

## Robust estimation in canonical correlation analysis for multivariate functional data

Mirosław Krzyśko\* and Łukasz Smaga<sup>†‡</sup>

### Abstract

In this paper, the canonical correlation analysis for multivariate functional data is considered. The analysis is based on the basis functions representation of the data. The use of non-orthogonal bases is available in contrast to the approach given in the literature. The robust estimation methods of the covariance matrix are also studied in the multivariate functional canonical correlation analysis. Simulation studies and breakdown analysis suggest that the proposed methods may perform better than the classical estimator under non-normal models and in the presence of outlying observations.

**Keywords:** Basis functions representation, Canonical correlation, Functional data analysis, Multivariate analysis, Robust covariance estimation.

*Mathematics Subject Classification (2010):* 62H20, 62F35

*Received :* 20.11.2017 *Accepted :* 28.06.2018 *Doi :* 10.15672/HJMS.2018.613

### 1. Introduction

Nowadays, quick and accurate measurement procedures are developed caused by the technological progress. This results in obtaining new types of (usually large) data. In many areas, the observations of random variables are taken over a continuous interval or in larger discretizations of such interval. In functional data analysis (FDA), such data observed longitudinally are expressed as smooth functions or curves, and then the information is drawn from the collection of functions or curves, called functional data. FDA has received considerable attention in such fields of applications as chemometrics, economics, environmental studies, image recognition, spectroscopy, and many others [30, 12, 2, 26].

---

\*Interfaculty Institute of Mathematics and Statistics, The President Stanisław Wojciechowski State University of Applied Sciences in Kalisz, Nowy Świat 4, 62-800 Kalisz, Poland, Email: mkrzysko@amu.edu.pl

<sup>†</sup>Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland, Email: ls@amu.edu.pl

<sup>‡</sup>Corresponding Author.

Comprehensive surveys about functional data analysis can be found in [31, 13, 19, 45]. Some new developments in FDA can be found, for instance, in [1, 4] and in recent special issue on *Econometrics and Statistics* [23]. In the literature, the following problems of analysis of univariate and multivariate functional data are of particular interest: analysis of variance [16, 17], canonical correlation analysis [15], classification [21, 11], cluster analysis [44], outlier detection [12], principal component analysis [3, 14], regression [6, 18, 34], repeated measurements analysis [28, 41], variable selection [29, 7].

This paper concerns the canonical correlation analysis for multivariate functional data (MFCCA). The aim of this analysis is to identify and quantify the relations between a  $p$ -dimensional stochastic process  $\mathbf{X}(s)$  and a  $q$ -dimensional stochastic process  $\mathbf{Y}(t)$ . The associations between  $\mathbf{X}(s)$  and  $\mathbf{Y}(t)$  are measured by the correlations between linear combinations of both sets of processes. For the case  $p = q = 1$ , the so-called canonical correlation analysis for (univariate) functional data was developed, for example, in the monograph by Ramsay and Silverman [31]. Recently, Górecki et al. [15] extended this analysis to multivariate functional data ( $p \geq 1$  and  $q \geq 1$ ). Their method is based on orthonormal basis functions representation of the data and the sample covariance matrix as estimator of unknown covariance matrix, which however may cause in poorer results of the canonical correlation analysis. Namely, some non-orthogonal basis may be more appropriate for particular type of data, and what is perhaps more important, the classical estimator of covariance matrix may be sensitive to non-normal data and outlying observations. In this paper, we extend the multivariate functional canonical correlation analysis proposed by Górecki et al. [15] to be available for using non-orthogonal bases and consider robust estimation of covariance matrix in this analysis. Simulation results suggest that the proposed modifications of MFCCA perform promisingly and may be alternatives to existing methods in practical applications.

The remainder of the paper is organized as follows. In Section 2, the canonical correlation analysis for multivariate functional data based on the basis functions representation of the data is considered. Section 3 raises the problem of estimation of unknown covariance matrices in canonical correlation analysis. A brief review of robust estimators of multivariate location and scatter is also given there. In Section 4, efficiency and robustness of the covariance matrix estimation methods in MFCCA are investigated by means of a simulation study and breakdown plots. Conclusions are provided in Section 5.

## 2. Canonical correlation analysis for multivariate functional data

In this Section, we formally present the canonical correlation analysis for multivariate functional data and show how to deal with it by using the basis functions representation of the data. The obtained results are more general than those of [15].

Let  $\mathbf{X}(s) = (X_1(s), \dots, X_p(s))^\top$  and  $\mathbf{Y}(t) = (Y_1(t), \dots, Y_q(t))^\top$  be random processes belonging to spaces  $L_2^p(I_1)$  and  $L_2^q(I_2)$ , where  $L_2^p(I)$  is a Hilbert space of  $p$ -dimensional vectors of square integrable functions on the set  $I$  and  $I_1 = [a, b]$  and  $I_2 = [c, d]$ ,  $a, b, c, d \in \mathbb{R}$ . Moreover, without loss of generality, we can assume that  $E(\mathbf{X}(s)) = \mathbf{0}_p$  for  $s \in I_1$  and  $E(\mathbf{Y}(t)) = \mathbf{0}_q$  for  $t \in I_2$ . By Górecki et al. [15], the functional canonical variables  $U$  and  $V$  for processes  $\mathbf{X}(s)$  and  $\mathbf{Y}(t)$  are defined in the following way:

$$U = \langle \mathbf{u}, \mathbf{X} \rangle = \int_{I_1} \mathbf{u}^\top(s) \mathbf{X}(s) ds, \quad V = \langle \mathbf{v}, \mathbf{Y} \rangle = \int_{I_2} \mathbf{v}^\top(t) \mathbf{Y}(t) dt,$$

where  $\mathbf{u} \in L_2^p(I_1)$  and  $\mathbf{v} \in L_2^q(I_2)$  are the vector weight functions chosen to maximize the coefficient

$$\rho_{U,V} = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U^{(N)}) \text{Var}(V^{(N)})}},$$

where  $\text{Var}(U^{(N)}) = \text{Var}(U) + \lambda \text{PEN}_2(\mathbf{u})$ ,  $\text{Var}(V^{(N)}) = \text{Var}(V) + \lambda \text{PEN}_2(\mathbf{v})$ ,  $\lambda > 0$  and

$$\text{PEN}_2(\mathbf{u}) = \int_{I_1} \left( \frac{\partial^2 \mathbf{u}(s)}{\partial s^2} \right)^\top \frac{\partial^2 \mathbf{u}(s)}{\partial s^2} ds, \quad \text{PEN}_2(\mathbf{v}) = \int_{I_2} \left( \frac{\partial^2 \mathbf{v}(t)}{\partial t^2} \right)^\top \frac{\partial^2 \mathbf{v}(t)}{\partial t^2} dt,$$

subject to the constraint that

$$(2.1) \quad \text{Var}(U^{(N)}) = \text{Var}(V^{(N)}) = 1.$$

Without adding roughness penalty terms in constraints, functional canonical correlation analysis may not produce a meaningful result as described, for instance, in [15] and [31] (Chapter 11). Further, we define the  $k$ th functional canonical correlation  $\rho_k$  and the associated vector weight functions  $\mathbf{u}_k(s)$  and  $\mathbf{v}_k(t)$  as follows:

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

subject to the restrictions given in (2.1), and the  $k$ th pair of canonical variables  $(U_k, V_k) = (\langle \mathbf{u}_k, \mathbf{X} \rangle, \langle \mathbf{v}_k, \mathbf{Y} \rangle)$  is uncorrelated with the first  $k-1$  canonical variables. The above analysis is called (smoothed) canonical correlation analysis for multivariate functional data.

