

Up hill, down dale: quantitative genetics of curvaceous traits

Karin Meyer^{1,*} and Mark Kirkpatrick²

¹*Animal Genetics and Breeding Unit[†], University of New England, Armidale,
New South Wales 2351, Australia*

²*Section of Integrative Biology, University of Texas, 1 University Station C-0930, Austin, TX 78712, USA*

‘Repeated’ measurements for a trait and individual, taken along some continuous scale such as time, can be thought of as representing points on a curve, where both means and covariances along the trajectory can change, gradually and continually. Such traits are commonly referred to as ‘function-valued’ (FV) traits. This review shows that standard quantitative genetic concepts extend readily to FV traits, with individual statistics, such as estimated breeding values and selection response, replaced by corresponding curves, modelled by respective functions. Covariance functions are introduced as the FV equivalent to matrices of covariances.

Considering the class of functions represented by a regression on the continuous covariable, FV traits can be analysed within the linear mixed model framework commonly employed in quantitative genetics, giving rise to the so-called random regression model. Estimation of covariance functions, either indirectly from estimated covariances or directly from the data using restricted maximum likelihood or Bayesian analysis, is considered. It is shown that direct estimation of the leading principal components of covariance functions is feasible and advantageous. Extensions to multi-dimensional analyses are discussed.

Keywords: function-valued traits; quantitative genetics; covariance function; random regression model

1. INTRODUCTION

Quantitative genetics has typically been concerned with traits that, in essence, are points, i.e. measured at a specific location or time. Clearly, however, there are traits which are not only recorded repeatedly per individual, but are also considered to be changing, gradually, up or down, and continually, between measurements.

The most common of these are longitudinal data, i.e. records taken at successive times or ages, such as weights or body size measurements, and we will concentrate on this case in the following. Alternatively, these could be records taken along some spatial scale, or any other continuous covariable, referred to as ‘control variable’ in the following. For instance, we might want to relate changes in backfat thickness of farm animals to weight at recording, or assess the effect of ambient temperature on growth rates of caterpillars (Kingsolver & Gomulkiewicz 2003). In this case, our traits of interest are complete curves or trajectories rather than individual points, and we want to quantify genetic values and their dispersion structure among records for the complete range of the control variable.

Today, such traits are commonly referred to as function-valued (FV) traits, where FV emphasizes that

the corresponding curves are described by a mathematical function (rather than implying some notion of functionality), and we will adopt this nomenclature. Earlier, Kirkpatrick *et al.* (1990) called these ‘infinite-dimensional’ traits, as they could, hypothetically, be measured infinitely many times along the continuous scale of interest. This paper reviews the role and analysis of FV traits in quantitative genetics.

2. GENETICS AND FUNCTIONAL DATA ANALYSIS

Functional data analysis (FDA) is the branch of statistics concerned with the analysis of FV traits, albeit generally at a phenotypic level. This extends to more than one control variable, e.g. for two spatial coordinates, we might deal with images. Loosely speaking, FDA attempts to characterize patterns in variation of curves and extract common features. This might include horizontal and vertical shifts, i.e. variation in phase and amplitude of curves, and an investigation of the associated eigenfunctions or ‘eigenfaces’. A brief introduction to FDA is given by Ramsay & Silverman (2001), and details can be found in their books (Ramsay & Silverman 1997, 2002). Applications span a wide range of fields, e.g. biology, economics and psychology. For example, Rice & Silverman (1991) considered the analysis of human gait patterns, Spitzner *et al.* (2003) examined tactile perception and Ramsay (2000) presented an analysis of variation in handwriting.

In quantitative genetics, models and methods for analyses of data that represent curves have obtained

*Author for correspondence (kmeyer@pobox.une.edu.au).

[†]A joint venture of NSW Department of Primary Industries and the University of New England.

One contribution of 16 to a Theme Issue ‘Population genetics, quantitative genetics and animal improvement: papers in honour of William (Bill) Hill’.

insights and stimuli from both evolutionary biology and animal breeding (Hill 1998).

(a) *Covariance functions*

Kirkpatrick & Heckman (1989) and Kirkpatrick *et al.* (1990) introduced the infinite-dimensional model for traits measured repeatedly per individual, and suggested to model genetic covariances of trajectories through covariance functions.

Let the continuous scale along which repeated observations are taken be time t . As the name implies, a covariance function (CF) then gives the covariance between two measurements taken at times t_j and $t_{j'}$ as a function of these times. CFs for FV traits can be interpreted as the equivalent to variance components for standard, non-FV traits. While CFs conceptually involve infinitely many parameters, Kirkpatrick & Heckman (1989) showed that good estimates can be obtained by approximating the CFs as the weighted sum of a relatively small number of basis functions, such as orthogonal polynomials. This yields reduced rank and smoothed estimates of the corresponding covariance matrices. Kirkpatrick *et al.* (1994) illustrated the scope for estimation of genetic CFs, for records taken at individual test days during the course of lactation of dairy cows.

CFs based on orthogonal polynomials or similar functions do not involve any assumptions about the shape of the trajectory or the dispersion structure other than implicit in the choice of order of approximation and, to a lesser extent, of basis functions. In contrast, CFs commonly considered in other areas of statistics are often parametric and restrictive in the structure they permit and are sometimes difficult to justify biologically. However, they generally involve only a small number of parameters.

For the parametric approach, a CF is usually decomposed into a variance function, describing changes in variance with t , and a correlation function. Commonly used correlation functions are stationary, i.e. correlations for records taken at a given lag, $\Delta t = |t_{j'} - t_j|$, are the same for all t_j . A well-known example is the auto-correlation structure with a single parameter ρ , which involves an exponential decay in correlation, $r_{j'j} = \rho^{\Delta t} = \exp\{-\theta\Delta t\}$, with $\theta = -\log(\rho)$. Other functions with one or two parameters exist (e.g. Pletcher & Geyer 1999). Additional parameters can add more flexibility. For instance, a Box-Cox power transformation of t accommodates some non-stationarity (Núñez-Antón & Zimmerman 2000). Pletcher & Geyer (1999) presented an application of such models for the estimation of genetic CF, referring to them as 'character process' models; see also Jaffrézic & Pletcher (2000) for a review.

(b) *Random regression models*

In analyses of data from livestock improvement schemes, measurements along curves have often been treated simply as repeated measurement of the same trait, potentially applying some correction for time or age at recording. In addition, large ranges of observed times have been subdivided into intervals representing individual, correlated traits. A typical example is the weights of cattle, for which we usually distinguish

between birth, weaning, yearling and later weights. A major drawback of these approaches is that they do not fully use the information provided by the times of recording, and thus do not account for any constraints they might impose on the covariance structure.

Alternatively, curves of interest, in particular growth and lactation curves, have been fitted at a phenotypic level, and the parameters of the curve have subsequently been analysed as new traits. Drawbacks of this approach are that estimates of curves may be influenced by systematic environmental effects not taken into account, and that constraints are imposed on the families of curves.

These considerations have motivated the random regression (RR) approach. Estimates of breeding values of livestock, i.e. of the additive genetic values of animals, are usually obtained from mixed model analyses. Fitting animals' additive genetic effects as random effects, analyses involve a linear regression of observations on indicator variables, which have values of unity or zero for animals that do or do not have a record, respectively, for the trait considered. This framework extends readily to other and multiple covariables (Henderson 1982), such as functions of t . This allows any trajectories that can be described in a regression equation to be modelled directly within the mixed model analysis. While the covariables are usually nonlinear functions of t , such as polynomials or splines, the model is linear in the regression coefficients to be estimated. Fitting sets of RR coefficients for each individual and random factor, e.g. additive genetic and permanent environmental effects, then yields estimates of the corresponding trajectories. This is the so-called RR model, also referred to as the random coefficient model in other areas of statistics.

Initial applications of the RR model were in genetic evaluation of dairy cows, using records from individual test days to model the lactation curve (Schaeffer & Dekkers 1994; Jamrozik *et al.* 1997). Over the last decade, the RR model has become a standard for analyses of repeated records from animal breeding schemes, such as test day, growth and feed intake data. A recent review of applications has been given by Schaeffer (2004).

Treating sets of regression coefficients as random effects implies a matrix of covariances among RR coefficients. In turn, these covariances determine the covariance structure among all records along the trajectory (Jamrozik & Schaeffer 1997), i.e. define a CF. In other words, fitting a RR model and estimating CFs approximated by a limited number of terms are equivalent, assuming the corresponding functions of t are fitted (Meyer & Hill 1997; Meyer 1998b). This can be exploited to estimate CFs efficiently (Meyer 1998a).

3. MODELLING FUNCTION VALUED TRAITS

Classical quantitative genetics assumes phenotypic observations are determined by the individuals' additive genetic effects as well as so-called permanent environmental effects, which may encompass any non-additive genetic effects. In addition, there may be systematic environmental effects, and a temporary environmental effect, specific to each record. The same

assumptions are invoked for FV traits, but with most effects replaced by corresponding curves.

(a) The ‘functional’ linear model

Let y_{ij} denote the j th record for individual i taken at time t_j . Assume the trait considered is continuous and has a multivariate normal distribution. A ‘functional’ model is then

$$y_{ij} = M(t_j) + F_{ij} + g_i(t_j) + p_i(t_j) + \varepsilon_{ij}, \tag{3.1}$$

where $M(t_j)$ represents any systematic, fixed effects affecting y_{ij} which change with t . This can simply be a mean trajectory, or represent sets of fixed curves nested within some classification. For instance, we might want to allow for different mean growth curves of male and female animals. F_{ij} denotes any other fixed effects, and ε_{ij} is the temporary environmental effect. Additive genetic and permanent environmental effects are given by random functions $g_i(\cdot)$ and $p_i(\cdot)$ evaluated at t_j , with the subscript i (omitted in the following) emphasizing that these are curves specific to individual i . The additive genetic and permanent environmental covariance functions, $G(t_j, t_{j'})$ and $P(t_j, t_{j'})$, respectively, describe the covariances between effects at different ages.

