Genet. Sel. Evol. 37 (2005) 473–500

© INRA, EDP Sciences, 2005
DOI: 10.1051/gse:2005012

Original article

Random regression analyses using B-splines to model growth of Australian Angus cattle

Karin MEYER*

Animal Genetics and Breeding Unit[1], University of New England,
Armidale NSW 2351, Australia

(Received 4 October 2004; accepted 2 May 2005)

Abstract – Regression on the basis function of B-splines has been advocated as an alternative to orthogonal polynomials in random regression analyses. Basic theory of splines in mixed model analyses is reviewed, and estimates from analyses of weights of Australian Angus cattle from birth to 820 days of age are presented. Data comprised 84 533 records on 20 731 animals in 43 herds, with a high proportion of animals with 4 or more weights recorded. Changes in weights with age were modelled through B-splines of age at recording. A total of thirteen analyses, considering different combinations of linear, quadratic and cubic B-splines and up to six knots, were carried out. Results showed good agreement for all ages with many records, but fluctuated where data were sparse. On the whole, analyses using B-splines appeared more robust against "end-of-range" problems and yielded more consistent and accurate estimates of the first eigenfunctions than previous, polynomial analyses. A model fitting quadratic B-splines, with knots at 0, 200, 400, 600 and 821 days and a total of 91 covariance components, appeared to be a good compromise between detailedness of the model, number of parameters to be estimated, plausibility of results, and fit, measured as residual mean square error.

covariance function / growth / beef cattle / random regression / B-splines

1. INTRODUCTION

Random regression (RR) analyses have become a standard procedure for the genetic analysis of "repeated records" on individuals which are recorded along a continuous scale, such as time. A large proportion of applications considered test day production of dairy cows, but RR analyses of growth records or data on feed intake of meat producing animals are finding increasing use. A recent review of applications has been given by Schaeffer [35]. The underlying idea in RR analyses is that subject-specific curves can be described as the weighted

* Corresponding author: kmeyer@didgeridoo.une.edu.au
[1] AGBU is a joint venture between the NSW Department of Primary Industries and the University of New England.

Article published by EDP Sciences and available at http://www.edpsciences.org/gse or http://dx.doi.org/10.1051/gse:2005012

sum of a set of functions of the continuous covariable, the so-called basis functions. The majority of RR analyses so far fitted polynomial of time or age at recording as basis functions. In particular, following Kirkpatrick *et al.* [22], Legendre polynomials have been widely used.

Higher order polynomials are flexible and have been found to be capable of modelling changes in means and variances along a continuous scale well. However, such polynomials often put a high emphasis on observations at the extremes, and are notoriously problematic 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, and, moreover, that errors were predominantly due to oscillations at the extremes of the curve (*e.g.* [2], Chapter 2). Not surprisingly, RR analyses fitting cubic, quartic or even higher order polynomials thus have frequently encountered "end-of-range" problems, resulting in erratic and implausible estimates of variance components and genetic parameters. This has been especially prevalent at the highest ages with least records, for data sets with many more records at earlier than later ages, and for analyses where substantial proportions of animals had fewer records than the order of polynomials fitted.

An alternative to high degree polynomials are "piece-wise polynomials", *i.e.*, curves constructed from pieces of lower degree polynomials, joined at selected points, so-called knots. Such curves are commonly referred to as spline curves, and are widely used, in particular in non-parametric analyses involving smoothing of curves (*e.g.* [13,34]). An application of splines for the analysis of test day records of dairy cows has been presented by White *et al.* [41]. A particular type of spline curve is the so-called B-spline [2]. B-splines yield the same fit as splines based on truncated power functions, but have better numerical properties [9,34]; pertinent details are reviewed below. Shi *et al.* [36] found changes in CD4 counts of children over time to be inadequately modelled by regressing on quadratic polynomials of time, and proposed B-splines as a more flexible alternative. Recently, Rice and Wu [33] described the use of B-splines to model random effects curves in mixed model analyses, and demonstrated their efficacy in estimating covariance functions. Similarly, James *et al.* [17] employed B-spline functions to model growth curves for longitudinal data with sparse, irregular measurements. All these applications of RR analyses using B-splines were at the phenotypic level, but use of B-splines can readily be extended to genetic RR analyses. Torres and Quaas [39] performed a RR analysis using B-splines to model test day records of dairy cows, but there have been no applications to growth data of farm animals.

A number of RR analyses of weight data for beef cattle have been reported. Previous, large scale analyses of records from birth to approximately two years of age considered a model fitting cubic regressions on age at recording for all random effects, *i.e.*, direct and maternal, genetic and permanent environmental effects [26, 32]. This choice was based on a compromise between computational demands and detailedness of the model, the fact that a cubic polynomial is the lowest degree polynomial for which both first and second derivatives, *i.e.* growth rate and acceleration could vary with age, and limits in the numbers of records available per animal. Recent analyses of a set of field data from Australian Angus cattle, however, found this model to represent a problematic choice for such data: cubic polynomials were most likely to yield erratically high estimates of variances at the highest ages. In addition, estimates of the first genetic eigenfunctions differed markedly from those fitting a quadratic or quartic polynomial to model changes in animals' additive genetic effects with age [28].

After a brief review of spline functions, this paper presents a set of analyses using RR on basis functions of B-spline to model growth trajectories of beef cattle, from birth to just over two years of age. Data are field records, analysed previously fitting regressions on Legendre polynomials of age at recording, and results from the two sets of analyses are contrasted.

2. REVIEW OF SPLINE BASICS

Splines are generally defined as curves which consist of individual segments which are joined smoothly; the segments are given by polynomials and the points at which they join are referred to as knots. Often the polynomials are cubic and first or second derivatives are constrained to be smooth [5].

The name spline originates from the long thin, flexible strip of wood traditionally used in drawing curves. When fixed at given points ("knots"), the wood takes the shape which minimises the energy required for bending, thus yielding the smoothest shape possible. Similarly, bivariate splines are commonly called "thin plate" splines, referring to a thin, bendable sheet of metal used to approximate the surface. There is a variety of different spline functions, and a large body of literature. Numerous textbooks covering the subject are available (*e.g.* [2, 4, 13, 34]).

In the simplest case, a spline function consists of linear segments. This is commonly described as a "broken-stick" curve, and is a straightforward extension of the parametric, linear regression. Let y_i denote some observations recorded at times t_i, with $i = 1, \ldots, n$, and assume we have knots T_k. In fitting

a linear spline, we then assume a non-parametric regression model

$$y_i = \beta_0 + \beta_1 t_i + \sum_k \beta_{1k} (t_i - T_k)_+ + e_i \tag{1}$$

for y_i, with β_0 and β_{1j} denoting the intercept and linear regression coefficients, respectively, e_i the residual error pertaining to y_i, and $(x)_+ = \max(0, x)$ equal to x if x is positive and equal to 0 otherwise [34]. This gives a slope of β_1 for the first segment, *i.e.*, for $t_i \leq T_1$, a slope of $\beta_1 + \beta_{11}$ for the second segment with $T_1 \leq t_i \leq T_2$, and a slope of $\beta_1 + \sum_{k=1}^{m} \beta_{1k}$ for the segment between bordered by T_m and T_{m+1}. For splines consisting of polynomial segments of degree p, Eq. (1) is expanded to

$$y_i = \beta_0 + \beta_1 t_i + \ldots + \beta_p t_i^p + \sum_k \beta_{pk} (t_i - T_k)_+^p + e_i, \tag{2}$$

where β_p are the $p-$th degree regression coefficient. $(x)_+^p$ are known as truncated power functions, *i.e.*, Eq. (2) gives a spline with truncated power basis of degree p [34].

These are the "grafted" polynomials considered by Fuller [11]. For $p = 2$, El Faro *et al.* [10] used these to model lactation curves of dairy cows, and Meyer [25] applied such splines in modelling seasonal changes in variances of mature cow weights in a RR analysis.

2.1. Penalised and smoothing splines

Fitting spline functions as given above can yield "wiggly" estimates of curves, especially if p is low and there are many knots. Moreover, the choice of knots can have a substantial influence on estimates. These problems can be alleviated by imposing a penalty, which reduces the influence of the regression coefficients β_{pk}. This yields a smoother curve, and the penalty is thus referred to as "roughness" penalty [13]. Ruppert *et al.* [34] suggested to constrain the sum of squared regression coefficients β_{pk} in Eq. (2), imposing a penalty of $\lambda \sum_k \beta_{pk}^2$, added to the criterion to be minimised in least-squares or maximum likelihood estimation, to smooth the resulting curve. The smoothing parameter, λ, governs the degree of smoothing, small values yielding curves close to the unpenalised fit, and large values resulting in estimates more similar to the corresponding parametric regression. With smoothing, choice of knots is less crucial than for $\lambda = 0$.

