Ann Inst Stat Math (2018) 70:919–950
https://doi.org/10.1007/s10463-017-0613-1

Testing equality between several populations covariance operators

Graciela Boente[1] · Daniela Rodriguez[2] · Mariela Sued[3]

Received: 14 December 2015 / Revised: 22 May 2017 / Published online: 20 August 2017

Abstract In many situations, when dealing with several populations, equality of the covariance operators is assumed. An important issue is to study whether this assumption holds before making other inferences. In this paper, we develop a test for comparing covariance operators of several functional data samples. The proposed test is based on the Hilbert–Schmidt norm of the difference between estimated covariance operators. In particular, when dealing with two populations, the test statistic is just the squared norm of the difference between the two covariance operators estimators. The asymptotic behaviour of the test statistic under both the null hypothesis and local alternatives is obtained. The computation of the quantiles of the null asymptotic distribution is not feasible in practice. To overcome this problem, a bootstrap procedure is considered. The performance of the test statistic for small sample sizes is illustrated through a Monte Carlo study and on a real data set.

Keywords Asymptotic distribution · Bootstrap calibration · Covariance operators · Functional data analysis · Local alternatives

✉ Mariela Sued
marielasued@gmail.com

Graciela Boente
gboente@dm.uba.ar

Daniela Rodriguez
drodrig@dm.uba.ar

[1] Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and IMAS, CONICET, Ciudad Universitaria, Pabellón 1, 1428 Buenos Aires, Argentina

[2] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and IMAS, CONICET, Ciudad Universitaria, Pabellón 2, 1428 Buenos Aires, Argentina

[3] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Ciudad Universitaria, Pabellón 2, 1428 Buenos Aires, Argentina

1 Introduction

In many applications, we study phenomena that are continuous in time or space and can be considered as smooth curves or functions. The data collected in such situations can be viewed as realizations of a stochastic process, often assumed to be in $L^2(\mathcal{I})$, with $\mathcal{I}$ a bounded interval, and are usually called functional data in the literature. Statistical procedures to deal with such functional data may be found, for instance, in Ramsay and Silverman (2005), Ferraty and Vieu (2006) and Ferraty and Romain (2010). We also refer to Horváth and Kokoszka (2012) and Hsing and Eubank (2015) where a description of different procedures for functional data and their properties is given, and to Cuevas (2014) and Goia and Vieu (2016) who present a summary of recent advances in infinite-dimensional statistics. For functional data, most of the literature on hypothesis testing deals with tests on the mean function including the functional linear model. However, inference on the covariance operators recently emerged due to its multiple applications. In what follows, we review several papers where the comparison among covariance operators in practical problems is an issue. To analyse a data set from food industry quality control, Ferraty et al. (2007) considered tests for comparing groups of curves based on their covariances. Benjamini and Yekutieli (2001) proposed two-sample bootstrap tests for specific aspects of the spectrum of functional data, such as the equality of a subset of eigenfunctions of the covariance operators, and presented an application of the method to the implied volatilities of European options on the German stock index. On the other hand, Panaretos et al. (2010) gave an example based on DNA minicircles in which mean functions comparisons detected no differences, whereas covariance structures presented differences between the two groups. As mentioned therein, the comparison of the mean functions was related to the shape of the minicircles, while that of covariance operators was associated to the flexibility or stiffness of the minicircles, which was in fact the scientific hypothesis. Pigoli et al. (2014) explained that, in linguistic problems, language comparison using their phonetic structure is related to the analysis of their covariance operators and developed a two-sample test using different distances between covariance operators. Their procedure is based on a permutation test. These results have been recently extended from the two-sample case to the several population setting in Cabassi et al. (2017), where the proposed procedure is applied to compare the evolution of locomotor behaviour in mice. Furthermore, as pointed out in Fremdt et al. (2013), when dealing with two functional samples, if some of the parameters between populations are different, estimating them using the pooled sample may lead to inappropriate conclusions. For instance, Donoghue et al. (2008) studied human movement data of the Achilles tendon on injured and healthy individuals using a functional principal analysis of the combined centred data. However, to allow for a clear interpretation of the results, this approach implicitly assumes equality between covariance operators or at least equality of the eigenfunctions with the eigenvalue order preserved across populations. This point was also discussed in Coffey et al. (2011), who provide a different analysis based on functional common principal components, revealing differences in the variation of movement patterns of injured versus control subjects which were not detected by the analysis of the combined data. The previous discussion motivates the need of defining procedures to test equality among covariance operators and providing a deep analysis on their asymptotic behaviour, in particular, to study their capability to detect local alternatives. Even if several procedures have been presented to deal with this problem, most of them only study the null asymptotic distribution to define critical values. Among others, we can mention Fremdt et al. (2013), who considered an approach based on the projection of the data over a suitable chosen finite-dimensional space, such as that defined by the functional principal components. These results generalized those provided in Panaretos et al. (2010), who assumed that the processes have a Gaussian distribution. Further details may be found in Horváth and Kokoszka (2012). On the other hand, Gaines et al. (2011) presented a different approach to test equality of two covariance operators, where an univariate likelihood ratio test is combined with Roy's union–intersection principle for testing the equality of two covariance operators, and derived its asymptotic behaviour under both the null hypothesis and a set of local alternatives converging to the null hypothesis with rate $n^{1/2}$, where n stands for the total sample size. Even if the permutation test defined in Pigoli et al. (2014) has been adapted to the situation of more than two populations in Cabassi et al. (2017), one drawback of such tests is that they rely strongly on the exchangeability assumption, under the null hypothesis. In particular, if the populations have the same underlying distribution except for changes in their means and covariance operators, this assumption corresponds to assuming known means across populations, for otherwise the procedure is not exact but asymptotic. Moreover, as far as we know, the asymptotic behaviour under local alternatives of the test statistic proposed in Pigoli et al. (2014) and generalized in Cabassi et al. (2017) has not been studied yet, nor their asymptotic null distribution when means are unknown.

This paper aims to not only propose a test statistic to compare covariance operators of k populations, but also provide a theoretical framework which clarifies the ability of the test statistic to detect local alternatives and their rate of convergence. Hence, our results extend the approaches based on distances between covariance operators estimators given in the case of two independent samples to the several samples situation and provide a full asymptotic analysis not only under the null but also under local alternatives converging at a root$-n$ rate, which include, for instance, the functional common principal components model. The asymptotic distribution of the test statistic proposed for comparing covariance operators among multiple populations depends on unknown quantities. For this reason, in order to implement the test we introduce a bootstrap calibration method whose validity is studied. It is worth noticing that even if the existing literature devoted to theoretical advances in bootstrapping for functional data is scarce, some bootstrap-based testing procedures have been previously considered in the literature. Contributions in functional nonparametric regression were given in Ferraty et al. (2010, 2012) and more recently in Raña et al. (2016), who provided bootstrap procedures to construct pointwise confidence intervals when the covariate is functional. As previously mentioned, bootstrap procedures for testing equality of means and equality of a fixed number of eigenfunctions for two populations have been studied in Benko et al. (2009). On the other hand, Chang and Ogden (2009) gave general results for sums of independent but not identically distributed processes which are applied to brain imaging. However, the results given therein cannot be applied to our setting. Finally, Paparoditis and Sapatinas (2016) presented a bootstrap-based scheme which allows for testing either equality of mean functions or equality of covariance operators for functional data, obtaining its null behaviour for the two population case. Our contribution provides a new path in the literature of bootstrapping for functional data, since it applies a bootstrap procedure to calibrate the critical values of the test statistic rather than constructing bootstrap samples as done recently in Paparoditis and Sapatinas (2016). This approach enhances previous proposals by allowing to study the null test performance but also its power for root$-n$ local alternatives when comparing the covariance operators of several populations.

The paper is organized as follows. Section 2 presents the notation and reviews some basic concepts. In Sect. 3, we introduce the test statistic and derive its asymptotic distribution under the null hypothesis. An important issue is to describe a set of local alternatives that the proposed statistic is able to detect. For that purpose, the asymptotic distribution under a set of local alternatives converging to the null hypothesis at rate $n^{1/2}$ is studied in Sect. 4. These alternatives include the functional common principal component model. A bootstrap calibration for the null distribution of the test statistic is described in Sect. 5. The results of a Monte Carlo study are summarized in Sect. 6, while Sect. 7 presents a real data example. Proofs are relegated to Appendix.

2 Preliminaries and notation

From now on, $\mathcal{H}$ stands for a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|u\| = \langle u, u\rangle^{1/2}$. Let $\mathbf{H} : \mathcal{H} \to \mathcal{H}$ be a compact operator. The operator $\mathbf{H}$ is said to be a trace class operator if $\sum_{\ell=1}^{\infty} \langle \mathbf{H} u_\ell, u_\ell\rangle < \infty$ for any orthonormal basis $\{u_\ell : \ell \geq 1\}$ of $\mathcal{H}$, while it is said to be Hilbert–Schmidt if $\sum_{\ell=1}^{\infty} \|\mathbf{H} u_\ell\|^2 < \infty$. The Hilbert space of Hilbert–Schmidt operators over $\mathcal{H}$ is denoted as $\mathcal{F}$, while $\mathbf{H}^*$ stands for the adjoint of the operator $\mathbf{H}$. Given $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{H}$ Hilbert–Schmidt operators, the inner product in $\mathcal{F}$ is defined as $\langle \mathbf{H}_1, \mathbf{H}_2\rangle_{\mathcal{F}} = \text{trace}(\mathbf{H}_1^* \mathbf{H}_2) = \sum_{\ell=1}^{\infty} \langle \mathbf{H}_1 u_\ell, \mathbf{H}_2 u_\ell\rangle$, while the norm equals $\|\mathbf{H}\|_{\mathcal{F}} = \langle \mathbf{H}^*, \mathbf{H}\rangle_{\mathcal{F}}^{1/2} = \{\sum_{\ell=1}^{\infty} \|\mathbf{H} u_\ell\|^2\}^{1/2}$, with $\{u_\ell : \ell \geq 1\}$ any orthonormal basis of $\mathcal{H}$. These definitions are independent of the basis choice. Besides, as is well known, Hilbert–Schmidt operators have a countable number of eigenvalues, all of them being real when the operator is self-adjoint. Hence, given a nonnegative and self-adjoint operator $\mathbf{H}$ and choosing the eigenfunctions of $\mathbf{H}$ as the orthonormal basis, we get that $\|\mathbf{H}\|_{\mathcal{F}}^2 = \sum_{\ell=1}^{\infty} \lambda_\ell^2$, where $\{\lambda_\ell : \ell \geq 1\}$ are the eigenvalues of $\mathbf{H}$ ordered so that $\lambda_\ell \geq \lambda_{\ell+1}$.

Let us consider independent random elements $X_1, \ldots, X_k$ in $\mathcal{H}$ and assume that $\mathbb{E}\|X_i\|^2 < \infty$. Denote by $\mu_i \in \mathcal{H}$ the mean of X_i, $\mu_i = \mathbb{E}(X_i)$ and by $\boldsymbol{\Gamma}_i : \mathcal{H} \to \mathcal{H}$ the covariance operator of X_i. Let $\otimes$ stand for the tensor product on $\mathcal{H}$, e.g., for $u, v \in \mathcal{H}$, the operator $u \otimes v : \mathcal{H} \to \mathcal{H}$ is defined as $(u \otimes v)w = \langle v, w\rangle u$. With this notation, the covariance operator $\boldsymbol{\Gamma}_i$ can be written as $\boldsymbol{\Gamma}_i = \mathbb{E}\{(X_i - \mu_i) \otimes (X_i - \mu_i)\}$. The operator $\boldsymbol{\Gamma}_i$ is a linear, self-adjoint and compact operator with finite trace, so it is a Hilbert–Schmidt operator. From now on, we denote as $\{\phi_{i,\ell} : \ell \geq 1\}$ the eigenfunctions of $\boldsymbol{\Gamma}_i$ related to the eigenvalues $\{\lambda_{i,\ell} : \ell \geq 1\}$, ordered as a non-increasing sequence, i.e., $\lambda_{i,\ell} \geq \lambda_{i,\ell+1}$. Recall that the trace of $\boldsymbol{\Gamma}_i$ is given by $\sum_{\ell=1}^{\infty} \lambda_{i,\ell}$.

When $\mathcal{H} = L^2(\mathcal{I})$ for some bounded interval $\mathcal{I}$ and $\langle u, v\rangle = \int_{\mathcal{I}} u(s)v(s)ds$, it is well known that the covariance operator is defined through the covariance function of

X_i, $\gamma_i(s,t) = Cov(X_i(s), X_i(t))$, $s, t \in \mathcal{I}$ as $(\boldsymbol{\Gamma}_i u)(t) = \int_{\mathcal{I}} \gamma_i(s,t) u(s) ds$. Besides, $\|\boldsymbol{\Gamma}_i\|_{\mathcal{F}}^2 = \sum_{\ell=1}^{\infty} \lambda_{i,\ell}^2 = \|\gamma_i\|^2 = \int_{\mathcal{I}} \int_{\mathcal{I}} \gamma_i^2(t,s) \mathrm{d}t\, \mathrm{d}s$.

Our goal is to test whether the covariance operators $\boldsymbol{\Gamma}_i$ of several populations are equal or not. For that purpose, let us consider independent samples of each population, i.e., let us assume that we have independent observations $X_{i,1}, \ldots, X_{i,n_i}$, $1 \le i \le k$ such that $X_{i,j} \sim X_i$, $1 \le j \le n_i$. A natural way to estimate the covariance operators $\boldsymbol{\Gamma}_i$ is through their empirical versions. The sample covariance operator $\widehat{\boldsymbol{\Gamma}}_i$ is defined as

$$\widehat{\boldsymbol{\Gamma}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(X_{i,j} - \overline{X}_i\right) \otimes \left(X_{i,j} - \overline{X}_i\right),$$

where $\overline{X}_i = (1/n_i) \sum_{j=1}^{n_i} X_{i,j}$. Dauxois et al. (1982) obtained the asymptotic behaviour of $\widehat{\boldsymbol{\Gamma}}_i$. In particular, they have shown that when $\mathbb{E}(\|X_i\|^4) < \infty$, $\sqrt{n_i}\left(\widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i\right)$ converges in distribution to a zero mean Gaussian random element of $\mathcal{F}$, $\mathbf{U}_i$, with covariance operator $\boldsymbol{\Upsilon}_i$ given by

$$\begin{aligned}\boldsymbol{\Upsilon}_i = &\sum_{m,r,o,p} \lambda_{im}^{1/2} \lambda_{ir}^{1/2} \lambda_{io}^{1/2} \lambda_{ip}^{1/2} \mathbb{E}\left(f_{im} f_{ir} f_{io} f_{ip}\right) \phi_{i,m} \otimes \phi_{i,r} \tilde{\otimes} \phi_{i,o} \otimes \phi_{i,p} \\ &- \sum_{m,r} \lambda_{im} \lambda_{ir}\, \phi_{i,m} \otimes \phi_{i,m} \tilde{\otimes} \phi_{i,r} \otimes \phi_{i,r} \end{aligned} \tag{1}$$

where $\tilde{\otimes}$ stands for the tensor product in $\mathcal{F}$ and, as mentioned before, $\{\phi_{i,\ell} : \ell \ge 1\}$ is an orthonormal basis of eigenfunctions of $\boldsymbol{\Gamma}_i$ with associated eigenvalues $\{\lambda_{i,\ell} : \ell \ge 1\}$ such that $\lambda_{i,\ell} \ge \lambda_{i,\ell+1}$. The random variables f_{im} are the standardized coordinates of $X_i - \mu_i$ on the basis $\{\phi_{i,\ell} : \ell \ge 1\}$, that is, $f_{im} = \lambda_{i,m}^{-1/2} \langle X_i - \mu_i, \phi_{i,m} \rangle$ and $\mathbb{E}(f_{im}) = 0$. Using that $Cov\left(\langle u, X_i - \mu_i \rangle, \langle v, X_i - \mu_i \rangle\right) = \langle u, \boldsymbol{\Gamma}_i v \rangle$, we get that $\mathbb{E}(f_{im}^2) = 1$, $\mathbb{E}(f_{im}\ f_{is}) = 0$ for $m \ne s$. In particular, the Karhunen–Loéve expansion leads to

$$X_i = \mu_i + \sum_{\ell=1}^{\infty} \lambda_{i,\ell}^{\frac{1}{2}} f_{i\ell}\, \phi_{i,\ell}\,. \tag{2}$$

It is worth noticing that, since $\mathbb{E}\|\mathbf{U}_i\|_{\mathcal{F}}^2 < \infty$, $\boldsymbol{\Upsilon}_i$ is a linear operator over $\mathcal{F}$ with finite trace, so it is also a Hilbert–Schmidt operator. Thus, any linear combination of the operators $\boldsymbol{\Upsilon}_i$, $\boldsymbol{\Upsilon} = \sum_{i=1}^{k} a_i \boldsymbol{\Upsilon}_i$, with $a_i \ge 0$, will be trace class operator, that is, if $\{\theta_\ell\}_{\ell \ge 1}$ stands for the eigenvalues of $\boldsymbol{\Upsilon}$ ordered in decreasing order, we have that $\theta_\ell \ge 0$ and $\sum_{\ell \ge 1} \theta_\ell < \infty$. This property will be used later in Theorem 1.

