Parameter Estimates of a Random Regression Test Day Model for First Three Lactation Somatic Cell Scores

Z. Liu, F. Reinhardt and R. Reents
United Datasystems for Animal Production (VIT), Heideweg 1, D-27280 Verden, Germany

Introduction

Somatic cell counts have being routinely recorded in most dairy countries. The availability of the data, uniform trait definition and reasonable connectedness across countries make it feasible to evaluate animals on international level (Mark et al., 2000). For national genetic evaluations both single trait repeatability model applied to lactation averages of somatic cell scores (SCS) as well as multiple trait test day models have being used. For estimating parameters of test day yields or SCS, the covariance function approach incorporated with an iterative two-step algorithm has been proven to be an efficient way, as it enables analysing very large data set (Liu et al. 2000). The objectives of this study were 1) to estimate parameters of first three lactation test day SCS via the covariance function approach, and 2) to derive parameters of lactation SCS using the test day parameter estimates.

Material and Methods

Original data from August 2000 German Holstein genetic evaluation were selected matching the following criteria: herd-test-date-parity (HTD) classes with at least five records, supervised monthly testing with two times milkings only, and calving years for first three lactations no earlier than 1993, 1994 and 1995 respectively. One test day record was randomly chosen in case of multiple records within each of the six lactation stages defined on days in milk (DIM): 5-50, 51-105, 106-160, 161-215, 216-259, 260-305. Only full lactations were kept for estimating parameters. Cows were allowed to have later lactation records missing in order to remove the bias caused by selection. Sires with fewer than 30 daughters were discarded to achieve a better data structure. The original pedigree file from the routine genetic evaluation was used to extract pedigree information for all ancestors of cow sires. Table 1 shows the structure of the final test day data set and sire pedigree file used in parameter estimation. For each of the three lactations, 60 fixed lactation curves were defined based on three calving seasons, five classes of age at calving and four breed-region classes. Of the total number of test day records, 49%, 32% and 19% belong to first, second and third lactations, respectively.

Table 1. Description of the final data set and sire pedigree file for parameter estimation

Factors	Cows	Sires of cow	Test day records in total	HTD of all lactations	Fixed lactation curves in total	Animals in sire pedigree file
Size	1,727,682	5,415	17,161,866	3,336,178	180	10,645

A multi-trait sire model was applied to the first three lactation test day SCS to estimate (co)variance components of the six lactation stages:

$$y_{ijklmn} = \mu_{lm} + HTD_{il} + \sum_{p=1}^{5} \beta_{jlp} v_{pd} + s_{klm} + e_{ijklmn} \quad [1]$$

where

y_{ijklmn} is test day SCS at lactation stage m of lactation l of cow n, μ_{lm} is general mean for lactation stage m of lactation l, HTD_{il} is the i-th herd-test-date effect of lactation l, v_{pd} is the

p-th parameter of Ali-Schaeffer function for DIM *d*, β_{jlp} is the *p*-th fixed regression coefficient for lactation *l* specific to subclass *j* of fixed lactation curves, s_{klm} is additive genetic effect of sire *k* for lactation stage *m* of lactation *l*, and e_{ijklmn} is the residual effect. Test day SCS from different lactation stages are treated as genetically distinct traits in the above model.

The iterative two-step algorithm (Liu et al., 2000) was applied to estimate the (co)variance components of lactation stages for three lactations jointly. This approach was proven to yield identical results as one-step approach (Royle and Berliner, 1999), provided the iteration process is converged. Residual maximum likelihood estimates of the (co)variance components for lactation stages were obtained via VCE (Neumaier and Groeneveld, 1998). Due to the large number of components to be estimated for three lactations, the estimation task was partitioned into seven 9-trait analyses to obtain parameter estimates of all 18 traits. The iteration process was stopped when all (co)variance components and estimated breeding values (EBV) were converged upon predefined convergence criteria. After the iteration process was completed, simple averages of the (co)variance estimates from the seven parallel runs were calculated for derivation of (co)variances of random regression coefficients (RRC).

Third order normalised orthogonal Legendre polynomial was chosen to smooth the estimated (co)variance matrices of lactation stages in order to derive (co)variances of RRC. The extended Kirkpatrick et al.'s weighted least squares method (Liu et al., 2000) was applied to estimate (co)variances of genetic RRC and to separate the (co)variances of permanent environmental RRC from error effects, because the extended weighted least squares method can account for different accuracies of the (co)variance estimates of lactation stages and also appears to guarantee the positive definiteness of the derived (co)variance matrices of RRC.

