BIOMETRICS 62, 392–401
June 2006

DOI: 10.1111/j.1541-0420.2005.00500.x

The Analysis of Longitudinal Data Using Mixed Model L-Splines

Sue J. Welham,[1,*] Brian R. Cullis,[2] Michael G. Kenward,[3] and Robin Thompson[1]

[1]Rothamsted Research, Harpenden AL5 2JQ, U.K.
[2]Wagga Agricultural Institute, Wagga Wagga, NSW 2650, Australia
[3]London School of Hygiene and Tropical Medicine, London WC1E 7HT, U.K.
[*]*email:* sue.welham@bbsrc.ac.uk

SUMMARY. L-splines are a large family of smoothing splines defined in terms of a linear differential operator. This article develops L-splines within the context of linear mixed models and uses the resulting mixed model L-spline to analyze longitudinal data from a grassland experiment. In the spirit of time-series analysis, a periodic mixed model L-spline is developed, which partitions data into a smooth periodic component plus smooth long-term trend.

KEY WORDS: Longitudinal data; L-splines; Mixed models; Residual maximum likelihood; Smoothing splines.

1. Introduction

This article develops mixed model L-splines, motivated by an experiment to investigate grassland response to different treatment regimes. Within this experiment, data on available pasture dry matter (termed food on offer) were collected from individual plots over 4 years, which showed a strong seasonal pattern as well as longer-term trend. The aim was to partition the data for each treatment into periodic and nonperiodic components, both modeled as a continuous function of time. As the plot profiles across time did not correspond to any simple parametric form, smoothing splines provided a convenient method of curve fitting, which could be built into a standard linear mixed model for longitudinal data.

Cubic smoothing splines were introduced into the mixed model setting by Wang (1998), Zhang et al. (1998), Brumback and Rice (1998), and Verbyla et al. (1999), who all used the mathematical equivalence between the penalized sum of squares used to fit a cubic spline and best linear unbiased predictor (BLUP) estimation in a particular mixed model for a given smoothing parameter. Within this mixed model, the cubic spline corresponds to fitting a model with linear trend plus a set of random covariates whose covariance matrix is known apart from a scaling constant, called the smoothing variance component. In addition, residual maximum likelihood (REML) estimation of the smoothing variance component in the linear mixed model is equivalent to the generalized maximum likelihood method for smoothing parameter estimation of Wahba (1990). Within the linear mixed model context, a cubic spline can be included within a general treatment structure and fitted at different levels of the structure, for example, as a common spline across all treatments, or as separate splines for different treatment combinations (Verbyla et al., 1999). Additional random terms can easily be added to the model to account for all sources of variation in the data. Other types of smoothing splines have also been used within mixed models, with the same motivation of building nonparametric terms into a general and flexible family of models. Eilers and Marx (1996) developed P-splines, which used a B-spline basis with an approximate discrete penalty and a reduced set of knots. Although they fitted these models using cross-validation, Eilers noted in the discussion of Verbyla et al. (1999) that P-splines could also be fitted as mixed model splines with the smoothing parameter estimated by REML, as demonstrated by Currie and Durban (2002). Parise et al. (2001) proposed a slightly different mixed model spline, using truncated power basis functions for polynomial splines with an identity penalty matrix, and called these functions penalized splines. These spline models also used a reduced set of knots and the smoothing parameter was estimated using REML. Wand (2003) gives a general review of penalized spline models and some extensions.

In the mixed model context the cubic spline is partitioned into a fixed linear component plus a random component, with zero expectation, representing smooth deviations about the linear trend. Although the standard interpretation of a spline is as a single smooth trend, we find this partition useful to interpret the difference between L-splines with polynomial and periodic core functions. For a given value of the smoothing parameter, the fitted cubic spline is determined by minimizing the residual sum of squares for the model subject to a penalty consisting of the integrated squared second derivative of the fitted curve, i.e., of deviations from linear trend, scaled by the smoothing parameter. However, in many examples the underlying trend is not linear. In the grassland experiment considered later, the underlying pattern is a seasonal cycle with some linear trend, which does not reflect the implicit assumptions of the mixed model cubic spline. It then seems more natural to use L-splines that fit an appropriate underlying form plus smooth deviations about this underlying form. The amount of smoothing is controlled by a penalty constructed to penalize departures from the appropriate underlying form. The theory behind L-splines is given briefly by Wahba (1990), and more accessibly by Gu (2002) or Ramsay and Silverman (1997). Ramsay and Dalzell (1991), Ramsay and Silverman (1997), and Heckman and Ramsay (2000) all used L-splines in preference to cubic splines where the underlying form of the data was not linear, and particularly for periodic data.

In this article, we show that L-splines can be fitted as mixed models and extend the model to partition deviations from the underlying form into periodic and nonperiodic components. We demonstrate our approach using mixed model L-splines to model data from the grassland experiment described in Section 2. Section 3 gives a brief overview of the construction of L-splines, and Section 4 shows the mixed model form of the L-spline, extends the definition to a reduced set of knots, and evaluates the performance of an L-spline, in terms of mean squared error of prediction, in a small simulation study. In Section 5, the development of a periodic L-spline term is described and its performance is evaluated. Finally, in Section 6 the methods are used to analyze data from the grassland experiment and obtain results, with discussion in Section 7.

2. Grassland Experiment

The data that motivated the analysis in this article were kindly provided by D. L. Garden of NSW DPI (Canberra) and arose from a grassland experiment designed to investigate the effects of superphosphate application and grazing on the production of native grasslands in the high rainfall zone of south-eastern Australia. Further details of the experimental approach, design, and analysis of additional variables can be found in Garden et al. (2003). The experiment was designed to test the effect of six fertilizer treatments: four levels of superphosphate application (nil, low, medium, high), high superphosphate application with lime, or sewage ash application. The experimental design was laid out in the field as a randomized block design with two replicates of six plots. The plots were grazed continuously with a similar stocking rate across all plots at the start of the experiment. Stocking rates were adjusted from time to time for each treatment with the aim of maintaining individual sheep liveweights at similar levels across treatments so that increases in productivity were reflected in the number of animals per hectare. Fertilizer treatment and stocking rates were therefore confounded. However, the amount of food on offer (FOO) per plot in the presence of grazing was still of interest as a measure of the success of the stocking rate strategy for each treatment. FOO was defined as herbage mass in kg dry matter per hectare (DM/ha) and was measured by sampling 30 0.5 m × 0.5 m quadrats per plot on eight occasions, 6–7 weeks apart, within each year. All plots were sampled on each occasion and the samples were taken at approximately the same dates within each year. The log-transformed FOO data are shown in Figure 1, which clearly shows yearly cycles with substantial downward trend.

