Contrasting Models for Lactation Curve Analysis

F. Jaffrezic,*,† I. M. S. White,* R. Thompson,‡ and P. M. Visscher*
*Institute of Cell, Animal and Population Biology,
University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT
†INRA-SGQA, Jouy-en-Josas 78352, France
‡Rothamsted Experimental Station,
IACR, Harpenden, Herts AL5 2JQ, UK
and
Roslin Institute (Edinburgh),
Roslin, Midlothian EH25 9PS, UK

ABSTRACT

Several statistical models have been proposed for the genetic evaluation of production traits in dairy cattle based on test-day records. Three main approaches have been put forward in the literature: random regression, orthogonal polynomials, and, more recently, character process models. The aim of this paper is to show how these different approaches are related, to compare their performance for the genetic analysis of lactation curves, and to assess equivalence between sire and animal models for repeated measures analyses. It was found that, with an animal model, a character process model with 11 parameters performed better, regarding the likelihood criterion, than a quartic random regression model (with 31 parameters). However, although the likelihood was higher, the genetic variance was very different with the character process model from the unstructured model, which raises important issues concerning model selection criteria. There are advantages in combining methodologies. A quadratic random regression model for the environmental part, combined with a character process model for the residual, performed better than the quartic random regression model and had fewer parameters. A character process structure allowing for a correlation pattern modeled the residual better than a simple quadratic variance, and had only one extra parameter.
(**Key words:** lactation curve, random regression, orthogonal polynomials, character process models)

INTRODUCTION

Several methodologies have already been proposed for genetic evaluation of production traits for dairy

cattle based on test-day records. Currently, the most commonly used test-day models are random regression models (Diggle et al., 1994; Jamrozik and Schaeffer, 1997). The idea of these models is to consider a mean curve in the population, which can be either parametric or non-parametric, and to model individual deviations from this mean curve for each animal. These deviations are usually modeled with polynomial functions and, more specifically, orthogonal polynomials that have desirable numerical properties. Estimates of genetic values at each time are directly obtained from these individual curves. Another approach, called character process models, recently proposed by Pletcher and Geyer (1999), concentrates on the modeling of the covariance structure. If a completely unstructured matrix were considered, which corresponds to a multivariate analysis, the number of parameters to be evaluated would be very large. The character process approach aims at reducing the number of parameters in the covariance structure by considering appropriate parametric functions for the variance and correlation. The aim of this paper is to investigate how character process models can be incorporated into the mixed model framework well-known by animal breeders, to contrast them with the random regression approach, and to assess equivalence between sire and animal models for repeated measures analyses.

MATERIALS AND METHODS

Theory

Mixed model framework. Each individual has several observed measurements over time. For simplicity, individuals are assumed to have the same number of measurements J taken at the same times (t_1, …, t_j), although this assumption is not required for the models considered. Assuming a sire model, for individual i within sire s, the model can be written:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_s + \mathbf{p}_i + \mathbf{e}_i \qquad [1]$$

where $\mathbf{y}_i$ is the vector of observed measurements for individual i: $\mathbf{y}_i = (y_{i1}, \ldots, y_{i,J})'$, $\mathbf{X}_i\boldsymbol{\beta}$ represent the fixed effects, $\mathbf{u}_s$ is the vector of additive genetic values of sire s at the different times of measurement: $\mathbf{u}_s = (u_{s1}, \ldots, u_{sJ})'$, $\mathbf{p}_i$ is the permanent environmental effect (for example, the cow effect in the case of dairy cattle evaluation), and $\mathbf{e}_i$ is a residual term that can also be considered as a short-term environmental effect. As measurements taken on a same individual are correlated, the random vectors are assumed to follow multivariate normal distributions: $\mathbf{u}_s \sim N(0, \mathbf{G})$, $\mathbf{p}_i \sim N(0, \mathbf{P})$, $\mathbf{e}_i \sim N(0, \mathbf{E})$, and $\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V})$ where $\mathbf{V} = \mathbf{G} + \mathbf{P} + \mathbf{E}$. Two different approaches have been proposed to model the covariance structure.

