September 2003

Shrinkage Estimation for Functional Principal Component Scores with Application to the Population Kinetics of Plasma Folate

Fang Yao,[1] Hans-Georg Müller,[1,*] Andrew J. Clifford,[2] Steven R. Dueker,[2] Jennifer Follett,[2] Yumei Lin,[2] Bruce A. Buchholz,[3] and John S. Vogel[3]

The main contributions of this article are, first, the proposed shrinkage estimates of the FPC scores, which improve upon the estimates obtained by the integration method, especially when the measurements are contaminated with noise, as is usually the case in practical applications. A consequence is improved prediction of individual trajectories. Second, we demonstrate with a new data set, consisting of a sample of longitudinal measurements on the kinetics of plasma folate, the usefulness of these methods for biostatistical data.

Our approach is related to that of Staniswalis and Lee (1998), who also used scatterplot smoothing to obtain mean and covariance functions, and proposed modifications to allow for additional measurement errors. We also propose an improved estimate for the variance of these errors through improved estimation in the neighborhood of and at the diagonal of the covariance surface, by fitting local quadratic components along the direction perpendicular to the diagonal. In practice, smoothing the covariance surface guarantees a symmetric but not always nonnegative-definite estimate. We implement a simple modification where we neglect the negative eigenvalues and corresponding eigenfunctions to obtain a nonnegative-definite estimate of the covariance, without changing the main characteristics of the covariance estimate. We use one-curve-leave-out cross-validation for choosing such auxiliary parameters as the degree of smoothing and the model dimension, corresponding to the number of eigenfunctions to be included, similar to a proposal of Rice and Silverman (1991).

In our application, the time courses of ^{14}C-folate in plasma were recorded for 13 healthy adults that were administered a small oral dose of ^{14}C-folic acid (80 nmol, 100 nCi). Use of labeled nutrients and drugs is common in nutritional and pharmacologic research, because it provides the only realistic way of tracking their overall fate in the body in an appropriate time frame. The time courses of plasma folate of these thirteen subjects are unknown *a priori;* this motivates the use of nonparametric methods for exploratory analysis. The numbers and locations of the time points on which measurements are available are irregular for these 13 subjects. Measurements are very dense during the first day, and then become increasingly sparse. This motivates a preprocessing step that consists of a time transformation. We found that a square-root logarithm transformation of time (days) is appropriate.

The remainder of the article is organized as follows: In Section 2, we present the FPCA model incorporating measurement errors. Section 3 contains a description of the estimation of the mean, covariance surface, and eigenfunctions. The proposed shrinkage estimates for the FPC scores are discussed in Section 4. A simulation study is included in Section 5, to demonstrate the performance of the proposed methods. The application to longitudinal folate data is described in Section 6, and concluding remarks are given in Section 7.

2. Modeling Trajectories through Functional Principal Components

We model the sample curves or trajectories as independent realizations of a stochastic process $X(t)$ that has mean $\mathrm{E}\{X(t)\} = \mu(t)$ and covariance function $\mathrm{cov}\{X(s), X(t)\} = G(s,t)$. We assume that there is an orthogonal expansion (in the L^2 sense) of G in terms of eigenfunctions ϕ_k and nonincreasing eigenvalues λ_k:

$$G(s,t$$

(t

diagonal of the raw covariances should be removed, i.e., only $\mathrm{C}_i(t_{ij},\ t_{il})$, $j \neq l$, should be included as predictors in the smoothing step (Staniswalis and Lee, 1998). We again use one-curve-leave-out cross-validation, minimizing $\mathrm{CV}(h) = \sum_{i=1}^{N}\sum_{j\neq l}\{\mathrm{C}_i(t_{ij},t_{il}) - \widehat{G}^{(-i)}(t_{ij},t_{il};h)\}^2/N$, to choose the smoothing parameter h in the surface smoothing step. Here $\widehat{G}^{(-i)}(s,t;h)$ is the smoothed covariance function obtained by removing the ith individual curve, using bandwidth h.

We note that the estimate $\widehat{G}(s,t)$ will always be symmetric if a symmetric weight function is used in the local linear smoothers, but not necessarily nonnegative-definite for finite samples, as local linear smoothers may assign negative weights in some local windows. We implement a simple modification by ignoring negative estimates of eigenvalues and the corresponding eigenfunctions in the usual expansion of the covariance function into eigenvalues/eigenfunctions. In this way, one obtains an estimate that is guaranteed to be nonnegative-definite, without changing the characteristics of the covariance estimate for larger sample sizes, where estimated eigenvalues are closer to the nonnegative true eigenvalues.

The variance σ^2 of measurement errors is of special interest for our proposed shrinkage estimates. Since the covariance of $X(t)$ is maximal along the diagonal, we expect that the shape of the surface in the direction orthogonal to the diagonal can be better approximated by a local quadratic rather than a local linear fit. Indeed, we found that the standard fitting of local planes around the diagonal leads to overestimation of σ^2. To improve the estimation in the neighborhood of and at the diagonal, we 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 easily by rotating the coordinates by 45°.

