Journal of Animal and Feed Sciences, 10, 2001, 559 – 569

Estimates of genetic parameters of milk traits using REML and Gibbs sampling

E. Ptak, M. Buś, M. Morek-Kopeć and A. Żarnecki

Department of Genetics and Animal Breeding, Cracow Agricultural University
Al. Mickiewicza 24/28, 30-059 Kraków, Poland

(Received 11 July 2001; accepted 6 November 2001)

ABSTRACT

Two methods of variance component estimation: Restricted Maximum Likelihood (REML) and Gibbs Sampling (GS), were used to analyze data consisting of 305d first lactation milk and fat yield and fat content of 47,574 cows calved from 1989 through 1996. Two three-trait linear models were applied, both including fixed effect of HYS, random effect of animal and error. One included fixed effect of age at calving class, and the other regression on age at calving.

The estimates of genetic correlations and heritabilities were higher when Gibbs Sampling was used; the phenotypic correlations did not depend on the method applied (GS or REML). The heritability estimates for milk yield were 0.28-0.29 by GS and 0.24 by REML. For fat yield they were 0.24 with GS and 0.19 with REML. For fat content they were 0.44-0.45 and 0.37, respectively. The biggest difference between h^2 estimates was found for the fat content. The magnitude of heritabilities was not influenced by the linear model used (with age classes or regression on age).

KEY WORDS: variance components, Gibbs sampling, restricted maximum likelihood, milk traits

INTRODUCTION

Estimates of genetic parameters such as heritabilities and genetic correlations are necessary for evaluation of breeding values and designing breeding programmes. The variance and covariance components needed for these estimates require multivariate analysis which accounts more accurately for selection than univariate analysis (Meyer, 1991). These components can be obtained by a variety of methods among which restricted maximum likelihood (REML) has been most frequently used in animal breeding during the last fifteen years (Hartley and Rao, 1967; Patterson and Thompson, 1971). Several software packages for REML

variance/covariance estimation are available (Meyer, 1988; Jensen and Madsen, 1992; Misztal, 1994; Boldman et al., 1995).

During the last decade the Bayesian approach based on Markov Chain Monte Carlo methods (MCMC) was introduced to estimate parameters of more complicated models in animal breeding (Gianola and Fernando, 1986; Jensen et al., 1994). A variant of MCMC methods known as the Gibbs sampler is used to obtain ML estimates of variances in linear models (Gelfand and Smith, 1990; Casella and George, 1992). In animal breeding, Gibbs sampling was used for the first time by Wang et al. (1993, 1994).

As a general numerical integration method, the Gibbs sampler is particularly useful for high-dimensional integration in situations where ML methods have failed. The method generates random samples from the marginal posterior distribution of all parameters in the model through sampling from and updating conditional posterior distributions (Wang et al., 1994). Using these random samples, the point estimates of variance components and their errors can be calculated. This is the main advantage of Gibbs sampling over other methods of estimation. Although computationally intensive and complicated, the procedure is simple for programming because all calculations are expressed in scalar algebra and no matrix inversion is needed.

This study compares estimates of genetic parameters for milk production traits obtained by Gibbs sampling (GS) and the REML methods, both applied to multiple trait (MT) animal model.

MATERIAL AND METHODS

The data consisted of milk and fat 305d first lactation yield and fat content from 47,574 cows, daughters of 2,504 sires and 43,055 dams. The pedigree file contained 96,606 animals including cows, their parents and grandparents. Cows calved for the first time from 1989 through 1996. There were 371 herds with a minimum of 35 cows, 850 HYS subclasses and 2 seasons of calving: April-September and October-March.

The following linear models was used for multiple trait analysis:

$$\mathbf{y}_i = \mathbf{X}_1\mathbf{b}_{1i} + \mathbf{Z}\mathbf{u}_i + \mathbf{e}_i \quad (1)$$

or

$$\mathbf{y}_i = \mathbf{X}_2\mathbf{b}_{2i} + \mathbf{Z}\mathbf{u}_i + \mathbf{e}_i \quad (2)$$

where $\mathbf{y}_i$ is the vector of observations for trait i (i=1, 2, 3 for milk yield, fat yield and fat content, respectively), $\mathbf{b}_{1i}$, $\mathbf{b}_{2i}$ are vectors of fixed effects, $\mathbf{u}_i$ is the vector of random animal effects and $\mathbf{e}_i$ is the vector of random residual effects. $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{Z}$ are corresponding incidence matrices. The difference between $\mathbf{b}_{1i}$ and $\mathbf{b}_{2i}$ is that

