Approximate accuracy of dairy cattle evaluation from multiple lactation random regression model for Polish Black-and-White cattle[*]

T. Strabel[1] and J. Jamrozik[2,3]

*[1]August Cieszkowski Agricultural University, Department of Genetics and Animal Breeding
Wołyńska 33, 60-637 Poznań, Poland*
*[2]Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science,
University of Guelph
Guelph, ON, N1G 2W1, Canada*
*[3]National Research Institute of Animal Production, Department of Animal Genetics and Breeding
32-083 Balice, Poland*

ABSTRACT

A method is described to obtain approximate reliabilities for breeding values calculated with a multiple-lactation random regression test day model. The method was based on a concept of an equivalent number of progeny for animals with records, which is further used to derive reliabilities for related animals. The procedure accounted for the size of contemporary groups. Approximate reliabilities were calculated for estimated breeding values of the average milk yield in the first three lactations, using the data of daily milk yield of Polish Black-and-White cows in randomly selected herds. Results of approximation were analysed by comparison with corresponding exact reliabilities, obtained from the inverse of the coefficients matrix of the mixed model equations. High correlation of 0.98 for bulls between reliabilities from both methods and very low computer requirements facilitate the implementation of the approximate method into a routine genetic evaluation procedure.

KEY WORDS: accuracy of evaluation, dairy cattle, genetic value, random regressions, test day model

INTRODUCTION

Accuracy of genetic evaluations is usually provided along with breeding values and can influence the selection decision. The measure of accuracy, in terms of reliability of evaluation, can be calculated as the function of prediction error variance. This requires an inversion of the coefficient matrix of mixed model equations (MME) when the evaluation is based on the Best Linear Unbiased Prediction (BLUP). Genetic evaluations of dairy cattle using test day records and random regression models implies setting up a very large system of MME. The coefficient matrix of such a system cannot be inverted directly due to its size. Hence, methods for approximation of accuracy have to be used.

A number of procedures for the approximation of diagonal elements of the inverse of MME matrix have been proposed for a single trait (Misztal and Wiggans, 1988; Meyer, 1989) and also for multiple trait models (Tier et al., 1991; Strabel et al., 2001). The source of information for these methods is data on animals with records, which is further used to derive approximations for the remaining animals. An idea of an equivalent number of progeny (ENP) is based on converting the number of records on an animal into a corresponding ENP that would give the same accuracy if progeny were the only information available, and to accumulate this quantity over all progeny of an animal (Koots et al., 1997). This method was developed by Jamrozik et al. (2000) for random regression models. Although it was shown that this approximation provides relatively accurate reliabilities, it does not account for the size of contemporary groups and hence it may be less effective for the Polish dairy population where herds of a small size are common.

The objective of this study was to adopt the procedure for approximating the reliability of genetic evaluations from RRM proposed by Jamrozik et al. (2000) to the multiple lactation random regression test day model for Polish Black-and-White cattle in a way that it would account for the number of a cow's contemporaries on a given test. Results of the approximation were compared with accuracies obtained by the exact method.

MATERIAL AND METHODS

The model

The random regression model proposed for the Polish Black-and-White cattle (Strabel et al., 2005) was extended to the three lactation form as:

$$y = Xb + Uq + Wp + Za + e$$

where: y is a vector of observations, b is a vector of fixed regression coefficients for age - season of calving classes and herd-years, q is a vector of random

herd-test-date effects, **p** is a vector of permanent environmental regression coefficients, **a** is a vector of additive genetic regression coefficients, **e** is a vector of residuals, and **X**, **U**, **W**, and **Z** are incidence matrices relating observations to effects. Both sets of random regressions, **a** and **p**, were modeled with third order Legendre polynomials. Fixed regressions were modeled with Legendre polynomials of order five and three for age-season of calving and herd-years, respectively.

The covariance structure of the model is:

$$Var(\mathbf{y}) = \mathbf{UHU'} + \mathbf{WPW'} + \mathbf{ZGZ'} + \mathbf{R}$$

where: $\mathbf{H}=\mathbf{I}\otimes\mathbf{H_0}$, $\mathbf{P}=\mathbf{I}\otimes\mathbf{P_0}$, $\mathbf{G}=\mathbf{A}\otimes\mathbf{G_0}$; $\mathbf{H_0}$ is a diagonal matrix with variance for herd-test-date effect for each lactation, $\mathbf{P_0}$ and $\mathbf{G_0}$ are covariance matrices for permanent environmental and additive genetic random regression coefficients, respectively; **R** is a diagonal matrix of residual variance for each lactation, and **A** is an additive genetic relationship matrix.

Approximation method

The procedure for approximation reliabilities used in this study was described in details by Jamrozik et al. (2000). The method consists of three steps:
   1. estimation of ENP due to the animal's own records
   2. accumulation of progeny contribution to parents
   3. accumulation of contributions of remaining relatives for each animal.

