

## Estimated Heterogeneity of Phenotypic Variance of Test-Day Yield with a Structural Variance Model

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 20705-2350

### ABSTRACT

First-lactation test-day milk, fat, and protein yields from New York, Wisconsin, and California herds from 1990 through 2000 were adjusted additively for age and lactation stage. A random regression model with third-order Legendre polynomials for permanent environmental and genetic effects was used. The model included a random effect with the same polynomial regressions for 2 yr of calvings within herd (herd-time effect) to provide herd-specific lactation curves that can change every 2 yr. (Co)variance components were estimated using expectation-maximization REML simultaneously with phenotypic variances that were modeled using a structural variance model. Maximum heritability for test-day milk yield was estimated to be ~20% around 200 to 250 d in milk; heritabilities were slightly lower for test-day fat and protein yields. Herd-time effects explained 12 to 20% of phenotypic variance and had the greatest impact at start of lactation. Variances of test-day yields increased with time, subclass size, and milking frequency. Test month had limited influence on variance. Variance increased for cows in herds with low and high milk yields and for early and late lactation stages. Repeatabilities of variances observed for a given class of herd, test-day, and milking frequency were 14 to 17% across nested variance subclasses based on lactation stage.

**(Key words:** heterogeneity of variance, (co)variance structure, test-day yield)

**Abbreviation key:** EM = expectation maximization.

### INTRODUCTION

Although a common assumption of genetic evaluation models is homogeneity of (co)variances, this assumption is often incorrect across time or herds (Boldman

and Freeman, 1990) and has the potential to bias rankings of animals. Heterogeneity of (co)variance can be addressed through adjustment of data either prior to analysis (Wiggans and VanRaden, 1991) or as a part of analysis (Meuwissen et al., 1996). In both cases, the adjustment is generally done from the observed variance toward a population variance.

Structural variance models, in which observed heterogeneity is modeled, are seldom used for yield traits but have been used successfully for type traits (Weigel and Lawlor, 1994). One important advantage of such models, especially for research purposes, is that they allow explanation of heterogeneity of observed variance, whereas current methods to account for heterogeneous variance of yield traits function much like a black box: observed heterogeneous variance is adjusted without explanation.

Research on 305-d lactation data (Brotherstone and Hill, 1986) showed that the most important factor that affected variance was yield level of the herd. A general positive correlation of yield level with estimated genetic and nongenetic variances was reported. With test-day models, another reason for unequal variance was lactation stage (Vargas et al., 1998). Dodenhoff and Swalve (1998) found that regional differences also affected yield variance. The method proposed by Meuwissen et al. (1996) to describe heterogeneity of variance accounts for covariances among observations while jointly modeling the mean and phenotypic variance of the input data.

The objective of this study was to document the existence and nature of unequal phenotypic variances by modeling heterogeneity of variance through a structural variance model jointly with (co)variance estimation for milk, fat, and protein yields.

### MATERIALS and METHODS

#### Data

First-lactation test-day yields from New York, Wisconsin, and California herds during 1990 through 2000 were adjusted additively for age and lactation stage using adjustment factors of Bormann et al. (2002). Data

Received August 20, 2003.

Accepted December 20, 2003.

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

**Table 1.** Characteristics of 3 random data sets used to model heterogeneity of variance for test-day yield.

Characteristic	Data set		
	1	2	3
Records, no.	73,456	72,582	76,641
Mean DIM	166.0	167.3	166.8
Mean test-day yield			
Milk, kg	30.2	29.0	30.8
Fat, g	1090	1064	1114
Protein, g	893	857	913

were adjusted prior to analysis so that results would be comparable with those from other studies that investigated methodology for calculating genetic evaluations of US data based on a test-day model (Bormann et al., 2002, 2003; Gengler and Wiggans, 2002; Wiggans et al., 2002). Because the adjustment of data prior to analysis was additive, variance estimates should not be affected. Herds were randomly assigned to three data sets (Table 1). The data sets were similar for size (72,582 to 76,641 records) and mean test-day milk yield (29.0 to 30.8 kg).

### (Co)variance Component Estimation Model

(Co)variance components were estimated with an accelerated expectation-maximization (EM) REML algorithm (Gengler et al., 1999) and the random regression model

$$\mathbf{y} = \mathbf{Xt} + \mathbf{Q}(\mathbf{Hh} + \mathbf{Z}^*\mathbf{a} + \mathbf{Zp}) + \mathbf{e},$$

which can be rewritten as

$$\mathbf{y} = \mathbf{Xt} + \mathbf{Qr} + \mathbf{e}$$

by setting

$$\mathbf{r} = \mathbf{Hh} + \mathbf{Z}^*\mathbf{a} + \mathbf{Zp}$$

and where  $\mathbf{y}$  = vector of test-day records for milk, fat, or protein yield;  $\mathbf{t}$  = vector of fixed-class effects for herd, test day, and milking frequency;  $\mathbf{h}$  = vector of random effects for 2-yr period within herd;  $\mathbf{p}$  = vector of random permanent environmental effects;  $\mathbf{a}$  = vector of animal effects (breeding values);  $\mathbf{e}$  = residual effect;  $\mathbf{X}$  = incidence matrix that links  $\mathbf{y}$  and  $\mathbf{t}$ ;  $\mathbf{r}$  = vector of regressions;  $\mathbf{Q}$  = matrix of constant, linear, and quadratic modified Legendre polynomials (Gengler et al., 1999):  $\mathbf{I}_0 = \mathbf{1}$ ,  $\mathbf{I}_1 = 3^{0.5}\mathbf{x}$ , and  $\mathbf{I}_2 = (5/4)^{0.5}(3\mathbf{x}^2 - 1)$ , where  $\mathbf{x} = -1 + 2[(\text{DIM} - 1)/(365 - 1)]$ , that link  $\mathbf{y}$  and  $\mathbf{r}$ ; and  $\mathbf{H}$ ,  $\mathbf{Z}$ , and  $\mathbf{Z}^*$  are incidence matrices that link  $\mathbf{y}$  with  $\mathbf{h}$ ,  $\mathbf{p}$ , and  $\mathbf{a}$ , respectively. A herd-time effect ( $\mathbf{h}$ ) was introduced

because a previous study (Gengler and Wiggans, 2001) of the same data found that the portion of total variance explained by such an effect was not negligible. Inclusion of  $\mathbf{h}$  provides herd-specific lactation curves that can change every 2 yr.