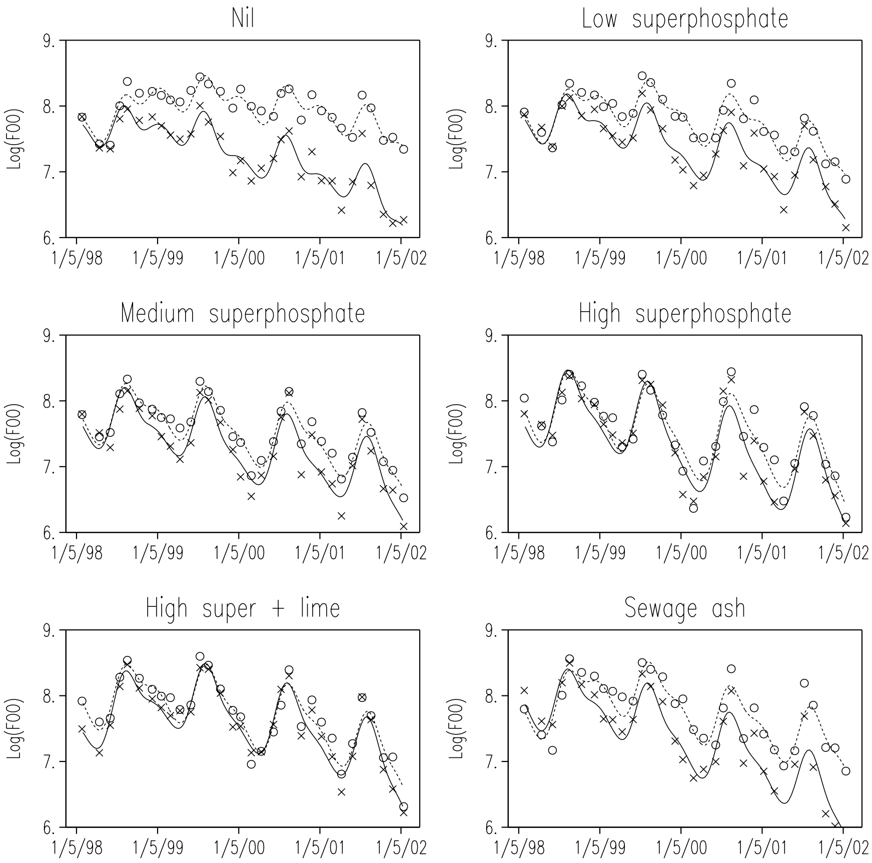

Figure 1. Log(FOO) measurements (as log kg dry matter per hectare) across time for each plot of each treatment (× = replicate 1, ∘ = replicate 2) with predictive component of the final model from Table 3 (solid line = replicate 1, dashed line = replicate 2).

3. L-Splines

Suppose we have data $\boldsymbol{y}$ $(n \times 1)$ and an explanatory variable $\boldsymbol{x}$ observed at unique values $x_1, x_2, \ldots, x_n$ in the range $[a,\ b]$. An L-spline is defined in terms of the underlying form of the data, described by a set of core functions $\{f_j;\ j = 1, \ldots, m\}$, and associated linear differential operator L of order m, which annihilates the core functions, i.e., $Lf = 0$ if and only if f is a linear combination of the f_j or $f \equiv 0$. For example, the linear differential operator $L = D^4 + \omega^2 D^2$ annihilates the set of core functions

$$f_1(t) = 1; \quad f_2(t) = t; \quad f_3(t) = \cos(\omega t); \quad f_4(t) = \sin(\omega t).$$

The polynomial splines (of odd degree $k = 2m - 1$) form a subset of L-splines where $L = D^m$ for positive integers m. In general, the functions f may depend on unknown parameters. In this article, we assume that all such parameters (e.g., the period parameter ω) are known.

For a single spline term, we fit a model

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{e},$$

where $\boldsymbol{g}$ is the realization of an unknown smooth function $g(t)$ at the data points $\boldsymbol{x}$, i.e., $\boldsymbol{g} = g(\boldsymbol{x})$, and $\boldsymbol{e}$ is a vector of errors with $\boldsymbol{e} \sim N(\boldsymbol{0},\ \sigma^2\boldsymbol{R})$, for some general covariance matrix $\boldsymbol{R}$. The L-spline is the function $g(t)$ that minimizes the penalized sum of squares,

$$(\boldsymbol{y} - \boldsymbol{g})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{g}) + \lambda \int_a^b [Lg(s)]^2\, ds, \qquad (1)$$

for a given value of λ, the smoothing parameter. The smoothing parameter determines the balance between fidelity to the data (measured by the first term, a residual sum of squares) and fidelity to the underlying form (measured by the second term). Ramsay and Silverman (1997, Section 15.2) prove that, for any basis $\{f_j;\ j = 1, \ldots, m\}$ for the set of core functions, the function g minimizing the penalized sum of squares (1) has the form

$$g(t) = \sum_{j=1}^{m} \tau_j f_j(t) + \sum_{i=1}^{n} c_i k_2(x_i, t),$$

where k_2 is the reproducing kernel function for a subspace defined in terms of the operator L and a set of boundary conditions. We used initial value boundary constraints (see Ramsay and Silverman, 1997, Section 13.5.1). Ramsay and Silverman (1997, Section 15.3) further show that equation (1) can then be reexpressed as

$$(\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{Kc})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{Kc}) + \lambda \boldsymbol{c}'\boldsymbol{Kc}, \qquad (2)$$

where $[\boldsymbol{X}]_{ij} = f_j(x_i)$ for $i = 1, \ldots, n,\ j = 1, \ldots, m$, $[\boldsymbol{K}]_{ij} = k_2(x_j, x_i)$ for $i, j = 1, \ldots, n$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)'$, and $\boldsymbol{c} = (c_1, \ldots, c_n)'$. Gu (2002) shows that the symmetric matrix $\boldsymbol{K}$ is nonnegative definite. Heckman and Ramsay (2000) and Dalzell and Ramsay (1993) give recipes for constructing reproducing kernel functions k_2. For $L = \omega^2 D^2 + D^4$ and $a = 0$ the reproducing kernel function k_2 can be written as

$$\begin{aligned} k_2(x_i, t) = \frac{1}{\omega^6}\Bigg\{ & \frac{\omega^2 t^2}{6}(3x_i - t) - (x_i - t) + x_i\cos(\omega t) \\ & + t\cos(\omega x_i) - \frac{1}{\omega}\sin(\omega t) - \frac{1}{\omega}\sin(\omega x_i) \\ & + \frac{1}{2}t\cos[\omega(x_i - t)] + \frac{1}{\omega}\sin[\omega(x_i - t)] \\ & - \frac{1}{2\omega}\cos(\omega x_i)\sin(\omega t)\Bigg\} \quad \text{for} \quad t < x_i \\ k_2(x_i, t) = \frac{1}{\omega^6}\Bigg\{ & \frac{\omega^2 x_i^2}{6}(3t - x_i) - (t - x_i) + t\cos(\omega x_i) \\ & + x_i\cos(\omega t) - \frac{1}{\omega}\sin(\omega x_i) - \frac{1}{\omega}\sin(\omega t) \\ & + \frac{1}{2}x_i\cos[\omega(t - x_i)] + \frac{1}{\omega}\sin[\omega(t - x_i)] \\ & - \frac{1}{2\omega}\cos(\omega t)\sin(\omega x_i)\Bigg\} \quad \text{for} \quad x_i < t. \end{aligned} \qquad (3)$$

