Stat Biosci (2017) 9:582–604
https://doi.org/10.1007/s12561-015-9137-5

Quantifying Infinite-Dimensional Data: Functional Data Analysis in Action

Kehui Chen[1,2] · **Xiaoke Zhang**[3] · **Alexander Petersen**[4] · **Hans-Georg Müller**[4]

Received: 20 June 2015 / Revised: 04 September 2015 / Accepted: 30 October 2015 /
Published online: 20 November 2015

Abstract Functional data analysis (FDA) is concerned with inherently infinite-dimensional data objects and therefore can be viewed as part of the methodology for big data. The size of functional data may vary from terabytes as encountered in functional magnetic resonance imaging (fMRI) and other applications in brain imaging to just a few kilobytes in longitudinal data with small or modest sample sizes. In this contribution, we highlight some applications of FDA methodology through various data illustrations. We briefly review some basic computational tools that can be used to accelerate implementations of FDA methodology. The analyses presented in this paper illustrate the principal analysis by conditional expectation (PACE) package for FDA, where our applications include both relatively simple and more complex functional data from the biomedical sciences. The data we discuss range from functional data that result from daily movement profile tracking and that are modeled as repeatedly observed functions per subject, to medfly longitudinal behavior profiles, where the goal is to predict remaining lifetime of individual flies. We also discuss the quantification of connectivity of fMRI signals that is of interest in brain imaging

Dedicated to the memory of Bitao Liu.

Bitao Liu graduated with a Ph.D. in statistics from UC Davis in 2008 on topics in functional data analysis and made substantial contributions to the PACE package. She worked at Affymetrix and suffered a premature and unexpected death in October 2014.

✉ Kehui Chen
khchen@pitt.edu

1 Department of Statistics, University of Pittsburgh, Pittsburgh, USA

2 Department of Psychiatry, University of Pittsburgh, Pittsburgh, USA

3 Department of Applied Economics and Statistics, University of Delaware, Newark, USA

4 Department of Statistics, University of California, Davis, USA

and the prediction of continuous traits from high-dimensional SNPs in genomics. The methods of FDA that we demonstrate for these analyses include functional principal component analysis, functional regression and correlation, the modeling of dependent functional data, and the stringing of high-dimensional data into functional data and can be implemented with the PACE package.

Keywords Functional principal components · Functional regression · Repeated functional data · PACE · Medfly activity profiles · SNPs · Connectivity in fMRI · High-dimensional data · Stringing

1 Introduction

Functional data analysis (FDA) is an area of statistics where one studies models and analysis methods for data recorded over a continuum for each subject from a sample of subjects. Equivalently, it can be described as the study of a sample of trajectories or time courses. A feature that distinguishes FDA from time series methodology is that in FDA one considers repeated observations of the time courses and does not rely on any stationarity assumptions, whereas in time series analysis one typically has only one realization that is assumed to be stationary, although the boundaries between these fields are increasingly blurry. FDA is particularly suited for the analysis of time dynamic and longitudinal data that are abundantly found in biomedical applications. There are a number of books and reviews available on FDA [11,17,19,23–25,31].

In general, any time-dependent data that are repeatedly observed for many independent individuals or units can be analyzed by FDA methods. A basic paradigm of FDA is that the observed data correspond to or are derived from an independent identically distributed random sample of a stochastic process that generates the observed data; however, dependencies between the realizations of the stochastic process can also be incorporated. The underlying stochastic process is usually assumed to be smooth over a continuum, is usually assumed to lie in the space L^2 or sometimes in a reproducing kernel Hilbert space or constrained subspace, and is the target of interest in FDA.

A basic problem is that the smooth underlying process rarely is fully observed, and the available discrete observations that are thought to be generated by the process are often noisy. In some cases, the data are also sparsely observed, a frequently encountered scenario for longitudinal data. Over the last decade, various FDA methods, blending stochastic process theory, smoothing methods, and multivariate techniques, have been developed for increasingly complex types of functional data. This includes flexible methods to address sparse data, scenarios where one observes multiple or repeated functions per unit or subject, or scenarios where functions are part of a time series [1]. These methods have been utilized successfully for numerous applied problems, where inferring the structure of repeatedly observed trajectories and their relationship with covariates often leads to insights into the underlying dynamics of time-dependent processes.

Functional data are thought to be derived from the smooth realizations of an underlying stochastic process, which is an inherently infinite-dimensional data object, and in this sense functional data are part of big data. A core principle when dealing with high-

dimensional or infinite-dimensional data is dimension reduction. In some instances, functional data are not only complex but also large. An example is data originating from brain imaging such as PET and EEG/MEG signals. For example, in fMRI the blood oxygenation level-dependent (BOLD) signals may be recorded over 240 time points at 100,000 voxels per subject for say $n = 1000$ individuals. With pre-processed versions, such data can quickly reach into the terabyte range. Another area where truly large functional data arise includes data recordings from sensors that monitor certain variables of interest and produce continuous time recordings, such as recordings from weather stations and sensors that monitor the functioning of technical equipment, or data generated by mobile tracking devices that are integrated into wearable or portable electronic devices and can be used to monitor exercise levels, health, and behavior. The analysis of such complex and high-dimensional time course data is a challenging task, and FDA methodology is expected to have a major impact.

In this article, we illustrate some applications of the PACE package, which has been designed to implement various FDA methods, including FDA for both sparsely and densely sampled random trajectories and repeatedly observed functional data. PACE is based on the Principal Analysis by Conditional Expectation (PACE) algorithm [38] and is geared towards the analysis of data that have been generated by a sample of underlying (but often not fully observed) random trajectories. It does not rely on pre-smoothing of trajectories, which is problematic if functional data are sparsely sampled. For functional data that are densely sampled on a regular grid, simple cross-sectional averaging to obtain mean and autocovariance functions of the underlying stochastic process is often a good choice and is an option included in PACE.

In addition to functional principal component analysis (FPCA), PACE provides options to implement various models for functional regression and correlation, for functional conditional distribution and quantile analysis, for functional manifold analysis, for the analysis of linear and non-linear empirical dynamics, and for other techniques such as time synchronization, curve warping, and curve clustering. The development of PACE has been supported by various NSF grants and it is written in Matlab, while an R version is currently under development. The current version is available at http://anson.ucdavis.edu/~mueller/data/pace.html.

In the following, we illustrate some basic and popular FDA methods as implemented in PACE through data applications and discuss methodological and computational challenges for large functional data. Among these illustrations, we also include the stringing of high-dimensional data into functional data, which is also part of the functional methodology implemented in PACE.

2 Functional Principal Component Analysis

Due to the infinite dimensionality of the underlying random trajectories that are the objects of interest in FDA, practically feasible approaches must include some form of dimensionality reduction. This is usually achieved by expanding the underlying random process X into a basis of the function space, often considered to be $L^2(T)$, where T is the domain of the random functions, usually a finite interval. The random functions are then typically reduced to the Fourier coefficients of this basis, of which

one takes a finitely truncated sequence to represent the functions. A common choice for the basis is pre-specified orthonormal trigonometric or polynomial basis functions. An alternative is the choice of a data based orthonormal system that has some optimality properties. A common data-adaptive orthonormal basis selection is obtained via FPCA, which aims at the basis that consists of the orthonormal eigenfunctions of the autocovariance operator of the underlying stochastic process [13,14,33].

The projections of the centered processes on the first K eigenfunctions then explain most of the variance of the underlying process among all projections on K components. This optimal dimension reduction feature provides motivation for using FPCA, which provides a foundation for many other methods of FDA. In practical applications, FPCA has turned out to be very successful to the extent that it has become the most popular technique in FDA.

Formally, for a square-integrable random process $X(t)$, $t \in \mathcal{T} \subset \mathcal{R}$, with mean $\mu(t)$ and covariance function $G(s,t)$, the autocovariance operator is

$$(Af)(t) = \int_{s\in\mathcal{T}} f(s)G(s,t)\,\mathrm{d}s,$$

with orthonormal eigenfunctions ϕ_k and ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. The well-known Karhunen–Loève expansion then gives the representation

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \tag{1}$$

where $\{\xi_k,\ k \geq 1\}$ is a sequence of uncorrelated random variables, with $E(\xi_k) = 0$ and $\mathrm{var}(\xi_k) = \lambda_k$, and functional principal components (FPCs) are given by

$$\xi_k = \int_{t\in\mathcal{T}} \Big(X(t) - \mu(t)\Big)\phi_k(t)\mathrm{d}t. \tag{2}$$

Stochastically, X can be represented by the sequence of scores $\{\xi_1, \xi_2, \ldots\}$. For any fixed K, the first K terms in (1) yield the best K-dimensional linear approximation for $X(t)$ in $L_2(\mathcal{T})$, i.e., it is the unique linear representation which explains the highest fraction of variance in the data with a given number of components, where $\sum_{k=1}^{K} \lambda_k$ is the amount of total variation or process variance that is explained by the first K FPCs.

