STATISTICS IN TRANSITION new series, September 2016
Vol. 17, No. 3, pp. 449–466

AN EXTENSION OF THE CLASSICAL DISTANCE CORRELATION COEFFICIENT FOR MULTIVARIATE FUNCTIONAL DATA WITH APPLICATIONS

Tomasz Górecki[1], **Mirosław Krzyśko**[2], **Waldemar Ratajczak**[3], **Waldemar Wołyński**[4]

ABSTRACT

The relationship between two sets of real variables defined for the same individuals can be evaluated by a few different correlation coefficients. For the functional data we have one important tool: canonical correlations. It is not immediately straightforward to extend other similar measures to the context of functional data analysis. In this work we show how to use the distance correlation coefficient for a multivariate functional case.
The approaches discussed are illustrated with an application to some socio-economic data.

Key words: multivariate functional data, functional data analysis, correlation analysis.

1. Introduction

In recent years methods for data representing functions or curves have received much attention. Such data are known in the literature as functional data (Ramsay & Silverman (2005), Horváth & Kokoszka (2012)). Examples of functional data can be found in several application domains, such as medicine, economics, meteorology and many others. In a great number of applications it is necessary to use statistical methods for objects characterized by many features observed in many time points (double multivariate data). In this case such data are called multivariate functional data. The pioneering theoretical paper was Besse (1979), in which random variables have values in a general Hilbert space. Berrendero et al. (2011), Górecki et

[1]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. Department of Statistics, Colorado State University, USA. E-mail: tomasz.gorecki@amu.edu.pl
[2]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: mkrzysko@amu.edu.pl
[3]Faculty of Geographical and Geological Sciences, Adam Mickiewicz University, Poland. E-mail: walrat@amu.edu.pl
[4]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: wolynski@amu.edu.pl

al. (2014) and Jacques & Preda (2014), present an analysis of multivariate functional data from the point of view of Multivariate Principal Component Analysis (MPCA). Also functional regression models have been extensively studied; see for example James (2002), Müller and Stadmüller (2005), Reiss and Ogden (2007) and Matsui et al. (2008). Various basic classification methods have been adapted to functional data, such as linear discriminant analysis (Hastie et al. (1995)), logistic regression (Rossi et al. (2002)), penalized optimal scoring (Ando (2009)), *k*nn (Ferraty and Vieu (2003)), SVM (Rossi and Villa (2006)), and neural networks (Rossi et al. (2005)). Moreover, the theory of combining classifiers has been extended to functional data (Ferraty and Vieu (2009)). Górecki et al. (2015) discussed the problem of classification via regression for multivariate functional data.

In this paper we focus on correlation analysis for multivariate functional data. In the literature, there are different strategies to explore the association between two sets of variables (p dimensional $\boldsymbol{X}$ and q dimensional $\boldsymbol{Y}$). Historically, the first approach was put forward by Hotelling (1936), who proposed the canonical correlation in the framework of Canonical Correlation Analysis (CCA). The CCA is a reference tool concerned with describing linear dependencies between two sets of variables; it seeks a linear combination of the variables of the first group which is maximally correlated with a linear combination of the variables of the second group. The correlation coefficient thus obtained is said to be canonical and the linear combinations are called canonical variables. Leurgans et al. (1993), He et al. (2004), Krzyśko & Waszak (2013) discussed this analysis in the context of functional data.

Another approach is to consider each set of variables through its individual cloud, and to compare the structures (i.e. the shapes) of the two point clouds. In this way, the so-called rV coefficient (Escoufier (1970, 1973), Robert & Escoufier (1976), Escoufier & Robert (1979)) provides an insight into the global association between the two sets of variables.

Székely et al. (2007), Székely & Rizzo (2009, 2012, 2013) defined a measure of dependence between random vectors: the distance correlation (dCor) coefficient. The authors showed that for all random variables with finite first moments, the dCor coefficient generalizes the idea of correlation in two ways. Firstly, this coefficient can be applied when $\boldsymbol{X}$ and $\boldsymbol{Y}$ are of any dimensions and not only for the simple case where $p = q = 1$. They constructed their coefficient as a generalization of the simple correlation coefficient without reference to the earlier literature on the rV coefficient. Secondly, the dCor coefficient is equal to zero if and only if there is independence between the random vectors. Indeed, a correlation coefficient measures linear relationships and can be equal to 0 even when the random variables are dependent. This can be seen as a major shortcoming of the correlation coefficient and the rV coefficient.

The rest of this paper is organized as follows. We first review the concept of transformation of discrete data into multivariate functional data (Section 2). Section 3 contains the functional version of canonical correlation coefficients analysis. Section 4 describes our extension of the distance correlation coefficient to the functional case. In Section 5 the accuracy of the proposed methods is demonstrated using some empirical data. Conclusions are given in Section 6.

2. Smoothing of stochastic processes

Let us assume that $\boldsymbol{X} \in L_2^p(I_1)$ and $\boldsymbol{Y} \in L_2^q(I_2)$ are random processes, where $L_2(I)$ is a Hilbert space of square integrable functions on the interval I.

We also assume that $\mathrm{E}(\boldsymbol{X}(s)) = \mathbf{0},\ s \in I_1$ and $\mathrm{E}(\boldsymbol{Y}(t)) = \mathbf{0},\ t \in I_2$.

