Commission on Cattle Production, Free Communications, Session 35, CPh6.6
56th Annual Meeting of the European Association for Animal Production (EAAP),
June 5-8, 2005, Uppsala, Sweden

Oestrus detection in dairy cows using control charts and neural networks

J. Krieter (jkrieter@tierzucht.uni-kiel.de) , E. Stamer and W. Junge

Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, D-24098 Kiel

Abstract
Exponentially weighted moving average control charts and neural networks were used for oestrus detection in dairy cows. The analysis involved 373 cows, each with one verified oestrus event. Model inputs were the traits activity, measured by pedometer, and the period (days) since last oestrus. In total 10,386 records were available, which were partitioned into training and validation subsets to train and test the neural network (multifold cross-validation). When the trained neural network was applied to the validation sets, the averaged sensitivity, specificity and error rate were 77.5, 99.6 and 9.1%, respectively. Performance for the same data with the univariate control chart was less successful. Neural networks are useful tools to improve computerised oestrus detection in dairy cows.
Keywords: Oestrus detection, Dairy cattle, Control chart, Neural network

1. Introduction

Modern dairy farming is generally characterised by extended herd sizes and narrowed income margins per unit of output. In consequence, the economic results become more and more sensitive to minor changes in farm performance. Reproductive efficiency, e.g. oestrus detection, rebreeding and calving interval, has a strong impact on farmers' income. Therefore, modern farming imposes increasing demand for reliable management systems supporting the decision of the farmer. Schofield et al. (1991) denoted the trait activity as easy to obtain and strongly correlated with the incidence of oestrus. The detection rate of 78% as calculated by Maatje et al. (1997) confirmed the suitability of the trait activity for oestrus detection, but the error rate of 32% indicated a high number of falsely detected oestrus. A review about the automation of oestrus detection in dairy cows is given by Firk et al. (2002).

In order to improve oestrus detection, multivariate analyses with activity, milk yield, milk temperature, electrical conductivity and flow-rate were performed (Yang, 1998; De Mol et al., 1997; Firk et al., 2003a). But none of the presented combinations showed an appreciable improvement in error rate. Firk et al. (2003b) combined the trait activity and the period since last oestrus into a fuzzy logic model and observed a strong reduction in the number of false positive warnings.

The first objective of the present study was to develop a neural network model to classify oestrus alerts. Desirable properties of neural networks recommending their use for classification or signal processing are supervised learning and adaptivity. The natural architecture makes it useful in adaptive pattern classification or signal processing. Multiple layers with nonlinear transfer function allow the network to learn nonlinear and linear relationships between model inputs and outputs. A network response can also be, to a degree, robust and fault tolerant against noise and distortion and extract the underlying pattern (Haykin, 1999). The second objective of this research was to compare the classification performance of the neural network with more conventional methods from statistical quality control, i.e. control charts.

2. Material and methods

2.1 Material

The collection of data was performed on a commercial dairy farm in 1998. During this period 373 inseminations were verified as oestrus cases by a following calving. Accordingly, each oestrus originated from one cow. In order to have a reference for evaluation of the oestrus detection model, oestrus detection was performed on the basis of these objectively correct oestrus cases. For oestrus detection time series consisting of 15 days before oestrus, the day of oestrus and 15 days after oestrus were analysed. The selected period of 15 days before oestrus resulted from the knowledge that the duration of the mean oestrus cycle is 21 days, with a variation from 18 to 23 days (Sambraus, 1978). Accordingly, in this period no additional oestrus cases, which could influence the obtained results, were expected.

The trait activity was measured by conventional pedometers, which were attached to the left foreleg of each cow. The pedometers were recorded at the entrance of the rotary milking parlour twice daily for cows with high milk yield or in early lactation thrice daily. The activity values for further analyses were calculated from the difference between two successive pedometer readings, divided by the period of time between these readings. In agreement with Firk et al. (2003a), who indicated daily activity values as superior to values for each milking, a daily activity value was calculated for each cow.