In principle, $M(\cdot)$, $g(\cdot)$ and $p(\cdot)$ can be chosen among many types of continuous functions, including, for instance, exponential functions such as Gompertz’s or Brody’s curves, which are commonly fitted to model growth curves, and different families of function can be used to model different trajectories in equation (3.1). However, we will restrict our discussion to functions that can be expressed as regression equations, i.e. which are linear in the coefficients of the curve. This assumes that $g(\cdot)$ and $p(\cdot)$ can be expressed as weighted sums of a set of so-called basis functions of t , e.g. polynomials.

Let $\phi_r(t_j)$ denote the r th basis function, evaluated for t_j , and α_{ir} and γ_{ir} be the regression coefficients for additive genetic and permanent environmental effects for the i th individual. Assume we fit k_α and k_γ regression coefficients, respectively. Our model then becomes

$$y_{ij} = M(t_j) + F_{ij} + \sum_{r=1}^{k_\alpha} \alpha_{ir} \phi_r(t_j) + \sum_{r=1}^{k_\gamma} \gamma_{ir} \phi_r(t_j) + \varepsilon_{ij}, \tag{3.2}$$

where the first sum represents the additive genetic, $g(\cdot)$, and the second sum the permanent environmental, $p(\cdot)$, component of the phenotype. Omitting details, $M(\cdot)$, of course, can be expanded in the same fashion.

Theoretically, there are infinitely many terms in the expansion of $g(\cdot)$ or $p(\cdot)$. Hence, equation (3.2) represents a truncated, reduced-rank approximation. In practice, orders of fit k_α and k_γ are often substantially smaller than the number of distinct values of t observed.

(b) Basis functions

A common choice for the basis functions are orthogonal polynomials, as advocated by Kirkpatrick & Heckman (1989). In particular, the Legendre polynomials used by Kirkpatrick *et al.* (1990) have been

used extensively in RR analyses. While polynomials are flexible and have been proven capable of modelling a range of covariance structures, they are also frequently associated with numerical problems, especially for high orders of fit. ‘Runge’s phenomenon’ describes the observation that the error of polynomial approximation of a curve increased with the order of polynomial fit. Moreover, errors were predominantly due to oscillations at the extremes (de Boor 2001). RR analyses fitting higher degree polynomials have frequently encountered erratic estimates of variance components at the highest ages, in particular, for longitudinal data with few records at the extremes.

An alternative to high degree polynomials are ‘piecewise polynomials’, i.e. curves constructed from pieces of low degree polynomials, which are joined smoothly at selected points, so-called splines. A particular type of spline is the B-spline (de Boor 2001). Rice & Wu (2001) suggested the use of the basis functions of a B-spline as covariables to model random curves in mixed model analyses. Although B-splines are not automatically orthogonal, they have good numerical properties that make them attractive (Ruppert *et al.* 2003). While following the lead of White *et al.* (1999), cubic-smoothing splines have been used in a number of studies (e.g. White *et al.* 1999; Huisman *et al.* 2002; DeGroot *et al.* 2003), use of B-splines in RR analyses so far has been limited (Torres & Quaas 2001; Meyer 2005).

(c) Covariance structure

Let α_i , of length k_α , and γ_i , of length k_γ , denote the vectors of RR coefficients for individual i . Assuming a multivariate normal distribution for the data,

$$\alpha_i \sim N(\mathbf{0}, \mathbf{K}_\alpha) \quad \text{and} \quad \gamma_i \sim N(\mathbf{0}, \mathbf{K}_\gamma), \tag{3.3}$$

where $\mathbf{K}_\alpha = \{K_{\alpha lm}\}$ and $\mathbf{K}_\gamma = \{K_{\gamma lm}\}$ are the matrices of covariances between RR coefficients, and define $\text{Cov}(\alpha_i, \gamma_i) = \mathbf{0}$.

Temporary environmental effects are assumed to be independently distributed, but variances are allowed to change with t . This is modelled through a variance function, i.e. $V(\varepsilon_{ij}) = \sigma_j^2 = e(t_j, \sigma^2)$. In the simplest case, $e(\cdot)$ could be a step function. Alternatively, polynomials or spline function of t can be used to model heteroscedasticity of residual variances. Often, logarithmic values of variances are modelled in this way to account for ‘scale effects’, i.e. dependencies between means and variances.

The covariances between RR coefficients determine the covariance between two records for animal i , taken at times t_j and $t_{j'}$.

$$\begin{aligned} \text{Cov}(y_{ij}, y_{ij'}) &= \sum_{r=1}^{k_\alpha} \sum_{s=1}^{k_\alpha} \phi_r(t_j) \phi_s(t_{j'}) K_{\alpha rs} \\ &+ \sum_{r=1}^{k_\gamma} \sum_{s=1}^{k_\gamma} \phi_r(t_j) \phi_s(t_{j'}) K_{\gamma rs} + \delta_{jj'} \sigma_j^2, \end{aligned} \tag{3.4}$$

with $\delta_{jj'} = 1$ for $j=j'$, and zero otherwise. The first part on the right-hand side of equation (3.4) is the additive genetic covariance function $G(\cdot, \cdot)$, evaluated for t_j

and t_j , and the second part is its permanent environmental counterpart, $\mathcal{P}(\cdot, \cdot)$.

(d) Matrix formulation

For N animals with n_i records each, the model is

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{3.5}$$

with \mathbf{y} and $\boldsymbol{\varepsilon}$ the vectors of observations and residuals, respectively, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ the vectors of genetic and permanent environmental RR coefficients, and \mathbf{b} the vector of fixed effects fitted. Expectations of these vectors are $E[\boldsymbol{\alpha}] = \mathbf{0}$, $E[\boldsymbol{\gamma}] = \mathbf{0}$, $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $E[\mathbf{y}] = \mathbf{X}\mathbf{b}$. \mathbf{X} , \mathbf{Z} and \mathbf{W} are the design or incidence matrices. For effects modelled as trajectories, the design matrices contain the basis functions evaluated for the time pertaining to each record, $\phi_r(t_j)$, i.e. are considerably denser than their counterparts in ‘standard’ analyses, where each row of \mathbf{Z} or \mathbf{W} would only have a single non-zero element of unity. The model accommodates any pattern of values of t_j , i.e. individuals may have different numbers of records and records may be taken at arbitrary times.

Generally, analyses include genetic effects for animals which are parents only, i.e. do not have any records. These are included in $\boldsymbol{\alpha}$, so that $\boldsymbol{\alpha}$ has length $N_A k_\alpha$ with $N_A \geq N$ the total number of animals, but corresponding rows of \mathbf{Z} are zero.

Random vectors $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are defined to be uncorrelated, with covariance matrices

$$V(\boldsymbol{\alpha}) = \mathbf{A} \otimes \mathbf{K}_\alpha = \mathbf{G} \quad V(\boldsymbol{\gamma}) = \mathbf{I}_N \otimes \mathbf{K}_\gamma = \mathbf{P}$$

$$V(\boldsymbol{\varepsilon}) = \text{Diag}\{\sigma_k^2\} = \mathbf{R},$$

where \mathbf{A} is the numerator relationship matrix between animals, \mathbf{I}_N an identity matrix of size N , and ‘ \otimes ’ denotes the direct matrix product. Diagonal elements of \mathbf{A} are unity, augmented by the inbreeding coefficient for individuals, and off-diagonal elements give the degree of relationship, e.g. one-half for parents and their offspring or one-fourth for half-sibs. This gives variance of \mathbf{y}

$$\begin{aligned} \mathbf{V} &= \mathbf{Z}(\mathbf{A} \otimes \mathbf{K}_\alpha)\mathbf{Z}' + \mathbf{W}(\mathbf{I}_N \otimes \mathbf{K}_\gamma)\mathbf{W}' + \text{Diag}\{\sigma_m^2\} \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{W}\mathbf{P}\mathbf{W}' + \mathbf{R}. \end{aligned} \tag{3.6}$$

4. ESTIMATION OF BREEDING VALUES

The central tasks of quantitative genetic analyses, estimation of breeding values and estimation of covariance components and genetic parameters generally invoke the linear, mixed model. The so-called mixed model equations (MME) for the model given by equation (3.5) are

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{K}_\alpha^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{I} \otimes \mathbf{K}_\gamma^{-1} \end{pmatrix} \times \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}. \tag{4.1}$$

Assuming variances are known, solutions to equation (4.1) are generalized least-squares (LSQ) estimates of

the fixed effects \mathbf{b} , and best linear unbiased predictors of the random effects $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ (Henderson 1973). For national genetic evaluation schemes, especially multi-trait analyses for dairy or beef cattle, the MME can comprise millions of equations. Hence equation (4.1) is generally solved iteratively. Specialized strategies suitable for large scale, RR-test day models have been discussed by Lidauer *et al.* (1999), Gengler *et al.* (2000) and Jamrozik & Schaeffer (2000).

The MME (equation 4.1) play a central role in statistical analyses in quantitative genetics, and in particular, of animal breeding data. Standard generalized LSQ analyses would require inverting the variance matrix \mathbf{V} , but only inverses \mathbf{R}^{-1} and \mathbf{A}^{-1} are needed in the MME. While \mathbf{R} has size equal to the number of records, assuming zero environmental covariances between records on different animals, it is generally blockdiagonal for animals, if not, as in our case, diagonal. \mathbf{R} is thus readily invertible, and the data part of the MME can be accumulated considering records for one animal at a time. \mathbf{A} has size proportional to the number of animals in the analysis (including parents without records), but \mathbf{A}^{-1} can be set up directly and easily from a list of pedigree records (Henderson 1976; Quaas 1976).

