

Selected statistical methods of data analysis for multivariate functional data

Tomasz Górecki¹  · Mirosław Krzyśko¹ ·
Łukasz Waszak¹ · Waldemar Wołyński¹

Received: 13 October 2014 / Revised: 3 February 2016 / Published online: 23 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Data in the form of a continuous vector function on a given interval are referred to as multivariate functional data. These data are treated as realizations of multivariate random processes. The paper is devoted to three statistical dimension reduction techniques for multivariate data. For the first one, principal components analysis, the authors present a review of a recent paper (Jacques and Preda in, *Comput Stat Data Anal*, 71:92–106, 2014). For two others one, canonical variables and discriminant coordinates, the authors extend existing works for univariate functional data to multivariate. These methods for multivariate functional data are presented, illustrated and discussed in the context of analyzing real data sets. Each of these techniques is applied on real data set.

Keywords Multivariate functional data · Functional data analysis · Principal component analysis · Discriminant coordinates · Canonical correlation analysis

1 Introduction

In recent years, methods for representing data by functions or curves have received much attention. Such data are known in the literature as functional data (Bongiorno et al. 2014; Ferraty and Vieu 2006; Horváth and Kokoszka 2012; Ramsay and Silverman 2005). Examples of functional data can be found in various application domains, such as medicine, economics, meteorology and many others. In previous papers on functional data analysis, objects are characterized only by one feature observed at

✉ Tomasz Górecki
tomasz.gorecki@amu.edu.pl

¹ Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland

many time points. Methods of functional data analysis are becoming increasingly popular, e.g. in the cluster analysis (Jacques and Preda 2013; James and Sugar 2003; Peng and Müller 2008), classification (Chamroukhi et al. 2013; Delaigle and Hall 2012; Mosler and Mozharovskiy 2015; Rossi and Villa 2006) and regression (Ferraty et al. 2012; Goia and Vieu 2014; Kudraszow and Vieu 2013; Peng et al. 2015; Rachdi and Vieu 2006; Wang et al. 2015). Unfortunately, multivariate data methods cannot be directly used for functional data, because of the problem of dimensionality and difficulty in putting functional data into order. In many applications there is a need to use statistical methods for objects characterized by many features observed at many time points (double multivariate data), and such data are called multivariate functional data. A pioneering theoretical works were Besse (1979), Besse and Ramsay (1986), where random variables take values in a general Hilbert space. Saporta (1981) presents an analysis of multivariate functional data from the point of view of factorial methods (principal components and canonical analysis). Berrendero et al. (2011), Jacques and Preda (2014) and Panaretos et al. (2010) discussed principal component analysis for multivariate functional data (MFPCA). In this paper we extend the construction of principal component analysis to other projective dimension reduction techniques, i.e. discriminant coordinates and canonical correlation analysis for multivariate functional data.

Dimension reduction is a very active field of statistical research. We focused only on projective dimension reduction techniques (Burgess 2009): principal components analysis (PCA), canonical correlation analysis (CCA) and Fisher discriminant analysis (DCA). These procedures are a transformation that allows us to obtain a linear projection of our data, originally in \mathbf{R}^p onto \mathbf{R}^k , where $k < p$. Along with reducing the data dimensions, the data are also projected in a different orientation. Altogether, this transformation presents the data in a manner that stresses out the trends in it facilitating its interpretation.

The rest of this paper is organized as follows. We first review the concept of transformation of discrete data to multivariate functional data (Sect. 2). Section 3 contains a review principal components analysis for multivariate functional data. Sections 4 and 5 contain our extension of existing works for univariate functional data to multivariate, respectively for CCA and discriminant coordinates. Section 6 contains the results of our experiments on the real data set. Conclusions are given in Sect. 7.

2 Transformation of discrete data to multivariate functional data

Let $X(t)$ be a stochastic process with continuous parameter $t \in I$. Moreover, assume that $X \in L_2(I)$, where $L_2(I)$ is a Hilbert space of square integrable functions on the interval I and that the process $X(t)$ has the following representation:

$$X(t) = \sum_{b=0}^B c_b \varphi_b(t), \quad t \in I, \quad (1)$$

where $\{\varphi_b\}$ are orthonormal basis functions, and c_0, c_1, \dots, c_B are the random coefficients.

Many financial, meteorological and other data are recorded at discrete moments in time. Let x_j denote an observed value of process $X(t)$ at the j th time point t_j , where I is a compact set such that $t_j \in I$, for $j = 1, \dots, J$. Then our data consist of J pairs (t_j, x_j) . This discrete data can be smoothed by continuous function $x(t)$, where $t \in I$ (Ramsay and Silverman 2005).

Let $\mathbf{x} = (x_1, x_2, \dots, x_J)'$, $\mathbf{c} = (c_0, c_1, \dots, c_B)'$ and $\Phi(t)$ be a matrix of dimension $J \times (B + 1)$ containing the values $\varphi_b(t_j)$, $b = 0, 1, \dots, B$, $j = 1, 2, \dots, J$. The coefficient \mathbf{c} in (1) is estimated by the least squares method, that is, so as to minimize the function:

$$S(\mathbf{c}) = (\mathbf{x} - \Phi(t)\mathbf{c})' (\mathbf{x} - \Phi(t)\mathbf{c}).$$

Differentiating $S(\mathbf{c})$ with respect to the vector \mathbf{c} , we obtain the least squares method estimator

$$\hat{\mathbf{c}} = (\Phi'(t)\Phi(t))^{-1} \Phi'(t)\mathbf{x}.$$

Then

$$x(t) = \sum_{b=0}^B \hat{c}_b \varphi_b(t), \quad t \in I. \quad (2)$$

The degree of smoothness of the function $x(t)$ depends on the value B (a small value of B causes more smoothing of the curves). The optimum value for B is selected using the Bayesian information criterion (BIC):

$$\text{BIC} = \ln \left(\sum_{j=0}^J (x_j - \sum_{b=0}^B \hat{c}_b \varphi_b(t_j))^2 \right) + (B + 1) \left(\frac{\ln J}{J} \right).$$

We decided to use such criterion because Akaike Information Criterion (AIC) better measures predictive accuracy while BIC better measures goodness of fit (Berk 2008; Shmueli 2010; Sober 2002).

Let us assume that there are n independent pairs of values (t_{ij}, x_{ij}) , $j = 1, \dots, J$, $i = 1, \dots, n$. These discrete data are smoothed to continuous functions in the following form:

$$x_i(t) = \sum_{b=0}^{B_i} \hat{c}_{ib} \varphi_b(t), \quad i = 1, \dots, n, \quad t \in I.$$

Among all the B_1, B_2, \dots, B_n one common value of B is chosen, as the modal value of the numbers B_1, B_2, \dots, B_n .

The set of functions $\{x_1(t), \dots, x_n(t) : t \in I\}$ obtained in this way is called functional data (see Ramsay and Silverman 2005). Note that in some cases it could be interesting

to discretize smooth functions. This is at the center of variable selection methods in functional data analysis (Aneiros and Vieu 2014).

So far we have been dealing with data characterized by one feature. Our considerations can be generalized to the case of $p \geq 2$ features. Then our data consist of n independent vector functions $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \dots, x_{ip}(t))'$, $t \in I$, $i = 1, \dots, n$. The data $\{\mathbf{x}_1(t), \dots, \mathbf{x}_n(t) : t \in I\}$ will be called multivariate functional data. Multivariate functional data can conveniently be treated as realizations of a finite multidimensional stochastic process $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_p(t))'$ with continuous parameter $t \in I$. We will further assume, that $\mathbf{X} \in L_2^p(I)$, where $L_2(I)$ is a Hilbert space of square integrable functions on the interval I equipped with the following inner product:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \int_I \mathbf{u}'(t) \mathbf{v}(t) dt.$$

We consider the case, where the d th component of process $\mathbf{X}(t)$ can be represented by a finite number of orthonormal basis functions $\{\varphi_b\}$

$$X_d(t) = \sum_{b=0}^{B_d} c_{db} \varphi_b(t), \quad t \in I, \quad d = 1, 2, \dots, p,$$

where c_{db} are random variables. Let

$$\begin{aligned} \mathbf{c} &= (c_{10}, \dots, c_{1B_1}, \dots, c_{p0}, \dots, c_{pB_p})', \\ \Phi(t) &= \begin{bmatrix} \varphi'_{B_1}(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \varphi'_{B_2}(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \varphi'_{B_p}(t) \end{bmatrix}, \end{aligned} \quad (3)$$

where $\varphi_{B_d}(t) = (\varphi_0(t), \dots, \varphi_{B_d}(t))'$, $d = 1, \dots, p$. Then

$$\mathbf{X}(t) = \Phi(t) \mathbf{c}, \quad t \in I.$$

3 Principal component analysis for multivariate functional data

The idea of principal component analysis is to reduce the dimensionality of a data set consisting of a large number of correlated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables. Suppose that we are observing a p -dimensional random vector $\mathbf{X} = (X_1, X_2, \dots, X_p)' \in \mathbb{R}^p$. In the first step we look for a linear combination $U_1 = u_{11}X_1 + u_{12}X_2 + \dots + u_{1p}X_p = \mathbf{u}'_1 \mathbf{X}$ of the elements of vector \mathbf{X} having maximum variance. The variable U_1 is called the first principal component. Next, we look

for a linear combination $U_2 = \mathbf{u}'_2 \mathbf{X}$, uncorrelated with the first principal component U_1 , having maximum variance, and so on, so that at the k th stage a linear combination $U_k = \mathbf{u}'_k \mathbf{X}$, called the k th principal component, is found that has maximum variance subject to being uncorrelated with the first $k - 1$ principal components (Jolliffe 2002). The observations can be presented graphically as points on a plane (U_1, U_2) . The functional case of PCA (FPCA) is a more informative way of looking at the variability structure in the variance-covariance operator for one dimensional functional data (Górecki and Krzyśko 2012). In this section we present PCA for multivariate functional data (Jacques and Preda 2014).

Without loss of generality we will further assume, that $E(\mathbf{X}) = \mathbf{0}$. In principal component analysis in the multivariate functional case, we are interested in finding the inner product

$$U = \langle \mathbf{u}, \mathbf{X} \rangle = \int_I \mathbf{u}'(t) \mathbf{X}(t) dt$$

having maximal variance for all $\mathbf{u} \in L_2^p(I)$ such, that $\langle \mathbf{u}, \mathbf{u} \rangle = 1$. Let

$$\lambda_1 = \sup_{\mathbf{u} \in L_2^p(I)} \text{Var}(\langle \mathbf{u}, \mathbf{X} \rangle) = \text{Var}(\langle \mathbf{u}_1, \mathbf{X} \rangle),$$

where $\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = 1$. The inner product $U_1 = \langle \mathbf{u}_1, \mathbf{X} \rangle$ will be called the first principal component, and the vector function \mathbf{u}_1 will be called the first vector weight function. Subsequently we look for the second principal component $U_2 = \langle \mathbf{u}_2, \mathbf{X} \rangle$, which maximizes $\text{Var}(\langle \mathbf{u}, \mathbf{X} \rangle)$, is such that $\langle \mathbf{u}_2, \mathbf{u}_2 \rangle = 1$, and is not correlated with the first functional principal component U_1 , i.e. is subject to the restriction $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = 0$.

