Title
Bayesian Analysis of Curves Shape Variation through Registration and Regression

Author
Telesca, Donatello

Publication Date
2015-06-09

Bayesian Analysis of Curves Shape Variation through Registration and Regression.

Donatello Telesca
Department of Biostatistics, University of California Los Angeles

April 4, 2015

Abstract

This manuscript reviews the use of Bayesian hierarchical curve registration in Biostatistics and Bioinformatics. Several models allowing for unit-specific random time scales are discussed and applied to longitudinal data arising in biomedicine, pharmacokinetics and time-course genomics. We consider representations of random functionals based on P-spline priors. Under this framework, straightforward posterior simulation strategies are outlined for inference. Beyond curve registration, we discuss joint regression modeling of both random effects and population level functional quantities. Finally, the use of mixture priors is discussed in the setting of differential expression analysis.

1 Introduction

Longitudinal studies in Biostatistics often aim to characterize time-dependent dynamics associated with the evolution of specific biological or bio-behavioral processes. Several examples are reported, for example, in Pinheiro and Bates (2000). A more comprehensive treatment of statistical analysis strategies for longitudinal designs has been discussed by Wakefield (2012).

In cases where observed outcomes arise as the realization of non-linear stochastic processes, some care is needed in the characterization of its variability. In particular, it is reasonable to expect that outcomes will be observed over unit-specific random time scales, resulting in phase-varying random curves. Ignoring phase variability, may lead to inconsistent estimates of time-dependent quantities (Kneip and Gasser 1992; Kneip and Gasser 1988), as well as hard to interpret inferential summaries.

We illustrate this point by describing two simple studies and aiming to provide a simple point estimate of the mean over time. Figure 1 (a), reports the growth velocity, intended as the yearly change in height, for a sample of 39 boys and 54 girls from the Berkeley growth study (Tuddenham and Snyder 1954). We observe an overall deceleration trend in growth from infancy to adulthood, with acceleration-deceleration pulses in velocity. In particular, the most prominent velocity pulse corresponds to the pubertal spurt. Even though children experience a similar sequence of hormonal events affecting growth, such events do not occur at the same rate/time in all children. Ignoring the individual timing of growth pulses, a naïve point estimate of the average growth profile is the cross-sectional mean. Clearly, this estimate appears immediately inadequate, as it misrepresents the amplitude and length of typical pubertal growth spurts.

Figure 1 (b), shows the blood concentration trajectories of the drug Remifentanil for 65 post-surgical patients receiving i.v. infusion of the drug for up to 20 minutes until reaching a target sedation level. As the drug is infused for a length of time which varies across patients, we observe pharmacokinetic (PK) profiles over subject-specific time scales. This case study illustrates even more transparently the inadequacy of the cross-sectional mean as an estimate of average PK dynamics. In particular, timing artifacts in the study design seem to induce a two-phase excretion rate in the mean, which is not justified from a physiological perspective and clearly atypical when compared to subject-level PK profiles.

These illustrative examples highlight how different experimental and ob-

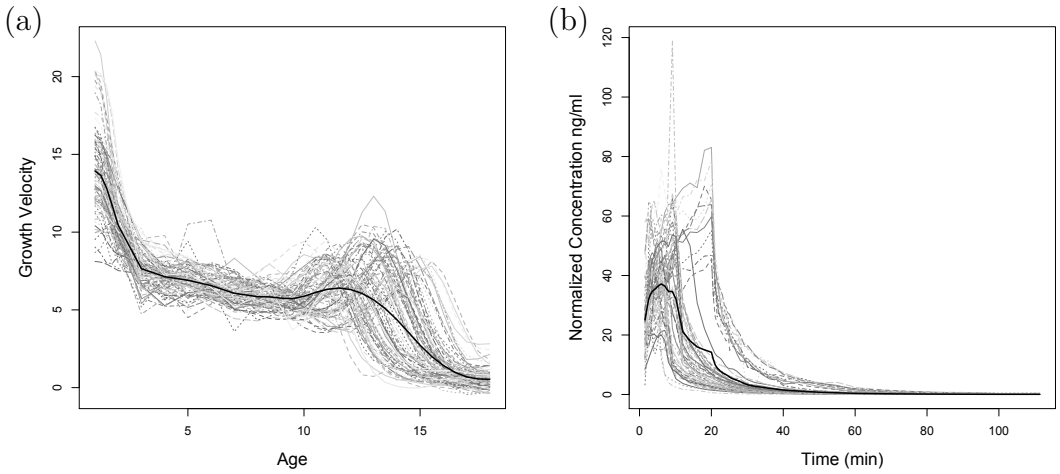

Figure 1: **Phase variability of non-linear profiles.** (a) Berkeley growth study: growth velocity for 93 subjects, defined as yearly changes in the subjects height. The cross-sectional mean growth velocity is superimposed. (b) Pharmacokinetics of Remifentanil: normalized drug blood concentration (ng/ml) monitored over phases of infusion and absorption. The cross-sectional mean PK curve is superimposed.

servational settings may define a differential genesis of phase variation in the observed outcome. In cases where variability in timing is related to the measurement process or the design of the study itself, technical or experimental information may be useful in devising pre-processing techniques aimed at removing timing artifacts. However, most commonly, variability in the timing of subject-level functional features relates to the very nature of the observed process and a rigorous approach to alignment strategies is needed to provide valid inference.

The structure of this manuscript is as follows. We introduce the problem of curve registration in Section 2 and review Bayesian hierarchical curve registration in Section 3. In relation to this model, we introduce a simplified regression strategy in Section 4. More involved regression approaches, based on varying-coefficient models are introduced in Section 5. Finally, in Section 6 we introduce and discuss applications of curve registration models to the analysis of time-course genomic data.

2 Phase variability and curve registration

The case studies summarized in Figure 1 illustrate how naïve estimation of a functional mean may result in inadequate inferential summaries. In order to formalize these concepts, let us define a notation for observations $y_i(t)$, as the observed outcome for subject i, $(i = 1, 2, \ldots, n)$ at time $t \in T$. Typically one only observes $\mathbf{y}_i = (y_i(t_{i1}), \ldots, y_i(t_{im_i}))'$ over a discrete sampling grid. However, for ease of notation, we entertain the possibility of observing y continuously. A standard working assumption sees observations arising as a realization of the following compound stochastic process:

$$y_i(t) = f_i(t) + \epsilon_i(t) = \mu\{u_i(t)\} + \epsilon_i(t), \tag{1}$$

where $\mu(\cdot)$ is some shape function, $u_i(\cdot)$ is a random, unit-specific time transformation function and $\epsilon_i(\cdot)$ a mean-zero, stationary stochastic process. An illustration of how random time scales $u_i(t)$ act on $\mu(t)$ is reported in Figure 2.

We are interested in estimating $\mu(\cdot)$ as a function quantifying a representative shape. In this setting, it may be tempting to estimate this quantity using the cross-sectional mean $\bar{y}(t) = \frac{1}{n}\sum_i y_i(t)$. However, in general, $E\{y_i(t)\} \neq \mu(t)$, therefore $\bar{y}(t)$ is not a consistent estimator of $\mu(t)$. Similar considerations extend to estimators associated with more ambitious statistical analyses, including functional PCA (Rice and Silverman 1991) and functional regression (Guo 2002; Yao et al. 2005).

This simple observation, is perhaps a reflection of the deeper tension between marginal and conditional models in longitudinal data analysis. When interest centers on non-linear dynamics, inference naturally focuses on quantities summarizing typical time-dependent evolutions (Wakefield 2012). Therefore, explicit modeling of all random components of variation is needed for meaningful inference.

There are number of proposal that deal with the problem of phase variability. A common methodological thread aims to estimate the processes $u_i(t)$ and then compute aligned profiles $y_i^*(t) = y_i\left\{u_i^{(-1)}(t)\right\}$. The shape function $\mu(t)$ is then estimated by the structural mean $\hat{\mu}(t) = \bar{y}^*(t)$.

