Bayesian hierarchical modelling of growth curve derivatives via sequences of quotient differences

Garritt L. Page,  
*Brigham Young University, Provo, USA, and Basque Center for Applied Mathematics, Bilbao, Spain*  

María Xosé Rodríguez-Álvarez  
*Basque Center for Applied Mathematics, Bilbao, and Basque Foundation for Science, Bilbao, Spain*  

and Dae-Jin Lee  
*Basque Center for Applied Mathematics, Bilbao, Spain*  

[Received August 2018. Revised December 2019]

**Summary.** Growth curve studies are typically conducted to evaluate differences between group or treatment-specific curves. Most analyses focus solely on the growth curves, but it has been argued that the derivative of growth curves can highlight differences between groups that may be masked when considering the raw curves only. Motivated by the desire to estimate derivative curves hierarchically, we introduce a new sequence of quotient differences (empirical derivatives) which, among other things, are well behaved near the boundaries compared with other sequences in the literature. Using the sequence of quotient differences, we develop a Bayesian method to estimate curve derivatives in a multilevel setting (a common scenario in growth studies) and show how the method can be used to estimate individual and group derivative curves and to make comparisons. We apply the new methodology to data collected from a study conducted to explore the effect that radiation-based therapies have on growth in female children diagnosed with acute lymphoblastic leukaemia.

**Keywords:** Bayesian hierarchical models; Growth studies; Longitudinal data; Penalized splines; Smoothing

### 1. Introduction

Growth studies are quite common in a variety of scientific fields (biology, ecology, etc.). These types of studies produce what is often referred to as longitudinal data since growth is measured on each experimental unit or subject over time (see Fitzmaurice *et al.* (2008, 2011) and Diggle *et al.* (2013)). This is the case for the 618 children who suffered from acute lymphoblastic leukaemia (ALL) in the study that we consider (Dalton *et al.*, 2003). One of the late effects of treatments taken for childhood ALL is a decrease in growth (or short stature), and a motivation in conducting the study was to determine whether there is an ALL treatment that minimizes this adverse effect. Because cranial radiation has been connected with a deficiency in hormones that are associated with the development of growth, the three treatments that were studied were...
(a) intrathecal therapy with no radiation,
(b) intrathecal therapy with standard radiation and
(c) intrathecal therapy with twice-daily radiation (hyperfractionated).

Between 1987 and 1995 height was measured on each subject at diagnosis and approximately every 6 months thereafter (see Dalton et al. (2003) for more details). The resulting height measurements for 197 female subjects who were diagnosed with ALL between ages 2 and 9 years are provided in Fig. 1. A rich literature is dedicated to methods that have been developed to fit curves such as those found in Fig. 1. To provide context associated with this literature, we introduce the following general framework. For $j = 1, \ldots, m$ and $i = 1, \ldots, n_j$, let $(y_{ij}, t_{ij})$ denote the $i$th measured response from individual $j$ at time point $t_{ij}$ and consider

$$y_{ij} = f_j(t_{ij}) + \epsilon_{ij},$$

where $\{f_j\}_{j=1}^m$ are subject-specific unknown functions and $\epsilon_{ij} \sim \text{iid} \ N(0, \sigma_j^2)$. The challenge is to estimate each of the $f_j$ and then possibly deviations from a population curve that is often defined by a treatment or group assignment. Some methods that were found in the literature that carry this out are based on semiparametric and non-parametric frequentist approaches, such as smoothing splines (Brumback and Rice, 1998), penalized regression splines (Durban et al., 2005; Djundje and Currie, 2010) and functional data analysis (Ramsay and Silverman, 2005; Yao et al., 2005). In particular, Durban et al. (2005) first analysed the ALL data that are discussed here by means of a mixed effects model formulation of truncated line basis that enables us to estimate flexibly both subject-specific and treatment-specific growth curves. From a Bayesian perspective there appear to be two approaches to model equation (1) hierarchically. The first is to employ a hierarchy where each of the $f_j$ is modelled with a Gaussian process that is centred on a group curve which in turn is modelled with a Gaussian process (e.g. Behseta et al. (2005), Page et al. (2013) or Yang et al. (2016)). The second is to express $f_j$ as a linear combination of subject-specific basis functions and to assign a prior distribution to the corresponding basis function coefficients. Subsequently the posterior distribution of basis function coefficients produces an estimate of $f_j$ (see Crainiceanu et al. (2005), Botts and Daniels (2008) and Page and Quintana (2015)).

The preceding discussion underscores the fact that there are methods that might be employed to fit the data that are found in Fig. 1. However, since a motivation for conducting the study was to determine whether an ALL treatment mitigates the undesirable side effect of reduced growth, it seems reasonable to consider treatment’s effect on the growth velocity or rate of growth (i.e. the first derivative of $f_j$ which we denote with $f'_j$). As a result, for our application, there is considerable interest in estimating, in addition to growth curves, the rate of growth.

It would be straightforward to estimate derivative curves if parametric forms for $f'_j(\cdot)$ for $j = 1, \ldots, m$ are assumed to be known (so long as the parametric forms are differentiable; Zhang et al. (2012)). However, as with all types of parametric models, assuming that the exact form of $f'_j(\cdot)$ being known is quite restrictive and, if the model is misspecified, inferences will be wrong. Therefore, we focus here on non-parametric approaches to curve estimation and, as a consequence, derivative curve estimation. Also, in practice non-parametric methods can be used to confirm and thoroughly examine conclusions that are obtained by applying parametric methods. Therefore, we do not see non-parametric methods as merely alternatives to existing parametric methods but rather as supplements that could offer additional information and insights.

To the best of our knowledge, modelling and estimating derivative curves in the hierarchical setting found in Fig. 1 has received little attention in the statistical literature (see, for example,
Fig. 1. For the childhood ALL study, subject-specific raw height measurements for individuals who received one of the three treatments (a) intrathecal therapy with no radiation, (b) intrathecal therapy with standard radiation and (c) intrathecal therapy with twice-daily radiation (hyperfractionated): age is measured in years and height in centimetres.
Simpkin et al. (2018)). In fact, non-parametric methods that were developed to estimate a single derivative curve $f^{(1)}$ are sparse compared with that dedicated to estimating $f$. One possible reason for this is that $f^{(1)}$ is rarely (if ever) directly measured, introducing complexities that are associated with modelling and estimating $f^{(1)}$ that do not exist for $f$ (Ramsay and Silverman, 2005). Although sparse, the literature contains methods for estimating a single $f^{(1)}$ that are based on splines (see, for example, Ramsay and Silverman (2005), Sangalli et al. (2009), Grajeda et al. (2016) and Song (2016)), local polynomial regression (Gasser and Müller, 1984; Fan and Gijbels, 1996) or weighted sequences or quotient differences of the observed data (Müller et al., 1987; Härdle, 1999; De Brabanter et al., 2011, 2013; Wang and Lin, 2015; Dai et al., 2016; Charnigo et al., 2011; De Brabanter and Liu, 2015). Thus, one approach to estimate derivative curves in a hierarchical setting would be to apply one of the methods that were previously listed independently for each subject. A group mean derivative curve could then be estimated by using the derivative curves derived from each subject’s model. Although simple and straightforward, Kelley and Maxwell (2008) pointed out that a group mean derivative curve can be very different from that obtained by averaging individual curves when growth curves are non-linear. Thus, a hierarchical modelling approach would be undoubtedly appealing. In addition, it would naturally provide uncertainty estimates and permit borrowing of strength among the subject-specific derivative curves when estimating a treatment or group curve.

