Institut für Tierzucht und Tierhaltung der Christian-Albrechts-Universität zu Kiel, Germany[1]
PIC Deutschland, Schleswig, Germany[2]

JOACHIM KRIETER[1], KATRIN KIRCHNER[2], JULIA ENGLER[1] and
KARL-HEINZ TÖLLE[1]

Computer-based analysis of sow herd performance

Dedicated to Prof. Dr. Dr. h.c. mult. Ernst Kalm on the occasion of his 65th birthday

Abstract

Increased herd sizes and narrowed income margins are common characteristics of modern swine farming. Therefore good management is becoming more and more important for the economic results. The paper describes a computer-based weak-point analysis of individual sow herd performance. Three stages were distinguished: (1) Tracing deviations between farm performance and a given standard in order to detect trends in the production process. It was shown that modified exponentially weighted moving average control charts are an effective tool in detecting small performances shifts. (2) Weighting the deviations by calculating the statistical and economic relevance allows the ranking of different traits independent of scales and units. (3) Finding the causes for the performance shifts. Decision tree algorithm was investigated to gain more insight in the critical points of production. A decision tree starts with the root node representing the traits which mostly influenced the target attribute (critical point), followed by internal nodes. The generated graphical decision trees are transparent and the outputs are easy to interpret for the farm manager or the consultant. Methods were applied to simulated and real sow herd datasets.

Key Words: pigs, sow, management systems, control charts, decision trees

Zusammenfassung

Titel der Arbeit: **Computergestützte Schwachstellenanalyse in der Ferkelerzeugung**
Die künftige Ferkelerzeugung ist durch größere Bestände und sinkende Erlöse gekennzeichnet, selbst marginale Änderungen im Produktionsprozess wirken sich nachhaltig auf die ökonomischen Ergebnisse der Betriebe aus. Daher sind effiziente Kontrollstrategien für das Betriebsmanagement unerlässlich, um Schwachstellen frühzeitig zu identifizieren und zu beseitigen. Der vorliegende Beitrag befasst sich mit der Entwicklung einer computergestützten Schwachstellenanalyse in der Ferkelerzeugung. Drei Stufen werden unterschieden: (1) Identifizierung von Abweichungen im Produktionsverlauf: „Control Charts" stellen effektive Werkzeuge dar, um frühzeitig auch marginale Änderungen im Produktionsprozess aufzuzeigen. (2) Skalierung und Gewichtung der Abweichungen: An Beispielen wird demonstriert, wie mit der statistischen und ökonomischen Relevanz der Abweichungen die Bedeutung der unterschiedlichen Merkmale rangiert werden kann. (3) Analyse der Schwachstellen: Das Entscheidungsbaumverfahren ist ein wirksames Instrument, die Ursachen für die Abweichungen aufzuklären. Die einzelnen Elemente des Baumes werden als Wurzel, Äste, Blätter und Knoten bezeichnet. Ausgehend von der Wurzel, über die Knoten bis hin zu den Blättern werden Entscheidungsregeln aufgelistet und in ihrer Bedeutung gekennzeichnet. Die Methoden wurden an simulierten und realen Datensätzen angewendet.

Schlüsselwörter: Schwein, Sau, Managementsysteme, „Control Charts", Entscheidungsbaumverfahren

Introduction

Commercial swine farming has changed considerably in the last decades. In Germany, for instance, the total number of swine breeding farms decreased from 77,000 in 1994 to 44,000 in 2001 (ZMP, 2003). Within these years, the average number of sows per

farm increased, in 1994 40% of sows were kept in herds with more than 100 sows, in 2001 this portion amounted to 64%. The structural change was accompanied by decreasing revenues per unit of output (ZMP, 2003). The total profit of individual farms has come to depend more and more on production farm performance. Small shifts in farm performance can have a great impact on farmers' income (KRIETER, 2002). Therefore the demand for effective management information systems is growing with the increasing demands on farmer's management skills.

In many farms, information technology enables farmers to collect, treat and process farm data at individual animal level. But in most instances, effective computerised tools for tracing deviations from target specification (identification of strong and weak elements) and further analysis of deviations finding the underlying causes are not available. This paper presents a systematic approach for computer-based farm analysis. It can be used for early detection of weak elements and supports the decision making of the farmer.

General concept

Computer-based analysis of individual sow herd performance requires accurate and consistent data. These data are normally provided by an information system used on the farm. The primary objective of individual farm analysis is to exploit strengths and eliminate or improve weak elements. False positive signals for the farmer should be minimised and real problems indicated. Further examinations of weak points should analyse the underlying causes. Therefore, analysis of individual sow herd performance requires a systematic approach (HUIRNE, 1990).

Process control needs standards for comparisons. If the standards are derived from the farm itself the analysis is called internal farm analysis and if several years are included, the analysis is called "trend analysis" (HUIRNE, 1990). Standard values are the historical performance of the farm or target specification by the management. If standards are derived from other (but similar) farms the analysis is called external or "comparative analysis" (HUIRNE, 1990), the main objective is to determine the relative position of the farm.

