Methods of Canonical Analysis for Functional Data[1]

Guozhong He[2], Hans-Georg Müller[3] and Jane-Line Wang[3]

Revision July 2002

Running title: Canonical Analysis for Functional Data

Abstract

We consider estimates for functional canonical correlations and canonical weight functions. Four computational methods for the estimation of functional canonical correlation and canonical weight functions are proposed and compared, including one which is a slight variation of the spline method proposed by Leurgans, Moyeed and Silverman (1993). We propose dimension reduction and dimension augmentation procedures to address the dimensionality problems of functional canonical analysis (FCA) that are associated with computational break-down. Cross-validation is used for the automatic selection of tuning parameters, based on the minimax property of FCA. This allows to estimate several canonical correlations and canonical weight functions simultaneously and reasonably well as we show in simulations. The proposed estimation methods are compared and their use is demonstrated with medfly mortality data.

MSC: primary 62G05, 62H25; secondary 60G12, 62M10

Key words: Canonical correlation, canonical weight function, covariance operator, functional least squares, stochastic process, L_2-process, cross-validation, functional data analysis

---

[1]Research supported by NSF Grants DMS 98-03627, DMS 99-71602, and NIH Grant 99-SC-NIH-1028.

[2]University of California, San Francisco, Diabetes Control Program/DHS, 601 North 7th Street, MS 675, P.O. Box 942732, Sacramento, CA 94234-7302, USA

[3]Department of Statistics, University of California, One Shields Ave., Davis, CA 95616

1. Introduction

Functional Canonical Correlation Analysis (FCA) is a tool to quantify correlations between pairs of observed random curves for which a sample is available. Functional or curve data are increasingly common, and the analysis of correlations between functional data has applications in biology (see the medfly data in Section 4), the medical sciences (see Leurgans, Moyeed and Silverman, 1993), ecology, and the environmental sciences (compare Service, Rice and Chavez, 1998)). Hannan (1961) studied functional canonical analysis for stationary Gaussian processes, Dauxois and Pousse (1976) described canonical analysis in general Hilbert spaces, and Brillinger (1975) characterized canonical analysis for multivariate stationary time series. In the time series setting, invoking stationarity reduces this problem to classical multivariate canonical analysis. A sample version of functional canonical correlation with spline smoothing was proposed in Leurgans et al. (1993). These authors showed that there is a need for regularization in FCA in order to overcome a break-down problem which is caused by the high dimensionality of functional data. The need for regularization arises as in infinite-dimensional settings, relevant autocovariance operators and their inverses do not exist; another way of putting it is to say that one can find canonical weight functions to make the canonical correlation equal to 1. In Leurgans et al. (1993), regularization was implemented via modified smoothing splines. Functional canonical analysis is an ill-posed problem, and hence requires certain assumptions for the existence of canonical correlations and weight functions. Such conditions were studied in He, Müller and Wang (2002). We assume throughout this paper that the existence condition given in Theorem 4.8 of that paper is in force. A canonical decomposition for pairs of square integrable random processes was also developed there.

We begin with a formal definition of FCA, which includes functional correlation coefficients and associated weight functions. Let I be an interval, and $L_2(I)$ the Hilbert space of square integrable functions on I with respect to Lebesgue measure μ, equipped with the scalar product $u^T v = v^T u = \int_I u(s)v(s)d\mu(s)$. This assumption can be easily relaxed to cover the case of countable index sets and unequal intervals. Suppose $X, Y \in L_2(I)$ are jointly distributed L_2-processes, with

$$\int E \parallel Z \parallel^2 = E[Z^T Z] = E \int_I (Z(s))^2 ds < \infty, \quad \text{for } Z = X \text{ or } Z = Y.$$

Extending the notion of canonical correlation from p-vectors to stochastic processes, we define *functional canonical correlation* as follows.

The first *canonical correlation* ρ_1 and associated *weight functions* u_1 and v_1 for L_2-processes X and Y are defined as

$$\rho_1 = \sup_{u,v \in L_2(I)} \text{Cov}(u^T X, v^T Y) = \text{Cov}(u_1^T X, v_1^T Y), \tag{1}$$

where u and v are subject to

$$\text{Var}(u^T X) = 1, \text{and } \text{Var}(v^T Y) = 1. \tag{2}$$

The k-th canonical correlation ρ_k and associated weight functions u_k, v_k for the pair of processes $(X(t), Y(t))$, for $k > 1$, are defined as

$$\rho_k = \sup_{u,v \in L_2(I)} \text{Cov}(u^T X, v^T Y) = \text{Cov}(u_k^T X, v_k^T Y), \tag{3}$$

where u and v are subject to (2), and for $i = 1, \ldots, k-1$, $\text{Cov}(u^T X, u_i^T X) = \text{Cov}(v^T Y, v_i^T Y) = 0$. Thus, the k-th pair of canonical variables, (U_k, V_k) is uncorrelated with the (k-1) pairs $\{(U_i, V_i), i = 1, \ldots, k-1\}$, where $U_k = u_k^T X$ and $V_k = v_k^T Y$ are the extensions of k-th order canonical variates. We shall call $(\rho_k, u_k, v_k, U_k, V_k)$ the *k-th canonical components*.

Suppose that for the i-th subject, a pair of random curves $(X_i(\cdot), Y_i(\cdot))$ is sampled at a typically large number of discrete time points, where $X_i = (X_i(t_1), \ldots, X_i(t_{m_x}))$ are vectors of dimension m_x and $Y_i = (Y_i(t_1), \ldots, Y_i(t_{m_y}))$ are vectors of dimension m_y. A crucial issue, to be addressed in Section 2, is the reduction of dimensions m_x and m_y, which are large. Besides dimension reduction, another tool that we employ is dimension augmentation. Dimension augmentation serves to convert canonical weight vectors obtained at relatively few support points and therefore of low dimension into canonical weight functions at the original data dimension, by means of a smoothing step. An initial dimension reduction step is an integral part of two of the four FCA methods to be discussed here. These two methods include Method (1), a new proposal based on the extension of multivariate canonical correlation analysis, where smoothing with local polynomial fitting provides regularization and dimension augmentation; and Method (2), a modified version of the smoothing spline approach of Leurgans et al. (1993) that includes dimension augmentation.

He et al. (2000) showed that random processes with finite basis expansion have simple canonical structures, analogously to the case of random vectors. This motivates to implement regularization by projecting random processes on a finite number of basis functions. The idea to project processes on the first k basis functions is already mentioned in Castro, Lawton and Sylvestre (1976) and has been further discussed in Rice and Silverman (1991) and Ramsay and Silverman (1997). In our proposed Method (3), this projection is on a prespecified orthornormal basis, such as the Fourier basis or the wavelet basis. We note that this method is related to the basis approach described in Section 12.6 of Ramsay and Silverman (1997), except that there a fixed base size and additional penalty are used, similarly to the approach of Leurgans et al. (1993). Another alternative is our proposed Method (4), where the projection is not on a pre-specified basis but rather on the basis defined by the eigenfunctions of the autocovariance operators of X and Y.

Our four proposed methods (1)-(4) are introduced in Section 2. A simulation study is conducted in Section 3 to compare the finite-sample behaviors of these four methods. Since the data were generated with Fourier bases, it is not surprising that method (3) performs the best overall. However, method (4), based on projection on eigenfunctions, comes very close to method (3) and includes estimation of the basis functions. Our recommended method is thus method (4).