Estimated regression coefficients $\hat{\alpha}_i$ for animal i define its genetic merit along the complete trajectory considered. In some cases, we may be interested in estimated breeding values for target times t_k . These are obtained simply by evaluating the regression equation

$$\text{EBV}(t_k)_i = \sum_{r=1}^{k_\alpha} \hat{\alpha}_i \phi_r(t_k). \tag{4.2}$$

Other functions of the estimated curve may be of interest. For example, integrals of estimated lactation curves to provide an estimate of total lactation genetic merit for dairy cows. For growth curves, first and second derivatives give velocity and acceleration, respectively, and turning points of estimated curves may help distinguish between early and late maturing animals.

Sampling covariances among estimated regression coefficients are given by the inverse of the coefficient matrix in equation (4.1). For most genetic evaluation schemes, however, this matrix is too large to be inverted. Approximation procedures for the inverse and ‘accuracies’ of estimated breeding values for RR analyses have been described by Jamrozik *et al.* (2000) and Tier & Meyer (2004).

5. ESTIMATION OF COVARIANCE FUNCTIONS: I. ‘INDIRECTLY’

Estimates of genetic and environmental CFs can be obtained either directly from the data (see §6), or indirectly. Indirect estimation is, in essence, a two-step procedure. It requires a matrix of covariances for a set of points along the trajectory, to be estimated in a first step. Once such estimates of covariances are available, estimates of CFs for different choices of basis functions and orders of fit can be obtained and compared quickly.

However, in estimating CFs from covariance matrices, usually only one source of variation is considered. This

ignores sampling correlations, for example, between genetic and residual covariances. Hence, results tend to differ somewhat from those obtained in corresponding analyses estimating all CFs simultaneously from the data (e.g. van der Werf *et al.* 1998; Kettunen *et al.* 2000), as the latter provide greater scope for a repartitioning of variation to or from other random effects when order(s) of fit are varied.

Let \mathbf{S} denote the matrix of covariance estimates for a grid of s values t_j , and Φ_S , of size $s \times k$, the matrix of basis functions evaluated for the t_j , with $k \leq s$ the order of fit for the CF. We then assume

$$\mathbf{S} = \Phi_S \mathbf{K} \Phi_S' + \mathbf{O}, \tag{5.1}$$

with \mathbf{K} the matrix of coefficients of the CF, and \mathbf{O} a matrix of residuals.

(a) Least-squares

Kirkpatrick *et al.* (1990) described a weighted LSQ procedure to estimate the coefficients of a genetic CF from known genetic covariances. Kirkpatrick *et al.* (1994) proposed a modified scheme for a phenotypic CF, which allowed temporary environmental variances to be separated from permanent environmental variances. Other studies have employed simple LSQ estimation (e.g. van der Werf *et al.* 1998; Veerkamp & Goddard 1998).

Define \mathbf{X}_S as the $s(s+1)/2 \times k(k+1)/2$ matrix with elements $(2 - \delta_{uv})\phi_u(t_j)\phi_v(t_l)$ corresponding to the jl th element of \mathbf{S} and the uv th element of \mathbf{K} , and δ_{uv} as defined above. Weighted LSQ estimates of the coefficients of the CF are then obtained as

$$\text{vech}(\hat{\mathbf{K}}) = (\mathbf{X}'_S \Omega^{-1} \mathbf{X}_S)^{-1} \mathbf{X}'_S \Omega^{-1} \text{vech}(\mathbf{S}), \tag{5.2}$$

where ‘vech’ is the operator that stacks the columns of the lower triangle of a symmetric matrix into a vector (e.g. Harville 1997), and Ω is a matrix of weights.

(b) Maximum likelihood

A drawback of the LSQ approach is that it does not guarantee the estimated CFs to be positive semi-definite, i.e. to yield estimates within the parameter space (Kirkpatrick *et al.* 1990). This can be overcome in a maximum likelihood framework, as proposed by Mäntysaari (1999).

In essence, this involves treating the estimated covariances (\mathbf{S}) as if they were crossproducts arising from data, and equating them to their ‘expected’ values, shown above (equation (5.1)), comprising sums of products of coefficients of the CF and base functions evaluated at the grid points. Let $\mathbf{F} = E[\mathbf{S}]$. The log likelihood ($\log \mathcal{L}$) to be maximized is then (Thompson 1976)

$$\log \mathcal{L} = \text{const.} - \frac{1}{2} d(\log |\mathbf{F}| + \text{tr}(\mathbf{S}\mathbf{F}^{-1})), \tag{5.3}$$

where d are the degrees of freedom associated with \mathbf{S} . Thompson (1976) outlined an iterative procedure to maximize equation (5.3). Let θ denote the vector of parameters to be estimated, and $\hat{\theta}^q$ the estimate from

the q th iterate.

$$\begin{aligned} \hat{\theta}^{q+1} &= (\mathbf{H}(\hat{\theta}^q))^{-1} \mathbf{g}(\hat{\theta}^q) \\ &\text{with } \mathbf{H}(\hat{\theta}^q) = \{d \text{tr}(\mathbf{F}^{-1}(\partial \mathbf{F} / \partial \theta_u) \mathbf{F}^{-1}(\partial \mathbf{F} / \partial \theta_v))\} \\ &\text{and } \mathbf{g}(\hat{\theta}^q) = \{d \text{tr}(\mathbf{F}^{-1} \mathbf{S} \mathbf{F}^{-1}(\partial \mathbf{F} / \partial \theta_u))\}, \end{aligned} \tag{5.4}$$

where $\mathbf{H}(\hat{\theta}^q)$ is the matrix of expected values of second derivatives and $\mathbf{g}(\hat{\theta}^q)$ the vector of the data part of the first derivatives of $\log \mathcal{L}$ with respect to the elements of θ , both evaluated for \mathbf{F} as given by $\hat{\theta}^q$. Alternatively, the expectation-maximization approach, used to estimate permanent environmental CFs and temporary environmental variances from matrices of residual covariances (Mäntysaari 1999; Emmerling *et al.* 2002; Koivula *et al.* 2004), may be useful to estimate genetic CFs where we impose a structure on \mathbf{K} , such as considering the largest eigenvalues only (see §7).

(c) Constructed CF

Estimation of CF from covariance matrices also has been found useful for cases where we might have reliable estimates for some variance components, but have to rely on literature values or assumed relationships (such as the auto-correlation function described above) for estimates of some correlations. Misztal *et al.* (2000) outlined an approach to construct matrices of covariances \mathbf{S} for this scenario, proposing linear interpolation or variance function of standard shapes to derive missing components, and to obtain LSQ estimates of CFs. An application for CFs for growth of beef cattle has been presented by Legarra *et al.* (2004).

6. ESTIMATION OF COVARIANCE FUNCTIONS: II. ‘DIRECTLY’

Alternatively, CFs can be estimated in a single step, directly from the data. As shown above, the coefficients of CFs are equal to the covariances between RR coefficients in a functional, linear mixed model. Hence, standard mixed-model-based variance component procedures can be used to estimate CFs directly from the data.

(a) Restricted maximum likelihood

Early applications of restricted maximum likelihood (REML) used a simple reparametrization of standard, multivariate models to estimate CFs (Meyer & Hill 1997). This required MME of size proportional to the number of distinct values of t to be manipulated, and thus did not accommodate data from livestock improvement schemes involving records at ‘all ages’. Hence, today the coefficients of CFs are predominantly estimated as the covariances among RR coefficients.

REML estimation maximizes the part of the likelihood independent of fixed effects. Assuming unstructured covariance matrices, i.e. imposing no restrictions on \mathbf{K}_α and \mathbf{K}_γ other than that they should be positive semi-definite and symmetric, the vector of parameters to be estimated, θ , comprises the $k_\alpha(k_\alpha + 1)/2$ unique elements of \mathbf{K}_α , the $k_\gamma(k_\gamma + 1)/2$ elements of \mathbf{K}_γ , and k_e parameters modelling variances due to temporary environmental effects. The REML log likelihood

(log \mathcal{L}) is then given by

$$\begin{aligned}
 -2 \log \mathcal{L}(\theta) &= \text{const.} + \log |V^{-1}| + \log |X'V^{-1}X| \\
 &\quad + (\mathbf{y} - X\hat{\beta})'V^{-1}(\mathbf{y} - X\hat{\beta}) \\
 &= \text{const.} + N_A \log |K_\alpha| + k_\alpha \log |A| \\
 &\quad + N \log |K_\gamma| + \log |R| + \log |C| + \mathbf{y}'P\mathbf{y},
 \end{aligned}
 \tag{6.1}$$

where C is the coefficient matrix in the MME (see equation (4.1)), and P a projection matrix, $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$, so that $P\mathbf{y} = R^{-1}(\mathbf{y} - X\hat{\beta} - Z\hat{\alpha} - W\hat{\gamma})$. The first line of equation (6.1) is the standard form of log \mathcal{L} (e.g. Harville 1977), expressed in terms of the phenotypic variance matrix V . The formulation of the second line, however, is computationally advantageous. K_α and K_γ are small, of size $k_\alpha \times k_\alpha$ and $k_\gamma \times k_\gamma$, respectively, and their determinants are readily evaluated. Log $|A|$ arises as a by-product when setting up A^{-1} from a list of pedigree information. The last two terms, $\mathbf{y}'P\mathbf{y}$ and $\log |C|$, can be computed simultaneously through a Cholesky factorization of C augmented by the vector of right-hand sides in equation (4.1) and its transpose, and the residual sum of squares (Graser *et al.* 1987).

