Statistical Papers (2026) 67:74
https://doi.org/10.1007/s00362-026-01853-8

REGULAR ARTICLE

Two approaches to multiple canonical correlation analysis for repeated measures data

Tomasz Górecki[1] · **Mirosław Krzyśko**[2] · **Felix Gnettner**[3] · **Piotr Kokoszka**[4]

Received: 4 November 2025 / Revised: 25 March 2026

Abstract

In classical canonical correlation analysis (CCA), the goal is to determine the linear transformations of two random vectors into two new random variables that are most strongly correlated. Canonical variables are pairs of these new random variables, while canonical correlations are correlations between these pairs. In this paper, we propose and study two generalizations of this classical method: (1) Instead of two random vectors, we study more complex data structures that appear in important applications. In these structures, there are L features, each described by p_l scalars, $1 \leq l \leq L$. We observe n such objects over T time points. We derive a suitable analog of the CCA for such data. Our approach relies on embeddings into Reproducing Kernel Hilbert Spaces, and covers several related data structures as well. (2) We develop an analogous approach for multidimensional random processes. In this case, the experimental units are multivariate continuous, square-integrable functions over a given interval. These functions are modeled as elements of a Hilbert space, so in this case, we define the multiple functional canonical correlation analysis, MFCCA. We justify our approaches by applying them to two datasets and by appealing to suitably large-sample theory. We derive consistency rates for the related transformation and correlation estimators, and show that it is possible to relax two common assumptions on the compactness of the underlying cross-covariance operators and the independence of the data.

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

Mirosław Krzyśko
m.krzysko@uniwersytetkaliski.edu.pl

Felix Gnettner
Felix.Gnettner@sdstate.edu

Piotr Kokoszka
Piotr.Kokoszka@colostate.edu

1 Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznan, Poland

2 Interfaculty Department of Mathematics and Statistics, University of Kalisz, Kalisz, Poland

3 Department of Mathematics and Statistics, South Dakota State University, Brookings, SD, USA

4 Department of Statistics, Colorado State University, Fort Collins, CO, USA

Published online: 27 May 2026

Keywords Multiple kernel canonical variables · Multiple functional canonical variables · Multivariate repeated measures data · Multivariate functional data · Non-compact cross-covariance operator · Dependent data · Consistency rates

1 Introduction

Canonical Correlation Analysis (CCA), proposed by Hotelling (1936), is often used to study relationships between two feature sets (Adrover and Donato 2015; Krafty and Hall 2013; Langworthy et al. 2021; Ma and Li 2020; Shu et al. 2020). Linear transformations of the features in both sets (canonical variables) are constructed so that they are uncorrelated within each set, while the correlations between them (canonical correlations) are maximal. The objective is to maximize the correlation between data projections in the feature spaces. This reveals the underlying structural relationships between the two feature sets, quantifying how much variability in one set is explained by the other. Such approaches are useful for determining whether there exists a linear, or, in more advanced cases, a nonlinear mapping that transforms one set of features into another.

In this paper, the concepts and techniques of CCA are analyzed in the case of more than two feature sets. Multiple CCA aims to identify underlying patterns of correlation among more than two feature sets by finding transformations of the variables within each set that are maximally correlated across the sets. Multiple CCA extends the concept of CCA to identify shared relationships in complex multi-set data, such as across multiple data matrices or different measurement types, thereby revealing common structures or patterns that might be missed by analyzing pairs of sets individually. After estimating the optimal transformations, pairwise scatter plots of the transformed feature sets offer a powerful tool to visualize these patterns and structures.

A motivating data set

To motivate the methodology developed in this paper, we consider the Global Competitiveness Index (GCI) dataset. This example is developed further in Sect. 4.2. For $n = 115$ countries, $L = 12$ economic competitiveness features have been recorded. Two of them are institutions ($l = 1$) and infrastructure ($l = 2$). To each feature l belongs a number p_l of scalar indicators. The data have been recorded annually for $T = 10$ years.

For a feature l in country k, these observations can be encoded in a $T \times p_l$ matrix

$$\boldsymbol{A}_l[k] = \begin{pmatrix} a_{11} & \cdots & a_{1p_l} \\ \vdots & \vdots & \vdots \\ a_{T1} & \cdots & a_{Tp_l} \end{pmatrix}, \; k \in \{1, \ldots, n\}, \quad l \in \{1, \ldots, L\}.$$

The columns of the matrix $\boldsymbol{A}_l[k]$ can be thought of as time courses, or repeated measurements, of the p_l indicators for country k. The number and nature of scalar indicators

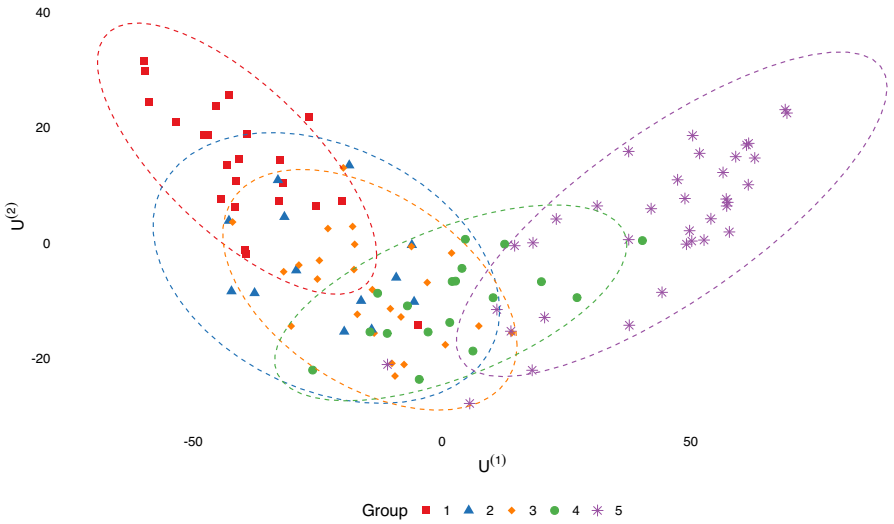

Fig. 1 Scatterplot for the optimally transformed feature pairs in the GCI data set (115 countries and five groups) in the system of the first two multiple kernel canonical variables ($U^{(1)}$, $U^{(2)}$) (with 95% confidence normal ellipses). The optimal transformations were determined by multiple kernel CCA for multivariate repeated measures data, as described in Sect. 2.

may vary from feature to feature, but the experimental units and the number T must be the same across all features.

The most elementary version of the multiple canonical correlations problem in the above setting is the optimization problem

$$\max_{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_L} \sum_{i=1}^{L}\sum_{j=1}^{L} \mathrm{Corr}(\boldsymbol{A}_i[1]\boldsymbol{u}_i, \boldsymbol{A}_j[1]\boldsymbol{u}_j), \tag{1}$$

where for each feature $l \in \{1,\ldots,L\}$, only $n = 1$ country at only $T = 1$ time point are considered. This optimization problem is a population version; the correlations are not estimated. Each vector $\boldsymbol{u}_l$ is called a "weight vector", and the vector $\boldsymbol{y}_l = \boldsymbol{A}_l[1]\boldsymbol{u}_l$ is called a "component". The objective is to find weight vectors

$$[\boldsymbol{u}_1^*,\ldots,\boldsymbol{u}_L^*] = \arg\max_{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_L} \sum_{i=1}^{L}\sum_{j=1}^{L} \mathrm{Corr}(\boldsymbol{A}_i[1]\boldsymbol{u}_i, \boldsymbol{A}_j[1]\boldsymbol{u}_j)$$

that lead to linear feature combinations within each block with maximum correlation across all subjects.

For most applications, it is impossible to solve (1), because the underlying population correlation values are unknown. Thus, relaxing the problem in (1) by replacing the correlation expressions with estimators of the corresponding covariances, based on all $n = 115$ countries, and introducing a suitable normalization constraint is required. For a concise mathematical expression, we refer to (8).

The solution of such a sample-based problem is an estimator of the corresponding population optimum $[\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_L^*]$, which is required to compute the transformation mappings

$$U^{(l)}(\cdot) = \langle \boldsymbol{u}_l^*, \cdot \rangle, \qquad l \in \{1, \ldots, L\}. \tag{2}$$

Plugging each data point into this mapping yields a representation of the multiple CCA results. For a feature pair of interest, a scatter plot of the transformed data is produced, as depicted in Fig. 1 for the pair (1, 2). The different colors in this plot arise from the fact that the countries were, independently of the analysis, divided into 5 groups by experts of the World Economic Forum. The multiple CCA was performed globally across all countries, and it is clear that the clusters in Fig. 1 partially correspond to the group assignments.

Connections to previous research

The multiple CCA problem was first considered by Horst (1961) whose solution is called the "maximum correlation method". It, however, suffers from severe problems with numerical convergence (Ten Berge 1988). Kettenring (1971) proposed a different solution (also for $T = 1$) to this problem, and named it the SUMCOR. Also relevant to our work, Carroll (1968), Hwang et al. (2013), Hwang et al. (2012) extended the CCA to the case of several data sets. Their method is known as the "generalized canonical correlation analysis (GCCA)" or the "multiple-set canonical correlation analysis (MCCA)". It has been considered by many authors (e.g., Gower 1989; Gloaguen et al. 2020; Lafosse 1989; Park and Huh 1996; Tenenhaus and Tenenhaus 2011, 2014; Tenenhaus et al. 2015; Gifi 1990; Yanai 1998; Tenenhaus et al. 2017).

Canonical correlation analysis for univariate functional data was introduced in Leurgans et al. (1993), who studied two functional data sets. It is also explained in Chapter 4 of Horváth and Kokoszka (2012) where references to more recent research in the case of two functional data sets are given. The CCA of two data sets of Multidimensional Functional Data (MFCCA) was studied in Górecki et al. (2017), Górecki et al. (2018), Krzyśko and Smaga (2019). There has been extensive research on various aspects of multivariate and longitudinal functional data, including sparse observations (see e.g. Happ and Greven 2018; Li et al. 2021; Sartini et al. 2025), among many others. These studies, however, are not directly related to the multiple functional CCA studied in this paper. A recent contribution (Girka et al. 2024) extends CCA to a tensor setting. Sort et al. (2024) study functional generalized canonical correlation analysis for multiple longitudinal variables from an optimization perspective: the focus is on proving that a gradient-based procedure converges to a solution of the underlying optimization problem if the number of steps grows. The authors conjecture that the estimators of the canonical functions obtained in this way are consistent with the standard nonparametric rate, but they do not prove this. Consistency of such estimates is the main focus of our paper.

Contributions

We extend the methodology of multiple CCA to the new setting of time-dependent observations of the multiple features of interest obtained from different experimental units. We propose two approaches. Our first approach relies on kernel embeddings of each block into a Reproducing Kernel Hilbert space. We derive consistency rates that cover data scenarios in which the cross-covariance operators are not compact, or the observations in the sample have some dependence structure among the experimental units. Neither case, to the best of our knowledge, has been studied, but they are relevant for data scenarios that motivate our work. Assuming independence and compactness is often an assumption of convenience, so our theory has practical relevance. Our second approach is an extension of multiple CCA to functional data indexed by time, which we think is particularly suitable for the problems that motivate this work. In the functional context, we view the repeated measurements as smooth time courses. We show that the consistency rates derived for the multiple kernel CCA carry over to this extension. To illustrate the functional context, in the competitiveness index example, suppose $Y_{l,j}^{(k)}(t)$ denotes the value of indicator j in the feature l for country k in year t. Unless a country experiences a dramatic change, like a government overthrow or a financial crash, it is reasonable to assume that the values of the $Y_{l,j}^{(k)}(t)$ evolve smoothly from year to year. This will be true for other examples, like the performance of sectors of the economy or agriculture. In such cases, a regularization of the time trajectories might lead to better-performing procedures.

Unlike in regression, no specific response variables exist in the data structures we study. Our methodology does not rely on assigning the roles of predictor and response variables. The consistency of multiple CCA has not yet been theoretically analyzed in the literature.

Our framework provides a unified view of several extensions of CCA, allowing both finite-dimensional and functional data to be treated within the same operator-theoretic setting. This perspective facilitates a rigorous comparison between kernel- and function-based approaches. The proposed approaches may serve as a foundation for advances in consistent inference for high-dimensional dependent data and regularized and sparse variants of multiple CCA.