This procedure goes under the name of curve registration, also known as curve alignment in biology, or time warping in the engineering literature. Several time warping methods have been devised to date. In the engineering literature, Sakoe and Chiba (1978) pioneered a registration technique

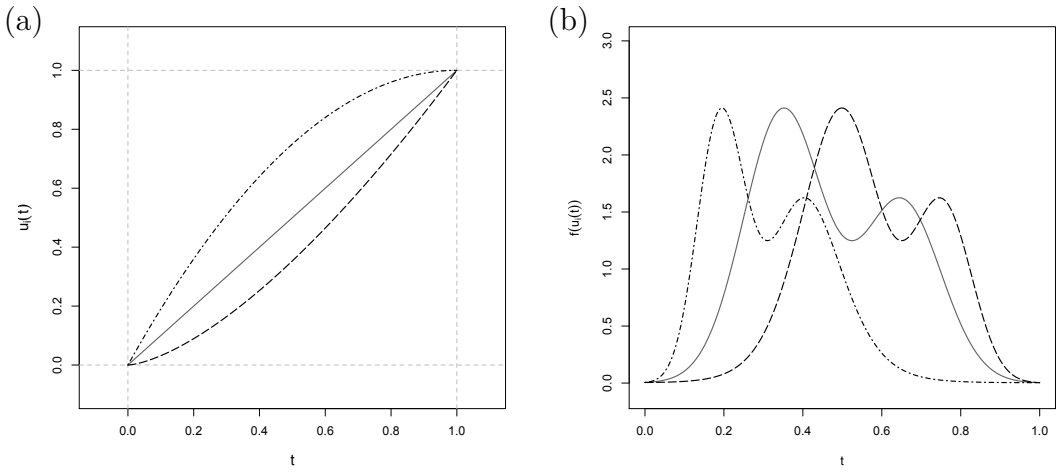

Figure 2: **Time transformation.** (a) Time transformation functions for three random profiles. In solid grey, we report the identity transform. (b) A function of time evaluated over the random time scales in (a).

called dynamic time warping for pairwise alignment. Wang and Gasser (1997) introduced the technique to the statistical literature and provided large sample properties of the time transformation estimators (Wang and Gasser 1999). Gasser and Kneip (1995) proposed the landmark registration method, which consists of identifying the timing of certain features (landmarks) in the curves. Profiles are then aligned so that they occur at the same transformed times. More recently, alignment models have focused on representing time transformation functions as continuous monotone transformations (Ramsay and Li 1998; Kneip et al. 2000). Improved estimation has been reported in Gervini and Gasser (2004); Brumback and Lindstrom (2004) as well as Liu and Müller (2004).

Bayesian approaches to the area of curve registration are more recent. Telesca and Inoue (2008) introduced a hierarchical representation of curve registration. A more recent perspective, focusing on invariance, was introduced by Cheng et al. (2013).

3 Bayesian hierarchical curve registration

The problem of curve registration admits a natural probabilistic representation in terms of hierarchical models (Telesca and Inoue 2008). A Bayesian approach to the problem confers extended flexibility in modeling both the shape $\mu(t)$ and time-transformation functions $u_i(t)$. This modeling framework is well adapted to both intensive and sparse sampling time grids as information across curves is shared via partial exchangeability assumptions. As usual, extended flexibility and a formal inferential structure come at the cost of having to specify a full probability model. If the focus of analysis is more exploratory several alternatives are available as reviewed in Section 2.

3.1 Hierarchical model

Let $y_i(t)$ denote the observed level of the i^{th} curve at time t, with $i = 1, 2, \ldots, n$ and $t \in T = [t_0, t_m] \subset \mathbb{R}$. In order to allow for linear shifts in the timing of functional features, we define an extended evaluation time window $\mathcal{T} \subset \mathbb{R}$, compact, with $T \subset \mathcal{T}$. The data generating mechanism in (1) is naturally represented via the following three–stage hierarchical model.

Stage One. Let $\mu(t)$ be a real valued function, s.t. $\mu(t) : \mathcal{T} \to \mathbb{R}$; let $c_i \in \mathbb{R}$ and $a_i > 0$, be two scalars. Furthermore, let $u_i(t)$ be a monotone smooth function, s.t. $u_i(t) : T \to \mathcal{T}$. The observed value of each curve i at time t is modeled as:

$$y_i(t) \;=\; c_i + a_i\, \mu(t) \circ u_i(t) + \epsilon_i(t) = c_i + a_i\, \mu\{u_i(t)\} + \epsilon_i(t); \qquad (2)$$

where, for any $t \in T$, $\epsilon_i(t) \sim \mathcal{N}(0, \sigma^2_\epsilon)$. In practice we only observe $y_i(t)$ over a discrete sampling time grid, which translates into standard assumptions of iid Normal random errors. In (2), $\mu(t)$ denotes a common shape function generating the individual curves and $u_i(t)$ denotes a curve-specific time transformation function. Unit-specific variability in the level and amplitude of individual profiles are modeled through c_i and a_i respectively.

Upon registration of the curves, we identify the i^{th} aligned curve at time t as

$$y^*_i(t) = y_i(t) \circ u_i^{-1}(t). \qquad (3)$$

All functional quantities are easily represented in finite dimensional form using linear combinations of appropriate basis functions. Additional considerations about specific modeling choices are deferred to Section 3.2.

Stage Two: Given a common shape function $\mu(t)$, individual curves may exhibit different scales and levels of response. Assuming $c_i \sim \mathcal{N}(0; \sigma_c^2)$ and $a_i \sim \mathcal{N}(1; \sigma_a^2) I\{a_i > 0\}$, defines a straightforward mechanism for curve-specific random affine transformations. For interpretation and identifiability purposes, it is often useful to assume $\sum_i c_i = 0$ and $\sum_i a_i = n$. Normality assumptions enable conjugacy with the likelihood. Moreover, the assumption of strictly positive amplitudes can be relaxed. For example, Telesca et al. (2009) consider a mixture prior in an application to time course expression data.

Curve–specific random time transformation functions $u_i(t)$ are assumed to be smooth realizations of a functional stochastic process, monotone increasing with probability one. Additional image and identifiability constraints are usually needed for implementation. In particular, defining $0 < \delta < (t_m - t_0)/2$, it is appropriate to require $u_i(t_0) \in [-\delta, \delta]$ and $u_i(t_m) \in [t_m - \delta, t_m + \delta]$.

Stage Three. The hierarchical model is completed with priors over population level parameters. Common assumptions exploit conditional conjugacy to define Gamma priors [1] over precision parameters:

$$1/\sigma_a^2 \sim Ga(a_a; b_a), \quad 1/\sigma_c^2 \sim Ga(a_c; b_c), \quad 1/\sigma_\epsilon^2 \sim Ga(a_\epsilon; b_\epsilon).$$

Additional priors are needed in the definition of random functional quantities. Specific choices are discussed in Section 3.2.

3.2 Penalized regression splines representation of random functionals

The shape function $\mu(t)$ is the principal object of inference in curve registration exercises. Specific parametric or semi-parametric forms may indeed be suggested by the application at hand. A general non-parametric approach is based on representing $\mu(t)$ as a random smooth function using linear combinations of B-spline basis functions (De Boor 1978).

Specifically, representation of the common shape function $\mu(t)$ may proceed selecting a set of knots $(\kappa_1, \kappa_2, \ldots, \kappa_p)$ partitioning the extended evaluation interval $\mathcal{T}$ into $p+1$ subintervals. Using piecewise polynomials of degree r and given the set of interior knots, we define $\boldsymbol{\mathcal{S}}_\mu(t)$ as a $K-$dimensional design vector of B–spline basis evaluated at time t, with $K = p + r + 1$. In

[1] In our development, $X \sim Ga(a; b)$ is parametrized so that $E[X] = a/b$.

this framework, letting $\boldsymbol{\beta}$ be a $K-$dimensional vector of basis coefficients, we represent the shape function as the following linear combination

$$\mu(t) = \mu(t; \boldsymbol{\beta}) = \boldsymbol{\mathcal{S}}'_{\mu}(t)\boldsymbol{\beta}.$$

Similarly, given a set of interior knots $(\omega_1, \omega_2, \ldots, \omega_h)$, partitioning the sampling interval T into $h+1$ subintervals, we may represent the individual time transformation functions $u_i(t)$ following the same strategy. In particular, let $\boldsymbol{\mathcal{S}}_u(t)$ be a $Q-$dimensional vector of B-spline bases of degree r evaluated at time t, with $Q = h + 1 + r$. Defining $\boldsymbol{\phi}_i$ as a Q-dimensional vector of spline coefficients, curve-specific time transformation functions may then be represented according to the following linear combination

$$u_i(t) = u_i(t;\ \boldsymbol{\phi}_i) = \boldsymbol{\mathcal{S}}'_u(t)\boldsymbol{\phi}_i.$$

Monotonicity and boundary conditions are insured by the following constraints on $\boldsymbol{\phi}_i$:

$$(t_1 - \delta) \leq \phi_{i1} < \cdots < \phi_{iq} < \phi_{i(q+1)} < \cdots < \phi_{iQ} \leq (t_m + \delta). \tag{4}$$

Similar strategies may be adopted to impose structural constraints on the form of the shape function $\mu(t; \boldsymbol{\beta})$. For an example requiring unimodality of the common shape see (Telesca et al. 2012).