The aim of this paper is to present a new hierarchical approach that permits jointly and coherently estimating subject- and group-specific derivative curves and making comparisons between them. The method that we develop employs a new way of constructing empirical derivatives based on quotient differences of the observed data, and a Bayesian method that enables estimating curve derivatives in a multilevel setting. The use of quotient differences for estimating a single derivative curve $f^{(1)}$ can be found in Charnigo et al. (2011), De Brabanter et al. (2013) or De Brabanter and Liu (2015). All these approaches are based on what Ramsay and Silverman (2005) called ‘central differences’. In this paper, we propose a variation of what Ramsay and Silverman (2005) called ‘forward differences’ as it seems more natural that each empirical derivative be anchored to the time point at which the derivative is being estimated. As will be seen, the manner in which we employ the forward differences mitigates boundary effects compared with central differences. Finally, what we develop can accommodate subject-specific derivative curves of varying lengths when estimating group curves, which is something that is crucial for the study that we consider.

The remainder of the paper is organized as follows. In Section 2 we detail a new approach to constructing a sequence of quotient differences based on forward and backward differences and a Bayesian hierarchical model that can be employed to estimate subject-specific and group-specific (population) derivative curves. Section 3 contains results from two simulation studies and in Section 4 we consider two applications that illustrate the utility of the methodology proposed. The paper closes with a brief discussion in Section 5.

2. Hierarchical modelling of curve derivatives

In this section, we describe our approach to modelling and estimating derivative curves hierarchically. As mentioned, an overarching complication that is associated with estimating a derivative curve is that the derivative is rarely if ever directly measured. Thus, we begin by detailing our construction of a new sequence of quotient differences which essentially produces noisy pointwise derivative measurements and we then describe the specific components of a Bayesian model that permits estimating group- and subject-specific derivative curves. To construct empirical derivatives and to model the derivative curve simultaneously, we introduce
unknown subject-specific parameters. These unknown objects are embedded in a hierarchical model permitting us to construct a unique sequence of empirical derivatives for each subject (something that is not trivial from a frequentist perspective). This will require estimating parameters that are akin to Box–Cox transformation parameters simultaneously with unknown model parameters which has been considered from a Bayesian perspective (e.g. Gottardo and Raftery (2009)).

### 2.1. Sequence of empirical derivatives

Recall that we consider \( m \) independent subjects, with observations measured on subject \( j \) denoted as \((y_{ij}, t_{ij}), i = 1, \ldots, n_j\). A natural estimate of the first derivative for subject \( j \) at time \( t_{ij} \) would be

\[
\hat{f}^{(1)}_j(t_{ij}) = \frac{y_{(i+1)j} - y_{ij}}{t_{(i+1)j} - t_{ij}}.
\]  

Unfortunately, this estimator has been shown to be notoriously noisy (Ramsay and Silverman, 2005; De Brabanter et al., 2011) with the variability depending on the distance \( t_{(i+1)j} - t_{ij} \).

To reduce this noise, we adopt ideas that were motivated by \( D \) nearest neighbours and kernel smoothing by computing a weighted average of \( D \)-lagged forward and backward differences. Employing both forward and backward differences will help to mitigate boundary effects.

Let \((y_{(i+d)j} - y_{ij})/(t_{(i+d)j} - t_{ij})\) and \((y_{ij} - y_{(i-d)j})/(t_{ij} - t_{(i-d)j})\) be the \( d \)th lagged forward and lagged backward quotient difference. Further, let \( K_u(d) \) denote a kernel weight function based on bandwidth parameter \( u \) that assigns weights to the \( d \)-lagged quotient differences that diminish as \( d \) increases (note that the lag value is weighted: not necessarily the distance from \( t_{ij} \)).

In what follows we set \( K_u(d) = \phi(d; 1, u) \), where \( \phi(d; 1, u) \) denotes a Gaussian density function with mean 1 and standard deviation \( u \). By centring the kernel weight function at 1, 1-lagged forward and backward quotient differences receive the most weight as \( u \) decreases. In addition to bandwidth parameter \( u \), we also introduce a window or neighbourhood parameter \( D \) and consider only at most the nearest \( D \leq \lceil n/2 - 1 \rceil \) forward and backward lagged quotients. With this in mind, the sequence of quotient differences that we propose is

\[
y^{(1)}_{ij}(D_j, u_j) = \sum_{d=1}^{D_j^U} K_{u_j}(d) \frac{y_{ij} - y_{(i-d)j}}{t_{ij} - t_{(i-d)j}} + \sum_{d=1}^{D_j^L} K_{u_j}(d) \frac{y_{(i+d)j} - y_{ij}}{t_{(i+d)j} - t_{ij}},
\]  

where \( D_j^L = D_j - (D_j - i + 1)_+ \) and \( D_j^U = D_j - (D_j - n_j + i)_+ \) with \((\cdot)_+\) denoting the positive part. Note that \( D \) and \( u \) are subject specific, which provides greater flexibility in modelling individual derivative curves compared with assuming that each subject has the same \( D \) and \( u \).

The objects \( D \) and \( u \) regulate the variance–bias trade-off that exists in the sequence of empirical derivatives. As \( D \) and \( u \) increase more lagged quotient differences are included when computing \( y^{(1)}_{ij}(D_j, u_j) \), making the sequence more global, which reduces variability at the cost of introducing bias. As the \( D \) and \( u \) decrease, the elements of the sequence converge to a version of estimate (2) that includes a backward lagged quotient difference, making the sequence more local, which increases the variability but reduces bias.

A possibly simpler approach to constructing a sequence of quotient differences that is a special case of that just described is to include all possible forward and backward lagged empirical
derivative estimates at time point \( t_{ij} \) and to assign each diminishing weight via \( K_u(d) \). That is to say,

\[
\gamma_{ij}(u_j) = \frac{1}{\sum_{d=1}^{n_j-i} K_{u_j}(d) + \sum_{d=1}^{n_j-i} K_{u_j}(d)} \sum_{d=1}^{i-1} K_{u_j}(d) \frac{y_{ij} - y_{(i-d)j}}{t_{ij} - t_{(i-d)j}} + \sum_{d=1}^{n_j-i} K_{u_j}(d) \frac{y_{(i+d)j} - y_{ij}}{t_{(i+d)j} - t_{ij}},
\]

where for \( i = 1 \) and \( i = n_j \) we adopt the convention that corresponding summands are empty. In this case \( D \) plays no role. Both sequences of empirical derivatives are considered in the simulation study of Section 3.

Although for a simpler case than that considered in this paper, we mention that, if \( t_{ij} - t_{(i-1)j} = t_{kj} - t_{(k-1)j} \) for all \( i \neq k \), then \( \gamma_{ij}^{(1)}(D_j, u_j) \) reduces to the quotient differences that were developed in De Brabanter et al. (2013) for \( i \in \{ D_j + 1, \ldots, n_j - D_j \} \) but with different weights. For exterior points, however, the quotient differences are very different, with ours including more summands. The simulation study that is discussed in Section 3.1 suggests that this characteristic reduces boundary effects. In addition, web appendix A contains a simple example that illustrates how \( D_j \) and \( u_j \) influence the variance–bias that is associated with sequences (3) and (4) and how they handle boundary effects relative to the sequence that was found in De Brabanter et al. (2013).

