Bayesian Hierarchical Curve Registration

Donatello Telesca & Lurdes Y. T Inoue

Bayesian Hierarchical Curve Registration

Donatello TELESCA and Lurdes Y. T. INOUE

Functional data often exhibit a common shape, but with variations in amplitude and phase across curves. The analysis often proceeds by synchronization of the data through curve registration. In this article we propose a Bayesian hierarchical model for curve registration. Our hierarchical model provides a formal account of amplitude and phase variability while borrowing strength from the data across curves in the estimation of the model parameters. We discuss extensions of the model by using penalized B-splines in the representation of the shape and time-transformation functions, and by allowing temporal misalignment of the curves. We discuss applications of our model to simulated data, as well as to two data sets. In particular, we use our model in a nonstandard analysis aimed at investigating regulatory network in time course microarray data.

KEY WORDS: Functional data; Markov chain Monte Carlo; Shape function; Temporal misalignment; Time transformation function.

1. INTRODUCTION

Functional data analysis is concerned with the analysis of random curves, known as functional data, which often exhibit a common pattern but with variations in amplitude and phase across curves. The monographs of Ramsay and Silverman (1997, 2002) provide a broad range of examples of functional data, as well as methods for analysis.

In this article we analyze two sets of functional data, shown in Figure 1. Panel (a) shows a sample of height velocity curves for 39 boys estimated from individual growth curves. We observe an overall deceleration trend in growth from infancy to adulthood, with some few acceleration–deceleration pulses in velocity. In particular, the most prominent velocity pulse corresponds to the pubertal spurt. The goal of the analysis lies in modeling both the amplitude and timing of features in the individual curves to estimate the velocity curve. Panel (b) shows a sample of time-course mRNA gene expression measurements in the yeast cell cycle experiment described by Spellman et al. (1998). Time-course gene expression microarray experiments are thought to provide insights into the functional state of an organism by mapping the activity of the entire genome as a biological process that unfolds over time. One of the goals of analysis is to learn about network relationships between genes. In this sense, we are specifically interested in modeling the timing of individual expression profile features (e.g., maxima, minima, plateaus, inflections), which may be suggestive of gene regulatory relationships of activation, inhibition, and coexpression.

An important aspect of functional data is that time variation explains a large portion of the variability in the data. For example, 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. Such time variation is accounted for with curve registration methods for functional data. Specifically, curve registration, also known as curve alignment in biology and time warping in the engineering literature, refers to a class of methods that involve finding, for each observed curve, a warping function that synchronizes all curves. Often, after aligning curves, an average curve is estimated, a process called structural averaging by Kneip and Gasser (1988, 1992).

Several time warping methods have been devised to date. In the engineering literature, Sakoe and Chiba (1978) developed a registration technique called dynamic time warping to align two curves with different dynamics and applied it to speech analysis and recognition. Under this method, dynamic programming is used to optimize the alignment between two curves through minimization of a cost function. Wang and Gasser (1997, 1999) introduced this technique to the statistical literature and provided large-sample properties of the time-transformation estimators. In particular, they proposed an alternative cost function that used amplitude and derivatives of the curves and extended dynamic time warping to align more than two curves. Central to their development, however, was the assumption that the curves were smooth.

Gasser and Kneip (1995) proposed the landmark registration method, which involves identifying the timing of certain features (landmarks) in the curves which are then aligned so that they occur at the same transformed times. Alternatively, Silverman (1995) developed a method for curve registration that did not require landmarks. Optimizing a fitting criterion, Silverman estimated time-shift transformations. The method was later extended by Ramsay and Li (1998) to a continuous monotone transformation family and by Kneip, Li, MacGibbon, and Ramsay (2000), who proposed curve registration with locally estimated monotone time transformations.

More recently, some authors have explored curve registration using self-modeling regression (SEMOR) methods. In their classical version, the SEMOR methods of Lawton, Sylvestre, and Maggio (1972) and Kneip and Gasser (1988) are semiparametric models in which the subject-specific regression function is a parametric transformation of a common smooth regression function. Specifically, let $y_i(t)$ denote the observed curve for subject i at time t. The general SEMOR model can be expressed as $y_i(t) = m_i(t) + e_i$, with $m_i(t) = g(m, \theta_i)(t)$, where m is the common shape function, g is the transformation that generates the individual regression functions, and e_i is the random error. The name "self-modeling" comes from the fact that in SEMOR, both the shape function g and model parameter θ_i are estimated from the data. A common class of SEMOR models with transformations of both the x and y axes with a shape-invariant shift has $m_i(t) = \alpha_i + \beta_i m(\gamma_i t + \delta_i)$, referred to as

Donatello Telesca is Post-Doctoral Fellow, Department of Biostatistics, University of Texas, M.D. Anderson Cancer Center, Houston, TX 77030 and Research Associate, Fred Hutchinson Cancer Research Center, Seattle, WA 98109 (E-mail: *donatello.telesca@gmail.com*). Lurdes Y. T. Inoue is Assistant Professor, Department of Biostatistics, University of Washington, Seattle, WA 98195 and Affiliate Investigator, Fred Hutchinson Cancer Research Center, Seattle, WA 98109. Partial support was provided by the National Cancer Institute (grant P50 CA 97186) and the Career Development Fund, Department of Biostatistics, University of Washington. The authors also acknowledge insightful suggestions by the associate editor and reviewers. Web Appendixes and Supplementary Material may be found at *http://www.amstat.org/publications/jasa/supplemental_materials*.
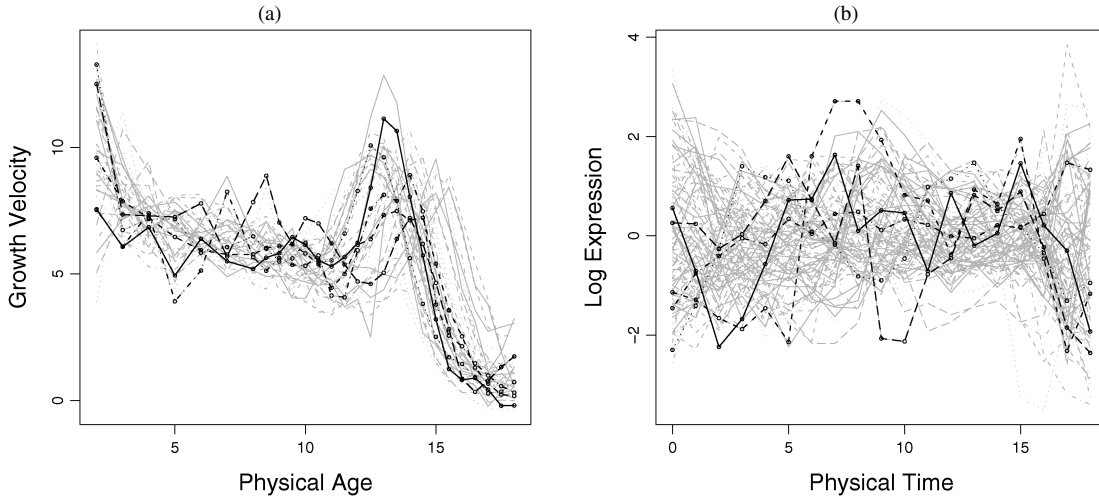
Figure 1. Examples of functional data. (a) Individual growth velocity curves for 39 boys. (b) Yeast cell cycle mRNA expression measurements (Spellman et al. 1998) for a sample of genes, where each time unit corresponds to 7 minutes.

the shape-invariant model (SIM) (Lawton et al. 1972; Kneip and Gasser 1988).

Ronn (2001), considering a version of the foregoing SIM model in which $m_i(t) = m(t + \delta_i)$, where the shifts δ_i are random effects, proposed nonparametric maximum likelihood registration of random time-shifted curves. Brumback and Lindstrom (2004) built on the SIM models, replacing the linear time transformation function to allow for more flexible, monotone random time transformations. Specifically, in their formulation, curve-specific amplitude parameters as well as curve-specific time transformations are assumed random. Moreover, the time transformation and the common shape function are modeled with B-splines.

With a similar idea toward developing flexible warping functions, Gervini and Gasser (2004) proposed a self-modeling warping function with the components of the warping function estimated from the data using B-spline basis functions. But in their formulation, the time transformations are not random. Liu and Müller (2004) proposed time synchronization of random processes where the observed random curves are assumed to be generated by a bivariate stochastic process consisting of a stochastic monotone time transformation function and an unrestricted random amplitude function.

In this article we investigate curve registration from a Bayesian perspective and propose a Bayesian hierarchical curve registration (BHCR) model. Although our model builds on earlier work (Ronn 2001; Brumback and Lindstrom 2004; Gervini and Gasser 2004), to the best of our knowledge, it is the first to address curve registration from a Bayesian standpoint, with the hierarchical model providing a formal account of amplitude and phase variability while borrowing strength from the data across curves in the estimation of the model parameters. Novel aspects of the model include (a) the extended family of time-transformation functions allowing for temporal misalignment, (b) the use of penalized B-splines in the representation of the shape and time-transformation functions, and (c) exact inferences using Markov chain Monte Carlo (MCMC) methods in estimating the posterior distribution of the model parameters.

