Aust. N. Z. J. Stat. **49**(1), 2007, 1–23 doi: 10.1111/j.1467-842X.2006.00454.x

A COMPARISON OF MIXED MODEL SPLINES FOR CURVE FITTING

SUE J. WELHAM[1*], BRIAN R. CULLIS[2], MICHAEL G. KENWARD[3] AND ROBIN THOMPSON[1]

Rothamsted Research, Wagga Wagga Agricultural Institute and London School of Hygiene and Tropical Medicine

Summary

Three types of polynomial mixed model splines have been proposed: smoothing splines, P-splines and penalized splines using a truncated power function basis. The close connections between these models are demonstrated, showing that the default cubic form of the splines differs only in the penalty used. A general definition of the mixed model spline is given that includes general constraints and can be used to produce natural or periodic splines. The impact of different penalties is demonstrated by evaluation across a set of functions with specific features, and shows that the best penalty in terms of mean squared error of prediction depends on both the form of the underlying function and the signal:noise ratio.

Key words: best linear unbiased prediction, mixed models, penalized splines, P-splines, residual maximum likelihood, smoothing splines.

1. Introduction

In recent years, smoothing splines have been introduced into the linear mixed model framework with the smoothing parameter estimated using residual (restricted) maximum likelihood (REML, Patterson & Thompson, 1971). This formulation enables flexible non-linear curves to be easily fitted in the presence of random effects and correlated errors. Several different forms of mixed model spline have been suggested. Verbyla *et al.* (1999), Brumback & Rice (1998), Zhang *et al.* (1998) and Wang (1998) all developed models using cubic smoothing splines. Eilers & Marx (1996) defined the class of P-splines, based on reduced-knot B-spline bases with discrete differencing penalties, motivated as a computationally efficient approximation to the cubic smoothing spline model, and this was set into the mixed model framework by Eilers (1999) and Currie & Durban (2002). Brumback, Ruppert & Wand (1999) and Ruppert, Wand & Carroll (2003) described penalized spline models, based on reduced-knot truncated power function bases with penalties on the untransformed coefficients, fitted as a mixed model, and motivated as a simple low-rank smoothing spline. There are close connections between the three types of model. This paper reviews these spline methods, shows the connections between the different models, and sets them within a unified framework that

*Author to whom correspondence should be addressed.

[1]Biomathematics and Bioinformatics, Rothamsted Research, Harpenden AL5 2JQ, UK.
e-mail: sue.welham@bbsrc.ac.uk

[2]Wagga Wagga Agricultural Institute, Wagga Wagga, NSW 2650, Australia.

[3]Medical Statistics Unit, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK.

Acknowledgments. SJW, BRC and RT acknowledge funding by the Australian Grains Research and Development Corporation.

allows for general constraints. In addition, we examine their behaviour over a contrasting set of functions to investigate the impact of different penalties on the fitted curve.

2. Mixed model polynomial splines

In the simplest case, data $\boldsymbol{y} = (y_1, \ldots, y_n)$ are to be modelled in terms of an explanatory variable $\boldsymbol{x} = (x_1, \ldots, x_n)$, which takes $p \leq n$ distinct values $\kappa_1 < \kappa_2 < \cdots < \kappa_p$ within a defined range $[a, b]$. The model for the data is

$$\boldsymbol{y} = g(\boldsymbol{x}) + \boldsymbol{e}$$

for some unknown smooth function $g(x)$ that will be estimated using a polynomial spline, and a vector of random errors, $\boldsymbol{e}$, with $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}(\boldsymbol{\phi}))$ for some variance-covariance matrix $\boldsymbol{R}$ which is a function of unknown parameters $\boldsymbol{\phi}$.

A polynomial spline $g(x)$ of degree k for $a \leq x \leq b$, with r knots $a < t_1 < \cdots < t_r < b$, is a piecewise polynomial of degree $\leq k$ on each interval $[t_j, t_{j+1}]$, $j = 1, \ldots, r-1$, $[a, t_1]$ and $[t_r, b]$, with derivatives up to order $k-1$ continuous for $x \in [a, b]$. Natural polynomial splines of degree $k = 2m - 1$ for $m \in \mathbb{Z}^+$ have derivatives of order $\geq m$ constrained to be zero for $x = a$ and $x = b$. A set of basis functions for $g(x)$ is then a set of polynomial splines of degree k, $s_j(x)$, $j = 1, \ldots, q$, with the specified knot points such that

$$g(x) = \sum_{j=1}^{q} c_j s_j(x)$$

for suitably chosen coefficients $c_1, \ldots, c_q$. For polynomial splines of degree k with r knots, this requires $q = r + k + 1$ basis functions, as the spline $g(x)$ uses $(r+1)(k+1)$ coefficients with rk constraints. A natural spline has an additional $2(k - m + 1) = k + 1$ constraints, leaving $q = r$ basis functions.

For the definition of some bases considered later, additional knots $t_{-k}, \ldots, t_0$ and $t_{r+1}, \ldots, t_{r+k+1}$ are required such that $t_j < t_{j+1}$ for $j = -k, \ldots, r + k + 1$ with $t_0 \leq a$ and $t_{r+1} \geq b$. The position of the additional knots is arbitrary subject to these conditions.

2.1. Polynomial smoothing splines

The polynomial smoothing spline of degree $k = 2m - 1$ is the function $g(x)$ that minimizes the penalized sum of squares

$$S_k(g) \;=\; (\boldsymbol{y} - g(\boldsymbol{x}))^\top \boldsymbol{R}^{-1} (\boldsymbol{y} - g(\boldsymbol{x})) \;+\; \lambda \int_a^b \left(g^{(m)}(x)\right)^2 dx \tag{1}$$

for a given value of the smoothing parameter λ. It can be shown that this function is a natural polynomial spline with knots at the p distinct values of the explanatory variable, i.e. $r = p$, $t_j = \kappa_j$ for $j = 1, \ldots, p$ (Green & Silverman, 1994, Chapter 2). Minimization of this criterion balances deviations of the fitted spline from the data against the roughness of the fitted spline, where roughness is quantified by the mth derivative of the function. The balance between the two components is controlled by the smoothing parameter λ.

The cubic smoothing spline was used within the linear model as a method of fitting flexible data-driven functions by Wahba (1984) and Green (1985) and was extended to the generalized linear model as generalized additive models by Hastie & Tibshirani (1986), with the smoothing parameter estimated by generalized cross validation (GCV) or specification of the spline's effective degrees of freedom. Models based on these methods have become widespread in practical data analysis, see e.g. Hastie & Tibshirani (1990) or, more recently, Hastie, Tibshirani & Friedman (2001).

The natural polynomial spline of degree k can be written in terms of a basis

$$g(x) \;=\; \sum_{j=0}^{m-1} \tau_{s,j} x^j \;+\; \sum_{j=1}^{r-m} \nu_j Q_j(x) \tag{2}$$

where $\{Q_j(x)\,;\, j=1,\ldots,r-m\}$ is a set of natural polynomial splines, defined later in (9). The model can thus be regarded as an underlying polynomial trend of degree $m-1$ (with zero penalty) plus smooth piecewise polynomial deviations (of degree k) about the trend (with non-zero penalty). In matrix form, this model becomes

$$g(\boldsymbol{x}) \;=\; \boldsymbol{X}_s\boldsymbol{\tau}_s + \boldsymbol{Z}_s\boldsymbol{\nu}$$

where $\boldsymbol{X}_s = [\boldsymbol{1},\ \boldsymbol{x}\ ,\ldots,\ \boldsymbol{x}^{m-1}]$, with unknown coefficients $\boldsymbol{\tau}_s = (\tau_{s,0},\ \tau_{s,1},\ldots,\tau_{s,m-1})$, and $\boldsymbol{Z}_s = [Q_1(\boldsymbol{x}),\ldots,Q_{r-m}(\boldsymbol{x})]$ is a design matrix for the basis functions $Q_j(x)$ evaluated at the covariate values $\boldsymbol{x}$, with unknown coefficients $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_{r-m})$. The penalty can be written as

$$\int_a^b \left(g^{(m)}(x)\right)^2 dx \;=\; \boldsymbol{\nu}^\top \boldsymbol{H}_s^{-1}\boldsymbol{\nu}$$

for a matrix $\boldsymbol{H}_s^{-1}$ with $(i,\,j)$th element defined by

$$\left[\boldsymbol{H}_s^{-1}\right]_{ij} \;=\; \int_a^b Q_i^{(m)}(x)Q_j^{(m)}(x)dx.$$

The penalized sum of squares for the polynomial smoothing spline can then be expressed as

$$(\boldsymbol{y} - \boldsymbol{X}_s\boldsymbol{\tau}_s - \boldsymbol{Z}_s\boldsymbol{\nu})^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}_s\boldsymbol{\tau}_s - \boldsymbol{Z}_s\boldsymbol{\nu}) \;+\; \lambda\boldsymbol{\nu}^\top\boldsymbol{H}_s^{-1}\boldsymbol{\nu}. \tag{3}$$

For a cubic smoothing spline, this simplifies to a model with $\boldsymbol{X}_s = [\boldsymbol{1}\ \ \boldsymbol{x}]$, $\boldsymbol{Z}_s = 6[P_2(\boldsymbol{x}),\ldots,P_{r-1}(\boldsymbol{x})]$, for $P_j(x)$ defined by

$$6P_j(x) = \begin{cases} 0 & x \le t_{j-1} \\ h_{j-1}^{-1}(x-t_{j-1})^3 & t_{j-1} < x \le t_j \\ h_{j-1}^{-1}(x-t_{j-1})^3 - \left(h_{j-1}^{-1} + h_j^{-1}\right)(x-t_j)^3 & t_j < x \le t_{j+1} \\ (h_{j-1}+h_j)(3x - t_{j-1} - t_j - t_{j+1}) & t_{j+1} < x \end{cases},$$

where $h_j = t_{j+1} - t_j$, $j = 1,\ldots,r-1$. Then $\boldsymbol{\nu} = \boldsymbol{\delta}/6$, where $\boldsymbol{\delta}$ is the second derivative of the fitted spline at the internal knots $t_2,\ldots,t_{r-1}$ and $36\boldsymbol{H}_s = \boldsymbol{G}_s$, where the rows and columns of $\boldsymbol{G}_s$ are labelled $2, \ldots, r-1$, and $\boldsymbol{G}_s$ is defined by

$$[\boldsymbol{G}_s^{-1}]_{ij} = \begin{cases} \frac{1}{6}h_{i-1} & j = i-1, \quad i = 3, \ldots, r-1 \\ \frac{1}{3}(h_i + h_{i-1}) & j = i, \quad i = 2, \ldots, r-1 \\ \frac{1}{6}h_i & j = i+1, \quad i = 2, \ldots, r-2 \\ 0 & \text{otherwise.} \end{cases}$$

For a given value of the smoothing parameter, the linear mixed model

$$\boldsymbol{y} \; = \; \boldsymbol{X}_s\boldsymbol{\tau}_s + \boldsymbol{Z}_s\boldsymbol{v} + \boldsymbol{e},$$

with $\boldsymbol{v} \sim N(\boldsymbol{0}, \sigma_s^2\boldsymbol{H}_s)$ and $\lambda = \sigma^2/\sigma_s^2$, has the property that the polynomial smoothing spline is the best linear unbiased predictor (BLUP) $\tilde{g}(\boldsymbol{x}) = \boldsymbol{X}_s\hat{\boldsymbol{\tau}}_s + \boldsymbol{Z}_s\tilde{\boldsymbol{v}}$ from the fitted model. If both terms were fitted as fixed effects, the spline model would interpolate the data. Because the deviations are fitted as random effects, they are shrunk to give a smooth curve with the degree of shrinkage determined by the smoothing parameter. The smoothing parameter can also be estimated by REML via the variance component σ_s^2. REML estimation then selects the smoothing parameter value that gives the best fit to the implicit variance model defined by the basis functions and penalty matrix $\boldsymbol{H}_s$.