In practical applications, one almost never has fully observed functions without noise, rather needs to assume Y_{ij} from a data model

$$Y_{ij} = X_i(t_{ij}) + \epsilon_{ij},\ \ 1 \leq i \leq n,\ 1 \leq j \leq N_i, \tag{3}$$

where one often assumes that the ϵ_{ij} are zero-mean i.i.d. measurement errors, with $\mathrm{var}(\epsilon_{ij}) = \sigma^2$, independent of all other random components. Here the t_{ij} are either recorded on a dense grid, $t_{i1}, \ldots, t_{iN_i}$, with $N_i \to \infty$, or they are random times and their number N_i per subject is random and finite. The first case is referred to as *Dense Design* and it applies to many functional data with dense recordings such as monitoring or sensor data, while the second case is usually referred to as *Sparse Design* and it is

also commonly encountered, for example in the case of irregularly spaced longitudinal data as they abound in biomedical applications.

If the recording points t are densely and regularly spaced, i.e., $t_{ij} = t_j$, the standard approach is to use an empirical estimator by averaging the data Y_{ij} over the subject index i for the n subjects and interpolating between design points. This scheme is also applicable to dense irregular designs by adding a pre-smoothing step and then sampling the smoothed functions at a dense regular grid to obtain regularly spaced observations of the underlying functions. If the measurement times t_{ij}, where observations are made, are sparsely and randomly spaced, cross-sectional empirical estimators are not an option, neither is pre-smoothing of individual trajectory data. In this situation, a preferred approach is to estimate the mean function μ by smoothing the pooled data [38], with a local linear smoother. Employing kernel functions κ and smoothing bandwidths h, resp. b, defining $\kappa_h(x) = \frac{1}{h}\kappa(\frac{x}{h})$, this leads to $\hat{\mu}(s,t) = \hat{a}_0$, where

$$(\hat{a}_0, \hat{a}_1) = \arg\min \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{ \left[Y_{ij} - a_0 - a_1(t_{ij} - t) \right]^2 \times \kappa_h(t_{ij} - t) \right\}.$$

In order to obtain a consistent estimate of $G(s,t)$, one can pool products for pairs of observations from the same subject and then implement a two-dimensional local linear smoothing step to obtain $\hat{G}(t_1, t_2) = \hat{a}_0$, where

$$\begin{aligned}(\hat{a}_0, \hat{a}_1, \hat{a}_2) = \arg\min \sum_{i=1}^{n} \sum_{j \neq l} \Big\{ \Big[\big(Y_{ij} - \hat{\mu}(t_{ij})\big)\big(Y_{il} - \hat{\mu}(t_{il})\big) - a_0 - a_1 (t_{ij} - t_1) \\ - a_2(t_{il} - t_2) \Big]^2 \times \kappa_b(t_{ij} - t_1)\kappa_b(t_{il} - t_2) \Big\}.\end{aligned}$$

Denoising is achieved by separating out the diagonal [34].

An additional step is to project the initial smoothed covariance estimates on the space of non-negative definite covariance surfaces [15] to ensure that the resulting covariance surfaces are non-negative definite. Under appropriate regularity conditions and considering uniform convergence, these smoothing estimators achieve a convergence rate of $(\log n/n)^{1/2}$ rate for dense data. If in the sparse design case the number of points sampled for each subject N_i is bounded, the rate for the mean function turns out to be $O(h^2 + [\log n/(nh)]^{1/2})$, and for the covariance function, it is $O(b^2 + [\log n/(nb^2)]^{1/2})$ [14,21,29,38].

These steps have been implemented in the Matlab software PACE, http://www.stat.ucdavis.edu/PACE/. Given $\hat{\mu}(t)$ and $\hat{G}(s,t)$, the eigenfunctions $\hat{\phi}_k(t)$ and eigenvalues $\hat{\lambda}_k$ are obtained through the discretized version of the eigen-equations. PACE then implements two methods to estimate FPCs. One option is to estimate the FPC scores or simply FPCs ξ_k through a numerical approximation of the integral $\int_{t\in\mathcal{T}}(X(t) - \hat{\mu}(t))\hat{\phi}_k(t)\mathrm{d}t$, which obviously only works for dense designs. The other option is to use the conditional expectation derived under a Gaussian assumption on both random processes and errors, which then gives the best unbiased predictor; if Gaussianity

does not hold, this approach targets the best linear unbiased predictor (BLUP). This approach works for both dense designs and sparse designs and is as follows.

Setting $\boldsymbol{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{iN_i}))^T$, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T$, $\boldsymbol{\mu}_i = (\mu(t_{i1}), \ldots, \mu(t_{iN_i}))^T$, and $\boldsymbol{\phi}_{ik} = (\phi_k(t_{i1}), \ldots, \phi_k(t_{iN_i}))^T$, one obtains by the joint Gaussianity of ξ and Y

$$E[\xi_{ik}|\boldsymbol{Y}_i] = \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_{\boldsymbol{Y}_i}^{-1}(\boldsymbol{Y}_i - \mu_i),$$

where

$$\boldsymbol{\Sigma}_{Y_i} = \mathrm{cov}(\boldsymbol{Y}_\mathrm{i}, \mathrm{Y_i}) = \mathrm{cov}(\boldsymbol{X}_\mathrm{i}, \mathrm{X_i}) + \sigma^2 \boldsymbol{I}_{\mathrm{m_i}}.$$

The PACE method has proven to be quite robust and generally works well also for the case of non-Gaussian data, in which case one deals with best linear predictors for the FPCs. It follows from results in [23] that this conditional method to obtain the FPCs is asymptotically the same as the numerical approximation of integration (Eq. 2) as sparse designs converge to dense designs.

3 Computational Aspects for Large Functional Data

Assume that for each functional datum X_i, $i = 1, \ldots, n$, measurements are available on an equispaced grid $t_1 < \cdots < t_p$, $\mathcal{T} = [t_1, t_p]$, where such measurements are possibly contaminated by noise, as in (3), and p is relatively large. In this dense design situation with an equidistant measurement grid, we may obtain cross-sectional means at each time point, $\bar{X}_j = n^{-1} \sum_{i=1}^n X_i(t_j)$, $j = 1, \ldots, p$, and the centered $n \times p$ data matrix W with elements $W_{ij} = X_i(t_j) - \bar{X}_j$. The covariance function G can be estimated at all pairs of the gridpoints (t_j, t_k), $j, k = 1, \ldots, p$, by computing the $p \times p$ matrix $H = \frac{1}{n} W^T W$ with elements H_{jk}, i.e., $\hat{G}(t_j, t_k) = H_{jk}$. The eigenvalues and eigenvectors of H can then be appropriately scaled to yield estimates of the eigenvalues and eigenfunctions of A, evaluated at the grid $(t_j,\ j = 1, \ldots, p)$ [3,7,9].

In the case of densely observed functional data, the dimension p is generally quite large and this computational step can become costly. As densely observed functional data are commonly assumed to arise from smooth processes, one technique for reducing computational complexity is to reduce the number of gridpoints to $q < p$ by binning the data values. Let $\tilde{t}_1 < \cdots < \tilde{t}_q$ be a coarser equispaced grid for $\mathcal{T}$. A simple binning scheme is to identify, for each gridpoint in the finer grid, the closest gridpoint in the coarser grid via

$$c_j = \underset{1 \le k \le q}{\mathrm{argmin}} \left| \tilde{t}_k - t_j \right|, \quad 1 \le j \le p,$$

and grouping the finer gridpoints together in the sets $I_k = \{j;\ c_j = k, 1 \le j \le p\}$.

For each subject, data values corresponding to gridpoints with indices in the same set I_k are then averaged to form the new $n \times q$ data matrix $\tilde{W}$ with elements

$$\tilde{W}_{ik} = \frac{1}{|I_k|} \sum_{j \in I_k} W_{ij}.$$

Discarding information by binning will affect the accuracy of the estimation, however. Specifically, binning will introduce a bias especially when some of the random trajectories may vary rapidly over short intervals and $q \ll p$. Another simple scheme to accelerate computations is to work with random subsamples of size $n' < n$. While this subsampling method also accelerates computations, the price to pay is not an increase in bias but rather an increase in variance and loss of efficiency.

In the commonly encountered case of very densely sampled data, where $p \gg n$, there is yet another well-known simple relation [12,20] that can be exploited to gain computational advantage. This is to compute the eigendecomposition for the $n \times n$ matrix WW^T instead of the eigendecomposition for the $p \times p$ matrix $W^T W$, as described above. This will save substantial computing time when $n \ll p$, i.e., the number of measurements per subject is much larger than the number of subjects or units.

