Application of a Random Regression Model to Genetic Evaluations of Test Day Yields and Somatic Cell Scores in Dairy Cattle

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

Introduction

Genetic evaluations using test day yields or somatic cell scores (SCS), instead of lactation records, in dairy cattle population have become a common practice in an increasing number of countries (Reents et al., 1998; Schaeffer et al., 2000; Stranden and Lidauer, 2001). Because random regression models make more realistic assumptions on the (co)variance structure of test day data than fixed regression models, research projects, such as parameter estimations for test day yields and SCS (Liu et al., 2000a; 2000b), have been conducted in Germany for implementing a random regression test day model (RRTDM) in national genetic evaluations. In this paper we present a new iteration scheme for solving mixed model equations (MME) of a RRTDM and a new reliability approximation method based on the concept of multiple trait effective daughter contribution (EDC) .

Material and Methods

Model: Test day yields on a 24-hour daily basis from first three lactations are analysed with the following random regression model for milk, fat, protein yield or SCS:

$$\mathbf{y} = \mathbf{X}_1\mathbf{h} + \mathbf{X}_2\mathbf{f} + \mathbf{Zp} + \mathbf{Za} + \mathbf{e}, \qquad [1]$$

where $\mathbf{y}$ is a vector of test day yields adjusted for heterogeneous herd variance (Reents et al., 1998) or SCS of first three lactations of a cow, **h** is a vector of fixed herd-test-date-parity-milking-frequency (HTD) effects, **f** is a vector of regression coefficients of fixed lactation curve effects that are defined by class of calving year, age at calving, season of calving, calving interval classes (not defined for SCS), and breed-region classes (Reents et al., 1998), **p** and **a** represent random regression coefficients (RRC) of permanent environmental (p.e.) and additive genetic effects of the cow, respectively, **e** is a vector of residual effects, $\mathbf{X}_1$, $\mathbf{X}_2$ and **Z** are design matrices for **h**, **f**, and **p** and **a**, respectively. For modelling **p** and **a** the normalised orthogonal third-order Legendre polynomials have been chosen (Liu et al., 2000a),

$$a_1 + a_2\sqrt{3}z + a_3\tfrac{1}{2}\sqrt{5}(3z^2-1) = a_1d_1 + a_2d_2 + a_3d_3$$

with z representing standardised days in milk (DIM) and a_i is the i-th regression coefficient of z on test day yield or SCS (Liu et al., 2000a; 2000b).

Data: Test day data from May 2001 national genetic evaluation for German Holstein and Red breeds were used for testing the genetic evaluation system based on model [1]. Table 1 presents the size of data set after all edits and averaged annual growth rates of data from 1998 to 2001. For the test run the total number of equations to be solved amounted to c.a. 225 millions per trait. With the assumption of the annual growth rates shown in Table 1, the total number of equations would be approximately 260 millions per trait in May 2005.

Milk recording programmes in Germany comprise mainly supervised and unsupersived monthly testing, with increasingly alternate morning and evening testing (AM-PM) programmes (Liu et al., 2000c), and also Lactocorder scheme in which daily milk yield is collected according to monthly testing programme and components according to AM-PM programme. Table 2 shows the weights on daily yields or estimates from the testing programmes in the RRTDM genetic evaluation system. Error variance associated with test day records in genetic evaluations depends on lactation number, DIM and testing programme. Up to now unsupervised tests are assigned equal weights as supervised tests. For test day SCS no differentiated weights are considered in genetic evaluations. Since all test day records collected by milk recording agencies in Germany have milk

yield as well as two components, and even SCS from 1992 onwards, recorded or estimated on a 24-hour daily basis, the little extra gain in accuracy of estimated breeding values (EBV) by applying a multiple trait RRTDM to genetic evaluations is not justified by the enormous increase in computing cost of solving a much larger MME.

