

Pairwise Curve Synchronization for Functional Data

BY RONG TANG

Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration, Rockville, Maryland 20850, U.S.A.

rong.tang@fda.hhs.gov

AND HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, California 95616, U.S.A.

mueller@wald.ucdavis.edu

SUMMARY

Increasingly, data collected by scientists in different fields are in the form of trajectories or curves. These curves can often be viewed as realizations of a composite process driven by both amplitude and time variation. We consider the situation where functional variation is dominated by time variation, and develop a curve-synchronization method that uses every trajectory in the sample as a reference to obtain pairwise warping functions in a first step. These initial pairwise warping functions are then used to create improved estimators of the underlying individual warping functions in a second step. A truncated averaging process is used to obtain robust estimation of individual warping functions. The method compares well with other available warping approaches and is illustrated with Berkeley growth data and gene expression data for multiple sclerosis.

Some key words: Alignment; Curve registration; Functional data analysis; Gene expression profiles; Multiple sclerosis; Time warping.

1. INTRODUCTION

Functional data have become very popular in many scientific fields, leading to the development of functional data analysis (Ramsay & Silverman, 2005). A challenge

is raised by the possibility of confounding between amplitude and time variation when observed trajectories share the same underlying amplitude function, but each evolves on its own time-scale. For example, growth velocity curves, see Fig. 1, of the Berkeley growth study (Ramsay & Silverman, 2002) exhibit a similar growth pattern throughout, but the velocity peaks occur at different times for each child. Traditionally, the cross-sectional mean is used to represent group profiles. However, because of the variation in the timing of growth spurts or of peaks in microarray gene expression profiles (Kaminski & Bar-Joseph, 2006), the cross-sectional mean may give an inaccurate summary of the underlying functional shapes.

A solution to this problem is to recognize time warping and to include it in the statistical modelling of functional data. Time warping can be viewed as a transformation that maps each trajectory from an internal trajectory-specific clock to absolute time (Liu & Müller, 2005). Warping was first used to synchronize speech signals (Sakoe & Chiba, 1978), and subsequently Gasser & Kneip (1992, 1995) proposed a landmark registration method that aligns continuous curves by matching locations of individual landmarks to the average location across the entire sample. This method works very well when all the observed curves have clearly identifiable common features. However, it is not easy to define features common to every curve, especially with noisy data. Visual inspection of the entire dataset and manual identification of the landmarks in each curve is usually required, making this method impractical for large datasets. Another well-known curve registration method is the Procrustes method (Ramsay & Li, 1998) that iteratively aligns each curve to the current cross-sectional mean of the sample and therefore relies heavily on the initial cross-sectional mean.

Other recent developments in curve registration and warping include self-modelling warping (Gervini & Gasser, 2004); nonparametric maximum-likelihood warping (Rønn, 2001; Gervini & Gasser, 2005); functional convex synchronization (Liu & Müller, 2004); dynamic time warping (Wang & Gasser, 1997, 1999), and a synchronization method based on a novel definition of moments of curves (James, 2007).

Here, we introduce a version of pairwise time synchronization that makes full use of information across samples and can be used to provide a robust structural mean estimator.

2. MODELS AND METHODS

2.1. Time-warped functional data

Let Y_1, Y_2, \dots, Y_n denote n continuous curves, defined on a closed real interval $\mathcal{T} = [0, T]$. These curves are observed at discrete time points t_j , $j = 1, \dots, m$, and the observed data for each curve are (t_j, y_{ij}) . The model for y_{ij} is

$$y_{ij} = Y_i(t_j) + \epsilon_{ij} = X_i\{h_i^{-1}(t_j)\} + \epsilon_{ij}, \quad t_j \in \mathcal{T}, \quad (1)$$

where the error terms ϵ_{ij} are independently distributed random noise with $E(\epsilon) = 0$ and $E(\epsilon^2) = \sigma^2 < \infty$, and X is a random function defined on \mathcal{T} with independent realizations X_i . The time-synchronized random functions X_i are modelled as

$$X_i(t) = \mu(t) + \delta Z_i(t), \quad \text{for } t \in \mathcal{T}, \quad (2)$$

where μ is a fixed function and Z_i are independent random trajectories, which are realizations of a process Z with $E\{Z(t)\} = 0$ and $E\{Z^2(t)\} < \infty$ for $t \in \mathcal{T}$. The coefficient δ is assumed to be a small constant that converges to 0 with increasing sample size. The functions $h_i : \mathcal{T} \mapsto \mathcal{T}$ in model (1) are time-synchronization or warping functions, corresponding to realizations of a random synchronization function h . The inverse mapping h^{-1} describes the internal time scale of a trajectory.

The random warping functions h are assumed to satisfy the conditions of common endpoints, i.e. $h(0) = 0$, $h(T) = T$, strict monotonicity, i.e. $h(t_1) < h(t_2)$, for $0 \leq t_1 < t_2 \leq T$, and average identity, i.e. $E\{h(t)\} = t$, for $t \in \mathcal{T}$. The monotonicity condition requires all warping functions to comply with the fact that time is always moving forward and ensures that h is invertible.

2.2. Modelling warping functions

In the proposed method, the warping functions h are modeled by piecewise-linear functions. More general splines could be considered and this assumption is primarily

made to ensure easy implementation of the constraints such as monotonicity. The linear splines also work well in practice. The corresponding warping functions are defined as the linear interpolation of p pre-selected equidistant knots (a_1, \dots, a_p) , where $a_j = Tj/(p+1)$, $j = 1, \dots, p$, for a fixed $p \geq 1$. If $\tau_j = h(a_j)$, the warping function h is

$$h(t) = \begin{cases} \tau_1 t/a_1, & \text{for } 0 \leq t < a_1, \\ (\tau_j - \tau_{j+1})(t - a_j)/(a_j - a_{j+1}) + \tau_j, & \text{for } a_j \leq t < a_{j+1}, j = 1, \dots, p-1, \\ (\tau_p - T)(t - T)/(a_p - T) + T, & \text{for } a_p \leq t \leq T. \end{cases} \quad (3)$$

If we set $a_0 = 0$, $a_{p+1} = T$, and require that $\tau_{j-1} < \tau_j$, $j = 1, \dots, p+1$, then $\tau_0 = h(a_0) = 0$, and the common endpoint and strict monotonicity constraints will be satisfied. Noting that (3) can be equivalently expressed as $h(t) = \sum_{j=1}^{p+1} \tau_j A_j(t)$, where $A_j(t) = B_j(t) - C_{j+1}(t)$, $j = 1, \dots, p+1$, are linear basis functions, with $B_j(t) = (t - a_{j-1})1_{[a_{j-1}, a_j]}/(a_j - a_{j-1})$, and $C_{j+1}(t) = (t - a_{j+1})1_{[a_j, a_{j+1})}/(a_{j+1} - a_j)$, $C_{p+2} = 0$, $j = 1, \dots, p+1$, we obtain with $\Theta = (\tau_1, \dots, \tau_{p+1})^\top$ and $A(t) = \{A_1(t), \dots, A_{p+1}(t)\}^\top$ that

$$h(t) = \Theta^\top A(t). \quad (4)$$

The parameter space of the coefficient vectors of the warping functions (4) is

$$\Omega = \{\Theta \in \mathcal{T}^{p+1} \mid \Theta = (\tau_1, \dots, \tau_{p+1})^\top, 0 < \tau_1 < \dots < \tau_{p+1} \equiv T\},$$

with corresponding family of warping functions $\mathcal{W} = \{h : \mathcal{T} \mapsto \mathcal{T} \mid h = \Theta^\top A, \Theta \in \Omega\}$. The norms of $\Theta \in \Omega$ and of $h \in \mathcal{W}$ are $\|\Theta\| = (\sum_{j=1}^{p+1} \tau_j^2)^{1/2}$ and $\|h\| = \{\int_{\mathcal{T}} h^2(t) dt\}^{1/2}$. The model in (4) reduces the problem of estimating the entire function h to estimating the p spline coefficients.

