The Journal of Animal & Plant Sciences, 22(3): 2012, Page: 537-544
ISSN: 1018-7081

A NEW SELECTION PROCEDURE IN ANIMAL BREEDING: CLASSIFICATION, ERROR RATES AND COMPARISON WITH BLUP

A. Pala

Canakkale Onsekiz Mart University, Department of Animal Science, office:222, Canakkale/Turkey
Corresponding Author e-mail: akin@comu.edu.tr, dr.akin.pala@gmail.com

ABSTRACT

Classification techniques can be used as an alternative to the selection index, especially for indirect selection of multiple traits. Prior information for this technique provides to increase accuracy of selection. Two sets of simulations were used to estimate the error rates of the classification techniques in selection of livestock and to compare the method with the animal model, an example of Best Linear Unbiased Predictions (BLUP). The first set of analyses considered the traits milk yield, milk fat, milk protein and somatic cell count (SCC). Frequent milking was simulated and the high and low means were created within the standard deviation limits. Using these different means, observations and standard deviations, various scenarios were produced to simulate various data sets for validation to check how correctly the classification method worked. The different scenarios included **A.** Same means and σ, different observations, **B.** Different means and observations, same σ, **C.** Different means (standardized) and observations, same σ, **D.** Same means, different σ (±5 for milk yield, milk fat and protein, ±3 for SCC) and observations, **E.** Same means, different σ (±10 for milk yield, milk fat and protein, ±6 for SCC) and observations. The procedure selected the correct animals and produced a very low error rate in almost all simulations. In the second set of simulations, the traits considered were milk yield and SCC for two goat breeds, Saanen and German Fawn. A total of 10 000 dams, 3174 sires and 10 000 progeny were simulated. The family variance was 4.5. The sires had 3.15 progeny on average, and sires had minimum 1 and maximum 15 progeny. Estimated breeding values from the MTDFREML (Multi Trait Derivative Free Estimated Maximum Likelihood) program were used in the culling decisions. Animals selected using the animal model of BLUP with MTDFREML and animals selected using the classification method with SAS were analyzed employing the rank correlation, and the calculated value was very high 0.99 ($P<0.01$). If used with the necessary standardizations, the method seems promising enough to be considered as an alternative to existing selection methods and a new method for indirect, multiple trait selection in animal breeding. In addition, the method can be used in cases (no extensive pedigree, small number of animals/sires etc.) where it is difficult to calculate correct genetic parameters.

Key words: selection index, classification, BLUP, prior and posterior probabilities, animal model.

INTRODUCTION

In animal breeding, there are three methods for multitrait selection and the index method is the most efficient (Hazel and Lush, 1942). However, one disadvantage of the index method is that calculations are somewhat complex. Hazel *et al.* (1994) stated that except for large corporate breeding organizations, it generally is not feasible for individual breeders to develop their own selection indexes since these involve accurate estimates of heritability, genetic variation and genetic and phenotypic correlations. Mrode (1996) noted that multiple trait BLUP analyses require reliable estimates of genetic and phenotypic correlations and these may not be readily available or may be hard to calculate due to genetic correlations requiring a large number of animals.

Classification technique can perform indirect selection for multiple traits and does not require a relationship matrix, accurate estimates of heritability or genetic correlations. Though lack of these parameters may bias the results, the accuracy is improved by other means such as prior information. Johnson and Wichern (1998) defined classification as the method that sorts objects (observations) into two or more labeled classes. This method is different from cluster analysis because it uses prior information to arrive at new decisions. Some "rules" are defined using prior information and new animals are assigned to classes such as selected or not. In a statistical analysis or a selection procedure, it is beneficial to use all the information available. If significant prior information is available, then it should be added to the information at hand to increase accuracy (Gianola and Fernando, 1986).

The index selection is based on the discrimination technique, which was first introduced by R. A. Fisher (1936). Johnson and Wichern (1998) noted that the goal of discrimination is to explore observed differences. However, the goal of the classification method is to sort observations into labeled classes and develop a rule for assigning new observations to those classes.

