56th Annual Meeting of the European Association of Animal Production, 5/8 June 2005, Uppsala, Sweden
Commission on Animal Management and Health, Session M 6.7

Mastitis detection in dairy cows by application of fuzzy logic

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), E-28040 Madrid, Spain

Abstract

The aim of the present research was to develop a fuzzy logic model for classification and control of mastitis for cows milked in an automatic milking system. A data set of 403,537 milkings from 478 cows was accumulated. The dataset was divided in two data sets, training and test data.
Cases of mastitis were determined according to three different definitions by taking into account udder treatments and somatic cell counts.
The alerts were generated by a fuzzy logic model connecting electrical conductivity, milk yield, milk flow rate and time between milkings. Sensitivity for a mastitis case was set to be at least 80%. The parameters specificity and error rate were applied for evaluation of the reliability of the detection model. Specificities ranged between 93.9% and 75.8% and the error rate varied between 95.5% and 41.9% depending on mastitis definition. The results for the test data not only verified those for the training data, but mostly were also slightly better.
The obtained results were satisfying for sensitivity and specificity. However, the error rate still remained very high. Fuzzy logic was used to develop a detection model for mastitis that can be used in the future as an easy system for the farmer, without need of great expertise.

Introduction

Mastitis is the most costly disease of dairy cattle and remains one of the major problems concerning the dairy industry. Timely detection of mastitis is very important, not only because of the economic impact due to production loss, but also because of the negative effects on the animals' welfare. In herds with an Automatic Milking System (AMS), identification of udder infections is no longer based on visual observation. In contrast, control programs managing the health status of the cows have been introduced based on sensor measurements.
The cow status monitoring model is based on alarms arising at the milking robot when abnormalities are detected by a combination of the measured parameters. However, to be useful for the farmer, a detection system should not supply too many false alarms, which is a major problem concerning this model.
Fuzzy logic was already used for oestrus detection with good results (*Firk et al., 2003, Yang 1998*), moreover, it was also used to improve sensitivity and specificity of systems using conductivity as the main information source for mastitis detection (*de Mol and Woldt, 2001*).

Fuzzy systems nowadays are used to handle data because of cheap sensors and the possibility to store huge amounts of data. Fuzzy methods are applied to extract knowledge from data in a representative form that is very convenient and understandable even for people without a strong mathematical background (*Palm and Kruse, 1997*). In the present research a fuzzy logic model was developed for classification and control of mastitis for cows milked in an AMS.

Materials and Methods

Data were recorded at the experimental farm "Karkendamm" of the University of Kiel. Collection of data was performed between July 2000 and March 2004. During this period 403,537 observations from milkings from 478 Holstein Friesian cows with a total of 645 lactations were accumulated. The mean herd size was 124 cows on average per day. Milking took place in an AMS with 4 boxes. The average number of milkings per cow per day was 2.4 and the 305-day milk yield was 9,200 kg on average. Descriptive statistical information about the sensor traits ais shown in Table 1.

Table 1. Means ($\bar{x}$) and standard deviations (s) for the traits milk yield, milk flow rate, time between milkings and electrical conductivity.

Trait	Unit	Number of observations	$\bar{x}$	s
Milk yield	kg/milking	390,900	12.35	4.06
Average milk flow rate	kg/min	390,694	2.60	0.92
Highest milk flow rate	kg/min	390,517	3.90	1.38
Time between milkings	h	403,537	9,87	2,61
Electrical conductivity	mS/cm	398,326	5.38	0.57

The data set was randomly divided into two data subsets with different cows. Two thirds of the original data were training data and used to develop the fuzzy logic model. The other part of the data were test data and used to test if the developed model could be generalized.
In order to asses cases of mastitis, appropriate laboratory examinations of milk samples and observations by herdsmen and veterinarians were carried out. Mastitis treatments and weekly somatic cell counts (SCC) were the reference data used to evaluate the alerts. Several definitions of mastitis are found in literature, therefore three variants of mastitis definition were used in this investigation:

1) Treat: treatment performed without consideration of SCC,
2) Treat+100: treatment performed or a SCC > 100,000 cells/ml,
3) Treat+400: treatment performed or a SCC > 400,000 cells/ml.