Prior to this smoothing step, the diagonal elements of the raw covariances, that is, $C_i(t_{ij},t_{ij})$, $i = 1,\ldots,N$, $j = 1,\ldots,n_i$, are removed. Denote the diagonal of the resulting surface estimate by $\tilde{G}(t)$. An estimate $\hat{V}(t)$ of $\{G(t,\ t) + \sigma^2\}$ is then obtained in a second local weighted linear smoothing step, applied to the scatterplot $\{t_{ij},\ C_i(t_{ij},\ t_{ij})\}$ with pooled raw variances $C_i(t_{ij},\ t_{ij})$, $i = 1,\ldots,N$, $j = 1,\ldots,n_i$. To mitigate against 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). The resulting estimate of σ^2, adjusted to avoid the possibility of negative estimates, is

$$\hat{\sigma}^2 = \frac{2}{T}\int_{T/4}^{3T/4}\{\hat{V}(t) - \tilde{G}(t)\}\,dt, \tag{4}$$

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

3.2 *Estimating the Eigenfunctions and Eigenvalues*

The estimates of eigenfunctions and eigenvalues correspond to the solutions $\hat{\phi}_k$ and $\hat{\lambda}_k$ of the eigenequations, $\int_0^T \widehat{G}(s,t)\hat{\phi}_k(s)\,ds = \hat{\lambda}_k\hat{\phi}_k(t)$, where $\hat{\phi}_k$ are subject to $\int_0^T \hat{\phi}_k(t)^2\,dt = 1$ and $\int_0^T \hat{\phi}_k(t)\hat{\phi}_m(t)\,dt = 0$ for $m < k$. We estimate the eigenfunctions by discretizing the smoothed covariance, as previously described in Rice and Silverman (1991) and Capra and Müller (1997). The FPC scores $\xi_{ik} = \int\{X_i(t) - \mu(t)\}\phi_k(t)\,dt$ have traditionally been estimated by numerical integration. However, the presence of additional contaminating errors, which are pervasive in practice, motivates shrinkage estimates for the FPC scores ξ_{ik} that have not yet been considered. Since the Y_{ij} are only available at discrete times t_{ij}, the integrals in the definition of the FPC scores ξ_{ik} are usually approximated by sums, substituting Y_{ij}, as defined in (3), for $X_i(t_{ij})$.

For notational convenience, we define $Y_i(t)$ as a step function with jumps at midpoints between neighboring t_{ij}, using Y_{ij} as the size of the steps, and analogously for $\epsilon_i(t)$. Set $\tilde{X}_i(t) = Y_i(t) - \epsilon_i(t)$, and let $\tilde{\xi}_{ik} = \int_0^T\{\tilde{X}_i(t) - \mu(t)\}\phi_k(t)\,dt$ be the discrete version of the FPC scores ξ_{ik}. Since neither X_i nor $\tilde{X}_i$ is available, due to the contaminating errors ϵ_i, we instead consider the approximations $\eta_{ik} = \int_0^T\{Y_i(t) - \mu(t)\}\phi_k(t)\,dt$. In practice, μ and ϕ_k are also unknown and must be estimated from the data. It is common practice, using estimates $\hat{\mu}(t_{ij})$ for $\mu(t_{ij})$ and $\hat{\phi}_k(t_{ij})$ for $\phi_k(t_{ij})$, to estimate η_{ik} by approximating sums, letting $t_{i0} = 0$,

$$\begin{aligned}\hat{\eta}_{ik} &= \int_0^T\{Y_i(t) - \hat{\mu}(t)\}\hat{\phi}_k(t)\,dt \\ &\approx \sum_{j=1}^{n_i}\{Y_{ij} - \hat{\mu}(t_{ij})\}\hat{\phi}_k(t_{ij})(t_{ij} - t_{i,j-1}).\end{aligned} \tag{5}$$

To choose the number of principal component curves that will provide a reasonable approximation to the infinite-dimensional process, we use the cross-validation score based on the one-curve-leave-out prediction error. Let $\hat{\mu}_i^{(-i)}$ and $\hat{\phi}_k^{(-i)}$ be the estimated mean function and eigenfunctions, respectively, after removing the ith individual's curve. Then, we choose the number of components K to be included in the model so as to minimize the cross-validation scores

$$\mathrm{CV}(K) = \frac{1}{N}\sum_{i=1}^{N}\left\|Y_i - \widehat{Y}_i^{(-i)}\right\|^2, \tag{6}$$

where $\|\cdot\|$ here and in the following is defined by approximating the usual L^2 distance through $\|Y_i - \widehat{Y}_i^{(-i)}\|^2 = \sum_{j=1}^{n_i}\{Y_{ij} - \widehat{Y}_i^{(-i)}(t_{ij})\}^2(t_{ij} - t_{i,j-1})$. Note that $\widehat{Y}_i^{(-i)}$ is the predicted curve for the ith subject after removing this sample curve from fitting model (3), i.e., $\widehat{Y}_i^{(-i)}(t) = \hat{\mu}^{(-i)}(t) + \sum_{k=1}^{K}\hat{\eta}_{ik}^{(-i)}\hat{\phi}_k^{(-i)}(t)$.