Random regression models. The idea of these models is to fit a mean curve, either parametric or nonparametric, in the population and to model individual deviations from this mean curve. For example, in the case of lactation curve analysis, an exponential curve of Wilmink (1987) can be assumed for the general curve of the population:

$$g(t) = \alpha_0 + \alpha_1 t + \alpha_2 \exp(-Dt) \qquad [2]$$

where t stands for days in milk. Individual deviations (genetic and environmental) from this curve are modeled with parametric functions of time, for example, polynomials. If a linear deviation is assumed for the genetic part, then

$$u_{sj} = a_s + b_s t_j, \qquad [3]$$

where a_s and b_s are assumed to be normally distributed and correlated. In this case, u_{sj} corresponds to the genetic value of sire s at time t_j, and the genetic covariance between observations at two ages t_i and t_j is given by:

$$G(t_i, t_j) = \text{Cov}(u_{si}, u_{sj}) = \text{Var}(a_s) \qquad [4]$$
$$+ \text{Cov}(a_s, b_s)(t_i + t_j) + \text{Var}(b_s)t_i t_j$$

The aim is to try to find the best parametric curve to model individual deviations. The covariance structure is then derived from the chosen regression model. In practice, orthogonal polynomials are often used because of their desirable numerical properties.

Character process models. These models concentrate on the modeling of the covariance structure itself. The idea is that if a completely unstructured form were considered for the covariance matrices, the number of parameters to be estimated would be far too large in practice. Therefore, parametric functions are assumed to model the variance and correlation structures and to reduce the number of parameters. The covariance function between observations at two given times t_i and t_j can be written as:

$$C(t_i, t_j) = v(t_i)\, v(t_j)\, \rho(t_i, t_j) \qquad [5]$$

where $v^2(t_i)$ is a variance function, which is an arbitrary function of time, and is usually assumed to be a polynomial, and $\rho(t_i, t_j)$ is a correlation function. As the variance function has to be positive, we model its logarithm. For example, if a linear variance function is assumed:

$$\ln v^2(t_i) = a + bt_i \qquad [6]$$

Pletcher and Geyer (1999) proposed a list of several possible correlation functions, such as a standard normal $\rho(t_i, t_j) = \exp(-\theta(t_i - t_j)^2)$ where $\theta > 0$, a Cauchy $\rho(t_i, t_j) = 1/(1 + \theta(t_i - t_j)^2)$ where $\theta > 0$, or an exponential correlation $\rho(t_i, t_j) = \theta^{|t_i - t_j|}$ where $0 < \theta < 1$ can be considered. Therefore, in the case of a log-linear variance and exponential correlation, for example, the genetic covariance between observations at two ages t_i and t_j can be written as:

$$\mathbf{G}(t_i, t_j) = \exp\!\left(a + b\!\left(\frac{t_i + t_j}{2}\right)\right)\theta^{|t_i - t_j|} \qquad [7]$$

Pletcher and Geyer (1999) originally assumed stationarity in correlation, i.e., the correlation between two ages is a function only of the time distance $|t_i - t_j|$ between them. This assumption can be relaxed by implementing a non-linear transformation upon the time axis, f(t), such that correlation stationarity holds on the transformed scale, but on the original scale the correlation is non-stationary (Jaffrezic and Pletcher, 2000; Nunez-Anton, 1998). The correlation function is then defined as $\rho(t_i, t_j) = \rho(|f(t_i) - f(t_j)|)$, and the correlation functions suggested by Pletcher and Geyer (1999) remain valid. As proposed by Nunez-Anton and Zimmerman (2000), a Box-Cox transformation seems to offer good flexibility with only one extra parameter:

$$f(t,\lambda) = (t^\lambda - 1)/\lambda \qquad \text{if } \lambda \neq 0$$
$$= \ln t \qquad\qquad \text{if } \lambda = 0$$

Considering an exponential correlation function, $\rho(t_i, t_j) = \theta^{|t_i - t_j|}$, the correlations on the subdiagonals are monotone increasing if $\lambda < 1$ or monotone decreasing if $\lambda > 1$. If $\lambda = 1$, the non-stationary model reduces to a stationary one.

In the general case, the inverse of covariance matrices defined by character process models are not sparse. However, when considering an exponential correlation function, the inverse of the covariance matrix is tridiagonal. This model is, in fact, equivalent to the structured antedependence models (Nunez-Anton and Zimmerman, 2000) of order 1 and corresponds to a continuous and non-stationary generalization of autoregressive model of order 1. Sparseness of the inverse can be useful when analyzing large data sets.

