Genetic Analysis of Milk Production Traits of Polish Black and White Cattle Using Large-Scale Random Regression Test-Day Models

T. Strabel^{*1} and J. Jamrozik^{†‡}

^{*}Agricultural University of Poznań, Department of Animal Genetics and Breeding, Wołyńska 33, 60-637 Poznań, Poland

[†]Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, ON, N1G 2W1, Canada

[‡]National Research Institute of Animal Production, Department of Animal Genetics and Breeding, 32-083 Balice k. Krakowa, Poland

ABSTRACT

Genetic parameters for milk, fat, and protein yield and persistency in the first 3 lactations of Polish Black and White cattle were estimated. A multiple-lactation model was applied with random herd-test-day effect, fixed regressions for herd-year and age-season of calving, and random regressions for the additive genetic and permanent environmental effects. Three data sets with slightly different edits on minimal number of days in milk and the size of herd-year class were used. Each subset included more than 0.5 million test-day records and more than 58,000 cows. Estimates of covariance components and genetic parameters for each trait were obtained by Bayesian methods using the Gibbs sampler. Due to the large size and a good structure of the data, no differences in estimates were found when additional criteria for record selection were applied. More than 95% of the genetic variance for all traits and lactations was explained by the first 2 principal components, which were associated with the mean yield and lactation persistency. Heritabilities of 305-d milk yield in the first 3 lactations (0.18, 0.16, 0.17) were lower than those for fat (0.12, 0.11, 0.12) and protein (0.13, 0.14, 0.15). Estimates of daily heritabilities increased in general with days in milk for all traits and lactations, with no apparent abnormalities at the beginning or end of lactation. Genetic correlations between yields in different lactations ranged from 0.74 (fat yield in lactations 1 and 3) to 0.89 (milk yield in lactations 2 and 3). Persistency of lactation was defined as the linear regression coefficient of the lactation curve. Heritability of persistency increased with lactation number for all traits and genetic correlations between persistency in different lactations were smaller than those for 305d yield. Persistency was not genetically correlated with the total yield in lactation.

Key words: dairy cattle, test-day yield, persistency, random regression model

INTRODUCTION

Estimation of genetic parameters, or equivalently covariance components, for random regression models (**RRM**) of dairy production traits may cause several problems. Models are computationally very demanding due to large numbers of parameters and, therefore, relatively small data sets have usually been used for multiple lactations or multiple traits (Jamrozik et al., 1998; Zavadilová et al., 2005). Research results indicated that the use of large amounts of data is essential for obtaining reliable estimates. Insufficiently large data sets (as demonstrated with subsets taken from the same population) yield estimates of genetic parameters that can sometimes be inconsistent (Druet et al., 2005; Strabel et al., 2005).

A number of studies have shown that estimates of daily variances and heritabilities at the beginning and the end of lactation could be unreasonably high and genetic correlations between yields at the opposite ends of lactation could be negative when small datasets are used (Miszta et al., 2000). Several authors explained such abnormalities as consequences of a lack of information to properly describe variability, especially at the end of lactation (Strabel and Misztal, 1999; Pool et al., 2000). Anomalous estimates have been presented as artifacts of functions (e.g., high-order orthogonal polynomials) used in RRM (Kettunen et al., 1998; Pool et al., 2000). Problems were reported, however, when simplified methods were applied to model environmental variation, such as assuming constant permanent environmental (**PE**) effects across DIM (Jamrozik and Schaeffer, 1997) or using a low-order polynomial for this effect (López-Romero et al., 2003).

The use of random herd-test-date effects in the models was recommended for Polish Black and White populations with small average herd size (Strabel et al., 2005). Additional regressions for herd-year of calving classes allowed for better modeling of daily yield variability. Modeling herd (or herd-year) lactation curves has been shown to have a positive impact on the shape of daily variance and heritability curves (Jamrozik et al., 2001; de Roos et al., 2004).

Received December 21, 2005.

Accepted March 6, 2006.

¹Corresponding author: strabel@man.poznan.pl

Persistency of lactation is a trait of interest for both genetic improvement and management purposes and many measures of this trait have been proposed in the past (Swalve and Gengler, 1999). Random regressions and the properties of orthogonal polynomials allow for a simple and effective differentiation between yield and persistency: the first parameter of genetic regression describes the total yield in lactation, and the second parameter is related to persistency (Druet et al., 2005).

The objective of this research was to estimate genetic parameters for milk, fat, and protein yield in the first 3 lactations of Polish Black and White cattle by using RRM and large data sets.

