Linear manifold modeling of multivariate functional data

Jeng-Min Chiou

Academia Sinica, Taiwan

Hans-Georg Müller

University of California, Davis

March 2013

Address for correspondence: Jeng-Min Chiou, Institute of Statistical Science, Academia Sinica, 128 Section 2 Academia Road, Taipei 11529, TAIWAN. E-mail: jmchiou@stat.sinica.edu.tw

Abstract

Multivariate functional data are increasingly encountered in data analysis, while statistical models for such data are not well developed yet. Motivated by a case study where one aims to quantify the relationship between various longitudinally recorded behavior intensities for Drosophila flies, we propose a functional linear manifold model. This model reflects the functional dependency between the components of multivariate random processes and is defined through data-determined linear combinations of the multivariate component trajectories, which are characterized by a set of varying coefficient functions. The time-varying linear relationships that govern the components of multivariate random functions yield insights about the underlying processes and also lead to noise-reduced representations of the multivariate component trajectories. The proposed functional linear manifold model is put to the task for an analysis of longitudinally observed behavioral patterns of flying, feeding, walking and resting over the lifespan of Drosophila flies and is also investigated in simulations.

Keywords: Behavior; Drosophila; Functional data analysis; Longitudinal data; Multivariate random functions; Principal component analysis.

1 Introduction

Multivariate functional data derived from realizations of multivariate random functions typically arise from a collection of simultaneous recordings of several time courses for a sample of subjects or units. Such multivariate longitudinal data are recorded in various areas, including road traffic monitoring, where vehicle speed, flow rate and occupancy are continuously monitored, as well as across the biological and social sciences. While in engineering applications such as traffic monitoring there is often an underlying theory that may lead to postulated relations between the different components of multivariate time courses (see, e.g., Bickel et al. 2007, Li 2008), in biological applications this is typically not the case, as usually not much, if anything, is known a priori about the relationships between various longitudinally recorded components and one is therefore interested in learning what these relationships are.

A typical example is provided by longitudinal behavioral monitoring experiments involving Drosophila flies with the goal to study the relationship between various behaviors as flies age and to characterize age-specific behavioral patterns (Carey et al. 2006). The quantification of age-related changes in the relationship between the intensities of four simultaneously recorded key behaviors of flies (*Flying*, *Feeding*, *Walking*, *Resting*) is of particular interest to determine behavioral patterns that are associated with the process of aging. Smoothed trajectories reflecting the levels of occurrence of these four behaviors in dependence on age are shown in Figure 1 for a randomly selected subsample of 62 flies.

Inspection of these random trajectories indicates substantial random variation across individuals, supporting the concept that these are realizations of a stochastic process with high variability at the individual level. At the same time, there is indication of patterns of behavioral relationships in terms of age-related changes and in the association between various behaviors with each other. For example, flying and feeding behaviors appear to follow more similar patterns than resting and walking. The proposed model will be used to extract and quantify such connections and their relation with aging. Further details about this analysis can be found in Section 5 below.

In addition to behavioral monitoring, there are many other situations in which one encounters multivariate functional data. The modeling and analysis of such data is a practically important statistical problem that has received surprisingly little attention so far. Besides functional regression (see, e.g.,

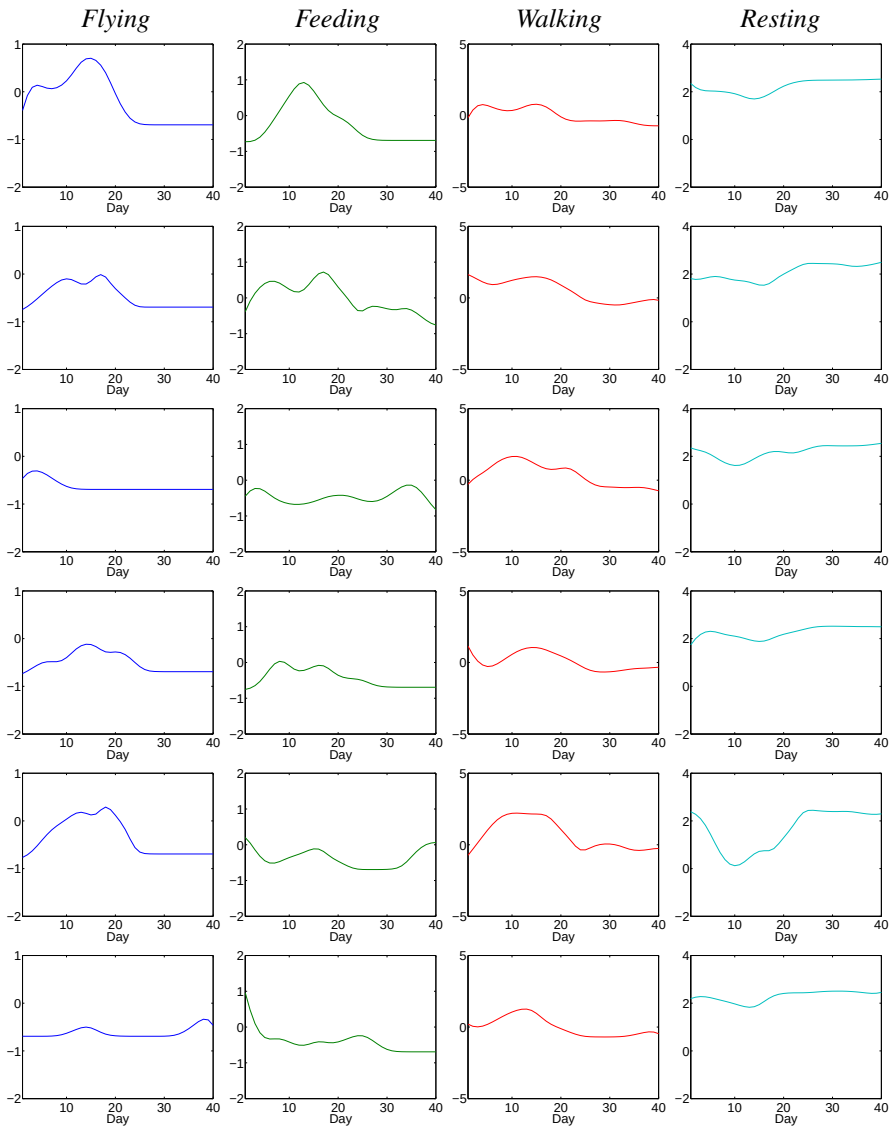

Figure 1: Behavioral trajectories for a random sample of six flies (corresponding to rows). Smoothed trajectories are shown in logarithmic scale for processes U_1 (*Flying*), U_2 (*Feeding*), U_3 (*Walking*), and U_4 (*Resting*) (columns from left to right).

Ramsay and Silverman 2005; Li and Hsing 2007), a focus in multivariate functional data analysis has been the development of measures to quantify the dependency of components of multivariate functional data, such as dynamical correlation (Dubin and Müller 2005) and canonical correlation analysis (Leurgans, Moyeed and Silverman 1993; He, Müller and Wang 2003), with further developments in Cupidon

et al. (2008) and Eubank and Hsing (2008). Related methods include factor analysis in multivariate time series analysis (Pan and Yao 2008).

In this article, we develop statistical methodology for modeling the relations between the components of multivariate functional data. For samples of trajectories corresponding to realizations of multivariate processes, we aim to learn their dynamic dependency structure. For this purpose, we propose a linear manifold model that is characterized by a set of data-determined linear combinations of multivariate component trajectories. These linear combinations are subject to the condition that the integral of their variance vanishes, so that the linear manifolds are global features over the entire time domain. They are determined by varying coefficient functions that satisfy orthonormality constraints and define time-varying relationships between the multivariate functional components.

The coefficient functions may be approximated by expansions into basis functions, such as those derived from the Karhunen-Loève representation. Under mild conditions, the integrated variance of the linear manifold can then be expressed as a quadratic form that is characterized by the cross-covariance between the functional components. This result facilitates the construction of linear manifold models. The dimension of the functional linear manifold is typically small and usually the manifold will correspond to interpretable relationships between the individual component functions of the multivariate functional vector. In addition, the functional linear manifold approach leads to noise-reduced and sparse representations of multivariate longitudinal trajectories.

The article is organized as follows. In Section 2 we present the setting and underlying ideas of modeling time-dynamic relationships for multivariate random trajectories, including estimation of the coefficient functions and the derivation of the noise-reduced multivariate process representation that results from the linear manifold. Consistency properties of the estimated model components are provided in Section 3. The proposed methods are illustrated with an application to multivariate behavioral trajectories in Section 4, followed by a simulation study presented in Section 5. Discussion and concluding remarks can be found in Section 6 and technical derivations are compiled in the Appendix.

2 Functional linear manifold models

2.1 *Preliminaries: Representing multivariate component functions*

Suppose that the observed sample trajectories $\{(X_{1i}, \ldots, X_{Mi}); i = 1, \ldots, n\}$ are sampled from the random M-component functions $(U_1, \ldots, U_M)$, where each component function U_i lies in the Hilbert space $L^2(\mathcal{T})$ of square integrable functions for a compact domain $\mathcal{T}$. We aim to study the underlying dynamic dependency between these M component functions, which we assume to be intrinsically related through K time-dynamic relationships, where K is unknown. Assume that each component function U_j, $j = 1, \ldots, M$, has an unknown smooth mean $\mu_j(t) = EU_j(t)$ and auto-covariance function $G_j(s,t) = \mathrm{cov}(U_j(s), U_j(t))$, for $s, t \in \mathcal{T}$.