Selection based on an index considers records of all traits, tandem selection only one trait at a time no matter how good other traits are, and selection by independent culling level each trait independently so that a poor record on any trait will lead to culling (Van Vleck *et al.*, 1987). The classification procedure takes prior information into consideration as well as records of all traits. The method develops a classification criterion to classify each observation into one of the groups on the basis of one or more quantitative variables. The derived criterion from this data set can be applied to a second data set, which is called the training or calibration data set. Classification can be done by either a parametric or a nonparametric method (SAS, 1999).

The selection index method calculates an index value for each animal. Similarly, this theory can be used to calculate a posterior probability for each animal. The animals are then ranked based on the posterior probabilities and the selection is based on the probabilities instead of the index values. Eisen (1998) noted that results from selection theory progress more rapidly compared to those from artificial selection experiments. Computer simulation studies are used to test selection theory to speed up the process and to compensate for this difference (Eisen, 1998). Major purpose of this paper was to investigate the use of classification method in animal breeding for selection of domestic animals using two sets of simulations.

MATERIALS and METHODS

Using the classification method to separate the animals for selection; the multivariate problem can be converted into a univariate problem, thus drawing one dependent variable line in space where the intersection of dependent variables is maximum. Therefore, to be able to obtain the best view of the three or more dimensional problem in two dimensions, the maximum "L", where $Y_{new} = L^t * Y$ is calculated as follows:

$$\text{Max L} = \frac{\left[L^t\left(\mu^{(1)} - \mu^{(2)}\right)\right]^2}{L^t \sum L}$$

where $L = \sum^{-1}\left(\mu^{(1)} - \mu^{(2)}\right)$

The distributions are:

Selected: N ~ ($L^t \mu^{(1)}$, $L^t \sum L$)

Not Selected: N ~ ($L^t \mu^{(2)}$, $L^t \sum L$)

where N defines the distribution (normal), $\sum$ is the covariance matrix, Y is individual's performance vector with all the traits, superscript "t" indicates transpose of the matrix or vector. $\mu^{(1)}$ is mean vector of traits for population 1 in the preparation data and $\mu^{(2)}$ is that for population 2.

Finding the maximum L makes it possible to separate the groups (selected, and not selected) as far apart as possible, thus making it conceivable to see all dependent variables in two dimensions. In two dimensions, it is convenient to find a midpoint and classify the individuals as "selected" if $L^t * Y >$ midpoint and "not selected" if $L^t * Y <$ midpoint. The midpoint can be found using this equation:

$$\text{Midpoint} = ½ \mu^{(1)t} \sum^{-1} \mu^{(1)} - ½ \mu^{(2)t} \sum^{-1} \mu^{(2)}$$

One way to perform classification calculations is to find the probability density functions f_1 (x) and f_2 (x) related to populations 1 and 2. According to Johnson and Wichern (1998), if region R_1 is the set of values for which one classifies objects as population 1, then:

R_1 : density ratio ≥ prior probability ratio

If R_2 is the set of values for which one classifies objects as population 2, then:

R_2 : density ratio < prior probability ratio

Here, density ratio is: f_1 (x)/ f_2 (x) and prior probability ratio is p_2/p_1, where $p_1 + p_2 = 1$. From here, the posterior probability of allocating a new observation (x_0) to, for example, selected animals can be calculated using Bayes' rule. The new observations with the larger posterior probabilities of belonging to selected animals population can be classified into the group of selected animals. The posterior probability will be:

$$\text{P (population } 1|x_0) = \frac{P_1 f_1(x_0)}{P_1 f_1\ (x_0) + P_2 f_2(x_0)}$$

(Johnson and Wichern, 1998)

Milk yield, Milk Fat, Milk Protein and Somatic Cell Count in Dairy Cows: All of the data sets were simulated using SAS V8 (1999). The traits considered were milk yield, milk fat, milk protein and somatic cell count (SCC). The lactation records were assumed to be calculated using the best prediction method and standardized for environmental effects such as age, calving month and previous days open. It was reported by many authors that genetic evaluation of dairy systems is usually based on 305 day lactation milk yields and 305 day milk yield is an important measurement for the dairy industry (Ulutas *et al.*, 2010; Yilmaz *et al*, 2011).

In the present study, records in progress and other records shorter than 305 days were simulated as records extended to 305 days. Component percentages were assumed to be computed from average standardized yields for milk and components (USDA, 2004).

