*Biometrika* (2016), **xx**, x, *pp*. 1–26

*Printed in Great Britain*

# A pairwise interaction model for multivariate functional and longitudinal data

BY JENG-MIN CHIOU

*Institute of Statistical Science, Academia Sinica, 128 Section 2 Academia Road, Nankang, Taipei 11529, Taiwan*

jmchiou@stat.sinica.edu.tw

AND HANS-GEORG MÜLLER

*Department of Statistics, University of California, Davis, One Shields Avenue, Davis, California 95616, U.S.A.*

hgmueller@ucdavis.edu

## SUMMARY

Functional data vectors consisting of samples of multivariate data, where each component is a random function, are increasingly encountered but have not yet been comprehensively investigated. We introduce a simple pairwise interaction model that leads to an interpretable and straightforward decomposition of multivariate functional data and of their variation into component-specific processes and pairwise interaction processes. The latter quantify the degree of pairwise interactions between the components of the functional data vectors, while the component-specific processes reflect the functional variation of a particular functional vector component that cannot be explained by the other components. Thus the proposed model offers an extension of the usual notion of a covariance or correlation matrix for multivariate vector data to functional data vectors and generates an interpretable functional interaction map. The decomposition provided by the model can also serve as a basis for subsequent analysis, such as the study of the network structure of functional data vectors. The decomposition of the total variance into component-wise and interaction contributions can be quantified by an $R^2$-like de-

composition. We provide consistency results for the proposed methods and illustrate the model by applying it to sparsely sampled longitudinal data from the Baltimore Longitudinal Study of Ageing, examining the relationships between body mass index and blood fats.

*Some key words*: Covariance modeling; Functional data analysis; Functional vector; Longitudinal data; Multivariate stochastic process; Variance decomposition.

## 1. INTRODUCTION

Multivariate functional data that can be interpreted as a sample of the repeated realisations of a multivariate stochastic process are ubiquitous and are encountered in applications ranging from traffic monitoring (Chiou, 2012) to biodemography (Carey et al., 2006). Such data arise in many longitudinal studies in the biomedical or social sciences. While classical functional data are perceived as samples of continuously observed random trajectories, longitudinal studies typically generate measurements that are sparsely located over time and are often corrupted by additional measurement errors. Bridging the gap between classical functional and longitudinal data, in recent years a paradigm has emerged in which longitudinal data are viewed as generated by an underlying multivariate stochastic process in continuous time, from which they are sparsely sampled (Rice, 2004; Yao et al., 2005; Li & Hsing, 2010; Müller & Yang, 2010; Kim & Zhao, 2013). This unified perspective enables the development of functional models that address data-analytic needs across a wide spectrum, including continuously observed or densely sampled functional data as well as longitudinal data that are sparsely sampled with noisy measurements.

Previous approaches to multivariate functional/longitudinal data have emphasised dimension reduction along the multivariate direction through various approaches, including dynamic factor models (Hays et al., 2012), that have been studied primarily in time series analysis (Pan & Yao, 2008; Lam et al., 2011). In this setting, one works within a framework where only one realisation of a process is available and correspondingly strong structural assumptions are needed. A second distinct setting is functional modeling, where one has repeated realization of the multivariate process. In this setting a linear manifold representation has been recently studied (Chiou & Müller, 2014). Here we introduce a different approach to the functional/longitudinal setting. The proposed model is intended for both functional, i.e., densely sampled, and longitudinal, i.e., sparsely sampled, data. It provides an interpretable decomposition of the underlying non-stationary multi-

variate stochastic process into component-specific processes and pairwise interaction processes. This leads to an implied decomposition of the overall variation into sums of component- and interaction-specific variations.

The proposed nonparametric approach draws from the rapidly expanding arsenal of functional methodology, which tends to be highly flexible and unconstrained by the typically strong modeling assumptions that are needed for well-established parametric random effects models in longitudinal data analysis (Verbeke et al., 2014). We consider multivariate functional/longitudinal data for which all components are measured on a continuous scale. For some of the procedures and results, we assume that the underlying stochastic processes are multivariate Gaussian. If the components correspond to repeated generalised observations, an underlying Gaussian process can often still be assumed (Hall et al., 2008; Serban et al., 2013). Previous nonparametric approaches for multivariate functional/longitudinal data have focused on the application of smoothing techniques for individual curves (Xiong & Dubin, 2010; Xiang et al., 2013). The adaptation of more traditional parametric random effects models to multivariate longitudinal data was discussed in Fieuws & Verbeke (2006). Few specific functional approaches have been developed for multivariate functional data, except for the bivariate case. For this special case, the notion of a functional correlation aiming at a meaningful correlation coefficient has been widely investigated, including extensions of canonical correlation to the functional case and alternative approaches of dynamic and singular correlation (Leurgans et al., 1993; Dubin & Müller, 2005; Yang et al., 2011). Similarly, function-to-function regression, in which one functional component is regressed on the other, has been well investigated (Ramsay & Dalzell, 1991; Faraway, 1997; Şentürk & Müller, 2010; Gervini, 2014).

Bivariate processes $(X, Y)$ have traditionally been represented through separate functional principal component analyses of the component processes, then introducing a truncation threshold and including the first few eigencomponents, thus reducing the functional data to finite-dimensional data vectors. This application of functional principal components for quantifying the relationships of multivariate processes is relatively straightforward (Zhou et al., 2008; Berrendero et al., 2011). For functional data with a hierarchical structure, a multilevel functional principal component analysis approach has been shown to be useful to model intra- and inter-subject variation (Di et al., 2009).

In contrast to componentwise approaches to dimension reduction via separate principal components or other basis expansions, the proposed approach is based on the cross-covariance surfaces between the component pairs. A key feature is that these surfaces can be consistently estimated in both densely and sparsely sampled situations, thus providing the statistician with essential information about the dependency structure of the various component processes. Our approach relies on previous work on cross-covariance operators and related estimators for various functional/longitudinal settings (Baker, 1974; Gualtierotti, 1979; Yang et al., 2011; Xu & Mackenzie, 2012) for the proposed decomposition of multivariate processes and component estimators. This decomposition is simple and its components provide insight into the structure of the underlying but often unobservable multivariate process. We discuss consistency of the estimated components and illustrate our methods with longitudinal data from the Baltimore Longitudinal Study on Ageing.

## 2. A functional pairwise interaction model

The observed data can be viewed as being generated from an underlying $p$-dimensional stochastic process $\mathbb{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$, $p > 1$, where the component processes $X_j$ are defined on a domain $\mathcal{T} = [0, T]$. We assume that these component processes are square integrable and smooth, i.e., they are at least twice continuously differentiable, and that they follow the model

$$X_j(t) = \mu_j(t) + Y_j(t) + \sum_{k=1, k\neq j}^{p} Z_{jk}(t), \quad (j = 1, \ldots, p). \tag{1}$$

The various components of this model have the following interpretations that relate to their means and covariances.

1. The processes $Z_{jk}$ $(j, k = 1, \ldots, p, j \neq k)$ represent pairwise interaction processes that are assumed to be mutually independent and independent of all processes $Y_j$. Here, $Z_{jk} = Z_{kj}$. Specifically, for $j_1 \neq k_1$ and $j_2 \neq k_2$, $Z_{j_1k_1}$ and $Z_{j_2k_2}$ are mutually independent if either $j_1 \neq j_2$ or $k_1 \neq k_2$. The pairwise interaction processes have means and covariances

$$E\{Z_{jk}(t)\} = 0, \quad \operatorname{cov}\{Z_{jk}(s), Z_{jk}(t)\} = G_{jk}(s, t),$$

where the $G_{jk}$ are referred to as the auto-covariance functions of the interaction processes $Z_{jk}$.

2. The component-specific processes $Y_j$ represent the independent parts of the random components $X_j$ that cannot be explained by interactions. The processes $Y_j$ are assumed to be mutually independent and also independent of the processes $Z_{jk}$, with

$$E\{Y_j(t)\} = 0, \quad \operatorname{cov}\{Y_j(s), Y_j(t)\} = H_j(s,t).$$

3. The component processes $X_j$ have the marginal mean functions $\mu_j$, auto-covariance functions $C_j$ and cross-covariance functions $C_{jk}$,

$$E\{X_j(t)\} = \mu_j(t), \quad \operatorname{cov}\{X_j(s), X_j(t)\} = C_j(s,t), \quad \operatorname{cov}\{X_j(s), X_k(t)\} = C_{jk}(s,t). \quad (2)$$

From the assumption that the $Y_j$ are mutually independent and also independent of the processes $Z_{jk}$, it follows that

$$C_j(s,t) = H_j(s,t) + \sum_{k=1,k\neq j}^{p} G_{jk}(s,t), \quad (3)$$

$$C_{jk}(s,t) = G_{jk}(s,t), \quad (j,k = 1,\ldots,p). \quad (4)$$

The component processes $X_j$ possess a Karhunen–Loève representation, which allows us to expand them in a basis that consists of the eigenfunctions of the linear covariance operator with the kernel $C_j$. Writing $X_j^c(t) = X_j(t) - \mu_j(t)$ and denoting these eigenfunctions by $\chi_{jm}$ ($j = 1,\ldots,p;\ m \geq 1$), we obtain the representations

$$X_j^c(t) = \sum_{m=1}^{\infty} \xi_{jm}\,\chi_{jm}(t), \quad (j = 1,\ldots,p;\ t \in \mathcal{T}), \quad (5)$$

where $\xi_{jm} = \int X_j^c(s)\,\chi_{jm}(s)\mathrm{d}s$ are the functional principal components of $X_j$, and $\operatorname{var}(\xi_{jl}) = \theta_{jl}$. Then the corresponding representation for covariance surfaces $C_j$ is

$$C_j(s,t) = \sum_{m=1}^{\infty} \theta_{jm}\,\chi_{jm}(s)\,\chi_{jm}(t), \quad (j = 1,\ldots,p;\ s,t \in \mathcal{T}). \quad (6)$$