This fact does not cause loss of generality, because functional correlation coefficients are calculated on the basis of the covariance functions of processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ of the form

$$\mathrm{Cov}\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \boldsymbol{\Sigma}(s,t) = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{XX}}(s,t) & \boldsymbol{\Sigma}_{\boldsymbol{XY}}(s,t) \\ \boldsymbol{\Sigma}_{\boldsymbol{YX}}(s,t) & \boldsymbol{\Sigma}_{\boldsymbol{YY}}(s,t) \end{bmatrix},\ s \in I_1,\ t \in I_2,$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{XX}}(s,t) &= \mathrm{E}[\boldsymbol{X}(s)\boldsymbol{X}'(t)],\ s,t \in I_1, \\ \boldsymbol{\Sigma}_{\boldsymbol{XY}}(s,t) &= \mathrm{E}[\boldsymbol{X}(s)\boldsymbol{Y}'(t)],\ s \in I_1,\ t \in I_2, \\ \boldsymbol{\Sigma}_{\boldsymbol{YX}}(s,t) &= \mathrm{E}[\boldsymbol{Y}(s)\boldsymbol{X}'(t)],\ s \in I_2,\ t \in I_1, \\ \boldsymbol{\Sigma}_{\boldsymbol{YY}}(s,t) &= \mathrm{E}[\boldsymbol{Y}(s)\boldsymbol{Y}'(t)],\ s,t \in I_2. \end{aligned}$$

We will further assume that each component X_g of process $\boldsymbol{X}$ and Y_h of process $\boldsymbol{Y}$ can be represented by a finite number of orthonormal basis functions $\{\varphi_e\}$ and $\{\varphi_f\}$ of space $L_2(I_1)$ and $L_2(I_2)$, respectively:

$$\begin{aligned} X_g(s) &= \sum_{e=0}^{E_g} \alpha_{ge}\varphi_e(s), s \in I_1, g = 1,2,...,p, \\ Y_h(t) &= \sum_{f=0}^{F_h} \beta_{hf}\varphi_f(t), t \in I_2, h = 1,2,...,q. \end{aligned}$$

The choice of the basis seems not crucial. We can use various orthonormal basis, but Fourier basis seems the most appropriate in most cases (Górecki & Krzyśko

(2012)) and the most common in practice. The degree of smoothness of functions X_g and Y_h depends on values E_g and F_h respectively (small values cause more smoothing). The choice of the truncation parameters is critical for the proper representation of general stochastic process. The optimal number of basis elements could be determined using the Bayesian Information Criterion (BIC) for each function separately through finding the most frequent value (modal value) over all functions. We should prefer this value to be large, particularly when the stochastic processes are observed at high frequency with little noise.

We introduce the following notation:

$$\boldsymbol{\alpha} = (\alpha_{10}, ..., \alpha_{1E_1}, ..., \alpha_{p0}, ..., \alpha_{pE_p})',$$
$$\boldsymbol{\beta} = (\beta_{10}, ..., \beta_{1F_1}, ..., \beta_{q0}, ..., \beta_{qF_q})',$$

$$\boldsymbol{\varphi}_{E_g}(s) = (\varphi_0(s), \dots, \varphi_{E_g}(s))', s \in I_1, g = 1, 2, ..., p,$$
$$\boldsymbol{\varphi}_{F_h}(t) = (\varphi_0(t), \dots, \varphi_{E_g}(t))', t \in I_2, h = 1, 2, ..., q,$$

$$\boldsymbol{\Phi}_1(s) = \begin{bmatrix} \boldsymbol{\varphi}'_{E_1}(s) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}'_{E_2}(s) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}'_{\mathrm{E}_p}(s) \end{bmatrix},$$
$$\boldsymbol{\Phi}_2(t) = \begin{bmatrix} \boldsymbol{\varphi}'_{F_1}(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}'_{F_2}(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}'_{F_q}(t) \end{bmatrix}.$$

Using the above matrix notation, processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be represented as:

$$\boldsymbol{X}(s) = \boldsymbol{\Phi}_1(s)\boldsymbol{\alpha}, \quad \boldsymbol{Y}(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{\beta}.$$

This means that the realizations of processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ are in finite dimensional subspaces of $L_2^p(I_1)$ and $L_2^q(I_2)$ respectively. We will denote these subspaces by $\mathscr{L}_2^p(I_1)$ and $\mathscr{L}_2^q(I_2)$.

For random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ we have:

$$\mathrm{E}(\boldsymbol{\alpha}) = \mathbf{0}, \ \mathrm{E}(\boldsymbol{\beta}) = \mathbf{0}$$

and

$$\text{Cov}\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{bmatrix} = \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\alpha\alpha}} & \boldsymbol{\Sigma}_{\boldsymbol{\alpha\beta}}\\ \boldsymbol{\Sigma}_{\boldsymbol{\beta\alpha}} & \boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}}\end{bmatrix},$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha\alpha}} = \text{E}(\boldsymbol{\alpha\alpha}')$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha\beta}} = \text{E}(\boldsymbol{\alpha\beta}')$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta\alpha}} = \text{E}(\boldsymbol{\beta\alpha}')$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}} = \text{E}(\boldsymbol{\beta\beta}')$.

Note that

$$\boldsymbol{\Sigma}_{\boldsymbol{XX}}(s,t) = \text{E}[\boldsymbol{\Phi}_1(s)\boldsymbol{\alpha\alpha}'\boldsymbol{\Phi}_1'(t)] = \boldsymbol{\Phi}_1(s)\,\text{E}(\boldsymbol{\alpha\alpha}')\boldsymbol{\Phi}_1'(t) = \boldsymbol{\Phi}_1(s)\boldsymbol{\Sigma}_{\boldsymbol{\alpha\alpha}}\boldsymbol{\Phi}_1'(t).$$

Similarly

$$\boldsymbol{\Sigma}_{\boldsymbol{XY}}(s,t) = \boldsymbol{\Phi}_1(s)\boldsymbol{\Sigma}_{\boldsymbol{\alpha\beta}}\boldsymbol{\Phi}_2'(t),$$

$$\boldsymbol{\Sigma}_{\boldsymbol{YX}}(s,t) = \boldsymbol{\Phi}_2(s)\boldsymbol{\Sigma}_{\boldsymbol{\beta\alpha}}\boldsymbol{\Phi}_1'(t),$$

$$\boldsymbol{\Sigma}_{\boldsymbol{YY}}(s,t) = \boldsymbol{\Phi}_2(s)\boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}}\boldsymbol{\Phi}_2'(t).$$

In fact, the correlation analysis for random processes is based on matrices $\boldsymbol{\Sigma}_{\boldsymbol{\alpha\alpha}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha\beta}}$ which are unknown. We have to estimate them on the basis of n independent realizations $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ and $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_n$ of random processes $\boldsymbol{X}$ and $\boldsymbol{Y}$. We have $\boldsymbol{X}_i(s) = \boldsymbol{\Phi}_1(s)\boldsymbol{\alpha}_i$ and $\boldsymbol{Y}_i(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{\beta}_i$, $i = 1,2,\ldots,n$. This problem has been extensively studied in the literature, e.g. Beutler (1970), Lee (1976) and Masry (1978).

