Assessment of Homogeneity vs. Heterogeneity of Residual Variance in Random Regression Test-Day Models in a Bayesian Analysis

P. López-Romero,* R. Rekaya,† and M. J. Carabaño*

*Departamento de Mejora Genética Animal. INIA.
Carretera de A Coruña km 7. 28040 Madrid, Spain

†Department of Animal and Dairy Science
University of Georgia, Athens 30602

ABSTRACT

Test-day first-lactation milk yields from Holstein cows were analyzed with a set of random regression models based on Legendre polynomials of varying order on additive genetic and permanent environmental effects. Homogeneity and heterogeneity of residual variance, assuming three and 30 arbitrary measurement error classes of different length were considered. Unknown parameters were estimated within a Bayesian framework. Bayes factors and a checking function for the cross-validation predictive densities of the data were the tools chosen for selecting among competing models. Residual variances obtained from 30 arbitrary intervals were nearly constant between d 70 and 300 and tended to increase towards the extremes of the lactation, especially at the onset. In early lactation, the temporary measurement errors were found to be larger and highly variable. A high order of the regression submodels employed for modeling the permanent environmental deviations tended to strongly correct the heterogeneity of the residual variance. Accordingly, the assumption of homogeneity of residual variance was the most plausible specification under both comparison criteria when the number of random regression coefficients was set to five. Otherwise, the heterogeneity assumption, using three or 30 error classes, was better supported, depending on the criterion and on the order of the submodel fitted for the permanent environmental effect.

(Key words: random regression test-day models, heterogeneity of residual variance, Bayesian analysis)

Abbreviation key: AG = additive genetic, BF = Bayes factor, DGV = daily additive genetic variance, DPV = daily permanent environmental variance, ESS = effective

sample size, IRV = intervals of residual variance, IW = inverse Wishart distribution, LMD = log marginal density of the data, MCV = Monte Carlo variance, MDD = marginal density of the data, MVN = multivariate normal distribution, PE = permanent environmental, RPT = repeatability model, RRC = random regression coefficients, RRM = random regression models, RV = residual variance, TD = test day.

INTRODUCTION

In a previous analysis, 22 alternative random regression models (RRM) for the genetic evaluation of first lactation test-day (TD) production traits for the Spanish Holstein-Friesian population were investigated using REML and different classical model comparison criteria (López-Romero and Carabaño, 2003). Models fitting at least three coefficients for the additive genetic (AG) effects and five for the permanent environmental (PE) effects were deemed necessary. All the models assumed constant residual variance (RV) along DIM. However, heterogeneous RV has been found in the course of lactation in several studies (Jamrozik et al., 1996, 1997; Olori et al., 1999a, 1999b; Brotherstone et al., 2000; Rekaya et al., 2000). The heterogeneity of RV is related to the lactation stage and is larger at the extremes of the lactation. This is likely due to a set of nonspecified factors in the model equation (days open, pregnancy status, characteristics of the dry period, body condition at calving, etc.) that make the temporary measurement errors larger and highly variable at the beginning and at the end of the lactation. The inclusion of all of these effects in the model equation might be difficult, mainly because of the lack of information.

Olori et al. (1999b) stated that the assumption of homogeneity of RV along the lactation leads to biases in the RV estimates in early lactation, but does not have a significant influence on the rest of the variance components. These authors conveyed that this could have consequences on the estimates of the total variance and hence in the heritability. Rekaya et al. (2000)

Received October 1, 2002.

Accepted March 14, 2003.

Corresponding author: P. Lopez-Romero; e-mail: plomero@cnb.uam.es.

also mentioned that assuming a homogeneous RV would directly affect the genetic evaluation through the different weights assigned to the information depending on the stage of the lactation. If homogeneity of RV is assumed, the information coming from intervals where the RV is actually larger than the assumed homogeneous value would be given more weight than it really has, and vice versa.

Several approaches can be considered to account for the heterogeneity of the RV along the lactation. The use of continuous functions allows description of the heterogeneity of measurement errors using a potentially small number of parameters and description of the changes of the RV during the lactation in a continuous way, gradually and relatively smoothly, as it is naturally produced. Simpler functions for RV than for the rest of components of variance could be used given that covariances among errors at different times are expected to be null (Olori et al., 1999a). However, it is not obvious what kind of function should be used. One choice would be to assign a function to the RV in a random regression approach, yet random regressions have been found to provide values that are somewhat large at the extremes of the trajectories when it comes to defining the AG and PE daily variances. Most likely this arises as a consequence of 'artefacts' of the random regression (Van der Werf et al., 1998; Rekaya et al., 1999; Misztal et al., 2000; Lopez-Romero and Carabaño, 2003). Jaffrezic et al. (2000) and Schnyder et al. (2001) have used a log link function to model RV, arguing that the RV is likely to change due to scale effects. Rekaya et al. (2000) implemented the change point identification technique (Stephens, 1994), which allows the estimation of the RV in a continuous fashion. This technique also allows inference of the times at which changes in the RV trajectory take place given the number of change points.

The use of continuous functions usually involves complex computations and the need to define a predetermined submodel for the RV. Alternatively, several authors have used different numbers of discontinuous arbitrary intervals of RV (IRV) with homogeneous RV within each interval and heterogeneity among them (Jamrozik et al., 1997; Olori et al., 1999b; Brotherstone et al., 2000). This approach is easy to implement and can be useful in the sense that it provides information on the expected pattern of RV changes along the lactation, as a previous step prior to the use of continuous functions. Caution should be taken when defining the different arbitrary classes of heterogeneity of RV because the proposed model might not be correct, if the identification of the IRV with a similar RV pattern is not accurate. Olori et al. (1999a) reported that a correct identification of the measurement error classes is more

important than the number of intervals. Shorter intervals seem to be needed at the beginning and end of the lactation because of the large variability of the temporary measurement errors found during those periods (Jamrozik et al., 1997; Olori et al., 1999a, 1999b; Brotherstone et al., 2000).

In this paper, some of the best performing RRM used in a previous study (López-Romero and Carabaño, 2003) will be compared, allowing heterogeneous RV with different number of IRV in order to assess the assumption of heteroskedasticity of the RV. In the previous study, goodness of fit for competing models was assessed using statistics based on the likelihood value and on the estimated residuals for each observation. Predictive ability was judged using simple cross-validation techniques based on the prediction of TD records excluded at random for each animal. All criteria tended to select the most complex model except the Bayesian information criteria, which uses a penalized likelihood in which the penalty term is a function of the number of unknowns and observations. In this paper, the major objective is to compare the competing models within a Bayesian framework. The use of alternative model selection criteria is a secondary objective.

