Commission on Animal Management and Health **Session M 4.9**

55th Annual Meeting of the European Association of Animal Production
05. – 09. September 2004, Bled, Slovenia

Analysing serial data for mastitis detection by means of local regression

D. Cavero[1], K.-H. Tölle[1], C. Buxadé[2], and J. Krieter[1]

[1]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University of Kiel, D-24098 Kiel, Germany
[2]Department of Animal Production, ETSIA, Polytechnic University of Madrid (UPM), 28040 Madrid, Spain

Abstract

The objective of this study was to detect incidence of mastitis in an automatic milking system using serial information. Data from 109,739 milkings of the research dairy herd Karkendamm of the University of Kiel were available. The incidence of mastitis was determined by therapies carried out and by the weekly somatic cell count measurements. The time series of electric conductivity of quarter milk were analysed to find deviations as a sign for mastitis. Three methods were used to detect mastitis cases. First, a local regression method using the SAS-procedure LOESS, second a moving average system and third an exponentially weighted moving average were applied.

All methods provided similar results. The realibility of alerts varied dependent on the threshold value. A low threshold (3%) led to a sensitivity of nearly 100%, however the specificity was only about 36% and thus the error rate was high (about 70%). With increasing thresholds (7%) sensitivity decreased to 70% and specificity increased to 84%. Error rate was slightly reduced to 60%.

Introduction

Udder health makes up 15% of total losses, making it the second main reason for culling in commercial dairy farms in Germany (ADR, 2004). Product safety and animal welfare in dairy management systems can be improved by means of early detection of diseases such as mastitis.

Detection of mastitis can be automated by using sensor measurements. Milk electrical conductivity (EC) was already used in the last decade in computerised systems, that were developed in order to detect mastitis (HAMANN and ZECCONI, 1998; DVG 2002). EC is determined by the concentration of anions and cations. Where mastitis is present the concentration of sodium and chloride in milk increases while concentration in lactose and potassium decrease. As a result of this process the value of EC of milk from the infected quarter increases.

Relevant and reliable information is essential in order to manage a dairy farm. With conventional milking a lot of this information is obtained visually around and during a milking. While the AMS are equipped with features that collect large amounts of data, it is impossible to analyse at such a vast amount of data. This should later be transformed with the aid of appropriated software into useful information for management support.

The objective of this research is to develop an automatic mastitis detection system. Several algorithms will be used to connect the serial sensor measurements and information from the dairy farm in order to establish a model providing clear alerts for mastitis. Such a system would allow early detection of sick cows at an initial stage of the mastitis. This could reduce the losses caused by this disease.

Such management aid requires minimum labour and should identify the sick cows with a high accuracy. This will help the efficiency of farmer decisions, and therefore not only animal welfare but also economic decisions will be optimised.

Material and methods

The data was collected on the experimental farm "Karkendamm" of the Institute for Animal Breeding and Husbandry, Christian-Albrechts-University (Kiel, Germany). Data was recorded from July 2000 till September 2001. During this period 109,739 observations from 160 Holstein Friesian cows with 191 lactations were accumulated. Milking took place in an Automatic Milking System (AMS) with 4 boxes and an extra cleaning box from Westfalia Landtechnik GmbH, where the number of milkings per day could vary and the milking interval were irregular. In total the information of 44,074 cow-days were included in the study, where around 85% were originated from cows in first lactation. The mean days in milk was 158. The trait milk yield showed a mean of 11.31 kg per milking and the cows were selected in the AMS aprox. 2.5 times a day for milking. The treatments and diseases were recorded according to a detailed scheme.

Besides milk yield several on-line measurements during milking were available. These were milk temperature, electrical conductivity and milk flow rate. For this work, first of all the electrical conductivity of the milk is examined as an indicator for development of detection models.

