

May 8, 2003

# Random regression models for analyses of longitudinal data in animal breeding

Karin Meyer

Animal Genetics and Breeding Unit<sup>1</sup>, University of New England,  
Armidale, NSW 2351, Australia

E-mail : kmeyer@didgeridoo.une.edu.au

Phone : +61 2 6773 3331

Fax : +61 2 6773 3266

## Summary

Statistical analyses of data in animal breeding are concerned with genetic differences between individuals and the proportion of variation attributable to genetic and environmental effects. Often traits are recorded repeatedly per individual and may change, gradually and continually, as time progresses, so that they are thought of as points on a trajectory. A typical example is growth of meat producing animals, such as beef cattle. Random regression (coefficient) analyses allow for variation in growth (or similar) curves for individual animals, modelled as regression on functions of age, e.g. orthogonal polynomials or splines. Covariances among the regression coefficients then define the covariance function between records at any pair of ages. With functions linear in the coefficients to be estimated, random regression analyses are readily carried out in the linear, mixed model framework commonly employed in animal breeding. Estimation of breeding values using best linear unbiased prediction, and estimation of the corresponding, genetic and environmental covariance components by restricted maximum likelihood using an average information algorithm or Bayesian analysis using Gibbs sampling are reviewed, with emphasis on formulations based on the mixed model equations. Interpretation of genetic covariance functions is discussed and results from an analysis of beef cattle data are given as an example.

**Keywords :** Repeated records, mixed model, random coefficients, covariance function, quantitative genetics

---

<sup>1</sup>AGBU is a joint venture of NSW Agriculture and the University of New England

# 1 Introduction

Animal breeding is concerned with the genetic improvement of farmed livestock, applying quantitative genetics principles. This involves the identification of genetically superior animals as parents of future generations and design of breeding programmes to optimise genetic progress for traits of economic interest. These tasks require knowledge of genetic parameters such as heritabilities, correlations and variances for the traits concerned. Hence, estimation of genetic parameters or, equivalently, (co)variance components is a central task in statistical analyses of animal breeding data.

Data sets originating from livestock recording schemes are typically highly unbalanced, subject to multiple systematic effects, and often large. In addition, data may be subject to selection. Statistical techniques used rely heavily on linear, mixed models, employing Best Linear Unbiased Prediction (BLUP) to obtain estimates of animals' genetic merit and restricted maximum likelihood (REML) or, more recently, Bayesian techniques to estimate variance components. Algorithms used are frequently formulated in terms of the so-called mixed model equations which are of dimension proportional to the number of fixed and random effects fitted.

Whilst most traits of interest are measured only once per animal, others are recorded repeatedly and may change, gradually and continually, as time progresses. Individual records can then be thought of as points on a trajectory. Typical examples are test day records for dairy cows, with milk production at the beginning and end of lactation having quite different means and variances but high genetic correlations, and growth of meat-producing animals. Earlier analyses treated such data simply as repeated records assuming homogeneous variances and compound symmetry for covariances. Generally this was combined with some adjustment for time of recording, for instance, by fitting time as a covariable with the corresponding regression coefficients being fixed effects in the model of analysis. Alternatively, for data spanning larger periods of time with marked changes in means and variances, measurements in different periods were treated as different traits in multivariate analyses. For example, for growth of beef cattle up to about two years of age, we often distinguish between birth, weaning, yearling and final weights, allowing ranges of age at recording of 200 days or more for each.

Recently, random regression (RR) models, often referred to as random coefficient models elsewhere, have become popular for the analysis of data representing repeated measures along some continuous scale and curve. Impetus has come more or less at the same time from animal breed-

ing, in particular genetic evaluation of dairy cattle, and evolutionary biology. Utilising ideas of Henderson (1982), Schaeffer & Dekkers (1994) and Jamrozik et al. (1997) modelled changing genetic effects for milk production over the course of lactation by fitting a set of random regression coefficients for each animal, with polynomial functions of days in milk as covariables. Considering traits measured potentially many times along a continuous scale, Kirkpatrick and co-workers (Kirkpatrick & Heckman, 1989; Kirkpatrick et al., 1990) proposed an 'infinite-dimensional' model, where covariances between records at any points are given as functions of the continuous scale. This introduced the concept of covariance functions (CF), as the infinite-dimensional equivalent to covariance matrices, to quantitative genetics.

Kirkpatrick et al. (1990) outlined a generalised least-squares procedure to estimate genetic CFs as functions of orthogonal polynomials of the continuous covariable from a matrix of genetic covariances between individual points. Whilst computationally simple, this required estimates of genetic covariances between all ages considered to be known, and did not guarantee non-negative definite estimates of CFs. Alternatively, estimates of CFs can be obtained directly from the data using restricted maximum likelihood (REML) or Bayesian analyses. Meyer & Hill (1997) presented an application to animal breeding data, describing direct REML estimation of genetic and environmental CFs. As this employed a reparameterisation of a multivariate model treating each age as a separate trait, its usefulness for data involving many and irregularly spaced ages, such as data from animal breeding applications, was limited. Subsequently, it was recognised that RR models implicitly define CFs, and thus allow the coefficients of CFs to be estimated simply as the covariances among RR coefficients using REML (Meyer, 1998) or Bayesian analysis via Gibbs sampling (Jamrozik & Schaeffer, 1997; Rekaya et al., 1999).

This paper reviews RR models as used in analyses of data from livestock improvement schemes, considering both genetic evaluation and variance component estimation. Whilst RR models are suitable to model changes along any continuous scale, they are most commonly applied to longitudinal data. Hence, age at recording will be used in the following to represent this scale.

## 2 Random regression model

Typically, genetic models for repeated records distinguish between animals' additive genetic and permanent environmental effects. Hence RR models for genetic analyses include at least two sets

79 of RR coefficients for each animal  $i$ , representing the direct, additive genetic ( $\alpha_{im}$ ) and permanent  
80 environmental ( $\gamma_{im}$ ) effects of the animal. If other random effects are to be taken into account,  
81 such as litter effects or maternal effects, corresponding sets of additional RR coefficients can be  
82 fitted. Without loss of generality, only the former two effects are considered in the following.

83 Let  $y_{ij}$  denote the  $j$ -th record for animal  $i$  taken at age (or time)  $a_{ij}$ . Covariables are functions  
84 of age,  $\phi_m(a_{ij})$ . Following suggestions of Kirkpatrick & Heckman (1989), orthogonal polynomials  
85 are the most common choice, as they require few assumptions about the shape of the trajectory to  
86 be modelled, but other functions, such as spline or trigonometric functions have been considered  
87 (e.g. Rice & Wu, 2001). This gives the linear model

$$y_{ij} = F_{ij} + \sum_{m=0}^{k_A-1} \alpha_{im} \phi_m(a_{ij}) + \sum_{m=0}^{k_R-1} \gamma_{im} \phi_m(a_{ij}) + \varepsilon_{ij} \quad (1)$$

88 where  $F_{ij}$  denotes the fixed effects and  $\varepsilon_{ij}$  the temporary environmental effect or ‘measurement  
89 error’ affecting  $y_{ij}$ . Fixed effects generally include corresponding regressions on age to model mean  
90 trajectories, fitted overall or within levels of systematic effects for which we expect differential  
91 growth curves, e.g. sex of animal or season of birth. The number of RR coefficients fitted for  $\alpha_i$   
92 and  $\gamma_i$  is given by  $k_A$  and  $k_R$ , respectively. As each record is assumed to be determined by the  
93 genetic effects of the animal on which it was taken, this is commonly called an ‘animal model’.  
94 In matrix notation,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{\Phi}_A \boldsymbol{\alpha} + \mathbf{\Phi} \boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (2)$$