MATERIALS AND METHODS

Over 16 million test-day (TD) records of milk, fat, and protein yields from the first 3 lactations of Polish Black and White cows calving from 1995 to 2001 were available for this study. Cows in their second or third lactations had to have their preceding lactations in the data set. Moreover, the calving date of primiparous cows was restricted in such a way that all the cows had the chance to complete second and third lactations. Edits included restrictions on daily milk yield between 0.1 and 90 kg, and on the range of age at calving between 18 and 49 mo, 29 and 67 mo, and 39 and 83 for the first, the second, and the third lactations, respectively. Three final data sets (A, B, and C) were created by applying additional sequential edits. All data sets were required to contain records from herds with a minimum of 50 first lactations and with TD records in herd-year classes covering at least 80 DIM. Additionally, for data sets B and C, the minimum size of herd-year classes was 5 TD records. To improve the distribution of records from different DIM, only lactations of at least 150 d were included in data sets A and B, whereas for data set C, this condition was extended to 250 d. After applying these criteria, no random selection of herds was necessary; all the remaining records were used for the estimation process. A detailed description of the data sets is in Table 1. The application of selection criteria slightly decreased the number of records and lactations in consecutive data sets and also increased the average daily milk yield from 14.5 kg for data sets A and B to 15.1 kg in data set C. Distribution of records across DIM in the first 3 lactations for data set A is presented in Figure 1. Even at the very end of the third lactation, at least 400 records were available at every single DIM.

Cows were assigned to subclasses for age-season of calving. Two seasons (April–September and October–March), and 5 classes for age at calving for primiparous cows (below 25 mo, 25 to 26, 27 to 28, 29 to 30, and over 30 mo) were defined, 4 classes for the second lactation

Table 1. Characteristics of 3 subsets of data (A, B, C)¹

	Data set A	Data set B	Data set C
Minimum no. of first lactations per herd	50	50	50
Minimum length of lactation	150	150	250
Minimum no. of test days per herd-year	0	5	5
Total no. of test-day records	1,229,941	1,126,231	988,848
No. of records per lactation 1/2/3	592,617 / 400,327 / 236,997	543,165 / 366,678 / 216,388	522,543 / 307,548 / 158,757
No. of animals with records per lactation 1/2/3	69,952 / 48,070 / 29,042	64,248 / 44,018 / 26,477	58,896 / 34,766 / 18,107
No. of animals without records	60,743	56,403	52,129
Lactations with test-day records after 250 DIM (%)	78	78	97
No. of herds	657	606	546
No. of daughters per bull 1/2/3	17.1 / 13.2 / 9.3	18.0 / 13.9 / 9.8	16.3 / 11.1 / 7.2
Average no. of test-day records per herd test day per lactation 1/2/3	17.4 / 13.1 / 9.3	17.2 / 13.0 / 9.2	17.7 / 11.4 / 7.1
Average no. of test days per cow per lactation 1/2/3	8.4 / 8.3 / 8.1	8.4 / 8.3 / 8.1	8.8 / 8.8 / 8.7
Mean daily yield (kg) per lactation 1/2/3	14.5 / 16.8 / 18.5	14.5 / 16.8 / 18.5	15.1 / 17.6 / 19.5
Standard deviation of milk yield per lactation 1/2/3	5.48 / 6.93 / 7.48	5.46 / 6.90 / 7.46	5.39 / 6.91 / 7.50
Fat yield (kg) per lactation 1/2/3	0.59 / 0.70 / 0.77	0.59 / 0.70 / 0.77	0.62 / 0.73 / 0.81
Standard deviation of fat yield per lactation 1/2/3	0.23 / 0.30 / 0.34	0.23 / 0.30 / 0.34	0.23 / 0.30 / 0.34
Protein yield (kg) per lactation 1/2/3	0.46 / 0.55 / 0.60	0.46 / 0.55 / 0.60	0.48 / 0.58 / 0.63
Standard deviation of protein yield per lactation 1/2/3	0.17 / 0.21 / 0.23	0.17 / 0.21 / 0.23	0.17 / 0.21 / 0.23

¹A and B = minimum 150 DIM; C = minimum 250 DIM; B and C = minimum 5 records in herd-year classes.

Figure 1. Distribution of records across DIM in the first (-----), second (—), and third (—) lactations for data set A. A = minimum 150 DIM, no requirements for number of records in herd-year classes.

(below 39 mo, 39 to 41, 42 to 44, and over 44 mo), and 3 for the third lactation (below 52 mo, 52 to 55 and over 55).

The model used in this study was the multiple-lactation generalization of the model presented by Strabel et al. (2005). The model equation was:

$$Y_{ijklmo} = htd_{io} + \sum_{n=1}^6 b_{jno} z_{mnlo} + \sum_{n=1}^4 c_{kno} z_{mnlo} + \sum_{n=1}^4 a_{mno} z_{mnlo} + \sum_{n=1}^4 p_{mno} z_{mnlo} + e_{ijklmo}$$

where y_{ijklm} is milk, fat, or protein yield l of cow m from lactation o within herd-test-day effect i , belonging to the herd-year of calving k and j th class of age-season of calving; HTD_i is a random herd-test-day effect; b_{jno} are fixed regression coefficients specific to age-season subclass j ; c_{kno} are fixed regression coefficients specific to the herd-year k ; a_{mno} are genetic random regression coefficients specific to animal (AG) m ; p_{mno} are random regression coefficients for the PE effect; e_{ijklmo} is the residual effect for each observation; and z_{mnlo} are covariates. Fixed lactation curves for age-seasons were modeled using Legendre polynomials of order 5 (6 covariates). The same function with 4 parameters was used for all remaining regressions.