Days were divided into days of mastitis and days of health, respectively. Two days before and two days after each treatment were classified as days of mastitis. In variant 1 all other days were set to days of health. Additionally, in variants 2 and 3 periods of two days before and two days after each SCC were graded analogous to the SCC thresholds. Days between two equally classified periods got the same category. Days between different periods got the status "unknown" and were not considered for estimation of classification parameters. An uninterrupted period of mastitis days was defined as mastitis block. Distribution of days of health and days of mastitis subject to definition of mastitis is shown in Table 2.
In this study a multivariate model was used to improve the mastitis detection. The sensor data were the input for the detection model. The alerts were generated by using a fuzzy logic system. This system performs a combination of conditions with electrical conductivity, milk yield, time between milkings and milk flow rate using MATLAB software (*2003*). The preparation of data and the calculation of the classification parameters were performed by using *SAS (2004).*

Table 2. Number of days of health, days of mastitis, or unknown days according to the three different considered mastitis definitions.

Training data	Days of mastitis	Days of health	Unknown	Mastitis cows/day	Healthy cows/day
1) Treat	651	109,690	4,307	0.48	80.5
2) Treat+100	37,719	68,538	8,391	27.7	50.3
3) Treat+400	6,607	102,476	5,565	4.85	75.2
Test data	Days of mastitis	Days of health	Unknown	Mastitis cows/day	Healthy cows/day
1) Treat	348	51,588	2,163	0.25	37.9
2) Treat+100	20,713	29,349	4,037	15.2	21.6
3) Treat+400	3,505	47,771	2,823	2.6	35.1

Parameters for evaluation of models

The traits for mastitis detection were recorded at each milking. If the output of the fuzzy logic models exceeded a determined value, the cow was reported to suffer from mastitis. Model outcomes were available for each milking, however, the tests of the models were calculated on daily basis. Each day inside a mastitis period was classified as true positive (TP) if one or more alerts were given and as false positive (FP) if no alerts occurred. Each day outside a mastitis period was classified as true negative (TN) if no alerts were given and as false negative (FN), otherwise. Accuracy of mastitis detection by the fuzzy logic model was expressed by sensitivity, specificity and error rate.

$$sensitivity = \frac{TP}{TP + FN} \times 100 \;;\; specificity = \frac{TN}{TN + FP} \times 100 \;;\; error\ rate = \frac{FP}{FP + TP} \times 100$$

It was expected that the detection model would supply alarms in the first stage of the mastitis period. Each mastitis case was considered as a true positive case (TPc) if one or more alerts were given in the first five days of a mastitis block, otherwise the case was false negative (FNc). The block-sensitivity represents the percentage of early mastitis cases detected correctly.

$$block\text{ - }sensitivity = \frac{TPc}{TPc + FNc} \times 100$$

For the calculation of specificity and error rate, data from all cows were taken into account, without considering whether the cow had been previously ill.

Fuzzy logic

Fuzzy logic translates a natural language knowledge into a formal mathematical modelling, so that it is suitable for computer processing (*Biewer, 1997*). The basic concept underlying fuzzy logic is that of a linguistic variable, a variable whose values are words rather than numbers. Although words are less precise than numbers, their use is closer to human intuition.
Unconventional modelling methods make better use of uncertain or imprecise data and vague knowledge about model components. The Fuzzy Set theory is based on an extension of the classical meaning of the term “set” and formulates specific logical and arithmetical operations for processing imprecise and uncertain information (*Zadeh, 1965*). Fuzzy sets can be used to handle uncertainty of data and fuzzy logic to handle inexact reasoning to produce a model based on human knowledge. The three steps of a fuzzy logic system are the fuzzification, fuzzy inference and the defuzzification.

Fuzzification

The first step is to transform the input variables into fuzzy values by the linguistic interpretation through membership functions and the grade of membership, with a range of [0,1]. Each trait is transformed in to a linguistic variable.

Fuzzy inference

The linguistic combination of the traits was performed in the fuzzy inference. The utilized rules result from human knowledge and have the form: if condition, then conclusion. The degree to which each part of the condition has been satisfied for each rule is known by the corresponding grades of membership. The outcome of combined traits in this investigation was the determination of the health status of the cow with the membership functions "sure mastitis", "very high probability of mastitis", "middle probability of mastitis", "low probability of mastitis" and "no mastitis".