When $\mathcal{H} = L^2(\mathcal{I})$, smooth estimators $\widehat{\boldsymbol{\Gamma}}_{i,h}$ of the covariance operators were studied in Boente and Fraiman (2000). The smoothed operator is the operator induced by the smooth covariance function

$$\widehat{\gamma}_{i,h}(t,s) = \frac{1}{n_1} \sum_{j=1}^{n_i} \left(X_{i,j,h}(t) - \overline{X}_{i,h}(t)\right)\left(X_{i,j,h}(s) - \overline{X}_{i,h}(s)\right),$$

where $X_{i,j,h}(t) = \int_{\mathcal{I}} K_h(t - x) X_{i,j}(t) \mathrm{d}t$ are the smoothed trajectories, $K_h(\cdot) = h^{-1} K(\cdot/h)$ is a nonnegative kernel function, and h a smoothing parameter. Boente and Fraiman (2000) have shown that the smooth estimators have the same asymptotic distribution as the empirical version, under mild conditions.

3 The test statistic

To motivate our test statistic, we first consider the two sample setting, that is, the problem of testing the hypothesis

$$H_0 : \boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_2 \quad \text{against} \quad H_1 : \boldsymbol{\Gamma}_1 \neq \boldsymbol{\Gamma}_2, \tag{3}$$

from two independent samples $X_{1,1}, \ldots, X_{1,n_1}$ and $X_{2,1}, \ldots, X_{2,n_2}$. A natural approach is to consider $\widehat{\boldsymbol{\Gamma}}_i$ as the empirical covariance operators of each population and construct a statistic T_n based on the difference between the covariance operators estimators, i.e., to define

$$T_n = n \|\widehat{\boldsymbol{\Gamma}}_1 - \widehat{\boldsymbol{\Gamma}}_2\|_{\mathcal{F}}^2 , \tag{4}$$

where $n = n_1 + n_2$, $n_i/n \to \tau_i$ with $\tau_i \in (0, 1)$. As mentioned in Pigoli et al. (2014), the null hypothesis can be written as $d(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) = \|\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2\|_{\mathcal{F}} = 0$ while the alternative corresponds to $\|\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2\|_{\mathcal{F}} > 0$. Thus, if $\widehat{\boldsymbol{\Gamma}}_j$ are consistent estimators of $\boldsymbol{\Gamma}_j$ for $j = 1, 2$, any test based on the distance $d(\widehat{\boldsymbol{\Gamma}}_1, \widehat{\boldsymbol{\Gamma}}_2)$ between $\widehat{\boldsymbol{\Gamma}}_1$ and $\widehat{\boldsymbol{\Gamma}}_2$ will be consistent.

To generalize the procedure to several populations, let $\boldsymbol{\Gamma}_i$ stand for the covariance operator of the $i-$th population. We wish to test the null hypothesis

$$H_0 : \boldsymbol{\Gamma}_1 = \cdots = \boldsymbol{\Gamma}_k \quad \text{against} \quad H_1 : \exists\, i \neq j \text{ such that } \boldsymbol{\Gamma}_i \neq \boldsymbol{\Gamma}_j\,. \tag{5}$$

The null hypothesis is equivalent to $\sum_{j=2}^{k} \|\boldsymbol{\Gamma}_j - \boldsymbol{\Gamma}_1\|_{\mathcal{F}}^2 = 0$ allowing to construct a consistent test using consistent covariance operator estimators. To be more precise, let $X_{i,1}, \ldots, X_{i,n_i}$, $1 \leq i \leq k$, be independent samples, $n = n_1 + \cdots + n_k$ and assume that $n_i/n \to \tau_i$, $0 < \tau_i < 1$, $\sum_{i=1}^{k} \tau_i = 1$. Denote by $\widehat{\boldsymbol{\Gamma}}_i$ the sample covariance operator of $i-$th population. A natural generalization of the statistic defined in (4) is to consider the test statistic

$$T_{k,n} = n \sum_{j=2}^{k} \|\widehat{\boldsymbol{\Gamma}}_j - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 . \tag{6}$$

To define the test, we need the asymptotic distribution of $T_{k,n}$ under the null hypothesis, which is derived in Corollary 1.

The following result allows to study the null asymptotic behaviour of $n \sum_{j=2}^{k} \|\widetilde{\boldsymbol{\Gamma}}_j - \widetilde{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2$ when considering a general class of covariance estimators $\widetilde{\boldsymbol{\Gamma}}_i$ rather than the sample covariance operators.

Theorem 1 *Let $X_{i,1}, \ldots, X_{i,n_i}$, for $1 \leq i \leq k$, be independent observations from k independent distributions in $\mathcal{H}$, with mean μ_i and covariance operator $\boldsymbol{\Gamma}_i$. Assume that*

$n_i/n \to \tau_i$ with $\tau_i \in (0,1)$ where $n = \sum_{i=1}^k n_i$. Let $\widetilde{\boldsymbol{\Gamma}}_i$ be the independent estimators of the $i-$th population covariance operator such that $\sqrt{n_i}\left(\widetilde{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i\right) \xrightarrow{D} \mathbf{U}_i$, with $\mathbf{U}_i$ a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_i$. Denote $\boldsymbol{\Upsilon}_w = \left(\boldsymbol{\Upsilon}_{w,1}, \ldots, \boldsymbol{\Upsilon}_{w,k-1}\right)$ the trace operator $\boldsymbol{\Upsilon}_w : \mathcal{F}^{k-1} \to \mathcal{F}^{k-1}$ with $i-$th component defined as

$$\boldsymbol{\Upsilon}_{w,i}(u_1, \ldots, u_{k-1}) = \frac{1}{\tau_{i+1}} \boldsymbol{\Upsilon}_{i+1}(u_i) + \frac{1}{\tau_1} \boldsymbol{\Upsilon}_1 \left(\sum_{\ell=1}^{k-1} u_\ell \right) \qquad \text{for } 1 \le i \le k-1\,. \tag{7}$$

Let $\{\theta_\ell\}_{\ell \ge 1}$ stand for the sequence of eigenvalues of $\boldsymbol{\Upsilon}_w$ ordered in decreasing order. Then, we have that

$$n \sum_{j=2}^{k} \|(\widetilde{\boldsymbol{\Gamma}}_j - \boldsymbol{\Gamma}_j) - (\widetilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{\Gamma}_1)\|_{\mathcal{F}}^2 \xrightarrow{\mathcal{D}} \sum_{\ell \ge 1} \theta_\ell Z_\ell^2\,,$$

with $Z_\ell \sim N(0,1)$ independent. In particular, if $H_0 : \boldsymbol{\Gamma}_1 = \cdots = \boldsymbol{\Gamma}_k$ holds, we have that $n \sum_{j=2}^k \|\widetilde{\boldsymbol{\Gamma}}_j - \widetilde{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 \xrightarrow{\mathcal{D}} \sum_{\ell \ge 1} \theta_\ell Z_\ell^2$.

When $\mathbb{E}(\|X_i\|^4) < \infty$, the results in Theorem 1 apply in particular to the sample covariance operator, i.e., when $\widetilde{\boldsymbol{\Gamma}}_i = \widehat{\boldsymbol{\Gamma}}_i$, leading to the asymptotic distribution of $T_{k,n}$ under the null hypothesis stated in Corollary 1. However, it also allows to use other covariance estimators to define the test statistic, such as the smooth ones $\widehat{\boldsymbol{\Gamma}}_{i,h}$ defined in Boente and Fraiman (2000).

Corollary 1 *Let $X_{i,1}, \ldots, X_{i,n_i}$, for $1 \le i \le k$, be independent observations from k independent distributions in $\mathcal{H}$, with mean μ_i and covariance operator $\boldsymbol{\Gamma}_i$ such that $\mathbb{E}(\|X_i\|^4) < \infty$. Let $\widehat{\boldsymbol{\Gamma}}_i$ be the sample covariance operator of the $i-$th population. Assume that $n_i/n \to \tau_i$ with $\tau_i \in (0,1)$ where $n = \sum_{i=1}^k n_i$. Denote $\boldsymbol{\Upsilon}_w = \left(\boldsymbol{\Upsilon}_{w,1}, \ldots, \boldsymbol{\Upsilon}_{w,k-1}\right)$ the trace operator $\boldsymbol{\Upsilon}_w : \mathcal{F}^{k-1} \to \mathcal{F}^{k-1}$ where $\boldsymbol{\Upsilon}_{w,i}$ is defined in (7) with $\boldsymbol{\Upsilon}_i$ given in (1). Let $\{\theta_\ell\}_{\ell \ge 1}$ stand for the sequence of eigenvalues of $\boldsymbol{\Upsilon}_w$ c$H_0 : \boldsymbol{\Gamma}_1 = \cdots = \boldsymbol{\Gamma}_k$, we have*

$$n \sum_{j=2}^{k} \|\widehat{\boldsymbol{\Gamma}}_j - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 \xrightarrow{\mathcal{D}} \sum_{\ell \ge 1} \theta_\ell Z_\ell^2\,, \tag{8}$$

with $Z_\ell \sim N(0,1)$ independent.

Remark 1 (a) The fact that $\mathbb{E}(\|X_i\|^4) < \infty$ entails that $\mathbb{E}(\|(X_i - \mu_i) \otimes (X_i - \mu_i)\|^2) < \infty$. Thus, the covariance operator $\boldsymbol{\Upsilon}_i$ of $(X_i - \mu_i) \otimes (X_i - \mu_i)$ is well defined and $\sum_{\ell \ge 1} \theta_\ell < \infty$. Hence, for any sequence of integers $\{q_n\}$ such that $q_n \to \infty$, the sequence $\mathcal{U}_n = \sum_{\ell=1}^{q_n} \theta_\ell Z_i^2$ is Cauchy in $L^2(\mathbb{P})$, so the limit $\mathcal{U} = \sum_{\ell \ge 1} \theta_\ell Z_\ell^2$ is well defined. In fact, analogous arguments to those considered in Neuhaus (1980) allow us to show that the series converges almost surely. Moreover, since $Z_1^2 \sim \chi_1^2$, $\mathcal{U}$ has a continuous distribution function $F_{\mathcal{U}}$ entailing that the distribution function $F_{\mathcal{U}_n}$ of $\mathcal{U}_n$ converges to $F_{\mathcal{U}}$ uniformly (see, for instance, Lemma 2.11 in Vaart (2000)).

(b) It is worth noticing that Corollary 1 is a natural extension of its analogous in the finite-dimensional case. To be more precise, let $\mathbf{Z}_{ij} \in \mathbb{R}^p$ with $1 \leq i \leq k$ and $1 \leq j \leq n_i$ be independent random vectors and let $\widehat{\mathbf{\Sigma}}_i$ be their sample covariance matrix. Then, $\sqrt{n_i}\mathbf{V}_i = \sqrt{n_i}(\widehat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}_i)$ converges to a multivariate normal distribution with mean zero and covariance matrix Υ_i. Let

$$\mathbf{A} = \begin{pmatrix} -\mathbf{I}_p & \mathbf{I}_p & 0 & \dots & 0 \\ -\mathbf{I}_p & 0 & \mathbf{I}_p & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\mathbf{I}_p & 0 & \dots & 0 & \mathbf{I}_p \end{pmatrix},$$

where $\mathbf{I}_p$ stands for the identity matrix of order p. Then, straightforward calculations allow us to show that $\sqrt{n}\mathbf{A}(\mathbf{V}_1, \dots, \mathbf{V}_k)^t \xrightarrow{D} N(0, \Upsilon)$, where

$$\Upsilon = \begin{pmatrix} \tau_1{}^{-1}\Upsilon_1 + \tau_2{}^{-1}\Upsilon_2 & \tau_1{}^{-1}\Upsilon_1 & \dots & \tau_1{}^{-1}\Upsilon_1 \\ \tau_1{}^{-1}\Upsilon_1 & \tau_1{}^{-1}\Upsilon_1 + \tau_3{}^{-1}\Upsilon_3 & \dots & \tau_1{}^{-1}\Upsilon_1 \\ \vdots & \vdots & \vdots & \vdots \\ \tau_1{}^{-1}\Upsilon_1 & \tau_1{}^{-1}\Upsilon_1 & \dots & \tau_1{}^{-1}\Upsilon_1 + \tau_k{}^{-1}\Upsilon_k \end{pmatrix}.$$

Therefore, under the null hypothesis of equality of the covariance matrices $\mathbf{\Sigma}_i$, we have that $n\sum_{i=2}^{k} \|\widehat{\mathbf{\Sigma}}_i - \widehat{\mathbf{\Sigma}}_1\|^2 = \|\sqrt{n}\mathbf{A}\mathbf{V}\|^2 \xrightarrow{D} \sum_{\ell=1}^{kp^4} \theta_\ell Z_\ell^2$ where $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_k)$ and $\theta_1, \theta_2, \dots, \theta_{kp^4}$ are the eigenvalues of Υ. It is worth noticing that the matrix Υ is the finite-dimensional version of the operator $\boldsymbol{\Upsilon}_w$.

(c) From Corollary 1 we have that, under the null hypothesis $H_0 : \boldsymbol{\Gamma}_1 = \dots = \boldsymbol{\Gamma}_k$, the test statistic $T_{k,n} = n\sum_{j=2}^{k} \|\widehat{\boldsymbol{\Gamma}}_j - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2, \xrightarrow{D} \mathcal{U} = \sum_{\ell \geq 1} \theta_\ell Z_\ell^2$. Hence, an asymptotic test may be based on $T_{k,n}$, rejecting the null hypothesis for large values of $T_{k,n}$. To obtain the critical values, the distribution of $\mathcal{U}$ and thus the eigenvalues of $\boldsymbol{\Upsilon}_w$ need to be estimated. In particular, when $k = 2$, the test statistic $T_{k,n}$ equals $T_n = n\|\widehat{\boldsymbol{\Gamma}}_1 - \widehat{\boldsymbol{\Gamma}}_2\|_{\mathcal{F}}^2$ and $\boldsymbol{\Upsilon}_w = \tau_1^{-1}\boldsymbol{\Upsilon}_1 + \tau_2^{-1}\boldsymbol{\Upsilon}_2$. As mentioned before, the distribution function of $\mathcal{U}$ can be uniformly approximated by that of $\mathcal{U}_n$ allowing us to approximate the critical values by the $(1-\alpha)-$quantile of $\mathcal{U}_n$. Gupta and Xu (2006) provide an approximation for the distribution function of any finite mixture of χ_1^2 independent random variables that can be used in the computation of the $(1-\alpha)-$quantile of $\sum_{\ell=1}^{q_n} \widehat{\theta}_\ell Z_\ell^2$, where $\widehat{\theta}_\ell$ are estimators of θ_ℓ. It is also worth noticing that, under $H_0 : \boldsymbol{\Gamma}_1 = \dots = \boldsymbol{\Gamma}_k$, the operator $\boldsymbol{\Upsilon}_i$ given in (1) reduces to

$$\boldsymbol{\Upsilon}_i = \sum_{m,r,o,p} \lambda_m^{1/2}\lambda_r^{1/2}\lambda_o^{1/2}\lambda_p^{1/2}\mathbb{E}[f_{im} f_{ir} f_{io} f_{ip}]\, \phi_m \otimes \phi_r \tilde{\otimes} \phi_o \otimes \phi_p$$
$$-\sum_{m,r} \lambda_m \lambda_r\, \phi_m \otimes \phi_m \tilde{\otimes} \phi_r \otimes \phi_r$$

for $i = 1, \dots, k$, where, for the sake of simplicity, we denote as λ_m the $m-$th largest eigenvalue of $\boldsymbol{\Gamma}_1$ and ϕ_m its corresponding eigenfunction.
In particular, if all the populations have the same underlying distribution except for their means and covariance operators, as it happens when comparing the covariance operators of Gaussian processes, the random functions f_{im}, $i = 2, \dots, k$, have the same distribution as f_{1m}, so, in this case, $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Upsilon}_i$, for $i = 2, \dots, k$, under H_0.