MATERIALS AND METHODS

Data

A dataset composed of 47,472 TD milk records corresponding to 4362 complete first lactations (more than 10 controlled test days) of Spanish Holstein-Friesian cows recorded from 1988 to 1998 was used in this study. These records were from herds included in the national database that were randomly selected to obtain a computationally manageable file. Data editing for the intervals between TD records in successive controls was based on the Spanish official milk recording system, where an interval between 5 and 37 d from parturition to the first TD and intervals between 26 and 33 d for the successive controls are required. The Spanish milk recording scheme extends these intervals up to 67 d during holiday periods. In the analyzed data, only four first TD were recorded after 50 d. Thus, very little influence in the parameter estimates is expected to arise from the first TD being recorded after peak production had been reached. Data after 335 DIM were discarded; and only records from cows with age at calving between 18 and 40 mo of age were used. The pedigree file consisted of 10,688 animals, including animals that have genetic ties to recorded animals within the past two generations. Table 1 has means and standard deviations for DIM, milk production, and numbers of TD records within successive 30-d periods.

Table 1. Mean and standard deviation (SD) for DIM and milk production and number of test-day records (TD) in successive periods of 30 d length during lactation.

Period	DIM		Milk yield (kg)		No. TD
	Mean	SD	Mean	SD	
1	18.04	8.28	25.33	6.50	4544
2	48.87	7.95	28.45	6.46	4284
3	79.20	7.92	28.80	6.77	4315
4	109.44	7.87	28.16	6.71	4252
5	139.69	7.91	27.38	6.78	4275
6	169.96	7.81	26.68	6.87	4285
7	200.18	7.75	25.85	6.89	4270
8	230.58	7.85	24.74	6.64	4305
9	260.82	7.87	23.88	6.54	4278
10	291.09	7.91	22.60	6.58	4313
11	321.56	8.00	21.18	6.77	4351

Methods

Three different RRM based on the normalized series of Legendre polynomials (Kirpatrick et al., 1990) of varying number of random regression coefficients (**RRC**) to describe the trajectory of the AG and PE effects along the lactation were used. Homogeneity and heterogeneity of RV in three and 30 arbitrary IRV along the lactation were considered.

Models. The general equation for all models analyzed was the following:

$$y_{ijkl} = \text{HTD}_i + \sum_{m=1}^5 \beta_{jm} X_m(t) + \sum_{m=1}^r \alpha_{km} Z_{a,m}(t) + \sum_{m=1}^s \omega_{km} Z_{p,m}(t) + e_{ijkl},$$

where, y_{ijkl} is the l th TD record of the cow k , HTD_i is the i th level of the Herd-TD effect (4767 levels), β_{jm} is the m th systematic regression coefficient associated with the m th covariate, $X_m(t)$ depending on $\text{DIM} = t$, of the Ali and Schaeffer (1987) function, which describes average lactation curves nested to the j th class of age-season of calving effect (seven levels); the terms α_{km} and ω_{km} are the RRC describing the AG and PE deviations, respectively, along with the covariates $Z_{a,m}(t)$ and $Z_{p,m}(t)$. The number of RRC, r and s , will be referred as the order of fit for the AG and PE effects, respectively. Finally, e_{ijkl} is the temporary measurement error, which is assumed to be independently normally distributed.

To analyze the RV structure along DIM, the heterogeneity of the RV was first analyzed using 30-IRV of different length. The time span for each interval ranged from 5 to 30 d, with shorter intervals at the extremes of the lactation when the measurement errors are highly variable. Details of the IRV characteristics for the 30-

IRV are shown in Table 2. From the RV structure observed in the 30-IRV analysis, an attempt of reducing the number of IRV was carried out assuming only three IRV. The three IRV were defined from 5 to 75 DIM (10,414 records), from 76 to 265 DIM (27,025 records) and from 266 to 335 DIM (10,033 records).

The notation that will subsequently be used to refer to the alternative RRM is $L(i,j)$, where i and j indicate the order of fit for the AG and PE effects, respectively. Based on previous results (López-Romero and Carabaño, 2003), only models $L(3,3)$, $L(3,5)$ and $L(5,5)$ were investigated. A repeatability model (**RPT**) defined as $L(1,1)$ was also investigated as a baseline or reference model.

The data were assumed to be generated from a multivariate normal distribution (**MVN**), according to the stochastic process:

$$\mathbf{y} \mid \mathbf{b}, \mathbf{a}, \mathbf{p}, \sigma_{ei}^2 (i=1, \dots, \delta) \sim \text{MVN}(\mathbf{Xb} + \mathbf{Za} + \mathbf{Wp}, \mathbf{R}),$$

where $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{W}$ are the corresponding incidence matrices, $\mathbf{b}$ is the vector containing the systematic effects previously defined, $\mathbf{a}$ is the vector that contains r RRC fitted to the AG effect for all animals, $\mathbf{p}$ is the vector that includes s RRC associated with the PE effect for all cows with records, $\mathbf{e}$ is the vector containing the residual terms, and $\mathbf{R}$ is the error (co)variance matrix of order $n = \text{number of records}$, whose diagonal elements, $\sigma_{ei}^2 (i=1, 2, \dots, \delta)$, with δ as the number of IRV considered in the model, are the variances of the residual term for all of the observations measured in the i th interval.

Conjugate priors were used for location and scale parameters and proper-vague priors were employed in order to both assign low degree of belief to the prior information and to allow the computation of the marginal density of the data and the Bayes factor (**BF**), which is not well defined for improper priors (Gelman

Table 2. Dry matter intake for the upper limit (UL) of the interval of residual variance (IRV) and number of test-day records (TD) per interval when 30 IRV were assumed in the model. The lactation period spanned from 5 to 335 DIM.

IRV	UL	No. TD	IRV	UL	No. TD	IRV	UL	No. TD
1	10	1037	11	70	1130	21	230	4217
2	15	912	12	80	1725	22	260	4200
3	20	852	13	90	1434	23	280	2716
4	25	779	14	100	1080	24	290	1548
5	30	562	15	110	1695	25	300	1628
6	35	402	16	120	1466	26	310	1109
7	40	794	17	130	1077	27	320	1517
8	45	862	18	150	3182	28	325	851
9	50	850	19	170	2713	29	330	794
10	60	1381	20	200	4219	30	335	740

et al., 1995). For the location parameters, MVN distributions were assumed:

$$\mathbf{b}|\xi \sim \text{MVN}(\mathbf{0}, \mathbf{I}\xi); \mathbf{a} | \Sigma_a, \mathbf{A} \sim \text{MVN}(\mathbf{0}, \mathbf{G});$$

$$\mathbf{p} | \Sigma_p \sim \text{MVN}(\mathbf{0}, \mathbf{P}),$$

where ξ is a positive, scalar hyper parameter and large enough to give small weight to prior information (more precisely, $\xi = 10^6$); $\mathbf{G} = \Sigma_a \otimes \mathbf{A}$ and $\mathbf{P} = \Sigma_p \otimes \mathbf{I}$ are the AG and PE (co)variance matrices, respectively; $\mathbf{A}$ is the relationship matrix and Σ_a and Σ_p are the matrices that contain the (co)variances among the AG and PE RRC, respectively.