Penalised splines are readily accommodated in a mixed model framework. In essence, imposing a roughness penalty yields a system of equations where

the penalised regression coefficients are estimated as random effects, *i.e.*, shrinking them towards their mean, whilst the unpenalised coefficients are treated as fixed effects. The shrinkage is determined by λ, which is equivalent to the ratio of error variance and variance due to the regression coefficients. This implies that a suitable value of λ can be estimated from the data in a REML analysis. A widely used spline is the cubic smoothing spline, considered in depth by Green and Silverman [13] and Verbyla *et al.* [40], which generally fits one knot at each observation. This is the spline function incorporated in the AsReml software package [12], and most RR analyses in quantitative genetic analysis using splines have fitted this type, though often with a reduced set of knots (*e.g.* [3, 15, 41]).

White *et al.* [41] reviewed the results of Verbyla *et al.* [40], extending them to the case where knots are chosen independently of the data locations, and presented an application to the analysis of test-day records on dairy cows. They showed that a mixed model analysis fitting splines for each animal is equivalent to a RR analysis fitting unity, time at recording, and segment-specific functions of time as covariables. However, in contrast to "standard" RR analyses, this analysis considers the matrix of covariances among RR coefficients to be highly structured. Whilst the first two RR coefficients (β_0 and β_1) are assumed correlated and have different variance, all other spline terms are assumed to have the same variance and to be independently distributed. This results in a total of only four covariance components among RR coefficients to be estimated. White *et al.* [41] emphasized that this fitted separate splines with the same degree of smoothing for each animal, but questioned whether the model might be improved by allowing for non-zero covariances between the first two and the segment-specific coefficients.

2.2. B-splines

Alternatives to the truncated power base considered above exist. A common choice are B-spline functions, which yield equivalent fits, but have better numerical properties [34]. Here, the "B" stands for basis [2]. B-splines comprise a set of overlapping, smooth and non-negative functions, which are unimodal and sum to unity for all values of t. Like splines with a truncated power basis, they can have different degree p.

B-spline functions can be defined recursively. Basis functions of degree $p = 0$ have values of unity for all points in a given interval, and zero otherwise.

For the $k-$th interval given by knots T_k and T_{k+1} with $T_k \leq T_{k+1}$,

$$B_{k,0}(t) = \begin{cases} 1 & \text{if} \quad T_k \leq t \leq T_{k+1} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Higher degree basis functions, $B_{k,p}$ for $p > 0$, are then determined from the values of the lower degree basis functions, already evaluated, and the width of the adjoining intervals between knots. The general relationship is [2]:

$$B_{k,p}(t) = \frac{t - T_k}{T_{k+p} - T_k} B_{k,p-1}(t) + \frac{T_{k+p+1} - t}{T_{k+p+1} - T_{k+1}} B_{k+1,p-1}(t). \tag{4}$$

For each p, there are a limited number of non-zero basis functions of lower order, which can be exploited when evaluating Eq. (4). Efficient strategies are described in the relevant literature (*e.g.* [2]). Alternatively, for equally spaced knots, B-spline functions can be obtained as the difference between splines with a basis of truncated power functions; see Eilers and Marx [9] for details.

For a given range of t, T_0 to T_m, division into m intervals requires specification of $m - 1$ internal knots. Together with the two external knots (T_0 and T_m), this yields $m+1$ knot points and $m+p$ non-zero functions $B_{k,p}$ to be considered. Evaluation of Eq. (4), however, requires p additional knots to be specified at each side of the interval, so that, for calculation, we have $2p + m + 1$ knots in total. For "uniform" B-splines with knots placed at equal intervals Δ, Eilers and Marx [9] suggest to place these additional knots at equal intervals outside the range of t, *i.e.* at $T_0 - p\Delta, \ldots, T_0 - \Delta$ and $T_m + \Delta, \ldots, T_m + p\Delta$. This would result in all $B_{k,p}$ for a given p to have the same shape, *i.e.*, simply be horizontally shifted copies of each other. Alternatively, we could define the two external knots to have multiplicity $p + 1$, *i.e.*, set the p additional knots either side to the value of the corresponding external knot. This is common practice for implementation of B-splines in statistical packages, for example in R [16]. For $p \geq 2$, this would yield $B_{k,p}$ of different shapes for those involving only internal knots and those spanning external and additional knots, even if all internal knots are spaced at equal intervals.

Like splines with a truncated power basis, B-splines can be penalised to obtain smoother curves. The penalty for the former, described above, resulted in a formulation analogous to a ridge regression. Corresponding penalties for a B-spline basis, suggested by Eilers and Marx [8], are difference based, where the degree of fit and order of difference for the penalty can be chosen independently. Eilers and Marx [8] refer to the resulting functions as "P-splines", whilst other authors use the term P-splines for any penalised splines, regardless of base. Again, penalised B-splines are readily estimated within the mixed

model framework [7]. The covariance matrix among random spline coefficient is again assumed to be highly structured, with the structure determined by the difference matrix used in defining penalties. A detailed review of the properties of B-splines and P-splines is given by Eilers and Marx [9].

2.3. Splines in random regression analyses

Penalised splines work best with a certain degree of "overfitting", *i.e.*, a relatively large number of knots compared to the number of distinct values of the continuous covariable t or number of observations [34]. In that case, fitting random, subject-specific spline curves (*e.g.* [6, 14, 40]) may yield a reasonable description of the covariance structure along the trajectory, in spite of the rigid structure imposed on the covariance matrix among RR coefficients. However, computational requirements for such models involving many coefficients to be estimated may be large, in particular in genetic evaluation.

Hence, random regression analyses of animal breeding data have mainly been carried out fitting polynomial basis functions with a narrow basis, *i.e.*, few regression coefficients. A logical alternative to penalised splines as described above, then is to fit a random regression on a limited number of B-spline basis functions, assuming the covariance matrix among RR coefficients is unstructured, as outlined by Rice and Wu [33]. The same approach has been taken by Shi *et al.* [36] and James *et al.* [17], who combined it with reduced rank estimation to avoid overparameterisation.

This yields the standard RR model commonly applied in quantitative genetic analyses of longitudinal or "repeated" records. Whilst such models are likely to entail more coefficients than a corresponding RR analysis with polynomial basis, the matrices in the mixed model are sparser. For a B-spline of degree p, each row in the design matrix has at most $p + 1$ non-zero coefficients, *i.e.*, computational requirements are less than for corresponding analyses with the same number of coefficients for a polynomial basis.

Few guidelines on the choice of knots are available. For "grafted polynomials", the obvious recommendation was the place knots where the curve to be modelled was expected to change [11]. Most applications of B-splines to model flexible random curves [17, 33, 36] so far used cubic splines with equally spaced knots, and cross-validation or information criteria to choose between models involving different numbers of knots. Torres [38] chose a quadratic spline for 10 equally spaced intervals to model lactation records of dairy cows, resulting in 12 basis function and RR coefficients. Eilers and Marx [9] emphasized that with penalised B-splines, equal placement of knots is all that is

required, and demonstrated good interpolation for areas with few observations when the distribution of data over the range of t was highly uneven. However, they considered a scenario with relatively many knots. Spline curves with fewer knots tend to be smoother, but give a less good fit to local details than those with more knot points. Concerns about the potential impact of the number and placement of knots for the model suggested by Rice and Wu [33] have been voiced [6]. Considering penalised splines with truncated power basis, Ruppert *et al.* [34] strongly advocated placement of knots at equal quantiles of t.

Covariance matrices among RR coefficients can be estimated using standard REML methodology or Bayesian analyses; see Meyer and Kirkpatrick [31] for a recent review. For a B-spline basis, the covariance functions defined by the matrices among RR coefficients are tensor products of B-splines [33], *i.e.*, bivariate B-splines (if $f(x)$ is a function of x and $g(y)$ is a function of y, the bivariate function $p(x, y) = f(x)g(y)$ is called a tensor product).

In addition to covariance functions *per se*, we are interested in their eigenfunctions and corresponding eigenvalues [31]. As emphasised by Kirkpatrick and Heckman [20], these can be estimated directly from the estimated matrix of coefficients of the covariance function, *i.e.*, the matrix of covariances among RR coefficient, if we fit orthogonal basis functions, such as Legendre polynomials, as covariables. As B-splines functions are not orthogonal, an eigendecomposition of the covariance matrix of RR coefficient does not yield the eigenfunctions and -values of the covariance function. Hence, Rice and Wu [33] suggested to extract the required quantities numerically, by evaluating the covariance function on a fine grid over the range of t, and calculating the eigenvalues and -vectors of the resulting covariance matrix.