Our approach provides extended flexibility in modeling both the shape and time-transformation functions. The borrowing of information across curves allows us to fit our models to sparsely observed functional data. Moreover, our models introduce a "temporal misalignment window" that naturally allows for the registration of curves not observed at the same sampling times. Our penalized representation of the B-splines of the shape and time transformation functions replaces the selection of the smoothing parameter with an "automatic" model-based smoothing. This procedure allows for inferences that are more robust than those with fixed B-splines. Finally, our model allows us to derive exact inferences about a richer set of quantities of interest, such as the posterior distribution of the pubertal and mid growth spurts, that cannot be easily addressed using existing methods.

This article is organized as follows. In Section 2 we present our model. Specifically, in Section 2.1 we present our hierarchical model. We discuss flexible representations of the shape and time-transformation functions using B-splines in Section 2.2 and using penalized B-splines in Section 2.3. In Section 2.4 we discuss an extension of the model that allows temporal misalignment in the time transformation function. In Section 3 we illustrate our methods through two simulation studies and two case studies. We assess model fit using simulated data in Section 3.1. In Section 3.2 we compare BHCR with landmark registration (LR) and self-modeling registration (SMR) methods. We analyze the height velocity data in Section 3.3, and apply our model to time course microarray data in Section 3.4. In particular, in the latter example we introduce network inference from BHCR models. Finally, we provide a discussion in Section 4.

2. MODEL FORMULATION

2.1 Hierarchical Model

Let $y_i(t)$ denote the observed level of the ith curve at time t, with $i = 1, 2, \ldots, N$ and $t \in T = [t_1, t_n]$. We introduce a three-stage hierarchical model.

Stage One. Write

$$m_i(t) = m(t; \boldsymbol{\theta}_i) = c_i + a_i m(t; \boldsymbol{\beta}) = c_i + a_i \mathcal{B}'_m(t)\boldsymbol{\beta}, \quad (1)$$

$$\mu_i(t) = \mu(t; \boldsymbol{\phi}_i) = \mathcal{B}'_{\boldsymbol{\mu}}(t)\boldsymbol{\phi}_i, \quad (2)$$

and the composite function

$$\tilde{m}_i(t) = m_i(t) \circ \mu_i(t) = m(\mu_i(t); \boldsymbol{\theta}_i). \tag{3}$$

The observed value of each curve i at time t is modeled as

$$y_i(t) = \tilde{m}_i(t) + \epsilon_i, \qquad i = 1, \ldots, N, \tag{4}$$

where $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0; \sigma_\epsilon^2)$; that is, we assume that the error terms are independent and normally distributed with null mean and common variance σ_ϵ^2.

In the foregoing, $m(t; \boldsymbol{\beta})$ denotes a common shape function generating the individual curves and $\mu_i(t)$ denotes a curve-specific time-transformation function. Both functions are modeled as linear combinations of an appropriate set of basis functions, $\mathcal{B}_m(t)$ and $\mathcal{B}_\mu(t)$, and a set of basis coefficients, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}_i$. Additional discussion about the representation of these functions is deferred to Sections 2.2 and 2.3. On registration of the curves, we identify the ith aligned curve at time t as

$$y_i^*(t) = y_i(t) \circ \mu_i^{-1}(t). \tag{5}$$

Stage Two. Given a common shape function $m(t; \boldsymbol{\beta})$, individual curves may exhibit different scales and levels of response. We assume that $c_i \sim \mathcal{N}(c_0; \sigma_c^2)$ and $a_i \sim \mathcal{N}(a_0; \sigma_a^2) \times I\{a_i > 0\}$, thus defining curve-specific random linear transformations.

We note that the normality assumptions enable conjugacy with the likelihood. Moreover, the foregoing assumption of strictly positive amplitude can be relaxed. For example, in our application to time course expression data, the amplitude parameters are allowed to vary over the entire real line, because expression levels for genes with opposing regulation could have antisymmetric profiles.

Individual curves may exhibit features occurring at different times. Thus we use curve-specific random time-transformation functions $\mu_i(t)$, parameterized with curve-specific parameters $\boldsymbol{\phi}_i$, to characterize the timing features of each curve. We start by considering this function as a smooth map defined over the design interval T, subject to the monotonicity and image constraints

$$t_1 < \cdots < \mu_i(t_j) < \mu_i(t_k) < \cdots < t_n,$$

$$\forall i \in \{1, \ldots, N\}, \; t_j < t_k, \tag{6}$$

and

$$\mu_i(t_1) = t_1, \qquad \mu_i(t_n) = t_n, \qquad \forall i \in \{1, \ldots, N\}, \tag{7}$$

where (6) restricts the time-transformation functions to maintain the original ordering of the sampling times (no time reversion) and allows for a one-to-one correspondence between physical times and transformed times, whereas constraint (7) corresponds to the assumption that the stochastic time transformation occurs between fixed starting and termination time points, coinciding with the boundaries of the sampling interval T. The latter assumption may not be adequate in some cases—for example, when some trajectories exhibit a simple linear shift. We discuss relaxing assumption (7) in Section 2.4.

We assume that the time-transformation function coefficients have a multivariate normal distribution, $\boldsymbol{\phi}_i \overset{\text{iid}}{\sim} \text{N}(\boldsymbol{\Upsilon}; \boldsymbol{\Sigma}_\phi)$, where $\boldsymbol{\Upsilon}$ is the vector associated with the identity time-transformation function, that is, $\mu(t; \boldsymbol{\Upsilon}) = t$ (see Web Appendix A.1 for more details).

Stage Three. We assume that $a_0 \sim \mathcal{N}(m_{a0}; \sigma_{a0}^2)$ and $c_0 \sim \mathcal{N}(m_{c0}; \sigma_{c0}^2)$. Moreover, for precision parameters,

$$1/\sigma_a^2 \sim Gamma(a_a; b_a),$$

$$1/\sigma_c^2 \sim Gamma(a_c; b_c),$$

and

$$1/\sigma_\epsilon^2 \sim Gamma(a_\epsilon; b_\epsilon).$$

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

In addition, we assume that the shape function coefficient vector $\boldsymbol{\beta}$ has a multivariate normal distribution, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}; \boldsymbol{\Sigma}_\beta)$. The precision matrix for the shape coefficients is specified as $\boldsymbol{\Sigma}_\beta^{-1} = \frac{1}{\lambda}\boldsymbol{\Omega}$, where $1/\lambda \sim Gamma(a_\lambda; b_\lambda)$. Similarly, the precision matrix for the time-transformation parameters is specified as $\boldsymbol{\Sigma}_\phi^{-1} = \mathbf{P}/\sigma_\phi^2$, with $1/\sigma_\phi^2 \sim Gamma(a_\phi; b_\phi)$. We discuss choices of $\boldsymbol{\Omega}$ and $\mathbf{P}$ under penalized B-spline basis functions in Section 2.3.

2.2 Fixed-Knots B–Spline Implementation

In this section we discuss the specification of the shape function $m(t; \boldsymbol{\beta})$ and the time-transformation functions $\mu(t; \boldsymbol{\phi}_i)$ using B-spline basis functions (de Boor 1978).

Because of their flexibility, B-spline basis functions have been previously used to represent the common shape function and/or the individual time-transformation functions (Brumback and Lindstrom 2004; Gervini and Gasser 2004). Specifically, to represent the common shape function, we select a set of knots $(\kappa_1, \kappa_2, \ldots, \kappa_p)$ partitioning the sampling interval T into $p + 1$ subintervals. Using piecewise cubic polynomials and given the set of interior knots denoted by $\{\kappa\}_p$, we define $\mathcal{B}_m(t)$ as a K-dimensional design vector of B-spline basis evaluated at time t, with $K = p + 4$. In this framework, we then define $m(t; \boldsymbol{\beta}) = \mathcal{B}_m'(t)\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the K-dimensional vector of basis coefficients defining the common shape function.

Similarly, given a set of interior knots $\{\omega\}_h$, we may represent the individual time-transformation functions as $\mu(t; \boldsymbol{\phi}_i) = \mathcal{B}_\mu'(t)\boldsymbol{\phi}_i$, where $\mathcal{B}_\mu(t)$ is a Q-dimensional design vector of B-spline basis evaluated at time t, where $Q = h + 4$, and $\boldsymbol{\phi}_i$ is the Q-dimensional vector of basis coefficients. Because the time-transformation function needs to respect monotonicity and boundary conditions (6) and (7), we impose the following constraints on the time-transformation coefficients $\boldsymbol{\phi}_i$:

$$t_1 < \phi_{i2} < \phi_{i3} < \cdots < t_n, \qquad \forall i \in \{1, \ldots, N\}$$

$$\text{(monotonicity),} \tag{8}$$

and

$$\phi_{i1} = t_1, \qquad \phi_{iQ} = t_n, \qquad \forall i \in \{1, \ldots, N\} \quad \text{(image).} \tag{9}$$

Equation (8) provides a sufficient condition for monotonicity, derived from basic properties of B-spline derivatives (de Boor 1978).