For the scale parameters, scaled inverse chi-squared and inverse Wishart (**IW**) distributions were used:

$$\sigma_{ei}^2 | \nu_{ei}, S_{ei}^2 \sim \chi^{-2}(\nu_{ei}, \nu_{ei} S_{ei}^2), i = 1, \dots, \delta$$

$$\Sigma_a | \nu_a, \mathbf{S}_a^2 \sim \text{IW}(\nu_a, \nu_a \mathbf{S}_a^2); \Sigma_p | \nu_p, \mathbf{S}_p^2 \sim \text{IW}(\nu_p, \nu_p \mathbf{S}_p^2),$$

where parameters ν_i and S_i are interpreted as degrees of belief and a priori values for the residual variances in every IVR and for the (co)variance matrices of AG and PE effects. A value of four was assigned to the degrees of belief for the χ^{-2} distributions. Degrees of belief in the IW distributions were 12, being always larger than the respective matrix dimensions in order to avoid degenerate forms (Gelman et al., 1995). Values for the scalars S_{ei}^2 and the matrices $\mathbf{S}_a^2$ and $\mathbf{S}_p^2$ were obtained from previous REML estimates (López-Romero and Carabaño, 2003).

Marginal inferences on parameters of interest were drawn from their corresponding conditional posterior distribution through a Gibbs sampling scheme.

Joint posterior and conditional posterior distributions. The joint posterior density of the parameters was obtained in terms of proportionality as the product of the sampling distribution and the joint prior density. Assuming independence among the prior distributions,

the joint posterior distribution can be written in terms of proportionality as:

$$p(\mathbf{b}, \mathbf{a}, \mathbf{p}, \Sigma_a, \Sigma_p, \sigma_{ei}^2, i=(1, \dots, \delta) | \mathbf{y})$$

$$\propto |\mathbf{R}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{e}' \mathbf{R}^{-1} \mathbf{e}\right) |\mathbf{I}\xi|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{b}' (\mathbf{I}\xi)^{-1} \mathbf{b}\right)$$

$$|\Sigma_a \otimes \mathbf{A}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{a}' (\Sigma_a \otimes \mathbf{A})^{-1} \mathbf{a}\right)$$

$$|\Sigma_p \otimes \mathbf{I}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{p}' (\Sigma_p \otimes \mathbf{I})^{-1} \mathbf{p}\right) \quad [1]$$

$$|\Sigma_a|^{-\frac{(\nu_a+r+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left[\Sigma_a^{-1}(\nu_a \mathbf{S}_a^2)^{-1}\right]\right)$$

$$|\Sigma_p|^{-\frac{(\nu_p+s+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}\left[\Sigma_p^{-1}(\nu_p \mathbf{S}_p^2)^{-1}\right]\right)$$

$$\prod_{i=1}^{\delta} \frac{1}{(\sigma_{ei}^2)^{\left(\frac{\nu_{ei}}{2}+1\right)}} \exp\left(-\frac{\nu_{ei} S_{ei}^2}{2\sigma_{ei}^2}\right).$$

The conditional posterior distributions needed for the implementation of the Gibbs sampling algorithm can be derived easily after some reordering of [1]:

$$\mathbf{b} | \mathbf{y}, \mathbf{a}, \mathbf{p}, \Sigma_a, \Sigma_p, \sigma_{ei}^2, i=(1, \dots, \delta) \sim \text{MVN}(\hat{\mathbf{b}}, (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} + \mathbf{I}\xi^{-1})^{-1}),$$

with

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} + \mathbf{I}\xi^{-1})^{-1} \mathbf{X}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{Z}\mathbf{a} - \mathbf{W}\mathbf{p})$$

$$\mathbf{a} | \mathbf{y}, \mathbf{b}, \mathbf{p}, \Sigma_a, \Sigma_p, \sigma_{ei}^2, i=(1, \dots, \delta) \sim \text{MVN}(\hat{\mathbf{a}}, (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}),$$

with

$$\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{W}\mathbf{p})$$

$$\mathbf{p} | \mathbf{y}, \mathbf{b}, \mathbf{a}, \Sigma_a, \Sigma_p, \sigma_{ei}^2, i=(1, \dots, \delta) \sim \text{MVN}(\hat{\mathbf{p}}, (\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{P}^{-1})^{-1}),$$

with

$$\hat{\mathbf{p}} = (\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{P}^{-1})(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a})$$

$$\sigma_{ei}^2|\mathbf{y}, \mathbf{b}, \mathbf{a}, \mathbf{p}, \Sigma_{\mathbf{a}}, \Sigma_{\mathbf{p}}, \sigma_{ei}^2, (j \neq i) \sim \chi^{-2}(N_i + \nu_{ei}, \nu_{ei}\mathbf{S}_{ei}^2 + \mathbf{e}_i'\mathbf{e}_i),$$

with i and $j = 1, \dots, \delta$.

$$\Sigma_{\mathbf{a}}|\mathbf{y}, \mathbf{b}, \mathbf{a}, \mathbf{p}, \Sigma_{\mathbf{p}}, \sigma_{ei}^2 (i=1, \dots, \delta) \sim \text{IW}(\mathbf{q} + \nu_{\mathbf{a}}, [\mathbf{U}'\mathbf{A}^{-1}\mathbf{U} + (\nu_{\mathbf{a}}\mathbf{S}_{\mathbf{a}}^2)^{-1}]^{-1}),$$

where $\mathbf{q}$ is the number of animals in the pedigree file and $\mathbf{U}_{(q \times r)}$ is the matrix including the r column vectors that contain the solutions of the RRC for the AG effects for the q animals, and,

$$\Sigma_{\mathbf{p}}|\mathbf{y}, \mathbf{b}, \mathbf{a}, \mathbf{p}, \Sigma_{\mathbf{u}}\sigma_{ei}^2 (i=1, \dots, \delta) \sim \text{IW}(\mathbf{c} + \nu_{\mathbf{p}}, [\Omega'\Omega + (\nu_{\mathbf{p}}\mathbf{S}_{\mathbf{p}}^2)^{-1}]^{-1}),$$

where $\mathbf{c}$ is the number of animals with records, and $\Omega_{(c \times s)}$ the matrix containing the s column vectors with the solutions for the RRC of the PE effects for all the animals with records.