In matrix notation the model can be presented as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Uq} + \mathbf{Wp} + \mathbf{Za} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of observations, $\mathbf{b}$ is a vector of fixed regression coefficients for age-season of calving classes and herd-years, $\mathbf{q}$ is a vector of random herd-test-date effects, $\mathbf{p}$ is a vector of PE regression coefficients, $\mathbf{a}$ is a vector of AG regression coefficients, $\mathbf{e}$ is a vector of residuals, and $\mathbf{X}$, $\mathbf{U}$, $\mathbf{W}$, and $\mathbf{Z}$ are incidence matrices relating observations to effects.

The covariance structure of the model was:

$$V \begin{bmatrix} \mathbf{q} \\ \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \otimes \mathbf{H}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \otimes \mathbf{G}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \otimes \mathbf{P}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \otimes \mathbf{R}_0 \end{bmatrix},$$

where $\mathbf{A}$ is the additive genetic relationship matrix; $\mathbf{H}_0$ and $\mathbf{R}_0$ are diagonal matrices of order 3×3 , with herd-test-date and residual variances for 3 lactations, respectively; and $\mathbf{P}_0$ and $\mathbf{G}_0$ are covariance matrices for PE and AG random regression coefficients, respectively, each of order 12×12 .

The Bayesian method with Gibbs sampling was used to generate 120,000 samples for each trait and data set (Jamrozik et al., 1998). Flat priors and normal distributions were assumed for all fixed and random effects, respectively, and inverted Wishart distributions with minimal numbers of degrees of belief were the prior distributions for all covariance components. The resulting conditional distributions were either multivariate normal (location parameters) or inverted Wishart (dispersion parameters). Convergence of Gibbs chains

Figure 2. Sample values of additive genetic variance for the first regression parameter of milk yield in the first lactation (data set B); B = minimum 150 DIM, minimum 5 records in herd-year classes.

was monitored by inspecting plots of samples for selected parameters. Effective sample size (**ESS**) for covariance components was estimated for all traits and data sets using the method of initial monotone sequence estimator (Geyer, 1992). Posterior means of covariance components, heritabilities, and correlations were calculated using 100,000 samples after discarding the first 20,000 iterations as a burn-in period. Estimates of covariance components from different data sets were compared using a matrix norm technique. Principal components analysis was performed on AG and PE covariances within each lactation. Daily and lactation heritability estimates were calculated according to formulas presented by Jamrozik and Schaeffer (1997).

RESULTS AND DISCUSSION

The sample plot of variance for the AG intercept for milk yield in the first lactation estimated on data set B is shown in Figure 2. Similar behavior of the chain was observed for data sets A and C. Plots of realizations for several other covariance components (results not shown) indicated that 20,000 iterations were sufficient as a burn-in for all traits and data sets.

The ESS for the estimated covariance components varied across effects and traits. Table 2 shows simple statistics of the ESS for covariance components for milk yield. The effective chain length was much smaller for

AG and PE effects than for HTD and residual variances. The ESS for AG components averaged slightly more than 50 samples (all traits and data sets) and they were approximately 2 times larger for PE components. Relatively high variation in the average ESS for HTD and residual variance across traits was observed. The ESS for milk yield and data set B were 11,493 and 11,460 for HTD and residual variance, respectively. The pattern of differences in ESS for covariance components did not indicate any particular association between data structure and the ESS. Discrepancies between standard deviation of ESS for these effects between data sets could also be observed. For example, SD of ESS for fat yield and data set A were 1,278 and 5,784 for HTD and residual effects, respectively, and the corresponding values for data set B were 4,361 and 945. Differences in ESS for AG and PE components were negligible for most combinations of traits and data sets.

Table 3 presents Euclidean norms for AG, PE, HTD, and residual covariance matrices for all 3 data sets. In general, all data sets gave very similar norms of respective covariance matrices. The most extreme norm of a difference between data sets did not exceed 10% of norms obtained with data set A. The range of norms of differences and their ratios to norms for data set A indicated a large degree of similarity between covariance components from all 3 data sets.

Table 2. Effective sample size for the additive genetic (AG), permanent environmental (PE), herd-test-day (HTD), and residual (E) effects for milk yield from different data sets¹

Data set	Effect	No. of parameters	Effective sample size			
			Mean	Minimum	Maximum	SD
A	AG	78	58	29	127	20
	PE	78	129	49	266	53
	HTD	3	11,527	10,964	12,311	699
	E	3	12,418	6,602	18,929	6,192
B	AG	78	60	17	125	22
	PE	78	130	24	272	49
	HTD	3	11,492	10,483	12,112	881
	E	3	11,459	8,504	15,929	3,937
C	AG	78	62	23	129	21
	PE	78	131	42	267	47
	HTD	3	11,916	9,857	13,040	1,785
	E	3	11,450	5,485	17,451	5,983

¹A and B = minimum 150 DIM; C = minimum 250 DIM; B and C = minimum 5 records in herd-year classes.