2.3 Pairwise and global warping

For a pair of curves Y_i and Y_k , the pairwise-synchronization or warping function g_{ik} is a transformation of the time scale of curve Y_i towards that of Y_k , defined by

$$g_{ik}(t) = h_i\{h_k^{-1}(t)\}. \quad (5)$$

Similarly, the pairwise-warping function of curve Y_k towards Y_i is $g_{ki}(t) = h_k\{h_i^{-1}(t)\}$. We assume that the g_{ik} can also be represented by (4), such that $g_{ik}(t) = \Theta_{g_{ik}}^T A(t)$, for suitable coefficient vectors $\Theta_{g_{ik}}$, $1 \leq i, k \leq n$. Initially, we consider the estimation of pairwise warping functions when the entire trajectories Y_i , $i = 1, \dots, n$, are observed. Estimation based on observed discrete data will be discussed in §2.5. Additional assumptions are as follows.

Assumption 1. The Z_i are independent, such that $E\{Z_i(t)\} = 0$ for $t \in \mathcal{T}$, and the Z_i are independent of h_i , $i = 1, \dots, n$.

Assumption 2. The Z_i are twice continuously differentiable on \mathcal{T} , and there exists $0 < C < \infty$ such that $E\{Z_i(t)^2\} < C$, $E\{Z_i'(t)^2\} < C$ and $E\{Z_i''(t)^2\} < C$, for all $t \in \mathcal{T}$.

Assumption 3. The mean function μ is a bounded and twice continuously differentiable function with $\int_{\mathcal{T}} \mu''(t)^2 dt < \infty$. For any nondegenerate interval $\mathcal{T}_o \subseteq \mathcal{T}$, $0 < \int_{\mathcal{T}_o} \mu'(t)^2 dt < \infty$.

Assumption 4. For any $t_1, t_2 \in \mathcal{T}$ and $t_1 < t_2$, there exist $0 < \omega_1 < \omega_2 < \infty$ and $\omega_1 \omega_2 \leq 1$ such that $\omega_1 \leq \{g(t_1) - g(t_2)\}/(t_1 - t_2) \leq \omega_2$, for any $g \in \mathcal{W}$. This assumption prevents plateaus and abrupt jumps in the pairwise warping functions and their inverse functions.

A preliminary estimator of $\Theta_{g_{ik}}$ can be obtained by minimizing a distance measure between curves Y_i and Y_k with respect to $\Theta \in \Omega$; that is

$$\tilde{\Theta}_{g_{ik}} = \arg \min_{\Theta \in \Omega} C_\lambda(Y_i, Y_k, \Theta),$$

$$C_\lambda(Y_i, Y_k, \Theta) = E\left(\int_{\mathcal{T}} \left[\{Y_i(\Theta^T A(t)) - Y_k(t)\}^2 + \lambda\{\Theta^T A(t) - t\}^2\right] dt \mid Y_i, Y_k\right), \quad (6)$$

where $\lambda \geq 0$. The penalty term $\lambda\{\Theta^T A(t) - t\}^2$ is introduced to regularize the estimator and to avoid shape distortion. Consequently, a preliminary estimator of the pairwise warping function can be obtained by $\tilde{g}_{ik}(t) = \tilde{\Theta}_{g_{ik}}^T A(t)$.

One can replace Y_i and Y_k in (6) by their derivatives. An example for synchronization in the first derivative is provided in §4.1. In the pairwise synchronization

step, each curve Y_k in the sample serves as template for all other curves in turn. The resulting \tilde{g}_{ik} , $i = 1, \dots, n$, then lead to the global warping function estimate h_k for curve Y_k , as described below. Constructing global warping from pairwise information within a simple time-shift framework has been considered in Leng & Müller (2006).

Since h_k corresponds to a map between an individual-specific warped time to absolute time, we refer to h_k as a global warping function. For $s = h_k^{-1}(t)$, $t \in \mathcal{T}$, one has $h_i(s) = h_i\{h_k^{-1}(t)\}$. Because warping functions are assumed to have average identify, $E[h_i\{h_k^{-1}(t)\}|h_k] = h_k^{-1}(t)$, and, as $g_{ik}(t) = h_i\{h_k^{-1}(t)\}$ (5), we find that $h_k^{-1}(t) = E\{g_{ik}(t)|h_k\}$, motivating estimators

$$\tilde{h}_k^{-1}(t) = \frac{1}{n} \sum_{i=1}^n \tilde{g}_{ik}(t) \quad (7)$$

for the inverse global warping functions; by simple inversion, one obtains the corresponding \tilde{h}_k and aligned curves $Y_i^*(t) = Y_i\{\tilde{h}_i(t)\}$.

2.4 Estimation

If only discrete and possibly noisy measurements are available for the trajectories, one may use weighted local linear regression (Fan & Gijbels, 1996) for presmoothing. If observations (y_{i1}, \dots, y_{im}) for the i th trajectory are recorded at $t_j = Tj/m$, the weighted local linear fits are the minimizers of

$$\sum_{j=1}^m K\left(\frac{t-t_j}{b}\right)[y_{ij} - \{\alpha_0 + \alpha_1(t_{ij} - t)\}]^2, \quad (8)$$

with respect to α_0, α_1 , leading to estimates $\hat{Y}_i(t) = \hat{\alpha}_0(t)$. We require two more assumptions.

Assumption 5. The bandwidth sequence b satisfies $b \rightarrow 0$ and $mb \rightarrow \infty$ as $m \rightarrow \infty$.

Assumption 6. The nonnegative kernel function K is Lipschitz continuous with support $[-1, 1]$.

Replacing Y_i in (6) with the fitted curves \hat{Y}_i , we obtain

$$\hat{C}_\lambda(Y_i, Y_k, \Theta) = \int_{\mathcal{T}} \left[\{\hat{Y}_i(\Theta^T A(t)) - \hat{Y}_k(t)\}^2 + \lambda \{\Theta^T A(t) - t\}^2 \right] dt,$$

$$\widehat{\Theta}_{g_{ik}} = \arg \min_{\Theta \in \Omega} \widehat{C}_\lambda(Y_i, Y_k, \Theta), \quad (9)$$

and the resulting estimated pairwise warping functions

$$\widehat{g}_{ik}(t) = \widehat{\Theta}_{g_{ik}}^\top A(t). \quad (10)$$

Theorem 1. *Under Assumptions 1-6, for any $1 \leq i, k \leq n$ and with bandwidths $b = b_{\text{opt}} \sim m^{-1/5}$ as $m \rightarrow \infty$, the constrained minimizer $\widehat{\Theta}_{g_{ik}}$ (9) obtained from y_{ij} satisfies*

$$\|\widehat{\Theta}_{g_{ik}} - \Theta_{g_{ik}}\| = O\{(\delta^2 + \lambda)^{1/2}\} + O_p(m^{-1/5}), \quad 1 \leq i, k \leq n, \quad (11)$$

and the estimated pairwise warping functions \widehat{g}_{ik} (10) satisfy

$$\sup_{t \in \mathcal{T}} |\widehat{g}_{ik}(t) - g_{ik}(t)| = O\{(\delta^2 + \lambda)^{1/2}\} + O_p(m^{-1/5}), \quad 1 \leq i, k \leq n. \quad (12)$$

Estimators for the inverse warping functions h_k^{-1} are then obtained as

$$\widehat{h}_k^{-1}(t) = \frac{1}{n} \sum_{i=1}^n \widehat{g}_{ik}(t), \quad k = 1, \dots, n, \quad (13)$$

and for $h_k(t)$ by exchanging the coordinates.

