33	34.87	34.48	37.91	3.43	36.19	34.75	37.84	3.11	36.22
-9	38.58	33.45	35.68	2.23	34.64	33.25	35.71	2.46	34.72	34.79	37.98	3.17	36.44	34.86	37.91	3.05	36.35
-7	38.46	32.92	35.21	2.29	34.07	32.43	35.25	2.82	34.11	34.36	37.67	3.32	35.89	34.32	37.61	3.27	35.77
-5	38.56	32.28	35.38	3.10	34.04	32.75	35.49	2.74	34.39	34.35	37.81	3.45	35.92	34.25	37.72	3.48	35.97
-3	38.82	32.56	35.37	2.80	34.09	33.49	35.63	2.15	34.80	34.47	37.87	3.39	35.93	34.17	37.69	3.54	35.76
-1	38.68	33.25	35.25	2.00	34.29	32.21	35.32	3.11	34.10	34.23	37.87	3.63	35.95	33.93	37.69	3.75	35.91
0	38.36	32.97	35.33	2.35	34.29	32.77	35.50	2.73	34.29	33.91	37.64	3.71	35.70	34.15	37.76	3.61	35.63
1	38.28	33.27	35.52	2.25	34.43	33.08	35.17	2.09	34.19	34.08	37.63	3.54	35.74	34.10	37.63	3.53	35.65
2	38.26	*	*	*	*	*	*	*	*	34.49	37.65	3.14	36.01	34.41	37.71	3.31	35.84
3	38.44	33.36	35.35	1.99	34.45	33.06	35.43	2.37	34.39	34.27	37.54	3.29	35.81	34.31	37.67	3.34	35.79
5	38.5	30.33	35.41	5.08	33.61	30.33	35.01	4.67	33.12	34.13	37.73	3.57	35.70	33.99	37.61	3.63	35.51
7	38.56	32.62	35.15	2.53	33.99	32.49	35.38	2.89	34.20	34.12	37.53	3.41	35.64	34.15	37.59	3.45	35.64
9	39.8	33.14	35.43	2.29	34.46	32.72	35.59	2.87	34.45	34.14	37.81	3.67	35.87	34.39	37.77	3.37	35.89
11	40.38	31.95	35.98	4.03	34.33	33.60	36.37	2.77	35.28	33.98	38.59	4.60	36.33	33.49	38.59	5.09	36.02
11.5	41.32	*	*	*	*	*	*	*	*	34.14	38.57	4.43	36.54	34.01	38.72	4.71	36.14
13	41.1	32.17	37.15	4.97	34.77	34.01	37.37	3.35	36.08	35.47	39.85	4.37	37.57	34.31	39.82	5.51	36.67
15	39.54	34.22	37.17	2.95	35.73	34.89	37.13	2.23	36.34	35.43	39.81	4.39	37.59	34.49	39.73	5.23	37.00
17	38.72	33.45	36.35	2.91	34.90	33.99	36.31	2.31	35.47	35.11	39.67	3.53	37.11	34.39	38.67	4.27	36.67
19	38.6	33.33	35.94	2.61	34.75	33.75	35.47	1.73	34.89	34.86	37.95	3.10	36.39	34.19	37.83	3.63	36.17
21	38.94	31.96	35.45	3.49	33.98	33.35	35.21	1.86	34.62	34.74	37.61	2.89	36.25	34.16	37.63	3.49	36.12
23	38.88	30.30	35.00	4.70	32.77	32.75	35.11	2.36	34.29	34.01	37.57	3.56	36.03	33.58	37.59	4.01	35.99
24	39.2	32.81	35.39	2.58	34.27	32.81	35.85	3.04	34.78	34.17	37.33	3.16	36.01	34.16	37.76	3.58	36.00

reference period

period after challenge

3. Appendix 3: Tables and Histograms of normality tests

Table 39: Table of results of D'Agostino-Pearson normality test (omnibus K2) of data obtained by automatic and manual evaluation method, P-values > 0.05 are shown in bold.

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Avg' – Average temperature values

	Automatic			Manual		
	Min	Max	Avg	Min	Max	Avg
n	154	154	154	154	154	154
Minimum	27.80	34.83	32.65	26.03	36.40	33.63
First Quartile	30.97	35.87	34.15	31.62	36.97	34.93
Median	31.90	36.33	34.56	32.55	37.27	35.35
Third Quartile	32.71	36.77	35.07	33.25	37.58	35.73
Maximum	35.67	38.00	36.84	35.77	38.60	37.27
Mean	31.78	36.33	34.60	32.27	37.27	35.35
Standard Deviation	1.41	0.63	0.68	1.65	0.43	0.57
D'Agostino-Pearson normality test						
P value	0.31	0.89	0.11	< 0,0001	1.00	0.07
Skewness	-0.23	0.00	0.24	-1.17	0.00	0.19
Kurtosis	0.32	-0.22	0.76	2.03	-0.05	1.01

Table 40: Table of results of D'Agostino-Pearson normality test (omnibus K2) of data obtained by evaluation method using the teats and evaluation method using the udder, P-values > 0.05 are shown in bold.

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Avg' – Average temperature values

	Teats			Udder		
	Min	Max	Avg	Min	Max	Avg
n	154	154	154	160	160	160
Minimum	26.47	33.80	30.43	31.93	36.57	34.17
First Quartile	31.88	34.80	33.56	33.54	37.33	35.37
Median	32.70	35.27	34.03	34.13	37.67	35.73
Third Quartile	33.42	35.87	34.70	34.70	37.93	36.10
Maximum	35.97	37.27	36.70	36.33	38.87	37.57
Mean	32.44	35.33	34.03	34.10	37.62	35.73
Standard Deviation	1.69	0.70	1.03	0.83	0.43	0.54
D'Agostino-Pearson normality test						
P value	< 0.0001	0.25	0.00	0.98	0.89	0.11
Skewness	-1.13	0.31	-0.65	0.00	0.00	0.15
Kurtosis	1.81	-0.21	1.56	-0.13	-0.22	0.91

