ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 1, *pp.* 177–196

Modelling function-valued stochastic processes, with applications to fertility dynamics

Kehui Chen,

University of Pittsburgh, USA

Pedro Delicado

Universitat Politècnica de Catalunya, Barcelona, Spain

and Hans-Georg Müller

University of California at Davis, USA

[Received October 2014. Final revision November 2015]

Summary. We introduce a simple and interpretable model for functional data analysis for situations where the observations at each location are functional rather than scalar. This new approach is based on a tensor product representation of the function-valued process and utilizes eigenfunctions of marginal kernels. The resulting marginal principal components and product principal components are shown to have nice properties. Given a sample of independent realizations of the underlying function-valued stochastic process, we propose straightforward fitting methods to obtain the components of this model and to establish asymptotic consistency and rates of convergence for the estimates proposed. The methods are illustrated by modelling the dynamics of annual fertility profile functions for 17 countries. This analysis demonstrates that the approach proposed leads to insightful interpretations of the model components and interesting conclusions.

Keywords: Asymptotics; Demography; Functional data analysis; Marginal kernel; Product principal component analysis; Tensor product representation

1. Introduction

In various applications one encounters stochastic processes and random fields that are defined on temporal, spatial or other domains and take values in a function space, which is assumed to be the space of square integrable functions L^2. More specifically, for $\mathcal{S} \subset \mathbb{R}^{d_1}$ and $\mathcal{T} \subset \mathbb{R}^{d_2}$, we consider the stochastic process $X : \mathcal{T} \to L^2(\mathcal{S})$ and denote its value at *time* $t \in \mathcal{T}$ by $X(\cdot, t)$, which is a square integrable random function with argument $s \in \mathcal{S}$. A key feature of our approach is that we consider the case where we have n independent observations of the functional stochastic process.

A specific example that we shall discuss in detail (Section 5) is that of female fertility profile functions $X(\cdot, t)$, which are available annually ($t =$ year) for $n = 17$ countries, with age as argument s. The starting point is the *age-specific fertility rate* ASFR $X(s, t)$ for a specific country, which is defined as

Address for correspondence: Kehui Chen, Department of Statistics and Department of Psychiatry, University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA 15260, USA.
E-mail: khchen@pitt.edu

 1369–7412/17/79177

$$X(s,t) = \mathrm{ASFR}(s,t) = \frac{\text{births during year } t \text{ given by women aged } s}{\text{person-years lived during year } t \text{ by women aged } s}. \quad (1)$$

Fig. 1 illustrates the ASFR-data for the USA from 1951 to 2006. Fig. 1(a) shows $\mathrm{ASFR}(\cdot,t)$ for $t = 1960, 1980, 2000$. The image plot representing $\mathrm{ASFR}(s,t)$ for all possible values of s and t in Fig. 1(b) provides a visualization of the dynamics of fertility in the USA over the whole period.

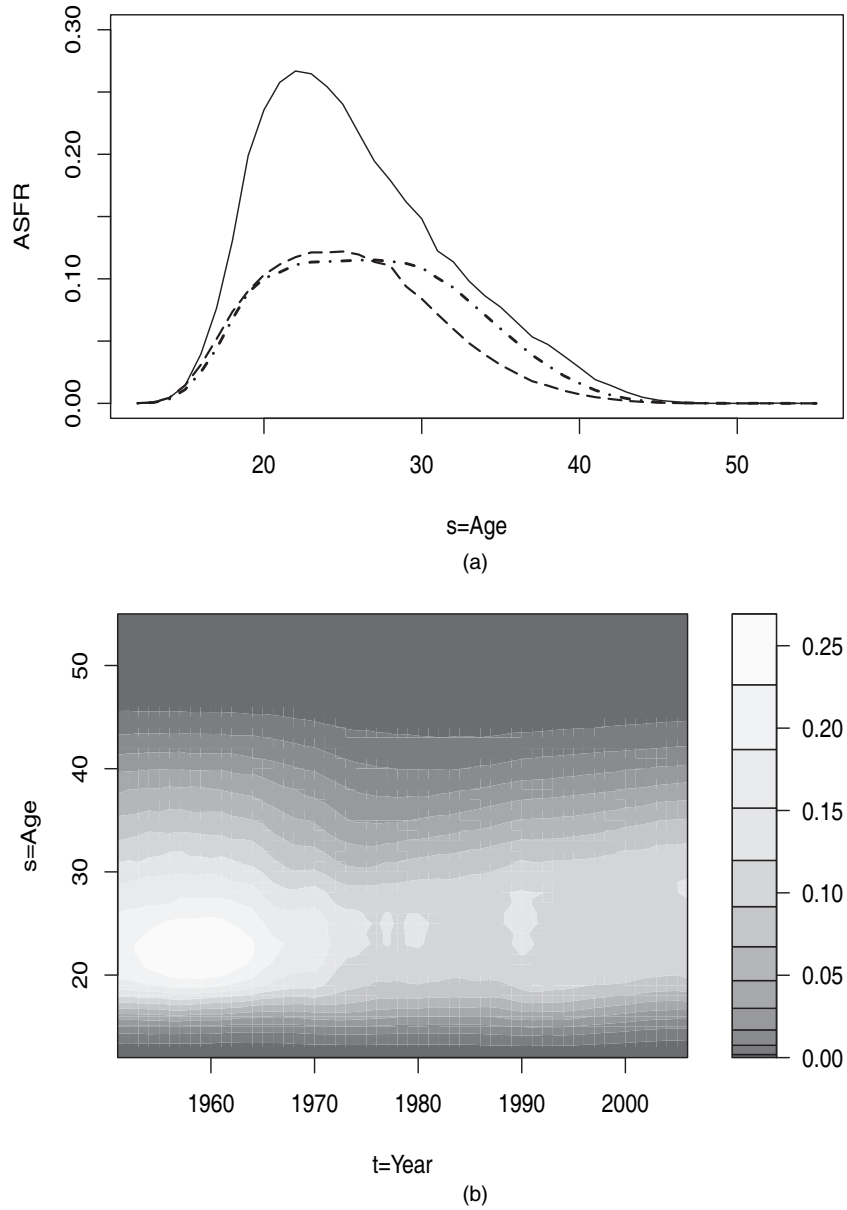


Fig. 1. Age-specific fertility rate for the USA: (a) profiles for three calendar years (———, $t = 1960$; – – –, $t = 1980$; ·–·–·–, $t = 2000$); (b) image plot representing $\mathrm{ASFR}(s,t)$ for all possible values of s and t

For data structures where we observe only one realization of a function-valued process, related modelling approaches have been discussed previously (Delicado *et al*., 2010; Nerini *et al*., 2010; Gromenko *et al*., 2012, 2013; Huang *et al*., 2009). Similarly, Hyndman and Ullah (2007) and Hyndman and Shang (2009) considered functional time series in a setting where only one realization is observed. In related applications such as mortality analysis, the decomposition into age and year has been studied by Eilers and Marx (2003), Currie *et al*. (2004, 2006) and Eilers *et al*. (2006), by using *P*-splines. The case where independently and identically distributed samples are available for random fields has been much less studied. Multilevel functional models and functional mixed effects models have been investigated by Morris and Carroll (2006), Crainiceanu *et al*. (2009), Greven *et al*. (2010) and Yuan *et al*. (2014), among others, whereas Chen and Müller (2012) developed a 'double functional principal component (FPC)' method and studied its asymptotic properties.

Our approach applies to general dimensions of both the domain of the underlying random process, with argument t, as well as of the domain of the observed functions, with argument s, and we emphasize the case of function-valued observations for stochastic processes on a one-dimensional time domain. This is the most common case and it often allows for particularly meaningful interpretations. Consider processes $X(s,t)$ with mean $\mu(s,t) = E\{X(s,t)\}$ for all $s \in \mathcal{S} \subseteq \mathbb{R}^{d_1}$ and all $t \in \mathcal{T} \subseteq \mathbb{R}^{d_2}$, and covariance function

$$C\{(s,t),(u,v)\} = E\{X(s,t)X(u,v)\} - \mu(s,t)\mu(u,v) = E\{X^{\mathrm{c}}(s,t)\,X^{\mathrm{c}}(u,v)\}, \tag{2}$$

where here and in what follows we denote the centred processes by X^{c}.

A well-established tool of functional data analysis is functional principal component analysis (FPCA) (Ramsay and Silverman, 2005) of the random process $X(s,t)$, which is based on the Karhunen–Loève expansion

$$X(s,t) = \mu(s,t) + \sum_{r=1}^{\infty} Z_r \gamma_r(s,t), \qquad s \in \mathcal{S}, \quad t \in \mathcal{T}. \tag{3}$$

Here $\{\gamma_r : r \geqslant 1\}$ is an orthonormal basis of $L^2(\mathcal{S} \times \mathcal{T})$ that consists of the eigenfunctions of the covariance operator of X, and $\{Z_r = \int \gamma_r(s,t)\,X^{\mathrm{c}}(s,t)\,\mathrm{d}s\,\mathrm{d}t : r \geqslant 1\}$ are the (random) coefficients. This expansion has the optimality property that the first K terms form the K-dimensional representation of $X(s,t)$ with the smallest unexplained variance.

A downside of the two- or higher dimensional Karhunen–Loève representation (3) is that it allows only for a joint symmetric treatment of the arguments and therefore is not suitable for analysing the separate (possibly asymmetric) effects of s and t. An additional technical drawback is that an empirical version of expression (3) requires the estimation of the covariance function C in equation (2) that depends on dimension $2(d_1 + d_2)$ and, for the case of sparse designs, this then requires performing non-parametric regression depending on at least four variables, with associated slow computing, curse of dimensionality and loss of asymptotic efficiency. Finally, Karhunen–Loève expansions for functional data depending on more than one argument are non-standard and suitable software is difficult to obtain.