Corollary 1. *Under Assumptions 1-6, the estimated warping functions \widehat{h}_k , $k = 1, \dots, n$, satisfy*

$$\sup_{t \in \mathcal{T}} |\widehat{h}_k(t) - h_k(t)| = O\{(\delta^2 + \lambda)^{1/2}\} + O_p(m^{-1/5}), \quad k = 1, \dots, n. \quad (14)$$

A more robust alternative to (13) can be obtained through a truncated averaging procedure in which only the pairs whose distance after pairwise synchronization is smaller than a threshold are used in the estimation of the global warping function h_k . This idea is implemented by

$$\widehat{h}_k^{-1*}(t) = \frac{\sum_{i=1}^n \widehat{g}_{ik}(t) 1_{\{d_{pw}(\widehat{Y}_i, \widehat{Y}_k) < D_{k_r}\}}}{\sum_{i=1}^n 1_{\{d_{pw}(\widehat{Y}_i, \widehat{Y}_k) < D_{k_r}\}}}, \quad (15)$$

where $d_{pw}(\widehat{Y}_i, \widehat{Y}_k) = [\int_{\mathcal{T}} \{\widehat{Y}_i(\widehat{g}_{ik}(t)) - \widehat{Y}_k(t)\}^2 dt]^{1/2}$, $1 \leq i, k \leq n$, and D_{k_r} is the r th quantile of $D_k = \{d_{pw}(\widehat{Y}_1, \widehat{Y}_k), \dots, d_{pw}(\widehat{Y}_n, \widehat{Y}_k)\}$, $k = 1, \dots, n$, for a suitably chosen

$0 < r < 1$. A good empirical choice for the threshold is $r = 0.9$. In all implementations of pairwise synchronization, we use the above version (15).

In applications, different curves might have different amplitude scales. In such situations, a more realistic model for the data would be $y_{ij} = \alpha_i X_i\{h_i^{-1}(t_j)\} + \epsilon_{ij}$, where α_i are random amplitudes. A simple approach in this setting is to normalize the functions before the synchronization step. We adopt the sup-norm $\|Y\|_\infty = \sup_{t \in \mathcal{T}} |Y(t)|$ to obtain normalized trajectories $\check{Y}_i(t) = Y_i(t)/\|Y_i\|_\infty$. A viable approach for accelerating the necessary computations compared to the full pairwise synchronization is to select a random subset of partner functions for a given function. The pairwise synchronization and averages (7) and (15) are then computed based on these subsets only. This subset method exacts a price in terms of slower convergence, but will reduce the computing burden substantially, especially for larger sample sizes n .

Corollary 1 also implies consistency of estimated synchronized curves. From (1), based on the estimates \hat{Y}_i and \hat{h}_k above, the estimate for the i th time-synchronized curve X_i becomes

$$\hat{X}_i(t) = \hat{Y}_i\{\hat{h}_i(t)\}. \quad (16)$$

Under the basic assumptions and adopting the necessary regularity conditions for uniform consistency in nonparametric regression (Mack & Silverman, 1982; Müller & Stadtmüller, 1987), and assuming in addition that $b \sim m^{-1/5}$ and $\|X_i^{(\nu)}\|_\infty = O_p(1)$ for $\nu = 1, 2$, we find

$$\|\hat{X}_i - X_i\|_\infty \leq \|\hat{Y}_i - Y_i\|_\infty + O_p(1)\|\hat{h}_i - h_i\|_\infty = O_p(\delta^2 + \lambda^{1/2} + m^{-1/5}),$$

i.e., the same rate as for the warping functions. This is a consequence of the uniform rate of convergence of the smoothed trajectories $\|\hat{Y}_i - Y_i\|_\infty = O_p(\log n/(nb) + b^2)$.

3. SIMULATIONS

In this section, the performance of the proposed pairwise synchronization is compared with several other existing warping methods, including Procrustes (Ramsay & Li, 1998), self-modelling registration (Gervini & Gasser, 2004) and nonparametric

maximum likelihood (Gervini & Gasser, 2005) methods. In each simulation, we assume $n = 20$ sample curves, and the reported results are based on 100 Monte Carlo runs.

Sample curves were generated on $[0, 1]$ according to (1); that is

$$y_{ij} = X_i\{h_i^{-1}(t_j)\} + \epsilon_{ij}, \quad i = 1, \dots, n, \quad t_j = j/m, \quad j = 0, \dots, m,$$

with $m = 50$. The error terms ϵ_{ij} were generated independently from $N(0, \sigma^2)$ with σ chosen such that the noise-to-signal ratio was 0.25. The components in (2) were chosen as follows: the random function Z was $Z(t_j) = \varsigma_{1j} + \varsigma_{2j}\sqrt{2}\sin(2\pi t_j)$, with $\varsigma_{1j} \sim N(0, 1)$ and $\varsigma_{2j} \sim N(0, 1/4)$ independent of each other, δ was 0.01, and the mean shape functions were

$$\begin{aligned} \mu_1(t) = & 0.88\exp\{-20(t - 0.7)^2\} - 0.5\exp\{-50(t - 0.45)^2\} + 0.6\exp\{-100(t - 0.3)^2\} \\ & - 0.6\exp\{-150(t - 0.2)^2\} + 0.5\exp\{-200(t - 0.15)^2\} \end{aligned} \quad (17)$$

and

$$\begin{aligned} \mu_2(t) = & 0.88\exp\{-20(t - 0.7)^2\} - 0.5\exp\{-50(t - 0.45)^2\} + 0.6\exp\{-100(t - 0.3)^2\} \\ & - 0.6\exp\{-150(t - 0.2)^2\} + 0.5\exp\{-200(t - 0.15)^2\} - 2.863\phi\{(t - 0.8)/0.01\} \\ & - 0.5\phi\{(t - 0.2)/0.2\}, \end{aligned} \quad (18)$$

where ϕ represents the density function of standard normal distribution. Both mean functions have three peaks, μ_2 providing a more difficult setting since its two early peaks have very similar heights; see Fig. 2.

The warping functions were generated by the area under a linear combination of cubic B -spline basis functions with equidistant knots: $l_i(t) = \sum_{j=1}^5 \alpha_{ij}B_j(t)$. To ensure that $h_i(0) = 0$, $\alpha_{i,1}$ was set to 0, and the other α_{ij} were randomly drawn from a uniform distribution on the interval $[0, 100]$, with the i th warping function generated as $h_i(t) = \int_0^t l_i(s)ds / \int_0^1 l_i(s)ds$. These warping functions are monotone increasing because cubic B -spline basis functions $l_i(t)$ are positive on the domain spanned by

the knots and zero elsewhere (Eilers & Marx, 1996), and correspond to considerable time distortion.

For the comparison of time-warping methods, we calculated two performance measures for time-synchronized curve estimates \hat{X}_i as in (16) and for estimated warping functions \hat{h}_i , namely

$$\text{FMISE} = \frac{1}{n} \sum_{i=1}^n \int_0^1 \{\hat{Y}_i^*(t) - \mu(t)\}^2 dt, \quad \text{HMISE} = \frac{1}{n} \sum_{i=1}^n \int_0^1 \{\hat{h}_i(t) - h_i(t)\}^2 dt. \quad (19)$$

For pairwise synchronization, the number of knots p and the penalty parameter λ were fixed at 3 and 10^{-3} respectively. These choices seemed reasonable, and the results for pairwise synchronization are not sensitive to the choice of p and λ , as demonstrated in Table 1. In general, choosing the number of knots p between 3 and 5 was adequate. Too many knots not only increase the computational burden, but may also cause shape distortion (Ramsay & Li, 1998).

For the Procrustes method, we used 5 break-points and the penalty parameter was set to $\lambda = 10^{-4}$, as in Ramsay & Li, (1998). The number of basis functions p was suggested to be approximately 3 – 4 times the number of features for self-modelling registration (Gervini & Gasser, 2004), motivating the choices 4 and 12. Finally, for the nonparametric maximum likelihood method, the location and scale parameters would be, ideally, chosen by maximum likelihood. However, this was not practical because of the large number of parameters, and we choose four equidistant points as the location parameters and $(1, 1, 1, 1)$ as the scale parameters (Gervini & Gasser, 2005).