Both internal and external analyses imply three steps: (1) identification of relevant deviations, (2) weighting deviations and (3) further analysis of deviations. The following chapters present methods and describe how these tools form a cohesive, practical framework for individual farm analysis. The presentation here only focuses on internal analysis, but methods can be easily extended for external analysis.

Identification of relevant deviations

In any production process a certain amount of inherent or natural variability always exists (Fig. 1a). This natural variability or "background noise" is the effect of many small causes. A process that operates with only random variation present is said to be in statistical control, the chance causes are an inherent part of the process. Other kinds of variability may occasionally be present in the output of the process: for instance, errors by the staff, unsufficient climatisation. These sources of variability that are not part of the chance cause pattern are called "assignable causes". A process that operates in the presence of assignable causes is said to be out-of-control (MONTGOMERY, 1997). A major objective of process control is to quickly detect the occurrence of relevant shifts (deviations) in the process so that investigation of the process and

corrective action may be undertaken before problems cumulate. Control charts are an on-line processcontrol technique widely used in the industry for this purpose. A typical control chart is shown in Figure 1b, which illustrates the pattern of total piglets born per litter over a time period of 52 weeks.
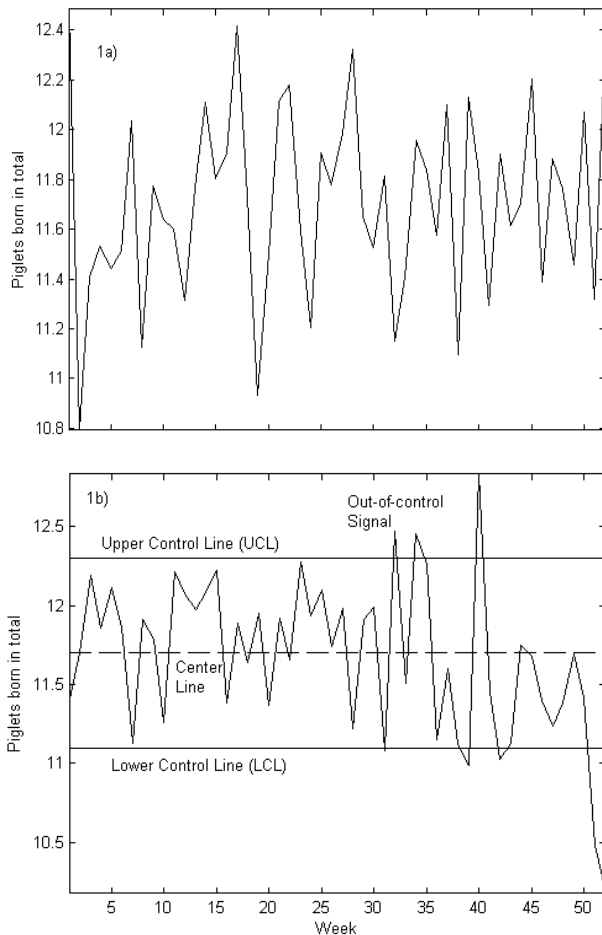

Fig. 1: Illustration of a control chart (Aufbau eines Control Chart)

The chart contains a center line that represents the target specification of the process. Two other horizontal lines, called the upper control limit (UCL) and the lower control limit (LCL) are also shown in the chart. These control limits are chosen so that if the process is in-control, nearly all of the observations range between them and no corrective action is necessary. A shift of the production process outside the control limits indicates that the process is out-of-control and investigations are required to improve the process.

Control charts can be found in a number of versions adopted to different monitoring and requirements. The Shewhart chart is the oldest and still most frequently applied among the control charts. The Shewhart control chart uses only the information concerning the last plotted observation ignoring any information given by the entire sequence of observations. Therefore the Shewhart chart is relatively insensitive to small shifts. There are two very effective alternatives which may be used when small deviations are of interest: the cumulative-sum (cusum) controlled chart and the exponentially weighted moving-average (EWMA) control chart (MONTGOMERY, 1997). The cusum charts use the unweighted sum of all previous observations, this

chart has a rather long memory. DE VRIES and CONLIN (2003) applied cusum control charts to oestrus detection in dairy cows. In EWMA control charts the process is monitored using a weighted mean of all previous observations. The weights decline exponentially as the observations get older and older. The EWMA was used in the present paper.

Exponentially weighted moving-average (EWMA) control chart
The EWMA is defined as

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1}$$

where λ is a constant satisfying $0 < \lambda \le 1$ (MONTGOMERY, 1997), x_i represents the sequence of independent observations and z_i denotes the EWMA statistic at time i. EWMA utilises all previous observations, but the weight attached to data declines exponentially as the observations get older. If $\lambda = 0.2$ then the weight assigned to the current mean is 0.2, and weights given to the preceding values are 0.16, 0.128, 0.1024, and so forth.