Aiming to address these difficulties and with a view towards interpretability and simplicity of modelling, we propose in this paper the representation

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \xi_j(t)\,\psi_j(s) = \mu(s,t) + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \chi_{jk}\,\phi_{jk}(t)\,\psi_j(s), \tag{4}$$

where $\{\psi_j : j \geqslant 1\}$ are the eigenfunctions of the operator in $L^2(\mathcal{S})$ with kernel

$$G_{\mathcal{S}}(s,u)=\int_{\mathcal{T}} C\{(s,t),(u,t)\}\,\mathrm{d}t, \tag{5}$$

and $\{\xi_j(t): j\geqslant 1\}$ are the (random) coefficients of the expansion of the centred processes $X^{\mathrm{c}}(\cdot,t)$ in $\psi_j(s)$, and $\xi_j(t)=\Sigma_{k=1}^{\infty}\chi_{jk}\phi_{jk}(t)$ is the Karhunen–Loève expansion of the random functions $\xi_j(t)$ in $L^2(\mathcal{T})$ with eigenfunctions ϕ_{jk} and FPCs χ_{jk}.

We refer to $G_{\mathcal{S}}$ as *the marginal covariance function*, and to equation (4) as the *marginal Karhunen–Loève representation* of X that leads to the *marginal FPCA* and note that the product basis functions $\phi_{jk}(t)\psi_j(s)$ are orthogonal to each other. Hence the scores χ_{jk} can be optimally estimated by the inner product of X^{c} with the corresponding basis. Also, for each $j\geqslant 1$, we have $E(\chi_{jk}\chi_{jk'})=0$ for $k\neq k'$. In related settings, marginal covariance functions very recently have also been utilized by other researchers (Park and Staicu, 2015; Aston *et al.*, 2015). In theorem 1 below we establish the optimality of the marginal eigenfunctions ψ_j under a well-defined criterion and show in theorem 2 that the finite expansion based on the marginal FPCA approach nearly minimizes the variance between all representations of the same form.

When using representation (4), the effects of the two arguments s and t can be analysed separately, which we shall show in greater detail below in Sections 2 and 5. We also note that the estimation of the marginal representation (4) requires only estimation of covariance functions that depend on $2d_1$ or $2d_2$ real arguments. In particular, when $d_1=d_2=1$, only two-dimensional surfaces need to be estimated and marginal FPCA can be easily implemented by using standard functional data analysis packages.

Motivated by a common principal component perspective, we also introduce a simplified version of equation (4): the *product FPCA*,

$$X(s,t)=\mu(s,t)+\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\chi_{jk}\phi_k(t)\psi_j(s), \tag{6}$$

where the ϕ_k, $k\geqslant 1$, are the eigenfunctions of the marginal kernel $G_{\mathcal{T}}(s,u)$, which is analogous to $G_{\mathcal{S}}(t,v)$, with supporting theory provided by theorem 4 and theorem 5.

Sections 2 and 3 provide further details on the model and estimation. Theoretical considerations are in Section 4. In Section 5, we compare the performance of the proposed marginal FPCA, product FPCA and the conventional two-dimensional FPCA in the context of an analysis of the fertility data. Simulation results are described in Section 6 and conclusions can be found in Section 7. Detailed proofs, additional materials and the analysis of an additional human mortality data example have been relegated to the on-line supplement.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Marginal functional principal component analysis

2.1. Modelling

Consider the standard inner product $\langle f,g\rangle=\int_{\mathcal{S}}\int_{\mathcal{T}} f(s,t)\,g(s,t)\,\mathrm{d}t\,\mathrm{d}s$ in the separable Hilbert space $L^2(\mathcal{S}\times\mathcal{T})$ and the corresponding norm $\|\cdot\|$. In what follows, X is in $L^2(\mathcal{S}\times\mathcal{T})$ with mean $\mu(s,t)$. Using the covariance function $C\{(s,t),(u,v)\}$ as kernel for the Hilbert–Schmidt covariance operator $\Gamma(f)(s,t)=\int_{\mathcal{S}}\int_{\mathcal{T}} C\{(s,t),(u,v)\}\,f(u,v)\,\mathrm{d}v\,\mathrm{d}u$ with orthonormal eigenfunctions γ_r, $r\geqslant 1$, and eigenvalues $\lambda_1\geqslant\lambda_2\geqslant\ldots$ then leads to the Karhunen–Loève representation of X in expression (3), where $E(Z_r)=0$ and $\mathrm{cov}(Z_r,Z_l)=\lambda_r\delta_{rl}$, with $\delta_{rl}=1$ for $r=l$ and $\delta_{rl}=0$ otherwise; see Horváth and Kokoszka (2012) and Cuevas (2013).

Since the marginal kernel $G_{\mathcal{S}}(s,u)$ as defined in equation (5) is a continuous symmetric positive definite function (see lemma 1 in on-line supplement A), denoting its eigenvalues and eigenfunctions by τ_j and ψ_j, $j \geqslant 1$, respectively, the following representation for X emerges:

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \xi_j(t)\,\psi_j(s), \tag{7}$$

where $\xi_j(t) = \langle X(\cdot,t) - \mu(\cdot,t), \psi_j \rangle_{\mathcal{S}}$, $j \geqslant 1$, is a sequence of random functions in $L^2(\mathcal{T})$ with $E\{\xi_j(t)\} = 0$ for $t \in \mathcal{T}$, and $E(\langle \xi_j, \xi_k \rangle_{\mathcal{T}}) = \tau_j \delta_{jk}$ (see lemma 2 in on-line supplement A). Theorem 1 in Section 4 shows that the above representation has an optimality property.

The marginal Karhunen–Loève representation (7) provides new functional data, the score functions $\xi_j(t)$, which are random functions that depend on only one argument. For each $j \geqslant 1$, the ξ_j have their own covariance functions $\Theta_j(t,v) = E\{\xi_j(t)\,\xi_j(v)\}$, $t, v \in \mathcal{T}$, $j \geqslant 1$, with eigencomponents (eigenvalues and eigenfunctions) $\{\eta_{jk}, \phi_{jk}(t) : k \geqslant 1\}$. The continuity of the covariance function C implies that the $\Theta_j(t,v)$ are also continuous functions. The random functions $\xi_j(t)$ then admit their own Karhunen–Loève expansions,

$$\xi_j(t) = \sum_{k=1}^{\infty} \chi_{jk}\,\phi_{jk}(t), \qquad\qquad j \geqslant 1, \tag{8}$$

with $E(\chi_{jk}) = 0$ and $E(\chi_{jk}\chi_{jr}) = \eta_{jk}\delta_{kr}$. From equations (7) and (8) we obtain the representation for $X(s,t)$ in equation (4), $X(s,t) = \mu(s,t) + \Sigma_{j=1}^{\infty}\Sigma_{k=1}^{\infty}\chi_{jk}\phi_{jk}(t)\psi_j(s)$. As already mentioned, this expansion does not coincide with the standard Karhunen–Loève expansion of X and it is not guaranteed that χ_{jk} and χ_{lr} are uncorrelated for $j \neq l$. But the product functions $\phi_{jk}(t)\psi_j(s)$ remain orthonormal in the sense that $\int_{\mathcal{S},\mathcal{T}} \phi_{jk}(t)\psi_j(s)\phi_{lh}(t)\,\psi_l(s)\,\mathrm{d}s\,\mathrm{d}t = \delta_{jk,lh}$, where $\delta_{jk,lh} = 1$ when $j = l$ and $k = h$; $\delta_{jk,lh} = 0$ otherwise.

2.2. Estimating procedures

Time- or space-indexed functional data consist of a sample of n independent subjects or units. For the ith subject, $i = 1, \ldots, n$, random functions $X_i(\cdot,t)$ are recorded at a series of time points t_{im}, $m = 1, \ldots, M_i$. Ordinarily, these functions are not continuously observed but instead are observed at a grid of functional design points s_l, $l = 1, \ldots, L$. In this paper we focus on the case where the grid of s is dense, regular and the same across all subjects. The case of sparse designs in s will be discussed in Section 7. Our proposed marginal FPCA procedure consists of three main steps.

Step 1: centre the data to obtain $\hat{X}_i^{\mathrm{c}}(s,t) = X_i(s,t) - \hat{\mu}(s,t)$. Obtain an estimator of $\mu(s,t)$ by pooling all the data. If the recording points t are densely and regularly spaced, i.e. $t_{im} = t_m$, an *empirical estimator* by averaging over n subjects and interpolating between design points can be used. This scheme is also applicable to dense irregular designs by adding a presmoothing step and sampling smoothed functions at a dense regular grid. Alternatively, one can recover the mean function μ by smoothing the pooled data (Yao *et al.*, 2005), for example with a local linear smoother, obtaining a *smoothing estimator* $\hat{\mu}(s,t) = \hat{a}_0$, where

$$\begin{aligned}(\hat{a}_0, \hat{a}_1, \hat{a}_2) = \arg\min \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{M_i} \sum_{l=1}^{L_{im}} [\{X_i(t_{im}, s_{iml}) - a_0 - a_1(s_{iml} - s) - a_2(t_{im} - t)\}^2 \\ \times K_{h_s}(s_{iml} - s) K_{h_t}(t_{im} - t)].\end{aligned} \tag{9}$$

Step 2: use the centred data $\hat{X}_i^{\mathrm{c}}(s,t)$ from step 1 to obtain estimates of the marginal covariance function $G_{\mathcal{S}}(s,u)$ as defined in equation (5), its eigenfunctions $\psi_j(s)$ and the corresponding

FPC score functions $\xi_{i,j}(t)$. For this, we pool the data $\{\hat{X}_i^c(\cdot, t_{im}), i = 1, \ldots, n, m = 1, \ldots, M_i\}$ and obtain estimates

$$\hat{G}_{\mathcal{S}}(s_j, s_l) = \frac{|\mathcal{T}|}{\sum_{i=1}^{n} M_i} \sum_{i=1}^{n} \sum_{m=1}^{M_i} \hat{X}_i^c(s_j, t_{im}) \hat{X}_i^c(s_l, t_{im}), \tag{10}$$

where $1 \leqslant j \leqslant l \leqslant L$ and $|\mathcal{T}|$ is the Lebesgue measure of $\mathcal{T}$, followed by interpolating between grid points to obtain $\hat{G}_{\mathcal{S}}(s, u)$ for $(s, u) \in \mathcal{S} \times \mathcal{S}$. We then obtain the eigenfunctions $\hat{\psi}_j$ and eigenvalues $\hat{\tau}_j$ by standard methods (Yao *et al.*, 2005) as implemented in the PACE package (`http://www.stat.ucdavis.edu/PACE`) or as in Kneip and Utikal (2001), and the FPC function estimates $\hat{\xi}_{i,j}(t)$ by interpolating numerical approximations of the integrals $\hat{\xi}_{i,j}(t_{im}) = \int \hat{X}_i^c(s, t_{im}) \hat{\psi}_j(s) \mathrm{d}s$. Theorem 3 shows that $\hat{G}_{\mathcal{S}}$ in equation (10) and $\hat{\psi}_j$ are consistent estimates of the marginal covariance function $G_{\mathcal{S}}$ and its eigenfunctions and that estimates $\{\hat{\xi}_{i,j}(t), i = 1, \ldots, n, \}$ converge uniformly to the target processes $\{\xi_{i,j}(t), j \geqslant 1\}$.