Table 1: Number of observations (n), means ($\bar{x}$), standard deviations (s), minimum (min) and maximum (max) of the recorded traits

	n	$\bar{x}$	s	min	max
Milk yield (kg)*	105,518	11.31	3.77	0.40	30
Fat content (%)	14,745	3.98	0.98	1.52	8.50
Protein content (%)	14,745	3.4	0.30	2.24	4.92
E. conductivity (mS)*	108,260	5.3	0.57	2.09	8.50
Temperature (°C)*	108,063	36.4	0.82	31.0	42.0
SCC (1,000/ml)	14,298	165	402	3	9216

* per milking

Establishment of mastitis

Udder health was classified on the basis of the cows SCC, which was measured weekly from pooled quarter milk samples taken from each cow, as well as information on udder treatments. A total of 14,298 SCC tests were carried out with 165,000 cells/ml on average. The “Deutsche Veterinärmedizinische Gesellschaft e.V.” (DVG, 2002) stated a value of 100,000 cells/ml as threshold for mastitis. This threshold of 100,000 cells/ml is taken into account in this trial.

The milking days were organised as “healthy days“ or “diseased days“. If two succeeding SCC measurements either both exceeded the threshold or both did not, all days between these measurements were defined as “disease days” or “healthy days”, respectively. In the other case the day where the SCC was recorded and two days after and two days before were defined according to this SCC-value and the days in the middle were set to uncertain days. In this context a day were a treatment take place and two days before and two days later were set as “disease day”. Ten days after the last treatment were set to uncertain days. A mastitis-block is defined as uninterrupted sequence of “disease days”.

With this definition 571 mastitis blocks were found, with a mean mastitis length of 24.5 days. On Average 24.2 cows suffered from mastitis per day.

Univariate methods

On modern dairy farms routinely measured traits represent serial data. Three different univariate time series analyses for the trait electrical conductivity were compared. Recent observations contain more information than older ones about what will happen in the future. These procedures are based on the estimation of an expected value from the last available data, that afterwards would be compared with the real value.

In some Management-Information-Systems “Moving-Average-Models“ (MA) are already used to monitor udder health. With this procedure a new estimate (Y'_t) on each milking is worked out from the mean of the N last recordings, so that each milk recording has the same weight in the forecasted mean. The smooting effect of the moving average increases with the increasing number of considered observations in history. Analyses for the trait were performed with 5, 10, 20 and 30 values in history. In this investigation for size of history exemplarily a value of 10 was used.

$$Y'_t = \frac{1}{N}\sum_{k=1}^{N} Y_{t-k}$$

With the „Exponentially-Weighted-Moving-Average“ (EWMA) the expected value is calculated by taking all the preceding milk recordings into account. However, the closer previous values of the actual observation value are exponentially stronger weighted for the calculation. Through variation of a factor (α) the weighting can be changed. The higher the value of α, the stronger are weighted the last values. Small values of α mean a strength smoothness. Different weights were given to previous observations (α =0.2, 0.4, 0.6, 0.8). In this paper only an α-value of 0.2 was described.

$$Y'_{t+1} = \alpha \cdot Y_{t-1} + (1-\alpha)\cdot Y'_{t-1}$$

The third procedure is the local regression by means of LOESS-Method (COHEN, 1999), that is more descriptively known as locally weighted polynomial regression. This procedure performs a nonparametric method for estimating regression surfaces. It allows great flexibility because no assumptions about the parametric form of the regression surface are needed. At each point in the data set a low-degree polynomial is fit to a subset of the data, with explanatory variable values near the point whose response is being estimated. The expected value is calculated by a local regression parameter, which determines how many observations are used for the regression. The smaller the smoothing parameter is, the more closely the fitted line traces the data points. The level of closeness of the fitted line to the data points is estimated by iteratively testing different values of “smoothing parameter”. At the same time the individual time distance of the observations are taken into account. The polynomial is fit using a weighted least squares smoothing function, giving decreasing weights with increasing distance between history value and actual observation. This procedure was modified in such a way that only previous observations were used for forecasting the actual value. Once the best fitted line is determined with LOESS. The data value of the current milking is compared against the past historical LOESS values based on a set of predetermined criteria.