(d) Assume that the processes are Gaussian. Then, using that $\mathbb{E}(f_{im} f_{ir} f_{io} f_{ip})$ equals 1 when pairs of indices are equal, 3 when $m = r = o = p$ and 0 otherwise, we have that, under the null hypothesis

$$\begin{aligned}\boldsymbol{\Upsilon}_i = \boldsymbol{\Upsilon}_1 &= \sum_{i \neq j} \lambda_i \lambda_j \, \phi_i \otimes \phi_j \tilde{\otimes} \phi_j \otimes \phi_i + \sum_{i \neq j} \lambda_i \lambda_j \, \phi_i \otimes \phi_j \tilde{\otimes} \phi_i \otimes \phi_j \\ &\quad + 2 \sum_i \lambda_i^2 \, \phi_i \otimes \phi_i \tilde{\otimes} \phi_i \otimes \phi_i \\ &= 2 \sum_i \lambda_i^2 \, \phi_i \otimes \phi_i \tilde{\otimes} \phi_i \otimes \phi_i \\ &\quad + \sum_{i<j} \lambda_i \lambda_j (\phi_i \otimes \phi_j + \phi_j \otimes \phi_i) \tilde{\otimes} (\phi_i \otimes \phi_j + \phi_j \otimes \phi_i)\,.\end{aligned}$$

Using that $\phi_i \otimes \phi_i$ and $(\phi_i \otimes \phi_j + \phi_j \otimes \phi_i)/\sqrt{2}$, for $i < j$, constitutes a complete orthonormal basis of the space of self-adjoint Hilbert–Schmidt operators, we conclude that they are the eigenfunctions of $\boldsymbol{\Upsilon}_1$ associated to the eigenvalues $2\lambda_i^2$ and $2\lambda_i \lambda_j$, respectively. Furthermore, if $\tau_i = 1/k$ for $i = 1 \dots, k$, we get that for $1 \leq i \leq k - 1$

$$\begin{aligned}\boldsymbol{\Upsilon}_{w,i}(u_1, \dots, u_{k-1}) &= k \left[\boldsymbol{\Upsilon}_1(u_i) + \boldsymbol{\Upsilon}_1 \left(\sum_{\ell=1}^{k-1} u_\ell \right) \right] \\ &= k \left[\boldsymbol{\Upsilon}_1(u_i) + \sum_{\ell=1}^{k-1} \boldsymbol{\Upsilon}_1\left(u_\ell\right) \right], \end{aligned} \tag{9}$$

which entails that $\theta_{i,i} = 2k^2 \lambda_i^2$ and $\theta_{i,j} = 2k^2 \lambda_i \lambda_j$, for $i < j$, are eigenvalues of $\boldsymbol{\Upsilon}_w = (\boldsymbol{\Upsilon}_{w,1}, \dots, \boldsymbol{\Upsilon}_{w,k-1})$, related to the eigenfunctions $v_{i,i} = (\phi_i \otimes \phi_i, \dots, \phi_i \otimes \phi_i)$ and $v_{i,j} = ((\phi_i \otimes \phi_j + \phi_j \otimes \phi_i)/\sqrt{2}, \dots, (\phi_i \otimes \phi_j + \phi_j \otimes \phi_i)/\sqrt{2})$, respectively. On the other hand, if α is an eigenvalue of $\boldsymbol{\Upsilon}_w$, α/k^2 is an eigenvalue of $\boldsymbol{\Upsilon}_1$, meaning that we have obtained all the eigenvalues of $\boldsymbol{\Upsilon}_w$.

4 Behaviour under local alternatives

In this section, we study the behaviour of the test statistic $T_{k,n}$ under a set of local alternatives. It is clear that, as in the multivariate situation, there are many ways in which the covariance operators may differ, one of them being the functional common principal model in which discrepancies from the null hypothesis arise only in the eigenvalues and not in the eigenfunctions of the covariance operators. Our results include that setting but also a situation in which the processes can be written as sums of two independent processes, one of them having the same covariance operator along populations.

We decided to fix the distribution of the first population, while that of the remaining ones will depend on the sample size, in such a way that for each fixed n the alternative assumption holds but, as is usual for local alternatives, when the sample sizes increase, the alternatives considered converge to the null hypothesis at a given rate. To avoid burdening the notation, in this section, for $1 \leq j \leq n_i$, $2 \leq i \leq k$, we will use $X_{i,j}$ to denote the observations under the local alternatives $X_{i,j}^{(n)}$ when it is clear. Similarly, we denote by X_i instead of $X_i^{(n)}$ the random element with common distribution, that is, $X_{i,j} \sim X_i$.

As in Sect. 3, the following result presents a general framework which allows to study the distribution of the test statistic under root$-n$ local alternatives. Theorem 2 together with Propositions 1 and 2 allows to derive the behaviour of the test statistic $T_{k,n}$ under the local alternatives described before. However, Theorem 2 may also be applied when considering covariance estimators other than the sample covariance estimators.

Theorem 2 *Let $X_{i,1}, \ldots, X_{i,n_i}$ for $i = 1, \ldots, k$ be independent observations from k independent distributions in $\mathcal{H}$ with covariance operators $\boldsymbol{\Gamma}_i$ such that, for $i \geq 2$, $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_{i,n} = \boldsymbol{\Gamma}_1 + n^{-1/2}\boldsymbol{\Delta}_i$. Assume that $\boldsymbol{\Delta}_i$ is a self-adjoint trace operator such that $\boldsymbol{\Gamma}_{i,n}$ is nonnegative. Denote as $\boldsymbol{\Delta}^{(k-1)} = (\boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_k)^t \in \mathcal{F}^{k-1}$, $n = \sum_{i=1}^k n_i$ and assume that $n_i/n \to \tau_i \in (0,1)$. Let $\widetilde{\boldsymbol{\Gamma}}_i$ be the independent estimators of the $i-$th population covariance operator such that, for $1 \leq i \leq k$, $\sqrt{n_i}\left(\widetilde{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1\right) \xrightarrow{D} \mathbf{U}_i + \tau_i^{1/2}\boldsymbol{\Delta}_i$ where $\mathbf{U}_i$ is a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_i$ and $\boldsymbol{\Delta}_1 = \mathbf{O}$ stands for the null operator. Define $\boldsymbol{\Upsilon}_w = (\boldsymbol{\Upsilon}_{w,1}, \ldots, \boldsymbol{\Upsilon}_{w,k-1})$ where $\boldsymbol{\Upsilon}_{w,i}$ is given in (7) and let $\{\upsilon_\ell\}_{\ell \geq 1}$ be an orthonormal basis of eigenfunctions of $\boldsymbol{\Upsilon}_w$ related to the eigenvalues $\{\theta_\ell\}_{\ell \geq 1}$ ordered in decreasing order. Then,*

$$n \sum_{i=2}^{k} \|\widetilde{\boldsymbol{\Gamma}}_i - \widetilde{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 \xrightarrow{D} \sum_{\ell \geq 1} \theta_\ell \left(Z_\ell + \frac{\eta_\ell}{\sqrt{\theta_\ell}} \right)^2,$$

where Z_ℓ are independent and $Z_\ell \sim N(0,1)$ and $\eta_\ell = \langle \boldsymbol{\Delta}^{(k-1)}, \upsilon_\ell \rangle_{\mathcal{F}^{k-1}}$, i.e., $\boldsymbol{\Delta}^{(k-1)} = \sum_{\ell \geq 1} \eta_\ell \upsilon_\ell$.

The requirement that $\boldsymbol{\Delta}_i$ is a self-adjoint trace operator is needed to guarantee that $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_{i,n}$ is a valid covariance operator. Besides, since $\boldsymbol{\Delta}_i$ has finite trace, we have that $\boldsymbol{\Delta}^{(k-1)} \in \mathcal{F}^{k-1}$, so $\sum_{\ell \geq 1} \eta_\ell^2 < \infty$. Furthermore, if $\eta_\ell \neq 0$ for some ℓ, then the test based on the asymptotic null distribution of $n \sum_{i=2}^k \|\widetilde{\boldsymbol{\Gamma}}_i - \widetilde{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2$ is consistent.

As mentioned at the beginning of this section, we will consider two scenarios where the assumptions of Theorem 2 are satisfied. The first one is a generalization of Example 2.2 in Gaines et al. (2011) and assumes that, for $i = 2, \ldots, k$, the observations from the i-th population can be written as the sum of two independent processes, the first one having the same covariance operator as X_1. Namely, we assume that

$$X_{i,j} = X_{i,j}^{(n)} = W_{i,j} + n^{-1/4} R_{i,j}, \quad \text{for } 2 \le i \le k, \tag{10}$$

where $W_{i,j}$, $R_{i,j}$ are independent and such that $W_{i,j} \sim W_i$, $R_{i,j} \sim R_i$ and W_i has the same covariance operator as X_1, for $1 \le i \le k$. Notice that the distribution of the term R_i is free to vary across populations, for $2 \le i \le k$, as well as the distribution of W_i as long as W_i and X_1 share the same covariance operator.

From now on, let $\{\phi_\ell\}_{\ell \ge 1}$ be the eigenfunctions of $\boldsymbol{\Gamma}_1$, the covariance operator of X_1, and denote λ_ℓ the eigenvalues of $\boldsymbol{\Gamma}_1$ related to ϕ_ℓ, that is, we omit the subscript 1 in $\lambda_{1,\ell}$ and $\phi_{1,\ell}$.

Proposition 1 *Let $X_{i,1}, \ldots, X_{i,n_i}$, $i = 1, \ldots, k$ be independent observations from k independent distributions in $\mathcal{H}$ such that (10) holds. Assume that $n_i/n \to \tau_i \in (0, 1)$ with $n = \sum_{i=1}^k n_i$, $\mathbb{E}(\|X_1\|^4) < \infty$ and that, for $2 \le i \le k$, $\mathbb{E}(\|W_i\|^4) < \infty$ and $\mathbb{E}(\|R_i\|^4) < \infty$. Let $\boldsymbol{\Delta}_i$ be the covariance operator of R_i, for $i = 2, \ldots, k$ and assume that $\boldsymbol{\Gamma}_1 = \mathbb{E}\{(X_1 - \mu_1) \otimes (X_1 - \mu_1)\}$ is also the covariance operator of W_i, for $i = 2, \ldots, k$. Denote as $\widehat{\boldsymbol{\Gamma}}_i$ the sample covariance operator of the $i-$th population. Then, we have that $\sqrt{n_i}\left(\widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1\right) \xrightarrow{D} \mathbf{U}_i + \tau_i^{1/2} \boldsymbol{\Delta}_i$ with $\mathbf{U}_i$ a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_i$ given in (1), that is,*

$$\begin{aligned}\boldsymbol{\Upsilon}_i &= \sum_{m,r,o,p} \lambda_m^{1/2} \lambda_r^{1/2} \lambda_o^{1/2} \lambda_p^{1/2} \mathbb{E}\left[f_{im} f_{ir} f_{io} f_{ip}\right] \phi_m \otimes \phi_r \tilde{\otimes} \phi_o \otimes \phi_p \\ &\quad - \sum_{m,r} \lambda_m \lambda_r \, \phi_m \otimes \phi_m \tilde{\otimes} \phi_r \otimes \phi_r \,,\end{aligned} \tag{11}$$

where f_{im} are the standardized coordinates of $W_i - \mathbb{E}(W_i)$ on the basis $\{\phi_\ell : \ell \ge 1\}$, i.e., $\lambda_\ell^{1/2} f_{i\ell} = \langle W_i - \mathbb{E}(W_i), \phi_\ell \rangle$.

If W_i has the same distribution as X_1, then we have that $\boldsymbol{\Upsilon}_i = \boldsymbol{\Upsilon}_1$.

The second model for local alternatives to be considered in this section is the functional common principal model. These alternatives include, as a particular case, alternatives following the proportional model $\boldsymbol{\Gamma}_{i,n} = (1 + \rho_i/\sqrt{n})\boldsymbol{\Gamma}_1$. For details on the functional principal component model, we refer to Benko et al. (2009) and Boente et al. (2010) , for instance. By assuming local alternatives satisfying a functional common principal model, we get that the processes X_i, $1 \le i \le k$, can be written as

$$X_1 = \mu_1 + \sum_{\ell=1}^{\infty} \lambda_\ell^{\frac{1}{2}} f_{1\ell} \phi_\ell \quad \text{and} \quad X_i = X_i^{(n)} = \mu_i + \sum_{\ell=1}^{\infty} \lambda_{i,\ell}^{(n)\frac{1}{2}} f_{i\ell} \phi_\ell, \text{ for } i \ge 2 \tag{12}$$

with $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$, $\lambda_{i,\ell}^{(n)} \to \lambda_\ell$ at a given rate, while $f_{i\ell}$ are random variables such that $\mathbb{E}(f_{i\ell}) = 0$, $\mathbb{E}(f_{i\ell}^2) = 1$, $\mathbb{E}(f_{i\ell}\, f_{is}) = 0$ for $\ell \ne s$.

Proposition 2 gives the asymptotic behaviour of the sample covariance operators when choosing $\lambda_{i,\ell}^{(n)} = \lambda_\ell(1 + n^{-1/2}\Delta_{i,\ell})$ in (12). It is worth noting that if $(1 +$

$n^{-1/2}\Delta_{i,\ell}) \geq 0$ and some additional conditions on $\Delta_{i,\ell}$ to be stated below are fulfilled, then $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_{i,n} = \boldsymbol{\Gamma}_1 + n^{-1/2}\boldsymbol{\Delta}_i$, for $i \geq 2$, where $\boldsymbol{\Delta}_i = \sum_{\ell \geq 1} \Delta_{i,\ell}\lambda_\ell\phi_\ell \otimes \phi_\ell$. Hence, Proposition 2 together with Theorem 2 leads to the asymptotic behaviour of the test statistic $T_{k,n}$ under local alternatives following a functional common principal model, as stated in Corollary 2.

Proposition 2 *Let $X_{i,1}, \ldots, X_{i,n_i}$, $i = 1, \ldots, k$, be independent observations from k independent distributions in $\mathcal{H}$, such that $X_{i,j} \sim X_i$. Assume that X_i satisfy (12) with $\lambda_{i,\ell}^{(n)} = \lambda_\ell(1 + n^{-1/2}\Delta_{i,\ell})$ and that $n_i/n \to \tau_i \in (0, 1)$ where $n = \sum_{i=1}^k n_i$. Let $\widehat{\boldsymbol{\Gamma}}_i$ be the sample covariance operator of the $i-$th population. Furthermore, assume that $\mathbb{E}(\|X_1\|^4) < \infty$, $\sigma_{4,i,\ell}^2 = \mathbb{E}(f_{i\ell}^4) < \infty$, $\sum_{\ell=1}^\infty \lambda_\ell|\Delta_{i,\ell}| < \infty$, $\sum_{\ell=1}^\infty \lambda_\ell\Delta_{i,\ell}^2\sigma_{4,i,\ell} < \infty$, $\sum_{\ell=1}^\infty \lambda_\ell\Delta_{i,\ell}^2 < \infty$ and $\sum_{\ell=1}^\infty \lambda_\ell\sigma_{4,i,\ell} < \infty$, for $i = 2, \ldots, k$. Then, $\sqrt{n_i}\left(\widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1\right) \xrightarrow{D} \mathbf{U}_i + \tau_i^{1/2}\boldsymbol{\Delta}_i$, where $\boldsymbol{\Delta}_i = \sum_{\ell \geq 1} \Delta_{i,\ell}\lambda_\ell\phi_\ell \otimes \phi_\ell$ and $\mathbf{U}_i$ a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_i$ given by (11) where f_{im} are defined in (12).*

Remark 2 The conditions $\sum_{\ell=1}^\infty \lambda_\ell|\Delta_{i,\ell}| < \infty$ and $\lambda_\ell(1 + n^{-1/2}\Delta_{i,\ell}) \geq 0$ ensure that $\boldsymbol{\Delta}_i$ is a self-adjoint trace operator and that $\boldsymbol{\Gamma}_{i,n}$ is nonnegative, respectively. It is worth noticing that if the observations $X_{i,j}$ have a Gaussian distribution for all the populations, then $f_{i\ell} \sim N(0, 1)$, so $\sigma_{4,i,\ell}^2 = 3$. This implies that the conditions $\sum_{\ell=1}^\infty \lambda_\ell\Delta_{i,\ell}^2\sigma_{4,i,\ell} < \infty$, $\sum_{\ell=1}^\infty \lambda_\ell\Delta_{i,\ell}^2 < \infty$ and $\sum_{\ell=1}^\infty \lambda_\ell\sigma_{4,i,\ell} < \infty$ reduce to $\sum_{\ell=1}^\infty \lambda_\ell\Delta_{i,\ell}^2 < \infty$, since $\sum_{\ell=1}^\infty \lambda_\ell < \infty$. Moreover, when considering root$-n$ local proportional alternatives, i.e., when $\Delta_{i,\ell} = \rho_i$, the condition $\sum_{\ell=1}^\infty \lambda_\ell\Delta_{i,\ell}^2 < \infty$ is immediately fulfilled since $\boldsymbol{\Gamma}_1$ is a trace operator.

Theorem 2 and Propositions 1 and 2 lead immediately to the asymptotic distribution of the test statistic $T_{k,n}$ under the local alternatives studied before. We summarize this result in Corollary 2.

Corollary 2 *Let $X_{i,1}, \ldots, X_{i,n_i}$ for $i = 1, \ldots, k$ be independent observations from k independent distributions in $\mathcal{H}$, with mean μ_i and covariance operator $\boldsymbol{\Gamma}_i$ such that $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_{i,n} = \boldsymbol{\Gamma}_1 + n^{-1/2}\boldsymbol{\Delta}_i$, for $i \geq 2$. Let $\widehat{\boldsymbol{\Gamma}}_i$ be the sample covariance operator of the $i-$th population. Assume that the assumptions of Propositions 1 or 2 hold and denote $\boldsymbol{\Upsilon}_w = (\boldsymbol{\Upsilon}_{w,1}, \ldots, \boldsymbol{\Upsilon}_{w,k-1})$ where $\boldsymbol{\Upsilon}_{w,i}$ is defined in (7) with $\boldsymbol{\Upsilon}_i$ given in (11). Let $\{\upsilon_\ell\}_{\ell \geq 1}$ be the orthonormal eigenfunctions of $\boldsymbol{\Upsilon}_w$ related to the eigenvalues $\{\theta_\ell\}_{\ell \geq 1}$ ordered in decreasing order and $\eta_\ell = \langle \boldsymbol{\Delta}^{(k-1)}, \upsilon_\ell \rangle_{\mathcal{F}^{k-1}}$. Then, we have that*

$$T_{k,n} = n \sum_{i=2}^k \|\widehat{\boldsymbol{\Gamma}}_i - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 \xrightarrow{D} \sum_{\ell \geq 1} \theta_\ell \left(Z_\ell + \frac{\eta_\ell}{\sqrt{\theta_\ell}} \right)^2,$$

where Z_ℓ are independent, $Z_\ell \sim N(0, 1)$.

