Original article

Estimating covariance functions for longitudinal data using a random regression model

Karin Meyer*

Institute of Cell, Animal and Population Biology, Edinburgh University,
West Mains Road, Edinburgh EH9 3JT, Scotland, UK

Abstract – A method is described to estimate genetic and environmental covariance functions for traits measured repeatedly per individual along some continuous scale, such as time, directly from the data by restricted maximum likelihood. It relies on the equivalence of a covariance function and a random regression model. By regressing on random, orthogonal polynomials of the continuous scale variable, the coefficients of covariance functions can be estimated as the covariances among the regression coefficients. A parameterisation is described which allows the rank of estimated covariance matrices and functions to be restricted, thus facilitating a highly parsimonious description of the covariance structure. The procedure and the type of results which can be obtained are illustrated with an application to mature weight records of beef cows. © Inra/Elsevier, Paris

covariance functions / genetic parameters / longitudinal data / restricted maximum likelihood / random regression model

Résumé – Estimation des fonctions de covariance de données en séquence à partir d'un modèle à coefficients de régression aléatoires. On décrit une méthode d'estimation des fonctions de covariance génétique et non génétiques pour des caractères mesurés plusieurs fois par individu le long d'une échelle continue, comme le temps. Elle s'appuie directement sur les données à partir du maximum de vraisemblance restreint, en considérant l'équivalence entre fonction de covariance et modèle de régression aléatoire. Les coefficients figurant dans les fonctions de covariance peuvent être estimés comme des covariances entre les coefficients de régression des observations par rapport à des polynômes orthogonaux de la variable temporelle. On décrit un paramétrage qui permet de diminuer le rang des matrices et des fonctions de covariances, rendant ainsi possible une bonne description de la structure de covariance

* Correspondence and reprints: Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia.
E-mail: kmeyer@didgeridoo.une.edu.au

avec peu de paramètres. La procédure et le type de résultats qui peuvent être obtenus sont illustrés par un exemple concernant les poids vifs adultes de vaches allaitantes. © Inra/Elsevier, Paris

1. INTRODUCTION

Covariance functions have been recognized as a suitable alternative to the conventional multivariate mixed model to describe genetic and phenotypic variation for longitudinal data, i.e. typically data with many, 'repeated' measurements per individual recorded over time. They are especially suited for traits which are changing with time so that repeated measurements do not completely represent the same trait. The example considered here forth is growth of an animal with weights taken at a number of ages, but the concept is readily applicable to other characters and other continuous scales or 'meta-meters'.

In essence, covariance functions are the 'infinite-dimensional' equivalent to covariance matrices in a traditional, 'finite' multivariate analysis [15]. As the name indicates, a covariance function (CF) describes the covariance between records taken at certain ages as a function of these ages. A suitable function is a higher order polynomial. This implies that when fitting a CF model, we need to estimate the coefficients of the polynomial instead of the covariance components in a finite-dimensional analysis. The number of coefficients required is determined by the order of fit of the polynomials.

A finite-dimensional, multivariate analysis is equivalent to 'full fit' CF analysis where the order of fit is equal to the number of ages measured, i.e. the covariance matrices for the ages in the data generated by the estimated CFs are equal to the estimates that would have been obtained in a conventional, multivariate analysis. In practice, however, a reduced order fit often suffices. This reduces the number of parameters to be estimated and thus sampling errors, resulting in a smoothing of the estimated covariance structure.

Kirkpatrick et al. [15, 16] modelled CFs using orthogonal polynomials of age, choosing Legendre polynomials. Let Σ denote a covariance matrix of size $q \times q$, and Φ of size $q \times k$ the matrix of orthogonal polynomials evaluated at the given ages with elements $\phi_{ij} = \phi_j(t_i)$, the jth polynomial for the ith age t_i. The order of fit of the CF is given by $k \leqslant q$. This allows the covariance matrix to be rewritten as $\Sigma = \Phi K \Phi'$ with $K = \{K_{ij}\}$ a matrix of coefficients, and gives CF

$$\mathcal{F}(t_m, t_l) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \phi_i(t_m)\phi_j(t_l)K_{ij} = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} t_m^i t_l^j \omega_{ij} \qquad (1)$$

Here, t_m are the ages adjusted to the range for which the polynomial is defined. Let $\mathbf{t}_m$ with elements t_m^i for $i = 0, \ldots, k-1$ denote the row vector of

powers of t_m and Λ the matrix of polynomial coefficients. This gives the mth row of Φ as $\phi_m = t_m\Lambda$. For instance for $k = 3$ and Legendre polynomials

$$\Lambda = \begin{bmatrix} \sqrt{1/2} & 0 & -\sqrt{5/8} \\ 0 & \sqrt{3/2} & 0 \\ 0 & 0 & \sqrt{45/8} \end{bmatrix}$$

and

$$\phi_m = \begin{bmatrix} 0.7071 & 1.2247t_m & 2.3717t_m^2 - 0.7906 \end{bmatrix}$$

Thus equation (1) can be rewritten as $\mathcal{F}(t_m, t_l) = t_m\Lambda K\Lambda' t_l' = t_m\Omega t_l'$, i.e. the coefficient matrix Ω with elements ω_{ij} is obtained from K by including the terms of the polynomial chosen.

Kirkpatrick et al. [15] described a generalized least-squares procedure to determine the coefficients of a CF from an estimated covariance matrix. Often, however, this is not available or computationally expensive to obtain. Meyer and Hill [24] showed that the coefficients of CFs can be estimated directly from the data by restricted maximum likelihood (REML) through a simple reparameterization of existing, 'finite-dimensional' multivariate REML algorithms. For the special case of a simple animal model with equal design matrices computational requirements were restricted to the order of fit of the genetic CF. In the general case, however, their approach required a multivariate mixed model matrix proportional to the number of ages in the data to be set up and factored, even for a reduced order fit. This severely limited practical applications, especially for data with records at 'all ages'.

Polynomial regressions have been used to describe the growth of animals for a long time [35], but only recently has there been interest in random regression (RR) models. These have by and large been ignored in animal breeding applications so far, although they are common in other areas; see, for instance, Longford [19] for a general exposition. RRs in a linear mixed model context have been considered by Henderson [9]. Jennrich and Schluchter [12] included the 'random coefficients' model in their treatment of REML and maximum likelihood estimation for unbalanced repeated measures models with structured covariance matrices. Recent applications include the genetic evaluation of dairy cattle using test day records ([10, 11, 14] Van der Werf et al. unpublished), and the description of growth curves in pigs [2] and beef cattle [33].

This paper describes an alternative procedure for the estimation of covariance functions to that proposed by Meyer and Hill [24], which overcomes the limitations discussed above. It is shown that the CF model is equivalent to a RR model with polynomials of age as independent variables, and that REML estimates of the coefficients of the CF can be obtained as covariances among the regression coefficients. A mechanism is described to restrict the rank of the estimated covariance matrices (of regression coefficients) and thus the CFs, reducing the number of parameters to be estimated. The method is illustrated with an application to beef cattle data.

2. ESTIMATION OF COVARIANCE FUNCTIONS

2.1. Model of analysis

2.1.1. Finite-dimensional model

Consider an animal model

$$y_{ij} = F + a_{ij} + r_{ij} + \varepsilon_{ij} \tag{2}$$

with y_{ij} the observation for animal i at time j, a_{ij} and r_{ij} the corresponding additive genetic and permanent environmental effects due to the animal, respectively, ε_{ij} the measurement error (or temporary environmental effect) pertaining to y_{ij} and F some fixed effects. Furthermore, let t_{ij} denote the age (or equivalent) at which y_{ij} is recorded, and assume there are q_i records for animal i and a total of q different ages in the data.