The mean and variance of outcome measures FMISE and HMISE are summarized in Table 2. These deviation measures increase as the underlying mean shape function becomes more complicated. Regardless of the shape of the mean function, the pairwise synchronization method consistently produced results that were never inferior to those for the other methods and sometimes considerably better. Among the other methods, Procrustes and self-modelling registration had the next best performance.

Overall, pairwise synchronization is seen to work very well with noisy functional data that include large and irregular time variation. The simulation results are best taken with a grain of salt, as they are based on comparing the various methods at reasonable but not optimal parameter choices.

4. APPLICATIONS

4.1. Berkeley growth study

An important goal in growth studies is to obtain a good estimate of normal growth patterns. As is well known by now, the cross-sectional mean often fails adequately to reflect important growth features, as a result of time variation among individuals. Ramsay & Li (1998) and Gervini & Gasser (2004, 2005) were able to obtain better growth profile estimation by synchronizing growth curves before taking the cross-sectional average. We applied the proposed pairwise synchronization method to the Berkeley growth data (Tuddenham & Snyder, 1954), and compared the results with those for the Procrustes method, self-modelling registration, nonparametric maximum likelihood and cross-sectional mean.

The Berkeley growth data contain height measurements for 54 girls and 38 boys, with 31 measurements taken between the ages of 1 and 18 years. Of special interest is the instantaneous rate of change in height, that is, the velocity of the growth curve. The functional representation of height for each individual was obtained with a B -spline smoothing program from Ramsay & Silverman (2002, pp. 98-99), and the velocity curves by taking the first-order difference quotients of the smoothed growth curves. We report here the results for the 54 girls, whose velocity curves are plotted in Fig. 1. It is evident that all girls went through two growth spurts at different ages; a minor early peak, the so-called mid-growth spurt, followed by the prominent pubertal spurt. Ramsay & Silverman (2002) applied the Procrustes method to these data, using 5 break points and penalty parameter 0.5, and we adopted the same parameter values when implementing the proposed pairwise synchronization method. Gervini & Gasser (2005) applied self-modelling registration to the Zürich longitudinal

growth study with 4 components and 12 basis functions. These numbers also worked well for the Berkeley growth data. However, the parameters of the nonparametric maximum likelihood estimation method used in the Zürich growth study (Gervini & Gasser, 2004) performed poorly for this dataset. After trying various possible choices, we found that selecting location parameters (6, 12) and scale parameters (1, 1) worked reasonably well.

Comparing and evaluating various warping methods, we obtained the deviations between the resulting mean growth curves and the landmark registration mean curve (Kneip & Gasser 1992; Gasser & Kneip, 1995). Landmark registration serves as a benchmark for these data, as the growth velocity curves of the 54 girls are well structured with two well-defined peaks. As an overall measure of the difference between estimated mean growth curve and this benchmark we used the integrated square error $\text{ISE} = \int \{\bar{Y}_{\text{warp}}(t) - \bar{Y}_{\text{LM}}(t)\}^2 dt$, where \bar{Y}_{warp} is the average of all warped curves and \bar{Y}_{LM} is the benchmark result. Pairwise warping synchronization produces the smallest integrated square error, $\text{ISE} = 0.82$, followed by self-modelling registration, $\text{ISE} = 1.19$. The cross-sectional mean has the largest error, $\text{ISE} = 11.90$, as expected, and it also fails to capture the details of the pubertal spurt. The values of ISE of the Procrustes method and the nonparametric maximum likelihood method are 6.76 and 5.75, respectively, which are relatively large.

The estimated growth profile of the pairwise synchronization method deviates relatively little from the benchmark as can be seen from Fig. 3. The profile of self-modelling registration is also not far from the benchmark. The Procrustes method performs better around the pubertal spurt than the cross-sectional mean, but its onset timing is not very accurate, a deficiency shared by the nonparametric maximum likelihood method. In conclusion, all warping methods considered recover the two growth spurts and improve on the cross-sectional mean. Pairwise synchronization outperforms the other methods by producing a growth profile that closely resembles the benchmark in both intensity and timing of important growth events, and is

associated with the smallest value of ISE.

4.2. Multiple sclerosis data

Multiple sclerosis is an autoimmune disease in the central nerve system that often results in neurological disabilities (Zamvil & Steinman, 2003). To shed light on the mechanisms of a popular therapy, INF- β , at the genomic level, Weinstock-Guttman et al. (2002) conducted an experiment in which eight patients with active relapsing multiple sclerosis were treated with INF- β -1a. Peripheral blood samples were obtained from patients just before and at 1, 2, 4, 8, 24, 48, 120 and 168 hours after the treatment. The expression levels of 4324 gene probes in the blood samples were measured at these nine time-points for every patient. Weinstock-Guttman et al. (2003) investigated the expression patterns of a set of candidate genes, with the goal of identifying key genes that show a response to treatment, and subsequently Liang et al. (2005) refined the methodology to filter out nonresponsive genes. In both studies, the expression profile of each gene was estimated by averaging information across patients. While some of these gene expression trajectories display similar amplitude response patterns, indicating a common baseline expression coupled with time variation across patients, see Fig. 4 (a) and (b), others vary more widely across patients, see Fig. 4 (c) and (d), with no clearly discernible pattern of variation. Particularly in the first situation, we expect that estimation of an overall expression profile will benefit from a preliminary time-warping step.

Applying the proposed pairwise synchronization to gene expression profiles measured for each of 8 patients, using 24 genes that are known to be INF-stimulated, we calculated the between-patient variation before, VAR_b , and after, VAR_a , warping:

$$\text{VAR}_b = \frac{1}{8} \sum_{i=1}^8 [\widehat{Y}_i(t) - \text{ave}\{\widehat{Y}(t)\}]^2, \quad \text{VAR}_a = \frac{1}{8} \sum_{i=1}^8 [\widehat{X}_i(t) - \text{ave}\{\widehat{X}(t)\}]^2, \quad (20)$$

where $\text{ave}\{\widehat{Y}(t)\} = \sum_{i=1}^8 \widehat{Y}_i(t)/8$, and $\text{ave}\{\widehat{X}(t)\} = \sum_{i=1}^8 \widehat{X}_i(t)/8$, \widehat{Y}_i are the smoothed observed curves (8) and \widehat{X}_i are the time synchronized trajectory estimates (16). After pairwise synchronization, the patient response profiles are found to be more synchro-

nized, see Fig. 4 (e)-(h), and indeed the between-patient variation of the warped data becomes much smaller, especially for genes sharing similar baseline expression profiles, see Table 3. On average, there is a 39% reduction of between-patient variation after warping. The gene that shows the greatest relative reduction is Hs.274382, an interferon-inducible double-stranded RNA dependent gene, which is related to protein kinase, with a 92% reduction.

The reduced variation leads to improved gene expression profile estimation, which in turn may aid subsequent analysis such as identifying genes that are responsive to treatment. Moreover, the magnitude of between-patient variation before, VAR_b , and after, VAR_a , warping can be used to study the response behavior of different genes. Genes with large VAR_a , such as Hs.1279, Hs.79322, Hs.14623 and Hs.83795, see Table 3, tend to be more heterogeneous in their response patterns, see Fig. 4. Genes with large before warping variation VAR_b but small after warping variation VAR_a , such as Hs.833, Hs.241510 and Hs.62661 usually maintain similar baseline expression patterns across different patients and most of the observed between-patient variation is probably due to time variation. There are also genes displaying little variation both before and after warping, for example, Hs.80645, Hs.146360 or Hs.174195. These genes responded similarly in different patients in terms of both baseline profiles and expression timings. In the most extreme case, genes with very heterogeneous baseline profiles, see Fig. 4, may even have increased variation after warping, as exemplified by genes Hs.1279 and Hs.79322. This example demonstrates the potential of the pairwise synchronization method for microarray clinical studies. Aligned curves provide better estimates of gene response to treatment, and, by differentiating shape and time variation, one may gain a better understanding of the response patterns of gene expression.