Typically, data are recorded at discrete moments in time. The transformation of discrete data into functional data is performed for each realization and each variable separately. Let x_{gj} denote an observed value of feature X_g, $g = 1,2,\ldots p$ at the jth time point s_j, where $j = 1,2,...,J$. Similarly, let y_{hj} denote an observed value of feature Y_h, $h = 1,2,\ldots q$ at the jth time point t_j, where $j = 1,2,...,J$. Then, our data consist of pJ pairs of (s_j, x_{gj}) and of qJ pairs of (t_j, y_{hj}).

The coefficients $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are estimated by the least squares method. Let us denote these estimates by $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$, $i = 1,2,\ldots,n$.

As a result of the transformation process, we obtain functional data of the form:

$$\boldsymbol{x}_i(s) = \boldsymbol{\Phi}_1(s)\boldsymbol{a}_i,\ \boldsymbol{y}_i(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{b}_i, \tag{1}$$

where $s \in I_1$, $t \in I_2$ and $i = 1,2,\ldots,n$.

Let $\boldsymbol{A} = (\boldsymbol{a}_1', \boldsymbol{a}_2', \ldots, \boldsymbol{a}_n')'$, and $\boldsymbol{B} = (\boldsymbol{b}_1', \boldsymbol{b}_2', \ldots, \boldsymbol{b}_n')'$. Then

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\alpha}} = \frac{1}{n}\boldsymbol{A}'\boldsymbol{A}, \quad \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta\beta}} = \frac{1}{n}\boldsymbol{B}'\boldsymbol{B}, \quad \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\beta}} = \frac{1}{n}\boldsymbol{A}'\boldsymbol{B}. \tag{2}$$

3. Functional canonical correlation coefficient

Functional canonical variables U and V for stochastic processes $\boldsymbol{X} \in \mathcal{L}_2^p(I_1)$ and $\boldsymbol{Y} \in \mathcal{L}_2^q(I_2)$ are defined as follows:

$$U(s) = \boldsymbol{l}'(s)\boldsymbol{X}(s),\ \boldsymbol{l} \in \mathcal{L}_2^p(I_1),$$
$$V(t) = \boldsymbol{m}'(t)\boldsymbol{Y}(t),\ \boldsymbol{m} \in \mathcal{L}_2^q(I_2),$$

where $\boldsymbol{l}$ and $\boldsymbol{m}$ are weight functions.

We have

$$\mathrm{E}(U(s)) = \mathrm{E}(V(t)) = 0,\ s \in I_1,\ t \in I_2.$$

Let us denote the covariance matrix of processes U and V by

$$\boldsymbol{\Sigma}_{UV}(s,t) = \begin{bmatrix} \sigma_{UU}(s,t) & \sigma_{UV}(s,t) \\ \sigma_{VU}(s,t) & \sigma_{VV}(s,t) \end{bmatrix}.$$

Because $\boldsymbol{l} \in \mathcal{L}_2^p(I_1)$ and $\boldsymbol{m} \in \mathcal{L}_2^q(I_2)$ we have

$$\boldsymbol{l}(s) = \boldsymbol{\Phi}_1(s)\boldsymbol{\lambda},\ \boldsymbol{m}(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{\mu},$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{K_1+p}$, $\boldsymbol{\mu} \in \mathbb{R}^{K_2+q}$ and $K_1 = E_1 + ... + E_p$, $K_2 = F_1 + ... + F_q$.

Moreover

$$\begin{aligned} \sigma_{UU}(s,t) &= \mathrm{E}[U(s)U'(t)] = \mathrm{E}[\boldsymbol{l}'(s)\boldsymbol{X}(s)\boldsymbol{X}'(t)\boldsymbol{l}(t)] \\ &= \mathrm{E}[\boldsymbol{\lambda}'\boldsymbol{\Phi}_1'(s)\boldsymbol{\Phi}_1(s)\boldsymbol{\alpha}\boldsymbol{\alpha}'\boldsymbol{\Phi}_1'(t)\boldsymbol{\Phi}_1(t)\boldsymbol{\lambda}] \\ &= \boldsymbol{\lambda}'\boldsymbol{\Phi}_1'(s)\boldsymbol{\Phi}_1(s)\boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\boldsymbol{\Phi}_1'(t)\boldsymbol{\Phi}_1(t)\boldsymbol{\lambda}. \end{aligned}$$

Similarly

$$\begin{aligned} \sigma_{UV}(s,t) &= \boldsymbol{\lambda}'\boldsymbol{\Phi}_1'(s)\boldsymbol{\Phi}_1(s)\boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\boldsymbol{\Phi}_2'(t)\boldsymbol{\Phi}_2(t)\boldsymbol{\mu}, \\ \sigma_{VU}(s,t) &= \boldsymbol{\mu}'\boldsymbol{\Phi}_2'(s)\boldsymbol{\Phi}_2(s)\boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\alpha}}\boldsymbol{\Phi}_1'(t)\boldsymbol{\Phi}_1(t)\boldsymbol{\lambda}, \\ \sigma_{VV}(s,t) &= \boldsymbol{\mu}'\boldsymbol{\Phi}_2'(s)\boldsymbol{\Phi}_2(s)\boldsymbol{\Sigma}_{\boldsymbol{\beta}\boldsymbol{\beta}}\boldsymbol{\Phi}_2'(t)\boldsymbol{\Phi}_2(t)\boldsymbol{\mu}. \end{aligned}$$

The functional canonical coefficient $\rho_{\boldsymbol{X},\boldsymbol{Y}}$ is defined as

$$\rho_{\boldsymbol{X},\boldsymbol{Y}} = \max_{\boldsymbol{l},\boldsymbol{m}} \int_{I_1}\int_{I_2} \sigma_{UV}(s,t)dsdt = \max_{\boldsymbol{l},\boldsymbol{m}} \int_{I_1}\int_{I_2} \mathrm{E}[\boldsymbol{l}'(s)\boldsymbol{X}(s)\boldsymbol{Y}'(t)\boldsymbol{m}(t)]dsdt,$$

subject to the constraint

$$\int_{I_1}\int_{I_2}\sigma_{UU}(s,t)dsdt = \int_{I_1}\int_{I_2}\sigma_{VV}(s,t)dsdt = 1.$$