Thompson (1985) pointed out that the cubic smoothing spline could be fitted as a mixed model using REML estimation of the smoothing parameter, but did not evaluate the method. Speed (1991) noted that, for a given value of λ, the cubic smoothing spline is a BLUP from a linear mixed model with fixed intercept and linear terms plus a set of random spline basis functions with covariance matrix $\sigma_s^2\boldsymbol{G}_s$, and showed that REML estimation of the smoothing parameter was equivalent to the generalized maximum likelihood (GML) method of Wahba (1985). Cubic smoothing splines were studied in detail as mixed models by Verbyla *et al.* (1999), Brumback & Rice (1998), Zhang *et al.* (1998) and Wang (1998).

Because the penalty is designed to reflect desirable properties of the fitted curve, it does not necessarily follow that this will translate into an appropriate variance matrix for the data. The performance of REML estimation of the smoothing parameter may therefore be questionable. Wahba (1990) stated that splines with the smoothing parameter estimated using GML over-smoothed compared to splines with the smoothing parameter estimated by generalized cross-validation (GCV). Kohn, Ansley & Tharm (1991) reported an extensive simulation study that compared the performance of GCV and GML in terms of mean squared error of prediction (MSEP) in the estimation of a set of smooth functions across a large range of sample sizes and error variances, using both cubic and quintic smoothing splines. Their results showed that GML performed similarly in terms of MSEP to GCV over the full range of functions tested, even when the tails of the error distribution were either heavier or lighter than expected for the normal distribution. Similar results were reported from a smaller study by Ruppert *et al.* (2003, section 5.4). These results give reassurance that REML estimation of the smoothing parameter in mixed model splines performs adequately compared to other standard techniques. Smoothing splines can then be fitted within a general linear mixed model.

The polynomial smoothing spline requires a knot at each distinct covariate value, leading to computational difficulties when p is large. A reduced-knot form follows from the definition above for any specified set of r distinct knots with $a < t_1 < \cdots < t_r < b$. For a given value of the smoothing parameter, the reduced-knot form is no longer the function that gives an overall minimum of the penalized sum of squares criterion (1), although the fitted spline $\tilde{g}(\boldsymbol{x})$ minimizes the penalized sum of squares (3) for $\boldsymbol{Z}_s$ and $\boldsymbol{H}_s$ defined in terms of the specified knots. The difference between the full- and reduced-knot splines depends on the density and location of the knots used. The reduced-knot version can be considered as a penalized natural polynomial spline using a basis of monomials x^j, $j = 0, \dots, m-1$, with Q_j, $j = 1, \dots, r-m$, and penalty matrix $\boldsymbol{H}_s^{-1}$.

2.2. P-splines

P-splines were introduced by Eilers & Marx (1996) as a simple and computationally efficient alternative to smoothing splines. The computational efficiency comes from the use of both a reduced-knot B-spline basis, following work by Parker & Rice (1985) and O'Sullivan (1986, 1988), and a discrete penalty specified in terms of differencing operators. The user has the choice of number of knots (r), degree of B-spline basis (k) and order of differencing in the penalty (d). The knots are usually taken to be equally spaced. In contrast with smoothing splines, the B-spline basis is not constrained to be natural. The P-spline model is written as

$$g(x) = \sum_{j=-k}^{r} \alpha_j B_j(x)$$

for $q = r + k + 1$ unknown coefficients $\alpha_{-k}, \dots, \alpha_r$ and a full set of q B-spline basis functions $B_j(x)$, $j = -k, \dots, r$, of degree k for knots $t_1, \dots, t_r$, as defined in (8). The full augmented knot set $t_{-k}, \dots, t_{r+k+1}$ is required for definition of the basis functions. In matrix form,

$$g(\boldsymbol{x}) = \boldsymbol{B\alpha}$$

where $\boldsymbol{B} = [B_{-k}(\boldsymbol{x}), \dots, B_r(\boldsymbol{x})]$ and $\boldsymbol{\alpha} = (\alpha_{-k}, \dots, \alpha_r)$. A P-spline model with dth order differencing is fitted by minimizing the penalized sum of squares

$$(\boldsymbol{y} - \boldsymbol{B\alpha})^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{B\alpha}) + \lambda \boldsymbol{\alpha}^\top \boldsymbol{\Delta}_d^\top \boldsymbol{\Delta}_d \boldsymbol{\alpha} \tag{4}$$

for a given value of λ, where the $(q-d) \times q$ matrix $\boldsymbol{\Delta}_d$ denotes the dth order differencing matrix for knots $t_{-k}, \dots, t_{r+d}$ defined in the Appendix. For example, for a cubic P-spline ($k = 3$) with second-order differencing ($d = 2$) and equally spaced knots such that $h = t_{j+1} - t_j$ for $j = -3, \dots, r+3$, the matrix $\boldsymbol{\Delta}_2$ is defined as

$$[\boldsymbol{\Delta}_2]_{ij} = \frac{1}{h} \begin{cases} 1 & j = i,\ i = 1, \dots, r+2 \\ -2 & j = i+1,\ i = 1, \dots, r+2 \\ 1 & j = i+2,\ i = 1, \dots, r+2 \\ 0 & \text{otherwise.} \end{cases}$$

Eilers & Marx (1996) motivated a penalty based on second order differences with a cubic B-spline basis ($k = 3$, $d = 2$) as an approximation to the cubic smoothing spline penalty. In their context, the sparsity of both the B-spline basis and the penalty gives a sparse matrix representation of the model, allowing efficient minimization of (4).

Eilers (1999) stated that P-splines could be fitted as linear mixed models, and details and examples were given by Currie & Durban (2002). Let

$$\boldsymbol{I} - \boldsymbol{\Delta}_d^\top \left(\boldsymbol{\Delta}_d \boldsymbol{\Delta}_d^\top\right)^{-1} \boldsymbol{\Delta}_d \;=\; \boldsymbol{L}\boldsymbol{L}^\top \;,$$

where the $q \times d$ matrix $\boldsymbol{L}$ has full column rank. The mixed model formulation for a cubic P-spline with a dth order difference penalty can be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{B}\boldsymbol{L}\boldsymbol{L}^\top\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\Delta}_d^\top\left(\boldsymbol{\Delta}_d\boldsymbol{\Delta}_d^\top\right)^{-1}\boldsymbol{\Delta}_d\boldsymbol{\alpha} + \boldsymbol{e} \\ &= \boldsymbol{X}_b\boldsymbol{\tau}_b + \boldsymbol{Z}_b\boldsymbol{u}_b + \boldsymbol{e}. \end{aligned} \tag{5}$$

The matrix $\boldsymbol{BL}$ is a linear transformation of a polynomial model of degree $d - 1$ on the vector $\boldsymbol{x}$ (Currie & Durban, 2002), and so there exists a full rank $d \times d$ matrix $\boldsymbol{A}$ such that $\boldsymbol{BLA} = \boldsymbol{X}_b = [\boldsymbol{1}\ \boldsymbol{x}, \ldots, \boldsymbol{x}^{d-1}]$. Then $\boldsymbol{\tau}_b = \boldsymbol{A}^{-1}\boldsymbol{L}^\top\boldsymbol{\alpha}$ is a vector of unknown coefficients of length d, $\boldsymbol{Z}_b = \boldsymbol{B}\boldsymbol{\Delta}_d^\top(\boldsymbol{\Delta}_d\boldsymbol{\Delta}_d^\top)^{-1}$, and $\boldsymbol{u}_b = \boldsymbol{\Delta}_d\boldsymbol{\alpha}$ is a vector of unknown coefficients of length $q - d$, with $\mathrm{var}(\boldsymbol{u}_b) = \sigma_s^2\boldsymbol{I}_{q-d}$. The quantity $\gamma_s^{-1}\boldsymbol{u}_b^\top\boldsymbol{u}_b$ is then equal to the P-spline penalty for $\lambda = \gamma_s^{-1} = \sigma^2/\sigma_s^2$. The order of differencing in the penalty determines the form of the fixed part of the model. For dth order differencing, the fixed part of the model represents a polynomial of degree $d - 1$. Unfortunately, the partition into fixed and random components means that the sparse representation of the penalized sum of squares cannot easily be utilized in the mixed model context.

2.3. Penalized splines using a truncated power function basis

Penalized splines based on a truncated power function (TPF) basis were introduced by Brumback *et al.* (1999) and Ruppert & Carroll (2000), and a full exposition in the mixed model setting was given by Wand (2003) and Ruppert *et al.* (2003). The penalized spline model of degree k with r knots $t_1, \ldots, t_r$ is written as

$$g(x) \;=\; \sum_{j=0}^{k} \tau_{T,j} x^j + \sum_{j=1}^{r} u_{T,j}(x - t_j)_+^k \;, \tag{6}$$

or in matrix terms as

$$g(\boldsymbol{x}) \;=\; \boldsymbol{X}_T\boldsymbol{\tau}_T + \boldsymbol{Z}_T\boldsymbol{u}_T \;,$$

where $\boldsymbol{X}_T = [\boldsymbol{1},\ \boldsymbol{x}\ , \ldots,\ \boldsymbol{x}^k]$ is the design matrix for a k-degree polynomial with $k+1$ unknown coefficients $\boldsymbol{\tau}_T = (\tau_{T,0}, \ldots, \tau_{T,k})$, and $\boldsymbol{Z}_T = [T_1(\boldsymbol{x}), \ldots, T_r(\boldsymbol{x})]$ is a design matrix of k-degree truncated power functions, $T_j(x) = (x - t_j)_+^k$, $j = 1, \ldots, r$, with r unknown coefficients $\boldsymbol{u}_T = (u_{T,1}, \ldots, u_{T,r})$. The penalized sum of squares to be minimized is

$$(\boldsymbol{y} - \boldsymbol{X}_T\boldsymbol{\tau}_T - \boldsymbol{Z}_T\boldsymbol{u}_T)^\top\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}_T\boldsymbol{\tau}_T - \boldsymbol{Z}_T\boldsymbol{u}_T) \;+\; \lambda\boldsymbol{u}_T^\top\boldsymbol{u}_T \;,$$

which corresponds to fitting the mixed model

$$\boldsymbol{y} \;=\; \boldsymbol{X}_T\boldsymbol{\tau}_T + \boldsymbol{Z}_T\boldsymbol{u}_T + \boldsymbol{e}$$

with $\mathrm{var}(\boldsymbol{u}_T) = \sigma_s^2\boldsymbol{I}_r$ for $\lambda = \sigma^2/\sigma_s^2$.