Orthogonal polynomials. Kirkpatrick and Heckman (1989) proposed use of orthogonal polynomials as a non-parametric way of smoothing previously estimated covariance matrices. As originally presented, the orthogonal polynomial approach is similar in spirit to the character process models in that both methods forgo the shapes of individual deviations in favor of directly modeling the covariance structure. Meyer (1998) has shown that this approach is equivalent to considering orthogonal polynomials in a random regression model.

Estimation procedure. For both models, classical mixed model equations can be used to estimate fixed effects β and to predict genetic values **u**. An average information algorithm (Gilmour et al., 1995) can be applied for the REML estimation of variances and parameters in the character process model. Recent versions of the program ASREML (Gilmour et al., 2000) allow character process models to be used.

Comparison of Methods

Data set. These methodologies were applied to the genetic evaluation of first lactation milk production for dairy cattle. Lactation curves were fitted to test-day records for 9277 progeny of 464 Holstein-Friesian sires, assumed unrelated. Observations were made over two years (1993 and 1994). The lactation stage of animals at first test varied between 4 and 40 d, with successive tests at approximately 30-d intervals. All cows had 10 measurements. The fixed effects considered were the age at calving, the percentage of North American Holstein genes, and herd-test-month. An exponential curve of Wilmink (1987) was fitted as a fixed regression model for the general curve of the population (equation 2). Parameter D was assumed to be known and equal to 0.068, chosen based on previous studies (White et al., 1999).

In classical quantitative genetics theory, when unrelated sires are considered, sire and animal models are equivalent. In the case of repeated measures, more complex covariance structures are involved, and the equivalence is not straightforward. We therefore performed all the analyses with both a sire and an animal model.

Model comparisons. The aim was to compare the performance of random regression and character process in modeling genetic and permanent environmental parts for lactation curve analysis. In a first analysis, residual terms were assumed to be independent and with constant variance over time. In order to have a reference value for the likelihood, we fitted a saturated model for the covariance structure, i.e., we considered completely unstructured genetic and environmental covariance matrices, which corresponds to a multivariate analysis. The other models were compared to this value to evaluate their goodness-of-fit. Many different combinations of variance (polynomials up to quadratic) and correlation functions (exponential: $\theta^{|t_i - t_j|}$, Gaussian: $\exp(-\theta(t_i - t_j)^2)$, Cauchy: $1/(1 + \theta(t_i - t_j)^2)$), stationary or non-stationary, were considered for the character process approach. Polynomials up to the quartic order were fitted for the random regression models. All the analyses were performed using ASREML (Gilmour et al., 2000).

Residual structure. In a second analysis, the assumptions of independence and constant variance of the residuals were relaxed. A character process structure was considered at the residual level. This corresponds to the models proposed by Diggle et al. (1994) and Zhang et al. (1998), in which a random regression is considered to model within-subject variation and a stochastic process is added to model serial correlation.

RESULTS

In order to make the description easier, the different models have been numbered (see Table 1) and are referred to in the text using these numbers. In order to have an informal but quantitative measure for the goodness-of-fit of variance and correlation functions, the concordance correlation coefficient proposed by Vonesh et al. (1996 [see Appendix]), which considers the unstructured covariance matrix as reference, was used.

Model Comparisons

The model with a completely unstructured genetic covariance matrix did not converge properly with AS-REML. Estimates of the genetic parameters were obtained in this case with another REML program which uses a different algorithm (Meyer, 1985) for multivariate analysis.

Sire and animal models. In general, when analyzing single measurements, a sire model gives the same results as an animal model using a simple pedigree

Table 1. Model comparisons (likelihood values are given with reference to the unstructured model in absolute values).