Step 3: this is a standard FPCA of one-dimensional processes $\{\xi_{i,j}(t), j \geqslant 1\}$, where, for each fixed j, one obtains estimates for the FPCs χ_{jk} and eigenfunctions $\{\phi_{jk}(t) : k \geqslant 1\}$; see for example Ramsay and Silverman (2005) and Kneip and Utikal (2001) for designs that are dense in t and Yao *et al.* (2005) for designs that are sparse in t.

After selecting appropriate numbers of included components P and K_j, $j = 1, \ldots, P$, we obtain the overall representation

$$\hat{X}_i(s, t) = \hat{\mu}(s, t) + \sum_{j=1}^{P} \hat{\xi}_{i,j}(t) \hat{\psi}_j(s) = \hat{\mu}(s, t) + \sum_{j=1}^{P} \sum_{k=1}^{K_j} \hat{\chi}_{i,jk} \hat{\phi}_{jk}(t) \hat{\psi}_j(s). \tag{11}$$

The number of components P included can be selected via a fraction of variance explained (FVE) criterion, finding the smallest P such that $\Sigma_{j=1}^{P} \hat{\tau}_j / \Sigma_{j=1}^{M} \hat{\tau}_j \geqslant 1 - p$, where M is large and we choose $p = 0.15$ in our application. The number of components K_j included can be determined by a second application of the FVE criterion, where the variance V_{jk} explained by each term (j, k) is defined as

$$V_{jk} = \frac{(1/n) \sum_{i=1}^{n} \hat{\chi}_{i,jk}^2}{(1/n) \sum_{i=1}^{n} \|X(s, t) - \hat{\mu}(s, t)\|_{\mathcal{S} \times \mathcal{T}}^2}. \tag{12}$$

Note that V_{jk} does not depend on the choice of P in the first step, since it is the fraction of total variance explained. Here the total variance explained, $\Sigma_{k=1}^{K_j} \Sigma_{j=1}^{P} E(\chi_{jk}^2)$, cannot exceed the variance explained in the first step, $\Sigma_{j=1}^{P} \tau_j$.

We shall illustrate these procedures in Section 5. Since the functions $\psi_j(s) \phi_{jk}(t)$ are orthogonal, the unexplained variance $E\|X^c\|^2 - \Sigma_{j=1}^{P} \Sigma_{k=1}^{K_j} E(\chi_{jk}^2)$ and the reconstruction loss

$$E\left[\int_{\mathcal{S},\mathcal{T}} \{X^c(s, t) - \sum_{j=1}^{P} \sum_{k=1}^{K_j} \langle X^c, \phi_{jk} \psi_j \rangle \phi_{jk}(t) \psi_j(s)\}^2 \mathrm{d}s \mathrm{d}t \right]$$

are equivalent.

3. Product functional principal component analysis

In this section we discuss a simplified version of the marginal Karhunen–Loève representation

(4). A simplifying assumption is that the eigenfunctions ϕ_{jk} in the Karhunen–Loève expansion of $\xi_j(t)$ in expression (4) do not depend on j. This assumption leads to a more compact representation of X as given in equation (6), $X(s,t) = \mu(s,t) + \Sigma_{j=1}^{\infty}\Sigma_{k=1}^{\infty}\chi_{jk}\phi_k(t)\psi_j(s)$.

To study the properties of this specific product representation, we consider product representations with general orthogonal basis $X(s,t) = \mu(s,t) + \Sigma_{j=1}^{\infty}\Sigma_{k=1}^{\infty}\chi_{jk}\, f_k(t)\, g_j(s)$, where $\chi_{jk} = \langle X^{\mathrm{c}} f_k g_j\rangle$. For such general representations, the assumption

$$\begin{aligned} \operatorname{cov}(\chi_{jk}, \chi_{jl}) &= 0 \qquad \text{for } k \neq l, \\ \operatorname{cov}(\chi_{jk}, \chi_{hk}) &= 0 \qquad \text{for } j \neq h \end{aligned} \tag{13}$$

implies that the covariance kernel induced by $\xi_j(t) = \langle X^{\mathrm{c}}(t,\cdot), g_j\rangle_{\mathcal{S}}$ has common eigenfunctions $\{f_k(t), k \geqslant 1\}$, not depending on j, and the covariance kernel induced by $\xi_k(s) = \langle X^{\mathrm{c}}(\cdot, s), f_k\rangle_{\mathcal{T}}$ has common eigenfunctions $\{g_j(s), j \geqslant 1\}$, not depending on k. Therefore we refer to expression (13) as the *common principal component assumption*. We prove in theorem 4 below that if there are bases $\{g_j(s), j \geqslant 1\}$ and $\{f_k(t), k \geqslant 1\}$ such that assumption (13) is satisfied, then $g_j \equiv \psi_j$ and $f_k \equiv \phi_k$, the eigenfunctions of the marginal covariance $G_{\mathcal{S}}(s,u)$ and $G_{\mathcal{T}}(t,v)$ respectively, where $G_{\mathcal{T}}(t,v)$ is defined as

$$G_{\mathcal{T}}(t,v) = \int_{\mathcal{S}} C\{(s,t),(s,v)\}\, \mathrm{d}s, \qquad t, v \in \mathcal{T}. \tag{14}$$

Even without invoking assumption (13), in theorem 5 we show that the finite expansion based on the marginal eigenfunctions ϕ_k and ψ_j yields a nearly optimal solution in terms of minimizing the unexplained variance between all possible product expansions. This result provides additional theoretical support for the use of *product FPCA* based on the marginal kernels $G_{\mathcal{S}}$ and $G_{\mathcal{T}}$ under fairly general situations. Although the product functions $\phi_k(t)\psi_j(s)$ are orthonormal, without additional conditions, the scores χ_{jk} in general will not be uncorrelated. Product FPCA (6) is well suited for situations where the two arguments of $X(s,t)$ play symmetric roles. This simplified model retains substantial flexibility, as we shall demonstrate in the application to fertility data (see on-line supplement C).

The estimation procedures for this model are analogous to those described in the previous section. This also applies to the theoretical analysis of these estimates and their asymptotic properties. A straightforward approach to estimate the eigenfunctions appearing in equation (6) is to apply the estimation procedure that was described in Section 2.2 twice, first following the description there to obtain estimates of $G_{\mathcal{S}}$ and ψ_j and then changing the roles of the two arguments in a second step to obtain estimates of $G_{\mathcal{T}}$ and ϕ_k.

4. Theoretical properties

Detailed proofs of the results in this section are in on-line supplement A. We show that the optimal finite dimensional approximation property of FPCA extends to the proposed marginal FPCA under well-defined criteria. Theorem 1 establishes the optimality of the basis functions ψ_j, i.e. the eigenfunctions in equation (4) derived from the marginal covariance in equation (5). Theorem 2 shows the near optimality of the marginal representation (4), based on the eigenfunctions ϕ_{jk} and ψ_j, in terms of minimizing the unexplained variance between all functional expansions of the same form.

Theorem 1. For each $P \geqslant 1$ for which $\tau_P > 0$, the functions $g_1, \ldots, g_P$ in $L^2(\mathcal{S})$ that provide the best finite dimensional approximations in the sense of minimizing

$$E\left\{\int_{\mathcal{T}} \|X^{\mathrm{c}}(\cdot,t) - \sum_{j=1}^{P} \langle X^{\mathrm{c}}(\cdot,t), g_j\rangle_{\mathcal{S}} g_j\|_{\mathcal{S}}^2 \,\mathrm{d}t\right\}$$

are $g_j = \psi_j$, $j = 1, \ldots, P$, i.e. the eigenfunctions of $G_{\mathcal{S}}$. The minimizing value is $\Sigma_{j=P+1}^{\infty} \tau_j$.

Theorem 2. For $P \geqslant 1$ and $K_j \geqslant 1$, consider the loss minimization

$$\min_{f_{jk}, g_j} E\left[\int_{\mathcal{S},T} \{X^{\mathrm{c}}(s,t) - \sum_{j=1}^{P} \sum_{k=1}^{K_j} \langle X^{\mathrm{c}}, f_{jk} g_j\rangle\, f_{jk}(t)\, g_j(s)\}^2 \mathrm{d}s\mathrm{d}t\right],$$

with minimizing value Q^*, where the $g_j(s)$, $j \geqslant 1$, are orthogonal and, for each j, the $f_{jk}(t), k \geqslant 1$, are orthogonal. The marginal eigenfunctions $\psi_j(s)$ and $\phi_{jk}(t)$ achieve good approximation in the sense that

$$E\left[\int_{\mathcal{S},T} \{X^{\mathrm{c}}(s,t) - \sum_{j=1}^{P} \sum_{k=1}^{K_j} \langle X^{\mathrm{c}}, \phi_{jk}\psi_j\rangle\, \phi_{jk}(t)\, \psi_j(s)\}^2 \mathrm{d}s\,\mathrm{d}t\right] < Q^* + a\, E\|X^{\mathrm{c}}\|^2,$$

where $a = \max_{1 \leqslant j \leqslant P} a_j$, with $1 - a_j$ denoting the fraction of variance explained by K_j terms for each process $\xi_j(t) = \langle X^{\mathrm{c}}(\cdot,t)\psi_j\rangle_{\mathcal{S}}$.

