Genetic Evaluation of Somatic Cell Scores Using a Random Regression Test Day Model for a Very Large Dairy Cattle Population

Z. Liu, F. Reinhardt, A. Bünger, J. Jaitner and R. Reents
VIT, Heideweg 1, D-27283 Verden, Germany

Abstract

A random regression test day model has being routinely used for genetic evaluation of somatic cell scores for Austrian, German, and Luxembourgish Holstein, Jersey and Red dairy cattle breeds since May 2003. The model includes fixed herd-test-date-parity-milking-frequency effects, fixed lactation curves modelled with Wilmink function, and random genetic and permanent environmental effects modelled with Legendre polynomials. A lactation based iteration scheme is implemented for solving the very large equation system with more than 250 million equations. For reliability approximation, the multiple trait effective daughter contribution method is used. The genetic evaluation system was validated with Interbull validation method III by comparing proofs from a complete evaluation with those from an evaluation based on a data set excluding the most recent four years. The genetic trend estimate was in the allowed range and correlations between the two sets of proofs were very high. Additionally, the random regression model was compared to the previous genetic evaluation model, a fixed regression model. The correlations of proofs between both models were, as expected, very high for bulls with complete daughter information. The correlations of bulls decreased with increasing incompleteness of daughter performance information.

1. Introduction

For selection against mastitis in dairy cattle, somatic cell scores (SCS) are chosen as indicator trait in most countries (Mark *et al.,* 2000). An increasing number of countries have adopted a random regression test day model (RRTDM) for genetic evaluations of production traits and/or SCS (De Roos *et al.,* 2002; Haile-Mariam *et al.,* 2001; Liu *et al.,* 2001*b*; Schaeffer *et al.,* 2000; Strandén and Lidauer, 1999). Since May 2003, a RRTDM has being used for routine genetic evaluation of SCS for Austrian, German and Luxembourgish Holstein, Red and Jersey dairy cattle breeds. This paper describes the genetic evaluation system and presents some results from verification and validation studies of the system.

2. Materials and Methods

2.1. Data

All test day records collected since 1990 are used in routine SCS genetic evaluations subject to certain data edits. Allowed range of days in milk (DIM), 5 to 365, is imposed on all tests. Table 1 describes data sets used in a test run for February 1999 and a routine evaluation for August 2003 and shows annual growth rates of the data since February 1999. As a result of shorter history of test day than lactation data and order of fit for the time dependent effects, the number of equations of RRTDM increases much faster over time than that of lactation models. According to the estimated annual growth rates in Table 1, total number of equations would reach 300 millions in November 2005.

2.2. Genetic evaluation model

For joint evaluations of Austrian, German and Luxembourgish Holstein, Red and Jersey dairy breeds, a multiple lactation RRTDM is applied to test day SCS from first three lactations:

$$y_{ijklo} = h_{lk} + \sum_{m=1}^{3} \beta_{jlm} f_{jlm} + \sum_{m=1}^{3} b_{ilm} p_{ilm} + \sum_{m=1}^{3} b_{ilm} a_{ilm} + e_{ijklo},$$