This model can be considered as an underlying polynomial function of degree k, plus smooth piecewise polynomial deviations, also of degree k. Ruppert & Carroll (2000) and Crainiceanu & Ruppert (2004) state that for penalized splines of degree k, the penalty

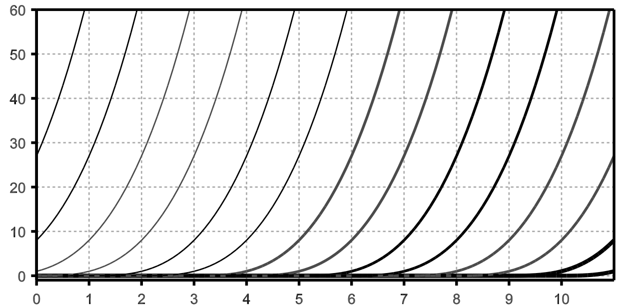

Figure 1. Truncated power function basis in the form of equation (7) for a cubic spline with knots 1, . . . ,10 on the range [0,11], with additional knots −3, −2, −1, 0 used to construct the basis.

$\boldsymbol{u}_T^\top \boldsymbol{u}_T$ is an approximation to a penalty on the $(k+1)$th derivative. Ruppert *et al.* (2003, Section 3.7) note that any form of penalty matrix could be used for the random effects within this framework, but in practice they usually work with an identity penalty matrix.

3. Mapping between different bases for polynomial splines

The connections between these three types of mixed model polynomial splines depend upon the relationships between the different spline bases. Some of these bases require the full augmented knot set $t_{-k}, \ldots, t_{r+k+1}$.

Consider the TPF basis of degree k $\{T_j(x);\ j = -k, \ldots, r\}$, where $T_j(x) = (x - t_j)_+^k$. A cubic TPF basis is shown in Figure 1.

Any polynomial spline of degree k with knots $t_1, \ldots, t_r$ in the range $[a, b]$ can be written in the form

$$g(x) = \sum_{j=-k}^{r} \beta_j T_j(x) \tag{7}$$

for suitable coefficients $\boldsymbol{\beta} = (\beta_{-k}, \ldots, \beta_r)$. The matrix $\boldsymbol{T} = [T_{-k}(\boldsymbol{x}), \ldots, T_r(\boldsymbol{x})]$ is used to represent this basis at the covariate values, and this is used here as the baseline parameterization of the polynomial spline.

The TPF basis is intuitively simple, but has bad computational properties in the sense that it can lead to ill-conditioned matrices that must be processed with care. One useful property of the TPF basis is that a knot at position t can be added to or dropped from the basis by adding or removing the corresponding basis function $(x - t)_+^k$, with the remaining basis functions unchanged.

Let $\boldsymbol{T}$ be partitioned as $[\boldsymbol{T}_1\, \boldsymbol{T}_2]$ where $\boldsymbol{T}_1 = [T_{-k}(\boldsymbol{x}), \ldots, T_0(\boldsymbol{x})]$ and $\boldsymbol{T}_2 = \boldsymbol{Z}_T$. The first $k+1$ basis functions are unconstrained polynomials of degree k for $a \le x \le b$ and can therefore be mapped onto the set of monomial functions $[1, x, \ldots, x^k]$ using a full rank transformation $\boldsymbol{A}_{xT}$ such that $[T_{-k}(\boldsymbol{x}), \ldots, T_0(\boldsymbol{x})] = [1,\ x, \ldots, x^k]\boldsymbol{A}_{xT}$. The matrix $\boldsymbol{A}_{xT}$ can be found for any set of knots $t_{-k}, \ldots, t_0$ by equating coefficients of x^j, $j = 0, \ldots, k$, across the two bases. This transformation provides a mapping from the full TPF basis onto the version (6) usually used in penalized splines, with

$$\begin{bmatrix} \boldsymbol{\tau}_T \\ \boldsymbol{u}_T \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{xT} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_r \end{bmatrix} \boldsymbol{\beta}.$$

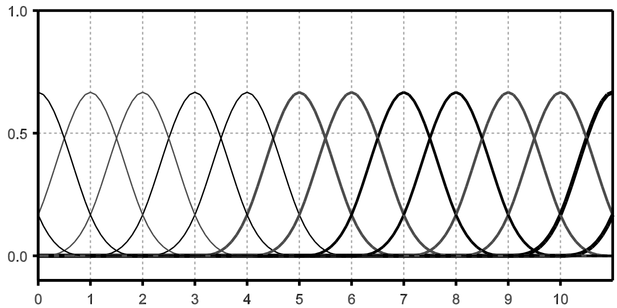

Figure 2. B-spline basis functions for cubic spline with knots $1, \ldots, 10$ on the range [0,11], with additional knots $-3, \ldots, 0$ and $11, \ldots, 14$ used to construct the basis.

A B-spline basis of degree k can be constructed by taking $(k+1)$th differences of the truncated power function basis of degree k (de Boor, 1978, chapter IX). A cubic B-spline basis is shown in Figure 2. More formally, the B-spline basis functions of degree k for knots $t_{-k}, \ldots, t_{r+k+1}$, are given by

$$B_j(x) \; = \; (-1)^{k+1}(t_{j+k+1} - t_j)[t_j, \ldots, t_{j+k+1}](x - \cdot)_+^k \, , \quad j = -k, \ldots, r \tag{8}$$

where $[t_j, \ldots, t_{j+d}]f$ denotes the dth divided difference of function f for points $t_j, \ldots, t_{j+d}$ and is defined in the Appendix. For $a \leq x \leq b$, $T_{r+j}(x) = 0$ for $j > 0$. Using results from the Appendix, with $q = r + k + 1$, it follows directly that

$$[B_{-k}(\boldsymbol{x}), \ldots, \; B_r(\boldsymbol{x})] = [T_{-k}(\boldsymbol{x}), \ldots, \; T_r(\boldsymbol{x})] \; \boldsymbol{D}_{k+1}$$

where the $q \times q$ matrix $\boldsymbol{D}_{k+1}$ is the $(k+1)$th order full rank differencing matrix for knots $t_{-k}, \ldots, t_{r+k+1}$. For the matrix $\boldsymbol{B} = [B_{-k}(\boldsymbol{x}), \ldots, \; B_r(\boldsymbol{x})]$, using the equivalence of the fitted spline across different bases

$$\boldsymbol{B\alpha} \; = \; \boldsymbol{T\beta},$$

then $\boldsymbol{D}_{k+1}\boldsymbol{\alpha} = \boldsymbol{\beta}$ and similarly $\boldsymbol{u}_{T} = \boldsymbol{\Delta}_{k+1}\boldsymbol{\alpha}$, where the $(q-d) \times q$ matrix $\boldsymbol{\Delta}_{k+1}$ is the $(k+1)$th order differencing matrix for knots $t_{-k}, \ldots, t_{r+k+1}$.

The B-spline basis functions are local, with $B_j(x) = 0$ for $x < t_j$ and $x > t_{j+k+1}$. The shape of each B-spline basis function depends on the knot positions, and adding or deleting a knot will change the $k+2$ members of the basis constructed using that knot. B-splines have the normalization property $\sum_j B_j(x) = 1$ for $a \leq x \leq b$. The fitted spline $\sum_j \alpha_j B_j(x)$ can therefore be considered as a weighted average of the coefficients $\alpha_{j-k}, \ldots, \alpha_j$ for $t_j < x < t_{j+1}$. The coefficients of a fitted B-spline model thus indicate the form of the fitted spline.

The natural polynomial spline in (2) was parameterized in terms of a polynomial of degree m plus a set of natural basis functions $Q_j(x)$, $j = 1, \ldots, r - m$, defined as

$$Q_j(x) \; = \; (-1)^m(t_{j+m} - t_j)[t_j, \ldots, t_{j+m}](x - \cdot)_+^k \, , \tag{9}$$

i.e. the mth divided differences of the TPF basis. This basis for a natural cubic spline is shown in Figure 3. To facilitate comparisons with non-natural splines, consider a related non-natural basis $\{Q_j(x); j = -k, \ldots, r\}$ for the augmented set of knots $t_{-k}, \ldots, t_{r+m+1}$. The polynomial

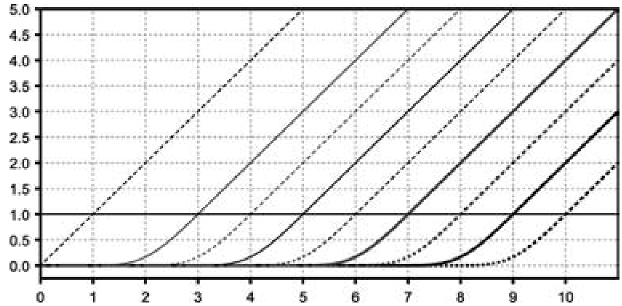

Figure 3. Basis for natural cubic splines derived as scaled second differences of the truncated power function basis with knots $1, \ldots, 10$ on the range [0,11].

spline can then be written as

$$g(x) = \sum_{j=-k}^{r} \eta_j Q_j(x) ,$$

where the coefficients for the natural polynomial spline form a subset of these coefficients, namely $\boldsymbol{\nu} = (\eta_1, \ldots, \eta_{r-m})$, and

$$[Q_{-k}(\boldsymbol{x}), \ldots, \ Q_r(\boldsymbol{x})] = [T_{-k}(\boldsymbol{x}), \ldots, \ T_r(\boldsymbol{x})]\boldsymbol{D}_m ,$$

where the $q \times q$ matrix $\boldsymbol{D}_m$ is the mth order full rank differencing matrix for knots $t_{-k}, \ldots, t_{r+m+1}$. Let $\boldsymbol{Q} = [Q_{-k}(\boldsymbol{x}), \ldots, Q_r(\boldsymbol{x})]$. Then

$$\boldsymbol{Q\eta} = \boldsymbol{T\beta} \quad \text{and} \quad \boldsymbol{D}_m\boldsymbol{\eta} = \boldsymbol{\beta}.$$