Organization of the paper

In Sect. 2, we derive the consistency rates of the multiple kernel CCA for repeated measures, while in Sect. 3, multiple functional CCA is introduced. Section 4 contains real data studies that demonstrate the usefulness of and compare the proposed methods. A summary and conclusions are presented in Sect. 5.

2 Multiple kernel canonical correlation analysis in the case of repeated measures data

In this section, we introduce multiple canonical correlation analysis and the corresponding estimators based on repeated measures data. We show that the consistency results in CCA can be extended to multiple CCA, including cases where

- there is dependence in the data among the experimental units. This often arises when the units represent different time periods or are related to each other because of their location.
- the underlying covariance operators are not compact. This is a common assumption in the literature, but it cannot be verified in practice. We present a simple example in which this assumption can be dropped.

Our theory builds on the classical kernel canonical correlation analysis problem (Akaho 2001; Alam et al. 2010; Alam and Fukumizu 2013, 2015; Bach and Jordan 2002; Bilenko and Gallant 2016; Fukumizu et al. 2007; Hardoon et al. 2004; Hardoon and Shawe-Taylor 2009; Lai and Fyfe 2000), where only two features are considered, i.e. $L = 2$. We begin the with a review of relevant definitions related to Reproducing Kernel Hilbert Spaces (RKHS) and their selected properties in order to facilitate the exposition that follows.

Suppose $[X_1, \ldots, X_L]$, are possibly dependent random elements with joint distribution $P^{[X_1,\ldots,X_L]}$, taking values in a general space $\times_{l=1}^{L} \mathcal{S}_l$. The random elements $[X_1, \ldots, X_L]$ correspond to $[\boldsymbol{A}_1[1], \ldots, \boldsymbol{A}_L[1]]$ in our motivating data example in the population case of $n = 1$, where no repeated measurements are available.

Definition 1 For each $l \in \{1, \ldots, L\}$ consider a measurable positive definite *kernel* $K_l : \mathcal{S}_l \times \mathcal{S}_l \to \mathbb{R}$, and the associated kernel embedding $x \mapsto K_l(\cdot, x)$ of $x \in \mathcal{S}_l$ into the reproducing kernel Hilbert space $\mathcal{H}_{K_l}$. The Moore-Aronszajn theorem Aronszajn (1950) guarantees that the inner product in $\mathcal{H}_{K_l}$ determines the kernel, i.e.

$$K_l(x, y) = \langle K_l(\cdot, x), K_l(\cdot, y)\rangle_{\mathcal{H}_{K_l}} \text{ for all } x, y \in \mathcal{S}_l.$$

Based on the notation introduced above, we can now define the population version of the optimization problem for multiple-kernel canonical correlation analysis.

Definition 2 The population multiple kernel CCA problem in the RKHS $\times_{l=1}^{L} \mathcal{H}_{K_l}$ is defined as

$$\max_{g_1 \in \mathcal{H}_{K_1}, \ldots, g_L \in \mathcal{H}_{K_L}} \sum_{i=1}^{L} \sum_{j \in \{1,\ldots,L\}\setminus\{i\}} \operatorname{Cov}(g_i(X_i), g_j(X_j)) \quad \text{s.t.} \sum_{k=1}^{L} \operatorname{Var}(g_k(X_k)) = L.$$

In contrast to Sect. 1, this optimization problem is formulated in the context of an RKHS rather than in the linear space spanned by the feature vectors. It thus allows us to model nonlinear dependencies among the features. For each feature $i \in \{1, \ldots, L\}$ the goal is to obtain a transformation $g_i : \mathcal{S}_i \to \mathbb{R}$ such that the optimization problem

in Definition 2 is solved. For $i \neq j, \boldsymbol{C}_{i,j} : \mathcal{H}_{K_j} \to \mathcal{H}_{K_i}$ denotes the cross-covariance operator between X_i and X_j and $\boldsymbol{C}_{i,i} : \mathcal{H}_{K_i} \to \mathcal{H}_{K_i}$ is the covariance operator of X_i.

Remark 1 Using the above covariance operator notation, the multiple kernel CCA problem can be written as

$$\rho_{\mathcal{F}} = \max_{g_1 \in \mathcal{H}_{K_1}, \ldots, g_L \in \mathcal{H}_{K_L}} \sum_{i=1}^{L} \sum_{j \in \{1,\ldots,L\} \setminus \{i\}} \langle g_i, \boldsymbol{C}_{i,j}\, g_j \rangle_{\mathcal{H}_{K_i}} \quad \text{s.t.} \sum_{k=1}^{L} \langle g_k, \boldsymbol{C}_{k,k}\, g_k \rangle_{\mathcal{H}_{K_k}} = L. \tag{3}$$

At first glance, it is unclear whether the optimization problem admits an easy-to-determine solution. The following remark provides a clarification. In particular, the constraint in (3) relaxes the classical formulations of CCA-related optimization problems in Tenenhaus et al. (2017) to ensure that the solutions have a structure that allows analyzing their consistency.

Remark 2 The solution of the multiple kernel CCA problem is given by the appropriately normalized eigenfunctions belonging to the largest eigenvalue $\rho_{\mathcal{F}}$ of the following generalized eigenvalue problem:

$$\sum_{j \in \{1,\ldots,L\} \setminus \{i\}} \boldsymbol{C}_{i,j}\, g_j = \rho \cdot \boldsymbol{C}_{i,i}\, g_i, \quad i \in \{1, \ldots, L\}. \tag{4}$$

This can be easily seen by computing the gradient of the Lagrangian of (3), i.e.

$$\nabla \mathcal{L}(g_1, \ldots, g_L; \lambda) = \begin{pmatrix} \sum_{j \in \{1,\ldots,L\} \setminus \{1\}} \boldsymbol{C}_{1,j}\, g_j \\ \vdots \\ \sum_{j \in \{1,\ldots,L\} \setminus \{L\}} \boldsymbol{C}_{L,j}\, g_j, \\ \sum_{k=1}^{L} \langle g_k, \boldsymbol{C}_{k,k}\, g_k \rangle_{\mathcal{H}_{K_k}} - L \end{pmatrix} + \lambda \cdot \begin{pmatrix} \boldsymbol{C}_{1,1}\, g_1 \\ \vdots \\ \boldsymbol{C}_{L,L}\, g_L \\ 0 \end{pmatrix}.$$

Solving $\nabla \mathcal{L}(g_1, \ldots, g_L; \lambda) = (0, \ldots, 0)^T$, where the last component is only for normalization purposes, leads to the generalized eigenvalue problem (4), which can be written as

$$\begin{pmatrix} 0 & \boldsymbol{C}_{1,2} & \boldsymbol{C}_{1,3} & \cdots & \boldsymbol{C}_{1,L} \\ \boldsymbol{C}_{2,1} & 0 & \boldsymbol{C}_{2,3} & \cdots & \boldsymbol{C}_{2,L} \\ \boldsymbol{C}_{3,1} & \boldsymbol{C}_{3,2} & 0 & \cdots & \boldsymbol{C}_{3,L} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{C}_{L,1} & \boldsymbol{C}_{L,2} & \boldsymbol{C}_{L,3} & \cdots & 0 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_L \end{pmatrix} = \rho \begin{pmatrix} \boldsymbol{C}_{1,1} & 0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{C}_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & \boldsymbol{C}_{3,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \boldsymbol{C}_{L,L} \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_L \end{pmatrix}.$$

If $\boldsymbol{C}_{1,1}, \ldots, \boldsymbol{C}_{L,L}$ are strictly positive definite, by setting $g_i = \boldsymbol{C}_{i,i}^{-\frac{1}{2}} f_i$ this generalized eigenvalue problem can be easily transformed into an ordinary eigenvalue problem, which reads

$$\begin{pmatrix} 0 & \boldsymbol{C}_{1,1}^{-1/2}\boldsymbol{C}_{1,2}\boldsymbol{C}_{2,2}^{-1/2} & \boldsymbol{C}_{1,1}^{-1/2}\boldsymbol{C}_{1,3}\boldsymbol{C}_{3,3}^{-1/2} & \cdots & \boldsymbol{C}_{1,1}^{-1/2}\boldsymbol{C}_{1,L}\boldsymbol{C}_{L,L}^{-1/2} \\ \boldsymbol{C}_{2,2}^{-1/2}\boldsymbol{C}_{2,1}\boldsymbol{C}_{1,1}^{-1/2} & 0 & \boldsymbol{C}_{2,2}^{-1/2}\boldsymbol{C}_{2,3}\boldsymbol{C}_{3,3}^{-1/2} & \cdots & \boldsymbol{C}_{2,2}^{-1/2}\boldsymbol{C}_{2,L}\boldsymbol{C}_{L,L}^{-1/2} \\ \boldsymbol{C}_{3,3}^{-1/2}\boldsymbol{C}_{3,1}\boldsymbol{C}_{1,1}^{-1/2} & \boldsymbol{C}_{3,3}^{-1/2}\boldsymbol{C}_{3,2}\boldsymbol{C}_{2,2}^{-1/2} & 0 & \cdots & \boldsymbol{C}_{3,3}^{-1/2}\boldsymbol{C}_{3,L}\boldsymbol{C}_{L,L}^{-1/2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{C}_{L,L}^{-1/2}\boldsymbol{C}_{L,1}\boldsymbol{C}_{1,1}^{-1/2} & \boldsymbol{C}_{L,L}^{-1/2}\boldsymbol{C}_{L,2}\boldsymbol{C}_{2,2}^{-1/2} & \boldsymbol{C}_{L,L}^{-1/2}\boldsymbol{C}_{L,3}\boldsymbol{C}_{3,3}^{-1/2} & \cdots & 0 \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_L \end{pmatrix}$$

$$= \rho \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_L \end{pmatrix}. \quad (5)$$

Remark 3 Considering the optimization problem

$$\max_{g_1 \in \mathcal{H}_{K_1}, \ldots, g_L \in \mathcal{H}_{K_L}} \sum_{i=1}^{L} \sum_{j \in \{1,\ldots,L\} \setminus \{i\}} \langle g_i, \boldsymbol{C}_{i,j}\, g_j \rangle_{\mathcal{H}_{K_i}} \quad \text{s.t. } \langle g_k, \boldsymbol{C}_{k,k}\, g_k \rangle_{\mathcal{H}_{K_k}} = 1 \text{ for all } k \in \{1, \ldots, L\} \quad (6)$$

instead of (3) leads to mathematical difficulties. As pointed out in (Nielsen 2002, Sect. A, p. 296), (6) cannot be reduced to an ordinary generalized eigenvalue problem.

In the setting of multiple kernel CCA with repeated measurements, which is the sample version of the problem in (3), we deal with observations $[X_1^{(1)}, \ldots, X_L^{(1)}], \ldots, [X_1^{(n)}, \ldots, X_L^{(n)}]$ that can be assumed to be iid for a gentle introduction into the topic. The independence assumption can be replaced by some notion of weak dependence, see Example 2.

Replacing the population covariance operators in (3) by corresponding estimators

$$\langle g_i, \widehat{\boldsymbol{C}}_{i,j}^{(n)}\, g_j \rangle_{\mathcal{H}_{K_i}} = \frac{1}{n} \sum_{k=1}^{n} \left\langle g_i, K_i(\cdot, X_i^{(k)}) - \frac{1}{n} \sum_{\ell=1}^{n} K_i(\cdot, X_i^{(\ell)}) \right\rangle_{\mathcal{H}_{K_i}} \cdot \left\langle g_j, K_j(\cdot, X_j^{(k)}) - \frac{1}{n} \sum_{\ell=1}^{n} K_j(\cdot, X_j^{(\ell)}) \right\rangle_{\mathcal{H}_{K_j}} \quad (7)$$

yields

$$\widehat{\rho}_{\mathcal{F}} = \max_{g_1 \in \mathcal{H}_{K_1}, \ldots, g_L \in \mathcal{H}_{K_L}} \sum_{i=1}^{L} \sum_{j \in \{1,\ldots,L\} \setminus \{i\}} \langle g_i, \widehat{\boldsymbol{C}}_{i,j}^{(n)}\, g_j \rangle_{\mathcal{H}_{K_i}} \quad \text{s.t. } \sum_{k=1}^{L} \langle g_k, \widehat{\boldsymbol{C}}_{k,k}^{(n)}\, g_k \rangle_{\mathcal{H}_{K_k}} = L. \quad (8)$$

In contrast to the previous population-based problems, (8) incorporates covariance estimators based on data from a sample of n experimental units.

