Fu [illegible] D A [illegible] S
L [illegible] D

F. [illegible] AO, H. -G [illegible] M LLER, . J. -L [illegible] ANG

We propose a nonparametric method to perform functional principal components analysis for the case of sparse longitudinal data. The method aims at irregularly spaced longitudinal data, where the number of repeated measurements available per subject is small. In contrast, classical functional data analysis requires a large number of regularly spaced measurements per subject. We assume that the repeated measurements are located randomly with a random number of repetitions for each subject and are determined by an underlying smooth random (subject-specific) trajectory plus measurement errors. Basic elements of our approach are the parsimonious estimation of the covariance structure and mean function of the trajectories, and the estimation of the variance of the measurement errors. The eigenfunction basis is estimated from the data, and functional principal components score estimates are obtained by a conditioning step. This conditional estimation method is conceptually simple and straightforward to implement. A key step is the derivation of asymptotic consistency and distribution results under mild conditions, using tools from functional analysis. Functional data analysis for sparse longitudinal data enables prediction of individual smooth trajectories even if only one or few measurements are available for a subject. Asymptotic pointwise and simultaneous confidence bands are obtained for predicted individual trajectories, based on asymptotic distributions, for simultaneous bands under the assumption of a finite number of components. Model selection techniques, such as the Akaike information criterion, are used to choose the model dimension corresponding to the number of eigenfunctions in the model. The methods are illustrated with a simulation study, longitudinal CD4 data for a sample of AIDS patients, and time-course gene expression data for the yeast cell cycle.

KEY WORDS: Asymptotics; Conditioning; Confidence band; Measurement error; Principal components; Simultaneous inference; Smoothing.

1. IN ROD C ION

We develop a version of functional principal components (FPC) analysis, in which the FPC scores are framed as conditional expectations. We demonstrate that this extends the applicability of FPC analysis to situations in longitudinal data analysis, where only few repeated and sufficiently irregularly spaced measurements are available per subject, and refer to this approach as principal components analysis through conditional expectation (PACE) for longitudinal data.

When the observed data are in the form of random curves rather than scalars or vectors, dimension reduction is mandatory, and FPC analysis has become a common tool to achieve this, by reducing random trajectories to a set of FPC scores. However, this method encounters difficulties when applied to longitudinal data with only few repeated observations per subject.

Beyond dimension reduction, FPC analysis attempts to characterize the dominant modes of variation of a sample of random trajectories around an overall mean trend function. There exists an extensive literature on FPC analysis when individuals are measured at a dense grid of regularly spaced time points. The method was introduced by Rao (1958) for growth curves, and the basic principle has been studied by Besse and Ramsay (1986), Castro, Lawton, and Sylvestre (1986), and Berkey, Laird, Valadian, and Gardner (1991). Rice and Silverman (1991) discussed smoothing and smoothing parameter choice in this context, whereas Jones and Rice (1992) emphasized applications. Various theoretical properties have been studied by Silverman (1996), Boente and Fraiman (2000), and Kneip and Utikal (2001). (For an introduction and summary, see Ramsay and Silverman 1997.) Staniswalis and Lee (1998) proposed kernel-based functional principal components analysis for repeated measurements with an irregular grid of time points. The case of irregular grids was also studied by Besse, Cardot, and Ferraty (1997) and Boularan, Ferré, and Vieu (1993). However, when the time points vary widely across subjects and are sparse, down to one or two measurements, the FPC scores defined through the Karhunen–Loève expansion are not well approximated by the usual integration method.

Shi, Weiss, and Taylor (1996), Rice and Wu (2000), James, Hastie, and Sugar (2001), and James and Sugar (2003) proposed B-splines to model the individual curves with random coefficients through mixed effects models. James et al. (2001) and James and Sugar (2003) emphasized the case of sparse data, postulating a reduced-rank mixed-effects model through a B-spline basis for the underlying random trajectories. In contrast, we represent the trajectories directly through the Karhunen–Loève expansion, determining the eigenfunctions from the data. Perhaps owing to the complexity of their modeling approach, James et al. (2001) did not investigate the asymptotic properties of the estimated components in relation to the true components, such as the behavior of the estimated covariance structure, eigenvalues, and eigenfunctions, especially for the sparse situation. Instead, they constructed pointwise confidence intervals for the individual curves using bootstrap. With our simpler and more direct approach, we are able to derive asymptotic properties, using tools from functional analysis. We can also derive both pointwise and simultaneous bands for predicted individual trajectories. This requires first obtaining the uniform convergence results for nonparametric function and surface estimates under dependence structure that follows from the longitudinal nature of the data. The dependence is a consequence of the assumed random nature of the observed sample of trajectories, which sets our work apart from previous results

Fang Yao is Assistant Professor, Department of Statistics, Colorado State University, Fort Collins, CO 80523 (E-mail:). Hans-Georg Müller is Professor (E-mail:) and Jane-Ling Wang is Professor (E-mail:), Department of Statistics, University of California, Davis, CA 95616. This research was supported in part by National Science Foundation grants DMS-98-03637, DMS-99-71602, DMS-02-04869, DMS-03-54448, and DMS-04-06430. The authors thank the associate editor and two referees for insightful comments on a previous version of this article that led to many improvements.

2005 A [illegible] S [illegible] A [illegible]
J-u A [illegible] S [illegible] A [illegible]
[illegible] 2005, V . 100, N . 470, T [illegible]
DOI 10.1198/016214504000001745

where either the observed functions are nonrandom with independent measurements (Kneip 1994), are random vectors of large but fixed dimensions (Ferré 1995), or are random trajectories sampled on dense and regular grids (Cardot, Ferraty, and Sarda 1999).

The contributions of this article are as follows. First, we provide a new technique, PACE, for longitudinal and functional data, a method designed to handle sparse and irregular longitudinal data for which the pooled time points are sufficiently dense. Second, the presence of additional measurement errors is taken into account, extending previous approaches of Staniswalis and Lee (1998) and Yao et al. (2003). Third, an emphasis is on the derivation of asymptotic consistency properties, by first establishing uniform convergence for smoothed estimates of the mean and covariance functions under mild assumptions. These uniform consistency results are developed for smoothers in the situation where repeated, and thus dependent, measurements are obtained for the same subject. Then we couple these results with the theory of eigenfunctions and eigenvalues of compact linear operators, to obtain uniform convergence of estimated eigenfunctions and eigenvalues. To our knowledge, there exist only few published asymptotic results for FPC (Dauxois, Pousse, and Romain 1982; Bosq 1991; Silverman 1996), and none for functional data analysis in the sparse situation. Fourth, we derive the asymptotic distribution needed to obtain pointwise confidence intervals for individual trajectories, and obtain asymptotic simultaneous bands for these trajectories.

The main novelty of our work is that we establish the conditional method for the case of sparse and irregular data, show that this provides a straightforward and simple tool for the modeling of longitudinal data, and derive asymptotic results for this method. Under Gaussian assumptions, the proposed estimation of individual FPC scores in PACE corresponds to the best prediction, combining the data from the individual subject to be predicted with data from the entire collection of subjects. In the non-Gaussian case, it provides an estimate for the best linear prediction. The proposed PACE method extends to the case of sparse and irregular data, provided that as the number of subjects increases, the pooled time points from the entire sample become dense in the domain of the data. We suggest one-curve-leave-out cross-validation for choosing auxiliary parameters, such as the degree of smoothing and the model dimension, corresponding to the number of eigenfunctions to be included, similar to the approach of to Rice and Silverman (1991). For faster computing, we also consider the Akaike information criterion (AIC) to select the number of eigenfunctions.

The remainder of the article is organized as follows. In Section 2 we introduce the PACE approach, that is, the proposed conditional estimates for the FPC scores. We present asymptotic results for the proposed method in Section 3, with proofs deferred to the Appendix. We discuss simulation results that illustrate the usefulness of the methodology in Section 4. Applications of PACE to longitudinal CD4 data and time-course gene expression data for yeast cell cycle genes are the theme of Section 5, followed by concluding remarks in Section 6 and proofs and theoretical results in the Appendix.

2. F NC IONAL PRINCIPAL COMPONEN ANAL I FOR PAR E DA A

2.1 M M E

We model sparse functional data as noisy sampled points from a collection of trajectories that are assumed to be independent realizations of a smooth random function, with unknown mean function $( ) = \mu( )$ and covariance function $\mathrm{cov}(( ), ( )) = (,)$. The domain of $(\cdot)$ typically is a bounded and closed time interval . Although we refer to the index variable as time, it could also be a spatial variable, such as in image or geoscience applications. We assume that there is an orthogonal expansion (in the ^{2} sense) of in terms of eigenfunctions φ and nonincreasing eigenvalues λ : $(,) = \sum \lambda \varphi ( )\varphi ( ), , \in$. In classical FPC analysis, it is assumed that the th random curve can be expressed as $( ) = \mu( ) + \sum \xi \varphi ( ), \in$, where the ξ are uncorrelated random variables with mean 0 and variance $\xi^2 = \lambda$, where $\sum \lambda < \infty, \lambda_1 \geq \lambda_2 \geq \cdots$.

We consider an extended version of the model that incorporates uncorrelated measurement errors with mean 0 and constant variance σ^2 to reflect additive measurement errors (see also Rice and Wu 2000). Let be the th observation of the random function $(\cdot)$, made at a random time , and let be the additional measurement errors that are assumed to be iid and independent of the random coefficients ξ , where $= 1, \ldots,$, $= 1, \ldots,$, $= 1, 2, \ldots$. Then the model that we consider is

$$= ( ) + = \mu( ) + \sum_{=1}^{\infty} \xi \varphi ( ) + , \quad \in , \qquad (1)$$

where $= 0$, $\mathrm{var}( ) = \sigma^2$, and the number of measurements made on the th subject is considered random, reflecting sparse and irregular designs. The random variables are assumed to be iid and independent of all other random variables.

2.2 E M C

We assume that mean, covariance, and eigenfunctions are smooth. We use local linear smoothers (Fan and Gijbels 1996) for function and surface estimation, fitting local lines in one dimension and local planes in two dimensions by weighted least squares. In a first step, we estimate the mean function μ based on the pooled data from all individuals. The formula for this local linear smoother is in (A.1) in the Appendix. Data-adaptive methods for bandwidth choice are available (see Müller and Prewitt 1993 for surface smoothing and Rice and Silverman 1991 for one-curve-leave-out cross-validation); subjective choices are often adequate. (For issues of smoothing dependent data, see Lin and Carroll 2000.) Adapting to estimated correlations when estimating the mean function did not lead to improvements (simulations not reported); therefore, we do not incorporate such adjustments.

Note that in model (1), $\mathrm{cov}(, | ,) = \mathrm{cov}(( ), ( )) + \sigma^2\delta$, where δ is 1 if $=$ and 0 otherwise. Let $(, ,) = (- \hat{\mu}( ))(- \hat{\mu}( ))$ be the "raw" covariances, where $\hat{\mu}( )$ is the estimated mean function obtained from the previous step. It is easy to see that $[(, ,)| ,] \approx \mathrm{cov}(( ), ( )) + \sigma^2\delta$. Therefore, the diagonal of the raw

covariances should be removed; that is, only $(\ ,\ ),\ \neq\ ,$ should be included as input data for the covariance surface smoothing step (as previously observed in Staniswalis and Lee 1998). We use one-curve-leave-out cross-validation to choose the smoothing parameter for this surface smoothing step.

The variance σ^2 of the measurement errors is of interest in model (1). Let $\widehat{\ }(\ ,\ )$ be a smooth surface estimate [see (A.2) in the App.] of $(\ ,\ ) = \mathrm{cov}(\ (\ ),\ (\ ))$. Following Yao et al. (2003), because the covariance of $(\ )$ is maximal along the diagonal, a local quadratic rather than a local linear fit is expected to better approximate the shape of the surface in the direction orthogonal to the diagonal. We thus fit a local quadratic component along the direction perpendicular to the diagonal and a local linear component in the direction of the diagonal; implementation of this local smoother is achieved by rotating the coordinates by 45 degrees and then minimizing weighted least squares [similar to (A.2)] in rotated coordinates with local quadratic and linear components, see (A.3) in the Appendix.