Consider functions defined piecewise on $[x_{i-1},\ x_i]$ as

$$\begin{aligned} & \beta_{1i} + \beta_{2i}t + \beta_{3i}t^2 + \beta_{4i}t^3 + \beta_{5i}\cos\omega t \\ & \quad + \beta_{6i}\sin\omega t + \beta_{7i}t\cos\omega t + \beta_{8i}t\sin\omega t, \end{aligned} \qquad (4)$$

for $x_{i-1} \le t \le x_i, i = 1, \ldots, n$ $(x_0 = a)$, with the requirement that the overall function is continuous and differentiable up to order 6 $(= 2m - 2)$ at the knots. It is straightforward to show that the functions defined in equation (3) are of this form. Together with the global functions $\{1,\ t, t^2,\ t^3,$ $\cos(\omega t), \sin(\omega t), t\cos(\omega t), t\sin(\omega t)\}$, these functions span the space of six-times differentiable functions that are piecewise (between knots) of the form (4), i.e., periodic with linearly varying amplitude and with added cubic trend.

4. L-Spline Mixed Models

The L-spline is fitted to the data $\boldsymbol{y}$ by finding the coefficient values $\hat{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{c}}$ that minimize the penalized sum of squares (2) for a given value of the smoothing parameter λ. Minimization of equation (2) requires solution of the equations

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{K} \\ \boldsymbol{K}\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{K}\boldsymbol{R}^{-1}\boldsymbol{K} + \lambda\boldsymbol{K} \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\boldsymbol{c}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{K}\boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix}.$$

These equations contain m implicit constraints that can be expressed as $\boldsymbol{X}'\tilde{\boldsymbol{c}} = \boldsymbol{0}$ (see, for example, Wahba, 1990, equation 1.3.17). This property has the consequence that the fitted spline takes the underlying form outside the range of the data, i.e., $Lg(t) = 0$ for $t < x_1$ or $t > x_n$. Following the terminology of polynomial splines, we call this a natural L-spline. For convenience, we make this implicit constraint on the parameter estimates explicit in our model as $\boldsymbol{X}'\boldsymbol{c} = \boldsymbol{0}$. The constraint $\boldsymbol{X}'\boldsymbol{c} = \boldsymbol{0}$ implies $\boldsymbol{c} = \boldsymbol{C\delta}$ for some vector $\boldsymbol{\delta}$ of length $n - m$ where $\boldsymbol{C}$ is any $n \times n - m$ matrix $\boldsymbol{C}$ of full column rank such that $\boldsymbol{X}'\boldsymbol{C} = \boldsymbol{0}$. Inserting this reparameterization into the penalized sum of squares (2) gives

$$(\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{Z\delta})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{Z\delta}) + \lambda\boldsymbol{\delta}'\boldsymbol{H}^{-1}\boldsymbol{\delta},$$

where $\boldsymbol{Z} = \boldsymbol{KC}$, $\boldsymbol{H}^{-1} = \boldsymbol{C}'\boldsymbol{KC}$, and it can be shown that $\boldsymbol{H}$ is positive definite. Minimizing this expression leads to equations

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}+\lambda\boldsymbol{H}^{-1} \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\boldsymbol{\delta}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix},$$

which are the mixed model equations for the model

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Z\delta} + \boldsymbol{e},$$

where $\boldsymbol{X\tau}$ represents the core functions as fixed terms in the model, $\boldsymbol{Z\delta}$ represents the constrained basis of reproducing kernel functions as a random model term, vector $\boldsymbol{e}$ represents a residual term, and $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{R}), \boldsymbol{\delta} \sim N(\boldsymbol{0}, \sigma_s^2\boldsymbol{H}), \operatorname{cov}(\boldsymbol{\delta}, \boldsymbol{e}) = \boldsymbol{0}, \lambda = \sigma^2/\sigma_s^2$.

As for cubic splines (Verbyla et al., 1999), the fitted L-spline can then be calculated as a BLUP from this mixed model for a given value of the smoothing parameter. The smoothing parameter can either be considered fixed at some predetermined value or estimated, via the L-spline variance component σ_s^2, as part of the variance model using REML estimation. In this article, we use REML estimation of smoothing parameters, as used by Verbyla et al. (1999) for cubic smoothing spline models. The extension to include additional fixed or random terms is achieved by simply adding these terms into the mixed model.

It is often convenient in practice to transform to independent random effects $\boldsymbol{u}$, where $\boldsymbol{u} = \boldsymbol{H}^{-1/2}\boldsymbol{\delta}$ and the random design matrix $\boldsymbol{Z}$ is replaced by $\boldsymbol{ZH}^{1/2}$ with $\operatorname{var}(\boldsymbol{u}) = \sigma_s^2\boldsymbol{I}_{n-m}$. The mixed model L-spline can then easily be fitted by most standard mixed model software. We write this final model as

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Z}_u\boldsymbol{u} + \boldsymbol{e},$$

where $\boldsymbol{Z}_u = \boldsymbol{ZH}^{1/2}$ with other terms as above. For interpretation, it is also advantageous to make the spline design matrix $\boldsymbol{Z}_u$ orthogonal to the fixed effect design matrix $\boldsymbol{X}$.

The change in logRL, the logarithm of the REML likelihood function, on adding the random L-spline term to the model containing the core functions, i.e., testing for a zero variance component, can be used to investigate whether there is evidence of deviations from the underlying form (Guo, 2002b). As a likelihood ratio test with the variance component constrained to remain positive, the asymptotic distribution of 2 × the change in logRL is a 50:50 mixture of a χ_0^2 and a χ_1^2 distribution under the null hypothesis (Stram and Lee, 1994). However, Crainiceanu and Ruppert (2004) showed that in many examples this approximation was very poor, and developed a better approximation for models with a single variance component in addition to the residual variance. Crainiceanu et al. (2005) suggested parametric bootstrap evaluation of the null distribution for more complex variance models. Zhang and Lin (2003) specified a score test for the smoothing parameter.