The sampling of the position parameters in $\Phi = (\mathbf{b}, \mathbf{a}, \mathbf{p})$ from the above-defined posterior conditional distributions was done on individual elements, Φ_i , following Sorensen (1996):

$$\Phi_i|\Phi_{(i)}, \Sigma_{\mathbf{a}}, \Sigma_{\mathbf{p}}, \sigma_{ei}^2 (i=1, \dots, \delta) \sim N(\hat{\Phi}_i, \text{var}(\Phi_i)), \quad [2]$$

where, Φ_i is the vector of the position parameters excluding the element Φ_i , which is being currently sampled. In [2], $\hat{\Phi}_i = \frac{\text{RHS}_i - \mathbf{C}_{(i)}'\Phi_{(i)}}{\mathbf{C}_i}$ and $\text{var}(\Phi_i) = \frac{1}{\mathbf{C}_i}$, where

RHS_i is the i th element of the vector $\mathbf{RHS} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$, and $\mathbf{C}_{(i)}$ is the i th row of the matrix of the

coefficients $\mathbf{C}$ excluding the i th diagonal element $\mathbf{C}_i$, being

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} + \xi^{-1}\mathbf{I} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \Sigma_{\mathbf{a}}^{-1} \otimes \mathbf{A}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \Sigma_{\mathbf{p}}^{-1} \otimes \mathbf{I} \end{bmatrix}.$$

The specifications for the Gibbs sampling implementation were based on the Raftery and Lewis criterion (1992), which was determined using the public domain program Gibbsit v.2.0 (Raftery and Lewis, 1995). Alternatively, the coupling method analysis (Johnson, 1996; García-Cortés et al., 1998) was performed using two chains of 5000 iterations each. The effective sample size (**ESS**) of the chains and the Monte Carlo variance (**MCV**) estimates were computed following Sorensen (1996).

Genetic and permanent daily variances. The AG and PE daily (co)variances were obtained from the following bivariate continuous function (Jamrozik and Schaeffer, 1997):

$$g(t_i, t_j) = \mathbf{z}(t_i)'\Sigma_{\mathbf{a}}\mathbf{z}(t_j), \quad [3]$$

where, $\mathbf{z}(t_i)$ is the vector of covariates calculated from the submodel fitted to the AG effect at time t_i . If we set $i = j$, a univariate continuous function, representing the change of the genetic variance across the lactation, will be obtained. The same can be applied to the PE effect. The daily AG and PE variances during lactation were computed from [3] from 5 to 335 DIM.

Model comparison. The BF (Newton and Raftery, 1994; Kass and Raftery, 1995) and the cross-validation predictive densities of the data (Gelfand et al., 1992) were computed to assess the performance of the alternative models.

The BF represents the evidence in favor of one model provided by the data. For two competing models, M_1 and M_2 , BF is formally defined as the ratio of the posterior ($p(M_i|\mathbf{y})$) odds for M_1 versus M_2 to its prior ($\pi(M_i)$) odds (Kass and Raftery, 1995, Gelfand, 1996). This expression can also be viewed as the ratio of the corresponding marginal densities of the data (**MDD**) for the two competing models:

$$\text{BF}_{12} = \frac{p(M_1|\mathbf{y})/p(M_2|\mathbf{y})}{\pi(M_1)/\pi(M_2)} = \frac{f(\mathbf{y}|M_1)}{f(\mathbf{y}|M_2)}. \quad [4]$$

Assuming that the prior probabilities of the models are the same, the BF is the ratio of the two posteriors distributions of the model. In this case, the BF is an indicator of the probability of a model compared with an alternative model, based on the observed data.

The MDD for a specific model M_k , $f(\mathbf{y}|M_k)$, has to be calculated by integrating over the whole parameter space (Kass and Raftery, 1995):

$$f(\mathbf{y}|M_k) = \int_{\phi} f(\mathbf{y}|\Phi_k, M_k) \pi(\Phi_k|M_k) d\Phi_k. \quad [5]$$

To do so, different approaches have been proposed (Newton and Raftery, 1994; Chib, 1995; Kass and Raftery, 1995; Gamerman, 1997). In this paper, the computation of the MDD was based on the estimator suggested by Newton and Raftery (1994):

$$\hat{\text{MDD}} = \left[\frac{1}{H} \sum_{i=1}^H \frac{1}{f(\mathbf{y}|\Phi^{(i)})} \right]^{-1},$$

where H is the total number of iterations after the burn in period; and $f(\mathbf{y}|\Phi^{(i)})$ is the conditional distribution of

the data given the parameters at iteration i . This result arises when solving [5] by means of an importance sampling, using the joint posterior distribution of the parameters as the importance distribution. This estimator presents a large Monte Carlo error due to the occasional occurrence of samples with small values of likelihood, but it converges to the correct value of $f(\mathbf{y}|\mathbf{M}_k)$ as H tends to infinity (Kass and Raftery, 1995). Additionally, this estimator has considerable computational advantages. The calibration suggested by Raftery (1996) was used to decide which model was more supported by the data. Model M_1 is strongly indicated when $2 \log BF_{12} > 10$.

The cross-validation predictive densities are defined as the set of univariate densities $f(\mathbf{Y}_r|\mathbf{y}_{(r)})$ with $r = 1, \dots, n$, (Gelfand et al., 1992; Gelfand, 1996), where, $\mathbf{y}_{(r)}$ is the vector of observations with data y_r deleted, and $\mathbf{Y}_r$ is a random variable, which may be considered to be a prediction of the observed y_r . The distribution $f(\mathbf{Y}_r|\mathbf{y}_{(r)})$ gives the values of the unobserved $\mathbf{Y}_r$ that would be more likely to be observed when the model has been fitted to $\mathbf{y}_{(r)}$ (Gelfand, 1996). This provides a way of comparing the observed versus the predicted data, resulting in a tool that is useful to assess the predictive ability of the competing models. Model comparisons are based on the computation of the expected value of a checking function, $g(\mathbf{Y}_r, y_r)$, weighted by its predictive density $f(\mathbf{Y}_r|\mathbf{y}_{(r)})$:

$$d_r = E_{\mathbf{Y}_r|\mathbf{y}_{(r)}} [g(\mathbf{Y}_r, y_r)],$$

for all of the data.