A simple calculation shows that for each $j$ ($j = 1,\ldots,p$) and all $l \geq 1$, $\operatorname{cov}(\xi_{jl}, \xi_{jr}) = \theta_{jl}$ if $r = l$ and 0 otherwise. For $j \neq k$ ($j,k = 1,\ldots,p$) and for any $l, r \geq 1$, $X_j$ and $X_k$ are mutually correlated with

$$\operatorname{cov}(\xi_{jl}, \xi_{kr}) = \int_{\mathcal{T}}\int_{\mathcal{T}} C_{jk}(s,t)\,\chi_{jl}(s)\,\chi_{kr}(t)\,\mathrm{d}s\mathrm{d}t. \quad (7)$$

Equation (3) reflects the decomposition of the auto-covariances of the component processes into those of the component-specific processes and the interaction processes, where according

to (4) the latter are identical with the cross-covariances of the component processes. This decomposition can be quantified in terms of fractions of variance explained by these model components for each fixed time $t$, as follows. Defining $R^2_j(t) = H_j(t,t)/C_j(t,t)$, $R^2_{j\cdot k}(t) = G_{jk}(t,t)/C_j(t,t)$ and

$$R^2_j = T^{-1}\int_{\mathcal{T}} R^2_j(t)\mathrm{d}t, \tag{8}$$

$$R^2_{j\cdot k} = T^{-1}\int_{\mathcal{T}} R^2_{j\cdot k}(t)\mathrm{d}t, \quad (k \neq j), \tag{9}$$

it follows from (3) that

$$R^2_j + \sum_{k=1,k\neq j}^{p} R^2_{j\cdot k} = 1, \quad (1 \leq j \leq p). \tag{10}$$

Here $R^2_j$ can be interpreted as the fraction of variance of $X_j$ explained by the component-specific processes $Y_j$ and $R^2_{j\cdot k}$ as the fraction explained by the interaction processes $Z_{jk}$ $(k \neq j)$. Because of the different denominators, $R^2_{j\cdot k}$ and $R^2_{k\cdot j}$ $(j \neq k)$ are not the same. The decompositions (10) hold without distributional assumptions such as Gaussianity and where independence was required for the model components above, uncorrelatedness will suffice.

## 3. Data model and estimation

The sample of multivariate longitudinal data consists of $n$ independent units and is assumed to be generated by a $p$–dimensional smooth stochastic process according to

$$W_{ijl} = X_{ij}(T_{ijl}) + \varepsilon_{ijl}, \quad (i = 1,\ldots,n;\ j = 1,\ldots,p;\ l = 1,\ldots,N_{ij}), \tag{11}$$

where $i$ is the subject index, $j$ the component index and $l$ the index of the repeated measurements of the assumed smooth underlying random trajectory $X_{ij}$. These measurements are taken at a random number of $N_{ij} \geq 2$ measurement times. In (11), the $\varepsilon_{ijl}$ are independently and identically distributed measurement errors with $E(\varepsilon_{ijl}) = 0$, $\mathrm{var}(\varepsilon_{ijl}) = \sigma^2$. The times $T_{ijl}$ at which the measurements are taken may follow an equidistant regular design, in which case $N_{ij} = m$ is fixed for all $i$ and $j$ and the spacing of $T_{ijl}$ is $T/(m-1)$.

In contrast to this regular design case, the measurements in many longitudinal studies are taken at random and irregular locations and also are often sparse. The underlying distributions of these locations may vary across components, but these distributions are assumed to be the same across subjects. However the numbers of measurements are not assumed to be the same across

subjects and while the distributions of the random number of observations made per component are assumed to be the same across subjects they may differ between components. We assume all random numbers of measurements to be larger or equal to 2. The proposed methods will be illustrated with an application to longitudinal data in Section 7. For this longitudinal case, the locations $T_{ijl}$ of the measurements and the $N_{ij}$ are assumed to be independent from all other random variables. The $T_{ijl} \in \mathcal{T}$ are assumed to have a smooth density $f_{ij}$, which for all $i, j$ is uniformly bounded away from zero.

The basic elements of the proposed pairwise interaction model are the mean and auto-covariance functions for each component process and their cross-covariance functions. Their estimation follows steps that have already been developed elsewhere. Different approaches are used for dense regular and sparse irregular designs. For dense regular designs, we simply estimate the means, auto- and cross-covariances by forming the corresponding cross-sectional averages across all data in the sample. For sparse irregular designs, smoothing methods are needed, as cross-sectional averaging is not possible if the measurements are made at random locations (Yao et al., 2005). Estimation for the case of sparse irregular designs depends on various smoothing steps that are implemented with one- or two-dimensional local least squares estimators. We choose the necessary bandwidths by generalised cross-validation for each smoothing application.

An important consideration is that the covariance surfaces are restricted to be symmetric and positively definite. While the symmetry of these covariance surfaces is inherent in the steps below, positive definiteness is ensured by reconstructing the estimated covariances $\tilde{C}$ that are obtained by smoothing. Estimates of the eigenvalues/eigenfunctions $\hat{\lambda}_j$, $\hat{\rho}_j$ are calculated numerically and then the estimated eigendecomposition $\tilde{C}(s,t) = \sum_{j=1}^{K} \hat{\lambda}_j \hat{\rho}_j(s)\hat{\rho}_j(t)$ for a large value of $K$ is projected on the space of positive definite matrices by omitting the eigencomponents with non-positive eigenvalues from the eigendecomposition. This projection is optimal in the $L^2$ sense (Hall et al., 2008) and results in the appropriately restricted estimate $\hat{C}(s,t) = \sum_{j=1,\hat{\lambda}_j>0}^{K} \hat{\lambda}_j \hat{\rho}_j(s)\hat{\rho}_j(t)$.

We assume that $m_j$ measurements are available for the $j$th unit, and abbreviate this to $m$ in the following. The various estimation and smoothing steps are briefly delineated as follows. Additional details can be found in the Supplementary Material .

1. For each $j$, obtain the estimates $\hat{\mu}_j$ of the component mean functions $\mu_j$ (2) by smoothing the scatterplot data $\{(T_{ijl}, W_{ijl});\ i = 1, \ldots, n,\ l = 1, \ldots, m\}$.
2. For each $j$, obtain the estimates $\tilde{C}_j(s,t)$ of $C_j(s,t)$, the component process auto-covariance surfaces in (2), which correspond to smoothed covariance surfaces. These estimates are based on two-dimensional local weighted least squares smoothing, applied to responses $\{W_{ijl} - \hat{\mu}_j(T_{ijl})\}\{W_{ijl} - \hat{\mu}_j(T_{ijl'})\}$ with associated two-dimensional predictors $(T_{ijl}, T_{ijl'})$, $(i = 1, \ldots, n; l, l' = 1, \ldots, m)$. The final estimate $\hat{C}_j(s,t)$ is obtained by projecting $\tilde{C}_j(s,t)$ on the space of positive and symmetric covariance surfaces, as described above.
3. For each $j$, obtain a representation $\hat{X}_{ij}^{KL}$ of $X_{ij}$ by the Karhunen–Loève expansion, based on the above estimators $\hat{\mu}_j(t)$ and $\hat{C}_j(s,t)$ and the associated estimated eigenbasis $\hat{\phi}_{jk}$, $k \geq 1$. The estimated eigenbasis, coupled with a suitable truncation point $K$, yields

$$\hat{X}_{ij}^{KL}(t) = \hat{\mu}_j(t) + \sum_{k=1}^{K} \hat{\xi}_{ij} \hat{\phi}_{jk}(t).$$

4. For each pair $(j,k)$, obtain the estimate $\tilde{C}_{jk}(s,t)$ based on two-dimensional local weighted least squares smoothing, applied to responses $\{W_{ijl} - \hat{\mu}_j(T_{ijl})\}\{W_{ikl'} - \hat{\mu}_k(T_{ikl'})\}$ with associated two-dimensional predictors $(T_{ijl}, T_{ikl'})$, $(i = 1, \ldots, n;\ l, l' = 1, \ldots, m)$, and then obtain the final estimate $\hat{C}_{jk}(s,t)$ by projecting $\tilde{C}_{jk}(s,t)$ onto the space of positive semi-definite and symmetric surfaces. In addition to playing a role in the basic model (1) as a cross-covariance, $C_{jk}$ is simultaneously a covariance surface and thus resides in the space of positive semi-definite and symmetric surfaces.
5. Set $\hat{G}_{jk}(s,t) = \hat{C}_{jk}(s,t)$.
6. Obtain $\tilde{H}_j(s,t) = \hat{C}_j(s,t) - \sum_{k=1,k\neq j}^{p} \hat{G}_{jk}(s,t)$ and the estimate $\hat{H}_j(s,t)$ of $H_j$ by projecting $\tilde{H}_j(s,t)$ as described above, such that $\hat{H}_j(s,t)$ is positive semi-definite.
7. For $C_j^*(t,t) = \hat{H}_j(t,t) + \sum_{k=1,k\neq j}^{p} \hat{G}_{jk}(t,t)$, let $\hat{R}_j^2(t) = \hat{H}_j(t,t)/C_j^*(t,t)$ and $\hat{R}_{j\cdot k}(t) = \hat{G}_{jk}(t,t)/C_j^*(t,t)$. Obtain the estimates

$$\hat{R}_j^2 = T^{-1} \int_{\mathcal{T}} \hat{R}_j^2(t) \mathrm{d}t, \tag{12}$$

$$\hat{R}_{j\cdot k}^2 = T^{-1} \int_{\mathcal{T}} \hat{R}_{j\cdot k}^2(t) \mathrm{d}t, \quad (k \neq j), \tag{13}$$

for (8) and (9).

With these estimates, the variation of the component curves can be decomposed into contributions from the interaction processes and contributions that are component-specific and cannot be predicted from interactions with other components.

## 4. Consistency

We consider here the longitudinal case with sparse and noisy measurements under similar assumptions as in Theorem 2 in Müller & Yao (2010). First, we list the regularity assumptions that are needed for the theoretical derivation in the Supplementary Material and include assumptions on the design points introduced in (11), where measurements are taken, the behaviour of eigenfunctions $\phi_m$ and eigenvalues $\lambda_m$ for $m \to \infty$ and also on the large sample behaviour of the smoothing bandwidths. Proofs and auxiliary results are in the Supplementary Material.