2.4. Illustration

Figure 1 shows the values of covariables for a continuous scale divided into four equal intervals. The first column represents the basis functions for $p = 0$. Clearly, fitting such covariables would be equivalent to a multivariate analysis, considering observations in each interval to be a different trait. Regressing on B-spline functions of degree p, an observation has a non-zero effect for at most $p + 1$ intervals, *i.e.*, has a local rather than global influence. As shown in the second column, a linear B-spline basis ($p = 1$) results in blending of information across two intervals, with values of the trajectory at the knots completely determined by the observations at these points. At a fixed effects level, linear B-spline have been suggested to blend information for various ages at

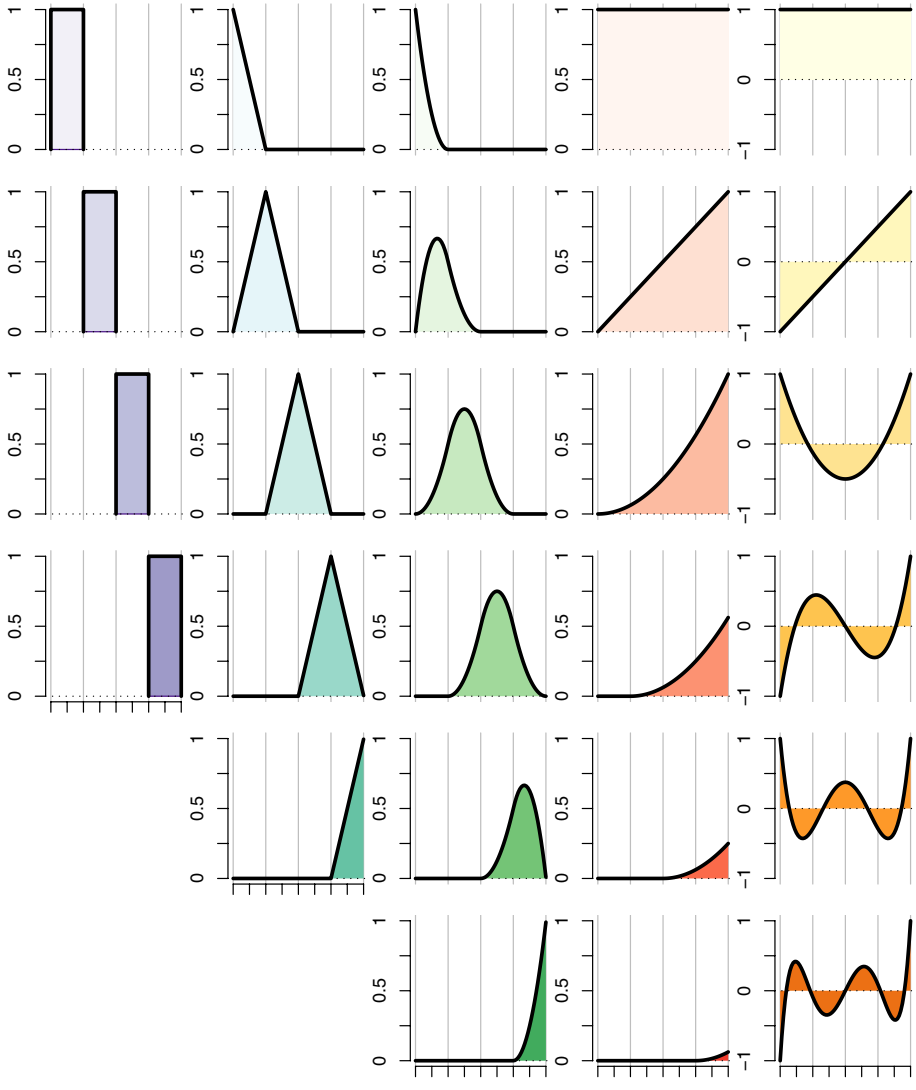

Figure 1. Values of covariables in random regression analyses using B-spline basis functions, dividing the longitudinal scale into four equal intervals (first column: degree 0, second column: degree 1, and third column: degree 2), together with corresponding values for a quadratic spline function with truncated power function base (fourth column), and Legendre polynomials (fifth column, polynomials of degree 0 to 5 shown).

recording within fixed contemporary groups, instead of the customary subdivision into distinct subclasses [1, 37].

Quadratic B-spline functions, given in the third column, have been calculated repeating the external knots as described above, resulting in three identical knots at each side. This gives weights of unity for the external knot points. If equal-distant additional knots outside the range had been used, the end points of the curve would have been the combination of two estimated splines curves, with equal weights of 0.5. Similarly, the curve at the mid-point of t, is the straight average of the estimates of the third and fourth spline curve. For the point in the middle of the second interval (fourth tick-mark), weights for the spline functions two, three and four are 0.125, 0.750 and 0.125, respectively. For a cubic B-spline (not shown), there would be a seventh basis function spanning all four intervals.

Corresponding functions for a quadratic spline with truncated power basis are given in the fourth column, for t standardised to the interval from 0 to 1. When fitting a regression on Legendre polynomials (fifth column; values shown are unnormalised Legendre polynomials), observations at all points (except for selected points where the polynomial has a value of zero) affect the estimate of each regression coefficient, *i.e.*, all observations have global influence. Moreover, weights at the extremes are consistently high.

3. MATERIALS AND METHODS

3.1. Data

Data comprised records for weights of Australian Angus cattle, measured from birth to 820 days of age. Data were extracted for herds with most records per animal available, deleting any records on animals with less than three observations and records in single record subclasses; see Meyer [28] for details. This yielded 84 533 records on 20 731 animals, which were progeny of 1348 sires and 8167 dams.

Figure 2 shows the distribution of records over ages at recording in 10-day intervals, together with corresponding mean weights. Not shown in Figure 2 are 17 891 birth weight records. Genetic evaluation of beef cattle in Australia considers weights at target ages of 200, 400 and 600 days with permissible age ranges of 80 to 300 days, 301 to 500 and 501 to 700 days, respectively. The distribution of records over ages of recording reflects this régime, with very few weights recorded between birth and 80 days of age and limited numbers of weights available after 700 days. On average there were 4.08 records per

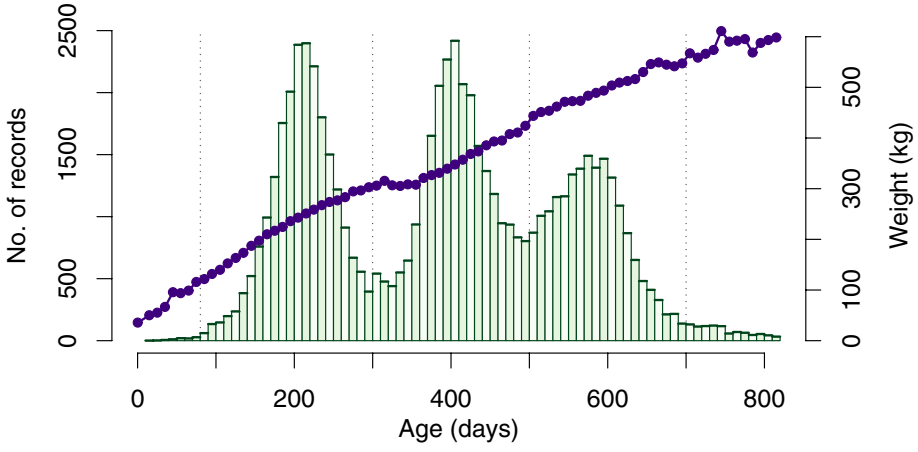

Figure 2. Numbers of records[a] (□) and mean weights (●) for ages at recording in 10-day intervals.
[a] omitting 17 891 birth weight records.

animal, and the maximum number of weights recorded per individual was 9. There were 6027, 8963, 4231, 1189 and 321 animals with 3, 4, 5, 6 and 7–9 records, respectively. This resulted in 70.9%, 27.7% and 7.3% of animals with at least 4, 5 and 6 weights recorded. At a genetic level, considering information from parents, progeny and sibs, corresponding proportions were 98.3%, 85.2% and 48.1%.

Pedigree information for animals in the data and their parents was extracted, considering up to five generations backwards. Prior to analyses, pedigrees were "pruned", *i.e.* any parents without records and a link to a single offspring only were treated as unknown. This was done recursively, resulting in 14 326 parents without records to be taken into account, *i.e.*, direct genetic effects for a total of 35 057 animals to be included in the analysis. Pedigrees for maternal genetic effects were "pruned" separately, eliminating any animals without an expression of their maternal genetic effects and linked to only one other animal. This reduced the number of animals in the maternal pedigree and relationship matrix to 20 463. The data comprised a proportion of animals which were the result of embryo transfer, and thus had different dams at the genetic and permanent environmental level. Often the foster dams had only one calf, resulting in 9670 permanent environmental dams to be considered. Of these, 5202 had one offspring only, 1782, 1012, 680, 376 and 619 cows had 2, 3, 4, 5 and 6 or more calves with records, respectively.