The computation of d_r implies the evaluation of the integral

$$d_r = \iint g(\mathbf{Y}_r, y_r) f(\mathbf{Y}_r|\boldsymbol{\Phi}, \mathbf{y}_{(r)}) \pi(\boldsymbol{\Phi}|\mathbf{y}_{(r)}) d\boldsymbol{\Phi} d\mathbf{Y}_r, \quad [6]$$

which can be approximated by the corresponding Monte Carlo estimate, $\hat{d}_r = \frac{1}{H} \sum_{i=1}^H g(\mathbf{Y}_r^{(i)}, y_r)$. However, the estimation of [6] in this way requires a Gibbs sampling scheme to be run for every single observation in the dataset to compute the $\hat{d}_r$ values. Thus, this procedure is not feasible, even for a reduced dataset. To overcome this problem, an importance sampling scheme suggested by Gelfand et al. (1992), using the joint posterior density of the parameters as an importance distribution, was implemented to evaluate [6] instead. In this fashion, $\hat{d}_r$ is obtained for all data directly from the outputs of the Gibbs sampling scheme implemented to draw the marginal posterior densities of the parameters. In this study, the checking function used to measure the discrepancy between the observed and predicted data was

$g(\mathbf{Y}_r, y_r) = \mathbf{Y}_r - y_r$. The best model was identified as the one with the minimum $D = \frac{1}{n} \sum_{i=1}^n d_r^2$.

RESULTS AND DISCUSSION

Pre-Gibbs Analysis

Table 3 has the results of convergence obtained by the Raftery and Lewis (1992) criterion and by the coupling method (Johnson, 1996; García-Cortés et al., 1998) and the total number of iterations required according to the Raftery and Lewis (1995) criterion for the AG and PE (co)variance components in model L(3,3) with three classes of RV. The number of iterations to be discarded in accordance with the Raftery and Lewis (1995) criterion was much smaller than that indicated by the coupling method. For G_{22} , the more demanding component, the Raftery and Lewis (1995) criterion indicates that convergence was reached after 277 and 423 iterations, for the 2.5 and 97.5 percentiles, respectively, versus the 4012 iterations, indicated by the coupling method. Similar results dealing with the comparison of these two methods were observed by García-Cortés et al. (1998) and Hernández et al. (1999). The method of Raftery and Lewis, based on the results obtained from a single chain, seems to produce less reliable results than multiple chain methods with respect to the burn-in period estimation (Gelman and Rubin, 1992; García-Cortés et al., 1998). Both criteria indicated that a much smaller burn-in period was needed for the RV components in all models analyzed. The total chain length according to the Raftery and Lewis criterion (1995) ranged from 6696 to 48,101 for the AG components and from 3100 to 39,138 iterations for the PE components. More complex models (not shown) did not necessarily require larger burn-in periods or larger total chain length. In fact, component G_{23} under homogeneity of RV was the parameter showing the largest requirements, demanding a total chain length of 44,064 iterations. Conservatively, single chains of 100,000 iterations for the RRM and 32,000 iterations for the RPT were run. The number of iterations discarded was 10,000 for all the RRM and 2000 for the RPT.

Parameter Estimates

Posterior features of the (co)variances among RRC for the AG and PE effects, as well as the MCV, ESS, and intrachain autocorrelation for model L(3,3) with three IRV are in Table 4. All posterior distributions (not shown) were unimodal symmetric densities. The autocorrelation was stronger for the AG than for the PE RRC, resulting in a smaller ESS and in a larger

Table 3. Number of iterations to be discarded for the burn-in period based on the Raftery and Lewis criterion (NID_{RL}) and on the coupling method (NID_{CM}), and total length of the chains (TLC) for the additive genetic and permanent environmental (co)variance components in a model fitting three coefficient Legendre polynomials for the additive genetic and permanent environmental effects, L(3,3), with three classes of residual variance.

	q = 0.025 ¹		q = 0.975 ¹		NID _{CM} ²
	NID _{RL}	TLC	NID _{RL}	TLC	
G ₁₁ ³	170	31,324	208	38,108	3851
G ₁₂	81	12,706	84	14,928	3706
G ₁₃	40	6696	102	17,904	3911
G ₂₂	277	48,101	423	37,623	4012
G ₂₃	168	26,904	126	20,646	3887
G ₃₃	96	16,740	206	20,744	3805
P ₁₁	423	37,623	15	2913	3912
P ₁₂	160	23,135	18	3114	3885
P ₁₃	30	4405	108	18,860	3850
P ₂₂	423	37,630	20	3100	3824
P ₂₃	24	4734	72	11,964	3790
P ₃₃	430	39,138	32	5196	3863

¹Raftery and Lewis criterion was applied to determine the 2.5th and 97.5th quantiles within ± 0.0125 with a 95.0% probability.

²Number of iterations reached when absolute differences between chains were smaller than 5,00E-5.

³G_{ij} and P_{ij} represent the i,j components of the (co)variance matrices between the additive genetic and permanent environmental random regression coefficients, respectively.

MCV estimate for the AG coefficients. The ESS ranged from 83 to 158 for the AG, and from 108 to 275 for the PE components as a consequence of the strong autocorrelation found between samples. Similar results (not shown) were found for the rest of the models. Smaller ESS for the AG than for the environmental components have also been reported by Schnyder et al. (2001). The ESS observed in this study are within the range of ESS values in other similar applications (Varona et al., 1998; Pool et al., 2000) and seem to be sufficient to provide small MCV.

Daily variances along lactation. Figure 1 shows the posterior mean estimates of the RV along DIM when 30-IRV were assumed for the four groups of models analyzed. The RPT model always showed larger RV values than the RRM. This was more evident in the two extremes of the lactation, particularly during the first 100 d.

As the order of fit for the PE effect was increased, the RV was significantly reduced, mainly in early lactation [models L(3,3) vs. L(3,5) or L(5,5)]. However, this reduction was not observed when the order of fit for the AG

Table 4. Posterior features of marginal posterior distributions of the (co)variances among the random regression coefficients for a model fitting three coefficient Legendre polynomials for the additive genetic and permanent environmental effects, L(3,3), with three classes of residual variance: mean (m) and standard deviation (SD), mode (M), quantiles 2.5, 50, and 97.5%, Monte Carlo variance (MCV), effective sample size (ESS) and autocorrelation between samples with a lag of 100 (L100) and 500 (L500) iterations.

	m	SD	M	Quantiles %			MCV	ESS	L100	L500
				2.5	50.0	97.5				
G ₁₁ ¹	4.144	0.758	3.201	2.751	4.118	5.795	6.86E-3	83	0.795	0.316
G ₁₂	0.202	0.266	0.164	-0.311	0.199	0.707	5.09E-4	133	0.759	0.342
G ₁₃	-0.708	0.243	-0.633	-1.199	-0.685	-0.263	6.07E-4	97	0.760	0.361
G ₂₂	0.835	0.182	0.722	0.519	0.820	1.221	3.52E-4	93	0.755	0.358
G ₂₃	-0.110	0.114	-0.113	-0.332	-0.113	0.124	8.21E-4	158	0.722	0.221
G ₃₃	0.788	0.133	0.803	0.549	0.778	1.075	1.25E-4	142	0.651	0.142
P ₁₁	11.317	0.677	11.083	9.974	11.330	12.608	4.20E-3	108	0.620	0.246
P ₁₂	1.021	0.265	0.873	0.521	1.024	1.535	3.31E-4	209	0.479	0.171
P ₁₃	-1.117	0.223	-1.127	-1.546	-1.127	-0.671	3.88E-4	131	0.549	0.209
P ₂₂	4.067	0.213	4.063	3.654	4.078	4.475	2.38E-4	188	0.463	0.153
P ₂₃	-0.300	0.122	-0.341	-0.548	-0.300	-0.060	5.88E-5	259	0.411	0.141
P ₃₃	1.402	0.135	1.311	1.129	1.405	1.662	6.61E-5	275	0.412	0.086

¹G_{ij} and P_{ij} represent the i,j components of the (co)variance matrices between the additive genetic and permanent environmental random regression coefficients, respectively.