Because

$$\int_{I_1}\int_{I_2}\sigma_{UV}(s,t)dsdt = \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\alpha\beta}}\boldsymbol{\mu},$$
$$\int_{I_1}\int_{I_2}\sigma_{UU}(s,t)dsdt = \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\alpha\alpha}}\boldsymbol{\lambda},$$
$$\int_{I_1}\int_{I_2}\sigma_{VV}(s,t)dsdt = \boldsymbol{\mu}'\boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}}\boldsymbol{\mu}$$

we have

$$\rho_{\boldsymbol{X},\boldsymbol{Y}} = \max_{\boldsymbol{\lambda},\boldsymbol{\mu}}\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\alpha\beta}}\boldsymbol{\mu} = \rho_{\boldsymbol{\alpha},\boldsymbol{\beta}},$$

subject to the restriction

$$\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\alpha\alpha}}\boldsymbol{\lambda} = \boldsymbol{\mu}'\boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}}\boldsymbol{\mu} = 1.$$

From the above we see that the functional canonical correlation coefficient $\rho_{\boldsymbol{X},\boldsymbol{Y}}$ of the pair of random processes $\boldsymbol{X}\in\mathscr{L}_2^p(I_1)$ and $\boldsymbol{Y}\in\mathscr{L}_2^q(I_2)$ is equivalent to the canonical correlation coefficient $\rho_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ of the pair of the random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Note that if $\boldsymbol{X}\in L_2^p(I_1)$ and $\boldsymbol{Y}\in L_2^q(I_2)$ there exist weight functions such that the functional canonical coefficient is equal to one. This means that, with an increasing size of a number of basis functions, the functional canonical coefficient will tend to one. To avoid this problem Leurgans et al. (1993) proposed some additional regularization. However, as for many correlation coefficients, it is difficult to evaluate the magnitude of the relationship just by considering its values.

The canonical correlation coefficient $\rho_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ of the pair of random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is based on matrices $\boldsymbol{\Sigma}_{\boldsymbol{\alpha\alpha}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha\beta}}$. If they are not known, we have to use their estimators (2).

Hence

$$\hat{\rho}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \max_{\boldsymbol{\lambda},\boldsymbol{\mu}}\boldsymbol{\lambda}'\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\beta}}\boldsymbol{\mu},$$

under the condition

$$\boldsymbol{\lambda}'\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\alpha}}\boldsymbol{\lambda} = \boldsymbol{\mu}'\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta\beta}}\boldsymbol{\mu} = 1.$$

4. Functional distance correlation

First, let us define the joint characteristic function of the pair of random processes $(\boldsymbol{X},\boldsymbol{Y})$. If for all functions $\boldsymbol{l}\in L_2^p(I_1)$ the integral $\int_{I_1}\boldsymbol{l}'(s)\boldsymbol{X}(s)ds$ converges for almost all realization of $\boldsymbol{X}$, and for all functions $\boldsymbol{m} \in L_2^q(I_1)$ the integral $\int_{I_2} \boldsymbol{m}'(t)\boldsymbol{Y}(t)dt$ converges for almost all realizations of $\boldsymbol{Y}$, then the characteristic function of the pair of random processes $(\boldsymbol{X},\boldsymbol{Y})$ has the following form:

$$\varphi_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{l},\boldsymbol{m}) = \mathrm{E}\{\exp[i < \boldsymbol{l},\boldsymbol{X} >_p + i < \boldsymbol{m},\boldsymbol{Y} >_q]\},$$

where

$$< \boldsymbol{l},\boldsymbol{X} >_p = \int_{I_1} \boldsymbol{l}'(s)\boldsymbol{X}(s)ds, \quad < \boldsymbol{m},\boldsymbol{Y} >_q = \int_{I_2} \boldsymbol{m}'(t)\boldsymbol{Y}(t)dt$$

and $i^2 = -1$. Moreover, we define the marginal characteristic function of $\boldsymbol{X}$ and $\boldsymbol{Y}$ as follows: $\varphi_{\boldsymbol{X}}(\boldsymbol{l}) = \varphi_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{l},\boldsymbol{0})$ and $\varphi_{\boldsymbol{Y}}(\boldsymbol{m}) = \varphi_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{0},\boldsymbol{m})$.

Now, let us assume that $\boldsymbol{X} \in \mathscr{L}_2^p(I_1)$ and $\boldsymbol{Y} \in \mathscr{L}_2^q(I_2)$. Then, the processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be represented as:

$$\boldsymbol{X}(s) = \boldsymbol{\Phi}_1(s)\boldsymbol{\alpha}, \quad \boldsymbol{Y}(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{\beta},$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{K_1+p}$ and $\boldsymbol{\beta} \in \mathbb{R}^{K_2+q}$.

In this case, we may assume (Ramsey & Silverman (2005)) that the vector function $\boldsymbol{l}$ and the process $\boldsymbol{X}$ are in the same space, i.e. function $\boldsymbol{l}$ can be written in the form

$$\boldsymbol{l}(s) = \boldsymbol{\Phi}_1(s)\boldsymbol{\lambda},$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{K_1+p}$.

We may assume the same for the vector function $\boldsymbol{m}$ and the process $\boldsymbol{Y}$. Then, we have

$$\boldsymbol{m}(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{\mu},$$

where $\boldsymbol{\mu} \in \mathbb{R}^{K_2+q}$.

