Functional Variance Processes

Hans-Georg MÜLLER, Ulrich STADTMÜLLER, and Fang YAO

We introduce the notion of a functional variance process to quantify variation in functional data. The functional data are modeled as samples of smooth random trajectories observed under additive noise. The noise is assumed to be composed of white noise and a smooth random process—the functional variance process—which gives rise to smooth random trajectories of variance. The functional variance process is a tool for analyzing stochastic time trends in noise variance. As a smooth random process, it can be characterized by the eigenfunctions and eigenvalues of its autocovariance operator. We develop methods to estimate these characteristics from the data, applying concepts from functional data analysis to the residuals obtained after an initial smoothing step. Asymptotic justifications for the proposed estimates are provided. The proposed functional variance process extends the concept of a variance function, an established tool in nonparametric and semiparametric regression analysis, to the case of functional data. We demonstrate that functional variance processes offer a novel data analysis technique that leads to relevant findings in applications, ranging from a seismic discrimination problem to the analysis of noisy reproductive trajectories in evolutionary biology.

KEY WORDS: Eigenfunction; Functional data analysis; Principal component; Random trajectory; Variance function.

1. INTRODUCTION

The need to model locally changing variances has long been recognized in nonparametric regression, generalized linear modeling, and the analysis of volatility. In these settings, a variance function is invoked to quantify heteroscedasticity and to achieve efficient estimation. Often variance functions are assumed to follow a parametric form, for example, in generalized linear models or quasi-likelihood models (Wedderburn 1974), where the variance is considered a known function of the mean. In other settings, such as quasi-likelihood regression models (Chiou and Müller 1999), the variance function is assumed to be a smooth but otherwise unspecified function. Variance functions play a role in semiparametric regression models (Müller and Zhao 1995), and their applications include residual analysis (Gasser, Sroka, and Jennen-Steinmetz 1986), construction of local confidence intervals under heteroscedasticity and local bandwidth selection (Müller and Stadtmüller 1987), and, more generally, statistical model building (Eubank and Thomas 1993). There now exists a sizeable literature on the nonparametric analysis of variance functions that includes work by Dette and Munk (1998), Fan and Yao (1998), Yao and Tong (2000), and Yu and Jones (2004), among others.

In nonparametric variance function estimation, it is assumed that observed data scatter randomly around a fixed regression function. The variance function then pertains to the variance of errors that are added to a smooth mean regression function g,

$$Y_j = g(t_j) + e_j(t_j), \qquad j = 1, \ldots, J.$$

Here $(t_j)_{j=1,\ldots,J}$ is a grid of design points and $v(t_j) = \mathrm{var}(e_j(t_j))$ is the variance function that typically is assumed to be smooth. If the predictors are random and the sample consists of bivariate data (X, Y), then the variance function is defined alternatively as $v(x) = E(Y^2|X = x) - [E(Y|X = x)]^2$.

Although the variance function traditionally is considered a nonrandom object targeted by function estimation methods such as kernel or spline smoothing, increasingly data of a more complex functional type are collected, and the goal is statistical analysis for a sample of observed random trajectories. Goals of analyzing this type of high-dimensional data include defining the characteristics of a given sample of curves, finding clusters of similar subgroups, and discriminating between different types of trajectories. An excellent overview on functional data analysis has been provided by Ramsay and Silverman (2002, 2005). In this article we aim to extend the concept of a variance function to a random variance process that appropriately reflects and quantifies the variation observed in functional data.

Our study is motivated by a discrimination problem in seismology that has been described by Shumway (2002) (see also Kakizawa, Shumway, and Tanaguchi 1998). The available data correspond to time courses of seismic activity as recorded in array stations, and a major goal is to infer the type of seismic event that caused the activity. There are two possibilities, explosion and earthquake. Typical examples of recorded activity for earthquakes and explosions are shown in Figure 1. Analysis of such data traditionally has been the domain of time series methodology. We add a new angle by approaching this problem within the framework of functional data analysis. This is feasible because the data consist of repeated realizations of time courses of seismic activity. Although discriminant analysis for functional data focusing on information contained in smooth random trajectories has been described in work by Hall, Poskitt, and Presnell (2001) and generally can be based on functional principal component scores, scrutinizing the time courses in Figure 1 indicates that relevant information is contained in locally varying patterns of variation rather than smooth signal trajectories. Aiming to quantify this random variability motivates us to introduce the concept of a functional variance process.

Because each recorded trajectory is a random process, the notion of a variance function, as described earlier, is not sufficient to quantify the locally varying variation of each individual random trajectory, which in itself is a random phenomenon. Therefore, for these and other data analysis problems involving curve data with a potentially informative variation structure, an extension of the usual modeling approaches currently available for functional data analysis is needed. We are aiming at a model

Hans-Georg Müller is Professor, Department of Statistics, University of California, Davis, CA 95616 (E-mail: *mueller@wald.ucdavis.edu*). Ulrich Stadtmüller is Professor, Department of Mathematics, University of Ulm, 89069 Ulm, Germany (E-mail: *stamue@mathematik.uni-ulm.de*). Fang Yao is Assistant Professor, Department of Statistics, Colorado State University, Fort Collins, CO 80523 (E-mail: *fyao@stat.colostate.edu*). The authors thank the editor and associate editor for their handling of the manuscript and two referees for very helpful remarks. Very special thanks are due to a referee for the excellent suggestion to simplify the original approach by introducing a presmoothing step. This research was supported in part by National Science Foundation grants DMS-03-54448 and DMS-05-05537.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000186

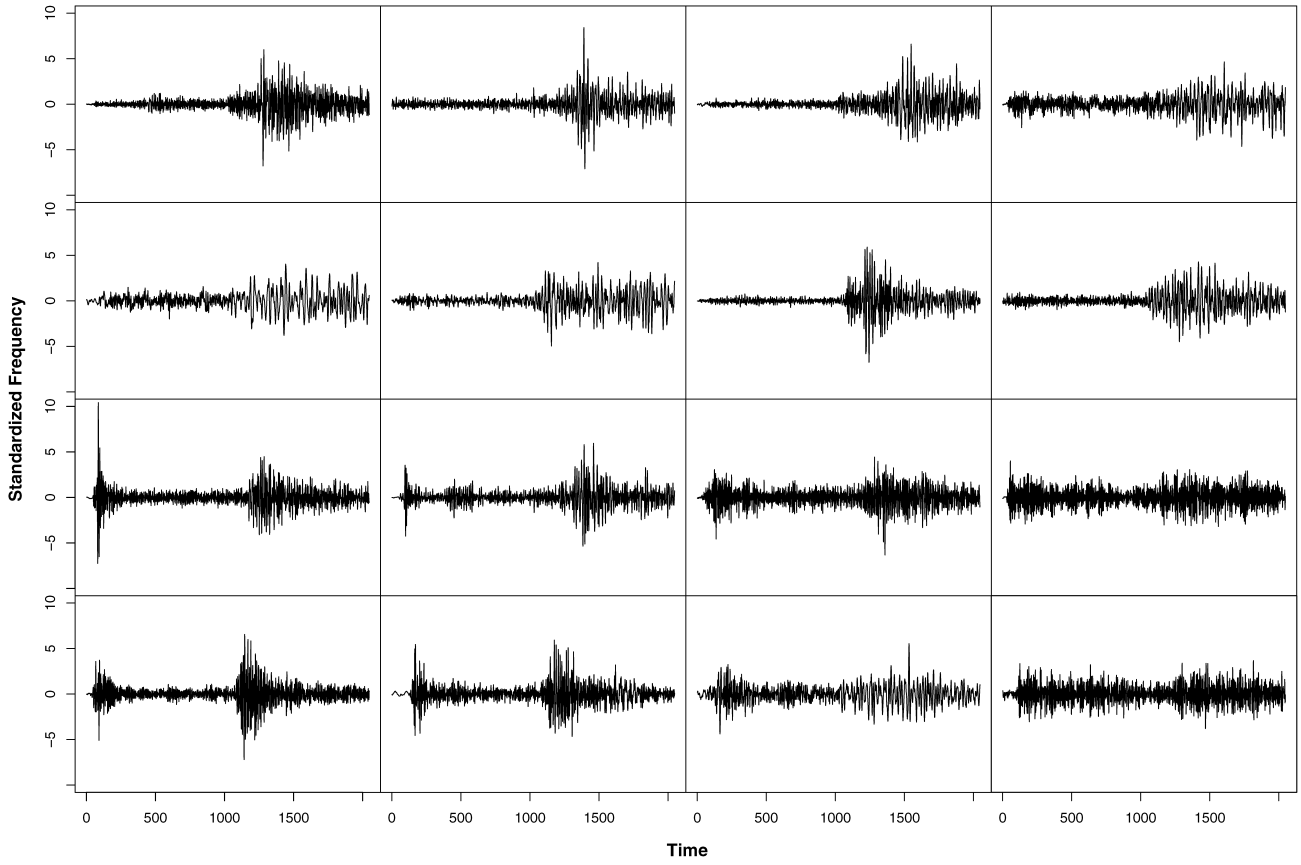

Figure 1. Data for Eight Standardized Explosions (first two rows), Seven Standardized Earthquakes (last two rows except for the bottom right panel), and One Unknown Event (bottom right panel). The first earthquake out of eight earthquakes is not shown. Time unit is .025 seconds.

that includes random components for variation. In this article we propose such an extension and demonstrate its usefulness for applications. We show that functional variance processes lead to sensible procedures for the seismic discrimination problem compared with other approaches of functional discriminant analysis and manifest themselves in random trajectories that quantify variation. One trajectory, corresponding to a realization of the variance process, is associated with each realization of the underlying random process, as shown in Figure 1. Functional variance processes generate smooth trajectories and jointly with pure noise components determine the additive errors in the discretely observed data.

Functional principal component (FPC) analysis is a major tool for the proposed development. FPC provides dimension reduction for functional data, where an eigenfunction base is used to parsimoniously describe observed random trajectories in terms of a number of random components, the FPC scores. The eigenfunctions or principal component functions are orthonormal functions that have been interpreted as the modes of variation of functional data (Castro, Lawton, and Sylvestre 1986). Early work on this concept was done by Grenander (1950) and Rao (1958), and lately it has assumed a central role in functional data analysis (Rice and Silverman 1991; Jones and Rice 1992; Ramsay and Silverman 2005; James, Hastie, and Sugar 2001; Yao et al. 2003).

The basic decomposition of the noise in the data that defines functional variance processes is presented in Section 2. Estimation of the characteristic eigenfunctions and eigenvalues of functional variance processes is described in Section 3, where estimates of individual trajectories of functional variance processes are also introduced. Section 4 is devoted to asymptotic results on the consistency of estimated residuals (Thm. 1), providing the basis for constructing trajectories of functional variance processes and convergence of estimated eigenfunctions and eigenvalues (Thm. 2), as well as convergence of individual estimated trajectories (Thm. 3) of the functional variance process.

Applications of the functional variance process technique to recorded seismic geophysical and reproductive biological random trajectories are the theme of Section 5, followed by concluding remarks. Details about estimation procedures are compiled in Appendix A, assumptions and notations as needed for the proofs are given in Appendix B, and proofs and auxiliary results are provided in Appendix C.