Model	Genetic	Environmental		NPCov[1]	Log-likelihood Sire	Animal
Unstructured						
1	US[2]	US		110		0
2	Quad RR[3]	US		61		22
3	Repeatability	US		57		103
Character process + constant residual variance						
4	Quad-ExpNS	Quad-ExpNS[4]		11	863	316
5	Quad-Exp	Quad-ExpNS		10	870	349
6	Quad-Exp	Quad-Exp		9	1370	501
7	Lin-Exp	Quad-ExpNS		9	885	623
8	Lin-Exp	Quad-Exp		8	1396	711
Random regression + constant residual variance						
9	Quartic	Quartic		31		506
10	Quad	Quartic		22		522
11	Cubic	Cubic		21		793
12	Quad	Quad		13		1362
13	Quad	Lin		10	2902	1397
Random regression + character process						
14	Quad	Quartic	Quad-Exp	25		85
15	Quad	Quad	Quad-Exp	16		326
16	Quad	Quad	Quad-Zero[5]	15		631

[1]Number of parameters in the covariance structure.

[2]Unstructured covariance matrix.

[3]Quadratic random regression.

[4]Quadratic variance – non-stationary exponential correlation.

[5]Quadratic residual variance – correlation set to 0.

with sires specified and dams unknown. In this case, the sire model is just a reparameterization of the animal model and vice-versa. The sire model provides estimates of between-sire variance $V(B) = (1/4)V(A)$ and within-sire variance $V(W) = (3/4)V(A)+V(E)$. The animal model estimates $V(A)$ and $V(E)$ directly.

With repeated measurements, the two models do not necessarily give the same results. In the case of random regression models, an animal model with an order p polynomial for the genetic effect and an order q polynomial for the environmental effect will give the same results as a sire model, with an order p polynomial modeling differences between sires and an order q polynomial within sires, provided $q \geq p$. (The within-sire differences are modeled as the sum of an order p and an unrestricted order q polynomial, equivalent to an unrestricted order q polynomial if $q \geq p$). If $q < p$, the within-sire differences are modeled as an order p polynomial. However, the leading coefficient is forced to be the same as that of the between-sire polynomial. The restricted nature of the within-sire polynomial in this case means that the animal model is not equivalent to a sire model with order p polynomials for between- and within-sire components.

With character process models, there is no equivalence between the two types of model. For example, if a first-order autoregressive structure is assumed for

genetic and environmental components, the corresponding within-sire structure would have to be a linear combination of two autoregressive structures.

For repeated measurements, the fact that the within-sire component is the sum of genetic and environmental components means that it may well be difficult to model with simple structures, and that in those cases in which the animal and sire models are not equivalent, we might expect the animal model to do better.

Model selection. The best fitting animal model for the likelihood criterion among all the character process and random regression models considered in the first analysis was a character process with quadratic variance and non-stationary exponential correlation for both genetic and environmental parts. With only 11 parameters, the likelihood was, in fact, higher than a quartic-quartic random regression model (difference of 190), which had 31 parameters.

Figure 1 shows that estimates obtained with the animal model for the best character process (Model 4) improved the fit for the within-sire variance compared to the sire model. Vonesh's coefficient (see Appendix) for the within-sire variance was 0.8 for the animal model and only 0.6 for the sire model. In this case, the animal model was equivalent to considering the sum of two character processes for the within-sire component,
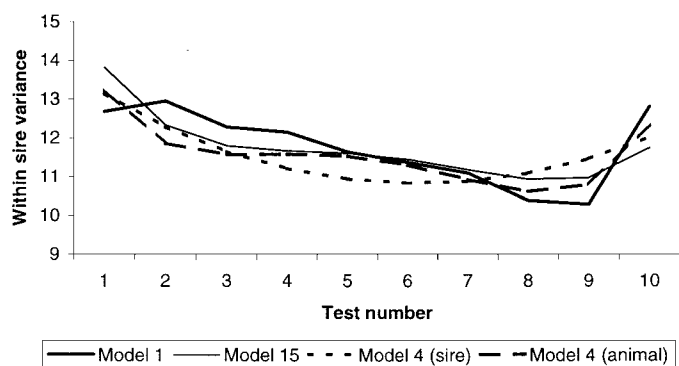
Figure 1. Within sire variance. Model 1: Unstructured. Model 15: Quadratic random regression + residual character process. Model 4: Character process with quadratic variance and nonstationary exponential correlation – Sire and animal models.