Hence

$$< \boldsymbol{l},\boldsymbol{X} >_p = \int_{I_1} \boldsymbol{l}'(s)\boldsymbol{X}(s)ds = \boldsymbol{\lambda}'[\int_{I_1} \boldsymbol{\Phi}_1'(s)\boldsymbol{\Phi}_1(s)ds]\boldsymbol{\alpha} = \boldsymbol{\lambda}'\boldsymbol{\alpha}$$

and

$$< \boldsymbol{m},\boldsymbol{Y} >_q = \int_{I_2} \boldsymbol{m}'(t)\boldsymbol{Y}(t)dt = \boldsymbol{\mu}'[\int_{I_2} \boldsymbol{\Phi}_2'(t)\boldsymbol{\Phi}_2(t)dt]\boldsymbol{\beta} = \boldsymbol{\mu}'\boldsymbol{\beta}$$

then

$$\varphi_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{l},\boldsymbol{m}) = \mathrm{E}\{\exp[i\boldsymbol{\lambda}'\boldsymbol{\alpha} + i\boldsymbol{\mu}'\boldsymbol{\beta}]\} = \varphi_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{\lambda},\boldsymbol{\mu}),$$

where $\varphi_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{\lambda},\boldsymbol{\mu})$ is the joint characteristic function of the pair of random vectors

$(\boldsymbol{\alpha},\boldsymbol{\beta})$.

On the basis of the idea of distance covariance between two random vectors (Székely et al. (2007)), we can introduce functional distance covariance between random processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ as a nonnegative number $v_{\boldsymbol{X},\boldsymbol{Y}}$ defined by

$$v_{\boldsymbol{X},\boldsymbol{Y}} = v_{\boldsymbol{\alpha},\boldsymbol{\beta}},$$

where

$$v^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{1}{C_{K_1+p}C_{K_2+q}} \int_{\mathbb{R}^{K_1+K_2+p+q}} \frac{|\phi_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{\lambda},\boldsymbol{\mu}) - \phi_{\boldsymbol{\alpha}}(\boldsymbol{\lambda})\phi_{\boldsymbol{\beta}}(\boldsymbol{\mu})|^2}{\|\boldsymbol{\lambda}\|^{K_1+p+1}_{K_1+p}\|\boldsymbol{\mu}\|^{K_2+q+1}_{K_2+q}} d\boldsymbol{\lambda} d\boldsymbol{\mu},$$

and $|z|$ denotes the modulus of $z \in \mathbb{C}$, $\|\boldsymbol{\lambda}\|_{K_1+p}$, $\|\boldsymbol{\mu}\|_{K_2+q}$ the standard Euclidean norms on the corresponding spaces, and

$$C_r = \frac{\pi^{\frac{1}{2}(r+1)}}{\Gamma(\frac{1}{2}(r+1))}.$$

The functional distance correlation between random processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ is a nonnegative number defined by

$$\mathscr{R}_{\boldsymbol{X},\boldsymbol{Y}} = \frac{v_{\boldsymbol{X},\boldsymbol{Y}}}{\sqrt{v_{\boldsymbol{X},\boldsymbol{X}} v_{\boldsymbol{Y},\boldsymbol{Y}}}}$$

if both $v_{\boldsymbol{X},\boldsymbol{X}}$ and $v_{\boldsymbol{Y},\boldsymbol{Y}}$ are strictly positive, and zero otherwise. For distributions with finite first moments, distance correlation characterizes independence in that $0 \leq \mathscr{R}_{\boldsymbol{X},\boldsymbol{Y}} \leq 1$ with $\mathscr{R}_{\boldsymbol{X},\boldsymbol{Y}} = 0$ if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent.

We can estimate functional distance covariance using functional data of the form (1).

On the basis of the result of Székely et al. (2007), we have

$$\hat{v}^2_{\boldsymbol{X},\boldsymbol{Y}} = \frac{1}{n^2} \sum_{k,l=1}^{n} A_{kl} B_{kl},$$

where $a_{kl} = \|\boldsymbol{a}_k - \boldsymbol{a}_l\|_{K_1+p}$, $\bar{a}_{k\cdot} = \frac{1}{n}\sum_{l=1}^{n} a_{kl}$, $\bar{a}_{\cdot l} = \frac{1}{n}\sum_{k=1}^{n} a_{kl}$, $\bar{a}_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}$ and $A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}$, and similarly for $b_{kl} = \|\boldsymbol{b}_k - \boldsymbol{b}_l\|_{K_2+q}$, $\bar{b}_{k\cdot}$, $\bar{b}_{\cdot l}$, $\bar{b}_{\cdot\cdot}$, and B_{kl}, where $\boldsymbol{a}_k$, $\boldsymbol{a}_l$, $\boldsymbol{b}_k$, $\boldsymbol{b}_l$ are given by (1) and $k,l = 1,\ldots,n$. Thus, the squared sample distance covariance equals an average entry in the component-wise or Schur product of the centered distance matrices for the two vectors.

The sample functional distance correlation is then defined by

$$\hat{\mathscr{R}}_{\boldsymbol{X},\boldsymbol{Y}} = \frac{\hat{v}_{\boldsymbol{X},\boldsymbol{Y}}}{\sqrt{\hat{v}_{\boldsymbol{X},\boldsymbol{X}}\hat{v}_{\boldsymbol{Y},\boldsymbol{Y}}}}$$

if both $\hat{v}_{\boldsymbol{X},\boldsymbol{X}}$ and $\hat{v}_{\boldsymbol{Y},\boldsymbol{Y}}$ are strictly positive, and zero otherwise.

The problem of testing the independence between the random processes $\boldsymbol{X} \in \mathscr{L}_2^p(I_1)$ and $\boldsymbol{Y} \in \mathscr{L}_2^q(I_2)$ is equivalent to the problem of testing $H_0\colon \mathscr{R}_{\boldsymbol{X},\boldsymbol{Y}} = 0$. Székely et al. (2007) showed that under the null hypothesis of independence, $n\hat{\mathscr{R}}_{\boldsymbol{X},\boldsymbol{Y}}$ converges to $\sum_{j=1}^{\infty} \nu_j Z_j^2$, where Z_j are i.i.d $N(0,1)$, and ν_j depends on the distribution of $(\boldsymbol{X},\boldsymbol{Y})$. In practice, permutation tests are used to assess the significance of the functional distance correlation (Josse & Holmes (2014)).