Commonly, under a 'finite-dimensional' model of analysis, data represented by equation (2) are analysed either assuming measures at different ages are different traits, i.e. carrying out a q-dimensional, multivariate analysis, or fitting the so-called repeatability model, i.e. assuming $a_{ij} = a_i$ and $r_{ij} =_i$ for all $j = 1, \ldots, q_i$ and carrying out a univariate analysis. In the former, fully parametric case, covariance matrices are taken to be unstructured. Fitting a covariance function model, however, we impose some structure on the covariance matrices. This implies the assumption that the series of (up to) q measurements represents k different 'traits' or variates, with $1 \leq k \leq q$ denoting the order of fit of the covariance function.

2.1.2. Random regression model

As shown below, the covariance function model is equivalent to a 'random regression' model fitting functions of age (or equivalent) as covariables.

Kirkpatrick et al. [15, 16] used the well-known Legendre polynomials (see, for instance Abramowitz and Stegun [1]) in fitting covariance functions. These have a range of -1 to 1. Let t_{ij}^* denote the jth age for animal i standardized to this interval, and let $\phi_m(t_{ij}^*)$ be the mth Legendre polynomial evaluated for t_{ij}^*. We can then rewrite equation (2) as a RR model

$$y_{ij} = F + \sum_{m=0}^{k_A-1} \alpha_{im}\phi_m(t_{ij}^*) + \sum_{m=0}^{k_R-1} \gamma_{im}\phi_m(t_{ij}^*) + \varepsilon_{ij} \tag{3}$$

with α_{im} and γ_{im} representing the mth additive genetic and permanent environmental *random regression* coefficients for animal i, respectively, and k_A and k_R denoting the respective orders of fit.

This formulation (3) implies that the vector of q breeding values in a 'finite-dimensional', multivariate analysis is replaced by the vector of k_A additive genetic, random regression coefficients. Note, however, that with k_A chosen appropriately (i.e. the minimum order of fit modelling the data adequately), there is virtually no loss of information. In other words, equation (3) can be

employed as an effective tool to reduce the number of traits to be handled (and breeding values to be reported) for 'traits' measured over a continuous time scale such as weights (e.g. birth, weaning, yearling, final and mature weight) in beef cattle or test day records for dairy cows. Moreover, the RR model (3) yields a description of the animal's genetic potential for the complete time period considered, for instance, an estimate of the growth or lactation curve.

2.1.3. Covariance structure

The covariance between two records for the same animal is then

$$
Cov(y_{ij}, y_{ij'}) = \sum_{m=0}^{k_A-1} \sum_{l=0}^{k_A-1} \phi_m(t_{ij}^*)\phi_l(t_{ij'}^*)Cov(\alpha_{im}, \alpha_{il})
$$

$$
+ \sum_{m=0}^{k_R-1} \sum_{l=0}^{k_R-1} \phi_m(t_{ij}^*)\phi_l(t_{ij'}^*)Cov(\gamma_{im}, \gamma_{il}) + Cov(\varepsilon_{ij}, \varepsilon_{ij'}) \quad (4)
$$

Generally measurement errors are assumed to be i.i.d. with variance σ_ε^2, so that $Cov(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma_\varepsilon^2$ for $j = j'$ and 0 otherwise, but other assumptions, such as heterogeneous variances or autoregressive errors, are readily accommodated.

Clearly, the first two terms in equation (4) are CF with the covariances between random regression coefficients equal to the coefficients of the corresponding covariance functions [24], see equation (1) above, i.e. the RR model is equivalent to a CF model. Conversely, the RR model provides an alternative strategy to estimate CFs. While the REML algorithm described by Meyer and Hill [24] required mixed model equations of size proportional to the total number of ages q to be set up and factored in the general case, requirements under the equivalent random regression model are proportional to the orders of fit, k_A and k_R. Hence, this approach offers considerably more scope to handle data coming in 'at all ages' and should be especially advantageous for k_A or $k_R \ll q$.

2.1.4. Fixed effects

In fitting a RR model it is generally assumed that systematic differences in age are taken into account by the fixed effects in the model of analysis. In most cases, these include a fixed regression of the same form as the random regression (e.g. [9, 11, 12]), which can be thought of as modelling the population trajectory, while the random regressions for each animal represent individuals' deviations from this curve.

2.2. REML estimation

Considering all animals, equation (3) can be written in matrix form as

$$
\mathbf{y} = \mathbf{Xb} + \mathbf{Z}^*\boldsymbol{\alpha} + \mathbf{Z}_D^*\boldsymbol{\gamma} + \boldsymbol{\varepsilon}
$$

with $\mathbf{y}$ the vector of N observations measured on N_D animals, $\mathbf{b}$ the vector of fixed effects, $\boldsymbol{\alpha}$ the vector of $k_A \times N_A$ additive-genetic random regression

coefficients ($N_A \geq N_D$ denoting the total number of animals in the analysis, including parents without records), γ the vector of $k_R \times N_D$ permanent environmental random regression coefficients, ε the vector of N measurement errors, and $\mathbf{X}$, $\mathbf{Z}^*$ and $\mathbf{Z}_D^*$ denoting the corresponding 'design' matrices.

Here $\mathbf{Z}_D^*$ is the non-zero part of $\mathbf{Z}^*$ (for $k_A = k_R$), i.e. the part of $\mathbf{Z}^*$ corresponding to animals in the data. The superscript '*' marks matrices incorporating orthogonal polynomial coefficients. Assuming $\mathbf{y}$ is ordered for animals, $\mathbf{Z}_D^*$ is blockdiagonal, the block for animal i is of dimension $q_i \times k_R$, and has elements $\phi_m(t_{ij}^*)$. Note that each observation gives rise to k_R (or k_A for $\mathbf{Z}^*$) non-zero elements rather than a single element of 1 in the usual, finite-dimensional model, i.e. the design matrices are considerably denser than in the latter case.

Let $\mathbf{K}_A$ with elements $K_{A_{ml}} = Cov(\alpha_m, \alpha_l)$ and $\mathbf{K}_R$ with elements $K_{R_{ml}} = Cov(\gamma_m, \gamma_l)$ denote the coefficient matrices for the additive genetic and permanent environmental covariance functions $\mathcal{A}$ and $\mathcal{R}$, respectively. In terms of analysis, this is analogous to treating RR coefficients as correlated 'traits'. Assume that the fixed part of the model accounts for systematic age effects, so that $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{K}_A \otimes \mathbf{A})$ and $\boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{K}_R \otimes \mathbf{I}_{N_D})$, and that $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are uncorrelated. For generality, let $V(\boldsymbol{\varepsilon}) = \mathbf{R}$, but assume $\mathbf{R}$ is blockdiagonal for animals with blocks equal to submatrices of the $q \times q$ matrix $\boldsymbol{\Sigma}_\varepsilon$. The mixed model matrix pertaining to equation (5) is then

$$\mathbf{M}^* =$$
$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}^* & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_D^* & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^{*'}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{*'}\mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{K}_A^{-1} \otimes \mathbf{A}^{-1} & \mathbf{Z}^{*'}\mathbf{R}^{-1}\mathbf{Z}_D^* & \mathbf{Z}^{*'}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_D^{*'}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_D^{*'}\mathbf{R}^{-1}\mathbf{Z}^* & \mathbf{Z}_D^{*'}\mathbf{R}^{-1}\mathbf{Z}_D^* + \mathbf{K}_R^{-1} \otimes \mathbf{I}_{N_D} & \mathbf{Z}_D^{*'}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z}^* & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z}_D^* & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$
$$(6)$$

where $\mathbf{A}$ is the numerator relationship matrix between animals, $\mathbf{I}_N$ is an identity matrix of size N, and $\otimes$ denotes the direct matrix product. $\mathbf{M}^*$ has $N_F + k_A N_A + k_R N_D + 1$ rows and columns (with N_F being the total number of levels of fixed effects fitted), i.e. its size and thus computational requirements are proportional to the order of fit of the CFs. For $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}$, σ_ε^2 can be be factored from $\mathbf{M}^*$, resulting in a matrix which can be set up as for a univariate analysis.