In what follows, $\|G_{\mathcal{S}}(s,u)\|_{\mathcal{S}} = \{\int_{\mathcal{S}}\int_{\mathcal{S}} G_{\mathcal{S}}(s,u)^2 \mathrm{d}s\mathrm{d}u\}^{1/2}$ is the Hilbert–Schmidt norm and $a \asymp b$ denotes that a and b are of the same order asymptotically. For the consistency of marginal FPCA (4) it is important that the covariance kernel $G_{\mathcal{S}}$ and its eigenfunctions ψ_j and eigenvalues τ_j can be consistently estimated from the data. Uniform convergence of the empirical working processes $\{\hat{\xi}_{i,j}(t_{im}), 1 \leqslant i \leqslant n, 1 \leqslant m \leqslant M_i\}$ to the target processes $\{\xi_{i,j}(t), t \in \mathcal{T}\}$ then guarantees the consistency of the estimates of the eigenfunctions ϕ_{jk} and the eigenvalues η_{jk} (Yao and Lee, 2006).

The following assumptions are needed to establish these results. We use $0 < B < \infty$ as a generic constant that can take different values at different places.

Assumption 1. $\sup_{s,t} |\mu(s,t)| < B$ and $\sup_s |\psi_j(s)| < B$ for all $1 \leqslant j \leqslant P$.

Assumption 2. $E\{\sup_{s,t} |X(s,t)|\} < B$ and $\sup_{s,t} E|X(s,t)|^4 < B$.

Assumption 3. $\sup_{(s,u)\in\mathcal{S}^2,(t_1,t_2)\in\mathcal{T}^2} |C\{(s,t_1),(u,t_1)\} - C\{(s,t_2),(u,t_2)\}| < B|t_1 - t_2|$.

Assumption 4. $\sup_{(s_1,u_1,s_2,u_2)\in\mathcal{S}^4} |G_{\mathcal{S}}(s_1,u_1) - G_{\mathcal{S}}(s_2,u_2)| < B(|s_1 - s_2| + |u_1 - u_2|)$.

Assumption 5. For all $1 \leqslant j \leqslant P$, $\delta_j > 0$, where $\delta_j = \min_{1 \leqslant l \leqslant j}(\tau_l - \tau_{l+1})$.

Assumption 6.

(a) The grid points $\{t_{im} : m = 1, \ldots, M\}$ are equidistant, and $n/M = O(1)$.
(b) The grid points $\{t_{im} : m = 1, \ldots, M_i\}$ are independently and identically distributed with uniform density, and $\min_i M_i \asymp \max_i M_i$.

Condition 1 generally holds for smooth functions that are defined on finite domains. Conditions 2 are commonly used moment conditions for $X(s,t)$. Conditions 3 and 4 are Lipschitz conditions for the joint covariance C and the marginal covariance $G_{\mathcal{S}}$ and quantify the smoothness of these covariance surfaces. Condition 5 requires non-zero eigengaps for the first P leading components and has been widely adopted in the literature (Hall *et al.*, 2006; Li and Hsing, 2010).

Condition 6, parts (a) and (b), correspond to two alternative scenarios for the design at which the underlying random process is sampled over t. Here condition 6, part (a), reflects the case of a *dense regular* design, where we observe functions $X(\cdot, t_m)$ at a dense and regular grid of $\{t_m : m = 1, \ldots, M\}$, with $n/M = O(1)$, whereas part (b) corresponds to the case of a random design, where we observe functions $X(\cdot, t_{im})$ at a series of random locations corresponding to the time points $\{t_{im} : m = 1, \ldots, M_i\}$, where the number of available measurements M_i may vary across subjects.

Theorem 3. If conditions 1–5 and 6, part (a), or 1–5 and 6, part (b), hold, $\max(s_l - s_{l-1}) = O(n^{-1})$, and $\hat{\mu}(s,t)$ obtained in step 1 above satisfies $\sup_{s,t}|\hat{\mu}(s,t) - \mu(s,t)| = O_p[\{\log(n)/n\}^{1/2}]$, we have the following results for $1 \leqslant j \leqslant P$:

$$\|\hat{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)\|_{\mathcal{S}} = O_p[\{\log(n)/n\}^{1/2}], \tag{15}$$

$$|\hat{\tau}_j - \tau_j| = O_p[\{\log(n)/n\}^{1/2}], \tag{16}$$

$$\|\hat{\psi}_j(s) - \psi_j(s)\|_{\mathcal{S}} = O_p[\{\log(n)/n\}^{1/2}], \tag{17}$$

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{1 \leqslant m \leqslant M_i} |\hat{\xi}_{i,j}(t_{im}) - \xi_{i,j}(t_{im})| = O_p[\{\log(n)/n\}^{1/2}]. \tag{18}$$

The *empirical estimator* and the *smoothing estimator* that are discussed in step 1 both satisfy $\sup_{s,t} |\hat{\mu}(s,t) - \mu(s,t)| = O_p[\{\log(n)/n\}^{1/2}]$ under appropriate conditions and an appropriate choice of the bandwidth in the *smoothing estimator*. We refer to Chen and Müller (2012), theorems 1 and 2 for detailed conditions and proofs. The following result establishes the uniqueness of the product representation with marginal eigenfunctions ψ_j and ϕ_j derived from expressions (5) and (14) under the common principal component assumption (13). An important implication of theorem 4 is that the product FPCA based on marginal eigenfunctions is optimal if the eigenfunctions of kernel $C(s,t;u,v)$ indeed can be written as products in their arguments.

Theorem 4. If there are orthogonal bases $\{g_j(s), j \geqslant 1\}$ and $\{f_k(t), k \geqslant 1\}$, under which the common principal component assumption (13) is satisfied, we have $g_j(s) \equiv \psi_j(s)$ and $f_k(t) \equiv \phi_k(t)$, with

$$G_{\mathcal{S}}(s,u) = \sum_{j=1}^{\infty} \tau_j \psi_j(s)\psi_j(u), \qquad \text{for all } s,u \in \mathcal{S}, \tag{19}$$

$$G_{\mathcal{T}}(t,v) = \sum_{k=1}^{\infty} \vartheta_k \phi_k(t)\phi_k(v), \qquad \text{for all } t,v \in \mathcal{T}, \tag{20}$$

where

$$\begin{aligned} \tau_j &= \sum_{k=1}^{\infty} \operatorname{var}(\chi_{jk}), \\ \vartheta_k &= \sum_{j=1}^{\infty} \operatorname{var}(\chi_{jk}), \\ \chi_{jk} &= \int_{\mathcal{S}}\int_{\mathcal{T}} \{X(s,t) - \mu(s,t)\}\psi_j(s)\phi_k(t)\,\mathrm{d}t\,\mathrm{d}s, \\ E(\chi_{jk}) &= 0, \\ \operatorname{cov}(\chi_{jk}, \chi_{jl}) &= \operatorname{var}(\chi_{jk})\delta_{kl}, \\ \operatorname{cov}(\chi_{jk}, \chi_{hk}) &= \operatorname{var}(\chi_{jk})\delta_{jh}. \end{aligned} \tag{21}$$

Theorem 5. For $P \geqslant 1$ and $K \geqslant 1$, consider the loss minimization

$$\min_{f_k, g_j} E\left[\int_{\mathcal{S},\mathcal{T}} \{X^{\mathrm{c}}(s,t) - \sum_{j=1}^{P} \sum_{k=1}^{K} \langle X^{\mathrm{c}}, f_k g_j\rangle f_k(t) g_j(s)\}^2 \mathrm{d}s\mathrm{d}t\right],$$

with minimizing value Q^*, where f_k, $k \geqslant 1$, are orthogonal, and g_j, $j \geqslant 1$, are orthogonal. The marginal eigenfunctions $\psi_j(s)$ of $G_{\mathcal{S}}(s,u)$ and $\phi_k(t)$ of $G_{\mathcal{T}}(t,v)$ achieve good approximation in the sense that

$$E\left[\int_{\mathcal{S},\mathcal{T}} \{X^{\mathrm{c}}(s,t) - \sum_{j=1}^{P} \sum_{k=1}^{K} \langle X^{\mathrm{c}}, \phi_k\psi_j\rangle \phi_k(t)\psi_j(s)\}^2 \mathrm{d}s\mathrm{d}t\right] < Q^* + a E\|X^{\mathrm{c}}\|^2,$$

where $a = \min(a_{\mathcal{T}}, a_{\mathcal{S}})$, with $1 - a_{\mathcal{T}}$ denoting the fraction of variance explained by K terms for $G_{\mathcal{T}}(t,v)$ and analogously for $a_{\mathcal{S}}$.

Similarly to the situation in theorem 2, the error term $a E\|X^{\mathrm{c}}\|^2$ depends on the loss involved in truncating just the (marginal) principal component decompositions, which also imposes a lower bound on Q^*.