5. Empirical application

In this Section we offer an illustrative example of applying correlation analysis to functional data. This method was employed here to cluster the twenty groups (pillars) of variables of 127 countries of the world in the period 2008-2014. The list of countries used in correlation analysis is contained in Table 1. Table 2 describes the variables used in the analysis divided into pillars. For this purpose, use was made of data published by the World Economic Forum (WEF) in its annual reports (http://www.weforum.org). Those are comprehensive data, describing exhaustively various socio-economic conditions or spheres of individual states. The data were transformed into functional data by the method described in Section 2. Calculations were performed using the Fourier basis. The time interval $[0,T] = [0,6]$ was divided into moments of time in the following way: $t_1 = 0.5(2008/2009), t_2 = 1.5(2009/2010), \ldots, t_6 = 5.5(2013/2014)$. Moreover, in view of the small number of time periods $(J = 6)$, for each variable the maximum number of basis components was taken to be equal to five. Table 3 contains the values of functional canonical correlation coefficients. As expected, they are all close to one. But a high value of this coefficient does not necessarily mean that there is a significant relationship between the two groups of variables. Table 4 contains the values of functional distance correlation coefficients. This time the values are rather moderate and easier to interpret. It is readily visible that the coefficients assume the highest values for the following pairs of pillars: 2 - infrastructure and 10 - marker size; 11 - business sophistication and 12 - innovation; 5 - higher education and training and 12 - innovation; 5 - higher education and training and 11 - business sophistication; as well as 6 - goods market efficiency and 11 - business sophistication. In turn, the coefficients have the lowest values for the pillars: 4 - health and primary education and

10 - market size, as well as 4 - health and primary education and 2 - infrastructure. Both the highest and the lowest values of distance correlation coefficients have an obvious empirical foundation.

We performed permutation tests for the correlation coefficients discussed. For all tests p-values were close to zero, so we can infer that we have some significant relationship between the groups (pillars) of variables.

Finally, we joined these pillars using Ward's hierarchical clustering algorithm with a distance measure of the form $1-\hat{\rho}_{\mathbf{X},\mathbf{Y}}$ and $1-\hat{\mathscr{R}}_{\mathbf{X},\mathbf{Y}}$ respectively. The results are shown in Figures 1 and 2. As can be observed, given the wide differences in the $\hat{\mathscr{R}}_{\mathbf{X},\mathbf{Y}}$ values, functional distance correlations permit arranging the various groups of variables into pillars in a logical way, e.g. $(4,9)$, $(11,12)$, etc. This allows analysing the examined reality in a deeper way, which is not possible when using canonical correlation coefficients.

During the numerical calculation process we used R software (R Core Team (2015)) and packages: `CCP` (Menzel (2012)), `energy` (Rizzo & Székely (2014)) and `fda` (Ramsay et al. (2014)).

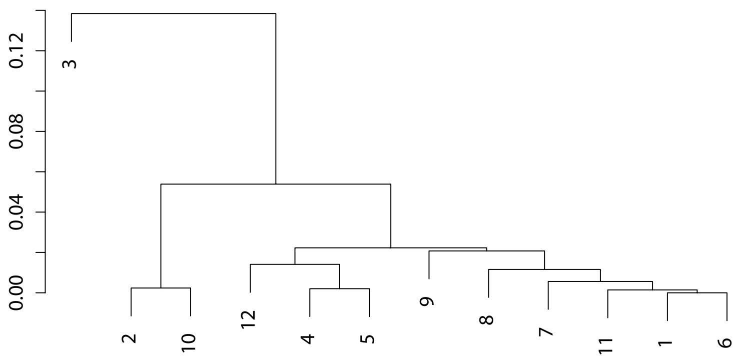

Figure 1: Dendrogram based on the functional canonical correlation coefficients.

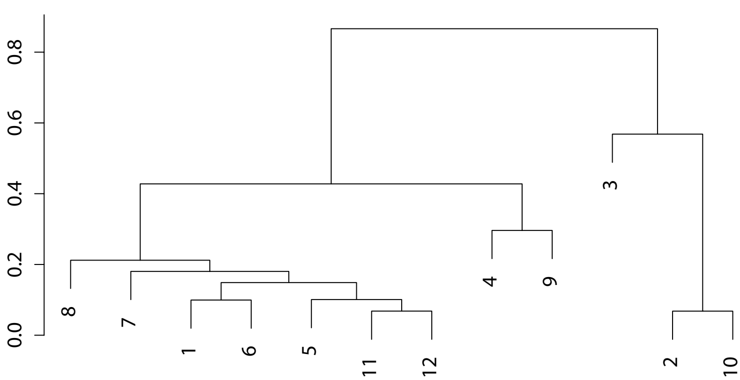

Figure 2: Dendrogram based on the functional distance correlation coefficients.

Table 1. Countries used in correlation analysis, 2008-2014

1	Albania	44	Germany	87	Nicaragua
2	Algeria	45	Ghana	88	Nigeria
3	Argentina	46	Greece	89	Norway
4	Armenia	47	Guatemala	90	Oman
5	Australia	48	Guyana	91	Pakistan
6	Austria	49	Honduras	92	Panama
7	Azerbaijan	50	Hong Kong SAR	93	Paraguay
8	Bahrain	51	Hungary	94	Peru
9	Bangladesh	52	Iceland	95	Philippines
10	Barbados	53	India	96	Poland
11	Belgium	54	Indonesia	97	Portugal
12	Benin	55	Ireland	98	Puerto Rico
13	Bolivia	56	Israel	99	Qatar
14	Bosnia and Herzegovina	57	Italy	100	Romania
15	Botswana	58	Jamaica	101	Russian Federation
16	Brazil	59	Japan	102	Saudi Arabia
17	Brunei Darussalam	60	Jordan	103	Senegal
18	Bulgaria	61	Kazakhstan	104	Serbia
19	Burkina Faso	62	Kenya	105	Singapore
20	Burundi	63	Korea Rep	106	Slovak Republic
21	Cambodia	64	Kuwait	107	Slovenia
22	Cameroon	65	Kyrgyz Republic	108	South Africa
23	Canada	66	Latvia	109	Spain
24	Chad	67	Lesotho	110	Sri Lanka
25	Chile	68	Lithuania	111	Sweden
26	China	69	Luxembourg	112	Switzerland
27	Colombia	70	Macedonia FYR	113	Taiwan China
28	Costa Rica	71	Madagascar	114	Tanzania
29	Côte d'Ivoire	72	Malawi	115	Thailand
30	Croatia	73	Malaysia	116	Timor-Leste
31	Cyprus	74	Mali	117	Trinidad and Tobago
32	Czech Republic	75	Malta	118	Turkey
33	Denmark	76	Mauritania	119	Uganda
34	Dominican Republic	77	Mauritius	120	Ukraine
35	Ecuador	78	Mexico	121	United Arab Emirates
36	Egypt	79	Mongolia	122	United Kingdom
37	El Salvador	80	Montenegro	123	United States
38	Estonia	81	Morocco	124	Uruguay
39	Ethiopia	82	Mozambique	125	Venezuela
40	Finland	83	Namibia	126	Vietnam
41	France	84	Nepal	127	Zambia
42	Gambia The	85	Netherlands		
43	Georgia	86	New Zealand		