Under mild conditions the component covariance functions can be expanded into an orthonormal system, defined by eigenfunctions $\{\phi_{jl}\}$ of the covariance operators with nondecreasing eigenvalues $\{\lambda_{jl}\}$ as l increases, for each $j = 1, \ldots, M$ and $l = 1, 2, \ldots$, as follows,

$$\mathrm{cov}(U_j(t), U_j(s)) = \sum_{l=1}^{\infty} \lambda_{jl}\phi_{jl}(t)\phi_{jl}(s), \tag{1}$$

subject to $\int \phi_{jl}(t)\phi_{jm}(t)dt = \delta_{lm}$, where here and below $\delta_{lm} = 1$ for $l = m$ and $\delta_{lm} = 0$ for $l \neq m$. For notational convenience, we denote the centered process for each component function by

$$\widetilde{U}_j(t) = U_j(t) - \mu_j(t), \ \ j = 1, \ldots, M,$$

where $E\widetilde{U}_j(t) = 0$. It follows by the Karhunen-Lòeve expansion (see, e.g. Ash and Gardner 1975) that

$$\widetilde{U}_j(t) = \sum_{l=1}^{\infty} \xi_{jl}\, \phi_{jl}(t), \tag{2}$$

where $\xi_{jl} = \int \widetilde{U}_j(t)\phi_{jl}(t)dt$ with $E(\xi_{jl}) = 0$ and $\mathrm{var}(\xi_{jl}) = \lambda_{jl}$. In practice, $\widetilde{U}_j(t)$ is approximated by a finite number of components, especially when noisy recordings are made at discrete times, leading to the approximated processes

$$\widetilde{U}_j^L(t) = \sum_{l=1}^{L_j} \xi_{jl}\, \phi_{jl}(t), \tag{3}$$

where the truncation points L_j need to be determined. We assume that by choosing L_j large enough the approximation error is negligible and note that representations in any other fixed basis system such as B-splines or trigonometric bases can alternatively be used to represent processes $\widetilde{U}_j$ and the linear manifolds in function space that we discuss below.

2.2 *Defining linear manifolds by constrained time-varying linear combinations*

To model the dynamic dependency between the M component functions, we define K constraints for the integrated variance of varying coefficient linear combinations of the components,

$$\int \mathrm{var}\left(\sum_{j=1}^{M} \beta_{kj}(t)\widetilde{U}_j(t)\right)dt = 0, \quad 1 \le k \le K. \tag{4}$$

Here the coefficient functions β_{kj} that define the time-varying linear combinations of the component processes for which integrated variances vanish are assumed to be smooth functions that are orthonormal as k varies, i.e.,

$$\sum_{j=1}^{M} \int \beta_{kj}(t)\beta_{lj}(t)dt = \delta_{kl}, \quad 1 \le k, l \le K, \tag{5}$$

where $\delta_{kl} = 1$ for $k = l$ and 0 otherwise. Noting that $E\{\sum_{j=1}^{M} \beta_{kj}(t)\widetilde{U}_j(t)\} = 0$ for all $t \in \mathcal{T}$, the constraints in (4) imply that $\sum_{j=1}^{M} \beta_{kj}(t)\widetilde{U}_j(t) = 0$ a.s. almost everywhere on $\mathcal{T}$.

Assume the coefficient functions $\beta_{kj}(t)$ corresponding to the jth component function of the kth constraint in (4) are expanded into basis functions $\{\psi_{jp}(t)\}_{p=1,2,\ldots}$ and that these coefficient functions are sufficiently well approximated when including the first P_j terms only, leading to the approximations

$$\beta_{kj}^{P}(t) = \sum_{p=1}^{P_j} b_{kjp}\psi_{jp}(t), \tag{6}$$

with suitable coefficients b_{kjp}. When approximating $\widetilde{U}_j(t)$ by $\widetilde{U}_j^L(t)$ and $\beta_{kj}(t)$ by $\beta_{kj}^P(t)$, the left-hand-side of the functional linear manifold constraints (4) becomes

$$\begin{aligned}
&\int \mathrm{var}\left(\sum_{j=1}^{M} \beta_{kj}^{P}(t)\widetilde{U}_j^L(t)\right)dt \\
&= \int \sum_{r=1}^{M}\sum_{s=1}^{M}\left\{\sum_{p=1}^{P_r} b_{krp}\psi_{rp}(t)\right\}\left\{\sum_{q=1}^{P_s} b_{ksq}\psi_{sq}(t)\right\}\left\{\sum_{l=1}^{L_s}\sum_{m=1}^{L_r} E(\xi_{rl}\xi_{sm})\phi_{rl}(t)\phi_{sm}(t)\right\}dt \\
&= \sum_{r=1}^{M}\sum_{s=1}^{M}\sum_{p=1}^{P_r}\sum_{q=1}^{P_s} b_{krp}b_{ksq}\left\{\sum_{l=1}^{L_r}\sum_{m=1}^{L_s} E(\xi_{rl}\xi_{sm})\int \psi_{rp}(t)\psi_{sq}(t)\phi_{rl}(t)\phi_{sm}(t)dt\right\}.
\end{aligned} \tag{7}$$

From now on we replace the original constraints in (4) by the analogous constraint for these approximations. Specifically, define P_j-vectors $\boldsymbol{b}_{kj} = (b_{kj1}, \ldots, b_{kjP_j})'$, $P_r \times P_s$ matrices $\boldsymbol{A}_{rs} = (a_{rspq})$ with elements

$$a_{rspq} = \sum_{l=1}^{L_r}\sum_{m=1}^{L_s} E(\xi_{rl}\xi_{sm})\int \psi_{rp}(t)\psi_{sq}(t)\phi_{rl}(t)\phi_{sm}(t)dt, \tag{8}$$

for $1 \le p \le P_r$ and $1 \le q \le P_s$, P^*-vectors $\boldsymbol{b}_k = (\boldsymbol{b}'_{k1}, \ldots, \boldsymbol{b}'_{kM})'$ and $P^* \times P^*$ constraint matrices $\boldsymbol{A} = (\boldsymbol{A}_{rs})$, $1 \le r, s \le M$, where $P^* = \sum_{j=1}^{M} P_j$. Then, in view of (7), the K constraints for the approximation to (4) that define the linear manifold model can be expressed in terms of the following quadratic form,

$$\begin{aligned} \int \mathrm{var}\left(\sum_{j=1}^{M} \beta_{kj}^{P}(t) \widetilde{U}_j^{L}(t) \right) dt &= \sum_{r,s=1}^{M} \boldsymbol{b}'_{kr} \boldsymbol{A}_{rs} \boldsymbol{b}_{ks} \\ &= \boldsymbol{b}'_k \boldsymbol{A} \boldsymbol{b}_k = 0 \quad \text{for} \quad 1 \le k \le K, \quad \text{subject to} \quad \boldsymbol{b}'_k \boldsymbol{b}_l = \delta_{kl}. \end{aligned} \tag{9}$$

Here the vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K$ are not identifiable, only the K-dimensional vector space that is spanned by these vectors and that defines the linear manifold (9). We may define $\mathcal{L}$ as orthogonal complement

$$\mathcal{L} = \mathrm{span}(\boldsymbol{b}_{K+1}, \ldots \boldsymbol{b}_{P^*})^{\perp}, \tag{10}$$

where eigenvectors $\boldsymbol{b}_{K+1}, \ldots, \boldsymbol{b}_{P^*}$ are uniquely defined if their corresponding eigenvalues are positive and unique. We note that a straightforward calculation shows that the normalization constraint in (9) guarantees the orthonormality of the β_{kj} in (5).

2.3 *Obtaining the functional linear manifold*

Observing that $a_{rspq} = a_{srqp}$, the constraint matrix $\boldsymbol{A}$ that defines the quadratic form in (9) is symmetric. From $\int \mathrm{var}(\sum_j \beta_{kj}(t) \widetilde{U}_j(t)) dt \ge 0$ it follows that $\boldsymbol{A}$ is also non-negative definite. Hence, its eigenvalues are real and non-negative. To find the constraints that define the linear manifold $\mathcal{L}$ in (9) for given K is then equivalent to determining the kernel of the linear map $\boldsymbol{A}$, which is the vector space $\{\boldsymbol{v} \in \mathcal{R}^{P^*} : \boldsymbol{A}\boldsymbol{v} = 0\}$. However, both K and $\boldsymbol{A}$ are unknown.

Finding consistent estimates $\hat{\boldsymbol{A}}$ for $\boldsymbol{A}$ is relatively straightforward. Various functional designs may be encountered in applications. In a first scenario, underlying processes X_{ji} are observed on the entire time domain and without error. Then estimates for the mean functions μ_j and auto-covariance functions G_j, $j = 1, \ldots, M$, are obtained simply by taking pointwise averages,

$$\hat{\mu}_j(t) = \frac{1}{n} \sum_{i=1}^{n} X_{ji}(t), \quad \widehat{G}_j(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} (X_{ji}(t_1) - \hat{\mu}_j(t_1))(X_{ji}(t_2) - \hat{\mu}_j(t_2)). \tag{11}$$

A second scenario is that the data are observed on a grid and may be contaminated with measurement noise. With $\{(Y_{ji}(s_p);\ j = 1, \ldots, M,\ i = 1, \ldots, n,\ p = 1, \ldots, P\}$ denoting observations of the random

trajectories corresponding to the ith random function of the jth functional component X_{ji}, recorded on a discrete grid of times s_p, the observed data then are $Y_{ji}(s_p) = X_{ji}(s_p) + \epsilon_{jip}$, where the measurement errors ϵ_{jip} are independently distributed with $E\epsilon_{jip} = 0$ and $\mathrm{var}(\epsilon_{jip}) = \sigma_j^2$, for each $1 \leq j \leq M$.