Estimates of the distinct elements of $\mathbf{K}_A$ and $\mathbf{K}_R$ and the parameters determining $\boldsymbol{\Sigma}_\varepsilon$ can be obtained by REML, applying existing procedures for multivariate analyses under a 'finite' model. This may involve a simple, derivative-free algorithm [21] or, more efficiently, a method utilizing information from derivatives of the likelihood, such as Johnson and Thompson's [13] 'average information' algorithm; see Madsen et al. [20] or Meyer [22] for a description of the latter in the multivariate case.

While true measurement errors are generally assumed to be i.i.d., there may be cases in which we need to allow for heterogeneous variances or correlations between 'temporary' environmental effects. This may, to some extent, compensate for suboptimal orders of fit for permanent environmental or genetic covariance functions. In other cases $\boldsymbol{\Sigma}_\varepsilon$ may include parameters, such as the

autocorrelation ρ for measurement errors following a stationary time series, for which $V(\mathbf{y})$ is non-linear and for which derivatives are thus not straightforward to evaluate. In these instances, a two-step procedure combining a derivative-free search (e.g. a quadratic approximation) for the 'difficult' parameter(s) with an average information algorithm to maximize $\log \mathcal{L}$ with respect to the 'linear' parameters can be envisaged. A similar strategy has been employed by Thompson [32] in estimating the regression on maternal phenotype as well as additive genetic and environmental components of variance. Alternatively, estimation may be carried out in a Baysian framework using a Monte Carlo based technique, see Varona et al. [33] for an application in a linear RR model.

Calculation of the log likelihood ($\mathcal{L}$) requires factoring $\mathbf{M}^*$ to calculate the log determinant of the coefficient matrix ($\log |\mathbf{C}^*|$) and the residual sums of squares ($\mathbf{y}'\mathbf{P}^*\mathbf{y}$) (see Meyer [21] for details). The likelihood is then

$$\log \mathcal{L} = -\frac{1}{2}(N_A \log |\mathbf{K}_A| + k_A \log |\mathbf{A}| + N_D \log |\mathbf{K}_R|$$
$$+ \log |\mathbf{C}^*| + \log |\mathbf{R}| + \mathbf{y}'\mathbf{P}^*\mathbf{y}) \quad (7)$$

For i.i.d. measurement errors, the error variance can be estimated directly as $\sigma_\varepsilon^2 = \mathbf{y}'\mathbf{P}^*\mathbf{y}/(N - r(\mathbf{X}))$, as for univariate analyses.

2.2.1. Extensions to other models

So far only the case of a simple, 'univariate' animal model has been considered. More complicated models, however, are readily accommodated in the framework described. For instance, additional random effects such as maternal genetic effects or litter effects can be taken into account analogously by modelling each as a series of random regression coefficients. Correlations between random effects, e.g. non-zero direct-maternal genetic covariances, can be modelled by allowing for covariances between the respective regression coefficients, which then yield a CF describing the covariance between random effects over time.

Similarly, 'multivariate' CF [24] for series of measurements for different traits (e.g. height and weight measured at different times) can be estimated simply by fitting sets of RR coefficients for each trait and allowing for covariances between corresponding sets for different traits. An expectation-maximization type algorithm for a bivariate analysis under a RR model has recently been described by Shah et al. [30]. As mentioned above, a variety of assumptions about the structure of the within-individual, temporary environmental covariance matrices can be accommodated; see, for instance, Wolfinger [36] for a description of some commonly used models.

2.3. Reduced rank covariance functions

For q correlated measurements, the information supplied (or most of it) can generally be summarized as a set of $k < q$ linear combinations. These can be determined by a singular value decomposition of the corresponding covariance matrix. Typically, this yields one or a few (k) large, dominating

eigenvalues with the remainder $(q - k)$ being small or zero. Setting the latter to zero and backtransforming (by pre- and postmultiplying the diagonal matrix of eigenvalues with the matrix of eigenvectors and its transpose, respectively) then yields a modified, reduced rank covariance matrix. In estimating covariance matrices, this could be used to reduce the number of parameters to be estimated and thus sampling variation. A parameterization to the elements of the eigenvalue decomposition and setting eigenvalues $k + 1, \ldots, q$ and the corresponding eigenvectors to zero would achieve this but reduce the number of parameters to be estimated only for $k < q/2$. Though not perceived for this explicit purpose, the 'symmetric coefficients' CF model of [15] provides an alternative way of estimating reduced rank covariance matrices [22].

As outlined by Kirkpatrick et al. [15], there is an equivalent to the eigenvalue decomposition of covariance matrices for covariance functions, with a corresponding interpretation. Estimates of the eigenvalues of a CF fitted to order k are simply the eigenvalues of the corresponding, estimated matrix of coefficients ($\mathbf{K}$). Similarly, estimates of the eigenfunctions of a CF, the infinite-dimensional equivalent to eigenvectors, can be obtained from the eigenvectors of $\mathbf{K}$. Let $\mathbf{v}_i$ denote the ith eigenvector of $\mathbf{K}$ with elements ν_{ij} and $\theta_j(t^*)$ the jth order Legendre polynomial. The ith eigenfunction of the CF is then [15]

$$\psi_i = \sum_{j=0}^{k-1} \nu_{ij} \, \theta_j(t^*) \tag{8}$$

Note that $\theta_j(t^*)$ is not evaluated for any particular age, but includes polynomials of the standardized age t^*. Hence, ψ_i is a continuous, polynomial function in t^*. As discussed by Kirkpatrick et al. [15], eigenfunctions of genetic CF are especially of interest, as they represent possible deformations of the mean (growth) trajectory which can be effected by selection, while the corresponding eigenvalues describe the amount of genetic variation in that direction. In particular, the eigenfunction associated with the largest eigenvalue gives the direction in which the mean trajectory will change most rapidly.

Fitting a CF to order k requires $k(k + 1)/2$ coefficients, i.e. covariances between random regression coefficients, to be estimated, and gives estimates of the first k eigenfunctions and eigenvalues of the CF. In some instances, one or several eigenvalues of the CF may be close to zero or small compared to the other eigenvalues. This implies that we require a kth order fit to model the shape of the (growth) curve adequately, but that a subset of m directions (= eigenfunctions) suffices. In other words, we might obtain a more parsimonious fit of the CF by estimating a reduced rank coefficient matrix, forcing $k - m$ eigenvalues of $\mathbf{K}$ to be zero.