In dairy cattle, higher milk production increases the probability that the animal will stay in the herd longer. If the increase in the second lactation is high, it is an indication that the cow will stay longer in the herd. For the calibration data, an average of 17±3 % increase in milk yield, 17±3 % increase in milk fat, 18 ±3 % increase in protein and 15±1.5 % increase in SCC in the second lactation was simulated for the selected animals. For those culled, increases in the second lactation simulated were an average of 12±3 % increase for milk yield, 12±3 % increase for milk fat, 13 ±3 % increase for protein and 20±1.5 % increase for SCC.

Two different milking frequencies were simulated; 2 or 4 times/day (2x or 4x, respectively). The distribution was normal within treatments and the data were balanced. In all simulations, the high and low means were created within the standard deviation limits. For example, if the increases were reported to be 17±3 % in the milk yield, then the high mean increase in the second lactation was 20 % and the low mean increase was 14 %. Similarly, 20±1.5 increase in SCC meant that the increase in the second lactation was 21.5 (high mean) and 18.5 (low mean) and the standard deviation was 1.5 overall.

Because 20 % of the animals was assumed to be culled, 8000 animal records were simulated and were marked as cows in milk, and 2000 animal records were simulated and were marked culled, with the purpose of using the data set for calibration.

Seeds are used in simulations so that a precise replicate can be reproduced later using the same seed. The same seeds were used for all the simulations related to milk yield, milk fat, milk protein and SCC in dairy cows. The seeds in the training/calibration data were 55555 for milk yield, 77777 for milk fat, 11111 for milk protein and 89898 for SCC. The seeds in the new data set (test data), were 123 for milk yield, 321 for milk fat, 432 for milk protein and 345 for SCC (Table 1). To accomplish the simulations, do statements were used in SAS along with the rannor statement. For example, z=mu+rannor(55555); was used to produce the training data for milk yield where mu was provided according to treatment averages and standard deviations needed using do statements. Range of the mu was extended as standard deviations needed increased. The data produced by the simulations were analyzed using the PROC DISCRIM procedure in SAS and the model used milk yield, milk fat, milk protein and SCC as dependent factors and the status as the class variable.

The performance of the classification technique was evaluated by probabilities of misclassification, that is, by estimating error rates. When only the calibration data set exists (no test data), the error rates can be estimated by cross validation (SAS, 1999). Crossvalidation classification of the data gives the probability of misclassification. Crossvalidation estimates are less biased since they do not use the observation, which was used to construct the data (SAS, 1999).

Comparison of the Classification Method and the Animal Model: Data were simulated using SAS V8 (SAS, 1999) for comparison of the technique with the animal model. The traits considered were milk yield and SCC for two goat breeds, Saanen and German Fawn. A total of 10 000 dams, 3174 sires and 10 000 progeny were simulated. The family variance was 4.5. The sires had 3.15 progeny on average, and sires had minimum 1 and maximum 15 progeny. The lactation records were assumed to be calculated using the best prediction method and standardized for environmental effects such as age, calving month and previous days open. The breed effects were not corrected in the simulation; it was left to the classification and BLUP methods to do the correction.

The SCC values were created with normal distribution and then were transformed to Somatic Cell Score for the analyses. Normally this transformation is performed so that the positively skewed field data will have normal distribution. In this simulation, the data were transformed so that the otherwise too large variances and covariance would fit to the MTDFREML program's format.

Data were analyzed with a derivative-free algorithm (Smith and Graser, 1986; Graser *et al.*, 1987) using MTDFREML (Boldman *et al*., 1995). Convergence criterion was 1E-6 and number of maximum evaluations was 2000. To ensure global convergence, the algorithm by Boldman *et al*. (1995) was restarted with estimates until the log likelihood did not change at the second decimal (Zulkadir et al., 2009).

The statistical model for BLUP in matrix notation was,

$Y = Xb + Za + e$,

where;

Y = observations matrix for milk yield and Somatic Cell Count,

b = vector of the fixed effect goat breeds,

a = vector of random animal effects,

e = vector of random residual effects,

X and Z= incidence matrices relating records to the fixed effect and random animal effects, respectively (Mrode, 1996).