Table 2. Variables used in correlation analysis, 2008-2014

No.	Variables	Pillars
1	Property rights	1. Institutions
2	Intellectual property protection	
3	Diversion of public funds	
4	Public trust of politicians	
5	Judicial independence	
6	Favoritism in decisions of government officials	
7	Wastefulness of government spending	
8	Burden of government regulation	
9	Transparency of government policymaking	
10	Business costs of terrorism	
12	Business costs of crime and violence	
11	Organized crime	
12	Reliability of police services	
13	Ethical behavior of firms	
14	Strength of auditing and reporting standards	
15	Efficacy of corporate boards	
16	Protection of minority shareholders' interests	
17	Quality of overall infrastructure	2. Infrastructure
18	Quality of roads	
19	Quality of port infrastructure	
20	Quality of air transport infrastructure	
21	Available airline seat kilometers	
22	Quality of electricity supply	
23	Inflation	3. Macroeconomic environment
24	Government debt	
25	Business impact of tuberculosis	4. Health and primary education
26	Tuberculosis incidence	
27	Business impact of HIV/AIDS	
28	HIV prevalence	
29	Infant mortality	
30	Life expectancy	
31	Quality of primary education	
32	Quality of the educational system	5. Higher education and training
33	Quality of math and science education	
34	Quality of management schools	
35	Internet access in schools	
36	Local availability of specialized research and training services	
37	Extent of staff training	
38	Intensity of local competition	6. Goods market efficiency
39	Extent of market dominance	
40	Effectiveness of anti-monopoly policy	
41	Agricultural policy costs	
42	Prevalence of trade barriers	
43	Prevalence of foreign ownership	
44	Business impact of rules on FDI	
45	Burden of customs procedures	
46	Degree of customer orientation	
47	Buyer sophistication	

Table 2. Variables used in correlation analysis, 2008-2014 (continuation)

No.	Variables	Pillars
48	Cooperation in labor-employer relations	7. Labor market efficiency
49	Flexibility of wage determination	
50	Hiring and firing practices	
51	Pay and productivity	
52	Reliance on professional management	
53	Female participation in labor force	
54	Financing through local equity market	8. Financial market development
55	Ease of access to loans	
56	Venture capital availability	
57	Soundness of banks	
58	Regulation of securities exchanges	
59	Availability of latest technologies	9. Technological readiness
60	Firm-level technology absorption	
61	FDI and technology transfer	
62	Internet users	
63	Domestic market size index	10. Market size
64	Foreign market size index	
65	GDP valued at PPP	
66	Exports as a percentage of GDP	
67	Local supplier quantity	11. Business sophistication
68	Local supplier quality	
69	State of cluster development	
70	Nature of competitive advantage	
71	Value chain breadth	
72	Control of international distribution	
73	Production process sophistication	
74	Extent of marketing	
75	Willingness to delegate authority	
76	Capacity for innovation	12. Innovation
77	Quality of scientific research institutions	
78	Company spending on R&D	
79	Government procurement of advanced technology products	
80	Availability of scientists and engineers	

Table 3. Functional canonical correlation coefficients

	1	2	3	4	5	6	7	8	9	10	11
2	0.9997										
3	0.9711	0.9137									
4	0.9999	0.9903	0.9036								
5	0.9993	0.9928	0.9186	0.9980							
6	1.0000	0.9927	0.9440	0.9921	0.9941						
7	0.9995	0.9795	0.8687	0.9822	0.9913	0.9947					
8	0.9988	0.9701	0.8773	0.9778	0.9781	0.9917	0.9878				
9	0.9988	0.9872	0.8714	0.9744	0.9927	0.9922	0.9846	0.9683			
10	0.9949	0.9976	0.8558	0.9518	0.9699	0.9828	0.9561	0.9408	0.9391		
11	1.0000	0.9934	0.9274	0.9924	0.9973	0.9982	0.9941	0.9885	0.9915	0.9816	
12	0.9984	0.9795	0.8782	0.9849	0.9915	0.9928	0.9794	0.9763	0.9795	0.9540	0.9937

Table 4. Functional distance correlation coefficients

	1	2	3	4	5	6	7	8	9	10	11
2	0.4961										
3	0.5116	0.5199									
4	0.6162	0.4064	0.4255								
5	0.8941	0.5128	0.5034	0.7062							
6	0.9006	0.5480	0.5550	0.6582	0.8972						
7	0.8480	0.5654	0.5760	0.6380	0.8443	0.8760					
8	0.8508	0.4921	0.4662	0.6371	0.8026	0.8681	0.8062				
9	0.7482	0.4524	0.4882	0.7038	0.8540	0.7510	0.6880	0.6750			
10	0.4752	0.9319	0.4938	0.4163	0.5066	0.5419	0.5506	0.4631	0.4400		
11	0.8671	0.5864	0.5381	0.7047	0.9008	0.9110	0.8398	0.8257	0.8118	0.5713	
12	0.8606	0.5616	0.5466	0.6466	0.9121	0.8652	0.8194	0.7757	0.7892	0.5585	0.9318

6. Conclusions

We proposed an extension of the classical correlation coefficients for two sets of variables for multivariate functional data. We suggested permutation tests to examine the significance of the results because the values of the proposed coefficients are rather hard to interpret. The presented method has been proved to be useful as it was tested on a real data set, in investigating the correlation between two sets of variables. This example confirms its usefulness in revealing the hidden structure of the co-dependence between groups (pillars) of variables representing various fields of socio-economic activity.