In general, the k th functional principal component $U_k = \langle \mathbf{u}_k, \mathbf{X} \rangle$ satisfies the conditions:

$$\begin{aligned} \lambda_k &= \sup_{\mathbf{u} \in L_2^p(I)} \text{Var}(\langle \mathbf{u}, \mathbf{X} \rangle) = \text{Var}(\langle \mathbf{u}_k, \mathbf{X} \rangle), \\ \langle \mathbf{u}_{\kappa_1}, \mathbf{u}_{\kappa_2} \rangle &= \delta_{\kappa_1 \kappa_2}, \quad \kappa_1, \kappa_2 = 1, \dots, k, \end{aligned}$$

where

$$\delta_{\kappa_1 \kappa_2} = \begin{cases} 1 & \text{if } \kappa_1 = \kappa_2 \\ 0 & \text{if } \kappa_1 \neq \kappa_2. \end{cases}$$

The expression $(\lambda_k, \mathbf{u}_k(t))$ will be called the k th principal system of the process $\mathbf{X}(t)$.

In sect. 2 we have shown, that the process $\mathbf{X}(t)$ can be represented as $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, $t \in I$. Now let us consider the principal components of the random vector \mathbf{c} . From $E(\mathbf{X}) = \mathbf{0}$ we have $E(\mathbf{c}) = \mathbf{0}$. Let us denote $\text{Var}(\mathbf{c}) = \Sigma$. The k th principal component $U_k^* = \langle \omega_k, \mathbf{c} \rangle$ of this vector satisfies the conditions:

$$\begin{aligned}
\gamma_k &= \sup_{\omega \in \mathbb{R}^{K+p}} \text{Var}(\langle \omega, \mathbf{c} \rangle) = \sup_{\omega \in \mathbb{R}^{K+p}} \omega' \text{Var}(\mathbf{c}) \omega \\
&= \sup_{\omega \in \mathbb{R}^{K+p}} \omega' \Sigma \omega = \omega'_k \Sigma \omega_k, \\
\omega'_{\kappa_1} \omega_{\kappa_2} &= \delta_{\kappa_1 \kappa_2},
\end{aligned}$$

where $\kappa_1, \kappa_2 = 1, \dots, k$, $K = B_1 + \dots + B_p$. The expression (γ_k, ω_k) will be called the k th principal system of vector \mathbf{c} .

Determining the k th principal system of vector \mathbf{c} is equivalent to solving for the eigenvalue and corresponding eigenvectors of the covariance matrix Σ of that vector, standardized so that $\omega'_{\kappa_1} \omega_{\kappa_2} = \delta_{\kappa_1 \kappa_2}$.

Theorem 1 *The k th principal system $(\lambda_k, \mathbf{u}_k(t))$ of the stochastic process $\mathbf{X}(t)$ is related to the k th principal system (γ_k, ω_k) of the random vector \mathbf{c} by the equations:*

$$\lambda_k = \gamma_k, \quad \mathbf{u}_k(t) = \Phi(t) \omega_k, \quad t \in I,$$

where $k = 1, \dots, s$ and $s = \text{rank}(\Sigma)$.

Proof It may be assumed (Ramsay and Silverman 2005), that the vector weight function $\mathbf{u}(t)$ and the process $\mathbf{X}(t)$ are in the same space, i.e. the function $\mathbf{u}(t)$ can be written in the form:

$$\mathbf{u}(t) = \Phi(t) \omega,$$

where $\omega \in \mathbb{R}^{K+p}$. Then

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{X} \rangle &= \int_I \mathbf{u}'(t) \mathbf{X}(t) dt = \int_I \omega' \Phi'(t) \Phi(t) \mathbf{c} dt \\
&= \omega' \int_I \Phi'(t) \Phi(t) dt \mathbf{c} = \omega' \mathbf{I}_{K+p} \mathbf{c} = \omega' \mathbf{c},
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}(\langle \mathbf{u}, \mathbf{X} \rangle) &= \omega' \mathbb{E}(\mathbf{c}) = \omega' \mathbf{0} = 0, \\
\text{Var}(\langle \mathbf{u}, \mathbf{X} \rangle) &= \omega' \mathbb{E}(\mathbf{c} \mathbf{c}') \omega = \omega' \Sigma \omega.
\end{aligned}$$

Let us consider the first functional principal component of process $\mathbf{X}(t)$:

$$\lambda_1 = \sup_{\mathbf{u} \in L_2^p(I)} \text{Var}(\langle \mathbf{u}, \mathbf{X} \rangle) = \text{Var}(\langle \mathbf{u}_1, \mathbf{X} \rangle),$$

where $\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = 1$. This is equivalent to saying that

$$\gamma_1 = \sup_{\omega \in \mathbb{R}^{K+p}} \text{Var}(\langle \omega, \mathbf{c} \rangle) = \sup_{\omega \in \mathbb{R}^{K+p}} \omega' \text{Var}(\mathbf{c}) \omega = \omega'_1 \Sigma \omega_1,$$

where $\omega'_1 \omega_1 = 1$.

This is the definition of the first principal component of the random vector \mathbf{c} . On the other hand, if we begin with the first principal system of the random vector \mathbf{c} defined by $(\gamma_1, \boldsymbol{\omega}_1)$, we will obtain the first principal system for the process $\mathbf{X}(t)$ from the equations

$$\lambda_1 = \gamma_1, \mathbf{u}_1(t) = \Phi(t)\boldsymbol{\omega}_1.$$

We may extend these considerations to the second principal system and so on. \square

Principal component analysis for random vectors \mathbf{c} is based on the matrix $\boldsymbol{\Sigma}$. In practice this matrix is unknown. We estimate it on the basis of n independent realizations $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t)$ of the form $\mathbf{x}_i(t) = \Phi(t)\hat{\mathbf{c}}_i$ of the random process $\mathbf{X}(t)$, where the vectors $\hat{\mathbf{c}}_i$ are centered, $i = 1, 2, \dots, n$.

Let $\hat{\mathbf{C}} = (\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, \dots, \hat{\mathbf{c}}_n)'$. Then

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \hat{\mathbf{C}}' \hat{\mathbf{C}}.$$

Let $\hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \dots \geq \hat{\gamma}_s$ be non-zero eigenvalues of matrix $\hat{\boldsymbol{\Sigma}}$, and $\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2, \dots, \hat{\boldsymbol{\omega}}_s$ the corresponding eigenvectors, where $s = \text{rank}(\hat{\boldsymbol{\Sigma}})$.

Moreover the k th principal system of the random process $\mathbf{X}(t)$ determined from a sample has the following form:

$$(\hat{\lambda}_k = \hat{\gamma}_k, \hat{\mathbf{u}}_k(t) = \Phi(t)\hat{\boldsymbol{\omega}}_k), \quad k = 1, \dots, s.$$

Hence the coefficients of the projection of the i th realization $\mathbf{x}_i(t)$ of process $\mathbf{X}(t)$ on the k th functional principal component are equal to:

$$\begin{aligned} \hat{U}_{ik} &= \langle \hat{\mathbf{u}}_k, \mathbf{x}_i \rangle = \int_I \hat{\boldsymbol{\omega}}_k' \Phi'(t) \Phi(t) \hat{\mathbf{c}}_i dt \\ &= \hat{\boldsymbol{\omega}}_k' \int_I \Phi'(t) \Phi(t) dt \hat{\mathbf{c}}_i = \hat{\boldsymbol{\omega}}_k' \hat{\mathbf{c}}_i, \end{aligned}$$

for $i = 1, 2, \dots, n, k = 1, 2, \dots, s$. Finally the coefficients of the projection of the i th realization $\mathbf{x}_i(t)$ of the process $\mathbf{X}(t)$ on the plane of the first two functional principal components from the sample are equal to $(\hat{\boldsymbol{\omega}}_1' \hat{\mathbf{c}}_i, \hat{\boldsymbol{\omega}}_2' \hat{\mathbf{c}}_i)$, $i = 1, 2, \dots, n$.

4 Discriminant coordinates for multivariate functional data

Now let us consider the case where the samples originate from L groups. We would often like to present them graphically, to see their configuration or to eliminate outlying observations. However it may be difficult to produce such a presentation even if only three features are observed. A different method must therefore be sought for presenting multidimensional data originating from multiple groups. To make the task easier, in the first step every p -dimensional observation $\mathbf{X} = (X_1, X_2, \dots, X_p)' \in \mathbb{R}^p$ can be transformed into a one-dimensional observation $U_1 = u_{11}X_1 + u_{12}X_2 + \dots +$

$u_{1p}X_p = \mathbf{u}'_1\mathbf{X}$, and the resulting one-dimensional observations can be presented graphically as points on a straight line. In the second step we can define a second linear combination $U_2 = \mathbf{u}_2'\mathbf{X}$ not correlated with the first, and present the observations graphically as points on a plane (U_1, U_2) . The space of discriminant coordinates is a space convenient for the use of various classification methods (methods of discriminant analysis). When $L = 2$ we obtain only one discriminant coordinate, coinciding with the well-known Fisher's linear discriminant function (Fisher 1936). The functional case of discriminant coordinates analysis (FDCA) and its kernel variant (KFDCA) are also well known (Górecki et al. 2014). In this section we propose FDCA for multivariate functional data (MFDCA). Let $\mathbf{x}_{l1}(t), \mathbf{x}_{l2}(t), \dots, \mathbf{x}_{ln_l}(t)$ be n_l independent realizations of a p -dimensional stochastic process $\mathbf{X}(t)$ belonging to the l th class, where $l = 1, 2, \dots, L$. Our purpose is to construct the discriminant coordinate based on multivariate functional data, i.e. to construct

$$U = \langle \mathbf{u}, \mathbf{X} \rangle = \int_I \mathbf{u}'(t)\mathbf{X}(t)dt$$

such that their between-class variance is maximal compared with the total variance, where $\mathbf{u} \in L^p_2(I)$. The vector function $\mathbf{u}(t) = (u_1(t), u_2(t), \dots, u_p(t))'$ is called the vector weight function.

More precisely, the first functional discriminant coordinate $U_1 = \langle \mathbf{u}_1, \mathbf{X} \rangle$ is defined as

$$\begin{aligned} \lambda_1 &= \sup_{\mathbf{u} \in L^p_2(I)} \frac{\text{Var}_B(\langle \mathbf{u}, \mathbf{X} \rangle)}{\text{Var}_T(\langle \mathbf{u}, \mathbf{X} \rangle)} \\ &= \frac{\text{Var}_B(\langle \mathbf{u}_1, \mathbf{X} \rangle)}{\text{Var}_T(\langle \mathbf{u}_1, \mathbf{X} \rangle)}, \end{aligned}$$

subject to the constraint

$$\text{Var}_T(\langle \mathbf{u}_1, \mathbf{X} \rangle) = 1, \quad (4)$$

where $\text{Var}_B(\langle \mathbf{u}, \mathbf{X} \rangle)$ and $\text{Var}_T(\langle \mathbf{u}, \mathbf{X} \rangle)$ are respectively the between-class and total variance of discriminant coordinate U_1 . Condition (4) ensures the uniqueness of the first discriminant coordinate U_1 .