The parameter “period since last oestrus” included information about previous inseminations and previous oestrus cases. In contrast to De Mol and Woldt (2001) no difference was made between previous insemination and oestrus cases, because the distinction between both cases is only denoted by the decision of the farmer to breed the cow or not. For each day considered in the analyses the period since last oestrus was calculated from the difference of the actual day and the day of previous information. In Fig. 1, the distribution for the trait period since last oestrus is given for 94.9% of cows with previous information. Another 5.1% of cows showed periods since last oestrus of longer than 75 days. As expected, most observations accumulated around the mean oestrus cycle length of 21 days. A smaller accumulation occurred for a period of doubled oestrus cycle length. Further cycles were not identifiable.

2.2 Methods

Exponentially weighted moving-average control chart (EWMA)

A control chart is a simple time plot of a sequence of observations or subgroups statistics (Fig. 2). The observations in the plot are compared to upper and lower control limits determining the range of variation due to common causes. If the process is in-control, nearly all observations fall between the control limits. A point outside of the control limits indicates an out-of-control signal, so more variation exits than can be attributed to the effect of common causes of variation (e.g. oestrus alert). In the present investigation, an EWMA control chart was used because it is flexible, easy to set up and operate. The EWMA utilises previous observations, the weight attached to data exponentially declines as observations get older and older. The EWMA is defined as

$$W_{X,t} = \lambda X_t + (1-\lambda) W_{X,t-1} \tag{1}$$

where $X_t \sim N(\mu_t, \sigma_X^2)$ and $W_{X,t}$ denotes the EWMA statistic at time t, usually $W_{X,0}$ is set equal to a target value (Montgomery, 1997). The parameter λ is a constant satisfying $0 < \lambda \leq 1$. The choice of λ determines the decline of the weights and therefore the memory of the chart. If $\lambda \rightarrow 1$, the EWMA puts all of its weight in the most recent observations. If $\lambda \rightarrow 0$, then the most recent observations are assigned a small weight and the weight attached to previous observations only slightly decreases with time.

The variance of $W_{X,t}$ is given by

$$\sigma^2_{W_{x,t}} = \sigma^2_X \left(\frac{\lambda}{2-\lambda} \right) \left[1 - (1-\lambda)^{2t} \right] \tag{2}$$

Using this expression the upper (UCL) and lower (LCL) control limits are as follows

$$UCL_t, OCL_t = \mu \pm L\sigma \sqrt{\left(\frac{\lambda}{2-\lambda} \right) \left[1 - (1-\lambda)^{2t} \right]} \tag{3}$$

The EWMA chart provides an out-of-control signal if the realisation of $W_{X,t}$ is larger then UCL_t or smaller than LCL_t. L is a constant with $L > 0$. The design parameters of the chart are the constants L and λ that need to be chosen by the designer of the control chart. In the present study the values of L and λ varied in order to determine the performance of the control chart.

Neural networks (NN)

The Multilayer perceptron which was used in the present study is the most widely used, studied and applied NN. In these networks there is a set of input nodes (input layers), whose only role is to feed input patterns into the rest of the network (Fig. 3). Following the input layer, and before the output layer, there are one or more intermediate layers of units. These units are called hidden units because they have no direct connection to the outside environment neither input nor output. In the feedforward networks there are no connections leading from a unit to units in a previous layer, nor to other units in the same layer nor to units more than one layer ahead. The output of every unit is connected only to the units in the next layer. Every unit is associated with a nonlinear function called the activation function. A commonly used form of nonlinearity is a sigmoid nonlinearity defined by the hyperbolic tangent

$$f(X) = \frac{e^X - e^{-X}}{e^X + e^{-X}} \tag{4}$$

Thus the output of each node j in the network is given by equation (4) where its input X is given by

$$X_j = \sum_i b_{ij} O_i \tag{5}$$

where i runs for all nodes in the previous layer. The values O_i are the outputs of the units in the previous layer, and b_{ij} are the weights associated with the node connections (Fig. 3). More details about the construction of NN are found in Haykin (1999) and Patterson (1997).

Once the network weights and biases has been initialised the network has to be trained. NN have a task associated with the optimisation of a defined error function. Usually this is defined as

$$E = \frac{1}{2} \sum_p \sum_k (T_{pk} - O_{pk})^2 \tag{6}$$

where k, p belongs to the sets of output nodes and training patterns respectively. The values O_{pk} and T_{pk} are the actual network outputs and the training sets target outputs respectively, of the nodes k in the output layer for the training pattern p.