Table 1. Size of data set used in May 2001 test run of genetic evaluation for German Holstein and Red breeds and annual growth rates of data averaged over the time period 1998 to 2001

	Number of				
	cows with records	animals in pedigree	Lactations	HTD levels	test day records
May 2001 evaluation	9,013,809	13,671,261	18,070,856	12,662,629	147,564,936
annual growth rate (%)	9.8	6.2	11.6	8.9	12

Table 2. Weights assigned to 24-hour daily yields or estimates in RRTDM genetic evaluations

	Milk yield		Fat yield		Protein yield	
	Morning	Evening	Morning	Evening	Morning	Evening
Monthly testing	1.00		1.00		1.00	
AM-PM testing	0.92	0.90	0.83	0.82	0.90	0.88
- Lactocorder	1.00	1.00	0.83	0.82	0.90	0.88

Scheme for solving MME: Ignoring pedigree information, equations pertinent to cow *i* are:

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{Z} & \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{Z} \\ & \mathbf{X}_2'\mathbf{R}_0^{-1}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}_0^{-1}\mathbf{Z} & \mathbf{X}_2'\mathbf{R}_0^{-1}\mathbf{Z} \\ & & \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{Z}+\mathbf{P}_0^{-1} & \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{Z} \\ \textit{symm.} & & & \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{Z}+a^{ii}\mathbf{G}_0^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{h}} \\ \hat{\mathbf{f}} \\ \hat{\mathbf{p}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{y} \\ \mathbf{X}_2'\mathbf{R}_0^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{y} \end{bmatrix}, \qquad [2]$$

where $\mathbf{G}_0$, $\mathbf{P}_0$ and $\mathbf{R}_0$ are (co)variance matrices of additive genetic, p.e. and error effects, respectively, a^{ii} is the diagonal element of the inverse of relationship matrix for cow *i*.

The iteration on data technique (Schaeffer and Kennedy 1986) is used to solve MME. For updating solutions of each effect at a given round of iteration, adjusted right-hand-sides (**ARHS**) are calculated using the most recent solutions of the other effects and then multiplied with corresponding inverted diagonal blocks according to the Gauss-Seidel algorithm. At round *t* of the iteration process **ARHS** are updated for data contribution in the following sequence:

$$\mathbf{ARHS}_h^{[t]} = \mathbf{X}_1'\mathbf{R}_0^{-1}(\mathbf{y}-\mathbf{X}_2\hat{\mathbf{f}}^{[t-1]}-\mathbf{Z}\hat{\mathbf{p}}^{[t-1]}-\mathbf{Z}\hat{\mathbf{a}}^{[t-1]}), \qquad [3]$$

$$\mathbf{ARHS}_f^{[t]} = \mathbf{X}_2'\mathbf{R}_0^{-1}(\mathbf{y}-\mathbf{X}_1\hat{\mathbf{h}}^{[t]}-\mathbf{Z}\hat{\mathbf{p}}^{[t-1]}-\mathbf{Z}\hat{\mathbf{a}}^{[t-1]}), \qquad [4]$$

$$\mathbf{ARHS}_p^{[t]} = \mathbf{Z}'\mathbf{R}_0^{-1}(\mathbf{y}-\mathbf{X}_1\hat{\mathbf{h}}^{[t]}-\mathbf{X}_2\hat{\mathbf{f}}^{[t]}-\mathbf{Z}\hat{\mathbf{a}}^{[t-1]}), \qquad [5]$$

$$\mathbf{ARHS}_a^{[t]} = \mathbf{Z}'\mathbf{R}_0^{-1}(\mathbf{y}-\mathbf{X}_1\hat{\mathbf{h}}^{[t]}-\mathbf{X}_2\hat{\mathbf{f}}^{[t]}-\mathbf{Z}\hat{\mathbf{p}}^{[t]}). \qquad [6]$$

Solutions of the effects are updated by solving equations, e.g. for p.e. effects $\hat{\mathbf{p}}^{[t]} = (\mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{Z} + \mathbf{P}_0^{-1})^{-1}\mathbf{Z}'\mathbf{R}_0^{-1}(\mathbf{y} - \mathbf{X}_1\hat{\mathbf{h}}^{[t]} - \mathbf{X}_2\hat{\mathbf{f}}^{[t]} - \mathbf{Z}\hat{\mathbf{a}}^{[t-1]})$ After the data contribution has been processed, pedigree is read from the youngest to oldest animals to accumulate pedigree contribution following the Gauss-Seidel algorithm. For each animal in pedigree file **ARHS** for genetic effects are computed in the following sequence:

$$\mathbf{ARHS}_a^{[t]} = \mathbf{ARHS}_a^{[t]} + \tfrac{1}{2} d_a \mathbf{G}_0^{-1}(\hat{\mathbf{s}}^{[t-1]} + \hat{\mathbf{d}}^{[t-1]}), \quad [7]$$