### Integrated Heterogeneous Variance Adjustment

Meuwissen et al. (1996) developed a method to allow joint estimation of breeding values and heterogeneous variances. The Meuwissen method is basically a multiplicative mixed model that scales milk production records toward a common phenotypic variance through computation of a heterogeneity parameter at each iteration. Adjustment factors are then obtained by modeling those heterogeneity parameters and extracting an expected variance estimate. The Meuwissen method is appealing because it accounts for (co)variances among observations and heterogeneity can be modeled in a flexible manner, e.g., through a structural variance model. Pool and Meuwissen (2000) applied a slightly modified version of the Meuwissen method in the estimation of (co)variance components to correct for unequal variances caused by lactation stage. However, their modification could be improved by using a structural variance model within the original Meuwissen method to describe heterogeneity.

Unfortunately, the Meuwissen method scales all fixed effects, including those that are not nested within a correction subclass, which can be problematic. Based on a suggestion by Pool and Meuwissen (2000), the Meuwissen method was modified in this study by correcting for fixed effects prior to every EM round. The general model solved in EM round  $n + 1$  can then be written as

$$\mathbf{y}_c^{n+1} = \mathbf{Xt}^{n+1} + \mathbf{Qr}_c^{n+1} + \mathbf{e}_c^{n+1},$$

with the corrected data vector ( $\mathbf{y}_c$ ) obtained from

$$\mathbf{y}_c^{n+1} = \mathbf{X}\hat{\mathbf{t}}^n + (\mathbf{I}^n)^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{t}}^n),$$

where  $\mathbf{I}$  = diagonal matrix of scaling coefficients  $\exp(\gamma_k/2)$  for every subclass  $k$  that is considered homogeneous for variances obtained from solutions of the model used to describe heterogeneous variances and  $\hat{\mathbf{t}}$  = estimated  $\mathbf{t}$ . As  $\mathbf{X}\hat{\mathbf{t}}^{n+1}$  approaches  $\mathbf{X}\hat{\mathbf{t}}^n$  and  $\mathbf{I}^{n+1}$  approaches  $\mathbf{I}^n$ , the model can be rewritten as

$$\mathbf{y}^{n+1} = \mathbf{Xt}^{n+1} + \mathbf{I}^{n+1}(\mathbf{Qr}_c^{n+1} + \mathbf{e}_c^{n+1}),$$

which is a random regression model with scaled random effects and a modified version of the Meuwissen model.

Because the proposed strategy is unable to model (co)variances among observations directly, phenotypic covariance between 2 observations can only change through changes in the associated phenotypic variances. Therefore, correlations between observations were assumed to be unaffected as a result of adjustment through the heterogeneous variance model.

### Heterogeneity Parameter

A feature of the Meuwissen method is that the modeling of the heterogeneity parameter uses a weighted mixed model and pseudovariates that were obtained by summing current  $\gamma_k$  with the remaining heterogeneity within variance subclass. Based on Meuwissen et al. (1996), a heterogeneity parameter ( $z$ ) for variance subclass  $k$  was developed:

$$z_k = (\{[\mathbf{y}_{c_k} - \mathbf{X}_k \hat{\mathbf{t}}_k' \hat{\mathbf{e}}_{c_k}] / \sigma_e^2\} - n_k) / 2$$

where  $\hat{\mathbf{e}}$  = estimated  $\mathbf{e}$ ,  $\sigma_e^2$  = residual variance, and  $n_k$  = number of records in variance subclass  $k$ . This formula is conceptually similar to a quadratic form but for a log-normal distribution. The variance associated with this heterogeneity parameter was estimated as

$$\text{Var}(z_k) = ([(\mathbf{Q}_k \hat{\mathbf{r}}_{c_k})' \mathbf{Q}_k \hat{\mathbf{r}}_{c_k} / \sigma_e^2] + 2n_k) / 4.$$

Variance subclasses were defined optimally as the portion of the records in a given herd, test day, and milking frequency class that were as homogeneous as possible for lactation stage but still with a specified number of records. Therefore, herd, test-day, and milking frequency classes were subdivided according to DIM (6 to 65, 66 to 125, ..., 306 to 365) into 60-d heterogeneity subclasses.

Although equal weights were assumed for every test-day record, the formulas can be easily modified. The weighted mixed model on pseudovariates

$$\text{diag}(\gamma_k) + \mathbf{W}^{-1}\mathbf{z}$$

was written as

$$(\mathbf{S}'\mathbf{W}\mathbf{S} + \mathbf{A}^{-1})\boldsymbol{\beta} = \mathbf{S}'\mathbf{W}[\text{diag}(\gamma_k) + \mathbf{W}^{-1}\mathbf{z}]$$

where  $\boldsymbol{\beta}$  = solutions,  $\mathbf{S}$  = design matrix linking pseudovariates and  $\boldsymbol{\beta}$ , and  $\mathbf{W}$  = diagonal matrix of iterative weights with  $\mathbf{W} = \text{diag}[\text{Var}(z_k)]$  and  $\text{Var}(\boldsymbol{\beta}) = \mathbf{A}$ . In contrast to Meuwissen et al. (1996),  $\gamma_k$  were scaled toward a common base:

$$\gamma_k = \mathbf{S}_k \boldsymbol{\beta} - \gamma^{\text{base}}$$

to correct variances toward this base.

