

## Adjustment for Heterogeneous Covariance due to Herd Milk Yield by Transformation of Test-Day Random Regressions

N. Gengler,<sup>1,2</sup> G. R. Wiggans,<sup>3</sup> and A. Gillon<sup>2</sup>

<sup>1</sup>National Fund for Scientific Research, B-1000 Brussels, Belgium

<sup>2</sup>Animal Science Unit, Gembloux Agricultural University, B-5030 Gembloux, Belgium

<sup>3</sup>Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705

### ABSTRACT

A method of accounting for differences in covariance components of test-day milk records was developed based on transformation of regressions for random effects. Preliminary analysis indicated that genetic and nongenetic covariance structures differed by herd milk yield. Differences were found for phenotypic covariances and also for genetic, permanent environmental, and herd-time covariances. Heritabilities for test-day milk yield tended to be lower at the end and especially at the start of lactation; they also were higher (maximum of ~25%) for high-yield herds and lower (maximum of 15%) for low-yield herds. Permanent environmental variances were on average 10% lower in high-yield herds. Relative herd-time variances were ~10% at start of lactation and then began to decrease regardless of herd yield; high-yield herds increased in midlactation followed by another decrease, and medium-yield herds increased at the end of lactation. Regressors for random regression effects were transformed to adjust for heterogeneity of test-day yield covariances. Some animal reranking occurred because of this transformation of genetic and permanent environmental effects. When genetic correlations between environments were allowed to differ from 1, some additional animal reranking occurred. Correlations of variances of genetic and permanent-environmental regression solutions within herd, test-day, and milking frequency class with class mean milk yields were reduced with adjustment for heterogeneous covariance. The method suggests a number of innovative solutions to issues related to heterogeneous covariance structures, such as adjusted estimates in multibreed evaluation.

**(Key words:** heterogeneous covariance, covariance structure, test-day yield, random regression)

**Abbreviation key:** EM = expectation maximization, HC = heterogeneous covariance, RRM = random regression model.

### INTRODUCTION

Accounting for heterogeneity of covariance among test-day observations is an important component of test-day model development. For lactation models, the issue of heterogeneous variance has been addressed by numerous studies (e.g., Dong and Mao, 1990; Short et al., 1990), and most genetic evaluation systems account for heterogeneity of variance through data adjustment prior to analysis (e.g., Wiggans and VanRaden, 1991) or direct estimation during analysis (e.g., Meuwissen et al., 1996). Only a few systems correct for heterogeneous variance components. One example is in the US, where heritability is adjusted (Wiggans and VanRaden, 1991).

For test-day models, most studies have focused on heterogeneity of phenotypic (e.g., Ibáñez et al., 1996; Pool and Meuwissen, 2000) or residual (e.g., Ibáñez et al., 1999; Rekaya et al., 1999; Jaffrezic et al., 2000) variance. However, heterogeneity of covariance components, which is more difficult to estimate, has received limited attention despite its importance. The assumed covariance structures among test-day yields are used for estimation over the whole lactation or across lactations, even if information is available only for a few test days (e.g., Pool and Meuwissen, 1999).

One feature of random coefficient models, also known as random regression models (RRM), is that they allow for modeling of covariances through regressions. That feature has been used in studies on heat stress (Ravagnolo and Misztal, 2000) and on reaction norm models (Strandberg et al., 2000). With current RRM, covariances are modeled as functions of regression and elementary covariances among regressions.

Simple, robust estimation procedures for heterogeneous covariance (HC) matrices currently are not available. The first objective of this study, therefore, was to estimate HC components according to herd milk yield. The second objective was to show that HC across herd yield levels can be modeled by adjusting *a priori* regressions for random effects. The final objective was to extend the method to study and to model genetic correlations between herd yield levels that can be <1.

Received December 31, 2003.

Accepted May 5, 2005.

Corresponding author: N. Gengler; e-mail: gengler.n@fsagx.ac.be.

## MATERIALS AND METHODS

### Data

Test-day milk yields (222,679) of first-parity Holstein cows in New York, Wisconsin, and California herds from 1990 through 2000 were adjusted additively to a constant age and lactation stage using the adjustment factors of Bormann et al. (2002). Those factors had been obtained from a much larger data set, of which the data for this study were a subset. The comparability of results with those from other investigations of test-day evaluation methodology with US data (Bormann et al., 2002, 2003; Gengler and Wiggans, 2002; Wiggans et al., 2002) and the availability of estimates for effect of age and lactation stage based on a large population were considered to be of sufficient benefit to offset possible effects on variance reduction for random effects from data adjustment prior to analysis. Eventual shifts in the overall mean for the data were accommodated by adjusting a fixed effect so that the mean was kept in the model.

This approach also allowed the direct use of mean herd yield levels. Four data subsets of similar size (55,604 to 55,685 records) were defined by mean herd yield. Herds could change yield levels after 2 yr. Difference in mean test-day milk yield of first-parity cows in the highest (37.4 kg) and lowest (23.3 kg) subsets for herd yield was 14.1 kg. The complete data set also was grouped into three randomly selected subsets, which had similar size (72,582 to 76,641 records) and mean test-day milk yield (29.0 to 30.7 kg). The three random data subsets were used to compute genetic correlations across environments, which were then averaged over the three data sets.

### Covariance Structure

Consider the following RRM:

$$\mathbf{y} = \mathbf{X}\mathbf{t} + \sum_i \mathbf{Q}_i \mathbf{u}_i + \mathbf{e},$$

where  $\mathbf{y}$  = vector of test-day records,  $\mathbf{t}$  = vector of fixed effects,  $\mathbf{X}$  = incidence matrix linking  $\mathbf{y}$  and  $\mathbf{t}$ ,  $\mathbf{Q}_i$  = matrix of regressors,  $\mathbf{u}_i$  = vector of random effects  $i$ , and  $\mathbf{e}$  = vector of residuals. The test-day record is nested in a given lactation of a given animal. The covariances among observations for that lactation and animal are as follows:

$$\text{Var}(\mathbf{y}) = \sum_i \mathbf{Q}_i \text{Var}(\mathbf{u}_i) \mathbf{Q}_i' + \text{Var}(\mathbf{e}),$$

which can be rewritten as

$$\text{Var}(\mathbf{y}) = \sum_i \mathbf{Q}_i \mathbf{G}_i \mathbf{Q}_i' + \mathbf{R},$$

where  $\mathbf{G}_i$  = elementary covariance matrix for random effects and  $\mathbf{R} = \text{Var}(\mathbf{e})$ ;  $\mathbf{Q}_i \mathbf{G}_i \mathbf{Q}_i'$  creates the covariance components linked to random effect  $i$  in  $\text{Var}(\mathbf{y})$ . At this stage, the matrices of regressors can be used to generate HC structures by modeling the covariances as functions of regression variables:

$$\text{Var}(\mathbf{y}_j) = \sum_i \mathbf{Q}_{ij} \mathbf{G}_{ij} \mathbf{Q}_{ij}' + \mathbf{R}_j,$$

where  $\mathbf{G}_{ij}$  = covariance matrix of effect  $i$  in environment  $j$ .