The numbers of measurements $N_{ij}$ that are available for the $j$th trajectory of the $i$th subject ($i = 1, \ldots, n$; $j = 1, \ldots, p$) are allowed to be very sparse, as we only require the following:

(A1) The $N_{ij}$ are random variables that are independently and identically distributed as $N_j$, where the $N_j$ are bounded positive discrete random variables with $Pr(N_j \geq 2) > 0$, and $\{(T_{ijl},\ l \in J_{ij}), (W_{ijl},\ j \in J_{ij})\}$ are independent and also independent of $N_{ij}$, for all subsets $J_{ij} \subseteq \{1, \ldots, N_{ij}\}$.

For the bandwidths used in the smoothing steps (S1)–(S3), as detailed in the Supplementary Material, we require that, as $n \to \infty$,

(A2) $h_{\mu_j} \to 0$, $h_{C_j} \to 0$, $h_{C_{jk}} \to 0$, $nh_{\mu_j} \to \infty$, $nh_{C_j}^2 \to \infty$, $nh_{C_{jk}}^2 \to \infty$, $(j, k = 1, \ldots, p)$.

Denote the densities of $T_{ij}$ and joint densities of $(T_{ij}, T_{ik})$ by $f_j(t)$ and $f_{jk}(t, s)$ $(j, k = 1, \ldots, p)$.

(A3) The derivatives up to order 4 of mean functions $\mu_j$ of processes $X_j$, auto-covariance functions $C_j$, cross-covariance functions $C_{jk}$ and densities $f_j, f_{jk}$ exist and are continuous. The densities $f_j, f_{jk}$ are strictly positive on the domain $\mathcal{T}$.

(A4) The kernel functions $\kappa_1, \kappa_2$ that are used in the smoothing steps with details in the Supplementary Material have vanishing moments of order 1 and their moments of order 2 are positive.

Denote the $L^2$ norm by $\|f\| = \{\int_{\mathcal{T}} f^2(t)dt\}^{1/2}$ and the Hilbert–Schmidt norm for functions in two arguments by $\|\psi\|_s = \{\int_{\mathcal{T}}\int_{\mathcal{T}}\{\psi^2(t,s)dtds\}^{1/2}$ and define the rates

$$\alpha_{nj} = (nh_{\mu_j})^{-1/2} + h_{\mu_j}^2, \quad \beta_{nj} = n^{-1/2}h_{C_j}^{-1} + h_{C_j}^2, \quad \gamma_{njk} = n^{-1/2}h_{C_{jk}}^{-1} + h_{C_{jk}}^2. \tag{14}$$

(A5) For the rates in (14) it holds that $\alpha_n/\beta_{nj} \to 0,\ \alpha_{nj}/\gamma_{njk} \to 0$.

Then we have the following consistency result for mean and covariance estimators $\hat{\mu}_j, \hat{C}_j$ and $\hat{C}_{jk}$, for which details are provided in the Supplementary Material:

Theorem 1. *Under (A1)-(A5), for $j,k = 1,\ldots,p,\ j \neq k$, and with $\alpha_{nj},\beta_{nj},\gamma_{njk}$ as in (14), estimators (S1)–(S3) satisfy*

$$\|\hat{\mu}_j - \mu_j\| = O_p(\alpha_{nj}), \quad \|\hat{C}_j - C_j\|_s = O_p(\beta_{nj}), \quad \|\hat{C}_{jk} - C_{jk}\|_s = O_p(\gamma_{njk}). \tag{15}$$

In the special case where one employs bandwidths $h_{\mu_j} \sim n^{-1/5}$, $h_{C_j} \sim n^{-1/6}$ and $h_{C_{jk}} \sim n^{-1/6}$, one obtains the usual nonparametric rates

$$\|\hat{\mu}_j - \mu_j\| = O_p(n^{-4/5}), \quad \|\hat{C}_j - C_j\|_s = O_p(n^{-2/3}), \quad \|\hat{C}_{jk} - C_{jk}\|_s = O_p(n^{-2/3}). \tag{16}$$

Applying (3) and (4), the representation $H_j = C_j - \sum_{k=1,k\neq j}^{p} G_{jk}$ motivates the estimators

$$\hat{H}_j = \hat{C}_j - \sum_{k=1,k\neq j}^{p} \hat{G}_{jk}, \tag{17}$$

where $\hat{G}_{jk} = \hat{C}_{jk}$. We then find that

$$\|\hat{H}_j - H_j\|_s = O_p\left\{\max(\beta_{nj},\ \gamma_{njk},\ \ j,k = 1,\ldots,p)\right\}. \tag{18}$$

Convergence rates for the eigenvalue and eigenfunction estimators for the corresponding component processes follow from these results.

Define eigenvalue/eigenfunction pairs as $(\lambda_m, \varphi_m)$ and covariance functions $G$ with estimators $\hat{G}$ for any one of the processes $X_j, Y_j, Z_{jk}$. Applying Lemma 4·3 of Bosq (2000) and setting

$$\delta_1 = \lambda_1 - \lambda_2, \quad \delta_m = \min(\lambda_{m-1} - \lambda_m, \lambda_m - \lambda_{m+1}) \quad (m \geq 2), \tag{19}$$

one obtains for the estimators $(\hat{\lambda}_m, \hat{\varphi}_m)$ of $(\lambda_m, \varphi_m)$

$$|\ \hat{\lambda}_m - \lambda_m\ | \leq \|\hat{G} - G\|_s, \quad \|\hat{\phi}_m - \phi_m\| \leq 2\sqrt{2}\delta_m^{-1}\|\hat{G} - G\|_s, \tag{20}$$

where the bounds on the right hand side are uniform over $m$. This implies that the estimators of the various eigencomponents are consistent under mild additional assumptions. The rates of

convergence depend on those of the covariance function estimators and on the rates of decline of the eigenvalues and of their spacings, with specific examples in the Supplementary Material.

Next turning to the consistency of the estimators (11) and (12) of the fractions of variance explained $R_j^2$ and $R_{j\cdot k}^2$ defined in (8) and (9), we make the additional assumption

(A6) It holds that $\inf_{t\in\mathcal{T}} C_j(t,t) > 0$ $(j = 1,\ldots,p)$.

and then define slightly modified estimators $\tilde{C}'_j$ for $C_j$, as follows: $\tilde{C}'_j(s,t) = \hat{C}_j(s,t)$ for $s \neq t$, $\tilde{C}'_j(t,t) = \max\{\hat{C}_j(t,t), (\log n)^{-1}\}$. These modified estimators have the same convergence properties as the original estimators $\hat{C}_j$ but prevent the estimators to assume too small values, as the modified estimators are bounded below by a non-random null sequence.

Consider $H_j(s,t)$ and $G_{jk}(s,t)$ $(j,k = 1,\ldots,p,\ j \neq k)$ defined in Section 1, and observe that as stated in Section 1 as part of the basic model, $G_{jk}$ also has the properties of a covariance of the interaction process $Z_{jk}$. We then obtain by Mercer's theorem the representations

$$H_j(s,t) = \sum_{m=1}^{\infty} \lambda_{Hjm}\phi_{Hjm}(s)\phi_{Hjm}(t), \quad G_{jk}(s,t) = \sum_{m=1}^{\infty} \lambda_{Gjkm}\phi_{Gjkm}(s)\phi_{Gjkm}(t),$$

where $(\lambda_{Hjm},\phi_{Hjm})_{m\geq 1}$, respectively, $(\lambda_{Gjkm},\phi_{Gjkm})_{m\geq 1}$, are the eigenvalues and eigenfunctions of the auto-covariance operators of $H_j$ and of $G_{jk}$. Denoting the eigenvalue and eigenfunction estimators derived from the estimators $\hat{H}_j$ in (15), respectively $\hat{G}_{jk}$ as described in estimation step 4 of §3, after the projection as in Hall et al. (2008) that has been discussed in §3 and in step 5 of the estimation procedures, we obtain estimators $\hat{C}_j(s,t) = \sum_{m=1}^{M(n)} \hat{\lambda}_{Cjm}\hat{\phi}_{Cjm}(s)\hat{\phi}_{Cjm}(t)$, $\hat{H}_j(s,t) = \sum_{m=1}^{M(n)} \hat{\lambda}_{Hjm}\hat{\phi}_{Hjm}(s)\hat{\phi}_{Hjm}(t)$, $\hat{G}_{jk}(s,t) = \sum_{m=1}^{M(n)} \hat{\lambda}_{Gjkm}\hat{\phi}_{Gjkm}(s)\hat{\phi}_{Gjkm}(t)$, which are consistent for $C_j, H_j$ and $G_{jk}$, respectively, if we choose a sequence $M(n) \to \infty$ as $n \to \infty$.

Define the remainder sums

$$\zeta_{Cn} = \sum_{m=M(n)}^{\infty} \lambda_{Cjm}, \quad \zeta_{Hn} = \sum_{m=M(n)}^{\infty} \lambda_{Hjm}, \quad \zeta_{Gn} = \sum_{m=M(n)}^{\infty} \lambda_{Gjkm},$$

where due to the $L^2$ integrability these are null sequences as $M(n) \to \infty$. With $\eta_n = \max(\beta_{nj},\ \gamma_{njk},\ j,k = 1,\ldots,p)$, we assume

(A7) $\zeta_{Cn}\log n \to 0$, $\zeta_{Gn}\log n \to 0$, $\zeta_{Hn}\log n \to 0$, $M(n)\eta_n \log n \to 0$, as $n \to \infty$.