95 with  $\mathbf{y}$ ,  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\gamma}$  and  $\boldsymbol{\varepsilon}$  the vectors of observations, RR coefficients and residuals, respectively,  
96 assumed to be ordered according to animals, and  $\mathbf{b}$  denoting the vector of fixed effects fitted.  $\mathbf{X}$ ,  
97  $\mathbf{\Phi}_A$  and  $\mathbf{\Phi}$  are the corresponding incidence matrices. Assume the trait of interest is continuous  
98 (rather than categorical) and that  $\mathbf{y}$  has normal distribution  $N(\mathbf{X}\mathbf{b}, \mathbf{V})$ . Further, assume that  
99  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\gamma}$  and  $\boldsymbol{\varepsilon}$  are uncorrelated and that

$$\begin{aligned} E[\boldsymbol{\alpha}] &= \mathbf{0} & E[\boldsymbol{\gamma}] &= \mathbf{0} & E[\boldsymbol{\varepsilon}] &= \mathbf{0} \\ \text{Var}(\boldsymbol{\alpha}) &= \mathbf{A} \otimes \mathbf{K}_A = \mathbf{G} & \text{Var}(\boldsymbol{\gamma}) &= \mathbf{I}_N \otimes \mathbf{K}_R = \mathbf{R} & \text{Var}(\boldsymbol{\varepsilon}) &= \text{Diag}\{\sigma_k^2\} = \boldsymbol{\Sigma}_\varepsilon \end{aligned}$$

100 with  $N$  denoting the number of animals with records, and  $\mathbf{I}_N$  an identity matrix of size  $N$ . This  
 101 gives variance of  $\mathbf{y}$

$$\mathbf{V} = \Phi_A (\mathbf{A} \otimes \mathbf{K}_A) \Phi_A' + \Phi (\mathbf{I}_N \otimes \mathbf{K}_R) \Phi' + \text{Diag} \{\sigma_k^2\} = \Phi_A \mathbf{G} \Phi_A' + \Phi \mathbf{R} \Phi' + \Sigma_\epsilon \quad (3)$$

102 Partitioning of animal effects into their genetic and non-genetic components requires records  
 103 for related animals and information on the degree of relationship between animals. Contrasts  
 104 between different types of relatives or, equivalently, resemblance of relatives with a given degree  
 105 of relationship allow inferences about additive genetic effects and variances; see, for instance,  
 106 Falconer & Mackay (1996) or Lynch & Walsh (1998) for details. Relationship information is  
 107 provided by the numerator relationship matrix  $\mathbf{A}$ . Off-diagonal elements of  $\mathbf{A}$  give the degree of  
 108 relationship or, equivalently, the proportion of genes animals are expected to have in common,  
 109 for example, 1/2 for a parent and an offspring, 1/2 for full-sibs, 1/4 for half-sibs and 1/8 for first  
 110 cousins. Diagonal elements of  $\mathbf{A}$  are unity plus the coefficient of inbreeding for the animal.

111 Invariably, there are animals which do not have any records but are ancestors of animals in the  
 112 data. Sets of RR coefficients for these animals are included in  $\alpha$ , which has length  $N_A \times k_A$  for  
 113  $N_A > N$ , with the corresponding rows of  $\Phi_A$  having elements of zero.  $\mathbf{K}_A = \{K_{Aij}\}$  and  $\mathbf{K}_R =$   
 114  $\{K_{Rij}\}$  are the matrices of covariances between RR coefficients, and  $\sigma_k^2$  for  $k = 1, \dots, p$  denote  
 115 the measurement error variances. In the simplest case,  $\varepsilon_{ij}$  are assumed to be *i.i.d.* distributed,  
 116 i.e.  $p = 1$ . In other cases,  $\sigma_k^2$  are assumed to change with age, for instance, due to scale effects or  
 117 reflecting seasonal variation. Heterogeneous  $\sigma_k^2$  are commonly modelled through a step function  
 118 of age or a continuous variance function linking age and  $\sigma_k^2$  or  $\log(\sigma_k^2)$ .

### 119 3 Estimation of Breeding Values

120 For known variances, solutions to the mixed model equations (MME)

$$\begin{bmatrix} \mathbf{X}'\Sigma_\epsilon^{-1}\mathbf{X} & \mathbf{X}'\Sigma_\epsilon^{-1}\Phi_A & \mathbf{X}'\Sigma_\epsilon^{-1}\Phi \\ \Phi_A'\Sigma_\epsilon^{-1}\mathbf{X} & \Phi_A'\Sigma_\epsilon^{-1}\Phi_A + \mathbf{A}^{-1} \otimes \mathbf{K}_A^{-1} & \Phi_A'\Sigma_\epsilon^{-1}\Phi \\ \Phi'\Sigma_\epsilon^{-1}\mathbf{X} & \Phi'\Sigma_\epsilon^{-1}\Phi_A & \Phi'\Sigma_\epsilon^{-1}\Phi + \mathbf{I} \otimes \mathbf{K}_R^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\Sigma_\epsilon^{-1}\mathbf{y} \\ \Phi_A'\Sigma_\epsilon^{-1}\mathbf{y} \\ \Phi'\Sigma_\epsilon^{-1}\mathbf{y} \end{bmatrix} \quad (4)$$

121 yield BLUPs of  $\alpha$  and  $\gamma$  (Henderson, 1973).

122 The MME (4) play a central rôle in statistical analyses of animal breeding data. Instead of the  
 123 inverse of the phenotypic variance matrix  $\mathbf{V}$  needed in a generalised least-squares formulation,  
 124 they require inverse matrices  $\Sigma_\epsilon^{-1}$  and  $\mathbf{A}^{-1}$ . Whilst both  $\mathbf{V}$  and  $\Sigma_\epsilon$  have dimensions equal to  
 125 the number of records, the latter is blockdiagonal for animals (assuming zero environmental  
 126 covariances between records for different animals), if not, as in this case, diagonal.  $\Sigma_\epsilon$  is thus  
 127 readily invertible, and the data part of the MME can be accumulated considering records for one  
 128 animal at a time.  $\mathbf{A}$  is of size  $N_A \times N_A$ , but  $\mathbf{A}^{-1}$  can be set up directly from a list of pedigree  
 129 records, following simple rules (Henderson, 1976; Quaas, 1976).

130 Equations (4) are generally solved iteratively, using schemes such as Gauss-Seidel iteration with  
 131 overrelaxation or a preconditioned conjugate gradient algorithm (Tsuruta et al., 2001). This  
 132 allows solutions to MME comprising millions of equations to be obtained routinely for national  
 133 genetic evaluation schemes.

134 Coefficients  $\alpha_{im}$  for animal  $i$  define its genetic merit for all ages within the range considered.  
 135 Hence, estimated breeding values for target ages  $a_t$  are obtained simply by evaluating the regres-  
 136 sion curve

$$\text{EBV}(a_t)_i = \sum_{m=0}^{k_A-1} \alpha_{im} \phi_m(a_t) \quad (5)$$

137 In addition, functions of the curve may be of interest, for example, integrals of estimated lactation  
 138 curves to estimate total lactation genetic merit for dairy cows, or turning points of growth curves  
 139 to distinguish between early and late maturing animals.