### Structural Variance Model

Meuwissen et al. (1996) used a multiplicative mixed model with a fixed mean effect and an autoregressive (co)variance structure among random effects. To allow examination of the factors that influence heterogeneity, a greater number of fixed effects were used in the variance model of heterogeneity subclasses: mean test-day milk yield (26 groups: <13, 14 to 15, 16 to 17, 18, 19, ..., 39, >39 kg); year-season of test [6-mo periods (January through June, July through December) during 1990 through 2000]; test month (12 mo); milking frequency (2 or  $\geq 3$  times per day); size (11 groups: 1, 2, ..., 6, 7 to 8, 9 to 11, 12 to 15, 16 to 20, >20 animals); and mean DIM (36 groups of 10 d starting at 6 DIM). Separate year-season and month effects rather than a year-month effect were included to allow modeling of separate time and month trends.

In addition to those fixed effects, a random effect for herd, test day, and milking frequency was fitted in the structural variance model but without the autoregressive structure of Meuwissen et al. (1996). Their autoregressive structure is an indirect approximation to model similarities across time, and those similarities were modeled through fixed effects for this study. In contrast to the model in this study for estimating (co)variance components, which included both fixed and random herd-time effects, the structural variance model included only a random herd-time effect so that the observed variance could be regressed toward the estimated variance. The structural variance model also allowed heterogeneity of variances to be explained as well as to be modeled. In addition, different fixed and random effects easily can be fitted to this model to accommodate future research findings (e.g., modification to account for regional differences).

Estimation of repeatabilities of random effects is required to solve the variance mixed model. Instead of the REML procedure used by Meuwissen et al. (1996), Method R (Reverter et al., 1994) and the accelerator described by Druet et al. (2001) were used to estimate repeatabilities, because Method R could be integrated easily into a procedure for model solution and (co)variance component estimation.

### Solution and Estimation

Solutions of the heterogeneous variance model and estimates of (co)variance components were obtained iteratively. After one round of EM REML,  $\text{diag}(\gamma_k) +$

$W^{-1}z$  was computed using the current solution. Then,  $\beta$  was computed from 2 Method R subsamples and the complete data sets, and  $A$  was updated until convergence of  $A$  was reached based on Method R regressions. Final  $\beta$  was computed from the complete data set; then,  $\Gamma$  was computed, and data in  $y$  were adjusted to  $y_c$ . Those steps were repeated from the initial round of EM REML until convergence of (co)variance components was attained.

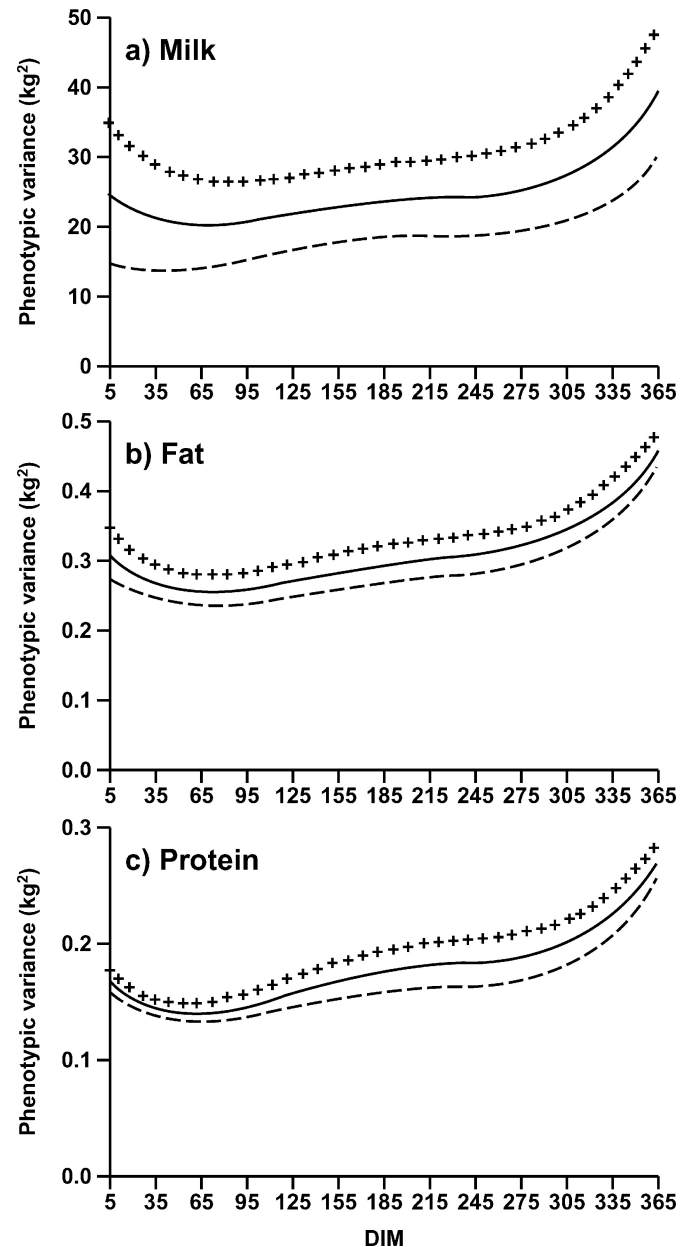
## RESULTS AND DISCUSSION

Heterogeneity of phenotypic variance across DIM was found for milk, fat, and protein test-day yields during first lactation (Figure 1). Numerous previous studies (e.g., Vargas et al., 1998) have shown such heterogeneity over lactation stages. For all 3 yield traits, phenotypic variance was highest at start and end of lactation. Although records from cows with >335 DIM or from cows culled early in lactation could have influenced results, similar patterns were found in studies that used random regression models and Legendre polynomials (Pool et al., 2000; Mayeres, 2002), which did not use test-day yields recorded after 335 DIM.