ACKNOWLEDGEMENT

Thanks are due to the editor, the associate editor and two anonymous referees for helpful remarks. This research was supported in part by two grants from the U.S.

National Science Foundation.

APPENDIX

Auxiliary results and proofs

Lemma A1. *For any $h, h^* \in \mathcal{W}$ that satisfy $\mu\{h(t)\} = \mu\{h^*(t)\}$ where $\mu : \mathcal{S} \mapsto \mathcal{R}$ satisfies Assumption 3, it holds that*

$$h(t) = h^*(t). \quad (\text{A1})$$

Proof The domain \mathcal{S} can be divided into two non-overlapping subsets, $\mathcal{S}_a = \{s \in \mathcal{S} \mid \mu^{-1}\{\mu(s)\} = s\}$ and $\mathcal{S}_b = \mathcal{S} \setminus \mathcal{S}_a$. Since $h(t)$ is a one-to-one map, the domain \mathcal{T} can be divided into non-overlapping subsets \mathcal{T}_a and \mathcal{T}_b such that $\mathcal{T}_a = \{t \in \mathcal{T} \mid h(t) \in \mathcal{S}_a\}$ and $\mathcal{T}_b = \{t \in \mathcal{T} \mid h(t) \in \mathcal{S}_b\}$. For any $h, h^* \in \mathcal{W}$ that satisfy $\mu\{h(t)\} = \mu\{h^*(t)\}$, one can verify (A1) separately on \mathcal{T}_a and \mathcal{T}_b .

Lemma A2. *For any $1 \leq i, k \leq n$, the coefficient vector $\Theta_{g_{ik}}$ for g_{ik} is also the unique constrained minimizer of $C_\mu(h_i, h_k, \Theta)$, where*

$$C_\mu(h_i, h_k, \Theta) = E \left(\int_{\mathcal{T}} [\mu\{h_i^{-1}(\Theta^T A(t))\} - \mu\{h_k^{-1}(t)\}]^2 dt \mid Y_i, Y_k \right),$$

subject to the constraints

$$I(\Theta) = \tau_{p+1} - T = 0, \quad J_j(\Theta) = \tau_{j-1} - \tau_j + \beta \leq 0, \quad j = 1, \dots, p+1. \quad (\text{A2})$$

Proof Since $C_\mu(h_i, h_k, \Theta) \geq 0$ and $\Theta_{g_{ik}} \in \Omega$, then $\Theta_{g_{ik}}$ is a constrained local minimizer of $C_\mu(h_i, h_k, \Theta)$. According to the Kuhn-Tucker condition (Kuhn & Tucker, 1951), there exist $\gamma \geq 0$, $\xi_j \geq 0$, for $j = 1, \dots, p+1$, and a Lagrangian $l_\mu(h_i, h_k, \Theta) = C_\mu(Y_i, Y_k, \Theta) + \gamma I(\Theta) + \sum_{j=1}^{p+1} \xi_j J_j(\Theta)$. A constrained local minimizer of $C_\mu(Y_i, Y_k, \Theta)$ must satisfy $\nabla l_\mu(h_i, h_k, \Theta) = 0$ and $\sum_{j=1}^{p+1} \xi_j J_j(\Theta) = 0$. From the Euler-Lagrange equation (Sagan, 1992), Assumption 4 and equation (3), $l_\mu(h_i, h_k, \Theta) = C_\mu(Y_i, Y_k, \Theta)$, and the necessary condition can be simplified to $\mu[h_i^{-1}\{\Theta^T A(t)\}] - \mu\{h_k^{-1}(t)\} = 0$, for $j = 1, \dots, p+1$. One then shows by contradiction that $\Theta_{g_{ik}}$ is the unique minimizer using the simplified necessary condition.

Lemma A3. Suppose $l_0 : \Omega \mapsto \mathcal{R}$ and $l_\varepsilon : \Omega \mapsto \mathcal{R}$ are continuous functions that satisfy $\sup_{\Theta \in \Omega} |l_\varepsilon(h_i, h_k, \Theta) - l_0(h_i, h_k, \Theta)| \leq \varepsilon$. In addition, assume that $l_0(h_i, h_k, \Theta)$ has an unique minimizer at Θ_0 and that for any $(p + 1)$ -dimensional ball $B_\varrho(\Theta_0)$ surrounding Θ_0 , there exists $c > 0$ such that

$$|l_0(h_i, h_k, \Theta) - l_0(h_i, h_k, \Theta_0)| > c\|\Theta - \Theta_0\|^2, \text{ for } \Theta \in B_\varrho(\Theta_0).$$

Then it holds for $\Theta_\varepsilon = \arg \min_{\Theta \in \Omega} l_\varepsilon(h_i, h_k, \Theta)$ that

$$\|\Theta_\varepsilon - \Theta_0\| = O(\varepsilon^{1/2}) \text{ as } \varepsilon \rightarrow 0.$$

A variant of this result can be found in Facer & Müller (2003) and the proof is omitted.

Lemma A4. Under Assumptions 1-4, for any $1 \leq i, k \leq n$, the minimizer $\tilde{\Theta}_{g_{ik}} = \arg \min_{\Theta \in \Omega} C_\lambda(Y_i, Y_k, \Theta)$, subject to the constraints in (A2), satisfies

$$\|\tilde{\Theta}_{g_{ik}} - \Theta_{g_{ik}}\| = O\{(\delta^2 + \lambda)^{1/2}\}, \quad (\text{A3})$$

and thus the estimated pairwise warping function \tilde{g}_{ik} satisfies

$$\sup_{t \in \mathcal{T}} |\tilde{g}_{ik}(t) - g_{ik}(t)| = \sup_{t \in \mathcal{T}} |\tilde{g}_{ik}(t) - h_i\{h_k^{-1}(t)\}| = O\{(\delta^2 + \lambda)^{1/2}\}.$$

Proof Since $\tilde{\Theta}_{g_{ik}}$ is a minimizer of a constrained optimization problem, for suitable $\kappa \geq 0$, $\zeta_j \geq 0$, $j = 1, \dots, p + 1$, the Lagrangian is

$$l_\lambda(h_i, h_k, \Theta) = C_\lambda(Y_i, Y_k, \Theta) + \kappa I(\Theta) + \sum_{j=1}^{p+1} \zeta_j J_j(\Theta), \quad (\text{A4})$$

where $I(\Theta)$ and $J(\Theta)$ are defined in (A2). The minimizer $\tilde{\Theta}_{g_{ik}}$ has to satisfy $\nabla l_\lambda(h_i, h_k, \tilde{\Theta}_{g_{ik}}) = 0$ and $\sum_{j=1}^{p+1} \zeta_j J_j(\tilde{\Theta}_{g_{ik}}) = 0$. From Lemma A2, one may conclude that $\sup_{\Theta \in \Omega} |l_\lambda(h_i, h_k, \Theta) - l_\mu(h_i, h_k, \Theta)| = O(\delta^2 + \lambda)$, because of $\sum_{j=1}^{p+1} \zeta_j J_j(\Theta) \leq 0$, Assumptions 1 and 2, and $\int_{\mathcal{T}} \{\Theta^T A(t) - t\}^2 dt \leq T^3/3$. From Lemma A3, $\|\tilde{\Theta}_{g_{ik}} - \Theta_{g_{ik}}\| = O\{(\delta^2 + \lambda)^{1/2}\}$. Finally, because $\sup_{t \in \mathcal{T}} |A_j(t)| \leq 1$, for $j = 1, \dots, p + 1$,

$$\sup_{t \in \mathcal{T}} |\tilde{g}_{ik}(t) - h_i\{h_k^{-1}(t)\}| = \sup_{t \in \mathcal{T}} |(\tilde{\Theta}_{g_{ik}} - \Theta_{g_{ik}})^T A(t)| \leq \|\tilde{\Theta}_{g_{ik}} - \Theta_{g_{ik}}\| = O\{(\delta^2 + \lambda)^{1/2}\}.$$