the former contains HYS effects and regressions on age at calving while the latter includes HYS and age at calving classes. The variance-covariance matrices for random effects were defined as

$$H=V\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}=G\otimes A,\quad R=V\begin{bmatrix}e_1\\e_2\\e_3\end{bmatrix}=E\otimes I,\quad V\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=ZHZ'+R \text{ and } Cov(\boldsymbol{u},\boldsymbol{e}')=\boldsymbol{0}$$

where **G** is the 3 × 3 genetic (co)variance matrix, **A** is the numerator relationship matrix, **E** is the 3 × 3 residual (co)variance matrix, **I** is the identity matrix and ⊗ denotes the Kronecker product function for matrices (Searle, 1982).

Age at calving was expressed in days in model (1). Six age groups were created for model (2). The characteristics of the data are given in Table 1.

Variance and covariance components were estimated for all three traits simultaneously, assuming one of the two linear models and using Gibbs Sampling (GS) or Restricted Maximum Likelihood (REML) algorithms. The four resulting combinations of models and procedures were as follows:

TABLE 1

Numbers of records, means ($\bar{x}$) and standard deviations (SD) for 305d milk and fat yield and fat content by year of calving, season of calving and age class

Year of calving	Number of records	Milk yield, kg		Fat yield, kg		Fat content, %	
		$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1989	7655	4205	1018	166.8	42.5	3.97	0.40
1990	4808	4063	936	164.1	39.6	4.04	0.36
1991	9101	4058	1047	163.7	43.8	4.04	0.36
1992	7831	4122	1056	164.0	44.6	3.98	0.35
1993	5168	4327	1125	176.0	48.2	4.07	0.40
1994	5700	4411	1202	179.0	51.9	4.06	0.42
1995	5132	4803	1218	195.2	51.8	4.07	0.37
1996	2179	4794	1293	196.9	55.8	4.11	0.39
Season of calving							
1 (April-Sept.)	21071	4205	1104	171.2	47.6	4.07	0.38
2 (Oct.-March)	26503	4336	1132	173.3	47.9	4.00	0.38
Age class, months							
1 (21-24)	5061	4132	1030	166.8	43.9	4.04	0.38
2 (25-27)	1777	4284	1132	172.9	47.4	4.04	0.39
3 (28-30)	1431	4329	1136	174.6	48.8	4.03	0.37
4 (31-33)	6919	4314	1134	173.2	49.0	4.01	0.38
5 (34-36)	2681	4191	1097	167.5	47.4	3.99	0.36
6 (37-39)	775	4134	1056	166.0	46.7	4.01	0.37
Total	47574	4278	1122	172.4	47.7	4.03	038

GS-1 - model with regression on age (1)
GS-2 - model with age classes (2)
REML-1 - model with regression on age (1)
REML-2 - model with age classes (2).

The MTGSAM (Multiple Trait Gibbs Sampling in Animal Models) programs of van Tassell and van Vleck (1995) were used for estimation of (co)variance components by the Bayesian approach. Prior distributions were needed to describe the Bayesian model. For the fixed effects, flat priors were used, indicating no prior knowledge of these effects; for (co)variances of genetic and residual effects an inverted Wishart distribution was assumed. There were 20,000 samples generated by the Gibbs Sampler with 1,000 rounds in the burn-in period. The REML estimates of genetic parameters were obtained using the MTC program of Misztal (1994) with an assumed convergence criterion of 10^{-8}. Standard errors of heritabilities were estimated using van Raden's algorithm (Misztal, 1994), standard errors of correlations were not calculated.

RESULTS AND DISCUSSION

There were 316 and 349 rounds of Gauss-Seidel iterations for GS-1 and GS-2, respectively, followed by 20,000 Gibbs samples generated for each model. In the case of REML-1 and REML-2, estimates of (co)variances were obtained after 101 and 61 iterations.

