Functional Variance Processes

Hans-Georg Müller, Ulrich Stadtmüller and Fang Yao

Final Version

Hans-Georg Müller is Professor, Department of Statistics, University of California, Davis, One Shields Ave., Davis, CA 95616 (E-mail: *mueller@wald.ucdavis.edu*). Ulrich Stadtmüller is Professor, Department of Mathematics, Abteilung für Zahlen- und Wahrscheinlichkeitstheorie, Universität 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*). We wish to thank the Editor and Associate Editor for their handling of the paper and two referees for very helpful remarks. Very special thanks are due to a referee for the excellent suggestion to simplify our original approach by introducing a presmoothing step. Research supported in part by NSF grants DMS03-54448 and DMS05-05537.

ABSTRACT

We introduce the notion of a functional variance process to quantify variation in functional data. The functional data are modelled as samples of smooth random trajectories which are 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 to analyze stochastic time-trends in noise variance. As a smooth random process it can be characterized by the eigenfunctions and eigenvalues of its auto-covariance 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 non- 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 to be 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 exists by now a sizable 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,

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

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), the variance function is defined alternatively as $v(x) = E(Y^2|X = x) - [E(Y|X = x)]^2$.

While the variance function traditionally is considered a non-random object that is targeted

by function estimation methods such as kernel or spline smoothing, increasingly data are collected that are of a more complex functional type, 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, or discriminating between different types of trajectories. An excellent overview on functional data analysis is provided in the books by Ramsay and Silverman (1997, 2002). In this article, we aim at extending 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 in 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 has caused the activity. There are two possibilities, explosion or 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 since the data consist of repeated realizations of time courses of seismic activity. While discriminant analysis for functional data focussing 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 at quantifying this random variability motivates us to introduce the concept of a functional variance process.

Since each recorded trajectory is a random process, the notion of a variance function, as described above, 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, in which one encounters curve data with potentially informative variation structure, an extension of the usual modeling approaches that are currently available for functional data analysis is needed. We are aiming at a model 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 when compared to 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 seen 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 analysis is a major tool for the proposed development. It 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 functional principal component 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 1987). Early work on this concept is due to 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 1997; James, Hastie and Sugar 2002; 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 also estimates of individual trajectories

of functional variance processes are introduced. Section 4 is devoted to asymptotic results on the consistency of estimated residuals (Theorem 1), providing the basis for constructing trajectories of functional variance processes, and convergence of estimated eigenfunctions and eigenvalues (Theorem 2) as well as convergence of individual estimated trajectories (Theorem 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.1, assumptions and notations as needed for the proofs in Appendix A.2, and proofs and auxiliary results in Appendix A.3.

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 functional principal component (FPC) scores and eigenfunctions.

The data are generated from a square integrable process S which produces a sample of n independent and identically distributed smooth random trajectories S_i, $i = 1, \ldots, n$. The observed measurements X_{ij} are available on a regular dense grid of support points t_{ij} on the domain $\mathcal{T} = [a_1, a_2]$ and are related to S by

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

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

$$ER_{ij} = 0,\ \mathrm{var}(R_{ij}) = \sigma^2_{Rij} < \infty.$$

Note that the noise R_{ij} within the same subject or item i may be correlated. Throughout this paper, "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 A.2 after (A2.5).

Focusing on the noise R_{ij}, we assume that squared errors R^2_{ij} 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 $V(t)$, i.e., $R^2_{ij} = \exp(V(t_{ij}))\exp(W_{ij})$. As in the case of regression residuals, the squared errors R^2_{ij} can be expected to carry relevant information about the random variation, and the exponential factors convey the nonnegativity restriction. The transformed errors $Z_{ij} = \log(R^2_{ij})$ 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 one hand, and the errors W_{ij} on the other hand. The W_{ij} are assumed to satisfy

$$E(W_{ij}) = 0, \quad \mathrm{var}(W_{ij}) = \sigma^2_W, \quad W_{ij} \perp W_{ik} \quad \text{for} \quad j \neq k, \tag{2}$$

furthermore $W \perp V$, $W \perp S$, where $Q \perp T$ means that r.v.s Q and T are independent.

The decomposition

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

implies

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

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

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

The auto-covariance operator associated with the symmetric kernel G_V, ,

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

is a linear integral operator with kernel G_V, mapping a function $f \in L^2(\mathcal{T})$ to the function $\boldsymbol{G}_V(f) \in L^2(\mathcal{T})$. It has smooth eigenfunctions ψ_k with nonnegative eigenvalues ρ_k, which are assumed to be ordered by size, $\rho_1 \geq \rho_2 \geq \ldots$. 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 functional principal component scores ζ_k, $k \geq 1$. These are random variables with $E\zeta_k = 0$, $\mathrm{var}(\zeta_k) = \rho_k$, that can be represented as

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

Observing (4), given the transformed errors Z_{ij}, one can obtain estimates of μ_V by pooling these errors for all n subjects and smoothing the resulting scatterplot. Furthermore, (2) implies

$$\mathrm{cov}(Z_{ij}, Z_{ik})) = \mathrm{cov}(V_i(t_{ij}), V_i(t_{ik})) = G_V(t_{ij}, t_{ik}),\ j \neq k. \tag{9}$$

Since the covariance kernel G_V is smooth, it can be estimated from the empirical covariances of the Z_{ij}. Here one needs to omit the diagonal since it is contaminated by the white noise error variance σ_W^2. Details on estimating such covariance surfaces can be found in Staniswalis and Lee

(1998) and Yao *et al.* (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 one has 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 $\sum_k \lambda_k < \infty$, $\sum_k \rho_k < \infty$. Take sequences of random variables ξ_k with $E(\xi_k) = 0, \mathrm{var}(\xi_k) = \lambda_k$ and ζ_k with $E(\zeta_k) = 0, \mathrm{var}(\zeta_k) = \rho_k$, where all of these r.v.s are independent. Selecting any smooth functions μ_S and μ_V on $\mathcal{T}$, we then set

$$S(t) = \mu_S(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \qquad V(t) = \mu_V(t) + \sum_{k=1}^{\infty} \zeta_k \psi_k(t). \tag{10}$$

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

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

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, ρ_k to be zero and by using bounded r.v.s ξ_k, 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 proceed to demonstrate below. As for recovering individual trajectories V_i of the functional variance process, according to (7) this requires to obtain the functional principal component scores ζ_k of V,

given in (8). As has been shown in Yao *et al.* (2003), these integrals can be approximated by Riemann sums, substituting $V(t_{ij})$ by $\hat{Z}_{ij}$ and μ_V, ψ_k by estimates $\hat{\mu}_k, \hat{\psi}_k$. Another component of the overall model that is of interest and needs to be determined is $\text{var}(W_{ij}) = \sigma_W^2$.

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 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 in Yao *et al.* (2005), compare also Staniswalis and Lee (1998). We provide only an outline here; for further details regarding the estimation steps we refer to Appendix A.1.