2. DECOMPOSING FUNCTIONAL DATA

The observed data are decomposed into a smooth process S that is sampled on a discrete dense grid and additive noise. The noise is assumed to be generated by the smooth functional variance process V and an independent white noise component. Individual trajectories of the functional variance process are modeled through the corresponding FPC scores and eigenfunctions.

The data are generated from a square-integrable process S that produces a sample of n iid smooth random trajectories S_i, $i = 1, \ldots, n$. The observed measurements X_{ij} are available on

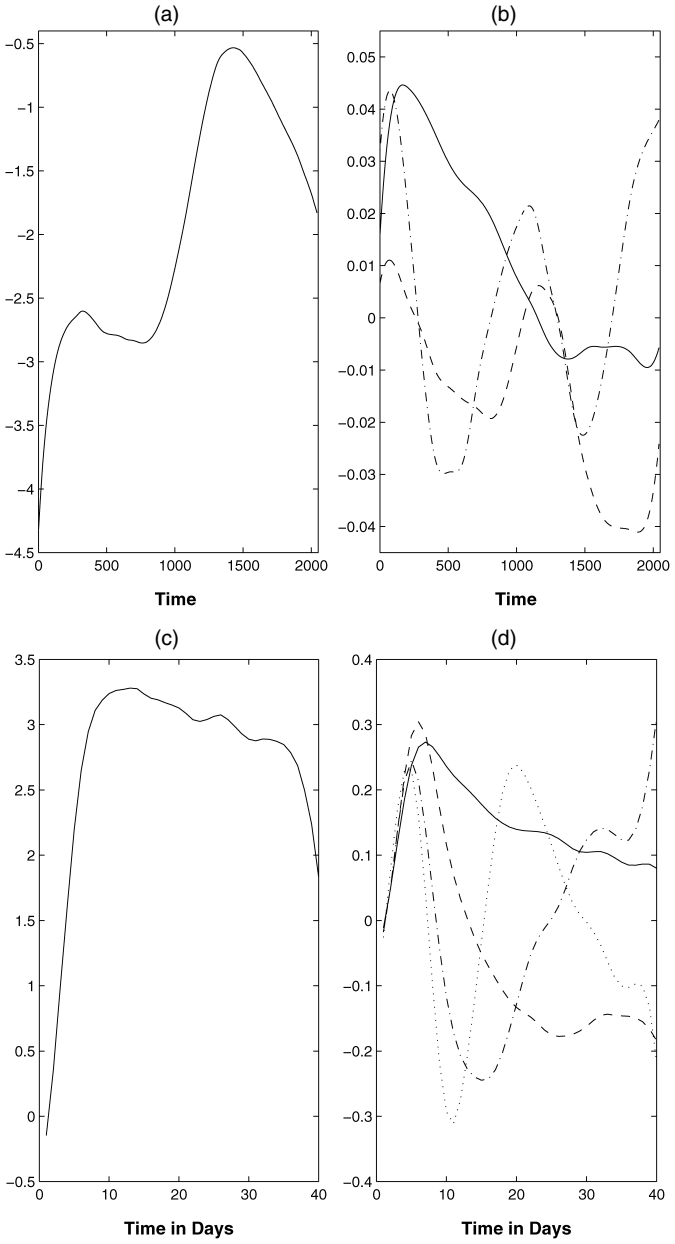

Figure 2. Components of Functional Variance Processes for the Earthquakes and Explosions Data [(a) and (b)] and the Egg-Laying Data [(c) and (d)]. (a) and (c) Smooth estimate of the mean function for the functional variance process V(t) for earthquakes and explosions data (a) and egg-laying data (c). (b) Smooth estimates of the first (——), second (- - - -), and third (-·-·-·-) eigenfunctions of V(t) for the earthquakes and explosions data, accounting for 62.8%, 23.6%, and 7.8% of total variation. (d) Smooth estimates of the first (——), second (- - - -), third (-·-·-·-), and fourth (·····) eigenfunctions of V(t) for egg-laying data, accounting for 48.3%, 21.0%, 11.6%, and 6.7% of total variation. Time unit is .025 seconds for (a) and (b), and is days for (c) and (d).

a regular dense grid of support points t_{ij} on the domain $\mathcal{T} = [a_1, a_2]$ and are related to S by

$$X_{ij} = S_i(t_{ij}) + R_{ij}, \qquad i = 1, \ldots, n, j = 1, \ldots, m. \tag{1}$$

The R_{ij} are additive noise, such that R_{ij} and $R_{i'k}$ are independent for all $i \neq i'$, and

$$ER_{ij} = 0, \qquad \operatorname{var}(R_{ij}) = \sigma_{Rij}^2 < \infty.$$

Note that the noise R_{ij} within the same subject or item i may be correlated. Throughout this article, "smooth" refers to twice continuously differentiable. The domain of S is assumed to be a compact interval $\mathcal{T} = [a_1, a_2]$. We remark that the assumptions of a dense grid of measurement times and of the same number of observations m made on each subject can be relaxed, as discussed in Appendix B after (A2.5).

Focusing on the noise R_{ij}, we assume that squared errors R_{ij}^2 are the product of two nonnegative components, one of which can be represented as an exponentiated white noise W_{ij}, and the other as an exponentiated random function $V(t)$, that is, $R_{ij}^2 = \exp(V(t_{ij}))\exp(W_{ij})$. As in the case of regression residuals, the squared errors R_{ij}^2 can be expected to carry relevant information about the random variation, and the exponential factors convey the nonnegativity restriction. The transformed errors $Z_{ij} = \log(R_{ij}^2)$ are then additively decomposed into the two components $V(t_{ij})$ and W_{ij}. The components of this decomposition are smooth random trajectories corresponding to realizations of the functional variance process V, which is our target, on the one hand and to the errors W_{ij} on the other hand. The W_{ij} are assumed to satisfy

$$E(W_{ij}) = 0, \qquad \operatorname{var}(W_{ij}) = \sigma_W^2, \qquad \text{and} \qquad W_{ij} \perp W_{ik} \quad \text{for } j \neq k. \tag{2}$$

Furthermore, $W \perp V$ and $W \perp S$, where $Q \perp T$ means that random variables Q and T are independent.

The decomposition

$$Z_{ij} = \log(R_{ij})^2 = V(t_{ij}) + W_{ij} \tag{3}$$

implies that

$$E(Z_{ij}) = E(V(t_{ij})) = \mu_V(t_{ij}), \tag{4}$$

where the functional variance process V is assumed to have a smooth mean function μ_V and a smooth covariance structure

$$G_V(s,t) = \operatorname{cov}(V(s), V(t)), \qquad s, t \in \mathcal{T}. \tag{5}$$

The autocovariance operator associated with the symmetric kernel G_V,

$$\mathbf{G}_V(f)(s) = \int_{\mathcal{T}} G_V(s,t) f(t)\, dt, \tag{6}$$

is a linear integral operator with kernel G_V mapping a function $f \in L^2(\mathcal{T})$ to the function $\mathbf{G}_V(f) \in L^2(\mathcal{T})$. It has smooth eigenfunctions ψ_k with nonnegative eigenvalues ρ_k, which are assumed to be ordered by size, $\rho_1 \geq \rho_2 \geq \cdots$. The covariance surface G_V of V can then be represented as $G_V(s,t) = \sum_k \rho_k \psi_k(s)\psi_k(t)$, $s, t \in \mathcal{T}$. A consequence is the Karhunen–Loève decomposition for random trajectories V,

$$V(t) = \mu_V(t) + \sum_{k=1}^{\infty} \zeta_k \psi_k(t), \tag{7}$$

with FPC scores ζ_k, $k \geq 1$. These are random variables with $E\zeta_k = 0$ and $\operatorname{var}(\zeta_k) = \rho_k$, which can be represented as

$$\zeta_k = \int_{\mathcal{T}} \big(V(t) - \mu_V(t)\big)\psi_k(t)\, dt. \tag{8}$$

Observing (4), given the transformed errors Z_{ij}, estimates of μ_V can be obtained by pooling these errors for all n subjects

and smoothing the resulting scatterplot. Furthermore, (2) implies that

$$\begin{aligned}\operatorname{cov}(Z_{ij}, Z_{ik}) &= \operatorname{cov}\big(V_i(t_{ij}), V_i(t_{ik})\big) \\ &= G_V(t_{ij}, t_{ik}), \qquad j \neq k. \end{aligned} \tag{9}$$

Because the covariance kernel G_V is smooth, it can be estimated from the empirical covariances of the Z_{ij}. Here the diagonal needs to be omitted because it is contaminated by the white noise error variance σ_W^2. Details on estimating such covariance surfaces have been given by Staniswalis and Lee (1998) and Yao, Müller, and Wang (2005). Once μ_V and G_V are available, the eigenfunctions are obtained by standard procedures (Rice and Silverman 1991).

Specific examples of how the assumed data structure might arise are easily constructed. Assume that we have two orthonormal systems on $\mathcal{T}$, ϕ_k and ψ_k, $k = 1, 2, \ldots$, both consisting of smooth functions, and two null sequences λ_k and ρ_k such that $\sum_k \lambda_k < \infty$ and $\sum_k \rho_k < \infty$. Take sequences of random variables ξ_k with $E(\xi_k) = 0$ and $\operatorname{var}(\xi_k) = \lambda_k$ and ζ_k with $E(\zeta_k) = 0$ and $\operatorname{var}(\zeta_k) = \rho_k$, where all of these random variables are independent. Selecting any smooth functions μ_S and μ_V on $\mathcal{T}$, we then set

$$\begin{aligned} S(t) &= \mu_S(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t) \qquad \text{and} \\ V(t) &= \mu_V(t) + \sum_{k=1}^{\infty} \zeta_k \psi_k(t). \end{aligned} \tag{10}$$

Consider random variables W_{ij} and $\varepsilon_{ij}, i = 1, \ldots, n$, $j = 1, \ldots, m$, which are independent among themselves and of all other random variables such that $E(W_{ij}) = 0, \operatorname{var}(W_{ij}) = \sigma_W^2$, and $P(\varepsilon_{ij} > 0) = P(\varepsilon_{ij} < 0) = \frac{1}{2}$. Observations X_{ij} that satisfy all of the properties mentioned earlier are then given by

$$X_{ij} = S_i(t_{ij}) + \operatorname{sign}(\varepsilon_{ij}) \big\{ \exp[V_i(t_{ij}) + W_{ij}] \big\}^{1/2}. \tag{11}$$

Bounds on the trajectories of S and V and the first two derivatives of S, as required for some of the asymptotic results, are easily achieved by choosing all but finitely many of the λ_k and ρ_k to be 0 and using bounded random variables ξ_k and ζ_k.

3. ESTIMATION OF MODEL COMPONENTS

The estimation procedures outlined in the previous section will work if the Z_{ij} can be reasonably well estimated from the available data, which indeed is the case, as we demonstrate here. As for recovering individual trajectories V_i of the functional variance process, according to (7), this requires obtaining the FPC scores ζ_k of V, given in (8). As has been shown by Yao et al. (2003), these integrals can be approximated by Riemann sums, substituting $V(t_{ij})$ by $\hat{Z}_{ij}$ and μ_V and ψ_k by estimates $\hat{\mu}_k$ and $\hat{\psi}_k$. Another component of the overall model that is of interest and must be determined is $\operatorname{var}(W_{ij}) = \sigma_W^2$.