The backpropagation algorithm is a popular algorithm for the training of a multilayer perceptron. In our study the training process followed a modified Levenberg-Marquardt algorithm (Bayesian regularisation, Foresee and Hagan, 1997) in order to minimise equation (6) and to produce a network that generalises well. The backpropagation algorithm was stopped when the absolute rate of change in the averaged squared error per iteration was sufficiently small. Finally, by comparing convergence, consistency and classification accuracy, a multilayer perceptron with one hidden layer was adopted. The input layer contained two nodes (activity, days since last oestrus), the hidden layer consisted of five nodes and the output layer had only one node (oestrus event yes/no).

Evaluation and validation

The classification performance of the EWMA chart and NN can be tested by analysing the number of correctly and incorrectly classified observations using the form of an error matrix (Congalton, 1991). The error matrix is a square array and consists of the number of classified examples: true positive (TP), false negative (FN), false positive (FP) and true negative (TN). The classification performance is expressed by the sensitivity, specificity and error rate. The sensitivity measures the number of correctly detected oestrus to all oestrus events.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \cdot 100 \tag{7}$$

The specificity denotes the number of false oestrus warnings in relation to number of true negative observations.

$$\text{Specificity} = \frac{TN}{TN + FP} \cdot 100 \tag{8}$$

The error rate describes the number of false oestrus warnings in proportion to the number of detected oestrus alerts.

$$\text{Error rate} = \frac{FP}{FP + TP} \cdot 100 \tag{9}$$

Multifold cross validation was used to evaluate the ability of the trained NN to accurately classify oestrus events. The available set on N examples was divided into $M = 5$ subsets. A NN model was trained on all the subsets, except for one, and the performance of the model was measured by testing it on the subset left out. The same training and validation subsets were utilised to derive the EWMA chart. The performance of the models was assessed by averaging sensitivity, specificity and error rate under training or validation over all trials of the experiment.

Preprocessing of data, deriving the EWMA control chart and building the neural network with training and validation was performed with MATLAB Version 7.0.1.24074 (2004).

3. Results and discussion

Using the EWMA control chart and the whole data set the sensitivity ranged from 63.9 to 90.3%, the error rate varied between 19.7 and 55.8% (Fig. 4). The specificity was always high (> 93.2%, not presented) due to the high number of the true negatives results. By moving the control limits further from the center line (L=3), the risk of an observation falling beyond the control limits, indicating an oestrus event when no oestrus is present, decreased. Widening the control limits also reduced the sensitivity, i.e. the number of correctly detected oestrus on all oestrus. If the control limits moved closer to the center line, the opposite effects were obtained: the sensitivity and error rate increased.

The second design parameter λ also determined the performance of the control chart. If λ increased, the sensitivity declined and the error rate was lower. Thus, larger λ-values diminish the memory of the control chart and more weight is given to the more recent observation which is appropriated when larger shifts should be detected (as the case with oestrus detection). Larger λ-values also widen the control limits, i.e. the sensitivity and error rate decreased.

Fig. 5 demonstrates the pattern of the mean squared error of the NN model for a selected training set according to the number of epochs elapsed. An epoch corresponds to the presentation of the set of training vectors to the network and the calculations of new weights

and biases. In the early period the mean squared error declined quickly. When the number of epochs increased the absolute rate of change in the mean squared error was small. Convergence was assumed if the rate of change was lower then 0.01% per epoch (approximately 300 epochs).
Using the NN with two input nodes (activity, period since last oestrus) for oestrus detection, sensitivity was 78.7% and error rate was 5.1% with the training sets (Table 1). The performance of the model was assessed by averaging the classification parameter under validation. The sensitivity remained more or less constant (77.5%), but the number of false oestrus warnings in relation to the number of detected oestrus increased (9.1%). These results confirmed that the model provided an adequate fit and generalised well. Compared to the univariate EWMA control chart with $\lambda = 0.6$ and L ranging from 2.5 to 3, sensitivity was only slightly enhanced, but an obvious improvement was found in the reduced number of false positive oestrus warnings. Using the same data set (n=10,386) Firk et al. (2003b) also observed a strong improvement in the error rate if the trait activity and period since last oestrus were combined by a fuzzy logic model (senitivity = 87.9%; error rate = 12.5%). In contrast De Mol and Woldt (2001) could not find any improvement by considering previous oestrus information. If the input of the NN model was restricted to the trait activity, the differences between the EWMA chart and the NN model were small (sensitivity = 76.7%, error rate = 15.2% using the training set; sensitivity = 75.3%, error rate = 17.8% using the validation set) indicating the benefit of previous oestrus detection.