Relative variances of model effects during first lactation were compared with total variance for milk, fat, and protein test-day yields (Figure 2) to assess which effects had the greatest impact on observed variation in phenotypic variance. For milk, fat, and protein, observed patterns were similar for all effects with some distribution differences for nongenetic variance.

Heritability estimates (Figure 2) were highest (12 to 22%) for test-day milk yields and lowest (8 to 17%) for test-day protein yields. Heritabilities for test-day milk yield generally were lower than those from a previous study with the same data grouped by annual milk yield of herd (Gengler and Wiggans, 2001). Data were not grouped by herd yield in the present study because the objective was to model that effect directly in the variance function. Maximum heritability for test-day milk yield was estimated to be ~20% around 200 to 250 DIM, which is low compared with estimates in the literature (Pool et al., 2000). Some previous studies (Lidauer and Mäntysaari, 1999; Gengler and Wiggans, 2001) showed that the introduction of random herd effects, such as done in this study, results in lower heritability estimates, primarily because of a reduction in unlikely genetic variance at start and end of lactation. Lower heritabilities may also be linked to data adjustment prior to analysis, which might have reduced additive variance.

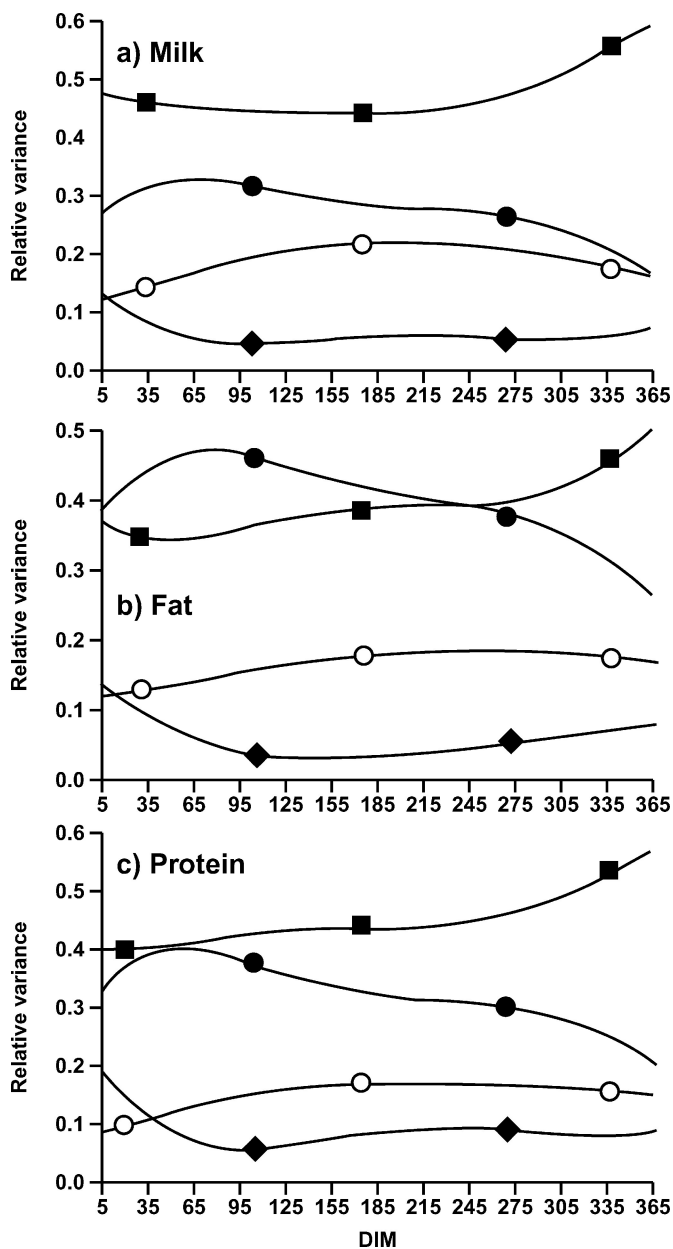
Herd-time effects explained some phenotypic variance (12 to 20%), especially at the start of lactation. Residual variances were kept constant over the entire



**Figure 1.** Mean phenotypic variance (—) plus (+ +) and minus (---) one standard deviation for test-day yield during first lactation by DIM for a) milk, b) fat, and c) protein.

lactation rather than allowed to be proportional to the mean. Therefore, changes in residual relative variance reflect changes in total variance. The sum of all nongenetic random effects represented by  $1 - \text{heritability}$  was relatively stable with only limited increases of its relative importance at start and end of lactation. All partitioning of environmental (co)variance components is basically arbitrary, because equivalent models can be developed that redirect some of those (co)variances





**Figure 2.** Heritability (○) and relative variances for herd-time (◆), permanent environmental (■), and residual (●) effects for test-day yield during first lactation by DIM for a) milk, b) fat, and c) protein.

toward a more complicated residual (co)variance structure.