If the observations x_i are independent random variables with variance σ^2, then the variance of z_i is

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) \left[1 - (1 - \lambda)^{2i} \right]$$

and the UCL and OCL are given as

$$UCL_i, OCL_i = \mu \pm L\sigma \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) \left[1 - (1 - \lambda)^{2i} \right]}$$

The constant L with $L>1$ determines the width of the control chart. Specifying the control limits is one of the critical decisions that must be made in designing a control chart. By moving the control limits further from the mean or target value, the risk of type I error decreases – that is the risk of an observation falling beyond the control limits, indicating an out-of-control condition when no cause is present. However, the risk of a type II error increases – that is the risk of an observation falling between the control limits when the process is really out-of-control. If the control limit is moved to the target specification, the opposite effect is obtained. If data is autocorrelated and the limits of the EWMA chart are widened, negative autocorrelation makes the control chart very insensitive, positive autocorrelation will result in many false out-of-control signals (WIERINGA, 1999). WIERINGA (1999) proposed several methods for accounting serial correlation in control charts, one is to modify the EWMA control limits with an adopted variance (for details see WIERINGA, 1999).

In the EWMA control chart, the control limits depend on λ and L. MONTGOMERY (1997) emphasises that in general values of λ in the interval $0.05 \le \lambda \le 0.25$ and $L = 3$ work well in industrial application. The performance of the control chart can be measured by the average time to signal (ATS). The ATS is the average number of time periods that occur until a signal is generated. It is desirable for the farmer to have a low ATS if the process is out-of-control. To gain more insight into the optimal EWMA designs in agriculture production processes without any side-effects, a simulation study was started.

Example 1 - simulated datasets

The trait considered was piglets born in total per litter over a time period of 52 weeks. Each week, 100 litters were generated, the number of replications was limited to 100. The mean value (center line) of piglets total born was 11.7±2.5. From week 32, a negative shift of size 0.1-0.2 σ in the mean was generated. EWMA control charts were derived for weekly subgroups, x_i were replaced by $\bar{x}$ and σ with $\sigma_{\bar{x}} = \sigma\sqrt{n}$ in the previous equation. The ATS and the false positive rate (FPR, false positive signals/n*100) were calculated to evaluate the EWMA designs (Fig. 2).

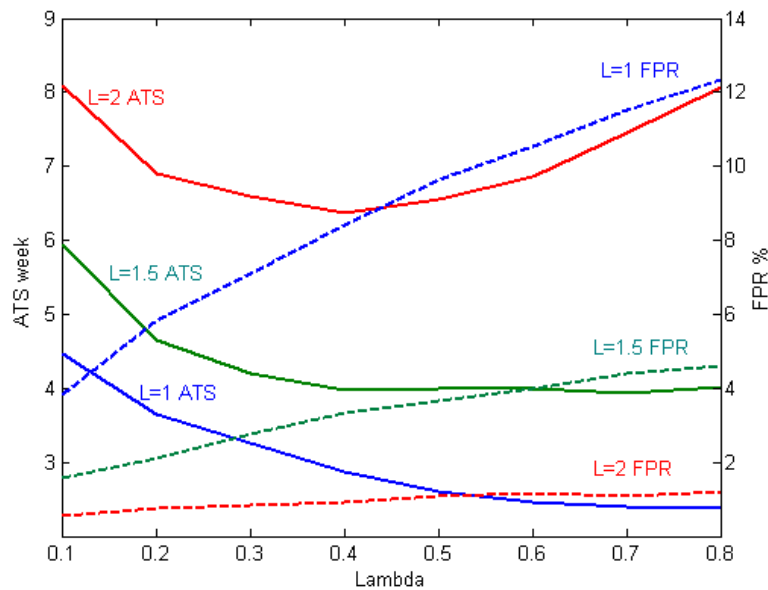Fig. 2: Performance of EWMA control charts depending on smoothing parameter λ, width of control limits L and on the shift in the mean of 0.1σ (ATS = average time to signal, weeks; FPR = false positive rate, %) (Leistungsfähigkeit des EWMA Control Chart in Abhängigkeit von dem Glättungsparameter λ, den Kontrollgrenzen L und der Veränderung des Mittelwertes von 0.1σ; ATS = mittlere Zeit bis zum Alarm, Wochen; FPR = Falschpositivrate, %)

If *L* increased from 1 to 2 the number of false positive signals decreased because the *LCL* moved away from the target value (shift in the mean of 0.1σ), the FPR varied between 1 and 14%. ATS ranged from 2 (*L*=1) to 8 (*L*=2) weeks because the probability of an observation falling beyond the control limits was reduced with higher *L*-values. If λ → 1, the EWMA placed all of its weight on the most recent observations. ATS was slightly reduced and the number of false positive signals increased. For the shift in the mean of 0.2 σ ATS declined to 1.5 and 4.7 weeks, the FPR was reduced to 0.2 and 9%.