At present, direct estimation of heterogeneous  $\mathbf{G}_{ij}$  in an RRM is too complex for available procedures. An indirect way to estimate heterogeneous  $\mathbf{G}_{ij}$  is to decompose the matrix into orthogonal components through a transformation matrix ( $\mathbf{T}$ ), which would render  $\mathbf{G}_{ij}$  independent of the heterogeneity strata ( $\mathbf{G}_{0i} = \mathbf{T}_{ij} \mathbf{G}_{ij} \mathbf{T}_{ij}'$ ) and result in

$$\text{Var}(\mathbf{y}_j) = \sum_i \mathbf{Q}_{ij} (\mathbf{T}_{ij})^{-1} \mathbf{G}_{0i} (\mathbf{T}_{ij}')^{-1} \mathbf{Q}_{ij}' + \mathbf{R}_j.$$

Conceptually, the simple transformation of regressors  $\mathbf{T}_{ij}^*$  "bends" the matrix of coefficients through  $\mathbf{Q}_{ij}^* = \mathbf{Q}_{ij} \mathbf{T}_{ij}^* = \mathbf{Q}_{ij} (\mathbf{T}_{ij})^{-1}$ . This approach allows replacement of  $\mathbf{G}_{ij}$ , which differs by environment  $j$  and effect  $i$ , with a single matrix  $\mathbf{G}_{0i}$  for every random effect  $i$ . Thus, HC structures can be modeled easily for both nongenetic and genetic random effects.

The initial underlying assumption is that genetic correlations between environments are unity for every transformed regression. Transformation of regressors was done independently for the different random effects. Possible dependencies among the variation of some of those random effects (e.g., genetic and permanent environmental) were not considered.

Although several possibilities exist for  $\mathbf{T}$ , an obvious candidate is the inverse of the lower Cholesky decomposition because  $\mathbf{G}_{0i}$  then becomes an identity matrix. The Cholesky matrix is also a matrix generalization of the square root of the covariances. The approach used was a simple generalization of the standardization of random effects approach used in France (Robert-Granié et al., 1999). The technique of rescaling random coefficients in mixed linear models so as to make them orthogonal via a Cholesky triangular transformation of the variance covariance matrix has been previously reported (e.g., Groeneveld, 1994). The advantage of doing this in a random regression or random coefficient models setting is that those models allow for the direct integra-

tion of the transformation. The order of random regressions can be chosen so that the first transformed regression is defined as the standardized constant term. Robert-Granié et al. (2002) extended this idea to heteroskedastic random regression models. For this study, heterogeneity in  $\mathbf{G}_{ij}$  was modeled by modeling  $\mathbf{T}_{ij}^*$ . However, instead of applying a generalized expectation-maximization (EM) algorithm (e.g., Foulley and Quaas, 1995),  $\mathbf{T}_{ij}^*$  was modeled *a posteriori* based on  $\mathbf{G}_{ij}$  matrices obtained from the different environments, where the distinction among environments was based on a continuous variate (e.g., yield level within heterogeneity strata). Integrated modeling similar to the methods proposed by Robert-Granié et al. (2002) is mathematically straightforward but was not used in the present study because of computing complexity.

**Estimation of covariance components.** Covariance components were estimated using a combination of EM and average-information REML. If positive definite values could not be obtained through average-information REML, estimates were obtained through a combination of EM and average-information REML (Druet et al., 2003).

**Modeling of covariance components based on herd yield.** Estimated covariance components ( $\mathbf{G}_{ij}$ ) were transformed into lower Cholesky triangular matrices  $\mathbf{L}_{ij}$ , where  $i$  = random effect and  $j$  = herd yield levels (environment). Every elementary element  $k$  of  $\mathbf{L}_{ijk}$  ( $l_{ijk}$ ) was then modeled as a constant, linear, and quadratic function of standardized milk yield  $s$  for class mean  $m$  based on herd, test-day, and milking frequency:

$$l_{ijk} = \alpha_{0ik} + \alpha_{1ik}s_j + \alpha_{2ik}s_j^2 + \varepsilon_{ijk},$$

where  $\alpha$  = regression coefficient;  $s = -1 + 2[(m - 23.3)/(36.8 - 23.3)]$  = standardized milk yield when 23.3 and 36.8 kg of milk were means for lowest and highest herd-time yield classes, respectively,  $-1 < s < 1$ , and  $m$  = mean herd milk yield for a 2-yr period; and  $\varepsilon$  = residual. In matrix algebra, for every effect  $i$ ,  $\mathbf{l}_i = \mathbf{S}\alpha_i + \varepsilon_i$ , or

$$\begin{bmatrix} \mathbf{l}_{i1} \\ \vdots \\ \mathbf{l}_{ik} \\ \vdots \\ \mathbf{l}_{in_k} \end{bmatrix} = [\mathbf{I}_{n_k} \otimes \mathbf{S}] \begin{bmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{ik} \\ \vdots \\ \alpha_{in_k} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{ik} \\ \vdots \\ \varepsilon_{in_k} \end{bmatrix},$$

where  $\otimes$  = Kronecker product and  $n_k$  = number of non-zero elements in  $\mathbf{L}$ .

Estimates of  $\alpha_{ik}$  ( $\hat{\alpha}_{ik}$ ) were obtained independently for every effect  $i$  and every coefficient  $k$  by solving  $\hat{\alpha}_{ik} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{l}_{ik}$ . The solutions allowed definition of the

transformation matrix as a function of standardized yield  $s$ . Observed covariances were regressed towards expected covariances based on herd yield. This regression towards expected variances is similar to the method described by Robert-Granié et al. (2002); however, their method was integrated, and the parameters of the dispersion models were estimated using generalized EM REML (e.g., Foulley and Quaas, 1995).

A second major difference from the method of Robert-Granié et al. (2002) was that the variances and covariances in this study were modeled with a global generalized square-root (Cholesky triangular) transformation of the entire matrix instead of a log transformation for variances and no transformation for correlations. Modeling under the Cholesky transformation guaranteed positive definiteness of the covariance matrices. The method of Robert-Granié et al. (2002) does not guarantee correlations in the parameter space (between  $-1$  and  $1$ ) but has the advantage of being an integrated approach. Further research should be able to merge the indirect method in this study with the direct method of Robert-Granié et al. (2002).