The representation of functional quantities via spline bases requires choosing the degree of local spline polynomials, the number of interior knots as well as the location of the knots for both the common shape function $\mu(t; \boldsymbol{\beta})$ and the individual time transformation functions $u_i(t; \boldsymbol{\phi}_i)$. This model selection problem is often addressed with the minimization of measures of prediction error (Hastie et al. 2001) and cross–validation procedures (Gervini and Gasser 2004).

An alternative modeling strategy relies on penalized regression splines (Eilers and Marx 1996; Ruppert and Carroll 2003). Specifically, a relatively large number of equidistant knots is selected in order to purposely overparametrize the model. A penalty, dependent on a smoothing parameter λ, is then placed on coefficients of adjacent B–splines. In a frequentist framework the choice of λ is usually made in the model selection stage and is based on cross–validation analysis. From a Bayesian perspective this strategy is equivalent to the definition of appropriate dependent priors for functional coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\phi}_i$.

In particular, following Lang and Brezger (2004), one may consider a second–order random walk shrinkage prior on the shape coefficients $\boldsymbol{\beta}$, so that, for $k = 1, \ldots, K$:

$$\beta_k = 2\beta_{k-1} - \beta_{k-2} + e_k, \quad e_k \sim \mathcal{N}(0; \lambda_\beta). \tag{5}$$

Assuming $\beta_{-1} = \beta_0 = 0$, conditional on λ_β, $\boldsymbol{\beta}$ has a multivariate Normal distribution with null mean vector and precision matrix $\boldsymbol{\Omega}/\lambda_\beta$. Under the above second–order random walk, $\boldsymbol{\Omega}$ is a banded precision penalization matrix

$$\boldsymbol{\Omega} = \begin{pmatrix} 6 & -4 & 1 & & & & 0 \\ -4 & 6 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ 0 & & & & 1 & -2 & 1 \end{pmatrix}. \tag{6}$$

Note that the random walk variance λ_β can be interpreted as the smoothing parameter. In particular, small values of λ_β shrink the shape function $\mu(t, \boldsymbol{\beta})$ towards a linear function of time. Following Lang and Brezger (2004) we place a relatively diffuse conjugate inverse gamma hyperprior on the variance, so that $\lambda_\beta \sim IG(a_{\lambda 1}; b_{\lambda 1})$.

A similar approach may be adopted to model time transformation functions $u_i(t; \boldsymbol{\phi}_i)$. Defining identity transform coefficients $\boldsymbol{\Upsilon}' = (\Upsilon_1, \ldots, \Upsilon_Q)$, s.t. $u_i(t, \boldsymbol{\Upsilon}) = t$; for all $i = 1, 2, \ldots, n$, $q = 1, \ldots, Q$:

$$(\phi_{iq} - \Upsilon_q) = (\phi_{i(q-1)} - \Upsilon_{q-1}) + \eta_q, \quad \eta_q \sim \mathcal{N}(0; \lambda_\phi). \tag{7}$$

Assuming that $(\phi_{i0} - \Upsilon_0) = 0$, it can be shown that $\boldsymbol{\phi}_i \sim \mathcal{N}(\boldsymbol{\Upsilon}; \boldsymbol{P}/\lambda_\phi)$, where $\boldsymbol{P}$ is a banded precision matrix and λ_ϕ is the smoothing parameter associated with the transformation functions $u_i(t; \boldsymbol{\phi}_i)$. Small values of λ_ϕ shrink $u_i(t; \boldsymbol{\phi}_i)$ towards the identity transformation. The model is completed with a prior for λ_ϕ, s.t. $\lambda_\phi \sim IG(a_{\lambda 2}; b_{\lambda 2})$.

3.3 Inference for hierarchical curve registration models

In applications we observe functional data over a finite sampling grid $\mathbf{t}_i' = (t_{i1}, \ldots, t_{ij}, \ldots, t_{im_i})$. Let $\mathbf{y}_i(\mathbf{t}_i)' = (y_i(t_{i1}), \ldots, y_i(t_{im_i}))$ be an m_i-dimensional

vector, representing the observed trajectory for unit i, $(i = 1, \ldots, n)$, over time. Using B-spline representations, the functional model in (2) simplifies into a standard hierarchical model involving random quantities of finite dimensional form. In particular, let $\boldsymbol{\mathcal{S}}_\mu(\mathbf{t}_i) : m_i \times K$ and $\boldsymbol{\mathcal{S}}_u(\mathbf{t}_i) : m_i \times Q$, be the shape and time transformation spline design matrices respectively. The sampling model can be expressed as

$$\mathbf{y}_i(\mathbf{t}_i) = c_i \, 1_{m_i} + a_i \, \boldsymbol{\mathcal{S}}_\mu(\mathbf{t}_i)\boldsymbol{\beta} \circ \boldsymbol{\mathcal{S}}_u(\mathbf{t}_i)\boldsymbol{\phi}_i + \epsilon_i(\mathbf{t}_i); \tag{8}$$

with $\epsilon_i(\mathbf{t}_i) \sim \mathcal{N}_{m_i}(0, \sigma^2_\epsilon I_{m_i})$.

Given priors on population level quantities $\boldsymbol{\beta}$ and σ^2_ϵ, and unit-specific parameters c_i, a_i and $\boldsymbol{\phi}_i$, $(i = 1, 2, \ldots, n)$; inference about all functionals of interest is directly available from their posterior distribution. In particular, MCMC simulation from the posterior is relatively straightforward. Given $\boldsymbol{\phi}_i$, for all i, simulation from all remaining parameters is easily implemented following any sampling strategy applicable to hierarchical linear models (Gelman et al. 2013). Some care is needed in the sampling of $\boldsymbol{\phi}_i$ as the support of these parameters is defined over random cuts insuring monotonicity of time transformation functions. However, for these quantities, relatively simple Metropolis Hastings transitions tend to work well in practice. Telesca and Inoue (2008) discuss implementation of these strategies in detail.

Let $a_i^{(j)}$, $c_i^{(j)}$, $\boldsymbol{\phi}_i^{(j)}$ and $\boldsymbol{\beta}^{(j)}$, $(j = 1, \ldots, M)$, denote M draws from the marginal posterior distributions of respective parameters. To register the observed curves one may use the posterior expectation $E\{u_i(t) \mid \mathbf{y}\}$ as a point estimate of the stochastic time scale for unit i. That is, given posterior samples from time transformation parameters $\boldsymbol{\phi}_i^{(j)}, (j = 1, \ldots, M)$, posterior samples for the functional quantity $u_i(t)$ are easily calculated as

$$u_i^{(j)}(t) \;=\; u_i(t; \boldsymbol{\phi}_i^{(j)}) \;=\; \boldsymbol{\mathcal{S}}_u(t)'\boldsymbol{\phi}_i^{(j)}. \tag{9}$$

Similarly, draws from the marginal posterior distribution of the shape function $\mu(t; \boldsymbol{\beta})$, for any time $t \in \mathcal{T}$, are given by:

$$\mu^{(j)}(t; \boldsymbol{\beta}) = \boldsymbol{\mathcal{S}}_\mu(t)' \, \boldsymbol{\beta}^{(j)}. \tag{10}$$

Clearly, inference about several functional summaries, including extrema, differentials, etc., are obtained in the same straightforward fashion.

Simultaneous credible bands for any function of interest, say $f(.)$, are easily approximated using a fine grid of evaluation time points $t_1 < \ldots < t_\ell$

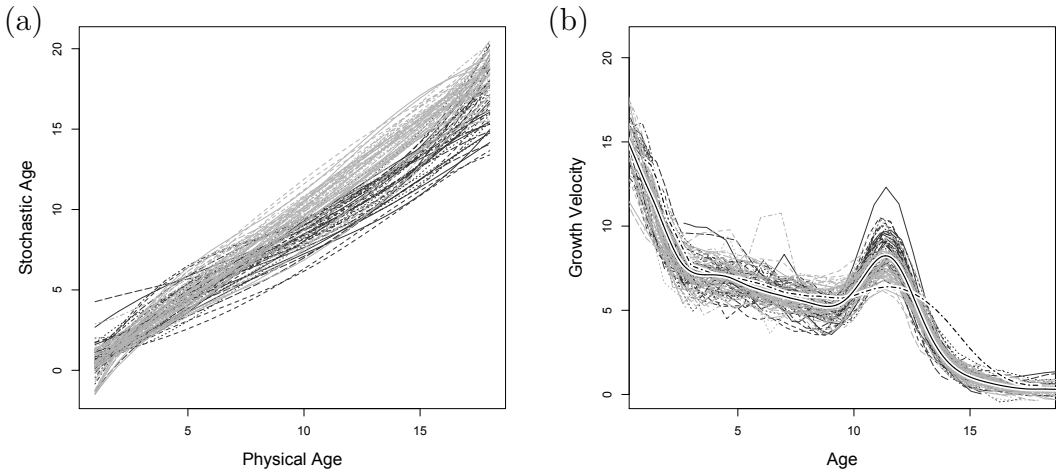

Figure 3: **Berkeley Growth Study.** (a) Posterior expectation of time transformation functions $u_i(t)$. (b) Aligned growth velocity profiles, with superimposed posterior expectation of $\mu(t)$ (solid line). The cross sectional mean for misaligned profiles is reported as the (dashed line). In both panels grey profiles identify girls, while black profiles identify boys.