In conclusion, varying the smoothing parameter has only a slight impact on the ATS and FPR. A good rule is to use smaller values of λ to detect smaller shifts. The parameter *L* strongly affects the performance of the EWMA schemes. For the given scenarios an L-value of 1.5 diminished the ATS with an acceptable FPR.

Example 2 – real datasets

In example 2 the EWMA control chart was applied to real pig farming datasets from a breeding herd with 1,000 sows. The data sets consisted of 4,342 litters over 24 months. EWMA control charts were calculated for the number of piglets weaned, the target value was 9.8. In Figure 3a, the process fluctuates randomly around the target value. In

Figure 3b, the process wanders away form the target specification. The ATS was 8.3 weeks. In this case, management action would be necessary to improve the process.
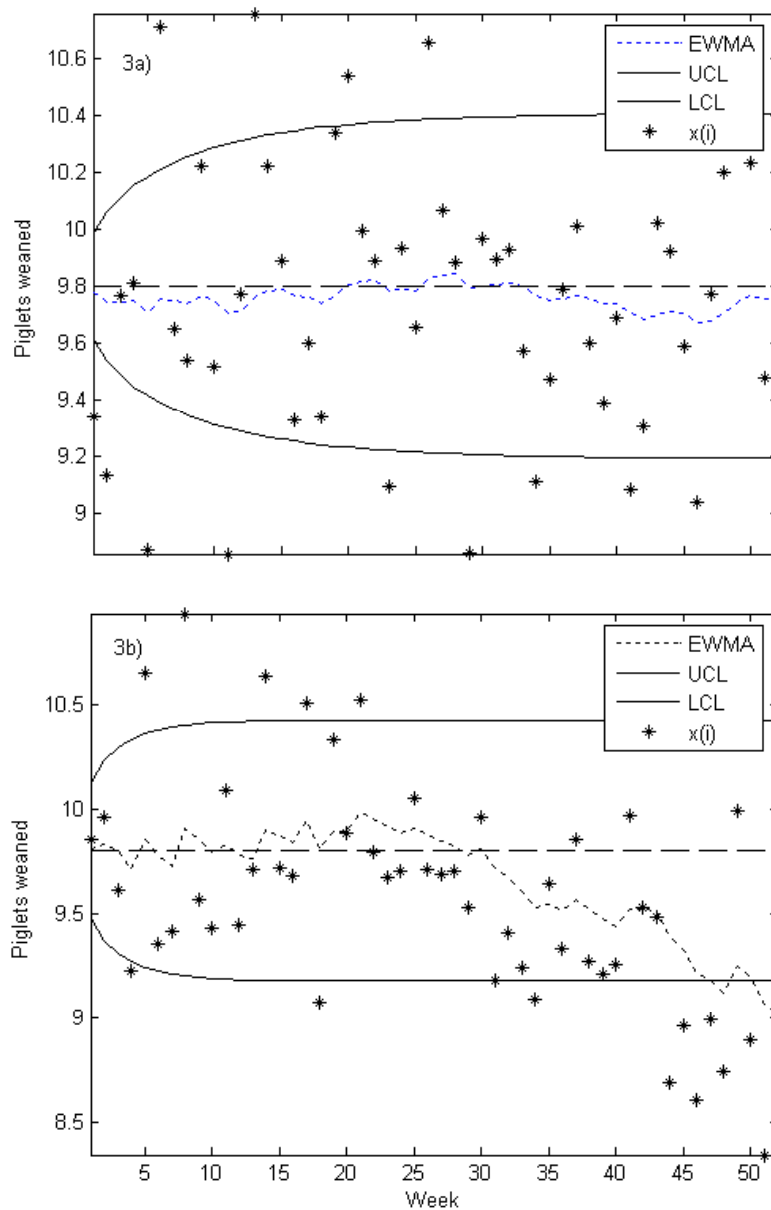

Fig. 3: EWMA control charts with random fluctuation (3a) and shift in the mean (3b), real pig dataset n=4,324 (EWMA Control Chart mit zufälliger Variation (3a) und Trend (3b), realer Datensatz)

Weighting deviations

As stated above, all shifts or deviations between actual performance and standards are initially assessed in their original dimensions. Additionally, the economic importance of one unit deviation will vary between variables, depending on their impact on total economic farm importance. By calculating the relevance of deviations (Hurine, 1990) all deviations are converted to the same units and comparisons can be made. The relevance of a deviation of variable i (RD_i) accounts for the economic (EI_i) and statistical importance (SI_i)

$$RD_i = EI_i * |SI_i|$$

SI$_i$ is the statistical importance of a deviation in variable i and is determined by

$$SI_i = TD_i / \sigma_i$$

TD$_i$ is the traced deviation and σ_i the standard deviation of variable i. The economic importance of a deviation equals the difference between the base and the new value of the total economic farm performance. An example, adopted from HUIRNE (1990), is given in Table 1.