Heterogeneous error variances were modified in a similar fashion by replacing  $\mathbf{Q}$  with an identity matrix. Mixed-model equations were then adjusted by weighting according to the assumed inverse of the residual covariance of a given record.

## Applied Models

Three different models were applied to the various data sets to estimate covariance components and to calculate EBV. Table 1 summarizes application of the models to the data sets.

**Covariance estimation based on herd yield.** The four data subsets defined by mean herd yield were used to estimate four sets of covariance components with the RRM

$$\mathbf{y} = \mathbf{X}\mathbf{t} + \mathbf{Q}_h\mathbf{h} + \mathbf{Q}_a\mathbf{a} + \mathbf{Q}_p\mathbf{p} + \mathbf{e}, \quad [1]$$

where  $\mathbf{y}$  = vector of test-day records for milk yield;  $\mathbf{t}$  = vector of fixed class effects for herd, test day, and milking frequency;  $\mathbf{h}$  = vector of random effects for 2-yr time period within herd (herd-time effects);  $\mathbf{a}$  = vector of animal effects (breeding values);  $\mathbf{p}$  = vector of random permanent environmental effects;  $\mathbf{e}$  = residual effect;  $\mathbf{X}$  = incidence matrix that links  $\mathbf{y}$  and  $\mathbf{t}$ ;  $\mathbf{Q}_h$ ,  $\mathbf{Q}_a$ , and  $\mathbf{Q}_p$  = matrices of constant, linear, and quadratic modified Legendre polynomials (Gengler et al., 1999):  $r_0 = 1$ ,  $r_1 = 3^{0.5}x$ , and  $r_2 = (5/4)^{0.5}(3x^2 - 1)$ , where  $x = -1 + 2[(\text{DIM} - 1)/(365 - 1)]$ , that link  $\mathbf{y}$  and  $\mathbf{h}$ ,  $\mathbf{a}$ , and  $\mathbf{p}$ , respectively. A previous study (Gengler and Wiggans, 2001) of the same data had found that the portion of total variance

**Table 1.** Applied models, data sets, and analysis results.

Applied model	Model description	Analysis results		
		All data	Subsets based on mean herd yield (n = 4)	Random subsets (n = 3)
1	No heterogeneous covariance adjustment; genetic correlation across environments = 1	EBV	Covariance component estimates	—
2	Heterogeneous covariance adjustment; genetic correlation across environments = 1	EBV	—	Covariance component estimates <sup>1</sup>
3	Heterogeneous covariance adjustment; genetic correlation across environments ≠ 1	EBV	—	Covariance component estimates

<sup>1</sup>Computations used for likelihood ratio tests to compare Models [2] and [3].

explained by a herd-time effect was not negligible; therefore, **h** was included to allow herd-specific lactation curves. The covariance structure for Model [1] can be summarized as

$$\text{Var} \begin{bmatrix} \mathbf{h} \\ \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_h \otimes \mathbf{H}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \otimes \mathbf{G}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_p \otimes \mathbf{P}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_n \sigma_e^2 \end{bmatrix},$$

and

$$\text{Var}(\mathbf{y}) = \mathbf{Q}_h(\mathbf{I}_h \otimes \mathbf{H}_0)\mathbf{Q}'_h + \mathbf{Q}_a(\mathbf{A} \otimes \mathbf{G}_0)\mathbf{Q}'_a + \mathbf{Q}_p(\mathbf{I}_p \otimes \mathbf{P}_0)\mathbf{Q}'_p + \mathbf{I}_n \sigma_e^2,$$

where **I** = identity matrix; **H**<sub>0</sub>, **G**<sub>0</sub>, and **P**<sub>0</sub> = elementary covariance matrices among the three random regressions for herd-time, genetic, and permanent environmental effects, respectively; **A** = additive relationship matrix, h = number of herd-time effects, p = number of animals with records, and n = number of test-day records.

Second-order polynomials were used as a compromise between model complexity and desire to achieve a reasonably good fit. Preliminary research had shown that the constant, linear, and quadratic polynomials were highly related to the first, second, and third eigenvectors, which explained a large part of the variances for all three random effects.

**Computation of EBV with and without HC adjustment.** The complete data set was analyzed with and without HC adjustment. To provide EBV without HC adjustment, the regular mixed-model equations from Model [1] were solved using mean coefficients ( $\bar{l}_{ijk} = \alpha_{0ik}$ ). To provide EBV with HC adjustment, mixed-model equations with transformed regressors based on standardized milk yield *s* were solved:

$$\mathbf{y} = \mathbf{Xt} + \mathbf{Q}_{h(s)}^* \mathbf{h}^* + \mathbf{Q}_{a(s)}^* \mathbf{a}^* + \mathbf{Q}_{p(s)}^* \mathbf{p}^* + w_{(s)} \mathbf{e}^*, \quad [2]$$

where **Q**<sub>h(s)</sub><sup>\*</sup>, **Q**<sub>a(s)</sub><sup>\*</sup>, and **Q**<sub>p(s)</sub><sup>\*</sup> = matrices of transformed regressors dependent on standardized herd yields *s* and linking **y** with **h**<sup>\*</sup>, **a**<sup>\*</sup>, and **p**<sup>\*</sup> and *w*<sub>(s)</sub> = square root of the inverse of the weight dependent on *s*. The associated covariance structure was

$$\text{Var} \begin{bmatrix} \mathbf{h}^* \\ \mathbf{a}^* \\ \mathbf{p}^* \\ \mathbf{e}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I}_h \otimes \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \otimes \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_p \otimes \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W} \end{bmatrix},$$

where **W** = **I**<sub>n</sub>*w*<sub>(s)</sub><sup>2</sup>, a diagonal matrix with diagonal elements equal to the inverse of the weight associated with the record. Covariance of the observations based on *s* was

$$\text{Var}(\mathbf{y}_{(s)}) = \mathbf{Q}_{h(s)}^*(\mathbf{I}_h \otimes \mathbf{I}_3)\mathbf{Q}_{h(s)}^{\prime*} + \mathbf{Q}_{a(s)}^*(\mathbf{A} \otimes \mathbf{I}_3)\mathbf{Q}_{a(s)}^{\prime*} + \mathbf{Q}_{p(s)}^*(\mathbf{I}_p \otimes \mathbf{I}_3)\mathbf{Q}_{p(s)}^{\prime*} + \mathbf{I}_n w_{(s)}^2.$$

**Genetic correlation across environments ≠ 1.** Although Model [2] allows for differences in genetic covariance across herd yield levels, it does not allow genetic correlation across environments to differ from 1. Recently, several studies (e.g., Castillo-Juarez et al., 2002) used RRM as an approach to address this issue.