To map back onto the usual parameterization of the polynomial spline, partition $\boldsymbol{Q} = [\boldsymbol{Q}_1\ \boldsymbol{Q}_2]$ for $\boldsymbol{Q}_1 = [Q_{-k}(\boldsymbol{x}), \ldots, Q_{-m}(\boldsymbol{x})]$ and $\boldsymbol{Q}_2 = [Q_{-m+1}(\boldsymbol{x}), \ldots, Q_r(\boldsymbol{x})]$. The vector $\boldsymbol{\eta}$ is partitioned conformally. For $j = -k, \ldots, -m$, $Q_j(x)$ are unconstrained polynomials of degree $m-1$ for $a \le x \le b$. Then

$$[Q_{-k}(\boldsymbol{x}), \ldots, Q_{-m}(\boldsymbol{x})] = [\boldsymbol{1},\ \boldsymbol{x}, \ldots, \boldsymbol{x}^{m-1}]\boldsymbol{A}_{xQ}$$

for some suitable full rank matrix $\boldsymbol{A}_{xQ}$. The polynomial spline can then be written as

$$g(x) = \sum_{j=0}^{m-1} \tau_{s,j} x^j + \sum_{j=-m+1}^{r} \eta_j Q_j(x)$$

or

$$g(\boldsymbol{x}) = \boldsymbol{X}_s\boldsymbol{\tau}_s + \boldsymbol{Q}_2\boldsymbol{\eta}_2.$$

For clarity, this parameterization will be referred to as penalized splines using a Q-basis. From $\boldsymbol{D}_m\boldsymbol{\eta} = \boldsymbol{\beta}$,

$$\boldsymbol{u}_T = [\boldsymbol{0}_{r,k+1}\ \boldsymbol{I}_r]\boldsymbol{D}_m\boldsymbol{\eta} = [\boldsymbol{0}_{r,m}\ \boldsymbol{\Delta}_m]\boldsymbol{\eta} = \boldsymbol{\Delta}_m\boldsymbol{\eta}_2 ,$$

where the $r \times (r+m)$ matrix $\boldsymbol{\Delta}_m$ is the mth order differencing matrix defined for knots $t_{-m+1}, \ldots, t_{r+m}$. These results give equivalent representations of the polynomial spline as

$$g(\boldsymbol{x}) \;=\; \boldsymbol{T\beta} \;=\; \boldsymbol{X}_{\scriptscriptstyle T}\boldsymbol{\tau}_{\scriptscriptstyle T} + \boldsymbol{Z}_{\scriptscriptstyle T}\boldsymbol{u}_{\scriptscriptstyle T} \;=\; \boldsymbol{B\alpha} \;=\; \boldsymbol{X}_b\boldsymbol{\tau}_b + \boldsymbol{Z}_b\boldsymbol{u}_b \;=\; \boldsymbol{Q\eta} \;=\; \boldsymbol{X}_s\boldsymbol{\tau}_s + \boldsymbol{Q}_2\boldsymbol{\eta}_2\ ,$$

with the following basic relationships between the sets of random coefficients:

$$\begin{aligned}
\boldsymbol{\beta} &= \boldsymbol{D}_{k+1}\boldsymbol{\alpha} & \qquad \boldsymbol{u}_{\scriptscriptstyle T} &= [\boldsymbol{0}_{r,k+1}\,\boldsymbol{I}_r]\boldsymbol{\beta}\\
\boldsymbol{\beta} &= \boldsymbol{D}_m\boldsymbol{\eta} & \qquad \boldsymbol{u}_b &= \boldsymbol{\Delta}_d\boldsymbol{\alpha}\\
\boldsymbol{\eta} &= \boldsymbol{D}_m^{-1}\boldsymbol{D}_{k+1}\boldsymbol{\alpha} & \qquad \boldsymbol{\eta}_2 &= [\boldsymbol{0}_{r+m,m}\,\boldsymbol{I}_{r+m}]\boldsymbol{\eta}\ ,
\end{aligned}$$

from which additional relationships such as $\boldsymbol{u}_{\scriptscriptstyle T} = \boldsymbol{\Delta}_{k+1}\boldsymbol{\alpha}$ and $\boldsymbol{u}_{\scriptscriptstyle T} = \boldsymbol{\Delta}_m\boldsymbol{\eta}_2$ can be derived.

These relationships can be used to directly compare penalties across the different spline models. For the penalized spline using a TPF basis, $\boldsymbol{u}_{\scriptscriptstyle T} = \boldsymbol{\Delta}_{k+1}\boldsymbol{\alpha} = \boldsymbol{\Delta}_m\boldsymbol{\eta}_2$, and the penalized spline penalty can be written as

$$\boldsymbol{u}_{\scriptscriptstyle T}^{\top}\boldsymbol{u}_{\scriptscriptstyle T} \;=\; \boldsymbol{\alpha}^{\top}\boldsymbol{\Delta}_{k+1}^{\top}\boldsymbol{\Delta}_{k+1}\boldsymbol{\alpha} \;=\; \boldsymbol{\eta}_2^{\top}\boldsymbol{\Delta}_m^{\top}\boldsymbol{\Delta}_m\boldsymbol{\eta}_2 .$$

The penalized spline of degree k using a TPF basis is therefore equivalent to a k-degree P-spline using a B-spline basis with differencing of order $k+1$, or to differencing of order m on the random coefficients of the k-degree penalized spline using a Q-basis. For example, the cubic penalized spline using a TPF basis is equivalent to a cubic P-spline with fourth order differencing, or to a penalty constructed from second order differences on the random coefficients of a cubic penalized spline using a Q-basis. The family of penalized splines using a TPF basis and identity penalty matrix is thus a subset of the family of P-splines using a B-spline basis and general differencing penalties. Within this subset, the penalty is determined by the order of the spline basis. This connection between the two types of splines was also made by Eilers & Marx (2004), although they only considered splines with equally spaced knots.

The equivalence between the penalized splines based on TPF or B-spline bases and polynomial smoothing splines is less obvious because of the naturalness constraints and the form of the matrix $\boldsymbol{H}_s^{-1}$ used in constructing the penalty. For equally spaced knots, the cubic smoothing spline penalty becomes

$$\boldsymbol{\delta}^{\top}\boldsymbol{G}_s^{-1}\boldsymbol{\delta} = \frac{2}{3}h\left(\sum_{j=2}^{r-1}\delta_j^2 + \frac{1}{2}\sum_{j=3}^{r-1}\delta_{j-1}\delta_j\right).$$

Using the relationship $\boldsymbol{D}_2\boldsymbol{\eta} = \boldsymbol{D}_4\boldsymbol{\alpha}$, then $\boldsymbol{\Delta}_2\boldsymbol{\alpha} = 6h(\eta_{-1}, \ldots, \eta_r)$ and the cubic P-spline penalty with $d = 2$ can be written as

$$\boldsymbol{\alpha}^{\top}\boldsymbol{\Delta}_2^{\top}\boldsymbol{\Delta}_2\boldsymbol{\alpha} \;=\; 36h^2\sum_{j=-1}^{r}\eta_j^2 \;=\; h^2\sum_{j=2}^{r-1}\delta_j^2 \;+\; 36h^2\big(\eta_{-1}^2 + \eta_0^2 + \eta_{r-1}^2 + \eta_r^2\big).$$

Ignoring the extra terms that are constrained to zero in the cubic smoothing spline, this penalty is proportional to an identity penalty matrix used on the second derivative scale. For equally spaced knots, the second difference penalty on the cubic B-spline scale can thus be considered as a discrete approximation to the cubic smoothing spline penalty.

4. A general model for polynomial splines

The different mixed model splines are defined by the choice of spline basis, number of knots and a penalty. The penalty is defined in terms of a penalty matrix applied to coefficients that may have been differenced. All of these spline models can be written in terms of a general mixed model spline, including constraints to transform unconstrained splines into natural splines.

The general model will be written in terms of B-splines because this basis generates a model that is easily interpretable. B-splines are local, so that coefficients can be related to behaviour at a specific location and are also widely available in software packages. Moreover, transformation to other bases is straightforward.

The general mixed model spline for $x \in [a, b]$ is defined in terms of a set of $q = r + k + 1$ (B-spline) basis functions $B_{-k}, \ldots, B_r$ of degree k, for a set of r knots $t_1, \ldots, t_r$, augmented with $t_{-k}, \ldots, t_0$ and $t_{r+1}, \ldots, t_{r+k+1}$, such that $t_j < t_{j+1}$ for $j = -k, \ldots, r + k$ and $t_0 \leq a$, $t_{r+1} \geq b$, with order of differencing d (which may be zero), and a penalty matrix $\boldsymbol{G}^{-1}$ (which may be the identity matrix). The matrix $\boldsymbol{G}^{-1}$ must be positive semi-definite, but may be less than full rank, e.g. an integrated squared second derivative penalty for a full basis of cubic splines. The spline can be written as

$$g(x) = \sum_{j=-k}^{r} \alpha_j B_j(x).$$

The general spline minimizes the penalized sum of squares

$$(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha})^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha}) + \lambda \boldsymbol{\alpha}^\top \boldsymbol{\Delta}_d^\top \boldsymbol{G}^{-1} \boldsymbol{\Delta}_d \boldsymbol{\alpha} \ ,$$

where $\boldsymbol{B} = [B_{-k}(\boldsymbol{x}), \ldots, B_r(\boldsymbol{x})]$, $\boldsymbol{\alpha} = (\alpha_{-k}, \ldots, \alpha_r)$ is a set of unknown coefficients, and the $(q - d) \times q$ matrix $\boldsymbol{\Delta}_d$ is the dth order differencing matrix for knots $t_{-k}, \ldots, t_{r+d}$. For positive-definite $\boldsymbol{G}$, the general spline is fitted using the mixed model (5) but now with $\operatorname{var}(\boldsymbol{u}_b) = \sigma_s^2 \boldsymbol{G}$.