(Baladandayuthapani et al. 2005). Let Γ_α denote the $100(1-\alpha)\%$ sample quantile of

$$\max_{1\leq i\leq \ell} |[f(t_i) - E\{f(t_i) \mid \mathbf{y}\}]/SD\{f(t_i) \mid \mathbf{y}\}| \,;$$

a simultaneous $100(1-\alpha)\%$ credible band for $f(t)$ is estimated as

$$I(t) = E\{f(t) \mid \mathbf{y}\} \pm \Gamma_\alpha \, SD\{f(t) \mid \mathbf{y}\}.$$

3.4 Case studies in Bayesian curve registration

We illustrate the application of Bayesian hierarchical registration techniques to the analysis of the two illustrative case studies reported in Figure 1. In particular, Figure 3 reports estimates for the posterior expected time transformation functions $u_i(t)$ (a) and aligned growth velocity profiles (b). In the same figure we report the posterior expectation for the structural average curve $\mu(t)$ (solid line). When compared to the naïve cross-sectional estimate,

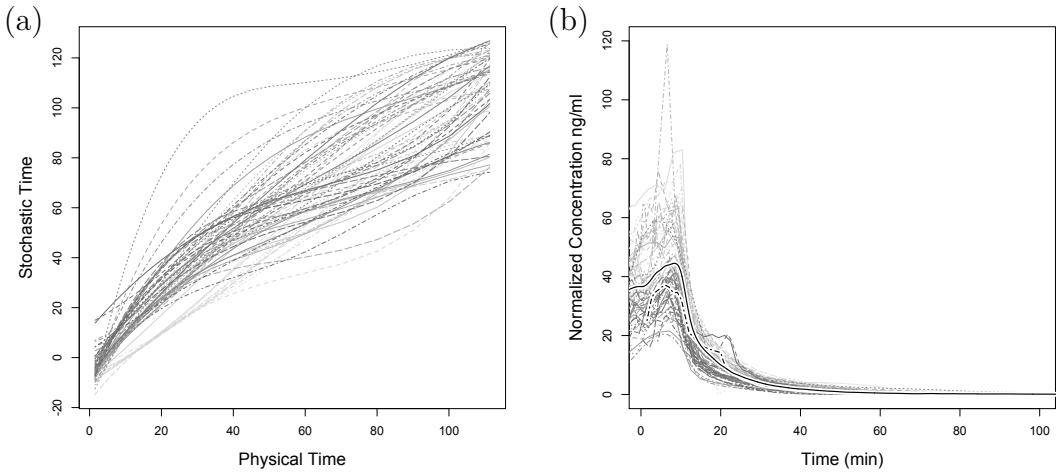

Figure 4: **Pharmacokinetincs of Remifentanil.** (a) Posterior expectation of time transfomation functions $u_i(t)$. (b) Aligned drug concentration trajectories with superimposed posterior expectation of $\mu(t)$ (solid line) and cross sectional mean of misaligned concentration dynamics (dashed line).

the structural average appears clearly as a better representation of typical growth velocity patterns.

Similarly Fig. 4, reports a hierarchical registration analysis of drug concentration dynamics associated with the drug Remifentanil. In panel (a) we plot posterior expected time transformation functions. As information about differing functional features becomes more sparse in later sampling time points, the estimated warping functions exhibit higher variance near the end of the sampling time domain. Panel (b) reports aligned drug concentration trajectories and superimposed posterior expected structural mean (solid line). As for the growth study, alignment removes artifacts in the cross sectional average and produces estimates of average concentration kinetics, which are more representative of typical individual profiles. The application of this technique to pharmacokinetic data is indeed non-standard as one often seeks to learn about compartment model parameters in a system of differential equations. Nevertheless, we find this analysis useful and essentially informative as a primitive exploration of drug concentration dynamics.

4 Regression models for timing and amplitude of functional features

Consider the growth study in Fig. 3. A more in depth look at Individual profiles is indeed made easier after removing phase variability. We code individual curves in grey-levels to reflect the subject sex. Black profiles indicate boys and grey profiles indicate girls. An explorative examination of the estimated time transformation functions (a) reveals that the time scale for girls tends to lie above the identity transform, whereas boys tend to lie below it. This observation indicates that girls tend to experience pubertal growth at earlier ages, when compared to boys. Beyond variation in timing, a visual examination of the aligned profiles in (b), allows for a clearer distinctions of sex-related amplitude variation patterns. In particular, pubertal spurts for girls tend to be attenuated when compared to pubertal spurts in boys.

These observations motivate a natural extension of curve registration models as formal tools to relate individual covariate information to simple and interpretable components of variation in functional data. Specifically, following Brumback and Lindstrom (2004) and Telesca et al. (2012) we develop a class of models aimed at explaining amplitude and phase variability in a sample of curves using individual-level predictors.

4.1 Generalized curve registration models

The hierarchical model in (2) need not be restricted to assumptions of Gaussian sampling. In fact this assumption may be relaxed to accommodate a wider range of sampling scenarios, usually encountered in biostatistical application. In particular, we consider Hidden Gaussian Random Fields (HGRF) models, as a suitable family amenable to straightforward adaptations of the formulation introduced in Section 3.

In HGRF observations $y_i(t_{ij})$ are equipped with mirroring latent Gaussian quantities, say $z_i(t_{ij})$. A sampling model for $y_i(t_{ij}) \mid z_i(t_{ij}), \boldsymbol{\psi}_i \sim F(z_i(t_{ij}); \boldsymbol{\psi}_i)$, is fully defined conditionally on $z_i(t_{ij})$ and a possible set of parameters $\boldsymbol{\psi}_i$. Registration is then achieved at the latent Gaussian level. More precisely, by assuming the stochastic dynamic generating $z_i(t_{ij})$ is centered around a compound process, defined as an affine transformation of a population mean trajectory $\mu(t)$, evaluated over subject-specific random time schedules $u_i(t)$, with random scales c_i and amplitudes a_i, $(i = 1, 2, \ldots, n)$ as

in (2).

For example, longitudinal counts arise naturally in many applications, like immunology, bioinformatics and behavioral studies. In this case one may follow the approach outlined in Telesca et al. (2012) and model

$$y_i(t_{ij}) \mid z_i(t_{ij}) \sim Poisson[\exp\{z_i(t_{ij})\}],$$

with $z_i(t_{ij}) \sim \mathcal{N}(g_i(t_{ij}),\, \sigma^2_\epsilon)$ and $g_i(t_{ij}) = c_i + a_i\, \mu(t_{ij}) \circ u_i(t_{ij})$.

Similarly Erosheva et al. (2014) apply this representation to model censored Gaussian observations. In particular, given a left censoring point η_0 and right censoring η_1, common for all subjects, $z_i(t_{ij})$ is defined as an uncensored latent variable. The outcome at time t_{ij}, for individual i is then modeled as

$$y_i(t_{ij}) = \min\Big[\max\{\eta_0,\, z_i(t_{ij})\},\, \eta_1\Big],$$

with $z_i(t_j) \sim \mathcal{N}(g_i(t_{ij}),\, \sigma^2_\epsilon)$. Other common applications of the latent Gaussian Field framework include models for binary and ordinal data (Albert and Chib 1993).

4.2 Amplitude and phase regression

Let X_i be a $p-$dimensional vector of subject specific covariate information. The assessment of how amplitude and phase variability are explained by predictors is naturally achieved at the second stage of the hierarchical model, through covariate-dependent priors for amplitude parameters a_i and time transformation coefficients $\boldsymbol{\phi}_i$.

Amplitude regression. Let $\mathbf{b}_a$ be a $p-$dimensional vector of amplitude regression coefficients, we explain amplitude variability by defining the hidden linear model:

$$a_i \sim N(1 + X_i'\mathbf{b}_a,\, \sigma^2_a) I(a_i > 0). \tag{11}$$

In the foregoing formula, regression coefficients are offset by a factor of 1, to define coefficients with respect to a reference amplitude. The coefficients $\mathbf{b}_a$ are interpreted as in common linear regression models. Additionally, it is customary to assume $\sum_i a_i = n$, for amplitude identifiability.