That these two matrices have the same nonzero eigenvalues can be evidenced by the following well-known argument. Let λ be an eigenvalue of WW^T with eigenvector u. Then $W^T(WW^T u) = \lambda W^T u$, so that $W^T u$ is an eigenvector of $W^T W$ with eigenvalue λ. Since the argument is symmetric, these matrices have the same nonzero eigenvalues. Furthermore, if $\|u\| = 1$, a unit eigenvector of $W^T W$ corresponding to the eigenvalue λ is $v = cW^T u$. To determine c, note that

$$1 = c^2 \left\| W^T u \right\|^2 = c^2 \lambda \|u\|^2 = c^2 \lambda,$$

so that $c = \lambda^{-1/2}$ and $v = \lambda^{-1/2} W^T u$.

Hence, the eigenvalues and eigenvectors of $H = \frac{1}{n} W^T W$ can be calculated from those of WW^T, which are easier to compute when $p \gg n$. This device is useful to accelerate the computational implementation of FPCA for the case of dense regular designs whenever there is a substantial imbalance between n and p.

4 Models for Functional Predictors and a Scalar Response

In this section, we illustrate functional regression models. These models are generally characterized by the inclusion of a functional component in either predictors or responses, along with scalar or vector components. Most research has been devoted to the functional linear regression model where a functional predictor is coupled with a scalar response. Extended models with functional predictors and scalar responses are the quadratic regression model [18,37], functional additive model [10,28,42], and continuously additive model [22,27]. These also include quantile regression models with functional predictors [4,6]. Going beyond functional models with single functional predictors and scalar responses, models of interest extend to cases with several functional predictors and functional responses, which are topics of recent interest.

We illustrate functional regression models with the simplest model, the functional linear model that associates a scalar response with a functional predictor. The data are daily observations of sexual signaling of male Mediterranean fruit flies (medflies, *Ceratitis capitata*) [30,40]. For each fly, sexual signaling was recorded every 10 min for a

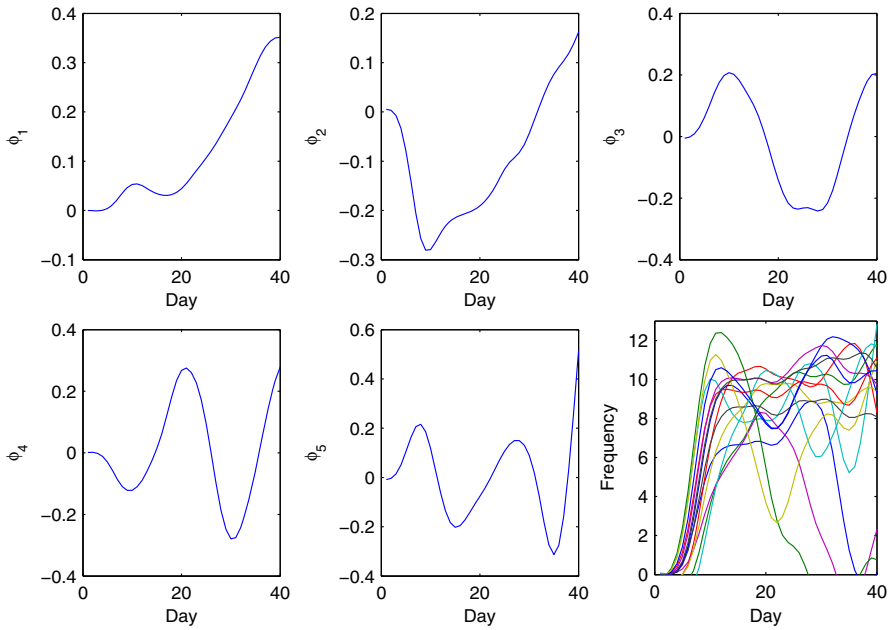

Fig. 1 FPCA estimates for medfly data. *Top Row* and *Bottom Left*, *Middle* estimated eigenfunctions ϕ_k for $k = 1, 2, 3, 4, 5$. *Bottom Right* smoothed predictor functions $X(t)$ for a subsample of 15 flies (Color figure online)

2-h period each day, resulting in 12 daily measurements. The functional predictor $X(t)$ corresponds to the number of times ($X(t) \in \{0, \ldots, 12\}$) the fly exhibited sexual signaling behavior on day t. Given the behavioral calling trajectories observed for each fly for ages $t \leq t_0 = 40$ days as functional data, we are interested to predict the logarithm of the remaining lifetime $Y = \log(T - t_0)$, where T is the total lifetime of the fly. The logarithmic transformation is applied due to the right skewness of remaining lifetimes in our sample. There were $n = 180$ flies in the sample which survived at least 40 days.

With $X^c(t) = X(t) - \mu(t)$, the functional linear regression model [31] is

$$E(Y|X) = \alpha + \int X^c(t)\beta(t)\mathrm{d}t.$$

The regression parameter function $\beta(t)$ is usually represented in a suitable basis, and if predictor functions X are represented in the same basis, and both expansions are truncated at a finite number of components, the above model converts into a finite-dimensional multiple linear regression model.

If for example both X and the regression parameter function β are expanded in the eigenbasis, as in the PACE package, then this linear regression model has uncorrelated predictors, which means that it can be decomposed into a series of simple linear regressions [26]. For the medfly data, the predictor curves $X(t)$ are represented by truncating the eigenexpansion in (1) at the first $K = 5$ terms. The corresponding first five eigenfunctions are plotted in Fig. 1, along with a subsample of the predictor functions that are represented in these eigenfunctions as a basis.

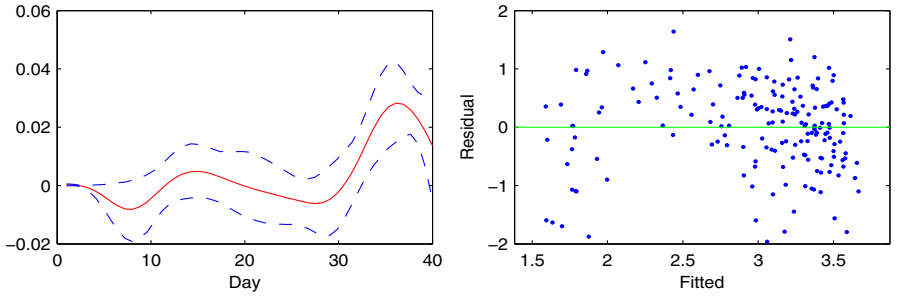

Fig. 2 Components of fitted linear regression model for medfly data. *Left* fitted regression parameter function β with pointwise 95 % confidence band, obtained via bootstrap. *Right* plot of residuals versus fitted values

The linear model fitted to the fly data yields a functional R^2 value of 0.356. Measures of variance explained such as R^2 are not straightforward and not unique in the case of functional predictors, and we use here the measure proposed in [39]. The fitted coefficient function $\beta(t)$ and the diagnostic residual plot [8] for the linear model are shown in Fig. 2. The coefficient function indicates that flies which peak in their signaling behavior after day 30 are associated with longer lifetimes. However, the residual plot provides evidence for a non-linear pattern and thus lack of fit.

This problem can be remedied by extending the model to a functional quadratic regression model [37]

$$E(Y|X) = \alpha + \int X^c(t)\beta(t)\mathrm{d}t + \iint \gamma(s,t)X^c(s)X^c(t)\mathrm{d}s\mathrm{d}t.$$

The quadratic model fitted to the fly data resulted in an R^2 value of 0.528, showing a steep improvement compared to the linear model. The residual plot, along with the estimated coefficient function $\beta(t)$ and surface $\gamma(s,t)$, is shown in Fig. 3. The linear coefficient $\beta(t)$ shows that high activity just prior to day 40 and lower activity at around day 30 are associated with higher longevity. The interpretation of the surface γ is more nuanced. For instance, the valleys around (15, 15) and (30, 30) reinforce the valleys seen in the coefficient β, while the off-diagonal valley and peak near (15, 35) and (20, 40), respectively, indicate that interactions between early- and late-life signaling behavior are significant in determining longevity. Importantly, the residual plot does not display a non-linear pattern, indicating an improved fit.