Lemma A5. Under Assumptions 1-6 and for bandwidth $b = b_{\text{opt}} \sim m^{-1/5}$ as $m \rightarrow \infty$, it holds that

$$\sup_{g \in \mathcal{W}} \left| E \left(\int_{\mathcal{T}} [\widehat{Y}_i^2\{g(t)\} - Y_i^2\{g(t)\}] dt \middle| Y_i \right) \right| = O_p(m^{-1/5}). \quad (\text{A5})$$

Proof Let $V = \int K^2(t)dt$, $B = \int \{K(t)t^2/2\}dt$, $V \neq 0$, $B \neq 0$, and assume that both V and B are bounded. Results about variance and bias for weighted local linear regression (Müller, 1987) lead to

$$\text{var}\{\widehat{Y}_i(t)|Y_i\} = (mb)^{-1}\{\sigma^2V + o(1)\} = O_p\{(mb)^{-1}\}. \quad (\text{A6})$$

Using properties of the smoothing weights $l_j(t)$, the bias term becomes

$$E\{\widehat{Y}_i(t) - Y_i(t)|Y_i\} \leq \left\{ \sum_{j=1}^m l_j^2(t) \right\}^{1/2} E \left[\sum_{j=1}^m \{Y_i(t_j) - Y_i(t)\}^2 1_{t \in [t_j - b, t_j + b]} | Y_i \right]^{1/2} = O_p(b),$$

since $Y_i(t)$ is uniformly Lipschitz continuous of order 1, and $E\{Y_i(t_j) - Y_i(t)|Y_i\}^2 \leq C'(t_j - t)^2$, where C' is a constant. Let $\mathcal{T}_j = [a_j, a_{j+1}]$, $j = 0, \dots, p$, and \mathcal{S}_j be the images of \mathcal{T}_j under the map $g \in \mathcal{W}$. For any $t \in \mathcal{T}_j$, there exists a unique $s \in \mathcal{S}_j$ such that $t = g^{-1}(s)$ and $dt = [g'\{g^{-1}(s)\}]^{-1}ds$. Using (A6), one finds that $\sup_{g \in \mathcal{W}} \int_{\mathcal{T}} \text{var}\{\widehat{Y}_i\{g(t)\}|Y_i\} dt = O_p\{1/(mb)\} = O_p(m^{-4/5})$, as $b = m^{-1/5}$ and $\sup_{g \in \mathcal{W}} \sum_{j=0}^p \int_{\mathcal{S}_j} [g'\{g^{-1}(s)\}]^{-1}ds$ is bounded according to Assumption 4. Similarly, $\sup_{g \in \mathcal{W}} \int_{\mathcal{T}} E^2\{\widehat{Y}_i\{g(t)\} - Y_i\{g(t)\}|Y_i\} dt = O_p(m^{-2/5})$. We then find that

$$\begin{aligned} \sup_{g \in \mathcal{W}} \int_{\mathcal{T}} (\text{var}\{\widehat{Y}_i\{g(t)\}|Y_i\} + E^2\{\widehat{Y}_i\{g(t)\} - Y_i\{g(t)\}|Y_i\}) dt &= O_p(m^{-2/5}), \\ \sup_{g \in \mathcal{W}} \int_{\mathcal{T}} (E[Y_i\{g(t)\}|Y_i]^2)^{1/2} (E[\widehat{Y}_i\{g(t)\} - Y_i\{g(t)\}|Y_i]^2)^{1/2} dt &= O_p(m^{-1/5}). \end{aligned}$$

The result follows by combining the last two bounds, and observing that the left-hand side of (A5) is bounded by

$$\begin{aligned} &2 \sup_{g \in \mathcal{W}} \int_{\mathcal{T}} (E[Y_i\{g(t)\}|Y_i]^2)^{1/2} (E[\widehat{Y}_i\{g(t)\} - Y_i\{g(t)\}|Y_i]^2)^{1/2} dt \\ &+ \sup_{g \in \mathcal{W}} \int_{\mathcal{T}} \text{var}\{\widehat{Y}_i\{g(t)\}|Y_i\} + (E[\widehat{Y}_i\{g(t)\} - Y_i\{g(t)\}|Y_i])^2 dt. \end{aligned}$$

Lemma A6. Under Assumptions 2-6, for bandwidths $b = b_{\text{opt}} \sim m^{-1/5}$, it holds that

$$\sup_{g \in \mathcal{W}} \left| E \left[\int_{\mathcal{T}} \widehat{Y}_i\{g(t)\} \widehat{Y}_k(t) - Y_i\{g(t)\} Y_k(t) dt \middle| Y_i \right] \right| = O_p(m^{-1/5}).$$

The proof is similar to that of Lemma A5 and is omitted.

Proof of Theorem 1 Observe that $\widehat{\Theta}_{g_{ik}}$, the constrained minimizer of $\widehat{C}_\lambda(Y_i, Y_k, \Theta)$, also minimizes $\widehat{l}_\lambda(Y_i, Y_k, \Theta) = \widehat{C}_\lambda(Y_i, Y_k, \Theta) + \kappa I(\Theta) + \sum_{j=1}^{p+1} \zeta_j J_j(\Theta)$, and that, according to the proof of Lemma A4, $\widetilde{\Theta}_{g_{ik}}$ minimizes $l_\lambda(h_i, h_k, \Theta)$, defined in (A4). According to Lemmas A5 and A6, $\sup_{\Theta \in \Omega} \left| \widehat{l}_\lambda(h_i, h_k, \Theta) - l_\lambda(h_i, h_k, \Theta) \right| = O_p(m^{-1/5})$, which together with Lemma A3 implies that $\|\widehat{\Theta}_{g_{ik}} - \widetilde{\Theta}_{g_{ik}}\| = O_p(m^{-2/5})$. Then with (A3) one obtains, for the left-hand side of (11),

$$\|\widehat{\Theta}_{g_{ik}} - \Theta_{g_{ik}}\| \leq \|\widehat{\Theta}_{g_{ik}} - \widetilde{\Theta}_{g_{ik}}\| + \|\widetilde{\Theta}_{g_{ik}} - \Theta_{g_{ik}}\| = O_p(m^{-2/5}) + O\{(\delta^2 + \lambda)^{1/2}\}.$$

The result of Theorem 1 follows because $\sup_{t \in \mathcal{T}} |A_j(t)| \leq 1$, for $j = 1, \dots, p+1$, and $\sup_{t \in \mathcal{T}} |\widehat{g}(t) - g(t)| = \sup_{t \in \mathcal{T}} |(\widehat{\Theta}_{g_{ik}}^\top - \Theta_{g_{ik}}^\top)A(t)|$.

Proof of Corollary 1 Using (7), (12), (13) and the generalized triangle inequality, we obtain

$$\sup_{t \in \mathcal{T}} |\widehat{h}_k^{-1}(t) - h_k^{-1}(t)| \leq \sup_{t \in \mathcal{T}} \sum_{i=1}^n \frac{|\widehat{g}_{ik}(t) - g_{ik}(t)|}{n} = O_p(m^{-1/5}) + O\{(\delta^2 + \lambda)^{1/2}\}.$$

Since $\widehat{h}_k^{-1}(t)$ and $\widehat{h}_k(t)$, $h_k^{-1}(t)$ and $h_k(t)$ are symmetric around the identity function, (14) follows and this concludes the proof.