In a first step, following the suggestion of an anonymous reviewer, we smooth the scatterplots $(t_{ij}, X_{ij}), \quad j = 1, \ldots, m$, separately for each trajectory S_i; any of a number of available smoothing methods can be used for this purpose, and also for the other subsequent smoothing steps. If one uses local linear smoothing, as in our implementation, one may apply a different bandwidth $b_{S,i}$ for each trajectory; see Appendices A.1 and A.2 for further details. We selected bandwidths $b_{S,i}$ by cross-validation, individually per subject, which yielded good results in applications, and avoids biases that may arise when using cross-panel smoothing techniques such as pooled cross-validation or an initial functional principal component expansion for smooth processes S. The resulting estimates $\hat{S}_i(t_{ij})$, see (19) below, 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

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

In a second step, we then apply the PART algorithm to the sample of transformed residuals $\hat{Z}_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, that were 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 (20) below. The bandwidth b_V is chosen data-adaptively by cross-validation.

2. This is followed by estimating the smooth covariance surface G_V (5), applying two-dimensional smoothing (21), fitting local planes by weighted least squares to empirical covariances, employing 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 $(t_{ij}, t_{ij'})$, $t_{ij} \neq t_{ij'}$, while the empirical variances obtained at the diagonal of the surface are omitted, as 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 (22).

4. Estimating the variance $\mathrm{var}(W_{ij}) = \sigma_W^2$ (2); this involves a one-dimensional smoothing step along the diagonal of the covariance surface, employing bandwidth b_{Q_W} in the direction of the diagonal, and then obtaining the estimate $\hat{\sigma}_W^2$ as in (25). Again, in our data-adaptive implementation bandwidths b_{Q_W} are chosen by cross-validation.

5. Estimating individual functional principal component scores ζ_j (8) by numerical integration as in (23).

The algorithm also provides an estimate of the necessary number M of functional principal components to approximate processes V, employing one-curve-leave-out cross-validation; see (24). Alternative selectors such as pseudo-AIC and BIC criteria (Yao *et al.* 2005) might be employed as well.

The output consists of estimated mean function $\hat{\mu}_V$, estimated eigenfunctions/eigenvalues

$\hat{\psi}_k, \hat{\rho}_k$, estimated noise variance $\hat{\sigma}_W^2$ and estimated functional principal component scores $\hat{\hat{\zeta}}_{ik}$. According to (7), if a number M of approximating components is chosen by the algorithm, this leads to fitted individual functional variance process trajectories

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

Examples of such estimated trajectories can be found in Figure 3.

4. ASYMPTOTIC RESULTS

In order 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} (3) and the observed transformed residuals $\hat{Z}_{ij}$ (12) that are becoming available after the initial smoothing step which aims at recovering the smooth trajectories S_i. In the following, we refer to bandwidths b_S for smoothing trajectories S_i (see (19)); the bandwidth sequence b_S represents bandwidths $b_{S,i}$ that 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 A.2. Bandwidths b_V, h_V and b_{Q_V} are used in the smoothing steps for $\hat{\mu}_V$ in (20), $\widehat{G}_V(s,t)$ in (21) and $\widehat{Q}_V(t)$ in (25), respectively. 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) can be found in Appendix A.2, while the proofs are in Appendix A.3.

Theorem 1. *Under conditions (A1), (A2), (B1.1) and (B2.1), it holds for smoothed trajectories*

$\widehat{S}_i(t)$ *that*

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

As a consequence of (14), if we apply the PART algorithm to the observed transformed residuals $\widehat{Z}_{ij}$, we expect to obtain consistent estimates of the components of the functional variance process, which is our target. The difficulty is that we do not observe the actual transformed errors Z_{ij} but only the approximate values $\widehat{Z}_{ij}$, 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 estimate $\hat{\mu}_V(t)$ of the mean function $\mu_V(t)$, the estimate $\widehat{G}_V(s,t)$ 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 (20), (21) and (22) and characterize the functional variance process. Consistent estimation of these components validates our approach asymptotically. In addition, consistency of the estimate $\hat{\sigma}_W^2$ (25) of the noise variance σ_W^2 (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*

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

Considering eigenvalues ρ_k of multiplicity one, $\hat{\psi}_k$ 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 (28), using definitions (26) and (27).

Another central result provides consistency for individually estimated trajectories $\widehat{V}_i$ (13) of functional variance trajectories V_i, such as those drawn in Figure 3. We proceed by first establishing consistency of estimates $\hat{\zeta}_{ik}$ (23) of individual functional principal components 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 functional principal components of functional variance processes V that*

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

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

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

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), while for the convergence of (17) in addition the number of included components satisfies $M(n) \to \infty$ 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 $\sup_{t\in\mathcal{T}} |\widehat{V}_i(t) - V_i(t)| = O(\vartheta^{(1)}_{in} + \vartheta^{(2)}_{in})$, where the $O(\cdot)$ terms hold uniformly in $1 \leq i \leq n$, and

$\vartheta_{in}^{(1)}$, $\vartheta_{in}^{(2)}$ are random variables as defined in (36) in Appendix A.3.

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 prior to 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 ($t = 1, \ldots, 1024$), and the secondary shear wave ($t = 1025, \ldots, 2048$).

Ratios and amplitudes of the two components as well as spectral ratios in different frequency bands have been used as features in previous attempts at feature-based discriminant analysis, see for example Kakizawa *et al.* (1998). Shumway (2002) proposed the use of time-varying spectra for classification and clustering of non-stationary 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 functional principal component (FPC) scores, as in (10), using for example the estimation methods described in Rice and Silverman (1991).

Since 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 (24)) 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 the two top panels of Figure 2, 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, while 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 expressed and emphasize somewhat earlier times in comparison to 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 (25). We note that in practice, with discrete data such as the explosions and earthquake data, it may happen that in (12) the term $(X_{ij} - \hat{S}_i(t_{ij}))^2$ is 0, and therefore we added 0.001 to the squared residuals $\hat{R}_{ij}^2$ before taking the log.

The estimates for the trajectories V_i for the same data as 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 comparing against the corresponding panels of Figure 1. An early peak in the variance process trajectories is quite noticeable for the earthquakes, while it is largely absent for the explosions.

Estimates $\hat{\zeta}_{ik}$, $k = 1, 2$; $i = 1, \ldots, 15$ (23) of the first two functional principal component scores (8) of processes V are presented in the right panel of Figure 4 and 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 the left panel of Figure 4. We find that the pattern of the functional principal component 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 scores of second versus first FPC score of S, displayed in the left panel of Figure

4, 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 $(\hat{\zeta}_{i1}, \hat{\zeta}_{i2})$ that are obtained for functional variance processes clearly distinguish explosions and earthquakes; one can draw a line that separates the two groups. In fact, the one-leave-out misclassification error for logistic discriminant analysis based on the scores for S led to seven misclassifications (out of 15 events), while the scores for the functional variance process led to zero 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 $V_i(t)$. As 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 1000 female medflies described in Carey et al (1998) with lifetimes no less than 40 days, and investigated the dynamics of the number of daily eggs laid 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 shape of the egg-laying curves vary quite a bit, but a general feature is a more or less rapid increase in egg-laying activity that is followed by a more protracted decline.