$$\mathbf{ARHS}_s^{[t]} = \mathbf{ARHS}_s^{[t]} + \tfrac{1}{2} d_a \mathbf{G}_0^{-1}\hat{\mathbf{a}}^{[t]} - \tfrac{1}{4} d_a \mathbf{G}_0^{-1}\hat{\mathbf{d}}^{[t-1]}, \quad [8]$$

$$\mathbf{ARHS}_d^{[t]} = \mathbf{ARHS}_d^{[t]} + \tfrac{1}{2} d_a \mathbf{G}_0^{-1}\hat{\mathbf{a}}^{[t]} - \tfrac{1}{4} d_a \mathbf{G}_0^{-1}\hat{\mathbf{s}}^{[t-1]}, \quad [9]$$

where subscripts *a*, *s* and *d* stand for animal, sire and dam, respectively; d_a is the diagonal element pertinent to animal *a* in matrix $\mathbf{D}^{-1}$ (Mrode, 1996 pp 28), and $\hat{\mathbf{a}}$, $\hat{\mathbf{s}}$, $\hat{\mathbf{d}}$ are solutions of genetic RRC for animal, sire and dam, respectively. Rewriting **ARHS** for **h** and for **a** due to data contribution:

$$\mathbf{ARHS}_h^{[t]} = \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{y} - \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{X}_2\hat{\mathbf{f}}^{[t-1]} - \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{Z}\hat{\mathbf{p}}^{[t-1]} - \mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{Z}\hat{\mathbf{a}}^{[t-1]}, \quad [10]$$

$$\mathbf{ARHS}_a^{[t]} = \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{y} - \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{X}_1\hat{\mathbf{h}}^{[t]} - \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{X}_2\hat{\mathbf{f}}^{[t]} - \mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{Z}\hat{\mathbf{p}}^{[t]}, \quad [11]$$

reveals that accumulating data contribution using equations [3] to [6] (Reents et al. 1995, Jamrozik and Schaeffer 2000) has to process through all test day records. This implicitly sets up the matrices: $\mathbf{X}_1'\mathbf{R}_0^{-1}$, $\mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{X}_2$, $\mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{Z}$, $\mathbf{X}_2'\mathbf{R}_0^{-1}\mathbf{Z}$, $\mathbf{Z}'\mathbf{R}_0^{-1}$, and $\mathbf{Z}'\mathbf{R}_0^{-1}\mathbf{Z}$ etc. for every cow several times during each round of iteration process. Such test day record based iteration scheme requires unnecessary and redundant operations, since the matrices do not change over the rounds of iteration. Therefore, a new efficient scheme for updating **ARHS** has been developed for random regression models.

As cows can have different number of lactations and some irregular pattern of missing lactations, updating **ARHS** on a cow basis may not improve the efficiency of the proposed iteration scheme much, compared to a lactation based scheme for updating **ARHS**. Applying equations [10] and [11] to a single lactation results in:

$$\mathbf{ARHS}_{hl}^{[t]} = \mathbf{X}_{1l}'\mathbf{R}_{0l}^{-1}\mathbf{y}_l - \mathbf{X}_{1l}'\mathbf{R}_{0l}^{-1}\mathbf{X}_{2l}\hat{\mathbf{f}}_l^{[t-1]} - \mathbf{X}_{1l}'\mathbf{R}_{0l}^{-1}\mathbf{Z}_l\hat{\mathbf{p}}_l^{[t-1]} - \mathbf{X}_{1l}'\mathbf{R}_{0l}^{-1}\mathbf{Z}_l\hat{\mathbf{a}}_l^{[t-1]}, \quad [12]$$

$$\mathbf{ARHS}_{al}^{[t]} = \mathbf{Z}_l'\mathbf{R}_{0l}^{-1}\mathbf{y}_l - \mathbf{Z}_l'\mathbf{R}_{0l}^{-1}\mathbf{X}_{1l}\hat{\mathbf{h}}_l^{[t]} - \mathbf{Z}_l'\mathbf{R}_{0l}^{-1}\mathbf{X}_{2l}\hat{\mathbf{f}}_l^{[t]} - \mathbf{Z}_l'\mathbf{R}_{0l}^{-1}\mathbf{Z}_l\hat{\mathbf{p}}_l^{[t]}, \quad [13]$$