4.1 *Reduced Knot Set*

For data sets with a large number of distinct covariate values, or with several spline terms in the model, the L-spline mixed model may generate a large number of random spline effects with a dense design matrix $\boldsymbol{Z}_u$. Solution of the mixed model equations may then require a large amount of computer workspace and processing time. Both can be reduced by using a (relatively) small number of knots, r say, defined at distinct values $\boldsymbol{t} = (t_1, t_2, \ldots, t_r)'$ with $a < t_1 < \cdots < t_r < b$. Our development here is similar to that of Parise et al. (2001) and Wand (2003), who used a reduced number of knots in generating polynomial spline basis functions as a low-rank approximation to the full basis.

The set of r L-spline basis functions is generated as $k_2(t_i, \cdot)$ for $i = 1, \ldots, r$, using the reduced set of knots, $\boldsymbol{t}$. The L-spline function then takes the form

$$g(t) = \sum_{j=1}^{m} \tau_j f_j(t) + \sum_{i=1}^{r} c_i k_2(t_i, t),$$

and the penalized sum of squares corresponding to this function is written as

$$(\boldsymbol{y}-\boldsymbol{g})'\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{g}) + \lambda \int_a^b [Lg(s)]^2\, ds$$
$$= (\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{K}_{xt}\boldsymbol{c})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{K}_{xt}\boldsymbol{c}) + \lambda\boldsymbol{c}'\boldsymbol{K}_{tt}\boldsymbol{c},$$

where now $[\boldsymbol{K}_{tt}]_{ij} = k_2(t_j, t_i),\ i, j = 1, \ldots, r$, because the penalty term is defined in terms of the r basis functions evaluated at the knots, $[\boldsymbol{K}_{xt}]_{ij} = k_2(t_j, x_i),\ i = 1, \ldots, n,\ j = 1, \ldots, r$, gives the value of the r basis functions at the n covariate values, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)'$ and $\boldsymbol{c} = (c_1, \ldots, c_r)'$. In order to retain the natural spline property achieved for L-splines with knots at all data points, we explicitly impose the constraint $\boldsymbol{X}_t'\boldsymbol{c} = \boldsymbol{0}$, where $[\boldsymbol{X}_t]_{ij} = f_j(t_i),\ i = 1, \ldots, r,\ j = 1, \ldots, m$, to get the amended penalized sum of squares

$$(\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{K}_{xt}\boldsymbol{C}_t\boldsymbol{\delta})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X\tau} - \boldsymbol{K}_{xt}\boldsymbol{C}_t\boldsymbol{\delta}) + \lambda\boldsymbol{\delta}'\boldsymbol{H}_{tt}^{-1}\boldsymbol{\delta}, \tag{5}$$

where $\boldsymbol{C}_t$ is an $r \times r - m$ matrix defined such that $\boldsymbol{X}_t'\boldsymbol{C}_t = \boldsymbol{0}$, $\boldsymbol{H}_{tt}^{-1} = \boldsymbol{C}_t'\boldsymbol{K}_{tt}\boldsymbol{C}_t$, and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{r-m})'$. Minimizing (5) yields mixed model equations for the model

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Z}_t\boldsymbol{\delta} + \boldsymbol{e}, \tag{6}$$

where $\boldsymbol{Z}_t = \boldsymbol{K}_{xt}\boldsymbol{C}_t$ and $\boldsymbol{\delta} \sim N(\boldsymbol{0}, \sigma_s^2\ \boldsymbol{H}_{tt})$.

The fitted spline, estimated smoothing parameter, and logRL are not invariant to the reduced set of knots chosen, although the fitted spline is usually similar at the data points. In addition, the reduced knot L-spline is no longer the function that produces the overall minimum of the penalized sum of squares (1) and care is required to ensure that the fitted spline is not sensitive to the knots used. Ruppert (2002) discusses choice of knots in the penalized spline context.

Verbyla et al. (1999), Wang (1998), and Brumback and Rice (1998) all utilized a hierarchical decomposition in their cubic spline mixed models by decomposing individual treatment splines into an overall spline, main effect splines, and interaction splines. In addition, Verbyla et al. (1999) used subject-specific splines to model the profile of individual subjects over time. A similar model was proposed by Guo (2002a). An exactly analogous decomposition can be used for L-spline mixed models.

4.2 *Evaluation of L-Splines versus Cubic Splines*

The cubic smoothing spline and L-splines give alternative methods to model smooth trend. It is, therefore, useful to consider whether use of an L-spline results in an improved fit to the data. We have investigated this via a small simulation study relevant to the log(FOO) data from the grassland experiment. For grassland, the amplitude, length, and timing of the growth phase depend on weather conditions that vary between years. In addition, the long-term trend may change according to grazing pressure and treatment regimes. To represent these features we simulated monthly data across 5 years, using known curves of the form

$$\begin{aligned} c_{jk}(t) &= s_{1jk}(t) + (s_{2jk}(t) - 0.03)(t - 30.5) \\ &\quad + [0.5 + s_{3jk}(t)]\sin[\omega_{jk}(t)(t + s_{4jk}(t))], \qquad (7) \end{aligned}$$

where $\omega_{jk}(t) = 2\pi/(12 + s_{5jk}(t))$, for $j = 1, \ldots, 9$, $k = 1, \ldots, 10$. The average amplitude and linear trend were chosen to match that of the log(FOO) data from the grassland experiment. The functions $s_{ijk}(t)$ were generated from independent realizations of an $N(0,\, 1)$ random walk, smoothed using a cubic spline with four effective degrees of freedom, then standardized to a predetermined range r_{ij}(so $r_{ij} = 0 \Rightarrow s_{ijk}(t) \equiv 0$). This generated smooth deviations in trend and in amplitude, length, and timing of the annual cycle. Nine combinations (j) of range values, shown in Table 1, were investigated. For each set, realizations $s_{ijk}(t)$ were generated for $k = 1, \ldots, 10$ and used to form a set of representative curves $c_{jk}(t)$. For each curve $c_{jk}(t)$, 500 sets of data were independently generated as $\boldsymbol{y}_{jkl} = c_{jk}(\boldsymbol{x}) + \boldsymbol{e}_{jkl}$, $l = 1, \ldots, 500$, with $\boldsymbol{x} = (1, \ldots, 60)'$, $\boldsymbol{e}_{jkl} \sim N(0,\, \sigma^2\boldsymbol{I}_{60})$. Two values of the residual variance σ^2 were assessed, with σ^2 equal to 0.01 or 0.001. Each data set was modeled using a cubic smoothing spline, a partial cubic spline (PCS) consisting of the cubic smoothing spline plus additional fixed periodic functions $\sin(\omega t)$ and $\cos(\omega t)$, and an L-spline with $L = \omega^2D^2 + D^4$. The splines were all fitted as mixed model splines using REML estimation of variance parameters, using several starting values to ensure that a global optimum was reached. The fit of the models to the data was evaluated using the average mean squared error of prediction (MSEP)