4. Shrinkage Estimation of Functional Principal Component Scores

We reconsider the estimation of the functional principal components in model (3). Setting $\tilde{\epsilon}_{ik} = \int_0^T \epsilon_i(t)\phi_k(t)\,dt$, we note that $\eta_{ik} = \tilde{\xi}_{ik} + \tilde{\epsilon}_{ik}$. We conclude that the best linear predictors of the approximate FPC scores $\tilde{\xi}_{ik}$ are not given by η_{ik}, but rather by

$$\frac{\tilde{\lambda}_k}{\tilde{\lambda}_k + \mathrm{var}(\tilde{\epsilon}_{ik})}\eta_{ik}, \tag{7}$$

where $\tilde{\lambda}_k = \mathrm{var}(\tilde{\xi}_{ik})$. This is a shrinkage formula, moving η_{ik} closer towards the origin, whenever $\mathrm{var}(\tilde{\epsilon}_{ik}) > 0$. If we assume that ξ_{ik} and ϵ_{ij} and also that the discrete versions $\tilde{\xi}_{ik}$ and $\tilde{\epsilon}_{ik}$ are independent and jointly Gaussian, the best predictors of the approximate FPC scores $\tilde{\xi}_{ik}$ are $\mathrm{E}(\tilde{\xi}_{ik}\,|\,\eta_{ik}) = \tilde{\lambda}_k\eta_{ik}/\{\tilde{\lambda}_k + \mathrm{var}(\tilde{\epsilon}_{ik})\}$, corresponding to predictors (7). The

variation of the shrinkage estimators is reduced, observing that $\mathrm{var}(\eta_{ik}) = \mathrm{var}\{\mathrm{E}(\tilde{\xi}_{ik} \mid \eta_{ik})\} + \mathrm{E}\{\mathrm{var}(\tilde{\xi}_{ik} \mid \eta_{ik})\} > \mathrm{var}\{\mathrm{E}(\tilde{\xi}_{ik} \mid \eta_{ik})\}$, if $\mathrm{E}\{\mathrm{var}(\tilde{\xi}_{ik} \mid \eta_{ik})\} > 0$.

The quantities $\tilde{\epsilon}_{ik}$ can be approximated by the sums $\sum_{j=1}^{n_i} \epsilon_{ij}\phi_k(t_{ij})(t_{ij} - t_{i,j-1})$. Note that $\mathrm{var}(\tilde{\xi}_{ik}) \approx \sigma^2 \sum_{j=1}^{n_i} \phi_k^2 \times (t_{ij})(t_{ij} - t_{i,j-1})^2$, and if the number and density of measurements are sufficiently large, $\mathrm{var}(\tilde{\epsilon}_{ik})$ can be approximated by $\mathrm{T}\sigma^2/n_i$, where n_i is the number of repeated measurements for the ith subject. A reasonable estimate of the shrinkage factor (7) is therefore

$$\hat{\xi}_{ik} = \frac{\hat{\lambda}_k}{\hat{\lambda}_k + \mathrm{T}\hat{\sigma}^2/n_i}\hat{\eta}_{ik}, \tag{8}$$

where $\hat{\sigma}^2$ is the sample variance estimate obtained from (4), $\hat{\lambda}_k$ is the estimated kth eigenvalue, and $\hat{\eta}_{ik}$ is the raw sample estimate for the kth FPC score of the ith individual given in (5). In what follows, we refer to (8) as Gaussian shrinkage, since it provides the appropriate shrinkage factor for the Gaussian case.

Since the presence of additional measurement errors is a main motivation of shrinkage estimation, stochastic variation in estimating the variance of the measurement errors can be expected to have an impact on the shrinkage estimates. Neglecting the variation in the estimation of eigenvalues λ_k, we note that the shrinkage factor $\lambda/\{\lambda + (\mathrm{T}\sigma^2/n_i)\}$ is a convex function of σ^2. Assume a multiplicative error model for the behavior of the variance estimate $\hat{\sigma}^2$, $\hat{\sigma}^2 = \varepsilon\mathrm{E}(\hat{\sigma}^2)$, where $\varepsilon > 0$ is a r.v. with $\mathrm{E}(\varepsilon) = 1$. Then, the targeted shrinkage factor is $\lambda/[\lambda + \{\mathrm{TE}(\hat{\sigma}^2)/n_i\}]$, and from Jensen's inequality, we have

$$\mathrm{E}\left\{\lambda \Big/ \left(\lambda + \frac{\mathrm{T}\hat{\sigma}^2}{n_i}\right)\right\} > \lambda \Big/ \left(\lambda + \frac{\mathrm{TE}(\hat{\sigma}^2)}{n_i}\right). \tag{9}$$

We see from (9) that the appropriate shrinkage factor is smaller than that given by (8), on the average, under reasonable assumptions about randomness in the error variance. This observation motivates a generalized shrinkage method, where we replace the Gaussian shrinkage factor $\lambda/\{\lambda + (\mathrm{T}\sigma^2/n_i)\}$ by a generalized shrinkage factor $\lambda/\{\lambda + (\rho/n_i)\}$ for an unknown ρ. To achieve the target $\lambda/[\lambda + \{\mathrm{TE}(\hat{\sigma}^2)/n_i\}]$ in the random error variance, we see that ρ should be greater than $\mathrm{T}\hat{\sigma}^2$ on average, which means more shrinkage is needed than provided by the Gaussian shrinkage method.