In this scenario of dense and noisy recordings, one may use the data $\{(s_p, Y_{ji}(s_p));\ i = 1, \ldots, n,\ p = 1, \ldots, P\}$ to implement estimators (11) first on the grids $\{s_p\}$ for μ_j and $\{s_p, s_q\},\ p, q = 1, \ldots, P,$ and then interpolating these estimators linearly to obtain estimates on the continuum. An alternative estimation scheme that also works for sparser and less regular data is the PACE procedure described in Yao *et al.* (2005a). The estimates of the eigenvalue and eigenfunction pairs $\{\lambda_{jl}, \phi_{jl}(t)\}$ in (1) correspond to the solutions $\{\hat{\lambda}_{jl}, \hat{\phi}_{jl}(t)\}$ of the eigenequations such that

$$\int_{\mathcal{T}} \widehat{G}_j(t_1, t_2)\, \hat{\phi}_{jl}(t_1)\, dt_1 = \hat{\lambda}_{jl}\, \hat{\phi}_{jl}(t_2), \tag{12}$$

where $\hat{\lambda}_{jl} \geq 0$, subject to $\int_{\mathcal{T}} \hat{\phi}_{jl}(t)\hat{\phi}_{jk}(t)dt = \delta_{lk}$. Estimation of the random coefficients ξ_{jl} in (2) are obtained by

$$\hat{\xi}_{jli} = \int (X_{ji}(t) - \hat{\mu}_j(t))\hat{\phi}_{jl}(t)dt, \tag{13}$$

or alternatively by the PACE method, which leads to the same estimators as the above in the case of fully observed trajectories (Müller 2005).

We note that

$$E(\xi_{rl}\xi_{sm}) = \int_{\mathcal{T}} \int_{\mathcal{T}} \phi_{rl}(t) G_{rs}(t, u) \phi_{sm}(u) du dt, \tag{14}$$

where $G_{rs}(t, u) = \mathrm{cov}(U_r(t), U_s(u))$ is the cross-covariance function of U_r and U_s. Estimates for the elements a_{rspq} of $\boldsymbol{A}$ are obtained by plugging in the empirical cross-covariance estimates of $G_{rs}(t, u)$ and the estimated eigenfunctions $\hat{\phi}_{rl}$ and $\hat{\phi}_{sm}$ obtained in (12). For the following, the orthonormal basis functions ψ_{rp}, into which the varying coefficient functions β_{kr} are expanded according to (6), are chosen by the user and therefore may be considered known. A user might for example select an orthonormalized version of the B-spline basis. Another straightforward choice, which we adopt in the implementation, is to use the eigenbasis, i.e., to set $\psi_{rp} = \hat{\phi}_{rp}$.

We then obtain estimates $\hat{\boldsymbol{A}} = (\hat{\boldsymbol{A}}_{rs})$, where the estimated matrix $(\hat{\boldsymbol{A}}_{rs}) = (\hat{a}_{rspq})$ is defined through its estimated elements

$$\hat{a}_{rspq} = \sum_{l=1}^{L_r} \sum_{m=1}^{L_s} \widehat{E}(\xi_{rl}\xi_{sm}) \int \hat{\psi}_{rp}(t)\hat{\psi}_{sq}(t)\hat{\phi}_{rl}(t)\hat{\phi}_{sm}(t)dt. \tag{15}$$

Here, relation (14) suggests estimates

$$\widehat{E}(\xi_{rl}\xi_{sm}) = \int\int \hat{\phi}_{rl}(t)\widehat{G}_{rs}(t,u)\hat{\phi}_{sm}(u)dudt \tag{16}$$

for $E(\xi_{rl}\xi_{sm})$, and $\widehat{G}_{rs}(t,u)$ is the estimated cross-covariance of $U_r(t)$ and $U_s(u)$, which is obtained in analogy to (11) or in the sparse sampling scenario as in Yao *et al.* (2005b).

According to (10), the functional linear manifold can be determined from estimates $(\hat{\boldsymbol{b}}_{K+1},\ldots,\hat{\boldsymbol{b}}_{P^*})$ of eigenvectors $(\boldsymbol{b}_{K+1},\ldots,\boldsymbol{b}_{P^*})$ of $\boldsymbol{A}$, yielding the estimate

$$\hat{\mathcal{L}} = \text{span}(\hat{\boldsymbol{b}}_{K+1},\ldots,\hat{\boldsymbol{b}}_{P^*})^{\perp} \quad \text{of the linear manifold} \quad \mathcal{L} = \text{span}(\boldsymbol{b}_{K+1},\ldots,\boldsymbol{b}_{P^*})^{\perp}. \tag{17}$$

In practice, as the matrix $\hat{\boldsymbol{A}}$ almost surely will not have eigenvalues of multiplicity more than 1, the eigenvectors $(\hat{\boldsymbol{b}}_1,\ldots,\hat{\boldsymbol{b}}_K)$ belonging to the K smallest eigenvalues of $\hat{\boldsymbol{A}}$ can be uniquely determined.

One may therefore alternatively use the estimates $\hat{\mathcal{L}} = \text{span}(\hat{\boldsymbol{b}}_1,\ldots,\hat{\boldsymbol{b}}_K)$. They are computed by the solving the eigensystem

$$\hat{\boldsymbol{A}}\hat{\boldsymbol{b}}_k = \hat{\gamma}_k\hat{\boldsymbol{b}}_k, \tag{18}$$

subject to $\hat{\boldsymbol{b}}_k'\hat{\boldsymbol{b}}_k = 1$, with the $\hat{\gamma}_k$ corresponding to the eigenvalues of $\hat{\boldsymbol{A}}$, ordered from smallest to largest, which are estimates of the eigenvalues γ_k defined by $\boldsymbol{A}\boldsymbol{b}_k = \gamma_k\boldsymbol{b}_k$. This is the implementation we adopt, expressing the manifold constraints conveniently in terms of the smallest K eigenvectors of $\hat{\boldsymbol{A}}$. This leads to the estimated constraint functions

$$\hat{\beta}_{kj}^{P}(t) = \sum_{p=1}^{P^*} \hat{b}_{kjp}\psi_{jp}(t), \tag{19}$$

where the vector components b_{kjp} are as defined after (8).

In practical implementations, it is convenient to select the basis functions ψ_{jp} that are used for the expansion of the coefficient functions β_{kj} as in (6) to be equal to the basis functions ϕ_{jp} that correspond to the eigenexpansions of the corresponding component $\widetilde{U}_j$ as in (3), setting $\psi_{jp} = \phi_{jp}$. Given eigenvalue-eigenvector pairs $(\hat{\gamma}_k, \hat{\boldsymbol{b}}_k)$ of $\hat{\boldsymbol{A}}$, $k = 1,\ldots,P^*$, arranged in nondecreasing order in terms of the eigenvalues $\hat{\gamma}_k$, one may base the selection of the number L_j of included components on the estimated proportion of total variance that is explained by the leading components in a corresponding functional principal component analysis,

$$L_j = \min\left\{L : \sum_{l=1}^{L}\hat{\lambda}_{jl}\Big/\sum_{l=1}^{\infty}\hat{\lambda}_{jl}I_{\{\hat{\lambda}_{jl}\geq 0\}} > \vartheta\right\}, \quad j = 1,\ldots,M, \tag{20}$$

where $0 < \vartheta < 1$ and ϑ is a preselected threshold value. We adopted the following approach, which is easy to implement and led to good practical results: Set $\vartheta = 0.9$ and use criterion (20) to select L_j and then choose $P_j = L_j$.

Another important step is to determine the number of constraints K that define the functional linear manifold. We proceed as follows. Observing from (9) that $\int \mathrm{var}\left(\sum_{j=1}^{M} \beta_{kj}^{P}(t)\widetilde{U}_j^{L}(t)\right) dt = \gamma_k$, and given a set of values $\{P_j = L_j\}_{j=1,\ldots,M}$, the number of constraints is chosen by

$$K^* = \max_{1 \le K \le P^*} \left\{ K : \sum_{k=1}^{K} \hat{\gamma}_k < \tau \sum_{k=1}^{P^*} \hat{\gamma}_k \right\}, \tag{21}$$

for a small threshold $\tau > 0$.

The value of τ determines the fraction of the (estimated) total variance that the integrated variance for the estimated linear manifold may assume. This fraction cannot be expected to exactly reach its target value of zero, due to deviations between estimates and true quantities. So while the correct value of τ depends on the unknown size of the deviations of the true eigenvalues from their estimates, it should be small. The smaller one chooses the fraction τ, the smaller is the dimension of the resulting manifold. In our data analysis a value of $\tau = 0.05$ provided a reasonable choice, but this is not a universal recommendation. The trade-off is between dimension reduction through the manifold representations described in section 2.4 and the corresponding prediction error for the component processes, which can be determined by k-fold cross-validation, for example.

2.4 *Representing the multivariate component functions on the functional linear manifold*

Recall that with eigenfunctions ϕ_{jl} and inclusion of L_j eigencomponents, the jth component function $\widetilde{U}_j(t)$ is assumed to be sufficiently well represented by $\widetilde{U}_j^{L}(t)$ as in (3),

$$\widetilde{U}_j(t) \approx \widetilde{U}_j^{L}(t) = \sum_{l=1}^{L_j} \xi_{jl}\phi_{jl}(t), \quad j = 1, \ldots, M,$$

with functional principal components ξ_{jl}. Once one has identified the functional linear manifold with K constraints on (or rather near) which the component functions are situated, given the constraints and equating them to zero, one only needs to represent the "remaining features" of the M component processes to obtain a complete (approximate) representation of the multivariate process. Thus, if the linear

manifold model holds, this constrained representation will be more parsimonious than unconstrained componentwise representations.

Formally, setting $\psi_{jp} = \phi_{jp}$ and assuming $L_j = P_j$, as described in Section 2.3,

$$\int \sum_{j=1}^{M} \beta_{kj}^{L}(t)\widetilde{U}_{j}^{L}(t)dt = \int \sum_{j=1}^{M} \left\{ \sum_{p=1}^{L_j} b_{kjp}\phi_{jp}(t) \right\} \left\{ \sum_{l=1}^{L_j} \xi_{jl}\phi_{jl}(t) \right\} dt = \sum_{j=1}^{M} \sum_{l=1}^{L_j} b_{kjl}\,\xi_{jl}, \tag{22}$$

where the b_{kjl} are the coefficients for the expansions of the varying coefficient functions β_{kj}, due to the orthonormality of the basis functions. In order to find the representation of the component processes constrained to be on the manifold, the components for $k = 1, \ldots, K$ in (22) are set to equal zero. This leads to the desired representation, which is given in equation (28), with estimated versions in equation (29) below.