$$\frac{1}{10}\sum_{k=1}^{10}\frac{1}{500}\sum_{l=1}^{500}[c_{jk}(\boldsymbol{x}) - \tilde{c}_{jkl}(\boldsymbol{x})]'\,[c_{jk}(\boldsymbol{x}) - \tilde{c}_{jkl}(\boldsymbol{x})],$$

where $\tilde{c}_{jkl}(t)$ is the relevant fitted spline for data $\boldsymbol{y}_{jkl}$. Table 1 shows the average MSEP for each set of parameters. For $j = 1$, where deviations from the annual cycle consisted only of changes in long-term trend, the PCS had the smallest MSEP value. In all but one case, where variation was introduced into the periodic cycle the L-spline had the smallest MSEP. The relative improvement for the L-spline increased for $\sigma^2 = 0.001$ and then the PCS always had smaller MSEP than the cubic spline alone. For $\sigma^2 = 0.01$ the cubic spline sometimes had a smaller MSEP value than the partial spline, and for $j = 5$ was smaller than both the cubic smoothing spline and L-spline. In this case, both the PCS and L-spline occasionally oversmoothed the data, using variance components close to zero and inflating the overall MSEP. In contrast the cubic smoothing spline consistently overfitted the data, severely underestimating the residual variance.

5. Modeling the Within-Year Component: Periodic Splines

The L-spline mixed model described above has a seasonal component consisting of the periodic core functions, with long-term trend modeled by linear trend plus the random spline terms. However, the data could also contain seasonal pattern not adequately described by core functions, or by any alternative parametric form. It is then appropriate to introduce a periodic spline term to separate any additional seasonal pattern from long-term trend. To stay within the L-spline framework, it is sensible to add a periodic L-spline to the mixed model. We adapt the approach of Zhang, Lin,

Table 1

Average MSEP across 5000 data sets (model 7) for a cubic smoothing spline (CSS), partial cubic spline (PCS), and L-spline model (with $L = D^4 + \omega^2D^2$). Average MSEP values are also shown as a percentage of minimum value for each row.

							Average MSEP × 1000 (% min value)		
σ^2	j	r_{1j}	r_{2j}	r_{3j}	r_{4j}	r_{5j}	CSS	PCS	L-spline
0.01	1	0.2	0.1	0	0	0	3.959 (219)	1.811 (100)	2.088 (115)
	2	0.2	0	0.4	0	0	4.109 (125)	6.403 (195)	3.280 (100)
	3	0.2	0.1	0.4	0	0	3.909 (151)	2.989 (116)	2.586 (100)
	4	0	0	0.4	0	0	4.018 (123)	7.102 (218)	3.259 (100)
	5	0	0	0	2	0	3.990 (100)	6.593 (165)	4.422 (111)
	6	0	0	0	0	2	3.658 (136)	3.475 (130)	2.676 (100)
	7	0	0	0.4	2	2	4.613 (141)	4.323 (132)	3.278 (100)
	8	0.2	0	0.4	2	2	4.140 (119)	3.995 (115)	3.481 (100)
	9	0.2	0.1	0.4	2	2	4.070 (135)	3.997 (132)	3.022 (100)
0.001	1	0.2	0.1	0	0	0	0.7532 (309)	0.2439 (100)	0.2463 (101)
	2	0.2	0	0.4	0	0	0.7865 (256)	0.3834 (125)	0.3071 (100)
	3	0.2	0.1	0.4	0	0	0.7561 (262)	0.3947 (137)	0.2885 (100)
	4	0	0	0.4	0	0	0.7134 (242)	0.3840 (131)	0.2936 (100)
	5	0	0	0	2	0	0.7609 (266)	0.4189 (146)	0.2865 (100)
	6	0	0	0	0	2	0.7806 (227)	0.6815 (198)	0.3440 (100)
	7	0	0	0.4	2	2	0.7406 (218)	0.6744 (199)	0.3392 (100)
	8	0.2	0	0.4	2	2	0.7731 (201)	0.7614 (198)	0.3849 (100)
	9	0.2	0.1	0.4	2	2	0.7572 (216)	0.7073 (202)	0.3501 (100)

and Sowers (2000) who used the full (nonnatural) basis for a cubic smoothing spline and applied periodic constraints. Clearly, fitting a full additional L-spline would lead to identifiability problems. We resolve this by fitting a single set of core functions with two sets of basis functions: the natural L-spline functions derived earlier, and a set derived from the full basis but constrained to be periodic. For $L = D^4 + \omega^2 D^2$ with $m = 4$ and p knots $(t_1, \ldots, t_p)'$ within the period $[0, T]$, $T = 2\pi/\omega$, we construct the periodic component from a set of p basis functions $k_2(t_i; t)$, defined using equation (3), augmented by the m noncore global basis functions $l_1(t) = t^2$, $l_2(t) = t^3$, $l_3(t) = t \cos(\omega t)$, $l_4(t) = t \sin(\omega t)$. Define

$$g_d(t) = \sum_{i=1}^{p} d_i k_2(t_i, t) + \sum_{j=1}^{m} d_{p+j} l_j(t),$$

with associated penalty matrix $\boldsymbol{K}_a$ constructed such that for $\boldsymbol{d} = (d_1, \ldots, d_{p+m})'$

$$\boldsymbol{d}'\boldsymbol{K}_a\boldsymbol{d} = \int_0^T [Lg_d(s)]^2\, ds.$$

Periodic constraints are applied as $\boldsymbol{G}'\boldsymbol{d} = \boldsymbol{0}$ where $\boldsymbol{G}$ is a $(p + m) \times (2m - 1)$ matrix

$$\boldsymbol{G}_{ij} = \begin{cases} k_2^{(j-1)}(t_i, T) - k_2^{(j-1)}(t_i, 0) & i = 1, \ldots, p, \\ l_{p+i}^{(j-1)}(T) - l_{p+i}^{(j-1)}(0), & i = 1, \ldots, m, \end{cases}$$

for j=1, ..., $2m$−1 with $2m - 1$ constraints applied to ensure the same degree of continuity as in the underlying basis functions. It follows that $\boldsymbol{d} = \boldsymbol{C}_p\boldsymbol{v}$ for a $(p + m) \times (p - m + 1)$ matrix $\boldsymbol{C}_p$ such that $\boldsymbol{G}'\boldsymbol{C}_p = \boldsymbol{0}$. The model (6) is extended to include a periodic component

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Z}_t\boldsymbol{\delta} + \boldsymbol{Z}_p\boldsymbol{v} + \boldsymbol{e},$$

with $\boldsymbol{Z}_p = \boldsymbol{K}_{xp}\boldsymbol{C}_p$, where $\boldsymbol{K}_{xp}$ evaluates the $p + m$ noncore basis functions at the covariate values, and $\boldsymbol{v} \sim N(\boldsymbol{0}, \sigma_p^2\boldsymbol{H}_{pp})$ for $\boldsymbol{H}_{pp}^{-1} = \boldsymbol{C}_p'\boldsymbol{K}_a\boldsymbol{C}_p$ with cov $(\boldsymbol{v}, \boldsymbol{\delta}) =$ cov $(\boldsymbol{v}, \boldsymbol{e}) = \boldsymbol{0}$.