Based on the (co)variances of RRC, genetic parameters for lactation records of any length can be derived. The heritability for lactation records is calculated as follows:

$$h_L^2 = \frac{\sum_i^L \sum_j^L \sigma_{g(i,j)}}{\sum_i^L \sum_j^L \sigma_{g(i,j)} + \sum_i^L \sum_j^L \sigma_{p(i,j)} + \sum_i^L \sigma_{e(i,i)}}$$

[2]

where

h_L^2 is heritability of a lactation record of length of *L* days,
$\sigma_{g(i,j)}$ is genetic covariance between DIM *i* and *j* (*i*, *j* = 1, ..., *L*),
$\sigma_{p(i,j)}$ is permanent environmental covariance between DIM *i* and *j*, and
$\sigma_{e(i,i)}$ is error variance at DIM *i*.

For lactations of length of *L* days, genetic correlation between two lactations is:

$$r_{g(k,l)} = \frac{\sum_i^L \sum_j^L \sigma_{g(i_k, j_l)}}{\sqrt{\sum_i^L \sum_j^L \sigma_{g(i_k, j_k)}} \sqrt{\sum_i^L \sum_j^L \sigma_{g(i_l, j_l)}}}$$

[3]

where

$r_{g(k,l)}$ is genetic correlation between lactation *k* and *l* of length *L* days,
$\sigma_{g(i_k, j_l)}$ is genetic covariance between DIM *i* of lactation *k* and DIM *j* of lactation *l*,
$\sigma_{g(i_k, j_k)}$ is genetic covariance between DIM *i* and *j* of lactation *k*, and
$\sigma_{g(i_l, j_l)}$ is genetic covariance between DIM *i* and *j* of lactation *l*.

Results and Discussion

Fortran 90 programs and Unix shell scripts were developed for estimating the parameters of the multiple trait sire model. The computation was conducted on a HP9000 K460 computer running HP-UX 11 and a Pentium III PC running Linux. For the first round of the iterative two-step algorithm, sire EBV from the routine SCS genetic evaluation with a fixed regression test day model were used as starting values. Four iterative steps were required to get both (co)variance estimates and sire EBV converged. Heritability, genetic and residual correlations did not vary at third decimal place between the third and fourth iterative steps. The whole estimation took considerable time and memory resources. To derive the (co)variances of RRC based on the (co)variance estimates of lactation stages, Maple 6 programs were written for Microsoft Windows system. For separating (co)variances of RRC for permanent environmental effects from error effects using the extended weighted least squares method, 25 rounds of iteration, with the (co)variance matrix of genetic RRC as starting value, were conducted to achieve high accuracy**.**

Derived heritability values: Figures 1 and 2 show derived daily heritability values and heritability values for lactation records, respectively. Fairly homogeneous daily heritability, with an average of 0.09, was observed throughout the course of lactation. Though second and third lactations have 0.01 higher heritability than first lactation, no significant differences in heritability values were found between lactations. The average daily heritability values agree with those estimates using a fixed regression test day model (Reents et al., 1995) as well as those using a multiple trait animal model (Reents et al., 1994), but these daily heritability estimates are strikingly lower than those obtained via a random regression model approach (Jamrozik et al., 1997). In Figure 2 it can be seen that the heritability of lactation record increases as lactation makes progress, which can be explained by higher genetic than phenotypic correlations between DIM of the same lactation. When a lactation record is completed, its heritability reaches approximately 0.20, that is higher than those estimates obtained based on lactation models (Mark et al., 2000). It can be concluded that the use of test day models, which account for environmental effects specific to each test day, leads to higher heritability estimates.

Figure 1. Daily heritability values of first three lactation SCS

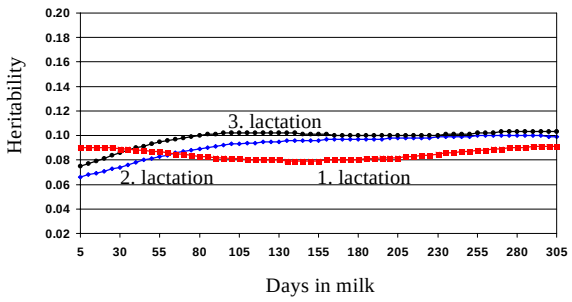
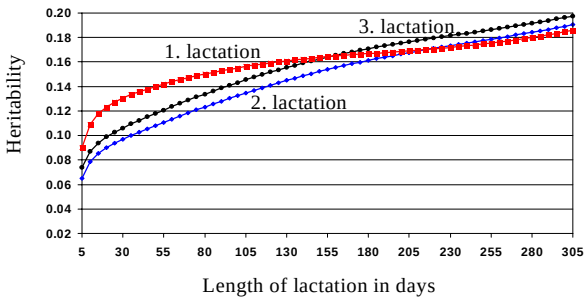