Model [2] could be generalized by including separate genetic effects for high and low yield. Every observation then potentially would be influenced by two sets of genetic effects. Genetic effects for every animal then could be defined continuously from high to low yield as **a**<sub>(s)</sub><sup>\*</sup> =  $\phi_{1(s)} \mathbf{a}_1^* + \phi_{2(s)} \mathbf{a}_2^*$ , where  $\phi_1$  and  $\phi_2$  are coefficients for environments defined as a function of *s* with  $\phi_{1(s)} = (1 + s)/2$  and  $\phi_{2(s)} = 1 - \phi_{1(s)} = (1 - s)/2$ . The coefficients  $\phi_{1(s)}$  and  $\phi_{2(s)}$  also would link observations with *s*. If an obser-

vation was at the maximal herd yield level ( $s = 1$ ), then  $\phi_{1(1)} = 1$  and  $\phi_{2(1)} = 0$ ; if an observation was at the minimal low herd yield level ( $s = -1$ ), then  $\phi_{1(-1)} = 0$  and  $\phi_{2(-1)} = 1$ .

Given those conventions, Model [2] easily was rewritten to allow differences in covariances across environments and also genetic correlations that differed from 1:

$$\mathbf{y}_{(s)} = \mathbf{Xt} + \mathbf{Q}_{h(s)}^* \mathbf{h} + \phi_{1(s)} \mathbf{Q}_{a(s)}^* \mathbf{a}_1^* + \phi_{2(s)} \mathbf{Q}_{a(s)}^* \mathbf{a}_2^* + \mathbf{Q}_{p(s)}^* \mathbf{p} + \mathbf{w}_{(s)} \mathbf{e}^* \quad [3]$$

Covariance matrices for Model [3] were as follows:

$$\text{Var} \begin{bmatrix} \mathbf{h}^* \\ \mathbf{a}_1^* \\ \mathbf{a}_2^* \\ \mathbf{p}^* \\ \mathbf{e}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I}_h \otimes \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \otimes \begin{bmatrix} \mathbf{I}_3 & \mathbf{D} \\ \mathbf{D} & \mathbf{I}_3 \end{bmatrix} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_p \otimes \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_n & \mathbf{0} \end{bmatrix},$$

and

$$\begin{aligned} \text{Var}(\mathbf{y}_{(s)}) &= \mathbf{Q}_{h(s)}^* (\mathbf{I}_h \otimes \mathbf{I}_3) \mathbf{Q}_{h(s)}^{\prime} \\ &+ \left[ \phi_{1(s)} \mathbf{Q}_{a(s)}^* \quad \phi_{2(s)} \mathbf{Q}_{a(s)}^* \right] \left( \mathbf{A} \otimes \begin{bmatrix} \mathbf{I}_3 & \mathbf{D} \\ \mathbf{D} & \mathbf{I}_3 \end{bmatrix} \right) \begin{bmatrix} \phi_{1(s)} \mathbf{Q}_{a(s)}^* \\ \phi_{2(s)} \mathbf{Q}_{a(s)}^* \end{bmatrix} \\ &+ \mathbf{Q}_{p(s)}^* (\mathbf{I}_p \otimes \mathbf{I}_3) \mathbf{Q}_{p(s)}^{\prime} + \mathbf{I}_n \mathbf{w}_{(s)}^2, \end{aligned}$$

where  $\mathbf{D} = \text{diag}[\phi_k]$  is a diagonal matrix of dimension 3 with the correlation between transformed regressors in the two environments. In Model [3], differences in covariances across environments were accounted for by the Cholesky transformation as in Model [2]; however, correlations across environments that differed from 1 were modeled based on separation into environmentally dependent genetic effects. Covariance of the total genetic effects could be written as

$$\begin{aligned} \text{Var}(\mathbf{a}_{p(s)}^*) &= [\phi_{1(s)} \quad \phi_{2(s)}] \left( \mathbf{A} \otimes \begin{bmatrix} \mathbf{I}_3 & \mathbf{D} \\ \mathbf{D} & \mathbf{I}_3 \end{bmatrix} \right) \begin{bmatrix} \phi_{1(s)} \\ \phi_{2(s)} \end{bmatrix} \\ &= \phi_{1(s)}^2 (\mathbf{A} \otimes \mathbf{I}_3) + (\phi_{2(s)}^2) (\mathbf{A} \otimes \mathbf{I}_3) + 2\phi_{1(s)}\phi_{2(s)} (\mathbf{A} \otimes \mathbf{D}). \end{aligned}$$

When the correlation between transformed regressors in the two environments tended to 1, covariance of the total genetic effects simplified to

$$\begin{aligned} \text{Var}(\mathbf{a}_{p(s)}^*) &= (\phi_{1(s)}^2 + \phi_{2(s)}^2 + 2\phi_{1(s)}\phi_{2(s)}) (\mathbf{A} \otimes \mathbf{I}_3) \\ &= (\phi_{1(s)} + \phi_{2(s)})^2 (\mathbf{A} \otimes \mathbf{I}_3) = \mathbf{A} \otimes \mathbf{I}_3 \end{aligned}$$

as in Model [2].

To determine if the introduction of a genetic correlation across environments that differed from 1 improved model fit, likelihood ratio tests were conducted with covariance components estimated from each of the three random data subsets using Models [2] and [3].

The estimated covariance components from Model [3] were applied to calculate EBV for the complete data set.

## Comparison of EBV

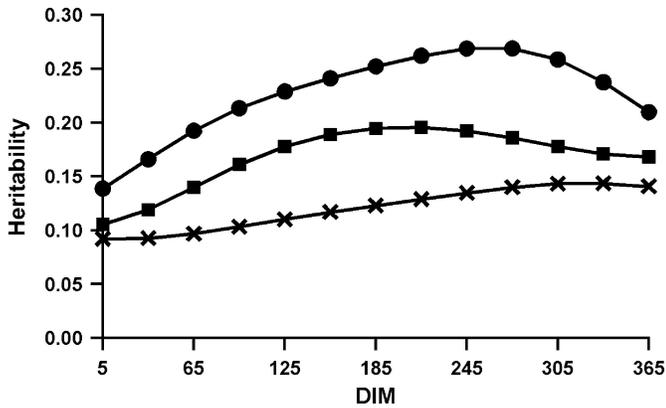
To demonstrate applicability of the methods and models, EBV were computed and expressed on a 305-d lactation basis; EBV from Models [2] and [3], which included transformation, were back-transformed to a mean scale. For cows, the same reverse transformation was done for the sum of EBV and permanent environmental effects. For genetic correlation  $\neq 1$ , EBV for every animal were defined continuously from high to low yield as  $\mathbf{a}_{(s)}^* = \phi_{1(s)} \mathbf{a}_1^* + \phi_{2(s)} \mathbf{a}_2^*$ , where  $\phi_{1(s)} + \phi_{2(s)} = 1$ , and reported for three environments: high herd yield ( $\phi_{1(1)} = 1$ ;  $\phi_{2(1)} = 0$ ), medium herd yield ( $\phi_{1(0)} = 0.5$ ;  $\phi_{2(0)} = 0.5$ ), and low herd yield ( $\phi_{1(-1)} = 0$ ;  $\phi_{2(-1)} = 1$ ). Rankings were created for cows and for sires with  $\geq 10$  daughters based on unadjusted EBV, HC-adjusted EBV with genetic correlation = 1, and HC-adjusted EBV with genetic correlation  $\neq 1$ .