$M =$

$$\begin{pmatrix}
 X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W & X'R^{-1}\mathbf{y} \\
 Z'R^{-1}X & Z'R^{-1}Z + A^{-1} \otimes K_\alpha^{-1} & Z'R^{-1}W & Z'R^{-1}\mathbf{y} \\
 W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + I \otimes K_\gamma^{-1} & W'R^{-1}\mathbf{y} \\
 \mathbf{y}'R^{-1}X & \mathbf{y}'R^{-1}Z & \mathbf{y}'R^{-1}W & \mathbf{y}'R^{-1}\mathbf{y}
 \end{pmatrix}
 \tag{6.2}$$

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

$$\log |C| = 2 \sum_{i=1}^{M-1} \log(l_{ii}) \quad \text{and} \quad \mathbf{y}'P\mathbf{y} = l_{MM}^2,
 \tag{6.3}$$

REML estimates of K_α and K_γ have to be positive semi-definite, and estimates of temporary environmental variances cannot be negative (Harville 1977). Hence, maximization of log \mathcal{L} , as given in equation (6.1), represents a constrained optimization problem. A reparametrization can be used to remove constraints on the parameter space. For instance, instead of estimating the unique elements of a covariance matrix K , we can estimate the elements of its Cholesky factor, taking logarithmic values of the diagonals (Meyer & Smith 1996; Pinheiro & Bates 1996). This not only allows use of unconstrained maximization procedures, but can also improve rates of convergence in an iterative estimation scheme (Groeneveld 1994).

The Newton–Raphson algorithm, with various modifications, is frequently applied in maximum likelihood estimation (Jennrich & Sampson 1976). It is an iterative scheme, which requires first and second derivatives of log \mathcal{L} . For Fisher’s ‘method of scoring’, the latter are replaced by their expected values, but, in most quantitative genetic applications, both are difficult to compute. Hence, today the so-called ‘average information’ (AI) algorithm (Gilmour *et al.* 1995), which approximates second derivatives with the average of observed and expected values, is generally

preferred. These averages are equivalent to second derivatives of the ‘data part’ of log \mathcal{L} , $\mathbf{y}'P\mathbf{y}$. They are relatively easy to calculate, while yielding convergence rates similar to those using second derivatives or their expectations; see Hofer (1998) for a review.

Algorithms for multivariate analyses via AI REML are readily adapted to the estimation of covariances among RR coefficients. Jensen *et al.* (1997) employ sparse matrix inversion of C to obtain first derivatives of log \mathcal{L} . An alternative, relying on automatic differentiation of the Cholesky factor of M (see equation (6.2); Smith 1995), is outlined by Meyer (1998a). This requires specifications of first derivatives of M with respect to the covariance components to be estimated, which have a simple form. An extension, accommodating variance functions to model changes in temporary environmental variances with t and parametric correlation structures for covariances among permanent environmental effects is given by Meyer (2001b). Other authors described expectation–maximization type algorithms, and their extensions to improve convergence, for RR models (Gengler *et al.* 1999; Foulley & van Dyk 2000; Foulley *et al.* 2000).

Most software packages available for mixed model analyses allow RR models to be fitted, and the corresponding covariance matrices to be estimated.

(b) Bayesian analysis

Even for comparatively low orders of fit, RR models involve substantial numbers of covariance components to be estimated. Typically, large datasets are required to estimate so many parameters with sufficient accuracy. However, high computational requirements limit the size of datasets and models that can be analysed using REML.

Bayesian estimation has become a standard for quantitative genetic analyses (Sorensen & Gianola 2002). In particular, schemes sampling from fully conditional posterior distributions of the parameters of interest are widely used. These are computationally easy to implement. While computing times required to sample long or multiple Markov chains may be large, memory requirements are considerably smaller than for corresponding REML analyses. This facilitates analyses of much larger datasets. In addition, as estimates of complete posterior distributions are obtained, sampling properties of estimates or functions thereof are readily examined.

RR analyses using a Gibbs sampler have been applied initially to the analysis of test day records of dairy cows. Full details are given by Jamrozik & Schaeffer (1997) and Rekaya *et al.* (1999). Early models fitted a simple step function for temporary environmental variances. Variance functions (e.g. Schnyder *et al.* 2001) or changepoint techniques (Lopez-Romero *et al.* 2004) were used subsequently, but required Metropolis–Hasting steps (Sorensen & Gianola 2002). Jamrozik (2004) discusses implementation issues of Markov chain Monte Carlo methods for RR analyses.

7. EIGENFUNCTIONS AND BEYOND

CFs can involve a large number of parameters. Hence estimation can be computationally difficult, and

estimates can be subject to large sampling variances. This has motivated the use of principal component methods to the analysis of FV traits (Kirkpatrick & Meyer 2004).

An eigenvalue decomposition of a covariance matrix \mathbf{K} gives the matrix as the product of the matrix of eigenvectors \mathbf{E} and the diagonal matrix of eigenvalues $\mathbf{\Lambda}$.

$$\mathbf{K} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}', \tag{7.1}$$

\mathbf{E} is orthonormal, i.e. $\mathbf{E}\mathbf{E}' = \mathbf{I}$. The columns of \mathbf{E} , or eigenvectors, define linear transformations of the original variables to new variables, which are uncorrelated and have variances as given by the corresponding eigenvalues. These variables are commonly referred to as ‘principal components’ (PC). Eigenvectors and -values are usually ordered in descending order of the latter, and the i th PC explains maximum variation given PCs $1, \dots, i-1$ (e.g. Jolliffe 1986).

The FV equivalent to eigenvectors are eigenfunctions, for curves, or eigenfaces, for images. Eigenfunctions are useful in visualizing and analysing patterns of variation of FV traits. Like CFs, eigenfunctions, theoretically, have infinite dimensions, but are, in practice, approximated as the weighted sum of a limited number of basis functions.

If the basis functions are orthogonal, estimates of the eigenfunctions of a CF can be obtained directly from the eigenvectors of the coefficient matrix of the CF (Kirkpatrick & Heckman 1989). Let \mathbf{K} of size $k \times k$ be the coefficient matrix of a CF. The estimates of eigenfunctions are then

$$\hat{k}_i(t) = \sum_{r=1}^k \phi_r(t)e_{ir}, \tag{7.2}$$

where e_{ir} are the elements of the i th eigenvector of \mathbf{K} . Evaluating equation (7.2) for all values of t yields a curve, the eigenfunction. Similarly, the eigenvalues of \mathbf{K} provide estimates of the eigenvalues of the CF. For non-orthogonal basis functions, we need to adjust for the basis (James *et al.* 2000). Alternatively, we can extract the eigenfunctions and -values numerically, by evaluating the CF for a fine grid of values of t , and calculating the eigendecomposition of the resulting covariance matrix (Rice & Wu 2001).

Eigenfunctions are one of the main characteristics of CFs examined in FDA. As discussed below (§8), eigenfunctions of genetic CFs provide an insight into the expected transformation of trajectories when subject to selection.

(a) Truncated expansion

Principal component analysis is widely employed as a dimension-reduction technique. Considering only the first m PCs with the largest eigenvalues, we reduce the number of variables, while still capturing a maximum of variation. If the eigenvalues for PCs $m+1, \dots, k$ are small, loss of information can be negligible. The same argument applies for FV traits. Moreover, we can choose to use the eigenfunctions as basis functions, and truncate our expansion by considering the most variable PCs only (Kirkpatrick & Meyer 2004). This is the Karhunen–Loève expansion favoured in FDA

(Ramsay & Silverman 1997). As it is based on PCs of the CF, it is the expansion which, for a given number of terms, minimizes the residual.

Let $\kappa_{\alpha r}(t)$ and $\kappa_{\gamma r}(t)$ denote the r th eigenfunction of $\mathcal{G}(\cdot, \cdot)$ and $\mathcal{P}(\cdot, \cdot)$, the additive genetic and permanent environmental CFs, respectively. The functional, linear model equation (3.2) then becomes

$$y_{ij} = M(t_j) + F_{ij} + \sum_{r=1}^{m_\alpha} \alpha_{ir}^* \kappa_{\alpha r}(t_j) + \sum_{r=1}^{m_\gamma} \gamma_{ir}^* \kappa_{\gamma r}(t_j) + \varepsilon_{ij}, \tag{7.3}$$

with α_{ir}^* and γ_{ir}^* the r th additive genetic value and permanent environmental effect for animal i . The corresponding CFs are directly given by the eigenfunctions

$$\begin{aligned} \mathcal{G}(t_j, t_{j'}) &= \sum_{r=1}^{m_\alpha} \lambda_{\alpha r} \kappa_{\alpha r}(t_j) \kappa_{\alpha r}(t_{j'}) \quad \text{and} \\ \mathcal{P}(t_j, t_{j'}) &= \sum_{r=1}^{m_\gamma} \lambda_{\gamma r} \kappa_{\gamma r}(t_j) \kappa_{\gamma r}(t_{j'}). \end{aligned} \tag{7.4}$$

As emphasized by Kirkpatrick & Meyer (2004), there are two levels of truncation. First, we approximate the CF by its first m eigenfunctions only, assuming infinite-dimensional CF and eigenfunctions. Second, we estimate the eigenfunctions using a limited number of terms, $k \geq m$, as shown in equation (7.2).

(b) Estimating eigenfunctions directly

Genetic PCs or eigenfunctions have usually been estimated by first obtaining full rank estimates of the genetic covariance matrix, and then carrying out an eigenvalue decomposition of the matrix. Alternatively, PCs have been determined at the phenotypic level, and genetic parameters of the new variables have been estimated (e.g. Chase *et al.* 2002). A better approach would be to estimate the genetic eigenfunctions directly, and, at the same time, to restrict estimation to the most important eigenfunctions (Kirkpatrick & Meyer 2004). Fortunately, this can readily be done within the mixed-model framework described above. In essence, it only requires a simple reparametrization of the functional linear model given by equation (7.3).