Similarly we can construct the k th functional discriminant coordinate

$$U_k = \langle \mathbf{u}_k, \mathbf{X} \rangle,$$

where the vector weight function $\mathbf{u}_k(t)$ is defined as

$$\begin{aligned} \lambda_k &= \sup_{\mathbf{u} \in L^p_2(I)} \frac{\text{Var}_B(\langle \mathbf{u}, \mathbf{X} \rangle)}{\text{Var}_T(\langle \mathbf{u}, \mathbf{X} \rangle)} \\ &= \frac{\text{Var}_B(\langle \mathbf{u}_k, \mathbf{X} \rangle)}{\text{Var}_T(\langle \mathbf{u}_k, \mathbf{X} \rangle)}, \end{aligned}$$

subject to the constraint

$$\text{Var}_T(\langle \mathbf{u}_k, \mathbf{X} \rangle) = 1.$$

Moreover the k th discriminant coordinate U_k is not correlated with the first $k - 1$ discriminant coordinates. The expression $(\lambda_k, \mathbf{u}_k(t))$ will be called the k th discriminant system of the process $\mathbf{X}(t)$.

Let us recall that the process $\mathbf{X}(t)$ can be represented as $\mathbf{X}(t) = \Phi(t)\mathbf{c}$, $t \in I$. Now let us consider the discriminant coordinates of the random vector \mathbf{c} . The k th discriminant coordinate $U_k^* = \langle \boldsymbol{\omega}_k, \mathbf{c} \rangle$ of this vector satisfies the condition:

$$\begin{aligned} \gamma_k &= \sup_{\boldsymbol{\omega} \in \mathbb{R}^{K+p}} \text{Var}_B(\langle \boldsymbol{\omega}, \mathbf{c} \rangle) = \text{Var}_B(\langle \boldsymbol{\omega}_k, \mathbf{c} \rangle) \\ &= \boldsymbol{\omega}_k' \text{Var}_B(\mathbf{c}) \boldsymbol{\omega}_k = \boldsymbol{\omega}_k' \mathbf{B} \boldsymbol{\omega}_k, \end{aligned}$$

subject to the restriction

$$\text{Var}_T(\langle \boldsymbol{\omega}_k, \mathbf{c} \rangle) = \boldsymbol{\omega}_k' \text{Var}_T(\mathbf{c}) \boldsymbol{\omega}_k = \boldsymbol{\omega}_k' \mathbf{T} \boldsymbol{\omega}_k = 1.$$

Additionally the k th discriminant coordinate U_k^* is not correlated with the first $k - 1$ discriminant coordinates, i.e.

$$\boldsymbol{\omega}_{\kappa_1}' \mathbf{T} \boldsymbol{\omega}_{\kappa_2} = \delta_{\kappa_1 \kappa_2}, \quad \kappa_1, \kappa_2 = 1, \dots, k.$$

The expression $(\gamma_k, \boldsymbol{\omega}_k)$ will be called the k th discriminant system of the random vector \mathbf{c} .

Theorem 2 *The k th discriminant system $(\lambda_k, \mathbf{u}_k(t))$ of the stochastic process $\mathbf{X}(t)$ is related to the k th discriminant system $(\gamma_k, \boldsymbol{\omega}_k)$ of the random vector \mathbf{c} by the equations:*

$$\lambda_k = \gamma_k, \quad \mathbf{u}_k(t) = \Phi(t) \boldsymbol{\omega}_k, \quad t \in I,$$

where $k = 1, \dots, s$, $s = \min(K + p, L - 1)$.

Proof We assume, that the vector weight function $\mathbf{u}(t)$ and the process $\mathbf{X}(t)$ are in the same space, i.e. the function $\mathbf{u}(t)$ can be written in the form:

$$\mathbf{u}(t) = \Phi(t) \boldsymbol{\omega}$$

where $\boldsymbol{\omega} \in \mathbb{R}^{K+p}$. Than

$$\langle \mathbf{u}, \mathbf{X} \rangle = \boldsymbol{\omega}' \mathbf{c}.$$

Hence the between-class variance of the inner product $\langle \mathbf{u}, \mathbf{X} \rangle$ is

$$\text{Var}_B(\langle \mathbf{u}, \mathbf{X} \rangle) = \boldsymbol{\omega}' \text{Var}_B(\mathbf{c}) \boldsymbol{\omega},$$

and the total variance

$$\text{Var}_T(<\mathbf{u}, \mathbf{X}>) = \boldsymbol{\omega}' \text{Var}_T(\mathbf{c})\boldsymbol{\omega},$$

where $\text{Var}_B(\mathbf{c})$ and $\text{Var}_T(\mathbf{c})$ are respectively the matrices of sum of squares and products of between-class and total variance.

For the first functional discriminant coordinate of the process $\mathbf{X}(t)$ we have:

$$\begin{aligned} \lambda_1 &= \sup_{\mathbf{u} \in L_2^p(I)} \frac{\text{Var}_B(<\mathbf{u}, \mathbf{X}>)}{\text{Var}_T(<\mathbf{u}, \mathbf{X}>)} \\ &= \sup_{\boldsymbol{\omega} \in \mathbb{R}^{K+p}} \frac{\boldsymbol{\omega}' \text{Var}_B(\mathbf{c})\boldsymbol{\omega}}{\boldsymbol{\omega}' \text{Var}_T(\mathbf{c})\boldsymbol{\omega}} = \frac{\boldsymbol{\omega}'_1 \text{Var}_B(\mathbf{c})\boldsymbol{\omega}_1}{\boldsymbol{\omega}'_1 \text{Var}_T(\mathbf{c})\boldsymbol{\omega}_1}, \end{aligned}$$

where $\boldsymbol{\omega}'_1 \text{Var}_T(\mathbf{c})\boldsymbol{\omega}_1 = 1$. This is equivalent to

$$\gamma_1 = \sup_{\boldsymbol{\omega} \in \mathbb{R}^{K+p}} \text{Var}_B(<\boldsymbol{\omega}, \mathbf{c}>) = \sup_{\boldsymbol{\omega} \in \mathbb{R}^{K+p}} \boldsymbol{\omega}' \text{Var}_B(\mathbf{c})\boldsymbol{\omega} = \boldsymbol{\omega}'_1 \mathbf{B}\boldsymbol{\omega}_1,$$

where $\boldsymbol{\omega}'_1 \boldsymbol{\omega}_1 = 1$.

This meets the definition of the first discriminant coordinate of the random vector \mathbf{c} . On the other hand, if the first discriminant system $(\gamma_1, \boldsymbol{\omega}_1)$ defines the first discriminant coordinate of the random vector \mathbf{c} , we will obtain the first discriminant system for the process $\mathbf{X}(t)$ from the equations

$$\lambda_1 = \gamma_1, \mathbf{u}_1(t) = \boldsymbol{\Phi}(t)\boldsymbol{\omega}_1.$$

We may extend these considerations to the second discriminant system and so on. \square

The matrices $\text{Var}_B(\mathbf{c})$ and $\text{Var}_T(\mathbf{c})$ are unknown and must be estimated based on the sample. Let $\mathbf{x}_{l1}(t), \mathbf{x}_{l2}(t), \dots, \mathbf{x}_{ln_l}(t)$ be a sample belonging to the l th class, where $l = 1, 2, \dots, L$. The function $\mathbf{x}_{li}(t)$ has the form

$$\mathbf{x}_{li}(t) = \boldsymbol{\Phi}(t)\hat{\mathbf{c}}_{li},$$

where $\hat{\mathbf{c}}_{lj} = \left(\hat{c}_{10}^{(lj)}, \dots, \hat{c}_{1K_1}^{(lj)}, \dots, \hat{c}_{p0}^{(lj)}, \dots, \hat{c}_{pK_p}^{(lj)} \right)', i = 1, 2, \dots, n_l, l = 1, 2, \dots, L$. Let

$$\bar{\mathbf{c}}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} \hat{\mathbf{c}}_{li}.$$

Then

$$\hat{\text{Var}}_B(\mathbf{c}) = \hat{\mathbf{B}} = \frac{1}{L} \sum_{l=1}^L n_l \bar{\mathbf{c}}_l \bar{\mathbf{c}}_l',$$

$$\hat{\text{Var}}_T(\mathbf{c}) = \hat{\mathbf{T}} = \frac{1}{n} \sum_{l=1}^L \sum_{i=1}^{n_l} n_l \hat{\mathbf{c}}_{li} \hat{\mathbf{c}}_{li}',$$

where $n = \sum_{l=1}^L n_l$. Next we find non-zero eigenvalues $\hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \dots \geq \hat{\gamma}_s$ and the corresponding eigenvectors $\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2, \dots, \hat{\boldsymbol{\omega}}_s$ of matrix $\hat{\mathbf{T}}^{-1} \hat{\mathbf{B}}$, where $s = \min(K + p, L - 1)$. Furthermore the k th discriminant system of the random process $\mathbf{X}(t)$ has the following form:

$$(\hat{\lambda}_k = \hat{\gamma}_k, \hat{\mathbf{u}}_k(t) = \Phi(t) \hat{\boldsymbol{\omega}}_k), \quad k = 1, \dots, s = \min(K + p, L - 1).$$

Hence the coefficients of the projection of the i th realization $\mathbf{x}_{li}(t)$ of process $\mathbf{X}(t)$ belonging to the l th class on the k th functional discriminant component are equal to:

$$\hat{U}_{lik} = \langle \hat{\mathbf{u}}_k(t), \mathbf{x}_{li}(t) \rangle = \hat{\boldsymbol{\omega}}_k' \hat{\mathbf{c}}_{li},$$

for $i = 1, 2, \dots, n_l, k = 1, 2, \dots, s, l = 1, 2, \dots, L$.

5 Canonical correlation analysis for multivariate functional data

Suppose now that we are observing two random vectors $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)' \in \mathbb{R}^p$ and $\mathbf{X} = (X_1, X_2, \dots, X_q)' \in \mathbb{R}^q$ and looking for the relationship between them. This is the one of the main problems of canonical correlation analysis. We search for weight vectors $\mathbf{u} \in \mathbb{R}^p$ and $\mathbf{v} \in \mathbb{R}^q$, such that the linear combinations $U_1 = u_{11}Y_1 + u_{12}Y_2 + \dots + u_{1p}Y_p = \mathbf{u}_1' \mathbf{Y}$ and $V_1 = v_{11}X_1 + v_{12}X_2 + \dots + v_{1q}X_q = \mathbf{v}_1' \mathbf{X}$, called the first pair of canonical variables, are maximally correlated.