One consequence of not applying adjustments for heterogeneity of covariance is that solutions in high-variance environments are more variable than in low-variance environments. To test if the proposed HC adjustment method corrects this problem, variances of regression solutions for genetic and permanent environmental effects were computed in every herd, test-day, and milking-frequency class and compared with mean milk yield for that class. If the HC adjustment was successful, correlation between those variances and class mean yield should be reduced.

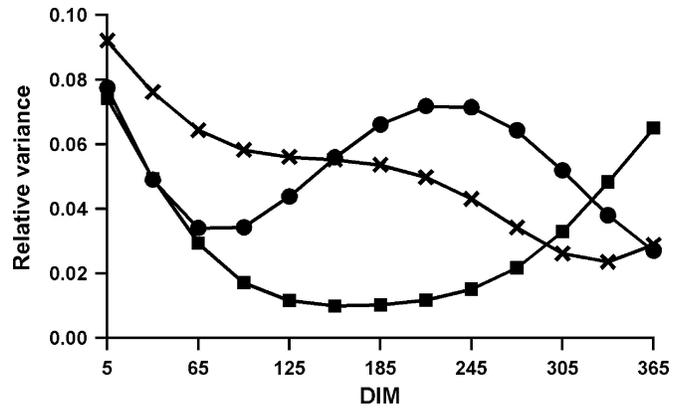
## RESULTS AND DISCUSSION

### Covariance Components Based on Herd Yield

Covariance components were estimated with Model [1] and then modeled and expressed as functions of  $s$ . For simplicity, only mean variances with  $s = 0$  (without HC adjustment) and extreme variances with  $s = -1$  or  $s = 1$  (with HC adjustment) are reported. Heritabilities for test-day milk yields (Figure 1) were substantially higher for high-yield than for low-yield herds and reached ~25% compared with ~15%, respectively. Medium-yield herds had intermediate heritability. However, the heritability trends were only somewhat similar to trends for permanent-environmental variance (Figure 2) as only high-yield herds differed substan-



**Figure 1.** Heritability of test-day milk yield by DIM for herds with low (×), medium (■), or high (●) yield.

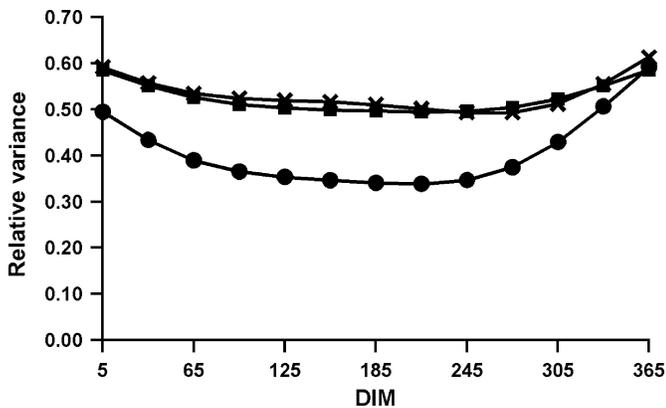


**Figure 3.** Relative variance of herd-time effect on test-day milk yield by DIM for herds with low (×), medium (■), or high (●) yield.

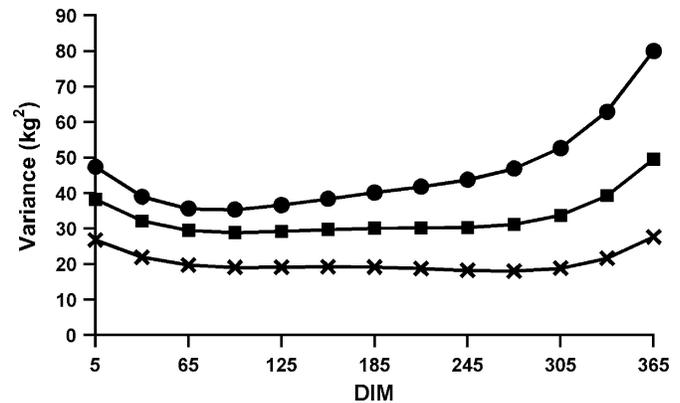
tially with lower relative permanent-environmental variance compared with herds with other yield levels. Combined variance for genetic and permanent environmental effects may be similar across herd yields, but a larger portion of that combined variance may be genetic for high-yield herds.

Relative herd-time variances (Figure 3) did not show similar patterns. Low-yield herds had higher herd-time variance at start of lactation, whereas variance for medium-yield herds was higher at start and end of lactation. For high-yield herds, variance was high at the start of lactation, decreased until about 65 DIM, and then increased until around 220 DIM to the same variance level as at the start of lactation, and again decreased through the end of lactation. No explanation was apparent for the differing relative variance patterns, and additional research is required to investigate possible negative effects.