The test data set were analyzed using the two methods. Classification method used the calibration data set to set up and apply the rules to the test data set, while BLUP method, applied by MTDFREML programs used genetic variances, genetic covariance, residual variances and residual covariance of the two traits as priors. MTDFREML program corrected for the fixed effect "breed" while the classification method used the rule from the training data set to distinguish the breeds. Classification analyses were carried out using the PROC DISCRIM procedure of SAS and the model used included milk yield, milk fat, milk protein and SCC as dependent factors and the status as the class variable.

Estimated breeding values (EBV) from the MTDFREML program were used in the culling decision. The 20 % of animals that had the lowest milk yield and highest SCC were marked "culled".

Selected and culled animals by using these two methods were given ones and zeros, respectively, to analyze for Spearman rank correlation. The correlation analysis was carried out using SAS (SAS, 1999).

RESULTS

To have high and low means in the data set, frequency of milking was simulated. A simulation with different seeds, but the same means and standard deviations (observations are different, means are the same) was run to produce a new data set to be used for validation to check how correctly the new set of animals was classified into the cows in milk or the culled group (Table 1).

In the new data set, those cows with large increases of milk yield, milk fat and protein and low increases of SCC in the second lactation were depicted the status selected and those with low increases in all traits except SCC were given the status culled. That is, the animals with the lowest 20 % performance were culled. Performance in SCC is measured by a decrease in SCC while the numbers in the other traits such as milk yield are expected to increase. Then the classification analysis was run and the error rates were calculated. There was only one observation misclassified in both groups; thus, the error rates were 0.0005 for culled cows and 0.0001 for cows in milk.

A more difficult scenario, using simulation was presented by increasing the means in the test herd. An average of 23±3 % increase for milk yield, 23±3 % increase for milk fat, 24 ±3 % increase in protein and 9±1.5 % increase in SCC was simulated in the second lactation for the selected animals. For those culled, an average of 18±3 % increase for milk yield, 18±3 % increase for milk fat, 19 ±3 % increase in protein and 14±1.5 % increase in SCC was simulated in the second lactation. All of these means except SCC were 6 points higher than those in the calibration data; SCC had a 6-point drop. Milk yield, milk fat and protein increases of culled cows in the test data were higher than those selected in the calibration data (17, 17 and 18 % vs. 18, 18 and 19 %) and SCC increase of culled cows in the test data were lower than SCC increase of selected cows in the calibration data (14 % vs. 15 %). The error rate increased to maximum for the culled, marking all animals selected (no standardization). There were no misclassifications for the selected animals. Then, the higher means were standardized back to the original test data means by subtracting 6 points of increase from all observations of milk yield, milk fat and protein, and by adding 6 points of increase to all observations of SCC. The error rates decreased to levels near zero; only one observation was misclassified in each of the groups; namely the culled and selected animals.

Variances of calibration data and test data may differ though the means may stay the same. If the means differ as well, they should be standardized as previously explained because more improvement may be desired in the test herd/generation. In this approach, the means were the same both in the calibration data and the test data while the standard deviations for the advances in the second lactation were increased to ±5 for milk yield, milk fat and protein, and were increased to ±3 for SCC both for selected and culled cows. The error rates were quite low; 7 culled animals out of 2000 were misclassified as selected (error rate=0.0035) and 11 selected animals out of 8000 were misclassified as culled (error rate=0.0014).

Then the standard deviations were increased to ±10 for milk yield, milk fat and protein, and were increased to ±6 for SCC. Increasing the standard deviations this much increased the error rate to 0.04 for the culled animals (81 out of the 2000 misclassified into selected) and to 0.02 for the selected animals (147 out of the 8000 misclassified into culled). Though the standard deviations were increased up to ±10, the misclassification rate did not go over 4 % in this simulation (Table 1).

Table 2 provides a summary of the classification for the last scenario shown in Table 1, by providing error count estimates in more detail. Namely, the number of observations and the percent classified into a certain status are given in Table 2 to show more detail. In the table, the number of observations shows the number of observations classified into one of the status and the percent classified provides the percentage of these numbers. The results show that around 96 per cent of the observations were labeled correctly to the status 0, culled, out of the 2000 observations that were labeled as the status 0. About 98 per cent of the 8000 observations that were labeled as status 1 (selected) were labeled correctly and 147 of the observations out of the 8000 were mislabeled as status 0.