Assume that data X_{ij} are observed on a regular and dense grid $(t_{ij}), i = 1, \ldots, n, j = 1, \ldots, m$, where i is the subject index and j is the measurement index, and that (1)–(7) hold. A core algorithm is principal analysis of random trajectories (PART). This algorithm is similar to a procedure described by Yao et al. (2005) (see also Staniswalis and Lee 1998). We provide only an outline here; for further details on the estimation steps, see Appendix A.

In a first step, following the suggestion of an anonymous reviewer, we smooth the scatterplots (t_{ij}, X_{ij}), $j = 1, \ldots, m$, separately for each trajectory S_i; any of a number of available smoothing methods can be used for this purpose and for the other subsequent smoothing steps. When using local linear smoothing, as in our implementation, we may apply a different bandwidth $b_{S,i}$ for each trajectory; see Appendixes A and B for further details. We selected bandwidths $b_{S,i}$ by cross-validation, individually per subject, which yields good results in applications and avoids biases that may arise when using cross-panel smoothing techniques, such as pooled cross-validation or an initial FPC expansion for smooth processes S. The resulting estimates $\hat{S}_i(t_{ij})$ [see (A.1) in App. A] are taken to approximate the true underlying smooth trajectory $S_i(t_{ij})$. Accordingly, we approximate the errors by the residuals $\hat{R}_{ij} = X_{ij} - \hat{S}_i(t_{ij})$ to obtain observed transformed residuals

$$\hat{Z}_{ij} = \log(\hat{R}_{ij}^2) = \log\big(X_{ij} - \hat{S}_i(t_{ij})\big)^2, \qquad i = 1, \ldots, n, j = 1, \ldots, m. \tag{12}$$

In a second step, we then apply the PART algorithm to the sample of transformed residuals $\hat{Z}_{ij}$, $i = 1, \ldots, n, j = 1, \ldots, m$, obtained in the first step. The main steps of the PART algorithm applied to these data are as follows:

1. Given the sample of all observed transformed residuals $\hat{Z}_{ij}$, estimate the mean function μ_V (4) using a univariate weighted least squares smoother with bandwidth b_V applied to the aggregated scatterplot of all observations; details are as in (A.2) in Appendix A. The bandwidth b_V is chosen data-adaptively by cross-validation.
2. Estimate the smooth covariance surface G_V [see (5)] by applying two-dimensional smoothing [see (A.3)], fitting local planes by weighted least squares to empirical covariances, and using bandwidth h_V, which in applications is chosen by cross-validation. The empirical covariances from which the covariance surface is obtained are constructed between all pairs of observations $(t_{ij}, t_{ij'})$, $t_{ij} \neq t_{ij'}$, whereas the empirical variances obtained at the diagonal of the surface are omitted, because these are contaminated with the residual variance σ_W^2; see (2).
3. From estimated covariance surface and mean function, obtain estimated eigenfunctions and eigenvalues using discretization and numerical algorithms; see (A.4).
4. Estimate the variance $\operatorname{var}(W_{ij}) = \sigma_W^2$ [see (2)]. This involves a one-dimensional smoothing step along the diagonal of the covariance surface, using bandwidth b_{Q_W} in the direction of the diagonal, and then obtaining the estimate $\hat{\sigma}_W^2$ as in (A.7). Again, in our data-adaptive implementation bandwidths, b_{Q_W} are chosen by cross-validation.
5. Estimating individual FPC scores ζ_j [see (8)] by numerical integration as in (A.5).

The algorithm also provides an estimate of the number of functional principal components M needed to approximate processes V, using leave-out-one-curve cross-validation;

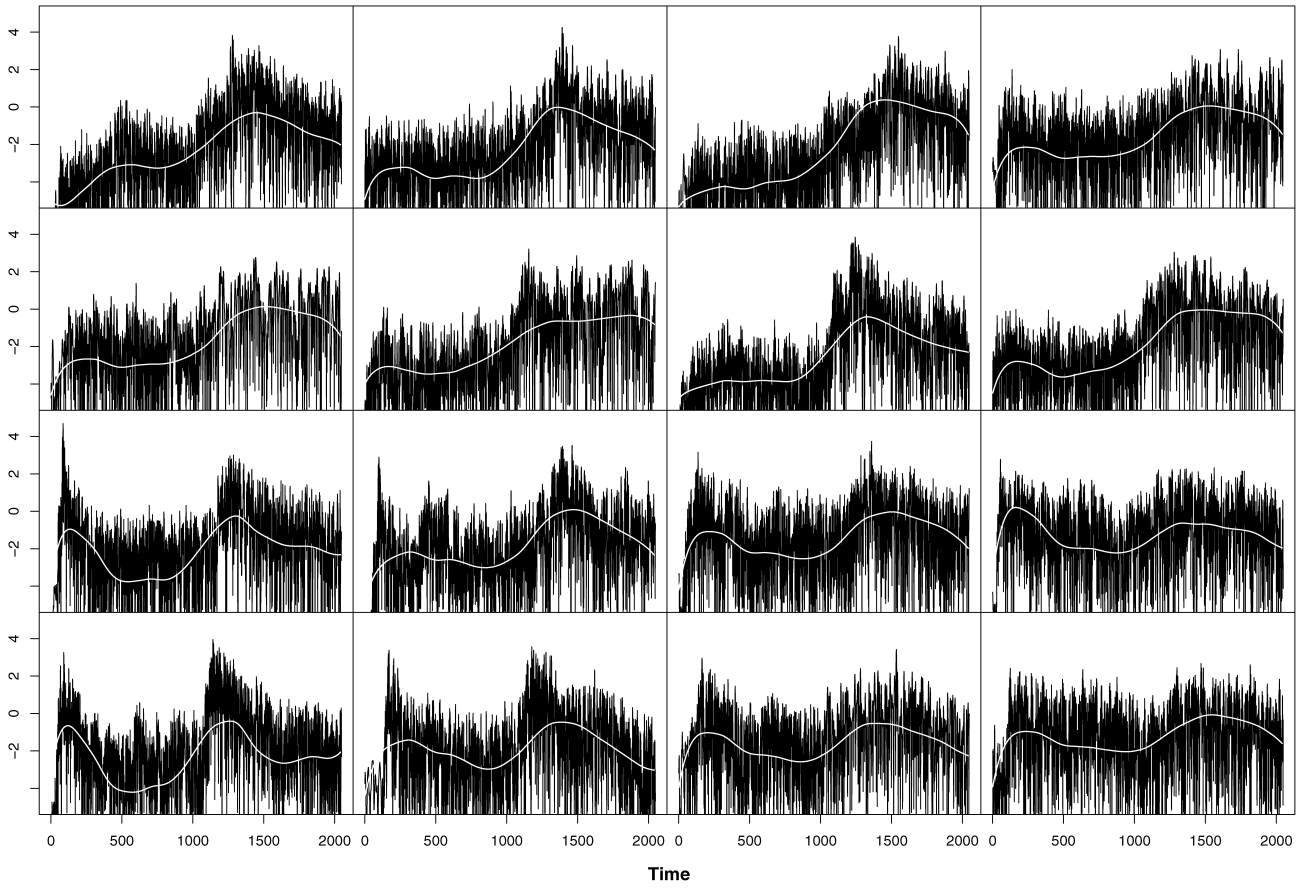

Figure 3. Observed Values of $Z_{ij} = \log(R_i^2(t_{ij}))$ (black) and Estimated Trajectories of Functional Variance Processes $\hat{V}_i(t)$ [see (13)] (white) for Eight Explosions (first two rows), Seven Earthquakes (last two rows except for the bottom right panel), and one Unknown Event (bottom right panel). The first earthquake out of eight earthquakes is not shown; the order of earthquakes and explosions is as in Figure 1. Time unit is .025 seconds.

see (A.6). Alternative selectors, such as pseudo–Akaike information and pseudo–Bayes information criteria (Yao et al. 2005), might be used as well.

The output consists of estimated mean function $\hat{\mu}_V$, estimated eigenfunctions/eigenvalues $\hat{\psi}_k$ and $\hat{\rho}_k$, estimated noise variance $\hat{\sigma}_W^2$, and estimated FPC scores $\hat{\xi}_{ik}$. According to (7), if a number M of approximating components is chosen by the algorithm, then this leads to fitted individual functional variance process trajectories

$$\hat{V}_i(t) = \hat{\mu}_V(t) + \sum_{k=1}^{M} \hat{\xi}_{ik}\hat{\psi}_k(t). \tag{13}$$

Examples of such estimated trajectories are shown in Figure 3.

4. ASYMPTOTIC RESULTS

To develop functional asymptotic results for the components of the expansion (13) of individual estimated trajectories of the functional variance process, a preliminary first step is to derive bounds for the differences between actual transformed errors Z_{ij} [see (3)] and the observed transformed residuals $\hat{Z}_{ij}$ [see (12)] that are becoming available after the initial smoothing step, which aims to recover the smooth trajectories S_i. In what follows, we refer to bandwidths b_S for smoothing trajectories S_i [see (A.1) in App. A]; the bandwidth sequence b_S represents bandwidths $b_{S,i}$, which are chosen separately for each individual trajectory. These bandwidths, $b_{S,i}$, are tied to a universal sequence of bandwidths, b_S, according to assumption (A2.1), such that the overall sequence, b_S, satisfies (A2.2); these assumptions are listed in Appendix B. Bandwidths b_V, h_V, and b_{Q_V} are used in the smoothing steps for $\hat{\mu}_V$ in (A.2), $\hat{G}_V(s,t)$ in (A.3), and $\hat{Q}_V(t)$ in (A.7) in Appendix A. These choices are governed by assumptions (A2.3)–(A2.5).

We obtain the following consistency properties for the random trajectories, where m is the number of measurements that are available for each trajectory. Assumptions (A) and (B) are given in Appendix B, and the proofs are provided in Appendix C.

Theorem 1. Under conditions (A1), (A2), (B1.1), and (B2.1), it holds for smoothed trajectories $\hat{S}_i(t)$ that

$$E\Big(\sup_{t\in\mathcal{T}} |\hat{S}_i(t) - S_i(t)|\Big) = O\Big(b_S^2 + \frac{1}{\sqrt{m}b_S}\Big). \tag{14}$$

As a consequence of (14), if we apply the PART algorithm to the observed transformed residuals $\hat{Z}_{ij}$, then we expect to obtain consistent estimates of the components of the functional variance process, which is our target. The difficulty here is that we do not observe the actual transformed errors Z_{ij}, but observe only the approximate values $\hat{Z}_{ij}$, corresponding to the transformed residuals from the initial smoothing step.

The next result establishes consistency of the estimates of the components of the functional variance process, namely the esti-

mate $\hat{\mu}_V(t)$ of the mean function $\mu_V(t)$, the estimate $\hat{G}_V(s,t)$ of the covariance function $G_V(s,t)$, and estimates $\hat{\rho}_k$ and $\hat{\psi}_k(t)$ of eigenvalues ρ_k and eigenfunctions ψ_k. These components are obtained as in (A.2), (A.3), and (A.4) in Appendix A and characterize the functional variance process. Consistent estimation of these components validates our approach asymptotically. In addition, consistency of the estimate $\hat{\sigma}_W^2$ [see (A.7)] of the noise variance σ_W^2 [see (2)] is also obtained.