Table I. Models of analysis fitted with residual mean square error (MSE), together with results for models with the same number of parameters fitting Legendre polynomials (from Meyer [28]).

Model	n[a]	Knots		MSE	Model[b]	MSE
		Direct	Maternal			
L4343	51	0, 300, 600, 821	0, 300, 601	49.1	M4343	36.4
L5454	69	0, 200, 400, 600, 821	0, 200, 400, 601	25.1		
L6565a	91	0, 160, 320, 480, 640, 821	0, 160, 320, 480, 601	23.9		
L6565b	91	0, 1, 200, 400, 600, 821	0, 1, 200, 400, 601	24.6		
Q3333	43	0, 821	0, 601	53.5	M3333	53.6
Q4444	59	0, 410, 821	0, 300, 601	34.3	M4444	35.4
Q5454	69	0, 300, 600, 821	0, 300, 601	29.4	M5454	28.7
Q6464	81	0, 80, 300, 600, 821	0, 300, 601	27.5		
Q6565a	91	0, 200, 400, 600, 821	0, 200, 400, 601	24.3		
Q6565b	91	0, 80, 300, 600, 821	0, 200, 400, 601	27.3		
Q7575	105	0, 80, 300, 500, 700, 821	0, 80, 300, 601	19.7		
Q7676	117	0, 160, 320, 480, 640, 821	0, 160, 320, 480, 601	15.9		
C7676	117	0, 200, 400, 600, 821	0, 200, 400, 601	16.3		

[a] Number of parameters to be estimated.
[b] Fitting Legendre polynomials.

3.2. Model

3.2.1. Random effects

The model of analysis fitted a RR on basis functions of B-splines of age at recording for four random effects. Linear (L), quadratic (Q) and cubic (C) polynomials for the individual segments, *i.e.* basis functions of degree $p = 1, 2$ and 3, were considered. The same degree of fit was selected for all random effects in the model. Knots were chosen approximately equally spaced, paying some attention to the current recording régime and the distribution of records available. Up to six knots per effect were specified, and the same knots were fitted for genetic and permanent environmental splines. Table I summarises the knot positions chosen for the 13 analyses performed. In addition to the m knots chosen to divide the range of ages into $m - 1$ intervals, the external knots (0 days and 821 or 601 days) were repeated to have multiplicity $p + 1$ as described above (Sect. 2), where p is the degree of polynomials. This resulted in m, $m + 1$ and $m + 2$ RR coefficients to model trajectories for linear, quadratic and cubic basis functions, respectively, and $m - 1$ intervals chosen.

In the following, analyses are described as "Xk_A k_M k_R k_C", where X = L, Q or C gives the degree of the polynomials segments, and k_A, k_M, k_R and k_C specify the number of regression coefficients fitted for direct (A) and maternal (M) additive genetic effects, and direct (R) and maternal (C) permanent environmental trajectories, respectively. For example, Q5454 denotes an analysis fitting a quadratic B-spline with three knots or four coefficients for both maternal effects, and four knots or five RR coefficients for direct genetic and permanent environmental effects. Analyses of the same degree and number of knots, but differing in the position of knots, are distinguished by suffixes 'a' and 'b' (see Tab. I).

Maternal effects (M and C) were restricted to affect records taken from 0 to 600 days only. Direct and maternal genetic effects were assumed to be uncorrelated. This is common practice in the analysis of Australian field data, as the data structure generally does not allow the direct-maternal genetic covariance to be estimated without bias; see Meyer [28] for details. Residuals were considered independently distributed with heterogeneous measurement error variances (σ_i^2). Changes in σ_i^2 with age were modelled as a step function with 19 classes (0, 1–30, …, 271–300, 301–360, …, 721–780 and 781–820 days) throughout. This gave a total of 43 (Q3333) to 117 (C7676) covariance components to be estimated.

3.2.2. Fixed effects

As in previous analyses [28], changes in mean with age were modelled through fixed, cubic regressions on Legendre polynomials of age, nested within sex, dam age class (in years, treating ages > 9 years as one class) and birth type (single *vs.* twin). This was done for comparability of results. More logically, mean trajectories could be modelled through splines as well when fitting such functions for the random effects in the model [33]. In addition, contemporary groups (CG) were fitted as cross-classified fixed effects (8822 levels). CG were defined as herd-sex-management group-year/month subclasses for birth weights, and herd-sex-management group-date of weighing subclasses for other weights. To reduce the range of ages compared directly, CG classes were further subdivided applying an "age slicing" of 45 days up to 300 days, and 60 days for higher ages. If this resulted in a small subclass with less than five records for the highest ages in the CG, this last subclass was merged with the previous age subclass, provided the range of ages did not exceed 54 days for weights up to 300 days, and 72 days otherwise.

This gave the formal model of analysis

$$y_{ijklmno} = cg_i + \sum_{r=0}^{3}\left(b^s_{rj} + b^{da}_{rk} + b^{bt}_{rl}\right)\phi_r(t_{ijklmno}) + \sum_{r=1}^{k_A}\alpha_{rm}B_r(t_{ijklmno}) + \sum_{r=1}^{k_R}\gamma_{rm}B_r(t_{ijklmno}) + \sum_{r=1}^{k_M}\tau_{rn}B_r(t_{ijklmno}) + \sum_{r=1}^{k_C}\omega_{ro}B_r(t_{ijklmno}) + \varepsilon_{ijklmno} \quad (5)$$

where $y_{ijklmno}$ is a record, taken at age $t_{ijklmno}$ in contemporary group i, for animal m which has sex j, is born in birth type class l, whose genetic dam is n, and who has permanent environmental dam o which belongs to age class k. cg_i is the fixed effect for the $i-$th contemporary group, and b^s_{rj}, b^{da}_{rk}, and b^{bt}_{rl} are the fixed regression coefficients for the $j-$th sex, $k-$th dam age, and $l-$th birth type class, respectively. Covariables $\phi_r(t_{ijklmno})$ are the $r-$th Legendre polynomial, evaluated for $t_{ijklmno}$. α_{rm} and γ_{rm} denote the $r-$th random regression coefficient for direct additive genetic and permanent environmental effects of animal m, and τ_{rn} and ω_{ro} are the corresponding coefficients for maternal genetic effects of dam n and maternal permanent environmental effects of dam o. Covariables $B_r(t_{ijklmno})$ are values of the $r-$th B-spline function, evaluated (see Eq. (4)) for $t_{ijklmno}$ and the chosen degree p and knots T_x in a particular analysis, summarised in Table I. Finally, $\varepsilon_{ijklmno}$ is the residual error pertaining to $y_{ijklmno}$.

3.3. Analyses

Estimates of (co)variance components were obtained by Bayesian analysis, employing a Gibbs sampling algorithm as implemented in program RRGIBBS [27], using its option to fit regressions on user-defined functions of the continuous scale (age) considered[1]. Covariables were the functions $B_{k,p}$ as given by Eq. (4), evaluated for the ages in the data. Calculations were carried out using a simple Fortran program, but could equally have been performed by standard statistical software, for instance the `bs` function in the `spline` library of the R package [16].

Three Markov chains with 130 000 samples each were drawn for each analysis, assuming flat bounded priors. Estimates for all statistics considered were obtained as posterior means, pooled across chains, disregarding the first

[1] An updated version of RRGIBBS which exploits sparseness in the covariables and a Fortran program to evaluate B-splines are available at http://agbu.une.edu.au/kmeyer/rrgibbs.html http://agbu.une.edu.au/~kmeyer/rrgibbs.html

30 000 samples in each chain as "burn-in" period. Pointwise confidence bands (CI) for estimates were estimated as approximate 95% highest posterior density regions.

Variances and genetic parameters for ages in 10 day intervals were calculated for every fourth sample. Eigenvalues and eigenfunctions of the estimated covariance functions were obtained numerically. This involved evaluating the covariance functions on a 10-day grid of ages, and calculating the eigenvalues and eigenvectors of the resulting covariance matrix. Scaling for grid size, the non-zero eigenvalues and corresponding eigenvectors of this matrix then provided estimates of eigenvalues and eigenfunctions. Preliminary investigations had shown little advantage in using a finer grid, but a marked increase in computational requirements.

A generalised least-squares analysis was carried out for each model, with the corresponding estimates of covariance components as variance parameters. This yielded "best linear unbiased predictors" (or estimators) for all effects in the model of analysis. Residuals ($\varepsilon_{ijklmno}$ in Eq. (5)) were then estimated by adjusting each record for the estimates of all effects assumed to affect it, and their mean squares (MSE) were calculated.