Solving this optimization problem via finding the solutions of a generalized eigenvalue problem as in Remark 2 requires some modification to keep it numerically

tractable. The inversion of the covariance operator estimates $\widehat{\boldsymbol{C}}_{i,i}^{(n)}$ is required for this purpose, but these usually do not have full rank. Consequently, some form of regularization is needed. Various regularizations have been studied in similar contexts (Alam et al. 2010; Alam and Fukumizu 2013; Bilenko and Gallant 2016; Tenenhaus and Tenenhaus 2011, 2014; Tenenhaus et al. 2015, 2017). Basically, without any regularization, it is possible to find an estimate $\widehat{\rho}_{\mathcal{F}}$ that indicates perfect dependence, and does not entail any meaningful information on the correlations of the features.

Replacing $\widehat{\boldsymbol{C}}_{i,i}^{(n)}$ by $\widehat{\boldsymbol{C}}_{i,i}^{(n)} + \epsilon_n \cdot \boldsymbol{I}$ with a positive decreasing sequence $(\epsilon_n)_{n\in\mathbb{N}}, \epsilon_n \to 0$, solves the invertibility issues. So, the regularized problem reads

$$\widehat{\rho}_{\mathcal{F}} = \max_{g_1\in\mathcal{H}_{K_1},\ldots,g_L\in\mathcal{H}_{K_L}} \sum_{i=1}^{L} \sum_{j\in\{1,\ldots,L\}\setminus\{i\}} \langle g_i, \widehat{\boldsymbol{C}}_{i,j}^{(n)}\, g_j\rangle_{\mathcal{H}_{K_i}} \text{ s.t. } \sum_{k=1}^{L} \langle g_k, (\widehat{\boldsymbol{C}}_{i,i}^{(n)} + \epsilon_n \cdot \boldsymbol{I})\, g_k\rangle_{\mathcal{H}_{K_k}} = L. \tag{9}$$

For the sake of lucidity, in the following, we state the eigenvalue problem (5) as

$$\mathfrak{C}_L\, \mathfrak{f} = \rho\, \mathfrak{f}, \tag{10}$$

and the regularized sample version (9) reads

$$\widehat{\mathfrak{C}}_{n,L}\, \widehat{\mathfrak{f}}_n = \widehat{\rho}\, \widehat{\mathfrak{f}}_n. \tag{11}$$

To obtain a simple numerical solution of (9), the Gram matrices of each feature are considered to get a computationally tractable formulation of the generalized eigenvalue problem.

Definition 3 Given a kernel K_l and a sample $X_l^{(1)}, \ldots, X_l^{(n)}$ with values in $\mathcal{S}_l$, the $n \times n$ matrix $\boldsymbol{G}_l$ with entries $K_l(X_l^{(i)}, X_l^{(n)})$ is called the *Gram matrix*, or the *kernel matrix*, of K_l with respect to the given sample. $\boldsymbol{G}_l$ is *positive semi-definite* if $\boldsymbol{c}^\top \boldsymbol{G}_l \boldsymbol{c} \geq 0$ holds for any $\boldsymbol{c} \in \mathbb{R}^n$

The centered kernel matrices $\widetilde{\boldsymbol{G}}_l$ are defined by Schölkopf et al. (1998) as

$$\widetilde{\boldsymbol{G}}_l = \boldsymbol{H}\boldsymbol{G}_l\boldsymbol{H}, \quad \boldsymbol{H} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top, \quad l \in \{1, \ldots, n\}. \tag{12}$$

The matrix $\boldsymbol{H}$ is called the *centering matrix*; $\boldsymbol{I}_n$ is the identity matrix of order n and $\mathbf{1}_n$ is the column-vector with n 1s. Elementary linear algebra shows that the entries of $\widetilde{\boldsymbol{G}}_l$ can be written as

$$(\widetilde{\boldsymbol{G}}_l)_{k,\ell} = K_l(X_l^{(k)}, X_l^{(\ell)}) - \frac{1}{n}\sum_{b=1}^{n} K_l(X_l^{(k)}, X_l^{(b)}) - \frac{1}{n}\sum_{a=1}^{n} K_l(X_l^{(a)}, X_l^{(\ell)}) + \frac{1}{n^2}\sum_{a=1}^{n}\sum_{b=1}^{n} K_l(X_l^{(a)}, X_l^{(b)}). \tag{13}$$

In our motivating data example, $\mathcal{S}_l = \mathbb{R}^{T\times p_l}$ and the repeated measurements $X_l^{(1)}, \ldots, X_l^{(n)}$ correspond to the blocks $\boldsymbol{A}_l[1], \ldots, \boldsymbol{A}_l[n]$ for each feature $l \in \{1, \ldots, n\}$. In this context, we consider Gaussian kernels $K_l : \mathbb{R}^{T\times p_l} \times \mathbb{R}^{T\times p_l} \to \mathbb{R}$, so that the corresponding Gram matrix has the entries

$$(\boldsymbol{G}_l)_{i,j} = K_l(\boldsymbol{A}_l[i], \boldsymbol{A}_l[j]) = \exp\left(-\gamma\|\boldsymbol{A}_l[i] - \boldsymbol{A}_l[j]\|_F^2\right), \quad i, j \in \{1, \ldots, n\}, \tag{14}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and γ is a fixed positive constant. Here, the kernel K_l can be regarded as a similarity measure between two elements $\boldsymbol{A}_l[i]$ and $\boldsymbol{A}_l[j]$, relevant to our specific real data analysis problem.

With the above preparation, we can formulate the regularized problem (9) by means of the centered kernel Gram matrices in (13). Since the range of of each $\widehat{\boldsymbol{C}}_{i,i}, i \in \{1, \ldots, L\}$, is spanned by $u_i^{(1)}, \ldots, u_i^{(n)}$ with $u_i^{(k)} = K_i(\cdot, X_i^{(k)}) - \frac{1}{n}\sum_{\ell=1}^n K_i(\cdot, X_i^{(\ell)})$, we can write each sololution $g_1, \ldots, g_l$ of the MCCA problem as a linear combination, i.e. there exist vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L \in \mathbb{R}^n$ such that

$$g_i = \sum_{k=1}^n w_i^{(k)} \cdot u_i^{(k)}, \quad i \in \{1, \ldots, L\}.$$

Thus, it holds

$$\begin{aligned}
\langle g_i, \widehat{\boldsymbol{C}}_{i,j}^{(n)}\, g_j\rangle_{\mathcal{H}_{K_i}} &= \frac{1}{n}\sum_{k=1}^n \left\langle g_i, u_i^{(k)}\right\rangle_{\mathcal{H}_{K_i}} \cdot \left\langle g_j, u_j^{(k)}\right\rangle_{\mathcal{H}_{K_j}} \\
&= \frac{1}{n}\sum_{k=1}^n \left\langle \sum_{l_1=1}^n w_i^{(l_1)} \cdot u_i^{(l_1)}, u_i^{(k)}\right\rangle_{\mathcal{H}_{K_i}} \cdot \left\langle \sum_{l_2=1}^n w_j^{(l_2)} \cdot u_j^{(l_2)}, u_j^{(k)}\right\rangle_{\mathcal{H}_{K_j}} \\
&= \frac{1}{n}\sum_{k=1}^n \sum_{l_1=1}^n \sum_{l_2=1}^n w_i^{(l_1)} \left\langle u_i^{(l_1)}, u_i^{(k)}\right\rangle_{\mathcal{H}_{K_i}} \left\langle u_j^{(l_2)}, u_j^{(k)}\right\rangle_{\mathcal{H}_{K_j}} w_j^{(l_2)} = \frac{\boldsymbol{w}_i^T \widetilde{\boldsymbol{G}}_i \widetilde{\boldsymbol{G}}_j \boldsymbol{w}_j}{n},
\end{aligned}$$

as well as

$$\langle g_i, (\widehat{\boldsymbol{C}}_{i,i} + \epsilon_n \cdot \boldsymbol{I})\, g_i\rangle_{\mathcal{H}_{K_i}} = \frac{\boldsymbol{w}_i^T \widetilde{\boldsymbol{G}}_i^2 \boldsymbol{w}_i}{n} + \epsilon_n \cdot \boldsymbol{w}_i^T \widetilde{\boldsymbol{G}}_i \boldsymbol{w}_i = \boldsymbol{w}_i^T \left(\frac{\widetilde{\boldsymbol{G}}_i^2}{n} + \epsilon_n \cdot \widetilde{\boldsymbol{G}}_i\right) \boldsymbol{w}_i,$$

where $\boldsymbol{w}_i^T = (w_i^{(1)}, \ldots, w_i^{(n)})$. We made use of the identity $\langle K_i(\cdot, X_i^{(l_1)}), K_i(\cdot, X_i^{(l_2)})\rangle_{\mathcal{H}_{K_i}} = K_i(X_i^{(l_1)}, X_j^{(l_2)})$, which implies that $\langle u_i^{(k)}, u_i^{(\ell)}\rangle_{\mathcal{H}_{K_i}} = (\widetilde{\boldsymbol{G}}_i)_{k,\ell}$.

These transformations yield the following formulation of the sample kernel MCCA problem (9):

$$\widehat{\rho}_{\mathcal{F}} = \max_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_L \in \mathbb{R}^n} \sum_{i=1}^L \sum_{j\in\{1,\ldots,L\}\setminus\{i\}} \frac{\boldsymbol{w}_i^T \widetilde{\boldsymbol{G}}_i \widetilde{\boldsymbol{G}}_j \boldsymbol{w}_j}{n} \quad \text{s.t.} \sum_{k=1}^L \boldsymbol{w}_i^T \left(\frac{\widetilde{\boldsymbol{G}}_i^2}{n} + \epsilon_n \cdot \widetilde{\boldsymbol{G}}_i\right) \boldsymbol{w}_i = L. \tag{15}$$

In block matrix notation, this reads

$$\frac{1}{n}\cdot\begin{pmatrix} 0 & \widetilde{\boldsymbol{G}}_1\widetilde{\boldsymbol{G}}_2 & \widetilde{\boldsymbol{G}}_1\widetilde{\boldsymbol{G}}_3 & \cdots & \widetilde{\boldsymbol{G}}_1\widetilde{\boldsymbol{G}}_L \\ \widetilde{\boldsymbol{G}}_2\widetilde{\boldsymbol{G}}_1 & 0 & \widetilde{\boldsymbol{G}}_2\widetilde{\boldsymbol{G}}_3 & \cdots & \widetilde{\boldsymbol{G}}_2\widetilde{\boldsymbol{G}}_L \\ \widetilde{\boldsymbol{G}}_3\widetilde{\boldsymbol{G}}_1 & \widetilde{\boldsymbol{G}}_3\widetilde{\boldsymbol{G}}_2 & 0 & \cdots & \widetilde{\boldsymbol{G}}_3\widetilde{\boldsymbol{G}}_L \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \widetilde{\boldsymbol{G}}_L\widetilde{\boldsymbol{G}}_1 & \widetilde{\boldsymbol{G}}_L\widetilde{\boldsymbol{G}}_2 & \widetilde{\boldsymbol{G}}_L\widetilde{\boldsymbol{G}}_3 & \cdots & 0 \end{pmatrix}\begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \boldsymbol{w}_3 \\ \vdots \\ \boldsymbol{w}_L \end{pmatrix}$$

$$= \rho\begin{pmatrix} \frac{\widetilde{\boldsymbol{G}}_1^2}{n} + \epsilon_n\cdot\widetilde{\boldsymbol{G}}_1 & 0 & 0 & \cdots & 0 \\ 0 & \frac{\widetilde{\boldsymbol{G}}_2^2}{n} + \epsilon_n\cdot\widetilde{\boldsymbol{G}}_2 & 0 & \cdots & 0 \\ 0 & 0 & \frac{\widetilde{\boldsymbol{G}}_3^2}{n} + \epsilon_n\cdot\widetilde{\boldsymbol{G}}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{\widetilde{\boldsymbol{G}}_L^2}{n} + \epsilon_n\cdot\widetilde{\boldsymbol{G}}_L \end{pmatrix}\begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \boldsymbol{w}_3 \\ \vdots \\ \boldsymbol{w}_L \end{pmatrix}.$$

Instead, the function space related representation (9), the above finite-dimensional representation is suitable for numerical solvers. We now formulate the assumptions for our result on the consistency of multiple kernel canonical correlations. Our theory is based on the proof strategies of Fukumizu et al. (2007), but we relax several of their assumptions.