Let $\mathbf{Z}^* = \mathbf{Z}(\mathbf{E}_\alpha \otimes \mathbf{I}_{N_A})$, $\mathbf{W}^* = \mathbf{W}(\mathbf{E}_\gamma \otimes \mathbf{I}_N)$, $\boldsymbol{\alpha}^* = (\mathbf{E}'_\alpha \otimes \mathbf{I}_{N_A})\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}^* = (\mathbf{E}'_\gamma \otimes \mathbf{I}_N)\boldsymbol{\gamma}$, where \mathbf{E}_α and \mathbf{E}_γ are the matrices of eigenvectors of \mathbf{K}_α and \mathbf{K}_γ , respectively. These matrices, \mathbf{E}_α and \mathbf{E}_γ , are orthonormal, so at full rank

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}^* \boldsymbol{\alpha}^* + \mathbf{W}^* \boldsymbol{\gamma}^* + \boldsymbol{\varepsilon}, \tag{7.5}$$

is an equivalent model to equation (3.5). The MME for equation (7.5) are of the same form as given in equation (4.1), but with \mathbf{Z} , \mathbf{W} , $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ replaced by \mathbf{Z}^* , \mathbf{W}^* , $\boldsymbol{\alpha}^*$ and $\boldsymbol{\gamma}^*$, respectively. Furthermore, we have diagonal matrices of eigenvalues $\mathbf{\Lambda}_\alpha$ and $\mathbf{\Lambda}_\gamma$, instead of \mathbf{K}_α and \mathbf{K}_γ .

Elements of the vector $\boldsymbol{\alpha}^*$, α_{ir}^* , are the additive genetic effects for the eigenfunctions. This implies that estimates of the α_{ir}^* can be used to select animals for desired transformations in the r th eigenfunction.

If we truncate \mathbf{E}_α to the first m_α columns only, $\boldsymbol{\alpha}^*$ has only m_α elements for each animal, corresponding to the genetic effects for the leading eigenfunctions.

Moreover, the number of equations in the MME can be greatly reduced, as the number of genetic effects fitted is now proportional to m_α rather than k_α . Analogously, we can restrict the order of fit for permanent environmental effects to $m_\gamma < k_\gamma$.

(c) *Reduced rank covariance functions*

Considering the leading eigenfunctions only, the corresponding CFs (equation (7.4)) have reduced rank, $m_\alpha < k_\alpha$ and $m_\gamma < k_\gamma$. We can estimate the CFs, i.e. the coefficients of the eigenfunctions and the corresponding eigenvalues, based on the reparametrized mixed model equation (7.5) in the same way as described in §6 for the general RR model. Only REML estimation is considered here.

For a trajectory modelled by k RR coefficients, the full-rank coefficient matrix has $k(k+1)/2$ covariances among RR to be estimated. If we fit the first m eigenfunctions only, the number of parameters describing the CF is reduced to $m(2k-m+1)/2$. While there are km elements in the first m eigenfunctions and m eigenvalues, the number of parameters is less as \mathbf{E} is orthonormal. To accommodate these constraints easily, Kirkpatrick & Meyer (2004) suggested replacing matrices \mathbf{E} above with $\mathbf{Q} = \mathbf{E}\mathbf{A}^{1/2}$, i.e. to scale the eigenfunctions by the square root of the corresponding eigenvalues. As shown in Meyer & Kirkpatrick (2005), the r th column of \mathbf{Q} then has $k+1-r$ 'free' elements, while the remaining $r-1$ elements are given by the orthogonality constraints on \mathbf{Q} , and can be determined as solutions to a linear system of equations.

This parametrization puts all parameters to be estimated into the design matrices, now $\mathbf{Z}^* = \mathbf{Z}(\mathbf{Q}_\alpha \otimes \mathbf{I}_{N_A})$ and $\mathbf{W}^* = \mathbf{W}(\mathbf{Q}_\gamma \otimes \mathbf{I}_N)$. As the scaled eigenfunctions α^* and γ^* have covariance matrices \mathbf{I}_{m_α} and \mathbf{I}_{m_γ} , the (log) likelihood to be maximized is reduced to

$$-2 \log \mathcal{L} = \text{const.} + m_A \log |\mathbf{A}| + \log |\mathbf{R}| + \log |\mathbf{C}^*| + \mathbf{y}' \mathbf{P} \mathbf{y}, \quad (7.6)$$

where \mathbf{C}^* is the coefficient matrix in the correspondingly reparametrized MME (see Meyer & Kirkpatrick 2005 for full details).

Strategies outlined above (§6) to evaluate and maximize $\log \mathcal{L}$ are directly applicable to the estimation of eigenfunctions. An AI-REML algorithm is described in Meyer & Kirkpatrick (2005). With the number of random effects fitted and the size of the MME to be handled proportional to m_α and m_γ rather than k_α and k_γ , computational requirements can be reduced dramatically compared with full-rank analyses.

Parametrizing to the 'free' elements of matrices \mathbf{Q} , maximization of $\log \mathcal{L}$ represents an unconstrained optimization problem. Estimates of the corresponding eigenvalues can be obtained by calculating the vector norm for the columns of \mathbf{Q} . However, with \mathbf{V} highly nonlinear in the parameters to be estimated, convergence of standard algorithms for this parametrization has been found to be slow (James *et al.* 2000; Meyer & Kirkpatrick 2005). This can be alleviated by a rotation of the parameter space (Smith *et al.* 2001).

Let $\mathbf{Q}^* = \mathbf{Q}\mathbf{T}$, where \mathbf{T} defines an orthogonal transformation. Since $\mathbf{T}\mathbf{T}' = \mathbf{I}$, this does not affect the

likelihood. A suitable choice might be $\mathbf{T} = \mathbf{E}'$, so that $\mathbf{Q}^* = \mathbf{E}\mathbf{A}^{1/2}\mathbf{E}'$, which is the matrix square root of \mathbf{K} . Alternatively, consider the Cholesky factorization $\mathbf{K} = \mathbf{L}_K \mathbf{L}_K'$, and the singular value decomposition of $\mathbf{L}_K = \mathbf{U}_1 \mathbf{D} \mathbf{U}_2'$. The left singular vectors of \mathbf{L}_K are equal to the eigenvectors of \mathbf{K} and the corresponding singular values are the square root of the eigenvalues (Harville 1997, p. 232), i.e. $\mathbf{Q} = \mathbf{U}_1 \mathbf{D}$. Hence, choosing the matrix of right singular vectors as our transformation, $\mathbf{T} = \mathbf{U}_2$, the rotated matrix of parameters to be estimated is $\mathbf{Q}^* = \mathbf{L}_K$. This implies that we can maximize $\log \mathcal{L}$ with respect to the non-zero elements of the first m columns of the Cholesky factor of \mathbf{K} , to obtain REML estimates of the first m eigenfunctions of \mathbf{K} . This parametrization has been found to have good convergence properties (Groeneveld 1994; Hofer 1998; Thompson & Mäntysaari 1999).

(d) *Sampling properties*

As well as reducing computational requirements, estimation of only the most important eigenfunctions can yield estimates with smaller errors. This is illustrated with a small simulation study.

Matrices of mean crossproducts for 500 unrelated sires with four progeny each and three or five records per progeny, were sampled from a Wishart distribution. Genetic effects were assumed to have CF as considered by Kirkpatrick *et al.* (1990), i.e. $\text{vech}(\mathbf{K}_\alpha)' = (1348.13, 66.55-111.68, 24.27-14.01, 14.51)$ with $k_\alpha = 3$ and eigenvalues $\lambda_{\alpha 1} = 1360.8$, $\lambda_{\alpha 2} = 24.5$ and $\lambda_{\alpha 3} = 1.5$. For simplicity, records were assumed to be taken at the same time for all animals and equally spaced. Permanent environmental effects were considered absent, and temporary environmental variances were assumed to be homogeneous with $\sigma^2 = 625$. Maximum likelihood estimates were obtained as described above (§5), considering increasing numbers of eigenfunctions, i.e. $m_\alpha = 1, 2$ and 3 . Accuracy of estimates was measured as relative error, $(\hat{\theta}/\theta - 1) \times 100$ (in %), for $\theta = \sigma^2, \lambda_1, \lambda_2, \lambda_3, \mathcal{G}$. The error in the genetic CF, \mathcal{G} , was evaluated by numerical integration. Errors in estimated eigenfunctions were measured as the angle between 'true' and estimated eigenvectors (\mathbf{e}_r) of \mathbf{K}_α , i.e. $(180/\pi) \arccos(\hat{\mathbf{e}}_r' \mathbf{e}_r / (|\hat{\mathbf{e}}_r| |\mathbf{e}_r|))$ (in °), where $|\cdot|$ is the norm of a vector; see Kirkpatrick & Meyer (2004) for details.

Results are summarized in table 1. When estimating the first eigenfunction only, estimates of the residual variance tended to be inflated. Similarly, estimates of eigenvalue $\lambda_{\alpha 1}$ were somewhat lower than the population values, though deviations were well within the range of sampling errors. The first eigenfunction, $\kappa_{\alpha 1}$, was estimated accurately for all analyses, with mean deviations from the true direction less than 1°. While error in estimates of \mathcal{G} were largest for $m_\alpha = 1$, differences to the analyses fitting more components were small. Results indicated that there was little advantage in fitting the second eigenfunction, and that there was nothing to be gained by estimating the third eigenfunction.

