An Approach to Compute EDC and DYD for Test-day Models

H. Täubert[*], Z. Liu[*], J. Tarrés[†], V. Ducrocq[‡]

Introduction

National breeding value estimations are based on complex models handling several traits at the same time. In dairy cattle random regression test-day models (RRTDM) are used to estimate breeding values (EBV) for milk, fat and protein yield in separate lactations. Reliabilities derived from these models are difficult to obtain. Single and multiple trait approaches have been used to get approximation (Koots et al., 1997, Jamrozik et al., 2000, Strabel et al., 2001, VanRaden, 2001)

For international genetic evaluations, estimated breeding values from each country are deregressed to obtain a value comparable to daughter yield deviations (DYD) for bulls that have daughters with records. Only single trait informations are currently used to obtain international EBV. A multiple trait MACE would use the complete information from national evaluations based on RRTDM. But EDC and DYD from multiple trait models are required to set up MACE equations. Furthermore, Lidauer et al. (2004) showed a validation method for genetic trends from test day models, which also requires DYD. This method is one out of three to evaluate the correctness of national breeding value estimation models by testing for potential biases in the estimation of genetic trend of a participating country. The mixed model equations of a multiple trait MACE model require the computation of effective daughter contribution (EDC) matrix (Tarrès et al., 2006).

The derivation of multiple trait EDC was described by Liu et al. (2004). In their paper, it was shown how to directly convert multiple trait reliabilities into EDC based on daughter contributions. Tarrés et al. (2007) described the validation of an approximate REML algorithm in a multiple trait MACE model (2007). It was shown in this paper, how the EDC matrix takes all sources of information into account. One problem was an unbiased calculation of parents' EDC values within this EDC-matrix. The approach proposed by Harris and Johnson (1998), to compute single trait reliabilities was extended to the RRTDM situation by Ducrocq and Schneider (2007). This resulted in unbiased symmetric EDC matrices and reliability matrices.
Ducrocq and Schneider (2008) proposed an approach to summarize any number of test-day records of old cows into a reduced number of "pseudo-records" in order to alleviate computations. This approach supposed the construction of a right hand side for each cow.
In this paper, the strategies used in Ducrocq and Schneider (2007, 2008) are combined to compute EDC matrices and DYD vectors for sires.

[*] Vereinigte Informationssyteme Tierhaltung (**vit**), Heideweg 1, 27283 Verden (Aller), Germany
[†]Grup de Recerca en Remugants, Departament de ciència animal i dels aliments, Universitat autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
[‡] INRA, 1313 – Génétique Animale et Biologie Intégrative, 78352 Jouy en Josas, France

Methods

Hypotheses and general strategy

Suppose a Random Regression (RR) animal model with a contemporary group effect (**h**) corresponding to the herd-test date (HTD) effect. The other fixed effects and random effects other than genetic and permanent environment effects (e.g., herd-year effects) will be considered as being accurately estimated and will be ignored for the DYD computation.
The general model is:

$\mathbf{y} = \mathbf{Xh} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e}$ with Var(**e**) = **R**

where **y** refers to the observations of cows. The corresponding mixed model equations are:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}W} \\ \mathbf{Z'R^{-1}W} & \mathbf{Z'R^{-1}Z} + \mathbf{G_0^{-1}} \otimes \mathbf{A^{-1}} & \mathbf{Z'R^{-1}W} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}Z} & \mathbf{W'R^{-1}W} + \mathbf{P_0^{-1}} \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \mathbf{a} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \\ \mathbf{W'R^{-1}y} \end{bmatrix} \quad (1)$$

Solutions $\hat{a}_i$ for all animals are supposed to be available (i.e., already computed),

We want to find sire EDC matrices **Ψ*** and DYD vectors **y***, such that for a particular sire s:
$\mathbf{y}_s^* = \mathbf{Z}_s^* \mathbf{a}_s + \mathbf{e}_s^*$. The mixed models equations are:

$$\left[\mathbf{\Psi}_s^* + \mathbf{G_0^{-1}} \otimes \mathbf{A_s^{-1}}\right]\left[\mathbf{a}_s\right] = \left[\mathbf{\Psi}_s^* \mathbf{y}*\right] \quad (2)$$

System (2) should give the same solutions as (1) for sires.

The strategy used is similar to the method used in Ducrocq and Schneider (2008):

step 1: compute $\mathbf{\Psi}_s^*$ for each sire s

step 2: compute the right hand side of (2) for each sire s

step 3: transform the EDC matrix $\mathbf{\Psi}_s^*$ and the right hand side for sire s into pseudo records and weights, so they can be used directly in a RR animal model.