Estimates of genetic and residual (co)variances obtained by the GS and REML methods are given in Tables 2 and 3. The variance components for animal effect

TABLE 2

Variance components (on diagonal) and covariance components (above diagonal) for first lactation milk yield, fat yield and fat content of Polish Black-and-White cows estimated by Gibbs Sampling (GS) using the model with regression on age (1) or with age classes (2)

Model	GS-1			GS-2		
Trait	Milk, kg	Fat, kg	Fat, %	Milk, kg	Fat, kg	Fat, %
Genetic component						
Milk, kg	134472.6	4206.1	-27.69	133784.3	4186.2	-27.57
Fat, kg		200.8	0.74		199.4	0.73
Fat, %			0.047			0.046
Residual component						
Milk, kg	339717.9	13136.4	-9.32	340715.9	13173.2	-9.36
Fat, kg		636.7	2.27		638.7	2.28
Fat, %			0.058			0.058

TABLE 3

Variance components (on diagonal) and covariance components (above diagonal) for first lactation milk yield, fat yield and fat content of Polish Black-and-White cows estimated by Restricted Maximum Likelihood (REML) using the model with regression on age (1) or with age classes (2)

Model	REML-1			REML-2		
Trait	Milk, kg	Fat, kg	Fat, %	Milk, kg	Fat, kg	Fat, %
Genetic component						
Milk, kg	111733.4	3283.5	-27.94	111207.0	3261.2	-27.97
Fat, kg		154.8	0.51		153.8	0.51
Fat, %			0.038			0.038
Residual component						
Milk, kg	356158.9	13805.9	-8.84	357171.8	13850.3	-8.76
Fat, kg		670.4	2.44		672.4	2.44
Fat, %			0.065			0.065

obtained by the GS method were 20-30% higher and the residual components 5-10% lower than the corresponding REML estimates. The covariances between negatively correlated traits (milk and fat %) obtained by the two methods were very similar. Also, there were no differences in the magnitude of variances and covariances estimated from the two models, indicating that they describe the analysed traits similarly.

The estimates of heritabilities and genetic and phenotypic correlations for milk production traits are presented in Tables 4 and 5. The GS estimates of heritabilities for milk yield were 0.28 and 0.29, depending on the model, slightly higher than those estimated by REML (0.24). The heritabilities of fat yield (0.24 for GS and 0.19 for REML) and of fat % (0.44-0.45 for GS and 0.37 for REML) were also higher when GS was used as the method of estimation. The simple differences between the magnitudes of parameters estimated by GS and REML were small, only 0.05, on average (Tables 4 and 5). The genetic variance compo-

TABLE 4

Heritabilities (on diagonal), genetic correlations (above diagonal) and phenotypic correlations (below diagonal) for milk yield, fat yield and fat content estimated by Gibbs sampling (GS-1) and REML (REML-1) for model with regression on age (1)

Model	GS-1			REML-1		
Trait	Milk, kg	Fat, kg	Fat, %	Milk, kg	Fat, kg	Fat, %
Milk, kg	0.28	0.81	-0.33	0.24	0.79	-0.43
Fat, kg	0.87	0.24	0.27	0.89	0.19	0.21
Fat, %	-0.16	0.32	0.45	-0.16	0.32	0.37

TABLE 5

Heritabilities (on diagonal), genetic correlations (above diagonal) and phenotypic correlations (below diagonal) for milk yield, fat yield and fat content estimated by Gibbs sampling (GS-2) and REML (REML-2) for model with age classes (2)

Model	GS-2			REML-2		
Trait	Milk, kg	Fat, kg	Fat, %	Milk, kg	Fat, kg	Fat, %
Milk, kg	0.29	0.81	-0.36	0.24	0.79	-0.43
Fat, kg	0.87	0.24	0.22	0.89	0.19	0.21
Fat, %	-0.17	0.32	0.44	-0.16	0.32	0.37

nents obtained by GS were larger than those obtained by REML (Table 2); thus the latter gave smaller heritabilities. The biggest differences between estimated heritabilities were found for the trait with high heritability, i.e. fat content. Within the same method of estimation the magnitude of heritabilities was not influenced by how the age effects were defined in the model.

The genetic correlations estimated by REML were slightly lower than the GS estimates. The biggest difference was found between correlations for milk yield and fat content. The phenotypic correlations were the same no matter which method and model were used. Milk and fat yields were highly correlated (genetic correlation 0.79-0.81 and phenotypic correlation 0.87-0.89), whereas the correlation between fat % and fat yield was low (genetic correlation 0.21-0.27 and phenotypic correlation 0.32), implying that selection for fat % would not increase fat yield. An increase in fat yield could be achieved by direct selection or through selection for milk yield. The genetic correlation between fat content and milk yield was rather low and negative (-0.33 to -0.43). The phenotypic correlation between those two traits was even lower and also negative (-0.16 to -0.17).