The estimated mean function $\hat{\mu}_V$ (20) for the functional variance processes V and the first four eigenfunctions for these processes are depicted in the two bottom panels of Figure 2. Here four is the number of components chosen by one-curve-leave-out cross-validation (24). As men-

tioned 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 therefore 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, i.e., the mean function and the first eigenfunction appear to be 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 (25) 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 are typically increasing rapidly from 0 up to a large level, and then tend to stay at that level with only slight decline. This seems to imply that the behavior after the initial peak is quite different between processes S and V. While the trajectories of smooth components S are for the most part monotonically declining after the initial peak in egg-laying, the trajectories of the variance processes remain more or less at a constant level and stay elevated, with individual variations.

These findings provide some evidence that the variance structure of these data is not of a simple Poisson type as one could have surmised based on the idea that the data are counts. What we see instead is that as the average counts decrease, their individual variability relative to the mean count increases as flies are aging. 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. While the overall reproductive activity of flies is declining 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 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 functional principal component 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 that has been the traditional target of functional principal components analysis, and that we approximate for our purposes here by an initial smoothing step. Alternatively, the trajectories S_i could be represented in the form of a functional principal component 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 principal component scores of the functional variance process are useful for functional modeling and may serve as input for classification or functional regression.

While 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 (for example in a pre-smoothing step as described in this paper, or alternatively by another functional principal component analysis) will lead to changes in the estimated functional principal component scores and estimated trajectories of the functional variance process. In the application to seismic data, we found that the discriminating ability of the functional principal component scores was not particularly sensitive to these choices. Generally it will depend on the final goal of the analysis how big of a role these choices will play.

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 functional principal components 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 non- and semiparametric regression, functional variance processes may be useful to obtain more efficient functional methodology and for the construction of 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 the modeling of volatility for financial market data.

Appendix

A.1 Estimation procedures

Let $\kappa_1(\cdot)$ and $\kappa_2(\cdot,\cdot)$ be nonnegative univariate and bivariate kernel functions that are used as weights for locally weighted least squares smoothing in one and two dimensions and satisfy assumptions (B2.1) and (B2.2) below. 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 $S_i\,,\ i = 1,\ldots,n$ from data (t_{ij}, X_{ij}), $j = 1,\ldots,m$ with bandwidths $b_{S,i}$ are obtained through minimizing

$$\sum_{j=1}^{m} \kappa_1(\frac{t_{ij}-t}{b_{S,i}})\{X_{ij} - \beta_{i,0} - \beta_{i,1}(t-t_{ij})\}^2 \tag{19}$$

with respect to $\beta_{i,0}$, $\beta_{i,1}$. 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) below.

For the estimation of μ_V, the first step in the PART algorithm, we also employ local linear smoothing, by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \kappa_1(\frac{t_{ij}-t}{b_V})\{\hat{Z}_{ij} - \beta_0 - \beta_1(t-t_{ij})\}^2 \tag{20}$$

with respect to β_0, 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 $G_V(s,t)$ through minimizing

$$\sum_{i=1}^{n} \sum_{1\le j_1\ne j_2\le m} \kappa_2(\frac{t_{ij_1}-s}{h_V}, \frac{t_{ij_2}-t}{h_V})\{\mathrm{G}_i(t_{ij_1}, t_{ij_2}) - f(\beta,(s,t),(t_{ij_1},t_{ij_2}))\}^2, \tag{21}$$

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 $\widehat{G}_V(s,t) = \hat{\beta}_0(s,t)$.

The estimates of $\{\rho_k, \psi_k\}_{k\geq 1}$ correspond to the solutions $\{\hat{\rho}_k, \hat{\psi}_k\}_{k\geq 1}$ of the eigenequations,

$$\int_{\mathcal{T}} \widehat{G}_V(s,t)\hat{\psi}_k(s)ds = \hat{\rho}_k\hat{\psi}_k(t), \tag{22}$$

with orthonormality constraints on $\{\hat{\psi}_k\}_{k\geq 1}$ and positive definiteness constraints (for the latter, see Yao *et al.* (2003)). We use a simple discrete integral approximation to estimate the first M functional principal component scores ζ_{ik} (8),

$$\hat{\zeta}_{ik} = \sum_{j=2}^{m}(\hat{Z}_{ij} - \hat{\mu}_V(t_{ij}))\hat{\psi}_k(t_{ij})(t_{ij} - t_{i,j-1}), \quad i = 1,\ldots,n,\ k = 1,\ldots,M. \tag{23}$$

Let $\hat{\mu}_V^{(-i)}$ and $\hat{\psi}_k^{(-i)}$ be the estimated mean and eigenfunctions after removing the data for the i-th subject. One-curve-leave-out cross-validation aims at minimizing

$$\mathrm{CV}_V(M) = \sum_{i=1}^{n}\sum_{j=1}^{m}\{\hat{Z}_{ij} - \widehat{V}_i^{(-i)}(t_{ij})\}^2, \tag{24}$$

with respect to the number of included components M, where $\widehat{V}_i^{(-i)}(t) = \hat{\mu}_V^{(-i)}(t) + \sum_{k=1}^{M}\hat{\zeta}_{ik}^{(-i)}\hat{\psi}_k^{(-i)}(t)$, and $\hat{\zeta}_{ik}^{(-i)}$ is obtained by (23). 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 $\widehat{G}_V^*(t)$, and a local linear smoother focusing on diagonal values $\{G_V(t,t) + \sigma_W^2\}$ by $\widehat{Q}_V(s)$, 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

$$\hat{\sigma}_W^2 = \frac{1}{|\mathcal{T}_1|}\int_{\mathcal{T}_1}\{\widehat{Q}_V(t) - \widehat{G}_V^*(t)\}_+\,dt, \tag{25}$$

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

A.2 Assumptions and Notations

Processes S and V are assumed to be twice continuously differentiable and to possess the following properties,

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

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

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 $\widehat{S}_i$ (19), $\hat{\mu}_V$ (20), $\widehat{G}_V$ (21) and $\widehat{Q}_V(t)$ in (25). 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, c_2, $0 < c_1 < \inf_i b_{S,i}/b_S \leq \sup_i b_{S,i}/b_S < c_2 < \infty$.

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

(A2.3) $b_V \to 0, \quad b_{Q_V} \to 0, \quad n b_V^4 \to \infty, \quad n b_{Q_V}^4 \to \infty, \quad \text{limsup}_n\, n b_V^6 < \infty, \quad \text{limsup}_n\, n b_{Q_V}^6 < \infty$.

(A2.4) $h_V \to 0, \quad n h_V^6 \to \infty, \quad \text{limsup}_n\, n h_V^8 < \infty$.

(A2.5) $\text{limsup}_n\, n^{1/2} b_V m^{-1} < \infty, \quad \text{limsup}_n\, n^{1/2} b_{Q_V} m^{-1} < \infty, \quad \text{limsup}_n\, n^{1/2} h_V m^{-1} < \infty$.