Another flexible extension of the functional linear model is the functional additive model [28]. With FPCs ξ_k, $k \geq 1$, as defined above, the centered predictor process X^c can be equivalently represented by the sequence of FPCs ξ_j, $j \geq 1$, according to the Karhunen–Loève representation. The functional additive model for the regression of a scalar response Y on a functional predictor X can then be represented as

$$E(Y|X) = \mu_Y + \sum_{k=1}^{\infty} f_k(\xi_k),$$

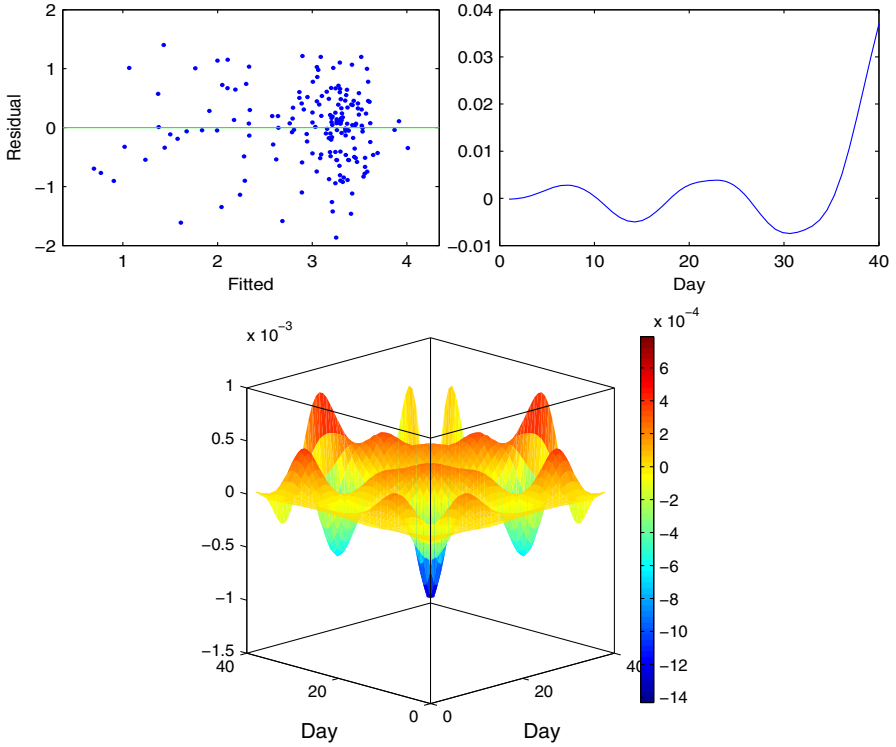

Fig. 3 Components of fitted quadratic regression model for medfly data. *Top Left* plot of residuals versus fitted values. *Top Right* fitted regression parameter function β. *Bottom* fitted regression parameter surface γ (Color figure online)

where the additive functions $f_k(\cdot)$ are smooth functions that satisfy the constraints $E(f_k(\xi_k)) = 0$.

When predictor processes are Gaussian, the scores ξ_k are independent and this implies that $f_k(\xi_k) = E(Y - \mu_Y \mid \xi_k)$, which means that these functions correspond to nonparametric regressions with a one-dimensional predictor. Motivated by this finding, [28] propose to estimate the additive functions $f_k(\xi_k)$ by individual nonparametric smoothing of the scatterplot $\{(\xi_{ik}, Y_i - \mu_Y), i = 1, \ldots, n\}$, instead of using the common backfitting approach for fitting an additive linear model.

After suitable truncation at K included terms, we obtain estimates

$$\hat{E}(Y \mid X) = \hat{\mu}_Y + \sum_{k=1}^{K} \hat{f}_k(\xi_k),$$

where $\hat{\mu}_Y$ is the sample mean of the Y_i. Fitting this model to the fly data gave an R^2 of 0.305, even lower than that of the linear model. The fitted additive functions f_k, $k = 1, \ldots, 5$, and the residual plot for this fit are shown in Fig. 4. It is clear that the first two components are the most influential in determining expected lifetime. We can

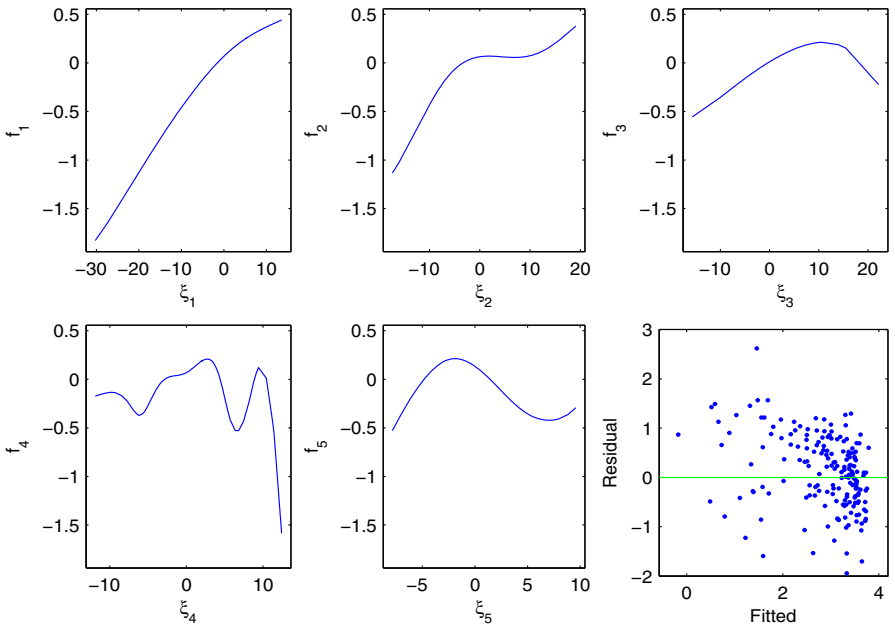

Fig. 4 Components of fitted additive regression model for medfly data. *Top Row* and *Bottom Left*, *Middle* fitted additive function $f_k(\xi_k)$ for $k = 1, 2, 3, 4, 5$. *Bottom Right* plot of residuals versus fitted values

interpret these effects by examining the corresponding eigenfunctions in Fig. 1. Since the first eigenfunction is strictly positive and increasing, and f_1 is increasing, this indicates that male medflies which exhibit above-average sexual signaling behavior across their lifetime, but particularly later in life, are associated with higher longevity.

The interpretation of the second additive function f_2 in conjunction with the second eigenfunction additionally indicates that flies which are highly active at around day 10 and then decline in activity tend to have shorter lifespans. While the residuals from the additive model do not indicate clear lack of fit, there is no clear improvement when compared to the quadratic model. Hence, the quadratic model seems to give the best fit for these data. This provides an example of how careful model selection for functional regression aids in the interpretation of complex time dynamic biological phenomena.

In addition to conditional mean estimation, which is the usual regression task, in some applications, one may be interested in estimating the conditional distribution and conditional quantiles of Y given the predictors. This is a more challenging task in the case of functional predictors X. The problem of extending the mean regression with functional predictors to the case of a conditional distribution has been studied using different approaches [4,6]. Specifically, [6] have proposed estimating the conditional distribution of the responses Y given predictor trajectories X as a first step. This can be done by the following generalized functional regression approach:

$$F(y|X) = P(Y \leq y|X) = E(I(Y \leq y)|X) = g^{-1}\left(\alpha(t) + \int X^c(t)\beta(y,t)\mathrm{d}t\right),$$

where g is a binomial regression link function such as the logit link.

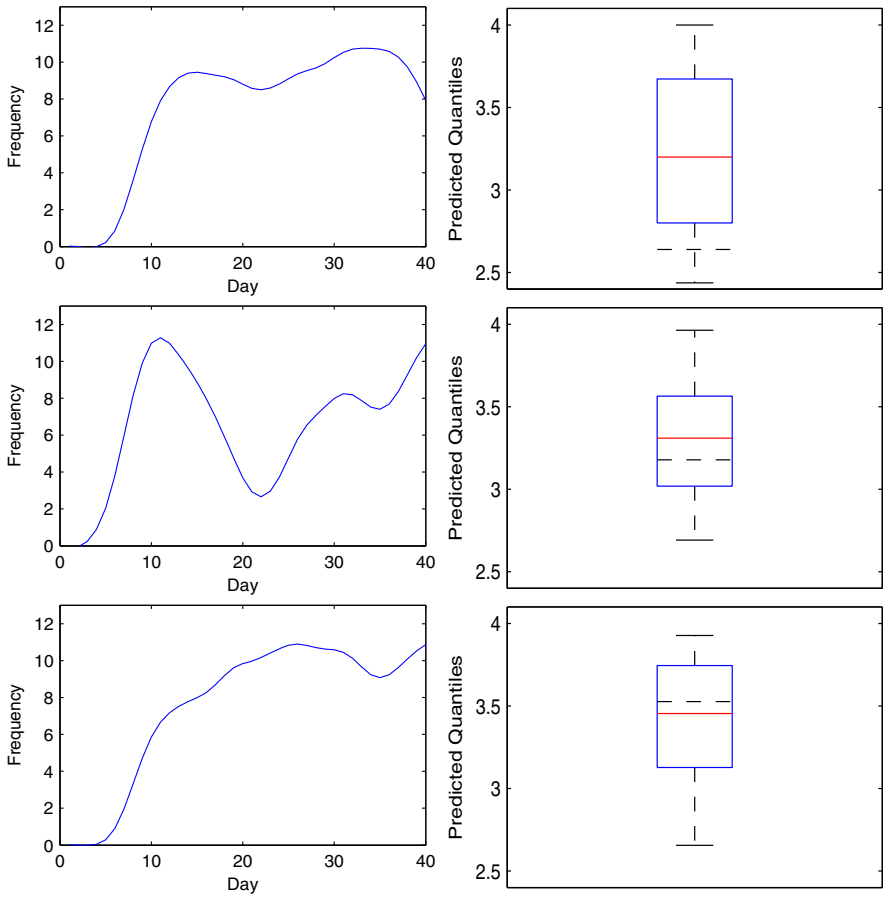

Fig. 5 *Left Column* prediction functions for three specific male medflies. *Right Column* corresponding estimated conditional quantiles ($\tau = 0.1, 0.25, 0.5, 0.75, 0.9$) of Y given the respective predictor function in the *left* adjacent panel. Observed Y values are shown as a *horizontal dashed line*

The conditional quantiles are obtained from the inverse of the fitted conditional distribution. Applying this nonparametric quantile regression approach to the medfly data, we examine the fitted quantiles for three particular flies corresponding to the first quartile, median, and third quartile of observed lifetime. The predictor functions for these three flies are shown in Fig. 5, along with the fitted median, first and third quartiles, and 0.1th and 0.9th quantiles in the form of a boxplot. Overall, one finds that these conditional quantiles indeed enclose the observations. Additionally, the distributions for the two longer-living flies show less spread, with that of the longest-living fly also showing a right skew not present in the other conditional distributions.