Another interesting finding described in Section 3 is that the estimates for the first canonical correlation can be improved greatly for all four methods by using a new procedure, termed first *Empirical Canonical Correlation*, abbreviated by EC_1 and defined in Section 2.5.2. Briefly, we find optimal tuning parameters (number of dimensions, smoothing parameters) by maximizing first a one-leave-out empirical correlation measure. Then we use this empirical correlation itself, with the corresponding estimated tuning parameters plugged in, as estimate for the canonical correlation.

The proposed procedures are illustrated in Section 4 through the correlation of random trajectories of mortality, obtained for cohorts of Mediterranean fruit flies (medflies). The goal in this application is to correlate hazard function for male and female medfly cohorts that are raised in the same cage. It is of biological interest to quantify and describe the nature of the mutual influence of mortality patterns between male and female flies through the weight functions associated with canonical correlation.

2. Methods for Functional Canonical Analysis

We begin this section with a discussion of two basic tools, dimension reduction and dimension augmentation.

2.1. Initial Dimension Reduction

In practice, functional data typically consist of a large number of observed points for each sample curve while the number of observed sample curves is much smaller. One example is the study of pinch force of human fingers by Ramsay, Wang and Flanagan (1995). There were recordings of the finger force for $n = 20$ subjects. The force lasted for about 0.3 seconds and was sampled at a rate of 2000 times per second, leading to 600-dimensional vectors. In practice such high dimensional data cause problems with matrix inversion and eventually lead to computational breakdown. In the pinch force study, Ramsay et al. therefore selected a subset of 41 points from each series of 600 force measurements for the analysis, and thus implemented an initial dimension reduction step (without providing details).

The simple and straightforward approach to FCA, namely to apply multivariate canonical correlation analysis (MCA) directly (see Mardia, Kent and Bibby, 1979) to the high-dimensional data vectors, not only leads to computational breakdown but even if computable, produces estimates of canonical correlation that are close to 1, due to nearly unrestricted flexibility in the choice of weight functions. This feature was already observed by Leurgans et al. (1993) and is confirmed in our own simulation studies (not reported here). Another unwelcome restriction of the classical FCA method is the requirement that the number of measurements (and implicitly also the location of the measurements) is the same from subject to subject. The methods which we propose here are not subject to such restrictions as they involve an initial smoothing or projection step and therefore will perform reasonably well as long as the designs are not too irregular from subject to subject.

Methods for reducing the dimension of a high-dimensional vector of observations are: (a) Systematic sub-sampling which produces equidistant sub-sampled points; (b) Binning which divides the observed points into a number of bins and produces new observations of reduced dimension, corresponding to the averages of the observations falling into the bins, with assumed locations in the middle of each bin; (c) Smoothing the observations and equidistant sub-sampling from the smoothed observations, at an equidistant grid of lower dimension.

Method (a), systematic sub-sampling, provides a natural and simple way for dimension reduction. However, the data points not sampled are entirely ignored. Another concern is that the starting point for the systematic sub-sampling may affect the data analysis. In simulation studies not reported here, we found that this procedure in fact is sensitive to the choice of the starting point, see He (1999), Section 5.3.

Binning is a widely used dimension reduction method. Since it is based on averages over bins, no data are thrown away. However, when the full dimension cannot be divided by a given number of bins with equal length, one would face bins with unequal sizes. For example, for $m = 50$, and a target dimension of 9, one can consider the following bin selections (the numbers represent the number of points in a bin): (6,6,6,6,6,6,6,6,2), (2,6,6,6,6,6,6,6,6), and (6,5,6,5,6,5,6,5,6). Simulations (not reported here, see He, 1999) show that these three different bin selections may lead to quite different outcomes; such effects are familiar from the behavior of histograms.

Smoothing and subsequent sub-sampling avoids some of these problems, and performed well in our simulations (not reported here). An example of pre-smoothing can be found in the gait study by Rice and Silverman (1991). The data consist of the angular rotations of the knee and hip for 39 normal 5-year-old children. A number of frames were taken by cameras over a gait cycle consisting of one complete step taken by each child. For each cycle, 16 to 22 measurements were digitized, and 20 points per cycle were produced by linear interpolation for data analysis. The linear interpolation is the smoothing method used in this example.

This example also illustrates how pre-smoothing may be employed to convert irregularly sampled functional (or curve) data to a set of data with identical sampling points. Smoothing provides a simple and straightforward method to deal with the irregularity in the data that is often caused by missing observations, and to obtain smooth weight function estimates. For these reasons, we include smoothing and subsequent sub-sampling steps in our procedures, implementing the smoothing steps by local polynomial fitting (Stone, 1977, Müller, 1987, and Fan and Gijbels, 1996).

Let $t_i, i = 1, \ldots, n$, be a (fixed) grid of predictors which have a design density (see Müller, 1984), and assume that observations W_i at t_i are generated by $W_i = m(t_i) + \varepsilon_i$, where ε_i is a measurement error, $E(\varepsilon_i) = 0, \mathrm{Var}(\varepsilon_i) = \sigma$, and $m(\cdot)$ is a smooth regression function. Let $K(\cdot) \geq 0$ be a kernel function which is usually assumed to be symmetric, and define scaled versions $K_h(z) \equiv K(z/h)/h$ with a scale factor or bandwidth $h > 0$. Then consider the weighted sum of squares expression

$$\mathrm{WSS}(z; \beta_0, \ldots, \beta_p) = \sum_{i=1}^{n} \{W_i - \sum_{j=0}^{p} \beta_j (t_i - z)^j\}^2 K_h(t_i - z).$$

Denote by $\hat{\beta}_j(z)(j = 0, \ldots, p)$ the minimizers of $\mathrm{WSS}(z; \beta_0, \ldots, \beta_p)$. Finally, we set $\hat{m}(z) = \hat{\beta}_0(z)$, for which an explicit formula is available.

We use the notation

$$\hat{m}(z) = S(z, (t_i, W_i)_{i=1,\ldots,n}, h)$$

to indicate that we smooth the scatter plot (t_i, W_i), using bandwidth h and evaluating the smoothed function at the point z. If the kernel K has support $[-1, 1]$, only data (t_i, W_i) with t_i falling into the window $[z-h, z+h]$ will be used in the computation of the smoothed estimate $\hat{m}(z)$. The order of the polynomial that is fitted locally is often chosen as $p = 1$, corresponding to local linear fitting.

We implement here local linear fitting with the box kernel

$$K(t) = 0.5 \text{ for } |t| < 1, \; = 0 \text{ otherwise}, \tag{4}$$

for the initial dimension reduction step. This kernel choice corresponds to using unweighted sums of squares in the local least squares step, and is the simplest possible choice, especially as smoothness considerations do no play a role in this step.

Let

$$X_i = \{X_i(t_j)\}_{j=1}^{m_x} \text{ and } Y_i = \{Y_i(r_j)\}_{j=1}^{m_y}, \text{ for } i = 1, \ldots, n,$$

be the observed sample processes, S_{DR} be the local polynomial smoother with the box kernel (4) used for dimension reduction, and let m_1 and m_2 be the target dimensions. Assuming that $m_1 = m_2 = m$, we obtain reduced data vectors $\tilde{X}_i, \tilde{Y}_i$ of dimension m,

$$\tilde{X}_i = S_{DR}(\boldsymbol{s}, (t_j, X_i(t_j))_{j=1,\ldots,m_x}, b_x) = \{\tilde{X}_i(s_j)\}_{j=1,\ldots,m} \in \mathcal{R}^m, i = 1, \ldots, n,$$

$$\tilde{Y}_i = S_{DR}(\boldsymbol{s}, (r_j, Y_i(r_j))_{j=1,\dots,m_y}, b_x) = \{\tilde{Y}_i(s_j)\}_{j=1,\dots,m} \in \mathcal{R}^m, i = 1, \dots, n,$$

where the evaluation points $\mathbf{s} = (s_1, \dots, s_m)$ are chosen as the mid-points of m equal- sized sub-intervals that partition $[0, T]$. We will use this notation throughout the following.