### Genetic and Phenotypic Correlations

Mean genetic (Table 2, above diagonal) and phenotypic (Table 3, above diagonal) correlations between test days at various DIM were calculated based on estimated (co)variances for milk, fat, and protein test-day

yields from the 3 data sets. For all 3 yield traits, genetic effects for test-day yields at midlactation (165 DIM) had correlations of  $>0.50$  with test-day yields at start and end of 365-d lactations. Test days at lactation start and end were less correlated with each other; however, their genetic correlations were  $>0.35$ , which seemed reasonable compared with results from the literature for multitrait models (Vargas et al., 1998). Few studies included data through 365 DIM; therefore, comparison should be based on correlations for 305 DIM. Correlations from this study were clearly higher than some results obtained with random regression models (Gengler et al., 1999) but were similar to estimates obtained with multitrait models (Vargas et al., 1998).

Phenotypic correlations (Table 3, above diagonal) showed a pattern similar to that for genetic correlations for all 3 yield traits, with correlations of  $\geq 0.80$  for test-day yield at 165 DIM with other midlactation test days (65 to 285 DIM). Between test-day yields at start and end of lactation, correlations were  $>0.22$ . Correlations between test days at start of lactation and test days closer to 305 DIM were  $>0.25$ , which were still lower than correlations from multitrait models (Vargas et al., 1998).

A fundamental difference between random regression and multitrait models is that the latter concentrates nongenetic (co)variance in an unstructured residual (co)variance matrix. As shown by several researchers (Tijani et al., 1999), progressing from multitrait residuals to a nongenetic (co)variance structure in random regression models is not straightforward. The inclusion of additional nongenetic random effects in random regression models as pioneered by Lidauer and Mäntysaari (1999), proposed by Gengler and Wiggans (2001), and done in this study provides a more complete description of this nongenetic structure.

Variability of estimated parameters was indicated by observed standard deviations of the genetic correlations for the 3 samples (Tables 2 and 3, below diagonal).

### Structural Variance Model Solutions

Solutions for fixed effects of the variance model were expressed as the logarithm of variance, which can be transformed into multiplicative scaling effects that are computed as  $\exp(-\text{solution}/2)$ . Log variances for mean test-day yield were  $0.25 \pm 0.05$  for milk,  $0.26 \pm 0.02$  for fat, and  $0.19 \pm 0.04$  for protein. Variance increased for all 3 yield traits over time (Figure 3) except from 1992 through 1994. Because herd yield and size are not expected to be confounded and because they were modeled as other effects in the variance model, this trend should not be an artifact of increased yield or herd size over time.

**Table 2.** Genetic correlations between test-day yields at various DIM (above diagonal) and standard deviations of those correlations (below diagonal) for milk, fat, and protein.

Yield trait	DIM	DIM											
		5	25	45	65	85	125	165	205	245	285	325	365
Milk	5		0.98	0.92	0.85	0.79	0.68	0.61	0.57	0.54	0.52	0.48	0.41
	25	0.008		0.98	0.94	0.90	0.82	0.76	0.72	0.68	0.64	0.56	0.45
	45	0.026	0.006		0.99	0.96	0.91	0.86	0.82	0.78	0.72	0.62	0.46
	65	0.048	0.020	0.004		0.99	0.96	0.92	0.89	0.85	0.78	0.66	0.47
	85	0.068	0.036	0.013	0.002		0.99	0.96	0.93	0.89	0.82	0.68	0.48
	125	0.093	0.058	0.029	0.012	0.004		0.99	0.97	0.94	0.86	0.72	0.51
	165	0.093	0.062	0.034	0.016	0.006	0.001		0.99	0.97	0.90	0.77	0.56
	205	0.078	0.051	0.026	0.012	0.006	0.003	0.001		0.99	0.94	0.83	0.63
	245	0.046	0.023	0.008	0.011	0.014	0.010	0.005	0.002		0.98	0.90	0.73
	285	0.029	0.024	0.032	0.035	0.033	0.024	0.016	0.009	0.003		0.97	0.85
	325	0.090	0.084	0.079	0.070	0.061	0.045	0.035	0.026	0.016	0.005		0.96
	365	0.160	0.146	0.127	0.108	0.092	0.070	0.057	0.047	0.034	0.018	0.004	
	Fat	5		0.98	0.93	0.86	0.79	0.66	0.56	0.51	0.49	0.48	0.48
25		0.006		0.98	0.94	0.89	0.79	0.71	0.65	0.62	0.60	0.57	0.51
45		0.016	0.002		0.99	0.96	0.88	0.82	0.77	0.73	0.70	0.64	0.54
65		0.019	0.003	0.001		0.99	0.94	0.89	0.85	0.82	0.77	0.69	0.56
85		0.015	0.006	0.007	0.002		0.98	0.95	0.91	0.87	0.82	0.72	0.58
125		0.018	0.032	0.030	0.019	0.008		0.99	0.97	0.94	0.88	0.77	0.60
165		0.039	0.054	0.051	0.037	0.021	0.003		0.99	0.97	0.92	0.81	0.64
205		0.045	0.061	0.058	0.044	0.027	0.007	0.001		0.99	0.95	0.86	0.70
245		0.035	0.051	0.049	0.037	0.023	0.006	0.002	0.001		0.98	0.92	0.78
285		0.021	0.027	0.026	0.018	0.012	0.011	0.010	0.006	0.003		0.97	0.88
325		0.044	0.035	0.035	0.038	0.041	0.040	0.034	0.026	0.017	0.007		0.97
365		0.091	0.094	0.099	0.103	0.103	0.093	0.080	0.068	0.054	0.034	0.011	
Protein		5		0.97	0.91	0.83	0.76	0.64	0.57	0.53	0.50	0.47	0.43
	25	0.005		0.98	0.94	0.89	0.80	0.74	0.69	0.66	0.60	0.52	0.40
	45	0.019	0.005		0.99	0.96	0.90	0.85	0.81	0.77	0.70	0.58	0.41
	65	0.036	0.016	0.003		0.99	0.96	0.92	0.89	0.84	0.76	0.62	0.42
	85	0.052	0.028	0.010	0.002		0.99	0.96	0.93	0.88	0.80	0.65	0.43
	125	0.069	0.043	0.020	0.007	0.002		0.99	0.97	0.93	0.85	0.69	0.46
	165	0.069	0.044	0.022	0.008	0.002	0.001		0.99	0.96	0.89	0.74	0.52
	205	0.055	0.034	0.015	0.005	0.004	0.003	0.001		0.99	0.94	0.81	0.60
	245	0.029	0.016	0.011	0.013	0.013	0.009	0.005	0.002		0.98	0.89	0.72
	285	0.023	0.026	0.029	0.028	0.025	0.020	0.016	0.010	0.004		0.96	0.84
	325	0.066	0.058	0.049	0.040	0.034	0.031	0.031	0.025	0.015	0.004		0.96
	365	0.110	0.090	0.067	0.046	0.033	0.030	0.034	0.032	0.023	0.012	0.004	