Theorem 2. Under conditions (A1)–(A8) and (B1.1)–(B2.2), it holds for the estimates of the components of the functional variance process that

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} |\hat{\mu}_V(t) - \mu_V(t)| \\
&\quad = O_p\left(b_S^2 + \frac{1}{\sqrt{m}b_S} + \frac{1}{\sqrt{n}b_V}\right), \\
&\sup_{s,t\in\mathcal{T}} |\hat{G}_V(s,t) - G_V(s,t)| \\
&\quad = O_p\left(b_S^2 + \frac{1}{\sqrt{m}b_S} + \frac{1}{\sqrt{n}h_V^2}\right), \\
&|\hat{\sigma}_W^2 - \sigma_W^2| \\
&\quad = O_p\left(b_S^2 + \frac{1}{\sqrt{m}b_S} + \frac{1}{\sqrt{n}h_V^2} + \frac{1}{\sqrt{n}b_{Q_V}}\right).
\end{aligned} \tag{15}$$

Considering eigenvalues ρ_k of multiplicity 1, $\hat{\psi}_k$ can be chosen such that

$$\sup_{t\in\mathcal{T}} |\hat{\psi}_k(t) - \psi_k(t)| \xrightarrow{p} 0, \qquad \hat{\rho}_k \xrightarrow{p} \rho_k. \tag{16}$$

The rates of convergence of the estimated eigenvalues $\hat{\rho}_k$ and eigenfunctions $\hat{\psi}_k$ can be obtained as $\sup_{t\in\mathcal{T}} |\hat{\psi}_k(t) - \psi_k(t)| = O_p(\alpha_{nk} + \alpha_{nk}^*)$ and $|\hat{\rho}_k - \rho_k| = O_p(\alpha_{nk} + \alpha_{nk}^*)$, where α_{nk} and α_{nk}^* are defined in (C.1) in Appendix C, using definitions (B.1) and (B.2).

Another central result provides consistency for individually estimated trajectories $\hat{V}_i$ [see (13)] of functional variance trajectories V_i, such as those drawn in Figure 3. We proceed by first establishing the consistency of estimates $\hat{\zeta}_{ik}$ [see (A.5)] of individual FPCs ζ_{ik} of functional variance processes V. This result provides asymptotic justification for the proposed estimates of individual trajectories of functional variance processes.

Theorem 3. Under conditions (A1)–(A8) and (B1.1)–(B2.2), it holds for the estimates of the FPCs of functional variance processes V that

$$\sup_{1\le k\le M} |\hat{\zeta}_{ik} - \zeta_{ik}| \xrightarrow{p} 0, \tag{17}$$

where for the number of components in expansion (13), $M = M(n) \to \infty$, as $n \to \infty$. Furthermore, for estimated trajectories $\hat{V}_i(t)$ of the functional variance process V, it holds that for $1 \le i \le n$,

$$\sup_{t\in\mathcal{T}} |\hat{V}_i(t) - V_i(t)| \xrightarrow{p} 0. \tag{18}$$

We note that for the convergence in (16), the conditions on the number of observed trajectories n and on the number of points m at which each trajectory is sampled must satisfy $n, m \to \infty$ under conditions (A2) and (A3), whereas for the convergence of (17), the number of included components also must satisfy $M(n) \to \infty$ and, furthermore, conditions (A5)–(A8) must hold. These conditions amount to upper limits on the speed at which $M(n) \to \infty$. To conclude this section, we remark that the rates of convergence of estimated trajectories $\hat{V}_i$ in (18) depend on properties of the underlying processes S and V and can be determined as $\sup_{t\in\mathcal{T}} |\hat{V}_i(t) - V_i(t)| = O(\vartheta_{in}^{(1)} + \vartheta_{in}^{(2)})$, where the $O(\cdot)$ terms hold uniformly in $1 \le i \le n$ and $\vartheta_{in}^{(1)}$ and $\vartheta_{in}^{(2)}$ are random variables as defined in (C.9) in Appendix C.

5. APPLICATIONS OF FUNCTIONAL VARIANCE PROCESSES

5.1 Earthquake and Mining Exploration Series

The series in Figure 1 represent typical earthquake and mining explosion seismic data from a suite of eight earthquakes and eight explosions and an event of unknown mechanism originating on the Scandinavian peninsula, as recorded by seismic arrays. We standardized each series by dividing by the sample standard deviation for the entire series before analysis. The general problem of interest for these data is to distinguish or discriminate between waveforms generated by earthquakes and those generated by explosions. Note that both series contain two phases, the initial body wave [so-called "arrivals" ($t = 1, \ldots, 1{,}024$)] and the secondary shear wave ($t = 1{,}025, \ldots, 2{,}048$).

Ratios and amplitudes of the two components, as well as spectral ratios in different frequency bands, have been used in previous attempts at feature-based discriminant analysis (see, e.g., Kakizawa et al. 1998). Shumway (2002) proposed using time-varying spectra for classification and clustering of nonstationary time series. Our proposal is to apply functional data analysis methods to perform discriminant analysis. This can be done in the standard way by targeting the smooth random process $S(t)$ [see (1)] and its decomposition into eigenfunctions and FPC scores, as in (10), using, for example, the estimation methods described by Rice and Silverman (1991).

Because for these data the major information of interest appears to reside in the random variation, the application of the newly introduced functional variance process is of interest. Three eigenfunctions are chosen by cross-validation [see (A.6)] to represent the dominant modes of variation for V. The estimated mean function $\mu_V(\cdot)$ and estimated first three eigenfunctions for the functional variance process V are displayed in Figures 2(a) and 2(b), with the mean function on the left and the eigenfunctions on the right. The first eigenfunction is broadly associated with the size of the body wave, whereas the second eigenfunction forms two contrasts, one between the early and late phases of the body wave and the other between the early and late phases of the shear wave. The third eigenfunction also forms two contrasts, which are more clearly expressed and emphasize somewhat earlier times compared with the second eigenfunction. Another quantity of interest is the constant variance of the white noise process W, estimated here as $\hat{\sigma}_W^2 = 3.07$, using (A.7). We note that in practice, with discrete data such as the explosions and earthquake data, it may be that in (12) the

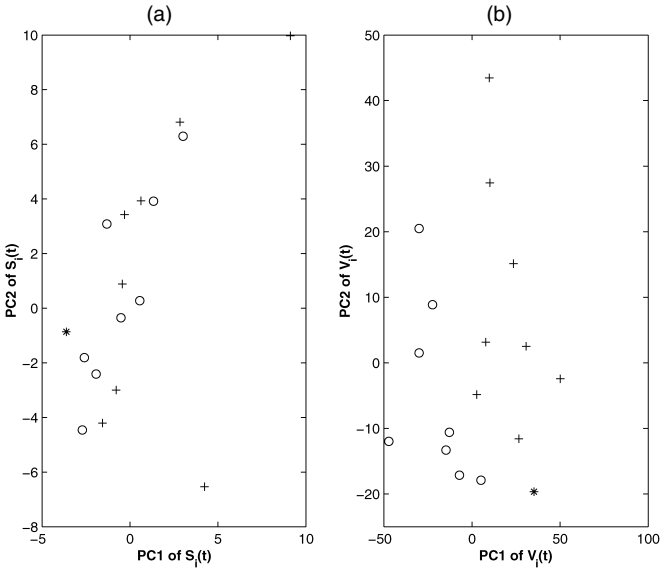

Figure 4. Representation of the First Two Estimated Functional Principal Component Scores, PC2 versus PC1, Obtained for the Smooth Processes S_i (a) and for the Functional Variance Processes V_i (b), for Earthquakes and Explosions Data (+, earthquakes; ∘, explosions; ∗, unknown event).

term $(X_{ij} - \hat{S}_i(t_{ij}))^2$ is 0, and thus we added .001 to the squared residuals $\hat{R}_{ij}^2$ before taking the log.

The estimates for the trajectories V_i for the same data shown in Figure 1 are depicted in Figure 3. These estimated random trajectories correspond to realizations of the functional variance process and visually reflect the local variation of the data when compared with the corresponding panels of Figure 1. An early peak in the variance process trajectories is quite noticeable for the earthquakes, whereas it is largely absent for the explosions.

Estimates $\hat{\zeta}_{ik}$, $k = 1, 2$, $i = 1, \ldots, 15$ [see (A.5)], of the first two FPC scores [see (8)] of processes V, presented in Figure 4(b), show a clear separation between the two types of events. The corresponding estimates of the first two FPC scores of a more traditional FPC analysis of processes S (Rice and Silverman 1991, implemented here following Yao et al. 2003) are shown in Figure 4(a). We see that the pattern of the FPC scores obtained for the variance processes V_i is much more striking than that obtained for processes S_i. This clearly indicates that using the scores obtained for functional variance processes V_i here leads to a more illuminating analysis.

Visually, the second versus first FPC scores of S, shown in Figure 4(a), do not distinguish between explosions and earthquakes, indicating that processes S, which are the commonly used basis for FPC analysis, do not contain much information for discriminating between the two groups. In contrast, the scores $(\hat{\zeta}_{i1}, \hat{\zeta}_{i2})$ obtained for functional variance processes clearly distinguish explosions and earthquakes; a line can be drawn to separate the two groups. In fact, the leave-one-out misclassification error for logistic discriminant analysis based on the scores for S led to 7 misclassifications (out of 15 events), whereas the scores for the functional variance process led to 0 misclassifications, thus demonstrating the usefulness of the functional variance process approach.

The last event from an unknown origin is classified as an explosion if we use the scores from S, and as an earthquake based on the scores $\hat{\zeta}_{ik}$, $k = 1, 2$, for the functional variance process trajectories $V_i(t)$. Because the classification based on functional variance processes is clearly more reliable, we conclude from this analysis that the unknown event is an earthquake.

5.2 Egg-Laying Data

To illustrate the application of functional variance processes to a biological problem, we selected 359 medflies from a study of 1,000 female medflies described by Carey, Liedo, Müller, Wang, and Chiou (1998) with lifetimes no less than 40 days, and investigated the dynamics of the number of eggs laid daily during the first 40 days. The estimated trajectories S obtained from the initial smoothing step for eight randomly selected flies are shown in the top eight panels of Figure 5. The shapes of the egg-laying curves vary quite a bit, but a general feature is a more or less rapid increase in egg-laying activity followed by a more protracted decline.

The estimated mean function $\hat{\mu}_V$ [see (A.2)] for the functional variance processes V and the first four eigenfunctions for these processes are depicted in Figures 2(c) and 2(d). Here four components were chosen by leave-out-one-curve cross-validation [see (A.6)]. As mentioned before, we note that in practice it may happen that in (12), the term $(X_{ij} - \hat{S}_i(t_{ij}))^2$ is 0, and thus we added 1 to the squared residuals $\hat{R}_{ij}^2$ before taking the log.

A major component of variation in the egg-laying curves is seen to occur (more or less) along the direction of the mean egg-laying curve; that is, the mean function and the first eigenfunction appear somewhat aligned. The second eigenfunction emphasizes an early sharp peak in variation and then forms a contrast with protracted high values, and the higher-order eigenfunctions align with more complex contrasts, while also emphasizing the initial rise. The variance of the noise process $W(t)$ using (A.7) is found to be $\hat{\sigma}_W^2 = 1.78$.