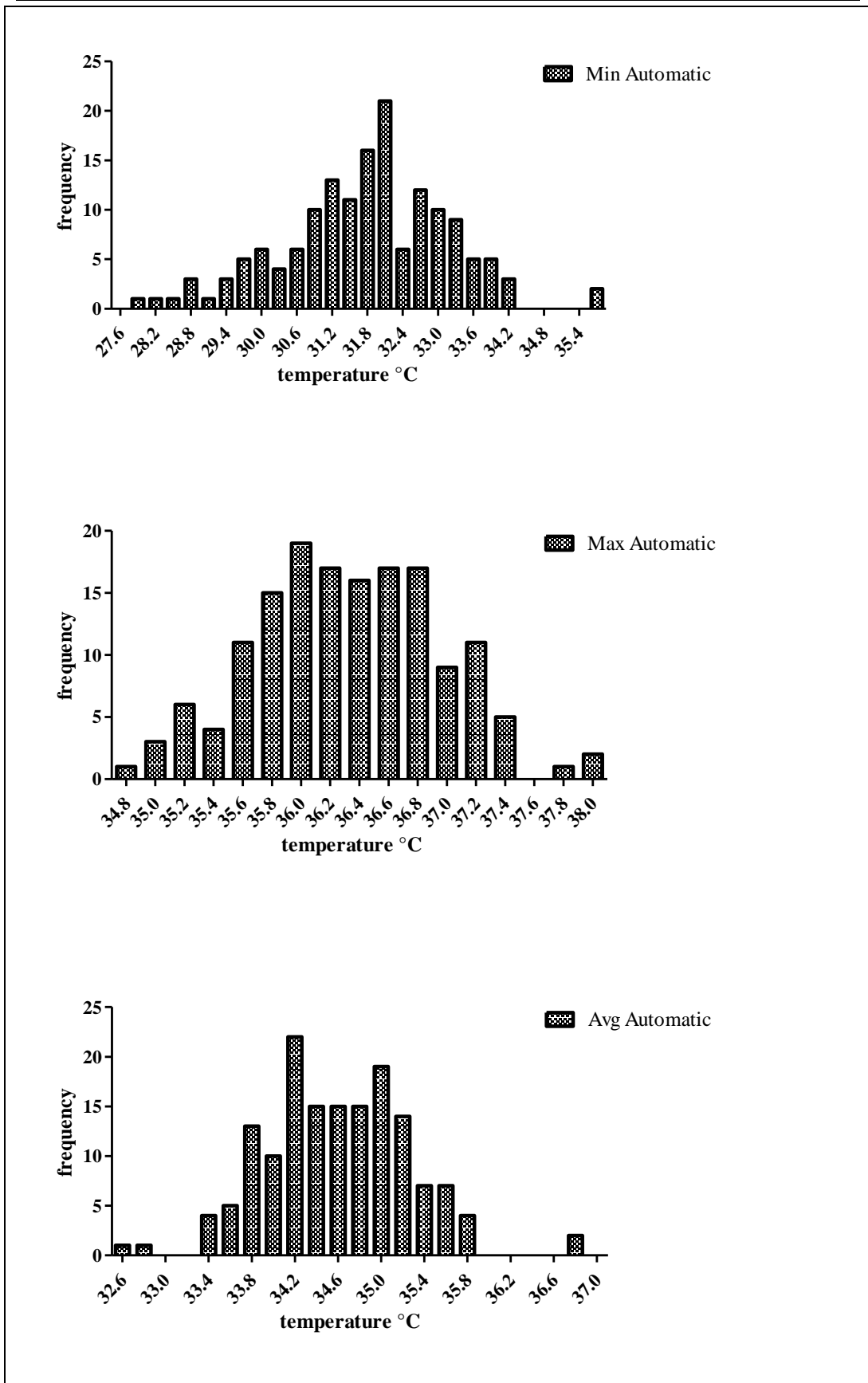


Figure 23: Histogram analysis of temperature parameters obtained by automatic evaluation method

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Avg' – Average temperature values

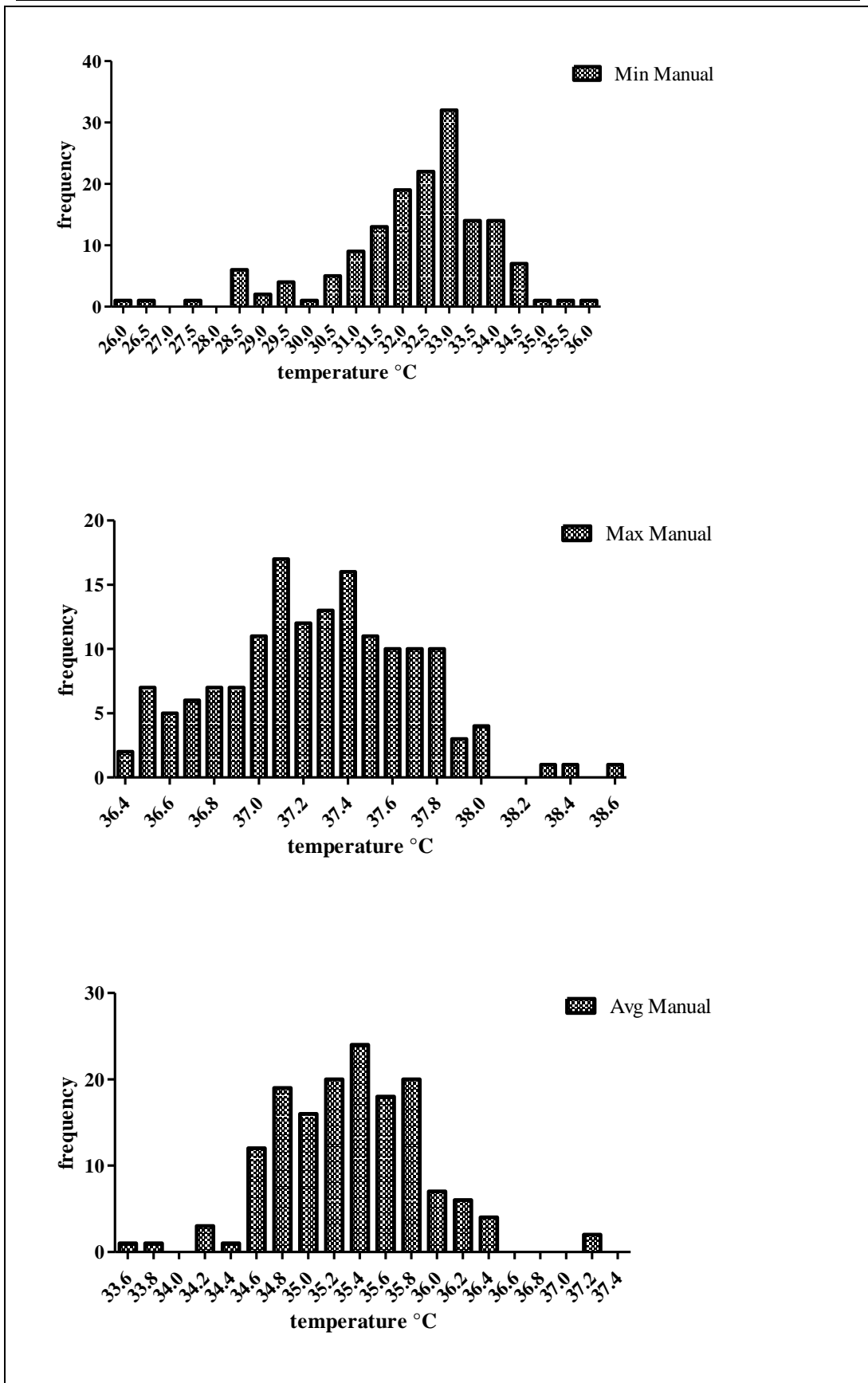


Figure 24: Histogram analysis of temperature parameters obtained by manual evaluation method

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Avg' – Average temperature values