Phase regression. Let $\mathbf{b}_\phi$ be a $p-$dimensional vector of phase regression coefficients, we explain phase variability by defining the hidden autoregressive

linear model:

$$\gamma_{iq} = \Upsilon_q + X_i' \mathbf{b}_\phi, \tag{12}$$

$$\phi_{iq} - \gamma_{iq} = \phi_{i(q-1)} - \gamma_{i(q-1)} + \eta_{iq}; \tag{13}$$

with $\eta_{iq} \sim \mathcal{N}(0, \sigma_\phi^2) I(\mathcal{M})$ and $\mathcal{M} = \{\phi_{iq} : \ \phi_{i(q+1)} > \phi_{iq}, \ \phi_{i1} \geq (t_1 - \delta), \ \phi_{iQ} \leq (t_m + \delta), \ q = 1, 2, \ldots, Q\}$. As for the case of amplitude, regression coefficients are offset by the identity transform coefficients Υ, in order to fix a reference time scale. Regression coefficients $\mathbf{b}_\phi$ are then interpreted as changes in the average time scale associated with changes in predictor values.

Random scales c_i are most commonly treated as nuisance parameters and simply modeled as $c_i \sim \mathcal{N}(0, \sigma_c^2)$, with $\sum_i c_i = 0$ for scale identifiability.

In the setting of HGRF models, prior distributions for regression coefficients may still exploit conditional conjugacy. For example, if we denote the covariates matrix with $\mathbf{X} : n \times p$, standard Zellner priors may be considered for amplitude and phase regression as follows:

$$\mathbf{b}_a \mid \sigma_a^2 \sim \mathcal{N}\left(0, \, n\sigma_a^2 (X'X)^{-1}\right), \tag{14}$$

$$\mathbf{b}_\phi \mid \sigma_\phi^2 \sim \mathcal{N}\left(0, \, n\sigma_\phi^2 (X'X)^{-1}\right). \tag{15}$$

Variance components σ_a^2 and σ_ϕ^2 are commonly assigned conditionally conjugate Inverse Gamma priors.

4.3 Growth velocities and drug concentrations revisited

We apply the model introduced in Section 4.2 two our two case study datasets. Regression results are reported in Table 1.

For the Berekely growth study, we consider sex as a predictor of amplitude and phase variation in growth velocity. Our intuition being the plot in Fig. 3 is confirmed, in that girls tend to experience both attenuated amplitude (-0.031) and accelerated timing (-0.39), when compared to boys. Our formal analysis, however, highlights that there is too much uncertainty around amplitude and phase variation in growth, therefore no significant group differences are detected.

A similar question may be asked of the PK dynamics of Remifentanil, that is, are drug concentration amplitude and phase variability explained

Table 1: **Amplitude and phase regression.**

Berkeley Growth Study

	Amplitude		Phase (years)	
Predictor	$E(b_a \mid \mathbf{y})$	95% CI	$E(b_\phi \mid \mathbf{y})$	95% CI
Baseline				
Male	0.019	[-0.027, 0.064]	0.44	[-0.65, 1.49]
Main Effects				
Female	-0.031	[-0.093, 0.031]	-0.39	[-1.76, 1.06]

Pharmacokinetics of Remifentanil

	Amplitude		Phase (years)	
Predictor	$E(b_a \mid \mathbf{y})$	95% CI	$E(b_\phi \mid \mathbf{y})$	95% CI
Baseline				
Female	-0.03	[-0.19, 0.13]	-0.67	[-13.23, 12.27]
Main Effects				
Male	0.05	[-0.19, 0.282]	-0.77	[19.22, -18.22]
Age	0.01	[-0.01, 0.022]	0.02	[-0.33, 0.36]
Weight	-0.01	[-0.02,-0.001]*	-0.03	[-0.63, 0.59]

by sex, when adjusting for potential confouding factors? In this case we perform a regression analysis involving patients sex, body weight and age. This analysis reveals that weight plays a possible role in the concentration dynamic of Remifentanil, with heavier patients experiencing lower (-0.01) concentration amplitude per Kg.

For these analyses to produce conclusive evidence, we often require large amounts of data (large n), as high variability often characterizes estimates of individual level amplitude and phase. At the same time, these results warn against the meaningfulness of step-wise regression approaches, where asymptotic validity may not result in acceptable finite-sample conclusions.

5 Joint functional regression and registration

From a statistical perspective our goal is to develop models that: (1) deal with registration by aligning response trajectories, so that they are defined over a standardized time scale and (2) allow for the estimation of covariate effects on a functional response, that are representative of typical response patterns. Some progress can be made using the approach outlined in Section 4. However, regression coefficients obtained using this technique, average across the entire function evaluation domain and are likely to miss more nuanced, time-dependent effects. As an example, consider the growth data in Fig. 3. Timing differences between girls and boys are most evident between the ages of 10 and 15. However, the regression analysis reported in Table 1 averages across periods of homogenous timing, revealing no conclusive difference between sexes.

A more nuanced approach to joint regression and registration finds motivation in the original generative model in (1). In particular, one often assumes that individual trajectories are centered around common structural mean function $\mu(t)$ evaluated over unit-specific time scales $u_i(t)$. When predictors X_i are available in the form of a p-dimensional vector of subject-level covariates, the assumption of a common mean, may be relaxed and a new generative model for a sample of random trajectories is more realistically represented as:

$$y_i(t) = \mu(t, X_i) \circ u_i(t), \tag{16}$$

where the form of $\mu(t, X_i)$ is made dependent on the vector or predictors X_i.

In the following we review common approaches to the estimation of functional regression coefficients (Hastie and Tibshirani 1993; Guo 2002; Morris

and Carroll 2006). Finally we discuss a natural extension of Bayesian hierarchical curve registraton (Telesca and Inoue 2008) to a unified framework for functional mixed effects modeling and curve registration. We name this class of models Functional Mixed Registration Models (FMRM).

5.1 Functional regression and mixed models

Approaches extending linear models to the functional context often build on the idea of varying-coefficients (Hastie and Tibshirani 1993). Varying-coefficient models are linear in the regressors, but their coefficient are allowed to vary smoothly with the value of other variables, known as effect–modifiers. Given a set of p predictors $x_1, \ldots, x_p$ and p effect modifiers $r_1, \ldots, r_p$, varying coefficient models consider a general link function as $\eta = s_0 + \sum_{j=1}^{p} x_j \, S_j(r_j)$. Hastie and Tibshirani (1993) showed that additive and generalized additive models represent a special case of varying–coefficient models. Several authors have extended these modeling approaches to incorporate intra-curve dependence (Hart and Wehrly 1986 Wypij et al. 1993; Zeger and Diggle 1994; Wang and Gasser 1999; Verbyla et al. 1999; Silverman 1995).

From a mixed effects perspective, Functional Mixed Models (FMM), as proposed by Shi et al. (1996), extend the work of Laird and Ware (1982) to functional data by leaving the forms of the fixed and random effect functions unspecified. These models inherit the flexibility of mixed effects models in handling complex designs and correlation structures. We review a general and flexible view of the problem as discussed by Guo (2002) who used smoothing splines to model both fixed and random effects.

Let $y_i(t)$ denote the value of curve i ($i = 1, ..., n$), at time $t \in \mathcal{T}$ compact. Following Guo (2002) we define:

$$y_i(t) = \mathbf{X}_i' \mathbf{B}(t) + \mathbf{Z}_i' \mathbf{U}_i(t) + \epsilon_i(t) \,, \tag{17}$$

where, for any time $t \in \mathcal{T}$, $\mathbf{B}(t) = (\beta_1(t), \ldots, \beta_p(t))'$ is a p–dimensional vector of fixed effect functions with corresponding design vector $\mathbf{X}_i = (X_{i1}, ..., X_{ip})'$ and $\mathbf{U}_i(t) = (U_{i1}(t), \ldots, U_{im}(t))'$ is an m–dimensional vector of random effect functions with corresponding design vector $\mathbf{Z}_i = (Z_{i1}, ..., Z_{im})'$. Finally, $\epsilon_i(t)$ represents the residual error process for curve i at time t.

Guo (2002) discussed FMM in the context of smoothing splines proposing two estimation approaches based on restricted maximum likelihood and Kalman filters. Morris and Carroll (2006) discussed wavelet-based FMM in

a Bayesian framework proposing inferences based on posterior samples of the functions of interest. FMM include several other models like linear mixed effect models, functional regression and functional ANOVA, as special cases.