Table 1
Relevance of deviations for some variables (HUIRNE, 1990) (Bedeutung der Abweichungen für ausgewählte Merkmale)

Variable i[1]		AV$_i$	SV$_i$	TD$_i$	|SI$_i$|	EI$_i$[2]	RD$_i$[2]
Litters	pspy	2.01	2.15	0.14	1.40	-44.96	-62.95
Piglets born alive	pl	10.80	10.67	0.13	0.26	7.75	2.01
Fe pigs sold	pl	8.86	9.05	-0.19	0.32	-13.35	-4.22
Pig feed	pp	25.00	29.04	-4.04	1.68	27.08	45.50

[1] pspy = per sow per year, pl = per litter, fe = feeder, pp = per pig 2) Euro

The actual value (AV) of litters per sow per year was 2.01 and standard value (SV) 2.15. The traced deviation amounts to -0.14 and the statistical importance is 1.40. The product of the economic (-44.96) and the statistical importance yields the relevance of the deviation (-62.95). The concept of the relevance of traced deviation enables a ranking of variables because all deviations are on the same scale.

Further analysis of deviations (shifts)

The final step in individual farm analysis is the further analysis of the relevant deviation in order to find the underlying causes of changes and weak points. The computerised detection of weak elements in a production process requires an analysis of the relationships between target criterion and other traits (attributes) which are usually provided by an information system used on the farm. Data Mining methods are useful tools for checking these relationships in large datasets. Data Mining has been used extensively in e.g. medical diagnosis, marketing and credit approvals (KAMBER et al., 1997), its application in animal production has been limited. DeWAR and McQUEEN (1995) tried to calculate the optimal replacement strategy of dairy cows, MITCHELL et al. (1996) analysed oestrus events in sows, PIETERSMA et al. (2003) investigated dairy lactation curves. KIRCHNER et al. (2004a,b; 2005) applied decision tree techniques to simulated and real pig farming datasets.

Decision tree building is one of the machine learning tools which belong to the Data Mining methods. The decision tree-based methods expressed their results in graphical presentation of decision rules. A decision tree contains a root node, internal nodes representing the attributes, branches characterising the attributes values and leaves expressing the binary decision. The examples and analysis presented in this paper use the C4.5 algorithm for generating decision trees (QUINLAN, 1986, 1993). Trees were calculated with the open source program package WEKA 3-2-3 developed at the University of Waikato (New Zealand).

In Figure 4, the data flow of the decision tree computing is shown (KIRCHNER et al., 2004b). In phase I, the raw datasets were pre-processed and controlled for plausibility and missing values. Phase II describes the construction of the model. The C4.5 algorithm performs the top-down induction of the decision tree on the basis of a

training set. The descending order of the attributes within the tree is calculated by the gain-ratio criterion. The procedure to classify the observations (instances) starts with the determination of the root attributes, followed by tests on further attributes to build the subordinate nodes. Furthermore, the algorithm calculates the split values of the attributes represented by the branches. The branches end in the leaves indicating the classification of the decision they present (QUINLAN, 1993).

After generating the tree, an error-based pruning method is used to simplify the tree by discarding one or more sub-trees and replacing them with leaves or branches (QUINLAN, 1993). In phase III, the generated model is tested with respect to its explanatory power with the stratified n-fold cross-validation method. The whole dataset is portioned randomly into n subsets and the C4.5algorithm runs for n times. For each time, different training sets and test sets are used and the results are validated. The classification accuracy is calculated using the "confusion matrix". This matrix consists of the numbers of true positive (TP), false negative (FN), false positive (FP) and true negative (TN) classified instances. The classification performance of a decision tree is measured by the sensitivity ($SE = TP / (TP + FN)$), specificity ($SP = TN / (TN + FP)$) and error rate ($ER = FP / (FP + TP)$).
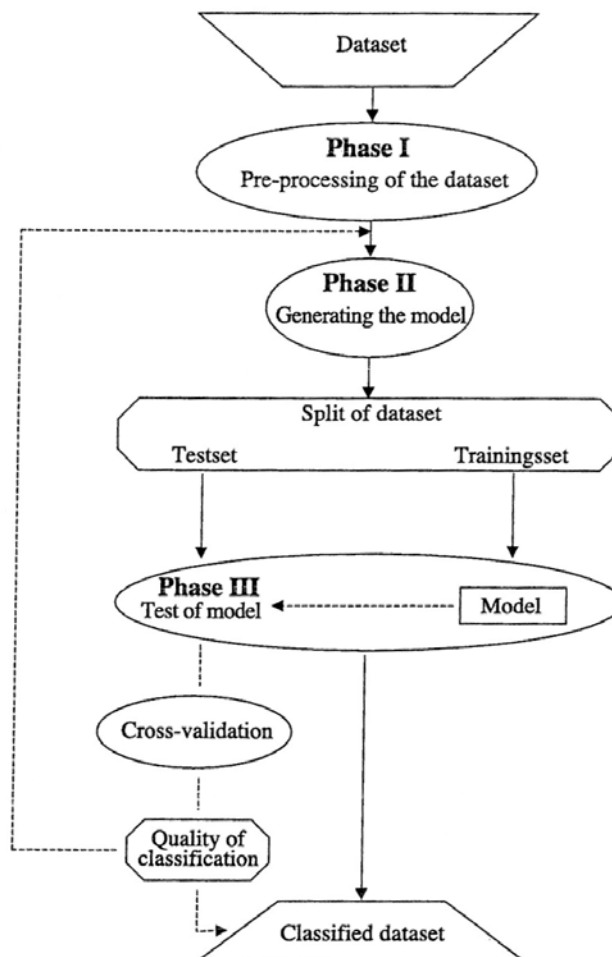