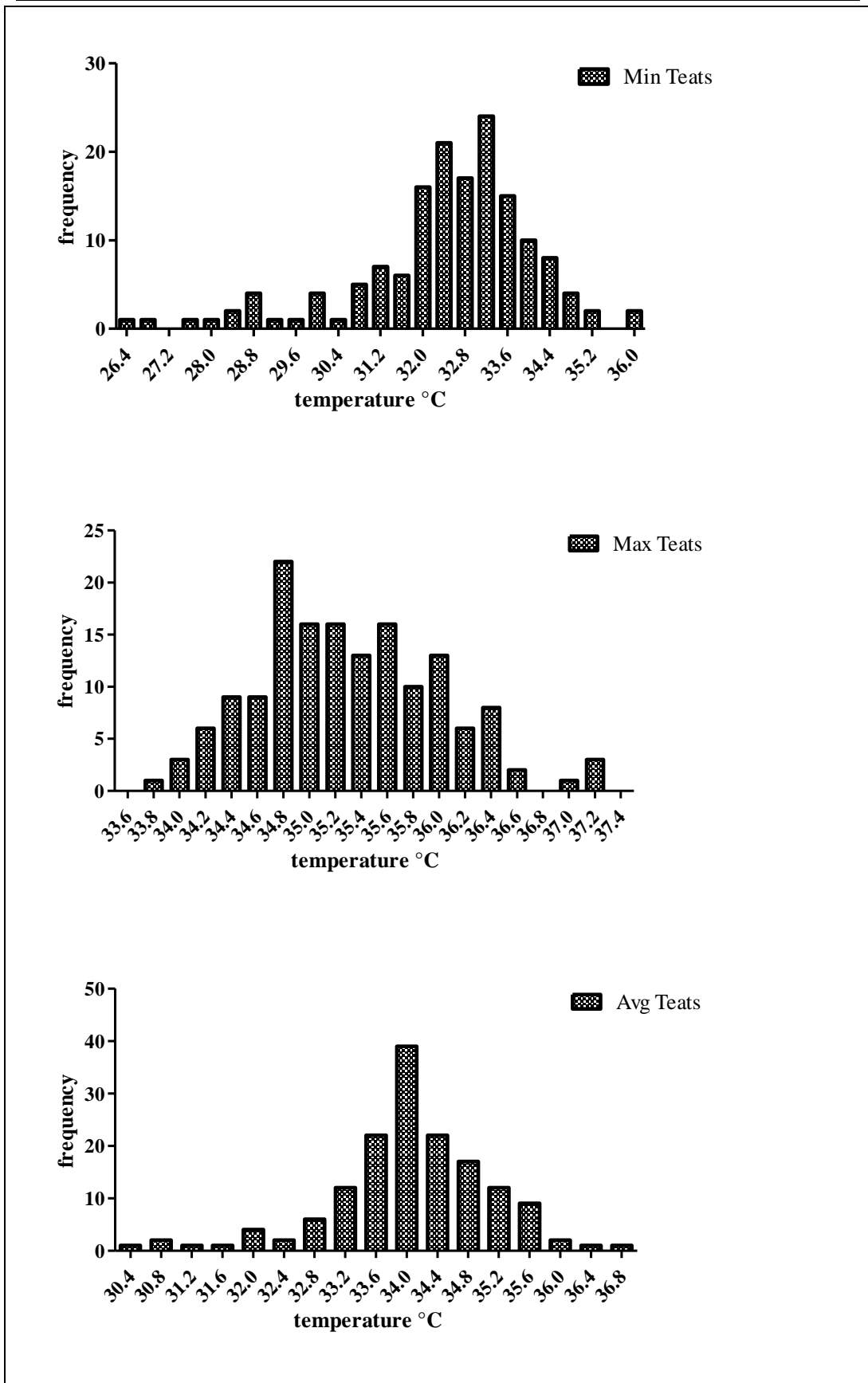


Figure 25: Histogram analysis of temperature parameters obtained by evaluation method using the teats

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Avg' – Average temperature values

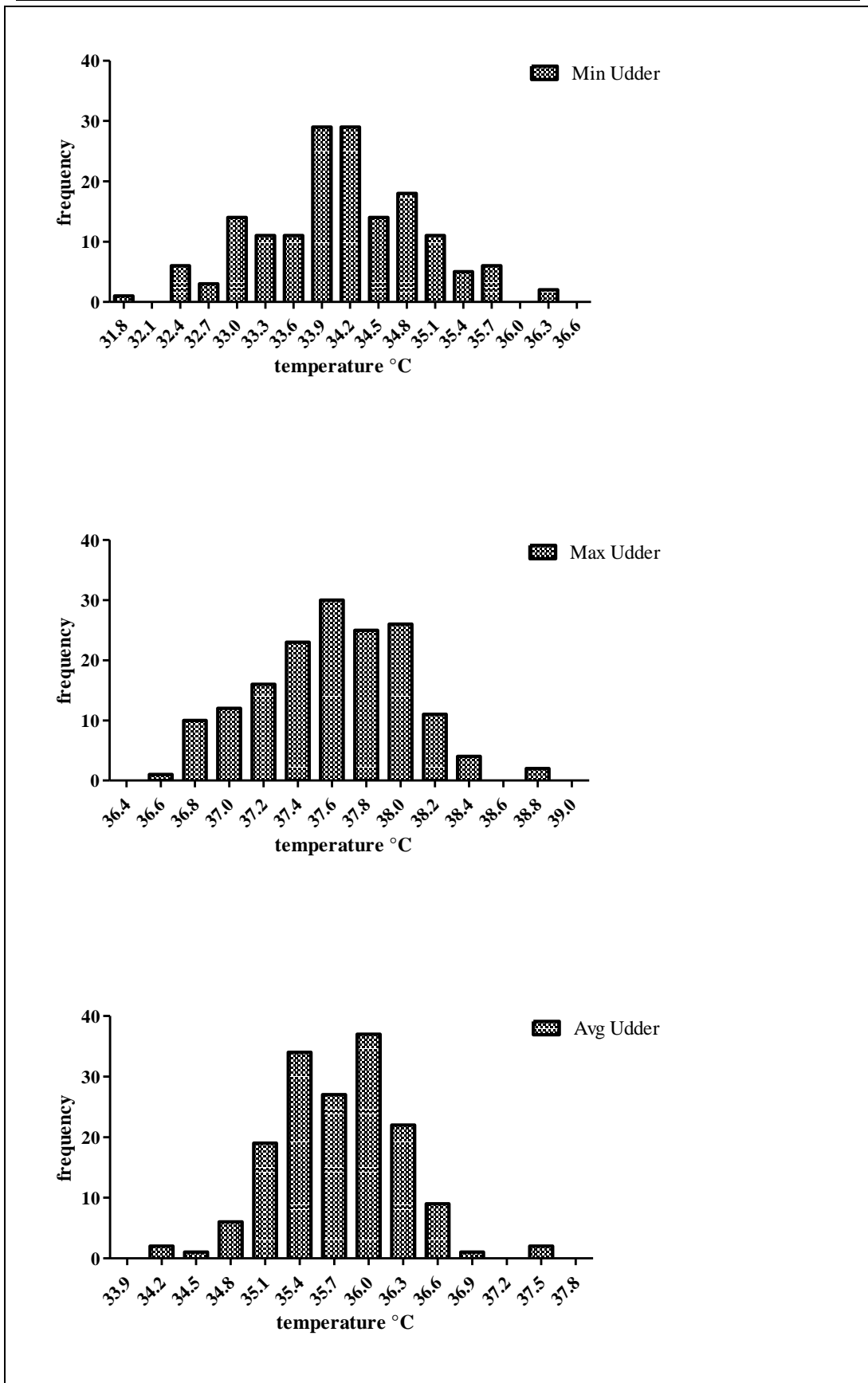


Figure 26: Histogram analysis of temperature parameters obtained by evaluation using the udder

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Avg' – Average temperature values

4. Appendix 4: Tables of precision analysis

Table 41: Median, q1 and q3 of coefficient of variation of the different parameters in automatic and manual evaluation. Medians are shown in bold.