The foregoing implementation requires specification of the number of interior knots as well as the location of the knots for both the common shape function $m(t; \boldsymbol{\beta})$ and the individual time-transformation functions $\mu(t; \boldsymbol{\phi}_i)$. This model selection problem is often addressed with the minimization of a measure of prediction error (see, e.g., Hastie, Tibshirani, and Friedman 2001) and cross-validation procedures (see, e.g., Gervini and Gasser 2004).

2.3 Penalized Regression Splines Implementation

As pointed out earlier, B-splines offer a flexible modeling tool. An issue for its implementation arises with the choice of the positions and number of interior knots. We propose an alternative that relies on penalized regression splines (Eilers and Marx 1996; Ruppert, Wand, and Carrol 2003). Specifically, under this formulation, a relatively large number of equidistant knots is selected and a penalty, dependent on a smoothing parameter λ, is 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. We take the Bayesian approach and explicitly include the smoothing parameter in the model.

Consider the problem of estimating the common shape function $m(t, \boldsymbol{\beta})$. Given a relatively large number of equidistant interior knots $\{\kappa\}_p$, we represent the common shape function as a linear combination of B-spline basis, $m(t; \boldsymbol{\beta}) = \mathcal{B}'_m(t)\boldsymbol{\beta}$. Following Lang and Brezger (2004), we place a first-order random-walk shrinkage prior on the shape coefficients $\boldsymbol{\beta}$, so that

$$\beta_k = \beta_{k-1} + e_k, \qquad e_k \sim \mathcal{N}(0; \lambda). \quad (10)$$

We assume that $\beta_0 = 0$. It can be shown that, conditional on λ, $\boldsymbol{\beta}$ has a multivariate normal distribution with null mean vector and precision matrix $\boldsymbol{\Omega}/\lambda$. Under the foregoing first-order random walk, $\boldsymbol{\Omega}$ is a banded precision penalization matrix,

$$\boldsymbol{\Omega} = \begin{pmatrix} 2 & -1 & 0 & & & & 0 \\ -1 & 2 & -1 & \ddots & & & \\ 0 & -1 & \ddots & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & -1 & 0 \\ & & & \ddots & -1 & 2 & -1 \\ 0 & & & & 0 & -1 & 1 \end{pmatrix}. \quad (11)$$

Note that the random-walk variance λ can be interpreted as the smoothing parameter. We place a relatively diffuse conjugate inverse gamma hyperprior on the variance, that is, $\lambda \sim IG(a_\lambda; b_\lambda)$. A similar approach may be adopted to model the time-transformation functions $\mu(t; \boldsymbol{\phi}_i)$ so that $\forall i = 1, \ldots, N$,

$$(\phi_{ik} - \Upsilon_k) = (\phi_{i(k-1)} - \Upsilon_{k-1}) + \eta_k, \qquad \eta_k \sim \mathcal{N}(0; \sigma_\phi^2). \quad (12)$$

Assuming that $(\phi_{i0} - \Upsilon_0) = 0$, it can be shown that $\boldsymbol{\phi}_i \sim N(\boldsymbol{\Upsilon}; \sigma_\phi^2 \mathbf{P})$, where $\mathbf{P}$ is a precision penalization matrix as in (11) and σ_ϕ^2 is the smoothing parameter associated with the transformation functions $\mu(t; \boldsymbol{\phi}_i)$.

2.4 Stochastic Time Model

As we discussed in Section 2.1, the assumptions imposed by (7) imply that the stochastic time transformation occurs between fixed starting and termination time points, coinciding with the boundaries of the sampling interval T. But in the case of time course microarray data, for example, such a restriction may not be adequate. For instance, consider two expression profile signals such that $E(Y_1(t)) = m_1(t)$ and $E(Y_2(t)) = m_1(t + \delta_2)$, differentiated by a simple linear offset. The time-transformation function cannot be analytically represented by a time transformation with fixed image T.

To account for such cases, we model the time-transformation functions as monotone maps defined over the sampling interval T with values in a random, curve-specific interval, $\mathcal{T}_i = [t_1 + \delta_{i1}, t_n + \delta_{in}]$. Because we are modeling the common shape function $m(t; \boldsymbol{\beta})$ in a semiparametric fashion, we assume that

$$(\delta_{i1}, \delta_{in})' \sim \mathcal{N}(\mathbf{m}_\delta; \boldsymbol{\Lambda}) I\{(\delta_{i1} \geq t_1 - \Delta; \delta_{in} \leq t_n + \Delta)\}, \quad (13)$$

where $\mathbf{m}_\delta = (t_1, t_n)'$, $\boldsymbol{\Lambda}$ is the covariance matrix implied by the penalized formulation in (12), and $\Delta < \infty$ is the window of temporal misalignment.

The shape function $m(t; \boldsymbol{\beta})$ and the individual time-transformation functions $\mu_i(t)$ can be modeled again as linear combinations of (penalized) B-spline basis functions. Note that the set of basis defining $m(t, \boldsymbol{\beta})$ now need to span a functional space over the expanded domain $\tilde{T} = [t_1 - \Delta, t_n + \Delta]$. For the individual time-transformation functions to take values over random intervals of the type $\mathcal{T}_i$, the two terminal time-transformation coefficients are allowed to vary within some intervals. Operationally, we set $(\delta_{i1} = \phi_{i1})$ and $(\delta_{in} = \phi_{iQ})$ and proceed as in Section 2.2, modifying the set of constraints (8) and (9) so that, $\forall i \in \{1, \ldots, N\}$,

$$(t_1 - \Delta) \leq \phi_{i1} < \cdots < \phi_{iq} < \phi_{i(q+1)}$$
$$< \cdots < \phi_{iQ} \leq (t_n + \Delta). \quad (14)$$

The ordering in (14) guarantees monotonicity of the time-transformation functions $\mu(t, \boldsymbol{\phi}_i)$ and ensures that the set of random intervals $\{\mathcal{T}_i\}_{(i=1,\ldots,N)}$ has compact support $\tilde{T}$. We note that this model introduces a flexible family of smooth monotone time-transformation functions, establishing a general framework in which linear time shifting, time warping, and their combinations are special cases obtained under additional constraints on (14). For example, we can represent time shifting imposing $Q = 2$ so that $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2})'$ with $(t_1 - \Delta) < \phi_{i1} < (t_1 + \Delta)$ and $\phi_{i2} = \phi_{i1} + t_n$. Warping requires only $t_1 = \phi_{i1} < \cdots < \phi_{iQ} = t_n$, whereas linear shift and warping require $(t_1 - \Delta) < \phi_{i1} < \min(\phi_{12}, (t_1 + \Delta))$ and $\phi_{iQ} = \phi_{i1} + t_n$.

2.5 Posterior Inference Through Markov Chain Monte Carlo

2.5.1 Model Implementation. Under our model formulation, the full parameter vector is

$$\boldsymbol{\theta} = (\mathbf{c}', \mathbf{a}', \boldsymbol{\beta}', \boldsymbol{\phi}', c_0, a_0, \sigma_\epsilon^2, \sigma_c^2, \sigma_a^2, \lambda, \sigma_\phi^2)',$$

where $\mathbf{c} = (c_1, \ldots, c_N)'$, $\mathbf{a} = (a_1, \ldots, a_N)'$, and $\boldsymbol{\phi} = (\boldsymbol{\phi}_1', \ldots, \boldsymbol{\phi}_N')'$ is an NQ vector of individual time-transformation parameters. We fully specify the Bayesian model with priors on the parameter vector $\boldsymbol{\theta}$ as specified in Section 2.1. The observation vector is $\mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_N')'$, that is, a vector where each component $\mathbf{Y}_i$ is itself a vector containing the observed levels of curve i, $i = 1, \ldots, N$.

Thus, given $\mathbf{Y}$, the posterior distribution then follows from

$$f(\boldsymbol{\theta}|\mathbf{Y}) = f(\mathbf{c}; \mathbf{a}; \boldsymbol{\beta}; \boldsymbol{\phi}; c_0; a_0; \sigma_\epsilon^2; \sigma_c^2; \sigma_a^2; \lambda; \sigma_\phi^2 | \mathbf{Y})$$
$$\propto f(\mathbf{Y}|\mathbf{c}; \mathbf{a}; \boldsymbol{\beta}; \boldsymbol{\phi}; \sigma_\epsilon^2) f(\mathbf{c}; \mathbf{a}|c_0; a_0; \sigma_c^2; \sigma_a^2)$$
$$\times f(\boldsymbol{\beta}|\lambda) f(\boldsymbol{\phi}|\sigma_\phi^2) f(c_0|\sigma_{c_0}^2) f(a_0|\sigma_{a_0}^2)$$
$$\times f(\sigma_\epsilon^2|a_\epsilon; b_\epsilon) f(\sigma_c^2|a_c; b_c) f(\sigma_a^2|a_a; b_a)$$
$$\times f(\lambda|a_\lambda; b_\lambda) f(\sigma_\phi^2|a_\phi; b_\phi). \quad (15)$$

The joint posterior density is analytically intractable. Therefore, our posterior inferences are based on MCMC simulation (Gamerman 1997).

Let θ_j denote a generic jth element of θ and let θ_{-j} denote the complement vector, that is, the vector containing all components of θ except θ_j. Using the foregoing notation, $[\theta_j | \theta_{-j}, \mathbf{Y}]$ is the full conditional distribution of θ_j. We can show that most components of our parameter vector θ allow for closed-form full conditional distributions. The exception is for parameter ϕ. The full conditional distributions for our model parameters (for components of θ that admit closed-form distributions) are shown in Web Appendix A.2.