The time points $\{t_{ij}\}_{i=1,\ldots,n;j=1,\ldots,m}$, 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 below. We assume that for all i and $j = 1, \ldots, m-1$, $t_{ij} < t_{i,j+1}$ 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, c_2 such that all 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 i-th subject may differ from subject to subject, as long as there is a sequence $m \to \infty$ such that for suitable constants c_1, c_2, $0 < c_1 < \inf_i \frac{N_i}{m} < \sup_i \frac{N_i}{m} < c_2 < \infty$; our analysis is focusing on the case where $N_i = m$, so that we refer only to m in the following, while 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 $t \in \mathcal{T}$, i.e.,

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

Background on linear operators in Hilbert space as needed for the following can be found for example in Courant and Hilbert (1989). Define the operator $\big(f \otimes g\big)(h) = \langle f, h\rangle g$, 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 $\boldsymbol{G}_V$ defined in (6) and its estimate $\widehat{\boldsymbol{G}}_V$, generated by kernels G_V (5), respectively $\hat{G}_V$ (21) are Hilbert-Schmidt operators. Let $\mathcal{I}_i = \{j : \rho_j = \rho_i\}$, $\mathcal{I}' = \{i : |\mathcal{I}_i| = 1\}$, where $|\mathcal{I}_i|$ denotes the number of elements in $\mathcal{I}_i$. Let $\boldsymbol{P}_j^V = \sum_{k\in\mathcal{I}_j} \psi_k \otimes \psi_k$, and $\widehat{\boldsymbol{P}}_j^V = \sum_{k\in\mathcal{I}_j} \hat{\psi}_k \otimes \hat{\psi}_k$ denote the true and estimated orthogonal projection operators from H to the subspace spanned by $\{\psi_k : k \in \mathcal{I}_j\}$, respectively

$\{\hat{\psi}_k : k \in \mathcal{I}_j\}$. For fixed j, let

$$\delta_j^V = \frac{1}{2}\min\{|\rho_l - \rho_j| : l \notin \mathcal{I}_j\}, \tag{26}$$

and let $\mathbf{\Lambda}_{\delta_j^V} = \{z \in \mathcal{C} : |z - \rho_j| = \delta_j^V\}$, where $\mathcal{C}$ stands for the set of complex numbers. The resolvents of $\boldsymbol{G}_V$, $\widehat{\boldsymbol{G}}_V$, respectively, are denoted by $\boldsymbol{R}_V$, $\widehat{\boldsymbol{R}}_V$, i.e., $\boldsymbol{R}_V(z) = (\boldsymbol{G}_V - zI)^{-1}$, $\widehat{\boldsymbol{R}}_V(z) = (\widehat{\boldsymbol{G}}_V - zI)^{-1}$.

Let

$$A_{\delta_j^V} = \sup\{\|\boldsymbol{R}_V(z)\|_F : z \in \mathbf{\Lambda}_{\delta_j^V}\} \tag{27}$$

and $M = M(n)$ denote the numbers of components, corresponding to the eigenfunctions that are included to approximate $V(t)$, i.e., $\widehat{V}_i(t) = \hat{\mu}_V(t) + \sum_{m=1}^{M(n)} \hat{\zeta}_{im}\hat{\psi}_m(t)$ (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, i.e., 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 $\sum_{j=1}^{M} \|\psi_j\|_\infty \|\psi_j'\|_\infty = o(m)$.

We note that assumptions (A5), (A6) do not require that eigenfunctions ψ_j or their derivatives are uniformly bounded, rather that their growth is bounded as the index j increases. Defining $\theta_m = b_S^2 + (\sqrt{m}b_S)^{-1}$, processes V are assumed to possess 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 $1 \le i \le n$, $\theta_m \sum_{k=1}^{M} \|\psi_k\|_\infty^2 = o_p(1)$ 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, j_2, let $g(x;t)$ denote the density function of X_{ij}, and $g_2(x_1, x_2; t_1, t_2)$ 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

(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, i.e., $\int |\chi_1(t)|dt < \infty$.

(B2.2) Kernel κ_2 is a compactly supported density function, $\|\kappa_2\|^2 = \int\int \kappa_2^2(u, v)dudv < \infty$, and κ_2 is a symmetric kernel function with zero mean and finite variance in both arguments u and v. In addition, $\chi_2(t, s)$ is absolutely integrable, i.e., $\int\int |\chi_2(t, s)|dtds < \infty$.

A.3 Proofs

Proof of Theorem 1. Since W is independent from 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 below) 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, one obtains

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

Studying the dependency of the r.h.s. on ω_1, one finds 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,

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

which implies the result (14).

For the following proofs, we need Lemma 1 below. Denote the estimates for process V that one would obtain from input data $\{t_{ij}, Z_{ij}\}$, i.e., based on the unknown true (rather than estimated) transformed residuals, by $\tilde{\mu}_V$, $\widetilde{G}_V$, $\tilde{\sigma}_W^2$, $\tilde{\psi}_k$, $\tilde{\rho}_k$ and $\tilde{\zeta}_{ik}$, defined analogously to (20), (21), (22) and (23), respectively, and let

$$\alpha_{nk}=(\delta_k^V A_{\delta_k^V})/(\sqrt{n}h_V^2-A_{\delta_k^V}),\qquad \alpha_{nk}^*=(\delta_k^V A_{\delta_k^V})/(\theta_m^{-1}-A_{\delta_k^V}), \tag{28}$$

where $\theta_m=b_S^2+(\sqrt{m}b_S)^{-1}$, and $\delta_k^V, A_{\delta_k^V}$ are as in (26), (27).

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

$$\sup_{t\in\mathcal{T}}|\tilde{\mu}_V(t)-\mu_V(t)|=O_p(\frac{1}{\sqrt{n}b_V}),\qquad \sup_{s,t\in\mathcal{T}}|\widetilde{G}_V(s,t)-G_V(s,t)|=O_p(\frac{1}{\sqrt{n}h_V^2}). \tag{29}$$

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

$$\sup_{t\in\mathcal{T}}|\tilde{\psi}_k(t)-\psi_k(t)|=O_p(\alpha_{nk}),\qquad |\tilde{\rho}_k-\rho_k|=O_p(\alpha_{nk}), \tag{30}$$

where $\alpha_{nk}\to 0$ as $n\to\infty$, k fixed, α_{nk} is defined in (28), and the $O_p(\cdot)$ terms in (30) hold uniformly over all $1\le k\le M$. As a consequence of (29),

$$\sup_{t\in\mathcal{T}}|\tilde{\sigma}_W^2(t)-\sigma_W^2(t)|=O_p(\max\{\frac{1}{\sqrt{n}h_V^2},\frac{1}{\sqrt{n}b_{Q_V}}\}). \tag{31}$$

Under (A1)-(A7) and (B1.1)-(B2.2),

$$\sup_{1\le k\le M}|\tilde{\zeta}_{ik}-\zeta_{ik}|\xrightarrow{p}0, \qquad \sup_{t\in\mathcal{T}}|\sum_{k=1}^{M}\tilde{\zeta}_{ik}\tilde{\psi}_k(t)-\sum_{k=1}^{\infty}\zeta_{ik}\psi_k(t)|\xrightarrow{p}0, \tag{32}$$

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

Proof of Lemma 1. Results (29), (30) and (32) are immediate from Lemma 1 and Theorem 1 of Yao and Lee (2005). Note that $\tilde{\sigma}_W^2(t)=\{\widehat{Q}_V(t)-\widetilde{G}_V^*(t)\}_+$, where the estimate $\widehat{Q}_V(t)$ targets $\{G_V(t,t)+\sigma_W^2(t)\}$ and the estimate $\widetilde{G}_V^*(t)$ targets $G_V(t,t)$ with the same rate as $\widetilde{G}_V(t)$. 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 (30) and the above result, we may conclude $\widetilde{Q}_V(t)\ge\widetilde{G}_V^*(t)$ uniformly in t, with probability converging to 1 as the sample size n increases. This leads to (31).