Consider the Cholesky decomposition of $\mathbf{K}$, pivoting on the largest diagonal

$$\mathbf{K} = \mathbf{LDL}' = \sum_{i=1}^{k} d_i \, \mathbf{l}_i \, \mathbf{l}_i' \tag{9}$$

where $\mathbf{L}$ is a lower diagonal matrix with diagonal elements of unity, $\mathbf{l}_i$ the ith column vector of $\mathbf{L}$, and $\mathbf{D}$ is a diagonal matrix. For a covariance matrix $\mathbf{K}$, the

ith element of $\mathbf{D}$, d_i, can be interpreted as the conditional variance of variable i, given variables $1, \ldots, i-1$. A reparameterization to the non-zero off-diagonal elements of $\mathbf{L}$ and the diagonal elements of $\mathbf{D}$ has been advocated for REML estimation of covariance components to remove constraints on the parameter space or improve rate of convergence in an iterative estimation scheme [6, 18, 25]. Other parameterizations in this context, based on the eigenstructure of the covariance matrix, have been considered by Pinheiro and Bates [26].

An alternative form of the Cholesky decomposition is $\mathbf{K} = \mathbf{L}^*\mathbf{L}^{*'}$ where $\mathbf{L}^*$ has diagonal elements $l_{ii}^* = \sqrt{d_i}$. $\mathbf{L}^*$ is often interpreted as $\mathbf{K}^{1/2}$. The eigenvalues of the power of a matrix are equal to the power of the eigenvalues of the matrix, and the eigenvalues of a triangular matrix are equal to its diagonal elements [5]. Hence, the estimate of $\mathbf{K}$ can be forced to have rank m by assuming elements d_{m+1} to d_k in equation (9) are zero (elements d_i are assumed to be in descending order). This yields a modified matrix

$$\mathbf{K}^+ = \sum_{i=1}^{m} d_i\, \mathbf{l}_i\, \mathbf{l}_i' \qquad (10)$$

The vectors $\mathbf{l}_i$ corresponding to the zero d_i are then not needed, i.e. $\mathbf{K}^+$ is described by $km - m(m-1)/2$ parameters, m elements d_i and $(k-1)m - m(m-1)/2$ elements of l_{ij} $(j > i)$ of the $\mathbf{l}_i$. Clearly, this is not equivalent to fitting a (full rank) CF to the order m (which would involve $m(m+1)/2$ parameters) – for instance for $k = 4$ and $m = 2$ we fit a cubic regression assuming there are only two independent directions in which the trajectory is likely to change, while for $k = m = 2$ we fit a linear regression.

Strictly speaking, equation (6) has to be of full rank. Hence, for practical computations, d_i are set to a small positive value (e.g. 10^{-4}). Alternatively, a REML algorithm which allows for a semi positive definite covariance matrix of random effects could be employed, c.f. Harville [8] or Frayley and Burns [3]. Obviously, this parameterization can also be used to estimate reduced rank covariance matrices for finite-dimensional, multivariate analyses.

3. APPLICATION

3.1. Material and methods

Meyer and Hill [24] fitted covariance functions to January weights of 913 beef cows, weighed from 2 to 6 years of age, 2 795 records in total with up to five records per cow available. Their analysis used age at weighing in years and fitted measurement errors and fixed effects for each age separately. These data were re-analysed using the random regression model and fitting age at weighing in months. Analyses were carried out using program DXMRR [23], employing a derivative-free algorithm to maximize log $\mathcal{L}$.

There were a total of 22 ages in the data, ranging from 19 to 70 months. *Figure 1* gives the mean weight and number of records for each age class. Analyses were carried out fitting a separate measurement error variance component for each year of age (five variances). Fixed effects fitted were year-paddock of
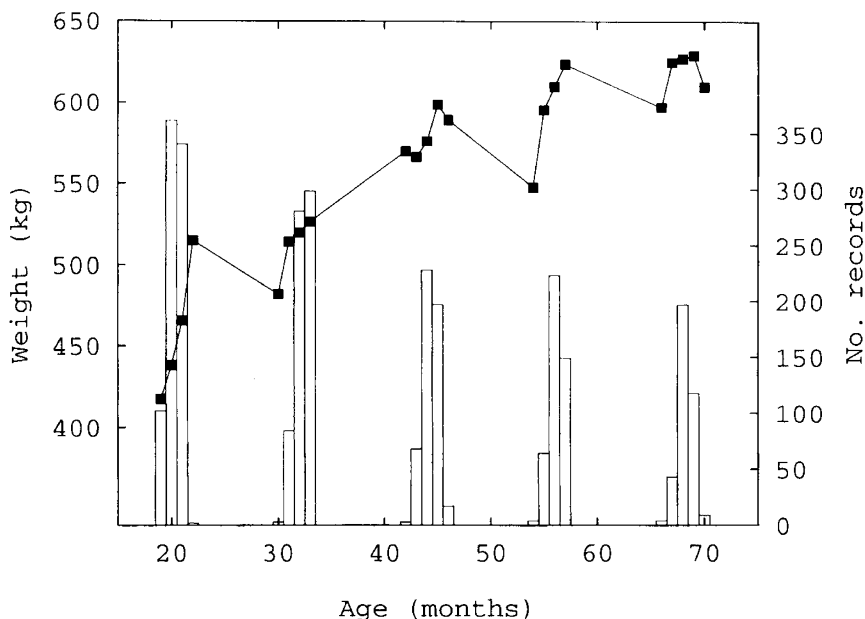
Figure 1. Mean weights (■, in kg) and numbers of records for ages in the data.

weighing subclasses (86 levels), year of birth effects (16 levels) and a cubic regression on age at weighing. The model for fixed effects was 'univariate', i.e. the effects were assumed to be similar for cows of all ages. Additive genetic and permanent environmental covariance functions were fitted to the same order throughout ($k_A = k_R = k$). Orders of fit considered ranged from 1 to 6. In addition, the usual 'repeatability model' was fitted, i.e. a CF model with $k = 1$ and a single measurement error variance, assumed to be the same for all ages.

For each order of fit, the number of non-zero eigenvalues allowed for each coefficient matrix to be estimated was set to the same value (r) for $\mathbf{K}_A$ and $\mathbf{K}_R$, considering values of $r \leq k$ of 1 to 3. In several instances, analyses resulted in estimates of $\mathbf{K}_R$ with one small eigenvalue. In these cases, the rank of $\mathbf{K}_R$ was reduced by one. In every case, this yielded a further improvement in log $\mathcal{L}$ when continuing the analysis, i.e. earlier convergence had been to a false maximum as the search procedure had become 'stuck' at the bounds of the parameter space.

In general, analyses took a considerable time to converge, markedly longer for an order of fit of k than for a comparable k-variate, finite-dimensional analysis. Furthermore, several restarts were required for each analysis before likelihoods stabilized. Convergence was especially slow when attempting to estimate 'unnecessary' parameters, i.e. an order of fit or rank of CF with one or more eigenvalues close to zero.

3.2. Results

3.2.1. Likelihoods

Maximum likelihood values from all analyses together with eigenvalues of the estimated coefficient matrices and estimates of the measurement error variances are summarized in *table I*. Clearly, a repeatability model (first line) was inappropriate in this case, estimates of σ_ε^2 (for $k = 1$) being considerably higher for older cows than for 2-year-old cows. Forcing the rank r of an estimated coefficient matrix – and thus CF – to be 1 is equivalent to the assumption that all corresponding correlations have a value of unity. For $r = 1$ and $k > 1$, the higher order coefficients then model heterogeneity of variances.