More specifically, we concatenate the L_j-vectors $\boldsymbol{\xi}_{ji} = (\xi_{j1i}, \ldots, \xi_{jL_ji})'_{L_j\times 1}$, where $\xi_{jli} = \int (X_{ji}(t) - \mu_j(t))\phi_{jl}(t)dt$, that consist of the L_j leading principal component scores of $X_{ji}(t)$, the jth component for the ith subject, to L_M^*-vectors $\boldsymbol{\xi}_i = (\boldsymbol{\xi}'_{1i}, \ldots, \boldsymbol{\xi}'_{Mi})'_{L_M^*\times 1}$, where $L_M^* = \sum_{j=1}^{M} L_j$. Then by (22),

$$\zeta_{ki} = \int \sum_{j=1}^{M} \beta_{kj}^{L}(t)X_{ji}^{L}(t)dt = \sum_{j=1}^{M} \boldsymbol{b}'_{kj}\,\boldsymbol{\xi}_{ji} = \boldsymbol{b}'_{k}\,\boldsymbol{\xi}_{i}, \tag{23}$$

where $X_{ji}^{L}(t) = \mu_j(t) + \sum_{l=1}^{L_j} \xi_{jli}\phi_{jl}(t)$. Setting $\Upsilon_i = (\zeta_{1i}, \ldots, \zeta_{L_M^*i})'_{L_M^*\times 1}$ and $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{L_M^*})_{L_M^*\times L_M^*}$, one finds $\Upsilon_i = \boldsymbol{B}'\boldsymbol{\xi}_i$ and as $\boldsymbol{B}$ is an orthogonal matrix, furthermore

$$\boldsymbol{\xi}_i = \boldsymbol{B}\,\Upsilon_i = \sum_{k=1}^{L_M^*} \boldsymbol{b}_k\,\zeta_{ki}. \tag{24}$$

For $\boldsymbol{\Upsilon} = (\Upsilon_1, \ldots, \Upsilon_n)_{L_M^*\times n}$ and $\boldsymbol{\Xi} = (\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^n)_{L_M^*\times n}$, one has

$$\boldsymbol{\Upsilon} = \boldsymbol{B}'\,\boldsymbol{\Xi}, \quad \boldsymbol{\Xi} = \boldsymbol{B}\,\boldsymbol{\Upsilon}. \tag{25}$$

To construct the constraint-reduced component functions, let $\Upsilon_i^K = (\zeta_{(K+1)i}, \ldots, \zeta_{L_M^*i})'_{(L_M^*-K)\times 1}$, and $\boldsymbol{B}^K = (\boldsymbol{b}_{K+1}, \ldots, \boldsymbol{b}_M)_{L_M^*\times L_{M-K}^*}$. The K constraints defining the linear manifold model (4) imply that we may set $\{\zeta_{ki}\} = 0$ for $k \le K$. Observing (24), obtain constrained principal scores

$$\boldsymbol{\xi}_i^K = \boldsymbol{B}^K\,\Upsilon_i^K = \sum_{k=K+1}^{L_M^*} \boldsymbol{b}_k\zeta_{ki}. \tag{26}$$

For $\boldsymbol{\Upsilon}^K = (\Upsilon_1^K, \ldots, \Upsilon_n^K)_{L^*_{M-K} \times n}$ and $\boldsymbol{\Xi}^K = (\boldsymbol{\xi}_1^K, \ldots, \boldsymbol{\xi}_n^K)_{L^*_{M-K} \times n}$, it follows from (26) that

$$\boldsymbol{\Xi}^K = \boldsymbol{B}^K \boldsymbol{\Upsilon}^K. \tag{27}$$

Finally, the reduced versions $\boldsymbol{\xi}_i^K$ in (26) and $\boldsymbol{\Xi}^K$ in (27) are used to construct the constrained reduced functional components and representations,

$$X_{ji}^K(t) = \mu_j(t) + \sum_{l=1}^{L_j} \xi_{jli}^K \phi_{jl}(t). \tag{28}$$

Plugging estimates $\hat{\phi}_{jl}(t)$, $\hat{\xi}_{jli}^K$, $\hat{\mu}_j(t)$ and $\widehat{G}_j(t,t)$ into (28), one obtains estimated representations

$$\widehat{X}_{ji}^K(t) = \hat{\mu}_j(t) + \sum_{l=1}^{L_j} \hat{\xi}_{jli}^K \hat{\phi}_{jl}(t). \tag{29}$$

3 Consistency properties

We assume that component processes are fully observed and estimates (11) are used, furthermore that all processes are continuous on their finite domain, possess uniformly bounded moments up to the fourth moment and that all basis functions are continuous.

Using arguments based on moments, it is straightforward to see that

$$\int \{\hat{\mu}_r(t) - \mu_r(t)\}^2 \, dt = O_p(n^{-1}), \quad 1 \le r \le M, \tag{30}$$

$$\begin{aligned} &\int\int \left\{\widehat{G}_{rs}(t,u) - G_{rs}(t,u)\right\}^2 \, dtdu \\ = &\int\int \left\{\frac{1}{n}\sum_{i=1}^{n} \{X_{ri}(t) - \hat{\mu}_r(t)\}\{X_{si}(u) - \hat{\mu}_s(t)\} - G_{rs}(t,u)\right\}^2 dtdu = O_p(n^{-1}),\ 1 \le r,s \le M. \end{aligned} \tag{31}$$

The following notation will be used. Set $L = \max\{L_r, P_r;\ 1 \le r \le M\}$ and for the eigenvalues λ_{jl}, $1 \le j \le M$, $1 \le l \le L$, for the jth component process X_j, which are assumed to be decreasingly ordered for each j, define $\rho_{j1} = \lambda_{j1} - \lambda_{j2}$, $\rho_{jk} = \min_{l \le k}(\lambda_{j(l-1)} - \lambda_{jl}, \lambda_{jl} - \lambda_{j(l+1)})$, $k \ge 2$, and $\Delta_k = \min_{1 \le j \le M}\{\rho_{jk}\}$. Note that the Δ_k correspond to lower bounds of the minimum spacings between adjacent eigenvalues of the component processes. In the following, $\|\cdot\|$ stands for the Euclidean norm for vectors or matrices and estimates $\hat{\boldsymbol{A}}$ of $\boldsymbol{A}$ refer to estimates (15), $\hat{\boldsymbol{b}}_k$ of $\boldsymbol{b}_k$ to estimates (18) of the eigenvectors of $\boldsymbol{A}$ and $\hat{\gamma}_k$ to estimates of the corresponding eigenvalues γ_k.

Theorem. If $\Delta_k \geq \Delta_0$ for some small value $\Delta_0 > 0$ and all $1 \leq k \leq L$, the following holds:

$$(\hat{\lambda}_{rl} - \lambda_{rl})^2 = O_p(n^{-1}), \quad \int \left\{\hat{\phi}_{rl}(t) - \phi_{rl}(t)\right\}^2 dt = O_p(n^{-1}), \tag{32}$$

$$\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|^2 = O_p(n^{-1}). \tag{33}$$

Results analogous to (32) have been established before by Bosq (2000) and others under various conditions. Result (33) establishes the consistency of the components of the multivariate processes and of the estimated matrix $\hat{\boldsymbol{A}}$. The main application is the following.

Corollary. Under the assumptions of the theorem, if the first K eigenvalues of $\boldsymbol{A}$ vanish and all other eigenvalues are non-zero with eigenspaces of dimension 1, then

$$|\hat{\gamma}_k - \gamma_k|^2 = O_p(n^{-1}), \quad \|\hat{\boldsymbol{b}}_k - \boldsymbol{b}_k\|^2 = O_p(n^{-1}), \quad k = K+1, \ldots, P^*, \tag{34}$$

and the linear manifold estimate $\hat{\mathcal{L}}$ converges to the linear manifold $\mathcal{L}$ as defined in (17) in the sense that

$$\max\left\{\sup_{u\in\hat{\mathcal{L}},\,\|u\|=1} \inf_{v\in\mathcal{L},\,\|v\|=1} \|u-v\|^2, \sup_{u\in\mathcal{L},\,\|u\|=1} \inf_{v\in\hat{\mathcal{L}},\,\|v\|=1} \|u-v\|^2\right\} = O_p(n^{-1}). \tag{35}$$

This result shows that the Hausdorff set distance between the two unit balls in $\mathcal{L}$ and $\hat{\mathcal{L}}$ converges to 0 at rate n^{-1} and that therefore the distances between corresponding points of the two spaces $\mathcal{L}$ and $\hat{\mathcal{L}}$ vanish asymptotically, which means that the two subspaces $\mathcal{L}$ and $\hat{\mathcal{L}}$ of $\mathcal{R}^{P^*}$ become asymptotically indistinguishable. More specifically, denoting the orthogonal projection on $\mathcal{L}$ by $P_{\mathcal{L}}$ and analogously the orthogonal projection on $\hat{\mathcal{L}}$ by $P_{\hat{\mathcal{L}}}$, result (34) implies that

$$\sup_{u\in\hat{\mathcal{L}}} \|u - P_{\mathcal{L}}(u)\|^2 \leq \|u\|^2 O_p(n^{-1}), \quad \sup_{u\in\mathcal{L}} \|u - P_{\hat{\mathcal{L}}}(u)\|^2 \leq \|u\|^2 O_p(n^{-1}),$$

implying that the distances between any elements of the estimated (true) manifold and their orthogonal projections onto the true (estimated) manifold converge pointwise to 0 at rate n^{-1}, with uniform convergence over any compact set.