Table 1. Error Count Estimates for the selected or culled animals in various simulation models.

Simulation	**Same means and σ, different observations**		**Different means and observations, same σ**		**Different means (standardized) and observations, same σ**		**Same means, different σ (±5, ±3) and observations**		**Same means, different σ (±10, ±6) and observations**	
Status	Culled	Selected	Culled	Selected	Culled	Selected	Culled	Selected	Culled	Selected
Error estimate, %	0.001	0.001	1.000	0.000	0.001	0.001	0.003	0.001	0.041	0.018

Table 2. Classification Summary for the “same means, different σ (±10, ±6) and observations” scenario.

From status		0	1	Total
0	Number of Observations	1919	81	2000
	Percent Classified	95.95	4.05	100
1	Number of Observations	147	7853	8000
	Percent Classified	1.84	98.16	100
Total	Number of Observations	2066	7934	10000
	Percent Classified	20.66	79.34	100
	Priors	0.2	0.8	

Table 3. Classification Results using Linear Discriminant Function, showing Posterior Probability of Membership in status

Obs	From status	Classified into status	0	1
1	0	0	1.0000	0.0000
2	0	1	0.0369	0.9631
3	0	0	1.0000	0.0000
4	0	0	0.7973	0.2027
5	0	0	1.0000	0.0000

The first five observations of the simulation are shown in Table 3, which indicates that the first five observations of the 10 000 are mostly in status zero. The last two columns indicate the posterior probabilities that the observations should be classified in one of the classes. The second observation is one of the rare ones that were mislabeled as selected, while it should have been labeled as culled, as indicated by the star, realized as the result of the crossvalidation procedure. The other four observations were correctly labeled in this example of the first five observations.

Although evaluation of the classification method in selection using the error rate method is valuable, a direct comparison of the BLUP method and the classification method may be desirable. The observations were simulated so that Saanen goats would have higher milk yield and higher SCC (Table 4). The probability that an animal will have mastitis tends to increase with the increased milk yield. The status in Table 4 indicates whether an animal was selected or not, according to their milk yield.

The standard deviations of the means were set so that the breeds would be within one standard deviation. For example, overall mean of the milk yield for selected animals was set to be 700±100 and thus, Saanen breed had 800 kg and German Fawn had 600 kg milk in the lactation. For those not selected, the overall mean was 500±100. Average milk yield of culled Saanen and selected German Fawn goats were the same. The prior heritabilities used in MTDFREML were h^2=0.15 for milk yield and h^2=0.11 for SCC, and the genetic and phenotypic correlations were 0.36 and 0.14, respectively. Both methods provided very similar outcomes. The rank correlation analysis indicated that the correlation coefficient was 0.99875 (P<0.01). This is a very high coefficient, a strong indication that both methods provided similar results.

Table 4. Simulated means of milk and SCC for selected and non-selected categories in Saanen and German Fawn goat breeds

Status	Breed	Milk	SCC
Selected	German Fawn	600	350000
	Saanen	800	400000
Non-Selected	German Fawn	400	650000
	Saanen	600	700000

DISCUSSION

The first simulation with the same means was an ideal situation where everything was controlled and the rules derived applied to the new data exactly because they conformed to the same rules as the previous data set. However, in reality, differences may arise because the expected result may not exactly be bound by the rules of the previous data, called the training or calibration data. For example, a cow with high increases in milk yield, protein ratio and fat ratio and a low increase in SCC could have a low life-long milk yield simply due to a breed effect. A Holstein breed of cow that has a lower increase of milk yield in the second lactation than a Jersey cow may have a higher lifetime milk yield, though the rules of the training data would suggest otherwise. Environmental effects such as days in milk, frequency of milking and age must be accounted for before classification is performed. This can be done using SAS. Further, the same means may not apply to a different herd because the herd’s genetic level may be more advanced (different breeds or the more advanced-next generation) or the herd may be on a higher nutritional level. The new, calibration data may consist of cows with higher lifetime production expectancy, even culled cows in the test data may have higher increases than those selected in the calibration data. Therefore, a more difficult scenario, using simulation was presented by increasing the means in the test herd and the results suggested that standardizations may be required for the classification effects if there are multiple factors in the data set for the classification techniques to succeed completely.