Assumption 1 $\mathbb{E}(K_i(X_i, X_i)) < \infty$ holds for all $i \in \{1, \ldots, L\}$.

As we consider covariance operators in an RKHS setting, Assumption 1 is required to guarantee that the random variables $g_i(X_i)$ have finite second moments for $g_i \in \mathcal{H}_{K_i}, i \in \{1, \ldots, L\}$.

Assumption 2 The covariance operators $\boldsymbol{C}_{i,i}$, $i \in \{1, \ldots, L\}$, are strictly positive definite.

Assumption 2 ensures that the generalized eigenvalue problem can be transformed into the ordinary eigenvalue problem.

Assumption 3 For the sequence $(\epsilon_n)_{n\in\mathbb{N}}$ it holds $\epsilon_n \to 0$ as well as $n^{1/3}\epsilon_n \to \infty$, as $n \to \infty$.

As explained previously, some regularization is required because the covariance operator estimates do not have full rank and thus are not invertible. For our asymptotic considerations, the regularization sequence must converge to 0, but not too fast. In practice, the regularization parameter ϵ_n acts as a tuning parameter controlling the bias–variance trade-off in the estimation of the covariance operators. The theoretical conditions $\epsilon_n \to 0$ and $n^{1/3}\epsilon_n \to \infty$ ensure that the regularization decreases with the sample size but not too rapidly. In empirical applications, a convenient choice is $\epsilon_n = c\, n^{-\alpha}$ with $\alpha \in (0, 1/3)$ and a moderate constant c. Alternatively, data-driven procedures such as cross-validation or grid search may be used to select ϵ_n based on predictive or reconstruction criteria.

Assumption 4 The data $[X_1^{(1)}, \ldots, X_L^{(1)}], \ldots, [X_1^{(n)}, \ldots, X_L^{(n)}]$ have a dependence structure such that for any $(i, j) \in \{1, \ldots, n\}$ the estimator $\widehat{\boldsymbol{C}}_{i,j}$ is weakly $\sqrt{n}$-consistent in operator norm, i.e. $O_P(1/\sqrt{n})$,

By the multivariate CLT, Assumption 4 holds for under the usual assumption of independence and identical distribution across n, but also under notions of weak dependence that imply weak invariance principles.

While Assumptions 1–3 are fairly standard, Assumption 4 is more delicate. Below, we present two examples in which Assumption 4 is satisfied. Example 1 highlights that it is possible to drop the requirement of compact cross-covariance operators, which is imposed throughout the literature. Example 2 is focused on incorporating dependence between the experimental units. In fact, this is more realistic than assuming independence for most real data settings.

Example 1 Consider a bounded sample $[Z_1^{(1)}, Z_2^{(1)}], \ldots, [Z_1^{(n)}, Z_2^{(n)}]$ of iid pairs of random variables. $Z_1^{(i)}$ and $Z_2^{(i)}$ take values in L^2, and $\|Z_1^{(i)}\|_{\mathrm{L}^2} \leq a_1$, $\|Z_1^{(i)}\|_{\mathrm{L}^2} \leq a_2$ holds for some fixed constants $a_1, a_2 > 0$. Then, the covariance estimator

$$\widehat{\boldsymbol{C}}_{1,2} = \left(\frac{1}{n}\sum_{i=1}^{n} Z_1^{(i)} Z_2^{(i)\top}\right) - \left(\frac{1}{n}\sum_{j=1}^{n} Z_1^{(j)}\right)\left(\frac{1}{n}\sum_{\ell=1}^{n} Z_2^{(\ell)}\right)^{\top}$$

is weakly $\sqrt{n}$-consistent. This can be seen by

$$\begin{aligned}
&\|\widehat{\boldsymbol{C}}_{1,2} - \boldsymbol{C}_{1,2}\|_{\mathrm{op}} \\
&\leq \left\| \frac{1}{n}\sum_{i=1}^{n} Z_1^{(i)} Z_2^{(i)\top} - \mathbb{E}\left(Z_1^{(1)} Z_2^{(1)\top}\right) \right\|_{\mathrm{op}} + \left\| \left(\frac{1}{n}\sum_{j=1}^{n} Z_1^{(j)}\right)\left(\frac{1}{n}\sum_{\ell=1}^{n} Z_2^{(\ell)} - \mathbb{E}\left(Z_2^{(1)}\right)\right)^{\top} \right\|_{\mathrm{op}} \\
&\quad + \left\| \left(\frac{1}{n}\sum_{j=1}^{n} Z_1^{(j)} - \mathbb{E}\left(Z_1^{(1)}\right)\right)\left(\mathbb{E}\left(Z_2^{(1)}\right)\right)^{\top} \right\|_{\mathrm{op}} \\
&\leq \left\| \frac{1}{n}\sum_{i=1}^{n} Z_1^{(i)} Z_2^{(i)\top} - \mathbb{E}\left(Z_1^{(1)} Z_2^{(1)\top}\right) \right\|_{\mathrm{op}} + a_1 \cdot \left\| \left(\frac{1}{n}\sum_{\ell=1}^{n} Z_2^{(\ell)} - \mathbb{E}\left(Z_2^{(1)}\right)\right)^{\top} \right\|_{\mathrm{L}^2} \\
&\quad + a_2 \cdot \left\| \left(\frac{1}{n}\sum_{j=1}^{n} Z_1^{(j)} - \mathbb{E}\left(Z_1^{(1)}\right)\right)^{\top} \right\|_{\mathrm{L}^2},
\end{aligned}$$

because the spectral norm of a dyadic product equals the product of the L^2-norms of the vectors. For the first term on the right hand side, applying McDiarmid's inequality, respectively its generalization to random elements in Banach spaces (Pinelis 1994), implies weak $\sqrt{n}$-consistency, as the mean satisfies the bounded differences property (in the spectral norm) with upper bound

$$\frac{1}{n}\|Z_1^{(i)} Z_2^{(i)\top} - \widetilde{Z}_1^{(i)} \widetilde{Z}_2^{(i)\top}\|_{\mathrm{op}} \leq \frac{2a_1 a_2}{n}.$$

This is because the spectral norm of a dyadic product equals the product of the L^2-norms of the vectors. Thus, for any $\epsilon > 0$ McDiarmid's inequality entails

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} Z_1^{(i)} Z_2^{(i)\top} - \mathbb{E}\left(Z_1^{(1)} Z_2^{(1)\top}\right)\right\|_{\text{op}} > \epsilon\right) \leq \exp\left(-\frac{n\epsilon^2}{2a_1^2 a_2^2}\right),$$

and the layer cake representation of the expected value then gives that

$$\mathbb{E}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} Z_1^{(i)} Z_2^{(i)\top} - \mathbb{E}\left(Z_1^{(1)} Z_2^{(1)\top}\right)\right\|_{\text{op}} > \epsilon\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

Applying a similar argument with respect to the L^2-norm to the other summands entails the same rate for them. Hence, $\mathbb{E}(\|\widehat{\boldsymbol{C}}_{1,2} - \boldsymbol{C}_{1,2}\|_{\text{op}}) = O(1/\sqrt{n})$. Thus, it is possible to not assume compactness of the cross-covariance operators $\boldsymbol{C}_{i,j}$ if the sample, respectively, the embeddings of the sample are bounded. Non-compact cross-covariance operators can actually arise from very simple dependence structures, e.g. $Z_1^{(i)} = Z_2^{(i)}, i \in \{1, \ldots, N\}$. The latter is a very trivial example in the context of CCA, but useful to point out that there exist more complex dependence structures that cannot be described by compact cross-covariance operators. This was observed in Fukumizu et al. (2007), who provided a sufficient condition for a cross-covariance operator to be Hilbert-Schmidt. However, this condition is not satisfied in general.

Example 2 Consider a bounded sequence $[Z_1^{(1)}, Z_2^{(1)}], \ldots, [Z_1^{(n)}, Z_2^{(n)}]$ of identically distributed L^4-m-approximable pairs of random variables, whose cross-covariance operator is Hilbert Schmidt. Then, Assumption 4 holds, as shown in Theorem 3 of Kokoszka and Reimherr (2013). The reasoning in the latter is for ordinary covariance operators in Hilbert spaces, but carries over to cross-covariance operators. Specific examples include linear and nonlinear moving averages, autoregressive and stochastic volatility models (see e.g. Aue et al. 2009; Hörmann and Kokoszka 2010; Kokoszka et al. 2025).

Based on the above assumptions, we can now state the main result of this section, which provides consistency rates for the outputs of multiple kernel CCA. In particular, it entails a joint consistency rate for the optimal transformations $[\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_L^*]$ that are required in (2) to create plots as in Fig. 1. The operator $\mathfrak{C}_L$ belongs to the population version of the CCA-related eigenvalue problem (10), see also (4) for its detailed representation.

Theorem 1 *Let Assumptions 1–4 hold. Furthermore, we assume that $\mathfrak{C}_L$ is closed and bounded. Let $\widehat{\rho}_n$ be the estimator, arising from (11) for an isolated eigenvalue ρ of $\mathfrak{C}_L$ with finite multiplicity and $\widehat{\mathfrak{f}}$ the corresponding normalized eigenvector estimator for $\mathfrak{f}$. Then,*

$$|\widehat{\rho}_n - \rho| = O_P(\max\{\epsilon_n, \epsilon_n^{-\frac{3}{2}} \cdot n^{-\frac{1}{2}}\}),$$

$$\|\widehat{\mathfrak{f}}_n - \mathfrak{f}\| = O_P(\max\{\epsilon_n, \epsilon_n^{-\frac{3}{2}} \cdot n^{-\frac{1}{2}}\}).$$

Proof If a sequence of bounded linear operators $(T_n)_{n\in\mathbb{N}}$ converges in the operator norm $\|\cdot\|_{\text{op}}$ to some bounded limit element T, i.e. $\|T_n - T\|_{\text{op}} \to 0$, this implies strong stability on the resolvent of T (Chatelin, 1981, Proposition 2.11). As the eigenvalue ρ of our considered covariance operators is isolated by assumption, Proposition 4.1 from Chatelin (1981) entails that there exists a projection operator $\mathcal{P}$ such that

$$|\widehat{\rho}_n - \rho| = O_P(\|(\widehat{\mathfrak{C}}_{n,L} - \mathfrak{C}_L)\mathcal{P}\|_{\text{op}}), \quad \|\widehat{\mathfrak{f}}_n - \mathfrak{f}\| = O_P(\|(\widehat{\mathfrak{C}}_{n,L} - \mathfrak{C}_L)\mathcal{P}\|_{\text{op}}).$$

Since

$$\begin{aligned}\|(\widehat{\mathfrak{C}}_{n,L} - \mathfrak{C}_L)\mathcal{P}\|_{\text{op}} &\leq \|\widehat{\mathfrak{C}}_{n,L} - \mathfrak{C}_L\|_{\text{op}} \\ &\leq \sum_{i,j=1, i\neq j}^{L} \|(\widehat{\boldsymbol{C}}_{i,i} + \epsilon_n \cdot \boldsymbol{I})^{-1/2}\, \widehat{\boldsymbol{C}}_{i,j}\, (\widehat{\boldsymbol{C}}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-1/2} \\ &\quad - \boldsymbol{C}_{i,i}^{-1/2}\, \boldsymbol{C}_{i,j}\, \boldsymbol{C}_{j,j}^{-1/2}\|_{\text{op}},\end{aligned}$$

it suffices to derive the consistency rate for the right hand side in order to obtain consistency rates for the eigenvector and eigenvalue estimators. Under Assumption 4, Lemma 6 of Fukumizu et al. (2007) also holds for non-iid data $[X_1^{(1)}, \ldots, X_L^{(1)}], \ldots, [X_1^{(n)}, \ldots, X_L^{(n)}]$. Thus, whe have for $i \neq j$:

$$\begin{aligned}\|(\widehat{\boldsymbol{C}}_{i,i} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}}\, \widehat{\boldsymbol{C}}_{i,j} (\widehat{\boldsymbol{C}}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}} - (\boldsymbol{C}_{i,i} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}}\, \boldsymbol{C}_{i,j} (\boldsymbol{C}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}}\|_{\text{op}} \\ = O_P\left(\epsilon_n^{-\frac{3}{2}} \cdot n^{-\frac{1}{2}}\right). \quad (16)\end{aligned}$$