Month of test had limited influence (Figure 4). Differences that might have originated from seasonal yield changes appeared to have been corrected by simultaneous fitting of other effects. Variances increased with subclass size as expected (Figure 5).

For all 3 traits, variance increased as mean milk yield of the heterogeneity subclass increased from 23 kg (Figure 6), which was expected because the literature reports a moderate correlation of variance to herd mean (Meinert et al., 1988). An even greater increase in variance was observed as mean subclass yield fell below 23 kg. However, this increase was unexpected based on the literature. The increased variability for high- and low-yield herds probably is not an artifact from the use of polynomials because an effect for DIM was included in the structural variance model.

Variances as a function of lactation stage (Figure 7) tended to be higher during late and especially during early lactation. Some of that pattern may be caused by

the use of polynomials even if the method used should normally describe only variance that is not explained by the random regression model.

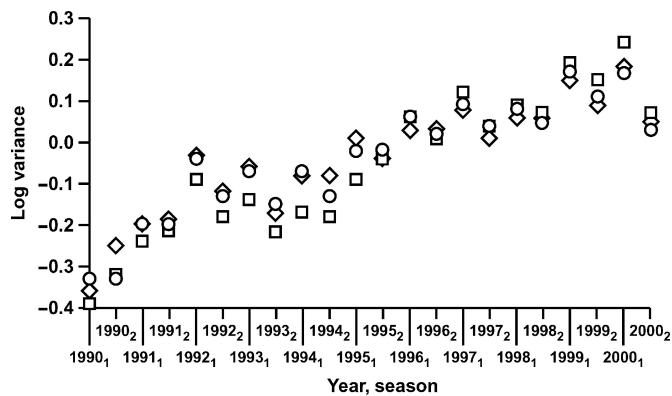
Variances increased with milking frequency (Table 4). No confounding between milking frequency and production level should occur because those effects are estimated jointly. Repeatabilities of variances for a given herd, test day, and milking frequency were  $0.14 \pm 0.02$  for milk,  $0.17 \pm 0.02$  for fat, and  $0.14 \pm 0.01$  for protein, across nested variance subclasses based on lactation stage.

### CONCLUSIONS

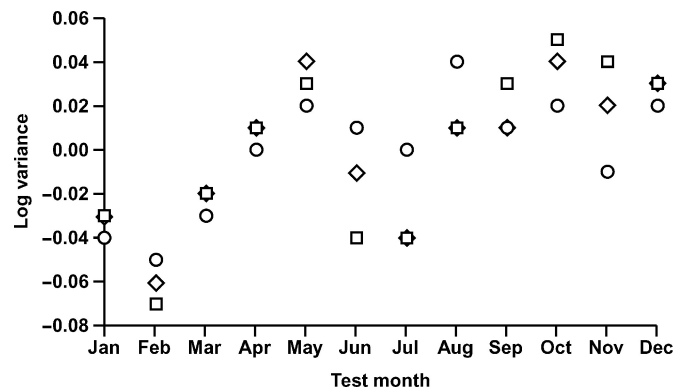
A method to model and to estimate (co)variances and variance heterogeneity jointly in test-day models was developed. Variances increased with time, subclass size, and milking frequency. Test month had extremely limited influence. As expected, test-day yield for herds

**Table 3.** Phenotypic correlations between test-day yields at various DIM (above diagonal) and standard deviations of those correlations (below diagonal) for milk, fat, and protein.