5.1 *Evaluation of Extended Model with Periodic Spline Component*

The simulation study of Section 4.2 was modified to investigate the performance of the periodic spline component, again using curves of relevance to the log(FOO) data from the grassland experiment. Using the notation of Section 4.2, curves were generated as

$$\begin{aligned} c_{jk}(t) &= s_{1jk}(t) + (s_{2jk}(t) - 0.03)(t - 30.5) \\ &\quad + (0.5 + s_{3jk}(t)) \sin(\omega t) \; + \; r_{6j} \sin(2\omega t), \end{aligned} \tag{8}$$

where $\omega = 2\pi/12$ for $j = 1, \ldots, 6$, $k = 1, \ldots, 10$. The functions $s_{1jk}(t)$, $s_{2jk}(t)$ used $r_{1j} = 0.2$, $r_{2j} = 0.1$ for all j to provide smooth nonlinear trend. The amplitude of the underlying periodic cycle was either kept constant ($r_{3j} = 0$) or allowed to vary ($r_{3j} = 0.4$). Extra periodic components were added to the curve as $\sin(2\omega t)$ to mimic the extra periodic pattern found in the log(FOO) data, and as a component orthogonal to the underlying periodic function. The effect of adding this term was assessed using coefficient $r_{6j} = 0$, 0.2, or 0.3. For each curve $c_{jk}(t)$, 500 sets of data were generated for σ^2 equal to 0.01 or 0.001. Each data set was modeled using the splines assessed in Section 4.2 and an L-spline with additional periodic L-spline, both using $L = \omega^2 D^2 + D^4$. Table 2 shows the average MSEP for each set of parameters. In every case where $r_{6j} > 0$, the MSEP for the model including the periodic spline was substantially smaller than for the other models. For $j = 3 - 6$ with $\sigma^2 = 0.01$ both the PCS and L-spline had very large MSEP values. The MSEP value for the L-spline with $j = 3$ and $\sigma^2 = 0.001$ was also very large. In these cases, the splines allocated the extra periodic term as noise rather than signal. In general, as the periodic frequency of deviations about the underlying form increased or as their amplitude decreased, it was increasingly likely that the deviations would be attributed as noise rather than fitted as smooth trend (signal). Different splines vary in the point at which the deviations are deemed to change from noise to signal. The L-spline was slower to detect the $\sin(2\omega t)$ component than the PCS as the

Table 2

Average MSEP across 5000 *data sets* (*model* 8) *for a cubic smoothing spline* (*CSS*), *partial cubic spline* (*PCS*), *L-spline, and L-spline with additional periodic spline* (*LP-spline*). *Average MSEP values are also shown as a percentage of the minimum value for each row.*

				Average MSEP × 1000 (% of minimum value)			
σ^2	j	r_{3j}	r_{6j}	CSS	PCS	L-spline	LP-spline
0.01	1	0	0	3.95 (220)	1.80 (100)	2.08 (116)	2.17 (121)
	2	0.4	0	4.00 (163)	3.05 (124)	2.46 (100)	2.53 (103)
	3	0	0.2	5.30 (208)	19.13 (751)	21.11 (829)	2.55 (100)
	4	0.4	0.2	5.32 (180)	17.21 (581)	22.02 (743)	2.96 (100)
	5	0	0.3	6.24 (253)	30.45 (1236)	45.66 (1854)	2.46 (100)
	6	0.4	0.3	6.28 (209)	14.46 (481)	47.85 (1590)	3.01 (100)
0.001	1	0	0	0.752 (311)	0.242 (100)	0.248 (103)	0.256 (106)
	2	0.4	0	0.770 (277)	0.384 (139)	0.277 (100)	0.285 (103)
	3	0	0.2	0.991 (322)	0.967 (314)	3.309 (1075)	0.308 (100)
	4	0.4	0.2	0.994 (284)	0.973 (278)	1.144 (327)	0.350 (100)
	5	0	0.3	0.999 (298)	0.999 (298)	0.642 (165)	0.335 (100)
	6	0.4	0.3	0.999 (272)	0.999 (272)	0.642 (150)	0.367 (100)

coefficient r_{6j} increased, but both performed badly in some cases. The presence of the periodic spline in the model guards against this problem.

The ability of the periodic spline to separate periodic deviations from long-term trend was evaluated by examining the periodic spline and long-term L-spline components of the fitted model, and residuals from the fitted model. For the long-term L-spline component and residuals, each series was detrended (using a smoothing spline with 4 degrees of freedom) and then the periodic component extracted by fitting a linear model with effects for the 12 months of the year. The sum of squares accounted for by month was used as a measure of periodic trend in each component. The variation in the periodic spline was measured by its total sum of squares. The total periodic variation was measured by the total of the three sums of squares. For $r_{6j} > 0$, the proportion of periodic variation accounted for by the periodic spline increased as r_{6j} increased and as σ^2 decreased, and was in the range 88–97%, with 1–8% accounted for in the long-term L-spline and 0–4% in the residual.

6. Analysis of the Grassland Experiment

The FOO data were analyzed using a logarithmic transformation to stabilize the variance of the data. The L-spline with core functions $\{1, t, \sin(\omega t), \cos(\omega t)\}$ was used to model long-term trend, with an additional set of periodic basis functions as described in Section 5. The structure of the experiment can be written symbolically (using the notation of Wilkinson and Rogers, 1973) as $(rep/plot) * sample$, where "$*$" is a crossing operator (i.e., $rep * sample = rep + sample + rep.sample$), "/" is a nesting operator (i.e., $rep/plot = rep + rep.plot$), "." denotes an interaction, *rep* indicates replicates in the design, *plot* indicates plots within replicates, and *sample* indicates the 32 sample dates. The treatment structure can be written as $trt * sample$, where *trt* labels the six fertilizer treatments. We wish to model the pattern across sample dates in terms of the underlying variable time, decomposing the term into linear trend $lin(t)$, sine/cosine periodic term $sin(\omega t)$, $cos(\omega t)$, L-spline deviations $lspl(t)$, and periodic L-spline deviations $pspl(t)$. To complete this decomposition, we need the residual (or lack of fit) term, *sample*, to represent pattern not accounted for by other terms, which can be separated from random variation due to the replicate plots measured at each sample date. The lack-of-fit term was fitted as a random term with independent errors, analogous to the usual lack-of-fit term in a regression model. The full model is shown in Table 3. In the bottom stratum, Rep.Plot.Sample, residual error could be considered as a composite of variation due to intrinsic plot effects on pasture growth, which would be correlated over time, and random measurement error. This composite structure was decomposed into separate independent and correlated components, with correlation within plots between samples d days apart modeled as ϕ^d, with a common parameter $0 < \phi < 1$ estimated across all plots. Eight values were chosen to represent typical sample dates within years. These eight points were used as knots for the periodic spline, and the corresponding 32 dates across the 4 years were used as knots for the nonperiodic L-spline.