Regarding bandwidths b_x and b_y, the heuristic choices $b_x = T/2m, b_y = T/2m$ gave reasonable results in simulations. These simulations, not reported here, provide evidence that S_{DR} overall is the most attractive option for dimension reduction. After the initial dimension reduction one may perform classical multivariate canonical correlation analysis for the resulting m-dimensional data.

2.2. Dimension Augmentation

Smooth canonical weight functions can be obtained as vectors with the same dimension as the original functional data by performing another smoothing step. In this smoothing step, the number of points in the output grid will be larger than the number of points in the input grid. We refer to this as a data augmentation step. More specifically, if k-th order canonical weight vectors of reduced dimension m are given,

$$\{\breve{u}_k(\boldsymbol{s}), \breve{v}_k(\boldsymbol{s})\} = \{\breve{u}_k(s_j), \breve{v}_k(s_j)\}_{j=1,\dots,m} \in \mathcal{R}^{2m},$$

where $s_j \in [0, T]$, for $j = 1, \dots, m$, and $\boldsymbol{s} = (s_1, \dots, s_m)$, we can augment the dimensions of these vectors with the aim to restore the initial dimensions through such a data augmentation step. If S_{DA} denotes the augmentation smoother, this can be formally defined as

$$\tilde{u}_{\alpha,k} = S_{DA}(\boldsymbol{t}, (s_j, \breve{u}_k(s_j))_{j=1,\dots,m}, h) = \{\breve{u}_{\alpha,k}(t_j)\}_{j=1,\dots,m_x} \in \mathcal{R}^{m_x},$$

$$\tilde{v}_{\alpha,k} = S_{DA}(\boldsymbol{r}, (s_j, \breve{v}_k(s_j))_{j=1,\dots,m}, h) = \{\breve{v}_{\alpha,k}(r_j)\}_{j=1,\dots,m_y} \in \mathcal{R}^{m_y},$$

where again S_{DA} may be implemented via local linear fitting.

If the smoothing spline method of FCA is employed, as proposed by Leurgans et al. (1993), the resulting smoothed canonical weight functions are anchored on the reduced dimension m. A dimension augmentation step can then be implemented for this method as well, in order to obtain reasonably smooth weight function estimates and to calculate the functional canonical variates U_i and V_i (see Method 2 below).

2.3. Procedures Using Initial Dimension Reduction and Dimension Augmentation

Two of our methods fall under this rubric. They differ mainly in the way the smoothing steps are implemented. Method 1 is using local polynomial smoothing while Method 2 is a combination of the smoothing spline method of Leurgans et al. (1993) and dimension reduction/augmentation.

2.3.1. Method 1. Functional Canonical Analysis with Local Polynomial Smoothing (FCA-LP)

(i) Reduce the dimension to m with the dimension reduction smoother S_{DR} in section 2.1, with bandwidth choices as described there;
(ii) Obtain canonical weight vectors with MCA, performed on the m-dimensional data;
(iii) Estimate smooth weight functions by employing the dimension augmentation smoother S_{DA} in section 2.2. We use the Epanechnikov kernel,

$$K(t) = \frac{3}{4}(1-t^2) \text{ for } |t| < 1, \text{and } = 0 \text{ otherwise}, \tag{5}$$

for data augmentation. Using this kernel provides us with desirable continuous function estimates for the canonical weight functions. Bandwidths for the data augmentation step are chosen by cross-validation (see Section 2.5 below).
(iv) Estimate functional canonical correlations by the plug-in formula with the smoothed canonical weight functions

$$\tilde{r}_{\alpha,k} = \frac{|\tilde{u}_{\alpha,k}^T \hat{R}_{XY} \tilde{v}_{\alpha,k}|}{\sqrt{(\tilde{u}_{\alpha,k}^T \hat{R}_{XX} \tilde{u}_{\alpha,k})(\tilde{v}_{\alpha,k}^T \hat{R}_{YY} \tilde{v}_{\alpha,k})}}. \tag{6}$$

Here,

$$\hat{R}_{XX} = \frac{\mathbf{X}^T\mathbf{X}}{n-1},\ \hat{R}_{YY} = \frac{\mathbf{Y}^T\mathbf{Y}}{n-1},\ \text{and } \hat{R}_{XY} = \frac{\mathbf{X}^T\mathbf{Y}}{n-1},$$

are the sample covariance operators from the observed functional data matrices

$$\mathbf{X} = \{X_i(t_j) - \bar{X}(t_j)\}_{j=1}^{m_x}, \mathbf{Y} = \{Y_i(r_j) - \bar{Y}(r_j)\}_{j=1}^{m_y}, i = 1, \ldots, n,$$

with

$$\bar{X}(t_j) = \frac{1}{n}\sum_{i=1}^{n} X_i(t_j), \bar{Y}(r_j) = \frac{1}{n}\sum_{i=1}^{n} Y_i(r_j),$$

and $\tilde{u}_{\alpha_{1,k}}, \tilde{v}_{\alpha_{1,k}}$ are the canonical weight functions obtained in step (iii). This formula is motivated by an equivalent definition of the targeted functional canonical correlation as

$$\rho_k = \frac{|u_k^T R_{XY} v_k|}{\sqrt{(u_k^T R_{XX} u_k)(v_k^T R_{YY} v_k)}},$$

that is provided in He et al. (2000).

We note that the dimension reduction step includes necessary regularization as the resulting canonical correlations are obtained in a relatively low-dimensional space and are therefore stable; the reduced dimensions m will have to increase with $m = m(n) \to \infty$ in order to obtain consistency. The subsequent smoothing step serves the purpose to achieve continuous function estimates while preserving consistency . The resulting smoothed weight functions do not exactly conform to the orthonormality conditions. But the discrepancy to orthonormality with respect to R_{XX} vanishes as $m(n) \to \infty$ and the smoothed weight functions move close to the true weight functions. It is also possible, but will lead to little gain while being computationally expensive, to orthonomalize the estimated weight functions with respect to R_{XX} by adding a numerical Gram-Schmidt orthonormalization step.

2.3.2. Method 2. Functional Canonical Analysis with Smoothing Splines (FCA-SP)

(i) Reduce the data dimension to m with dimension reduction smoothers S_{DR} analogous to (i) of Method 1;

(ii) Obtain canonical weight functions with smoothing splines for the reduced dimension data, following the procedure of Leurgans et al. (1993), see also Ramsay and Silverman (1997);

(iii) Data augmentation step, the same as step (iii) of Method 1, but choosing fixed bandwidths $b_x = T/2m, b_y = T/2m$, the same bandwidths as for the dimension reduction step;

(iv) Estimate the functional canonical correlation by means of formula (5) of Leurgans et al. (1993). Substituting the smoothed canonical weight functions, this leads to

$$\hat{\rho}_\alpha(u, v) = \frac{(u^T \hat{R}_{12} v)^2}{(u^T \hat{R}_{11} u + \alpha \int u''^2)(v^T \hat{R}_{22} v + \alpha \int v''^2)},$$

where α is the smoothing parameter, corresponding to a roughness penalty.