Denote the diagonal of the resulting surface estimate by $\widetilde{\ }(\ )$ and a local linear smoother focusing on diagonal values $\{\ (\ ,\ ) + \sigma^2\}$ by $\widehat{\ }(\ )$, obtained by (A.1) with $\{\ (\ ,\ )\}$ as input. To mitigate boundary effects, we cut off the two ends of the interval to get a more stable estimate, following a suggestion of Staniswalis and Lee (1998). Let $|\ |$ denote the length of $\ $, and let $_1$ be the interval $_1 = [\inf\{\ :\ \in\ \} + |\ |/4, \sup\{\ :\ \in\ \} - |\ |/4]$. The proposed estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{2}{|\ |}\int_{\mathcal{T}_1}\{\widehat{\ }(\ ) - \widetilde{\ }(\ )\} \tag{2}$$

if $\hat{\sigma}^2 > 0$ and $\hat{\sigma}^2 = 0$ otherwise.

The estimates of eigenfunctions and eigenvalues correspond to the solutions $\hat{\varphi}$ and $\hat{\lambda}$ of the eigenequations,

$$\int_{\mathcal{T}}\widehat{\ }(\ ,\ )\hat{\varphi}\ (\ )\ \ = \hat{\lambda}\ \hat{\varphi}\ (\ ), \tag{3}$$

where the $\hat{\varphi}$ are subject to $\int_{\mathcal{T}}\hat{\varphi}\ (\ )^2\ \ = 1$ and $\int_{\mathcal{T}}\hat{\varphi}\ (\ )\times \hat{\varphi}\ (\ )\ \ = 0$ for $\ <\ $. We estimate the eigenfunctions by discretizing the smoothed covariance, as previously described by Rice and Silverman (1991) and Capra and Müller (1997).

2.3 F . . . P . . C η . . A . . C . . . E . .

The FPC scores $\xi\ = \int(\ (\ ) - \mu(\ ))\varphi\ (\ )$ have traditionally been estimated by numerical integration, which works well when the density of the grid of measurements for each subject is sufficiently large. Because in our model the are available only at discrete random times , reflecting the sparseness of the data, the integrals in the definition of the FPC scores ξ accordingly would be approximated by sums, substituting as defined in (1) for $(\ )$ and estimates $\hat{\mu}(\ )$ for $\mu(\ )$ and $\hat{\varphi}\ (\ )$ for $\varphi\ (\ )$, leading to $\hat{\xi}\ = \sum_{=1}(\ - \hat{\mu}(\ ))\hat{\varphi}\ (\ )(\ - \ _{,-1})$, setting $_0 = 0$. For sparse functional data, $\hat{\xi}$ will not provide reasonable approximations to ξ, for example, when one has only two observations per subject. Moreover, when the measurements are contaminated with errors, the underlying random process cannot be directly observed. Substituting for $(\ )$ then leads to biased FPC scores. These considerations motivate the alternative PACE method to obtain the FPC scores.

Assume that in (1), ξ and are jointly Gaussian. In all of what follows, the results pertaining to expectations are always conditional on the observation times , $= 1,\ldots,$, $= 1,\ldots,$. For simplicity, the dependence on is suppressed. Write $\widetilde{\mathbf{X}} = (\ (\ _1),\ldots,\ (\ ))$, $\widetilde{\mathbf{Y}} = (\ _1,\ldots,\ )$, $\ = (\mu(\ _1),\ldots,\mu(\ ))$, and $\ = (\varphi\ (\ _1),\ldots,\varphi\ (\ ))$. The best prediction of the FPC scores for the th subject, given the data from that individual, is the conditional expectation, which, under Gaussian assumptions [also given in (A5) in Sec. 3], is found to be (see, e.g., thm. 3.2.4 in Mardia, Kent, and Bibby 1979)

$$\tilde{\xi}\ \ =\ \ [\xi\ |\widetilde{\mathbf{Y}}\ ] = \lambda\ \ \ \mathbf{\Sigma}_{\mathbf{Y}}^{-1}(\widetilde{\mathbf{Y}}\ -\ \ ), \tag{4}$$

where $\mathbf{\Sigma}_{\mathbf{Y}}\ = \mathrm{cov}(\widetilde{\mathbf{Y}}\ ,\widetilde{\mathbf{Y}}\ ) = \mathrm{cov}(\widetilde{\mathbf{X}}\ ,\widetilde{\mathbf{X}}\ ) + \sigma^2\mathbf{I}$; that is, the $(\ ,\ )$ entry of the $\times$ matrix $\mathbf{\Sigma}_{\mathbf{Y}}$ is $(\mathbf{\Sigma}_{\mathbf{Y}}\ )$

the linear functions of $\tilde{\mathbf{Y}}$, we have $[\tilde{\ }, \tilde{\ }] = [\tilde{\ }, \tilde{\ }]$, that is, $\mathrm{var}(\tilde{\ }) - \mathrm{var}(\tilde{\ }) = \ ,$ where $\ = \mathbf{\Lambda} - \mathbf{H}\mathbf{\Sigma}_{\mathbf{Y}}^{-1}\mathbf{H}$ and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda\ \}$. Then, under Gaussian assumptions, $(\tilde{\ } - \ ) \sim (\mathbf{0}, \ )$.

We construct asymptotic pointwise confidence intervals for individual trajectories as follows. Let $\hat{\ } = \hat{\mathbf{\Lambda}} - \hat{\mathbf{H}}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}^{-1}\hat{\mathbf{H}}$, where $\hat{\mathbf{\Lambda}} = \mathrm{diag}\{\hat{\lambda}_1, \ldots, \hat{\lambda}\ \}$ and $\hat{\mathbf{H}} = (\hat{\lambda}_1\hat{\ }_1, \ldots, \hat{\lambda}\ \hat{\ })$. For $\in$, let $\ = (\varphi_1(\ ), \ldots, \varphi\ (\ ))$, $\hat{\ } = (\hat{\varphi}_1(\ ), \ldots, \hat{\varphi}\ (\ ))$, and $\hat{\ }(\ ) = \hat{\mu}(\ ) + \ \hat{\ }$. Theorem 4 establishes that the distribution of $\{\hat{\ }(\ ) - \ (\ )\}$ may be asymptotically approximated by $(0, \hat{\ }\ \hat{\ }\ \hat{\ })$. Because we assume that can be approximated sufficiently well by the first eigenfunctions, we may construct the $(1-\alpha)$ asymptotic pointwise interval for $(\ )$,

$$\hat{\ }(\ ) \pm \ ^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\ \hat{\ }\ \hat{\ }\ \hat{\ }}, \tag{7}$$

where is the standard Gaussian cdf. These confidence intervals are constructed by ignoring the bias that results from the truncation at in $\hat{\ }$.

Next, consider the construction of asymptotic simultaneous confidence bands. Let $\ (\ ) = \mu(\ ) + \sum_{=1} \xi\ \varphi\ (\ )$. Theorem 5 provides the asymptotic simultaneous band for $\{\hat{\ }(\ ) - \ (\ )\}$, for a given fixed . The Karhunen–Loève theorem implies that $\sup_{\in\mathcal{T}} [\ (\ ) - \ (\ )]^2$ is small for fixed and sufficiently large . Therefore, ignoring a remaining approximation error that may interpreted as a bias, we may construct $(1-\alpha)$ asymptotic simultaneous bands for $(\ )$ through

$$\hat{\ }(\ ) \pm \sqrt{\chi^2_{,1-\alpha}\ \hat{\ }\ \hat{\ }\ \hat{\ }}, \tag{8}$$

where $\chi^2_{,1-\alpha}$ is the $100(1-\alpha)$th percentile of the chi-squared distribution with degrees of freedom. Because $\sqrt{\chi^2_{,1-\alpha}} > \ ^{-1}(1-\alpha/2)$ for all ≥ 1, the asymptotic simultaneous band is always wider than the corresponding asymptotic pointwise confidence intervals.

We obtain simultaneous intervals for all linear combinations of the FPC scores analogously. Given , let $\subseteq \Re$ be a linear space with dimension $\leq$. Then, asymptotically, it follows from the uniform result in Corollary 2 in Section 3 that for all linear combinations $\mathbf{l}$, , simultaneously, where $\mathbf{l} \in$,

$$\mathbf{l}\ , \ , \in \mathbf{l}\ \hat{\ }\ , \pm \sqrt{\chi^2_{,1-\alpha}\mathbf{l}\ \hat{\ }\ \mathbf{l}}, \tag{9}$$

with approximate probability $(1-\alpha)$.

2.5 N η E

To choose the number of eigenfunctions that provides a reasonable approximation to the infinite-dimensional process, we may use the cross-validation score based on the one-curve-leave-out prediction error (Rice and Silverman 1991). Let $\hat{\mu}^{(-)}$ and $\hat{\varphi}^{(-)}$ be the estimated mean and eigenfunctions after removing the data for the th subject. Then we choose so as to minimize the cross-validation score based on the squared prediction error,

$$\mathrm{CV}(\ ) = \sum_{=1}\sum_{=1}\{\ - \hat{\ }^{(-)}(\ )\}^2, \tag{10}$$

where $\hat{\ }^{(-)}$ is the predicted curve for the th subject, computed after removing the data for this subject, that is, $\hat{\ }^{(-)}(\ ) = \hat{\mu}^{(-)}(\ ) + \sum_{=1} \hat{\xi}^{(-)}\hat{\varphi}^{(-)}(\ )$, where $\hat{\xi}$ is obtained by (5).

One can also adapt AIC-type criteria (Shibata 1981) to this situation. In simulations not reported here, we found that AIC is computationally more efficient while the results are similar to those obtained by cross-validation. A pseudo-Gaussian log-likelihood, summing the contributions from all subjects, conditional on the estimated FPC scores $\hat{\xi}$ (5), is given by

$$\hat{\ } = \sum_{=1}\left\{-\frac{}{2}\log(2\pi) - \frac{}{2}\log\hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2}\left(\tilde{\mathbf{Y}} - \hat{\ } - \sum_{=1}\hat{\xi}\ \hat{\ }\right) \times \left(\tilde{\mathbf{Y}} - \hat{\ } - \sum_{=1}\hat{\xi}\ \hat{\ }\right)\right\}, \tag{11}$$

where we define $\mathrm{AIC} = -\hat{\ } + \ $.

3. A MP O IC PROPER IE

We derive consistency and distribution results demonstrating the consistency of the estimated FPC scores $\hat{\xi}$ in (5) for the true conditional expectations $\tilde{\xi}$ in (4). Uniform convergence of the local linear estimators of mean and covariance functions on bounded intervals plays a central role in obtaining these results and thus is established first (Thm. 1). Proofs are deferred to the Appendix.

The data $(\ , \ )$, $= 1, \ldots, \ , \ = 1, \ldots, \ $, coming from model (1), are assumed to have the same distribution as $(\ , \ )$, with joint density $(\ , \ )$. Assume that the observation times are iid with marginal density $(\ )$, but that dependence is allowed between observations and , coming from the same subject or cluster. The following assumptions pertain to the number of observations made on the th subject or cluster:

(A1.1) The number of observations made for the th subject or cluster is a random variable with $\overset{\text{iid}}{\sim}$, where > 0 is a positive discrete random variable, with $< \infty$ and $\{\ > 1\} > 0$.

The observation times and measurements are assumed to be independent of the number of measurements, that is, for any subset $\subseteq \{1, \ldots, \ \}$ and for all $= 1, \ldots, \ ,$

(A1.2) $(\{\ : \in \ \}, \{\ : \in \ \})$ is independent of .

Writing $\tilde{\mathbf{T}} = (\ _1, \ldots,$

definitions of these smoothers.] Kernel $\kappa_1(\cdot)$ is also used for obtaining the estimate $\widehat{\ }$ for $\{ (\ ,\ ) + \sigma^2\}$ with the local linear smoother. Let $_\mu$, , and be the bandwidths for estimating $\hat{\mu}$, $\widehat{\ }$, and $\widehat{\ }$. Assume that κ_1 and κ_2 are compactly supported densities with properties (B2.1a) and (B2.2a) and (B2.1b) and (B2.2b). We develop asymptotics as the number of subjects $\to \infty$, and require the following:

(A2.1) $_\mu \to 0$, $^4_\mu \to \infty$, and $^6_\mu < \infty$.
(A2.2) $\to 0$, $^6 \to \infty$, and $^8 < \infty$.
(A2.3) $\to 0$, $^4 \to \infty$, and $^6 < \infty$.

Define the Fourier transforms of $\kappa_1(\ )$ and $\kappa_2(\ ,_\prime)$ by $\zeta_1(\ ) = \int^{\ -\ } \kappa_1(\ )$ and $\zeta_2(\ ,\ ) = \int^{\ -(\ +_\prime\ )} \kappa_2(\ ,_\prime)\ _\prime$. They satisfy the following:

(A3.1) $\zeta_1(\ )$ is absolutely integrable, that is, $\int |\zeta_1(\ )| \times < \infty$.
(A3.2) $\zeta_2(\ ,\ )$ is absolutely integrable, that is, $\iint |\zeta_2(\ ,\ )| < \infty$.