We deal with the canonical correlation analysis for multivariate functional data by assuming that the random processes  $\mathbf{X}(s)$  and  $\mathbf{Y}(t)$  belong to finite dimensional subspaces  $\mathcal{L}_2^p(I_1)$ ,  $\mathcal{L}_2^q(I_2)$  of  $L_2^p(I_1)$ ,  $L_2^q(I_2)$ , respectively, and their components can be represented by a finite number of basis functions  $\{\varphi_{il}\}$  and  $\{\psi_{jm}\}$ ,  $i = 1, \dots, p$ ,  $l = 1, \dots, j = 1, \dots, q$ ,  $m = 1, \dots$ , i.e.,

$$X_i(s) = \sum_{l=1}^{K_i} \alpha_{il} \varphi_{il}(s), \quad Y_j(t) = \sum_{m=1}^{L_j} \beta_{jm} \psi_{jm}(t)$$

where  $s \in I_1$ ,  $t \in I_2$ ,  $i = 1, \dots, p$ ,  $j = 1, \dots, q$  and  $\alpha_{il}$  and  $\beta_{jm}$  are random variables of mean zero and finite variance. Let

$$(2.2) \quad \begin{aligned} \boldsymbol{\alpha} &= (\alpha_{11}, \dots, \alpha_{1K_1}, \dots, \alpha_{p1}, \dots, \alpha_{pK_p})^\top, \\ \boldsymbol{\beta} &= (\beta_{11}, \dots, \beta_{1L_1}, \dots, \beta_{q1}, \dots, \beta_{qL_q})^\top, \\ \boldsymbol{\Phi}(s) &= \begin{bmatrix} \boldsymbol{\varphi}_1^\top(s) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2^\top(s) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}_p^\top(s) \end{bmatrix}, \end{aligned}$$

$$(2.3) \quad \boldsymbol{\Psi}(t) = \begin{bmatrix} \boldsymbol{\psi}_1^\top(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\psi}_2^\top(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\psi}_q^\top(t) \end{bmatrix},$$

where  $\boldsymbol{\varphi}_i^\top(s) = (\varphi_{i1}(s), \dots, \varphi_{iK_i}(s))$  and  $\boldsymbol{\psi}_j^\top(t) = (\psi_{j1}(t), \dots, \psi_{jL_j}(t))$ . Moreover, we assume that  $\text{E}(\boldsymbol{\alpha}) = \mathbf{0}$ ,  $\text{E}(\boldsymbol{\beta}) = \mathbf{0}$ ,  $\text{Cov}(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}_{11} > 0$ ,  $\text{Cov}(\boldsymbol{\beta}) = \boldsymbol{\Sigma}_{22} > 0$  and  $\text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\Sigma}_{12}$ , where the matrices  $\boldsymbol{\Sigma}_{11}$ ,  $\boldsymbol{\Sigma}_{22}$  and  $\boldsymbol{\Sigma}_{12}$  are unknown parameters. Therefore, in matrix notation, the processes can be expressed as follows:

$$(2.4) \quad \mathbf{X}(s) = \boldsymbol{\Phi}(s)\boldsymbol{\alpha}, \quad \mathbf{Y}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\beta}.$$

We may assume that the vector weight functions  $\mathbf{u}(s)$  and  $\mathbf{v}(t)$  belong to the same spaces as the processes  $\mathbf{X}(s)$  and  $\mathbf{Y}(t)$ , respectively, i.e.,  $\mathbf{u}(s) = \boldsymbol{\Phi}(s)\boldsymbol{\mu}$  and  $\mathbf{v}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\nu}$ , where

$$\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{1K_1}, \dots, \mu_{p1}, \dots, \mu_{pK_p})^\top, \quad \boldsymbol{\nu} = (\nu_{11}, \dots, \nu_{1L_1}, \dots, \nu_{q1}, \dots, \nu_{qL_q})^\top.$$

Then, we have

$$U = \langle \mathbf{u}, \mathbf{X} \rangle = \int_{I_1} \mathbf{u}^\top(s) \mathbf{X}(s) ds = \int_{I_1} \boldsymbol{\mu}^\top \boldsymbol{\Phi}^\top(s) \boldsymbol{\Phi}(s) \boldsymbol{\alpha} ds = \boldsymbol{\mu}^\top \mathbf{J}_\Phi \boldsymbol{\alpha},$$

where

$$\mathbf{J}_\Phi := \int_{I_1} \boldsymbol{\Phi}^\top(s) \boldsymbol{\Phi}(s) ds = \text{diag} \left( \mathbf{J}_{\varphi_1}, \dots, \mathbf{J}_{\varphi_p} \right)$$

is the block diagonal matrix of  $K_i \times K_i$  cross product matrices

$$\mathbf{J}_{\varphi_i} = \int_{I_1} \varphi_i(s) \varphi_i^\top(s) ds, \quad i = 1, \dots, p.$$

Analogously, we obtain  $V = \mathbf{v}^\top \mathbf{J}_\Psi \boldsymbol{\beta}$ , where

$$\mathbf{J}_\Psi := \int_{I_2} \boldsymbol{\Psi}^\top(t) \boldsymbol{\Psi}(t) dt = \text{diag} \left( \mathbf{J}_{\psi_1}, \dots, \mathbf{J}_{\psi_q} \right)$$

and  $\mathbf{J}_{\psi_j} = \int_{I_2} \psi_j(t) \psi_j^\top(t) dt$ ,  $j = 1, \dots, q$ . Therefore,

$$\mathbf{E}(U) = \boldsymbol{\mu}^\top \mathbf{J}_\Phi \mathbf{E}(\boldsymbol{\alpha}) = 0,$$

$$\mathbf{E}(V) = \mathbf{v}^\top \mathbf{J}_\Psi \mathbf{E}(\boldsymbol{\beta}) = 0,$$

$$\text{Var}(U) = \boldsymbol{\mu}^\top \mathbf{J}_\Phi \text{Cov}(\boldsymbol{\alpha}) \mathbf{J}_\Phi \boldsymbol{\mu} = \boldsymbol{\mu}^\top \mathbf{J}_\Phi \boldsymbol{\Sigma}_{11} \mathbf{J}_\Phi \boldsymbol{\mu},$$

$$\text{Var}(V) = \mathbf{v}^\top \mathbf{J}_\Psi \text{Cov}(\boldsymbol{\beta}) \mathbf{J}_\Psi \mathbf{v} = \mathbf{v}^\top \mathbf{J}_\Psi \boldsymbol{\Sigma}_{22} \mathbf{J}_\Psi \mathbf{v},$$

$$\text{Cov}(U, V) = \boldsymbol{\mu}^\top \mathbf{J}_\Phi \text{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \mathbf{J}_\Psi \mathbf{v} = \boldsymbol{\mu}^\top \mathbf{J}_\Phi \boldsymbol{\Sigma}_{12} \mathbf{J}_\Psi \mathbf{v}.$$

By Górecki et al. [15], we conclude that

$$\text{Var}(U^{(N)}) = \boldsymbol{\mu}^\top (\mathbf{J}_\Phi \boldsymbol{\Sigma}_{11} \mathbf{J}_\Phi + \lambda \mathbf{R}_\Phi) \boldsymbol{\mu}, \quad \text{Var}(V^{(N)}) = \mathbf{v}^\top (\mathbf{J}_\Psi \boldsymbol{\Sigma}_{22} \mathbf{J}_\Psi + \lambda \mathbf{R}_\Psi) \mathbf{v},$$

where

$$\mathbf{R}_\Phi = \begin{bmatrix} \int_{I_1} \frac{\partial^2 \varphi_1(s)}{\partial s^2} \frac{\partial^2 \varphi_1^\top(s)}{\partial s^2} ds & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \int_{I_1} \frac{\partial^2 \varphi_2(s)}{\partial s^2} \frac{\partial^2 \varphi_2^\top(s)}{\partial s^2} ds & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \int_{I_1} \frac{\partial^2 \varphi_p(s)}{\partial s^2} \frac{\partial^2 \varphi_p^\top(s)}{\partial s^2} ds \end{bmatrix}$$

and

$$\mathbf{R}_\Psi = \begin{bmatrix} \int_{I_2} \frac{\partial^2 \psi_1(t)}{\partial t^2} \frac{\partial^2 \psi_1^\top(t)}{\partial t^2} dt & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \int_{I_2} \frac{\partial^2 \psi_2(t)}{\partial t^2} \frac{\partial^2 \psi_2^\top(t)}{\partial t^2} dt & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \int_{I_2} \frac{\partial^2 \psi_q(t)}{\partial t^2} \frac{\partial^2 \psi_q^\top(t)}{\partial t^2} dt \end{bmatrix}.$$

We have thus proved the following theorem.