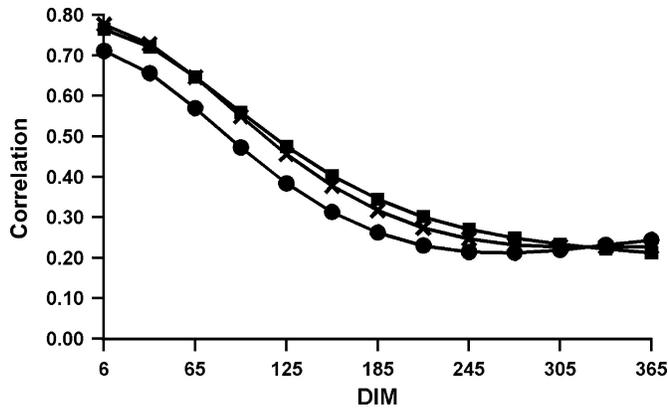
Relative variance patterns should be considered together with the pattern for phenotypic variance (Figure 4) over lactation. Plots for phenotypic variance were similar in shape but clearly not identical across herd yield levels. For low-yield herds, variances were nearly constant with rather limited increases at start and end of lactation. Compared with low-yield herds, phenotypic variances for medium-yield herds tended to be higher and increase more at the end of lactation. For high-yield herds, overall phenotypic variance and rate of increase in variance with DIM was substantially greater than for the other yield levels. The variance increase with herd yield level could result primarily from better management in high-yield herds, which allowed cows to express differences. The large heritability difference seems to confirm that animals in high-yield herds express relatively more genetic variance than do those in low-yield herds. The results of this study support that lactation stage and herd yield level should



**Figure 2.** Relative variance of permanent environmental effect on test-day milk yield by DIM for herds with low (×), medium (■), or high (●) yield.



**Figure 4.** Phenotypic variance of test-day milk yield by DIM for herds with low (×), medium (■), or high (●) yield.



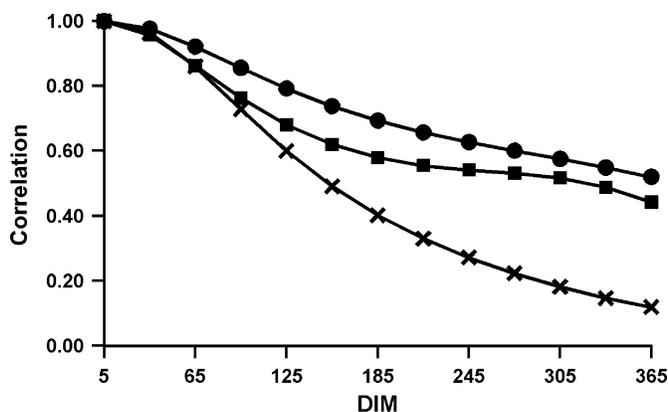
**Figure 5.** Phenotypic correlation of test-day milk yield at 5 DIM with test-day yield at other DIM for herds with low (x), medium (■), or high (●) yield.

be considered when developing adjustments for heterogeneity of phenotypic covariance.

Test-day milk yield at 5 DIM was compared with test-day yield at other DIM. Although phenotypic correlations (Figure 5) were remarkably stable, genetic correlations (Figure 6) decreased with herd yield level, especially for low-yield herds. Using inflated correlations could impact animal rankings, especially for dairy bulls with early evaluations based primarily on daughter records from early lactation in low-yield herds.

### Estimation of Genetic Correlations Across Environments

Likelihood ratio tests for the three random data subsets used to compare Models [2] and [3] showed that in all cases the introduction of additional parameters in



**Figure 6.** Genetic correlation of test-day milk yield at 5 DIM with test-day yield at other DIM for herds with low (x), medium (■), or high (●) yield.

the models significantly ( $P < 0.001$ ) improved the fit; likelihood ratios were 75.93, 84.24, and 65.15.

Means of estimated REML genetic correlations across environments from the three random data subsets were 0.972, 0.799, and 0.968 for the three Legendre coefficients. Standard deviations were 0.025, 0.211, and 0.041, respectively, which indicated a rather large degree of uncertainty in the estimation of the correlation for the second regression. Because of the variation in subset genetic correlations, no definitive conclusions can be made about genetic  $\times$  environmental interactions. Genetic differences across environments were reported by Veerkamp and Goddard (1998). In this study, the definition of environments and data sampling based solely on mean herd yield did not allow identification of the primary reason for genetic correlations of  $<1$ . A recent study by Raffrenato et al. (2003) suggests that regional differences can be a factor, and data for this study were pooled from three states with quite different environmental conditions.

### Comparison of Rankings With and Without HC Adjustment

Rank correlations of cow evaluations with and without HC adjustment were  $>0.99$  for EBV and  $>0.98$  for permanent environmental effect. However, some reranking did occur for the top 10 cows (Table 2) and for the top 10 bulls with  $\geq 10$  daughters with records (Table 3). The most reranking occurred for EBV plus permanent environmental effects (Table 2). Although EBV were quite stable (probably because families of animals seldom were concentrated in one environment and ties existed through the relationship matrix), HC adjustment resulted in some reranking of the top bulls based on evaluations without adjustment. Four (genetic correlation = 1) and 5 (genetic correlation  $\neq 1$ ) bulls of the original top 10 were eliminated.

For genetic correlation  $\neq 1$ , all animals had breeding values across all environments because of the continuous description of genetic effects as a function of standardized milk yield. As shown in Tables 2 and 3, animal rankings differed by mean herd yield. Evaluations of some of the top 10 bulls based on evaluations without HC adjustment were greatly affected by HC adjustment, with changes up to 121 kg between low- and high-yield environments. With additional research to verify the level of correlation across environments, this observed difference could lead to the use of the proposed HC adjustment method to create a ranking of bulls specific to a herd based on the yield level of that herd (Castillo-Juarez et al., 2002).

**Table 2.** Comparison of EBV, EBV plus permanent environmental (PE) effects, and rankings for evaluations with and without heterogeneous covariance (HC) adjustment and considering genetic correlation across environments and mean herd yield (low, medium, or high) for top 10 cows.

Evaluation without HC adjustment	Evaluation with HC adjustment			
	Genetic correlation = 1	Genetic correlation $\neq$ 1		
		High	Medium	Low
EBV (kg) and rank (in parentheses)	EBV (kg) and rank (in parentheses)			
1245 (1)	1398 (1)	1410 (1)	1414 (1)	1419 (1)
1197 (2)	1203 (3)	1172 (3)	1191 (3)	1210 (3)
1182 (3)	1242 (2)	1200 (2)	1213 (2)	1227 (2)
1154 (4)	1155 (4)	1144 (6)	1148 (4)	1153 (4)
1149 (5)	1148 (5)	1133 (7)	1137 (7)	1140 (5)
1143 (6)	1135 (7)	1166 (4)	1141 (5)	1115 (9)
1106 (7)	1081 (11)	1040 (19)	1085 (11)	1131 (6)
1101 (8)	1146 (6)	1150 (5)	1140 (6)	1129 (7)
1084 (9)	1071 (13)	1056 (15)	1060 (17)	1064 (16)
1079 (10)	1042 (17)	1022 (20)	1034 (20)	1046 (20)
EBV + PE (kg) and rank (in parentheses)	EBV + PE (kg) and rank (in parentheses)			
4606 (1)	4563 (6)	4554 (6)	4592 (6)	4631 (5)
4538 (2)	4238 (9)	4239 (9)	4276 (9)	4314 (8)
4501 (3)	4627 (5)	4615 (5)	4619 (5)	4623 (6)
4480 (4)	4720 (2)	4741 (2)	4751 (2)	4760 (2)
4411 (5)	5090 (1)	5087 (1)	5100 (1)	5113 (1)
4365 (6)	3621 (24)	3611 (24)	3601 (25)	3590 (28)
4328 (7)	4675 (3)	4680 (3)	4699 (3)	4718 (3)
4204 (8)	3991 (13)	4037 (11)	3942 (14)	3848 (18)
4195 (9)	4655 (4)	4647 (4)	4676 (4)	4706 (4)
4050 (10)	4483 (7)	4466 (7)	4476 (7)	4486 (7)