Figure 2. Heritability values of lactation SCS as lactation progresses



Derived genetic covariance structure: DIM 30 at the beginning, DIM 150 in the middle, and DIM 250 at the end of lactation were chosen to describe the genetic covariance structure of test day SCS and the results for first and second lactations are shown in Figures 3 and 4, respectively. Genetic correlations between two ends of lactation, e.g. between DIM 30 and 305 or between DIM 5 and 250, are 0.71 or 0.54 for first lactation SCS, which are higher than for test day yields (Liu et al., 2000). Daily SCS of second lactation are less correlated than daily SCS of first lactation, but the difference in within-lactation genetic correlations is less evident between lactations than test day yields (Liu et al., 2000). The genetic correlation structure was almost identical between second and third lactations.

Figure 5 presents genetic correlations between the same DIM of two lactations. Between first and second lactations the genetic correlations range from 0.84 at DIM 305 to 0.94 at DIM 105. Compared to the correlations between first and second lactations, lower genetic correlations were observed between first and third lactations, and this was not observed in the parameters of test day yields (Liu et al., 2000). It can be seen that the middle stages of lactation are more highly correlated between lactations than the two ends of lactation. The genetic correlations of the same DIM between second and third lactation are quite high, above 0.95, which indicates high genetic similarity between second and third lactations.

Figure 3. Genetic correlations between a given DIM and the remaining part of lactation for first lactation SCS

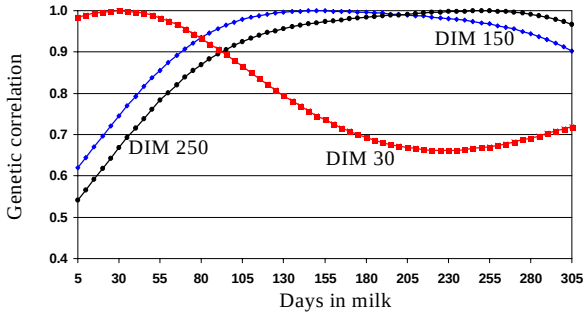


Figure 4. Genetic correlations between a given DIM and the remaining part of lactation for second lactation SCS

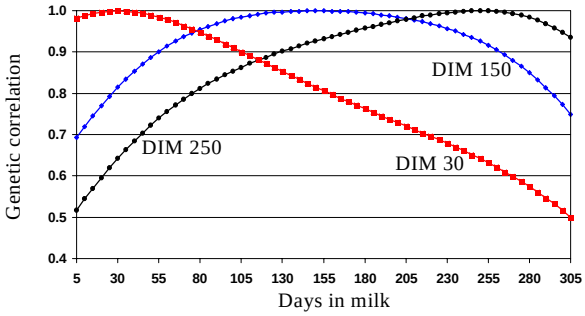


Figure 6 displays the ratio of genetic standard deviations of two lactations for test day SCS. The ratio curves differ drastically to those of test day yields (Liu et al., 2000), suggesting difference in genetic (co)variance structures between SCS and yield traits. The ratio of later to first lactation increases from early stage up to the middle of lactation and then decreases towards the end of lactation.

Figure 5. Genetic correlations between the same DIM of two lactations for test day SCS

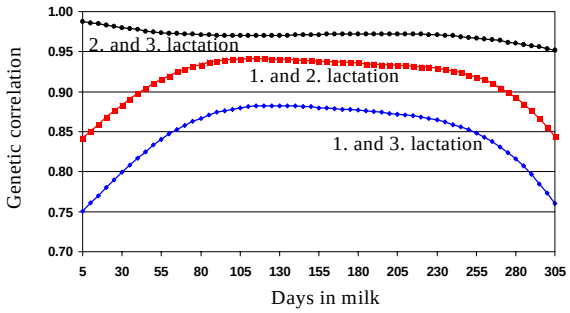
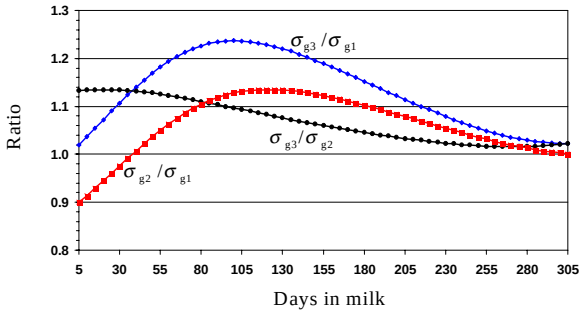


Figure 6. Ratio of genetic standard deviations between two lactations for test day SCS