Under the local alternatives $\boldsymbol{\Gamma}_{i,n} = \boldsymbol{\Gamma}_1 + n^{-1/2}\boldsymbol{\Delta}_i$, for $i \geq 2$, and, in particular, under those given in Propositions 1 and 2, similar arguments to those considered in the proof of Proposition 4 in Boente and Fraiman (2000) allow us to show that, if

$h = h_n \to 0$, the smooth estimator $\widehat{\boldsymbol{\Gamma}}_{i,h}$ has the same asymptotic behaviour as $\widehat{\boldsymbol{\Gamma}}_i$, i.e., that $\sqrt{n_i}\|(\widehat{\boldsymbol{\Gamma}}_{i,h} - \boldsymbol{\Gamma}_{1,h}) - (\widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1)\|_{\mathcal{F}} \xrightarrow{p} 0$, where $\boldsymbol{\Gamma}_{1,h}$ is the smoothed covariance operator. On the other hand, Proposition 3 in Boente and Fraiman (2000) entails that $\sqrt{n}\|\boldsymbol{\Gamma}_{1,h} - \boldsymbol{\Gamma}_1\|_{\mathcal{F}} \to 0$ if, in addition, $n\,h \to 0$, the kernel K has finite first moment and the covariance kernel γ_1 satisfies the following Lipschitz condition $|\gamma_1(t,u) - \gamma_1(t,t)| \leq C|t-u|$, so that the asymptotic distribution of the statistic test $T_{k,n,h} = n\sum_{j=2}^{k}\|\widehat{\boldsymbol{\Gamma}}_{j,h} - \widehat{\boldsymbol{\Gamma}}_{1,h}\|_{\mathcal{F}}^2$ is that given in Corollary 2.

Remark 3 Proportional alternatives of the form $\boldsymbol{\Gamma}_{i,n} = (1 + \rho_i/\sqrt{n})\boldsymbol{\Gamma}_1$ are obtained taking $\Delta_{i,\ell} = \rho_i$ in Proposition 2, so that $\boldsymbol{\Delta}_i = \rho_i\boldsymbol{\Gamma}_1$. In this particular case, we have that

$$\left\langle \boldsymbol{\Delta}^{(k-1)}, v_{i,i} \right\rangle = \sum_{j=2}^{k} < \rho_j \boldsymbol{\Gamma}_1, \phi_i \otimes \phi_i >= \lambda_i \sum_{j=2}^{k} \rho_j$$

and

$$\left\langle \boldsymbol{\Delta}^{(k-1)}, v_{i,j} \right\rangle = \frac{1}{\sqrt{2}} \sum_{j=2}^{k} < \rho_j \boldsymbol{\Gamma}_1, \phi_i \otimes \phi_j + \phi_j \otimes \phi_i >= 0\,,$$

where $\boldsymbol{\Gamma}_1 = \sum \lambda_i \phi_i \otimes \phi_i$. Moreover, if the processes are Gaussian, using Remark 1, we get that the asymptotic distribution given in Theorem 2 can be written as

$$W_k = 2k^2 \sum_{i \geq 1} \lambda_i^2 \left(Z_i + \frac{\sum_{j=2}^{k} \rho_j}{k\sqrt{2}} \right)^2 + 2k^2 \sum_{i \geq 1} \sum_{j \geq 1} \lambda_i \lambda_{i+j} Z_{i,j}^2 \tag{13}$$

and it depends only on the eigenvalues of $\boldsymbol{\Gamma}_1$ different from zero.

Figure 1 contains the theoretical power computed using Monte Carlo for different number of populations and alternatives, when the underlying processes are Brownian motions. In Fig. 1a–c, we choose identical values of ρ_j, i.e., we considered the alternatives $\Gamma_i = (1 + \rho n^{-1/2})\Gamma_1$, for $2 \leq i \leq k$. On the other hand, Figure 1d corresponds to a three population situation in which $\Gamma_i = (1 + \rho_i n^{-1/2})\Gamma_1$, for $2 \leq i \leq 3$, and shows the surface plot of the theoretical power $\pi(\rho_2, \rho_3)$.

To numerically compute the power, we have truncated the statistic W_k defined in (13) as

$$W_k = 2k^2 \sum_{i=1}^{20} \lambda_i^2 \left(Z_i + \frac{\sum_{j=2}^{k} \rho_j}{k\sqrt{2}} \right)^2 + 2k^2 \sum_{1 \leq i < j \leq 20} \lambda_i \lambda_j Z_{i,j}^2\,.$$

The value 20 was chosen since the proportion of explained variance $\sum_{i=1}^{20} \lambda_i / \sum_{i \geq 1} \lambda_i$ is approximately 0.9898. Figure 1a–c displays the theoretical power $\pi(\rho)$ as a function of ρ for different values $\rho \in [0, 10]$ and different number of populations. More precisely, Figure 1a corresponds to $k = 2, 3, 4$, (b) to $k = 5, 6, 7$ and (c) to $k = 8, 9, 10$. The solid lines correspond to $k = 2, 5, 8$, the circles to $k = 3, 6, 9$ and the triangles $k = 4, 7, 10$. On the other hand, Figure 1d corresponds to $k = 3$

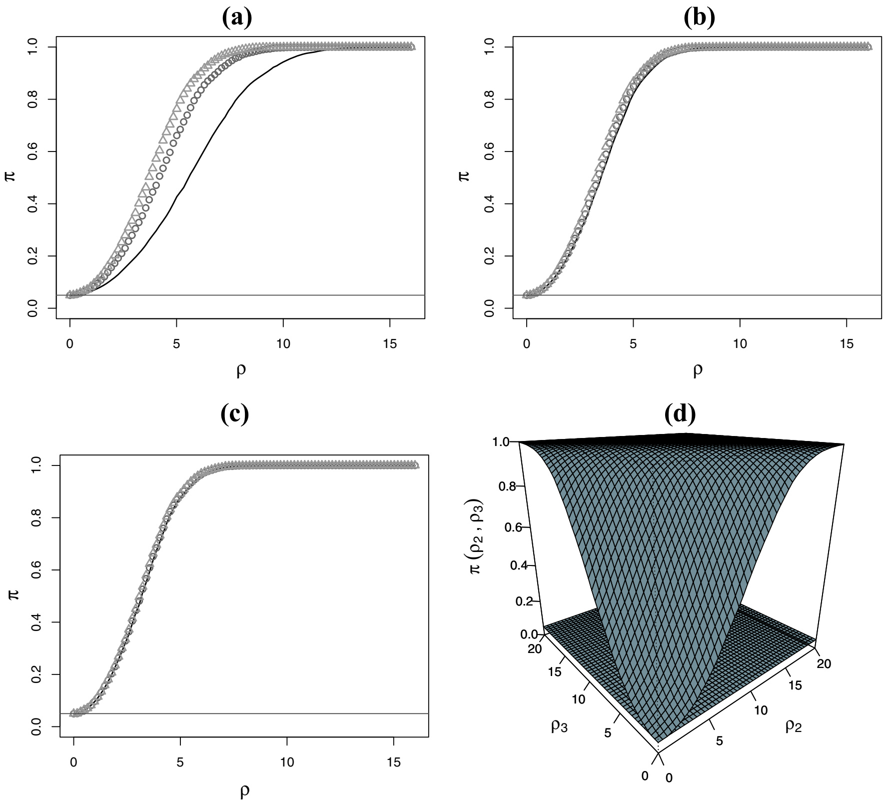

Fig. 1 Theoretical power for proportional Brownian motions. Figures **a** to **c** correspond to the situation $\rho_1 = \cdots = \rho_k = \rho$, where $k = 2$ to 4 in (**a**), $k = 5$ to 7 in (**b**) and $k = 8$ to 10 in (**c**). The *solid lines* correspond to $k = 2, 5, 8$, the *circles* to $k = 3, 6, 9$ and the *triangles* $k = 4, 7, 10$. Figure **d** corresponds to $k = 3$ and $\rho_2, \rho_3 \in [0, 20]$

populations and provides a surface plot for the theoretical power $\pi(\rho_2, \rho_3)$ when $\rho_i \in [0, 20]$ for $i = 2, 3$. The horizontal grey line in (a) to (c) and the horizontal grey plane in (d) correspond to the level 0.05. These plots show that the test improves its performance considerably when $k = 3$ populations are compared instead of two populations. Besides, the power is quite stable for values of k larger than 5 and for the proportional alternatives considered it shows an important detection capability, when $k \geq 4$.

When the stronger condition $\sup_{n\geq 1} \mathbb{E}\|X_i^{(n)}\|^{4+\delta} < \infty$ holds, Theorem 2.1 in Gaines et al. (2011) together with Theorem 2 leads immediately to the asymptotic distribution of test statistic $T_{k,n}$ under root$-n$ local alternatives as stated in Proposition 3.

Proposition 3 *Let $X_{i,1}^{(n)}, \ldots, X_{i,n_i}^{(n)}$ for $i = 1, \ldots, k$ be independent observations from k independent distributions in $\mathcal{H}$, with covariance operators $\boldsymbol{\Gamma}_i$ such that, for $i \geq 1$, $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_{i,n} = \boldsymbol{\Gamma}_1 + n^{-1/2}\boldsymbol{\Delta}_i$, where $\boldsymbol{\Delta}_i$ is a trace operator and $\sup_{n\geq 1} \mathbb{E}\|X_i^{(n)}\|^{4+\delta} < \infty$, with $X_{i,j}^{(n)} \sim X_i^{(n)}$. Assume that, for $i \geq 2$, the covariance operator of $\mathbf{Y}_i^{(n)} =$*

$(X_i^{(n)} - \mathbb{E}(X_i^{(n)})) \otimes (X_i^{(n)} - \mathbb{E}(X_i^{(n)}))$ converges to an operator $\boldsymbol{\Upsilon}_i$ in trace norm. Denote $\boldsymbol{\Upsilon}_w = (\boldsymbol{\Upsilon}_{w,1}, \ldots, \boldsymbol{\Upsilon}_{w,k-1})$ where $\boldsymbol{\Upsilon}_{w,i}$ is defined in (7) and $\boldsymbol{\Upsilon}_1$ is given in (11). Let $\{\upsilon_\ell\}_{\ell \geq 1}$ be the orthonormal eigenfunctions of $\boldsymbol{\Upsilon}_w$ related to the eigenvalues $\{\theta_\ell\}_{\ell \geq 1}$ ordered in decreasing order and $\eta_\ell = \langle \boldsymbol{\Delta}^{(k-1)}, \upsilon_\ell \rangle_{\mathcal{F}^{k-1}}$. Then, if $\widehat{\boldsymbol{\Gamma}}_i$ stands for the sample covariance operator of the i-th population, we have that

$$T_{k,n} = n \sum_{i=2}^{k} \|\widehat{\boldsymbol{\Gamma}}_i - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 \xrightarrow{D} \sum_{\ell \geq 1} \theta_\ell \left(Z_\ell + \frac{\eta_\ell}{\sqrt{\theta_\ell}} \right)^2,$$

where Z_ℓ are independent, $Z_\ell \sim N(0, 1)$.

It is worth noting that if $X_{i,j}$ satisfy (10) and $\mathbb{E}\|W_i\|^{4+\delta} < \infty$ and $\mathbb{E}\|R_i\|^{4+\delta} < \infty$, the proof of Proposition 1 is a consequence of Theorem 2.1 in Gaines et al. (2011). Similarly, if $\Delta_{i,\ell} \geq 0$ and $\mathbb{E}\|X_i^{(1)}\|^{4+\delta} < \infty$, the proof of Proposition 1 can also be derived from Theorem 2.1 in Gaines et al. (2011) through straightforward calculations. However, in both cases, we prefer to avoid imposing higher moment conditions and/or to consider more general alternatives and for that reason we have included their proof in Appendix.

5 Bootstrap calibration

The asymptotic null behaviour derived in Sect. 3 motivates the use of the bootstrap methods given that the asymptotic distribution obtained in (8) depends on the unknown eigenvalues θ_ℓ. For that reason, we will consider a general bootstrap method to approximate the distribution of the test which can be described as follows.

Step 1. For $1 \leq i \leq k$, and given the sample $X_{i,1}, \ldots, X_{i,n_i}$, let $\widehat{\boldsymbol{\Upsilon}}_i$ be consistent estimators of $\boldsymbol{\Upsilon}_i$. Define $\widehat{\boldsymbol{\Upsilon}}_w = (\widehat{\boldsymbol{\Upsilon}}_{w,1}, \ldots, \widehat{\boldsymbol{\Upsilon}}_{w,k-1})$ where

$$\widehat{\boldsymbol{\Upsilon}}_{w,i}(u_1, \ldots, u_{k-1}) = \frac{1}{\widehat{\tau}_{i+1}} \widehat{\boldsymbol{\Upsilon}}_{i+1}(u_1) + \frac{1}{\widehat{\tau}_1} \widehat{\boldsymbol{\Upsilon}}_1 \left(\sum_{i=1}^{k-1} u_i \right),$$

and $\widehat{\tau}_i = n_i / \sum_{s=1}^{k} n_s$. In particular, if $k = 2$, $\widehat{\boldsymbol{\Upsilon}}_w = \widehat{\tau}_1^{-1} \widehat{\boldsymbol{\Upsilon}}_1 + \widehat{\tau}_2^{-1} \widehat{\boldsymbol{\Upsilon}}_2$ with $\widehat{\tau}_i = n_i/(n_1 + n_2)$.
Step 2. For $1 \leq \ell \leq q_n$ denote by $\widehat{\theta}_\ell$ the positive eigenvalues of $\widehat{\boldsymbol{\Upsilon}}_w$.
Step 3. Generate $Z_1^*, \ldots, Z_{q_n}^*$ i.i.d. such that $Z_i^* \sim N(0, 1)$ and let $\mathcal{U}_n^* = \sum_{j=1}^{q_n} \widehat{\theta}_j Z_j^{*2}$.
Step 4. Repeat **Step 3** N_b times, to get N_b values of $\mathcal{U}_{nr}^*$ for $1 \leq r \leq N_b$.

The $(1 - \alpha)-$quantile of the asymptotic null distribution of $T_{k,n}$ can be approximated by the $(1 - \alpha)-$quantile of the empirical distribution of $\mathcal{U}_{nr}^*$ for $1 \leq r \leq N_b$. The $p-$value can be estimated by $\widehat{p} = s/N_b$ where s is the number of $\mathcal{U}_{nr}^*$ which are larger than or equal to the observed value of $T_{k,n}$.

Remark 4 It is worth noticing that this procedure depends only on the asymptotic distribution of $\widehat{\boldsymbol{\Gamma}}_i$. For the sample covariance estimator, the covariance operator $\boldsymbol{\Upsilon}_i$ to be estimated in **Step 1** is given in (1). Assume that all the populations have a Gaussian distribution, then $\boldsymbol{\Upsilon}_i$ can be estimated using the eigenvalues and eigenfunctions of the sample covariance, since f_{ij} are independent and $f_{ij} \sim N(0, 1)$. For non Gaussian samples, $\boldsymbol{\Upsilon}_i$ can be estimated noticing that it is the covariance operator of $\mathbf{Y}_i = (X_i - \mu_i) \otimes (X_i - \mu_i)$. When considering other asymptotically normally estimators of $\boldsymbol{\Gamma}_i$, such as the smoothed estimators $\widehat{\boldsymbol{\Gamma}}_{i,h}$ for $L^2(\mathcal{I})$ trajectories, the estimators need to be adapted.

Taking into account that the space of covariance operators of random elements on $\mathcal{H}$ is a Hilbert space with the inner product defined in $\mathcal{F}$, we have that the covariance of any covariance operator estimator is also an element of a Hilbert space, which we denote as $\mathcal{G}$. Then, for instance, $\boldsymbol{\Upsilon}_i$ and $\widehat{\boldsymbol{\Upsilon}}_i$ in **Step 1** belong to $\mathcal{G}$, while $\widehat{\boldsymbol{\Upsilon}}_w$ and $\boldsymbol{\Upsilon}_w$ are random elements of the product Hilbert space $\mathcal{G}^{k-1}$ with norm denoted as $\|\cdot\|_{\mathcal{G}^{k-1}}$.

The following theorem entails the validity of the bootstrap calibration method. More precisely, let $\widetilde{X}_n = (X_{1,1}, \ldots, X_{1,n_1}, \ldots, X_{k,1}, \ldots, X_{k,n_k})$, Theorem 3 states that the conditional distribution function of $\mathcal{U}_n^*$ given $\widetilde{X}_n$, converges to the asymptotic null distribution of T_n. Denote as $F_{\mathcal{U}_n^*|\widetilde{X}_n}$ the distribution function of $\mathcal{U}_n^*$ given the sample, i.e., $F_{\mathcal{U}_n^*|\widetilde{X}_n}(t) = \mathbb{P}(\mathcal{U}_n^* \leq t|\widetilde{X}_n)$. It is worth noticing that $F_{\mathcal{U}_n^*|\widetilde{X}_n}$ is a sequence of random distribution functions depending on $\widetilde{X}_n$, so $\rho_k(F_{\mathcal{U}_n^*|\widetilde{X}_n}, F_{\mathcal{U}})$ is a sequence of random variables depending only on the given observations.