The eight estimated variance process trajectories $\hat{V}_i$ for the eight flies whose egg-laying trajectories are displayed in the top panels of Figure 5 are shown in the bottom panels of this figure. They typically increase rapidly from 0 up to a high level, and then tend to stay at that level with only a slight decline. This seems to imply that the behavior after the initial peak is quite different between processes S and V. Although the trajectories of smooth components S for the most part monotonically decline after the initial egg-laying peak, the trajectories of the variance processes remain more or less constant and elevated, with individual variations.

These findings provide some evidence that the variance structure of these data is not of a simple Poisson type, as could have been surmised based on the concept of the data as counts. What we see instead is that as the average counts decrease, their individual variability relative to the mean count increases as flies age. The observed high variability of the reproductive activity of older flies may be a characteristic of the aging process itself. It reflects surprisingly large oscillations in old-age reproduction of medflies. Although the overall reproductive activity of flies declines with age, it becomes less predictable at the individual level due to these large oscillations.

6. DISCUSSION AND CONCLUSIONS

Functional variance processes are a new tool in the emerging field of functional data analysis. They extend the notion of a

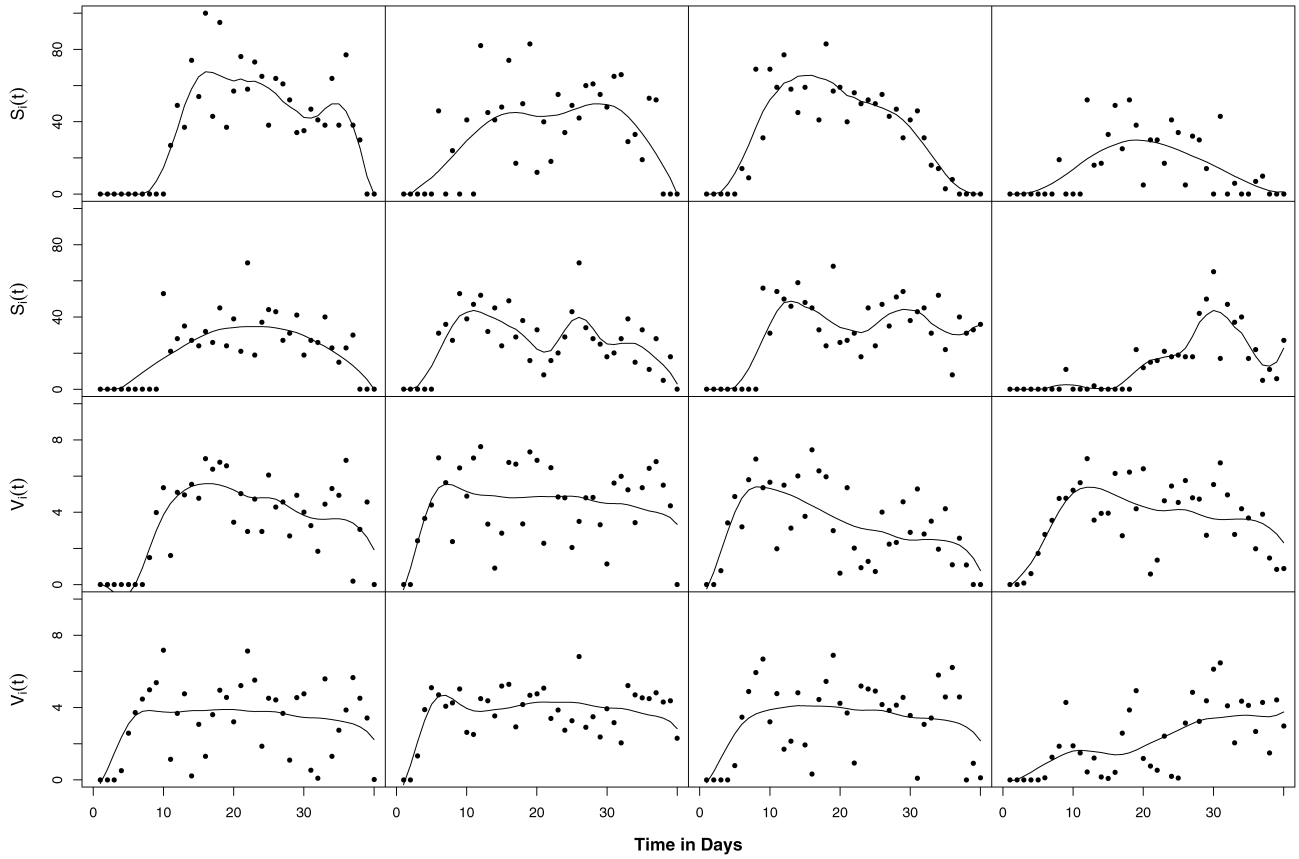

Figure 5. Observed Egg-Laying Counts (•) and Smoothed Individual Trajectories $\hat{S}_i$ for Eight Medflies for the First 40 Days of Age, With the Bandwidths Chosen by Leave-One-Out Cross-Validation for Each Subject (top two rows), and Observed Values for $Z_{ij} = \log(R^2(t_{ij}))$ (•) and Estimated Smooth Random Trajectories $\hat{V}_i$ (13) of the Functional Variance Process for the Same Eight Medflies (bottom two rows).

variance function as it is commonly used in semiparametric and generalized regression modeling to the case of functional data and random variance trajectories. As we have demonstrated, this concept and its associated statistical tools are useful to gain an understanding of complex functional data, including longitudinal data and panels of time series, and may provide novel insights into the structure of such data. In our approach, functional variance processes are characterized by their mean and eigenfunctions, which convey information about the underlying data structure. An individual trajectory of this process is obtained for each observed longitudinal data series and is characterized by its FPC scores. These quantities are shown to be estimable with consistent estimators.

The functional variance process approach leads to a representation of each longitudinal series by two trajectories. The first of these is the trajectory S_i corresponding to the smooth process S, which has been the traditional target of FPC analysis and which we approximate for our purposes here by an initial smoothing step. Alternatively, the trajectories S_i could be represented in the form of an FPC analysis of the process S, especially if summarizing the trajectories S_i into a few random coefficients is desired; we note that our theoretical analysis can be extended to cover this case. The second trajectory characterizing the data is V_i, corresponding to the realization of the smooth functional variance process. These trajectories can be visualized and interpreted in a meaningful way in applications. The FPC scores of the functional variance process are useful for functional modeling and may serve as input for classification or functional regression.

Although our algorithms lead to relatively stable and easily applicable procedures that can be implemented in a fully automatic data-adaptive way, further investigations into the practical effects of smoothing parameter choices and longitudinal designs will be of interest. We note that changing the number of included components, the smoothing parameters, or the manner in which the smooth processes S are handled (e.g., in a presmoothing step as described in this article or, alternatively, by another FPC analysis) will lead to changes in the estimated FPC scores and estimated trajectories of the functional variance process. In the application to seismic data, we found that the discriminating ability of the FPC scores was not particularly sensitive to these choices. Generally, how big a role these choices will play will depend on the final goal of the analysis.

Another area of future research is the development of inference procedures for variance processes in both asymptotic and practical situations. A possibility for practical applications is to derive inference from a functional bootstrap. Theoretical developments along these lines will depend on a careful analysis of the properties of the FPCs for variance processes. Functional variance processes are likely to play a significant role in generalized functional modeling, where they may be included as additional predictor or response processes in functional regression models. They also serve a valuable purpose in functional discriminant analysis, as has been demonstrated for the seismic data example. In analogy to the situation in nonparametric and

semiparametric regression, functional variance processes may be useful in obtaining more efficient functional methodology, in constructing confidence regions, and, more generally, for inference in functional models. Functional models with variance processes as response may be of special interest in applications where changes in variance over time are of prime interest, such as in modeling volatility for financial market data.

APPENDIX A: ESTIMATION PROCEDURES

Let $\kappa_1(\cdot)$ and $\kappa_2(\cdot,\cdot)$ be nonnegative univariate and bivariate kernel functions used as weights for locally weighted least squares smoothing in one and two dimensions that satisfy assumptions (B2.1) and (B2.2). Let $b_V = b_V(n)$ and $h_V = h_V(n)$ be the bandwidths for estimating μ_V in (4) and G_V in (5) in steps 1 and 2 of the PART algorithm applied to the transformed residuals Z_{ij}.

Local linear scatterplot smoothers (Fan and Gijbels 1996) for estimating individual trajectories S_i, $i = 1,\dots,n$, from data (t_{ij}, X_{ij}), $j = 1,\dots,m$, with bandwidths $b_{S,i}$ are obtained through minimizing

$$\sum_{j=1}^{m} \kappa_1\left(\frac{t_{ij}-t}{b_{S,i}}\right)\{X_{ij}-\beta_{i,0}-\beta_{i,1}(t-t_{ij})\}^2 \tag{A.1}$$

with respect to $\beta_{i,0}$ and $\beta_{i,1}$. The resulting estimates are $\hat{S}_i(t_{ij}) = \hat{\beta}_{i,0}(t_{ij})$. Note that individual bandwidths $b_{S,i}$ are tied to an overall bandwidth sequence b_S in assumption (A2.1).

For estimating μ_V, the first step in the PART algorithm, we also use local linear smoothing by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \kappa_1\left(\frac{t_{ij}-t}{b_V}\right)\left\{\hat{Z}_{ij}-\beta_0-\beta_1(t-t_{ij})\right\}^2 \tag{A.2}$$

with respect to β_0 and β_1, leading to $\hat{\mu}_V(t) = \hat{\beta}_0(t)$. Let $\mathrm{G}_i(t_{ij_1}, t_{ij_2}) = (\hat{Z}_i(t_{ij_1}) - \hat{\mu}_V(t_{ij_1}))(\hat{Z}_i(t_{ij_2}) - \hat{\mu}_V(t_{ij_2}))$, and define the local linear surface smoother for $G_V(s,t)$ by minimizing

$$\sum_{i=1}^{n}\sum_{1\le j_1\ne j_2\le m} \kappa_2\left(\frac{t_{ij_1}-s}{h_V}, \frac{t_{ij_2}-t}{h_V}\right) \times \left\{\mathrm{G}_i\left(t_{ij_1}, t_{ij_2}\right) - f\left(\beta, (s,t), \left(t_{ij_1}, t_{ij_2}\right)\right)\right\}^2, \tag{A.3}$$

where $f(\beta, (s,t), (t_{ij_1}, t_{ij_2})) = \beta_0 + \beta_{11}(s - t_{ij_1}) + \beta_{12}(t - t_{ij_2})$, with respect to $\beta = (\beta_0, \beta_{11}, \beta_{12})$, yielding $\hat{G}_V(s,t) = \hat{\beta}_0(s,t)$.