We thus implement our MCMC using a Gibbs-within-Metropolis algorithm. Specifically, we use a componentwise updating in which we directly sample components with closed-form full conditionals using Gibbs steps; otherwise, we use Metropolis–Hastings steps. The latter case is applied to sampling the time-transformation coefficients ϕ_i, $i = 1, \ldots, N$. In detail, each component of ϕ_i is updated individually using uniform proposals on their constrained support with the range of the proposal calibrated at the burn-in of the MCMC runs to ensure an acceptance rate between 35% and 75%.

In most of our applications, our posterior inferences are based on 15,000 posterior samples obtained after discarding the initial 50,000 MCMC iterations for burn-in. Convergence diagnostics are performed with the R package BOA (Smith 2005).

2.5.2 Posterior Inference. Given M draws from the posterior distribution of the model parameters θ, we can derive the posterior distribution of any parameter/function of interest, as we describe. To register the observed curves, we use (5), replacing $\mu_i(t)$ with a posterior mean estimate; that is, given posterior samples $\phi_i^{(j)}$, $j = 1, \ldots, M$, we calculate

$$\mu_i^{(j)}(t) = \mu\big(t; \phi_i^{(j)}\big) = \mathcal{B}'_\mu(t)\phi_i^{(j)}.$$

The foregoing procedure gives posterior samples for the time-transformation function at time t. Next, we obtain the posterior samples of the warping function at time t [i.e., the inverse $\mu_i^{-1(j)}(t)$] and estimate the posterior mean as $\hat{\mu}_i^{-1}(t) = \frac{1}{M}\sum_{j=1}^M \mu_i^{-1(j)}(t)$. Finally, we plug this posterior estimate into (5) to register curve i.

Besides registering the observed curves, we may be interested in the scaled common shape function, which we define as

$$m_s(t; a_0; \beta; c_0) = c_0 + a_0 \mathcal{B}'_m(t)\beta, \qquad (16)$$

that is, the common shape function rescaled by population level parameters a_0 and c_0. Let $a_0^{(j)}$, $\beta^{(j)}$, and $c_0^{(j)}$, $j = 1, \ldots, M$, denote M draws from the marginal posterior distributions of a_0, β, and c_0. Draws from the marginal posterior distribution of the scaled shape function, at time t, are given by

$$m_s^{(j)}(t; a_0; \beta; c_0) = c_0^{(j)} + a_0^{(j)} \mathcal{B}'_m(t)\beta^{(j)}, \qquad j = 1, \ldots, M. \tag{17}$$

In Section 3.3, one of the goals of analysis lies in estimating the age distribution for the mid and pubertal growth spurts. These parameters correspond to the local maxima of the common shape function. To make inferences about these quantities,

we can use our draws from the posterior distribution. Specifically, using (17), for each $\beta^{(j)}$, $j = 1, \ldots, M$, we evaluate the common shape function on a fine grid of values of $t \in T$. For each posterior curve, we record the two local maxima, denoted by $l_1^{(j)}$ and $l_2^{(j)}$, which can be found using, for example, numerical procedures. This provides a posterior sample of the age of the mid and pubertal growth spurts.

Using the posterior draws, we also can obtain $100(1 - \alpha)\%$ credible intervals for any quantity of interest. Let θ denote a generic scalar parameter (among those in our model) and assume that $\theta^1, \ldots, \theta^M$ are M draws from the marginal posterior distribution of θ. A $100(1 - \alpha)\%$ posterior credible interval for θ is given by $CI(\theta; \alpha) = [\tilde{\theta}_{\alpha/2}, \tilde{\theta}_{1-\alpha/2}]$, where $\tilde{\theta}_q$ denotes the qth posterior sample quantile.

The foregoing idea also can be extended to functions using pointwise credible intervals. Let $f(\cdot)$ denote a function (e.g., the shape function or time-transformation function). Then for each time t, we can examine the posterior distribution of the function at time t and use the posterior sample quantiles (as described earlier) to determine the credible intervals at each time point.

Alternatively, we can obtain the simultaneous credible band for $f(\cdot)$. This idea was considered by Baladandayuthapani, Mallick, and Carroll (2005). Using a fine grid of time points $t_1 = \tau_1 < \tau_2 < \cdots < \tau_L = t_n$, let M_α denote the $100(1 - \alpha)$ sample quantile of

$$\max_{1 \le i \le L} \big|\big[f(\tau_i) - E\{f(\tau_i)|\mathbf{Y}\}\big]/SD\{f(\tau_i)|\mathbf{Y}\}\big|.$$

A simultaneous $100(1 - \alpha)$ credible band for $f(\tau)$ is

$$CI(f(\tau); \alpha) = E\{f(\tau)|\mathbf{Y}\} \pm M_\alpha SD\{f(\tau)|\mathbf{Y}\}.$$

If the grid is sufficiently fine $(L \gg K)$, given the smoothness properties of the B-spline representation, then the derived credible bands are likely to represent a good approximation to the theoretical $100(1 - \alpha)\%$ credible set of interest.

The foregoing description illustrates how posterior simulation facilitates inferences about a richer set of quantities of scientific interest. We apply these ideas in the next section.

3. APPLICATIONS

3.1 Simulation Study 1: Evaluating Model Fit

To assess estimation by BHCR, we simulated 30 curves (Fig. 2) with the common shape function $m(t) = \cos(t/4) + \sin(t)/4$ evaluated at 50 equally spaced time points in the design interval $T = [0, 30]$. Note that the common shape function for the simulated curves does not follow our model with B-spline representation. Each simulated curve is of the form $y_{ij} = c_i + a_i m(u_i(t_j; \phi_i)) + \epsilon_{ij}$, where we chose, $c_i \overset{iid}{\sim} N(0; 1)$ and $a_i \overset{iid}{\sim} N(1; 1)I\{a_i > 0\}$, $i = 1, \ldots, 30$. The time-transformation functions $u_i(t; \phi_i)$ were generated from a linear combination of B-spline basis over the sampling interval T. In particular, we defined $u_i(t; \phi_i) = \mathcal{B}'(t)\phi_i$, with $\mathcal{B}(t) = \{g_1, \ldots, g_5\}$, corresponding to a single interior knot placed at $t = 15$, and $\phi_i \sim N_5(\mathbf{\Upsilon}; 2\mathbf{I})$ defined over the constrained space as given by (14). We fitted the BHCR model to a set of noisy curves in which $\epsilon_{ij} \overset{iid}{\sim} N(0; \sigma_\epsilon = .3)$.
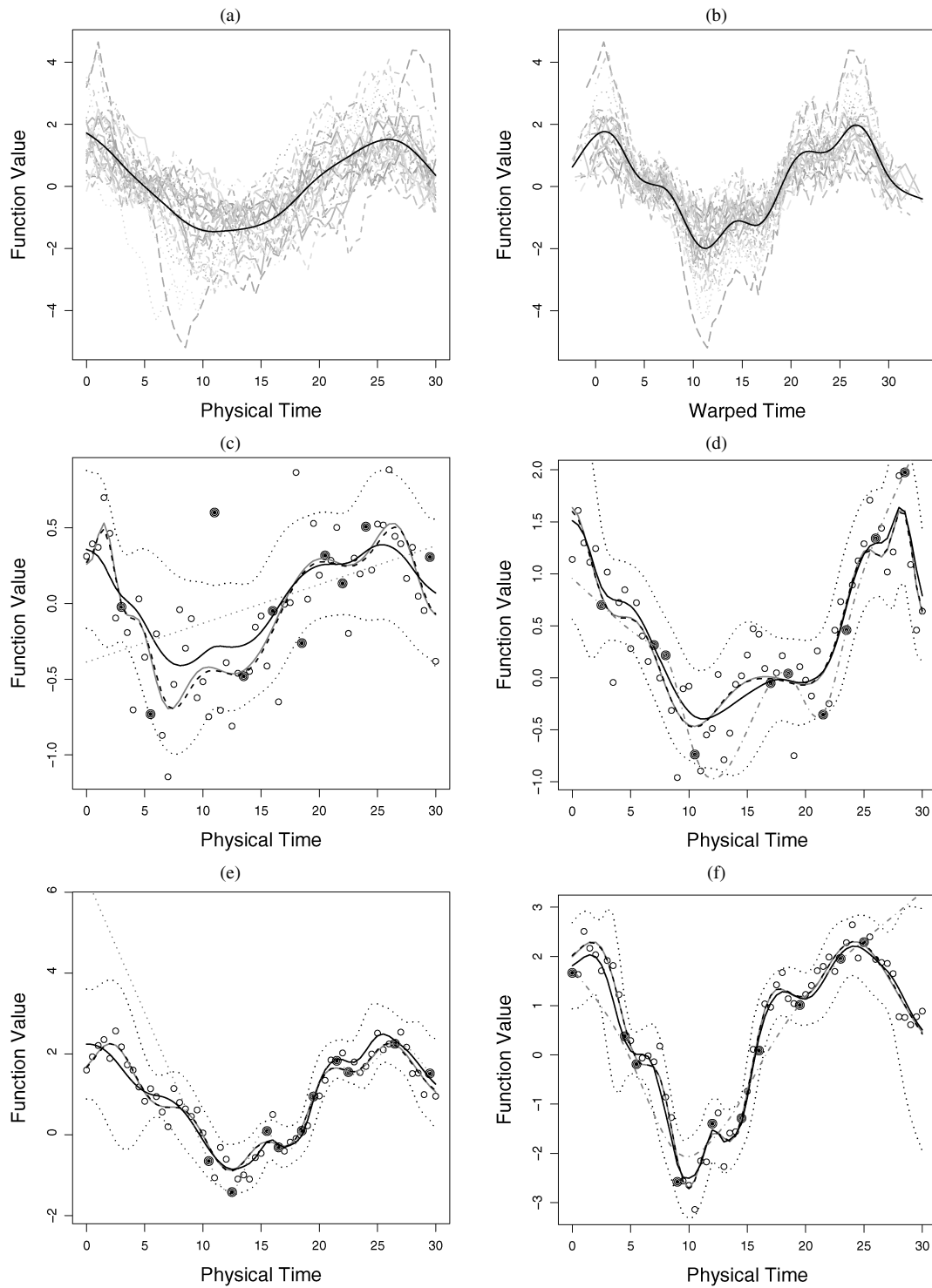
Figure 2. Simulation study 1: Assessing model fit. (a) Thirty simulated curves, superimposed with the cross-sectional mean. (b) Thirty registered sample curves superimposed with the posterior mean of the scaled shape function. (c)–(f) Four randomly selected trajectories with full design (○) and sparse design (●). For each random curve, we plot the true signal (solid gray) and posterior mean trajectory for the full design (dashed black) and the sparse design (solid black, with its respective 95% simultaneous credible bands in dotted lines). Also shown is the smooth spline fit to each curve (dot-dashed gray).