Figure 1. Posterior mean estimates of the residual variance along DIM when 30 intervals are defined for this component under models L(1,1), L(3,3), L(3,5), and L(5,5), fitting Legendre polynomials of varying order for the additive genetic (1 and 3) and permanent environmental (1, 3, and 5) components.

effect was increased [models L(3,5) vs. L(5,5)], although model L(5,3) should have been analyzed as well to be more precise. According to these results, there seems to be a trade-off between the permanent and temporary environmental effects at the beginning and end of lactation. This might be because there are more environmental factors acting at the onset and end of lactation, which are not accounted for in the model. Most of these factors, such as the body condition of the cow at calving, the animal's physiological response to the onset or end of the lactation process, and the pregnancy status, are expected to be linked to the animal. If the order of fit for PE effects is not large enough to adapt to changes in the true trajectory, the variance of the temporary environment may be erroneously overestimated. Moreover, from the results obtained in this study, it seems that more than three coefficients are needed for the PE effects in order to reduce the RV estimates in the extremes of the lactation. Olori et al. (1999a) also observed a decline in the RV estimates as the order of the regression submodels increased. On the other hand, Brotherstone et al. (2000) reported that RV estimates were fairly consistent for the different orders of regression submodels.

The influence of the different PE regression submodels over the variability of the residual terms was also explored by plotting the estimated residuals from the models under study against DIM. Figure 2 shows these plots for models L(1,1) and L(5,5), both under 30-IRV. It can be noticed that the heterogeneity of the RV is almost corrected under the fifth order of fit for the PE effects. The same result (not shown) was observed for the model L(3,5). For model L(3,3) (not shown), a reduction of the heterogeneity of the RV in comparison to

Figure 2. Residual plots for the simple repeatability model, L(1,1) (top), and for a model fitting five coefficient Legendre polynomials for the additive genetic and permanent environmental components, L(5,5), both with 30 intervals of heterogeneity of residual variance.

model L(1,1) was achieved. However, heterogeneity of RV still persisted at the extremes of the lactation.

From Figure 1, three different stages in the RV trajectory can be clearly distinguished for the four groups of models analyzed. Focusing only on models L(3,3), L(3,5), and L(5,5), we observed large RV in early lactation. Then, in mid- and late lactation, approximately from 70 to 310 DIM, smaller and relatively constant values of the RV were observed. Afterwards, beyond 310 DIM, the RV tended to increase for models L(3,3) and L(5,5), and to decrease for model L(5,5), probably linked to the large values of the daily AG variance that was estimated under this model, as will be shown. Similar patterns of the RV have been reported by Olori et al. (1999b) and Brotherstone et al. (2000) when fitting Legendre polynomials or the Ali and Schaeffer (1987) function to the AG and PE components (only in Brotherstone et al., 2000). On the other hand, other authors (Jamrozik et al., 1997; Brotherstone et al., 2000; Rekaya et al., 2000) found a different RV structure at the beginning of the period, obtaining smaller values for the RV in the very early lactation period. Jamrozik et al. (1997) used the Ali and Schaeffer (1987) function as a submodel to define the AG effect and constant PE effects, which produced anomalously large estimates of the AG

Table 5. Posterior mean and standard deviation for the residual variance under the assumption of allowing for heterogeneity of the residual variance across three intervals, for alternative models for the additive genetic and permanent environmental components.

Model ¹	Residual variance interval		
	5–75 d	76–265 d	266–335 d
L(1,1)	21,524 ± 0,342	9,496 ± 9,73E-2	13,705 ± 0,236
L(3,3)	13,358 ± 0,259	8,071 ± 8,53E-2	8,032 ± 0,165
L(3,5)	10,167 ± 0,204	7,202 ± 8,19E-2	6,621 ± 0,166
L(5,5)	10,064 ± 0,237	7,164 ± 8,16E-2	6,421 ± 0,157

¹Model L(i,j) refers to Legendre polynomials with i and j coefficients fitted to the additive genetic and permanent environmental components, respectively.

variance during early lactation. This was probably the cause of the low estimates of RV. Rekaya et al. (2000) used Wood's nonlinear function (Wood, 1967) to define individual trajectories with a AG and PE components and Brotherstone et al. (2000) used the Wilmink's function (Wilmink, 1987) for both the AG and PE components. These contradictory results seem to indicate that accommodating the true variance patterns at the early stages of the lactation is difficult. This might be due to a lower amount of information available in this period, in terms of quantity and in terms of unknown and unaccounted factors, and to a bad behavior of the regression fit at the extremes of the trajectory.

Due to the pattern found for the RV in this and other studies, where a long period of relatively constant RV was observed, a reduction in the number of IRV can be considered. In this study, three IRV, defined according to the three lactation stages observed in the 30-IRV case, were subsequently employed in order to evaluate alternative RRM with a substantially reduced number of parameters. Posterior mean estimates of the RV under three IRV for the four groups of RRM previously proposed are presented in Table 5. For the three IRV estimates, as in the 30-IRV case, the RV was reduced as the order the AG and PE regression submodels was increased, especially when increased for the PE submodel. The midinterval under three IRV was very similar to the average of values estimated in the same period under the 30-IRV assumption. However, except for the RPT model, the RV estimate for the third IRV was not larger than the estimated value for the mid interval, as it would have been expected from the 30-IRV analyses. Because the estimates of RV in the three-IRV case should represent an average of the estimates in the 30-IRV case weighted by the number of data in the corresponding period, it seems that the larger values found at the very end of lactation are not large enough to offset the smaller estimates obtained with more observations elsewhere. This result might indicate that the long midinterval should be enlarged even more, probably until day 310, approximately, and the last

interval reduced to include the part of lactation where increases of RV are more evident.