For a singular penalty matrix, a different algorithm is required. Also, constraints on the coefficients may be required to produce a natural or periodic spline. Constraints $\boldsymbol{C}^\top \boldsymbol{\alpha} = 0$, where $\boldsymbol{C}$ is a $q \times c$ matrix of full rank representing c constraints, can be imposed directly via $\boldsymbol{\alpha} = \boldsymbol{S}\boldsymbol{\zeta}$ where $\boldsymbol{S}$ is a $q \times (q - c)$ matrix of full column rank such that $\boldsymbol{C}^\top \boldsymbol{S} = \boldsymbol{0}$. The penalized sum of squares to be minimized becomes

$$(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{S}\boldsymbol{\zeta})^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{S}\boldsymbol{\zeta}) + \lambda \boldsymbol{\zeta}^\top \boldsymbol{S}^\top \boldsymbol{\Delta}_d^\top \boldsymbol{G}^{-1} \boldsymbol{\Delta}_d \boldsymbol{S}\boldsymbol{\zeta}.$$

The spectral decomposition

$$\boldsymbol{S}^\top \boldsymbol{\Delta}_d^\top \boldsymbol{G}^{-1} \boldsymbol{\Delta}_d \boldsymbol{S} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$$

is used where the $(q - c) \times (q - c - s)$ matrix $\boldsymbol{U}$ is orthonormal, $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_{q-c-s}$ for $s \geq 0$, and the diagonal matrix $\boldsymbol{D} = \operatorname{diag}(d_1, \ldots, d_{q-c-s})$, $d_1 \geq \cdots \geq d_{q-c-s} > 0$ contains all the non-zero eigenvalues. Then

$$\boldsymbol{B}\boldsymbol{S}\boldsymbol{\zeta} = \boldsymbol{B}\boldsymbol{S}\boldsymbol{L}\boldsymbol{\tau} + \boldsymbol{B}\boldsymbol{S}\boldsymbol{U}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{u} \ ,$$

where $\boldsymbol{\tau} = \boldsymbol{L}^\top \boldsymbol{\zeta}$ is an $s \times 1$ vector of fixed effects, $\boldsymbol{L}$ is a $(q-c) \times s$ matrix of full column rank such that $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{I}_{q-c} - \boldsymbol{U}\boldsymbol{U}^\top$, and $\boldsymbol{u} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}^\top\boldsymbol{\zeta}$ is a vector of $q-c-s$ random effects. The penalty term is proportional to $\boldsymbol{u}^\top\boldsymbol{u}$ and the constrained spline fitted as

$$\boldsymbol{y} \;=\; \boldsymbol{B}\boldsymbol{S}\boldsymbol{L}\boldsymbol{\tau} \;+\; \boldsymbol{B}\boldsymbol{S}\boldsymbol{U}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{u} \;+\; \boldsymbol{e}$$

with $\operatorname{var}(\boldsymbol{u}) = \sigma_s^2 \boldsymbol{I}_{q-c-s}$. The number of fixed effects depends upon the rank of and interactions between the constraints $\boldsymbol{C}\boldsymbol{\alpha} = \boldsymbol{0}$, the differencing matrix $\boldsymbol{\Delta}_d$ and the penalty matrix $\boldsymbol{G}^{-1}$. Transformation between bases follows directly as a special case of constraints: if $\boldsymbol{T}\boldsymbol{\beta} = \boldsymbol{B}\boldsymbol{\alpha}$ with $\boldsymbol{T} = \boldsymbol{B}\boldsymbol{A}$ for some full rank matrix $\boldsymbol{A}$, then $\boldsymbol{\alpha} = \boldsymbol{A}\boldsymbol{\beta}$ and calculation follows as above with $\boldsymbol{S} = \boldsymbol{A}$.

This general form of mixed model spline is a broader family than those previously considered that includes general knot positions, differencing, a penalty matrix and constraints. The mixed model P-splines considered by Currie & Durban (2002) and Eilers & Marx (2004) take this form with equally spaced knots and an identity penalty matrix. The polynomial smoothing splines and the penalized splines with a TPF basis used by Ruppert *et al.* (2003) correspond to the general model with the order of differencing determined by the order of the spline. In addition in the case of the polynomial smoothing splines, knots are positioned at the covariate values, the penalty matrix must contain the integrated cross-products of mth derivatives of the basis functions and naturalness constraints are required.

4.1. Example: cubic smoothing spline with periodic constraints

Consider the construction of a periodic cubic smoothing spline on the range $[0, \omega]$ with knots at $\omega/r, 2\omega/r, \ldots, \omega$. Additional knots $-3\omega/r, \ldots, 0$ and $(r+1)\omega/r, \ldots, (r+4)\omega/r$ are used to construct the basis. Construction of the cubic smoothing spline requires the penalty matrix $\boldsymbol{G}_B^{-1}$ such that

$$\boldsymbol{\alpha}^\top \boldsymbol{G}_B^{-1} \boldsymbol{\alpha} = \int_0^\omega (g''(x))^2 dx.$$

This can be achieved by setting the $(i+4, j+4)$th element of $\boldsymbol{G}_B^{-1}$ to $\int_0^\omega B_i''(x) B_j''(x) dx$ for $i, j = -3, \ldots, r$. Calculation is simpler starting from the TPF basis, where elements of the penalty $\boldsymbol{G}_T^{-1}$ are equal to

$$\begin{aligned}\int_0^\omega T_i''(x) T_j''(x) dx &= \int_0^\omega (x - t_i)_+ (x - t_j)_+ \, dx \\ &= \frac{1}{3}\big(\omega^3 - t_k^3\big) - \frac{1}{2}\big(\omega^2 - t_k^2\big)(t_i + t_j) + (\omega - t_k) t_i t_j\end{aligned}$$

where $t_k = \max(t_i, t_j, 0)$. The $(r+4) \times (r+4)$ penalty matrix $\boldsymbol{G}_T^{-1}$ with $(i+4, j+4)$th element derived from the TPF basis has rank $(r+2)$ and, using $\boldsymbol{\beta} = \boldsymbol{D}_4\boldsymbol{\alpha}$, can be transformed onto the B-spline basis as $\boldsymbol{G}_B^{-1} = \boldsymbol{D}_4^\top \boldsymbol{G}_T^{-1} \boldsymbol{D}_4$.

Periodic constraints require $g^{(k)}(0) = g^{(k)}(\omega)$ for $k = 0, 1, 2$ in order to get continuity at the ends of the period equal to that within the spline. This translates to using the $(r+4) \times 3$ constraint matrix $\boldsymbol{C}$ with entries

$$[\boldsymbol{C}]_{ij} \;=\; B_{i-4}^{(j-1)}(\omega) - B_{i-4}^{(j-1)}(0)$$

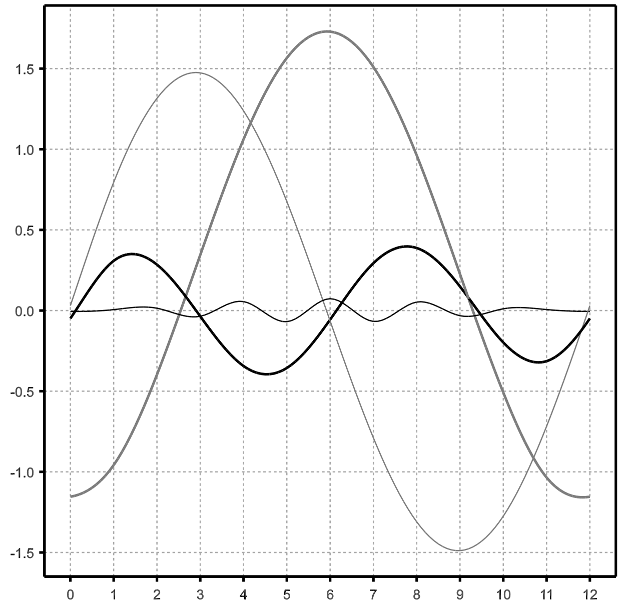

Figure 4. A subset of basis functions for a cubic smoothing spline constrained to be periodic on the range [0,12] with knots at $1, \dots, 12$.

for $i = 1, \dots, r + 4$, $j = 1, 2, 3$. This procedure results in $s = 1$, with the constant as the only fixed term in the model. A subset of the r periodic basis functions is shown in Figure 4. This method is equivalent to the algebraic procedure used by Zhang, Lin & Sowers (2000) to construct periodic cubic smoothing splines.

5. Properties of the model and the fitted spline

It is helpful to consider separately the different choices that can be made in defining the penalized splines and their impact on the resulting model.

5.1. Choice of spline basis

For a reasonable number of knots and a given order of differencing, the degree of the spline basis usually has little influence on the fitted spline at the knot points, although interpolation between the knots will take the form of the underlying basis. The degree of the spline basis chosen should therefore be appropriate to the aims of the analysis. For example, if growth rates are of interest, then the basis should be continuous and differentiable, i.e. at least quadratic. Cubic bases are often chosen because of these considerations, as they are flexible and twice differentiable at the knots without being overly complex. An example where there are clear differences between linear and cubic bases is shown in Section 6.

5.2. Number of knots

There is a choice of both number and position of knots. Reducing the number of knots reduces the flexibility of the fitted spline, and increasing the density of knots in different regions of the covariate allows increased flexibility within those regions. Choice of knot set has been widely discussed in the regression spline literature, although for regression splines there is a need to minimize the number of knots to avoid over-fitting. In the penalized spline context, use of a penalty avoids over-fitting and the object is to use sufficient knots to allow a flexible spline without incurring too large a computational load. This problem has been considered for penalized splines by Ruppert (2002) and Ruppert *et al.* (2003, Chapter 5), who give a recipe for knot positions based on quantiles of the covariate. For a covariate $\boldsymbol{x}$ with p distinct values, they recommend using r knots where

$$r = \min\left(\frac{p}{4}, 35\right),$$

and place knots at the $(i+1)/(r+2)$ quantiles of the data for $i = 1, \ldots, r$.

However, Eilers & Marx (2004) strongly recommend knots with equal spacing and give examples where, with large gaps in the set of covariate values, knots based on quantiles of the covariate may give undesirable results. Large differences between these two approaches can only occur when the covariate is unevenly distributed across the range, and it may be appropriate to consider the impact of both approaches in these cases. It should also be remembered that low-rank approximations based on a reduced number of knots may be poor, and both Eilers & Marx (1996, 2004) and Ruppert *et al.* (2003) emphasize that the knot set should be sufficiently dense to avoid loss of information. Ruppert *et al.* (2003, Section 5.6) propose algorithms to determine the minimum acceptable number of knots.

5.3. Order of differencing and penalty matrix

The order of differencing and the penalty matrix combine together to constrain the basis function coefficients. In practice, penalty matrices and differencing have rarely been used together. Penalty matrices are generally used as the evaluation of an integral-based penalty, as in cubic smoothing splines or the wider class of L-splines (Wahba, 1990; Welham *et al.* 2006). Differencing matrices are often used as an approximation to integral-based penalties. The impact of non-identity penalty matrices in combination with differencing does not appear to have been studied.