In the first step of the procedure the coefficient matrix $\mathbf{C_i}$ for each animal with record is created using the following formula:

$$\mathbf{C}_i = \mathbf{Z}_i' \mathbf{R}^{-1} \mathbf{Z}_i + \mathbf{G_0}^{-1}$$

where $\mathbf{Z_i}$ is a part of matrix **Z** corresponding to animal i. In order to account for a contemporary group (herd - test day) size the residual variance corresponding to a particular observation was adjusted by using the weight

$$w_j = (cgs_j - 1) / cgs_j$$

where $\mathbf{cgs_j}$ is the number of animals in the same lactation, tested on the same day in the same herd. Further absorption of environmental effects and the calculation of prediction error variance to determine ENP are carried out according to procedure by Jamrozik et al. (2000). Equivalent numbers of progeny for each lactation are assigned equal weights, and the reliabilities of the estimated breeding value for the average yield in the first three lactations are then calculated.

Implementation

The data set consisted of test day records from the first three lactations formed by randomly selecting 317 herds from the population of Polish Black-and-White cattle. There were 45 640, 25 822, and 13 741 test day records of the first, second and third lactation milk yield, respectively. Pedigree data included 5693 cows with records and 1385 bulls with an average number of daughters with data equal to 3.74. Only 112 sires had at least 10 daughters. Reliabilities were calculated for the estimated breeding value of average lactation milk yield for all animals in the data. Average daily milk yield heritability was equal to 0.18, 0.17, and 0.18 for the first, second and third lactation, respectively (Strabel et al., 2005).

To examine the effectiveness of the proposed method, approximate reliabilities were compared with exact reliabilities obtained by the inversion method using the BLUPF90 software package (Misztal et al., 2002). There were 242 415 equations in the mixed model equations system. For comparison purposes, correlations between reliabilities from both methods, regression of approximation on exact reliabilities and simple statistics of differences between both reliabilities were calculated.

RESULTS AND DISCUSSION

Correlations between reliabilities obtained by approximation and the exact method are presented in Table 1 for different groups of animals. The values of these correlations were very high, larger than 0.98 and 0.94 for bulls and for cows

Table 1. Correlations between exact and approximate reliabilities

Item	Contemporary group size	
	with correction	no correction
Bulls (n=1385)	0.981	0.972
Cows with records (n=5693)	0.943	0.822
All animals (n=10887)	0.978	0.939

with records, respectively. These estimates were clearly lower when the procedure did not account for the number of a given cow's contemporaries in the herd-test-day class. The corresponding correlations were equal to 0.97 for bulls and 0.82 for cows with records. That confirms the importance of accounting for small herds in the process of calculating genetic evaluations along with their reliabilities. The higher sensitivity of cows' reliabilities to the contemporary group size not being taken into account is due to the lower number of records contributing to reliabilities. Those records more often may belong to single observation classes.

Summary statistics of the differences between reliabilities obtained by the approximation and the exact methods are presented in Table 2. Positive mean

Table 2. Simple statistics of differences between exact and approximate reliabilities

Item	Mean	SD	Min	Max
Bulls (n=1385)	0.0337	0.0405	-0.0480	0.4711
Cows with records (n=5693)	0.0637	0.0323	-0.1367	0.2259
All animals (n=10887)	0.0407	0.0386	-0.1759	0.4711

values of the differences implied that the approximation method lead to an overestimation of results, although relatively small standard deviations of differences were noticed (0.04 for bulls and 0.03 for cows with records). Overestimation of reliabilities was also found by Jamrozik et al. (2000), who used the same procedure, but they did not account for the size of herd - test day classes. These authors suggested that overestimation was caused by the fact that the procedure for calculating ENP did not take into account the distribution of records in contemporary groups. The maximum difference between approximate and exact reliabilities for bulls was equal to 0.47. A closer investigation of the differences between both reliabilities showed that overestimation occurs mainly for sires, which daughters were the only cows in a particular herd - test day class.