Eigenvalues for the within-lactation AG and PE covariance matrices for all the analyzed traits from data set A and relative differences of corresponding values with the remaining data sets are shown in Table 4. The first principal component of the AG effect explained from 81.0 to 88.5% of variation in milk yield in all lactations. This proportion was similar for fat and protein yields. The first principal component of the PE effect explained relatively less variation and the third component for this effect described from 6.6 to 8.9% of variation. Relative differences of eigenvalues from data sets B and C with corresponding values from data set A were small. Differences for the first 2 eigenvalues for all traits did not exceed 6 and 11% for data set B and C, respectively. Small relative differences for most of the remaining eigenvalues confirmed almost identical

nature of variability for all 3 data sets. The pattern of these differences did not show a clear relationship between variance partition and the data structure.

The majority of variance explained by the first 2 principal components confirmed that the first 2 parameters, associated with the level of production and its persistency, play a key role in explaining variation of yield traits throughout lactations (Jamrozik et al., 2002; Druet et al., 2005). This was also in agreement with the study of Pool et al. (2000) who found that for the genetic part, the first 3 components explained over 98% of the variation, but for the PE part, 4 components were needed to reach this level. Similarly to our results, Druet et al. (2005) found that the first 2 components explained more than 95% of variation. Results of Jamrozik et al. (2002) showed that differences between pop-

Table 3. Euclidean norms for additive genetic (AG), permanent environmental (PE), herd-test day (HTD), and residual (R) covariance matrices for 3 data sets (A, B, C)¹

		Data set A	Data set B	Data set C	B – A	C – A	(B – A)/A (%)	(C – A)/A (%)
Milk yield	AG	5.99	5.97	6.00	–0.02	0.01	–0.3	0.1
	PE	11.98	11.89	10.88	–0.09	–1.10	–0.7	–9.1
	R	7.71	7.73	8.02	0.03	0.31	0.3	4.0
	HTD	3.64	3.68	3.51	0.04	–0.13	1.0	–3.6
Fat yield ²	AG	0.99	1.03	1.00	0.03	0.01	3.5	1.2
	PE	2.39	2.36	2.31	–0.03	–0.07	–1.0	–3.0
	R	2.80	2.80	3.04	0.00	0.24	–0.1	8.4
	HTD	0.72	0.73	0.71	0.01	–0.02	0.9	–2.3
Protein yield ²	AG	0.52	0.52	0.52	0.00	–0.01	–0.4	–1.2
	PE	1.16	1.16	1.08	0.00	–0.09	–0.3	–7.4
	R	0.97	0.97	1.04	0.00	0.07	–0.1	7.1
	HTD	0.49	0.49	0.45	0.00	–0.04	0.6	–8.6

¹A and B = minimum 150 DIM; C = minimum 250 DIM; B and C = minimum 5 records in herd-year classes.

²Values multiplied by 100.

Table 4. Eigenvalues and their relative differences for additive genetic (AG) and permanent environmental (PE) covariance matrices for milk, fat, and protein yield in the first 3 lactations for data sets A, B, and C¹

Trait and variance component	Lactation	Eigenvalues				Relative difference, (B – A)/A, of eigenvalues (%)				Relative difference, (C – A)/A, of eigenvalues (%)			
		1	2	3	4	1	2	3	4	1	2	3	4
Milk yield AG	1	1.65	0.13	0.04	0.02	–1.1	4.6	7.5	11.5	1.2	8.9	16.7	12.5
	2	2.17	0.29	0.06	0.04	1.4	–3.7	3.6	5.9	–0.9	–4.8	11.5	3.0
	3	2.72	0.41	0.13	0.09	–0.3	–0.3	4.4	–7.2	4.9	–0.7	23.9	8.1
PE	1	3.92	1.01	0.38	0.19	0.4	–0.5	0.4	–0.3	–10.8	–4.9	0.1	–6.6
	2	6.26	1.63	0.57	0.26	–1.2	0.2	–1.2	–0.6	–8.4	–2.9	1.7	–3.2
	3	7.16	1.84	0.71	0.28	–0.7	0.7	0.6	3.0	–8.3	–2.2	1.7	–7.4
Fat yield ² AG	1	0.23	0.02	0.01	0.01	1.9	1.5	9.9	7.6	2.4	–3.1	19.8	9.9
	2	0.35	0.06	0.01	0.01	5.8	–4.1	4.9	–0.1	–2.9	–10.1	16.2	–5.2
	3	0.51	0.11	0.03	0.02	2.5	–4.3	4.8	–3.4	8.9	–4.1	43.6	–0.1
PE	1	0.72	0.20	0.08	0.03	–0.5	–0.3	–0.3	–0.1	–8.3	–3.3	–3.7	–5.6
	2	1.19	0.34	0.13	0.03	–1.5	0.9	–0.3	1.7	–4.2	2.3	2.3	–13.7
	3	1.47	0.42	0.17	0.02	–0.7	1.7	0.1	0.3	0.2	0.4	2.8	–8.2
Protein yield ² AG	1	0.12	0.01	0.01	0.01	–0.6	7.2	6.5	19.4	2.7	9.6	9.6	11.1
	2	0.19	0.02	0.01	0.01	1.5	–3.7	–2.7	1.5	–2.4	–3.4	11.5	–3.5
	3	0.25	0.03	0.02	0.01	–1.4	0.2	1.0	–3.6	1.8	4.3	16.9	9.8
PE	1	0.38	0.10	0.04	0.02	0.2	–0.5	–0.5	2.0	–10.1	–1.9	2.3	–6.6
	2	0.58	0.17	0.06	0.03	–1.0	1.2	–0.7	–1.6	–8.1	0.7	2.2	–9.2
	3	0.67	0.19	0.08	0.03	–0.1	0.4	0.4	3.5	–5.8	1.6	3.6	–0.7

¹A and B = minimum 150 DIM; C = minimum 250 DIM; B and C = minimum 5 records in herd-year classes.