The estimates of $\{\rho_k, \psi_k\}_{k\ge1}$ correspond to the solutions $\{\hat{\rho}_k, \hat{\psi}_k\}_{k\ge1}$ of the eigenequations

$$\int_{\mathcal{T}} \hat{G}_V(s,t)\hat{\psi}_k(s)\,ds = \hat{\rho}_k\hat{\psi}_k(t), \tag{A.4}$$

with orthonormality constraints on $\{\hat{\psi}_k\}_{k\ge1}$ and positive definiteness constraints (for the latter, see Yao et al. 2003). We use a simple discrete integral approximation to estimate the first M FPC scores ζ_{ik} (8),

$$\hat{\xi}_{ik} = \sum_{j=2}^{m}\left(\hat{Z}_{ij} - \hat{\mu}_V(t_{ij})\right)\hat{\psi}_k(t_{ij})(t_{ij} - t_{i,j-1}), \qquad i = 1,\dots,n, k = 1,\dots,M. \tag{A.5}$$

Let $\hat{\mu}_V^{(-i)}$ and $\hat{\psi}_k^{(-i)}$ be the estimated mean and eigenfunctions after removing the data for the ith subject. Leave-out-one-curve cross-validation aims to minimize

$$\mathrm{CV}_V(M) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left\{\hat{Z}_{ij} - \hat{V}_i^{(-i)}(t_{ij})\right\}^2, \tag{A.6}$$

with respect to the number of included components M, where $\hat{V}_i^{(-i)}(t) = \hat{\mu}_V^{(-i)}(t) + \sum_{k=1}^{M}\hat{\xi}_{ik}^{(-i)}\hat{\psi}_k^{(-i)}(t)$ and $\hat{\xi}_{ik}^{(-i)}$ is obtained by (A.5). The proposed estimates for individual smooth trajectories V_i are then given by (13).

For the estimation of the white noise variance σ_W^2, we first fit a local quadratic component orthogonal to the diagonal of G_V, and a local linear component in the direction of the diagonal. Denote the diagonal of the resulting surface estimate by $\hat{G}_V^*(t)$ and a local linear smoother focusing on diagonal values $\{G_V(t,t) + \sigma_W^2\}$ by $\hat{Q}_V(s)$, using bandwidth b_{Q_V}. As $\mathcal{T} = [a_1, a_2]$, let $|\mathcal{T}| = a_2 - a_1$ and $\mathcal{T}_1 = [a_1 + |\mathcal{T}|/4, a_2 - |\mathcal{T}|/4]$. Then we obtain the estimate

$$\hat{\sigma}_W^2 = \frac{1}{|\mathcal{T}_1|}\int_{\mathcal{T}_1}\{\hat{Q}_V(t) - \hat{G}_V^*(t)\}_+\,dt \tag{A.7}$$

if $\hat{\sigma}_W^2 > 0$ and $\hat{\sigma}_W^2 = 0$ otherwise. To attenuate boundary effects, removing intervals of lengths $|\mathcal{T}|/4$ near both boundaries was empirically found to produce good results (Yao et al. 2003).

APPENDIX B: ASSUMPTIONS AND NOTATIONS

Processes S and V are assumed to be twice continuously differentiable and to have the following properties:

(A1) There exists a constant $C > 0$ such that trajectories of processes S and V satisfy

$$\sup_t\left|S^{(\nu)}(t)\right| < C \quad \text{for } \nu = 0,1,2 \qquad \text{and}$$
$$\sup_t |V(t)| < C.$$

Recall that $b_{S,i} = b_{S,i}(n)$, $b_V = b_V(n)$, $h_V = h_V(n)$, and $b_{Q_V} = b_{Q_V}(n)$ are the bandwidths for estimating $\hat{S}_i$ [see (A.1)], $\hat{\mu}_V$ [see (A.2)], $\hat{G}_V$ [see (A.3)], and $\hat{Q}_V(t)$ [see (A.7)]. We develop asymptotics as the number of subjects n and the number of observations per subject m both tend to infinity, under the following assumptions on the smoothing parameters:

(A2.1) Regarding bandwidths $b_{S,i}$, there exists a common sequence of bandwidths b_S such that for constants c_1 and c_2, $0 < c_1 < \inf_i b_{S,i}/b_S \le \sup_i b_{S,i}/b_S < c_2 < \infty$.

(A2.2) $m \to \infty$, $b_S \to 0$, and $mb_S^2 \to \infty$.

(A2.3) $b_V \to 0$, $b_{Q_V} \to 0$, $nb_V^4 \to \infty$, $nb_{Q_V}^4 \to \infty$, $\limsup_n n \times b_V^6 < \infty$, and $\limsup_n nb_{Q_V}^6 < \infty$.

(A2.4) $h_V \to 0$, $nh_V^6 \to \infty$, and $\limsup_n nh_V^8 < \infty$.

(A2.5) $\limsup_n n^{1/2}b_V m^{-1} < \infty$, $\limsup_n n^{1/2}b_{Q_V}m^{-1} < \infty$, and $\limsup_n n^{1/2}h_V m^{-1} < \infty$.

The time points $\{t_{ij}\}_{i=1,\dots,n;j=1,\dots,m}$ at which the observations are sampled are considered to correspond to a dense regular design and are the same for all subjects. The results can be easily extended to the case of more irregular designs, as detailed later. We assume that for all i and $j = 1,\dots,m-1$, $t_{ij} < t_{i,j+1}$ and that there exists a smooth design density f satisfying $\int_{\mathcal{T}} f(t)\,dt = 1$ and $\inf_{t\in\mathcal{T}} f(t) > 0$ that generates the time points t_{ij} according to $t_{ij} = F^{-1}(\frac{j-1}{m-1})$, where F^{-1} is the inverse of $F(t) = \int_{a_1}^{t} f(s)\,ds$. These assumptions reflect the notion of a dense regular design. They can be further relaxed at additional notational expense. For example, we may include situations where the design densities generating the times t_{ij} depend on the subject i, as long as there exist constants c_1 and c_2 such that all of these design densities f_i are uniformly smooth and satisfy $0 < c_1 < \inf_i \inf_{t\in\mathcal{T}} f_i(t) < \sup_i \sup_{t\in\mathcal{T}} f_i(t) < c_2$. Furthermore, the number of measurements N_i made on the ith subject may differ from subject to subject, as long as there is a sequence $m \to \infty$ such that for suitable constants c_1 and c_2, $0 < c_1 < \inf_i \frac{N_i}{m} < \sup_i \frac{N_i}{m} < c_2 < \infty$; our analysis focuses on the case where $N_i = m$, so that we refer only to m in the following, whereas the more general cases are covered by analogous arguments.

The assumptions imply that for $\Delta_n = \max\{t_{ij} - t_{i,j-1} : j = 2, \ldots, m\}$, it holds that

(A3) $\Delta_n = O(m^{-1})$, as $n, m \to \infty$.

Assume that the fourth moments of the X_{ij} and Z_{ij} are uniformly bounded for all $t \in \mathcal{T}$, that is,

(A4) $\sup_j E[X_{ij}^4] < \infty$ and $\sup_j E[Z_{ij}^4] < \infty$.

Background on linear operators in Hilbert space as needed for the following can be found in, for example, Courant and Hilbert (1989). Define the operator $(f \otimes g)(h) = \langle f, h\rangle g$ for $f, g, h \in H$, and denote the separable Hilbert space generated by the Hilbert–Schmidt operators on H by $F \equiv \sigma_2(H)$, endowed with the inner product $\langle T_1, T_2\rangle_F = \mathrm{tr}(T_1 T_2^*) = \sum_j \langle T_1 u_j, T_2 u_j\rangle_H$ and the norm $\|T\|_F^2 = \langle T, T\rangle_F$, where $T_1, T_2, T \in F$, T_2^* is the adjoint of T_2, and $\{u_j : j \geq 1\}$ is any complete orthonormal system in H. The covariance operator $\mathbf{G}_V$ defined in (6) and its estimate $\hat{\mathbf{G}}_V$, generated by kernels G_V [see (5)], respectively $\hat{G}_V$ [see (A.3)], are Hilbert–Schmidt operators. Let $\mathcal{I}_i = \{j : \rho_j = \rho_i\}$ and $\mathcal{I}' = \{i : |\mathcal{I}_i| = 1\}$, where $|\mathcal{I}_i|$ denotes the number of elements in $\mathcal{I}_i$. Let $\mathbf{P}_j^V = \sum_{k\in\mathcal{I}_j} \psi_k \otimes \psi_k$ and $\hat{\mathbf{P}}_j^V = \sum_{k\in\mathcal{I}_j} \hat{\psi}_k \otimes \hat{\psi}_k$ denote the true and estimated orthogonal projection operators from H to the subspace spanned by $\{\psi_k : k \in \mathcal{I}_j\}$ and $\{\hat{\psi}_k : k \in \mathcal{I}_j\}$. For fixed j, let

$$\delta_j^V = \frac{1}{2}\min\{|\rho_l - \rho_j| : l \notin \mathcal{I}_j\}, \tag{B.1}$$

and let $\mathbf{\Lambda}_{\delta_j^V} = \{z \in \mathcal{C} : |z - \rho_j| = \delta_j^V\}$, where $\mathcal{C}$ represents the set of complex numbers. The resolvents of $\mathbf{G}_V$ and $\hat{\mathbf{G}}_V$ are denoted by $\mathbf{R}_V$ and $\hat{\mathbf{R}}_V$, that is, $\mathbf{R}_V(z) = (\mathbf{G}_V - zI)^{-1}$ and $\hat{\mathbf{R}}_V(z) = (\hat{\mathbf{G}}_V - zI)^{-1}$.

Let

$$A_{\delta_j^V} = \sup\bigl\{\|\mathbf{R}_V(z)\|_F : z \in \mathbf{\Lambda}_{\delta_j^V}\bigr\} \tag{B.2}$$

and $M = M(n)$ denote the numbers of components, corresponding to the eigenfunctions that are included to approximate $V(t)$, that is, $\hat{V}_i(t) = \hat{\mu}_V(t) + \sum_{m=1}^{M(n)} \hat{\zeta}_{im}\hat{\psi}_m(t)$ [see (13)]. Denote by $\|\pi\|_\infty = \sup_{t\in\mathcal{T}} |\pi(t)|$ the sup-norm for an arbitrary function $\pi(\cdot)$ with support $\mathcal{T}$. We assume that mean functions μ_V and eigenfunctions ψ_j are smooth, that is, twice continuously differentiable. The sequences $M = M(n)$ are assumed to depend on n and m and in such a way that as $n \to \infty$,

(A5) $\tau_n = \sum_{j=1}^M (\delta_j^V A_{\delta_j^V}\|\psi_j\|_\infty)/(\sqrt{n}h_V^2 - A_{\delta_j^V}) \to 0$, as $M = M(n) \to \infty$;

(A6) $\sum_{j=1}^M \|\psi_j\|_\infty = o(\min\{\sqrt{n}b_V, \sqrt{m}\})$ and $\sum_{j=1}^M \|\psi_j\|_\infty \times \|\psi_j'\|_\infty = o(m)$.

We note that assumptions (A5) and (A6) do not require that eigenfunctions ψ_j or their derivatives be uniformly bounded, but rather their growth be bounded as the index j increases. Defining $\theta_m = b_S^2 + (\sqrt{m}b_S)^{-1}$, processes V are assumed to have the following properties:

(A7) $E\{[\sup_{t\in\mathcal{T}} |V(t) - V^{(M)}(t)|]^2\} = o(n)$, where $V^{(M)}(t) = \mu_V(t) + \sum_{k=1}^M \zeta_k\psi_k(t)$.

(A8) For any $1 \leq i \leq n$, $\theta_m \sum_{k=1}^M \|\psi_k\|_\infty^2 = o_p(1)$ and $\gamma_n = \sum_{k=1}^M (\delta_k^V A_{\delta_k^V}\|\psi_k\|_\infty)/(\theta_m^{-1} - A_{\delta_k^V}) \to 0$ as $n \to \infty$.