Canonical correlation analysis has been extended to the case of multivariate time series (Brillinger 2001), under the assumption of stationarity, and extension of canonical correlation to functional data has been proposed in Leurgans et al. (1993), where the need for regularization was pointed out. He et al. (2000) showed that random processes with finite basis expansion have simple canonical structures, analogously to the case of random vectors. This motivates to implement regularization by projecting random processes on a finite number of basis function. The idea to project processes on the finite k basis functions has been discussed in He et al. (2004). This projection is on a prespecified orthonormal basis. In this Section, we consider the canonical correlations for the multivariate functional data. The proposed method is a generalization of the method presented in He et al. (2004). Other generalization is presented by Dubin and Müller (2005).

Let $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ are stochastic processes. We will further assume that $\mathbf{Y} \in L_2^p(I_1)$, $\mathbf{X} \in L_2^q(I_2)$ and each component $Y_g(t)$ of process $\mathbf{Y}(t)$ and $X_h(t)$ of process $\mathbf{X}(t)$

can be represented by a finite number of orthonormal basis functions $\{\varphi_e\}$ and $\{\varphi_f\}$ respectively:

$$Y_g(t) = \sum_{e=0}^{E_g} \alpha_{ge} \varphi_e(t), t \in I_1, g = 1, 2, \dots, p,$$

$$X_h(t) = \sum_{f=0}^{F_h} \beta_{hf} \varphi_f(t), t \in I_2, h = 1, 2, \dots, q.$$

Moreover let $E(\mathbf{Y}) = \mathbf{0}$, $E(\mathbf{X}) = \mathbf{0}$. This fact does not cause loss of generality, because functional canonical variables are calculated based on the covariance functions of processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$.

We introduce the following notation:

$$\begin{aligned} \boldsymbol{\alpha} &= (\alpha_{10}, \dots, \alpha_{1E_1}, \dots, \alpha_{p0}, \dots, \alpha_{pE_p})', \\ \boldsymbol{\beta} &= (\beta_{10}, \dots, \beta_{1F_1}, \dots, \beta_{q0}, \dots, \beta_{qF_q})', \\ \Phi_1(t) &= \begin{bmatrix} \varphi'_{E_1}(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \varphi'_{E_2}(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \varphi'_{E_p}(t) \end{bmatrix}, \\ \Phi_2(t) &= \begin{bmatrix} \varphi'_{F_1}(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \varphi'_{F_2}(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \varphi'_{F_q}(t) \end{bmatrix}, \end{aligned}$$

where $\varphi_{E_1}, \dots, \varphi_{E_p}$ and $\varphi_{F_1}, \dots, \varphi_{F_q}$ are orthonormal basis functions of space $L_2(I_1)$ and $L_2(I_2)$, respectively, and $K_1 = E_1 + E_2 + \dots + E_p$, $K_2 = F_1 + F_2 + \dots + F_q$. Using the above matrix notation the processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ can be represented as:

$$\mathbf{Y}(t) = \Phi_1(t)\boldsymbol{\alpha}, \quad \mathbf{X}(t) = \Phi_2(t)\boldsymbol{\beta}.$$

Functional canonical variables U and V for stochastic processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ are defined as follows

$$U = \langle \mathbf{u}, \mathbf{Y} \rangle = \int_{I_1} \mathbf{u}'(t) \mathbf{Y}(t) dt, \quad V = \langle \mathbf{v}, \mathbf{X} \rangle = \int_{I_2} \mathbf{v}'(t) \mathbf{X}(t) dt,$$

where the vector functions $\mathbf{u}(t)$ and $\mathbf{v}(t)$ are called the vector weight functions. The weight functions $\mathbf{u}(t)$ and $\mathbf{v}(t)$ are chosen to maximize the coefficients

$$\rho = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U) \text{Var}(V)}} \in (0, 1],$$

subject to the constraint that

$$\text{Var}(U) = \text{Var}(V) = 1. \quad (5)$$

The coefficient ρ is called the canonical correlation coefficient. However, simply carrying out this maximization does not produce a meaningful result. The correlation ρ achieved by the function $\mathbf{u}(t)$ and $\mathbf{v}(t)$ is equal to 1. The canonical variate weight functions $\mathbf{u}(t)$ and $\mathbf{v}(t)$ do not give any meaningful information about the data and clearly demonstrate the need for a technique involving smoothing. A straightforward way of introducing smoothing is to modify the constraints (5) by adding roughness penalty terms to give (Ramsay and Silverman 2005):

$$\text{Var}\left(U^{(N)}\right) = \text{Var}\left(\int_{I_1} \mathbf{u}'(t) \mathbf{Y}(t) dt\right) + \lambda \text{PEN}_2(\mathbf{u}) = 1, \quad (6)$$

$$\text{Var}\left(V^{(N)}\right) = \text{Var}\left(\int_{I_2} \mathbf{v}'(t) \mathbf{X}(t) dt\right) + \lambda \text{PEN}_2(\mathbf{v}) = 1, \quad (7)$$

where the roughness function PEN_2 is the integrated squared second derivative

$$\text{PEN}_2(\mathbf{u}) = \int_{I_1} \left(\frac{\partial^2 \mathbf{u}(t)}{\partial t^2} \right)' \frac{\partial^2 \mathbf{u}(t)}{\partial t^2} dt.$$

Assuming that the vector weight function $\mathbf{u}(t)$ and the process $Y(t)$ are in the same space, i.e. the function $\mathbf{u}(t)$ can be written in the form:

$$\mathbf{u}(t) = \Phi_1(t) \boldsymbol{\omega}$$

we have

$$\begin{aligned} \text{PEN}_2(\mathbf{u}) &= \int_{I_1} \left(\frac{\partial^2 \Phi_1(t) \boldsymbol{\omega}}{\partial t^2} \right)' \frac{\partial^2 \Phi_1(t) \boldsymbol{\omega}}{\partial t^2} dt \\ &= \boldsymbol{\omega}' \int_{I_1} \left(\frac{\partial^2 \Phi_1(t)}{\partial t^2} \right)' \frac{\partial^2 \Phi_1(t)}{\partial t^2} dt \boldsymbol{\omega} \\ &= \boldsymbol{\omega}' \mathbf{R}_1 \boldsymbol{\omega}, \end{aligned}$$

where

$$\mathbf{R}_1 = \int_{I_1} \left(\frac{\partial^2 \Phi_1(t)}{\partial t^2} \right)' \frac{\partial^2 \Phi_1(t)}{\partial t^2} dt. \quad (8)$$

Similarly assuming that

$$\mathbf{v}(t) = \Phi_2(t) \mathbf{v}$$

we can obtain

$$\text{PEN}_2(\mathbf{v}) = \mathbf{v}' \mathbf{R}_2 \mathbf{v},$$

where

$$\mathbf{R}_2 = \int_{I_2} \left(\frac{\partial^2 \Phi_2(t)}{\partial t^2} \right)' \frac{\partial^2 \Phi_2(t)}{\partial t^2} dt. \quad (9)$$

Now the first functional canonical correlation ρ_1 and corresponding vector weight functions $\mathbf{u}_1(t)$ and $\mathbf{v}_1(t)$ are defined as

$$\rho_1 = \sup_{\mathbf{u} \in L_2^p(I_1), \mathbf{v} \in L_2^q(I_2)} \frac{\text{Cov}(\langle \mathbf{u}, \mathbf{Y} \rangle, \langle \mathbf{v}, \mathbf{X} \rangle)}{\sqrt{\text{Var}(U^{(N)}) \text{Var}(V^{(N)})}},$$

subject to the constraint that

$$\text{Var}(U^{(N)}) = \text{Var}(V^{(N)}) = 1.$$

In general, the k th functional canonical correlation ρ_k and the associated vector weight functions $\mathbf{u}_k(t)$ and $\mathbf{v}_k(t)$ are defined as follows:

$$\begin{aligned} \rho_k &= \sup_{\mathbf{u} \in L_2^p(I_1), \mathbf{v} \in L_2^q(I_2)} \text{Cov}(\langle \mathbf{u}, \mathbf{Y} \rangle, \langle \mathbf{v}, \mathbf{X} \rangle) \\ &= \text{Cov}(\langle \mathbf{u}_k, \mathbf{Y} \rangle, \langle \mathbf{v}_k, \mathbf{X} \rangle), \end{aligned}$$

where $\mathbf{u}_k(t)$ and $\mathbf{v}_k(t)$ are subject to the restrictions (6) and (7), and the k th pair of canonical variables (U_k, V_k) is not correlated with the first $k - 1$ canonical variables, where

$$U_k = \langle \mathbf{u}_k, \mathbf{Y} \rangle, \quad V_k = \langle \mathbf{v}_k, \mathbf{X} \rangle$$

are canonical variables. We refer to this procedure as smoothed canonical correlation analysis. The expression $(\rho_k, \mathbf{u}_k(t), \mathbf{v}_k(t))$ will be called the k th canonical system of the pair of processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$.

Let

$$\begin{aligned} \text{Var}(\boldsymbol{\alpha}) &= \text{E}(\boldsymbol{\alpha}\boldsymbol{\alpha}') = \boldsymbol{\Sigma}_{11}, \\ \text{Var}(\boldsymbol{\beta}) &= \text{E}(\boldsymbol{\beta}\boldsymbol{\beta}') = \boldsymbol{\Sigma}_{22}, \\ \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \text{E}(\boldsymbol{\alpha}\boldsymbol{\beta}') = \boldsymbol{\Sigma}_{12}. \end{aligned}$$

Let us consider the canonical variables $U^* = \langle \boldsymbol{\omega}, \boldsymbol{\alpha} \rangle$ and $V^* = \langle \mathbf{v}, \boldsymbol{\beta} \rangle$ of random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively. The k th canonical correlation γ_k and associated vector weights $\boldsymbol{\omega}_k$ and \mathbf{v}_k are defined as

$$\gamma_k = \sup_{\omega \in \mathbb{R}^{K_1+p}, \mathbf{v} \in \mathbb{R}^{K_2+q}} \text{Cov}(\langle \omega, \alpha \rangle, \langle \mathbf{v}, \beta \rangle) = \omega'_k \Sigma_{12} \mathbf{v}_k,$$

subject to the restriction

$$\begin{aligned} \omega'_k (\Sigma_{11} + \lambda R_1) \omega_k &= 1, \\ \mathbf{v}'_k (\Sigma_{22} + \lambda R_2) \mathbf{v}_k &= 1, \end{aligned}$$

where R_1 and R_2 are given by (8) and (9) respectively, and the k th canonical variables (U_k^*, V_k^*) are not correlated with the first $k - 1$ canonical variables. The expression $(\gamma_k, \omega_k, \mathbf{v}_k)$ will be called the k th canonical system of the random vectors α and β .

Theorem 3 *The k th canonical system $(\rho_k, \mathbf{u}_k(t), \mathbf{v}_k(t))$ of the pair of random processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ is related to the k th canonical system $(\gamma_k, \omega_k, \mathbf{v}_k)$ of the pair of the random vectors α and β by the equations:*

$$\rho_k = \gamma_k, \mathbf{u}_k(t) = \Phi_1(t) \omega_k, t \in I_1, \mathbf{v}_k(t) = \Phi_2(t) \mathbf{v}_k, t \in I_2,$$

where $1 \leq k \leq \min(K_1 + p, K_2 + q)$, $K_1 = E_1 + \dots + E_p$, $K_2 = F_1 + \dots + F_q$.