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Max-Min' – Range of temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

	Min HL	Max HL	Max-Min HL	Avg HL	Min HR	Max HR	Max-Min HR	Avg HR
AUTOMATIC								
q1	0.0054	0.0022	0.0419	0.0011	0.0074	0.0027	0.0614	0.0010
Median	0.0094	0.0041	0.0807	0.0016	0.0141	0.0043	0.1203	0.0018
q3	0.0240	0.0055	0.1898	0.0027	0.0328	0.0078	0.2381	0.0025
MANUAL								
q1	0.0033	0.0015	0.0190	0.0016	0.0031	0.0015	0.0256	0.0000
Median	0.0049	0.0016	0.0441	0.0016	0.0061	0.0016	0.0458	0.0016
q3	0.0089	0.0031	0.0657	0.0017	0.0110	0.0041	0.0704	0.0017

Table 42: Median, q1 and q3 of coefficient of variation of the different parameters in evaluation using the teats ('Teats') and evaluation using the udder ('Udder'). Medians are shown in bold.

'Min' – Minimum temperature values

'Max' – Maximum temperature values

'Max-Min' – Range of temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

	Min HL	Max HL	Max-Min HL	Avg HL	Min HR	Max HR	Max-Min HR	Avg HR
TEATS								
q1	0.0035	0.0016	0.0373	0.0017	0.0037	0.0016	0.0533	0.0017
Median	0.0054	0.0028	0.0628	0.0029	0.0064	0.0017	0.0726	0.0029
q3	0.0087	0.0043	0.1204	0.0045	0.0093	0.0030	0.1097	0.0045
UDDER								
q1	0.0018	0.0015	0.0204	0.0000	0.0017	0.0015	0.0188	0.0000
Median	0.0034	0.0015	0.0346	0.0016	0.0029	0.0016	0.0333	0.0016
q3	0.0060	0.0027	0.0640	0.0016	0.0045	0.0027	0.0465	0.0016

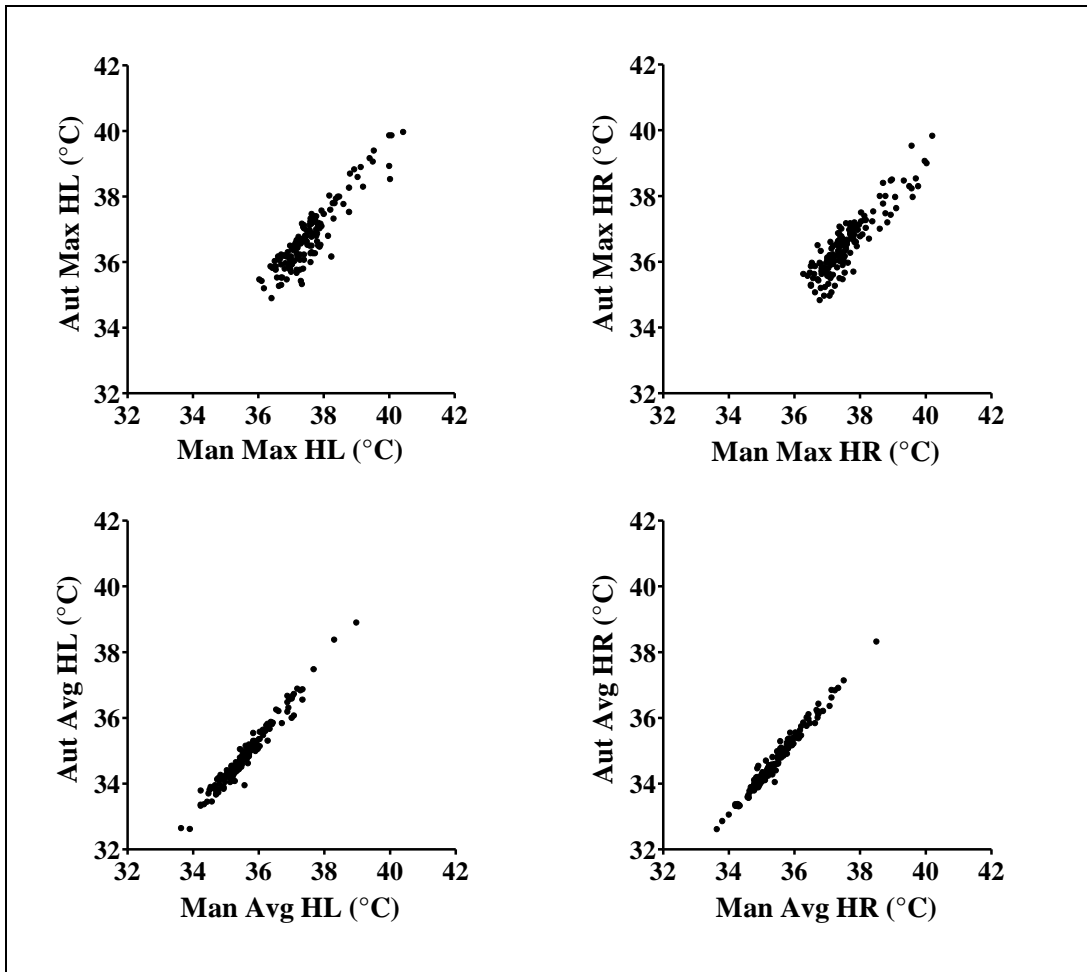
5. Appendix 5: Figures of method comparison

Figure 27: Scatter plots of values of maximum temperature ('Max') and average temperature ('Avg') in automatic ('Aut') compared to manual ('Man') evaluation method, separated by left ('HL') and right hindquarter ('HR').