While the FMM approach is flexible enough to account for curve–specific variability, it fails to discriminate between the different sources of variation in functional data, namely, amplitude and phase variability. If phase variability is ignored, FMM tend to provide an estimate of the covariate effect which oversmoothes with respect to curve-specific functional features occurring on a stochastic time scale.

5.2 Functional mixed registration

We introduce our Functional Mixed Registration Models (FMRM) as a natural extension of the Bayesian hierarchical curve registration framework (Telesca and Inoue 2008).

Following the notation introduced in Section 5.1, we model a sample of curves $y_i(t)$ $(i = 1, \ldots, n,\ t \in \mathcal{T})$ as:

$$y_i(t) \;=\; \{\mathbf{X}_i'\mathbf{B}(t) + \mathbf{Z}_i'\mathbf{U}_i(t)\,\} \circ u_i(t) + \epsilon_i(t)\,, \tag{18}$$

so that,

$$y_i(t) \;=\; \mathbf{X}_i'\mathbf{B}\{u_i(t)\} + \mathbf{Z}_i'\mathbf{U}_i\{u_i(t)\} + \epsilon_i(t)\,, \tag{19}$$

where $u_i(t)$ is a smooth monotone time transformation function as defined in Section 3.

The FMRM framework includes naturally several existing modeling strategies. In fact, given specific configurations of the time transformation functions or the covariate set, we may obtain the following models as special cases:

a. *Functional Mixed Effect Models.* By setting the time transformation functions μ_i to the identity transformation so that $\mu_i(t) = t$, for any $t \in \mathcal{T}$, our FMRM reduces to the Functional Mixed Model.

b. *Hierarchical Curve Registration Models.* By setting the random effect functions $\mathbf{U}_i(t) = c_i + a_i\,\mathbf{B}(t)$, with $\mathbf{X}_i = 1$ and $\mathbf{Z}_i = 1$ $(i = 1, ..., N)$, $t \in \mathcal{T}$, our FMRM reduces to:

$$\begin{aligned} y_i(t) \;&=\; \{\,\mathbf{B}(t) + (c_i + a_i\,\mathbf{B}(t)\,\} \circ u_i(t) + \epsilon_i(t), && (20)\\ &=\; (c_i + (1 + a_i)\mathbf{B}(t)) \circ u_i(t) + \epsilon_i(t), && (21) \end{aligned}$$

which is equivalent to (2).

The FMRM in equation (19) is, however, not identifiable. Given any fixed effect function $\mathbf{B}(t)$, a number of combinations of random effects $\mathbf{U}_i(t)$ and time transformation functions $u_i(t)$ may, in fact, lead to the same likelihood or posterior density. The identifiability issue is mainly due to the arbitrary flexibility with the random effects functions. Choosing a reference curve or considering constrained formulations may help with the identification problem. However, this is not usually a straightforward task. Here we choose to focus on random effects which are assumed to have a strictly parametric form. In particular, we will only allow for individual random scale or amplitude transformations, so that model (19) can be rewritten as:

$$y_i(t) = c_i + a_i \, \mathbf{X}_i' \mathbf{B}(t) \circ u_i(t), \quad i = 1, ..., N; \tag{22}$$

where c_i is a curve–specific scale parameter and a_i is a curve–specific amplitude parameter.

The finite dimensional representation of functional quantities in (22), may follow the penalized B-spline formulation introduced in Section 3.2. More precisely, given a K–dimensional set of kernels $\boldsymbol{\mathcal{S}}_\beta(t)$, evaluated at time $t \in \mathcal{T}$, and a $p \times K$ matrix of regression coefficients $\boldsymbol{\beta}$, one may represent the fixed effect functions $\mathbf{B}(t)$ as $\mathbf{B}(t) = \boldsymbol{\beta} \, \boldsymbol{\mathcal{S}}_\beta(t)$.

More generally, of course, specific choices for the kernels $\boldsymbol{\mathcal{S}}_\beta(t)$ may depend on the study and should reflect reasonable assumptions about the functional form of $\mathbf{B}(t)$. For example, in the analysis of the Berkeley growth study, it is reasonable to consider functions that are smooth and continuous. Thus, one may choose $\boldsymbol{\mathcal{S}}_\beta(t)$ to belong to the spline family. On the other hand, if the outcome consists of a set of long time series, characterized by highly localized features, such as in mass spectrometry data, then $\boldsymbol{\mathcal{S}}_\beta(t)$ could be represented by wavelet basis functions (Morris and Carroll 2006).

Given the representation in (22), prior settings and MCMC simulation strategies may follow the same approach outlined in Section 3.

5.3 Functional mixed registration of drug concentrations

We apply the FMRM approach to the pharmacokinetics of Remifentanil. Our analysis replicates the regression exercise attempted in Section 4.3. The goal

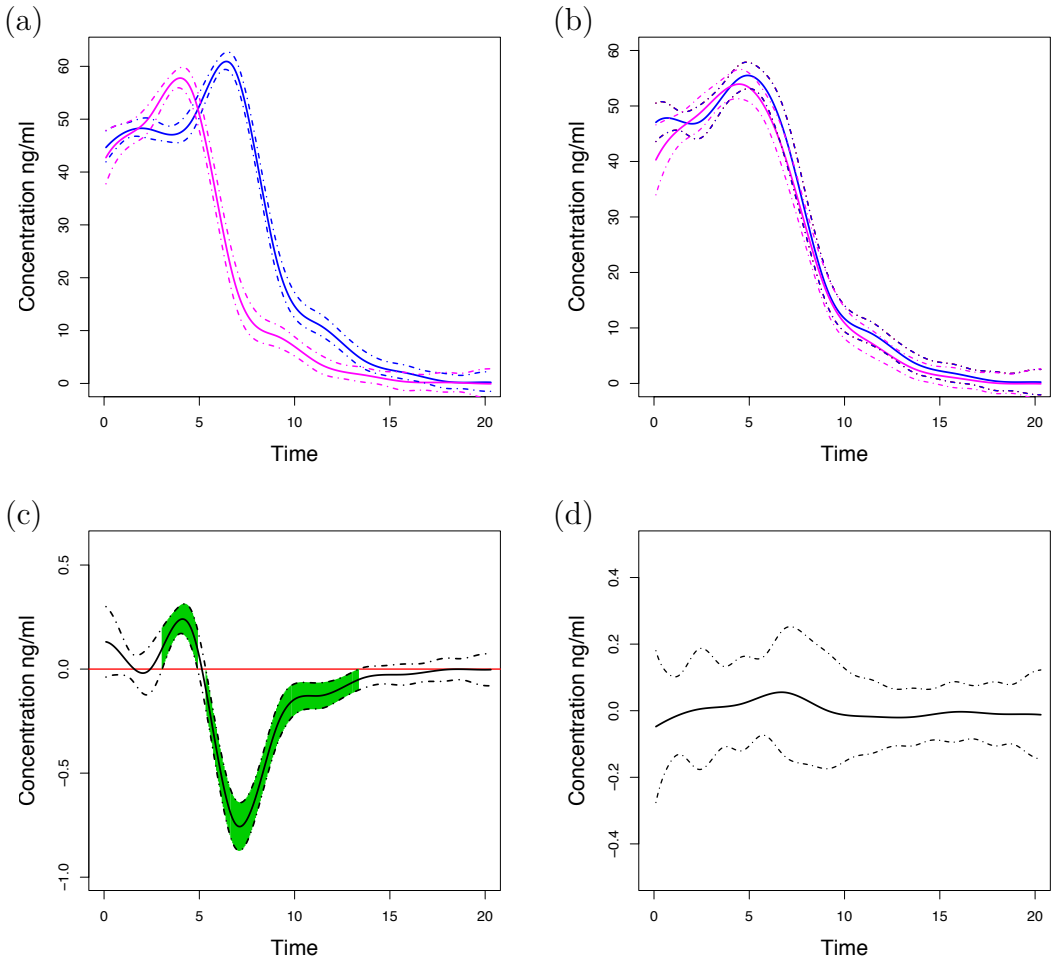

Figure 5: **FMRM analysis of drug concentration**. (a) Unadjusted mean posterior drug concentrations for males (blue) and females (magenta). (b) Posterior mean drug concentration trajectories for males (blue) and females (magenta), adjusted for age and body weight. (c) Time varying effect of body weight. (d) Time varying effect of age. In all panels we report simultaneous 95% credible bands.

of our analysis is to assess differences in the pharmacokinetics of Remifentanil between males and females adjusting for age and body weight.