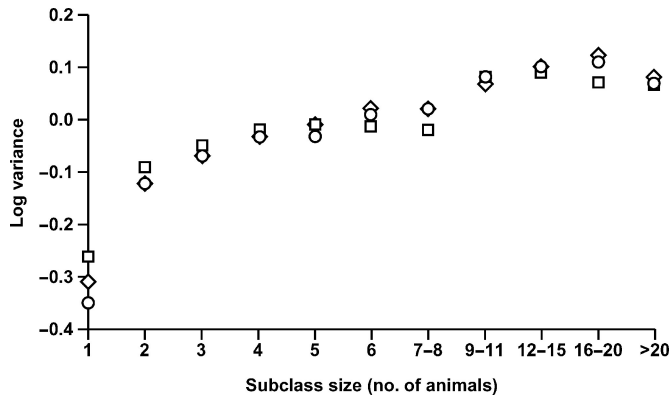
Yield trait	DIM	DIM											
		5	25	45	65	85	125	165	205	245	285	325	365
Milk	5		0.98	0.93	0.86	0.78	0.63	0.51	0.44	0.39	0.35	0.31	0.28
	25	0.005		0.98	0.94	0.88	0.76	0.65	0.57	0.51	0.44	0.35	0.26
	45	0.018	0.004		0.99	0.95	0.86	0.77	0.70	0.62	0.51	0.38	0.25
	65	0.030	0.011	0.002		0.99	0.93	0.86	0.79	0.70	0.58	0.41	0.24
	85	0.038	0.016	0.006	0.001		0.97	0.92	0.86	0.77	0.63	0.44	0.25
	125	0.041	0.025	0.018	0.012	0.006		0.98	0.94	0.86	0.72	0.51	0.28
	165	0.042	0.035	0.033	0.026	0.017	0.004		0.98	0.93	0.80	0.59	0.36
	205	0.039	0.039	0.040	0.035	0.026	0.010	0.002		0.98	0.88	0.70	0.48
	245	0.028	0.031	0.036	0.035	0.029	0.015	0.006	0.001		0.96	0.84	0.65
	285	0.013	0.009	0.018	0.020	0.018	0.011	0.004	0.002	0.002		0.95	0.83
	325	0.032	0.025	0.020	0.017	0.016	0.015	0.018	0.022	0.019	0.008		0.96
	365	0.061	0.062	0.062	0.060	0.058	0.053	0.053	0.055	0.050	0.032	0.009	
	Fat	5		0.98	0.92	0.83	0.73	0.55	0.42	0.36	0.33	0.33	0.34
25		0.001		0.98	0.92	0.85	0.70	0.59	0.52	0.47	0.44	0.39	0.33
45		0.005	0.002		0.98	0.94	0.83	0.73	0.66	0.60	0.53	0.43	0.32
65		0.011	0.007	0.002		0.99	0.92	0.84	0.78	0.71	0.61	0.47	0.31
85		0.022	0.017	0.009	0.002		0.97	0.92	0.86	0.79	0.67	0.50	0.30
125		0.046	0.040	0.028	0.015	0.006		0.99	0.95	0.88	0.75	0.55	0.31
165		0.059	0.055	0.042	0.026	0.013	0.002		0.99	0.94	0.82	0.62	0.37
205		0.061	0.057	0.044	0.029	0.016	0.003	0.001		0.98	0.89	0.71	0.48
245		0.051	0.047	0.035	0.022	0.011	0.005	0.004	0.001		0.96	0.84	0.64
285		0.029	0.025	0.017	0.010	0.009	0.010	0.007	0.004	0.002		0.95	0.82
325		0.014	0.013	0.014	0.016	0.017	0.016	0.016	0.017	0.014	0.005		0.96
365		0.036	0.035	0.033	0.030	0.027	0.026	0.031	0.035	0.033	0.021	0.006	
Protein		5		0.98	0.91	0.82	0.73	0.55	0.43	0.36	0.31	0.28	0.25
	25	0.004		0.98	0.92	0.85	0.71	0.60	0.52	0.46	0.39	0.31	0.23
	45	0.015	0.004		0.98	0.94	0.84	0.75	0.67	0.59	0.49	0.36	0.22
	65	0.029	0.013	0.003		0.99	0.92	0.85	0.78	0.70	0.58	0.41	0.23
	85	0.041	0.023	0.009	0.002		0.97	0.92	0.86	0.77	0.64	0.44	0.23
	125	0.057	0.037	0.020	0.009	0.003		0.99	0.95	0.87	0.73	0.52	0.28
	165	0.060	0.041	0.024	0.013	0.006	0.001		0.99	0.93	0.81	0.60	0.36
	205	0.053	0.036	0.021	0.012	0.007	0.004	0.001		0.98	0.89	0.71	0.48
	245	0.038	0.023	0.012	0.009	0.009	0.007	0.003	0.000		0.96	0.84	0.64
	285	0.017	0.006	0.004	0.009	0.010	0.007	0.006	0.006	0.003		0.95	0.82
	325	0.019	0.019	0.020	0.020	0.020	0.023	0.026	0.026	0.020	0.008		0.96
	365	0.039	0.038	0.038	0.039	0.042	0.048	0.053	0.055	0.047	0.028	0.007	



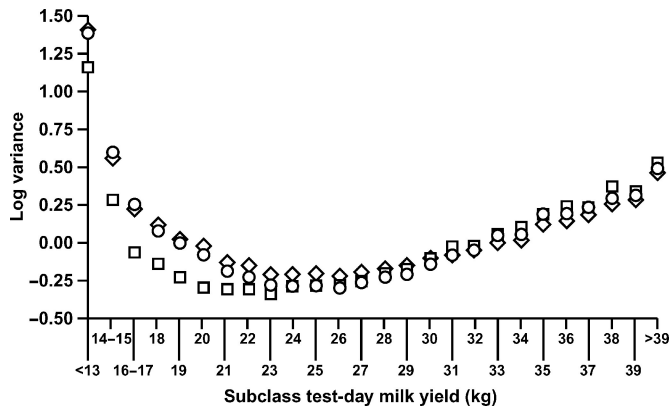
**Figure 3.** Variance model solutions for effect of year and season (1 = January to June; 2 = July to December) of heterogeneity subclass for test-day milk (○), fat (□), and protein (◇) yield.



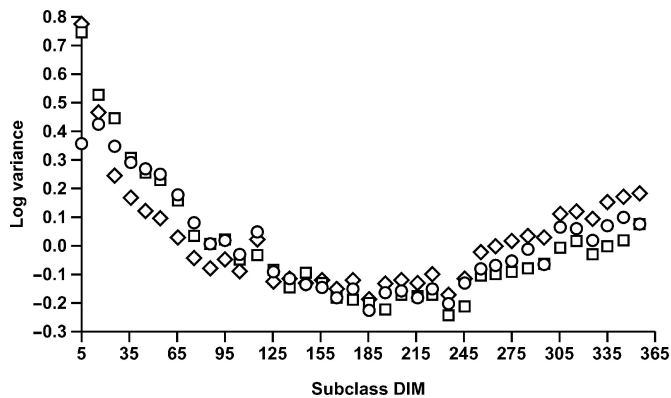
**Figure 4.** Variance model solutions for effect of test month of heterogeneity subclass for test-day milk (○), fat (□), and protein (◇) yield.