8. SELECTION ON CURVES

An important component of quantitative genetics is the selection of animals with desired characteristics, and

Table 1. Means (\bar{x}) and standard deviations (s.d.) of relative errors in parameter estimates (50 000 replicates).

fit	relative error in estimates of								
	σ^2 (%)	$\lambda_{\alpha 1}$ (%)	$\lambda_{\alpha 2}$ (%)	$\lambda_{\alpha 3}$ (%)	$\kappa_{\alpha 1}$ (°)	$\kappa_{\alpha 2}$ (°)	$\kappa_{\alpha 3}$ (°)	\mathcal{G} (%)	
<i>three records/animal</i>									
$m_\alpha=1$	\bar{x}	6.96	-3.03	—	—	0.80	—	—	5.11
	s.d.	2.38	4.85	—	—	0.43	—	—	2.78
$m_\alpha=2$	\bar{x}	0.44	-0.19	-0.95	—	0.73	8.40	—	4.28
	s.d.	2.95	4.93	32.17	—	0.39	7.24	—	2.61
$m_\alpha=3$	\bar{x}	-1.83	0.98	15.07	156.32	0.75	8.23	8.22	4.53
	s.d.	4.32	5.18	40.04	354.45	0.39	6.80	6.81	2.79
<i>five records/animal</i>									
$m_\alpha=1$	\bar{x}	4.28	-1.01	—	—	0.63	—	—	3.94
	s.d.	1.65	3.95	—	—	0.34	—	—	1.97
$m_\alpha=2$	\bar{x}	0.26	-0.06	-0.26	—	0.60	6.84	—	3.49
	s.d.	1.79	3.95	24.25	—	0.32	5.57	—	2.11
$m_\alpha=3$	\bar{x}	-0.31	0.10	3.25	48.25	0.61	6.64	6.63	3.50
	s.d.	1.93	3.96	24.73	185.53	0.32	5.28	5.30	2.11

the prediction of response to selection. Standard methods to quantify selection and predict response, developed for traits which are points, extend naturally to FV traits.

(a) Quantifying selection

Two related measures for the strength of selection on quantitative traits are in use. The first is the selection differential function, which is the difference in the means of the selected individuals ($\bar{M}(t)^*$) and the population as a whole ($\bar{M}(t)$),

$$s(t) = \bar{M}(t)^* - \bar{M}(t), \tag{8.1}$$

If only differentials at individual points are known, we may estimate $s(t)$ by interpolation. With artificial selection, the selection differential is controlled by the breeder, while in studies of natural populations it needs to be estimated, for instance, from data on survival of individuals and their trait values.

The second measure is the selection gradient function, $\psi(\cdot)$. It is a more informative measure in a number of situations, and has a simple interpretation. The value of the selection gradient, $\psi(t)$, reflects the intensity of selection acting to increase or decrease the population's mean at point t . In contrast, the selection differential, $s(t)$ also includes the effects of selection acting at other points.

It is not always possible to observe the selection differential directly. For example, we may be interested in a trait that changes with age, and individuals die at different ages. We cannot observe the trait values at a given age for individuals that die at a younger age, so we cannot estimate the selection differential at that age. We can, however, still estimate $\psi(t)$. The selection gradient is equal to the partial regression of relative fitness on to the trait value (Lande & Arnold 1983). We therefore estimate $\psi(t)$ at each of several ages, then link these points into a continuous function by interpolation (Kirkpatrick & Lofsvold 1992). A similar situation appears in studies of reaction norms, where an individual's phenotype depends on the environment. If each individual can only be observed in a single environment, we cannot observe the selection differential function $s(\cdot)$ directly. However, we can estimate

$\psi(t)$ within each of several environments t by regressing the trait value on to fitness, and then interpolate, taking into account the frequency with which the environments are encountered (Gomulkiewicz & Kirkpatrick 1992).

Another use of the selection gradient is in predicting evolutionary trajectories, based on assumptions about how a trait affects fitness. Given a fitness function and the distribution of the trait, we can calculate mean fitness, \bar{W} , in a population. When relative fitnesses for all individuals are constant over time, the selection gradient is equal to the gradient in mean fitness as a function of the trait mean (Gomulkiewicz & Beder 1996)

$$\psi(t) = d \ln(\bar{W})/d\bar{M}(t). \tag{8.2}$$

Intuitively, the strength of directional selection acting at each age is equal to the rate at which changing the mean for the trait at that age would increase the population's mean fitness.

(b) Response to selection

Given a population's additive genetic covariance function and either the selection differential or the selection gradient, we can calculate the population's mean function in the next generation (Kirkpatrick & Heckman 1989; Beder & Gomulkiewicz 1998)

$$\bar{M}'(t) = \bar{M}(t) + \int \mathcal{G}(t, x)\psi(x) dx. \tag{8.3}$$

If an estimate of $s(\cdot)$ is available, the selection gradient function can be obtained by evaluating both $s(t)$ and the phenotypic covariance function, $\mathcal{V}(t, x)$ on a fine grid of points, to form \mathbf{s} and \mathbf{V} , respectively. An estimate of $\psi(t)$ is then obtained by interpolating the vector $\mathbf{p} = \mathbf{V}^{-1}\mathbf{s}$.

As outlined above (§7), it is advantageous to formulate the functional linear model in terms of its genetic and environmental PCs. Likewise, response to selection can be described by the genetic PCs. We can rewrite the selection gradient function as

$$\psi(t) = \sum_{r=1}^{k_g} w_r \phi_r(t), \tag{8.4}$$

where $\phi_r(t)$ are the same basis functions as used to estimate the PCs, and $\mathbf{w} = \{w_r\}$ are a set of regression coefficients. These can be estimated, for instance, using LSQ. For orthonormal basis functions, such as Legendre polynomials, the predicted response to selection for m_α genetic PCs, estimated by corresponding eigenvectors $\mathbf{e}_{\alpha ir} = \{e_{\alpha ir}\}$, is then

$$\int \mathcal{G}(t, x) \psi(x) dx = \sum_{i=1}^{m_\alpha} \left[\left(\sum_{r=1}^{k_\alpha} e_{\alpha ir} w_r \right) \left(\lambda_{\alpha i} \sum_{r=1}^{k_\alpha} e_{\alpha ir} \phi_r(t) \right) \right]. \quad (8.5)$$

The form of the right-hand side of equation (8.5) is convenient, as it does not involve integrals. Furthermore, it shows how selection response is determined by the interactions between the selection gradient and the genetic PCs. The first product, $\mathbf{e}_{\alpha i}^T \mathbf{w}$ in vector form, measures how strongly selection ‘loads’ on to the respective PCs, i.e. to what extent the deformation of the growth trajectory favoured by selection coincides with that mode of genetic variation. The second product, $\lambda_{\alpha i} \mathbf{e}_{\alpha i}^T \mathbf{K}_{\alpha i}$, represents the amount of genetic variation available for the i th direction.

(c) Selection strategies

When selecting for several, correlated traits, selection indices are used to combine estimates of genetic merit for individual traits, and to identify animals which maximize the overall response to selection. Index weights are derived as a function of the phenotypic covariance matrix among the selection criteria, and the genetic covariances with the selection objective. Again, this extends readily to FV traits (Kirkpatrick & Bataillon 1999; Kirkpatrick 2002). Togashi & Lin (2003) demonstrated equivalence of selection based on an index of estimated breeding values for part-lactations and selection based on estimates of RR coefficients, to improve persistency of lactation in dairy cows.

Genetic changes in the mean trajectory due to selection can be quantified using the PC approach described above (§7). Genetic PCs or eigenfunctions that have large eigenvalues represent the deformations in the shape of the mean trajectory for which the population has large amounts of genetic variation, i.e. they show the kind of changes readily achieved by selection (Kirkpatrick *et al.* 1990). Further, these PCs indicate the genetic changes likely to occur as correlated responses to selection. For example, the leading PC for growth of animals is typically positive for all ages. This implies that an increase or decrease of the mean at all ages is relatively easy to produce. Conversely, selection to increase (or decrease) mean size at any one age will result in a corresponding change in size for all other ages as a correlated response to selection. PCs that change sign indicate trade-offs: selection that increases the mean at one age will decrease the mean at other ages. Optimal breeding strategies for FV traits can readily be designed by deriving selection index weights for individual PCs (van der Werf 2002).

9. EXTENSIONS

For simplicity, the model considered so far was univariate and comprised random effects of the animal

only. Extensions to more complicated scenarios, involving additional random effects, multiple traits or more than one control variable, are conceptually straightforward. Implementations, however, have been hampered by computational and data structure problems and require further research.

Models including additional random effects have been used in the analysis of data measured on pigs where litter effects were to be modelled (Schnyder *et al.* 2001; Huisman *et al.* 2002), and growth data from beef cattle (e.g. Albuquerque & Meyer 2001; Meyer 2001a; Nobre *et al.* 2003), where both additive genetic and permanent environmental effects of the dam needed to be taken into account.

(a) Multivariate analyses

In some cases, we may have records on several FV traits. Commonly, these have the same control variable t . Examples are morphological characters that change with age (Gomulkiewicz & Kirkpatrick 1992), or feed intake and live weight of animals (e.g. Veerkamp & Thompson 1999). Each trait is then modelled by a set of RR regression coefficients, and we need to estimate covariance matrices among RR coefficients of size equal to the total number of coefficients across all traits. Even for small numbers of traits and low orders of fit, this can result in substantial numbers of parameters to be estimated. Reduced rank estimation, as described in (§7), then becomes all the more appealing. The coefficients of the CFs for individual traits are then given by the diagonal blocks of the estimated covariance matrices, while the off-diagonal blocks determine the cross-covariance functions between traits (Meyer & Hill 1997).

This encompasses the scenario where we have a mixture of FV traits and traits that have a single record only. At the genetic level, we can simply treat the latter as ‘curves’, which are described by a single RR coefficient. At the environmental level, we cannot separate temporary from permanent environmental effects for the traits with single records. Hence we need to parametrize the mixed model, so that only the sum of environmental effects and their covariances are required. In a REML framework of estimation, this is readily done by fitting an equivalent model that incorporates permanent environmental effects in the residual; see Meyer (2001b) for details. For Bayesian analyses via Gibbs sampling, constraints on parameters have been used for this purpose (Schnyder *et al.* 2002).