A second motivation for the generalized shrinkage method is minimization of the squared prediction error over a class of linear shrinkage factors. Let $\hat{X}_i(\rho)$ be the prediction for the true process X_i obtained by using the linear shrinkage estimates $\hat{\xi}_{ik} = \hat{\lambda}_k\hat{\eta}_{ik}/\{\hat{\lambda}_k + (\rho/n_i)\}$. The orthonormality of the eigenfunctions (analogous to the derivation of Parseval's equality; see Courant and Hilbert (1953)) leads to $\|\hat{X}_i(\rho) - X_i\|^2 = \sum_k[\hat{\lambda}_k\hat{\eta}_{ik}/\{\hat{\lambda}_k + (\rho/n_i)\} - \xi_{ik}]^2$. Setting $Z(\rho) = 1/N\sum_{i=1}^N \|\hat{X}_i(\rho) - X_i\|^2$, $f(Z) = \min_\rho Z(\rho)$, the best linear predictor property of Gaussian shrinkage implies that $\rho = \mathrm{T}\hat{\sigma}^2$ is an approximate minimizer of $\mathrm{E}\{Z(\rho)\}$. Due to the concavity of f, using Jensen's inequality, we find that

$$\mathrm{E}\{\min_\rho Z(\rho)\} \le \min_\rho \mathrm{E}\{Z(\rho)\} \approx \mathrm{E}\{Z(\mathrm{T}\hat{\sigma}^2)\}. \tag{10}$$

Therefore, on average, generalized shrinkage which corresponds to minimizing $Z(\rho)$ leads to an improved estimate as compared to Gaussian shrinkage. For large sample sizes, $Z(\rho)$ will be closer to $\mathrm{E}\{Z(\rho)\}$ and therefore the relative gain of generalized over Gaussian shrinkage is predicted to be more pronounced for smaller samples by this argument. Indeed, this is what we found in simulations (Section 5).

The initial form of the generalized shrinkage formula is therefore given by $\hat{\xi}_{ik} = [\hat{\lambda}_k/\{\hat{\lambda}_k + (\rho/n_i)\}]\hat{\eta}_{ik}$, where $\rho > 0$ is an unspecified shrinkage parameter. We note that Gaussian shrinkage (8) is a special case with $\rho = \mathrm{T}\hat{\sigma}^2$. The generalized shrinkage parameter ρ is chosen to minimize the estimated prediction error, where the estimated prediction error is obtained by one-curve-leave-out cross-validation. Let $\hat{Y}_i^{(-i)}(\cdot, \rho)$ be the predicted curve for the ith subject, using shrinkage parameter ρ and removing the data of the ith subject. We then minimize the cross-validated integrated prediction error, with respect to $\rho \ge 0$, to obtain

$$\hat{\rho} = \arg\min_{\rho \ge 0} \frac{1}{N}\sum_{i=1}^N \left\|Y_i - \hat{Y}_i^{(-i)}(\rho)\right\|^2, \tag{11}$$

leading to the generalized shrinkage formula

$$\hat{\xi}_{ik} = \frac{\hat{\lambda}_k}{\hat{\lambda}_k + \hat{\rho}/n_i}\hat{\eta}_{ik}. \tag{12}$$

The generalized shrinkage method leads to the best possible shrinkage estimates in this class, in terms of prediction error. Thus, in general, it will lead to improvements over Gaussian shrinkage (8), according to (10). Gaussian shrinkage in turn leads to improvements over the customary FPC scores $\hat{\eta}_{ik}$, which are motivated by approximating the integrals that define the FPC scores (see Section 3.2). The customary FPC scores $\hat{\eta}_{ik}$ will nearly always overestimate the FPC scores that are optimal for prediction. Gaussian shrinkage, while improving on this situation, often will not produce enough shrinkage, especially in small sample situations.

5. Simulation Results

To illustrate the advantage of shrinkage estimation of FPC scores compared to customary estimation without shrinkage, we devised a simulation study, using 100 i.i.d. normal and 100 i.i.d nonnormal samples consisting of $N = 10$, 20, and 50 random trajectories. The data were generated following model (3). The simulated process had mean function $\mu(t) = t + \sin(2\pi t)$, $0 \le t \le 1$. We construct the covariance function of the process according to (1) from two orthonormal functions $\phi_1(t) = -2^{1/2}\cos(\pi t)$ and $\phi_2(t) = 2^{1/2}\sin(\pi t)$, $0 \le t \le 1$. For the ith partially observed path in a single sample, the number of observations, n_i, was randomly chosen to be between 30 to 40. For an equally spaced grid $\{\mathrm{c}_1, \ldots, \mathrm{c}_{n_i}\}$ on $[0, 1]$ with $c_1 = 0$, $c_{n_i} = 1$, $d = 1/(n_i - 1)$, the t_{ij} were uniform on $[c_j - d/2, c_j + d/2]$, for $j = 2, \ldots, n_i - 1$, t_{i1} uniform on $[0, d/2]$, and t_{in_i} uniform on $[1 - d/2, 1]$, allowing for nonequidistant "jittered" designs. We chose $\lambda_1 = 2$ and $\lambda_2 = 1$ as the eigenvalues, $\lambda_k = 0$, $j \ge 3$, and $\sigma^2 = 0.25$ as the variance of the additional measurement errors ϵ_{ij} in (3), which were assumed to be normal with mean 0. Note that $\mathrm{T} = 1$.

Table 1

Results of 100 *Monte Carlo runs with* $N = 10$, 20, and 50 *trajectories per sample. Shown are averages of estimates* $\rho = \hat{\sigma}^2$ (4) *for Gaussian* (8) *and of estimates* $\rho = \hat{\rho}$ (11) *for generalized* (12) *shrinkage, cross-validation scores* (11)*, squared prediction errors* (13) *and average squared errors for the two functional principal component scores* ξ_1 *and* ξ_2.