²Values multiplied by 100.

ulations were small with respect to principal component analyses of the genetic covariance matrices. Corresponding values for the PE effect were in general lower than for the AG effect and they were similar to other studies (from 72 to 85%; Pool et al., 2000; Jamrozik et al., 2002; Druet et al., 2005). Druet et al. (2005) used only 2 eigenvectors, associated with the first 2 eigenvalues, as covariates to estimate covariance components of daily milk traits. Decisions concerning the rank reduction of the model should take into account the proportion of variance explained by the omitted parameters. For the PE effect in the second lactation, the third principal component explained between 9 and 13% of the total variation, which can be regarded as a substantial amount of variation and is possibly important at the beginning and end of the lactation.

Convergence behavior, effective sample size, norms, and principal component analysis of covariance matrices indicated very similar patterns of variability for all 3 data sets. Therefore, the remaining part of the paper will concentrate mainly on results obtained from the largest and the least edited data set (A).

Heritability

Table 5 gives heritabilities and correlations for the first 2 random regression coefficients for different parities. The first coefficient of orthogonal Legendre polynomials describes the total yield in lactation; therefore,

heritability of this coefficient is equivalent to heritability of 305-d lactation yield. Estimates of heritability for milk were higher than for the other traits and they were equal to 0.18, 0.16, and 0.17 for the first, second, and third lactation, respectively. The second regression coefficient can be viewed as the persistency measure. Heritabilities of persistency were lower than corresponding estimates for the total yield. In general, they increased steadily with lactation number, eventually reaching the magnitude of heritability of 305-d yield. This tendency was consistent for all traits and data sets (results for data set B and C not shown). Fat persistency had heritabilities equal to 0.12, 0.17, and 0.20; and protein persistency estimates were 0.11, 0.15, and 0.18 for lactations 1, 2, and 3, respectively. Although variance of both the first and the second parameter in general increased with lactation number, relative proportions of variance increased substantially for later lactations. Production in the second and third lactations is usually higher; and lactation curves therefore must be steeper, which leaves space for higher variability and offers a chance for improvement of genetic persistency. Jamrozik et al. (2002) reported various genetic relationships between the first 2 parameters of the AG effect for different populations, which can be caused by varying levels of production and the location of the peak yield. In such a case, the average level of persistency may, to a large extent, depend on the average level of production in the population.

Table 5. Posterior means (SD in parentheses) of heritabilities of the first 2 regression parameters of additive genetic effects and the ratio of variance of the first 2 regression parameters of permanent environmental variance for milk, fat, and protein yield in the first 3 lactations for data set A¹

Parameter ²	Milk yield			Fat yield			Protein yield		
	Lactation 1	Lactation 2	Lactation 3	Lactation 1	Lactation 2	Lactation 3	Lactation 1	Lactation 2	Lactation 3
a ₀	0.18 (0.01)	0.16 (0.01)	0.17 (0.01)	0.12 (0.00)	0.11 (0.01)	0.12 (0.01)	0.13 (0.01)	0.14 (0.01)	0.15 (0.01)
a ₁	0.11 (0.01)	0.16 (0.01)	0.18 (0.01)	0.12 (0.01)	0.17 (0.01)	0.20 (0.01)	0.10 (0.01)	0.16 (0.01)	0.18 (0.01)
p ₀	0.44 (0.01)	0.48 (0.01)	0.46 (0.01)	0.37 (0.01)	0.37 (0.01)	0.35 (0.01)	0.41 (0.01)	0.43 (0.01)	0.42 (0.01)
p ₁	0.88 (0.01)	0.83 (0.01)	0.81 (0.01)	0.87 (0.01)	0.83 (0.01)	0.79 (0.01)	0.89 (0.01)	0.83 (0.01)	0.81 (0.01)

¹A = minimum 150 DIM, no requirements for number of records in herd-year classes.

²Parameters: a₀ and a₁ and p₀ and p₁ are the first and second parameters, respectively, of Legendre polynomials of additive genetic and permanent environmental effects.

Although heritabilities for genetic persistency were found to be relatively low, the genetic correlation between the first and the second regression parameters for the AG effect was almost zero. Moderate heritabilities and a lack of relation to the total 305-d production will allow for an efficient selection for the shape of the lactation curve in Polish Black and White cattle.

The ratio of variance (PE over sum of AG and PE) for the first 2 parameters of the PE effect and the correlation between them are presented in Table 5 for all traits and lactations. The estimates of variance ratio were much higher than corresponding heritabilities. They were equal to 0.44, 0.48, and 0.46 for the first PE coefficient of milk yield, for the first, second, and third lactations, respectively. For the second PE parameter, these ratios were doubled on average when compared with the intercept. Only 2 variance components (AG and PE) were included in the calculation of heritability and the PE variance ratio for regression coefficients. This corresponded to the partition of variance in the second stage of linear hierarchical model for TD yields as in Jamrozik et al. (2001).