4. RESULTS

4.1. Variance components

Estimates of variance components for 10 analyses, together with corresponding estimates for two analyses fitting Legendre polynomials as basis functions (from [28]) are summarised in Figure 3 and Figure 4. The latter fitted quadratic polynomials for maternal effects, quartic polynomials for animals' permanent environmental effects, and quadratic (M3353) or quartic (M5353) polynomials for direct genetic effects. In Figure 4, results for model M5454, which fitted cubic rather than quadratic polynomials for maternal effects, are shown to provide estimates for two different orders of fit.

As for previous analyses regressing on Legendre polynomials of age at recording [28], estimates from all analyses agreed well for ages with large numbers of observations. Conversely, for ages with few records estimates were largely determined by the covariables chosen to model the trajectory. For weights from birth to about 120 days of age, analyses were, in essence, interpolating. Hence, if knots were placed so that the corresponding segment of the curve did not include sufficient records, as in analyses Q6565b and Q7575 (not shown), estimates were implausible and had large CIs (Fig. 3). Knots for

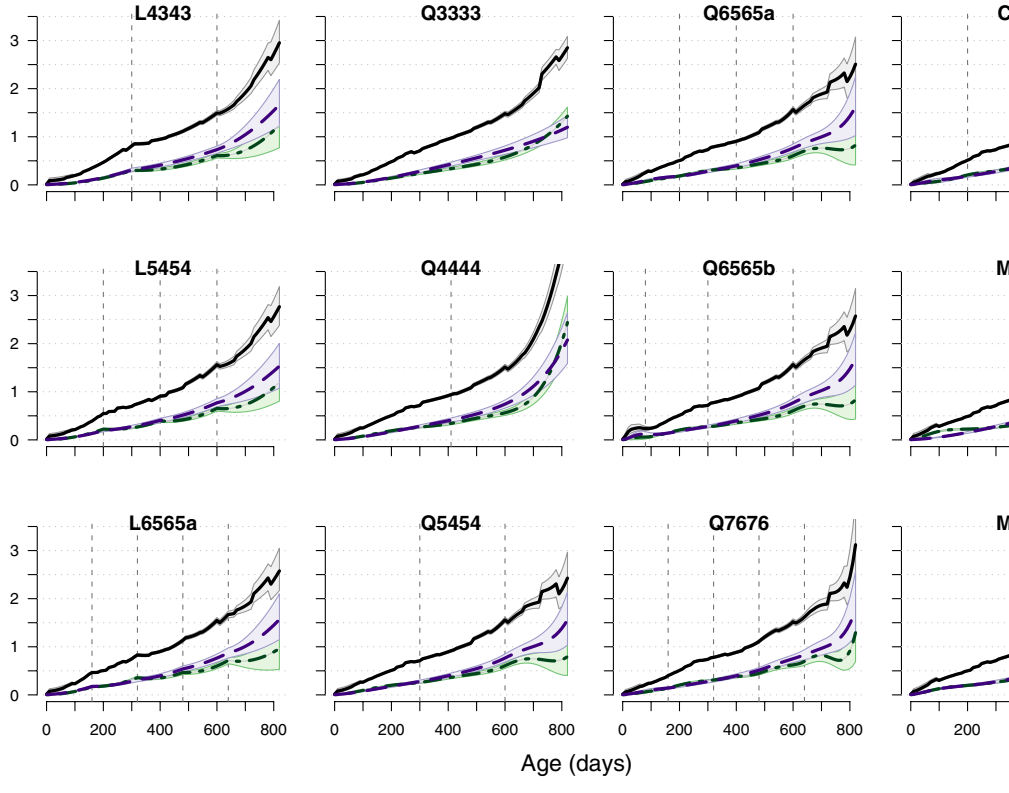

Figure 3. Estimates of phenotypic (solid line), direct genetic (dashed line) and direct permanent environment
(in 1000 kg^2) from selected analyses, together with their approximate 95% confidence regions (shaded areas)
the position of internal knots.

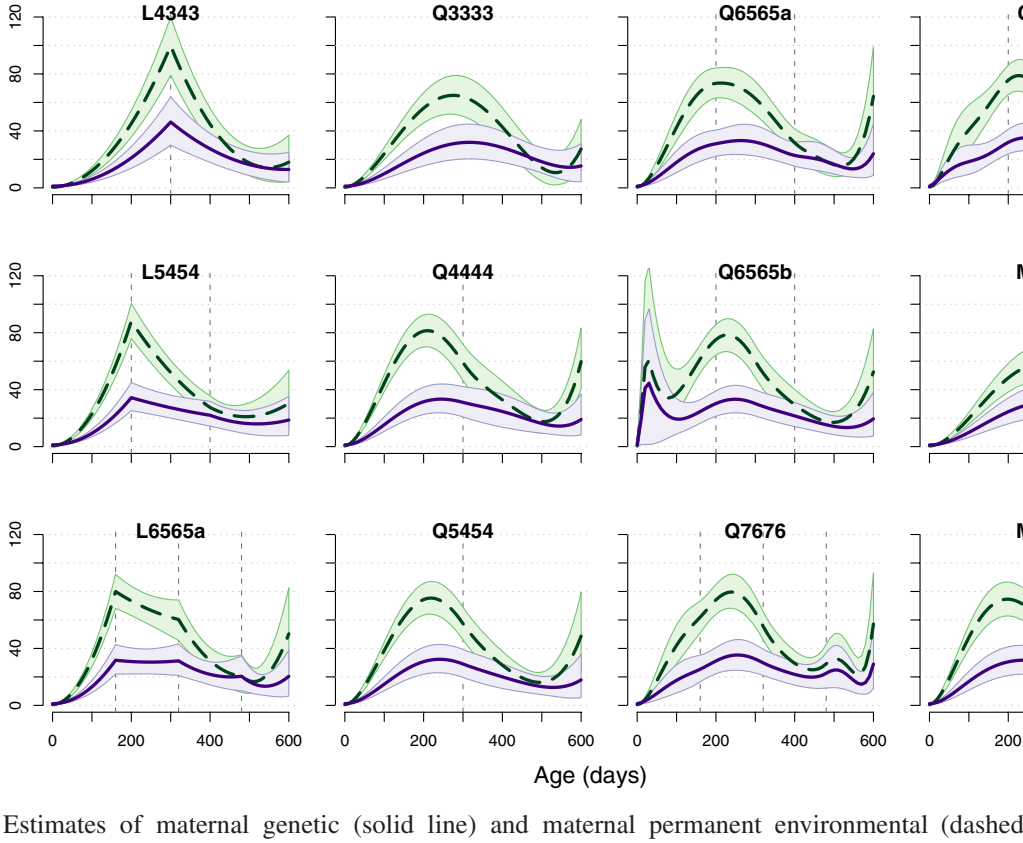

Figure 4. Estimates of maternal genetic (solid line) and maternal permanent environmental (dashed
together with approximate 95% confidence intervals (shaded areas), for selected analyses (vertical dash
of internal knots).

analyses Q7575 were chosen to coincide with the limits distinguishing between different traits in the current multi-trait genetic evaluation (see Sect. 3.1). However, these did not represent a useful choice. For records beyond 700 days, estimated curves would frequently be extrapolations and thus be subject to big sampling variances and CIs.

Analysis Q3333 did not fit any internal knots, *i.e.*, was equivalent to fitting quadratic polynomials for all effects, and, as expected, results were virtually identical to those obtained fitting Legendre polynomials of order 3 [28]. Fitting a curve consisting of two quadratic segments with a single knot at the mean of the range of ages, analysis Q4444, yielded estimates which increased sharply at the highest ages, in particular for permanent environmental variances due to animals (σ^2_R). Furthermore, estimates were quite similar to those from earlier analyses fitting cubic polynomials for direct genetic and permanent environmental effects (not shown), obtained by Meyer [28].

Similarly, fitting a cubic B-spline (analysis C7676) gave estimates of σ^2_R and thus the phenotypic variance (σ^2_P), which rose steeply and implausibly at the highest ages. However, this did not occur until ages were close to 800 days. Results indicate that the erratic behaviour observed, both here and in earlier analysis, was due to the cubic RR coefficient. Whilst effects can be reduced by fitting a B-spline curve where individual observation are restricted to a local influence, B-splines are no guarantee against all end-of-range problems.