4. Conclusions

A neural network model was developed for oestrus detection using the activity measurements and the period since last oestrus. A feedforward three-layer perceptron provided an adequate fit and generalised well. Oestrus detection by a conventional univariate EWMA control chart was less successful.
Neural networks learn from experience, generalise from previous examples to new ones and extract essential characteristics from inputs containing noisy data. Once trained, a network response can be insensitive to minor variation in its input. Additionally, their possible hardware implementation due to their inherently parallel nature makes them ideal for real time application. Therefore neural networks are useful tools for computerised decision support systems in order to improve oestrus detection, but also mastitis evaluation in dairy cows.

References

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35-46

De Mol, R.M., Woldt, W.E., 2001. Application of fuzzy logic in automated cow status monitoring. J. Dairy Sci. 84, 400-410.

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.

Firk, R., Stamer, E., Junge, W., Krieter, J., 2002. Automation of oestrus detection in dairy cows: a review. Livest. Prod. Sci. 75, 219-232

Firk, R., Stamer, E., Junge, W., Krieter, J., 2003a. Oestrus detection in dairy cows based on serial measurements using univariate and multivariate analysis. Arch. Tierz., 46, 127-142

Firk, R., Stamer, E., Junge, W., Krieter, J., 2003b. Improving oestrus detection by combination of activity measurements with information about previous oestrus cases. Livest. Prod. Sci. 82, 97-103

Foresee, F.D., Hagan, M.T., 1997. Gauss-Newton approximation to Bayesian regularization. Proceedings of the 1997 International-Joint Conference on Neural Networks, 1930-1935

Haykin, S., 1999: Neural networks. A comprehensive foundation. Prentice Hall International, New Jersey, 842 pp.

Matlab, 2004. The MathWorks, Inc., Version 7.0.1.24074 (Release 14).

Maatje, K., Loeffler, S.H., Engel, B., 1997. Predicting optimal time of insemination in cows that show visual signs of oestrus by estimating onset of oestrus with pedometers. J. Dairy Sci. 80, 1098-1105.

Montgomery, D.C., 1997. Introduction to statistical quality control. John Wiley & Sons, 677 pp.

Patterson, D., 1997. Künstliche Neuronale Netze: das Lehrbuch. Prentice Hall Verlag GmbH, Haar bei München, 506 pp.

Sambraus, H.H., 1978. Nutztierethologie. Verlag Paul Parey, Berlin und Hamburg, 315 pp.

Schofield, S.A., Phillips, C.J.C., Owens, A.R., 1991. Variation in the milk production, activity rate and electrical impedance of cervical mucus over the oestrus period of dairy cows. Anim. Reprod. Sci. 24. 231-248.

Yang, Y., 1998. Rechnergestützte Östrusüberwachung bei Milchkühen unter Anwendung der Fuzzy – Logic – Methode. Herbert Utz Verlag, Wissenschaft, München.

Table 1: Classification performance (%) of the neural network using training and validation set

	Training sets[1]			Validation sets[1]		
	Sensitivity	Specificity	Error rate	Sensitivity	Specificity	Error rate
Neural network	78.7	99.5	5.1	77.5	99.6	9.1
EWMA chart[2]						
L = 3.0	71.3	99.3	17.1	66.9	99.3	18.8
L = 2.5	77.3	99.2	18.1	70.6	99.1	20.9