Proof of Theorem 2. Noting $\theta_m=b_S^2+(\sqrt{m}b_S)^{-1}$, we find for $\hat{Z}_{ij}$ (12) that

$$E\left(\sup_{1\le j\le m}|\hat{Z}_{ij}-Z_{ij}|\right)=O(\theta_m). \tag{33}$$

Let $\theta_{im}=\sup_{1\le j\le m}|\widehat{Z}_{ij}-Z_{ij}|$. As linear smoothers, including those based on local polynomial fitting, are weighted averages, as (33) implies that $E\theta_{im}\to 0$ and as θ_{im} are i.i.d. across all subjects, one has $\bar{\theta}_n=O_p(\theta_m)\xrightarrow{p}0$, where $\bar{\theta}_n=\sum_{i=1}^n\theta_{im}/n$. It follows that

$$\sup_{t\in\mathcal{T}}|\hat{\mu}_V(t)-\tilde{\mu}_V(t)|=O_p(\theta_m), \quad \sup_{s,t\in\mathcal{T}}|\widehat{G}_V(s,t)-\widetilde{G}_V(s,t)|=O_p(\theta_m), \tag{34}$$

and $|\hat{\sigma}_W^2-\tilde{\sigma}_W^2|=O_p(\theta_m)$. In view of (29) and (31), this implies (15). Analogous to the derivation of (30), we conclude that

$$\sup_{t\in\mathcal{T}}|\hat{\psi}_k(t)-\tilde{\psi}_k(t)|=O_p(\alpha_{nk}^*), \quad |\hat{\rho}_k-\tilde{\rho}_k|=O_p(\alpha_{nk}^*) \tag{35}$$

for sufficiently large n, where α_{nk}^* is as in (28), implying (16).

In preparation for the next proof, consider the random variables $\vartheta_{in}^{(1)}$ and $\vartheta_{in}^{(2)}$, using the auxiliary quantities δ_k^V(26), $A_{\delta_k^V}$ (27), θ_m in the line below (28), Δ_n (A3), τ_n (A5), and γ_n (A8),

$$\begin{aligned}
\vartheta_{in}^{(1)} &= \theta_m \sum_{k=1}^{M} \|\psi_k\|_\infty^2 + \gamma_n \sum_{j=2}^{m} |Z_{ij}|(t_{ij} - t_{i,j-1}) + \sum_{k=1}^{M} \frac{\delta_k^V A_{\delta_k^V} |\zeta_{ik}|}{\theta_m^{-1} - A_{\delta_k^V}}, \\
\vartheta_{in}^{(2)} &= \tau_n \{\|V_i\|_\infty \|V_i'\|_\infty \Delta_n + \sum_{j=2}^{m} |W_{ij}|(t_{ij} - t_{i,j-1})\} + \sum_{k=1}^{M} \|\psi_k\|_\infty (\frac{1}{\sqrt{n} b_V} + \sqrt{\Delta_n}) \\
&\quad + \sum_{k=1}^{M} \frac{\delta_k^V A_{\delta_k^V} |\zeta_{ik}|}{\sqrt{n} h_V^2 - A_{\delta_k^V}} + \Delta_n \sum_{k=1}^{M} \|\psi_k\|_\infty \|\psi_k'\|_\infty (\|V_i\|_\infty + \|V_i'\|_\infty) + \sup_{t \in \mathcal{T}} |\sum_{k=M+1}^{\infty} \zeta_{ik} \psi_k(t)| \quad (36)
\end{aligned}$$

Proof of Theorem 3. Without loss of generality, assume that $\|\psi_k\|_\infty \geq 1$, which implies that $\tilde{\tau}_n = \sup_{1 \leq k \leq M} \delta_k^V A_{\delta_k^V} / (\sqrt{n} h_V^2 - A_{\delta_k^V}) \to 0$ in view of (A5) and $\tilde{\gamma}_n = \sup_{1 \leq k \leq M} \delta_k^V A_{\delta_k^V} / (\theta_m^{-1} - A_{\delta_k^V}) \to 0$ 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 $\tilde{\zeta}_{ik} = \sum_{j=2}^{m} (Z_{ij} - \tilde{\mu}_V(t_{ij})) \tilde{\psi}_k(t_{ij})(t_{ij} - t_{i,j-1})$ and using (A8), (30), (34) and (35),

$$\begin{aligned}
\max_{1 \leq k \leq M} |\hat{\zeta}_{ik} - \tilde{\zeta}_{ik}| &\leq \sup_{1 \leq k \leq M} \{ |\sum_{j=2}^{m} (\widehat{Z}_{ij} - Z_{ij} + \tilde{\mu}_V(t_{ij}) - \hat{\mu}_V(t_{ij})) \hat{\psi}_k(t_{ij})(t_{ij} - t_{i,j-1})| \\
&\qquad + |\sum_{j=2}^{m} (Z_{ij} - \tilde{\mu}_V(t_{ij}))(\hat{\psi}_k(t_{ij}) - \tilde{\psi}_k(t_{ij}))(t_{ij} - t_{i,j-1})| \} \\
&\leq [\sup_{1 \leq j \leq m} |\widehat{Z}_{ij} - Z_{ij}| + \sup_{t \in \mathcal{T}} |\hat{\mu}_V(t) - \tilde{\mu}_V(t)|] |\mathcal{T}| (\max_{1 \leq k \leq M} \|\psi_k\|_\infty + \tilde{\tau}_n + \tilde{\gamma}_n) \\
&\quad + \max_{1 \leq k \leq M} \sup_{t \in \mathcal{T}} |\hat{\psi}_k(t) - \tilde{\psi}_k(t)| \{ |\mathcal{T}| (\|\mu_V\|_\infty + \sup_{t \in \mathcal{T}} |\tilde{\mu}_V(t) - \mu_V(t)|) + \sum_{j=2}^{m} |Z_{ij}|(t_{ij} - t_{i,j-1}) \} \\
&\leq C_1 \theta_{im} \max_{1 \leq k \leq M} \|\psi_k\|_\infty + \tilde{\gamma}_n \{ C_2 + \sum_{j=2}^{m} |Z_{ij}|(t_{ij} - t_{i,j-1}) \} \xrightarrow{p} 0, \quad (37)
\end{aligned}$$

where we observe that $\sum_{j=2}^{m} |Z_{ij}|(t_{ij} - t_{i,j-1}) = O_p(1)$ by taking expectations. Analogously to (32), we obtain $\max_{1 \leq k \leq M} |\tilde{\zeta}_{ik} - \zeta_{ik}| \xrightarrow{p} 0$, whence (17) follows.