**Figure 5.** Variance model solutions for effect of size of heterogeneity subclass for test-day milk (○), fat (□), and protein (◇) yield.



**Figure 6.** Variance model solutions for effect of mean milk yield of heterogeneity subclass for test-day milk (○), fat (□), and protein (◇) yield.



**Figure 7.** Variance model solutions for effect of DIM of heterogeneity subclass for test-day milk (○), fat (□), and protein (◇) yield.

**Table 4.** Variance model solutions for effect of milking frequency by yield trait.

Milking frequency	Log variance		
	Milk	Fat	Protein
2 times/d	-0.10	-0.13	-0.09
≥3 times/d	0.20	0.27	0.19

with high milk yield had increased variance, but, unexpectedly, herds with low milk yield also had increased variance. The increased variance for low-yield herds could have resulted from sick animals or from identification or data-recording errors, but those causes could not be verified. Test-day yields during late and especially during early lactation also were more variable. Repeatabilities of variances for a given class of herd, test day, and milking frequency were 14 to 17%, which indicated that a moderate degree of information came from directly observed variance.

Jointly estimated (co)variance components followed the patterns reported in earlier studies (Gengler and Wiggans, 2001). Future studies should account for heritability differences that are associated with herd yield level. Herd-time effects explained some phenotypic variance, especially at start of lactation.

The method developed to model and to adjust phenotypic variance can also be useful for routine genetic evaluation. However, heterogeneity of (co)variance components, which has been reported by earlier studies of the same data (Gengler and Wiggans, 2001), cannot be corrected through such a method. Further research may overcome this limitation.

**ACKNOWLEDGMENTS**

Nicolas Gengler, who is Chercheur Qualifié of the National Fund for Scientific Research (Brussels, Belgium), acknowledges support through Grant 2.4.507.02 F (2) of the National Fund for Scientific Research. Manuscript review by S. M. Hubbard, A. H. Sanders, and L. L. M Thornton (Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD) is appreciated. The research was based on modified versions of EM REML programs that were provided by I. Misztal (University of Georgia, Athens).

**REFERENCES**

Boldman, K. G., and A. E. Freeman. 1990. Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. *J. Dairy Sci.* 73:503–512.  
 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.* 85:263. [Online.]



- 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.
- Brotherstone, S., and W. G. Hill. 1986. Heterogeneity of variance amongst herds for milk production. *Anim. Prod.* 42:297–303.
- Dodenhoff, J., and H. H. Swalve. 1998. Heterogeneity of variances across regions of northern Germany and adjustment in genetic evaluation. *Livest. Prod. Sci.* 53:225–236. [Online.] Available: [http://dx.doi.org/10.1016/S0301-6226\(97\)00129-2](http://dx.doi.org/10.1016/S0301-6226(97)00129-2).
- Druet, T., I. Misztal, M. Duangjinda, A. Reverter, and N. Gengler. 2001. Estimation of genetic covariances with Method R. *J. Anim. Sci.* 79:605–615.
- Gengler, N., A. Tijani, G. R. Wiggans, and I. Misztal. 1999. Estimation of (co)variance function coefficients for test day yield with an expectation-maximization restricted maximum likelihood algorithm. *J. Dairy Sci.* 82:1849. [Online.]
- 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. [Online.] Available: <http://www-interbull.slu.se/bulletins/bulletin29/Gengler.pdf>.
- Lidauer, M., and E. A. Mäntysaari. 1999. Multiple trait reduced rank random regression test-day model for production traits. *Interbull Bull.* 22:74–80. [Online.] Available: <http://www-interbull.slu.se/bulletins/bulletin22/LIDAUER.PDF>.
- Mayeres, P. 2002. Appui technique à la gestion des troupeaux de bovins laitiers à travers la mise en évidence d'influences spécifiques au niveau troupeaux et des vaches sur les performances et l'efficacité biologique. M.S. Thesis, Gembloux Agric. Univ., Belgium.
- Meinert, T. R., R. E. Pearson, W. E. Vinson, and B. G. Cassell. 1988. Prediction of daughter's performance from dam's cow index adjusted for within-herd variance. *J. Dairy Sci.* 71:2220–2231.
- 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., L. L. G. Janss, and T. H. E. Meuwissen. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.* 83:2640–2649.
- 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. [Online.] Available: [http://dx.doi.org/10.1016/S0301-6226\(99\)00166-9](http://dx.doi.org/10.1016/S0301-6226(99)00166-9).
- Reverter, A., B. L. Golden, R. M. Bourdon, and J. S. Brinks. 1994. Method R variance components procedure: Application on the simple breeding value model. *J. Anim. Sci.* 72:2247–2253.
- Tijani, A., G. R. Wiggans, C. P. Van Tassell, J. C. Philpot, and N. Gengler. 1999. Use of (co)variance functions to describe (co)variances for test day yield. *J. Dairy Sci.* 82:226. [Online.]
- Vargas, B., E. Perez, and J. A. M. Van Arendonk. 1998. Analysis of test day yield data of Costa Rican dairy cattle. *J. Dairy Sci.* 81:255–263.
- Weigel, K. A., and T. J. Lawlor. 1994. Adjustment for heterogeneous variance in genetic evaluations for conformation of United States Holsteins. *J. Dairy Sci.* 77:1691–1701.
- 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.* 85:264.