For given $t = t_{ij}, t_1 = t_{ij_1}$, and $t_2 = t_{ij_2}$, for some i, j, j_1, and j_2, let $g(x; t)$ denote the density function of X_{ij} and let $g_2(x_1, x_2; t_1, t_2)$ denote the density of (X_{ij_1}, X_{ij_2}). Similarly, let $f(z; t)$ and $f_2(z_1, z_2; t_1, t_2)$ denote the densities of Z_{ij} and (Z_{ij_1}, Z_{ij_2}). We assume these densities can be extended to smooth families of densities $g(\cdot; t), f(\cdot; t), t \in \mathcal{T}$, and $g_2(\cdot; t_1, t_2), f_2(\cdot; t_1, t_2), t_1, t_2 \in \mathcal{T}$, that satisfy the following regularity conditions:

(B1.1) $(d^2/dt^2)g(x; t)$ and $(d^2/dt^2)f(z; t)$ exist and are uniformly continuous on $\Re \times \mathcal{T}$.

(B1.2) $(d^2/dt_1^{\ell_1}\, dt_2^{\ell_2})g_2(x_1, x_2; t_1, t_2)$ and $(d^2/dt_1^{\ell_1}\, dt_2^{\ell_2})f_2(z_1, z_2; t_1, t_2)$ exist and are uniformly continuous on $\Re^2 \times \mathcal{T}^2$, for $\ell_1 + \ell_2 = 2, 0 \leq \ell_1, \ell_2 \leq 2$.

The assumptions for kernel functions $\kappa_1 : \Re \to \Re$ and $\kappa_2 : \Re^2 \to \Re$ are as follows. Fourier transforms of $\kappa_1(u)$ and $\kappa_2(u, v)$ are denoted by $\chi_1(t) = \int e^{-iut}\kappa_1(u)\, du$ and $\chi_2(t, s) = \int e^{-(iut+ivs)}\kappa_2(u, v)\, du\, dv$. Then we require the following:

(B2.1) The kernel κ_1 is a compactly supported symmetric density function, $\|\kappa_1\|^2 = \int \kappa_1^2(u)\, du < \infty$, such that this density has finite variance. Furthermore, $\chi_1(t)$ is absolutely integrable, that is, $\int |\chi_1(t)|\, dt < \infty$.

(B2.2) Kernel κ_2 is a compactly supported density function, $\|\kappa_2\|^2 = \int\int \kappa_2^2(u, v)\, du\, dv < \infty$, and κ_2 is a symmetric kernel function with mean 0 and finite variance in both arguments u and v. In addition, $\chi_2(t, s)$ is absolutely integrable, that is, $\int\int |\chi_2(t, s)|\, dt\, ds < \infty$.

APPENDIX C: PROOFS

Proof of Theorem 1

Because W is independent of both S and V, we may factor the probability space $\Omega = \Omega_1 \times \Omega_2$. We write E^* for expectations with regard to the probability measure on Ω_2 only. Given the data for a single subject (i.e., for a specific realization of S and V), this corresponds to fixing a value $\omega_1 \in \Omega_1$. For each fixed ω_1, we note that $R_j = R_j(\omega_1)$ [omitting the index i in R_{ij} in (1) and in the formulas that follow] are mutually independent in Ω_2 for different j, with $E^*(R_j) = 0$ and $\sup_j E^*(R_j^2) < C_1$ for a suitable constant C_1. Combining assumption (A1) with the arguments given in the proof of lemma 3 and theorem 2 of Schuster and Yakowitz (1979) for kernel estimators, which are easily extended to local linear smoothing in the regular fixed-design case, checking that the assumptions for these results are satisfied, we obtain that

$$E^*\Bigl(\sup_{t\in\mathcal{T}} |\hat{S}(t) - S(t)|\Bigr)(\omega_1) = O\Bigl(b_S^2 + \frac{1}{\sqrt{m}b_S}\Bigr)(\omega_1).$$

Studying the dependency of the right side on ω_1, we find that only bounds on $|S^{(\nu)}(\omega_1)|$, $\nu = 0, 1, 2$, and on $|V(\omega_1)|$ play a role. Under (A1), these bounds are uniform in all ω_1; therefore,

$$\sup_{\omega_1\in\Omega_1} E^*\Bigl(\sup_{t\in\mathcal{T}} |\hat{S}(t) - S(t)|\Bigr)(\omega_1) = O\Bigl(b_S^2 + \frac{1}{\sqrt{m}b_S}\Bigr),$$

which implies the result (14).

For the following proofs, we need Lemma C.1. Denote the estimates for process V that would be obtained from input data $\{t_{ij}, Z_{ij}\}$ (i.e., based on the unknown true rather than estimated transformed residuals) by $\tilde{\mu}_V$, $\tilde{G}_V$, $\tilde{\sigma}_W^2$, $\tilde{\psi}_k$, $\tilde{\rho}_k$, and $\tilde{\zeta}_{ik}$, defined analogously to (A.2), (A.3), (A.4), and (A.5) and let

$$\alpha_{nk} = \frac{\delta_k^V A_{\delta_k^V}}{\sqrt{n}h_V^2 - A_{\delta_k^V}} \quad \text{and} \quad \alpha_{nk}^* = \frac{\delta_k^V A_{\delta_k^V}}{\theta_m^{-1} - A_{\delta_k^V}}, \tag{C.1}$$

where $\theta_m = b_S^2 + (\sqrt{m}b_S)^{-1}$ and δ_k^V and $A_{\delta_k^V}$ are as in (B.1) and (B.2).

Lemma C.1. Under assumptions (A2.1)–(A2.3), (A3)–(A5), and (B1.1)–(B2.2),

$$\begin{aligned} &\sup_{t\in\mathcal{T}} |\tilde{\mu}_V(t) - \mu_V(t)| = O_p\Bigl(\frac{1}{\sqrt{n}b_V}\Bigr) \quad \text{and} \\ &\sup_{s,t\in\mathcal{T}} |\tilde{G}_V(s, t) - G_V(s, t)| = O_p\Bigl(\frac{1}{\sqrt{n}h_V^2}\Bigr). \end{aligned} \tag{C.2}$$

Considering eigenvalues ρ_k of multiplicity 1, $\hat{\psi}_k$ can be chosen such that

$$\sup_{t\in\mathcal{T}} |\tilde{\psi}_k(t) - \psi_k(t)| = O_p(\alpha_{nk}) \qquad \text{and}$$
$$|\tilde{\rho}_k - \rho_k| = O_p(\alpha_{nk}), \tag{C.3}$$

where $\alpha_{nk} \to 0$ as $n \to \infty$, k fixed, α_{nk} is defined in (C.1), and the $O_p(\cdot)$ terms in (C.3) hold uniformly over all $1 \le k \le M$. As a consequence of (C.2),

$$\sup_{t\in\mathcal{T}} |\tilde{\sigma}^2_W(t) - \sigma^2_W(t)| = O_p\left(\max\left\{\frac{1}{\sqrt{n}h_V^2}, \frac{1}{\sqrt{n}b_{Q_V}}\right\}\right). \tag{C.4}$$

Under (A1)–(A7) and (B1.1)–(B2.2),

$$\sup_{1\le k\le M} |\tilde{\zeta}_{ik} - \zeta_{ik}| \xrightarrow{p} 0 \qquad \text{and}$$
$$\sup_{t\in\mathcal{T}} \left| \sum_{k=1}^{M} \tilde{\zeta}_{ik}\tilde{\psi}_k(t) - \sum_{k=1}^{\infty} \zeta_{ik}\psi_k(t) \right| \xrightarrow{p} 0, \tag{C.5}$$

as the number M of included components $M = M(n) \to \infty$ as $n \to \infty$.

Proof of Lemma C.1. Results (C.2), (C.3), and (C.5) are immediate from lemma 1 and theorem 1 of Yao and Lee (2006). Note that $\tilde{\sigma}^2_W(t) = \{\hat{Q}_V(t) - \tilde{G}^*_V(t)\}_+$, where the estimate $\hat{Q}_V(t)$ targets $\{G_V(t,t) + \sigma^2_W(t)\}$ and the estimate $\tilde{G}^*_V(t)$ targets $G_V(t,t)$ with the same rate as $\tilde{G}_V(t)$. Analogous to the convergence of $\tilde{\mu}_V(t)$, $\sup_{t\in\mathcal{T}} |\tilde{Q}_V(t) - Q_V(t)| = O_p(n^{-1/2}b_{Q_V}^{-1})$, where b_{Q_V} is the bandwidth used in the smoothing step for $\hat{Q}_V(t)$. From (C.3) and the foregoing result, we may conclude that $\tilde{Q}_V(t) \ge \tilde{G}^*_V(t)$ uniformly in t, with probability converging to 1 as the sample size n increases. This leads to (C.4).

Proof of Theorem 2

Noting that $\theta_m = b_S^2 + (\sqrt{m}b_S)^{-1}$, we find that for $\hat{Z}_{ij}$ [see (12)],

$$E\Big(\sup_{1\le j\le m} |\hat{Z}_{ij} - Z_{ij}|\Big) = O(\theta_m). \tag{C.6}$$

Let $\theta_{im} = \sup_{1\le j\le m} |\hat{Z}_{ij} - Z_{ij}|$. Because linear smoothers, including those based on local polynomial fitting, are weighted averages, because (C.6) implies that $E\theta_{im} \to 0$, and because θ_{im} are iid across all subjects, we have $\bar{\theta}_n = O_p(\theta_m) \xrightarrow{p} 0$, where $\bar{\theta}_n = \sum_{i=1}^n \theta_{im}/n$. It follows that

$$\sup_{t\in\mathcal{T}} |\hat{\mu}_V(t) - \tilde{\mu}_V(t)| = O_p(\theta_m),$$
$$\sup_{s,t\in\mathcal{T}} |\hat{G}_V(s,t) - \tilde{G}_V(s,t)| = O_p(\theta_m), \tag{C.7}$$

and $|\hat{\sigma}^2_W - \tilde{\sigma}^2_W| = O_p(\theta_m)$. In view of (C.2) and (C.4), this implies (15). Analogous to the derivation of (C.3), we conclude that

$$\sup_{t\in\mathcal{T}} |\hat{\psi}_k(t) - \tilde{\psi}_k(t)| = O_p(\alpha^*_{nk}) \qquad \text{and}$$
$$|\hat{\rho}_k - \tilde{\rho}_k| = O_p(\alpha^*_{nk}) \tag{C.8}$$

for sufficiently large n, where α^*_{nk} is as in (C.1), implying (16).