Sample size	Model shrinkage type	Normal			Mixture		
		None	Gaussian ($\rho = \hat{\sigma}^2$)	Generalized ($\rho = \hat{\rho}$)	None	Gaussian ($\rho = \hat{\sigma}^2$)	Generalized ($\rho = \hat{\rho}$)
$N = 10$	True σ^2	—	0.250	—	—	0.250	—
	Ave. ρ	—	0.273	0.945	—	0.271	1.34
	Ave. CV	1.35	1.24	1.19	1.32	1.26	1.16
	Ave. SPE	0.971	0.909	0.873	0.964	0.908	0.862
	ASE(ξ_1)	0.386	0.361	0.354	0.371	0.353	0.338
	ASE(ξ_2)	0.442	0.414	0.398	0.453	0.438	0.409
$N = 20$	Ave. ρ	—	0.261	0.881	—	0.267	1.15
	Ave. CV	1.14	1.06	1.03	1.12	1.04	0.985
	Ave. SPE	0.866	0.814	0.793	0.910	0.876	0.831
	ASE(ξ_1)	0.352	0.334	0.327	0.336	0.328	0.315
	ASE(ξ_2)	0.417	0.396	0.389	0.425	0.409	0.384
$N = 50$	Ave. ρ	—	0.257	0.421	—	0.259	0.578
	Ave. CV	0.961	0.893	0.887	0.948	0.887	0.875
	Ave. SPE	0.812	0.767	0.762	0.830	0.781	0.774
	ASE(ξ_1)	0.343	0.329	0.319	0.331	0.320	0.308
	ASE(ξ_2)	0.406	0.387	0.371	0.408	0.398	0.381

For the 100 normal samples, the FPC scores ξ_{ik} were generated from $\mathcal{N}(0, \lambda_k)$, while ξ_{ik} in the nonnormal samples were generated from a mixture of two normals, $\mathcal{N}\{(\lambda_k/2)^{1/2}, \lambda_k/2\}$ with probability 1/2 and $\mathcal{N}\{-(\lambda_k/2)^{1/2}, \lambda_k/2\}$ with probability 1/2. We used the cross-validation procedures described earlier for selection of bandwidths, number of eigenfunctions, and generalized shrinkage parameter $\hat{\rho}$ (11).

To demonstrate the performance of the proposed shrinkage estimates, we report in Table 1 the averages of the estimates of the variances $\hat{\sigma}^2$ in (4), the mean shrinkage factor estimates for $\rho = \hat{\sigma}^2$ in the case of Gaussian shrinkage (8), and for $\rho = \hat{\rho}$, in the case of generalized shrinkage (12), and also cross-validation scores in (11) and squared prediction errors

$$\mathrm{SPE}(\rho) = \frac{1}{N}\sum_{i=1}^{N} \|Y_i - \hat{Y}_i(\rho)\|^2. \tag{13}$$

We find that on average, the generalized shrinkage (12) leads to larger downsizing of the raw estimators than Gaussian shrinkage (8) in both normal and mixture distribution situations.

For sample sizes $N = 10/20/50$, and underlying normal distribution, Gaussian shrinkage results in decreases in cross-validation (CV) of about 9%/8%/7% and in SPE of about 7%/6%/6%, while for generalized shrinkage, the corresponding decreases are 12%/10%/8% and 10%/8%/6%. Just looking at decreases in SPE for the mixture distribution, Gaussian shrinkage leads to decreases of 6%/4%/6% for $N = 10/20/50$, and generalized shrinkage to corresponding decreases of 11%/9%/7%. We draw the following conclusions: The gains are largest for small samples, and are always larger for generalized shrinkage, as compared to Gaussian shrinkage. The differences in gains between the two shrinkage methods decrease as the sample size gets larger, as was predicted in the discussion following (10) in Section 4. The generalized shrinkage parameter $\hat{\rho}$ also is seen to move closer to $\hat{\sigma}^2$ with increasing sample size. We find that the gains do not depend much on the nature of the distribution of the principal components.

Another outcome measure of interest is the average squared error for the two FPC scores, $\mathrm{ASE}(\xi_k) = \sum_{i=1}^{N}(\hat{\xi}_{ik} - \xi_k)^2/N, k = 1, 2$, also listed in Table 1. These errors show similar behavior to the prediction errors. With regard to all measures, the gains obtained from shrinkage as compared to no shrinkage remain substantial, even for large sample sizes. Among the two shrinkage methods considered, generalized shrinkage uniformly achieves additional gains, which are particularly pronounced for small sample sizes.

6. Application to Longitudinal Plasma Folate Data

6.1 ^{14}C-*Folate Specific Activity*

In an experiment conducted at the University of California-Davis, repeated measurements of the fraction of labeled folate among total folate in plasma were obtained for 13 healthy adult volunteers. Measurements were labeled by time since the volunteers orally ingested a small tracer dose of ^{14}C-folic acid. The fraction of ^{14}C-folate-among total folate in plasma (the so-called plasma ^{14}C-folate-specific activity) was measured in about 20 plasma specimens drawn during the first day after dosing. After the first day, the drawing of additional blood specimens became less frequent, with a total of about 50 specimens being taken during the 200-day period after dosing.