The following consistency result is for the modified versions of estimators (8) and (9),

$$\hat{R}_j^2 = \frac{1}{T}\int_{\mathcal{T}} \frac{\hat{H}_j(t,t)}{\tilde{C}'_j(t,t)}\mathrm{d}t, \quad \hat{R}_{j\cdot k}^2 = \frac{1}{T}\int_{\mathcal{T}} \frac{\hat{G}_{jk}(t,t)}{\tilde{C}'_j(t,t)}\mathrm{d}t, \quad (k \neq j).$$

THEOREM 2. *Under (A1)-(A7), for $j,k = 1,\ldots,p$, $j \neq k$, the estimators $\hat{R}_j^2$, $\hat{R}_{j\cdot k}^2$ are consistent for $R_j^2$, $R_{j\cdot k}^2$, respectively,*

$$|\, \hat{R}_j^2 - R_j^2 \,| = O_p\{\zeta_{Cn}\log n + \zeta_{Hn}\log n + M(n)\eta_n \log n\}, \tag{21}$$

$$|\, \hat{R}_{j\cdot k}^2 - R_{j\cdot k}^2 \,| = O_p\{\zeta_{Cn}\log n + \zeta_{Gn}\log n + M(n)\gamma_{njk}\log n\}. \tag{22}$$

As a specific example, consider using the optimal rates in (16) by choosing appropriate bandwidths for the smoothing steps and assume that the eigenvalue sequences for $H_j$ and $G_{jk}$ decline polynomially, say $\lambda_{Cjm} \sim \lambda_{Hjm} \sim \lambda_{Gjkm} \sim m^{-\alpha}$ for $\alpha > 1$. Then $\zeta_{Cn} \sim \zeta_{Hn} \sim \zeta_{Gn} \sim M(n)^{1-\alpha}$. The fastest convergence is then achieved with the choice $M(n) = n^{2/(3\alpha)}$, which gives the rate $n^{-2(\alpha-1)/(3\alpha)}\log n$ in (21) and (22). For $\alpha = 2$ this rate is $n^{-1/3}\log n$, for $\alpha = 4$ it is $n^{-1/2}\log n$.

## 5. REPRESENTING THE COMPONENT TRAJECTORIES

In addition to the eigencomponents of the component processes, we seek representations of the component-specific processes $Y_j$ and of the pairwise interaction processes $Z_{jk}$, based on the indirect information about these processes that is contained in the measurements that are available for processes $X_{ij}$. However, even processes $X_{ij}$ are not directly observed in the case of longitudinal studies, where only sparse and potentially noisy measurements of these processes are available. We discuss the additional estimation problem this poses for the estimation of individual process trajectories in the context of the application to longitudinal biomedical data in the next section.

We use a contemporary model for relating the observed data that are generated by processes $X_{ij}$ to the component-specific processes $Y_j$. In the case where all of the component processes $Y_j$ and $Z_{jk}$ $(j,k = 1,\ldots,p)$ are jointly Gaussian, the basic model assumptions (1) imply $E\left\{Z_{jk}(t) \mid X_j(t), X_k(t)\right\} = \beta_{jk1}(t)\, X_j^c(t) + \beta_{jk2}(t)\, X_k^c(t)$ with varying coefficient functions $\beta_{jk1}, \beta_{jk2}$. A somewhat weaker assumption is to postulate the concurrent linear model

$$Z_{jk}(t) = \beta_{jk1}(t)\, X_j^c(t) + \beta_{jk2}(t)\, X_k^c(t) + \sum_{m=1}^{\infty} \zeta_{jkm}\varphi_{jkm}(t), \tag{23}$$

where the $\zeta_{jkm}$ are random coefficients that expand the unexplained residual $Z_{jk}(t) - E\left\{Z_{jk}(t) \mid X_j(t), X_k(t)\right\}$ for the varying coefficient model, with $E(\zeta_{jkm}) = 0$ and $\text{var}(\zeta_{jkm}) = \lambda_{jkm}$. Here $\lambda_{jkm}$ and $\varphi_{jkm}$, $m \geq 1$, correspond to the eigenvalues and eigenfunctions of the conditional covariance function of $Z_{jk}$, defined below in (27). The varying coefficient functions $(\beta_{jk1}, \beta_{jk2})$ only depend on the covariance and cross-covariance functions of the processes $X_j(t)$ and $X_k(t)$.

Under the assumption in (23),

$$V_{jk}(t) = E\left\{Z_{jk}(t) \mid X_j(t), X_k(t)\right\} = \beta_{jk1}(t)\, X_j^c(t) + \beta_{jk2}(t)\, X_k^c(t), \tag{24}$$

with $E\{Z_{jk}(t)\} = 0$. For $U_j(t) = E\left\{Y_j(t) \mid X_1(t), \ldots, X_p(t)\right\}$, (24) leads to

$$U_j(t) = X_j(t) - \mu_j(t) - \sum_{k=1, k\neq j}^{p} E\left\{Z_{jk}(t) \mid X_j(t), X_k(t)\right\}, \tag{25}$$

and then to the overall decomposition of $X_j(t)$ for all $t \in \mathcal{T}$ into conditional processes

$$X_j(t) = \mu_j(t) + U_j(t) + \sum_{k=1, k\neq j}^{p} V_{jk}(t). \tag{26}$$

For the conditional processes $V_{jk}$, the expectation of the conditional covariance is

$$E\left[\text{cov}\left\{Z_{jk}(s), Z_{jk}(t) \mid X_j(t), X_k(t)\right\}\right] = \sum_{m=1}^{\infty} \lambda_{jkm}\, \varphi_{jkm}(s)\, \varphi_{jkm}(t), \tag{27}$$

and the covariance function of the conditional expectation

$$\text{cov}\left\{V_{jk}(s), V_{jk}(t)\right\} = \beta_{jk}(s)^{\mathrm{T}}\, \Gamma_{jk}(s,t)\, \beta_{jk}(t), \tag{28}$$

where $\beta_{jk}(t) = \{\beta_{jk1}(t), \beta_{jk2}(t)\}^{\mathrm{T}}$ and $\Gamma_{jk}(s,t)$ is a $2 \times 2$ matrix with the (1,1), (2,2), (1,2) and (2,1) entries $\Gamma_{jk,11}(s,t) = C_j(s,t)$, $\Gamma_{jk,22}(s,t) = C_k(s,t)$, $\Gamma_{jk,12}(s,t) = \Gamma_{jk,21}(s,t) = C_{jk}(s,t)$.

From (27) and (28),

$$\text{cov}\{Z_{jk}(s), Z_{jk}(t)\} = \beta_{jk}(s)^{\mathrm{T}}\, \Gamma(s,t)\, \beta_{jk}(t) + \sum_{m=1}^{\infty} \lambda_{jkm}\, \varphi_{jkm}(s)\, \varphi_{jkm}(t). \tag{29}$$

Since $\text{cov}\{Z_{jk}(t), X_j(t)\} = C_{jk}(t,t)$, where $C_{jj}(t,t) = C_j(t,t)$, defining the $2 \times 2$ matrix function $M_{jk}(t)$ with elements

$$m_{jk,11}(t) = C_j(t,t), \quad m_{jk,22}(t) = C_k(t,t), \quad m_{jk,12}(t) = m_{jk,21}(t) = C_{jk}(t,t),$$

and the vectors $v_{jk}(t) = \{C_{jk}(t,t), C_{jk}(t,t)\}^{\mathrm{T}}$, leads to

$$\left\{\beta_{jk1}(t), \beta_{jk2}(t)\right\}^{\mathrm{T}} = \{M_{jk}(t)\}^{-1} v_{jk}(t). \tag{30}$$

When $C_j(t,t) = C_k(t,t) = 1$, this simplifies to $\beta_{jk1}(t) = \beta_{jk2}(t) = C_{jk}(t,t)/\{1 + C_{jk}(t,t)\}$.

The corresponding estimators $\{\hat{\beta}_{jk1}(t), \hat{\beta}_{jk2}(t)\}$ are easily obtained by substituting the estimators $\hat{C}_j$, $\hat{C}_{jk}$ for $C_j, C_{jk}$, which were obtained in steps 2 and 4 of the estimation procedure described in Section 3. Given the observed or imputed observations $X_j(t)$ and $X_k(t)$, the estimate of the conditional mean (24) is then given by

$$\hat{V}_{jk}(t) = \hat{\beta}_{jk1}(t)\, X_j^c(t) + \hat{\beta}_{jk2}(t)\, X_k^c(t). \tag{31}$$

Substituting the estimators for the population quantities as before, (26) leads to the following estimators $\hat{U}_j(t)$ of $E\left\{Y_j(t) \mid X_1(t), \ldots, X_p(t)\right\}$,

$$\hat{U}_j(t) = X_j(t) - \hat{\mu}_j(t) - \sum_{k=1, k\neq j}^{p} \hat{V}_{jk}(t), \tag{32}$$

where all of these quantities are defined for the $i$th subject ($i = 1, \ldots, n$). More explicitly, for the $i$th subject, replacing $X_j$ in (31) and (32) by $X_{ij}$ or its best linear predictor given the observations, we obtain $\hat{V}_{ijk}(t)$ and $\hat{U}_{ij}(t)$. In conjunction with the concurrent linear model assumption in (23), this results in approximate fitted trajectories

$$\hat{X}_{ij}(t) = \hat{\mu}_j(t) + \hat{U}_{ij}(t) + \sum_{k=1, k\neq j}^{p} \hat{V}_{ijk}(t), \quad (j = 1, \ldots, p). \tag{33}$$

## 6. Application to the Baltimore Longitudinal Study of Ageing

### 6·1. *Auto- and Cross-Covariance Functions*

The proposed model is demonstrated for the analysis of multivariate longitudinal data that were collected in the Baltimore Longitudinal Study of Ageing (Shock et al., 1984; Pearson et al., 1997). We focus on the longitudinal multivariate relationships between the time courses of weight, total cholesterol and triglyceride levels in blood. The relationship between weight and blood lipids is of general interest, as elevated levels for each of these components are believed to have negative cardiovascular consequences (Ford et al., 2009). The data were collected for a sample of $n = 1370$ subjects on the age interval $[17, 97]$. The included subjects were followed for up to 32 years between 1958 and 1991.