### Comparison of Class Variances With and Without HC Adjustment

Correlations of variances of random regression solutions for genetic and permanent environmental effects within herd, test-day, and milking-frequency class with class mean yields (Table 4) were smaller with HC adjustment than without it. The reduction in correlation was much smaller for genetic than for permanent environmental solutions (for which correlations became nearly 0). The anticipated reason for the difference in

the effect of HC adjustment for genetic and permanent environmental effects was the assumption of a perfect genetic correlation across environments. However, even with genetic correlation  $\neq$  1 (Table 4), a similar pattern was observed. If the effect of HC adjustment was small, only a few animal rankings would change as was observed in the example data sets (Tables 2 and 3). Correlations for genetic solution variances with class mean yield were reduced somewhat with HC adjustment and were smallest for low-yield herds.

**Table 3.** Comparison of EBV and rankings for evaluations with and without heterogeneous covariance (HC) adjustment and considering genetic correlation across environments and mean herd yield (low, medium, or high) of daughter for top 10 bulls with  $\geq$ 10 daughters with records.

EBV (kg) and rank (in parentheses) based on evaluation without HC adjustment	Daughters (no.)	EBV (kg) and rank (in parentheses) based on evaluation with HC adjustment			
		Genetic correlation = 1	Genetic correlation $\neq$ 1		
			High	Medium	Low
1099 (1)	54	1111 (1)	1168 (1)	1132 (1)	1097 (1)
984 (2)	67	921 (2)	961 (2)	929 (2)	896 (3)
926 (3)	159	900 (3)	920 (3)	907 (3)	893 (5)
898 (4)	10	851 (7)	832 (7)	844 (8)	855 (9)
869 (5)	21	776 (18)	769 (17)	783 (16)	796 (16)
867 (6)	141	869 (5)	862 (4)	857 (5)	851 (10)
861 (7)	10	842 (9)	839 (6)	838 (9)	836 (12)
856 (8)	222	803 (11)	750 (20)	790 (15)	829 (13)
829 (9)	12	825 (10)	773 (15)	832 (10)	892 (6)
823 (10)	16	756 (21)	771 (16)	776 (19)	782 (21)

**Table 4.** Correlations of variances of random regression solutions for genetic and permanent environmental effects within herd, test-day, and milking-frequency class with class mean yields with and without heterogeneous covariance (HC) adjustment and considering genetic correlation across environments and mean herd yield (low, medium, or high).

Model Effect	Legendre regression coefficient <sup>1</sup>	Correlation with class mean yield				
		No HC adjustment	Genetic correlation = 1	HC Adjustment		
				Genetic correlation ≠ 1	High	Medium
Genetic	r <sub>0</sub>	0.47	0.42	0.45	0.42	0.39
	r <sub>1</sub>	0.57	0.41	0.52	0.40	0.27
	r <sub>2</sub>	0.56	0.39	0.44	0.40	0.36
Permanent environmental	r <sub>0</sub>	0.48	-0.02	-0.02	-0.02	-0.02
	r <sub>1</sub>	0.56	0.12	0.13	0.13	0.13
	r <sub>2</sub>	0.54	0.10	0.10	0.10	0.10

$$^1r_0 = 1, r_1 = 3^{0.5}x, \text{ and } r_2 = (5/4)^{0.5}(3x^2 - 1), \text{ where } x = -1 + 2[(\text{DIM} - 1)/(365 - 1)].$$

## CONCLUSIONS

Currently, the methods used for HC adjustment in genetic evaluations with test-day models are often pre-adjustments (International Bull Evaluation Service, 2004). Some evaluation centers are testing or considering methods (e.g., Lidauer and Mäntysaari, 2001) based on the approach of Meuwissen et al. (1996), but no country is yet adjusting regressions. Although this study was not directly related to current HC adjustment methods, some of its results could influence the choice of future methods. Genetic and nongenetic covariance structures were found to be different according to herd milk yield. Differences were found not only for phenotypic covariances but also for heritability, permanent environmental, and herd-time variances. Current adjustment methods used by all major dairy countries except the US and The Netherlands (International Bull Evaluation Service, 2004) consider the variance ratios to be constant. High-yield herds had higher heritabilities for test-day milk yields and lower relative permanent environmental variances.

All currently used adjustment methods either correct data prior to analysis or have been integrated into the evaluation system and affect variances. This study showed that a method based on transformed regressors for random regression effects can be used to address the issue of heterogeneity of test-day yield covariances. As shown in the example data sets, some animal re-ranking occurred because of the effect of this transformation on both genetic and permanent environmental effects.

A challenge in the developed HC adjustment method is that nongenetic and genetic covariance matrices have to be estimated for different environments prior to calculation of genetic evaluations. Those additional calculations could require substantial computing resources and time, and the estimates could have large sampling

errors. However, as shown with Model [3], the method can be adapted to allow genetic correlations between environments to differ from 1, which produced animal re-ranking in the example data sets. Correlations of regression coefficient variances for genetic and permanent environmental effects within herd, test-day, and milking frequency class with class mean milk yield were reduced with HC adjustment.

The HC adjustment method that was developed suggests innovative solutions for a number of issues related to heterogeneity of covariances and their impact on genetic evaluation systems. First, the general concept can be used for data adjustment both prior to analysis (single transformation of regressors) and during analysis (transformation and update of transformation matrices). Because every regression of each test-day yield of a given cow can be adjusted, extreme flexibility can be achieved within the modeling process. For example, differences in covariance structures among breeds can be accommodated for multibreed evaluation. Crossbred animals can then be included by interpolation based on the proportion of genes from each breed of ancestors. This particular benefit could be especially important if breeds are to be evaluated together because of their simultaneous presence in contemporary groups or the presence of crossbreds in contemporary groups (e.g., Jerseys and Holsteins in the US and dual-purpose Belgian Blues and Holsteins in Belgium). The method developed also allows genetic correlations between environments to differ from 1 and has potential use if different bull rankings are needed according to source of covariance differences.