Moreover, the upper bound

$$\begin{aligned}&\|(\boldsymbol{C}_{i,i} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}}\, \boldsymbol{C}_{i,j} (\boldsymbol{C}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}} - \boldsymbol{C}_{i,i}^{-\frac{1}{2}}\, \boldsymbol{C}_{i,j} \boldsymbol{C}_{j,j}^{-\frac{1}{2}}\|_{\text{op}} \quad (17) \\ &\leq \|((\boldsymbol{C}_{i,i} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}} - \boldsymbol{C}_{i,i}^{-\frac{1}{2}})\, \boldsymbol{C}_{i,j}\, (\boldsymbol{C}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}}\|_{\text{op}} \\ &\quad + \|\boldsymbol{C}_{i,i}^{-\frac{1}{2}}\, \boldsymbol{C}_{i,j}\, ((\boldsymbol{C}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}} - \boldsymbol{C}_{j,j}^{-\frac{1}{2}})\|_{\text{op}} \\ &= O(\max\{\|(\boldsymbol{C}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}} - \boldsymbol{C}_{i,i}^{-\frac{1}{2}}\|_{\text{op}}, \|(\boldsymbol{C}_{j,j} + \epsilon_n \cdot \boldsymbol{I})^{-\frac{1}{2}} - \boldsymbol{C}_{j,j}^{-\frac{1}{2}}\|_{\text{op}}\}).\end{aligned}$$

holds, because the covariance operators are bounded and invertible, which implies boundedness of their inverse square root. By setting $\boldsymbol{A} = \boldsymbol{C}_{i,i} + \epsilon_n \cdot \boldsymbol{I}$ and $\boldsymbol{B} = \boldsymbol{C}_{i,i}$ in

$$\|\boldsymbol{A}^{-\frac{1}{2}} - \boldsymbol{B}^{-\frac{1}{2}}\|_{\text{op}} = \|\boldsymbol{A}^{-\frac{1}{2}}\, (\boldsymbol{B}^{\frac{3}{2}} - \boldsymbol{A}^{\frac{3}{2}})\, \boldsymbol{B}^{-\frac{3}{2}} + (\boldsymbol{A} - \boldsymbol{B})\, \boldsymbol{B}^{-\frac{3}{2}}\|_{\text{op}},$$

and combining it with the fact from (Fukumizu et al. 2007, Lemma 8) that there exists a constant $\lambda > 0$ for the choice of $\boldsymbol{A}$ and $\boldsymbol{B}$ such that

$$\|\boldsymbol{A}^{\frac{3}{2}} - \boldsymbol{B}^{\frac{3}{2}}\|_{\text{op}} \leq 3\lambda^{\frac{1}{2}} \|\boldsymbol{A} - \boldsymbol{B}\|_{\text{op}},$$

we obtain that the term on the right hand side of (17) is $O(\epsilon_n)$. This concludes the proof. □

Remark 4 While Fukumizu et al. (2007) proves consistency of the estimators $\widehat{\rho}_n$ and $\widehat{\mathfrak{f}}_n$, but do not provide rates for them, Zhou and Chen (2020) gives rates that directly rely on the decay of the eigenvalues of the underlying covariance operators, which is usually unknown. Their assumption on the decay requires compactness of the underlying covariance operators, which is not assumed in this paper. The results in Hsing and Eubank (2015) do not cover regularized estimators, which are required in the sample case.

Remark 5 The number of features L can also be allowed to slowly grow with the sample size n. In this case, an additional factor L^2 appears in the consistency rates. Nevertheless, this only leads to a convergence statement of the form $\|\widehat{\mathfrak{C}}_{n,L} - \mathfrak{C}_L\| \xrightarrow{P} 0$ in which the operator norm also depends on L. Without defining a proper limit element of $\mathfrak{C}_L$, the interpretation of such a mathematical result is unclear.

3 Multiple functional CCA

Instead of embedding the matrix-valued observations $\boldsymbol{A}_l[1], \ldots, \boldsymbol{A}_l[n], l \in \{1, \ldots, L\}$, from our motivating example in Sect. 1 into an RKHS, we consider a different approach in this section. We transform each data block $\boldsymbol{A}_l[i]$ into a function $\boldsymbol{Y}_l[i]$, i.e. we are smoothing every block in order to lift it into a function space, before performing CCA with these transformed data.

Many smoothing procedures are well-known (see e.g. Górecki et al. 2018; Ramsay and Silverman 2005). We assume our smoothed observations to be of the form $\boldsymbol{Y}_l[i] = (Y_{l,1}^{(i)}, \ldots, Y_{l,p_l}^{(i)})^\top$, where each $Y_{l,j}^{(i)}$ is a function in $L^2(I)$, and I is an interval. For the sake of lucidity, we only use the notation $\boldsymbol{Y}_l[i](t) = (Y_{l,1}^{(i)}(t), \ldots, Y_{l,p_l}^{(i)}(t))^\top$, when pointwise evaluations are required.

The population model of a smoothed block is

$$Y_{l,j}(t) = \sum_{b=1}^{B_{l,j}} c_{l,j,b} \cdot \varphi_{l,j,b}(t), \quad l \in \{1, 2, \ldots, L\}, \ j \in \{1, 2, \ldots, p_l\}, \tag{18}$$

where the $\varphi_{l,j,1}(t), \ldots, \varphi_{l,j,B_{l,j}}(t)$, $j \in \{1, \ldots, p_l\}$, are known orthonormal basis functions and $c_{l,j,1}, \ldots, c_{l,j,B_{l,j}}$ are unknown random coefficients. We can rewrite (18) as

$$\boldsymbol{Y}_l(t) = \boldsymbol{\Phi}_l(t) \cdot \mathfrak{c}_l, \tag{19}$$

where

$$\boldsymbol{\Phi}_l(t) = \begin{pmatrix} \boldsymbol{\varphi}_{l1}^\top(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_{l2}^\top(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}_{lp_l}^\top(t) \end{pmatrix},$$

$\boldsymbol{\varphi}_{lj}(t) = (\varphi_{l,j,1}(t), \dots, \varphi_{l,j,B_{l,j}}(t))^\top$, $\mathbf{c}_l = (c_{l,1,1}, \dots, c_{l,1,B_{l,1}}, \dots, c_{l,p_l,1}, \dots, c_{l,p_l,B_{l,p_l}})^\top$, $t \in I$, $j \in \{1, \dots, p_l\}$, $l \in \{1, 2, \dots, L\}$.

For the sample case, we replace $\mathbf{c}_l$ in (19) by the least-squares estimators $\widehat{\mathbf{c}}_l[i]$ such that each block $\boldsymbol{A}_l[i]$ is encoded by an estimate $\widehat{\mathbf{c}}_l[i]$.

In the population case, functional canonical variables are defined as

$$U_l = \langle \boldsymbol{u}_l, \boldsymbol{Y}_l \rangle = \int_I \boldsymbol{u}_l^\top(t) \boldsymbol{Y}_l(t) dt, \quad l \in \{1, \dots, L\},$$

where $\boldsymbol{u}_l$ is the vector weight function that lives in the same function space as $\boldsymbol{Y}_l$, i.e. it has a representation

$$\boldsymbol{u}_l(t) = \boldsymbol{\Phi}_l(t) \cdot \boldsymbol{w}_l. \tag{20}$$

In the competitiveness index example, l represents a feature, e.g., the institutions, so U_l is a weighted average of the indicators in that feature, averaged over time. The value of U_l depends on the country.

One of the objectives of our analysis is to obtain the mappings

$$U^{(l)}(\cdot) = \langle \boldsymbol{u}_l^*, \cdot \rangle, \tag{21}$$

and to evaluate them for each datum in the sample, and attempt to cluster subjects by their values. This is typically done by drawing a scatterplot on the $(U^{(i)}, U^{(j)})$ plane. Subjects with similar similar $(U^{(i)}, U^{(j)})$ values are considered as sharing important features.

As $\boldsymbol{Y}_l$ and $\boldsymbol{u}_l$ have representations as in (19) and (20), we can write

$$U_l = \langle \boldsymbol{u}_l, \boldsymbol{Y}_l \rangle = \boldsymbol{w}_l^\top \mathbf{c}_l, \quad l \in \{1, \dots, L\}. \tag{22}$$

Thus, the form of the functional canonical variable corresponding to the random process $\boldsymbol{Y}_l$ is determined by the vectors $\mathbf{c}_l$ and $\boldsymbol{w}_l$.

In the population case, multiple functional canonical correlation analysis can be presented as the following optimization problem:

$$\rho = \max_{\boldsymbol{u}_1, \dots, \boldsymbol{u}_L} \sum_{i=1}^{L} \sum_{j \in \{1, \dots, L\} \setminus \{i\}} \mathrm{Cov}(U_i, U_j) \text{ subject to } \sum_{i=1}^{L} \mathrm{Var}(U_i) = L, \tag{23}$$

where

$$(\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_L^*) = \arg \max_{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_L} \sum_{i=1}^{L} \sum_{j\in\{1,\ldots,L\}\setminus\{i\}} \operatorname{Cov}(U_i, U_j) \text{ subject to } \sum_{i=1}^{L} \operatorname{Var}(U_i) = L,$$

By the definition of the U_i in (22), it is easy to see that this problem is equivalent to

$$\rho = \max_{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_L} \sum_{i=1}^{L} \sum_{j\in\{1,\ldots,L\}\setminus\{i\}} \operatorname{Cov}(\langle \boldsymbol{u}_i, \boldsymbol{Y}_i\rangle, \langle \boldsymbol{u}_j, \boldsymbol{Y}_j\rangle) \text{ subject to } \sum_{i=1}^{L} \operatorname{Var}(\langle \boldsymbol{u}_i, \boldsymbol{Y}_i\rangle) = L,$$

and thus can be written as in (3) by using covariance operators the $\boldsymbol{C}_{i,j}$, $i, j \in \{1, \ldots, L\}$. We call the coefficient ρ the canonical functional correlation coefficient, but we emphasize that our definition is valid for $L > 2$. Previous work (e.g. Leurgans et al. 1993), studied only the case of $L = 2$. In the competitiveness example, $L > 2$ allows us to study more than two features of competitiveness.