The main steps of the proofs can be found in Appendix A. While we have assumed here that processes X_i are fully observed, these results can be extended to the case where processes are observed on a dense regular grid of measurement times $s_1, \ldots, s_p$. If the number of equidistant sampling times

$p = p(n)$ satisfies $p \to \infty$ as $n \to \infty$ and $\liminf p(n)/n > 0$, it is not difficult to see that forming cross-sectional averages at grid points s_p, subsequently interpolated between the grid points, gives rise to estimates $\hat{\mu}_r$ that satisfy (30), using the approximation of integrals by Riemann sums. Analogously, empirical covariances can be defined for the measurements taken at any two grid points t_p, t_q, and these covariances can be interpolated to form a continuous covariance surface that satisfies (31). This guarantees the extension of the results to densely sampled situations. In addition, the results also generalize to the case a dense gridded design where measurements are contaminated by i.i.d. errors with zero mean and finite variance, as described after (11). This is a straightforward extension of the arguments provided in the Appendix, obtained by calculating expectations.

In the above results we have ignored the approximation error that is incurred by truncating the processes at a finite number of included components. When the approximation error of the component processes is additionally considered, the consistency results in the theorem and the corollary hold with different rates of convergence depending on the assumptions one invokes on the convergence behavior of higher order eigenvalues when $p \to \infty$ as $n \to \infty$. Applicable assumptions can for example be found in the online supplement of Müller and Yao (2008), and the extension to more general rates can be obtained in analogy to the derivations given there.

Finally, we remark that the above results are generally applicable to the problem of identifying a manifold that is defined by the eigenspace of $\boldsymbol{A}$ that belongs to the K smallest eigenvalues, as long as the $(K + 1)$-st eigenvalue is larger than any preceding eigenvalue.

4 Application to multivariate longitudinal behavior data

The study of behavioral changes associated with aging has found considerable interest lately, for example to gauge the effects of increasing longevity on patterns of old-age disability (Manton et al. 1997; Crimmins 2004; Christensen et al. 2009). Comprehensive studies of behavioral patterns over entire lifespans are based on cohort studies of animals and have been frequently conducted with short-lived species such as flies, usually restricted to the study of just one behavior, e.g. egg-laying, for which one-dimensional functional data analysis methods have been used (Papadopoulos et al. 2002; Chiou, Müller and Wang 2004; Papadopoulos et al. 2004). Multiple behaviors at densely spaced observation

times over the entire lifespan of organisms were recorded in Carey et al. (2006), where cohorts of flies were raised in the laboratory under controlled conditions and daily observations of the frequency of behaviors within 10-minute intervals during a fixed two hour daily window were made. We analyze here the data for Drosophila melanogaster regarding the four behaviors *Flying*, *Feeding*, *Walking* and *Resting*, which were found to vary with age in Carey et al. (2006), and refer to this paper for further details about the experimental protocol of this study.

Since the impact of behavior on old-age survival and longevity was of interest and because of the need to avoid biases caused by drop-outs, we restricted our analysis to flies that were observed beyond 40 days of age, yielding a sample size of $n = 62$ flies without drop-outs on the time domain from 0 to 40 days of age. To reduce the range of the log count data and enable taking logarithms, the data were preprocessed by adding 0.5 to the observed daily frequencies of each observed behavior. Taking logarithms was followed by a pre-smoothing step for which we used local linear smoothers with a pre-selected relatively small bandwidth that was chosen to be the same for all trajectories.

In the following, $U_1, \ldots, U_4$ denote trajectories of the log intensities of the behaviors in dependence of age, where U_1 stands for *Flying*, U_2 for *Feeding*, U_3 for *Walking* and U_4 for *Resting*. While Carey et al. (2006) were mainly interested in determining the relationship between the intensity of behaviors and subsequent survival, our goal here is to uncover dynamic relationships, allowing for changes in the nature of the multivariate relationship between the various behaviors as flies age. Dynamic relationships between various behaviors are visually hard to assess, due to the multivariate nature of the functional data. Therefore a formal method for deriving underlying relationships will be useful for situations where there is little or no prior information available about time-dynamic multivariate relationships.

Examples of smoothed behavioral trajectories for *Flying* (with trajectories U_1), *Feeding* (trajectories U_2), *Walking* (trajectories U_3) and *Resting* (trajectories U_4) are visualized in Figure 1 in Section 1. Figure 2 displays the estimated mean functions $E(U_j(t))$ and variance functions $\mathrm{var}(U_j(t))$. Strong age effects are clearly visible in these mean curves, with flying and walking declining at very old ages while feeding and resting increase. The trends of the mean functions for *Flying*, *Feeding* and *Walking* are similar, while the mean function of *Resting* points in an opposite direction, indicating that very young flies tend to rest frequently and resting behavior climbs to an elevated plateau as flies are aging.

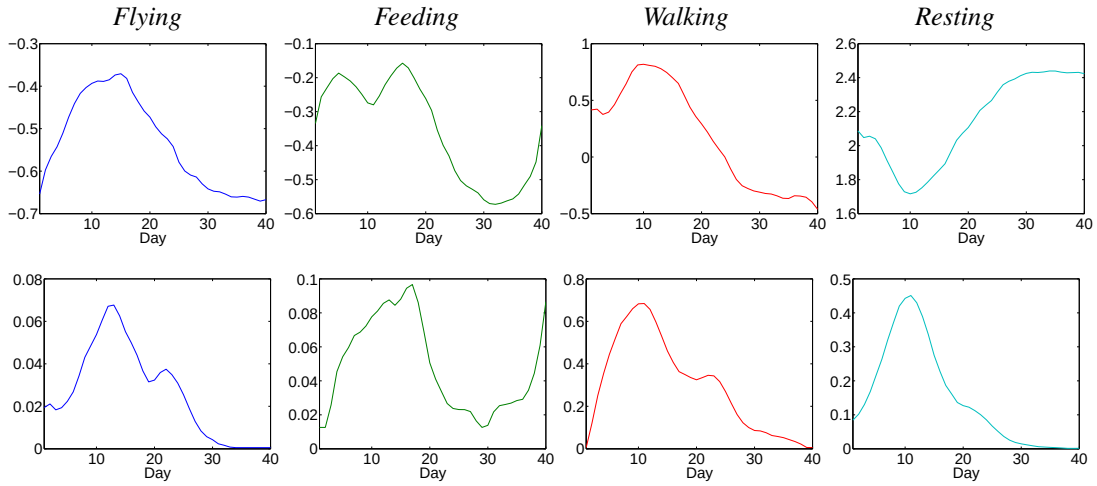

Figure 2: Mean function estimates $\hat{\mu}_j(t)$ (upper panels) and variance function estimates $\hat{G}_j(t,t)$ (lower panels), $j = 1, ..., 4$, for U_1 (*Flying*), U_2 (*Feeding*), U_3 (*Walking*), and U_4 (*Resting*).

The variance functions for *Flying*, *Walking* and *Resting* generally follow the pattern of the mean functions and are proportional to the mean functions for *Flying*, *Feeding* and *Walking*, whereas the mean and variance functions for *Resting* are inversely proportional to each other. This implies that resting becomes a dominant behavior at older ages, while flying and walking show a strong decline across the population.

The associated leading eigenfunctions of the auto-covariances are illustrated in Figure 3. Three eigenfunctions are selected to explain at least 90% of the total variability for all four behavioral variables. Specifically, the numbers of included components selected by criterion (20) with $\vartheta = 0.90$ are $\{L_j\} = \{P_j\} = \{2, 3, 2, 2\}$ and $P^* = 9$. Interestingly, the patterns of the eigenfunctions for the four component processes are quite similar, reflecting similarity of their auto-covariance structures.

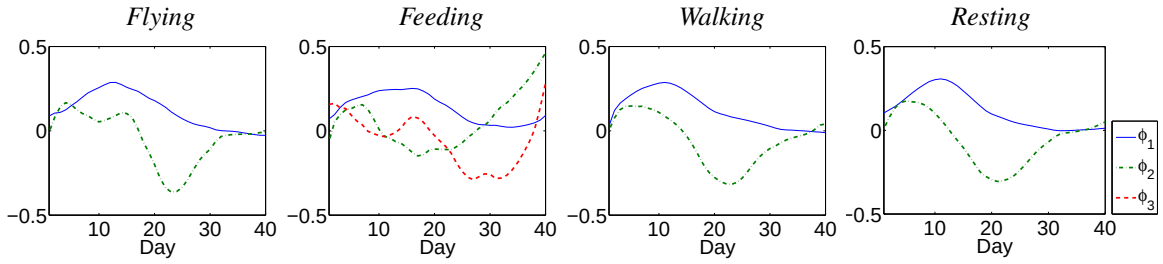

Figure 3: Estimated leading eigenfunctions for U_1 (*Flying*), U_2 (*Feeding*), U_3 (*Walking*), and U_4 (*Resting*).

Applying the selection criterion (21) to the estimated eigenvalues of the matrix $\boldsymbol{A}$ in the quadratic form (9) that defines the linear manifold leads to the choice $K^* = 3$ for $\tau = 0.05$. The corresponding estimates for the three sets of time-varying coefficient functions $\beta_{kj}(t)$, $j = 1, 2, 3, 4$ in (6) for $K^* = 3$ time-varying linear constraints are displayed in Figure 4, for $k = 1, 2, 3$ in left, middle and right panels, respectively. The sets of coefficient functions $\beta_{1j}(t)$, $\beta_{2j}(t)$ and $\beta_{3j}(t)$ that correspond to each of the three smallest eigenvalues and associated eigenvectors of $\boldsymbol{A}$, respectively, are associated with residual variances of size 0.83%, 0.88% and 1.67%, which together account for a total of 3.38% of the total variance, which falls well below the 0.05 cut-off criterion.