2.4. Procedures Using Projections on Basis Functions

The methods described above achieve regularization by an initial dimension reduction step. Another approach which implicitly provides dimension reduction is regularization with basis functions, projecting the L_2-processes X and Y onto a finite number of basis functions. Similar ideas have been discussed in Ramsay and Silverman (1997). We discuss two classes of pertinent basis functions. On one hand, one may use a pre-specified basis such as the Fourier or wavelet basis; on the other hand, data-driven basis functions such as the eigenfunctions of the autocovariance operator (see He et al., 2001) are of interest.

2.4.1. Method 3. Functional Canonical Analysis Using the Fourier Basis (FCA-FB)

(i) Obtain the coefficients of truncated basis expansions of X and Y as follows: Let $\{\theta_l(t)\}_{l=1}^{\infty}$ be an orthonormal basis, e.g. the Fourier basis of $L_2([0,T])$. Then approximate the L_2-processes X and Y by

$$\tilde{X}_i(t_j) = \sum_{l=1}^{k_1} \xi_{il}\theta_l(t_j), j = 1, \ldots, m_x, \tilde{Y}_i(r_j) = \sum_{l=1}^{k_2} \zeta_{il}\theta_l(r_j), j = 1, \ldots, m_y,$$

with Fourier coefficients ξ_{il} and ζ_{il}. Assume for simplicity that $k_1 = k_2 = k$. In matrix notation, the projection on the first k elements of the Fourier basis can be expressed as

$$\tilde{\mathbf{X}} = \boldsymbol{\Xi}\boldsymbol{\Theta}_{\mathbf{1}}^T, \; \tilde{\mathbf{Y}} = \mathbf{Z}\boldsymbol{\Theta}_{\mathbf{2}}^T,$$

where $\tilde{\mathbf{X}} = (\tilde{X}_i(t_j))$, $\tilde{\mathbf{Y}} = (\tilde{Y}_i(r_j))$, $\boldsymbol{\Xi} = (\xi_{il})$, $\mathbf{Z} = (\zeta_{il})$, $\boldsymbol{\Theta}_{\mathbf{1}} = (\theta_l(t_j))$, and $\boldsymbol{\Theta}_{\mathbf{2}} = (\theta_l(r_j))$. With $\mathbf{X} = (X_i(t_j))$, $\mathbf{Y} = (Y_i(r_j))$, the Fourier coefficient matrices may be estimated by

$$\hat{\boldsymbol{\Xi}} = \frac{T}{m_x}\mathbf{X}\boldsymbol{\Theta}_{\mathbf{1}}{}^T, \; \hat{\mathbf{Z}} = \frac{T}{m_y}\mathbf{Y}\boldsymbol{\Theta}_{\mathbf{2}}{}^T.$$

(ii) Obtain canonical correlations $\hat{\rho}$ and weight vectors $\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}$ with MCA, using the k-dimensional coefficient vectors of the Fourier expansion;

(iii) Defining $\boldsymbol{\theta}(t) = (\theta_1(t), \ldots, \theta_k(t))^T$, obtain smoothed weight functions by

$$\hat{\boldsymbol{u}}(t) = \hat{\boldsymbol{u}}^T \boldsymbol{\theta}(t),\ \hat{\boldsymbol{v}}(t) = \hat{\boldsymbol{v}}^T \boldsymbol{\theta}(t).$$

2.4.2. Method 4. Functional Canonical Analysis with Eigenbase Functions (FCA-EB)

(i) According to the Karhunen-Loève decomposition (see Ash and Gardner, 1975), we can approximate the observed processes by

$$\breve{X}_i(t_j) = \sum_{l=1}^{k_1} \xi_{il} \breve{\theta}_l(t_j), j = 1, \ldots, m_x,\ \breve{Y}_i(r_j) = \sum_{l=1}^{k_2} \zeta_{il} \breve{\phi}_l(r_j), j = 1, \ldots, m_y,$$

where $\{\breve{\theta}_l(t_j)\}_{l=1}^{k_1}$ and $\{\breve{\phi}_l(r_j)\}_{l=1}^{k_2}$ are the first k_1 resp. k_2 eigenvectors for the estimated covariance matrices of X and Y.

(ii) Applying a smoothing step to these eigenvectors with the local polynomial smoother S_{LP}, one obtains estimated smooth eigenfunctions

$$\tilde{\theta}_{h,l} = S_{LP}(\boldsymbol{s}, (t_j, \breve{\theta}_l(t_j))_{j=1,\ldots,m_x}, h) = \{\breve{\theta}_{h,l}(s_j)\}_{j=1,\ldots,m_x} \in \mathcal{R}^{m_x},$$

$$\tilde{\phi}_{h,l} = S_{LP}(\boldsymbol{t}, (r_j, \breve{\phi}_l(r_j))_{j=1,\ldots,m_y}, h) = \{\breve{\phi}_{h,l}(t_j)\}_{j=1,\ldots,m_y} \in \mathcal{R}^{m_y}.$$

Here we employ the Epanechnikov kernel (5) and cross-validation for bandwidth choice. Next we expand the sample processes in terms of the smoothed eigenfunctions to obtain estimated principal components

$$\hat{\boldsymbol{\Xi}} = \frac{T}{m_x} \mathbf{X} \tilde{\boldsymbol{\Theta}}^T,\ \hat{\mathbf{Z}} = \frac{T}{m_y} \mathbf{Y} \tilde{\boldsymbol{\Phi}}^T.$$

We note as above, that estimated eigenfunctions are not exactly orthonormal due to the smoothing step. But as the smoothed functions converge to the true functions, this problem vanishes asymptotically, and moreover, experience shows that in almost all cases, it is practically negligible.

(iii) Obtain estimated functional canonical correlations $\hat{\rho}_h$ and weight vectors $\hat{\boldsymbol{u}}_h, \hat{\boldsymbol{v}}_h$ by MCA for these estimated coefficients vectors;

(iv) Obtain smooth weight function estimates

$$\hat{\boldsymbol{u}}_h(t) = \hat{\boldsymbol{u}}_h^T \tilde{\boldsymbol{\theta}}_h(t),\ \hat{\boldsymbol{v}}_h(t) = \hat{\boldsymbol{v}}_h^T \tilde{\boldsymbol{\phi}}_h(t),$$

where $\tilde{\boldsymbol{\theta}}_h(t) = (\tilde{\theta}_{h,1}(t), \ldots, \tilde{\theta}_{h,m_x}(t))^T$, $\tilde{\boldsymbol{\phi}}_h(t) = (\tilde{\phi}_{h,1}(t), \ldots, \tilde{\phi}_{h,m_y}(t))^T$.

2.5. Functional Canonical Analysis with Cross-Validation

Similar to other functional data analysis methods, the Functional Canonical Analysis (FCA) methods proposed in the previous section require the selection of a tuning parameter. For example, for FCA-LP, we need to choose the target dimension m in the initial dimension reduction step and also the bandwidth for the smoother S_{DR}. For FCA-SP, we need to choose the target dimension m and the smoothing parameter of the smoothing spline. For FCA-FB, since there is no need for initial dimension reduction, we only need to choose the number of Fourier base functions determining the expansion. For FCA-EB, we need to choose the number of eigenfunctions and a smoothing bandwidth as well. All of these choices may be derived from a version of leave-one-out cross-validation.