Table 3

Structural decomposition showing full model for log(FOO) data, terms indicated as fixed (F) or random (R). Degrees of freedom (DF) shown for each stratum and for fixed terms. Scaled estimated variance components are shown for random terms retained in the final model.

Stratum Term	Type	DF	In final model? Yes/No	Variance component
mean	F	1	Y	n/a
Rep		1		
rep	R[a]		Y	0.0583
Rep.Plot		10		
trt	F	5	Y	n/a
rep.plot	R[b]		Y	0.0190
Sample		31		
$lin(t)$	F	1	Y	n/a
$sin(\omega t)$	F	1	Y	n/a
$cos(\omega t)$	F	1	Y	n/a
$lspl(t)$	R		Y	0.0617
$pspl(t)$	R		Y	0.0201
sample	R		Y	0.0307
Rep.Sample		31		
$rep.lin(t)$	R[a]		Y	0.0097
$rep.sin(\omega t)$	R		N	–
$rep.cos(\omega t)$	R[a]		Y	0.0039
$rep.lspl(t)$	R		N	–
$rep.pspl(t)$	R		N	–
rep.sample	R		Y	0.0040
Rep.Plot.Sample		372		
$trt.lin(t)$	F	6	N	n/a
$trt.sin(\omega t)$	F	6	Y	n/a
$trt.cos(\omega t)$	F	6	Y	n/a
$trt.lspl(t)$	R		N	–
$trt.pspl(t)$	R		N	–
trt.sample	R		Y	0.0042
$rep.plot.lin(t)$	R[b]		Y	0.0063
$rep.plot.sin(\omega t)$	R		N	–
$rep.plot.cos(\omega t)$	R		N	–
$rep.plot.lspl(t)$	R		Y	0.0061
$rep.plot.pspl(t)$	R		N	–
rep.plot.sample	R		N	–
rep.plot.exp (sample)	R[c]		Y	0.0088

[a] Indicates correlated terms fitted using rank 1 model.

[b] Indicates correlated terms fitted using unstructured model.

[c] exp(·) indicates exponential correlation model.

Model selection started with the full model shown in Table 3. Variance components relating to random terms were individually assessed for significance using REML likelihood ratio tests. Although a score test or simulation to determine the distribution under the null hypothesis would be more appropriate, computational requirements for this complex model made this difficult. Instead, we took a less exact approach based on the Stram and Lee (1994) approximation, and dropped nine variance parameters, which were close to zero (combined change of 0.72 in logRL), from the model and retained the remaining variance parameters. This process resulted in the final random model shown in Table 3. Changes in logRL achieved by dropping individual terms from the final model are shown in Table 4. Additional correlation parameters were required between related random regression terms to ensure invariance to change of origin in the time variable. As

Table 4

LogRL for models for log(FOO) data. The final fixed/random models refer to Table 3. *LogRL for model* 14 *is not comparable with other models as a different set of fixed terms was fitted.*

Model	Terms included in model	LogRL	Number of variance parameters	Change in logRL from model (2)
1	All fixed and random terms in Table 3	466.99	23	0.72
2	All fixed terms, final random model	466.27	14	–
3	Model (2) without term *pspl*(*t*)	463.04	13	−3.23
4	Model (2) without term *rep.plot.lspl*(*t*)	460.48	13	−5.78
5	Model (2) without term *lspl*(*t*)	463.71	13	−2.56
6	Model (2) without term *trt.sample*	456.13	13	−10.14
7	Model (2) without term *rep.sample*	449.04	13	−17.23
8	Model (2) without term *sample*	450.76	13	−15.51
9	Model (2) without term *rep*	462.12	13	−4.15
10	Model (2) without term *rep.lin*(*t*)	463.79	13	−2.48
11	Model (2) without term *rep.cos*(*t*)	458.45	13	−7.82
12	Model (2) without term *rep.plot*	452.82	12	−13.45
13	Model (2) without term *rep.plot.lin*(*t*)	453.13	12	−13.14
14	Final fixed and random models	484.27	14	–

there were only two reps, a full unstructured model could not be fitted between terms *rep*, *rep.lin*(*t*), and *rep.cos*(ωt), but a rank 1 model with correlation 1 increased logRL compared to independent terms. A correlation of 0.89 was estimated between the plot intercept (*rep.plot*) and linear (*rep.plot.lin*(*t*)) terms. The REML estimate of $\hat{\phi} = 0.997$ was equivalent to a correlation of 0.35 between samples taken 45 days apart from the same plot. Once the final variance model was established, the fixed model was simplified using Wald tests in a backward selection strategy, respecting marginality. The final fixed model allowed for a separate intercept and periodic pattern for each fertilizer treatment, with a common linear trend. GenStat (Payne, 2003) and ASREML (Gilmour et al., 2002) were used for the analysis.

The REML estimates of the variance components are shown in Table 3 multiplied by the mean value of the diagonal of the matrix $\boldsymbol{Z}_i\boldsymbol{Z}_i'$ in order to indicate the average contribution of the term to the variance of a data value. Variation due to the random regressions on *rep*/(*lin*(*t*) + *cos*(*t*)) and *rep.plot*/*lin*(*t*) was relatively large due to differences between replicates (see Figure 1). As the stocking rate was the same for both replicates, the replicate with less initial FOO would be expected to show a greater decline, and hence positive correlation between intercept and linear terms. As there were also initial differences in available pasture between replicate plots for each treatment, and as stocking rate was changed on plots according to the treatment, a similar correlation structure with highly correlated intercept and linear terms was observed for plots within replicates. The variance components corresponding to overall spline and lack-of-fit terms also made a significant contribution to the variation of the data compared to the residual component. The presence of the periodic L-spline (*pspl*(*t*)) indicated a common seasonal pattern of deviations about the underlying model for all treatments. However, the nonperiodic L-spline (*lspl*(*t*)) and lack-of-fit (*sample*) terms had larger variance components, indicating substantial nonperiodic smooth trend over time with some lack of fit. There was no evidence that separate splines were required for different fertilizer treatments, although the small *rep.plot.lspl*(*t*) term indicated differences in long-term patterns between plots. There was also some lack of fit to individual treatment profiles, indicated by the *trt.sample* term.