For all orders of fit, increasing r from 1 to 2 yielded significant increases in log $\mathcal{L}$. Increasing r further to 3, however, did not raise log $\mathcal{L}$ significantly, increases of 5.33 ($k = 5$) and 5.75 ($k = 6$) being outweighed by the number of additional parameters (6 and 8, respectively). For $r = 2$, log $\mathcal{L}$ increased significantly until k reached 4. Estimated CFs for this model were

$$
\mathcal{A}(t_i, t_j) = \begin{bmatrix} 1 & t_i & t_i^2 & t_i^3 \end{bmatrix} \begin{bmatrix} 2162.1 & 136.7 & -602.3 & 533.8 \\ 136.7 & 118.0 & -128.0 & -11.4 \\ -602.3 & -128.0 & 241.8 & -111.6 \\ 533.8 & -11.4 & -111.6 & 150.4 \end{bmatrix} \begin{bmatrix} 1 \\ t_j \\ t_j^2 \\ t_j^3 \end{bmatrix}
$$

$$
\mathcal{R}(t_i, t_j) = \begin{bmatrix} 1 & t_i & t_i^2 & t_i^3 \end{bmatrix} \begin{bmatrix} 528.8 & 425.2 & 159.0 & -156.5 \\ 425.2 & 505.7 & -59.2 & -607.1 \\ 159.0 & -59.2 & 261.5 & 502.5 \\ -156.5 & -607.1 & 502.5 & 1460.2 \end{bmatrix} \begin{bmatrix} 1 \\ t_j \\ t_j^2 \\ t_j^3 \end{bmatrix}
$$

with t_i denoting the ith standardized age. When fitting polynomials, however, we should also examine an order of fit of $k + 2$, i.e. the next order of fit with the same type (even versus uneven) of exponent, as an odd-degree polynomial is likely to contribute little when an even-degree polynomial fits the data [4]. Indeed, while adding a quartic coefficient ($k = 5$) did not increase log $\mathcal{L}$ significantly over a cubic polynomial ($k = 4$), fitting a quintic term ($k = 6$) did.

Fitting years rather than months of age as meta-meter for the CFs, Meyer and Hill [24] found log $\mathcal{L}$ not to increase significantly beyond $k = 2$. As shown in *figure 1*, there was only a small spread of ages within each year, i.e. the change in scale was expected to have little effect on estimates. The model employed by Meyer and Hill [24], however, was multivariate, i.e. fitted separate fixed effects for each year of age. Presumably, the stronger trend in covariances observed in this analysis can be attributed to less variation being removed by fixed effects. In addition, likelihood ratio tests were carried out considering full rank CFs and the associated number of parameters.

3.2.2. Phenotypic variation

Figure 2 shows the estimated phenotypic standard deviations for the ages in the data. For $k = 1$, deviations from a horizontal line reflected differences in

Table I. Maximum log likelihood ($\log \mathcal{L} + 11{,}800$), the first three eigenvalues of the estimated genetic ($\mathcal{A}$) and permanent environmental ($\mathcal{R}$) covariance function (λ_{A_i} and λ_{R_i}, $i = 1, 3$ for $\mathcal{A}$ and $\mathcal{R}$, respectively), and estimates of measurement error variances ($\sigma^2_{\varepsilon_i}$, $i = 1, 5$) for different orders of polynomial fit (k) and rank of the estimated coefficient matrices.[a]

k	p[b]	$\log \mathcal{L}$	r_A	λ_{A_1}	λ_{A_2}	λ_{A_3}	r_R	λ_{R_1}	λ_{R_2}	λ_{R_3}	$\sigma^2_{\varepsilon_1}$	$\sigma^2_{\varepsilon_2}$	$\sigma^2_{\varepsilon_3}$	$\sigma^2_{\varepsilon_4}$	$\sigma^2_{\varepsilon_5}$
1	3	−237.621	1	2 402.2			1	1 040.9			1 395	1 395	1 395	1 395	1 395
1	7	−138.480	1	1 810.5			1	842.3			480	2 415	2 274	2 266	2 044
2	9	−52.709	1	2 953.2	0.0		1	1 779.4			646	2 286	1 928	1 626	857
2	11	−48.300	2	3 129.4	82.7		2	1 560.4	16.4		533	2 296	1 897	1 612	846
3	11	−45.411	1	3 716.9	0.0	0.0	1	1 502.4	0.0	0.0	706	2 183	1 768	1 633	779
3	13	−39.508	2	3 679.8	58.9	0.0	2	1 433.9	0.0	0.0	560	2 199	1 733	1 572	924
3	15	−39.016	2	3 662.2	49.5	0.0	2	1 433.7	57.6	0.0	513	2 200	1 724	1 586	799
3	17	−39.026	3	3 685.4	49.5	0.0	3	1 420.4	51.4	0.0	516	2 200	1 726	1 584	818
4	13	−43.139	1	4 190.4	0.0	0.0	1	1 254.7	0.0	0.0	709	2 144	1 769	1 784	2
4	19	−33.647	2	3 667.0	63.4	0.0	2	1 521.2	121.8	0.0	442	2 129	1 719	1 617	0
4	21	−33.597	3	3 648.9	66.0	7.0	2	1 525.1	121.5	0.0	434	2 123	1 715	1 621	0
4	23	−33.599	3	3 655.4	61.5	3.7	3	1 520.9	121.6	7.7	433	2 120	1 715	1 615	0
5	15	−39.680	1	3 566.4	0.0	0.0	1	1 708.9	0.0	0.0	711	2 148	1 506	1 744	552
5	23	−30.472	2	3 606.5	135.4	0.0	2	1 658.2	115.2	0.0	506	2 152	1 570	1 419	95
5	29	−25.144	3	3 655.0	254.1	53.0	3	1 547.6	130.3	50.2	427	2 086	1 341	1 501	0
6	17	−39.022	1	3 676.1	0.0	0.0	1	1 645.4	0.0	0.0	707	2 146	1 489	1 758	486
6	27	−23.675	2	3 346.8	197.0	0.0	2	1 789.3	179.4	0.0	290	2 120	1 429	1 798	0
6	35	−17.922	3	3 685.1	226.8	125.0	3	1 469.4	196.9	56.2	245	1 970	1 476	1 560	0

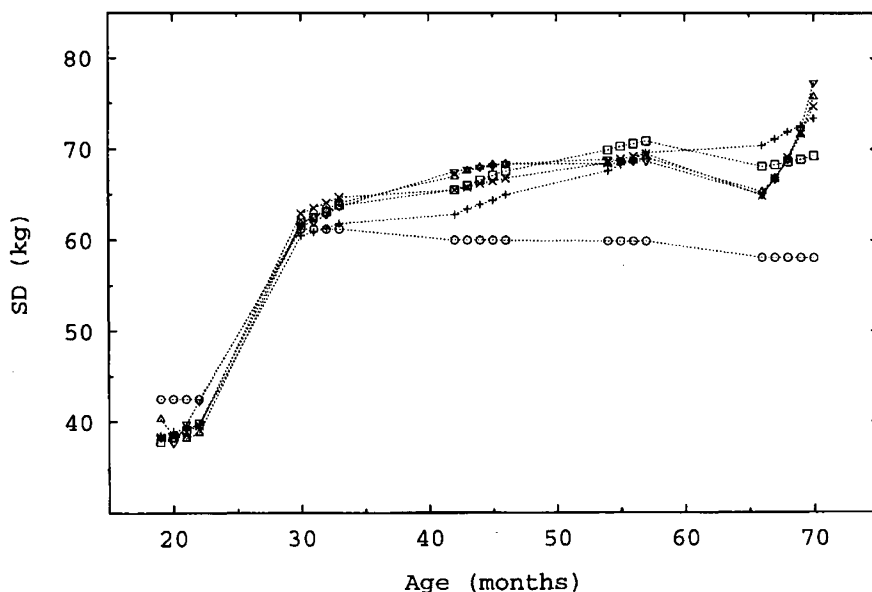[a] r_A: rank of $\mathbf{K}_A$; r_R: rank of $\mathbf{K}_R$.   [b] Number of parameters.