A problem that arises for the analysis of these data and other longitudinal data is that the measurements are sparse and irregular. Frequently, not all longitudinal component processes are observed at the same time points, for instance if at patient visits only a subset of the longitudinal

characteristics of interest is measured or there are missing data for some of the components. In addition to being recorded at irregular time points $T_{ijl}$, the observed data $W_{ijl}$ are assumed to be contaminated by noise, according to model (11). In the Baltimore Longitudinal Study on Ageing, the measurements for the different components were all recorded at the same times, varying by subject, so that here $T_{ijm} = T_{ikm}$ for all $i$ and $m$.

As the units of measurement for weight and blood lipids are vastly different, to implement our approach it is beneficial to first normalise the data for the various components prior to fitting the interaction model, where it is assumed that the interactions and therefore cross-covariances of the component processes $X_j$ are symmetric. Fixing $j$, we obtain normalised observations $W^s_{ijl} = \{W_{ijl} - \hat{\mu}_j(T_{ijl})\}/\hat{C}_j(T_{ijl}, T_{ijl})^{1/2}$, where estimates $\hat{\mu}_j$ and $\hat{C}_j$ are obtained using only the component-specific observations $\{W_{ijl}\}$. The normalised $\{W^s_{ijl}\}$ are then used as observations to implement the model component estimation steps as described in Section 3, setting the mean of the normalised observations to zero. As an additional last step in the analysis, one may transform the fitted model components back to their original scale by multiplying with $\hat{C}_j(t,t)^{1/2}$.

For our analysis, we select a subset of the data that consists of subjects who survived beyond 85 years of age and use their measurements collected between ages 55–85, where we choose the domain $\mathcal{T} = [61, 80]$ as a subinterval to reduce boundary effects. Our analysis includes $n = 278$ subjects, excluding 7 outlying individuals and subjects who do not have at least two measurements in $\mathcal{T}$ or whose maximal weight, cholesterol or triglyceride exceeds 125kg, 350 dL/mg or 500 dL/mg, respectively. The longitudinal observations for weight, cholesterol and triglycerides along with the estimated mean functions and the auto-covariance functions are shown in Fig. 1, demonstrating that there is great variability across subjects, accompanied by substantial noise in the measurements and sparsity in the designs.

The mean functions of all three variables generally decrease with age, except for cholesterol, which shows an increasing trend and a flat peak between 63 and 67 years of age. The auto-covariance surfaces indicate larger variances near 80 years of age for weight and triglycerides, whereas the auto-covariance of cholesterol has an earlier variance peak and then tapers off. After implementing the normalisation step, the estimated auto-covariance surfaces $\hat{C}_j$ of the normalised observations that can be found in the three diagonal panels of Fig. 2 exhibit roughly similar patterns across the three components.

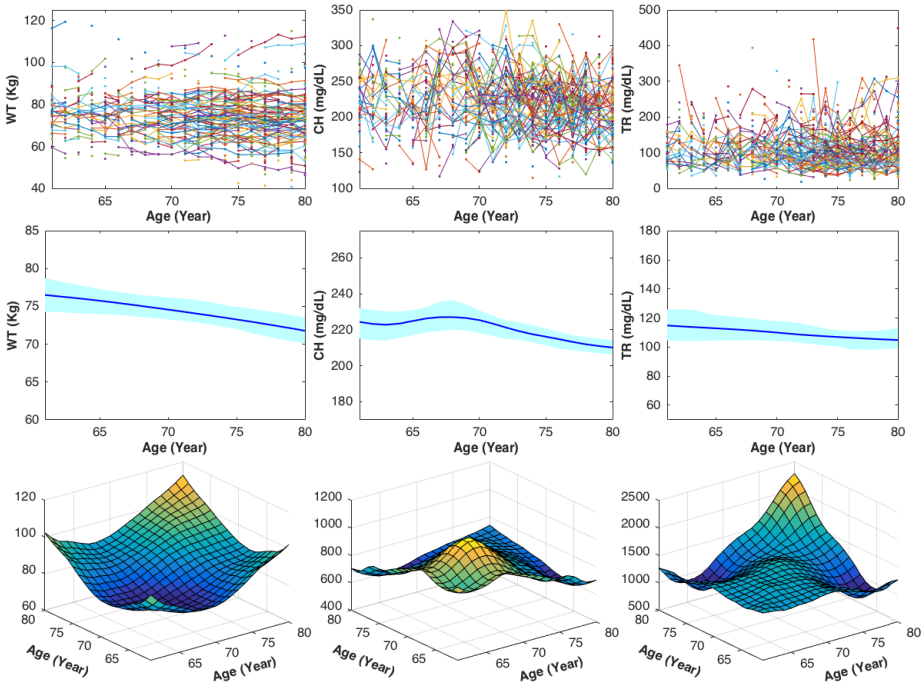


Fig. 1: Longitudinal observations of weight (WT), cholesterol (CH) and triglycerides (TR), top panels from left to right, estimated mean functions $\hat{\mu}_j(t)$, including 95% pointwise bootstrap confidence bands, middle, based here and in the following on resampling the data with replacement 200 times, and covariance functions $\hat{C}_j(t,s)$, bottom.

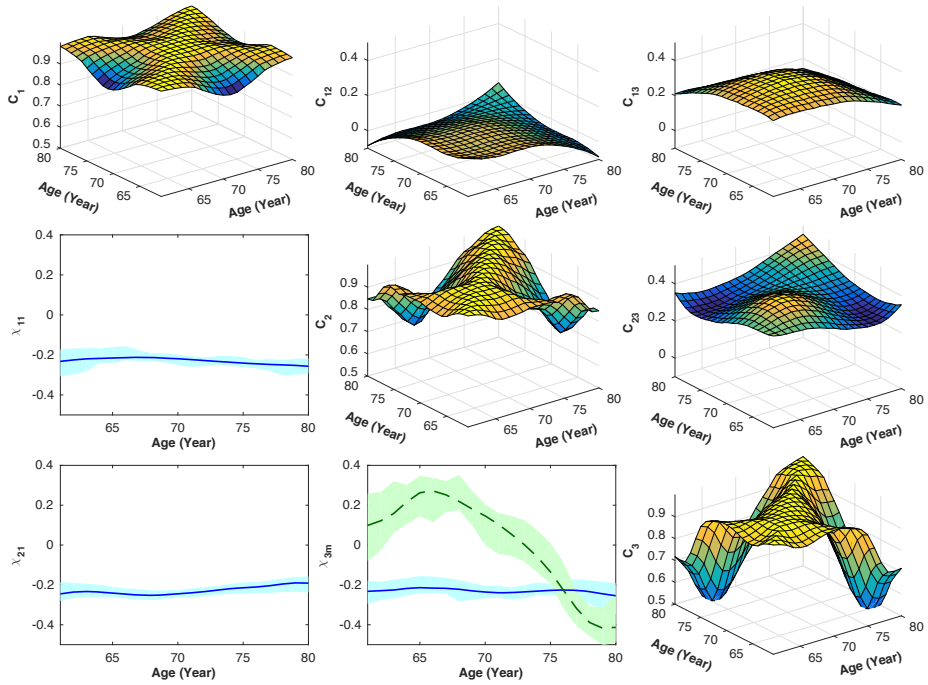


Fig. 2: Estimates of the normalised auto-covariance functions of $\hat{C}_j$, $j = 1,2,3$, for weight (WT), cholesterol (CH) and triglycerides (TR) in the diagonal panels from left to right, the cross-covariance functions of $\hat{C}_{jk}$ in the panels forming the upper right triangle and the eigenfunctions $\hat{\chi}_{11}$ for WT, $\hat{\chi}_{21}$ for CH, and $\hat{\chi}_{31}$ (solid) and $\hat{\chi}_{32}$ (dashed) for TR, including 95% pointwise bootstrap confidence bands in the panels forming the lower left triangle.

We observe high correlations between measurements collected both in the beginning or near the end of the domain and relatively smaller correlations between those made at early and mid-range ages. The cross-covariance surfaces $\hat{C}_{jk}$ in the three right and upper off-diagonal panels are near zero and overall display little variation. One finds some negative correlations between early and late ages for weight and cholesterol, middle upper panel, a flat surface at a level around 0·2 between weight and triglycerides, right upper panel, and an overall larger and somewhat more structured surface with peaks at early and late ages for the correlation between cholesterol and triglycerides, middle right panel.

Decomposing the auto-covariances for the component processes into their respective eigenfunctions, we chose the number of included components to explain a fraction of total variance exceeding 90%. The resulting eigenfunctions derived from the cross-covariances are in the lower and left off-diagonal panels, indicating that a major mode of variation for all components is the variation of the overall component levels, as reflected in flat eigenfunctions. There is evidence for additional time-dynamic components for the triglycerides, where the second mode of variation corresponds to a contrast of triglyceride levels before and after the age of 75.

### 6·2. *Decomposition of Variance*

The proportions of variance explained by the component-specific processes $Y_j$ along with their pairwise interaction terms $Z_{jk}$ are quantified by the estimators of $R^2_j$ and $R^2_{j\cdot k}$ in (8) and (9), which are given in Table 1. They indicate that the specific component weight with $\hat{R}^2_1 = 0{\cdot}654$ makes the largest and cholesterol with $\hat{R}^2_2 = 0{\cdot}559$ the second largest contribution to the overall variation. These component-specific variations are idiosyncratic in the sense that they constitute contributions to overall variation that cannot be explained by interactions. In comparison, the component-specific contribution of triglycerides is smaller with $\hat{R}^2_3 = 0{\cdot}303$. This means the observed variation of weight is least explained by interactions with the other variables, and that of triglycerides is mostly explained by interactions.

The interactions between weight and cholesterol are negligible and do not contribute to overall variation, while those between weight and triglycerides have $R^2$ values of approximately 0·3 and those between the two blood lipids of approximately 0·4. This confirms what was graphically evident from the cross-covariance surfaces in Fig. 2. The non-interacting fractions of weight and

cholesterol carry strong contributions to overall variation, while the non-interacting fraction of triglycerides and the interactions between weight and triglycerides and those between cholesterol and triglycerides make sizeable and roughly similar contributions. For example, the variation of triglycerides can be attributed in approximately equal parts to the idiosyncratic component of triglycerides alone, their interaction with cholesterol and their interaction with weight, while two thirds of the variation of weight is idiosyncratic and one third associated with cholesterol.