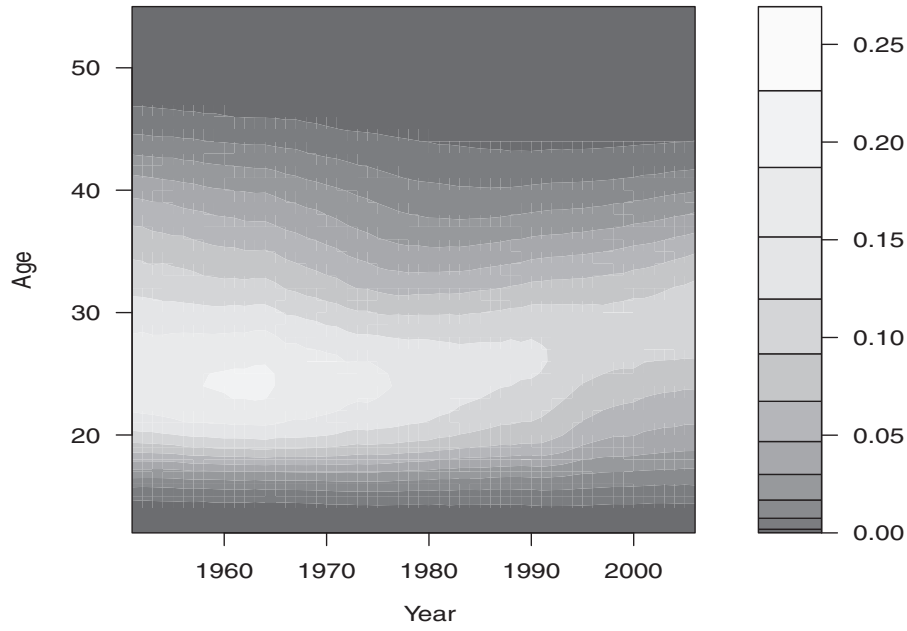
5. Functional data analysis of fertility

Human fertility naturally plays a central role in demography (Preston *et al.*, 2001) and its analysis recently has garnered much interest due to declining birth rates in many developed countries and associated subreplacement fertilities (Takahashi, 2004; Ezeh *et al.*, 2012). The Human Fertility Database (Human Fertility Database, 2013) contains detailed period and cohort fertility annual data for 22 countries (plus five subdivisions: two for Germany and three for the UK). We are interested in *age-specific fertility rates* ASFR, considered as functions of women's age in years, s, and repeatedly measured for each calendar year t for various countries. These rates (see equation (1)) constitute the functional data $X(s,t) = \mathrm{ASFR}(s,t)$.

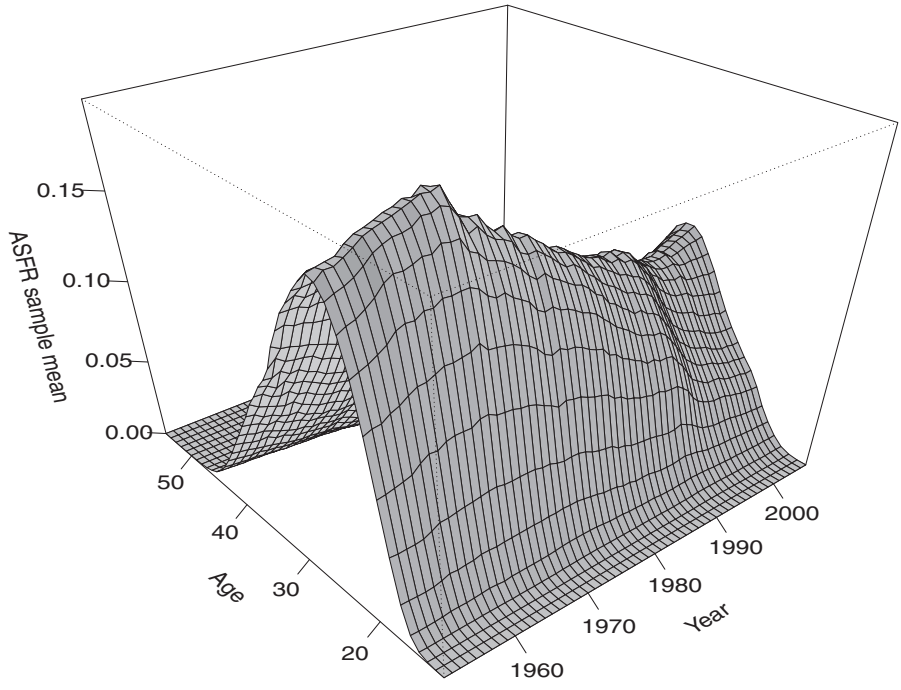
A detailed description how ASFR is calculated from raw demographic data can be found in the Human Fertility Database methods protocol (Jasilioniene *et al.*, 2012). The specific definition of ASFR that we are using corresponds to period fertility rates by calendar year and age (Lexis squares; age in completed years). In Human Fertility Database (2013), $\mathrm{ASFR}(s,t)$ is included for mothers of ages $s = 12$–55 years; thus the domain $\mathcal{S}$ is an interval of length $L = 44$ years. The interval of calendar years with available ASFR varies by country. Aiming at a compromise between the length M of the studied period $\mathcal{T}$ and the number n of countries that can be included, we choose $\mathcal{T}$ as the interval from 1951 to 2006. There are $n = 17$ countries (or territories) with available ASFR-data during this time interval (see Table 4 and Fig. 5 in on-line supplement B for the list of $n = 17$ countries included and heat maps depicting individual functions ASFR_i).

The sample means $\overline{\mathrm{ASFR}}(s,t)$ of the ASFR-functions for 17 countries that are displayed in Fig. 2 shows that fertility rates are, on average, highest for women aged between 20 and 30 years and are decreasing with increasing calendar year; this overall decline is interspersed with two periods of increasing fertility before 1965, corresponding to the baby boom, and after 1995 with a narrow increase for ages between 30 and 40 years, is narrowing in terms of the age range with high fertility and displays an increase in regard to the ages of women where maximum fertility occurs. We applied marginal FPCA, product FPCA and two-dimensional FPCA to quantify the variability across individual countries and summarize the main results here. Additional details can be found in on-line supplement C.

The fertility data include one fertility curve over age per calendar year and per country and



(a)



(b)

Fig. 2. Sample means of the 17 fertility rate functions by calendar year

are observed on a regular grid spaced in years across both co-ordinates age s and calendar year t, which means that the empirical estimators that were described in Section 2 can be applied to these data. Fig. 6 in on-line supplement B displays the $nM = 952$ centred functional data $\mathrm{ASFR}_i^c(s_l, t_m) = \mathrm{ASFR}_i(s_l, t_m) - \overline{\mathrm{ASFR}}(s_l, t_m)$, for $l = 1, \ldots, L = 44$, $m = 1, \ldots, M = 56$ and $i = 1, \ldots, n = 17$, demonstrating that there is substantial variation across countries and calendar years. The results of the proposed marginal FPCA are summarized in Figs 3 and 4 for the first three eigenfunctions $\hat{\psi}_j(s)$, $j = 1, 2, 3$, resulting in an FVE of 95.8%. From Fig. 3, the first eigenfunction $\hat{\psi}_1(s)$ can be interpreted as a contrast between fertility before and after the age of 25 years, representing the direction from mature fertility (negative scores) to young fertility (positive scores).

The second eigenfunction $\hat{\psi}_2(s)$ takes positive values for all ages s, with a maximum at age $s = 24$ years. The shape of $\hat{\psi}_2(s)$ is similar to that of the mean function $\overline{\mathrm{ASFR}}(s, t)$ for a fixed year t (see Fig. 2(b)). Therefore $\hat{\psi}_2(s)$ can be interpreted as a *size* component: country-years with positive score in the direction of this eigenfunction have higher fertility ratios than the mean function for all ages. The third eigenfunction $\hat{\psi}_3(s)$ represents a direction from more concentrated fertility around the age of 25 years to a more dispersed age distribution of fertility.

Examining the score functions $\hat{\xi}_{i,j}(t)$, $t \in \mathcal{T}$, which are country-specific functions of calendar year, we find from Fig. 3 for $\hat{\xi}_{i,1}(t)$ that there are countries, such as the USA, Bulgaria or Slovakia, for which $\xi_1(t)$ is positive for all calendar years t, which implies that these countries always have higher fertility rates for young women and vice versa for mature women, relative to the mean function. Countries from eastern Europe such as Bulgaria, the Czech Republic, Hungary and Slovakia have high scores until the end of the 1980s when there is a sudden decline, implying that the relationship of fertility between younger and more mature women has reversed for these countries. Also notable is a declining trend in the dispersion of the score functions since 1990, implying that the fertility patterns of the 17 countries are converging.

The score functions $\hat{\xi}_{i,2}(t)$ corresponding to the size component indicate that Canada and the USA had a particularly strong baby boom in the 1960s, whereas Portugal and Spain had later baby booms during the 1970s. In contrast, Hungary had a period of relatively low fertility during the 1960s. Again, the dispersion of these size score functions declines towards 2006. The patterns of the score functions $\hat{\xi}_{i,3}(t_h)$ indicate that Japan has by far the largest degree of concentrated fertility at ages from 22 to 29 years, from 1960–1980, but lost this exceptional status in the 1990s and beyond. There is also a local anomaly for Japan in 1966. Takahashi (2004) reported that in 1966 the total fertility in Japan declined to the lowest value ever recorded, because 1966 was the year of the *Hinoe-Uma* ('fire horse', which is a calendar event that occurs every 60 years), associated with the superstitious belief of bad luck for girls who are born in such years.