Figure 2. Estimated phenotypic standard deviations for increasing order (k) of polynomial fit ($\circ$: $k = 1$; $+$: $k = 2$; $\square$: $k = 3$; $\times$: $k = 4$; $\triangle$: $k = 5$; ∇: $k = 6$).

estimates of σ_ε^2 over years. Estimates for $k > 3$ were similar, the cubic term for $k \geq 4$ causing estimates for 6-year-old cows to rise sharply. As shown in *table I*, this was accompanied by estimates of $\sigma_{\varepsilon_5}^2$ of zero, i.e. presumably to some extent due to a restriction imposed on the parameter space, forcing $\sigma_\varepsilon^2 \geq 0$. Except for these last age classes, estimates agreed closely with those from a finite-dimensional multivariate analysis treating records at different years of age as separate traits, which were 40.0, 60.5, 65.8, 67.3 and 71.0 (kg) for average ages of 20.3, 32.3, 44.3, 56.2 and 68.2 months, respectively [24].

3.2.3. Eigenfunctions

As for phenotypic standard deviations, estimates of the first eigenvalue (*table I*) and corresponding eigenfunction of $\mathcal{A}$ (*figure 3*) were similar for orders of fit $k \geq 3$. Values of the eigenfunction for individual ages were roughly proportional to the genetic variance for the age, as determined from the corresponding estimate of genetic CF $\mathcal{A}$. Positive values throughout imply that selection for increased weight at any age is likely to increase weight at all other ages. Estimates for the second eigenvalue and -function, however, changed considerably from $k = 3, 4$ to $k = 5, 6$, in particular for $r = 3$. This was accompanied by a significant increase in log $\mathcal{L}$ from $k = 4$ to $k = 5$ and from $k = 5$ to $k = 6$ for $r = 3$ but not $r = 2$. Eigenvalues of estimates of $\mathcal{A}$ and $\mathcal{R}$ up to orders of fit of $k = 4$ were similar to those reported by Meyer and Hill [24], suggesting that the third sizeable eigenvalues appearing or $k \geqslant 5$ might be an artefact of the order of the polynomial function and the 'univariate' fixed effects part of the model of analysis.
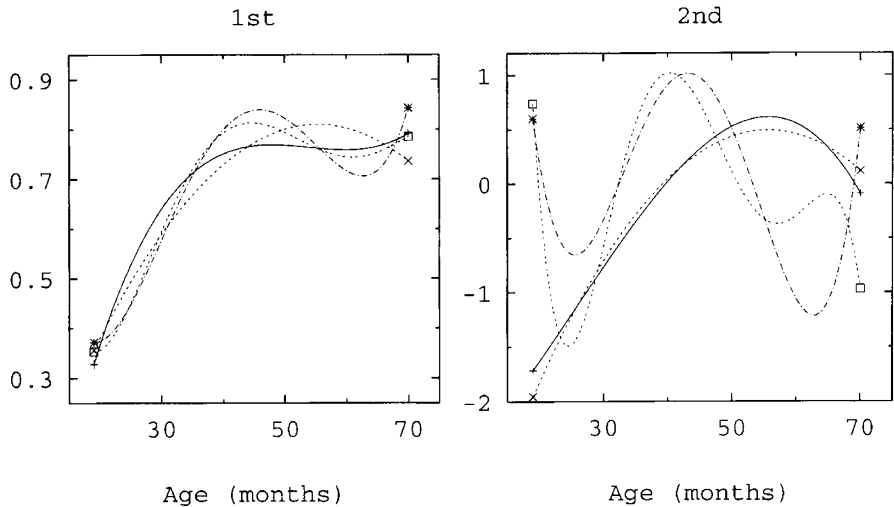
Figure 3. Estimated first and second eigenfunction of the genetic covariance function, for orders of polynomial fit of 3 ($\times$), 4 ($+$), 5 ($*$) and 6 ($\square$), respectively (rank 3 estimates of the coefficient matrices).

3.2.4. Genetic covariances

Figure 4 shows genetic covariances obtained from estimates of $\mathcal{A}$ for $k = 3, 6$ ($r = 3$), evaluated for the range of ages spanned by data. For $k = 3$ and $k = 4$, the resulting surfaces were very smooth, clearly following a quadratic and cubic function, while for $k \geq 5$ surfaces became 'wiggly', following the data more closely. A peak in genetic variance at 4 years of age with a subsequent decline was consistent with estimates of 672, 1 277, 2 221, 722 and 1 988 (kg^2) for ages 2, 3, 4, 5 and 6 years, respectively, from a finite-dimensional, multivariate analysis [24].

Corresponding genetic correlations for $k = 4$ and 6 are shown in *figure 5*. For $k = 4$, the surface was smooth with a 'plateau' close to unity for ages from 3 years onwards, and correlations between weights at 2 years and later ages decreasing with increasing time between measurements. This agrees with our biological expectations for the trait under consideration. In contrast, correlations for $k = 5$ (not shown) and $k = 6$, fluctuated considerably, especially at the edges of the surface, greatly magnifying sampling variation in the genetic covariances for these orders of fit, exhibited in *figure 4*.

3.2.5. Permanent environmental variation

As shown in *figure 6*, estimates of $\mathcal{R}$ produced permanent environmental variances which, except for weights at 6 years of age, increased considerably less over time than their genetic counterparts. Moreover, surfaces were considerably smoother for higher orders of fit ($k > 4$). For the last ages in the data, estimates of the permanent environmental variances increased dramatically, forcing estimates of the corresponding measurement error variance to zero,
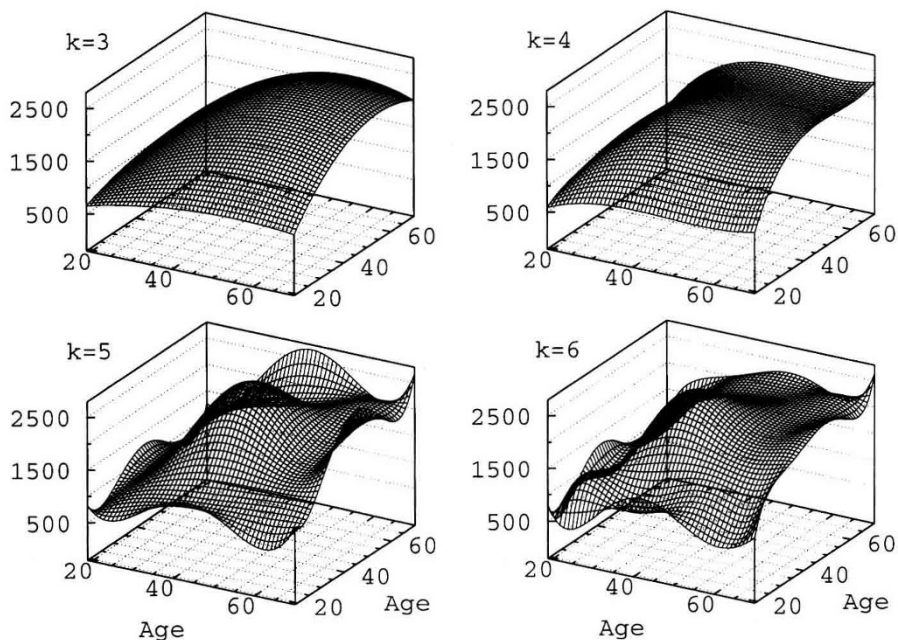
Figure 4. Genetic covariances (in kg^2) from estimates of the genetic covariance function for orders of polynomial fit (k) of 3 to 6 (rank 3).
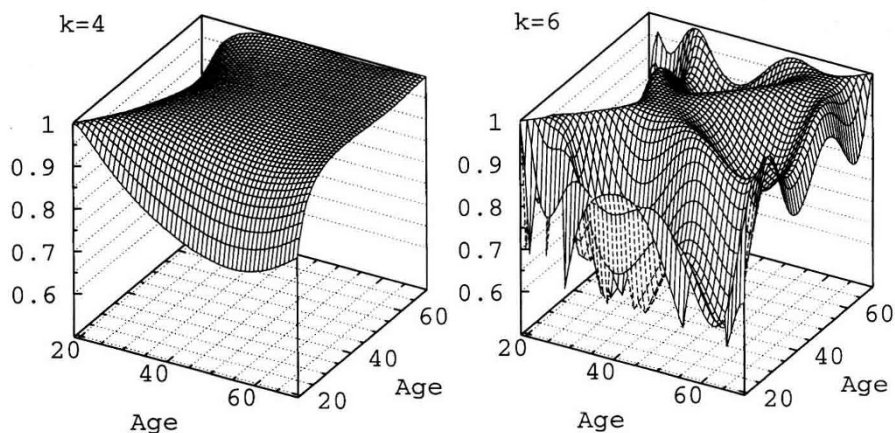