Our main interest is characterizing the dynamic behavior of plasma ^{14}C-folate-specific activity in healthy adults. Methods for the collection, processing, and laboratory analyses of specimens were as described in Clifford et al. (1998). The plot of the plasma ^{14}C-folate-specific activity in the first four subjects versus time (after dosing) at which these measurements

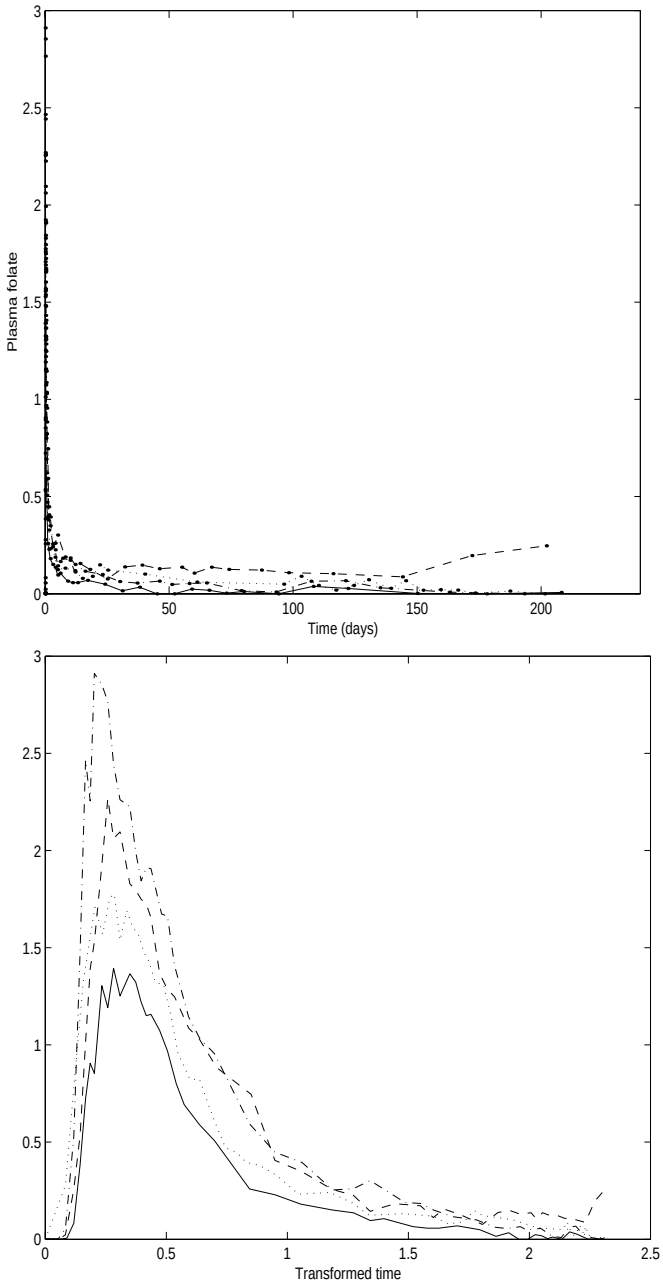


Figure 1. Top: Observed individual trajectories for the first four subjects, 1 (solid), 2 (dashed), 3 (dashdot), 4 (dotted), in the original time scale. Bottom: The same four individual trajectories shown in the transformed time scale.

were taken is shown in the top panel of Figure 1; it reveals some common patterns that are described in Section 6.4.

6.2 *Time Transformation to Address Sparseness*

We note that the numbers and locations of the time points at which measurements were taken are not the same for these 13 subjects. Measurements are very dense during the first day, and then become increasingly sparse. From the top panel of Figure 1, using the original time scale, it is not easy to observe the shapes of the curves. Moreover, because we use global bandwidths in the smoothing steps for mean function and covariance surface, the increasing time lag between the measurements for individual subjects leads to oversmoothing when choosing an overall appropriate smoothing parameter.

Our solution to this problem is preprocessing of the data with the time transformation $t' = \{\log\,(1+t)\}^{1/2}$. In fact, we obtain a reasonably even distribution of measurement times after this transformation. Folate measurements versus transformed time are shown in the bottom panel of Figure 1 for the first four subjects. The transformed time scale is more conducive for observing the shapes of the curves and for data smoothing. For the transformed time scale, $\mathrm{T} = 2.31$.

6.3 *Diagnostics for Trajectories and Outlier Detection*

Initially we applied our methods to the plasma folate data of the entire sample consisting of 13 healthy adult volunteers. Whether the trajectories recorded for the subjects come from the same population, and whether there are outliers is of biological interest. To address such questions, we desire quantities that are analogous to residuals or deviances, but apply to entire trajectories rather than traditional scatterplots. As a summary measure, we consider the integrated squared residual error in the original time scale. This criterion reflects the ability of the model to predict an observed sample curve.

Let $Y_i(t)$ be the step function determined by Y_{ij} in the original time scale for the ith subject, as described in Section 3.2; let $\hat{Y}_i(t)$ be the predicted curve and $[0,\, T']$ the original time range for all subjects. We define the integrated squared residual error for the ith subject by

$$\mathrm{RE}_i = \int_0^{T'} \{Y_i(s) - \hat{Y}_i(s)\}^2\, ds. \tag{14}$$

A large value of RE_i, $i = 1, \ldots, N$, may point to an outlying sample curve that is poorly fitted by the model.

We calculated these integrated squared residual errors for all 13 subjects, and found that $\mathrm{RE}_2 = 1.6$ is much larger than $\mathrm{RE}_i \leq 0.55$, $i \neq 2$. This provides some evidence that this subject is poorly fitted by the proposed model, and may be an outlier. The predicted curve and observed values of subject 2 in the original time scale are shown in Figure 2. The predicted curve is clearly underestimating the observed data after about 25 days, especially in the right tail, while the predicted curves for the other subjects are fairly close to the observations. Indeed, it was found that subject 2 had been exposed to ^{14}C earlier through participation in an unrelated medical trial several years prior to the present study. This fact would have disqualified this subject from participating in the present study, but had not been known. This finding attests not only to the astonishing accuracy of the ^{14}C detection methods used in this study, but also to the usefulness of the FPC approach for data screening and outlier detection. Since subject 2 was thus confirmed to be a clear outlier, the data of subject 2 were not used in the subsequent analysis. Using the integrated squared residual error RE_i in the transformed time scale leads to the same conclusion.