Trends over calendar time for particular countries can be visualized by *track plots*, which depict the changing vectors of score functions $(\xi_{i,1}(t), \ldots, \xi_{i,K}(t))$, parameterized in $t \in \mathcal{T}$, as one-dimensional curves in $\mathcal{R}^K$. Track plots are most useful for pairs of score functions and are shown in the form of planar curves for the pairs $(\xi_{i,1}(t), \xi_{i,2}(t))$ and $(\xi_{i,1}(t), \xi_{i,3}(t))$, $t \in \mathcal{T}$, in Fig. 4 for selected countries and in Fig. 7 in on-line supplement B for all countries. Fig. 4(a) with the track plot illustrating the evolution in calendar time of first and second FPCs shows predominantly vertical movements: From 1951 to 2006 for most countries there are more changes in total fertility than changes in the distribution of fertility over the different ages of mothers. Exceptions to this are Portugal, Spain, the Czech Republic, and the USA, with considerable variation over the years in the first FPC score. There was more variation in fertility patterns between the countries that were included in this analysis in 1951 than in 2006, indicating a 'globalization' of fertility patterns. In the track plot corresponding to the first and third eigenfunctions in Fig. 4(b), the anomalous behaviour of Japan stands out. The third step

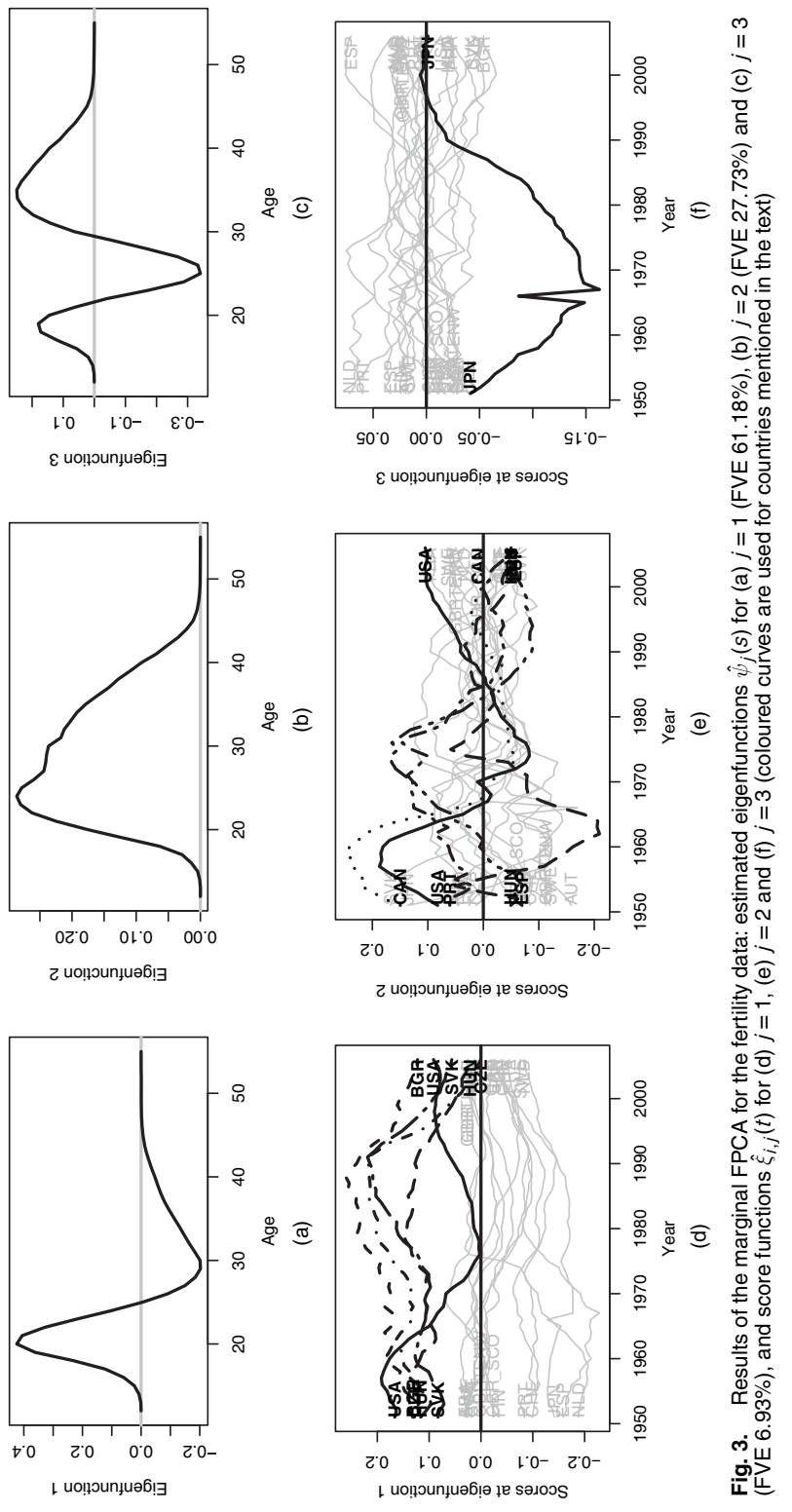
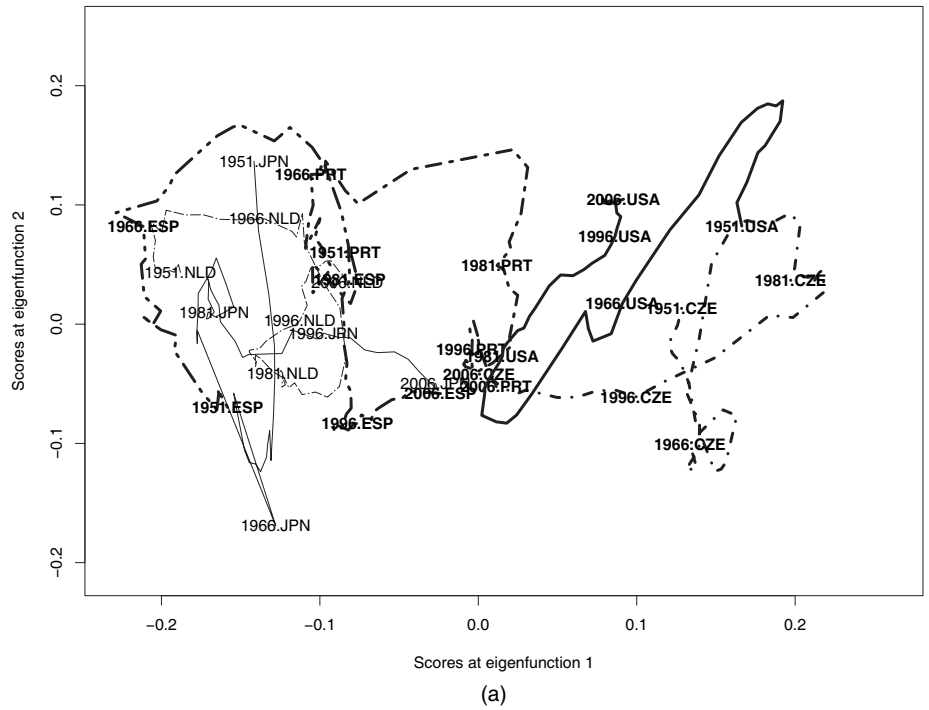
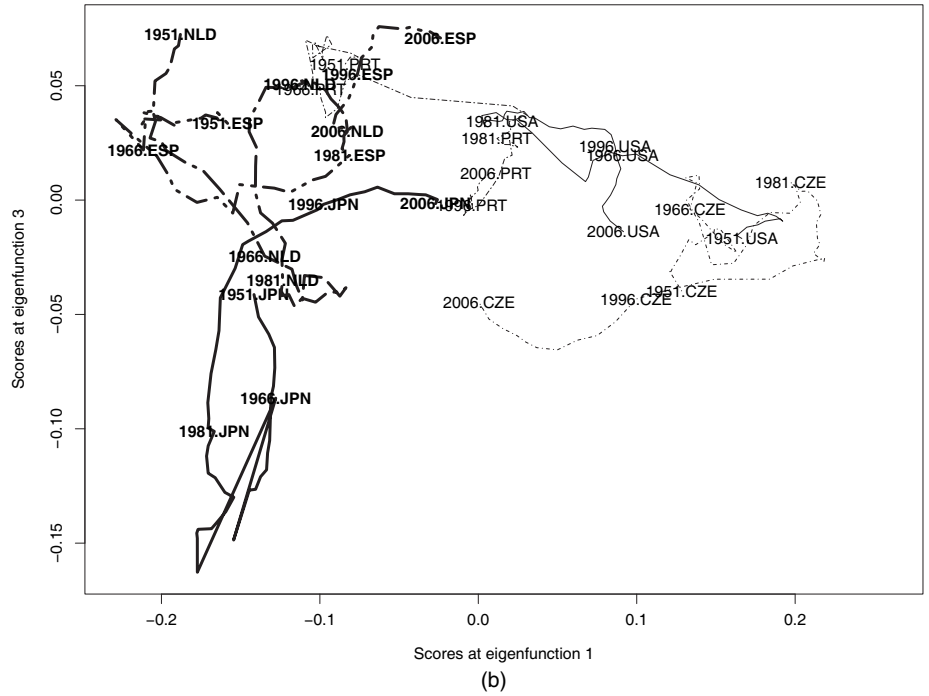


Fig. 3. Results of the marginal FPCA for the fertility data: estimated eigenfunctions $\hat{\psi}_j(s)$ for (a) $j = 1$ (FVE 61.18%), (b) $j = 2$ (FVE 27.73%) and (c) $j = 3$ (FVE 6.93%), and score functions $\hat{\xi}_{i,j}(t)$ for (d) $j = 1$, (e) $j = 2$ and (f) $j = 3$ (coloured curves are used for countries mentioned in the text)



(a)



(b)

Fig. 4. Track plots (a) $\{(\hat{\xi}_{i1}(t), \hat{\xi}_{i2}(t)) : t = 1951, \ldots, 2006\}$ and (b) $\{(\hat{\xi}_{i,1}(t), \hat{\xi}_{i3}(t)) : t = 1951, \ldots, 2006\}$, indexed by calendar time t, where $\xi_{i,j}(t)$ is the jth score function for country i (for selected countries) as in equation (4)

Table 1. FVE of ASFR(s, t) for the leading terms in the proposed marginal FPCA, product FPCA and two-dimensional FPCA†