Test procedure

Model outcomes (alerts for mastitis) were compared with actual occurrences of mastitis. Each mastitis case was considered as a true positive case (TP) if one or more alerts were given in the first five days of this observed case, otherwise the case was false negative (FN).

For the evaluation the basis time cluster was one day. Each milking day outside a mastitis period was considered a true negative case (TN) if no alerts were given and a false positive case (FP) if an alert was given.

Afterwards the accuracy of these procedures were judged by the parameters sensitivity, specificity and error rate, as it was already used for evaluation of oestrus detection by FIRK et al. (2003).

The sensitivity represents the number of correctly detected mastitis cases:

$$\text{sensitivity} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \times 100$$

The specificity indicates the percentage of correctly found healthy days from all the healthy days:

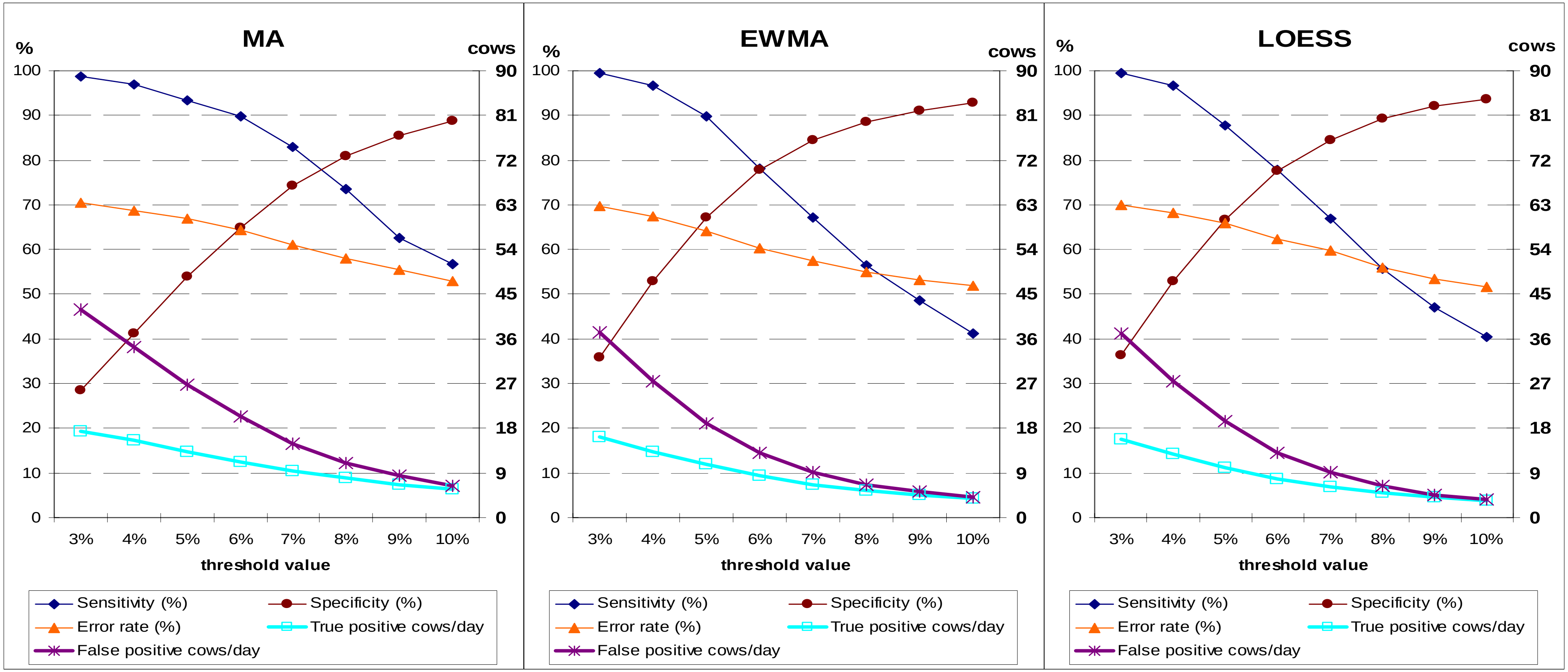
$$\text{specificity} = \frac{\text{true negative}}{\text{true negative} + \text{false positive}} \times 100$$