## 140 4 Estimation of Covariance Functions

141 The RR model directly defines the covariances between records at any observed ages. These  
 142 are functions of the covariables,  $\phi_m(a_t)$ , evaluated at the ages of interest, and the covariances  
 143 among random regression coefficients. Let  $\phi_{Qt}$  denote the vector of covariables evaluated for  
 144 age  $a_t$  and random effect  $Q$ , of length  $k_Q$ . Covariance functions for genetic ( $\mathcal{A}$ ) and permanent  
 145 environmental ( $\mathcal{R}$ ) covariances between ages  $a_i$  and  $a_j$  are then

$$\begin{aligned}
\mathcal{A}(a_i, a_j) &= \sum_{m=0}^{k_A-1} \sum_{n=0}^{k_A-1} \phi_m(a_i) \phi_n(a_j) K_{Amn} = \boldsymbol{\phi}'_{Ai} \mathbf{K}_A \boldsymbol{\phi}_{Aj} \\
\mathcal{R}(a_i, a_j) &= \sum_{m=0}^{k_R-1} \sum_{n=0}^{k_R-1} \phi_m(a_i) \phi_n(a_j) K_{Rmn} = \boldsymbol{\phi}'_{Ri} \mathbf{K}_R \boldsymbol{\phi}_{Rj}
\end{aligned} \tag{6}$$

146 With  $\mathcal{A}$  and  $\mathcal{R}$  generally fitted to reduced order, i.e.  $k_A$  and  $k_R$  smaller, often much smaller, than  
147 the number of ages in the data, the resulting estimates of covariance matrices among observations  
148 are smoothed and have reduced rank, equal to the rank of the corresponding matrix of covariances  
149 among RR coefficients.

## 150 4.1 REML estimation

### 151 4.1.1 REML Likelihood

152 In its general form, based on the variance matrix of the observations, the REML log likelihood  
153  $(\log \mathcal{L})$  pertaining to model (2) is given by (e.g. Harville, 1977)

$$-2\log \mathcal{L} = \text{const} + \log |\mathbf{V}| + \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \tag{7}$$

154 Equivalently,  $\log \mathcal{L}$  can be written involving terms of the MME. For the orthogonal projection  
155 matrix  $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}^{-1}$ ,

$$\mathbf{P}\mathbf{y} = \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \boldsymbol{\Sigma}_\varepsilon^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \boldsymbol{\Phi}_A \hat{\boldsymbol{\alpha}} - \boldsymbol{\Phi} \hat{\boldsymbol{\gamma}}) = \boldsymbol{\Sigma}_\varepsilon^{-1} \hat{\boldsymbol{\varepsilon}} \tag{8}$$

156 (Harville, 1977), i.e.  $\mathbf{y}'\mathbf{P}\mathbf{y}$  is the weighted sum of squares of residuals. Let  $\mathbf{C}$  denote the coefficient  
157 matrix in (4), and rewrite (7) as (Graser et al., 1987)

$$-2\log \mathcal{L} = \text{const} + \log |\mathbf{G}| + \log |\mathbf{R}| + \log |\boldsymbol{\Sigma}_\varepsilon| + \log |\mathbf{C}| + \mathbf{y}'\mathbf{P}\mathbf{y} \tag{9}$$

158 The first three terms in (9) can be evaluated indirectly, requiring only the determinants of the  
159 covariance matrices of RR coefficients

$$\log |\mathbf{G}| = N_A \log |\mathbf{K}_A| + k_A \log |\mathbf{A}| \quad \log |\mathbf{R}| = N \log |\mathbf{K}_R| \quad \text{and} \quad \log |\Sigma_\varepsilon| = \sum_{k=1}^p N_k \log(\sigma_k^2) \quad (10)$$

160 with  $N_k$  the number of records assumed to have measurement error variance  $\sigma_k^2$ . For a given  
 161 analysis,  $\log |\mathbf{A}|$  is a constant and not required in locating the maximum of  $\log \mathcal{L}$ . If needed, for  
 162 instance for model comparisons, it can be readily obtained as a by-product of the procedure to  
 163 set up  $\mathbf{A}^{-1}$ . The other two terms,  $\mathbf{y}'\mathbf{P}\mathbf{y}$  and  $\log |\mathbf{C}|$ , can be evaluated simultaneously by factoring  
 164 the so-called 'mixed model matrix' (Graser et al., 1987), which is  $\mathbf{C}$  augmented by the vector of  
 165 right hand sides in (4) and its transpose, respectively, and the residual sum of squares.

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{X} & \mathbf{X}'\Sigma_\varepsilon^{-1}\Phi_A & \mathbf{X}'\Sigma_\varepsilon^{-1}\Phi & \mathbf{X}'\Sigma_\varepsilon^{-1}\mathbf{y} \\ \Phi_A'\Sigma_\varepsilon^{-1}\mathbf{X} & \Phi_A'\Sigma_\varepsilon^{-1}\Phi_A + \mathbf{A}^{-1} \otimes \mathbf{K}_A^{-1} & \Phi_A'\Sigma_\varepsilon^{-1}\Phi & \Phi_A'\Sigma_\varepsilon^{-1}\mathbf{y} \\ \Phi'\Sigma_\varepsilon^{-1}\mathbf{X} & \Phi'\Sigma_\varepsilon^{-1}\Phi_A & \Phi'\Sigma_\varepsilon^{-1}\Phi + \mathbf{I} \otimes \mathbf{K}_R^{-1} & \Phi'\Sigma_\varepsilon^{-1}\mathbf{y} \\ \mathbf{y}'\Sigma_\varepsilon^{-1}\mathbf{X} & \mathbf{y}'\Sigma_\varepsilon^{-1}\Phi_A & \mathbf{y}'\Sigma_\varepsilon^{-1}\Phi & \mathbf{y}'\Sigma_\varepsilon^{-1}\mathbf{y} \end{bmatrix} \quad (11)$$

166 For  $\mathbf{M} = \mathbf{L}\mathbf{L}'$ , of size  $M \times M$ , and  $l_{ii}$  the  $i$ -th diagonal element of  $\mathbf{L}$

$$\log |\mathbf{C}| = 2 \sum_{i=1}^{M-1} \log(l_{ii}) \quad \text{and} \quad \mathbf{y}'\mathbf{P}\mathbf{y} = l_{MM}^2 \quad (12)$$

#### 167 4.1.2 First derivatives

168 First derivatives of  $\log \mathcal{L}$  can be obtained similarly, differentiating (9) rather than (7). Let  
 169  $\mathbf{D}_Q^{mn} = \partial \mathbf{K}_Q / \partial K_{mn}$  denote the matrix of size  $k_Q \times k_Q$  with elements of unity in the  $mn$ -th and  
 170  $nm$ -th position and zero otherwise, with  $Q$  standing in turn for  $A$  and  $R$ . For  $\theta_k$  the  $k$ -th  
 171 variance component to be estimated

$$\frac{\partial \log |\mathbf{G}|}{\partial \theta_k} = N_A \operatorname{tr} (\mathbf{K}_A^{-1} \mathbf{D}_A^{mn}) = (2 - \delta_{mn}) N_A K_A^{mn} \quad \text{for } \theta_k = K_{Amn} \quad (13)$$

$$\frac{\partial \log |\mathbf{R}|}{\partial \theta_k} = N \operatorname{tr} (\mathbf{K}_R^{-1} \mathbf{D}_R^{mn}) = (2 - \delta_{mn}) N K_R^{mn} \quad \text{for } \theta_k = K_{Rmn} \quad (14)$$

$$\frac{\partial \log |\Sigma_\varepsilon|}{\partial \theta_k} = \sum_{i=1}^N \sigma_i^{-2} \partial \sigma_i^2 / \partial \theta_k = N_m \sigma_m^{-2} \quad \text{for } \theta_k = \sigma_m^2 \quad (15)$$



with  $K_Q^{mn}$  the  $mn$ -th element of  $\mathbf{K}_Q^{-1}$  (for  $Q = A, R$ ) and  $\delta_{mn}$  Kronecker's delta, i.e.  $\delta_{mn} = 1$  for  $m = n$  and 0 otherwise. Corresponding derivatives to (15) for the case where changes in  $\sigma^2$  with age are modelled through a polynomial variance function are given in Meyer (2001b).