The daily AG (DGV) and PE (DPV) variances along DIM were calculated from equation [3]. Figure 3 shows the values for the DGV and DPV obtained under model L(5,5) with different numbers of IRV. No differences in the DGV under different numbers of IRV were observed. For the DPV, the alternative definitions of IRV seem to affect this component only in early lactation. The DPV was 18.85 and 24.92 kg² for IRV = 30 and IRV = 1, respectively, at DIM = 5. This difference tended to disappear as the lactation progressed. Similar results (not shown) were observed for models L(3,3) and L(3,5). Therefore, increasing the number of IRV causes no effect on the DGV estimates and tends to decrease the DPV estimates in the early lactation, as expected, since a compensatory effect exists between the two environmental components, as was mentioned earlier. Olori et al. (1999b) found little effect on the estimates of the daily variances associated with both the AG and PE effects under different definitions of IRV, although

Figure 3. Posterior means of the daily additive genetic variance (DGV) (below) and permanent environmental (DPV) variances along DIM for a model fitting five coefficient Legendre polynomials for both components, L(5,5), with one (IRV = 1), three (IRV = 3), and 30 (IRV = 30) error classes.

Figure 4. Posterior means of the daily additive genetic variance (DGV) (below) and permanent environmental (DPV) variances along DIM for models fitting either three or five coefficient Legendre polynomials for those components, L(3,3), L(3,5), and L(5,5), with 30 error classes.

these authors studied data comprised between 4 and 40 wk.

Figure 4 shows the posterior means of the DGV and DPV for models L(3,3), L(3,5), and L(5,5) under 30-IRV. Differences among models are now due to differences in the orders of the regression submodels, which led to quite different trajectories of the daily variances along DIM especially at the extremes of the lactation. Models fitting three coefficients for the AG effect [models L(3,3) and L(3,5)] produced a nearly equal and almost constant trajectory for that component regardless of the order fitted for the PE component. The pattern for the DPV was also flatter when three coefficients were fitted for that effect [model L(3,3)] than when five coefficients were employed [model L(5,5)]. Some trade-off effect in the extremes of lactation was observed between the DGV and DPV when different orders of fit for the regression submodels describing the AG and PE deviations were used. The DPV estimated in the extremes of lactation under the L(5,5) specification was smaller than the DPV estimated by the L(3,5) model due to the greater values of the DGV estimated under the fifth-order AG submodel in the extremes of the lactation. Hence, increasing order of fit for the AG submodel led to a reduction in the DPV at the boundaries of the lactation, indicating a possible confounding between AG and PE factors in those parts of the lactation. Artefacts of RRM might be the cause of the abnormally large values of the AG and PE variances at the extremes of the lactation. The same patterns, with similar values for the DGV and DPV, have been observed and more extensively discussed by López-Romero and Carabaño (2003) in different Spanish datasets analyzed with a wider range of RRM assuming homogeneous RV.

Figure 5. Posterior means of the heritability (h^2) for models fitting either three or five coefficient Legendre polynomials for the additive genetic and permanent environmental components, L(3,3), L(3,5), and L(5,5), with one (above) and 30 error classes.

Figure 5 shows posterior means of the heritabilities estimated for models L(3,3), L(3,5), and L(5,5) under IRV = 1 and IRV = 30. The heritability estimates for a given RRM were practically the same under different RV assumptions. Although different hypotheses of RV along DIM led to different estimates of the RV, especially near the beginning of the lactation, significant and opposed differences for the DPV were found as well in early lactation under the different RV hypotheses (Figure 3). As a consequence of this trade-off between RV and DPV and the apparent lack of effect of the definition of the RV pattern on the DGV, the total variance, and hence the heritability, were hardly affected when different numbers of IRV were postulated in the RRM. The differences in the heritability estimates that can be seen in Figure 5 are mostly due to differences on the AG submodel. According to these results, given the data analyzed and the models explored, ignoring the heterogeneity of RV will have little impact on genetic evaluations when a model fitting more than three coefficients to the PE effect is used.

Table 6. Log of the marginal density of the data (LMD) and D statistic used for the cross-validation predictive densities under different intervals of residual variance (IRV). Best results for each group of models are in bold.

Model ¹	IRV	LMD	D
L(1,1)	1	-128,518	15.378
	3	-127,221	15.263
	30	-126,648	15.266
L(3,3)	1	-120,072	12.881
	3	-120,019	12.803
	30	-120,054	12.791
L(3,5)	1	-115,910	11.837
	3	-116,256	11.987
	30	-116,670	12.011
L(5,5)	1	-115,715	11.651
	3	-115,909	11.896
	30	-116,120	11.932

¹Model L(i,j) refers to Legendre polynomials with i and j coefficients fitted to the additive genetic and permanent environmental components, respectively.

Model Comparison

Table 6 shows the results for the two statistics employed in the model assessment. Both statistics, the log of the MDD (**LMD**) and the D predictive measure, showed a similar behaviour in terms of ranking the competing models. The LMD seems to have a larger ability to discriminate among competing models, always showing very strong evidence in favor of one of the two competing models according to the calibration scale of Raftery (1996). On the other hand, the differences in the D statistic under two alternative models were negligible for some models [e.g., models L(1,1) under three or 30-IRV].

Changes in the submodel applied to the AG or PE components had a much larger impact in the model performance than modifying the RV specification. As expected, results were worst for the RPT models and significantly better for the RRM. Within the RRM, the L(3,3) models were poorest, whereas L(5,5) performed best. Increasing the order of fit for the PE effect resulted in a clear improvement of both statistics employed in the model comparison. However, when increasing the order of fit for the AG effect, the model improvement was smaller. Similar results were observed by López-Romero and Carabaño (2003) using comparison criteria based on the likelihood and classical cross-validation techniques.

The improvement of the plausibility of the models when changes in the RV were allowed was only achieved for RPT and L(3,3) models. This general result agrees with the reduction in the heterogeneity of RV showed for models L(3,5) and L(5,5), as a consequence of increasing the order of the regression submodels employed for the PE components. As it was observed in

Figures 1 and 2, there is a strong heterogeneity of RV under the model L(1,1), so the LMD and D statistic are improved under either assumption of heterogeneous RV in this case. Nevertheless, in models L(3,5) and L(5,5), the heterogeneity of the RV is significantly corrected by PE. Homogeneity of RV is the most plausible specification under both criteria.

The statistical criteria used to measure the plausibility of the models indicated that homogeneity of RV along first lactation might be adequate when using a RRM fitting at least three coefficients for the AG and PE effects. However, the changes of RV observed at the beginning and, to a lesser extent, at the end of lactation, might justify fitting alternative models on the RV. The use of continuous functions, which make more biological sense than discontinuous arbitrary intervals, may be worth investigating in order to find a more parsimonious model. Given the kind of trajectory estimated in this and other studies, a procedure allowing for different patterns at the beginning, mid-, and late lactation, such as the change point identification technique or the use of splines, might be good options for modeling the RV along lactation.