2.5.1. Selecting the Tuning Parameters

Let α stand for the parameters (usually there are several) to be selected. For FCA-LP, we have α = (target dimension, smoothing bandwidth); for FCA-SP, α = (target dimension, smoothing parameter); for FCA-FB, α = number of base functions; for FCA-EB, α = (smoothing bandwidth, number of base functions). For computational simplicity we use the same tuning parameters for both X and Y processes. Let

$$CV_{\alpha,j} = \mathrm{Corr}(\{\tilde{u}_{\alpha,j}^{(-i)T} X_i\}_{i=1}^n,\ \{\tilde{v}_{\alpha,j}^{(-i)T} Y_i\}_{i=1}^n), \tag{7}$$

where $j = 1, \ldots, m,\ (X_i, Y_i)$ is the i-th pair of observed curves, and $\tilde{u}_{\alpha,j}^{(-i)}$ and $\tilde{v}_{\alpha,j}^{(-i)}$ are the j-th weight functions obtained by a FCA procedure, using tuning parameters α from the sample data with the i-th curve omitted. Here, Corr is the sample correlation.

For the simulations, we choose $j = 1$ and 4. We refer to the method that selects the value of maximizing $CV_{\alpha,1}$ as MAX-CV1, and the method that selects the value of minimizing $CV_{\alpha,4}$ as MIN-CV4, leading to

$$\hat{\alpha}_{\max} = \arg\max_{\alpha} CV_{\alpha,1} \text{ for the MAX-CV1 method, and} \tag{7a}$$

$$\hat{\alpha}_{\min} = \arg\min_{\alpha} CV_{\alpha,4}, \text{ for the MIN-CV4 method.} \tag{7b}$$

The corresponding estimate for ρ_j based on (6) is

$$\hat{R}_j = \tilde{r}_{\hat{\alpha},j}, \tag{8}$$

where $\hat{\alpha} = \hat{\alpha}_{\max}$ or $\hat{\alpha} = \hat{\alpha}_{\min}$. We use both $\hat{\alpha}_{\max}$ and $\hat{\alpha}_{\min}$ to estimate the first three canonical correlations simultaneously. The motivation for the MIN-CV4 method is based on a minimax property of canonical correlation that was established in He et al. (2000). The MIN-CV4 method minimizes the maximal correlation between the two sub-spaces $\text{span}\{\tilde{u}_{\alpha,j}\}_{j\geq 4}$ and $\text{span}\{\tilde{v}_{\alpha,j}\}_{j\geq 4}$ and thus maximizes the canonical correlations between $\text{span}\{\tilde{u}_{\alpha,j}\}_{j=1,2,3}$ and $\text{span}\{\tilde{v}_{\alpha,j}\}_{j=1,2,3}$. This motivates the choice $\hat{\alpha}_{\min}$ when estimating ρ_1, ρ_2, ρ_3 simultaneously.

A cross-validation procedure similar to MAX-CV1 was proposed by Leurgans et al. (1993). This method is motivated by the definition of the first canonical correlation (5) in Leurgans et al. and is aimed at estimating the first canonical correlation ρ_1. The well-know variability of cross-validation tuning parameter choice can be a problem and can negatively impact performance, but we found that one-leave-out cross-validation worked well overall in the framework of FCA.

2.5.2. Empirical Canonical Correlations

In addition to estimates (8), we observe that the observed one-leave-out correlations $CV_{\alpha,j}$ themselves can be used to find estimates for canonical correlations. Especially in the case $j = 1$, when using parameter selection $\hat{\alpha}_{\max}$ as in (7a), this method is promising. We call these *Empirical Canonical Correlations*. For ρ_1, the first *Empirical Canonical Correlation* (EC_1) is given by

$$\hat{\rho}_1 = EC_1 = CV_{\hat{\alpha}_{\max},1}. \tag{9}$$

We note that in the determination of EC_1, one aims at selecting smoothing parameters and number of base functions $\hat{\alpha}_{\max}$ with the property that the resulting one-leave-out weight function estimates lead to maximal empirical canonical correlations (7) (for $j = 1$).

In our simulations (Section 3), this turned out to be by far the best estimator for the first canonical correlation, clearly superior over $\hat{R}_1$ in (8) or $\tilde{r}_{\alpha,1}$ in (6).

2.5.3. Alternative Tuning Parameter Selections

Besides $\hat{R}_j$ in (8), $j = 1, 2, \ldots$, alternatives for estimating the first k, say $k = 3$, canonical correlations simultaneously can be based on different choices of the tuning parameters. Consider for example

$$\hat{\alpha} = \arg\max_{\alpha} \frac{\sum_{j=1}^{3} \tilde{r}^2_{\alpha,j} CV_{\alpha,j}}{\sum_{j=1}^{3} \tilde{r}^2_{\alpha,j}}, \tag{10}$$

where initial canonical correlations $\{\tilde{r}_{\alpha,1}, \tilde{r}_{\alpha,2}, \tilde{r}_{\alpha,3}\}$ are estimated by (6) with an FCA procedure and a preliminary value for the parameter α. The $\tilde{r}^2_{\alpha,j}$'s in (10) act as the weights in a sum of weighted CV-values. However, we prefer the MIN-CV4 method to the method based on (10), due to its simplicity.

Except for FCA-FB, all methods involve two-dimensional tuning parameters α , and two search schemes present themselves for obtaining estimates $\hat{\alpha}_{\max}, \hat{\alpha}_{\min}$ (7a) and (7b).

(a) *Two-way search.* Search directly for a maximum or minimum among all possible values in a two-dimensional range. The two-way search maps the entire parameter surface and therefore is guaranteed to locate extrema, but it is computing intensive.

(b) *Two-stage search.* First search for an extremum along only one component of the tuning parameter, and once located, fix the value of the first component where the extremum occurs and search along the second component. This requires to first determine which of the components is the more relevant one. As it turns out, dimension is the important parameter for FCA-LP and FCA-SP, and base size is the influential parameter for FCA-EB.

3. Comparing the Finite Sample Behavior of FCA Methods

We briefly discuss the generation of pseudo-random processes with given canonical correlation structure. Following a suggestion in He et al. (2001), we define processes

$$X(t) = \sum_{i=1}^{k} \xi_i \theta_i(t) \quad \text{and} \quad Y(t) = \sum_{i=1}^{k} \zeta_i \theta_i(t), \;\; t \in [0, T], \tag{11}$$

where $k = 21, T = 50$, and $\{\theta_i(t)\}$ is the Fourier basis on [0, T], with

$$\theta_1(t) = \sqrt{1/T},\ \theta_2(t) = \sqrt{2/T}\sin((t - T/2)2\pi/T),$$
$$\theta_3(t) = \sqrt{2/T}\cos((t - T/2)2\pi/T),\ \ldots,\ \theta_{20}(t) = \sqrt{2/T}\sin((t - T/2)20\pi/T), \tag{12}$$
$$\theta_{21}(t) = \sqrt{2/T}\cos((t - T/2)20\pi/T).$$

Furthermore, let $\boldsymbol{\xi}= \{\xi_i\}$ and $\boldsymbol{\zeta}= \{\zeta_i\}$ be Gaussian vectors with covariance matrices

$$\mathrm{Cov}(\boldsymbol{\xi}) = R_{11},\ \mathrm{Cov}(\boldsymbol{\zeta}) = R_{22},\ \ \text{and}\ \ \mathrm{Cov}(\boldsymbol{\xi},\boldsymbol{\zeta}) = R_{12},$$

where $R_{11}, R_{22}, \mathrm{and} R_{12}$ are chosen as diagonal matrices with

$$\mathrm{diag}(R_{11}) = \mathrm{diag}(R_{22}) = \{10 \cdot 0.75^i\},\ i = 0, 0, 0, 1, 2, \ldots, 18,$$

and

$$\mathrm{diag}(R_{12}) = \{7, 3, 1, 0, 0, \ldots, 0\}.$$

Then the canonical correlations for random processes X and Y are

$$\rho_1 = .7,\ \rho_2 = .3,\ \rho_3 = .1,\ \rho_4 = \ldots = \rho_{21} = 0,$$

and the canonical weight functions are

$$u_k(t) = v_k(t) = \sqrt{.1}\theta_k(t), \quad \text{for } k = 1, 2, 3, \tag{13}$$

and 0 otherwise.