Step 1: computation of $\mathbf{\Psi}_s^*$

First, the block corresponding to the own records of any cow with data is computed as described in Ducrocq and Schneider (2007). Contemporary group effect and permanent environment effects are absorbed into genetic effect equations, leading to a block matrix B_i^* for each animal i with records. These blocks should be stored for later use. As shown in Liu et al. (2004), the reliability matrix $\mathbf{\Psi}_s^*$ requires these blocks to be absorbed into the block of sire s of animal i, as well as a component involving the full reliability matrix of the dam of i. Therefore, the whole procedure described in Ducrocq and Schneider (2007) will be completed. At the end, reliability matrices of all animals are available, which are needed for all dams for animals with records: $\mathbf{R_d}$

The next task is to successively absorb the equations corresponding to all the daughters of sire s as well as their dam into the sire block. This is a situation similar to equation (11) in Tarrés et al. (2007). Tarrés et al. (2007) developed the submatrix for the triplet sire s – dam d – animal i:

$$\begin{bmatrix} \boldsymbol{\Psi}^*_{\mathbf{s-i}} + (\frac{1}{4}\mathbf{d_i} + \mathbf{d_s})\mathbf{G_0^{-1}} & \frac{1}{4}\mathbf{d_i}\mathbf{G_0^{-1}} & -\frac{1}{2}\mathbf{d_i}\mathbf{G_0^{-1}} \\ & \boldsymbol{\Psi}^*_{\mathbf{d-i}} + (\frac{1}{4}\mathbf{d_i} + \mathbf{d_d})\mathbf{G_0^{-1}} & -\frac{1}{2}\mathbf{d_i}\mathbf{G_0^{-1}} \\ \textbf{symm.} & & \boldsymbol{\Psi}^*_{\mathbf{i}} + \mathbf{d_i}\mathbf{G_0^{-1}} \end{bmatrix} \quad (3)$$

where d_i, d_d, d_s are the mendelian sampling terms corresponding to i, d and s.

Note than in (3), $\boldsymbol{\Psi}^*_{\mathbf{s-i}}$ is building up as more equations of daughters and mates are absorbed. d_s can be ignored at this stage because it will appear later in (2). $\boldsymbol{\Psi}^*_{\mathbf{d-i}}$ is the EDC matrix of dam d after taking into account all sources of information except the contribution of daughter i. Ducrocq and Schneider (2007) described how to obtain $\boldsymbol{\Psi}^*_{\mathbf{d-i}}$ from $\boldsymbol{\Psi}^*_{\mathbf{d}}$ (or similarly, from $\mathbf{R_d}$). For the dam, first $\boldsymbol{\Psi}^*_{\mathbf{d}}$, is obtained from the final $\mathbf{R_d}$, then $\boldsymbol{\Psi}^*_{\mathbf{d-i}}$ is computed using Harris and Johnson's (1998) approach.

In other words, to compute $\boldsymbol{\Psi}^*_{\mathbf{s}}$, we start with $\boldsymbol{\Psi}^*_{\mathbf{s}} = \mathbf{0}$. For all animals with records, we construct (3), then absorb the last two rows and columns of blocks in to the first one:

$$\boldsymbol{\Psi}^*_{\mathbf{s}} = \boldsymbol{\Psi}^*_{\mathbf{s-i}} + \frac{1}{4}\mathbf{d_i}\mathbf{G_0^{-1}} - \begin{bmatrix} \frac{1}{4}\mathbf{d_i}\mathbf{G_0^{-1}} & -\frac{1}{2}\mathbf{d_i}\mathbf{G_0^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi}^*_{\mathbf{d-i}} + (\frac{1}{4}\mathbf{d_i} + \mathbf{d_d})\mathbf{G_0^{-1}} & -\frac{1}{2}\mathbf{d_i}\mathbf{G_0^{-1}} \\ -\frac{1}{2}\mathbf{d_i}\mathbf{G_0^{-1}} & \boldsymbol{\Psi}^*_{\mathbf{i}} + \mathbf{d_i}\mathbf{G_0^{-1}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{4}\mathbf{d_i}\mathbf{G_0^{-1}} \\ -\frac{1}{2}\mathbf{d_i}\mathbf{G_0^{-1}} \end{bmatrix} \quad (4)$$

In practice, rewrite (4) as:

$$\boldsymbol{\Psi}^*_{\mathbf{s=}}\boldsymbol{\Psi}^*_{\mathbf{s-i}} + \frac{1}{4}\mathbf{d_i}\mathbf{G_0^{-1}} - \mathbf{F'}\mathbf{E^{-1}}\mathbf{F} \quad (5)$$

Let **E=TT'.** Then $\mathbf{F'E^{-1}F} = \mathbf{F'T^{-T}T^{-1}F} = \mathbf{K'K}$ where **K** is solution of **TK=F**. So once **E** is computed, one can compute its Cholesky factor **K**, solve **TK = F** for **G** and compute **K'K**.