6.4 *Mean and Covariance Functions*

The estimated mean curve obtained using local weighted linear smoothing for the remaining 12 subjects is shown in the

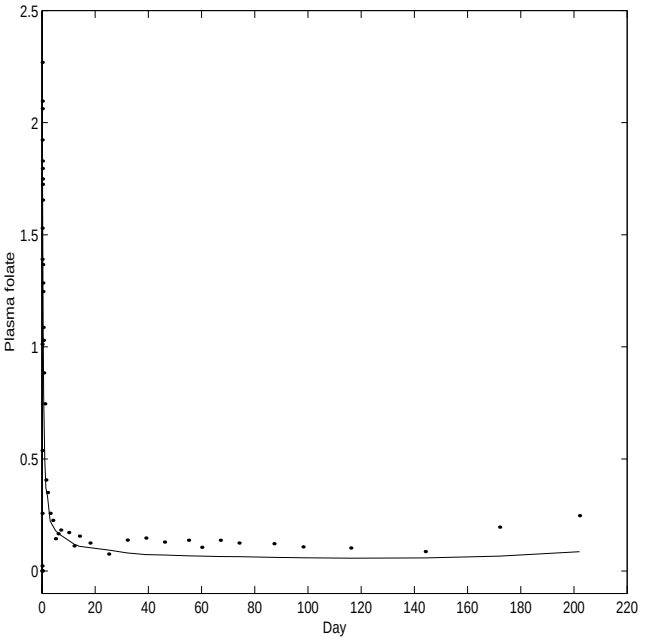


Figure 2. Observed values (dots) and predictions (solid) for subject 2 in the original time scale.

top panel of Figure 3. The bandwidth $b = 0.11$ was chosen, guided by cross-validation. The estimated mean curve reflects the overall trend of the individual curves. One notes a short slow increase (about 4 minutes in the original time scale) in the appearance of ^{14}C-folate in plasma. This delay corresponds to the time needed for the dose to reach and be absorbed from the small intestine. Then, the mean plasma ^{14}C-folate specific activity by time since dosing exhibits a sharp initial rise to a peak at about 2 hours (in the original time frame) as the administered dose enters the blood plasma pool. Then, it drops at a decreasing rate as it enters cells where the ^{14}C-folate is sequestered and converted to other chemical forms of folate. Finally, the pattern changes into a smooth, slow decline toward zero.

The estimated covariance surface obtained using local weighted linear smoothing after the diagonal was removed is shown in the bottom panel of Figure 3. The bandwidth was $h = (0.15, 0.15)$, chosen by one-curve-leave-out cross-validation. We note that the two smallest eigenvalues of this covariance estimate were found to be negative. To obtain a nonnegative-definite covariance surface, we implemented the simple modification described in Section 3.1, namely, omitting the components with negative eigenvalues. The covariance surface shows high variability at the beginning, with a sharp increase followed by a rapid decline. The estimate $\hat{\sigma}^2$ (4) for the variance σ^2 of the measurement error was $\hat{\sigma}^2 = 0.0037$.

6.5 *Eigenfunctions*

The smooth estimates of the first three eigenfunctions are presented in Figure 4 (in transformed time scale). The cross-validation scores level off when more than three eigenfunctions are used, suggesting that the first three principal component curves are sufficient to describe the modes of variability. These three principal component functions explain about 98.5% of the total variation.

The first eigenfunction indicates that a large portion of the variability between subjects is roughly in the direction of the amplitude of the mean curve, as the first eigenfunction has a similar shape to the mean function. In particular, 89.48% of the total variability is "explained" by the first eigenfunction, which indicates that this mode of variation is dominant. The second eigenfunction takes the shape of an approximate