where y_{ijklo} is SCS of the o-th test day of lactation l of cow i; h_{lk} is fixed effects of the k-th herd-test-date-parity-milking-frequency (HTD) for lactation l; f_{jlm} represents the m-th regression coefficient of the j-th fixed lactation curve of lactation l; β_{jlm} is the m-th term of Wilmink function with $\beta_{.1} = 1, \beta_{.2} = d$ and $\beta_{.3} = e^{-0.05d}$ and d denoting DIM; a_{ilm} and p_{ilm} are the m-th random regression coefficient of lactation l of cow i for genetic and permanent environmental effects, respectively; b_{ilm} is the m-th term of Legendre polynomials with three parameters $b_{.1} = 1, b_{.2} = \sqrt{3}z$ and $b_{.3} = \frac{1}{2}\sqrt{5}(3z^2 - 1)$ and $z = (d-5)/150 - 1$; and e_{ijklo} is error effect. Fixed lactation curves are defined by parity number, calving year, calving age, calving season, F$_1$ crossbred and breed effects. HTD effects of second and third lactations are combined in order to avoid too small contemporary groups. No adjustment for heterogeneous herd variances is conducted for SCS genetic evaluations. Genetic groups are defined for unknown parents of animal based on breed, sex, year of birth and country of origin of the animal. At least six generations of pedigree are traced back from the cows in test day data set. Genetic parameters for SCS evaluations, expressed on average daily basis, are given in Table 2, which were estimated using the covariance function approach incorporated with an iterative two-step estimation procedure (Liu *et al.*, 2000). For first three lactations the economic weights, 0.26, 0.37, and 0.37, are used to combine lactation estimated breeding values (EBV).

Table 1. Description of the data sets used in genetic evaluations of February 1999 test and August 2003 official runs and estimated annual growth rates of the data since February 1999.

Genetic evaluation	No. of cows with records	No. of animals in pedigree	No. of lactations	No. of HTD classes	No. of test day records	No. of equations
February 1999	7,103,041	11,698,794	13,972,569	10,158,954	114,908,280	179,386,933
August 2003	10,836,038	15,655,705	22,120,601	14,883,894	187,698,637	253,317,142
Annual growth rate (%)	9.8	6.7	10.7	8.9	11.5	8.0

Table 2. Parameters of somatic cell scores on average daily basis: heritabilities on diagonal, genetic correlations above diagonal and phenotypic correlations below diagonal.

Lactation	1	2	3
1	.16	.95	.89
2	.34	.16	.97
3	.28	.42	.17

For solving the very large mixed model equation system, a lactation based iteration scheme (Liu *et al.,* 2001*b*) is implemented, which has been proven to be twice as fast as the traditional test day based iteration scheme, because it eliminates operations on test day levels during the iteration process. The solving program makes efficient use of RAM, c.a. 5.6 Gb for the August 2003 evaluation. For approximating reliabilities of EBV, the multiple trait effective daughter contribution (MTEDC, Liu *et al.,* 2001*a* and 2001*b*) is used. Average yield deviations of cows and daughter yield deviations of bulls are calculated for SCS, too.

3. Results and Discussion

The genetic evaluation system was tested and has been routinely in use since May 2003 on a HP Unix server HP9000/L2000 equipped with three 420Mhz CPUs and a total of 16 Gb RAM. The lactation based iteration scheme took 18.1 minutes CPU time per round of iteration on a single CPU in August 2003 evaluation. A total of 400 rounds of iteration were performed in routine genetic evaluations, starting with solutions from the previous evaluation. The convergence criterion, defined as sum of squares of differences in solutions between two consecutive rounds of iteration divided by sum of squares of solutions from the last round, reached $10^{-9.6}$ in August 2003 evaluation for genetic effect that had the slowest convergence rate among all the effects.

For validating genetic trend using Interbull Method III (Boichard *et al.,* 1995), a test evaluation using data till February 1999 was compared to the February 2003 genetic evaluation. The trend estimate was 0.09% of genetic standard deviations, which satisfies the requirement for genetic trend estimates by Interbull. Table 3 shows the correlations of EBV from both evaluations for Black and White Holstein bulls. Bulls born from 1984 to 1989 had already complete daughter information in February 1999 evaluation, therefore, the correlations between both evaluations are above 0.99. With increasing incompleteness of daughter performance information of bulls, correlations between both evaluations decreased. Because bulls born in 1993 had only first lactation daughters at the February 1999 evaluation, correlations of later lactation EBV between both evaluations were lower than the correlation of first lactation EBV.

Table 3. Correlations of EBV of the RRTDM between evaluations February 1999 and February 2003 by birth years for Black and White Holstein bulls with at least 35 daughters in at least 10 herds.