In order to give a better interpretation of these alerts, the error rate was defined as another indicator (here true positive cases referred to the alerts that was given inside a mastitis period). This rate represents the percentage of days outside the mastitis periods, from all the days where the threshold value were exceeded:

$$\text{error rate} = \frac{\text{false positive}}{\text{false positive} + \text{true positive}} \times 100$$

Results and discussion

Observations which caused an exceeding of the given threshold value were considered as mastitis warning. Comparing these alerts with the actual cases of mastitis the model performance could be assessed.



Detection has to take place within the first 5 days of the mastitis block, because it was not so important that all the days in a mastitis block were recognized, but it was decisive that a mastitis case was early detected. True positive and false negative cows/day means the number of right and wrong diseased registered cows per day respectively and thus directly the effort of the farmer for mastitis monitoring.

Regarding the parameters verifying the model, the three methods differed only slightly from each other. Sensitivity decreased with increasing threshold value (figure 1). The best results for sensitivity were reached by MA with 98.7 to 56.7 %. Lower sensitivities were calculated by EWMA and LOESS, with sensitivities between 99.6 and 40.5%.

Specificity, on the contrary, increased with increasing threshold value. Independent from the threshold value the lowest specificities were obtained by MA with 28.5 to 88.9 %. The best results were reached by EWMA and LOESS with sensitivities between 36.0 to 93.7%. The error rate decreased moderately with increasing threshold value. It ranged for the three methods between 70.4 and 51.6%.

The number of false positive cows/day had a stronger decrease than the number of the true positive cows/day using either of the methods. Slightly higher number of the true positive cows/day were calculated by MA, it ranged between 17.5 and 5.8 in comparison with the other methods, where it varied between 16.2 and 3.4. The number of the false positive cows/day by MA ranged between 41.8 and 6.5, 37.4 and 4.2 for EWMA and between 37.1 and 3.7 for LOESS.

One reason for the high error rates is the great number of "healthy days" compared with "disease days", so that the likelihood of the appearance of false reports in comparison with the right positive alerts is high. However the number of false reported cows per day decreases fast with increasing threshold value. Using a threshold of 5% the number of false positive registered cows is 10 animals per day, while the total number of the herd for the estimation is about 80 cows per day. Another reason for the large error rate is probably the measurement of the electrical conductivity. Only an improvement on the sensor technique could lead to reliable estimations. This information is not reliable enough to automatically monitor whether a cow suffers from mastitis. Most other investigations also show high error rates and low specificities, respectively. HAMMAN and ZECCONI (1998) reported an average sensitivity and specificity of 68 and 82% for detection of clinical mastitis based on electrical conductivity (the SCS threshold was 1,000,000 cells/ml or 500,000 cells/ml and presence of bacteria). VAN ASSELDONK et al. (1998) found (based on expert knowledge on a farm with activity measurements, electrical conductivity of quarter milk, automated concentrate feeders, milk yield and milk temperature measurements) that the sensitivity and specificity was 73 and 87% for clinical mastitis and 58 and 82% for subclinical mastitis respectively. MELE et al. (2001) obtained with application of the tracking signal method that the sensitivity was 65 to 83 % for clinical mastitis and 84% for subclinical mastitis (the SCS threshold was 300,000 cells/ml), while the specificity was 97%. In contrast to the present study and to other investigations DE MOL and OUWELTJES (2001) estimated a very high sensitivity (with a time-series model for milk yield and EC of milk and a SCC Threshold of 500,000 cells/ml), 100% and a rather high specificity, 87.4 to 98%.Different results between the studies are not only caused of distinctions between sensors and between statistical analyses. A further important reason is the definition of mastitis. Whereas mastitis was defined on the basis of treatments and on a threshold of 100,000 cells/ml, other authors defined a mastitis if threshold was important higher (500,000 cells/ml). More estimation of classification parameters depends on data basis. While specificity for mastitis was calculated by regarding all cows in this study, basis were only cows without any mastitis case during the test period in the studies of DE MOL and OUWELTJES (2001) and MELE et al. (2001).