In some herds, even adding a simple recording system can improve traits such as lactation milk yield (Reiad *et al.,* 2010). However, methods such as selection index or animal model, taking estimated genetic parameters into account, or the classification method considering prior information are expected to make a larger contribution to genetic improvement. To establish a challenging data set for these elaborate systems to solve, frequency of milking was simulated to include high and low means in the data set. Dahl *et al.* (2004) reported that cows milked more frequently for 21 days at the beginning of lactation had higher milk yield and lower somatic cell count throughout lactation, though differences were not significant after test day 6. Similarly, Koyuncu and Pala (2008) reported that increasing frequency of milking for a short period in Saanen goats increased milk yield with no negative effects on udder health, measured by Somatic Cell Count. In addition, Pala and Koyuncu (2007) reported that increasing milk frequency increased persistency, which was calculated using methods reported by Pala and Savas (2005). Unalan and Cankaya (2010) reported that as the parity increased, 305-day milk yield increased in cows, although they also reported that the first lactation is a useful indicator for the succeeding lactations. In this study, an increase in later lactations was simulated to produce realistic simulations. The first simulation included the same means and standard deviations, but different observations and had only one observation misclassified in both groups. In this scenario, the classification method had very high success. A tougher scenario included different means, a test herd with higher performances than the calibration herd. The results indicated that without standardization, the culled animals may not be selected correctly. This outcome may be acceptable if the level of improvement, for example lifetime production, is the same for both the calibration herd and the test herd. However, if the lifetime production is expected to be greater in the test herd, than the means in the second herd needs to be standardized. If selection in the calibration generation was successful and the test herd is ahead of the original population in terms of some traits, then the means should be standardized for each trait separately, so that more improvement can be obtained. For example, milk yield may increase in the test herd while milk fat decrease or protein ratio may increase while SCC may also increase. In this case, the traits should be standardized separately and then the classification analyses should be run. In a second simulation, standardization was applied to all traits and the error levels decreased significantly, which is a strong indication that the method can work very well, if the necessary standardizations are applied.

Another approach were tried by keeping the means the same both in the calibration data and the test data, but the standard deviations for the advances in the second lactation were increased. The error rates were very low, well within acceptable limits as they were below 0.01. Then standard deviations were increased further, twice as before (±10), and the error rates increased some. However, the rates were still very low; they were 0.02 for the selected and 0.04 for the culled. This indicates that if means are corrected accordingly, increased variances contribute little to the misclassification rate.

All results of the simulations were summarized in Table 1. The error rate increased further as the standard deviations deviated from the calibration data. This increase was more pronounced in culled animals, because there were fewer animals in that group, compared to the cows in milk. The great increase in the error rate in the non-standardized data set and the drop of the rates in the standardized data set indicate the great need for standardization. Overall, the error rates were small and the accuracy of the method was high for classifying the animals into the culled or the selected category.

A direct comparison of the BLUP method and the classification method was carried out using prior heritabilities. Cilek and Bakir (2010) reported that because heritabilities of milk yield related traits are low, other selection methods, such as family selection needs to be used, as opposed to using simple mass selection. Using animal model or the classification method can be a solution to the low heritability problem. The rank correlation indicated that the results of the animal method and the classification methods were very similar. It can be speculated that a real data set with unaccountable factors (larger error term) may provide more realistic results. Though a real selection experiment may provide results that are more realistic, this simulation is certainly a good incentive for using the classification method and continuing the investigation of the method to be used as a selection tool.

Breeders will frequently meet the situation of antagonistic selection (Ieiri *et al.*, 2004). If it is wished to increase one trait and decrease others, a minus sign can be placed in front of all the numbers of the dependent variables. Thus, the calculations result in assigning higher posterior probabilities for low numbers in absolute value (high value is considered negative) and still conserve the percentage to be selected.