Theorem 3 *Let q_n such that $q_n/\sqrt{n} \to 0$. Denote by $F_{\mathcal{U}_n^*|\widetilde{X}_n}(\cdot) = \mathbb{P}(\mathcal{U}_n^* \leq \cdot\, |\widetilde{X}_n)$ and by $F_{\mathcal{U}}$ the distribution function of $\mathcal{U} = \sum_{\ell\geq 1} \theta_\ell Z_\ell^2$, where Z_ℓ are i.i.d. and $Z_\ell \sim N(0, 1)$. Assume that $\mathbb{E}(\|X_i\|^4) < \infty$ and $n_i/n \to \tau_i$ with $\tau_i \in (0, 1)$ and $n = \sum_{i=1}^k n_i$. Then, if*

$$\sqrt{n}\|\widehat{\boldsymbol{\Upsilon}}_w - \boldsymbol{\Upsilon}_w\|_{\mathcal{G}^{k-1}} = O_{\mathbb{P}}(1)\,, \tag{14}$$

we have that $\rho_k(F_{\mathcal{U}_n^*|\widetilde{X}_n}, F_{\mathcal{U}}) \xrightarrow{p} 0$, where $\rho_k(F, G)$ stands for the Kolmogorov distance between the distribution functions F and G.

Remark 5 From Theorem 3, we have that the asymptotic significance level of the test based on the bootstrap critical value is indeed α, since the $(1-\alpha)-$quantile of the limiting null distribution of $T_{k,n}$ is well approximated by the $(1-\alpha)-$quantile of the empirical distribution of $\mathcal{U}_{nr}^*$ for $1 \leq r \leq N_b$. It is worth noticing that Theorem 3 holds whenever $\sqrt{n}\|\widehat{\boldsymbol{\Upsilon}}_w - \boldsymbol{\Upsilon}_w\|_{\mathcal{G}^{k-1}} = O_{\mathbb{P}}(1)$, regardless of whether the null hypothesis is true or not. For instance, (14) holds under the sequence of local alternatives considered in (10) or (12), when taking $\widehat{\boldsymbol{\Upsilon}}_i$ the sample covariance operator of $\mathbf{Y}_i = (X_i - \mu_i) \otimes (X_i - \mu_i)$, so the power of bootstrap test mimic that of the infeasible test constructed from the asymptotic null distribution given in Sect. 3. Furthermore, under (14), the bootstrap procedure leads to a consistent test.

6 Simulation study

This section contains the results of two simulation studies carried on with $k = 2$ and $k = 3$ populations and designed to illustrate the finite-sample performance of the test procedure described in Sect. 5, under the null hypothesis and under different alternatives. In all scenarios, we generate $NR = 1000$ samples of size n_i, $1 \leq i \leq k$ and each trajectory was observed at $m = 100$ equidistant points in the interval $[0, 1]$. To analyse the dependence on the sample size, we choose $n_i = 50$, 100 and 200, for $1 \leq i \leq k$ which allows to study the behaviour of the test in terms of level approximation as well as power performance depending on the sample size. To summarize the test performance, we compute the observed frequency of rejections over replications with nominal level $\alpha = 0.05$.

6.1 Simulation settings

Under the null hypothesis, we consider infinite-dimensional processes generating independent centred Brownian motion processes, denoted from now on as $\mathcal{BW}(0, 1)$. On the other hand, to check the test power performance, we consider root$-n$ local alternatives. To be more precise, when comparing two populations, we generate independent observations $X_{1,j} \sim X_1$, $1 \leq j \leq n_1$, and $X_{2,j} \sim X_2$, $1 \leq j \leq n_2$, such that $X_1 \sim \mathcal{BW}(0, 1)$ and $X_2 \sim W_1 + \delta_n W_2^2$, where W_1 and W_2 are independent $W_i \sim \mathcal{BW}(0, 1)$, $i = 1, 2$ and $\delta_n = \rho n^{-1/4}$ with $n = n_1 + n_2$. The situation $\rho = 0$ corresponds to the null hypothesis, while to study the test power ρ takes values from 1 to 10. The set of alternatives considered corresponds to the local alternatives studied in Proposition 1 since the covariance operator of $X_{2,1}$ equals $\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}_1 + \rho^2 n^{-1/2} \boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is the covariance operator of W_2^2.

On the other hand, for the three populations case, we consider a proportional model taking independent observations $X_{i,j} \sim X_i$, $1 \leq j \leq n_i$, $1 \leq i \leq k$, such that $X_1 \sim \mathcal{BW}(0, 1)$, while $X_i \sim (1 + \delta_n)^{1/2} \mathcal{BW}(0, 1)$, for $i = 2, 3$, where $\delta_n = \rho n^{-1/2}$ with $n = \sum_{i=1}^{3} n_i$. The parameter ρ takes values on an equidistant grid of points between 0 and 20 of size 11. In this case, the covariance operators of X_2 and X_3 equal $\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}_3 = (1 + \rho n^{-1/2}) \boldsymbol{\Gamma}_1$ corresponding to the proportional alternatives described in Remark 3.

6.2 Testing procedures

We study the behaviour of the test based on $T_{k,n}$ defined in (6) using the bootstrap calibration described in Sect. 5 with $N_b = 5000$ bootstrap replications. To perform the bootstrap calibration, we project the centred data onto the M largest principal components of the pooled sample covariance matrix $n^{-1} \sum n_i \widehat{\boldsymbol{\Gamma}}_i$. We then estimate the covariance operator $\widehat{\boldsymbol{\Upsilon}}_w$ through a finite-dimensional matrix. To evaluate the dependence on the number of principal components chosen, we select $M = 3, 10, 20$ and 30. In this situation, the value q_n used in **Step 2** equals $q_n = M(M + 1)/2$. The percentage of total variance explained by the selected number of principal components is reported in Table 1, while the fre-

Table 1 Percentage of the total variance explained by the first M principal components

k	ρ	$n_i = 50$ M				$n_i = 100$ M				$n_i = 200$ M			
		3	10	20	30	3	10	20	30	3	10	20	30
2	0	0.935	0.982	0.993	0.996	0.934	0.981	0.991	0.995	0.934	0.980	0.991	0.994
3	0	0.934	0.981	0.991	0.995	0.934	0.980	0.991	0.995	0.934	0.981	0.991	0.994

quencies of rejection at the 5% level, for $k = 2$ and $k = 3$, are given in Figs. 2 and 3, respectively. The frequencies of rejection corresponding to $M = 3, 10, 20$ and 30 are given in circles, upper, lower triangles and squares, respectively. The obtained values for the rejection frequencies can be seen in Boente et al. (2014).

Taking into account the fact that, under the null hypothesis, the processes are Gaussian, Remark 1.d) entails that $\theta_{i,i} = 2k^2\lambda_i^2$ and $\theta_{i,j} = 2k^2\lambda_i\lambda_j$, for $i < j$. Then, from the eigenvalues $\widehat{\lambda}_\ell$ of the pooled sample covariance matrix, one may easily provide estimators $\widehat{\theta}_j$ of θ_j to replace those considered in **Step 2**. This approximation, referred as *Gaussian*, is plotted with dashed lines in Figs. 2 and 3 and was computed using the

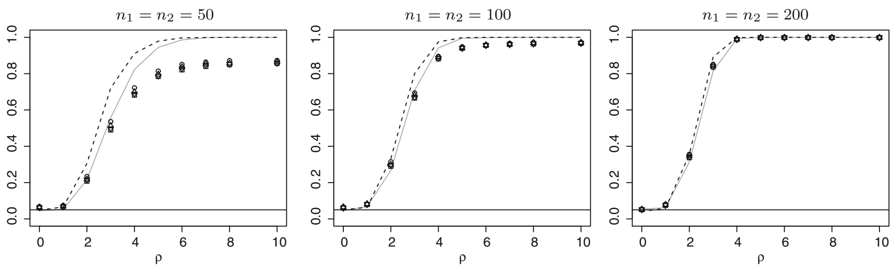

Fig. 2 Frequency of rejection when $k = 2$ for the bootstrap test $\Phi_{b,M}$, $\Phi_{b,g}$, and $\Phi_{p,5000}$. The *solid* and *dashed lines* correspond to $\Phi_{p,5000}$ and $\Phi_{b,g}$, respectively, while the circles, upper, lower triangles and the square correspond to $\Phi_{b,M}$, with $M = 3$, 10, 20 and 30, respectively

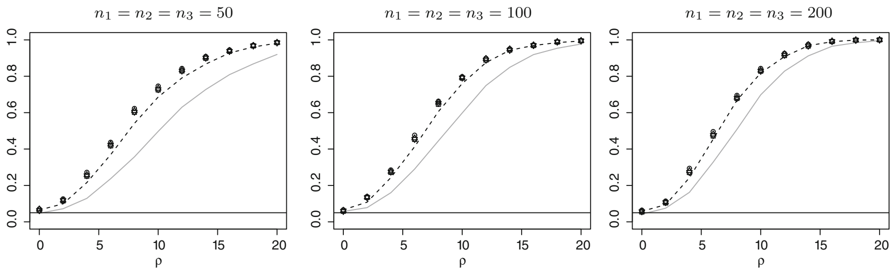

Fig. 3 Frequency of rejection when $k = 3$ for the bootstrap test $\Phi_{b,M}$, $\Phi_{b,g}$, and $\Phi_{p,5000}$. The *solid* and *dashed lines* correspond to $\Phi_{p,5000}$ and $\Phi_{b,g}$, respectively, while the *circles*, *upper*, *lower triangles* and the *square* correspond to $\Phi_{b,M}$, with $M = 3$, 10, 20 and 30, respectively

fact that the trajectories were generated over a grid of 100 points for all the sample sizes leading to at most 100 non-null values $\widehat{\lambda}_\ell$.

We also compare the behaviour of our test statistic with the permutation test introduced in Pigoli et al. (2014) when $k = 2$. Our choice for the permutation test is based on the numerical study reported in Pigoli et al. (2014), where it is shown that the permutation test provides a good competitor to the tests introduced in Panaretos et al. (2010) and Fremdt et al. (2013). We perform the permutation test taking the same discrepancy measure between covariance operators used for $T_{k,n}$, i.e., $d(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) = \|\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2\|_{\mathcal{F}}$. The results obtained when using $N_p = 5000$ random permutations are given in solid lines in Fig. 2, while those corresponding to $N_p = 1000$ are available in Boente et al. (2014). In the case of $k = 3$ populations, a permutation test was also considered taking $D = d(\widehat{\boldsymbol{\Gamma}}_1, \widehat{\boldsymbol{\Gamma}}_2, \widehat{\boldsymbol{\Gamma}}_3) = \|\widehat{\boldsymbol{\Gamma}}_2 - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 + \|\widehat{\boldsymbol{\Gamma}}_3 - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 + \|\widehat{\boldsymbol{\Gamma}}_3 - \widehat{\boldsymbol{\Gamma}}_2\|_{\mathcal{F}}^2$ as test statistic. As in Pigoli et al. (2014), we first centre the samples using the sample mean and then, we consider N_p random permutations of the labels 1, 2, 3 on the centred sample curves. For each permutation j, we compute $D_j = d(\widehat{\boldsymbol{\Gamma}}_1^{(j)}, \widehat{\boldsymbol{\Gamma}}_2^{(j)}, \widehat{\boldsymbol{\Gamma}}_3^{(j)})$, for $j = 1, \ldots, N_p$, where $\widehat{\boldsymbol{\Gamma}}_i^{(j)}$ is the sample covariance operator of the group indexed with label i in the given permutation. As in the two populations case, the $p-$value of the test is the proportion of D_j greater than or equal to D.

We also used this approach taking as test statistic $D^\star = d(\widehat{\boldsymbol{\Gamma}}_1, \widehat{\boldsymbol{\Gamma}}_2, \widehat{\boldsymbol{\Gamma}}_3) = \|\widehat{\boldsymbol{\Gamma}}_2 - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 + \|\widehat{\boldsymbol{\Gamma}}_3 - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2$, which corresponds to $T_{k,n}$, but is not invariant by permutation of the labels. The results for $D^\star$ are similar to those obtained for D and are not reported here.

From now on, we denote as $\Phi_{b,M}$, for $M = 3$, 10, 20 and 30 the bootstrap calibration of $T_{k,n}$ computed using M principal components, $\Phi_{b,g}$ the bootstrap calibration of $T_{k,n}$ computed using the Gaussian approximation for $\theta_{i,j}$ and Φ_{p,N_p} the permutation test computed using N_p random permutations.

6.3 Simulation results

Regarding the bootstrap calibration described in Sect. 5 for the test based on $T_{k,n}$, Figures 2 and 3, as well as the results reported in Boente et al. (2014), show the improvement attained in level when the Gaussian approximation is used, both for $k = 2$ and $k = 3$ populations. Also, when we project the data on the first M principal components, the empirical size of the test based on the bootstrap calibration is quite close to the nominal one. To analyse the significance of the empirical size, we study if it is significantly different from the nominal level $\alpha = 0.05$. To be more precise, for a test Φ_n based on a statistic T_n, let π be such that $\pi_{H_0}(\Phi_n) \xrightarrow{p} \pi$. Then, using the central limit theorem, the hypothesis $H_{0,\pi} : \pi = \alpha$ is rejected at level γ versus $H_{1,\pi} : \pi \neq \alpha$ if $\pi_{H_0}(\Phi_n) \notin [a_1(\alpha), a_2(\alpha)]$ where $a_j(\alpha) = \alpha + (-1)^j z_{\gamma/2} \{\alpha(1-\alpha)/NR\}^{1/2}$, $j = 1, 2$. If $H_{0,\pi} : \pi = \alpha = 0.05$ is not rejected, the testing procedure based on T_n is considered accurate, whereas if $\pi_{H_0}(\Phi_n) < a_1(\alpha)$ the testing procedure is conservative and when $\pi_{H_0}(\Phi_n) > a_2(\alpha)$ the test is liberal. In all the considered scenarios for $k = 2$, the test is accurate with significance level $\gamma = 0.01$. On the other hand, for $k = 3$ populations, the test is liberal when $n_1 = n_2 = n_3 = 50$ and $M = 3$

or 10, being accurate in all other scenarios. Hence, in almost all considered situations the proposed method has a good level performance.

Regarding the power behaviour, the bootstrap test detects the considered alternatives for different values of M and also when using the Gaussian approximation to the eigenvalues $\theta_{i,j}$. As expected, the observed frequencies of rejection converge to 1 as ρ increases. Since local alternatives are taken, the power is almost similar for all choices of sample sizes and shows the test's capability to detect the selected local alternatives. However, it is worth noticing that the test shows a slower power convergence for $k = 2$ and $n_1 = n_2 = 50$.

Figures 2 and 3 also show that the permutation test is an accurate test for both $k = 2$ and $k = 3$. When comparing the power of the permutation test and the bootstrap calibration, we notice that both tests lead to similar results. However, the permutation test has a better power performance for $k = 2$ when large values of ρ and small sample sizes are combined. On the contrary, for $k = 3$ populations a better power is attained with the bootstrap calibration.

To help in the effective comparison of the power performance of the two tests, we compute the size-corrected relative exact powers $\rho_{H_1}(\Phi_{b,M}, \Phi_{p,N_p})$ and $\rho_{H_1}(\Phi_{b,g}, \Phi_{p,N_p})$, where as mentioned before, $\Phi_{b,M}$ stands for the bootstrap calibration of $T_{k,n}$ computed using M principal components, $\Phi_{b,g}$ denotes the bootstrap calibration of $T_{k,n}$ computed using the Gaussian approximation, and Φ_{p,N_p} is the permutation test computed using N_p random permutations. For two test ϕ_1 and ϕ_2 and an alternative H_1, the size-corrected relative exact power $\rho_{H_1}(\phi_1, \phi_2)$ was defined in Morales et al. (2004) as

$$\rho_{H_1}(\phi_1, \phi_2) = \left(\frac{D_{H_1}(\phi_1)}{D_{H_1}(\phi_2)} - 1 \right) \times 100 \,,$$

with $D_{H_1}(\phi) = \pi_{H_1}(\phi) - \pi_{H_0}(\phi)$, where $\pi_{H_1}(\phi)$ and $\pi_{H_0}(\phi)$ denote the power of the test ϕ under H_1 and the null hypothesis, respectively. This measure allows to clarify the fluctuations in the powers which are harder to observe in Figs. 2 and 3, since large negative values of $\rho_{H_1}(\phi_1, \phi_2)$ indicate that ϕ_2 outperforms ϕ_1, while large positive values show that ϕ_1 is preferable.

Figure 4 gives the plots of $\rho_{H_1}(\Phi_{b,M}, \Phi_{p,N_p})$ and $\rho_{H_1}(\Phi_{b,g}, \Phi_{p,N_p})$, for two populations, when the permutation test Φ_{p,N_p} is computed with $N_p = 5000$ random permutations. As expected the test obtained using the Gaussian approximation outperforms the permutation test specially for local alternatives close to the null hypothesis. On the other hand, the permutation test shows its advantage for $n_1 = n_2 = 50$, in particular when $\rho = 1$, since the asymptotic approximation leads to some loss of power in the bootstrap test. The better performance for $\rho = 1$ is also observed when $n_1 = n_2 = 100$, while for $n_1 = n_2 = 200$ the test defined in Sect. 5 is much better than the permutation test. In general, the bootstrap test shows its advantage, for large sample sizes. The worse behaviour of the permutation test for large samples may be due to the fact that the number of random permutations should increase with the sample size.