### 2.2. Data and smoothing model

In principle, given \( u_j \) and/or \( D_j \) we can use the sequences \( \{ (\gamma_{ij}^{(1)}(D_j, u_j), t_{ij}) \}_{i=1}^{n_j} \) or \( \{ (\gamma_{ij}^{(1)}(u_j), t_{ij}) \}_{i=1}^{n_j} (j = 1, \ldots, m) \) together with any number of semiparametric and nonparametric methods (both frequentist and Bayesian) to estimate subject’s derivative curves (i.e. \( f_j^{(1)} \)). A complication is how to handle the unknowns \( D_j \) and \( u_j \). Because of this, we take a Bayesian approach that facilitates treating \( D_j \) and \( u_j \) as unknowns and lends itself to the hierarchical modelling that is detailed in this section. For simplicity, we describe our method using \( y_j^{(1)}(D, u) \) (equation (3)) as sequence \( y_j^{(1)}(u) \) (equation (4)) essentially follows the same arguments except that those associated with \( D \) are dropped. Specifically, we propose the following model to estimate individual derivative curves:

\[
y_{ij}^{(1)}(D_j, u_j) = f_j^{(1)}(t_{ij}) + \epsilon_{ij}^{\text{HD}} \sim N(0, \sigma_j^2).
\]

It can be argued that incorporating a more sophisticated error model (e.g. auto-regressive) may be warranted here given that the \( y_{ij}^{(1)}(D_j, u_j) \) are not independent. This lack of independence is particularly troublesome for estimating \( D_j \) and \( u_j \) via cross-validation. However, from a Bayesian perspective dependence among the \( y_{ij}^{(1)}(D_j, u_j) \) does not change the sampling mechanism that is associated with \( D_j \) or \( u_j \).

A fairly popular method of characterizing an unknown function such as \( f_j^{(1)} \) is to define a collection of basis functions (e.g. wavelet or polynomial) and to assume that \( f_j^{(1)} \) lies in their span. We adopt this method and employ a \( B \)-spline basis as it has some attractive computational properties and its local behaviour (see, for example, Eilers et al. (2015) for details) is crucial to modelling group-specific curves in unbalanced design studies. We note that the approach that we take is similar in spirit to the model that was found in Page and Quintana (2015) for curve (not derivative) estimation.

Let \( B_l(t, \xi) \) denote the \( l \)th \( B \)-spline basis function evaluated at \( t \) for knots \( \xi \). To facilitate estimating group-specific derivative curves in an unbalanced design, \( \xi \) is comprised of the same knots for all subjects. Knot values span the range between the first and last measurements times,
regardless of subject. Then, we can express \( f_j^{(1)}(t_{ij}) = \sum_{l=1}^{q_l} \beta_{jl} B_l(t_{ij}, \xi) \). In matrix notation, model (5) is thus expressed as

\[
y_j^{(1)}(D_j, u_j) = B_j \beta_j + \epsilon_j,
\]

where \( y_j^{(1)}(D_j, u_j) = (y_1^{(1)}(D_j, u_j), \ldots, y_n^{(1)}(D_j, u_j))^\top \) denotes the \( n \times 1 \) vector of quotient differences for the \( j \)th subject, \( B_j \) is the \( n \times q \) B-spline design matrix, \( \beta_j = (\beta_{j1}, \ldots, \beta_{jq})^\top \) is the \( q \times 1 \) vector of the B-spline basis coefficients and \( \epsilon_j \sim N_n(0, \sigma_j^2 I) \). We note that \( q \) (the length of \( \beta_j \)) depends on the dimension of \( \xi \) and the degree of the B-spline basis.

To estimate group-specific curves, we assume that each \( \beta_j \) is drawn from a distribution that is centred at the corresponding group-specific B-spline coefficients. More specifically, let \( g_j \in \{1, \ldots, C\} \) denote the \( j \)th subject’s group label where \( C \) is the number of groups and consider the following subject-specific coefficient model:

\[
\beta_j | g_j \sim N(\theta_{gj}, \lambda_j^2 I),
\]

where \( (\theta'_1, \ldots, \theta'_C) \) are the collection of \( C \) group-specific coefficient vectors with \( \theta_h = (\theta_{h1}, \ldots, \theta_{hq})^\top \) (\( h = 1, \ldots, C \)). Implicit in this model is the fact that the group-specific curves take on the same basis as employed for each of the subject-specific curves. Therefore, the decisions that are associated with the number and location of inner knots in \( \xi \) influence both sets of B-spline coefficients (\( \beta_j \) and \( \theta_h \)).

Information that can guide inner knot selection is rarely known but is crucial to producing an attractive curve without overfitting. To overcome this problem, we select a fixed number of equally spaced knots within the time domain and employ the Bayesian penalized spline (P-spline) technology of Lang and Brezger (2004). Bayesian P-splines are the Bayesian analogue of B-splines penalized by \( b \)-order differences (Eilers and Marx, 1996) and are constructed around \( b \)-order Gaussian random walks (Lang and Brezger, 2004). We opt to employ this technology at the group level which results in smooth group-specific derivative curves and allow individual derivative curves to vary flexibly around their group counterpart (see equation (6)). Now, a first-order Gaussian random walk for \( \theta_h \) (i.e. \( b = 1 \)) is defined as

\[
\theta_{hl} = \theta_{hl-1} + \nu_{hl}, \quad \text{for } l = 2, \ldots, q \text{ and } h = 1, \ldots, C,
\]

with \( \nu_{hl} \sim N(0, \tau_h^2) \). Typically \( p(\theta_{h1}) \propto 1 \) and this together with the previous equation produces the following hierarchical prior for \( \theta_h \):

\[
p(\theta_h | \tau_h^2) \propto \exp\left( -\frac{1}{\tau_h^2} \theta_h'K\theta_h \right),
\]

\[
\tau_h^2 \sim IG(a_\tau, b_\tau),
\]

where \( K \) is a known penalty matrix whose entries are determined by the random-walk order and \( IG(\cdot, \cdot) \) denotes an inverse gamma distribution with rate \( b_\tau \). The newly introduced penalized complexity priors of Ventrucci and Rue (2016) could provide a more intuitively appealing prior specification for \( \tau_h^2 \) (compared with the inverse gamma distribution). However, since the penalized complexity prior cannot be employed as currently constructed in the hierarchical model we opt to follow suggestions that were made in Lang and Brezger (2004).

Regarding \( \lambda^2_j \) (see equation (6)), we assign \( \lambda_h \sim \text{UN}(\lambda_h;0,A) \), where \( \text{UN}(\cdot;0,A) \) denotes a uniform density on interval \((0,A)\) and \( A \) is a user-supplied upper bound on the standard deviation of \( \beta_j \). This parameter regulates the amount of flexibility that is afforded subject-specific curves to vary around group-specific mean curves. For \( \sigma^2_j \) we assign the commonly used
conjugate prior $\sigma_j^2 \sim \text{IG}(a_\sigma, b_\sigma)$ with $a_\sigma$ and $b_\sigma$ being user-supplied values and $b_\sigma$ corresponding to the rate.