All estimates were similar to those cited in the literature (Żuk et al., 1981; Chauhan and Hayes, 1991; Visscher and Thompson, 1992; Jamrozik and Żarnecki, 1993). The h^2 values for fat yield were lower than for milk yield but still within the range of values reported by other authors (Meyer, 1985; Jamrozik and Żarnecki, 1993). Genetic (0.72-0.76) and phenotypic (0.73-0.81) correlations between milk and fat yields were slightly higher than those obtained by Cue et al. (1987) or Visscher and Thompson (1992) but lower than reported by Żuk et al. (1981) and Jamrozik and Żarnecki (1993). Meyer (1985) found values of 0.76, -0.39 and 0.30 for genetic correlations between milk and fat yields, milk yield and fat content, and fat yield and fat content, very similar to the figures in the present study. A genetic correlation of 0.56 between fat and fat content, a lower genetic correlation of 0.45 between milk and fat yield and a high heritability of 0.65 for fat content were found by Chauhan and Hayes (1991). Their results differed from the genetic parameters estimated for the Polish population.

The standard errors of the heritability estimates were less than 0.001 for GS and less or equal to 0.01 for REML method. These errors made up less than 0.40% of the GS estimated values and were 2.5 to 3% of all REML estimates. In the case of GS the standard errors of heritabilities were calculated on the basis of generated samples, while for the REML method they were approximated.

The errors of all estimates derived from GS were significantly smaller than those from REML, giving more accurate estimates. Van Tassell et al. (1995) compared the standard errors of estimates obtained by these two methods and concluded that GS gave smaller standard errors than REML because of the impact of the prior distribution of variance components. This difference will decrease when the amount of data increases. They also stated that GS may have advantages over REML for large data sets because it allows calculation of parameters without approximations or normality assumptions. Jensen et al. (1994) pointed out that REML analysis includes only an approximation of standard errors of estimates, while GS yields the full marginal posterior distribution permitting direct computation of standard errors as well as many other features of this distribution. Wang et al. (1994) stated that an important advantage of GS is that it provides all parameter estimates always within the permissible parameter space. A serious problem of REML analysis is that estimates can be outside of the parameter space.

This study found that genetic parameters could be estimated using a model with age classes or a model with regression on age, alternatively. When the GS and REML estimation methods were compared, the application of Gibbs Sampling (GS) proved more desirable because of the smaller errors of the heritability estimates and because we could directly calculate not only the point characteristics but also the confidence intervals or any other functions of the posterior distribution of the genetic parameters (Gianola et al., 1991; van Tassell et al., 1995). On the other hand, GS is much more time-consuming than REML and the values of the parameters depend on priors. Van Tassell et al. (1995) showed an increase in the bias of GS estimates when the prior value of h^2 was assumed above the true value. They stated also that the dependence on priors lessens for highly heritable traits. The power of Bayesian methods in animal breeding applications might be evident with more complicated problems solved only approximately by traditional methods.

REFERENCES

Boldman K.G., van Vleck L.D., van Tassell C.P., Kachman S.D., 1995. Manual for the Use of MTDFREML. A Set of Programs to Obtain Estimates of Variances and Covariances. USDA, Clay Center, NE

Casella G., George E.I., 1992. Explaining-the Gibbs sampler. Amer. Statist. 46, 167-174

Chauhan V.P.S., Hayes J.F., 1991. Genetic parameters for first lactation milk production and composition traits for Holsteins using multivariate restricted maximum likelihood. J. Dairy Sci. 74, 603-610

Cue R.I., Monardes H.G., Hayes J.F., 1987. Correlations between production traits in first lactation Holstein cows. J. Dairy Sci. 70, 2132-2137

Gelfand A.E., Smith A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398-409

Gianola D., Fernando R.F., 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63, 217-244

Gianola D., Foulley J.L., Fernando R.L., 1991. Bayesian estimation of genetic parameters. Abstracts of the 42th EAAP Meeting, Berlin (Germany), p. 48

Hartley H.O., Rao J.N., 1967. Maximum likelihood estimation for the mixed model analysis of variance model. Biometrika 54, 93-108

Jamrozik J., Żarnecki A., 1993. Genetic parameters of Polish Black-and-White cattle for milk yield and milk constituents. Prace Mat. Zoot. 44, 45-50

Jensen J., Madsen P., 1992. A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models. Mimeo. The National Institute of Animal Science, Foulum (Denmark)

Jensen J., Wang C.S., Sorensen D.A., Gianola D., 1994. Bayesian inference on variance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. Acta Agr. Scand., Sect. A 44, 193-201