Table 1: *Estimated $R^2_j$, diagonal elements, and $R^2_{j\cdot k}$, off-diagonal elements, for the data from the Baltimore Longitudinal Study on Ageing. According to the definitions* (8) *and* (9)*, these fractions of variance explained are not symmetric; rows sum to 1. The first row and column, (j,k=1), correspond to weight, the second row and column, (j,k=2), to cholesterol and the third row and column, (j,k=3), to triglycerides.*

| | | $k$ | |
|---|---|---|---|
| $j$ | 1 | 2 | 3 |
| 1 | 0·654 | 0·046 | 0·300 |
| 2 | 0·047 | 0·559 | 0·394 |
| 3 | 0·304 | 0·393 | 0·303 |

### 6·3. *Decomposition of Trajectories*

To decompose the observed trajectories, or in the sparse case, the latent trajectories that generate the data, into the components corresponding to the interaction and component-specific processes according to the concurrent varying coefficient model (31), the first step is to estimate the varying coefficient functions $\beta_{jk1}$ and $\beta_{jk2}$. The estimators are shown in Fig. 3. Here $\beta_{jk1} = \beta_{jk2}$ since $C_j(t,t) = C_k(t,t) = 1$ when the model component estimators are based on normalised observations. As in the previous analysis, the varying coefficient function $\beta_{121}$ that is used to recover the interaction process $Z_{12}$ to quantify the interaction of weight and cholesterol is seen to be quite small. In contrast, the function $\beta_{231}$ to recover the interaction of triglycerides and cholesterol is more substantial and has larger values around ages 68 and 80, where the interaction appears to be stronger than at other ages. Similarly, the substantial interaction between weight and triglycerides also gives rise to a sizeable interaction function with a flat peak also around 68, then tapering off to small values near age 80. Therefore, this interaction diminishes and triglycerides become more independent from weight as subjects move beyond age 75.

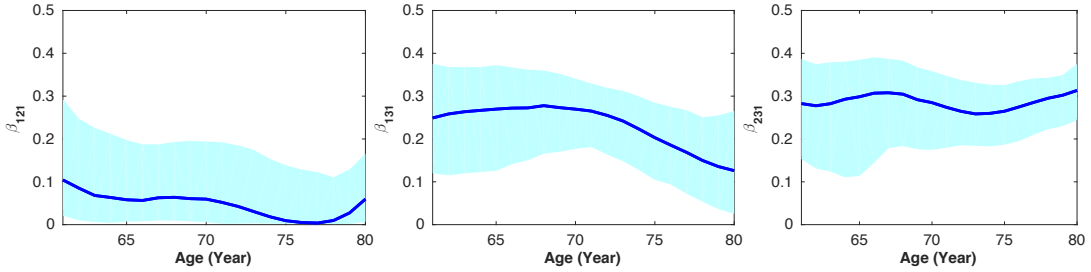


Fig. 3: Regression coefficient function estimators $\beta_{jk1}(t)$ of each pair of component functions, $\beta_{121}(t)$ for weight and cholestrerol, $\beta_{131}(t)$ for weight and triglycerides and $\beta_{231}(t)$ for cholesterol and triglycerides, with 95% bootstrap pointwise confidence intervals, from left to right.

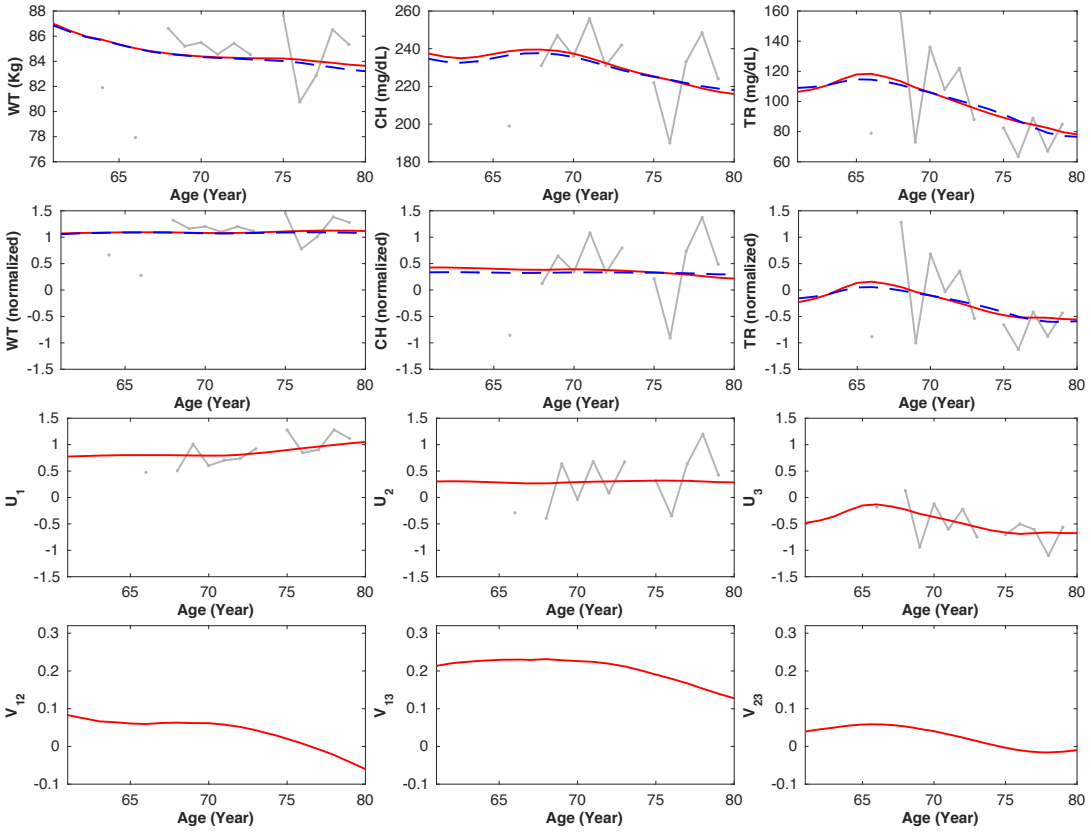


Fig. 4: Sparsely observed data (grey) for one subject, with weight (WT) in the left column, cholesterol (CH) in the middle column and triglycerides (TR) in the right column, overlaid with the respective Karhunen–Loève representation (dashed) and the fitted interaction model representation as in (33) (solid), on the original scale in the top row and the normalised scale in the second row. The fitted conditional subject-specific component functions $U_j(t)$ resulting from the decomposition (26) into conditional components are depicted in the third panel (solid) on the normalised scale along with the data. The estimated conditional interaction components $V_{12}(t)$, $V_{13}(t)$ and $V_{23}(t)$ are in the bottom panel, in the order WT and CH, WT and TR, CH and TR, from left to right.

We illustrate the decomposition at the individual level in Fig. 4 for the data of one subject. The data and functional principal component approach for sparse longitudinal data are shown in Fig. 4 for both original and normalised scales. This figure also illustrates the fitted trajectories $\hat{X}_{ij}$ that result from the proposed interaction model, as described in (26) and (33). The fit of the interaction model is seen to be in good agreement with the conditional Karhunen–Loève expansion. The fitted component-specific functions $\hat{U}_j$ and pairwise interaction functions $\hat{V}_{jk}$ are displayed in the two bottom rows of the figure. The conditional component-specific functions are quite flat, in contrast to the interaction functions which exhibit temporal variation. Especially the interaction function of weight and triglycerides in the middle panel of the bottom row shows positive interaction with a peak at around age 67, while the interaction between cholesterol and triglycerides in the right panel of the bottom row indicates a sign change around age 74.

### 6·4. *Model Comparisons*

For comparison purposes, we consider a linear mixed model for longitudinal data using a set of pre-determined basis functions $b_j(t) = \left(\varphi_{j1}(t), \ldots, \varphi_{jL}(t)\right)^{\mathrm{T}}$ as a vector of covariates with both fixed and the random effects,

$$X_{ij}(t) = b_j(t)^{\mathrm{T}}\alpha_j + b_j(t)^{\mathrm{T}}\gamma_{ij} + \epsilon_j(t),\ (i = 1, \ldots, n;\ j = 1, \ldots, p). \tag{34}$$

Here $\alpha_j = (\alpha_1, \ldots, \alpha_L)^{\mathrm{T}}$ is a vector of fixed regression coefficients and $\gamma_{ij} = \left(\gamma_{ij1}, \ldots, \gamma_{ijL}\right)^{\mathrm{T}}$ is a vector of subject-specific random effects following a multivariate normal distribution with mean zero, auto-covariance $\mathrm{cov}(\gamma_{ij}, \gamma_{ij}) = \Gamma_{jj}$ and cross-covariance $\mathrm{cov}(\gamma_{ij}, \gamma_{ik}) = \Gamma_{jk}$, for $j \neq k$. The random errors $\epsilon_j(t)$, considered over a discrete grid of points $t$, are identically and independently distributed with mean zero and constant variance $\sigma^2$.

We evaluate the prediction performance for the $j$th component function using root mean squared error,

$$\mathrm{RMSE}(X_j) = \left[ n^{-1} \sum_{i=1}^{n} {m_i}^{-1} \sum_{l=1}^{m_i} \left\{ W_{ijl} - \hat{X}_{ij}(T_{ijl}) \right\}^2 \right]^{1/2}, \tag{35}$$

where $W_{ijl}$ is the observation of $X_i(T_{ij})$ contaminated with a random measurement error. We compare the proposed functional pairwise interaction model with traditional functional principal component analysis and the linear mixed model, as described in (34), for two cases, where in the first case the basis functions are chosen as B-spline functions with equally spaced knots,

and in the second case as Legendre polynomials $\phi_{jl} = P_{l-1}^{j-1}$, choosing $L = 3$, 5 and 10 as the number of base functions. Table 2 shows that while functional principal component analysis and the proposed model behave similarly, both perform clearly better than the two versions of the linear mixed model, irrespective of the included number of basis functions.