Defuzzification

By defuzzification, fuzzy values were transformed into a single number, representing the real variable, e.g., a cow suffers from mastitis or not. The grades of membership, calculated in the fuzzification step, and the rules of inference result in special areas below the membership functions (μ) of the output variable. By calculation of the center of gravity of these areas, the fuzzy values are transformed back in order to resolve a single output value from the set:

$$x_o = \frac{\int_X x \cdot \mu(x) \cdot dx}{\int_X \mu(x) \cdot dx}$$

Results and discussion

The training data contained 127 cases of mastitis for variant 1 (Treat), 1,612 for variant 2 (Treat+100) and 620 for variant 3 (Treat+400). The test data contained 70, 736 and 322 cases of mastitis for variant 1, 2 and 3 respectively.

The aim of using training data was to develop a fuzzy model with sufficient accuracy. Therefore, the block-sensitivity was set to be at least 80%, thus the threshold for the value of fuzzy output for the alarm occurrence was selected for each variant (Table 3). The other important performance criterion was reliability. The parameters specificity and error rate were applied for evaluation of the reliability of the detection model.

As shown in Table 3, the specificities were 94.0%, 77.5% and 89.1% and the error rates were 96.1%, 46.5% and 77.2% for variants 1,2 and 3 respectively. For the test data specificities were 93.9%, 75.8% and 88.1% and the error rate were 95.5%, 41.9% and 75.7% for variants 1, 2 and 3 respectively. The results obtained for the test data not only verified those for the training data, but were also slightly better.

Table 3. Classification parameters of mastitis detection from the training data and test data by the fuzzy logic models using the information: electrical conductivity, milk yield, milk flow and time between milkings.

Training data	Threshold	Block-sensitivity	Specificity	Error rate	TP cows/day	FP cows/day
1) Treat	0.64	81.1	94.0	96.1	0.2	4.8
2) Treat+100	0.36	80.1	77.5	46.5	12.9	10.2
3) Treat+400	0.50	81.2	89.1	77.2	2.4	8.0
Test data	Threshold	Block-sensitivity	Specificity	Error rate	TP cows/day	FP cows/day
1) Treat	0.64	92.9	93.9	95.5	0.1	2.4
2) Treat+100	0.36	83.2	75.8	41.9	7.2	5.2
3) Treat+400	0.50	83.9	88.1	75.7	1.3	4.1

Additionally true positive and false negative cows/day were determined, that means the number of cows per day classified rightly and wrongly as diseased respectively, and thus directly farmers' effort for mastitis monitoring. The number of TP cows/day for the training data were 0.2, 12.9 and 2.4 and the FP cows/day were 4.8, 10.2 and 8.0 for variants 1, 2 and 3, respectively. The average herd size for the training data was 84.2 cows/day, in consequence around 6%, 12% and 10% of the herd would be identified wrongly as diseased. For the test data the number of TP cows/day were 0.1, 7.2 and 1.3 and the FP cows/day were 2.4, 5.2 and 4.1 for variants 1, 2 and 3. The average herd size for the test data was 39.7 cows/day, therefore 6%, 13% and 10% of the cows were erroneously detected as diseased.

Three variants of mastitis definition were used in this investigation. Each definition contains advantages and disadvantages. On the one hand, the main disadvantage in variant 1 is that there may be cows being ill, but not considered as such. This results in a higher probability of FP, since there may be alarms, although the cows may not be considered as ill, therefore, resulting in higher error rates. Moreover, there is also a higher probability of TN, since fewer cows are considered ill, most negatives are true. Consequently, the specificity is higher for this variant. On the other hand, the variant 2 (Treat+100), which is the definition recommended by the *DVG (2002)*, is the most stringent definition. Cows with relatively low cell counts are considered to have mastitis, but other changes may not occur to the cow, so that clear changes on the parameters do not take place. This means a smaller proportion of TN and higher TP, leading to a low specificity and a low error rate. The third variant, could be discussed as an intermediate case.