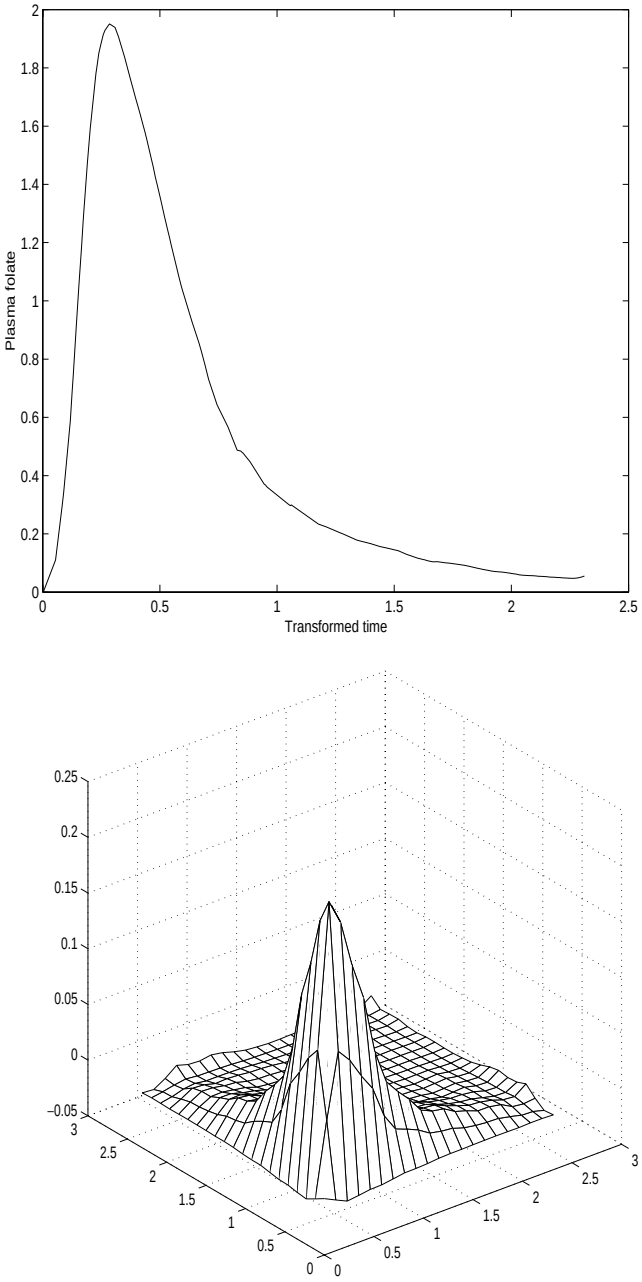
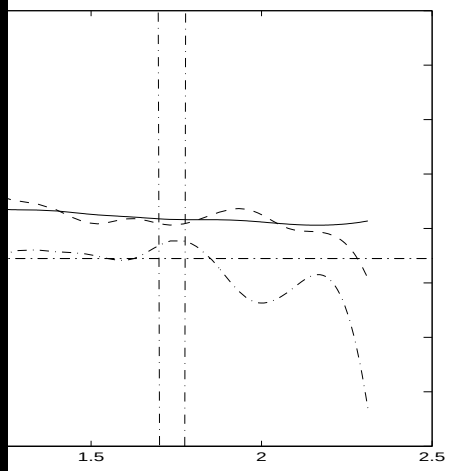
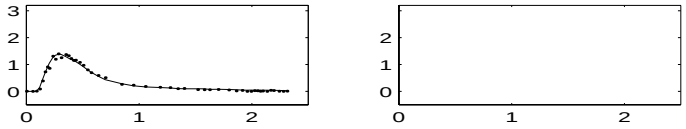


Figure 3. Top: Smooth estimate of the mean function (transformed time scale). Bottom: Smooth estimate of the covariance surface (transformed time scale).


1.5
2
2.5


3
2
1
0
0
1
2
2
1
0
0
1
2

de ce modèle. Nous proposons d'améliorer l'estimation au voisinage et sur la diagonale de la surface de covariance où les erreurs de mesures sont réfléchies. La présence d'erreurs de mesure additives justifie les estimations par réduction pour les scores des composantes principales fonctionnelles. Les estimation par réduction sont développées par meilleure prédiction linéaire, et dans une version généralisée visent à minimiser l'erreur de prédiction par écart-d'une-courbe (one-curve-leave-out). L'estimation d'une trajectoire individuelle combine les données de l'individu avec celles des autres individus. Nous appliquons nos méthodes à des données relatives à l'analyse du niveau de folates marqué au ^{14}C comme fonction du temps pour des adultes sains avec une petite dose de traçage d'acide folique marqué au ^{14}C. On a introduit une transformation sur le temps pour tenir compte de l'irrégularité des temps auxquels les mesures ont été prises. La méthodologie proposée, incorporant la réduction et des propriétés adaptées aux données apparaît bien adaptée pour décrire les cinétiques de population pour l'activité spécifique des folates marqués au ^{14}C, ainsi que les effets aléatoires, et peut être aussi appliquée à d'autres problèmes d'analyse de données fonctionnelles.

Berkey, C. S. and Kent, R. L., Jr. (1983). Longitudinal principal components and non-linear regression models of early childhood growth. *Annals of Human Biology* **10,** 523–536.

Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika* **51,** 285–311.

Capra, W. B. and Müller, H. G. (1997). An accelerated-time model for response curves. *Journal of the American Statistical Association* **92,** 72–83.

Castro, P. E., Lawton, W. H., and Sylvestre, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28,** 329–337.

Clifford, A. J., Arjomand, A., Dueker, S. R., Schneider, P. D., Buchholz, B. A., and Vogel, J. S. (1998). The dynamics of folate acid metabolism in an adult given a small tracer dose of ^{14}C-folic acid. *Advances in Experimental Medicine and Biology* **445,** 239–251.

Courant, R. and Hilbert, D. (1953). *Methods of Mathematical Physics*. New York: Interscience.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Application*. London: Chapman and Hall.

Gasser, T., Müller, H. G., Köhler, W., Molinari, L., and Prader, A. (1984). Nonparametric regression analysis of growth curves. *Annals of Statistics* **12,** 210–229.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* **95,** 520–534.

Müller, H. G. and Prewitt, K. (1993). Multiparameter bandwidth processes and adaptive surface smoothing. *Journal of Multivariate Analysis* **47,** 1–21.

Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. New York: Springer.

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

Rice, J. and Wu, C. (2000). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57,** 253–259.

Silverman, B. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics* **24,** 1–24.

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **93,** 1403–1418.

Received July 2002. *Revised February* 2003.
Accepted February 2003.