instead of one for the sire model. However, the between-sire component in the animal model was very different from the estimate from the unstructured model (Model 1), as shown in Figure 2. This seems to be due to the fact that the model attached more weight to the within-sire component and neglected the between-sire part. When calculating the information matrices, it could be shown that 80% of the information for the genetic variance in the animal model came from the within-sire component. Moreover, it seems that the likelihood only reflected how well the within sire component was fitted and was very little influenced by the goodness-of-fit of the between-sire part. In fact, the difference between estimates obtained with Model 4 and the unstructured model (Model 1) was not reflected in the likelihood. This may be a very important problem in model selection as animal breeders are mainly interested in modeling of the genetic
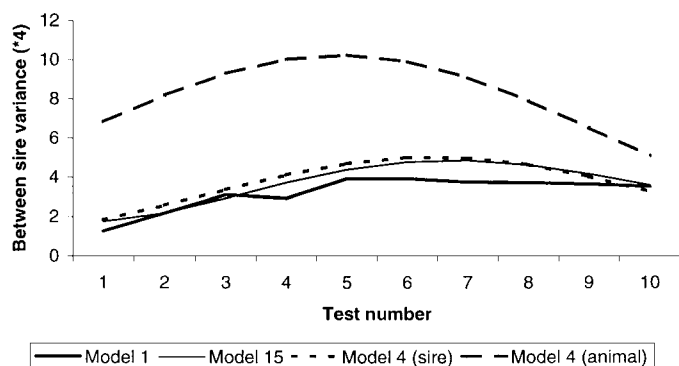

Figure 2. Between sire variance. Model 1: Unstructured. Model 15: Quadratic random regression. Model 4: Character process with quadratic variance and nonstationary exponential correlation – Sire and animal models.

part. Likelihood may, therefore, not be the most appropriate criterion to select a model.

Residual Structure

In order to avoid misspecification of the between-sire component, we considered a more complex structure that would better capture the within-sire part. As proposed by Diggle et al. (1994) and Zhang et al. (1998), we considered an additional stochastic process at the residual level to model variance and correlation that could not be taken into account in the previous models. The residual variance was assumed to change according to a quadratic function (Jaffrezic et al., 2000), and we considered an exponential correlation function.

Table 1 shows that a very significant improvement of the likelihood was reached with this structure compared to the simple random regression models. When considering, for example, the quadratic-quartic random regression model—Model 10— with only three additional parameters (Model 14), the likelihood was much closer to that obtained by the unstructured model—difference of 85—whereas the difference was 522 without the additional stochastic process. Moreover, the likelihood of a quadratic-quadratic random regression model with the additional character process structure (Model 15) was higher than the quadratic-quartic model (Model 10) with a difference in likelihood of 196 with fewer parameters (16 instead of 22). The likelihood was also higher when considering a character process structure on the residual rather than only a quadratic variance (assuming independence of the residuals), as in Model 16. For a quadratic-quadratic random regression model, the parameter $\hat{\theta}$ in the correlation function was 0.14, which was significantly different from zero, and with this extra parameter the difference of likelihood was 305 compared to a simple quadratic residual variance (Models 15 and 16). This means that there are additional short-term environmental effects that could not be taken into account with the models considered in the first analysis.

The same estimation pattern as previously observed for between-sire component in an animal model (Model 4) was obtained when a character process was added at the residual level to the previously considered character process models. Based upon the likelihood criterion, the model that would be chosen is a quadratic random regression for the genetic part, quartic for the permanent environmental part, and character process model with quadratic variance and exponential correlation for the residuals (25 parameters), as in Model 14. However, in practice it would be preferable to keep

Table 2. Genetic correlation matrix.

Model 1: Unstructured

1									
0.96	1								
0.89	0.96	1							
0.80	0.90	0.98	1						
0.75	0.87	0.97	0.98	1					
0.70	0.83	0.94	0.97	0.99	1				
0.69	0.81	0.94	0.97	0.99	0.99	1			
0.62	0.73	0.88	0.90	0.94	0.97	0.97	1		
0.58	0.64	0.82	0.85	0.87	0.90	0.93	0.97	1	
0.69	0.73	0.88	0.89	0.90	0.93	0.94	0.97	0.99	1

the number of parameters as small as possible while keeping a reasonable fit, and therefore the model with quadratic polynomial for both the genetic and environmental parts and a character process with quadratic variance and exponential correlation for the residual part (16 parameters), Model 15, would be favored.