When considering $k = 3$ populations, Fig. 5 shows the size-corrected values $\rho_{H_1}(\Phi_{b,M}, \Phi_{p,N_p})$ and $\rho_{H_1}(\Phi_{b,g}, \Phi_{p,N_p})$ when $N_p = 5000$. In this setting, the boot-

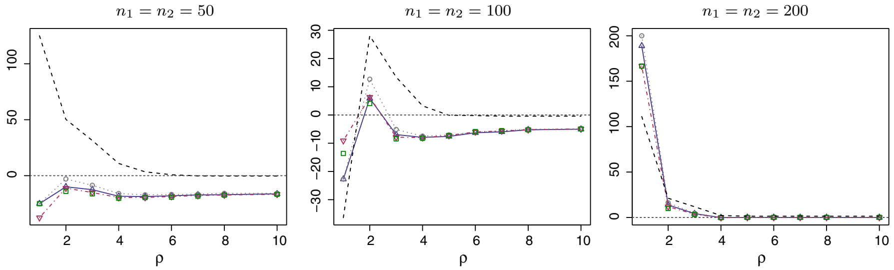

Fig. 4 Size-corrected relative exact powers, $\rho_{H_1}(\Phi_{b,M}, \Phi_{p,N_p})$ and $\rho_{H_1}(\Phi_{b,g}, \Phi_{p,N_p})$, for the bootstrap tests $\Phi_{b,M}$ ($M = 3$, 10, 20 and 30) and $\Phi_{b,g}$ (in *dashed black lines*) with respect to the permutation test Φ_{p,N_p} with $N_p = 5000$ random permutations, for $k = 2$. The lines with the circles (in *grey*), upper (in *blue*), lower triangles (in *maroon*) and the square (in *green*) correspond to $\Phi_{b,M}$, with $M = 3$, 10, 20 and 30, respectively

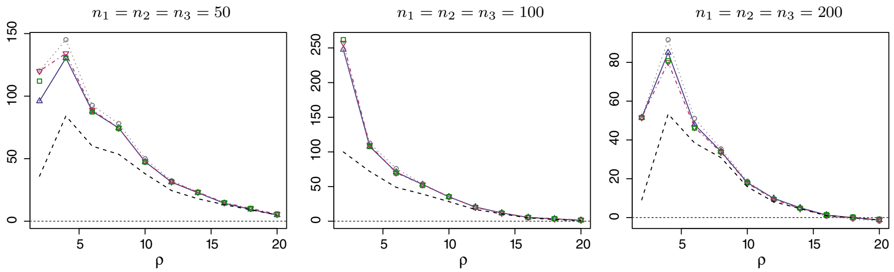

Fig. 5 Size-corrected relative exact powers, $\rho_{H_1}(\Phi_{b,M}, \Phi_{p,N_p})$ and $\rho_{H_1}(\Phi_{b,g}, \Phi_{p,N_p})$, for the bootstrap tests $\Phi_{b,M}$ ($M = 3$, 10, 20 and 30) and $\Phi_{b,g}$ (in dashed black lines) with respect to the permutation test Φ_{p,N_p} based on $D = \|\widehat{\boldsymbol{\Gamma}}_2 - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 + \|\widehat{\boldsymbol{\Gamma}}_3 - \widehat{\boldsymbol{\Gamma}}_1\|_{\mathcal{F}}^2 + \|\widehat{\boldsymbol{\Gamma}}_3 - \widehat{\boldsymbol{\Gamma}}_2\|_{\mathcal{F}}^2$ with $N_p = 5000$ random permutations when $k = 3$. The lines with the circles (in *grey*), upper (in *blue*), lower triangles (in *maroon*) and the square (in green) correspond to $\Phi_{b,M}$, with $M = 3$, 10, 20 and 30, respectively

strap calibration test always outperforms the permutation test, specially for alternatives close to the null hypothesis. The better performance may be explained by the fact that the asymptotic behaviour of the tests and so its bootstrap calibration detects alternatives following a proportional model more easily than those considered in the two population case. The higher capability of $\Phi_{b,M}$ to detect proportional local alternatives for three populations is related to power performance described in Remark 3. Furthermore, the obtained results suggest that as the number of populations increases the number of permutations needed to attain a good power performance also needs to be increased considerably leading to a larger computing time.

Although a formal computational analysis of the different test statistics is beyond the scope of this paper, we tested the speed of our R codes using an Intel i7-2600K CPU (3.4 GHz) machine running Windows 7. Table 2 reports the average time in CPU seconds of the different test procedures computed over 10 random samples generated as in the simulation settings under H_0 and for the sample sizes n_i considered before. The results obtained show that the computing time increases linearly as the number of permutations increases and in all situations Φ_{p,N_p} is much more time expensive than $\Phi_{b,M}$. On the other hand, as expected, the number M of principal components

Table 2 Average timing (in seconds) of the test procedures

	$k=2$ $n_i=50$	$n_i=100$	$n_i=200$	$k=3$ $n_i=50$	$n_i=100$	$n_i=200$
$\Phi_{b,3}$	0.053	0.059	0.090	0.055	0.072	0.114
$\Phi_{b,10}$	0.125	0.120	0.151	0.164	0.173	0.215
$\Phi_{b,20}$	0.334	0.309	0.367	0.693	0.693	0.828
$\Phi_{b,30}$	0.867	0.906	1.069	3.510	3.580	3.822
$\Phi_{b,g}$	3.424	3.363	3.379	3.317	3.315	4.413
$\Phi_{p,5000}$	5.831	9.493	17.825	15.544	25.957	47.575

used considerably increases the computation time. However, the computing time of $\Phi_{b,M}$ is quite stable through sample sizes, for a fixed number of populations and a fixed M. The Gaussian approximation takes almost the same computing time in all the considered scenarios and shows a larger average time than $\Phi_{b,M}$, except when $M=30$ and $k=3$, in which they both give similar average timings.

From the obtained results, we see that our procedure is, in terms of level and power behaviour, a good competitor for the permutation test introduced for two populations in Pigoli et al. (2014). On the other hand, when $k=3$ it has a better detection capability with a much lower computing time. Furthermore, our method has the advantage of allowing to develop a theory regarding its asymptotic behaviour as described in Sects. 4 and 5.

7 Real data example

The following real data set corresponds to an example of speech recognition. This data set was introduced in Hastie et al. (1995). A completed description and the data are available at http://www-stat.stanford.edu/ElemStatLearn. The data consist of 2000 log-periodograms of length 150 divided in five groups according to five speech frames. The classes correspond to five phonemes transcribed as “sh” as in “she”(group 1); “iy” as in “she”(group 2); “dcl” as in “dark”(group 3); “aa” as the vowel in “dark”(group 4); “ao” as the first vowel in “water”(group 5).

Functional data analysis has been used previously to study linguistic and phonetic problems which are beyond speech recognition. As mentioned in Pigoli et al. (2014), comparison of covariance operators has its own interest in this setting, since they provide a good characterization of language. For that reason, we compare the covariance operators of the five phoneme data sets using the bootstrap test described in Sect. 5 with 5000 bootstrap replications. Different values of M, the number of principal components of the pooled estimated covariance operator, are considered. As shown in Table 3, a large number of components is needed to explain an 85% of the total variance. When testing equality between five groups of phonemes, the obtained $p-$values are reported in Table 3. In all cases, the results are consistent for all choices of M, rejecting the hypothesis of equality among the covariance operator between the phoneme groups.

Table 3 p-values for the bootstrap calibrated test to compare the five class of phonemes and percentage of variance explained for different values of the number M of components

	4	8	16	24	32	40	48	56	64
$p-$value	0	0	0	0	0	0	0	0	0
% variance	0.4812	0.5697	0.6674	0.7255	0.7480	0.7856	0.8170	0.8437	0.8665

As shown in our simulation results, the test is quite stable with respect to the selection of M. However, in a practical analysis, a key point is the selection of the number of components M. Usually, the value of M is chosen so that the M principal components explain a given amount of the total variability. Nevertheless, it may be of interest to study the dependence of the obtained results on the number of components and a global $p-$value may be obtained controlling the False Discovery Rate (FDR). Recall that if M different statistical hypotheses are tested, then the FDR is the expected proportion of wrongly rejected hypotheses. In our situation, all the hypotheses are equal; hence, the FDR coincides with the level of the procedure. As in Cuesta-Albertos and Febrero-Bande (2010), to provide a corrected $p-$value, one may proceed as follows. Assume that we select ℓ values, $M_1, \ldots, M_\ell$, for the number of components M and for each value M_j, $1 \leq j \leq \ell$, we perform the test obtaining a $p-$value p_j. Denote $p^{(1)} \leq \cdots \leq p^{(\ell)}$ the ordered $p-$values. The results in Benjamini and Yekutieli (2001) allow us to reject the null hypothesis for every level α such that $\alpha \geq \inf_{1 \leq i \leq \ell} \ell\, p^{(i)}/i$. Therefore, the corrected $p-$value may be taken as the quantity $\inf_{1 \leq i \leq \ell} \ell\, p^{(i)}/i$. Even though this procedure is conservative, it is less conservative than the Bonferroni one. It is clear from Table 3 that, in this example, the corrected $p-$value equals any of the obtained $p-$values for the different dimension choices.

As mentioned in Sect. 6, a possible competitor of the bootstrap test is the permutation test based on $\sum_{j<s} \|\widehat{\boldsymbol{\Gamma}}_s - \widehat{\boldsymbol{\Gamma}}_j\|_{\mathcal{F}}^2$, which is the extension of the procedure described in Pigoli et al. (2014) to several populations. The test introduced in Fremdt et al. (2013) cannot be applied in this setting, since it allows comparisons only between two covariance operators. Its extension to the several population case is beyond the scope of this paper and for that reason, we only considered the permutation test which leads to a $p-$value of 0.000. As mentioned in Sect. 6.3, the permutation test is computationally more expensive than the proposed bootstrap procedure, in particular, when dealing with more than two populations. However, its main disadvantage is that it relies on the samples exchangeability under the null hypothesis which means that, in particular, all the populations have the same underlying distribution except for changes in their mean and covariance operators. To explore the validity of this assumption, we projected the samples over a random direction u generated according to a Brownian motion $\mathcal{BW}(0, 1)$. The random variables $\langle X_{i,j}, u\rangle$. $1 \leq i \leq 5$, $1 \leq j \leq 400$, are then centred and standardized within each group, leading to samples $\mathcal{Y}_i = \{Y_{i,j} : 1 \leq j \leq 400\}$, $1 \leq i \leq 5$. If the five phoneme samples have the same underlying distribution except for changes in their means and covariance operators, the samples $\mathcal{Y}_i$ should have approximately the same distribution. Figure 6

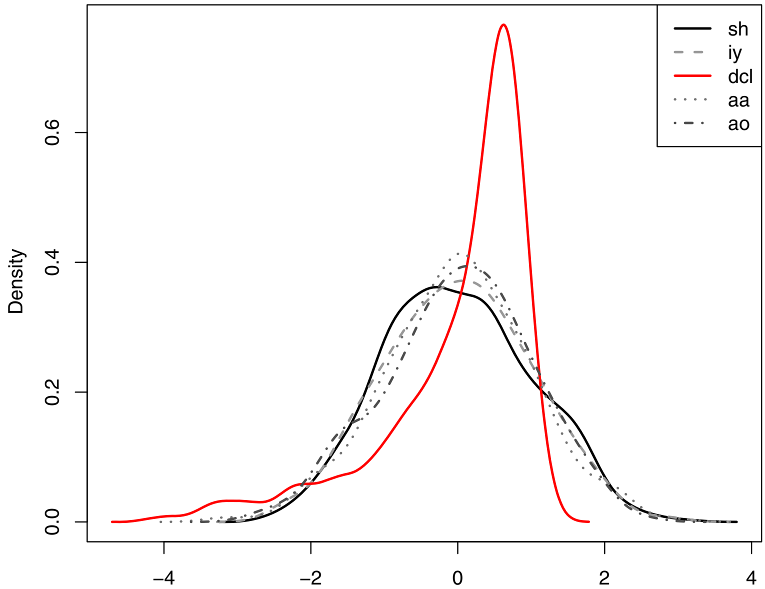

Fig. 6 Estimators of the density of a random projection of the phoneme data. The projected data are first centred and standardized within each group before computing the kernel density estimator

shows the obtained density estimators computed using a Gaussian kernel and Silverman's *rule of thumb* bandwidth, see Silverman (1986). The plot shows that the density estimator corresponding to the data group 3 and related to the phoneme "dcl" has a different shape that those of the other groups. Indeed, its density is strongly skewed showing a long left tail, while that of group 1, corresponding to the phoneme "sh" is slightly asymmetric to the right, the other ones being more symmetric. It is also worth noticing that the MAD, i.e., the normalized median of the absolute deviation with respect to the median of $\mathcal{Y}_i$ is close to 1 for all groups except when $i = 3$, in which case it is equal to 0.578, which also suggests that the samples have a different underlying distribution. To test if the group labelled as 3 has the same distribution as the other ones up to changes in means and covariance operators, we have compared, for each $i \neq 3$, the distribution of the standardized projected data $\mathcal{Y}_3$ with that of $\mathcal{Y}_i$ using the Kolmogorov–Smirnov test statistic. As expected, the Kolmogorov–Smirnov test rejects the null hypothesis for each $i \neq 3$. The obtained $p-$values $(7 \times 10^{-6}, 5 \times 10^{-5}, 7 \times 10^{-6}, 0.00013)$ reject with level 0.001 the null hypothesis of equality between the distribution of the data related to the phoneme "dcl"and the other phonemes, except for changes in their means and covariance operators. These results suggest that in this case, the permutation test may not be an appropriate procedure. The bootstrap procedure introduced in this paper is more robust in the sense that it does not require normality of the samples or the same underlying probability measure, up to mean and covariance operator changes, among populations. It only assumes $\mathbb{E}(\|X_i\|^4) < \infty$ to guarantee the asymptotic normality of the sample covariance operators.

8 Conclusions

In this paper, we have studied a procedure to test equality among several populations covariance operators. The test statistic is based on the Hilbert–Schmidt distance between consistent estimators of $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Gamma}_1$, for $2 \leq i \leq k$. The analysis of the asymptotic distribution of the test statistic reveals that the testing procedure is consistent against local alternatives converging to the null hypothesis at rate $n^{-1/2}$ when the sample covariance operators are used. These results also hold for the smoothed covariance operators defined in Boente and Fraiman (2000), under mild conditions. The asymptotic null behaviour obtained motivates the use of bootstrap methods, since it depends on the eigenvalues of an unknown operator. For that reason, we also provide a general bootstrap calibration method whose validity is derived. Our numerical studies have shown that the bootstrap calibration has a good practical behaviour and is a good competitor for the permutation test defined in Pigoli et al. (2014) for two populations and the considered alternatives. On the other hand, when $k = 3$ and for proportional alternatives considered, the bootstrap test has shown a better detection capability. Another advantage of the bootstrap test over the permutation test is its lower computing time for the sample sizes considered.

Acknowledgements The authors wish to thank the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper. This research was partially supported by Grants PIP 112-201101-00339 and 112-201101-00742 from CONICET, PICT 2014-0351 and 2012-1641 from ANPCYT and 20020130100279BA and 20020120200244BA from the Universidad de Buenos Aires at Buenos Aires, Argentina.

Appendix

Proof of Theorem 1. Denote as $\mathcal{F}^k = \mathcal{F} \times \cdots \times \mathcal{F}$ the $k-$th dimensional product space of identical copies of $\mathcal{F}$ and consider the process $\mathbf{V}_{k,n} = \left(\sqrt{n}(\widetilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{\Gamma}_1), \ldots, \sqrt{n}(\widetilde{\boldsymbol{\Gamma}}_k - \boldsymbol{\Gamma}_k)\right)^t$. Using that $\sqrt{n_i}\left(\widetilde{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i\right) \xrightarrow{D} \mathbf{U}_i$, the independence of the estimated operators and the fact that $n_i/n \to \tau_i \in (0, 1)$, we get that $\mathbf{V}_{k,n} \xrightarrow{D} \mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_k)^t$, where $\mathbf{V}_i = \tau_i^{-1/2}\mathbf{U}_i$ are independent random processes of $\mathcal{F}$ with covariance operators ${\tau_i}^{-1}\boldsymbol{\Upsilon}_i$. Hence, $\mathbf{V}_{k,n}$ converges in distribution to a zero mean Gaussian random element $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_k)^t \in \mathcal{F}^k$ with covariance operator $\widetilde{\boldsymbol{\Upsilon}} = \text{DIAG}\left({\tau_1}^{-1}\boldsymbol{\Upsilon}_1, \ldots, {\tau_k}^{-1}\boldsymbol{\Upsilon}_k\right)$.

Let $A : \mathcal{F}^k \to \mathcal{F}^{k-1}$ be the linear operator defined as $A(V_1, \ldots, V_k) = (V_2 - V_1, \ldots, V_k - V_1)$ and denote as $A^* : \mathcal{F}^{k-1} \to \mathcal{F}^k$ its adjoint operator. The continuous map theorem guarantees that $A\mathbf{V}_{k,n} \xrightarrow{D} \mathbf{W}$, where $\mathbf{W} = (W_1, \ldots, W_{k-1})^t = A\mathbf{V}$ is a zero mean Gaussian random element of $\mathcal{F}^{k-1}$ with covariance operator $\boldsymbol{\Upsilon}_w = A\widetilde{\boldsymbol{\Upsilon}} A^*$. Moreover, we also obtain that $n \sum_{j=2}^k \|(\widetilde{\boldsymbol{\Gamma}}_j - \boldsymbol{\Gamma}_j) - (\widetilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{\Gamma}_1)\|_{\mathcal{F}}^2 \xrightarrow{D} \sum_{j=1}^{k-1} \|W_j\|_{\mathcal{F}}^2 = \|\mathbf{W}\|_{\mathcal{F}^{k-1}}^2$. Let $\upsilon_\ell \in \mathcal{F}^{k-1}$ be the orthonormal eigenfunctions of $\boldsymbol{\Upsilon}_w$ related to the eigenvalues θ_ℓ ordered in decreasing order. Since $\mathbf{W}$ is a zero mean Gaussian random element of $\mathcal{F}^{k-1}$ with covariance operator $\boldsymbol{\Upsilon}_w$, $\mathbf{W}$ can be written as $\sum_{\ell \geq 1} \theta_\ell^{1/2} Z_\ell\, \upsilon_\ell$ where Z_ℓ are i.i.d. random variables such that $Z_\ell \sim N(0, 1)$. Hence, $\|\mathbf{W}\|_{\mathcal{F}^{k-1}}^2 = \sum_{\ell \geq 1} \theta_\ell Z_\ell^2$, which leads to the desired result.