Confidence bands for the coefficient functions as shown in Figure 4 were computed by bootstrap resampling, with the resulting pointwise 95% confidence intervals shown in Figure 5. These confidence intervals are based on 200 bootstrap samples that were obtained by resampling the observed trajectories with replacement. The width of these intervals varies considerably between ages and coefficients, indicating that some caution is needed in interpreting the coefficient functions across ages.

The constraints visualized in Figure 4 pertain to log-trajectories $X_{ji}(t)$, for $j = 1, 2, 3, 4$ and refer to the deviations from the component mean functions after components have been transformed to the log scale. We note that the vectors $(\beta_{k1}(t), \ldots, \beta_{k4}(t))'$ can be expected to be close to normal vectors of the linear manifold in function space at most times t; these vectors depend on t. Similar values of $\beta_{kj_1}(t)$ and $\beta_{kj_2}(t)$ for two components j_1, j_2 indicate that the constraint describes a dynamic effect such that $U_{m_1}(t)$ tends to decrease as $U_{m_2}(t)$ increases, and vice versa, so that $U_{j_1}(t)$ and $U_{j_2}(t)$ are negatively associated; they are positively associated if $\beta_{km_1}(t)$ and $\beta_{km_2}(t)$ are of opposite sign. This simple observation aids in the interpretation of the coefficient functions.

The first constraint (left plot in Figure 4) corresponds to a contrast between *Flying* and *Feeding* and thus reflects synergistic effects between these two behaviors, especially at older ages, while *Resting* and *Walking* are less constrained. The constraint gains force after 20 days of age and thus primarily reflects a behavioral characteristic of older flies, where flies that feed more also tend to fly more, reflecting higher nutritional demands for "frequent fliers".

The second constraint, associated with an only slightly smaller eigenvalue, also comes into play mostly for older ages and also mainly involves *Flying* and *Feeding*, but in a somewhat more antagonistic

fashion. In advanced mid-age, around day 25, *Flying* is subject to stronger constraints and is positively associated with *Walking* and negatively with *Feeding*, so that moving around takes time away from *Feeding*, which in turn constrains *Feeding* at the oldest ages more than *Flying*. This constraint thus reflects a competition between moving around and *Feeding*.

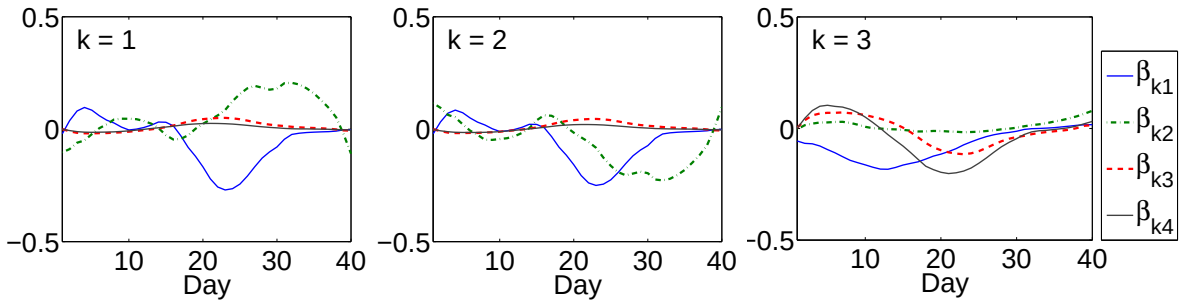

Figure 4: Estimated time-varying coefficient functions $\beta_{kj}(t)$ for $k = 1, 2, 3$ and $j = 1, 2, 3, 4$, i.e. the behavioral components *Flying*, *Feeding*, *Walking* and *Resting*, that define the three constraints that determine the linear manifold for log intensities of behaviors in dependency on age for Drosophila. Here k is the index of the constraint.

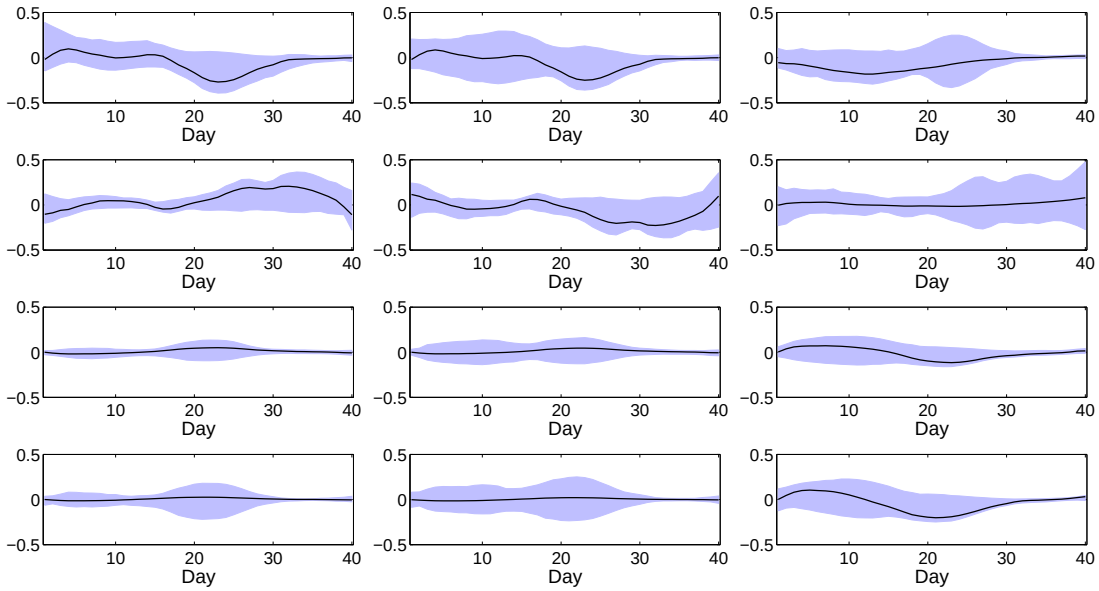

Figure 5: 95% confidence intervals for varying coefficient functions $\beta_{kj}(t)$ for $k = 1, 2, 3$, where k is the index of the constraint (from left to right) and $j = 1, 2, 3, 4$, i.e. *Flying*, *Feeding*, *Walking* and *Resting* (top-down).

The third constraint is dominated by *Resting*, which is competing with *Walking* and to a lesser

extent with *Feeding* for young flies until around day 15, but forms a contrast and thus is positively associated with *Flying* during this time. At mid-age and older ages, *Resting* is competing with the movement activities of *Flying* and *Walking*. This may indicate that *Resting* is energizing flies at young ages, which enables them to fly more, but at older ages it becomes a sign of general age decline, so that flies that are aging faster fly less and rest more, while they also feed more, as *Feeding* and *Resting* form a contrast at old age. From this analysis, *Flying*, *Feeding* and *Resting* emerge as signature behaviors for assessing the aging process.

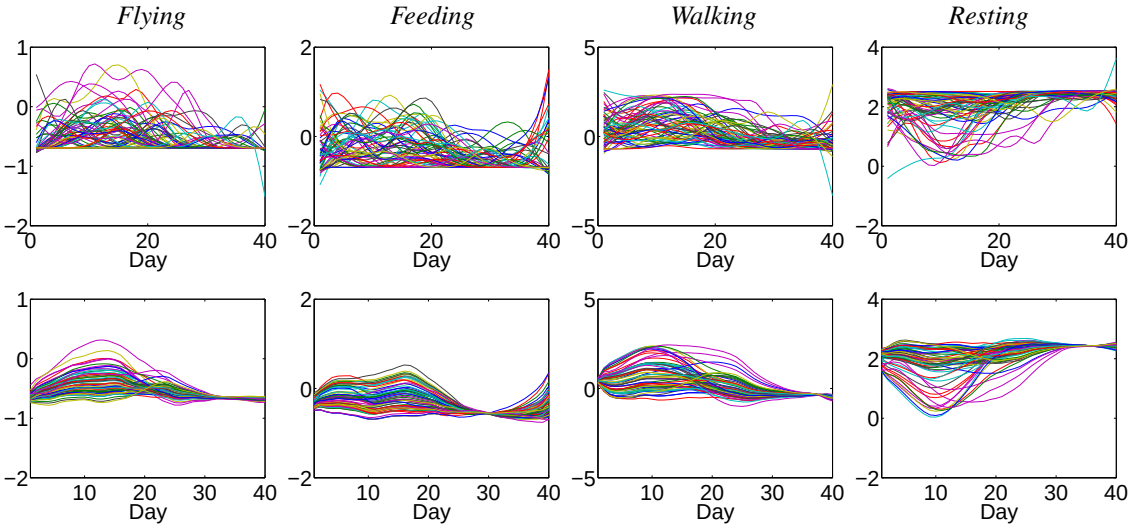

Figure 6: Smoothed X_{ji} via locally weighted least squares method (upper panels) and the corresponding fitted $\hat{X}_{ji}^K$ (29) (lower panels).

The constraints implied by the functional linear manifold model can be utilized to obtain representations of the original multivariate component functions X_{ji} in the form $\widehat{X}_{ji}^K(t)$, which are estimated versions of (28). The resulting smooth data fits $\widehat{X}_{ji}^K(t)$ are displayed in Figure 6 (lower panels), in comparison with the smoothed X_{ji} (upper panel), using locally weighted least squares. This indicates that this low-dimensional representation provides a satisfactory fit for the data.

5 Simulations

To explore how the functional linear manifold model works for finite sample situations, we simulated data from a model that generally although not in all details mimics the fitted model for the multivariate

behavioral data. Here, the estimated linear manifold model for the Drosophila behavioral data served as true model for generating samples of multivariate process trajectories. For each generated sample, we applied the proposed methods to determine the plausibility of a linear manifold model, the number of constraints, and the performance of the estimated coefficient functions that represent each constraint.