Figure 1 presents the plot of approximate by exact reliabilities. Most of the reliabilities are in the lower range because relatively small numbers of records used
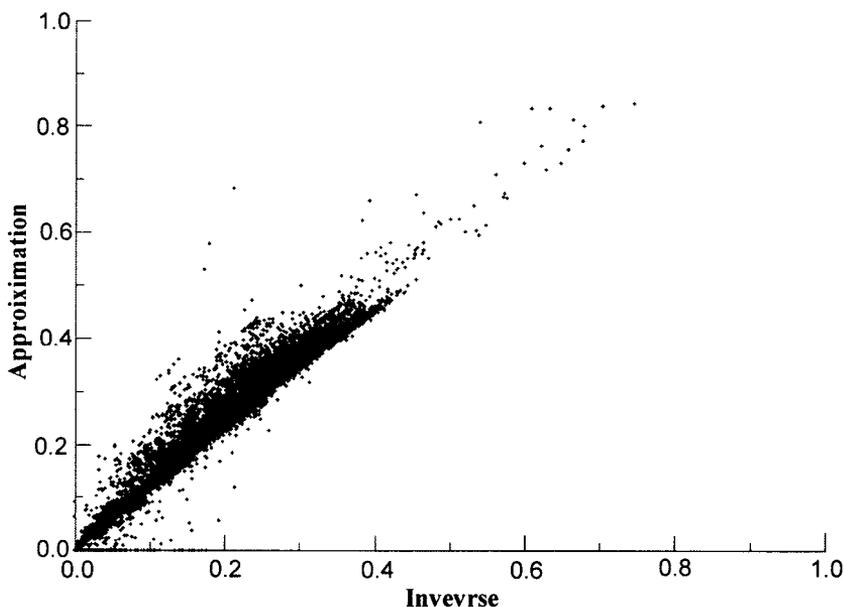

Figure 1. Approximate by exact reliabilities for all animals (n=10887)

in multiple-lactation analyses resulted in most animals having limited information. Most of the points are located above the diagonal confirming overestimation of approximate reliabilities. On the other hand, several approximations located on or very close to the X-coordinate are examples of underestimation of approximation reliabilities. A detailed analysis of this underestimation revealed that the problem is associated with animals having all records in single HTD classes. For such animals no contribution from records is taken into account in an approximation method. The inversion method, however, accounts for the fact that the HTD effect is random and therefore even cows with no contemporaries provide certain information for calculating reliabilities.

Table 3 presents results of regression analysis of exact on approximated reliabilities for different groups of animals. Although intercept values were very

Table 3. Comparison of exact and approximated reliabilities

Item	Bulls	Cows with records	All animals
Regression coefficient	0.776	0.837	0.787
Intercept	-0.001	-0.014	0.002
R^2, %	96.3	89.0	95.6

close to zero, regression coefficients were smaller than 1, which confirms a general problem of a minor overestimation tendency of the approximation procedure. Tier and Meyer (2004) also analysed regression of exact reliabilities on approximations of Jamrozik et al. (2000) for growth traits in the beef cattle evaluation carried out in single herds. Regression coefficients of exact on approximation reliabilities were larger (0.80-0.87) and the larger part of the variation (97.4-99.1) was explained by the regression. In the current study regression described a smaller amount of variation (95.6 for all animals). The differences were probably caused by the structure of the data set used in this study, where the average size of the contemporary group was very small, much lower than in the example herd analysed by Tier and Meyer (2004).

Discrepancies between exact and approximation reliabilities should be much lower in an actual national genetic evaluation system. A relatively small data set had to be used in this study so that exact reliabilities could be obtained by the inversion method. This requirement had definitely worsened the problem of small contemporary groups.

CONCLUSIONS

Reliabilities of dairy cattle test day genetic evaluations based on the multiple-lactation random regression model obtained by the approximation method were satisfactorily accurate. The main difference from exact reliabilities calculated by

the inversion method were caused by the poor structure of the data sets with large numbers of small herd-test-day classes, leading to contemporary groups filled out by half-sibs only. Very low computer requirements make this method feasible for routine applications.

REFERENCES

Jamrozik J., Schaeffer L.R., Jansen G.B., 2000. Approximate accuracies of prediction from random regression models. Livest. Prod. Sci. 66, 85-92

Koots K.R., Schaeffer L.R., Jansen G.B., 1997. Approximate accuracy of genetic evaluation under and animal model. J. Dairy Sci. 80, Suppl. 1, 226 (Abstr. P325)

Meyer K., 1989. Approximate accuracy of genetic evaluation under an animal model. Livest. Prod. Sci. 21, 87-100

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H., 2002. BLUPF90 and related programs (BGF90). Proceedings of the 7th World Congress on Genetic Applied Livestock Production. Montpellier (France), p. 28

Misztal I., Wiggans G.R., 1989. Approximation of prediction error variance in large-scale animal models. J. Dairy Sci. 71, Suppl. 2, 27-32

Strabel T., Misztal I., Bertrand J.K., 2001. Approximation of reliabilities for multiple-trait model with maternal effects. J. Anim. Sci. 79, 833-839

Strabel T., Szyda J., Ptak E., Jamrozik J., 2005. Comparison of random regression test-day models for Polish Black-and-White cattle. J. Dairy Sci. 88, 3688-369

Tier B., Meyer K., 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. J. Anim. Breed. Genet. 121, 77-89

Tier B., Schneeberger M., Hammond K., Fuchs W., 1991. Determining the accuracy of estimated breeding values in Multiple Trait Animal Models. Proceedings of the 9th AAABG Conference, pp. 239-242