**2.1. Theorem.** *Under the above assumptions and notation, we have*

$$U_k = \boldsymbol{\mu}_k^\top \mathbf{J}_\Phi \boldsymbol{\alpha}, \quad V_k = \mathbf{v}_k^\top \mathbf{J}_\Psi \boldsymbol{\beta}, \quad \mathbf{u}_k(s) = \boldsymbol{\Phi}(s) \boldsymbol{\mu}_k, \quad \mathbf{v}_k(t) = \boldsymbol{\Psi}(t) \mathbf{v}_k,$$

$s \in I_1$ ,  $t \in I_2$ , and

$$\rho_k = \sup_{\boldsymbol{\mu} \in \mathbb{R}^K, \mathbf{v} \in \mathbb{R}^L} \boldsymbol{\mu}^\top \mathbf{J}_\Phi \boldsymbol{\Sigma}_{12} \mathbf{J}_\Psi \mathbf{v} = \boldsymbol{\mu}_k^\top \mathbf{J}_\Phi \boldsymbol{\Sigma}_{12} \mathbf{J}_\Psi \mathbf{v}_k, \quad k = 1, \dots, \min\{K, L\},$$

where  $K = K_1 + \dots + K_p$ ,  $L = L_1 + \dots + L_q$ , subject to the restrictions

$$\boldsymbol{\mu}_k^\top (\mathbf{J}_\Phi \boldsymbol{\Sigma}_{11} \mathbf{J}_\Phi + \lambda \mathbf{R}_\Phi) \boldsymbol{\mu}_k = 1$$

and

$$\mathbf{v}_k^\top (\mathbf{J}_\Psi \boldsymbol{\Sigma}_{22} \mathbf{J}_\Psi + \lambda \mathbf{R}_\Psi) \mathbf{v}_k = 1.$$

By Theorem 2.1, MFCCA reduces to the canonical correlation analysis for  $K$ -dimensional random vector  $\boldsymbol{\alpha}_*$  and  $L$ -dimensional random vector  $\boldsymbol{\beta}_*$  with  $E(\boldsymbol{\alpha}_*) = \mathbf{0}_K$ ,  $E(\boldsymbol{\beta}_*) = \mathbf{0}_L$ ,  $\text{Cov}(\boldsymbol{\alpha}_*) = \mathbf{J}_\Phi \boldsymbol{\Sigma}_{11} \mathbf{J}_\Phi + \lambda \mathbf{R}_\Phi$ ,  $\text{Cov}(\boldsymbol{\beta}_*) = \mathbf{J}_\Psi \boldsymbol{\Sigma}_{22} \mathbf{J}_\Psi + \lambda \mathbf{R}_\Psi$  and  $\text{Cov}(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*) = \mathbf{J}_\Phi \boldsymbol{\Sigma}_{12} \mathbf{J}_\Psi$ . It is well known that the CCA optimization problem has a fairly simple solution. Namely, the vectors  $\boldsymbol{\mu}_k$  and  $\mathbf{v}_k$  are the eigenvectors corresponding to the eigenvalues  $\rho_1^2 \geq \dots \geq \rho_{\min\{K,L\}}^2 > 0$  of the matrices

$$\begin{aligned} & (\mathbf{J}_\Phi \boldsymbol{\Sigma}_{11} \mathbf{J}_\Phi + \lambda \mathbf{R}_\Phi)^{-1} (\mathbf{J}_\Phi \boldsymbol{\Sigma}_{12} \mathbf{J}_\Psi) (\mathbf{J}_\Psi \boldsymbol{\Sigma}_{22} \mathbf{J}_\Psi + \lambda \mathbf{R}_\Psi)^{-1} (\mathbf{J}_\Psi \boldsymbol{\Sigma}_{12}^\top \mathbf{J}_\Phi), \\ & (\mathbf{J}_\Psi \boldsymbol{\Sigma}_{22} \mathbf{J}_\Psi + \lambda \mathbf{R}_\Psi)^{-1} (\mathbf{J}_\Psi \boldsymbol{\Sigma}_{12}^\top \mathbf{J}_\Phi) (\mathbf{J}_\Phi \boldsymbol{\Sigma}_{11} \mathbf{J}_\Phi + \lambda \mathbf{R}_\Phi)^{-1} (\mathbf{J}_\Phi \boldsymbol{\Sigma}_{12} \mathbf{J}_\Psi), \end{aligned}$$

respectively. In practice, we have to estimate the unknown parameters, i.e., the matrices  $\boldsymbol{\Sigma}_{11}$ ,  $\boldsymbol{\Sigma}_{22}$  and  $\boldsymbol{\Sigma}_{12}$ . This problem is discussed in the next Section. The smoothing parameter  $\lambda$  can be chosen subjectively or by using an automatic procedure as, for example, cross-validation one analogous to that considered by Ramsay and Silverman [31] (Chapter 11) in the canonical correlation analysis for univariate functional data.

### 3. Robust estimation in MFCCA

In this Section, we consider the estimation problem of the unknown matrices  $\boldsymbol{\Sigma}_{11}$ ,  $\boldsymbol{\Sigma}_{22}$  and  $\boldsymbol{\Sigma}_{12}$  to make the theory of Section 2 applicable. Robust estimators are discussed as competitors to the classical estimator.

Assume that we have  $n$  independent realizations  $\mathbf{x}_1(s), \dots, \mathbf{x}_n(s)$  and  $\mathbf{y}_1(t), \dots, \mathbf{y}_n(t)$  of random processes  $\mathbf{X}(s)$  and  $\mathbf{Y}(t)$ ,  $s \in I_1$ ,  $t \in I_2$ . They are represented as in (2.4), i.e.,  $\mathbf{x}_r(s) = \Phi(s) \boldsymbol{\alpha}_r$  and  $\mathbf{y}_r(t) = \Psi(t) \boldsymbol{\beta}_r$ ,  $r = 1, \dots, n$ . The vectors  $\boldsymbol{\alpha}_r$  and  $\boldsymbol{\beta}_r$  can be estimated by the least squares method or by the roughness penalty approach (see [31] Chapter 5). Let  $\hat{\boldsymbol{\alpha}}_r$  and  $\hat{\boldsymbol{\beta}}_r$  denote the estimates of  $\boldsymbol{\alpha}_r$  and  $\boldsymbol{\beta}_r$ , respectively. The optimum numbers of basis functions  $K_i$  and  $L_j$  ( $i = 1, \dots, p$ ,  $j = 1, \dots, q$ ) can be selected by using Akaike or Bayesian information criterion (see, for example, [15]). The Bayesian information criterion is usually preferred, since it measures goodness of fit better than Akaike information criterion [40].

The unknown matrices  $\boldsymbol{\Sigma}_{11}$ ,  $\boldsymbol{\Sigma}_{22}$  and  $\boldsymbol{\Sigma}_{12}$  are estimated based on the vectors  $\hat{\boldsymbol{\alpha}}_r$  and  $\hat{\boldsymbol{\beta}}_r$ ,  $r = 1, \dots, n$ . More precisely, we estimate the joint covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

of  $(\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$  by using the vectors  $(\hat{\boldsymbol{\alpha}}_r^\top, \hat{\boldsymbol{\beta}}_r^\top)^\top$ ,  $r = 1, \dots, n$  as observations. Górecki et al. [15] used the sample covariance matrix to estimate  $\boldsymbol{\Sigma}$ . However, it is known that this classical estimator is sensitive to non-normal data and outlying observations. The poor behavior of this estimator may result in worse performance of MFCCA based on it. For this reason, we propose to use robust estimators of the matrix  $\boldsymbol{\Sigma}$ , which seems to result in better performance of MFCCA than using classical one, as we will see in Section 4. In that Section, we compare the sample covariance matrix with commonly used robust estimates of covariance matrix, which are briefly reviewed in the remainder of the present Section. Other estimation methods of  $\boldsymbol{\Sigma}$  can also be applied (see, for example, [42, 46]).