To generate pseudo-random versions of processes X and Y, we start with a $2k$-dimensional standard pseudo-Gaussian vector $\boldsymbol{z}$ and then, defining

$$A = \begin{pmatrix} R_{11} & R_{12} \\ R_{12}^T & R_{22} \end{pmatrix},$$

obtain $(\boldsymbol{\xi}^T, \boldsymbol{\zeta}^T)^T = A^{1/2}\boldsymbol{z}$, from which processes (11) are generated.

In our simulation studies, we generate each sample with 50 pairs of observed pseudo-random processes, $n = 50$, and with 50 observed points on each observed process, $m_x = m_y = 50$. We also choose the observed points $\{t_j\} \subset [0, T]$, where $T = 50$, to be equi-spaced. The finite sample comparisons are based on 100 Monte Carlo samples.

The following table lists means and MSEs for estimates for the first canonical correlation obtained by using the best possible parameters, where Dim=dimension after dimension reduction, BW=bandwidth, SP=spline smoothing parameter, and BS=number of basis functions. For example, the best dimension choice over all Monte Carlo runs was Dim=4 for FCA-LP and Dim=13 for FCA-SP, and analogously for the other choices. These parameters are not selected from the data. Thus these results are not achievable in practice where the optimal values of the parameters are unknown. Methods FCA-LP, FCA-SP, FCA-FB, and FCA-EB and the corresponding method-specific parameters are as defined in Section 2.

Table 1. Finite sample behavior of various FCA methods with optimal parameters based on 100 Monte Carlo runs. True value $\rho_1 = .7$.

Method	Parameters	mean $\hat{\rho}_1$	MSE
FCA-LP	Dim=4, BW=22	.7303	.0045
FCA-SP	Dim=13, SP=28	.7023	.0036
FCA-FB	BS=1	.7082	.0038
FCA-EB	BS=3, BW=2.51	.7328	.0047

Results for the four methods using data-based parameter selections are shown in Table 2, which also contains results for different parameter selection techniques. Theses methods and parameter selections based on two-way or two-stage search have been defined in Section 2.5. Here the AISE of the weight function is given by the average of $\| u_1 - \hat{u}_1 \|^2 + \| v_1 - \hat{v}_1 \|^2$ over the 100 simulations, and the mean squared errors (MSE) of the various canonical correlation estimates are given by the averages of $(\hat{\rho}_i - \rho_i)^2, i = 1, 2, 3,$ over the 100 simulations.

Table 2. Finite sample behavior of various FCA methods under data-based parameter choices for 100 Monte Carlo runs. True Canonical correlations are .7, .3, and .1, then 0.

FCA Method	Selection Method	Target Criterion	MSE $\hat{R}_1$	MSE $\hat{R}_2$	MSE $\hat{R}_3$	MSE EC_1	AISE Weight Functions
FCA-LP	Two-Way	Max-CV1	.0129	.0563	.0826	.0057	.7834
		Min-CV4	.0416	.0518	.0766		.3570
	Two-Stage	Max-CV1	.0146	.0648	.0987	.0058	.8331
		Min-CV4	.0297	.1073	.1554		.5279
FCA-SP	Two-Way	Max-CV1	.1217	.0655	.0585	.0045	.7349
		Min-CV4	.0301	.0843	.1223		.5775
	Two-Stage	Max-CV1	.0216	.0580	.0798	.0061	.8965
		Min-CV4	.0310	.1150	.1698		.5975
FCA-FB	One-Way	Max-CV1	.0195	.1906	.4288	.0049	.0875
		Min-CV4	.0394	.2395	.3847		.1671
FCA-EB	Two-Way	Max-CV1	.0305	.1481	.2466	.0053	.2336
		Min-CV4	.0275	.1731	.2719		.1670
	Two-Stage	Max-CV1	.0169	.0770	.1188	.0052	.1045
		Min-CV4	.0225	.1442	.2246		.1232

Table 2 suggests that (1) the newly proposed estimator EC_1 for estimating the first canonical correlation is strongly preferable over the conventional estimate $\hat{R}_1$; (2) the estimation of canonical correlation gets harder as the order of the correlation increases, i.e., the estimates of the second canonical correlation have a considerably larger error than the estimates of the first canonical correlation; (3) MIN-CV4 is consistently better in regard to the AISE of the weight functions than MAX-CV1 (with two exceptions, the two-stage FCA-EB and the FCA-FB procedures), but is usually worse regarding estimation of the canonical correlations; (4) Comparing the methods in terms of weight function estimation, FCA-FB provides the best weight function estimates, while FCA-LP and FCA-SP are giving the worst results, with about 10 times larger AISE as compared to FCA-FB. This is perhaps not surprising as the Fourier basis is the true underlying basis here; (5) Comparing the methods in terms of estimating canonical correlations with estimators $\hat{R}_1 - \hat{R}_3$, based on Max-CV1, two-way FCA-LP produces throughout the best results, except for $\hat{R}_3$, where the best results are achieved by FCA-SP. Both FCA-FB and FCA-EB are less competitive when using these estimators (one should remember that estimators $\hat{R}_1$ for ρ_1 are clearly suboptimal when compared to EC_1, however); (6) Comparing the methods with regard to the behavior of EC_1, by far the best estimator for the first canonical correlation, one finds that FCA-SP two-way and FCA-FB one-way are best, closely followed by FCA-EB two-way and two-stage. The other methods are somewhat less competitive.

Our conclusion is that overall, the newly proposed estimator EC_1 is clearly the best estimator for the first functional correlation coefficient ρ_1. Among the FCA methods, the FCA-EB two-stage procedure overall is the best; while FCA-FB results are also good, this may be due to the fact the processes were generated based on the Fourier basis itself and therefore this base is expected to perform well. If one plans to use estimates $\hat{R}_1, \hat{R}_2, \hat{R}_3$, one can also consider FCA-LP two-way; if one strictly plans to estimate ρ_1 via EC_1, then FCA-SP two-way is a good if expensive option. If estimation of the weight functions is of interest, two-stage FCA-EB is clearly the winner. The estimation of the first three weight functions with method FCA-EB two-stage is illustrated in Figure 1 for three typical samples. It can be seen that the estimates for the first three canonical weight functions are fairly accurate, except possibly by a scale factor. Our overall recommendation based on this limited simulation study is to use FCA-EB two-stage (which has the additional benefit that it is computationally less expensive than two-way methods).