[1] means of replications, M = 5 subsets
[2] $\lambda = 0.60$

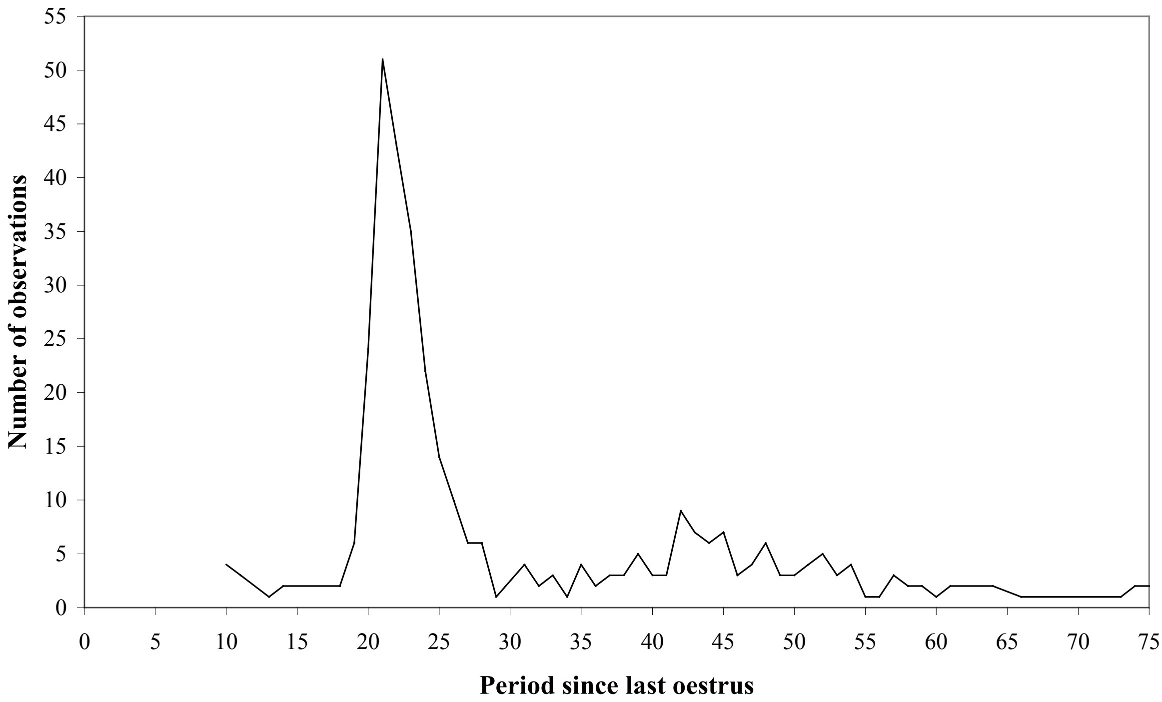

Figure 1: Distribution of the trait period since last oestrus (n = 373)