Similar to the multiple kernel CCA in Sect. 2, in the sample case, we replace $\operatorname{Cov}(\langle \boldsymbol{u}_i, \boldsymbol{Y}_i\rangle, \langle \boldsymbol{u}_j, \boldsymbol{Y}_j\rangle)$ by an estimator $\langle \boldsymbol{u}_i, \widehat{\boldsymbol{C}}_{i,j}\, \boldsymbol{u}_j\rangle$, and $\operatorname{Var}(\langle \boldsymbol{u}_i, \boldsymbol{Y}_i\rangle)$ by a regularized estimator $\langle \boldsymbol{u}_i, (\widehat{\boldsymbol{C}}_{i,i} + \epsilon_n \cdot \boldsymbol{I})\, \boldsymbol{u}_i\rangle$. Using the standard sample covariance estimator $\widehat{\boldsymbol{C}}_{i,j}$ and (22), it is easy to see that

$$\langle \boldsymbol{u}_i, \widehat{\boldsymbol{C}}_{i,j}\, \boldsymbol{u}_j\rangle = \langle \boldsymbol{w}_i, \widehat{\boldsymbol{C}}_{i,j}\, \boldsymbol{w}_j\rangle,$$

with

$$\widehat{\boldsymbol{C}}_{i,j} = \frac{1}{n-1} \sum_{k=1}^{n} \widehat{\mathfrak{c}}_i[k] \widehat{\mathfrak{c}}_j[k]^\top - \frac{n}{n-1} \left(\frac{1}{n} \sum_{l=1}^{n} \widehat{\mathfrak{c}}_i[k] \right) \cdot \left(\frac{1}{n} \sum_{\ell=1}^{n} \widehat{\mathfrak{c}}_j[\ell] \right)^\top.$$

Similarly, for the variance, it holds

$$\langle \boldsymbol{u}_i, (\widehat{\boldsymbol{C}}_{i,i} + \epsilon_n \cdot \boldsymbol{I})\, \boldsymbol{u}_i\rangle = \left\langle \boldsymbol{w}_i, \left(\widehat{\boldsymbol{C}}_{i,i} + \epsilon_n \cdot \boldsymbol{I}\right)\, \boldsymbol{w}_i \right\rangle.$$

Thus, the sample version of the multiple CCA problem for functional data is

$$\widehat{\rho} = \max_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_L} \sum_{i=1}^{L} \sum_{j\in\{1,\ldots,L\}\setminus\{i\}} \boldsymbol{w}_i^\top \widehat{\boldsymbol{C}}_{i,j} \boldsymbol{w}_j \quad \text{subject to} \quad \sum_{i=1}^{L} \boldsymbol{w}_i^\top (\widehat{\boldsymbol{C}}_{i,i} + \epsilon \cdot \boldsymbol{I}) \boldsymbol{w}_i = L,$$

and the arg max of this optimization problem gives estimators for (21). As in Sect. 2, the above problem can be reduced to the following generalized eigenvalue problem whose solution provides the weights $\boldsymbol{w}_l$:

$$\begin{pmatrix} \boldsymbol{0} & \widehat{\boldsymbol{C}}_{1,2} & \widehat{\boldsymbol{C}}_{1,3} & \ldots & \widehat{\boldsymbol{C}}_{1,L} \\ \widehat{\boldsymbol{C}}_{2,1} & \boldsymbol{0} & \widehat{\boldsymbol{C}}_{2,3} & \ldots & \widehat{\boldsymbol{C}}_{2,L} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \widehat{\boldsymbol{C}}_{L,1} & \widehat{\boldsymbol{C}}_{L,2} & \widehat{\boldsymbol{C}}_{L,3} & \ldots & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \ldots \\ \boldsymbol{w}_L \end{pmatrix}$$

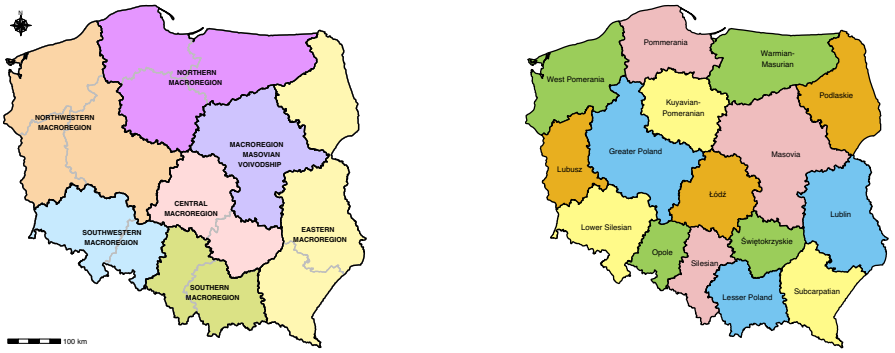

Fig. 2 Macroregions (left) and voivodeships (right) in Poland (2018)

$$= \rho \begin{pmatrix} \widehat{\mathcal{C}}_{1,1} + \varepsilon \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \widehat{\mathcal{C}}_{2,2} + \varepsilon \boldsymbol{I} & \mathbf{0} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \widehat{\mathcal{C}}_{L,L} + \varepsilon \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \dots \\ \boldsymbol{w}_L \end{pmatrix}.$$

Theorem 2 *Let the random variables $Y_l[k]$ take values in a separable Hilbert space, have finite second moments, and Assumptions 2–4 hold. Moreover, the conditions on the eigenvalues and operators imposed in Theorem 1 are satisfied. Then, the results of Theorem 1 directly carry over to multiple functional CCA described in this section.*

4 Illustrative examples

4.1 Agriculture data set for Polish voivodeships

We use agricultural data for Polish voivodeships to demonstrate practical aspects of the described methodology.[1] Figure 2 presents the administrative division of Poland into macroregions and voivodeships. Macroregions are broadly equivalent to NUTS 1 units.[2] In comparison, voivodeships are equivalent to NUTS 2 units.

The data consists of yields of thirty crops, expressed in quintals per hectare. These crops were recorded in 2003–2016 ($T = 14$ years) in $n = 16$ Polish voivodeships (administrative units). The analyzed plants are divided into $L = 3$ groups:

- Group 1—cereals and root crops ($p_1 = 9$ variables): barley, buckwheat, millet, oat, potatoes, rye, sugar beet, triticale, and wheat.
- Group 2—fodders ($p_2 = 6$ variables): clover, field crops, legume fodder, lucerne, root fodder, and serradella.
- Group 3—fruits and vegetables ($p_3 = 15$ variables): apples, cabbage, carrot, cauliflower, cherries, cucumbers, currants, gooseberry, onion, pears, plums, raspberries, sweet cherries, strawberries, and tomatoes.

[1] http://stat.gov.pl.

[2] The NUTS classification is a territorial standard valid for the statistical division of the European Union member countries.

Table 1 Top-3 biggest multiple canonical correlations for the Polish voivodeships data set

No.	Multivariate repeated measures data ($\hat{\rho}$)	Multivariate functional data ($\hat{\rho}_F$)
1	0.45	0.58
2	0.29	0.37
3	0.25	0.34

4.1.1 Multivariate repeated measures data

The input data are, therefore, the multivariate repeated measures data $[\boldsymbol{A}_1[1], \boldsymbol{A}_2[1], \boldsymbol{A}_3[1]], \ldots, [\boldsymbol{A}_1[16], \boldsymbol{A}_2[16], \boldsymbol{A}_3[16]]$, with $\boldsymbol{A}_1[k] \in \mathbb{R}^{14\times 9}$, $\boldsymbol{A}_2[k] \in \mathbb{R}^{14\times 6}$, $\boldsymbol{A}_3[k] \in \mathbb{R}^{14\times 15}$.

Based on these data, multiple kernel canonical correlation analysis was performed (Table 1). The multivariate repeated measures data for 16 voivodeships in the system of the first two multiple kernel canonical variables $(U^{(1)}, U^{(2)})$ are shown in Fig. 3.

It is important to note that the reported values are not classical correlation coefficients, but generalized canonical correlations defined by the optimization problems (8), and (23). For $L = 2$, they coincide with the usual canonical correlations and are bounded by one. For $L > 2$, however, they quantify the overall strength of association across several sets and are not restricted to the unit interval. Therefore, these quantities should be interpreted comparatively rather than on an absolute scale. Larger values indicate stronger overall dependence among the feature sets. In practice, the relative magnitude and decay of successive canonical correlations are typically used to assess how many components capture the dominant joint structure in the data, similarly to the interpretation of eigenvalues in principal component analysis. The values in Tables 1 and 3 should therefore be interpreted as generalized correlation coefficients, which are always non–negative but may exceed one without indicating a normalization issue.

Figure 3 shows that the voivodeships belonging to the same macroregion are located close to one another on the $(U^{(1)}, U^{(2)})$ plot. This is reasonable because the voivodeships belonging to a given macroregion have similar temperature, rainfall, and sun exposure, influencing crop yields. This shows that our algorithm produces beneficial results when some prior knowledge is available. It gives us confidence that its output will be helpful in situations where no prior information can be obtained. The only outlier voivodeship is Opole. This is unsurprising because this voivodeship is considered the best in Poland in terms of agricultural production. Its climate is characterized by hot summers, mild and short winters, early springs, and long, mild autumns. Moreover, 62% of its area consists of rich brown and clay soils and productive soils in numerous lowland river valleys. These conditions are ideal for cultivating cereals such as wheat, barley, rapeseed, and sugar beets. The agricultural valorization index[3] (WWRPP) places this region as the best in Poland. The WWRPP for Opole is 81.6 points, compared to the national average of 66.6 points (Stuczyński et al. 2007), see Fig. 4. The fact that our algorithm identified this outlier further validates its usefulness.

[3] WWRPP reflects the potential of agricultural production space resulting from natural conditions. It is an integrated indicator that assesses individual habitat elements such as soil quality and suitability, soil water relations, terrain relief, and agroclimate.

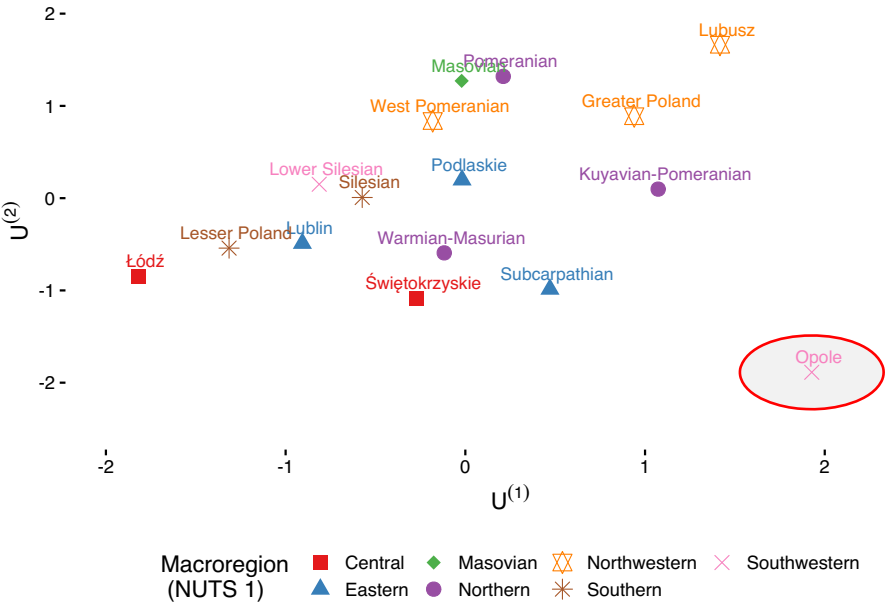

Fig. 3 Scatterplot for the optimally transformed feature pairs in the agriculture data set (16 voivodeships and seven regions) in the system of the first two multiple kernel canonical variables $(U^{(1)}, U^{(2)})$. The optimal transformations were determined by multiple kernel CCA for repeated measures data, as described in Sect. 2

4.1.2 Multivariate functional data

We used a Fourier basis with nine components ($B_{ij} = 9,\ i \in \{1, \ldots, L\},\ j \in \{1, \ldots, p_i\}$) to express the agriculture data as functional data. Using the transformed data, multiple functional canonical correlation analysis was performed (Table 1). The multivariate functional data for the 16 voivodeships in the system of the first two multiple functional canonical variables $(U^{(1)}, U^{(2)})$ are shown in Fig. 5, which shows that in the system of multiple canonical variables, the voivodeships are grouped into compact clusters belonging to specific macroregions. Additionally, from the canonical correlations (Table 1), we observe higher values for the multivariate functional approach. Thus, for these data, multiple functional canonical variables have stronger discrimination ability than the multiple kernel canonical variables.

4.2 Global competitiveness index (GCI) data set

In the second example, we study the relationships involving $n = 115$ countries over $T = 10$ years (2008–2017), based on $L = 12$ features. For this purpose, data published by the World Economic Forum (WEF) in its annual reports[4] is used. Established in 1979, the Global Competitiveness Report by the WEF stands as the most enduring and thorough evaluation of the factors influencing economic development. These are

[4] http://www.weforum.org.