Curves of AG and PE variances of daily milk yields for the first 3 lactations (Figure 3) were characterized by patterns analogous to those presented by Strabel et al. (2005) for the same population analyzed with the same model for first-lactation data. Increased estimates for AG, and especially for PE, effects at the peripheries

of the trajectory were found. A tendency toward an increase of genetic variance with DIM, noticed for all the traits and data sets, was in agreement with results of Druet et al. (2005). However, this propensity was not observed in the first lactation, possibly due to the low total variance. Differences among shapes for different data sets were negligible. Heritability estimates for daily yields were low and ranged in general between 0.1 and 0.2 for all lactations and data sets (Figure 4). Estimates were generally higher for milk than for fat and protein yields. Lower heritabilities of daily milk yields than those for fat and protein yields were reported in the previous study of Polish Black and White population (Strabel and Misztal, 1999) and for other populations (Liu et al., 2000; Jakobsen et al., 2002; Druet et al., 2005). Lidauer et al. (2003) also obtained higher daily estimates for milk (0.30) than for fat (0.22) and protein (0.23) for the Finnish population, similar to the results of Jakobsen et al. (2002) for the Danish population (0.42, 0.37, and 0.36 for milk, fat, and protein, respectively). The highest heritabilities were obtained for milk yield and the lowest for fat yield in the study of Zavadilová et al. (2005) and Rekaya et al. (1999).

Relatively flat shapes of heritability curves were found in this study. Several different patterns, however, can be distinguished. For example, first-lactation milk and protein yield heritabilities were higher in the

Table 6. Posterior means (SD in parentheses) of correlations between lactations for the first 2 regression parameters of additive genetic and permanent environmental effects for milk, fat, and protein yield in the first 3 lactations for data set A¹

Parameter	Milk yield			Fat yield			Protein yield		
	1 – 2	2 – 3	1 – 3	1 – 2	2 – 3	1 – 3	1 – 2	2 – 3	1 – 3
a ₀	0.84 (0.01)	0.89 (0.01)	0.80 (0.02)	0.82 (0.01)	0.85 (0.01)	0.73 (0.02)	0.83 (0.01)	0.87 (0.01)	0.78 (0.02)
a ₁	0.59 (0.04)	0.65 (0.03)	0.50 (0.04)	0.55 (0.04)	0.65 (0.03)	0.43 (0.04)	0.52 (0.04)	0.62 (0.03)	0.44 (0.04)
p ₀	0.40 (0.01)	0.40 (0.01)	0.31 (0.01)	0.42 (0.01)	0.43 (0.01)	0.32 (0.01)	0.44 (0.01)	0.44 (0.01)	0.31 (0.01)
p ₁	0.13 (0.01)	0.24 (0.01)	0.11 (0.01)	0.10 (0.01)	0.23 (0.01)	0.10 (0.01)	0.13 (0.01)	0.22 (0.01)	0.10 (0.01)

¹A = minimum 150 DIM, no requirements for number of records in herd-year classes.

²Parameters: a₀ and a₁ and p₀ and p₁ are the first and second parameters of Legendre polynomials of additive genetic and permanent environmental effects, respectively.

Figure 3. Additive genetic (black curves) and permanent environmental (gray curves) variances for milk yield estimated for analyzed data sets (A, B, and C) and the first (-----), second (—), and third (—) lactations; A and B = minimum 150 DIM; C = minimum 250 DIM; B and C = minimum 5 records in herd-year classes.

middle part of lactation than at the beginning or end, which was in contrast to fat yield. Heritabilities for second and third lactations showed an increasing trend with DIM. No undesired extreme estimates at the peripheries of lactations were found for any of the analyzed traits and lactations. Daily heritabilities for later lactations based on data set C showed a tendency toward slightly lower values at the very end of lactation.

Table 7. Posterior means (SD in parentheses) of correlations between first 2 regression parameters of additive genetic effect for milk, fat, and protein yield obtained for data set A¹

Lactation	Milk yield	Fat yield	Protein yield
1	0.06 (0.051)	−0.12 (0.053)	0.19 (0.055)
2	0.26 (0.050)	0.21 (0.054)	0.41 (0.048)
3	0.21 (0.050)	0.07 (0.054)	0.33 (0.048)

¹A = minimum 150 DIM, no requirements for number of records in herd-year classes.