Goodness-of-Fit Analysis

Genetic structure (between-sire component). It seems that simple models, such as quadratic random regression, were able to capture both the genetic correlation and variance patterns. The genetic correlation remained quite high over time, as shown in Table 2, and Jaffrezic and Pletcher (2000) showed that this pattern can easily be captured with classical models. The Vonesh's coefficient for the chosen model, with a quadratic random regression, compared to the unstructured one (Models 15 and 1) was 0.9. The genetic variance also seemed to be quite well modeled with the chosen model ($r_c = 0.8$), as shown in Figure 2.

Although simple models seemed to capture the genetic covariance structure, the likelihood criterion should be considered with caution when choosing the genetic model, as shown in the animal model analysis for character processes.

Environmental structure (within-sire component). The likelihood-based criterion led to the choice of more sophisticated models for the environmental covariance structure. Table 3 gives the unstructured environmental correlation and covariance matrices. For the chosen model presented in the previous analysis (Model 15), the fit seemed to be good for the correlation structure ($r_c = 0.99$). The overall variance-covariance matrix was also well-fitted by the chosen model ($r_c = 0.99$).

However, the difficulty of modeling the environmental covariance structure, and therefore the requirement for complicated models, seemed to be mainly due to the substantial increase in variance at the end of the lactation (Figure 1). This pattern has been observed in previous studies (Rekaya et al., 1999), and does not seem to be well-accommodated with standard polynomial functions. However, removing the last test from the analysis allows much simpler models to be used. In this case, even a simple character process model with linear variance and exponential correlation for the genetic part and quadratic variance with non-stationary exponential correlation for the environmental part (Model 7 in Table 1), with only 9 parameters, seemed to accommodate the covariance structure well. The likelihood difference with the unstructured model was 192 in the 9-test analysis, whereas it was 885 with 10 tests for the sire model.

DISCUSSION

In classical quantitative genetics theory, when unrelated sires are considered, sire and animal models

Table 3. Correlation (below diagonal) and covariance (above diagonal) matrices for the within-sire component.

Model 1: Unstructured

12.7	7.6	6.6	6.0	5.7	5.3	5.0	4.6	4.1	3.6
0.59	12.9	8.6	8.0	7.5	7.0	6.6	6.0	5.5	4.8
0.53	0.68	12.3	8.6	8.0	7.5	7.1	6.5	6.1	5.3
0.49	0.63	0.70	12.1	8.6	8.2	7.7	7.1	6.5	5.8
0.47	0.61	0.68	0.72	11.6	8.7	8.2	7.6	6.9	6.2
0.44	0.58	0.64	0.70	0.76	11.4	8.6	7.9	7.1	6.4
0.42	0.55	0.61	0.66	0.72	0.77	11.1	8.3	7.5	6.7
0.40	0.52	0.58	0.63	0.69	0.72	0.77	10.4	7.8	7.0
0.36	0.48	0.54	0.58	0.63	0.66	0.70	0.75	10.3	7.8
0.28	0.37	0.43	0.46	0.51	0.53	0.56	0.61	0.68	12.8

are equivalent. This is, however, in general no longer the case with repeated measures. For random regression models, sire and animal models are only equivalent when a lower- or equal-order polynomial is considered for the genetic compared to the environmental part. In general, the equivalence does not hold for character process models. With an animal model, a character process model with 11 parameters performed better, regarding the likelihood criterion, than a quartic-quartic random regression model (with 31 parameters). However, although the likelihood was higher (difference of 190), the genetic variance was misspecified with the character process model. This was because the model attached more weight to the within-sire component and neglected the between-sire part. This discrepancy could not be detected with the likelihood, which may not be the most appropriate criterion for model selection. Other criteria should be investigated, such as choosing the model that offers the most accurate genetic values or which gives the best response to selection.

These analyses showed that additional flexibility could be obtained when combining both random regression and character process models in the analysis, as proposed by Zhang et al. (1998). Even the simple quadratic random regression model with an additional stochastic process at the residual level performed better than a quartic polynomial, with fewer parameters. It was also shown that a character process structure, allowing for a correlation pattern, performed better than a simple quadratic residual variance with only one extra parameter. This may be due to a short-term environmental effect, which cannot be taken into account by the random regression model. This additional effect may be even more important for daily record analysis.