Derivatives of  $\log |\mathbf{C}|$  and  $\mathbf{y}'\mathbf{P}\mathbf{y}$  can be determined through an 'automatic differentiation' of the Cholesky factor of  $\mathbf{M}$ , as described by Smith (1995). This requires the derivatives of  $\mathbf{M}$  with respect to the variance components to be estimated.

$$\frac{\partial \mathbf{M}}{\partial K_{A mn}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_A & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{with } \mathbf{Q}_A = \mathbf{A}^{-1} \otimes (-\mathbf{K}_A^{-1} \mathbf{D}_A^{mn} \mathbf{K}_A^{-1}) \quad (16)$$

$$\frac{\partial \mathbf{M}}{\partial K_{R mn}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{with } \mathbf{Q}_R = \mathbf{I}_N \otimes (-\mathbf{K}_R^{-1} \mathbf{D}_R^{mn} \mathbf{K}_R^{-1}) \quad (17)$$

$$\frac{\partial \mathbf{M}}{\partial \sigma_m^2} = \begin{bmatrix} \mathbf{X}'\mathbf{Q}_\varepsilon \mathbf{X} & \mathbf{X}'\mathbf{Q}_\varepsilon \Phi_A & \mathbf{X}'\mathbf{Q}_\varepsilon \Phi & \mathbf{X}'\mathbf{Q}_\varepsilon \mathbf{y} \\ \Phi_A' \mathbf{Q}_\varepsilon \mathbf{X} & \Phi_A' \mathbf{Q}_\varepsilon \Phi_A & \Phi_A' \mathbf{Q}_\varepsilon \Phi & \Phi_A' \mathbf{Q}_\varepsilon \mathbf{y} \\ \Phi' \mathbf{Q}_\varepsilon \mathbf{X} & \Phi' \mathbf{Q}_\varepsilon \Phi_A & \Phi' \mathbf{Q}_\varepsilon \Phi & \Phi' \mathbf{Q}_\varepsilon \mathbf{y} \\ \mathbf{y}' \mathbf{Q}_\varepsilon \mathbf{X} & \mathbf{y}' \mathbf{Q}_\varepsilon \Phi_A & \mathbf{y}' \mathbf{Q}_\varepsilon \Phi & \mathbf{y}' \mathbf{Q}_\varepsilon \mathbf{y} \end{bmatrix} \quad \text{with } \mathbf{Q}_\varepsilon = -\Sigma_\varepsilon^{-1} \frac{\partial \Sigma_\varepsilon}{\partial \sigma_i^2} \Sigma_\varepsilon^{-1} \quad (18)$$

Matrices  $-\mathbf{K}_Q^{-1} \mathbf{D}_Q^{mn} \mathbf{K}_Q^{-1}$  have elements  $-(2 - \delta_{mn})/2 (K_Q^{rm} K_Q^{ns} + K_Q^{rn} K_Q^{ms})$  for  $r, s = 1, \dots, k_Q$  and  $Q = A, R$  (Meyer, 2001b). Similarly,  $\mathbf{Q}_\varepsilon$  is diagonal with elements  $-(\sigma_m^2)^{-2}$  for records which have variance  $\sigma_m^2$  and zero otherwise.

For  $\partial \mathbf{M} / \partial \theta_k = (\partial \mathbf{L} / \partial \theta_k)(\partial \mathbf{L}' / \partial \theta_k)$ , derivatives of  $\log |C|$  and  $\mathbf{y}'\mathbf{P}\mathbf{y}$  are simple functions of the diagonal element of  $\partial \mathbf{L} / \partial \theta_k$ ,  $\partial l_{ii} / \partial \theta_k$  (Meyer & Smith, 1996),

$$\partial \log |\mathbf{C}| / \partial \theta_k = 2 \sum_{i=1}^{M-1} l_{ii}^{-1} \partial l_{ii} / \partial \theta_k \quad \text{and} \quad \partial \mathbf{y}'\mathbf{P}\mathbf{y} / \partial \theta_k = l_{MM} \partial l_{MM} / \partial \theta_k \quad (19)$$

### 4.1.3 ‘Average’ information

Second derivatives of  $\log \mathcal{L}$  can be obtained analogously (Smith, 1995; Meyer & Smith, 1996), but computations are demanding. As noted by Gilmour et al. (1995) and Johnson & Thompson (1995), the average of observed and expected information (AI) is proportional to second derivatives of  $\mathbf{y}'\mathbf{P}\mathbf{y}$ , the ‘data part’ of  $\log \mathcal{L}$ , and thus considerably easier to compute than either of the former,

$$\frac{\partial^2 \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial \theta_k \partial \theta_m} = \mathbf{b}_k' \mathbf{P} \mathbf{b}_m \quad (20)$$

with

$$\mathbf{b}_k = \frac{\partial \mathbf{V}}{\partial \theta_k} \mathbf{P} \mathbf{y} = \begin{cases} \Phi_A (\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{mn}) \hat{\boldsymbol{\alpha}} & \text{for } \theta_k = K_{A mn} \\ \Phi (\mathbf{I}_N \otimes \mathbf{D}_R^{mn}) \hat{\boldsymbol{\gamma}} & \text{for } \theta_k = K_{R mn} \\ \mathbf{D}_m^* (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \Phi_A \hat{\boldsymbol{\alpha}} - \Phi \hat{\boldsymbol{\gamma}}) & \text{for } \theta_k = \sigma_m^2 \end{cases} \quad (21)$$

where  $\mathbf{D}_m^*$  is a diagonal matrix with elements  $\sigma_m^{-2}$  for records which have measurement error variance  $\sigma_m^2$ , and zero otherwise.

Let  $\mathbf{B}$  be the matrix of column vectors  $\mathbf{b}_k$ , and expand  $\mathbf{M}$  to  $\mathbf{M}_B$  by replacing  $\mathbf{y}$  and  $\mathbf{y}'$  in  $\mathbf{M}$  by  $\mathbf{B}$  and  $\mathbf{B}'$ , respectively. Absorbing rows 1 to  $M-1$  of  $\mathbf{M}_B$  then replaces elements of  $\mathbf{B}'\Sigma_\varepsilon^{-1}\mathbf{B}$  with  $\mathbf{b}_k' \mathbf{P} \mathbf{b}_m$  (Meyer, 1997). With  $\mathbf{M}$  already factored in evaluating  $\log \mathcal{L}$ , additional computations needed are undemanding. Alternatively, as outlined by Johnson & Thompson (1995), terms  $\mathbf{P} \mathbf{b}_k$  can be obtained by replacing  $\mathbf{y}$  in (4) by  $\mathbf{b}_k$ , solving (4) and computing the vector of residuals, pre-multiplied by  $\Sigma_\varepsilon^{-1}$  (see (8)). Elements of the average information matrix, given by (20) can then be evaluated as simple vector products.

### 4.1.4 Maximising the likelihood

The maximum of  $\log \mathcal{L}$  (9) can be located using a derivative-free search strategy (Meyer, 1989). However, with many, highly correlated parameters to be estimated, methods which utilise information from derivatives of the likelihood are advantageous. If first derivatives and AI are available, they can be employed in a modified Newton-Raphson procedure to maximise  $\log \mathcal{L}$ , using a modification of step size, either through step halving (Jennrich & Schluchter, 1986) or Marquardt (1963)’s method, to ensure  $\log \mathcal{L}$  is increased in round of each iteration.