Derived phenotypic covariance structure: Figure 7 shows phenotypic variances during the course of lactation for test day SCS. Compared to test day yield traits, SCS has also a decreasing phase of phenotypic variance at early stage of lactation but no increasing phase at the end of lactation. Phenotypic variances of test day SCS vary little from the middle through the late stage. Figure 8 presents derived phenotypic correlations between DIM 30, 150 or 250 and the remaining part of lactation for first lactation SCS. It should be noted that the phenotypic correlation at the selected DIM 30, 150 or 250 corresponds to the repeatability value at the same DIM. Daily repeatability values are about 0.5, on average, for all lactations. The rather high repeatability at DIM 30 may be caused partially by the

assumption of homogeneous error variance throughout the course of lactation (Liu et al., 2000).

In general, permanent environmental effects at different DIM of the same lactation are less correlated than additive genetic effects, and consequently the within-lactation phenotypic correlations are smaller than genetic correlations. No negative within-lactation correlations were found for genetic, permanent environmental and phenotypic effects. On a full lactation basis, genetic correlations are 0.95, 0.89 and 0.97 between first and second, between first and third, and between second and third lactations, respectively. These genetic correlations between full lactations are higher than their corresponding averaged daily correlations, because daily SCS are all positively correlated.

Figure 7. Daily phenotypic variances of first three lactation SCS

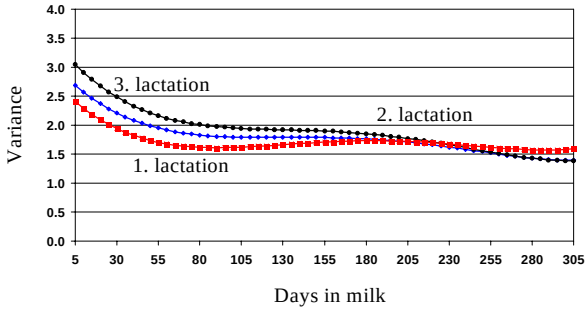
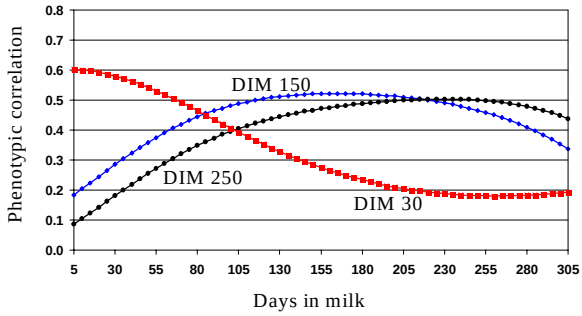


Figure 8. Phenotypic correlations between a given DIM and the remaining part of lactation for first lactation SCS



Summary

The covariance function approach incorporated with an iterative two-step algorithm was applied to estimate genetic parameters of a random regression test day model for first three lactation SCS. Daily heritability values of SCS were found to be homogeneous throughout the course of lactation and between lactations, about 0.09. Test day SCS has both higher within-lactation and between-lactation genetic correlations than test day yield traits. For completed SCS lactation records, heritability can reach as high as 0.20, which is larger than those obtained from a repeatability model applied to lactation averages of SCS. Genetic correlations between full lactations were very high but significantly less than one.

Literature Cited

Jamrozik, J., Schaeffer, L.R., Liu, Z. & Jansen, G. 1997. Multiple trait random regression test day model for production traits. *Interbull Bulletin No. 16*, 43-47.

Mark, T., Fikse, W.F., Sigurdsson, A. & Philipsson, J. 2000. Feasibility of international genetic evaluations of dairy sires for somatic cell count and clinical mastitis. *Interbull Bulletin No. 25*, 154-162.

Liu, Z., Reinhardt, F. & Reents, R. 2000. Estimating parameters of a random regression test day model for first three lactation milk production traits using the covariance function approach. *Interbull Bulletin No. 25*, 74-80.

Neumaier, A. & Groeneveld, E. 1998. Restricted maximum likelihood estimation of covariances in sparse linear models. *Genet. Sel. Evol. 30*, 3-26.

Reents, R., Dekkers, J.C.M. & Schaeffer, L.R. 1994. Genetic parameters of test day somatic cell counts and production traits. *Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Canada. Vol. 17*, 120-123.

Reents, R., Jamrozik, J., Schaeffer, L.R. & Dekkers, J.C.M. 1995. Estimation of genetic parameters for test day records of somatic cell score. *J. Dairy Sci. 78*, 2847.

Royle, J. A. & Berliner, L. M. 1999. A hierarchical approach to multivariate spatial modelling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics. Vol. 1*, 29-56.