As the order of differencing increases, the average rate of change in adjacent coefficient values is progressively constrained, leading to smoother splines for a given smoothing parameter value. In practice, the smoothing parameter is selected (by REML) to optimize the fit of the model to the data, so that spline models with higher orders of differencing tend to use smaller smoothing parameters (larger variance components) in order to achieve a similar fit. The impact of the order of differencing on the fitted spline is examined below.

There is currently no satisfactory objective method of choosing the optimal order of differencing for a data set within the mixed model framework. The mixed model splines are fitted by maximizing the REML log-likelihood function, but the change in the REML log-likelihood cannot be used to compare mixed models with different fixed terms (Welham and Thompson, 1997). As the fixed model depends on the order of differencing, this criterion cannot be applied here. In the context of P-splines, Currie & Durban (2002) suggested use of the corrected Akaike Information criterion (AICc) criterion of Hurvich, Simonoff & Tsai (1998) to choose the degree of spline basis k, order of differencing d, and number of knots r, and then estimated the smoothing parameter for the chosen spline using REML in a mixed model. Although Currie & Durban (2002) used this technique, they did not evaluate it except in the case of correlated data, where they showed that it led to under-smoothing compared to REML estimation. The AICc criterion was developed in the context of models with a single source of independent errors, and requires modification to be used in the general mixed model setting or with REML estimation procedures.

6. A comparison of mixed model splines

It is useful to explore the impact of the penalty on the fitted spline to determine the implications for the choice of an appropriate model. In this section, the fit of various polynomial splines to known functions is evaluated in terms of mean squared error of prediction (MSEP).

A simulation study was used to compare the behaviour of several mixed model splines across four known functions, $f_A(x)$ to $f_D(x)$, see Figure 5. Function f_A is a linear combination of beta functions, used by Wang and Wahba (1995), that gives a smooth symmetric curve. Functions f_B to f_D were used by Hurvich *et al.* (1998) to evaluate methods of smoothing parameter selection for splines. Function f_B is a sine function with 2.5 cycles on [0, 1]. Function f_C is a linear combination of two exponential functions of x^2 which has different curvature for different covariate values. Function f_D takes separate exponential forms below and above $x = 1/3$, and is continuous but not differentiable at $x = 1/3$. All functions

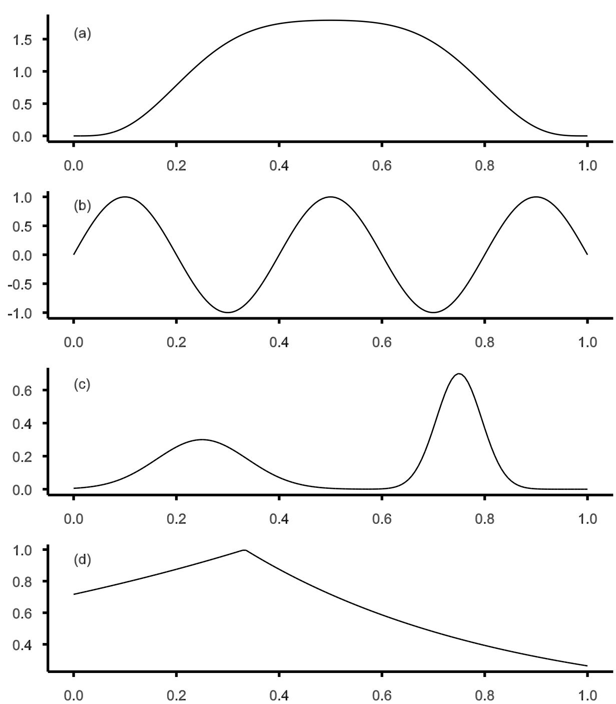

Figure 5. Functions used for comparison of smoothing spline models in simulation study: (a) f_A, (b) f_B, (c) f_C and (d) f_D.

are defined on the range [0, 1]. Four different error variances were tested for each function, defined in relation to the function range for $x \in [0, 1]$. Covariate values were defined as $p = 32$ or 128 equally spaced points on the range [0, 1] with no replication.

For each value of p, 500 sets of standard normal variables of length p were generated at random to form error vectors $\boldsymbol{e}_{pk}$, $k = 1, \ldots, 500$. Data sets $\boldsymbol{y}_{ijkp}$ were constructed as

$$\boldsymbol{y}_{ijkp} = f_i(\boldsymbol{x}_p) + \sigma_j R_i \boldsymbol{e}_{pk}$$

where $i = A, \ldots, D$, $j = 1, \ldots, 4$ with $\sigma_1 = 0.5, \sigma_2 = 0.25, \sigma_3 = 0.1, \sigma_4 = 0.05, k = 1, \ldots,$ 500, and R_i is the range of function $f_i(x)$ for $x \in [0, 1]$.

For each data set, cubic P-splines with differencing order $d = 1, 2, 3$ or 4, a cubic smoothing spline (CSS) and a linear penalized spline (LPS) based on a TPF basis were fitted. Knots were positioned at the p distinct covariate values. This set of models allowed a comparison between cubic P-splines with increasing orders of differencing, and also a comparison of different splines with similar penalties, i.e. the cubic P-spline with $d = 2$, the linear penalized spline and the cubic smoothing spline.

The fitted spline $\tilde{\boldsymbol{g}} = \tilde{g}(\boldsymbol{x}_p)$ at the covariate values and the estimated residual variance $\hat{\sigma}^2$ were saved for each model and each data set. The fit of each model for each data set was assessed by the mean squared error of prediction, MSEP, at the covariate values, calculated as

$$\text{MSEP}_{ijkp} = (f_i(\boldsymbol{x}_p) - \tilde{g}_{ijkp}(\boldsymbol{x}_p))^\top (f_i(\boldsymbol{x}_p) - \tilde{g}_{ijkp}(\boldsymbol{x}_p))/p.$$

The average MSEP across 500 sets of random errors assessed the performance of the splines in recapturing the underlying curve at the data points.

6.1. Results

The average value of MSEP (with standard deviation) assessed over the 500 data sets is shown in Table 1 for the four P-spline models. Results for the cubic smoothing spline and linear penalized spline were usually similar to those for the cubic P-spline with $d = 2$ and so are omitted here, although cases where these splines differed are discussed later. In each case, the best model was defined as the model with minimum average MSEP. To eliminate differences due to scale, the average MSEP values are presented as a percentage of the MSEP for the best model. The order of differencing that gave the minimum MSEP differed between functions and tended to increase as the signal:noise ratio increased, i.e. as the error variance decreased, or as p increased.

P-splines with higher orders of differencing tended to over-smooth the data for complex curves for larger error variances. For the smaller error variances, the overall smooth pattern was clearer, but P-splines with low orders of differencing tended to under-smooth and follow the data at the expense of the underlying pattern. Figure 6 shows that the difference between fitted splines with $d = 1$, 2 and 4 decreased as the error variance decreased. Although the fitted curves were very similar for smaller error variances, the second differences of the fitted splines show how the roughness of the curves increased as d decreased (see Figure 7).

The average estimated error variance for data is shown in Figures 8 and 9 as a percentage of the true value. Values much smaller than 100 indicate over-fitting, as some error is attributed to trend. Values much larger than 100 indicate over-smoothing. In most cases, the estimates decreased as the error variance decreased, and increased as the order of differencing increased.

TABLE 1

Average mean squared error of prediction (MSEP) with standard error for cubic P-spline models with $d = 1, 2, 3, 4$ over 500 data sets generated with four error standard deviations, σ_j, from known functions $f_A(x)$ to $f_D(x)$ with $p = 32$ or 128 covariate values

			Average MSEP as % of best model (s.e.)				
p	Function	σ_j	$d = 1$	$d = 2$	$d = 3$	$d = 4$	best model
32	f_A	0.50	126.6 (3.6)	100.0 (2.9)	104.2 (2.7)	116.6 (3.1)	0.11195
		0.25	126.6 (2.9)	100.0 (2.2)	125.8 (2.2)	112.3 (2.5)	0.03660
		0.10	141.7 (2.6)	100.0 (2.0)	112.4 (2.7)	110.6 (2.0)	0.00849
		0.05	202.6 (3.1)	111.0 (2.2)	100.0 (2.1)	109.8 (2.4)	0.00222
	f_B	0.50	100.0 (1.7)	119.4 (1.5)	119.7 (1.5)	111.6 (1.9)	0.38901
		0.25	119.4 (2.9)	110.4 (4.3)	105.4 (4.5)	100.0 (3.0)	0.08186
		0.10	182.8 (2.7)	116.3 (2.0)	100.0 (1.9)	101.4 (1.9)	0.01315
		0.05	249.6 (2.7)	134.3 (2.1)	105.9 (1.8)	100.0 (1.8)	0.00361
	f_C	0.50	100.0 (1.3)	111.3 (1.3)	116.4 (1.4)	116.2 (1.6)	0.03409
		0.25	100.0 (2.7)	131.5 (2.9)	153.2 (2.5)	150.2 (2.1)	0.01358
		0.10	102.9 (1.5)	101.7 (1.7)	157.4 (3.6)	258.9 (4.1)	0.00261
		0.05	140.0 (1.6)	100.0 (1.4)	109.6 (1.8)	180.8 (7.7)	0.00073
	f_D	0.50	101.0 (3.1)	100.0 (2.6)	100.7 (2.9)	111.2 (3.2)	0.01798
		0.25	103.4 (2.7)	100.1 (2.7)	102.4 (2.4)	102.7 (2.5)	0.00568
		0.10	130.2 (2.7)	100.0 (2.1)	122.6 (2.1)	139.2 (2.4)	0.00114
		0.05	145.6 (2.5)	100.0 (1.8)	119.7 (2.2)	135.5 (2.5)	0.00037
128	f_A	0.50	128.5 (2.7)	100.0 (2.4)	120.1 (2.3)	115.3 (2.6)	0.03642
		0.25	135.8 (2.3)	100.0 (2.0)	130.3 (3.0)	112.0 (2.2)	0.01264
		0.10	205.8 (2.7)	110.8 (2.1)	100.0 (2.1)	105.1 (2.2)	0.00232
		0.05	284.1 (3.0)	119.9 (2.0)	100.0 (1.9)	107.9 (1.8)	0.00064
	f_B	0.50	132.3 (2.1)	108.7 (2.3)	100.0 (2.3)	105.2 (2.0)	0.07170
		0.25	175.6 (2.3)	114.6 (1.9)	100.0 (1.9)	103.4 (2.0)	0.01966
		0.10	264.1 (2.7)	133.3 (1.9)	105.5 (1.8)	100.0 (1.8)	0.00364
		0.05	361.3 (3.1)	153.6 (2.0)	112.0 (1.7)	100.0 (1.7)	0.00101
	f_C	0.50	100.0 (2.0)	120.2 (2.3)	145.9 (2.3)	151.7 (2.1)	0.01167
		0.25	105.0 (1.4)	100.1 (1.6)	124.4 (1.9)	165.5 (2.5)	0.00376
		0.10	147.2 (1.6)	100.0 (1.3)	101.3 (1.4)	113.6 (1.6)	0.00073
		0.05	200.6 (1.8)	111.2 (1.4)	100.0 (1.3)	103.6 (1.4)	0.00020
	f_D	0.50	100.6 (2.5)	100.0 (2.7)	101.5 (2.5)	102.1 (2.5)	0.00583
		0.25	122.8 (2.5)	100.0 (2.3)	114.8 (2.3)	125.8 (2.2)	0.00171
		0.10	145.8 (2.2)	100.0 (1.9)	119.5 (2.2)	133.5 (2.4)	0.00039
		0.05	166.9 (2.1)	100.0 (1.6)	115.9 (1.7)	139.6 (1.9)	0.00013

Using $p = 32$, for $d = 1$ there was substantial under-estimation as the error variance decreased, supporting the earlier interpretation of patterns in MSEP. For $d = 2$, there was some over-estimation for larger error variances, and some under-estimation for smaller error variances, but most of the estimates were acceptable. For $d = 3$ and $d = 4$, the error variance tended to be over-estimated, but there were few very large over-estimates, except in the case of function $f_C(x)$. A similar pattern held for $p = 128$, although the estimates were then much closer to the true value.