Let  $\{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  be the data set in  $\mathbb{R}^d$ . Hence  $n$  stands for the number of objects and  $d$  for the number of variables. First, we consider the minimum covariance determinant

(MCD) estimator, which is very popular in the literature. It was introduced by Rousseeuw [35]. The MCD estimates for multivariate location and scatter are the mean and a multiple of the sample covariance matrix of  $h$  observations  $\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_h}$ , for which the determinant of the covariance matrix is minimum, i.e.,

$$\mathbf{L}_{\text{MCD}} = \frac{1}{h} \sum_{j=1}^h \mathbf{z}_{i_j}, \quad \mathbf{S}_{\text{MCD}} = \frac{c_1 c_2}{h-1} \sum_{j=1}^h (\mathbf{z}_{i_j} - \mathbf{L}_{\text{MCD}})(\mathbf{z}_{i_j} - \mathbf{L}_{\text{MCD}})^\top.$$

The consistency correction factor  $c_1$  and the small sample correction factor  $c_2$  are selected for consistency at the multivariate normal model and unbiasedness at small samples of  $\mathbf{S}_{\text{MCD}}$ . With  $h = \lfloor (n+d+1)/2 \rfloor$ , the breakdown point of the MCD estimator is high, and hence such a choice of  $h$  is recommended. The computation of the MCD estimate by the naive algorithm is possible in sensible time only for very small data sets. However, fast computing algorithms of the estimators are available as that by Rousseeuw and Van Driessen [36], which is usually used in practice. Let  $(\mathbf{L}_1, \mathbf{S}_1)$  be an approximation of the MCD estimator and let  $d_1, \dots, d_n$  denote the distances of the observations  $\mathbf{z}_1, \dots, \mathbf{z}_n$  related to this approximation, i.e.,

$$d_r = \sqrt{(\mathbf{z}_r - \mathbf{L}_1)^\top \mathbf{S}_1^{-1} (\mathbf{z}_r - \mathbf{L}_1)}, \quad r = 1, \dots, n.$$

Then, the C-step of this algorithm moves from  $(\mathbf{L}_1, \mathbf{S}_1)$  to the next approximation  $(\mathbf{L}_2, \mathbf{S}_2)$  by computing it for those  $h$  data points, which have smallest distances. By this step, lower determinant  $\det(\mathbf{S}_2)$  may be obtained. Here, ‘‘C’’ stands for ‘‘concentration’’, because we look for the  $h$  observations with smallest distances, and  $\mathbf{S}_2$  is more concentrated and has possibly lower determinant than  $\mathbf{S}_1$ . The estimators  $\mathbf{L}_{\text{MCD}}$  and  $\mathbf{S}_{\text{MCD}}$  are not very efficient at normal models. This is especially evident, when  $h$  is chosen to achieve maximal breakdown point. This low efficiency drawback may be overcome by using reweighted MCD estimators (see, for instance, [8]). Weight  $w_r$  assigned to observation  $\mathbf{z}_r$  is defined as

$$w_r = I_{(-\infty, \chi_{d,0.975}^2]} \left( (\mathbf{z}_r - \mathbf{L}_{\text{MCD}})^\top \mathbf{S}_{\text{MCD}}^{-1} (\mathbf{z}_r - \mathbf{L}_{\text{MCD}}) \right)$$

related to the raw estimators  $\mathbf{L}_{\text{MCD}}$  and  $\mathbf{S}_{\text{MCD}}$ , where  $I_A$  stands for the usual indicator function on the set  $A$ . Then, the reweighted MCD estimators are as follows:

$$\mathbf{L}_{\text{MCD}}^R = \frac{1}{w} \sum_{r=1}^n w_r \mathbf{z}_r, \quad \mathbf{S}_{\text{MCD}}^R = \frac{c_1^R c_2^R}{w-1} \sum_{r=1}^n w_r (\mathbf{z}_r - \mathbf{L}_{\text{MCD}}^R)(\mathbf{z}_r - \mathbf{L}_{\text{MCD}}^R)^\top,$$

where  $w = w_1 + \dots + w_n$ .

Now, we describe the S-estimators introduced by Davies [9] and further investigated by Lopuhaä [24]. In fact, Davies [9] extended an idea of S-estimators by Rousseeuw and Yohai [37] in regression. The S-estimators of location and scatter are the vector  $\mathbf{L}$  and the positive definite symmetric matrix  $\mathbf{S}$  minimizing the determinant  $\det(\mathbf{S})$ , subject to

$$\frac{1}{n} \sum_{r=1}^n \rho(d_r) \leq b_0,$$

where the function  $\rho$  is non-negative, symmetric and continuously differentiable and strictly increasing on  $[0, c_0]$  with  $\rho(0) = 0$  and constant on  $[c_0, \infty)$  for some  $c_0 > 0$ , and

$$(3.1) \quad d_r = \sqrt{(\mathbf{z}_r - \mathbf{L})^\top \mathbf{S}^{-1} (\mathbf{z}_r - \mathbf{L})}, \quad r = 1, \dots, n$$

and  $b_0 < \sup \rho$ . The S-estimators are asymptotically normal and  $\sqrt{n}$ -consistent. Moreover, these estimates can have a very high breakdown point  $\lfloor (n-d+1)/2 \rfloor / n$  and be highly efficient at normal models. However, a high breakdown point and a high efficiency at the normal models may be not simultaneously attained by the S-estimators. Lopuhaä

[25] and Davies [10] proposed some modifications of the S-estimators, which can overcome this drawback. The S-estimates may be computed, for example, by a fast algorithm of [38] and that similar to the one proposed by Salibian-Barrera and Yohai [39] for the regression setting (see [42]).

Finally, we consider multivariate M-estimators defined by Maronna [27], who extended the idea of the univariate M-estimators by Huber [20]. The M-estimators are defined as the vector  $\mathbf{L}$  and the positive definite symmetric matrix  $\mathbf{S}$ , which are solutions of the following equations:

$$\frac{1}{n} \sum_{r=1}^n u_1(d_r)(\mathbf{z}_r - \mathbf{L}) = \mathbf{0}, \quad \frac{1}{n} \sum_{r=1}^n u_2(d_r^2)(\mathbf{z}_r - \mathbf{L})(\mathbf{z}_r - \mathbf{L})^\top = \mathbf{S},$$

where  $u_1$  and  $u_2$  are weight functions satisfying certain conditions and  $d_r$  are given in (3.1). Unfortunately, the M-estimators may have relatively low breakdown points. However, they can be quite efficient at normal and other models. Modified M-estimators, which perform better than the standard ones, are proposed by Tyler [43] and Kent and Tyler [22]. Lopuhaä [24] shows that M-estimators have a close connection to S-estimators.

In the next Section, the performance of considered estimators of the covariance matrix in the canonical correlation analysis for multivariate functional data is studied under finite samples.

## 4. Simulation study

In this Section, the methods of estimation of covariance matrix  $\Sigma$  for MFCCA considered in Section 3 are compared by means of a simulation study.

**4.1. Experimental setup.** The numbers of observations are  $n = 100, 500, 1000$  and the number of simulations within each setup was  $nr = 500$ . The dimensions of the processes  $\mathbf{X}(s)$  and  $\mathbf{Y}(t)$  are  $p = 2$  and  $q = 2, 4$ . The functional observations corresponding to these processes are represented by their values in an equally spaced grid of 50 points  $s_1 = t_1, \dots, s_{50} = t_{50}$  in  $I_1 = I_2 = [0, 1]$ , which are generated in the following way:

$$\begin{bmatrix} \mathbf{x}_r(s_u) \\ \mathbf{y}_r(t_u) \end{bmatrix} = \begin{bmatrix} \Phi(s_u) & \mathbf{0} \\ \mathbf{0} & \Psi(t_u) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_r \\ \boldsymbol{\beta}_r \end{bmatrix} + \boldsymbol{\varepsilon}_{ru},$$

where  $r = 1, \dots, n$ ,  $u = 1, \dots, 50$ , the matrices  $\Phi(s)$  and  $\Psi(t)$  are given in (2.2)-(2.3) and contain the Fourier basis functions only and  $K_i = 5$ ,  $i = 1, \dots, p$ ,  $L_j = 5$ ,  $j = 1, \dots, q$ ,  $(\boldsymbol{\alpha}_r^\top, \boldsymbol{\beta}_r^\top)^\top$  are  $5(p+q)$ -dimensional random vectors, and  $\boldsymbol{\varepsilon}_{ru}^\top = (\varepsilon_{ru1}, \dots, \varepsilon_{ru,p+q})$  are the measurement errors such that  $\varepsilon_{ruv} \sim N(0, 0.025a_{rv})$  and  $a_{rv}$  is the range of the  $v$ -th row of the following matrix:

$$\begin{bmatrix} \Phi(s_1)\boldsymbol{\alpha}_r & \dots & \Phi(s_{50})\boldsymbol{\alpha}_r \\ \Psi(t_1)\boldsymbol{\beta}_r & \dots & \Psi(t_{50})\boldsymbol{\beta}_r \end{bmatrix}.$$

We consider the covariance matrices  $\Sigma$  with  $\Sigma_{11} = \mathbf{I}_{5p}$ ,  $\Sigma_{22} = \mathbf{I}_{5q}$ , and