Figure 5. Estimates of genetic correlations for orders of polynomial fit (k) of 4 and 6.

and causing the increase in estimated phenotypic standard deviation observed above. While observation in the four age classes concerned were more variable than in the earlier years, this appears to be spurious and can only be attributed to smaller numbers of records and the large influence data points at the extremes can have on the estimates of regression coefficients.
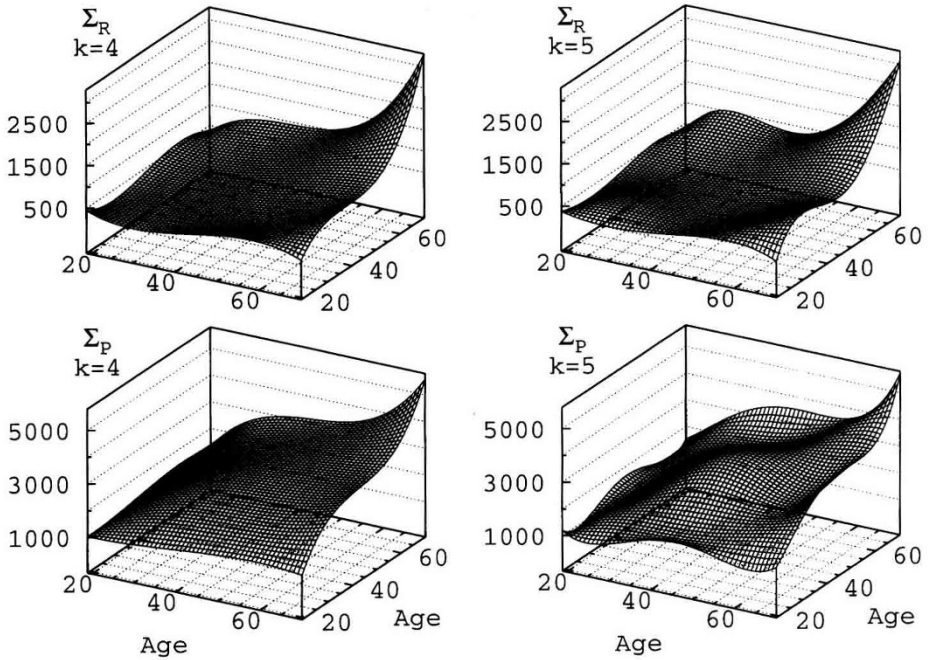
Figure 6. Permanent environmental (Σ_R) and genetic plus permanent environmental (Σ_P) covariances (in kg^2), for orders of polynomial fit (k) of 4 and 5.

4. DISCUSSION

There is a wealth of statistical literature on modelling of 'repeated records' in general, and structured covariance matrices in particular (see, for instance, Lindsey [17] or Vonesh and Chinchilli [34] and their extensive bibliographies). Applications in animal breeding, however, have until recently been limited to the extremes of a simple repeatability model or an unstructured, multivariate model. An added complication of quantitative genetic analyses, not shared by other fields, is the fact that we want to partition the between subject variation into its genetic and environmental components.

Covariance function and random regression models enable us to model the covariance structure of such records more adequately, alleviating the problems associated with an oversimplification (repeatability model) or an over-parameterization (multivariate model) for traits which change gradually along some continuous scale, such as time. Moreover, each source of variation, genetic or environmental, can be modelled. While a regression model is usually thought of in the context of modelling trends in means, covariance functions have been perceived to describe and smooth dispersion matrices. When regression coefficients are treated as random, however, they implicitly impose a covariance structure among the observations. Thus the two approaches converge and indeed, as shown, the covariance function model of Kirkpatrick et al. [15] is

equivalent to a special class of random regression models. While not common practice, covariance functions can be formulated for the sources of variation modelled by random regression coefficients in any RR model.

Regression models, fixed or random, require some assumptions about the parametric form of the regression equation. Regressing on orthogonal polynomials of time (or any other meta-meter) in order to fit Kirkpatrick et al.'s [15] CF model, only invokes the assumption that the trajectory to be modelled can be described by such polynomials. As shown above, in a maximum likelihood framework we can let the data determine the order of polynomial fit required and the rank of the CF which suffices.

Likelihood ratio tests carried out to determine the order of fit of covariance functions required should account for the fact that we are at the boundary of the parameter space and thus have 'non-standard' conditions where the likelihood ratio test criterion λ does not follow a χ^2 distribution [29], when testing whether additional random regression coefficients might have zero variance. Stram and Lee [31] considered tests of variance components for longitudinal data in a mixed model. They showed that the large sample distribution of λ is a 50:50 mixture of χ^2 distributions with q and $q + 1$ degrees of freedom, respectively, when we test the hypothesis that a matrix $\mathbf{D}$ is a $q \times q$ positive definite matrix against the hypothesis that it is a $(q + 1) \times (q + 1)$ matrix. Hence 'naive' tests, assuming λ has a χ^2 distribution with degrees of freedom equal to the number of parameters tested $(q+1)$ are too conservative. Stram and Lee [31] argue though that for such simple tests (of one additional parameter), resulting biases are likely to be small. In practical applications, the error probability α has been 'doubled' to account for this conservatism, i.e. λ has been contrasted against $\chi^2_{q+1,2\alpha}$ instead of $\chi^2_{q+1,\alpha}$ [27].

Alternatively, we can decide on an order of fit a priori or decide only to use the first n eigenfunction of the CF to describe the covariance structure. This choice may depend on the computational resources available, or may be influenced by the knowledge that higher order polynomials are notorious for magnifying small sampling errors and producing 'wiggly' functions. Even if this choice does not maximize the likelihood, we are likely to obtain more appropriate estimates than under the simple repeatability model when this clearly does not apply, both in terms of estimates of genetic parameters and prediction of breeding values. Further research is required to develop guidelines as how to choose the 'best' model for particular situations.