Step 2: right-hand-side computation

The right hand side **rhs** of (2) can be computed as in Ducrocq and Schneider (2008) by multiplying the left hand side by the vector of known sire EBV.

$$\mathbf{rhs} = \left[\boldsymbol{\Psi}^*_{\mathbf{s}} + \mathbf{G_0^{-1}} \otimes \mathbf{A_s^{-1}}\right]\hat{\mathbf{a}}_{\mathbf{s}} \quad (6)$$

Step 3: transformation of the right hand side into DYD.

A simple transformation converts the right hand side (**rhs)** into pseudo-records or DYD. First, decompose $\boldsymbol{\Psi}^*_{\mathbf{s}}$ as $\boldsymbol{\Psi}^*_{\mathbf{s}} = \sigma_e^2 \sum_{j=1}^{j=rank(\Psi^*_s)} \mathbf{u_j}\mathbf{u_j'}$ (=equation (6) in the appendix of Ducrocq and Schneider (2008)). Take $\mathbf{Z}^*_{\mathbf{s}} = \begin{bmatrix} \mathbf{u_1'} \\ \vdots \\ \mathbf{u_J'} \end{bmatrix}$. Then, the DYD vector for sire s is:

$$\mathbf{y}^*_{\mathbf{s}} = \left(\mathbf{Z}^{*\prime}_{\mathbf{s}}\mathbf{R^{-1}}\right)^{-}\mathbf{rhs}$$

Validation

If the derivation is correct, fitting a random regression model $\mathbf{y}_s^* = \mathbf{Z}_s^* \mathbf{a}_s + \mathbf{e}_s^*$ (where $\mathbf{y}_s^*$ are records and the rows of $\mathbf{Z}_s^*$ are the corresponding regression coefficient) should give the same sire solutions as in (1).

Discussion

Because of the size of the equations system, calculating reliabilities or EDC for random regression EBV is a major computing task. Approximations of single trait reliabilities are often used and they are easy to calculate. But for multiple trait models at national and international level, they are more difficult to compute and often biased. The approach used in this paper describes the progressive computation of sire EDC values as a sum of daughter contributions corrected for their dams. These sire EDC matrices can be used to set up the right-hand side for sires and combining them with the national sire EBV (equation (6)); they lead to multiple trait DYD values. The advantage of this system is that computations are done on a single animal basis. Large matrices are not needed to be set up or inverted. The potential applications of these EDC matrices and DYD vectors are the same as in the single trait situation. For example, they can be used in the estimation of genetic correlations between countries applying different RRTDM evaluations with the approximate EM-REML approach described by Tarrés et al, (2007) There is no loss of information, in contrast with approaches which first reduce the multiple trait problem to a single trait situation.

The procedure was implemented into the French genetic evaluation software *Genekit*.

Acknowledgements:

The financial support by German Holstein Association (DHV) and Union Nationale de Coopératives d'Elevage et d'Insémination Animale (UNCEIA) of France is appreciated.

References

Ducrocq V., Schneider, M.dP. (2007). *Interbull Bulletin* 37, 82-87.
Ducrocq, V., Schneider, M.dP. (2008). *Interbull Bulletin* 38, 27-31.
Harris, B., Johnson, D. (1998). *J. Dairy Sci.*, 81:2723-2728.
Jamrozik, J., Schaeffer, L.R., Jansen, G.B. (2000). *Livest. Prod. Sci.* 66:85-92
Koots, K.R., Schaeffer, L.R., Jansen, G.B. (1997). *J. Dairy Sci.*, 80:(Suppl. 1):226
Lidauer, M., Mäntysaari, E.A., Stranden, I. (2004). *Interbull Bulletin* 32, 142-145.
Liu Z., Reinhardt, F., Bünger, A., Reents, R., (2004) *J. Dairy Sci.*, 87: 1896-1907.
Strabel, T., Misztal, I., Bertrand, J.K., (2001). *J. Anim. Sci.*, 79:833-839
Tarrés, J., Liu, Z., Ducrocq, V., Reinhardt, F., Reents, R., (2007). *J. Dairy Sci.*, 90:4846-4855.
VanRaden, P.M. (2001). *J. Dairy Sci.*, 84:(E. Suppl.):E47-E55