Estimates of both maternal variances, shown in Figure 4, declined from peaks at about 200 days to about 500 days of age. Estimates of the maternal permanent environmental variance (σ^2_C), however, then increased again, steeply so at the upper end of the range of ages for which maternal effects were assumed to be effective. This occurred for all orders of splines fitted, though least for linear B-splines. An additional analysis (not shown) for model Q6565a considered maternal effects to be absent after 500 days. However, the resulting estimate of the variance function for σ^2_C was of similar shape to that shown in Figure 4 for Q6565a (rising steeply at the new end of range), *i.e.*, simply reducing the upper limit for maternal effects did not provide a cure for erratic estimates at the upper bound. Problems were least evident for model M3353, *i.e.*, a simply quadratic polynomial to model maternal effects. Clearly, current RR models for maternal components are not quite satisfactory. Future research should examine the scope to constrain the corresponding trajectories, so that they approach a value of zero at the upper limit.

Fitting a linear spline is equivalent to approximating the growth trajectory by a "broken-stick" curve. Whilst overall yielding sensible estimates of variances, some angularity of the variance functions could be observed, especially

for σ^2_R and both maternal variances. Increasing the number of knots tended to reduce variance estimates (not maternal components) at the highest ages slightly. Fitting quadratic or cubic segments yielded smooth estimates of variance functions, without discernible changes at the knots, except for σ^2_C from analysis C7676.

Estimates of MSE for the different analyses are given in Table I. Clearly, there was a counterbalance between the number of knots and the order of fit. The influence of the number of knots was largest for the lowest degree curves, and, conversely, the order of fit was most important for small numbers of knots. MSEs were reduced from 49.1 (L4343) to 29.4 (Q5454) by fitting quadratic rather than linear splines, when there were two internal knots for direct effects at 300 and 600 days, respectively. However, allowing knots at 200, 400 and 600 days, corresponding values were 25.1 (L5454) and 24.3 (Q6565a), *i.e.* there was much less additional variation modelled by the quadratic coefficients. Introducing cubic coefficients for the same knots (C7676), reduced MSE substantially to 16.3. At equal number of RR coefficients fitted, MSE for the linear model (L6565a) was slightly less than that for a quadratic model (Q6565a). Other criteria might have ranked models differently [18], especially those including a penalty for the number of parameters fitted. Contrasting MSEs to those for polynomial analyses with equal numbers of parameters, showed little difference in the amount of variation explained, except for the analysis with the least number parameters (L4343). However, this was to a some extent due to high, erratic variance estimates at the highest ages for model M4343, which "picked up" all variation at those points.

4.2. Eigenvalues and eigenfunctions

Eigenvalues and -functions of a covariance function summarise both the variance and the correlation structure. Hence, these quantities can be used to compare estimates from different models of analyses efficiently. Estimates of the first three eigenvalues and their CIs of the direct genetic and permanent environmental covariance functions for are shown in Figure 5, together with corresponding values for analyses M3333, M4444, M3353 and M5353 (from [28]).

Estimates of the first (λ_1) and second (λ_2) genetic eigenvalue for all analyses were very similar, with overlapping CIs. For analyses Q3333, M3333 and M3353, the third (λ_3) genetic eigenvalue is the last value fitted. It seems to be a common occurrence in RR analyses that estimates of this last value are close to zero, regardless of the number of RR coefficients. Estimates of λ_3 for the

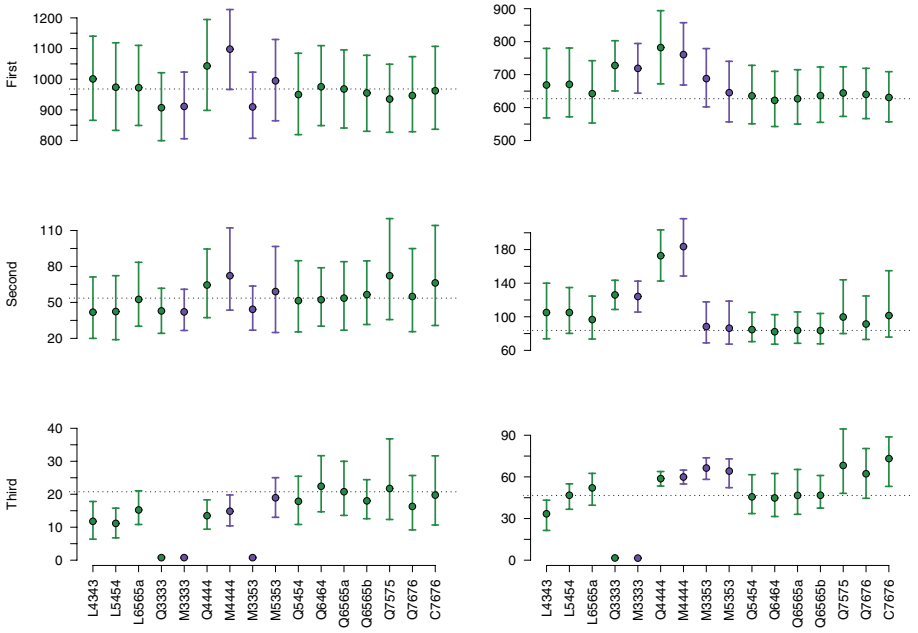

Figure 5. Estimates of the first three eigenvalues, together with approximate 95% confidence intervals, for direct genetic (left) and permanent environmental (right) covariance functions[a].
[a] dotted horizontal line gives estimates for analysis Q6565a.

other analyses showed relatively more differences than those for λ_1 and λ_2. In particular, values for a linear spline model with few knots (L4343 and L5454) were lower than the others, suggesting that genetic trajectories might not have been modelled adequately by a "broken-stick" curve. For direct, permanent environmental covariance functions, estimates of λ_1 and λ_2 for analyses Q3333 and Q4444 differed from the rest. This suggested that the comparative low number of parameters for these models was not sufficient to fully model variation in permanent environmental effects with age.

Estimates of the first (ψ_1) and second (ψ_2) direct genetic eigenfunction for four analyses are displayed in Figure 6. Overall, estimates of ψ_1 from all analyses were highly consistent, with narrow confidence regions up to about 700 days of age, even for models comprising six or more direct, genetic RR coefficients. As reported in other studies [22, 24], values for ψ_1 were positive throughout, *i.e.*, selection on the first eigenfunction is likely to change weight at all ages in the same direction. In addition, estimates agreed closely with those obtained fitting a quadratic (M3353, not shown) or quartic (M5353)

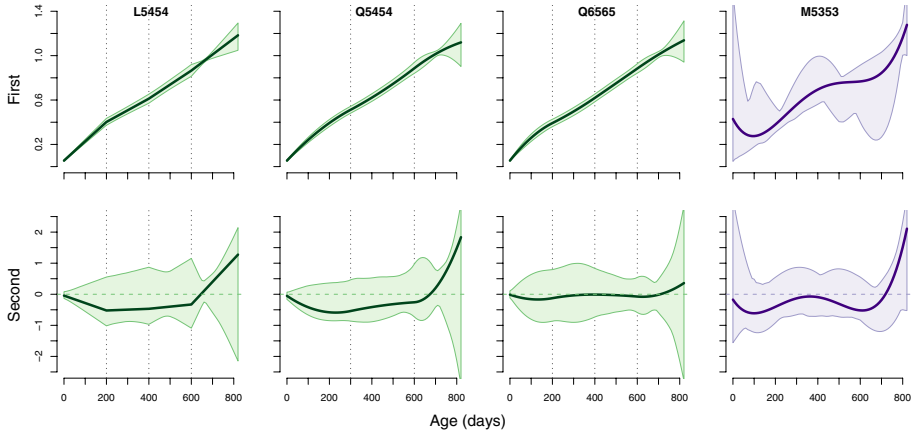

Figure 6. Estimates of the first (top row) and second (bottom row) direct genetic eigenfunctions (solid line), together with approximate 95% confidence region (filled area), for analyses L5454 (first column), Q5454 (second column), Q6565 (third column) and model M5353[a] fitting Legendre polynomials; vertical dotted lines indicate positions of internal knots.
[a] from Meyer [28].

polynomial. In contrast, however, the latter, analyses yielded estimates of ψ_1 with large, irregular CIs. Moreover, fitting cubic Legendre polynomials to model direct genetic trajectories, gave very different estimates of ψ_1 [28]. Results for the cubic B-spline model (C7676, not shown) were virtually the same as their counterparts for quadratic and linear B-splines, albeit with slightly increased CIs. This emphasized that RR analyses fitting basis functions of B-spline as covariables were more robust than those fitting orthogonal polynomials.

For the second genetic eigenfunction, ψ_2, however, CIs regions were large and irregularly shaped for all spline analyses, except Q3333 and Q4444. Consequently, estimates fluctuated considerably between analyses, but were reasonably consistent in such that values for ψ_2 were just below zero at birth, were slightly to moderately negative until 600 to 650 days, and then increased to become moderately to strongly positive. Similar fluctuations in estimates and CIs (not shown) were observed for the third genetic eigenfunction.