Great care has to be taken when defining and interpreting RR models. As shown in the numerical example, data points at the extremes of the independent variable can have a big influence on estimates of regression coefficients, yielding quite erratic estimates of CF or measurement error variances. Various authors estimated genetic parameters for test day milk yields of dairy cows fitting a random regression model. In most cases, resulting heritability estimates were high at the beginning and end of lactation and low in the middle of lactation, in marked contrast to previous estimates under a 'finite-dimensional' model, and thus regarded with justified scepticism ([10, 14]; Van der Werf et al., unpublished). This emphasizes the sensitivity of RR models to the effects and orders of CFs fitted. It has to be stressed that judicious modelling of fixed and random effects, careful choice of the orders of fit of CFs, and meticuluous screening of the data are paramount for successful RR model analyses.

As emphasized by Kirkpatrick et al. [15], many families of functions can be employed in the estimation of CF – while the choice of orthogonal function does not affect the estimate of the CF at the ages in the data, it does affect interpolation. Following their choice, we have used Legendre polynomials. These, however, generate a weight function with comparatively heavy emphasis on records at the outer parts of the interval for which they are defined (compared to, say, weights following a kernel from a Normal distribution) ([7]; section 3.3). Other functions might thus be more suitable. Alternatively, some scaling of the observed ages can be envisaged, for instance, a logarithmic transformation should reduce the impact of few observations on very old animals.

5. CONCLUSIONS

Random regression models provide a valuable tool to model repeated records in animal breeding adequately, especially if traits measured change gradually. They allow covariance functions to be formulated which describe genetic and environmental covariances among records over time. Moreover, they impose a structure on covariance matrices. Fitting regressions on orthogonal polynomials of time (or equivalent) we can estimate genetic covariance functions as suggested by Kirkpatrick et al. [15], whose eigenvalues and eigenfunctions provide an insight into the way selection is likely to affect the mean trajectory of the records considered and can be used to characterize differences between populations, e.g. breeds of animals.

6. SOFTWARE

A program is available for the estimation of covariance functions by REML, fitting a random regression animal model, as described above. 'DXMRR' is written in Fortran 90 and is part of the DFREML package, version 3.0; see Meyer [23] for further details. This is contained on the CD-ROM of the Sixth World Congress on Genetics Applied to Livestock Production, or can be downloaded from the DFREML home page at
http: //agbu.une.edu.au/~kmeyer/dfreml.html.

REFERENCES

[1] Abramowitz M., Stegun I.A., Handbook of Mathematical Functions, Dover, New York, 1965.
[2] Anderson S., Pedersen B., Growth and food intake curves for group-housed gilts and castrated male pigs, Anim. Sci 63 (1996) 457–464.

[3] Frayley C., Burns P.J., Large-scale estimation of variance and covariance components, SIAM J. Sci. Comp. 16 (1995) 192–209.

[4] Graybill F.A., An Introduction to Linear Statistical Models. Volume I, McGraw-Hill, Inc., New York, 1961.

[5] Graybill F.A., Matrices with Applications in Statistics, 2nd ed., Wadsworth Inc., 1983.

[6] Groeneveld E., A reparameterisation to improve numerical optimisation in multivariate REML (co)variance component estimation, Genet. Sel. Evol. 26 (1994) 537–545

[7] Härdle W., Applied Nonparametric Regression, Cambridge University Press, 1990.

[8] Harville D.A., Maximum likelihood approaches to variance component estimation and related problems, J. Am. Stat. Assoc. 72 (1977) 320–338.

[9] Henderson C.R. Jr., Analysis of covariance in the mixed model: Higher-level, non-homogeneous and random regressions, Biometrics 38 (1982) 623–640.

[10] Jamrozik J., Schaeffer L.R., Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins, J. Dairy Sci. 80, 1997) 762–770.

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

[12] Jennrich R.I., Schluchter M.D., Unbalanced repeated-measures models with structured covariance matrices, Biometrics 42 (1986) 805–820.

[13] Johnson D.L., Thompson R., Restricted Maximum Likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information, J. Dairy Sci. 78 (1995) 449–456.

[14] Kettunen A., Mäntysaari E., Stranden I., Pösö J., Genetic parameters for test day milk yields of Finnish Ayrshires with random regression model, J. Dairy Sci. 80 Suppl. 1, 197 (1997) (abstr.).

[15] Kirkpatrick M., Lofsvold D., Bulmer M., Analysis of the inheritance, selection and evolution of growth trajectories, Genetics 124 (1990) 979–993.

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

[17] Lindsey J.K., Models for Repeated Measurements, Oxford Statistical Science Series, Clarendon Press, Oxford, 1993.

[18] Lindstrom M.J., Bates D.M., Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, J. Am. Stat. Assoc. 83 (1988) 1014–1022.

[19] Longford N.T., Random Coefficient Models, Oxford Statistical Science Series, Clarendon Press, Oxford, 1993.

[20] Madsen P., Jensen J., Thompson R., Estimation of (co)variance components by REML in multivariate mixed linear models using average of observed and expected information, Proc. 5th World Congr. Genet. Appl. Livest. Prod. Vol. 22 (1994) 19–22.

[21] Meyer K., Estimating variances and covariances for multivariate animal models by Restricted Maximum Likelihood, Genet. Sel. Evol. 23 (1991) 67–83.

[22] Meyer K., An "average information" Restricted Maximum Likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices, Genet. Sel. Evol. 29 (1997) 97–116.

[23] Meyer K., 'DXMRR' – A program to estimate covariance functions for longitudinal data by Restricted Maximum Likelihood, Proc. 6th World Congr. Genet. Appl. Livest. Prod. Vol. 27 (1998) 465–466.

[24] Meyer K., Hill W.G., Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood, Livest. Prod. 47 (1997) 185–200.

[25] Meyer K., Smith S.P., Restricted Maximum Likelihood estimation for animal models using derivatives of the likelihood, Genet. Sel. Evol. 28 (1996) 23–49.

[26] Pinheiro J.C., Bates D.M., Unconstrained parameterizations for variance-covariance matrices, Stat. Comp. 6 (1996) 289–296.

[27] Robert-Granié C., Ducrocq V., Foulley J.L., Heterogeneity of variance for type traits in the Montbeliarde cattle breed, Genet. Sel. Evol. 29 (1997) 545–570.

[28] Schaeffer L.R., Dekkers J.C.M., Random regressions in animal models for test-day production in dairy cattle, Proc. 5th World Congr. Genet. Appl. Livest. Prod. Vol. 18 (1994) 443–446.

[29] Self S.G., Liang K.Y., Asymptotic properties of the maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, J. Am. Stat. Assoc. 82 (1987) 605–610.

[30] Shah A., Laird N., Schoenfeld D., A random-effects model for multiple characteristics with possibly missing data, J. Am. Stat. Assoc. 92 (1997) 775–779.

[31] Stram D.A., Lee J.W., Variance component testing in the longitudinal model, Biometrics 50, (1994) 1171–1177.

[32] Thompson R., The estimation of maternal genetic variance, Biometrics 32 (1976) 903–917.

[33] Varona L., Moreno C., Garcia-Cortes L.A., Altarriba J., Multiple trait genetic analysis of underlying biological variables of production functions, Livest. Prod. Sci. 47 (1997) 201–210.

[34] Vonesh E.F., Chinchilli V.M., Linear and Nonlinear Models for the Analysis of Repeated Measurements, Marcel Dekker Inc, New York, 1997.

[35] Wishart J., Growth-rate determinations in nutrition studies with the bacon pig and their analysis, Biometrics 30 (1938) 16–28.

[36] Wolfinger R.D., Heterogeneous variance-covariance structures for repeated measures data, J. Agric. Biol. Environ. Stat. 1 (1996) 205–230.