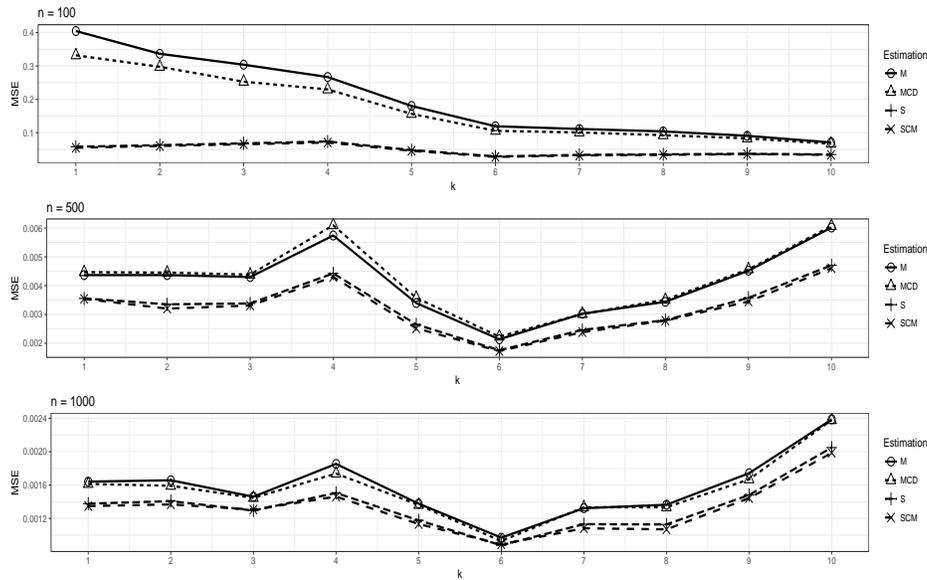
$$\Sigma_{12} = \begin{cases} \text{diag}(0.9, 0.8, 0.7, 0.6, 0.55, 0.5, 0.4, 0.3, 0.2, 0.1), & \text{for } q = 2, \\ (\text{diag}(0.9, 0.8, 0.7, 0.6, 0.55, 0.5, 0.4, 0.3, 0.2, 0.1), \mathbf{0}_{5p \times 5p}), & \text{for } q = 4. \end{cases}$$

Similarly to Branco et al. [5], we consider the following four cases of generating random vectors  $(\boldsymbol{\alpha}_r^\top, \boldsymbol{\beta}_r^\top)^\top$ ,  $r = 1, \dots, n$ :

NOR – normal distribution:  $(\boldsymbol{\alpha}_r^\top, \boldsymbol{\beta}_r^\top)^\top \sim N_{5(p+q)}(\mathbf{0}_{5(p+q)}, \Sigma)$ ,

T3 – multivariate  $t_3$ -distribution with scatter parameter  $\Sigma$ :  $(\boldsymbol{\alpha}_r^\top, \boldsymbol{\beta}_r^\top)^\top = \mathbf{T}_r / \sqrt{C_r/3}$ , where  $\mathbf{T}_r \sim N_{5(p+q)}(\mathbf{0}_{5(p+q)}, \Sigma)$  and  $C_r \sim \chi_3^2$ ,

SCN – symmetric contamination:  $(\boldsymbol{\alpha}_r^\top, \boldsymbol{\beta}_r^\top)^\top \sim N_{5(p+q)}(\mathbf{0}_{5(p+q)}, \Sigma)$  for  $r = 1, \dots, 0.95n$ , and  $(\boldsymbol{\alpha}_r^\top, \boldsymbol{\beta}_r^\top)^\top \sim N_{5(p+q)}(\mathbf{0}_{5(p+q)}, 9\Sigma)$  for  $r = 0.96n, \dots, n$ ,



**Figure 1.** Mean squared error for canonical correlations in NOR case ( $p = 2, q = 4, nr = 500$ , SCM – sample covariance matrix).

ACN – asymmetric contamination:  $(\alpha_r^\top, \beta_r^\top)^\top \sim N_{5(p+q)}(\mathbf{0}_{5(p+q)}, \Sigma)$  for  $r = 1, \dots, 0.95n$ , while for  $r = 0.96n, \dots, n$ ,  $(\alpha_r^\top, \beta_r^\top)^\top = 5\mathbf{1}_{5(p+q)}, 7.5\mathbf{1}_{5(p+q)}$  when  $q = 2, 4$ , respectively.

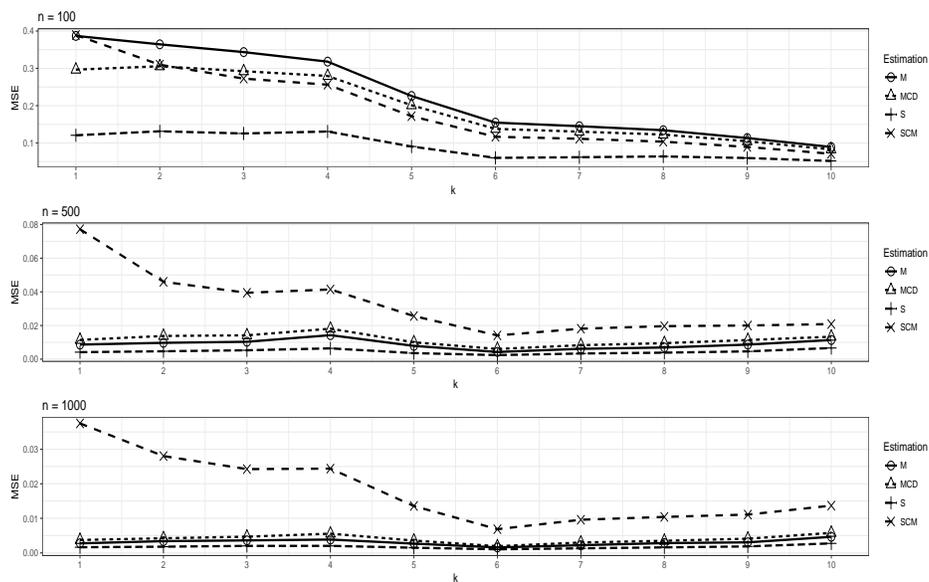
Therefore, we consider the classical (NOR) and non-normal (T3) models as well as the models with different outlying observations (SCN, ACN).

We set  $\lambda = 0$ . Then, Theorem 2.1 shows that the canonical correlations for the random processes  $\mathbf{X}(s)$  and  $\mathbf{Y}(t)$  generated as above are determined by the matrix  $\Sigma$ , i.e., they are equal to 0.9, 0.8, 0.7, 0.6, 0.55, 0.5, 0.4, 0.3, 0.2, 0.1. Since the Fourier basis is used to generate the data, the MFCCA is performed with the B-spline basis with  $K_i = 5, i = 1, \dots, p, L_j = 5, j = 1, \dots, q$  (Other basis does not change the canonical correlations.). To compare the true canonical correlations with their estimators  $\hat{\rho}_{kl}$ ,  $k = 1, \dots, 5p, l = 1, \dots, nr$ , obtained in simulation replications, we compute the following measure of mean squared error (MSE) as in [5]:

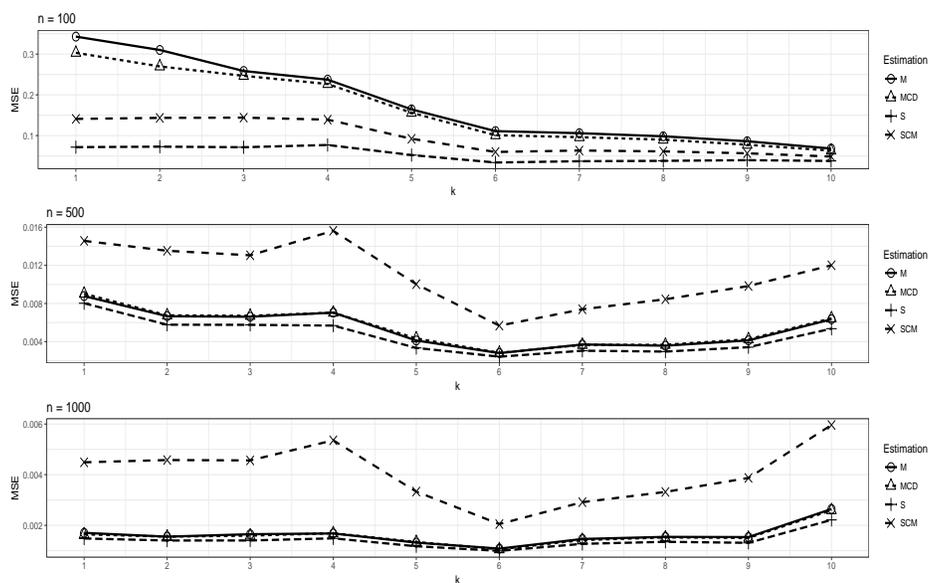
$$(4.1) \quad \text{MSE}(\hat{\rho}_k) = \frac{1}{nr} \sum_{l=1}^{nr} (\phi(\hat{\rho}_{kl}) - \phi(\rho_k))^2,$$

where  $\phi(x) = \tanh^{-1}(x)$  is the Fisher transformation.