Proof Without loss of generality we may assume that the covariance matrices Σ_{11} and Σ_{22} are of full column rank. As in the proof of Theorem 1 it may be assumed, that the vector weight function $\mathbf{u}(t)$ and the process $\mathbf{Y}(t)$ are in the same space, i.e. the function $\mathbf{u}(t)$ can be written in the form:

$$\mathbf{u}(t) = \Phi(t) \omega,$$

where $\omega \in \mathbb{R}^{K_1+p}$. Than

$$\langle \mathbf{u}, \mathbf{Y} \rangle = \omega' \alpha.$$

Similarly for $\mathbf{v} \in L_2^q(I_2)$

$$\langle \mathbf{v}, \mathbf{X} \rangle = \mathbf{v}' \beta,$$

where $\mathbf{v} \in \mathbb{R}^{K_2+q}$. Hence

$$\begin{aligned} E(\langle \mathbf{u}, \mathbf{Y} \rangle) &= \omega' E(\alpha) = \alpha' \mathbf{0} = 0, \\ E(\langle \mathbf{v}, \mathbf{X} \rangle) &= \mathbf{v}' E(\beta) = \beta' \mathbf{0} = 0, \\ \text{Var}(\langle \mathbf{u}, \mathbf{Y} \rangle) &= \omega' \Sigma_{11} \omega, \\ \text{Var}(\langle \mathbf{v}, \mathbf{X} \rangle) &= \mathbf{v}' \Sigma_{22} \mathbf{v}, \\ \text{Cov}(\langle \mathbf{u}, \mathbf{Y} \rangle, \langle \mathbf{v}, \mathbf{X} \rangle) &= \omega' \Sigma_{12} \mathbf{v}. \end{aligned}$$

Let us consider the first canonical correlation between the processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$:

$$\begin{aligned}\rho_1 &= \sup_{\mathbf{u} \in L_2^p(I_1), \mathbf{v} \in L_2^q(I_2)} \text{Cov}(\langle \mathbf{u}, \mathbf{Y} \rangle, \langle \mathbf{v}, \mathbf{X} \rangle) \\ &= \text{Cov}(\langle \mathbf{u}_1, \mathbf{Y} \rangle, \langle \mathbf{v}_1, \mathbf{X} \rangle),\end{aligned}$$

where $\mathbf{u}(t)$ and $\mathbf{v}(t)$ are subject to the restrictions (6) and (7). This is equivalent to saying that

$$\gamma_1 = \sup_{\boldsymbol{\omega} \in \mathbb{R}^{K_1+p}, \mathbf{v} \in \mathbb{R}^{K_2+q}} \boldsymbol{\omega}' \boldsymbol{\Sigma}_{12} \mathbf{v} = \boldsymbol{\omega}'_1 \boldsymbol{\Sigma}_{12} \mathbf{v}_1,$$

subject to the restriction

$$\begin{aligned}\boldsymbol{\omega}'_1 (\boldsymbol{\Sigma}_{11} + \lambda \mathbf{R}_1) \boldsymbol{\omega}_1 &= 1, \\ \mathbf{v}'_1 (\boldsymbol{\Sigma}_{22} + \lambda \mathbf{R}_2) \mathbf{v}_1 &= 1,\end{aligned}$$

where \mathbf{R}_1 and \mathbf{R}_2 are given by (8) and (9) respectively. This is the definition of the first canonical correlation between the random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

On the other hand, if we begin with the first canonical system $(\gamma_1, \boldsymbol{\omega}_1, \mathbf{v}_1)$ of the pair of random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we will obtain the first canonical system for the processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ from the equation

$$\rho_1 = \gamma_1, \mathbf{u}_1(t) = \boldsymbol{\Phi}_1(t) \boldsymbol{\omega}_1, \mathbf{v}_1(t) = \boldsymbol{\Phi}_2(t) \mathbf{v}_1.$$

We may extend these considerations to the second canonical system and so on. \square

Canonical correlation analysis for the random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is based on the matrices $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ and $\boldsymbol{\Sigma}_{12}$, which are unknown. We estimate them on the basis of n independent realizations $\mathbf{y}_1(t), \mathbf{y}_2(t), \dots, \mathbf{y}_n(t)$ of the form $\mathbf{y}_i(t) = \boldsymbol{\Phi}_1(t) \hat{\boldsymbol{\alpha}}_i$ of random process $\mathbf{Y}(t)$, and $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_n(t)$ of the form $\mathbf{x}_i(t) = \boldsymbol{\Phi}_2(t) \hat{\boldsymbol{\beta}}_i$ of random process $\mathbf{X}(t)$, $i = 1, 2, \dots, n$, where

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_i &= (\hat{\alpha}_{10}^{(i)}, \dots, \hat{\alpha}_{1E_1}^{(i)}, \dots, \hat{\alpha}_{p0}^{(i)}, \dots, \hat{\alpha}_{pE_p}^{(i)})', \\ \hat{\boldsymbol{\beta}}_i &= (\hat{\beta}_{10}^{(i)}, \dots, \hat{\beta}_{1F_1}^{(i)}, \dots, \hat{\beta}_{q0}^{(i)}, \dots, \hat{\beta}_{qF_q}^{(i)})'.\end{aligned}$$

Let

$$\begin{aligned}\hat{\mathbf{A}} &= (\hat{\boldsymbol{\alpha}}_1, \dots, \hat{\boldsymbol{\alpha}}_n)', \\ \hat{\mathbf{B}} &= (\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_n)'. \end{aligned}$$

Finally the estimators of the matrices Σ_{11} , Σ_{22} and Σ_{12} have the form:

$$\begin{aligned}\hat{\Sigma}_{11} &= \frac{1}{n} \hat{\mathbf{A}}' \hat{\mathbf{A}}, \\ \hat{\Sigma}_{22} &= \frac{1}{n} \hat{\mathbf{B}}' \hat{\mathbf{B}}, \\ \hat{\Sigma}_{12} &= \frac{1}{n} \hat{\mathbf{A}}' \hat{\mathbf{B}}.\end{aligned}$$

Let $\hat{\mathbf{C}} = \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12}$ and $\hat{\mathbf{D}} = \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$, where $\hat{\Sigma}_{21} = \hat{\Sigma}_{12}'$. The matrices $\hat{\mathbf{C}}\hat{\mathbf{D}}$ and $\hat{\mathbf{D}}\hat{\mathbf{C}}$ have the same nonzero eigenvalues $\hat{\gamma}_k^2$, and their corresponding eigenvectors $\hat{\mathbf{w}}_k$ and $\hat{\mathbf{v}}_k$ are given by the equations:

$$\begin{aligned}(\hat{\mathbf{C}}\hat{\mathbf{D}} - \hat{\gamma}_k^2 \mathbf{I}_{K_1+p})\hat{\mathbf{w}}_k &= \mathbf{0}, \\ (\hat{\mathbf{D}}\hat{\mathbf{C}} - \hat{\gamma}_k^2 \mathbf{I}_{K_2+q})\hat{\mathbf{v}}_k &= \mathbf{0}.\end{aligned}$$

$1 \leq k \leq \min(K_1 + p, K_2 + q)$.

Hence the coefficients of the projection of the i th realization $\mathbf{y}_i(t)$ of process $\mathbf{Y}(t)$ on the k th functional canonical variable are equal to

$$\hat{U}_{ik} = \langle \hat{\mathbf{u}}_k, \mathbf{y}_i \rangle = \int_{I_1} \hat{\mathbf{u}}_k(t) \mathbf{y}_i(t) dt = \hat{\boldsymbol{\alpha}}_i' \hat{\mathbf{w}}_k,$$

Analogously the coefficients of the projection of the i th realization $\mathbf{x}_i(t)$ of process $\mathbf{X}(t)$ on the k th functional canonical variable are equal to

$$\hat{V}_{ik} = \hat{\boldsymbol{\beta}}_i' \hat{\mathbf{v}}_k,$$

where $i = 1, 2, \dots, n$, $k = 1, \dots, \min(K_1 + p, K_2 + q)$.

6 Example

The following data (Fig. 1) come from the online database of the World Bank (<http://data.worldbank.org/>). For the analysis, fifty-four countries of the world were chosen ($n = 54$). These were recorded in the years 1972–2009 ($J = 38$). Each country belongs to one of four classes ($L = 4$):

1. Low-income economies (GDP \$1,025 or less), $n_1 = 3$
2. Lower-middle-income economies (GDP \$1,026 to \$4,035), $n_2 = 19$
3. Upper-middle-income economies (GDP \$4,036 to \$12,475), $n_3 = 14$
4. High-income economies (GDP \$12,476 or more), $n_4 = 18$

and was characterized by four variables:

1. X_1 : GDP growth (annual %)—Annual percentage growth rate of GDP at market prices based on constant local currency. Aggregates are based on constant 2000

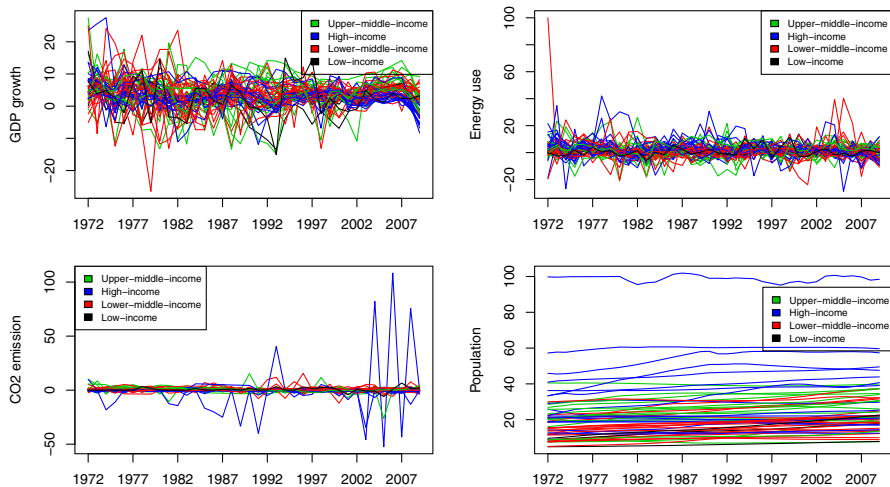


Fig. 1 Data set trajectories

U.S. dollars. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.

2. X_2 : Energy use (rate of growth in kg of oil equivalent per capita)—Energy use refers to use of primary energy before transformation to other end-use fuels, which is equal to indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport.
3. X_3 : CO₂ emissions (rate of growth in kt)—Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
4. X_4 : Population in urban agglomerations of more than 1 million (% of total population)—Population in urban agglomerations of more than one million is the percentage of a country's population living in metropolitan areas that in 2000 had a population of more than one million people.

The data were transformed to functional data by the method described in Sect. 2. The calculations were performed using the Fourier basis system what is a typical selection. But others such as splines, polynomials or wavelets can also be used. Optimum values of B , selected using the BIC criterion, for X_1 , X_2 , X_3 and X_4 take the values 2, 2, 2 and 6 respectively. The time interval $[0, T] = [0, 38]$ was divided into moments of time in the following way: $t_1 = 0.5(1972)$, $t_2 = 1.5(1973)$, ..., $t_{38} = 37.5(2009)$.