Heritability curves of various shapes have been presented in literature (Misztal et al., 2000). Daily heritability graphs with high values at the beginning and at the end of lactations (border or wave effect) were the features of early applications of RRM, in which the PE effect was constant during lactation (Jamrozik and Schaeffer, 1997) or relatively small data sets were analyzed with single-trait models (Strabel and Misztal, 1999; Ptak et al., 2004). Heritability curves with oscillatory patterns lack clear biological explanations and were assumed to be artifacts of using functions of high order to explain random effects in the model. The insufficient amount of data to model complex variability was raised as a possible reason of biased estimation of variances at the end of lactation. Results of this study could support this hypothesis when compared with estimates of Strabel and Misztal (1999) based on much smaller data sets. A tendency toward an increase of genetic

Figure 4. Milk, fat, and protein yield heritabilities for different DIM of the first 3 lactations obtained for data sets A (---), B (— · —), and C (—); A and B = minimum 150 DIM; C = minimum 250 DIM; B and C = minimum 5 records in herd-year classes.

variance with DIM, noticed for all the traits and data sets, is in agreement with results of Druet et al. (2005). However, this tendency was not observed in the first lactation, possibly due to the low total variance in the first lactation.

A limited number of records used for parameter estimation may also lead to accidental results, which are not confirmed by repeated estimations based on different random sample of data (Druet et al., 2003; Strabel et al., 2005). Similarly, Pool et al. (2000) stated that their preliminary results from 8,000 lactations were inconsistent and indicated that more lactations should have been used for the estimation. Additional editing criteria (e.g., excluding shorter lactations from the data set C) did not change the shape of heritability curves or the magnitude of estimates.

Correlations

Correlations of the first AG regression coefficient (305-d yield) between consecutive lactations for data sets A and B ranged between 0.8 and 0.9 for all the traits (Table 6). They were lower between the first and the third lactation (slightly below 0.8) with a similar trend of lower values for data set C (results not shown). Genetic correlations for persistency (second regression coefficient) were much lower and more variable than for the first parameter. Estimates were as low as 0.32 between the first- and the third-lactation milk and fat yield and 0.66 for the fat yield in the second and the third lactations. Patterns of persistency correlations were similar to those for total production in lactation: the lowest correlations were observed between the first and the third and the highest between the second and

Figure 5. Genetic correlations between the first three lactations for milk, fat, and protein yields for different DIM obtained for data sets A (---), B (— — —), and C (——). From the left, curves represent the following pairs of lactations: 1 and 2, 2 and 3 and 1 and 3; A and B = minimum 150 DIM; C = minimum 250 DIM; B and C = minimum 5 records in herd-year classes.

the third lactations. Correlations among PE regression coefficients were lower than those for the AG effect and they ranged between 0.32 and 0.45 and between 0.10 and 0.25, for the intercept and the linear term, respectively.

Correlations between the first and the second random regression parameter for the AG effect for all the traits were small and they were practically equal to zero for first-lactation milk yield (Table 7). The lowest, negative

estimate (-0.12) was obtained for the first-lactation fat yield and the highest for the second-lactation traits.

Genetic correlations among milk yields at the same DIM in various lactations were all positive and greater between consecutive lactations than between the first and the third lactation (Figure 5). The pattern for correlations across the trajectory was similar for all the traits and lactations with the smallest estimates at the peripheries of lactation. However, lower correlations be-

tween the first and the third lactation were observed in the first half of lactation. For all traits between the first- and the second-parity correlations were below 0.7 at the beginning of the lactation, not exceeding 0.9 in the middle part, and dropping below 0.7 at the end of lactation for data sets A and B. Data set C can be characterized by a similar pattern, but clearly lower estimates were obtained between the second and the third lactation. Estimates of genetic correlations smaller than 0.5 between yields at the peripheries of the first and other lactations could be regarded as artifacts of the model. This hypothesis, however, could not be supported by results from this study in which a large number of records was used for parameter estimation. One of the key points of breeding programs is selection of young bulls, usually based on their daughters' performance in the first lactation. Because longevity, and hence performance in later lactations is also of interest for breeders, the effectiveness of selection schemes will depend on the genetic correlations between the first and later lactations. Results from this study showed, in general, lower heritabilities in the first lactation than in the second and the third due to a lower genetic variance in the first lactation. That also reflects a lower correlation between the first and the second lactations and even lower between the first and the third than between second and third lactations estimated in this study for all the traits. Similar patterns of correlation estimates were presented in recent studies by Zavadilová et al. (2005) and by Muir et al. (accepted).

General

The problem of selecting records for the data set used for variance component estimations has been discussed in several studies. Druet et al. (2003) suggested the use of the pooling method to decrease data sensitivity. In a previous study on the same population, Strabel et al. (2005) suggested using all the available data when estimating genetic parameters for the routine genetic evaluation model, due to the relatively large impact of the data set on variance component estimation. In comparison with that study, in which estimation was based on the REML method, employing the Gibbs sampling technique made it possible to use all the available data. Applied edits improved the structure of the data; in particular, numbers of records from different lactations were much better balanced. Pool and Meuwissen (1999) suggested using complete lactations, which were regarded as the interval of a maximum 50 d, at least one test day before 80 and one after 280 DIM. For that reason, we imposed a constraint on data set C to include lactations with records covering at least 250 DIM. This restriction had no clear and consistent impact on the