To conduct simulation experiments, the R program was used [33]. The cross product matrices in  $\mathbf{J}_\Phi$  and  $\mathbf{J}_\Psi$  and the roughness penalty matrices in  $\mathbf{R}_\Phi$  and  $\mathbf{R}_\Psi$  can be approximated by using the functions `inprod()` and `getbasispenalty()` from the `fda` package [32]. The functions `create.fourier.basis()` and `eval.basis()` (resp. `create.bspline.basis()` and `Data2fd()`) available in this package were used to compute the values of the Fourier basis functions (resp. estimate the coefficients of the basis functions representation of the data). To estimate the covariance matrix  $\Sigma$  by the MCD, S- and M-estimators, the functions `CovMcd()`, `CovSest()` and `CovMest()` from the `rrcov` package [42] were applied. The default values of these functions were used.

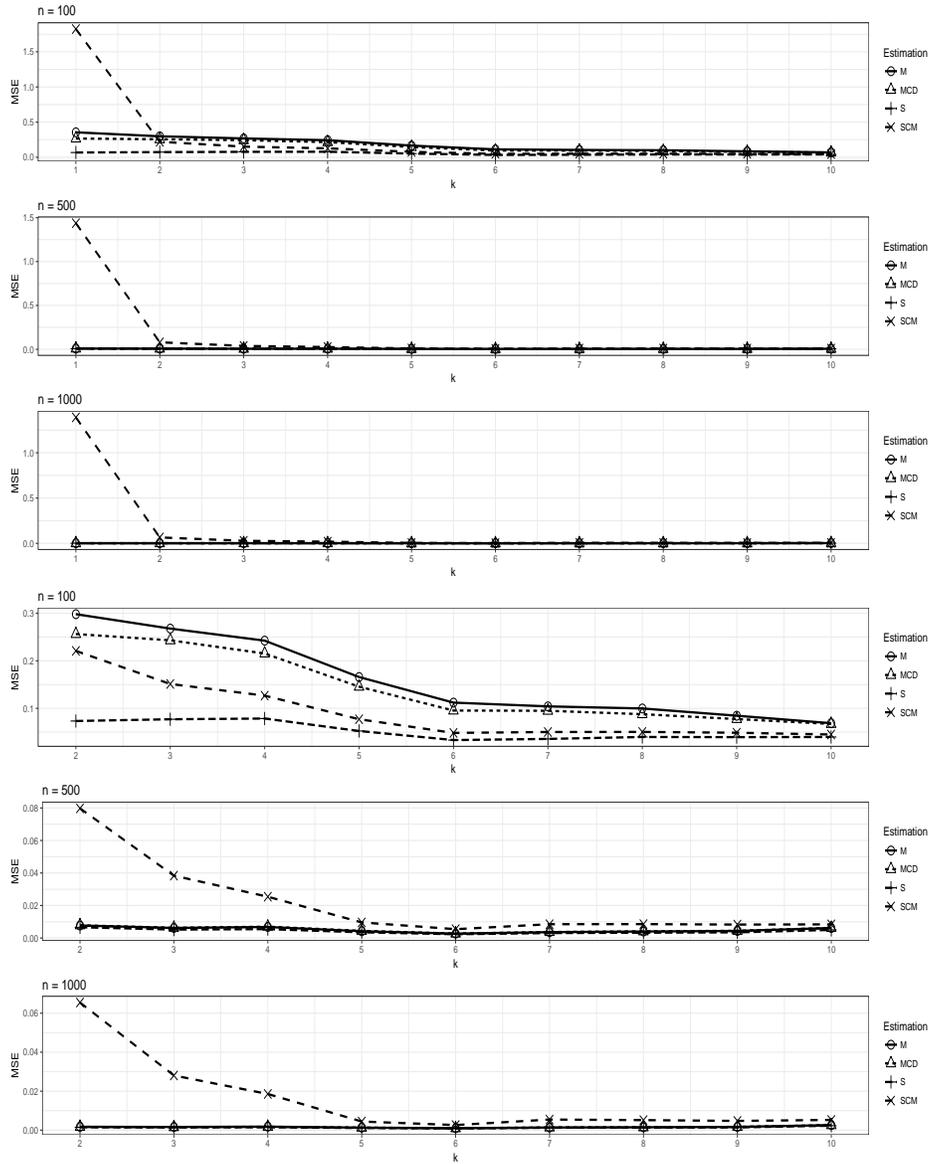


**Figure 2.** Mean squared error for canonical correlations in T3 case ( $p = 2, q = 4, nr = 500$ , SCM – sample covariance matrix).



**Figure 3.** Mean squared error for canonical correlations in SCN case ( $p = 2, q = 4, nr = 500$ , SCM – sample covariance matrix).

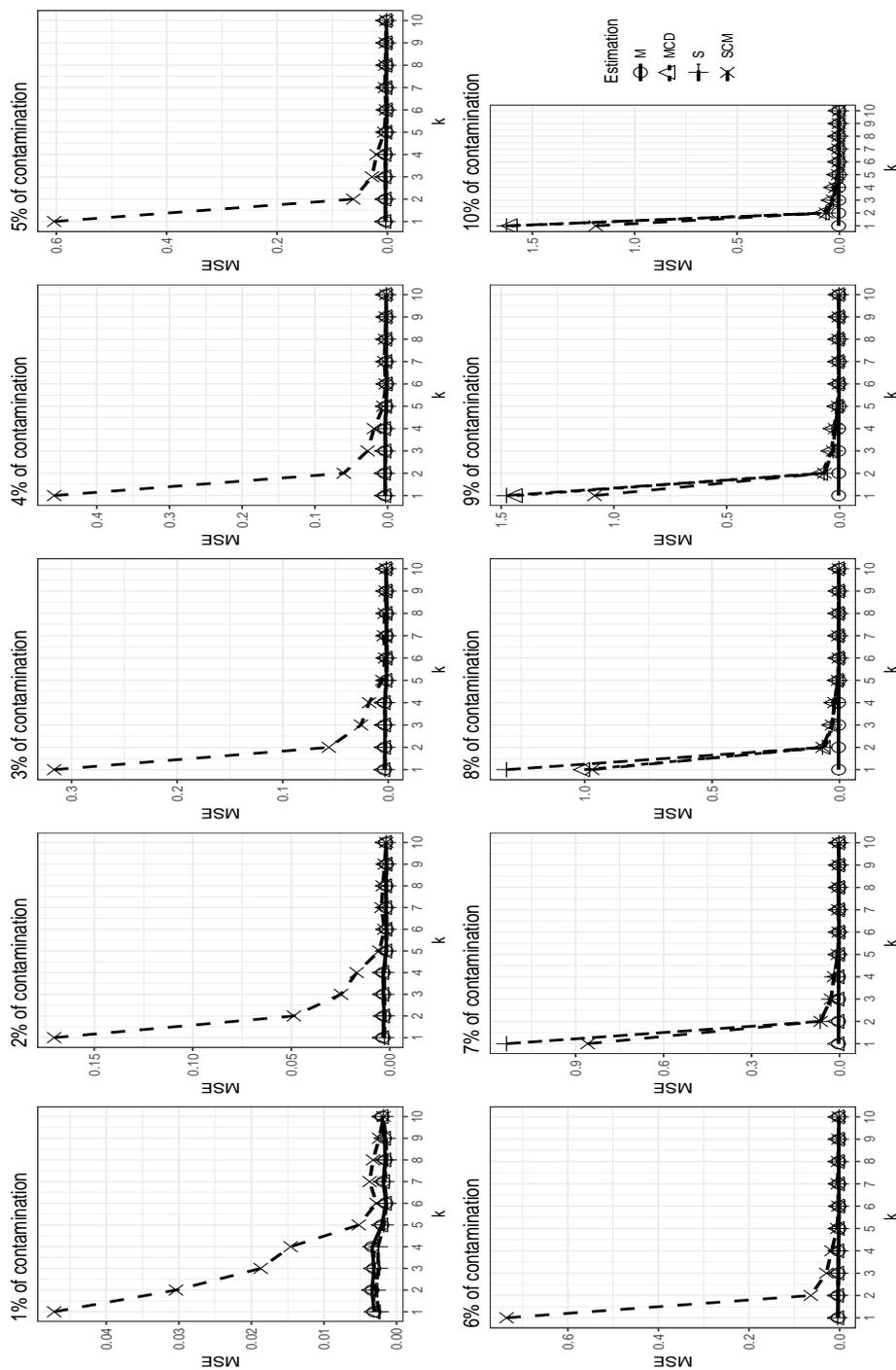
**4.2. Simulation results.** The simulation results are depicted in Figures 1-4. They show the mean squared error for canonical correlations (4.1) obtained by using the sample