4.3. Genetic parameters

On the basis of MSE and number of parameters to be estimated, a quadratic B-spline with internal knots at 200, 400 and 600 days (Q6565a) appeared an

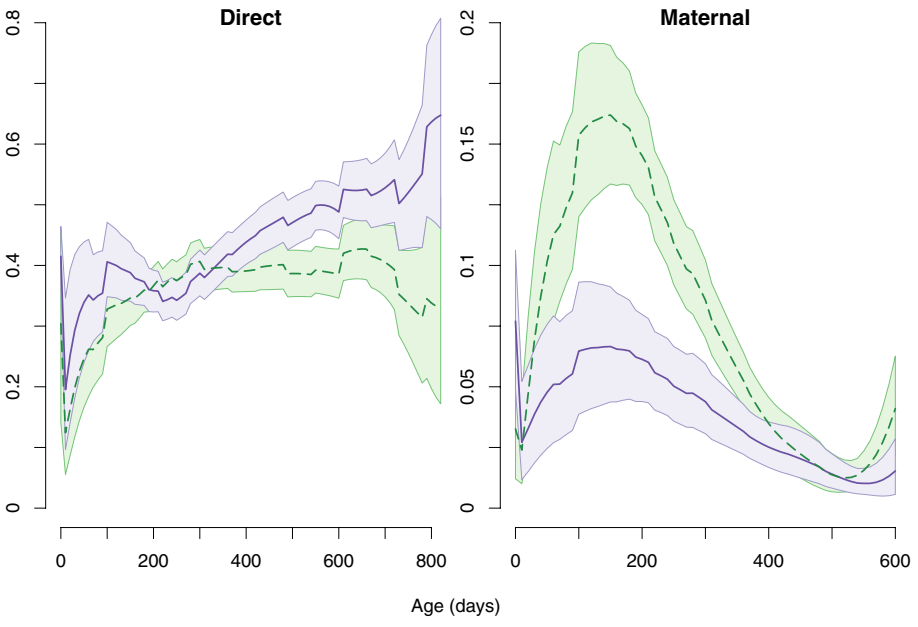

Figure 7. Estimates of heritabilities (solid line) and the proportion of variance due to permanent environmental effects (dashed line) from analysis Q6565a for direct (left) and maternal (right) effects, together with approximate 95% confidence regions (filled area).

adequate choice to model growth of beef cattle from birth to about two years of age. Estimates of heritabilities (h^2) and the proportions of σ_P^2 due to permanent environmental effects (p^2) for this model are depicted in Figure 7. As discussed previously [28], h^2 values, especially for ages corresponding to standard weaning and yearling weights, were higher than most literature values (*e.g.* [23]), and higher than corresponding univariate estimates from the same data [28]. Repeatability estimates (not shown) rose from about 70% at birth and from 120 days onwards to about 90% shortly after 600 days. At the highest ages, estimates of h^2 rose and values for p^2 decreased, whilst their sum remained more or less constant, pinpointing problems in disentangling genetic and permanent environmental variation.

Similarly, estimates of the maternal heritability (m^2) around weaning were somewhat lower than found in other analyses of Angus data, whilst the proportions of total variance explained by maternal permanent environmental effects (c^2) were somewhat higher. Again though, the sum, $m^2 + c^2$, was well within the range of literature values [23]. Both m^2 and c^2 dropped to values

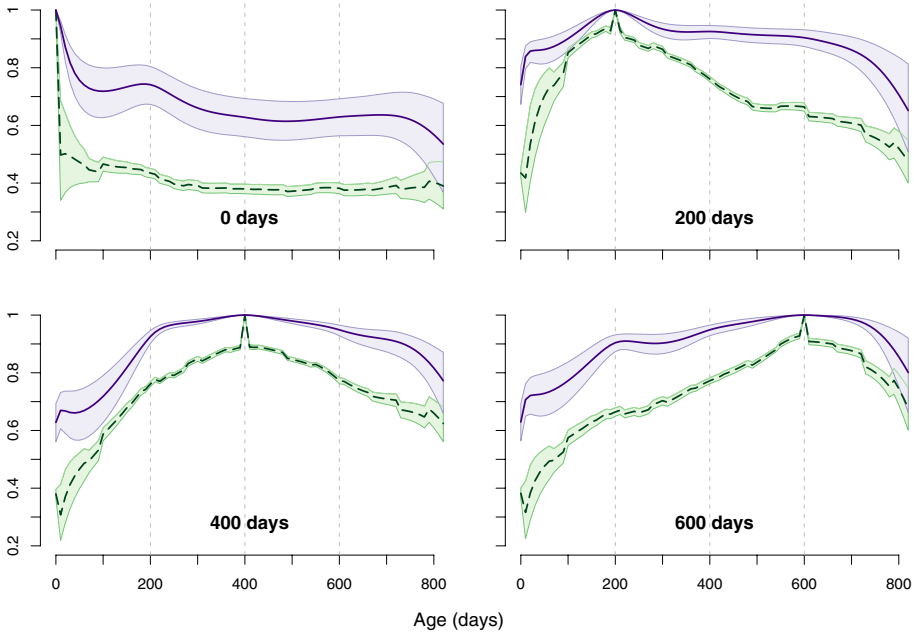

Figure 8. Estimates of direct genetic (solid line) and phenotypic (dashed line) correlations, together with approximate confidence regions (filled area), for weight at 0 days (top left), 200 days (top right), 400 days (bottom left) and 600 days (bottom right) with weights at all other ages, for analysis Q6565a; vertical dotted lines indicate positions of internal knots.

close to zero at about 500 days, suggesting that maternal effects would be unimportant from then on. Small subsequent increases in estimates of c^2 had large CIs, and should be regarded as spurious.

Figure 8 shows estimates of direct genetic (r_A) and phenotypic (r_P) correlations of weights at the target ages of 0, 200, 400 and 600 days with all other ages (analysis Q6565a). A contour plot, summarising all estimates for r_A and direct permanent environmental (r_R) correlations, is displayed in Figure 9. Estimates agreed, by and large, with previous results [28], with r_R decreasing consistently with increasing time between measurements, and r_A amongst most ages beyond 300 days above 0.9. However there were some unexpected, albeit small, oscillations in contour levels for both r_R and r_A. Whilst not noticeably related to knot position, some effects of joints between pieces of polynomials on estimates of r_A are evident in Figure 8, in particular for the knot at 200 days of age. This may have contributed to the oscillations in contour levels observed in Figure 9.

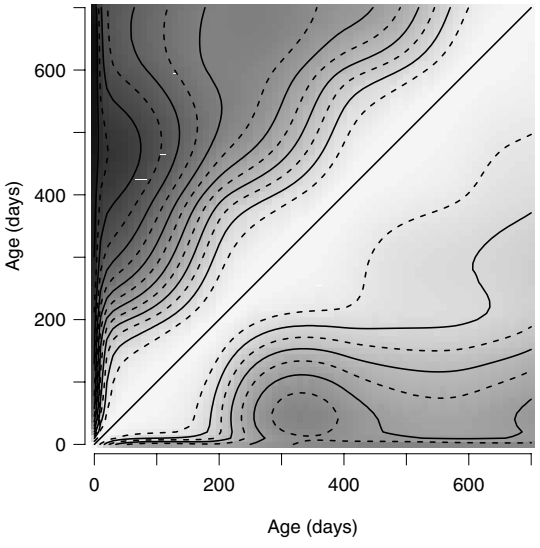

Figure 9. Contour plots for estimates of direct genetic (lower triangle) and permanent environmental (upper triangle) correlations, for ages from birth to 700 days and analysis Q6565a; contour lines given are in steps of 0.05 (solid: 1.0 (diagonal), $0.9, 0.8, \ldots$, dashed: $0.95, 0.85, \ldots$).

5. DISCUSSION

Results show that RR models fitting basis functions of B-splines of age at recording as covariables are well suited to modelling growth trajectories of beef cattle and their dispersion structure. In particular, they were less susceptible to the "end-of-range" problems frequently observed for such analyses when fitting polynomial regressions. This was due to the lower degree of polynomials in the individual segments as well as the local rather than global influence of individual observations. However, as illustrated, B-spline RR models do not provide a panacea for such problems, and are affected by sparsity and irregular distribution of records. Care must be taken in choosing appropriate knots and degrees of B-splines.

Only a limited number of models has been investigated. Other combinations of knots, both number and position, and degree of B-splines may well be more appropriate. For instance, some authors recommend knots placed at equal quantiles instead of equidistant knots [34].

The number of RR coefficients to be fitted for a random effect is determined by the number of knots chosen and the degree of the B-spline. Even for a small number of knots, this is typically larger than the number of RR coefficients which would have been fitted in a high-order polynomial model. As the number of covariance components to be estimated increases quadratically with the number of RR coefficients, this can have substantial ramifications for both computational demands and accuracy of estimation.