Relatively diffuse *Gamma*(.1; 1) priors were considered for the amplitude, scale, and shape precision parameters. To estimate the common shape function, we placed 41 equally spaced interior knots between -5 and 35 with a temporal misalignment window of $\Delta = 6$. To estimate the individual time-

transformation functions, we partitioned the interval T into four subintervals, placing three interior knots at $\kappa = (7.5, 10, 22.5)$. Because the time-transformation parameters are constrained with maximum support of length $(t_n - t_1 + \Delta) = 30 + 6 = 36$, we consider a *Gamma*(.5; *rate* = 20) prior for the precision of

the time-transformation parameters. Our inferences are based on 15,000 samples from the posterior distribution obtained after discarding the initial 50,000 MCMC iterations for burn-in.

Figure 2 shows some results from our model. Panel (a) shows 30 simulated trajectories superimposed with a cross-sectional mean. The cross-sectional mean curve does not resemble any of the individual curves, because phase variation is not taken into account.

Using draws from the posterior distribution of the model parameters we computed the posterior scaled shape function following the procedure described in Section 2.5. Computation of the registered trajectories follows a similar procedure replacing $\mu_i^{-1}(t)$ from (5) with its posterior mean analog. Panel (b) shows registered trajectories superimposed with the posterior mean scaled shape function along with the true signal. The figure shows that, accounting for phase variability, we are now able to recover subtle features of the originating signal. In fact, both are indistinguishable in the figure.

Panels (c)–(f) indicate the quality of signal recovery when borrowing strength across curves in the fit of four randomly selected trajectories. In particular, we compare the estimated posterior mean curve obtained from a "full" design (50 equally spaced samples per curve) with the posterior mean curve obtained from a sparse design (with curves observed at 10 randomly selected time points). The figure shows that the posterior mean trajectories for the full design (dashed black) are indistinguishable from the true signal (solid gray). When we consider the sparse design, the posterior mean trajectories (solid black) are still able to recover many features of the original true signal, due to the borrowing of information. The benefits of the hierarchical formulation become particularly evident when comparing our posterior mean trajectories with a smoothing spline fit (dot-dashed gray). Here we chose the smoothness degree for the spline fit by cross-validation. Also shown are simultaneous 95% credible bands associated with the sparse design (dotted).

To assess the sensitivity of the results to our prior choices, we reestimated the model considering an approximate twofold increase in the variability of the prior variances. The resulting estimates did not change substantially and thus are omitted. We also assessed estimation with noisier curves by doubling the error variance. We could still successfully estimate the model parameters. For example, we found that the true time-transformation functions were still contained within the 90% posterior credible bands (figure omitted). Finally, our inferences were robust to the number of basis used to represent the common shape (with a minimum of 45 and a maximum of 85 bases) and individual warping functions (with a minimum of 5 and a maximum of 9 bases) (results not given).

3.2 Simulation Study 2: Comparison of Methods

In this section we carry out a simulation study aimed at comparing BHCR with other curve registration techniques, namely landmark registration (Gasser and Kneip 1995) and self-modeling warping functions (Gervini and Gasser 2004).

Following Gervini and Gasser (2004), we define the common shape function as $m(t) = \sin(2\pi t)$, with $t \in [0, 1]$, characterized by two landmarks: a local maximum at time $\ell_{01} = .25$ and a local minimum at time $\ell_{02} = .75$. We simulated N curves over n equally spaced time points in $[0, 1]$ from model (4).

The individual scale c_i was set to 0, and the individual amplitude was chosen so that $a_i \sim \mathcal{N}(1; \sigma_a = .15)I\{a_i > 0\}$. For each curve, we simulated two random landmarks (ℓ_{i1}, ℓ_{i2}) so that $\ell_{ik} \sim \mathcal{N}(\ell_{i0}; \sigma_\ell = .1)I\{\mathcal{M}\}$, where $\mathcal{M}$ ensures that $0 < \ell_{i1} < \ell_{i2} < 1$, for all $i = 1, \ldots, N$. We defined the individual inverse time-transformation functions (also known as warping functions) $\mu_i^{-1}(t)$ as smooth interpolations to the random landmarks so that $\mu_i^{-1}(\ell_{ik}) = \ell_{ik}$, $\mu_i^{-1}(0) = 0$, and $\mu_i^{-1}(1) = 1$.

Considering sample sizes $N = 15, 30$ and $n = 15, 30$ time points, with 200 replications for each combination, and error standard deviation $\sigma_\epsilon = .5$, we compared three estimators of the common shape function $m(t)$:

- LR. Landmarks are estimated as the zero-crossings of the first derivative of the smooth spline-fitted curve. Moreover, we used the true common shape function as the reference curve.
- SM. We use two components and a six B-spline (order = 3) basis (all defined by equally spaced interior knots) for the warping functions, as suggested by Gervini and Gasser (2004). We use their Matlab algorithm (available at *http://www.uwm.edu/%7Egervini/programs.html*) for the registration of the curves.
- BHCR. We use 6 B-spline (order = 3) basis for the time-transformation functions and 19 B-spline (order = 3) basis for the common shape function (all defined by equally spaced interior knots). In the Bayesian framework, we define $\hat{m}(t)$ as the scaled posterior mean shape function $E[m_s(t; a_0, \boldsymbol{\beta}, c_0)|\mathbf{Y}]$. We choose relatively diffuse *Gamma*(.1; 1) priors for the amplitude and scale parameters and *Gamma*(.01; 10) for the shape and time-transformation precision parameters.

Simulation results are displayed in Figure 3. Here we show boxplots of the root mean squared error, $RMSE = [\frac{1}{n} \times \sum_{j=1}^{n}\{\hat{m}(t_j) - m(t_j)\}^2]^{1/2}$, for $\hat{m}(t)$. The figure shows that LR and SM have a comparable performance for larger values of N. This conclusion is in agreement with the findings of Gervini and Gasser (2004). The figure also shows that BHCR has a consistently superior performance across different combinations of the number of curves or observation times.

3.3 Case Study 1: Registering Height Acceleration Functions

The Berkeley Growth Study (Tuddenham and Snyder 1954) recorded the height of 39 boys for 27 time points between age 2 and 18 years. The goal of our analysis is to model both the amplitude and timing of features of the individual curves to estimate, using the notation introduced in Section 2.5.2, the scaled common shape function.

Figure 4(a) shows growth velocity curves superimposed with the cross-sectional mean. As pointed out by Ramsay and Li (1998) and Gervini and Gasser (2004), the cross-sectional mean underestimates the amplitude of the local maxima and overestimates the amplitude of local minima in a systematic fashion, because it does not account for phase variation.