Assume that the fourth moment of centered at $\mu(\ )$ is finite, that is,

(A4) $[(\ - \mu(\ ))^4] < \infty$.

Then we obtain uniform convergence rates for local linear estimators $\hat{\mu}(\ )$ of $\mu(\ )$ and $\widehat{\ }(\ ,\ )$ of $(\ ,\ )$ on compact sets and ^{2}.

Under (A1.1)–(A4) and (B1.1)–(B2.2b) with $\nu = 0$, $= 2$ in (B2.2a) and $\nu = (0,0)$, $= 2$ in (B2.2b),

$$\sup_{\in\mathcal{T}} |\hat{\mu}(\ ) - \mu(\ )| = \left(\frac{1}{\sqrt{\ _\mu}}\right) \tag{12}$$

and

$$\sup_{\ ,\ \in\mathcal{T}} |\widehat{\ }(\ ,\ ) - \ (\ ,\ )| = \left(\frac{1}{\sqrt{\ ^2}}\right). \tag{13}$$

The consistency of $\hat{\sigma}^2$ (2) is obtained as a consequence.

Under (A1.1)–(A4) and (B1.1)–(B2.2b) with $\nu = 0$, $= 2$ in (B2.2a) and $\nu = (0,0)$, $= 2$ in (B2.2b),

$$|\hat{\sigma}^2 - \sigma^2| = \left(\frac{1}{\sqrt{\ }}\left(\frac{1}{\ ^2} + \frac{1}{\ }\right)\right). \tag{14}$$

We note that the rates of convergence provided in (12) and (13) are slower than the optimal ones known for the case of smoothing functions or surfaces from sufficiently densely spaced independent measurements. These rates would be of order $(\sqrt{\log\ /(\ _\mu)})$ for function estimates and $(\sqrt{\log\ /(\ ^2)})$ for surface estimates. It is an interesting question whether these rates remain optimal for the present dependent-data setting and whether they can be attained in the situation of dependent and sparse data that we are dealing with.

Next, consider the real separable Hilbert space $^2(\ ) \equiv$ endowed with inner product $\langle\ ,\ \rangle = \int_{\mathcal{T}} (\ )\ (\ )$ and norm $\|\ \| = \sqrt{\langle\ ,\ \rangle}$ (Courant and Hilbert 1953). Let $'$ denote the set of indices of the eigenfunctions φ corresponding to eigenvalues λ of multiplicity 1. We obtain the consistency of the $\hat{\lambda}$ in (3) for λ , the consistency of $\hat{\varphi}$ in (3) for φ in the ^{2} norm $\|\cdot\|$, by choosing $\hat{\varphi}$ appropriately when λ is of multiplicity 1, and furthermore the uniform consistency of $\hat{\varphi}$ for φ on the bounded interval .

Under (A1.1)–(A4) and (B1.1)–(B2.2b) with $\nu = 0$, $= 2$ in (B2.2a) and $\nu = (0,0)$, $= 2$ in (B2.2b),

$$|\hat{\lambda}\ - \lambda\ | = \left(\frac{1}{\sqrt{\ ^2}}\right); \tag{15}$$

$$\|\hat{\varphi}\ - \varphi\ \| = \left(\frac{1}{\sqrt{\ ^2}}\right), \qquad \in\ '; \tag{16}$$

and

$$\sup_{\in\mathcal{T}} |\hat{\varphi}\ (\ ) - \varphi\ (\ )| = \left(\frac{1}{\sqrt{\ ^2}}\right), \qquad \in\ '. \tag{17}$$

We remark that the rates (15)–(17) are direct consequences of the rates (12) and (13), as is evident from the proofs. If the rates in (12) and (13) are both $(\alpha\ )$, then the rates in (15)–(17) will also be $(\alpha\ )$.

For the following results, we require Gaussian assumptions:

(A5) The FPC scores ξ and measurement errors in (1) are jointly Gaussian.

We also assume that the data asymptotically follow a linear scheme:

(A6) The number, location, and values of measurements for a given subject or cluster remain unaltered as $\to \infty$.

The target trajectories that we aim to predict are

$$\tilde{\ }(\ ) = \mu(\ ) + \sum_{=1}^{\infty} \tilde{\xi}\ \ \varphi\ (\ ), \qquad = 1, \ldots, \ , \tag{18}$$

with $\tilde{\xi}$ as defined in (4). We note that $\tilde{\ }$ may be defined as a limit of random functions $\tilde{\ }(\ ) = \mu(\ ) + \sum_{=1} \tilde{\xi}\ \ \varphi\ (\ )$, as $\sup_{\in\mathcal{T}} [\tilde{\ }(\ ) - \tilde{\ }(\ )]^2 \to 0$ (see Lemma A.3 in the App.). For any ≥ 1, the target curve $\tilde{\ }(\ )$ is then estimated by

$$\widehat{\ }(\ ) = \hat{\mu}(\ ) + \sum_{=1} \hat{\xi}\ \ \hat{\varphi}\ (\ ), \tag{19}$$

with $\hat{\xi}$ as in (5).

Assume (A1.1)–(A6) and (B1.1)–(B2.2b) with $\nu = 0$, $= 2$ in (B2.2a) and $\nu = (0,0)$, $= 2$ in (B2.2b). Then

$$\lim_{\to\infty} \hat{\xi}\ = \tilde{\xi} \quad \text{in probability}, \tag{20}$$

and for all $\in$,

$$\lim_{\to\infty} \text{à})$$

( )

Applying Theorems 1 and 2, the estimate $\hat{\omega}(\ ,\ )$ is consistent for $\omega(\ ,\ )$ for all $\ ,\ \in\ $; that is, $\lim_{\to\infty}\lim_{\to\infty}\hat{\omega}(\ ,\ )=\omega(\ ,\ )$ in probability.

Assume (A1.1)–(A7) and (B1.1)–(B2.2b) with $\nu=0,\ =2$ in (B2.2a) and $\nu=(0,0),\ =2$ in (B2.2b). For all $\in$ and $\in\Re$,

$$\lim_{\to\infty}\lim_{\to\infty}\left\{\frac{\widehat{\ }(\ )-\ (\ )}{\sqrt{\hat{\omega}(\ ,\ )}}\le\right\}=\ (\ ), \tag{22}$$

where is the standard Gaussian cdf.

The number of random components and eigenfunctions that are needed in Theorems 3 and 4 to approximate the trajectory $\widetilde{\ }(\ )$ depends primarily on the complexity of the covariance structure $(\ ,\ )$ and on the number and location of the measurements observed for a given subject. It also depends on the sample size , through the eigenfunction and covariance estimates. Although data-based choices for are available through (10) and (11) and are successful in practical applications, results (21) and (22) indicate that for large , the number of components needs to be increased to obtain consistency, but these results do not provide further guidance as to how should be chosen in dependence on .

We next establish $(1-\alpha)$ asymptotic simultaneous inference for $\{\widehat{\ }(\ )-\ (\ )\}$ on the domain , where $(\ )=\mu(\ )+\sum_{=1}\xi\ \varphi(\ )$. For these results, we are providing not functional asymptotics, but instead finite-dimensional asymptotics, because the number of included components is considered fixed, whereas the sample size $\to\infty$ as before. If is chosen such that only trajectories truncated at the first components $(\cdot)$ of their expansion are of interest, then the following two results provide simultaneous confidence bands, as well as simultaneous confidence sets for the first random effects. Corollary 2 is a variation of Scheffé's method.

Under (A1.1)–(A7) and (B1.1)–(B2.2b) with $\nu=0,\ =2$ in (B2.2a) and $\nu=(0,0),\ =2$ in (B2.2b), for fixed number of components ,

$$\lim_{\to\infty}\left\{\sup_{\in\mathcal{T}}\frac{|\widehat{\ }(\ )-\ (\ )|}{\sqrt{\hat{\omega}(\ ,\ )}}\le\sqrt{\chi^2_{,1-\alpha}}\right\}\ge 1-\alpha, \tag{23}$$

where $\chi^2_{,1-\alpha}$ is the $(1-\alpha)$th percentile of the chi-squared distribution with degrees of freedom.

Assuming components, let $\subseteq\Re$ be a linear space with dimension $\le$. By arguments analogous to the proof of Theorem 5, we obtain the asymptotic simultaneous $(1-\alpha)$ confidence region for all linear combinations $\mathbf{l}_{,}\hat{\ }_{,}$, where $\mathbf{l}\in$.

Under the assumptions of Theorem 5,

$$\lim_{\to\infty}\left\{\sup_{\mathbf{l}\in\mathcal{A}}\frac{|\mathbf{l}\,(\hat{\ }_{,}-_{,}\ _{,})|}{\sqrt{\mathbf{l}_{,}\widehat{\ }_{,}\mathbf{l}}}\le\sqrt{\chi^2_{,1-\alpha}}\right\}\ge 1-\alpha, \tag{24}$$

where $\chi^2_{,1-\alpha}$ is the $(1-\alpha)$th percentile of the chi-squared distribution with degrees of freedom.

4. IM LA ION DIE

To illustrate the implementation of sparse FPC analysis by PACE, we construct 100 iid normal and 100 iid nonnormal samples each consisting of $=100$ random trajectories. The simulated processes have mean function $\mu(\ )=\ +\sin(\ )$ and covariance function derived from two eigenfunctions, $\varphi_1(\ )=-\cos(\pi\ /10)/\sqrt{5}$ and $\varphi_2(\ )=\sin(\pi\ /10)/\sqrt{5}$, $0\le\ \le 10$. We chose $\lambda_1=4$, $\lambda_2=1$, and $\lambda\ =0,\ \ge 3$, as eigenvalues and $\sigma^2=.25$ as the variance of the additional measurement errors in (1), which are assumed to be normal with mean 0. For the smoothing steps, we use univariate and bivariate Epanechnikov kernel functions, that is, $\kappa_1(\ )=3/4(1-\ ^2)\mathbb{1}_{[-1,1]}(\ )$ and $\kappa_2(\ ,\ )=9/16(1-\ ^2)(1-\ ^2)\mathbb{1}_{[-1,1]}(\ )\mathbb{1}_{[-1,1]}(\ )$, where $\mathbb{1}\ (\ )=1$ if $\in$ and 0 otherwise for any set . For an equally spaced grid $\{\ _0,\ldots,\ _{50}\}$ on $[0,10]$ with $\ _0=0$ and $\ _{50}=10$, let $=\ +\ $, where are iid with $(0,.1)$, $=0$ if <0, and $=10$ if >10, allowing for nonequidistant "jittered" designs. Each curve was sampled at a random number of points, chosen from a discrete uniform distribution on $\{1,\ldots,4\}$, and the locations of the measurements were randomly chosen from $\{\ _1,\ldots,\ _{49}\}$ without replacement. For the 100 normal samples, the FPC scores ξ were generated from $(0,\lambda\ )$, whereas the ξ for the nonnormal samples were generated from a mixture of two normals, $(\sqrt{\lambda\ /2},\lambda\ /2)$ with probability 1/2 and $(-\sqrt{\lambda\ /2},\lambda\ /2)$ with probability 1/2.

To demonstrate the superior performance of the conditional method, Table 1 reports the average mean squared error (MSE) for the true curves , $\mathrm{MSE}=\sum_{=1}\int_0^{10}\{\ (\ )-\widehat{\ }(\ )\}^2\ /\ $, where $\widehat{\ }(\ )=\hat{\mu}(\ )+\sum_{=1}\hat{\xi}\ \hat{\varphi}(\ )$ and the $\hat{\xi}$'s were obtained using either the proposed PACE method (5) or the integration method. The number of eigenfunctions in each run was chosen by the AIC (11). In each simulation consisting of 100 Monte Carlo runs (for a total of 400 runs, normal/mixture and sparse/nonsparse), there were always more than 95 runs in which two eigenfunctions were chosen.

Another outcome measure of interest is the average squared error (ASE) for the two FPC scores, $\mathrm{ASE}(\xi\ )=\sum_{=1}(\hat{\xi}\ -\xi\ )^2/\ $, $=1,2$, also listed in Table 1. We also compared the two methods for irregular but nonsparse simulated data, where the number of observations for each curve was randomly chosen from $\{30,\ldots,40\}$, with results given in Table 1. We find that the gains in the sparse situation are dramatic when switching from the traditional method to the PACE method. For the case of an underlying normal distribution, the MSE was reduced by 43% using the PACE method (5) as compared with

1. R FPC A. C. E. (CE, PACE). I. (IN) M 100 M. C R. N = 100 R. n n , G. R. n C n .

$N=100$ FPC		N n			M_i		
		M E	A E(ξ_1)	A E(ξ_2)	M E	A E(ξ_1)	A E(ξ_2)
	CE	1.33	.762	.453	1.30	.737	.453
	IN	2.32	1.58	.622	2.25	1.53	.631
N .	CE	.259	.127	.110	.256	.132	.105
	IN	.286	.159	.115	.286	.168	.114

NO E: , A E(ξ), $=1,2$, n n , M E, . 4. n n M. C AIC (11).