REFERENCES

ANDO, T., (2009). Penalized optimal scoring for the classification of multi-dimensional functional data. Statistcal Methodology 6, 565–576.

BERRENDERO, J. R., JUSTEL, A., SVARC, M., (2011). Principal components for multivariate functional data. Computational Statistics & Data Analysis 55(9), 2619–2634.

BESSE, P., (1979). Étude descriptive d'un processus: Aproximation et interpolation, Ph.D. thesis, Université Paul Sabatier, Toulouse III.

BEUTLER, F. J., (1970). Alias-free randomly timed sampling of stochastic process. IEEE Transactions on Information Theory 16, 147–152.

ESCOUFIER, Y., (1970). Echantillonnage dans une population de variables aléatoires réelles, Ph.D thesis, Université des sciences et techniques du Languedoc, Montpellier.

ESCOUFIER, Y., (1973). Le traitement des variables vectorielles. Biometrics 29(4), 751–760.

ESCOUFIER, Y., ROBERT, P., (1979). Choosing variables and metrics by optimizing the RV coefficient. In Optimizing Methods in Statistics, Rustagi, J.S., Ed., Academic: New York, 205–219.

FERRATY, F., VIEU, P., (2003). Curve discrimination. A nonparametric functional approach. Computational Statistics & Data Analysis 44 161–173.

FERRATY, F., VIEU, P., (2009). Additive prediction and boosting for functional data. Computational Statistics & Data Analysis 53(4), 1400–1413.

GÓRECKI, T., KRZYŚKO, M., WASZAK, Ł., WOŁYŃSKI, W., (2014). Methods of reducing dimension for functional data. Statistics in Transition new series 15(2), 231–242.

GÓRECKI, T., KRZYŚKO, M., WOŁYŃSKI, W., (2015). Classification problems based on regression models for multi-dimensional functional data. Statistics in Transition new series 16(1), 97–110.

HASTIE, T. J., TIBSHIRANI, R. J., BUJA, A., (1995). Penalized discriminant analysis. Annals of Statistics 23, 73–102.

HE, G., MULLER, H. G., WANG, J. L., (2004). Methods of canonical analysis for functional data. Journal of Statistical Planning and Inference 122(1-2), 141–159.

HOTELLING, H., (1936). Relation between two sets of variables. Biometrika 28(3/4), 321–377.

HORVÁTH, L., KOKOSZKA, P., (2012). Inference for Functional Data with Applications, Springer.

JAMES, G. M., (2002). Generalized linear models with functional predictors. Journal of the Royal Statistical Society 64(3), 411–432.

JACQUES, J., PREDA, C., (2014). Model-based clustering for multivariate functional data. Computational Statistics & Data Analysis 71(C), 92–106.

JOSSE, J., HOLMES, S.,(2014). Tests of independence and beyond. arXiv:1307.7383v3.

KRZYŚKO, M., WASZAK, Ł., (2013). Canonical correlation analysis for functional data. Biometrical Letters 50(2), 95–105.

LEE, A. J., (1976). On band-limited stochastic processes. SIAM Journal on Applied Mathematics 30, 169–177.

LEURGANS, S. E., MOYEED, R. A., SILVERMAN, B. W., (1993). Canonical correlation analysis when the data are curves. Journal of the Royal Statistical Society. Series B (Methodological) 55(3), 725–740.

MASRY, E., (1978). Poisson sampling and spectral estimation of continuous time processes. IEEE Transactions on Information Theory 24, 173–183.

MATSUI, H., ARAKI, Y., KONISHI, S., (2008). Multivariate regression modeling for functional data. Journal of Data Science 6, 313–331.

MENZEL, U., (2012). CCP: Significance Tests for Canonical Correlation Analysis (CCA). R package version 1.1. http://CRAN.R-project.org/package=CCP.

MÜLLER, H. G., STADMÜLLER, U., (2005). Generalized functional linear models. Annals of Statistics 33, 774–805.

R CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RAMSAY, J. O., SILVERMAN, B. W. (2005). Functional Data Analysis, Second Edition, Springer.

RAMSAY, J. O., WICKHAM, H., GRAVES, S., HOOKER, G., (2014). fda: Functional Data Analysis. R package version 2.4.4. http://CRAN.R-project.org/package=fda.

REISS, P. T., OGDEN, R.T., (2007). Functional principal component regression and functional partial least squares. Journal of the American Statistcal Assosiation 102(479), 984–996.

RIZZO, M. L., SZÉKELY, G. J., (2014). energy: E-statistics (energy statistics). R package version 1.6.2. http://CRAN.R-project.org/package=energy.

ROBERT, P., ESCOUFIER, Y., (1976). A unifying tool for linear multivariate statistical methods: the RV coefficient. Journal of the Royal Statistical Society. Series C (Applied Statistics) 25(3), 257–265.

ROSSI, F., DELANNAYC, N., CONAN-GUEZA, B., VERLEYSENC, M., (2005). Representation of functional data in neural networks. Neurocomputing 64, 183–210.

ROSSI, F., VILLA, N., (2006). Support vector machines for functional data classification. Neural Computing 69, 730–742.

SZÉKELY, G. J., RIZZO, M. L., BAKIROV, N.K., (2007). Measuring and testing dependence by correlation of distances. The Annals of Statistics 35(6), 2769–2794.

SZÉKELY, G. J., RIZZO, M. L., (2009). Brownian distance covariance. Annals of Applied Statistics 3(4), 1236–1265.

SZÉKELY, G. J., RIZZO, M. L., (2012). On the uniqueness of distance covariance. Statistical Probability Letters 82(12), 2278–2282.

SZÉKELY, G. J., RIZZO, M. L., (2013). The distance correlation t-test of independence in high dimension. Journal of Multivariate Analysis 117, 193–213.