205 A reparameterisation to elements of the Cholesky decompositions of  $\mathbf{K}_A$  and  $\mathbf{K}_R$ , taking loga-  
 206 rithms of their diagonal elements and also of  $\sigma_k$  removes constraints on the parameters and allows  
 207 an unconstrained optimisation such as a Newton-Raphson procedure to be used. A gradient vec-  
 208 tor and Hessian matrix for a Cholesky factor of a matrix can readily be obtained from derivatives  
 209 with respect to elements of the matrix (Lindstrom & Bates, 1988) by pre- and postmultiplying  
 210 with the appropriate Jacobian matrix; see Meyer & Smith (1996) for details.

## 211 4.2 Bayesian estimation

212 Bayesian estimates for the simple RR model considered here are readily obtained using Gibbs  
 213 Sampling. This is computationally easy to implement, and, whilst computing time required to  
 214 sample long or multiple Markov chains may be large, memory requirements are much smaller  
 215 than for corresponding REML analysis, facilitating analyses of large data sets. This is particu-  
 216 larly important for models involving multiple random effects and high orders of fit for RRs, i.e  
 217 requiring numerous variance components to be estimated. In addition, as estimates of complete  
 218 posterior distributions are obtained, sampling properties of estimates or functions thereof are  
 219 readily examined. More complicated models can be accommodated, but may require sampling  
 220 from non-standard distributions involving Metropolis-Hastings steps (Sorensen, 1996; Sorensen  
 221 & Gianola, 2002). Gibbs Sampling schemes for RR models applied to test day records of dairy  
 222 cows have been described in detail by Jamrozik & Schaeffer (1997) and Rekaya et al. (1999).

### 223 4.2.1 Conditional distributions

224 Location parameters  $\mathbf{b}$ ,  $\boldsymbol{\alpha}$  and  $\boldsymbol{\gamma}$  can be sampled sequentially from their fully conditional distri-  
 225 butions. For RR coefficients which typically are highly correlated, it is advantageous to sample  
 226 blocks of coefficients pertaining to an individual simultaneously. Let  $\mathbf{u}$  denote the vector of lo-  
 227 cation parameters and  $\mathbf{r}$  the vector or right hand sides in (4), so that the MME can be written  
 228 as  $\mathbf{C}\hat{\mathbf{u}} = \mathbf{r}$ . Consider the subvector  $\mathbf{u}_s$  to be sampled, and let  $\mathbf{u}_{-s}$  denote its complement. Con-  
 229 ceptually reorder equations so that  $\mathbf{u}' = (\mathbf{u}'_s | \mathbf{u}'_{-s})$ , and partition  $\mathbf{r}$  in the same way as  $\mathbf{u}$  and  
 230  $\mathbf{C}$  accordingly into  $\mathbf{C}_{s,s}$ ,  $\mathbf{C}_{s,-s}$ ,  $\mathbf{C}_{-s,s}$  and  $\mathbf{C}_{-s,-s}$ . For  $\mathbf{y}$  normally distributed and assuming a  
 231 uniform prior distribution for  $\mathbf{b}$ , the fully conditional distribution of  $\mathbf{u}_s$  is (Sorensen, 1996)

$$\mathbf{u}_s \sim N(\mathbf{C}_{s,s}^{-1}(\mathbf{r}_s - \mathbf{C}_{s,-s}\hat{\mathbf{u}}_{-s}), \mathbf{C}_{s,s}^{-1}) \quad (22)$$

Variances are sampled from scaled, inverted Wishart (IW) or  $\chi^2$  distributions. Assume prior inverse Wishart distributions,  $IW_k(\Sigma^{-1}, \nu)$ , with  $\Sigma_\alpha$  and  $\Sigma_\gamma$  and  $\nu_\alpha$  and  $\nu_\gamma$  for  $\mathbf{K}_A$  and  $\mathbf{K}_R$ , respectively, and let  $\mathbf{S}_\alpha = \{\boldsymbol{\alpha}'_k \mathbf{A}^{-1} \boldsymbol{\alpha}_m\}$  and  $\mathbf{S}_\gamma = \{\boldsymbol{\gamma}'_k \boldsymbol{\gamma}_m\}$  denote corresponding matrices of sums of squares and crossproducts, with  $\boldsymbol{\alpha}_m$  and  $\boldsymbol{\gamma}_m$  denoting the subvectors of  $\boldsymbol{\alpha}$  and  $\boldsymbol{\gamma}$  gathering RR coefficients  $\alpha_{im}$  (for  $i = 1, \dots, N_A$ ) and  $\gamma_{im}$  (for  $i = 1, \dots, N$ ). This gives conditional distributions (Jamrozik & Schaeffer, 1997),

$$\mathbf{K}_A \sim IW_{k_A}((\mathbf{S}_\alpha + \Sigma_\alpha)^{-1}, N_A + \nu_\alpha) \quad \mathbf{K}_R \sim IW_{k_R}((\mathbf{S}_\gamma + \Sigma_\gamma)^{-1}, N + \nu_\gamma) \quad (23)$$

Similarly, let  $\nu_{\varepsilon m}$  and  $\tilde{\sigma}_m^2$  denote the hyperparameters for the prior distribution of the  $m$ -th measurement error variance. This gives fully condition distribution (Jamrozik & Schaeffer, 1997)

$$\sigma_m^2 \sim \chi_{\nu_{\varepsilon m} + N_m}^{-2}(\tilde{\sigma}_m^2 \nu_{\varepsilon m} + \hat{\boldsymbol{\varepsilon}}'_m \hat{\boldsymbol{\varepsilon}}_m) \quad (24)$$

with  $\hat{\boldsymbol{\varepsilon}}_m$  of length  $N_m$  the subvector of residuals for records assumed to have error variance  $\sigma_m^2$ .

#### 4.2.2 Sampling scheme

The Gibbs sampler then consists of repeated, sequential sampling of all effects fitted and variance components from their fully conditional distributions. Location parameters can be sampled blockwise as described above or simultaneously (García-Cortés & Sorensen, 1996) to reduce auto-correlations between samples. Estimates are generally obtained as posterior means, i.e. means of samples disregarding an initial “burn-in” phase (Sorensen, 1996).

## 5 Eigenfunctions and beyond

In addition to defining genetic covariances for all points along the continuous scale considered, genetic CFs provide an insight into how selection is likely to change the trait or trajectory. This is of interest to evolutionary biologist who are concerned with the effects of natural selection,

and animal breeders who want to predict how effective their selection of parents of the next generation will be and how selection at any point might affect the remainder of the curve.

An eigenvalue decomposition of the covariance matrix of genetic RR coefficients

$$\mathbf{K}_A = \mathbf{\Omega} \text{Diag}\{\lambda_k\} \mathbf{\Omega}' = \sum_{k=1}^{k_A} \lambda_k \boldsymbol{\omega}_k \boldsymbol{\omega}_k' \quad (25)$$

yields estimates of the eigenvalues  $\lambda_k$  and eigenfunctions  $\xi_k$  of  $\mathcal{A}$  (Kirkpatrick & Heckman, 1989)

$$\xi_k(a_i) = \boldsymbol{\omega}_k' \boldsymbol{\phi}_{Ai} = \sum_{m=0}^{k_A-1} \omega_{km} \phi_m(a_i) \quad (26)$$

with  $\boldsymbol{\omega}_k = \{\omega_{km}\}$  the eigenvector of  $\mathbf{K}_A$  pertaining to the  $k$ -th eigenvalue, and  $\mathbf{\Omega}$  the orthogonal matrix of eigenvectors.