To finish the model specification, prior distributions for $D_j$ and $u_j$ (parameters related to quotient differences) need to be assigned. For $D_j$ we employ

$$\Pr(D_j = k) = \pi_k, \quad \text{for } k = 1, \ldots, \lceil n/2 - 1 \rceil,$$

where $\Sigma_{k=1}^{\lfloor n/2 - 1 \rfloor} \pi_k = 1$. To optimize the variance–bias trade-off (as $D_j$ increases, variance decreases but bias increases; see web appendix A), the $\pi_k$s should favour smaller values of $D_j$ a priori for derivative curves that are wiggly and moderate values of $D_j$ for those that are less up and down. Selecting an $\lceil n/2 - 1 \rceil$-dimensional sequence of $\pi_k$s that reflect this desired characteristic would be challenging. One way of simplifying the prior elicitation while maintaining the desire to place the majority of the prior mass on small or moderate values of $D_j$ with diminishing yet appreciable prior mass on large $D_j$ values is to employ stick breaking ideas (Sethuraman, 1994; Ongaro and Cattaneo, 2004). More specifically, set $\pi_1 = p$, $\pi_2 = p(1 - \pi_1) = p(1 - p)$, $\pi_3 = p(1 - \pi_2)(1 - \pi_1) = p(1 - p)^2$, etc. Thus, for an arbitrary $k$, $\pi_k = p\Pi_{j<k}(1 - \pi_j) = p(1 - p)^{k-1}$.

To ensure that the $\pi_k$s sum to 1, we set $\pi_{\lceil n/2 - 1 \rceil} = \Pi_{k=1}^{\lceil n/2 - 1 \rceil} (1 - \pi_k) = (1 - p)^{\lceil n/2 - 1 \rceil}$. Now instead of selecting $\lceil n/2 - 1 \rceil$ $\pi_k$-values it is necessary to select only a value for $p$ (the probability that $D_j = 1$). Note that as $p$ approaches 1 the sequence of empirical derivatives becomes more local, making it more variable but with less bias (essentially approaching estimate (2)). Regarding $u_j$, we assume $u_j \sim \text{gamma}(a_u, b_u)$ where $b_u$ is a scale parameter (i.e. $E(u_j) = a_u b_u$). For derivative curves that are wiggly small values of $u_j$ are desirable (i.e. small values of $a_u$ or $b_u$ or both) as it creates a sequence of empirical derivatives that is local, whereas for less wiggly derivative curves a more global sequence and thus large values for $u_j$ are preferable (i.e. large values of $a_u$ or $b_u$ or both). The effect that values selected for $p$, $a_u$ and $b_u$ have on model fit is explored extensively in the simulation that is provided in Section 3 and web appendix C. Finally, as a means to visualize all the moving parts of the model, we provide it in its entirety:

$$y^{(1)}_j (D_j, u_j) = B_j \beta_j + \epsilon_j, \quad \epsilon_j \sim N(0, \sigma_j^2 I), \quad \sigma_j^2 \sim \text{IG}(a_\sigma, b_\sigma),$$

$$\beta_j | g_j \sim N(\theta_{g_j}, \lambda_{g_j}^2 I), \quad \lambda_h \sim \text{UN}(\lambda_h; 0, A),$$

$$p(\theta_h | \tau_h^2) \propto \exp\left(-\frac{1}{\tau_h} \theta_h^\top K \theta_h\right),$$

$$\tau_h^2 \sim \text{IG}(a_\tau, b_\tau),$$

$$\Pr(D_j = k) = p(1 - p)^{k-1},$$

$$u_j \sim \text{gamma}(a_u, b_u).$$

2.3. Estimation of subject-specific curves in unbalanced designs

The full conditional of $\beta_j$ gives insight into how the local property of the $B$-spline basis accommodates subjects with an unequal number of measurements when estimating $\beta_j$. Through a little matrix algebra and using well-known arguments it can be shown that the mean of the full conditional of $\beta_j$ is the following $q$-dimensional vector (here $E[\beta_j|\cdot]$ denotes expectation with respect to the distribution of $\beta_j$ given all other unknowns and $y_j$):

$$E[\beta_j|\cdot] = \left(\frac{1}{\sigma_j^2} B_j^\top B_j + \frac{1}{\lambda_{g_j}^2} I_q\right)^{-1} \left\{ \frac{1}{\sigma_j^2} y^{(1)}_j (D_j, u_j) B_j + \frac{1}{\lambda_{g_j}^2} \theta_{g_j} \right\}.$$

Because of the local structure of the $B$-spline basis, all columns of $B_j$ that are associated with
Modelling of Growth Curve Derivatives

$B$-spline basis functions based on knots that are beyond each of the $t_{1j}, \ldots, t_{nj}$ are simply zero vectors. Therefore, the entries of $y_j^{(1)}(D_j, u_j)'B_j$ that correspond to the zero columns of $B_j$ are also 0 and, as a result, the corresponding entries of $E[\beta_j|\cdot]$ are completely informed by $\theta_{gj}$. Thus, the compact support of the $B$-spline basis carries out a very natural updating scheme for $\beta_j$, mainly that, if measurements exist, use them to estimate the corresponding coefficients in $\beta_j$; otherwise, use the group coefficients $\theta_{gj}$.

2.4. Computation

The full posterior distribution $p(D, u, \beta, \theta, \sigma^2, \tau^2|y)$ is not analytically tractable; therefore we resort to numerical techniques based on Markov chain Monte Carlo (MCMC) sampling to sample from it. The algorithm that we develop is a Gibbs sampler with Metropolis steps. Exact details that are associated with updating each of the parameters are provided in appendix B. We briefly mention that updating $D_j$ and $u_j$ requires a little more care as their full conditionals are not of recognizable form. We employ a random-walk Metropolis step for $u_j$ and an independent Metropolis step for $D_j$. The algorithm behaved well with good mixing and rapid convergence. See web appendix B for more details.

2.5. Accommodating missing response values

The hierarchical model that is specified in expression (7) assumes that a response is recorded for each subject at all planned measurement times, i.e. what Daniels and Hogan (2008) called the full data are the same as the observed data. However, it is common in longitudinal-type studies that the full data are not available. In the presence of missing response values, it would be necessary to add a level to expression (7) that models the full data (i.e. observed responses, missing responses and the missingness mechanism). Daniels and Hogan (2008) described some possible models for an array of missingness mechanisms. Since the full data model can be specified independently of expression (7) our fully Bayesian approach provides benefit as no new inference techniques are needed (see Ibrahim et al. (2005)). Within a Bayesian framework the missing observations are treated as unknowns and as a result would be updated accordingly within the MCMC algorithm that is detailed in web appendix B. Thus, in the presence of missing response values, our procedure operates as before except that the sequence of empirical derivatives is constructed with a mix of observed response values and updated values for missing observations.

3. Simulation studies

To study the performance of our derivative curve estimation procedure, we conducted two separate simulation studies. The first was designed to study the behaviour of the new sequences of empirical derivatives that were described in Section 2.1, whereas the second explored the performance of our hierarchical modelling approach. All simulations were done using R (R Core Team, 2018). Extra simulation results can be found in the on-line supporting material.