ACKNOWLEDGMENTS

REFERENCES

Diggle, P. J., K. Y. Liang, and S. L. Zeger. 1994. Analysis of Longitudinal Data. Oxford University Press, Oxford.

Gilmour, A. R., R. Thompson, and B. R. Cullis. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics. 51:1440–1450.

Gilmour, A. R., R. Thompson, B. R. Cullis, and S. J. Welham. 2000. AS-REML Manual. New South Wales Department of Agriculture, Orange, 2800, Australia.

Jaffrezic, F., and S. D. Pletcher. 2000. Statistical models for estimating the genetic basis of repeated measures and other function-values traits. Genetics. 156:913–922.

Jaffrezic, F., I. M. S. White, R. Thompson, and W. G. Hill. 2000. A link function approach to model heterogeneity of residual variances over time in lactation curve analyses. J. Dairy Sci. 83:1089–1093.

Jamrozik, J., and L. 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.

Kirkpatrick, M., and N. Heckman. 1989. A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. J. Mathematical Biology. 27:429–450.

Meyer, K. 1985. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics. 41:153–165.

Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. Genet. Sel. Evol. 30:221–240.

Nunez-Anton, V. 1998. Longitudinal data analysis: Non-stationary error structures and antedependent models. Applied Stochastic Models and Data Analysis. 13:279–287.

Nunez-Anton, V., and D. L. Zimmerman. 2000. Modeling nonstationary longitudinal data. Biometrics. 56:699–705.

Pletcher, S. D., and C. J. Geyer. 1999. The genetic analysis of age-dependent traits: Modeling a character process. Genetics. 153:825–833.

Rekaya, R., M. J. Carabano, 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.

Vonesh, E., V. Chinchilli, and K. Pu. 1996. Goodness-of-fit in generalized nonlinear mixed-effects models. Biometrics. 52:572–587.

White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82:632–638.

Wilmink, J. B. M. 1987. Adjustment of test day milk, fat and protein yield for age, season and stage of lactation. Livest. Prod. Sci. 16:335–348.

Zhang, D., X. Lin, J. Raz, and M. F. Sowers. 1998. Semiparametric stochastic mixed models for longitudinal data. J. Am. Stat. Assoc. 93:710–719.

APPENDIX

Vonesh's Coefficient

The concordance correlation coefficient r_c described by Vonesh et al. (1996) was used to evaluate the goodness-of-fit for both the variance and correlation functions estimated by the models when compared to the unstructured covariance matrix, which corresponds to a multivariate analysis. For the correlation structure, for instance, we consider:

$$r_c = 1 - \frac{\sum_{i=1}^{T-1} \sum_{j=i+1}^{T} (y_{ij} - \hat{y}_{ij})^2}{\sum_{ij}(y_{ij} - \bar{y})^2 + \sum_{ij}(\hat{y}_{ij} - \hat{y})^2 + T(T-1)(\bar{y} - \hat{y})^2/2} \quad [8]$$

where $\hat{y}_{ij}$ represents the estimated correlation between observations at times t_i and t_j given by the model, and y_{ij} is the correlation between observations at times t_i and t_j in the unstructured matrix. T represents the total

number of times at which measurements were taken. $\overline{y}$ and $\hat{y}$ are the mean of the correlation values for the unstructured matrix and the model, respectively. The concordance coefficient for the variance estimate is much simpler and given by

$$r_c = 1 - \frac{\Sigma_{i=1}^{T}(y_i - \hat{y}_i)^2}{\Sigma_i(y_i - \overline{y})^2 + \Sigma_i(\hat{y}_i - \hat{y})^2 + T(\overline{y} - \hat{y})^2} \quad [9]$$

where y_i now refers to variances rather than correlations.

The coefficient r_c directly measures the level of agreement (concordance) between y_{ij} and $\hat{y}_{ij}$, and its value is reflected in how well a scatter plot y_{ij} versus $\hat{y}_{ij}$ falls about the line identity. The possible values of r_c are in the range $-1 \leq r_c \leq 1$, with a perfect fit corresponding to a value of 1 and a lack of fit to values ≤ 0.