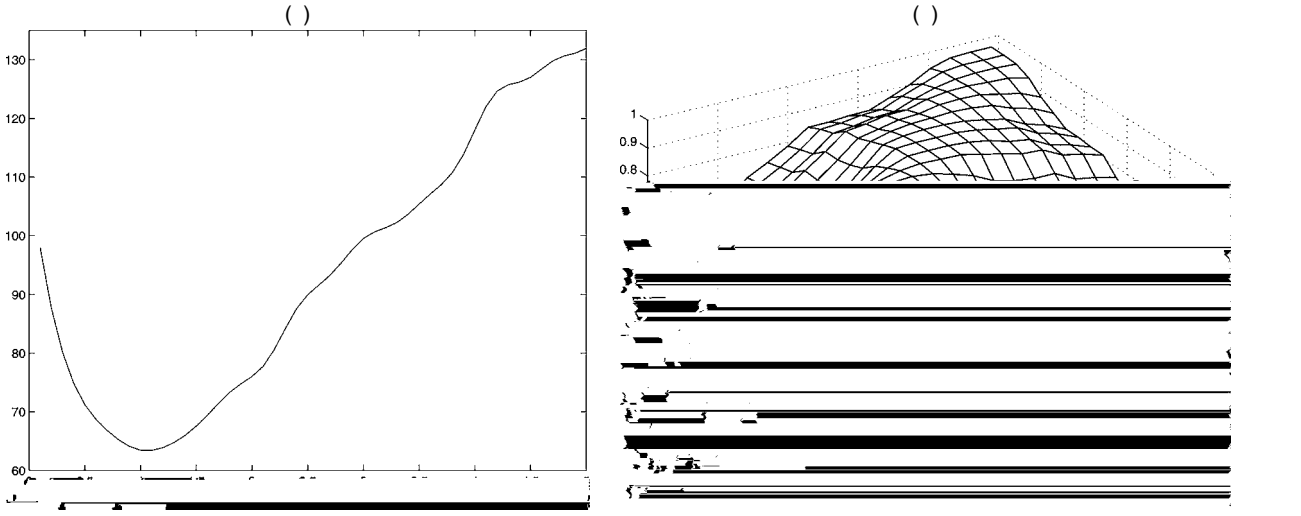
the traditional method; the ASE(ξ) were reduced by 52%/27% ($= 1, 2$). For the mixture distribution case, the decreases were still 42% for MSE and 52%/28% for ASE(ξ) ($= 1, 2$). In nonsparse situations, the traditional estimates provide reasonable approximations to the underlying integrals, but nevertheless PACE still produces better estimates, with improvements of 10%/10% for MSE and 20%/21%, 5%/8% for ASE(ξ), $= 1, 2$, for normal/nonnormal samples. We conclude that the gains obtainable using PACE are substantial for sparse data and also extend to the case of dense and non-Gaussian data.

5. APPLICA ION

5.1 L . . CD4 C .

Because CD4 counts constitute a critical assessment of the status of the immune system and are used as an important marker in describing the progress to AIDS in adults, CD4 cell counts and CD4 percentages (i.e., CD4 counts divided by the total number of lymphocytes) are commonly used markers for the health status of human immunodeficiency virus (HIV) infected persons. The dataset considered here is from the Multicenter AIDS Cohort Study, which includes repeated measurements of physical exams, laboratory results, and CD4 percentages for 283 homosexual men who became HIV-positive between 1984 and 1991. All individuals were scheduled to have their measurements made at semiannual visits. However, because many individuals missed scheduled visits and the HIV infections happened randomly during the study, the data are sparse, with unequal numbers of repeated measurements per subject and different measurement times, , per individual. The number of observations per subject ranged from 1 to 14, with a median of 6. The trajectories in their entirety are assembled in Figure 1(a).

That the data from such a classical longitudinal study, with measurements intended to be spaced at regular 6-month intervals, are quite well suited for analysis by PACE is illustrated by Figure 2. As this figure shows, the assembled pairs (,) measurements54.8(t)-2820-256densed-)096002 91 T4 5cau9592 cm /Im1 Do Q

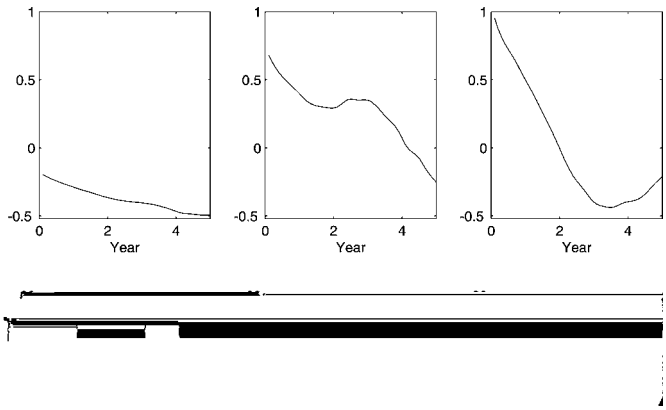


F 3. ( ) E F . CD4 C . D . ( ) E C . F . , E , R D F . D . .

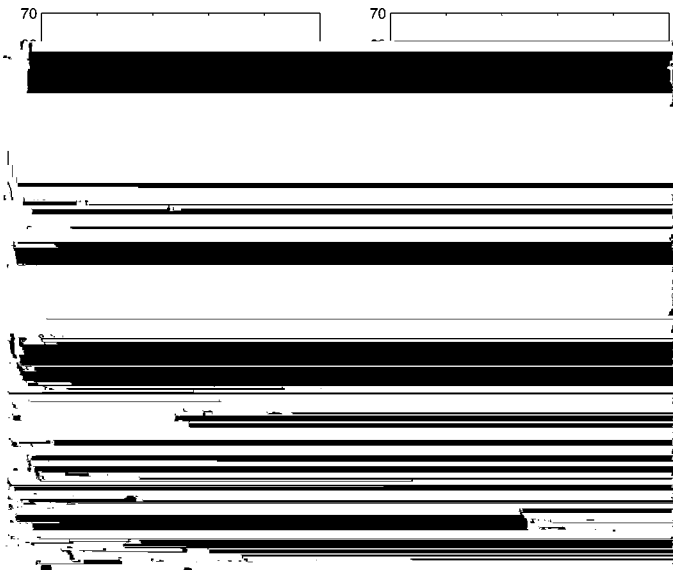
early times, decreasing until about 1 year and then increasing again. Measurements made on the same subject are strongly correlated, irrespective of the time difference. However, the correlation between very early and late counts dies off relatively rapidly, whereas for middle and later times, the dependence patterns persist more strongly. These features would be difficult to anticipate in a traditional parametric model; they would not be produced, by, for example, linear random-effects models.

Next, consider the eigenfunction decomposition of the estimated covariance surface. Three eigenfunctions shown in the upper panels of Figure 4 are used to approximate the infinite-dimensional process. The choice $= 3$ emerges as a reasonable choice, supported both by the AIC (11) and one-curve-leave-out cross-validation. The first eigenfunction is somewhat similar to the mean function, the second corresponds to a contrast between very early and very late times, and the third corresponds to a contrast between the early and the medium plus later times. These eigenfunctions account for 76.9%, 12.3%, and 8.1% of the total variation. Most of the variability is thus in the direction of overall CD4 percentage level. In exploring such data, extreme individual cases are difficult to detect by visual examination due to irregular sampling and substantial noise. One way to explore the variability in the sample and to single out extreme cases is to identify cases that exhibit large principal component scores in the directions of a few leading eigenfunctions (Jones and Rice 1992). Three such cases, corresponding to the largest absolute values of the projections on the first three eigenfunctions, are shown in the lower panels of Figure 4.

The predicted curves and 95% pointwise and simultaneous confidence bands for four randomly chosen individuals are displayed in Figure 5, where the principal component scores of

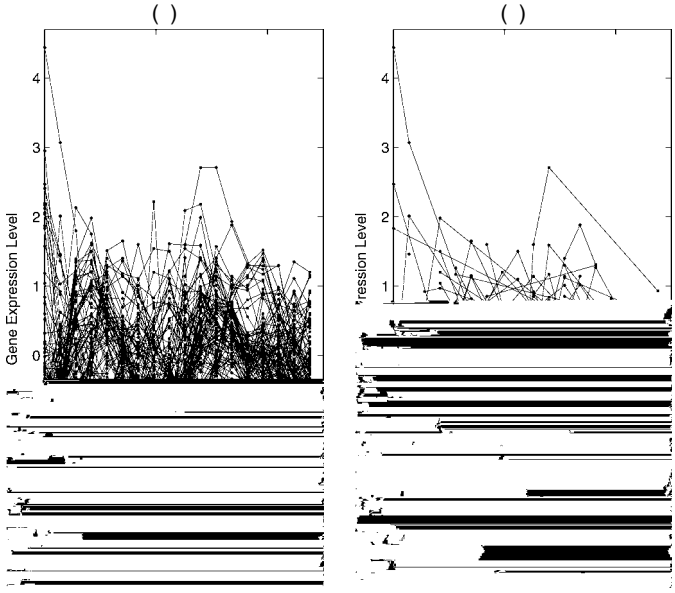


F 4. E F E . . . CD4 C . D (.) . O . ( ) . P . (.) I . L . P . . R E . . . A , O . O E M . F . . (.) ( ) (.).



F 5. O . (.), P . (.) . 95% P . (.) . (.) B . F R . C . I . , CD4 C . D .

each subject are estimated using the PACE method. The predicted curves are seen to be reasonably close to the observations. Even for individuals with very sparse measurements, one is still able to effectively recover their random trajectories, combining the information from that individual and the entire collection. For example, the PACE principle of borrowing strength from the entire sample for predicting individual trajectories makes it feasible to predict trajectories and construct corresponding prediction bands for those cases where only one observation available per subject, as exemplified in the lower left panel of Figure 5. The predictions based on only one observation per subject work reasonably well, as is demonstrated in the second example described in Section 5.2 (see the lower right panel in Fig. 9, where only one single measurement enclosed in the circle is used for the prediction of the trajectory). Because we need to be able to consistently estimate the covariance structure, it is, however, not feasible to apply the method if there is only one observation available per subject for all subjects. Note that the 95% simultaneous bands show a widening near the endpoints due to end effects and increased variance near the ends, and that all observed data fall within these bands.



F . 6. C n M n . ( ) G . E . P . . R . n ( ) 92 C C .

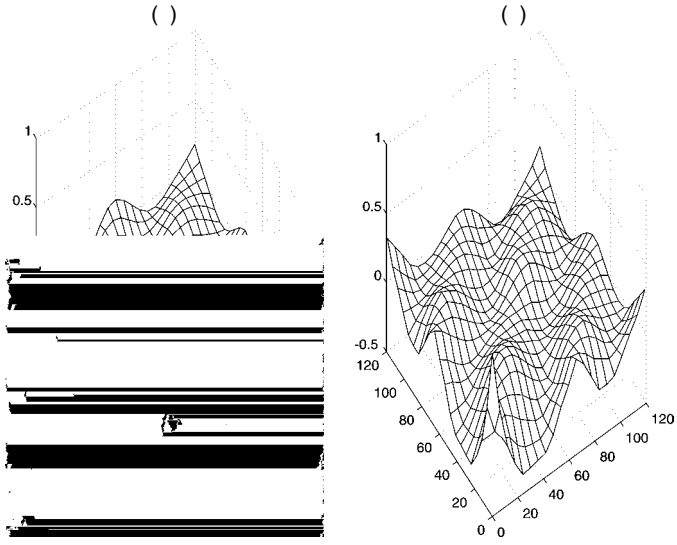
5.2 C C G . E . P

Time-course gene expression data (factor-synchronized) for the yeast cell cycle were obtained by Spellman et al. (1998). The experiment started with a collection of yeast cells, whose cycles were synchronized (α factor-based) by a chemical process. There are 6,178 genes in total, and each gene expression profile consists of 18 data points, measured every 7 minutes between 0 and 119 minutes, covering two cell cycles. Of these genes, 92 had sufficient data and were identified by traditional methods, of which 43 are related to G1 phase regulation and 49 are related to non-G1 phase regulation (i.e., S, S/G2, G2/M, and M/G1 phases) of the yeast cell cycle; these genes serve as a training set. The gene expression level measurement at each time point is obtained as a logarithm of the expression-level ratio.