3.1. Simulation 1: derivative curve estimation based on sequences of empirical derivatives

This subsection details a simulation study that was carried out to investigate the utility of our approach in estimating derivative curves based on the sequence of empirical derivatives that is developed in this paper. The main aim of this study was twofold:

(a) to explore how values that are selected for $p$, $a_u$ and $b_u$ influence derivative curve fit and
(b) to compare our approach with established alternatives.
Because potential competitors were not developed in a hierarchical model framework, this simulation study focused on estimating a single derivative curve. This required using the following simplified version of the Bayesian hierarchical model (7):

\[
\begin{align*}
\mathbf{y}^{(1)}(D, u) &= \mathbf{B}\mathbf{\beta} + \mathbf{\epsilon}, \\
\mathbf{\epsilon} &\sim N(0, \sigma^2 I), \\
\mathbf{\beta} &\sim \text{IG}(a_{\beta}, b_{\beta}), \\
p(\mathbf{\beta}|\tau^2) &\propto \exp\left(-\frac{1}{2\tau^2}\mathbf{\beta}'\mathbf{K}\mathbf{\beta}\right), \\
\tau^2 &\sim \text{IG}(a_{\tau}, b_{\tau}), \\
Pr(D = k) &= p(1 - p)^{k-1}, \\
u &\sim \text{gamma}(a_u, b_u).
\end{align*}
\]

Model (9) was fitted by using the sequences of quotient differences that are proposed in De Brabanter et al. (2013), which is not based on empirical derivatives. To compare directly the performance of the sequences of quotient differences we propose those described in De Brabanter et al. (2013) in two ways. The first employed the same Bayesian model as detailed above, and the second used the kernel-based estimation method that was suggested in De Brabanter et al. (2013). In both cases, the weights for constructing the sequence were those proposed in De Brabanter et al. (2013).

Since our overarching goal is to model derivative curves hierarchically we are mainly interested here in showing that our sequence of empirical derivatives coupled with model (9) produces derivative curve estimates that are competitive relative to competitors for a number of combinations of \(p, a_u\) and \(b_u\). Showing this would imply that extending the procedure to a hierarchical setting would be reasonable, which would be appealing as extending other existing methods (e.g. local polynomials) to a hierarchical setting is not obvious.

### 3.1.1. Scenarios and set-up

We considered the following two data-generating mechanisms:

(a) \(y_t = t^2 + N(0, 3^2)\) for \(t \in [-1, 1]\) (thus \(f(t) = t^3\) and \(f^{(1)}(t) = 3t^2\));

(b) \(y_t = \sin(t) + N(0, s^2)\) for \(t \in [-5, 5]\) (thus \(f(t) = \sin(t)\) and \(f^{(1)}(t) = \cos(t)\)).

These functions were considered as the derivative in (a) is smoother than that in (b). Using each data-generating mechanism, we created data sets with \(n \in \{20, 50, 100\}\) equally spaced time points. To study how noise relative to signal influences derivative fits we set \(s \in \{0.1, 0.5\}\). When fitting the model we fixed \(p \in \{0.1, 0.5, 0.9\}\) along with \(a_u \in \{0.1, 1, 10\}\) and \(b_u \in \{0.1, 1, 10\}\).

For each factor combination (i.e. \(f(\cdot), n, s, p, a_u\) and \(b_u\)) 100 data sets were generated. For each data set generated, model (9) was fitted by collecting 2000 MCMC draws after discarding the first 1000 as burn-in. For the \(P\)-spline model described in equation (9), we used a second-order Gaussian random walk and 40 evenly spaced knots in the domain of \(t\). We set \(a_{\beta} = b_{\beta} = a_{\tau} = 1\). As mentioned by Jullion and Lambert (2007), selecting appropriate values for \(b_{\tau}\) depends on sample size, signal-to-noise ratio and the knot configuration. In an attempt to remove the effect that this prior has when making comparisons across data-generating scenarios, we specify \(b_{\tau}\) so that the induced prior on the effective degrees of freedom is centred approximately at 4 with standard deviation about 1.5 for the cubic scenario and 10 with standard deviation 3.8 for the sine scenario. Exact values of \(b_{\tau}\) are provided in Table S.1 of the on-line supplementary material. For the kernel-based estimation procedure that was described in De Brabanter et al. (2013),
3.1.2. Results of simulation study 1

Derivative fit was assessed by computing the integrated mean-squared error (IMSE) (i.e. 
\(\frac{1}{n} \sum_{i=1}^{n} \left\{ f^{(1)}(t_i) - \hat{f}^{(1)}(t_i) \right\}^2 \)). To assess performance of procedures near the boundaries, we calculated the IMSE for the interior 90% of observations separately from the outer 10%. For derivative fits that employed Bayesian model (9), the average pointwise 95% credible interval width and average pointwise coverage were also calculated. For ease of exposition, the majority of numerical and graphical results are provided in web appendix C, and we focus here on discussing the main findings in the case that \(a_u = 1\) (similar patterns arose for other values of \(a_u\)). With respect to coverage and credible interval width (Tables S.2 and S.3 in web appendix C), it appears that for many of the prior value combinations the two sequences of quotient differences that are proposed in this paper (\(y^{(1)}(D, u)\) and \(y^{(1)}(u)\)) perform better in terms of coverage with comparable interval widths than that proposed by De Brabanter et al. (2013). This is particularly true on the boundaries (again, see Tables S.2 and S.3 in web appendix C). Generally speaking, the performance of the procedure when using \(y^{(1)}(u)\) worsens as \(b_u\) increases. The results also suggest that increasing \(p\) when using \(y^{(1)}(D, u)\) mitigates this effect.

Regarding the results that are associated with the IMSE (Figs S.2–S.5 in web appendix C), large values of \(p\) seem to provide benefit for the sine function scenario, but not the cubic function scenario. As with coverage, for \(y^{(1)}(u)\), large \(b_u\)-values seem to impact the IMSE negatively in both data-generating scenarios but has less influence for \(y^{(1)}(D, u)\) for \(p > 0.1\). Overall, the simulation study highlights the fact that for a number of combinations of \(p\) and \(b_u\) (and setting \(a_u = 1\)) using the Bayesian hierarchical model based on \(y^{(1)}(D, u)\) or \(y^{(1)}(u)\) performs better relative to the sequence that was proposed by De Brabanter et al. (2013) (using either the Bayesian model or the kernel-based approach) and is very competitive with the method that was proposed by Dai et al. (2016) and the local polynomial approach (which could be considered the non-parametric gold standard estimation procedure in the cubic scenario). To see this a little more clearly, we provide Figs 2(a) and 2(b). These figures display the true derivative curve versus the average estimated derivative curves (averaged across the 100 data sets), along with 2.5 and 97.5 simulation estimation percentiles for both scenarios. The factors selected are \(n = 50\), \(s = 0.5\), \(a_u = 1\), \(b_u = 0.1\) and \(p = 0.9\). (For other factors see Figs S.6–S.29 in web appendix C.) These plots illustrate the benefit that \(y^{(1)}(D, u)\) and \(y^{(1)}(u)\) provide in derivative curve estimation near the boundaries and overall derivative estimate. Note that, especially for the sine function scenario, derivative fits based on \(y^{(1)}(D, u)\) and \(y^{(1)}(u)\) present a lower variance (but slightly larger bias) than the alternative procedures. Although we showed improved performance on the boundaries for many combinations of \(b_u\), and \(p\), even more important to this paper is that the sequence of empirical derivatives that we develop is easily employed in a hierarchical setting, which we address in the next section.
Fig. 2. True derivative curve (——) versus the average of estimated derivative curves (---) over the 100 synthetic data sets for $n = 50$, $s = 0.5$, $a_u = 1$, $b_u = 0.1$ and $p = 0.9$ (the interval bands represent the 2.5- and 97.5-percentiles of the estimated derivative curves; $y^{(1)}(D, u)$ and $y^{(1)}(D)$ refer to the results by using the Bayesian model proposed in this paper, DAI to the proposal by Dai et al. (2016), LP to the local polynomial approach and $\bar{y}_B^{(1)}(D)$ and $\bar{y}_K^{(1)}(D)$ to the sequence of quotient differences proposed by De Brabanter et al. (2013) using respectively the Bayesian model described in this paper and the kernel-based approach proposed in De Brabanter et al. (2013)): (a)–(f) data-generating mechanism where $f(t) = t^3$ and as a result $f^{(1)}(t) = 3t^2$; (g)–(l) data-generating mechanism where $f(t) = \sin(t)$ and as a result $f^{(1)}(t) = \cos(t)$
On the basis of this simulation study (again, see web appendix C for details), we conclude that, generally speaking, our approach is fairly robust to values that are employed for $p$ and $b_u$ (after setting $a_u = 1$) so long as they are reasonably selected based on the wiggliness of the raw curves and the signal-to-noise ratio. Just as in other methods that employ regularization or smoothing parameters (i.e. smoothing splines), $b_u$ and $p$ will need to be selected case by case. That said, the simulation seems to suggest that it is preferable to err on the side of sequences of empirical derivatives that are local (which are less biased). Thus, we recommend setting $a_u = 1$ and $b_u = 1$ preliminarily and to adjust $b_u$ if necessary depending on the wiggliness of the curve. For $y^{(1)}(D, u)$ we suggest setting $p = 0.9$ as this seems to protect the sequence of empirical derivatives from becoming too global. Values that are selected for $a_\tau$ and $b_\tau$ also impact the derivative curve fits. Since their selection is not the focus of this paper, we direct interested readers to the growing literature that is dedicated to properly specifying the smoothing priors in Bayesian penalized splines (see for example Jullion and Lambert (2007)).