The eigenvector corresponding to the largest eigenvalue  $\lambda_1$  gives the linear function of RR coefficients which explains most genetic variation, and the corresponding eigenfunction  $\xi_1(a_i)$  summarises the direction of changes in the trajectory when selecting on the basis of this quantity. This implies that greatest response to selection is achieved if the selection objective is of similar form than  $\xi_1(a_i)$ , and, conversely, that it may be difficult to achieve changes corresponding to eigenfunctions with low eigenvalues. In livestock improvement schemes we commonly attach different importance, i.e. economic weights, to different traits considered. In a RR framework, it seems logical to consider selection on the basis of individual eigenfunctions, choosing appropriate weights for each  $\xi_i$  to achieve the desired deformation of the whole trajectory. van der Werf (2002) gives an example for growth of beef cattle.

Selection pressure is commonly measured as selection differential, i.e. the difference in means between selected individuals and all candidates for selection. Selection at any point of the trajectory creates correlated differentials at other points. Evolutionary biologist thus define a selection gradient function ( $\zeta$ ), which summarises the force of directional selection along the complete trajectory. Kirkpatrick et al. (1990), Beder & Gomulkiewicz (1998) and de Jong & Bijma (2002) consider estimation of  $\zeta$  and its approximation when relative fitness cannot be measured directly, and Kingsolver et al. (2001) give a detailed example. The estimated change in the mean trajectory is then  $\Delta = \sum_{i=0}^{k_A-1} \lambda_i (\boldsymbol{\omega}_i' \zeta) \boldsymbol{\omega}_i$  (Kirkpatrick & Heckman, 1989).

From (25) it follows that the first few eigenfunctions are generally sufficient to explain most genetic variation. Hence little information is lost, if eigenfunctions corresponding to eigenvalues

277 close to zero are ignored in selection decisions. This can lead to substantial computational savings.  
 278 This is readily implemented by reparameterising (4) to estimate  $\boldsymbol{\alpha}^* = (\mathbf{I}_{N_A} \otimes \boldsymbol{\Omega}') \boldsymbol{\alpha}$ , replacing  
 279  $\boldsymbol{\Phi}_A$  with  $\boldsymbol{\Phi}_A^* = \boldsymbol{\Phi}_A (\mathbf{I}_{N_A} \otimes \boldsymbol{\Omega})$  and  $\mathbf{K}_A$  with  $\boldsymbol{\Lambda}$ , and deleting equations for  $\alpha_{im}^*$  corresponding  
 280 to the smallest eigenvalues  $\lambda_m$ . This is analogous to the reduced-rank approach of James et al.  
 281 (2000). Note that  $\boldsymbol{\Phi}_A^* \boldsymbol{\alpha}^* = \boldsymbol{\Phi}_A \boldsymbol{\alpha}$  when no equations are deleted, i.e. that of breeding value  
 282 estimates at individual ages are the same as those obtained above (5). Further, using  $\boldsymbol{\Phi}_A^*$   
 283 eigenvectors are incorporated in the incidence matrix, i.e. this parameterisation directly yields  
 284 estimates of genetic values for the eigenfunctions. An analogous reparameterisation and reduction  
 285 in the number of effects to be estimated can be applied to other random effects such as  $\gamma$ . This  
 286 has been considered as a computational strategy for RR analyses of test-day records of dairy cows  
 287 (Miszta et al., 2000). Alternatively, the number of effects to be estimated can be reduced fitting  
 288 an equivalent model which incorporates permanent environmental effects due to the animal in  
 289 the vector of residuals (Meyer, 2001b).

## 290 6 Example

291 Application of a RR model analysis to model growth of beef cattle is illustrated using results from  
 292 an analysis by Meyer (2001a). A similar analysis considering a much larger data set and range  
 293 of ages and using Bayesian estimation, demonstrating the kind of analyses feasible, is presented  
 294 by Meyer (2002).

295 REML estimates of variance components were obtained as outlined above for weights of beef  
 296 calves in a selection experiment, recorded at monthly intervals from birth to 280 days of age.  
 297 The RR model fitted regressions on Legendre polynomials of age at recording for four random  
 298 effects. These were animals' direct genetic and environmental effects as in (1), modelled through  
 299 quartic ( $k_A = 5$ ) and quintic ( $k_R = 6$ ) polynomials, respectively. In addition, maternal genetic  
 300 (M) and permanent environmental (C) effects had to be taken into account, as records were  
 301 predominantly taken prior to weaning. Hence two additional sets of RR with  $k_M = k_C = 3$   
 302 (quadratic polynomials) were included in the model of analysis. Changes in  $\sigma_k^2$  were modelled as  
 303 a quadratic function of age at weighing, yielding a total of 52 parameters to be estimated. This  
 304 was one of numerous models considered, which was selected on the basis of the REML form of  
 305 the so-called Bayesian Information Criterion (Wolfinger, 1993) and plausibility of the resulting  
 306 variance estimates along the trajectory.

307 Data consisted of 21,053 records for  $N = 3,417$  calves. These were offspring of 1,023 dams and  
 308 174 sires, and including parents without records yielded a total of  $N_A = 3,794$  animals in the  
 309 analysis. Figure 1 shows estimates of the direct variance components for the ages in the data  
 310 derived from estimated CFs and measurement error variances, together with estimates of the first  
 311 three eigenfunctions for direct genetic effects and estimates of direct genetic correlations with  
 312 contour lines from 0.95,  $\dots$ , 0.65 in steps of 0.05. Results show increasing variances with age and  
 313 size (means not shown), with most variation attributable to permanent environmental effects of  
 314 animals. Estimates of genetic correlations between weights at different ages were high to very  
 315 high except for weights recorded at birth and during the first month of life. Consequently, the  
 316 first eigenfunction explained 95.9% of the total genetic variation, with eigenvalues for the second  
 317 and third eigenfunction amounting to 3.1% and 0.7%, respectively. As found in other studies of  
 318 growth traits (e.g. Kirkpatrick et al., 1990), the first eigenfunction was positive for the complete  
 319 trajectory considered, i.e. selection for increased weight at any age is expected to increase weights  
 320 at all ages.

## 321 7 Discussion

322 As outlined above, uptake of RR models in animal breeding has been stimulated by evolutionary  
 323 biology, in particular the work of Kirkpatrick and co-workers (e.g. Kirkpatrick & Heckman, 1989;  
 324 Kirkpatrick et al., 1990), and the desire to model complete lactation curves in genetic evaluation  
 325 of dairy cattle.

326 In contrast to animal breeding applications, analyses of longitudinal or similar data in other  
 327 areas of applied statistics have long since dealt with covariance functions and patterned covari-  
 328 ance matrices. Most commonly, this involved assumptions of a parametric correlation structure  
 329 for ‘within-subject’ covariances between repeated records for an individual. Stationary models  
 330 described by one or few parameters, such as the well known auto-regressive correlation struc-  
 331 ture, were favoured and variances were often assumed homogeneous (Jennrich & Schluchter, 1986;  
 332 Wolfinger, 1996), though more recently non-stationary models with heterogeneous variances have  
 333 attracted increased attention, especially the so-called structured ante-dependence model (Núñez-  
 334 Antón & Zimmerman, 2000) and related parameterisations (Pourahmadi, 1999; Daniels & Zhao,  
 335 2003). Whilst RR models, generally referred to a ‘random coefficient’ models, have been con-  
 336 sidered, they have been less widely used, as they were often found to be of little advantage,

337 requiring high orders of fit and thus many parameters to be estimated, and yielding less readily  
338 interpretable results than a parametric correlation function.