Using our BHCR model, to estimate the common shape function, we placed 27 equally spaced interior knots between -3 and 23 and considered a maximum expansion constraint $\Delta = 7$.
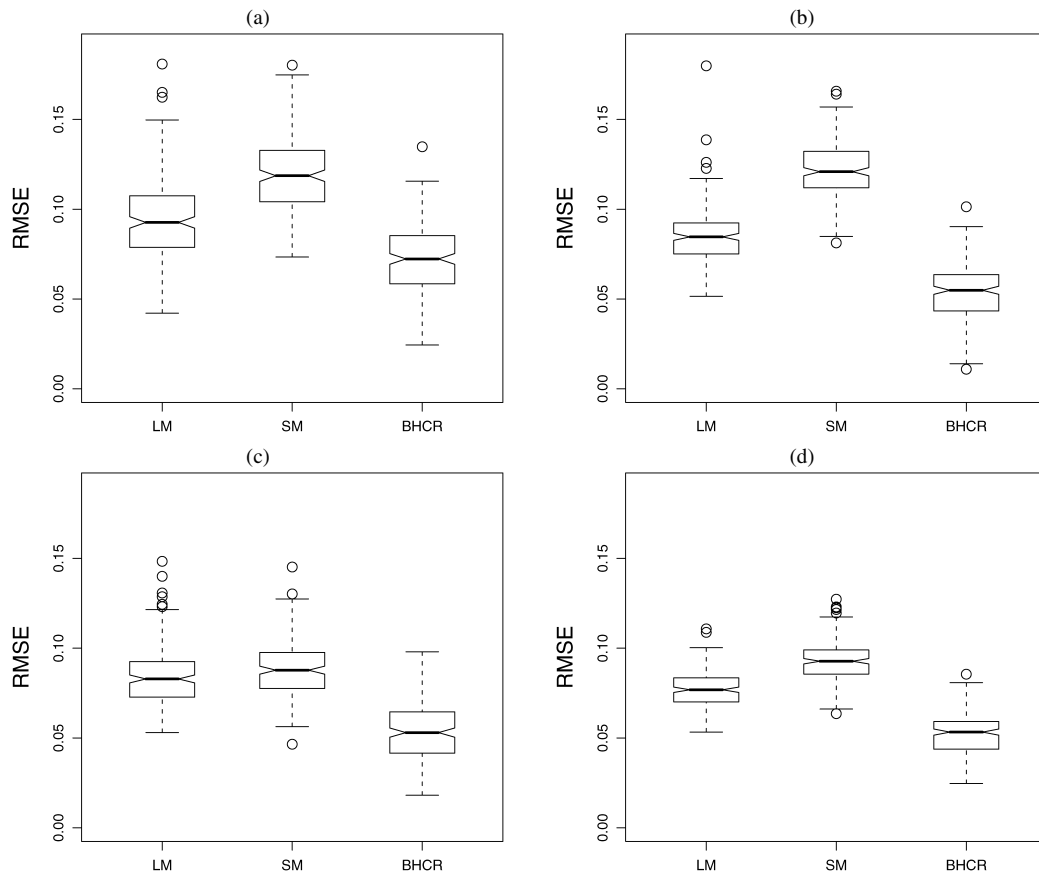
Figure 3. Simulation study 2: Comparison of methods. RMSE of $\hat{m}(t)$ for the LR, SM, and BHCR. (a) $N = 15$, $n = 15$; (b) $N = 15$, $n = 30$; (c) $N = 30$, $n = 15$; (d) $N = 30$, $n = 30$.

We modeled the individual time-transformation functions, partitioning the interval $T = [2, 18]$ into five subintervals and placing four interior knots at $\kappa = (5.2, 8.2, 11.6, 14.8)$. We considered a *Gamma*$(.5; rate = 20)$ prior for the time-transformation precision and placed *Gamma*$(.1; 1)$ priors on the amplitude, scale, and shape precision parameters. Our inferences are based on 15,000 samples from the posterior distribution obtained after discarding the initial 50,000 MCMC iterations for burn-in.

Figures 4(b)–4(d) shows the results from our analysis. Panel (b) shows the curve-specific posterior mean time-transformation functions, whereas panel (c) shows the registered curves superimposed with the scaled posterior mean shape function. The time-transformation functions reveal that curve registrations occur over monotone transformations combined with temporal misalignment. Panel (d) shows the velocity curves aligned by SM (Gervini and Gasser 2004), superimposed on the posterior mean common shape function. Comparing the estimates obtained by BHCR and SM [panel (e)] shows that BHCR allows for clearer identification of the mid growth spurt. Finally, panel (e) shows the scaled posterior mean common shape function with associated simultaneous 95% credible bands and posterior densities for the mid growth (mean = 6.90, SD = .84) and pubertal (mean = 13.30, SD = .56) spurts.

To assess the sensitivity of the results to our prior choices, we also considered a twofold increase of the prior variances. The resulting estimates did not differ substantially from those previously reported in Figure 4 and thus are omitted here.

It is interesting to note that whereas both BHCR and SM identify the pubertal growth spurt, only BHCR clearly identifies the mid growth spurt. Other studies support the existence of the mid growth spurt. Ramsay, Bock, and Gasser (1995), using (presmoothed) data arising from Fels, Zurich, and Berkeley growth studies, derived growth acceleration curves (i.e., the derivative of the velocity curve). They justified the procedure by observing that "other phenomena, and specially the mid spurt, are transient changes in growth most easily seen in the acceleration of height" (Ramsay et al. 1995, p. 416). Using landmark registration on acceleration curves, they found that in boys, the mid growth spurt is on average at age 7.34 (SD = 1.03) and the pubertal spurt is on average at age 13.59 (SD = 1.07).

For a similar data set in the Zurich Growth Study, the SM mean registered trajectory for the leg velocity growth reported by Gervini and Gasser (2004) also showed two spurts. However, in their analysis, the original measurements were presmoothed and the sampling frequency was inflated, by interpolation, from the original 30 to 100 measurements. The authors did not report estimates of mid and pubertal growth spurts; however, examination of their figure 1(c) shows that our point estimates seem to be similar to theirs.

In conclusion, our estimates are similar to those reported in the literature, but our method does not rely on presmoothing of the data. Moreover, we are able to identify the mid spurt without additional data transformation.
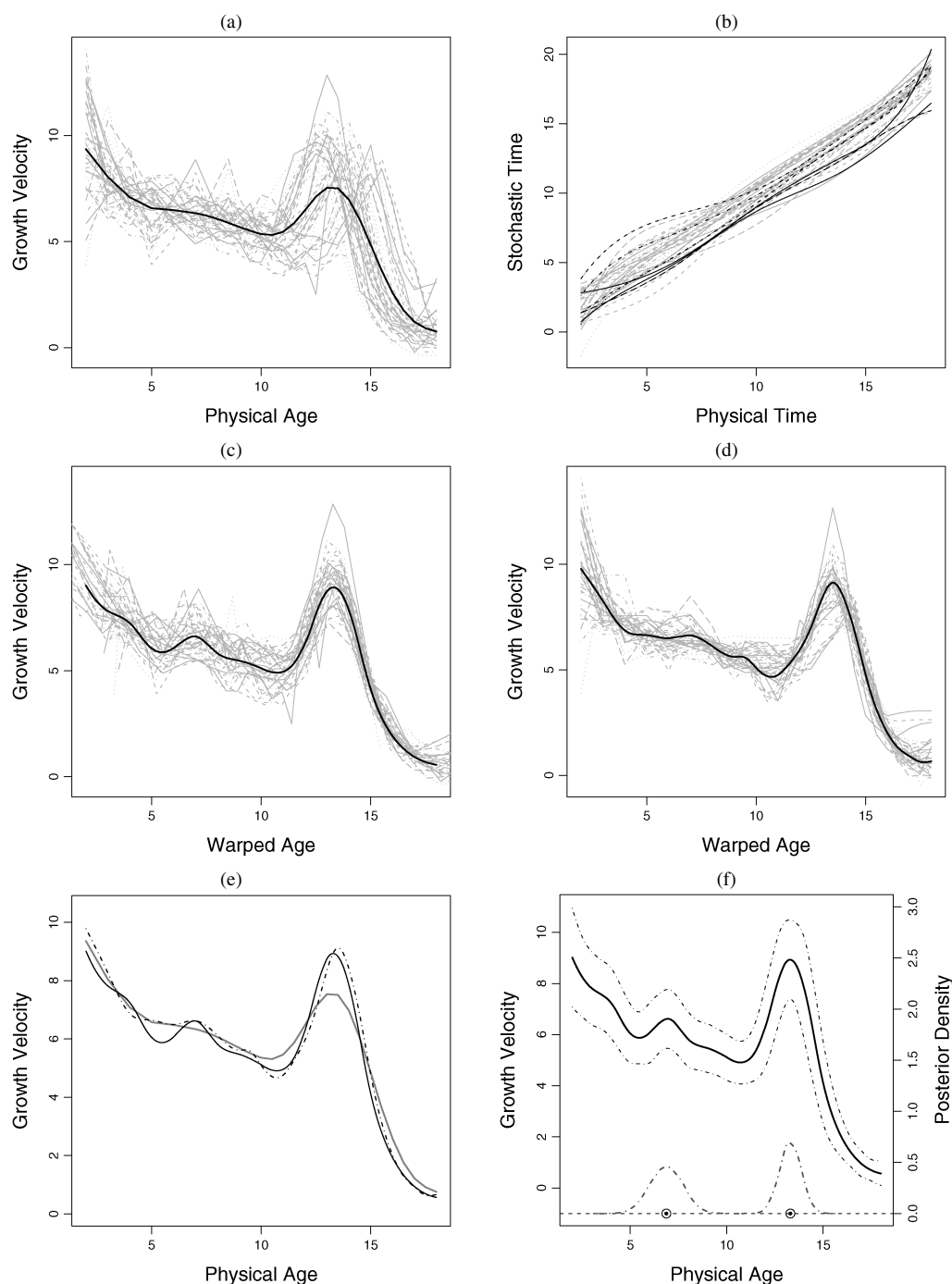
Figure 4. Case study 1: Height velocity data analysis. (a) Individual unregistered growth velocity curves for 39 boys, with cross-sectional mean in solid black. (b) Estimated curve-specific posterior mean time-transformation functions. (c) Thirty-nine curves registered by BHCR, with associated scaled posterior mean shape function. (d) Thirty-nine curves registered by SM, with associated registered mean. (e) Cross-sectional (solid gray), posterior mean shape from BHCR (solid black) and registered mean from SM (dashed black). (f) Posterior mean growth velocity from BHCR with 90% simultaneous credible bands, and posterior distributions of the growth spurts (enclosed circles represent the posterior mean age of the spurts).