3.2. Simulation 2: performance of hierarchical model in estimating group derivative curves

The purpose of the second simulation was to investigate how coupling the hierarchical model (7) with the new sequences of empirical derivatives that were described in Section 2.1 performed in estimating group derivative curves. On the basis of the ALL data, we created synthetic data sets in the following way. For each treatment, the individual measured growth curves were used to compute three empirical treatment mean curves. The empirical mean curves were calculated by using means that were computed cross-sectionally after grouping growth measurements into 20 time intervals or bins (this was done since subjects were not all measured at the same time points). The empirical mean curves were then fitted to a Richards model (or generalized logistic curve) (Richards, 1959) to produce 'theoretical' treatment–group curves. This was done by using non-linear least squares. A Richards model was used as it fitted the empirical treatment curves well. Then synthetic growth curves for subjects were generated by using a Gaussian process with an exponential covariance function and the estimated theoretical treatment curves as mean functions. The range parameter of the exponential covariance function was set to 10 and the variance to 4. These values ensured that the synthetic growth curves were essentially monotone.

As a means to visualize the type of data that were generated in the simulation, we provide Fig. 3. As suggested by the first simulation study, for this study we fixed $a_u = 1$ and considered $b_u \in \{0.01, 0.1, 1, 10, 100\}$ and $p \in \{0.1, 0.9\}$. We set $a_\tau = 1$ and $b_\tau = 2$, as the derivative curves were similar to those from the sine scenario from the previous simulation. Lastly we considered $A = \{1, 5\}$. Using the data-generating mechanism just described, we created synthetic data sets with $m = 30$ subjects per group and $n \in \{20, 50\}$ observations per subject, at equally spaced time points. In addition to generating data that are balanced, we also considered unbalanced data (a characteristic of the ALL data). This was done by randomly selecting 50% of the subjects and then removing a random number of time points. The number of time points to be removed was determined by using a shifted Poisson distribution (shifted up by 1) with a mean of 4. To each data set generated, we fit model (7) by collecting 1000 MCMC samples after discarding the first 3000 as burn-in and thinning by 2. For each factor combination (i.e. $n$, $p$ and $b_u$) and balanced or unbalanced design, we run a total of 100 repetitions. Performance was assessed by computing the IMSE, the width of the credible bands and the coverage of the credible bands for the estimated group curves.

Results for the unbalanced case can be found in Fig. 4 (the balanced case produced very similar results). It seems that we can recover well (in terms of coverage) the group curves for
Fig. 3. (a) Example of the synthetic data from the generalized logistic model used in the simulation study along with (b) the true group growth curves and (c) derivative curves.
small values of $b_u$ and large $p$ at the cost of credible interval width. This implies that small variance–high bias estimates when $b_u$ is large (and $p$ is small) are not desirable. It seems that $A$ does not impact estimation of group curves. This is to be expected as the effect of $A$ is felt at the subject curve level. The upshot regarding the simulation study is that for specific values of $b_u$ and $p$ our methodology does well in recovering group curves in settings like that of the ALL data.

4. Applications

We now consider the ALL data that were introduced in Section 1. To compare the growth curve derivatives of the ALL study subjects with those of healthy subjects, we also consider growth
curve derivative estimation by using the well-known Berkeley growth study. This data set is publicly available in the \texttt{fda} package by Ramsay \textit{et al.} (2014) in R.

4.1. Growth curve derivatives for Berkeley growth study
The Berkeley growth study consists of 93 subjects, 39 of whom are male and 54 female. On each subject 31 growth measurements were taken. The first was taken during the first year of life and the last during the 18th year. Fig. 5(a) displays each subject-specific growth curve.

We fitted the hierarchical model that was detailed in the previous section to these data by using the sequence of quotient differences $y^{(1)}(D, u)$. For the $P$-spline model we considered a second-order Gaussian random walk and 40 evenly spaced knots. Further, we set $a_\sigma = b_\sigma = 1$ and, since these growth curves are more similar to the sine scenario of the first simulation, set

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5}
\caption{For the Berkeley growth study ( girls; boys), (a) growth curves for the 93 subjects in the study, (b) estimated subject-specific growth curve derivatives based on the Bayesian hierarchical model, (c) derivative curve estimates for each gender with pointwise error bands, and (d) estimate of differences between gender-specific derivative curves.}
\end{figure}
\( a_\tau = 1 \) and \( b_\tau = 1/0.5 \). The MCMC algorithm was used to collect 1000 MCMC iterates after discarding the first 10000 as burn-in and thinning by 10. Guided by the simulation study of Section 3.2 we set \( p = 0.9, a_u = b_u = 1 \) and \( A = 1.0 \). We fitted the hierarchical model by using other specified prior values and found derivative curve fits to be robust.

The results are provided in Fig. 5. Fig. 5(b) contains the subject-specific derivative curve estimates and Fig. 5(c) displays gender-specific derivative curve estimates. As expected, the rate of growth decreases as toddlers approach pubescence, regardless of gender. As individuals enter pubescence, gender differences in the rate of growth become evident as females enter the pubescent growth spurt earlier than males and that of males lasts longer. This pattern is quite stable for both individual and gender-specific derivative curves: something that is absent in the ALL data that are considered next. Differences between gender-specific derivative curves are more clearly seen in Fig. 5(d). This plot contains the difference of the derivative curves and displays the sharp change in growth rate difference between males and females between 12 and 15 years of age.