5 Repeated Functional Data

Repeated functional data exhibit an increased level of complexity, due to the presence of dependencies between the functional recordings. They often will fall under the

rubric of large and complex functional data. Here one faces a situation where curves are repeatedly recorded for a sample of n subjects. Specifically, for the ith subject ($i = 1, \ldots, n$) one has measurements at multiple time points $\{s_{ij} : j = 1, \ldots, m_i\}$; at each time s_{ij}, a curve/function $X_i(\cdot \mid s_{ij})$ is recorded. In practice, these functions $X_i(\cdot \mid s_{ij})$ are usually observed at a grid of discrete time points $\{t_{ijl} : l = 1, \ldots, L_{ij}\}$. Therefore, the repeated functional data we observe are $\{X_i(t_{ijl} \mid s_{ij}) : i = 1, \ldots, n;\ j = 1, \ldots, m_i;\ l = 1, \ldots, L_{ij}\}$.

Functional principal component analysis (FPCA) as described in Sect. 2 provides a starting point for modeling repeatedly observed functional data. An extension to repeated functional data is the double FPCA approach [7], which is based on a two-step Karhunen–Loève expansion. For the ith subject, the function $X_i(\cdot \mid s)$ observed at a given longitudinal time s is assumed to admit the following decomposition:

$$X_i(t \mid s) = \mu(t \mid s) + \sum_{k=1}^{\infty} \xi_{ik}(s)\phi_k(t \mid s), \qquad \text{(Step 1 FPCA)} \tag{4}$$

where $\mu(\cdot \mid s)$ is the mean function at s, and $\phi_k(\cdot \mid s)$ is the kth eigenfunction of the repeated functions at s with the corresponding FPC $\xi_{ik}(s)$, i.e., $\phi_k(\cdot \mid s)$ and $\xi_{ik}(s)$ are the kth eigenfunction and FPC of the covariance $G(t_1, t_2 \mid s) = \mathrm{cov}(X(t_1 \mid s), X(t_2 \mid s))$.

For each k, $\xi_{ik}(s)$ can be further decomposed by employing a second Karhunen–Loève expansion:

$$\xi_{ik}(s) = \sum_{p=1}^{\infty} \zeta_{ikp}\psi_{kp}(s), \qquad \text{(Step 2 FPCA)} \tag{5}$$

with eigenfunctions $\psi_{kp}(\cdot)$ and corresponding FPCs ζ_{ikp}, i.e., $\psi_{kp}(\cdot)$ and ζ_{ikp} are, respectively, the pth eigenfunction and FPC of the autocovariance operator with covariance kernel $R_k(s_1, s_2) = \mathrm{cov}(\xi_k(s_1), \xi_k(s_2))$. Combining (4) and (5), we have

$$\begin{aligned} X_i(t \mid s) &= \mu(t \mid s) + \sum_{k=1}^{\infty} \xi_{ik}(s)\phi_k(t \mid s) \\ &= \mu(t \mid s) + \sum_{k=1}^{\infty} \left(\sum_{p=1}^{\infty} \zeta_{ikp}\psi_{kp}(s) \right) \phi_k(t \mid s) \\ &= \mu(t \mid s) + \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} \zeta_{ikp}\varphi_{kp}(t \mid s), \end{aligned}$$

where $\varphi_{kp}(t \mid s) = \psi_{kp}(s)\phi_k(t \mid s)$. Therefore, the total variation of X_i can be decomposed into the variation conditional on longitudinal time s and the variation along the longitudinal time s with random effects ζ_{ikp}.

Estimation procedures for the unknown components of the two-step Karhunen–Loève representation, including $\mu(t \mid s)$, $G(t_1, t_2 \mid s)$, $\phi_k(t \mid s)$, $\xi_k(s)$, $\psi_{kp}(s)$,

and ζ_{kp}, were developed in the study of [7], where also various sampling plans and designs for the measurement time locations were considered. These include *dense regular designs*, *dense random designs*, and *sparse random designs* in s. For dense regular designs, where both t and s are measured on regular and dense time grids, $\mu(t \mid s)$ and $G(t_1, t_2 \mid s)$ can be estimated by taking a cross-sectional mean; for dense random designs and sparse random designs, where the measurement time points are irregular, smoothing is needed, which can be implemented by local linear smoothers to estimate both $\mu(t \mid s)$ and $G(t_1, t_2 \mid s)$. Based on consistent estimates of $\mu(t \mid s)$ and $G(t_1, t_2 \mid s)$, one then obtains consistent estimates of $\phi_k(t \mid s)$, $\xi_k(s)$, $\psi_{kp}(s)$, and ζ_{kp}.

We implemented this double FPCA approach using repeated recordings of daily movements of Mexican fruit flies (mexflies, *Anastrepha ludens*). Additional details about data collection and further descriptions can be found in [43]. Movement data were measured for 16 mexflies on a full diet (protein and sugar) and 16 mexflies on a sugar-only diet. Each fly was continuously monitored at 1-min intervals, and the spatial $X-$, $Y-$, Z-coordinates of the location of the fly were recorded every 0.2 s. Recording started after the eclosion of flies and lasted throughout their lifetime.

We focus on quantifying movement by computing the distances between successive location measurements for each fly during daytime (between 7 am and 7 pm daily) for the first 50 days of the fly's lifespan. We removed one fly from the analysis which died before 50 days so eventually we had 31 flies in total, 15 on the full diet, and 16 on the sugar diet. The longitudinal time s represents age in days (1–50 days), while time t represents daytime, which was measured as a fraction of 24 h, with measurements recorded on a regular grid of length 36 between 0.2917 and 0.7917, corresponding to 7am and 7pm. For each fly, based on the recorded $X-$, $Y-$, Z-coordinates, we obtained the approximate Euclidean distance traveled for each minute while it was monitored. We interpolated the distances onto an equidistant time grid, leading to distance observations $\{X_i(t \mid s) : i = 1, \ldots, 31; t = 0.2917, \ldots, 0.7917; s = 1, \ldots, 50\}$. Some flies had missing values for a whole day, so the design is best characterized as a dense random design. The movement trajectories of eight randomly selected flies, four with full diet and the other four with sugar diet, are illustrated in Fig. 6.

The fitted overall mean surface $\mu(t \mid s)$ as shown in Fig. 7 demonstrates that on average, the distance moved by a fly per minute is larger in the afternoon than in the morning within a day starting at the age of 10 days, which indicates that the flies are in general more active in the afternoon. In addition, the overall movement activity increases as a fly ages and reaches a peak at around the age of 25 days. The level of activity slightly fluctuates afterwards but generally remains steady.