We used the R package *fda* (Ramsay et al. 2009) to create Fourier system and in order to convert raw data into a functional object. Other procedures were implemented by us (Table 1).

6.1 Multivariate functional principal component analysis (MFPCA)

The statistical objects in the functional principal component analysis are 54 countries ($n = 54$) characterized by four ($p = 4$) pieces of functional data $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), x_{i3}(t), x_{i4}(t))'$, $t \in [0, 38]$, $i = 1, 2, \dots, 54$. No account is taken of an objects membership of one of the four defined groups of countries. The vector functions $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_{54}(t)$ have the form

$$\mathbf{x}_i(t) = \Phi(t)\hat{\mathbf{c}}_i,$$

where $\Phi(t)$ is a matrix with the form of (3), and the vector $\hat{\mathbf{c}}_i$ has the form

$$\hat{\mathbf{c}}_i = (\hat{c}_{i10}(t), \dots, \hat{c}_{i1B_1}(t), \dots, \hat{c}_{i40}(t), \dots, \hat{c}_{i4B_4}(t))',$$

where $B_1 = B_2 = B_3 = 2$, $B_4 = 6$, $i = 1, 2, \dots, 54$. In the first step, from the vectors $\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, \dots, \hat{\mathbf{c}}_{54}$ we build the matrix $\hat{\Sigma}$, and next we find its eigenvalues $\hat{\gamma}_k$ and the corresponding vectors $\hat{\omega}_k$. The ratios of the particular eigenvalues to the sum of all eigenvalues, expressed as percentages, are shown in Fig. 2. It can be seen from Fig. 2 that 94.8 % of the total variation is accounted for by the first functional principal component. In the second step we form the vector weight functions

$$\hat{\mathbf{u}}_k(t) = \Phi(t)\hat{\omega}_k,$$

where $k = 1, \dots, 16$, and the corresponding functional principal components in the form

$$\hat{U}_k = \langle \hat{\mathbf{u}}_k, \mathbf{X} \rangle.$$

The graphs of the four components of the vector weight functions for the first and second functional principal components appear in Fig. 3. The values of the coefficients of the vector weight functions corresponding to the first and second functional principal components are given in Table 2. At a given time point t the greater is the absolute value of a component of the vector weight function, the greater is the contribution, in the structure of the given functional principal component, from the process $X(t)$ corresponding to that component. From Fig. 3 (left) it can be seen that the greatest contribution in the structure of the first functional principal component comes from process $X_4(t)$, and this holds for all of the observation years considered. Figure 3 (right) shows that, on specified time intervals, the greatest contribution in the structure of the second functional principal component comes alternately from the processes $X_2(t)$ and $X_1(t)$. The total contribution of a particular original process $X_i(t)$ in the structure of a given functional principal component is equal to the area under the module weighting function corresponding to this process. These contributions for the four components of the vector process $\mathbf{X}(t)$, and the first and second functional principal components are given in Table 2. The relative positions of the 54 countries in the system of the first two functional principal components are shown in Fig. 4. The system of the first two functional principal components retains 96.3 % of the total variation. From Fig. 4, we see that 54 countries form a relatively homogeneous group, with the exception of Singapore (SGP), Korea Rep. (KOR) and China (CHN).

6.2 Multivariate functional discriminant coordinates (MFDCA)

In the construction of functional discriminant coordinates, by contrast with the construction of functional principal components, we take account additionally of the information concerning the division of the 54 countries into four disjoint groups ($L = 4$). From the vectors $\hat{\mathbf{c}}_i$ we build the estimator $\hat{\mathbf{B}}$ of the matrix of between-class variation, and the estimator $\hat{\mathbf{T}}$ of the matrix of total variation, and then we find the non-zero eigenvalues $\hat{\gamma}_k$ of the matrix $\hat{\mathbf{T}}^{-1} \hat{\mathbf{B}}$ and the corresponding vectors $\hat{\boldsymbol{\omega}}_k, k = 1, 2, 3$. The ratios of particular eigenvalues to the sum of all eigenvalues, expressed as percentages, are shown in Fig. 5. It can be seen from Fig. 5 that 74.7 % of the total variation is accounted for by the first functional discriminant coordinate. In the second step we form the vector weight functions

$$\hat{\mathbf{u}}_k(t) = \Phi(t) \hat{\boldsymbol{\omega}}_k,$$

where $k = 1, 2, 3$, and the corresponding functional discriminant coordinates in the form

$$\hat{U}_k = \langle \hat{\mathbf{u}}_k, \mathbf{X} \rangle.$$

The graphs of the four components of the vector weight function for the first and second functional discriminant coordinates appear in Fig. 6. The values of the coefficients of the vector weight functions corresponding to the first and second functional discriminant coordinates are given in Table 3. At a given time point t the greater is the absolute value of a component of the vector weight function, the greater is the contribution, in the structure of the given functional discriminant coordinate, from the process $X(t)$ corresponding to that component. Figure 6 (left) shows that the greatest contribution in the structure of the first and second functional discriminant coordinates comes from process $X_4(t)$, and this holds for all of the observation years considered. Similarly as in the case of the functional principal components, the total contribution of a particular original process $X_i(t)$ in the structure of a particular functional discriminant coordinate can be estimated using the area under the module weighting function corresponding to this process. These contributions for the four components of the vector process $\mathbf{X}(t)$ and the first and second functional discriminant coordinates are given in Table 3. The relative positions of the 54 countries in the system of the first two functional discriminant coordinates are shown in Fig. 7. The system of the first two functional discriminant coordinates retains 93.9 % of the total variation. Compared with the projection onto the first two functional principal components, the division into four groups is better visible here. State clearly different from the other countries is the Democratic Republic of Kongo (COD). In the group of developed countries, Finland (FIN) is clearly different from other countries.

6.3 Multivariate functional canonical analysis (MFCCA)

In the construction of functional canonical variables we do not take account of the division of the 54 countries into four groups, and we divide the four-dimensional stochastic process into two parts: $\mathbf{Y}(t) = (X_2(t), X_3(t))'$ and $\mathbf{X}(t) = (X_1(t), X_4(t))'$.

In our case $p = q = 2$. We are interested in the relationship between the processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$. We build estimators of the matrices Σ_{11} , Σ_{22} and Σ_{12} , and we then find the non-zero eigenvalues $\hat{\gamma}_k^2$ and corresponding vectors $\hat{\omega}_k$ of the matrix $\hat{\mathbf{C}}\hat{\mathbf{D}}$, and the eigenvalues $\hat{\gamma}_k^2$ and corresponding vectors $\hat{\mathbf{v}}_k$ of the matrix $\hat{\mathbf{D}}\hat{\mathbf{C}}$, where $\hat{\mathbf{C}} = \hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$ and $\hat{\mathbf{D}} = \hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$, $\hat{\Sigma}_{21} = \hat{\Sigma}_{12}'$, $k = 1, \dots, 6$. The eigenvalues $\hat{\gamma}_k$, called canonical correlations, are shown in Fig. 8. In the second step we form vector weight functions

$$\hat{\mathbf{u}}_k(t) = \Phi_1(t)\hat{\omega}_k, \quad \hat{\mathbf{v}}_k(t) = \Phi_2(t)\hat{\mathbf{v}}_k,$$

corresponding to the processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$, where $k = 1, \dots, 6$. Corresponding to these functions are the functional canonical variables in the form

$$\hat{U}_k = \langle \hat{\mathbf{u}}_k, \mathbf{Y} \rangle, \quad \hat{V}_k = \langle \hat{\mathbf{v}}_k, \mathbf{X} \rangle,$$

corresponding to the processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$. The graphs of the two components of the vector weight function for the first and second functional canonical variables of the processes $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ are shown in Figs. 9, 10. Table 4 contains the values of the coefficients of the vector weight functions, together with the total contribution from each process in the structure of the corresponding functional canonical variable. The relative positions of the 54 countries in the systems (\hat{U}_1, \hat{V}_1) of functional canonical variables are shown in Fig. 11. The strong correlation ($\rho_1 = 0.951$) between the processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ means that in the system of canonical variables (\hat{U}_1, \hat{V}_1) points representing individual countries are almost on a straight line. In terms of the correlation between the processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, 54 countries form a relatively homogeneous group with the exception of Singapore (SGP) and Saudi Arabia (SAU).

7 Conclusions and future work

This paper introduces and analyzes a new methods of constructing canonical variables and discriminant coordinates for multivariate functional data. In addition, we reviewed principal components analysis for such data (Jacques and Preda 2014). FDA is an important tool that can be used for exploratory data analysis. A primary advantage is the ability do assess continuous data without reducing the signal into discrete variables. By representing each curve as a function, it is possible to use functional analogue of classical methods. Functional methods (1) allow more complex dynamics than classical methods; (2) they utilize a nonparametric smoothing technique to reduce the observational error; and (3) they solve the inverse and multicollinearity problems caused by the "curse of dimensionality".

Proposed methods was applied to geographic economic multivariate time series. Our research has shown, on this example, that the use of a multivariate projective dimension reduction techniques gives good results and provide an attractive method for flexibly analyse such data. Of course, the performance of the algorithms needs to be further evaluated on additional real and artificial data sets.

In a similar way, we would like to extend manifold dimension reduction techniques like multidimensional scaling (Borg and Groenen 2005), isometric feature mapping

(Tenenbaum et al. 2000) or maximum variance unfolding (Weiss 1999) for univariate functional data to multivariate case. This is the direction of our future research.

Acknowledgements We would like to thank the editor and two anonymous reviewers for the very useful comments and suggestions which help us improve the quality of our paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and Tables 1, 2, 3, 4.