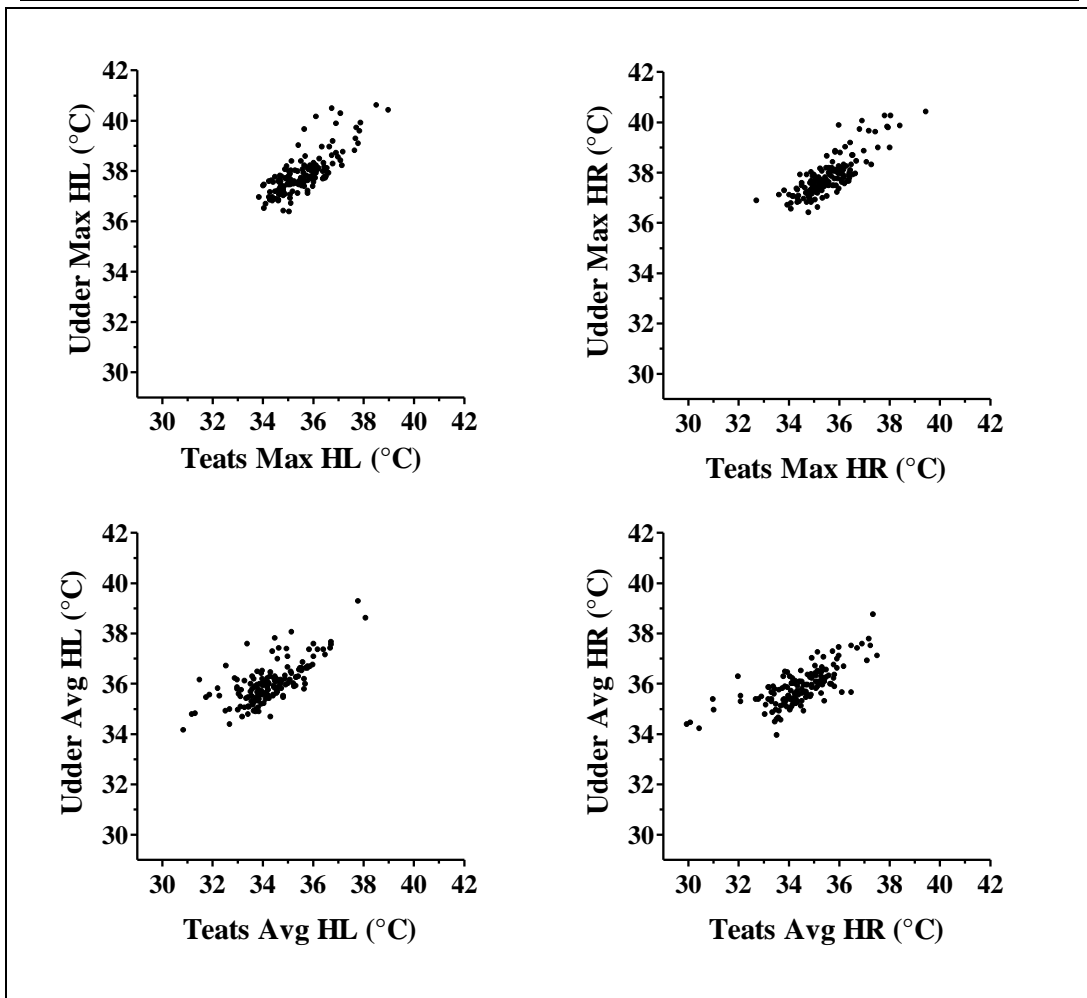


Figure 28: Scatter plots of values of maximum temperature ('Max') and average temperature ('Avg') in evaluation using the teats ('Teats') compared to evaluation using the udder ('Udder'), separated by left ('HL') and right hindquarter ('HR').

Table 43: Median difference (diff.), first quartile (q1) and third quartile (q3) of temperature values in automatic evaluation in degrees Celsius ($^{\circ}\text{C}$). The larger the difference is, the darker it is colored red. Differences are calculated for values measured after challenge ('-l' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Automatic evaluation													
time	-1	0	1	3	5	7	9	11	13	15	17	19	21
Median diff. of													
'Max HL'	0.37	0.23	-0.07	-0.10	0.53	-0.07	0.30	0.63	2.43	1.97	1.13	0.43	0.00
q1	0.17	-0.07	-0.20	-0.73	0.33	-0.43	-0.10	0.43	2.30	1.90	0.50	-0.07	-0.40
q3	0.63	0.33	0.30	-0.07	0.57	0.07	0.77	1.90	2.57	2.30	1.47	0.60	0.10
Median diff. of													
'Max HR'	0.67	0.03	-0.03	0.20	0.37	-0.03	0.93	0.93	2.13	1.83	1.33	-0.17	0.23
q1	0.40	-0.20	-0.07	-1.13	0.03	-0.30	0.50	0.77	1.57	1.73	1.13	-0.57	0.00
q3	0.70	0.70	-0.03	0.43	0.57	0.03	0.97	1.00	2.33	2.03	1.43	0.27	0.33
Median diff. of													
'Avg HL'	0.59	0.34	0.21	-0.32	-0.04	0.07	0.46	0.38	1.47	0.99	1.19	0.76	0.37
q1	0.44	0.32	0.09	-0.55	-0.48	-0.66	0.33	0.13	1.17	0.83	1.12	-0.05	0.19
q3	0.66	0.41	0.33	0.65	0.22	0.09	0.57	0.73	1.82	1.52	1.51	0.84	0.61
Median diff. of													
'Avg HR'	0.79	0.26	-0.08	0.37	-0.01	0.04	0.60	0.80	0.78	1.08	1.01	-0.12	0.32
q1	0.15	0.25	-0.26	-0.74	-0.51	-0.11	0.60	-0.03	0.15	0.67	0.79	-0.14	-0.09
q3	1.03	0.62	0.49	0.67	0.01	0.22	0.70	0.86	0.95	1.35	1.19	0.50	0.63

Table 44: Median difference (diff.), first quartile (q1) and third quartile (q3) of temperature values in manual evaluation in degrees Celsius ($^{\circ}\text{C}$). The larger the difference is, the darker it is colored red. Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Manual evaluation													
time	-1	0	1	3	5	7	9	11	13	15	17	19	21
Median diff. of													
'Max HL'	0.43	0.10	0.27	-0.23	0.17	-0.03	0.23	0.73	2.10	1.87	0.80	0.60	-0.43
q1	0.33	-0.13	0.27	-0.43	-0.10	-0.10	0.00	0.17	1.90	1.87	0.50	-0.40	-0.50
q3	0.47	0.63	0.30	0.13	0.53	0.13	0.47	1.37	2.17	2.00	1.50	1.30	0.07
Median diff. of													
'Max HR'	0.50	0.20	-0.10	0.10	-0.03	0.37	0.60	1.03	1.93	1.77	1.20	0.20	-0.20
q1	0.43	0.00	-0.33	-0.23	-0.20	0.27	0.57	0.80	1.87	1.73	0.57	0.17	-0.27
q3	0.50	0.53	0.20	0.27	0.20	0.43	0.63	1.33	2.63	2.17	1.40	1.17	0.77
Median diff. of													
'Avg HL'	0.47	0.30	0.23	-0.30	0.17	-0.07	0.50	0.23	1.33	0.87	0.90	0.67	0.17
q1	0.33	0.10	0.20	-0.67	-0.77	-0.50	0.13	0.20	1.30	0.80	0.90	-0.10	0.13
q3	0.60	0.57	0.47	0.50	0.33	0.07	0.57	0.73	1.70	1.33	1.57	0.87	0.47
Median diff. of													
'Avg HR'	0.67	0.53	0.33	0.23	0.00	0.10	0.40	0.67	0.67	0.83	0.73	0.03	0.37
q1	0.23	0.17	-0.23	-0.57	-0.60	0.00	0.40	0.03	0.43	0.70	0.53	-0.17	0.00
q3	0.77	0.53	0.40	0.47	0.03	0.17	0.80	0.70	0.73	1.20	1.20	0.30	0.37

Table 45: Median difference (diff.), first quartile (q1) and third quartile (q3) of temperature values in evaluation using the teats in degrees Celsius (°C). The larger the difference is, the darker it is colored red. Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Evaluation using the teats													
time	-1	0	1	3	5	7	9	11	13	15	17	19	21
Median diff. of													
'Max HL'	0.00	0.23	-0.07	-0.33	-0.07	-0.13	0.23	0.43	1.10	1.33	0.83	0.33	0.20
q1	-0.10	-0.03	-0.33	-0.80	-0.53	-0.23	-0.07	0.17	0.03	1.10	0.50	0.23	-0.20
q3	0.53	0.47	0.40	0.13	0.63	0.03	0.30	0.83	2.40	1.63	1.80	1.10	0.33
Median diff. of													
'Max HR'	0.23	0.40	-0.57	-0.10	-0.77	0.13	0.30	0.67	0.93	1.47	0.90	0.27	-0.30
q1	-0.10	0.37	-0.57	-0.13	-1.10	0.10	0.27	0.33	0.77	1.17	0.77	-0.67	-0.87
q3	0.97	0.70	0.20	0.23	0.73	0.30	0.53	0.97	1.53	1.67	1.20	0.53	0.03
Median diff. of													
'Avg HL'	-0.03	0.63	-0.17	-0.03	-0.27	-0.13	0.50	0.27	-0.40	0.97	0.53	0.87	-0.17
q1	-0.23	0.60	-0.23	-0.50	-2.47	-0.27	0.03	-1.03	-0.40	0.80	0.47	0.03	-0.17
q3	0.40	2.00	0.53	0.33	0.93	0.23	0.83	0.97	1.10	1.07	1.50	1.37	-0.07
Median diff. of													
'Avg HR'	0.37	0.70	-0.23	0.20	-0.13	0.60	0.23	1.30	1.40	1.50	1.27	0.43	0.30
q1	-0.80	0.67	-1.00	0.13	-3.37	-0.10	0.17	0.77	0.93	1.10	0.83	0.33	-0.33
q3	0.93	1.90	-0.10	0.33	1.20	0.63	0.83	2.10	1.83	2.20	1.30	0.63	0.33