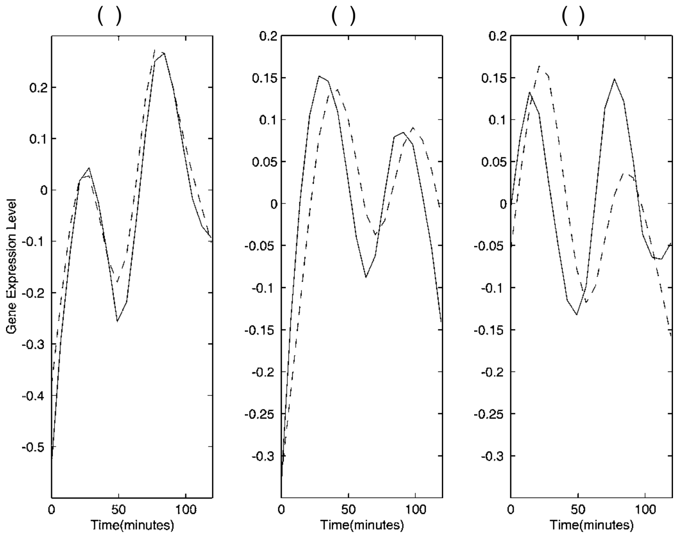
To demonstrate the usefulness of the PACE method for sparse functional data, we artificially "sparsify" the measurements made for the genes in the training data, then compare the results obtained from this "sparsified" data with those obtained from the complete data. To sparsify the expression measurements made for the th gene expression profile, the number of measurements is randomly chosen between 1 and 6 with equal probability, and the measurement locations are then randomly selected from the 18 recorded gene expression measurements per profile. The median number of observations per gene expression profile for the resulting sparse data is just 3.

Analyses of both complete and sparsified yeast cell cycle profile data are illustrated in Figures 6–8. The two mean function estimates for the sparse and complete data, obtained by local linear smoothing of the pooled data, are close to each other and demonstrate periodicity [see Fig. 8(a), presenting two cell cycles]. The two smooth covariance surface estimates revealing the structure of the underlying process are displayed in Figure 7. Both surfaces are very similar and exhibit periodic features. We use the first two eigenfunctions to approximate the expression profiles [Figs. 8(a) and 8(c)]. The estimates of the first two eigenfunctions obtained from both sparse and complete data are also close and reflect periodicity, explaining approximately 75% of the total variation.

We randomly select four genes, and present the predicted profiles obtained from both sparse and complete data and the confidence bands using only the sparse data in Figure 9. We note that the trajectories obtained for the complete data are enclosed in the simultaneous 95% confidence bands constructed from the sparse data. The predictions obtained from the sparse data are similar to those constructed from the complete data and are reasonable when compared with the complete measurements. This demonstrates that the PACE method allows us to effectively recover entire individual trajectories from fragmental data.



F . 7. n E n Ĝ (A.2) C . F . - . O F n C n D ( ) . F n D ( ) C C G . E . P .

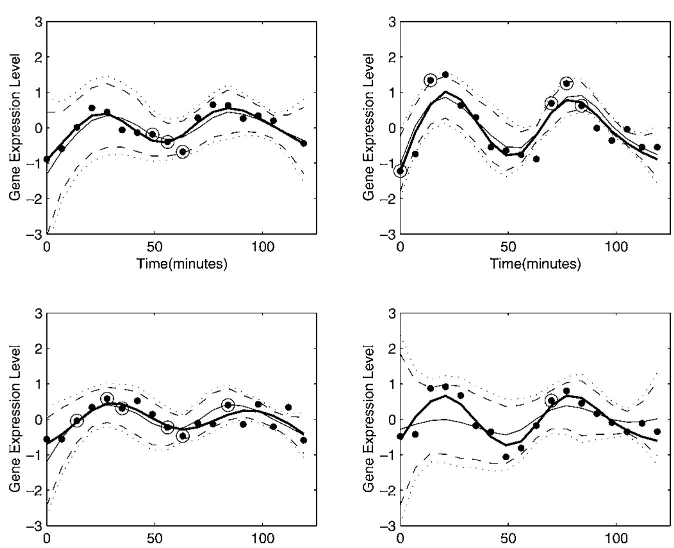


F 8. E M . F . . ( ), F ( ) . . ( ) E . . , O F ( ) . C ( ) G . E . D .

6. CONCL DING REMARK

Besides the general application to FPC analysis for sparse and irregular data, an application of our proposed PACE method to impute missing data in longitudinal studies is also feasible. Consider a regular design where for some subjects many data are missing. The PACE method can then be used to impute the missing data from predicted trajectories.

An interesting finding from the simulation study is that the PACE method improves on traditional FPC analysis even under dense and regular designs. This improvement is due to replacing integrals by conditional expectations when determining FPC scores. The conditioning step can be interpreted as shrinkage of these random effects toward 0. The observed improvement indicates that PACE can also be used to advantage for regularly spaced data, which enhances this method's appeal. We conclude that the underlying principle of borrowing strength from an entire sample of curves to predict individual trajectories shows promise in applications.



F 9. P G . E . P O F C M ( ) . F M . ( ) F R . G . A . 95% ( ) . ( ) . , . .

APPENDI : PROOF AND A ILIAR RE L

We assume regularity conditions for the marginal and joint densities ( ), (,), and $_2($ $_1,$ $_2,$ $_1,$ $_2)$. Let v_1, v_2, and be given integers, with $0 \le v_1 + v_2 <$. Then the following conditions apply:

(B1.1) (/) ( ) exists and is continuous on $\mathcal{T}$ with ( ) > 0 on $\mathcal{T}$.

(B1.2) (/) (,) exists and is uniformly continuous on $\mathcal{T} \times \Re$.

(B1.3) (/($_1^{\ 1}$ $_2^{\ 2}$)) $_2($ $_1,$ $_2,$ $_1,$ $_2)$ exists and is uniformly continuous on $\mathcal{T}^2 \times \Re^2$, for $_1 +$ $_2 =$, $0 \le$ $_1,$ $_2 \le$.

The assumptions for kernel functions $\kappa_1 : \Re \to \Re$ and $\kappa_2 : \Re^2 \to \Re$ are as follows. We say that a bivariate kernel function κ_2 is of order $(v,\ )$, where v is a multi-index $v = (v_1, v_2)$, if

$$\iint \ \ ^1 \ ^2 \kappa_2(\ ,) \qquad ,$$

$$= \begin{cases} 0, & 0 \le \ _1 + \ _2 < \ , \ _1 \ne v_1, \ _2 \ne v_2 \\ (-1)^{|v|}|v|!, & _1 = v_1, \ _2 = v_2 \\ \ne 0, & _1 + \ _2 = \ , \end{cases} \qquad \text{(A.1)}$$

where $|v| = v_1 + v_2$. A univariate kernel κ_1 is of order $(v,\ )$ for a univariate $v = v_1$, if (A.1) holds with $_2 = 0$ on the right side, integrating only over the argument on the left side.

(B2.1a) κ_1 is compactly supported, $\|\kappa_1\|^2 = \int \kappa_1^2(\ ) \quad < \infty$.

(B2.2a) κ_1 is a kernel function of order $(v,\ )$.

(B2.1b) κ_2 is compactly supported, $\|\kappa_2\|^2 = \int\int \kappa_2^2(\ ,\ ) \times \quad < \infty$.

(B2.2b) κ_2 is a kernel function of order $(v,\ )$.

We define the local linear scatterplot smoother for $\mu(\ )$ by minimizing

$$\sum_{=1} \sum_{=1} \kappa_1\left(\frac{\ -\ }{\mu}\right)\{\ \ - \beta_0 - \beta_1(\ -\ )\}^2 \qquad \text{(A.2)}$$

with respect to β_0 and β_1. The estimate of $\mu(\ )$ is then $\hat{\mu}(\ ) = \hat{\beta}_0(\ )$. The local linear surface smoother for (,) is defined by minimizing

$$\sum_{=1} \sum_{1 \le \ \ne \ \le} \kappa_2\left(\frac{\ -\ }{\ }, \frac{\ -\ }{\ }\right)$$

$$\times \{\ (\ ,\ ) - (\ ,(\ ,\ ),(\ ,\ ))\}^2, \qquad \text{(A.3)}$$

where (,(,),(,)) $= \beta_0 + \beta_{11}(\ -\ ) + \beta_{12}(\ -\ )$. Minimization is with regard to $= (\beta_0, \beta_{11}, \beta_{12})$, yielding the estimate $\widehat{\ }(\ ,\ ) = \hat{\beta}_0(\ ,\ )$. To obtain the adjusted estimate of (,) on the diagonal [i.e., $\tilde{\ }(\ )$], we first rotate both the -axis and -axis by 45 degrees clockwise and obtain the coordinates of (,) in the rotated axes, denoted by $(\ ^*,\ ^*)$, that is, $\binom{\ ^*}{\ ^*} = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}\binom{\ }{\ }$. We then define the surface estimate $\bar{\ }(\ ,\ )$ by minimizing the weighted least squares,

$$\sum_{=1} \sum_{1 \le \ \ne \ \le} \kappa_2\left(\frac{\ ^* -\ }{\ }, \frac{\ ^* -\ }{\ }\right)$$

$$\times \{\ (\ ^*,\ ^*) - \ ,(\ ,\ ),(\ ^*,\ ^*))\}^2, \qquad \text{(A.4)}$$

where $\gamma(\cdot,\cdot,(\cdot^*,\cdot^*)) = \gamma_0 + \gamma_1(\cdot - \cdot^*) + \gamma_2(\cdot - \cdot^*)^2$. Minimization is with respect to $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, leading to $\bar{\cdot}(\cdot,\cdot) = \hat{\gamma}_0(\cdot,\cdot)$. Because of the rotation, the estimate of the covariance surface on the diagonal, $\tilde{\cdot}(\cdot)$, is now indeed $\bar{\cdot}(0, \cdot/\sqrt{2})$, obtained with the rotated coordinates.

The following auxiliary results provide the weak uniform convergence rate for univariate weighted averages defined later (cf. Bhattacharya and Müller 1993). For a positive integer ≥ 1, let $(\psi_{\cdot})_{=1,\ldots,}$ be a collection of real functions $\psi_{\cdot} : \Re^2 \to \Re$, that satisfy the following conditions:

(C1.1a) $\psi_{\cdot}$ are uniformly continuous on $\mathcal{T} \times \Re$.

(C1.2a) The functions $(\cdot/\cdot)\psi_{\cdot}(\cdot,\cdot)$ exist for all arguments $(\cdot,\cdot)$ and are uniformly continuous on $\mathcal{T} \times \Re$.

(C1.3a) $\int \psi_{\cdot}^2(\cdot,\cdot)(\cdot,\cdot) < \infty$.

Bandwidths $\mu = \mu(\cdot)$ used for one-dimensional smoothers are assumed to satisfy the following:

(C2.1a) $\mu \to 0$, $\nu_{\mu}^{\nu+1} \to \infty$, $\nu_{\mu}^{2+2} < \infty$, as $\to \infty$.

Define the weighted averages

$$= (\cdot)$$

$$= \frac{1}{\nu_{\mu}^{\nu+1}} \sum_{=1} \frac{1}{} \sum_{=1} \psi_{\cdot}(\cdot,\cdot)\kappa_1\left(\frac{-}{\mu}\right), \qquad = 1,\ldots,,$$

and the quantity

$$\mu = \mu(\cdot)$$

$$= \frac{\nu}{\nu} \int \psi_{\cdot}(\cdot,\cdot)(\cdot,\cdot) \,, \qquad = 1,\ldots,.$$

Under (A1.1), (A1.2), (A3.1), (B1.1), (B1.2), (B2.1a), (B2.2a), (C1.1a)–(C1.3a), and C(2.1a), $\tau = \sup_{\in\mathcal{T}} | - \mu | = (1/(\sqrt{\nu_{\mu}^{\nu+1}}))$.

Note that $|\tau| \leq \sup | - \mu | + \{\sup | - |\}$, where takes values in $\mathcal{T}$ and $|\tau| = (1/(\sqrt{\nu_{\mu}^{\nu+1}}))$ implies that $\tau = (1/(\sqrt{\nu_{\mu}^{\nu+1}}))$.