Conclusion

- As is currently required, milk that enters the bulk tank must be from cows that are not visibly ill (EU, 2000), indicating that before the milk is stored a general health check has to be performed. The development of high-performance mastitis detection programs for effective controlling is important in connection with automatic milking systems, where there is no detection by visual observation in the milking barn during milking.
- The three univariate methods used showed similar results. That may be explained by an insufficient description of the trait electrical conductivity (is not enough correlated with our definition of mastitis, SCC).
- An decrease in sensitivity with increasing threshold leads to a decrease in specificity. An optimum level may be determined by minimizing the sum of the cost of false positive and false negative alerts.
- The univariate analysis of the trait electrical conductivity by the three methods was satisfying for sensitivity and specificity. However the corresponding error rate were too high.
- Approaches to reduce the false positive alerts are:
    - Further research to develop and to test multivariate models, where other additional variables could be taken into account.
    - An improvement of sensors in an AMS measuring more explanatory traits. Some of the extra information that will be recorded could be useful to detect mastitis and other abnormalities in the cow.

Acknowledgements

The authors are grateful to H. Wilhelm Schaumann foundation for financial support to this project. The technical assistance provided by the staff of the research farm Karkendamm (Kiel, Germany) is gratefully acknowledged.

References

ADR (2004): Rinderproduktion in der Bundesrepublik Deutschland 2003.

COHEN, R.A. (1999): An introduction to Proc Loess for Local Regression. Proceedings of the 24th SAS Users Group International Conference, paper 273.

DE MOL, R.M. (2000): Automated detection of oestrus and mastitis in dairy cows. Ph.D. Thesis Wageningen, The Netherlands.

DE MOL, R.M., OUWELTJES, W. (2001): Detection model for mastitis in cows milked in an automatic milking system. Prev. Vet. Med. 49, 71-82.

DVG (2002): Leitlinien zur Bekämpfung der Mastitis des Rindes als Bestandsproblem. Sachverständigenausschuss „Subklinische Mastitis".

EU (2000) http://europa.eu.int/eur-lex/en/com/pdf/2000/en_500PC0438_01.pdf

FIRK, R.; STAMER, E.; JUNGE, W.; KRIETER, J. (2003): Improving oestrus detection by combination of activity measurements with information about previous oestrus cases. Livest. Prod. Sci. 82, 97–103.

HAMANN, J.; ZECCONI, A. (1998): Evaluation of the electrical conductivity of milk as a mastitis indicator. Bull. IDF 334, 5–26.

MELE, M.; SECCHIARI, P.; SERRA, A. FERRUZZI, G.; PAOLETTI, F.; BIAGIONI, M. (2001): Application of the 'tracking signal' method to the monitoring of udder health and oestrus in dairy cows. Livest. Prod. Sci. 72, 279–284.

OUWELTJES, W. (2004): Demands and oportunities for operational management support. Operational management on farms with automatic milking systems. Deliverable D28, EU-project QLK5-2000-31006. www.automaticmilking.nl

VAN ASSELDONK, M.A.P.M., HUIRNE, R.B.M., DIJKHUIZEN, A.A. (1998): Quantifying characteristics of information technology applications based on experts Knowledge for detection of oestrus and mastitis in dairy cows. Prev. Vet. Med. 66, 273-286.