3.4 Case Study 2: Analysis of Yeast Cell Cycle Gene Expression Data

Spellman et al. (1998) conducted a series of microarray experiments to create a comprehensive catalog of yeast genes with transcription levels that vary periodically within the cell cycle. In our analysis, we considered a sample of 100 differentially expressed genes from yeast cultures synchronized through the

α-factor arrest. We collected mRNA expression measurements every 7 minutes for 126 minutes, for a total of 18 measurements per gene.

Motivated by earlier work on the existence of time-delayed relationships as seen, for example, between transcription factors and their targets (see, e.g., Qian, Dolled-Filhart, Lin, Yu, and Gerstein 2001; Yu, Luscombe, Qian, and Gerstein 2003),

we used curve registration to investigate time-varying network relationships between genes. This is a nonstandard application of curve registration. The BHCR model presented earlier is extended; here the individual amplitude parameters a_i are defined over the entire real line $\Re$, to allow for inverse regulatory relationships between genes.

Because timing of expression features may be indicative of network relationships, inferences about gene–gene interactions may be based on similarities in their time-transformation functions. Global or local (time-specific) measures of similarity may be based on distance measures, $D(u_i(t), u_j(t))$, between individual time-transformation functions $u_i(t)$ and $u_j(t)$. Information about the order of influence [e.g., gene i affecting gene j $(i \rightarrow j)$] can be assessed with $\text{sign}(u_i(t) - u_j(t))$, where a positive sign indicates that $i \rightarrow j$ and a negative sign indicates that $j \rightarrow i$.

In our analysis, each time unit corresponds to 7 minutes in the original time scale. We estimated the common shape function, placing 27 equally spaced interior knots between -4 and 22 and considering a maximum expansion constraint $\Delta = 6$. To estimate the individual time-transformation functions, we partitioned the interval $T = [0, 18]$ into four subintervals, placing three interior knots at $\kappa = (4.5, 9.0, 13.5)$. We placed a $Gamma(.2; rate = 2)$ prior on the scale and amplitude preci-

sions and a $Gamma(.5; rate = 20)$ prior on the precision of the time-transformation parameters. We choose $Gamma(.1; 1)$ prior for the shape precision parameter. We based our inference on 15,000 samples from the posterior distribution obtained after discarding the initial 50,000 MCMC iterations for burn-in.

Figure 5(a) shows the curve-specific posterior mean time-transformation functions for 100 genes. Because all expression profiles are defined as random transformations of a common shape function, these time-transformation functions are informative about similarities between expression profiles. Up to scale, the closer together these functions, the more similar the fitted expression profiles. Panel (b) shows a random sample of gene expression profiles with associated posterior mean fit and corresponding 90% simultaneous credible bands. Once the top 50 genes with highest absolute mean amplitude are selected, panel (c) shows the clustered posterior mean absolute differences $E(D_{ij}(u_i; u_j)|\mathbf{y})$ between each pair of genes with $D_{ij}(u_i; u_j) = \sum_t |u_i(t) - u_j(t)|$ for all pairs $i \neq j$. The closer the time-transformation functions, the darker the area in the gene–gene association matrix.

To show how similarities may help us identify interacting genes, we consider gene *YPR119W*, which in this experiment seems to exhibit a fairly distinct expression profile. Looking at the row or column values for this gene in the gene–gene as-
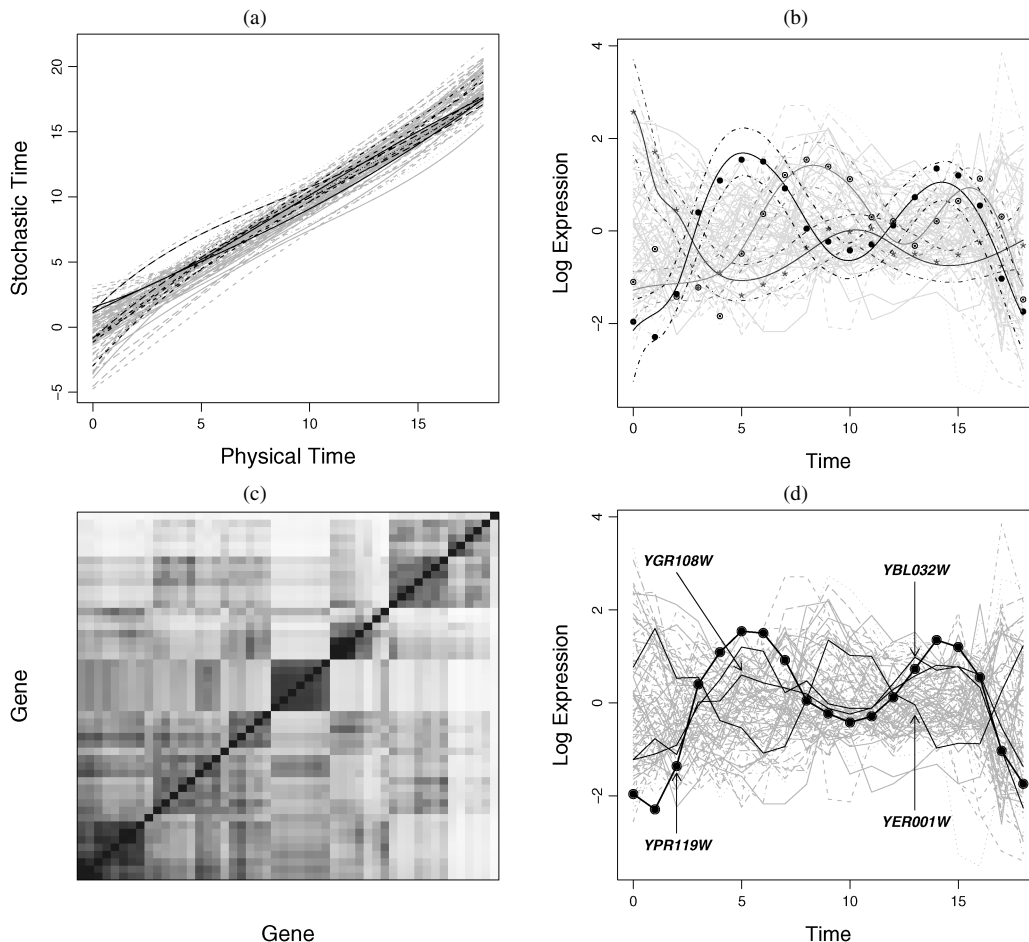

Figure 5. Cell cycle data analysis. (a) Posterior mean time-transformation functions for 100 genes. (b) Posterior mean expression profiles and 90% CI for a random sample of three genes. (c) Posterior mean absolute distance matrix (darker areas indicate closer time-transformation distances). (d) Gene YPR119W (dark solid+bullets) with its three best matches: YGR108W, YER001W, and YBL032W (dark solid trajectories).

Table 1. Top gene profile similarities (lowest posterior mean distance)

Gene 1 ORF (Gene)	Gene 2 ORF (Gene)	Relationship	Notes
YJL159W (*HSP150*)	YKL163W (*PIR3*)	Coexpression	Both members of the PIR family with similar functions.
YGR109C (*CLB6*)	YKL164C (*PIR1*)	Inhibition	No known interactions.
	YBL003C (*HTA2*) YNL030W (*HHF2*)		All histones, associated with cell cycle and DNA processing.
YDR225W (*HTA1*)	YBR010W (*HHT1*) YNL031C (*HHT2*) YDR224C (*HTB1*)	Coexpression	All present evidence of physical interactions from mass spectrometry studies.
YBR010W (*HHT1*)	YNL030W (*HHF2*)	Coexpression	Both histones with similar functions. Physical interactions were observed in gel retardation experiments.
YPL031C (*SSG3*)	YFR015C (*RME*)	Inhibition	Both related to cell cycle and DNA processing.
YPL031C (*SSG3*)	YGR044C (*GSY1*)	Coexpression	Both related to metabolism and energy.
YKR013W (*PRY2*)	YPL163C (*SVS1*)	Coexpression	PRY2 has no known function.
YNL030W (*HHF2*)	YNL031C (*HHT2*)	Coexpression	Physically interacting histones.
YKL164C (*PIR1*)	YKL096W (*CWP1*)	Inhibition	Both genes play a role in the biogenesis of cellular components.
YDR224C (*HTB1*)	YPL127C (*HH01*)	Coexpression	Histones with similar functions.
YKL185W (*ASH1*)	YDR055W (*PST1*)	Coexpression	Similar cellular functions.