In most cases, the MSEP and estimated error variance were almost identical for the cubic P-spline with $d = 2$, cubic smoothing spline and linear penalized spline. This was not the case for functions f_A, f_B and f_C with $p = 32$ at the smallest error variance, where the linear penalized spline had a much smaller estimate of error variance (and increased MSEP) compared to the cubic P-spline with $d = 2$, with the cubic smoothing spline intermediate (see Figure 8).

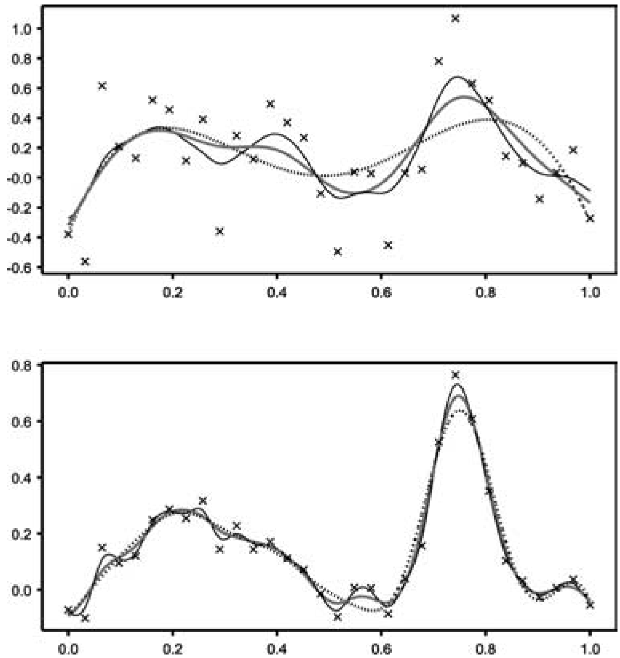

Figure 6. Simulated data generated from function f_C with $\sigma_j = 0.5$ (top) or 0.1 (bottom) with fitted splines from cubic P-spline mixed models with $d = 1$ (thin black line), $d = 2$ (thick grey line) or $d = 4$ (dotted line).

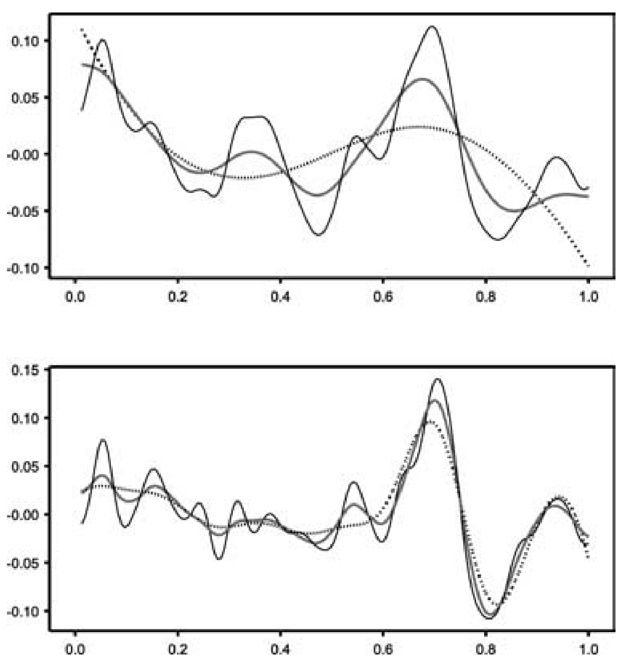

Figure 7. Second differences of fitted mixed model splines in Figure 6. Fitted splines generated from cubic P-spline mixed models with $d = 1$ (thin black line), $d = 2$ (thick grey line) or $d = 4$ (dotted line).

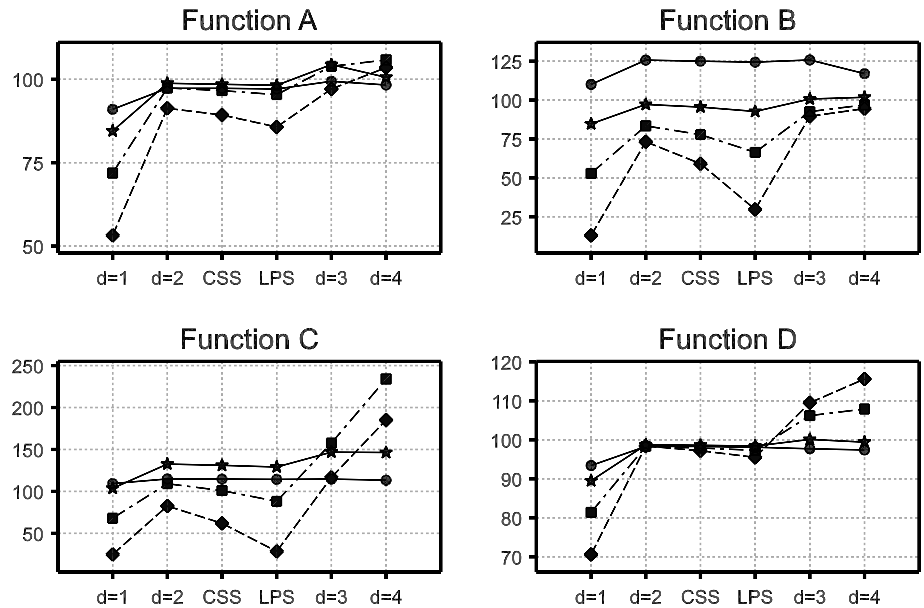

Figure 8. Average estimated residual variance (as percentage of true value) for cubic P-spline models with order of differencing $d = 1, 2, 3, 4$, cubic smoothing spline (CSS) and linear penalized spline (LPS) over 500 data sets generated from known functions with $p = 32$ covariate values and four error standard deviations, $\sigma_j = 0.5$ (●), $\sigma_j = 0.25$ (⋆), $\sigma_j = 0.1$ (■), $\sigma_j = 0.05$ (♦).

In these cases, the linear penalized spline fitted the data points much more closely. As the cubic P-spline and linear penalized spline used the same penalty, the differences must arise from the change from a cubic to a linear basis. The linear basis is required to be continuous at the knots, but there is no continuity constraint on derivatives, whereas the cubic basis is required to be continuous in both first and second differentials. In these examples, the resulting fitted linear spline had a roughness undesirable in many contexts.

To summarize, none of the splines evaluated here performed well over the whole set of simulations. Although cubic P-splines with second order differencing had reasonable overall performance, there were still cases where the average MSEP was $>50\%$ greater than for the best overall model. In the absence of an objective criteria, starting with a second differencing penalty (or close equivalent), plotting the fitted spline and modifying the penalty where undesirable characteristics are present is suggested.

7. Discussion

The area of low-rank mixed model penalized splines has recently received much attention. For many practical reasons, these splines are becoming a popular method of including non-linear curves within more general mixed models. This paper is intended to highlight connections between the different types of penalized spline currently available. The choice of knot points has been studied previously, but the impact of the choice of penalty on the fitted spline has received much less attention. For smoothing splines and for penalized splines

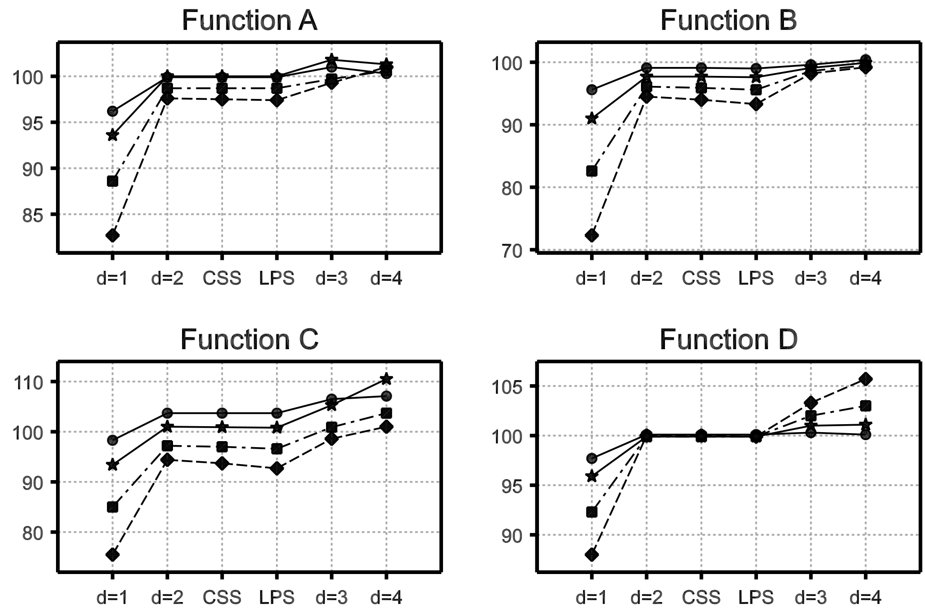

Figure 9. Average estimated residual variance (as percentage of true value) for cubic P-spline models with order of differencing $d = 1, 2, 3, 4$, cubic smoothing spline (CSS) and linear penalized spline (LPS) over 500 data sets generated from known functions with $p = 128$ covariate values and four error standard deviations, $\sigma_j = 0.5$ (●), $\sigma_j = 0.25$ (⋆), $\sigma_j = 0.1$ (■), $\sigma_j = 0.05$ (♦).

based on a TPF basis with identity penalty matrix, the penalty is determined by the spline order. This may not be obvious to the naive practitioner, who might otherwise be attracted by the simplicity of the TPF basis. Using the P-spline formulation with a B-spline basis, the penalty has to be chosen explicitly via the order of differencing. As it has been shown here that the choice of penalty can have a major impact on the fit of the model, this is an important advantage. However, the general model of Section 4 and connections between the bases developed in Section 3 can be used to translate a differencing penalty based on a B-spline basis into an equivalent form for a TPF basis, allowing both transparency and flexibility. Unfortunately, methods to determine the appropriate penalty do not currently exist within the mixed model framework using REML estimation, and further work is required in this area.