Fig. 4: Flow chart for developing a decision tree (KIRCHNER et al., 2004b) (Flow Chart für die Entwicklung eines Entscheidungsbaums)

Example 3

Decision tree technique was used to detect threshold values of management decisions relating to sows' replacement. In order to generate side-effect-free data, three pig

porduction herds, each at a different performance level, were created using Monte Carlo simulation (Table 2).

Table 2
Means of selected performance parameter calculated by the simulation program for different herds (Ausgewählte Merkmale für unterschiedliche Betriebsleistungsklassen)

Variable		Level of pig production systems[1]		
		Low	Medium	High
Number of litters	nl	4	4	4
Piglets born in total per litter	nbt	11	12	13
Piglets stillborn per litter	nsb	2	1	1
Piglets born alive per litter	nba	9	11	12
Piglets weaned per litter	nw	8	10	11
Weaning to oestrus interval (days)	woe	8	6	5
Re-breeding (%)		20	15	10

[1] values are rounded

Each herd contained 500 sows. Selection of a number of sows for culling was based either on (1) fertility problems (number of matings, me; weaning to oestrus interval, woe), (2) clinical problems (locomotion, diseases, peripartum problems, sudden death), (3) low performance (production index), or (4) age (number of litters, nl). The accuracy of the classification at the three herd performance levels is illustrated in Table 3.

Table 3
Accuracy of the classification at three sow herd performance levels (Genauigkeit der Klassifizierung für unterschiedliche Betriebsleistungsklassen)

Dataset – Performance	Sensitivity %	Specificity %	Error rate %	Number of leaves	Number of Nodes
Low	72.7	98.8	6.0	10	19
Medium	60.3	98.0	13.9	14	27
High	58.3	97.6	15.2	14	27

Each dataset contains 500 sows

In the results, sensitivity (proportion of correctly detected culled sows of all culled sows) ranged from 58 to 73%, specificity (proportion of detected retained sows of all retained sows) was always high (>97%) and error rate (proportion of false positively classified sows to all positively classified sows) varied between 6 and 15%. The dataset with the low herd performance level (L) showed the best classification parameters, which could be explained by the fact that most sows were culled to fertility criteria and high age.

These culling reasons were very explicit for the C4.5-algorithm. However, the selection of sows due to clinical problems was only correlated with parity and it was fixed for all three sow herd performance levels. This situation was not identifiable by the algorithm and sows were classified in relation to their productivity. The explanation for the worse classification of M is that most sows show similar performance parameters. A strong difference between the culling reasons was not very obvious. The size of the tree increased with the performance level. An example is given in Figure 5. Every node is shown in a circle, the branches are labelled with the split values of the attributes and the leaves are shown as rectangles. The tree began

with the attribute number of matings (me) followed by number of weaned piglets and number of litters.
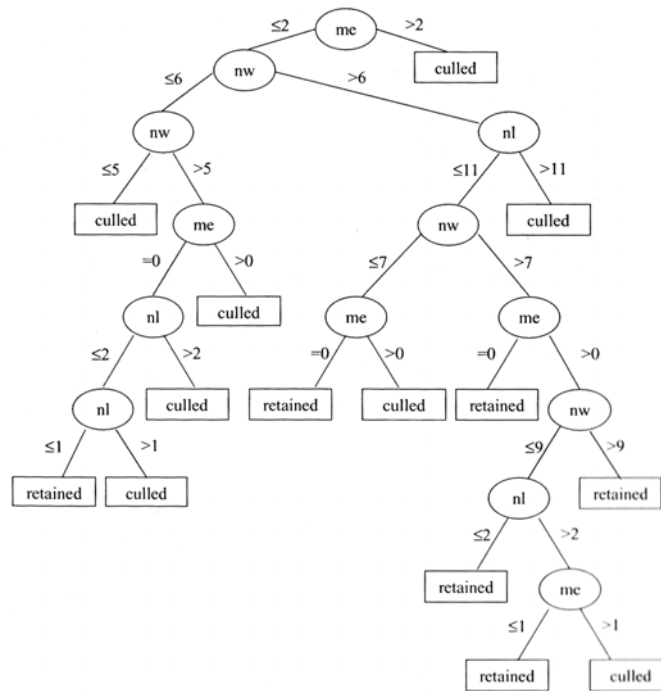

Fig. 5: Decision tree showing the detected threshold values of retaining/culling decisions in herd M (medium performance, minimum 20 instances per class, number of sows: 500; me: number of mating events; nw = number weaned piglets; nl: number of litters) (Entscheidungsbaum für die Ersatzentscheidung in Betrieb M; mittleres Leistungsniveau, mindestens 20 Beobachtungen je Subzelle, Anzahl Sauen: 500; me: Anzahl Umrauscher; nw = Anzahl der aufgezogenen Ferkel; nl: Wurfnummer)

These results and the results from other simulation scenarios (for details see KIRCHNER et al., 2004a, 2004c) demonstrate that the decision tree method is a suitable method for detecting relationships and patterns in simulated pig farming datasets.