NOTE: Descriptions extracted from *http://mips.gsf.de/genre/proj/yeast/*.

sociation matrix represented in Figure 5(c), we find that the best three matches with the gene of interest are *YGR108W*, *YER001W*, and *YBL032W*. Panel (d) shows the expression profile for our gene of interest, together with its three best matches. We note that the model identifies relationships between genes that exhibit similar timing in their profile features, relating genes that appear to be coexpressed as well as genes that seem to exhibit patterns of inhibition. In particular, the interaction between *YPR119W* and *YGR108W* can be explained by the fact that both are cell-cycle regulators involved with a common complex. The interactions with *YER001W* and *BYL032W* have not been previously reported in the literature. Table 1 reports the top 15 relationships with the lowest expected posterior time-transformation distance. For most known genes, the inferred relationships have biological support, because they share similar cellular functions and/or have been reported to interact with each other.

4. DISCUSSION

In this article we have proposed a flexible Bayesian hierarchical model for curve registration. As opposed to most methods for curve registration, our model does not require preliminary smoothing of the data, and it provides a natural framework for assessing uncertainty in the estimated time-transformation and shape functions and allows us to derive exact inferences about a richer set of quantities of interest.

Our simulation study showed that BHCR can adequately capture true signals and time-transformation functions even in the presence of noisy or sparse data. Moreover, our simulation study showed that our model has better performance compared with existing methods for curve registration. We also applied our BHCR model to two data sets. In the application to the

height velocity data set, our analysis indicated that the subject-specific curves had temporal misalignments. The latter feature has not been previously investigated in related data sets. We also applied our model to a nonstandard problem investigating network relationships using time-course microarray data. The advantage of this approach over traditional methods to network analysis (see, e.g., Friedman, Linial, Nachman, and Pe'er 2000; Inoue, Neira, Nelson, Gleave, and Etzioni 2007) is that the complexity of the problem increases only linearly with the number of genes. Moreover, unlike traditional methods, the time-transformation functions allow us to investigate time-varying networks without increasing the complexity of the model.

Although our model allows for model-based smoothing with the penalized B-splines, it requires specification of the number of the B-spline basis representing the common shape and time-transformation functions. In our applications, when modeling the common shape function, we considered a relatively large number of equally spaced knots, spanning the extended domain $[t_1 - \Delta, t_n + \Delta]$. In our experience, a number of knots reproducing the frequency of the sampling design achieved a good smoothing performance. In modeling the time-transformation functions, the number of knots can be more parsimonious because of the property of monotonicity, which carries along structural smoothness of the function. In our simulation studies and data analysis, between one and five interior knots was usually sufficient for registering curves. We performed additional simulation studies increasing the number of knots in the time-transformation or shape functions to assess the effect of these choices in our posterior inferences. The results (not shown) were robust to different choices of the number of knots. The choice of the temporal misalignment window can be set in a model selection framework considering measures such as the

Akaike or Bayesian information criterion. The value of Δ is constrained by the length of the sampling design T. In our applications we took Δ less or equal to $(t_n - t_1)/2$. Finally, priors for the variance hyperparameters should be relatively diffuse in relation to the scale of the original data $(\lambda, \sigma_a^2, \sigma_c^2)$ and to the time-length of the sampling interval (σ_ϕ^2); for example, if the sampling design spans the interval $T = [0, 50]$, then we can choose a prior with the property that about 60% of the prior density (in the SD scale) is below 50.

Our BHCR model can be extended in various directions. First, the shrinkage procedure outlined in this article may be inappropriate when the underlying shape function is highly oscillating (which was not the case in any of our applications). A possible solution, offering flexible shrinkage, is to replace the global variance (smoothing) parameter for the shape coefficients with a set of local smoothing parameters. Second, in our hierarchical model, we assume that curve-specific parameters are drawn from a common set of population distributions. Dependence may differ across curves, however. For example, suppose that an analysis of growth curves for boys and girls. Although it may be reasonable to assume dependence between growth curves for a given gender, it may not be reasonable to assume that the dependence is the same for both groups. To address this issue, we may consider additional levels of hierarchy in the model or include covariates. In fact, this procedure can be used when the assumption of a common shape function is not valid across all curves but is valid between subgroups identified with measured covariates. In a related issue, in our application to the time course microarray data set, our common shape assumption may be reasonable, because yeast genes might be regulated in a manner that coincides with the cell cycle (Spellman et al. 1998). However, it may not be reasonable when considering genes from other systems. In particular, we note that we may not be able to measure covariates that identify subgroups of genes that share a common shape function. Thus a possible alternative is to consider a mixture model of the time-transformed shape functions. Finally, in relation to the application to network inference, standard procedures that control for the posterior expected Bayesian false discovery rate, as well as decision-theoretic techniques, may be used to address the issue of multiple comparisons when testing network relationships.

[Received November 2006. Revised August 2007.]

REFERENCES

Baladandayuthapani, V., Mallick, B. K., and Carroll, R. J. (2005), "Spatially Adaptive Bayesian P–Splines With Heteroscedastic Errors," *Journal of Computational & Graphical Statistics*, 14, 378–394.

Brumback, C. L., and Lindstrom, J. M. (2004), "Self-Modeling With Flexible, Random Time Transformations," *Biometrics*, 60, 461–470.

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

Eilers, P., and Marx, B. (1996), "Flexible Smoothing Using B–Splines and Penalized Likelihood" (with discussion), *Statistical Science*, 11, 1200–1224.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000), "Using Bayesian Networks to Analyze Expression Data," *Journal of Computational Biology*, 7, 601–620.

Gamerman, D. (1997), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, London: Chapman & Hall.

Gasser, T., and Kneip, A. (1995), "Searching for Structure in Curve Samples," *Journal of the American Statistical Association*, 90, 1179–1188.

Gervini, D., and Gasser, T. (2004), "Self-Modelling Warping Functions," *Journal of the Royal Statistical Society*, Ser. B, 66, 959–971.

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

Inoue, L. Y. T., Neira, M., Nelson, C., Gleave, M., and Etzioni, R. (2007), "Cluster-Based Network Model," *Biostatistics*, 8, 507–525.

Kneip, A., and Gasser, T. (1988), "Convergence and Consistency Results for Self-Modeling Nonlinear Regression," *The Annals of Statistics*, 16, 82–112.

——— (1992), "Statistical Tools to Analyze Data Representing a Sample of Curves," *The Annals of Statistics*, 20, 1266–1305.

Kneip, A., Li, X., MacGibbon, K. B., and Ramsay, J. O. (2000), "Curve Registration by Local Regression," *The Canadian Journal of Statistics*, 28, 19–29.

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

Lawton, W. H., Sylvestre, E. A., and Maggio, M. S. (1972), "Self-Modeling Nonlinear Regression," *Technometrics*, 14, 513–532.

Liu, X., and Müller, H. (2004), "Functional Averaging and Synchronization for Time-Warped Random Curves," *Journal of the American Statistical Association*, 99, 687–699.

Qian, J., Dolled-Filhart, M., Lin, J., Yu, H., and Gerstein, M. (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., Bock, R. D., and Gasser, T. (1995), "Comparison of Height Acceleration Curves in the Fels, Zurich and Berkeley Growth Data," *Annals of Human Biology*, 22, 413–426.

Ramsay, J. O., and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society*, Ser. B, 60, 351–363.

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

——— (2002), *Applied Functional Data Analysis: Methods and Case Studies*, Berlin: Springer-Verlag.

Ronn, B. (2001), "Nonparametric Maximum Likelihood Estimation for Shifted Curves," *Journal of the Royal Statistical Society*, Ser. B, 63, 243–259.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge, U.K.: Cambridge University Press.

Sakoe, H., and Chiba, S. (1978), "Dynamic Programming Optimization for Spoken Word Recognition," *IEEE Transactions on Acoustic, Speech and Signal Processing*, 26, 43–49.

Silverman, B. W. (1995), "Incorporating Parametric Effects Into Functional Principal Components Analysis," *Journal of the Royal Statistical Society*, Ser. B, 57, 673–689.

Smith, B. J. (2005), *Bayesian Output Analysis Program* (*BOA*), Version 1.1.5 (online), The University of Iowa, *http://www.public-health.uiowa.edu/boa*.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998), "Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast *Saccharomyces cerevisiae* by Microarray Hybridization," *Molecular Biology of the Cell*, 9, 3273–3297.

Tuddenham, R. D., and Snyder, M. M. (1954), "Physical Growth of California Boys and Girls From Birth to Eighteen Years," *University of California Publications in Child Development*, 1, 183–364.

Wang, K., and Gasser, T. (1997), "Alignment of Curves by Dynamic Time Warping," *The Annals of Statistics*, 27, 1251–1276.

——— (1999), "Synchronizing Sample Curves Nonparametrically," *The Annals of Statistics*, 27, 439–460.

Yu, H., Luscombe, N. M., Qian, J., and Gerstein, M. (2003), "Genomic Analysis of Gene Expression Relationships in Transcriptional Regulatory Networks," *Trends in Genetics*, 19, 422–427.