where subscript *l* denotes lactation number. For a higher efficiency of the lactation based iteration scheme, fixed lactation curve effects are modelled with the same mathematical function as p.e. and genetic effects, i.e. $\mathbf{X}_2$=$\mathbf{Z}$. Prior to the iteration process, the following matrices $\mathbf{R}_{0l}^{-1}$, $\mathbf{Z}_l'\mathbf{R}_{0l}^{-1}$, and $\mathbf{Z}_l'\mathbf{R}_{0l}^{-1}\mathbf{Z}_l$ (note that $\mathbf{X}_{1l} = \mathbf{I}$ and $\mathbf{X}_{2l} = \mathbf{Z}$ for the RRTDM) are computed and stored for every lactation, which are then read in during the iteration process for updating **ARHS** due to data contribution. Because solution and **ARHS** arrays for all the effects usually are defined as multi-dimensional, repeatedly addressing such arrays can be time consuming. Therefore, all solution and **ARHS** arrays are also represented with one-dimensional vectors that share the same memory storage with their corresponding multi-dimensional arrays via Fortran command *equivalence*. The positions of elements of $\mathbf{h}_l$, $\mathbf{f}_l$, $\mathbf{p}_l$, and $\mathbf{a}_l$ in their corresponding solution or **ARHS** arrays are calculated and stored for each lactation before the iteration process.

Diagonal blocks for each level of **h** and **f** and diagonal blocks for **p** and **a** for each cow are inverted and stored prior to iteration process. As diagonal blocks for animals without records ($a^{ii}\mathbf{G}_0^{-1}$) differ only in a^{ii}, different integer values of $10000 * a^{ii}$ are counted and corresponding diagonal blocks are inverted and

stored (Tier and Graser, 1991). Since the number of different $10000 * a^{ii}$ usually is much smaller than the number of animals without records, it is feasible to keep the different inverted diagonals for animals without records in RAM to achieve faster iteration. For every animal in pedigree, positions of elements of **a**, **s**, and **d** in their multi-dimensional solution or **ARHS** arrays are also stored prior to iteration process and will be read in during each round of iteration.

Approximating reliabilities: The multiple trait EDC (Liu et al. 2001) has been applied to approximate reliability values of EBV of the RRTDM. The approximation method consists of three major steps in the sequence: computing own data, progeny and parental contributions.

Computing data contribution for a cow involves absorbing all fixed effects and her p.e. effects into her additive genetic effects. By absorbing the major fixed HTD and ignoring the minor effect of fixed lactation curve effects, the left-hand-side of equations [2] becomes:

$$\begin{bmatrix} \mathbf{Z'MZ} + \mathbf{P}_0^{-1} & \mathbf{Z'MZ} \\ \mathbf{Z'MZ} & \mathbf{Z'MZ} + a^{ii}\mathbf{G}_0^{-1} \end{bmatrix},$$

where $\mathbf{M} = \mathbf{R}_0^{-1} - \mathbf{R}_0^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{R}_0^{-1}\mathbf{X}_1)^{-}\mathbf{X}_1'\mathbf{R}_0^{-1}$. Further absorbing p.e. effects gives $\left[\mathbf{Z'WZ} + a^{ii}\mathbf{G}_0^{-1}\right]$ with $\mathbf{W} = \mathbf{M} - \mathbf{MZ}(\mathbf{Z'MZ} + \mathbf{P}_0^{-1})^{-1}\mathbf{Z'M}$.

Since error covariances between lactations are null, the resulting matrix after the absorption is block-diagonal: $\mathbf{Z'MZ} = diag\{\mathbf{Z}_1'\mathbf{MZ}_1, \mathbf{Z}_2'\mathbf{MZ}_2, \mathbf{Z}_3'\mathbf{MZ}_3\}$ with diagonal block for lactation *I* expressed by:

$$\mathbf{Z}_i'\mathbf{MZ}_i = \begin{bmatrix} \sum_{j=1}^{n_i} d_1^2\sigma_{e_j}^{-2}(1-\frac{\sigma_{e_j}^{-2}}{h_j}) & \sum_{j=1}^{n_i} d_1 d_2\sigma_{e_j}^{-2}(1-\frac{\sigma_{e_j}^{-2}}{h_j}) & \sum_{j=1}^{n_i} d_1 d_3\sigma_{e_j}^{-2}(1-\frac{\sigma_{e_j}^{-2}}{h_j}) \\ & \sum_{j=1}^{n_i} d_2^2\sigma_{e_j}^{-2}(1-\frac{\sigma_{e_j}^{-2}}{h_j}) & \sum_{j=1}^{n_i} d_2 d_3\sigma_{e_j}^{-2}(1-\frac{\sigma_{e_j}^{-2}}{h_j}) \\ \textit{symm.} & & \sum_{j=1}^{n_i} d_3^2\sigma_{e_j}^{-2}(1-\frac{\sigma_{e_j}^{-2}}{h_j}) \end{bmatrix},$$

where n_i is the number of test day records of lactation i for the cow, $\sigma_{e_j}^2$ is error variance associated with the *j*-th test, $h_j = \sum \sigma_e^{-2}$, and h_j is the inverted diagonal for the HTD class the *j*-th test belongs to. The further absorption of p.e. into genetic effects gives: $\mathbf{Z'WZ} = \mathbf{Z'MZ} - \mathbf{Z'MZ}(\mathbf{Z'MZ} + \mathbf{P}_0^{-1})^{-1}\mathbf{Z'MZ}$. The diagonal structure of the matrix $\mathbf{Z'MZ}$ can be taken into consideration to make the calculation of $\mathbf{Z'WZ}$ more efficient. The information contributed by cow's own test day records is $\Psi_d = \mathbf{Z'WZ}$, from which reliability is computed using $\Re_Y = \mathbf{I} - (\Psi_d\mathbf{G}_0 + \mathbf{I})^{-1}$ and then the reliability is converted to EDC using $\Psi_Y = 4[(\mathbf{I} - \Re_Y)^{-1} - \mathbf{I}]\mathbf{G}_0^{-1}$.

Collecting EDC from all progeny is done by processing pedigree file from the youngest to oldest animals (VanRaden and Wiggans 1991). Animal's reliability contributed by progeny adjusted for mate effect is computed as:

$$\Re_{P\text{-}M} = \tfrac{1}{4}[\mathbf{E} - \mathbf{E}(\mathbf{E} + (\mathbf{I} - \tfrac{1}{4}\Re_M^*)^{-1})^{-1}\mathbf{E}],$$

where $\Re_M^*$ is mate's reliability calculated excluding EDC of this progeny and $\mathbf{E} = (\mathbf{I} - \Re_P^*)^{-1} - \mathbf{I}$ with $\Re_P^*$ being the progeny's reliability including information from its performance and its progeny but not from its parents (Liu et al. 2001). Then the reliability is converted to EDC with $\Psi_{P\text{-}M} = 4[(\mathbf{I} - \Re_{P\text{-}M})^{-1} - \mathbf{I}]\mathbf{G}_0^{-1}$ for each progeny. The calculation from the youngest to oldest animals ensures that EDC from all progeny of an animal have been accumulated before calculating the animal's contribution to its parents.

Once EDC from all progeny have been accumulated, reliability is calculated from the oldest to youngest animals to collect parental contribution (VanRaden and Wiggans 1991). This ensures parental reliability is available before progeny reliability is calculated. To avoid double counting animal's contribution to its parents, EDC contributed by the animal to its parents must first be subtracted from its parental EDC. Then the resulting EDC (Ψ^*) are converted to reliability using $\Re^* = \mathbf{I} - (\Psi^* \mathbf{G}_0 + \mathbf{I})^{-1}$ for computing parental contribution to that animal. Formulae $\Re_{\mathrm{PA}} \approx \frac{1}{4}(\Re^*_{sire} + \Re^*_{dam})$ and $\Psi_{\mathrm{PA}} = 4[(\mathbf{I} - \Re_{\mathrm{PA}})^{-1} - \mathbf{I}]\mathbf{G}_0^{-1}$ are used to compute EDC contributed from parent average to the animal (Liu et al., 2001).

When the three steps are done, the total EDC for each animal (Ψ_{T}) is computed with $\Psi_{\mathrm{T}} = \Psi_{\mathrm{PA}} + \Psi_{\mathrm{Y}} + \sum \Psi_{\mathrm{P\text{-}M}}$ (Liu et al., 2001) and converted to reliability with $\Re_{\mathrm{T}} = \mathbf{I} - (\frac{1}{4}\Psi_{\mathrm{T}}\mathbf{G}_0 + \mathbf{I})^{-1}$.