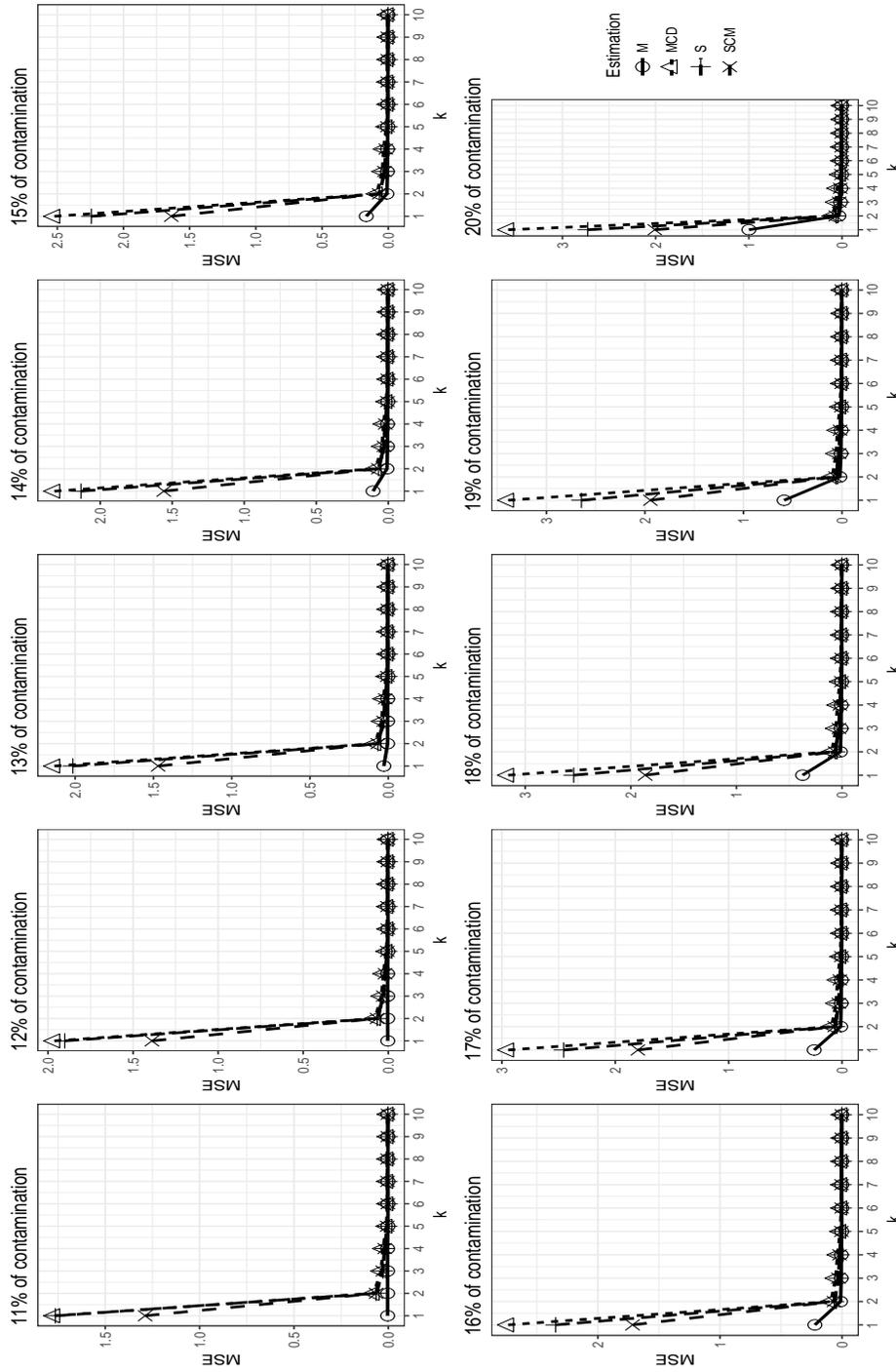
**Figure 4.** Mean squared error for canonical correlations in ACN case ( $p = 2, q = 4, nr = 500$ , SCM – sample covariance matrix).

covariance matrix (SCM) and the MCD, M-, S-estimators for estimating the covariance matrix, when  $p = 2$  and  $q = 4$ . For the dimensions  $p = q = 2$ , similar results were obtained, and therefore they are not presented. We only mention that the mean squared errors for  $p = q = 2$  are usually smaller than those for  $p = 2$  and  $q = 4$ .

In general, the mean squared errors obtained in ACN case are greater than those for T3 case, which are greater than the MSEs for SCN case. As expected, the smallest



**Figure 5.** Breakdown plot: Mean squared error for canonical correlations and for the percentage of contamination, ranging from 1 to 10%, in ACN case ( $p = q = 2$ ,  $n = 500$ ,  $nr = 500$ , SCM – sample covariance matrix).



**Figure 6.** Breakdown plot: Mean squared error for canonical correlations and for the percentage of contamination, ranging from 11 to 20%, in ACN case ( $p = q = 2$ ,  $n = 500$ ,  $nr = 500$ , SCM – sample covariance matrix).

mean squared errors are obtained in the normal model (NOR). We also observe that the MSEs for all estimation methods decrease as the number of observations increases. In cases T3 and ACN, the estimates of lower order canonical correlations lead usually to the largest MSEs. In the other cases, the largest MSE's may be also obtained for higher order canonical correlations.

In the normal model (NOR), the sample covariance matrix is the most precise. However, the S-estimator gives only slightly worse results than the SCM. The other robust estimators perform very similarly. When deviating from the normal model (the cases T3, SCN, ACN), the sample covariance matrix is overcome by at least one of the robust methods. For  $n = 100$ , the smallest mean squared errors are obtained by using the S-estimator, and the classical method works similarly to or better than the MCD and M-estimates. For greater number of observations ( $n = 500, 1000$ ), all robust estimators significantly outperform the sample covariance matrix. The robust methods give similar results, but the S-estimator seems to perform slightly better than the other ones.

Finally, we study the robustness of the covariance matrix estimators in MFCCA by breakdown analysis. Namely, we investigate the sensitivity of considered estimators to increasing amounts of contamination (proportion of atypical points) in the data. For this purpose, we carried out additional simulation study in ACN case, when  $p = q = 2$ ,  $n = 500$  and the percentage of contamination ranges from 1 to 20%. The resulting mean squared errors for all methods are presented in Figures 5-6. We observe that the MSEs of the sample covariance matrix rapidly increase in presence of contamination. The classical method performs poorly even when the percentage of contamination is very small. The S- and MCD estimators are more stable. Their MSEs remain small up to 6% and 7% of contamination, respectively, but then they also go up, especially for the first canonical correlation. In general, the M-estimator performs best and seems to be very stable up to about 16% of contamination.

## 5. Conclusions

We have considered the canonical correlation analysis for multivariate functional data based on the basis functions representation of the data. In contrast to Górecki et al. [15], we have developed this analysis in such a way to be available also for using non-orthogonal bases. Moreover, the robust estimation methods of the covariance structure were investigated to increase the performance of the multivariate functional canonical correlation analysis under non-normal models and in the presence of outlying observations. This performance was studied in simulations and breakdown analysis. Their results have indicated that the new estimation methods perform usually better in the presence of outliers and are more robust to deviations from the normal model than the sample covariance matrix, which is the classical estimator.