To prove (18), noting

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

it is sufficient to show that $Q_{i1}(n) \xrightarrow{p} 0$ and $Q_{i2}(n) \xrightarrow{p} 0$. Analogously to the derivation of (32), one has $Q_{i2}(n) \xrightarrow{p} 0$ under (A1)-(A7), and indeed $Q_{i2}(n) = O(\vartheta_{in}^{(2)})$, where the $O(\cdot)$ term holds uniformly in $1 \leq i \leq n$. Focusing on $Q_{i1}(n)$,

$$Q_{i1}(n) \leq \sup_{t \in \mathcal{T}} \{ \sum_{k=1}^{M} |\hat{\zeta}_{ik} - \tilde{\zeta}_{ik}| \cdot |\hat{\psi}_k(t)| + \sum_{k=1}^{M} |\tilde{\zeta}_{ik}| \cdot |\hat{\psi}_k(t) - \tilde{\psi}_k(t)| \}. \tag{38}$$

Similarly to (37), the first term on the right hand side of (38) is bounded by

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

The second term on the right hand side of (38) 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 $E\{\sum_{k=1}^{M} \delta_k^V A_{\delta_k^V} E|\zeta_{ik}|/(\theta_m^{-1} - A_{\delta_k^V})\} \leq \sum_{k=1}^{M} \delta_k^V A_{\delta_k^V} \sqrt{\rho_k}/(\theta_m^{-1} - A_{\delta_k^V}) \leq \gamma_n$ by observing $\rho_k \to 0$, the second term also converges to zero 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, i.e., $\sup_{t \in \mathcal{T}} |\widehat{V}_i(t) - V_i(t)| = O_p(\vartheta_{in}^{(1)} + \vartheta_{in}^{(2)})$, where again the $O_p(\cdot)$ terms are uniform in $1 \leq i \leq n$.

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.

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

Chiou, J.M. and Müller, H.G. (1999), "Nonparametric quasi-likelihood," *Annals of Statistics*, 27, 36-64.

Dette, H. and Munk, A. (1998), "Testing heteroscedasticity in nonparametric regression," *Jour-*

nal of the Royal Statistical Society Series B, 60, 693-708.

Eubank, R. and Thomas, W. (1993), Detecting heteroscedasticity in nonparametric regression," *Journal of the Royal Statistical Society Series B*, 55, 145-155.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman and 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," *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 Associatoion*, 441, 328-340.

Müller, H.G. and Stadtmüller, U. (1987), " Estimation of heteroscedasticity in regression analysis," *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," *Annals of Statistics* 23, 946-967.

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

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

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 Series B*, 53, 233-243.

Schuster, E. and Yakowitz, S. (1979), "Contributions to the theory of nonparametric regression, with application to system identification," *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, 373-379.

Staniswalis, J. G. and Lee, J. J. (1998), "Nonparametric regression analysis of longitudinal data," *Journal of the American Statistical Associatoion*, 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. (2005), "Penalized spline models for functional principal component analysis," *Journal of the Royal Statistical Society Series B*, in press.

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 Associatoion*, 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 Associatoion*, 99, 139-144.

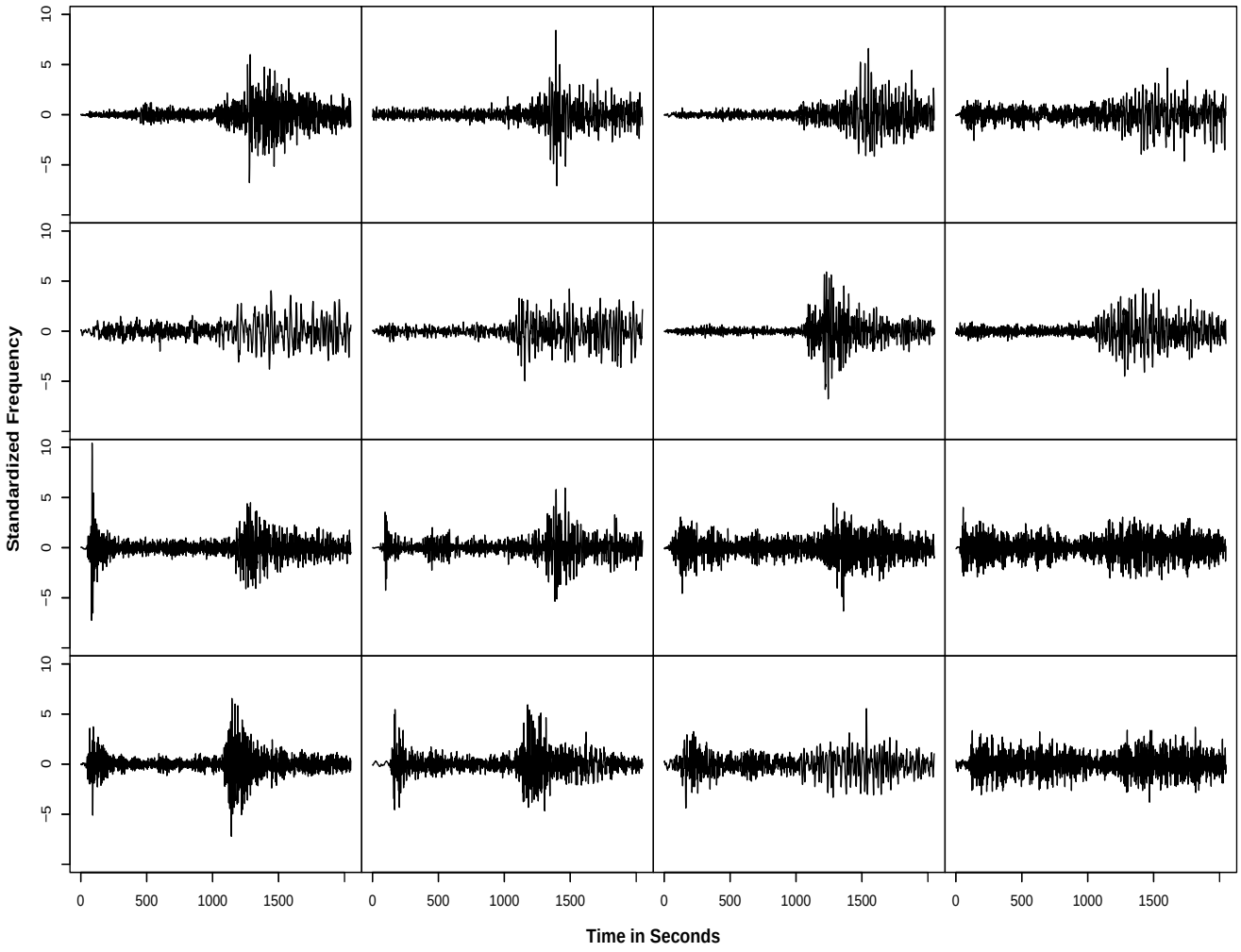

Figure 1: Data for eight standardized explosions (first two rows), seven standardized earthquakes (last two rows except for the bottom right panel) and one unknown event (bottom right panel) (first earthquake out of eight earthquakes not shown).