Figure 5, panels (a) through (d), shows the results from our analysis. All figures are plotted over a transformed time scale $(\log\ t)^2$ in order to better display differences between curves. Panel (a) shows the unadjusted posterior common scaled shape functions for the blood concentration trajectories of male (blue) and female (magenta) patients. Without adjusting for other predictors it appears that females have a faster metabolism and excretion of the drug. Panel (b) shows the mean posterior pharmacokinetic profile for males (blue) and (females) adjusted by age (years) and body weight (Kg). Panel (c) shows the time varying effect of body weight and panel (d) shows the time varying effect of age. We highlight in green areas where the effect of the predictors are significantly different from 0. We note that the differences in the metabolism of Remifentanil between males and females are now fully accounted for by differences in body weight. As one may reasonably expect, we no longer see significant sex-related effects on the pharmacokinetics of the drug.

6 Differential expression and gene profile similarities

In this Section we discuss the application of registration models in Bioinformatics. Specifically, time course genomics data, consist of measurements from a common set of genes collected at different time points and provide new opportunities into the understanding of gene regulation.

In particular, clues about the temporal structure of expression may be informative about co-regulation and gene-gene relationships (Qian et al. 2001; Leng and Müller 2006). In this section we discuss the approach of Telesca et al. (2009), who introduced a model-based selection of differentially expressed genes, and a probabilistic framework for the investigation of regulatory relationships between genes.

6.1 A functional mixture model for differential expression

Following the formulation of Section 3, we let $y_i(t)$ denote the observed expression level of gene i at time t where $i = 1, 2, \ldots, n$ and $t \in \mathcal{T}$. We assume that gene-specific expression profiles arise following the same stochastic generative mechanism in (2). Given a common shape function $\mu(t)$, individual curves may exhibit different levels and amplitudes of response and different timing schedules associated with time-dependent expression features.

In this setting, the parameters a_i describes the amplitude of the mRNA signal for gene i. A formalization of our statistical definition of differentially expressed genes may be achieved via a mixture approach. This idea follows naturally from similar formalizations introduced by Parmigiani et al. (2002) and extended by Telesca et al. (2012).

For each gene ($i = 1, 2, \ldots, n$), we specify the following prior for the amplitude of the expression signal,

$$a_i = \pi^- N(a_0^-, \sigma^2_{a^-}) I(a_i < 0) + \pi^+ N(a_0^+, \sigma^2_{a^+}) I(a_i > 0) + \pi^0 N(0, \sigma^2_{a^0}); \quad (23)$$

with $(\pi^- + \pi^0 + \pi^+) = 1$. Here π^0 identifies the overall proportion of genes in their normal range of variation, while $(\pi^- + \pi^+)$ identifies the proportion of overly active genes. The mixture characterization with two truncated normals (that is, $N^-(\cdot, \cdot) I(a_i < 0)$ and $N^+(\cdot, \cdot) I(a_i > 0)$) allows us to account for genes with a synchronous expression signal of opposite sign (negative dependence).

From an inferential perspective, a decision to flag specific genes as being differentially expressed corresponds to testing the following set of hypotheses for all $i = 1, 2, \ldots, n$:

$$\begin{aligned} H_{0i} &: a_i \sim N(0, \sigma^2_{a^0}) \\ H_{1i} &: a_i \sim N(a_0^+, \sigma^2_{a^+}) \text{ or } a_i \sim N(a_0^-, \sigma^2_{a^-}). \end{aligned} \quad (24)$$

Give posterior samples from $a_i \mid \mathbf{y}$, decision rules controlling for pre-defined error rates, like the False Discovery Rate (FDR) of Benjamini and Hochberg 1995, are easily derived, for example, following the approach described by Müller et al. (2006).

6.2 Posterior measures of profile similarities

The underlying idea for the investigation of gene networks using time course microarray data is that genes that share similar expression profiles may share similar biological functions and thus, could be related. Posterior inference about gene-specific time transformation functions may be used to derive measures of gene-gene relationships which are based on functional similarities.

In the context of the model described in Section 6.1, for differentially expressed profiles, local measures of profile similarity may be derived as follows:

Local warping distance. Let $\tau \subset T$, we define a *local distance* $d_{ik}(\tau)$ between genes i and k $(i \neq k)$ as

$$d_{ik}(\tau) = \int_{\tau} \mid u_i(t) - u_k(t) \mid dt, \tag{25}$$

that is, as the absolute distance between the time transformation functions of genes i and k along time points $t \in \tau$. This measure may be interpreted as the average difference in the timing of expression features between the expression of two genes over a period of time τ. From a global perspective one may of course consider a warping distance integrating over the entire sampling window T.

Relevant summaries from the marginal posterior distribution of time transformation functions may be extracted to draw inference about gene-gene relationships. In particular, one may formalize inference about profile similarities as the series of hypotheses:

$$H_{0ik} : d_{ik}(\tau) \geq \gamma, \text{ vs. } H_{1ik} : d_{ik}(\tau) < \gamma; \text{ for all } i \neq k.$$

For this series of decisions, optimal rules controlling for error rates are derived as discussed in Section 6.1.

While recognizing the importance of the timing characteristics of gene expression, the selection of an appropriate timing envelope γ must, however, be aided by biological knowledge about the timing of gene–gene regulation in the specific process under investigation. For example, in cell cycle experiments, regulatory envelopes of interest may span only a few minutes, while in the study of androgen refractory tumors the timing of interest is of the order of days (Pound et al. 1999).

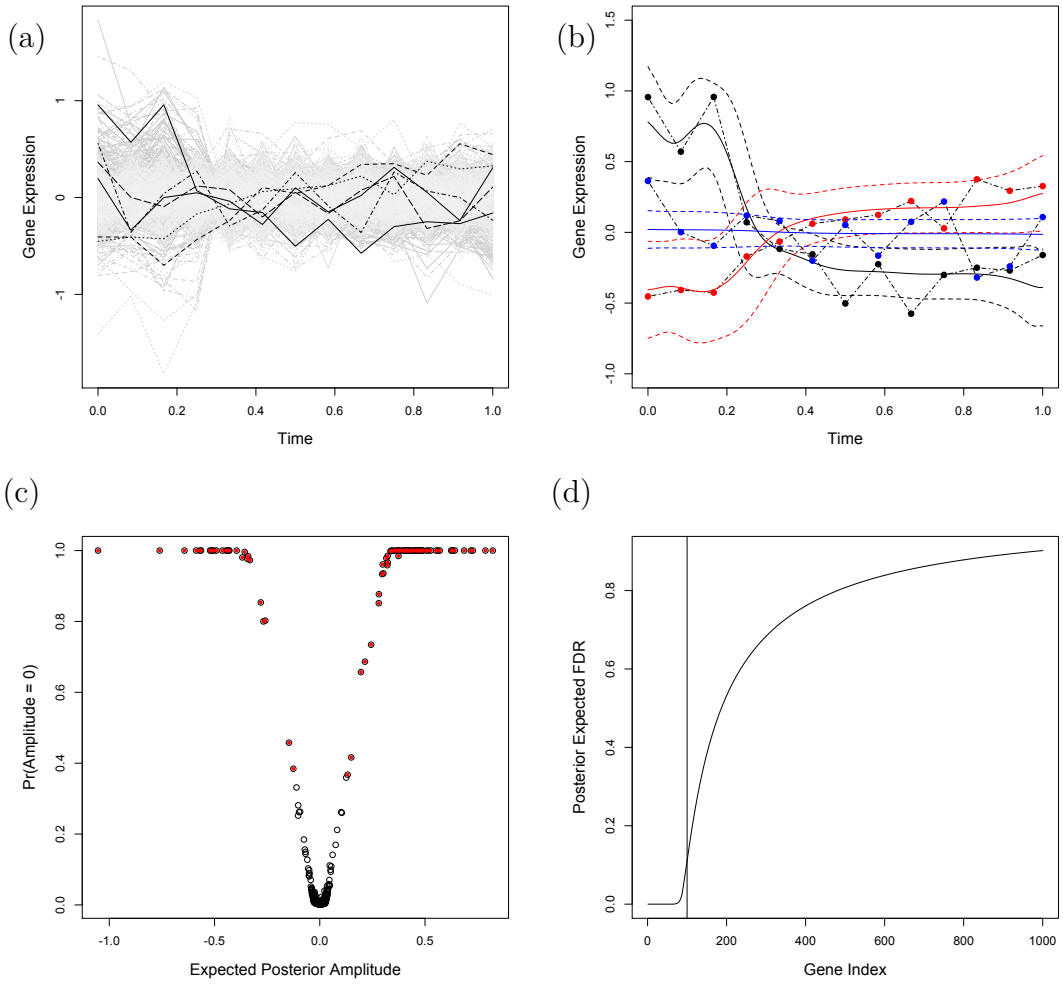

Figure 6: **Time-course gene expression.** (a) Time course gene-expression profiles. (b) Individual model fit for a representative set of genes. (c) Volcano plot of posterior expected amplitude vs. posterior probability of no time-dependent expression. (d) Posterior expected FDR vs. gene index.