Table 4

Means ($\bar{x}$) and standard deviations (s) of reproductive parameters for datasets A and B (Mittelwerte und Standardabweichungen der Fruchtbarkeitsmerkmale für die Datensätze A und B)

Parameter	.	Dataset A, n= 14,897		Dataset B, n= 21,818	
		$\bar{x}$	s	$\bar{x}$	s
Number of litters per sow	nl	3.7	2.4	3.6	2.2
Piglets born in total per litter	nbt	11.6	3.3	10.5	2.5
Piglets stillborn per litter	nsb	1.1	1.6	0.5	0.7
Piglets born alive per litter	nba	10.5	3.1	10.1	2.3
Piglets weaned per litter	nw	9.5	1.5	9.1	2.2
Weaning-to-conception interval	wco	10.2	-	13.9	-
Re-breeding rate		10.9	-	13.9	-

Example 4

In example 4, real datasets from two herds (A and B) were used to generate decision trees. The objective was to classify the binary farmer decision regarding replacing or not replacing a sow with a gilt. Table 4 summarises the reproductive parameters for datsets A and B.

The main reasons for culling were lameness and leg weakness (A: 20.4%; B: 30.8%), performance (A: 15.8%; B: 17.5%), age (A: 12.6%) and rebreeding (A:11.4%; B: 9.7%).

The application of the C4.5 algorithm on the two datasets showed different results referring to the quality of classification (Table 5). Dataset B reached better evaluation parameters than A. The sensitivity demonstrated a classification of the culled sows with 46.8% and an error rate of 15.3%. A' (n=7,057) and B' (n=12,149) are reduced datasets, they consisted only of instances which s described unambiguous culling reasons for the algorithm (fertility and reproductive problems). Observations with for the algorithm blurred culling reasons (e.g. aggressive or stressful sow's behaviour) were excluded. These culling reasons are not handled as an extra attribute and are not classified. As expected all classification parameters of the reduced datasets reached better results than the original datasets. This was affected by excluding all culling decisions which were blurry for the algorithm.

Table 5
Evaluation parameters for the sow herds A and B (Genauigkeit der Klassifizierung für die Datensätze A und B)

Dataset[1]	Sensitivity %	Specificity %	Error rate %	Number of leaves	Number of nodes
A	40.3	97.4	18.9	14	27
B	46.8	97.8	15.3	12	23
A'[2]	74.2	96.7	15.0	7	13
B'	87.5	96.1	14.9	9	17

[1] A = dataset A (n = 14,897); B = dataset B (n = 21.818); 100 instances per class
2) A' (n=7,057), B' (n=12,149), reduced datasets with only unambiguous culling reasons

The trees of dataset A and B differ clearly (Fig. 6). The tree of dataset B presents the attributes number of litters, piglets weaned and born instead of piglets born alive to tree A. The attributes piglets stillborn, weaning to conception interval and number of matings although available for the classification, do not appear in any of the generated decision trees. The generated trees reflect the different culling strategies adopted by the farmer. Regardless of the datasets, the ranking of the attributes in the generated trees is reasonable. It is conceivable the farmer reached a decision regarding sow replacement following the pattern exhibited by these trees.
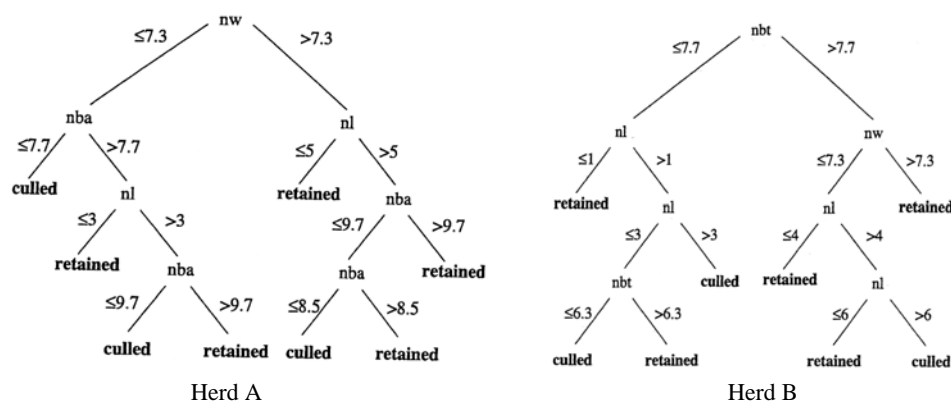