4.2. Growth curve derivatives for childhood acute lymphoblastic leukaemia

We now turn our attention to the ALL study. Fig. 6 is a replica of Fig. 1 except that we highlight nine individuals who will be used to illustrate and communicate results. These data correspond to 197 females who were diagnosed with childhood ALL between 2 and 9 years of age. Growth was measured for each subject at unequal intervals with a total of 1988 observations. Note from Fig. 6 that, unlike the Berkeley growth study, in the ALL study not all subjects recorded the same number of growth measurements (the number of observations per subject ranges from 1 to 21). Since estimating derivative curves based on one or two measurements is not possible, we include only subjects with at least three growth measurements. Also, for five subjects, multiple height measurements were recorded at the same age. When this occurred, we used the average height. This results in 1968 observations for 186 subjects.

We fitted the model based on \( y^{(1)}(D, u) \) to the childhood ALL data by collecting 1000 draws after discarding the initial 10000 as burn-in and thinning by 5. Since there are many subjects for the ALL data with fewer than 10 observations, we used a second-order Gaussian random walk with only 20 evenly spaced knots. Motivated by the simulation studies, we set \( a_\tau = 1, b_\tau = 1/0.5, p = 0.9 \) and \( a_u = b_u = 1 \). As before, \( a_\sigma = b_\sigma = 1 \). Since subject curves are more variable here than in the Berkeley data, we set \( A = 2 \) to allow subject derivative curves added flexibility to vary around the mean derivative curve. The estimated subject-specific derivative curves are provided in Figs 6(d)–6(f) whereas the treatment-specific derivative curves are in Fig. 7.

From Figs 6(d)–6(f) the first thing to note is that, regardless of treatment, females who were diagnosed with ALL and assigned one of the treatments have drastically different growth rate curves compared with females in the Berkeley study even though the raw growth curves appear very similar (compare Figs 1 and 5). In fact, the estimated derivative curves in Fig. 6 display patterns that are very foreign to expected growth velocity curves (as seen in Fig. 5). That said, the estimated derivative curves of the nine subjects who are highlighted in Figs 6(d)–6(f) are very reasonable when comparing them with the actual changes of growth rate that exist visually over time in the raw growth curves.

We compare the nine highlighted estimated derivative curves in Fig. 5 with those obtained by the local polynomial approach (Fan and Gijbels, 1996) applied to each subject independently. The results are found in Fig. 8 and verify the general shape of the derivative curves. However, the hierarchical model provides more reasonable fits on the boundaries and slightly smoother derivative curves relative to employing the local polynomial approach to each subject independently.
Fig. 6. For the childhood ALL study, raw subject-specific growth curves for individuals separated by the three ALL treatments (a) no radiation, (b) standard radiation and (c) hyperfractionated radiation, and the individual derivative curves estimates based on the Bayesian hierarchical model for the ALL treatments (d) no radiation, (e) standard radiation and (f) hyperfractionated radiation: nine subjects are highlighted to illustrate how variable derivative curves can be even when growth curves are quite similar.

From Fig. 7 there are clear differences in the treatment-specific derivative curves with no radiation treatment having more accelerated growth initially and maintaining growth longer (something that is expected). The double radiation therapy seems to impact growth rate negatively more than the standard radiation therapy as it peaks earlier and declines faster. These differences are further highlighted in Fig. 7(e) where the two radiation therapies are compared with the non-radiation therapy. Note that differences between the two radiation therapies and non-radiation therapy are more pronounced during years when the preadolescent growth spurt typically begins for females.

The estimated derivative curve shape seems to corroborate studies that were conducted in developed countries that have shown that children who are diagnosed with ALL and prescribed radiation-type therapies have deceleration of growth in the initial phase of therapy (Ahmed et al., 1997). In fact, for these individuals, the growth rate continues to increase up through
For the childhood ALL study, (a)–(d) treatment-specific derivative curve (rate of change) estimates (---, ---, ---) with pointwise error bands (- - - - - - - -) (---, ---, ---, ---, ---, ---), intrathecal therapy with no radiation; --- intrathecal therapy with standard radiation; ---, ---, ---, --- intrathecal therapy with twice-daily radiation (hyperfractionated)): (e) comparison of standard and double radiation treatments with radiation-free treatment.
Fig. 8. For the childhood ALL study and for each of the nine subjects highlighted in Fig. 6, estimated empirical differences $y^{(1)}(D, u)$ (●) with estimated derivative curves by using the Bayesian hierarchical model (●) and the local polynomial approach (●) (for the local polynomial approach, the derivative curves were estimated independently for each subject): (a) subject 3; (b) subject 18; (c) subject 20; (d) subject 27; (e) subject 56; (f) subject 80; (g) subject 109; (h) subject 122; (i) subject 134
year 5 albeit at a much slower rate. Therefore patients with childhood ALL had accelerated growth relative to healthy subjects, but they still grew much less.

5. Discussion

We have developed a statistical methodology that takes advantage of the hierarchical structure that typically exists in growth studies to estimate subject-specific derivative curves and then, borrowing strength among them, estimate group-specific derivative curves. The methodology that was developed is based on a new approach to constructing sequences of quotient differences (empirical derivative estimates) that incorporates both forward and backward differences. We showed through simulations that this new approach is better able to balance variance–bias that is inherent in sequences of empirical derivatives and greatly reduces boundary bias relative to central differences.

Employing the methodology showed that considering derivative curves in growth studies provides valuable information beyond that available in the raw growth curves. In particular, raw growth curves that are produced by individuals in the Berkeley study appeared to be fairly similar to those corresponding to individuals who had been diagnosed with ALL, but the growth rate curves were very dissimilar. In addition, the treatment-specific derivative curves tended to corroborate results found in Durban et al. (2005) while highlighting time points where differences between radiation treatments occur.

Even though the methodology that is presented here was motivated by a specific application of estimating the first derivative of growth curves, it can provide utility in any application or field that considers derivatives and routinely produces data with a hierarchical structure (i.e. biology, physics or chemistry). Additionally, it is completely plausible that higher order derivatives are also valuable (e.g. mass spectrometry). The methodology that we developed can be straightforwardly extended to incorporate higher order derivatives (something explored in De Brabanter et al. (2013)). Although not formally considered in our application, it seems reasonable that the differences that are seen in the derivative curves (and absent in raw curves) can be useful when there is interest in curve clustering. It is interesting also that the derivative curves seem to discriminate between genders more so than the raw curves (although the scale between them is much different). As a result, using derivative information for classification purposes seems promising and is the focus of future research.

Lastly, an R package (HDCurves) containing functions that were employed to fit the hierarchical model described in this paper can be download from the Comprehensive R Archive Network.

Acknowledgements

We would like to thank María Durban for sharing the growth curve ALL data and Wenlin Dai for sharing the R code to perform the comparisons in Section 3.1. This research was supported by the Basque Government through the ‘BERC 2018-2021 program’, by the Spanish Ministry of Science, Innovation and Universities through ‘BCAM Severo Ochoa accreditation SEV-2017-0718’ and through project MTM2017-82379-R funded by ‘(AEI/FEDER, UE)’.

References

Ahmed, S., Wallace, W. and Kelnar, C. (1997) An anthropometric study of children during intensive chemotherapy for acute lymphoblastic leukemia. Horm. Res. Paediatr., 48, 178–188.
Behseta, S., Kass, R. K. and Wallstrom, G. L. (2005) Hierarchical models for assessing variability among functions. *Biometrika*, 92, 419–434.

Botts, C. and Daniels, M. (2008) A flexible approach to Bayesian multiple curve fitting. *Computat Statist. Data Anal.*, 51, 5100–5120.