Breeding value on 305-day lactation basis is defined as: $u_{L_i} = \sum_j^{305} u_{ij} = \sum_j^{305} \mathbf{d}_j' \mathbf{a}_i = \mathbf{1'D}\mathbf{a}_i$, and combined lactation breeding value as:

$$u_{comb} = \mathbf{w}' \begin{bmatrix} u_{L_1} \\ u_{L_2} \\ u_{L_3} \end{bmatrix} = \mathbf{w}' \begin{bmatrix} \mathbf{1'D} & & \\ & \mathbf{1'D} & \\ & & \mathbf{1'D} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} = \mathbf{v}'\mathbf{a}$$

where u_{L_i} is 305-day lactation breeding value for lactation *i*, $\mathbf{d}_j$ contains *d*'s for DIM *j*, $\mathbf{a}_i$ contains regression coefficients of genetic effects for lactation *i*, **w** contains economic weights of first three lactations, and $\mathbf{v}' = \mathbf{w}' diag\{\mathbf{1'D}, \mathbf{1'D}, \mathbf{1'D}\}$. Since var(**a**) = $\mathbf{G_0}$ and var($\hat{\mathbf{a}}$)= $\mathbf{G_0}\Re_{\mathrm{T}}$, reliability of combined lactation EBV is $R^2_{\hat{u}_{comb}} = \frac{\mathbf{v}'\mathbf{G_0}\Re_{\mathrm{T}}\mathbf{v}}{\mathbf{v}'\mathbf{G_0}\mathbf{v}}$. By setting the *i*-th element of **w** to one and the others to null, reliability for the *i*-th lactation EBV $R^2_{\hat{u}_{L_i}}$ can be calculated using the same formula as for $R^2_{\hat{u}_{comb}}$.

Results and Discussion

The RRTDM genetic evaluation system has been set up using Fortran 90 and automated via Unix shell scripts. Library and object files are managed via makefile. Record layouts for input and output files are uniquely defined and can be easily modified for further development of the system. Only one parameter file must be changed prior to each genetic evaluation. Further details are available upon request. The system has being tested on a HP Unix computer HP9000/L2000 equipped with two CPUs and 8Gb RAM.

For the test run the following components were kept in RAM: solutions for all the effects, **ARHS** for all but p.e. effect, inverted diagonal blocks for both fixed effects and genetic effects for animals without records. For the lactation based iteration scheme, $\mathbf{Z}_l'\mathbf{R}_{0l}^{-1}\mathbf{Z}_l$ for each lactation was held in RAM, whereas $\mathbf{R}_{0l}^{-1}$ and $\mathbf{Z}_l'\mathbf{R}_{0l}^{-1}$ were read from disk during iteration. The positions of elements of $\mathbf{h}_l, \mathbf{f}_l, \mathbf{p}_l,$ and $\mathbf{a}_l$ for each lactation in their corresponding solution or **ARHS** arrays were kept in RAM and the positions of elements of **a**, **s**, and **d** for each animal in pedigree were read repeatedly from disk. In total, the RAM usage was about 5.4Gb per trait for May 2001 test evaluation. Disk consumption was minimized by using binary storage type.

Figures 1 and 2 show the convergence criterion (CC), that is defined as the sum of squares of differences in solutions between two consecutive rounds of iteration divided by the sum of squares of solutions from the last round, for every effect at each round of iteration for milk and fat yield, respectively. No starting values were used in the test run for any of the effects. At round 433, 298, and 371 CC satisfied the predefined limit 10^{-8} for all the effects for milk, fat and protein yield, respectively. The different rates of convergence among the traits can be partially explained by the fact that RRC of fat yield are least correlated with one another and RRC of milk yield are most correlated. It is noteworthy that the effects converge at a rather different speed, with HTD being fastest followed by p.e., fixed lactation curve and genetic effects. Thus, the genetic evaluation system will allow, after a certain number of initial rounds, variable iteration intensity to be imposed on each individual effect, which can maximize the rate of convergence

within a given time frame in routine genetic evaluations. In the RRTDM environmental trends are represented in HTD effects and fixed lactation curve effects through calving year classes, while genetic trends are shown in breeding values. From Figure 1 it can be seen that highly accurate partitioning of phenotypic trend into genetic and environmental components is reflected in slow improvement in cc for all but p.e. effects in later rounds of iteration. Since p.e. effect does not contain any trend, it converged fastest in later rounds. The slow convergence for fixed lactation curve effects was also caused by small size of some classes and the relatively large number of classes.