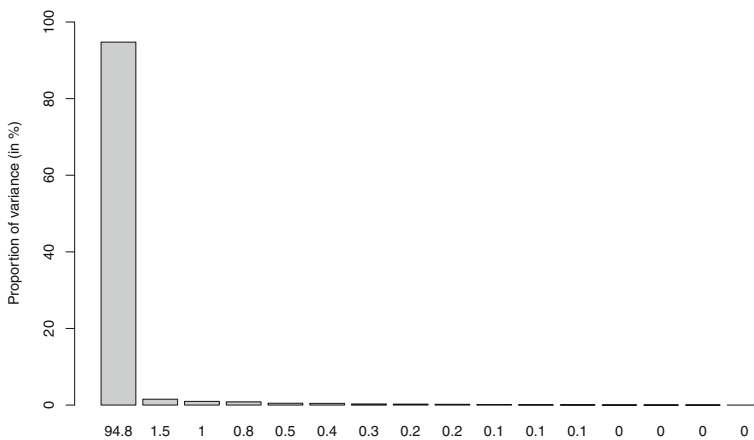


Fig. 2 Proportions of variance for multivariate functional principal components (MFPCA)

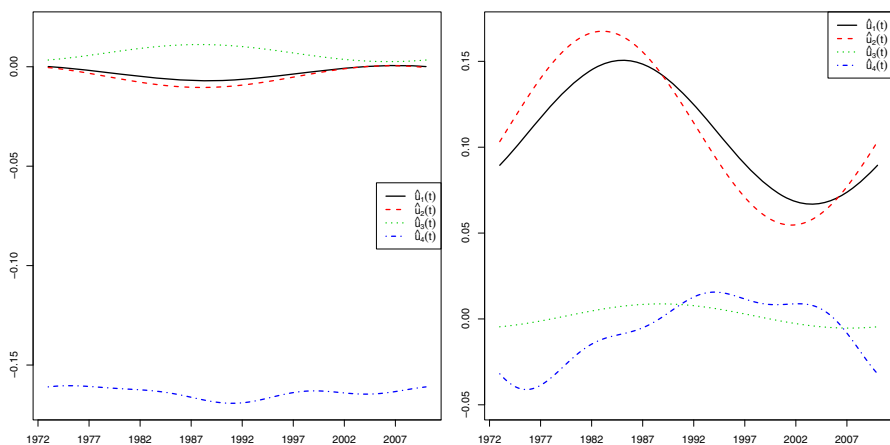


Fig. 3 Weight functions for first functional principal component \hat{U}_1 (left). Weight functions for second functional principal component \hat{U}_2 (right)

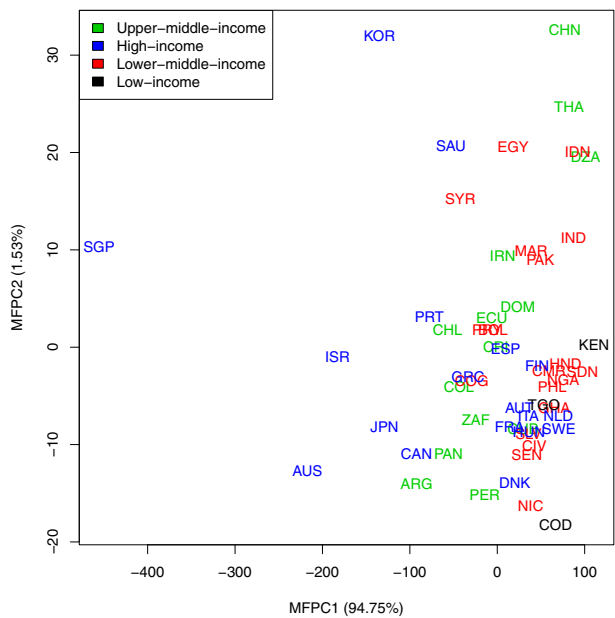


Fig. 4 Projection of the 54 selected countries on the plane (\hat{U}_1, \hat{U}_2) —multivariate functional principal components (MFPCA)

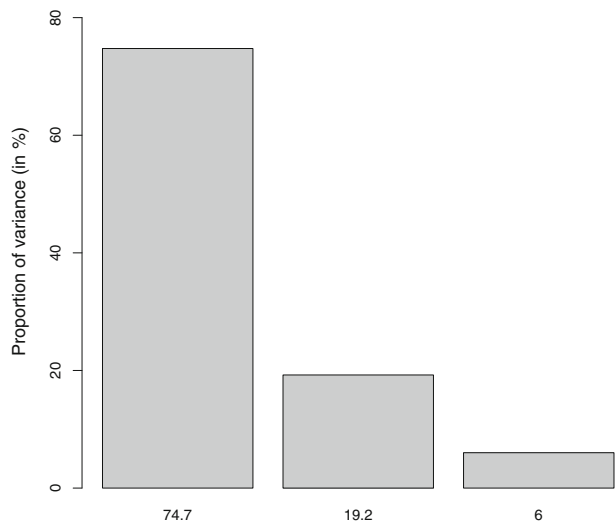


Fig. 5 Proportions of variance for multivariate functional discriminant coordinates (MFDCA)

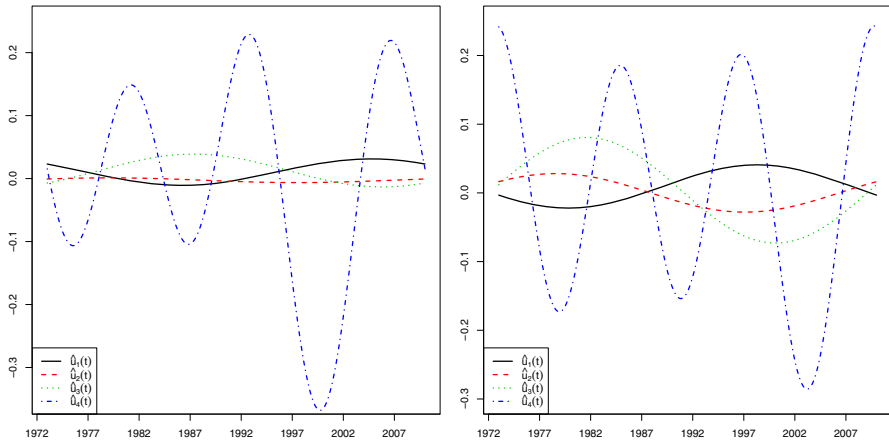


Fig. 6 Weight functions for first functional discriminant coordinate \hat{U}_1 (left). Weight functions for second functional discriminant coordinate \hat{U}_2 (right)

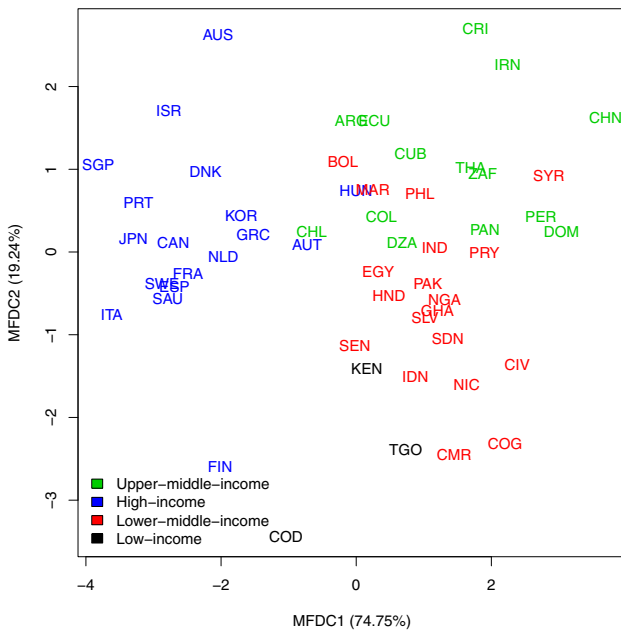


Fig. 7 Projection of the 54 selected countries on the plane (\hat{U}_1, \hat{U}_2) - multivariate functional discriminant coordinates (MFDCA)

Fig. 8 Canonical correlation for multivariate functional canonical variables (MFCCA)

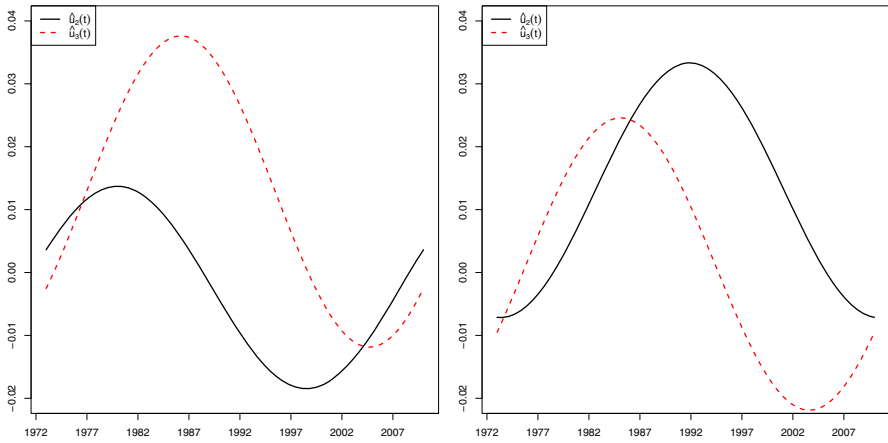
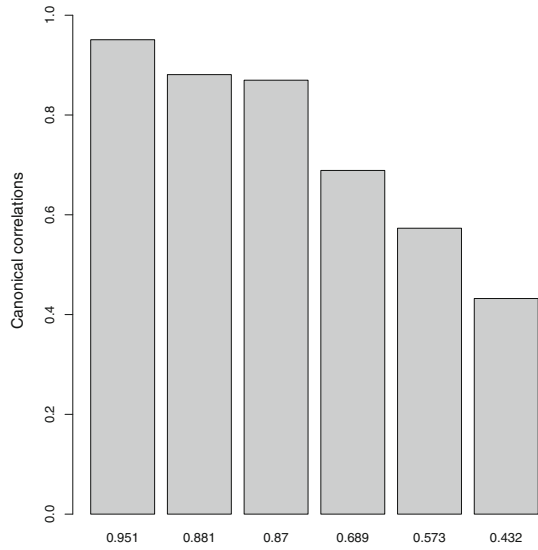


Fig. 9 Weight functions for first functional canonical variable \hat{U}_1 (left). Weight functions for second functional canonical variable \hat{U}_2 (right)

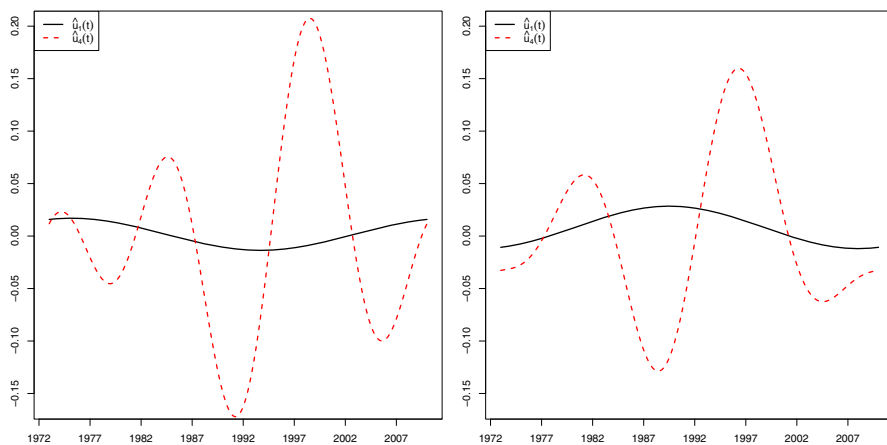


Fig. 10 Weight functions for first functional canonical variable \hat{V}_1 (left). Weight functions for second functional canonical variable \hat{V}_2 (right)