In dairy cattle, longevity is usually predicted by reproduction rate and milk yield. Higher milk production increases the probability that the animal will stay in the herd longer. Usually the first lactation is held constant and increase in the second lactation is measured. If that is high, it is an indication that the cow will have a long productive life. Increased age at first calving can be related with lower lifetime production due to decreased time, although the life span may increase slightly. According to findings of Westendorp and Kirkwood (1998) in human females, there is a tradeoff between longevity and reproduction. The authors wrote that they found a positive correlation between age at first child birth and female longevity. One can determine, for example, in Holstein cows, what age at first calving gives the optimum lifetime milk production, milk fat and milk protein. Since results from the literature data (how long they lived and their lifetime performance) is known, the method should be able to give the posterior probabilities on what population (selected or culled) the new animal will belong to. Then, animals at hand can be evaluated, and selection can be performed for their lifetime performance, without waiting for their productive life to end.

The classification method is used in various fields, for example, in colleges to select students. The administration looks at high school graduates' GPA (Grade Point Average) and SAT (Scholastic Assessment Test) scores and tries to estimate if the students will be successful in the college, based on these previous data. Because they have the GPA and SAT scores of the previous students, and they already know if these students were successful or not, they can derive rules and apply these to the new candidates. Similarly, one can try to estimate battery life looking at various factors that affects battery life, such as carbon percentage. This is similar to trying to estimate a cow's lifetime production by looking at various factors. In an earlier data set, there are some factors and the cows' lifetime productions. For the contemporary cows, a rule is derived from that earlier herd and the cows' lifetime production is estimated using the same factors.

Dairy cows' longevity in a herd can be calculated using the classification method. Independent variables affecting longevity such as milk production, reproduction traits (conception rate, pelvic size) and frequency of diseases can be used to define the rules. Then, selection can be applied using the posterior probabilities produced.

Economic weight procedures are not included in this paper, but the theory exists and it can be incorporated. The "cost" of selecting or culling an animal has to be included in the equation. Different traits may have different "costs" so that the data must overwhelmingly support one trait to be selected while animals that excel in another trait can be selected with little data support. This is similar to, for example the bankruptcy of a firm. The data must overwhelmingly support the argument that the company is in economic straits so that it can be overtaken and sold. Similarly, the cost of classifying a heart patient as sick is smaller than the cost of classifying a heart patient as healthy. Here, the evidence must strongly support that the patient is well. If trait A gets more weight while trait B is of little value, then the cost of culling based on trait A can be set to a high number while cost of culling based on B can be set to a low number. Animals with poor performance of trait B can be culled easily while the data must overwhelmingly show poor performance of A to cull that animal.

Waldron and Thomas (1992) predicted response to direct and indirect selection in litter size in sheep by using variance and covariance estimates. They estimated that a selection index using litter size and ovulation rate produced 123% as much response in litter size in contrast to selection on litter size alone. If animals are selected on an index of litter size, ovulation rate, and scrotal circumference, the same authors estimated that it would result in 133% as much response in litter size as opposed to directly selecting for litter size. Classification methods can be used to add this kind of prior information to the early measurements, thus improving the accuracy of selection.

The method can especially be useful in traits where receiving accurate information requires killing the animal. Meat quality traits are in this category. Normally, it is necessary to slaughter the animal to be able to evaluate it for such traits. It is appropriate to look at the relatives such as parents or sibs in such cases. It can also be done using previous data from another farm. Cattle are usually selected without killing them for meat quality, using ultrasound. Subcutaneous fat and loin eye measurements are used for selection. Although these may give good estimates, using prior information will increase the accuracy of selection. Because the animals providing the prior information are slaughtered and the results are known in terms of the meat quality (meat quality can be easily measured after killing the animal using methods such as KPH and tasting experiments), they provide valuable information. If marbling is measured by ultrasound, for example, the prior data (pilot study, literature, relatives etc.) will provide information on how ultrasound measures deviate from the accuracy in the new group.

Conclusion: Classification methods may offer more variety to breeders on the selection methodology and may improve results of selection because of the use of prior information. The method can be useful in cases where it is difficult to calculate genetic parameters with accuracy (low number of animals or non-existing/inaccurate relationship matrix or low number of relationship pedigree records).

REFERENCES

Boldman, K. G., L. A. Kriese, L. D. Van Vleck, C. P. Van Tassell, S. D. Kachman (1995). A manual for the use of MTDFREML. ARS, USDA, Clay Center, NE.

Cilek, S. and G. Bakir (2010). Milk Yield Traits of Brown Cows Reared at Malya State Farm and Effects of Some Environmental Factors on These Traits. Kafkas Univ Vet Fak Derg. 16 (2): 347-350, 2010.