In preparation for the next proof, consider the random variables $\vartheta^{(1)}_{in}$ and $\vartheta^{(2)}_{in}$, using the auxiliary quantities δ^V_k [see (B.1)], $A_{\delta^V_k}$ [see (B.2)], θ_m in the line below (C.1), Δ_n [see (A3)], τ_n [see (A5)], and γ_n [see (A8)],

$$\vartheta^{(1)}_{in} = \theta_m \sum_{k=1}^{M} \|\psi_k\|^2_\infty + \gamma_n \sum_{j=2}^{m} |Z_{ij}|(t_{ij} - t_{i,j-1}) + \sum_{k=1}^{M} \frac{\delta^V_k A_{\delta^V_k}|\zeta_{ik}|}{\theta_m^{-1} - A_{\delta^V_k}}$$

and

$$\vartheta^{(2)}_{in} = \tau_n \left\{ \|V_i\|_\infty \|V_i'\|_\infty \Delta_n + \sum_{j=2}^{m} |W_{ij}|(t_{ij} - t_{i,j-1}) \right\}$$
$$+ \sum_{k=1}^{M} \|\psi_k\|_\infty \left(\frac{1}{\sqrt{n}b_V} + \sqrt{\Delta_n} \right) + \sum_{k=1}^{M} \frac{\delta^V_k A_{\delta^V_k}|\zeta_{ik}|}{\sqrt{n}h_V^2 - A_{\delta^V_k}}$$
$$+ \Delta_n \sum_{k=1}^{M} \|\psi_k\|_\infty \|\psi_k'\|_\infty (\|V_i\|_\infty + \|V_i'\|_\infty)$$
$$+ \sup_{t\in\mathcal{T}} \left| \sum_{k=M+1}^{\infty} \zeta_{ik}\psi_k(t) \right|. \tag{C.9}$$

Proof of Theorem 3

Without loss of generality, assume that $\|\psi_k\|_\infty \ge 1$, which implies that $\tilde{\tau}_n = \sup_{1\le k\le M} \delta^V_k A_{\delta^V_k}/(\sqrt{n}h_V^2 - A_{\delta^V_k}) \to 0$ in view of (A5) and $\tilde{\gamma}_n = \sup_{1\le k\le M} \delta^V_k A_{\delta^V_k}/(\theta_m^{-1} - A_{\delta^V_k}) \to 0$ in view of (A8). For sufficiently large n and m and positive constants C_1 and C_2 that do not depend on i and k, recalling that $\tilde{\zeta}_{ik} = \sum_{j=2}^m (Z_{ij} - \tilde{\mu}_V(t_{ij}))\tilde{\psi}_k(t_{ij}) \times (t_{ij} - t_{i,j-1})$ and using (A8), (C.3), (C.7), and (C.8),

$$\begin{aligned}
&\max_{1\le k\le M} |\hat{\zeta}_{ik} - \tilde{\zeta}_{ik}| \\
&\le \sup_{1\le k\le M} \left\{ \left| \sum_{j=2}^{m} \big(\hat{Z}_{ij} - Z_{ij} + \tilde{\mu}_V(t_{ij}) - \hat{\mu}_V(t_{ij})\big)\hat{\psi}_k(t_{ij})(t_{ij} - t_{i,j-1}) \right| \right. \\
&\qquad \left. + \left| \sum_{j=2}^{m} (Z_{ij} - \tilde{\mu}_V(t_{ij}))\big(\hat{\psi}_k(t_{ij}) - \tilde{\psi}_k(t_{ij})\big)(t_{ij} - t_{i,j-1}) \right| \right\} \\
&\le \Big[\sup_{1\le j\le m} |\hat{Z}_{ij} - Z_{ij}| + \sup_{t\in\mathcal{T}} |\hat{\mu}_V(t) - \tilde{\mu}_V(t)| \Big] |\mathcal{T}| \\
&\qquad \times \Big(\max_{1\le k\le M} \|\psi_k\|_\infty + \tilde{\tau}_n + \tilde{\gamma}_n \Big) \\
&\qquad + \max_{1\le k\le M} \sup_{t\in\mathcal{T}} |\hat{\psi}_k(t) - \tilde{\psi}_k(t)| \\
&\qquad\quad \times \left\{ |\mathcal{T}| \Big(\|\mu_V\|_\infty + \sup_{t\in\mathcal{T}} |\tilde{\mu}_V(t) - \mu_V(t)| \Big) \right. \\
&\qquad\qquad \left. + \sum_{j=2}^{m} |Z_{ij}|(t_{ij} - t_{i,j-1}) \right\} \\
&\le C_1 \theta_{im} \max_{1\le k\le M} \|\psi_k\|_\infty + \tilde{\gamma}_n \left\{ C_2 + \sum_{j=2}^{m} |Z_{ij}|(t_{ij} - t_{i,j-1}) \right\} \\
&\xrightarrow{p} 0,
\end{aligned} \tag{C.10}$$

where we observe that $\sum_{j=2}^m |Z_{ij}|(t_{ij} - t_{i,j-1}) = O_p(1)$ by taking expectations. Analogously to (C.5), we obtain $\max_{1\le k\le M} |\tilde{\zeta}_{ik} - \zeta_{ik}| \xrightarrow{p} 0$, from which (17) follows.

To prove (18), noting that

$$\sup_{t\in\mathcal{T}} \left\{ \left| \sum_{k=1}^{M} \hat{\zeta}_{ik}\hat{\psi}_k(t) - \sum_{k=1}^{\infty} \zeta_{ik}\psi_k(t) \right| \right\}$$
$$\le \sup_{t\in\mathcal{T}} \left\{ \left| \sum_{k=1}^{M} \hat{\zeta}_{ik}\hat{\psi}_k(t) - \sum_{k=1}^{M} \tilde{\zeta}_{ik}\tilde{\psi}_k(t) \right| \right\}$$

$$+ \sup_{t\in\mathcal{T}}\left\{\left|\sum_{k=1}^{M}\tilde{\zeta}_{ik}\tilde{\psi}_k(t) - \sum_{k=1}^{\infty}\zeta_{ik}\psi_k(t)\right|\right\}$$
$$\equiv Q_{i1}(n) + Q_{i2}(n),$$

it is sufficient to show that $Q_{i1}(n) \xrightarrow{p} 0$ and $Q_{i2}(n) \xrightarrow{p} 0$. Analogously to the derivation of (C.5), we have $Q_{i2}(n) \xrightarrow{p} 0$ under (A1)–(A7), and indeed $Q_{i2}(n) = O(\vartheta_{in}^{(2)})$, where the $O(\cdot)$ term holds uniformly in $1 \le i \le n$. Focusing on $Q_{i1}(n)$,

$$Q_{i1}(n) \le \sup_{t\in\mathcal{T}}\left\{\sum_{k=1}^{M}|\hat{\zeta}_{ik} - \tilde{\zeta}_{ik}|\cdot|\hat{\psi}_k(t)| + \sum_{k=1}^{M}|\tilde{\zeta}_{ik}|\cdot|\hat{\psi}_k(t) - \tilde{\psi}_k(t)|\right\}. \quad \text{(C.11)}$$

Similarly to (C.10), the first term on the right side of (C.11) is bounded by

$$C_1\theta_{im}\sum_{k=1}^{M}\|\psi_k\|_\infty^2 + \tilde{\gamma}_n\left\{C_2 + \sum_{j=2}^{m}|Z_{ij}|(t_{ij} - t_{i,j-1})\right\} \xrightarrow{p} 0.$$

The second term on the right side of (C.11) has an upper bound, $O_p\{\sum_{k=1}^{M}\delta_k^V A_{\delta_k^V}E|\zeta_{ik}|/(\theta_m^{-1} - A_{\delta_k^V})\}$. As $E\{\sum_{k=1}^{M}\delta_k^V A_{\delta_k^V}E|\zeta_{ik}|/(\theta_m^{-1} - A_{\delta_k^V})\} \le \sum_{k=1}^{M}\delta_k^V A_{\delta_k^V}\sqrt{\rho_k}/(\theta_m^{-1} - A_{\delta_k^V}) \le \gamma_n$, by observing $\rho_k \to 0$, the second term also converges to 0 in probability, and in fact $Q_{i1}(n) = O_p(\vartheta_{in}^{(1)})$, where the $O_p(\cdot)$ terms are uniform in i. The result (18) follows, that is, $\sup_{t\in\mathcal{T}}|\hat{V}_i(t) - V_i(t)| = O_p(\vartheta_{in}^{(1)} + \vartheta_{in}^{(2)})$, where again the $O_p(\cdot)$ terms are uniform in $1 \le i \le n$.

[Received December 2004. Revised December 2005.]

REFERENCES

Carey, J., Liedo, P., Müller, H. G., Wang, J. L., and Chiou, J. M. (1998), "Relationship of Age Patterns of Fecundity to Mortality, Longevity and Lifetime Reproduction in a Large Cohort of Mediterranean Fruit Fly Females," *Journal of Gerontology–Biological Sciences*, 53, 245–251.

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.

Chiou, J. M., and Müller, H. G. (1999), "Nonparametric Quasi-Likelihood," *The Annals of Statistics*, 27, 36–64.

Courant, R., and Hilbert, D. (1989), *Methods of Mathematical Physics*, New York: Wiley.

Dette, H., and Munk, A. (1998), "Testing Heteroscedasticity in Nonparametric Regression," *Journal of the Royal Statistical Society*, Ser. B, 60, 693–708.

Eubank, R., and Thomas, W. (1993), "Detecting Heteroscedasticity in Nonparametric Regression," *Journal of the Royal Statistical Society*, Ser. B, 55, 145–155.

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

Fan, J. Q., and Yao, Q. W. (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85, 645–660.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986), "Residual Variance and Residual Patterns in Nonlinear Regression," *Biometrika*, 73, 625–633.

Grenander, U. (1950), "Stochastic Processes and Statistical Inference," *Arkiv för Matematik*, 195–276.

Hall, P., Poskitt, D. S., and Presnell, B. (2001), "A Functional Data-Analytic Approach to Signal Discrimination," *Technometrics*, 43, 1–9.

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

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

Kakizawa, Y., Shumway, R. H., and Tanaguchi, M. (1998), "Discrimination and Clustering for Multivariate Time Series," *Journal of the American Statistical Association*, 441, 328–340.

Müller, H. G., and Stadtmüller, U. (1987), "Estimation of Heteroscedasticity in Regression Analysis," *The Annals of Statistics*, 15, 610–625.

Müller, H. G., and Zhao, P. L. (1995), "On a Semiparametric Variance Function Model and a Test for Heteroscedasticity," *The Annals of Statistics*, 23, 946–967.

Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Data Analysis*, New York: Springer-Verlag.

——— (2005), *Functional Data Analysis*, New York: Springer-Verlag.

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

Rice, J. A., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data Are Curves," *Journal of the Royal Statistical Society*, Ser. B, 53, 233–243.

Schuster, E., and Yakowitz, S. (1979), "Contributions to the Theory of Nonparametric Regression, With Application to System Identification," *The Annals of Statistics*, 7, 139–149.

Shumway, R. H. (2002), "Time-Frequency Clustering and Discriminant Analysis," in *Computing Science and Statistics—Geoscience and Remote Sensing, Proceedings of the Interface*, Vol. 34, eds. E. J. Wegman and A. Braverman, Fairfax Station: Interface Foundation of North America, pp. 373–379.

Staniswalis, J. G., and Lee, J. J. (1998), "Nonparametric Regression Analysis of Longitudinal Data," *Journal of the American Statistical Association*, 93, 1403–1418.

Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models and Gauss–Newton Method," *Biometrika*, 61, 439–447.

Yao, F., and Lee, T. C. M. (2006), "Penalized Spline Models for Functional Principal Component Analysis," *Journal of the Royal Statistical Society*, Ser. B, 68, 3–25.

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

Yao, F., Müller, H. G., and Wang, J. L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590.

Yao, Q. W., and Tong, H. (2000), "Nonparametric Estimation of Ratios of Noise to Signal in Stochastic Regression," *Statistica Sinica*, 10, 751–770.

Yu, K., and Jones, M. C. (2004), "Likelihood-Based Local Linear Estimation of the Conditional Variance Function," *Journal of the American Statistical Association*, 99, 139–144.