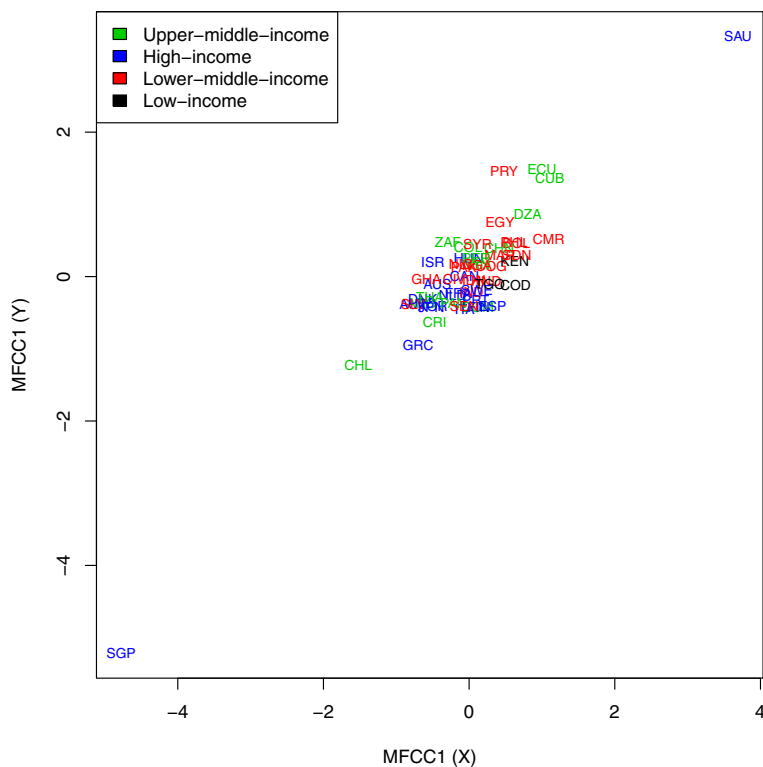


Fig. 11 Projection of the 54 selected countries on the plane (\hat{U}_1, \hat{V}_1) —multivariate functional canonical variables (MFCCA)

Table 1 Country prefix and class description

Class	Country code	Country name	Class	Country code	Country name
1	COD	Congo, Dem. Rep.	3	CRI	Costa Rica
	KEN	Kenya		CUB	Cuba
	TGO	Togo		DOM	Dominican Republic
2	BOL	Bolivia	4	ECU	Ecuador
	CMR	Cameroon		IRN	Iran, Islamic Rep.
	COG	Congo, Rep.		PAN	Panama
	CIV	Cote d'Ivoire		PER	Peru
	EGY	Egypt, Arab Rep.		ZAF	South Africa
	SLV	El Salvador		THA	Thailand
	GHA	Ghana		AUS	Australia
	HND	Honduras		AUT	Austria
	IND	India		CAN	Canada
	IDN	Indonesia		DNK	Denmark
	MAR	Morocco		FIN	Finland
	NIC	Nicaragua		FRA	France
	NGA	Nigeria		GRC	Greece
	PAK	Pakistan		HUN	Hungary
	PRY	Paraguay		ISR	Israel
	PHL	Philippines		ITA	Italy
	SEN	Senegal		JPN	Japan
	SDN	Sudan		KOR	Korea, Rep.
	SYR	Syrian Arab Republic		NLD	Netherlands
3	DZA	Algeria		PRT	Portugal
	ARG	Argentina		SAU	Saudi Arabia
	CHL	Chile		SGP	Singapore
	CHN	China		ESP	Spain
	COL	Colombia		SWE	Sweden

Table 2 Values of coefficients of the vector weight function (MFPCA)

First functional principal component								
Process	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\omega}_5$	$\hat{\omega}_6$	Area $ u(t) $
X_1	−0.019	−0.005	0.015					0.12
X_2	−0.030	−0.009	0.021					0.19
X_3	0.042	0.008	−0.017					0.25
X_4	−0.998	0.005	0.012	0.003	−0.006	0.003	0.004	6.07
Second functional principal component								
Process	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\omega}_5$	$\hat{\omega}_6$	Area $ u(t) $
X_1	0.661	0.144	−0.109					4.02
X_2	0.676	0.231	−0.074					4.11
X_3	0.010	0.009	−0.029					0.17
X_4	−0.040	−0.080	−0.066	−0.025	−0.008	−0.024	−0.003	0.57

Table 3 Values of coefficients of the vector weight function (MFDCA)

First functional discriminant coordinate								
Process	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\omega}_5$	$\hat{\omega}_6$	Area $ u(t) $
X_1	0.062	−0.061	0.066					0.55
X_2	−0.016	0.015	0.006					0.11
X_3	0.078	0.057	−0.096					0.69
X_4	−0.003	0.300	−0.013	0.002	0.469	−0.811	−0.019	4.92
Second functional discriminant coordinate								
Process	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\omega}_5$	$\hat{\omega}_6$	Area $ u(t) $
X_1	0.057	−0.131	−0.033					0.77
X_2	0.000	0.109	0.052					0.66
X_3	0.023	0.330	−0.021					1.81
X_4	0.013	0.124	−0.047	0.183	0.150	0.217	0.856	4.95

Table 4 Values of coefficients of the vector weight function (MFCCA)

First functional canonical variable ($\boldsymbol{X}(t)$)								
Process	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\omega}_5$	$\hat{\omega}_6$	Area $ u(t) $
X_1	0.010	0.034	0.056					0.36
X_4	−0.001	−0.134	0.019	0.052	−0.380	0.271	0.303	2.74
First functional canonical variable ($\boldsymbol{Y}(t)$)								
Process	\hat{v}_0	\hat{v}_1	\hat{v}_2	\hat{v}_3	\hat{v}_4	\hat{v}_5	\hat{v}_6	Area $ v(t) $
X_2	−0.014	0.068	0.015					0.38
X_3	0.078	0.071	−0.079					0.66

References

- Aneiros G, Vieu P, Aneiros G, Vieu P (2014) Variable selection in infinite-dimensional problems. *Stat Probab Lett* 94:12–20
- Berk RA (2008) *Statistical learning from a regression perspective*. Springer, New York
- Berrenderoa JR, Justela A, Svareb M (2011) Principal components for multivariate functional data. *Comput Stat Data Anal* 55:2619–2634
- Besse P (1979) *Étude descriptive d'un processus: approximation et interpolation*. Ph.D. thesis, Université Paul Sabatier (Toulouse)
- Besse P, Ramsay JO (1986) Principal components analysis of sampled functions. *Psychometrika* 51(2):285–311
- Bongiorno EG, Goia A, Salinelli E (2014) *Contributions in infinite-dimensional statistics and related topics*. Societa Editrice Esculapio, Bologna
- Borg I, Groenen P (2005) *Modern multidimensional scaling: theory and applications*. Springer, New York
- Brillinger DR (2001) *Time series: data analysis and theory*. Society for Industrial and Applied Mathematics, Philadelphia
- Burges CJC (2009) Dimension reduction: a guided tour. *Found Trends Mach Learn* 2(4):275–365
- Chamroukh F, Glotin H, Samé A (2013) Model-based functional mixture discriminant analysis with hidden process regression for curve classification. *Neurocomputing* 18:153–163
- Delaigle A, Hall P (2012) Achieving near perfect classification for functional data. *J R Stat Soc* 74(2):267–286
- Dubin JA, Müller HG (2005) Dynamical correlation for multivariate longitudinal data. *J Am Stat Assoc* 100(471):872–881
- Ferraty F, Vieu P (2006) *Nonparametric functional data analysis*. Springer, New York
- Ferraty F, Gonzalez-Manteiga W, Martinez-Calvo A, Vieu P (2012) Presmoothing in functional linear regression. *Stat Sin* 22:69–94
- Fisher RA (1936) The use of multiple measurements in taxonomic problem. *Ann Eugen* 7:179–188
- Goia A, Vieu P (2014) Some advances on semi-parametric functional data modelling. In: *Contributions in infinite-dimensional statistics and related topics*, Esculapio, Bologna
- Górecki T, Krzyśko M (2012) Functional principal components analysis. In: Pociecha J, Decker R (eds) *Data analysis methods and its applications*. C.H. Beck, Munich, pp 71–87
- Górecki T, Krzyśko M, Waszak Ł (2014) Functional discriminant coordinates. *Commun Stat—Theory Methods* 43(5):1013–1025
- He G, Müller HG, Wang JL (2000) Extending correlation and regression from multivariate to functional data. *Asymptotics in statistics and probability*. VSP, Zeist, pp 197–210
- He G, Müller HG, Wang JL (2004) Methods of canonical analysis for functional data. *J Stat Plan Inference* 122:141–159
- Horváth L, Kokoszka P (2012) *Inference for functional data with applications*. Springer, New York

- Jacques J, Preda C (2013) Funclust: a curves clustering method using functional random variables density approximation. *Neurocomputing* 112:164–171
- Jacques J, Preda C (2014) Model-based clustering for multivariate functional data. *Comput Stat Data Anal* 71:92–106
- James GM, Sugar CA (2003) Clustering for sparsely sampled functional data. *J Am Stat Assoc* 98(462):397–408
- Jolliffe IT (2002) *Principal component analysis*, 2nd edn. Springer, New York
- Kudraszow NL, Vieu P (2013) Uniform consistency of kNN regressors for functional variables. *Stat Probab Lett* 83(8):1863–1870
- Leurgans SE, Moyeed RA, Silverman BW (1993) Canonical correlation analysis when the data are curves. *J R Stat Soc* 55(3):725–740
- Mosler K, Mozharovskyi P (2015) Fast DD-classification of functional data. *Stat Papers*, doi:[10.1007/s00362-015-0738-3](https://doi.org/10.1007/s00362-015-0738-3)
- Panaretos VM, Kraus D, Maddocks JH (2010) Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J Am Stat Assoc* 105(490):670–682
- Peng J, Müller HG (2008) Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *Ann Appl Stat* 2(3):1056–1077
- Peng Q, Zhou J, Tang N (2015) Varying coefficient partially functional linear regression models. *Stat Papers*, doi:[10.1007/s00362-015-0681-3](https://doi.org/10.1007/s00362-015-0681-3)
- Rachdi M, Vieu P (2006) Nonparametric regression for functional data: automatic smoothing parameter selection. *J Stat Plan Inference* 137:2784–2801
- Ramsay JO, Silverman BW (2005) *Functional data analysis*, 2nd edn. Springer, New York
- Ramsay JO, Hooker G, Graves S (2009) *Functional data analysis with R and MATLAB*. Springer, New York
- Rossi F, Villa N (2006) Support vector machine for functional data classification. *Neurocomputing* 69(7–9):730–742
- Saporta G (1981) *Méthodes exploratoires d'analyse de données temporelles*. Ph.D. thesis, Université Pierre et Marie Curie (Paris)
- Shmueli G (2010) To explain or to predict? *Stat Sci* 25(3):289–310
- Sober E (2002) Instrumentalism, parsimony, and the Akaike framework. *Philos Sci* 69:112–123
- Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323
- Wang G, Zhou J, Wu W, Chen M (2015) Robust functional sliced inverse regression. *Stat Papers*, doi:[10.1007/s00362-015-0695-x](https://doi.org/10.1007/s00362-015-0695-x)
- Weiss Y (1999) Segmentation using eigenvectors: a unifying view. In: *Proceedings of the IEEE international conference on computer vision*. Los Alamitos, pp 975–982