This paper dealt only with the use of penalized splines for curve fitting. Various extensions to the methods have been made. Currie, Durban & Eilers (2004) considered the use of P-splines for forecasting, albeit with warnings that extrapolation should always be approached with caution, and showed that the form of the penalty also controls the structure of the extrapolated spline. These authors also extended the P-spline model to two (or more) dimensions and, in later papers, develop efficient algorithms for fitting multi-dimensional P-splines (Eilers, Currie & Durban, 2006; Currie, Durban & Eilers, 2006), although not using mixed model methods. A two-dimensional version of penalized splines based on TPF basis functions and using REML estimation within the mixed model framework was developed by Kammann & Wand (2003).

Appendix: Divided differences and general differencing matrices

For a set of values $t_j, \dots, t_{j+d}$, the dth order divided difference of a function f, denoted $[t_j, \dots, t_{j+d}]f$, can be calculated recursively from the formulae

$$[t_j, t_{j+1}]f = \frac{f(t_j) - f(t_{j+1})}{t_j - t_{j+1}}$$

$$[t_j, \dots, t_{j+l}]f = \frac{[t_j, \dots, t_{j+l-1}]f - [t_{j+1}, \dots, t_{j+l}]f}{t_j - t_{j+l}} \quad \text{for } l = 2, \dots, d.$$

Here, the scaled divided differences $(-1)^d\,(t_{j+d} - t_j)\,[t_j, \dots, t_{j+d}]f$ are of interest. Define coefficients $\xi_{d,j,l}$, $l = 0, \dots, d$, such that

$$(-1)^d (t_{j+d} - t_j)[t_j, \dots, t_{j+d}]f \;=\; \sum_{l=0}^{d} \xi_{d,j,l}\, f(t_{j+l}).$$

For a set of points $t_m, \dots, t_{m+q+1}$, define the full rank $q \times q$ first order differencing matrix $\boldsymbol{D}_1$ as

$$[\boldsymbol{D}_1]_{ij} \;=\; \begin{cases} 1 & i = j,\ i = 1, \dots, q \\ -1 & j = i-1,\ i = 2, \dots, q \\ 0 & \text{otherwise.} \end{cases}$$

Define diagonal scaling matrices $\boldsymbol{S}_l$, $l = 1, \dots, d$, for points $t_m, \dots, t_{m+q+l}$ as

$$\boldsymbol{S}_l \;=\; \operatorname{diag}\left\{(t_{j+l} - t_j)^{-1};\, j = m, \dots, m+q\right\}.$$

For equally spaced knots, where $t_{j+1} - t_j = h$, $\boldsymbol{S}_l = \boldsymbol{I}_q/(lh)$. The full rank $q \times q$ general differencing matrix $\boldsymbol{D}_d$ for $d > 1$ on points $t_m, \dots, t_{m+q+d}$ can be recursively defined from

$$\boldsymbol{D}_l \;=\; \boldsymbol{D}_{l-1}\boldsymbol{S}_{l-1}\boldsymbol{D}_1 \qquad \text{for } l = 2, \dots, d.$$

The inverses for the differencing matrices can also be defined recursively from

$$[\boldsymbol{D}_1^{-1}]_{ij} = \begin{cases} 1 & i \ge j,\ i, j = 1, \dots, q \\ 0 & \text{otherwise} \end{cases}$$

and

$$\boldsymbol{D}_l^{-1} = \boldsymbol{D}_1^{-1}\boldsymbol{S}_{l-1}^{-1}\boldsymbol{D}_{l-1}^{-1}, \qquad 1 < l \le d.$$

It is straightforward to verify that the (i, j)th element of $\boldsymbol{D}_d$ is

$$[\boldsymbol{D}_d]_{ij} \;=\; \begin{cases} \xi_{d,m+j-1,i-j} & \text{for } 0 \le i - j \le d,\ \ i, j = 1, \dots, q \\ 0 & \text{otherwise.} \end{cases}$$

Finally, the dth order general differencing matrix $\boldsymbol{\Delta}_d$ for points $t_m, \dots, t_{m+q+d}$ is defined as the $q - d \times q$ sub-matrix consisting of the last $(q - d)$ rows of $\boldsymbol{D}_d$.

References

DE BOOR, C. (1978). *A practical guide to splines*. New York: Springer.

BRUMBACK, B. A. & RICE, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *J. Amer. Statist. Assoc.* **93**, 961–994.

BRUMBACK, B. A., RUPPERT, D. & WAND, M. P. (1999). Comment on: Variable selection and function estimation in additive nonparametric regression using a data-based prior (by Shively, Kohn and Wood). *J. Amer. Statist. Assoc.* **94**, 794–797.

CRAINICEANU, C. M. & RUPPERT, D. (2004). Restricted likelihood ratio tests in nonparametric longitudinal models. *Statist. Sinica* **12**, 713–729.

CURRIE, I. D. & DURBAN, M. (2002). Flexible smoothing with P-splines: a unified approach. *Statist. Modelling* **2**, 333–349.

CURRIE, I. D., DURBAN, M. & EILERS, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statist. Modelling* **4**, 279–298.

CURRIE, I., DURBAN, M. & EILERS, P. (2006). Generalized linear array models with applications to multidimensional smoothing. *J. Roy. Statist. Soc. Ser. B* **68**, 259–280.

EILERS, P. H. C. (1999). Comment on: The analysis of designed experiments and longitudinal data by using smoothing splines (by Verbyla *et al.*) *Appl. Statist.* **48**, 307–308.

EILERS, P. H. C. & MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11**, 89–121.

EILERS, P. H. C. & MARX, B. D. (2004). Splines, knots and penalties. Available from: http://www.stat.lsu.edu/faculty/marx/splines_knots_penalties.pdf (submitted).

EILERS, P. H. C., CURRIE, I. D. & DURBAN, M. (2006). Fast and compact smoothing on large multidimensional grids. *Comput. Statist. Data Anal.* **50**, 61–76.

GREEN, P. J. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika* **72**, 527–537.

GREEN, P. J. & SILVERMAN, B. W. (1994). *Nonparametric regression and generalized linear models*. London: Chapman and Hall.

HASTIE, T. J. & TIBSHIRANI, R. J. (1986). Generalized additive models (with discussion). *Statist. Sci.* **1**, 297–318.

HASTIE, T. J. & TIBSHIRANI, R. J. (1990). *Generalized additive models*. London: Chapman and Hall.

HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (2001). *The elements of statistical learning*. New York: Springer.

HURVICH, C. M., SIMONOFF, J. S. & TSAI, C-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. Roy. Statist. Soc. Ser. B* **60**, 271–293.

KAMMANN, E. E. & WAND, M. P. (2003). Geoadditive models. *Appl. Statist.* **52**, 1–18.

KOHN, R., ANSLEY, C. F. & THARM, D. (1991). The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *J. Amer. Statist. Assoc.* **86**, 1042–1050.

O'SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statist. Sci.* **1**, 505–527.

O'SULLIVAN, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM J. Sci. Statist. Comput.* **9**, 363–379.

PARISE, H., WAND, M. P., RUPPERT, D. & RYAN, L. (2001). Incorporation of historical controls using semi-parametric mixed models. *Appl. Statist.* **50**, 31–42.

PARKER, R. L. & RICE, J. A. (1985). Comment on: Some aspects of the spline smoothing approach to nonparametric regression curve fitting (by Silverman). *J. Roy. Statist. Soc. Ser. B* **47**, 41–42.

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **31**, 100–109.

RUPPERT, D. (2002). Selecting the number of knots for penalized splines. *J. Comput. Graph. Statist.* **11**, 735–757.

RUPPERT, D. & CARROLL, R. J. (2000). Spatially adaptive penalties for spline fitting. *Aust. N. Z. J. Stat.* **42**, 205–223.

RUPPERT, D., WAND, M. P. & CARROLL, R. J. (2003). *Semiparametric regression*. Cambridge: Cambridge University Press.

SPEED, T. P. (1991). Comment on: That BLUP is a good thing: the estimation of random effects (by Robinson). *Statist. Sci.* **6**, 44.

THOMPSON, R. (1985). Comment on: Some aspects of the spline smoothing approach to nonparametric regression curve fitting (by Silverman). *J. Roy. Statist. Soc. Ser. B* **47**, 43–44.

VERBYLA, A. P., CULLIS, B. R., KENWARD, M. G. & WELHAM, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Appl. Statist.* **48**, 269–311.

WAHBA, G. (1984). Partial spline models for the semi-parametric estimation of functions of several variables. In: *Statistical Analysis of Time Series, Proceedings of the Japan US Joint Seminar*, Tokyo.

WAHBA, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline problem. *Ann. Statist.* **13**, 1378–1402.

WAHBA, G. (1990). *Spline models for observational data*. Philadelphia: SIAM.

WAND, M. P. (2003). Smoothing and mixed models. *Comput. Statist.* **18**, 223–249.

WANG, Y. (1998). Mixed effects smoothing spline analysis of variance. *J. Roy. Statist. Soc. Ser. B* **60**, 159–174.

WANG, Y. & WAHBA, G. (1995). Bootstrap confidence intervals for smoothing splines and their comparison to Bayesian confidence intervals. *J. Statist. Comput. Sim.* **51**, 263–279.

WELHAM, S. J., CULLIS, B. R., KENWARD, M. G. & THOMPSON, R. (2006). The analysis of longitudinal data using mixed model L-splines. *Biometrics* **62**, 392–401.

WELHAM, S. J. & THOMPSON, R. (1997). Likelihood ratio tests for fixed model terms using residual maximum likelihood. *J. Roy. Statist. Soc. Ser. B* **59**, 701–714.

ZHANG, D., LIN, X., RAZ, J. & SOWERS, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *J. Amer. Statist. Assoc.* **93**, 710–719.

ZHANG, D., LIN, X. & SOWERS, M. (2000). Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics* **56**, 31–39.