The scatter plot of ζ_{i12} against ζ_{i11} shows that ζ_{i11} may represent the distinguishing feature between the flies on the two different food diets. The flies on the full diet mostly have positive ζ_{i11} values, while those on the sugar diet mostly have negative values. Together with the fact that $\varphi_{11}(t \mid s)$ is always negative, the flies on the sugar diet are seen to be on average more active than those on the full diet, which is also confirmed by [43]. While $\varphi_{11}(t \mid s)$ reflects an overall activity level, the second conditional eigenfunction $\varphi_{11}(t \mid s)$ reflects an increase in overall activity at all times during a day as a fly ages, where half of the flies show such a nearly uniform increase, while

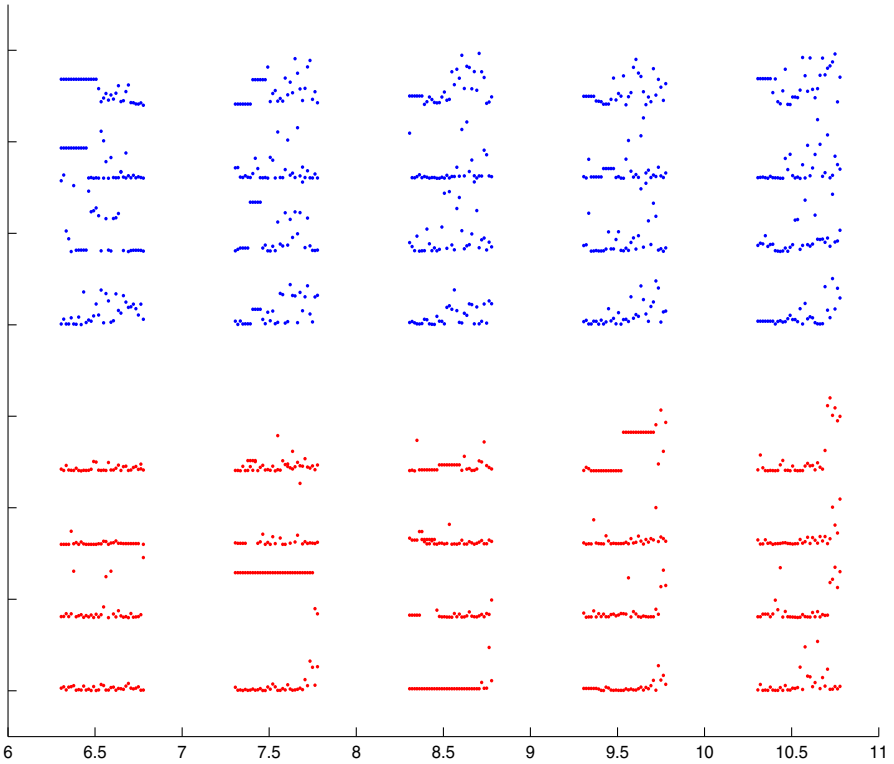

Fig. 6 Distances traveled per minute for eight randomly selected Mexflies between day 6 and day 11. The four flies at the *top* are on a sugar diet, while the four flies at the *bottom* are on a full diet

the other half of the flies shows a decrease in activity, as can be seen from the range of the levels of ζ_{i11}.

6 Quantifying Functional Connectivity in the Brain

Quantifying functional connectivity in the brain and identifying regions with strong connections has been a major focus in neuroscience research over the past two decades, accelerating steadily with the increasing availability of data from fMRI scans. One such study, conducted at UC Davis [16], consists of over 700 fMRI scans, each of which represents time courses for more than 10^5 voxels in the brain over a period of 480 s. Applying standard preprocessing steps results in a data set easily exceeding 1 TB in size.

The strength of functional connectivity between two voxels or regions is quantified by some similarity measure between representative time courses. In the neuroscience literature, the most common measure is (temporal) correlation, so that signals with similar fluctuation patterns are considered to be strongly connected. Methods in FDA provide several alternatives for measuring connectivity which are well supported theoretically.

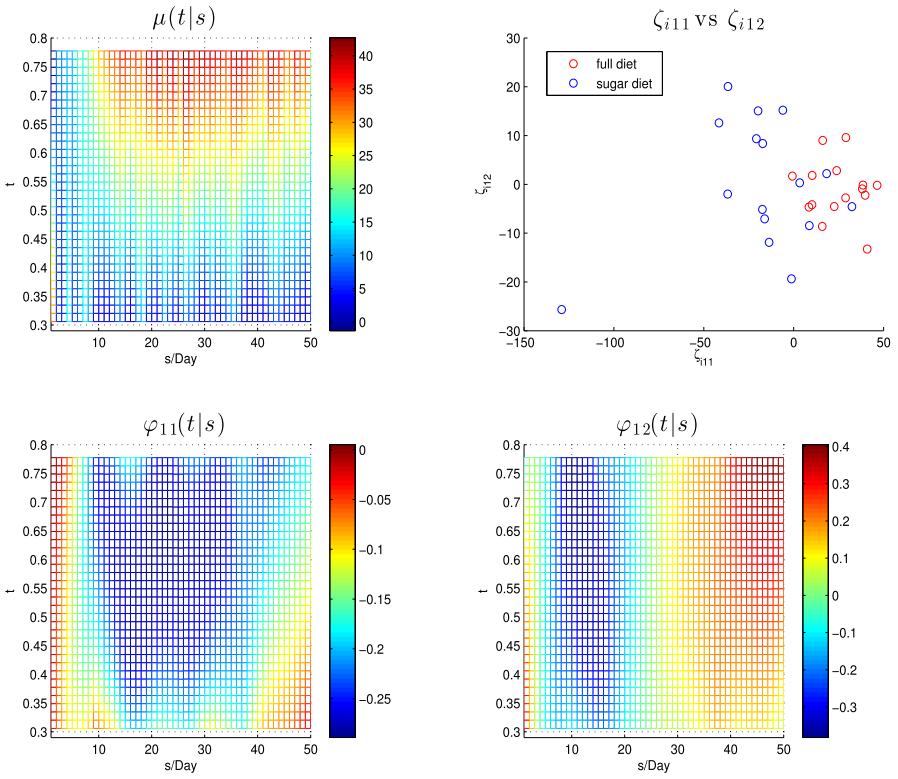

Fig. 7 Results of double FPCA for daily movement of mexflies. *Top left* fitted mean surface, *top right* scatter plot of FPC scores ζ_{i12} against ζ_{i11}; *bottom left* and *right* fitted φ_{11} and φ_{12} (Color figure online)

One such method which has been found useful in practice is functional singular component analysis (FSCA, [36]). Given two zero-mean random processes X and Y defined on a compact interval $\mathcal{T} \subset \mathcal{R}$, FSCA defines the (first) covariance as

$$\sigma = \sup_{\|u\|=\|v\|=1} \operatorname{cov}\big(\langle u, X\rangle, \langle v, Y\rangle\big),$$

where u and v range over functions in $L^2(\mathcal{T})$ and $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ are the L^2 inner product and norm, respectively. The functions u^* and v^* which attain the maximum are called singular component functions, and the coefficients $\zeta = \langle X, u^*\rangle$ and $\xi = \langle Y, v^*\rangle$ are called the singular component scores. Defining $C_{XY}(s,t) = E(X(s)Y(t))$ and the corresponding operator $\mathcal{C}_{XY} : L^2(\mathcal{T}) \to L^2(\mathcal{T})$ as

$$\mathcal{C}_{XY}(f)(s) = \int_{\mathcal{T}} C_{XY}(s,t) f(t)\,\mathrm{d}t, \quad f \in L^2(\mathcal{T}),$$

the covariance and singular functions are the solutions to

$$\mathcal{C}_{XY}(v) = \sigma u, \quad \|u\| = \|v\| = 1,$$

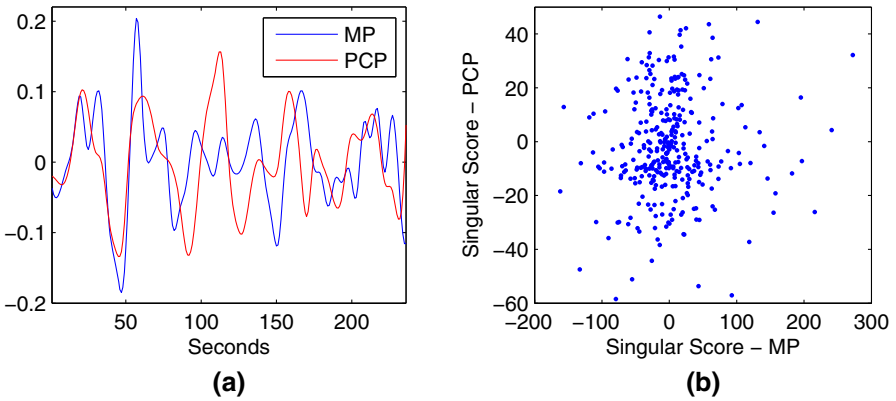

Fig. 8 Singular component from FSCA on time signals from the medial prefrontal (MP) and posterior cingulate/precuneus (PCP) regions. **a** Singular component functions. **b** Scatterplot of singular component scores (Color figure online)

whence the term "singular."

A sequence of singular component functions and scores can be constructed iteratively, where the optimization takes place over the orthogonal complement of the space spanned by the previous singular functions. Thus, FSCA provides a means of achieving joint dimensionality reduction for a pair of processes by reducing these infinite-dimensional objects to a list of pairs (ζ_k, ξ_k).

To apply FSCA to the study of connectivity, we considered two regions within the so-called default mode network, one located within the medial prefrontal cortex (MP) and the other in the posterior cingulate/precuneus area (PCP). Changes in the connectivity between these regions have been observed previously in subjects with Alzheimer's disease [35] or even mild cognitive impairment [41].

Using a subsample of scans of 306 subjects from the UC Davis study, the singular component functions and corresponding scores (ζ, ξ) for these two regions were computed (Fig. 8).