(b) Multiple control variables

In other cases, we may have more than one control variable, for instance, time or age and an environmental variable, such as temperature, season or level of production in a herd. Alternatively, we may have spatial data, described by longitude and latitude. Unless we can assume the different control variables to act independently, we need to model the resulting surface by regressing on bivariate basis functions. The CFs are then functions of four variables. More control variables are conceptually possible, but are intrinsically difficult and have not been considered. Again, the number of parameters to be estimated increases rapidly with increasing dimension. Moreover, computational

difficulties and incidence of numerical problems are likely to be exacerbated. As above, direct estimation of the leading eigenfunctions only may make analyses more tractable.

(c) Reaction norms

A particular application of the FV approach involves the so-called reaction norms, i.e. the curves that describe how a phenotype is expressed as a function of an environmental variable. These have long been considered in evolutionary biology, where both standard (Via & Lande 1985) and FV traits have been studied (Gomulkiewicz & Kirkpatrick 1992; Kirkpatrick 1993). Only recently, however, have reaction norms attracted interest in modelling of data from livestock improvement programmes, in particular, in relation to genotype \times environment interactions (de Jong & Bijma 2002). This concept assumes that a given genotype has a different phenotype in different environments, i.e. that it has phenotypic 'plasticity' (Falconer 1952). Considering non-FV traits, importance of and models for variation in environmental variability linked to genetic effects have recently been examined by Hill (2002) and Hill & Zhang (2004).

Reaction norms are readily modelled in an RR framework, if a suitable environmental variable can be identified. Corresponding RR analyses have recently been carried out for data from livestock improvement programmes (e.g. Kolmodin *et al.* 2002; Calus & Veerkamp 2003). Optimization of selection programmes considering environmental sensitivity described by a reaction norm has been considered by Kirkpatrick & Bataillon (1999), de Jong & Bijma (2002), Kolmodin *et al.* (2003) and Kolmodin & Bijma (2004).

10. CONCLUSIONS

I believe the random regression and covariance function methods will completely replace the use of specific time point correlations (Hill 1998, p.34)

Random regression models provide a powerful and—with hindsight—obvious framework for quantitative genetic analyses of data that represent curves. In essence, they are straightforward extensions of the linear models invoked for traits that are points. Hence, standard quantitative genetic theory applies, requiring only minor extensions to accommodate the 'infinite-dimensional' nature of function valued traits.

Moreover, models that are linear in the regression coefficients are readily accommodated in the linear mixed model, which is a standard for quantitative genetic analyses, especially for data from livestock improvement programmes. Consequently, regular methods for multivariate estimation in the mixed model can be used, i.e. best linear unbiased prediction for the estimation of breeding values and REML or Bayesian procedures for the estimation of covariance functions, again only needing minor modifications.

Of particular interest is the truncated expansion of trajectories, based on the leading eigenfunctions of the corresponding covariance function. For highly correlated traits, the first few eigenfunctions are likely to

explain the bulk of variation and thus give a sufficiently accurate approximation of the trajectory. This can result in very parsimonious models. As shown, this representation can be accommodated in the mixed-model-based analyses, and is thus likely to see increasing use in the analysis of function valued traits in future.

This work was supported by grants DEB-9973221 and EF-0328594 from the National Science Foundation (NSF) and grant NER/A/S/2002/00857 from the Natural Environment Research Council (NERC; M.K.).

REFERENCES

- Albuquerque, L. G. & Meyer, K. 2001 Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. *J. Anim. Sci.* **79**, 2776–2789.
- Beder, J. H. & Gomulkiewicz, R. 1998 Computing the selection gradient and evolutionary response of an infinite-dimensional trait. *J. Math. Biol.* **36**, 299–319.
- Calus, M. P. L. & Veerkamp, R. F. 2003 Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model. *J. Dairy Sci.* **86**, 3756–3764.
- Chase, K., Carrier, D. R., Adler, F. R., Jarvik, T., Ostrander, E. A., Lorentzen, T. D. & Lark, K. G. 2002 Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton. *Proc. Natl Acad. Sci. USA* **99**, 9930–9935.
- de Boor, C. 2001 *A practical guide to splines*. Springer series in applied mathematics, vol. 27, 2nd edn. New York: Springer.
- de Jong, G. & Bijma, P. 2002 Selection and phenotypic plasticity in evolutionary biology and animal breeding. *Livest. Prod. Sci.* **78**, 195–214.
- DeGroot, B. J., Keown, J. F., Kachman, S. D. & Van Vleck, L. D. 2003 Estimates of genetic parameters and lactation curves with a cubic spline model for Holstein cows treated with bovine somatotropin. *J. Anim. Sci.* **81**(Suppl. 1), 37. (Abstr.)
- Emmerling, R., Mäntysaari, E.A., Lidauer, M. 2002 Reduced rank covariance functions for a multi-lactation test-day model. In *Proc. Seventh World Congress on Genetics Applied to Livestock Production*. Communication No. 17-03. (CD-ROM)
- Falconer, D. S. 1952 The problem of environment and selection. *Am. Nat.* **86**, 293–298.
- Foulley, J. L. & van Dyk, D. A. 2000 The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genet. Select. Evol.* **32**, 143–163.
- 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.
- Gengler, N., Tijani, A., Wiggans, G. R. & Misztal, I. 1999 Estimation of (co)variance function coefficients for test day yield with an expectation-maximization restricted maximum likelihood algorithm. *J. Dairy Sci.* **82**, 1849. (See <http://jds.fass.org/cgi/content/abstract/82/8/1849>.)
- Gengler, N., Tijani, A. & Wiggans, G. R. 2000 Use of sequential estimation of regressions and effects on regressions to solve large multi-trait test day models. *J. Dairy Sci.* **83**, 369. (See <http://jds.fass.org/cgi/content/abstract/83/2/369>.)
- 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.

- Gomulkiewicz, R. & Beder, J. H. 1996 The selection gradient of an infinite-dimensional trait. *SIAM J. Appl. Math.* **56**, 509–523.
- Gomulkiewicz, R. & Kirkpatrick, M. 1992 Quantitative genetics and the evolution of reaction norms. *Evolution* **46**, 390–411.
- 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.
- Groeneveld, E. 1994 A reparameterisation to improve numerical optimisation in multivariate REML (co)variance component estimation. *Genet. Select. Evol.* **26**, 537–545.
- Harville, D. A. 1977 Maximum likelihood approaches to variance component estimation and related problems. *J. Am. Stat. Assoc.* **72**, 320–338.
- Harville, D. A. 1997 *Matrix algebra from a statistician's perspective*. Berlin: Springer.
- Henderson, C. R. 1973 Sire evaluation and genetic trends. In *Proc. Animal Breeding Genetics Symposium in Honor of Dr. J. L. Lush*, pp. 10–41. Champaign, Illinois: ASAS and ADSA.
- Henderson, C. R. 1976 A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**, 69–83.
- Henderson Jr, C. R. 1982 Analysis of covariance in the mixed model: higher-level, non-homogeneous and random regressions. *Biometrics* **38**, 623–640.
- Hill, W. G. 1998 Inferences from evolutionary biology to livestock breeding. In *Proc. Sixth World Congress on Genetics Applied to Livestock Production* **23**, 32–39. Australia: Armidale.
- Hill, W. G. 2002 Direct effects of selection on phenotypic variability of quantitative traits. In *Seventh World Congress on Genetics Applied to Livestock Production*. Communication no. 19-02. (CD-ROM)
- Hill, W. G. & Zhang, X.-S. 2004 Effects of phenotypic variability of directional selection arising through genetic differences in residual variability. *Genet. Res.* **83**, 121–132.
- Hofer, A. 1998 Variance component estimation in animal breeding: a review. *J. Anim. Breed. Genet.* **115**, 247–265.
- Huisman, A. E., Veerkamp, R. F. & van Arendonk, J. A. M. 2002 Genetic parameters for various random regression models to describe the weight data of pigs. *J. Anim. Sci.* **80**, 575–582.
- Jaffrézic, F. & Pletcher, S. D. 2000 Statistical models for estimating the genetic basis of repeated measures and other function-valued traits. *Genetics* **156**, 913–922.
- James, G. M., Hastie, T. J. & Sugar, C. A. 2000 Principal component models for sparse functional data. *Biometrika* **87**, 587–602.
- Jamrozik, J. 2004 Implementation issues for Markov Chain Monte Carlo methods in random regression test-day models. *J. Anim. Breed. Genet.* **121**, 1–13.
- Jamrozik, J. & Schaeffer, L. R. 1997 Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *J. Dairy Sci.* **80**, 762–770.
- Jamrozik, J. & Schaeffer, L. R. 2000 Comparison of two computing algorithms for solving mixed model equations for multiple trait random regression test day models. *Livest. Prod. Sci.* **67**, 143–153.
- Jamrozik, J., Schaeffer, L. R. & Dekkers, J. C. M. 1997 Genetic evaluation of dairy cattle using test day yields and random regression model. *J. Dairy Sci.* **80**, 1217–1226.
- Jamrozik, J., Schaeffer, L. R. & Jansen, G. B. 2000 Approximate accuracies of prediction from random regression models. *Livest. Prod. Sci.* **66**, 85–92.
- Jennrich, R. I. & Sampson, P. F. 1976 Newton–Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics* **18**, 11–17.
- Jensen, J., Mäntysaari, E. A., Madsen, P. & Thompson, R. 1997 Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *J. Ind. Soc. Agric. Stat.* **49**, 215–225.
- Jolliffe, I. T. 1986 *Principal component analysis*. Springer series in statistics. New York: Springer.
- Kettunen, A., Mäntysaari, E. A. & Pösö, J. 2000 Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models. *Livest. Prod. Sci.* **66**, 251–261.
- Kingsolver, J. G. & Gomulkiewicz, R. 2003 Environmental variation and selection on performance curves. *Integr. Comp. Biol.* **43**, 470–477.
- Kirkpatrick, M. 1993 The evolution of size in size-structured populations. In *The dynamics of size-structured populations* (ed. B. Ebenman & L. Persson), pp. 13–28. Heidelberg: Springer.
- Kirkpatrick, M. 2002 Selection on age dependent traits. In *Seventh World Congress on Genetics Applied to Livestock Production*. Communication no. 19-06. (CD-ROM)
- Kirkpatrick, M. & Bataillon, T. 1999 Artificial selection on phenotypically plastic traits. *Genet. Res.* **74**, 265–270.
- Kirkpatrick, M. & Heckman, N. 1989 A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol.* **27**, 429–450.
- Kirkpatrick, M. & Lofsvold, D. 1992 Measuring selection and constraint in the evolution of growth. *Evolution* **46**, 954–971.
- Kirkpatrick, M. & Meyer, K. 2004 Simplified analysis of complex phenotypes: direct estimation of genetic principal components. *Genetics* **168**, 2295–2306.
- Kirkpatrick, M., Lofsvold, D. & Bulmer, M. 1990 Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* **124**, 979–993.
- Kirkpatrick, M., Hill, W. G. & Thompson, R. 1994 Estimating the covariance structure of traits during growth and aging, illustrated with lactations in dairy cattle. *Genet. Res.* **64**, 57–69.
- Koivula, M., Negussie, E. & Mäntysaari, E. A. 2004 Genetic parameters for test-day somatic cell count at different lactation stages of Finnish dairy cattle. *Livest. Prod. Sci.* **90**, 145–157.
- Kolmodin, R. & Bijma, P. 2004 Response to mass selection when genotype by environment interaction is modelled as a linear reaction norm. *Genet. Select. Evol.* **36**, 435–454.
- Kolmodin, R., Strandberg, E., Madsen, P., Jensen, J. & Jorjani, H. 2002 Genotype by environment interactions in Nordic dairy cattle studied using reaction norms. *Acta Agric. Scand. A* **52**, 11–24.
- Kolmodin, R., Strandberg, E., Jorjani, H. & Danell, B. 2003 Selection in the presence of a genotype by environment interaction: response in environmental sensitivity. *Anim. Sci.* **76**, 375–385.
- Lande, R. & Arnold, S. J. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226.
- Legarra, A., Misztal, I. & Bertrand, J. K. 2004 Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle. *J. Anim. Sci.* **82**, 1564–1571.
- Lidauer, M., Strandén, I., Mäntysaari, E. A., Pösö, J. & Kettunen, A. 1999 Solving large test-day models by iteration on data and preconditioned conjugate gradient. *J. Dairy Sci.* **82**, 2788–2796.
- Lopez-Romero, P., Rekaya, R. & Carabano, M. J. 2004 Bayesian comparison of test-day models under different assumptions of heterogeneity for the residual variance: the change point technique versus arbitrary intervals. *J. Anim. Breed. Genet.* **121**, 14–25.