Dahl, G. E., R. L. Wallace, R. D. Shanks, D. Lueking Hot Topic (2004). Effects of frequent milking in early lactation on milk yield and udder health. J Dairy Sci, 87: 882–885.

Eisen, E. J. (1998). Selection Theory and Experiments. 6th World Congress on Genetics Applied to Livestock Production. 26: 67.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems, Ann Eug, 7: 179-188.

Gianola, D, R. L. Fernando (1986). Bayesian methods in animal breeding theory. J Anim Sci, 63: 217-244.

Graser, H. U., S. P. Smith and B. Tier (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci. 64: 1362-1370.

Hazel, L. N., G. E. Dickerson and A. E. Freeman (1994). The selection index—Then, now, and for the future. J Dairy Sci, 77: 3236-3251.

Hazel, L. N., and J. L. Lush (1942). The efficiency of the three methods of selection. J Heredity, 33: 393-399.

Ieiri, S., T. Nomura, H. Hirooka and M. Satoh (2004). A comparison of restricted selection procedures to control genetic gains. J Anim Breed Genet, 121: 90–100.

Johnson, R. A., D. W. Wichern (1998). Applied multivariate statistical analysis. Prentice Hall, Upper Saddle River, New Jersey.

Koyuncu, E. and A. Pala (2008). Effects of Short Period Frequent Milking on Milk Yield and Udder Health in Turkish Saanen Goats. Anim Sci J, 79 (1): 111-115.

Mrode, R. A. (1996). Linear Models for the Prediction of Animal Breeding Values, p78. Wallingford, CAB International, UK.

Pala, A., and E. Koyuncu (2007). Effects of short period frequent milking on persistency of milk yield and SCS in Turkish Saanen Goats. Anim Sci J, 78 (4): 400-406.

Pala, A, Savas T (2005). Persistency within and between lactations in morning, evening and daily test day milk in dairy goats. Archiv fur Tierzucht-Arch Anim Breed, 48(4): 396-403.

Reiad, K., W. Al-Azzawi, K. Al-Najjar, Y. Masri, S. Salhab, Z. Abdo, I. El-Herek, H. Omed, M. Saatci (2010). Factors influencing the milk production of Awassi sheep in a flock with the selected lines at the Agricultural Scientific Research Centre. Kafkas Univ Vet Fak Derg, 16 (3): 425-430.

SAS Institute Inc. (1999). SAS OnlineDoc®, Version 8, Cary, NC, USA.

Smith, S. P., H. U. Graser (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. J Dairy Sci, 69: 1156-1165.

Ulutas, Z., M. Sezer, M. Saatci, A. Şahin (2010). Estimation of genetic and phenotypic trends of 305-day milk yield for Simmentals reared in Kazova State Farm in Turkey. Kafkas Univ Vet Fak Derg, 16 (4): 533-536.

Unalan, A. and S. Cankaya (2010). Genetic parameters and correlations for lactation milk yields according to lactation numbers in Jersey cows. Kafkas Univ Vet Fak Derg, 16 (6): 995-1000.

USDA(2004). United States Department of Agriculture. Available from URL: http://www.aipl.arsusda.gov/publish/dhi/dhi03/laall.shtml.

Van Vleck, L. D., J. E. Pollak and E. A. B. Oltenacu (1987). Genetics for the Animal Sciences, p. 339. W. H. Freeman company, New York.

Waldron, D. F. and D. L. Thomas (1992). Increased litter size in Rambouillet sheep: II. Expected responses from alternative selection criteria. J Anim Sci, 70: 3345-3350.

Westendorp, R. G. J. and T. B. L. Kirkwood (1998). Human longevity at the cost of reproductive success, Nature 396: 743-746.

Yilmaz, I, E. Eyduran, A. Kaygisiz and K. Javed (2011). Estimates of genetic parameters for lactation shape parameters with multivariate statistical technique in brown swiss cattle. Int J Agric Biol 13: 174–178.

Zulkadir, U., I. Aytekin and A. Pala (2009). Genetic Analyses for Milk Yield, Lactation Period and Fat Percentage in Brown Swiss Cattle. J Anim Vet Adv, 8 (5): 857-862.