Table 46: Median difference (diff.), first quartile (q1) and third quartile (q3) of temperature values in evaluation using the udder in degrees Celsius ($^{\circ}\text{C}$). The larger the difference is, the darker it is colored red. Differences are calculated for values measured after challenge ('-1' to '21' hours, time (h)) and values measured 24 hours earlier.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

Evaluation using the udder													
time	-1	0	1	3	5	7	9	11	13	15	17	19	21
Median diff. of													
'Max HL'	0.30	0.20	0.13	-0.43	0.03	0.00	0.33	0.77	2.07	1.87	0.80	0.37	-0.20
q1	0.30	-0.33	0.10	-0.53	-0.27	-0.33	0.13	0.33	2.03	1.73	0.67	-0.57	-0.40
q3	0.70	0.50	0.23	0.20	0.20	0.03	0.37	1.50	2.27	1.87	1.30	0.57	0.07
Median diff. of													
'Max HR'	0.43	0.03	0.17	0.17	0.03	-0.03	0.43	0.63	2.17	1.83	0.83	0.13	0.17
q1	0.17	-0.13	0.10	-0.17	-0.17	-0.10	0.27	0.60	1.93	1.63	0.60	-0.37	-0.27
q3	0.63	0.60	0.27	0.33	0.23	0.33	0.57	1.57	2.23	2.00	1.30	0.27	0.27
Median diff. of													
'Avg HL'	0.43	0.27	0.20	-0.27	0.17	-0.10	0.40	0.23	1.43	0.90	1.07	0.67	0.30
q1	0.43	0.10	0.20	-0.77	-0.60	-0.40	0.17	0.20	1.27	0.87	0.90	-0.17	0.20
q3	0.57	0.57	0.43	0.43	0.30	0.10	0.57	0.80	1.93	1.37	1.67	0.87	0.50
Median diff. of													
'Avg HR'	0.63	0.57	0.27	0.30	-0.07	0.00	0.43	0.60	0.67	0.80	0.83	-0.07	0.33
q1	0.30	0.07	-0.07	-0.47	-0.57	-0.20	0.37	0.03	0.27	0.70	0.60	-0.27	-0.03
q3	0.80	0.67	0.43	0.43	0.07	0.07	0.70	0.83	0.70	1.17	1.13	0.27	0.37

6. Appendix 6: Graphs of ROC analysis

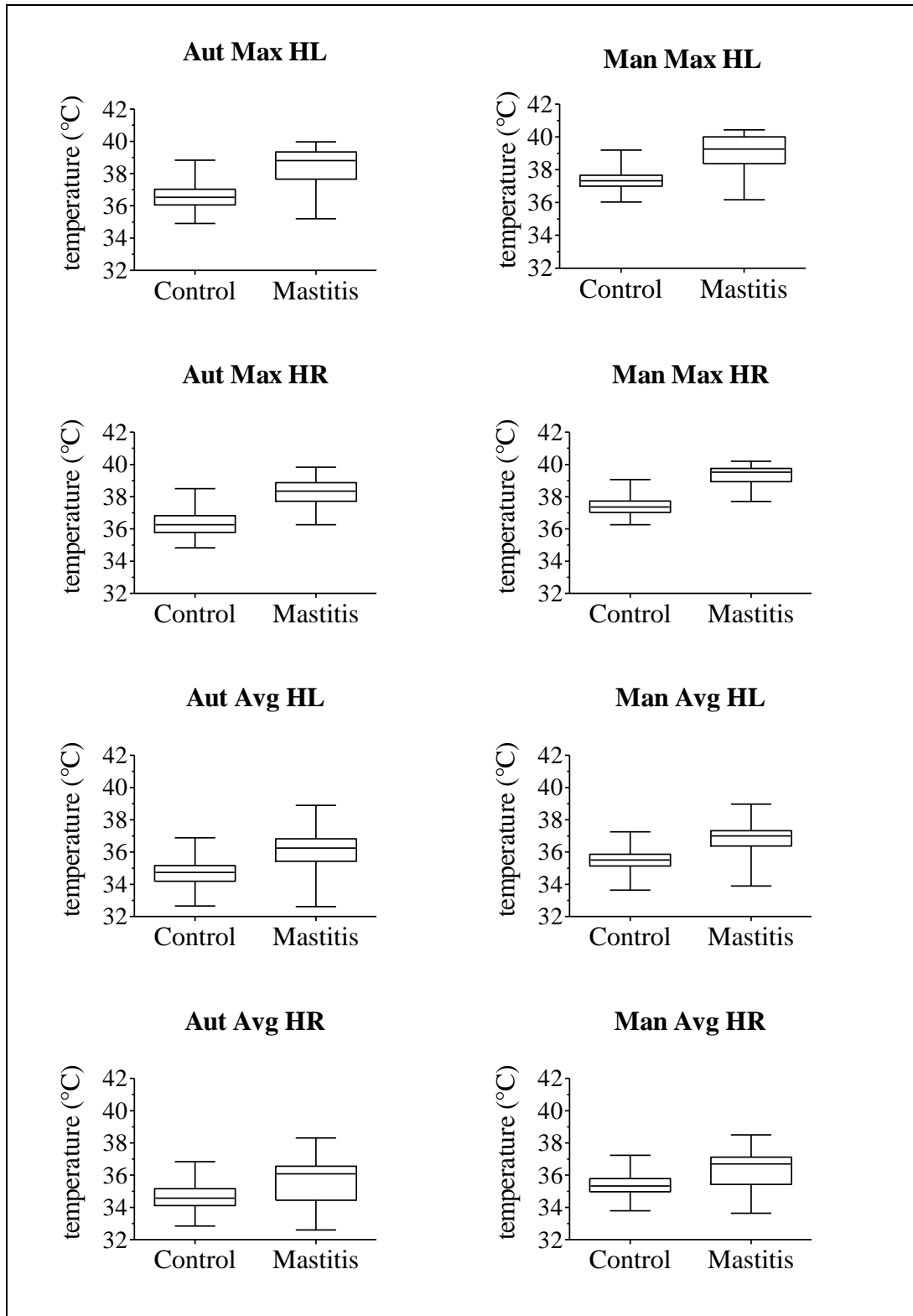


Figure 29: Box-Plot Graphs of temperature values in automatic ('Aut') and manual ('Man') evaluation in the clinically healthy udder ('Control') and in the udder suffering from clinical mastitis ('Mastitis'). The bars depict the range of the values, the boxes depict all values from first to third quartile. The line marks the median value.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