Using a Taylor expansion to order , it is easy to show that $= \mu + (\mu^{-\nu})$, where the remainder term is uniform in , observing that $(\cdot/\cdot)\psi_{\cdot}(\cdot,\cdot)$ and $(\cdot/\cdot)(\cdot,\cdot)$ are uniformly continuous. It remains to show that $\{\sup | - |\} = (1/(\sqrt{\nu_{\mu}^{\nu+1}}))$. Recall that the inverse Fourier transform is $\zeta_1(\cdot) = \int^{-} \kappa_1(\cdot)$. We may insert $\kappa_1((-)/\mu) = \int \cdot^{(-)/\mu} \zeta_1(\cdot)\cdot/(2\pi)$ into . Letting

$$\phi(\cdot) = \frac{1}{} \sum_{=1} \frac{1}{} \sum_{=1} \psi_{\cdot}(\cdot,\cdot),$$

we obtain

$$= \frac{1}{\nu_{\mu}^{\nu+1}} \sum_{=1} \frac{1}{} \sum_{=1} \kappa_1\left(\frac{-}{\mu}\right)\psi_{\cdot}(\cdot,\cdot)$$

$$= \frac{1}{2\pi\,\nu_{\mu}^{\nu}} \int \phi(\cdot) - \zeta_1(\mu) \,,$$

and thus

$$\sup | - | \leq \frac{1}{2\pi\,\nu_{\mu}^{\nu}} \int |\phi(\cdot) - \phi(\cdot)| \cdot |\zeta_1(\mu)| \,.$$

Note that $|\phi(\cdot) - \phi(\cdot)| \leq \sqrt{[\phi(\cdot) - \phi(\cdot)]^2}$, and because $\{\widetilde{\mathbf{T}}, \widetilde{\mathbf{Y}}, \}$ are iid, using the Cauchy–Schwarz inequality,

$$\operatorname{var}(\phi(\cdot)) = \frac{1}{}\operatorname{var}\left\{\frac{1}{}\sum_{=1} \psi_{\cdot}(\cdot,\cdot)\right\}$$

$$\leq \frac{1}{}\left\{\left(\frac{1}{}\sum_{=1} \psi_{\cdot}(\cdot,\cdot)\right)^2\right\}$$

$$\leq \frac{1}{}\left\{\frac{1}{(\cdot)^2}\left(\sum_{=1} {}^2\right)\left(\sum_{=1}\psi_{\cdot}^2(\cdot,\cdot)\right)\right\}$$

$$\leq \frac{1}{}\left\{\frac{}{(\cdot)^2}\sum_{=1}\left(\psi_{\cdot}^2(\cdot,\cdot)|\right)\right\} = \frac{1}{}\,\psi^{\mathbf{2}}(\cdot,\cdot),$$

implying that

$$\left\{\sup | - |\right\}$$

$$\leq \frac{1}{2\pi\,\nu_{\mu}^{\nu}} \int \{|\phi$$

P n 1

From (A.2), the local linear estimator $\hat{\mu}(\ )$ of the mean function $\mu(\ )$ can be written explicitly as

$$\hat{\mu}(\ ) = \hat{\beta}_0(\ ) = \frac{\sum \frac{1}{} \sum}{\sum \frac{1}{} \sum} - \frac{\sum \frac{1}{} \sum \quad (\quad - \ )}{\sum \frac{1}{} \sum} \hat{\beta}_1(\ ), \quad \text{(A.5)}$$

where

$$\hat{\beta}_1(\ ) = \Bigg(\sum \frac{1}{} \sum \quad (\quad - \ ) - \Bigg(\sum \frac{1}{} \sum \quad (\quad - \ ) \sum \frac{1}{} \sum \quad \Bigg) \Big/ \Bigg(\sum \frac{1}{} \sum \quad \Bigg) \Bigg) \times \Bigg(\sum \frac{1}{} \sum \quad (\quad - \ )^2 - \Bigg(\sum \frac{1}{} \sum \quad (\quad - \ ) \Bigg)^2 \Big/ \Bigg(\sum \frac{1}{} \sum \quad \Bigg) \Bigg)^{-1}. \quad \text{(A.6)}$$

Here $\quad = \kappa_1((\ - \quad)/\ \mu)/(\quad \mu)$, where κ_1 is a kernel function of order $(0, 2)$ satisfying (B2.1a) and (B2.2a), and $\hat{\beta}_1(\ )$ is an estimator for the first derivative $\mu'(\ )$ of μ at $\ $.

Considering the Nadaraya–Watson estimator of μ, $\hat{\mu} \quad (\ ) = (\sum \sum \quad / \quad)/(\sum \sum \quad / \quad)$ and $\hat{\ }(\ ) = \sum \sum \quad / \quad$, we choose $\nu = 0$, $\quad = 2$, $\quad = 2$, $\psi_1(\ ,\ ) = \quad$, and $\psi_2(\ ,\ ) \equiv 1$ in Lemma A.1. Obviously, $\hat{\mu} \quad (\ ) = (\ _1, \ _2)$, with $(\ _1, \ _2) = \ _1/\ _2$ and $\hat{\ }(\ ) = \ _2$. Using Slutsky's theorem and Lemma A.1, it follows that $\sup_{\ \in \mathcal{T}} |\hat{\mu} \quad (\ ) - \mu(\ )| = \quad (1/(\sqrt{\ } \mu))$ and $\sup_{\ \in \mathcal{T}} |\hat{\ }(\ ) - \ (\ )| = \quad (1/(\sqrt{\ } \mu))$.

For the uniform consistency of $\hat{\beta}_1$ as an estimator of the derivative μ', define $\quad, 1 \le \ \le 3$, $\sigma_{\kappa_1}^2 = \int \ ^2 \kappa_1(\ ) \quad$, and the kernel function $\tilde{\kappa}_1(\ ) = -\ \kappa_1(\ )/\sigma_{\kappa_1}^2$; furthermore, $\psi_1(\ ,\ ) = \quad, \psi_2(\ ,\ ) \equiv 1$ and $\psi_3(\ ,\ ) = \ - \quad$. Observe that $\tilde{\kappa}_1$ is of order $(1, 3)$, $\sup_{\ \in \mathcal{T}} |\hat{\ }(\ ) - \ (\ )| = \quad (1/(\sqrt{\ } \mu))$, and define

$$\tilde{\ }(\ _1, \ _2, \ _3) = \frac{\ _1 - \ _2 \hat{\mu} \quad (\ )}{\ _3 - \ _\mu^2 \ _2^2 / \hat{\ }(\ ) \cdot \sigma_{\kappa_1}^2} \quad \text{and}$$

$$(\ _1, \ _2, \ _3) = \frac{\ _1 - \ _2 \mu(\ )}{\ _3}.$$

Then

$$\hat{\beta}_1(\ ) = \tilde{\ }(\ _1, \ _2, \ _3) = \left[(\ _1, \ _2, \ _3) + \frac{\ _2 \ (\mu(\ ) - \hat{\mu} \quad (\ ))}{\ _3} \right] \times \frac{\ _3}{\ _3 + \ _\mu^2 \ _2^2 / \hat{\ }(\ ) \cdot \sigma_{\kappa_1}^2}.$$

Note that $\mu_1 = (\mu' + \ ')(\ )$, $\mu_2 = \ '(\ )$, and $\mu_3 = \ (\ )$, implying that $\sup_{\ \in \mathcal{T}} | \quad - \mu \ | = \quad (1/(\sqrt{\ } \mu))$, for $\quad = 1, 2, 3$, by Lemma A.1. Using the uniform version of Slutsky's theorem, $\sup_{\ \in \mathcal{T}} | \ (\ _1, \ _2, \ _3) - \mu'(\ )| = \quad (1/(\sqrt{\ } \mu))$ follows.

Considering the uniform convergence of $\hat{\beta}_0$ for μ, note that

$$\hat{\beta}_0(\ ) = \hat{\mu} \quad (\ ) + \frac{\ _2 \ \hat{\beta}_1(\ )}{\hat{\ }(\ )} \ _\mu^2.$$

Because $\sup_{\ \in \mathcal{T}} | \ _2 \ - \ '(\ )| = \quad (1/(\sqrt{\ } \ _\mu^2))$, $\sup_{\ \in \mathcal{T}} |\hat{\beta}_1(\ ) - \mu'(\ )| = \quad (1/(\sqrt{\ } \ _\mu^2))$, and $\sup_{\ \in \mathcal{T}} |\hat{\ }'(\ ) - \ (\ )| = \quad (1/(\sqrt{\ } \mu))$, we have $\sup_{\ \in \mathcal{T}} | \ _2 \ \hat{\beta}_1(\ ) \ _\mu^2 / \hat{\ }(\ )| = \quad (\ _\mu^2) = \quad (1/(\sqrt{\ } \mu))$, as $\ _\mu^6 < \infty$. As $\sup_{\ \in \mathcal{T}} |\hat{\mu} \quad (\ ) - \mu(\ )| = \quad (1/(\sqrt{\ } \mu))$, the result (12) follows.

We proceed to show (13). In the local linear estimator for the covariance $\quad (\ ,\ )$, we used the raw observations, $\quad (\ ,\ ) = (\ - \hat{\mu}(\ ))(\ - \hat{\mu}(\ ))$, instead of $\tilde{\ }(\ ,\ ) = (\ - \mu(\ ))(\ - \mu(\ ))$. Note that

$$(\ ,\ ) = \tilde{\ }(\ ,\ ) + (\ - \mu(\ ))(\mu(\ ) - \hat{\mu}(\ )) + (\ - \mu(\ ))(\mu(\ ) - \hat{\mu}(\ )) + (\mu(\ ) - \hat{\mu}(\ ))(\mu(\ ) - \hat{\mu}(\ )).$$

Because $\sup_{\ \in \mathcal{T}} |\hat{\mu}(\ ) - \mu(\ )| = \quad (1/(\sqrt{\ } \mu))$ by (12), letting $\theta_1(\ _1, \ _2, \ _1, \ _2) = (\ _1 - \mu(\ _1))(\ _2 - \mu(\ _2))$, $\theta_2(\ _1, \ _2, \ _1, \ _2) = \ _1 - \mu(\ _1)$, and $\theta_3(\ _1, \ _2, \ _1, \ _2) \equiv 1$, then $\sup_{\ , \ \in \mathcal{T}} | \quad | = \quad (1)$, for $\quad = 1, 2, 3$, by Lemma A.2. This implies that $\sup_{\ , \ \in \mathcal{T}} | \ _2 \ | \ (1/(\sqrt{\ } \mu)) = \quad (1/(\sqrt{\ } \mu))$ and $\sup_{\ , \ \in \mathcal{T}} | \ _3 \ | \ (1/(\sqrt{\ } \mu)) = \quad (1/(\sqrt{\ } \mu))$. Because $\sup_{\ \in \mathcal{T}} |\hat{\mu}(\ ) - \mu(\ )|^2 = \quad (1/(\ _\mu^2))$ are negligible compared with $\ _1$, the local linear estimator, $\hat{\ }(\ ,\ )$, of $(\ ,\ )$ obtained from $\quad (\ ,\ )$ is asymptotically equivalent to that obtained from $\tilde{\ }(\ ,\ )$, denoted by $\tilde{\ }(\ ,\ )$. Analogously to the proof of (12), using Lemma A.2 and the uniform version of Slutsky's theorem, we obtain the uniform consistency of the local linear estimator $\hat{\ }(\ ,\ )$.

P C 1

Because $\hat{\ }(\ )$ is a uniformly consistent estimator of $\{ \quad (\ ,\ ) + \sigma^2\}$, analogously to (12), (14) follows by applying (13).

P n 2

Define the rank-one operator $\otimes \ = \langle \ ,\ \rangle$ for $\ ,\ \in \ $, and denote the separable Hilbert space of Hilbert–Schmidt operators on $\ $ by $\ \equiv \sigma_2(\ )$, endowed by $\langle \ _1, \ _2 \rangle \ = \mathrm{tr}(\ _1 \ _2^*) = \sum \langle \ _1 \ , \ _2 \ \rangle$ and $\| \ \|^2 = \langle \ ,\ \rangle$, where $\ _1, \ _2, \ \in \ $, $\ _2^*$ is the adjoint of $\ _2$ and $\{ \ : \ge 1\}$ is any complete orthonormal system in $\ $. The covariance operator $\mathbf{G}$ (resp. $\widehat{\mathbf{G}}$) is generated by the kernel $\quad$ (resp. $\hat{\ }$), that is, $\mathbf{G}(\ ) = \int_{\mathcal{T}} \quad (\ ,\ ) \ (\ ) \quad$ [resp. $\widehat{\mathbf{G}}(\ ) = \int_{\mathcal{T}} \hat{\ }(\ ,\ ) \ (\ ) \quad$]. It is obvious that $\mathbf{G}$ and $\widehat{\mathbf{G}}$ are Hilbert–Schmidt operators, and (13) implies that $\|\widehat{\mathbf{G}} - \mathbf{G}\| \ = \quad (1/(\sqrt{\ } \ ^2))$.