It only remains to show (7). Straightforward calculations allow to show that the adjoint operator $A^* : \mathcal{F}^{k-1} \to \mathcal{F}^k$ is given by $A^*(w_1, \ldots, w_{k-1}) = (-\sum_{i=1}^{k-1} w_i, w_1, \ldots, w_{k-1})$. Hence, as $\mathbf{U}_1, \ldots, \mathbf{U}_k$ are independent, we obtain that

$$\begin{aligned}
\boldsymbol{\Upsilon}_w(w_1, \ldots, w_{k-1}) &= (A\widetilde{\boldsymbol{\Upsilon}}\, A^*)(w_1, \ldots, w_{k-1}) \\
&= \left(\frac{1}{\tau_2}\boldsymbol{\Upsilon}_2(w_1) + \frac{1}{\tau_1}\boldsymbol{\Upsilon}_1\left(\sum_{i=1}^{k-1} w_i\right), \ldots, \frac{1}{\tau_k}\boldsymbol{\Upsilon}_k(w_{k-1})\right. \\
&\quad \left. + \frac{1}{\tau_1}\boldsymbol{\Upsilon}_1\left(\sum_{i=1}^{k-1} w_i\right)\right),
\end{aligned}$$

concluding the proof. □

Proof of Corollary 1. Consider the process $\mathbf{U}_{i,n_i} = \sqrt{n_i}(\widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i)$. The independence of the samples and among populations together with the results stated in Dauxois et al. (1982), allow to show that $\mathbf{U}_{i,n_i}$ are independent and converge in distribution to independent zero mean Gaussian random elements $\mathbf{U}_i$ of $\mathcal{F}$ with covariance operator $\boldsymbol{\Upsilon}_i$ defined in (1). The result follows now from Theorem 1. □

Proof of Theorem 2. Using that $n_i/n \to \tau_i$, we get immediately that $\sqrt{n}\left(\widetilde{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1\right) \xrightarrow{D} \boldsymbol{\Delta}_i + (1/\sqrt{\tau_i})\mathbf{U}_i$ where $\mathbf{U}_i$ is a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_i$ and for $i = 1$, $\boldsymbol{\Delta}_1 = \mathbf{O}$ is the null operator. The fact that the estimators are independent implies that $\mathbf{U}_i$ can be chosen to be independent so, as in the proof of Theorem 1, we have that $\mathbf{V}_{k,n} = \left(\sqrt{n}(\widetilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{\Gamma}_1), \ldots, \sqrt{n}(\widetilde{\boldsymbol{\Gamma}}_k - \boldsymbol{\Gamma}_1)\right)^t \xrightarrow{D} \mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_k)^t$, where $\mathbf{V}_i = \boldsymbol{\Delta}_i + (1/\sqrt{\tau_i})\mathbf{U}_i$ are independent random processes of $\mathcal{F}$ with mean $\boldsymbol{\Delta}_i$ and covariance operators ${\tau_i}^{-1}\boldsymbol{\Upsilon}_i$. Hence, $\mathbf{V}_{k,n}$ converges in distribution to a Gaussian random element $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_k)^t \in \mathcal{F}^k$ with mean $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_k)^t$ and covariance operator $\widetilde{\boldsymbol{\Upsilon}} = \text{DIAG}\left({\tau_1}^{-1}\boldsymbol{\Upsilon}_1, \ldots, {\tau_k}^{-1}\boldsymbol{\Upsilon}_k\right)$.

As in the proof of Theorem 1, define $A : \mathcal{F}^k \to \mathcal{F}^{k-1}$ as the linear operator $A(V_1, \ldots, V_k) = (V_2 - V_1, \ldots, V_k - V_1)$. Then, $A\mathbf{V}_{k,n} \xrightarrow{D} \mathbf{W}$, where $\mathbf{W} = (W_1, \ldots, W_{k-1})^t = A\mathbf{V}$ is a Gaussian random element of $\mathcal{F}^{k-1}$ with mean $A\boldsymbol{\Delta}$ and covariance operator $A\widetilde{\boldsymbol{\Upsilon}}\,A^*$. Using that $\boldsymbol{\Delta}_1$ is the null operator, we obtain that $A\boldsymbol{\Delta} = (\boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_k) = \boldsymbol{\Delta}^{(k-1)}$. Moreover, from the proof of Theorem 1 we get that $A\widetilde{\boldsymbol{\Upsilon}}\,A^* = \boldsymbol{\Upsilon}_w$. Let $\upsilon_\ell \in \mathcal{F}^{k-1}$ be the orthonormal eigenfunctions of $\boldsymbol{\Upsilon}_w$ related to the eigenvalues θ_ℓ ordered in decreasing order. Since $\mathbf{W} - \boldsymbol{\Delta}^{(k-1)}$ is a zero mean Gaussian random element of $\mathcal{F}^{k-1}$ with covariance operator $\boldsymbol{\Upsilon}_w$, $\mathbf{W} - \boldsymbol{\Delta}^{(k-1)}$ can be written as $\sum_{\ell\geq 1} \theta_\ell^{1/2} Z_\ell\, \upsilon_\ell$ where Z_ℓ are i.i.d. random variables such that $Z_\ell \sim N(0, 1)$. On the other hand, we have the expansion $\boldsymbol{\Delta}^{(k-1)} = \sum_{\ell\geq 1} \eta_\ell\, \upsilon_\ell$, so that $\mathbf{W} = \sum_{\ell\geq 1}\left(\eta_\ell + \theta_\ell^{1/2} Z_\ell\right)\upsilon_\ell$ and $\|\mathbf{W}\|^2_{\mathcal{F}^{k-1}} = \sum_{\ell\geq 1}\left(\eta_\ell + \theta_\ell^{1/2} Z_\ell\right)^2 = \sum_{\ell\geq 1} \theta_\ell \left(\eta_\ell\, \theta_\ell^{-1/2} + Z_\ell\right)^2$, which concludes the proof since $T_{k,n} = n\sum_{j=2}^{k} \|(\widehat{\boldsymbol{\Gamma}}_j - \boldsymbol{\Gamma}_1) - (\widehat{\boldsymbol{\Gamma}}_1 - \boldsymbol{\Gamma}_1)\|^2_{\mathcal{F}} \xrightarrow{D} \sum_{j=1}^{k-1} \|W_j\|^2_{\mathcal{F}} = \|\mathbf{W}\|^2_{\mathcal{F}^{k-1}}$. □

Proof of Proposition 1. The results in Dauxois et al. (1982) entail that $\sqrt{n_1}\left(\widehat{\boldsymbol{\Gamma}}_1 - \boldsymbol{\Gamma}_1\right) \xrightarrow{D} \mathbf{U}_1$, where $\mathbf{U}_1$ a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_1$ so, we only have to prove the result for $i \geq 2$. We have the following decomposition for $\sqrt{n_i}(\widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1)$,

$$\sqrt{n_i}(\widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1) = \sqrt{n_i}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}(X_{i,j} - \overline{X}_i) \otimes (X_{i,j} - \overline{X}_i) - \boldsymbol{\Gamma}_1\right)$$
$$= \sqrt{n_i}(\widetilde{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1) + n^{-1/4}\sqrt{n_i}\,\widehat{\boldsymbol{\Gamma}}_{i,WR} + n^{-1/4}\sqrt{n_i}\,\widehat{\boldsymbol{\Gamma}}_{i,RW} + n^{-1/2}\sqrt{n_i}\,\widehat{\boldsymbol{\Delta}}_i$$

where

$$\widetilde{\boldsymbol{\Gamma}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(W_{i,j} - \overline{W}_i) \otimes (W_{i,j} - \overline{W}_i), \quad \widehat{\boldsymbol{\Delta}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(R_{i,j} - \overline{R}_i) \otimes (R_{i,j} - \overline{R}_i),$$

$$\widehat{\boldsymbol{\Gamma}}_{i,WR} = \frac{1}{n_i}\sum_{j=1}^{n_i}(W_{i,j} - \overline{W}_i) \otimes (R_{i,j} - \overline{R}_i) \quad \text{and}$$

$$\widehat{\boldsymbol{\Gamma}}_{i,RW} = \frac{1}{n_i}\sum_{j=1}^{n_i}(R_{i,j} - \overline{R}_i) \otimes (W_{i,j} - \overline{W}_i).$$

Using that $W_{i,j} \sim W_i$ and that the covariance operator of W_i is $\boldsymbol{\Gamma}_1$, from the results in Dauxois et al. (1982) we get that $\sqrt{n_i}\left(\widetilde{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_1\right) \xrightarrow{D} \mathbf{U}_i$, where $\mathbf{U}_i$ a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_i$ given in (11).

It is worth noticing that $\widehat{\boldsymbol{\Gamma}}_{i,WR}$ and $\widehat{\boldsymbol{\Gamma}}_{i,RW}$ are estimators of the cross covariance operators $\boldsymbol{\Gamma}_{i,WR} = \mathbb{E}\{(W_i - \mathbb{E}W_i) \otimes (R_i - \mathbb{E}R_i)\}$ and $\boldsymbol{\Gamma}_{i,RW} = \mathbb{E}\{(R_i - \mathbb{E}R_i) \otimes (W_i - \mathbb{E}W_i)\}$, respectively. The independence between W_i and R_i entails that $\boldsymbol{\Gamma}_{i,WR}$ is the null operator, which implies that $\sqrt{n_i}\,\widehat{\boldsymbol{\Gamma}}_{i,WR}$ is bounded in probability, so that $n^{-1/4}\sqrt{n_i}\,\widehat{\boldsymbol{\Gamma}}_{i,WR} \xrightarrow{p} 0$. Similarly, we obtain that $n^{-1/4}\sqrt{n_i}\,\widehat{\boldsymbol{\Gamma}}_{i,RW} \xrightarrow{p} 0$.

Finally, using the law of large numbers we have that $\widehat{\boldsymbol{\Delta}}_i$, the empirical covariance operator of R_i, converges in probability to $\boldsymbol{\Delta}_i$, so $n^{-1/2}\sqrt{n_i}\,\widehat{\boldsymbol{\Delta}}_i \xrightarrow{p} \tau_i^{1/2}\boldsymbol{\Delta}_i$, concluding the proof of a). □

Proof of Proposition 2. As in the proof of Proposition 1, we only have to prove the result for $i \geq 2$. Using the Karhunen–Loéve representation, we can write

$$X_{1,j} = \mu_1 + \sum_{\ell=1}^{\infty} \lambda_\ell^{\frac{1}{2}} f_{1\ell j}\,\phi_\ell, \quad 1 \leq j \leq n_1$$

$$X_{i,j} = \mu_i + \sum_{\ell=1}^{\infty} \lambda_\ell^{\frac{1}{2}} \left(1 + \frac{\Delta_{i,\ell}}{\sqrt{n}}\right)^{\frac{1}{2}} f_{i\ell j}\,\phi_\ell, \quad 1 \leq j \leq n_i, \quad 2 \leq i \leq k,$$

where $f_{i\ell j} \sim f_{i\ell}$ in (12). For $1 \leq j \leq n_i$, let $Z_{i,j} = \mu_i + \sum_{\ell=1}^{\infty} \lambda_\ell^{\frac{1}{2}} f_{i\ell j}\,\phi_\ell = \mu_i + Z_{0,i,j}$. Denote as

$$V_{i,j} = X_{i,j} - Z_{i,j} = \sum_{\ell=1}^{\infty} \lambda_\ell^{\frac{1}{2}} \left[\left(1 + \frac{\Delta_{i,\ell}}{\sqrt{n}}\right)^{\frac{1}{2}} - 1\right] f_{i\ell j}\,\phi_\ell .$$

Define the following operators that will be used in the sequel $\widetilde{\boldsymbol{\Gamma}}_i = (1/n_i) \sum_{j=1}^{n_i} \left(X_{i,j} - \mu_i\right) \otimes \left(X_{i,j} - \mu_i\right)$, $\widehat{\boldsymbol{\Gamma}}_{Z_0} = (1/n_i) \sum_{j=1}^{n_i} Z_{0,i,j} \otimes Z_{0,i,j}$, $\widehat{\boldsymbol{\Gamma}}_V = (1/n_i) \sum_{j=1}^{n_i} V_{i,j} \otimes V_{i,j}$ and finally, $\widetilde{\mathbf{A}} = (1/n_i) \sum_{j=1}^{n_i} (Z_{0,i,j} \otimes V_{i,j} + V_{i,j} \otimes Z_{0,i,j})$, where we avoid the index i for the sake of simplicity. Using that $X_{i,j} - \mu_i = Z_{0,i,j} + V_{i,j}$, we obtain the following expansion $\widetilde{\boldsymbol{\Gamma}}_i = \widehat{\boldsymbol{\Gamma}}_{Z_0} + \widehat{\boldsymbol{\Gamma}}_V + \widetilde{\mathbf{A}}$.

The proof will be carried out in several steps, by showing that

$$\sqrt{n_i}(\widehat{\boldsymbol{\Gamma}}_i - \widetilde{\boldsymbol{\Gamma}}_i) = o_{\mathbb{P}}(1) \tag{15}$$

$$\sqrt{n_i}\,\widehat{\boldsymbol{\Gamma}}_V = o_{\mathbb{P}}(1) \tag{16}$$

$$\sqrt{n_i}\,\widetilde{\mathbf{A}} \xrightarrow{p} \tau_i^{\frac{1}{2}} \boldsymbol{\Delta}_i \tag{17}$$

$$\sqrt{n_i}\,(\widehat{\boldsymbol{\Gamma}}_{Z_0} - \boldsymbol{\Gamma}_1) \xrightarrow{D} \mathbf{U}_i, \tag{18}$$

where $\mathbf{U}_i$ is a zero mean Gaussian random element with covariance operator $\boldsymbol{\Upsilon}_i$. Using that, for all $2 \le i \le k$, the covariance operator of $Z_{0,i,j}$ is $\boldsymbol{\Gamma}_1$, (18) follows from Dauxois et al. (1982).

We will derive (15). Noticing that $\widehat{\boldsymbol{\Gamma}}_i - \widetilde{\boldsymbol{\Gamma}}_i = -\left(\overline{X}_i - \mu_i\right) \otimes \left(\overline{X}_i - \mu_i\right)$, it will be enough to prove that $\sqrt{n_i}\left(\overline{X}_i - \mu_i\right) = \sqrt{n_i}\left(\overline{Z}_{0,i} + \overline{V}_i\right) = O_{\mathbb{P}}(1)$, with $\overline{Z}_{0,i} = (1/n_i) \sum_{j=1}^{n_i} Z_{0,i,j}$ and $\overline{V}_i = (1/n_i) \sum_{j=1}^{n_i} V_{i,j}$.

By the central limit theorem in Hilbert spaces, we get that $\sqrt{n_i}\,\overline{Z}_{0,i}$ converges in distribution, which entails that the process is tight, i.e., $\sqrt{n_i}\,\overline{Z}_{0,i} = O_{\mathbb{P}}(1)$.

We have that

$$\left(1 + \frac{\Delta_{i,\ell}}{\sqrt{n}}\right)^{\frac{1}{2}} - 1 = \frac{1}{\sqrt{n}} \frac{\Delta_{i,\ell}}{\left(1 + \frac{\Delta_{i,\ell}}{\sqrt{n}}\right)^{\frac{1}{2}} + 1} = a_{i,\ell,n} \frac{\Delta_{i,\ell}}{\sqrt{n}} \tag{19}$$

where $0 \le a_{i,\ell,n} \le 1$.

To derive that $\sqrt{n_i}\,\overline{V}_i = O_{\mathbb{P}}(1)$, we will further show that $\sqrt{n_i}\,\overline{V}_i = o_{\mathbb{P}}(1)$. To do so, notice that $\mathbb{E}\,\|\overline{V}_i\|^2 = (1/n_i) \sum_{\ell=1}^{\infty} \lambda_\ell \left[\left(1 + (\Delta_{i,\ell}/\sqrt{n})\right)^{\frac{1}{2}} - 1\right]^2$. Using (19), we get that $\mathbb{E}(\|\sqrt{n_i}\,\overline{V}_i\|^2) \le (1/n) \sum_{\ell=1}^{\infty} \lambda_\ell \Delta_{i,\ell}^2$, concluding the proof of (15).