Marginal FPCA	*FVE (%)*	*Product FPCA*	*FVE (%)*	*2nd FPCA*	*FVE (%)*
Six terms	87.49	Seven terms	87.38	Four terms	89.73
$\hat{\phi}_{11}(t)\,\hat{\psi}_1(s)$	54.33	$\hat{\phi}_1(t)\,\hat{\psi}_1(s)$	53.69	$\hat{\gamma}_1(s,t)$	58.93
$\hat{\phi}_{21}(t)\,\hat{\psi}_2(s)$	13.04	$\hat{\phi}_2(t)\,\hat{\psi}_2(s)$	8.10	$\hat{\gamma}_2(s,t)$	13.71
$\hat{\phi}_{22}(t)\,\hat{\psi}_2(s)$	6.88	$\hat{\phi}_1(t)\,\hat{\psi}_2(s)$	8.08	$\hat{\gamma}_3(s,t)$	11.04
$\hat{\phi}_{12}(t)\,\hat{\psi}_1(s)$	4.62	$\hat{\phi}_3(t)\,\hat{\psi}_2(s)$	5.51	$\hat{\gamma}_4(s,t)$	6.05
$\hat{\phi}_{23}(t)\,\hat{\psi}_2(s)$	4.40	$\hat{\phi}_2(t)\,\hat{\psi}_1(s)$	4.47		
$\hat{\phi}_{31}(t)\,\hat{\psi}_3(s)$	4.22	$\hat{\phi}_4(t)\,\hat{\psi}_2(s)$	3.85		
		$\hat{\phi}_1(t)\,\hat{\psi}_3(s)$	3.68		

†The number of terms in each case is selected to achieve an FVE of more than 85%.

of the marginal FPCA that was described in Section 2 consists of performing a separate FPCA for the estimated score functions $\hat{\xi}_{i,j}(t)$, $i = 1, \ldots, n$, for $j = 1, 2, 3$, with estimated eigenfunctions $\hat{\phi}_{jk}$ shown in Fig. 8 in on-line supplement B. The interpretation of these eigenfunctions is relative to the shape of the $\psi_j(s)$.

The results in Table 1 for estimated representations (11) justify the inclusion of only the six terms with the highest FVE in the final model, leading to a cumulative FVE of 87.49%, where the FVE for each term (j, k) is estimated by equation (12). The corresponding six product functions $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$ are shown in Fig. 9 in on-line supplement B. Regarding the comparative performance of standard two-dimensional FPCA, product FPCA (with detailed results in on-line supplement C) and marginal FPCA, we have the following findings.

(a) As expected, standard FPCA based on the two-dimensional Karhunen–Loève expansion requires fewer components to explain a given amount of variance, as four eigenfunctions lead to an FVE of 89.73% (see Table 1), whereas the marginal FPCA representation achieves an FVE of 87.49% with six terms, and product FPCA needs seven terms to explain 87.38%.
(b) Product FPCA and marginal FPCA represent the functional data as a sum of terms that are products of two functions, each depending on only one argument. This enables much better interpretability and makes it possible to discover patterns in functional data that are not found when using standard FPCA. For instance, the second eigenfunction ψ_2 in the first step of the marginal FPCA could be characterized as a *fertility size* component, with a country-specific time varying multiplier $\xi_2(t)$. Standard FPCA does not pinpoint this feature, which is an essential characteristic of demographic changes in fertility.
(c) Marginal FPCA makes it much easier than standard FPCA to analyse the time dynamics of the fertility process.

Specifically, the plots in Figs 3(d)–3(f) or the track plots in Fig. 4 are informative about the fertility evolution over calendar years.

(a) The relative balance between young and mature fertility at each country changes over the years. The graphical representation of functional score functions $\hat{\xi}_{i,1}(t)$ allows us to characterize and quantify this phenomenon.
(b) The track plot in Fig. 4(a) indicates that in general it is much more common that fertility

rates rise or decline across all ages compared with transfers of fertility between different age groups.

(c) The fertility patterns of the various countries are much more similar in 2006 than in 1951.

All three approaches to FPCA for function-valued stochastic processes, namely standard FPCA (3), marginal FPCA (4) and product FPCA (6), can be used to produce country scores which can be plotted against each other. They turn out to be similar for these approaches; as an example the standard FPCA scores are shown in Fig. 12 in on-line supplement C. We conclude that standard FPCA, marginal FPCA and product FPCA complement each other. Our recommendation is to perform all whenever feasible, to gain as much insight about complex functional data as possible.

6. Simulations

We conducted two simulation studies: one to investigate the estimation procedure for marginal FPCA, and a second study to evaluate the performance of product FPCA. Both were conducted in a scenario that mimics the fertility data. For simulation 1, we generated data following a truncated version of equation (4), where we used the estimated mean function $\overline{\mathrm{ASFR}}(s,t)$ from the country fertility data (Section 5) as mean function and the estimated product functions $\hat{\phi}_{jk}(t)\,\hat{\psi}_j(s)$, $1 \leqslant j,k \leqslant 4$, as base functions in equation (4). Random scores χ_{jk} were generated as independent normal random variables with variances λ_{jk}, corresponding to the estimates derived from the fertility data, $\lambda_{jk} = (1/n)\,\Sigma_{i=1}^{n} \hat{\chi}_{i,jk}^2$. We also added independently and identically distributed noise to the actual observations $Y_i(s_l,t_h) = X(s_l,t_h) + \varepsilon_{i,lh}, l = 1,\ldots,44, h = 1,\ldots,56$, where $\varepsilon_{i,lh} \sim N(0,\sigma^2)$ with $\sigma = 0.005$ to mimic the noise level of the fertility data.

Estimated and true functions $\psi_j(s)$ and $\phi_{jk}(t)$ obtained for one sample run with $n = 50$ are shown in Fig. 10 in on-line supplement B, demonstrating very good recovery of the true basis functions. To quantify the quality of the estimates of $\mu(s,t)$, we use the relative squared error

$$\mathrm{RE} = \frac{\|\mu(s,t) - \hat{\mu}(s,t)\|^2}{\|\mu(s,t)\|^2}, \tag{22}$$

where $\|\mu(s,t)\|^2 = \int\int \mu(s,t)^2 \mathrm{d}s\mathrm{d}t$, analogously for $\hat{X}_i^{\mathrm{c}}(s,t)$ and $\hat{\phi}_{jk}(t)\,\hat{\psi}_j(s)$. The relative squared errors over 200 simulation runs, which are reported in Table 2, were found to be quite small for

Table 2. Results for simulation 1, reporting median relative errors RE, as defined in equation (22) (with median absolute deviation in parentheses), for various components of the model and varying sample sizes n

Component	*FVE (%)*	*REs for the following values of n:*		
		$n = 50$	$n = 100$	$n = 200$
μ		0.0012 (0.0008)	0.0006 (0.0004)	0.0003 (0.0002)
X^{c}		0.1523 (0.0228)	0.1483 (0.0168)	0.1435 (0.0091)
$\phi_{11}(t)\,\psi_1(s)$	53.6967	0.0092 (0.0071)	0.0045 (0.0040)	0.002 (0.0016)
$\phi_{21}(t)\,\psi_2(s)$	12.9333	0.0584 (0.0538)	0.0280 (0.0243)	0.0133 (0.0110)
$\phi_{22}(t)\,\psi_2(s)$	6.7450	0.1306 (0.1267)	0.0660 (0.0619)	0.0311 (0.0287)
$\phi_{12}(t)\,\psi_1(s)$	4.5367	0.0222 (0.0178)	0.0129 (0.0091)	0.005 (0.0037)
$\phi_{23}(t)\,\psi_2(s)$	4.1917	0.0999 (0.0904)	0.0469 (0.0417)	0.0296 (0.0238)
$\phi_{31}(t)\,\psi_3(s)$	4.0400	0.0283 (0.0238)	0.0127 (0.0100)	0.0077 (0.0062)

Table 3. Results for simulation 2, reporting median relative errors RE, as defined in equation (22) (with median absolute deviations in parentheses) for various components of the model and varying sample sizes n

Component	*FVE (%)*	*REs for the following values of n:*		
		$n=50$	$n=100$	$n=200$
μ		0.0011 (0.0006)	0.0006 (0.0003)	0.0003 (0.0002)
X^c		0.1464 (0.0216)	0.1390 (0.0138)	0.1324 (0.0080)
$\phi_1(t)\,\psi_1(s)$	53.5400	0.0099 (0.0075)	0.0053 (0.0048)	0.0022 (0.0017)
$\phi_2(t)\,\psi_2(s)$	8.1500	0.0559 (0.0525)	0.0342 (0.0275)	0.0174 (0.0180)
$\phi_1(t)\,\psi_2(s)$	8.0817	0.0109 (0.0078)	0.0064 (0.0051)	0.0026 (0.0018)
$\phi_3(t)\,\psi_2(s)$	5.4300	0.0776 (0.0635)	0.0389 (0.0331)	0.0208 (0.0204)
$\phi_2(t)\,\psi_1(s)$	4.2783	0.0543 (0.0502)	0.0328 (0.0271)	0.0173 (0.0180)
$\phi_4(t)\,\psi_2(s)$	3.7817	0.0368 (0.0293)	0.0167 (0.0131)	0.0089 (0.0072)
$\phi_1(t)\,\psi_3(s)$	3.5917	0.0108 (0.0075)	0.0056 (0.0039)	0.0028 (0.0019)

μ, X_i^c and for the six product functions $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$ with largest FVEs, which are the same six functions as in Fig. 9 in on-line supplement B. The errors decline with increasing sample size n, as expected. The FVEs for each term (j,k) are also in Table 2, averaged over simulation runs and over the different sample sizes, as they were similar across varying sample sizes.