References

- EILERS, P. & MARX, B. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.* **11**, 89–121.
- FACER, M. & MÜLLER, H. G. (2003). Nonparametric estimation of the peak location in a response surface. *J. Mult. Anal.* **87**, 191–217.
- FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modeling and its Applications*. London: CRC Press.
- GASSER, T. & KNEIP, A. (1992). Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20**, 1266–350.

- GASSER, T. & KNEIP, A. (1995). Searching for structure in curve samples. *J. Am. Statist. Assoc.* **90**, 1179–88.
- GERVINI, D. & GASSER, T. (2004). Self-modelling warping functions. *J. R. Statist. Soc. B* **66**, 959–71.
- GERVINI, D. & GASSER, T. (2005). Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika* **92**, 801–20.
- JAMES, G. (2007). Curve alignment by moments. *Ann. Appl. Statist.* **1**, 480–501.
- KAMINSKI, N. & BAR-JOSEPH, Z. (2007). A patient-gene model for temporal expression profiles in clinical studies. *J. Comput. Biol.* **14**, 324–38.
- KUHN, H. & TUCKER, A. (1951). Nonlinear programming. In *Proc 2nd Berkeley Symp. Math. Statist. Prob.* Ed. J Neyman, pp. 481–92. Berkeley, CA: University of California Press.
- LENG, X. & MÜLLER, H. (2006). Time ordering of gene co-expression. *Biostatistics* **7**, 569–84.
- LIANG, Y., TAYO, B., CAI, X. & KELEMEN, A. (2005). Differential and trajectory methods for time course gene expression data. *Bioinformatics* **21**, 3009–16.
- LIU, X. & MÜLLER, H. (2004). Functional convex averaging and synchronization for time-warped random curves. *J. Am. Statist. Assoc.* **99**, 687–99.
- MACK, Y. & SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Prob. Theory Rel. Fields* **61**, 405–15.
- MÜLLER, H. G. & STADTMÜLLER, U. (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15**, 182–201.
- RAMSAY, J. & SILVERMAN, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. New York: Springer-Verlag.
- RAMSAY, J. & SILVERMAN, B. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag.

- RAMSAY, J. O. & LI, X. (1998). Curve registration. *J. R. Statist. Soc. B* **60**, 351–63.
- RØNN, B. (2001). Nonparametric maximum likelihood estimation for shifted curves. *J. R. Statist. Soc. B* **63**, 243–59.
- SAGAN, H. (1992). *Introduction to the Calculus of Variations*. Mineola, New York: Dover Publications.
- SAKOE, H. & CHIBA, C. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. on Acoust., Speech, Sig. Proces.* **26**, 43–9.
- TUDDENHAM, R. & SNYDER, M. (1954). Physical growth of California boys and girls from birth to age 18. *Calif. Publ. Child Deve.* **1**, 183–364.
- WANG, K. & GASSER, T. (1997). Alignment of curves by dynamic time warping. *Ann. Statist.* **25**, 1251–76.
- WANG, K. & GASSER, T. (1999). Synchronizing sample curves nonparametrically. *Ann. Statist.* **27**, 439–60.
- WEINSTOCK-GUTTMAN, B., BADGETT, D., PATRICK, K., HARTRICH, L., SANTOS, R., HALL, D., BAIER, M., FEICHTER, J. & RAMANATHAN, M. (2002). Genomic effects of INF- beta in multiple sclerosis patients. *J. Immunol.* **171**, 1503–8.
- ZAMVIL, S. & STEINMAN, L. (2003). Diverse targets for intervention during inflammatory and neurodegenerative phases of multiple sclerosis. *Neuron* **38**, 685–8.

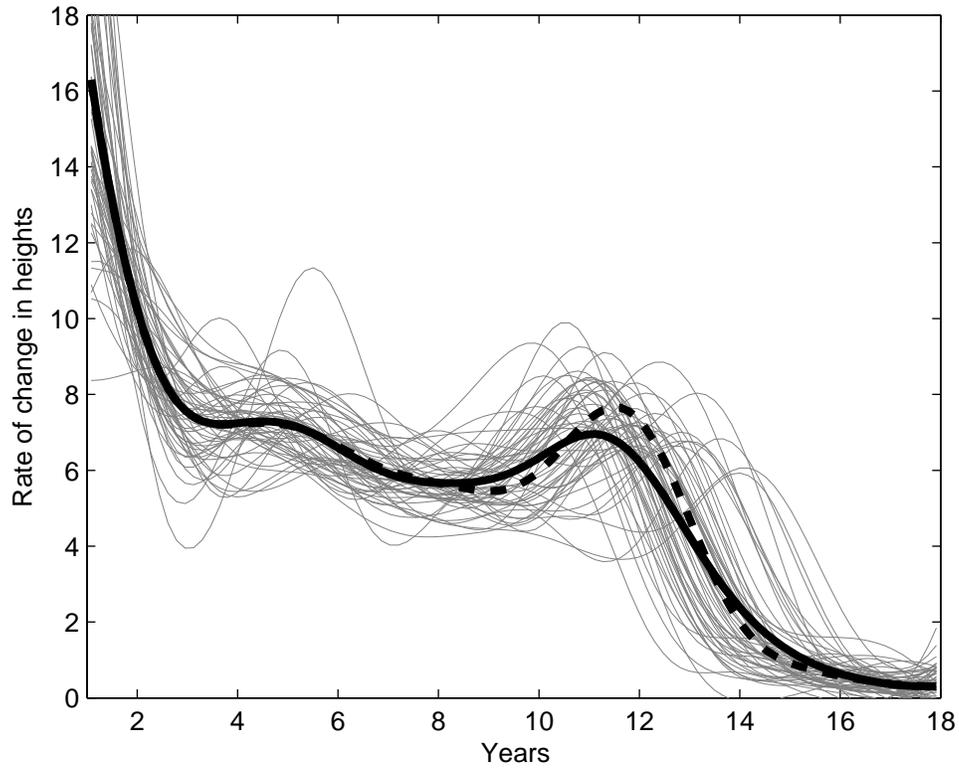


Figure 1: Female growth velocity curves of 54 girls in the Berkeley growth study (grey solid lines), the cross-sectional mean (solid line), and the landmark registration mean (dashed line).

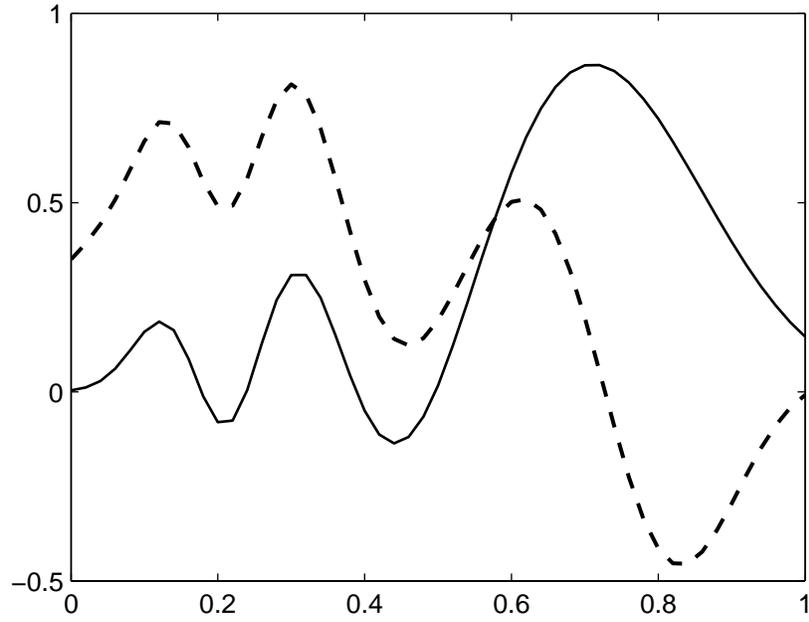


Figure 2: Mean shape functions $\mu(t)$ used in two simulations: $\mu_1(t)$ from equation (17) (solid) and $\mu_2(t)$ from equation (18) (dashed line).