To obtain random process trajectories $X_{ji}(t)$, we generated functional principal component scores $\{\xi_{jl,i}\}$ as in (2) and (3) with the following properties: $(\xi_{1l,i}, \xi_{2l,i}, \xi_{3l,i}, \xi_{4l,i})$ has a multivariate normal distribution with mean zero and covariance given by $E(\xi_{rl}\xi_{sm})$, $1 \le r, s \le 4$ and $1 \le l, m \le 3$, where

$$E(\xi_{rl}\xi_{sm}) = \begin{cases} \lambda_{rl}\,, & \text{if } r = s \text{ and } l = m \\ 0\,, & \text{if } r = s \text{ and } l \neq m \\ \sigma_{rlsm}\,, & \text{if } r \neq s. \end{cases}$$

Here λ_{rl} is the eigenvalue that corresponds to the estimate obtained in the data analysis and $\sigma_{rlsm} = \hat{E}(\xi_{rl}\xi_{sm})$ is chosen as $n^{-1}\sum_{i=1}^{n}\hat{\xi}_{rl,i}\hat{\xi}_{sm,i}$, where $\{\hat{\xi}_{rl,i}, i = 1, \ldots, n\}$ are the empirical estimates obtained in the data analysis. As the matrix obtained through these specifications is not necessarily positive definite, as required for a non-degenerate covariance matrix, we additionally applied a spectral decomposition, truncating the expansion to include only components with positive eigenvalues, then reconstructing the covariance from the retained components.

Using the estimated eigenfunctions $\hat{\phi}_{jl}(t)$ of the auto-covariance kernels for the processes U_j as shown in Figure 3, along with the generated functional principal component scores $\{\xi_{jl,i}\}$ as described above, the process trajectories $X_{ji}(t) = \hat{\mu}_j(t) + \sum_{l=1}^{L_j}\xi_{jl,i}\hat{\phi}_{jl}(t)$ are generated by choosing L_j as the number of strictly positive $\{\hat{\lambda}_{jl}, l \ge 1\}$ that were obtained in the application in Section 4, for each $j = 1, \ldots, 4$. In the simulations, the sample size is assumed to be $n = 62$, and the number of equally spaced time grid points is chosen to be 40 on $[1, 40]$, mimicking the situation for the Drosophila behavioral data. The simulation results are reported for 500 Monte Carlo runs.

We consider two scenarios in the simulation for selecting the number of functional principal components L_j in (3). In scenario 1 we use a fixed number of components $(2, 3, 2, 2)$ for L_j, $j = 1, \ldots, 4$, while in scenario 2 we select this number by the criterion of fraction of total variance explained as in (20) with $\vartheta = 0.9$. For both scenarios, using a pre-determined threshold value $\tau = 0.05$, the true underlying number of constraints K in (4) is 3. Table 1 provides summary results for the number of

Table 1: Number of constraints K chosen in 500 simulation replicates. (to be updated)

Scenario	$K^* = 2$	$K^* = 3$	$K^* = 4$
1	6 (1.2%)	493 (98.6%)	1 (0.2%)
2	12 (2.4%)	487 (97.4%)	1 (0.2%)

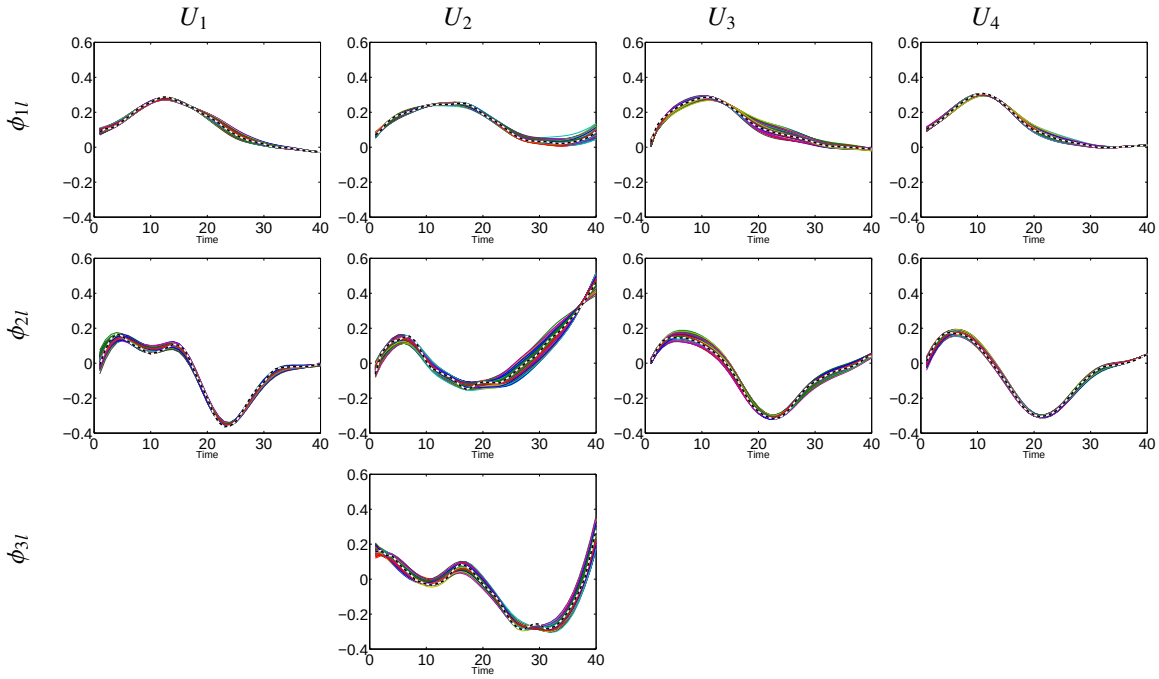

Figure 7: Estimated eigenfunctions ϕ_{jl} for the four component processes $U_j(t)$, $j = 1, \ldots, 4$ (from left to right) for the first three components, $l = 1, 2, 3$ (top-down), superimposed on the true eigenfunctions (white dotted lines), for 500 simulation runs.

constraints that are selected by the algorithm. We see that in Scenario 1 the percentage of selecting the correct number of constraints reaches 98.6%, and it is 97.4% in Scenario 2, indicating high accuracy in choosing the number of constraints.

The estimated eigenfunctions ϕ_{jl} for component processes $X_j(t)$ are displayed in Figure 7 and are seen to be very accurate. The resulting estimated varying coefficient functions $\hat{\beta}_{kj}$ that define the constraints as in eq. (4) are presented in Figure 8. We note here that the true regression coefficient functions $\beta_{kj}(t)$ in the simulation are similar but not the same as the estimated functions $\hat{\beta}_{kj}(t)$ obtained

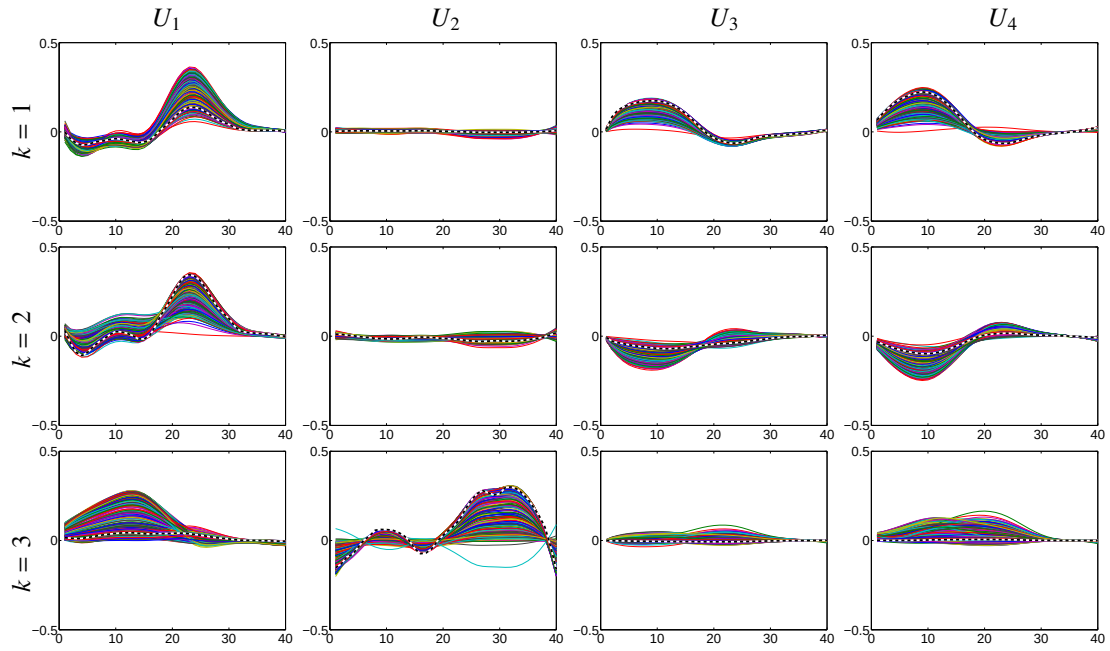

Figure 8: Estimated constraint functions $\beta_{kj}(t)$ that define the functional linear manifolds, based on 500 simulation replicates, for components $U_j(t)$, $j = 1, \ldots, 4$ (from left to right) and constraint indices $k = 1, 2, 3$ (top-down), superimposed on the true functions (white dotted lines).

in Section 4. The estimated varying coefficient functions are seen to be relatively well behaved and overall are reasonably good reflections of the true functions.

6 Discussion

The proposed linear manifold model aims to identify linear combinations of the components of multivariate functional data that are characterized by low variances. It is linear in the functional components and is defined by time-varying coefficient functions. This model is simple, interpretable and can be implemented with suitable transformations that may be applied to each functional component. With such transformations included, this approach can adequately reflect complex nonlinear relationships on the original scale. We exemplify this approach with logarithmic transformations of four behavioral time courses observed over the lifespan of Drosophila flies.