## ACKNOWLEDGMENTS

Nicolas Gengler, who is a research associate of the National Fund for Scientific Research (Brussels, Belgium), acknowledges the Fund's financial support and

the facilitation of computations through Grant 2.4507.02 F (2). Partial financial support for A. Gillon was provided by the Luxembourgian Herdbook Federation, a breeders' cooperative. The authors gratefully acknowledge computer programs provided by I. Misztal (University of Georgia, Athens) and T. Druet (Institut National de la Recherche Agronomique, Jouy-en-Josas, France) and manuscript review by L. L. M. Thornton and S. M. Hubbard (Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD).

## REFERENCES

- Bormann, J., G. R. Wiggans, T. Druet, and N. Gengler. 2002. Estimating effects of permanent environment, lactation stage, age, and pregnancy on test-day yield. *J. Dairy Sci.* 85(Jan.). Online. Available: <http://jds.fass.org/>.
- Bormann, J., G. R. Wiggans, T. Druet, and N. Gengler. 2003. Within-herd effects of age at test day and lactation stage on test-day yields. *J. Dairy Sci.* 86:3765–3774.
- Castillo-Juarez, H., P. A. Oltenacu, and E. G. Cienfuegos-Rivas. 2002. Genetic and phenotypic relationships among milk production and composition traits in primiparous Holstein cows in two different herd environments. *Livest. Prod. Sci.* 78:223–231.
- Dong, M. C., and I. L. Mao. 1990. Heterogeneity of (co)variance and heritability in different levels of intraherd milk production variance and of herd average. *J. Dairy Sci.* 73:843–851.
- Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86:2480–2490.
- Foulley, J. L., and R. L. Quaas. 1995. Heterogeneous variances in Gaussian linear mixed models. *Genet. Sel. Evol.* 27:211–228.
- Gengler, N., A. Tijani, G. R. Wiggans, and I. Misztal. 1999. Estimation of (co)variance function coefficients for test day yield with expectation-maximization restricted maximum likelihood algorithm. *J. Dairy Sci.* 82(June). Online. Available: <http://jds.fass.org/>.
- Gengler, N., and G. R. Wiggans. 2001. Variance of effects of lactation stage within herd by herd yield. *J. Dairy Sci.* 84(Suppl. 1):216. (Abstr.)
- Gengler, N., and G. R. Wiggans. 2002. Adjustment for heterogeneous genetic and non-genetic (co)variance structures in test-day models using a transformation on random regression effect regressors. *Interbull Bull.* 29:79–83.
- Groeneveld, E. 1994. A reparameterization to improve numerical optimization in multivariate REML (co)variance component estimation. *Genet. Sel. Evol.* 26:537–545.
- Ibáñez, M. A., M. J. Carabaño, and R. Alenda. 1999. Identification of sources of heterogeneous residual and genetic variances in milk yield data from the Spanish Holstein-Friesian population and impact on genetic evaluation. *Livest. Prod. Sci.* 59:33–49.
- Ibáñez, M. A., M. J. Carabaño, J. L. Foulley, and R. Alenda. 1996. Heterogeneity of herd-period phenotypic variances in the Spanish Holstein-Friesian cattle: Sources of heterogeneity and genetic evaluation. *Livest. Prod. Sci.* 45:137–147.
- International Bull Evaluation Service. 2004. Description of National Genetic Evaluation systems for dairy cattle traits as applied in different Interbull member countries. Available: [http://www-interbull.slu.se/national\\_ges\\_info2/framesida-ges.htm](http://www-interbull.slu.se/national_ges_info2/framesida-ges.htm). Accessed June 20, 2004.
- Jaffrézic, F., I. M. S. White, R. Thompson, and W. G. Hill. 2000. A link function approach to model heterogeneity of residual variances over time in lactation curve analyses. *J. Dairy Sci.* 83:1089–1093.
- Lidauer, M., and E. A. Mäntysaari. 2001. A multiplicative random regression model for test-day data with heterogeneous variances. *Interbull Bull.* 27:167–171.
- Meuwissen, T. H. E., G. de Jong, and B. Engel. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci.* 79:310–316.
- Pool, M. H., and T. H. E. Meuwissen. 1999. Prediction of daily milk yields from a limited number of test days using test day models. *J. Dairy Sci.* 82:1555–1564.
- Pool, M. H., and T. H. E. Meuwissen. 2000. Reduction of the number of parameters needed for a polynomial random regression test-day model. *Livest. Prod. Sci.* 64:133–145.
- Raffrenato, E., R. W. Blake, P. A. Oltenacu, J. Carvalheira, and G. Licitra. 2003. Genotype by environment interaction for yield and somatic cell score with alternative environmental definitions. *J. Dairy Sci.* 86:2470–2479.
- Ravagnolo, O., and I. Misztal. 2000. Genetic component of heat stress in dairy cattle, parameter estimation. *J. Dairy Sci.* 83:2126–2130.
- Rekaya, R., M. J. Carabaño, and M. A. Toro. 1999. Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. *Livest. Prod. Sci.* 57:203–217.
- Robert-Granié, C., B. Bonaïti, D. Boichard, and A. Barbat. 1999. Accounting for variance heterogeneity in French dairy cattle genetic evaluation. *Livest. Prod. Sci.* 60:343–357.
- Robert-Granié, C., B. Heude, and J. L. Foulley. 2002. Modelling the growth curve of Maine-Anjou beef cattle using heteroskedastic random coefficients models. *Genet. Sel. Evol.* 34:423–445.
- Short, T. H., R. W. Blake, R. L. Quaas, and L. D. Van Vleck. 1990. Heterogeneous within-herd variance. 1. Genetic parameters for first and second lactation milk yields of grade Holstein cows. *J. Dairy Sci.* 73:3312–3320.
- Strandberg, E., R. Kolmodin, P. Madsen, J. Jensen, and H. Jorjani. 2000. Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. *Interbull Bull.* 25:41–45.
- Veerkamp, R. F., and M. E. Goddard. 1998. Covariance functions across herd production levels for test day records on milk, fat, and protein yields. *J. Dairy Sci.* 81:1690–1701.
- Wiggans, G. R., and P. M. VanRaden. 1991. Method and effect of adjustments for heterogeneous variance. *J. Dairy Sci.* 74:4350–4357.
- Wiggans, G. R., P. M. VanRaden, J. Bormann, J. C. Philpot, T. Druet, and N. Gengler. 2002. Deriving lactation yields from test-day yields adjusted for lactation stage, age, pregnancy, and herd test date. *J. Dairy Sci.* 85(Jan.). Online. Available: <http://jds.fass.org/>.