- Mäntysaari, E. A. 1999 Derivation of multiple trait reduced random regression (RR) model for the first lactation test day records of milk, protein and fat. *50th Annual Meeting of the European Association for Animal Production*, Mimeo.
- Meyer, K. 1998a Estimating covariance functions for longitudinal data using a random regression model. *Genet. Select. Evol.* **30**, 221–240.
- Meyer, K. 1998b Modeling 'repeated' records: covariance functions and random regression models to analyse animal breeding data. In *Proc. Sixth World Congress on Genetics Applied to Livestock Production* **25**, 517–520. Australia: Armidale.
- Meyer, K. 2001a Estimates of direct and maternal covariance functions for growth of Australian beef calves from birth to weaning. *Genet. Select. Evol.* **33**, 487–514.
- Meyer, K. 2001b Estimating genetic covariance functions assuming a parametric correlation structure for environmental effects. *Genet. Select. Evol.* **33**, 557–585.
- Meyer, K. 2005 Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet. Select. Evol.* **37**. (In the press.)
- Meyer, K. & Hill, W. G. 1997 Estimation of genetic and phenotypic covariance functions for longitudinal data by restricted maximum likelihood. *Livest. Prod. Sci.* **47**, 185–200.
- Meyer, K. & Kirkpatrick, M. 2005 Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. *Genet. Select. Evol.* **37**, 1–30.
- Meyer, K. & Smith, S. P. 1996 Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genet. Select. Evol.* **28**, 23–49.
- Misztal, I., Strabel, T., Jamrozik, J., Mäntysaari, E. A. & Meuwissen, T. H. E. 2000 Strategies for estimating the parameters needed for different test-day models. *J. Dairy Sci.* **83**, 1125–1134.
- Nobre, P. C. R., Misztal, I., Tsuruta, S., Bertrand, J. K., Silva, L. O. C. & Lopes, P. S. 2003 Analyses of growth curves of Nelore cattle by multiple-trait and random regression models. *J. Anim. Sci.* **81**, 918–926.
- Núñez-Antón, V. & Zimmerman, D. L. 2000 Modelling non-stationary longitudinal data. *Biometrics* **56**, 699–705.
- Pinheiro, J. C. & Bates, D. M. 1996 Unconstrained parameterizations for variance-covariance matrices. *Stat. Comput.* **6**, 289–296.
- Pletcher, S. D. & Geyer, C. J. 1999 The genetic analysis of age-dependent traits: modeling the character process. *Genetics* **153**, 825–835.
- Quaas, R. L. 1976 Computing the diagonal elements of a large numerator relationship matrix. *Biometrics* **32**, 949–953.
- Ramsay, J. O. 2000 Functional components of variation in handwriting. *J. Am. Stat. Assoc.* **95**, 9–15.
- Ramsay, J. O. & Silverman, B. W. 1997 *Functional data analysis*. New York: Springer.
- Ramsay, J. O. & Silverman, B. W. 2001 Functional data analysis. In *International encyclopedia of the social & behavioural sciences* (ed. J. Smelsen & P. B. Baltes), pp. 5822–5828. New York: Pergamon Press.
- Ramsay, J. O. & Silverman, B. W. 2002 *Applied functional data analysis*. Berlin: Springer.
- Rekaya, R., Carabaño, M. J. & Toro, M. A. 1999 Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. *Livest. Prod. Sci.* **57**, 203–217.
- Rice, J. A. & Silverman, B. W. 1991 Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. B* **53**, 233–243.
- Rice, J. A. & Wu, C. O. 2001 Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.
- Ruppert, D., Wand, M. P. & Carroll, R. J. 2003 *Semiparametric regression*. New York: Cambridge University Press.
- Schaeffer, L. R. 2004 Application of random regression models in animal breeding. *Livest. Prod. Sci.* **86**, 35–45.
- Schaeffer, L. R. & Dekkers, J. C. M. 1994 Random regressions in animal models for test-day production in dairy cattle. In *Proc. Fifth World Congress on Genetics Applied to Livestock Production* **18**, 443–446.
- Schnyder, U., Hofer, A., Labroue, F. & Künzi, N. 2001 Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and large white growing pigs. *Genet. Select. Evol.* **33**, 635–658.
- Schnyder, U., Hofer, A., Labroue, F. & Künzi, N. 2002 Multiple trait model combining random regressions for daily feed intake with single measured performance traits of growing pigs. *Genet. Select. Evol.* **34**, 61–81.
- Smith, S. P. 1995 Differentiation of the Cholesky algorithm. *J. Comput. Graph. Stat.* **4**, 134–147.
- Smith, A. B., Cullis, B. R. & Thompson, R. 2001 Analysing variety by environment data using multiplicative mixed models and adjustments for spatial field trends. *Biometrics* **57**, 1138–1147.
- Sorensen, D. & Gianola, D. 2002 *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Berlin: Springer.
- Spitzner, D. J., Marron, J. S. & Essick, G. K. 2003 Mixed-model functional ANOVA for studying human tactile perception. *J. Am. Stat. Assoc.* **98**, 263–272.
- Thompson, R. 1976 The estimation of maternal genetic variance. *Biometrics* **32**, 903–917.
- Thompson, R. & Mäntysaari, E. A. 1999 Prospects for statistical methods in dairy cattle breeding. *Interbull Bull.* **20**, 70–77.
- Tier, B. & Meyer, K. 2004 Approximating prediction error covariances in multiple-trait and random regression models. *J. Anim. Breed. Genet.* **121**, 77–89.
- Togashi, K. & Lin, C. Y. 2003 Modifying the lactation curve to improve lactation milk and persistency. *J. Dairy Sci.* **86**, 1487–1493.
- Torres, R. A. A. & Quaas, R. L. 2001 Determination of covariance functions for lactation traits on dairy cattle using random-coefficient regressions on B-splines. *J. Anim. Sci.* **79**(Suppl. 1), 112. Abstract.
- van der Werf, J. H. J. 2002 Optimising selection for traits along trajectories. In *Seventh World Congress on Genetics Applied to Livestock Production*. Communication no. 16-07. (CD-ROM)
- van der Werf, J. H. J., Goddard, M. & Meyer, K. 1998 The use of covariance functions and random regression for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* **81**, 3300–3308.
- Veerkamp, R. F. & Goddard, M. E. 1998 Covariance functions across herd production levels for test day records on milk, fat and protein yields. *J. Dairy Sci.* **81**, 1690–1701.
- Veerkamp, R. F. & Thompson, R. 1999 A covariance function for feed intake, live weight and milk yield estimated using a random regression model. *J. Dairy Sci.* **82**, 1565–1573.
- Via, S. & Lande, R. 1985 Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution* **39**, 505–522.
- White, I. M. S., Thompson, R. & Brotherstone, S. 1999 Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* **82**, 632–638.