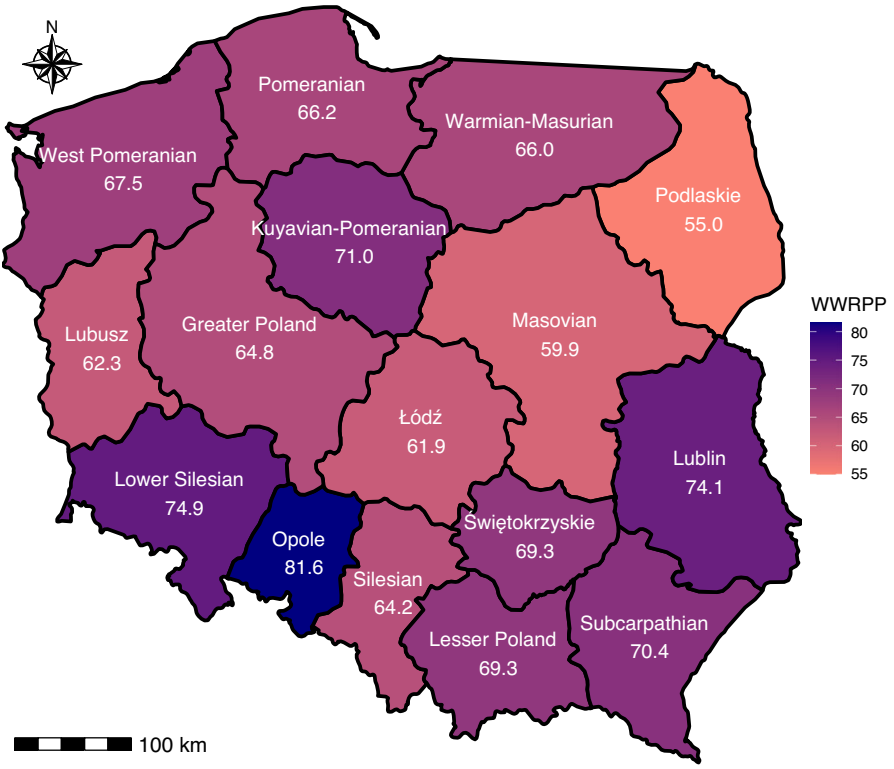

Fig. 4 WWRPP index for Poland (country mean = 66.6 points)

comprehensive data, exhaustively describing various socio-economic conditions of countries for which relevant data are available. Table 2 describes the features and many scalar indicators (a detailed description can be found in Górecki et al. 2016) in each feature employed in the analysis. WEF experts have divided the countries (115) into five groups (Fig. 6). These groups are not used in the analysis, but they illustrate the meaning of the Global Competitiveness Index and are used at the end of this section to validate the clustering implied by our analysis.

We performed the multivariate and the functional analyses, similarly as in Sect. 4.1. This time, we have $L = 12$ groups, $n = 115$ countries, and $T = 10$ years. In the case of the MFCCA, observations are converted to functions by utilizing the Fourier basis with five basis functions ($B_{lj} = 5,\ l \in \{1, \ldots, L\},\ j \in \{1, \ldots, p_l\}$). Table 3 displays the results of the two analyses. It shows that, as in the previous example, the correlations are slightly higher for the functional approach. This is likely due to the fact that crop yields and competitiveness data generally evolve smoothly from year to year, and this can be better captured by smooth functional basis expansions than by kernel models. The corresponding projections on the plane can be found in Figs. 1 and 7.

How is the representation quality assessed, and how is the whole analysis? For this purpose, we decided to use the information that the data is divided into 5 groups.

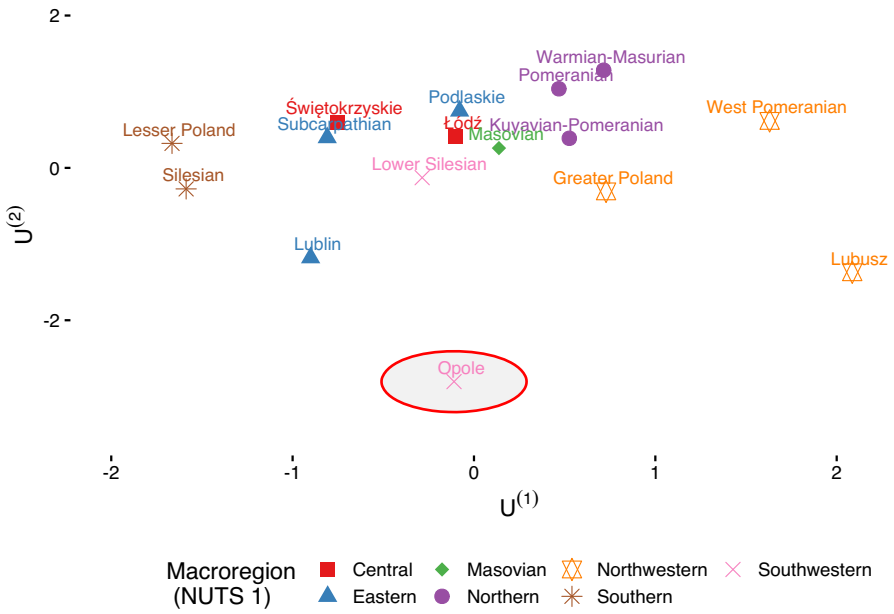

Fig. 5 Scatterplot for the optimally transformed feature pairs in the agriculture data set (16 voivodeships and seven regions) in the system of the first two multiple functional canonical variables ($U^{(1)}$, $U^{(2)}$). The optimal transformations were determined by multiple functional CCA, as described in Sect. 3

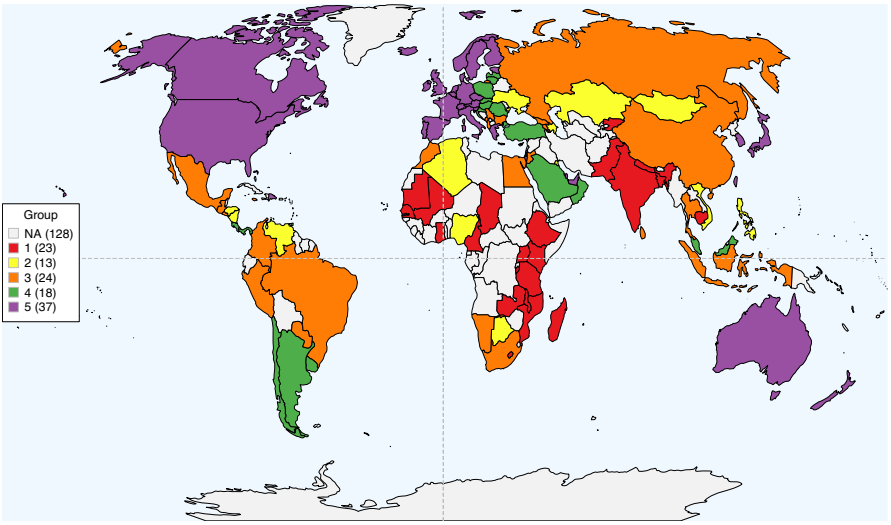

Fig. 6 The 115 countries used in our analysis are highlighted by colors. They are divided into five groups based on the GCI value. The count of countries in each group is given in parentheses (NA stands for missing data)

Table 2 The 12 features used in the analysis of the GCI data set

No. (l)	Feature	Number of variables, i.e. scalar indicators (p_l)
1	Institutions	16
2	Infrastructure	6
3	Macroeconomic environment	3
4	Health and primary education	7
5	Higher education and training	6
6	Goods market efficiency	10
7	Labour market efficiency	6
8	Financial market development	5
9	Technological readiness	4
10	Market size	4
11	Business sophistication	9
12	Innovation	5

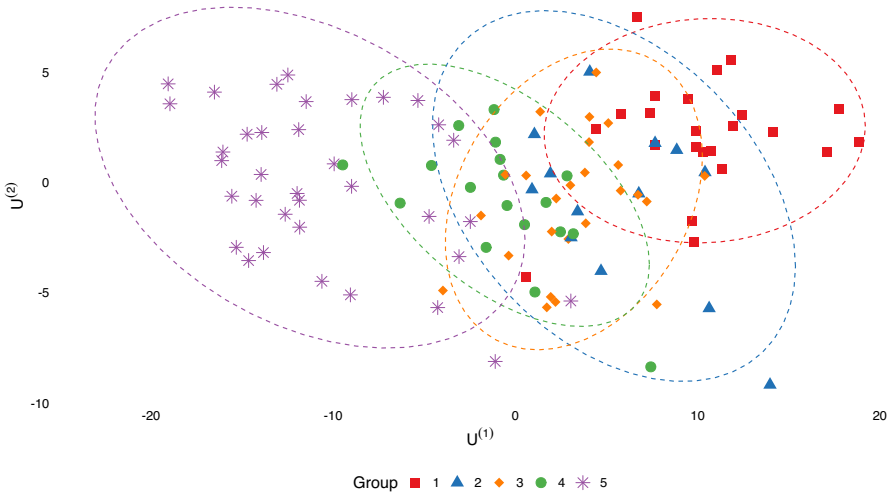

Fig. 7 Scatterplot for the optimally transformed feature pairs in the GCI data set (115 countries and five groups) in the system of the first two multiple functional canonical variables $(U^{(1)}, U^{(2)})$ (with 95% confidence normal ellipses). The optimal transformations were determined by multiple functional CCA, as described in Sect. 3

Table 3 Top-3 biggest multiple canonical correlations for GCI data set

No.	Multivariate repeated measures data ($\hat{\rho}$)	Multivariate functional data ($\hat{\rho}_F$)
1	0.74	0.76
2	0.28	0.29
3	0.11	0.11

To evaluate clusterability, we used the Hopkins statistic (Hopkins and Skellam 1954; Lawson and Jurs 1990). This statistic is employed to assess the clustering tendency of a data set. Let X be the data set of n points in the d-dimensional space. We want to test the pair of hypotheses:

Denote by C_X the smallest convex hull that contains X. The Hopkins statistic is calculated with the following algorithm:

(1) Sample randomly one observation from the data set X and set the counter to $i = 1$. Denote by w_i the Euclidean distance from this observation to the nearest-neighbor observation in X.
(2) Generate one new point uniformly distributed in C_X and denote by u_i the Euclidean distance from this point to the nearest-neighbor observation in X.
(3) Repeat steps (1) and (2) $m \ll n$ times (typically $m \approx 0.1 \cdot n$).
(4) Compute the Hopkins statistic

$$H = \frac{\sum_{i=1}^{m} u_i^d}{\sum_{i=1}^{m} (u_i^d + w_i^d)}.$$

Under the null hypothesis, this statistic follows the Beta(m, m) distribution. If the data has little structure, the average distance between real points will be similar to that from a uniformly distributed random point to a real point, resulting in a Hopkins statistic value of approximately 0.5. Conversely, if the data are tightly clustered, the distances w_i will be much smaller than those u_i, leading to a Hopkins statistic value close to 1.0. The interpretation of H can be understood through the following guidelines (Lawson and Jurs 1990; Wright 2022):

- Low values of H suggest that the observations in X are repelling each other.
- Values of H close to 0.5 indicate that the observations in X are spatially random.
- High values of H suggest possible clustering of the observations in X.
- Values of H greater than 0.75 indicate a clustering tendency at the 90% confidence level.

Results are presented in Fig. 8. Due to sampling variability, it is standard to calculate H multiple times and take the average. On the plot, we present for each $k \in \{1, \ldots, L\}$, with $L = 12$ (number of multiple canonical correlations), the average value of H for 100 replications. Multiple functional canonical variables are superior to multiple kernel canonical ones. This method exhibits better clusterability in just two components, indicating that the group structure has been more accurately mapped. For both techniques, the Hopkins statistic approaches one as the number of components increases. However, using only one component ($k = 1$) is not sufficient.

4.3 Calculations details

All computations were performed in `R` 4.5.1 (R Core Team 2025) using our own implementation of the proposed methods, provided in the `multiCCA` package (Górecki 2026).

The number of basis functions in the functional representation acts as a tuning parameter. In our analysis, several choices of the basis dimension were examined,

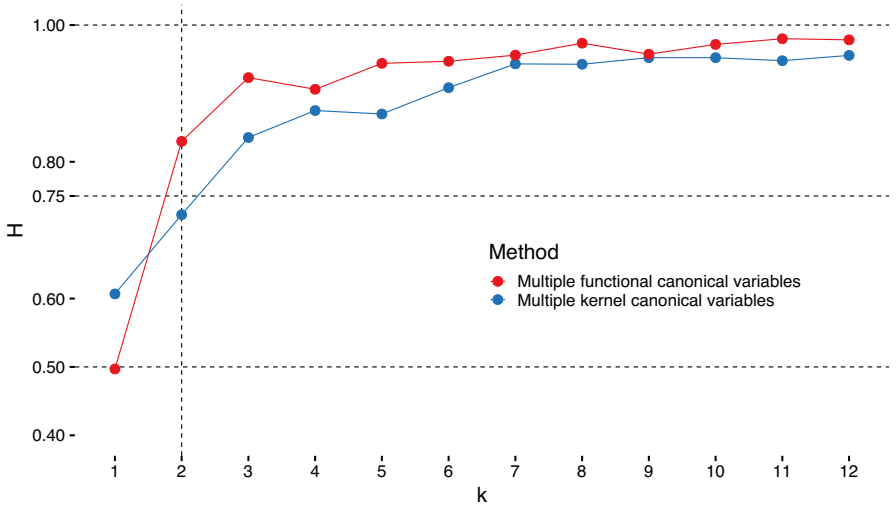

Fig. 8 The Hopkins statistic for the GCI data ($m = \lceil 115/10 \rceil = 12$) vs the number of components. The horizontal dashed lines indicate the standard thresholds: 0.5, 0.75, and 1. The vertical line at 2 indicates the projection onto the plane

and the main results remained stable over a reasonable range of values. The reported numbers of basis functions were selected from this range to provide a clear illustrative representation of the data.