To obtain (16), notice that (19) entails that $V_{i,j} \otimes V_{i,j} = (1/n) \sum_{\ell,s} \lambda_\ell^{\frac{1}{2}} \lambda_s^{\frac{1}{2}} a_{i,\ell,n} a_{i,s,n} \Delta_{i,s} \Delta_{i,\ell} f_{i\ell j} f_{isj} \phi_\ell \otimes \phi_s$,

so if we denote as $U_{\ell s} = (1/n_i) \sum_{j=1}^{n_i} f_{i\ell j}\ f_{isj}$, we get that

$$\widehat{\boldsymbol{\Gamma}}_V = \frac{1}{n} \sum_{\ell,s} \lambda_\ell^{\frac{1}{2}} \lambda_s^{\frac{1}{2}} a_{i,\ell,n}\, a_{i,s,n} \Delta_{i,s} \Delta_{i,\ell}\, U_{\ell s}\ \phi_\ell \otimes \phi_s\,.$$

Recall that $f_{i\ell j} \sim f_{i\ell}$ and $\mathbb{E}(f_{i\ell} f_{is}) = \delta_{\ell s}$, where $\delta_{\ell s} = 1$ if $\ell = s$ and 0 otherwise. Hence, we have that $\mathbb{E}(U_{\ell s}) = \delta_{\ell s}$ which implies that

$$\begin{aligned}\mathbb{E}(U_{\ell s}^2) &= Var(U_{\ell s}) + \mathbb{E}^2(U_{\ell s}) = \frac{1}{n_i} Var(f_{i\ell} f_{is}) + \delta_{\ell s} \\ &\leq \frac{1}{n_i}\mathbb{E}(f_{i\ell}^2 f_{is}^2) + \delta_{\ell s} \leq \frac{1}{n_i}\sigma_{4,i,\ell}\sigma_{4,i,s} + \delta_{\ell s},\end{aligned} \tag{20}$$

where the last bound follows from the Cauchy–Schwartz inequality and the fact that $\sigma_{4,i,s}^2 = \mathbb{E}(f_{is}^4)$. Hence, using (20) and the fact that $0 \leq a_{i,\ell,n} \leq 1$, we obtain the bound

$$\begin{aligned}\mathbb{E}(n_i \|\widehat{\boldsymbol{\Gamma}}_V\|_{\mathcal{F}}^2) &\leq \frac{n_i}{n^2}\sum_{\ell,s} \lambda_\ell \lambda_s \Delta_{i,s}^2 \Delta_{i,\ell}^2 \, \mathbb{E}(U_{\ell s}^2) \\ &\leq \frac{n_i}{n^2}\sum_{\ell,s} \lambda_\ell \lambda_s \, \Delta_{i,\ell}^2 \Delta_{i,s}^2 \left(\frac{1}{n_i}\sigma_{4,i,\ell}\sigma_{4,i,s} + \delta_{\ell s}\right) \\ &= \frac{1}{n^2}\left(\sum_\ell \lambda_\ell \, \Delta_{i,\ell}^2 \sigma_{4,i,\ell}\right)^2 + \frac{1}{n}\sum_\ell \lambda_\ell^2 \, \Delta_{i,\ell}^4 \, .\end{aligned}$$

Therefore, from the fact that $\sum_\ell \lambda_\ell^2 \, \Delta_{i,\ell}^4 \leq \left(\sum_\ell \lambda_\ell \, \Delta_{i,\ell}^2\right)^2 < \infty$ we get that $\mathbb{E}(n_i \|\widehat{\boldsymbol{\Gamma}}_V\|_{\mathcal{F}}^2) \to 0$, concluding the proof of (16).

Finally, to derive (17) we perform the decomposition

$$\widetilde{\mathbf{A}} = \frac{1}{\sqrt{n}}\sum_{\ell,s} \lambda_\ell^{\frac{1}{2}} \lambda_s^{\frac{1}{2}} a_{i,s,n} \Delta_{i,s} \, U_{\ell s} \; (\phi_\ell \; \otimes \phi_s + \phi_s \; \otimes \phi_\ell) = \widetilde{\mathbf{A}}_1 + \widetilde{\mathbf{A}}_2$$

where $U_{\ell s} = (1/n_i)\sum_{j=1}^{n_i} f_{i\ell j} \;\, f_{isj}$, as before. We will show that $\sqrt{n_i}\left(\widetilde{\mathbf{A}}_j - \mathbb{E}(\widetilde{\mathbf{A}}_j)\right) \xrightarrow{p} 0$, for $j = 1, 2$ which entails that $\sqrt{n_i}\left(\widetilde{\mathbf{A}} - \mathbb{E}(\widetilde{\mathbf{A}})\right) \xrightarrow{p} 0$. We will only prove that $\sqrt{n_i}\left(\widetilde{\mathbf{A}}_1 - \mathbb{E}(\widetilde{\mathbf{A}}_1)\right) \xrightarrow{p} 0$, since the other one follows similarly. From (20), we get that $Var(U_{\ell s}) \leq (1/n_i)\sigma_{4,i,\ell}\sigma_{4,i,s}$ which together with the fact that $0 \leq a_{i,\ell,n} \leq 1$ leads to

$$\begin{aligned}\mathbb{E}(n_i \|\widetilde{\mathbf{A}}_1 - \mathbb{E}(\widetilde{\mathbf{A}}_1)\|_{\mathcal{F}}^2) &= \frac{n_i}{n}\sum_{\ell,s} \lambda_\ell \lambda_s a_{i,s,n}^2 \Delta_{i,s}^2 Var(U_{\ell s}) \\ &\leq \frac{1}{n}\sum_{\ell,s} \lambda_\ell \lambda_s \Delta_{i,s}^2 \sigma_{4,i,\ell}\sigma_{4,i,s} \\ &= \frac{1}{\sqrt{n}}\left(\sum_\ell \lambda_\ell \sigma_{4,i,\ell}\right)\left(\sum_\ell \lambda_\ell \sigma_{4,i,\ell} \Delta_{i,\ell}^2\right)\end{aligned}$$

so that $\sqrt{n_i}\left(\widetilde{\mathbf{A}}_1 - \mathbb{E}(\widetilde{\mathbf{A}}_1)\right) \xrightarrow{p} 0$, as desired. Besides, using that $\mathbb{E}(U_{\ell s}) = \delta_{\ell s}$, we get that

$$\mathbb{E}(\sqrt{n_i}\,\widetilde{\mathbf{A}}) = \frac{2\sqrt{n_i}}{\sqrt{n}} \sum_{\ell} \lambda_\ell\, a_{i,\ell,n} \Delta_{i,\ell}\, \phi_\ell \otimes \phi_\ell \to \tau_i^{\frac{1}{2}} \sum_{\ell=1}^{\infty} \lambda_\ell \Delta_{i,\ell} \phi_\ell \otimes \phi_\ell = \tau_i^{\frac{1}{2}} \boldsymbol{\Delta}_i$$

where we have used that $a_{i,\ell,n} \to 1/2$, as $n \to \infty$ and $\sum_{\ell=1}^{\infty} \lambda_\ell |\Delta_{i,\ell}| < \infty$. This concludes the proof of (17). The proof of Proposition 2a) follows now combining (15) to (18). □

Proof of Theorem 3. Recall that $\widetilde{X}_n = (X_{1,1}, \ldots, X_{1,n_1}, \ldots, X_{k,1}, \ldots, X_{k,n_k})$. Let $\widetilde{Z}_n = (Z_1, \ldots, Z_{q_n})$ and $\widetilde{Z} = \{Z_\ell\}_{\ell \geq 1}$ with $Z_i \sim N(0,1)$ independent. Define $\widehat{\mathcal{U}}_n(\widetilde{X}_n, \widetilde{Z}_n) = \sum_{\ell=1}^{q_n} \widehat{\theta}_\ell Z_\ell^2$, $\mathcal{U}_n(\widetilde{Z}_n) = \sum_{\ell=1}^{q_n} \theta_\ell Z_\ell^2$ and $\mathcal{U}(\widetilde{Z}) = \sum_{\ell=1}^{\infty} \theta_\ell Z_\ell^2$. It is worth noticing that $\widehat{\mathcal{U}}_n$ has the same distribution as $\mathcal{U}_n^*$.

First notice that, for any ℓ, $|\widehat{\theta}_\ell - \theta_\ell| \leq \|\widehat{\boldsymbol{\Upsilon}}_w - \boldsymbol{\Upsilon}_w\|_{\mathcal{G}^{k-1}}$ (see, for instance, Kato (1966)), which implies that

$$\sum_{\ell=1}^{q_n} |\widehat{\theta}_\ell - \theta_\ell| \leq \frac{q_n}{\sqrt{n}} \sqrt{n} \|\widehat{\boldsymbol{\Upsilon}}_w - \boldsymbol{\Upsilon}_w\|_{\mathcal{G}^{k-1}}\,. \tag{21}$$

On the other hand, we have

$$\mathbb{E}\left[|\widehat{\mathcal{U}}_n - \mathcal{U}||\widetilde{X}_n\right] = \mathbb{E}\left[|\widehat{\mathcal{U}}_n - \mathcal{U}_n + \mathcal{U}_n - \mathcal{U}|\;|\tilde{X}_n\right] \leq \sum_{\ell=1}^{q_n} |\widehat{\theta}_\ell - \theta_\ell| + \sum_{\ell > q_n} \theta_\ell$$

which together with (21), the fact that $\sqrt{n}\|\widehat{\boldsymbol{\Upsilon}}_w - \boldsymbol{\Upsilon}_w\| = O_{\mathbb{P}}(1)$, $q_n/\sqrt{n} \to 0$ and $\sum_{\ell \geq 1} \theta_\ell < \infty$ implies that

$$\mathbb{E}\left[|\widehat{\mathcal{U}}_n - \mathcal{U}|\;|\widetilde{X}_n\right] \xrightarrow{p} 0\,. \tag{22}$$

We also have the following inequalities

$$\begin{aligned}
\mathbb{P}(\widehat{\mathcal{U}}_n \leq t|\widetilde{X}_n) &= \mathbb{P}(\widehat{\mathcal{U}}_n \leq t \cap |\widehat{\mathcal{U}}_n - \mathcal{U}| < \epsilon\;|\widetilde{X}_n) + \mathbb{P}(\widehat{\mathcal{U}}_n \leq t \cap |\widehat{\mathcal{U}}_n - \mathcal{U}| > \epsilon\;|\widetilde{X}_n) \\
&\leq \mathbb{P}(\mathcal{U} \leq t + \epsilon) + \mathbb{P}(|\widehat{\mathcal{U}}_n - \mathcal{U}| > \epsilon\;|\widetilde{X}_n) \\
&\leq F_{\mathcal{U}}(t+\epsilon) + \frac{1}{\epsilon}\mathbb{E}(|\widehat{\mathcal{U}}_n - \mathcal{U}|\;|\widetilde{X}_n) \leq F_{\mathcal{U}}(t) + \Delta_\epsilon(t) + \frac{1}{\epsilon}\mathbb{E}(|\widehat{\mathcal{U}}_n - \mathcal{U}|\;|\widetilde{X}_n),
\end{aligned}$$

where $\Delta_\epsilon(t) = \sup_{|\delta| \leq \epsilon} |F_{\mathcal{U}}(t+\delta) - F_{\mathcal{U}}(t)|$. Besides,

$$\begin{aligned}
\mathbb{P}(\widehat{\mathcal{U}}_n \leq t\;|\widetilde{X}_n) &= \mathbb{P}(\widehat{\mathcal{U}}_n \leq t \cap |\widehat{\mathcal{U}}_n - \mathcal{U}| < \epsilon\;|\widetilde{X}_n) + \mathbb{P}(\widehat{\mathcal{U}}_n \leq t \cap |\widehat{\mathcal{U}}_n - \mathcal{U}| > \epsilon\;|\widetilde{X}_n) \\
&\geq \mathbb{P}(\mathcal{U} \leq t - \epsilon \cap |\widehat{\mathcal{U}}_n - \mathcal{U}| < \epsilon\;|\widetilde{X}_n) \\
&\geq F_{\mathcal{U}}(t-\epsilon) - \frac{1}{\epsilon}\mathbb{E}(|\widehat{\mathcal{U}}_n - \mathcal{U}||\widetilde{X}_n) \geq F_{\mathcal{U}}(t) - \Delta_\epsilon(t) - \frac{1}{\epsilon}\mathbb{E}(|\widehat{\mathcal{U}}_n - \mathcal{U}||\widetilde{X}_n)\,.
\end{aligned}$$

Therefore,

$$|\mathbb{P}(\widehat{\mathcal{U}}_n \le t \,|\widetilde{X}_n) - F_{\mathcal{U}}(t)| \le \Delta_\epsilon(t) + \frac{1}{\epsilon}\mathbb{E}(|\widehat{\mathcal{U}}_n - \mathcal{U}| \,|\widetilde{X}_n)\;.$$

As we mentioned in Remark 1, $F_{\mathcal{U}}$ is a continuous distribution function on $\mathbb{R}$ and so uniformly continuous; hence, $\lim_{\epsilon\to 0}\,\sup_{t\in\mathbb{R}}\;\Delta_\epsilon(t) = 0$, which together with (22) implies that $\rho_k(F_{\mathcal{U}_n^*|\widetilde{X}_n}, F_{\mathcal{U}}) = \sup_t |\mathbb{P}(\widehat{\mathcal{U}}_n \le t \,|\tilde{X}_n) - F_{\mathcal{U}}(t)| \xrightarrow{p} 0.$ □

References

Benjamini, Y., Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*, 1165–1188.

Benko, M., Härdle, P., Kneip, A. (2009). Common functional principal components. *Annals of Statistics*, *37*, 1–34.

Boente, G., Fraiman, R. (2000). Kernel-based functional principal components. *Statistics and Probability Letters*, *48*, 335–345.

Boente, G., Rodriguez, D., Sued, M. (2010). Inference under functional proportional and common principal components models. *Journal of Multivariate Analysis*, *101*, 464–475.

Boente, G., Rodriguez, D., Sued, M. (2014). Testing equality between several covariance operators. Preprint available at https://arxiv.org/pdf/1404.7080v2.pdf

Cabassi, A., Pigoli, D., Secchi, P., Carter, P. A. (2017). Permutation tests for the equality of covariance operators of functional data with applications to evolutionary biology. Preprint available at https://arxiv.org/pdf/1701.05870.pdf

Chang, C., Ogden, R. T. (2009). Bootstrapping sums of independent but not identically distributed continuous processes with applications to functional data. *Journal of Multivariate Analysis*, *100*, 1291–1303.

Coffey, N., Harrison, A. J., Donaghue, O. A., Hayes, K. (2011). Common functional principal components analysis: A new approach to analyzing human movement data. *Human Movement Science*, *30*, 1144–1166.

Cuesta-Albertos, J., Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data. *Test*, *19*, 537–557.

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, *147*, 1–23.

Dauxois, J., Pousse, A., Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, *12*, 136–154.

Donoghue, O., Harrison, A. J., Coffey, N., Hayes, K. (2008). Functional data analysis of running kinematics in chronic Achilles tendon injury. *Medicine and Science in Sports and Exercise*, *40*, 1323–1335.

Ferraty, F., Romain, Y. (2010). *The oxford handbook of functional data analysis*. Oxford: Oxford University Press.

Ferraty, F., Vieu, Ph. (2006). *Nonparametric functional data analysis*: *Theory and practice*. Springer Series in Statistics New York: Springer.

Ferraty, F., Vieu, Ph, Viguier-Pla, S. (2007). Factor-based comparison of groups of curves. *Computational Statistics and Data Analysis*, *51*, 4903–4910.

Ferraty, F., Van Keilegom, I., Vieu, Ph. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics*, *37*, 286–306.

Ferraty, F., Van Keilegom, I., Vieu, Ph. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, *109*, 10–28.

Fremdt, S., Steinebach, J. G., Horváth, L., Kokoszka, P. (2013). Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, *40*, 138–52.

Gaines, G., Kaphle, K., Ruymgaart, F. (2011). Application of a delta-method for random operators to testing equality of two covariance operators. *Mathematical Methods of Statistics*, *20*, 232–245.

Goia, A., Vieu, P. (2016). An introduction to recent advances in high/infinite-dimensional statistics. *Journal of Multivariate Analysis*, *146*, 1–6.

Gupta, A., Xu, J. (2006). On some tests of the covariance matrix under general conditions. *Annals of the Institute of Statistical Mathematics*, *58*, 101–114.

Hastie, T., Buja, A., Tibshirani, R. (1995). Penalized discriminant analysis. *Annals of Statistics*, *23*(1), 73–102.

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

Hsing, T., Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. New York: Wiley.

Kato, T. (1966). *Perturbation theory for linear operators*. New York: Springer.

Morales, D., Pardo, L., Pardo, M. C., Vajda, I. (2004). Rényi statistics for testing composite hypotheses in general exponential models. *Statistics*, *38*, 133–147.

Neuhaus, G. (1980). A note on computing the distribution of the norm of Hilbert space valued Gaussian random variables. *Journal of Multivariate Analysis*, *10*, 19–25.

Panaretos, V. M., Kraus, D., Maddocks, J. H. (2010). Second-Order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles. *Journal of the American Statistical Association*, *105*, 670–682.

Paparoditis, E., Sapatinas, T. (2016). Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika*, *103*, 727–733.

Pigoli, D., Aston, J. A., Dryden, I., Secchi, P. (2014). Distances and inference for covariance operators. *Biometrika*, *101*, 409–422.

Ramsay, J. O., Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.

Raña, P., Aneiros, G., Vilar, J. (2016). Bootstrap confidence intervals in functional nonparametric regression under dependence. *Electronic Journal of Statistics*, *10*, 1973–1999.

Silverman, B. W. (1986). *Density estimation*. London: Chapman and Hall.

Van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge: Cambridge University Press.