Table 2: *Root mean squared errors,* RMSE$(X_j)$, $j = 1, 2, 3$ *for the Baltimore Longitudinal Study of Ageing data*

| RMSE | FPCA | FPIM | $LMM_1$ ($L = 3$) | $LMM_2$ ($L = 3$) | $LMM_1$ ($L = 5$) | $LMM_2$ ($L = 5$) | $LMM_1$ ($L = 10$) | $LMM_2$ ($L = 10$) |
|---|---|---|---|---|---|---|---|---|
| $X_1$ | 2·06 | 2·19 | 40·71 | 34·63 | 15·09 | 14·60 | 25·12 | 14·50 |
| $X_2$ | 20·16 | 18·88 | 147·85 | 132·89 | 88·50 | 87·47 | 106·10 | 87·28 |
| $X_3$ | 25·57 | 25·37 | 87·42 | 80·16 | 67·44 | 64·94 | 66·40 | 64·22 |

FPCA, Functional principal component model; FPIM, The proposed functional pairwise interaction model; $LMM_1$, Linear mixed model using B-spline basis functions as covariates; $LMM_2$, Linear Mixed Model using Legendre polynomial bases as covariates.

## 7. Simulations

### 7·1. *Data generation*

We consider two underlying models for generating simulation data to study the finite sample behavior of the proposed functional pairwise interaction model. In Scenario 1, data generation follows the proposed functional pairwise interaction model, while in Scenario 2 it is based on linear mixed modeling, adapted to functional data. For Scenario 1, to be consistent with the method used for the Baltimore Longitudinal Study on Ageing, we consider the functional pairwise interaction model based on standardised trajectories $X_j^s(t) = v_j^{-1/2}(t)\left\{X_j(t) - \mu_j(t)\right\}$. The generation of the data when choosing $p = 3$ functional components proceeds as follows.

1. Set $\tilde{\theta}_1^Y = (9, 4, 1)$, $\tilde{\theta}_2^Y = (10, 5)$, and $\tilde{\theta}_3^Y = (8, 3)$; $\tilde{\phi}_1^Y(t) = \{P_0^0(2t-1), P_1^0(2t-1), P_2^0(2t-1)\}$, where $P_l^k(x)$ is the $l$th-degree Legendre polynomial, $k = 0, \ldots, l$, $\tilde{\phi}_2^Y(t) = \{1, \cos(\pi t)\}$, and $\tilde{\phi}_3^Y(t) = [\exp(t)/\{1 + \exp(t)\}, \log(t + 1)]$. In addition, set $\tilde{\theta}_{12}^Z = (7, 3)$, $\tilde{\theta}_{13}^Z = (8, 3, 1)$, $\tilde{\theta}_{23}^Z = (8, 2)$; $\tilde{\phi}_{12}^Z = \tilde{\phi}_2^Y$, $\tilde{\phi}_{13}^Z = \tilde{\phi}_1^Y$, and $\tilde{\phi}_{23}^Z = \tilde{\phi}_3^Y$.

2. Set the positive semidefinite covariance function for $v_j^{1/2}Y_j$ as $\tilde{H}_j(s,t) = \sum_l \tilde{\theta}_{jl}^Y \tilde{\phi}_{jl}^Y(s)\tilde{\phi}_{jl}^Y(t)$ $(j = 1, \ldots, p)$ and the positive semidefinite auto-covariance function for $v_j^{1/2}Z_{jk}$ as $\tilde{G}_{jk}(s,t) = \sum_r \tilde{\theta}_{jkr}^Z \tilde{\phi}_{jkr}^Z(s)\tilde{\phi}_{jkr}^Z(t)$, $(j,k = 1,2,3,\ j \neq k)$.
3. Compute the positive semidefinite auto-covariance function of $X_j$ by $C_j(s,t) = \tilde{H}_j(s,t) + \sum_{k=1,k\neq j}^p \tilde{G}_{jk}(s,t)$ with $C_j(t,t) = v_j(t)$, and the cross-covariance function of $X_j$ and $X_k$ by $C_{jk}(s,t) = \tilde{G}_{jk}(s,t)$ for $j \neq k$.
4. Compute the auto-covariance of $X_j^s$ by $C_j^s(s,t) = v_j^{-1/2}(s)C_j(s,t)v_j^{-1/2}(t)$, the auto-covariance of $Y_j$ by $H_j(s,t) = v_j^{-1/2}(s)\tilde{H}_j(s,t)v_j^{-1/2}(t)$, and the auto-covariance of $Z_{jk}$ by $G_{jk}(s,t) = v_j^{-1/2}(s)\tilde{G}_{jk}(s,t)v_j^{-1/2}(t)$.
5. Obtain the eigen-pairs $\{\theta_{jl}^Y, \phi_{jl}^Y(\cdot)\}$ of the covariance function $H_j(\cdot,\cdot)$ for all positive eigenvalues $\theta_{jl}^Y$, and the eigen-pairs $\{\theta_{jkr}^Z, \phi_{jkr}^Z(\cdot)\}$ of $G_{jk}(\cdot,\cdot)$ for all positive eigenvalues $\theta_{jkr}^Z$ by the respective spectral decomposition.
6. Let $\xi_{ij}^Y$ be a vector with elements $(\xi_{ijl}^Y)$ and $\xi_{ijk}^Z$ a vector with elements $(\xi_{ijkr}^Z)$. Generate realizations $\{(\xi_{i1}^Y)^\mathrm{T}, (\xi_{i2}^Y)^\mathrm{T}, (\xi_{i3}^Y)^\mathrm{T}, (\xi_{i12}^Z)^\mathrm{T}, (\xi_{i13}^Z)^\mathrm{T}, (\xi_{i23}^Z)^\mathrm{T}\}^\mathrm{T}$ from the multivariate normal distributions with mean zero and block-diagonal variance-covariance matrix $\mathrm{diag}\{(\theta_1^Y)^\mathrm{T}, (\theta_2^Y)^\mathrm{T}, (\theta_3^Y)^\mathrm{T}, (\theta_{12}^Z)^\mathrm{T}, (\theta_{13}^Z)^\mathrm{T}, (\theta_{23}^Z)^\mathrm{T}\}$. Then generate realizations of $Y_j(t)$ and $Z_{jk}(t)$ evaluated at $t = \{T_{ijl}\}$ by $Y_{ij}(t) = (\xi_{ij}^Y)^\mathrm{T}\phi_j^Y(t)$ and $Z_{ijk}(t) = (\xi_{ijk}^Z)^\mathrm{T}\phi_{jk}^Z(s)$ $(j \neq k)$. Here $Y_j(t)$ and $Z_{jk}(t)$ are uncorrelated.
7. Set $\mu_j(t) = 0$ and obtain $X_{ij}(t) = v_j^{1/2}(t)X_{ij}^s(t)$, evaluated at $t = (T_{ijl})$, where $X_{ij}^s(t) = Y_{ij}(t) + \sum_{k=1,k\neq j}^p Z_{ijk}(t)$. Then generate observations $W_{ijl} = X_{ij}(T_{ijl}) + \epsilon_{ijl}$ with random measurement errors.

We consider sample sizes $n = 100$, 200, and 400. The observations for each subject $i$ and component $j$ are sampled at $(T_{ijl})$ from a set of equally spaced time points on $[0,1]$, where the number of observations for each subject is $N_{ij} = m_i$. For the sparse design, $m_i$ follows a discrete uniform distribution in $[20,30]$, whereas $m_i = 51$ for the dense design. The measurement errors $\epsilon_{ijl}$ in (11) are independently sampled from $N(0, 0{\cdot}5)$. Additional results and details for the case of $p = 5$ functional components can be found in Supplementary Material.

For Scenario 2, where data are generated by the linear mixed model in (34), we set $\sigma^2 = 0{\cdot}5$ and $\alpha_j = 0$, $j = 1, \ldots, p$, and select the basis functions $\varphi_{jl}$ as Legendre polynomials. The covariance matrix $\Gamma$ of $\gamma_i = (\gamma_{i1}^\mathrm{T}, \ldots, \gamma_{ip}^\mathrm{T})^\mathrm{T}$ is generated by the squared exponential covariance

function such that the $(j,k)$ element of $\Gamma$ is $\Gamma_{jk} = \exp\{-(s_j - s_k)^2/2\}$, where $s_j$ $(j = 1, \ldots, pL)$ are generated by a discrete uniform distribution in (0, 5). We set $L = 5$ for the number of basis functions that serve as covariates when fitting the linear mixed model.