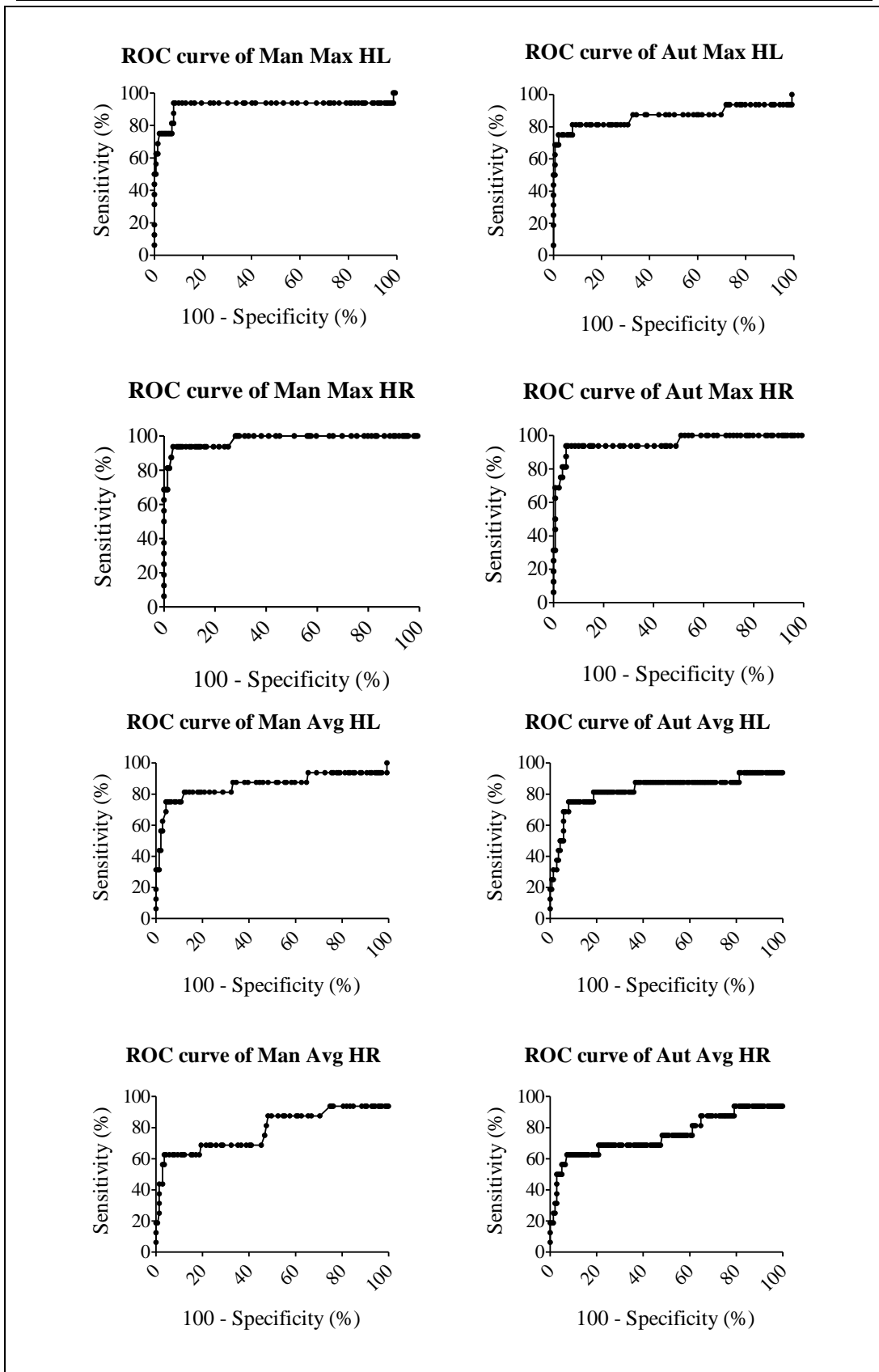


Figure 30: ROC (Receiver-Operating-Characteristics) curves of evaluation parameters in automatic ('Aut') and manual ('Man') evaluation.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

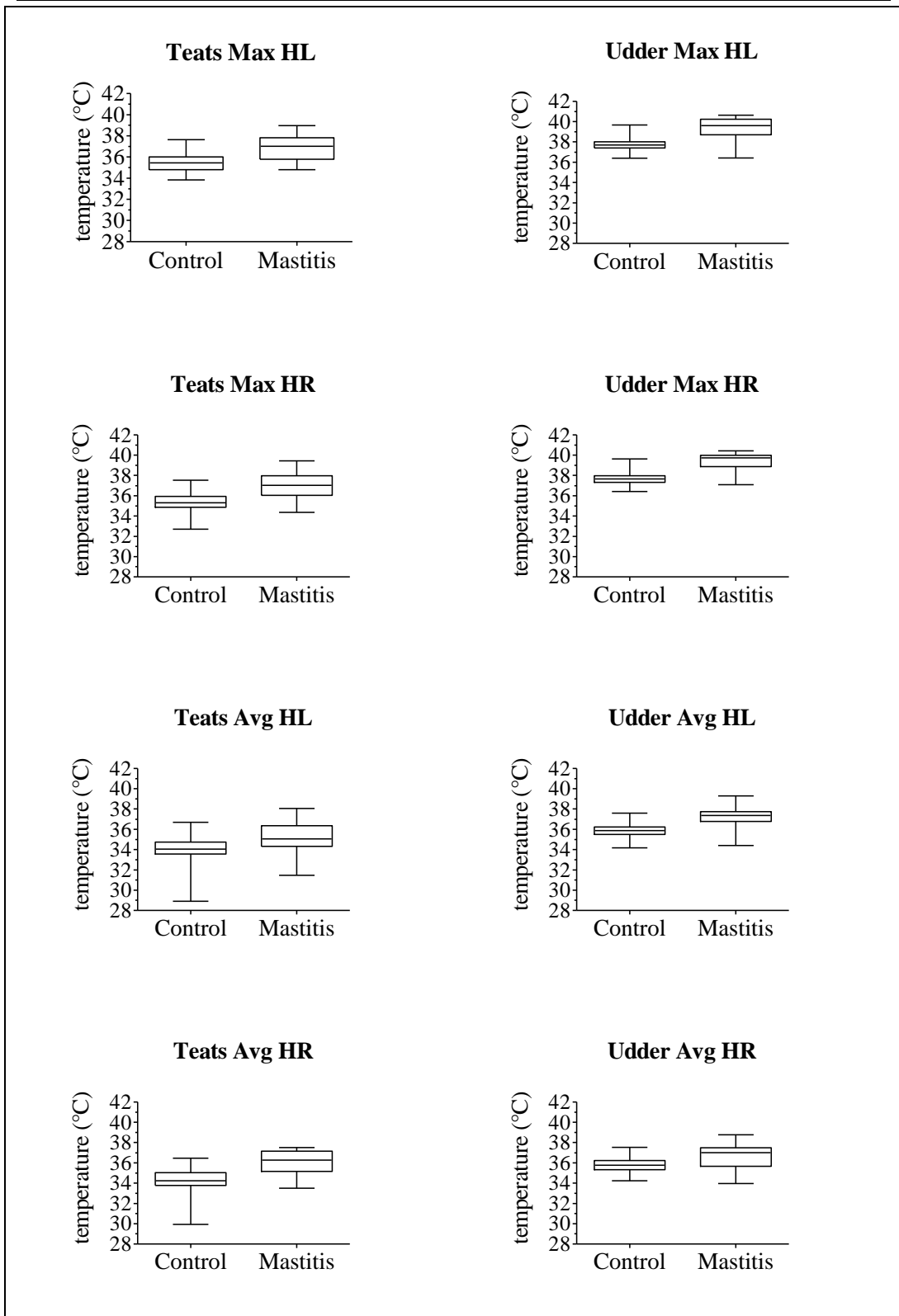


Figure 31: Box-Plot Graphs of temperature values in evaluation using the teats ('Teats') and evaluation using the udder ('Udder') in the clinically healthy udder ('Control') and in the udder suffering from clinical mastitis ('Mastitis'). The bars depict the range of the values, the boxes depict all values from first to third quartile. The line marks the median value.