For simulation 2, data were generated according to a truncated product FPC model

$$X(s,t)=\mu(s,t)+\sum_{j=1}^{4}\sum_{k=1}^{4}\chi_{jk}\phi_k(t)\psi_j(s),$$

where $\mu(s,t)$ and $\phi_k(t)\psi_j(s)$ for $1\leqslant j,k\leqslant 4$ are substituted by the estimates obtained from the fertility data. As in simulation 1, the random scores χ_{jk} were generated as independent normal random variables with variances estimated from the data. Estimated and true functions $\psi_j(s)$ and $\phi_k(t)$ obtained for one sample run with $n=50$ are shown in Fig. 11 in on-line supplement B. The relative squared errors over 200 simulation runs, for μ, X_i^c and for the seven product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$ with largest FVEs (among 16 total product functions), which are the same seven functions as in Fig. 18 in on-line supplement C, are reported in Table 3. Both Fig. 18 and the numbers demonstrate good performance of the method.

7. Discussion

The marginal FPCA and product FPCA proposed provide a simple and straightforward representation of function-valued stochastic processes. This holds especially in comparison with a previously proposed two-step expansion for repeatedly observed functional data (Chen and Müller, 2012), in which processes X are represented as

$$X(s,t)=\mu(s,t)+\sum_{j=1}^{\infty}\nu_j(t)\rho_j(s|t)=\mu(s,t)+\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\theta_{jk}\omega_{jk}(t)\rho_j(s|t), \tag{23}$$

where $\rho_j(\cdot|t)$ is the jth eigenfunction of the operator in $L^2(\mathcal{S})$ with kernel $G_{\mathcal{S}}(s,u|t)=C\{(s,t),(u,t)\}$, $\nu_j(t)=\langle X(\cdot,t),\rho_j(\cdot|t)\rangle_{\mathcal{S}}$ and $\Sigma_{k=1}^{\infty}\theta_{jk}\omega_{jk}(t)$ is the Karhunen–Loève expansion of $\nu_j(t)$ as a random function in $L^2(\mathcal{T})$. This method can be characterized as a conditional FPCA approach (we note that in Chen and Müller (2012) the notation of s and t is reversed compared with in the

present paper). Similarly to the proposed marginal approach this conditional method enables asymmetric handling of arguments s and t of X and is a two-step procedure which is composed of iterated one-dimensional FPCAs.

Key differences between the marginal FPCA and the conditional FPCA are as follows.

(a) The first step of the conditional FPCA approach (23) requires performing a separate FPCA for each $t \in \mathcal{T}$, whereas in the marginal approach (4) only one FPCA is required, with lower computational cost, and, most importantly, using all the data rather than the data in a window around t.
(b) In equation (23), the eigenfunctions $\rho_j(s|t)$ depend on both arguments, making it difficult to separate and interpret the effects of s and t in conditional FPCA, in contrast with marginal FPCA, where the eigenfunctions in equation (4) depend on s only.
(c) For sparse designs, conditional FPCA requires a smoothing estimator of the function $G_{\mathcal{S}}(s, u|t)$ that depends on $2d_1 + d_2$ univariate arguments. This improves on the standard two-argument FPCA (3), where the corresponding covariance functions depend on $2d_1 + 2d_2$ arguments. The improvement is, however, even greater for marginal FPCA, where the covariance function depends on only $2d_1$ or $2d_2$ arguments, leading to faster convergence.

The proposed marginal FPCA improves on standard FPCA by providing interpretable components and making it possible to treat the index of the stochastic process asymmetrically in the arguments of the random functions that constitute the values of the process. Although we have discussed in detail the case of time-indexed function-valued processes, and our example also conforms with this simplest setting, extensions to spatially indexed function-valued processes or processes which are indexed on a rectangular subdomain of $\mathcal{R}^p$ are straightforward. Marginal FPCA also is supported by theoretical optimality properties as per theorem 1 and theorem 2.

A promising simplified version of the marginal FPCA is product FPCA, motivated by a common principal component assumption; see theorem 4. Additional motivation is its near optimality even without the common principal component assumption, as per theorem 5. In our fertility data example, the loss of flexibility is quite limited and may be outweighed by the simplicity and interpretability of this model. In general, the explanatory power of the product FPCA model depends on the structure of the double-indexed array $\eta_{jk} = \mathrm{var}(\chi_{jk})$. When one of the marginal kernels does not have fast decaying eigenvalues, relatively large values of η_{jk} might show up in large j or large k and in such situations the product FPCA might have limited explanatory power, and it would be better to apply marginal FPCA or two-dimensional FPCA. The eigenvalues of the marginal kernels can be directly estimated and can be used to diagnose this situation in data applications.

In this paper we mainly focus on the case where the argument of the functional values s is densely and regularly sampled. In practical applications with designs that are sparse in s, one may obtain $\hat{G}_{\mathcal{S}}$ by pooling the data $\{\hat{X}_i^{\mathrm{c}}(\cdot, t_{im}), i = 1, \ldots, n, m = 1, \ldots, M_i\}$, and utilizing a two-dimensional smoothing estimator of the covariance (Yao *et al.*, 2005). The FPCs can be obtained through principal analysis by conditional expectation under Gaussian assumptions; software is available from `http://www.stat.ucdavis.edu/PACE/`. For this case, one can only show that $\hat{\xi}_{i,j}(t) \to E\{\xi_{i,j}(t)|\mathrm{data}\}$, almost surely, under Gaussian assumptions.

Acknowledgements

This research was supported by National Science Foundation grants DMS-1104426, DMS-1228369 and DMS-1407852, by the Spanish Ministry of Education and Science, and Fondo

Europeo de Desarrollo Regional grant MTM2010-14887. The main part of this work was done when Pedro Delicado was visiting the University of California at Davis, USA, with the financial support of the Spanish Government (Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i).

References

Aston, J. A., Pigoli, D. and Tavakoli, S. (2015) Tests for separability in nonparametric covariance operators of random surfaces. *Preprint arXiv:1505.02023*. University of Cambridge, Cambridge.

Chen, K. and Müller, H.-G. (2012) Modeling repeated functional observations. *J. Am. Statist. Ass.*, **107**, 1599–1609.

Crainiceanu, C., Staicu, A. and Di, C. (2009) Generalized multilevel functional regression. *J. Am. Statist. Ass.*, **104**, 1550–1561.

Cuevas, A. (2013) A partial overview of the theory of statistics with functional data. *J. Statist. Planng Inf.*, **147**, 1–23.

Currie, I. D., Durban, M. and Eilers, P. H. (2004) Smoothing and forecasting mortality rates. *Statist. Modllng*, **4**, 279–298.

Currie, I. D., Durban, M. and Eilers, P. H. C. (2006) Generalized linear array models with applications to multidimensional smoothing. *J. R. Statist. Soc.* B, **68**, 259–280.

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010) Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21**, 224–239.

Eilers, P. H., Currie, I. D. and Durbán, M. (2006) Fast and compact smoothing on large multidimensional grids. *Computnl Statist. Data Anal.*, **50**, 61–76.

Eilers, P. H. and Marx, B. D. (2003) Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometr. Intell. Lab. Syst.*, **66**, 159–174.

Ezeh, A. C., Bongaarts, J. and Mberu, B. (2012) Global population trends and policy options. *Lancet*, **380**, 142–148.

Greven, S., Crainiceanu, C., Caffo, B. and Reich, D. (2010) Longitudinal functional principal component analysis. *Electron. J. Statist.*, **4**, 1022–1054.

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2012) Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Ann. Appl. Statist.*, **6**, 669–696.

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2013) Nonparametric estimation in small data sets of spatially indexed curves with application to temporal trend determination. *Computnl Statist. Data Anal.*, **59**, 82–94.

Hall, P., Müller, H.-G. and Wang, J.-L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, **34**, 1493–1517.

Horváth, L. and Kokoszka, P. (2012) *Inference for Functional Data with Applications*. New York: Springer.

Huang, J. Z., Shen, H. and Buja, A. (2009) The analysis of two-way functional data using two-way regularized singular value decompositions. *J. Am. Statist. Ass.*, **104**, 1609–1620.

Human Fertility Database (2013) Human Fertility Database. Max Planck Institute for Demographic Research, Rostock, and Vienna Institute of Demography. (Available from `http://www.humanfertility.org`.)

Hyndman, R. J. and Shang, H. L. (2009) Forecasting functional time series. *J. Kor. Statist. Soc.*, **38**, 199–211.

Hyndman, R. J. and Ullah, S. (2007) Robust forecasting of mortality and fertility rates: a functional data approach. *Computnl Statist. Data Anal.*, **51**, 4942–4956.

Jasilioniene, A., Jdanov, D. A., Sobotka, T., Andreev, E. M., Zeman, K. and Shkolnikov, V. M. (2012) Methods protocol for the Human Fertility Database. Max Planck Institute for Demographic Research, Rostock, and Vienna Institute of Demography.

Kneip, A. and Utikal, K. (2001) Inference for density families using functional principal component analysis. *J. Am. Statist. Ass.*, **96**, 519–542.

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

Morris, J. S. and Carroll, R. J. (2006) Wavelet-based functional mixed models. *J. R. Statist. Soc.* B, **68**, 179–199.

Nerini, D., Monestiez, P. and Manté, C. (2010) Cokriging for spatial functional data. *J. Multiv. Anal.*, **101**, 409–418.

Park, S. and Staicu, A. (2015) Longitudinal functional data analysis. *Stat*, **4**, 212–226.

Preston, S. H., Heuveline, P. and Guillot, M. (2001) *Demography: Measuring and Modeling Population Processes*. New York: Blackwell Publishing.

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

Takahashi, S. (2004) Demographic investigation of the declining fertility process in Japan. *Jpn J. Popln*, **2**, 93–116.

Yao, F. and Lee, T. C. M. (2006) Penalized spline models for functional principal component analysis. *J. R. Statist. Soc.* B, **68**, 3–25.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *J. Am. Statist. Ass.*, **100**, 577–590.

Yuan, Y., Gilmore, J. H., Geng, X., Martin, S., Chen, K., Wang, J.-L. and Zhu, H. (2014) Fmem: functional mixed effects modeling for the analysis of longitudinal white matter tract data. *Neuroimage*, **84**, 753–764.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Online supplement'.