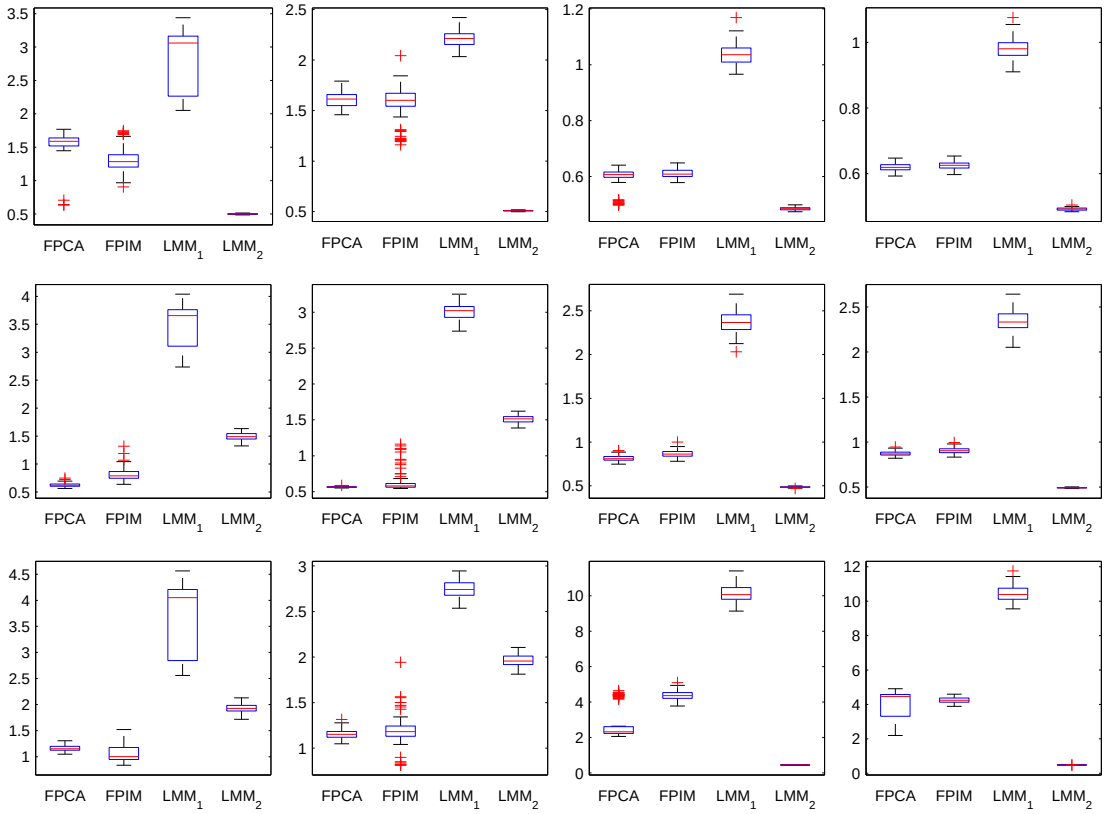


Fig. 5: Boxplots of $\text{RMSE}(X_j)$ in (35), for $j = 1, 2, 3$ from top to bottom, for the following combinations: Scenario 1, sparse design (column 1); Scenario 1, dense design (column 2); Scenario 2, sparse design (column 3); and Scenario 2, dense design (column 4), based on 100 simulation runs. The methods are as described in Table 2.

## 7·2. *Simulation results*

Figure 5 displays the boxplots of $\text{RMSE}(X_j)$, $j = 1, 2, 3$, comparing the proposed functional pairwise interaction model, functional principal component analysis, and the two linear mixed models that were described in the previous section, for the case of $p = 3$, $n = 200$. Table 3 contains the average values of RMSE. The mixed linear models are based on $L = 5$ basis functions, the same number as used in the data generation for the simulations. Additionally, for Scenario 2 and the second linear mixed model the assumed and the simulated basis functions are the same,

so that this model has an inherent advantage, while for the first linear mixed model the five basis functions are chosen differently.

In Scenarios 1, the functional pairwise interaction and functional principal component models perform quite well and indeed much better than the first and overall somewhat better than the second linear mixed model. In Scenarios 2, the second linear mixed model has the best predictions, as expected from the fact that in this situation this is a correctly specified parametric random effects model, while the first linear mixed model performs the worst. We conclude that the linear mixed model approach is non-robust as it strongly depends on a good choice of the basis functions.

Table 3: *Averages of the sample means (with standard errors* $\times 10^{-2}$ *in parentheses) of* $\mathrm{RMSE}(X_j)$ *in* (35) *averaged over* $j = 1, 2, 3$ *for various methods under Scenarios 1 and 2 for sparse (S) and dense (D) designs for sample size* $n = 200$ *and 100 Monte Carlo runs.*

| Scenario | FPCA | FPIM | $\mathrm{LMM}_1$ | $\mathrm{LMM}_2$ |
|---|---|---|---|---|
| 1(S) | 1·11 (0·61) | 1·07 (0·87) | 3·37 (2·85) | 1·31 (0·36) |
| 1(D) | 1·11 (0·30) | 1·13 (0·91) | 2·65 (0·54) | 1·33 (0·28) |
| 2(S) | 1·40 (3·01) | 1·95 (0·88) | 4·53 (1·67) | 0·47 (0·03) |
| 2(D) | 1·82 (3·22) | 1·93 (0·54) | 4·58 (1·53) | 0·49 (0·02) |

(S), sparse design; (D), dense design. FPCA, Functional principal component model; FPIM, The proposed functional pairwise interaction model; $\mathrm{LMM}_1$, Linear mixed model using B-spline basis functions as covariates; $\mathrm{LMM}_2$, Linear Mixed Model using Legendre polynomial bases as covariates.

The functional principal component approach is optimized to minimize the squared prediction errors, which in some cases gives it a slight advantage over the fits of the functional pairwise interaction model, where the interaction modeling is the goal and not prediction. Additional simulation results that include different numbers of eigencomponents in the generated processes, different sample sizes and an evaluation of the varying coefficient functions in the concurrent regression model in (23) as well as of the various estimators of the fractions of variance in (12) and (13) can be found in the Supplementary Material.

## Acknowledgements

The authors wish to thank the editor, the associate editor and the referees for helpful remarks. This research was supported by grants from the National Science Foundation, U.S.A., and the Academia Sinica and the Ministry of Science and Technology, Taiwan.

## Supplementary material

Supplementary material available at *Biometrika* online includes additional details on estimation and on theoretical considerations, including proofs of Theorems 1 and 2, as well as other simulation results.

## References

Baker, C. R. (1974). Joint measures and cross-covariance operators. *Transac. Am. Math. Soc.* **186**, 273–289.

Berrendero, J., Justel, A. & Svarc, M. (2011). Principal components for multivariate functional data. *Computat. Statist. Data Anal.* **55**, 2619–2634.

Bosq, D. (2000) *Linear Processes in Function Spaces.* Lecture notes in statistics **149**. 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 behaviour patterns in *Drosophila melanogaster* and the Mediterranean fruit fly, *Ceratitis capitata*. *Exp. Gerontol.* **41**, 93–97.

Chiou, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *Ann. Appl. Statist.* **6**, 1588–1614.

Chiou, J.-M. & Müller, H.-G. (2014). Linear manifold modelling of multivariate functional data. *J. Roy. Statist. Soc.* B **76**, 605–626.

Di, C.-Z., Crainiceanu, C. M., Caffo,B. S. & Punjabi, N. M. (2009). Multilevel principal component analysis. *Annals of Applied Statistics* **3**, 458-488.

Dubin, J. A. & Müller, H.-G. (2005). Dynamical correlation for multivariate longitudinal data. *J. Am. Statist. Assoc.* **100**, 872–881.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39**, 254–261.

Fieuws, S. & Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* **62**, 424–431.

Ford, E. S., Li, C., Zhao, G., Pearson, W. S. & Mokdad, A. H. (2009). Hypertriglyceridemia and its pharmacologic treatment among U.S. adults. *Arch. Intern. Med.* **169**, 572–578.

Gervini, D. (2014). Dynamic retrospective regression for functional data. *Technometrics* **57**, 26–34.

Gualtierotti, A. F. (1979). On cross-covariance operators. *SIAM J. Appl. Math.* **37**, 325–329.

Hall, P., Müller, H.-G. & Yao, F. (2008). Modelling sparse generalised longitudinal observations with latent Gaussian processes. *J. Roy. Statist. Soc.* B **70**, 730–723.

HAYS, S., SHEN, H., HUANG, J. Z. ET AL. (2012). Functional dynamic factor models with application to yield curve forecasting. *Ann. Appl. Statist.* **6**, 870–894.

KIM, S. & ZHAO, Z. (2013). Unified inference for sparse and dense longitudinal models. *Biometrika* **100**, 203–212.

LAM, C., YAO, Q. & BATHIA, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98**, 901–918.

LEURGANS, S. E., MOYEED, R. A. & SILVERMAN, B. W. (1993). Canonical correlation analysis when the data are curves. *J. Roy. Statist. Soc.* B **55**, 725–740.

LI, Y. & HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38**, 3321–3351.

MÜLLER, H.-G. & YANG, W. (2010). Dynamic relations for sparsely sampled Gaussian processes. *Test* **19**, 1–29.

MÜLLER, H.-G. & YAO, F. (2010). Empirical dynamics for longitudinal data. *Ann. Statist.* **38**, 3458–3486.

PAN, J. & YAO, Q. (2008). Modelling multiple time series via common factors. *Biometrika* **95**, 365–379.

PEARSON, J. D., MORRELL, C. H., BRANT, L. J. & LANDIS, P. K. (1997). Age-associated changes in blood pressure in a longitudinal study of healthy men and women. *J. Gerontol. - Med. Sci.* **52**, 177–183.

RAMSAY, J. O. & DALZELL, C. J. (1991). Some tools for functional data analysis. *J. Roy. Statist. Soc.* B **53**, 539–572.

RICE, J. A. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statist. Sinica* **14** 631–647.

ŞENTÜRK, D. & MÜLLER, H.-G. (2010). Functional varying coefficient models for longitudinal data. *J. Am. Statist. Assoc.* **105**, 1256–1264.

SERBAN, N., STAICU, A.-M. & CARROLL, R. J. (2013). Multilevel cross-dependent binary longitudinal data. *Biometrics* **69**, 903–913.

SHOCK, N. W., GREULICH, R. C., ANDRES, R., LAKATTA, E. G., ARENBERG, D. & TOBIN, J. D. (1984). Normal human aging: The Baltimore Longitudinal Study of Aging. In *NIH Pub. No. 84-2450.* Washington, D.C.: U.S. Government Printing Office.

VERBEKE, G., FIEUWS, S., MOLENBERGHS, G. & DAVIDIAN, M. (2014). The analysis of multivariate longitudinal data: A review. *Statist. Meth. Med. Res.* **23**, 42–59.

XIANG, D., QIU, P. & PU, X. (2013). Nonparametric regression analysis of multivariate longitudinal data. *Statist. Sin.* **23**, 769–789.

XIONG, X. & DUBIN, J. A. (2010). A binning method for analyzing mixed longitudinal data measured at distinct time points. *Statist. Med.* **29**, 1919–1931.

XU, J. & MACKENZIE, G. (2012). Modelling covariance structure in bivariate marginal models for longitudinal data. *Biometrika* **99**, 649–662.

YANG, W., MÜLLER, H.-G. & STADTMÜLLER, U. (2011). Functional singular component analysis. *J. Roy. Statist. Soc.* B **73**, 303–324.

YAO, F., MÜLLER, H.-G. & WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Am. Statist. Assoc.* **100**, 577–590.

ZHOU, L., HUANG, J. & CARROLL, R. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601–619.