CONCLUSIONS

The RV pattern along the lactation depends largely on the regression submodel fitted to the PE component. Significant reductions were observed in the RV at the beginning and at the end of the lactation when a fifth-order polynomial submodel was fitted to the PE component. However, the order of fit for the AG submodel had little impact on the RV values along lactation. Conversely, changing the assumptions on the RV pattern did not seem to affect the estimates of the daily AG variance and only affected the estimates of daily PE variance in the first part of the lactation (before 50 DIM, approximately). Heritability values followed the AG variance pattern with almost no influence of the RV definition.

From the RV estimates obtained under the different models studied, three stages in the RV pattern were identified. First, there is a period at the beginning of lactation in which the RV tends to be larger at the onset and then decreases until up to 60 to 100 DIM. Then there is a mid- and late-lactation period, up to almost 300 DIM, when RV stays nearly constant. Finally there is a very late-lactation period, where RV tends to increase until the end of lactation.

The statistical criteria used to assess model adequacy indicates that changing the order of fit of the AG and PE components has a larger impact on the plausibility of the models and their predictive ability than different assumptions about the RV patterns. Allowing for het-

erogeneous RV only improves the adequacy of the model to the observed or predicted data for models fitting a low order to the PE submodel. Larger improvements in all the criteria can be expected from increasing the order of PE effects than from increasing the order of AG effects.

ACKNOWLEDGMENTS

This work was supported by a grant of Program Sectorial I+D of the Ministry of Agriculture, Fishery and Food of Spain (SCOO-039). We are grateful to the Spanish Friesian Association (CONAFE) for supplying the data. The first author is also grateful to the Spanish Ministry of Science and Technology for providing financial support.

REFERENCES

- Ali, T. E., and L. R. Schaeffer. 1987. Accounting for covariances among test day milk yield in dairy cows. *Can. J. Anim. Sci.* 67:637–645.
- Brotherstone, S., I. M. S. White, and K. Meyer. 2000. Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. *Anim. Sci.* 70:407–415.
- Chib, S. 1995. Marginal likelihood from the Gibbs output. *J. Am. Statist. Ass.* 90:1313–1321.
- Gamerman, D. 1997. *Markov Chain Monte Carlo*. Chapman and Hall, New York, NY.
- García-Cortés, L. A., M. Rico, and E. Groeneveld. 1998. Using coupling with the Gibbs sampler to assess convergence in animal models. *J. Anim. Sci.* 76:441–447.
- Gelfand, A. E. 1996. Model determination using sampling-based methods. Pages 145–161 in *Markov Chain Monte Carlo in Practice*. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter eds. Chapman and Hall, New York, NY.
- Gelfand, A. E., D. K. Dey, and H. Chang. 1992. Model determination using predictive distributions with implementation via sampling-based methods. Pages 147–167 in *Bayesian Statistics 4*. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds. Oxford University Press, Oxford, UK.
- Gelfand, A. E., and D. K. Dey. 1994. Bayesian model choice: Asymptotic and exact calculations. *J. R. Statist. Soc. B.* 56:501–514.
- Gelman, A., and D. B. Rubin. 1992. A single series from the Gibbs sampler provides a false sense of security. Pages 625–632 in *J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds. Bayesian Statistics IV*. Oxford University Press, Oxford, U.K.
- Gelman, A., B. J. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian data analysis*. Chapman and Hall, New York, NY.
- Hernández, D., P. López-Romero, M. J. Carabaño, and R. Alenda. 1999. Aplicación de la metodología Bayesiana en la estima de parámetros genéticos para la producción de leche. VII Jornadas de Producción Animal. 11–13 Mayo, 1999, Zaragoza (Spain). ITEA Vol. Extra 20:318–320.
- Jaffrezik, F., I. M. S. White, R. Thompson, and W. G. Hill. 2000. A link function approach to model heterogeneity of residual variances over time in lactation curve analysis. *J. Dairy Sci.* 83:1089–1093.
- 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 J. C. M. Dekkers. 1996. Random regression models for production traits in Canadian Holsteins. *Interbull Bull.* No. 14:124–134.
- Jamrozik, J., J. Kistemaker, L. R. Schaeffer, and J. C. M. Dekkers. 1997. Comparison of possible covariates for use in a random regression model for analyses of test-day yields. *J. Dairy Sci.* 80:2550–2556.
- Johnson, V. E. 1996. Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths. *J. Am. Statist. Assoc.* 91:154–166.
- Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *J. Am. Statist. Assoc.* 90:773–795.
- Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection, and evolution of growth trajectories. *Genetics* 124:979–993.
- López-Romero, P., and M. J. Carabaño. 2003. Evaluating alternative random regression models to analyse first lactation daily milk yield data in Holstein-Friesian cattle. *Livest. Prod. Sci.* 82:81–96.
- 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.
- Newton, M. A., and A. E. Raftery. 1994. Approximate Bayesian Inference with the weighted likelihood bootstrap. *J. R. Statist. Soc. B.* 56:3–48.
- Olori, V. E., W. G. Hill, B. J. McGuirk, and S. Brotherstone. 1999a. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression model. *Livest. Prod. Sci.* 61:53–63.
- Olori, V. E., W. G. Hill, and S. Brotherstone. 1999b. The structure of the residual error variance of test day milk yield in random regression models. *Interbull Bull.* 20:103–108.
- 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.
- Raftery, A. E. 1996. Hypothesis testing and model selection. Pages 145–161 in *Markov Chain Monte Carlo in Practice*. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds. Chapman and Hall, New York.
- Raftery, A. E., and S. M. Lewis. 1992. How many iterations in the Gibbs sampler. Pages 763–773 in *Bayesian Statistics 4*. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds. Oxford University Press, Oxford, UK.
- Raftery, A., and S. M. Lewis. 1995. Gibbsit Version 2.0. <http://lib.stat.cmu.edu/general/gibbsit>.
- 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.
- Rekaya, R., M. J. Carabaño, and M. A. Toro. 2000. Assessment of heterogeneity of residual variances using changepoint techniques. *Genet. Sel. Evol.* 32:383–394.
- Schnyder, U., A. Hofer, F. Labroue, and N. Kunzi. 2001. Genetic parameters of random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs. *Genet. Sel. Evol.* 33:635–658.
- Sorensen, D. 1996. Gibbs sampling in quantitative genetics. Internal report no. 82 from the Danish Institute of Animal Science, Foulum, Denmark.
- Stephens, D. A. 1994. Bayesian retrospective multiple-changepoint identification. *Appl. Statist.* 43:159–178.
- Van der Werf, J. H. J., M. E. Goddard, and K. Meyer. 1998. The use of covariance function and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* 81:3300–3308.
- Varona, L., C. Moreno, L. A. García-Cortés, and J. Altarriba. 1998. Bayesian analysis of wood's curve for Spanish dairy cows. *J. Dairy Sci.* 81:1469–1478.
- Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164–165.