Let $\mathcal{I} = \{ \ : \lambda \ = \lambda \ \}$, $\mathcal{I}' = \{ \ : |\mathcal{I}| = 1\}$, where $|\mathcal{I}|$ denotes the number of elements in $\mathcal{I}$. To obtain (16), let $\mathbf{P} = \sum_{\ \in \mathcal{I}} \varphi \ \otimes \varphi \quad$ and $\widehat{\mathbf{P}} = \sum_{\ \in \mathcal{I}} \hat{\varphi} \ \otimes \hat{\varphi}$ denote the true and estimated orthogonal projection operators from $\ $ to the subspace spanned by $\{\varphi \ : \ \in \mathcal{I}\}$. For fixed $0 < \rho < \min\{|\lambda \ - \lambda \ | : \ \notin \mathcal{I}\}$, let $\mathbf{\Lambda}_{\rho, \ } = \{ \ \in \mathcal{C} : | \ - \lambda \ | = \rho\}$, where $\mathcal{C}$ stands for the complex numbers. The resolvent of $\mathbf{G}$ (resp. $\widehat{\mathbf{G}}$) is denoted by $\mathbf{R}$ (resp. $\widehat{\mathbf{R}}$), that is, $\mathbf{R}(\ ) = (\mathbf{G} - \ )^{-1}$ [resp. $\widehat{\mathbf{R}}(\ ) = (\widehat{\mathbf{G}} - \ )^{-1}$]. As $\widehat{\mathbf{R}}(\ ) = \mathbf{R}(\ )[\ + (\widehat{\mathbf{G}} - \mathbf{G})\mathbf{R}(\ )]^{-1} = \mathbf{R}(\ ) \sum_{\ =0}^{\infty} [(\widehat{\mathbf{G}} - \mathbf{G})\mathbf{R}(\ )] \ , \ \|\widehat{\mathbf{R}}(\ ) - \mathbf{R}(\ )\| \le (\|\widehat{\mathbf{G}} - \mathbf{G}\| \ \|\mathbf{R}(\ )\| \ )/(1 - \|\widehat{\mathbf{G}} - \mathbf{G}\| \ \|\mathbf{R}(\ )\| \ )$. Note that $\mathbf{P} = (2\pi \ )^{-1} \int_{\mathbf{\Lambda}_{\rho, \ }} \mathbf{R}(\ ) \quad , \ \widehat{\mathbf{P}} \ = (2\pi \ )^{-1} \int_{\mathbf{\Lambda}_{\rho, \ }} \widehat{\mathbf{R}}(\ ) \quad$. Let $\quad_{\rho, \ } = \sup\{\|\mathbf{R}(\ )\| \ : \ \in \mathbf{\Lambda}_{\rho, \ }\} < \infty$, and let $\ $ be such that $0 < \ < 1/(2 \quad_{\rho, \ })$; then

$$\|\widehat{\mathbf{P}} \ - \mathbf{P} \ \| \ \le \int_{\mathbf{\Lambda}_{\rho, \ }} \|\widehat{\mathbf{R}}(\ ) - \mathbf{R}(\ )\| \quad /(2\pi) \le \rho \frac{\|\widehat{\mathbf{G}} - \mathbf{G}\| \quad_{\rho, \ }}{1 - \|\widehat{\mathbf{G}} - \mathbf{G}\| \quad_{\rho, \ }} \le 2\rho \quad_{\rho, \ } .$$

Considering φ corresponding to $\in \mathcal{I}'$, choose $\hat{\varphi}$ such that $\langle \hat{\varphi}, \varphi \rangle > 0$. Then

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|^2 = 2\big(1 - \langle \hat{\varphi} \otimes \hat{\varphi}, \varphi \otimes \varphi \rangle\big) = 2\big(1 - \langle \hat{\varphi}, \varphi \rangle^2\big) \geq \|\hat{\varphi} - \varphi\|^2,$$

and (16) follows. Note that $\lambda = \langle \varphi, \mathbf{G}(\varphi) \rangle$ and $\hat{\lambda} = \langle \hat{\varphi}, \widehat{\mathbf{G}}(\hat{\varphi}) \rangle$; then (15) follows by applying Slusky's theorem. To obtain (17), for fixed $\in \mathcal{I}'$,

$$\begin{aligned}
&|\hat{\lambda}\ \hat{\varphi}(\ ) - \lambda\ \varphi(\ )| \\
&= \left| \int_0^{\mathcal{T}} \widehat{\ }(\ ,\ )\hat{\varphi}(\ )\ - \int_0^{\mathcal{T}} (\ ,\ )\varphi(\ ) \right| \\
&\leq \int_0^{\mathcal{T}} |\widehat{\ }(\ ,\ ) - (\ ,\ )| \cdot |\hat{\varphi}(\ )| \\
&\quad + \int_0^{\mathcal{T}} |(\ ,\ )| \cdot |\hat{\varphi}(\ ) - \varphi(\ )| \\
&\leq \sqrt{\int_0^{\mathcal{T}} (\widehat{\ }(\ ,\ ) - (\ ,\ ))^2} + \sqrt{\int_0^{\mathcal{T}} {}^2(\ ,\ )}\ \|\hat{\varphi} - \varphi\|.
\end{aligned}$$

Due to (13) and (16), assuming $\lambda > 0$ without loss of generality, we have $|\hat{\lambda}\ \hat{\varphi}(\ )/\lambda - \varphi(\ )| = (1/(\sqrt{\ }^2))$, uniformly in $\in \mathcal{I}$. Then (17) follows by applying (15).

The next result ensures that the target trajectory $\tilde{\ }$ is well defined.

For the positive definite covariance operator $\mathbf{G}$ generated by the continuous symmetric function $(\ ,\ )$ on $\mathcal{T}^2$, as $\to \infty$,

$$\sup_{\in \mathcal{I}} [\tilde{\ }(\ ) - \tilde{\ }(\ )]^2 \longrightarrow 0. \tag{A.7}$$

Because the covariance operator $\mathbf{G}$ generated by the continuous symmetric function $(\ ,\ )$ is positive definite, by Mercer's theorem, $\sum_{=}^{\infty} \lambda\ \varphi(\ )\varphi(\ )$ converges to 0 uniformly in $(\ ,\ ) \in \mathcal{T}^2$. Note that $\tilde{\ }(\ ) - \tilde{\ }(\ ) = [\sum_{=+1}^{\infty} \xi\ \varphi(\ )|\widetilde{\mathbf{Y}}]$. From

$$\begin{aligned}
&\sup_{\in \mathcal{I}} \operatorname{var}\left(\sum_{=+1}^{\infty} \xi\ \varphi(\ ) \right) \\
&= \sup_{\in \mathcal{I}} \left\{ \left[\left[\sum_{=+1}^{\infty} \xi\ \varphi(\ )|\widetilde{\mathbf{Y}} \right]^2 \right] \right. \\
&\qquad \left. + \left[\operatorname{var}\left(\sum_{=+1}^{\infty} \xi\ \varphi(\ )|\widetilde{\mathbf{Y}} \right) \right] \right\} \\
&= \sup_{\in \mathcal{I}} \sum_{=+1}^{\infty} \lambda\ \varphi^2(\ ) \longrightarrow 0,
\end{aligned}$$

and $[\operatorname{var}(\sum_{=+1}^{\infty} \xi\ \varphi(\ )|\widetilde{\mathbf{Y}})] \geq 0$, (A.7) follows.

P n 3

Recall that $\hat{\xi} = \hat{\lambda}\ \hat{} \ \widehat{\mathbf{\Sigma}}_{\mathbf{Y}}^{-1}(\widetilde{\mathbf{Y}} - \hat{\ })$, where the $(\ ,\ )$th entry of the $\times$ matrix $\widehat{\mathbf{\Sigma}}_{\mathbf{Y}}$ is $(\widehat{\mathbf{\Sigma}}_{\mathbf{Y}})_{,} = \widehat{\ }(\ ,\ ) + \hat{\sigma}^2\delta$ with $\delta = 1$ if $=$ and 0 if $\neq$. Applying Theorems 1 and 2, Corollary 1, and Slutsky's theorem, (20) follows. We next prove (21) for each fixed $\in \mathcal{T}$. Let $\tilde{\ }(\ ) = \mu(\ ) + \sum_{=1} \tilde{\xi}\ \varphi(\ )$, where $\tilde{\xi}$ is as defined in (4). Note that

$$|\widehat{\ }(\ ) - \tilde{\ }(\ )| \leq |\widehat{\ }(\ ) - \tilde{\ }(\ )| + |\tilde{\ }(\ ) - \tilde{\ }(\ )|.$$

Lemma A.3 implies that $\tilde{\ }(\ ) \to \tilde{\ }(\ )$ as $\to \infty$. For fixed $$, observing that $\hat{\xi} \to \tilde{\xi}$ as $\to \infty$, $\sup_{\in \mathcal{I}} |\widehat{\ }(\ ) - \tilde{\ }(\ )| \to 0$ as $\to \infty$ by (12), (17), and Slutsky's theorem. This implies that for given $, \delta > 0$, there exists $_0$ such that for $\geq {}_0$, $\{|\tilde{\ }(\ ) - \tilde{\ }(\ )| > /2\} \leq \delta/2$. For each $$, there exists $_0(\ )$ such that for $\geq {}_0(\ )$, $\{|\widehat{\ }(\ ) - \tilde{\ }(\ )| \geq /2\} \leq \delta/2$. Thus for $\geq {}_0$ and $\geq {}_0(\ )$, $\{|\widehat{\ }(\ ) - \tilde{\ }(\ )| \geq \} \leq \{|\widehat{\ }(\ ) - \tilde{\ }(\ )| > /2\} + \{|\tilde{\ }(\ ) - \tilde{\ }(\ )| \geq /2\} \leq \delta$, which leads to (21).

P n 4

Under the Gaussian assumption, for any fixed ≥ 1, from Section 2.4, we have $(\tilde{\ }_{,} -_{,} \ _{,}) \sim \mathcal{N}(\mathbf{0}, \ )$. Observing (12), (17), and (20), $\lim_{\to\infty} \sup_{\in \mathcal{I}} |\widehat{\ }(\ ) - \tilde{\ }(\ )| \to 0$. Because $\widehat{\ }(\ ) - (\ ) = \widehat{\ }(\ ) - \tilde{\ }(\ ) + \tilde{\ }(\ ) - (\ )$ for fixed $$, it follows that $\{\widehat{\ }(\ ) - (\ )\} \xrightarrow{\mathcal{D}} \sim \mathcal{N}(0, \omega(\ ,\ ))$. Under condition (A7), letting $\to \infty$ leads to $\xrightarrow{\mathcal{D}} \sim \mathcal{N}(0, \omega(\ ,\ ))$. From the Karhunen–Loève theorem, $|(\ ) - (\ )| \to 0$, as $\to \infty$. Therefore, $\lim_{\to\infty} \lim_{\to\infty} \{\widehat{\ }(\ ) - (\ )\} \overset{\mathcal{D}}{=}$. From Theorems 1 and 2, it can be shown that $\hat{\omega}(\ ,\ ) \to \omega(\ ,\ )$ as $\to \infty$. Under condition (A7), it follows that $\lim_{\to\infty} \lim_{\to\infty} \hat{\omega}(\ ,\ ) = \omega(\ ,\ )$ in probability. Applying Slutsky's theorem, (22) follows.

P n 5

We first prove

$$\left\{ \sup_{\in \mathcal{I}} \frac{|\tilde{\ }(\ ) - (\ )|}{\sqrt{\omega(\ ,\ )}} \leq \sqrt{\chi^2_{,1-\alpha}} \right\} \geq 1 - \alpha. \tag{A.8}$$

It is obvious that $\tilde{\ }(\ ) - (\ ) =$, $(\tilde{\ }_{,} -_{,} \ _{,})$. Due to orthogonality, $\mathcal{F} =$ $: \in \mathcal{T}\}$ is a -dimensional compact set. Because , is positive definite, there exists a $\times$ nonsingular matrix $\mathbf{U}$ such that $\mathbf{U}_{,} \ \mathbf{U} = \mathbf{I}$. Let $_{;} = \mathbf{U}_{,} \ _{,}$ and $\tilde{\ }_{;} = \mathbf{U}_{,} \tilde{\ }_{,}$; then $(\tilde{\ }_{;} -_{;}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. This leads to $(\tilde{\ }_{;} -_{;})\ (\tilde{\ }_{;} -_{;}) \sim \chi^2$ and $\{(\tilde{\ }_{;} -_{;})\ (\tilde{\ }_{;} -_{;})\} = 1 - \alpha$. We use the following result, known from linear algebra.