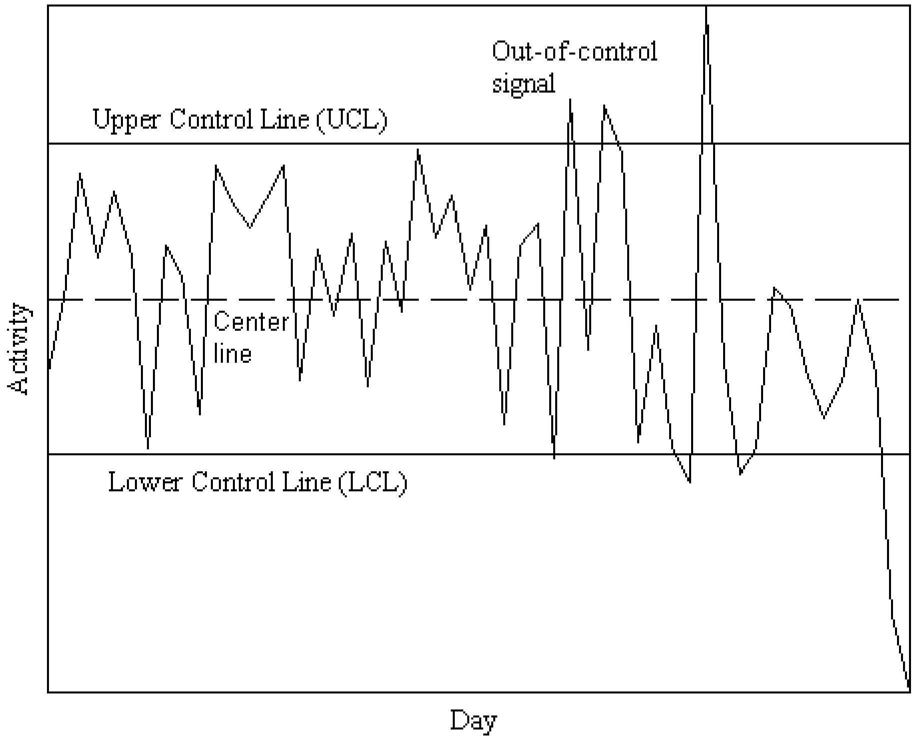

Figure 2: Illustration of a control chart

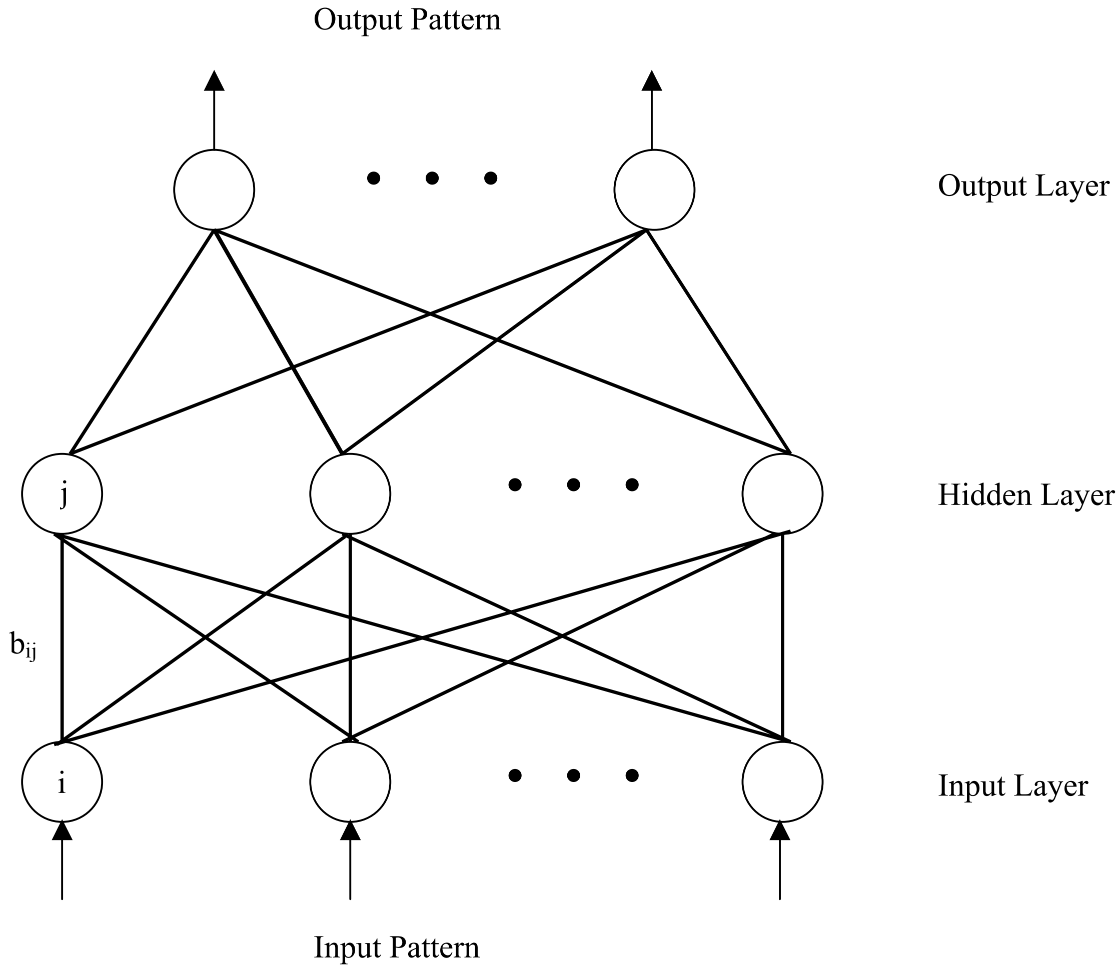

Figure 3: An example of a multilayer feedforward neural network

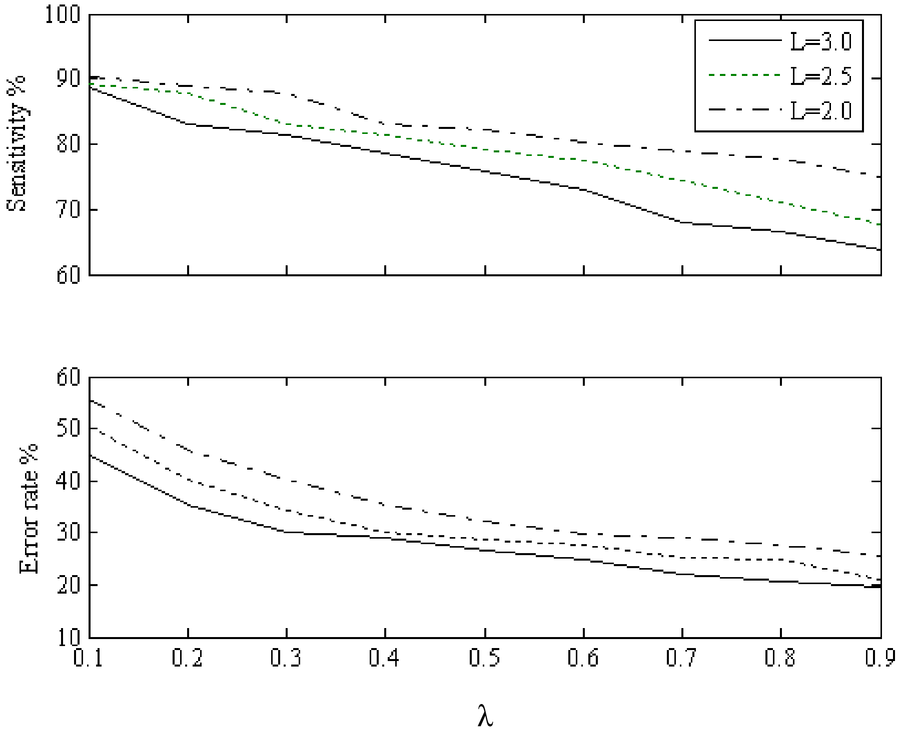