## References

- [1] Aneiros, G., Bongiorno, E.G., Cao, R. and Vieu, P. (Eds) *Functional statistics and related fields*. Contributions to Statistics (Springer, 2017).
- [2] Bobelyn, E., Serban, A.S., Nicu, M., Lammertyn, J., Nicolai, B.M. and Saeys, W. *Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance*. Postharvest Biol. Technol. **55**, 133–143, 2010.
- [3] Boente, G., Barrera, M.S. and Tyler, D.E. *A characterization of elliptical distributions and some optimality properties of principal components for functional data*. J. Multivariate Anal. **131**, 254–264, 2014.
- [4] Bongiorno, E.G., Salinelli, E., Goia, A. and Vieu, P. *Contributions in infinite-dimensional statistics and related topics* (Societa Editrice Esculapio, 2014).

- [5] Branco, J.A., Croux, C., Filzmoser, P. and Oliveira, M.R. *Robust canonical correlations: A comparative study*. *Comput. Statist.* **20**, 203–229, 2005.
- [6] Chen, D., Hall, P. and Müller, H.G. *Single and multiple index functional regression models with nonparametric link*. *Ann. Statist.* **39**, 1720–1747, 2011.
- [7] Collazos, J.A.A., Dias, R. and Zambom, A.Z. *Consistent variable selection for functional regression models*. *J. Multivariate Anal.* **146**, 63–71, 2016.
- [8] Croux, C. and Haesbroeck, G. *Influence function and efficiency of the minimum covariance determinant scatter matrix estimator*. *J. Multivariate Anal.* **71**, 161–190, 1999.
- [9] Davies, P.L. *Asymptotic behavior of  $S$ -estimators of multivariate location parameters and dispersion matrices*. *Ann. Statist.* **15**, 1269–1292, 1987.
- [10] Davies, P.L. *An efficient Fréchet-differentiable high breakdown multivariate location and dispersion estimator*. *J. Multivariate Anal.* **40**, 311–327, 1992.
- [11] Delaigle, A. and Hall, P. *Achieving near perfect classification for functional data*. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74**, 267–286, 2012.
- [12] Febrero-Bande, M., Galeano, P. and González-Manteiga, W. *Outlier detection in functional data by depth measures, with application to identify nonnormal  $NO_x$  levels*. *Environmetrics* **19**, 331–345, 2008.
- [13] Ferraty, F. and Vieu, P. *Nonparametric functional data analysis: Theory and practice* (Springer, New York, 2006).
- [14] Fremdt, S., Horváth, L., Kokoszka, P. and Steinebach, J.G. *Functional data analysis with increasing number of projections*. *J. Multivariate Anal.* **124**, 313–332, 2014.
- [15] Górecki, T., Krzyśko, M., Waszak, Ł. and Wołyński, W. *Selected statistical methods of data analysis for multivariate functional data*. *Statist. Papers* **59**, 153–182, 2018.
- [16] Górecki, T. and Smaga, Ł. *A comparison of tests for the one-way ANOVA problem for functional data*. *Comput. Stat.* **30**, 987–1010, 2015.
- [17] Górecki, T. and Smaga, Ł. *Multivariate analysis of variance for functional data*. *J. Appl. Stat.* **44**, 2172–2189, 2017.
- [18] Hilgert, N., Mas, A. and Verzelen, N. *Minimax adaptive tests for the functional linear model*. *Ann. Statist.* **41**, 838–869, 2013.
- [19] Horváth, L. and Kokoszka, P. *Inference for functional data with applications* (Springer, New York, 2012).
- [20] Huber, P.J. *Robust estimation of a location parameter*. *Ann. Math. Statist.* **35**, 73–101, 1964.
- [21] James, G.M. and Hastie, T.J. *Functional linear discriminant analysis for irregularly sampled curves*. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63**, 533–550, 2001.
- [22] Kent, J.T. and Tyler, D.E. *Constrained  $M$ -estimation for multivariate location and scatter*. *Ann. Statist.* **24**, 1346–1370, 1996.
- [23] Kokoszka, P., Oja, H., Park, B. and Sangalli, L. *Special issue on functional data analysis*. *Econometrics and Statistics* **1**, 99–100, 2017.
- [24] Lopuhaä, H.P. *On the relation between  $S$ -estimators and  $M$ -estimators of multivariate location and covariance*. *Ann. Statist.* **17**, 1662–1683, 1989.
- [25] Lopuhaä, H.P. *Multivariate  $\tau$ -estimators for location and scatter*. *Can. J. Statist.* **19**, 307–321, 1991.
- [26] Long, W., Li, N., Wang, H. and Cheng, S. *Impact of US financial crisis on different countries: Based on the method of functional analysis of variance*. *Procedia Computer Science* **9**, 1292–1298, 2012.
- [27] Maronna, R.A. *Robust  $M$ -estimators of multivariate location and scatter*. *Ann. Statist.* **1**, 51–67, 1976.
- [28] Martínez-Camblor, P. and Corral, N. *Repeated measures analysis for functional data*. *Comput. Statist. Data Anal.* **55**, 3244–3256, 2011.
- [29] Matsui, H. and Konishi, K. *Variable selection for functional regression models via the  $L_1$  regularization*. *Comput. Statist. Data Anal.* **55**, 3304–3310, 2011.
- [30] Ogden, R.T., Miller, C.E., Takezawa, K. and Ninomiya, S. *Functional regression in crop lodging assessment with digital images*. *J. Agric. Biol. Environ. Stat.* **7**, 389–402, 2002.
- [31] Ramsay, J.O. and Silverman, B.W. *Functional data analysis*, 2nd edition. (Springer, New York, 2005).

- [32] Ramsay, J.O., Wickham, H., Graves, S. and Hooker, G. `fda` - *Functional data analysis*. R package version 2.4.7, 2017. <http://CRAN.R-project.org/package=fda>
- [33] R Core Team *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. <https://www.R-project.org/>
- [34] Reimherr, M. *Functional regression with repeated eigenvalues*. *Statist. Probab. Lett.* **107**, 62–70, 2015.
- [35] Rousseeuw, P.J. *Multivariate estimation with high breakdown point*. in: W. Grossmann, G. Pflug, I. Vincze, W. Wertz (Eds), *Mathematical Statistics and Applications*, volume B (Reidel Publishing, Dordrecht, 1985), 283–297.
- [36] Rousseeuw, P.J. and Van Driessen, K. *A fast algorithm for the minimum covariance determinant estimator*. *Technometrics* **41**, 212–223, 1999.
- [37] Rousseeuw, P.J. and Yohai, V.J. *Robust regression by means of S-estimators*. in: *Robust and Nonlinear Time Series Analysis*. *Lecture Notes in Statistics* **26** (Springer, New York, 1984), 256–272.
- [38] Ruppert, D. *Computing S-estimators for regression and multivariate location/dispersion*. *J. Comput. Graph. Statist.* **1**, 253–270, 1992.
- [39] Salibian-Barrera, M. and Yohai, V. *A fast algorithm for S-regression estimates*. *J. Comput. Graph. Statist.* **15**, 414–427, 2006.
- [40] Shmueli, G. *To explain or to predict?* *Statist. Sci.* **25**, 289–310, 2010.
- [41] Smaga, Ł. *Repeated measures analysis for functional data using Box-type approximation - with applications*. *REVSTAT*, 2017. (To appear)
- [42] Todorov, V. and Filzmoser, P. *An object-oriented framework for robust multivariate analysis*. *Journal of Statistical Software* **32**, 1–47, 2009.
- [43] Tyler, D.E. *Finite sample breakdown points of projection based multivariate location and scatter statistics*. *Ann. Statist.* **22**, 1024–1044, 1994.
- [44] Yamamoto, M. and Terada, Y. *Functional factorial K-means analysis*. *Comput. Statist. Data Anal.* **79**, 133–148, 2014.
- [45] Zhang, J.T. *Analysis of variance for functional data* (Chapman & Hall, London, 2013).
- [46] Zuo, Y. *Robust location and scatter estimators in multivariate analysis*. in: J. Fan and H.L. Koul (Eds) *Frontiers of Statistics (in honor of Professor P.J. Bickel's 65th Birthday)*, Imperial College, 467–490, 2006.