The transformation functions can also be estimated from the data, for example by selecting the transformation from the Box-Cox family. For situations where the component functions have different degrees of variability or units of measurements, one may also consider normalized component functions

$Z_j(t) = \{U_j(t) - \mu_j(t)\}/G_j(t,t)^{1/2}$, $j = 1, \ldots, M$, so that the integrated variance as in the constraints (4) will not be dominated by certain component functions. It depends on the specific situation whether such normalization makes sense, which it likely does when the component processes have different scales or units.

As a by-product, we obtain parsimonious fits of the multivariate process components by using the properties of the linear manifold model coupled with the orthonormality of the coefficient functions. We illustrate this by demonstrating that the resulting representations of the trajectories provide smooth versions of the observed trajectories that furnish significant dimension reduction over unconstrained fitting of multivariate functional data.

There is intrinsic interest in uncovering changing relationships between behavioral variables in aging, exemplified here with a biodemographic study of Drosophila flies for which the proposed approach is useful, as it will be in many other applications. For example, linear manifold models may aid in designing traffic monitoring systems in an efficient way. More generally, the linear manifold model provides a useful tool for the modeling of multivariate functional data.

Appendix A: Proofs

Lemma 4.3 in Bosq (2000) implies that

$$\begin{aligned} |\hat{\lambda}_{rl} - \lambda_{rl}|^2 &\leq \int\int \left\{\widehat{G}_r(t,u) - G_r(t,u)\right\}^2 dtdu, \;\; r = 1, \ldots, M, \\ \int \left\{\hat{\phi}_{rl}(t) - \phi_{rl}(t)\right\}^2 dt &\leq 2\sqrt{2}\,\Delta_l^{-1}\left[\int\int \left\{\widehat{G}_r(t,u) - G_r(t,u)\right\}^2 dtdu\right], \; l = 1, \ldots, L. \end{aligned} \tag{36}$$

Results (32) then follow from (36) and the assumptions on Δ_k.

Results (33) will be implied if one can show

$$|\hat{a}_{rspq} - a_{rspq}| = O_p(n^{-1/2}), \quad 1 \leq r, s \leq M,\ 1 \leq p \leq P_r,\ 1 \leq q \leq P_s. \tag{37}$$

To establish (37), observe

$$\begin{aligned}
&\int \psi_{rp}(t)\psi_{sq}(t)\hat{\phi}_{rl}(t)\hat{\phi}_{sm}(t)dt - \int \psi_{rp}(t)\psi_{sq}(t)\phi_{rl}(t)\phi_{sm}(t)dt \\
= &\int \psi_{rp}(t)\psi_{sq}(t)\left[\left\{\hat{\phi}_{rl}(t) - \phi_{rl}(t)\right\}\hat{\phi}_{sm}(t) + \left\{\hat{\phi}_{sm}(t) - \phi_{sm}(t)\right\}\phi_{rl}(t)\right]dt \\
\leq &\left[\int \left\{\psi_{rp}(t)\psi_{sq}(t)\hat{\phi}_{sm}(t)\right\}^2 dt \int \left\{\hat{\phi}_{rl}(t) - \phi_{rl}(t)\right\}^2 dt\right]^{1/2} \\
&+\left[\int \left\{\psi_{rp}(t)\psi_{sq}(t)\phi_{rl}(t)\right\}^2 dt \int \left\{\hat{\phi}_{sm}(t) - \phi_{sm}(t)\right\}^2 dt\right]^{1/2} = O_p(n^{-1/2}),
\end{aligned}$$

using (32).

Next, observe

$$\begin{aligned}
\widehat{E}(\xi_{rl}\xi_{sm}) - E(\xi_{rl}\xi_{sm}) &= \int\int \left\{\widehat{G}_{rs}(t,u) - G_{rs}(t,u)\right\}\hat{\phi}_{rl}(t)\hat{\phi}_{sm}(u)dudt \\
&\quad + \int\int \left\{\hat{\phi}_{rl}(t) - \phi_{rl}(t)\right\}G_{rs}(t,u)\hat{\phi}_{sm}(u)dudt \\
&\quad + \int\int \left\{\hat{\phi}_{su}(t) - \phi_{su}(t)\right\}G_{rs}(t,u)\phi_{rl}(u)dudt \\
&= I + II + III,
\end{aligned}$$

where

$$\begin{aligned}
I &\leq \left[\int \left\{\int \{\widehat{G}_{rs}(t,u) - G_{rs}(t,u)\}\hat{\phi}_{rl}(t)dt\right\}^2 du \int \hat{\phi}_{sm}(u)^2 du\right]^{1/2} \\
&= O_p\left(\left[\int\int \left\{\widehat{G}_{rs}(t,u) - G_{rs}(t,u)\right\}^2 dtdu\right]^{1/2}\right).
\end{aligned}$$

Terms II and III are handled analogously, and the result in (33) follows from (32).

The results in the Corollary follow from (33). The first result in (34) is an application of the Bauer-Fike theorem (p. 192 in Stewart and Sun (1990), see also Li (2007), p. 15-2). The second result in (34) then follows using results of Davis and Kahan (1970), see also Parlett (1998) p. 244 and Li (2007) p. 15-3). Finally, (35) is a consequence of the convergence of the $\hat{\boldsymbol{b}}_k$ to $\boldsymbol{b}_k$, $k = K, \ldots, P^*$ in (35), due to the definition of $\mathcal{L}$, $\hat{\mathcal{L}}$ as orthogonal complement of the space spanned by the $\boldsymbol{b}_k$ respectively $\hat{\boldsymbol{b}}_k$, $k = K, \ldots, P^*$.

Acknowledgements

We wish to thank two reviewers for constructive comments. This research was supported in part by National Science Foundation grants DMS-1104426 and DMS-1228369, National Science Council grant NSC-98-2118-M-001-019-MY3 and Academia Sinica grant 99-ASIS-02.

References

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

Bickel, P. J., Chen, C., Kwon, J., Rice, J., Van Zwet, E. and Varaiya, P. (2007) Measuring Traffic. *Statist. Sci.*, 22, 581–597.

Bosq, D. (2000) *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer.

Carey, J.R., Papadopoulos, N.T., Kouloussis, N.A., Katsoyannos, B.I., Müller, H.G., Wang, J.L., Tseng, Y.K. (2006). Age-specific and lifetime behavior patterns in Drosophila melanogaster and the Mediterranean fruit fly, Ceratitis capitata. *Experimental Gerontology* **41**, 93–97.

Chiou, J.M., Müller, H.G., Wang, J.L. (2004). Functional response models. *Statistica Sinica* **14**, 675–693.

Christensen, K., Doblhammer, G., Rau, R., Vaupel, J.W. (2009). Ageing populations: the challenges ahead. *Lancet* 372, 1196–1208.

Crimmins, E.M. (2004). Trends in the health of the elderly. *Annual Review of Public Health* 25, 79-98.

Cupidon, J., Eubank, R., Gilliam, D. and Ruymgaart, F. (2008) Some properties of canonical correlations and variates in infinite dimensions. *J. Multiv. Anal.*, 99, 1083–1104.

Davis, C. and Kahan, W. (1970) The rotation of eigenvectors by a perturbation. III. *SIAM J. Numerical Analysis* 7, 1–46.

Dubin, J. and Müller, H. G. (2005) Dynamical correlation for multivariate longitudinal data. *J. Am. Statist. Ass.*, 100, 872–881.

Eubank, R. L. and Hsing, T. (2008) Canonical correlation for stochastic processes. *Stochastic Processes and their Applications*, 118, 1634–1661.

He, G., Müller, H.-G. and Wang, J.-L. (2003) Functional canonical analysis for square integrable

stochastic processes. *J. Multiv. Anal.*, 85, 54–77.

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.

Li, M. Z. F. (2008), A generic characterization of equilibrium speed-flow curves. *Transportation Science*, 42, 220–235.

Li, R.C. (2007) *Matrix Perturbation Theory.* Chapter 15 in Handbook of Linear Algebra, L. Hogben, R. Brualdi, A. Greenbaum and R. Mathias (editors). New York: Chapman and Hall/CRC

Li, Y. and Hsing, T. (2007) On rates of convergence in functional linear regression. *J. Multiv. Anal.*, 98, 1782–1804.

Manton, K.G., Corder, L., Stallard, E. (1997) Chronic disability trends in elderly United States populations: 19821994. *Proceedings of the National Academy of Science* 94, 2593-2598.

Müller, H.G. (2005) Functional modelling and classification of longitudinal data. *Scand. J. Statist.*, 32, 223–240.

Müller, H.G. and Yao, F. (2008) Functional additive models. *J. Am. Statist. Ass.*, 103, 1534–1544.

Pan, J. and Yao, Q. (2008) Modeling multiple time series via common factors. *Biometrika*, 95, 365–379.

Papadopoulos, N.T., Carey, J.R., Katsoyannos, B.I., Kouloussis, N.A., Müller, H.G., Liu, X. (2002) Supine behaviour predicts time to death in Mediterranean fruit flies (Ceratitis capitata). *Proceedings of the Royal Society B* **269**, 1633–1637.

Papadopoulos, N.T., Katsoyannos, B.I., Kouloussis, N.A., Carey, J.R., 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.

Parlett, B.N. (1998) *The Symmetric Eigenvalue Problem*. SIAM: Philadelphia.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis* (2nd ed.). New York: Springer.

Stewart, G.W. and Sun, J.G. (1990) *Matrix Perturbation Theory*. Boston: Academic Press.

Yao, F., Müller, H.G. and Wang, J.L. (2005a) Functional data analysis for sparse longitudinal data. *J. Am. Statist. Ass.*, 10, 577–590.

Yao, F., Müller, H.G. and Wang, J.L. (2005b) Functional linear regression analysis for longitudinal

data. *Ann. Statist.*, 33, 2873–2903.