Unsurprisingly, the singular component functions for the two regions are similar, while the correlation between the scores across subjects is only $r = 0.078$. The scalar scores can be used in place of the fMRI time courses for the two regions in connectivity analyses, facilitating a great reduction in computational cost, as the raw time courses are high dimensional.

7 Stringing High-Dimensional Data for Functional Analysis

Modeling and prediction for very high-dimensional data of dimension p, when p is large, is well known to be a challenging problem. However, if the p-dimensional observations are ordered and can be represented as discretized and noisy observations that originate from a hidden smooth stochastic process, one can utilize various functional data techniques to take advantage of the underlying smoothness. This idea has been conceptualized in a "stringing" method that reorders the components of the high-

dimensional vectors and transforms the observations into functional data. Established techniques from FDA can be applied for further statistical analysis, once an underlying stochastic process and the corresponding random trajectory for each subject have been identified. Stringing can be implemented by distance-based metric Multidimensional Scaling, mapping high-dimensional data to locations on a real interval such that predictors that are close in a suitable sample metric also are located close to each other on the interval.

We illustrate the stringing idea with a single-nucleotide polymorphism (SNP) data set and a gene data set, and also explore the idea of variable selection based on the stringed order. The SNP genotype data for our first illustration of stringing are from a marker trait association study where 74 SNPs in 23 candidate genes have been selected from the literature, focusing on different pathways associated with the folate, lipid, and vitamins A, E, and B-12 metabolism [5]. We only retain SNPs that have less than 10 missing values, and subjects in the Caucasian group, resulting in $n = 374$ subjects and 64 SNPs. The SNP genotype takes the values 0 (AA), 1 (Aa), and 2 (aa), thus the SNPs can be viewed as ordinal data. We base our analysis on the 64×64 distance matrix of pairwise Euclidian distances of SNP genotypes and then apply Stringing.

The SNP data in the stringed order are presented in Fig. 9, overlaid with recovered underlying processes $X_i(t)$ for nine subjects. These recovered processes have been obtained by FPCA, as described in Sect. 2. Following this stringing step, one can further apply FDA techniques. For example, one can build functional linear regression models as described in Sect. 3, using stringed SNP functions $X_i(t)$ to predict a relevant response Y, such as plasma homocysteine (Hcy). A next goal is then to identify SNPs that are significant for Hcy level regulation.

For a second illustration of stringing, we present data from a study of the survival of patients with diffuse large-B-cell lymphoma (DLBCL) [32]. Here one is interested in predicting survival from individual high-dimensional microarray gene expression data. The data consist of $n = 240$ patients, for each of whom $p = 7399$ gene expression levels were measured. For initial screening, we follow the same approach as described in [5], and select the top 240 genes that have the largest individual Cox scores, where Cox score is defined as the score test statistic for gene effect in univariate Cox regression models. The patients are randomly divided into training (160 subjects) and test (80 subjects) groups; only the training group data are used to perform stringing and to establish the stringed order of genes. The genes in stringed order for four randomly selected subjects are visualized in Fig. 10.

Next, we explore an idea of variable selection based on the stringed order. The starting point for this selection is that we expect smoothness in gene profiles, specifically that after stringing genes placed close to each other on the interval have a similar effect on the response. While a direct full search of all possible subsets of genes is impossible, and the expected continuity motivates to consider subintervals as units, which is computationally much faster.

Specifically, we further divide the 160 subjects randomly into subsets of 110 ($S1$) and 50 ($S2$) subjects. We use $S1$ to train a Cox model, and $S2$ to evaluate the prediction performance of the model. In a first step, we divide the ordered genes into eight adjacent intervals and implement a full search of any possible combinations of these 8 intervals as predictors in a Cox model. We build Cox regression models using the center gene in

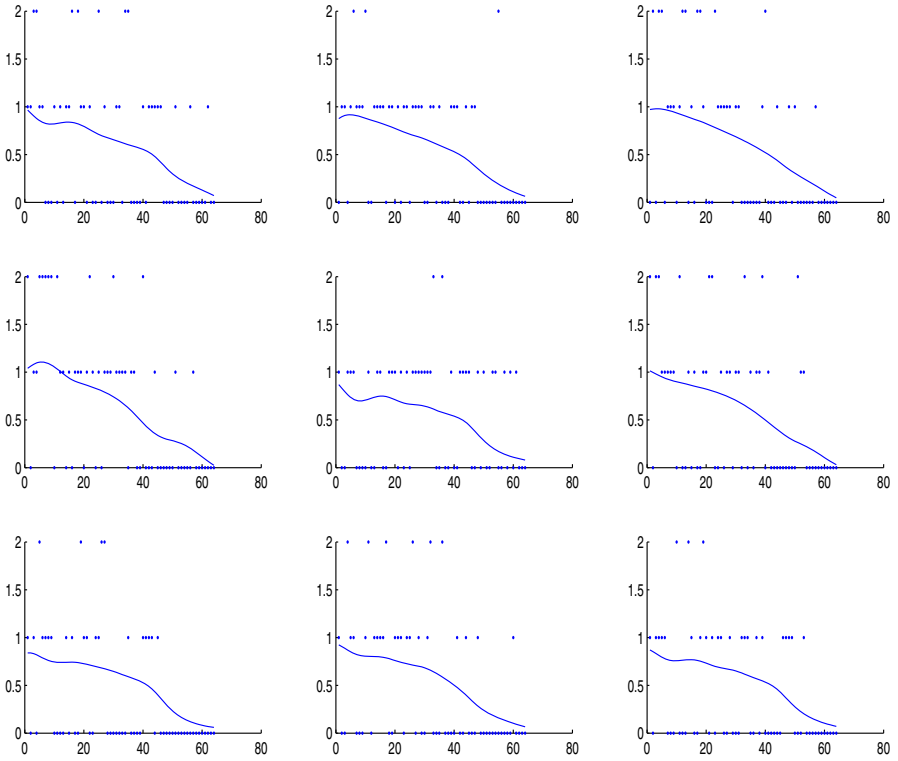

Fig. 9 SNP data represented in stringed order with fitted *curves* using functional principal component analysis for nine randomly selected subjects

the corresponding interval as predictors and evaluate the performance of each model on the subset $S2$. Then the model with the smallest deviance for the Cox model on $S2$ is selected, which leads to the selection of regions 1, 2, and 8. The regions are marked blue in Fig. 11. We further divide each of the selected regions into 6 regions, resulting in 18 regions (each contains five genes). Then a second search over models that contain k regions, for $k \leq 6$, is performed. The best model contains the center gene from five sub-regions marked red in Fig. 11. Since we use center genes in each region when building the regression model, where, for example, 18 is the center gene of the region [16,17,18,19, 20], we could further search around those center genes to fine tune the model. The final best model contains genes with indices [19, 23, 29, 38, 228] out of the 240 genes. They correspond to gene numbers [5301, 1188, 3810, 3811, 1638] in the original gene index set (7399 genes).

To evaluate the model with the selected five genes, we look at the deviance (DEV) on the testing sample ($n = 80$). For 50 simulations based on random splits of training and testing sets, the median deviance, using the selected 5 genes, is -10.1. [2] reported the median deviance for three previously used methods based on the same simulation design. These comparison methods are principal component regression with $DEV = -3$, ridge regression with $DEV = -8.5$, and Lasso with -4.5. The new variable