339 Animal breeders, however, by and large have embraced RR models for the analysis of longitudinal  
340 data. There are various reasons. Firstly, quantitative genetic analyses are concerned with the  
341 variation between animals, while other areas of statistics are often content with modelling within-  
342 subject covariances only. Fitting a set of additive genetic RR coefficients provides estimates of  
343 genetic merit for the whole range of ages considered, and allows ranking of animals to change  
344 with time. RR models are thus an obvious choice if we are concerned with (genetic) differences  
345 between individuals. Assuming a RR model and resulting covariance structure on a genetic  
346 level, it seems natural to apply the same model to other random effects such as permanent  
347 environmental effects. RR models account for changes in variances with time and do not require  
348 specific assumptions about the shape of the resulting CF, other than implied by the choice of  
349 covariables.

350 Secondly, estimates of genetic covariance matrices arising from RR model analyses can be thought  
351 of as smoothed versions of corresponding estimates from an unstructured, multivariate analysis  
352 treating records at different ages as different traits. Estimates of the eigenvalues and eigenfunc-  
353 tions of CFs can be obtained directly from estimates of covariances among RR coefficients. For  
354 genetic covariance functions, these statistics provide valuable insight into the effects of selection  
355 for the trait considered, and facilitate design of selection schemes optimising response along the  
356 trajectory.

357 Last, but not least, RR models fit well into the methodological framework commonly employed by  
358 animal breeders, and, moreover, provide a computationally feasible way to estimate CF for large  
359 data sets with records coming in at ‘all ages’, as are typical for data from livestock recording  
360 schemes. Estimating coefficients of CFs as the matrix of covariances between RR coefficients  
361 requires manipulation of mixed model equations of size proportional to the number of regression  
362 coefficients to be manipulated, rather than proportional to the number of ages or even the number  
363 of records.

364 Nevertheless, computational requirements of RR analyses can be large, in particular for variance  
365 component estimation as, typically, the number of parameters to be estimated is substantial.  
366 In addition, the higher order polynomials commonly fitted in RR analysis are notoriously ill-  
367 behaved, in particular at the limits of the range of ages considered. Kirkpatrick et al. (1994), for



instance, encountered ‘wiggly’ covariance function estimates for test-day records in dairy cattle, and erratically high estimates of variances for the highest ages with least records have been reported repeatedly (e.g. Meyer, 2002).

Hence attempts have been made recently to apply the more parsimonious parametric models to genetic analyses. Pletcher & Geyer (1999) modelled both genetic and environmental covariances between records at different ages through a variety of one- or two parameter correlation function, combined with polynomial variance functions, but emphasized computational limitations in the number of distinct ages which could be accommodated. Applying such model to test-day records from dairy cows, Jaffrézic et al. (2002) found higher likelihoods for parametric models than RR models involving more parameters, but encountered problems with substantial overestimates of the genetic variance from parametric models fitted in conjunction with animal model analyses. Foulley et al. (2000) combined a RR model for genetic effects with an auto-regressive covariance structure for residual effects, and Meyer (2001b) extended this approach to non-stationary correlation functions and heterogeneous variances. Results from an application to beef cattle data showed good agreement with estimates of variances and correlations from corresponding RR analyses, and alleviated problems of somewhat erratic estimates at the highest ages with very few records (Albuquerque & Meyer, 2002), suggesting that such hybrid models may be well suited to genetic analyses of longitudinal data.

Spline functions may provide a better behaved alternative than polynomial function. B-splines have been suggested (Rice & Wu, 2001) but have had limited applications for RR analyses in animal breeding so far (e.g. Torres & Quaas, 2001). Cubic smoothing splines integrate well into the mixed model framework and require few parameters to be estimated, as it is assumed that all only intercepts and linear coefficients are correlated and that all other coefficients are uncorrelated and have equal variance (Verbyla et al., 1999). They are finding increasing use (e.g. White et al., 1999; Huisman et al., 2002), especially as they are readily fitted within the ASREML software package (Gilmour et al., 2001). Other RR models imposing a structure on the covariance matrix of RR coefficients have been suggested. Robert-Granié et al. (2002), for example, modelled growth of beef cattle allowing for heterogeneous variances across environments, with changes in variances described by a link function. Attempting to reduce the number of parameters to be estimated, Meyer (1997) forced estimates of covariance matrices among RR to have reduced rank. More research on such structured models is required, in particular on the effects of the resulting covariance structure along the trajectory.

400

## 8 Conclusions

401

402

403

404

405

Analyses of longitudinal data or 'functional analyses' when data are curves is a rapidly evolving field of research. In animal breeding applications, where we are concerned with genetic differences between animals, random regression models provide a useful framework to not only distinguish between trajectories for individuals, but also account for changes in variances and correlations along the trajectory.

406

### Acknowledgments

407

This work was supported by grant BFGEN.100 of Meat and Livestock Australia (MLA).

408

## Résumé

409

410

On describe des modèles de régression aléatoire pour analyse des données longitudinales en génétique animale.

## References

- Albuquerque, L. G. & Meyer, K. (2002). Estimates of genetic covariance functions for growth assuming a parametric correlation structure for permanent environmental effects. *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.*. Communication No. 2–21.
- Beder, J. H. & Gomulkiewicz, R. (1998). Computing the selection gradient and evolutionary response of an infinite-dimensional trait. *J. Math. Biol.*, **36**, 299–319.
- Daniels, M. J. & Zhao, Y. D. (2003). Modelling the random effects covariance matrix in longitudinal data. *Statistics in Medicine*, **22**, 1631–1647.
- de Jong, G. & Bijma, P. (2002). Selection and phenotypic plasticity in evolutionary biology and animal breeding. *Livest. Prod. Sci.*, **78**, 195–214.
- Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Longman Group, London, 4th ed.
- Foulley, J. L., Jaffrézic, F. & Robert-Granié, C. (2000). EM-REML estimation of covariance parameters in Gaussian mixed models for longitudinal data analysis. *Genet. Select. Evol.*, **32**, 129–141.
- García-Cortés, L. & Sorensen, D. (1996). On a multivariate implementation of the Gibbs sampler. *Genet. Select. Evol.*, **28**, 121–126.
- Gilmour, A., Gogel, B., Cullis, B. R., Welham, S. J. & Thompson, R. (2001). *ASREML User Guide Release 1.0*. VSN International Ltd., Hemel Hempstead, HP1 1ES, U.K.
- Gilmour, A. R., Thompson, R. & Cullis, B. R. (1995). Average Information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, **51**, 1440–1450.
- Graser, H. U., Smith, S. P. & Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.*, **64**, 1362–1370.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and related problems. *J. Amer. Stat. Ass.*, **72**, 320–338.

438 Henderson, C. R. (1973). Sire evaluation and genetic trends. In *Proc. Anim. Breed. Genet. Symp.*  
439 *in Honor of Dr. J.L. Lush*. ASAS and ADSA, Champaign, Illinois, pages 10–41.

440 Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship  
441 matrix used in prediction of breeding values. *Biometrics*, **32**, 69–83.

442 Henderson, C. R. jnr. (1982). Analysis of covariance in the mixed model : Higher-level, non-  
443 homogeneous and random regressions. *Biometrics*, **38**, 623–640.

444 Huisman, A. E., Veerkamp, R. F. & van Arendonk, J. A. M. (2002). Genetic parameters for  
445 various random regression models to describe the weight data of pigs. *J. Anim. Sci.*, **80**,  
446 575–582.

447 Jaffrézic, F., White, I. M. S., Thompson, R. & Visscher, P. M. (2002). Contrasting models for  
448 lactation curve analysis. *J. Dairy Sci.*, **84**, 968–975.