The lactation based iteration scheme needed c.a. 12 minutes CPU time per round on a single CPU, which is significantly faster than the test day record based iteration scheme. Because the lactation based iteration scheme, which iterates on MME rather than on data, reduces the computing cost of RRTDM to ordinary multiple trait models, i.e. operations on RRC (namely setting up the matrices $\mathbf{R}_{0l}^{-1}$, $\mathbf{Z}_l'\mathbf{R}_{0l}^{-1}$, and $\mathbf{Z}_l'\mathbf{R}_{0l}^{-1}\mathbf{Z}_l$) are no longer required during iteration. Compared to the fixed regression test day model (Reents et al. 1998) iteration program implemented with the test day record based iteration scheme that required 3.8Gb RAM, the RRTDM iteration program using the new lactation based iteration scheme took only as much as 3.3 times CPU time on the same computer to solve as many as three times more equations. The genetic evaluation system enables the use of EBV from previous evaluation as starting values for iteration process, which can further enhance its run time performance.

Figure 1. Convergence criterion (CC) for the effects of the RRTDM applied to May 2001 test run of genetic evaluation for test day milk yield

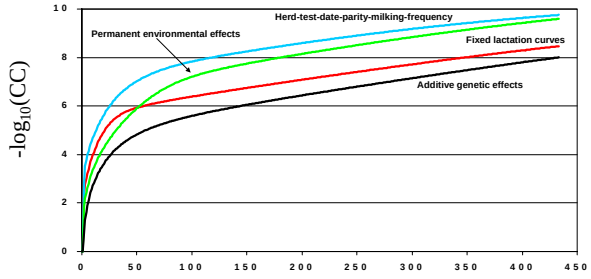
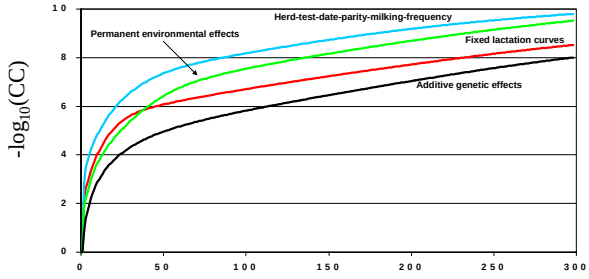


Figure 2. Convergence criterion (CC) for the effects of the RRTDM applied to May 2001 test run of genetic evaluation for test day fat yield



Figures 3 and 4 show reliability values for a cow and a progeny tested bull calculated using the approximation method. The lactation progress is described with increasing number of tests of first three lactations. For the reliability calculation, assumptions were made on reliability values of bull parents and mates as well as cow parents. It can be seen that reliabilities of both individual lactation EBV and combined lactation EBV increase as daughters of the bull or cow herself progress in lactation. The increase in reliability of each individual lactation EBV is slowed down when no more actual test day records are available, e.g. first (second) lactation reliability values have increased at a slower rate after first (second) lactation is completed. The reliability approximation method was developed using both computational techniques and genetic theory, thus it is both efficient and accurate (Liu et al., 2001). A useful feature of this reliability method is that relative weights on individual components of EBV, data and progeny and parental contributions, can be extracted for every EBV so that a possible re-construction of any EBV can be done in order to explain, verify and evaluate genetic evaluation results.