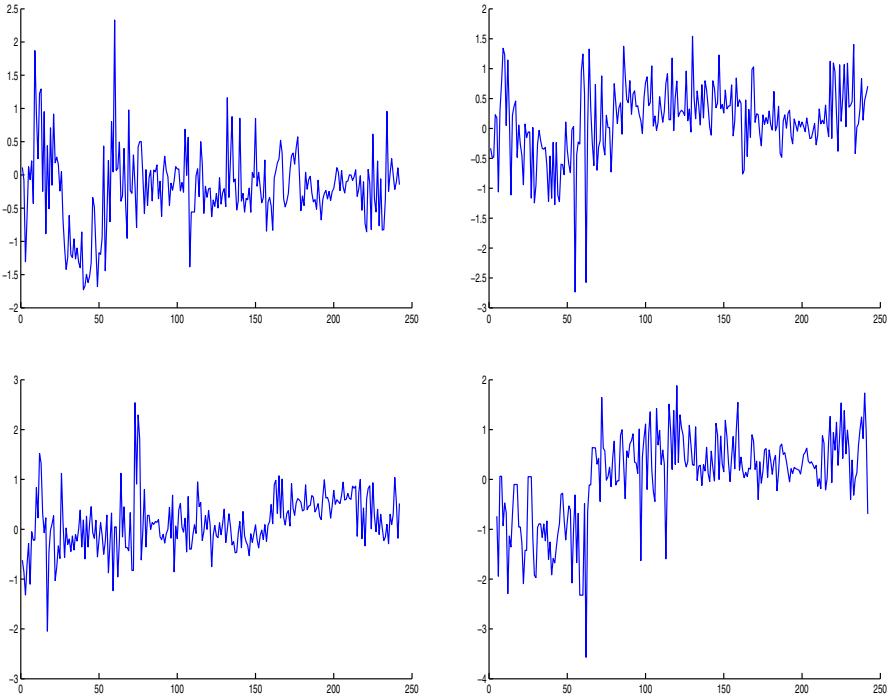

Fig. 10 Genes in stringed order for four randomly selected subjects

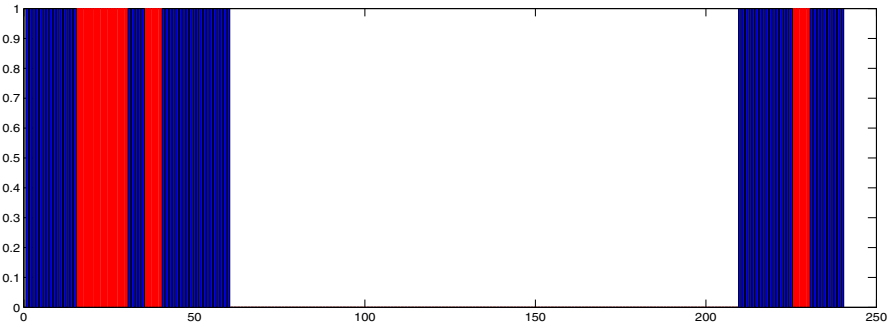

Fig. 11 Variable selection for DLBCL data: the regions selected from the first step are marked in *blue* and the subsequent selections are marked in *red* (Color figure online)

selection method based on stringed order outperforms these methods as it has a smaller deviance.

8 Conclusions

Functional Data Analysis is well suited for the analysis of large and complex biological data that include a time dynamic component and where the time dynamics is an

important feature. Such data abound in ecology and biodemography, monitoring and tracking, genomics, and many medical applications. Often one is interested in studying relationships between variables that consist of both vectors and functional components, and by now there exist a large number of functional regression models that aim at various scenarios of combinations of scalar and functional variables. Specifically, functional data analysis provides a flexible tool for sparsely measured longitudinal data as they are commonly encountered throughout the social and life sciences.

A highly versatile tool that has proven almost universally applicable is FPCA that is based on an eigenexpansion of the underlying covariance operators of the processes observed. This method is highly effective for implementing the necessary dimension reduction that reduces infinite-dimensional functional and longitudinal data to a vector of FPC scores that are then used for further statistical analysis. While there are various to accelerate the necessary computations for the case of very large sets of functional data, speeding-up of the underlying computations remains an important topic for future research.

Acknowledgements Research supported by NSF Grants DMS-1228369 and DMS-1407852.

References

1. Bosq D (2000) Linear processes in function spaces: theory and applications. Springer, New York
2. Bøvelstad H, Nygård S, Størvold H, Aldrin M, Borgan Ø, Frigessi A, Lingjærde O (2007) Predicting survival from microarray data—a comparative study. Bioinformatics 23:2080–2087
3. Cardot H (2000) Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. J Nonparametr Stat 12:503–538
4. Cardot H, Crambes C, Sarda P (2005) Quantile regression when the covariates are functions. J Nonparametr Stat 17:841–856
5. Chen K, Chen K, Müller H-G, Wang J (2011) Stringing high-dimensional data for functional analysis. J Am Stat Assoc 106:275–284
6. Chen K, Müller H-G (2012) Conditional quantile analysis when covariates are functions, with application to growth data. J R Stat Soc Ser B 74:67–89
7. Chen K, Müller H-G (2012) Modeling repeated functional observations. J Am Stat Assoc 107:1599–1609
8. Chiou J-M, Müller H-G (2007) Diagnostics for functional regression via residual processes. Comput Stat Data Anal 51:4849–4863
9. Dauxois J, Pousse A, Romain Y (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J Multivar Anal 12:136–154
10. Febrero-Bande M, González-Manteiga W (2013) Generalized additive models for functional data. Test 22:278–292
11. Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer, New York
12. Good IJ (1969) Some applications of the singular decomposition of a matrix. Technometrics 11:823–831
13. Hall P, Hosseini-Nasab M (2006) On properties of functional principal components analysis. J R Stat Soc Ser B 68:109–126
14. Hall P, Müller H-G, Wang J-L (2006) Properties of principal component methods for functional and longitudinal data analysis. Ann Stat 34:1493–1517
15. Hall P, Müller H-G, Yao F (2008) Modeling sparse generalized longitudinal observations with latent Gaussian processes. J R Stat Soc Ser B 70:703–723
16. Hinton L, Carter K, Reed BR, Beckett L, Lara E, DeCarli C, Mungas D (2010) Recruitment of a community-based cohort for research on diversity and risk of dementia. Alzheimer Dis Assoc Disord 24:234
17. Horvath L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York

18. Horváth L, Reeder R et al (2013) A test of significance in functional quadratic regression. Bernoulli 19:2120–2151
19. Hsing T, Eubank R (2015) Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley, Chichester
20. Kneip A, Utikal KJ (2001) Inference for density families using functional principal component analysis. J Am Stat Assoc 96:519–542
21. Li Y, Hsing T (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Ann Stat 38:3321–3351
22. McLean MW, Hooker G, Staicu A-M, Scheipl F, Ruppert D (2014) Functional generalized additive models. J Comput Graph Stat 23:249–269
23. Müller H-G (2005) Functional modelling and classification of longitudinal data. Scand J Stat 32:223–240
24. Müller H-G (2008) Functional modeling of longitudinal data. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds) Longitudinal data analysis (handbooks of modern statistical methods). Chapman & Hall, New York, pp 223–252
25. Müller H-G (2011) Functional data analysis, in international encyclopedia of statistical science. In: Lovric M (ed) Extended version available in StatProb: the encyclopedia sponsored by statistics and probability societies, id 242, Springer, Heidelberg, pp. 554–555
26. Müller H-G, Chiou J-M, Leng X (2008) Inferring gene expression dynamics via functional regression analysis. BMC Bioinform 9:60
27. Müller H-G, Wu Y, Yao F (2013) Continuously additive models for nonlinear functional regression. Biometrika 100:607–622
28. Müller H-G, Yao F (2008) Functional additive models. J Am Stat Assoc 103:1534–1544
29. Müller H-G, Yao F (2010) Empirical dynamics for longitudinal data. Ann Stat 38:3458–3486
30. Papadopoulos NT, Katsoyannos BI, Kouloussis NA, Carey JR, Müller H-G, Zhang Y (2004) High sexual calling rates of young individuals predict extended life span in male Mediterranean fruit flies. Oecologia 138:127–134
31. Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn., Springer series in statistic-sSpringer, New York
32. Rosenwald A, Wright G, Chan W, Connors J, Campo E, Fisher R, Gascoyne R, Muller-Hermelink H, Smeland E, Giltnane J, Hurt E, Zhao H, Averett L, Yang L, Wilson W, Jaffe E, Simon R, Klausner R, Powell J, Duffey P, Longo D, Greiner T, Weisenburger D, Sanger W, Dave B, Lynch J, Vose J, Armitage J, Montserrat E, Lpez-Guillermo A, Grogan T, Miller T, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt L, LMPP (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346:1937–1947
33. Silverman BW (1996) Smoothed functional principal components analysis by choice of norm. Ann Stat 24:1–24
34. Staniswalis JG, Lee JJ (1998) Nonparametric regression analysis of longitudinal data. J Am Stat Assoc 93:1403–1418
35. Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, Jiang T (2007) Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. Hum Brain Mapp 28:967–978
36. Yang W, Müller H-G, Stadtmüller U (2011) Functional singular component analysis. J R Stat Soc Ser B 73:303–324
37. Yao F, Müller H-G (2010) Functional quadratic regression. Biometrika 97:49–64
38. Yao F, Müller H-G, Wang J-L (2005) Functional data analysis for sparse longitudinal data. J Am Stat Assoc 100:577–590
39. Yao F, Müller H-G, Wang J-L (2005b) Functional linear regression analysis for longitudinal data. Ann Stat 33:2837–2903
40. Zhang Y, Müller H-G, Carey JR, Papadopoulos NT (2006) Behavioral trajectories as predictors in event history analysis: male calling behavior forecasts medfly longevity. Mech Ageing Dev 127:680–686
41. Zhang H-Y, Wang S-J, Xing J, Liu B, Ma Z-L, Yang M, Zhang Z-J, Teng G-J (2009) Detection of PCC functional connectivity characteristics in resting-state fMRI in mild Alzheimers disease. Behav Brain Res 197:103–108
42. Zhu H, Yao F, Zhang HH (2014) Structured functional additive regression in reproducing kernel Hilbert spaces. J R Stat Soc Ser B 76:581–603

43. Zou S, Liedo P, Altamirano-Robles L, Cruz-Enriquez J, Morice A, Ingram DK, Kaub K, Papadopoulos N, Carey JR (2011) Recording lifetime behavior and movement in an invertebrate model. PloS One 6:e18151