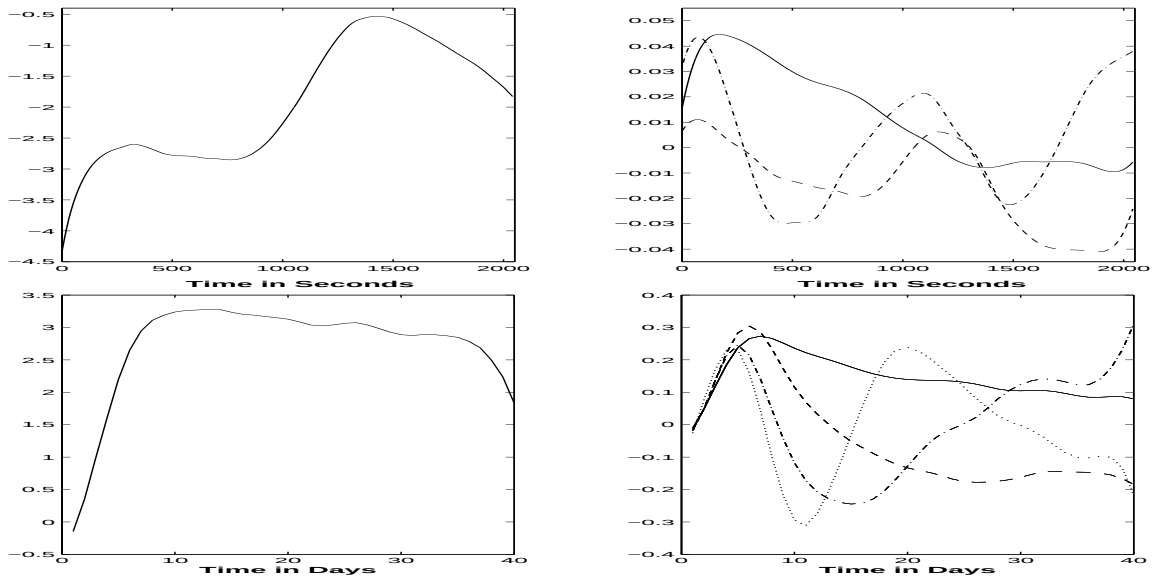

Figure 2: Components of functional variance processes for the earthquakes and explosions data (top panels) and for the egg-laying data (bottom panels). Left panels: Smooth estimate of the mean function for the functional variance process $V(t)$ for earthquakes and explosions data (top left) and egg-laying data (bottom left). Top right: Smooth estimates of the first (solid), second (dashed) and third (dash-dot) eigenfunctions of $V(t)$ for the earthquakes and explosions data, accounting for 62.8%, 23.6% and 7.8% of total variation. Bottom right: Smooth estimates of the first (solid), second (dashed), third (dash-dot) and fourth (dotted) eigenfunctions of $V(t)$ for egg-laying data, accounting for 48.3%, 21.0%, 11.6% and 6.7% of total variation.

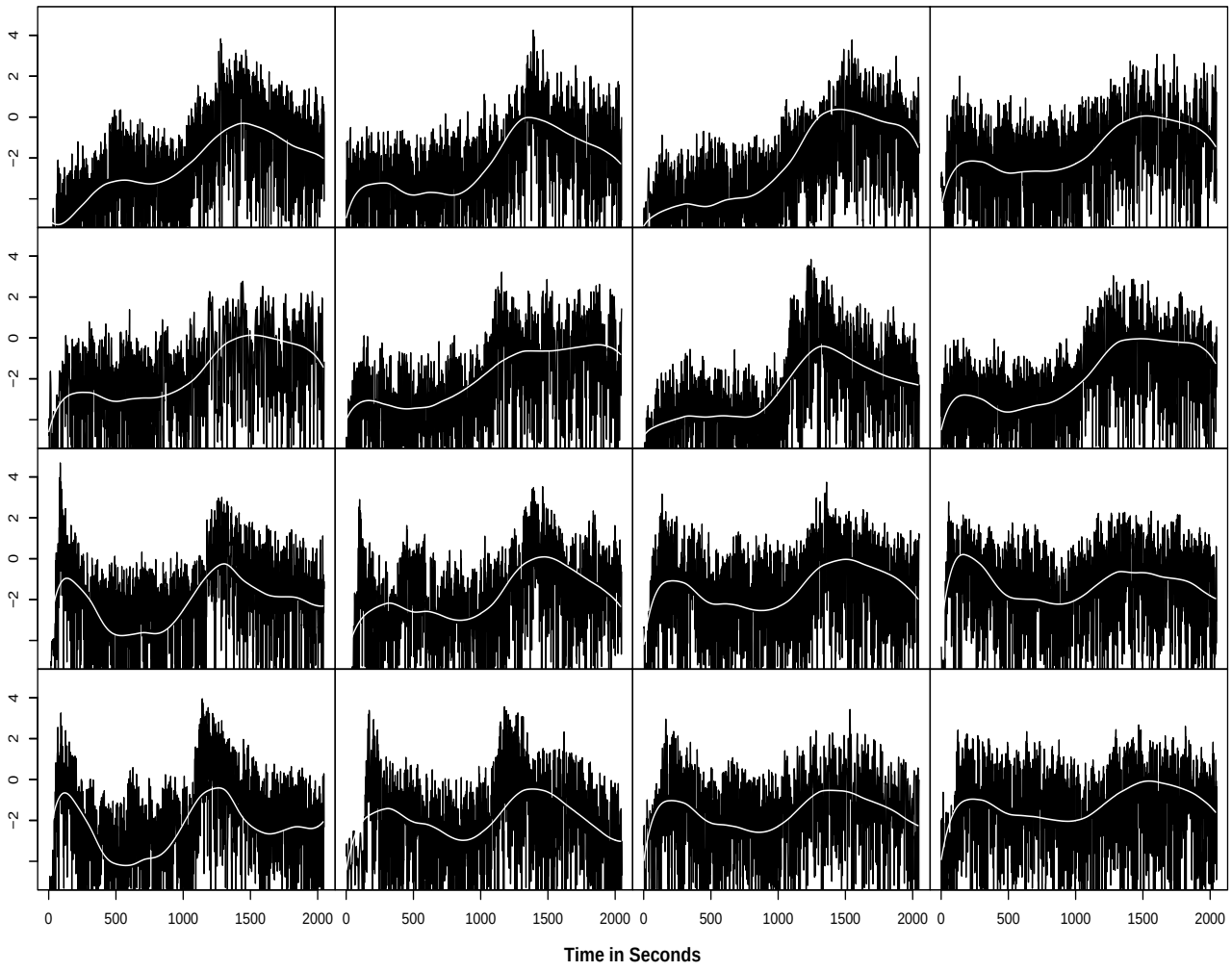

Figure 3: Observed values of $Z_{ij} = \log(R_i^2(t_{ij}))$ (black) and estimated trajectories of functional variance processes $\hat{V}_i(t)$ (13) (white) for eight explosions (first two rows), seven earthquakes (last two rows except for the bottom right panel) and one unknown event (bottom right panel) (first earthquake out of eight earthquakes not shown; order of earthquakes and explosions is as in Figure 1).