Figure 4: Classification performance of the EWMA control chart depending on parameters λ and L (n = 10,386)

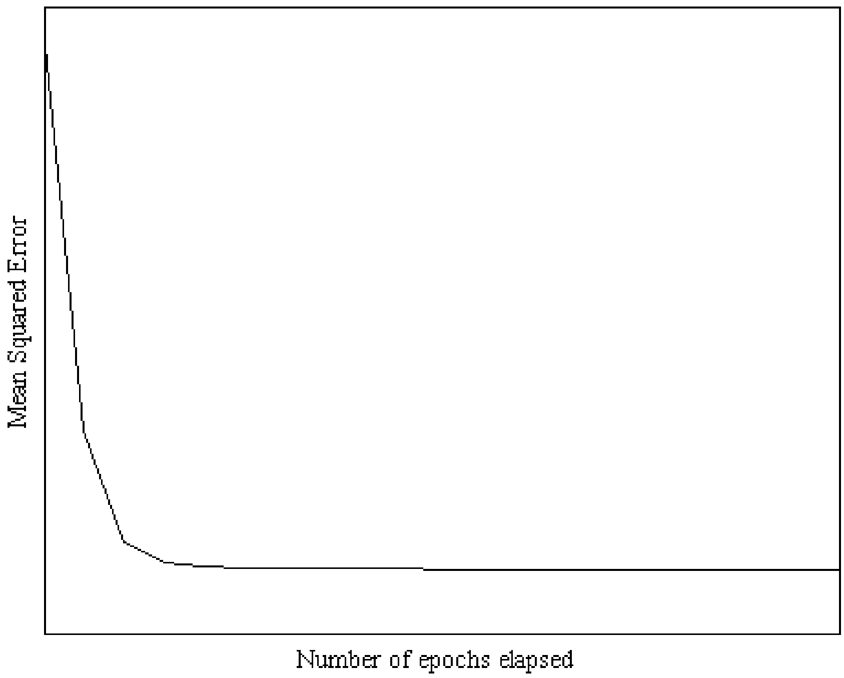

Figure 5: Mean squared error depending on epochs elapsed (approximately 300)