'Max' – Maximum temperature values

'Avg' – Average temperature values

'HL' – Region of interest, left hindquarter

'HR' – Region of interest, right hindquarter

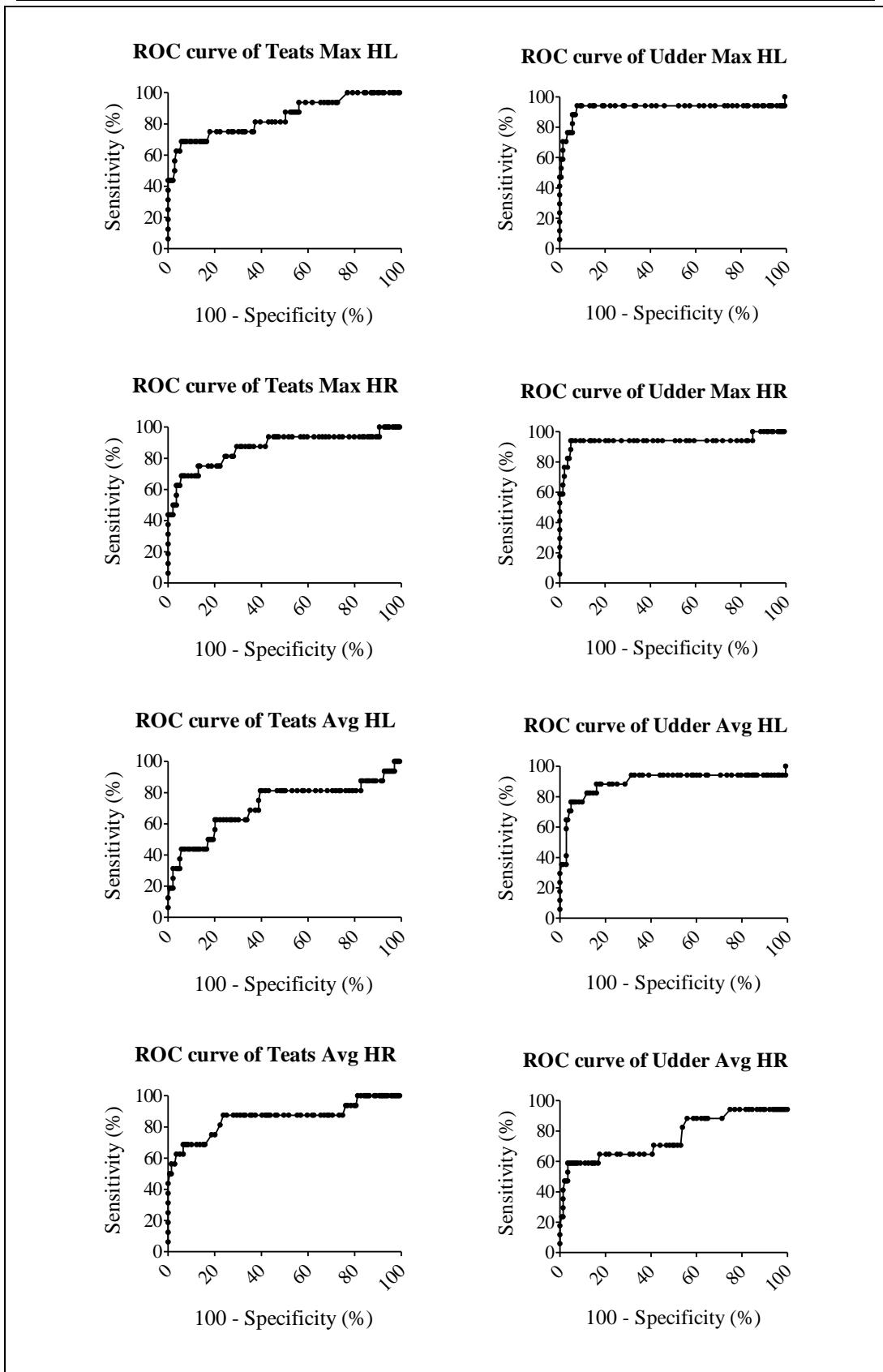


Figure 32: ROC (Receiver-Operating-Characteristics) curves of evaluation parameters in evaluation using the teats ('Teats') and evaluation using the udder ('Udder').
 'Max' – Maximum temperature values
 'Avg' – Average temperature values
 'HL' – Region of interest, left hindquarter
 'HR' – Region of interest, right hindquarter

XI. DANKSAGUNG

Ich möchte meinem Doktorvater, Herrn Prof. Dr. Zerbe, für die Überlassung dieses Dissertationsthemas und die Möglichkeit meine Arbeit auf zahlreichen Tagungen vorzustellen danken, sowie für die Bereitstellung meiner Arbeitsstelle als wissenschaftliche Hilfskraft an der Klinik für Wiederkäuer.

Mein besonderer Dank gilt auch Herrn Dr. Metzner, der mir das Vertrauen entgegengebracht hat an diesem Projekt selbstständig zu arbeiten, der aber ebenso für meine Fragen jederzeit erreichbar war und mir bei fachlichen und methodischen Anliegen umgehend unterstützend zur Seite stand. Ebenso möchte ich Dr. Wolfram Petzl für die gute Betreuung und die fachliche Expertise danken, ganz besonders dafür, dass ich zu jeder Zeit unmittelbar und unkompliziert Antworten auf meine Fragen bekommen habe. Ich habe mich während des gesamten Projektes ausgesprochen nett und kompetent betreut gefühlt.

Des Weiteren möchte ich Dr. Andrea Seemüller herzlichst für die Bereitstellung der ausführlichen Daten und Bilddateien ihres Dissertationsprojektes danken. Ebenso gebührt mein Dank Herrn Dr. Landgraf und Herrn Schröter, aus dem Fraunhofer-Institut für Verkehrs- und Infrastruktursysteme in Dresden, für die Zusammenarbeit und den Austausch von Daten. Die professionelle Entwicklung der Bilderkennungssoftware hat dieses Dissertationsprojekt überhaupt erst ermöglicht.

Ich möchte mich außerdem bei Anna Rieger für die geduldige und professionelle Unterstützung in statistischen Angelegenheiten bedanken, sowie für die oft kurzfristig vergebenen Termine.

Herzlicher Dank gebührt auch meinen Kollegen an der Klinik für Wiederkäuer, ganz besonders meinen Mitdoktoranden, für die freundschaftliche Zusammenarbeit und nicht zuletzt auch für die gegenseitige mentale Unterstützung.

Mein ganz besonderer Dank geht an meine Familie: An meine Eltern, die mir das Studium ermöglicht haben, an meine Schwestern, für ihren liebevollen Rückhalt und an meinen Freund Felix, auf dessen Unterstützung ich immer zählen kann.