From a computational perspective, the main cost of the proposed procedures comes from constructing covariance or kernel Gram matrices and solving the corresponding generalized eigenvalue problems. In the kernel approach, this involves matrices of size $n \times n$, so the computational complexity is mainly driven by the sample size n. In the functional approach, the dimensionality depends on the number of basis coefficients used to represent the functions, which is typically much smaller than the number of time points T. Consequently, the functional formulation often leads to smaller matrices and improved numerical efficiency. For very large datasets, scalability can be improved by using low-rank approximations to kernel matrices or randomized eigensolvers.

5 Concluding remarks

Hotelling's classical canonical correlation analysis has been generalized in two ways in this paper. First, random vectors are replaced by random matrices containing observations of multiple variables at T time points for the same experimental unit (in this case, we are dealing with multivariate repeated-measures data, also known as doubly multivariate data). Relationships between a finite number of random matrices, not limited to two, are considered. In this case, multiple kernel canonical variables are constructed. Secondly, we replace random vectors with multidimensional random processes. In this case, the experimental units are characterized by smooth functions over a time interval (elements of Hilbert space). The experimental units are characterized

by multivariate functional data. Moreover, as before, we consider the relationships between a finite number of multidimensional random processes, not limited to two. In this case, multiple functional canonical variables were constructed. For actual data on yields per hectare across three crop groups, multiple functional canonical variables proved superior. A similar result was obtained for the GCI data set.

The proposed methods apply when multiple feature sets are available and aim to examine their dependencies. Additionally, a third dimension is incorporated, such as time or space. The objectives of these techniques are twofold. On the one hand, we aim to present data on a plane to assess dependencies visually. On the other hand, obtaining vector representations of observations that account for time/space allows us to use them in other analyses where direct use may not be feasible. Canonical correlations for functional data can be applied to such data, but to date, they have been used only for two feature sets. Similarly, in such situations, an appropriate kernel canonical correlation technique can be employed, which we hope captures nonlinear dependencies more effectively. Choosing a specific method is challenging, and it is generally only possible with at least a preliminary analysis of the data set. If we have additional information about the groups to which observations belong, we can perform a clusterability assessment, as we do in our work.

We have justified our methodology under assumptions weaker than those used in previous research, even in simpler settings. Specifically, we showed that certain covariance operators need not be compact, and the observational units can be dependent. A high-level assumption is formulated that covers both cases.

The proposed operator-based framework for multiple CCA not only unifies existing approaches but also offers a principled way to handle dependence and non-compactness in high-dimensional or functional data. Future work may focus on extending these ideas toward regularized and sparse formulations, which could further improve stability and interpretability in complex empirical settings.

While the empirical section illustrates the proposed methodology on two real data sets, a broader numerical comparison with other multivariate functional methods, such as multivariate functional principal component analysis (MFPCA), would be of interest. These approaches address somewhat different goals: MFPCA focuses on explaining variance, whereas the proposed framework aims to capture dependence across multiple feature sets. A systematic simulation study and a more extensive empirical benchmarking will therefore be considered in future work.

Data availability The data that support the findings of this study are available at http://stat.gov.pl and http://www.weforum.org.

References

Adrover JG, Donato SM (2015) A robust predictive approach for canonical correlation analysis. J Multivar Anal 133:356–376

Akaho S (2001) A Kernel method for canonical correlation analysis. In: Proceedings of the international meeting of the psychometric society (IMPS2001)

Alam MA, Fukumizu K (2013) Higher-order regularized Kernel CCA. 12th international conference on machine learning and applications. pp 374–377

Alam MA, Fukumizu K (2015) Higher-order regularized kernel canonical correlation analysis. Int J Pattern Recognit Artif Intell 29(04):1551005

Alam MA, Nasser M, Fukumizu K (2010) A comparative study of Kernel and robust canonical correlation analysis. J Multimed 5(1):3–11

Aronszajn N (1950) Theory of reproducing Kernels. Trans Am Math Soc 68(3):337–404

Aue A, Hörmann S, Horváth L, Reimherr M (2009) Break detection in the covariance structure of multivariate time series models. Ann Stat 37(6B):4046–4087

Bach BR, Jordan MI (2002) Kernel independent component analysis. J Mach Learn Res 3:1–48

Bilenko NY, Gallant JL (2016) Pyrcca: regularized kernel canonical correlation analysis in python and its applications to neuroimaging. Front Neuroinform 10:49

Carroll JD (1968) Generalization of canonical correlation analysis to three or more sets of variables. In: Proceedings of the 76th annual convention of the American Psychological Association, vol 3. pp 227–228

Chatelin F (1981) The spectral approximation of linear operators with applications to the computation of eigenelements of differential and integral operators. SIAM Rev 23(4):495–522

Fukumizu K, Bach FR, Gretton A (2007) Statistical consistency of kernel canonical correlation analysis. J Mach Learn Res 8:361–383

Gifi A (1990) Nonlinear multivariate analysis. Wiley, New York

Girka F, Gloaguen A, Le Brusquet L, Zujovic V, Tenenhaus A (2024) Tensor generalized canonical correlation analysis. Inf Fusion 102:102045

Gloaguen A, Philippe C, Frouin V, Gennari G, Dehaene-Lambertz G, Le Brusquet L, Tenenhaus A (2020) Multiway generalized canonical correlation analysis. Biostatistics 23(1):240–256

Górecki, T. (2026). multiCCA: Multiple canonical correlation analysis (Kernel and Functional). R package version 0.1.0. https://CRAN.R-project.org/package=multiCCA

Górecki T, Krzyśko M, Ratajczak W, Wołyński W (2016) An extension of the classical distance correlation coefficient for multivariate functional data with applications. Stat Trans 17(3):449–466

Górecki T, Krzyśko M, Wołyński W (2017) Correlation analysis for multivariate functional data. In: Data science, studies in classification, data analysis, and knowledge organization. pp 243–258

Górecki T, Krzyśko M, Waszak Ł, Wołyński W (2018) Selected statistical methods of data analysis for multivariate functional data. Stat Pap 59(1):153–182

Gower JC (1989) Generalized canonical analysis. In: Coppi R, Bolasco S (eds) Multiway data analysis. North Holland, Amsterdam, pp 221–232

Happ C, Greven S (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. J Am Stat Assoc 113:649–659

Hardoon DR, Shawe-Taylor J (2009) Convergence analysis of kernel canonical correlation analysis: theory and practice. Mach Learn 74:23–38

Hardoon DR, Szedmak S, Shawe-Taylor J (2004) Canonical correlation analysis. An overview with application to learning methods. Neural Comput 16:2639–2664

Hopkins B, Skellam JG (1954) A new method for determining the type of distribution of plant individuals. Ann Bot 18(2):213–227

Hörmann S, Kokoszka P (2010) Weakly dependent functional data. Ann Stat 38(3):1845–1884

Horst P (1961) Relations among *m* sets of measures. Psychometrika 26:126–149

Horváth L, Kokoszka P (2012) Inference for functional data with application. Springer, New York

Hotelling H (1936) Relations between two sets of variates. Biometrika 28(3/4):321–377

Hsing T, Eubank RL (2015) Theoretical foundations of functional data analysis, with an intriduction to linear operators. Wiley, New York

Hwang H, Jung K, Takane Y, Woodward TS (2012) Functional multiple-set canonical correlation analysis. Psychometrika 77(1):48–64

Hwang H, Jung K, Takane Y, Woodward TS (2013) A unified approach to multiple-set canonical correlation analysis and principal components analysis. Br J Math Stat Psychol 66:308–321
Kettenring JR (1971) Canonical analysis of several sets of variables. Biometrika 58(3):433–451
Kokoszka P, Reimherr M (2013) Asymptotic normality of the principal components of functional time series. Stochastic Process Appl 123:1546–1562
Kokoszka P, Mohammadi N, Wang H (2025) Dependence properties of stochastic volatility models. J Time Ser Anal 46(3):421–431
Krafty RT, Hall M (2013) Canonical correlation analysis between time series and static outcomes, with application to the spectral analysis of heart rate variability. Ann Appl Stat 7(1):570–587
Krzyśko M, Smaga Ł (2019) Robust estimation in canonical correlation analysis for multivariate functional data. Hacet J Math Stat 48(2):521–535
Lafosse R (1989) Ressemblance et différence entre deux tableaux totalement appariés. Statistique et analyse des données 14(2):1–24
Lai PL, Fyfe C (2000) Kernel and nonlinear canonical correlation analysis. Int J Neural Syst 10(5):365–377
Langworthy BW, Stephens RL, Gilmore JH (2021) Canonical correlation analysis for elliptical copulas. J Multivar Anal 183:104715
Lawson RG, Jurs PC (1990) New index for clustering tendency and its application to chemical problems. J Chem Inf Comput Sci 30(1):36–41
Leurgans SE, Moyeed RA, Silverman BW (1993) Canonical correlation analysis when the data are curves. J Roy Stat Soc B 55(3):725–740
Li C, Xiao L, Luo S (2021) Joint model for survival and multivariate sparse functional data with application to a study of Alzheimer's disease. Biometrics 78(2):435–447
Ma Z, Li X (2020) Subspace perspective on canonical correlation analysis. Dimension reduction and minimax rates Bernoulli 26(1):432–470
Nielsen AA (2002) Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. IEEE Trans Image Process 11(3):293–305
Park M, Huh MH (1996) Quantification plots for several sets of variables. J Korean Stat Soc 25(4):589–601
Pinelis I (1994) Optimum bounds for the distributions of martingales in banach spaces. Ann Probab 22(4):1679–1706
R Core Team (2025) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer, New York
Sartini J, Zeger S, Crainiceanu C (2025) Bayesian multivariate sparse functional PCA. Preprint at http://arxiv.org/abs/2509.03512
Schölkopf B, Smola AJ, Müller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10(5):1299–1319
Shu H, Wang X, Zhu H (2020) D-CCA: a decomposition-based canonical correlation analysis for high-dimensional datasets. J Am Stat Assoc 115:242–306
Sort L, Le Brusquet L, Tenenhaus A (2024) Functional generalized canonical correlation analysis for studying multiple longitudinal variables. Biometrics 80(4):ujae113
Stuczyński T, Kozyra J et al (2007) Przyrodnicze uwarunkowania produkcji rolniczej w Polsce. Stud I Raporty IUNG-PIB 7:77–115
Ten Berge JMF (1988) Generalized approaches to the MAXBET problem and to the MAXDIFF problem, with applications to canonical correlations. Psychometrika 53:487–494
Tenenhaus A, Tenenhaus M (2011) Regularized generalized canonical correlation analysis. Psychometrika 76(2):257–284
Tenenhaus A, Tenenhaus M (2014) Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. Eur J Oper Res 238(2):391–403
Tenenhaus A, Philippe C, Frouin V (2015) Kernel generalized canonical correlation analysis. Comput Stat Data Anal 90(C):114–131
Tenenhaus M, Tenenhaus A, Groenen PJF (2017) Regularized generalized canonical correlation analysis. A framework for sequential multiblock components methods. Psychometrika 82(3):737–777
Wright K (2022) Will the real Hopkins statistic please stand up? R J 14(3):282–292
Yanai H (1998) Generalized canonical correlation analysis with linear constraints. In: Hayashi C, Ohsumi N, Yajima K, Tanaka Y, Bock HH, Baba Y (eds) Data science, classification, and related methods. Springer, Tokyo, pp 539–546

Zhou Y, Chen D (2020) The optimal rate of canonical correlation analysis for stochastic processes. J Stat Plan Inference 207:276–287