Figure 3. Development of reliabilities of first three lactation and combined EBV of SCS for a cow

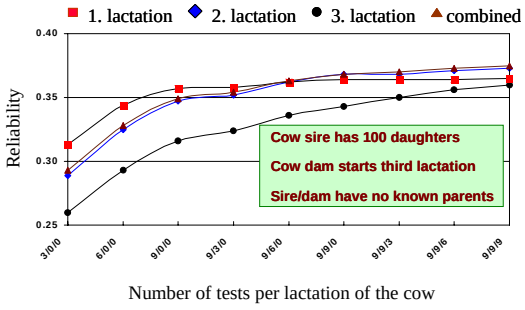
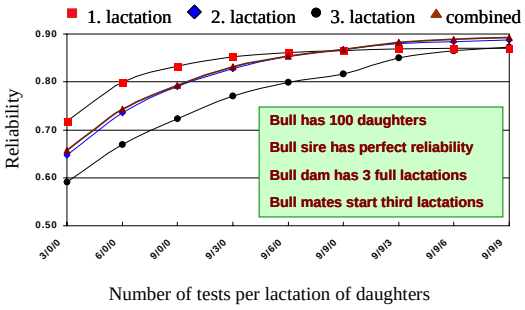


Figure 4. Development of reliabilities of first three lactation and combined EBV of SCS for a progeny tested bull.



Summary

Test day milk, fat, protein yield or SCS on 24-hour daily basis from first three lactations are analysed with a random regression model which includes HTD, fixed lactation curve effects, permanent environmental, and additive genetic effects. A highly automated and efficient genetic evaluation system has been set up for routine breeding value estimation for German Holstein and Red population, for which huge mixed model equations are to be solved, e.g. approximately 225 million equations per trait for the May 2001 test run. To solve such a huge equation system, an efficient lactation based iteration scheme has been developed and its efficiency confirmed with the test run. Compared to the traditional test day record based iteration scheme, the new iteration scheme avoids implicit building of random regression coefficient related matrices for every cow several times at each round of iteration, therefore the new iteration scheme reduces the computing cost of random regression models to ordinary multiple trait models. And this makes it possible for routine genetic evaluations for very large populations like German Holstein and Red breeds to be completed within a reasonable time frame. The lactation based iteration scheme improves run time performance of the system without extra requirement on RAM. The effects in the model converge at rather different rates during the iteration process, which suggests imposing variable iteration intensity to the individual effects in order to maximize the rate of convergence within a given time frame in routine genetic evaluations. A new method using the concept of multiple trait effective daughter contribution was outlined to approximate reliability of EBV from random regression models and was demonstrated with two examples. The method combines genetic theory and computational techniques, therefore it is both accurate and efficient for calculating reliability for random regression models applied to very large populations.

Literature Cited

Jamrozik, J. & Schaeffer L.R. 2000. Comparison of two computing algorithms for solving mixed model equations for multiple trait random regression test day models. *Livest. Prod. Sci. 67*, 143-153.

Liu, Z., Reinhardt, F., Reents, R. & Dopp, L. 1999. Derivation of weights on daughter and pedigree information for estimating breeding values of progeny tested bulls using a test day model. *Interbull Bulletin 22*, 81-87.

Liu, Z., Reinhardt, F. & Reents, R. 2000a. Estimating parameters of a random regression test day model for first three lactation milk production traits using the covariance function approach. *Interbull Bulletin 25*, 74-80.

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

Liu, Z., Reents, R., Reinhardt, F. & Kuwan, K. 2000c. Approaches to estimating daily yield from single milking testing schemes and use of a.m.-p.m. records in test-day model genetic evaluation in dairy cattle. *J. Dairy Sci. 83*, 2672-2682.

Liu, Z., Reinhardt, F. & Reents, R. 2001. The effective daughter contribution concept applied to multiple trait models for approximating reliability of estimated breeding values. *Interbull Meeting in Budapest*, Hungary, 2001.

Mrode, R.A. 1996. *Linear models for the prediction of animal breeding values.* CAB International, U.K.

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.

Reents, R., Dopp, L., Schmutz, M. & Reinhardt, F. 1998. Impact of application of a test day model to dairy production traits on genetic evaluations of cows. *Interbull Bulletin 17,* 49-54.

Schaeffer, L.R. & Kennedy, B.W. 1986. Computing strategies for solving mixed model equations. *J. Dairy Sci. 69,* 575-579.

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.

Stranden, I. & Lidauer M. 2001. Parallel computing applied to breeding value estimation in dairy cattle. *J. Dairy Sci. 84,* 276-285.

Tier, B. & Graser, H.U. 1991. Predicting breeding values using an implicit representation of the mixed model equations for a multiple trait animal model. *J. Anim. Breed. Genet. 108,* 81-88.

VanRaden, P.M. & Wiggans, G.R. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci. 74,* 2737-2746.