Fig. 6: Decision trees showing the detected threshold values of the farmer replacement strategy of dataset A (n=14,897) and B (n=21,818); nl: number of litters; nba: number of piglets born alive; nw: number of piglets weaned; nbt: number of piglets born in total) (Entscheidungsbäume für die Ersatzstrategien der Betriebe A und B; nl: Wurfnummer, nba: Anzahl der lebend geborenen Ferkel, nw: Anzahl der abgesetzten Ferkel, nbt: Anzahl der gesamt geborenen Ferkel)

The results of the decision tree technique applied to real pig farming datasets show comprehensible decision rules. The more information about the sows is available, e.g. information about the fundament after each farrowing, the better the quality of the classification.

Conclusion

Commercial swine farming is characterised by extended herd size, decreasing income margins and increasing demands on the farmer's management skills. Therefore, consistent information and decision support is becoming more and more important.

Computerised individual farm analysis involves three stages, (1) tracing deviations, (2) weighting deviations, and (3) further analysis of deviation.

EWMA control charts were used to trace shifts and deviations. These charts enable a graphical presentation of the results and modern computer technology has made it easy to implement control charts in on-line production control. The performance of the EWMA depends on the smoothing parameter λ and on the width of the control limits L. Simulation results showed that $\lambda = 0.2$ and $L = 1.5$ were appropriate choices for the given scenarios. But more research is needed (e.g. optimal design, behaviour of different types of charts) before the application of control charts can be generally recommended for tracing deviations.

Further analysis of the relevant (out-of-control) deviations should examine the underlying causes to improve the farm performance. The decision tree technique was checked to classify the farmer's sow replacement decision. Classifying other reproductive or economically important traits such as number of weaned piglets per sow per year is also conceivable. Having the graphical trees compared by an expert enables the farmer to detect weak elements of the production. An important issue for the application of the decision tree is data reliability. The more information is available, the better the quality of the classification and the decision rules.

References

DE VRIES, A.; CONLIN, B.J.:
    Design and performance of statistical process control charts applied to estrus detection efficiency. J. Dairy Sci. **86** (2003), 1970-1984
DEWAR, R.E.; MCQUEEN,R.J.:
    Analysing of cow culling data with a machine learning workbench. Working Paper Series 95/1. Department of Computer Science, University of Waikato, Hamilton, New Zealand (1995)
HUIRNE, R.B.M.:
    Computerized management support for swine breeding farms. Dissertation Thesis, University of Wageningen, The Netherlands (1990)
KAMBER, M.; WINSTONE, L.; GONG, W.; CHENG, S.; HAN, J.:
    Generalization and decision tree induction: efficient classification in data mining. In: Proceedings of the Seventh International Workshop on Issue of Data Engineering (RIDE'97), Birmingham, UK (1997), pp. 111-120
KIRCHNER, K.; TÖLLE, K.-H.; KRIETER, J.:
    The analysis of simulated sow herd datasets using decision tree technique. Comput. Electron. Agric. **42** (2004a), 111-127
KIRCHNER, K.; TÖLLE, K.-H.; KRIETER, J.:
    Decision tree technique applied to pig farming datasets. Livest. Prod. Sci., **90** (2004b), 191-200
KIRCHNER, K.; TÖLLE, K.-H.; KRIETER, J.:
    Optimisation of the decision tree technique applied to simulated sow herd datasets. Comput. Electron. Agric. (2005), submitted

KRIETER, J.:
Evaluation of different pig production systems including economic, welfare and environmental aspects. Arch. Tierz., Dummerstorf **45** (2002), 223-235

MITCHELL, R.S.; SHERLOCK, R.A.; SMITH, L.A.:
An investigation into the use of machine learning for determining oestrus in sows. Comput. Electron. Agric. **15** (1996), 195-213

MONTGOMERY, D.C.:
Introduction to statistical quality control., John Wiley & Sons, New York (1997)

PIETERSMA, D.; LACROIX, R.; LEFEBVRE, D.; WADE, K.M.:
Induction and evaluation of decision tree for lactation curve analysis. Comput. Electron. Agric. **38** (2003), 19-32

QUINLAN, J.R.:
Induction of decisions tree. Mach. Learn. **1** (1986), 81-106

QUINLAN, J.R.:
C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, USA (1993)

WIERINGA, J.E.:
Statistical process control for serially correlated data. Dissertation Thesis, University Groningen, The Netherlands (1999)

ZMP:
Vieh und Fleisch. Marktbilanz 2003. Zentrale Markt- und Preisberichtsstelle

Corresponding author
Prof. Dr. JOACHIM KRIETER
Institut f. Tierzucht u. Tierhaltung
Christian-Albrechts-Universität zu Kiel
Hermann-Rodewald-Str. 6
24118 KIEL

E-mail: jkrieter@tierzucht.uni-kiel.de