For a fixed -vector $\mathbf{x}$ and a constant > 0, $\mathbf{x}\ \mathbf{x} \leq {}^2$ if and only if $|\mathbf{a}\ \mathbf{x}| \leq \sqrt{\mathbf{a}\ \mathbf{a}}$, for all $\mathbf{a} \in \Re$.

Hence $\{|\mathbf{a}\ (\tilde{\ }_{;} -_{;})| \leq \sqrt{\chi^2_{,1-\alpha}\mathbf{a}\ \mathbf{a}}$: for all $\mathbf{a} \in \Re\ \} = 1 - \alpha$. Let $\mathcal{E} = \{\mathbf{a} \in \Re$: $_{,} = \mathbf{U}\ \mathbf{a}, \in \mathcal{T}\}$, which is a compact subset of $\Re$. Then $\{|\mathbf{a}\ (\tilde{\ }_{;} -_{;}) \leq \sqrt{\chi^2_{,1-\alpha}\mathbf{a}\ \mathbf{a}}$: for all $\mathbf{a} \in \mathcal{E}\} \geq 1 - \alpha$, that is,

$$\begin{aligned}
&\{| \ _{,}(\tilde{\ }_{,} -_{,} \ _{,})| \\
&\quad \leq \sqrt{\chi^2_{,1-\alpha}}\ _{,} \mathbf{U}^{-1}(\mathbf{U}\ )^{-1}\ _{,} : \text{for all } \in \mathcal{T}\} \\
&\geq 1 - \alpha.
\end{aligned}$$

Observing that $\mathbf{U}_{,} \ \mathbf{U} = \mathbf{I}$, (A.8) follows.

To prove (23), note that

$$\begin{aligned}
&\sup_{\in \mathcal{I}} \frac{|\widehat{\ }(\ ) - (\ )|}{\sqrt{\omega(\ ,\ )}} \\
&\leq \left(\sup_{\in \mathcal{I}} \frac{|\widehat{\ }(\ ) - \tilde{\ }(\ )|}{\sqrt{\omega(\ ,\ )}} + \sup_{\in \mathcal{I}} \frac{|\tilde{\ }(\ ) - (\ )|}{\sqrt{\omega(\ ,\ )}} \right) \sup_{\in \mathcal{I}} \sqrt{\frac{\omega(\ ,\ )}{\hat{\omega}(\ ,\ )}}.
\end{aligned}$$

Let $= \sup_{\in \mathcal{I}} |\widehat{\ }(\ ) - \tilde{\ }(\ )|/\sqrt{\omega(\ ,\ )}$, $= \sup_{\in \mathcal{I}} |\tilde{\ }(\ ) - (\ )|/\sqrt{\omega(\ ,\ )}$, and $= \sup_{\in \mathcal{I}} \sqrt{\omega(\ ,\ )/\hat{\omega}(\ ,\ )}$. Because $\omega(\ ,\ )$ is a continuous positive definite function on the bounded interval $\mathcal{T}$, it is bounded from above and below, say $0 < \leq \omega(\ ,\ ) \leq < \infty$. Because $\sup_{\in \mathcal{I}} |\widehat{\ }(\ ) - \tilde{\ }(\ )| \to 0$ as $\to \infty$, we have $\to 0$ as $\to \infty$. In the proof of (22), we established that $\hat{\omega}(\ ,\ ) \to \omega(\ ,\ )$, as $\to \infty$, implying that $\to 1$. We now show that

$$\lim_{\to\infty} \{(\ +\ ) \geq (\ + \sqrt{\chi^2_{,1-\alpha}})(1 + \ )\} \leq \alpha. \tag{A.9}$$

Note that

$$\begin{aligned}&\{(\quad + \quad) \quad \geq \big(\quad + \sqrt{\chi^2_{,1-\alpha}}\big)(1 + \quad)\}\\ &\quad \subseteq \{(\quad + \quad) \geq \big(\quad + \sqrt{\chi^2_{,1-\alpha}}\big)\} \cup \{ \quad \geq (1 + \quad)\}\\ &\quad \subseteq \{ \quad \geq \quad \} \cup \{ \quad \geq \sqrt{\chi^2_{,1-\alpha}}\} \cup \{ \quad \geq (1 + \quad)\}.\end{aligned}$$

Because $\quad \to 0$ and $\quad \to 1$ as $\quad \to \infty$, for sufficiently large $\quad$, $(\quad \geq \quad) \leq \tau/3$ and $(\quad - 1 \geq \quad) \leq \tau/3$. We have shown that $(\quad \geq \sqrt{\chi^2_{,1-\alpha}}) \leq \alpha$ in (A.8). This implies (A.9), and then (23), by letting $\to 0$.

P C 2

There exists a $\times$ matrix $\mathbf{Q}$ with rank $\leq$ such that $\mathcal{F}$ is spanned by the column vectors of $\mathbf{Q}$. Letting $, = \mathbf{Q}_{,}$ and $,\tilde{} = \mathbf{Q}_{,}\tilde{}$, for any $\mathbf{l} \in \mathcal{A}$, where $\mathcal{A} \subseteq \Re$ is a linear space with dimension , there exists a vector $\in \Re$ such that $\mathbf{l} = \mathbf{Q}$. Then

$$\mathbf{l}_{,}\tilde{} \quad , - \mathbf{l}_{,} \quad , = \quad ,\tilde{} - \quad , \sim \mathcal{N}(\mathbf{0}, \quad \mathbf{Q} \quad \mathbf{Q} \quad).$$

Because $\mathbf{Q}$ is of rank and $,$ is positive definite, which implies that $\mathbf{Q} \quad \mathbf{Q}$ is also positive definite, there exists a nonsingular $\times$ matrix $\mathbf{P}$ such that $\mathbf{P}\mathbf{Q}_{,} \quad \mathbf{Q} \ \mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is the $\times$ identity matrix. Letting $\flat = \mathbf{P}_{,}$ and $\natural = \mathbf{P}_{,}\tilde{}$, we have $(\natural - \flat) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, that is, $(\natural - \flat) \ (\natural - \flat) \sim \chi^2$. Therefore, $\{(\natural - \flat) \ (\natural - \flat) \leq \chi^2_{,1-\alpha}\} = 1 - \alpha$. Applying Lemma A.4, we obtain $\{|\mathbf{a} \ (\natural - \flat)| \leq \sqrt{\chi^2_{,1-\alpha}\mathbf{a} \ \mathbf{a}}$: for all $\mathbf{a} \in \Re \ \} = 1 - \alpha$. Because $\mathbf{P}$ is nonsingular and $\mathbf{Q}$ is of rank , there exists $\in \Re$ and $\mathbf{l} \in \mathcal{A}$, such that $= \mathbf{P} \ \mathbf{a}$ and $\mathbf{l} = \mathbf{Q}$. If $\mathbf{a}$ takes all values in $\Re$, then $\mathbf{l}$ will also take all values in $\mathcal{A}$, that is,

$$\begin{aligned}&\{|\mathbf{l} \ (\tilde{} \quad , -_{,} \quad ,)|\\ &\qquad \leq \sqrt{\chi^2_{,1-\alpha}\mathbf{l} \ (\mathbf{P}\mathbf{Q})^{-1}(\mathbf{Q} \ \mathbf{P} \ )^{-1}\mathbf{l}}\text{: for all } \mathbf{l} \in \mathcal{A}\} = 1 - \alpha.\end{aligned}$$

Because $\mathbf{P}\mathbf{Q}_{,} \quad \mathbf{Q} \ \mathbf{P} = \mathbf{I}$, the result (24) follows.

REFERENCE

Berkey, C. S., Laird, N. M., Valadian, I., and Gardner, J. (1991), "Modeling Adolescent Blood Pressure Patterns and Their Prediction of Adult Pressures," , 47, 1005–1018.

Besse, P., Cardot, H., and Ferraty, F. (1997), "Simultaneous Nonparametric Regression of Unbalanced Longitudinal Data," , 24, 255–270.

Besse, P., and Ramsay, J. O. (1986), "Principal Components Analysis of Sampled Functions," , 51, 285–311.

Bhattacharya, P. K., and Müller, H. G. (1993), "Asymptotics for Nonparametric Regression," , 55, 420–441.

Boente, G., and Fraiman, R. (2000), "Kernel-Based Functional Principal Components," , 48, 335–345.

Boularan, J., Ferré, L., and Vieu, P. (1993), "Growth Curves: A Two-Stage Nonparametric Approach," , 38, 327–350.

Bosq, D. (1991), "Modelization, Nonparametric Estimation and Prediction for Continuous Time Processes," in , ed. G. Roussas, Dordrecht, The Netherlands: Kluwer Academic, pp. 509–529.

Capra, W. B., and Müller, H. G. (1997), "An Accelerated-Time Model for Response Curves," , 92, 72–83.

Cardot, H., Ferraty, F., and Sarda, P. (1999), "Functional Linear Model," , 45, 11–22.

Castro, P. E., Lawton, W. H., and Sylvestre, E. A. (1986), "Principal Modes of Variation for Processes With Continuous Sample Curves," , 28, 329–337.

Courant, R., and Hilbert, D. (1953), , New York: Wiley.

Dauxois, J., Pousse, A., and Romain, Y. (1982), "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference," , 12, 136–154.

Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994), , Oxford, U.K.: Oxford University Press.

Fan, J., and Gijbels, I. (1996), , London: Chapman and Hall.

Fan, J., and Zhang, J. T. (2000), "Two-Step Estimation of Functional Linear Models With Applications to Longitudinal Data," , Ser. B, 62, 303–322.

Ferré, L. (1995), "Improvement of Some Multivariate Estimates by Reduction of Dimensionality," , 54, 147–162.

James, G., Hastie, T. G., and Sugar, C. A. (2001), "Principal Component Models for Sparse Functional Data," , 87, 587–602.

James, G., and Sugar, C. A. (2003), "Clustering for Sparsely Sampled Functional Data," , 98, 397–408.

Jones, M. C., and Rice, J. (1992), "Displaying the Important Features of Large Collections of Similar Curves," , 46, 140–145.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., and Rinaldo, C. R. (1987), "The Multicenter AIDS Cohort Study: Rationale, Organization and Selected Characteristics of the Participants," , 126, 310–318.

Kneip, A. (1994), "Nonparametric Estimation of Common Regressors for Similar Curve Data," , 22, 1386–1472.

Kneip, A., and Utikal, K. (2001), "Inference for Density Families Using Functional Principal Component Analysis," , 96, 519–532.

Lin, X., and Carroll, R. J. (2000), "Nonparametric Function Estimation for Clustered Data When the Predictor Is Measured Without/With Error," , 95, 520–534.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), , London: Academic Press.

Müller, H. G., and Prewitt, K. (1993), "Multiparameter Bandwidth Processes and Adaptive Surface Smoothing," , 47, 1–21.

Ramsay, J., and Silverman, B. (1997), , New York: Springer-Verlag.

Rao, C. R. (1958), "Some Statistical Methods for Comparison of Growth Curves," , 14, 1–17.

Rice, J., and Silverman, B. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data Are Curves," , Ser. B, 53, 233–243.

Rice, J., and Wu, C. (2000), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," , 57, 253–259.

Shi, M., Weiss, R. E., and Taylor, J. M. G. (1996), "An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves," , 45, 151–163.

Shibata, R. (1981), "An Optimal Selection of Regression Variables," , 68, 45–54.

Silverman, B. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," , 68, 45–54.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Tyer, V. R., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998), "Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast by Microarry Hybridization," , 9, 3273–3297.

Staniswalis, J. G., and Lee, J. J. (1998), "Nonparametric Regression Analysis of Longitudinal Data," , 93, 1403–1418.

Wu, C., and Chiang, C. (2000), "Kernel Smoothing on Varying Coefficient Models With Longitudinal Dependent Variable," , 10, 433–456.

Yao, F., Müller, H. G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A., and Vogel, J. S. (2003), "Shrinkage Estimation for Functional Principal Component Scores With Application to the Population Kinetics of Plasma Folate," , 59, 676–685.