Meyer K., 1985. Genetic parameters for dairy production of Australian Black and White cows. Livest. Prod. Sci. 12, 205-219

Meyer K., 1988. DFREML, a set of programs to estimate variance components under an individual animal model. J. Dairy Sci. 71, Supl.2, 33

Meyer K., 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. Genet. Sel. Evol. 23, 67-83

Misztal I., 1994. MTCAFS (MTC) - Multitrait REML Estimation of Variance Components Program by Canonical Transformation with Support for Multiple Random Effects. University of Georgia, Athens, GA

Patterson H.D., Thompson R., 1971. Recovery of interblock information when block sizes are unequal. Biometrika 54, 545-554

Searle S.R., 1982. Matrix Algebra Useful for Statistics. John Wiley and Sons, Inc., New York, NY

Van Tassell C.P., van Vleck L.D., 1995. A Manual for Use of MTGSAM. A Set of Programs to Apply Gibbs Sampling to Animal Models for Variance Component Estimation. US Department of Agriculture, Agricultural Research Service

Van Tassell C.P., Casella G., Polłak E.J., 1995. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. J. Dairy Sci. 78, 678-692

Visscher P.M., Thompson R., 1992. Univariate and multivariate parameter estimates for milk production traits using an animal model. I. Description and results of REML analyses. Genet. Sel. Evol. 24, 415-430

Wang C.S., Rutledge J.J., Gianola D., 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. Genet. Sel. Evol. 25, 41-62

Wang C.S., Rutledge J.J., Gianola D., 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. Genet. Sel. Evol. 26, 91-115

Żuk B., Nowicki B., Szyszkowski L, Filistowicz A., Zwolińska-Bartczak I., 1981. Genetic parameters of milk features of cattle in south-western Poland. I. Heritability and genetic and phenotypical correlations. Rocz. Nauk rol., B-100, 7-22

STRESZCZENIE

Parametry genetyczne cech wydajności mlecznej bydła oszacowane przy pomocy metody REML i próbkowania Gibbsa

Do obliczeń posłużyły dane o 305-dniowych wydajnościach mleka i tłuszczu oraz zawartości tłuszczu w mleku 47574 krów, cielących się po raz pierwszy w latach 1989-1996 i pochodzących po 2504 ojcach i 43055 matkach. Wybrano jedynie stada, w których było co najmniej 35 krów. Parametry genetyczne oszacowano przy pomocy dwóch metod: próbkowania Gibbsa (GS - Gibbs Sampling) oraz największej wiarogodności z ograniczeniem (REML- Restricted Maximum Likelihood). Metody te zastosowano dla dwóch trzycechowych modeli liniowych, w których uwzględniono stały efekt stada x wieku x sezonu ocielenia (HYS), losowy efekt zwierzęcia i błędu oraz regresję na wiek ocielenia (model 1) lub stały efekt klasy wieku (model 2).

Odziedziczalność wydajności mleka (w kg) wynosiła 0,28-0,29 (GS) oraz 0,24 (REML), nieco mniejsza była odziedziczalność wydajności tłuszczu (w kg) i równała się 0,24 (GS) oraz 0,19 (REML). Dla zawartości tłuszczu w mleku (w %) odziedziczalność wyniosła odpowiednio 0,44-0,45 (GS) oraz 0,37 (REML). Największa różnica między oszacowaniami tego parametru wystąpiła w procentowej zawartości tłuszczu, która była najwyżej odziedziczalną cechą.

Korelacje genetyczne oszacowane metodą REML były nieco niższe niż uzyskane metodą GS. Wydajności mleka i tłuszczu (w kg) były wysoce skorelowane (od 0,79 do 0,81), podczas gdy korelacja genetyczna między wydajnością mleka i procentową zawartością tłuszczu była niska i ujemna (od -0,33 do -0,43). Korelacja genetyczna między wydajnością tłuszczu (w kg) i zawartością tłuszczu (w %) była również niska, ale dodatnia, i wynosiła od 0,21 do 0,27, w zależności od metody i modelu liniowego.

Odziedziczalności i korelacje genetyczne były większe gdy zastosowano metodę próbkowania Gibbsa, natomiast wartości korelacji fenotypowych nie zależały od użytej metody szacowania komponentów (ko)wariancji. Nie stwierdzono różnic między oszacowaniami parametrów genetycznych otrzymanymi przy pomocy obydwóch modeli (z regresją na wiek ocielenia lub z klasami wieku).