The block-sensitivity was calculated for the first five days of mastitis (reference mastitis block). The length of the block was chosen because an early detection of the disease is critical and, moreover, because these are the days were more parameter changes occur. Therefore, the block-sensitivity was considered more relevant than the sensitivity, which was calculated for each day of the disease period.

The basis for the evaluation of the performance of the mastitis detection is the knowledge of the actual status of the cow on each day of observation, therefore the choice of the length of the reference mastitis block is crucial. In fact, the sensitivity would increase significantly if longer periods were considered. For instance, *Mele et al. (2001)* took 7 days for clinical and 10 days after and 10 days before for subclinical mastitis and *De Mol et al. (1997)* took 10 to 7 days for clinical mastitis and 14 days before and after for subclinical mastitis.

The main problem in the actual study was the error rate, which was too high. Probably, among other factors this was due to that the sensors are still not well developed. Moreover, the available parameters or how they are measured may not be adequate. In addition, the fact that there are many more days of health than days of mastitis causes a greater probability for FP to arise, which has an impact on the error rate.

The application of fuzzy logic gives the model the advantage of being easy to interpret and easy to modify and adapt, by changing the membership functions and the bases of the rules. The main problem of developing the fuzzy logic models was the appropriate choice of suitable membership functions. The optimal shape of the membership functions was found by trial-and-error attempts. The obtained results were satisfying for sensitivity and specificity, the specificity ranged between 93.9 for variant 1 and 75.8 for variant 2. However, the error rate still remained very high, only for the variant 2 error rate was smaller than 50%.

Sensitivity and error rate obtained with fuzzy logic in the current research were better than those estimated with univariate methods (*Cavero et al., 2004*). In that study, for a block-sensitivity about 80% the sensitivity was 38% and the error rate was 60.4% for a mastitis definition analogous to variant 2.

In order to be able to compare results among different studies, it would be convenient to establish standard parameters to detect the disease as well as the use of a standard definition of the condition itself.

Conclusion

The automatation of detection of mastitis in farms with AMS can be a promising alternative to visual observation. Fuzzy logic was used to develop a detection model for mastitis, that can be used in the future as an easier system for the farmer, without need of great expertise. With fuzzy logic models, better results are found than with the univariate methods. The error rates are nevertheless high. But this fact could be attributed to errors arrising from data instead of a problem in the model itself and from incertain mastitis definition. However, further studies are needed to confirm its applicability.

References

- **BIEWER, B., 1997**. Fuzzy-Methoden: Praxisrelevante Rechenmodelle und Fuzzy-Programmiersprachen. Springer, Berlin Heidelberg.
- **CAVERO, D., TÖLLE, K.H., KRIETER, J., 2004**. Analysing serial data for mastitis detection by means of local regression, 55. EAAP, Bled, Book of Abstracts (10) 153, M4.9
- **DE MOL, R. M., KROEZE, G.H., ACHTEN, J.M.F.H., MAATJE, K., ROSSING, W., 1997**. Results of a multivariate approach to automated oestrus and mastitis detection. Livest. Prod. Sci. 48, 219-227.
- **DE MOL, R. M. WOLDT W. E.** Application of Fuzzy Logic in Automated Cow Status Monitoring J. Dairy Sci. 84 (2001): 400 –410
- **DVG, 2002**. Leitlinien zur Bekämpfung der Mastitis des Rindes als Bestandsproblem. Sachverständigenausschuss "Subklinische Mastitis".
- **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.
- **MATLAB 2003.** MathsWorks, Fuzzy Logic Toolbox, for use with MATLAB© 6.5.1. User's Guide.
- **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.
- **PALM R., AND KRUSE, R., 1997**. Methods for data analysis in classification and control. Fuzzy Sets and Systems 85, 127-129.
- **SAS, 2004.** SAS/STAT User's Guide (Release 8). SAS Institute, Cary NC, USA.
- **YANG, Y., 1998**. Rechnergestützte Östrusüberwachung bei Milchkühen unter Anwendung der Fuzzy-Logic-Methode. Herbert Utz Verlag, Wissenschaft, München.
- **ZADEH, L.A., 1965**. Fuzzy sets. Information and Control 8, 338-353.