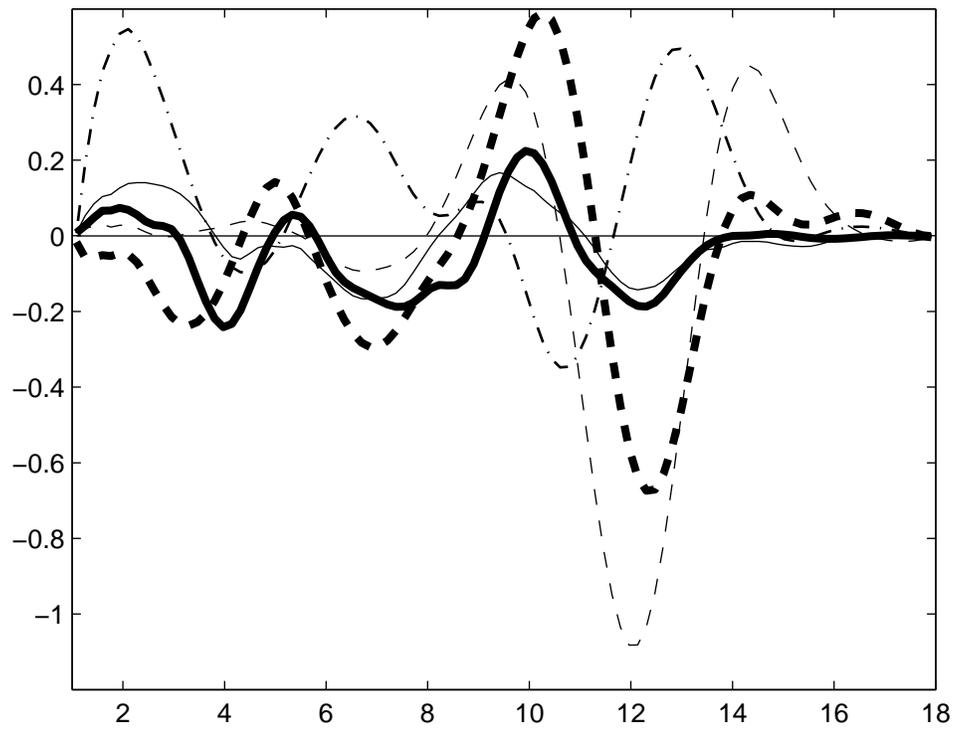


Figure 3: Deviations between warped mean functions and the landmark mean growth function for pairwise synchronization (solid line), self-modelling registration (bold solid line), nonparametric maximum likelihood estimation (dash-dot line), Procrustes method (bold dashed line); and cross-sectional mean with no warping (dashed line).

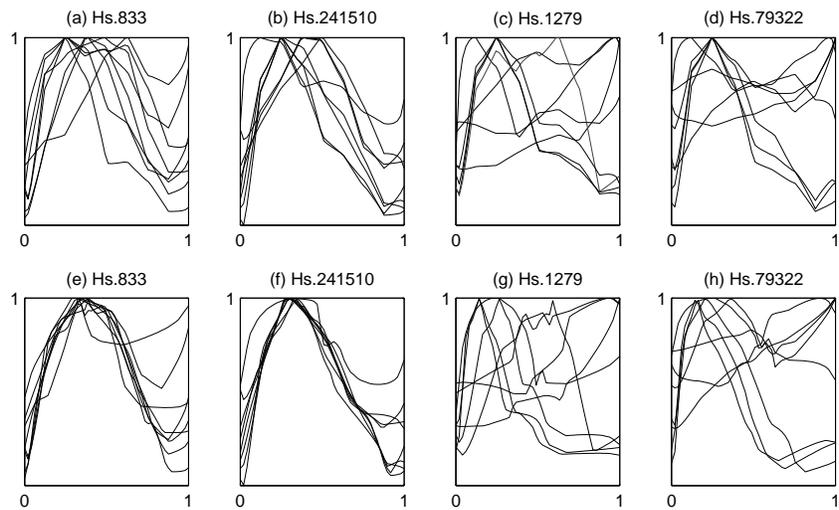


Figure 4: (a)-(d), smoothed and normalized expression profiles for 4 genes, measured on eight multiple sclerosis patients at nine time points; (e)-(h), smoothed, normalized and time-warped, using pairwise synchronization, expression profiles of these four genes, for the same patients.

Table 1: *Simulation results in terms of synchronized function, FMISE, and warping function, HMISE, mean integrated squared errors for pairwise synchronization under various combinations of auxiliary parameters p and λ , using mean shape function μ_1 in (17), for sample size $n = 20$ and 100 Monte Carlo runs.*

p	λ	10^{-1}		10^{-2}		10^{-3}		10^{-4}	
		\bar{x}	s^2	\bar{x}	s^2	\bar{x}	s^2	\bar{x}	s^2
3	FMISE	1.52	0.18	1.54	0.18	1.55	0.18	1.55	0.18
	HMISE	0.52	0.45	0.53	0.45	0.53	0.46	0.53	0.46
4	FMISE	1.54	0.19	1.56	0.18	1.57	0.18	1.56	0.18
	HMISE	0.52	0.45	0.54	0.47	0.54	0.47	0.54	0.46
5	FMISE	1.57	0.19	1.60	0.18	1.60	0.18	1.60	0.18
	HMISE	0.53	0.47	0.55	0.49	0.56	0.48	0.56	0.49

Table 2: Comparison of simulation results for different time warping methods: pairwise warping, PW, Procrustes, PRO, self-modelling registration, SR, and nonparametric maximum likelihood, NML. Shown are the mean and variance of the two deviation measures FMISE and HMISE, for mean shape functions μ_1 in (17) and μ_2 in (18), sample size $n = 20$ and 100 Monte Carlo runs.

		PW		PRO		SR		NML	
		\bar{x}	s^2	\bar{x}	s^2	\bar{x}	s^2	\bar{x}	s^2
μ_1	FMISE	1.55	0.20	1.93	0.24	1.98	0.24	2.50	0.31
μ_1	HMISE	0.16	0.004	0.21	0.03	0.23	0.008	0.47	0.014
μ_2	FMISE	3.40	0.13	3.49	0.17	3.62	0.26	3.67	0.34
μ_2	HMISE	0.19	0.01	0.18	0.01	0.23	0.01	0.39	0.02

Table 3: Comparison of between-patient variation before, VAR_b , and after, VAR_a , pairwise synchronization for 24 INF-stimulated genes, and the percentage of reduction in variation.

Gene	VAR_b	VAR_a	Ratio ¹	Gene	VAR_b	VAR_a	Ratio ¹
Hs.833	1.13	0.56	0.51	Hs.278613	1.25	0.67	0.47
Hs.183487	0.64	0.42	0.34	Hs.86958	0.49	0.33	0.33
Hs.82030	0.63	0.25	0.60	Hs.76391	0.77	0.50	0.34
Hs.241510	0.75	0.32	0.58	Hs.80645	0.22	0.17	0.22
Hs.146360	0.34	0.09	0.75	Hs.174195	0.37	0.26	0.31
Hs.1279	2.57	3.40	-0.32	Hs.171862	1.11	0.57	0.49
Hs.21486	0.87	0.13	0.86	Hs.155530	0.79	0.69	0.13
Hs.62661	1.07	0.14	0.87	Hs.77367	0.99	0.74	0.26
Hs.20315	0.62	0.09	0.85	Hs.83795	1.26	1.15	0.08
Hs.274382	0.47	0.04	0.92	Hs.277477	0.92	0.68	0.26
Hs.926	0.80	0.27	0.66	Hs.14623	1.46	1.23	0.16
Hs.79322	1.48	2.07	-0.39	Hs.75415	1.23	1.13	0.08

1. Ratio = $(\text{VAR}_b - \text{VAR}_a)/\text{VAR}_b$.