4. Functional Canonical Analysis for Medfly Data

In this section we present an application of our proposed FCA methods to lifetime data collected for cohorts of male and female medflies which were raised in the same cages. The experiment concerns the survival of cohorts of male and female Mediterranean fruit flies

(*Ceratitis capitata*). The experiment was carried out to investigate mortality patterns of medflies. The motive for this study stemmed from our interest in determining trajectories of mortality for male and female medflies from a sample involving a total of more than 1.2 million individual medflies (see Carey et al. 1992), all of which were sugar-fed. Of interest is the relation of mortality trajectories between male and female medflies which were raised in the same cage. One would like to know whether there is a cage factor of mortality and moreover whether the shapes of the weight functions provide clues about the nature of the influence that male and female mortality exercise on each other.

There were 167 cages with flies raised at different time periods. Since there may be a period effect, we focus our analysis on one period with 48 cages. Each of the 48 cohorts of medflies was kept in a separate cage that contained approximately 4000 male and 4000 female medflies. For each cohort, the number of flies alive was recorded daily, and all cohorts had survivors beyond day 40. The observed processes $X_i(t)$ and $Y_i(t),\ t = 1, \ldots, 40$, are the hazard functions estimated separately for male and female flies (Capra and Müller, 1997; for smoothing of hazard rates based on lifetables compare Wang, Müller and Capra, 1998). An outlier was removed from the sample of male hazard functions. Figure 2 shows the sample of estimated hazard functions, smoothed with bandwidth = 5 days, for male and female medflies. This set of paired male/female hazard functions, consisting of $n = 47$ pairs of functions with support [0, 40], is the sample of pairs of curves for which we wish to perform FCA. In this data analysis, we focus on the FCA-FB and FCA-EB methods.

Principal components analysis in the MDA sense was performed to check the base size for the finite basis function expansion. The first two eigenvalues of the male and female covariance matrices were found to account for more than 89% of the total variation explained, which we took to indicate that a finite expansion with the first two eigenfunctions was reasonable.

We also estimated canonical correlations at different fixed dimensions, bandwidths, and basis sizes to check how robust the estimates are with respect to changes in the tuning parameters. The methods show consistently that the first canonical correlation is relatively high, ranging from .8626 to 1.000. The value increases as the number of basis functions increases. The amount of smoothing has little influence on the values obtained with FCA-EB when the basis size is small (basis size less than 3, for example). The second and third canonical correlations estimated with FCA-FB are also relatively high. These results show that a data-driven choice of the tuning parameters such as cross-validation is a necessity.

Table 3 lists the estimates obtained with FCA-FB and FCA-EB, implemented with tuning parameter choice via cross-validation, for various search schemes and selection criteria.

Table 3. FCA with Cross-Validation For the Medfly Hazard Functions

Procedure	Search	Criterion	BW	BS	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_3$	EC_1
FCA-FB	One-Way	Max-CV1	5.00	39	1.000	.9999	.9982	.9987
		Min-CV4	5.00	15	.9921	.9715	.9243	NA
FCA-EB	Two-Way	Max-CV1	5.00	30	1.000	.9997	.9992	1.000
		Min-CV4	4.02	15	.9648	.9281	.8626	NA
	Two-Stage	Max-CV1	18.1	2	.9259	.6083	NA	.9181
		Min-CV4	4.02	15	.9648	.9281	.8626	NA

We note that the range of the selected number of basis functions by different cross-validation methods is quite large. We also note that the estimated weight functions have similar characteristics across varying tuning parameter choices. The results from FCA-EB two-stage with MAX-CV1 which selects two basis functions appears to be a good choice, judging from the simulation results reported previously. Because FCA-EB two-stage performs better than both FCA-EB two-way and FCA-FB one-way, we regard the estimate produced by FCA-EB two-stage with MAX-CV1 and with EC_1 as the best choice for this application. This yields the estimate $\hat{\rho}_1 = EC_1 = .9181$. For the second canonical correlation we use $\hat{R}_2 = .6083$.

For weight function estimation, we use the FCA-EB two-stage MIN-CV4 method which performed well in the simulation. Other choices for smoothing parameters led to weight functions with similar characteristics. The weight functions are useful to interpret the nature of the canonical correlation between male and female mortality trajectories. Figure 3 displays the first three pairs of estimated weight functions estimated by the FCA-EB two-stage MIN-CV4 method.

The u_1/v_1 plot shows that the male weight function stays relative constant near zero for the first 20 days with a small bump near day 12, and then gradually increases towards the tail. The female weight function decreases during the first 10 days, and then accelerates from day 10 to about day 28, crosses the male's weight function, and then stays relatively close to the male weight function to the end. The interpretation is that the male trajectory exercises little influence to determine the correlation during the first 10 days, with slight positive effects of heightened mortality around day 12, the time of maximum reproduction of the medflies, associated with lowered female mortality around that time. Possible this effect is a consequence of lowered "cost of mating" for females in the presence of heightened male mortality.

After day 20, female and male mortality exercise roughly equal influence on each other. We note that "influence" is not to be interpreted in a causal sense but in the sense of a correlation in an observational study. For example, it can reflect the differential and time-varying manifestation of a cage effect during the life histories of female and male medflies. Overall, these findings indicate that the observed functional canonical correlation between male and female hazard functions is determined for the most part by mortality after day 20, i.e., by the parallel development of male and female mortality in he right tail; this observation is in agreement with the hypothesis of a cage effect.

The second pair of weight functions basically parallels the two peaks of the male and female hazard functions. The male weight function peaks first and is closely followed by the female weight function.

Acknowledgments

We wish to thank two referees for careful reading and helpful comments and especially one referee for detailed scrutiny that led to several corrections and numerous improvements in the paper. Further thanks are due to James Carey for making available to us the medfly data that are used in the data illustration.

References

Ash, R. and Gardner, M. (1975). *Topics in Stochastic Processes*, New York: Academic Press.

Brillinger, D. R. (1985). *Time Series: Data Analysis and Theory*, Second Edition. Holden Day, San Francisco.

Capra, W. B. and Müller, H. G. (1997). An accelerated time model for response curves. *J. Am. Statist. Assoc.* **92**, 72-83.

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

Dauxois, J. and Pousse, A. (1976). Les analyses factorielles en calcul des probabilités et statistique: essai d'étude synthétique. *Thesis.* University of Toulouse, Toulouse.

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

Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J. Australian Math. Soc.* **21**, 229-242.

He, G. (1999). Functional canonical analysis and linear modelling. *Thesis.* University of California, Davis.

He, G., Müller, H-G, and Wang, J-L (2000). Extending correlation and regression from multivariate to functional data. *Asymptotics in Statistics and Probability*, ed. M. Puri, pp. 197-210.

He, G., Müller, H-G, and Wang, J-L (2002). Functional canonical analysis for square integrable stochastic processes. *J. Multiv. Analysis*, in press.

Leurgans, S.E., Moyeed, R.A. and Silverman, B.W. (1993). Canonical correlation analysis when the data are curves. *J. R. Statist. Soc.* **B** 55, 725-740.

Mardia, K. V, Kent, J. T. and Bibby, J. M. (1997). *Multivariate Analysis*, London; New York, Academic Press.

Müller, H.G. (1984). Optimal designs for nonparametric kernel regression. *Statist. Probab. Letters* **2**, 285-290.

Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Am. Statist. Assoc.* **82**, 231-238.

Ramsay, J.O., Wang, X. and Flanagan, R. (1995). A functional analysis of the pinch force of human fingers. *Appl. Statistics* **44**, 17-30.

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

Rice, J.A. and Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* **B** 53, 233-243.