obtained curves of variances, and, in consequence, on estimates of genetic parameters. Shifts between PE and error variances at the end of the second and the third lactation were noticed, as well as unexpected changes in the first part of lactation. Thus, when the data set is composed of large number of records from all lactations and when it is relatively well balanced, there is no need to exclude shorter lactations from the analysis. Pool et al. (2000) suggested that, although selection for at least 10 observations per herd-test-day might have favored lactations from large farms, the data set remained a representative sample of the Dutch cattle population. For these reasons we used 2 data sets with different requirements for the number of records per HY effect. No impact of this restriction was found on estimates of genetic parameters. The number of records from successive lactations is connected with the structure of the data set. A multiple-trait model generally leads to better results—fewer artifacts are observed (Strabel and Misztal, 1999; Zavadilová et al., 2005). Requirements regarding the data set, its size, and its structure, are higher when more small herds, a more complicated model (function of higher order for submodels), heterogeneity of variance across regions, and various levels of production exist. On the other hand, high restrictions on the data structure, especially the minimum number of records per herd or environmental effect subclasses, may weaken data representativeness. For this reason, and considering minor differences between covariance components for different data sets, we recommend using estimates for the largest data set (A).

CONCLUSIONS

Results of this study indicated that large data sets should be used for estimation of covariance components for the RRM. No restrictions are then necessary with regard to using only longer lactations and records from larger subclasses of fixed effect. Assuming that a homogeneous error variance does not need to lead to abnormal estimates of heritability at the edges of lactation if AG and PE effects are modeled with the same submodel of sufficiently large order.

ACKNOWLEDGMENTS

The Polish State Committee for Scientific Research, Project KBN No 6 P06D 019 20, supported this work.

REFERENCES

- de Roos, A. P. W., A. G. F. Harbers, and G. de Jong. 2004. Random herd curves in a test-day model for milk, fat, and protein production of dairy cattle in the Netherlands. *J. Dairy Sci.* 87:2693–2701.
- 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.
- Druet, T., F. Jaffrézic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. *Genet. Sel. Evol.* 37:257–271.
- Geyer, C. J. 1992. Practical Markov chain Monte Carlo (with discussion). *Stat. Sci.* 7:473–511.
- Jakobsen, J. H., P. Madsen, J. Jensen, J. Pedersen, L. G. Christensen, and D. A. Sorensen. 2002. Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. *J. Dairy Sci.* 85:1607–1616.
- Jamrozik, J., D. Gianola, and L. R. Schaeffer. 2001. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. *Livest. Prod. Sci.* 71:223–240.
- Jamrozik, J., and L. R. Schaeffer. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *J. Dairy Sci.* 80:762–770.
- Jamrozik, J., L. R. Schaeffer, and F. Grignola. 1998. Genetic parameters for production traits and somatic cell score of Canadian Holsteins with multiple trait random regression models. *Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia* 23:303–306.
- Jamrozik, J., L. R. Schaeffer, and K. A. Weigel. 2002. Estimates of genetic parameters for single-and multiple-country test-day models. *J. Dairy Sci.* 85:3131–3141.
- Kettunen, A., E. A. Mäntysaari, I. Strandén, J. Pösö, and M. Lidauer. 1998. Estimation of genetic parameters for first lactation test day milk production using random regression models. *Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia* 23:307–310.
- Lidauer, M., E. A. Mäntysaari, and I. Strandén. 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livest. Prod. Sci.* 79:73–86.
- Liu, Z., F. Reinhardt, and R. Reents. 2000. Estimating parameters of a random regression test day model for first three lactation milk production traits using the covariance function approach. *Interbull Bull.* 25:74–80.
- López-Romero, P., R. Rekaya, and M. J. Carabaño. 2003. Assessment of homogeneity vs. heterogeneity of residual variance in random regression test-day models in a Bayesian analysis. *J. Dairy Sci.* 86:3374–3385.
- Misztal, I., T. Strabel, J. Jamrozik, E. A. Mäntysaari, and T. H. E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. *J. Dairy Sci.* 83:1125–1134.
- Muir, B. L., G. Kistemaker, J. Jamrozik, and F. Canavesi. Genetic parameters for a multiple-trait multiple-lactation random regression test day model in Italian Holsteins. *J. Dairy Sci.* 89:(accepted).
- Pool, M. H., and T. H. E. Meuwissen. 1999. Reduction of the number of parameters needed for a polynomial random regression test day model. *Livest. Prod. Sci.* 64:133–145.
- 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.
- Ptak, E., W. Jagusiak, and J. Jamrozik. 2004. Estimates of heritability for first and second lactation test-day milk production in Polish Black-and-White cattle. *Anim. Sci. Pap. Rep.* 22:105–108.
- 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.
- Strabel, T., and I. Misztal. 1999. Genetic parameters for first and second lactation milk yields of Polish Black and White cattle with random regression test-day models. *J. Dairy Sci.* 82:2805–2810.
- Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of random regression test-day models for Polish Black and White cattle. *J. Dairy Sci.* 88:3688–3699.
- Swalve, H. H., and N. Gengler. 1999. Genetics of lactation persistency. *Occ. Publ. Br. Soc. Anim. Sci.* 24:75–82.
- Zavadilová, L., J. Jamrozik, and L. R. Schaeffer. 2005. Genetic parameters for test-day model with random regressions for production traits of Czech Holstein cattle. *Czech J. Anim. Sci.* 50:142–154.