Brumback, B. and Rice, J. (1998) Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Statist. Ass.*, 93, 961–994.

Cabrera, J. L. O. (2012) locpol: kernel local polynomial regression. *R Package Version 0.6-0*. (Available from https://CRAN.R-project.org/package=locpol.)

Charnigo, R., Hall, B. and Srinivasan, C. (2011) A generalized Cp criterion for derivative estimation. *Technometrics*, 53, 238–253.

Crainiceanu, C., Ruppert, D. and Wand, W. (2005) Bayesian analysis for penalized spline regression using WinBUGS. *J. Statist. Softw.*, 14, 1–24.

Dai, W., Tong, T. and Genton, M. G. (2016) Optimal estimation of derivatives in nonparametric regression. *J. Mach. Learn. Res.*, 17, 1–25.

Dalton, V. K., Rue, M., Silverman, L. B., Gelber, R. D., Asselin, B. L., Barr, R. D., Clavell, L. B., Gelber, R. D., Asselin, B. L., Barr, R. D., Clavell, L. A., Hurwitz, C. A., Moghrabi, A., Samson, Y., Schorin, M., Tarbell, N. J., Sallan, S. E. and Cohen, L. E. (2003) Height and weight in children treated for acute lymphoblastic leukemia: relationship to CNS treatment. *J. Clin. Oncol.*, 21, 2953–2960.

Daniels, M. J. and Hogan, J. W. (2008) *Missing Data in Longitudinal Studies Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton: Chapman and Hall.

De Brabanter, K., De Brabanter, J. and De Moor, B. (2011) Nonparametric derivative estimation. In *Proc. 23rd Benelux Conf. Artificial Intelligence, Gent* (eds P. De Causmaecker, J. Maevoet, T. Messelis, K. Verbeek and T. Vermeulen), pp. 75–81. Drongen: Nevelland.

De Brabanter, K., De Brabanter, J., De Moor, B. and Gijbels, I. (2013) Derivative estimation with local polynomial fitting. *J. Mach. Learn. Res.*, 14, 281–301.

De Brabanter, K. and Liu, Y. (2015) Smoothed nonparametric derivative estimation based on weighted difference sequences. In *Stochastic Models, Statistics and Their Applications* (eds A. Steland, E. Rafajlowicz and K. Szajowski), ch. 4, pp. 31–38. Cham: Springer.

Diggle, P., Heagerty, P., Liang, K.-Y. and Zeger, S. (2013) *Analysis of Longitudinal Data*, 2nd edn. Oxford: Oxford University Press.

Djeundje, V. A. and Currie, I. D. (2010) Appropriate covariance-specification via penalties for penalized splines in mixed models for longitudinal data. *Electron. J. Statist.*, 4, 1202–1224.

Durban, M., Harezlak, J., Wand, M. P. and Carroll, R. J. (2005) Simple fitting of subject-specific curves for longitudinal data. *Statist. Med.*, 24, 1153–1167.

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

Eilers, P., Marx, B. and Durban, M. (2015) Twenty years of P-splines. *Statist. Oper. Res. Trans.*, 39, 1–38.

Fan, J. and Gijbels, I. (1996) Local Polynomial Modelling and Its Applications. Boca Raton: Chapman and Hall.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008) *Longitudinal Data Analysis*. Boca Raton: Chapman and Hall–CRC.

Fitzmaurice, G., Laird, N. and Ware, J. (2011) *Applied Longitudinal Analysis*, 2nd edn. New York: Wiley.

Gasser, T. and Müller, H.-G. (1984) Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, 11, 171–185.

Gottardo, R. and Raftery, A. (2009) Bayesian robust transformations and variable selection: a unified approach. *Can. J. Statist.*, 37, 361–388.

Grajeda, L. M., Ivanescu, A., Saito, M., Crainiceanu, C., Jaganath, D., Gilman, R. H., Crabtree, J. E., Kelleher, D., Cabrera, L., Cama, V. and Checkley, W. (2016) Modelling subject-specific childhood growth using linear mixed-effect models with cubic regression splines. *Eur. J. Epidemiol.*, 31, 1–13.

Härdle, W. (1999) *Applied Nonparametric Regression*, 1st edn. Cambridge: Cambridge University Press.

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R. and Herring, A. H. (2005) Missing-data methods for generalized linear models. *J. Am. Statist. Ass.*, 100, 332–346.

Jullion, A. and Lambert, P. (2007) Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian p-splines models. *Computat Statist. Data Anal.*, 51, 2542–2558.

Kelley, K. and Maxwell, S. E. (2008) Delineating the average rate of change in longitudinal models. *J. Educ. Behav. Statist.*, 33, 307–332.

Lang, S. and Brezger, A. (2004) Bayesian p-splines. *J. Computat Graph. Statist.*, 13, 183–212.

Müller, H., Stadtmüller, U. and Schmitt, T. (1987) Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika*, 74, 743–749.

Ongaro, A. and Cattaneo, C. (2004) Discrete random probability measures: a general framework for nonparametric Bayesian inference. *Statist. Probab. Lett.*, 67, 33–44.

Page, G. L., Barney, B. J. and McGuire, A. T. (2013) Effect of position, usage rate, and per game minutes played on NBA player production curves. *J. Quant. Anal. Sprts*, 9, 337–345.

Page, G. L. and Quintana, F. A. (2015) Predictions based on the clustering of heterogenous functions via shape and subject-specific covariates. *Bayesian Anal.*, 10, 379–410.
Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.

Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G. (2014) fda: functional data analysis. *R Package Version 2.4.4*. (Available from https://CRAN.R-project.org/package=fda.)

R Core Team (2018) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Richards, F. J. (1959) A flexible growth function for empirical use. *J. Exptl Bot.*, 10, 290–300.

Sangalli, L. M., Secchi, P., Vantini, S. and Veneziani, A. (2009) Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines. *Appl. Statist.*, 58, 285–306.

Sethuraman, J. (1994) A constructive definition of Dirichlet priors. *Statist. Sin.*, 4, 639–650.

Simpkin, A. J., Durban, M., Lawlor, D. A., MacDonald-Wallis, C., May, M. T., Metcalfe, C. and Tilling, K. (2018) Derivative estimation for longitudinal data analysis: examining features of blood pressure measured repeatedly during pregnancy. *Statist. Med.*, 19, 2836–2854.

Song, Q. (2016) Non-parametric derivative estimation with confidence bands. *Communs Statist. Theory Meth.*, 45, 277–290.

Ventrucci, M. and Rue, H. (2016) Penalized complexity priors for degrees of freedom in Bayesian p-splines. *Statist. Modlng*, 16, 429–453.

Wang, W. and Lin, L. (2015) Derivative estimation based on difference sequence via locally weighted least squares regression. *J. Mach. Learn. Res.*, 16, 2617–2641.

Yang, J., Zhu, H., Choi, T. and Cox, D. D. (2016) Smoothing and mean-covariance estimation of functional data with a Bayesian hierarchical model. *Bayes Anal.*, 11, 649–670.

Yao, F., Müller, H. and Wang, J. (2005) Functional linear regression analysis for longitudinal data. *Ann. Statist.*, 33, 2873–2903.

Zhang, Z., McArdle, J. J. and Nesselroade, J. R. (2012) Growth rate models: emphasizing growth rate analysis through growth curve modeling. *J. Appl. Statist.*, 39, 1241–1262.

**Supporting information**

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary material’. 