Service, S.K., Rice, J.A. and Chavez, F.P. (1998). Relationship between physical and biological variables during the upwelling period in Monterey Bay, *CA. Deep-Sea Research* **II**, 1669-1685.

Stone, C.J. (1977), Consistent nonparametric regression. *Ann. Statist.* **5**, 595-645.

Wang, J. L., Müller, H. G. and Capra, W. B. (1998). Analysis of oldest-old mortality: lifetables revisited. *Annals of Statistics* **26**, 126-163.

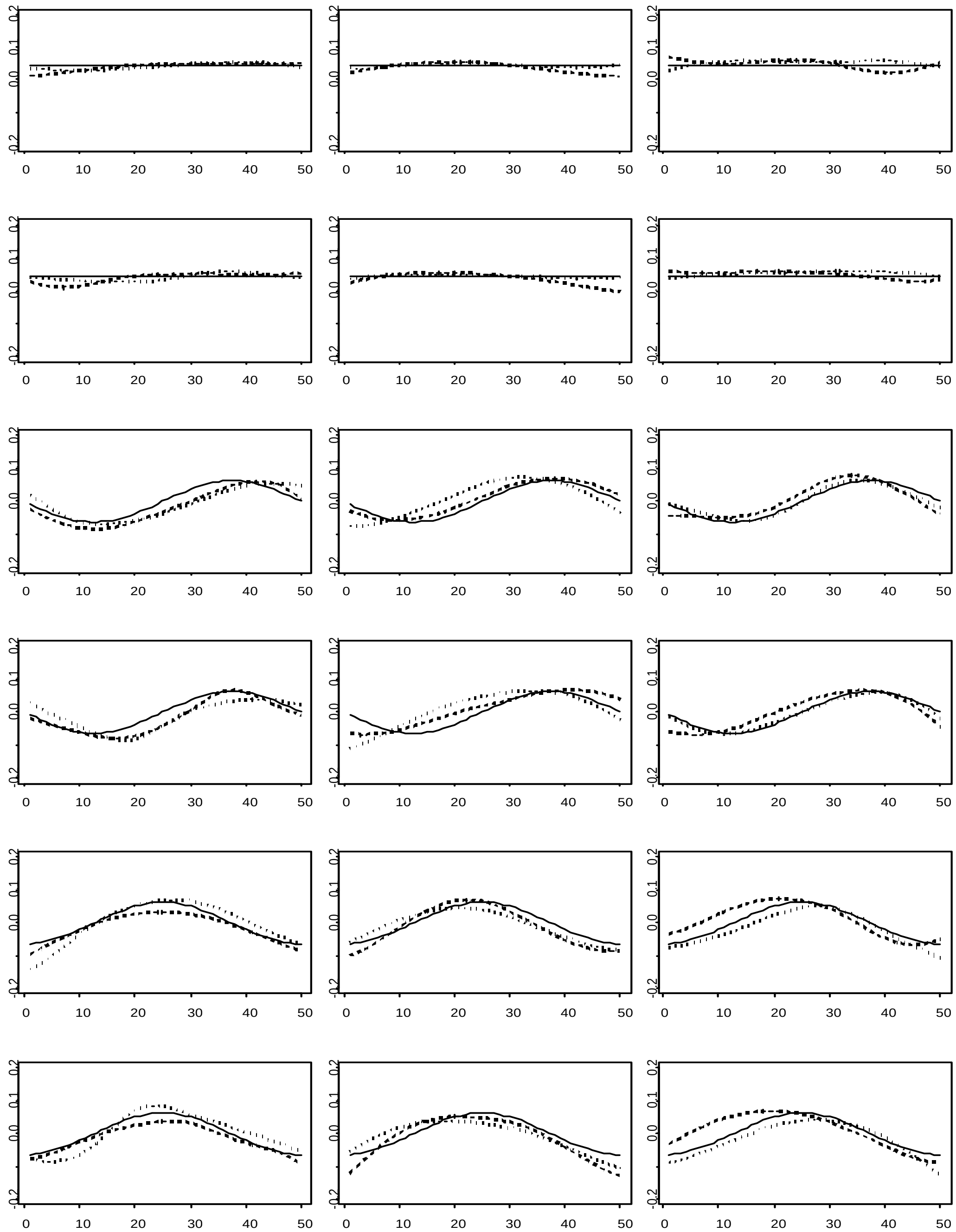

Figure 1: Targeted canonical weight functions (solid), estimated weight functions $u(\cdot)$ (dashed) and $v(\cdot)$ (dotted), based on the eigenbase method (FCA-EB), two-stage, for three random samples (each column corresponds to a random sample). First two rows from above: u_1, v_1, MAX-CV1 (first row) and MIN-CV4 (second row); rows three and four from above: u_2, v_2, MAX-CV1 (third row) and MIN-CV4y (fourth row); rows five and six from above: u_3, v_3, MAX-CV1 (row five) and MIN-CV4 (row six).

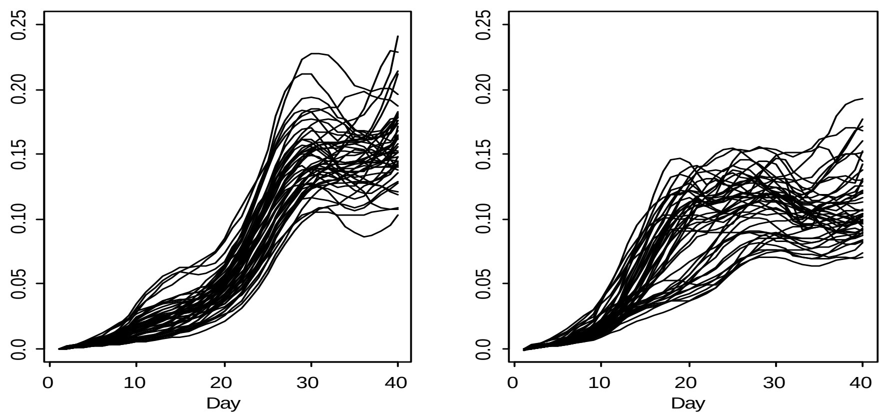

Figure 2: Smoothed hazard rates (using bandwidth 5 days) based on the lifetables for 47 cohorts of male (left) and 47 cohorts of female (right) medflies.

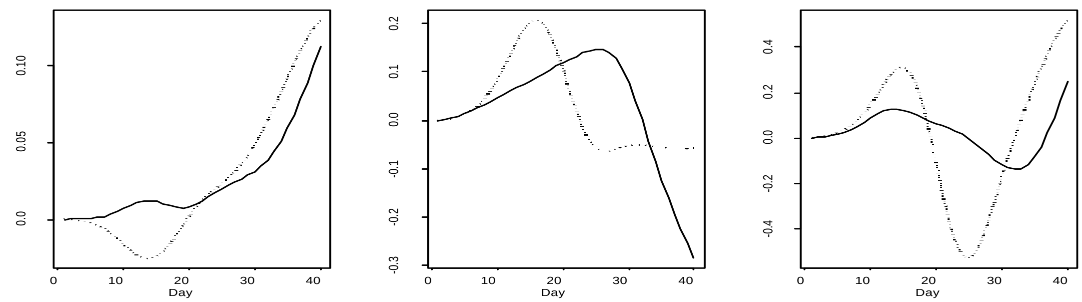

Figure 3: First three pairs (from left to right) of canonical weight functions for the medfly mortality data, using the eigenbasis method (FCA-EB), two-stage version. Male weight functions are solid and female weight functions are dotted.