6.3 A case study of time-course gene expression analysis

Here we illustrate the application of registration models with mixture priors, to the analysis of time-course expression data. In particular, we consider data on 100 gene reporters of 13 time-points mouse Affimetrix microarray gene expression from a study on primary mouse keratinocytes, with induced activation of the TRP63 transcription factor (Della Gatta et al. 2008).

The data has been processed using `rma` (affy) and the profiles are centred (zero-mean) across the time points. As the original data is composed of direct targets of TRP63, we expect all genes to be differentially expressed over time. For illustrative purposes, in order to test the performance of registration models for differential expression analysis, we augment the original data-set with 900 pseudo-genes of constant average expression.

Results are summarized in figure 6. In particular, panel (a) highlights a random sample of profiles, and panel (b) shows model fits and associated posterior predictive bands for a representative set of profiles. Even though a registration model of time-dependent expression makes what seem like restrictive assumptions bout possible gene-specific time-dependent dynamics, this figure illustrates the actual flexibility of the model in its ability to recover heterogenous time-course profiles. Our analysis of differential expression is reported in panels (c) and (d), where we show the model selection, aimed at controlling the posterior expected FDR at 10%. The model selects 96 genes as differentially expressed, all of which are in the original TRP63 target set.

A full analysis of profile similarities is beyond the scope of this manuscript. For more examples we refer the reader to Telesca et al. (2009).

7 Concluding Remarks

We have reviewed the application of curve registration techniques to the analysis of functional data arising in Biostatistics and Bioinformatics. Our review is by no means exhaustive and is clearly biased towards the author's expertise.

Modeling frameworks using the idea of stochastic time scales have a strong tradition in several fields and are indeed the subject of active research efforts (Zhang and Telesca 2014; Cheng et al. 2013)

Our discussion is focused on Bayesian inference with smoothing priors. In

the setting of kernel-based regression the selection of diffuse priors remains controversial and default choices are based on purely intuitive arguments. Attempts at formalization do exist, for example, Wakefield (2012), Chapter 11, discusses approaches based on effective degrees of freedom.

Finally, wile inference based on posterior simulation is straightforward, the application of standard MCMC techniques may be unrealistic for cases where the analysis involves a large number of subjects. Similarly, computational feasibility is often in question for studies where technology allows for highly intensive sampling of individual profiles. In these situation, one must consider careful computational nuances and the potential development of efficient approximation techniques.

References

Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*, 669–679.

Baladandayuthapani, V., B. K. Mallick, and R. J. Carroll (2005). Spatially adaptive Bayesian penalized regression splines (P-splines). *Journal of Computational and Graphical Statistics 14*(2), 378–394.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B 57*, 289–300.

Brumback, L. C. and M. J. Lindstrom (2004). Self modeling with flexible, random time transformations. *Biometrics 60*(2), 461–470.

Cheng, W., I. Dryden, and X. Huang (2013). Bayesian registrations of functions and curves. *eprint arXiv:1311.2105*.

De Boor, C. (1978). *A Practical Guide to Splines*. Berlin: Springer-Verlag.

Della Gatta, G., M. Bansal, A. Ambesi-Impiombato, D. Antonini, C. Missero, and D. di Bernardo (2008). Direct targets of the trp63 transcription factor revealed by a combination of gene expression profiling and reverse engineering. *Genome Research 18*(6), 939–948.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science 11*, 89–102.

Erosheva, E. A., R. L. Matsueda, and D. Telesca (2014). Breaking bad: Reviewing two decades of life course data analysis in criminology and beyond. *Annual Reviews of Statistics and Its Applications 1*, 301–332.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2013). *Bayesian Data Analysis* (3rd ed.). Chapman & Hall / CRC.

Gervini, D. and T. Gasser (2004). Self-modelling warping functions. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 66*(4), 959–971.

Guo, W. (2002). Functional mixed effects models. *Biometrics 58*(1), 121–128.

Hart, J. D. and T. E. Wehrly (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association 81*, 1080–1088.

Hastie, T. and R. Tibshirani (1993). Varying–coefficient models. *Journal of the Royal Statistical Society 55*, 757–796.

Hastie, T., R. Tibshirani, and J. H. Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer–Verlag Inc.

Kneip, A. and T. Gasser (1988). Convergence and consistency results for self-modeling nonlinear regression. *The Annals of Statistics 16*, 82–112.

Kneip, A. and T. Gasser (1992). Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics 20*, 1266–1305.

Kneip, A., X. Li, K. B. MacGibbon, and J. O. Ramsay (2000). Curve registration by local regression. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 28*(1), 19–29.

Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics 38*, 963–974.

Lang, S. and A. Brezger (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics 13*(1), 183–212.

Leng, X. and H. Müller (2006). Time ordering of gene co-expression. *Biostatistics 7*.

Liu, X. and H. Müller (2004). Functional averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association 99*, 687–699.

Morris, J. S. and R. J. Carroll (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 68*(2), 179–199.

Müller, P., G. Parmigiani, and K. Rice (2006). Fdr and Bayesian multiple comparisons rules. *Proceedings of the Valencia/ISBA 8th World Meeting on Bayesian Statistics (Oxford University Press)*..

Parmigiani, G., S. E. Garrett, R. Anbashgahn, and E. Gabrielson (2002). A statistical framework for expression-based molecular classification in cancer. *Journal of The Royal Statistical Society, Series B 64*, 717–736.

Pinheiro, J. C. and D. M. Bates (2000). *Mixed–Effects Models in S and S–Plus*. Springer–Verlag: New York.

Pound, C. R., A. W. Partin, M. A. Eisenberger, D. W. Chan, J. D. Pearson, and P. C. Walsh (1999). Natural history of progression after psa elevation following radical prostatectomy. *Journal of the American Medical Association 281*, 1591–1597.

Qian, J., M. Dolled-Filhart, J. Lin, H. Yu, and M. Gerstein (2001). Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. *Journal of Molecular Biology 314*, 1053–1066.

Ramsay, J. O. and X. Li (1998). Curve registration. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 60*, 351–363.

Rice, J. A. and B. W. Silverman (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B: Methodological 53*, 233–243.

Ruppert, D., W. M. and R. J. Carroll (2003). *Semiparametric Regression*. Cambridge University Press.

Sakoe, H. and S. Chiba (1978). Dynamic programming optimization for spoken word recognition. *IEEE Transactions of Acoustic, Speech and Signal Processing* (1), 43–49.

Shi, M., R. E. Weiss, and J. M. G. Taylor (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics 45*, 151–163.

Silverman, B. W. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society, Series B: Methodological 57*, 673–689.

Telesca, D., E. A. Erosheva, D. A. Kreager, and R. L. Matsueda (2012). Modeling criminal careers as departures from a unimodal popolation age-crime curve: The case of marijuana use. *JASA 107*(500), 1427–1440.

Telesca, D. and L. Y. T. Inoue (2008). Bayesian hierarchical curve registration. *Journal of the American Statistical Association 103 (481)*, 328–339.

Telesca, D., L. Y. T. Inoue, M. Neira, R. Etzioni, M. Gleave, and C. Nelson (2009). Differential expression and network inferences through functional data modeling. *Biometrics 65*, 793–804.

Telesca, D. adn Muller, P., G. Parmigiani, and F. RS (2012). Modeling dependent gene expression. *Annals of Applied Statistics 6*(2), 542–560.

Tuddenham, R. D. and M. M. Snyder (1954). Physical growth of califorfornia boys and girls from birth to eighteen years. *University of California Publications in Child Development 1*, 183–364.

Verbyla, A. P., P. Arunas, B. R. Cullis, M. G. Kenward, and S. J. Welham (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society, Series C: Applied Statistics 48*, 269–300.

Wakefield, J. (2012). *Bayesian and Frequentist regression methods*. Springer.

Wang, K. and T. Gasser (1997). Alignment of curves by dynamic time warping. *The Annals of Statistics 25*(3), 1251–1276.

Wang, K. and T. Gasser (1999). Synchronizing sample curves nonparametrically. *The Annals of Statistics 27*(2), 439–460.

Wypij, D., M. Pugh, and J. H. Ware (1993). Modeling pulmonary function growth with regression splines. *Statistica Sinica 3*, 329–350.

Yao, F., H. J. Müller, and J. L. Wang (2005). Functional data analysis of sparse longitudinal data. *JASA 100*(470), 577–590.

Zeger, S. L. and P. J. Diggle (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics 50*, 689–699.

Zhang, Y. and D. Telesca (2014). Joint clustering and registration of functional data. Technical report, UCLA.