449 James, G. M., Hastie, T. J. & Sugar, C. A. (2000). Principal component models for sparse  
450 functional data. *Biometrika*, **87**, 587–602.

451 Jamrozik, J. & Schaeffer, L. R. (1997). Estimates of genetic parameters for a test day model with  
452 random regressions for yield traits of first lactation Holsteins. *J. Dairy Sci.*, **80**, 762–770.

453 Jamrozik, J., Schaeffer, L. R. & Dekkers, J. C. M. (1997). Genetic evaluation of dairy cattle  
454 using test day yields and random regression model. *J. Dairy Sci.*, **80**, 1217–1226.

455 Jennrich, R. I. & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured  
456 covariance matrices. *Biometrics*, **42**, 805–820.

457 Johnson, D. L. & Thompson, R. (1995). Restricted maximum likelihood estimation of vari-  
458 ance components for univariate animal models using sparse matrix techniques and average  
459 information. *J. Dairy Sci.*, **78**, 449–456.

460 Kingsolver, J. G., Gomulkiewicz, R. & Carter, P. A. (2001). Variation, selection and evolution  
461 of function-valued traits. *Genetica*, **112–113**, 87–104.

462 Kirkpatrick, M. & Heckman, N. (1989). A quantitative genetic model for growth, shape, reaction  
463 norms, and other infinite-dimensional characters. *J. Math. Biol.*, **27**, 429–450.

464 Kirkpatrick, M., Lofsvold, D. & Bulmer, M. (1990). Analysis of the inheritance, selection and  
465 evolution of growth trajectories. *Genetics*, **124**, 979–993.

466 Kirkpatrick, M., Hill, W. G. & Thompson, R. (1994). Estimating the covariance structure of  
467 traits during growth and aging, illustrated with lactations in dairy cattle. *Genet. Res.*, **64**,  
468 57–69.

469 Lindstrom, M. J. & Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-  
470 effects models for repeated-measures data. *J. Amer. Stat. Ass.*, **83**, 1014–1022.

471 Lynch, M. & Walsh, J. B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Assoc.s,  
472 Inc., Sunderland, MA.

473 Marquardt, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters.  
474 *SIAM J.*, **11**, 431–411.

475 Meyer, K. (1989). Restricted maximum likelihood to estimate variance components for animal  
476 models with several random effects using a derivative-free algorithm. *Genet. Select. Evol.*,  
477 **21**, 317–340.

478 Meyer, K. (1997). An “average information” restricted maximum likelihood algorithm for esti-  
479 mating reduced rank genetic covariance matrices or covariance functions for animal models  
480 with equal design matrices. *Genet. Select. Evol.*, **29**, 97–116.

481 Meyer, K. (1998). Estimating covariance functions for longitudinal data using a random regression  
482 model. *Genet. Select. Evol.*, **30**, 221–240.

483 Meyer, K. (2001a). Estimates of direct and maternal covariance functions for growth of Australian  
484 beef calves from birth to weaning. *Genet. Select. Evol.*, **33**, 487–514.

485 Meyer, K. (2001b). Estimating genetic covariance functions assuming a parametric correlation  
486 structure for environmental effects. *Genet. Select. Evol.*, **33**, 557–585.

487 Meyer, K. (2002). Estimates of covariance functions for growth of Australian beef cattle from  
488 a large set of field data. *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.*.  
489 Communication No. 11–01.

490 Meyer, K. & Hill, W. G. (1997). Estimation of genetic and phenotypic covariance functions for  
491 longitudinal data by restricted maximum likelihood. *Livest. Prod. Sci.*, **47**, 185–200.

492 Meyer, K. & Smith, S. P. (1996). Restricted maximum likelihood estimation for animal models  
493 using derivatives of the likelihood. *Genet. Select. Evol.*, **28**, 23–49.

494 Misztal, I., Strabel, T., Jamrozik, J., Mäntysaari, E. A. & Meuwissen, T. H. E. (2000). Strategies  
495 for estimating the parameters needed for different test-day models. *J. Dairy Sci.*, **83**, 1125–  
496 1134.

497 Núñez-Antón, V. & Zimmerman, D. L. (2000). Modelling non-stationary longitudinal data.  
498 *Biometrics*, **56**, 699–705.

499 Pletcher, S. D. & Geyer, C. J. (1999). The genetic analysis of age-dependent traits : Modeling  
500 the character process. *Genetics*, **153**, 825–835.

501 Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data :  
502 Unconstrained parameterisation. *Biometrika*, **86**, 677–690.

503 Quaas, R. L. (1976). Computing the diagonal elements of a large numerator relationship matrix.  
504 *Biometrics*, **32**, 949–953.

505 Rekaya, R., Carabaño, M. J. & Toro, M. A. (1999). Use of test day yields for the genetic  
506 evaluation of production traits in Holstein-Friesian cattle. *Livest. Prod. Sci.*, **57**, 203–217.

507 Rice, J. A. & Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled  
508 noisy curves. *Biometrics*, **57**, 253–259.

509 Robert-Granié, C., Heude, B. & Foulley, J. L. (2002). Modelling the growth curve of Maine-  
510 Anjou beef cattle using heteroskedastic random regression models. *Genet. Select. Evol.*,  
511 **34**, 423–445.

512 Schaeffer, L. R. & Dekkers, J. C. M. (1994). Random regressions in animal models for test-day  
513 production in dairy cattle. In *Proceeding Fifth World Congr. Genet. Appl. Livest. Prod.*,  
514 Vol. 18. pages 443–446.

515 Smith, S. P. (1995). Differentiation of the Cholesky algorithm. *J. Comp. Graph. Stat.*, **4**, 134–147.

516 Sorensen, D. (1996). *Gibbs Sampling in Quantitative Genetics*. Internal report no. 82, Danish  
517 Institute of Animal Science.

- 518 Sorensen, D. & Gianola, D. (2002). *Likelihood, Bayesian and MCMC Methods in Quantitative*  
519 *Genetics*. Springer Verlag.
- 520 Torres, R. A. A. & Quaas, R. L. (2001). Determination of covariance functions for lactation  
521 traits on dairy cattle using random-coefficient regressions on B-splines. *J. Anim. Sci.*, **79**,  
522 Suppl.1, 112. Abstract.
- 523 Tsuruta, S., Misztal, I. & Strandén, I. (2001). Use of preconditioned conjugate gradient algorithm  
524 as generic solver for mixed-model equations in animal breeding applications. *J. Anim. Sci.*,  
525 **79**, 1166–1172.
- 526 van der Werf, J. H. J. (2002). Optimising selection for traits along trajectories. *CD-ROM Seventh*  
527 *World Congr. Genet. Appl. Livest. Prod.*. Communication No. 16–07.
- 528 Verbyla, A. R., Cullis, B. R., Kenward, M. G. & Welham, S. J. (1999). The analysis of designed  
529 experiments and longitudinal data by using smoothing splines (with discussion). *Appl.*  
530 *Stat.*, **48**, 269–311.
- 531 White, I. M. S., Thompson, R. & Brotherstone, S. (1999). Genetic and environmental smoothing  
532 of lactation curves with cubic splines. *J. Dairy Sci.*, **82**, 632–638.
- 533 Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Comm. Stat. -*  
534 *Simul. Comp.*, **22**, 1079–1106.
- 535 Wolfinger, R. D. (1996). Heterogeneous variance-covariance structures for repeated measures  
536 data. *J. Agric. Biol. Environ. Stat.*, **1**, 205–230.

**Figure 1.** Estimates of variance components (top left), the first three direct genetic eigenfunctions (bottom left), and direct genetic correlations (right) for growth of beef cattle from birth to 280 days.