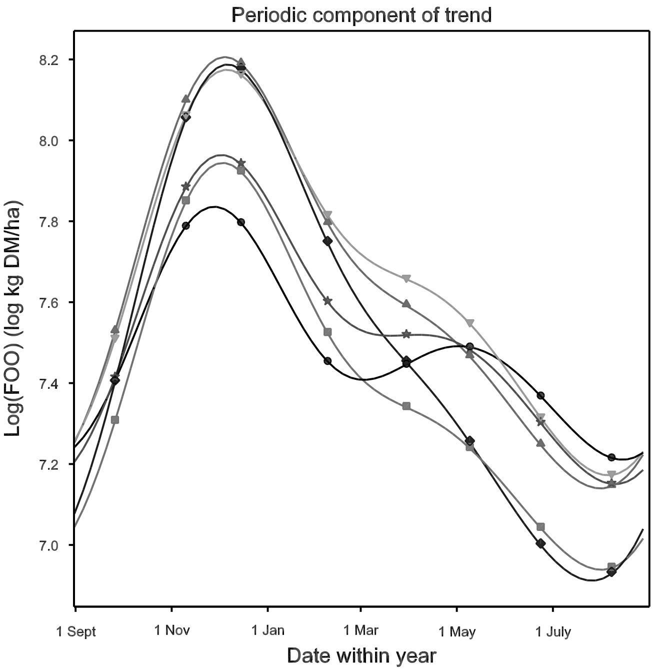

Figure 2. Total periodic component of final model (Table 3), with knot points marked, for each treatment: nil (●), low super (⋆), medium super (■), high super (♦), high super with lime (▲), sewage ash (▼). Average prediction error for treatment differences at knots is 0.088, minimum = 0.086, maximum = 0.091.

Figure 1 presents the log-transformed data with fitted curves calculated by excluding all lack-of-fit terms and the residual, as these terms represent unpredictable noise. There was good agreement between the data and the fitted curves. The nonperiodic trend over time did not depend on treatment and showed a steady decrease in FOO over time after the first year. Figure 2 shows the periodic component of the trend for each treatment, calculated from the treatment-specific constant, sine/cosine, and common periodic L-spline terms. Differences in phase and amplitude arise solely from the fixed terms in the model and it is clear that amplitude increased with superphosphate levels. However, applying superphosphate with lime or sewage ash appeared to provide extra herbage mass during the autumn and winter (May–August). The deviation away from the underlying sine/cosine shape was introduced by the periodic L-spline component. The main influence of the periodic L-spline was to sharpen the main peak during November and December, and to introduce some regrowth in May and June. This introduced an extra bump into the pattern for lower levels of superphosphate, and a reduced rate of decrease for higher levels. This regrowth was present in the raw data (Figure 1) but could not be modeled by the underlying fixed model.

7. Discussion

We have shown that L-splines with linear parameters in the core functions can be implemented within mixed models in a similar manner to cubic smoothing splines, and can be used to give ANOVA-type decompositions of the treatment structure. The extension to a periodic L-spline term is helpful in giving a familiar decomposition into periodic and nonperiodic components of the trend. The model with the periodic L-spline allows an explicit prediction of the shape of the periodic component; in the grassland example this can be used to predict the pattern of log(FOO) in an average year.

Heckman and Ramsay (2000) showed that use of the appropriate L-spline required a smoother with fewer degrees of freedom than a cubic spline, and Ramsay and Dalzell (1991) argued that the partitioning of the function space into orthogonal subspaces (with respect to the inner product) was in itself desirable. Our small simulation showed that a mixed model L-spline with core functions accounting for a linear trend with a periodic cycle can perform better than either cubic splines or PCS in terms of squared error of prediction when there are perturbations in the periodic cycle. Further work is required to establish the full range of conditions under which mixed model L-splines could be recommended.

Acknowledgements

We gratefully acknowledge funding from the Grains Research and Development Corporation of Australia for this work, and thank D. L. Garden for permission to use the data from the grassland experiment and advice on the interpretation of the analysis.

References

Brumback, B. A. and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association* **93,** 961–994.

Crainiceanu, C. M. and Ruppert, D. (2004). Restricted likelihood ratio tests in nonparametric longitudinal models. *Statistica Sinica* **12,** 713–729.

Crainiceanu, C. M., Ruppert, D., Claeskens, G., and Wand, M. P. (2005). Exact likelihood ratio tests for penalized splines. *Biometrika* **92,** 91–103.

Currie, I. D. and Durban, M. (2002). Flexible smoothing with P-splines: A unified approach. *Statistical Modelling* **4,** 333–349.

Dalzell, C. J. and Ramsay, J. O. (1993). Computing reproducing kernels with arbitrary boundary constraints. *SIAM Journal of Scientific Computing* **14,** 511–518.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11,** 89–121.

Garden, D. L., Ellis, N. J. S., Rab, A., Langford, C. M., Johnston, W. H., Shields, C., Murphy, T., Holmberg, M., Dassanayake, K. B., and Harden, S. (2003). Fertiliser and grazing effects on production and botanical composition of native grasslands in southeast Australia. *Australian Journal of Experimental Agriculture* **43,** 843–859.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2002). *ASREML User Guide Release 1.0.* Hemel Hempstead, U.K.: VSN International Ltd.

Gu, C. (2002). *Smoothing Spline ANOVA Models.* New York: Springer-Verlag.

Guo, W. (2002a). Functional mixed effects models. *Biometrics* **58,** 121–128.

Guo, W. (2002b). Inference in smoothing spline analysis of variance. *Journal of the Royal Statistical Society, Series B* **64,** 887–898.

Heckman, N. E. and Ramsay, J. O. (2000). Penalized regression with model-based penalties. *Canadian Journal of Statistics* **28,** 241–258.

Parise, H., Wand, M. P., Ruppert, D., and Ryan, L. (2001). Incorporation of historical controls using semi-parametric mixed models. *Applied Statistics* **50,** 31–42.

Payne, R. W., ed. (2003). *The Guide to Genstat, Part 2: Statistics.* Hemel Hempstead, U.K.: VSN International Ltd.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B* **53,** 539–572.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis.* New York: Springer-Verlag.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11,** 735–757.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects setting. *Biometrics* **50,** 1171–1177.

Verbyla, A. P., Cullis, B. R., Kenward, M. G., and Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Applied Statistics* **48,** 269–311.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics* **18,** 223–249.

Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society, Series B* **60,** 159–174.

Wilkinson, G. N. and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics* **22,** 392–399.

Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4,** 57–74.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93,** 710–719.

Zhang, D., Lin, X., and Sowers, M. (2000). Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics* **56,** 31–39.

Received December 2003. *Revised September* 2005.
Accepted September 2005.