Birth year	No. bulls	Lactation 1	Lactation 2	Lactation 3	Combined
1984	204	.996	.996	.995	.996
1985	415	.996	.996	.995	.996
1986	567	.995	.994	.993	.994
1987	625	.994	.994	.992	.994
1988	691	.993	.993	.991	.993
1989	560	.990	.990	.987	.990
1990	615	.981	.978	.975	.978
1991	644	.972	.973	.968	.973
1992	700	.975	.964	.950	.963
1993	779	.924	.891	.869	.892

The RRTDM was compared to the previous genetic evaluation model, a fixed regression test day model (Reents *et al.,* 1995) based on the same data used in February 2003 genetic evaluation. Table 4 shows correlations of lactation EBV between both test day models. For bulls born before 1994, correlations between both models are very high, about 0.99, with an exception of later lactation EBV of bulls born in 1986, this is due to the fact that later lactations of daughters of the bulls were not included in genetic

evaluations due to step-wise selection of test day data from different lactations by year. Young bulls in progeny testing program have lower correlations than bulls with complete daughter performance information. With increasing incompleteness of data, in terms of length of lactation and number of missing lactations, correlations between both models decreased, indicating the two models make different projection of lactation breeding values.

4. References

Biochard, D., Bonaiti, B., Barbat, A. & Mattalia, S. 1995. Three methods to validate the estimation of genetic trend for dairy cattle. *J. Dairy Sci. 78,* 431-437.

De Roos, A.P.W., Harbers, A.G.F. & de Jong, G. 2002. Herd specific random regression curves in a test-day model for protein yield in dairy cattle. *Proc. 7WCGALP,* Communication #01-05.

Haile-Mariam, M., Goddard, M.E. & Bowman, P.J. 2001. Estimates of genetic parameters for daily somatic cell count of Australian dairy cattle. *J. Dairy Sci. 84,* 1255-1264.

Liu, Z., Reinhardt, F. & Reents, R. 2000. Parameter estimates of a random regression test day model for first three lactation somatic cell scores. *Interbull Bulletin 26,* 61-66.

Liu, Z., Reinhardt, F. & Reents, R. 2001*a*. The effective daughter contribution concept applied to multiple trait models for approximating reliability of estimated breeding values. *Interbull Bulletin 27,* 41-47.

Liu, Z., Reinhardt, F., Bünger, A., Dopp, L. & Reents, R. 2001*b*. Application of a random regression model to genetic evaluations of test day yields and somatic cell scores in dairy cattle. *Interbull Bulletin 27,* 159-166.

Mark, T., Fikse, F., Banos, G., Emanuelson, U. & Philipsson, J. 2000. Summary of Mace pilot runs for somatic cell count and mastitis. *Interbull Bulletin 26,* 43-52

Reents, R., Dekkers, J.C.M. & Schaeffer, L.R. 1995. Genetic evaluation for somatic cell score with a test day model for multiple lactations. *J. Dairy Sci. 78,* 2858-2870.

Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J. & Van Doormaal, B.J. 2000. Experience with a test day model. *J. Dairy Sci. 83,* 1135-1144.

Strandén, I. & Lidauer, M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci. 82,* 2779-2787.

Table 4. Correlations of EBV of February 2003 evaluations between the random and fixed test day models by birth years for Black and White Holstein bulls with at least 35 daughters in at least 10 herds.

Birth year	No. bulls	Lactation 1	Lactation 2	Lactation 3	Combined
1986	567	.981	.976	.981	.982
1987	625	.987	.984	.986	.989
1988	692	.992	.985	.985	.989
1989	561	.994	.987	.985	.991
1990	618	.995	.988	.988	.992
1991	645	.994	.983	.982	.988
1992	702	.993	.982	.984	.988
1993	799	.992	.983	.985	.988
1994	868	.994	.984	.983	.988
1995	970	.994	.979	.976	.985
1996	1020	.992	.966	.962	.974
1997	989	.986	.961	.953	.967