For k RR coefficients, an "unstructured" covariance matrix, as estimated in this study, comprised $k(k + 1)/2$ parameters. However, as the analysis of the eigenstructure of estimated covariance functions showed, there was some redundancy. With the last eigenvalues close to zero, the first m eigenfunctions explained the bulk of variation. As emphasized by Kirkpatrick and Meyer [21], this implies that estimating the first m eigenfunctions only would be sufficient, reducing the number of parameters to be estimated to $m(2k-m+1)/2$. A simple simulation study to examine properties of reduced rank estimates of covariance function used estimates for direct covariances matrices among RR coefficients (A and R) and for residual variances from analysis Q7676 as population values. Results suggested that the first four genetic and five permanent environmental principal components sufficed to model the corresponding trajectories and their variation [29].

The scope for reduced rank analyses has also been recognised by Shi *et al.* [36]. Similarly, James *et al.* [17] considered reduced rank estimation for function-valued data, and demonstrated that it resulted in better estimates than full rank mixed model analyses. Both authors used maximum likelihood estimation. The reduced rank, principal component parameterisation is readily implemented in a restricted maximum likelihood framework for quantitative genetic analyses [30]. Whilst asymptotic sampling distributions of eigenvalues and -vectors of matrices with a Wishart distribution are simple [19], fully conditional distributions required for Bayesian estimation, however, are likely to be less straightforward, and have not been documented.

6. CONCLUSION

The basis functions of B-splines provide a useful alternatives to the orthogonal polynomials commonly used to model trajectories of function-valued traits recorded in livestock improvement programmes. They were found to be well suited to describe variation in growth curves of beef cattle, and appeared more robust against "end-of-range" problems than polynomial models. Moreover, they provided more consistent estimates of the first, most important genetic eigenfunction.

ACKNOWLEDGEMENTS

This work was supported by grant BFGEN.100 of Meat and Livestock Australia Ltd (MLA). We are indebted to the Australian Angus Society for the use of their data.

REFERENCES

[1] Cantet R.J.C., B-spline fitting of time as an alternative to contemporary group effects. CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod. (2002), Communication No. 20–11.

[2] de Boor C., A Practical Guide to Splines, vol. 27 of *Springer Series in Applied Mathematics*, Springer Verlag, New York, 2nd edn., 2001.

[3] De Groot B.J., Keown J.F., Kachman S.D., Van Vleck L.D., 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 (2003) 37.

[4] Dierkx P., Curve and Surface Fitting with Splines, Oxford University Press, Oxford and New York, 1993.

[5] Dodge Y., The Oxford Dictionary of Statistical Terms, Oxford University Press, Oxford and New York, 2003.

[6] Durbán M., Harezlak J., Wand M.P., Carroll R.J., Simple fitting of subject-specific curves for longitudinal data, Stat. Med. 24 (2005) 1153–1167.

[7] Eilers P.H.C., Discussion on paper by Verbyla, Cullis, Kenward and Welham, Appl. Stat. 48 (1999) 307–308.

[8] Eilers P.H.C., Marx B.D., Flexible smoothing with B-splines and penalties (with discussion), Stat. Sci. 11 (1996) 89–121.

[9] Eilers P.H.C., Marx B.D., Splines, knots, and penalties, J. Comput. Graph. Statist. (2005), in press. http://www.stat.lsu.edu/faculty/marx/splines_knots_penalties.pdf.

[10] El Faro L., Albuquerque L.G., Fries L.A., Adjusting lactation curves with segmented polynomials, in: Proceedings Sixth World Congr. Genet. Appl. Livest. Prod., Vol. 23, Armidale, Australia (1998), pp. 411–414.

[11] Fuller W.A., Grafted polynomials as approximating functions, Austr. J. Agric. Econ. 13 (1969) 35–46.

[12] Gilmour A., Gogel B., Cullis B.R., Welham S.J., Thompson R., ASREML User Guide Release 1.0. VSN International Ltd., Hemel Hempstead, HP1 1ES, UK (2001).

[13] Green P.J., Silverman B.W., Nonparametric regression and Generalized Linear Models. A Roughness Penalty Approach. Monographs in Statistics and Applied Probability 58, Chapman & Hall, London, 1994.

[14] Guo W., Functional mixed effects models, Biometrics 58 (2002) 121–128.

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

[16] Ihaka R., Gentleman R., R: A language for data analysis and graphics, J. Comput. Graph. Statist. 5 (1996) 299–314.

[17] James G.M., Hastie T.J., Sugar C.A., Principal component models for sparse functional data, Biometrika 87 (2000) 587–602.

[18] Jamrozik J., Schaeffer L.R., Bayesian comparison of random regression models for test-day yields in dairy cattle, CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod. (2002), Communication No. 1–03.

[19] Jollife I.T., Principal Component Analysis, Springer Series in Statistics, Springer Verlag, New York, 1986.
[20] Kirkpatrick M., Heckman N., A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters, J. Math. Biol. 27 (1989) 429–450.
[21] Kirkpatrick M., Meyer K., Simplified analysis of complex phenotypes: Direct estimation of genetic principal components, Genetics 168 (2004) 2295–2306.
[22] Kirkpatrick M., Lofsvold D., Bulmer M., Analysis of the inheritance, selection and evolution of growth trajectories, Genetics 124 (1990) 979–993.
[23] Koots K.R., Gibson J.P., Smith C., Wilton J.W., Analyses of published genetic parameter estimates for beef production traits. 1. Heritability, Anim. Breed. Abstr. 62 (1994) 309–337.
[24] Meyer K., Estimates of genetic and phenotypic covariance functions for post-weaning growth and mature weight of beef cows, J. Anim. Breed. Genet. 116 (1999) 181–205.
[25] Meyer K., Random regressions to model phenotypic variation in monthly weights of Australian beef cows, Livest. Prod. Sci. 65 (2000) 19–38.
[26] Meyer K., Estimates of covariance functions for growth of Australian beef cattle from a large set of field data. CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod. (2002), Communication No. 11–01.
[27] Meyer K., "RRGIBBS" – A program for simple random regression analyses via Gibbs sampling. CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod. (2002), Communication No. 28–27.
[28] Meyer K., Estimates of genetic covariance functions for growth of Australian Angus cattle, J. Anim. Breed. Genet. 122 (2005) 73–85.
[29] Meyer K., Sampling behaviour of reduced rank estimates of genetic covariance functions, Proc. Ass. Advan. Anim. Breed. Genet. 16 (2005) 226–289.
[30] Meyer K., Kirkpatrick M., Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices, Genet. Sel. Evol. 37 (2005) 1–30.
[31] Meyer K., Kirkpatrick M., Up hill, down dale: quantitative genetics of curvaceous traits, Phil. Trans. R. Soc. B 360 (2005) 1443–1455.
[32] Nobre P.C.R., Misztal I., Tsuruta S., Bertrand J.K., Silva L.O.C., Lopes P.S., Analyses of growth curves of Nellore cattle by multiple-trait and random regression models, J. Anim. Sci. 81 (2003) 918–926.
[33] Rice J.A., Wu C.O., Nonparametric mixed effects models for unequally sampled noisy curves, Biometrics 57 (2001) 253–259.
[34] Ruppert D., Wand M.P., Carroll R.J., Semiparametric Regression, Cambridge University Press, New York, 2003.
[35] Schaeffer L.R., Application of random regression models in animal breeding, Livest. Prod. Sci. 86 (2004) 35–45.
[36] Shi M., Weiss R.E., Taylor J.M.G., An analysis of pediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves, Appl. Stat. 45 (1996) 151–164.

[37] Strandberg E., Grandinson E., Adjusting for seasonal effects in an animal model using fuzzy classification, Proceedings Sixth World Congr., Genet. Appl. Livest. Prod. 25 (1998) 633–636.
[38] Torres R.A.A., Markov chain Monte Carlo methods for estimating the covariance structure of longitudinal data – an application to dairy cattle data, Ph.D. thesis, Cornell University, 2001.
[39] Torres R.A.A., Quaas R.L., Determination of covariance functions for lactation traits on dairy cattle using random-coefficient regressions on B-splines, J. Anim. Sci. 79 Suppl. 1 (2001) 112.
[40] Verbyla A.R., Cullis B.R., Kenward M.G., Welham S.J., The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion), Appl. Stat. 48 (1999) 269–311.
[41] White I.M.S., Thompson R., Brotherstone S., Genetic and environmental smoothing of lactation curves with cubic splines, J. Dairy Sci. 82 (1999) 632–638.

To access this journal online:
www.edpsciences.org