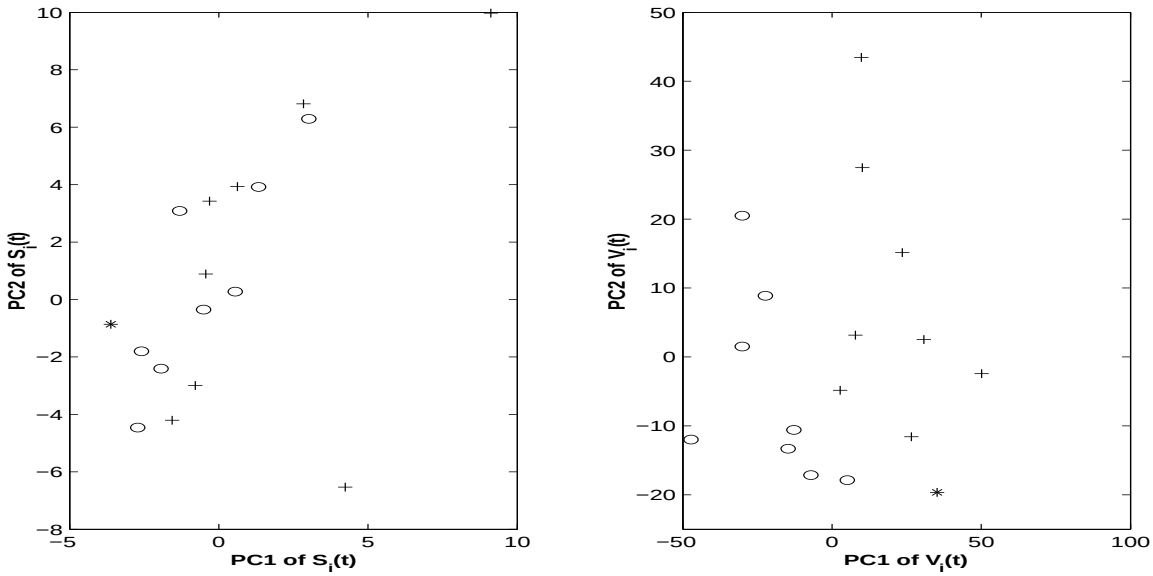

Figure 4: Representation of the first two estimated functional principal component scores, PC2 vs PC1, obtained for the smooth processes S_i (left panel) and for the functional variance processes V_i (right panel), for earthquakes and explosions data. Earthquakes are denoted by plus signs, explosions by circles, and the unknown event by an asterisk.

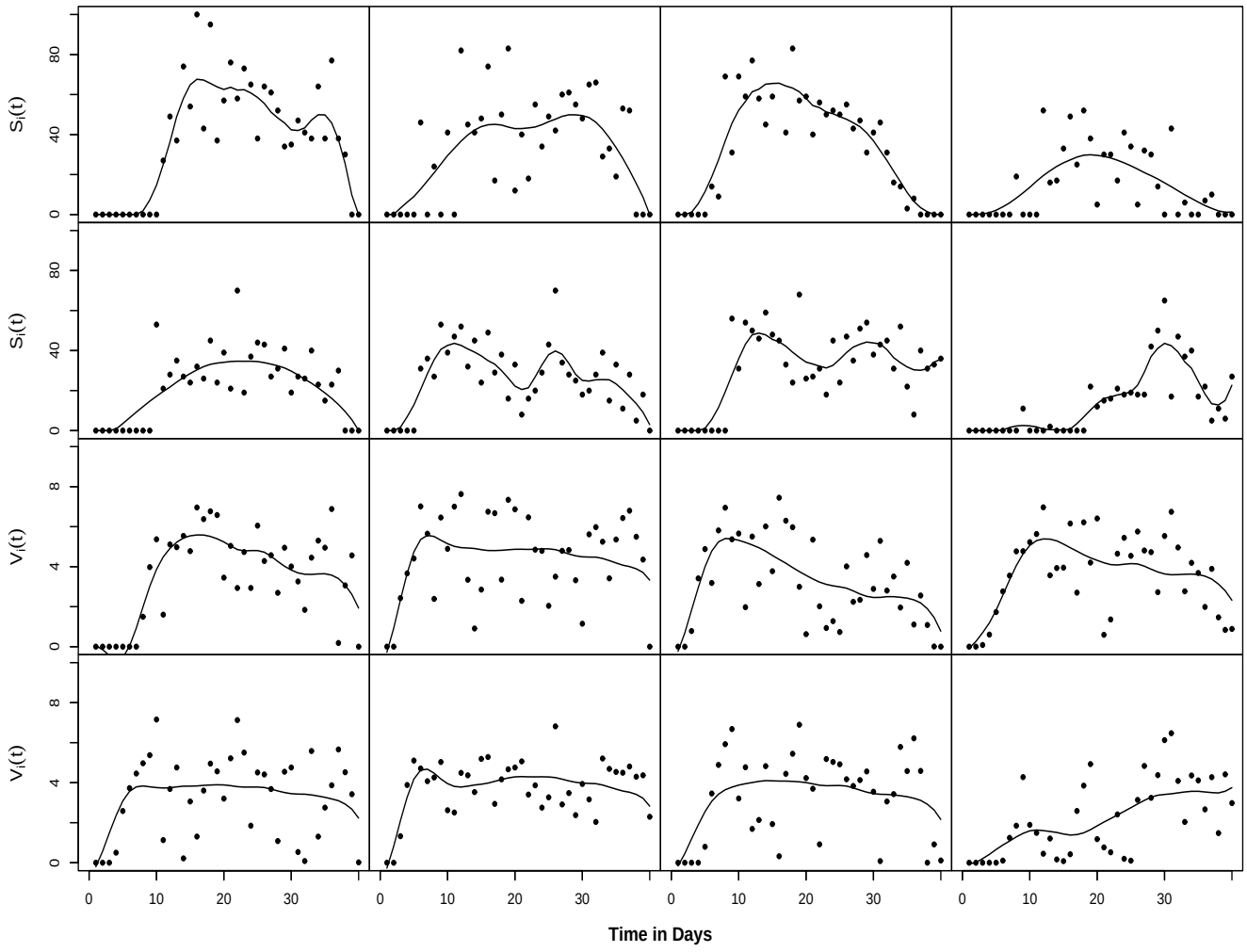

Figure 5: Observed egg-laying counts (dots) and smoothed individual trajectories $\widehat{S}_i$ for eight medflies for the first 40 days of age, where the bandwidths are chosen by leave-one-out cross-validation for each subject (top two rows). Observed values for $Z_{ij} = \log(R^2(t_{ij}))$ (dots) and estimated smooth random trajectories $\widehat{V}_i$ (13) of the functional variance process for the same eight medflies (bottom two rows).